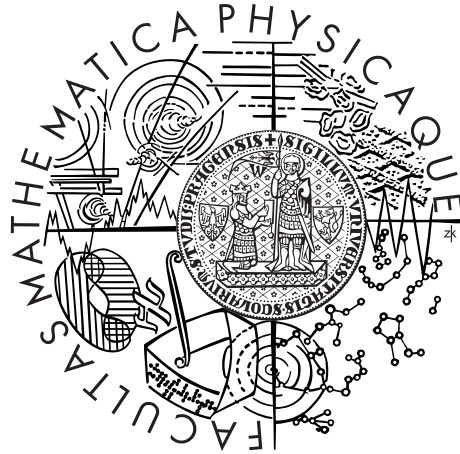


Charles University in Prague  
Faculty of Mathematics and Physics

## HABILITATION THESIS



Viktor Holubec

### Řízení pohyblivé aktivní hmoty: dynamika a energetika

### Steering motile active matter: dynamics and energetics

Department of Macromolecular Physics

Specialization: Theoretical Physics

Prague 2023

**Acknowledgements:** My deepest gratitude goes to my family, who allowed me to finish this work and tolerated the fact that it sometimes took away too much of our shared time. Turning to my mentors and collaborators, first and foremost, I would like to thank my Ph.D. supervisor, Prof. RNDr. Petr Chvosta CSc., for preparing me for international research in physics and imparting most of the theory tricks I know. Secondly, I express my gratitude to Prof. Dr. Klaus Kroy and Prof. Dr. Frank Cichos for allowing me to join their research groups as a postdoctoral researcher, introducing me to the field of active matter, and providing opportunities to contribute to their research projects and collaborate with their numerous Ph.D. and undergraduate students. Additionally, I would like to thank all the former and current doctoral students who contributed to this work, with special recognition given to Dr. Sven Auschra and Dr. Daniel Geiss. They not only assisted me with physics but also helped me settle into life as a postdoc in Germany. Finally, I extend my appreciation to my numerous coworkers, particularly RNDr. Artem Ryabov, Ph.D., for their invaluable help and collaboration throughout the years.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Active matter engines</b>	<b>3</b>
2.1	Active heat engines (Refs. <sup>51–56</sup> ) . . . . .	5
2.1.1	Effective temperature in overdamped active heat engines (Ref. <sup>51</sup> ) .	6
2.1.2	Effective temperature in underdamped active heat engines (Ref. <sup>56</sup> )	8
2.1.3	Results (in)valid when effective temperature exists (Refs. <sup>52,55</sup> ) . .	9
2.2	General results (Refs. <sup>53,54</sup> ): . . . . .	10
2.2.1	Quasi-static efficiency at finite power (Ref. <sup>53</sup> ) . . . . .	10
2.2.2	Maximum efficiency protocol for constrained driving (Ref. <sup>54</sup> ) . . .	11
2.3	Active ratchets (Refs. <sup>50,59–61</sup> ) . . . . .	12
2.3.1	Active Brownian particles in activity landscapes (Refs. <sup>59–61</sup> ) . . . .	13
2.3.2	Activity ratchet (Ref. <sup>50</sup> ) . . . . .	14
<b>3</b>	<b>Effects of time delay</b>	<b>17</b>
3.1	Equilibrium delay (Ref. <sup>88</sup> ) . . . . .	19
3.2	Feedback driven active Brownian particles (Ref. <sup>2,72,85,90,94</sup> ) . . . . .	20
3.2.1	Active Brownian molecules (Ref. <sup>85,90</sup> ) . . . . .	22
3.2.2	Delay-induced chirality in systems of micro-swimmers (Ref. <sup>72,94</sup> ) .	24
3.2.3	Machine learning with micro-swimmers (Ref. <sup>2</sup> ) . . . . .	26
3.3	Delay Vicsek model (Ref. <sup>80–82</sup> ) . . . . .	28
3.3.1	Finite-size scaling in delay Vicsek model (Ref. <sup>80</sup> ) . . . . .	30
3.3.2	Information propagation in delay Vicsek model (Ref. <sup>81,82</sup> ) . . . . .	31
<b>4</b>	<b>Final Remarks</b>	<b>35</b>
<b>5</b>	<b>References</b>	<b>37</b>
<b>6</b>	<b>List of discussed Papers</b>	<b>49</b>

<b>7</b>	<b>Original Papers</b>	<b>51</b>
7.1	Holubec V., Steffenoni S., Falasco G., and Klaus K., Active Brownian heat engines, PRR, 2, 043262 (2020). . . . .	53
7.2	Holubec V. and Rahul M., Underdamped active Brownian heat engine, PRE, 102, 060101(R) (2020). . . . .	77
7.3	Holubec V. and Ryabov A., Maximum efficiency of low-dissipation heat engines at arbitrary power, J. Stat. Mech. 073204 (2016). . . . .	83
7.4	Holubec V., Klaus K., and Steffenoni S., Physically consistent numerical solver for time-dependent Fokker-Planck equations, PRE, 99, 032117 (2019). . . . .	99
7.5	Holubec V. and Ryabov A., Cycling Tames Power Fluctuations near Optimum Efficiency, PRL, 121, 120601 (2018). . . . .	117
7.6	Ye Z., Cerisola F., Abiuso P., Anders J., Perarnau-Llobet M., and Holubec V., Optimal finite-time heat engines under constrained control, PRR, 4, 043130 (2022). . . . .	127
7.7	Söker N. A., Auschra S., Holubec V., Kroy K., and Cichos F., How Activity Landscapes Polarize Microswimmers without Alignment Forces, PRL, 126, 228001 (2021). . . . .	139
7.8	Auschra S., Holubec V., Söker N. A., Kroy K., and Cichos F., Polarization-density patterns of active particles in motility gradients, PRE, 103, 062601 (2021). . . . .	149
7.9	Auschra S. and Holubec V., Density and polarization of active Brownian particles in curved activity landscapes, PRE, 103, 062604 (2021). . . . .	169
7.10	Rein C., Kolář M., Kroy K. and Holubec V., Force-Free and Autonomous Active Brownian Ratchets, EPL, 142, 31001 (2023). . . . .	179
7.11	Holubec V., Ryabov A., Loos S. A. M., and Kroy K., Equilibrium stochastic delay processes, NJP, 24 023021 (2022). . . . .	187
7.12	Khadka U., Holubec V. Yang H., and Cichos F., Active particles bound by information flows, Nat. Commun, 9, 3864 (2018). . . . .	209
7.13	Geiss D., Kroy K., and Holubec V., Brownian molecules formed by delayed harmonic interactions, NJP, 21 093014 (2019). . . . .	231
7.14	Wang X., Chen P., Kroy K., Holubec V., and Cichos F., Spontaneous vortex formation by microswimmers with retarded attractions, Nat. Commun, 14, 56 (2023). . . . .	261
7.15	Chen P., Kroy K., Cichos F., Wang X. and Holubec V., Active particles with delayed attractions form quaking crystallites, EPL, 000 (2023). . . . .	287

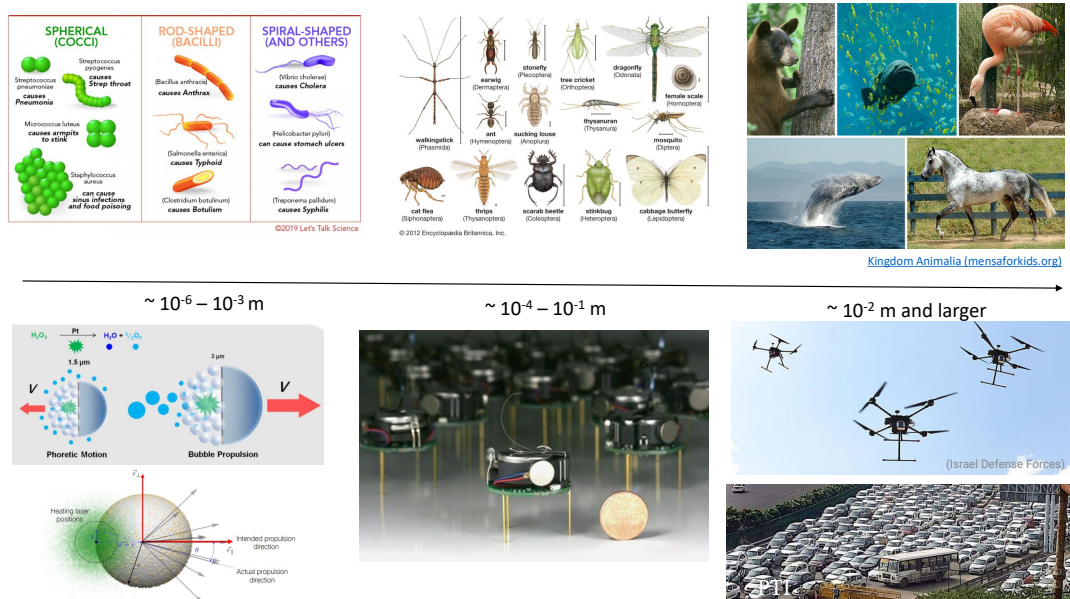
7.16	Muiños-Landin S., Fischer A., Holubec V., and Cichos F., Reinforcement learning with artificial microswimmers, <i>Sci. Rob.</i> , 6, eabd9285 (2021). . .	297
7.17	Holubec V., Geiss D., Loos S. A. M., Kroy K., and Cichos F., Finite-Size Scaling at the Edge of Disorder in a Time-Delay Vicsek Model, <i>PRL</i> , 127, 258001 (2021). . . . .	313
7.18	Geiss D., Kroy K., and Holubec V., Information conduction and convection in noiseless Vicsek flocks, <i>PRE</i> , 106, 014609 (2022). . . . .	335
7.19	Geiss D., Kroy K., and Holubec V., Signal propagation and linear response in the delay Vicsek model, <i>PRE</i> , 106, 054612 (2022). . . . .	345



# 1 Introduction

Motile active matter is a vibrant multidisciplinary field that brings together physicists, biologists, and even social engineers. It uses tools from theoretical and experimental physics to understand the dynamics of self-propelling particles in various environments, interactions among them, and emergent behaviors in their large assemblies<sup>4-7</sup>. As shown in FIGURE 1.1, systems of interest range from self-propelled colloids<sup>8-10</sup>, over motile cells, filaments, tissues and bacteria<sup>11,12</sup>, flocking insects<sup>13-15</sup> and birds<sup>16</sup>, and schools of fish<sup>17</sup> to the coordinate motion of ants<sup>18</sup> and the crowding of pedestrians<sup>19</sup>.

Active matter systems at all scales share three characteristic features. First, they are driven out of equilibrium on the level of single particles, which irreversibly transform some fuel into a directed motion. The nonequilibrium state is thus sustained by the



**Figure 1.1: Motile active matter.** Examples of natural (top) and artificial (bottom) active matter systems across length scales. Except for the chemically propelled Janus particles<sup>1</sup>, the optically steered symmetric active particles<sup>2</sup> (both bottom left), and the ‘vibrobots’<sup>3</sup> (bottom middle), the sources are given inside the individual figures.

inflow of the fuel or food into the system, rather than, e.g., heating and cooling the walls as in a boundary-driven nonequilibrium system. The second important ingredient of active matter is that its effective dynamics can disobey standard thermodynamic limitations, such as the fluctuation-dissipation theorem, or even more universal symmetries, such as reciprocity. This is because the ‘social’ or ‘feedback’ interactions result from a complicated coarsegraining of the microscopic degrees of freedom far from thermodynamic equilibrium. The third characteristic feature of active matter is that the interactions often involve a time delay. Intuitively, these delays result from limited speeds of information transfer between and inside the individuals, decision-making, and body transformation of the individuals. Mathematically, the delays derive from the coarsegraining of time-local dynamics of the microscopic degrees of freedom.

The ultimate goal of the field of active matter is to provide an understanding of evolutionary mechanisms which lead to the variety of behaviors observed in nature. A technical part of this task is to theoretically describe behaviors observed in nature by developing suitable generalizations of the tools of equilibrium statistical physics. Another more practical objective is to create well-controlled (not necessarily artificial) counterparts of natural active particles, able to, e.g., perform medical tasks on the level of individual cells<sup>20</sup> or to form distributed collectively communicating sensorial networks on the macroscale<sup>21</sup>.

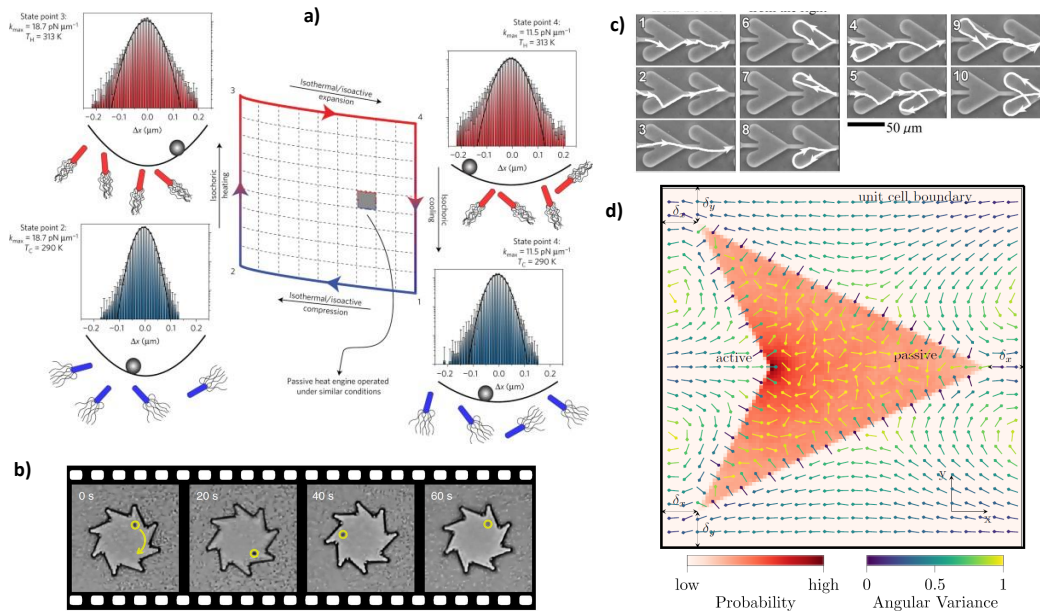
This habilitation thesis is divided into two major parts summarizing the author’s contributions to understanding the dynamics and energetics (or thermodynamics) of active matter. CHAPTER 2 investigates how to utilize the activity of single active particles or their ensembles to perform useful work or induce transport. CHAPTER 3 is devoted to the study of the effects of time-delayed interactions in active matter systems. Both these chapters are conceived as overviews of the corresponding parts of active matter research with a summary of the author’s contributions, reprinted in the CHAPTER 7 of this thesis in the same order in which they appear in the text.

Most of the papers discussed in the thesis were written or conceived during the post-doctoral stay of the author in the group of Prof. Klaus Kroy at Uni. Leipzig. Therefore, up to a few exceptions, the presented work aims to describe overdamped active particles, such as bacteria or driven colloids, which are investigated experimentally in the group of Prof. Frank Cichos from Uni. Leipzig. The thesis contains only works where the author’s contribution was significant. With a single exception, it does not contain the authors’ contributions to the study of noise-induced coherence<sup>22,23</sup>, maximum efficiency at fixed power<sup>24–28</sup>, unstable stochastic systems<sup>29–32</sup>, classical Brownian ratchets<sup>29–32</sup>, optimal driving of stochastic heat engines<sup>33,34</sup>, and work fluctuations in small systems<sup>35–46</sup>.



## 2 Active matter engines

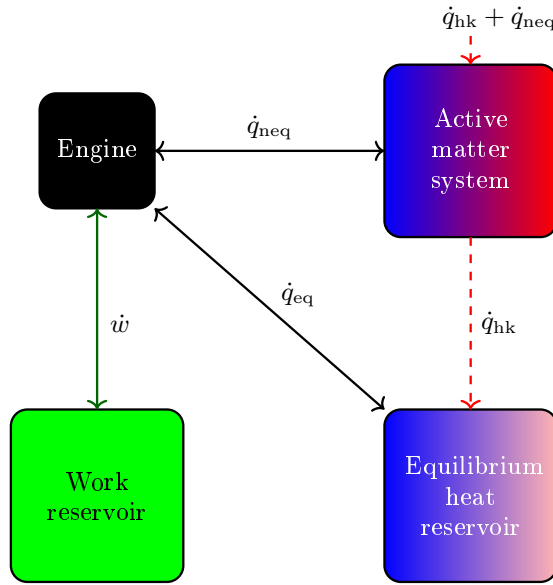
Microscopic active particles such as artificial active colloids or bacteria have been employed to perform useful work in two conceptually different ways. The first one, exemplified in FIGURE 2.1a, aims to treat the system of active particles as a non-equilibrium heat bath and to transform the disordered energy from this bath into useful work via so-called active Brownian heat engines<sup>47,51</sup>. The second approach, depicted in FIGURE 2.1b-d, aims to harvest the energy of the active motion more directly by rectifying the



**Figure 2.1: Extracting energy from active matter.** Panel a) shows a colloidal particle confined by a harmonic potential in an active bath composed of living bacteria in water<sup>47</sup>. In this setup, energy is extracted from the active bath by varying in time the bath's activity (e.g., by reducing food content in the solvent) and the stiffness of the potential. The remaining panels show various ways to rectify (or directionalize) erratic motion of bacteria. In panel b), the bacteria are trapped between a cog wheel's asymmetric tooth to rotate it<sup>48</sup>. In panel c), a similar asymmetry of channel walls induces a directed (average) motion of bacteria<sup>49</sup>. Panel d) shows how to create a likewise directed motion of active particles by making the particle speed position-dependent instead of using potentials or walls<sup>50</sup>.

direction of self-propulsion of otherwise randomly turning active Brownian particles via obstacles<sup>48,49</sup> or other ratchet-like mechanisms<sup>50</sup>.

Of course, in both these cases, the energy of the autonomous motion of the active particles is transformed into work. Based on the mechanism underlying the autonomous motion, this energy conversion can be identified as heat-to-work conversion (for thermophoretically propelled swimmers) or as work-to-work conversion (for chemically propelled swimmers). The conceptual distinction between active heat engines and other engines is thus motivated rather theoretically than practically. The energy extracted from active matter systems can be identified as heat only in special cases when one can identify an effective temperature of an equilibrium bath that would yield the same engine's performance, which allows attributing the energy flux from the bath with a valid (second law) entropy production. The effective temperature then replaces thermodynamic temperature in standard upper bounds on the engine performance, such as the Carnot's efficiency, which still limit active heat engines' performance. When the effective



**Figure 2.2: Energy fluxes during energy extraction from active matter.** An engine transforming the heat flux  $\dot{q} = \dot{q}_{neq} + \dot{q}_{eq}$  from a non-equilibrium active matter system (neq) and perhaps also an equilibrium (eq) heat reservoir into usable power  $\dot{w}$ . The corresponding energy fluxes relevant for the engine's operation are depicted by arrows. The dashed arrow depicts the housekeeping heat flux,  $\dot{q}_{hk}$ , flowing from the active bath to the infinite equilibrium reservoir, which prevents the active bath from overheating. This energy flux and also  $\dot{q}_{neq}$  are sustained by the energy influx  $\dot{q}_{hk} + \dot{q}_{neq}$  into the non-equilibrium bath, which keeps it in a non-equilibrium "active" steady state. Template for the figure was taken from Ref.<sup>51</sup> (PUBLICATION 7.1).

temperature does not exist, the active engines' efficiency is limited only by the trivial first law bound on the efficiency of work-to-work conversion, i.e., by one.

FIGURE 2.2 shows a diagram of energy fluxes involved in any energy extraction from an active matter system in a steady state. To stay active or, in other words, alive, the active matter system consumes per unit time the amount of energy  $\dot{q}_{hk} + \dot{q}_{neq}$ . If no energy is extracted from the active matter system, all this power has to be dissipated in a heat reservoir (or heat sink). Otherwise, the active matter system would overheat. Assuming that the power  $\dot{q}_{neq}$  is extracted from the active matter system, the power delivered to the heat sink is  $\dot{q}_{hk}$ . The energy transferred into the engine from the active matter system is thus  $\dot{q}_{neq}$  and, in its operation, the engine in general also dissipates a heat flux  $\dot{q}_{eq}$  into the equilibrium reservoir. In majority of active matter engines, the energy influx  $\dot{q}_{neq}$  depends just on the dynamics of the individual active particles, not on the type of their self-propelling mechanism. In particular,  $\dot{q}_{neq}$  is usually independent of the efficiency with which the engines in the individual active particles transform the overall energy influx  $\dot{q}_{hk} + \dot{q}_{neq}$  into their activity. Hence, it is reasonable to characterise the engine performance by the efficiency  $\eta = \dot{w}/\dot{q}_{neq}$  of conversion of  $\dot{q}_{neq}$  into the power  $\dot{w}$ , rather than the overall efficiency  $\dot{w}/(\dot{q}_{neq} + \dot{q}_{hk})$  of the engine and active matter system. The latter efficiency is strongly system-dependent and usually very close to 0. For an active heat engine,  $\eta$  is in general limited by the Carnot's efficiency  $\eta_C = 1 - T_c^{eff}/T_h^{eff}$  with largest and smallest values of the effective temperature  $T_c^{eff}$  and  $T_h^{eff}$  experienced by the engine. For an active engine, where an effective temperature cannot be defined,  $\eta < 1$ .

In the rest of this section, I first briefly introduce periodically driven active heat engines and review our results for them in SECTION 2.1. Next, in SECTION 2.2, I highlight some of our general results, which hold both for active and standard periodically driven (heat) engines. Finally, in SECTION 2.3, I present our results on active ratchets that autonomously rectify the motion of active particles.

## 2.1 Active heat engines (Refs.<sup>51–56</sup>)

As described above, the main theoretical difficulty in deciding whether an active engine can be treated as an active heat engine, and thus one can assess its performance using results valid for heat engines in contact with equilibrium heat reservoirs, is to determine if an effective temperature can describe the active bath. Motivated by the experimental realization of the 'Bacteria heat engine'<sup>47</sup>, depicted in FIGURE 2.1a, and the claims made in this work that its efficiency can surpass the second law upper bound on the

efficiency of the corresponding thermodynamic (Stirling) cycle, which is conceptually an erroneous statement as that would mean that the engine investigated in this work is not a heat engine \*, my colleagues and I wrote two papers<sup>51,56</sup> studying when the effective temperature in general exists.

### 2.1.1 Effective temperature in overdamped active heat engines (Ref.<sup>51</sup>)

In PUBLICATION 7.1<sup>51</sup> we have shown that effective temperature in general exists for engines described by Hamiltonian of the form

$$H = k(t)f(\mathbf{x}), \quad (2.1)$$

where  $k(t)$  is an externally controlled parameter periodically varied in time,  $\mathbf{x}$  denotes degrees of freedom of the engine, and  $f$  stands for a confining potential (such that the equilibrium partition function  $\int d\mathbf{x} \exp(-\beta H)$  is finite for any positive inverse temperature  $\beta$ ). The Hamiltonian describes the engine part of the compound bath-engine system; the existence of the effective temperature is, thus, in this case, independent of the details of the bath and the bath-engine coupling. Furthermore, this result is valid regardless of the details of the dynamics, which can thus be arbitrary, including non-Markovian, quantum, or other dynamics (even though the calculation of the effective temperature might sometimes be challenging). It thus proves that the efficiency analysis presented in Ref.<sup>47</sup> is unavoidably wrong not only conceptually but also numerically as the corresponding Hamiltonian  $H = k(t)(x^2 + y^2)$  is of the form (2.1) and thus there certainly exists an effective temperature that allows limiting the efficiency of the engine below the ultimate second law bound. This illustrates how our result can serve as a simple sanity check of measured or calculated efficiencies of active heat engines.

To understand why an effective temperature can be always found for Hamiltonians of the form (2.1) but not for more general ones, e.g., containing also a kinetic energy  $m\mathbf{p}^2/2$ , it is enough to write down expressions for average heat and work fluxes

$$\dot{q}(t) = k(t)\dot{\sigma}_x(t) + m\dot{\sigma}_p(t) \quad (2.2)$$

$$\dot{w} = \dot{k}(t)\sigma_x(t), \quad (2.3)$$

where  $\sigma_x(t) = \langle f(\mathbf{x}) \rangle$  and  $\sigma_p(t) = \langle \mathbf{p}^2 \rangle / 2$  ( $\langle \bullet \rangle$  denotes the ensemble average). These

---

\*Similar claims of beating second law efficiencies by using non-equilibrium ‘heat’ reservoirs (such as various quantum squeezed baths<sup>57</sup>) fall into the very same category, pointing to authors’ desire to sell their research well in high-impact journals rather than deep physics break troughs.

expressions follow from the first law<sup>40</sup> by identifying the changes of internal energy  $U = \langle H \rangle = \langle k(t)f(\mathbf{x}) + m\mathbf{p}^2/2 \rangle$  of the system related to the variation of the control parameter  $k(t)$  as work, and the rest of  $\dot{U}$  as heat. The work and heat fluxes are thus determined by the ‘response’ functions  $\sigma_x(t)$  and  $\sigma_p(t)$ . A non-equilibrium bath can be prescribed an effective temperature  $T_{eff}(t)$  if there is an equilibrium setup with the same time-dependent Hamiltonian  $H$  and equilibrium heat bath at a time-dependent temperature  $T_{eff}(t)$ , which yield the same heat and work fluxes  $\dot{w}$  and  $\dot{q}$  (and thus the response functions  $\sigma_x(t)$  and  $\sigma_p(t)$ ) as the setup with non-equilibrium bath. In general, the response functions are functionals of the driving parameters determined by details of the engine and bath dynamics, and the time-dependent effective temperature has to be such that the functionals for the two setups agree numerically. It is reasonable to assume that finding such a mapping should always be possible when one needs to match a single functional, e.g.,  $\sigma_x(t)$ . However, matching two or more functionals (such as when considering momentum degrees of freedom) by modifying the single effective temperature might not always be possible.

The simplest and most important situation demonstrating this conclusion is quasi-static driving. Then the distribution for  $\{\mathbf{x}, \mathbf{p}\}$  at any time  $t$  has the Boltzmann form  $p(\mathbf{x}, \mathbf{p}, t) = \frac{1}{Z} \exp\left(-\frac{k(t)f(\mathbf{x})}{k_B T_{eff}}\right) \exp\left(-\frac{m\mathbf{p}^2}{2k_B T_{eff}}\right)$  ( $Z$  stands for partition function and  $k_B$  the Boltzmann constant). The averages  $\sigma_x(t)$  and  $\sigma_p(t)$  then follow as integrals over  $p(\mathbf{x}, \mathbf{p}, t)$ . It is always possible to tune the effective temperature  $T_{eff}$  to match any given value of one of these averages. However, the resulting  $T_{eff}$  also automatically determines the other average. Thus, it is generally impossible to find an equilibrium setup that would match an arbitrary couple  $\sigma_x(t)$  and  $\sigma_p(t)$  resulting from the dynamics with a non-equilibrium bath that defies restrictions imposed by equilibrium dynamics, and similarly for more complicated settings. For an analysis when  $T_{eff}$  exists in settings with non-negligible momentum, see SUBSECTION 2.1.2.

In PUBLICATION 7.1<sup>51</sup> we have also shown how to calculate the time dependent effective temperature for the specific dynamics considered in Ref.<sup>47</sup> with arbitrary cycle duration and arbitrary protocols for the potential stiffness  $k(t)$  and parameters of the active bath. Concretely, we considered dynamics described by Langevin equation

$$\dot{x}(t) = -k(t)x(t)/\gamma + \eta(t), \quad (2.4)$$

where  $\gamma$  is friction coefficient and  $\eta$  is a zero-mean noise describing effects of the bacteria bath. For an equilibrium bath, the noise correlation function has to obey the fluctuation dissipation relation and thus it reads  $\langle \eta(t)\eta(t') \rangle = (2k_B T_{eff}(t)/\gamma)\delta(t-t')$ . The relevant

response function  $\sigma_x(t) = \langle x^2 \rangle$  obeys the dynamical equation

$$\dot{\sigma}_x(t) = 2k(t)/\gamma\sigma_x(t) + 2\langle x(t)\eta(t) \rangle \quad (2.5)$$

for a general noise  $\eta(t)$ , which translates to

$$\dot{\sigma}_x(t) = 2k(t)/\gamma\sigma_x(t) + 2k_B T_{eff}/\gamma \quad (2.6)$$

for the equilibrium bath. Comparing these two equations, one can conclude that the active bath can be in the general situation described by the effective temperature

$$T_{eff} = \gamma/k_B \langle x(t)\eta(t) \rangle. \quad (2.7)$$

When the non-equilibrium noise is exponentially correlated,  $T_{eff}(t)$  can be calculated explicitly. Interestingly, it strongly depends on the stiffness of the potential,  $k(t)$ . This dependence on the dynamics of the engine must be considered when assessing limits on the engine's performance using known results. For example, the efficiency can reach Carnot's bound with the effective temperature only if the engine is driven quasi-statically and protocols for  $k(t)$  and bath parameters are fine-tuned to yield constant  $T_{eff}(t)$  between the adiabatic strokes.

### 2.1.2 Effective temperature in underdamped active heat engines (Ref.<sup>56</sup>)

In PUBLICATION 7.2<sup>56</sup>, we studied the existence of effective temperature for engines with Hamiltonian of the form  $H = kx^n/n + mp^2/2$ . We assumed that the dynamics is described by the system of Langevin equations

$$\dot{x}(t) = p(t)/m \quad (2.8)$$

$$\dot{p}(t) = -kx(t)^n + F(t) + \eta(t), \quad (2.9)$$

where the 'friction'  $F(t)$  stands for the systematic force exerted on the particle by the active bath and noise  $\eta(t)$  for the stochastic component of that force. Since the bath is out of equilibrium,  $F(t)$  and  $\eta(t)$  are not interconnected by a fluctuation dissipation relation. It turns out that in this setting the existence of effective temperature can be proven for quasi-static drivings only. Under such conditions, the effective temperature consistently describing both the work and heat fluxes in Eqs. (2.2) and (2.3) exists if  $\langle x(t)(F(t) + \eta(t)) \rangle = 0$ , i.e., if the total force exerted by the bath at time  $t$  is independent of the position  $x(t)$  of the engine at the same time. This condition can be broken if the

interaction between the engine and the active bath is strong enough to correlate the two subsystems. For example, if the engine is based on a colloidal particle trapped in the potential  $x^n/n$  and the active particles in the bath interact with the colloid by a steric repulsion, slowly rotating active particles will accumulate close to the colloid, leading to nonzero  $\langle x(t)(F(t) + \eta(t)) \rangle$ . This shows that when momentum is taken into account, the existence of the effective temperature depends not only on the engine Hamiltonian but also on the engine-bath coupling (cf the discussion in SUBSECTION 2.1.1).

### 2.1.3 Results (in)valid when effective temperature exists (Refs. <sup>52,55</sup>)

As discussed above, when the active engine setup allows describing the active bath by an effective temperature, both finite time and quasi-static performance will be automatically limited by corresponding known results from settings with equilibrium reservoirs. While the quasi-static limitations on efficiency, such as Carnot's efficiency, are notorious, available limitations on the finite-time performance of heat engines are much less known. To give one specific example, when an overdamped Langevin equation describes the dynamics of the active heat engine, one can immediately write down limitations on the maximum available efficiency of this engine for any fixed value of its output power using results of PUBLICATION 7.3<sup>52</sup>.

Let  $\delta P \equiv (P - P^*)/P^*$  denote the deviation from the maximum power  $P^*$  attainable in the given engine under the conditions that (i) the cycle with the effective temperature comprises two (effective)isotherms and two adiabats and (ii) the driving is slow (but not quasi-static) or the probability distributions for position at the ends of the isotherms are fixed (we will return to this somewhat awkward condition in SECTION 2.2). Then our results in Ref.<sup>52</sup> shows that the maximum efficiency attainable by the engine for given  $\delta P$  obeys the inequalities

$$\frac{\eta_C}{2} \left( 1 + \sqrt{-\delta P} \right) \leq \eta \leq \eta_C \frac{1 + \sqrt{-\delta P}}{2 - (1 - \sqrt{-\delta P})\eta_C}, \quad (2.10)$$

where  $\eta_C = 1 - T_c/T_h$  and  $T_c/T_h$  is the ratio of 'cold' and 'hot' effective temperatures. The main asset of active baths is that their hot effective temperature (achieved, e.g., by providing bacteria with a lot of food) can be very large without any danger of evaporating the lab, and thus  $\eta_C$  can be close to 1. Over the years, we have derived many similar results for various thermodynamic machines<sup>24-28</sup> all of which can find application also in the field of active heat engines (or refrigerators, etc.), but this thesis contains only PUBLICATION 7.3<sup>52</sup> as an example.

As a warning, I stress again that the existence of effective temperature just means that there is a setup with an equilibrium bath that has the same average thermodynamic performances as the given setup with an active bath, nothing more. When one studies other features of the active system, there is thus no guarantee of any further correspondence with the equilibrium system. For example, even though average work and heat for the equilibrium and active setup are equal, fluctuations of these quantities can be completely different. Such differences can be studied using Brownian dynamics simulations. Nevertheless, we have developed an alternative numerical method<sup>55</sup> (PUBLICATION 7.4) which can, in some cases, overperform these simulations, in particular, if one needs to determine with high accuracy higher moments of fluctuating thermodynamic fluxes. The method can be applied to systems with overdamped dynamics. It is based on approximating the real dynamics by a thermodynamically consistent hopping process in the discretized state space. It allows calculating the probability distribution to find the engine in a given state (position) and characteristic functions for arbitrary stochastic functionals of that position, such as work and heat. Details of this ‘Matrix numerical method’ are rather technical and I invite the interested reader to read more in the attached PUBLICATION 7.4<sup>55</sup>.

## **2.2 General results (Refs.<sup>53,54</sup>):**

There are some results obtained for standard heat engines, which are also valid for active engines even when the effective temperature does not exist. This generally holds for results obtained without assuming equilibrium concepts such as detailed balance or (equivalently) fluctuation-dissipation relation. Here, I present two examples of such results from our kitchen.

### **2.2.1 Quasi-static efficiency at finite power (Ref.<sup>53</sup>)**

PUBLICATION 7.5<sup>53</sup> shows that any cyclically driven microscopic engine can operate at maximum quasi-static efficiency and simultaneously deliver nonzero power with vanishing (or at least limited) fluctuations. For heat engines in contact with an equilibrium heat bath, this result shows that they can be operated with Carnot’s efficiency while delivering finite, stable power. This can be interpreted as a Holy Grail of engineers, which was conjectured to be forbidden by recently discovered thermodynamic uncertainty relations<sup>58</sup> before our work was published. Nevertheless, we have shown that thermodynamic uncertainty relations only limit the performance of steady-state heat



engines, transforming a stationary heat flux from a hot to a cold reservoir into work.

The first main idea of our paper is that the work in cyclic heat engines (e.g., that in Eq. (2.3)) represents a different stochastic process than work in steady state heat engines<sup>40,53</sup>. Cyclic heat engines perform work when the engines' energy is decreased by externally modifying the potential. Changes in the engine's microstate are then related to heat interchanged with the bath. On the other hand, in steady-state heat engines, both heat and work are associated with the motion of particles in a fixed potential landscape, and thus both these quantities qualify as heat from the point of view of cyclic setups. The heat and work in cyclic engines have very different properties when the system is driven slowly. For very slow driving, the individual microstates are occupied according to the quasi-static probability density (Boltzmann distribution when the bath is in equilibrium), and the probability density for work per cycle is  $\delta(w - W_{qs})$ , where  $W_{qs}$  is the average quasi-static work. The work-type variables, in other words, self-average with increasing cycle time. According to the first law, heat plus work equals energy difference per cycle. With  $\delta$ -distributed work, this implies that heat fluctuations are those of internal energy, and thus they do not vanish regardless of the driving speed<sup>40,53</sup>.

With this insight, the only question remains whether one can drive a system quasi-statically in a finite time. For small systems, all relaxation times are under reasonable control. Thus one can make them very short (definitely shorter than overdamped timescales), for example, by increasing the stiffness in the potential (2.1). This is the second main idea of PUBLICATION 7.6<sup>54</sup>, to which I refer for more details.

## 2.2.2 Maximum efficiency protocol for constrained driving (Ref.<sup>54</sup>)

Our second general result on performance of cyclic engines is described in PUBLICATION 7.6<sup>54</sup>, where we have derived maximum efficiency protocol for any heat engine described by the Hamiltonian of the form (2.1) under the experimentally relevant conditions that (i) the stiffness  $k(t) \in [k_-, k_+]$ , (ii)  $T_{eff} \in [T_-, T_+]$ , (iii) cycle time is arbitrary but fixed. Our derivation is based on the definition of heat flux (2.2) with  $m = 0$ , and thus it is completely independent of the details of engine or bath dynamics. Results of such generality are rare in the field of optimal finite-time control of (stochastic) heat engines. In fact, this is the only optimal protocol that is valid for arbitrary dynamics known to the author. All other optimal protocols described in the literature are derived based on standard functional optimization techniques, such as Euler-Lagrange formalism (see references in<sup>54</sup> for more details), which cannot be applied without prescribing the dynamical equations.

Our derivation is based on the fact that, at second glance, the heat flux  $\dot{q}(t) = k(t)\dot{\sigma}_x(t)$  resembles the Clausius equality  $TdS$  valid in equilibrium thermodynamics. In equilibrium thermodynamics, the most efficient cycle operating between temperatures  $T_-$  and  $T_+$  is Carnot's cycle, which forms a rectangle in the  $T - S$  diagram and has efficiency  $\eta_C = 1 - T_-/T_+$ . Hence, the most efficient cycle under our conditions must form a rectangle in the  $k - \sigma_x$  diagram and has efficiency  $\eta = 1 - k_-/k_+$ . An important piece of the derivation is that the final formula for efficiency is independent of the system response  $\sigma_x$  (which cancels out between the nominator and denominator in the definition of efficiency). For power, this does not happen, and hence the piece-wise constant protocol for  $k(t)$  is not always optimal. Nevertheless, one can prove that the piece-wise constant  $k(t)$  maximizes power for slow enough driving and a small allowed range  $k_+ - k_-$  for  $k$ . For more details, see PUBLICATION 7.6<sup>54</sup>.

## 2.3 Active ratchets (Refs. <sup>50,59–61</sup>)

Qualitatively (and often even quantitatively), the motion of active Brownian particles such as bacteria or various active colloids is well described by the so-called active Brownian particle model. In two dimensions, it is described by the system of Langevin equations

$$\dot{x}(t) = v[x(t), y(t)] \cos[\theta(t)] + \sqrt{2D}\eta_x(t), \quad (2.11a)$$

$$\dot{y}(t) = v[x(t), y(t)] \sin[\theta(t)] + \sqrt{2D}\eta_y(t), \quad (2.11b)$$

$$\dot{\theta}(t) = \sqrt{2D_r}\eta_\theta(t), \quad (2.11c)$$

for position coordinates  $x(t)$  and  $y(t)$  and orientation  $\theta(t)$  of the active particle. The formulae above assume that the particle's speed  $v(x, y)$  can depend on its position. The mutually independent unbiased Gaussian white noises  $\eta_i(t)$ ,  $i = x, y, \theta$  of unit intensity ( $\langle \eta_i(t)\eta_j(t') \rangle = \delta_{ij}\delta(t - t')$ ) represent translational and rotational Brownian motion of the active particle, and  $D$  and  $D_r$  denote the corresponding diffusion coefficients.

The most important ingredient of the model is that the particles move persistently until their reorient due to the rotational diffusion. The average reorientation time of the particles is given by  $D_r$ . Thus the average distance a particle travels until it changes its direction can be estimated as  $v(x, y)/D_r$ . Per the same time window, the particles' average displacement due to the translational diffusion is  $\sqrt{D/D_r}$ . The ratio of these two length scales,  $v\sqrt{D/D_r}$ , measures the importance of active motion over diffusion and is often referred to as the Péclet number.

If confined by asymmetric walls (or even potentials), the active particles slide along walls due to their persistence until they get possibly trapped in wedge-shaped regions, or pockets, such as in FIGURE 2.1. The particles can then propel freely movable objects in the active bath toward the pockets (FIGURE 2.1b). Alternatively, orienting fixed pockets in one direction renders a global current of the active Brownian particles in the opposite direction (FIGURE 2.1c).

One can ask whether active Brownian particles can render a macroscopic current by themselves without a necessity for confinement or other complications such as time-dependent activity<sup>62</sup>. We have positively answered this question in the series of papers studying the motion of active particles with space-dependent activity<sup>50,59–61</sup>.

### 2.3.1 Active Brownian particles in activity landscapes (Refs.<sup>59–61</sup>)

We started this program by studying the dynamics of active particles in spatially varying activity landscapes experimentally in PUBLICATION 7.7<sup>60</sup> and theoretically in PUBLICATION 7.8<sup>61</sup> for a simple one-dimensional setup and in PUBLICATION 7.9<sup>59</sup> for radially symmetric two-dimensional geometry. Our main findings are summarized using a piecewise constant active-passive activity landscape in FIGURE 2.3.

Due to their persistence, active Brownian particles accumulate at the active-passive interface, pointing from the active to the passive region. In the steady state, this accumulation can be described by a simple approximate model that reduces the complete Fokker-Planck equation for the probability density for position and orientation,  $\tilde{\rho}(x, \theta)$ , corresponding to Eqs. (2.11), to equations for position density  $\rho(x) = \int d\theta \tilde{\rho}(x, \theta)$  and polarization  $p(x) = \int d\theta \cos(\theta) \tilde{\rho}(x, \theta)$ . Notably, the resulting approximate equations

$$\rho'(x) = p(x)v(x)/D, \quad (2.12a)$$

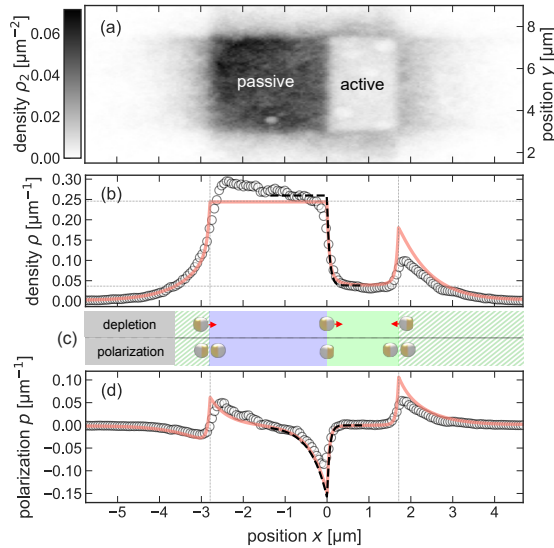
$$p''(x) = D/D_r p(x) + \rho(x)v'(x)/(2D), \quad (2.12b)$$

can be exactly mapped to equations for density and polarization in a model where the particle can have just two values of  $\theta$ , so that it points either to the left or to the right. That the approximate model describes a related model exactly implies that results obtained by solving Eqs. (2.12) should be qualitatively correct regardless of the chosen parameter regime. For more details concerning the polarization and density patterns near active-passive interfaces, particularly for the properties of the corresponding decay length, we refer to Refs.<sup>59–61</sup>. The results presented above can be used to construct an activity ratchet by imposing in the setup of FIGURE 2.3 instead of photon nudging boundaries periodic boundaries and letting the activity landscape travel from right to

left. Such time-dependent ratchets have already been described in the literature<sup>62</sup>.

### 2.3.2 Activity ratchet (Ref.<sup>50</sup>)

In PUBLICATION 7.10<sup>50</sup>, we show how the insights described above can be used to construct a ratchet just by a periodic spatial modulation of the particle activity. That such a ratchet can be constructed is nontrivial because, at long time scales, active Brownian particles usually behave like common (passive) ones, just with an increased diffusion coefficient  $D + v^2/(2D_r)$  (and thus also correspondingly increased effective temperature). And one can show that Brownian particles traveling through a bath locally equilibrated at a spatially modulated temperature can only induce thermophoretic flows from ‘hot



**Figure 2.3: Density and polarization at active-passive interface.** Panel a) shows the quasi-one-dimensional experimental setup with thermophoretically propelled Janus-type active particles sketched in panel c. When irradiated by a laser, these particles propel toward their polystyrene hemisphere. Naturally, the probability of finding the active particle is much larger in the passive than in the active region. Upon leaving the active-passive area, particles were steered back by photon nudging (they were irradiated by the laser only when pointing into the active-passive area with their polystyrene end). Panel b) shows the density in a) integrated over the  $y$  coordinate, and panel d) depicts the corresponding polarization (average orientation at a given position) of the active particle. Panel c) gives an intuitive explanation for the depletion of the active region and behavior polarization at the active-passive and passive-active interfaces. Symbols in b and d correspond to experimental data, dashed lines are analytical predictions, and solid lines were computed numerically using Ref.<sup>55</sup> (PUBLICATION 7.4). The figure is reprinted from Ref.<sup>60</sup> (PUBLICATION 7.7).

to cold' but not a macroscopic transfer under spatially periodic conditions.

Furthermore, one-dimensional static activity landscapes can be proven to generally fail to produce a global current as follows: (i) Eq. (2.11a) implies that the current  $j(x) = \langle \dot{x}(t) \rangle$  is proportional to the polarization. (ii) While activity landscapes can sort active particles according to their orientations, they can never reorient them and, hence, total orientation  $\int dx p(x) = 0$ . Physically, without external torques, polarization is a continuous function of position. Thus  $p(x)$  must be zero at least at a single position to make the overall polarization vanish. (iii) In a steady state where the ratchet operates, the one-dimensional continuity condition  $\partial j(x)/\partial x = 0$  implies that  $j(x)$  is the same for all positions and, hence, it must vanish for all  $x$ .

Even though points (i) and (ii) also hold in two spatial dimensions, this argument does not apply here because the two-dimensional continuity condition  $\partial j(x, y)/\partial x + \partial j(x, y)/\partial y = 0$  allows for nonzero global solutions with local zeros, corresponding to inevitable points of vanishing polarization. Around these points, the two-dimensional current forms vortices visible in FIGURE 2.1d. The piece-wise constant activity profile utilized in the ratchet depicted in this figure (and analyzed in Ref.<sup>50</sup>) consists of a densely populated passive region surrounded by an active region, where the particles move with a constant nonzero speed. The easiest way to understand the ratchet's operation is to consider the  $y$  dimension as a 'periodic time modulation' of the piece-wise constant profile from the preceding section. An alternative explanation can be based on the fact that particles get polarized along the whole active-passive interface; however, those localized inside the wedge region can leave the passive region much harder than those on the tip side. This leads to an overall 'leakage' of particles oriented to the left along the two edges of the wedge-shaped passive domain and, thus, to a global current to the left. For more details concerning the performance of this ratchet, we refer to PUBLICATION 7.10<sup>50</sup>.



### 3 Effects of time delay

To perform a useful task or to interact with a neighbor, both living and artificial agents must acquire and process information about their surroundings. This cannot be done instantaneously and thus response of active particles is always delayed after the stimuli (for delays of various animal species to various stimuli, see TABLE 3.1). Consequently, effects of time delay have already been thoroughly investigated from an engineering point of view in control theory<sup>73</sup>, which is a general framework for feedback systems with applications in life sciences, engineering, and sociology. The main insight is that time delays may induce oscillations, instabilities, and poor control performances, which should be familiar to everyone who experienced delayed hot water flow from a shower (see FIGURE 3.1). On the other hand, time delays are deliberately used in control theory to stabilize unstable periodic orbits in chaotic systems using, e.g., OGY or Pyragas control methods<sup>74,75</sup>.

Even though active matter research shares some goals (and hence also issues) – such as precise control of interacting self-propelling particles – with control theory, physical theories of retarded active matter are scarce: As a notable exception, effects of time delay are well understood for traffic models<sup>76</sup>. Besides, time delay was studied with respect to stability and formation of dynamical patterns in the active Brownian dynamics model<sup>77,78</sup> and in several models of bird flocks, including the Vicsek model<sup>79–82</sup> and CuckerSmale model<sup>83</sup>. All these studies suggest that moderate time delays foster order in the dynamics, while large delays induce disorder. The current interest of the active matter community in dynamics with time delay is mainly driven by the necessity to describe experiments involving feedback<sup>2,84–86</sup> and to adjust existing models to capture natural instances of retarded dynamics more accurately<sup>13,80,87</sup>. This has spurred the theoretical study of analytically tractable toy models capturing the main ingredients of experiments, and more detailed models to plan and analyze specific measurements and experiments.

In the rest of this section, I first briefly review stochastic delay differential equations, which describe the dynamics of feedback-driven active matter systems, and review our results for their solution in SECTION 3.1. Then, in SECTION 3.2, I review our results for

Animal	Stimulus/Response	Reaction Time [ms]	References
Human	auditory	140 – 160	63
	visual	180 – 200	63
	touch	~ 155	63
Fruit fly	roll perturbation	~ 5	64
	pitch perturbation	~ 12	65
	yaw perturbation	10 – 25	66
Starling	startling sound stimuli	64 – 80	67
	startling light stimuli	38 – 76	67
Teleost fish	startle response	5 – 10	68,69
Calanoida	stirring water	< 2.5	70
E. coli	chemical stimuli	~ 10 <sup>3</sup> – 10 <sup>4</sup>	71

**Table 3.1:** Typical reaction times measured between a stimulus and the corresponding discrete response strongly vary among species and the type of stimulus. Delay times comparable to the characteristic time scale of the stimulus may be expected to trigger qualitatively new effects in the dynamical response, similar to those analyzed in the present work. The table is taken from Ref.<sup>72</sup>.

feedback-driven systems of active Brownian particles. Finally, in SECTION 3.3, I review our findings about the effects of time-delayed interactions in the Vicsek model.



**Figure 3.1: Shower delay, a control problem from daily life.** Due to the delayed hot water flow, we usually open the hot water tap too much and get burned once the hot water finally comes out of the shower. To reach a comfortable temperature, we regulate the faucet up and down and induce the oscillating behavior typical for delay systems. The figure was drawn by Daniel Geiss.



### 3.1 Equilibrium delay (Ref.<sup>88</sup>)

The dynamics of most active matter systems treated considered in my work is strongly influenced by environmental noises such as thermal noise. Therefore they are described by stochastic delay differential equations<sup>89</sup>. These equations can be in general written in the form

$$\dot{\mathbf{x}}(t) = f[t, \mathbf{x}(t), \mathbf{x}(t - \tau)] + g[t, \mathbf{x}(t)]\boldsymbol{\xi}(t), \quad (3.1)$$

where  $\tau$  stands for the delay time,  $\mathbf{x}(t)$  describes the stochastic trajectory of the (possibly many-body) system,  $f[t, x(t), x(t - \tau)]$  and  $g[t, x(t)]$  are arbitrary real-valued functions, and  $\boldsymbol{\xi}(t)$  represents the noise, which is usually but not necessarily Gaussian and white. For vanishing noise, stochastic delay differential equations become delay differential equations, which are notoriously difficult to treat analytically. In fact, up to a few exceptions such as Eq. (3.7) in SUBSECTION 3.2.1, exact solutions to them are known only if they are linear ( $f[t, \mathbf{x}(t), \mathbf{x}(t - \tau)] = a + b\mathbf{x}(t) + c\mathbf{x}(t - \tau)$ ), where one can derive the Green's function for Eq. (3.1) using e.g., Laplace transform<sup>90</sup>. For the case when the noise is additive with constant intensity ( $g[t, \mathbf{x}(t)] = g$ ), this result can then be used for the derivation of exact expressions for the probability distribution for  $\mathbf{x}(t)$ ,  $\rho_1(\mathbf{x}, t)$ , and also for all higher joint probability distributions, e.g.,  $\rho_2(\mathbf{x}, t; \mathbf{x}', t')$ , etc. However, when the dynamical equation is nonlinear, it is not even known how to write a closed Fokker-Planck equation for  $\rho(\mathbf{x}, t)$ . Instead, one obtains an infinite hierarchy of equations for  $\rho_n$ ,  $n = 1, 2, = \dots$ <sup>91</sup>. How to close this hierarchy is currently an open problem. In fact, to the best of our knowledge, no exact solutions to nonlinear stochastic delay differential equations have been known until recently, when we made a moderate breakthrough<sup>88</sup> by deriving a class of (to some extend) exactly solvable nonlinear stochastic delay differential equation by imposing fluctuation-dissipation relation in Eq. (3.1).

To be specific, in PUBLICATION 7.11<sup>88</sup>, we consider stochastic delay differential equations (3.1) that can be written in the form of a system of Langevin equations (for simplicity in one dimension)

$$\dot{x}(t) = v(t), \quad (3.2)$$

$$m\dot{v}(t) = F(t, x(t)) + F_F(x(t), x(t - \tau)) + \eta(t), \quad (3.3)$$

where  $F(t, x(t))$  is an arbitrary time-local external force, and the friction  $F_F(x(t), x(t - \tau))$  and noise  $\eta(t)$  obey the fluctuation dissipation relation. That is, we assume that the

friction can be written using a memory Kernel  $\Gamma(t)$  as

$$F_F(x(t), x(t - \tau)) = - \int_{-\infty}^t dt' \Gamma(t') v(t'), \quad (3.4)$$

and the noise auto-correlation function fulfills the requirement

$$\langle \eta(t) \eta(t') \rangle = T \Gamma(t - t'), \quad (3.5)$$

with some temperature  $T$ . To restrict the analysis to real-valued processes only, we, in addition, assume that the noise power spectrum is positive:

$$S(\omega) = \int_{-\infty}^{\infty} dt \langle \eta(t) \eta(0) \rangle \exp(i\omega t) > 0. \quad (3.6)$$

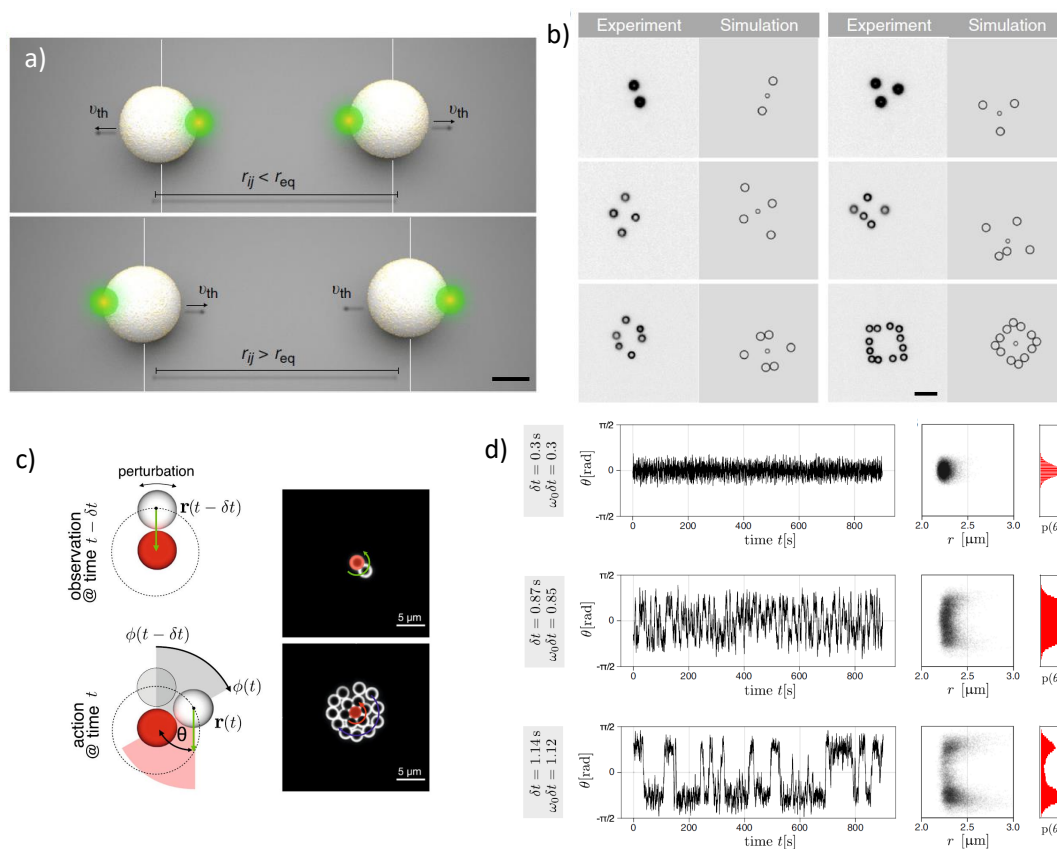
Under these conditions, the stochastic delay differential equation describes the dynamics of a particle dragged by the external force  $F$  through a non-Markovian but equilibrium bath. Hence, one can use all results valid under these conditions, such as fluctuation theorems, equilibrium linear response theory, etc. For example, when the external force is potential,  $F(\mathbf{x}(t)) = -\nabla U(\mathbf{x}(t))$ , the stationary probability distribution for  $\{x(t), v(t)\}$  is given by the Boltzmann distribution  $\rho(x, v) = \exp(-\beta H(x, v))$ , with the Hamiltonian  $H(x, v) = U(x) + mv^2/2$ , and inverse temperature  $\beta = 1/T$ . In PUBLICATION 7.11<sup>88</sup>, we detail two experimentally motivated examples of such processes. One potential issue that might constrain the practical applicability of our results is that the noise that fulfills the fluctuation-dissipation relation (3.5) is nontrivial and might be challenging to realize in experiments. Nevertheless, our numerical simulations suggest that this should be possible at least approximately.

Unfortunately, the described dynamical class involves only delay stochastic differential equations that are linear in the delayed position. Nonetheless, systems when the feedback is (at least approximately) linear in the time-delayed term are quite commonly used in practice. One example is the so-called feedback cooling<sup>92,93</sup>. Furthermore, I believe there is still untapped potential in using general results and symmetries of physics to derive solvable nonlinear stochastic delay differential equations. Currently, I am exploring ways to generalize these results.

## 3.2 Feedback driven active Brownian particles (Ref.<sup>2,72,85,90,94</sup>)

Natural microswimmers such as bacteria represent rather complex biophysical systems<sup>7</sup>. To better understand their behavior, mimic their functionality, or eventually utilize

them to perform useful tasks, researchers nowadays intensely study experimentally and theoretically artificial microswimmers. These microswimmers are often spherical particles made from two hemispheres with different physical properties, called Janus particles after the homonymous two-faced Roman god. Typical examples are the catalytic microswimmers in FIGURE 1.1 and hot (or thermophoretic) microswimmers in FIGURE 2.3c. However, as discussed in SECTION 2.3, the Janus microswimmers swim ballistically until they are reoriented by rotational diffusion, significantly limiting the experimental control over their trajectories<sup>95,96</sup>. Therefore, my experimental colleagues developed<sup>85,86</sup>



**Figure 3.2: Active particles with delayed attractive interactions.** a) When the particles are further away (closer) than  $r_{eq}$  they swim towards (away from) each other with a constant speed  $v_{th}$ . Due to these two-body interactions, the particles form dynamical ‘active molecules’ in b). c) The interaction rule from a) with  $r_{eq} = 0$  and one of the two particles fixed leads, for large enough delay times, to rotational motion of the active particle around the pinned one. d) Polar angles  $\theta = \int_{t-\delta t}^t dt' \omega(t') = \psi(t) - \psi(t - \delta t)$  traveled by the active particle per one delay time as functions of time for a fixed swimming speed and different delays. Panels a) and b) are taken from Ref.<sup>85</sup> and panels c) and d) from Ref.<sup>72</sup>.

a new type of symmetric thermophoretic microswimmers shown in FIGURE 3.2a. These particles are melamine resin spheres (radius  $1 \mu\text{m}$ ) covered by gold nanoparticles (radius  $10 \text{ nm}$ ). If placed in water and irradiated by laser at their circumference, they swim with a constant speed proportional to the laser intensity in the direction of the vector connecting the laser focus with the particle center. The swimming direction can be controlled simply by changing the position of the laser focus.

This improved experimental control allows steering the microswimmers with unprecedented precision, only hindered by the phenomenon that inevitably limits the accuracy of any feedback control, the delay of the feedback loop, i.e., the time required to measure the position, process the measurement in the computer, and change the laser focus. Even though the current (2023) variant of the experimental setup allows for short enough delays that hardly affect the dynamics, during the development of the experiment, we uncovered many surprising phenomena occurring for long enough delay times. In short, it turns out that trivial delayed interactions alone have potential to imply a large part of the complexity observed in motile active matter.

### 3.2.1 Active Brownian molecules (Ref. <sup>85,90</sup>)

In PUBLICATION 7.12, we investigated the level of control achievable using the symmetric active particles by steering them to form active molecules in FIGURE 3.2b. We achieved that by implementing the simple rule depicted in FIGURE 3.2b. When the particles were further away than a fixed nonzero distance  $r_{eq}$ , we propelled them with a fixed speed  $v_{th}$  towards each other. And when they were closer than  $r_{eq}$ , we propelled them with the same speed away of each other. As a result the dynamics of the center of mass of the particles obeyed the nonlinear stochastic delay differential equation

$$\dot{r}(t) = -2v_{th} \text{sign}(r(t - \tau) - r_{eq}) + \sqrt{4D}\eta(t). \quad (3.7)$$

For vanishing delay, this equation describes an overdamped Brownian particle in absolute value potential  $U(r) = |r - r_{eq}|$ . For nonzero delay, the particles are thus on average distant  $r_{eq}$  and their distance  $r$  exhibits exponentially decaying fluctuations following from the Boltzmann distribution  $P(r) \propto \exp[-U(r)/2D]$ . Interestingly, the nonlinear delay differential equation obtained by neglecting the noise can be solved exactly by a triangle wave with amplitude  $2v_{th}\tau$  and period  $4\tau$ . Due to the delay, the distance  $r$  is thus on average still given by  $r_{eq}$  but it oscillates around this value. Due to the noise, these oscillations have a finite correlation time, which can be predicted by an approximate solution of Eq. (3.7)<sup>85</sup>.

Considering more than two particles, the first term on the right-hand side of Eq. (3.7) is given by the average of two-particle interactions, normalized to  $v_{th}$ . For zero delay, the equation describes the diffusion in a multi-dimensional absolute value potential, which insight can be used to determine the average structure of the resulting ‘active molecules’ in FIGURE 3.2b. Similarly, as for two particles, these molecules are highly dynamic. They oscillate due to the delay, and due to the noise, the individual particles in the molecules can even interchange their positions. For more details, I refer to videos supplementing Ref.<sup>85</sup> and to its full text in PUBLICATION 7.12.

Since the interactions used in the experiments are strongly nonlinear already for two particles, the theoretical analysis is limited to highly approximate calculations. To get more insights, we considered in PUBLICATION 7.13 a similar system with linear interactions, i.e., we considered a harmonic potential for vanishing delay instead of the absolute value potential. The main physical difference between this setup and the experimental setup of PUBLICATION 7.12 is that the latter assumes constant particle speeds while the ‘harmonic potential’ implies that the speed grows linearly with increasing interparticle distance. As a result, the dynamics can be reasonably linearized even for more than two particles, and the resulting approximated dynamical equations can be solved exactly. Interestingly, one can even obtain a reasonable analytical estimate for transition rates between different possible conformations of the molecules formed by the delayed harmonic interactions. However, this is only possible when neglecting that the usually used absorbing boundary condition is no longer valid for non-Markovian dynamics. As a result, the predicted transition rates are reasonably accurate for intermediate delays only. While this outperforms the usually employed short delay approximation<sup>97</sup>, derivation of transition rates for non-Markovian dynamics still represents an interesting open problem.

The main qualitative difference between the ‘constant-force’ molecules of PUBLICATION 7.12<sup>85</sup> and ‘harmonic’ molecules of PUBLICATION 7.13<sup>90</sup> is that, due to the constant speed, the former are stable for arbitrarily long delay times. On the other hand, the oscillations in a harmonic potential get amplified for long delay times, leading to an exponential increase of inter-particle distances. Hence, the harmonic molecules exhibit both hallmark features of delay systems: instabilities and oscillations. Since the analysis in<sup>90</sup> is somewhat technical, I invite the interested reader to read more in the attached PUBLICATION 7.13.

### 3.2.2 Delay-induced chirality in systems of micro-swimmers (Ref.<sup>72,94</sup>)

In PUBLICATION 7.14<sup>72</sup>, we investigated what happens if the simple constant speed interaction from the previous subsection is used to propel the particles towards a fixed target particle. The situation is depicted in FIGURE 3.2c. The microswimmer detects the target's position vector  $-\mathbf{r}(t)$  at time  $t$  and at time  $t + \delta t$  swims in the direction  $-\mathbf{r}(t)/|\mathbf{r}(t)|$  with a constant speed  $v_0$  \*. As depicted in the figure, this delayed attraction makes the particle rotate around the target for long enough delay times. Mathematically, the rotational motion of a single microswimmer is well described by the nonlinear stochastic delay differential equation

$$\dot{\phi}(t) = \omega_0 \sin[\phi(t) - \phi(t - \delta t)] + \sqrt{D/(2a^2)}\eta(t), \quad (3.8)$$

where  $\phi(t)$  is the polar angle (see FIGURE 3.2c),  $D$  the transitional diffusion coefficient of the microswimmer,  $a$  its radius (the target particle is just a pinned, passive microswimmer), and  $\omega_0 = v_0\delta t/(2a)$ . This equation describes a Kuramoto oscillator<sup>98</sup> trying to synchronize with its own past position. Assuming that, for  $D = 0$ , the system eventually converges into a state with constant angular velocity  $\omega$ , one can use the formula  $\phi(t) = \int_{-\infty}^t dt' \omega$  to rewrite Eq. (3.8) as  $\omega = \omega_0 \sin(\omega\delta t)$  and find the stable solutions as functions of the control parameter  $\omega_0\delta t$  numerically. It turns out that, for  $\omega_0\delta t < 1$ , the only stable solution is  $\omega = 0$ , while there are two stable rotating states differing in the sign of  $\omega$  for  $\omega_0\delta t > 1$ . For increasing delay (or, equivalently, swimming speed  $v_0$ ), the system thus undergoes a normal supercritical pitchfork bifurcation<sup>99</sup>. Alternatively, approximating  $\omega(t)\delta t$  by the delay angle  $\theta(t) \equiv \int_{t-\delta t}^t dt' \omega(t')$ , and expanding the sine in Eq. (3.8) up to the third order in the delay time  $\delta t$  (and neglecting the term  $\ddot{\theta}(t)$ , which makes the resulting approximate equation unstable<sup>100</sup>), one arrives at the Markovian Langevin equation, which describes the diffusion of an overdamped Brownian particle in a quartic potential. For  $\omega_0\delta t < 1$ , the potential has a single minimum corresponding to the non-rotating state. And, for  $\omega_0\delta t > 1$ , it has two symmetric minima corresponding to the two rotating states. In addition to this insight, the Markovian Langevin equations allow one to predict relaxation times to the stable rotating states and, using the Kramers' theory<sup>101</sup>, the transition rates for changes between the two transiently-stable rotating states attained for  $\omega_0\delta t > 1$ . This is the simplest version of the theory, which

---

\*I apologize that the notation throughout the text is not unified. For example, in more experimentally motivated papers, we denote delay as  $\delta t$  and in the theoretical ones as  $\tau$ . I decided to reuse the notation employed in the attached publications in their commentary to help an interested reader to digest the publications more easily.

explains the behavior observed in experiments qualitatively. In PUBLICATION 7.14, we also developed a refined version of the above-described theory, which considers additional experimental details (most notably, another type of delay involved in the feedback loop). The refined theory gives even quantitative agreement with the experiments.

Interestingly, the single particle theory also fits the average angular velocity of multiple particles rotating around the target when each is attracted to the target by the same delayed attraction as the single particle. In addition, the individual particles interact sterically, hydrodynamically, and thermophoretically. Due to these interactions, the particles organize in concentric shells around the target, which rotate in the same direction for large delays, and can even counter-rotate for intermediate delays. While the corotation can be explained solely based on steric interactions, the counterrotation is caused by hydrodynamic coupling between the particles. Simply put, when the laser propels a particle, it also propels the water in the opposite direction. This backflow pushes particles in neighboring shells in the opposite direction for intermediate delays. For more details, I refer to PUBLICATION 7.14<sup>72</sup>.

As our current experimental setup is not capable of controlling more than 20 particles at once, we decided to study manybody systems with up to 200 particles using Brownian dynamics simulations. In the simulations, we took into account the steric repulsion between the particles but not the hydrodynamic coupling (taking into account hydrodynamics for such a large system represents a nontrivial numerical challenge). The results of these simulation, described in PUBLICATION 7.15<sup>94</sup>, are quite surprising. While the average angular velocity of the system still qualitatively obeys the single particle theory described above, the detailed dynamics of the system experiences a series of dynamical phase transitions. These transitions are induced by shear stress caused by unequal angular velocities of the individual particle layers around the target particle. When  $v_0\delta t$  is increased, the system goes through the following dynamical phases:

1. stable, non-rotating crystallite.
2. homogeneously rotating crystallite.
3. sheared or ‘quaking’ crystallite, where the outermost layers slide over (or lag behind) the inner layers.
4. ring phase, where the innermost layers are no longer in contact with the target particle.
5. a yin-yang phase, where the radial symmetry of the ring state is broken.

6. a blob phase, where the particles completely detach from the target and form a densely packed satellite orbiting around it while shaking from the shear stresses.

The shearing of the system is of slip and stick type observed in athermal granular materials, and thus it is accompanied by the formation of shear bands. For more details and for a detailed discussion of the individual dynamical phases, I refer to PUBLICATION 7.15<sup>94</sup>. A very good intuition about the behavior of this beautiful system can be obtained by watching videos of the individual phases, which can be found either in the paper’s supplementary material or on YouTube <sup>†</sup>.

### 3.2.3 Machine learning with micro-swimmers (Ref.<sup>2</sup>)

One of the ultimate aims of active matter research is to develop autonomous, perhaps even self-learning, artificial microswimmers with applications, e.g., in engineering or medicine. Motivated by this goal and also with the vision that understanding the adaptation of artificial microswimmers to real-world conditions might bring new insights into evolutionary mechanisms at work in the development of bacteria and similar natural microswimmers, we have investigated in PUBLICATION 7.16<sup>2</sup> how our symmetric artificial microswimmers can learn to orient in real-world arenas by using reinforcement learning. To the best of our knowledge, our work represents the first experimental application of reinforcement learning to a real-world navigation problem in a noisy environment. Our setup is halfway to the goal of autonomous self-learning microswimmers because the brain that learns the optimal strategy is not inside the individual particles but in a computer operating the feedback loop.

In the experiment, the microswimmers are confined between two glass cover slides, and thus they effectively move in two dimensions. To implement the learning, we divided the plane into  $7 \times 7$  equal squares shown in FIGURE 3.3a. Blue denotes the region through which the microswimmer can move to reach the green target state. When the swimmer entered the red absorbing boundary, it was returned back to its initial position at one of the blue states. To find the optimal policy to steer the swimmer from blue states to the target state using the set of allowed actions in FIGURE 3.3b, we implemented the reinforcement learning method called Q-learning<sup>102</sup>. In this method, one defines a Q-matrix where weight is given for performing the allowed actions in each blue state. Hence our matrix had  $9 \times (5 \times 5 - 1)$  entries. The policy described by the Q-matrix imposes the action with the lowest weight in each state.

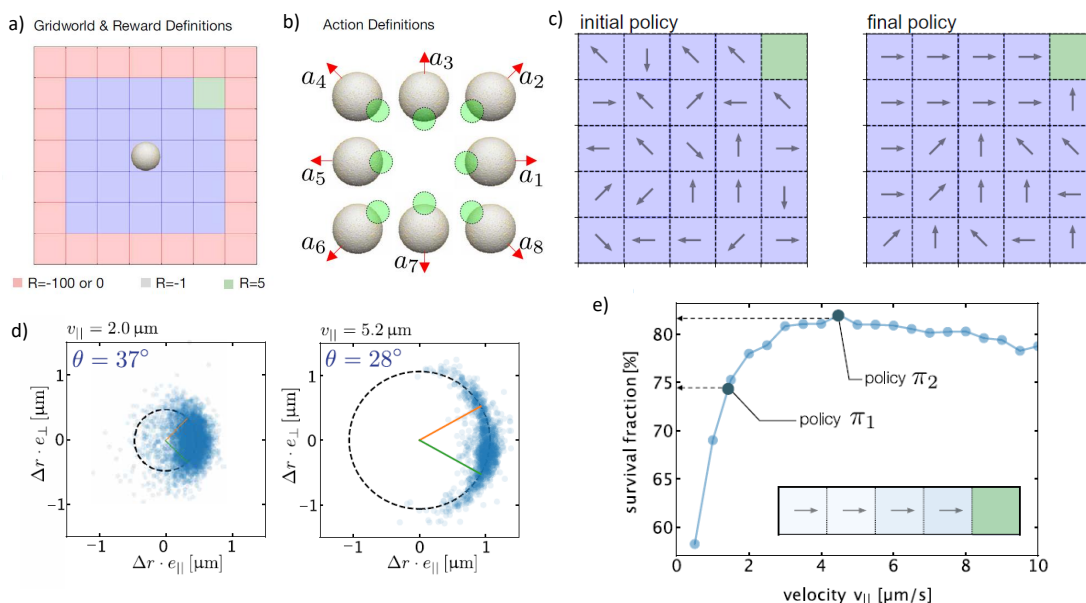
---

<sup>†</sup>One can either click the ‘YouTube’ above in the electronic version or use the link [https://www.youtube.com/watch?v=1Gfgq7FvfaA&list=PLDwaP\\_kIyigWI4637AQH1upD4seyYOynw](https://www.youtube.com/watch?v=1Gfgq7FvfaA&list=PLDwaP_kIyigWI4637AQH1upD4seyYOynw).



At the beginning of the learning, the Q-matrix is populated randomly, resulting in a random initial policy depicted in FIGURE 3.3c, left. During the learning, the Q-matrix is updated according to an algorithm, which gives a positive reward to actions leading the swimmer to the target and negative rewards to the actions which end up in one of the absorbing states (for more details, I refer to PUBLICATION 7.16<sup>2</sup>). The final policy obtained after the learning, depicted in FIGURE 3.3c, right, represents an optimal compromise between the fast approach to the target and staying in the blue arena in the noisy environment. We have tested that learning is more efficient if several microswimmers update the same Q-matrix.

My main job in this project was to explain why the optimal policy contains some unintuitive elements (e.g., those pointing to the left while the target is in the up di-



**Figure 3.3: Reinforcement learning with artificial microswimmers.** a) shows the grid world where the swimmer learns to navigate using the set of allowed actions in b). The red squares in a) are absorbing states, and the green square is the target state. c) Due to the learning, the initial random policy (left) transforms into an optimal policy to reach the target as fast as possible (right). d) due to the feedback loop delay between detecting the microswimmers' position and imposing the action and thermal noise, the real displacements  $\Delta r$  of the particle (blue circles) are symmetrically distributed along the desired displacement  $v_{\parallel} \delta t e_{\parallel}$  ( $v_{\parallel}$  is the particle swim speed,  $\delta t$  the time delay, and  $e_{\parallel}$  unit vector in a desired direction). This leads to the optimal swim speed (or delay time) in e) to reach a target position without being absorbed by the boundary. Figures were taken from Ref.<sup>2</sup>.

reaction). Partially, this can be explained by a weak drift in the experimental sample. However, there is also a more fundamental reason for such a policy, which should be considered by any device or animal navigating with time delay in a noisy environment. Within the delay time between the decision where the particle should swim and taking the corresponding action, the swimmer performs Brownian motion (and it also perhaps moves due to the previous action). As a result, its relative position to the target at the time of actual implementation of the action is stochastic, resulting in the set of actual displacements depicted in FIGURE 3.3d, which are randomly distributed around the desired displacement. The corresponding error increases with the noise intensity  $D$ , delay time  $\delta t$ , and the swimming speed  $v_{\parallel}$ . As a consequence, there is an optimal speed that guarantees that the target is reached with maximum probability (see FIGURE 3.3e). Using a simple model detailed in Ref.<sup>2</sup>, the optimal speed can be estimated as

$$v_{\parallel}^{opt} = \sqrt{\frac{2D}{\sinh \sigma_{\theta}^2 \delta t}}, \quad (3.9)$$

where  $\sigma_{\theta}^2$  is the variance of the aiming error angle  $\theta$ , depicted in FIGURE 3.3d. The variance depends on the previous action, noise intensity, and delay time, and in Ref.<sup>2</sup> we take it as a fit parameter.

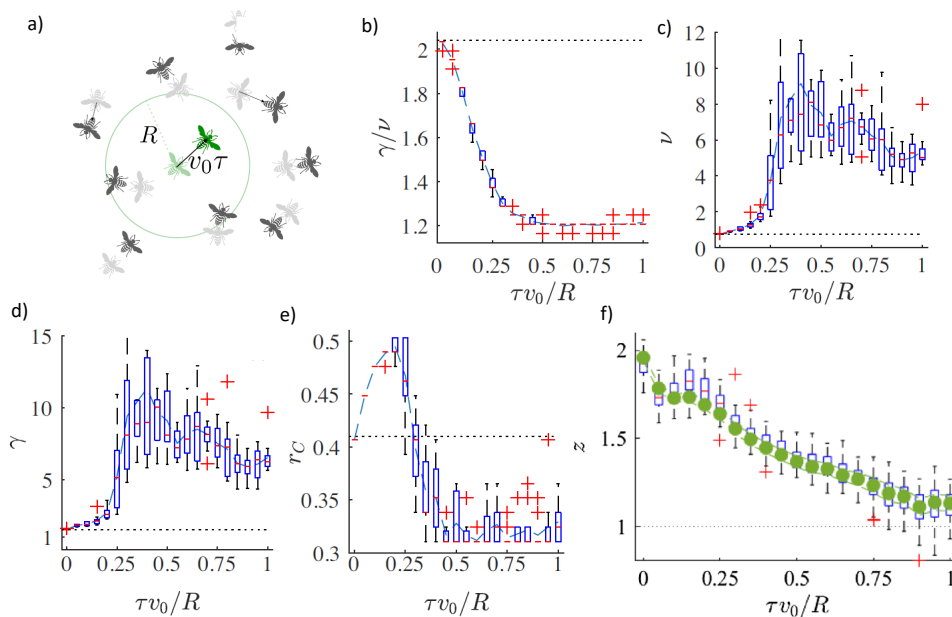
The above formula also predicts an optimal delay time  $\delta t$  for a fixed swimming speed. Interestingly, this is in accord with the recent finding that the precision of reaching a target by the run-and-tumble bacteria also exhibits an optimum as a function of the run-and-tumble times<sup>103,104</sup>, which play a similar role for the motion of bacteria as the single delay time  $\delta t$  in our experiments. For more details, I refer to PUBLICATION 7.16<sup>2</sup>.

### 3.3 Delay Vicsek model (Ref.<sup>80–82</sup>)

The Vicsek model<sup>105</sup> is one of the best-known toy models of active matter. Its original variant<sup>106</sup>, is a simple generalization of the XY model (in two dimensions) and Heisenberg model (in three dimensions) in which the individual spins (agents) move in discrete time with a fixed speed  $v_0$  in their direction. At each time  $t$ , the spin of agent  $i$  assumes the value of the average spin of its neighbors closer than an interaction radius  $R$  at time  $t - 1$  modified by a noise (alignment interaction). For low noise intensities, the Vicsek model exhibits a global order (aligned spins) even in two dimensions, and hence it beats the equilibrium limitation imposed by the Mermin-Wagner theorem. For an intense noise, the spins are disordered. The nature of the transition between these two phases was long debated. The current consensus is that the transition is discontinuous (or

first order), with a microphase separation into dense bands (or sheets) of aligned spins and low-density bands of disorder traveling through the sample. The bands form only for large enough systems, where the crossover system size<sup>107</sup> increases with decreasing speed  $v_0$ . However, even at high speeds, microphase separation only occurs in simulations involving a significant number of particles. That is why understanding the true nature of the transition had to wait for sufficiently fast computers.

In fact, not only the original paper<sup>106</sup> reported that the transition is continuous (second order), as in the Heisenberg model. Nowadays, it seems to be clear that the transition looks second order whenever the density fluctuations in the simulation are not too large<sup>15</sup>. Since these fluctuations grow with the particle number  $N$ , the transition can be considered as smooth for small enough  $N$  (for fixed  $v_0$ ). In this regime, the Vicsek model close to the transition exhibits a finite-size critical behavior<sup>13,87</sup> in the sense that a set of scaling functions and critical exponents describes susceptibility and space and time correlation functions<sup>108,109</sup>.



**Figure 3.4: Finite-size scaling in the delay Vicsek model.** a) In the delay Vicsek model, each agent assumes at time  $t + \tau$  average orientation of the particles, which were closer to it than  $R$  at time  $t$ . The agents move with constant speed  $v_0$  in discrete time (time-step 1) in the direction of their orientation. The static critical exponents (b-d), the critical nearest neighbor distance (e), and the dynamical exponent  $z$  dramatically change with increasing delay from their values for the classical Vicsek model towards long-delay asymptotic values. The figures were taken from Ref.<sup>80</sup>.

When comparing the predictions of the Vicsek model to data obtained for birds or insects, the finite size results in the region way below the crossover system size are of the main interest<sup>13,87,110</sup>. These comparisons show that the Vicsek model fails to predict the shape of time-correlation functions and scaling exponents found from experimental data for swarms of midges<sup>13,87</sup>, and also information spreading in bird flocks, e.g., when birds follow a leader or react to a local stimuli<sup>110</sup>.

The authors of Refs.<sup>13,15</sup> argued that these failures of the Vicsek model follow from the fact that it completely neglects inertia in changing the orientation of the individual agents. Hence, they introduced an improved model called as inertia spin model, where the alignment interaction acts on the agent orientation indirectly by an additional ‘spin’ variable controlled by the time derivative of the orientation. The refined model’s predictions nicely agree with the field observations<sup>13,15,110</sup>. Nevertheless, we see another important gap in the Vicsek model: it neglects time delay in the interactions. This is our main motivation for investigations of delay Vicsek model<sup>80</sup>, where the alignment interaction is not based on the particle’s neighbors at time  $t - 1$ , but at time  $t - 1 - \tau$  (see FIGURE 3.4a). Below, I present the results we have obtained for finite-size scaling (SUBSECTION 3.3.1) and information propagation (SUBSECTION 3.3.2) in the delay Vicsek model. I consider these results preliminary and am still actively working on their refinement.

### 3.3.1 Finite-size scaling in delay Vicsek model (Ref.<sup>80</sup>)

In PUBLICATION 7.17<sup>80</sup>, we report on our study of finite size scaling in the delay Vicsek model. In the study, we fixed noise intensity, particle speed  $v_0$ , and interaction radius  $R$  and considered various particle numbers  $N$  ranging from 64 to 2048 particles. We simulated the delay Vicsek model in a cube with edge  $L$  and periodic boundary conditions. Keeping in mind that our results were obtained for a specific set of parameters is important because it is known that scaling exponents in the Vicsek model are, in general, parameter dependent.

For each  $N$ , we varied  $L$  to obtain the susceptibility  $\chi$  of the system as a function of the nearest neighbor distance between the particles,  $r_1$ . Afterward, we computed the static critical exponents  $\gamma$  and  $\nu$  and the asymptotic nearest neighbor distance  $r_C$  by the best data collapse of the resulting curves by shifting and rescaling the x and y axis as  $(r_1 - r_C)N^{1/3\nu}$  and  $\chi N^{-\gamma/3\nu}$ . We have repeated this procedure for delay times  $\tau$  ranging from 0 to  $v_0\tau/R \sim 1$ . The resulting delay dependences of the critical exponents and  $r_C$  are given in FIGURE 3.4b-e. With increasing delay time, all the parameters converge

from their standard-Vicsek-model values to long delay plateau values. An analytical argument given in the supplementary information to<sup>80</sup> shows that, for long delays, the delay Vicsek model dynamics depends just on the combination  $v_0\tau$ . Thus, while the precise form of the delay dependence of the static critical parameters  $\gamma$ ,  $\nu$ , and  $r_C$  varies with the particle speed, they should converge to the same plateau values as in FIGURE 3.4 for arbitrary nonzero  $v_0$ . I am now working on a numerical check of these results. Unfortunately, the numerical simulations, mainly calculating susceptibilities and other correlation functions, are incredibly time-consuming.

The most interesting observations from the static finite size scaling are the unprecedentedly large values of the critical exponents  $\gamma$  and  $\nu$  for intermediate and long delays and the maximum found in the asymptotic nearest neighbor distance  $r_C$ . While we still haven't understood the meaning of the large scaling exponents, the maximum in  $r_C$  reflects the known<sup>79,82,83</sup> effect of stabilization of the flocking phase by short delays. Delayed reactions enhance the system's stability against random perturbations. On the other hand, too long delays prevent the agents from efficiently following their neighbors. The maximum in  $r_C$  results from a compromise between these two tendencies.

For each set  $N$ ,  $v_0$ ,  $R$ , noise intensity, and delay time  $\tau$ , the susceptibility exhibits a maximum that marks the system size (or, equivalently, the nearest neighbor distance  $r_1$ ) corresponding to the order-disorder transition. For the parameters at the transition, we have calculated space and time-correlation functions. For each  $\tau$ , the correlation functions for different  $N$  can again be collapsed to a single master curve by rescaling time as  $t/\tau_R = t/\xi^z$ , where  $\tau_R$  and  $\xi$  are the correlation time and length, and  $z$  is the dynamical exponent. The resulting dependence of  $z$  on the delay time is given in FIGURE 3.4f. For  $\tau = 0$ ,  $z = 2$  as in the standard Vicsek model with a small  $v_0$  and in the Heisenberg model. With increasing delay time,  $z$  converges to 1, which is the value reported for natural swarms in Ref.<sup>13</sup>, and which is also close to the prediction from the inertia spin model<sup>15</sup>. Besides, the shape of time correlation functions obtained from the delay Vicsek model is also similar to those found for natural swarms and the inertia spin model (for details, see PUBLICATION 7.17<sup>80</sup>). According to our analysis, time delay thus represents another way to explain the dynamical scaling observed in natural swarms. However, this work is still in progress.

### 3.3.2 Information propagation in delay Vicsek model (Ref.<sup>81,82</sup>)

When we found that the delay Vicsek model is capable of reproducing dynamical scaling and time correlations observed in natural swarms, we focused on the study of information

propagation in the model in a similar fashion as it was done for natural bird flocks<sup>110</sup>.

Even without delay, the study of information propagation in motile systems such as the Vicsek model is a complicated task, in particular when the source of information is moving, which is quite a generic situation in flocks following a leader bird. Therefore, we have started our investigation in PUBLICATION 7.18<sup>81</sup> with a simple lattice model, where a scalar field at a given site assumes at discrete time  $t + 1$  average value of itself and its neighbors at time  $t$ . At first glance, the model can be classified as a lattice variant of the Vicsek model with zero speed of the agents and a vanishing noise. However, it turns out that its different continuum limits converge to the spin-wave approximations of either the inertia spin model (when lattice constant is kept proportional to time step) or the Vicsek model (when lattice constant squared is kept proportional to time step).

Using the lattice model, we studied two types of local information sources called firm and lax leaders. Firm leaders are meant to describe a leader particle deliberately trying to influence the whole system and correspond to fixing the field's value at the origin during the whole evolution of the system. Lax leader describes the spreading of a random fluctuation through the system and corresponds to setting the field's initial value at the origin to a given value and letting the system evolve freely for  $t > 0$ . It turns out that a reasonable definition of signal speed is (distance from the leader)/(the time when the field changes most rapidly at that distance). Using this definition, we found that the information spreading in the lattice model is approximately diffusive for both types of perturbation, i.e., the distance traveled by the signal is proportional to  $\sqrt{t}$ . Interestingly, this result is obtained regardless of the fact the information spreading in the inertia spin model is predicted to be linear<sup>110</sup>.

Next, we considered the Vicsek model with very weak noise and the two types of perturbations from the lattice model, which, however, traveled through the system at the same speed as the other agents. For low speeds, the information spreads in the same way as in the lattice model. However, the information spreading is no longer purely conductive for larger speeds. As a result, the information spreads diffusively in the direction opposite to the leaders heading and approximately ballistically (distance traveled proportional to  $t$ ) in the direction of the leader.

In PUBLICATION 7.19<sup>82</sup>, we performed an analogous analysis of information spreading in the delay Vicsek model, together with the analysis of the ability of the system to follow a moving leader. Concerning the latter, we found in accord with the results described in the previous SUBSECTION 3.3.1 that delays foster the stability of the aligned state of the system against random perturbations but hinder the system's ability to follow a leader. Concerning information propagation, we found that the information

spreading in the direction opposite to the leader's motion is still diffusive. However, for a fixed low (but nonzero) speed of the agents, the information spreading in the leader's direction is diffusive for short delays but becomes increasingly linear as the delay is increased. Furthermore, the delay introduces oscillations into the dispersion relations. Again, I consider these results preliminary, and we are working intensely to improve our understanding of information spreading in Vicsek and related models.

Finally, in Ref.<sup>82</sup>, we have also studied linear response in the delay Vicsek model. Out of thermal equilibrium, it lacks its general properties as a response is no longer bound to be given by equilibrium (or stationary) correlation functions, and we indeed have not found such a general relation. Nevertheless, our analysis suggests that the response in the Vicsek model to a torque applied to a subgroup of agents is linear only in the parameter regime when the average polarization of the system is approximately conserved. For more details, I refer to PUBLICATION 7.19<sup>82</sup>.





## 4 Final Remarks

This thesis summarizes some of the advances in our understanding of the dynamics and thermodynamics of active matter, focusing on energy extraction from active self-propulsion and the effects of time-delayed interactions. These results represent tiny contributions to the knowledge acquired over the past years in this dynamic field. And even our results leave more loose ends than answers. For example, we have just started with investigations of the importance of delay in active matter systems. More importantly, most of our current models are rather based on observed phenomenology than on some deeper (bio)physical principles. Hence one of the natural topics for future investigation is to derive more reliable models of interparticle interactions by using a bottom-up approach based on the capabilities of the individuals in question. For example, our preliminary works show that the Vicsek model can be (approximately) derived by considering agents moving with a fixed speed and trying to maximize their local orientational correlations with their neighbors. Similarly, other types of agents combined with other local target functions might result in new models more suitable to a given situation than our present models. Besides continuing the study of the influence of delay on the dynamics of experimental Brownian active matter systems, studying such a bottom-up approach to the derivation of active matter systems is my main goal for the forthcoming years.



## 5 References

1. Ma, X., Jang, S., Popescu, M. N., Uspal, W. E., Miguel-López, A., Hahn, K., Kim, D.-P. & Sánchez, S. Reversed Janus Micro/Nanomotors with Internal Chemical Engine. *ACS Nano* **10**, 8751–8759. ISSN: 1936-0851. <https://doi.org/10.1021/acsnano.6b04358> (2016).
2. Muiños-Landin, S., Fischer, A., Holubec, V. & Cichos, F. Reinforcement learning with artificial microswimmers. *Science Robotics* **6**, eabd9285. <https://www.science.org/doi/abs/10.1126/scirobotics.abd9285> (2021).
3. Zastrow, M. Researchers create 1,000-robot swarm. *Nature*. ISSN: 1476-4687. <https://doi.org/10.1038/nature.2014.15714> (2014).
4. Romanczuk, P., Bär, M., Ebeling, W., Lindner, B. & Schimansky-Geier, L. Active Brownian particles. *The European Physical Journal Special Topics* **202**, 1–162. ISSN: 1951-6401. <https://doi.org/10.1140/epjst/e2012-01529-y> (2012).
5. Cates, M. E. & Tailleur, J. Motility-Induced Phase Separation. *Annual Review of Condensed Matter Physics* **6**, 219–244. <https://doi.org/10.1146/annurev-conmatphys-031214-014710> (2015).
6. Bechinger, C., Di Leonardo, R., Löwen, H., Reichhardt, C., Volpe, G. & Volpe, G. Active particles in complex and crowded environments. *Reviews of Modern Physics* **88**, 045006. <https://link.aps.org/doi/10.1103/RevModPhys.88.045006> (2016).
7. Gompper, G., Bechinger, C., Herminghaus, S., Isele-Holder, R., Kaupp, U. B., Löwen, H., Stark, H. & Winkler, R. G. Microswimmers – From Single Particle Motion to Collective Behavior. *The European Physical Journal Special Topics* **225**, 2061–2064. ISSN: 1951-6401. <https://doi.org/10.1140/epjst/e2016-60095-3> (2016).
8. Dreyfus, R., Baudry, J., Roper, M. L., Fermigier, M., Stone, H. A. & Bibette, J. Microscopic artificial swimmers. *Nature* **437**, 862–865. ISSN: 1476-4687. <https://doi.org/10.1038/nature04090> (2005).

9. Qian, B., Montiel, D., Bregulla, A., Cichos, F. & Yang, H. Harnessing thermal fluctuations for purposeful activities: the manipulation of single micro-swimmers by adaptive photon nudging. *Chemical Science* **4**, 1420–1429. <http://dx.doi.org/10.1039/C2SC21263C> (4 2013).
10. Lavergne, F. A., Wendehenne, H., Bäuerle, T. & Bechinger, C. Group formation and cohesion of active particles with visual perception-dependent motility. *Science* **364**, 70–74. <https://www.science.org/doi/abs/10.1126/science.aau5347> (2019).
11. Friedl, P. & Gilmour, D. Collective cell migration in morphogenesis, regeneration and cancer. *Nature Reviews Molecular Cell Biology* **10**, 445–457. ISSN: 1471-0080. <https://doi.org/10.1038/nrm2720> (2009).
12. Elgeti, J., Winkler, R. G. & Gompper, G. Physics of microswimmers—single particle motion and collective behavior: a review. *Reports on Progress in Physics* **78**, 056601. <https://dx.doi.org/10.1088/0034-4885/78/5/056601> (2015).
13. Cavagna, A., Conti, D., Creato, C., Del Castello, L., Giardina, I., Grigera, T. S., Melillo, S., Parisi, L. & Viale, M. Dynamic scaling in natural swarms. *Nature Physics* **13**, 914–918. ISSN: 1745-2481. <https://doi.org/10.1038/nphys4153> (2017).
14. Cavagna, A., Di Carlo, L., Giardina, I., Grandinetti, L., Grigera, T. S. & Pisegna, G. Dynamical Renormalization Group Approach to the Collective Behavior of Swarms. *Physical Review Letters* **123**, 268001. <https://link.aps.org/doi/10.1103/PhysRevLett.123.268001> (2019).
15. Cavagna, A., Di Carlo, L., Giardina, I., Grigera, T. S., Melillo, S., Parisi, L., Pisegna, G. & Scandolo, M. Natural swarms in 3.99 dimensions. *Nature Physics*. ISSN: 1745-2481. <https://doi.org/10.1038/s41567-023-02028-0> (2023).
16. Ballerini, M., Cabibbo, N., Candelier, R., Cavagna, A., Cisbani, E., Giardina, I., Lecomte, V., Orlandi, A., Parisi, G., Procaccini, A., Viale, M. & Zdravkovic, V. Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. *Proceedings of the National Academy of Sciences* **105**, 1232–1237. <https://www.pnas.org/doi/abs/10.1073/pnas.0711437105> (2008).
17. Katz, Y., Tunstrøm, K., Ioannou, C. C., Huepe, C. & Couzin, I. D. Inferring the structure and dynamics of interactions in schooling fish. *Proceedings of the*

- National Academy of Sciences* **108**, 18720–18725. <https://www.pnas.org/doi/abs/10.1073/pnas.1107583108> (2011).
18. Feinerman, O., Pinkoviezky, I., Gelblum, A., Fonio, E. & Gov, N. S. The physics of cooperative transport in groups of ants. *Nature Physics* **14**, 683–693. ISSN: 1745-2481. <https://doi.org/10.1038/s41567-018-0107-y> (2018).
  19. Vermuyten, H., Beliën, J., De Boeck, L., Reniers, G. & Wauters, T. A review of optimisation models for pedestrian evacuation and design problems. *Safety Science* **87**, 167–178. ISSN: 0925-7535. <https://www.sciencedirect.com/science/article/pii/S092575351630042X> (2016).
  20. Medina-Sánchez, M., Schwarz, L., Meyer, A. K., Hebenstreit, F. & Schmidt, O. G. Cellular Cargo Delivery: Toward Assisted Fertilization by Sperm-Carrying Micro-motors. *Nano Letters* **16**, 555–561. ISSN: 1530-6984. <https://doi.org/10.1021/acs.nanolett.5b04221> (2016).
  21. Iyengar, S., Brooks, R. & University, C. *Distributed Sensor Networks* ISBN: 9781439870785. <https://books.google.cz/books?id=efzRBQAAQBAJ> (CRC Press, 2004).
  22. Holubec, V. & Novotný, T. Effects of Noise-Induced Coherence on the Performance of Quantum Absorption Refrigerators. *Journal of Low Temperature Physics* **192**, 147–168. <https://link.springer.com/article/10.1007/s10909-018-1960-x> (2018).
  23. Holubec, V. & Novotný, T. Effects of noise-induced coherence on the fluctuations of current in quantum absorption refrigerators. *The Journal of Chemical Physics* **151**, 044108. <https://aip.scitation.org/doi/10.1063/1.5096275> (2019).
  24. Holubec, V. & Ryabov, A. Efficiency at and near maximum power of low-dissipation heat engines. *Physical Review E* **92**, 052125. <https://link.aps.org/doi/10.1103/PhysRevE.92.052125> (5 2015).
  25. Ryabov, A. & Holubec, V. Maximum efficiency of steady-state heat engines at arbitrary power. *Physical Review E* **93**, 050101. <https://link.aps.org/doi/10.1103/PhysRevE.93.050101> (2016).
  26. Holubec, V. & Ye, Z. Maximum efficiency of low-dissipation refrigerators at arbitrary cooling power. *Physical Review E* **101**, 052124. <https://link.aps.org/doi/10.1103/PhysRevE.101.052124> (2020).
  27. Ye, Z. & Holubec, V. Maximum efficiency of absorption refrigerators at arbitrary cooling power. *Physical Review E* **103**, 052125. <https://link.aps.org/doi/10.1103/PhysRevE.103.052125> (2021).

28. Ye, Z. & Holubec, V. Maximum efficiency of low-dissipation heat pumps at given heating load. *Physical Review E* **105**, 024139. <https://link.aps.org/doi/10.1103/PhysRevE.105.024139> (2022).
29. Ryabov, A., Berestneva, E. & Holubec, V. Brownian motion in time-dependent logarithmic potential: Exact results for dynamics and first-passage properties. *The Journal of Chemical Physics* **143**, 114117. <https://doi.org/10.1063/1.4931474> (2015).
30. Ornigotti, L., Ryabov, A., Holubec, V. & Filip, R. Brownian motion surviving in the unstable cubic potential and the role of Maxwell's demon. *Physical review E*. <https://link.aps.org/doi/10.1103/PhysRevE.97.032127> (2018).
31. Šiler, M., Ornigotti, L., Brzobohatý, O., Jákl, P., Ryabov, A., Holubec, V., Zemánek, P. & Filip, R. Diffusing up the Hill: Dynamics and Equipartition in Highly Unstable Systems. *Physical Review Letters* **121**. <https://link.aps.org/doi/10.1103/PhysRevLett.121.230601> (2018).
32. Ryabov, A., Holubec, V. & Berestneva, E. Living on the edge of instability. *Journal of Statistical Mechanics: Theory and Experiment* **2019**, 084014. <https://dx.doi.org/10.1088/1742-5468/ab333f> (2019).
33. Holubec, V. & Ryabov, A. Diverging, but negligible power at Carnot efficiency: Theory and experiment. *Physical Review E* **96**, 062107. <https://link.aps.org/doi/10.1103/PhysRevE.96.062107> (2017).
34. Abiuso, P., Holubec, V., Anders, J., Ye, Z., Cerisola, F. & Perarnau-Llobet, M. Thermodynamics and optimal protocols of multidimensional quadratic Brownian systems. *Journal of Physics Communications* **6**, 063001. <https://dx.doi.org/10.1088/2399-6528/ac72f8> (2022).
35. Chvosta, P., Einax, M., Holubec, V., Ryabov, A. & Maass, P. Energetics and performance of a microscopic heat engine based on exact calculations of work and heat distributions. *Journal of Statistical Mechanics: Theory and Experiment* **2010**, P03002. <https://dx.doi.org/10.1088/1742-5468/2010/03/P03002> (2010).
36. Chvosta, P., Holubec, V., Ryabov, A., Einax, M. & Maass, P. Thermodynamics of two-stroke engine based on periodically driven two-level system. *Physica E: Low-dimensional Systems and Nanostructures* **42**, 472–476. ISSN: 1386-9477. <https://www.sciencedirect.com/science/article/pii/S1386947709002380> (2010).

37. Holubec, V. *Non-equilibrium Energy Transformation Processes. Theoretical Description at the Level of Molecular Structures* Aufl. 2015. ENGLISCH, 142 S. in 1 Teil. ISBN: 3-319-07090-8 (Springer International Publishing, Cham, 2014).
38. Holubec, V. & Ryabov, A. Work and power fluctuations in a critical heat engine. *Physical Review E* **96**, 030102. <https://link.aps.org/doi/10.1103/PhysRevE.96.030102> (2017).
39. Holubec, V. An exactly solvable model of a stochastic heat engine: optimization of power, power fluctuations and efficiency. *Journal of Statistical Mechanics: Theory and Experiment* **2014**, P05022. <https://dx.doi.org/10.1088/1742-5468/2014/05/P05022> (2014).
40. Holubec, V. & Ryabov, A. Fluctuations in heat engines. *Journal of Physics A: Mathematical and Theoretical* **55**, 013001. <https://dx.doi.org/10.1088/1751-8121/ac3aac> (2021).
41. Nostheide, S., Holubec, V., Chvosta, P. & Maass, P. Unfolding kinetics of periodic DNA hairpins. *Journal of Physics: Condensed Matter* **26**, 205102. <https://dx.doi.org/10.1088/0953-8984/26/20/205102> (2014).
42. Holubec, V., Chvosta, P. & Maass, P. Dynamics and energetics for a molecular zipper model under external driving. *Journal of Statistical Mechanics: Theory and Experiment* **2012**, P11009. <https://dx.doi.org/10.1088/1742-5468/2012/11/P11009> (2012).
43. Holubec, V., Dierl, M., Einax, M., Maass, P., Chvosta, P. & Ryabov, A. Asymptotics of work distribution for a Brownian particle in a time-dependent anharmonic potential. *Physica Scripta* **2015**, 014024. <https://dx.doi.org/10.1088/0031-8949/2015/T165/014024> (2015).
44. Holubec, V., Lips, D., Ryabov, A., Chvosta, P. & Maass, P. On asymptotic behavior of work distributions for driven Brownian motion. *The European Physical Journal B* **88**, 340. ISSN: 1434-6036. <https://doi.org/10.1140/epjb/e2015-60635-x> (2015).
45. Chvosta, P., Lips, D., Holubec, V., Ryabov, A. & Maass, P. Statistics of work performed by optical tweezers with general time-variation of their stiffness. *Journal of Physics A: Mathematical and Theoretical* **53**, 275001. <https://dx.doi.org/10.1088/1751-8121/ab95c2> (2020).

46. Holubec, V., Chvosta, P., Einax, M. & Maass, P. Attempt time Monte Carlo: An alternative for simulation of stochastic jump processes with time-dependent transition rates. *Europhysics Letters* **93**, 40003. <https://dx.doi.org/10.1209/0295-5075/93/40003> (2011).
47. Krishnamurthy, S., Ghosh, S., Chatterji, D., Ganapathy, R. & Sood, A. K. A micrometre-sized heat engine operating between bacterial reservoirs. *Nature Physics* **12**, 1134–1138. ISSN: 1745-2481. <https://doi.org/10.1038/nphys3870> (2016).
48. Leonardo, R. D., Angelani, L., Dell’Arciprete, D., Ruocco, G., Iebba, V., Schippa, S., Conte, M. P., Mecerini, F., Angelis, F. D. & Fabrizio, E. D. Bacterial ratchet motors. *Proceedings of the National Academy of Sciences* **107**, 9541–9545. <https://www.pnas.org/doi/abs/10.1073/pnas.0910426107> (2010).
49. Elizabeth Hulme, S., DiLuzio, W. R., Shevkoplyas, S. S., Turner, L., Mayer, M., Berg, H. C. & Whitesides, G. M. Using ratchets and sorters to fractionate motile cells of *Escherichia coli* by length. *Lab Chip* **8**, 1888–1895. <http://dx.doi.org/10.1039/B809892A> (11 2008).
50. Rein, Constantin, Kolář, Martin, Kroy, Klaus & Holubec, Viktor. Force-free and autonomous active Brownian ratchets(a). *Europhysics Letters* **142**, 31001. <https://doi.org/10.1209/0295-5075/accca5> (2023).
51. Holubec, V., Steffenoni, S., Falasco, G. & Kroy, K. Active Brownian heat engines. *Physical Review Research* **2**. <https://journals.aps.org/prresearch/abstract/10.1103/PhysRevResearch.2.043262> (2020).
52. Holubec, V. & Ryabov, A. Maximum efficiency of low-dissipation heat engines at arbitrary power. *Journal of Statistical Mechanics: Theory and Experiment* **2016**, 073204. <https://dx.doi.org/10.1088/1742-5468/2016/07/073204> (2016).
53. Holubec, V. & Ryabov, A. Cycling Tames Power Fluctuations near Optimum Efficiency. *Physical Review Letters* **121**, 120601. <https://link.aps.org/doi/10.1103/PhysRevLett.121.120601> (2018).
54. Ye, Z., Cerisola, F., Abiuso, P., Anders, J., Perarnau-Llobet, M. & Holubec, V. Optimal finite-time heat engines under constrained control. *Physical Review Research* **4**, 043130. <https://link.aps.org/doi/10.1103/PhysRevResearch.4.043130> (2022).
55. Holubec, V., Kroy, K. & Steffenoni, S. Physically consistent numerical solver for time-dependent Fokker-Planck equations. *Physical Review E* **99**, 032117. <https://link.aps.org/doi/10.1103/PhysRevE.99.032117> (2019).



56. Holubec, V. & Marathe, R. Underdamped active Brownian heat engine. *Physical Review E* **102**. <https://journals.aps.org/pre/abstract/10.1103/PhysRevE.102.060101> (2020).
57. Niedenzu, W., Mukherjee, V., Ghosh, A., Kofman, A. G. & Kurizki, G. Quantum engine efficiency bound beyond the second law of thermodynamics. *Nature Communications* **9**, 165. ISSN: 2041-1723. <https://doi.org/10.1038/s41467-017-01991-6> (2018).
58. Pietzonka, P. & Seifert, U. Universal Trade-Off between Power, Efficiency, and Constancy in Steady-State Heat Engines. *Physical Review Letters* **120**, 190602. <https://link.aps.org/doi/10.1103/PhysRevLett.120.190602> (2018).
59. Auschra, S. & Holubec, V. Density and polarization of active Brownian particles in curved activity landscapes. *Physical Review E* **103**, 062604. <https://link.aps.org/doi/10.1103/PhysRevE.103.062604> (2021).
60. Söker, N. A., Auschra, S., Holubec, V., Kroy, K. & Cichos, F. How Activity Landscapes Polarize Microswimmers without Alignment Forces. *Physical Review Letters* **126**, 228001. <https://link.aps.org/doi/10.1103/PhysRevLett.126.228001> (2021).
61. Auschra, S., Holubec, V., Söker, N. A., Cichos, F. & Kroy, K. Polarization-density patterns of active particles in motility gradients. *Physical Review E* **103**, 062601. <https://link.aps.org/doi/10.1103/PhysRevE.103.062601> (2021).
62. Geiseler, A., Hänggi, P., Marchesoni, F., Mulhern, C. & Savel'ev, S. Chemotaxis of artificial microswimmers in active density waves. *Physical Review E* **94**, 012613. <https://link.aps.org/doi/10.1103/PhysRevE.94.012613> (2016).
63. Kosinski, R. J. A literature review on reaction time. *Clemson University* **10**, 337–344. <http://www.cognaction.org/cogs105/readings/clemson.rt.pdf> (2008).
64. Beatus, T., Guckenheimer, J. M. & Cohen, I. Controlling roll perturbations in fruit flies. *Journal of The Royal Society Interface* **12**, 20150075. <https://royalsocietypublishing.org/doi/10.1098/rsif.2015.0075> (2015).
65. Ristroph, L., Ristroph, G., Morozova, S., Bergou, A. J., Chang, S., Guckenheimer, J., Wang, Z. J. & Cohen, I. Active and passive stabilization of body pitch in insect flight. *Journal of The Royal Society Interface* **10**, 20130237. <https://royalsocietypublishing.org/doi/10.1098/rsif.2013.0237> (2013).

66. Ristroph, L., Bergou, A. J., Ristroph, G., Coumes, K., Berman, G. J., Guckenheimer, J., Wang, Z. J. & Cohen, I. Discovering the flight autostabilizer of fruit flies by inducing aerial stumbles. *Proceedings of the National Academy of Sciences* **107**, 4820–4824. <https://www.pnas.org/doi/full/10.1073/pnas.1000615107> (2010).
67. Pomeroy, H. & Heppner, F. Laboratory determination of startle reaction time of the starling (*Sturnus vulgaris*). *Animal Behaviour* **25**, 720–725. [https://doi.org/10.1016/0003-3472\(77\)90121-X](https://doi.org/10.1016/0003-3472(77)90121-X) (1977).
68. Eaton, R. C., Bombardieri, R. A. & Meyer, D. L. The Mauthner-initiated startle response in teleost fish. *Journal of Experimental Biology* **66**, 65–81. <https://doi.org/10.1242/jeb.66.1.65> (1977).
69. Eaton, R. C. *Neural mechanisms of startle behavior* (Springer Science & Business Media, 1984).
70. Lenz, P. & Hartline, D. Reaction times and force production during escape behavior of a calanoid copepod, *Undinula vulgaris*. *Marine Biology* **133**, 249–258. <https://doi.org/10.1007/s002270050464> (1999).
71. Segall, J. E., Block, S. M. & Berg, H. C. Temporal comparisons in bacterial chemotaxis. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 8987–8991. ISSN: 0027-8424. <https://doi.org/10.1073/pnas.83.23.8987> (1986).
72. Wang, X., Chen, P.-C., Kroy, K., Holubec, V. & Cichos, F. Spontaneous vortex formation by microswimmers with retarded attractions. *Nature Communications* **14**. <https://doi.org/10.1038/s41467-022-35427-7> (2023).
73. Zhong, Q. *Robust Control of Time-delay Systems* ISBN: 9781846282652. <https://books.google.cz/books?id=loj-eJ6cPAwC> (Springer London, 2006).
74. Ott, E., Grebogi, C. & Yorke, J. A. Controlling chaos. *Physical Review Letters* **64**, 1196–1199. <https://link.aps.org/doi/10.1103/PhysRevLett.64.1196> (1990).
75. Pyragas, K. Continuous control of chaos by self-controlling feedback. *Physics Letters A* **170**, 421–428. ISSN: 0375-9601. <https://www.sciencedirect.com/science/article/pii/0375960192907458> (1992).
76. Orosz, G., Wilson, R. E., Szalai, R. & Stépán, G. Exciting traffic jams: Nonlinear phenomena behind traffic jam formation on highways. *Physical Review E* **80**, 046205. <https://link.aps.org/doi/10.1103/PhysRevE.80.046205> (2009).

77. Mijalkov, M., McDaniel, A., Wehr, J. & Volpe, G. Engineering Sensorial Delay to Control Phototaxis and Emergent Collective Behaviors. *Physical Review X* **6**, 011008. <https://link.aps.org/doi/10.1103/PhysRevX.6.011008> (2016).
78. Leyman, M., Ogemark, F., Wehr, J. & Volpe, G. Tuning phototactic robots with sensorial delays. *Physical Review E* **98**, 052606. <https://link.aps.org/doi/10.1103/PhysRevE.98.052606> (2018).
79. Piwowarczyk, R., Selin, M., Ihle, T. & Volpe, G. Influence of sensorial delay on clustering and swarming. *Physical Review E* **100**, 012607. <https://link.aps.org/doi/10.1103/PhysRevE.100.012607> (2019).
80. Holubec, V., Geiss, D., Loos, S. A. M., Kroy, K. & Cichos, F. Finite-Size Scaling at the Edge of Disorder in a Time-Delay Vicsek Model. *Physical Review Letters* **127**, 258001. <https://link.aps.org/doi/10.1103/PhysRevLett.127.258001> (2021).
81. Geiß, D., Kroy, K. & Holubec, V. Information conduction and convection in noiseless Vicsek flocks. *Physical Review E* **106**, 014609. <https://link.aps.org/doi/10.1103/PhysRevE.106.014609> (2022).
82. Geiß, D., Kroy, K. & Holubec, V. Signal propagation and linear response in the delay Vicsek model. *Physical Review E* **106**, 054612. <https://link.aps.org/doi/10.1103/PhysRevE.106.054612> (2022).
83. Erban, R., Haškovec, J. & Sun, Y. A Cucker-Smale Model With Noise and Delay. *SIAM Journal on Applied Mathematics* **76**, 1535–1557. ISSN: 00361399. <http://www.jstor.org/stable/44028734> (2023) (2016).
84. Bechhoefer, J. Feedback for physicists: A tutorial essay on control. *Reviews of Modern Physics* **77**, 783–836. <https://link.aps.org/doi/10.1103/RevModPhys.77.783> (2005).
85. Khadka, U., Holubec, V., Yang, H. & Cichos, F. Active particles bound by information flows. *Nature Communications* **9**. <https://www.nature.com/articles/s41467-018-06445-1> (2018).
86. Fränzl, M., Muiños-Landin, S., Holubec, V. & Cichos, F. Fully Steerable Symmetric Thermoplasmonic Microswimmers. *ACS Nano* **15**, 3434–3440. ISSN: 1936-0851. <https://doi.org/10.1021/acsnano.0c10598> (2021).

87. Attanasi, A., Cavagna, A., Del Castello, L., Giardina, I., Melillo, S., Parisi, L., Pohl, O., Rossaro, B., Shen, E., Silvestri, E. & Viale, M. Finite-Size Scaling as a Way to Probe Near-Criticality in Natural Swarms. *Physical Review Letters* **113**, 238102. <https://link.aps.org/doi/10.1103/PhysRevLett.113.238102> (2014).
88. Holubec, V., Ryabov, A., Loos, S. A. M. & Kroy, K. Equilibrium stochastic delay processes. *New Journal of Physics* **24**, 023021. <https://dx.doi.org/10.1088/1367-2630/ac4b91> (2022).
89. Longtin, A. in *Complex time-delay systems* 177–195 (Springer, 2009).
90. Geiss, D., Kroy, K. & Holubec, V. Brownian molecules formed by delayed harmonic interactions. *New Journal of Physics* **21**, 093014. <https://dx.doi.org/10.1088/1367-2630/ab3d76> (2019).
91. Guillouzic, S., L'Heureux, I. & Longtin, A. Small delay approximation of stochastic delay differential equations. *Physical Review E* **59**, 3970–3982. <https://link.aps.org/doi/10.1103/PhysRevE.59.3970> (1999).
92. Bushev, P., Rotter, D., Wilson, A., Dubin, F. ç., Becher, C., Eschner, J., Blatt, R., Steixner, V., Rabl, P. & Zoller, P. Feedback Cooling of a Single Trapped Ion. *Physical Review Letters* **96**, 043003. <https://link.aps.org/doi/10.1103/PhysRevLett.96.043003> (4 2006).
93. Goldwater, D., Stickler, B. A., Martinetz, L., Northup, T. E., Hornberger, K. & Millen, J. Levitated electromechanics: all-electrical cooling of charged nano- and micro-particles. *Quantum Science and Technology* **4**, 024003. <https://dx.doi.org/10.1088/2058-9565/aaf5f3> (2019).
94. Chen, P.-C., Kroy, K., Cichos, F., Wang, X. & Holubec, V. *Active particles with delayed attractions form quaking crystallites* 2023.
95. Selmke, M., Khadka, U., Bregulla, A. P., Cichos, F. & Yang, H. Theory for controlling individual self-propelled micro-swimmers by photon nudging I: directed transport. *Physical Chemistry Chemical Physics* **20**, 10502–10520. <http://dx.doi.org/10.1039/C7CP06559K> (15 2018).
96. Selmke, M., Khadka, U., Bregulla, A. P., Cichos, F. & Yang, H. Theory for controlling individual self-propelled micro-swimmers by photon nudging II: confinement. *Physical Chemistry Chemical Physics* **20**, 10521–10532. <http://dx.doi.org/10.1039/C7CP06560D> (15 2018).

97. Sun, Y., Lin, W. & Erban, R. Time delay can facilitate coherence in self-driven interacting-particle systems. *Physical Review E* **90**, 062708. <https://link.aps.org/doi/10.1103/PhysRevE.90.062708> (6 2014).
98. Acebrón, J. A., Bonilla, L. L., Pérez Vicente, C. J., Ritort, F. & Spigler, R. The Kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of Modern Physics* **77**, 137–185. <https://link.aps.org/doi/10.1103/RevModPhys.77.137> (1 2005).
99. Strogatz, S. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering* <https://books.google.cz/books?id=NZZDnQEACAAJ> (Westview, 2000).
100. Insperger, T. On the Approximation of Delayed Systems by Taylor Series Expansion. *Journal of Computational and Nonlinear Dynamics* **10**. ISSN: 1555-1415. <https://doi.org/10.1115/1.4027180> (2015).
101. Hänggi, P., Talkner, P. & Borkovec, M. Reaction-rate theory: fifty years after Kramers. *Reviews of Modern Physics* **62**, 251–341. <https://link.aps.org/doi/10.1103/RevModPhys.62.251> (2 1990).
102. Watkins, C. J. *Learning from Delayed Rewards*. Cambridge, UK, King's College (Dissertation, 1989).
103. Romanczuk, P. & Salbreux, G. Optimal chemotaxis in intermittent migration of animal cells. *Physical Review E* **91**, 042720. <https://link.aps.org/doi/10.1103/PhysRevE.91.042720> (4 2015).
104. Diz-Muñoz, A., Romanczuk, P., Yu, W., Bergert, M., Ivanovitch, K., Salbreux, G., Heisenberg, C.-P. & Paluch, E. K. Steering cell migration by alternating blebs and actin-rich protrusions. *BMC Biology* **14**, 74. ISSN: 1741-7007. <https://doi.org/10.1186/s12915-016-0294-x> (2016).
105. Ginelli, F. The Physics of the Vicsek model. *The European Physical Journal Special Topics* **225**, 2099–2117. ISSN: 1951-6401. <https://doi.org/10.1140/epjst/e2016-60066-8> (2016).
106. Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I. & Shochet, O. Novel Type of Phase Transition in a System of Self-Driven Particles. *Physical Review Letters* **75**, 1226–1229. <https://link.aps.org/doi/10.1103/PhysRevLett.75.1226> (6 1995).

107. Chaté, H., Ginelli, F., Grégoire, G. & Raynaud, F. Collective motion of self-propelled particles interacting without cohesion. *Physical Review E* **77**, 046113. <https://link.aps.org/doi/10.1103/PhysRevE.77.046113> (4 2008).
108. Halperin, B. I. & Hohenberg, P. C. Scaling Laws for Dynamic Critical Phenomena. *Physical Review* **177**, 952–971. <https://link.aps.org/doi/10.1103/PhysRev.177.952> (2 1969).
109. Hohenberg, P. C. & Halperin, B. I. Theory of dynamic critical phenomena. *Reviews of Modern Physics* **49**, 435–479. <https://link.aps.org/doi/10.1103/RevModPhys.49.435> (3 1977).
110. Cavagna, A., Giardina, I. & Grigera, T. S. The physics of flocking: Correlation as a compass from experiments to theory. *Physics Reports* **728**, 1–62. ISSN: 0370-1573. <https://www.sciencedirect.com/science/article/pii/S0370157317303575> (2018).

## 6 List of discussed Papers

The following papers are discussed in this habilitation thesis. The best five of these papers, according to my personal taste, are marked with an asterisk (\*).

1. Holubec, V. & Ryabov, A. Maximum efficiency of low-dissipation heat engines at arbitrary power. *Journal of Statistical Mechanics: Theory and Experiment* **2016**, 073204. <https://dx.doi.org/10.1088/1742-5468/2016/07/073204> (2016).
2. \*Holubec, V. & Ryabov, A. Cycling Times Power Fluctuations near Optimum Efficiency. *Physical Review Letters* **121**, 120601. <https://link.aps.org/doi/10.1103/PhysRevLett.121.120601> (2018).
3. \*Ye, Z., Cerisola, F., Abiuso, P., Anders, J., Perarnau-Llobet, M. & Holubec, V. Optimal finite-time heat engines under constrained control. *Physical Review Research* **4**, 043130. <https://link.aps.org/doi/10.1103/PhysRevResearch.4.043130> (2022).
4. Holubec, V., Kroy, K. & Steffenoni, S. Physically consistent numerical solver for time-dependent Fokker-Planck equations. *Physical Review E* **99**, 032117. <https://link.aps.org/doi/10.1103/PhysRevE.99.032117> (2019).
5. Holubec, V., Steffenoni, S., Falasco, G. & Kroy, K. Active Brownian heat engines. *Physical Review Research* **2**. <https://journals.aps.org/prresearch/abstract/10.1103/PhysRevResearch.2.043262> (2020).
6. Holubec, V. & Marathe, R. Underdamped active Brownian heat engine. *Physical Review E* **102**. <https://journals.aps.org/pre/abstract/10.1103/PhysRevE.102.060101> (2020).
7. Holubec, V., Ryabov, A., Loos, S. A. M. & Kroy, K. Equilibrium stochastic delay processes. *New Journal of Physics* **24**, 023021. <https://dx.doi.org/10.1088/1367-2630/ac4b91> (2022).
8. Auschra, S. & Holubec, V. Density and polarization of active Brownian particles in curved activity landscapes. *Physical Review E* **103**, 062604. <https://link.aps.org/doi/10.1103/PhysRevE.103.062604> (2021).

9. \*Söker, N. A., Auschra, S., Holubec, V., Kroy, K. & Cichos, F. How Activity Landscapes Polarize Microswimmers without Alignment Forces. *Physical Review Letters* **126**, 228001. <https://link.aps.org/doi/10.1103/PhysRevLett.126.228001> (2021).
10. Auschra, S., Holubec, V., Söker, N. A., Cichos, F. & Kroy, K. Polarization-density patterns of active particles in motility gradients. *Physical Review E* **103**, 062601. <https://link.aps.org/doi/10.1103/PhysRevE.103.062601> (2021).
11. Rein, Constantin, Kolář, Martin, Kroy, Klaus & Holubec, Viktor. Force-free and autonomous active Brownian ratchets(a). *Europhysics Letters* **142**, 31001. <https://doi.org/10.1209/0295-5075/accca5> (2023).
12. Khadka, U., Holubec, V., Yang, H. & Cichos, F. Active particles bound by information flows. *Nature Communications* **9**. <https://www.nature.com/articles/s41467-018-06445-1> (2018).
13. Geiss, D., Kroy, K. & Holubec, V. Brownian molecules formed by delayed harmonic interactions. *New Journal of Physics* **21**, 093014. <https://dx.doi.org/10.1088/1367-2630/ab3d76> (2019).
14. \*Muiños-Landin, S., Fischer, A., Holubec, V. & Cichos, F. Reinforcement learning with artificial microswimmers. *Science Robotics* **6**, eabd9285. <https://www.science.org/doi/abs/10.1126/scirobotics.abd9285> (2021).
15. Wang, X., Chen, P.-C., Kroy, K., Holubec, V. & Cichos, F. Spontaneous vortex formation by microswimmers with retarded attractions. *Nature Communications* **14**. <https://doi.org/10.1038/s41467-022-35427-7> (2023).
16. Chen, P.-C., Kroy, K., Cichos, F., Wang, X. & Holubec, V. *Active particles with delayed attractions form quaking crystallites* 2023.
17. \*Holubec, V., Geiss, D., Loos, S. A. M., Kroy, K. & Cichos, F. Finite-Size Scaling at the Edge of Disorder in a Time-Delay Vicsek Model. *Physical Review Letters* **127**, 258001. <https://link.aps.org/doi/10.1103/PhysRevLett.127.258001> (2021).
18. Geiß, D., Kroy, K. & Holubec, V. Information conduction and convection in noiseless Vicsek flocks. *Physical Review E* **106**, 014609. <https://link.aps.org/doi/10.1103/PhysRevE.106.014609> (2022).
19. Geiß, D., Kroy, K. & Holubec, V. Signal propagation and linear response in the delay Vicsek model. *Physical Review E* **106**, 054612. <https://link.aps.org/doi/10.1103/PhysRevE.106.054612> (2022).



## **7 Original Papers**

The original papers discussed in this thesis are reprinted below.



## Active Brownian heat engines

Viktor Holubec<sup>1,2,\*</sup>, Stefano Steffenoni<sup>1,3</sup>, Gianmaria Falasco<sup>1,4</sup> and Klaus Kroy<sup>1</sup><sup>1</sup>*Institut für Theoretische Physik, Universität Leipzig, Postfach 100 920, D-04009 Leipzig, Germany*<sup>2</sup>*Charles University, Faculty of Mathematics and Physics, Department of Macromolecular Physics, V Holešovičkách 2, CZ-180 00 Praha, Czech Republic*<sup>3</sup>*Max Planck Institute for Mathematics in the Sciences, Inselstr. 22, D-04103 Leipzig, Germany*<sup>4</sup>*Complex Systems and Statistical Mechanics, Department of Physics and Materials Science, University of Luxembourg, L-1511 Luxembourg, Luxembourg*

(Received 28 January 2020; revised 9 September 2020; accepted 2 November 2020; published 19 November 2020)

When do nonequilibrium forms of disordered energy qualify as heat? We address this question in the context of cyclically operating heat engines feeding on nonequilibrium energy reservoirs that defy the zeroth law of thermodynamics into work. To consistently address a nonequilibrium bath as a heat bath in the sense of the second law of thermodynamics requires the existence of a precise mapping to an equivalent cycle with an equilibrium bath at a time-dependent effective temperature. We identify the most general setup for which this can generically be ascertained and thoroughly discuss an analytically tractable, experimentally relevant scenario: a Brownian particle confined in a periodically modulated harmonic potential and coupled to some nonequilibrium bath of variable activity. We deduce formal limitations for its thermodynamic performance, including maximum efficiency, efficiency at maximum power, and maximum efficiency at fixed power. The results can guide the design of new micromachines and clarify how much these can outperform passive-bath designs, which has been a debated issue for recent experimental realizations. To illustrate the practical implications of the general principles for quasistatic and finite-rate protocols, we further analyze a specific realization of such an active heat engine based on the paradigmatic active Brownian particle (ABP) model. This reveals some nonintuitive features of the explicitly computed dynamical effective temperature, illustrates various conceptual and practical limitations of the effective-equilibrium mapping, and clarifies the operational relevance of various coarse-grained measures of dissipation.

DOI: [10.1103/PhysRevResearch.2.043262](https://doi.org/10.1103/PhysRevResearch.2.043262)

## I. INTRODUCTION

The study of heat engines is as old as the industrialization of the world. Its practical importance has prompted physicists and engineers to persistently improve their experiments and theories to eventually establish the consistent theoretical framework of classical thermodynamics. It allows to quantify very generally, on a phenomenological level, how work is transformed to heat, and to what extent this process can be reversed. Heat is the most abundant but least valuable form of energy, namely “disordered” energy dispersed among unresolved degrees of freedom. And turning it into the coherent accessible form called work has been a central aim since the days of Carnot, Stirling, and other pioneers, after whom some common engine designs have been named.

Recent advances in technology have allowed and also required to extend this success story into two new major directions. First, towards microscopic designs that are so small

that their operation becomes stochastic rather than deterministic [1–5]. And secondly to cases where the degrees of freedom of the heat bath are themselves driven far from equilibrium, which potentially matters for small systems operating in a biological context, e.g., inside living cells or motile bacterial colonies [6].

The analysis of small systems requires an extension of the theory and basic notions of classical thermodynamics to stochastic dynamics, which goes under the name of stochastic thermodynamics [7–10]. It seeks to define heat, work, and entropy on the level of individual stochastic trajectories. The theory recovers the laws of thermodynamics for ensemble-averaged quantities but allows to additionally quantify the probability of rare large fluctuations [10]. Along these lines, many experimental [11–16] and theoretical [17–26] studies have recently been devoted to microscopic thermodynamic cycles. In this field, Brownian heat engines play a paradigmatic role [14–16,25,26]. They are usually based on a diffusing colloidal particle that represents the working substance. Its solvent provides a natural equilibrium heat bath, and a time-dependent confinement potential can be realized by optical tweezers [14,15,27].

Over the last few years, increasing effort has also been devoted to the second mentioned extension of the classical designs, namely to endow quantum [28,29] and classical (colloidal) [6,30–32] heat engines with so-called

\*viktor.holubec@mff.cuni.cz

“active” (nonequilibrium) baths. Some paradigmatic realizations of such active baths are provided by suspensions of self-propelling bacteria or synthetic microswimmers [6,33]. Remarkably, they are driven far from equilibrium on the level of the individual particles—not merely by externally imposed overall boundary or body forces. The corresponding “active heat engines” utilizing such baths can outperform classical designs by evading the zeroth law of thermodynamics, which would require interacting degrees of freedom to mutually thermalize. Engines that exploit this unconventional property can operate between hugely different (effective) temperatures and thereby at unconventionally high efficiencies, without risking the evaporation or freezing of the laboratory. While technically potentially desirable, the lack of thermalization jeopardizes the unambiguous distinction of heat from work (roughly speaking, as the disordered or contagiously spreading versus the concentrated, coherently accessible form of energy). Which means that one has to resort to the second law of thermodynamics, alone, for that purpose. But active heat engines have even prominently been claimed to transcend the universal performance bounds set by the second law [6], a notion that is critically examined below.

In the following, we first provide a general discussion of heat engines in contact with arbitrary nonequilibrium reservoirs. We simply refer to them as active heat engines, since active reservoirs represent a paradigmatic example of a nonequilibrium bath. The main claims are exemplified by an analytical discussion of a still quite general limiting case: the linear theory for a Brownian heat engine with a nonequilibrium bath. In particular, in Sec. IV, we derive the effective temperature (14) for this class of models, thereby establishing, as a main result, the explicit mapping of the active-bath engine to a classical engine with an equilibrium bath achieving the same thermodynamic output and performance. Finally, to illustrate and further elucidate our general results and conclusions, Sec. VI provides a detailed analysis of a specific realization of such linear active Brownian heat engine based on the standard minimal model for active particle suspensions, namely the so-called ABP (“active Brownian particle”) model [34]. To facilitate the distinction between the general linear theory and the exemplifying model, we refer to the latter by the reminiscent acronym ABE (“active Brownian engine”), in the following. It still allows for several alternative physical interpretations [35,36], detailed in Sec. VII. Their dissimilar contributions to the entropy production denounce the nonequilibrium character of the engine that persists during nominally reversibly operation. In Sec. VIII, we analyze the quasistatic and finite-time performance of the model and highlight some peculiarities of the effective temperature. For better readability, various technical details have been deferred to an Appendix.

## II. ACTIVE HEAT ENGINES

### A. Work-to-work versus heat-to-work conversion

Speaking of nonequilibrium heat baths that defy the zeroth law, an important qualification needs to be made as to how their energy is accessed, if it is no more obliged to spread indiscriminately by itself. Any thermodynamic entity

that can qualify as a nonequilibrium bath should be in a nonequilibrium steady state while being able to exchange some disordered form of energy with the so-called system or working medium. Importantly, the exchanged energy should not entirely be work in disguise. In other words, the internal nonequilibrium structure of the bath should not entirely be resolved by the device that feeds on it, in order to allow us to speak of an engine that operates by *heat-to-work conversion*. Yet, to exploit the advantages of active baths relative to conventional equilibrium baths, practical designs often rectify at least some of the bath energy by directly tapping some of the internal thermodynamic fluxes that are responsible for the nonequilibrium character of the bath. Typical examples are provided by so called *steady-state designs*, such as various flywheels and ratchet-like devices in active suspensions [37–41]. They geometrically rectify the persistent motion of active particles and thereby extract work from their (collective) motion against an external load [42,43]. Such rectification is reminiscent of the action of a Maxwell demon, but can be less sophisticated, since it feeds on palpable nonequilibrium fluxes rather than equilibrium fluctuations, which average to zero. Yet, it is not immediately obvious whether to classify it as *heat-to-work conversion* or *work-to-work conversion*. Especially if the rectified nonequilibrium flux in the active bath is driven mechanically or chemically, one is tempted to argue that the rectification should be addressed as a form of work-to-work conversion. However, any heat engine ultimately draws its power from a nonequilibrium thermodynamic flux, namely, a heat flux. So, in particular if the rectified flux in the active bath is ultimately caused by a temperature gradient, such as in hot Brownian motion or hot microswimmers [44–47], the notion of heat-to-work conversion in the spirit of two-temperature (Feynman-Smoluchowski) ratchets [48–53] also seems very justifiable.

In the following, we focus on the operational scheme of traditional heat engines, which cannot extract work from a single bath with time-independent parameters, and are therefore operated cyclically. We assume that the working medium of the engine is a small (i.e., Brownian) system described by an (overdamped) Hamiltonian  $\mathcal{H}(\mathbf{k}, \mathbf{x})$ , which depends on a set of stochastic coordinates  $\mathbf{x} = (x_1, \dots, x_{N_x})$  and a set of externally controlled parameters  $\mathbf{k} = (k_1, \dots, k_{N_k})$ , measuring, for example, height of a weight in a gravitational field. These parameters are used to extract work (“ordered” energy in the sense of the external handling) from the engine or to feed it from an external work source. Examples from this class of *cyclic engines* are various colloidal engines immersed in active fluids such as bacteria suspensions (see Ref. [6] for an experiment and Refs. [30,32,54–56] for theoretical works). We argue that, for these machines, there is a well-defined regime, where energy extracted from the nonequilibrium bath and transformed to work can unambiguously and quantitatively be interpreted as (a generalized form of) heat—namely, if there exists a precise mapping to an equivalent setup with an equilibrium bath at a suitable (finite) time-dependent effective temperature  $T_{\text{eff}}(t)$ . Due to the nonequilibrium character of the bath, such engines can still exploit similar “rectification loopholes” as the mentioned steady-state ratchets. But the effect is then fully quantified by  $T_{\text{eff}}(t)$ , which can, in a precise sense, interpolate between the limits of pure heat-to-work and

work-to-work conversion, attained for  $T_{\text{eff}}(t) \equiv T = \text{constant}$  and  $\max T_{\text{eff}}(t) - \min T_{\text{eff}}(t) \rightarrow \infty$ , respectively.

### B. Energetics and efficiency of cyclic heat engines

For arbitrary dynamics, the instantaneous internal energy  $\mathcal{H}(t) = \mathcal{H}(\mathbf{k}(t), \mathbf{x}(t))$  of the working medium of the engine changes as

$$\frac{d}{dt} \mathcal{H}(t) = \dot{w}(t) + \dot{q}(t) \quad (1)$$

with

$$\dot{w}(t) = \sum_{i=1}^{N_k} \frac{\partial}{\partial k_i} \mathcal{H}(t) \dot{k}_i(t), \quad (2)$$

$$\dot{q}(t) = \sum_{i=1}^{N_x} \frac{\partial}{\partial x_i} \mathcal{H}(t) \dot{x}_i(t) = \frac{d}{dt} \mathcal{H}(t) - \dot{w}(t). \quad (3)$$

The contribution  $\dot{w}$  corresponds to a change of the externally controlled parameters  $\mathbf{k} = \mathbf{k}(t)$  and thus it is naturally identified as work delivered to the working medium from the external work reservoir [6,9,10,14,15,25,26,56,57]. The remaining part of the energy change,  $q$ , is then acquired from what are the heat reservoirs according to standard heat-engine nomenclature. With the above-mentioned potential caveats in mind, one might speak more generally of energy influx into the working medium from these reservoirs. Conditions when this influx has thermodynamic properties of heat, namely when an effective temperature  $T_{\text{eff}}$  exists such that  $\dot{S}_{\text{R}}^{\text{eff}}(t) = \dot{q}(t)/T_{\text{eff}}(t)$  is a reasonable second-law entropy change in the bath, are discussed in the next section. From here on, we assume that such an effective temperature can be defined and simply call  $\dot{q}(t)$  the heat flux.

The above defined work and heat transfers are stochastic quantities that fluctuate due to the stochastic nature of the coordinates  $\mathbf{x}$ . One is often interested in their mean values both over a certain span of time and over the stochastic ensemble. Upon integration over time and ensemble averaging, the average total work exchanged between the system and its environment during the time interval  $(t_i, t_f)$  is given by

$$W(t_i, t_f) = \int_{t_i}^{t_f} dt \dot{W}(t) = \int_{t_i}^{t_f} dt \langle \dot{w}(t) \rangle \quad (4)$$

and the corresponding total heat by

$$Q(t_i, t_f) = \int_{t_i}^{t_f} dt \dot{Q}(t) = \int_{t_i}^{t_f} dt \langle \dot{q}(t) \rangle \quad (5)$$

From now on, we assume that the parameters of the Hamiltonian are varied periodically, with period  $t_p$ . The (ensemble-averaged) states of the system and the reservoirs are assumed to eventually attain a time-periodic limit cycle with the same period. If not explicitly written otherwise, all variables below will be evaluated on this limit cycle.

The net average work performed or *output work* by the engine per cycle is, with the above definitions, expressed as

$$W_{\text{out}} = -W(0, t_p). \quad (6)$$

As the *input heat*,  $Q_{\text{in}}$ , one usually identifies only the heat acquired during those parts of the cycle during which heat on

average flows from the bath into the system [58], i.e., when  $\Theta(\langle \dot{q} \rangle) > 0$ , where  $\Theta$  denotes the Heaviside step function. So we have

$$Q_{\text{in}} = \int_0^{t_p} dt \dot{Q}(t) \Theta[\dot{Q}(t)], \quad (7)$$

which may well differ from  $Q(0, t_p)$ . So while the definition may look a bit awkward, it allows us to write the input heat in a form that is independent of specific details of the driving protocol. For standard thermodynamic cycles such as Carnot or Stirling cycle with a hot and a cold equilibrium heat bath, it recovers the standard expressions for the heat afforded via the hot reservoir (irrespective of the amount of heat taken up by the cold reservoir).

Common measures of performance of a heat engine are its output power  $P$  and efficiency  $\eta$ :

$$P \equiv \frac{W_{\text{out}}}{t_p}, \quad \eta \equiv \frac{W_{\text{out}}}{Q_{\text{in}}}. \quad (8)$$

In accord with earlier works [6,30], the hereby defined efficiency measures how efficiently the engine transforms the energy  $Q_{\text{in}}$  actually acquired from the bath into work  $W_{\text{out}}$ . It doesn't measure how efficiently it collects energy from the nonequilibrium bath, which would be important if one would want to take into account also the housekeeping energy flux  $\dot{q}_{\text{hk}}$  that maintains nonequilibrium steady state of the nonequilibrium bath (see Fig. 1). In most practical settings, the housekeeping contribution would completely overshadow  $Q_{\text{in}}$ —rendering the efficiency tiny and dependent on the technical realization of the bath, which is usually not desirable. Moreover, the knowledge about the bath that would allow us to evaluate the housekeeping heat would also allow us to design more sophisticated ways to extract the bath energy than a heat engine. For studies utilizing definitions of efficiency taking into account dissipation caused by the active motion of the individual constituents of an active bath (and thus part of the housekeeping heat), we refer to Refs. [56,59].

If the engine communicates with an equilibrium bath at temperature  $T(t)$ , its efficiency is unambiguously restricted by the first and second law of thermodynamics to obey  $\eta \leq 1$  and  $\eta \leq \eta_{\infty} < \eta_C = 1 - \min(T)/\max(T) \leq 1$ , respectively. Here,  $\eta_{\infty}$  refers to the value obtained upon infinitely slow, reversible operation, and  $\eta_C$  is the Carnot efficiency. On the level of stochastic heat and work transfers, these constraints are moreover reflected by various fluctuation theorems for the corresponding probability distributions [10,60–63].

Allowing for an (additional) active bath, the interpretation of the conventional formalism may require some extra considerations. First, one can exploit the nonequilibrium state of the bath to effectively isolate certain degrees of freedom from the rest of the setup, thereby effectively circumventing the zeroth law. This allows one to emulate unusually high or low temperatures (for these degrees of freedom) without contaminating many others, and thus to reach exceptionally high efficiencies. An example would be a hot Brownian swimmer, which is actually laser heated relative to the solvent by only a few Kelvin, while executing a random motion as if it had been heated by thousands of Kelvin, which would technically be much more difficult to achieve for a conventional equilibrium bath obeying the zeroth law [47].

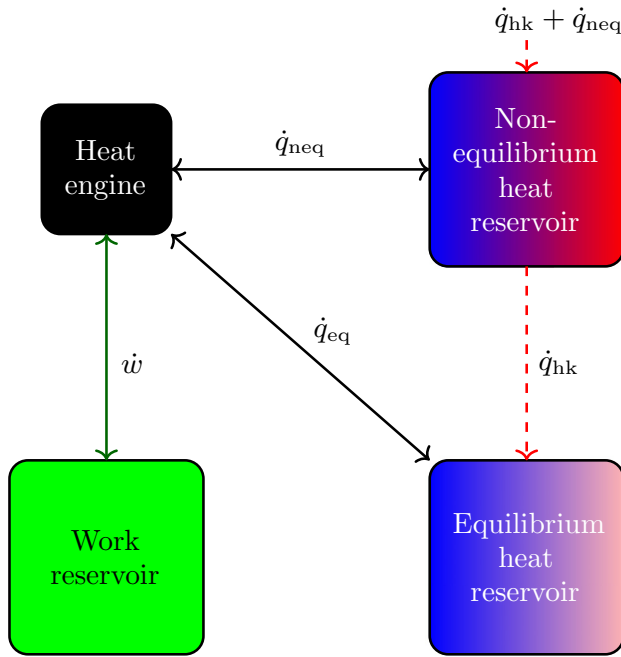


FIG. 1. Cyclic heat engine transforming the heat flux  $\dot{q} = \dot{q}_{\text{neq}} + \dot{q}_{\text{eq}}$  from a nonequilibrium (neq) and equilibrium (eq) heat reservoir into usable power  $\dot{w}$ . The corresponding energy fluxes relevant for the engine's operation are depicted by arrows. The dashed arrow depicts the housekeeping heat flux,  $\dot{q}_{\text{hk}}$ , flowing from the active bath to the infinite equilibrium reservoir, which prevents the active bath from overheating. This energy flux and also  $\dot{q}_{\text{neq}}$  are sustained by the energy influx  $\dot{q}_{\text{hk}} + \dot{q}_{\text{neq}}$  into the nonequilibrium bath, which keeps it in a nonequilibrium steady state. In this paper, we discuss the setup where the energy  $\mathcal{H}[\mathbf{k}(t)]$  of the working medium of the engine (e.g., a single trapped colloid) is periodically modulated by an external control parameter  $\mathbf{k}(t)$  and the temperature/activity (technically: the noise intensities) of the two heat reservoirs.

Secondly, one can extract net work from a single steady-state heat bath at constant activity, thus apparently beating the second law [56]. For this, one needs at least two control parameters, though, since quasistatically operating engines with a single control parameter  $\mathbf{k} = k$  allow the output power to be integrated  $\langle \dot{w}(t) \rangle dt = \langle \partial \mathcal{H} / \partial k \rangle dk = f(k) dk$ . Here,  $f(k)$  depends solely on  $k$  since all other parameters are held constant. A physically sensible one-dimensional function  $f(k)$  can always be written as a derivative  $f(k) \equiv dg(k)/dk$ . The output work per cycle then reads  $W_{\text{out}} = \int_0^{t_p} dt \langle \dot{w}(t) \rangle = g(k(t_p)) - g(k(0)) = 0$ , because of the periodicity of  $k(t)$ . This result is valid regardless of the properties of the steady-state bath, except that for nonquasistatic protocols the output work will be negative, due to finite-time losses. For two (and more) parameters, on the other hand,  $\langle \dot{w}(t) \rangle dt = \sum_{i=1}^{N_k} f_i[\mathbf{k}(t)] dk_i$ . Hence  $W_{\text{out}} = 0$  now only holds if an integrability condition is satisfied, namely that a function  $g[\mathbf{k}(t)]$  exists such that  $f_i = \partial g / \partial k_i$  is a gradient and thus  $\langle \dot{w}(t) \rangle dt = \nabla g \cdot d\mathbf{k}$ . Otherwise, internal currents may indeed allow the extraction of work from a single nonequilibrium bath at constant activity [56].

In both of the above examples of how to “beat” classical constraints on the performance of heat engines an equivalent

of temperature is seen to play a crucial role, namely the one characterizing the Brownian motion of the microswimmer and the one characterizing the constant activity of the active bath, respectively. Indeed, as we lay out in the following paragraph and in even greater detail in the remainder of this contribution, this notion can sometimes be made fully quantitative and then be used to explicitly compute meaningful efficiencies for active heat engines.

### C. Dynamic effective temperature

The crucial step is to construct a mapping for the power  $P$  and efficiency  $\eta$  of a heat engine in contact with a nonequilibrium bath to that of a heat engine in contact with an equilibrium bath. This can be achieved if one can define a temperature in the sense of the second law of thermodynamics [47,64]. Which is the case if, for a given protocol for varying the control parameters  $\mathbf{k}(t)$ , the energy fluxes  $\langle \dot{w}(t) \rangle$  and  $\langle \dot{q}(t) \rangle$  for the heat engine in contact with the nonequilibrium bath agree with those for a virtual heat engine in contact with an equivalent equilibrium bath maintained at a time-dependent temperature  $T_{\text{eff}}(t)$ . This then allows the application of known results for heat engines with equilibrium baths to meaningfully define and assess the performance of active engines. Per construction, their efficiencies are then bounded by the second law. (Further consequences of the mapping are discussed for a specific example in Sec. IV.)

The most general situation for which an appropriate effective temperature can always be found is when one can write the Hamiltonian in the form  $\mathcal{H} = k(t)h(\mathbf{x})$ , with an arbitrary function  $h(\mathbf{x})$  diverging at  $|\mathbf{x}| \rightarrow \infty$ . Then we have  $\langle \dot{w} \rangle = \dot{k}(t)f(t)$  and  $\langle \dot{q} \rangle = k(t)\dot{f}(t)$ , with  $f(t) = \langle h(\mathbf{x}) \rangle$ . In general, for a nonequilibrium bath described by a set of functions  $\mathbf{b}(t)$ ,  $f(t)$  is a functional of the external protocol  $k(t)$  and the bath parameters  $\mathbf{b}(t)$ , say  $f(t) = f_{\text{neq}}[\{k(t), \mathbf{b}(t)\}_{t=0}^{t_p}]$ . If the equilibrium mapping exists, this functional can be written as  $f(t) = f_{\text{eq}}[\{k(t), T_{\text{eff}}(t), \gamma(t)\}_{t=0}^{t_p}]$ , where all relevant effects of the bath parameters have been subsumed into the dynamic effective temperature  $T_{\text{eff}}(t)$  and possibly also a time-dependent friction  $\gamma(t)$ .

These two quantities are implicitly given by the functional identity  $f_{\text{eq}}(t) \equiv f_{\text{neq}}(t)$ , which has to be solved to derive their explicit form. In the next section, we discuss a specific scenario where this can be always achieved analytically. In general, our physical intuition suggests that the equation  $f_{\text{eq}}(t) \equiv f_{\text{neq}}(t)$  has at least one solution. While it may be difficult to rigorously prove its existence and uniqueness on such a general level, what matters most with respect to the thermodynamic performance is the case of quasistatic driving. In this limit,  $f_{\text{eq}}(t)$  is only a function of  $k(t)$  and  $T_{\text{eff}}(t)$ , which can be determined by calculating the average  $f_{\text{eq}} = \langle h(\mathbf{x}) \rangle$  over the Gibbs canonical distribution  $p(\mathbf{x}, t) = \exp[-k(t)h(\mathbf{x})/T_{\text{eff}}(t)]/Z$ , where  $Z$  is the normalization constant, and we have set the Boltzmann constant to unity,  $k_B \rightarrow 1$ , measuring energies in Kelvin. The resulting equation  $f_{\text{eq}}(t) \equiv f_{\text{neq}}(t)$  for  $T_{\text{eff}}(t)$  can then always be solved, because any value of the average  $\langle h(\mathbf{x}) \rangle$  (taken over the Gibbs distribution) can be assigned an effective temperature  $T_{\text{eff}}(t)$  varying between zero and infinity, thereby exhausting all possible

values of the average obtainable with an arbitrary nonequilibrium distribution.

For more general Hamiltonians of the form  $\mathcal{H} = k(t)h_1(\mathbf{x}) + h_2(\mathbf{x})$ , or even more complicated, we have  $\langle \dot{w} \rangle = \dot{k}f_1$  and  $\langle \dot{q} \rangle = k\dot{f}_1 + \dot{f}_2$ , with  $f_i = \langle h_i(\mathbf{x}) \rangle$  again being functionals of the driving and bath parameters. These two functionals need not consistently determine a single function  $T_{\text{eff}}$ , solving  $f_{i,\text{eq}} \equiv f_{i,\text{neq}}$  for both  $i = 1, 2$ . Then, equivalent cycles with equilibrium baths might still exist under specific circumstances, but they are not generally guaranteed or generically expected, anymore.

To sum up this introductory section and to answer the question asked at the beginning of our abstract, we note that the notion of heat can unambiguously be generalized to nonequilibrium situations where the zeroth law does not hold, but it is tied to the operational definition of an effective temperature in the sense of the second law. In other words, one has to require that the only energy that can be extracted from a nonequilibrium heat bath is of the disordered form that comes with a reduced work efficiency. Otherwise, one actually deals with some sort of work reservoir in disguise. Active heat engines coupled to such nonequilibrium baths and Hamiltonians proportional to a single control parameter can always be reinterpreted in terms of equivalent engines in contact with equilibrium baths, at some dynamic effective temperature. Their thermodynamic properties thus obey standard-second law bounds, with important consequences for the interpretation of experimental results.

#### D. Application to experimental data

A relevant real-world realization of a heat engine in contact with a nonequilibrium bath is the bacterial heat engine of Ref. [6]. In this impressive experimental study, a colloidal particle with Cartesian position  $\{x, y\}$  was trapped in a time-dependent harmonic potential,

$$\mathcal{H}(x, y, t) = \frac{1}{2}k(t)\mathbf{r}^2 = \frac{1}{2}k(t)(x^2 + y^2), \quad (9)$$

and immersed in a bath of self-propelled bacteria. Both the trap stiffness  $k(t)$  and the bacterial activity were quasistatically modulated to realize a Stirling-type active heat engine with a cycle composed of two isochoric and two isothermal state changes. These were technically implemented by changing the bacterial activity at constant trap stiffness  $k$  and vice versa, respectively. The ensuing colloid dynamics was observed to converge to a quasistatic limit cycle transforming energy absorbed from the disordered bacterial bath into colloidal work.

The authors measured the work done per cycle as well as the energy (heat) obtained per cycle from the bath and determined the efficiency of the machine as their ratio. From Eq. (9) the time-dependent average system energy reads

$$\langle \mathcal{H} \rangle = \frac{1}{2}k(t)[\sigma_x(t) + \sigma_y(t)] = \frac{1}{2}k(t)\sigma(t), \quad (10)$$

where  $\sigma_x = \langle x^2 \rangle$ ,  $\sigma_y = \langle y^2 \rangle$ , and  $\sigma = \langle \mathbf{r} \cdot \mathbf{r} \rangle$ . Due to the symmetry of the potential, the average particle displacements  $\langle x \rangle$  and  $\langle y \rangle$  vanish during the cycle, so that the mean square displacements  $\sigma_x(t)$  and  $\sigma_y(t)$  also determine the long-time variances for the  $x$  and  $y$  coordinates, respectively.

Based on their analysis of the apparent equipartition temperature  $T_{\text{eff}}(t) \equiv k(t)\sigma(t)/2$ , denoted by  $T_a$  in Ref. [6], its authors concluded that they had realized a Stirling cycle that allowed them to significantly surpass the maximum Stirling efficiency,  $[1 + 1/\ln(k_>/k_<)]$  attained for equilibrium heat baths with an infinite temperature difference ( $k_>$  and  $k_<$  denote maximum and minimum values of  $k(t)$  during the cycle). This extraordinary result was attributed to large non-Gaussian fluctuations in the nonequilibrium bacterial reservoir, which, according to the authors, cannot be captured by an effective temperature.

These conclusions are plainly at odds with the general analysis in the preceding paragraph. To see this, notice that the experimental heat engine corresponds to a Hamiltonian proportional to a single control parameter, for which one can always define an effective temperature so that the conventional bounds on the efficiency apply. Using Eqs. (2) and (3) (employed also in Ref. [6] to evaluate work and heat fluxes into the system), we obtain  $\langle \dot{w}(t) \rangle = \dot{k}(t)\sigma(t)/2$  and  $\langle \dot{q}(t) \rangle = k(t)\dot{\sigma}(t)/2$ . The equivalent heat engine with an equilibrium bath has the bath temperature  $T_{\text{eff}}(t)$ . It thus has the same energy input (heat), as correctly noted in Ref. [6], but also the same energy output (work). Accordingly, if  $T_{\text{eff}}(t)$  evolves along a Stirling cycle, the efficiency  $\eta$  of the active engine, determined by the ratio of output work over afforded heat, is necessarily bounded by the Stirling efficiency. The non-Gaussian fluctuations in the bath indeed affect the output work, input heat, and efficiency of the engine, but only via the mean square displacement  $\sigma$ , hence again via the appropriate effective temperature  $T_{\text{eff}}$ . Assuming that heat and work were accurately measured (which is supported by the correctly measured Stirling efficiency in the case of inactive bacteria), the observation of an efficiency surpassing the maximum value for Stirling engines calls into question the notion that the experimental engine realized a Stirling cycle with respect to  $T_{\text{eff}}$  (see also Ref. [30]). As we demonstrate next, the dynamic effective temperature  $T_{\text{eff}}(t)$  may generally indeed vary in time even while the ambient solvent temperature and the activity remain constant.

### III. LINEAR THEORY: DYNAMICS

Up to this point, we have not specified any particular system dynamics and thus the described results are valid for arbitrary time-evolution of the degrees of freedom  $\mathbf{x}$ . To provide better insight and to show that a nonintuitive behavior of effective temperatures can be expected, this section investigates a specific (but from the point of view of Brownian heat engines still quite generic) exactly solvable class of models. Concretely, we analytically derive the effective temperature for a class of one-parameter engines inspired by the experimental work described above. We detail the mapping to the equilibrium model and its consequences for the thermodynamics of the active heat engine. In particular, we reveal a nontrivial behavior of the effective temperature. This seems to be the first explicit result of its type.

From now on, we specialize our discussion to a heat engine consisting of a colloidal particle confined to a *time-dependent harmonic* potential with an externally controlled stiffness  $k(t)$ , as introduced in Eq. (9). We specify the dynamics by further

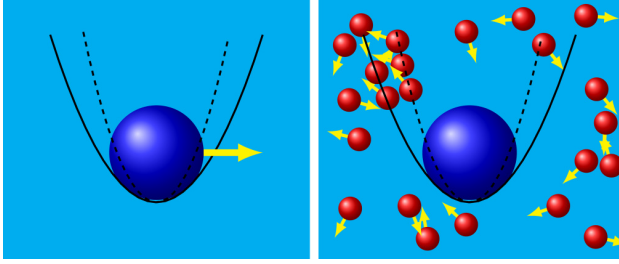


FIG. 2. Schematic designs of microscopic heat engines based on colloids in modulated harmonic traps, playing the roles of the working substance and the movable piston, respectively. (Left) Active particle in a “passive” equilibrium bath. (Right) Passive particle in an “active” nonequilibrium bath composed of energy consuming microswimmers immersed into a passive background fluid. To operate the heat engine, the bath temperature and/or activity as well as the confinement strength are modulated cyclically. Thereby “disordered” energy dispersed in the bath and randomly propelling the colloid against its confinement is concentrated in a degree of freedom that can be externally harnessed to perform (mechanical) work.

requiring that the colloid is immersed in a (possibly) nonequilibrium bath, which couples to it via a drag coefficient  $\mu^{-1}$  and a zero-mean *additive* noise  $\boldsymbol{\eta}(t)$ , so that its position  $\mathbf{r} = (x, y)^T$  obeys the *overdamped linear* Langevin equation

$$\dot{\mathbf{r}} = -\mu k(t)\mathbf{r}(t) + \boldsymbol{\eta}(t). \quad (11)$$

Depending on the noise correlations, which remain to be prescribed and need not be Markovian, this equation can describe various experimentally relevant situations. In Fig. 2, we depict two of them that we discuss further below: namely, an active particle or “microswimmer” immersed in a passive equilibrium bath (left) [65,66], and a (passive) colloid immersed in an active nonequilibrium bath that is itself composed of active particles swimming in a thermal background solvent (right) [57,67–70]. Further examples are provided by devices that share the same formal description on a suitably coarse-grained level, such as noisy electric circuits and similar Langevin systems [71].

In line with such realizations, the trapping potential (9) has the harmonic standard form experimentally created with the help of optical tweezers [6,14,15]. We have also taken advantage of the fact that such experiments are typically designed in a quasi-two-dimensional geometry, in narrow gaps between two glass coverslips. For simplicity, the particle mobility is represented by a constant scalar  $\mu$  and the two-time correlation matrix

$$C_{ij}(t, t') \equiv \langle \eta_i(t)\eta_j(t') \rangle \propto \delta_{ij} \quad (12)$$

of the noise  $\boldsymbol{\eta} = (\eta_x, \eta_y)$  by a diagonal form. Our analysis can of course straightforwardly be generalized to arbitrary dimensions and mobility matrices.

If  $\boldsymbol{\eta}$  in Eq. (11) stands for the white noise, the model provides a good description for existing experimental realizations of Brownian heat engines [14,15]. Their thermodynamics has been thoroughly analyzed in the literature [25,72,73]. An example for an experimental realization of the nonequilibrium-noise version is the active Brownian engine with a bacterial bath [6] discussed in the previous section. The performance

of a quasistatic Stirling heat engine based on the latter design was already nicely analyzed by Zakine *et al.* [30]. Its finite-time performance was numerically investigated in Refs. [32,54,55]. In contrast to these studies, which employ specific protocols, our approach is valid for arbitrary driving protocols at arbitrary speeds.

Our first main result, to be derived in the following, concerns the thermodynamics of the system described by Eq. (11) with arbitrary time-periodic driving  $k(t)$  and with a nonequilibrium noise  $\boldsymbol{\eta}$  with arbitrary time-periodic intensity. We refer to it as the (linear) *active* heat engine, and show that it can be mapped onto the well-investigated model with a passive equilibrium bath [25,72,73], to which we refer as the *passive* model:

$$\dot{\mathbf{r}}(t) = -\mu k(t)\mathbf{r}(t) + \sqrt{2D_{\text{eff}}(t)}\boldsymbol{\xi}(t). \quad (13)$$

Its bath is characterized by the Gaussian white noise  $\boldsymbol{\xi}(t)$  with zero mean,  $\langle \boldsymbol{\xi}(t) \rangle = 0$ , the unit correlation matrix,  $\langle \xi_i(t)\xi_j(t') \rangle = \delta_{ij}\delta(t-t')$ , and a time-dependent (effective) temperature<sup>1</sup>

$$T_{\text{eff}}(t) = \frac{D_{\text{eff}}(t)}{\mu} = \frac{1}{2\mu} \langle \mathbf{r}(t) \cdot \boldsymbol{\eta}(t) \rangle. \quad (14)$$

Below, the latter is shown to follow solely from the two-time correlation matrix  $C(t, t')$  of the noise  $\boldsymbol{\eta} = (\eta_x, \eta_y)$ . Since the passive model (13) and the corresponding temperature (14) describe the active model only effectively, in terms of its average thermodynamic properties, (13) and (14) are referred to as an effective passive model and an effective temperature, respectively.

The existence of this mapping immediately implies that the performance of the active heat engine in terms of its output power and efficiency is precisely that of the corresponding effective equilibrium model. Therefore the known bounds on (finite-time) performance of cyclic Brownian heat engines described by Eq. (13), such as the ultimate Carnot efficiency bound [74], the efficiency at maximum power [25], the maximum efficiency at arbitrary power [72,75], and the possibility to almost attain the reversible efficiency at nonzero power [73], directly carry over to the active heat engine. Furthermore, the effective equilibrium model also sets bounds on average thermodynamic variables for noncyclic and even transient processes. Yet, the nonequilibrium character of the underlying dynamics reveals itself upon closer inspection, as detailed in the remainder of the paper.

## IV. LINEAR THEORY: EFFECTIVE TEMPERATURE

### A. General initial conditions

It is a noteworthy property of the linear theory and the experiments that motivate it that thermodynamic quantities like work, heat, and efficiency are all determined solely by the variance  $\sigma(t)$  of the colloidal position, see Sec. IID. The

<sup>1</sup>Here and in the rest of the paper, we use the Stratonovich convention. See Sec. VID for an explicit calculation of the effective temperature for the ABE model.



variance  $\sigma(t)$  itself obeys the ordinary differential equation

$$\dot{\sigma}(t) = -2\mu k(t)\sigma(t) + 2\langle \mathbf{r}(t) \cdot \boldsymbol{\eta}(t) \rangle \quad (15)$$

which follows from Eq. (11) by taking the scalar product with  $\mathbf{r}$  on both sides and averaging over the noise. For arbitrary additive noise  $\boldsymbol{\eta}$ , Eq. (11) has the formal solution

$$\mathbf{r}(t) = \mathbf{r}_0 e^{-K(t,t_0)} + \int_{t_0}^t dt' \boldsymbol{\eta}(t') e^{-K(t,t')}, \quad (16)$$

with  $K(t, t') \equiv \mu \int_{t'}^t dt'' k(t'')$  and  $\mathbf{r}_0 \equiv \mathbf{r}(0)$  denoting an arbitrary initial position of the particle. With the two-time noise correlation matrix,  $C(t, t')$  from Eq. (12), the average in Eq. (15) evaluates to

$$\begin{aligned} \langle \mathbf{r}(t) \cdot \boldsymbol{\eta}(t) \rangle &= 2D_{\text{eff}}(t) \equiv \langle \mathbf{r}_0 \cdot \boldsymbol{\eta}(t) \rangle e^{-K(t,t_0)} \\ &+ \int_{t_0}^t dt' \text{Tr}[C(t, t')] e^{-K(t,t')}, \end{aligned} \quad (17)$$

where Tr denotes the trace operation. A crucial observation is that Eq. (15) therefore assumes a form that would also result from the Gaussian white noise  $\boldsymbol{\eta} = \sqrt{2D_{\text{eff}}(t)}\boldsymbol{\xi}(t)$  with the correlation matrix  $C_{ij}(t, t') = 2\sqrt{D_{\text{eff}}(t)D_{\text{eff}}(t')}\delta_{ij}\delta(t - t')$ .<sup>2</sup> This implies that the average thermodynamic behavior of the active model (11) with arbitrary additive noise is the same as that of the passive model (13) with an effective equilibrium bath temperature

$$T_{\text{eff}}(t) = \frac{D_{\text{eff}}(t)}{\mu} = \frac{\langle \mathbf{r}(t) \cdot \boldsymbol{\eta}(t) \rangle}{2\mu} = \frac{k(t)\sigma(t)}{2} + \frac{\dot{\sigma}(t)}{4\mu}. \quad (18)$$

The last expression follows from Eq. (15). It shows that also the effective temperature is uniquely given by the variance  $\sigma$ . Notably, the result (18) is valid arbitrarily far from equilibrium and it does not follow from any close-to-equilibrium linear-response approximation like in the Green-Kubo formula [76].

Also note that for positive effective temperature  $T_{\text{eff}}(t) \geq 0$ , Eq. (18) establishes the announced mapping between the active and passive heat engine and thus proves our main result. Negative effective temperatures can however be obtained, for example, during transients departing from initial conditions with  $\langle \mathbf{r}_0 \cdot \boldsymbol{\eta}(t) \rangle < 0$ . At late times, the sign of the effective temperature is determined by the integral in Eq. (17), which is positive for standard correlation matrices  $C(t, t')$  with non-negative diagonal elements. For a quasistatic process, where the system parameters vary slowly compared to the intrinsic relaxation times, one can neglect  $\dot{\sigma}(t)$  relative to the other term in Eq. (18). The effective temperature then reduces to the well-known form [6]

$$T_{\text{eff}}(t) = k(t)\sigma(t)/2. \quad (19)$$

For slowly driven systems, the effective temperature is thus always positive, thanks to the positivity of the trap stiffness  $k$  and variance  $\sigma$ .

<sup>2</sup>This can be seen by substituting this expression for matrix  $C$  into the right-hand side of Eq. (17) and evaluating the integral therein. We further assume that the initial condition is not correlated with the equilibrium noise,  $\langle \mathbf{r}_0 \cdot \boldsymbol{\xi} \rangle = 0$ , which is quite natural.

## B. Cyclic heat engines

The definition (18) of the effective temperature applies both under transient and stationary conditions. Cyclic heat engines operate time periodically by virtue of their periodic driving. Accordingly, we assume that the potential stiffness  $k(t)$  is a periodic function with period  $t_p$  and that the noise correlation matrix is of the form

$$C_{ij}(t, t') = 2\delta_{ij}I(t)I(t')f_i(t - t'), \quad (20)$$

where  $I(t)$  stands for a  $t_p$ -periodic intensity of the noise, and  $f_i(t)$  are arbitrary functions obeying  $f_i(0) = 1$  and decaying towards zero as  $t \rightarrow \infty$ . The system dynamics then settles onto a time-periodic attractor, independent of the initial condition  $\mathbf{r}_0$ , at late times. From now on, we assume that the engine operates in this ‘‘steady state’’ regime, to which we refer as the *limit cycle*. During the cycle, the effective temperature  $T_{\text{eff}}(t)$  takes the form [see Eqs. (17) and (18)]

$$\frac{1}{\mu}I(t) \int_{-\infty}^t dt' I(t') [f_x(t - t') + f_y(t - t')] e^{-K(t,t')}. \quad (21)$$

Importantly, for positive diagonal elements of the correlation matrix, the effective temperature is then manifestly positive, as required to map the active onto the passive model.

## C. (Im)possible generalizations

The simplifying power of the present approach crucially relies on two main features. Firstly, on the linearity of Eq. (11), and secondly on the fact that thermodynamics is predominantly concerned with average energetics.

For the active heat engines discussed in the present contribution, the pertinent microscopic degree of freedom is the position of the colloid. Its thermodynamics is contained in the variance  $\sigma = \langle \mathbf{r} \cdot \mathbf{r} \rangle$ , which controls the complete average energetics (work and heat) of the engine through Eqs. (4) and (5). However, the described mapping to a passive-bath model cannot be extended beyond such average energetics, since the active (11) and passive (13) heat engines differ in variables which depend on higher moments of the position  $\mathbf{r}$  or its complete distribution. This is for example the case for the total entropy or the fluctuations of work, heat and entropy. Without further investigation, one thus cannot take for granted the results obtained under the assumption of a perfect contact with an equilibrium bath, such as the Jarzynski equality [60], the Crooks fluctuation theorem [63], the Hatano-Sasa equality [77,78], and various inequalities containing higher moments of work, heat, and entropy, such as thermodynamic uncertainty relations [79–82].

Also note that, for a true equilibrium noise  $\boldsymbol{\eta}$ , the (effective) temperature  $T_{\text{eff}}$  in Eq. (14) would agree with all other possible definitions of temperature, thereby tying together many *a priori* unrelated dynamical quantities (e.g., by their structurally identical Boltzmann distributions or fluctuation-dissipation theorems, etc.). However, for a nonequilibrium noise, differently defined temperatures can (and generally will) have different values. We refer to Refs. [45,83–90] for various (complementary) approaches to effective temperatures and Refs. [47,64,91,92] for some reviews. Moreover, as illustrated by the ABP results (42) and (43) in Appendix D, typical nonequilibrium distributions deviate strongly from

Boltzmann's Gaussian equilibrium distribution, such as the one characterizing the long-time limit of the equilibrium process, Eq. (13), at constant  $T_{\text{eff}}$ —namely,  $\rho(\mathbf{r}) \propto \exp[-k\mathbf{r}^2/2T_{\text{eff}}]$ . Therefore, in order to build an effective thermodynamic description from a nonequilibrium statistical-mechanics model, one generally has to calculate precisely the effective temperatures corresponding to the relevant degrees of freedom, under the prescribed conditions.

This leads to the mentioned second limitation of the presented effective-temperature mapping, namely that it hinges on the linearity of the model. To make the point, let us consider a one-dimensional setting with the potential  $\mathcal{H}(x, t) = k(t)x^n/n$  when the Langevin equation for position  $x$  reads

$$\dot{x}(t) = -k(t)[x(t)]^{n-1} + \eta(t) \quad (22)$$

and the internal energy, work, and heat (per unit time) are given by  $\langle \mathcal{H}(t) \rangle = k(t)\langle [x(t)]^n \rangle$ ,  $\dot{W}(t) = k(t)\langle [x(t)]^n \rangle$ , and  $\dot{Q}(t) = k(t)d\langle [x(t)]^n \rangle/dt$ , respectively. In order to describe the average thermodynamics, we thus have to consider the dynamics of the  $n$ th moment  $\langle [x(t)]^n \rangle$ . Multiplying Eq. (22) by  $x^{n-1}$  and averaging the result over the noise, we find that

$$\frac{d}{dt}\langle [x(t)]^n \rangle = -nk(t)\langle [x(t)]^{2n-2} \rangle + n\langle [x(t)]^{n-1} \eta(t) \rangle. \quad (23)$$

Thus, in order to get an exact closed dynamical equation for  $\langle [x(t)]^n \rangle$ , we also need a dynamical equation for  $\langle [x(t)]^{2n-2} \rangle$  which, in turn, depends on the moment  $\langle [x(t)]^{3n-4} \rangle$ , and so on. However, out of equilibrium each degree of freedom (and, also each moment  $\langle [x(t)]^n \rangle$ ) has, in general, its own effective temperature, if such a set of effective temperatures can consistently be defined at all.

Recall that the development of a useful (finite-time) thermodynamic description based on a time-dependent effective temperature requires a system with equilibrium noise that yields the same (time-resolved) dynamics of the relevant moments, for which our above discussion of the variance of the linear model (11) provides the paradigm. This means that we would have to develop a passive model with an equilibrium noise that gives rise to precisely the same dynamics of all the moments in Eq. (23) as the original nonlinear active model. Even though our general discussion in Sec. II C shows that for the considered one-parameter potential (Hamiltonian) this should always be possible, this can get very difficult to achieve analytically [90] if the moments represent independent effective degrees of freedom so that their effective temperatures differ.<sup>3</sup> Nevertheless, in this case it should be possible to find the effective temperature numerically. As discussed in Sec. II C, for Hamiltonians that are not proportional to a single control parameter we are not able to give any general conclusions.

Despite these limitations, there are also many important properties that can successfully be captured by the effective-temperature mapping. In the next section, we review its consequences for the performance of active heat engines.

<sup>3</sup>A way to overcome this limitation, leading to an approximate analytical effective temperature, might be based on finding a suitable (approximate) closure for Eq. (23), so that it would only depend on a finite number of moments.

Experts in stochastic thermodynamics may wish to continue directly with Sec. VI, where we derive and discuss more specific analytical results based on the so-called active Brownian particle (ABP) model with an exponential correlation matrix.

## V. LINEAR THEORY: THERMODYNAMICS

### A. Effective entropy production

As described above, the dynamics of the variance in the active model (11) can be mimicked exactly by the effective passive model in (13) with an equilibrium bath at the time-dependent temperature  $T_{\text{eff}}(t)$ , as long as the latter does not transiently turn negative. The noise intensity  $D_{\text{eff}}(t)$  and the mobility  $1/\mu$  in Eq. (13) are thus related by the fluctuation-dissipation relation  $D_{\text{eff}} = \mu T_{\text{eff}}$ . Recall that the variance determines the average thermodynamics of the active engine in terms of work, heat, and efficiency. In particular, due to our interpretation of the thermodynamic variables, the (average) performance of the active heat engine is the same as that of a passive heat engine based on Eq. (13) and can thus be taken over from the known thermodynamics of classical heat engines [9,25,26]. In fact, such a (partial) thermodynamic framework based on the first and second law of thermodynamics is a crucial requirement for a consistent extension of the conventional notion of efficiency to conditions far from equilibrium.

For pedagogical reasons and for completeness, we gather the explicit expressions that summarize the thermodynamics of the linear active heat engine, here. The work reads

$$W(t_i, t_f) = \frac{1}{2} \int_{t_i}^{t_f} dt \underbrace{\dot{k}(t)\sigma(t)}_{\dot{W}(t)} = \frac{1}{2} \int_{k(t_i)}^{k(t_f)} dk \sigma, \quad (24)$$

and the exchanged total heat is given by

$$Q(t_i, t_f) = \frac{1}{2} \int_{t_i}^{t_f} dt \underbrace{k(t)\dot{\sigma}(t)}_{\dot{Q}(t)} = \frac{1}{2} \int_{\sigma(t_i)}^{\sigma(t_f)} d\sigma k. \quad (25)$$

The cycle output work and input heat are still given by Eqs. (6) and (7). Since  $k > 0$ , the latter now explicitly reads

$$Q_{\text{in}}(t_p, 0) = \frac{1}{2} \int_0^{t_p} dt k \dot{\sigma} \Theta(\dot{\sigma}) \quad (26)$$

A main result (to be derived below) is the explicit formulation of the second law of thermodynamics in terms of the mapping to the passive model. It states that the active engine has a nonnegative total effective (in the sense of the mapping to the passive model) entropy-production rate

$$\dot{S}_{\text{tot}}^{\text{eff}}(t) = \mu T_{\text{eff}}(t) \sigma(t) \left[ \frac{2}{\sigma(t)} - \frac{k(t)}{T_{\text{eff}}(t)} \right]^2 \geq 0. \quad (27)$$

Thermodynamically, the entropy production can always be decomposed into the contributions

$$\dot{S}_{\text{tot}}^{\text{eff}}(t) = \dot{S}^{\text{eff}}(t) + \dot{S}_{\text{R}}^{\text{eff}}(t) \quad (28)$$

due to the working substance itself and due to the entropy change in the (effective) heat bath, respectively. Since, by definition, the heat flow from/into an equilibrium heat bath is reversible, the entropy change of the bath obeys the Clausius

equality,

$$\dot{S}_R^{\text{eff}}(t) = -\frac{\dot{Q}(t)}{T_{\text{eff}}(t)} = -\frac{k(t)\dot{\sigma}(t)}{2T_{\text{eff}}(t)}. \quad (29)$$

For the system entropy, one merely has the weaker Clausius inequality

$$\dot{S}^{\text{eff}}(t) \geq -\dot{S}_R^{\text{eff}}(t) = \dot{Q}(t)/T_{\text{eff}}(t). \quad (30)$$

It can be turned into an equality if a quasistatic driving protocol is employed, which then also optimizes the thermodynamic efficiency of the active heat engine.

We now show how these results follow from the statistical-mechanics description. First and foremost, note that the linearity of Eq. (13) ensures that the stochastic process  $\mathbf{r}(t)$  is a linear functional of the Gaussian white noise  $\xi(t)$ . The probability density for the particle position  $\mathbf{r} = (x, y)$  at time  $t$  is therefore also Gaussian, namely

$$p^{\text{eff}}(x, y, t) = \frac{1}{\pi\sigma(t)} \exp\left[-\frac{(x^2 + y^2)}{\sigma(t)}\right], \quad (31)$$

and can easily be seen to solve the Fokker-Planck equation

$$\frac{\partial p^{\text{eff}}}{\partial t} = \nabla_{\mathbf{r}} \cdot [\mu \nabla_{\mathbf{r}} \mathcal{V}(\mathbf{r}) + D_{\text{eff}} \nabla_{\mathbf{r}}] p^{\text{eff}} \quad (32)$$

with  $\nabla_{\mathbf{r}} = (\partial_x, \partial_y)$ . The corresponding Gibbs-Shannon entropy

$$S^{\text{eff}}(t) = -\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy p^{\text{eff}} \ln p^{\text{eff}} = \ln \sigma(t) + \ln \pi + 1 \quad (33)$$

is thus solely determined by the variance  $\sigma(t)$  of the PDF (31), and therefore changes with the rate

$$\dot{S}^{\text{eff}}(t) = \frac{\dot{\sigma}(t)}{\sigma(t)}. \quad (34)$$

The second law in the form given in Eq. (27) now follows by inserting Eqs. (29) and (34) into Eq. (28) and using Eq. (15) for the time derivative of the variance in the form  $\dot{\sigma} = 4\mu T_{\text{eff}}\sigma(1/\sigma - k/2T_{\text{eff}})$ , after rearranging the resulting terms.

To make the entropy production vanish, which corresponds to the equal sign in Eqs. (27) and (30), one has to drive the engine quasistatically. This amounts to setting  $\dot{\sigma} = 0$  in Eq. (18), which yields

$$\sigma(t) = \sigma_{\infty}(t) \equiv 2T_{\text{eff}}(t)/k(t). \quad (35)$$

For a quasistatic driving, the rates of change (29) and (34) of the reservoir and system entropies also both vanish, since they are proportional to the vanishing time derivative  $\dot{\sigma} = 0$ . However, this feature alone might not be enough for concluding that the entire entropy

$$\Delta S_{\text{tot}}^{\text{eff}}(t_p) = \int_0^{t_p} dt' \dot{S}_{\text{tot}}^{\text{eff}}(t') \quad (36)$$

throughout the whole cycle vanishes as  $t_p \rightarrow \infty$ , since it depends on how large  $t_p$  must be to ensure quasistatic conditions, which in turn depends on the intrinsic relaxation behavior of the working substance (in our case the trapped colloid) [52].

It is a consequence of the fluctuation-dissipation relation fulfilled by the effective equilibrium model that the rates of change (29) and (34) of the reservoir and system entropies converge to each other fast enough that the whole quasistatic cycle is reversible and (36) vanishes for large  $t_p$ . We come back to this issue in Sec. VI, where we analyze an explicit model realization.

## B. Efficiency bounds

For an arbitrary cycle, the Clausius inequality (30) can, via standard manipulations [58], be rewritten in terms of the quasistatic (qs) bounds for the output work  $W_{\text{out}}$  and efficiency  $\eta$ , respectively,

$$W_{\text{out}} \leq W_{\text{out}}^{\text{qs}}, \quad (37)$$

$$\eta \leq \eta^{\text{qs}} \leq \eta_C = 1 - \frac{\min(T_{\text{eff}})}{\max(T_{\text{eff}})}. \quad (38)$$

According to the discussion in the previous section, these conditions identically constrain the active heat engine. Given any driving protocol for the variation of the control parameters  $k(t)$  and  $T_{\text{eff}}(t)$ , etc., along the cycle, the largest output work per cycle and the largest efficiency are thus attained for quasistatic driving with  $t_p \rightarrow \infty$ . The ultimate (Carnot) efficiency limit  $\eta_C$  for the active engine is thus reached in a quasistatic Carnot cycle composed of two ‘‘isothermal’’ branches, with constant  $T_{\text{eff}}$ , interconnected by two ‘‘adiabatic branches,’’ with constant entropy (33) and variance  $\sigma_{\infty}$ .

Similarly, the mapping to the passive model (13) implies that the finite-time performance of the active heat engine is the same as that of its effective passive replacement. For convenience, we summarize some consequences of this observation, here. The quasistatic conditions, needed to reach the upper bound  $\eta_C$  on efficiency exactly, imply infinitely slow driving and thus vanishing output power. Naturally, such powerless heat engines are uninteresting for practical purposes [72], where only finite-time processes are relevant, and, thus, other measures of engine performance have been proposed. A prominent role among them plays the maximum power condition. Schmiedl and Seifert [25] showed that overdamped Brownian heat engines deliver maximum power if they operate in the so called low-dissipation regime [93]. Their analysis implies that the efficiency at maximum power of the active heat engine is given by

$$\eta_{\text{MP}} = 1 - \sqrt{\frac{\min(T_{\text{eff}})}{\max(T_{\text{eff}})}}. \quad (39)$$

This result applies if the engine is driven along a finite-time Carnot cycle composed of two isotherms of constant  $T_{\text{eff}}$  and two infinitely fast adiabatic state changes at constant  $\sigma$ , with a suitable protocol for the trap stiffness  $k(t)$  that minimizes the work dissipated during the isothermal branches. We also note that the maximum-power condition was investigated for a specific class of active colloidal heat engines in Ref. [31].

Actual technical realizations of heat engines are usually designed for a certain desired power output. Thus, even more useful than the knowledge of the efficiency at maximum power is the knowledge of maximum efficiency at a given power. Like the former, the latter is, for a Brownian heat

engine of fixed design, attained when operating in the low-dissipation regime along a finite-time Carnot cycle [75,94,95]. The exact numerical and approximate analytical value of the maximum efficiency at arbitrary power for our setting can be obtained using the approach of Ref. [75]. Another universal result, applicable to the active Brownian heat engine, is that, for powers  $P$  close to the maximum power  $P^*$ , the efficiency increases infinitely fast with decreasing  $P$  (i.e.,  $|d\eta/dP|_{P \rightarrow P^*} \rightarrow \infty$ ) [50,75]. Therefore it is usually advantageous to operate heat engines close to maximum power conditions [small  $\delta P = (P^* - P)/P^*$ ], rather than exactly at these conditions ( $\delta P = 0$ ) [94]. Moreover, the results of Refs. [75,94,95] show that  $\eta_C$  can be attained only in the limit  $\delta P \rightarrow 1$ , where either the power  $P$  completely vanishes, or it is negligible with respect to the maximum power  $P^*$ . Recently, this insight led to a proposition of protocols yielding very large maximum power, thus allowing Brownian heat engines to operate close to (and practically with) Carnot's efficiency at large output power [72,73]. As discussed in the following paragraph, active Brownian heat engines offer an alternative route for achieving this.

In the following sections and in Appendix D, we explicitly analyze a specific realization of an active heat engine to illustrate the merits and limitations of the mapping to the "passive dynamics" (13), with an equilibrium bath.

$$C_{ij}(t, t') = \langle \eta_i(t) \eta_j(t') \rangle = \delta_{ij} \left[ 2\sqrt{D(t)D(t')} \delta(t - t') + \frac{1}{2} v(t)v(t') \exp \left\{ - \int_{\min(t, t')}^{\max(t, t')} dt'' D_r(t'') \right\} \right]. \quad (41)$$

Such exponential memory has indeed also been found in a weak-coupling model for a passive tracer in an active bath [96,98]. Besides, it is often employed as a tractable model for the complex correlations arising in strongly interacting systems.

For the following, we assume that the translational diffusion coefficient  $D(t)$  obeys the Einstein relation  $D(t) = \mu T(t)$ , but do not constrain the rotational diffusion coefficient  $D_r(t)$  in the same way. The latter describes the free diffusion of the particle orientation  $\mathbf{n}$  on a unit circle and is incorporated into the ABP equations of motion [65,99,100] through yet another independent zero-mean unit-variance Gaussian white noise  $\xi_\theta$ ,  $\langle \xi_\theta(t) \xi_\theta(t') \rangle = \delta(t - t')$ . The ABP equations then read

$$\dot{\mathbf{r}}(t) = -\mu k \mathbf{r}(t) + \mathbf{v}(t) + \sqrt{2D(t)} \boldsymbol{\xi}(t) \quad (42)$$

$$\dot{\theta}(t) = \sqrt{2D_r(t)} \xi_\theta(t). \quad (43)$$

That the ABP model provides a proper nonequilibrium active noise, as desired for Eq. (11), is not only apparent from the two-time correlation matrix (41), which fixes the average thermodynamics of the model in a way that is not consistent with a fluctuation-dissipation relation. It is further manifest in higher order correlation functions [101] that are sensitive to the non-Gaussian character of the noise (40). As illustrated in Appendix D, this for example allows for a bimodal distribution of the coordinates  $x$  and  $y$ , so that the ABP model captures some of the generically non-Gaussian character of

## VI. WORKED EXAMPLE: THE ABE MODEL

### A. Model definition

To exemplify the above findings for a specific model, we now consider the so-called ABP model. It is the standard minimal model for a particle embedded into an equilibrium bath at temperature  $T$  but actively propelling with velocity  $\mathbf{v}(t) = v(t)\mathbf{n}(\theta)$  in the direction determined by the diffusing unit vector  $\mathbf{n}(\theta)$  at angle  $\theta(t)$ . Encouraged by experimental evidence [6,57,67] and in accord with theoretical studies based on a rigorous elimination of (fast) active degrees of freedom [96,97], the ABP model with harmonic confinement [Fig. 2(a)] has recently also been used to model passive Brownian colloids embedded in an active bath [Fig. 2(b)] [30,32,68–70]. Indeed, within the formalism for a general additive noise outlined above, the ABP model provides us with a simple realization of Eq. (11) in terms of a trapped colloid driven by the nonequilibrium noise

$$\boldsymbol{\eta} = \sqrt{2D(t)} \boldsymbol{\xi} + \mathbf{v}(t). \quad (40)$$

Here the components of  $\boldsymbol{\xi} = (\xi_x, \xi_y)$  are mutually independent zero-mean unit-variance Gaussian white noises, but the velocity term  $\mathbf{v}$  prohibits a straightforward equilibrium interpretation. It contributes an exponential term to the total noise correlation matrix

nonequilibrium fluctuations, lost in another widely employed active-particle model that represents the active velocity as an Ornstein-Uhlenbeck process [66]. We note that these properties are essentially caused by the variable rotational noise  $\xi_\theta$  and persist in a constant-speed ( $v = \text{const.} \neq 0$ ) version of the model.

To emphasize the paradigmatic character of the heat engine described by the ABP Eqs. (42) and (43) with periodically driven parameters  $k(t)$ ,  $T(t)$ ,  $v(t)$ ,  $D_r(t)$ , we refer to it as the ABE model. It involves three ingredients that can potentially drive it far from equilibrium: (i) If the stiffness  $k(t)$  changes on time-scales shorter than the intrinsic relaxation time, the particle dynamics is not fast enough to follow the protocol adiabatically. (ii) If the rotational diffusion coefficient  $D_r$  is not constrained by the Einstein relation, the rotational degree of freedom can be considered connected to a second bath at a temperature distinct from  $T$ . In general, connecting a system to several reservoirs at different temperatures drives it out of equilibrium. (iii) Finally, the velocity term in the Langevin system is formally identical to a nonconservative force giving rise to persistent currents that prevent equilibration.

### B. Cyclic driving protocol

Our driving protocol involves a periodically modulated stiffness  $k$ , reservoir temperature  $T$ , rotational diffusion coefficient  $D_r$ , and active velocity  $v$ . We let the system evolve towards the limit cycle, where we analyze its performance.

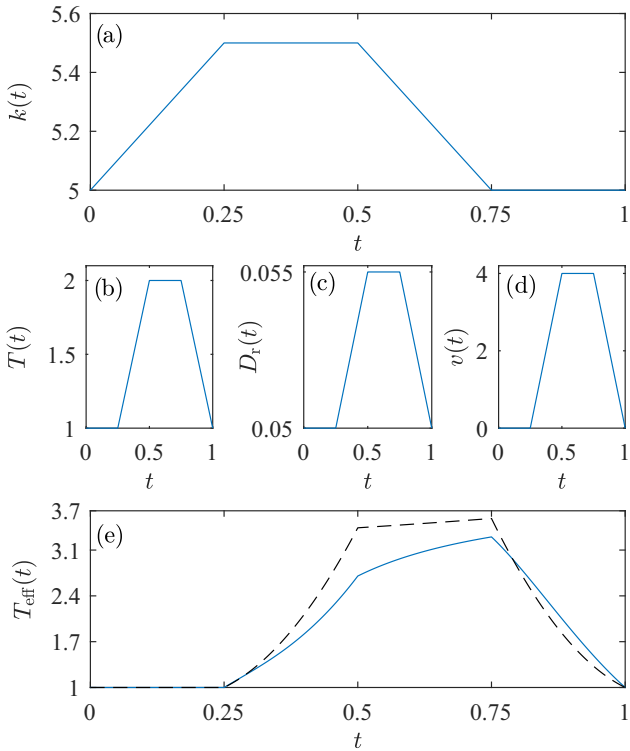


FIG. 3. The driving protocol of the ABE [(a)–(d)] and the effective temperature (e) that maps it to a passive model: (a) trap stiffness, (b) bath temperature, (c) rotational diffusion coefficient, and (d) active velocity, all as functions of time during the limit cycle. The full blue line in (e) depicts the effective temperature  $T_{\text{eff}}(t)$  of Eq. (51), the dashed line its limit (52) for a quasistatic (infinitely slow) driving. Parameters used:  $t_p = 1$ ,  $k_< = 5$ ,  $k_> = 5.5$ ,  $T_> = 2$ ,  $T_< = 1$ ,  $D_r^> = 0.055$ ,  $D_r^< = 0.05$ ,  $v_> = 4$ ,  $v_< = 0$ , and  $\mu = 1$ .

While the following theoretical discussion applies to arbitrary periodic driving, we exemplify our results with a specific Stirling-type protocol that mimics the experimental setup of Ref. [6] (see Fig. 3). It consists of four steps of equal duration ( $t_p/4$ ):

(i) “Isothermal” compression A  $\rightarrow$  B: the stiffness  $k$  increases linearly from  $k_<$  to  $k_>$  at constant noise strength corresponding to the temperature  $T = T_<$  and activity  $\{D_r, v\} = \{D_r^<, v_< \}$ .

(ii) “Isochoric” heating B  $\rightarrow$  C: the noise strength  $\{T, D_r, v\}$  increases linearly from  $\{T_<, D_r^<, v_< \}$  to  $\{T_>, D_r^>, v_> \}$  at constant stiffness  $k = k_>$ .

(iii) “Isothermal” expansion C  $\rightarrow$  D: the stiffness decreases linearly from  $k_>$  to  $k_<$  at constant noise strength  $\{T_>, D_r^>, v_> \}$ .

(iv) “Isochoric” cooling D  $\rightarrow$  A: the noise strength decreases back to its initial value at constant stiffness  $k = k_<$ .

Note that the “isothermal” state changes are characterized by constant bath temperature and activity, which in general corresponds to a varying effective temperature [see Fig. 3(e)]. As explained in Secs. II B and V A, the engine consumes (performs) work when  $\dot{k} > 0$  ( $\dot{k} < 0$ ), i.e., from A  $\rightarrow$  B (C  $\rightarrow$  D) as a standard Stirling engine. On the other hand, heat is absorbed (emitted) from (to) the reservoir when  $\dot{\sigma} > 0$  ( $\dot{\sigma} < 0$ )

and the corresponding portions of the cycle might be different than for the standard Stirling engine, depending on the behavior of the variance  $\sigma$ .

### C. Variance dynamics in the limit cycle

During the limit cycle, which is attained at late times, the dynamics of the variance  $\sigma(t) = 2\sigma_x(t) = 2\sigma_y(t)$  [due to the symmetry of Eq. (42)] is for arbitrary time-periodic driving governed by the two coupled ordinary differential equations

$$\dot{H}(t) = -[\mu k(t) + D_r(t)]H(t) + v(t), \quad (44)$$

$$\dot{\sigma}(t) = -2\mu k(t)\sigma(t) + 4D(t) + 2v(t)H(t). \quad (45)$$

Here, the term  $2D(t) + v(t)H(t)$  determines the long-time time-periodic behavior of the average  $\langle \mathbf{r}(t) \cdot \boldsymbol{\eta}(t) \rangle$ . See Appendix A for details of derivation of Eqs. (44) and (45). Their general solution reads

$$H = H_0 e^{-F(t,0)} + \int_0^t dt' v(t') e^{-F(t,t')}, \quad (46)$$

$$\sigma = \sigma_0 e^{-2K(t,0)} + 4 \int_0^t dt' D_{\text{eff}}(t') e^{-2K(t,t')} \quad (47)$$

with functions  $K(t, t_0) = \mu \int_{t_0}^t dt' k(t')$ ,  $F(t, t_0) = K(t, t_0) + \int_{t_0}^t dt' D_r(t')$ , and  $D_{\text{eff}}(t) = D(t) + v(t)H(t)/2$ . The constants

$$H_0 = \frac{\int_0^{t_p} dt' v(t') e^{-F(t_p,t')}}{1 - e^{-F(t_p,0)}}, \quad (48)$$

$$\sigma_0 = 4 \frac{\int_0^{t_p} dt' D_{\text{eff}}(t') e^{-2K(t_p,t')}}{1 - e^{-2K(t_p,0)}} \quad (49)$$

secure the time-periodicity of the solution and thus they are fixed by the conditions  $H(t_p) = H(0)$  and  $\sigma(t_p) = \sigma(0)$ .

Quasistatic conditions correspond to slow driving relative to the relaxation times  $\tau_H = 1/(\mu k + D_r)$  and  $\tau_\sigma = 1/(2\mu k)$  for  $H$  and  $\sigma$ , respectively. That allows the dynamics of the functions  $H$  and  $\sigma$  to be regarded as relaxed,  $\dot{H} = \dot{\sigma} = 0$ , from which one gets the quasistatic variance

$$\sigma(t) \rightarrow \sigma_\infty(t) \equiv \frac{2}{k} \left( T + \frac{v^2}{2\mu} \frac{1}{k\mu + D_r} \right). \quad (50)$$

The leading correction in the driving speed is derived in Appendix B. Conversely, if the driving is fast relative to the relaxation times  $\tau_H$  and  $\tau_\sigma$ , the colloid cannot respond to the changing parameters  $k$ ,  $T$ ,  $v$  and  $D_r$ , and its variance is given by Eq. (50) with time-averaged parameter values.

At intermediate rates, the complete expression (47) has to be used. To make sure that we calculate the nested integral correctly, we cross-check the obtained results with two independent methods, Brownian Dynamics (BD) simulations and numerical solutions [102]. The finite-time variances follow the quasistatic ones like carrot-chasing donkeys, i.e., the variance decreases (increases) if it is larger (smaller) than the stationary value  $\sigma_\infty$  corresponding to the given value of the control parameters, cf. Figs. 4(b)–4(d). The discrepancy between the quasistatic and the finite-time predictions increases for faster driving and moreover grows with the activity ratio  $v_>/v_<$ . As intuitively expected, and suggested by the

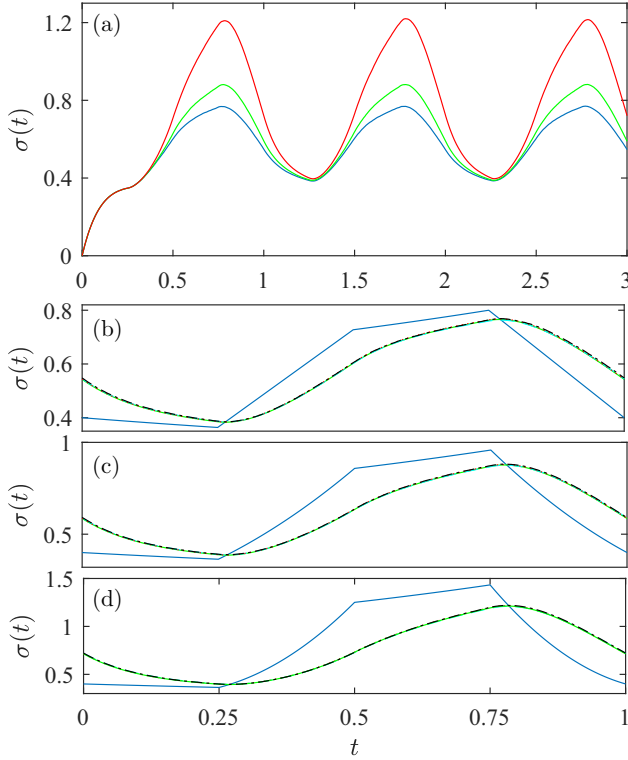


FIG. 4. Positional variance  $\sigma(t) = \langle \mathbf{r} \cdot \mathbf{r} \rangle$  over time for the protocol shown in Fig. 3. Increasing activity  $v_s = 0, 2, 4$  yields an increasing variance  $\sigma$ . (a) Brownian dynamics (BD) simulations of the relaxation to the limit cycle. [(b)–(d)] The dynamics on the limit cycle in BD simulations (solid green), numerical solutions [102] (dot-dashed black), and from the analytical formula (47) (dotted red), shows perfect agreement, despite considerable distance from the quasistatic limit (50) (broken blue lines).

role of  $v$  in Eq. (45), larger active velocities lead to larger variances.

#### D. Effective temperature

Comparing Eqs. (15), (18), and (45), we find for the effective temperature of the ABE on the limit cycle

$$T_{\text{eff}}(t) = \frac{D_{\text{eff}}(t)}{\mu} = T(t) + \frac{v(t)H(t)}{2\mu}. \quad (51)$$

Its value is always larger than the bath temperature  $T$ . Apart from the latter, it also depends on the activity  $v$ , mobility  $\mu$ , trap stiffness  $k$ , and rotational diffusion coefficient  $D_r$ . All the parameters, except for  $T$ , enter  $T_{\text{eff}}$  indirectly, and in a complex way, through the differential equation (44) for  $H$ . The effective temperature thereupon acquires the characteristic relaxation time,  $\tau_H = (\mu k + D_r)^{-1}$ . Its quasistatic limiting form (19) explicitly reads

$$T_{\text{eff}}(t) \rightarrow T_{\text{eff}}^{\infty}(t) \equiv \frac{k\sigma_{\infty}}{2} = T + \frac{v^2}{2\mu} \frac{1}{k\mu + D_r}. \quad (52)$$

The effective temperature possesses several counter-intuitive features. First, in case of periodically modulated activity or trap stiffness, it varies in time, even if the bath

temperature is held constant. Moreover, due to its dynamical nature and finite relaxation time, it generally does so even when the parameters  $T$ ,  $v$ ,  $D_r$ , and  $k$  are held constant. Hence, to realize a proper (effectively) isothermal process with constant  $T_{\text{eff}}$ , one has to carefully tune the control parameters. This is most easily achieved under quasistatic conditions, as demonstrated in Fig. 3(e). There we plot the effective temperature (51) (full blue line) and also its quasistatic limit (52), which would be obtained at very slow driving (black dotted line). For the chosen parameters, the quasistatic effective temperature (52) runs *approximately* along a Stirling cycle, in accord with the temperature  $T(t)$  and activity  $v(t)$  [Figs. 3(b)–3(d)]. Conversely, the finite-time effective temperature (51) exhibits substantially different behavior.

Before going into more details, we now outline three thermodynamically consistent interpretations of the ABE model and derive the corresponding entropy productions. In the discussion of quasistatic and finite-time performance of the engine in Secs. VIII A and VIII B, respectively, we utilize these entropy productions as examples of variables that are not captured by the effective-temperature mapping (13). Another example is the full distribution of the particle position, which we discuss in Appendix D.

## VII. ABE ENTROPY PRODUCTION

As a genuinely nonequilibrium system, any active heat engine always produces entropy, even if operated infinitely slowly. However, how much of that entropy we can (or care to) track depends on our experimental resolution (and interpretation of the engine).

### A. User perspective

On the coarsest level of description, which might be adopted by a *user* of the heat engine, only the supplied heat and the harvested output work matter. Their ratio is the natural measure of efficiency, which is bounded by the optimum (Carnot) efficiency determined by the effective temperature  $T_{\text{eff}}$ . As we have discussed, this temperature can experimentally be measured for the model of a trapped Brownian particle, namely by a device sensible to the variance  $\sigma$  of the particle position; see Fig. 5(a). The thermodynamics of the active heat engine is thereby mapped to that of an ordinary engine with an equilibrium bath and obeys the same limitations. Accordingly, the user would conclude that the total dissipated cycle entropy

$$\Delta S_{\text{tot}}^{\text{eff}} = \int_0^{t_p} dt \dot{S}_{\text{tot}}^{\text{eff}} = \int_0^{t_p} dt \dot{S}_R^{\text{eff}} \quad (53)$$

is given by the net entropy change per cycle in the bath, which thus solely controls the degree of irreversibility of the cycle. To compute the latter, the user would resort to the expression given in Eq. (29) of Sec. V A, namely,

$$\dot{S}_R^{\text{eff}} = -\dot{Q}/T_{\text{eff}} \equiv \dot{Q}_{\text{dis}}^{\text{eff}}/T_{\text{eff}}. \quad (54)$$

Since the particle dynamics is modelled within an overdamped Stokes approximation, the corresponding “effective” dissipation  $Q_{\text{dis}}^{\text{eff}}$  to the effective equilibrium bath is straightforwardly given by the force acting on the particle times its velocity

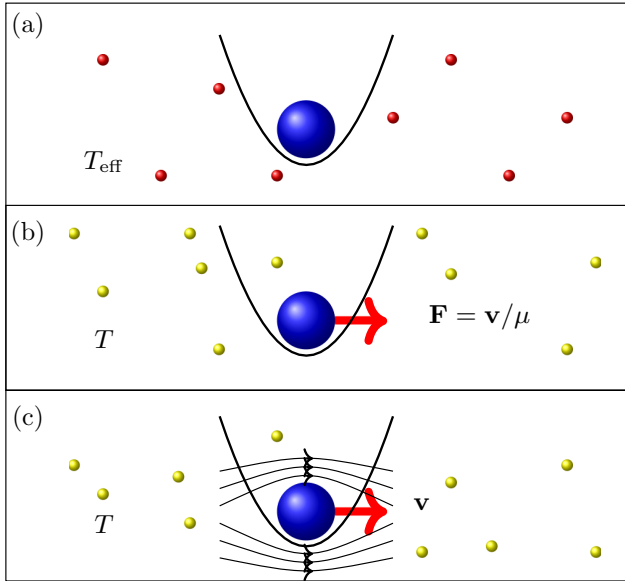


FIG. 5. Different levels of control over the system imply different changes in the bath entropy. (a) Mere “users” of an active heat engine are only concerned with its thermodynamic input/output characteristics. They judge reversibility and entropy changes with respect to an effective equilibrium bath (red particles) at a fictitious temperature  $T_{\text{eff}}$  larger than the temperature  $T$  of the background solvent [yellow particles in (b) and (c)]. More detailed knowledge about the engine’s internal working substance (here the ABP particle) and its dynamics uncovers the nonequilibrium character of the system, which depends on the time-reversal properties of its dynamics [35,36]. If the active velocity  $\mathbf{v}$  in the ABE results from dragging or pushing the particle through the liquid by an external force  $\mathbf{F}$  (b), the particle behaves like a sailboat and the change in bath entropy obeys Eq. (59). If the particle is self-propelled or advected by the surrounding liquid with velocity  $\mathbf{v}$  like a surfboard (c), the bath entropy obeys Eq. (60).

(averaged)

$$\dot{Q}_{\text{dis}}^{\text{eff}} = -\langle \nabla_{\mathbf{r}} \mathcal{H} \cdot \dot{\mathbf{r}} \rangle = -k(t)\dot{\sigma}(t)/2. \quad (55)$$

Importantly, the user is not concerned with other details of the nonequilibrium bath than the variance  $\sigma$  and the effective equilibrium temperature  $T_{\text{eff}}$  it provides. He would thus adopt the above expressions for arbitrary noise in Eq. (11), regardless of the underlying physics of the bath. For the specific ABE realization of the active heat engine, these expressions can explicitly be evaluated using Eqs. (45), (47), and (51). This notion of entropy production, directly derived from the notion of system entropy consistent with the second law for the supplied heat and the harvested output work, is the only one to safely yield efficiency bounds compatible with conventional definitions. It is thus arguably the most pertinent one in the context of active heat engines.

### B. Trajectory perspective

In contrast to the above user, a *heat engineer* would possibly consider the engine at a higher resolution and have access to the individual stochastic *trajectories* of the particle

position generated by Eq. (11). Thereby, she could uncover the nonequilibrium character of the active heat bath, which dissipates energy even if the engine operates under quasistatic conditions. To this end, she could evaluate the dissipation per cycle in the form  $\langle \ln P_{\text{F}}(\Gamma)/P_{\text{R}}(\Gamma^*) \rangle$ , exploiting a relation often referred to as local detailed balance condition. It relates the symmetry breaking between the path probabilities  $P_{\text{F}}(\Gamma)$  and  $P_{\text{R}}(\Gamma^*)$  for paths  $\Gamma$  and their time-reversed images  $\Gamma^*$  to dissipation. (For more details, see Appendix C and Refs. [103,104].) The method can in principle be applied regardless of the physics underlying the noise term in Eq. (11), if one can observe or otherwise guess the time-reversed dynamics. See Ref. [104] for an example of a successful application of such a strategy to biological systems. In general, this will however technically require assumptions or knowledge of the time-reversed noise dynamics, i.e., microscopic information beyond that of the stochastic (forward) trajectories of the particle position. Such information is seldom available outside the realm of detailed models of the mesoscopic physics. For specificity, we therefore now consider explicitly the ABE model, based on the concrete ABP model.

### C. ABP perspective: sailboats versus surfboards

For ABP’s, the noise  $\boldsymbol{\eta}$  comprises the (time-symmetric) equilibrium white noise  $\sqrt{2D}\xi$  together with the active propulsion  $\mathbf{v}$ . The colloid could be a randomly (self-) propelled active particle or a schematically modeled passive tracer in an active bath [6,30]. In any case, its active velocity  $\mathbf{v}$  is due to a dissipative process and admits two alternative interpretations, depending on its presumed time-reversal properties [35,36]. Namely, it can be understood as a Stokes velocity caused by an external (random) force  $\mathbf{v}(t)/\mu$ , the so-called swim force. This very common interpretation, depicted in Fig. 5(b), treats the particle like a sailboat blown around by erratic winds, which is why we refer to it as the “sailboat” interpretation. Or, in a second interpretation, depicted in Fig. 5(c), the active term  $\mathbf{v}(t)$  can be interpreted as the actual swim velocity of a microswimmer that either “sneaks” through the quiescent background solvent by an effective phoretic surface slip  $\mathbf{v}(t)$  [46,105,106] or is passively advected by a local flow field  $\mathbf{v}(t)$  [107,108]. We refer to it as the “surfboard” interpretation. It treats  $\mathbf{v}(t)$  as a proper dynamic velocity as opposed to the disguised force in the sailboat interpretation. Upon time reversal, forces usually do not change the sign, while velocities do. The detailed balance condition then implies that the rate of entropy change in the bath reads

$$\dot{S}_{\text{R}}^{\pm} = \dot{Q}_{\text{dis}}^{\pm}/T, \quad (56)$$

for sailboats (+) and surfboards (−), respectively [109–111]. The corresponding dissipation rates are

$$\dot{Q}_{\text{dis}}^{+} = \langle (\mathbf{v}/\mu - \nabla_{\mathbf{r}} \mathcal{H}) \cdot \dot{\mathbf{r}} \rangle = \dot{Q}_{\text{dis}}^{\text{eff}} + v^2/\mu - \langle \nabla_{\mathbf{r}} \mathcal{H} \cdot \mathbf{v} \rangle, \quad (57)$$

$$\dot{Q}_{\text{dis}}^{-} = \langle -\nabla_{\mathbf{r}} \mathcal{H} \cdot (\dot{\mathbf{r}} - \mathbf{v}) \rangle = \dot{Q}_{\text{dis}}^{\text{eff}} + \langle \nabla_{\mathbf{r}} \mathcal{H} \cdot \mathbf{v} \rangle. \quad (58)$$

We refer to Appendix C for details of the formal derivation, and discuss these results on a physical basis. In the dissipation rate  $\dot{Q}_{\text{dis}}^{+}$  for sailboats, the swim term is added as an additional

force (intuitively the wind drag) to the potential force. In contrast, for surfboards, it is subtracted from the particle velocity corresponding to a reformulation of the equation of motion in a frame that is freely co-moving with the flow velocity  $\mathbf{v}(t)$ .

Since  $\dot{Q}_{\text{dis}}^+(t)$  and  $\dot{Q}_{\text{dis}}^-(t)$  have different reference points (vanishing for sailboats blown against the quay and surfboards floating freely with the surf, respectively), the two dissipation rates can not generally be ordered according to their magnitude for the ABE, where both situations may (approximately) be encountered along the cycle. Also note that the detailed balance condition imposes that the heat is dissipated in the background solvent at temperature  $T(t)$ , which is natural from the point of view of the ABP model. As a consequence, also different amounts of entropy production will be assigned to the self-propulsion, dependent on the chosen ABP interpretation.

They can both be understood as composed of the effective dissipation  $\dot{Q}_{\text{dis}}^{\text{eff}}(t)$  over the solvent temperature  $T(t) \leq T_{\text{eff}}(t)$ , plus some extra (manifestly active) entropy production due to the particle's excursions off the surf or off the quay, respectively,

$$T\dot{S}_{\text{R}}^+ = \dot{Q}_{\text{dis}}^{\text{eff}} + v^2/\mu - 2\mu k(T_{\text{eff}} - T), \quad (59)$$

$$T\dot{S}_{\text{R}}^- = \dot{Q}_{\text{dis}}^{\text{eff}} + 2\mu k(T_{\text{eff}} - T), \quad (60)$$

were we used  $\langle \nabla_{\mathbf{r}} \mathcal{H} \cdot \mathbf{v} \rangle = k(\mathbf{r} \cdot \mathbf{v}) = k(\mathbf{r} \cdot (\boldsymbol{\eta} - \sqrt{2D(t)}\boldsymbol{\xi})) = 2\mu k(T_{\text{eff}} - T)$ , which follows from Eqs. (14) and (40). In the second case (surfboards), the additional propulsion contribution to the entropy production beyond  $\dot{S}_{\text{R}}^{\text{eff}}$  is manifestly positive, since  $T_{\text{eff}} \geq T$ . Intuitively, this is because any failure to float with the flow gives rise to dissipation. In the first case (sailboats), the minimum condition for  $\dot{S}_{\text{R}}^{\text{eff}}$  can only be guaranteed under quasistatic conditions. Intuitively, the ‘‘wind’’ may otherwise transiently prevent dissipation by ‘‘arresting the sailboat at the quay.’’

While the derivation of the expressions (59) and (60) relies on a deeper knowledge of the system dynamics than the behavior of the variance, it is worth noting that  $\sigma(t)$  is still sufficient for their evaluation. The dynamics of the variance thus suffices to evaluate the ‘‘total’’ entropy  $\Delta S_{\text{tot}}^{\pm}(t_p) = S_{\text{R}}^{\pm}(t_p) = \int_0^{t_p} dt \dot{S}_{\text{R}}^{\pm}$  produced per cycle of the operation of the ABE. In contrast, the change in the system entropy

$$S(t) = - \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \int_0^{2\pi} d\theta p \ln p, \quad (61)$$

which vanishes for a complete cycle but is necessary for evaluating the total entropy change within the cycle,  $\Delta S_{\text{tot}}^{\pm}(t) = S_{\text{R}}^{\pm}(t) + S(t) - S(0)$ , depends on the full probability distribution  $p(x, y, \theta, t)$  for the position of the particle at time  $t$ . The latter obeys the Fokker-Planck equation

$$\frac{\partial p}{\partial t} = \left( \nabla_{\mathbf{r}} \cdot [\mu \nabla_{\mathbf{r}} \mathcal{H}(\mathbf{r}) - \mathbf{v}] + D \nabla_{\mathbf{r}}^2 + D_{\text{r}} \frac{\partial^2}{\partial \theta^2} \right) p \quad (62)$$

corresponding to Eqs. (42) and (43). One can calculate the PDF  $p(x, y, \theta, t)$  either numerically, from Eq. (62), or using BD simulations of Eqs. (42) and (43) (see Appendix D for a detailed discussion of the results). The system entropy  $S(t)$  is thus the only variable of our thermodynamic analysis which

generally cannot be calculated using the mean square displacement  $\sigma$  alone.

The above results are suitable to fully quantify the engine's thermodynamic performance. In the following section, we evaluate the derived expressions and discuss their generic properties.

## VIII. ABE PERFORMANCE

In this section, we first focus on the quasistatic regime of operation of the ABE, where we demonstrate in more detail some peculiarities connected with the unintuitive behavior of the effective temperature. For vanishing entropy productions  $\Delta S_{\text{tot}}^{\pm}$ , as defined in the previous section, the nonequilibrium ABE bath is seen to admit a representation as an equilibrium bath. Then, we consider finite-time effects onto the performance of the ABE, and the additional entropy production due to the nonquasistatic operation.

### A. Quasistatic regime

In the quasistatic regime, the engine dynamics in terms of the variance  $\sigma(t)$  and the effective temperature  $T_{\text{eff}}(t)$  are given by Eqs. (50) and (52), respectively. They thus depend merely parametrically on the driving  $k(t)$ ,  $T(t)$ ,  $D_{\text{r}}(t)$ , and  $v(t)$ . The effective entropy production  $\Delta S_{\text{tot}}^{\text{eff}}$  (27) then vanishes, and the (effective) efficiency of the ABE is given by the classical result evaluated in terms of the stiffness  $k(t)$  and temperature  $T_{\text{eff}}(t)$ . In particular, a quasistatic cycle consisting of two branches with constant  $T_{\text{eff}}$  and two adiabats will thus operate with Carnot efficiency  $\eta_{\text{C}}$  (38). Equivalently, realizing a Stirling cycle in terms of  $k(t)$  and the effective temperature  $T_{\text{eff}}(t)$  will result in the (effective) Stirling efficiency  $\eta_{\text{C}} \ln a / (\eta_{\text{C}} + \ln a)$  with  $a = \min(k) / \max(k)$  [30]. And one could deal similarly with other thermodynamic cyclic protocols. However, using the simplifying analogy with the effective equilibrium bath, one should make sure to actually use  $k(t)$  and  $T_{\text{eff}}(t)$  as control parameters and not simply rely on an intuition about the behavior of the effective temperature based on the background solvent temperature  $T$ , activity  $v$  and rotational diffusivity  $D_{\text{r}}$ . Indeed, as mentioned in Sec. VID, what is a Stirling (or Carnot) cycle in terms of the effective temperature can be quite different from the one defined in terms of  $T$ ,  $v$ , and  $D_{\text{r}}$ . To quantify the difference, it is useful to introduce the parameter

$$\mathcal{K}(t) \equiv k\mu/D_{\text{r}} \quad (63)$$

which compares the characteristic timescales  $D_{\text{r}}^{-1}$  and  $(k\mu)^{-1}$  for relaxation of the orientation  $\theta$  and the position  $\mathbf{r}$ , respectively. The quasistatic effective temperature (52) can be written as

$$T_{\text{eff}}(t) = T + \frac{v^2}{2\mu D_{\text{r}}} \frac{1}{1 + \mathcal{K}}. \quad (64)$$

Only in the limiting cases  $\mathcal{K} \rightarrow 0$  and  $\mathcal{K} \rightarrow \infty$ , a *naive* quasistatic isothermal process (constant temperature  $T$ , activity  $v$ , and rotational diffusivity  $D_{\text{r}}$  and variable stiffness  $k$ ) corresponds to an effective equilibrium isothermal process (constant  $T_{\text{eff}}$ ).



Despite the equilibrium analogy, the bath actually corresponds to a driven system with its apparent equilibrium characteristics actively maintained by some dissipative processes. So even for quasistatic operation of the engine, closer inspection reveals this nonequilibrium nature of the bath. In particular, the sailboat/surfboard interpretations of the particle motion will reveal some of this entropy production, since

$$T\dot{S}_R^\pm = \dot{Q}_{\text{dis}}^{\text{eff}} + \frac{v^2}{\mu} \frac{1}{1 + \mathcal{K}^{\pm 1}} \geq \dot{Q}_{\text{dis}}^{\text{eff}} \quad (65)$$

and thus  $\Delta S_{\text{tot}}^\pm \geq \Delta S_{\text{tot}}^{\text{eff}} = 0$ .

For strong confinements,  $\mathcal{K} \gg 1$ , the active dynamics is highly persistent on the confinement scale, so that the particle moves quasiballistically in the potential. The effective temperature  $T_{\text{eff}}$  is therefore given by the temperature of the equilibrium solvent  $T$ , which is the only remaining source of noise. Using the sailboat interpretation of the ABP (for which  $\mathbf{v}$  is interpreted as an external force), we find that  $\Delta S_{\text{tot}}^+ = \Delta S_{\text{tot}}^{\text{eff}} = 0$  since the sailboat is trapped in a quay. The sailboat interpretation is thus consistent with the notion that the ABE operates reversibly. In contrast, a trapped surfboard (for which  $\mathbf{v}$  is interpreted as a velocity) is inhibited from moving with the surf, leading to dissipation:  $\Delta S_{\text{tot}}^- = \Delta S_{\text{tot}}^{\text{eff}} + \int_0^{t_p} dt v^2(t)/\mu = \int_0^{t_p} dt v^2(t)/\mu > 0$ .

For weak confinements,  $\mathcal{K} \ll 1$ , the particle's active motion randomizes on the confinement scale so that it can be subsumed into the  $\delta$ -correlated noise (40) via the effective temperature and the corresponding noise correlation matrix  $C_{ij}(t, t') = 2\sqrt{D_{\text{eff}}(t)D_{\text{eff}}(t')}\delta_{ij}\delta(t - t')$ . Its dynamics mimics Brownian motion in an effective equilibrium bath maintained at the (stiffness-independent) temperature  $T_{\text{eff}} = T + v^2/(2\mu D_r)$ . In this case, confinement and random active motion interfere in such a way that both the sailboat and surfboard interpretations can detect the positive entropy production,  $\Delta S_{\text{tot}}^\pm > \Delta S_{\text{tot}}^{\text{eff}} = 0$ , and the actual irreversibility of the operation. Only by imposing the additional limit  $v^2 \ll 2\mu D_r T$ , when the rotational motion completely obliterates the active swimming so that  $T_{\text{eff}} = T$ , surfboards cease to be bothered by the confinement and no longer dissipate, i.e.,  $\Delta S_{\text{tot}}^- = 0$ . In the sailboat interpretation, the release of the boat from the tug of war with the quay instead results in a complete waste of the efforts of the external swim force to haul the particle around in an enhanced random motion. The corresponding dissipation of the fully released sailboat thus precisely matches that of a fully trapped surfboard:  $\Delta S_{\text{tot}}^+ = \int_0^{t_p} dt v^2(t)/\mu > 0$ .

For intermediate values of  $\mathcal{K}$ , the effective temperature depends on the stiffness  $k(t)$  and the (traditional) definition of heat input along an individual step of the driving protocol may not actually yield the correct interpretation. It then also fails to yield a consistent measure of efficiency. Instead, one should carefully reconsider what is the actual heat input, based on Eq. (26). Heat thus flows into the system whenever the variance  $\sigma$ —and thus the effective system entropy (33)—increases, and vice versa.

To illustrate this point, recall the definition of the Stirling cycle in Sec. VI B. The standard Stirling cycle consists of two isochores (constant trap stiffness  $k$ ) and two isotherms (constant solvent temperature  $T$ ). Therefore it forms a rect-

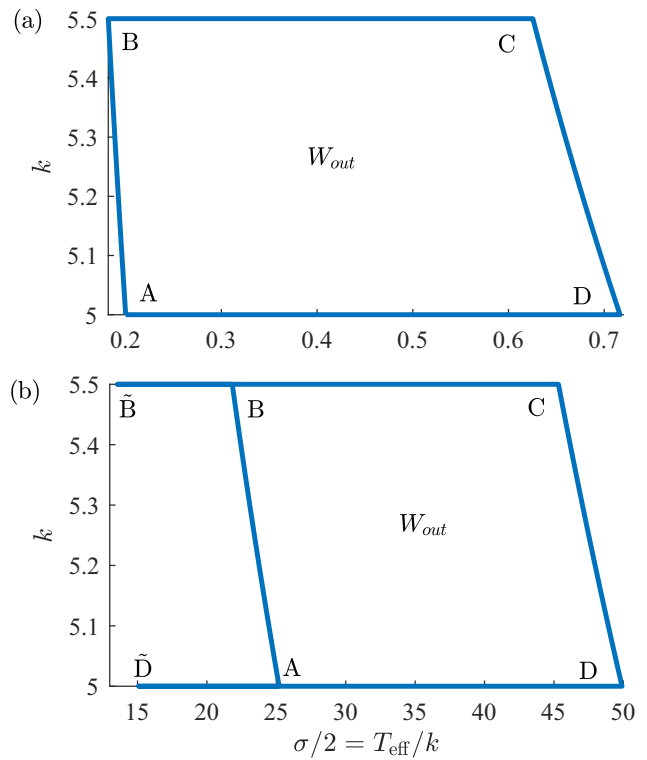


FIG. 6. Two quasistatic (generalized) Stirling cycles in terms of the trap stiffness  $k(t)$  and variance  $\sigma(t)$  of the particle positions ABCD and ABBCDD. (The corresponding energy flows are evaluated in Fig. 7.) (a) In the standard Stirling cycle ABCD, the heat flows from the bath into the system along the isochor BC and isotherm CD ( $\dot{Q} = k\dot{\sigma}/2 > 0$ ), and from the system into the bath otherwise ( $\dot{Q} < 0$ ). (b) In the “nonstandard Stirling” cycle ABBCDD, the heat flow reverses (outflow along  $B\tilde{B}$ , inflow along  $\tilde{B}C$ ) along the isochoric branch  $BC = B\tilde{B}C$  and similarly for the isochor  $DA = D\tilde{D}A$ . The output works  $W_{\text{out}} = \int_0^{t_p} \sigma dk/2$  of the individual cycles are given by the areas they enclose. Similarly, heat input and output can be visualized as areas below the curves.

angle in a  $k$ - $T$  diagram, translating to a shape similar to the ABCD cycle in Fig. 6, in a  $k$ - $T/k$  diagram. Actually, Fig. 6 is slightly more general, as it shows two possible interpretations of the quasistatic ABE-Stirling cycle in a  $k$ - $T_{\text{eff}}/k$  diagram. The “standard” protocol ABCD corresponds to the evolution of the thermodynamic variables as depicted in Fig. 7(a). Note that they, in turn, evolve strictly monotonically or remain constant during the individual steps of duration  $t_p/4$ . Hence, during a single step, heat is either only absorbed or only released by the system, and it is possible to write the input heat as  $Q_{\text{in}} = Q_{\text{BC}} + Q_{\text{CD}}$ , where  $Q_{\text{XY}}$  is the amount of heat absorbed between the points  $X$  and  $Y$ . Which corresponds to the conventional practice for a Stirling cycle.

Consider next the cycle ABBCDD corresponding to Fig. 7(b). In this case, the system releases heat during the segment  $B\tilde{B}$  ( $\sigma$  decreases from  $\sigma_B$  to  $\sigma_{\tilde{B}}$ ), but absorbs heat during the remainder of the state change BC ( $\sigma$  increases from  $\sigma_{\tilde{B}}$  to  $\sigma_C$ ). A similar situation occurs also at the end of the cycle. Hence, the conventional shorthand notion of heat input as heat exchanged between the system and the reservoir during an

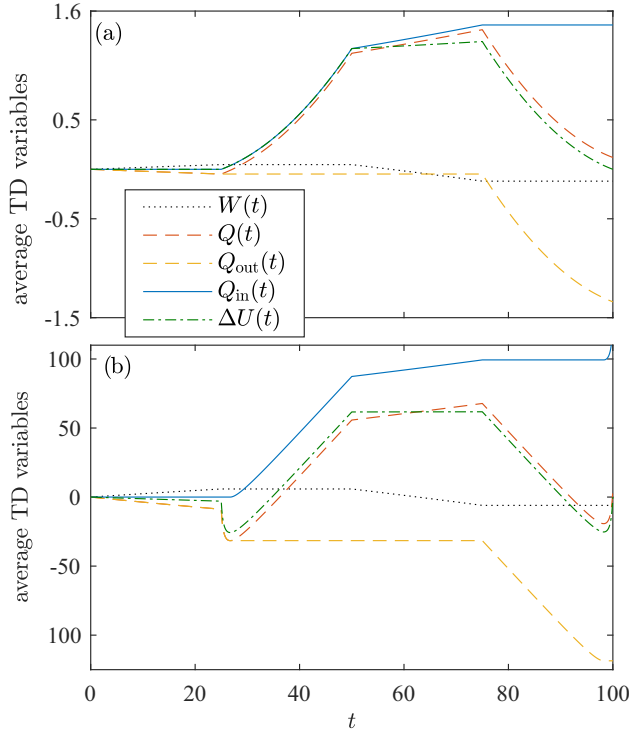


FIG. 7. Energy flows for the quasistatic Stirling cycles depicted in Fig. 6. Net work and heat  $W$ ,  $Q = Q_{\text{in}} + Q_{\text{out}}$ , heat influx and outflux  $Q_{\text{in}}$ ,  $Q_{\text{out}}$ , and internal energy change  $\Delta U \equiv \langle \mathcal{H}(t) - \mathcal{H}(0) \rangle$ , as defined in Secs. II B and V A, are traced out as functions of time during a quasistatic cycle of duration  $t_p = 100$  significantly larger than the relaxation times  $\tau_H = 1/(\mu k + D_r)$  and  $\tau_\sigma = 1/(2\mu k)$  for  $T_{\text{eff}}$  and  $\sigma$ , respectively. (a)  $v_s = 4$ . (b)  $v_s = 500$ ,  $v_c = 50$ ,  $D_r^> = 500$ , and  $D_r^< = 5$ ; other parameters as in Fig. 3.

entire step of the cycle is not appropriate, in this case. Instead, one has to use the definition (7), also utilized in Fig. 7. The dashed red, dashed yellow and full blue lines in Fig. 7(b) in the time interval from  $t = 25$  to  $50$ , also serve to illustrate the differences in the heat balance. For a further treatment of efficiency of Stirling engines operating in contact with active baths in the quasistatic regime, we refer to Ref. [30].

### B. Finite-time performance

Let us finally investigate the most complex case of non-quasistatic cycles for which the protocol from Sec. VIB is imposed with cycle durations  $t_p$  significantly shorter than the internal relaxation times  $\tau_H = 1/(\mu k + D_r)$  and  $\tau_\sigma = 1/(2\mu k)$  for  $T_{\text{eff}}$  and  $\sigma$ , respectively. The ABE model provides full control over the finite-time thermodynamics. To check our analytical results for the variance given in Sec. VIC, we compared it to direct numerical solutions of the equations of motion via the matrix numerical method of Ref. [102], and found perfect agreement. We also note that the new features observed in the analytical results for the toy model are generic, and should qualitatively also be observed for other heat engines in contact with nonequilibrium reservoirs.

The hallmark of nonquasistatic operation of any thermodynamic heat engine is the observation of a net entropy increase

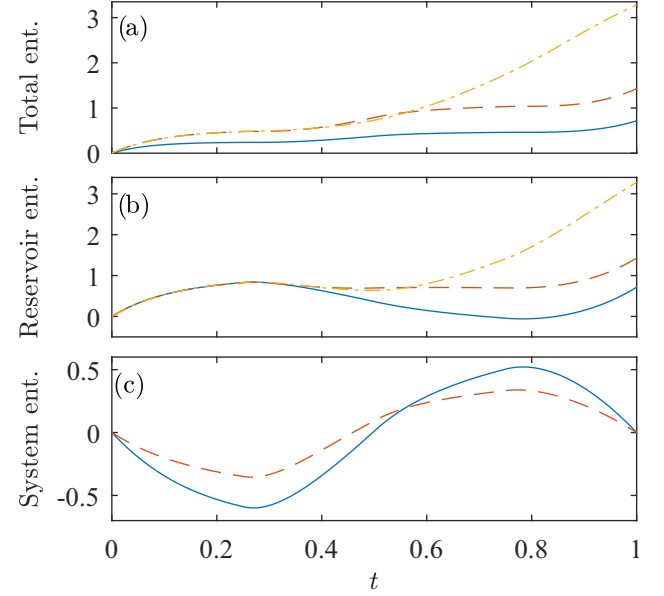


FIG. 8. Evolution of the various entropies discussed in the text as functions of time during the limit cycle, depicted in Fig. 3, with  $v_s = 4$ . (a) “Total” ABE entropy changes  $\Delta S_{\text{tot}}^+(t)$  for “sailboats”, (red dashed line),  $\Delta S_{\text{tot}}^-(t)$  for “surfboards” (yellow dot-dashed line), both from Sec. VII C, and the effective entropy change  $\Delta S_{\text{tot}}^{\text{eff}}(t)$  from Eq. (36) (solid blue line). (b) shows corresponding changes in the reservoir entropy  $\Delta S_{\text{R}}^+(t)$  from Eq. (59) (red dashed line),  $\Delta S_{\text{R}}^-(t)$  from Eq. (60) (yellow dot-dashed line), and  $\Delta S_{\text{R}}^{\text{eff}}(t)$  from integrating Eq. (29) (solid blue line), and panel c) in the system entropy  $\Delta S^{\text{eff}}(t)$  from Eq. (66) (solid blue line) and  $\Delta S(t)$  from Eq. (67) (red dashed line).

during the cycle. Therefore Fig. 8 depicts the individual entropy changes defined in Secs. V A and VII as functions of time during the limit cycle. Panel (a) shows that both the total effective entropy change  $\Delta S_{\text{tot}}^{\text{eff}}(t)$ , measured by the ABE user, and the total ABE entropy changes  $\Delta S_{\text{tot}}^{\pm}(t)$ , corresponding to the sailboat and surfboard interpretations, are nondecreasing functions of time. They thus meet the expectation for valid total entropies according to the second law of thermodynamics. It is noteworthy, that the ABE entropy changes  $\Delta S_{\text{tot}}^{\pm}(t)$  are larger than the effective entropy change  $\Delta S_{\text{tot}}^{\text{eff}}(t)$ , at all times, even during the first part of the cycle, given by  $t \in (0, 0.25)$ , where the active velocity  $v$  vanishes.

As gleaned from the panel (b), the rates of entropy change in the bath, with  $\dot{S}_{\text{R}}^{\text{eff}}$  given by Eq. (29) and  $\dot{S}_{\text{R}}^{\pm}(t)$  given by Eqs. (59) and (60), are in that case all equal. The inequality  $\Delta S_{\text{tot}}^{\text{eff}}(t) < \Delta S_{\text{tot}}^{\pm}(t)$  is then solely caused by the different changes of the system entropy

$$\Delta S^{\text{eff}}(t) = S^{\text{eff}}(t) - S^{\text{eff}}(0) = \ln \frac{\sigma(t)}{\sigma(0)}, \quad (66)$$

$$\Delta S(t) = S(t) - S(0), \quad (67)$$

shown in the panel c), with  $S^{\text{eff}}(t)$  and  $S(t)$  given by Eqs. (33) and (61), respectively. For the remaining time  $[t \in (0.25, 1)]$  of the cycle, even the changes in the bath entropies  $\Delta S_{\text{R}}^{\pm}(t)$  of the ABE are larger than  $\Delta S_{\text{R}}^{\text{eff}}(t)$ . While  $\dot{S}_{\text{R}}^-(t) \geq \dot{S}_{\text{R}}^{\text{eff}}(t)$  and  $\dot{S}_{\text{tot}}^{\pm} \geq \dot{S}_{\text{tot}}^{\text{eff}}$  always hold, we find that  $\dot{S}_{\text{R}}^+(t) < \dot{S}_{\text{R}}^{\text{eff}}(t)$  is not

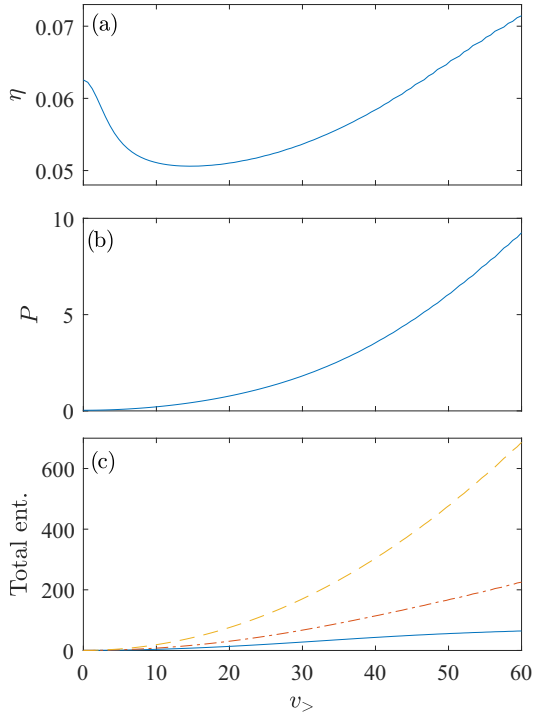


FIG. 9. (a) Efficiency and (b) power output, both from Eq. (8), and (c) total entropy production as functions of the maximum active velocity  $v_>$  for the protocol from Fig. 3. (c)  $\Delta S_{\text{tot}}^-$  (yellow dashed),  $\Delta S_{\text{tot}}^+$  (red dot-dashed), and  $\Delta S_{\text{tot}}^{\text{eff}}$  (solid blue), all from Eq. (68).

ruled out (detailed data now shown). The figure also corroborates the periodicity of the system entropies  $S^{\text{eff}}(t)$  and  $S(t)$ , so that the total entropy changes  $\Delta S_{\text{tot}}^{\text{eff}}(t_p)$  and  $\Delta S_{\text{tot}}^{\pm}(t_p)$  per cycle are solely determined by the (per cycle) entropy changes  $\Delta S_{\text{R}}^{\text{eff}}(t_p)$  and  $\Delta S_{\text{R}}^{\pm}(t_p)$  in the bath, as it should be.

To study the influence of activity on the ABE performance, in Fig. 9, we fix all the parameters according to Fig. 3 and vary the maximum active velocity  $v_>$ . For small values of  $v_>$ , the efficiency is decreased by the activity, while for large values of  $v_>$  it is increased, and eventually attains a constant maximum value. This behavior can be understood as follows. The efficiency of the heat engine quite generally increases with the largest difference in the effective temperature  $\max(T_{\text{eff}}) - \min(T_{\text{eff}})$ , similarly as in the Carnot formula. Even beyond the quasistatic regime one expects that the effective temperature is qualitatively described by Eq. (52). For small values of  $v_>$ , Eq. (52) implies that the temperature difference can be decreased by variations of the rotational diffusion coefficient, depicted in Fig. 3(c), while it increases with  $v_>$  for large  $v_>$ . More intuitive behavior is observed for the power [Fig. 9(b)] and the entropy productions  $\Delta S_{\text{tot}}^{\text{eff}}$  and  $\Delta S_{\text{tot}}^{\pm}$  [Fig. 9(c)] that monotonically increase with  $v_>$ .

Finally, we assess the effect of the finite-time driving on the ABE operation. Specifically, in Fig. 10, we depict performance of the ABE as function of the cycle duration  $t_p$  for three values of the maximum active velocity  $v_>$ . In panel (a), the efficiency monotonously increases with increasing  $t_p$  and eventually reaches the quasistatic limit (the red line). Notably, whether the efficiency is increased or decreased by

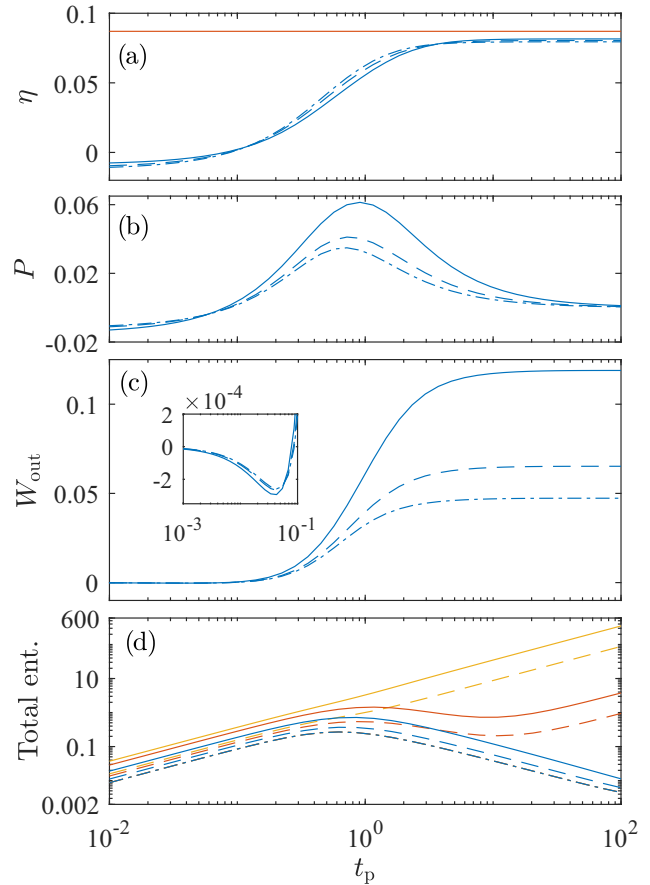


FIG. 10. (a) Efficiency, (b) output power, (c) output work, defined in Sec. II B, and (d) total entropy production for  $v_> = 0$  (dot-dashed lines),  $v_> = 2$  (dashed lines), and  $v_> = 4$  (solid lines) as functions of cycle duration  $t_p$ . The inset in (c) magnifies the initial part of the plot for  $t_p \in [10^{-3}, 10^{-1}]$ . (d)  $\Delta S_{\text{tot}}^-$  (yellow),  $\Delta S_{\text{tot}}^+$  (red), and  $\Delta S_{\text{tot}}^{\text{eff}}$  (blue); all according to Eq. (68). Other parameters as in Fig. 3.

the bath activity depends on the cycle duration, as evidenced by the dashed and solid lines wandering above and below the dot-dashed line. Namely, apart from enhancing the output work and power [panels (b) and (c)], the activity also provides an increased heat flow into the system. As expected, the output power vanishes for large cycle durations and exhibits a maximum for a certain value of  $t_p$ . On the contrary, the output work is, for large cycle times, an increasing function which converges to the quasistatic value, which monotonously increases with  $v_>$ . Interestingly, for  $10^{-2} \lesssim t_p \lesssim 10^{-1}$ , the output work exhibits a shallow negative excursion as revealed by the blowup in the inset. This implies a lower bound  $t_p \approx 10^{-1}$  on the cycle duration, below which the system ceases to operate as a heat engine.

As can be observed in Fig. 10(d), for small and large cycle durations, the cycle-time dependence of the total entropy productions  $\Delta S_{\text{tot}}^{\text{eff}}(t_p)$  and  $\Delta S_{\text{tot}}^{\pm}(t_p)$  exhibits asymptotic power-law behavior. Taylor expansions of the total entropy productions in  $t_p$  and  $1/t_p$ , respectively, give  $\Delta S_{\text{tot}}^{\pm} \propto \Delta S_{\text{tot}}^{\text{eff}} \propto t_p$  for short  $t_p$  and  $\Delta S_{\text{tot}}^{\pm} \propto 1/t_p$  for  $v \neq 0$ , and  $\Delta S_{\text{tot}}^{\text{eff}} \propto 1/t_p$

regardless of  $v$ , for long  $t_p$ . To be more specific, all the total entropy productions in question assume the form

$$\Delta S_{\text{tot}}^z = - \int_0^{t_p} dt \frac{1}{\mathcal{T}(t)} (\dot{Q} + \mathcal{F})(t), \quad (68)$$

where  $\mathcal{T} = T_{\text{eff}}$  and  $\mathcal{F} = 0$  for  $z = \text{eff}$ ,  $\mathcal{T} = T$  and  $\mathcal{F} = -2\mu k(T_{\text{eff}} - T)$  for  $z = -$ , and  $\mathcal{T} = T$  and  $\mathcal{F} = 2\mu k(T_{\text{eff}} - T) - v^2/\mu$  for  $z = +$ . For fast driving of the engine, ( $t_p$  much smaller than the intrinsic relaxation times), the colloid cannot react to the changing driving and settles on a time-independent state corresponding to a mean value of the driving. Hence, Eq. (68) can be approximated for all  $z$  by  $\Delta S_{\text{tot}}^z \approx -t_p (\dot{Q} + \mathcal{F})/\mathcal{T}$ , where the integrand is evaluated using the time-independent state attained for  $t_p \rightarrow 0$ .

For slow driving ( $t_p$  much larger than the intrinsic relaxation times), the colloid attains its steady state (50) independent of the cycle duration  $t_p$ . Substituting the integration time  $t$  in Eq. (68) by the dimensionless time  $\tau = t/t_p$  yields

$$\Delta S_{\text{tot}}^z = -t_p \int_0^1 d\tau \frac{1}{\tilde{\mathcal{T}}(\tau)} \left( \frac{1}{t_p} \frac{d\tilde{Q}(\tau)}{d\tau} + \tilde{\mathcal{F}}(\tau) \right), \quad (69)$$

where  $\tilde{\mathcal{T}}(\tau) = \mathcal{T}(\tau t_p)$ ,  $\tilde{Q}(\tau) = Q(\tau t_p)$ , and  $\tilde{\mathcal{F}}(\tau) = \mathcal{F}(\tau t_p)$ . The effective total entropy production  $\Delta S_{\text{tot}}^{\text{eff}}$  vanishes in the limit  $t_p \rightarrow \infty$ , and thus the leading contribution in Eq. (69) is expected to be of order  $1/t_p$ . Indeed, expanding under the integral, we obtain (for  $\mathcal{F} = 0$ )

$$\frac{1}{\tilde{\mathcal{T}}} \frac{d\tilde{Q}}{d\tau} \approx \frac{d}{d\tau} \ln \sigma_\infty + \frac{1}{t_p} C. \quad (70)$$

Since the first term represents a total derivative, the corresponding loop integral vanishes and what remains is the correction  $C/t_p$  with a  $t_p$  independent constant  $C$ . For  $z = \pm$ , the leading contribution to the integral (69) is simply determined by the nonzero value of  $\lim_{t_p \rightarrow \infty} \mathcal{F}$  and thus we find  $\Delta S_{\text{tot}}^\pm \propto t_p$  for large  $t_p$ . For  $v = 0$ , all three definitions of entropy production are equivalent since then  $\mathcal{T} = T = T_{\text{eff}}$  and  $\mathcal{F} = 0$ . This proves the scalings found in Fig. 10(d).

## IX. CONCLUSION AND OUTLOOK

We argued on a very general basis that energy extracted from nonequilibrium reservoirs by cyclically operating engines qualifies as heat only if there exists a precise mapping to an equivalent cycle with an equilibrium bath at a time-dependent effective temperature, which yields the same power and efficiency. We have discussed the most general setting when such a mapping always exists and explained that engines which do not allow for a consistent definition of effective temperature should rather be understood as (possibly loss-making) work-to-work converters than heat engines. A benefit of the effective-temperature mapping is that conventional bounds on both the finite-time and the quasistatic thermodynamic performance of machines, especially heat engines, become applicable to those with nonequilibrium (active) baths [6,30–32]. As a part of our discussion, we have therefore been able to provide a new perspective on recent claims of surprisingly high Stirling efficiencies (surpassing the second law bound corresponding to infinite temperature steps) in

a bacterial heat engine that was experimentally realized by Krishnamurthy *et al.* [6].

To exemplify the general findings, we have derived a simple strategy to map the average thermodynamics of a linear Langevin system with arbitrary additive noise to an effective equilibrium system. The mapping is based on the matching of the dynamical equations for the second moment of position, which happens to determine the (average) energetics. It is valid for arbitrary protocols imposed by the time-dependent model parameters. In the quasistatic limit, the (generally time-dependent) effective temperature  $T_{\text{eff}}(t)$  (14) that accomplishes the mapping recovers the known expression (19).

We have further exemplified these somewhat abstract general notions by a fully worked example of a specific engine design that we call the ABE, since the particle dynamics is based on the well-known active Brownian particle (ABP) model. Our qualitative conclusions should carry over to other designs, though. In particular, we find that the explicitly computed effective temperature  $T_{\text{eff}}$  has some nonintuitive features. (i) During the limit cycle, which is attained by the ABE at long times, it obeys a first-order differential equation and thus acquires some time dependence  $T_{\text{eff}}(t)$  with a technically relevant characteristic relaxation time. (ii) It is important to realize that it can therefore vary in time even during those parts of the cycle in which the model parameters are held constant. (iii) Even in the quasistatic limit,  $T_{\text{eff}}$  depends on the strength of the potential. This means that realizing specific thermodynamic conditions, like an “isothermal” process with respect to the effective temperature, is generally not trivial.

The ABE model is also instructive with respect to some limitations of the effective-bath mapping. Namely, by construction, the latter is blind to the potentially rich features of the nonequilibrium bath beyond the second moment of the particle position, which we identified as the working degree of freedom of the engine. The effective description thus misses the non-Gaussian shape of the positional probability density and the corresponding Shannon entropy, for example, and also all housekeeping heat fluxes required to maintain the bath activity. Accordingly, we could demonstrate that the entropy production in the effective model can be understood as a lower bound for all conceivable practical and theoretical realizations. Namely, it vanishes upon quasistatic operation, whereas any detailed model of the bath dynamics would, like the explicitly studied ABE, necessarily reveal some of the housekeeping heat fluxes and their associated entropy production.

As an outlook, it would be interesting to study possible generalizations of our analysis of the linear model for arbitrary time-dependent friction kernels and correlation matrices, thus also including under-damped dynamics, which does not belong into the class of systems where the effective temperature always exists. Another possible extension could be the application of the presented method to nonlinear systems, e.g., by deriving approximate time-dependent effective temperatures via suitable closures of the equations describing the relevant degrees of freedom. Our general analysis shows that at least for Hamiltonians of the form  $\mathcal{H} = k(t)h(\mathbf{x})$ , with an arbitrary function  $h(\mathbf{x})$  diverging at  $|\mathbf{x}| \rightarrow \infty$ , this should always be possible.

## ACKNOWLEDGMENTS

We acknowledge funding by Deutsche Forschungsgemeinschaft (DFG) via SPP 1726/1 and KR 3381/6-1, and by Czech Science Foundation (project No. 20-02955J). V.H. gratefully acknowledges support by the Humboldt foundation. S.S. acknowledges funding by International Max Planck Research Schools (IMPRS).

## APPENDIX A: ANALYTICAL SOLUTION FOR VARIANCE

Inserting the time correlation matrix (41) for the ABP model into Eq. (17), Eq. (15) yields the following dynamic equation for the variance  $\sigma = \langle \mathbf{r} \cdot \mathbf{r} \rangle = \langle x^2 + y^2 \rangle$ :

$$\dot{\sigma} + 2\mu k\sigma = 4\langle x_0\eta_x(t) + y_0\eta_y(t) \rangle e^{-K(t,t_0)} + 4D(t) + 2v(t) \int_{t_0}^t dt' v(t') e^{-F(t,t')}, \quad (\text{A1})$$

where

$$K(t, t_0) = \mu \int_{t_0}^t dt' k(t'), \quad (\text{A2})$$

$$F(t, t_0) = K(t, t_0) + \int_{t_0}^t dt' D_r(t'). \quad (\text{A3})$$

In order to explicitly evaluate the thermodynamics of the particular realization of an active Brownian heat engine described in Sec. VI, namely, the ABP-based engine that we refer to as the ABE model, we need the solution of Eq. (A1). More precisely, we can concentrate onto the time periodic

solution, which is attained by the system at late times, after transients have relaxed, so that it settles onto a limit cycle (cf. Fig. 4). Taking the limit  $t_0 \rightarrow -\infty$  in the formal solution to Eq. (A1), we obtain

$$\sigma(t) = 2 \lim_{t_0 \rightarrow -\infty} \int_{t_0}^t dt' [2D(t') + v(t')H(t')] e^{-2K(t,t')} \quad (\text{A4})$$

with

$$H(t) = \lim_{t_0 \rightarrow -\infty} \int_{t_0}^t dt' v(t') e^{-F(t,t')}. \quad (\text{A5})$$

For the numerical evaluation of Eq. (A4) it is useful to exploit that  $H(t)$  is a  $t_p$ -periodic function and to rewrite  $K(t, t_0)$  as  $K(t, t_0) = \lfloor (t - t_0)/t_p \rfloor K(t_p, 0) + K(t, t_0 + \lfloor (t - t_0)/t_p \rfloor t_p)$  using the  $t_p$ -periodicity of  $k(t)$  (the symbol  $\lfloor x \rfloor$  denotes the floor operation) and similarly for  $F(t, t_0)$ . Interestingly, using a simple trick, the time-periodic late-time limit can be found without considering the (numerically inconvenient) limit  $t_0 \rightarrow -\infty$ , just as in the case of memoryless dynamics [25,26]. The key insight is that, in the long-time regime, the functions  $\sigma$  and  $H$  obey two coupled ordinary differential equations, namely Eqs. (44) and (45) in Sec. VIC, which follow from Eqs. (A4) and (A5) by taking derivative with respect to  $t$ .

## APPENDIX B: SLOW DRIVING LIMIT OF VARIANCE

For slowly varying driving functions  $k(t)$ ,  $D(t)$ ,  $D_r(t)$  and  $v(t)$ , the variance (A4) can be approximated using a simple formula which follows from the Laplace type approximation of the integral [112,113]

$$\int_{t_0}^t dt f(t') e^{-\int_{t_0}^{t'} dt'' g(t'')} = \int_{t_0}^t dt f(t') e^{-t_p \int_{t_0}^{t'/t_p} dt'' g(t_p t'')} = \frac{f(t)}{g(t)} - \frac{1}{g^2(t)} \left[ f(t) - f(t) \frac{\dot{g}(t)}{g(t)} \right] + o(\dot{f}, \dot{g}). \quad (\text{B1})$$

Applying this approximation first on the function  $H(t)$  (A5) and then on the variance  $\sigma(t)$  (A4), we obtain the approximate result

$$\sigma(t) = \sigma_\infty - \frac{v^2}{k\mu\kappa^2} \left( \frac{\dot{v}}{v} - \frac{\dot{\kappa}}{\kappa} \right) - \frac{D}{k^2\mu^2} \left( \frac{\dot{D}}{D} - \frac{\dot{k}}{k} \right) - \frac{v^2}{2k^2\mu^2\kappa} \left( 2\frac{\dot{v}}{v} - \frac{\dot{\kappa}}{\kappa} - \frac{\dot{k}}{k} \right) + o(\dot{v}, \dot{D}, \dot{k}, \dot{\kappa}). \quad (\text{B2})$$

Here,  $\sigma_\infty$  is the variance (50) for infinitely slow driving and  $\kappa = \kappa(t) = k\mu + D_r$ . For discontinuous driving, the limiting solution  $\sigma_\infty$  is also discontinuous. The first order correction (B2) may also be discontinuous if the first derivatives of the driving functions exhibit jumps. In such a case, however, the assumption on the smallness of the derivatives used in the calculation leading to Eq. (B2) is not valid. In accord with the discussion below Eq. (D1) in Appendix D, Eq. (B2) reveals that activity-corrections are at least second order in  $v$ .

## APPENDIX C: ENTROPY PRODUCTION FROM PATH PROBABILITIES

The entropy

$$\Delta S_{R,\Gamma}(t) = \ln(P_F/P_R) \quad (\text{C1})$$

delivered to the bath by a particle moving along a trajectory  $\Gamma(t) = \{\mathbf{r}(t'), \theta(t')\}_{t'=0}^t$  of the stochastic process (42), (43) is given by the logarithm of the ratio of conditional probabilities  $P_F$  and  $P_R$  [103,114], for the trajectory conditioned with

respect to its initial point and its time-reversed image. Up to normalization, the forward probability is given by

$$P_F \propto e^{-2 \int_0^t dt' [\xi \cdot \xi + \xi_\theta^2]}, \quad (\text{C2})$$

where the noise terms  $\xi = [\dot{\mathbf{r}} + \mu \nabla_r \mathcal{H} - \mathbf{v}] / \sqrt{2D}$  and  $\xi_\theta = \dot{\theta} / \sqrt{2D_r}$  follow from Eqs. (42) and (43) [115]. The backward probability is given by a similar formula. One just has to change the sign before quantities which are odd with respect to time reversal.

Assuming the active velocity  $\mathbf{v} = v(\cos \theta, \sin \theta)$  to be time-reversal even, the odd variables in Eqs. (42) and (43) are only time derivatives, giving

$$(P_F/P_R)^+ = e^{-\int_0^t dt' (\nabla_r \mathcal{H} - \mathbf{v}/\mu) \cdot \dot{\mathbf{r}}/T}, \quad (\text{C3})$$

whereas, for time-reversal odd  $\mathbf{v}$ , we find

$$(P_F/P_R)^- = e^{-\int_0^t dt' \nabla_r \mathcal{H} \cdot (\dot{\mathbf{r}} - \mathbf{v})/T}. \quad (\text{C4})$$

The entropy delivered to the reservoir during time interval  $(0, t)$  follows as

$$\Delta S_R(t) = \langle \Delta S_{R,\Gamma}(t) \rangle_\Gamma = \langle \ln(P_F/P_R) \rangle_\Gamma, \quad (\text{C5})$$

where the average is taken over the individual realizations  $\Gamma$  of the stochastic process [114]. With Eq. (C3) for the time-even active velocity, it yields

$$\Delta S_R^+(t) = \int_0^t dt' \frac{1}{T} \left\langle \left( \frac{\mathbf{v}}{\mu} - \nabla_r \mathcal{H} \right) \cdot \dot{\mathbf{r}} \right\rangle, \quad (\text{C6})$$

and with Eq. (C4), for the time-odd active velocity,

$$\Delta S_R^-(t) = \int_0^t dt' \frac{1}{T} \langle (\dot{\mathbf{r}} - \mathbf{v}) \cdot (-\nabla_r \mathcal{H}) \rangle. \quad (\text{C7})$$

#### APPENDIX D: PROBABILITY DISTRIBUTIONS (PDFS)

In the three-dimensional Langevin system [(42) and (43)], the  $x$ - $y$  coordinates are coupled via the active velocity  $\mathbf{v}$ . The steady probability distribution (PDF) to find the particle with orientation  $\theta$  at position  $(x, y)$  thus cannot generally be written in the separated form  $p(x, y, \theta) = \chi(x, \theta)\iota(y, \theta) = \chi(x, \theta)\chi(y, \pi/2 - \theta)$ , where  $\chi(x, \theta)$  solves the two-dimensional Fokker-Planck equation

$$\partial_t \chi = [D\partial_x^2 + D_r\partial_\theta^2 + \partial_x(\mu k\partial_x x - v \cos \theta)]\chi. \quad (\text{D1})$$

Inserting the separation ansatz into the three-dimensional equation (62) and using the formula (D1) leads to the condition  $2D_r\partial_\theta\chi(x, \theta)\partial_\theta\iota(y, \theta) = 0$  that cannot be fulfilled in general. Nevertheless, one can still reduce the three-dimensional system to just two degrees of freedom by introducing the polar coordinates  $x = r \cos \phi$ ,  $y = r \sin \phi$ . Then, Eq. (42) transforms to

$$\dot{r} = -\mu kr + v \cos(\theta - \phi) + \sqrt{2D}\eta_r, \quad (\text{D2})$$

$$\dot{\phi} = \frac{v}{r} \sin(\theta - \phi) + \sqrt{\frac{2D}{r^2}}\eta_\phi, \quad (\text{D3})$$

while  $\theta$  still obeys Eq. (43). The symbols  $\eta_r$  and  $\eta_\phi$  denote independent, zero-mean, Gaussian white noises. Since Eqs. (D2) and (D3) only depend on the difference  $\theta - \phi$ , introducing the relative angle  $\psi = \theta - \phi$ , subject to the zero-mean, Gaussian white noise  $\eta_\psi$  renders them in the form

$$\dot{r} = -\mu kr + v \cos \psi + \sqrt{2D}\eta_r, \quad (\text{D4})$$

$$\dot{\psi} = -\frac{v}{r} \sin \psi + \sqrt{2\left(\frac{D}{r^2} + D_r\right)}\eta_\psi. \quad (\text{D5})$$

The corresponding Fokker-Planck equation for the PDF  $\rho = \rho(r, \psi, t)$  reads [116]

$$\begin{aligned} \partial_t \rho = & \left[ D\partial_r^2 + \left( \frac{D}{r^2} + D_r \right) \partial_\psi^2 \right] \rho - \cos \psi \partial_r (v\rho) \\ & - D\partial_r \left( \frac{\rho}{r} \right) + \mu k \partial_r (r\rho) + \frac{v}{r} \partial_\psi (\sin \psi \rho). \end{aligned} \quad (\text{D6})$$

In general, Eqs. (D1) and (D6) [or equivalently (62)] can not be solved analytically and thus we solved them using the numerical method described in Ref. [102]. We

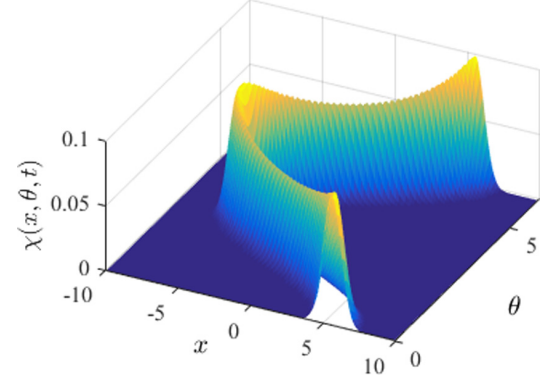


FIG. 11. Probability distribution  $\chi$  for particle position  $x$  and orientation  $\theta$  at the end of the hot isotherm ( $t = 3t_p/4$ , see Fig. 3). We take  $v_> = 30$  and  $t_p = 10^4$ . Other parameters are the same as in Fig. 3.

compared the numerical solution of Eq. (D6) to the separated ansatz  $p(x, y, \theta) = \chi(x, \theta)\chi(y, \pi/2 - \theta)$  and found out that, although not exact, the ansatz describes the full three-dimensional PDF  $p(x, y, \theta)$  sufficiently well. Since the two-dimensional PDF allows for a more intuitive discussion and exhibits the main qualitative features of  $p(x, y, \theta)$ , we restrict the following discussion to  $\chi(x, \theta)$ .

Figure 11 shows a snapshot of the PDF  $\chi(x, \theta, t)$ , solution of (D1), at the end of the third branch of a quasistatic cycle introduced in Sec. VIB (the hot “isotherm”). The figure reveals the typical shape of the PDF  $\chi$ , with two global maxima located at  $\theta = 0$  and  $\pi$ , which survives even for rapid driving protocols. Physically, the shape of the PDF can be understood as follows: (1) for any fixed orientation angle  $\theta$ , the PDF can be expected to exhibit a maximum at the position where the active velocity [which acts in the Langevin Eq. (42) for  $x$  as a force  $v \cos \theta / \mu$ ] is balanced by the force  $kx$  exerted by the parabolic potential; (2) the projection  $v \cos \theta / \mu$  of  $\mathbf{v}$  on the  $x$  coordinate changes slowest around its extrema (0 and  $\pi$ ), and thus most trajectories contribute to the surroundings of these points, making the extrema for 0 and  $\pi$  largest.

Figure 12 shows snapshots of the marginal PDF  $\rho(x, t) = \int d\theta \chi(x, \theta, t)$  for the position  $x$  at the beginning of the individual branches of the cycle, for four values of the maximum active velocity  $v_>$ . With increasing  $v_>$ , the resulting PDFs become increasingly non-Gaussian and finally even exhibit two separated peaks. Physically, this behavior can be understood by the wall accumulation effect due to the persistence of the active motion [117–119], which creates the double peak during the cycle branches with large  $v_>$ . (For similar PDFs, see Refs. [101,120].) Qualitatively similar results are also obtained in the quasistatic limit, as already apparent from Fig. 11.

To get some intuition about these results on analytical grounds, we now present several approximate solutions to Eq. (D1). Different from the standard diffusion ( $v = 0$ ) in an external potential, the quasistatic ( $\partial_t \chi = 0$ ) solution of the Fokker-Planck equation (D1) is not given by the Boltzmann PDF. This is because one cannot subsume the activity into a generalized potential  $\tilde{\mathcal{H}}$  which would act as a Lyapunov functional for the dynamics of  $x$  and  $\theta$ . Nevertheless, there

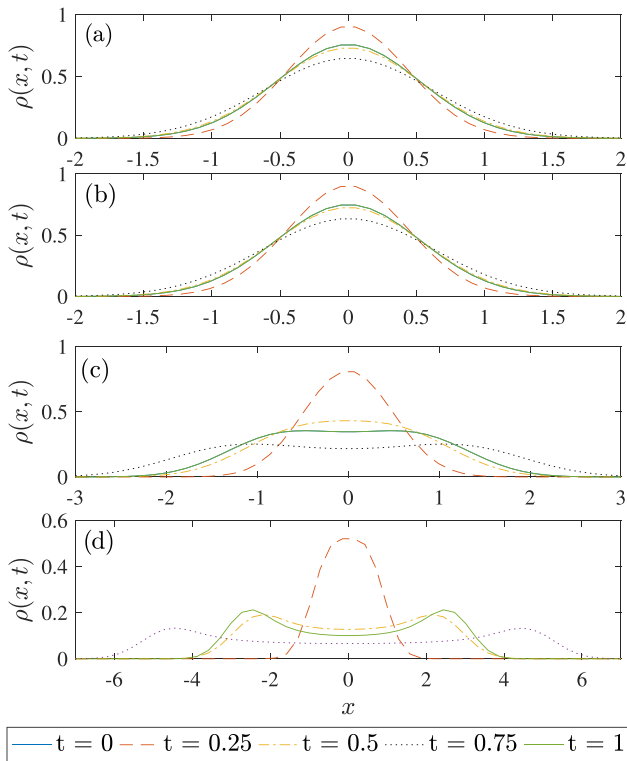


FIG. 12. Marginal distribution  $\rho$  for the particle position  $x$  at the end of the individual branches of the cycle for different values of the maximum active velocity (a)  $v_s = 0$ , (b) 1, (c) 10, and (d) 30. We have set  $t_p = 1$ , corresponding to nonstationary driving, other parameters as in Fig. 3. Note that the curves at  $t = 0$  and  $t = 1$  are equal, in accord with the time periodic operation.

are several limiting cases where the Boltzmann form  $\chi \propto \exp(-\tilde{\mathcal{H}}/T)$  is still a useful approximation.

The best analytical insight into the described qualitative properties of the presented numerical solutions to Eq. (D1) with time-dependent parameters is obtained for rotational diffusion coefficient  $D_r$  much smaller than  $k\mu$ , corresponding to the limit of large  $\mathcal{K}$  in Eq. (63). Then, the direction of the active velocity can be treated as quenched, so that the activity can be subsumed into a generalized potential  $\tilde{\mathcal{H}} = kx^2/2 - vx \cos \theta/\mu$ . The corresponding quasistatic solution

of Eq. (D1) then reads

$$\chi = \frac{1}{Z_\chi} \exp\left(\frac{vx \cos \theta}{\mu T} - \frac{kx^2}{2T}\right), \quad (\text{D7})$$

with a normalization constant  $Z_\chi$ . For each fixed value of the angle  $\theta$ , the PDF is then Gaussian with its maximum value  $\exp[v^2 \cos^2 \theta / (2T \mu^2 k)] / Z_\chi$  at the position  $v \cos \theta / (\mu k)$ . The PDF thus possesses two global maxima located at  $(x, \theta) = [v/(\mu k), 0]$  and  $(x, \theta) = [-v/(\mu k), \pi]$ , and is qualitatively similar to the PDF shown in Fig. 11.

The marginal PDF for  $x$  obtained from (D7) then reads

$$\rho(x, t) = \int d\theta \chi = \frac{1}{Z_\rho} \exp\left(-\frac{kx^2}{2T}\right) I_0\left(\frac{vx}{\mu T}\right). \quad (\text{D8})$$

Here,  $I_0(x)$  denotes the modified Bessel function of the first kind and  $Z_\rho$  is another normalization constant. The marginal PDF is Gaussian for  $v = 0$ , and becomes more and more non-Gaussian with increasing  $v/(\mu k)$ . For large values of  $v/(\mu k)$ , it can even become bimodal. This behavior can be traced back to the shift of the maxima of the PDF  $\chi$  with increasing  $v/(\mu k)$ . For small  $v/(\mu k)$ , the two maxima substantially overlap and the integration over the angle  $\theta$  yields a single peak which is nearly Gaussian. For large values of  $v/(\mu k)$ , the two peaks do not overlap any more and the marginal PDF thus also exhibits two peaks. The behavior of the marginal PDF obtained in the limit  $D_r \ll \mu k$  thus shows qualitatively the same behavior as the solution of Eq. (D1) shown in Fig. 12.

For  $D_r$  much larger than  $k\mu$ , corresponding to the limit of small  $\mathcal{K}$  in Eq. (63), the quasistatic PDF is given by  $\chi \propto \exp(-\mathcal{H}/T_{\text{eff}})$ . This is because the rotational diffusion obliterates any persistence of the active motion, and the nonequilibrium bath effectively behaves like an equilibrium one with the renormalized temperature  $T_{\text{eff}} = T + v^2/(2\mu D_r)$ . In this limit, the degrees of freedom  $x$  and  $y$  also become independent.

Yet another case admitting an analytical solution of Eq. (D1) is that of quasistatic driving at small active velocity. Then the quasistatic PDF  $\rho$  can be approximated by the McLennan-type form  $\chi \approx \exp(-\mathcal{H}/T)[1 - W(x)]$  [121–125]. Without going into details, the function  $W(x)$  is in general proportional to the (average) dissipation in the driven system [123], which, in our case, is given by the product of the active “force”  $\mu^{-1}v \cos \theta$  and the particle velocity  $\dot{x}$ . Since the average over the angle  $\theta$  of the active force is zero, the correction  $W(x)$  to the particle PDF is seen to be at least second order in  $v$ .

- [1] S. Sánchez, L. Soler, and J. Katuri, Chemically powered micro- and nanomotors, *Angew. Chem. Int. Ed.* **54**, 1414 (2015).
- [2] P. Hänggi and F. Marchesoni, Artificial brownian motors: Controlling transport on the nanoscale, *Rev. Mod. Phys.* **81**, 387 (2009).
- [3] G. A. Ozin, I. Manners, S. Fournier-Bidoz, and A. Arsenault, Dream nanomachines, *Adv. Mater.* **17**, 3011 (2005).
- [4] J. R. Baylis, J. H. Yeon, M. H. Thomson, A. Kazerooni, X. Wang, A. E. St. John, E. B. Lim, D. Chien, A. Lee, J. Q. Zhang, J. M. Piret, L. S. Machan, T. F. Burke, N. J. White, and C. J.

Kastrup, Self-propelled particles that transport cargo through flowing blood and halt hemorrhage, *Sci. Adv.* **1**, e1500379 (2015).

- [5] T. Speck, Stochastic thermodynamics for active matter, *Europhys. Lett.* **114**, 30006 (2016).
- [6] S. Krishnamurthy, S. Ghosh, D. Chatterji, R. Ganapathy, and A. K. Sood, A micrometre-sized heat engine operating between bacterial reservoirs, *Nat. Phys.* **12**, 1134 (2016).
- [7] U. Seifert, Entropy Production Along a Stochastic Trajectory and an Integral Fluctuation Theorem, *Phys. Rev. Lett.* **95**, 040602 (2005).


- [8] M. Esposito, U. Harbola, and S. Mukamel, Nonequilibrium fluctuations, fluctuation theorems, and counting statistics in quantum systems, *Rev. Mod. Phys.* **81**, 1665 (2009).
- [9] K. Sekimoto, *Stochastic Energetics*, Lecture Notes in Physics, Vol. 799 (Springer-Verlag, Berlin, Heidelberg, 2010).
- [10] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, *Rep. Prog. Phys.* **75**, 126001 (2012).
- [11] O. Abah, J. Roßnagel, G. Jacob, S. Deffner, F. Schmidt-Kaler, K. Singer, and E. Lutz, Single-Ion Heat Engine at Maximum Power, *Phys. Rev. Lett.* **109**, 203006 (2012).
- [12] J. Roßnagel, S. T. Dawkins, K. N. Tolazzi, O. Abah, E. Lutz, F. Schmidt-Kaler, and K. Singer, A single-atom heat engine, *Science* **352**, 325 (2016).
- [13] J.-P. Brantut, C. Grenier, J. Meineke, D. Stadler, S. Krinner, C. Kollath, T. Esslinger, and A. Georges, A thermoelectric heat engine with ultracold atoms, *Science* **342**, 713 (2013).
- [14] V. Blickle and C. Bechinger, Realization of a micrometre-sized stochastic heat engine, *Nat. Phys.* **8**, 143 (2012).
- [15] I. Martinez, É. Roldán, L. Dinis, D. Petrov, J. Parrondo, and R. R.A., Brownian carnot engine, *Nat. Phys.* **12**, 67 (2015).
- [16] I. A. Martinez, E. Roldan, L. Dinis, and R. A. Rica, Colloidal heat engines: A review, *Soft Matter* **13**, 22 (2017).
- [17] M. Esposito, R. Kawai, K. Lindenberg, and C. Van den Broeck, Quantum-dot carnot engine at maximum power, *Phys. Rev. E* **81**, 041106 (2010).
- [18] G. Verley, M. Esposito, T. Willaert, and C. Van den Broeck, The unlikely carnot efficiency, *Nat. Commun.* **5**, 4721 (2014).
- [19] K. E. Dorfman, D. V. Voronine, S. Mukamel, and M. O. Scully, Photosynthetic reaction center as a quantum heat engine, *Proc. Natl. Acad. Sci. USA* **110**, 2746 (2013).
- [20] V. Holubec and A. Ryabov, Work and power fluctuations in a critical heat engine, *Phys. Rev. E* **96**, 030102(R) (2017).
- [21] M. Campisi and R. Fazio, The power of a critical heat engine, *Nat. Commun.* **7**, 11895 (2016).
- [22] P. Chvosta, M. Einax, V. Holubec, A. Ryabov, and P. Maass, Energetics and performance of a microscopic heat engine based on exact calculations of work and heat distributions, *J. Stat. Mech.* (2010) P03002.
- [23] R. Kosloff and Y. Rezek, The quantum harmonic otto cycle, *Entropy* **19**, 136 (2017).
- [24] R. S. Whitney, Most Efficient Quantum Thermoelectric at Finite Power Output, *Phys. Rev. Lett.* **112**, 130601 (2014).
- [25] T. Schmiedl and U. Seifert, Efficiency at maximum power: An analytically solvable model for stochastic heat engines, *Europhys. Lett.* **81**, 20003 (2008).
- [26] V. Holubec, An exactly solvable model of a stochastic heat engine: Optimization of power, power fluctuations and efficiency, *J. Stat. Mech.* (2014) P05022.
- [27] J. M. Horowitz and J. M. Parrondo, Thermodynamics: A stirring effort, *Nat. Phys.* **8**, 108 (2012).
- [28] J. Roßnagel, O. Abah, F. Schmidt-Kaler, K. Singer, and E. Lutz, Nanoscale Heat Engine Beyond the Carnot Limit, *Phys. Rev. Lett.* **112**, 030602 (2014).
- [29] W. Niedenzu, V. Mukherjee, A. Ghosh, A. G. Kofman, and G. Kurizki, Quantum engine efficiency bound beyond the second law of thermodynamics, *Nat. Commun.* **9**, 165 (2018).
- [30] R. Zakine, A. Solon, T. Gingrich, and F. van Wijland, Stochastic stirling engine operating in contact with active baths, *Entropy* **19**, 193 (2017).
- [31] D. Martin, C. Nardini, M. E. Cates, and É. Fodor, Extracting maximum power from active colloidal heat engines, *Europhys. Lett.* **121**, 60005 (2018).
- [32] A. Kumari, P. S. Pal, A. Saha, and S. Lahiri, Stochastic heat engine using an active particle, *Phys. Rev. E* **101**, 032109 (2020).
- [33] G. Gompper, C. Bechinger, S. Herminghaus, R. Isele-Holder, U. B. Kaupp, H. Löwen, H. Stark, and R. G. Winkler, Microswimmers - from single particle motion to collective behavior, *Eur. Phys. J. Spec. Top.* **225**, 2061 (2016).
- [34] M. Cates, Diffusive transport without detailed balance in motile bacteria: Does microbiology need statistical physics? *Rep. Prog. Phys.* **75**, 042601 (2012).
- [35] L. Dabelow, S. Bo, and R. Eichhorn, Irreversibility in Active Matter Systems: Fluctuation Theorem and Mutual Information, *Phys. Rev. X* **9**, 021009 (2019).
- [36] E. Crosato, M. Prokopenko, and R. E. Spinney, Irreversibility and emergent structure in active matter, *Phys. Rev. E* **100**, 042613 (2019).
- [37] A. Sokolov, M. M. Apodaca, B. A. Grzybowski, and I. S. Aranson, Swimming bacteria power microscopic gears, *Proc. Natl. Acad. Sci. USA* **107**, 969 (2010).
- [38] R. Di Leonardo, L. Angelani, D. Dell'Arciprete, G. Ruocco, V. Iebba, S. Schippa, M. Conte, F. Mecarini, F. De Angelis, and E. Di Fabrizio, Bacterial ratchet motors, *Proc. Natl. Acad. Sci. USA* **107**, 9541 (2010).
- [39] N. Nikola, A. P. Solon, Y. Kafri, M. Kardar, J. Tailleur, and R. Voituriez, Active Particles with Soft and Curved Walls: Equation of State, Ratchets, and Instabilities, *Phys. Rev. Lett.* **117**, 098001 (2016).
- [40] C. J. O. Reichhardt and C. Reichhardt, Ratchet effects in active matter systems, *Annu. Rev. Condens. Matter Phys.* **8**, 51 (2017).
- [41] G. Vizsnyiczai, G. Frangipane, C. Maggi, F. Saglimbeni, S. Bianchi, and R. Di Leonardo, Light controlled 3d micromotors powered by bacteria, *Nat. Commun.* **8**, 15974 (2017).
- [42] P. Pietzonka, A. C. Barato, and U. Seifert, Universal bound on the efficiency of molecular motors, *J. Stat. Mech.* (2016) 124004.
- [43] P. Pietzonka, E. Fodor, C. Lohrmann, M. E. Cates, and U. Seifert, Autonomous Engines Driven by Active Matter: Energetics and Design Principles, *Phys. Rev. X* **9**, 041032 (2019).
- [44] D. Rings, R. Schachoff, M. Selmeke, F. Cichos, and K. Kroy, Hot Brownian Motion, *Phys. Rev. Lett.* **105**, 090604 (2010).
- [45] G. Falasco, M. V. Gnann, D. Rings, and K. Kroy, Effective temperatures of hot brownian motion, *Phys. Rev. E* **90**, 032131 (2014).
- [46] K. Kroy, D. Chakraborty, and F. Cichos, Hot microswimmers, *Eur. Phys. J. Spec. Top.* **225**, 2207 (2016).
- [47] D. Geiß and K. Kroy, Brownian thermometry beyond equilibrium, *ChemSystemsChem* **2**, e1900041 (2020).
- [48] M. Smoluchowski, Experimentell nachweisbare, der üblichen thermodynamik widersprechende molekularphänomene, *Pisma Mariana Smoluchowskiego* **2**, 226 (1927).
- [49] R. P. Feynman, R. B. Leighton, and M. Sands, *The New Millennium Edition: Mainly Mechanics, Radiation, and Heat*, The Feynman Lectures on Physics Vol. 1 (Basic books, New York, 2011).
- [50] A. Ryabov, V. Holubec, M. H. Yaghoubi, M. Varga, M. E. Foulaadvand, and P. Chvosta, Transport coefficients for a con-



- finned brownian ratchet operating between two heat reservoirs, *J. Stat. Mech.* (2016) 093202.
- [51] V. Holubec, A. Ryabov, M. H. Yaghoubi, M. Varga, A. Khodaei, M. E. Fouladadvand, and P. Chvosta, Thermal ratchet effect in confining geometries, *Entropy* **19**, 119 (2017).
- [52] J. S. Lee and H. Park, Carnot efficiency is reachable in an irreversible process, *Sci. Rep.* **7**, 10725 (2017).
- [53] P. Kalinay and F. Slanina, Feynman-smoluchowski ratchet in an effective one-dimensional picture, *Phys. Rev. E* **98**, 042141 (2018).
- [54] A. Saha, R. Marathe, P. S. Pal, and A. M. Jayannavar, Stochastic heat engine powered by active dissipation, *J. Stat. Mech.* (2018) 113203.
- [55] A. Saha and R. Marathe, Stochastic work extraction in a colloidal heat engine in the presence of colored noise, *J. Stat. Mech.* (2019) 094012.
- [56] T. Ekeh, M. E. Cates, and E. Fodor, Thermodynamic cycles with active matter, *Phys. Rev. E* **102**, 010101(R) (2020).
- [57] X. Zhao, K. K. Dey, S. Jeganathan, P. J. Butler, U. M. Córdova-Figueroa, and A. Sen, Enhanced diffusion of passive tracers in active enzyme solutions, *Nano Lett.* **17**, 4807 (2017).
- [58] H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics*, Student ed. (John Wiley and Sons, Singapore, 2006).
- [59] P. Malgaretti, P. Nowakowski, and H. Stark, Mechanical pressure and work cycle of confined active brownian particles, [arXiv:2008.05257](https://arxiv.org/abs/2008.05257).
- [60] C. Jarzynski, Nonequilibrium Equality for Free Energy Differences, *Phys. Rev. Lett.* **78**, 2690 (1997).
- [61] C. Jarzynski, Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach, *Phys. Rev. E* **56**, 5018 (1997).
- [62] G. E. Crooks, Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems, *J. Stat. Phys.* **90**, 1481 (1998).
- [63] G. E. Crooks, Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences, *Phys. Rev. E* **60**, 2721 (1999).
- [64] J. Casas-Vazquez and D. Jou, Temperature in non-equilibrium states: A review of open problems and current proposals, *Rep. Prog. Phys.* **66**, 1937 (2003).
- [65] Y. Fily and M. C. Marchetti, Athermal Phase Separation of Self-Propelled Particles with no Alignment, *Phys. Rev. Lett.* **108**, 235702 (2012).
- [66] G. Szamel, Self-propelled particle in an external potential: Existence of an effective temperature, *Phys. Rev. E* **90**, 012111 (2014).
- [67] X.-L. Wu and A. Libchaber, Particle Diffusion in a Quasi-Two-Dimensional Bacterial Bath, *Phys. Rev. Lett.* **84**, 3017 (2000).
- [68] L. Angelani, C. Maggi, M. L. Bernardini, A. Rizzo, and R. Di Leonardo, Effective Interactions between Colloidal Particles Suspended in a Bath of Swimming Cells, *Phys. Rev. Lett.* **107**, 138302 (2011).
- [69] J. Harder, S. A. Mallory, C. Tung, C. Valeriani, and A. Cacciuto, The role of particle shape in active depletion, *J. Chem. Phys.* **141**, 194901 (2014).
- [70] T. F. F. Farage, P. Krinninger, and J. M. Brader, Effective interactions in active brownian suspensions, *Phys. Rev. E* **91**, 042310 (2015).
- [71] W. T. Coffey, Y. P. Kalmykov, and J. T. Waldron, *The Langevin Equation* (World Scientific, Singapore, 2004).
- [72] V. Holubec and A. Ryabov, Diverging, but negligible power at carnot efficiency: Theory and experiment, *Phys. Rev. E* **96**, 062107 (2017).
- [73] V. Holubec and A. Ryabov, Cycling Tames Power Fluctuations Near Optimum Efficiency, *Phys. Rev. Lett.* **121**, 120601 (2018).
- [74] H. Callen, H. Callen, N. F. R. C. of Australia. Research Division, and W. Sons, *Thermodynamics and an Introduction to Thermostatistics* (John Wiley and Sons, Singapore, 1985).
- [75] V. Holubec and A. Ryabov, Maximum efficiency of low-dissipation heat engines at arbitrary power, *J. Stat. Mech.* (2016) 073204.
- [76] R. Kubo, The fluctuation-dissipation theorem, *Rep. Prog. Phys.* **29**, 255 (1966).
- [77] T. Hatano and S. I. Sasa, Steady-State Thermodynamics of Langevin Systems, *Phys. Rev. Lett.* **86**, 3463 (2001).
- [78] G. Szamel, Stochastic thermodynamics for self-propelled particles, *Phys. Rev. E* **100**, 050603(R) (2019).
- [79] T. R. Gingrich, J. M. Horowitz, N. Perunov, and J. L. England, Dissipation Bounds all Steady-State Current Fluctuations, *Phys. Rev. Lett.* **116**, 120601 (2016).
- [80] K. Proesmans and C. V. den Broeck, Discrete-time thermodynamic uncertainty relation, *Europhys. Lett.* **119**, 20001 (2017).
- [81] P. Pietzonka and U. Seifert, Universal Trade-Off between Power, Efficiency, and Constancy in Steady-State Heat Engines, *Phys. Rev. Lett.* **120**, 190602 (2018).
- [82] G. Falasco, M. Esposito, and J.-C. Delvenne, Unifying thermodynamic uncertainty relations, *New J. Phys.* **22**, 053046 (2020).
- [83] L. F. Cugliandolo, J. Kurchan, and L. Peliti, Energy flow, partial equilibration, and effective temperatures in systems with slow dynamics, *Phys. Rev. E* **55**, 3898 (1997).
- [84] G. Falasco, M. V. Gnann, and K. Kroy, Non-isothermal fluctuation-dissipation relations and brownian thermometry, [arXiv:1406.2116](https://arxiv.org/abs/1406.2116).
- [85] R. Wulfert, M. Oechsle, T. Speck, and U. Seifert, Driven brownian particle as a paradigm for a nonequilibrium heat bath: Effective temperature and cyclic work extraction, *Phys. Rev. E* **95**, 050103(R) (2017).
- [86] U. M. B. Marconi, A. Puglisi, and C. Maggi, Heat, temperature and clausius inequality in a model for active brownian particles, *Sci. Rep.* **7**, 46496 (2017).
- [87] S. Chaki and R. Chakrabarti, Entropy production and work fluctuation relations for a single particle in active bath, *Physica A* **511**, 302 (2018).
- [88] L. Caprini, U. Marini Bettolo Marconi, and A. Puglisi, Activity induced delocalization and freezing in self-propelled systems, *Sci. Rep.* **9**, 1386 (2019).
- [89] S. Chaki and R. Chakrabarti, Effects of active fluctuations on energetics of a colloidal particle: Superdiffusion, dissipation and entropy production, *Physica A* **530**, 121574 (2019).
- [90] L. Caprini, U. Marini Bettolo Marconi, A. Puglisi, and A. Vulpiani, Active escape dynamics: The effect of persistence on barrier crossing, *J. Chem. Phys.* **150**, 024902 (2019).
- [91] L. F. Cugliandolo, The effective temperature, *J. Phys. A: Math. Gen* **44**, 483001 (2011).

- [92] A. Puglisi, A. Sarracino, and A. Vulpiani, Temperature in and out of equilibrium: A review of concepts, tools and attempts, *Phys. Rep.* **709-710**, 1 (2017).
- [93] M. Esposito, R. Kawai, K. Lindenberg, and C. Van den Broeck, Efficiency at Maximum Power of Low-Dissipation Carnot Engines, *Phys. Rev. Lett.* **105**, 150603 (2010).
- [94] V. Holubec and A. Ryabov, Efficiency at and near maximum power of low-dissipation heat engines, *Phys. Rev. E* **92**, 052125 (2015).
- [95] Y.-H. Ma, D. Xu, H. Dong, and C.-P. Sun, Universal constraint for efficiency and power of a low-dissipation heat engine, *Phys. Rev. E* **98**, 042112 (2018).
- [96] C. Maes, On the second fluctuation–dissipation theorem for nonequilibrium baths, *J. Stat. Phys.* **154**, 705 (2014).
- [97] C. Maes and S. Steffenoni, Friction and noise for a probe in a nonequilibrium fluid, *Phys. Rev. E* **91**, 022128 (2015).
- [98] R. Zwanzig, Nonlinear generalized langevin equations, *J. Stat. Phys.* **9**, 215 (1973).
- [99] S. Steffenoni, K. Kroy, and G. Falasco, Interacting brownian dynamics in a nonequilibrium particle bath, *Phys. Rev. E* **94**, 062139 (2016).
- [100] U. M. B. Marconi, C. Maggi, and S. Melchionna, Pressure and surface tension of an active simple liquid: A comparison between kinetic, mechanical and free-energy based approaches, *Soft Matter* **12**, 5727 (2016).
- [101] X. Zheng, B. ten Hagen, A. Kaiser, M. Wu, H. Cui, Z. Silberli, and H. Löwen, Non-gaussian statistics for the motion of self-propelled janus particles: Experiment versus theory, *Phys. Rev. E* **88**, 032304 (2013).
- [102] V. Holubec, K. Kroy, and S. Steffenoni, Physically consistent numerical solver for time-dependent fokker-planck equations, *Phys. Rev. E* **99**, 032117 (2019).
- [103] C. Maes and K. Netočný, Time-reversal and entropy, *J. Stat. Phys.* **110**, 269 (2003).
- [104] C. Battle, C. P. Broedersz, N. Fakhri, V. F. Geyer, J. Howard, C. F. Schmidt, and F. C. MacKintosh, Broken detailed balance at mesoscopic scales in active biological systems, *Science* **352**, 604 (2016).
- [105] B. Qian, D. Montiel, A. Bregulla, F. Cichos, and H. Yang, Harnessing thermal fluctuations for purposeful activities: The manipulation of single micro-swimmers by adaptive photon nudging, *Chem. Sci.* **4**, 1420 (2013).
- [106] U. Khadka, V. Holubec, H. Yang, and F. Cichos, Active particles bound by information flows, *Nat. Commun.* **9**, 3864 (2018).
- [107] T. Speck, J. Mehl, and U. Seifert, Role of External Flow and Frame Invariance in Stochastic Thermodynamics, *Phys. Rev. Lett.* **100**, 178302 (2008).
- [108] E. Fodor, C. Nardini, M. E. Cates, J. Tailleur, P. Visco, and F. van Wijland, How Far from Equilibrium is Active Matter? *Phys. Rev. Lett.* **117**, 038103 (2016).
- [109] D. Chaudhuri, Entropy production by active particles: Coupling of odd and even functions of velocity, *Phys. Rev. E* **94**, 032603 (2016).
- [110] P. Pietzonka and U. Seifert, Entropy production of active particles and for particles in active baths, *J. Phys. A* **51**, 01LT01 (2018).
- [111] S. Shankar and M. C. Marchetti, Hidden entropy production and work fluctuations in an ideal active gas, *Phys. Rev. E* **98**, 020604(R) (2018).
- [112] N. Bleistein and R. A. Handelsman, *Asymptotic Expansions of Integrals* (Courier Corporation, New York, 1975).
- [113] G. Nemes, *Asymptotic Expansions for Integrals*, Master’s thesis, Loránd Eötvös University, 2012.
- [114] D. Luposchinsky and H. Hinrichsen, Entropy production in continuous phase space systems, *J. Stat. Phys.* **153**, 828 (2013).
- [115] H. Kleinert, *Path Integrals in Quantum Mechanics, Statistics, and Polymer Physics* (World Scientific, Singapore, 1995), Vol. 400 .
- [116] H. Risken, *The Fokker-Planck Equation* (Springer, Berlin, Heidelberg, 1996), pp. 63–95.
- [117] Y. Fily, A. Baskaran, and M. F. Hagan, Dynamics of self-propelled particles under strong confinement, *Soft Matter* **10**, 5609 (2014).
- [118] H. H. Wensink and H. Löwen, Aggregation of self-propelled colloidal rods near confining walls, *Phys. Rev. E* **78**, 031409 (2008).
- [119] R. Nosrati, P. J. Graham, Q. Liu, and D. Sinton, Predominance of sperm motion in corners, *Sci. Rep.* **6**, 26669 (2016).
- [120] J. Shin, A. G. Cherstvy, W. K. Kim, and V. Zaburdaev, Elasticity-based polymer sorting in active fluids: A brownian dynamics study, *Phys. Chem. Chem. Phys.* **19**, 18338 (2017).
- [121] J. A. McLennan, Statistical mechanics of the steady state, *Phys. Rev.* **115**, 1405 (1959).
- [122] T. S. Komatsu and N. Nakagawa, Expression for the Stationary Distribution in Nonequilibrium Steady States, *Phys. Rev. Lett.* **100**, 030601 (2008).
- [123] C. Maes and K. Netočný, Rigorous meaning of mclennan ensembles, *J. Math. Phys.* **51**, 015219 (2010).
- [124] D. A. Sivak and G. E. Crooks, Near-Equilibrium Measurements of Nonequilibrium Free Energy, *Phys. Rev. Lett.* **108**, 150601 (2012).
- [125] N. Nakagawa and S.-i. Sasa, Work relations for time-dependent states, *Phys. Rev. E* **87**, 022109 (2013).

## Underdamped active Brownian heat engine

Viktor Holubec<sup>1,2,\*</sup> and Rahul Marathe<sup>3,†</sup><sup>1</sup>*Institut für Theoretische Physik, Universität Leipzig, Postfach 100 920, D-04009 Leipzig, Germany*<sup>2</sup>*Charles University, Faculty of Mathematics and Physics, Department of Macromolecular Physics, V Holešovičkách 2, CZ-180 00 Prague, Czech Republic*<sup>3</sup>*Department of Physics, Indian Institute of Technology, Delhi, New Delhi 110016, India* (Received 14 October 2020; accepted 23 November 2020; published 10 December 2020)

Active Brownian engines rectify energy from reservoirs composed of self-propelling nonequilibrium molecules into work. We consider a class of such engines based on an underdamped Brownian particle trapped in a power-law potential. The energy they transform has thermodynamic properties of heat only if the nonequilibrium reservoir can be assigned a suitable effective temperature consistent with the second law and thus yielding an upper bound on the engine efficiency. The effective temperature exists if the total force exerted on the particle by the bath is not correlated with the particle position. In general, this occurs if the noise autocorrelation function and the friction kernel are proportional as in the fluctuation-dissipation theorem. But even if the proportionality is broken, the effective temperature can be defined in restricted, fine-tuned, parameter regimes, as we demonstrate on a specific example with harmonic potential.

DOI: [10.1103/PhysRevE.102.060101](https://doi.org/10.1103/PhysRevE.102.060101)

## I. INTRODUCTION

The surging field of active matter has recently attracted significant attention from researchers in various fields [1–7]. It studies the behavior of self-propelling agents ranging from suspensions of micrometer-sized Janus particles and bacteria to flocks of birds [8–10]. Besides focusing on dynamics of these systems [11–13], several studies attempted to put the active matter on firm thermodynamic footing [14–16]. The main aim of these efforts, which include investigation of proper definitions of the entropy production and derivation of corresponding fluctuation theorems for active matter [17–19], is to develop a consistent generalization of successful theoretical framework of stochastic thermodynamics [20,21] to active matter systems [22].

One of the most intriguing questions in this respect is when the energy extracted from active baths can be termed as heat [23]—the problem which never arises for equilibrium heat reservoirs providing only heat. While it is straightforward to identify the extracted energy as work for various ratchets [24–26] rectifying the directed active motion of the nonequilibrium constituents of the bath, it is not that simple for cyclically operating machines [23,27–34]. With respect to the latter, the extracted energy can be unambiguously termed as heat only if there exists an equivalent setup with an equilibrium bath and the same average energy currents as the active engine [23]. If such an equilibrium mapping exists, the nonequilibrium setups can outperform the equilibrium ones [15,35–45] only by achieving unnaturally high effective temperature differences (and thus efficiencies), allowed by the

lack of thermalization in the active bath. For example, the recently realized engine with a bath containing living bacteria was reported to operate at effective temperatures far beyond the boiling point of water while the background fluid was still at the room temperature [27]. If such a mapping does not exist, the active engines can easily break the second law limitations, with the most striking example being the cyclic energy extraction from a single bath reported in Ref. [32]. Such machines should thus be termed as a (lossy) work-to-work converters.

The most general setting where the equilibrium mapping generically exists are heat engines with working medium described by Hamiltonians of the form  $\mathcal{H} = k(t)h(\mathbf{x}, \mathbf{p})$ , where  $k(t)$  is an externally controlled parameter (e.g., playing the

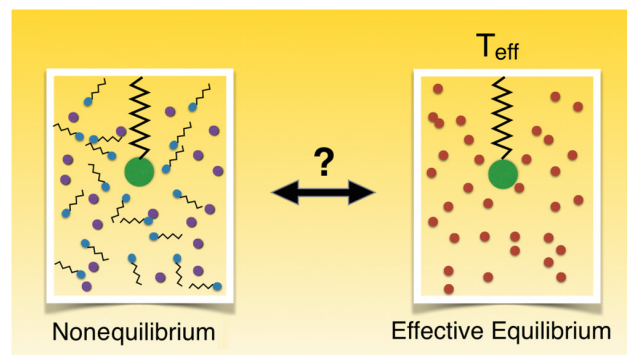


FIG. 1. Our setup. Left: An externally controlled working medium of an active Brownian engine (green particle attached to a spring) extracts energy from a nonequilibrium bath. Right: This energy can be termed as heat only if the energy fluxes in the engine can be realized using an equilibrium bath at an effective temperature  $T_{\text{eff}}$ .

\* viktor.holubec@mff.cuni.cz

† maratherahul@physics.iitd.ac.in

role of the inverse volume in macroscopic heat engines) and  $h$  is an arbitrary function increasing sufficiently fast with the absolute position  $|\mathbf{x}|$  and velocity  $|\mathbf{v}|$  [23].

A typical situation considered in this work is depicted in Fig. 1. We show when the equilibrium mapping exists for cyclic machines with the working medium described by the underdamped Hamiltonian

$$\mathcal{H}(x, v, t) = \mathcal{V}(x, t) + \frac{1}{2}mv^2, \quad (1)$$

where  $m$  is a constant mass and  $\mathcal{V}(x, t) = \frac{1}{n}k(t)x^n$ ,  $n = 2, 4, \dots$ , denotes a confining potential with periodically modulated stiffness  $k(t)$ . Existence of the equilibrium mapping in this case is generally not guaranteed, and we have to consider an explicit model. Inspired by recent experiments on Brownian heat engines [27,36,40], we assume that the working medium is an underdamped Brownian particle described by the system of Langevin equations

$$\dot{x} = v, \quad (2)$$

$$m\dot{v} = -kx^{n-1} + F + \eta. \quad (3)$$

The force  $-kx^{n-1} = -\partial_x \mathcal{V}$  corresponds to the potential and  $F + \eta$  is the total force applied on the particle by the bath. Its systematic component

$$F \equiv - \int_{-\infty}^t dt' \Gamma(t-t')v(t') \quad (4)$$

is a friction force with a friction kernel  $\Gamma(t)$ . And the additive noise with zero mean,  $\eta$ , denotes its stochastic component. We do not assume that the friction and noise fulfill the (second) fluctuation-dissipation (FDT) theorem [46–48]

$$k_B T \Gamma(|t|) = \langle \eta(t)\eta(0) \rangle \quad (5)$$

and thus the bath can be out of equilibrium. The dynamics (2)–(3) might, for example, describe an underdamped active Brownian particle, and thus, from now on, we term the considered engine the underdamped active Brownian engine (UABE).

The energy fluxes into the UABE can be identified from the change of the average internal energy of the working medium [20]

$$\frac{d}{dt} \langle \mathcal{H} \rangle = \frac{d}{dt} \left( \frac{1}{n} k \sigma_x + \frac{1}{2} m \sigma_v \right) = \dot{W} + \dot{Q}, \quad (6)$$

where the average is taken over realizations of the noise  $\eta$  and  $\sigma_x \equiv \langle x^n \rangle$  and  $\sigma_v \equiv \langle v^2 \rangle$ . The energy per unit time flowing into the system due to the external driving,

$$\dot{W} = \frac{1}{n} \dot{k} \sigma_x, \quad (7)$$

is the input work flux. The rest of the total energy influx,

$$\dot{Q} = \frac{d}{dt} \langle \mathcal{H} \rangle - \dot{W} = \frac{1}{n} k \dot{\sigma}_x + \frac{1}{2} m \dot{\sigma}_v, \quad (8)$$

originates in the nonequilibrium bath. While we denote it using the standard symbol for heat flux, it has thermodynamic properties of heat described by the second law only if there exists the equilibrium mapping [23], i.e., if there is an engine with equilibrium bath with the same energy fluxes (7) and (8)

as the UABE. In such a case, the ratio of the output work and the input heat,

$$E \equiv \frac{W_{\text{out}}}{Q_{\text{in}}} \equiv \frac{- \int_0^{t_p} dt \dot{W}(t)}{\int_0^{t_p} dt \dot{Q}(t) \theta[\dot{Q}(t)]}, \quad (9)$$

measuring the efficiency  $E$  of the UABE with period  $t_p$ , is bounded by the standard second-law bound corresponding to the given cycle realized by the stiffness  $k(t)$  and the effective temperature  $T_{\text{eff}}(t)$ . The Heaviside theta function  $\theta$  in the definition of the input heat  $Q_{\text{in}}$  ensures that the integral evaluates only the average heat flowing into the system [ $\dot{Q}(t) > 0$ ] [23].

In this paper, we focus on the upper bounds on thermodynamic efficiency of UABEs and thus we study existence of the equilibrium mapping in the quasistatic limit, where the control parameter  $k$  changes on timescales much longer than the system relaxation time, the frictional losses are minimal, and the total effective entropy  $\Delta S_{\text{tot}} \equiv - \int_0^{t_p} dt \dot{Q}(t)/T_{\text{eff}}(t)$  produced per cycle vanishes [23]. Then the efficiency (9) can be evaluated using the standard equilibrium thermodynamics. Namely, if the effective temperature and the stiffness are changed in such a way that the resulting cycle is composed of two branches with a constant effective temperature and two adiabatic branches when  $\dot{Q}(t) = 0$ , the efficiency of the cycle is given by the Carnot efficiency  $E_C = 1 - T_{\text{eff}}^-/T_{\text{eff}}^+$ , where  $T_{\text{eff}}^\pm$  denote the smallest (–) and largest (+) values of  $T_{\text{eff}}$  during the cycle. For arbitrary different driving, the efficiency is smaller than  $E_C$ . And it is still given by the standard (equilibrium) formula corresponding to the given cycle (such as the Stirling cycle), if the effective temperature is substituted for the real temperature in these formulas. For a more detailed discussion, see Ref. [23].

For a quasistatic driving, the dynamics of the moments  $\sigma_x$  and  $\sigma_v$  in all equilibrium models is described by the combination of equipartition  $2\langle \mathcal{T} \rangle \equiv \sigma_v = 2T_{\text{eff}}$  and virial  $2\langle \mathcal{T} \rangle = \langle x \frac{\partial \mathcal{V}}{\partial x} \rangle$  theorems:

$$T_{\text{eff}} = \frac{1}{2} k \sigma_x = \frac{1}{2} \sigma_v, \quad (10)$$

where  $T_{\text{eff}}$  denotes the temperature of the corresponding equilibrium baths. Above and in the rest of the paper, we use units in which  $k_B = 1$  and  $m = 1$ . If the noise in Eq. (3) can be described by an effective temperature allowing the above described standard equilibrium thermodynamic analysis of the engine efficiency, it must be given by Eq. (10). In the rest of this paper, we study when such an effective temperature exists.

## II. GENERAL RESULTS

Let us now find the general condition which must be fulfilled in the dynamics of the UABE so that the quasistatic equilibrium mapping (10) exists (for an example of an equilibrium mapping for nonquasistatic protocols, we refer to Ref. [23]). Multiplying Eq. (2) by  $v$ , Eq. (3) by  $x$ , taking the averages, and summing the results, we find that the moments  $\sigma_x$  and  $\sigma_v$  obey the equation

$$\frac{d}{dt} \langle xv \rangle = -k \sigma_x + \sigma_v + \langle (F + \eta)x \rangle. \quad (11)$$

In the quasistatic regime, we can neglect the time-derivative on the LHS obtaining the virial theorem

$$\langle \mathcal{T} \rangle = -\frac{1}{2} \langle F_{\text{tot}} x \rangle \quad (12)$$

for the whole system with the total force  $F_{\text{tot}} = -\partial \mathcal{V} / \partial x + F + \eta$  applied to the particle. The requirement (10) implies that the equilibrium mapping exists if the contribution from the bath to the virial vanishes,

$$\langle (F + \eta)x \rangle = 0. \quad (13)$$

This condition is quite reasonable since it means that the force exerted on the particle by the bath and the particle position are uncorrelated, i.e., that the bath is homogeneous. For an active bath, the condition (13) might be broken if the particle interacts strongly with the bath. For example, baths composed of active particles get polarized close to walls potentially leading to nonzero  $\langle (F + \eta)x \rangle$ .

The condition (13) is naturally fulfilled when the noise autocorrelation function and the friction kernel are proportional as in the FDT (5). Even though this situation might look trivial since it is mathematically equivalent to settings with equilibrium bath, it still represents an important class of nonequilibrium situations if  $T$  in (5) would be given by some effective temperature  $T_{\text{eff}}$ . As an example, consider a system connected to two independent equilibrium reservoirs,  $i = 1, 2$ , described by friction forces  $F_i = -\gamma_i v$  and white noises  $\eta_i = \sqrt{2D_i} \gamma_i \xi_i$ ,  $\langle \xi_i(t) \xi_j(t') \rangle = \delta_{ij} \delta(t - t')$ , with  $D_i = T_i / \gamma_i$ , as dictated by Einstein relation following from the FDT (5). Then the total friction force would be  $F = -(\gamma_1 + \gamma_2)v$  and the total noise  $\eta = \sqrt{2} \sqrt{D_1 \gamma_1^2 + D_2 \gamma_2^2} \xi$ ,  $\langle \xi(t) \xi(t') \rangle = \delta(t - t')$ , would be described by the friction kernel  $\Gamma(t) = (\gamma_1 + \gamma_2) \delta(t)$  and noise autocorrelation function  $\langle \eta(t) \eta(t') \rangle = 2(D_1 \gamma_1^2 + D_2 \gamma_2^2) \delta(t - t')$ . Even though the total noise and the friction obey the FDT (5) with the effective temperature  $T_{\text{eff}} = (D_1 \gamma_1^2 + D_2 \gamma_2^2) / (\gamma_1 + \gamma_2)$ , the system mediate a nonzero heat flux between the two baths and thus is out of equilibrium whenever  $T_1 \neq T_2$ .

Besides these equilibrium-like cases, there are also situations when the noise autocorrelation and friction kernel are not proportional but can be fine-tuned in such a way that Eq. (13) holds and thus the effective temperature can be defined. Also in these special parameter regimes, the effective temperature (10) can be used to derive upper bounds on efficiency of the corresponding machines. Thus they can serve as solid points for checking general solutions to these systems.

### III. SPECIFIC MODELS

To provide explicit analytical results, we now set  $n = 2$  in the general discussion of the previous section and thus we resort to the harmonic potential  $\mathcal{V}(x, t) = \frac{1}{2} k(t) x^2$ . In such a case, it was shown in detail in Ref. [23] how to calculate the effective temperature for one of the key toy models of the active matter, the active Brownian particle model [49]. We start by showing that in the underdamped version of this model, the effective temperature in general does not exist.

#### A. Underdamped active Brownian particle

The underdamped active Brownian particle (or, equivalently, underdamped active Ornstein-Uhlenbeck process) is described by the Langevin equations

$$\dot{x} = v, \quad (14)$$

$$\dot{v} = -\gamma v - kx + \eta, \quad (15)$$

where  $\gamma$  is a friction coefficient, and the stationary zero-mean noise  $\eta$  is exponentially correlated:

$$\langle \eta(t) \eta(t') \rangle = \frac{1}{2} u^2 \exp(-D_r |t - t'|). \quad (16)$$

Here  $u$  denotes the swimming velocity of the active particle and  $1/D_r$  is its orientation decorrelation time. Since the friction kernel is in this case given by  $-\gamma \Gamma(t) = -\gamma \delta(t)$ , the noise and the friction are clearly not related by the FDT (5).

The friction force in this case reads  $F = -\gamma v$ . Inserting it into the condition (13) for the existence of the equilibrium mapping, we find

$$-\gamma \langle vx \rangle + \langle \eta x \rangle = 0. \quad (17)$$

From Eq. (14) it follows that  $\langle vx \rangle = \dot{\langle x \rangle} / 2 = 0$  and thus the condition for existence of the equilibrium mapping in this case reads  $\langle \eta x \rangle = 0$ . In order to calculate this correlation, it is advantageous to rewrite the system (14) and (15) using the matrix notation as

$$\dot{\mathbf{X}} = \mathcal{M} \mathbf{X} + \eta \mathbf{e}, \quad (18)$$

where  $\mathbf{X} = (v, x)^T$ ,  $\mathbf{e} = (1, 0)^T$  are column vectors and

$$\mathcal{M} = \begin{pmatrix} -\gamma & -k \\ 1 & 0 \end{pmatrix}. \quad (19)$$

The long-time solution to Eq. (18) is given by

$$\mathbf{X}(t) = \int_{-\infty}^t dt' \mathcal{U}(t - t') \eta(t') \mathbf{e} \quad (20)$$

with  $\mathcal{U}(t) = \exp[\mathcal{M}(t - t')]$ . Multiplying it by  $\eta(t)$  and taking the average, we find

$$\begin{aligned} \langle \mathbf{X} \eta \rangle &= \lim_{t \rightarrow \infty} \int_0^t dt' \mathcal{U}(t - t') \langle \eta(t) \eta(t') \rangle \mathbf{e} \\ &= \frac{u^2}{2} \lim_{t \rightarrow \infty} \int_0^t dt' e^{(\mathcal{M} - D_r)(t - t')} \mathbf{e} = \frac{-1}{\mathcal{M} - D_r} \frac{u^2 \mathbf{e}}{2}, \end{aligned} \quad (21)$$

where the scalar terms are meant to be multiplied by the  $2 \times 2$  unit matrix. The second element of this vector is the desired correlation between the noise and position. It reads

$$\langle x \eta \rangle = \frac{u^2}{2 [k + D_r (D_r + \gamma)]}, \quad (22)$$

and it in general does not vanish. The effective temperature for the underdamped Brownian particle thus exists only in various special limiting situations when  $\langle x \eta \rangle \rightarrow 0$ . The simplest example is the limit of vanishing swimming velocity of the particle,  $u$ . In such a case, however, the system dynamics becomes deterministic. Other noteworthy limiting situations where  $\langle x \eta \rangle \rightarrow 0$  are the limits of infinitely strong potential,

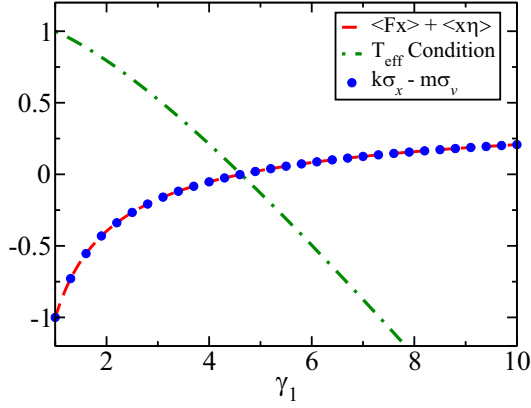


FIG. 2. The averages  $\langle(F + \eta)x\rangle$  and  $k\sigma_x - m\sigma_v$ , and the condition (34) as functions of  $\gamma_1$ . Parameters used are  $\gamma_0 = 4.0$ ,  $\alpha_0 = 0.5$ ,  $\alpha_1 = 1.0$ ,  $m = 1.0$ ,  $k = 0.1$ , and  $D = 1.0$ .

$k \gg u^2$  and  $k \gg D_r(D_r + \gamma)$ , and of infinitely fast orientation decorrelation,  $D_r^2 \gg k$  and  $D_r^2 \gg u^2$ , when  $\eta$  is a white noise.

### B. Exponential friction kernel and arbitrary noise

Let us now consider a slightly more general setup when the friction kernel is exponential,

$$\Gamma(t) = \gamma_0 \exp(-\gamma_1 |t|), \quad (23)$$

and the noise autocorrelation function is of the form

$$\langle\eta(t)\eta(t')\rangle = \alpha_0 \exp(-\alpha_1 |t - t'|) + 2D\delta(t - t'). \quad (24)$$

Also in this case, the proportionality in the FDT (5) is broken. This situation is quite common in experimental setups with active particles suspended in an aqueous medium [27], where the working substance experiences random forces from both the active and the solvent molecules.

Dynamics of this system is described by Eqs. (2) and (3) with the friction kernel (23). However, for analytical considerations, it is useful to note that the resulting friction force

$$F(t) = -\gamma_0 \int_{-\infty}^t dt' \exp[-\gamma_1(t - t')]v(t') \quad (25)$$

allows one to rewrite the dynamical equations in the form

$$\dot{x} = v, \quad (26)$$

$$\dot{v} = -kv + F + \eta, \quad (27)$$

$$\dot{F} = -\gamma_1 F - \gamma_0 v, \quad (28)$$

where we assume that all the variables are initialized at  $t' = -\infty$ .

To find when the condition (13) for the existence of the equilibrium mapping is fulfilled, we have to calculate the averages  $\langle Fx\rangle$  and  $\langle x\eta\rangle$ . To this end, we again rewrite the system (26)–(28) in the matrix form (18). This time with  $\mathbf{X} = (F, v, x)^T$ ,  $\mathbf{e} = (0, 1, 0)^T$ , and

$$\mathcal{M} = \begin{pmatrix} -\gamma_1 & -\gamma_0 & 0 \\ 1 & 0 & -k \\ 0 & 1 & 0 \end{pmatrix}. \quad (29)$$

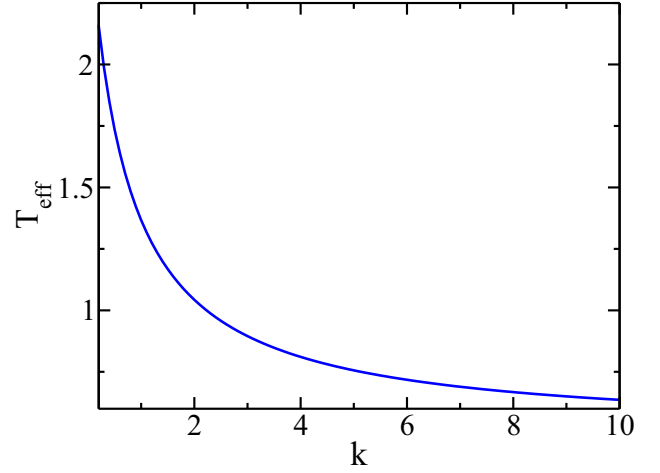


FIG. 3. The effective temperature (35) as a function of the trap strength  $k$  for parameters  $\gamma_0 = 4.0$ ,  $\gamma_1 = 0.5$ ,  $\alpha_0 = 2.0$ ,  $\alpha_1 = 0.1$ ,  $m = 1.0$ , and  $D = 1.0$ .

The long-time solution to the resulting matrix equation is again of the form (20) with  $\mathcal{U}(t) = \exp[\mathcal{M}(t - t')]$ . To calculate the averages involved in the condition (13), we evaluate the whole covariance matrix  $\langle\mathbf{X}(t)\mathbf{X}^T(t)\rangle$ . Taking its time-derivative derivative, expressing  $\dot{\mathbf{X}}$  from Eq. (18), and setting  $d\langle\mathbf{X}(t)\mathbf{X}^T(t)\rangle/dt = 0$  which follows from the steady-state assumption, we get

$$\mathcal{M}\langle\mathbf{X}(t)\mathbf{X}^T(t)\rangle + \langle\mathbf{X}(t)\mathbf{X}^T(t)\rangle\mathcal{M}^T + \langle\xi\mathbf{X}^T\rangle + \langle\mathbf{X}\xi^T\rangle = 0, \quad (30)$$

where we introduced the shorthand  $\xi = \eta\mathbf{e}$ . The correlation matrix  $\langle\xi\mathbf{X}^T\rangle^T = \langle\mathbf{X}\xi^T\rangle$  can be obtained by multiplying the formal solution (20) by  $\xi$  and taking the average, similarly as in Eq. (21).

The result is [50]

$$\begin{aligned} \langle\mathbf{X}\xi^T\rangle &= \lim_{t \rightarrow \infty} \int_{-\infty}^t dt' \exp[\mathcal{M}(t - t')] \langle\eta(t)\eta(t')\rangle \mathbf{e}\mathbf{e}^T \\ &= \left( \frac{\alpha_0}{\alpha_1 - \mathcal{M}} + D \right) \mathbf{e}\mathbf{e}^T. \end{aligned} \quad (31)$$

Solution of this linear system is straightforward but the full result is rather lengthy. For our purposes, it is enough to explicitly evaluate only the correlation  $\langle Fx\rangle$ , involved in the condition (13), and the variances  $\sigma_x$  and  $\sigma_v$ , defining the effective temperature (10), if the former condition is fulfilled. Evaluating also the correlation  $\langle\eta x\rangle$  given by the matrix element 3-2 of the matrix  $\langle\mathbf{X}\xi^T\rangle$  in Eq. (31), we find

$$\langle(F + \eta)x\rangle = k\sigma_x - \sigma_v = \frac{(\gamma_1 - \alpha_1)(\gamma_1 + \alpha_1)c - D}{\gamma_1}, \quad (32)$$

where

$$c = \frac{\alpha_0}{\gamma_0\alpha_1 + (\alpha_1 + \gamma_1)(k + \alpha_1^2)}. \quad (33)$$

The equilibrium mapping and the equipartition (10) in this system is thus fulfilled whenever

$$D - (\gamma_1 - \alpha_1)(\gamma_1 + \alpha_1)c = 0, \quad (34)$$

which can occur even when the FDT is broken. To check validity of our calculations, we also evaluated the terms  $\langle(F + \eta)x\rangle$  and  $k\sigma_x - \sigma_v$  in Eq. (32) using an independent numerical method. In Fig. 2 we plot the obtained result and also the explicit condition (34) as functions of the decay rate  $\gamma_1$ . The three lines indeed intersect at zero at the same point, proving our analytical calculations.

Setting,  $D = (\gamma_1 - \alpha_1)(\gamma_1 + \alpha_1)c$ , we find that the effective temperature  $T_{\text{eff}} = k\langle x^2 \rangle / k_B = \langle v^2 \rangle / k_B$  for the present model reads

$$T_{\text{eff}} = \frac{\alpha_0}{\gamma_0} \left[ \frac{\gamma_0 \gamma_1 + (\alpha_1 + \gamma_1)(k + \gamma_1^2)}{\gamma_0 \alpha_1 + (\alpha_1 + \gamma_1)(k + \alpha_1^2)} \right]. \quad (35)$$

Noteworthy, as shown in Fig. 3, this effective temperature depends on the trap stiffness  $k$ . Since similar dependence has been found also for the overdamped setting [23], this result seems quite general. It is thus puzzling why recent experimental work [27], considering a similar setup as ours, reported effective temperature independent of the trap stiffness.

#### IV. CONCLUSION

The existence of mapping to an equivalent equilibrium setup is an important condition for calling machines operating in contact with a nonequilibrium bath as heat engines. Only if such a mapping exists, the efficiency of these machines obeys standard second law limitations. We found that a broad class of machines based on an underdamped Brownian particle trapped in a power-law potential, which we called as UABEs, allows for the equilibrium mapping if the total force exerted

on the particle by the bath and the particle position are not correlated.

Validity of this condition can be easily checked in experiments. It is always fulfilled if the friction kernel and noise autocorrelation function in the Langevin equation for dynamics of the particle are proportional so that an effective variant of the second fluctuation-dissipation theorem holds. Besides this somewhat trivial case, equilibrium mapping also exists for special parameter regimes in systems where the proportionality is broken. These regimes may serve as solid points where the maximum efficiency of UABEs is known.

We have studied an explicit example, where such a special parameter regime exists and calculated the effective temperature of the corresponding equilibrium bath. This effective temperature depends on the strength of the applied potential, which, we believe, is a general feature of effective temperatures.

Our findings can guide theoretical analysis and serve as a sanity check of results measured for systems in contact with nonequilibrium reservoirs. As an outlook, it would be interesting to study extensions of our model to finite time regimes, where power delivered by the engines does not vanish. Furthermore, it would be interesting to extend our results to higher dimensions and more complicated potentials.

#### ACKNOWLEDGMENTS

V.H. gratefully acknowledges discussions with Klaus Kroy and support by the Humboldt foundation and by the Czech Science Foundation (Project No. 20-02955J). R.M. acknowledges financial support from Department of Science and Technology (DST), India under the MATRICS grant (Project No. MTR/2020/000349).

- 
- [1] J. Prost, F. Jülicher, and J.-F. Joanny, Active gel physics, *Nat. Phys.* **11**, 111 (2015).
  - [2] G. Popkin, The physics of life, *Nature News* **529**, 16 (2016).
  - [3] D. Needleman and Z. Dogic, Active matter at the interface between materials science and cell biology, *Nat. Rev. Mater.* **2**, 17048 (2017).
  - [4] A. Doostmohammadi, J. Ignés-Mullol, J. M. Yeomans, and F. Sagués, Active nematics, *Nat. Commun.* **9**, 3246 (2018).
  - [5] O. Feinerman, I. Pinkoviezky, A. Gelblum, E. Fonio, and N. S. Gov, The physics of cooperative transport in groups of ants, *Nat. Phys.* **14**, 683 (2018).
  - [6] X. Trepap and E. Sahai, Mesoscale physical principles of collective cell organization, *Nat. Phys.* **14**, 671 (2018).
  - [7] W. Xi, T. B. Saw, D. Delacour, C. T. Lim, and B. Ladoux, Material approaches to active tissue mechanics, *Nat. Rev. Mater.* **4**, 23 (2019).
  - [8] S. Ramaswamy, The mechanics and statistics of active matter, *Annu. Rev. Condens. Matter Phys.* **1**, 323 (2010).
  - [9] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe, Active particles in complex and crowded environments, *Rev. Mod. Phys.* **88**, 045006 (2016).
  - [10] M. Das, C. F. Schmidt, and M. Murrell, Introduction to active matter, *Soft Matter* **16**, 7185 (2020).
  - [11] X.-L. Wu and A. Libchaber, Particle Diffusion in a Quasi-Two-Dimensional Bacterial Bath, *Phys. Rev. Lett.* **84**, 3017 (2000).
  - [12] J. Tailleur and M. E. Cates, Statistical Mechanics of Interacting Run-and-Tumble Bacteria, *Phys. Rev. Lett.* **100**, 218103 (2008).
  - [13] M. E. Cates and J. Tailleur, Motility-induced phase separation, *Annu. Rev. Condens. Matter Phys.* **6**, 219 (2015).
  - [14] C. Maggi, M. Paoluzzi, N. Pellicciotta, A. Lepore, L. Angelani, and R. Di Leonardo, Generalized Energy Equipartition in Harmonic Oscillators Driven by Active Baths, *Phys. Rev. Lett.* **113**, 238303 (2014).
  - [15] E. Fodor, C. Nardini, M. E. Cates, J. Tailleur, P. Visco, and F. van Wijland, How Far From Equilibrium is Active Matter? *Phys. Rev. Lett.* **117**, 038103 (2016).
  - [16] G. Szamel, Stochastic thermodynamics for self-propelled particles, *Phys. Rev. E* **100**, 050603(R) (2019).
  - [17] D. Mandal, K. Klymko, and M. R. DeWeese, Entropy Production and Fluctuation Theorems for Active Matter, *Phys. Rev. Lett.* **119**, 258001 (2017).
  - [18] S. Chaki and R. Chakrabarti, Entropy production and work fluctuation relations for a single particle in active bath, *Physica A* **511**, 302 (2018).

- [19] L. Dabelow, S. Bo, and R. Eichhorn, Irreversibility in Active Matter Systems: Fluctuation Theorem and Mutual Information, *Phys. Rev. X* **9**, 021009 (2019).
- [20] K. Sekimoto, Langevin equation and thermodynamics, *Prog. Theor. Phys. Suppl.* **130**, 17 (1998).
- [21] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, *Rep. Prog. Phys.* **75**, 126001 (2012).
- [22] T. Speck, Stochastic thermodynamics for active matter, *Europhys. Lett.* **114**, 30006 (2016).
- [23] V. Holubec, S. Steffenoni, G. Falasco, and K. Kroy, Active Brownian heat engines, *Phys. Rev. Research* **2**, 043262 (2020).
- [24] R. Di Leonardo, L. Angelani, D. Dell'Arciprete, G. Ruocco, V. Iebba, S. Schippa, M. P. Conte, F. Mecarini, F. De Angelis, and E. Di Fabrizio, Bacterial ratchet motors, *Proc. Natl. Acad. Sci. USA* **107**, 9541 (2010).
- [25] C. O. Reichhardt and C. Reichhardt, Ratchet effects in active matter systems, *Annu. Rev. Condens. Matter Phys.* **8**, 51 (2017).
- [26] P. Pietzonka, E. Fodor, C. Lohrmann, M. E. Cates, and U. Seifert, Autonomous Engines Driven by Active Matter: Energetics and Design Principles, *Phys. Rev. X* **9**, 041032 (2019).
- [27] S. Krishnamurthy, S. Ghosh, D. Chatterji, R. Ganapathy, and A. K. Sood, A micrometre-sized heat engine operating between bacterial reservoirs, *Nat. Phys.* **12**, 1134 (2016).
- [28] R. Zakine, A. Solon, T. Gingrich, and F. Van Wijland, Stochastic Stirling engine operating in contact with active baths, *Entropy* **19**(5), 193 (2017).
- [29] A. Saha, R. Marathe, P. S. Pal, and A. M. Jayannavar, Stochastic heat engine powered by active dissipation, *J. Stat. Mech.: Theory Exp.* (2018) 113203.
- [30] D. Martin, C. Nardini, M. E. Cates, and É. Fodor, Extracting maximum power from active colloidal heat engines, *Europhys. Lett.* **121**, 60005 (2018).
- [31] A. Saha and R. Marathe, Stochastic work extraction in a colloidal heat engine in the presence of colored noise, *J. Stat. Mech.: Theory Exp.* (2019) 094012.
- [32] T. Ekeh, M. E. Cates, and E. Fodor, Thermodynamic cycles with active matter, *Phys. Rev. E* **102**, 010101(R) (2020).
- [33] A. Kumari, P. S. Pal, A. Saha, and S. Lahiri, Stochastic heat engine using an active particle, *Phys. Rev. E* **101**, 032109 (2020).
- [34] A. Kumari and S. Lahiri, Microscopic thermal machines using run-and-tumble particles, [arXiv:2008.01367](https://arxiv.org/abs/2008.01367) (2020).
- [35] T. Schmiedl and U. Seifert, Efficiency at maximum power: An analytically solvable model for stochastic heat engines, *Europhys. Lett.* **81**, 20003 (2007).
- [36] V. Blickle and C. Bechinger, Realization of a micrometre-sized stochastic heat engine, *Nat. Phys.* **8**, 143 (2012).
- [37] S. Rana, P. S. Pal, A. Saha, and A. M. Jayannavar, Single-particle stochastic heat engine, *Phys. Rev. E* **90**, 042146 (2014).
- [38] V. Holubec, An exactly solvable model of a stochastic heat engine: Optimization of power, power fluctuations and efficiency, *J. Stat. Mech.: Theory Exp.* (2014) P05022.
- [39] I. A. Martínez, E. Roldán, L. Dinis, D. Petrov, and R. A. Rica, Adiabatic Processes Realized with a Trapped Brownian Particle, *Phys. Rev. Lett.* **114**, 120601 (2015).
- [40] I. A. Martínez and E. Roldán, Brownian Carnot engine, *Nat. Phys.* **12**, 67 (2016).
- [41] I. A. Martínez, É. Roldán, L. Dinis, and R. A. Rica, Colloidal heat engines: A review, *Soft Matter* **13**, 22 (2017).
- [42] V. Holubec and A. Ryabov, Diverging, but negligible power at Carnot efficiency: Theory and experiment, *Phys. Rev. E* **96**, 062107 (2017).
- [43] V. Holubec and A. Ryabov, Cycling Tames Power Fluctuations Near Pptimum Efficiency, *Phys. Rev. Lett.* **121**, 120601 (2018).
- [44] P. Pietzonka and U. Seifert, Universal Trade-Off Between Power, Efficiency, and Constancy in Steady-State Heat Engines, *Phys. Rev. Lett.* **120**, 190602 (2018).
- [45] D. Wexler, N. Gov, K. O. Rasmussen, and G. Bel, Dynamics and escape of active particles in a harmonic trap, *Phys. Rev. Res.* **2**, 013003 (2020).
- [46] R. Kubo, The fluctuation-dissipation theorem, *Rep. Prog. Phys.* **29**, 255 (1966).
- [47] R. Kubo, M. Toda, and N. Hashitsume, *Statistical Physics II: Nonequilibrium Statistical Mechanics*, Springer Series in Solid-State Sciences, Vol. 31 (Springer, Berlin, 2012).
- [48] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, Oxford, 2001).
- [49] M. E. Cates, Diffusive transport without detailed balance in motile bacteria: Does microbiology need statistical physics?, *Rep. Prog. Phys.* **75**, 042601 (2012).
- [50] L. Caprini and U. M. B. Marconi, Inertial self-propelled particles, [arXiv:2009.14032](https://arxiv.org/abs/2009.14032) [cond-mat.stat-mech] (2020).



## Maximum efficiency of low-dissipation heat engines at arbitrary power

This content has been downloaded from IOPscience. Please scroll down to see the full text.

J. Stat. Mech. (2016) 073204

(<http://iopscience.iop.org/1742-5468/2016/7/073204>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 195.113.29.6

This content was downloaded on 11/07/2016 at 09:26

Please note that [terms and conditions apply](#).

# Maximum efficiency of low-dissipation heat engines at arbitrary power

Viktor Holubec and Artem Ryabov

Charles University in Prague, Faculty of Mathematics and Physics,  
Department of Macromolecular Physics, V Holešovičkách 2, CZ-180 00 Praha,  
Czech Republic

E-mail: [viktor.holubec@gmail.com](mailto:viktor.holubec@gmail.com) and [rjabov.a@gmail.com](mailto:rjabov.a@gmail.com)

Received 29 March 2016, revised 4 May 2016

Accepted for publication 3 June 2016

Published 11 July 2016



Online at [stacks.iop.org/JSTAT/2016/073204](http://stacks.iop.org/JSTAT/2016/073204)

[doi:10.1088/1742-5468/2016/07/073204](https://doi.org/10.1088/1742-5468/2016/07/073204)

**Abstract.** We investigate maximum efficiency at a given power for low-dissipation heat engines. Close to maximum power, the maximum gain in efficiency scales as a square root of relative loss in power and this scaling is universal for a broad class of systems. For low-dissipation engines, we calculate the maximum gain in efficiency for an arbitrary fixed power. We show that engines working close to maximum power can operate at considerably larger efficiency compared to the efficiency at maximum power. Furthermore, we introduce universal bounds on maximum efficiency at a given power for low-dissipation heat engines. These bounds represent direct generalization of the bounds on efficiency at maximum power obtained by Esposito *et al* (2010 *Phys. Rev. Lett.* **105** 150603). We derive the bounds analytically in the regime close to maximum power and for small power values. For the intermediate regime we present strong numerical evidence for the validity of the bounds.

**Keywords:** exact results, transport processes / heat transfer, molecular motors

**Contents**

<b>1. Introduction</b>	<b>2</b>
<b>2. Model</b>	<b>3</b>
<b>3. Efficiency at maximum power</b>	<b>4</b>
<b>4. Efficiency near maximum power</b>	<b>5</b>
<b>5. Maximum gain in efficiency for a fixed loss of power</b>	<b>6</b>
5.1. Exact numerical results. . . . .	6
5.2. Approximate analytical results . . . . .	7
5.3. Maximum $\delta_\eta$ and $\eta$ as functions of the parameter $A$ . . . . .	10
<b>6. Bounds on maximum gain in efficiency</b>	<b>11</b>
<b>7. Conclusions and outlooks</b>	<b>13</b>
<b>References</b>	<b>13</b>

J. Stat. Mech. (2016) 073204

**1. Introduction**

Since the dawn of heat engines people have struggled to optimize their performance [1]. One of the first theoretical results in the field was due to Carnot [2] and Clausius [3]: the maximum efficiency attainable by any heat engine operating between the temperatures  $T_h$  and  $T_c$ ,  $T_h > T_c$ , is given by the Carnot efficiency  $\eta_C = 1 - T_c/T_h$ . In order to attain  $\eta_C$ , the engine must work reversibly (infinitely slowly) and thus its output power is vanishingly small. Optimization of the power of *irreversible* Carnot cycles working under finite-time conditions was pioneered by Yvon [4], Novikov [5], Chambadal [6] and later by Curzon and Ahlborn [7]. Although the result obtained for the efficiency at maximum power (EMP),  $\eta_{CA} = 1 - \sqrt{T_c/T_h}$ , is not universal, nor does it represent a bound on the EMP [8–10], its close agreement with EMP for several model systems [11–24] ignited the search for universalities in performance of heat engines.

Up to the second order in  $\eta_C$  the EMP,  $\eta^*$ , is controlled by the symmetries of the underlying dynamics [25–28]. Further universalities were obtained for heat engines working in the low-dissipation regime [14], [29–33], where the work dissipated during the isothermal branches of the Carnot cycle grows in inverse proportion to the duration of these branches. In this regime, a general expression for the EMP has been published [14] and, subsequently, Esposito *et al* derived the bounds  $\eta_C/2 \leq \eta^* \leq \eta_C/(2 - \eta_C)$  on the EMP [29]. All these results were confirmed within the framework of irreversible thermodynamics [34, 35].

Recently, increased attention has been given to the optimization of heat engines which do not work at maximum power [24], [36–39]. Such studies are important for engineering practice, where not only powerful, but also economical devices should be

developed. Indeed, it had already been highlighted [40–42] that actual thermal plants and heat engines should not work at the maximum power  $P^*$ , where the corresponding efficiency  $\eta^*$  can be relatively small, but rather in a regime with slightly smaller power  $P$  and considerably larger efficiency  $\eta$ .

In the present paper, we introduce universal bounds on maximum efficiency at a given power for low-dissipation heat engines (LDHEs)

$$\frac{\eta_C}{2} \left(1 + \sqrt{-\delta_P}\right) \leq \eta \leq \eta_C \frac{1 + \sqrt{-\delta_P}}{2 - (1 - \sqrt{-\delta_P})\eta_C}, \quad (1)$$

where

$$\delta_P = (P - P^*)/P^*. \quad (2)$$

We derive these bounds analytically for small  $\delta_P$  and for  $\delta_P$  close to 1. For the intermediate regime we present strong numerical evidence that the bounds are valid for any  $\delta_P$ . The inequalities (1) represent direct generalization of the bounds on EMP obtained for  $\delta_P = 0$  by Esposito *et al* [29]. In the leading order in  $\eta_C$ , the left and the right bound coincide and the resulting maximum efficiency,  $\eta = \eta_C(1 + \sqrt{-\delta_P})/2$ , equates to that obtained using linear response theory in the strong coupling limit [43]. Both the bounds also coincide for vanishing power ( $\delta_P \rightarrow -1$ ), when they equal  $\eta_C$ , thus verifying Carnot's results.

We also study the maximum relative gain in efficiency

$$\delta_\eta = (\eta - \eta^*)/\eta^* \quad (3)$$

with respect to EMP of LDHEs [32, 37], [44–46] for arbitrary fixed power and show that it scales in the leading order of the relative loss of power  $\delta_P$  as

$$\delta_\eta \propto \sqrt{-\delta_P}. \quad (4)$$

The slope of the gain in efficiency  $\delta_\eta$  diverges at  $\delta_P = 0$  and hence LDHEs working close to maximum power operate at considerably larger efficiency than  $\eta^*$ . We show that both the diverging slope and the scaling (4) are direct consequences of the fact that the maximum power corresponds to  $\delta_P = 0$  and that these findings are valid for a broad class of systems (see the text below equation (24)). Indeed, the scaling (4) was already obtained in recent studies on quantum thermoelectric devices [38, 39], for a stochastic heat engine based on an underdamped particle diffusing in a parabolic potential [24] and also using linear response theory [43].

## 2. Model

We consider a non-equilibrium Carnot cycle composed of two isotherms and two adiabats working in the low-dissipation regime [14, 29, 37, 47–55]. During the hot (cold) isotherm the system is coupled to the reservoir at temperature  $T_h$  ( $T_c$ ). Let  $t_h$  ( $t_c$ ) denotes the duration of the hot (cold) isotherm. In the low-dissipation regime, it is assumed that the system relaxation time is short compared to  $t_h$  and  $t_c$ . Then it is possible to assume that the entropy production per cycle equals

$$\Delta S_{\text{tot}} = \frac{A_h}{t_h T_h} + \frac{A_c}{t_c T_c}, \quad (5)$$

where  $A_h, A_c$  are positive parameters. This means that the engine reaches reversible operation when duration of the cycle becomes very large ( $t_{h,c} \rightarrow \infty$ ). Another usual assumption, that we also adopt here, is that the duration of the adiabatic branches is short compared to  $t_h + t_c$  and thus the cycle duration can be well approximated by  $t_p = t_h + t_c$ .

The heat absorbed by the system during the hot isotherm,  $Q_h$ , and the heat delivered to the cold reservoir during the cold isotherm,  $Q_c$ , are given by

$$Q_h = T_h \Delta S - A_h/t_h, \quad (6)$$

$$Q_c = T_c \Delta S + A_c/t_c, \quad (7)$$

where  $\Delta S$  denotes the change of the system entropy during the hot isotherm. The positive parameters  $A_h$  and  $A_c$  thus measure the degree of irreversibility of the individual isotherms. They are given by the details of the dynamics of the system and can be easily measured [49].

We express  $t_h$  and  $t_c$  using the duration of the cycle,  $t_p$ , and its redistribution among the two isotherms,  $\alpha$ , as  $t_h = \alpha t_p$  and  $t_c = (1 - \alpha)t_p$ . Then the engine output power and its efficiency can be written as [37, 51]

$$P = \frac{Q_h - Q_c}{t_p} = \frac{(T_h - T_c)\Delta S}{t_p} - \frac{(1 - \alpha)A_h + \alpha A_c}{t_p^2 \alpha(1 - \alpha)}, \quad (8)$$

$$\eta = \frac{Q_h - Q_c}{Q_h} = \frac{\eta_C}{1 + T_c \Delta S_{\text{tot}}/(P t_p)}. \quad (9)$$

In general, interchanging the reservoirs at the ends of the isothermal branches brings the system out of equilibrium. During the subsequent relaxation, an additional positive contribution to the entropy production (5) arises, which may not vanish in the limit  $t_{h,c} \rightarrow \infty$ . This unavoidably results in a decrease of the efficiency at a fixed power (9). By considering cycles with a reversible limit, we assume this dissipation to be negligible. While this assumption is reasonable for large systems, it might require a delicate control of system dynamics in the case of microscopic heat engines [14], [29, 37, 47–49, 51], [56].

### 3. Efficiency at maximum power

Maximizing the power (8) as a function of  $t_p$  and  $\alpha$  gives [14] (values at maximum power are denoted by  $\star$ )

$$\alpha^* = \frac{A_h - \sqrt{A_h A_c}}{A_h - A_c}, \quad (10)$$

$$t_p^* = \frac{2}{T_h \eta_C \Delta S} (\sqrt{A_h} + \sqrt{A_c})^2, \quad (11)$$

$$P^* = \frac{1}{4} \left( \frac{T_h \eta_C \Delta S}{\sqrt{A_h} + \sqrt{A_c}} \right)^2, \quad (12)$$

$$\eta^* = \frac{\eta_C (1 + \sqrt{A_c/A_h})}{2(1 + \sqrt{A_c/A_h}) - \eta_C}. \quad (13)$$

Note that the EMP (13) does not depend on the individual parameters  $A_h$  and  $A_c$ , but only on their ratio  $A_h/A_c$ .

#### 4. Efficiency near maximum power

The operational point of maximum power (10)–(13) can be used to define the coordinate transformation

$$\tau = \frac{t_p}{t_p^*} - 1, \quad \tau \in [-1, \infty), \quad (14)$$

$$a = \frac{\alpha}{\alpha^*} - 1, \quad a \in [-1, \frac{1}{\alpha^*} - 1], \quad (15)$$

which decreases the number of parameters contained in the formulas (8)–(9) for power and efficiency by 2 [37] and thus makes the maximization of efficiency for a given power much easier. The point of maximum power corresponds in these coordinates to the origin, i.e.  $\tau = a = 0$ . The parameter  $\tau$  is larger than zero whenever  $t_p > t_p^*$  and similarly  $a > 0$  if  $\alpha > \alpha^*$ .

The relative loss of power (2) and the relative change in efficiency (3) in these new coordinates read

$$\delta_P = \frac{a^2}{(1+a)(a-A)(1+\tau)^2} - \left( \frac{\tau}{1+\tau} \right)^2, \quad (16)$$

$$\delta_\eta = -1 + \frac{2(1+A) - \eta_C}{a-A} \frac{a(2a-A+1) - A + 2(1+a)(a-A)\tau}{2(1+\tau)(1+a)(1+A) - \eta_C}, \quad (17)$$

where

$$A = \sqrt{A_c/A_h}. \quad (18)$$

Let us here stress that by using the symbol  $\delta$  in the notation we do not mean that the deviations from the maximum power measured by the functions (16) and (17) must be small.

The power exhibits a maximum at  $\tau = a = 0$  and thus  $\delta_P$  for small  $\tau$  and  $a$  varies very slowly. On the other hand, the efficiency can change much more rapidly and thus, for suitable parameters, the loss of power is much smaller than the gain in efficiency [24, 37–42]. We will now find the formula which describes this gain.

## 5. Maximum gain in efficiency for a fixed loss of power

For a fixed  $\delta_P$ , the parameters  $a$  and  $\tau$  are related due to equation (16) as

$$\tau = \frac{-\delta_P}{1 + \delta_P} \pm \frac{\sqrt{-a^2 - [(1 + a)A - a]\delta_P}}{\sqrt{(1 + a)(A - a)(1 + \delta_P)}}. \quad (19)$$

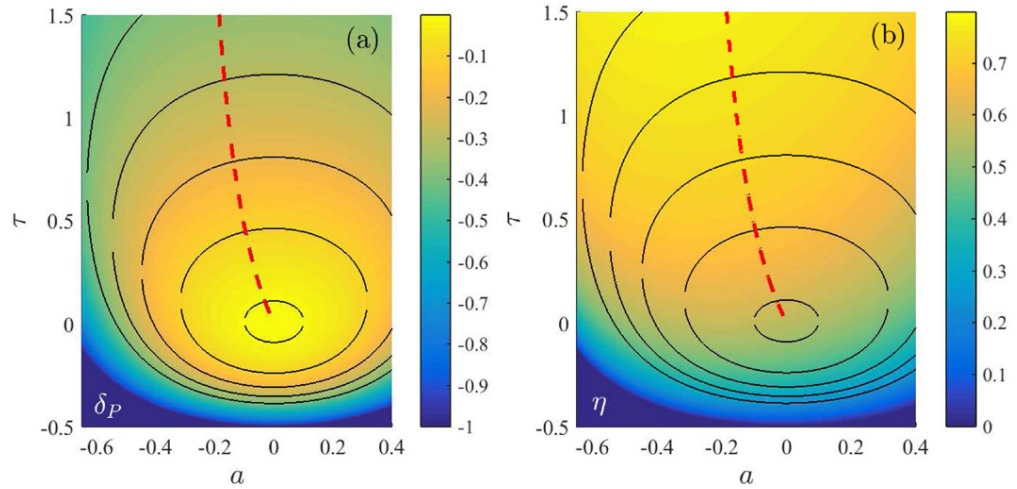
For five values of  $\delta_P$  and for  $A = 1$ , the curves defined by equation (19) are depicted by black lines in figure 1. Upper (lower) lines correspond to the upper (lower) sign on the right-hand side of equation (19). They mark the combinations of coordinates  $a, \tau$  which yield the same value of power. The power is larger the closer the curves are to the origin  $a = \tau = 0$ . In this figure, we also show the relative loss of power  $\delta_P$  (panel (a)) and the efficiency  $\eta$  (panel (b)) as functions of the parameters  $a$  and  $\tau$ .

### 5.1. Exact numerical results

Due to the algebraic complexity of equations (16) and (17), the analytical derivation of the maximum  $\delta_\eta$  for a given  $\delta_P$  is in general intractable and we perform this calculation only numerically. Examples of the results of such optimization are demonstrated in figures 1 and 2. In figure 1 the dashed red line denotes the values of  $a$  and  $\tau$  corresponding to the maximum efficiency (and thus also  $\delta_\eta$ ) for given values of  $\delta_P$ , which is the parameter of this curve. The dashed line intersects the upper solid black curves. Hence the optimal values of efficiency are obtained for the upper sign in equation (19). In figure 2 we show the maximum gain in efficiency for a fixed power (panel (a)), the maximum efficiency for a fixed power (panel (b)) and the corresponding optimal<sup>1</sup> values of the parameters  $\tau$  (panel (c)) and  $a$  (panel (d)) as functions of the relative loss of power  $\delta_P$ . The panels (a) and (b) in figure 2 demonstrate that the gain in efficiency when working close to maximum power ( $\delta_P = 0$ ) is indeed significant. The panels (c) and (d) in figure 2 and the red dashed line in figure 1 reveal that optimal values of  $\tau$  are always positive ( $t_p > t_p^*$ ) and the optimal values of  $a$  are always negative ( $\alpha < \alpha^*$ ). This result is quite intuitive.

For a fixed power, the efficiency (9) increases if the average entropy production rate during the cycle,  $\Delta S_{\text{tot}}/t_p$ , decreases. For a fixed  $\alpha$ , the total entropy production per cycle  $\Delta S_{\text{tot}}$  (5) decreases with increasing  $t_p$  (and thus  $\Delta S_{\text{tot}}/t_p$  decreases even faster). Physically, this is because slower processes are more reversible. On the other hand, for a fixed  $t_p$ ,  $\Delta S_{\text{tot}}$  is smaller for  $\alpha < \alpha^*$  than for  $\alpha \geq \alpha^*$ . To see this, let us expand  $\Delta S_{\text{tot}}$  into a Taylor series around the point of maximum power  $\alpha^*$ :

<sup>1</sup> In the following we will use the word ‘optimal’ as a synonym for ‘corresponding to the maximum efficiency for a fixed power’.



**Figure 1.** The relative loss of power (16) (panel (a)) and the efficiency  $\eta = \eta^*(\delta_\eta + 1)$  (panel (b)) as functions of the parameters  $a$  and  $\tau$ . In both panels, the upper black lines were calculated from equation (19) with the upper sign. Similarly, for calculation of the lower black lines we have used equation (19) with the lower sign. These lines connect the points with the same value of  $\delta_P$ . The red dashed lines correspond to the maximum efficiency for a fixed power, which is the parameter of this curve. In both panels we set  $A = 1$ ,  $\eta_C = 0.875$ .

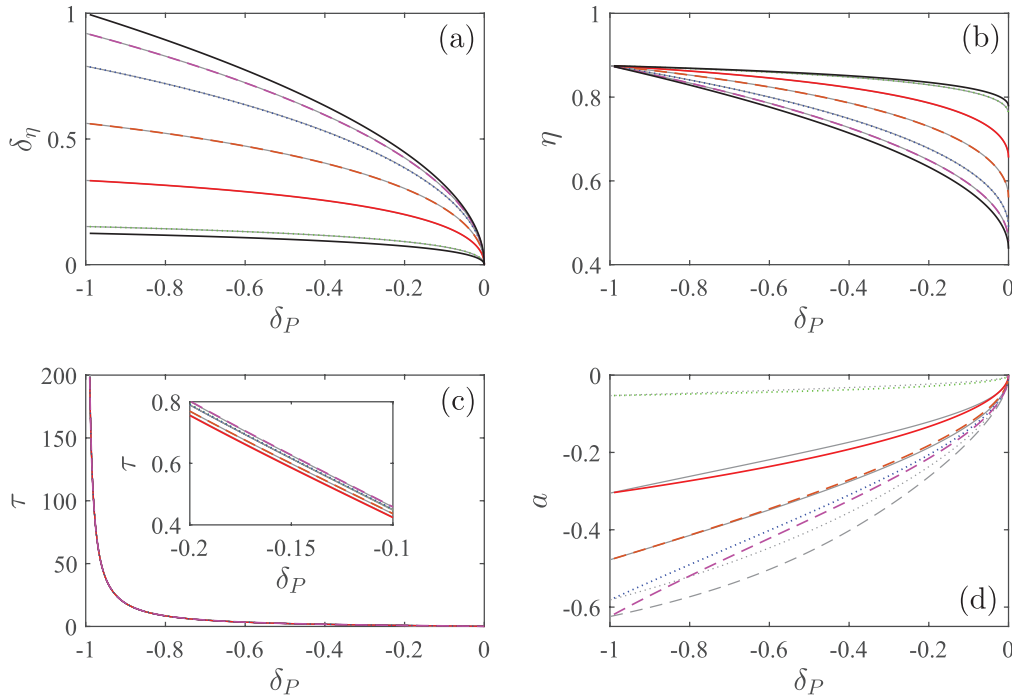
$\Delta S_{\text{tot}} = \Delta S_{\text{tot}}^* + (\Delta S'_{\text{tot}})^*(\alpha - \alpha^*) + O[(\alpha - \alpha^*)^2]$ , where  $\Delta S_{\text{tot}}^*$  is the total entropy production at maximum power and  $(\Delta S'_{\text{tot}})^* = (1 + A)^2 A_h (T_h - T_c) / (T_c T_h t_p) > 0$ . We thus have  $\Delta S_{\text{tot}} - \Delta S_{\text{tot}}^* < 0$  whenever  $\alpha < \alpha^*$ . Although this proof is valid only up to the linear order in  $\alpha - \alpha^*$ , the result holds generally. In order to get further physical intuition it is helpful to consider the symmetric situation  $A = 1$ . In such case for smaller  $\alpha$  (larger  $1 - \alpha$ ) more work is dissipated in a bath with large temperature  $T_h$ , where the same amount of dissipated work creates less entropy than it would generate in a cold bath (entropy produced in a bath is equal to (energy delivered to the bath)/(bath temperature)). For a fixed power,  $\alpha$  and  $t_p$  ( $a$  and  $\tau$ ) can not change independently and thus a compromise between an increased  $\tau$  and a decreased  $\alpha$  which verify equation (19) is chosen. In this compromise, depicted in figure 1 by the dashed red line, increasing  $\tau$  makes the cycle more reversible and decreasing  $\alpha$  causes more energy to be dissipated in the hot bath, which generates less entropy.

## 5.2. Approximate analytical results

Although the full analytical optimization of efficiency for a fixed power is in general beyond our reach, there are two limiting regimes when the analytical calculation is possible. The resulting simple analytical formulas (20), (23) and (24) yield the bounds (32) and (33) on maximum  $\delta_\eta$  and  $\eta$  for a fixed power. Comparison with exact numerics reveals that these bounds are valid also outside the two limiting regimes (see figure 2 and explanations below).

First, for  $\delta_P \rightarrow -1$  ( $P \rightarrow 0$ ), equation (19) yields  $\tau \rightarrow \infty$  ( $t_p \rightarrow \infty$ ). Then, we get from equation (17) that





**Figure 2.** Panel (a): maximum relative gain in efficiency  $\delta_\eta$  (3) as a function of the relative loss of power  $\delta_P$  (2) for  $\eta_C = 0.875$  and five values of the parameter  $A$ :  $A = \sqrt{0.001}$  (green dotted line),  $A = \sqrt{0.1}$  (red solid line),  $A = 1$  (black orange line),  $A = \sqrt{10}$  (blue dotted line) and  $A = \sqrt{100}$  (magenta dashed line). The dashed (full) black lines depict the lower (upper) bound on the maximum relative gain in efficiency (32). The corresponding efficiencies together with the bounds (33) are shown in panel (b). In panels (c) and (d) we show the corresponding optimal values of the parameters  $\tau$  and  $a$ . The colored lines are calculated using exact numerical optimization of efficiency for a fixed power. The thin gray lines are calculated using analytical optimization based on the approximate formula equation (21). Although the optimal values of the parameter  $a$  calculated in this approximation sometimes differ from the correct values (panel (d)), the resulting optimal parameter  $\tau$  (panel (c)) and, more importantly, the optimal efficiency (panel (b)) and the optimal gain in efficiency (panel (a)) are predicted so well that the individual gray and colored curves overlap.

$$\delta_\eta = \frac{\eta_C}{\eta^*} - 1 + O\left(\frac{1}{\tau}\right), \quad (20)$$

and thus  $\eta = \eta_C + O(1/\tau)$ . This means that, for large  $\tau$ , the efficiency depends on the parameter  $a$  only via the term proportional to  $1/\tau$ , which becomes negligible close to  $\delta_P \rightarrow -1$ .

The second analytically tractable situation, which is more important for practical reasons, is the case of small  $\delta_P$ . Close to the maximum power the parameters  $a$  and  $\tau$  are also small. This means that, instead of performing the derivation for a small  $\delta_P$ , one can perform it for a small  $a$ . Data from the exact numerical optimization shown in figure 2 demonstrate that the absolute values of the optimal parameter  $a$  are always either

small (for moderately small  $|\delta_P|$ ) or close to  $-1$  (for  $\delta_P \rightarrow -1$ , when  $\tau \gg 1$ ). This means that the optimization using the small  $a$  approximation may be close to the exact solution even for relatively large  $\delta_P$ . This is because the effect of  $a$  on the optimal efficiency is either well captured by the approximation (for moderate  $\delta_P$ ) or negligible ( $\delta_P \rightarrow -1$ , when  $\tau \gg 1$ ). Up to the second order in  $a$ , it follows from equation (19) that

$$\tau = \pm \frac{\sqrt{-\delta_P}}{1 \mp \sqrt{-\delta_P}} \mp \frac{a^2}{2A\sqrt{-\delta_P}}, \quad (21)$$

where the upper signs correspond to the upper sign in equation (19) and thus lead to the maximum efficiency for the fixed power. The rest of the calculation can be performed without any other approximation. The final results are depicted in figure 2 by the gray lines, which in the panels (a) (maximum  $\delta_\eta$  for a fixed power), (b) (maximum  $\eta$  for a fixed power) and (c) (the corresponding optimal parameter  $\tau$ ) overlap with the data obtained using exact numerical optimization. The only difference between the approximate analytical solution and the numerical results can be observed in panel (d), where we show the optimal values of the parameter  $a$ . Thus, as we have conjectured above, the results based on the approximate equation (21) if no other approximations are made describe very well the exact optimized values of  $\eta$  and  $\delta_\eta$ . Nevertheless, the formulas are quite involved and thus we will write in the rest of this section only the results up to the leading order in  $\delta_P$ .

Substituting  $\tau$  with the upper signs from equation (21) into equation (17) for  $\delta_\eta$ , taking the derivative with respect to  $a$  and solving the resulting equation  $d\delta_\eta/da = 0$  for  $a$ , we obtain in the leading order in  $\delta_P$

$$a = -\frac{1}{2} \frac{A \eta_C}{1 + A - \eta_C} \sqrt{-\delta_P} \leq 0. \quad (22)$$

The resulting optimal parameter  $a$  is thus negative ( $a < a^*$ ) in accordance with the discussion at the end of section 5.1. Inserting  $\tau$  from equation (21) and  $a$  from equation (22) into the formula (17) for  $\delta_\eta$ , we get up to the leading order in  $\delta_P$

$$\delta_\eta = f(A, \eta_C) \sqrt{-\delta_P}, \quad (23)$$

where

$$f(A, \eta_C) = \frac{1}{4} \left[ \frac{(A+1)A}{A - \eta_C + 1} + \frac{4(A+1)(A+2)}{-2A + \eta_C - 2} + A + 8 \right].$$

The corresponding maximum efficiency  $\eta = (\delta_\eta + 1)\eta^*$  reads

$$\eta = \eta^*(A, \eta_C) \left[ 1 + f(A, \eta_C) \sqrt{-\delta_P} \right]. \quad (24)$$

Equations (23)–(24) constitutes our first main result. The maximum relative gain in efficiency (23) and the maximum efficiency itself (24) are non-analytical functions of  $\delta_P$  with a diverging slope at  $\delta_P = 0$ , which clearly indicates that the gain in efficiency when working near maximum power is much larger than the loss of power, in accord with the findings of [37]. Both the diverging slope with  $\delta_P \rightarrow 0$  and the scaling  $\sqrt{-\delta_P}$  are direct

consequences of the fact that the power has maximum at  $\delta_P = 0$  and thus represent generic features of the maximum efficiency close to maximum power.

In order to understand how these results arise, assume that both power,  $P$ , and the corresponding maximum efficiency,  $\eta$ , are parametrized by the parameter vector  $\mathbf{x}$ , in the present setting  $\mathbf{x} = \{t_p, \alpha\}$ , and that they are analytical functions of all these parameters. Taylor expansions of  $P$  and  $\eta$  around the point of maximum power  $\mathbf{x} = \mathbf{x}^*$  (where  $P' = \nabla P|_{\mathbf{x}=\mathbf{x}^*} = 0$  denotes the gradient and  $P'' = \nabla^2 P|_{\mathbf{x}=\mathbf{x}^*} < 0$  the negative definite Hessian matrix evaluated at the point of maximum power),  $P = P^* + (\mathbf{x} - \mathbf{x}^*)^T P'' (\mathbf{x} - \mathbf{x}^*)/2$  and  $\eta = \eta^* + (\mathbf{x} - \mathbf{x}^*)^T \eta'$ , lead to  $\delta_\eta \propto \sqrt{-\delta_P}$ . The scaling (23) is thus universal whenever the Taylor expansions of power and efficiency used are valid. Indeed, the dependence (23) has already been obtained for quantum thermoelectric devices [38, 39], for a stochastic heat engine based on an underdamped particle diffusing in a parabolic potential [24] and also using linear response theory [43]. The next two terms in equation (23) are of the order  $\delta_P$  and  $(-\delta_P)^{3/2}$  and can be also accurately predicted if one departs from the approximate formula (21) for  $\tau$ .

### 5.3. Maximum $\delta_\eta$ and $\eta$ as functions of the parameter $A$

The optimal relative gain in efficiency (23) is an *increasing* function of  $A$  as can be proven by showing positivity of the derivative

$$\frac{\partial f(A, \eta_C)}{\partial A} = \eta_C \frac{g(A, \eta_C)}{4(1 + A - \eta_C)^2(-2 - 2A + \eta_C)^2} A. \quad (25)$$

The sign of this function is determined by the sign of the function  $g(A, \eta_C) = 8(1 + A)^2 - 2(1 + A)(7 + A)\eta_C + 5\eta_C^2 + \eta_C^3$ . The derivative of this expression with respect to  $A$ ,  $16(1 + A) - 4(1 + A)\eta_C - 12\eta_C$ , is positive for all  $\eta_C$ ,  $0 < \eta_C < 1$ . The function  $g(A, \eta_C)$  is thus an increasing function of  $A$  and hence we can demonstrate the positivity of  $g(A, \eta_C)$  by showing that  $g(0, \eta_C) > 0$ . To this end, we obtain  $g(0, \eta_C) = 8 - 14\eta_C + 5\eta_C^2 + \eta_C^3$ . This expression decreases with  $\eta_C$  and thus the function  $g(A, \eta_C)$  fulfills the inequality  $g(A, \eta_C) > g(0, 1) = 0$ , which proves positivity of the derivative  $\partial f(A, \eta_C)/\partial A$ . Therefore, for small values of  $\delta_P$ , the maximum relative gain in efficiency for a given power increases with  $A$ . Furthermore, the same can be observed from the full solution for the optimal  $\delta_\eta$ , using the exact numerical optimization and also using the analytical results for  $\delta_P \rightarrow -1$  (20). This means that the limit  $A \rightarrow 0$  of  $\delta_\eta$  yields a lower bound on the relative gain in efficiency for arbitrary  $\delta_P$ . The upper bound on  $\delta_\eta$  is then obtained in the limit  $A \rightarrow \infty$ .

Similar argumentation can be used also for the optimal efficiency at a given power. For small values of  $\delta_P$  the optimal  $\eta$  is a monotonously *decreasing* function of  $A$  as can be shown using the equation (24). According to this equation the derivative of the maximum efficiency with respect to  $A$  is given by

$$\frac{\partial \eta}{\partial A} = \frac{\partial \eta^*}{\partial A} + \left( \frac{\partial \eta^*}{\partial A} f + \eta^* \frac{\partial f}{\partial A} \right) \sqrt{-\delta_P}. \quad (26)$$

As can be seen directly from its definition (13),  $\eta^*$  decreases with  $A$ , i.e.  $\partial \eta^*/\partial A < 0$ . This means that  $\partial \eta/\partial A < 0$  and the maximum efficiency decreases with  $A$  for small values of

$\delta_P$ . Furthermore, the same behavior, but now for arbitrary  $\delta_P$ , is obtained using the full solution for the optimal efficiency and also using exact numerical optimization. Finally, for  $\delta_P \rightarrow -1$  the maximum efficiency equals to  $\eta_C$  for any  $A$ . The lower bound for the optimal efficiency is thus obtained for  $A \rightarrow \infty$  and corresponds to the upper bound for the optimal  $\delta_P$ . Similarly, the upper bound for the optimal  $\eta$  is obtained for  $A \rightarrow 0$  and corresponds to the lower bound for the optimal  $\delta_P$ .

Physically, this behavior can be understood if one realizes how the quantity  $A_c$  contributes to the total entropy production  $\Delta S_{\text{tot}}$ . At the end of section 5.1 we argued that, by decreasing  $\alpha$ , a larger part of the total dissipated work is delivered to the hot bath, where it produces a smaller amount of entropy than it would produce in the cold reservoir. For a fixed power, the parameters  $A_c$  and  $A_h$  are no longer independent since they satisfy equation (8). By changing these parameters one redistributes the total amount of dissipated work between the two reservoirs in the same way as by changing the parameter  $\alpha$ . If the parameter  $A_c$  is small, larger amount of work is dissipated in the hot bath and, similarly, for a large  $A_c$  more work is dissipated in the cold bath. This means that the efficiency decreases (entropy production increases) with increasing  $A = \sqrt{A_c/A_h}$  and vice versa.

Does this also imply that larger  $A$  leads to a larger gain in efficiency  $\delta_\eta = \eta/\eta^* - 1$ ? As we have argued above, both the EMP  $\eta^*$  and the maximum efficiency at a given power  $\eta$  are decreasing functions of  $A$ . The fact that  $\delta_\eta$  is an increasing function of  $A$  means that the decrease of  $\eta$  with  $A$  must be slower than the decrease of  $\eta^*$ . The EMP  $\eta^*$  is completely determined by the condition that the corresponding power is maximal (parameters  $a$  and  $\tau$  are fixed) and thus it has no freedom to be further optimized when the parameter  $A$  changes. On the other hand, the maximum efficiency  $\eta$  at a given power possesses such freedom and thus one may expect, that it will decay with increasing  $A$  slower than  $\eta^*$ . Our results for behavior of optimal  $\eta$  and  $\delta_\eta$  with  $A$  verify this conjecture (see figure 1). Now, let us focus on deriving the bounds for maximum gain in efficiency for a given power and for the maximum efficiency for a given power.

## 6. Bounds on maximum gain in efficiency

As we have discussed in section 5.3, the upper bound on  $\delta_\eta$  follows by taking the limit  $A \rightarrow \infty$  in equations (16)–(17). The result is

$$\lim_{A \rightarrow \infty} \delta_P = -\left(\frac{\tau}{1 + \tau}\right)^2, \quad (27)$$

$$\lim_{A \rightarrow \infty} \delta_\eta = \frac{\tau}{1 + \tau} = \sqrt{-\delta_P}. \quad (28)$$

The lower bound follows by taking the other total asymmetric limit  $A \rightarrow 0$ . Then  $\alpha^* = 1$  and thus  $a \in [-1, 0]$ . From equation (19) we get

$$\tau = \frac{-\delta_P}{1 + \delta_P} \pm \frac{\sqrt{a - \delta_P}}{\sqrt{1 + a(1 + \delta_P)}}, \quad (29)$$

where, for  $a \in [-1, 0]$ ,  $a - \delta_P > 0$  as can be shown directly from equation (16). Positive relative change in efficiency

$$\delta_\eta = \frac{2(1 - \eta_C) \left[ a - \delta_P + \sqrt{(1 + a)(a - \delta_P)} \right]}{2 \left[ a + 1 + \sqrt{(1 + a)(a - \delta_P)} \right] - (1 + \delta_P)\eta_C} \quad (30)$$

is obtained for the plus sign before the square root in equation (29). From  $a \in [-1, 0]$  and  $a - \delta_P > 0$  it follows that  $\partial\delta_\eta/\partial a > 0$  and thus  $\delta_\eta$  monotonously increases with  $a$ . This means that the maximum

$$\delta_\eta = \frac{2(1 - \eta_C)\sqrt{-\delta_P}}{2 - (1 - \sqrt{-\delta_P})\eta_C} \quad (31)$$

is obtained for the maximum possible value of  $a$ ,  $a = 0$ .

We have thus found that the maximum gain in efficiency at a given power obeys the inequalities

$$\frac{2(1 - \eta_C)\sqrt{-\delta_P}}{2 - (1 - \sqrt{-\delta_P})\eta_C} \leq \delta_\eta \leq \sqrt{-\delta_P}. \quad (32)$$

As we have discussed at the end of section 5.3, the upper bound (32) corresponds to the lower bound on maximum efficiency at a given power,  $\eta = (1 + \delta_\eta)\eta^*$ , and, similarly, the lower bound (32) yields the upper bound on  $\eta$ . For  $A \rightarrow \infty$ , we have  $\eta^* \rightarrow \eta_C/2$  and for  $A \rightarrow 0$ ,  $\eta^* \rightarrow \eta_C/(2 - \eta_C)$ . The bounds on efficiency thus read

$$\frac{\eta_C}{2} \left( 1 + \sqrt{-\delta_P} \right) \leq \eta \leq \eta_C \frac{1 + \sqrt{-\delta_P}}{2 - (1 - \sqrt{-\delta_P})\eta_C}. \quad (33)$$

The bounds (32)–(33) are our second main result. They represent direct generalizations of the bounds on EMP derived for  $\delta_P = 0$  by Esposito *et al* [29]. Note that for small temperature differences, i.e. up to the leading order in  $\eta_C$ , the lower and the upper bound on the maximum efficiency are equal and thus the maximum efficiency as a function of  $\delta_P$  is independent of the parameter  $A$ , which contains details about the system dynamics. It is given by

$$\eta = \frac{\eta_C}{2} \left( 1 + \sqrt{-\delta_P} \right). \quad (34)$$

The same formula for maximum efficiency has been recently obtained using linear response theory in the strong coupling limit [43].

In figure 2(a) we show the bounds (32) and in figure 2(b) we show the corresponding bounds on the maximum efficiency (33). From the figure, one can observe that the maximum efficiency interpolates between the EMP  $\eta^*$  (for  $\delta_P = 0$ ) and Carnot efficiency  $\eta_C$  (for  $\delta_P = -1$ ), which is, in accord with the bounds (33), reached irrespectively of the parameter  $A$ . Similar behavior of maximum efficiency was encountered for the underdamped particle diffusing in a parabolic potential [24].

## 7. Conclusions and outlooks

It is well known that real-world heat engines should not work at maximum power, but rather in a regime with slightly smaller power, but with considerably larger efficiency. For low-dissipation heat engines, we have introduced lower and upper bounds on the maximum efficiency at a given power (32) and the corresponding bounds on the maximum efficiency (33). We have also calculated the maximum relative gain in efficiency for an arbitrary fixed power. Close to maximum power, this gain scales as a square root from the relative loss of power  $\delta_P$  (23). This scaling is a direct consequence of the fact that power has a maximum at  $\delta_P = 0$  and thus it is universal for a broad class of systems. Indeed, the same scaling of maximum efficiency with the relative loss of power has been found recently for several models [24, 38, 39, 43]. Our results thus support the general statement about actual heat engines with quantitative arguments and reveal more practical limits on efficiency than the reversible one.

It would be interesting to investigate maximum gain in efficiency for a fixed power also for other models, such as endoreversible heat engines, or systems described by general Markov dynamics, i.e. by a Master equation, to see whether the behavior would be qualitatively the same as that obtained here and in the studies [24, 38, 39, 43]. Furthermore, one can ask if the functional form of the prefactor  $f$  in the formula for the gain in efficiency  $\delta_\eta = f\sqrt{-\delta_P} + O(\delta_P)$  is controlled by similar symmetries of the underlying dynamics as the EMP [25–28]. It would be also immensely interesting to find a heat engine where the square root scaling of the maximum gain in efficiency close to maximum power was not valid.

## References

- [1] Müller I 2007 *A History of Thermodynamics: the Doctrine of Energy and Entropy* (New York: Springer)
- [2] Carnot S 1978 *Réflexions sur la puissance motrice du feu (Académie Internationale d'histoire des Sciences. Collection des Travaux)* (Paris: J Vrin)
- [3] Clausius R 1856 X. On a modified form of the second fundamental theorem in the mechanical theory of heat *Phil. Mag. Ser. 4* **12** 81–98
- [4] Yvon J 1955 The Saclay reactor: two years experience on heat transfer by means of a compressed gas *Proc. Int. Conf. on Peaceful Uses of Atomic Energy (Geneva)* p 387
- [5] Novikov I I 1958 The efficiency of atomic power stations (a review) *J. Nucl. Energy* **7** 125–8
- [6] Chambadal P 1957 *Les centrales nucléaires* vol 321 (Paris: A Colin)
- [7] Curzon F L and Ahlborn B 1975 Efficiency of a Carnot engine at maximum power output *Am. J. Phys.* **43** 22–4
- [8] Hoffmann K H, Burzler J M and Schubert S 1997 Endoreversible thermodynamics *J. Non-Equilib. Thermodyn.* **22** 311–55
- [9] Berry R S, Kazakov V, Sieniutycz S, Szwast Z and Tsirlin A M 2000 *Thermodynamic Optimization of Finite-Time Processes* (New York: Wiley)
- [10] Salamon P, Nulton J D, Siragusa G, Andersen T R and Limon A 2001 Principles of control thermodynamics *Energy* **26** 307–19
- [11] De Vos A 1985 Efficiency of some heat engines at maximum-power conditions *Am. J. Phys.* **53** 570–3
- [12] Bejan A 1996 Entropy generation minimization: the new thermodynamics of finite-size devices and finite-time processes *J. Appl. Phys.* **79** 1191–218
- [13] Jiménez de Cisneros B and Calvo Hernández A 2007 Collective working regimes for coupled heat engines *Phys. Rev. Lett.* **98** 130602
- [14] Schmiedl T and Seifert U 2008 Efficiency at maximum power: an analytically solvable model for stochastic heat engines *Europhys. Lett.* **81** 20003

- [15] Izumida Y and Okuda K 2008 Molecular kinetic analysis of a finite-time carnot cycle *Europhys. Lett.* **83** 60003
- [16] Izumida Y and Okuda K 2009 Onsager coefficients of a finite-time Carnot cycle *Phys. Rev. E* **80** 021121
- [17] Allahverdyan A E, Johal R S and Mahler G 2008 Work extremum principle: structure and function of quantum heat engines *Phys. Rev. E* **77** 041118
- [18] Tu Z C 2008 Efficiency at maximum power of Feynman's ratchet as a heat engine *J. Phys. A: Math. Theor.* **41** 312003
- [19] Esposito M, Lindenberg K and Van den Broeck C 2009 Thermoelectric efficiency at maximum power in a quantum dot *Europhys. Lett.* **85** 60010
- [20] Rutten B, Esposito M and Cleuren B 2009 Reaching optimal efficiencies using nanosized photoelectric devices *Phys. Rev. B* **80** 235122
- [21] Esposito M, Kawai R, Lindenberg K and Van den Broeck C 2010 Quantum-dot Carnot engine at maximum power *Phys. Rev. E* **81** 041106
- [22] Zhou Y and Segal D 2010 Minimal model of a heat engine: information theory approach *Phys. Rev. E* **82** 011120
- [23] Tu Z-C 2012 Recent advance on the efficiency at maximum power of heat engines *Chin. Phys. B* **21** 020513
- [24] Dechant A, Kiesel N and Lutz E 2016 Underdamped stochastic heat engine at maximum efficiency arXiv:cond-mat/1602.00392
- [25] Esposito M, Lindenberg K and Van den Broeck C 2009 Universality of efficiency at maximum power *Phys. Rev. Lett.* **102** 130602
- [26] Izumida Y and Okuda K 2014 Work output and efficiency at maximum power of linear irreversible heat engines operating with a finite-sized heat source *Phys. Rev. Lett.* **112** 180603
- [27] Sheng S and Tu Z C 2015 Constitutive relation for nonlinear response and universality of efficiency at maximum power for tight-coupling heat engines *Phys. Rev. E* **91** 022136
- [28] Cleuren B, Rutten B and Van den Broeck C 2015 Universality of efficiency at maximum power *Eur. Phys. J. Spec. Top.* **224** 879–89
- [29] Esposito M, Kawai R, Lindenberg K and Van den Broeck C 2010 Efficiency at maximum power of low-dissipation Carnot engines *Phys. Rev. Lett.* **105** 150603
- [30] Sekimoto K and Sasa S I 1997 Complementarity relation for irreversible process derived from stochastic energetics *J. Phys. Soc. Japan* **66** 3326–8
- [31] Bonança M V S and Deffner S 2014 Optimal driving of isothermal processes close to equilibrium *J. Chem. Phys.* **140** 244119
- [32] de Tomas C, Roco J M M, Calvo Hernández A, Wang Y and Tu Z C 2013 Low-dissipation heat devices: unified trade-off optimization and bounds *Phys. Rev. E* **87** 012105
- [33] Muratore-Ginanneschi P and Schwieger K 2015 Efficient protocols for stirling heat engines at the micro-scale *Europhys. Lett.* **112** 20002
- [34] Izumida Y and Okuda K 2012 Efficiency at maximum power of minimally nonlinear irreversible heat engines *Europhys. Lett.* **97** 10004
- [35] Izumida Y, Okuda K, Calvo Hernández A and Roco J M M 2013 Coefficient of performance under optimized figure of merit in minimally nonlinear irreversible refrigerator *Europhys. Lett.* **101** 10005
- [36] Bauer M, Brandner K and Seifert U 2016 Optimal performance of periodically driven, stochastic heat engines under limited control *Phys. Rev. E* **93** 042112
- [37] Holubec V and Ryabov A 2015 Efficiency at and near maximum power of low-dissipation heat engines *Phys. Rev. E* **92** 052125
- [38] Whitney R S 2014 Most efficient quantum thermoelectric at finite power output *Phys. Rev. Lett.* **112** 130601
- [39] Whitney R S 2015 Finding the quantum thermoelectric with maximal efficiency and minimal entropy production at given power output *Phys. Rev. B* **91** 115425
- [40] Chen J, Yan Z, Lin G and Andresen B 2001 On the Curzon–Ahlborn efficiency and its connection with the efficiencies of real heat engines *Energy Convers. Manage.* **42** 173–81
- [41] De Vos A 1992 *Endoreversible Thermodynamics of Solar Energy Conversion* (Oxford: Oxford University Press)
- [42] Chen J 1994 The maximum power output and maximum efficiency of an irreversible Carnot heat engine *J. Phys. D: Appl. Phys.* **27** 1144
- [43] Ryabov A and Holubec V 2016 Maximum efficiency of steady-state heat engines at arbitrary power *Phys. Rev. E* **93** 050101
- [44] Long R and Liu W 2015 Unified trade-off optimization for general heat devices with nonisothermal processes *Phys. Rev. E* **91** 042127

- [45] Long R, Liu Z and Liu W 2014 Performance optimization of minimally nonlinear irreversible heat engines and refrigerators under a trade-off figure of merit *Phys. Rev. E* **89** 062119
- [46] Sheng S and Tu Z C 2013 Universality of energy conversion efficiency for optimal tight-coupling heat engines and refrigerators *J. Phys. A: Math. Theor.* **46** 402001
- [47] Zulkowski P R and DeWeese M R 2015 Optimal protocols for slowly driven quantum systems *Phys. Rev. E* **92** 032113
- [48] Zulkowski P R and DeWeese M R 2015 Optimal control of overdamped systems *Phys. Rev. E* **92** 032117
- [49] Martínez I A, Roldán É, Dinis L, Petrov D, Parrondo J M R and Rica R A 2015 Brownian Carnot engine *Nat. Phys.* **12** 67–70
- [50] Blickle V and Bechinger C 2011 Realization of a micrometre-sized stochastic heat engine *Nat. Phys.* **8** 143–6
- [51] Holubec V 2014 An exactly solvable model of a stochastic heat engine: optimization of power, power fluctuations and efficiency *J. Stat. Mech.* P05022
- [52] Rana S, Pal P S, Saha A and Jayannavar A M 2014 Single-particle stochastic heat engine *Phys. Rev. E* **90** 042146
- [53] Rana S, Pal P S, Saha A and Jayannavar A M 2015 Anomalous Brownian refrigerator *Physica A* **444** 783–98
- [54] Benjamin R and Kawai R 2008 Inertial effects in Büttiker–Landauer motor and refrigerator at the overdamped limit *Phys. Rev. E* **77** 051132
- [55] Tu Z C 2014 Stochastic heat engine with the consideration of inertial effects and shortcuts to adiabaticity *Phys. Rev. E* **89** 052148
- [56] Sato K, Sekimoto K, Hondou T and Takagi F 2002 Irreversibility resulting from contact with a heat bath caused by the finiteness of the system *Phys. Rev. E* **66** 016119



**Physically consistent numerical solver for time-dependent Fokker-Planck equations**Viktor Holubec,<sup>1,2,\*</sup> Klaus Kroy,<sup>1</sup> and Stefano Steffenoni<sup>1,3</sup><sup>1</sup>*Institut für Theoretische Physik, Universität Leipzig, Postfach 100 920, D-04009 Leipzig, Germany*<sup>2</sup>*Faculty of Mathematics and Physics, Department of Macromolecular Physics, Charles University, V Holešovičkách 2, CZ-180 00 Praha, Czech Republic*<sup>3</sup>*Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, D-04103 Leipzig, Germany*

(Received 5 July 2018; revised manuscript received 27 November 2018; published 11 March 2019)

We present a simple thermodynamically consistent method for solving time-dependent Fokker-Planck equations (FPE) for overdamped stochastic processes, also known as Smoluchowski equations. It yields both transition and steady-state behavior and allows for computations of moment-generating and large-deviation functions of observables defined along stochastic trajectories, such as the fluctuating particle current, heat, and work. The key strategy is to approximate the FPE by a master equation with transition rates in configuration space that obey a local detailed balance condition for arbitrary discretization. Its time-dependent solution is obtained by a direct computation of the time-ordered exponential, representing the propagator of the FPE, by summing over all possible paths in the discretized space. The method thus not only preserves positivity and normalization of the solutions but also yields a physically reasonable total entropy production, regardless of the discretization. To demonstrate the validity of the method and to exemplify its potential for applications, we compare it against Brownian-dynamics simulations of a heat engine based on an active Brownian particle trapped in a time-dependent quartic potential.

DOI: [10.1103/PhysRevE.99.032117](https://doi.org/10.1103/PhysRevE.99.032117)**I. INTRODUCTION**

Many natural phenomena exhibit a timescale separation between “slow” and “fast” degrees of freedom. The variables varying slowly in space or time can then be characterized by a self-contained coarse-grained dynamics, which is—for not too extreme coarse-graining—perceptibly perturbed by fluctuations arising from the noisy dynamics of the fast variables.

The Fokker-Planck equation (FPE) represents a most comprehensive description of such time-separated phenomena [1]. It predicts not only the average dynamics of the slow variables but directly addresses, in a technically manageable way, their complete probability distribution, which includes the relevant information about the fluctuations of the slow degrees of freedom induced by the fast ones. To achieve this, all of the slow variables need to be resolved explicitly in a so-called Markovian description, such that the remaining fast variables evolve without perceptible memory of the past dynamics.

Over the past century, the FPE has found applications in various scientific disciplines ranging from physics and chemistry to biology and ecology and even economy and finance [1–7]. Needless to say, only in very few special cases can it be solved analytically, so one usually has to resort to analytical approximations, computer simulations, and numerical methods [1, 8–12]. Both the Fokker-Planck formulation of stochastic dynamics and efficient techniques for its numerical solution become particularly relevant for situations far from equilibrium, where the slow variables are, as a rule, found to

exhibit non-Gaussian characteristic fluctuations that contain a crucial part of the information about the system of interest.

For the physical interpretation of this information, it is moreover crucial to also evaluate functionals defined along individual trajectories of the underlying stochastic process, which is one of the main tasks of stochastic thermodynamics. Important examples of such functionals are fluctuating particle currents and fluctuating heat and work in systems of Brownian particles, individual proteins, or living bacteria, which often operate under conditions far from equilibrium [13–15].

In this paper, we describe a simple thermodynamically consistent matrix numerical method (MNM) for solving overdamped FPEs with time-dependent coefficients, also known as Smoluchowski equations. Not only can the method resolve the transition and long-time behavior of probability distributions described by the FPE, but it is also naturally applicable to computations of moment-generating functions (MGFs) and large-deviation functions (LDFs) for various types of functionals defined along the trajectories of the stochastic process described by the FPE. This is achieved by a discretization that transforms the FPE into a master equation with transition rates that obey a local detailed balance condition. The time evolution of its solution is calculated from the time-ordered exponential, representing the discretized FPE propagator, by summing over all possible paths in the discretized configuration space. The MNM thus addresses all of the mentioned functions directly and gives physically reasonable results both from a probabilistic and from a thermodynamic point of view, for arbitrary discretization. Namely, the MNM is constructed to preserve the normalization and positivity of the initial distribution and to predict the correct entropy production of

\*viktor.holubec@mff.cuni.cz

the discrete models emerging upon discretization, at arbitrary resolution.

Toward the end of the paper, we test the MNM and illustrate its power by focusing on a specific example, namely a heat engine based on an overdamped active particle trapped in a time-dependent quartic potential and communicating with a heat bath at a time-dependent temperature. We investigate the dynamics of the particle and the fluctuations of work and heat exchanged with the bath, using both the proposed MNM and Brownian dynamics (BD) simulations, checking that both methods give the same results.

## II. PRINCIPLES OF THE MNM

For pedagogical reasons, we introduce the MNM for a two-dimensional stochastic system, parametrized by coordinates  $x$  and  $y$ . Although such a system can represent diffusion of an abstract object in an abstract energy landscape, we find it helpful to allude, in our description, to the intuitive paradigmatic example of an overdamped Brownian particle. Furthermore, we assume that the diffusion matrix  $\mathcal{D}$  and the mobility matrix  $\mu$  are diagonal:  $\mathcal{D} = \text{diag}(D_x, D_y)$  and  $\mu = \text{diag}(\mu_x, \mu_y)$ . An extension to higher dimensions and off-diagonal matrices  $\mathcal{D}$  and  $\mu$  is straightforward. The FPE for the probability density function (PDF)  $\rho(x, y, t)$  to find the system at time  $t$  in microstate  $(x, y)$  is the parabolic partial differential equation

$$\begin{aligned} \partial_t \rho(x, y, t) &= \mathcal{L}(x, y, t) \rho(x, y, t) \\ &= \partial_x [\partial_x D_x - \mu_x F_x] \rho(x, y, t) \\ &\quad + \partial_y [\partial_y D_y - \mu_y F_y] \rho(x, y, t) \end{aligned} \quad (1)$$

with generally time- and position-dependent diffusion coefficients  $D_x > 0$  and  $D_y > 0$ , mobilities  $\mu_x$  and  $\mu_y$ , and forces  $F_x$  and  $F_y$  in the  $x$  and  $y$  directions, respectively. The force  $\mathbf{F} = (F_x, F_y)$  does not need to be conservative, stemming from some potential  $U = U(x, y, t)$ , such that  $\mathbf{F} = -\nabla U = -(\partial_x U, \partial_y U)$ . Below, we show that the most general form of the FPE that can be solved using the MNM is Eq. (1) with time-dependent but position-independent diffusion coefficients (11). If one is willing to sacrifice the thermodynamic consistency of the MNM, its minimal modification moreover allows one to solve Eq. (1) in full generality, i.e., with all the coefficients  $D_x$ ,  $D_y$ ,  $\mu_x F_x$ , and  $\mu_y F_y$  time and position dependent.

The main idea, exploited in this paper, to solve the complicated time-dependent equation (1) is to approximate the underlying time-and-space continuous stochastic process by a time continuous hopping process in a discrete configuration space. To this end, we approximate the FPE (1) with the Fokker-Planck operator  $\mathcal{L}(x, y, t)$  by a master equation with a transition rate matrix  $\mathcal{R}$ :

$$\partial_t \rho(x, y, t) = \mathcal{L}(x, y, t) \rho(x, y, t) \rightarrow \dot{\mathbf{p}}(t) = \mathcal{R}(t) \mathbf{p}(t). \quad (2)$$

Here,  $\mathbf{p}(t)$  is the vector of probabilities of occupation of the individual discrete states which approximates the PDF  $\rho(x, y, t)$ , and  $\dot{\mathbf{p}}(t)$  denotes its total time derivative. In case of time-independent coefficients in  $\mathcal{L}$ , the master equation is simply solved by matrix exponentiation of the constant

rate matrix  $\mathcal{R}$ , namely  $\mathbf{p}(t) = \exp[(t - t_0)\mathcal{R}]\mathbf{p}(t_0)$ . In the case of time-dependent coefficients, the strategy is to construct a piecewise time-constant approximation  $\tilde{\mathcal{R}}(t)$  to the time-dependent rate matrix  $\mathcal{R}(t)$ , solve the master equation in the time intervals where  $\tilde{\mathcal{R}}(t)$  is constant using matrix exponentiation, and, finally, employ the Markov property of the stochastic process to construct an approximate solution by concatenation, i.e., by multiplying the matrix exponentials.

Simple variants of the MNM have already been used by one of the authors to investigate several model systems [16–18]. The main merits of the present paper are twofold. First, we generalize the previously used method to FPEs with time-dependent coefficients and show how to calculate MGFs and LDFs for various functionals in such a setting. Second, in the previous works [16–18] the MNM was always presented only as a minimal recipe in technical appendices. Here, we provide a comprehensive derivation and discussion of the method, including all its important aspects.

The following sections give a detailed description of the MNM. First, in Sec. II A, we specify the discretization mesh used throughout the paper. The precise meaning of thermodynamic consistency and the transition rates obeying the local detailed balance condition are described in Secs. II B and III A. In Sec. III B, we discuss several boundary conditions which can be implemented with the method. In Sec. IV, we show how to solve the approximate master equation. The long Sec. V is devoted to computations of MGFs and LDFs for various functionals defined along the trajectories of the stochastic process described by the FPE. The general presentation of the MNM is closed by a discussion of several practical issues and of its computational efficiency compared to other methods, in Sec. VI. After that, in Sec. VII, we show how to apply the general theory by guiding the reader through a solved example: a heat engine consisting of an active particle trapped in a time-dependent quartic potential and communicating with a bath with time-dependent temperature. We conclude in Sec. VIII. In Appendix A, we show why the (locally) detailed-balanced master equation, which is at the heart of the MNM, cannot be used for solving FPEs with position-dependent diffusion coefficients and what modifications of the MNM are necessary in order to solve Eq. (1) in full generality.

### A. Space discretization scheme

Our goal is to solve the FPE Eq. (1) numerically. In general, this can be done only within some finite space-and-time domain, which allows us to approximate the continuous space-time with a finite number of discrete points. For simplicity, we limit our presentation to rectangular domains of the form  $[t_0, \tau] \times [x_-, x_+] \times [y_-, y_+]$  only. The generalization to more complicated domains is straightforward. The time domain is naturally bounded by the initial time  $t_0$ , where we impose an initial PDF  $\rho(x, y, t_0)$ , and the final time of integration  $\tau$ . The finite space domain  $[x_-, x_+] \times [y_-, y_+]$  is defined by the boundary conditions imposed at boundaries  $x = x_{\pm}$  and  $y = y_{\pm}$ . The boundary conditions which can be handled by the MNM will be detailed in Sec. III B. Here, we present the discretization of the (configuration) space domain  $[x_-, x_+] \times [y_-, y_+]$  used in the rest of the paper.

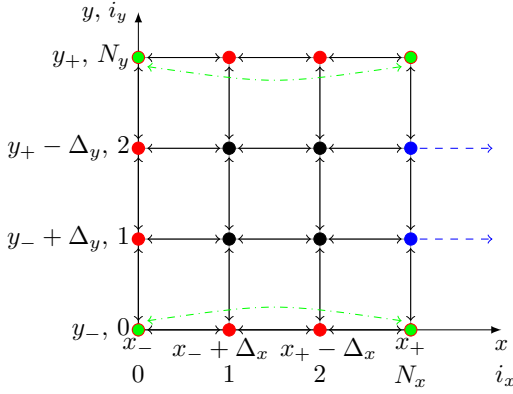


FIG. 1. Sketch of the configuration space discretization used for the numerical solution of the two-dimensional overdamped Fokker-Planck equation (1). The black points mark the states inside the domain  $[x_-, x_+] \times [y_-, y_+]$ , and the colored points form the boundary (see Sec. III B). The black full arrows depict the allowed transitions with the “bulk” transition rates (13) or (18) (horizontal transition) and (14) or (19) (vertical transitions). The red boundary is reflecting and thus the particles cannot cross the red states (hence no red arrows). The blue boundary is absorbing and thus particles can leave the system from these sites (depicted by one-way dashed blue arrows). The states in the corners of the domain require two boundary conditions. In the figure, we impose reflecting boundary condition in the  $y$  direction (depicted by red circumferences of the points) and periodic boundary conditions in the  $x$  direction (depicted by green interiors of the points). The periodic boundary allows the particles to leave the system in the  $x$  direction. The leaving particles reenter the system at the opposite side of the domain, as depicted by the green dot-dashed arrows.

For simplicity, we impose a rectangular discrete mesh with  $(N_x + 1)(N_y + 1)$  discrete configurations with coordinates  $\{i_x, i_y\}$ ,

$$i_x = \left\lfloor \frac{x - x_-}{\Delta_x} \right\rfloor, \quad \Delta_x = \frac{x_+ - x_-}{N_x}, \quad (3)$$

$$i_y = \left\lfloor \frac{y - y_-}{\Delta_y} \right\rfloor, \quad \Delta_y = \frac{y_+ - y_-}{N_y}, \quad (4)$$

$i_x = 0, 1, \dots, N_x$  and  $i_y = 0, 1, \dots, N_y$ , as illustrated in Fig. 1. The symbol  $\lfloor x \rfloor$  denotes the floor function. The generalization of the method to more complicated discretization meshes, which might be specifically adapted to some salient features of the coefficients in the FPE (1), is straightforward.

Let us denote as  $p_{i_x, i_y} = p_{i_x, i_y}(t)$  the occupation probabilities of the individual lattice points  $i_x, i_y$ . Allowing only transitions between neighboring lattice points (cf. the arrows in Fig. 1), the counterpart of the FPE (1) on the discrete lattice is the master equation [19]

$$\begin{aligned} \dot{p}_{i_x, i_y} = & r_{i_x+1 \rightarrow i_x}^{i_y} p_{i_x+1, i_y} + r_{i_x-1 \rightarrow i_x}^{i_y} p_{i_x-1, i_y} \\ & + l_{i_y+1 \rightarrow i_y}^{i_x} p_{i_x, i_y+1} + l_{i_y-1 \rightarrow i_y}^{i_x} p_{i_x, i_y-1} \\ & - (r_{i_x \rightarrow i_x+1}^{i_y} + r_{i_x \rightarrow i_x-1}^{i_y} + l_{i_y \rightarrow i_y+1}^{i_x} + l_{i_y \rightarrow i_y-1}^{i_x}) p_{i_x, i_y}, \end{aligned} \quad (5)$$

where the symbol  $r_{i_x \rightarrow i_x+1}^{i_y} = r_{i_x \rightarrow i_x+1}^{i_y}(t) \geq 0$  denotes the transition rate in the  $x$  direction from site  $(i_x, i_y)$  to site  $(i_x + 1, i_y)$  and  $l_{i_y \rightarrow i_y+1}^{i_x} = l_{i_y \rightarrow i_y+1}^{i_x}(t) \geq 0$  denotes the transition rate in the  $y$  direction from site  $(i_x, i_y)$  to site  $(i_x, i_y + 1)$ . These transition rates must be chosen in such a way that the occupation probabilities  $p_{i_x, i_y}$  determine the correct solution of the FPE (1) in the limit of an infinitely fine mesh:

$$\rho(x, y, t) = \lim_{\Delta_x \rightarrow 0} \lim_{\Delta_y \rightarrow 0} \frac{p_{i_x(x), i_y(y)}(t)}{\Delta_x \Delta_y}. \quad (6)$$

The master Eq. (5) possesses a simple probabilistic interpretation. For example, the expression  $r_{i_x+1 \rightarrow i_x}^{i_y}(t) p_{i_x+1, i_y}(t) dt$  stands for the probability to jump from the site  $(i_x + 1, i_y)$  at time  $t$  to the site  $(i_x, i_y)$  during the infinitesimally short time interval  $dt$ . The time derivative of the occupation probability in Eq. (5) is thus given by the probability to enter the site from neighboring sites [positive terms in (5)] minus the probability to leave it to neighboring sites [negative terms in (5)], during an infinitesimally short time interval.

## B. Thermodynamic consistency

The probabilistic interpretation of the master equation, (5), implies that solutions produced by the proposed discretization are by construction non-negative for any non-negative initial condition and conserve the normalization in absence of source- or sink-boundary conditions (cf. Sec. III B), regardless of the discretization parameters  $\Delta_x$  and  $\Delta_y$ .

There are various ways to write the rates for transitions between the lattice points depicted in Fig. 1 that lead to the same FPE (1) in the limit of infinitely fine discretization. Here we want to propose a mapping (2) guided by the aim to approximate the process described by the FPE (1) in a thermodynamically consistent way, for arbitrary mesh resolution. A discretization scheme with similar properties was proposed already in 1970 by Chang and Cooper [20]. Compared to their presentation, our interpretation of the discretization scheme in terms of master equations provides a clearer physical interpretation of the transition rates and a natural basis for studying various functionals, defined on realizations of the stochastic process, in terms of moment-generating functions and characteristic functions.

On the level of coarse-grained stochastic models, the time-reversal symmetry of the (standard) microscopic Hamiltonian dynamics manifests itself in a so-called *local detailed balance condition* [21–25]. This condition should therefore be expected to hold for any physically reasonable stochastic dynamics. In fact, it can be viewed as the most fundamental tool for devising consistent thermodynamic notions for a microscopically grounded stochastic system. It states that the logarithm of ratio of the (conditional) path probability  $P(\mathbf{r}_i \rightarrow \mathbf{r}_f, \mathbf{\Gamma}) = P(\mathbf{\Gamma})$  for the system to go from  $\mathbf{r}_i$  to  $\mathbf{r}_f$  along the path  $\mathbf{\Gamma}$  over the probability  $P^*(\mathbf{r}_f \rightarrow \mathbf{r}_i, \mathbf{\Gamma}^*) = P^*(\mathbf{\Gamma}^*)$  to return from  $\mathbf{r}_f$  to  $\mathbf{r}_i$  along the time-reversed path  $\mathbf{\Gamma}^*$  (with time-reversed dynamics) is proportional to the entropy change  $\Delta S_R(\mathbf{r}_i \rightarrow \mathbf{r}_f, \mathbf{\Gamma}) = \Delta S_R(\mathbf{\Gamma})$  in the reservoir to which the system is connected along the path  $\mathbf{\Gamma}$ , briefly

$$k_B \log \frac{P(\mathbf{\Gamma})}{P^*(\mathbf{\Gamma}^*)} = \Delta S_R(\mathbf{\Gamma}). \quad (7)$$

Let us consider an overdamped diffusion process where a particle communicates with a single global equilibrium bath at constant temperature  $T$  and is driven by a force  $\mathbf{F} = (F_x, F_y)$ . The fluctuation-dissipation theorem implies that the bath temperature is given by  $T = D_x/(k_B\mu_x) = D_y/(k_B\mu_y)$  with time-and-space constant diffusion coefficients  $D_x$  and  $D_y$  and mobilities  $\mu_x$  and  $\mu_y$ . The amount of entropy produced when the particle diffuses from  $\mathbf{r}_i = (x_i, y_i)$  to  $\mathbf{r}_f = (x_f, y_f)$  along the path  $\Gamma = [x(t), y(t)]$  parametrized by  $t \in [t_i, t_f]$  is given by the energy transferred to the reservoir along this process divided by the reservoir temperature  $T$ . For overdamped dynamics, the energy dissipated into the bath is given by the work  $W(\Gamma) = \int_{\Gamma} \mathbf{F}(\Gamma) \cdot d\Gamma = \int_{t_i}^{t_f} dt \mathbf{F}[x(t), y(t), t] \cdot [dx(t), dy(t)]/dt$  done by the force  $\mathbf{F}$  along  $\Gamma$ , and thus  $\Delta S_R(\Gamma) = W(\Gamma)/T$ .

This equation can be generalized to situations where we connect the system at every point  $(x, y)$  to one joint reservoir or even two independent reservoirs with time-dependent temperatures. The bath at temperature  $T_x(x, y, t) = D_x/(k_B\mu_x)$  induces diffusion in the  $x$  direction, and the one with temperature  $T_y(x, y, t) = D_y/(k_B\mu_y)$  induces diffusion in the  $y$  direction. Here, we again assumed that the diffusion and the mobility matrices  $\mathcal{D}$  and  $\mu$  in Eq. (1) are related by the fluctuation-dissipation theorem for each coordinate. In this generalized case, the total amount of entropy produced in all the reservoirs along the trajectory  $\Gamma$  reads

$$\begin{aligned} \Delta S_R(\Gamma) &= \int_{t_i}^{t_f} dt \left[ \frac{F_x(t)}{T_x(t)}, \frac{F_y(t)}{T_y(t)} \right] \cdot \left[ \frac{dx(t)}{dt}, \frac{dy(t)}{dt} \right] \\ &= \int_{\mathbf{r}_i}^{\mathbf{r}_f} \left[ \frac{F_x(t)}{T_x(t)}, \frac{F_y(t)}{T_y(t)} \right] \cdot [dx(t), dy(t)]. \end{aligned} \quad (8)$$

In order to find a reasonable form of the transition rates [transition probabilities  $P(\Gamma)$  per unit time] based on Eqs. (7) and (8), we assume that the explicit time dependence of the forces and temperatures can be neglected for the transition rates at time  $t_i$ . If such a timescale separation holds, we can evaluate the force and temperature fields in Eq. (8) at time  $t_i$ , thereby effectively approximating the process with time-dependent coefficients by a sequence of processes with time-independent coefficients. For an overdamped diffusion process with time-independent coefficients, the forward and reversed dynamics are identical, i.e.,  $P^*(\Gamma) = P(\Gamma)$ . Let us now use this formula to uncover the functional dependence of the transition probabilities, fulfilling the local detailed balance condition (7), on the entropy change  $\Delta S_R(\Gamma)$ .

Without loss of generality, we write the transition probability as  $P(\Gamma) = A(\Gamma) \exp[B(\Gamma)/k_B]$ , where  $A$  denotes a symmetric and  $B$  denotes an antisymmetric unknown function with respect to the path reversal, i.e.,  $A(\Gamma) = A(\Gamma^*)$  and  $B(\Gamma) = -B(\Gamma^*)$ . Inserting this ansatz into Eq. (7) and using the condition  $P^*(\Gamma) = P(\Gamma)$ , we find that  $B(\Gamma) - B(\Gamma^*) = 2B(\Gamma) = \Delta S_R(\Gamma)$ . We thus arrive at the expression

$$\frac{P(\Gamma)}{A(\Gamma)} = \frac{A(\Gamma)}{P(\Gamma^*)} = \exp \left[ \frac{\Delta S_R(\Gamma)}{2k_B} \right] \quad (9)$$

for the transition probabilities, the validity of which we assume for each transition and, consequently, also for arbitrary sequence of transitions. The prefactor  $A(\Gamma)$  depends on the

details of the dynamics and we determine it by inserting the transition rates fulfilling (9) into the FPE.

### III. IMPLEMENTATION OF THE MNM

#### A. Transition rates

The formulas (9) can be applied to arbitrary discretization meshes. Let us now identify the points  $\mathbf{r}_i$  and  $\mathbf{r}_f$  with neighboring sites of the rectangular lattice defined in Sec. II A and depicted in Fig. 1. We now take  $\mathbf{r}_i = (x, y)$  and  $\mathbf{r}_f = \mathbf{r}_f^x = (x + \Delta_x, y)$  for the horizontal transitions and  $\mathbf{r}_f = \mathbf{r}_f^y = (x, y + \Delta_y)$  for the vertical ones. The probabilities  $P(\mathbf{r}_i \rightarrow \mathbf{r}_f)$  will now determine the transition rates between the individual lattice points.

The formulas (9) imply that the necessary condition for the transition rates in the  $x$  direction in Eq. (5) to obey the local detailed balance principle is

$$\frac{r_{i_x \rightarrow i_x+1}}{A_{i_x+1/2}} = \frac{A_{i_x+1/2}}{r_{i_x+1 \rightarrow i_x}} = \exp \left[ \frac{\Delta S_R(\mathbf{r}_i \rightarrow \mathbf{r}_f^x)}{2k_B} \right], \quad (10)$$

where  $A_{i_x+1/2}$  is a symmetric prefactor, and similarly for the rates in the  $y$  direction. In Appendix A, we show that the transition rates satisfying these conditions can (even in one dimension) yield the FPE (1) only for position-independent diffusion coefficients. Hence, the most general FPE, which can be solved numerically using such transition rates, reads

$$\begin{aligned} \partial_t \rho(x, y, t) &= \mathcal{L}(x, y, t) \rho(x, y, t) \\ &= [D_x \partial_x^2 - \partial_x \mu_x F_x] \rho(x, y, t) \\ &\quad + [D_y \partial_y^2 - \partial_y \mu_y F_y] \rho(x, y, t). \end{aligned} \quad (11)$$

Nevertheless, in Appendix A, we also show how to modify the detailed-balanced transition rates in order to address the FPE (1) in its full generality. The resulting generalized MNM respects the local detailed balance condition in case of position-independent diffusion coefficients. For position-dependent diffusion coefficients, the local detailed balance condition and the underlying microreversibility, valid in the continuous FPE (1), are thus necessarily broken on the coarse-grained level of the master equation (5). This anticipates problems of attempts to mimic effects caused by spatially modulated mobilities using models with (temporally) diffusing diffusivities; see, for example, Ref. [26].

#### 1. Equilibrium dynamics

Whenever the quantities  $F_x/k_B T_x = \mu_x F_x/D_x$  and  $\mu_y F_y/k_B T_y = \mu_y F_y/D_y$  can be written using a dimensionless potential  $\tilde{U}(x, y, t)$ , such that

$$\left( \frac{F_x}{k_B T_x}, \frac{F_y}{k_B T_y} \right) = -\nabla \tilde{U} = -(\partial_x \tilde{U}, \partial_y \tilde{U}), \quad (12)$$

the formula (8) can be written as  $\Delta S_R(\Gamma) = \Delta S_R(\mathbf{r}_i \rightarrow \mathbf{r}_f) = k_B [\tilde{U}(\mathbf{r}_i, t) - \tilde{U}(\mathbf{r}_f, t)]$ . The transition rates satisfying the condition (9) and yielding the FPE (11) in the limit  $\Delta_x \rightarrow 0$ ,  $\Delta_y \rightarrow 0$  of the master equation (5) can then be found without

any further approximation by inserting the rates of the form (10) into the FPE, as in Appendix A. They read

$$r_{i_x \rightarrow i_x \pm 1}^{i_y} = \frac{D_x}{\Delta_x^2} \exp\left(-\frac{\tilde{U}_{i_x \pm 1, i_y} - \tilde{U}_{i_x, i_y}}{2}\right), \quad (13)$$

$$l_{i_y \rightarrow i_y \pm 1}^{i_x} = \frac{D_y}{\Delta_y^2} \exp\left(-\frac{\tilde{U}_{i_x, i_y \pm 1} - \tilde{U}_{i_x, i_y}}{2}\right), \quad (14)$$

with  $\tilde{U}_{i_x, i_y} = \tilde{U}_{i_x, i_y}(t) = \tilde{U}(x_- + \Delta_x i_x, y_- + \Delta_y i_y, t)$  and the symmetric prefactors  $D_x(t)/\Delta_x^2$  and  $D_y(t)/\Delta_y^2$ .

We refer to this as equilibrium dynamics because the FPE (11) with time-independent coefficients fulfilling (12) leads to the Boltzmann stationary distribution  $\rho(x, y, \infty) = \tilde{\rho}(x, y) \propto \exp[-\tilde{U}(x, y)]$ . This can be verified by the direct substitution of the Boltzmann distribution into Eq. (11). Similarly, the stationary solution of the master equation (5) reads  $p_{i_x, i_y}(\infty) = \tilde{p}_{i_x, i_y} \propto \exp(-\tilde{U}_{i_x, i_y})$ , regardless of the discretization.

Physically, the most important feature of the equilibrium stationary distribution is that in this state all mesoscopic probability currents in the system vanish. On the level of the FPE (11), this is reflected by the formulas  $j_x = -D_x \partial \tilde{\rho} + \mu_x F_x \tilde{\rho} = 0$  and  $j_y = -D_y \partial \tilde{\rho} + \mu_y F_y \tilde{\rho} = 0$ . On the level of the master equation (5), the probability current in the  $x$  direction reads  $j_x^{i_y}(i_x \rightarrow i_x + 1) = r_{i_x \rightarrow i_x + 1}^{i_y} p_{i_x, i_y} - r_{i_x + 1 \rightarrow i_x}^{i_y} p_{i_x + 1, i_y}$  and similarly for the probability current in the  $y$  direction. That these currents vanish for the Boltzmann distribution  $\tilde{p}_{i_x, i_y}$  is usually written in the form of the conventional *global detailed balance conditions*

$$\frac{r_{i_x \rightarrow i_x + 1}^{i_y}}{r_{i_x + 1 \rightarrow i_x}^{i_y}} = \exp\left[-(\tilde{U}_{i_x + 1, i_y} - \tilde{U}_{i_x, i_y})\right], \quad (15)$$

$$\frac{l_{i_y \rightarrow i_y + 1}^{i_x}}{l_{i_y + 1 \rightarrow i_y}^{i_x}} = \exp\left[-(\tilde{U}_{i_x, i_y + 1} - \tilde{U}_{i_x, i_y})\right]. \quad (16)$$

Let us stress that the ‘‘equilibrium dynamics’’ described in this section can sometimes be observed even though the system is not in equilibrium, for example, if the coefficients in the FPE (11) are time dependent and/or if the system relaxes from a nonequilibrium initial distribution  $\rho \neq \tilde{\rho}$ .

## 2. Nonequilibrium dynamics

If the quantities  $F_x/k_B T_x = \mu_x F_x/D_x$  and  $\mu_y F_y/k_B T_y = \mu_y F_y/D_y$  cannot be written using a single potential, one can still formally define (different) pseudopotentials for the individual degrees of freedom:

$$\left(\frac{F_x}{k_B T_x}, \frac{F_y}{k_B T_y}\right) = -(\partial_x \tilde{U}, \partial_y \tilde{V}). \quad (17)$$

In this case, it is not possible to get rid of the path dependence of the integral in Eq. (8) as was done for the equilibrium dynamics. Therefore, we now assume that for the transitions in the  $x$  direction the entropy change can be well approximated by  $\Delta S_R(\Gamma) = \Delta S_R[(x, y) \rightarrow (x + \Delta_x, y)] = \tilde{U}(x, y, t) - \tilde{U}(x + \Delta_x, y, t)$ . This means that, from all possible paths  $\Gamma$  between the points  $(x, y)$  and  $(x + \Delta_x, y)$ , we consider only the one with  $y$  coordinate fixed at  $y$ . We use a similar approximation also for the  $y$  direction. These

approximations become exact in the limit of vanishing  $\Delta_x$  and  $\Delta_y$ . The transition rates satisfying Eq. (9) under these approximations and leading to the FPE (5) as the  $\Delta_x \rightarrow 0$ ,  $\Delta_y \rightarrow 0$  limit of the master equation (5) read

$$r_{i_x \rightarrow i_x \pm 1}^{i_y} = \frac{D_x}{\Delta_x^2} \exp\left(-\frac{\tilde{U}_{i_x \pm 1, i_y} - \tilde{U}_{i_x, i_y}}{2}\right), \quad (18)$$

$$l_{i_y \rightarrow i_y \pm 1}^{i_x} = \frac{D_y}{\Delta_y^2} \exp\left(-\frac{\tilde{V}_{i_x, i_y \pm 1} - \tilde{V}_{i_x, i_y}}{2}\right), \quad (19)$$

with  $\tilde{U}_{i_x, i_y} = \tilde{U}_{i_x, i_y}(t) = \tilde{U}(x_- + \Delta_x i_x, y_- + \Delta_y i_y, t)$  and similarly for  $\tilde{V}_{i_x, i_y}$ . While for nonzero  $\Delta_x$  and  $\Delta_y$  these transition rates satisfy the local detailed balance condition (9) for the FPE only approximately, they satisfy it exactly on the discrete lattice depicted in Fig. 1, where the neighboring lattice points are interconnected exclusively by a single transition channel. On this discrete lattice, the process described by the rates (18) and (19) is thus perfectly thermodynamically consistent, yielding the correct entropy produced along the individual transitions regardless of the discretization.

For the nonequilibrium dynamics, not only is the time-dependent dynamics in general unknown, but also the characterization of the stationary distribution, attained in case of time-independent coefficients in the FPE (11) is a nontrivial task. The presence of persevering probability currents in such steady states implies that there might be stationary transport of particles, energy, etc. Formally, the transition rates (18) and (19) still obey a form reminiscent of the global detailed balance conditions (15) and (16), namely

$$\frac{r_{i_x \rightarrow i_x + 1}^{i_y}}{r_{i_x + 1 \rightarrow i_x}^{i_y}} = \exp\left[-(\tilde{U}_{i_x + 1, i_y} - \tilde{U}_{i_x, i_y})\right], \quad (20)$$

$$\frac{l_{i_y \rightarrow i_y + 1}^{i_x}}{l_{i_y + 1 \rightarrow i_y}^{i_x}} = \exp\left[-(\tilde{V}_{i_x, i_y + 1} - \tilde{V}_{i_x, i_y})\right], \quad (21)$$

but now with different potentials for the two degrees of freedom  $x$  and  $y$ . Intuitively, each of these conditions is trying to draw the system into the Boltzmann equilibrium corresponding to its own potential  $\tilde{U}$  or  $\tilde{V}$ , respectively. Globally, this competition leads to a nonequilibrium stationary state.

## B. Boundary conditions

The conditions at the boundaries of the configurational space domain  $[x_-, x_+] \times [y_-, y_+]$  require some extra care and give rise to modifications of the transition rates presented in the previous section. Briefly, while the rates (13) and (14) and the rates (18) and (19) are determined by the forces, temperatures, and mobilities explicitly appearing in the dynamic operator in the FPE (11), this is not necessarily true for the rates at the boundaries. The probabilistic interpretation of the master equation described below Eq. (6) allows a convenient implementation of arbitrary boundary conditions, which are thus also easily introduced into the MNM. We now show how to implement three basic types of boundary conditions.

(1) *Reflecting boundary condition*: The particle cannot cross the boundary.

(2) *Periodic boundary condition*: After crossing the boundary at one side of the domain, the particle returns to it, usually at its other side.

(3) *Absorbing boundary condition*: The particle is annihilated once it hits the boundary.

While the reflecting and periodic boundary conditions lead to the overall conservation of probability (no particles can leave the system), the absorbing boundary conditions lead to depletion of the system due to particle losses at the boundary. Besides using these three types of boundary conditions, one can use arbitrary combinations thereof (with some probability the particles can be allowed to leave the system or to appear at its other side, etc.).

### 1. Reflecting boundary conditions

Physically, the reflecting boundary condition corresponds to an infinite potential barrier. Overcoming such a barrier requires an infinite amount of energy from the reservoir which corresponds to an infinite change of entropy in Eq. (9) or potential in Eqs. (13) and (14) and Eqs. (18) and (19). The crossing rate across a reflecting barrier is thus 0, in accord with the rates Eqs. (13) and (14) and Eqs. (18) and (19).

Let us, for example, consider the situation depicted in Fig. 1, where the red points at the boundary obey reflecting boundary conditions. Specifically, we consider the point with coordinates (1,0) (the second one in the last line). Realizing that the transitions over the reflecting barrier are not allowed and that this point has only a single boundary toward negative  $i_y$ , the master equation (5) for this point reads

$$\dot{p}_{1,0} = r_{2 \rightarrow 1}^0 p_{2,0} + r_{0 \rightarrow 1}^0 p_{0,0} + l_{1 \rightarrow 0}^1 p_{1,1} - (r_{1 \rightarrow 2}^0 + r_{1 \rightarrow 0}^0 + l_{0 \rightarrow 1}^1) p_{1,0}. \quad (22)$$

Note that the transitions from (1,0) to (1, -1) and back occur with zero transition rate (and thus they do not show up in the equation). For other points with reflecting boundary, the master equation should be constructed in a similar manner.

### 2. Periodic boundary conditions

For periodic boundary conditions, the transition rates are still given by Eqs. (13) and (14) and Eqs. (18) and (19), one just needs to make the index periodic at the point where the periodic boundary condition is imposed. Consider for example the situation depicted in Fig. 1, where the upper left and upper right points are connected by the periodic boundary in the  $x$  direction. Then the rate to the right from the site  $(N_x, N_y)$  leads to the site  $(N_x + \Delta_x, N_y) = (0, N_y)$  and thus it reads

$$r_{N_x \rightarrow 0}^{N_y} = \frac{D_x}{\Delta_x^2} \exp\left(-\frac{\tilde{U}_{N_x + \Delta_x, N_y} - \tilde{U}_{N_x, N_y}}{2}\right). \quad (23)$$

In the expression for the transition rate, we used  $\tilde{U}_{N_x + \Delta_x, N_y} - \tilde{U}_{N_x, N_y} = \int_{x_- + N_x \Delta_x}^{x_- + (N_x + 1)\Delta_x} dx F_x / k_B T_x$  instead of  $\tilde{U}_{0, N_y} - \tilde{U}_{N_x, N_y}$ , because, although the sites  $(N_x + \Delta_x, N_y)$  and  $(0, N_y)$  coincide, the pseudopotential  $\tilde{U}$  may be discontinuous at the boundary for a nonconservative force  $F_x / k_B T_x$ .

Considering that the site  $(N_x, N_y)$  also possesses a reflecting boundary condition toward larger values of  $i_y$ , the

corresponding master equation reads

$$\dot{p}_{N_x, N_y} = r_{0 \rightarrow N_x}^{N_y} p_{0, N_y} + r_{N_x - 1 \rightarrow N_x}^{N_y} p_{N_x - 1, N_y} + l_{N_y - 1 \rightarrow N_y}^{N_x} p_{N_x, N_y - 1} - (r_{N_x \rightarrow 0}^{N_y} + r_{N_x \rightarrow N_x - 1}^{N_y} + l_{N_y \rightarrow N_y - 1}^{N_x}) p_{N_x, N_y}. \quad (24)$$

Other transitions across periodic boundaries should be handled in a similar manner.

### 3. Sources, sinks, and absorbing boundaries

Further examples are *source-sink boundary conditions*, meaning that particles can enter and leave the system across the boundary. They can be realized by connecting the boundary state to a particle reservoir. If the reservoir constantly feeds particles into the boundary state (the rate to go from the reservoir to the system is larger than the rate to go back), the boundary state behaves as a source. Vice versa, if the particles leave the boundary state toward the reservoir faster than they return, the boundary behaves as a sink.

The absorbing boundary condition represents a specific example of the sink condition with diverging rate to the reservoir and vanishing rate back. Physically, it corresponds to an infinitely deep potential cliff. When a particle hits such a boundary, it can be thought to release an infinite amount of energy that is dissipated to the bath, corresponding to a negatively infinite entropy change in Eq. (9) or an infinite change of the potential in Eqs. (13) and (14) and Eqs. (18) and (19). Under such circumstances, the transition rates (13) and (14) and rates (18) and (19) diverge.

In order to avoid including such infinite rates in the master equation, we take as ‘‘auxiliary’’ boundary points those bulk points next to the actual boundary. The transition rates from the bulk into this auxiliary boundary and from it to all neighboring grid points are given by Eqs. (13) and (14) or Eqs. (18) and (19), while the actual boundary points are assigned a vanishing back rate into the bulk. Consider, for example, the situation depicted in Fig. 1, where the point  $(N_x, 2)$  at the end of the second row from the top possesses an absorbing boundary in the  $x$  direction. From the discussion above, it follows that the corresponding master equation reads

$$\dot{p}_{N_x, 2} = r_{N_x - 1 \rightarrow N_x}^2 p_{N_x - 1, 2} + l_{3 \rightarrow 2}^{N_x} p_{N_x, 3} + l_{1 \rightarrow 2}^{N_x} p_{N_x, 1} - (r_{N_x \rightarrow N_x + 1}^2 + r_{N_x \rightarrow N_x - 1}^2 + l_{2 \rightarrow 3}^{N_x} + l_{2 \rightarrow 1}^{N_x}) p_{N_x, 2}. \quad (25)$$

Here, the transition rate  $r_{N_x \rightarrow N_x + 1}^2$  for transitions out of the system is given by Eqs. (13) and (14) or Eqs. (18) and (19). Since we assume that the absorbing boundary in the continuous space described by the FPE is located at  $x_+ + \Delta_x$ , the pseudopotentials  $\tilde{U}_{N_x + 1, 2}$  and  $\tilde{V}_{N_x + 1, 2}$  needed to evaluate the rates are well defined. Other transitions across absorbing boundaries should be handled in a similar manner.

## IV. SOLUTION OF THE MASTER EQUATION

Having described the transition rates in the approximate master equation (5), we will briefly explain how this equation can be solved in various situations. The key step always consists in rewriting the master equation (5) in the matrix form

$$\dot{\mathbf{p}}(t) = \mathcal{R}(t)\mathbf{p}(t), \quad (26)$$

where the  $(N_x + 1)(N_y + 1) \times (N_x + 1)(N_y + 1)$  matrix  $\mathcal{R}(t)$  contains the transition rates (13) and (14) or rates (13) and (19) in such a way that Eqs. (5) and (26) are equivalent. The elements of the  $(N_x + 1)(N_y + 1)$ -dimensional vector  $\mathbf{p}(t)$  are given by the occupation probabilities  $p_{i_x, i_y}(t)$ . One possible construction is [16]

$$\mathbf{p}(t) = (p_{0,0}, \dots, p_{N_x,0}, p_{0,1}, \dots, p_{N_x,1}, \dots, p_{N_x, N_y})^\top, \quad (27)$$

where  $\top$  denotes the transposition. In this case, the probability  $p_{i_x, i_y}$  is contained in the element  $j(i_x, i_y) = i_y(N_x + 1) + i_x + 1$  of the vector  $\mathbf{p}(t)$ . The inverse transformation reads

$$i_x(j) = j - i_y(j)(N_x + 1) - 1, \quad (28)$$

$$i_y(j) = \lfloor (j - 1)/(N_x + 1) \rfloor. \quad (29)$$

The time dependence of the rate matrix  $\mathcal{R}(t)$  comes directly from the time dependence of the coefficients  $D_x, D_y, \mu_x, \mu_y, F_x,$  and  $F_y$  in the FPE (11) appearing in the expressions for the transition rate. For the reflecting and periodic boundary conditions described in the preceding section, the matrix  $\mathcal{R}(t)$  is stochastic ( $\sum_i [\mathcal{R}(t)]_{ij} = 0$ ) and thus Eq. (26) conserves normalization of the probability vector  $\mathbf{p}(t)$ . All the following methods of solution for Eq. (26) in diverse situations are based on basic algebraic manipulations involving the rate matrix.

#### A. Time-independent coefficients

Let us start with the simplest situation of time-constant coefficients in the FPE (11) which leads to a time-independent rate matrix,  $\mathcal{R}(t) = \mathcal{R}$ . In this case, the Green's function (to which we also refer as the propagator throughout the text) for Eq. (26) is given by the matrix exponential

$$\mathcal{U}(t, t_0) = \exp[\mathcal{R}(t - t_0)] \quad (30)$$

and thus the time evolution of the probability vector  $\mathbf{p}(t)$  departing from the initial condition  $\mathbf{p}(t_0)$  is given by

$$\mathbf{p}(t) = \mathcal{U}(t, t_0)\mathbf{p}(t_0). \quad (31)$$

If the system state converges to a time-independent steady-state  $\mathbf{p}_\infty$  at late times, this steady state can be either determined from Eq. (31) as  $\mathbf{p}_\infty = \lim_{t \rightarrow \infty} \mathbf{p}(t)$ , or, much more conveniently, as an eigenvector of the rate matrix corresponding to the eigenvalue 0:

$$\dot{\mathbf{p}}_\infty = \mathcal{R}\mathbf{p}_\infty = 0. \quad (32)$$

Because only jumps between the neighboring sites are allowed (see Fig. 1), the time-independent jump matrix  $\mathcal{R}$  is sparse. Especially (but not solely) for the computation of the steady-state vector  $\mathbf{p}_\infty$  from Eq. (32) one can benefit from fast numerical procedures for sparse matrices (see Sec. VI for more details).

#### B. Time-dependent coefficients

The ability to calculate the propagator  $\mathcal{U}(t, t_0)$  for FPEs with time-constant coefficients eventually allows us to obtain the Green's function for Eq. (11) with arbitrary time-dependent coefficients. We discretize the relevant time interval  $[t_0, t_0 + \tau]$  into  $N_t$  time slices of length  $\Delta_t = \tau/N_t$ . We assume that the driving can be approximated by appropriately chosen constants during all of these intervals and that it

may change only stepwise from one interval to the next. In other words, we replace the actual time-dependent coefficients  $D_x, D_y, \mu_x, \mu_y, F_x,$  and  $F_y$  in Eq. (11) by their piecewise constant approximations  $\bar{D}_x(t) = D_x(t_0 + i_t \Delta_t)$ ,  $i_t = \lfloor (t - t_0)/\Delta_t \rfloor$ , and similarly for the other coefficients. The propagators for the individual time intervals, during which the driving is constant, can be obtained using the procedure described above. Denoting by  $\mathcal{U}_i$ ,  $i \geq 1$ , the propagator  $\mathcal{U}[t_0 + (i - 1)\Delta_t, t_0 + i\Delta_t] \equiv \exp[\mathcal{R}(t_0 + i\Delta_t)\Delta_t]$  corresponding to the  $i$ th time interval and by  $\mathcal{U}_0 \equiv \mathcal{I}$  the unit matrix, we obtain the approximate Green's function under continuous driving, for arbitrary  $t$ ,  $t_0 + \tau > t > t_0$ , as

$$\mathcal{U}(t, t_0) = \lim_{\Delta_t \rightarrow 0} \prod_{i=0}^{i_t(t)} \mathcal{U}_i. \quad (33)$$

With this Green's function, the time evolution of the probability vector  $\mathbf{p}(t)$  follows again from Eq. (31).

Let us note that the presented discretization of time is just one of many possible choices. While we evaluate the time-dependent parameters at time  $t' = t$  in order to compute the state of the system at time  $t + \Delta_t$ , one can also use values of the time-dependent parameters at any other time  $t'$  in the interval  $(t, t + \Delta_t)$ . What value  $t'$  suits best a specific situation depends on the relaxation time of the system. If it is long compared to  $\Delta_t$ , one should take  $t' = t$ . On the other hand, if the relaxation is fast compared to  $\Delta_t$ , one should rather take  $t' = t + \Delta_t$ .

### V. FUNCTIONALS DEFINED ALONG THE STOCHASTIC PROCESS

Besides computing the distribution  $\rho(x, y, t)$  and then using it to evaluate averages, moments, reduced distribution functions for  $x$  and  $y$ , and the mesoscopic probability currents  $j_x$  and  $j_y$ , the probabilistic interpretation of the discrete approximation (5) of the FPE (11) can moreover be employed to address the statistics of various stochastic variables, other than position, directly. Useful examples are microscopic currents or linear combinations thereof, and heat, work, or efficiency, which are much studied objects in stochastic thermodynamics.

Application of the MNM to probability currents was already suggested in Refs. [16,17], where it was employed in the calculation of the diffusion coefficient in a model of a two-dimensional Brownian ratchet. Here we discuss this approach in greater generality.

#### A. Probability currents

The probability current  $\mathbf{j}(x, y, t) = (j_x, j_y)$  at time  $t$  and position  $\mathbf{r} = (x, y)$  can be defined in two equivalent ways. First, one can define it *mesoscopically*, rewriting the FPE (11) as  $\partial_t \rho(x, y, t) = \mathcal{L}(x, y, t) = -\nabla \cdot \mathbf{j}(x, y, t)$ , leading to the expression

$$\mathbf{j}(x, y, t) = -(D_x \partial_x + \mu_x F_x, D_y \partial_y + \mu_y F_y) \rho. \quad (34)$$

On the level of the master equation (5), these expressions read

$$\begin{aligned} \dot{p}_{i_x, i_y}(t) = & j_x^{i_y}(i_x + 1 \rightarrow i_x) + j_x^{i_y}(i_x - 1 \rightarrow i_x) \\ & + j_y^{i_x}(i_y + 1 \rightarrow i_y) + j_y^{i_x}(i_y - 1 \rightarrow i_y) \end{aligned} \quad (35)$$

and

$$j_x^{i_y}(i_x \rightarrow i_x + 1) = r_{i_x \rightarrow i_x+1}^{i_y} p_{i_x, i_y} - r_{i_x+1 \rightarrow i_x}^{i_y} p_{i_x+1, i_y}, \quad (36)$$

$$j_y^{i_x}(i_y \rightarrow i_y + 1) = l_{i_y \rightarrow i_y+1}^{i_x} p_{i_x, i_y} - l_{i_y+1 \rightarrow i_y}^{i_x} p_{i_x, i_y+1}. \quad (37)$$

The mappings between the probability currents in the continuous space and those on the discrete lattice read

$$j_x(x, y, t) = \lim_{\Delta_x \rightarrow 0, \Delta_y \rightarrow 0} \frac{j_x^{i_y}(i_x \rightarrow i_x + 1, t)}{\Delta_y}, \quad (38)$$

$$j_y(x, y, t) = \lim_{\Delta_x \rightarrow 0, \Delta_y \rightarrow 0} \frac{j_y^{i_x}(i_y \rightarrow i_y + 1, t)}{\Delta_x}, \quad (39)$$

where  $x = x_- + \Delta_x i_x$  and  $y = y_- + \Delta_y i_y$ . The appearance of the factors  $\Delta_x$  and  $\Delta_y$  follows from discretization of the formula  $\partial_t \rho = -\nabla \cdot \mathbf{j} = \dot{p}_{i_x, i_y} / \Delta_x \Delta_y = \sum j / \Delta_x \Delta_y$ , valid in the limit of infinitely fine mesh, where  $\sum j$  stands for right-hand side of Eq. (35).

*Microscopically*, the current can be defined as  $\mathbf{j}(x, y, t) = \langle \delta[\mathbf{r}(t) - \mathbf{r}] \dot{\mathbf{r}}(t) \rangle = \langle \delta[x(t) - x] \delta[y(t) - y] \dot{\mathbf{r}}(t) \rangle$ , where the average is taken over many trajectories  $\mathbf{r}(t)$  of the underlying stochastic process. The quantity

$$\tilde{\mathfrak{J}}(x, y, t) = \tilde{\mathfrak{J}}(\mathbf{r}, t) = \delta[\mathbf{r}(t) - \mathbf{r}] \dot{\mathbf{r}}(t) \quad (40)$$

inside the average is what we call a *microscopic* current. In measurements, one can obtain not only the average current  $\mathbf{j}$ , but its full probability distribution. The MNM can be applied to investigate this distribution as well as other distributions of arbitrary variables that arise as linear combinations of the microscopic currents  $\tilde{\mathfrak{J}}(x, y)$  at different positions. An important example of such a variable from the field of stochastic thermodynamics is heat, as exemplified in the example in Sec. VII.

The lattice equivalents of the microscopic definitions of the mesoscopic currents are the formulas  $j_x^{i_y}(i_x \rightarrow i_x + 1) = \langle \frac{di_x(t)}{dt} \delta_{i_x(t), i_x} \delta_{i_y(t), i_y} \rangle$  and  $j_y^{i_x}(i_y \rightarrow i_y + 1) = \langle \frac{di_y(t)}{dt} \delta_{i_x(t), i_x} \delta_{i_y(t), i_y} \rangle$ . The  $x$  current measures the number of jumps to the right from the lattice point minus the number of jumps from the right to the lattice point, and similarly for the  $y$  current.

### B. Moment-generating functions for observables proportional to integrated currents

In this section, we calculate the moment-generating function  $\chi_A$  for an observable which is given by an arbitrary linear combination of the microscopic currents (40)

$$\begin{aligned} A(t_0 + \tau, t_0) &= \int_{t_0}^{t_0+\tau} dt \int dx \int dy \mathbf{c}(\mathbf{r}, t) \cdot \tilde{\mathfrak{J}}(\mathbf{r}, t) \\ &= \int_{t_0}^{t_0+\tau} dt \mathbf{c}[\mathbf{r}(t), t] \cdot \dot{\mathbf{r}}(t), \end{aligned} \quad (41)$$

where  $\mathbf{c}(\mathbf{r}, t) = (\partial_x g, \partial_y h)$  is a vector of space- and time-dependent coefficients. The MGF  $\chi_A = \int_{-\infty}^{\infty} dA \exp(-s_A A) p(A)$  is defined as a two-sided Laplace transform of the probability distribution  $p(A)$ .

In Appendix B, we discuss in detail the computation of the MGF  $\chi_{\tilde{\mathfrak{J}}}$  for the time-averaged probability current  $\tilde{\mathfrak{J}}(\mathbf{r}, \tau) = \frac{1}{\tau} \int_{t_0}^{t_0+\tau} dt \tilde{\mathfrak{J}}(\mathbf{r}, t)$ . The MGF  $\chi_A$  can be computed along similar

lines as  $\chi_{\tilde{\mathfrak{J}}}$  and thus we here omit the details and present the main results only.

The key ingredient in the computation of the MGF is the construction of the so-called tilted matrix  $\tilde{\mathcal{R}}_{s_A}(t)$ . In the present case, the rate matrix  $\mathcal{R}(t)$  must be tilted proportionally to the coefficients  $\partial_x g$  and  $\partial_y h$  in the vector  $\mathbf{c}(\mathbf{r}, t)$ . Namely, the rates  $r_{i_x \rightarrow i_x+1}^{i_y}(t)$  must be multiplied by

$$\exp\{-s_A [g(x_- + (i_x + 1)\Delta_x, t) - g(x_- + i_x \Delta_x, t)]\}, \quad (42)$$

the rates  $r_{i_y+1 \rightarrow i_y}^{i_x}(t)$  by

$$\exp\{-s_A [h(y_- + i_y \Delta_y, t) - h(y_- + (i_y + 1)\Delta_y, t)]\}, \quad (43)$$

and similarly for all other transition rates. The MGF for  $A(t_0 + \tau, t_0)$  is then obtained from Eq. (B8) with the only difference that the tilted matrices  $\tilde{\mathcal{R}}_{s_A}(t)$  involved in the equation are substituted by the tilted matrices  $\tilde{\mathcal{R}}_{s_A}(t)$  just described above. Namely,

$$\chi_A(s_A, t, t_0) = \lim_{\Delta_t \rightarrow 0} \mathbf{p}_+^\top \prod_{i=0}^{i_i(t)} \tilde{\mathcal{U}}_i(s_A) \mathbf{p}(t_0), \quad (44)$$

where  $\mathbf{p}_+^\top$  is a vector of ones effecting the summation over the final states at time  $t = t_0 + \tau$ , and  $\tilde{\mathcal{U}}_i(s_A) = \exp[\tilde{\mathcal{R}}_{s_A}(t_0 + i\Delta_t)\Delta_t]$  if  $i > 1$  and the unit matrix  $\mathcal{I}$  otherwise. For problems with a time-independent tilted rate matrix  $\tilde{\mathcal{R}}_{s_A}$ , the product in Eq. (44) simplifies to  $\prod_{i=0}^{i_i(t)} \tilde{\mathcal{U}}_i(s_A) = \exp[\tilde{\mathcal{R}}_{s_A} \tau] = \tilde{\mathcal{U}}(s_A, t_0 + \tau, t_0)$  and the moment-generating function is thus given by

$$\chi_A(s_A, t, t_0) = \mathbf{p}_+^\top \tilde{\mathcal{U}}(s_A, t, t_0) \mathbf{p}(t_0). \quad (45)$$

Some examples of physically relevant observables of the type (41) are time-averaged probability currents  $j_x = \int dt \int dx \int dy \tilde{\mathfrak{J}}_x(x, y, t) / \tau$  flowing through the system in the  $x$  direction [here  $\mathbf{c}(x, y, t) = (1, 0) / \tau$ ]; time-averaged probability currents  $j_y = \int dt \int dx \int dy \tilde{\mathfrak{J}}_y(x, y, t) / \tau$  flowing through the system in the  $y$  direction [here  $\mathbf{c}(x, y, t) = (0, 1) / \tau$ ]; the total heat flux  $Q = \int dt \int dx \int dy \nabla U(x, y, t) \cdot \tilde{\mathfrak{J}}(x, y, t) / \tau$  flowing from the reservoirs into a Brownian ratchet [16,17] [here  $\mathbf{c}(x, y, t) = \nabla U(x, t) / \tau$ , where  $U(x, y)$  is a potential energy]; and the heat flux  $Q_x = \int dt \int dx \int dy \partial_x U(x, y, t) \cdot \tilde{\mathfrak{J}}(x, y, t) / \tau$  flowing into the ratchet from the reservoirs connected to the  $x$  coordinate only [here  $\mathbf{c}(x, y, t) = (\partial_x U(x, t), 0) / \tau$ ].

For the observables  $A$  where the scalar product  $\mathbf{c}(\mathbf{r}, t) \cdot \tilde{\mathfrak{J}}(\mathbf{r}, t)$  in Eq. (41) can be written in the form of a total time derivative  $df[x(t), y(t), t] / dt = \partial f / \partial t + \nabla f \cdot \dot{\mathbf{r}}$ , the formula (41) can be simplified as

$$\begin{aligned} A(t_0 + \tau, t_0) &= \int_{t_0}^{t_0+\tau} dt \frac{d}{dt} f[\mathbf{r}(t), t] \\ &= f[\mathbf{r}(t_0 + \tau), t_0 + \tau] - f[\mathbf{r}(t_0), t_0] \end{aligned} \quad (46)$$

and thus depends only on the initial and final times and positions. Also in this case, the calculation of the MGF for  $A$  can be simplified as in the step from Eq. (44) to Eq. (45). Now, the matrix  $\tilde{\mathcal{U}}(s_A, t_0 + \tau, t_0) = \tilde{\mathcal{U}}(s_A)$  has elements

$$[\tilde{\mathcal{U}}(s_A)]_{kl} = [\mathcal{U}(t_0 + \tau, t_0)]_{kl} e^{-s_A \Delta_c(k, l, t_0 + \tau, t_0)}, \quad (47)$$



where  $\Delta_c(k, l, t_0 + \tau, t_0) = f[\mathbf{r}_f, t_0 + \tau] - f[\mathbf{r}_i, t_0]$ ,  $\mathbf{r}_f = [x_- + \Delta_x i_x(k), y_- + \Delta_y i_y(k)]$ , and  $\mathbf{r}_i = [x_- + \Delta_x i_x(l), y_- + \Delta_y i_y(l)]$ . Here, the coefficients  $i_x(k)$  and  $i_y(k)$  are given by Eqs. (28) and (29). A typical example of such an observable is the above-mentioned heat in case the potential  $U(x, t)$  does not depend on  $t$  explicitly. However, since we treat time-dependent protocols using a piecewise constant approximation (see Sec. IV B), this simplification is important also for time-dependent potentials. If the product  $\mathbf{c}(\mathbf{r}, t) \cdot \mathfrak{J}(\mathbf{r}, t)$  can be written as a total derivative  $df/dt$  only for a time-independent vector  $\mathbf{c}(\mathbf{r}, t) = \mathbf{c}(\mathbf{r})$ , the moment-generating function for  $A$  with the explicitly time-dependent  $\mathbf{c}(\mathbf{r}, t)$  can be calculated from Eq. (44) with  $\tilde{\mathcal{U}}_i(s_A)$ ,  $i > 1$ , redefined using Eq. (47) as  $\tilde{\mathcal{U}}[s_A, t_0 + (i+1)\Delta_t, t_0 + i\Delta_t]$ .

### C. Moment-generating functions for observables not proportional to integrated currents

Above, we have focused solely on observables which can be written as linear combinations (41) of microscopic probability currents. The integrand in these observables vanishes if the particle does not move ( $\dot{\mathbf{r}} = 0$ ). However, in driven systems, there are also important observables with nonzero increments even if the particle stands still. The MNM can also be used to calculate MGFs and LDFs for observables of the form

$$\begin{aligned} B(t_0 + \tau, t_0) &= \int_{t_0}^{t_0 + \tau} dt \int dx \int dy \delta[\mathbf{r}(t) - \mathbf{r}] \partial_t b(\mathbf{r}, t) \\ &= \int_{t_0}^{t_0 + \tau} dt \partial_t b[\mathbf{r}(t), t]. \end{aligned} \quad (48)$$

The observable  $B$  vanishes if the function  $b$  is constant in time. The best-known example of a physically relevant observable of the type (48) is the stochastic work done on the system due to a deterministic external driving, which changes the potential  $U$ . Then we have  $b[\mathbf{r}(t), t] = b[x(t), y(t), t] = U(x(t), y(t), t)$ . Another example is the occupation time for a position  $\mathbf{r}_a$ , in which case  $b(\mathbf{r}(t), t) = \delta[\mathbf{r}(t) - \mathbf{r}_a]t$ , or the occupation time for a region  $\Omega$ , in which case  $b(\mathbf{r}(t), t) = I_\Omega[\mathbf{r}(t)]t$ , where  $I_\Omega(\mathbf{r})$  is an indicator function equal to one if  $\mathbf{r} \in \Omega$  and 0 otherwise.

For observables of the above type  $B$ , the tilted matrix must be constructed using the time discretization, already introduced to derive Eq. (33). We define the piecewise constant approximation of the function  $b$  as  $\bar{b}(\mathbf{r}, t) = b(\mathbf{r}, t_0 + \Delta_t i_t)$ ,  $i_t = \lfloor (t - t_0)/\Delta_t \rfloor$ . For this approximate function, the variable  $B$  in Eq. (48) does not change during the time intervals  $[t_0 + \Delta_t i_t, t_0 + \Delta_t(i+1)]$ , where  $\bar{b}(\mathbf{r}, t)$  is constant for constant  $\mathbf{r}$ , and it abruptly jumps from  $B(t)$  to  $B(t) + b[\mathbf{r}(t), t+] - b[\mathbf{r}(t), t-]$  at time instants  $t = t_0 + \Delta_t i$ , where  $\bar{b}(\mathbf{r}, t)$  changes infinitely fast. Here  $b[\mathbf{r}(t), t \pm] = \lim_{\epsilon \rightarrow 0} b[\mathbf{r}(t), t \pm \epsilon]$ ,  $\epsilon > 0$ .

Let us now turn to the discrete approximation of the full process using the discrete lattice of Fig. 1. Using the notation of Eq. (33) and assuming that the system is in microstate  $[i_x(l), i_y(l)]$  at time  $t_0 + \Delta_t i$  and in microstate  $[i_x(k), i_y(k)]$  at time  $t_0 + \Delta_t(i+1)$  [see Eqs. (28) and (29) for definitions of  $i_x(l)$  and  $i_y(l)$ ], the PDF for  $B$  is

given by

$$[\tilde{\mathcal{U}}_i(B)]_{kl} = [\mathcal{U}_i]_{kl} \delta[B - \Delta_b(k, t_0 + \Delta_t i)], \quad i \geq 1. \quad (49)$$

Here we used the shorthand  $\Delta_b(k, t) = b[\mathbf{r}, t+] - b[\mathbf{r}, t-]$ ,  $\mathbf{r} = [x_- + \Delta_x i_x(k), y_- + \Delta_y i_y(k)]$  and  $\mathcal{U}_0(B) = \mathcal{I}$ . The PDF for  $B$  during the whole time interval  $[t_0, t_0 + \tau]$  is thus given by a multiple convolution of the form  $\lim_{\Delta_t \rightarrow 0} \mathbf{p}_+^\top [\tilde{\mathcal{U}}_{i(t)} \star \tilde{\mathcal{U}}_{i(t-1)} \star \cdots \star \tilde{\mathcal{U}}_0](B) \mathbf{p}(t_0)$ . The MGF for  $B$  and thus also the corresponding tilted matrix is obtained by the Laplace transform of the last expression with respect to  $B$ :

$$\chi_B(s_B, t_0 + \tau, t_0) = \lim_{\Delta_t \rightarrow 0} \mathbf{p}_+^\top \prod_{i=0}^{i(t_0 + \tau)} \tilde{\mathcal{U}}_i(s_B) \mathbf{p}(t_0), \quad (50)$$

where the matrix  $\tilde{\mathcal{U}}_i(s_B)$  is obtained as the Laplace transform of the matrix  $\tilde{\mathcal{U}}_i(B)$  (we again just substitute the  $\delta$  functions  $\delta[B - \Delta_b(k, t_0 + \Delta_t i)]$  for exponentials  $\exp[-s_B \Delta_b(k, t_0 + \Delta_t i)]$ ).

The MNM can also be applied to variables which are defined as combinations of the variables of the types  $A$  and  $B$ . An example of such a variable is the increase of internal energy  $\Delta U = W + Q$ , which consists of heat  $Q$  (type  $A$  variable) and work  $W$  (type  $B$  variable). Let us consider a general variable  $C$  decomposed as  $C = A + B$ . Then the corresponding MGF  $\chi_C$  is given by Eq. (50) with the tilted matrices  $\tilde{\mathcal{U}}_i(s_C)$  given by

$$[\tilde{\mathcal{U}}_i(s_C)] = [\tilde{\mathcal{B}}_i(s_C)]_{kl} \exp[-s_C \Delta_b(k, t_0 + \Delta_t i)], \quad (51)$$

where  $\tilde{\mathcal{B}}_i(s_C)$  is the tilted matrix  $\tilde{\mathcal{U}}_i$  for  $A$ , defined in Eq. (44). Similarly to the case of the variables of type  $A$ , also the computation of  $\chi_C$  may simplify if the variable  $C$  has a suitable structure.

### D. Moments and cumulants

The MGF  $\chi_X(s, t, t_0)$  allows one to access all moments of the stochastic variable  $X$  at time  $t$  simply by taking derivatives with respect to the Laplace variable  $s$ :

$$\langle X^n(t) \rangle = (-1)^n \left. \frac{d^n \chi_X(s, t, t_0)}{ds^n} \right|_{s=0}. \quad (52)$$

The zeroth moment is just a normalization  $\chi_X(0, t, t_0) = 1$  and it can be used as a first test of the calculated MGF. The first moment equals the average  $\langle X(t) \rangle$  of the quantity  $X$  and it can be calculated from the probability distribution for position  $\rho(x, y, t)$  [or from its approximation  $\mathbf{p}(t)$ ]. For the variable  $A$  defined in Eq. (41), it reads

$$\langle A(t) \rangle = \int_{t_0}^t dt' \int dx \int dy \mathbf{c}(x, y, t') \cdot \mathbf{j}(x, y, t'), \quad (53)$$

where the average current  $\mathbf{j}(x, y, t)$  is given either by Eq. (34) or by Eqs. (38) and (39). For the variable  $B$  defined in Eq. (48), we get

$$\langle B(t) \rangle = \int_{t_0}^t dt' \int dx \int dy \partial_t b[x, y, t'] \rho(x, y, t'). \quad (54)$$

The formulas (53) and (54) can be used as another test of calculated MGFs.

In a manner similar to moments, the MGF can be used for calculating all cumulants  $C_n(X, t)$  of the variable  $X$  at time  $t$ :

$$C_n(X, t) = (-1)^n \frac{d^n \log \chi_X(s, t, t_0)}{ds^n} \Big|_{s=0}. \quad (55)$$

The cumulants reflect the shape of the probability distribution for  $X$ . First, four of them can be written in terms of moment as  $C_0 = 0$ ,  $C_1 = \langle X \rangle$ ,  $C_2 = \langle X^2 \rangle - \langle X \rangle^2$  and  $C_3 = \langle X^3 \rangle - 3\langle X^2 \rangle \langle X \rangle + 2\langle X \rangle^3$  and thus for a centered random variable with  $\langle X \rangle = 0$  the first three cumulants are equal to the first three moments. In general, moments and cumulants can be related by the recursion relation

$$C_n(X, t) = \langle X^n \rangle - \sum_{m=1}^{n-1} \binom{n-1}{m-1} C_m \langle X^{n-m} \rangle. \quad (56)$$

The numerical computation of the derivatives in Eqs. (53) and (55) may lead to various problems, especially at higher orders. Alternatively, the moments and cumulants can be calculated via the derivative-free method introduced in Ref. [27].

Although the moments and cumulants provide rich information about the PDF for  $X$ , to reconstruct the whole distribution requires knowledge of all the moments and/or cumulants and is thus rarely achievable in practice. For long times  $\tau$ , however, a very general method for calculating the (approximate) PDF from the MGF can be applied. This method is based on the so-called large deviation theory.

### E. Large deviation functions

If the time domain  $\tau$  of the time integrals in Eqs. (41) and (48) gets very large, the PDFs  $\rho(X, t_0 + \tau, t_0) = \rho(X, \tau)$ ,  $X = A, B$  can assume the so-called large-deviation form [28]

$$\log \rho(X, \tau) \sim \tau J\left(\frac{X}{\tau}\right), \quad (57)$$

where the function  $J(x) \leq 0$  is the large deviation function. The symbol  $\sim$  means that Eq. (57) is an asymptotic representation of  $\log \rho(X, t_0 + \tau, t_0)$  valid for large times  $\tau$ , where the terms omitted in the formula are typically proportional to  $\log \tau$ .

The large deviation function can be calculated from the MGF by Laplace's method. Namely, assuming that  $\tau$  is large and Eq. (57) holds, the MGF can be written as

$$\begin{aligned} \log \chi(s_X, \tau) &= \log \int dX e^{-s_X X} \rho(X, \tau) \\ &\approx \log \int dX e^{-\tau[s_X X/\tau - J(X/\tau)]} \\ &\approx \tau \max_x [J(x) - s_X x]. \end{aligned} \quad (58)$$

The large deviation function  $J(x)$  can hence be calculated by a Legendre–Fenchel transform

$$J(x) = \min_{s_X} [\lambda(s_X) + s_X x], \quad (59)$$

where

$$\lambda(s_X) = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \log \chi(s_X, \tau) \quad (60)$$

denotes the so-called scaled cumulant-generating function. Here, we assume that the scaled cumulant-generating function is differentiable. Otherwise, the formula (59) does not universally hold, and one has to resort to a more involved procedure for calculation of the LDF, if it exists at all [28].

For problems with time-independent coefficients and the moment-generating function determined by Eq. (45) with the tilted Green's function given by  $\tilde{U}(s_X, t_0 + \tau, t_0) = [\tilde{\mathcal{R}}_{s_X} \tau]$ , the scaled cumulant-generating function (45) can be calculated as

$$\begin{aligned} \frac{1}{\tau} \log \chi(s_X, \tau) &= \frac{1}{\tau} \log [\mathbf{p}_+^\top \exp(\tilde{\mathcal{R}}_{s_X} \tau) \mathbf{p}(t_0)] \\ &= \frac{1}{\tau} \log \left\{ \sum_i c_i(s_X) \exp[\tau \lambda_i(s_X)] \right\} \\ &\approx \lambda_{\max}(s_X). \end{aligned} \quad (61)$$

In the calculation, we used the eigenvalue decomposition of the matrix  $\tilde{\mathcal{R}}_{s_X}$  which allowed us to rewrite the product  $\mathbf{p}_+^\top \exp(\tilde{\mathcal{R}}_{s_X} \tau) \mathbf{p}(t_0)$  using the coefficients  $c_i$  arising from products of the vectors  $\mathbf{p}_+^\top$ ,  $\mathbf{p}(t_0)$  and eigenvectors of the matrix  $\tilde{\mathcal{R}}_{s_X}$ . In the final step, we took the limit  $\tau \rightarrow \infty$  in which the sum is dominated by its largest term  $c_{\max} \exp(\tau \lambda_{\max})$  corresponding to the largest eigenvalue  $\lambda_{\max}$ . In short, the LDF  $J(a)$  is in this case determined by the largest eigenvalue  $\lambda_{\max}(s_X)$  of the tilted rate matrix  $\tilde{\mathcal{R}}_{s_X}(s_X)$  as

$$J(x) = \min_{s_X} [\lambda_{\max}(s_X) + s_X x]. \quad (62)$$

For problems with time-dependent coefficients, where the moment-generating function is determined by the product form (44) or (50), the large deviation principle (57) does not generally hold, unless the time dependence is periodic and we are interested in the PDF for the stochastic variable attained after many cycles  $N$  [29]. For a single cycle starting at  $t_0$  and ending at  $t_0 + t_c$ , where  $t_c$  denotes the duration of a single period, the moment-generating functions (44) and (50) are then determined by the propagator

$$\tilde{U}(s) = \lim_{\Delta_t \rightarrow 0} \prod_{i=0}^{i(t_0+t_c)} \tilde{U}_i(s_X). \quad (63)$$

Hence, the moment-generating function  $\chi(s_{X_N})$  for the variable  $X_N = X(t_0 + \tau, t_0)$ ,  $\tau = N t_c$  [see Eqs. (41) and (48)], is given by  $\chi(s_{X_N}) = \mathbf{p}_+^\top \tilde{U}(s_{X_N})^N \mathbf{p}(t_0)$ . A calculation similar to the one in Eq. (61) leads to the scaled cumulant-generating function for  $X_N$ :

$$\begin{aligned} \frac{1}{\tau} \log \chi(s_{X_N}) &= \frac{1}{\tau} \log [\mathbf{p}_+^\top \tilde{U}(s_{X_N})^N \mathbf{p}(t_0)] \\ &= \frac{1}{\tau} \log \left\{ \sum_i c_i(s_{X_N}) [\alpha_i(s_{X_N})]^N \right\} \\ &\approx \frac{1}{t_c} \log \alpha_{\max}(s_{X_N}). \end{aligned} \quad (64)$$

Here,  $\alpha_i(s_{X_N})$  denote eigenvalues of the propagator for a single cycle  $\tilde{U}(s_{X_N})$  and the coefficients  $c_i$  arise from the products of the vectors  $\mathbf{p}_+^\top$ ,  $\mathbf{p}(t_0)$  and eigenvectors of the matrix  $\tilde{U}(s_{X_N})$ . Now, the LDF  $J(a)$  is determined by the logarithm of the

largest eigenvalue  $\alpha_{\max}(s_{X_N})$  of the matrix  $\tilde{U}(s_{X_N})$  as

$$J(x) = \min_{s_{X_N}} \left[ \frac{1}{t_c} \log \alpha_{\max}(s_{X_N}) + s_{X_N} x \right]. \quad (65)$$

In the fully solved example given in Sec. VII, we compute this function for a simple model stochastic heat engine.

## VI. DISCRETIZATION AND EFFICIENCY

There are several ways how to determine suitable discretization parameters  $N_x$ ,  $N_y$ , and  $N_t$  and the boundaries  $x_{\pm}$  and  $y_{\pm}$  without knowing the exact solution. In general, if not fixed by the physics of the problem in question, these parameters should be chosen in such a way that their further refining affects the computed results only negligibly. A second way of choosing the discretization mesh, pursued in the example below, is to compare the numerical results with results obtained using Brownian dynamics (BD) simulations of the stochastic process described by the FPE (1). Then the mesh can be refined until both methods give the same results.

For a given discretization, the efficiency (defined as precision of calculation over the computation time) of the MNM is comparable to standard numerical methods based on substituting finite differences for partial derivatives in the FPE (1) such as the one described in Ref. [20]. It can be increased by adapting the discretization mesh to the salient features of the time-dependent driving, i.e., by putting the time-discretization parameter  $\Delta_t$  roughly inversely proportional to the first derivative of the driving (with some fixed upper bound) and similarly for  $\Delta_x$  and  $\Delta_y$ .

Main merits of the MNM are the following: (1) Versatility—similar implementations can be used for calculating probability distributions, moment-generating functions, and large deviation functions, both for time-independent and time-dependent problems. (2) Easy implementation—it is enough to construct the transition rate matrix using the expressions (13) and (14) or expressions (18) and (19), and the rest can be handled using matrix operations, which are usually well implemented in the current programming languages used in physics. (3) Thermodynamic consistency—qualitatively reasonable predictions of the system dynamics and thermodynamics are obtained with very coarse meshes, as soon as these meshes capture all qualitative features of the forces, potentials, and their time dependence. These coarse meshes can thus be used to find interesting effects for a given problem quickly, and thus to reserve time-consuming precise computations for the fraction of model parameters giving the most interesting results. As an example, we refer to Ref. [17], where all key effects occurring in a complex model of a two-dimensional continuous system were captured by a simple discrete six-level system.

The main limitation of the method concerns its generalization to higher dimensional problems. Namely, the available RAM determines the largest matrix that can swiftly be handled by the computer. The rate matrix  $\mathcal{R}$  in Eq. (26) has at most  $\prod_{i=1}^d (N_i + 1)(1 + 2d)$  nonzero elements, where  $d$  denotes the dimensionality of the problem and  $N_i + 1$  denotes the number of discrete points considered for the  $i$ th dimension. This is because each site in Fig. 2 is connected to

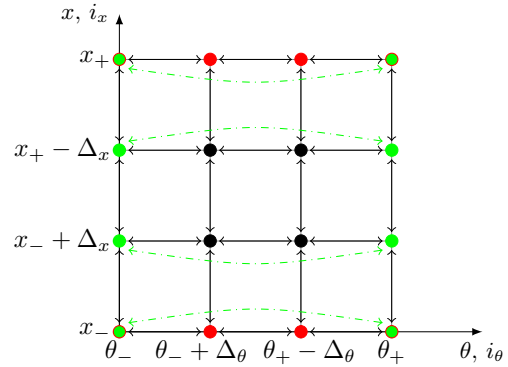


FIG. 2. Sketch of the phase-space discretization used for the numerical solution of the two-dimensional overdamped Fokker-Planck equation (1) in case of the driven active particle (Sec. VII). The meaning of the arrows and point colors is the same as in Fig. 1.

at most  $2d$  neighbors and each of the  $\prod_{i=1}^d (N_i + 1)$  rows of  $\mathcal{R}$  thus contains  $2d$  rates for transitions into the given site and 1 outward rate. On the other hand, the propagators  $\mathcal{U}(t, t_0)$  (30) already contain  $\prod_{i=1}^d (N_i + 1)^2$  nonzero matrix elements. The largest number of nonzero elements which can be handled by our computer (8 GB RAM) is approximately  $10^6$ . In practice, problems that can be solved solely using the rate matrix  $\mathcal{R}$ , such as the computation of a (nonequilibrium) stationary solution of Eq. (26), can usually be attacked with acceptable precision in higher dimensions, whereas fully time-dependent problems require additional resources.

## VII. EXAMPLE: DRIVEN ACTIVE PARTICLE

An example of a typical application of the MNM can be found in Refs. [16,17], investigating a two-dimensional Brownian ratchet in contact with two reservoirs at different constant temperatures. In this case, the authors used periodic and reflecting boundary conditions. Another example of usage of the MNM is the work [18], where the MNM was used to calculate probability distributions of a particle surviving in a constant unstable cubic potential. In this case, the authors implemented absorbing and reflecting boundary conditions.

In the present section, we consider a FPE with time-dependent coefficients and show that the MNM can be used both for describing the dynamics of the probability distribution and for evaluating MGFs and LDFs of stochastic functionals of the underlying stochastic process. For the sake of simplicity, all physical quantities in this section are represented in suitable natural units that render them dimensionless.

We consider an active particle self-propelling with a velocity of magnitude  $v(t) \cos \theta(t)$  and driven by a time-dependent quartic potential

$$U(x, t) = k(t)x^4/4 \quad (66)$$

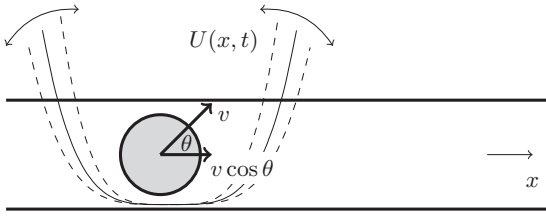


FIG. 3. Active particle confined to a single dimension and driven by the quartic potential (66) of time-dependent strength  $k(t)$ .

in the  $x$  direction, as shown in Fig. 3. We assume that the particle motion is overdamped and thus its position  $x(t)$  and orientation  $\theta(t)$  obey the first-order Langevin equations

$$\dot{x} = -kx^3 + v \cos \theta + \sqrt{2D_x} \eta_x, \quad (67)$$

$$\dot{\theta} = \sqrt{2D_\theta} \eta_\theta. \quad (68)$$

Here,  $\eta_x$  and  $\eta_\theta$  denote independent, zero-mean unit-variance Gaussian white noises. If we denote the angular variable  $\theta$  as  $y$ , the system of (67) and (68) corresponds to the FPEs (1) and (11) with  $\mu_x F_x = -kx^3 + v \cos \theta$ ,  $\mu_y F_y = 0$ ,  $D_x$ , and  $D_y = D_\theta$ , i.e.,

$$\partial_t \rho = [D_x \partial_x^2 + D_\theta \partial_\theta^2 - k \partial_x x^3 + v \cos \theta \partial_x] \rho, \quad (69)$$

where  $\rho = \rho(x, \theta, t)$ . Such schematic models of active particles are often considered as idealized caricatures of artificial or biological microswimmers [30–32]. In fact, they have acquired the status of a major new paradigmatic model of nonequilibrium statistical mechanics. While currently most studies resort to simulations when analytical approximations cease to work [12], the MNM could in the future provide a welcome alternative approach. To illustrate its application to the above model, we consider a specific nonequilibrium situation that is of interest for its own sake. Namely, motivated by recent studies interpreting trapped Brownian particles as microscopic heat engines [11, 14, 33–38], we choose the potential stiffness  $k(t)$ , the particle active velocity  $v(t)$ , and the diffusion coefficients  $D_x(t)$  and  $D_\theta(t)$  to be 1-periodic functions, as depicted in Fig. 4. This choice of parameters leads to a positive net work produced by the system per period.

To understand the thermodynamics of the system, it is helpful to first assume that the particle is not active ( $v = 0$ ) and can thus be understood as a system coupled only to a single bath with time-dependent temperature  $D_x(t)$ . During some parts of the cycle, the heat flows into the bath, and during others it flows from the bath to the system. The reservoir with a time-dependent temperature thus serves as a heat source during some parts of the cycle and as a heat sink during the rest of the cycle. Alternatively, one can understand this setup in such a way that there are many reservoirs at different temperatures and the system is at each time connected to one of them. In such a case, we would have many heat sources and many heat sinks. In both cases, the laws of thermodynamics allow us to transform heat into work and to operate the system as a heat engine. More details for heat engines of this type can be found in Refs. [11, 39]. If the particle is active, the basic principle of the engine operation is the same as described

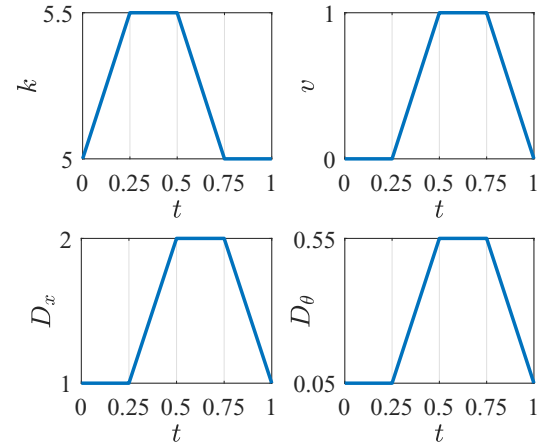


FIG. 4. The parameters of the microscopic heat engine consisting of the periodically driven active particle depicted in Fig. 3, during one period of the cyclic driving protocol. Boxes comprise maximum and minimum values of the corresponding variables during the cycle.

above; nevertheless, there are some significant differences. Most importantly, the source of the disordered energy (called heat) is now not only the heat bath itself, but also the active self-propulsion of the particle. For more details, we refer the interested reader to Refs. [35, 40].

### A. Dynamics

To compute the dynamical and statistical properties of the heat engine using the MNM, we consider the discretization depicted in Fig. 2 with  $\theta_- = 0$ ,  $\theta_+ = 2\pi - \Delta_\theta$ ,  $\Delta_\theta = 2\pi/(N_\theta + 1)$ , and  $x_+ = -x_-$ ,  $\Delta_x = (x_+ - x_-)/N_x$ . The positions  $x_\pm$  of the  $x$  boundaries of the discrete mesh, where we impose reflecting boundary conditions, are chosen in such a way that the probabilities at the boundaries turn out to be negligible. The discretization parameters  $N_\theta$  and  $N_x$  are chosen in such a way that their further refinement would barely affect the solution.

On the discrete lattice, we determine the matrix  $\mathcal{U}(t, 0)$  in Eq. (33), which represents the approximate Green's function for the FPE (69) of the model, during one driving cycle. For an arbitrary initial condition  $\mathbf{p}_0 = \mathbf{p}(0)$  at time 0, the matrix  $\mathcal{U}(t, 0)$  provides us with the distribution at time  $t$  as

$$\mathbf{p}(t) = \mathcal{U}(t - N, 0) [\mathcal{U}(1, 0)]^N \mathbf{p}_0, \quad (70)$$

where  $N = \lfloor t \rfloor$  is the number of full cycles done during the time interval  $(0, t)$ . After a transient relaxation period, the distribution  $\mathbf{p}(t)$  becomes independent of the initial condition. As a consequence of the periodicity of the driving, it converges to a 1-periodic vector in the long-time limit.

This time-dependent long-time solution  $\mathbf{p}_{lc}(t)$  of the master equation (FPE) with periodic transition rates is called the limit cycle. Using its periodicity, it can be determined using the eigenvector of the Green's function  $\mathcal{U}(t, 0)$  corresponding to the eigenvalue 1 as  $\mathbf{p}_{lc}(1) = \mathcal{U}(1, 0) \mathbf{p}_{lc}(0) = \mathbf{p}_{lc}(0)$ ,

$$\mathbf{p}_{lc}(t) = \mathcal{U}(t, 0) \mathbf{p}_{lc}(0). \quad (71)$$

From this approximate solution and the relation (6), we compute the approximate probability distribution  $\rho(x, \theta, t)$  of the

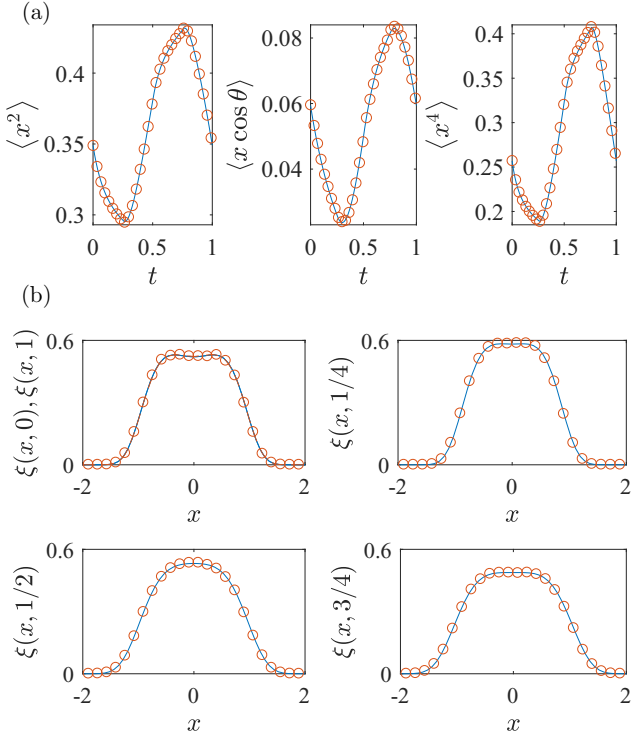


FIG. 5. Comparison of observables for the periodically driven active particle, depicted as a function of time during one limit cycle, as computed from BD simulations (symbols) and the MNM (lines): (a) averages  $\langle x^2 \rangle$ ,  $\langle x \cos \theta \rangle$ , and  $\langle x^4 \rangle$  and (b) marginal probability density  $\xi(x, t)$  for the particle position  $x$  at five time instants  $t$  during the cycle. The PDF  $\xi(x, 0)$  at the initial time 0 (dashed blue line) and  $\xi(x, 1)$  at the final time 1 (full orange line) in the first panel of (b) coincide, because the system operates in the limit cycle as described by Eq. (71).

active particle during the engine's operation. We use it to numerically compute the averages  $\langle x^2 \rangle$ ,  $\langle x \cos \theta \rangle$ , and  $\langle x^4 \rangle$  as functions of time and the marginal distribution for the  $x$  position  $\xi(x, t) = \int_0^{2\pi} d\theta \rho(x, \theta, t)$  at five time instants  $t = 0, 1/4, 1/2, 3/4, 1$ , during the limit cycle. We also independently evaluated these quantities using a BD simulation of the system of (67) and (68). The comparisons of the averages and the marginal distributions are shown in Figs. 5(a) and 5(b), respectively. The MNM results, depicted by full lines, perfectly overlap with those of the BD (symbols). The MNM results were calculated using the discretization parameters  $N_x = 51$ ,  $N_\theta = 21$ ,  $N_t = 76$ , and  $x_\infty = 2.4$ . Already for  $N_x = 31$ ,  $N_\theta = 15$ ,  $N_t = 76$ , and  $x_\infty = 2.4$ , one obtains curves that are visually indistinguishable from those depicted in Fig. 5, while the calculation is approximately  $10\times$  faster than with the finer mesh. For the BD, we generated  $10^6$  trajectories with the integration step  $10^{-3}$ .

Besides checking the correctness of our implementation of the MNM by BD, we have also tested our numerical results against analytical results available for the presented model in two limiting situations. Specifically, we tested that the computed PDF attains the form  $\rho(x, \theta, t) \propto \exp[-U(x, t)/T_{\text{eff}}(t)]$ ,  $T_{\text{eff}} = T + v^2/(2D_\theta)$  for a quasistatic driving, and

$D_\theta \gg 1$ . In this case, the particle rotates so fast that the term  $v \cos \theta$  in Eq. (67) becomes equivalent to a further white noise with the effective temperature  $v^2/(2D_\theta)$ . As a second benchmark, we considered quasistatic driving with  $D_\theta \rightarrow 0$ , where the active velocity can be treated as constant and thus  $\rho(x, \theta, t) \propto \exp\{-[U(x, t) - vx \cos \theta]/T\}$ .

### B. Moment-generating functions

Besides computing the distribution  $\rho(x, \theta, t)$  to evaluate averages, moments, and reduced distribution functions for  $x$  and  $\theta$ , the MNM can also be applied directly to other comprehensive representations of the stochastic thermodynamics encoded in the FPE. In the following, we apply the MNM to directly compute moment-generating functions (MGFs) and large-deviation functions (LDFs) of work and heat. From the point of view of stochastic thermodynamics, these MGFs and LDFs are of interest in studies of work fluctuations in microscopic heat engines operating close to the reversible efficiency [41–43] or of the fluctuating efficiency [36–38], both intensely investigated during the past few years.

In stochastic thermodynamics of externally driven systems, work and heat are usually defined from the first law of thermodynamics, as follows [13,14]. The energy  $U(x, t)$  of the particle in a fixed microstate  $(x, \theta)$  can change in the course of time in two fundamentally different ways, one called work  $w$  and the other heat  $q$ . Formally, we can write  $dU(x, t)/dt = \dot{w}(x, t) + \dot{q}(x, t)$ , where

$$\dot{w}[x(t), t] = \dot{w}(t) \equiv \dot{k}(t)x^4(t)/4, \quad (72)$$

$$\dot{q}[x(t), t] = \dot{q}(t) \equiv k(t)x^3(t)\dot{x}(t). \quad (73)$$

The work done on the particle per unit time,  $\dot{w}$ , is thus nonzero only if the potential is externally changed [ $\dot{k}(t) \neq 0$ ]. A heat exchange  $|\dot{q}| > 0$  occurs if the particle moves in the potential and either dissipates its kinetic energy or transforms energy acquired from the bath or from the active self-propulsion into potential energy. Since the considered particle is active, there is necessarily also some dissipated energy [mostly much larger than (73)] related to the self-propulsion mechanism. This energy is usually called housekeeping heat, and we neglect it here, treating it as an intrinsic property of the system.

Work and heat flowing to the particle during the time interval  $(0, \tau)$  are defined as integrals over the respective rates (72) and (73):

$$w(\tau) = \int_0^\tau dt \dot{w}(t) = \int_0^\tau dt \partial_t U[x(t), t], \quad (74)$$

$$q(\tau) = \int_0^\tau dt \nabla U[x(t), t] \cdot [x(t), \theta(t)] \\ = [U(x(\tau), \tau) - U(x(0), 0)] - w(\tau). \quad (75)$$

They correspond to the cumulative external work performed on the active particle by the device varying the confinement strength, and the cumulative heat transferred to it from the thermal reservoir at the time-dependent temperature  $T$ . Additionally, the energy gained due to the self-propulsion of the swimmer is counted as (“internal” or “active”) heat supply. The cumulative work is an example of a variable that is

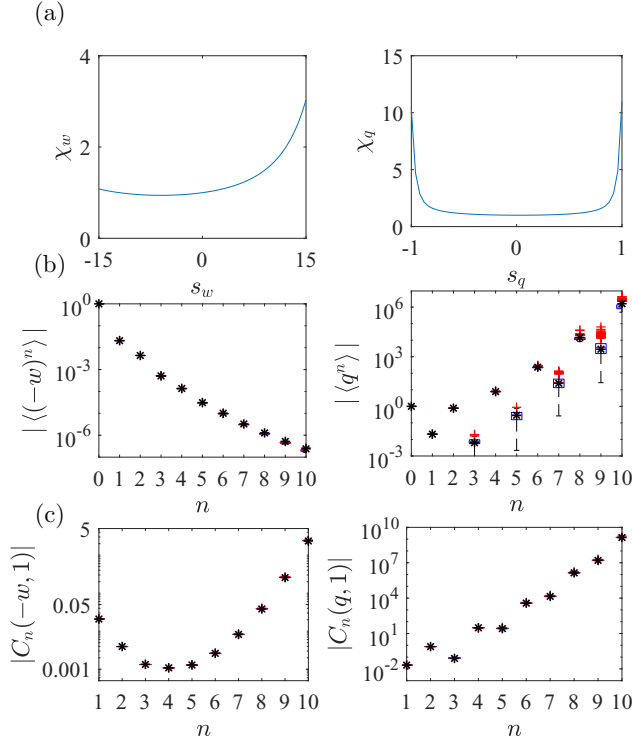


FIG. 6. Application of the MNM to moment-generating functions (MGFs): (a) the MGFs  $\chi_w$  and  $\chi_q$  for work and heat, (b) the first 11 raw moments, and (c) the first 10 cumulants [c] of the net extracted work  $-w$  and net supplied heat  $q$  per cycle, as calculated from the MGFs depicted in panel (a) (\*) and from corresponding BD simulations of  $200 \times 10^6$  trajectories (box plots).

not proportional to the probability current, with the function  $b[x(t), \theta(t), t]$  in Eq. (48) given by the instantaneous potential energy of the particle multiplied by the total time  $\tau$ ,  $b = \tau U[x(t), t]$ . The cumulative heat, on the other hand, is an example of an observable proportional to the current, with the vector  $\mathbf{c}(x, y, t)$  in Eq. (41) given by  $\tau \nabla U[x(t), t]$ .

Consider now the driving protocol depicted in Fig. 4 and the discretized time according to Sec. IV B. Using the formula (50) of Sec. VII B, we calculated the MGF  $\chi_w = \chi_w(s_w)$  for the work  $w(1)$  [Eq. (74)] transferred to the active particle during one limit cycle. The corresponding MGF  $\chi_q = \chi_q(s_q)$  for the heat  $q(1)$  [Eq. (75)] follows from formula (44) with tilted matrices  $\tilde{U}_i(s_q) = \exp[\tilde{\mathcal{R}}_{s_q}(t_0 + i\Delta_t)\Delta_t]$  if  $i \geq 1$  and  $\tilde{U}_0 = \mathcal{I}$  otherwise. For the parts of the piecewise constant protocol with time-independent potential, the tilted matrices can also be computed from the formula (47).

The resulting moment generating functions are shown in Fig. 6(a). The MGFs were sampled for  $s_w \in (-15, 15)$  with the step  $\Delta_{s_w} = 3/5$  for work and for  $s_q \in (-1, 1)$  with the step  $\Delta_{s_q} = 2/50$  for heat. To check the results, we computed the first 11 raw moments using the formula (52) and the first 10 cumulants using the formula (55). For the numerical evaluation of the derivatives in these equations, we used the

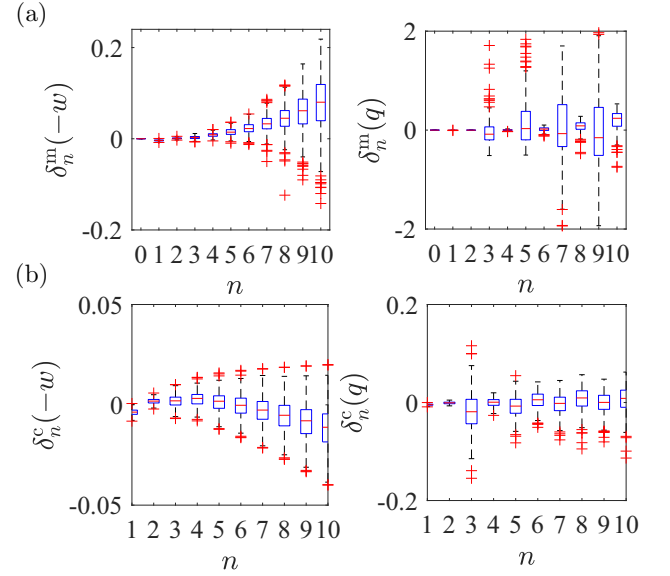


FIG. 7. Box plots of relative differences between moments (77) [panels (a)] and cumulants (78) [panels (b)] for work (left) and heat (right) as computed using MNM and BD, respectively [see Figs. 6(b) and 6(c)].

central difference scheme

$$\frac{d^n f(z)}{dz^n} \approx \sum_{i=0}^n (-1)^i \binom{n}{i} f\left[z + \left(\frac{n}{2} - i\right)\Delta_z\right], \quad (76)$$

where  $f$  is given by  $\chi_w$  for moments and cumulants of work and by  $\chi_q$  for moments and cumulants of heat. The parameters  $z$  and  $\Delta_z$  are given by  $s_w$  and  $2\Delta_{s_w}$  for  $\chi_w$  and  $s_q$  and  $2\Delta_{s_q}$  for  $\chi_q$ .

The resulting moments are depicted in Fig. 6(b) with an asterisk (\*) together with the corresponding results obtained from the BD simulations (depicted using box plots [44]). In order to assess the error of the latter, we simulated each moment 200 times using  $10^6$  trajectories yielding a box plot for each  $n$  in the figure. For the exchanged work, all data \* from the MNM and the corresponding box plots from BD perfectly superimpose so that the box plots are hardly visible, for all values of  $n$  [Fig. 6(b), left]. For heat, the results from both methods either coincide, or the MNM results lie within the boxes indicating the 25th and 75th percentiles of the BD data [Fig. 6(b), right]. The cumulants resulting from the MNM depicted in Fig. 6(c) (\*) together with the corresponding results obtained from the BD simulations (box plots) agree both for work [Fig. 6(c), left] and for heat [Fig. 6(c), right]. Note that the computation of cumulants from BD simulation is much less demanding than the computation of moments due to suppressed fluctuations.

To get better insight into the precision of these results, we show in Fig. 7 box plots of relative differences

$$\delta_n^m(x) = \frac{\langle x^n \rangle_a - \langle x^n \rangle_s}{\langle x^n \rangle_a + \langle x^n \rangle_s} \quad (77)$$

of computed and simulated moments for work [ $x = -w$ , Fig. 7(a), left] and heat [ $x = q$ , Fig. 7(a), right] and relative

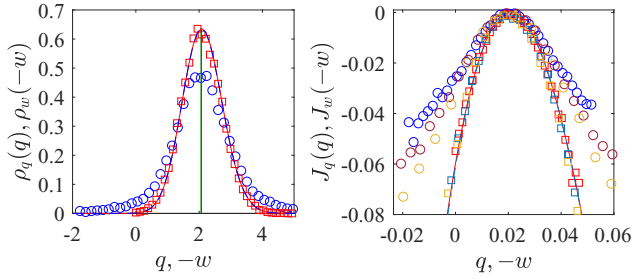


FIG. 8. Large-deviation limit of heat and work distributions. BD simulations of the cumulative distributions  $\rho_q$  and  $\rho_w$  of the net heat  $q$  supplied ( $\circ$ ) and the net work  $-w$  extracted ( $\square$ ) over 100 cycles show that the work distribution has converged to the common limiting form obtained from the MNM (superimposing lines, with the vertical line indicating the average), while the heat distribution has not (left panel). The Legendre-Fenchel transformed logarithms of the MGF of heat and work (right panel) elucidate the unequal convergence toward the common large-deviation function (LDF)  $J_q(q) \sim J_w(-w)$  with the number of cycles  $N = 30, 50, 100$ . While the work distribution ( $\square$ ) has already converged for  $N = 30$ , the heat distribution ( $\circ$ ) keeps evolving (top to bottom).

differences

$$\delta_n^s(x) = \frac{C_n^a(x, 1) - C_n^s(x, 1)}{C_n^a(x, 1) + C_n^s(x, 1)}, \quad (78)$$

of computed and simulated cumulants of work [ $x = -w$ , Fig. 7(b), left] and heat [ $x = q$ , Fig. 7(b), right]. The relative differences for work increase with  $n$  showing a trend toward positive relative differences for moments and negative relative differences for cumulants. These trends are caused by the chosen discretization. For heat, the data from BD are much more noisy than those for work and therefore no trend in the relative differences is detectable. Even with the obvious trends in the relative differences for work, all the data shown in Fig. 7 are relatively well centered around 0, showing a good agreement between the results computed using the MNM and the BD.

### C. Large-deviation functions

Let us now investigate fluctuations of work  $w(\tau N) = w = \int_0^{\tau N} dt \dot{w}(t)$  and heat  $q(\tau N) = q = \int_0^{\tau N} dt \dot{q}(t)$  integrated over many cycles  $N \gg 1$  of duration  $\tau = 1$  [see Eqs. (74) and (75)]. According to the large deviation theory [28] reviewed in Sec. V E, in such situation the PDFs for work and heat assume the form (57) with  $X/\tau = w/\tau N$  and  $X/\tau = q/\tau N$  for work and heat, respectively, on the right-hand side, i.e.,

$$\rho_w(w) \sim \exp\left[\tau N J_w\left(\frac{w}{\tau N}\right)\right], \quad (79)$$

$$\rho_q(q) \sim \exp\left[\tau N J_q\left(\frac{q}{\tau N}\right)\right]. \quad (80)$$

The LDFs  $J_w(w)$  and  $J_q(q)$  are determined by the largest eigenvalues of the tilted propagators used in the previous section for the MGFs; see Sec. V E and Eqs. (63)–(65) for details.

In Fig. 8, we show the LDFs  $J_w(w)$  and  $J_q(q)$  computed using the MNM. For  $N \gg 1$ , the nonextensive boundary term  $U[x(t), t] - U[x(0), 0]$  in Eq. (75) can be neglected as

compared to  $-w(\tau N)$ , so that  $\rho_x(x) \sim \exp[\tau N J_x(x/\tau N)]$  for  $x = q, -w$ , and  $J_q(q) = J_w(-w)$ , as is verified by our MNM results (superimposing lines). However, the data obtained from  $10^6$  BD trajectories (symbols) shows that only the work distribution ( $\square$ ) attains the large deviation limit quickly, while the heat distribution ( $\circ$ ) has not converged, even for  $N = 100$  cycles. This is because, for the parameters considered in our numerical study, heat fluctuates much more than work, as already suggested by the moments and cumulants shown in Fig. 6(b) and 6(c). Let us note that while we have computed the LDFs using the standard BD, which was much more time consuming than the evaluation of the MNM, there are various optimized simulation algorithms [45–47] for computing of LDFs that can render BD simulations more competitive.

## VIII. CONCLUSION AND OUTLOOK

We have presented a numerical scheme for overdamped FPEs with time-dependent coefficients, based on the mapping between the FPE and a master equation with detailed-balanced transition rates. The resulting numerical method yields thermodynamically consistent results for arbitrary discretizations. It can be used for solving the FPE and also for computing MGFs and LDFs for functionals defined along the trajectories of the stochastic process underlying the FPE.

The performance of the method for solving the FPE is similar to other numerical methods relying on approximating the derivatives by finite differences. However, due to its thermodynamic robustness, the method predicts well the qualitative behavior of the studied system already for coarse meshes that capture merely the salient features of the force field and potential landscape. Thus, the MNM can safely be used for a fast scanning of the parameter space if one looks for interesting effects.

The presented numerical scheme shares basic notions with so-called Markov-state models of molecular kinetics, which have been employed for interpreting data from single-molecule experiments and molecular-dynamics simulations [48]. Both methods exploit the mapping of stochastic processes occurring in continuous space and time to discrete state-space Markov processes. While the kinetic Markov-state models are often based on special protocols, such as time-periodic driving [49], our formulation can in principle handle arbitrary time-dependent protocols.

Unfortunately, the MNM cannot easily be generalized to underdamped systems because it relies on the mapping (6) between the FPE (11) and the master equation (5), which is restricted to overdamped dynamics. The difficulties with the underdamped limit can be anticipated from the transition rates (13) and (14) and rates (18) and (19), which are all of the form  $D \exp(\pm A/D)$ . Thus, some of them necessarily diverge if the diffusion coefficient  $D$  goes to zero. The only variables with vanishing diffusion coefficient (noise) in the Langevin equation [see, e.g., Eqs. (67)–(68) for variables with nonzero diffusion coefficients in the Langevin equation] tractable by the MNM in its present form are variables like work and heat [see Eqs. (72)–(72)], which do not feed back onto the dynamics of the noisy variables. For such variables, the MNM yields MGFs and LDFs. The presented form of the MNM is thus limited to those underdamped situations where

the momentum in the underdamped Langevin equation does not depend on the position. A promising way to generalize the MNM for general underdamped dynamics could build on the path integral method suggested in Ref. [50], which shares with the MNM the important property of summing over all possible paths of the stochastic process and thus allows naturally incorporating calculations of various path-dependent stochastic variables. Another possible pathway to generalize the MNM to underdamped systems may be to incorporate into the MNM the ideas used in the formulation of the lattice Boltzmann method [51,52].

#### ACKNOWLEDGMENTS

We thank M. Žonda and H. Touchette for valuable comments on a preliminary version of the paper. We also thank the two anonymous referees whose detailed reports helped us to improve the readability of the manuscript. V.H. gratefully acknowledges support by the Humboldt Foundation and by the Czech Science Foundation (Project No. 17-06716S). S.S. acknowledges funding by International Max Planck Research Schools (IMPRS).

#### APPENDIX A: SPACE-DEPENDENT DIFFUSION COEFFICIENT

In this Appendix, we will show that the transition rates obeying the local detailed balance condition (9) can be used for solving only FPEs with position-independent diffusion coefficients. For the proof, it suffices to consider the one-dimensional FPE

$$\partial_t \rho(x, t) = [\partial_x^2 D_x(x, t) - \partial_x \mu_x F_x(x, t)] \rho(x, t) \quad (\text{A1})$$

and the corresponding master equation

$$\begin{aligned} \dot{p}_{i_x} = & r_{i_x+1 \rightarrow i_x} p_{i_x+1} + r_{i_x-1 \rightarrow i_x} p_{i_x-1} \\ & - (r_{i_x \rightarrow i_x+1} + r_{i_x \rightarrow i_x-1}) p_{i_x} \end{aligned} \quad (\text{A2})$$

on the discrete lattice with points indexed by  $i_x = \lfloor \frac{x-x_-}{\Delta_x} \rfloor$  and the lattice parameter  $\Delta_x = \frac{x_+ - x_-}{N_x}$ . We assume that  $D_x$  and  $F_x$  in (A1) depend on time  $t$  and position  $x$  and we look for transition rates in (A2) fulfilling the condition (10) and yielding Eq. (A1) in the leading order in the discretization parameter  $\Delta_x$  if we set  $\rho(x, t) = \lim_{\Delta_x \rightarrow 0} p_{i_x} / \Delta_x$ .

In one dimension, the entropy production  $\Delta S_R(x \rightarrow x + \Delta_x) = \Delta S_R = \int_x^{x+\Delta_x} dx' \frac{F_x(x')}{T_x(x')}$  along the transition from  $x$  to  $x + \Delta_x$  can be written as

$$\Delta S_R / k_B = -[\tilde{U}(x + \Delta_x, t) - \tilde{U}(x, t)], \quad (\text{A3})$$

where  $\tilde{U}$  is a dimensionless potential such that  $F_x / k_B T_x = \mu_x F_x / D_x = -\partial_x \tilde{U}$ . The transition rates satisfying the detailed balance condition (9) can thus in general be written as

$$r_{i_x \rightarrow i_x+1} = \frac{A_{i_x+1/2}}{\Delta_x^2} \exp\left[-\frac{\tilde{U}_{i_x+1} - \tilde{U}_{i_x}}{2}\right], \quad (\text{A4})$$

$$r_{i_x+1 \rightarrow i_x} = \frac{A_{i_x+1/2}}{\Delta_x^2} \exp\left[\frac{\tilde{U}_{i_x+1} - \tilde{U}_{i_x}}{2}\right], \quad (\text{A5})$$

where  $\tilde{U}_{i_x} = \tilde{U}(x_- + \Delta_x i_x, t)$  and  $A_{i_x+1/2} = A[x_- + \Delta_x(i_x + 1/2), t]$  is some space- and time-dependent function deter-

mining the prefactor of the transition rates. Inserting the rates (A4) and (A5) in the master equation (A2), we obtain up to the leading order in  $\Delta_x$  a partial differential equation of the form

$$\partial_t \rho = A'(\rho \tilde{U}' + \rho') + A(\rho' \tilde{U}' + \rho'' + \rho \tilde{U}''), \quad (\text{A6})$$

where  $\rho' \equiv \partial_x \rho(x, t)$ . The first nonzero correction to Eq. (A6) is of order  $\Delta_x^2$ . Comparing Eq. (A6) with the desired Eq. (A1) and using  $\tilde{U}' = -\mu_x F_x / D_x$ , we find that it is not possible to choose  $A(x)$  in such a way that the two equations are identical, unless the diffusion coefficient is position independent ( $D'_x = 0$ ) and  $A = D_x$ .

The main problem of why the transition rates of the forms (A4) and (A5) cannot yield the FPE (A1) with space-dependent coefficients are the prefactors  $A_{i_x+1/2}$ , which must be the same for the transitions  $i_x \rightarrow i_x + 1$  and  $i_x + 1 \rightarrow i_x$ . If we relax this assumption [and thus we do not consider only the rates strictly fulfilling the local detailed balance condition (9)], it is not difficult to find transition rates which can be used for solving the FPE (A1) in its full generality. They read

$$r_{i_x \rightarrow i_x+1} = \frac{A_{i_x}}{\Delta_x^2} \exp\left[-\frac{\tilde{U}_{i_x+1} - \tilde{U}_{i_x}}{2}\right], \quad (\text{A7})$$

$$r_{i_x+1 \rightarrow i_x} = \frac{A_{i_x+1}}{\Delta_x^2} \exp\left[\frac{\tilde{U}_{i_x+1} - \tilde{U}_{i_x}}{2}\right], \quad (\text{A8})$$

with  $\tilde{U}_{i_x} = \tilde{U}(x_- + \Delta_x i_x, t)$  and  $A_{i_x} = A(x_- + \Delta_x i_x, t)$ , where  $\tilde{U}' = -\mu_x F_x / D_x$  and  $A(x, t) = D_x(x, t)$ . Inserting these transition rates into the master equation (A2), we obtain up to the leading order in  $\Delta_x$  the FPE (A1). The first nonzero correction is of the order of  $\Delta_x^2$ . Although the rates (A7) and (A8) do not obey the strict local detailed balance condition (9), they still describe dynamics that conserves positivity and normalization. Furthermore, for position-independent diffusion coefficients, the detailed-balanced rates (A4) and (A5) and the rates (A7) and (A8) are identical. The generalization of the transition rates (A7) and (A8) to higher dimensions is straightforward.

#### APPENDIX B: MOMENT-GENERATING FUNCTION FOR TIME-AVERAGED CURRENT

In this Appendix, we calculate the moment-generating function  $\chi_{\tilde{\mathfrak{J}}(\mathbf{r}_a)}(\mathbf{s}_{\tilde{\mathfrak{J}}}, \tau) = \chi_{\tilde{\mathfrak{J}}(\mathbf{r}_a)}(s_{\tilde{\mathfrak{J}}_x}, s_{\tilde{\mathfrak{J}}_y}, \tau)$  for the time-averaged particle current through the position  $\mathbf{r}_a = (x_a, y_a)$  at time  $t_0$ :

$$\tilde{\mathfrak{J}}(\mathbf{r}_a, \tau) = \tilde{\mathfrak{J}}(\mathbf{r}_a, \tau, t_0) = \frac{1}{\tau} \int_{t_0}^{t_0+\tau} dt \tilde{\mathfrak{J}}(\mathbf{r}_a, t). \quad (\text{B1})$$

In the limit  $\tau \rightarrow 0+$ , this random variable converges to the microscopic current  $\tilde{\mathfrak{J}}(\mathbf{r}_a, t)$  defined in Eq. (40). The following strategy for calculating  $\chi_{\tilde{\mathfrak{J}}(\mathbf{r}_a)}(\mathbf{s}_{\tilde{\mathfrak{J}}}, \tau)$  can be easily generalized to more complex random variables discussed in Sec. VB.

The MGF  $\chi_{\tilde{\mathfrak{J}}(\mathbf{r}_a)}(\mathbf{s}_{\tilde{\mathfrak{J}}}, \tau)$  is defined as the two-sided Laplace transform

$$\chi_{\tilde{\mathfrak{J}}(\mathbf{r}_a)}(\mathbf{s}_{\tilde{\mathfrak{J}}}, \tau) = \int d\tilde{\mathfrak{J}}_x \int d\tilde{\mathfrak{J}}_y e^{-s_{\tilde{\mathfrak{J}}_x} \tilde{\mathfrak{J}}_x - s_{\tilde{\mathfrak{J}}_y} \tilde{\mathfrak{J}}_y} p_{\tilde{\mathfrak{J}}(\mathbf{r}_a)}(\tilde{\mathfrak{J}}, \tau) \quad (\text{B2})$$



of the probability distribution  $p_{\tilde{\mathcal{J}}(\mathbf{r}_a)}(\tilde{\mathcal{J}}, \tau)$  for  $\tilde{\mathcal{J}}(\mathbf{r}_a)$ . To calculate an approximation to  $\chi_{\tilde{\mathcal{J}}(\mathbf{r}_a)}(s_{\tilde{\mathcal{J}}}, \tau)$  using the discrete model of Fig. 1, we count the number  $n_x^r$  of jumps to the right from the site  $(i_{x_a}, i_{y_a})$  during the time interval  $(t_0, t_0 + \tau)$  and also the corresponding number  $n_x^l$  of jumps to the left from the site  $(i_{x_a} + 1, i_{y_a})$  to get the net transport

$$n_x = n_x(\tau, t_0) = n_x^r - n_x^l = \lim_{\Delta_x \rightarrow 0, \Delta_y \rightarrow 0} (\tau \Delta_y \tilde{\mathcal{J}}_x). \quad (\text{B3})$$

Here, the factor  $\Delta_y$  comes from Eq. (38). Similarly, the numbers  $n_y^u$  and  $n_y^d$  of jumps up from  $(i_{x_a}, i_{y_a})$  and down from  $(i_{x_a}, i_{y_a} + 1)$ , respectively, determine  $n_y = n_y^u - n_y^d = \lim_{\Delta_x \rightarrow 0, \Delta_y \rightarrow 0} (\tau \Delta_x \tilde{\mathcal{J}}_y)$ .

Let us now consider a time interval  $dt$  so short that only a single jump can occur and investigate the PDF for  $\mathbf{n} = (n_x, n_y)$ , which can be mapped to the PDF for the time-averaged current  $\tilde{\mathcal{J}}$ . At the initial time  $t_0$ , the distribution of the particle position is described by the vector  $\mathbf{p}(t_0)$ , and the number of jumps is  $n_x = n_y = 0$ . The joint PDF that the particle dwells in a specific site and that the current has a certain value is thus initially given by  $\tilde{\mathbf{p}}(\mathbf{n}, t_0, t_0) = \mathbf{p}(t_0)\delta(\mathbf{n})$ . After time  $dt$ , the number of jumps attains nonzero values solely by jumps described by the transition rates:

- (1)  $r_{i_{x_a} \rightarrow i_{x_a}+1}^{i_{y_a}}(t_0)$ ,  $n_x$  increases by 1,
- (2)  $r_{i_{x_a}+1 \rightarrow i_{x_a}}^{i_{y_a}}(t_0)$ ,  $n_x$  decreases by 1,
- (3)  $l_{i_{y_a} \rightarrow i_{y_a}+1}^{i_{x_a}}(t_0)$ ,  $n_y$  increases by 1,
- (4)  $l_{i_{y_a}+1 \rightarrow i_{y_a}}^{i_{x_a}}(t_0)$ ,  $n_y$  decreases by 1.

Using the master equation (26), the vector of occupation probabilities at time  $t_0 + dt$  can, for short  $dt$ , be written as  $\mathbf{p}(t_0 + dt) = \mathcal{U}(t_0 + dt, t_0)\mathbf{p}(t_0)$ , where  $\mathcal{U}(t_0 + dt, t_0) = [\mathcal{I} + dt\mathcal{R}(t_0)]$  and  $\mathcal{I}$  denotes the identity matrix. The joint PDF for the dimensionless current and position at time  $t_0 + dt$  can be written as

$$\tilde{\mathbf{p}}(\mathbf{n}, t_0 + dt, t_0) = \tilde{\mathcal{U}}(t_0 + dt, t_0, \mathbf{n})\mathbf{p}(t_0). \quad (\text{B4})$$

Here, all the matrix elements of  $\tilde{\mathcal{U}}(t_0 + dt, t_0, \mathbf{n})$  are given by the matrix elements of  $\mathcal{U}(t_0 + dt, t_0)$ , which are multiplied by  $\delta(n_x)\delta(n_y)$  except for the four elements containing the above-mentioned transition rates. The corresponding nonvanishing currents  $n_{x,y} \neq 0$  are represented by shifted  $\delta$  functions. For example, the element of  $\mathcal{U}$  containing the rate  $r_{i_{x_a} \rightarrow i_{x_a}+1}^{i_{y_a}}$  is in  $\tilde{\mathcal{U}}$  multiplied by  $\delta(n_x - 1)$ , the element of  $\mathcal{U}$  containing the rate  $r_{i_{x_a}+1 \rightarrow i_{x_a}}^{i_{y_a}}$  is in  $\tilde{\mathcal{U}}$  multiplied by  $\delta(n_x + 1)$ , and similarly for the other two elements.

Using the definitions (27)–(29) of  $\mathbf{p}(t_0)$ , the matrix element  $[\tilde{\mathcal{U}}(t_0 + dt, t_0, \mathbf{n})]_{mn}dn_xdn_y$  stands for the joint probability that a particle starting at time  $t_0$  from site  $[i_x(n), i_y(n)]$  will arrive to site  $[i_x(m), i_y(m)]$  at time  $t_0 + dt$  given that the numbers of jumps  $n_x$  and  $n_y$  at site  $[i_{x_a}, i_{y_a}]$  during the interval  $[t_0, t_0 + dt]$  assume values from the intervals  $(n_x, n_x + dn_x)$  and  $(n_y, n_y + dn_y)$ . The matrix  $\tilde{\mathcal{U}}(t_0 + 2dt, t_0 + dt, \mathbf{n})$  allows us to construct the joint PDF  $\tilde{\mathbf{p}}(\mathbf{n}, t_0 + 2dt, t_0)$  from  $\tilde{\mathbf{p}}(\mathbf{n}, t_0 + dt, t_0)$  in a manner similar to  $\tilde{\mathbf{p}}(\mathbf{n}, t_0 + dt)$  from  $\mathbf{p}(t_0)$ . The only difference is that now the distribution for  $\mathbf{n}$  is more involved. Namely, to get the PDF for the current at time  $t_0 + 2dt$ , we need to integrate over all possible combinations of the initial  $\mathbf{n}$  and the

increase in  $\mathbf{n}$  during the time interval  $dt$ :  $\tilde{\mathbf{p}}(\mathbf{n}, t_0 + 2dt, t_0) = \int dn_x' \int dn_y' \tilde{\mathcal{U}}(t_0 + 2dt, t_0 + dt, \mathbf{n}')\tilde{\mathbf{p}}(\mathbf{n} - \mathbf{n}', t_0 + dt, t_0) = [\tilde{\mathcal{U}}(t_0 + 2dt, t_0 + dt) \star \tilde{\mathbf{p}}(t_0 + dt, t_0)](\mathbf{n}) = [\tilde{\mathcal{U}}(t_0 + 2dt, t_0 + dt) \star \tilde{\mathcal{U}}(t_0 + dt, t_0)](\mathbf{n})\mathbf{p}(t_0)$ , where  $\star$  denotes convolutions in  $n_x$  and  $n_y$ . In a similar manner, one can construct the joint PDF  $\tilde{\mathbf{p}}(\mathbf{n}, t_0 + \tau, t_0)$  for the whole time interval  $(t_0, t_0 + \tau)$ . The obvious technical difficulty here lies in the fact that such a PDF would contain many convolutions.

To circumvent this issue, it is advantageous to focus on moment-generating functions instead of PDFs. According to the definition (B2), the MGF is a Laplace transform of the PDF, which transforms convolutions of original functions into products of transformed functions. The joint PDF  $\tilde{\mathbf{p}}(\mathbf{n}, t_0 + 2dt, t_0)$  is thus transformed in  $\mathbf{p}_s(\mathbf{s}_n, t_0 + 2dt, t_0) = \tilde{\mathcal{U}}(t_0 + 2dt, t_0 + dt, \mathbf{s}_n)\tilde{\mathcal{U}}(t_0 + dt, t_0, \mathbf{s}_n)\mathbf{p}(t_0)$ , where the matrices  $\tilde{\mathcal{U}}(t + dt, t, \mathbf{s}_n)$  are given by Laplace transform of the matrices  $\tilde{\mathcal{U}}(t + dt, t, \mathbf{n})$ . These matrices are called *tilted* matrices and they are identical to  $\tilde{\mathcal{U}}(t + dt, t, \mathbf{n})$  except for the  $\delta$  functions  $\delta(n_x \mp 1)$  and  $\delta(n_y \mp 1)$  in  $\tilde{\mathcal{U}}(t + dt, t, \mathbf{n})$  that are transformed to the exponentials  $\exp(\mp s_{n_x})$  and  $\exp(\mp s_{n_y})$  and the  $\delta$  functions  $\delta(n_x)$  and  $\delta(n_y)$  that are both transformed to 1. The vector  $\mathbf{p}_s(\mathbf{s}_n, t_0)$  thus obeys a similar dynamical equation as the probability vector  $\mathbf{p}(t_0)$ :

$$\frac{d}{dt}\mathbf{p}_s(\mathbf{s}_n, t, t_0) = \tilde{\mathcal{R}}_{\mathbf{s}_n}(t)\mathbf{p}_s(\mathbf{s}_n, t, t_0), \quad (\text{B5})$$

where the *tilted* rate matrix  $\tilde{\mathcal{R}}_{\mathbf{s}_n}(t) = [\tilde{\mathcal{U}}(t + dt, t, \mathbf{s}_n) - \mathcal{I}]/dt$  can be obtained from the rate matrix  $\mathcal{R}(t)$  multiplying the rate  $r_{i_{x_a} \rightarrow i_{x_a}+1}^{i_{y_a}}(t)$  by  $\exp(-s_{n_x})$ , the rate  $r_{i_{x_a}+1 \rightarrow i_{x_a}}^{i_{y_a}}(t)$  by  $\exp(s_{n_x})$ , the rate by  $l_{i_{y_a} \rightarrow i_{y_a}+1}^{i_{x_a}}(t)$  by  $\exp(-s_{n_y})$ , and the rate  $l_{i_{y_a}+1 \rightarrow i_{y_a}}^{i_{x_a}}(t)$  by  $\exp(s_{n_y})$ , and keeping all other rates unchanged. For a given  $\mathbf{s}_n$ , the formula (B5) can be solved in a manner similar to the formula for (26) for  $\mathbf{p}(t)$ . For a time-independent tilted rate matrix  $\tilde{\mathcal{R}}_{\mathbf{s}_n}$ , the solution to Eq. (B5) is given by a matrix exponential

$$\mathbf{p}_s(\mathbf{s}_n, t, t_0) = \exp[\tilde{\mathcal{R}}_{\mathbf{s}_n}(t - t_0)]\mathbf{p}(t_0), \quad (\text{B6})$$

while for a time-dependent rate matrix  $\tilde{\mathcal{R}}_{\mathbf{s}_n}(t)$  the solution should be constructed using the time discretization analogous to the one used in Eq. (33) with  $\Delta_t = (t_0 + t)/N_t$ . We get

$$\mathbf{p}_s(\mathbf{s}_n, t, t_0) = \lim_{\Delta_t \rightarrow 0} \prod_{i=0}^{i_t(t)} \tilde{\mathcal{U}}_i(\mathbf{s}_n)\mathbf{p}(t_0), \quad (\text{B7})$$

where  $\tilde{\mathcal{U}}_i(\mathbf{s}_n) = \exp[\tilde{\mathcal{R}}_{\mathbf{s}_n}(t_0 + i\Delta_t)\Delta_t]$  if  $i \geq 1$  and  $\tilde{\mathcal{U}}_0 = \mathcal{I}$ . The vectors (B6) and (B7) give moment-generating functions for  $n_x$  and  $n_y$  conditioned on the final state of the system during the evolution. The unconditional generating function is thus obtained by summing over all final states,

$$\chi(\mathbf{s}_n, t, t_0) = \mathbf{p}_+^\top \cdot \mathbf{p}_s(\mathbf{s}_n, t, t_0), \quad (\text{B8})$$

where  $\mathbf{p}_+^\top$  is a vector of ones.

For fine discretizations, the moment-generating function  $\chi_n(\mathbf{s}_n, t_0 + \tau, t_0) = \chi_n(s_{n_x}, s_{n_y})$  finally approximates the MGF  $\chi_{\tilde{\mathcal{J}}}(\mathbf{s}_{\tilde{\mathcal{J}}}, t_0 + \tau, t_0) = \chi_{\tilde{\mathcal{J}}}(s_{\tilde{\mathcal{J}}_x}, s_{\tilde{\mathcal{J}}_y})$  for the time-averaged current:

$$\chi_{\tilde{\mathcal{J}}}(s_{\tilde{\mathcal{J}}_x}, s_{\tilde{\mathcal{J}}_y}) = \lim_{\Delta_x \rightarrow 0, \Delta_y \rightarrow 0} \chi_n\left(\frac{s_{n_x}}{\tau \Delta_y}, \frac{s_{n_y}}{\tau \Delta_x}\right). \quad (\text{B9})$$

- [1] H. Risken, *The Fokker-Planck Equation Methods of Solution and Applications* (Springer, Berlin, 1996).
- [2] N. V. Kampen, *Phys. Rep.* **124**, 69 (1985).
- [3] R. Friedrich, J. Peinke, M. Sahimi, and M. R. R. Tabar, *Phys. Rep.* **506**, 87 (2011).
- [4] A. Traulsen, J. C. Claussen, and C. Hauert, *Phys. Rev. E* **85**, 041901 (2012).
- [5] F. Slanina, *Essentials of Econophysics Modelling* (Oxford University Press, Oxford, UK, 2013).
- [6] W. Paul and J. Baschnagel, *Stochastic Processes: From Physics to Finance* (Springer, Berlin, 2010).
- [7] A. A. Drăgulescu and V. M. Yakovenko, *Quant. Financ.* **2**, 443 (2002).
- [8] S. Narayanan and P. Kumar, in *IUTAM Symposium on Nonlinear Stochastic Dynamics and Control*, edited by W. Q. Zhu, Y. K. Lin, and G. Q. Cai (Springer, Dordrecht, 2011), pp. 77–86.
- [9] B. Sepehrian and M. K. Radpoor, *Appl. Math. Comput.* **262**, 187 (2015).
- [10] L. Pichler, A. Masud, and L. A. Bergman, in *Computational Methods in Stochastic Dynamics*, edited by M. Papadrakakis, G. Stefanou, and V. Papadopoulos (Springer, Dordrecht, 2013), Vol. 2, pp. 69–85.
- [11] V. Holubec, *Non-equilibrium Energy Transformation Processes: Theoretical Description at the Level of Molecular Structures* (Springer, Berlin, 2014).
- [12] S. Das, G. Gompper, and R. G. Winkler, *New J. Phys.* **20**, 015001 (2018).
- [13] K. Sekimoto, *Stochastic Energetics*, Lecture Notes in Physics Vol. 799 (Springer, Heidelberg, 2010).
- [14] U. Seifert, *Rep. Prog. Phys.* **75**, 126001 (2012).
- [15] T. Speck, *EPL* **114**, 30006 (2016).
- [16] A. Ryabov, V. Holubec, M. H. Yaghoubi, M. Varga, M. E. Foulaadvand, and P. Chvosta, *J. Stat. Mech: Theory Exp.* (2016) 093202.
- [17] V. Holubec, A. Ryabov, M. H. Yaghoubi, M. Varga, A. Khodae, M. E. Foulaadvand, and P. Chvosta, *Entropy* **19**, 119 (2017).
- [18] L. Ornigotti, A. Ryabov, V. Holubec, and R. Filip, *Phys. Rev. E* **97**, 032127 (2018).
- [19] M. F. Weber and E. Frey, *Rep. Prog. Phys.* **80**, 046601 (2017).
- [20] J. Chang and G. Cooper, *J. Comput. Phys.* **6**, 1 (1970).
- [21] C. Jarzynski, *J. Stat. Phys.* **98**, 77 (2000).
- [22] C. Maes and K. Netočný, *Scholarpedia* **8**, 9664 (2013).
- [23] C. Maes and K. Netočný, *J. Stat. Phys.* **110**, 269 (2003).
- [24] T. S. Komatsu, N. Nakagawa, S.-i. Sasa, and H. Tasaki, *Phys. Rev. Lett.* **100**, 230602 (2008).
- [25] N. Nakagawa and S.-i. Sasa, *Phys. Rev. E* **87**, 022109 (2013).
- [26] A. V. Chechkin, F. Seno, R. Metzler, and I. M. Sokolov, *Phys. Rev. X* **7**, 021002 (2017).
- [27] M. Baiesi, C. Maes, and K. Netočný, *J. Stat. Phys.* **135**, 57 (2009).
- [28] H. Touchette, *Phys. Rep.* **478**, 1 (2009).
- [29] A. C. Barato and R. Chetrite, *J. Stat. Mech. Theor. Exp.* (2018) 053207.
- [30] P. Romanczuk, M. Bär, W. Ebeling, B. Lindner, and L. Schimansky-Geier, *Eur. Phys. J. Special Topics* **202**, 1 (2012).
- [31] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe, *Rev. Mod. Phys.* **88**, 045006 (2016).
- [32] G. Gompper, C. Bechinger, S. Herminghaus, R. Isele-Holder, U. B. Kaupp, H. Löwen, H. Stark, and R. G. Winkler, *Eur. Phys. J. Special Topics* **225**, 2061 (2016).
- [33] V. Blickle and C. Bechinger, *Nat. Phys.* **8**, 143 (2012).
- [34] I. A. Martínez, É. Roldán, L. Dinis, D. Petrov, J. M. R. Parrondo, and R. A. Rica, *Nat. Phys.* **12**, 67 (2016).
- [35] S. Krishnamurthy, S. Ghosh, D. Chatterji, R. Ganapathy, and A. K. Sood, *Nat. Phys.* **12**, 1134 (2016).
- [36] G. Verley, M. Esposito, T. Willaert, and C. Van den Broeck, *Nat. Commun.* **5**, 4721 (2014).
- [37] M. Poletti, G. Verley, and M. Esposito, *Phys. Rev. Lett.* **114**, 050601 (2015).
- [38] K. Proesmans, B. Cleuren, and C. V. den Broeck, *EPL* **109**, 20004 (2015).
- [39] T. Schmiel and U. Seifert, *EPL* **81**, 20003 (2008).
- [40] R. Zakine, A. Solon, T. Gingrich, and F. van Wijland, *Entropy* **19**, 193 (2017).
- [41] V. Holubec and A. Ryabov, *Phys. Rev. Lett.* **121**, 120601 (2018).
- [42] P. Pietzonka and U. Seifert, *Phys. Rev. Lett.* **120**, 190602 (2018).
- [43] T. Koyuk, U. Seifert, and P. Pietzonka, *J. Phys. A: Math. Theor.* **52**, 02LT02 (2019).
- [44] On each blue box, the central red mark indicates the median, and the bottom and top blue edges of the box indicate the 25th and 75th percentiles, respectively. The black dashed whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the red + symbol. Taken from MATLAB documentation, <https://www.mathworks.com/help/stats/boxplot.html>.
- [45] C. Giardinà, J. Kurchan, and L. Peliti, *Phys. Rev. Lett.* **96**, 120603 (2006).
- [46] T. Nemoto, F. Bouchet, R. L. Jack, and V. Lecomte, *Phys. Rev. E* **93**, 062123 (2016).
- [47] G. Ferré and H. Touchette, *J. Stat. Phys.* **172**, 1525 (2018).
- [48] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, *J. Chem. Phys.* **134**, 174105 (2011).
- [49] H. Wang and C. Schütte, *J. Chem. Theory Comput.* **11**, 1819 (2015).
- [50] A. Lyasoff, *Mathematica J.* **9**(2) (2004).
- [51] X. He and L.-S. Luo, *Phys. Rev. E* **56**, 6811 (1997).
- [52] S. Chen and G. D. Doolen, *Annu. Rev. Fluid Mech.* **30**, 329 (1998).

## Cycling Tames Power Fluctuations near Optimum Efficiency

Viktor Holubec<sup>1,2,\*</sup> and Artem Ryabov<sup>2</sup><sup>1</sup>*Institut für Theoretische Physik, Universität Leipzig, Postfach 100 920, D-04009 Leipzig, Germany*<sup>2</sup>*Charles University, Faculty of Mathematics and Physics, Department of Macromolecular Physics, V Holešovičkách 2, CZ-180 00 Praha, Czech Republic*

(Received 16 May 2018; published 17 September 2018)

According to the laws of thermodynamics, no heat engine can beat the efficiency of a Carnot cycle. This efficiency traditionally comes with vanishing power output and practical designs, optimized for power, generally achieve far less. Recently, various strategies to obtain Carnot's efficiency at large power were proposed. However, a thermodynamic uncertainty relation implies that steady-state heat engines can operate in this regime only at the cost of large fluctuations that render them immensely unreliable. Here, we demonstrate that this unfortunate trade-off can be overcome by designs operating cyclically under quasistatic conditions. The experimentally relevant yet exactly solvable model of an overdamped Brownian heat engine is used to illustrate the formal result. Our study highlights that work in cyclic heat engines and that in quasistatic ones are different stochastic processes.

DOI: 10.1103/PhysRevLett.121.120601

**Introduction.**—Conversion of disordered energy (heat) into directed motion (work) propels not only industry but also nature itself through photosynthesis. According to the laws of thermodynamics, the efficiency  $\eta = W/Q_h$  of this conversion is bounded from above by Carnot's efficiency  $\eta_C = 1 - T_c/T_h$  [1]. The average heat  $Q_h$  from a heat source can at most yield the average work  $W = \langle w \rangle = \eta_C Q_h$ , the remaining energy must be transferred into a heat sink. The upper bound is saturated if the temperatures of the hot and cold heat reservoirs assume constant values  $T_h$  and  $T_c$ , respectively, and if the heat engine (HE) operates reversibly. Also, it is frequently argued that  $\eta_C$  can be reached only if the engine operates on an infinite timescale  $t_p$  with vanishing output power  $P = W/t_p$ . Recently, this claim has been seriously challenged [2–15].

It was shown that either using a special coupling between subsystems [3], working substances close to criticality [4,6], or scalings leading to vanishing system relaxation times [7–9], it is possible to asymptotically reach  $\eta_C$  with  $P > 0$ . Although the HEs used for derivation of the last-mentioned results obey the trade-off bounds  $P \leq C(\eta_C - \eta)$  [10,11,13], they can operate with  $\eta = \eta_C$  and  $P > 0$  since the parameter  $C$  generally diverges with a vanishing system relaxation time [16].

However, it was suggested that the price one has to pay for overcoming the trade-off between power and efficiency are large power fluctuations [6,9]. In the critical heat engine [4], the fluctuations almost surely dominate the averages [6] and also steady state HEs (SSHEs) exhibit large power fluctuations [9].

Here, we show that such a trade-off does not exist for quasistatic cyclic HEs (CHEs) with controllable relaxation

times. These machines can work with  $\eta$  asymptotically close to  $\eta_C$  at  $P > 0$  with vanishing fluctuations. Specifically, we show that both the work and power fluctuation  $\tilde{\sigma}_P = \sigma_W/W = \sqrt{\langle w^2 \rangle - W^2}/W$  and the Fano factor for work  $\sigma_W^2/W$  are finite and can even vanish.

Our results highlight that the work done by CHEs and the work done by SSHEs are two different stochastic processes. Although their mean values can be equal [17–19], their fluctuations are qualitatively different. The work in the SSHEs obeys thermodynamic uncertainty relations [20–24] which imply that the Fano factor for the output work diverges if the efficiency reaches  $\eta_C$  [9]. The work in the CHEs obeys no such relation and it is possible to construct a CHE operating with Carnot's efficiency and delivering a persistent deterministic power output.

**Cyclic heat engines.**—Consider a periodically driven HE operating along a quasistatic Carnot cycle composed of two isotherms connected by two adiabats. For concreteness, we consider a one-dimensional system with the Hamiltonian

$$H(x, t) = k(t)x^{2n}/2n, \quad n = 1, 2, \dots, \quad (1)$$

where  $k = k(t)$  controls its stiffness and  $x = x(t)$  is a continuous stochastic process describing the microstate of the system. The Hamiltonian (1) serves as a mere illustration. Our main results are valid for arbitrary thermodynamic systems which can operate quasistatically, including many-dimensional systems with momentum degrees of freedom and systems with discrete state space.

The operational cycle of the engine is depicted in Fig. 1. During the hot isotherm at  $T_h$  (branch 1) and during the subsequent adiabat (branch 2), the Hamiltonian

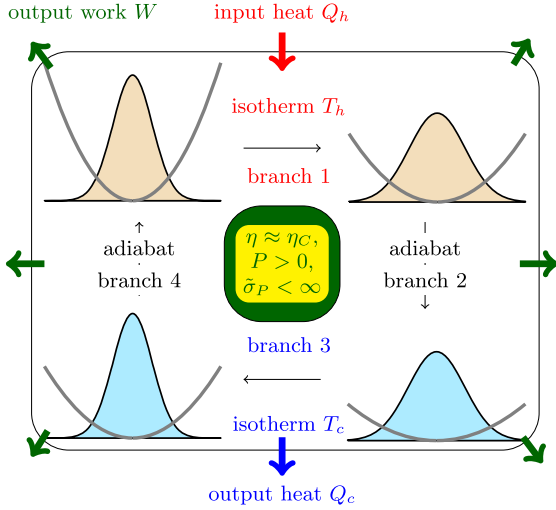


FIG. 1. The operational cycle of the considered cyclic heat engines. Gray lines depict the Hamiltonian (1) and the shaded areas stand for the probability density of the particle position during the cycle.

opens ( $\dot{k} \geq 0$ ) and the system performs work  $w_h = -\int_{1,2} dt \partial_t H(x, t) = -\int_{1,2} dt \dot{k}(t) x(t)^{2n} / 2n$  on the environment (the integration runs over the branches 1 and 2). During the rest of the cycle, the Hamiltonian closes ( $\dot{k} \leq 0$ ) and the engine consumes work  $-w_c = \int_{3,4} dt \partial_t H(x, t) = -\int_{3,4} dt \dot{k}(t) x(t)^{2n} / 2n$ . The heat on average enters the system during the hot isotherm and leaves it during the cold one (branch 3). We denote the duration of the  $i$ th branch as  $t_i$  and as  $t_p = t_1 + t_2 + t_3 + t_4$  the duration of the whole cycle.

The average thermodynamics of the engine observed after averaging work and heat over many cycles is that of a standard reversible Carnot cycle. Namely, a combination of the first and the second law of thermodynamics implies that the average output work  $W$  is given by [1,25]

$$W = \langle w \rangle = \langle w_h + w_c \rangle = Q_h - Q_c = (T_h - T_c) \Delta S, \quad (2)$$

where  $\Delta S$  is the change of the system entropy during the hot isotherm. On the other hand, the work fluctuations depend both on the details of the Hamiltonian and on the way how the adiabatic branches are realized.

By definition, no heat flows into the system during adiabatic branches. This condition can be realized in two physically different ways. (i) One ensures that no heat *at all* flows between the system and the bath by performing the adiabats very fast, or by disconnecting the system from the reservoir. During these adiabatic branches, the system evolves deterministically regardless of the dynamics of the baths. In general, reconnecting the bath and the system at the end of such adiabat brings the system far from equilibrium. To keep the cycle quasistatic, it is necessary to

secure that the system state just before the reconnection is identical with the equilibrium state corresponding to the bath temperature and system Hamiltonian at the time of reconnection. (ii) One ensures that no heat is interchanged *on average* only by carefully controlling the system connected to the reservoir with varying temperature [39,40]. Because of the coupling to the bath, the system evolves during such adiabats stochastically.

We start with the traditional adiabatic branches (i) where no heat at all is exchanged leading to a deterministic evolution of the system during the adiabats. Then the work PDF  $p(w)$  can be expressed as an average over the distributions for internal energy increases  $\Delta H_2$  and  $\Delta H_4$  along the adiabatic branches 2 and 4, respectively [25]:

$$p(w) = \langle \delta\{w - [W - \widetilde{\Delta H}_2 - \widetilde{\Delta H}_4]\} \rangle, \quad (3)$$

where  $\widetilde{\Delta H}_i = \Delta H_i - \langle \Delta H_i \rangle$ ,  $i = 2, 4$ . The PDF for  $\Delta H_2$  and  $\Delta H_4$  can be constructed from the Boltzmann distribution  $\rho(x, \tau_i) = \exp[-H(x, \tau_i)/k_B T(\tau_i)]/Z(\tau_i)$  corresponding to the system Hamiltonian and bath temperature at times  $\tau_i$ ,  $i = 1, \dots, 4$  delimiting the adiabatic branches. Here,  $k_B$  denotes the Boltzmann constant and  $Z$  is the partition function.

The PDF (3) allows us to calculate all moments of work:  $\langle w^n \rangle = \int_{-\infty}^{\infty} dw w^n \rho(w)$ . For the case of infinitely fast adiabatic branches ( $t_2 \rightarrow 0$  and  $t_4 \rightarrow 0$ ), the microstate of the system during the adiabatic branches does not change. Assuming that the particle is at a microstate  $x$  at the beginning of the first adiabat and at a microstate  $y$  at the beginning of the second one, the energy differences in Eq. (3) read  $\Delta H_2 = H(x, t_1 + t_2) - H(x, t_1)$  and  $\Delta H_4 = H(y, t_p) - H(y, t_p - t_4)$  and the average therein must be taken over the PDF  $\rho(x, t_1)\rho(y, t_p)$ . The work and power fluctuation evaluated for the Hamiltonian (1) are then given by [25]

$$\tilde{\sigma}_w = \frac{\sigma_w}{W} = \frac{\sqrt{\langle w^2 \rangle - W^2}}{W} = \frac{1}{\sqrt{n}} \frac{k_B}{\Delta S}. \quad (4)$$

The function  $\tilde{\sigma}_w = \tilde{\sigma}_P$ , which quantifies observability of the average work and power at the Carnot efficiency, is thus finite and decreases both with the exponent  $n$  in the Hamiltonian (1) and with the change of the system entropy during the hot isotherm  $\Delta S$ .

During the adiabatic branches (i) performed in a finite time with the disconnected heat bath, the system undergoes a nontrivial evolution determined by the Hamiltonian (through Hamiltonian equations for classical systems and the Schrödinger equation in quantum cases). To get an analytical result valid for arbitrary  $H$ , we use the approximation that microstates occupied by the system at the beginning of the adiabats are independent from those occupied at their ends. Then, the assumption that the system is in equilibrium both before the beginning and

after the end of the adiabats allows us to calculate the work fluctuation along similar lines as in the previous case. The result is [25]

$$\tilde{\sigma}_w = \frac{1}{\sqrt{n}} \frac{k_B}{\Delta S} \frac{\sqrt{1 + (1 - \eta_C)^2}}{\eta_C} \geq \frac{1}{\sqrt{n}} \frac{k_B}{\Delta S}. \quad (5)$$

Compared to the work fluctuation (4),  $\tilde{\sigma}_w$  now depends on the temperatures of the two baths via the Carnot efficiency  $\eta_C$ . The additional factor is always greater than 1 and thus Eq. (4) for the cycle with instantaneous adiabatic branches sets the lower bound on Eq. (5).

The work fluctuations (4) and (5) are always nonzero. Their origin can be mapped to disconnecting the system from the baths during the adiabatic branches. According to its definition  $w = -\int_0^t dt \partial_t H(x, t)$ , the work is in CHEs done only if the Hamiltonian changes in time. Along a quasistatic process, the reservoir causes many transitions in the system on the timescale on which the external parameter corresponding to the work (e.g., the stiffness  $k$  here, a piston position in thermodynamics) is varied. The time spent by arbitrary quasistatic trajectory  $x(t)$  in a microstate  $y$  within the time window  $[t, t + dt]$  is determined by the Boltzmann distribution  $\rho(y, t)$ . The work  $w$  done during a quasistatic process along each trajectory is hence given by the average work  $W = -\int dx \int_0^t dt \partial_t H(x, t) \rho(x, t) = (T_h - T_c) \Delta S$  [25,26,41,42].

Quasistatic Carnot cycles with adiabatic branches (ii) where the system can interchange heat with the bath thus yield sharp work PDF,

$$p(w) = \delta(w - W), \quad (6)$$

with vanishing variance  $\sigma_w^2$  and fluctuation  $\tilde{\sigma}_w$ . Different from Eqs. (4) and (5), this result does not depend on the system Hamiltonian. As one consequence, the large power fluctuations found in the critical heat engine [4,6] can be avoided by utilizing this type of quasistatic adiabatic branches.

*Comparison with steady state heat engines.*—Steady state HEs are connected to the hot and to the cold reservoir simultaneously and operate in a nonequilibrium steady state. They obey the current fluctuation relations [20–24] which can be used to derive the inequality for the relative work and power variance [9]

$$\tilde{\sigma}_{w_t}^2 \geq \frac{2k_B T_c}{W_t} \frac{\eta}{\eta_C - \eta} = \frac{2k_B}{\Delta S_t}. \quad (7)$$

Here,  $W_t$  and  $\Delta S_t$  are the work and entropy generated during time window  $[0, t]$ . The formula (7) is valid in the long time limit  $t \rightarrow \infty$ , when the PDF for work attains the large deviation form.

The formula (7) implies that it is not possible to construct a SSHE working with Carnot's efficiency  $\eta = \eta_C$ ,

delivering work with a finite fluctuation  $\tilde{\sigma}_{w_t}$ , and operating reversibly with  $\Delta S_t = 0$ , at the same time. The SSHEs operating with  $\eta_C$  must either dissipate ( $\Delta S_t > 0$ ) or yield diverging work fluctuations ( $\tilde{\sigma}_{w_t} \rightarrow \infty$ ). This observation is a HE analogy of the result obtained for Brownian clocks [43].

Another striking difference between the CHEs and the SSHEs is revealed if we rewrite our findings for CHEs in terms of the Fano factor for work  $\sigma_w^2/W$ , which equals to the ratio of constancy  $\Delta_P = \sigma_P^2 t$ ,  $t \gg 1$  [9] to the output power  $P = W/t$ . The formula (4) for a CHE operating with Carnot efficiency gives

$$\frac{\Delta_P}{P} = \frac{\sigma_w^2}{W} = \frac{1}{n} \frac{T_h \eta_C k_B^2}{\Delta S} \quad (8)$$

and thus the Fano factor is in this case finite. Equation (5) yields analogous results and the Fano factor corresponding to the work PDF (6) even vanishes.

On the other hand, Eq. (7) for the SSHEs leads to

$$\frac{\Delta_{P_t}}{P_t} = \frac{\sigma_{w_t}^2}{W_t} \geq 2k_B T_c \frac{\eta}{\eta_C - \eta}, \quad (9)$$

which diverges whenever  $\eta \rightarrow \eta_C$ . The work and power fluctuations in the CHEs and in the SSHEs operating with  $\eta_C$  thus significantly differ.

One may object that these conclusions are based on a comparison of incompatible quantities—variables measured per cycle for CHEs and variables measured over a long time for SSHEs. Nevertheless, measuring the quantities for the CHEs over many cycles or over many independent systems does not alter the main conclusions. More precisely, averaging over  $N$  independent CHEs or, equivalently, over  $N$  cycles of a single CHEs, both the average output work  $W$  and its variance  $\sigma_w^2$  scale as  $N$ . Therefore, although the fluctuation  $\tilde{\sigma}_w$  scales as  $1/\sqrt{N}$ , the ratio  $\Delta_P/P = \sigma_w^2/W$  remains constant.

The difference between work in CHEs and SSHEs lies in the very definitions of these variables. Work in CHEs is done only when an external parameter changes and under quasistatic conditions it is independent of the initial microstate of the system [25]. On the contrary, work in SSHEs is usually done when the microstate  $x$  of the system changes. During this thermally induced transition, the system internal energy is increased in ratchets [5], particles are transferred against gradients of chemical potential in thermochemical heat engines [7,9], etc. Such defined work depends on the initial and final points of the stochastic trajectory  $\{x(t)\}_{t=0}^t$ , which, e.g., determine the increase in the internal energy in a ratchet, and thus it always fluctuates. Work in SSHEs hence lacks the self-averaging property of the work done in CHEs. It is rather similar to the heat  $Q = \int_0^t dt \partial_x H(x, t) \dot{x}$  in CHEs which is

interchanged with the bath also only if the system microstate changes.

Our analysis implies that the work done in SSHEs and that in CHEs represent two different stochastic processes which cannot be directly mapped onto each other. Nevertheless, such a mapping might be constructed if the different definitions of work in the two classes of HEs would be taken into account.

*Cyclic Brownian heat engine.*—Let us now propose an actual CHE operating close to Carnot's efficiency while delivering a stable power output. Its engineering is rather straightforward, it can be performed with an arbitrary thermodynamic system capable of quasistatic operation. In order to further demonstrate that such a HE can operate in finite time, delivering a nonzero output power, we need a system with controllable relaxation time. A paradigmatic example of such a system from the field of stochastic thermodynamics [44,45] is the overdamped Brownian HE [8,27,46].

The HE is based on an overdamped Brownian particle diffusing in a harmonic potential [47]  $U(x, t) = H(x, t) = k(t)x^2/2$ , whose dynamics obeys the Langevin equation

$$\dot{x} = -kx/\gamma + \sqrt{2k_B T/\gamma}\zeta. \quad (10)$$

Here,  $\zeta$  is the Gaussian white noise with  $\langle \zeta \rangle = 0$  and  $\langle \zeta(t)\zeta(t') \rangle = \delta(t-t')$ . The relaxation time for the position,  $\tau_x = \gamma/k$ , can be easily controlled in experiments through the trap stiffness  $k$  (yet it may depend on the temperature). The model is valid if this relaxation time is much longer than the relaxation time for the momentum,  $\tau_p = m/\gamma$ , given by the ratio of the particle mass  $m$  to the friction  $\gamma$ . The model (10) is exactly solvable and it has been thoroughly investigated both theoretically [8,27] and experimentally, using optical tweezers for generation of the potential [46,48,49].

To demonstrate our results for instantaneous adiabatic branches (i), we periodically modulate the bath temperature  $T$  and the trap stiffness  $k$  using the Carnot-like driving depicted in Fig. 1 with infinitely fast adiabatic branches. If the cycle is performed in a finite time  $t_p$ , with a non-vanishing relaxation time  $\tau_x$ , the system is during the cycle inevitably out of equilibrium and the HE efficiency is smaller than  $\eta_C$ . In order to realize the quasistatic Carnot cycle using a finite  $t_p$ , we thus need to use a very stiff trap, which makes  $\tau_x \ll t_p$ .

In Figs. 2(a) and 2(b), we introduce a suitable scaling of the cycle duration  $t_p$  and minimum and maximum trap stiffness  $k$  during the cycle which, in the limit of infinite scaling parameter  $\sigma_\infty$ , leads to a HE operating with Carnot's efficiency and delivering an infinite power with fluctuation given by Eq. (4) with  $n = 1$ . The convergence of the output power, the power fluctuation, and the efficiency to these values as the cycle becomes gradually

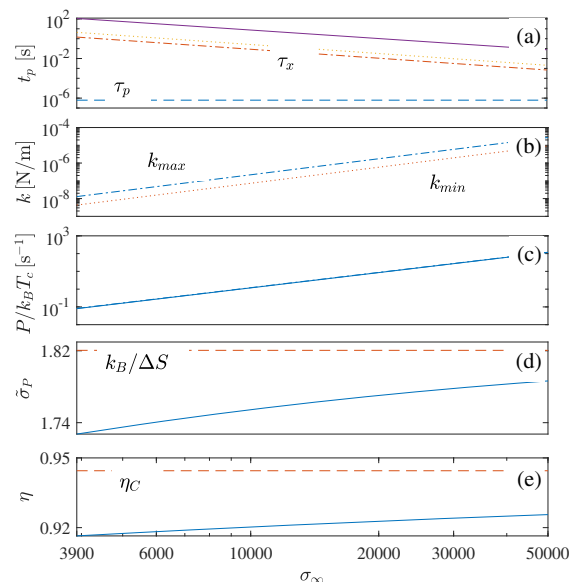


FIG. 2. Behavior of the overdamped Brownian HE with the scaling parameter  $\sigma_\infty$ . Timescale separation between the cycle duration  $t_p$  and the relaxation times  $\tau_x$  (we show its smallest and largest value during the cycle) and  $\tau_p$  is depicted in panel (a). In panel (b),  $k_{\max}$  ( $k_{\min}$ ) stand for the maximum or minimum value of the stiffness during the cycle. The shown values of cycle durations  $t_p$  and trap stiffnesses  $k$  are reasonable from experimental perspective. In panels (c) and (d) we demonstrate divergence of output power  $P$  and convergence of the relative power fluctuation  $\bar{\sigma}_P$  to  $k_B/\Delta S$  as the efficiency  $\eta$ , shown in panel (e), converges to  $\eta_C$  for large values of  $\sigma_\infty$ .

quasistatic with increasing  $\sigma_\infty$  is plotted in Figs. 2(c)–2(e), respectively. The curves are plotted using experimentally motivated values of the model parameters [46]. Further details are given in Supplemental Material [25]. The rest of our results can be tested along similar lines.

*Concluding remarks.*—Unlike steady state heat engines (SSHEs), cyclic heat engines (CHEs) can theoretically operate reversibly with Carnot's efficiency  $\eta_C$ , delivering a large and stable power output  $P$  with finite fluctuation and Fano factor. The main difference between the two classes of heat engines lies in the definitions of work in the two models. While the transitions caused in the system due to the contact with the bath lead to averaging of work in CHEs, such an averaging is not available for SSHEs. In the latter case, the work always depends on the initial and final point of a trajectory and thus inevitably fluctuates. The recently proposed one-to-one mappings between SSHEs and CHEs [17–19] thus break down on the level of work fluctuations.

In practice, the described strategy does not allow us to realize the strict limit  $\eta = \eta_C$  at  $P > 0$  without breaking the system-reservoir timescale separation used in standard thermodynamic models [12]. But it is possible to find parameter regimes where realizable systems operate with

efficiencies close to  $\eta_C$  and deliver large power  $P$  with small fluctuation. Experimental realizations of such HEs are possible using current micromanipulation techniques such as optical tweezers [48,49]. Finally, we stress that our results are valid for general HEs, including intensively studied quantum models [50].

We thank P. Pietzonka and A. Dechant for stimulating correspondence and to K. Kroy for helpful discussions. We gratefully acknowledge financial support by the Czech Science Foundation (Project No. 17-06716S). V. H. thanks the Humboldt Foundation for support.

\*viktor.holubec@gmail.com

- [1] H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics*, 2nd ed. (Wiley, New York, 1985).
- [2] G. Benenti, K. Saito, and G. Casati, *Phys. Rev. Lett.* **106**, 230602 (2011).
- [3] A. E. Allahverdyan, K. V. Hovhannisyanyan, A. V. Melkikh, and S. G. Gevorgian, *Phys. Rev. Lett.* **111**, 050601 (2013).
- [4] M. Campisi and R. Fazio, *Nat. Commun.* **7**, 11895 (2016).
- [5] J. S. Lee and H. Park, *Sci. Rep.* **7**, 10725 (2017).
- [6] V. Holubec and A. Ryabov, *Phys. Rev. E* **96**, 030102 (2017).
- [7] M. Polettini and M. Esposito, *Europhys. Lett.* **118**, 40003 (2017).
- [8] V. Holubec and A. Ryabov, *Phys. Rev. E* **96**, 062107 (2017).
- [9] P. Pietzonka and U. Seifert, *Phys. Rev. Lett.* **120**, 190602 (2018).
- [10] K. Brandner, K. Saito, and U. Seifert, *Phys. Rev. X* **5**, 031019 (2015).
- [11] N. Shiraishi, K. Saito, and H. Tasaki, *Phys. Rev. Lett.* **117**, 190601 (2016).
- [12] N. Shiraishi and H. Tajima, *Phys. Rev. E* **96**, 022138 (2017).
- [13] A. Dechant and S.-i. Sasa, *Phys. Rev. E* **97**, 062101 (2018).
- [14] A. P. Solon and J. M. Horowitz, *Phys. Rev. Lett.* **120**, 180605 (2018).
- [15] N. Shiraishi, K. Funo, and K. Saito, *Phys. Rev. Lett.* **121**, 070601 (2018).
- [16] For example, the constant  $\chi = C/T_c\eta$  in the bound (97) in Ref. [13] diverges for vanishing relaxation time of the momentum  $1/\gamma$ , and the bound derived in Ref. [11] diverges for diverging transition rates [7,9].
- [17] O. Raz, Y. Subaşı, and C. Jarzynski, *Phys. Rev. X* **6**, 021022 (2016).
- [18] G. M. Rotskoff, *Phys. Rev. E* **95**, 030101 (2017).
- [19] S. Ray and A. C. Barato, *Phys. Rev. E* **96**, 052120 (2017).
- [20] A. C. Barato and U. Seifert, *Phys. Rev. Lett.* **114**, 158101 (2015).
- [21] P. Pietzonka, A. C. Barato, and U. Seifert, *Phys. Rev. E* **93**, 052145 (2016).
- [22] P. Pietzonka, F. Ritort, and U. Seifert, *Phys. Rev. E* **96**, 012101 (2017).
- [23] T. R. Gingrich, J. M. Horowitz, N. Perunov, and J. L. England, *Phys. Rev. Lett.* **116**, 120601 (2016).
- [24] J. M. Horowitz and T. R. Gingrich, *Phys. Rev. E* **96**, 020103 (2017).
- [25] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.121.120601> for the derivation of the work PDF for quasistatic processes, the derivation of Eqs. (3)–(6), and a detailed description of the cyclic Brownian HE, including Refs. [8,12,26–38].
- [26] V. Holubec, *Non-equilibrium Energy Transformation Processes*, Springer Theses (Springer International Publishing, New York, 2014).
- [27] T. Schmiedl and U. Seifert, *Europhys. Lett.* **81**, 20003 (2008).
- [28] T. Speck and U. Seifert, *J. Stat. Mech.* (2007) L09002.
- [29] M. Esposito, R. Kawai, K. Lindenberg, and C. Van den Broeck, *Phys. Rev. Lett.* **105**, 150603 (2010).
- [30] A. del Campo, *Phys. Rev. Lett.* **111**, 100502 (2013).
- [31] E. Torrontegui, S. Ibez, S. Martínez-Garaot, M. Modugno, A. del Campo, D. Gury-Odelin, A. Ruschhaupt, X. Chen, and J. G. Muga, in *Advances in Atomic, Molecular, and Optical Physics*, Advances In Atomic, Molecular, and Optical Physics, edited by E. Arimondo, P. R. Berman, and C. C. Lin (Academic Press, New York, 2013), Vol. 62, p. 117.
- [32] Z. C. Tu, *Phys. Rev. E* **89**, 052148 (2014).
- [33] V. Holubec, *J. Stat. Mech.* (2014) P05022.
- [34] V. Holubec and A. Ryabov, *Phys. Rev. E* **92**, 052125 (2015).
- [35] I. A. Martínez, A. Petrosyan, D. Guéry-Odelin, E. Trizac, and S. Ciliberto, *Nat. Phys.* **12**, 843 (2016).
- [36] L. Dinis, I. A. Martínez, É. Roldán, J. M. R. Parrondo, and R. A. Rica, *J. Stat. Mech.* (2016) 054003.
- [37] G. Li, H. T. Quan, and Z. C. Tu, *Phys. Rev. E* **96**, 012144 (2017).
- [38] V. Holubec, K. Kroy, and S. Steffenoni, arXiv:1804.01285.
- [39] I. A. Martínez, E. Roldán, L. Dinis, D. Petrov, and R. A. Rica, *Phys. Rev. Lett.* **114**, 120601 (2015).
- [40] D. Arold, A. Dechant, and E. Lutz, *Phys. Rev. E* **97**, 022131 (2018).
- [41] T. Speck and U. Seifert, *Phys. Rev. E* **70**, 066112 (2004).
- [42] J. Hoppenau and A. Engel, *J. Stat. Mech.* (2013) P06004.
- [43] A. C. Barato and U. Seifert, *Phys. Rev. X* **6**, 041053 (2016).
- [44] K. Sekimoto, *Stochastic Energetics*, Lecture Notes in Physics (Springer, Heidelberg, 2010), Vol. 799.
- [45] U. Seifert, *Rep. Prog. Phys.* **75**, 126001 (2012).
- [46] I. A. Martínez, É. Roldán, L. Dinis, and R. A. Rica, *Soft Matter* **13**, 22 (2017).
- [47] We leave aside the discussion whether the Hamiltonian  $H(x, t) = k(t)x^2/2$ , or rather the full Hamiltonian  $H(x, t) = k(t)x^2/2 + p^2/2m$  including the particle mass  $m$  and momentum  $p$  should be used for describing the thermodynamics of the overdamped particle [39,40]. It is not important for our demonstrative purposes and in both cases one can construct a Brownian HE operating close to  $\eta_C$  at  $P > 0$  with finite fluctuation.
- [48] V. Blickle and C. Bechinger, *Nat. Phys.* **8**, 143 (2012).
- [49] I. A. Martínez, É. Roldán, L. Dinis, D. Petrov, J. M. Parrondo, and R. A. Rica, *Nat. Phys.* **12**, 67 (2016).
- [50] R. Kosloff and Y. Rezek, *Entropy* **19**, 136 (2017).

# Supplemental Material

## Cycling tames power fluctuations near optimum efficiency

Viktor Holubec<sup>1,2,\*</sup> and Artem Ryabov<sup>1</sup>

<sup>1</sup>*Charles University, Faculty of Mathematics and Physics,*

*Department of Macromolecular Physics, V Holešovičkách 2, 180 00 Praha 8, Czech Republic*

<sup>2</sup>*Universität Leipzig, Institut für Theoretische Physik, Postfach 100 920, D-04009 Leipzig, Germany*

(Dated: August 23, 2018)

### WORK PDF FOR A QUASI-STATIC PROCESS

In this Section, we show that for any quasi-static process the PDF of work (defined through the time derivative of the Hamiltonian) converges to the delta function. The derivation generalizes the one given in Ref. [1] to any process  $x(t)$ , for which the PDF of  $x(t)$  evolves according to the equation

$$\partial_t \rho(x, t) = \nu \mathcal{L}(t)[\rho(x, t)]. \quad (\text{S1})$$

For a continuous Markovian dynamics the operator  $\mathcal{L}(t)$  is a linear Fokker-Planck operator and for a discrete one it is a transition rate matrix. However, Eq. (S1) can also be a generalized Master equation for a non-Markovian process [2].

During the operational cycle of the engine, the evolution operator  $\mathcal{L}(t)$  varies from  $\mathcal{L}(0)$  to  $\mathcal{L}(t_p)$  as both the temperature  $T(t)$  and the Hamiltonian  $H(x, t)$  change. Naturally, we assume that the system with the fixed Hamiltonian  $H$  in contact with a heat bath at constant temperature  $T$ , will eventually relax to the Boltzmann distribution  $\rho(x, \infty) = \rho_B(x) = \exp(-H/k_B T)/Z$ , where  $k_B$  denotes the Boltzmann constant and  $Z$  is the partition function. The Boltzmann distribution thus satisfies  $\mathcal{L}(t)[\rho_B(x, t)] = 0$ . In Eq. (S1), the relaxation time to equilibrium is measured by the prefactor  $\nu$ . The relaxation is fast (slow) for  $\nu$  large (small).

The process is quasi-static if the evolution operator  $\mathcal{L}(t)$  changes on a time-scale much longer than the relaxation time. Since we are interested in finite-time processes only (in order to obtain a non-zero output power of the engine), we consider the limit of infinitely fast relaxation  $\nu \rightarrow \infty$ . At any instant  $t$  during such a process, the PDF for  $x$  is given by the Boltzmann distribution

$$\rho(x, t) \approx \rho_B(x, t) = \frac{1}{Z(t)} \exp \left[ -\frac{H(x, t)}{k_B T(t)} \right], \quad (\text{S2})$$

where  $H(x, t)$  and  $T(t)$  are values of the Hamiltonian and temperature at time  $t$ . Of course, the formula (S2) is in general valid only if the system is in contact with the heat bath. Once the bath is disconnected, the evolution becomes deterministic, determined by the specific form of the Hamiltonian. The formula (S2) is thus valid along the whole cycle with adiabatic branches of the type (ii), where the system never disconnects from the baths (see

the main text). On the other hand, for systems with adiabatic branches of the type (i), where the system is disconnected from the bath, the formula (S2) holds during the isothermal branches only.

The work done by the system during the time interval  $[0, t]$  is defined as

$$w(t) = - \int_0^t d\tau \dot{H}[x(\tau), \tau]. \quad (\text{S3})$$

The dot denotes the partial derivative with respect to the time argument in  $H(x, t)$ ,  $\dot{H}(x, t) = \partial_t H(x, t)$ . We will now derive the probability distribution for this stochastic functional assuming that the underlying stochastic process  $x(t)$  is quasi-static with the PDF given by Eq. (S2).

The evolution equation for the joint probability density for  $x(t)$  and  $w(t)$  reads [2]

$$\partial_t \xi(x, w, t) = \dot{H}(x, t) \partial_w \xi(x, w, t) + \nu \mathcal{L}(x, t)[\xi(x, w, t)]. \quad (\text{S4})$$

The desired PDF for work,  $p(w, t)$ , obtained from  $\xi(x, w, t)$  by integration over  $x$ ,  $p(w, t) = \int dx \xi(x, w, t)$ , obeys the evolution equation

$$\partial_t p(w, t) = \int dx \dot{H}(x, t) \partial_w \xi(x, w, t) + \int dx \nu \mathcal{L}(t)[\xi(x, w, t)]. \quad (\text{S5})$$

The equation cannot be solved in general and its solution is known only in a few special cases [1]. Nevertheless, we are interested only in the quasi-static limiting case  $\nu \rightarrow \infty$ , where the equation can be solved by the Ansatz  $\xi(x, w, t) \approx \rho_B(x, t) p(w, t)$ . After inserting the Ansatz into Eq. (S5), the last term on the right-hand side vanishes due to the quasi-static condition  $\mathcal{L}(t)[\rho_B(x, t)] = 0$ , and Eq. (S5) reduces to the pure convection equation

$$\partial_t p(w, t) = \left[ \int dx \dot{H}(x, t) \rho_B(x, t) \right] \partial_w p(w, t), \quad (\text{S6})$$

with the initial condition  $p(w, 0) = \delta(w)$ , where  $\delta$  denotes the Dirac  $\delta$ -function (using the definition of stochastic work (S3) we assume that no work has been done before  $t = 0$  with probability 1). Solution to this convection equation is

$$p(w, t) = \delta(w - W(t)), \quad (\text{S7})$$



where

$$W(t) = - \int_0^t d\tau \int dx \dot{H}(x, \tau) \rho_B(x, \tau) \quad (\text{S8})$$

is the average work done by the system during the quasi-static process taking account during the time window  $[0, t]$ . The validity of the solution can be checked by the direct substitution of  $p(w, t)$  given by Eq. (S7) with  $W(t)$  given by Eq. (S8) into the Eq. (S6). Alternatively, the solution (S7) can be found for example by Laplace transform [1].

### DERIVATION OF FORMULAS (3)–(6) FOR CYCLIC HEAT ENGINES

The work PDF (S7) characterizes the statistics of work done by the periodically driven heat engine (HE) described in the main text only if the bath is not decoupled from the working medium during the adiabatic branches. Then the PDF  $p(w, t_p)$  for work done per cycle is given by Eq. (S7) with  $W(t_p)$  defined in Eq. (S8). Denoting as  $p(w) = p(w, t_p)$  the PDF for the work per cycle and as  $W = W(t_p)$  the average work done per cycle, we obtain Eq. (6) from the main text.

Let us now further derive the formulas (3)–(5) in the main text for the stochastic work  $w = - \int_0^{t_p} dt \dot{H}[x(t), t]$  done by periodically driven heat engines (HEs) disconnected from the reservoirs during the adiabatic branches. The considered Carnot cycle is quasi-static and the PDF for  $w$  can be calculated solely using the Boltzmann distribution (S2).

Due to the self-averaging property of work done during quasi-static processes where the system is in contact with the reservoirs, which we described in the preceding section, the work PDFs for the isothermal branches are given by  $\delta$ -functions located at the positions of the reversible works. For the hot isotherm, the reversible work can be calculated using the combination of the first and the second law of thermodynamics in the form  $W_1 = T_h \Delta S - \langle \Delta H_1 \rangle$ , where  $\Delta S = \Delta S_1$  and  $\langle \Delta H_1 \rangle$  denote the change of the system entropy and the change of the average internal energy of the system during the hot isotherm, respectively. The reversible work done during the cold isotherm can be calculated along similar lines. It reads  $W_3 = -T_c \Delta S - \langle \Delta H_3 \rangle$ . The minus sign before the entropy term follows from the fact that the change of the system entropy during the cold isotherm is given by  $\Delta S_3 = -\Delta S$  as follows from the condition that the change of the system entropy per cycle,  $\Delta S_1 + \Delta S_3$ , vanishes. To sum up, the PDF for work done along the hot isotherm reads  $p_1(w) = \delta(w - W_1)$  and the one for the cold isotherm is given by  $p_3(w) = \delta(w - W_3)$ .

During the adiabatic branches when the reservoir is disconnected from the system, the work lacks the self-averaging property of quasi-static processes which slightly

complicates the calculation of the corresponding PDFs. Since no heat can be interchanged during such adiabats, the stochastic work done during the adiabats is simply given by the decrease of the internal energy along these branches. Concretely, we get  $w_2 = -\Delta H_2$  for the first adiabat and  $w_4 = -\Delta H_4$  for the second one. The PDFs for these works are thus determined by the PDFs for the changes of energy. Formally, we can write the work PDF for the first adiabat as  $p_2(w) = \langle \delta(w + \Delta H_2) \rangle$  and as  $p_4(w) = \langle \delta(w + \Delta H_4) \rangle$  for the second one. These averages must be taken over the PDFs for  $\Delta H_2$  and  $\Delta H_4$ , respectively. Due to the quasi-staticity of the considered cycle, these PDFs are independent and can be constructed from the Boltzmann distribution (S2).

For infinitely fast adiabatic branches the microstate of the system does not change during the adiabats. Assuming that the system dwells in a microstate  $x$  at the beginning of the first adiabat and at a microstate  $y$  at the beginning of the second one, the changes of the internal energy are given by  $\Delta H_2 = H(x, t_1 + t_2) - H(x, t_1)$  and  $\Delta H_4 = H(y, t_p) - H(y, t_p - t_4)$ , where  $t_i$ ,  $i = 1, \dots, 4$  denote durations of the individual branches (see Fig. 1 in the main text). In this case, the average in the PDF  $p_2(w)$  must be calculated over the PDF  $\rho_B(x, t_1) = \rho_B(x, t_1 + t_2)$  and the one in  $p_4(w)$  over the PDF  $\rho_B(y, t_p - t_4) = \rho_B(y, t_p)$ .

If the reservoirs are disconnected from the system for a finite time interval, the system microstate and thus also the corresponding PDF during the adiabatic branches deterministically changes. In order to avoid bringing the system out of equilibrium after the ends of such adiabats, the driving must be chosen in such a way that the PDF for  $x$  just before the end of an adiabat is equivalent to the equilibrium PDF corresponding to the Hamiltonian and bath temperature after reconnecting the system and the bath when the adiabat ends. Let us denote as  $x_2$  the microstates occupied at the beginning of the first adiabat. Because the system evolution during the adiabat is deterministic, the microstate at its end,  $y_2 = y_2(x_2)$ , is a function of  $x_2$ . The corresponding change of the internal energy reads  $\Delta H_2 = H(y_2(x_2), t_1 + t_2) - H(x_2, t_1)$  and the average in the PDF  $p_2(w)$  must be calculated over the PDF  $\rho_B(x_2, t_1)$ . To get explicit results, it is necessary to know the specific form of the mapping  $y_2(x_2)$ , i.e. to solve concrete dynamical equations for the microstate during the adiabat.

To avoid this, we here consider two situations which can be treated without specifying the dynamics. First, one can drive the system in such a way that  $y_2(x_2) = x_2$ , and return to the situation of infinitely fast adiabatic branches. Second, one can use the approximation that  $x_2$  and  $y_2$  are independent of each other. In this case, the average in the PDF  $p_2(w)$  must be calculated over the PDF  $\rho_B(x_2, t_1) \rho_B(y_2, t_1 + t_2)$ . Using similar notation, the change of the internal energy during the second adiabat can be written as  $\Delta H_4 = H(y_4, t_p) - H(x_4, t_p - t_4)$ . If we

again use the approximation that  $x_4$  and  $y_4$  are independent, the average in the PDF  $p_4(w)$  must be calculated over the PDF  $\rho_B(x_4, t_p - t_4)\rho_B(y_4, t_p)$ .

The work PDF for the whole cycle is given by the convolution of the work PDFs for the individual branches,  $p(w) = p_1 \star p_2 \star p_3 \star p_4(w) = \langle \delta \{w - [W_1 + W_2 - \Delta H_2 - \Delta H_4]\} \rangle$ . Using the condition on zero change of the system average internal energy per cycle,  $\sum_{i=1}^4 \langle \Delta H_i \rangle = 0$ , the expression  $W_1 + W_2 - \Delta H_2 - \Delta H_4$  can be rewritten as  $(T_h - T_c)\Delta S - (\Delta H_2 - \langle \Delta H_2 \rangle) - (\Delta H_4 - \langle \Delta H_4 \rangle)$ . The PDF for the work done per cycle thus reads

$$p(w) = \left\langle \delta \left\{ w - \left[ W - \widetilde{\Delta H}_2 - \widetilde{\Delta H}_4 \right] \right\} \right\rangle, \quad (\text{S9})$$

where  $\widetilde{\Delta H}_i = \Delta H_i - \langle \Delta H_i \rangle$ ,  $i = 2, 4$  and  $W = (T_h - T_c)\Delta S$ . Integrals over this function yield all moments of work:  $\langle w^m \rangle = \int_{-\infty}^{\infty} dw w^m p(w)$ . The first one ( $m = 1$ ) is given by

$$\langle w \rangle = W = (T_h - T_c)\Delta S \quad (\text{S10})$$

and the second one ( $m = 2$ ) reads

$$\langle w^2 \rangle = W^2 + \langle \Delta H_2^2 \rangle - \langle \Delta H_2 \rangle^2 + \langle \Delta H_4^2 \rangle - \langle \Delta H_4 \rangle^2. \quad (\text{S11})$$

Equations (S9)–(S11) are valid for an arbitrary Carnot cycle with adiabatic branches where the system and reservoir are disconnected. To get specific expressions for specific models, the averages must be taken over the proper PDFs for  $\Delta H_2$  and  $\Delta H_4$ , as described above.

For the instantaneous adiabatic branches, the changes in the internal energy assumes the form  $\Delta H_2 = H(x, t_1 + t_2) - H(x, t_1)$  and  $\Delta H_4 = H(y, t_p) - H(y, t_p - t_4)$  and the averages must be taken over the PDF  $\rho_B(x, t_1)\rho_B(y, t_p)$ . Using the Hamiltonian  $H(x, t) = k(t)x^{2n}/2n$ , Eqs. (S10) and (S11) gives the work variance

$$\sigma_w^2 = \langle w^2 \rangle - W^2 = \frac{k_B^2}{n}(T_h - T_c)^2 \quad (\text{S12})$$

which, together with  $W = (T_h - T_c)\Delta S$ , implies Eq. (4) in the main text for the work fluctuation.

For the finite time adiabats, under the assumption that the microstates occupied at the beginnings and at the ends of the adiabats are independent, the changes in the internal energy assumes the form  $\Delta H_2 = H(y_2, t_1 + t_2) - H(x_2, t_1)$  and  $\Delta H_4 = H(y_4, t_p) - H(x_4, t_p - t_4)$  and the averages must be taken over the PDF  $\rho_B(x_2, t_1)\rho_B(y_2, t_1 + t_2)\rho_B(x_4, t_p - t_4)\rho_B(y_4, t_p)$ . Using the Hamiltonian  $H(x, t) = k(t)x^{2n}/2n$ , we obtain from Eqs. (S10) and (S11) the formula

$$\sigma_w^2 = \frac{k_B^2}{n}(T_h + T_c)^2 \quad (\text{S13})$$

for the work variance and the formula (5) in the main text for the work fluctuation.

## CYCLIC BROWNIAN HEAT ENGINE

The HE is based on an overdamped Brownian particle diffusing in a harmonic potential

$$U(x, t) = H(x, t) = k(t)x^2/2. \quad (\text{S14})$$

Its position  $x = x(t)$  obeys the Langevin equation

$$\dot{x} = -kx/\gamma + \sqrt{2k_B T/\gamma}\zeta. \quad (\text{S15})$$

Here,  $\gamma$  is the friction coefficient and  $\zeta$  is the Gaussian white noise with  $\langle \zeta \rangle = 0$  and  $\langle \zeta(t)\zeta(t') \rangle = \delta(t - t')$ . The bath temperature  $T$  and the stiffness  $k$  are varied along the Carnot cycle depicted in Fig. 1 in the main text. Specifically, we use the stiffness protocol

$$k(t) = \begin{cases} \frac{1}{\sigma_0^2} \frac{k_B T_h}{(1+b_1 t)^2} - \frac{\gamma b_1}{1+b_1 t}, & t \in [0, t_1), \\ \frac{1}{\sigma_f^2} \frac{k_B T_c}{[1+b_2(t-t_1)]^2} - \frac{\gamma b_2}{1+b_2(t-t_1)}, & t \in [t_1, t_p), \end{cases} \quad (\text{S16})$$

maximizing the work done by the engine per cycle once the system entropy change during the hot isotherm  $\Delta S = k_B \log \sigma_f/\sigma_0$  and durations  $t_1$  and  $t_3$  of the two isotherms are fixed [3, 4]. In Eq. (S16), the parameter  $\sigma_0^2$  ( $\sigma_f^2$ ) stands for the variance of the particle position at the beginning (end) of the hot isotherm. The adiabatic branches are assumed to be instantaneous ( $t_2 = t_4 = 0$ ) and thus  $t_p = t_1 + t_3$  denotes the duration of the whole cycle. The constants  $b_1$  and  $b_2$  are given by  $b_1 = (\sigma_f/\sigma_0 - 1)/t_1$  and  $b_2 = (\sigma_0/\sigma_f - 1)/t_3$ . For  $t_1$  and  $t_3$  much larger than the system relaxation time  $\tau_x = k/\gamma$ , the driving produces a quasi-static process even though it changes infinitely fast during the adiabats. This is because the jumps of the stiffness during these branches balance the jumps in the temperature keeping constant the ratio  $k/T$  and thus also the equilibrium position distribution  $\rho(x) = \exp(-kx^2/2k_B T)/Z_x$ .

Besides the protocol (S16) is optimal, it also leads to concise expressions for power and efficiency [3, 5]

$$P = \frac{(T_h - T_c)\Delta S}{t_p} - \frac{A}{t_1 t_p} - \frac{A}{t_3 t_p}, \quad (\text{S17})$$

$$\eta = \frac{W_{\text{out}}}{Q_h} = \frac{\eta_C}{1 + T_c \Delta S_{\text{tot}}/W}, \quad (\text{S18})$$

where the parameter  $A = \gamma(\sigma_f - \sigma_0)^2$  measures the work dissipated due to the irreversible realization of the cycle for  $t_1 \approx t_3 \approx \tau_x$  and the entropy produced per cycle  $\Delta S_{\text{tot}} = A(1/T_h t_1 + 1/T_c t_3)$  assumes the so called low-dissipation form [3, 6]. General calculation of the work/power fluctuation  $\tilde{\sigma}_w$  is more involved. One can either calculate it with the help of Brownian dynamics simulation of the Langevin equation, or numerically, for example using the method suggested in Ref. [7]. We use the analytical method presented in Ref. [4].

In the considered setting, the Carnot efficiency is reached in the limit of infinitely fast cycles with infinitely

$t_p$	$\langle w \rangle$	$\langle w^2 \rangle$	$P$	$\langle P^2 \rangle$	$\Delta S_{tot}$	$\Delta S_{tot}/t_p$	$\eta_C - \eta$
$\sigma_\infty^{(\chi-1)\xi}$	$\sigma_\infty^0$	$\sigma_\infty^0$	$\sigma_\infty^{(1-\chi)\xi}$	$\sigma_\infty^{2(1-\chi)\xi}$	$\sigma_\infty^{-\chi\xi}$	$\sigma_\infty^{(1-2\chi)\xi}$	$\sigma_\infty^{-\chi\xi}$

TABLE I. Scalings of the main variables in the Brownian HE with the harmonic Hamiltonian (S14). The engine can work with the Carnot efficiency at nonzero power with finite fluctuation whenever  $\xi > 0$  and  $\chi \in (0, 1]$ .

strong driving [8]. More precisely, introducing the scaling

$$\sigma_f \propto \sigma_\infty^{-\xi}, \quad t_p \propto \sigma_\infty^{(\chi-1)\xi} \quad (\text{S19})$$

while keeping constant the ratio  $\sigma_f/\sigma_0 > 0$  yields in the limit  $\sigma_\infty$  the Carnot efficiency at a nonzero power for  $\xi > 0$  and  $\chi \in (0, 1]$ . Under the scaling (S19) the position relaxation time  $\tau_x = \gamma/k \propto \sigma_\infty^{-\xi}$  vanishes faster than the cycle duration  $t_p \propto \sigma_\infty^{(\chi-1)\xi}$ , and thus, although fast, the cycle can be regarded as quasi-static. Different from the shortcuts to adiabaticity and equilibrium presented in the literature [9–13], the fast relaxation is caused by decreasing the system relaxation time  $\tau_x$  by increasing the stiffness  $k$  rather than by devising a clever protocol.

In Tab. I, we sum scalings of the most important parameters of the Brownian HEs with  $\sigma_\infty$ . Let us note that the cycle duration  $t_p^*$ , corresponding to the maximum power  $P^*$  attainable in the HE once all its parameters except for  $t_p$  are fixed, scales as  $1/k$  and thus it is proportional to the position relaxation time  $\tau_x$ . Interestingly, the maximum power  $P^*$  is always much larger than  $P$  at  $\eta_C$  [8].

According to Ref. [8], the overdamped Brownian HEs work with  $\eta = \eta_C$  and  $P > 0$  whenever they operate in the low-dissipation regime, where the work dissipated during the hot (cold) isotherm scales as  $1/t_1$  ( $1/t_3$ ). This is achieved if the state of the system depends during the hot (cold) isotherm on time solely through the combination  $t/t_1$  ( $t/t_3$ ). In fact, the Carnot bound at  $P > 0$  can be reached under much milder conditions of: 1) Short cycle times  $t_p$  which are nevertheless long compared to the relaxation times for position  $\tau_x$  and for the momentum  $\tau_p = m/\gamma$ . In practice, the time-scale separation  $t_p \gg \tau_x, \tau_p$  must be realized in such a way that the time-scales  $\tau_x$  and  $\tau_p$  are much larger than the reservoir relaxation time [14]. 2) Driving leading to nonzero reversible work  $W = (T_h - T_c)\Delta S$ . 3) A Carnot cycle composed of two adiabats and two isotherms.

The curves plotted in Fig. 2 in the main text are calculated for a round glass particle with the density  $\rho_g = 2800$  kg/m<sup>3</sup> and radius  $R = 10^{-6}$  m diffusing in water with dynamical viscosity  $\kappa = 0.001$  Pa s, which is assumed to be temperature independent. The friction coefficient  $\gamma = 6\pi R\kappa$ , is calculated according to the Stokes law.

We take  $\sigma_f = c\sigma_\infty^{-\xi}$ ,  $c = 1$  m, and  $\sigma_f^2/\sigma_0^2 = 3$  for the boundary variances,  $t_1 = t_3 = t_p/2$  for durations of the isothermal branches,  $T_c = 293.15$  K and  $T_h = 5273.15$  K for the reservoir temperatures and  $\xi = 1.5$  and  $\chi = 0.05$  for the exponents defined in Eq. (S19).

The values of parameters taken are readily realizable in experiments with optical tweezers as can be checked by comparing their values with the values used in the experiments reported in Refs. [15, 16]. Especially, we point out that the seemingly unrealistically large hot bath temperature  $T_h = 5273.15$  K can be realized by applying on the particle a further electrostatic force that mimics a thermal bath much hotter than the actual temperature of the water surrounding the particle. The hottest effective temperature achieved using this method in Ref. [16] was 6000 K.

---

\* viktor.holubec@gmail.com

- [1] V. Holubec, *Non-equilibrium Energy Transformation Processes*, Springer Theses (Springer International Publishing, 2014).
- [2] T. Speck and U. Seifert, *J. Stat. Mech: Theory Exp.* **2007**, L09002 (2007).
- [3] T. Schmiedl and U. Seifert, *EPL* **81**, 20003 (2008).
- [4] V. Holubec, *J. Stat. Mech: Theory Exp.* **2014**, P05022 (2014).
- [5] V. Holubec and A. Ryabov, *Phys. Rev. E* **92**, 052125 (2015).
- [6] M. Esposito, R. Kawai, K. Lindenberg, and C. Van den Broeck, *Phys. Rev. Lett.* **105**, 150603 (2010).
- [7] V. Holubec, K. Kroy, and S. Steffenoni, arXiv preprint arXiv:1804.01285 (2018).
- [8] V. Holubec and A. Ryabov, *Phys. Rev. E* **96**, 062107 (2017).
- [9] A. del Campo, *Phys. Rev. Lett.* **111**, 100502 (2013).
- [10] E. Torrontegui, S. Ibez, S. Martínez-Garaot, M. Modugno, A. del Campo, D. Gury-Odelin, A. Ruschhaupt, X. Chen, and J. G. Muga, in *Advances in Atomic, Molecular, and Optical Physics*, Advances In Atomic, Molecular, and Optical Physics, Vol. 62, edited by E. Arimondo, P. R. Berman, and C. C. Lin (Academic Press, 2013) pp. 117 – 169.
- [11] Z. C. Tu, *Phys. Rev. E* **89**, 052148 (2014).
- [12] G. Li, H. T. Quan, and Z. C. Tu, *Phys. Rev. E* **96**, 012144 (2017).
- [13] I. A. Martínez, A. Petrosyan, D. Guéry-Odelin, E. Trizac, and S. Ciliberto, *Nat. Phys.* **12**, 843 (2016).
- [14] N. Shiraishi and H. Tajima, *Phys. Rev. E* **96**, 022138 (2017).
- [15] I. A. Martínez, É. Roldán, L. Dinis, D. Petrov, J. M. Parrondo, and R. A. Rica, *Nat. Phys.* **12**, 67 (2016).
- [16] L. Dinis, I. A. Martínez, É. Roldán, J. M. R. Parrondo, and R. A. Rica, *J. Stat. Mech: Theory Exp.* **2016**, 054003 (2016).



## Optimal finite-time heat engines under constrained control

Zhuolin Ye <sup>1,\*</sup>, Federico Cerisola,<sup>2,3</sup> Paolo Abiuso <sup>4,5</sup>, Janet Anders <sup>3,6</sup>, Martí Perarnau-Llobet <sup>5</sup> and Viktor Holubec <sup>7,†</sup>

<sup>1</sup>*Institut für Theoretische Physik, Universität Leipzig, Postfach 100 920, D-04009 Leipzig, Germany*

<sup>2</sup>*Department of Materials, University of Oxford, Parks Road, Oxford OX1 3PH, United Kingdom*

<sup>3</sup>*Department of Physics and Astronomy, University of Exeter, Stocker Road, Exeter EX4 4QL, United Kingdom*

<sup>4</sup>*ICFO-Institut de Ciències Fotòniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain*

<sup>5</sup>*Department of Applied Physics, University of Geneva, 1211 Geneva, Switzerland*

<sup>6</sup>*Institut für Physik und Astronomie, University of Potsdam, 14476 Potsdam, Germany*

<sup>7</sup>*Department of Macromolecular Physics, Faculty of Mathematics and Physics, Charles University, V Holešovičkách 2, CZ-180 00 Praha, Czech Republic*



(Received 25 February 2022; accepted 24 October 2022; published 22 November 2022)

We optimize finite-time stochastic heat engines with a periodically scaled Hamiltonian under experimentally motivated constraints on the bath temperature  $T$  and the scaling parameter  $\lambda$ . We present a general geometric proof that maximum-efficiency protocols for  $T$  and  $\lambda$  are piecewise constant, alternating between the maximum and minimum allowed values. When  $\lambda$  is restricted to a small range and the system is close to equilibrium at the ends of the isotherms, a similar argument shows that this protocol also maximizes output power. These results are valid for arbitrary dynamics. We illustrate them for an overdamped Brownian heat engine, which can experimentally be realized using optical tweezers with stiffness  $\lambda$ .

DOI: [10.1103/PhysRevResearch.4.043130](https://doi.org/10.1103/PhysRevResearch.4.043130)

### I. INTRODUCTION

The unprecedented improvement in experimental control over microscopic Brownian [1] and quantum systems [2–4] has induced a revolution in the study of heat engines [5,6]. It aims to generalize equilibrium and finite-time thermodynamics [7–15] to the nanoscale, where thermal and quantum fluctuations render thermodynamic variables such as work and heat stochastic [16]. Intense effort is devoted to uncover optimal performance of stochastic heat engines [16–41]. However, optimal control protocols are only known under approximations of fast [34–36] or slow [28,37–41] driving, or for specific microscopic models: engines based on overdamped Brownian particles in harmonic [24] or log-harmonic [42] potential, and underdamped harmonic Brownian heat engines [43]. Furthermore, most of these exact results are obtained under constraints on the state of the working medium [44], instead of experimentally motivated constraints on the control parameters [45,46]. An exception is Ref. [47], showing that reaching maximum efficiency of slowly driven cyclic heat engines requires control over the scaling of the full Hamiltonian to avoid heat leakages.

In this paper, we optimize finite-time thermodynamic cycles under constraints on control parameters such as trap stiffness of optical tweezers  $\lambda$  and bath temperature  $T$ . We show that, different from constraining the response such as the width  $\sigma$  of the phase distribution, constraining the control allows for surprisingly simple and general derivation of maximum-efficiency and maximum-power protocols. Besides other stark differences, for constrained control of Brownian heat engines, these protocols significantly outperform the protocol optimized for power and efficiency under constraints on  $\sigma$  [24].

The paper is organized as follows. In Sec. II, we introduce the considered setup with a periodically scaled Hamiltonian under experimentally motivated constraints. In Sec. III, we derive the corresponding maximum-efficiency protocol. In Sec. IV, we prove that the maximum-efficiency protocol yields, under certain conditions, also maximum output power. In Sec. V, we present a case study of optimization of power and efficiency for constrained control by considering a specific overdamped Brownian heat engine. Besides illustrating the general results derived in Secs. III and IV, we provide numerical evidence that the maximum-power protocol is, in this case, piecewise linear. We conclude in Sec. VI.

### II. SETUP

Following Ref. [47], we assume that the Hamiltonian of the system that serves as a working medium of the stochastic heat engine is of the form

$$H(x, t) = \lambda(t)f(x), \quad (1)$$

\*zhuolinYe@foxmail.com

†viktor.holubec@mff.cuni.cz

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

where the control parameter  $\lambda(t)$  periodically expands and shrinks the energy spectrum in time, and  $f(x)$  is an arbitrary function of the system degrees of freedom  $x$  such that the equilibrium partition function  $Z(t) = \int dx \exp[-H(x, t)/(k_B T)]$  is finite for all  $k_B T \geq 0$  ( $k_B$  denotes the Boltzmann constant). This class of Hamiltonians generalizes the well-known “breathing” parabola model [24] for an overdamped particle trapped in a parametrically driven harmonic potential. It also includes semiclassical two-level (or multilevel) systems with controlled gaps between the individual energy levels [16], and quantum spins, where the control parameter is an externally controlled magnetic field [17].

We connect the system to a heat bath and periodically alter its temperature  $T(t)$  with the same finite period  $t_p$  as  $\lambda(t)$ . The parameters under experimental control are thus  $\lambda(t)$  and  $T(t)$  and our aim is to find optimal  $t_p$ -periodic protocols for them under the experimentally motivated constraints [48]

$$\lambda(t) \in [\lambda_-, \lambda_+], \quad T(t) \in [T_-, T_+]. \quad (2)$$

### III. MAXIMUM-EFFICIENCY PROTOCOL

Our first main result is a general geometric proof that the maximum-efficiency finite-time cycle under the constraints (2) is a Carnot-Otto cycle composed of two isotherms/isochores interconnected by two adiabats. The maximum-efficiency protocol  $\{T(t), \lambda(t)\}$  is thus piecewise constant:

$$\{T(t), \lambda(t)\}_\eta = \begin{cases} \{T_+, \lambda_+\}, & 0 < t < t_+, \\ \{T_-, \lambda_-\}, & t_+ < t < t_p. \end{cases} \quad (3)$$

And the maximum efficiency is given by

$$\eta = 1 - \frac{\lambda_-}{\lambda_+}. \quad (4)$$

The proof relies just on the definition of heat and it is thus independent of the details of the system dynamics, including the times  $t_+$  and  $t_p$ . It holds both for situations when the heat bath is memoryless (Markovian) and non-Markovian. The nonequilibrium dynamics of the system communicating with a Markovian bath can be described by Fokker-Planck or master equations for the probability density for  $x$  [49]. Except for a few exactly solvable settings [16,49], these equations are usually hard to solve analytically for non-quasi-static time-dependent protocols. However, in the non-Markovian case, a corresponding closed deterministic description might not be available at all [50]. Then one has to resort to stochastic descriptions, such as a generalized Langevin equation, making even a numerical optimization challenging. The derivation also holds in situations with a nonequilibrium bath, such as in recently intensely studied cyclic active Brownian heat engines [51–54].

Let us now derive Eqs. (3) and (4). Under reasonable assumptions, any periodic variation of the control parameters eventually induces a periodic average response of the system,  $\sigma(t) = \langle f[x(t)] \rangle$ . This ensemble average is a functional of  $T(t)$  and  $\lambda(t)$  specified by dynamical equations of the system. Due to the factorized structure of the Hamiltonian (1), the average internal energy of the system  $\langle H(x, t) \rangle$  is given by  $\lambda(t)\sigma(t)$ . Decomposing its infinitesimal change into a component corresponding to the external variation of the control  $\lambda$

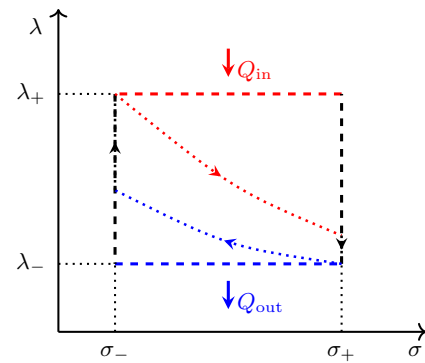


FIG. 1. The maximum-efficiency protocol (3) under the constraints in Eq. (2) (dashed line) compared to a suboptimal cycle (dotted line).

(work) and the rest (heat) [5,6], it follows that output work and input heat increments are given by  $dW_{\text{out}}(t) = -\sigma(t)d\lambda(t)$  and  $dQ(t) = \lambda(t)d\sigma(t)$ , respectively. Per cycle, the engine transforms the fraction

$$\eta = \frac{W_{\text{out}}}{Q_{\text{in}}} = 1 - \frac{Q_{\text{out}}}{Q_{\text{in}}} \quad (5)$$

of the heat

$$Q_{\text{in}} = \int_0^{t_p} \lambda(t)\theta[d\sigma(t)]d\sigma(t) \quad (6)$$

from the heat source into output work

$$W_{\text{out}} = - \int_0^{t_p} \sigma(t)d\lambda(t), \quad (7)$$

and dumps the remaining heat  $Q_{\text{out}} = Q_{\text{in}} - W_{\text{out}} = \int_0^{t_p} \lambda(t)\theta[-d\sigma(t)]d\sigma(t)$  into the heat sink. [The Heaviside step function  $\theta(\bullet) = 1$  when the heat flows on average into the system, i.e.,  $d\sigma > 0$ .]

Consider now the  $\lambda$ - $\sigma$  diagram of the cycle depicted in Fig. 1. We seek the shape of the cycle which yields maximum efficiency  $\eta$  under the constraints (2).<sup>1</sup> The cycle must run clockwise to secure that  $Q_{\text{in}} > Q_{\text{out}}$ . Next, we note that maximizing  $\eta$  amounts to minimizing the ratio  $Q_{\text{out}}/Q_{\text{in}}$ . For given boundary values  $\sigma_{\pm}$  of  $\sigma$ , this is obviously achieved by setting  $\lambda = \lambda_+$  when  $d\sigma > 0$  and  $\lambda = \lambda_-$  when  $d\sigma < 0$ . In such a case,  $Q_{\text{in}} = \lambda_+\Delta\sigma$ ,  $Q_{\text{out}} = \lambda_-\Delta\sigma$ , and the efficiency is given by Eq. (4). The increase in the system response  $\Delta\sigma = \sigma_+ - \sigma_-$ , which can be a complicated functional of the protocol  $\{T(t), \lambda(t)\}$ , canceled out. Equation (4) is thus valid for arbitrary  $\sigma_{\pm}$ , and it represents the maximum efficiency of a heat engine based on Hamiltonian (1) under the constraints (2). The corresponding maximum-efficiency protocol for  $\lambda$  forms a rectangle ranging from  $\lambda_-$  to  $\lambda_+$  in the  $\lambda$ - $\sigma$  diagram regardless the cycle duration and dynamical equations of the

<sup>1</sup>A similar optimization problem is often solved in courses on classical thermodynamics to show that maximum efficiency of an equilibrium cycle under the constraint  $T(t) \in [T_-, T_+]$  on the bath temperature is the Carnot efficiency. However, in our case, the system can be arbitrarily far from equilibrium.

system. The only constraint on these control parameters is that the cycle runs in the  $\lambda$ - $\sigma$  diagram clockwise.

When not driven, a system out of equilibrium relaxes towards the equilibrium state corresponding to the instantaneous values of the fixed control parameters. For cyclically varied control parameters, the system can no longer relax to equilibrium and its nonequilibrium state “lags behind” the quasistatic cycle specified by the instantaneous values of the control parameters. In our setting,  $\sigma(t)$  lags behind  $\sigma^{\text{eq}}(t) = \int dx f(x) \exp\{-\lambda(t)f(x)/[k_B T(t)]\}/Z(t)$ . In Appendix B 1, we show that  $\sigma^{\text{eq}}(t)$  is a monotonically increasing function of  $T/\lambda$ . Denoting as  $t_+$  the duration of the  $\lambda = \lambda_+$  branch, clockwise cycles with  $\Delta\sigma > 0$  are thus obtained for temperature protocols  $T(t)$  which obey (i)  $\dot{T}(t) \geq 0$  when  $\lambda = \lambda_+$ , (ii)  $\dot{T}(t) \leq 0$  when  $\lambda = \lambda_-$ , and (iii)  $T(t_+ -)/\lambda_+ > T(t_p -)/\lambda_-$ , where  $T(t-) \equiv \lim_{\epsilon \rightarrow 0} T(t - |\epsilon|)$ . The last condition implies that the maximum efficiency (4) obeys the standard second-law inequality  $\eta \leq 1 - T(t_p -)/T(t_+ -) \leq 1 - T_-/T_+$ . It saturates for the “compression ratio”  $\lambda_-/\lambda_+ = T_-/T_+$ . Even for a finite cycle time  $t_p$ , output power, in this case, vanishes because  $\sigma^{\text{eq}}(t)$  becomes constant, yielding an infinitesimal quasistatic cycle with a vanishing output work. In the maximum-efficiency protocol (3), we use the specific protocol for  $T(t)$  that maximizes the upper bound on  $\eta$ . In Appendix B 1, we argue that this temperature protocol also maximizes the output work of the engine regardless of  $\lambda(t)$  because it yields the largest temperature differences between the bath and the system when they exchange heat. However, we reiterate that the maximum efficiency (4) can be achieved for an arbitrary protocol for  $T(t)$  that obeys the above conditions (i)-(iii). This freedom in  $T(t)$  can be exploited in setups where precise control of the bath (effective) temperature is difficult, such as in active Brownian heat engines [52].

The adiabatic branches connecting the isotherms in the protocol (3) can be realised using several qualitatively different approaches [31]. (i) One can disconnect the system from the heat bath, which might be impractical for microscopic engines. (ii) One may keep the system in thermal contact with the bath and vary the control parameters  $T$  and  $\lambda$  in such a way that the response  $\sigma$  does not change [55]. This approach allows circumventing some of the shortcomings of overdamped thermodynamics [56], where the heat fluxes through the momentum degrees of freedom are neglected. (iii) One can realize the adiabatic branches by changing the control parameters much faster than the relaxation time of the response  $\sigma$  [57]. In the specific maximum-efficiency protocol (3), we employ the last possibility. It minimizes the cycle time  $t_p$  and thus maximize the output power  $P \equiv W_{\text{out}}/t_p$ . Besides, it allows for a direct comparison with the maximum-efficiency protocols derived for Brownian heat engines under constraints on  $\sigma$  [24]. However, other realisations of the adiabatic branches yield the same maximum efficiency (4). We reiterate that also the choice of the durations  $t_+$  and  $t_p - t_+$  of the isotherms in (3) do not affect the maximum  $\eta$ .

#### IV. MAXIMUM-POWER PROTOCOL

If the durations of the isotherms are long enough compared to the relaxation time of the system, i.e.,  $\Delta\sigma$  is close to its equilibrium value, and the compression ratio  $\lambda_-/\lambda_+$  is large,

the maximum-efficiency protocol (3) also yields maximum output work  $W_{\text{out}}$  (7) and power

$$P = \frac{W_{\text{out}}}{t_p} \quad (8)$$

under the constrained control (2). This is our second main result. To prove it, consider the generally unreachable geometric loose upper bound on the output work  $\max W_{\text{out}} = \Delta\lambda \max \Delta\sigma^{\text{eq}} = (\lambda_+ - \lambda_-)[\sigma^{\text{eq}}(T_+/\lambda_-) - \sigma^{\text{eq}}(T_-/\lambda_+)]$ , which follows from the broadly valid assumption  $\max \Delta\sigma < \max \Delta\sigma^{\text{eq}}$  and the insight that  $W_{\text{out}}$  is given by the area enclosed by the cycle in the  $\lambda$ - $\sigma$  diagram. Expanding  $\max W_{\text{out}}$  in  $\Delta\lambda$  yields  $\max W_{\text{out}} = \Delta\lambda[\sigma^{\text{eq}}(T_+/\lambda_+) - \sigma^{\text{eq}}(T_-/\lambda_-)] + \mathcal{O}(\Delta\lambda^2)$ . Up to the leading order in  $\Delta\lambda$  and under the condition that the system has relaxed at the ends of the two isotherms to equilibrium, this upper bound is saturated by the protocol (3), which completes the proof. We note that (i) the condition  $\Delta\sigma = \Delta\sigma^{\text{eq}}$  does not mean that the cycle is slow as the system has to be close to equilibrium at the ends of the two isotherms only and can be arbitrarily far from equilibrium otherwise. (ii) This condition allows one to analytically calculate the whole probability distribution for the output work regardless of additional details of the system dynamics [16,58]. Interestingly, for semiclassical systems, piecewise constant protocols with two or more branches also maximize output power when the cycle time is much shorter than the system relaxation time [34,35,59].

Beyond these regimes,  $W_{\text{out}}$  and  $P$  strongly depend on all details of the dynamics through  $\sigma(t)$  and cycle time  $t_p$ . While  $W_{\text{out}}$  and  $P$  are still optimized by the temperature protocol and the choice of fast adiabats in (3), optimal protocols for  $\lambda(t)$  under the constraints (2) are no longer piecewise constant and they have to be identified for each system separately. Similarly to the derivation of maximum-efficiency and maximum-power protocols under constraints on the system state [24,42–44], this often involves functional optimization or extensive numerical work which are both nontrivial tasks.

In the next section, we illustrate the main features of maximum-efficiency and maximum-power protocols under constrained control on an engine based on an overdamped Brownian particle in a harmonic potential. This model describes experimental realizations of microscopic heat engines using optical tweezers [57,60,61]. Besides, the corresponding maximum-efficiency and maximum-power protocols under the constrained response are known [24], allowing for a direct comparison with our results obtained under constrained control.

#### V. CASE STUDY: OVERDAMPED BROWNIAN HEAT ENGINE

Let us now consider the specific Brownian heat engine based on an overdamped particle with mobility  $\mu$  diffusing in a controlled harmonic potential. The Hamiltonian (1) now reads  $H(x, t) = \lambda(t)x^2/2$ , with  $x$  the position of the particle. The response of the system  $\sigma(t) = \langle x^2/2 \rangle$  is proportional to the position variance and it obeys the first-order differential equation [16,24,62]

$$d\sigma(t)/dt = -2\mu\lambda(t)\sigma(t) + \mu k_B T(t). \quad (9)$$

TABLE I. Considered classes of protocols with free parameters  $a, b, c, d$  to be determined by the optimization (protocols  $\lambda_{\text{pwc}}$  and  $\lambda_{\text{S}}$  have only two free parameters). The protocols are in general discontinuous at times  $t_+$  and  $t_p$ . The piecewise constant protocol  $\lambda_{\text{pwc}}(t)$  is a variant of the maximum-efficiency protocol  $\lambda_{\eta}(t)$  (3), where  $\lambda_{\text{pwc}}(t)$  does not have to reach the boundary values  $\lambda_-$  and  $\lambda_+$ . The piecewise linear protocol  $\lambda_{\text{pwl}}(t)$  has zero curvature. The protocol  $\lambda_{\text{slow}}(t)$  minimizes the irreversible losses during isothermal branches under close-to-equilibrium conditions. Such protocols can be derived for Brownian heat engines with Hamiltonians of the form  $\lambda(t)x^n/n$  (for details, see Appendix C). The protocol  $\lambda_{\text{S}}(t)$  maximizes both power and efficiency under the constraint that  $\sigma(0) = \sigma(t_p) \equiv a$  and  $\sigma(t_+) \equiv b$  [24]. The corresponding response  $\sigma_{\text{S}}(t)$  is given by Eq. (D2). For  $b = d = 0$ ,  $\lambda_{\text{pwl}}$  and  $\lambda_{\text{slow}}$  reduce to  $\lambda_{\text{pwc}}$ .

	$\lambda_{\text{pwc}}(t)$	$\lambda_{\text{pwl}}(t)$	$\lambda_{\text{slow}}(t)$	$\lambda_{\text{S}}(t)$
$t < t_+$	$a$	$a + bt$	$\frac{a}{(1+bt)^2}$	$\frac{T_+}{2\sigma_{\text{S}}(a,b,t)} - \frac{\sqrt{b}-\sqrt{a}}{\mu t + \sqrt{\sigma_{\text{S}}(a,b,t)}}$
$t > t_+$	$c$	$c + dt$	$\frac{c}{(1+dt)^2}$	$\frac{T_-}{2\sigma_{\text{S}}(a,b,t)} + \frac{\sqrt{b}-\sqrt{a}}{\mu t - \sqrt{\sigma_{\text{S}}(a,b,t)}}$

In Sec. III, we proved that the maximum-efficiency protocol under the constraints (2) should, in this case, be the protocol (3). In this section, we illustrate this results by direct numerical optimization. In addition, we ask which protocol for  $\lambda$  yields the largest output power under the constraints (2). Even though the model (9) is exactly solvable [16], the corresponding optimal  $\lambda(t)$  has to be found numerically, e.g., by the method in Ref. [63]. To keep the optimization transparent, we instead consider the specific set of families of protocols for the isothermal strokes in Table I and numerically optimize over their free parameters. When such classes are chosen suitably, the resulting suboptimal performance will be close to the global optimum [64,65]. Besides, we use the protocol for temperature and adiabatic branches from Eq. (3), and fix the durations of the two isotherms and thus  $t_p$ . The durations can be further optimized once the optimal variation of  $\lambda$  is known. The solutions to Eq. (9) can involve exponentials of very large or small numbers, which can lead to numerical instabilities inducing large losses of precision, and thus they have to be treated with care. To secure that our solutions are always precise enough, we have solved Eq. (9) in our analysis also numerically.

For the protocols in Table I and the temperature protocol in Eq. (3), we thus numerically optimized the efficiency (5) and output power (8) as functions of the parameters  $\{a, b, c, d\}$  under constraints on  $\lambda(t)$ . For constrained response  $\sigma(t)$ , we additionally verified in Appendix D that the protocol  $\lambda_{\text{S}}$  obtained from Ref. [24] indeed yields both the maximum power and maximum efficiency.

The results of optimizing efficiency under the constrained control are depicted in Fig. 2. For all of the trial protocols from Table I except for  $\lambda_{\text{S}}$  the optimal values of parameters  $b$  and  $d$  were 0. All these protocols thus collapsed to the piecewise constant maximum-efficiency protocol  $\lambda_{\eta}$  (3), illustrating our general theoretical result. Notably, the efficiency achieved by the maximum-efficiency protocol is significantly larger than that provided by usage of the protocol  $\lambda_{\text{S}}$ , which gives maximum efficiency under constrained response.

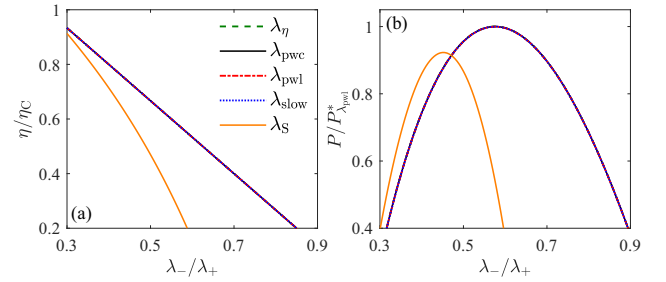


FIG. 2. Numerical optimization of the efficiency of the Brownian heat engine under constrained control verifies that the maximum-efficiency protocol is given by Eq. (3). (a) Maximum efficiency and (b) the corresponding power (in units of the ultimate maximum power  $P_{\lambda_{\text{pwl}}^*}$  for  $\lambda_{\text{pwl}}$ ) as functions of  $\lambda_-/\lambda_+$ . All protocols except for  $\lambda_{\text{S}}$  perfectly overlap. Parameters used are  $t_+ = t_- = 1$ ,  $k_{\text{B}}T_+ = 1$ ,  $k_{\text{B}}T_- = 0.25$  (thus Carnot efficiency  $\eta_{\text{C}} \equiv 1 - T_-/T_+ = 0.75$ ),  $\lambda_+ = 0.5$ , and  $\mu = 1$ .

Main results of the optimization of output power under the constrained control are summarized in Fig. 3. (i) With increasing minimum compression ratio  $\lambda_-/\lambda_+$  allowed by the constraints (2), maximum power for all considered protocols in (a) is first constant and then, at an optimal compression ratio  $r^*$ , decreases. The decreasing part corresponds to protocols which span between the allowed boundary values, i.e.,  $\max \lambda(t) = \lambda(0+) = \lambda_+$  and  $\min \lambda(t) = \lambda(t_+) = \lambda_-$ . At the plateau, the boundary values of the protocols are chosen within the bounds (2) to keep the optimal compression ratio  $r^*$ . (ii) Values of maximum power obtained for the protocols which have enough free parameters are indistinguishable within our numerical precision. As the corresponding optimized protocols seem to have minimum possible curvature  $\ddot{\lambda}(t)$ , we conclude that the maximum-power protocol is  $\lambda_{\text{pwl}}$ . (iii) Only the protocol  $\lambda_{\text{S}}$ , optimized for constrained response  $\sigma$ , yields notably smaller power than other protocols. (iv) In agreement with our above discussion, for large enough values of  $\lambda_-/\lambda_+ \geq 0.59$ , the optimized parameters for protocols  $\lambda_{\text{pwc}}$ ,  $\lambda_{\text{pwl}}$ , and  $\lambda_{\text{slow}}$  are  $b = d = 0$ ,  $a = \lambda_+$ , and  $c = \lambda_-$ , reducing them to  $\lambda_{\eta}$  (3). (v) The maximum powers for the protocols  $\lambda_{\eta}$  and  $\lambda_{\text{pwl}}$  differ just by 1%.

In Fig. 4, we further show that the relative difference in maximum power for  $\lambda_{\text{pwl}}$  and  $\lambda_{\eta}$  is small for a broad range of values of  $T_-/T_+$  and  $t_-/t_+$ . From panels (c)–(f) we conclude that the optimal ratio  $t_-/t_+$  is between 1 and 2, which is in agreement with the results of Appendix B 2 b [see Eq. (B19) below]. Thus, for branch durations that optimize output power, the relative difference  $\delta P$  in (a) is always below 12%, decreasing with the temperature ratio. These results indicate that when one can optimize  $W_{\text{out}}$  and  $P$  over  $\lambda_-$ , the maximum-efficiency protocol (3) often yields almost the maximum power.

The optimization over  $\lambda_-$  is natural for experimental platforms with limitations on the maximum strength of the potential only. The maximum power regime of the maximum-efficiency protocol (3) can be, to a large extent, investigated analytically. First, assuming again that durations of the isotherms are long enough that the system is close to equilibrium at times  $t_+$  and  $t_p$ , we have  $W_{\text{out}} =$



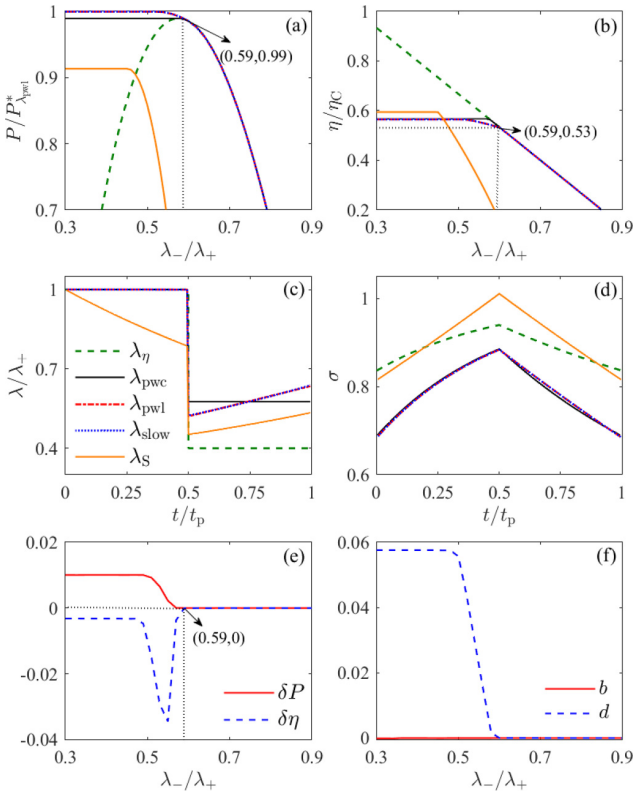


FIG. 3. Numerical optimization of the output power of the Brownian heat engine illustrates that the maximum-efficiency protocol  $\lambda_\eta$  (3) also yields maximum power when the compression ratio  $\lambda_-/\lambda_+$  is large and the durations  $t_+ = t_- = 1$  of the two isotherms are comparable to the relaxation times  $1/(2\mu\lambda_\pm)$  for  $\sigma$ . (a) Powers (in units of the ultimate maximum power  $P_{\lambda_{pwl}}^*$  for  $\lambda_{pwl}$ ) and (b) the corresponding efficiencies obtained using  $\lambda_\eta$  (3) and the protocols in Table I. For  $\lambda_-/\lambda_+ \geq 0.59$  all protocols except for  $\lambda_S$  coincide. (c) and (d) show the protocols and the resulting response for  $\lambda_-/\lambda_+ = 0.4$ . (e) The relative differences  $\delta X = (X_{\lambda_{pwl}} - X_{\lambda_{pwc}})/X_{\lambda_{pwl}}$  of power ( $X = P$ ) and efficiency ( $X = \eta$ ) for  $\lambda_{pwl}$  and  $\lambda_{pwc}$ . (f) The optimal values of parameters  $b$  and  $d$  for  $\lambda_{pwl}$ . We used the same parameters as in Fig. 2.

$(\lambda_+ - \lambda_-)[\sigma^{\text{eq}}(T_+/\lambda_+) - \sigma^{\text{eq}}(T_-/\lambda_-)]$ . For  $f(x) = |x|^n$  in Eq. (1), we then find that the optimal compression ratio is  $\lambda_-/\lambda_+ = \sqrt{T_-/T_+}$ , which leads to the output work  $W_{\text{out}} = k_B T_+(2\eta_{\text{CA}} - \eta_C)/n$  and Curzon-Ahlborn efficiency  $\eta = \eta_{\text{CA}} = 1 - \sqrt{T_-/T_+}$  (see Appendix B 2 a for details). For other than power-law Hamiltonians, the efficiency at maximum power can differ from  $\eta_{\text{CA}}$  but it can still be determined numerically regardless details of dynamical equations for the system (for details, see Appendix B 2 a). Relaxing the assumption of slow (but not quasistatic) isotherms, the optimization of  $W_{\text{out}}$  with respect to  $\lambda_-$  requires specification of the dynamics. In Fig. 5, we show that the efficiency at maximum power of the Brownian heat engine described by Eq. (9) and driven by the maximum-efficiency protocol (3) is bounded between the Curzon-Ahlborn efficiency, achieved for slow isotherms, and the efficiency  $2 - \sqrt{4 - 2\eta_C} < \eta_{\text{CA}}$ , reached in the limit  $t_p \rightarrow 0$ .

In closing this section, we summarize the strong effects of the constraints (constrained control versus constrained

response [24]). First, constraining the control allowed us to derive much more generally valid results than constraining the response. Second, for the constrained response, the power and efficiency can be optimized simultaneously, whereas for the constrained control this is, in general, not possible. Third, the resulting functional forms of the optimal protocols and the corresponding optimal performance strongly differ. Fourth, the change of boundary conditions alters the optimal allocation of cycle duration between hot and cold isotherms,  $t_+/t_-$ , as we show below Eq. (B19) in the Appendix.

## VI. CONCLUSION

We have optimized the thermodynamic performance of finite-time overdamped stochastic heat engines under the constraint that control parameters, such as potential strength or bath temperature, can be varied only over a limited range. This optimization problem is experimentally motivated and differs from previously studied optimization studies performed with constraints on the system's state. We have found that, for working fluids described by the experimentally most common “breathing” Hamiltonians proportional to a control parameter, the maximum efficiency is reached by piecewise constant modulation of the control parameters, independently of the detailed dynamics of the system. When the control parameter can only be changed over a small range and the system is close to equilibrium at the ends of the isotherms, the maximum-efficiency protocol also yields maximum output power. But outside this regime, the maximization of power requires specifying the dynamical equations of the working fluid. For engines based on an overdamped Brownian particle trapped in a harmonic potential, we numerically found that the maximum-power protocol is linear. Nevertheless, the global maxima of the maximum-power and maximum-efficiency protocols are in this setting close, suggesting that the maximum-efficiency protocol provides a reasonable estimate of the output power.

The main strength of the presented derivations of the maximum-efficiency and maximum-power protocols under constrained control is their simplicity and unprecedented generality. Their possible extension to more complicated Hamiltonians is sketched in Appendix A. While more general extensions remain to be explored in future work, the validity of our results for Brownian heat engines is already of experimental relevance. These engines are often realized using optical tweezers with strict bounds on the trap stiffness  $\lambda$ : too small  $\lambda$  leads to losing the Brownian particle while too large  $\lambda$  can induce its overheating. Interestingly, the achievable trap stiffnesses are well above  $10^{-6}$  N/m [31]. For spherical Brownian particles with the radius of  $10^{-6}$  m in water, the Stokes law predicts the mobility of  $\mu \approx 0.5 \times 10^8$  m/Ns, leading to the relaxation time  $1/(2\mu\lambda)$  of the response  $\sigma$  on the order of  $10^{-2}$  s. The assumption that the durations of the isotherms are longer than the response relaxation time, used in our derivation of the maximum-power protocol, is thus, in this setup, natural. Besides, we believe that extensions of our results can find applications in more involved optimization tasks, e.g., performed using machine learning algorithms [66,67] or geometric methods [68,69], as well as in quantum setups [39,70,71].

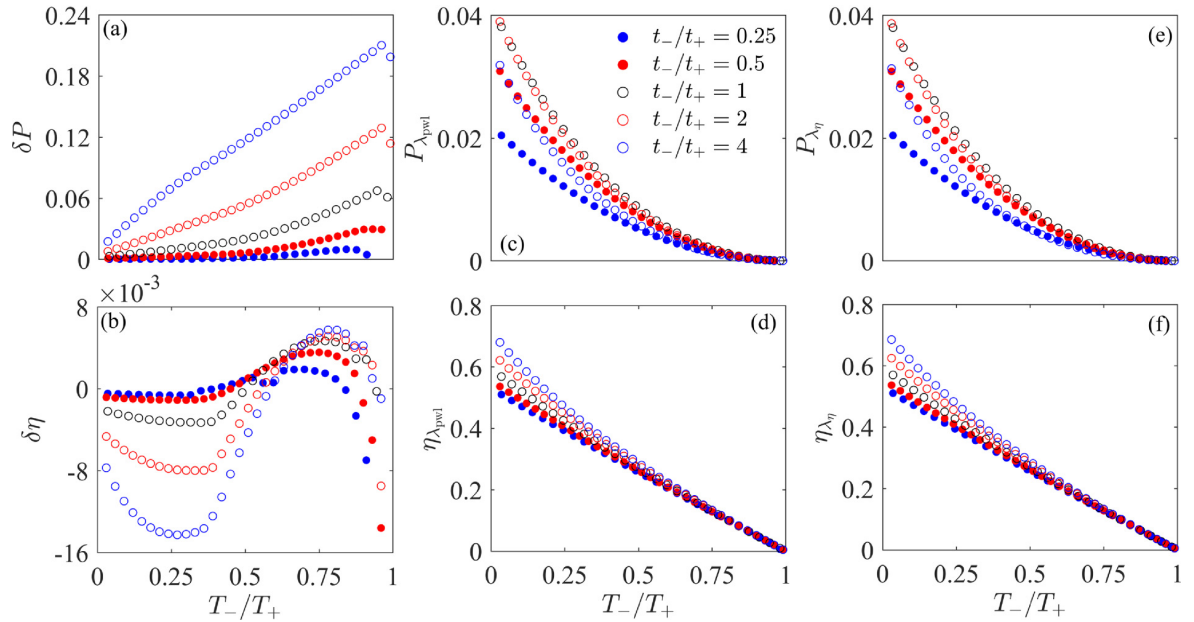


FIG. 4. The relative differences  $\delta X = (X_{\lambda_{\text{pwl}}} - X_{\lambda_{\eta}})/X_{\lambda_{\text{pwl}}}$  of (a) maximum power ( $X = P$ ) optimized with respect to  $\lambda_-$  and (b) the corresponding efficiency ( $X = \eta$ ) for the linear protocol  $\lambda_{\text{pwl}}$  and the maximum-efficiency protocol (3) for different values of  $T_-/T_+$  and  $t_-/t_+$ . (c)–(f) show the corresponding values of maximum power and efficiency. The piecewise constant protocol  $\lambda_{\text{pwc}}$  and the maximum-efficiency protocol  $\lambda_{\eta}$  (3) are in this case equal. We used the same parameters as in Fig. 2.

#### ACKNOWLEDGMENTS

Z.Y. is grateful for the sponsorship of China Scholarship Council (CSC) under Grant No. 201906310136. F.C. gratefully acknowledges funding from the Fundational Questions Institute Fund (FQXi-IAF19-01). P.A. is supported by la Caixa Foundation (ID 100010434, Grant No. LCF/BQ/DI19/11730023), and by the Government of Spain

(FIS2020-TRANQI and Severo Ochoa CEX2019-000910-S), Fundacio Cellex, Fundacio Mir-Puig, Generalitat de Catalunya (CERCA, AGAUR SGR 1381. J.A. acknowledges funding from the Engineering and Physical Sciences Research Council (EPSRC) (EP/R045577/1) and thanks the Royal Society for support. M.P.L. acknowledges financial support from the Swiss National Science Foundations (Ambizione Grant No. PZ00P2-186067). V.H. gratefully acknowledges support by the Humboldt foundation and by the Czech Science Foundation (Project No. 20-02955J).

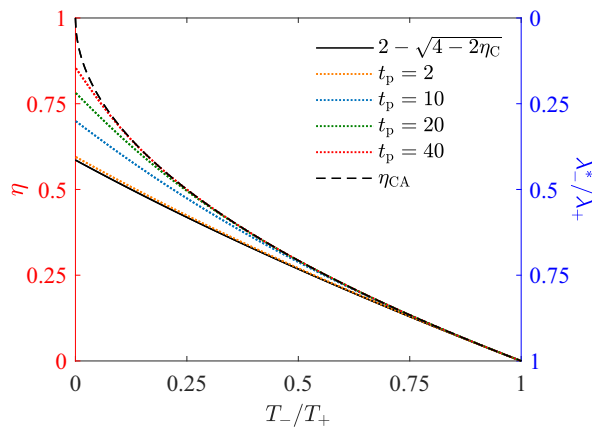


FIG. 5. Efficiency at maximum output power ( $\eta$ , red axis) and the corresponding optimal compression ratio ( $\lambda^*/\lambda_+$ , blue axis) for the maximum-efficiency protocol (3) as functions of the temperature ratio  $T_-/T_+$  for six values of cycle duration  $t_p$  (colored curves) and  $t_- = t_+$ . The cycle time,  $t_p$ , is measured in units of the system relaxation time during the hot isotherm. The corresponding ratio during the cold isotherm reads  $t_p \lambda_-^*/\lambda_+$ . We used the same parameters as in Fig. 2.

#### APPENDIX A: MAXIMUM-EFFICIENCY PROTOCOL FOR MULTITERM HAMILTONIANS

Consider a heat engine with a working fluid described by the Hamiltonian

$$H(x, t) = \sum_i \lambda_i(t) f_i(x) \quad (\text{A1})$$

with control parameters  $\lambda_i(t)$ ,  $i = 1, \dots, N$ . As in the main text, we now aim to derive the finite-time protocol for the constrained control parameters,  $\lambda_i(t) \in (\lambda_i^-, \lambda_i^+)$ , which would yield maximum efficiency of the engine. It will turn out that if the compression ratios  $\lambda_i^-/\lambda_i^+$  for all the control parameters equal, the geometric argument from the main text still applies.

The heat increment is for the Hamiltonian (A1) given by  $dQ = \sum_i \lambda_i(t) d\sigma_i(t)$  with the response functions  $\sigma_i(t) = \langle f_i(x) \rangle$ . For arbitrary fixed maximum changes  $\Delta\sigma_i$  in the response functions during the cycle, geometric upper and lower bounds on  $Q_{\text{in}}$  and  $Q_{\text{out}}$  and thus on efficiency are achieved by clockwise rectangular cycles in the individual  $\lambda_i$ - $\sigma_i$  diagrams. These hypothetical cycles yield the following geometric upper

bound on efficiency:

$$\eta = 1 - \frac{Q_{\text{out}}}{Q_{\text{in}}} \leq 1 - \frac{\sum_i \Delta\sigma_i \lambda_i^-}{\sum_i \Delta\sigma_i \lambda_i^+}. \quad (\text{A2})$$

We use the term ‘‘geometric’’ to stress that this bound follows from the analysis of the cycle in the  $\lambda$ - $\sigma$  diagram, without considering the relation between the protocol  $[\lambda_1(t), \dots, \lambda_N(t)]$  and the response  $[\sigma_1(t), \dots, \sigma_N(t)]$  imposed by dynamical equations of the working fluid. This means that the given set of  $\Delta\sigma_i$  might not be achievable by the piecewise constant protocol and thus the bound in (A2) is loose. Furthermore, we seek an optimal protocol constrained just by the conditions on  $\lambda_i$ , and the upper bound in (A2) in general strongly depends on the fixed values of  $\Delta\sigma_i$ . For the single-term Hamiltonian  $H(x, t) = \lambda(t)f(x)$  used in the main text, this has not been an issue because then  $\Delta\sigma$  in the nominator and denominator in (A2) cancel out and the upper bound becomes independent of the details of the dynamics. The optimal protocol for efficiency is then the piecewise constant protocol for  $\lambda(t)$  because it saturates the geometric upper bound. To sum up, the bound in (A2) allows one to derive the maximum-efficiency protocol only if it happens to be independent of  $\Delta\sigma_i$ . In the opposite case, the optimal protocol cannot be determined without considering the dynamical equations and performing the corresponding functional optimization.

Let us now investigate when the upper bound in (A2) becomes independent of the system response,  $\Delta\sigma_i$ . Defining the set of ‘‘probabilities’’  $p_i = \Delta\sigma_i \lambda_i^+ / \sum_i \Delta\sigma_i \lambda_i^+$ , the ratio in the upper bound in (A2) can be rewritten as the average

$$\frac{\sum_i \Delta\sigma_i \lambda_i^-}{\sum_i \Delta\sigma_i \lambda_i^+} = \sum_i p_i \frac{\lambda_i^-}{\lambda_i^+}. \quad (\text{A3})$$

This expression becomes independent of  $\sigma_i$  only if all the compression ratios  $\lambda_i^-/\lambda_i^+$  are equal. In such a case, the maximum-efficiency protocol is thus a piecewise constant protocol for each of  $\lambda_i$  and yields the efficiency

$$\eta = 1 - \lambda_i^-/\lambda_i^+. \quad (\text{A4})$$

Besides this result, the probabilistic interpretation (A3) of the upper bound in (A2) also yields the dynamics independent (but in general loose) *upper bound* on efficiency,

$$\eta \leq 1 - \min_i \frac{\lambda_i^-}{\lambda_i^+}. \quad (\text{A5})$$

To close this section, we note that a piecewise constant protocol for  $\lambda_i$  will always yield the efficiency  $1 - (\sum_i \Delta\sigma_i \lambda_i^-) / (\sum_i \Delta\sigma_i \lambda_i^+)$ , with values of  $\Delta\sigma_i$  induced by the dynamical equations of the system. Within the class of piecewise constant protocols, the upper bound (A5) is then tight if the constraints on all the control parameters  $\lambda_i$  allow to achieve the minimum compression ratio  $\min_i \frac{\lambda_i^-}{\lambda_i^+}$ . Furthermore, for such protocols, Eq. (A3) also implies the *lower bound* on the efficiency,

$$\eta \geq 1 - \max_i \frac{\lambda_i^-}{\lambda_i^+}, \quad (\text{A6})$$

which is always tight.

## APPENDIX B: PROPERTIES OF MAXIMUM-EFFICIENCY PROTOCOL

In this section, we provide further details concerning the maximum-efficiency protocol for the Hamiltonian,  $H(x, t) = \lambda(t)f(x)$ , discussed in the main text. First, we argue that the maximum-efficiency protocol that yields maximum output work for the given piecewise constant  $\lambda(t)$  requires piecewise constant variation of temperature. Then, we investigate output power of the maximum-efficiency protocol as a function of the lower bound on the control parameter  $\lambda(t)$ .

### 1. Temperature protocol

In the main text, we have shown that the maximum-efficiency protocol for the control parameter  $\lambda(t)$  is piecewise constant and the corresponding efficiency  $\eta = 1 - \lambda_-/\lambda_+$ . The only condition on the temperature protocol was that the cycle is performed clockwise in the  $\lambda$ - $\sigma$  diagram. Nevertheless, in order to allow the engine to operate at Carnot efficiency and to maximize its output work, we have chosen the protocol (3).

For this choice of  $T(t)$ , the working medium of the engine operates with the largest possible temperature gradient during the whole cycle. This maximizes the heat flux through the engine, which can be utilized to yield the maximum amount of work  $W_{\text{out}} = \eta Q_{\text{in}}$ . Besides, the engine efficiency  $\eta$  is also known to increase with the bath temperature difference [see also Figs. 4(c)–4(f)].

Let us now provide an alternative and more technical argument that the choice of  $T(t)$  in Eq. (3) maximizes the output work. We restrict this argument to the maximum-efficiency protocol for  $\lambda$  in Eq. (3). However, generalizations to other protocols are straightforward. The main idea is that connecting the system to the hottest possible bath when  $\dot{\sigma} > 0$  and to the coldest possible bath when  $\dot{\sigma} < 0$  maximizes the extent of the cycle in the  $\sigma$  direction in the  $\sigma$ - $\lambda$  diagram and thus also  $W_{\text{out}}$ .

For the protocol (3), the output work is given by

$$W_{\text{out}} = \Delta\lambda \Delta\sigma, \quad (\text{B1})$$

with  $\Delta\lambda = \lambda_+ - \lambda_-$  and the maximum change in the response parameter during the cycle  $\Delta\sigma = \sigma_+ - \sigma_-$ . To maximize  $W_{\text{out}}$ , we thus need to maximize  $\Delta\sigma$ . To this end, it is reasonable to assume that

$$\Delta\sigma \leq \Delta\sigma^{\text{eq}}, \quad (\text{B2})$$

where  $\Delta\sigma^{\text{eq}} = \max \sigma^{\text{eq}} - \min \sigma^{\text{eq}}$  is the maximum change in the response parameter  $\sigma$  during the cycle with isochoric branches (constant  $\lambda$ ) longer than the system relaxation time. This assumption is in particular valid for arbitrary overdamped dynamics, where  $\sigma$  always converges to its equilibrium value ( $k_B$  denotes the Boltzmann constant)

$$\sigma^{\text{eq}}(t) = \sum_x f(x) \frac{\exp\{-\lambda(t)f(x)/[k_B T(t)]\}}{\sum_x \exp\{-\lambda(t)f(x)/[k_B T(t)]\}}, \quad (\text{B3})$$

corresponding to the instantaneous values of the control parameters  $\{T(t), \lambda(t)\}$ . Noticing that  $\sigma^{\text{eq}}(t) = U(t)/\lambda(t)$ , where  $U(t) = \langle H(x, t) \rangle$  is the thermodynamic internal energy

of the system, the positivity of heat capacity

$$C_v = \frac{\partial U}{\partial T} = \frac{\partial \sigma^{\text{eq}}}{\partial(T/\lambda)} > 0 \quad (\text{B4})$$

implies that  $\sigma^{\text{eq}}$  is a monotonously increasing function of the ratio  $T/\lambda$ .

From Fig. 1 in the main text, it follows that  $\max \sigma^{\text{eq}}$  and  $\min \sigma^{\text{eq}}$  are the values of  $\sigma^{\text{eq}}$  at the ends of the isochores with  $\lambda = \lambda_+$  and  $\lambda = \lambda_-$ , respectively. The upper bound on  $\Delta\sigma$  is thus given by

$$\max \sigma^{\text{eq}} - \min \sigma^{\text{eq}} = \sigma^{\text{eq}}(T_+/\lambda_+) - \sigma^{\text{eq}}(T_-/\lambda_-). \quad (\text{B5})$$

It is attained for slow isochores when  $T = T_+$  for  $\lambda = \lambda_+$  and  $T = T_-$  for  $\lambda = \lambda_-$ . As long as  $\dot{\sigma}^{\text{eq}} > 0$  for  $\lambda = \lambda_+$  and  $\dot{\sigma}^{\text{eq}} < 0$  for  $\lambda = \lambda_-$  (so that the used definitions of input and output heat hold), details of the temperature protocol during the isochores in this limit do not alter the value of  $\Delta\sigma^{\text{eq}}$  and thus  $W_{\text{out}} = \Delta\lambda\Delta\sigma^{\text{eq}}$ . However, these details become important for finite-time cycles.

A typical dynamical equation for an overdamped degree of freedom has the form

$$\dot{\sigma}(t) = t_R^{-1}[\sigma^{\text{eq}} - \sigma(t)]. \quad (\text{B6})$$

For constant values of control parameters  $T(t)$  and  $\lambda(t)$ , which enter the relaxation time  $t_R$  and the equilibrium state  $\sigma^{\text{eq}}(t)$  defined in Eq. (B3), this equation describes an exponential relaxation of  $\sigma$  to  $\sigma^{\text{eq}}$  (for a specific example, see Sec. B 2 b). For a cyclic variation of the control parameters,  $\sigma$  lags behind  $\sigma^{\text{eq}}$  [72]. More precisely,  $\sigma \leq \sigma^{\text{eq}}$  and  $\dot{\sigma} \geq 0$  for  $\lambda = \lambda_+$ , when  $\sigma^{\text{eq}}$  increases to  $\max \sigma^{\text{eq}}$ , and  $\sigma \geq \sigma^{\text{eq}}$  and  $\dot{\sigma} \leq 0$  for  $\lambda = \lambda_-$ , when  $\sigma^{\text{eq}}$  decreases to  $\min \sigma^{\text{eq}}$ . The change in the response  $\Delta\sigma = \int_0^{t_+} \dot{\sigma} dt = -\int_{t_+}^{t_-} \dot{\sigma} dt$  and thus it can be maximized by maximizing (minimizing) the instantaneous rate of change of the response,  $\dot{\sigma}$ , during the first (second) isochore. From Eq. (B6), it follows that this is achieved by setting  $\sigma^{\text{eq}} = \max \sigma^{\text{eq}}$  during the first isochore and  $\sigma^{\text{eq}} = \min \sigma^{\text{eq}}$  during the second one. Altogether, this suggests that the piecewise constant temperature protocol in Eq. (3) yields maximum  $\Delta\sigma$  and thus output work  $W_{\text{out}}$  (B1) for arbitrary cycle duration.

## 2. Efficiency at maximum power

Let us now turn to the task of maximizing the output work  $W_{\text{out}} = (\lambda_+ - \lambda_-)\Delta\sigma$  with respect to  $\lambda_-$ . Analytical results can be obtained in the limits of slow and fast isotherms.

### a. Slow isotherms

When the duration of the isotherms is longer than the relaxation time of the response  $\sigma$ , one can approximate  $\sigma_+$  and  $\sigma_-$  in  $\Delta\sigma$  by their equilibrium values. Using Eq. (B1), the output work then reads

$$W_{\text{out}} = \Delta\lambda\Delta\sigma^{\text{eq}}. \quad (\text{B7})$$

Equation (B4) implies that the partial derivative of  $\sigma^{\text{eq}}$  with respect to the control parameter  $\lambda$  ( $T$  is constant) is given by

$$\frac{\partial}{\partial\lambda}\sigma^{\text{eq}}(T/\lambda) = -\frac{T}{\lambda^2}C_v. \quad (\text{B8})$$

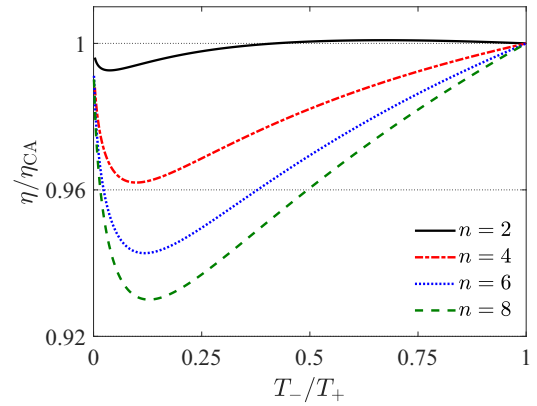


FIG. 6. Efficiency at maximum output work obtained using the Hamiltonian  $H = \lambda(t)(|x|^n/n - \ln|x|)$  as a function of  $T_-/T_+$ . Parameters used are  $k_B T_+ = 1$  and  $\lambda_+ = 0.5$ .

The condition on the extreme of  $W_{\text{out}}$  (B7) with respect to  $\lambda_-$  thus reads

$$\begin{aligned} \frac{\partial W_{\text{out}}}{\partial\lambda_-} &= (\lambda_+ - \lambda_-)\frac{T_-}{\lambda_-^2}C_v(T_-/\lambda_-) - \frac{U(T_+/\lambda_+)}{\lambda_+} \\ &\quad + \frac{U(T_-/\lambda_-)}{\lambda_-} = 0, \end{aligned} \quad (\text{B9})$$

where we additionally used the relation  $\sigma^{\text{eq}} = U/\lambda$  between  $\sigma^{\text{eq}}$  and the internal energy  $U$ .

For power law Hamiltonians of the form  $H = \lambda|x|^n/n$  where  $C_v = k_B/n$  and  $U = k_B T/n$ , this equation can be solved explicitly. The resulting optimal compression ratio is given by  $\lambda_-/\lambda_+ = \sqrt{T_-/T_+}$ . The corresponding efficiency at the maximum output work is given by the Curzohn-Ahlborn efficiency,

$$\eta = 1 - \frac{\lambda_-}{\lambda_+} = 1 - \sqrt{\frac{T_-}{T_+}} \equiv \eta_{\text{CA}}, \quad (\text{B10})$$

and the maximum output work is (Carnot efficiency  $\eta_C = 1 - T_-/T_+$ )

$$W_{\text{out}} = \frac{k_B T_+}{n}(2\eta_{\text{CA}} - \eta_C). \quad (\text{B11})$$

Let us now consider the asymmetric Hamiltonian  $H = \lambda(t)(|x|^n/n - \ln|x|)$ . In this case, the internal energy and heat capacity are given by

$$U = \frac{k_B T + \lambda \left[ 1 + \ln \frac{\lambda}{nk_B T} - \psi^{(0)}\left(\frac{\lambda + k_B T}{nk_B T}\right) \right]}{n}, \quad (\text{B12})$$

$$C_v = \frac{nk_B T(k_B T - \lambda) + \lambda^2 \psi^{(1)}\left(\frac{\lambda + k_B T}{nk_B T}\right)}{n^2 k_B^2 T^2}, \quad (\text{B13})$$

where  $\psi^{(m)}(z)$  denotes the polygamma function of order  $m$ . In this case, Eq. (B9) is transcendental and we solved it numerically. In Fig. 6, we show the resulting efficiency at the maximum output work as a function of  $T_-/T_+$ . Even though the resulting efficiency is still close to  $\eta_{\text{CA}}$ , it can be both slightly larger and smaller than that.

### b. Fast isotherms

Let us now assume that the duration of the isothermal branches are much shorter than the system relaxation time. In such a case, the work optimization cannot be done without specifying the dynamical equation for the response  $\sigma$ . To this end, we assume that it obeys the overdamped equation (B6) with the equilibrium value  $\sigma^{\text{eq}}$  and relaxation time  $t_R$  determined by the values of the control parameters  $\{T(t), \lambda(t)\}$  at time  $t$ . The most prominent examples of systems described by this formula are a two-level system [35] and an overdamped particle trapped in a harmonic potential [24].

Solving Eq. (B6) for the maximum-efficiency protocol (3), we find that

$$\sigma(t) = \begin{cases} \sigma_0 e^{-\frac{t}{t_R}} + \sigma_{\text{eq}}^+ \left(1 - e^{-\frac{t}{t_R}}\right), & 0 < t < t_+, \\ \sigma_1 e^{-\frac{t-t_+}{t_R}} + \sigma_{\text{eq}}^- \left(1 - e^{-\frac{t-t_+}{t_R}}\right), & t_+ < t < t_p, \end{cases} \quad (\text{B14})$$

where  $\sigma_0 \equiv \sigma(0)$  and  $\sigma_1 \equiv \sigma(t_+)$  are determined by the condition that  $\sigma(t)$  must be a continuous function of time. The variables corresponding to the first (second) isotherm are denoted by max (min). It turns out that

$$\begin{aligned} \Delta\sigma &= \sigma_+ - \sigma_- = \sigma_1 - \sigma_0 \\ &= \Delta\sigma^{\text{eq}} \frac{\sinh(t_+/t_R^+) \sinh(t_-/t_R^-)}{\sinh(t_+/t_R^+ + t_-/t_R^-)} \leq \Delta\sigma^{\text{eq}}. \end{aligned} \quad (\text{B15})$$

Substituting this result into the expression for the output work (B1) and expanding the result up to the leading order in the ratios of duration of the individual isotherms to the corresponding relaxation times,  $t_+/t_R^+$  and  $t_-/t_R^-$ , we find that

$$W_{\text{out}} = \Delta\lambda \Delta\sigma^{\text{eq}} \frac{\frac{t_+ t_-}{t_R^+ t_R^-}}{\frac{t_+}{t_R^+} + \frac{t_-}{t_R^-}}. \quad (\text{B16})$$

To maximize the output work, we need to choose a specific model to determine the dependence of the equilibrium values of response and relaxation times on the control parameters. To this end, we consider the paradigmatic model of stochastic thermodynamics,  $\dot{\sigma}(t) = -2\mu\lambda(t)\sigma(t) + \mu k_B T$ , describing an overdamped Brownian particle with mobility  $\mu$  in a harmonic trap [24,62]. In this case,  $\sigma^{\text{eq}} = T/(2\mu\lambda)$  and  $t_R = 1/(2\mu\lambda)$ , and the maximum output work (B16) is produced for

$$\frac{\lambda_-}{\lambda_+} = \sqrt{(\alpha+1)(\alpha+1-\eta_C)} - \alpha, \quad (\text{B17})$$

where  $\alpha \equiv t_+/t_-$ . The corresponding efficiency reads

$$\eta = 1 - \frac{\lambda_-}{\lambda_+} = \alpha + 1 - \sqrt{(\alpha+1)(\alpha+1-\eta_C)}, \quad (\text{B18})$$

which reduces to  $\eta_{\text{CA}}$  for  $\alpha \rightarrow 0$  and  $\eta_C/2$  for  $\alpha \rightarrow \infty$ . Assuming that  $\alpha = 1$  ( $t_+ = t_-$ ), Eq. (B18) is given by the formula

$$\eta = 2 - \sqrt{4 - 2\eta_C} = \frac{\eta_C}{2} + \frac{\eta_C^2}{16} + \mathcal{O}(\eta_C^3) \quad (\text{B19})$$

used in the main text. The corresponding expansion for the Curzohn-Ahlborn efficiency,  $\eta_{\text{CA}} \approx \frac{\eta_C}{2} + \frac{2\eta_C^2}{16}$ , has an identical linear term and a twice larger quadratic term.

Last but not least, with respect to  $\alpha$ , the output power  $W_{\text{out}}/t_p$  using Eq. (B16) develops a peak at  $\alpha = \alpha^* = \sqrt{\frac{\lambda_-}{\lambda_+}} < 1$ . This also contradicts the situation with constrained  $\sigma$ , where maximum power is attained when the durations of the isotherms are equal ( $\alpha = \alpha^* = 1$ ) [24].

## APPENDIX C: OPTIMAL DRIVING FOR SYSTEMS CLOSE TO EQUILIBRIUM

In this Appendix, we consider optimization of a slowly driven heat engine based on an overdamped Brownian particle trapped in the power-law potential  $H = \lambda(t)x^n/n$  with  $n = 2, 4, \dots$ . We use the temperature protocol from Eq. (3) and impose fixed values of the response  $\sigma$  (or, in the slow driving limit equivalently also the control  $\lambda$ ) at the ends of the two isotherms. Dynamics of the particle position is described by the Langevin equation

$$\dot{x} = -\mu\lambda(t)x^{n-1} + \sqrt{2D(t)}\xi(t), \quad (\text{C1})$$

where  $D(t) = \mu k_B T(t)$  denotes the diffusion coefficient. From Eq. (C1) and its formal solution

$$x(t) = -\mu \int dt \lambda(t)x^{n-1}(t) + \sqrt{2D(t)} \int dt \xi(t), \quad (\text{C2})$$

we find that  $\langle x(t)\xi(t) \rangle = \sqrt{D/2}$  and thus

$$\frac{d}{dt} \langle x^2(t) \rangle = -2\mu\lambda(t)\langle x^n(t) \rangle + 2D. \quad (\text{C3})$$

Let us now assume that the control parameters  $\{T(t), \lambda(t)\}$  vary slowly with respect to the relaxation time of the system, such that, during the limit cycle, the system is always close to equilibrium, and solve this equation up to the first order in  $\dot{\lambda}(t)$ . To this end, we consider the ansatz  $\langle x^2(t) \rangle = \langle x^2(t) \rangle_0$  and  $\langle x^n(t) \rangle = \langle x^n(t) \rangle_0 + \langle x^n(t) \rangle_\lambda$ , where

$$\langle x^m(t) \rangle_0 = \int_{-\infty}^{\infty} dx x^m \frac{\exp(-\mu \frac{\lambda x^n}{nD})}{Z}, \quad (\text{C4})$$

with the partition function  $Z = 2[nD/\mu\lambda(t)]^{1/n}\Gamma(1+1/n)$ , is the value of the moment  $\langle x^m(t) \rangle$  corresponding to the infinitely slow driving, and  $\langle x^n(t) \rangle_\lambda$  is the correction proportional to  $\dot{\lambda}$ . We find that

$$\langle x^n(t) \rangle_0 = \frac{D(t)}{\mu\lambda(t)}, \quad (\text{C5})$$

$$\langle x^2(t) \rangle_0 = \left[ \frac{nD(t)}{\mu\lambda(t)} \right]^{2/n} \frac{\Gamma(3/n)}{\Gamma(1/n)}, \quad (\text{C6})$$

and

$$\langle x^n(t) \rangle_\lambda = -\frac{1}{2\mu\lambda(t)} \frac{d}{dt} \langle x^2(t) \rangle_0. \quad (\text{C7})$$

We reiterate that this solution is valid only for protocols  $\{T(t), \lambda(t)\}$  which are changing slowly with respect to the relaxation time of the system so that the system is, during the whole cycle, close to equilibrium. However, as we know from the previous discussion, both the piecewise constant

maximum-efficiency protocol for constrained control and the optimal protocol (D1) for the constrained response contain discontinuities, where  $\{T(t), \lambda(t)\}$  changes abruptly. To be able to use the slow driving approximation for the derivation of optimal cyclic protocols, we thus need to additionally assume that during these jumps the system is not driven far from equilibrium. To this end, we assume that the ratio  $\lambda(t)/T(t)$  in the Boltzmann factor is during the jumps at the ends of the isotherms constant. This additional assumption fixes the state of the system  $\sigma$  at the ends of the isotherms and thus the present optimization scheme is only suitable for the optimization under the constrained response. Let us now proceed with the optimization.

Work done on the system during the time interval  $t_i \leq t \leq t_f$  for the given Hamiltonian reads

$$W = \frac{1}{n} \int_{t_i}^{t_f} dt \dot{\lambda}(t) \langle x^n(t) \rangle \equiv W(t_i, t_f). \quad (\text{C8})$$

Having fixed the state of the system at the ends of the isotherms, it is enough to maximize the work during these branches. For an isothermal process, the work Eq. (C8) can be written as  $W = \Delta F + W_{\text{irr}}$ , where the first term, denoting the nonequilibrium free energy difference [24], comes from  $\langle x^n(t) \rangle_0$ , and the second term reads

$$W_{\text{irr}} = \frac{1}{n} \int_{t_i}^{t_f} dt \dot{\lambda}(t) \langle x^n(t) \rangle_{\dot{\lambda}} = \frac{1}{n^2 \mu} \left( \frac{nD}{\mu} \right)^{2/n} \times \frac{\Gamma(3/n)}{\Gamma(1/n)} \int_{t_i}^{t_f} dt \dot{\lambda}(t)^2 \lambda(t)^{-2(1+n)/n}. \quad (\text{C9})$$

As  $\Delta F$  is fixed by the imposed boundary conditions on the state of the system  $\sigma$ , to maximize the output work  $-W$  means to minimize the irreversible work  $W_{\text{irr}}$  as a functional of  $\lambda(t)$ . This leads to the Euler-Lagrange equation

$$\ddot{\lambda}(t) \lambda(t) - \frac{1+n}{n} \dot{\lambda}(t)^2 = 0, \quad (\text{C10})$$

which has the general solution

$$\lambda_{\text{slow}}(t) = \frac{a}{(1+bt)^n}. \quad (\text{C11})$$

We thus come to an interesting conclusion that the optimal slow protocol for the constrained response scales with the same exponent as the potential. The values of  $a$  and  $b$  can be expressed in terms of the boundary conditions for  $\lambda_{\text{slow}}(t)$ , i.e.,  $\lambda_{\text{slow}}(t_i) \equiv \lambda_i$  and  $\lambda_{\text{slow}}(t_f) \equiv \lambda_f$ . The optimal slow protocol (C11) then reads

$$\lambda_{\text{slow}}(t) = \frac{\lambda(t_i)}{\left[ 1 + \left( \sqrt{\frac{\lambda(t_i)}{\lambda(t_f)}} - 1 \right) \frac{t-t_i}{t_f-t_i} \right]^n}. \quad (\text{C12})$$

And the corresponding irreversible work and input work are given by

$$W_{\text{irr}} = \frac{\Gamma(3/n)}{\Gamma(1/n)} \left[ \frac{nD}{\mu \lambda_i} \right]^{2/n} \left( \sqrt{\frac{\lambda_i}{\lambda_f}} - 1 \right)^2, \quad (\text{C13})$$

$$W = \frac{\Gamma(3/n)}{\Gamma(1/n)} \left[ \frac{nD}{\mu \lambda_i} \right]^{2/n} \left( \sqrt{\frac{\lambda_i}{\lambda_f}} - 1 \right)^2 - \frac{D}{n\mu} \ln \frac{\lambda_i}{\lambda_f}. \quad (\text{C14})$$

These results are valid for the individual isothermal branches of the cycle. Importantly, the obtained optimized values of

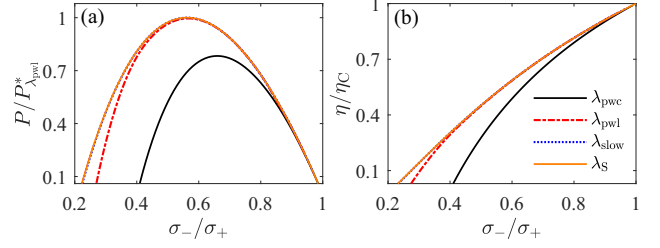


FIG. 7. Optimal performance for fixed boundary values of the response:  $\sigma_- \equiv \sigma(0) = \sigma(t_p)$  and  $\sigma_+ \equiv \sigma(t_+) = 0.5$ . (a) Maximum power (in units of the ultimate maximum power  $P_{\text{pwl}}^*$ ) for  $\lambda_{\text{pwl}}$  and (b) maximum efficiency as functions of  $\sigma_-/\sigma_+$ . Lines corresponding to  $\lambda_S$  (orange solid) and  $\lambda_{\text{slow}}$  (blue dotted) perfectly overlap. The maximum-efficiency protocol (3) and the piecewise constant protocol  $\lambda_{\text{pwc}}$  are in this case equal. We used the same parameters as in Fig. 2.

the irreversible work are correct up to the order  $1/(t_f - t_i)$ , which is their exact dependence on the process duration [24]. These results are thus exact even though they were obtained from the approximate optimal protocol. According to Refs. [24,73], these irreversible works determine the optimal performance of the engine under the constraints on  $\sigma$ , i.e., they give the maximum output work  $W_{\text{out}} = -W(0, t_+) - W(t_+, t_p)$  and efficiency  $\eta = W_{\text{out}}/[T_h \Delta S - W_{\text{irr}}(0, t_+)]$  ( $\Delta S$  is the increase in entropy of the system during the hot isotherm). Also this performance is thus from the approximate analysis based on the slow driving obtained exactly.

#### APPENDIX D: CONSTRAINED RESPONSE

To test our numerical procedure, in this Appendix we check numerically that the protocol  $\lambda_S$  obtained from Ref. [24] is indeed optimal for both power and efficiency under the constraints on  $\sigma$ . When the values of the response (position variance)  $\sigma$  at the ends of the two isotherms are fixed, i.e.,  $\sigma_- \equiv \sigma(0) = \sigma(t_p)$  and  $\sigma_+ \equiv \sigma(t_+)$ , the protocol which yields both maximum efficiency and power reads [24]

$$\lambda_S = \begin{cases} \frac{T_+}{2\sigma_S} - \frac{\sqrt{\sigma_+} - \sqrt{\sigma_-}}{\mu_+ \sqrt{\sigma_S}}, & 0 < t < t_+, \\ \frac{T_-}{2\sigma_S} + \frac{\sqrt{\sigma_+} - \sqrt{\sigma_-}}{\mu_- \sqrt{\sigma_S}}, & t_+ < t < t_p, \end{cases} \quad (\text{D1})$$

with

$$\sigma_S = \begin{cases} \frac{\sigma_-}{2} \left[ 1 + \left( \sqrt{\frac{\sigma_+}{\sigma_-}} - 1 \right) \frac{t}{t_+} \right]^2, & 0 < t < t_+, \\ \frac{\sigma_+}{2} \left[ 1 + \left( \sqrt{\frac{\sigma_-}{\sigma_+}} - 1 \right) \frac{t-t_+}{t_-} \right]^2, & t_+ < t < t_p. \end{cases} \quad (\text{D2})$$

However, this protocol is no longer optimal when one imposes just maximum and minimum values on the response, i.e.,  $\sigma(t) \in [\sigma_-, \sigma_+]$ . Then, our analysis shows that the maximum-efficiency and maximum-power protocol is still of the above form, but with  $\sigma_- < \sigma(0) = \sigma(t_p) < \sigma(t_+) < \sigma_+$ .

In Fig. 7, we show the maximum power (a) and maximum efficiency (b) for the trial protocols under the constraint  $\sigma_- \equiv \sigma(0) = \sigma(t_p)$  and  $\sigma_+ \equiv \sigma(t_+)$ . As expected, power and

efficiency corresponding to the protocol  $\lambda_S$  are largest from all the protocols. In particular, the figure demonstrates that the linear protocol, which was found to maximize output power for constrained  $\lambda$ , yields smaller output power than  $\lambda_S$ . And the piecewise constant protocol yields smaller efficiency than  $\lambda_S$ . Nevertheless, it is interesting to note that the performance

of the protocol  $\lambda_{\text{slow}}(t)$ , which optimizes both output power and efficiency for slow driving (see Sec. C for details), is for the chosen parameters indistinguishable from that of  $\lambda_S$ . This means that the chosen cycle is slow enough. Finally, for small enough cycles (small  $\sigma_-/\sigma_+$ ) performances of all the protocols are equal.

- 
- [1] G. Pesce, P. H. Jones, O. M. Maragò, and G. Volpe, Optical tweezers: theory and practice, *Eur. Phys. J. Plus* **135**, 949 (2020).
- [2] D. J. Wineland, Nobel lecture: Superposition, entanglement, and raising Schrödinger's cat, *Rev. Mod. Phys.* **85**, 1103 (2013).
- [3] S. Haroche, Nobel lecture: Controlling photons in a box and exploring the quantum to classical boundary, *Rev. Mod. Phys.* **85**, 1083 (2013).
- [4] N. M. Myers, O. Abah, and S. Deffner, Quantum thermodynamic devices: from theoretical proposals to experimental reality, *AVS Quantum Sci.* **4**, 027101 (2022).
- [5] K. Sekimoto, *Stochastic Energetics*, Lecture Notes in Physics Vol. 799 (Springer, Berlin, 2010).
- [6] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, *Rep. Prog. Phys.* **75**, 126001 (2012).
- [7] F. L. Curzon and B. Ahlborn, Efficiency of a Carnot engine at maximum power output, *Am. J. Phys.* **43**, 22 (1975).
- [8] M. H. Rubin, Optimal configuration of a class of irreversible heat engines. I, *Phys. Rev. A* **19**, 1272 (1979).
- [9] P. Salamon, A. Nitzan, B. Andresen, and R. S. Berry, Minimum entropy production and the optimization of heat engines, *Phys. Rev. A* **21**, 2115 (1980).
- [10] B. Andresen, R. S. Berry, M. J. Ondrechen, and P. Salamon, Thermodynamics for processes in finite time, *Acc. Chem. Res.* **17**, 266 (1984).
- [11] B. Andresen, P. Salamon, and R. S. Berry, Thermodynamics in finite time, *Phys. Today* **37**(9), 62 (1984).
- [12] M. Mozurkewich and R. S. Berry, Finite-time thermodynamics: Engine performance improved by optimized piston motion, *Proc. Natl. Acad. Sci. USA* **78**, 1986 (1981).
- [13] C. Van den Broeck, Thermodynamic Efficiency at Maximum Power, *Phys. Rev. Lett.* **95**, 190602 (2005).
- [14] B. Jiménez de Cisneros and A. C. Hernández, Collective Working Regimes for Coupled Heat Engines, *Phys. Rev. Lett.* **98**, 130602 (2007).
- [15] K. H. Hoffmann, An introduction to endoreversible thermodynamics, *AAPP Phys. Math. Nat. Sci.* **86**, 1 (2008).
- [16] V. Holubec and A. Ryabov, Fluctuations in heat engines, *J. Phys. A: Math. Theor.* **55**, 013001 (2022).
- [17] E. Geva and R. Kosloff, A quantum-mechanical heat engine operating in finite time. a model consisting of spin-1/2 systems as the working fluid, *J. Chem. Phys.* **96**, 3054 (1992).
- [18] A. Parmeggiani, F. Jülicher, A. Ajdari, and J. Prost, Energy transduction of isothermal ratchets: Generic aspects and specific examples close to and far from equilibrium, *Phys. Rev. E* **60**, 2127 (1999).
- [19] T. Hondou and K. Sekimoto, Unattainability of Carnot efficiency in the Brownian heat engine, *Phys. Rev. E* **62**, 6021 (2000).
- [20] T. Feldmann and R. Kosloff, Performance of discrete heat engines and heat pumps in finite time, *Phys. Rev. E* **61**, 4774 (2000).
- [21] R. D. Astumian and P. Hänggi, Brownian motors, *Phys. Today* **55**(11), 33 (2002).
- [22] J. M. R. Parrondo and B. J. de Cisneros, Energetics of brownian motors: a review, *Appl. Phys. A* **75**, 179 (2002).
- [23] P. Reimann, Brownian motors: noisy transport far from equilibrium, *Phys. Rep.* **361**, 57 (2002).
- [24] T. Schmiedl and U. Seifert, Efficiency at maximum power: An analytically solvable model for stochastic heat engines, *Europhys. Lett.* **81**, 20003 (2008).
- [25] Z. C. Tu, Efficiency at maximum power of Feynman's ratchet as a heat engine, *J. Phys. A: Math. Theor.* **41**, 312003 (2008).
- [26] M. Esposito, K. Lindenberg, and C. Van den Broeck, Universality of Efficiency at Maximum Power, *Phys. Rev. Lett.* **102**, 130602 (2009).
- [27] M. Esposito, K. Lindenberg, and C. V. den Broeck, Thermoelectric efficiency at maximum power in a quantum dot, *Europhys. Lett.* **85**, 60010 (2009).
- [28] M. Esposito, R. Kawai, K. Lindenberg, and C. Van den Broeck, Efficiency at Maximum Power of Low-Dissipation Carnot Engines, *Phys. Rev. Lett.* **105**, 150603 (2010).
- [29] O. Abah, J. Roßnagel, G. Jacob, S. Deffner, F. Schmidt-Kaler, K. Singer, and E. Lutz, Single-Ion Heat Engine at Maximum Power, *Phys. Rev. Lett.* **109**, 203006 (2012).
- [30] O. Raz, Y. Subaşı, and R. Pugatch, Geometric Heat Engines Featuring Power that Grows with Efficiency, *Phys. Rev. Lett.* **116**, 160601 (2016).
- [31] V. Holubec and A. Ryabov, Cycling Tames Power Fluctuations near Optimum Efficiency, *Phys. Rev. Lett.* **121**, 120601 (2018).
- [32] K. Brandner and K. Saito, Thermodynamic Geometry of Microscopic Heat Engines, *Phys. Rev. Lett.* **124**, 040602 (2020).
- [33] P. Terrén Alonso, P. Abiuso, M. Perarnau-Llobet, and L. Arrachea, Geometric optimization of nonequilibrium adiabatic thermal machines and implementation in a qubit system, *PRX Quantum* **3**, 010326 (2022).
- [34] V. Cavina, P. A. Erdman, P. Abiuso, L. Tolomeo, and V. Giovannetti, Maximum-power heat engines and refrigerators in the fast-driving regime, *Phys. Rev. A* **104**, 032226 (2021).
- [35] P. A. Erdman, V. Cavina, R. Fazio, F. Taddei, and V. Giovannetti, Maximum power and corresponding efficiency for two-level heat engines and refrigerators: optimality of fast cycles, *New J. Phys.* **21**, 103049 (2019).
- [36] A. Das and V. Mukherjee, Quantum-enhanced finite-time Otto cycle, *Phys. Rev. Res.* **2**, 033083 (2020).
- [37] P. Abiuso and M. Perarnau-Llobet, Optimal Cycles for Low-Dissipation Heat Engines, *Phys. Rev. Lett.* **124**, 110606 (2020).

- [38] T. E. Humphrey, R. Newbury, R. P. Taylor, and H. Linke, Reversible Quantum Brownian Heat Engines for Electrons, *Phys. Rev. Lett.* **89**, 116801 (2002).
- [39] P. Abiuso, H. J. Miller, M. Perarnau-Llobet, and M. Scandi, Geometric optimisation of quantum thermodynamic processes, *Entropy* **22**, 1076 (2020).
- [40] H. J. D. Miller, M. H. Mohammady, M. Perarnau-Llobet, and G. Guarneri, Thermodynamic Uncertainty Relation in Slowly Driven Quantum Heat Engines, *Phys. Rev. Lett.* **126**, 210603 (2021).
- [41] V. Cavina, A. Mari, and V. Giovannetti, Slow Dynamics and Thermodynamics of Open Quantum Systems, *Phys. Rev. Lett.* **119**, 050601 (2017).
- [42] V. Holubec, An exactly solvable model of a stochastic heat engine: optimization of power, power fluctuations and efficiency, *J. Stat. Mech. Theory Exp.* **2014**, P05022.
- [43] A. Dechant, N. Kiesel, and E. Lutz, Underdamped stochastic heat engine at maximum efficiency, *Europhys. Lett.* **119**, 50003 (2017).
- [44] Y. Zhang, Optimization of stochastic thermodynamic machines, *J. Stat. Phys.* **178**, 1336 (2020).
- [45] P. Abiuso, V. Holubec, J. Anders, Z. Ye, F. Cerisola, and M. Perarnau-Llobet, Thermodynamics and optimal protocols of multidimensional quadratic Brownian systems, *J. Phys. Commun.* **6**, 063001 (2022).
- [46] A. Zhong and M. R. DeWeese, Limited-control optimal protocols arbitrarily far from equilibrium, *Phys. Rev. E* **106**, 044135 (2022).
- [47] M. Bauer, K. Brandner, and U. Seifert, Optimal performance of periodically driven, stochastic heat engines under limited control, *Phys. Rev. E* **93**, 042112 (2016).
- [48] C. A. Plata, D. Guéry-Odelin, E. Trizac, and A. Prados, Optimal work in a harmonic trap with bounded stiffness, *Phys. Rev. E* **99**, 012140 (2019).
- [49] H. Risken, Fokker-planck equation, in *The Fokker-Planck Equation: Methods of Solution and Applications* (Springer, 1996), pp. 63–95.
- [50] V. Holubec, A. Ryabov, S. A. M. Loos, and K. Kroy, Equilibrium stochastic delay processes, *New J. Phys.* **24**, 023021 (2022).
- [51] S. Krishnamurthy, S. Ghosh, D. Chatterji, R. Ganapathy, and A. K. Sood, A micrometre-sized heat engine operating between bacterial reservoirs, *Nat. Phys.* **12**, 1134 (2016).
- [52] V. Holubec, S. Steffenoni, G. Falasco, and K. Kroy, Active Brownian heat engines, *Phys. Rev. Res.* **2**, 043262 (2020).
- [53] A. Kumari, P. S. Pal, A. Saha, and S. Lahiri, Stochastic heat engine using an active particle, *Phys. Rev. E* **101**, 032109 (2020).
- [54] G. Gronchi and A. Puglisi, Optimization of an active heat engine, *Phys. Rev. E* **103**, 052134 (2021).
- [55] I. A. Martínez, E. Roldán, L. Dinis, D. Petrov, and R. A. Rica, Adiabatic Processes Realized with a Trapped Brownian Particle, *Phys. Rev. Lett.* **114**, 120601 (2015).
- [56] D. Arold, A. Dechant, and E. Lutz, Heat leakage in overdamped harmonic systems, *Phys. Rev. E* **97**, 022131 (2018).
- [57] V. Blickle and C. Bechinger, Realization of a micrometre-sized stochastic heat engine, *Nat. Phys.* **8**, 143 (2012).
- [58] V. Holubec and A. Ryabov, Diverging, but negligible power at carnot efficiency: Theory and experiment, *Phys. Rev. E* **96**, 062107 (2017).
- [59] V. Cavina, A. Mari, A. Carlini, and V. Giovannetti, Optimal thermodynamic control in open quantum systems, *Phys. Rev. A* **98**, 012139 (2018).
- [60] I. A. Martínez, É. Roldán, L. Dinis, D. Petrov, J. M. R. Parrondo, and R. A. Rica, Brownian carnot engine, *Nat. Phys.* **12**, 67 (2016).
- [61] I. A. Martínez, É. Roldán, L. Dinis, and R. A. Rica, Colloidal heat engines: a review, *Soft Matter* **13**, 22 (2017).
- [62] T. Schmiedl and U. Seifert, Optimal Finite-Time Processes In Stochastic Thermodynamics, *Phys. Rev. Lett.* **98**, 108301 (2007).
- [63] H. Then and A. Engel, Computing the optimal protocol for finite-time processes in stochastic thermodynamics, *Phys. Rev. E* **77**, 041105 (2008).
- [64] V. Holubec and A. Ryabov, Efficiency at and near maximum power of low-dissipation heat engines, *Phys. Rev. E* **92**, 052125 (2015).
- [65] T. R. Gingrich, G. M. Rotskoff, G. E. Crooks, and P. L. Geissler, Near-optimal protocols in complex nonequilibrium transformations, *Proc. Natl. Acad. Sci. USA* **113**, 10263 (2016).
- [66] P. A. Erdman and F. Noé, Identifying optimal cycles in quantum thermal machines with reinforcement-learning, *NPJ Quantum Inf.* **8**, 1 (2022).
- [67] I. Khait, J. Carrasquilla, and D. Segal, Optimal control of quantum thermal machines using machine learning, *Phys. Rev. Res.* **4**, L012029 (2022).
- [68] A. Dechant and Y. Sakurai, Thermodynamic interpretation of Wasserstein distance, [arXiv:1912.08405](https://arxiv.org/abs/1912.08405).
- [69] T. Van Vu and Y. Hasegawa, Geometrical Bounds of the Irreversibility in Markovian Systems, *Phys. Rev. Lett.* **126**, 010601 (2021).
- [70] M. Carrega, L. M. Cangemi, G. De Filippis, V. Cataudella, G. Benenti, and M. Sassetti, Engineering dynamical couplings for quantum thermodynamic tasks, *PRX Quantum* **3**, 010323 (2022).
- [71] N. Pancotti, M. Scandi, M. T. Mitchison, and M. Perarnau-Llobet, Speed-Ups to Isothermality: Enhanced Quantum Thermal Machines through Control of the System-Bath Coupling, *Phys. Rev. X* **10**, 031015 (2020).
- [72] V. Holubec, *Non-equilibrium Energy Transformation Processes: Theoretical Description at the Level of Molecular Structures*, Springer Theses (Springer International, Cham, 2014).
- [73] V. Holubec and A. Ryabov, Maximum efficiency of low-dissipation heat engines at arbitrary power, *J. Stat. Mech. Theory Exp.* (2016) 073204.



## How Activity Landscapes Polarize Microswimmers without Alignment Forces

Nicola Andreas Söker<sup>1,\*</sup>, Sven Auschra<sup>2,†</sup>, Viktor Holubec<sup>2,3,‡</sup>, Klaus Kroy<sup>2,§</sup> and Frank Cichos<sup>1,||</sup>

<sup>1</sup>*Peter Debye Institute for Soft Matter Physics, Leipzig University, 04103 Leipzig, Germany*

<sup>2</sup>*Institute for Theoretical Physics, Leipzig University, 04103 Leipzig, Germany*

<sup>3</sup>*Charles University, Faculty of Mathematics and Physics, V Holešovičkách 2, CZ-180 00 Prague, Czech Republic*



(Received 12 January 2021; accepted 20 April 2021; published 1 June 2021)

Active-particle suspensions exhibit distinct polarization-density patterns in activity landscapes, even without anisotropic particle interactions. Such polarization without alignment forces is at work in motility-induced phase separation and betrays intrinsic microscopic activity to mesoscale observers. Using stable long-term confinement of a single thermophoretic microswimmer in a dedicated force-free particle trap, we examine the polarized interfacial layer at a motility step and confirm that it does not exert pressure onto the bulk. Our observations are quantitatively explained by an analytical theory that can also guide the analysis of more complex geometries and many-body effects.

DOI: [10.1103/PhysRevLett.126.228001](https://doi.org/10.1103/PhysRevLett.126.228001)

Active matter can succinctly be characterized as matter made from “animalcules,” a type of nonequilibrium molecules [1]. As a consequence, it can exhibit unusual material properties that would be strictly forbidden in conventional materials by symmetries implicit in the condition of (local) thermal equilibrium. The plethora of emergent phenomena observed in living matter has stimulated numerous laboratory studies of better controlled synthetic active fluids ([2], Table I), and the development of analytically tractable toy models [3,4] to describe them. A key feature vital for both biological and artificial microswimmers is their ability to adjust their motility in response to environmental cues [2,5–8]. The resulting heterogeneous motility [9–16] can be used to trap active particles [17,18] or to mimic quorum sensing [19]. In the case of intermittent “run-and-tumble” dynamics, an inverse relation between active-particle density and swim speed was predicted [20] and experimentally confirmed [21,22]. Dynamic clustering transitions known as motility-induced phase separation [23,24] have stimulated intense discussions of the “swim pressure” generated by active particles on solid boundaries [25–27] or between regions of different activity [28], and of the concomitant spontaneous polarization effects [14]. Despite recent progress in the theoretical description, precise experimental verifications remain challenging and plagued by confounding effects, such as hydrodynamic and complex physicochemical interactions among active particles and with physical boundaries.

In this Letter, we overcome these experimental difficulties by exploiting the precise control of autonomous Janus-type microswimmers via photon nudging [17,29–31]. With the help of a dedicated theoretical model (see also the companion paper [32]), we precisely quantify the colocalization of density modulations and polarization that

comes hand in hand with inhomogeneous swim speeds [10–12,19–22,26,28].

Importantly, apart from weak gravitational forces and interactions with the container walls, which are largely irrelevant to our experiment, our setup is entirely force free. A two-dimensional rectangular arena is divided into two regions of diverse activity, separated by a sharp activity step, as found as a concomitant feature of most actual physical boundaries. Here, however, it is realized without inserting any physical wall, as otherwise often done in experiments and computer simulations. We thereby avoid possible unintended (hydrodynamic, electrostatic, steric, ...) side effects, which could uncontrollably alter the particle density, current, and orientation. This allows us to experimentally confirm, on the single-particle level, that the interfacial polarization is emerging from unbalanced hidden bulk currents [14] rather than “controlling” the bulk states [33]. Our data reveal rich mutual relations between the particle density, polarization, and motility, which are all quantitatively explained by our analytical theory, which extends the mentioned asymptotic reciprocal relation of particle density and swim speed [20] to situations with continuous translational and rotational diffusion.

*Materials and methods.*—Janus microswimmers are constructed of a 1.5  $\mu\text{m}$  diameter polystyrene core (microParticles GmbH) and a 50 nm thin gold hemisphere. The particles are propelled by optically controlled self-thermophoresis [17,29–31]. The sample consists of a 2.4  $\mu\text{m}$  thin water film confined by two microscopy cover slips coated with Pluronic F127 to prevent particle adsorption, and sealed with polydimethylsiloxane to prohibit evaporation. The particle’s in-plane motion was observed in an inverted microscope Olympus (IX-73) under dark-field illumination (Olympus DF condenser) using 1 ms duration LED flashes (Thorlabs SOLIS-3C). The

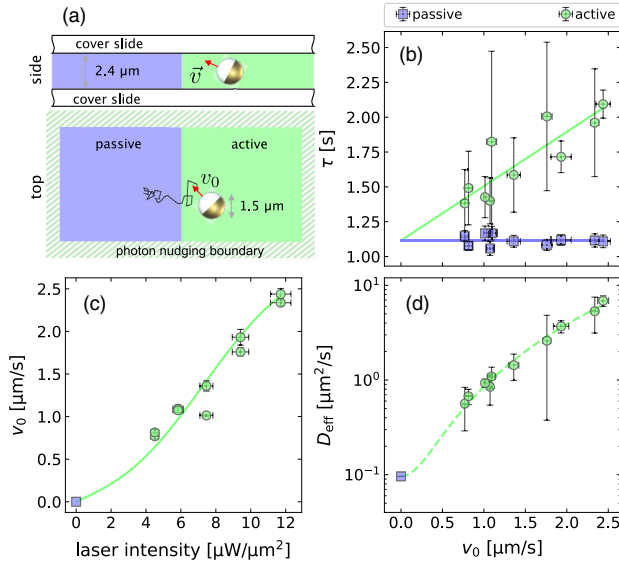


FIG. 1. Setup and parameter measurement. (a) Sketch of the rectangular arena, in which a  $1.5 \mu\text{m}$  Janus particle is confined by photon nudging. Position and in-plane orientation  $\mathbf{n}$  are observed in dark-field microscopy. (b) Orientational correlation time  $\tau$  for the active-passive bulk region as a function of the in-plane propulsion speed  $v_0$ . The horizontal and linear fits serve to identify  $\tau_{p,a}$ . (c) In-plane propulsion speed  $v_0$  as a function of the incident laser intensity, with a fit accounting for the weakly variable particle-wall alignment. (d) Effective diffusion coefficient  $D_{\text{eff}}$  in the active bulk, obtained from the slopes of the late-time MSD, and Eq. (1) (line) using  $\tau_a$  from (b). In (b)–(d) squares and circles correspond to the passive and active region, respectively, and error bars show 95% confidence intervals for Gaussian error propagation.

illumination was synchronized with a CCD camera (Hamamatsu Orca-Flash4.0 V2) at an inverse frame rate of 20 ms. The activity of the particle was adjusted by a  $\lambda = 532 \text{ nm}$  laser with a homogeneous intensity over the whole sample area. The illumination intensity is varied by an acousto-optic modulator (Isomet 1260C) between 0 and  $12 \mu\text{W} \mu\text{m}^{-2}$ . The corresponding feedback loop, which detects particle orientation and position in real time and adjusts the corresponding laser intensity is implemented in LABVIEW as described in Ref. [17]. The setup's overall feedback latency time amounts to 10 ms, which is negligible for the experimental results.

Photon nudging [29] as feedback control technique allowed us to define a virtual rectangular arena of  $6 \mu\text{m} \times 5 \mu\text{m}$  for the swimmer [Fig. 1(a)]. Outside the arena, the heating laser is turned on only if the swim direction of the Janus particle points toward the arena, yielding a mean propulsion speed of  $2.5 \mu\text{m} \text{ s}^{-1}$  in the on state. The central area of the arena is divided into a passive and an active region. All detected crossings of the particle between them trigger the laser to be switched off or on, irrespective of

particle orientation, corresponding to an activity landscape with a step profile  $v(x) = v_0 \Theta(x)$ . The step function  $\Theta(x)$  is 0 in the passive ( $p$ ,  $x < 0$ ) and 1 in active ( $a$ ,  $x > 0$ ) region. The accuracy of the experimental intensity step is given by the localization accuracy (about 40 nm) of the real-time tracking. The experiments were performed with a single Janus particle, with observation times of about 2 days for each of the laser powers depicted in Fig. 1(c).

**Results.**—The particle's in-plane motion is characterized by its translational and rotational diffusion coefficients  $D$ ,  $D_r$ , and, in the active region, its propulsion speed  $v_0$ . The parameters are estimated from the recorded in-plane mean-square displacement (MSD) and orientational autocorrelation function (ACF). Since the particle frequently commutes between the different regions, short trajectories need to be judiciously concatenated into longer ones, as detailed in the Supplemental Material [34], to assess the late-time dynamics. The deduced value of the diffusion coefficient  $D = 0.094 \mu\text{m}^2 \text{ s}^{-1} \pm 0.002 \mu\text{m}^2 \text{ s}^{-1}$  is roughly 1/3 of the expected Stokes–Einstein value  $k_B T / 6\pi\eta R$  in a bulk solvent, which we ascribe to hydrodynamic interactions of the swimmer with the confining cover slips [37]. Here,  $T = 295 \text{ K} \pm 2 \text{ K}$  is the solvent temperature,  $\eta = 0.9 \text{ mPa s}$  the dynamic viscosity,  $R = 0.77 \mu\text{m} \pm 0.04 \mu\text{m}$  the particle radius, and  $k_B$  the Boltzmann constant. The particle's in-plane orientational ACF is intrinsically multi-exponential [30] and affected by the Janus particle's bottom heaviness, causing preferential vertical alignment [38,39]. We therefore extract the longest decay time  $\tau$  of the orientational ACF by fitting the late-time ACFs with a single exponential function, to get an estimate for  $D_r$ . In the passive region, we obtained  $\tau = \tau_p = 1.12 \text{ s} \pm 0.05 \text{ s}$  [Fig. 1(b)], which is indeed close to the rotational correlation time  $1/(2D_r) = 1.27 \text{ s}$ , expected for a freely rotating sphere according to the (rotational) Einstein relation  $D_r \rightarrow k_B T (8\pi\eta R^3)^{-1} = 0.39 \text{ s}^{-1} \pm 0.06 \text{ s}^{-1}$  [40]. Restricting the analysis of the ACF in the passive region to data points where the particle exhibits strong in-plane alignment, we find a longer relaxation time  $\tau = 2.40 \text{ s}^{-1} \pm 0.05 \text{ s}^{-1}$ , in agreement with the expectation  $1/D_r$  [30] for free in-plane rotation. In the active region,  $\tau = \tau_a$  increases approximately linearly with the laser intensity from  $\tau_a = 1.4 \text{ s}$  to  $\tau_a = 2.0 \text{ s}$ , which can mainly be attributed to the particle's tendency to increasingly orient in plane in response to the thermo-osmotic flow fields generated by the heating in the confined geometry, and the torques exerted by the radiation pressure [41,42]. Finally, the effective diffusivity [26,43]

$$D_{\text{eff}}(x) = D + \frac{v(x)^2}{2D_r(x)} \xrightarrow{x \in \text{bulk}} D + \frac{v_0^2 \tau_a}{2}, \quad (1)$$

in the active region was deduced from the MSD. It grows nonlinearly in the laser intensity [Fig. 1(d)], and provides a consistency check for the parameters  $v_0$ ,  $D$ ,  $\tau_a$ .

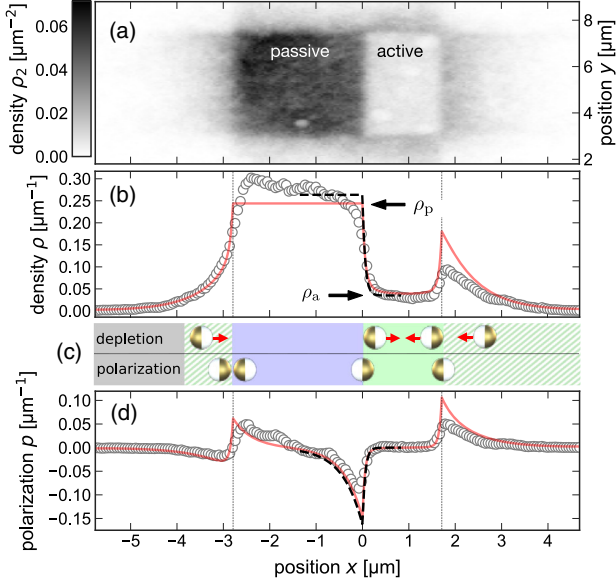


FIG. 2. Particle density and polarization. (a) Particle density in the sample plane for  $v_0 = 2.34 \mu\text{m s}^{-1}$  and  $\tau_a = 1.96 \text{ s}$  with the active-passive interface at  $x = 0$ . (b) Its integral along the  $y$  direction (excluding a  $0.7 \mu\text{m}$  lateral boundary region). The bulk densities  $\rho_p, \rho_a$  (upper, lower arrow) in the passive-active region were determined at about one decay length  $\lambda_{p,a}$  from the passive-active interface. The numerical solution [44] of Eq. (3) (solid line) employs reflecting outer boundaries. The (approximate) analytic solution from Eqs. (4)–(8) (dashed line) was normalized over half of the nominal width of the active and passive regions. (c) Sketch of the processes creating the interfacial polarization and depletion layers in a simplified model with binary particle orientations. (d) The experimental and theoretical (numerical and analytical) particle polarization [legend as in (b)]. The thin dotted vertical lines in (b) and (d) mark the borders to the confining photon-nudging region [cf. Fig. 1(a)]. Similar plots for other laser intensities can be found in the Supplemental Material [34].

Because of the spatially heterogeneous laser heating in the sample plane, the swim speed of the particle as well as the probability density  $\rho_2(x, y)$  to find it at a position  $(x, y)$  are spatially heterogeneous. The particle spends considerably more time in the passive than in the active region [Fig. 2(a)]. Outside of the central arena, the density decays due to photon nudging. Small imperfections in the data can most likely be attributed to mild statistical fluctuations due to the limited measurement time and localized defects in the Pluronic F127 coating of the cover slips. Integrating the density  $\rho_2(x, y)$  along the  $y$  direction, we obtain the marginal density  $\rho(x)$  shown in Fig. 2(b). It exhibits a pronounced step between essentially homogeneous active and passive bulk plateaus of height  $\rho_{a,p}$ . Further inspection reveals a colocalized excess of Janus orientations  $\mathbf{n}$  at the density step, pointing along  $-\mathbf{e}_x$ , toward the passive region [Fig. 2(d)]. This amounts to a negative average particle polarization

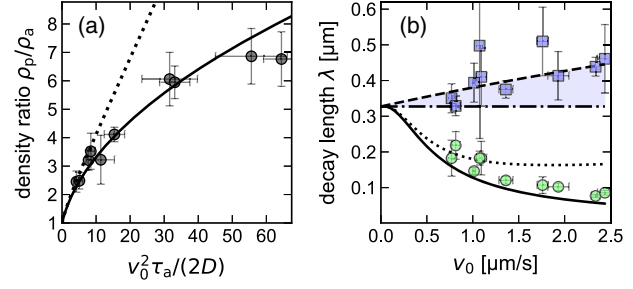


FIG. 3. Bulk density ratio and interface widths. (a) Measured ratio  $\rho_p/\rho_a$  of passive and active bulk densities as function of the (experimental) Péclet number  $v_0^2\tau_a/2D$  (circles). Our analytical prediction (8) with  $D_r = 1/\tau_a$  (solid line) improves that of Refs. [19,45], namely  $\rho_a/\rho_p = D_p/D_a$  (dotted line). (b) Decay lengths  $\lambda_{a,p}$  in the active (circles) and passive (squares) bulk regions as functions of  $v_0$ . The solid and dashed lines show the theoretical prediction (6) with  $D_r = 1/\tau_a$ , while the dot-dashed line assumes  $D_r = 1/\tau_p$ , with  $\tau_{a,p}$  from the fits in Fig. 1(b). The solid line for  $\lambda_a$  improves a prediction of Refs. [19,45] (dotted line). Error bars indicate the 95% confidence intervals of  $\rho$  and the Gaussian propagation of uncertainties for  $v_0, \tau_a, D$  from Fig. 1, respectively.

$$p(x) = \langle \mathbf{n} \cdot \mathbf{e}_x \rangle \rho(x), \quad (2)$$

where  $\langle \cdot \rangle$  denotes the time average. It decays approximately exponentially with the distance from the activity step. The characteristic decay length on the passive side is substantially longer than that on the active side,  $\lambda_p \gg \lambda_a$ . Similar polarization peaks are seen to occur at the interfaces to the photon-nudging boundary regions, which are regions of a different (orientation-dependent) type of activity [32]. Figure 3(a) shows that the ratio  $\rho_p/\rho_a$  of the passive and active bulk density plateaus increases as a function of the experimental estimate  $v_0^2\tau_a/(2D)$  for the Péclet number, which characterizes the activity of the Janus particle in the bulk region. Also, the decay lengths  $\lambda_a$  and  $\lambda_p$  depend on the activity contrast—the former decreasing and the latter slightly increasing (presumably due to the transient wall alignment, thus lower  $D_r$ , of trespassers) with  $v_0$ , as seen in Fig. 3(b). In summary, the main conclusion drawn from our experiments is that abrupt activity steps are accompanied by (i) smooth but pronounced steps in the particle density and (ii) the formation of skewed interfacial polarization layers of distinct widths and height.

*Theory.*—Our findings can be substantiated by a simple active-Brownian-particle model for the dynamic probability density  $f(x, \theta, t)$  to find the Janus particle at time  $t$  and position  $x$  with orientation angle  $\theta \equiv \arccos(\mathbf{n} \cdot \mathbf{e}_x)$  relative to the  $x$  axis. Its Fokker–Planck equation reads

$$\dot{f} = D\partial_x^2 f - \partial_x[fv(x)\cos\theta] + D_r\partial_\theta^2 f, \quad (3)$$

where  $D$  is the translational diffusion coefficient,  $v(x)$  the local propulsion speed, and  $1/D_r$  the orientational

correlation time. The stationary solution  $f(x, \theta)$  to Eq. (3) can be approximated by truncating the exact moment expansion with respect to the particle orientation [32,43] after the first two terms, yielding two coupled equations

$$\rho'(x) = p(x)v(x)/D, \quad (4)$$

$$p''(x) = p(x)/\lambda^2(x) + \rho(x)v'(x)/(2D), \quad (5)$$

for the particle density  $\rho(x) \equiv \int_0^{2\pi} d\theta f(x, \theta)$  and the polarization  $p(x) \equiv \int_0^{2\pi} d\theta f(x, \theta) \cos \theta$ . For the experimentally realized activity step (see Ref. [32] for a similar analysis at nudging interfaces) we find  $\lambda = \lambda_p \Theta(-x) + \lambda_a \Theta(x)$ , with

$$\lambda_p = (D_r/D)^{-1/2}, \quad \lambda_a = (D_r/D + v_0^2/2D^2)^{-1/2}. \quad (6)$$

In the two regions of constant activity, Eqs. (4) and (5) can be solved exactly [32]. With the matching condition across the interface, this yields the polarization profile

$$\frac{p_{a,p}(x)}{\rho_p} = -\frac{v_0}{2D} \frac{\lambda_a \lambda_p}{\lambda_a + \lambda_p} e^{-|x|/\lambda_{a,p}}, \quad (7)$$

and (assuming homogeneous  $D_r$ ) the bulk density ratio,

$$\frac{\rho_a}{\rho_p} = \sqrt{\frac{D_p}{D_a}} = \left(1 + \frac{v_0^2}{2DD_r}\right)^{-1/2} = \frac{\lambda_a}{\lambda_p}. \quad (8)$$

Observe that the ratio of the interfacial layer widths  $\lambda_{p,a}$ , the bulk-density ratio  $\rho_p/\rho_a$ , and the reduced peak polarization  $p_{a,p}(0)/\rho_{a,p}$  are kinetically determined quantities that only depend on the Péclet number  $v_0^2/(2DD_r)$ . In Ref. [32], we detail how the density ratio can be understood in terms of a detailed balance of two fluxes, a diffusion with the effective diffusivity (1), and a nonequilibrium flux due to its spatial heterogeneity.

*Discussion.*—A convincing parameter-free comparison of our analytical and experimental findings is obtained if we identify the parameter  $D_r$  of the model with the experimental  $1/\tau_a$ . This choice is not entirely trivial, since the theory describes planar rotational motion, whereas the rotational motion of the experimental particle is affected by its mass anisotropy, spatially heterogeneous optical forces, and the excited thermo-osmotic flows. Yet, assuming that uncertainties in position tracking (of about 40 nm) and the finite feedback-iteration time cause some low-pass filtering of the experimental curves, the comparison with the idealized analytical theory in Figs. 2(b) and 2(d) is very reasonable.

Intuitively, the (negatively) polarized boundary layer can most easily be understood by a caricature of the above modeling approach in terms of a two-species model that only admits left (−) and right (+) particle orientations, as

sketched in Fig. 2(c). The net polarization at the activity step is then immediately understood from the quasiballistic motion in the active region: it quickly drives the (+) particles to the right edge of the arena and the (−) particles across the interface into the passive region, where they get stuck and cause the negative interface polarization. The active region is thereby depleted relative to the passive region. The (up to a spurious factor 1/2) exact analogy between our analytical solutions of Eq. (3) and the schematic two-species model vindicates our approximation scheme in Eqs. (4) and (5) [32]. It also explains why, despite the superimposed Brownian motion, Eq. (8) coincides with the prediction for a run-and-tumble process [20]. Our results, moreover, improve recent theoretical predictions for quorum sensing [19,45].

Finally, also the skewed shape of the polarization layer is readily explained within the schematic two-species picture. Namely, during the characteristic reorientation time  $(2D_r)^{-1}$ , an initially perfectly polarized particle starting at the interface diffuses about a distance  $\sqrt{2D_r} = (D/D_r)^{1/2}$  into the passive region. On the active side, however, the same process is superimposed by self-propulsion, which acts like a sedimentation pressure. It provides a second channel to deplete the negative boundary layer polarization by driving trespassing (−) particles back across the interface. As a result, the interfacial polarization layer on the active side is diminished according to Eq. (6), which adds the two decay channels together. The total polarization

$$P_{\text{tot}} = \int_{-\infty}^{\infty} dx p(x) = -\frac{v_0 \rho_a}{2D_r} < 0 \quad (9)$$

is completely defined by the magnitude of bulk currents (here only  $v_0 \rho_a \neq 0$ ) and  $D_r$ , as a glance at Eqs. (6)–(8) confirms [14,32]. It is equivalent to the polarization expected at a solid boundary [14], where it represents the swim pressure [27] when multiplied by  $v_0/\mu$ ,  $\mu$  being the translational mobility. The abrupt motility step allows us to infer the swim pressure in the active bulk noninvasively, without a physical wall. Also note that it is not exerted across the interface onto the neighboring passive bulk phase. The lack of an alternative explanation in terms of a static equilibrium analogy [46] suggests that the precisely quantified polarization-density patterns can play the role of a smoking gun for particle-level activity.

In summary, we have employed the sophisticated technique of photon nudging to set up a boundary-free activity arena, thereby establishing a potent test bed to address fundamental open issues in active-particle physics. It enables us to observe active-particle motion in activity landscapes over several days. We found that motility gradients are accompanied by characteristic skewed interfacial polarization profiles. We showed them to arise from unbalanced dissipative active-particle fluxes (hidden in the bulk), thus not admitting any straightforward analogy with

equilibrium phase coexistence [16]. Our experiments are well described by a precise quantitative theory that advances previous work, can be generalized to photon nudging, and can deal with further phenomena such as wall accumulation and quorum sensing in active phase transitions [32]. It allows the swim pressure and essential microscopic parameters such as swim speed and (effective) translational and rotational diffusion coefficients to be inferred from accessible mesoscopic observables, namely the bulk particle-density and interfacial polarization profiles. And it suggests that similar polarization-density patterns are a hallmark of all microswimmer suspensions in heterogeneous activity landscapes, far beyond our artificial model system. As our experimental approach is capable of handling a controlled number of active particles simultaneously, a challenging but interesting avenue for future research would be to try and extend our experiments and theory to interacting many-body problems.

We acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG) through the priority program “Microswimmers” (SPP 1726, Project No. 237143019) and by the Czech Science Foundation (GACR Project No. 20-02955J). V.H. was supported by the Humboldt Foundation and N. A. S. by the European Social Fund (ESF), European Union (EU) and the Free State of Saxony (ESF Project No. 100327895).

\*nicola.soeker@uni-leipzig.de

†sven.auschra@itp.uni-leipzig.de

‡victor.holubec@mff.cuni.cz

§klaus.kroy@uni-leipzig.de

||frank.cichos@uni-leipzig.de

- [1] N. Lane, *Phil. Trans. R. Soc. B* **370**, 20140344 (2015).
- [2] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe, *Rev. Mod. Phys.* **88**, 045006 (2016).
- [3] M. E. Cates, *Rep. Prog. Phys.* **75**, 042601 (2012).
- [4] S. Contera, *Nano Comes to Life: How Nanotechnology is Transforming Medicine and the Future of Biology* (Princeton University Press, Princeton, 2019).
- [5] W. Poon, *Proc. Int. Sch. Phys. “Enrico Fermi”* **184**, 317386 (2013).
- [6] A. Zttl and H. Stark, *J. Phys. Condens. Matter* **28**, 253001 (2016).
- [7] A. Vikram Singh and M. Sitti, *Curr. Pharm. Des.* **22**, 1418 (2016).
- [8] D. Patra, S. Sengupta, W. Duan, H. Zhang, R. Pavlick, and A. Sen, *Nanoscale* **5**, 1273 (2013).
- [9] Y. Fily and M. C. Marchetti, *Phys. Rev. Lett.* **108**, 235702 (2012).
- [10] A. Sharma and J. M. Brader, *Phys. Rev. E* **96**, 032604 (2017).
- [11] A. Sharma and J. M. Brader, *J. Chem. Phys.* **145**, 161101 (2016).
- [12] H. Merlitz, H. D. Vuijk, J. Brader, A. Sharma, and J. U. Sommer, *J. Chem. Phys.* **148**, 194116 (2018).
- [13] H. D. Vuijk, A. Sharma, D. Mondal, J. U. Sommer, and H. Merlitz, *Phys. Rev. E* **97**, 042612 (2018).
- [14] S. Hermann and M. Schmidt, *Phys. Rev. Research* **2**, 022003(R) (2020).
- [15] L. Caprini and U. Marini Bettolo Marconi, *Soft Matter* **14**, 9044 (2018).
- [16] L. Caprini, U. Marini Bettolo Marconi, and A. Puglisi, *Phys. Rev. Lett.* **124**, 078001 (2020).
- [17] A. P. Bregulla, H. Yang, and F. Cichos, *ACS Nano* **8**, 6542 (2014).
- [18] S. Jahanshahi, C. Lozano, B. Liebchen, H. Lwen, and C. Bechinger, *Commun. Phys.* **3**, 127 (2020).
- [19] A. Fischer, F. Schmid, and T. Speck, *Phys. Rev. E* **101**, 012601 (2020).
- [20] M. J. Schnitzer, *Phys. Rev. E* **48**, 2553 (1993).
- [21] J. Arlt, V. A. Martinez, A. Dawson, T. Pilizota, and W. C. Poon, *Nat. Commun.* **9**, 768 (2018).
- [22] G. Frangipane, D. DellArciprete, S. Petracchini, C. Maggi, F. Saglimbeni, S. Bianchi, G. Vizsnyiczai, M. L. Bernardini, and R. di Leonardo, *eLife* **7**, 1 (2018).
- [23] J. Palacci, S. Sacanna, A. P. Steinberg, D. J. Pine, and P. M. Chaikin, *Science* **339**, 936 (2013).
- [24] I. Buttinoni, J. Bialké, F. Kümmel, H. Löwen, C. Bechinger, and T. Speck, *Phys. Rev. Lett.* **110**, 238301 (2013).
- [25] S. C. Takatori, W. Yan, and J. F. Brady, *Phys. Rev. Lett.* **113**, 028103 (2014).
- [26] M. E. Cates and J. Tailleur, *Annu. Rev. Condens. Matter Phys.* **6**, 219 (2015).
- [27] A. P. Solon, J. Stenhammar, R. Wittkowski, M. Kardar, Y. Kafri, M. E. Cates, and J. Tailleur, *Phys. Rev. Lett.* **114**, 198301 (2015).
- [28] H. Row and J. F. Brady, *Phys. Rev. E* **101**, 062604 (2020).
- [29] B. Qian, D. Montiel, A. Bregulla, F. Cichos, and H. Yang, *Chem. Sci.* **4**, 1420 (2013).
- [30] M. Selmke, U. Khadka, A. P. Bregulla, F. Cichos, and H. Yang, *Phys. Chem. Chem. Phys.* **20**, 10502 (2018).
- [31] M. Selmke, U. Khadka, A. P. Bregulla, F. Cichos, and H. Yang, *Phys. Chem. Chem. Phys.* **20**, 10521 (2018).
- [32] S. Auschra, N. Söker, V. Holubec, F. Cichos, and K. Kroy, companion paper, *Phys. Rev. E* **103**, 062601 (2021).
- [33] A. P. Solon, J. Stenhammar, M. E. Cates, Y. Kafri, and J. Tailleur, *Phys. Rev. E* **97**, 020602(R) (2018).
- [34] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.126.228001> for more details, which includes Refs. [35,36].
- [35] A. P. Solon, M. E. Cates, and J. Tailleur, *Eur. Phys. J. Special Topics* **224**, 1231 (2015).
- [36] M. Raible and A. Engel, *Appl. Organomet. Chem.* **18**, 536 (2004).
- [37] J. Happel and H. Brenner, Low Reynolds number hydrodynamics, in *Mechanics of Fluids and Transport Processes* Vol. 1 (Springer, Dordrecht, Netherlands, 1981).
- [38] D. E. O’Reilly, *J. Phys. Chem.* **74**, 3277 (1970).
- [39] Y. P. Kalmykov, *Phys. Rev. A* **45**, 7184 (1992).
- [40] P. Debye, *Polar Molecules* (Dover, New York, 1929).
- [41] S. Das, A. Garg, A. I. Campbell, J. Howse, A. Sen, D. Velegol, R. Golestanian, and S. J. Ebbens, *Nat. Commun.* **6**, 8999 (2015).
- [42] J. Simmchen, J. Katuri, W. E. Uspsal, M. N. Popescu, M. Tasinkevych, and S. Sánchez, *Nat. Commun.* **7**, 10598 (2016).

- [43] M. E. Cates and J. Tailleur, *Europhys. Lett.* **101**, 20010 (2013).
- [44] V. Holubec, K. Kroy, and S. Steffenoni, *Phys. Rev. E* **99**, 032117 (2019).
- [45] A. Fischer, F. Schmid, and T. Speck, *Phys. Rev. E* **102**, 059903(E) (2020).
- [46] A. P. Solon, J. Stenhammar, M. E. Cates, Y. Kafri, and J. Tailleur, *New J. Phys.* **20**, 075001 (2018).

# Supplemental Material for Active-Particle Polarization Without Alignment Forces

Nicola Andreas Söker,<sup>1,\*</sup> Sven Auschra,<sup>2,†</sup> Viktor Holubec,<sup>2,3,‡</sup> Klaus Kroy,<sup>2,§</sup> and Frank Cichos<sup>1,¶</sup>

<sup>1</sup>*Peter Debye Institute for Soft Matter Physics, Leipzig University, 04103 Leipzig, Germany*

<sup>2</sup>*Institute for Theoretical Physics, Leipzig University, 04103 Leipzig, Germany*

<sup>3</sup>*Charles University, Faculty of Mathematics and Physics, V Holešovičkách 2, CZ-180 00 Praha, Czech Republic*

(Dated: January 12, 2021)

The following two figures depict the experimentally obtained density and polarization profiles,  $\rho(x)$  and  $p(x)$ , for various particle activities, characterized by the Péclet number  $Pe \equiv v_0^2 \tau_a / (2D)$ . Here,  $v_0$ ,  $D$  and  $\tau_a$  denote the measured in-plane propulsion speed, translational diffusion coefficient and characteristic decay time of the swimmer's in-plane orientation within the active region, respectively. The solid lines depict the exact numerical solutions obtained from solving the Fokker-Planck equation for the ABP model, whereas the dashed curves correspond to our approximate analytic solutions discussed in the main text. The experimental profiles for  $Pe = 45.46$  are the ones displayed in Fig. 2 of the main text.

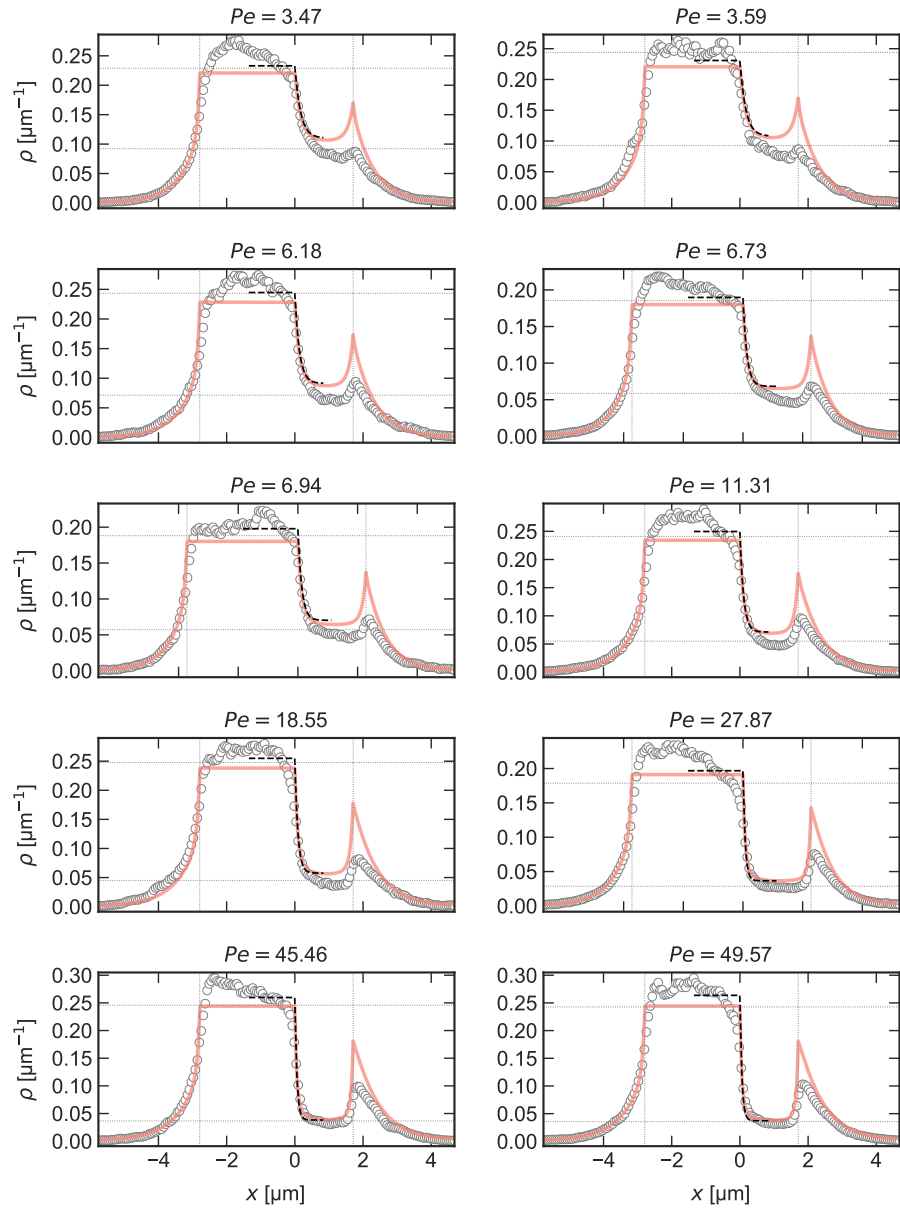


FIG. 1. Density profiles



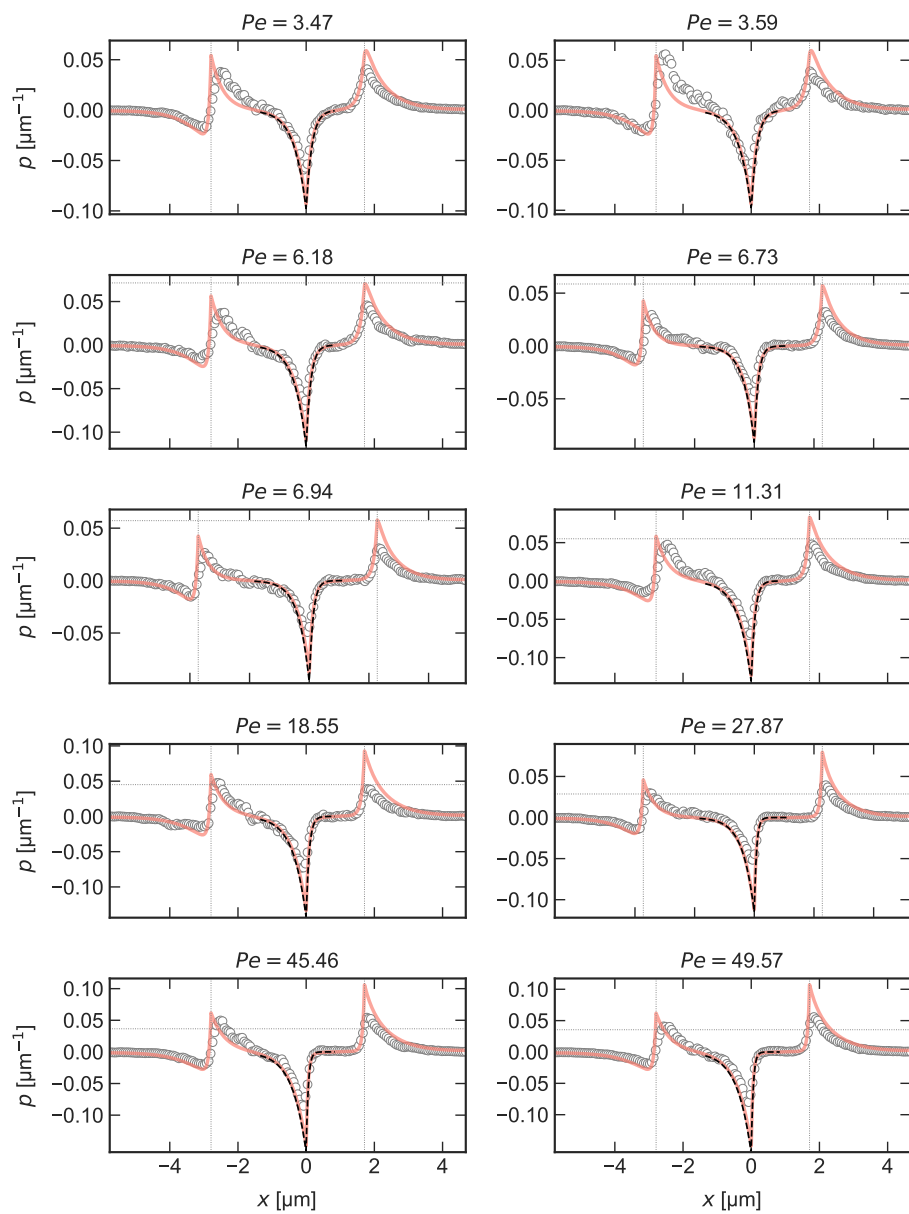


FIG. 2. Polarization profiles



**Polarization-density patterns of active particles in motility gradients**Sven Auschra<sup>1,\*</sup>, Viktor Holubec<sup>1,2,†</sup>, Nicola Andreas Söker<sup>1,3,‡</sup>, Frank Cichos<sup>3,§</sup> and Klaus Kroy<sup>1,||</sup><sup>1</sup>*Institute for Theoretical Physics, Leipzig University, 04103 Leipzig, Germany*<sup>2</sup>*Charles University, Faculty of Mathematics and Physics, V Holešovičkách 2, CZ-180 00 Prague, Czech Republic*<sup>3</sup>*Peter Debye Institute for Soft Matter Physics, Leipzig University, 04103 Leipzig, Germany*

(Received 12 January 2021; accepted 21 April 2021; published 1 June 2021)

The colocalization of density modulations and particle polarization is a characteristic emergent feature of motile active matter in activity gradients. We employ the active-Brownian-particle model to derive precise analytical expressions for the density and polarization profiles of a single Janus-type swimmer in the vicinity of an abrupt activity step. Our analysis allows for an optional (but not necessary) orientation-dependent propulsion speed, as often employed in force-free particle steering. The results agree well with measurement data for a thermophoretic microswimmer presented in the companion paper [Söker *et al.*, *Phys. Rev. Lett.* **126**, 228001 (2021)], and they can serve as a template for more complex applications, e.g., to motility-induced phase separation or studies of physical boundaries. The essential physics behind our formal results is robustly captured and elucidated by a schematic two-species “run-and-tumble” model.

DOI: [10.1103/PhysRevE.103.062601](https://doi.org/10.1103/PhysRevE.103.062601)**I. INTRODUCTION**

The surging field of active matter [1–3] aims for a microscopic understanding and control of the material properties of assemblies of interacting energy-consuming elements [4]. In particular, examples of *motile* active matter are ubiquitous in nature, ranging from flocks of birds [5], to swarms of insects [6], to colonies of bacteria such as *Escherichia coli* [7]. The wealth of observed natural phenomena has stimulated many laboratory studies of artificial active fluids of suspended inanimate microswimmers (see Ref. [4], Table I). Such “active-particle systems” often consist of simple colloidal particles propelled by a form of self-phoresis [8–15]. Numerous interesting features have been observed already on the level of a single or a few active particles [10,16–22], which open a wide range of potential applications [23–34]. Microswimmers moreover exhibit rich collective dynamics, ranging from mesoscopic turbulence via collective oscillations to macroscopic motility-induced phase separation (MIPS) [35–47].

A key feature vital for both biological and synthetic microswimmers is their ability to adjust their motility in response to environmental cues [4,48–51]. The omnipresence of heterogeneous motility on all scales of active matter has inspired many theoretical and experimental studies [19,43,47,52–74].

Inhomogeneous swim velocities go hand in hand with modulations in particle density and polarization [43,54–58,75]. These patterns might be exploited to polarize and

“transport” active particles along activity waves [19,63,64], find applications in the context of target finding [59,69], quorum-sensing [65,66], or to build motility traps [60]. Their precise mathematical characterization could also help to identify mesoscopic signatures of a motile-particle fraction [75].

The paradigmatic problem for a single active particle moving in a heterogeneous activity/motility field was experimentally studied in Ref. [75] (referred to hereafter as paper I). The crux of the experimental setup is that it allows for long-time observations of a single autonomous microswimmer [76] near an abrupt activity step. The latter may be thought of as a concomitant feature of most physical boundaries (e.g., sedimentation [72,77–79], wall adsorption [70,80–82]), and even of collective phenomena such as MIPS interface formation [70,73,83–87]. Importantly, the setup, schematically sketched in Fig. 1, confines the active particle to a planar arena by photon nudging [24,25], hence without imposing any lateral physical boundaries or confinement forces. It thereby enables the experimental study of the emerging interfacial patterns at the central activity step without any of the fundamentally unrelated perturbations usually encountered in practical applications. In many ways, the setup can thus be likened to the idealized textbook quantum-mechanics problem of a particle in a potential well. The main finding in Ref. [75] is an emerging characteristic polarization-density pattern in an interfacial layer around the motility step, which can serve as a distinctive trait of active versus passive Brownian particle motion. And the main tasks of the present contribution are its precise theoretical computation and the discussion of its physical implications.

To this end, we employ the (standard) active-Brownian-particle (ABP) model [3,39,88–90] for a single motile spherical particle.

We show that the well-known, intuitive, and experimentally confirmed [61,62] result that (one-particle) “density”  $\times$

\*sven.auschra@itp.uni-leipzig.de

†viktor.holubec@mff.cuni.cz

‡nicola.soeker@uni-leipzig.de

§cichos@physik.uni-leipzig.de

||klaus.kroy@itp.uni-leipzig.de

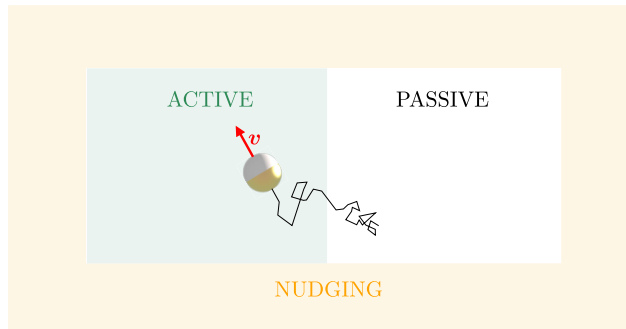


FIG. 1. Sketch of the experimental setup studied in paper I [75] (not to scale). Photon nudging [24,25] is employed to confine a self-thermophoretic Janus swimmer within a rectangular arena without physical boundaries or confining forces. Its in-plane orientation and position are observed by microscopy. Upon entering from the PASSIVE into the ACTIVE region, the particle's Brownian motion gets boosted by self-propulsion with velocity  $\mathbf{v}$  along its symmetry axis. This is a special, exactly solvable case of the general setup depicted in Fig. 2 covered by the theory presented in this paper.

“swim speed” = const [54] for run-and-tumble particles does not generally hold in situations of continuous translational and rotational diffusion. We furthermore extend the analysis to the case of orientation-dependent propulsion speeds [68], as often employed for force-free particle steering [23,91]. No such orientation dependence is needed, however, either in the motility or in the physical interactions, to generate the experimentally observed characteristic polarization-density patterns. This becomes explicitly clear from our precise analytical expressions for the case of a sharp motility step. They supersede a number of literature results. For example, the Green-Kubo approach [56] of Ref. [55] would lead to a symmetric polarization profile, described by a single characteristic length scale, while the interfacial layers actually have distinct activity-dependent widths. Further, comparing analytical and exact numerical solutions, we demonstrate the superiority of our approach over attempts to push the moment expansion for the one-particle density to higher orders [65]. Our explicit results moreover verify a general relation derived in Ref. [70], namely that the total polarization associated with a sudden activity step, which is accompanied by a corresponding jump in the osmotic pressure [58], is uniquely determined by bulk quantities, whereas the reverse statement [85] does not generally hold. While the corresponding “sum rule” [70] promotes a nonequilibrium flux (difference) to the status of a thermodynamic state variable, the interfacial polarization and pressure retain a fundamentally different status from that of molecules or passive colloids at an equilibrium phase boundary. The latter point was recently also emphasized for the collective velocity alignment in MIPS [47]. We finally show that our highly accurate but approximate analytic theory can be mapped onto an exact solution of a two-species run-and-tumble model. This link provides an intuitive physical picture that elucidates our key findings, and a physical explanation for the high accuracy and the broad range of applicability of our analytical theory, which thereby suggests itself as an efficient practical tool for the approximate reconstruction of the full picture from

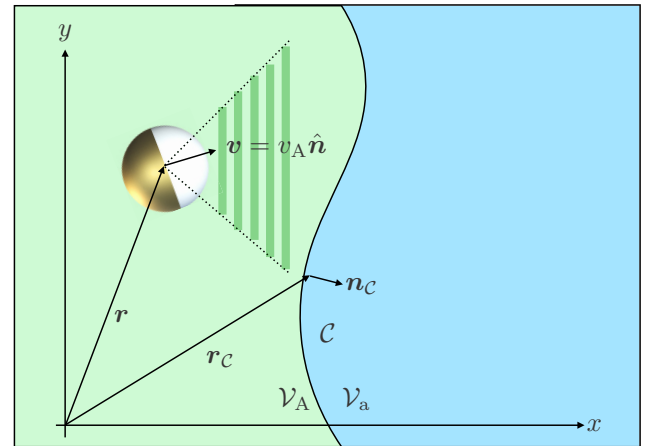


FIG. 2. A Janus particle with orientation  $\hat{\mathbf{n}}$  and position  $\mathbf{r}$ . The planar arena is divided into two subregions  $\mathcal{V}_{A,a}$  with distinct propulsion speeds,  $v_{A,a}$ . An optionally restricted acceptance range of particle orientations (shaded area) allows us to account for photon nudging. The interface  $\mathcal{C}$  between the regions  $\mathcal{V}_{A,a}$  is parametrized by its position vector  $\mathbf{r}_c$  and normal vector  $\mathbf{n}_c$  (pointing toward  $\mathcal{V}_a$ ).

incomplete and coarse-grained active-particle data for more complex geometries and interacting many-body problems.

## II. GENERAL THEORY

### A. Moment equations

We idealize the experimental thermophoretic microswimmer by the standard ABP model [3,39,88–90] of an overdamped particle, whose propulsion speed  $v(\mathbf{r}, \hat{\mathbf{n}})$  depends on its position  $\mathbf{r}$  and optionally also on its orientation  $\hat{\mathbf{n}}$ , according to the Langevin equations

$$\partial_t \mathbf{r} = v(\mathbf{r}, \hat{\mathbf{n}}) \hat{\mathbf{n}} + \sqrt{2D} \boldsymbol{\xi}_r, \quad \partial_t \hat{\mathbf{n}} = \sqrt{2D_r} \boldsymbol{\xi}_r \times \hat{\mathbf{n}}. \quad (1)$$

Here,  $D$  and  $D_r$  are the diffusion coefficients corresponding to the independent, unit variance, unbiased Gaussian white-noise processes  $\boldsymbol{\xi}_{r,r}(t)$  pertaining to the particle's translation and rotation, respectively. Some notation is illustrated by Fig. 2 for a piecewise constant activity profile in a planar setup.

The time evolution of the dynamic probability density  $f(\mathbf{r}, \hat{\mathbf{n}}, t)$  for finding the Janus swimmer at time  $t$  at position  $\mathbf{r}$  with the in-plane orientation  $\hat{\mathbf{n}}$  is described by the Fokker-Planck equation (FPE) [39,90,92]

$$\partial_t f = D \nabla^2 f + D_r \nabla_n^2 f - \nabla \cdot [f v(\mathbf{r}, \hat{\mathbf{n}}) \hat{\mathbf{n}}] \quad (2)$$

for dimensionality  $d = 2, 3$ . Here,  $\partial_t$  denotes the partial time derivative, and  $\nabla^2$  and  $\nabla_n^2$  are the translational and rotational parts of the Laplacian acting on  $\mathbf{r}$  and  $\hat{\mathbf{n}}$ , respectively. To extract measurable predictions from the model, we truncate the (exact) moment expansion of  $f$  with respect to the orientation  $\hat{\mathbf{n}}$  [39,92,93]:

$$f(\mathbf{r}, \hat{\mathbf{n}}, t) = \frac{1}{S_d} [\rho(\mathbf{r}, t) + d \mathbf{p}(\mathbf{r}, t) \cdot \hat{\mathbf{n}}]. \quad (3)$$

Here we used the following abbreviations:

$$S_d \equiv \int d\hat{\mathbf{n}} \quad (\text{unit sphere surface}), \quad (4)$$

$$\rho(\mathbf{r}) \equiv \int d\hat{\mathbf{n}} f(\mathbf{r}, \hat{\mathbf{n}}) \quad (\text{particle density}), \quad (5)$$

$$\mathbf{p}(\mathbf{r}) \equiv \int d\hat{\mathbf{n}} \hat{\mathbf{n}} f(\mathbf{r}, \hat{\mathbf{n}}) \quad (\text{polarization density}). \quad (6)$$

Motivated by the experimental setup, the activity profile is modeled as  $v(\mathbf{r}, \hat{\mathbf{n}}) = v(\mathbf{r})\chi_{\mathcal{A}}(\hat{\mathbf{n}})$ . Here,  $v(\mathbf{r})$  reflects the occurrence of inhomogeneous activity in position space, and the indicator function  $\chi_{\mathcal{A}}(\hat{\mathbf{n}})$ , which is unity if  $\hat{\mathbf{n}} \in \mathcal{A} \subseteq \mathcal{S}_d$  and zero otherwise, accounts for a limitation of the activity to an acceptance range  $\mathcal{A}$  of the particle orientation. This enables us to include the effect of photon nudging. We also note that the truncation in Eq. (3) is not systematic with respect to a small parameter but is physically motivated. Namely, the moment equations for  $\rho(x)$  and  $\mathbf{p}(x)$  derived from it in Sec. II C for one-dimensional activity profiles  $v(x)$  can be mapped onto an exact description of a one-dimensional run-and-tumble model, which captures the relevant physics for arbitrary activity profiles and dimensions (see Sec. III A).

Multiplying Eq. (2) by 1 or  $\hat{\mathbf{n}}$ , using  $\nabla_{\hat{\mathbf{n}}}^2 \hat{\mathbf{n}} = -(d-1)\hat{\mathbf{n}}$ , and integrating the result over the orientational degrees of freedom then yields the two moment equations [2]

$$\partial_t \rho(\mathbf{r}, t) = -\nabla \cdot \mathbf{J}(\mathbf{r}, t), \quad (7)$$

$$\partial_t \mathbf{p}(\mathbf{r}, t) = -(d-1)D_r \mathbf{p}(\mathbf{r}, t) - \nabla \cdot \mathbf{M}(\mathbf{r}, t). \quad (8)$$

Here, we introduced the (orientation-averaged) flux,

$$\mathbf{J}(\mathbf{r}, t) \equiv -D\nabla \rho(\mathbf{r}, t) + v(\mathbf{r})[\mathcal{I}_1 \rho(\mathbf{r}, t) + \mathcal{I}_2 \mathbf{p}(\mathbf{r}, t)], \quad (9)$$

and the matrix flux,

$$\mathbf{M}(\mathbf{r}, t) \equiv -D\nabla \mathbf{p}(\mathbf{r}, t) + v(\mathbf{r})[\mathcal{I}_3 \rho(\mathbf{r}, t) + \mathcal{I}_4 \mathbf{p}(\mathbf{r}, t)]. \quad (10)$$

The quantities  $\mathcal{I}_k$ ,  $k = 1, \dots, 4$ , account for a possibly restricted acceptance range  $\mathcal{A}$  for propulsion, and they are defined as

$$\begin{aligned} \mathcal{I}_1 &\equiv \frac{1}{\mathcal{S}_d} \int_{\mathcal{A}} d\hat{\mathbf{n}} \hat{\mathbf{n}}, \\ \mathcal{I}_2 &= d \mathcal{I}_3 \equiv \frac{d}{\mathcal{S}_d} \int_{\mathcal{A}} d\hat{\mathbf{n}} \hat{\mathbf{n}} \hat{\mathbf{n}}, \\ \mathcal{I}_4 &\equiv \frac{d}{\mathcal{S}_d} \int_{\mathcal{A}} d\hat{\mathbf{n}} \hat{\mathbf{n}} \hat{\mathbf{n}} \hat{\mathbf{n}}. \end{aligned} \quad (11)$$

Note that for orientationally unrestricted propulsion, i.e.,  $\mathcal{A} = \mathcal{S}_d$  and  $\chi_{\mathcal{A}} = 1$ , the only contributing integrals are

$$\mathcal{I}_2 = d \mathcal{I}_3 = \mathbf{1}, \quad (12)$$

with the unit matrix  $\mathbf{1}$ .

### B. Steady state and continuity conditions

For the remainder of this section, we focus on the steady-state particle density and polarization. Vanishing time-derivatives in Eqs. (7) and (8) provide the stationarity conditions

$$D\nabla^2 \rho(\mathbf{r}) = \nabla \cdot \{v(\mathbf{r})[\mathcal{I}_1 \rho(\mathbf{r}) + \mathcal{I}_2 \mathbf{p}(\mathbf{r})]\}, \quad (13)$$

$$\begin{aligned} D\nabla^2 \mathbf{p}(\mathbf{r}) &= (d-1)D_r \mathbf{p}(\mathbf{r}) \\ &+ \nabla \cdot \{v(\mathbf{r})[\mathcal{I}_3 \rho(\mathbf{r}) + \mathcal{I}_4 \mathbf{p}(\mathbf{r})]\}. \end{aligned} \quad (14)$$

Furthermore, it turns out that for all setups considered here, the no-flux boundary conditions imply that the steady-state flux  $\mathbf{J}(\mathbf{r})$  vanishes at each point in space. Equation (9) then implies

$$\nabla \rho = \frac{v}{D} (\mathcal{I}_1 \rho + \mathcal{I}_2 \mathbf{p}), \quad (15)$$

which we substitute into Eq. (14).

Let us now consider two domains of constant activity,  $\mathcal{V}_A$  and  $\mathcal{V}_a$ , whose interface is described by a hyperplane  $\mathcal{C}$ , as sketched in Fig. 2. For  $\mathbf{r} \in \mathcal{V}_{A,a}$ , the Janus particle propels at constant swim speed  $v_{A,a}$ , given that its orientation lies within the acceptance range of nudging. Upon crossing the interface  $\mathcal{C}$ , the swimmer experiences a sudden change in its activity. The respective solutions  $\rho_{A,a}(\mathbf{r})$  and  $\mathbf{p}_{A,a}(\mathbf{r})$  of the steady-state moment Eqs. (13) and (14) within each domain  $\mathcal{V}_{A,a}$  have to be matched at the interface  $\mathcal{C}$ . Besides continuity of  $\rho$  and  $\mathbf{p}$  itself, we demand the normal components  $\mathbf{J} \cdot \mathbf{n}_C$  and  $\mathbf{M} \cdot \mathbf{n}_C$  of both fluxes to be continuous at each point  $\mathbf{r}_C$  along the interface  $\mathcal{C}$  [94]. The surface normal  $\mathbf{n}_C$  is defined to point toward  $\mathcal{V}_a$ . Computing the limits  $\lim_{|\epsilon| \rightarrow 0} r_C \pm \epsilon$  in Eqs. (9) and (10) delivers the following two continuity conditions:

$$\mathbf{n}_C \cdot (\mathbf{J}_a - \mathbf{J}_A) = \mathbf{0}, \quad \mathbf{n}_C \cdot (\mathbf{M}_a - \mathbf{M}_A) = \mathbf{0}. \quad (16)$$

While the first of these relations is obvious from  $\mathbf{J} \equiv \mathbf{0}$ , both follow formally by integrating the (stationary) moment Eqs. (13) and (14) over an infinitesimal area around some point on the interface  $\mathcal{C}$  and exploiting the divergence theorem.

In the next paragraph, we discuss analytical solutions for the density  $\rho$  and polarization  $\mathbf{p}$  for a straight-interface geometry as depicted in Fig. 1.

### C. Activity step in an infinite planar arena

Consider a situation in which the Janus swimmer faces an orientation-independent activity with a step at an infinite straight interface in  $d = 2$  dimensions. For  $x < x_{if}$ , the particle propels at a swim speed  $v_A$ , which abruptly reduces to  $v_a < v_A$  upon crossing the interface at  $x = x_{if}$ . Due to the translational symmetry of the system in the  $y$ -direction, we project the particle dynamics onto the  $x$ -axis by replacing  $(\mathbf{r}, \hat{\mathbf{n}})$  by  $(x, \cos \theta)$  in the general equations above.

In this situation, the only nonzero coefficients (11) are given by Eq. (12), and thus the flux balance (15) reads

$$\rho'(x) = \frac{v(x)}{D} \rho(x). \quad (17)$$

Plugging this relation into the moment Eq. (14) yields

$$p''(x) = \frac{p(x)}{\lambda^2(x)} + \frac{v'(x)\rho(x)}{2D}, \quad (18)$$

where we defined the natural relaxation length

$$\lambda(x) \equiv \left[ \frac{D_r}{D} + \frac{v^2(x)}{2D^2} \right]^{-1/2}. \quad (19)$$

For the considered piecewise constant activity profile, Eq. (18) boils down within each half-space to  $p''_{A,a}(x) = p_{A,a}(x)/\lambda_{A,a}^2$  with the characteristic length scales  $\lambda_{A,a}$  defined by Eq. (19).

Their relation can be expressed as

$$\frac{\lambda_A}{\lambda_a} = \sqrt{\frac{1 + \mathcal{P}_a}{1 + \mathcal{P}_A}}, \quad \mathcal{P}_{A,a} \equiv \frac{v_{A,a}^2}{2DD_r}, \quad (20)$$

with the Péclet numbers  $\mathcal{P}_{A,a}$  weighing active versus diffusive transport rates in the respective regions.

The general solutions of the polarization profiles are

$$p_{A,a}(x) = C_{A,a}^{(1)} e^{x/\lambda_{A,a}} + C_{A,a}^{(2)} e^{-x/\lambda_{A,a}} \quad (21)$$

and the density profile follows by integrating Eq. (17) from an arbitrary reference point  $x_0$ :

$$\rho(x) = \rho(x_0) + \int_{x_0}^x dx \frac{v(x)}{D} p(x). \quad (22)$$

The boundary term  $\rho(x_0)$  follows from the normalization condition for the density, and the integration constants  $C_{A,a}^{(1,2)}$  are determined by the boundary and matching conditions at the activity step. The latter read

$$\rho_A(x_{if}) = \rho_a(x_{if}), \quad p_A(x_{if}) = p_a(x_{if}), \quad (23)$$

$$p'_a(x_{if}) - p'_A(x_{if}) = \frac{\rho_{A,a}(x_{if})}{2D} (v_a - v_A), \quad (24)$$

where the second line follows from Eq. (16) while using  $v'(x) = (v_a - v_A)\delta(x - x_{if})$ , with the delta function  $\delta(x)$ .

The natural boundary conditions in an infinite arena are [95]  $p(|x| \rightarrow \infty) = 0$ . The continuity condition (23) then implies that the polarization profiles around an activity step at the origin,  $x_{if} = 0$ , take the form

$$p_{A,a}(x) = p_{\max} e^{-|x|/\lambda_{A,a}}, \quad (25)$$

with an unknown maximum polarization  $p_{\max}$ . The density profile follows via Eq. (22) as

$$\rho_A(x) = \rho_A + \frac{p_{\max}}{D} v_A \lambda_A e^{x/\lambda_A}, \quad (26)$$

$$\rho_a(x) = \rho_a + \frac{p_{\max}}{D} (v_A \lambda_A + v_a \lambda_a - v_a \lambda_a e^{-x/\lambda_a}), \quad (27)$$

so that  $\rho(x \rightarrow \pm\infty)$  attains the regional constant bulk densities  $\rho_{a,A}$ .

A suitable order parameter for the polarization at the interface is the relative maximum polarization  $p_{\max}/\rho(0)$ . Using Eq. (24), this ratio can be expressed as

$$\frac{p_{\max}}{\rho(0)} = \frac{v_A - v_a}{2D} \frac{\lambda_A \lambda_a}{\lambda_A + \lambda_a} = \frac{1}{\sqrt{2}} \frac{\sqrt{\mathcal{P}_A} - \sqrt{\mathcal{P}_a}}{\sqrt{1 + \mathcal{P}_A} + \sqrt{1 + \mathcal{P}_a}}. \quad (28)$$

Its sign is via Eq. (25) shared by the whole polarization profile  $p(x)$  and solely determined by the difference in the swim speeds  $v_A - v_a$ . The Janus swimmer thus preferably points from the more into the less active region. The maximum  $1/\sqrt{2}$  of  $|p_{\max}|/\rho(0)$  is reached for  $\mathcal{P}_A \rightarrow \infty$  and  $\mathcal{P}_a = 0$  (or vice versa). It is less than 1, because we consider particle rotations in two dimensions in a projection onto the ( $x$ -)axis of the activity gradient.

By Eqs. (25)–(28),  $p_{A,a}(x)/\rho(0)$  and  $[\rho(x) - \rho_A]/\rho(0)$  are uniquely determined. The bulk density ratio

$$\frac{\rho_a}{\rho_A} = \frac{\lambda_a}{\lambda_A} = \sqrt{\frac{1 + \mathcal{P}_A}{1 + \mathcal{P}_a}} \quad (29)$$

is derived in Appendix A 1.

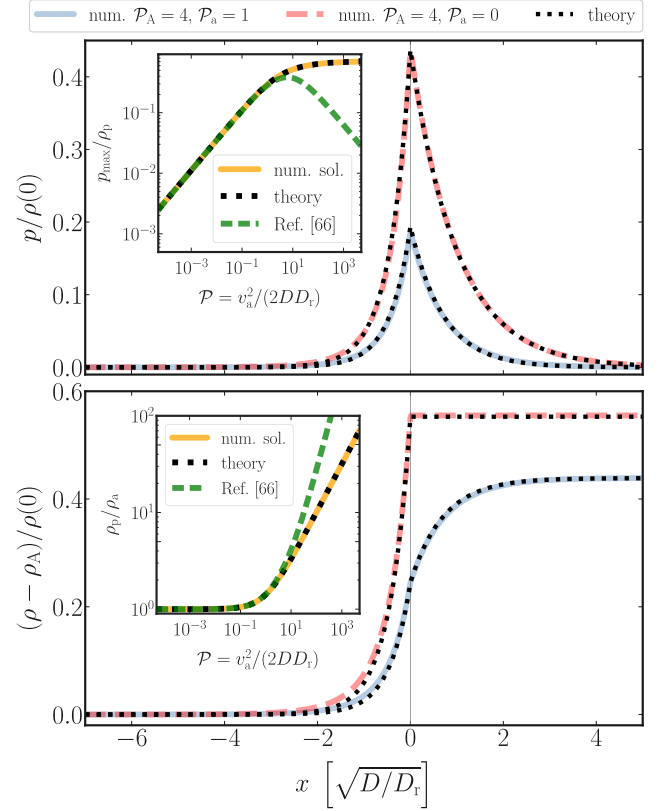


FIG. 3. Particle density and polarization profiles—approximate theory vs exact numerical solutions at an activity step at  $x = 0$ , between a high-activity ( $A$ ) and a low-activity ( $a$ ) or entirely passive ( $p$ ) region. Upper panel: The reduced polarization from Eqs. (25) and (28) closely matches the exact numerical solutions (dashed and solid lines). Inset: the polarization peak at the active-passive interface as a function of the active Péclet number  $\mathcal{P}$  from Eq. (28) (dotted) precisely follows the exact numerical solution (solid line), improving upon earlier predictions [65,66] (dashed). Lower panel: The reduced density ratio  $(\rho - \rho_A)/\rho(0)$  from Eqs. (26)–(28) (dotted) closely matches the exact numerical solutions (dashed and solid lines). Inset: the density ratio  $\rho_p/\rho_a$  for the active-passive interface as a function of the Péclet number  $\mathcal{P}$  from Eq. (29) (dotted line) precisely follows the numerical results (solid line), improving upon earlier predictions [65,66] (dashed line).

In the remainder of this paragraph and in Sec. III E, we show that this result holds far beyond the scope of the applied approximation [truncated moment expansion (3)], and for arbitrary activity profiles mediating between the two bulk states.

Figure 3 compares the approximate theory profiles following from Eqs. (25)–(29) to the (exact) numerical solutions obtained from Eq. (2) using the method of Ref. [96]. The two cases  $0 < v_a < v_A$  and  $0 = v_a < v_A$  are considered. The characteristic length scales  $\lambda_i$ , bulk densities  $\rho_i$ , and Péclet numbers  $\mathcal{P}_i$  corresponding to the highly active ( $x < 0$ ) and less active/completely passive regions ( $x \geq 0$ ) are distinguished by the subscripts  $A$ ,  $a$ , and  $p$ , respectively. The analytic solutions for the polarization and density profiles in

the main panels agree nicely with the numerical solutions, with slight deviations appearing only for substantial activity.

Such an impressive performance of the approximate model is unexpected. Intuitively, the approximation (3) should break down for  $\mathcal{P} \simeq 1$ . Its remarkable accuracy far beyond this limit becomes more plausible by observing that its moment equations (17) and (18) map onto an exactly solvable two-species model (one-dimensional run-and-tumble [54,97,98] accompanied by diffusion [67,99]; see Sec. III A). The latter robustly captures the pertinent physics (though not all quantitative details) at the motility step for arbitrary Péclet numbers. Indeed, the quantitative agreement between the approximate analytical and the exact numerical profiles, which we obtain using the method described in Ref. [96], remains very good up to Péclet numbers on the order of 100, as independently discovered in Ref. [100]. For even higher activities, higher moments in the moment expansion (3) become nonzero. However, formally pushing the moment expansion up to the third order actually leads to less precise numerical predictions than our physically motivated second-order approximation, already for intermediate Péclet numbers on the order of 1.

From the upper panel in Fig. 3, we infer that the relative polarization  $p_{\max}/\rho(0)$  peaks exactly at the motility step with a magnitude given by Eq. (28), which improves the predictions of Refs. [65,66] (see the inset). The polarization decays exponentially as a function of the distance from the interface. The polarization layers extend over the characteristic length scale  $\lambda$ , defined in Eq. (19). Its limiting forms for vanishing and diverging self-propulsion are  $\lambda \sim \sqrt{D/D_r}$  and  $\lambda \sim \sqrt{2D/v(x)}$ , respectively, i.e., the boundary layer is compressed by the particle's self-propulsion. (For a physical interpretation, see the discussion in Sec. III.) The inset provides strong numerical evidence that Eq. (28) for the reduced polarization peak generally holds for arbitrarily large activity steps. The lower panel of Fig. 3 presents the corresponding density profiles. For the active-passive interface, the density profile remains constant at the bulk density  $\rho = \rho_p$  throughout the whole passive region by virtue of Eq. (17). On crossing the interface, it decays to the bulk density  $\rho_A < \rho_p$  pertaining to the active region over a length scale  $\lambda_A$ . In the case of finite activity in both regions ( $0 < v_a < v_A$ ), the density still displays a kink at the interface, but it decays over the length scales  $\lambda_{a,A} > 0$  toward its bulk density values  $\rho_{a,A}$ . The inset of the lower panel of Fig. 3 compares the analytical expression (29) for the bulk density ratio  $\rho_p/\rho_a$  at an active-passive interface with the exact numerical solution. Both curves perfectly overlap over a vast range of Péclet numbers  $\mathcal{P}$ , demonstrating that Eq. (29) holds far beyond the scope of the approximation (3), and again improving predictions by Refs. [65,66].

The following paragraph points out a number of interesting consequences of the above findings and puts them into a broader context.

#### D. Various consequences and ramifications: Wall accumulation, MIPS, and swim pressure

##### 1. Wall accumulation

Equations (17) and (18) can serve to calculate (approximate) polarization and density profiles at various kinds of interfaces or boundaries, which enter the governing equations

only via different boundary conditions. As a concrete example, consider an active particle confined by hard reflecting walls [94]. Our model captures the well-known effects of polarization toward and accumulation at walls [67,70,80–82,99]. The simplest way to compute specific profiles is by exploiting the equivalence between the approximate model, Eqs. (17) and (18), and the two-species model (see Sec. III A), and we compute  $\rho(x)$  and  $p(x)$  within the latter, as done in [67,99]. The polarization and density layers near the wall are still determined by the same boundary-layer width  $\lambda$  in Eq. (19), and they exhibit the same physics as those at motility steps (see the discussion in Sec. III B).

## 2. MIPS

An extensively studied feature of active-Brownian-particle suspensions is the so-called motility-induced phase separation (MIPS) [39,43,73,83–85,101]. In a nutshell, it relies on a positive feedback between (i) the slowing down of particles in “crowded” areas [54], and (ii) active-particle accumulation in low-motility areas [43]. Our results for the polarization and density profiles of a single (overdamped) Janus swimmer can, at least on a qualitative level, closely mimic the effects observed for MIPS if applied to a gas of noninteracting active Janus spheres experiencing an inhomogeneous activity profile. As also pointed out in Ref. [100], such a gas exhibits the following features observed in MIPS:

(i) The system divides into two bulk areas with different bulk densities. Particles move slower (faster) in the denser (less dense) regions [43,85] [cf. Eq. (29) and the lower panel of Fig. 3].

(ii) An interfacial region forms between the two bulk areas, where particles, on average, point into the denser phase [73,83,85,102] (cf. the upper panel of Fig. 3). Note that for active Lennard-Jones particles, an opposite polarization was observed [87].

(iii) Smaller activity corresponds to a wider polarization layer [101] [cf. Eq. (19)].

In the particular case of quorum-sensing particles, adjusting their activity according to the local density (e.g., via chemical signaling) [103], Eq. (29) for the bulk density ratio can be used to improve the corresponding literature results [65,66] based on a dynamic mean-field theory.

## 3. Interface polarization

The total polarization  $P_{\text{tot}}$ , defined as the integrated local polarization profile  $p(x)$  over a suitable (sub)volume, was rigorously shown to obey an exact global sum rule by Hermann and Schmidt [70]. (Note that, without alignment interactions, the integral over the whole space always vanishes, because the polarization arises from spatial sorting of differently oriented particles, without changing their overall numbers—see Sec. III B for the sorting mechanism.) For our active-passive interface between two bulk regions with vanishing polarization,  $p(x_{\text{bulk}}^{A,a}) = 0$ , it is reasonable to define the total interface polarization as the integral from the bulk region  $x_{\text{bulk}}^A$  to the left of the interface to the bulk region  $x_{\text{bulk}}^a$  to its right. Using Eqs. (25) and (29), we find that within our approximate model the total interface polarization is then given by

(Appendix A 1)

$$P_{\text{tot}} = \int_{x_{\text{bulk}}^A}^{x_{\text{bulk}}^a} dx p(x) = p_{\text{max}}(\lambda_a + \lambda_A) = \frac{v_A \rho_A - v_a \rho_a}{2D_r}. \quad (30)$$

This exactly agrees with the sum rule of Ref. [70], which is thus verified within our approximate theory. The total interface polarization is accordingly determined by the difference of the otherwise “hidden” bulk fluxes  $v_{A,a}\rho_{A,a}$ , and both may be addressed as thermodynamic state variables, even though, as emphasized in Ref. [47], the dissipative, nonequilibrium origin of this kind of force-free velocity alignment of active Brownian particles sets it apart from any equilibrium interface polarization.

According to Eq. (30),  $P_{\text{tot}}$  points from the region of higher activity into the region of lower activity and only vanishes for homogeneous activity, i.e.,  $\mathcal{P}_A = \mathcal{P}_a$ . This becomes more obvious using Eqs. (20) and (29), for which one finds

$$\frac{2D_r}{v_A \rho_A} P_{\text{tot}} = 1 - \sqrt{\frac{1 + \mathcal{P}_A^{-1}}{1 + \mathcal{P}_a^{-1}}} > 0, \quad (31)$$

where the last inequality follows from  $v_A > v_a$ .

#### 4. Mechanical swim pressure

Suspended active Brownian particles contribute a “swim pressure” [104,105] to the overall mechanical stress. Like the hidden bulk fluxes that are responsible for it, it usually only becomes apparent when it is (internally) unbalanced, e.g., at confining soft or hard container walls. For a homogeneous and isotropic suspension of active Brownian spheres with constant swim speed  $v_0$  and bulk density  $\rho_0$ , this pressure is given by [106,107]

$$\frac{v_0}{\mu} \int_{x_{\text{bulk}}}^{\infty} dx p(x) = \frac{v_0^2 \rho_0}{2\mu D_r}, \quad (32)$$

where  $\mu$  denotes the particle’s (translational) mobility. The underlying rationale is that an arrested oriented swimmer turns into an oriented pump that emits a Stokeslet corresponding to the Stokes friction  $v_0/\mu$ , which balances the stopping force. For our active-passive motility step, far away from any confining wall, the total interface polarization (30) is  $v_A \rho_A / (2D_r)$ . Multiplication by  $v_A/\mu$  makes it formally equivalent to the swim pressure in Eq. (32) with  $v_A^2 \rho_A$  in place of  $v_0^2 \rho_0$ . One can then interpret the product  $\rho_A v_A^2 / 2\mu D_r$  as the (unbalanced) swim pressure *at the interface*, although, due to the abrupt motility step, this pressure is actually not exerted *across the interface* onto the adjacent passive bulk.

### E. Generalizations

We now show that Eq. (29) for the bulk density ratio actually holds for the full ABP model (2) and arbitrary one-dimensional activity variations, and we extend the above discussion to finite domains.

#### 1. Density ratio

Consider an arbitrary activity profile  $v(x)$  that mediates between two bulk regions of constant activity. Using Eq. (2),

the corresponding stationary FPE for the probability density  $f(x, \theta)$  reads

$$0 = -\partial_x \mathcal{J} - \partial_\theta \mathcal{J}_\theta = D \partial_x^2 f + D_r \partial_\theta^2 f - \partial_x (v \cos \theta f), \quad (33)$$

with the (angle-resolved) translational and rotational currents

$$\mathcal{J}(x, \theta) \equiv -D \partial_x f(x, \theta) + v(x) \cos \theta f(x, \theta), \quad (34)$$

$$\mathcal{J}_\theta(x, \theta) \equiv -D_r \partial_\theta f(x, \theta). \quad (35)$$

In contrast to the truncation applied in Eq. (3), we now consider the full moment expansion of  $f$  with respect to orientation. For one orientational degree of freedom, this is equivalent to expanding  $f$  into the Fourier series

$$f(x, \theta) = \frac{\rho(x)}{2\pi} + \frac{1}{\pi} \sum_{n=1}^{\infty} f_n(x) \cos(n\theta), \quad (36)$$

with coefficients  $f_n = \langle \cos(n\theta) \rangle = \int_0^{2\pi} d\theta \cos(n\theta) f(x, \theta)$ . In accord with our previous discussion, the zeroth and the first coefficients are given by the density  $\rho \equiv \langle 1 \rangle$  and the polarization  $p \equiv f_1 = \langle \cos \theta \rangle$ , respectively. The orientation-averaged flux thus reads

$$J(x) \equiv \langle \mathcal{J}(x, \theta) \rangle = -D \rho'(x) + v(x) p(x). \quad (37)$$

Similarly, multiplication of the FPE (33) by  $\cos \theta$  and integration over  $\theta$  yields the differential equation for the polarization,

$$0 = D p'' - D_r p - \frac{v}{2} \rho' - \frac{\rho}{2} v' - \frac{1}{2} \partial_x (v f_2), \quad (38)$$

where we used the identity  $2 \cos^2 \theta = 1 + \cos(2\theta)$ .

Note that the last term in Eq. (38) was absent in the above discussions based on the truncated moment expansion (3), which amounts to the closure relation  $f_2 = \langle \cos(2\theta) \rangle = 0$ .

Isolating  $p$  from Eq. (38) and plugging it into Eq. (37) yields

$$J(x) = -D_{\text{eff}}(x) \rho'(x) - \frac{1}{2} \rho D_{\text{eff}}' + \tilde{J}(x), \quad (39)$$

with the position-dependent effective diffusivity

$$D_{\text{eff}}(x) \equiv D + \frac{v^2(x)}{2D_r} \quad (40)$$

and the flux

$$\tilde{J} \equiv \frac{D}{D_r} v(x) p'(x) - \frac{v}{2D_r} \partial_x (v f_2). \quad (41)$$

Equation (39) constitutes a generalized version of Fick’s law. The first contribution,  $-D_{\text{eff}} \rho'$ , accounts for isotropic diffusive transport. The effective diffusivity  $D_{\text{eff}} \geq D$  is enhanced by the swimmer’s short-term ballistic motion. The second term,  $\rho D_{\text{eff}}'/2$ , accounts for the spatial dependency of this effective diffusivity, and its prefactor 1/2 for the directionality of the active velocity. The last contribution,  $\tilde{J}$ , represents the influence of the polarization  $p(x)$  and higher moments  $f_n$ ,  $n > 1$ , on the local density.

The condition  $J(x) = 0$  of a vanishing steady-state flux (39) yields

$$\frac{\rho'}{\rho} = -\frac{1}{2} \frac{D_{\text{eff}}'}{D_{\text{eff}}} + \frac{\tilde{J}}{\rho D_{\text{eff}}}. \quad (42)$$



Integrating this equation from a reference point  $x_0$  up to an arbitrary position  $x$ , we find

$$\frac{\rho(x)}{\rho(x_0)} = \sqrt{\frac{D_{\text{eff}}(x_0)}{D_{\text{eff}}(x)}} \exp\{\mathcal{U}[v](x_0, x)\}, \quad (43)$$

where we introduced the functional

$$\mathcal{U}[v](x_0, x) \equiv \int_{x_0}^x d\tilde{x} \frac{\tilde{J}(\tilde{x})}{\rho(\tilde{x})D_{\text{eff}}(\tilde{x})}. \quad (44)$$

Equation (43) shows that density ratios are determined by the ratio of the corresponding effective diffusion coefficients, corrected by the exponential of a complicated functional  $\mathcal{U}[v](x_0, x)$  of  $\tilde{J}/(\rho D_{\text{eff}})$  and thus the activity profile  $v(x)$ . It is therefore not suitable for a computation of the full density profile. However, if the integration bounds  $x_0$  and  $x$  in (43) and (44) are sufficiently far away from the activity variations, such that  $\rho(x)$  and  $\rho(x_0)$  correspond to the bulk densities, one can show that the functional vanishes,  $\mathcal{U}[v](x_0, x) = 0$ , irrespective of the activity profile  $v(x)$  (see Appendix A 2). We therefore find that the bulk density ratio is generally given by

$$\frac{\rho(x)}{\rho(x_0)} = \sqrt{\frac{D_{\text{eff}}(x_0)}{D_{\text{eff}}(x)}}. \quad (45)$$

By the definition of the Péclet number in Eq. (20), this result is seen to coincide with Eq. (29). For highly persistent swimmers, i.e.,  $v^2(x) \gg 2DD_r$ , Eq. (45) reduces to the well-known relation [2,54]  $\rho(x)/\rho(x_0) = v(x_0)/v(x)$ . To conclude, the bulk density ratio  $\rho(x)/\rho(x_0)$  is generally independent of the exact shape and magnitude of the activity variations.

## 2. Finite domains

The experiment described in paper I [75] was performed in a finite arena. We therefore next derive the appropriate analytic solutions  $\rho(x)$  and  $p(x)$  for an activity step in a rectangular domain of length  $2L$ , comprising a central active region of length  $2x_{if}$  and two adjacent passive regions interconnected by periodic boundaries, as sketched in Fig. 4. The corresponding activity profile is given by  $v(x) = v_a \Theta(x_{if} - |x|)$ , with the Heaviside step function  $\Theta(x)$ . The symmetry of this setup allows us to consider only the positive half-space with the active-passive interface at  $x = x_{if}$ . The respective polarization and density profiles in the active and passive regions are still given by Eqs. (21) and (22). By virtue of the system's symmetry and the imposed periodic boundary conditions at  $x = \pm L$ , the polarization must obey  $p_a(0) = 0$  and  $p_p(L) = 0$ . Hence, the polarization profiles in the active and passive regions read

$$p_a(x) = C_a \sinh\left(\frac{x}{\lambda_a}\right), \quad p_p(x) = C_p \sinh\left(\frac{L-x}{\lambda_p}\right). \quad (46)$$

Taking  $x_0 = 0$  as a reference point in Eq. (22), the density profile on the active side ( $0 \leq x < x_{if}$ ) reads

$$\rho_a(x) = \rho_a(0) + \frac{C_a v_a \lambda_a}{D} \left[ \cosh\left(\frac{x}{\lambda_a}\right) - 1 \right]. \quad (47)$$

The corresponding density profile on the passive side ( $x_{if} \leq x \leq L$ ) is constant,  $\rho_p(x) \equiv \rho_a(x_{if})$ . The integration constants

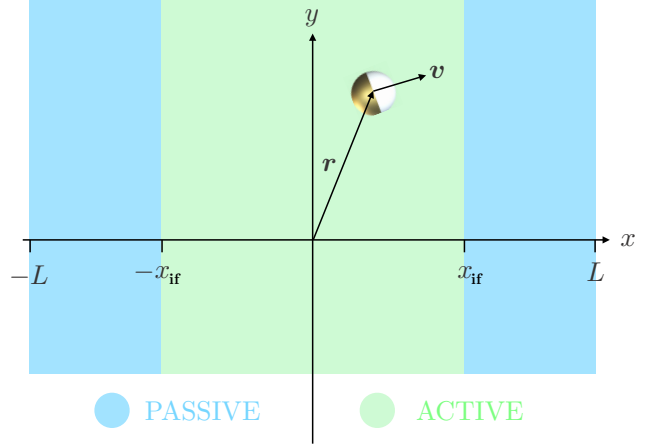


FIG. 4. Sketch of a Janus particle with orientation  $\hat{n} = (\cos \theta, \sin \theta)$  and position  $\mathbf{r} = (x, y)$ . It propels actively with velocity  $\mathbf{v} = v_a \hat{n}$  along its symmetry axis as long as  $|x| < x_{if}$ . Otherwise it undergoes ordinary translational and rotational diffusion. The system has periodic boundaries located at  $\pm L$ .

$C_a$  and  $C_p$  and the reference density  $\rho_a(0)$  are uniquely determined by the continuity conditions

$$p_p(x_{if}) = p_a(x_{if}), \quad (48)$$

$$p_a'(x_{if}) - p_p'(x_{if}) = \frac{v_a}{2D} \rho(x_{if}), \quad (49)$$

and the normalization condition on  $\rho(x)$ . The detailed calculation can be found in Appendix A 3.

Let us assume that the active and passive regions are sufficiently large compared to the decay lengths  $\lambda_{a,p}$  to maintain a scale separation between boundaries and bulk. Then, the polarization and density profiles can be written as

$$p_a(x) = \frac{1}{2L} \frac{P_{\text{max}}}{1 - (1 - r_\rho)^{\frac{x_{if} - \lambda_a}{L}}} \frac{\sinh(x/\lambda_a)}{\sinh(x_{if}/\lambda_a)}, \quad (50)$$

$$p_p(x) = \frac{1}{2L} \frac{P_{\text{max}}}{1 - (1 - r_\rho)^{\frac{x_{if} - \lambda_a}{L}}} \frac{\sinh[(L-x)/\lambda_p]}{\sinh[(L-x_{if})/\lambda_p]}, \quad (51)$$

$$\rho_a(x) = \frac{1}{2L} \left[ \frac{1}{1 - (1 - r_\rho)^{\frac{x_{if} - \lambda_a}{L}}} + \frac{1 - r_\rho}{1 - (1 - r_\rho)^{\frac{x_{if} - \lambda_a}{L}}} \left( \frac{\cosh(x/\lambda_a)}{\sinh(x_{if}/\lambda_a)} - 1 \right) \right], \quad (52)$$

$$\rho_p(x) \equiv \rho_a(x_{if}) = \frac{1}{2L} \frac{1}{1 - (1 - r_\rho)^{\frac{x_{if} - \lambda_a}{L}}}, \quad (53)$$

where we employed the short-hand notation

$$P_{\text{max}} \equiv \frac{v_a}{2D} \frac{\lambda_a \lambda_p}{\lambda_a + \lambda_p} = \frac{1}{\sqrt{2}} \frac{\sqrt{\mathcal{P}}}{1 + \sqrt{1 + \mathcal{P}}}, \quad (54)$$

$$r_\rho \equiv \frac{\lambda_a}{\lambda_p} = \frac{1}{\sqrt{1 + \mathcal{P}}}, \quad (55)$$

for the maximum (relative) polarization and density ratio, respectively. In Fig. 5, we show that these results agree nicely with the exact numerical solutions. In paper I [75], we further show that they also describe well the experimental data.

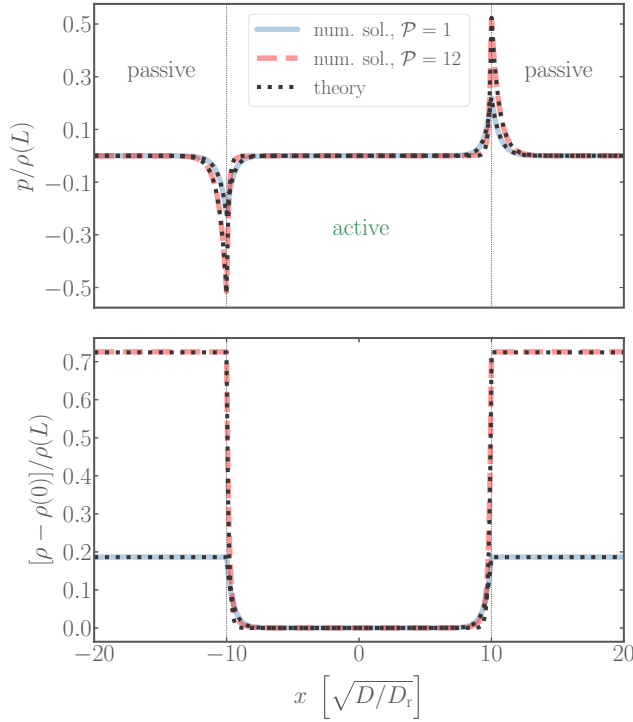


FIG. 5. Polarization-density patterns for a finite active domain bounded at  $x = \pm 10\lambda_p$  by two passive margins extending to  $\pm 20\lambda_p$ . Our approximate theory (50)–(53) for the particle density and polarization (dotted) fares well compared to the exact numerical solutions (dashed and solid lines).

Note that the total polarization of the system in Fig. 5,  $P_{\text{tot}} = \int_{-L}^L dx p(x)$ , vanishes by virtue of the antisymmetry between the polarization profiles on the left and right activity step. This is in agreement with Ref. [70], stating that  $P_{\text{tot}} = 0$  for systems with vanishing fluxes at their boundary. This means that inhomogeneous activity profiles in closed systems merely operate as local spatial sorting mechanisms for particles of different polarization.

### F. Nudging

To mimic the technique of photon nudging [23–25,91] exploited by the experimental setup described in paper I [75] (see also Fig. 1), we need to allow the activity to depend also on the particle orientation. In the nudging regions, the “fuel” for the particle’s autonomous propulsion is restricted not only spatially but also by an acceptance range of particle orientations. We now discuss this additional complication. The angular dependence of the activity profile is modeled as  $v(x, \theta) = v(x)\Theta(\alpha - |\theta|)$ , where  $\alpha$  represents the acceptance angle. Then, the no-flux condition (15) and the moment from Eq. (14) take the modified forms

$$\rho' = \frac{v}{D}(\mathcal{I}_1\rho + \mathcal{I}_2p), \quad (56)$$

$$p'' = \frac{D_r}{D}p + \frac{v}{D}(\mathcal{I}_3\rho' + \mathcal{I}_4p') + \frac{v'}{D}(\rho\mathcal{I}_3 + p\mathcal{I}_4). \quad (57)$$

According to the definitions (11), the constants  $\mathcal{I}_k$ ,  $k = 1, \dots, 4$ , which represent the influence of the restricted acceptance angle  $\alpha$ , read

$$\mathcal{I}_1 = \frac{1}{2\pi} \int_{-\alpha}^{\alpha} d\theta \cos\theta = \frac{\sin\alpha}{\pi}, \quad (58)$$

$$\mathcal{I}_2 = \frac{1}{\pi} \int_{-\alpha}^{\alpha} d\theta \cos^2\theta = \frac{\alpha}{\pi} + \frac{\sin(2\alpha)}{2\pi}, \quad (59)$$

$$\mathcal{I}_3 = \frac{\mathcal{I}_2}{2} = \frac{\alpha}{2\pi} + \frac{\sin(2\alpha)}{4\pi}, \quad (60)$$

$$\mathcal{I}_4 = \frac{1}{\pi} \int_{-\alpha}^{\alpha} d\theta \cos^3\theta = \frac{9\sin\alpha + \sin(3\alpha)}{6\pi}. \quad (61)$$

The swimmer is nudged to the right if  $\alpha < \pi$  and to the left by formally replacing  $v \rightarrow -v$ . The case  $\alpha = \pi$  corresponds to no nudging (orientation-independent activity). We let the sudden activity step again be located at  $x = x_{if} \equiv 0$ , and we assume the particle is nudged (“n”) to the right for  $x \leq 0$ . Upon crossing the interface to  $x > 0$ , it enters a fully active (a) or passive region (p), where its activity does not depend on its orientation. Within each region, the swim speed  $v_i \geq 0$ ,  $i \in \{n, a, p\}$ , is constant.

Plugging the steady-state condition (56) into the moment Eq. (57) yields an equation of the form

$$\mathbf{X}' = \mathbf{\Lambda}\mathbf{X}, \quad (62)$$

where

$$\mathbf{X} \equiv (p', p, \rho)^\top, \quad (63)$$

$$\mathbf{\Lambda} \equiv \begin{pmatrix} \frac{v_i}{D}\mathcal{I}_4 & \frac{D_r}{D} + \frac{v_i^2}{D^2}\mathcal{I}_3\mathcal{I}_2 & \frac{v_i^2}{D^2}\mathcal{I}_1\mathcal{I}_3 \\ 1 & 0 & 0 \\ 0 & \frac{v_i}{D}\mathcal{I}_2 & \frac{v_i}{D}\mathcal{I}_1 \end{pmatrix}. \quad (64)$$

For  $\alpha < \pi$ , all the integrals  $\mathcal{I}_k$  from Eqs. (58)–(61) give nonzero contributions. In Appendix B 1 we explicitly calculate the eigenvalues  $\lambda_{n_i}^{-1}$  of the matrix  $\mathbf{\Lambda}$ . All of them are real and mutually distinct. The general solution to Eq. (62) thus has the structure

$$\mathbf{X} = \sum_{i=1}^3 C_i \mathbf{w}_i e^{\lambda_{n_i}^{-1}x}, \quad (65)$$

where  $\mathbf{w}_i$  denotes the eigenvector pertaining to the eigenvalue  $\lambda_{n_i}^{-1}$ . The coefficients  $C_i$  are determined by boundary and matching conditions. The intuitive relations

$$\rho_n(0) = \rho_a(0), \quad p_n(0) = p_a(0), \quad (66)$$

at a nudging-active interface are complemented by the matching condition

$$p'_a(0) - p'_n(0) = \frac{v_a/2 - v_n\mathcal{I}_3}{D}\rho(0) - \frac{v_n\mathcal{I}_4}{D}p(0). \quad (67)$$

It follows from Eq. (16) while using  $I_{\rho_a}^{(2)} = 1/2$  and  $I_{p_a}^{(2)} = 0$  within the active region [see Eq. (12) and the sentence above]. The matching condition for a nudging-passive interface is included as the case  $v_a = 0$ .

We again require that the polarization vanishes for  $x \rightarrow \pm\infty$ , so that the density attains its constant bulk values  $\rho_n = 0$  and  $\rho_{a,p} > 0$  in the nudging and active/passive bulk, respectively. Hence, inside the nudging region ( $x \leq 0$ ), only positive

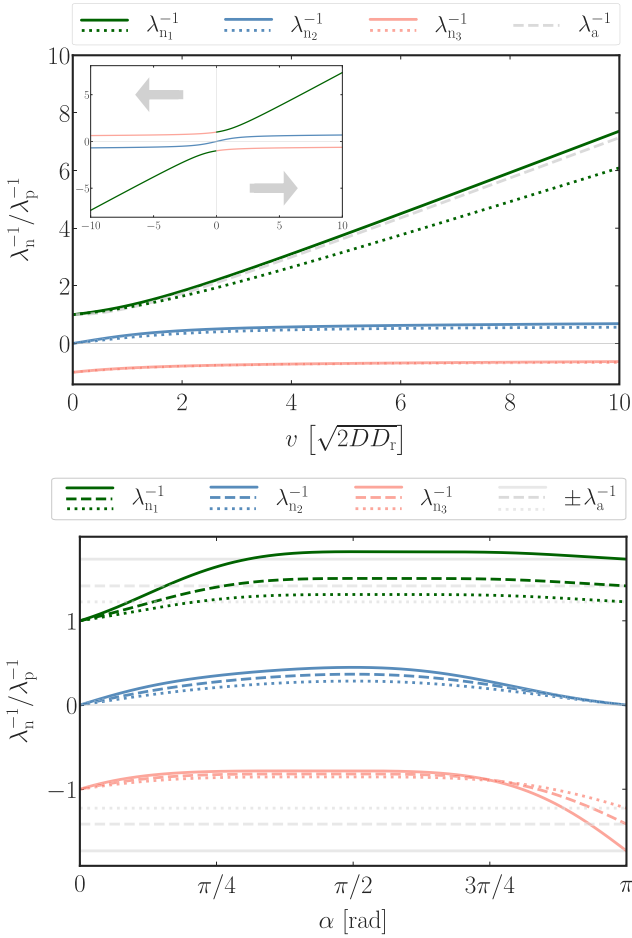


FIG. 6. Dependence of eigenvalues  $\lambda^{-1}$  on particle propulsion speed  $v$  and the nudging acceptance angle  $\alpha$ , with  $\lambda_{n_i}^{-1}$ ,  $i \in \{1, 2, 3\}$ , denoting the eigenvalues of the matrix  $\mathbf{\Lambda}$  defined in Eq. (64). The inverse interfacial layer widths  $\lambda_{a,p}^{-1}$  for active/passive layer widths are those from Eq. (19). See also Fig. 11, which presents a complementary analysis. Upper panel: Eigenvalues were calculated for fixed acceptance angles  $\alpha = 45^\circ$  (dotted curves) and  $\alpha = 90^\circ$  (solid curves). Inset: Dependence of eigenvalues on the nudging direction (indicated by gray arrows). Lower panel: Eigenvalues calculated for the fixed Péclet numbers  $\mathcal{P} \in \{1/2, 1, 2\}$  (dotted, dashed, solid curve).

eigenvalues  $\lambda_{n_i}^{-1} > 0$  will contribute to the general solution (65). Since the analytical expressions for the eigenvalues  $\lambda_{n_i}^{-1}$  given in Appendix B 1 are not very enlightening, we discuss their behavior graphically in Fig. 6. In both panels, the eigenvalues are measured in units of the inverse characteristic length  $\lambda_p^{-1} = \sqrt{D_r/D}$  of a passive boundary layer. From the main plot of the upper panel, one infers  $\lambda_{n_1}^{-1} \geq \lambda_p^{-1}$ ,  $\lambda_{n_2}^{-1} \geq 0$ ,  $\lambda_{n_3}^{-1} < 0$  for all acceptance angles  $0 \leq \alpha \leq \pi$  (lower panel and discussion in Appendix B 3). Therefore, only  $\lambda_{n_1}^{-1}$  and  $\lambda_{n_2}^{-1}$  contribute to the general solution (65). In contrast to purely active-passive interfaces, which are characterized by a single natural length scale  $\lambda_{a,p}$  for each side of the activity step, two characteristic lengths  $\lambda_1$  and  $\lambda_2$  determine the shape of the polarization and density profiles within the nudging layer. Both

$\lambda_{n_1}^{-1}$  and  $\lambda_{n_2}^{-1}$  grow monotonically with increasing propulsion speed  $v$ ,  $\lambda_{n_2}^{-1}$  less than  $\lambda_{n_1}^{-1}$ . While  $\lambda_{n_2}^{-1}$  remains strictly below the (inverse) natural length  $\lambda_a^{-1}$  of a fully active polarization layer,  $\lambda_{n_1}^{-1}$  might even exceed it, depending on swim speed and acceptance angle. For slow swim speeds, i.e., small  $\mathcal{P}$ , one generally has  $\lambda_{n_1}^{-1} > \lambda_a^{-1}$ , as detailed in Appendix B 3.

Purely active or passive polarization layers are captured within this framework (eigenvalues of matrix  $\mathbf{\Lambda}$ ) as well. In the limiting case of vanishing activity ( $v \rightarrow 0$ ), one finds  $\lambda_{n_2}^{-1} \rightarrow 0$ , corresponding to a constant (bulk) density in (65), whereas  $\lambda_{n_{1,3}}^{-1} \rightarrow \pm \lambda_p^{-1}$ . The positive/negative sign refers to a polarization layer in the negative/positive polarization region. Only one of both eigenvalues contributes, whereas the other vanishes for the natural boundary conditions  $p(|x| \rightarrow \infty) = 0$ . Similarly, for a fully active region ( $\alpha \rightarrow \pi$ ), the eigenvalues  $\lambda_{n_{1,3}}^{-1}$  approach  $\pm \lambda_a^{-1}$ , which is indicated by the faint gray lines in the lower panel of Fig. 6. Here, a positive/negative sign also refers to a polarization layer in the negative/positive region. Again only one of them contributes, due to the boundary conditions. The eigenvalue  $\lambda_{n_2}^{-1}$  vanishes, corresponding to a constant (bulk) density in Eq. (65). The inset of the upper panel of Fig. 6 contains information about the behavior of all three eigenvalues upon inverting the direction of the nudging process. We find a completely symmetric picture when replacing  $v \rightarrow -v$ . Only the roles of the eigenvalues change, as we now only allow negative eigenvalues to contribute (particles are nudged to the left if  $x > 0$ ). So  $\lambda_{n_1}^{-1}$  becomes  $\lambda_{n_3}^{-1}$  and thus does not contribute to  $p(x)$  and  $\rho(x)$  anymore, while, in return,  $\lambda_{n_3}^{-1} \rightarrow \lambda_{n_1}^{-1}$ . The eigenvalue  $\lambda_{n_2}^{-1}$  changes sign upon inverting the propulsion direction and thus keeps its role.

Having the eigenvalues  $\lambda_{n_{1,2}}^{-1}$  and the corresponding eigenvectors  $\mathbf{w}_{1,2}$  of matrix  $\mathbf{\Lambda}$  that contribute to the general solution (65), we obtained the polarization and density profiles for nudging-active and nudging-passive interfaces. Polarization and density profiles are matched to those in the respective passive/active regions by the matching conditions (66) and (67). The resulting profiles are depicted in Fig. 7 for two acceptance angles at  $\mathcal{P} = 1$  in both the active and the nudging region. (We note in passing that  $\mathcal{P} = 1$  corresponds to an effectively lower overall activity in the nudging region compared to the active region, due to the restricted acceptance angle.)

All our approximate analytical solutions nicely follow the exact numerical results. An intuitive physical explanation of the density and polarization profiles is provided in Secs. III C and III D.

The lower panels of Fig. 7 depict the relative polarization profiles  $p(x)/\rho(x)$  in both scenarios. Peaking exactly at the interface, they decay over a length scale (inverse eigenvalue)  $\lambda_{n_1} < \lambda_{n_2}$  into the nudging region (left) and approach a constant nonzero value. The (relative) bulk polarization  $(p/\rho)_n$  within the nudging region can be calculated explicitly. We therefor rewrite Eq. (56) as

$$\frac{p(x)}{\rho(x)} = \frac{D}{v(x)\mathcal{I}_2} \frac{\rho'(x)}{\rho(x)} - \frac{\mathcal{I}_1}{\mathcal{I}_2}. \quad (68)$$

Given  $\lambda_{n_1} < \lambda_{n_2}$ , and exploiting that the density profile within the nudging region can be written as a linear combination of  $e^{x/\lambda_{n_1}}$  and  $e^{x/\lambda_{n_2}}$  [see Eq. (65) and boundary conditions],

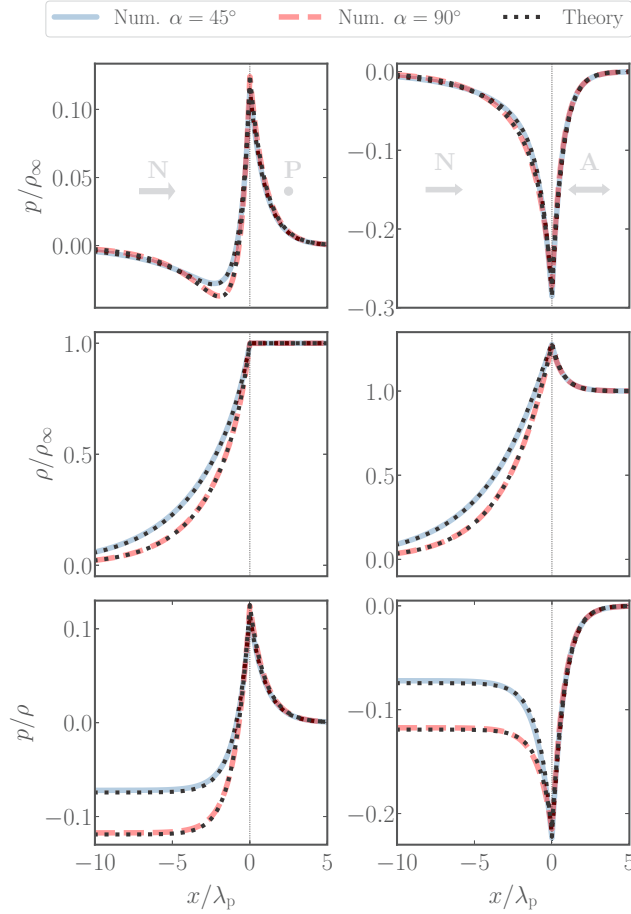


FIG. 7. Particle polarization, density, and relative polarization (from top to bottom) according to Eq. (65) vs the exact numerical solution, for  $\mathcal{P} = 1$ . Left panels: nudging-passive interface. Right panels: nudging-active interface. The coefficients  $C_i$  are determined by the boundary and matching conditions (66) and (67).

one finds  $\rho'/\rho \sim \lambda_{n_2}^{-1}$  for  $|x|$  sufficiently greater than  $\lambda_{n_1}$ . The relative polarization in the nudging bulk is thus given by

$$\left(\frac{p}{\rho}\right)_n = \frac{D}{v(x)\lambda_{n_2}\mathcal{I}_2} - \frac{\mathcal{I}_1}{\mathcal{I}_2}. \quad (69)$$

For a nudging layer in the region with  $x < 0$ , one finds that  $(p/\rho)_n \leq 0$  and that it is minimal for  $\alpha = \pi/2$ , as can be inferred from the upper panel of Fig. 8. Also, the higher the swimmer's activity  $\mathcal{P}$ , the deeper the minimum in the bulk polarization. The middle panel of Fig. 8 shows the agreement between the approximate theoretical prediction (69) and exact numerical solutions up to Péclet numbers  $\mathcal{P}$  on the order of 50. We attribute the small but systematically growing mismatch between both to the breakdown of the one-to-one mapping between our approximate analytical theory and the exact description of a simple two-state caricature of our physical setup, in the case of a nudging interface (see Sec. III A).

As explicitly checked above and proven to hold generally [70], the total polarization  $P_{\text{tot}} = \int_{-\infty}^{\infty} dx p(x)$  is determined by the difference of the respective bulk fluxes  $v_i \rho_i$  divided by  $2D_i$ ; cf. Eq. (30). The bulk fluxes associated with the

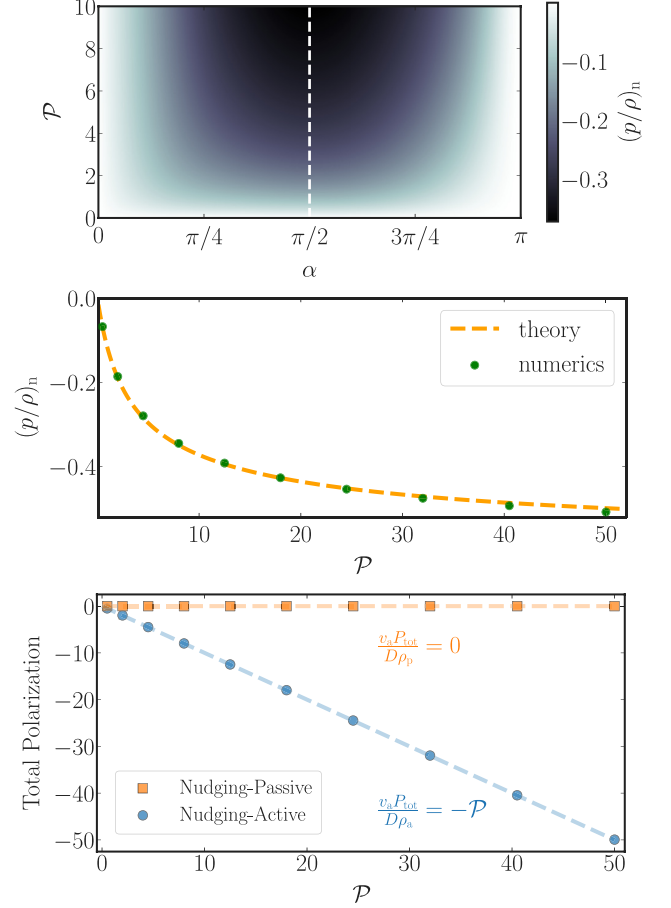


FIG. 8. Relative bulk polarization  $(p/\rho)_n$  within the nudging region and total interface polarization. Upper panel: Heat map of  $(p/\rho)_n$  as a function of the Péclet number  $\mathcal{P}$  and acceptance angle  $\alpha$  according to Eq. (69). The vertical dashed line depicts the maximum of  $|(p/\rho)_n|$  with respect to the acceptance angle  $\alpha$ . Middle panel: approximate theory (69) (dashed line) vs exact numerical solution (circles) for  $\alpha = \pi/2$ . Lower panel: Reduced total interface polarization from Eq. (30), calculated via numerical integration of the exact numerical polarization profiles (symbols) for various Péclet numbers compared to theory (dashed lines).

nudging and the passive region are both zero. The former vanishes because of  $\rho_n = 0$  in the bulk, the latter because of  $v = 0$ . Thus, for a nudging-passive interface,  $P_{\text{tot}} = 0$ . Vanishing bulk fluxes therefore explain the change of sign in the polarization profile for the nudging-passive interface (see the upper left panel of Fig. 7). For the nudging-active interface, one finds  $P_{\text{tot}} = -\rho_a v_a / (2D_r)$ , which coincides with the total polarization at a passive-active interface; cf. Eq. (30). We verified these results for  $P_{\text{tot}}$  at the considered nudging interfaces analytically and numerically, as shown in the lower panel of Fig. 8.

Note that only the total polarization is determined by bulk quantities, but not the local polarization profile  $p(x)$  itself. A thorough discussion and physical interpretation of the (active, passive, nudging) polarization layers is provided in the following section.

### III. INTUITIVE PHYSICAL INTERPRETATION

#### A. The two-species run-and-tumble model

The emerging polarization and inhomogeneous density distribution in the vicinity of an activity step are easily intuitively understood within a simple one-dimensional two-species run-and-tumble model [2,67,97,99,108]. It constrains the swimmer to either orient parallel or antiparallel to the  $x$ -axis, while it may randomly flip its orientation (“tumble”) at a rate  $k$  (corresponding to a dichotomous Markov process). During the “run phases,” besides its thermal diffusion, the particle propels actively with a position-dependent swim speed  $v_{\pm}(x)$ , which might also depend on the orientation ( $\pm$ ) of the particle in order to mimic the nudging. We define the probability densities  $n_{+}(x, t)$  and  $n_{-}(x, t)$  for encountering the particle at time  $t$  at position  $x$  with orientation parallel (+) or antiparallel (−) to the  $x$ -axis. The corresponding fluxes  $J_{\pm}(x, t)$  contain contributions both from thermal agitation and from active propulsion, and they are given by

$$J_{+}(x, t) = -Dn'_{+}(x, t) + v_{+}(x)n_{+}(x, t), \quad (70)$$

$$J_{-}(x, t) = -Dn'_{-}(x, t) + v_{-}(x)n_{-}(x, t). \quad (71)$$

For the (total) density  $\rho \equiv n_{+} + n_{-}$  and the polarization  $p \equiv n_{+} - n_{-}$ , we keep the notation of the continuous-angle model. In the steady state, the total flux  $J_{+} + J_{-}$  vanishes, which yields the balance condition

$$D\rho' = \frac{v_{+} + v_{-}}{2}\rho + \frac{v_{+} - v_{-}}{2}p. \quad (72)$$

The time evolution of the densities is given by

$$\begin{aligned} \dot{n}_{+} &= -J'_{+} - k(n_{+} - n_{-}) \\ &= Dn''_{+} - (v_{+}n_{+})' - k(n_{+} - n_{-}), \end{aligned} \quad (73)$$

$$\begin{aligned} \dot{n}_{-} &= -J'_{-} + k(n_{+} - n_{-}) \\ &= Dn''_{-} - (v_{-}n_{-})' + k(n_{+} - n_{-}). \end{aligned} \quad (74)$$

In the steady state ( $\dot{n}_{\pm} = 0$ ), subtracting Eq. (74) from Eq. (73) yields

$$p'' = \frac{2k}{D}p + \left( \frac{v_{+} - v_{-}}{2D}\rho + \frac{v_{+} + v_{-}}{2D}p \right)'. \quad (75)$$

##### 1. Fully active/passive

The case of a symmetrically active/passive particle (regardless of the orientation) is captured by setting  $v_{+} = -v_{-} \equiv v$ . Equations (72) and (75) then reduce to

$$\rho' = \frac{v}{D}p, \quad p'' = \frac{p}{\lambda^2} + \frac{\rho}{D}v', \quad (76)$$

with the characteristic length scale

$$\lambda(x) \equiv \left[ \frac{2k}{D} + \frac{v^2(x)}{D^2} \right]^{-1/2}. \quad (77)$$

The above equations for  $\rho$ ,  $p$ , and  $\lambda$  are structurally equal to Eqs. (17), (18), and (19) in the (approximate) model for an active particle that can rotate continuously in the plane. Upon mapping  $2k \rightarrow D_r$ ,  $v \rightarrow v/\sqrt{2}$ , and  $\rho \rightarrow \rho/\sqrt{2}$ , the two models become equivalent. One therefore applies the

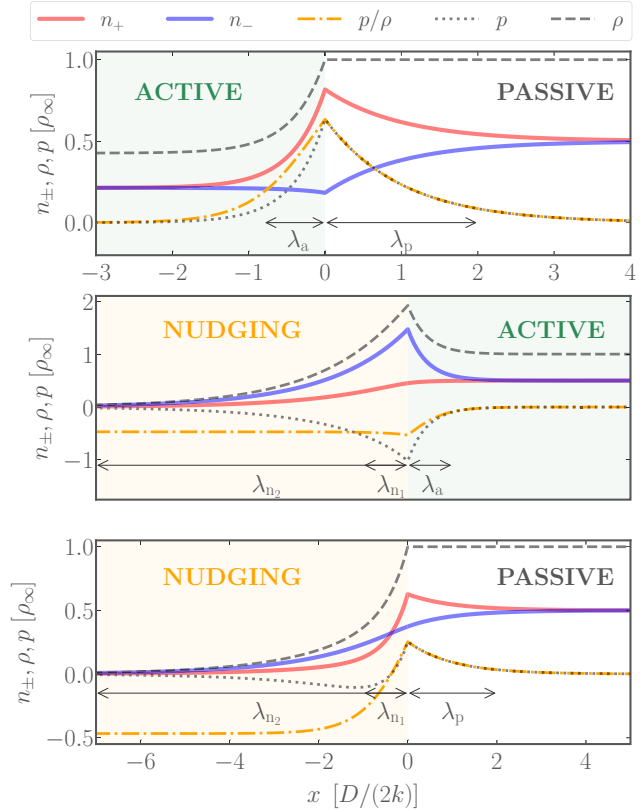


FIG. 9. Exact analytical solution of the two-species model: polarization and density profiles,  $p(x)$  and  $\rho(x)$ , their ratio  $p/\rho$ , and the species concentrations  $n_{\pm}(x) = [\rho(x) \pm p(x)]/2$  of right- and left-oriented particles around an active-passive, nudging-active, and nudging-passive interface located at  $x = 0$ . All quantities are normalized by the respective constant bulk density  $\rho_{\infty} = \rho(x \rightarrow \infty)$ . The natural widths of the interfacial polarization layers are denoted by  $\lambda_a$ ,  $\lambda_p$ , and  $\lambda_{n1,2}$  in the active, passive, and nudging regions, respectively.

same methods as above to obtain the analytical solutions for  $p(x)$  and  $\rho(x)$ . Both functions as well as the species densities  $n_{\pm}(x) = [\rho(x) \pm p(x)]/2$  and the relative polarization  $p/\rho$  are plotted in the upper panel of Fig. 9. All quantities are normalized by the bulk density  $\rho_{\infty} = \rho(x \rightarrow \infty)$  in the passive region.

##### 2. Nudging

Within the framework of the two-species model, a nudging process is modeled by setting  $v_{+} \equiv v$  and  $v_{-} \equiv 0$  for nudging to the right and *vice versa* for nudging to the left. The flux-balance condition (72) and Eq. (75) for the polarization then become

$$\rho' = \frac{v}{D}\rho + \frac{v}{D}p, \quad (78)$$

$$p'' = \frac{2k}{D}p + \frac{v}{D}\left(\frac{\rho'}{2} + \frac{p'}{2}\right) + \frac{v'}{D}\left(\frac{\rho}{2} + \frac{p}{2}\right). \quad (79)$$

These equations also have the same structure as their counterparts (56) and (57) for the continuous-rotation model. Thus, the methods used in Sec. II F can be applied in order to deter-

mine the polarization and density profiles. Note, however, that here the two-species and the continuous-angle model cannot be mapped onto each other. There exists no unique acceptance angle  $\alpha$  to ensure  $v/2 = v_a \mathcal{I}_k$  for all coefficients  $\mathcal{I}_k$ ,  $k = 1, \dots, 4$ , defined in (58)–(61), where  $v_a$  denotes the Janus swimmer’s propulsion speed in the continuous-angle model. Nevertheless, at least qualitatively, polarization and density profiles in the two-species model display the same features as their counterparts in the continuous-angle model, as can be inferred from the second and third panel of Fig. 9 (cf. Fig. 7). A more thorough discussion of the nudging process within the framework of the two-state model, and a comparison to the continuous model, can be found in Appendixes B 2 and B 3.

### B. Active-passive interface

Via the mutual mappings discussed in the previous paragraph, the two-species model supplies a straightforward intuitive interpretation of the results derived from the continuous-angle model. For example, to gain an intuitive understanding of the emerging polarization layer at an active-passive interface, focus on the upper panel of Fig. 9. For the sake of brevity, we denote particles pointing to the left and right by  $L$  and  $R$ , respectively. We first consider the situation on the active side, where the particle motion might be regarded as quasiballistic. While  $R$ -particles get stuck at the interface due to the ceasing propulsion,  $L$ -particles quickly “escape” the interfacial region. This sorting mechanism leads to a majority of  $R$ -particles at the interface and thereby reveals the otherwise hidden bulk currents [70]. We now qualitatively explain the shape of two polarization layers. In the passive region, the characteristic decay length  $\lambda_p = \sqrt{D/(2k)}$  of the polarization is obtained by setting  $v(x) = 0$  in Eq. (77). It is intuitively understood since the particle’s motion on the passive side is ordinary diffusion with diffusivity  $D$ . The spreading of the excess polarization into the passive region is limited by the characteristic flipping time  $(2k)^{-1}$ . Therefore,  $\lambda_p$  basically coincides with the mean-squared displacement of a passive particle during this time. On the active side, a kind of “sedimentation pressure” joins the game. Its cause is an “active swim force”  $\zeta v$  (with friction  $\zeta$ ) directed toward the interface for the  $R$ -particles and away from it for  $L$ -particles. One thus might regard  $R$ -particles as “heavy” and  $L$ -particles as “buoyant.” The effective sedimentation process therefore compresses the extension  $\lambda_a$  of the polarization layer according to Eq. (77) on the active side, which, for highly persistent motion [ $v^2/(2kD) \gg 1$ ], is described by the barometer formula  $\exp(-v\zeta x/\zeta D)$ . The swim force takes the role of gravity and  $\zeta D$  of thermal energy  $k_B T$ , according to the Sutherland-Einstein relation. In the same limit follows the motility-induced density suppression  $\rho_a/\rho_p = \lambda_p/\lambda_a \propto v^{-1}$ , which is a well-known result for run-and-tumble particles [2,54].

### C. Nudging-active interface

The middle panel of Fig. 9 depicts  $p(x)$ ,  $\rho(x)$ , their ratio, and the species concentrations  $n_{\pm}(x)$  in the vicinity of a nudging-active interface. While the propulsion speed is the same in both regions, only the  $R$ -particles propel actively inside the nudging region. The  $R$ -species, therefore, does not

display a sudden change (kink) in its species concentration  $n_+(x)$  upon crossing the interface, as no abrupt activity drop is experienced. In the active region,  $R$ -particles quickly “escape” the interfacial area into the active region, whereas  $L$ -particles get stuck at the interface and venture only diffusively into the nudging region. We thus observe an excess of  $L$ -particles at the interface, and therefore a negative polarization. The extent  $\lambda_a$  of the polarization layer on the active side is again determined by the interplay between Brownian motion and the effective sedimentation pressure, this time with “heavy”  $L$ - and “buoyant”  $R$ -particles. The symmetry between “heavy” and “buoyant” particle species is broken in the nudging region. There,  $R$ -particles are “heavy” and therefore nudged toward the interface, whereas the  $L$ -species are passive Brownian particles. Close to the interface, the decay of density and polarization into the nudging region is therefore characterized by a length scale  $\lambda_{n_1} \neq \lambda_a$ . Referring back to the upper panel of Fig. 6, we infer that  $\lambda_{n_1}$  is still quite similar to the characteristic decay length  $\lambda_a$  pertaining to a purely active region. Notice that, over a distance  $\lambda_{n_1}$ , the relative polarization  $p/\rho$  approaches a constant value given by Eq. (69), which is a distinctive feature of the bulk in the nudging region, characterized by the swim speed and the acceptance angle of the nudging procedure. In contrast to purely active or passive regions, both the particle polarization  $p(x \rightarrow \infty)$  and the bulk density  $\rho(x \rightarrow \infty)$  in the nudging region decay to zero, since every particle is inevitably nudged toward the interface until it crosses it. The decay of the absolute polarization and density profiles toward zero is described by the second characteristic length scale  $\lambda_{n_2} > \lambda_{n_1}$ .

### D. Nudging-passive interface

Finally, consider the polarization and density profiles and the respective species concentrations presented in the last panel of Fig. 9. Now, only  $R$ -particles propel actively inside the nudging region, while both particle species behave like ordinary Brownian particles in the passive region. Therefore, the  $L$ -particles can cross the interface smoothly, and their density  $n_-(x)$  does not display a kink. We observe an excess of  $R$ -particles at the interface by virtue of the inherently biased nudging process, and thus a positive polarization. The width  $\lambda_p$  of the polarization layer on the passive side is determined by the distance  $\sqrt{D/(2k)}$  covered by thermal diffusion during the characteristic time scale  $(2k)^{-1}$  for “tumbling.” The imbalance between the “heavy”  $R$ - and neutral  $L$ -particles determines the spreading of the polarization into the nudging region. Due to the “removal” of  $R$ -particles toward the interface by virtue of the nudging procedure, the polarization even changes sign and becomes negative over the characteristic length scale  $\lambda_{n_1}$  before it converges to zero over the length scale  $\lambda_{n_2}$  as discussed in the previous scenario. This distinctive shape of the polarization can be seen as indicative of the hidden bulk currents that are generally understood to cause the interfacial polarization layers [70], as already discussed in Sec. III F.

## IV. CONCLUSION

In this article, we have studied the behavior of a single Janus-type swimmer in the vicinity of a motility step. Within

an approximate ABP model, we derived analytical expressions for the polarization and density profiles of a Janus particle at planar activity steps. We showed that they agree well with exact numerical solutions and experimental data (see paper I [75]). Key features of polarization and density profiles at motility steps were discussed and shown to exhibit important similarities to those observed for MIPS. As a consistency check, we also explicitly demonstrated both analytically and numerically that the total polarization induced by the motility step is determined by the difference of (hidden) bulk fluxes, and it obeys an exact global sum rule. We further showed that the bulk density ratio between two regions of distinct but constant activity is determined by the ratio of the respective effective diffusion coefficients and independent of the shape of the activity profile that mediates between the bulk regions. Motivated by the versatile experimental technique of photon nudging, we moreover generalized our theoretical results to the situation of orientation-dependent propulsion speeds. We conclude that the colocalization of polarization and density patterns in activity gradients, as they naturally occur at various interfaces, is a characteristic phenomenological trait to robustly distinguish motile-particle suspensions from thermal and athermal passive suspensions.

## ACKNOWLEDGMENTS

We thank Paul Cervenak and Anton Stall for discussions and contributions during the early stages of this work. We acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG) through the priority program ‘‘Microswimmers’’ (SPP 1726, Project No. 237143019) and Project No. 432421051 (DFG-GACR cooperation) as well as the by Czech Science Foundation (Project No. 20-02955J). V.H. gratefully acknowledges support by the Humboldt Foundation. N.A.S. acknowledges support by the European Social Fund (ESF), the European Union (EU) and the Free State of Saxony (ESF Project No. 100327895).

## APPENDIX A: PLANAR ACTIVITY STEP

### 1. Density ratio and total polarization

#### a. Density ratio

Introducing the auxiliary quantities  $\beta_{A,a} \equiv p_{\max} v_{A,a} \lambda_{A,a} / [D\rho(0)]$ , Eqs. (26) and (27) evaluated at  $x = 0$  and  $x \rightarrow \infty$ , respectively, yield

$$\rho(0) - \rho_A = \rho(0)\beta_A, \quad (\text{A1})$$

$$\rho_a - \rho_A = \rho(0)(\beta_a + \beta_A). \quad (\text{A2})$$

Using these equations, the density ratio can be expressed as  $\rho_a/\rho_A = (1 + \beta_a)/(1 - \beta_A)$ . Further using

$$\lambda_{A,a} = \left( \frac{D_r}{D} + \frac{v_{A,a}^2}{2D^2} \right)^{-1/2}, \quad (\text{A3})$$

$$\frac{p_{\max}}{\rho(0)} = \frac{v_A - v_a}{2D} \frac{\lambda_A \lambda_a}{\lambda_A + \lambda_a}, \quad (\text{A4})$$

from Eq. (28), implies  $\rho_a/\rho_A = \lambda_a/\lambda_A$ , as given in Eq. (29).

#### b. Total polarization

From Eq. (30) we know that the total polarization  $P_{\text{tot}}$  is given by  $p_{\max}(\lambda_A + \lambda_a)$ . The coefficient  $p_{\max}$  can be expressed as  $D(\rho_a - \rho_A)/(v_A \lambda_A + v_a \lambda_a)$  by virtue of Eq. (A2). Hence, the total polarization reads

$$P_{\text{tot}} = D(\rho_a - \rho_A) \frac{\lambda_a + \lambda_A}{v_A \lambda_A + v_a \lambda_a}. \quad (\text{A5})$$

Using  $\rho_a/\rho_A = \lambda_a/\lambda_A$ , we find that

$$P_{\text{tot}} = D \frac{(\rho_a - \rho_A)(\rho_a + \rho_A)}{v_A \rho_A + v_a \rho_a}, \quad (\text{A6})$$

$$= D \frac{\rho_a^2 - \rho_A^2}{v_a^2 \rho_a^2 - v_A^2 \rho_A^2} (v_a \rho_a - v_A \rho_A). \quad (\text{A7})$$

Using the formula  $\rho_a/\rho_A = \lambda_a/\lambda_A$  and the definition (A3) of  $\lambda_{A,a}$ , the factor in front of the term in parentheses turns out to be equal to  $-1/(2D_r)$ . One thus obtains  $P_{\text{tot}} = (v_A \rho_A - v_a \rho_a)/(2D_r)$ , as stated in Eq. (30).

## 2. Vanishing integral

Our intermediate result (43) for the density ratio,

$$\frac{\rho(x)}{\rho(x_0)} = \sqrt{\frac{D_{\text{eff}}(x_0)}{D_{\text{eff}}(x)}} \exp\{\mathcal{U}[v](x_0, x)\}, \quad (\text{A8})$$

depends on the functional  $\mathcal{U}$  defined in Eq. (44). In the following, we prove that  $\mathcal{U}$  vanishes if the densities  $\rho(x_0)$  and  $\rho(x)$  correspond to the constant bulk densities outside the interface region with changing activity.

First, we consider the two-species model of Sec. III A with an arbitrary activity profile  $v(x)$  whose inhomogeneities are localized around a finite region beyond which the velocity assumes a single constant value; see Fig. 10(a). We will refer to the corresponding constant bulk densities as  $\rho_i$ ,  $i = 1, 2$ . In the bulk, the polarization vanishes and thus the concentrations  $n_{\pm}^{(i)} = \rho_i/2$  of the individual species coincide. The corresponding fluxes between the two bulk regions can be expressed as

$$J_{12} = \frac{\rho_1}{2} (v_{12}^+ + v_{12}^-), \quad (\text{A9})$$

$$J_{21} = \frac{\rho_2}{2} (v_{21}^+ + v_{21}^-), \quad (\text{A10})$$

where we introduced the transition rates  $v_{ij}^{\pm}$  of particle species ( $\pm$ ) between the bulk region  $i$  and  $j$ . In the steady state, the (anti)symmetry between the two particle species implies  $v_{12}^+ = v_{21}^-$  and  $v_{12}^- = v_{21}^+$ . Flux balance  $J_{12} = J_{21}$  then induces  $\rho_1 = \rho_2$ . For an arbitrary activity profile  $v(x)$  with bulk activities given by a single constant, Eq. (A8) thus implies that  $1 = \rho_1/\rho_2 = e^{\mathcal{U}}$  and hence  $\mathcal{U} = 0$ .

Now, we consider activity profiles of the form sketched in Fig. 10(b). Two bulk regions with constant activity  $v_1$  are interconnected via an intermediate bulk region with a constant activity  $v_2 \neq v_1$  by two arbitrary activity profiles,  $v_{12}(x)$  and  $v_{23}(x)$ . From the first part of the proof, we know that the density ratio  $\rho_1/\rho_3 = 1$  as the activity in the outer bulk regions is equal and constant. We thus know that  $0 = \mathcal{U}[v_{12}, v_{23}] = \mathcal{U}_{12}[v_{12}] + \mathcal{U}_{23}[v_{23}]$ , where we introduced the integrals  $\mathcal{U}_{12}$ ,  $\mathcal{U}_{23}$  pertaining to the two parts  $v_{12}(x)$  and  $v_{23}(x)$

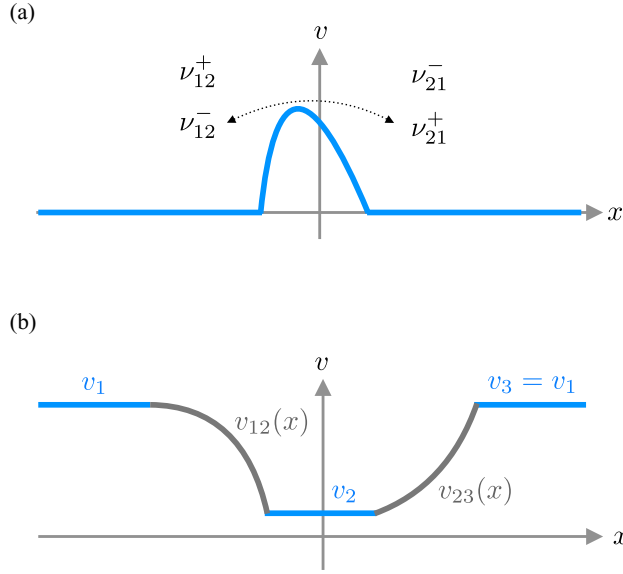


FIG. 10. Activity profiles  $v(x)$  considered to demonstrate that the integral  $\mathcal{U}$  vanishes. In the upper plot,  $v_{ij}^\pm$  denote the transition rates of particles species ( $\pm$ ) from the left to the right bulk region and vice versa. In the lower plot,  $v_{1,2}$  denote the constant activities in the middle and the outer regions, respectively, which are connected by the profiles  $v_{12}(x)$  and  $v_{23}(x)$ .

of the activity profile. Since the total integral has to vanish for any intermediate set of activity gradients, the integrals  $\mathcal{U}_{12}$  and  $\mathcal{U}_{23}$  can only depend on the bulk values of the activity, not on the details of the intermediate activity profiles. This means that the same  $\mathcal{U}$  would be obtained for the sharp activity step as for any other profile mediating between the two bulk activities. But for the sharp step, our discussion in Sec. II C applies, showing that  $\mathcal{U} = 0$ . Therefore, for the two-state model, the integral  $\mathcal{U}$  evaluated between two bulk positions  $x_1$  and  $x_2$  always vanishes and  $\rho_1/\rho_2 = \sqrt{D_{\text{eff}}(x_2)/D_{\text{eff}}(x_1)}$ .

In the two-species model, the particle orientation hops between the two orientations with a transition rate. In the continuous model, the orientation diffuses in a continuum of orientations. However, this difference does not break the symmetry of the model, which is the key for our proof. For each particle orientation there is still an antioriented particle with the same absolute value of projection of the velocity on the  $x$ -axis and the opposite velocity. Realizing this fact, the proof for the two-species model applies also to the continuum model.

The generalization can be formalized by introducing a  $(2N)$ -species model,  $N > 1$ . We now assume that the fluxes (A9) and (A10) correspond to the auxiliary process of left and right jumping particles, which have  $2N$  contributions. We exploit that in the steady state, each particle species has an equivalent “antispecies” with an equal transition rate to jump

in the opposite direction. As for the two-species model, flux balance then again induces  $\rho_1 = \rho_2$ , and thus  $\mathcal{U} = 0$ . The generalization to the continuous situation is done by taking the limit  $N \rightarrow \infty$  and introducing effective transition rates

$$v_{12}^{\text{eff}} \equiv \lim_{N \rightarrow \infty} \frac{1}{2N} \sum_{k=0}^{2N} v_{12}^{(k)}, \quad (\text{A11})$$

$$v_{21}^{\text{eff}} \equiv \lim_{N \rightarrow \infty} \frac{1}{2N} \sum_{k=0}^{2N} v_{21}^{(k)}, \quad (\text{A12})$$

where the index  $k$  runs over all particle species. As both effective transition rates are equal in the steady state, we can proceed with the proof as in the previously considered discrete scenarios.

This proves also for the continuous model that the functional  $\mathcal{U}$  depends on the bulk values of the activity profiles only. Despite its impressive accuracy, our analytical theory for this case is only an approximation and thus is not able to guarantee that  $\mathcal{U} = 0$  holds exactly for the activity step, but supportive numerical evidence is provided in Fig. 3.

### 3. Finite system—Determining the coefficients

The general solutions (46) and (47) to the particle’s density distribution  $\rho_{a/p}(x)$  and polarization  $p_{a/p}(x)$  within the active ( $0 \leq x \leq x_{if}$ ) and passive ( $a < x \leq L$ ) region, respectively, read

$$p_a(x) = C_a \sinh\left(\frac{x}{\lambda_a}\right), \quad \lambda_a = \left(\frac{D_r}{D} + \frac{v_a^2}{2D^2}\right)^{-1/2}, \quad (\text{A13})$$

$$p_p(x) = C_p \sinh\left(\frac{L-x}{\lambda_p}\right), \quad \lambda_p = \sqrt{\frac{D}{D_r}}, \quad (\text{A14})$$

$$\rho_a(x) = \rho_a(0) + \frac{C_a \lambda_a v_a}{D} \left[ \cosh\left(\frac{x}{\lambda_a}\right) - 1 \right], \quad (\text{A15})$$

$$\rho_p(x) = \rho_a(x_{if}) = \text{const.} \quad (\text{A16})$$

The emerging integration constants  $C_a$  and  $C_p$  are fixed by the matching conditions

$$p_p(x_{if}) = p_a(x_{if}), \quad (\text{A17})$$

$$p_a'(x_{if}) - p_p'(x_{if}) = \frac{v_a}{2D} \rho(x_{if}), \quad (\text{A18})$$

which follow from Eq. (16). The first one allows us to express one integration constant in terms of the other:

$$C_p = C_a \frac{\sinh[x_{if}/\lambda_a]}{\sinh[(L-x_{if})/\lambda_p]}. \quad (\text{A19})$$

Before we employ the condition (A18), we exploit the normalization condition  $\int_{-L}^L dx \rho(x) = 1$  to calculate the density at the interface  $\rho(x_{if})$ . Respecting the symmetry of the considered problem, we get

$$\begin{aligned} \frac{1}{2} &= \int_0^L dx \rho(x) = \int_0^{x_{if}} dx \rho_a(x) + \int_{x_{if}}^L dx \rho_p(x) = \int_0^{x_{if}} dx \rho_a(x) + (L-x_{if})\rho_a(x_{if}) \\ &= x_{if} \rho_a(0) + \frac{C_a \lambda_a v_a}{D} \left[ \lambda_a \sinh\left(\frac{x_{if}}{\lambda_a}\right) - x_{if} \right] + (L-x_{if}) \left\{ \rho_a(0) + \frac{C_a \lambda_a v_a}{D} \left[ \cosh\left(\frac{x_{if}}{\lambda_a}\right) - 1 \right] \right\}. \end{aligned} \quad (\text{A20})$$



Solving this equation for  $\rho_a(0)$ , we find

$$\rho_a(0) = \frac{1}{2L} - \left(1 - \frac{x_{if}}{L}\right) \frac{C_a \lambda_a v_a}{D} \left[ \cosh\left(\frac{x_{if}}{\lambda_a}\right) - 1 \right] - \frac{x_{if}}{L} \frac{C_a \lambda_a v_a}{D} \left[ \frac{\lambda_a}{x_{if}} \sinh\left(\frac{x_{if}}{\lambda_a}\right) - 1 \right]. \quad (\text{A21})$$

Substituting this relation into Eq. (A15) yields

$$\rho_a(x_{if}) = \frac{1}{2L} + \frac{x_{if}}{L} \frac{C_a \lambda_a v_a}{D} \left[ \cosh\left(\frac{x_{if}}{\lambda_a}\right) - \frac{\lambda_a}{x_{if}} \sinh\left(\frac{x_{if}}{\lambda_a}\right) \right]. \quad (\text{A22})$$

Finally, plugging this result into Eq. (A18) and using Eq. (A19) for  $C_p$  renders a linear equation for  $C_a$ , which is straightforwardly solved. With the definitions (54) and (55),

$$P_{\max} \equiv \frac{v_a}{2D} \frac{\lambda_a \lambda_p}{\lambda_a + \lambda_p}, \quad r_\rho \equiv \frac{\lambda_a}{\lambda_p}, \quad (\text{A23})$$

one finds that

$$C_a = \frac{1}{2L \sinh(x_{if}/\lambda_a) \coth\left(\frac{L-x_{if}}{\lambda_p}\right) - \frac{1-r_\rho}{L} [x_{if} \coth\left(\frac{L-x_{if}}{\lambda_p}\right) - \lambda_a]} P_{\max}. \quad (\text{A24})$$

The approximation  $\coth[(L-x_{if})/\lambda_p] \approx 1$  in the above equations then leads to the polarization and density profiles (50)–(53) given in the main text.

## APPENDIX B: NUDGING LAYER

### 1. Continuous-angle model

As derived in Sec. II F, within a nudging region, the vector  $\mathbf{X}(x) = [p'(x), p(x), \rho(x)]^\top$  composed of the polarization profile  $p(x)$ , its derivative, and the density  $\rho(x)$  obeys an equation of the form  $\mathbf{X}' = \mathbf{\Lambda} \mathbf{X}$ . In contrast to the discussion in the main text, we now resort to a dimensionless description by introducing  $\lambda_p = \sqrt{D/D_r}$  as a natural unit of length. The matrix  $\mathbf{\Lambda}$  can then be expressed in terms of a single parameter—the Péclet number  $\mathcal{P} = v_a^2/(2DD_r)$ —and it reads

$$\mathbf{\Lambda} = \begin{pmatrix} \sqrt{2\mathcal{P}}\mathcal{I}_4 & 1 + 2\mathcal{P}\mathcal{I}_3\mathcal{I}_2 & 2\mathcal{P}\mathcal{I}_1\mathcal{I}_3 \\ 1 & 0 & 0 \\ 0 & \sqrt{2\mathcal{P}}\mathcal{I}_2 & \sqrt{2\mathcal{P}}\mathcal{I}_1 \end{pmatrix}. \quad (\text{B1})$$

The quantities  $\mathcal{I}_k$ ,  $k = 1, \dots, 4$ , defined in Eqs. (58)–(61), characterize the influence of the restricted acceptance angle  $\alpha$  on the heating laser. Clearly, the eigenvalues  $\lambda_{n_i}^{-1}$  of the matrix  $\mathbf{\Lambda}$  determine the general solution  $\mathbf{X}(x)$ . The characteristic equation  $|\mathbf{\Lambda} - \lambda^{-1}\mathbf{1}| = 0$  renders the cubic equation  $\lambda^{-3} + a\lambda^{-2} + b\lambda^{-1} + c = 0$ , with

$$a \equiv -(\mathcal{I}_1 + \mathcal{I}_4)\sqrt{2\mathcal{P}}, \quad (\text{B2})$$

$$b \equiv 2(\mathcal{I}_1\mathcal{I}_4 - 2[\mathcal{I}_3]^2)\mathcal{P} - 1, \quad (\text{B3})$$

$$c \equiv \mathcal{I}_1\sqrt{2\mathcal{P}}. \quad (\text{B4})$$

Using the Tschirnhaus-Vieta approach to the solution of cubic equations, one finds that all three roots are real-valued and can be written in the form

$$\lambda_{n_1}^{-1} = -\frac{a}{3} + 2\sqrt{-q} \cos\left(\frac{\gamma}{3}\right), \quad (\text{B5})$$

$$\lambda_{n_2}^{-1} = -\frac{a}{3} + 2\sqrt{-q} \cos\left(\frac{\gamma}{3} + \frac{4\pi}{3}\right), \quad (\text{B6})$$

$$\lambda_{n_3}^{-1} = -\frac{a}{3} + 2\sqrt{-q} \cos\left(\frac{\gamma}{3} + \frac{2\pi}{3}\right), \quad (\text{B7})$$

where we introduced the auxiliary quantities

$$q \equiv \frac{3b - a^2}{9}, \quad (\text{B8})$$

$$r \equiv \frac{9ab - 27c - 2a^3}{54}, \quad (\text{B9})$$

$$\gamma \equiv \arccos\left(\frac{r}{\sqrt{-q}}\right). \quad (\text{B10})$$

The dependence of the eigenvalues  $\lambda_{n_i}^{-1}$  on the particle's propulsion speed (or Péclet number) and the acceptance angle  $\alpha$  are graphically discussed in Sec. II F of the main text as the analytical expressions above lack an immediate physical insight.

### 2. Two-species model

As derived in Sec. III A, the governing equations for the polarization and density profiles within the framework of the two-state model are structurally equivalent to those of the previously discussed continuous-angle model, namely  $\mathbf{X} = \mathbf{\Lambda}_2 \mathbf{X}$ . In a dimensionless description [lengths expressed in units of  $\sqrt{D/(2k)}$ ], the matrix  $\mathbf{\Lambda}_2$  reads

$$\mathbf{\Lambda}_2 = \begin{pmatrix} \frac{\sqrt{\mathcal{P}_2}}{2} & 1 + \frac{\mathcal{P}_2}{4} & \frac{\mathcal{P}_2}{4} \\ 1 & 0 & 0 \\ 0 & \frac{\sqrt{\mathcal{P}_2}}{2} & \frac{\sqrt{\mathcal{P}_2}}{2} \end{pmatrix}, \quad (\text{B11})$$

where we introduced the Péclet number  $\mathcal{P}_2 \equiv v^2/(2kD)$  corresponding to the two-species model. The characteristic polynomial  $|\mathbf{\Lambda}_2 - \lambda^{-1}\mathbf{1}| = 0$  delivers the cubic equation  $\lambda^{-3} + a\lambda^{-2} + b\lambda^{-1} + c = 0$ , with

$$a \equiv -\sqrt{\mathcal{P}_2}, \quad b \equiv -1, \quad c \equiv \sqrt{\mathcal{P}_2}/2. \quad (\text{B12})$$

The solutions are obtained using the same method as in the previous section (Tschirnhaus-Vieta approach).

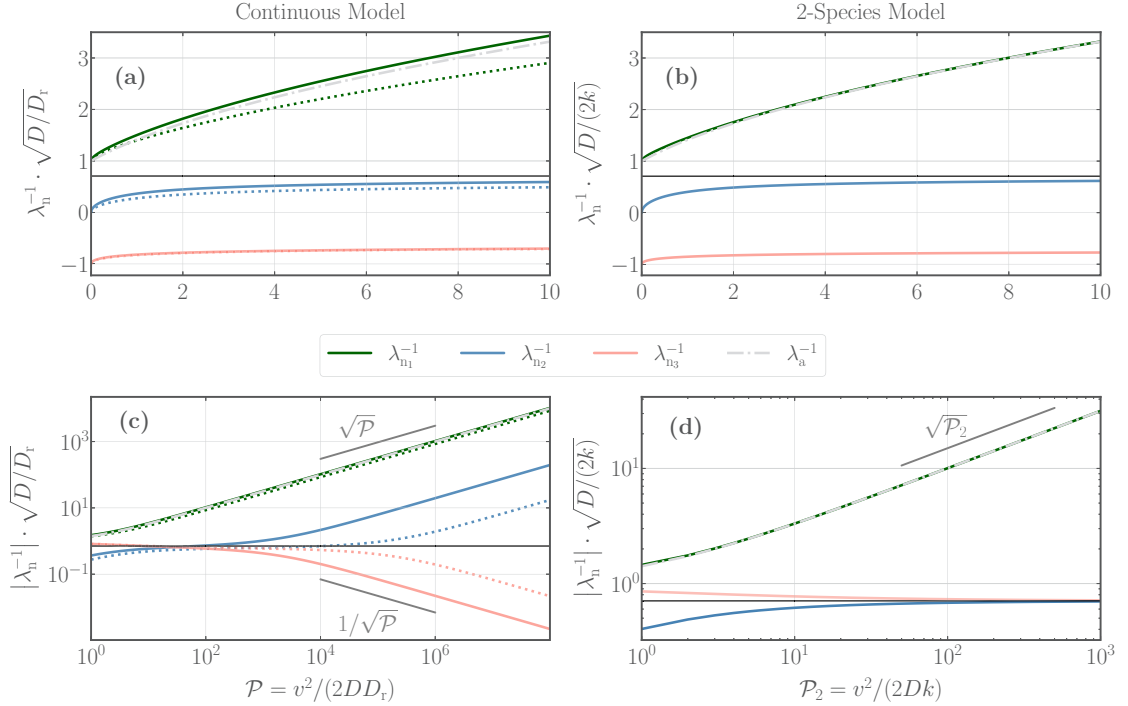


FIG. 11. Comparison of eigenvalues. Left panels: In (a), eigenvalues  $\lambda_{n_i}^{-1}$  (in units of the inverse length  $\lambda_p^{-1} = \sqrt{D_r/D}$  of a passive layer) pertaining to the continuous-angle model are plotted against the Péclet number  $\mathcal{P}$  using Eqs. (B5)–(B7) for two fixed acceptance angles,  $\alpha = 45^\circ$  (dotted curves) and  $\alpha = 90^\circ$  (solid curves). The dashed gray curve corresponds to the inverse length scale (or eigenvalue)  $\lambda_a^{-1} = \sqrt{1 + \mathcal{P}}$  of a symmetrically active polarization layer. In (c), the absolute values of eigenvalues  $\lambda_{n_i}^{-1}$  are presented on a double-logarithmic scale. Right panels: The same eigenvalues pertaining to the two-species model (cf. Sec. III A) are plotted against the corresponding Péclet number  $\mathcal{P}_2$  on a linear-linear (b) and a double-logarithmic scale (d). Eigenvalues are measured in units of the corresponding inverse length  $\sqrt{2k/D}$  of a passive layer. The gray dashed curve corresponds to  $\lambda_a^{-1} = \sqrt{1 + \mathcal{P}_2}$ , the (inverse) natural size of a fully active polarization layer. The horizontal lines in all four panels correspond to the value  $1/\sqrt{2}$ .

### 3. Comparison

The upper panels of Fig. 11 compare the eigenvalues  $\lambda_{n_i}^{-1}$  pertaining to the continuous model (a) to those corresponding to the two-state model (b). We infer that, within the plot range, the eigenvalues of both models display the same qualitative behavior. This observation proves the intuitive conjunction that the two-species model can serve as a simple model to explain the physics underlying the polarization and accumulation effects. There are, nevertheless, quantitative and qualitative differences between the two models. This becomes obvious from the lower panels of Fig. 11, where we plotted the absolute value of the eigenvalues  $\lambda_{n_i}^{-1}$  of both models on a double-logarithmic scale to visualize their behavior for large Péclet numbers. We will qualitatively compare both models in the following. Besides Fig. 11, we will refer to the content of Table I showing the limiting behavior of all eigenvalues for low Péclet numbers. During the following discussion, we will use the notion  $\mathcal{P}_{(2)}$  in order to refer to both Péclet numbers  $\mathcal{P}$  and  $\mathcal{P}_2$ . First, we focus on the eigenvalue  $\lambda_{n_1}^{-1}$ , which, for low Péclet numbers, increases proportionally to  $\sqrt{\mathcal{P}_{(2)}}$ , irrespective of the underlying model. The (inverse) characteristic size  $\lambda_a^{-1}$  of a fully active polarization layer grows only as  $O(\mathcal{P}_{(2)})$  to leading order in both models. Hence, up to order  $\sqrt{\mathcal{P}_{(2)}}$ , the eigenvalue  $\lambda_{n_1}^{-1} > \lambda_a^{-1}$ . On the other end of the

spectrum, for  $\mathcal{P}_2 \gg 1$ , eigenvalue  $\lambda_{n_1}^{-1} \sim \lambda_a^{-1} \sim \sqrt{\mathcal{P}_2}$  for the two-species model, as can be inferred from Figs. 11(b) and 11(d). The situation is more complicated for the continuous model, as can be seen in Figs. 11(a) and 11(c). Depending on the acceptance angle  $\alpha$ ,  $\lambda_{n_1}^{-1}$  can be smaller or larger than  $\lambda_a^{-1}$ , for moderate and large  $\mathcal{P}$ . As numerically determined, choosing  $\alpha \equiv \alpha_2 \approx 0.373\pi$ , one has  $\lambda_{n_1}^{-1} \sim \sqrt{\mathcal{P}}$  for  $\mathcal{P} \gg 1$ , similar to the two-species model. For  $\alpha \leq \alpha_2$  one has  $\lambda_{n_1}^{-1} \leq \lambda_a^{-1}$  for sufficiently large  $\mathcal{P}$ . Note, however, that the limit  $\mathcal{P} \gg 1$  must be treated with great caution as the orientationally continuous model is based on an approximation (3) that loses its justification for large Péclet numbers. Next, we focus on the eigenvalue  $\lambda_{n_2}^{-1}$ . For both the continuous and the

TABLE I. Behavior of the eigenvalues  $\lambda_{n_i}^{-1}$  and  $\lambda_a^{-1}$  for the continuous model (column  $\mathcal{P}$ ) as well as the two-species model (column  $\mathcal{P}_2$ ) for small Péclet numbers.

Eigenvalues	$\mathcal{P} \ll 1$	$\mathcal{P}_2 \ll 1$
$\lambda_{n_1}^{-1}$	$\mathcal{I}_4 \sqrt{\mathcal{P}/2} + 1$	$\sqrt{\mathcal{P}_2}/4 + 1$
$\lambda_a^{-1}$	$O(\mathcal{P})$	$O(\mathcal{P}_2)$
$\lambda_{n_2}^{-1}$	$\mathcal{I}_1 \sqrt{\mathcal{P}}$	$\sqrt{\mathcal{P}_2}/2$
$\lambda_{n_3}^{-1}$	$\mathcal{I}_4 \sqrt{\mathcal{P}/2} - 1$	$\sqrt{\mathcal{P}_2}/4 - 1$

two-species model,  $\lambda_{n_2}^{-1}$  grows proportionally to the square root of the respective Péclet number in the case  $\mathcal{P}_{(2)} \ll 1$ . As can be inferred from Fig. 11(d), in the limit  $\mathcal{P}_2 \rightarrow \infty$ , the eigenvalue  $\lambda_{n_2}^{-1} \rightarrow 1/\sqrt{2}$  (horizontal line) for the two-species model. Thus, for infinite activity, the nudging layer decays exponentially (since  $\lambda_{n_1} \rightarrow 0$ ) over a characteristic length  $\lambda_{n_2}$  proportional to the extent of a passive layer. This limiting behavior becomes intuitively clear by the observation that one particle species is instantaneously removed from the nudging region ( $\lambda_{n_1} = 0$ ) while the other species undergoes ordinary diffusion ( $\lambda_{n_2} \propto \lambda_p$ ) until its orientation flips. Regarding the behavior of  $\lambda_{n_2}^{-1}$  within the continuous model, we refer to Fig. 11(c). The eigenvalue  $\lambda_{n_2}^{-1}$  first seems to approach a constant value close to  $1/\sqrt{2}$  as well, but eventually starts to grow again for further increasing  $\mathcal{P}$ . Similar to  $\lambda_{n_1}^{-1}$ ,  $\lambda_{n_2}^{-1}$  grows proportionally to  $\sqrt{\mathcal{P}}$  for  $\mathcal{P} \gg 1$ . Both eigenvalues differ, however, by two to three orders of magnitude in this

limit, depending on the choice of the acceptance angle  $\alpha$ . We emphasize that the limiting behavior of  $\lambda_{n_2}^{-1}$  for  $\mathcal{P} \gg 1$  is unphysical. At infinite propulsion speed, particles are instantaneously nudged back to the interface as soon as their orientation lies within the acceptance range. The distance covered by Brownian motion until proper reorientation is proportional to  $\sqrt{D/D_r}$ . Therefore, as for the two-species model, the nudging layer should decay exponentially over said length scale for  $\mathcal{P} \rightarrow \infty$ . Figure 11(c) shows that the unphysical increase of  $\lambda_{n_2}^{-1}$  sets in at  $\mathcal{P} \approx 50$ –100, depending on the choice of the acceptance angle  $\alpha$ . Finally, the eigenvalue  $\lambda_{n_3}^{-1}$  remains negative for all Péclet numbers  $\mathcal{P}_{(2)}$  within both models. While approaching the value  $-1/\sqrt{2}$  in the limit  $\mathcal{P}_2 \rightarrow \infty$  for the two-species model,  $\lambda_{n_3}^{-1}$  approaches zero as  $1/\sqrt{\mathcal{P}}$  for  $\mathcal{P} \gg 1$  within the continuous model. The behavior of  $\lambda_{n_3}^{-1}$  for large Péclet numbers is unphysical for the same reason as for  $\lambda_{n_2}^{-1}$ .

- 
- [1] S. Ramaswamy, The mechanics and statistics of active matter, *Annu. Rev. Condens. Matter Phys.* **1**, 323 (2010).
- [2] M. E. Cates, Diffusive transport without detailed balance in motile bacteria: Does microbiology need statistical physics?, *Rep. Prog. Phys.* **75**, 042601 (2012).
- [3] P. Romanczuk, M. Bär, W. Ebeling, B. Lindner, and L. Schimansky-Geier, Active brownian particles, *Eur. Phys. J.: Spec. Top.* **202**, 1 (2012).
- [4] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe, Active particles in complex and crowded environments, *Rev. Mod. Phys.* **88**, 045006 (2016).
- [5] T. Vicsek and A. Zafeiris, Collective motion, *Phys. Rep.* **517**, 71 (2012).
- [6] S. Sponberg, The emergent physics of animal locomotion, *Phys. Today* **70**, 34 (2017).
- [7] H. Berg, *E. coli in Motion* (Springer, New York, 2004).
- [8] J. Anderson, Colloid transport by interfacial forces, *Annu. Rev. Fluid Mech.* **21**, 61 (1989).
- [9] H.-R. Jiang, N. Yoshinaga, and M. Sano, Active Motion of a Janus Particle by Self-Thermophoresis in a Defocused Laser Beam, *Phys. Rev. Lett.* **105**, 268302 (2010).
- [10] G. Falasco, R. Pfaller, A. P. Bregulla, F. Cichos, and K. Kroy, Exact symmetries in the velocity fluctuations of a hot Brownian swimmer, *Phys. Rev. E* **94**, 030602(R) (2016).
- [11] I. Buttinoni, G. Volpe, F. Kümmel, G. Volpe, and C. Bechinger, Active Brownian motion tunable by light, *J. Phys.: Condens. Matter* **24**, 284129 (2012).
- [12] M. Yang and M. Ripoll, Thermophoretically induced flow field around a colloidal particle, *Soft Matter* **9**, 4661 (2013).
- [13] J. L. Moran, P. M. Wheat, and J. D. Posner, Locomotion of electrocatalytic nanomotors due to reaction induced charge autoelectrophoresis, *Phys. Rev. E* **81**, 065302(R) (2010).
- [14] R. Golestanian, T. B. Liverpool, and A. Ajdari, Propulsion of a Molecular Machine by Asymmetric Distribution of Reaction Products, *Phys. Rev. Lett.* **94**, 220801 (2005).
- [15] J. R. Howse, R. A. L. Jones, A. J. Ryan, T. Gough, R. Vafabakhsh, and R. Golestanian, Self-Motile Colloidal Particles: From Directed Propulsion to Random Walk, *Phys. Rev. Lett.* **99**, 048102 (2007).
- [16] Z. Shen, A. Würger, and J. S. Lintuvuori, Hydrodynamic interaction of a self-propelling particle with a wall, *Eur. Phys. J. E* **41**, 39 (2018).
- [17] P. Bayati, M. N. Popescu, W. E. Uspsal, S. Dietrich, and A. Najafi, Dynamics near planar walls for various model self-phoretic particles, *Soft Matter* **15**, 5644 (2019).
- [18] T. Bickel, G. Zecua, and A. Würger, Polarization of active Janus particles, *Phys. Rev. E* **89**, 050303(R) (2014).
- [19] A. Geiseler, P. Hänggi, and F. Marchesoni, Taxis of artificial swimmers in a spatio-temporally modulated activation medium, *Entropy* **19**, 97 (2017).
- [20] J. D. Olarte-Plata and F. Bresme, Orientation of janus particles under thermal fields: The role of internal mass anisotropy, *J. Chem. Phys.* **152**, 204902 (2020).
- [21] S. Saha, S. Ramaswamy, and R. Golestanian, Pairing, waltzing and scattering of chemotactic active colloids, *New J. Phys.* **21**, 063006 (2019).
- [22] B. Nasouri and R. Golestanian, Exact axisymmetric interaction of phoretically active Janus particles, *J. Fluid Mech.* **905**, A13 (2020).
- [23] B. Qian, D. Montiel, A. Bregulla, F. Cichos, and H. Yang, Harnessing thermal fluctuations for purposeful activities: The manipulation of single micro-swimmers by adaptive photon nudging, *Chem. Sci.* **4**, 1420 (2013).
- [24] M. Selmke, U. Khadka, A. P. Bregulla, F. Cichos, and H. Yang, Theory for controlling individual self-propelled micro-swimmers by photon nudging I: Directed transport, *Phys. Chem. Chem. Phys.* **20**, 10502 (2018).
- [25] M. Selmke, U. Khadka, A. P. Bregulla, F. Cichos, and H. Yang, Theory for controlling individual self-propelled micro-swimmers by photon nudging II: Confinement, *Phys. Chem. Chem. Phys.* **20**, 10521 (2018).

- [26] G. Volpe, I. Buttinoni, D. Vogt, H. J. Kümmerer, and C. Bechinger, Microswimmers in patterned environments, *Soft Matter* **7**, 8810 (2011).
- [27] D. Takagi, J. Palacci, A. B. Braunschweig, M. J. Shelley, and J. Zhang, Hydrodynamic capture of microswimmers into sphere-bound orbits, *Soft Matter* **10**, 1784 (2014).
- [28] S. Das, A. Garg, A. I. Campbell, J. Howse, A. Sen, D. Velegol, R. Golestanian, and S. J. Ebbens, Boundaries can steer active Janus spheres, *Nat. Commun.* **6**, 1 (2015).
- [29] W. E. Uspal, M. N. Popescu, S. Dietrich, and M. Tasinkevych, Guiding Catalytically Active Particles with Chemically Patterned Surfaces, *Phys. Rev. Lett.* **117**, 048002 (2016).
- [30] S. Palagi, A. G. Mark, S. Y. Reigh, K. Melde, T. Qiu, H. Zeng, C. Parmeggiani, D. Martella, A. Sanchez-Castillo, N. Kapernaum *et al.*, Structured light enables biomimetic swimming and versatile locomotion of photoresponsive soft microbots, *Nat. Mater.* **15**, 647 (2016).
- [31] J. Simmchen, J. Katuri, W. E. Uspal, M. N. Popescu, M. Tasinkevych, and S. Sánchez, Topographical pathways guide chemical microswimmers, *Nat. Commun.* **7**, 1 (2016).
- [32] L. Baraban, D. Makarov, R. Streubel, I. Mönch, D. Grimm, S. Sanchez, and O. G. Schmidt, Catalytic janus motors on microfluidic chip: Deterministic motion for targeted cargo delivery, *ACS Nano* **6**, 3383 (2012).
- [33] J. Burdick, R. Laocharoensuk, P. M. Wheat, J. D. Posner, and J. Wang, Synthetic nanomotors in microchannel networks: Directional microchip motion and controlled manipulation of cargo, *J. Am. Chem. Soc.* **130**, 8164 (2008).
- [34] S. Ahmed, W. Wang, L. O. Mair, R. D. Fraleigh, S. Li, L. A. Castro, M. Hoyos, T. J. Huang, and T. E. Mallouk, Steering acoustically propelled nanowire motors toward cells in a biologically compatible environment using magnetic fields, *Langmuir* **29**, 16113 (2013).
- [35] H. H. Wensink, J. Dunkel, S. Heidenreich, K. Drescher, R. E. Goldstein, H. Lowen, and J. M. Yeomans, Meso-scale turbulence in living fluids, *Proc. Natl. Acad. Sci. (USA)* **109**, 14308 (2012).
- [36] R. Grossmann, P. Romanczuk, M. Bär, and L. Schimansky-Geier, Vortex Arrays and Mesoscale Turbulence of Self-Propelled Particles, *Phys. Rev. Lett.* **113**, 258104 (2014).
- [37] S. Saha, R. Golestanian, and S. Ramaswamy, Clusters, asters, and collective oscillations in chemotactic colloids, *Phys. Rev. E* **89**, 062316 (2014).
- [38] J. Bialké, H. Löwen, and T. Speck, Microscopic theory for the phase separation of self-propelled repulsive disks, *Europhys. Lett.* **103**, 30008 (2013).
- [39] M. E. Cates and J. Tailleur, When are active Brownian particles and run-and-tumble particles equivalent? Consequences for motility-induced phase separation, *Europhys. Lett.* **101**, 20010 (2013).
- [40] T. Speck, J. Bialké, A. M. Menzel, and H. Löwen, Effective Cahn-Hilliard Equation for the Phase Separation of Active Brownian Particles, *Phys. Rev. Lett.* **112**, 218304 (2014).
- [41] A. Wysocki, R. G. Winkler, and G. Gompper, Cooperative motion of active brownian spheres in three-dimensional dense suspensions, *Europhys. Lett.* **105**, 48004 (2014).
- [42] A. Zöttl and H. Stark, Hydrodynamics Determines Collective Motion and Phase Behavior of Active Colloids in Quasi-Two-Dimensional Confinement, *Phys. Rev. Lett.* **112**, 118101 (2014).
- [43] M. E. Cates and J. Tailleur, Motility-induced phase separation, *Annu. Rev. Condens. Matter Phys.* **6**, 219 (2015).
- [44] R. Großmann, P. Romanczuk, M. Bär, and L. Schimansky-Geier, Pattern formation in active particle systems due to competing alignment interactions, *Eur. Phys. J.: Spec. Top.* **224**, 1325 (2015).
- [45] M. Mijalkov, A. McDaniel, J. Wehr, and G. Volpe, Engineering Sensorial Delay to Control Phototaxis and Emergent Collective Behaviors, *Phys. Rev. X* **6**, 011008 (2016).
- [46] S. C. Takatori and J. F. Brady, Forces, stresses and the (thermo?) dynamics of active matter, *Curr. Opin. Colloid Interface Sci.* **21**, 24 (2016).
- [47] L. Caprini, U. Marini Bettolo Marconi, and A. Puglisi, Spontaneous Velocity Alignment in Motility-Induced Phase Separation, *Phys. Rev. Lett.* **124**, 078001 (2020).
- [48] W. Poon, From clarkia to escherichia and janus: The physics of natural and synthetic active colloids, *Proceedings of the International School of Physics "Enrico Fermi"* (IOS Press, Amsterdam, 2013), Vol. 184, p. 317.
- [49] A. Zöttl and H. Stark, Emergent behavior in active colloids, *J. Phys.: Condens. Matter* **28**, 253001 (2016).
- [50] A. Vikram Singh and M. Sitti, Targeted drug delivery and imaging using mobile milli/microbots: A promising future towards theranostic pharmaceutical design, *Curr. Pharmaceut. Des.* **22**, 1418 (2016).
- [51] D. Patra, S. Sengupta, W. Duan, H. Zhang, R. Pavlick, and A. Sen, Intelligent, self-powered, drug delivery systems, *Nanoscale* **5**, 1273 (2013).
- [52] M. Leyman, F. Ogemark, J. Wehr, and G. Volpe, Tuning phototactic robots with sensorial delays, *Phys. Rev. E* **98**, 052606 (2018).
- [53] M. A. Fernandez-Rodriguez, F. Grillo, L. Alvarez, M. Rathlef, I. Buttinoni, G. Volpe, and L. Isa, Feedback-controlled active brownian colloids with space-dependent rotational dynamics, *Nat. Commun.* **11**, 4223 (2020).
- [54] M. J. Schnitzer, Theory of continuum random walks and application to chemotaxis, *Phys. Rev. E* **48**, 2553 (1993).
- [55] A. Sharma and J. M. Brader, Brownian systems with spatially inhomogeneous activity, *Phys. Rev. E* **96**, 032604 (2017).
- [56] A. Sharma and J. M. Brader, Communication: Green-kubo approach to the average swim speed in active brownian systems, *J. Chem. Phys.* **145**, 161101 (2016).
- [57] H. Merlitz, H. D. Vuijk, J. Brader, A. Sharma, and J. U. Sommer, Linear response approach to active brownian particles in time-varying activity fields, *J. Chem. Phys.* **148**, 194116 (2018).
- [58] H. Row and J. F. Brady, Reverse osmotic effect in active matter, *Phys. Rev. E* **101**, 062604 (2020).
- [59] J. A. Kromer, N. de la Cruz, and B. M. Friedrich, Chemokinetic Scattering, Trapping, and Avoidance of Active Brownian Particles, *Phys. Rev. Lett.* **124**, 118101 (2020).
- [60] S. Jahanshahi, C. Lozano, B. Liebchen, H. Löwen, and C. Bechinger, Realization of a motility-trap for active particles, *Commun. Phys.* **3**, 127 (2020).
- [61] J. Arlt, V. A. Martinez, A. Dawson, T. Pilizota, and W. C. Poon, Painting with light-powered bacteria, *Nat. Commun.* **9**, 1 (2018).

- [62] G. Frangipane, D. Dell'Arciprete, S. Petracchini, C. Maggi, F. Saglimbeni, S. Bianchi, G. Vizsnyiczai, M. L. Bernardini, and R. di Leonardo, Dynamic density shaping of photokinetic E. Coli, *eLife* **7**, 1 (2018).
- [63] A. Geiseler, P. Hänggi, F. Marchesoni, C. Mulhern, and S. Savel'ev, Chemotaxis of artificial microswimmers in active density waves, *Phys. Rev. E* **94**, 012613 (2016).
- [64] A. Geiseler, P. Hänggi, and F. Marchesoni, Self-polarizing microswimmers in active density waves, *Sci. Rep.* **7**, 41884 (2017).
- [65] A. Fischer, F. Schmid, and T. Speck, Quorum-sensing active particles with discontinuous motility, *Phys. Rev. E* **101**, 012601 (2020).
- [66] A. Fischer, F. Schmid, and T. Speck, Erratum: Quorum-sensing active particles with discontinuous motility [Phys. Rev. E **101**, 012601 (2020)], *Phys. Rev. E* **102**, 059903(E) (2020).
- [67] K. Malakar, V. Jemseena, A. Kundu, K. Vijay Kumar, S. Sabhapandit, S. N. Majumdar, S. Redner, and A. Dhar, Steady state, relaxation and first-passage properties of a run-and-tumble particle in one-dimension, *J. Stat. Mech.: Theor. Expt.* (2018) 043215.
- [68] A. R. Sprenger, M. A. Fernandez-Rodriguez, L. Alvarez, L. Isa, R. Wittkowski, and H. Löwen, Active brownian motion with orientation-dependent motility: Theory and experiments, *Langmuir* **36**, 7066 (2020).
- [69] H. D. Vuijk, A. Sharma, D. Mondal, J. U. Sommer, and H. Merlitz, Pseudochemotaxis in inhomogeneous active Brownian systems, *Phys. Rev. E* **97**, 042612 (2018).
- [70] S. Hermann and M. Schmidt, Active interface polarization as a state function, *Phys. Rev. Research* **2**, 022003(R) (2020).
- [71] Y. Fily and M. C. Marchetti, Athermal Phase Separation of Self-Propelled Particles with no Alignment, *Phys. Rev. Lett.* **108**, 235702 (2012).
- [72] S. Hermann and M. Schmidt, Active ideal sedimentation: exact two-dimensional steady states, *Soft Matter* **14**, 1614 (2018).
- [73] S. Hermann, P. Krinninger, D. de las Heras, and M. Schmidt, Phase coexistence of active Brownian particles, *Phys. Rev. E* **100**, 052604 (2019).
- [74] L. Caprini and U. Marini Bettolo Marconi, Active particles under confinement and effective force generation among surfaces, *Soft Matter* **14**, 9044 (2018).
- [75] N. A. Söker, S. Auschra, V. Holubec, K. Kroy, and F. Cichos, How Activity Landscapes Polarize Microswimmers without Alignment Forces, *Phys. Rev. Lett.* **126**, 228001 (2021).
- [76] G. Gompper, C. Bechinger, S. Herminghaus, R. Isele-Holder, U. B. Kaupp, H. Löwen, H. Stark, and R. G. Winkler, Microswimmers—From single particle motion to collective behavior, *Eur. Phys. J.: Spec. Top.* **225**, 2061 (2016).
- [77] J. Vachier and M. G. Mazza, Dynamics of sedimenting active brownian particles, *Eur. Phys. J. E* **42**, 11 (2019).
- [78] M. Enculescu and H. Stark, Active Colloidal Suspensions Exhibit Polar Order Under Gravity, *Phys. Rev. Lett.* **107**, 058301 (2011).
- [79] F. Ginot, A. Solon, Y. Kafri, C. Ybert, J. Tailleur, and C. Cottin-Bizonne, Sedimentation of self-propelled janus colloids: polarization and pressure, *New J. Phys.* **20**, 115001 (2018).
- [80] T. Speck and R. L. Jack, Ideal bulk pressure of active Brownian particles, *Phys. Rev. E* **93**, 062605 (2016).
- [81] C. G. Wagner, M. F. Hagan, and A. Baskaran, Steady-state distributions of ideal active brownian particles under confinement and forcing, *J. Stat. Mech.: Theor. Expt.* (2017) 043203.
- [82] J. Elgeti and G. Gompper, Wall accumulation of self-propelled spheres, *Europhys. Lett.* **101**, 48003 (2013).
- [83] S. Paliwal, J. Rodenburg, R. van Roij, and M. Dijkstra, Chemical potential in active systems: predicting phase equilibrium from bulk equations of state?, *New J. Phys.* **20**, 015003 (2018).
- [84] A. P. Solon, J. Stenhammar, M. E. Cates, Y. Kafri, and J. Tailleur, Generalized thermodynamics of phase equilibria in scalar active matter, *Phys. Rev. E* **97**, 020602(R) (2018).
- [85] A. P. Solon, J. Stenhammar, M. E. Cates, Y. Kafri, and J. Tailleur, Generalized thermodynamics of motility-induced phase separation: phase equilibria, laplace pressure, and change of ensembles, *New J. Phys.* **20**, 075001 (2018).
- [86] V. Prymidis, S. Paliwal, M. Dijkstra, and L. Fillion, Vapour-liquid coexistence of an active lennard-jones fluid, *J. Chem. Phys.* **145**, 124904 (2016).
- [87] S. Paliwal, V. Prymidis, L. Fillion, and M. Dijkstra, Non-equilibrium surface tension of the vapour-liquid interface of active lennard-jones particles, *J. Chem. Phys.* **147**, 084902 (2017).
- [88] U. Erdmann, W. Ebeling, L. Schimansky-Geier, and F. Schweitzer, Brownian particles far from equilibrium, *Eur. Phys. J. B* **15**, 105 (2000).
- [89] F. Schweitzer, *Brownian Agents and Active Particles* (Springer, Berlin, 2007).
- [90] A. P. Solon, M. E. Cates, and J. Tailleur, Active brownian particles and run-and-tumble particles: A comparative study, *Eur. Phys. J.: Spec. Top.* **224**, 1231 (2015).
- [91] A. P. Bregulla, H. Yang, and F. Cichos, Stochastic localization of microswimmers by photon nudging, *ACS Nano* **8**, 6542 (2014).
- [92] R. Golestanian, Collective Behavior of Thermally Active Colloids, *Phys. Rev. Lett.* **108**, 038303 (2012).
- [93] E. Bertin, M. Droz, and G. Grégoire, Boltzmann and hydrodynamic description for self-propelled particles, *Phys. Rev. E* **74**, 022101 (2006).
- [94] H. Risken, *The Fokker-Planck Equation—Methods of Solution and Applications*, 2nd ed., edited by H. Haken (Springer-Verlag, Berlin, 1989).
- [95] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, 2nd ed., edited by H. Haken (Springer-Verlag, Berlin, 1985).
- [96] V. Holubec, K. Kroy, and S. Steffenoni, Physically consistent numerical solver for time-dependent Fokker-Planck equations, *Phys. Rev. E* **99**, 032117 (2019).
- [97] J. Tailleur and M. E. Cates, Statistical Mechanics of Interacting Run-And-Tumble Bacteria, *Phys. Rev. Lett.* **100**, 218103 (2008).
- [98] J. Tailleur and M. E. Cates, Sedimentation, trapping, and rectification of dilute bacteria, *Europhys. Lett.* **86**, 60002 (2009).
- [99] N. Razin, Entropy production of an active particle in a box, *Phys. Rev. E* **102**, 030103(R) (2020).
- [100] A. J. Rodenburg, Thermodynamic variables for active Brownian particles: Pressure, surface tension, and chemical potential, Ph.D. thesis, Utrecht University (2020).
- [101] A. Patch, D. M. Sussman, D. Yllanes, and M. C. Marchetti, Curvature-dependent tension and tangential flows at the interface of motility-induced phases, *Soft Matter* **14**, 7435 (2018).

- [102] A. K. Omar, Z.-G. Wang, and J. F. Brady, Microscopic origins of the swim pressure and the anomalous surface tension of active matter, *Phys. Rev. E* **101**, 012604 (2020).
- [103] M. B. Miller and B. L. Bassler, Quorum sensing in bacteria, *Annu. Rev. Microbiol.* **55**, 165 (2001).
- [104] S. C. Takatori, W. Yan, and J. F. Brady, Swim Pressure: Stress Generation in Active Matter, *Phys. Rev. Lett.* **113**, 028103 (2014).
- [105] R. G. Winkler, A. Wysocki, and G. Gompper, Virial pressure in systems of spherical active brownian particles, *Soft Matter* **11**, 6680 (2015).
- [106] A. P. Solon, J. Stenhammar, R. Wittkowski, M. Kardar, Y. Kafri, M. E. Cates, and J. Tailleur, Pressure and Phase Equilibria in Interacting Active Brownian Spheres, *Phys. Rev. Lett.* **114**, 198301 (2015).
- [107] A. P. Solon, Y. Fily, A. Baskaran, M. E. Cates, Y. Kafri, M. Kardar, and J. Tailleur, Pressure is not a state function for generic active fluids, *Nat. Phys.* **11**, 673 (2015).
- [108] G. H. Weiss, Some applications of persistent random walks and the telegrapher's equation, *Physica A* **311**, 381 (2002).

**Density and polarization of active Brownian particles in curved activity landscapes**Sven Auschra<sup>1,\*</sup> and Viktor Holubec<sup>1,2,†</sup><sup>1</sup>*Institute for Theoretical Physics, Leipzig University, 04103 Leipzig, Germany*<sup>2</sup>*Department of Macromolecular Physics, Faculty of Mathematics and Physics, Charles University, V Holešovičkách 2, CZ-180 00 Praha, Czech Republic*

(Received 13 January 2021; accepted 12 April 2021; published 4 June 2021)

Suspensions of motile active particles with space-dependent activity form characteristic polarization and density patterns. Recent single-particle studies for planar activity landscapes identified several quantities associated with emergent density-polarization patterns that are solely determined by bulk variables. Naive thermodynamic intuition suggests that these results might hold for arbitrary activity landscapes mediating bulk regions, and thus could be used as benchmarks for simulations and theories. However, the considered system operates in a nonequilibrium steady state and we prove by construction that the quantities in question lose their simple form for curved activity landscapes. Specifically, we provide a detailed analytical study of polarization and density profiles induced by radially symmetric activity steps, and of the total polarization for the case of a general radially symmetric activity landscape. While the qualitative picture is similar to the planar case, all the investigated variables depend not only on bulk variables but also comprise geometry-induced contributions. We verified that all our analytical results agree with exact numerical calculations.

DOI: [10.1103/PhysRevE.103.062604](https://doi.org/10.1103/PhysRevE.103.062604)**I. INTRODUCTION**

To feed, hide, or proliferate, both macroscopic [1–3] and microscopic [4–8] living organisms actively adjust their motion to mechanical, optical, or chemical stimuli. The ability to change motility based on the state of the environment is also vital for artificial *motile* active matter ranging from robots [9,10] to microscopic active particles [11–14], where some of the ultimate goals are noninvasive drug delivery and microsurgery [15,16]. On a much lower level of sophistication, large assemblies of active particles exhibit motility-induced phase separation (MIPS) [17] into a dense and slow, and dilute and fast phase [18–20]. Typically, but not exclusively, this separation is a consequence of a density-dependent propulsion speed [21–24].

The inhomogeneous or space-dependent activity comes hand in hand with characteristic modulations in the local density and polarization [7,8,22,25–28]. Given the omnipresence of inhomogeneous activity at all scales of active matter, the latter can serve as mesoscale indicator for intrinsic microscopic activity in the system [29,30]. In spite of that, a thorough investigation of characteristic patterns in the local density and polarization attracted a focused attention of the active matter community only recently.

It was shown [29,30] that density and polarization for a single micrometer-sized Janus swimmer in water are well captured by the active Brownian particle model [31–35] for noninteracting active spheres in a noisy environment. For a single swimmer and a planar activity interface [30], this model

allows us to identify three quantities that are solely determined by bulk diffusion coefficients, swim speeds, and system size, and thus acquire the status of thermodynamic state variables. Namely, (i) the local polarization peak at the interface, (ii) the ratio of densities of the bulk regions on either side of the interface, and (iii) the total polarization caused by the activity step. The latter two maintain this property irrespective of the shape of the (one-dimensional) activity modulations, as long as they mediate between two bulk regions [30,36]. If generally valid, these simple relations can serve as consistency checks for simulations and benchmarks for theories [36].

In this paper, we prove by construction that these results, in general, do not hold for other than planar activity profiles. Concretely, we applied the theoretical framework of Ref. [30] to radially symmetric activity steps and investigated in detail the resulting polarization and density patterns. Our analytical results show that the quantities (i)–(iii) depend on the nonzero curvature of the interface and thus on the geometry of the setup. We also investigate the (radial) total polarization for general radially symmetric motility modulations and show that it acquires a geometry-induced nonlocal contribution and hence is no longer determined only by bulk variables. In the limit of vanishing curvature, the obtained results converge to those for planar activity steps [30,36]. Our theoretical results can be readily tested using the experimental setup used in Ref. [37].

Our results would be surprising for a system in thermodynamic equilibrium with solid walls, where their shape does not affect bulk properties. However, they might be expected for the active-matter system at hand, as it operates in a nonequilibrium steady state. Indeed, the dependence of bulk properties in active-matter systems on the shape of their physical boundaries has been observed in Refs. [38–41]. The

\*sven.auschra@gmail.com

†viktor.holubec@mff.cuni.cz

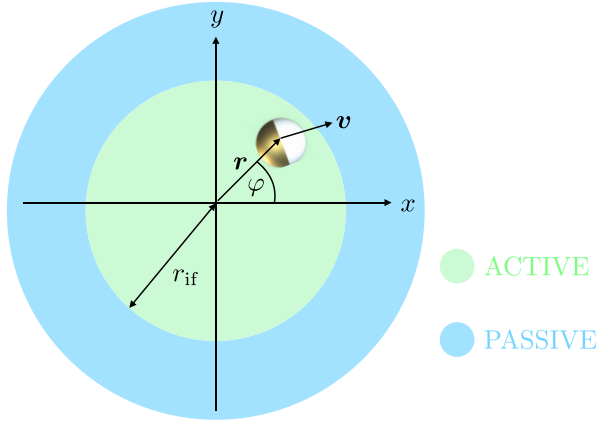


FIG. 1. Janus particle with coordinates  $x = r \cos \phi$ ,  $y = r \sin \phi$  subjected to a radially symmetric activity profile. The particle propels actively along its orientation  $\mathbf{v}/v = (\cos \theta, \sin \theta)^\top$  for  $r \equiv |\mathbf{r}| < r_{\text{if}}$ . Otherwise, its swim speed  $v$  is zero.

dependence on the interface curvature found here can be compared to the Laplace pressure [42], e.g., in soap bubbles. The main difference between the two setups is that the increased pressure inside a bubble is caused by a physical force applied in the form of the surface tension by the soap film on the bubble interior. The activity interface in our setup is fixed and the observed influence of its curvature can be traced to geometry-induced imbalance of probability currents across the curved interface.

## II. THE MODEL

Consider an overdamped Janus swimmer with space-dependent propulsion speed (activity)  $v(x, y)$  and orientation parametrized by the angle  $\theta$  confined in a plane. For a piecewise constant radially symmetric activity profile, we depict the system in Fig. 1. We model the particle dynamics by the active Brownian particle model [43] described by the system of Langevin equations:

$$\partial_t x = v(x, y) \cos \theta + \sqrt{2D} \xi_x, \quad (1)$$

$$\partial_t y = v(x, y) \sin \theta + \sqrt{2D} \xi_y, \quad (2)$$

$$\partial_t \theta = \sqrt{2D_r} \xi_\theta. \quad (3)$$

The translational and rotational diffusion coefficients  $D$  and  $D_r$ , respectively, measure intensities of independent, unit variance, and unbiased Gaussian white noise processes  $\xi_{x,y,\theta}(t)$ .

In the following section, we utilize the framework of Ref. [30] to derive approximate differential equations for the stationary probability density  $\rho(\mathbf{r})$  to find the particle at position  $\mathbf{r}$  and the corresponding polarization  $\mathbf{p}(\mathbf{r})$ .

## III. MOMENT EQUATIONS

The dynamic probability density  $f(\mathbf{r}, \hat{\mathbf{n}}, t)$  for finding the Janus swimmer at time  $t$  at position  $\mathbf{r}$  with the orientation  $\hat{\mathbf{n}} = (\cos \theta, \sin \theta)^\top$ , corresponding to the system of stochastic differential equations (1)–(3), obeys the Fokker-Planck equation (FPE) [34,35,44]:

$$\partial_t f = D \nabla^2 f + D_r \partial_\theta^2 f - \nabla \cdot [f v(\mathbf{r}) \hat{\mathbf{n}}]. \quad (4)$$

Here,  $\partial_t \equiv \partial/\partial t$ , and  $\nabla$  represents the Nabla operator with respect to  $\mathbf{r}$ . The exact moment expansion of  $f$  in terms of  $\hat{\mathbf{n}}$  [34,44,45] truncated after the second term reads [30]

$$f(\mathbf{r}, \hat{\mathbf{n}}, t) = \frac{1}{2\pi} [\rho(\mathbf{r}, t) + 2\mathbf{p}(\mathbf{r}, t) \cdot \hat{\mathbf{n}}], \quad (5)$$

where

$$\rho(\mathbf{r}, t) \equiv \int d\hat{\mathbf{n}} f(\mathbf{r}, \hat{\mathbf{n}}, t), \quad (6)$$

$$\mathbf{p}(\mathbf{r}, t) \equiv \int d\hat{\mathbf{n}} \hat{\mathbf{n}} f(\mathbf{r}, \hat{\mathbf{n}}, t) \quad (7)$$

denote time-resolved density and polarization, respectively. Multiplying Eq. (4) by 1 or  $\hat{\mathbf{n}}$ , integrating over orientational degrees of freedom, and using the definitions (6) and (7), we obtain the moment equations [43,44]:

$$\partial_t \rho(\mathbf{r}, t) = -\nabla \cdot \mathbf{J}(\mathbf{r}, t), \quad (8)$$

$$\partial_t \mathbf{p}(\mathbf{r}, t) = -D_r \mathbf{p}(\mathbf{r}, t) - \nabla \cdot \mathbf{M}(\mathbf{r}, t). \quad (9)$$

Here, we introduced the (orientation averaged) flux,

$$\mathbf{J}(\mathbf{r}, t) \equiv -D \nabla \rho(\mathbf{r}, t) + v(\mathbf{r}) \mathbf{p}(\mathbf{r}, t), \quad (10)$$

and the matrix flux,

$$\mathbf{M}(\mathbf{r}, t) \equiv -D \nabla \mathbf{p}(\mathbf{r}, t) + \frac{v(\mathbf{r})}{2} \rho(\mathbf{r}, t) \mathbf{1}, \quad (11)$$

with the unit matrix  $\mathbf{1}$ .

Throughout the rest of this paper, we will focus on the steady-state solutions  $\rho(\mathbf{r})$  and  $\mathbf{p}(\mathbf{r})$  of Eqs. (8) and (9), which obey  $\partial_t \rho = \partial_t \mathbf{p} = 0$  and thus

$$D \nabla^2 \rho(\mathbf{r}) = \nabla \cdot [v(\mathbf{r}) \mathbf{p}(\mathbf{r})], \quad (12)$$

$$D \nabla^2 \mathbf{p}(\mathbf{r}) = D_r \mathbf{p}(\mathbf{r}) + \frac{1}{2} \nabla [v(\mathbf{r}) \rho(\mathbf{r})]. \quad (13)$$

Moreover, we assume that, under no-flux boundary conditions, the stationary flux  $\mathbf{J}(\mathbf{r})$  vanishes. While this assumption is not generally valid in two (or higher) dimensions, it holds for all setups considered below. Exploiting the no-flux condition in Eq. (10) and substituting the resulting formula

$$\nabla \rho(\mathbf{r}) = \frac{v(\mathbf{r})}{D} \mathbf{p}(\mathbf{r}), \quad (14)$$

into Eq. (13), we obtain

$$\nabla^2 \mathbf{p}(\mathbf{r}) = \frac{\mathbf{p}(\mathbf{r})}{\lambda^2(\mathbf{r})} + \frac{\rho(\mathbf{r})}{2D} \nabla v(\mathbf{r}), \quad (15)$$

where we have introduced the length scale

$$\lambda(\mathbf{r}) \equiv \left[ \frac{D_r}{D} + \frac{v^2(\mathbf{r})}{2D^2} \right]^{-1/2}. \quad (16)$$

A thorough discussion and physical interpretation of this characteristic length scale in the case of a planar motility step is given in Refs. [29,30] and we omit it here.



#### IV. ACTIVE-PASSIVE INTERFACE

We will now solve Eqs. (14) and (15) for  $\rho$  and  $\mathbf{p}$  for the radially symmetric activity step sketched in Fig. 1. In this setup, the swim speed  $v(r) \equiv v_0$ , for  $r \equiv \sqrt{x^2 + y^2} < r_{\text{if}}$ , and is zero otherwise.

Polarization and density must reflect the radial symmetry of the activity profile leading to  $\rho = \rho(r)$  and  $\mathbf{p} = p(r)\hat{\mathbf{r}}$ , where  $\hat{\mathbf{r}} \equiv \mathbf{r}/r = (\cos \varphi, \sin \varphi)^\top$ . Using this ansatz, the flux-balance condition (14) reduces to

$$\rho'(r) = \frac{v(r)}{D} p(r), \quad (17)$$

where  $\rho'(r) \equiv \partial \rho / \partial r$ . Exploiting this relation in the moment Eq. (15) yields

$$p''(r) = -\frac{p'(r)}{r} + \frac{p(r)}{r^2} + \frac{p(r)}{\lambda^2(r)} + \frac{v'(r)\rho(r)}{2D}. \quad (18)$$

For the planar setup of Refs. [29,30] corresponding to  $r_{\text{if}} \rightarrow \infty$ , the first two terms on the right-hand side (r.h.s.) of this equation are zero, which suggests that the polarization and its derivative decay with the distance from the interface faster than  $1/r^2$  and  $1/r$ , respectively. In general, the last term of Eq. (18) vanishes everywhere except for  $r = r_{\text{if}}$ , since  $v'(r) = -v_0 \delta(r - r_{\text{if}})$ , with the Dirac delta function  $\delta(r)$ . Within the active ( $r \leq r_{\text{if}}$ ) and passive region ( $r > r_{\text{if}}$ ), Eq. (18) reduces to the modified Bessel equation [7,8]. Its general solution reads [46]

$$p_{a,p}(r) = C_{a,p}^{(1)} I_1(r/\lambda_{a,p}) + C_{a,p}^{(2)} K_1(r/\lambda_{a,p}), \quad (19)$$

where  $I_m(x)$  and  $K_m(x)$  are the modified Bessel functions of the first and second kinds, respectively. The characteristic length scales

$$\lambda_a \equiv \left( \frac{D_r}{D} + \frac{v_0^2}{2D^2} \right)^{-1/2}, \quad \lambda_p \equiv \left( \frac{D_r}{D} \right)^{-1/2}, \quad (20)$$

follow from Eq. (16) evaluated in the active and passive region, respectively.

To create bulk regions with constant density and vanishing polarization both in the active and in the passive region, we demand in the following that the active-passive interface is far enough both from the origin and from the system's boundary at  $r = R$ . That is, we assume that  $r_{\text{if}}$  and  $R - r_{\text{if}}$  are several times greater than  $\lambda_a$  and  $\lambda_p$ , respectively. This allows us to apply the boundary conditions

$$p_a(r=0) = 0, \quad p_p(r=R) = 0. \quad (21)$$

Then the general solution (19) simplifies to

$$p(r) = \begin{cases} p_a(r) = C_a I_1(r/\lambda_a) & \text{for } r \leq r_{\text{if}} \\ p_p(r) = C_p K_1(r/\lambda_p) & \text{for } r > r_{\text{if}}. \end{cases} \quad (22)$$

Integration of Eq. (17) delivers the corresponding density

$$\rho(r) = \begin{cases} \rho_a + \frac{v C_a \lambda_a}{D} [I_0(r/\lambda_a) - 1] & \text{for } r \leq r_{\text{if}} \\ \rho_a + \frac{v C_p \lambda_a}{D} [I_0(r_{\text{if}}/\lambda_a) - 1] \equiv \rho_p & \text{for } r > r_{\text{if}}, \end{cases} \quad (23)$$

which assumes the bulk value  $\rho_a \equiv \rho(0)$  in the active and  $\rho_p \equiv \rho(r_{\text{if}}) \equiv \rho(R)$  in the passive region. The constants  $C_a$  and  $C_p$  in Eq. (22) can be determined from continuity conditions on  $p$  and the corresponding flux  $\mathbf{M}$  at the active-passive

interface [47]. Demanding the polarization  $p(r)$  and the projection

$$\mathbf{M} \cdot \hat{\mathbf{r}} = \left[ -D p'(r) + \frac{v(r)\rho(r)}{2} \right] \hat{\mathbf{r}} \quad (24)$$

of the matrix flux (11) onto the radial direction to be continuous at  $r = r_{\text{if}}$  renders

$$p_a(r_{\text{if}}) = p_p(r_{\text{if}}), \quad (25)$$

$$p'_a(r_{\text{if}}) - p'_p(r_{\text{if}}) = \frac{v_0}{2D} \rho(r_{\text{if}}). \quad (26)$$

The density (23) satisfies  $\rho_a(r_{\text{if}}) = \rho_p(r_{\text{if}})$  by construction. The normal component  $\mathbf{J} \cdot \hat{\mathbf{r}}$  is continuous due to the imposed no-flux condition  $\mathbf{J} \equiv \mathbf{0}$ . The constant  $\rho_a$  in Eq. (23) follows from the normalization condition

$$\int_0^{2\pi} d\phi \int_0^R dr r \rho(r) = 1. \quad (27)$$

Figures 2(a) and 2(b) show nice agreement of the approximate analytic density and polarization profiles (22) and (23) with exact numerical solutions [48] for two distinct particle activities, expressed in terms of the Péclet number

$$\mathcal{P} \equiv \frac{v_0^2}{2DD_r}. \quad (28)$$

We observe nice agreement with exact numerical solutions [48] in both cases. In Sec. VB, we show that the approximate solutions deviate from the exact results for much smaller Péclet numbers than  $\mathcal{P} = 100$  found for straight planar interfaces [49]. This is because Eq. (18) cannot be mapped onto an exactly solvable (two-species) run-and-tumble model [27] providing the same phenomenology as the full model [50], which was the key ingredient for success of the approximate solutions at planar interfaces [29,30].

The qualitative behavior of density and polarization profiles is the same as for a planar activity step [29,30]. Namely, an increased activity step induces a higher polarization and a larger ratio  $\rho_p/\rho_a$  of bulk densities of the passive and active regions. The polarization peaks exactly at the active-passive interface and decays over characteristic lengths  $\lambda_p$  and  $\lambda_a < \lambda_p$  into the passive and active region, respectively. The density profile remains constant at the bulk density  $\rho = \rho_p$  throughout the whole passive region. On crossing the interface, it decays to the bulk density  $\rho_a < \rho_p$  pertaining to the active region over a length scale  $\lambda_a$ . We refer to Refs. [29,30] for a more detailed physical interpretation and discussion of the emerging polarization and density variations. Here, we focus on the influence of curvature on these profiles.

In Figs. 2(c) and 2(d), we show the (reduced) density and polarization profiles corresponding to the inverse setup for which the particle is passive for  $r < r_{\text{if}}$  and active otherwise. The derivation of the analytic profiles (solid and dashes lines) is similar to the above calculations, and is detailed in Appendix B. The approximate theory profiles overlap with the corresponding exact numerical solutions. The qualitative picture is similar to the situation shown in Figs. 2(a) and 2(b), we flipped the active and passive region, and a negative polarization in the vicinity of the interface, confirming that the particle preferably points into the passive region [29,30]. Note, however, that the convexity or concavity of the activity

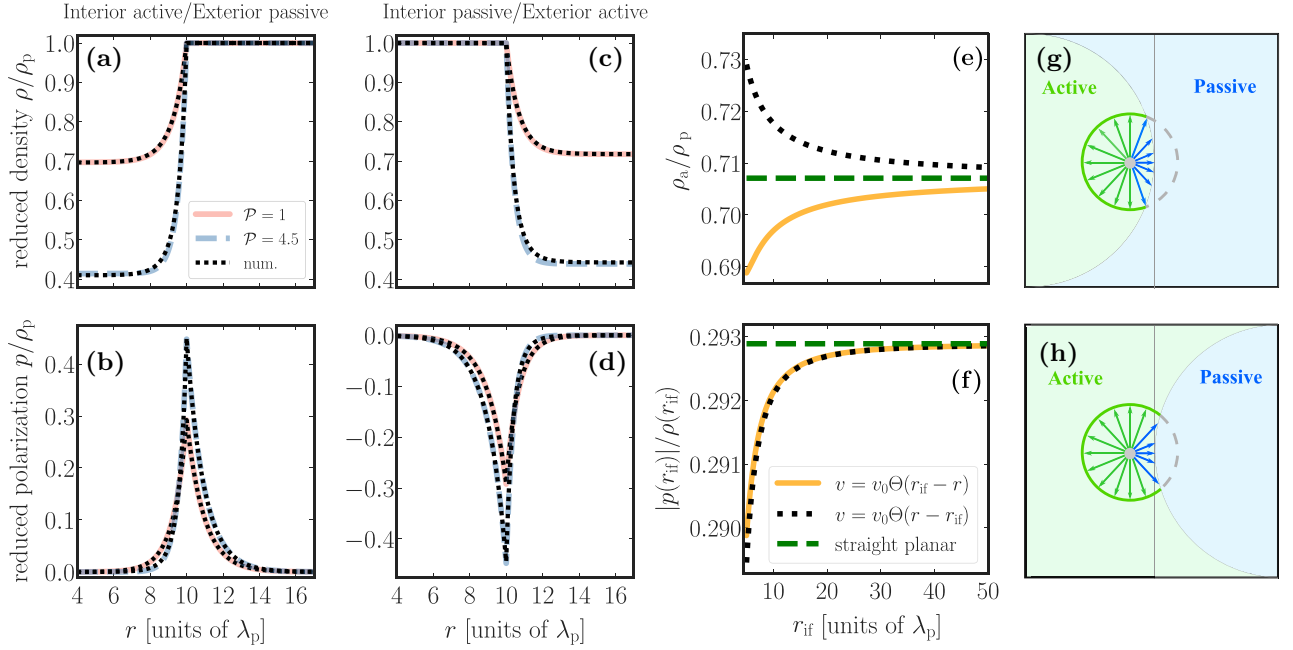


FIG. 2. (a)–(d) Reduced density (top) and polarization (bottom) profiles near a radially symmetric active-passive [(a),(b)] and passive-active [(c),(d)] interfaces at  $r = r_{if}$ . Lengths are measured in units of  $\lambda_p = \sqrt{D/D_r}$  and velocities in units of  $\sqrt{2DD_r}$ . This corresponds to a dimensionless theoretical description in terms of the Péclet number  $\mathcal{P}$  (28). Theory profiles (solid and dashed curves) were calculated for two distinct  $\mathcal{P}$  and  $r_{if} = 10\lambda_p$  using Eqs. (22)–(27) in (a) and (b) and Eqs. (B1)–(B6) in (c) and (d), and compared against exact numerically obtained [48] profiles (dotted lines). (e), (f) Density ratio,  $\rho_a/\rho_p$ , and magnitude of the (reduced) polarization at the interface,  $|p(r_{if})|/\rho(r_{if})$ , for  $\mathcal{P} = 1$  as functions of the radial distance  $r_{if}$  of the interface. Plotted curves correspond to the following analytical expressions (setups): Solid curves (interior active and exterior passive): Eqs. (29) and (30). Dotted curves (interior passive and exterior active): Eqs. (B8) and (B9). Dashed lines (straight planar case): Eqs. (33) and (35). (g), (h) Particle inside the active region in the vicinity of a concave (convex) active-passive interface. In both panels, the vertical line corresponds to a straight active-passive interface. Relative to the latter case, for a concave (convex) geometry, the particle has a higher (lower) chance to end up in the passive region.

interface also leads to quantitative differences between the two cases.

To grasp the influence of the curvature of the activity step more quantitatively, we compare the maximum relative polarization,  $p(r_{if})/\rho(r_{if})$ , which constitutes a suitable order parameter for the polarization at the interface, and the bulk density ratio,  $\rho_a/\rho_p$ , for circular and straight planar interfaces. For the setup where the particle is active for  $r < r_{if}$  and passive otherwise [Figs. 2(a) and 2(b)], these quantities are given by

$$\frac{p(r_{if})}{\rho(r_{if})} = \frac{v_0}{2D} G_A(r_{if}), \quad (29)$$

$$\frac{\rho_a}{\rho_p} = 1 - \frac{v_0^2 \lambda_a}{2D^2} G_A(r_{if}) \left( \frac{I_0}{I_1} - \frac{1}{I_1} \right), \quad (30)$$

as detailed in Appendix A. Here, the (geometry) function reads

$$G_A(r_{if}) \equiv \left( \frac{I_0 + I_2}{2\lambda_a I_1} + \frac{K_0 + K_2}{2\lambda_p K_1} \right)^{-1}, \quad (31)$$

with  $I_m \equiv I_m(r_{if}/\lambda_a)$  and  $K_m \equiv K_m(r_{if}/\lambda_p)$ . The subscript A indicates that  $G_A$  corresponds to the case where the particle is *active* for  $r < r_{if}$ . The geometry function  $G_P(r_{if})$  for the inverted setup is derived in Appendix B. Exploiting the asymptotic expansions  $I_n(z) \sim e^z/\sqrt{2\pi z}$  and

$K_n(z) \sim e^{-z}/\sqrt{2z/\pi}$ , valid for  $z \gg 1$  irrespective of the order  $n$  [46], one finds

$$G_A(r_{if}) \sim \frac{\lambda_a \lambda_p}{\lambda_a + \lambda_p}. \quad (32)$$

For  $r_{if} \gg \lambda_p > \lambda_a$ , the maximum (relative) polarization therefore approaches

$$\frac{p(r_{if})}{\rho(r_{if})} \sim \frac{v_0}{2D} \frac{\lambda_a \lambda_p}{\lambda_a + \lambda_p} = \frac{1}{\sqrt{2}} \frac{\sqrt{\mathcal{P}}}{1 + \sqrt{1 + \mathcal{P}}}, \quad (33)$$

which coincides with the expression found for planar interfaces [30]. The corresponding asymptotic behavior of the density ratio (30),

$$\frac{\rho_a}{\rho_p} \sim 1 - \frac{v_0^2 \lambda_a}{2D^2} \frac{\lambda_a \lambda_p}{\lambda_a + \lambda_p} \left( 1 - \sqrt{\frac{2\pi r_{if}}{\lambda_a}} e^{-r_{if}/\lambda_a} \right), \quad (34)$$

still displays an exponential decaying with  $r_{if}$ . By taking the limit  $r_{if} \rightarrow \infty$ , it reduces to the result found at planar activity steps [30]:

$$\frac{\rho_a}{\rho_p} = \frac{\lambda_a}{\lambda_p} = \frac{1}{\sqrt{1 + \mathcal{P}}}. \quad (35)$$

The analytic expressions for  $p(r_{if})/\rho(r_{if})$  and  $\rho_a/\rho_p$  for the inverted setup [Figs. 2(c) and 2(d)] are derived in Appendix B along similar lines.

Figures 2(e) and 2(f) show the dependence of these quantities on the radius of curvature  $r_{\text{if}}$  of the interface for both circular setups, as well as their counterparts (33) and (35) for straight planar interfaces (dashed horizontal lines). With increasing  $r_{\text{if}}$ , the polarization peaks  $|p(r_{\text{if}})|/\rho(r_{\text{if}})$  for the circular setup approach—the one for the straight motility step from below. For the setup where the particle is active for  $r < r_{\text{if}}$  (solid curve), the peak is slightly larger ( $\approx 0.02\%$  for  $r_{\text{if}} = 10 \lambda_p$ ) than for the inverted setup (dotted curve). Turning to the bulk density ratio, if the interior is active (solid line), the bulk density ratio  $\rho_a/\rho_p$  is smaller as compared to the straight planar case, whereas it exceeds it for the inverse circular setup (dotted curve).

This behavior can be intuitively understood as a result of a geometry-induced imbalance in probability fluxes across the curved interface. Consider the situations sketched in Figs. 2(g) and 2(h). In both panels, a Janus particle (gray dot) is situated inside an active region, in a close proximity to an adjacent passive region. In Fig. 2(g), the active-passive interface is concave, whereas in Fig. 2(h), it is convex. In both panels, the vertical lines correspond to a straight active-passive interface. For simplicity, consider a quasiballistic particle motion denoted by the arrows in both panels. As indicated by the number of blue arrows relative to the green ones, the particle's chance to enter the passive region is higher for the concave [g)] than convex [h)] geometry. The setup with active interior [g)] thus yields a larger bulk density ratio  $\rho_a/\rho_p$  than the inverse setup [h)]. It follows that the density ratio corresponding to a concave (convex) active-passive interface is always smaller (larger) than its counterpart for a straight planar interface. As the curvature of the circular activity interface decreases, i.e., for  $r_{\text{if}} \rightarrow \infty$ , the bulk density ratio for straight interfaces is approached.

To gain an intuition on why the magnitude of the reduced polarization,  $|p(r_{\text{if}})|/\rho(r_{\text{if}})$ , is always larger in the straight planar case than for a circular interface [see Fig. 2(f)] is more difficult. The absolute value of the polarization  $|p(r_{\text{if}})|$  depends on the probability that the particle with a given orientation hits the interface. Compared to the planar case, for the concave interface shown in Fig. 2(g), there are less active particles in the bulk to hit the interface with a broader range of polarizations, and vice versa for the convex interface shown in Fig. 2(h). Hence we observe three competing ingredients that determine the absolute polarization in the concave (convex) case: low (high) bulk density in the active region, large (small) probability for a given particle to hit the interface, and large (small) average polarization of particles which hit the interface, where the strength of the individual ingredients is compared to the planar case. Furthermore, the magnitude of the reduced polarization is obtained as absolute polarization divided by density of the passive bulk, which is high for the concave and low for the convex case. We thus find in both circular setups two ingredients leading to an increase and two leading to a decrease of  $|p(r_{\text{if}})|/\rho(r_{\text{if}})$ . Our analytical results show that they compensate each other in such a way that the magnitude of the reduced polarization for circular interfaces is always lower than in the planar case. Unfortunately, it seems impossible to guess the influence of these ingredients based on physical intuition.

The maximum polarization and density ratios, which are for planar motility steps solely determined by bulk variables ( $v_0, D, D_r$ ) [29,30] thus depend on the interface radius  $r_{\text{if}}$  in case of a circular activity step. This suggests that arbitrarily curved activity steps generally yield geometry-induced contributions to the emergent density-polarization patterns and the corresponding maximum polarization and bulk density ratio. To provide further evidence for this conjecture, we now study the total polarization, which is also solely determined by bulk quantities in the case of planar interfaces [36], for arbitrary radially symmetric activity modulations.

## V. TOTAL POLARIZATION

Without alignment forces, local polarization in active-matter systems arises from spacial sorting of particles with different orientations. Therefore the total polarization vector,  $\mathbf{P}_{\text{tot}}$ , defined as the integral

$$\mathbf{P}_{\text{tot}} \equiv \int_{\mathcal{V}} d\mathbf{r} p(\mathbf{r}) \quad (36)$$

over the whole space  $\mathcal{V}$ , universally vanishes for systems with no-flux boundary conditions [36].

A more appropriate definition of total polarization induced by activity landscapes that mediate between two bulk regions is to restrict the domain of integration  $\mathcal{V}$  so it connects the two bulk regions. For planar activity profiles, it is natural to integrate along a ray of fixed width parallel to the  $x$  axis and thus perpendicular to the interface. Then, the magnitude  $P_{\text{tot}}$  of such defined total polarization is proportional to the difference in strengths of fluxes,  $v\rho$ , corresponding to the two bulk regions [29,30], and thus it acquires the status of a thermodynamic state variable. For an arbitrary radial activity profile  $v(r)$  that mediates between two bulk regions of, respectively, constant activity, the radial symmetry implies that the local polarization profile must be of the form  $\mathbf{p} = p(r)\hat{\mathbf{r}}$ , with  $\hat{\mathbf{r}} = (\cos\phi, \sin\phi)^\top$ . To match the planar definition, we define the magnitude of the total radial polarization as the integral

$$P_{\text{tot}} \equiv \int_{R_1}^{R_2} dr p(r) \quad (37)$$

of the projection  $p(r)$  of the polarization vector onto the radial axis over a ray of fixed infinitesimal width, perpendicular to the interface, and mediating the inner bulk region at radius  $R_1$  and the outer one at  $R_2$  (see Fig. 3). Alternatively, and more naturally from the point of view of polar coordinates, one could integrate the polarization over an infinitesimal wedge mediating the two bulks. This would correspond to substituting  $rp(r)$  for  $p(r)$  in the definition (37). However, in this case, the width of the integration region increases with  $r$ .

Below we show that  $P_{\text{tot}}$  is generally composed of a contribution proportional to the difference  $v(R_1)\rho(R_1) - v(R_2)\rho(R_2)$  of the flux strengths, as for planar interfaces [36], and a second nonlocal contribution induced by the nonzero curvature of the interface. A similar expression also holds for the alternative definition of  $P_{\text{tot}}$  with  $rp(r)$ .

### A. Derivation of total polarization

We introduce polar coordinates,  $x = r \cos\phi$  and  $y = r \sin\phi$ , and the angular variable  $\psi \equiv \theta - \phi$ , which measures

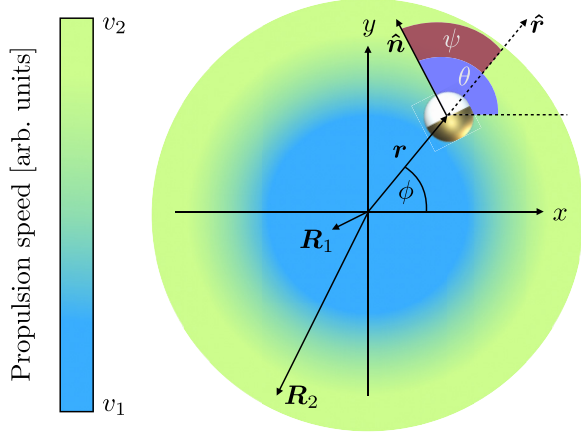


FIG. 3. A Janus particle at position  $\mathbf{r} = (r \cos \phi, r \sin \phi)^\top$  actively propelling along its orientation  $\hat{\mathbf{n}} = (\cos \theta, \sin \theta)^\top$ . The particle's orientation relative to the radial unit vector  $\hat{\mathbf{r}}$  is measured by the angle  $\psi$ . The swimmer's propulsion speed follows a radially symmetric activity profile (color coded), which mediates between an inner (dark blue,  $R_1$ ) and outer (light green,  $R_2$ ) bulk region with different constant activities.

the particle orientation relative to the radial axis (see Fig. 3). Under this transformation, the system of three Langevin equations (1)–(3) reduces to the two-dimensional set

$$\partial_t r = v(r) \cos \psi + \sqrt{2D} \xi_r, \quad (38)$$

$$\partial_t \psi = -\frac{v(r)}{r} \sin \psi + \sqrt{2\left(\frac{D}{r^2} + D_r\right)} \xi_\psi, \quad (39)$$

where  $\xi_r$  and  $\xi_\psi$  denote independent, zero-mean, unbiased Gaussian white noise processes. The associated FPE for the stationary probability density  $\tilde{f}(r, \psi)$  for finding the particle at distance  $r$  with relative orientation  $\psi$  reads

$$0 = -\partial_r \tilde{\mathcal{J}} + \left(\frac{D}{r^2} + D_r\right) \partial_\psi^2 \tilde{f} + \frac{v(r)}{r} \partial_\psi (\sin \psi \tilde{f}), \quad (40)$$

where we have introduced the (angle-resolved) flux:

$$\tilde{\mathcal{J}}(r, \psi) \equiv -D \partial_r \tilde{f} + \frac{D}{r} \tilde{f} + \cos \psi v(r) \tilde{f}. \quad (41)$$

Due to the radial symmetry,  $\tilde{f}$  and  $\tilde{\mathcal{J}}$  must be even functions of  $\psi$ , i.e.,  $\tilde{f}(r, \psi) = \tilde{f}(r, -\psi)$ , and  $\tilde{\mathcal{J}}(r, \psi) = \tilde{\mathcal{J}}(r, -\psi)$ . The Fourier expansion of the flux  $\tilde{\mathcal{J}}$  thus reads

$$\tilde{\mathcal{J}}(r, \psi) = \sum_{n=0}^{\infty} \tilde{\mathcal{J}}_n(r) \cos(n\psi), \quad (42)$$

where  $\tilde{\mathcal{J}}_n(r)$  is the  $n$ th Fourier coefficient. Plugging this series into the FPE (40) and integrating twice over the angle from 0 to  $\psi$  yields

$$\begin{aligned} \left(\frac{D}{r^2} + D_r\right) \tilde{f}(r, \psi) &= \tilde{f}_0(r) - \sum_{n=1}^{\infty} \partial_r \tilde{\mathcal{J}}_n(r) \frac{\cos(n\psi)}{n^2} \\ &\quad - \frac{v(r)}{r} \int_0^\psi d\tilde{\psi} \sin \tilde{\psi} \tilde{f}(r, \tilde{\psi}). \end{aligned} \quad (43)$$

The unknown function  $\tilde{f}_0(r)$  stems from the second integration. The integration constant from the first integration renders, after the second integration, a term linear in  $\psi$ , and thus it must be zero to maintain periodicity.

The radial component of the local polarization vector  $\mathbf{p}(r)$  is defined by

$$\mathbf{p}(r) \equiv \int_0^{2\pi} d\psi \cos \psi \tilde{f}(r, \psi). \quad (44)$$

The corresponding tangential component  $\int_0^{2\pi} d\psi \sin \psi \tilde{f}(r, \psi)$  vanishes due to the radial symmetry. Multiplying Eq. (43) by  $\cos \psi$ , integrating over  $\psi$  from 0 to  $2\pi$ , and using orthogonality of trigonometric functions renders

$$\begin{aligned} \left(\frac{D}{r^2} + D_r\right) \mathbf{p}(r) \\ = -\pi \partial_r \tilde{\mathcal{J}}_1(r) + \frac{v(r)}{r} \int_0^{2\pi} d\psi \sin^2 \psi \tilde{f}(r, \psi). \end{aligned} \quad (45)$$

The integral on the r.h.s. was obtained by interchanging the order of the double integration [51]. Using the definitions (41) and (44), we find that the first coefficient  $\tilde{\mathcal{J}}_1$  of the Fourier series (42) in terms of  $\tilde{f}$  and  $\mathbf{p}$  reads

$$\tilde{\mathcal{J}}_1 = \frac{1}{\pi} \int_0^{2\pi} d\psi \tilde{\mathcal{J}}(r, \psi) \cos \psi, \quad (46)$$

$$= \frac{1}{\pi} \left(\frac{D}{r} \mathbf{p} - D \partial_r \mathbf{p}\right) + \frac{v}{\pi} \int_0^{2\pi} d\psi \tilde{f}(r, \psi) \cos^2 \psi. \quad (47)$$

The distributions  $\tilde{f}$  and  $\mathbf{p}$  and the corresponding polarizations  $\mathbf{p}$  and  $p$  are connected via the Jacobian [47]  $|\partial(x, y)/\partial(r, \phi)| = r$ , i.e.,  $\tilde{f} = r f$  and  $\mathbf{p} = r p$ . Plugging these transformations and Eq. (47) into Eq. (45) yields

$$D_r p = D \partial_r \left(\frac{p}{r} + \partial_r p\right) - \partial_r (v \langle \cos^2 \psi \rangle) - \frac{v}{r} \langle 2 \cos^2 \psi - 1 \rangle. \quad (48)$$

Here, the averaging is defined as

$$\langle \bullet \rangle \equiv \int_0^{2\pi} d\psi \bullet f(r, \psi). \quad (49)$$

Finally, Eqs. (37) and  $2 \cos^2 \psi - 1 = \cos(2\psi)$  render the closed expression for the total polarization,

$$P_{\text{tot}} = \frac{D}{D_r} \left(\frac{p}{r} + \partial_r p\right) \Big|_{R_1}^{R_2} - \frac{v}{D_r} \langle \cos^2 \psi \rangle \Big|_{R_1}^{R_2} - \mathcal{I}[v](R_1, R_2), \quad (50)$$

where we introduced the functional

$$\mathcal{I}[v](R_1, R_2) \equiv \int_{R_1}^{R_2} dr \frac{v(r)}{r D_r} \langle \cos(2\psi) \rangle. \quad (51)$$

Within bulk regions, we have  $p(R_{1/2}) = \partial_r p(R_{1/2}) = 0$  and  $f(R_{1/2}, \psi) = \rho(R_{1/2})/(2\pi)$ . Hence, the first term on the r.h.s. of Eq. (50) vanishes and the second one simplifies to  $\langle \cos^2 \psi \rangle(R_{1/2}) = \rho(R_{1/2})/2$ . The total polarization between two bulk regions thus reads

$$P_{\text{tot}} = \frac{v(R_1)\rho(R_1) - v(R_2)\rho(R_2)}{2D_r} - \mathcal{I}[v](R). \quad (52)$$

The first summand above coincides with the total polarization found for planar interfaces [29,30]. The second term is a nonlocal contribution attributed to the nonzero curvature of the considered activity profile. It vanishes when the activity profile becomes effectively planar, i.e., when its radius diverges while the thickness measured by the distance between the two bulk regions remains finite.

The (radial) total polarization for curved activity profiles is thus not solely determined by stationary properties of the bulk and, in this sense, loses its status of a state variable. Using Eq. (45), one can show along similar lines as above that a similar expression would be obtained for the alternative definition of the total polarization using  $r\rho$  instead of  $\rho$  in Eq. (37). Specifically, the first term in this total polarization follows from Eq. (52) after substituting  $p$  for  $p$  in the first term on the r.h.s. The curvature-dependent term changes more but also vanishes for diverging radius of the interface.

We now demonstrate that  $\mathcal{I}[v](R_1, R_2)$  in Eq. (52) vanishes when truncating the exact moment expansion of  $f(r, \psi)$  after two terms as in Eq. (5) and quantify deviations between the exact solution and approximate solution of Sec. III.

### B. Approximate global sum rule and deviations

Plugging  $\rho = \rho(r)$  and  $\mathbf{p} = p(r)\hat{\mathbf{r}}$  into the truncated moment expansion (5), the approximate distribution function can be written as

$$f(r, \psi) = \frac{1}{2\pi} [\rho(r) + p(r) \cos \psi]. \quad (53)$$

Using the average (49), the density  $\rho(r)$  and polarization  $p(r)$  are given by (1) and  $\langle \cos \psi \rangle$ , respectively. All higher (angular) moments within the approximation (53) vanish. In particular,  $\langle \cos(2\psi) \rangle = 0$  and the functional  $\mathcal{I}$  (51) gives zero. The total polarization (52) then reads

$$P_{\text{tot}} = \frac{v(R_1)\rho(R_1) - v(R_2)\rho(R_2)}{2D_r}, \quad (54)$$

which is the result found for one-dimensional activity landscapes [30,36]. For the radial activity profiles considered here, it only holds within the approximation  $\langle \cos(n\psi) \rangle = 0$  for  $n > 1$ .

To verify the exact analytical result (52) for the total polarization and to assess the scope of the approximation (54), we numerically calculated the exact distributions  $f(r, \psi)$  [48] for several radially symmetric activity steps. As in Sec. IV, the particle propels actively if its radial distance  $r < r_{\text{if}}$  and its swimming mechanism is switched off otherwise (see Fig. 1).

The upper panel of Fig. 4 depicts the moments  $\langle \cos(n\psi) \rangle$ ,  $n = 0, 1, 2$  calculated from Eq. (49) using the exact distribution  $f(r, \psi)$  for Péclet number  $\mathcal{P} = 40$ . The moments are normalized by the bulk density  $\rho_a$  in the active region. In the active region, the second moment  $\langle \cos(2\psi) \rangle$  is up to the active-passive interface largely negative, rendering its contribution (51) to the total polarization (52) positive. Since the amplitude of the second moment increases with Péclet number, the approximate result for the total polarization (54) underestimates its exact value the more the larger the particle activity, as can be inferred from the lower panel of Fig. 4. It shows that the relative deviation between the two reaches

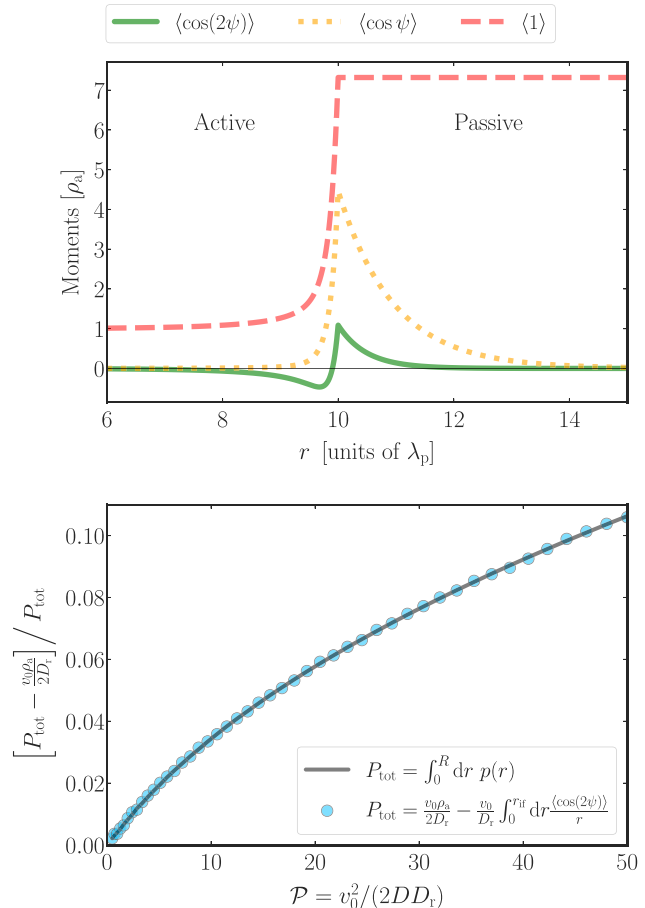


FIG. 4. Moments [Eq. (49), top] and relative deviations of the total polarization (Eq. (37), bottom) from its approximate value (54) in the vicinity of a radially symmetric active-passive interface at  $r = r_{\text{if}} = 10 \lambda_p$ . The particle propels actively if its radial distance  $r < r_{\text{if}}$  and only diffuses otherwise. Lengths are measured in units of  $\lambda_p = \sqrt{D/D_r}$  and velocities in units of  $\sqrt{2DD_r}$ . The overall radial extension of the system was chosen  $R = 20 \lambda_p$ , and in the top figure we took the Péclet number  $\mathcal{P} = v_0 / (2DD_r) = 40$ . The presented data was obtained using the exact numerically determined distribution  $f(r, \psi)$  [48]. The solid line and circles in the bottom panel were determined from Eqs. (37) and (52), respectively.

roughly 10% for  $\mathcal{P} = 50$  and thus the approximation (54) is reasonably accurate for experimentally realizable Péclet numbers [29]. The lower panel of Fig. 4 also shows perfect agreement between the analytical result (52) and polarization evaluated from Eq. (37) using the numerically determined density  $f(r, \psi)$ .

## VI. CONCLUSION

We derived approximate analytical formulas for polarization and density profiles induced by a radially symmetric motility step. These results nicely agree with numerically determined exact solutions even beyond the limit of small activities. We further evaluated the effect of nonzero curvature of the active-passive interface on the polarization and density.

Reduced polarization is smaller than for planar activity steps, whereas the contrast in the bulk densities is smaller (larger) for concave (convex) active-passive interfaces. Both the maximum polarization and the bulk density ratio depend on the curvature of the interface.

Furthermore, we derived an exact formula for the (radial) total polarization induced by an arbitrary radially symmetric activity landscape. Compared to the result for one-dimensional activity landscapes, the total polarization contains a nonlocal, geometry-induced correction. The total polarization is thus no longer determined solely by bulk variables. This result proves that curved active-passive interfaces generally yield a geometry-induced contribution to the emergent density and polarization profiles and the associated total polarization and bulk density ratio.

#### ACKNOWLEDGMENTS

We thank Klaus Kroy for valuable discussions. We acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG) through the priority program Microswimmers (SPP 1726, Project No. 237143019) and by Czech Science Foundation (Project No. 20-02955J). V.H. gratefully acknowledges support by the Humboldt Foundation. The authors acknowledge financial support by the German Research Foundation Project No. 432421051 (DFG-GACR cooperation).

#### APPENDIX A: POLARIZATION PEAK AND DENSITY RATIO

Within the active ( $r < r_{if}$ ) and passive region ( $r > r_{if}$ ) the general solution of the respective polarization and density profiles read [Eqs. (22) and (23)]

$$p_a(r) = C_a I_1(r/\lambda_a), \quad \lambda_a = \left( \frac{D_r}{D} + \frac{v_0^2}{2D^2} \right)^{-1/2}, \quad (\text{A1})$$

$$p_p(r) = C_p K_1(r/\lambda_p), \quad \lambda_p = \sqrt{\frac{D}{D_r}}, \quad (\text{A2})$$

$$\rho_a(r) = \rho_a + \frac{C_a \lambda_a v_0}{D} [I_0(r/\lambda_a) - 1], \quad (\text{A3})$$

$$\rho_p(r) \equiv \rho_a(r_{if}) \equiv \rho_p = \text{const}, \quad (\text{A4})$$

with  $\rho_a \equiv \rho_a(0)$ . The emerging integration constants  $C_a$  and  $C_p$  follow from the continuity conditions

$$p_p(r_{if}) = p_a(r_{if}), \quad p'_a(r_{if}) - p'_p(r_{if}) = \frac{v_0}{2D} \rho(r_{if}), \quad (\text{A5})$$

given Eqs. (25) and (26) in the main text. The first condition implies that

$$C_p = C_a \frac{I_1(r_{if}/\lambda_a)}{K_1(r_{if}/\lambda_p)}. \quad (\text{A6})$$

For the sake of brevity, we introduce the abbreviations  $I_n \equiv I_n(r_{if}/\lambda_a)$  and  $K_n \equiv K_n(r_{if}/\lambda_p)$ . The second condition in Eq. (A5) yields

$$p'_a - p'_p = C_a \frac{I_0 + I_2}{2\lambda_a} + C_p \frac{K_0 + K_2}{2\lambda_p} = \frac{v_0}{2D} \rho_p \quad (\text{A7})$$

and, using Eq. (A6), we get

$$\frac{C_a}{\rho_p} = \frac{v_0}{2D} \left( \frac{I_0 + I_2}{2\lambda_a} + \frac{I_1}{K_1} \frac{K_0 + K_2}{2\lambda_p} \right)^{-1}. \quad (\text{A8})$$

Knowing that the polarization peaks exactly at the interface (see Fig. 2), it follows from Eqs. (B1) and (B3) that the maximum relative polarization  $p(r_{if})/\rho(r_{if})$  is given by  $C_a I_1/\rho_p$ . Using Eq. (A8) and introducing the geometry function (31),

$$G_A(r_{if}) = \left( \frac{I_0 + I_2}{2\lambda_a I_1} + \frac{K_0 + K_2}{2\lambda_p K_1} \right)^{-1}, \quad (\text{A9})$$

the maximum relative polarization reads

$$\frac{p(r_{if})}{\rho(r_{if})} = \frac{v_0}{2D} G_A(r_{if}), \quad (\text{A10})$$

as given in Eq. (29). The relative density profile,  $(\rho - \rho_a)/\rho_p$ , and thus also the bulk density ratio,  $\rho_p/\rho_a$ , can be obtained using Eqs. (B4), (B3), (A8), and a similar approach.

#### APPENDIX B: INTERIOR PASSIVE-EXTERIOR ACTIVE

We now consider the case where the particle is passive for  $r < r_{if}$ , and otherwise active. In analogy to the calculations in Appendix A, one now has

$$p_p(r) = C_p I_1(r/\lambda_p), \quad (\text{B1})$$

$$p_a(r) = C_a K_1(r/\lambda_a), \quad (\text{B2})$$

$$\rho_p(r) \equiv \rho_a(r_{if}) \equiv \rho_p = \text{const}, \quad (\text{B3})$$

$$\rho_a(r) = \rho_p + \frac{C_a \lambda_a v_0}{D} [K_0(r_{if}/\lambda_a) - K_0(r/\lambda_a)]. \quad (\text{B4})$$

Note that the argument of  $I_1$  ( $K_1$ ) now carries  $\lambda_p$  ( $\lambda_a$ ) as a characteristic length scale. Integration constants  $C_a$  and  $C_p$  are determined by the same continuity conditions (A5) as in Appendix A, yielding

$$C_p = C_a \frac{K_1(r_{if}/\lambda_a)}{I_1(r_{if}/\lambda_p)}, \quad (\text{B5})$$

$$\frac{C_a}{\rho_p} = -\frac{v_0}{2D} \left( \frac{K_1}{I_1} \frac{I_0 + I_2}{2\lambda_p} + \frac{K_0 + K_2}{2\lambda_a} \right)^{-1}, \quad (\text{B6})$$

where we used the abbreviations  $I_n \equiv I_n(r_{if}/\lambda_p)$  and  $K_n \equiv K_n(r_{if}/\lambda_a)$ . Reduced density and polarization profiles,  $\rho/\rho_p$  and  $p/\rho_p$ , are plotted in Figs. 2(c) and 2(d). Introducing the geometry function

$$G_P(r_{if}) \equiv \left( \frac{I_0 + I_2}{2\lambda_p I_1} + \frac{K_0 + K_2}{2\lambda_a K_1} \right)^{-1}, \quad (\text{B7})$$

the (negative) polarization peak and the bulk density ratio are given by

$$\frac{p(r_{if})}{\rho(r_{if})} = -\frac{v_0}{2D} G_P(r_{if}), \quad (\text{B8})$$

$$\frac{\rho_a}{\rho_p} = 1 - \frac{v_0^2 \lambda_a}{2D^2} G_P(r_{if}) \frac{K_0}{K_1}. \quad (\text{B9})$$

Both quantities are plotted in Figs. 2(e) and 2(f) (dotted lines). For  $r_{if} \rightarrow \infty$ , both approach their counterparts at straight planar active-passive interfaces [Eqs. (33) and (35)].

- [1] N. Uchida, A. Kepecs, and Z. F. Mainen, Seeing at a glance, smelling in a whiff: Rapid forms of perceptual decision making, *Nat. Rev. Neurosci.* **7**, 485 (2006).
- [2] S. Sponberg, The emergent physics of animal locomotion, *Phys. Today* **70**(9), 34 (2017).
- [3] B. Moulia, C. Coutand, and J.-L. Julien, Mechanosensitive control of plant growth: Bearing the load, sensing, transducing, and responding, *Front. Plant Sci.* **6**, 52 (2015).
- [4] U. B. Kaupp, J. Solzin, E. Hildebrand, J. E. Brown, A. Helbig, V. Hagen, M. Beyermann, F. Pampaloni, and I. Weyand, The signal flow and motor response controlling chemotaxis of sea urchin sperm, *Nat. Cell Biol.* **5**, 109 (2003).
- [5] B. M. Friedrich and F. Jülicher, Chemotaxis of sperm cells, *Proc. Natl. Acad. Sci.* **104**, 13256 (2007).
- [6] M. B. Miller and B. L. Bassler, Quorum sensing in bacteria, *Annu. Rev. Microbiol.* **55**, 165 (2001).
- [7] A. Fischer, F. Schmid, and T. Speck, Quorum-sensing active particles with discontinuous motility, *Phys. Rev. E* **101**, 012601 (2020).
- [8] A. Fischer, F. Schmid, and T. Speck, Erratum: Quorum-sensing active particles with discontinuous motility, *Phys. Rev. E* **102**, 059903(E) (2020).
- [9] M. Mijalkov, A. McDaniel, J. Wehr, and G. Volpe, Engineering Sensorial Delay to Control Phototaxis and Emergent Collective Behaviors, *Phys. Rev. X* **6**, 011008 (2016).
- [10] C. Virágh, G. Vásárhelyi, N. Tarcai, T. Szörényi, G. Somorjai, T. Nepusz, and T. Vicsek, Flocking algorithm for autonomous flying robots, *Bioinspiration Biomimetics* **9**, 025012 (2014).
- [11] J. Anderson, Colloid transport by interfacial forces, *Annu. Rev. Fluid Mech.* **21**, 61 (1989).
- [12] W. C. K. Poon, From Clarkia to Escherichia and Janus: The physics of natural and synthetic active colloids, *Physics of Complex Colloids: Volume 184*, Proceedings of the International School of Physics “Enrico Fermi” (IOS Press, 2013), pp. 317–386.
- [13] A. Zöttl and H. Stark, Emergent behavior in active colloids, *J. Phys.: Condens. Matter* **28**, 253001 (2016).
- [14] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe, Active particles in complex and crowded environments, *Rev. Mod. Phys.* **88**, 045006 (2016).
- [15] D. Patra, S. Sengupta, W. Duan, H. Zhang, R. Pavlick, and A. Sen, Intelligent, self-powered, drug delivery systems, *Nanoscale* **5**, 1273 (2013).
- [16] A. Vikram Singh and M. Sitti, Targeted drug delivery and imaging using mobile milli/microrobots: A promising future towards theranostic pharmaceutical design, *Curr. Pharm. Des.* **22**, 1418 (2016).
- [17] M. E. Cates and J. Tailleur, Motility-induced phase separation, *Annu. Rev. Condens. Matter Phys.* **6**, 219 (2015).
- [18] Y. Fily and M. C. Marchetti, Athermal Phase Separation of Self-Propelled Particles with no Alignment, *Phys. Rev. Lett.* **108**, 235702 (2012).
- [19] I. Buttinoni, J. Bialké, F. Kümmel, H. Löwen, C. Bechinger, and T. Speck, Dynamical Clustering and Phase Separation in Suspensions of Self-Propelled Colloidal Particles, *Phys. Rev. Lett.* **110**, 238301 (2013).
- [20] G. S. Redner, M. F. Hagan, and A. Baskaran, Structure and Dynamics of a Phase-Separating Active Colloidal Fluid, *Phys. Rev. Lett.* **110**, 055701 (2013).
- [21] A. P. Solon, J. Stenhammar, M. E. Cates, Y. Kafri, and J. Tailleur, Generalized thermodynamics of phase equilibria in scalar active matter, *Phys. Rev. E* **97**, 020602(R) (2018).
- [22] A. P. Solon, J. Stenhammar, M. E. Cates, Y. Kafri, and J. Tailleur, Generalized thermodynamics of motility-induced phase separation: phase equilibria, laplace pressure, and change of ensembles, *New J. Phys.* **20**, 075001 (2018).
- [23] J. Bialké, H. Löwen, and T. Speck, Microscopic theory for the phase separation of self-propelled repulsive disks, *Europhys. Lett.* **103**, 30008 (2013).
- [24] T. Speck, A. M. Menzel, J. Bialké, and H. Löwen, Dynamical mean-field theory and weakly non-linear analysis for the phase separation of active Brownian particles, *J. Chem. Phys.* **142**, 224109 (2015).
- [25] M. J. Schnitzer, Theory of continuum random walks and application to chemotaxis, *Phys. Rev. E* **48**, 2553 (1993).
- [26] A. Sharma and J. M. Brader, Brownian systems with spatially inhomogeneous activity, *Phys. Rev. E* **96**, 032604 (2017).
- [27] K. Malakar, V. Jemseena, A. Kundu, K. Vijay Kumar, S. Sabhapandit, S. N. Majumdar, S. Redner, and A. Dhar, Steady state, relaxation and first-passage properties of a run-and-tumble particle in one-dimension, *J. Stat. Mech.: Theory Exp.* (2018) 043215.
- [28] S. Hermann, P. Krinninger, D. de las Heras, and M. Schmidt, Phase coexistence of active Brownian particles, *Phys. Rev. E* **100**, 052604 (2019).
- [29] N. A. Söker, S. Auschra, V. Holubec, K. Kroy, and F. Cichos, How Activity Landscapes Polarize Microswimmers without Alignment Forces, *Phys. Rev. Lett.* **126**, 228001 (2021).
- [30] S. Auschra, V. Holubec, N. A. Söker, F. Cichos, and K. Kroy, Polarization-density patterns of active particles in motility gradients, *Phys. Rev. E* **103**, 062601 (2021).
- [31] U. Erdmann, W. Ebeling, L. Schimansky-Geier, and F. Schweitzer, Brownian particles far from equilibrium, *Eur. Phys. J. B* **15**, 105 (2000).
- [32] F. Schweitzer, *Brownian Agents and Active Particles* (Springer, Berlin, 2007).
- [33] P. Romanczuk, M. Bär, W. Ebeling, B. Lindner, and L. Schimansky-Geier, Active Brownian particles, *Eur. Phys. J.: Spec. Top.* **202**, 1 (2012).
- [34] M. E. Cates and J. Tailleur, When are active Brownian particles and run-and-tumble particles equivalent? Consequences for motility-induced phase separation, *Europhys. Lett.* **101**, 20010 (2013).
- [35] A. P. Solon, M. E. Cates, and J. Tailleur, Active Brownian particles and run-and-tumble particles: A comparative study, *Eur. Phys. J.: Spec. Top.* **224**, 1231 (2015).
- [36] S. Hermann and M. Schmidt, Active interface polarization as a state function, *Phys. Rev. Research* **2**, 022003(R) (2020).
- [37] N. A. Söker, Heat driven Janus particles with spatially varying propulsion velocity, Master’s thesis, University of Leipzig, 2018.
- [38] N. Nikola, A. P. Solon, Y. Kafri, M. Kardar, J. Tailleur, and R. Voituriez, Active Particles with Soft and Curved Walls: Equation of State, Ratchets, and Instabilities, *Phys. Rev. Lett.* **117**, 098001 (2016).
- [39] R. Wittmann, F. Smallenburg, and J. M. Brader, Pressure, surface tension, and curvature in active systems: A touch of equilibrium, *J. Chem. Phys.* **150**, 174908 (2019).

- [40] Y. Fily, A. Baskaran, and M. F. Hagan, Dynamics and density distribution of strongly confined noninteracting nonaligning self-propelled particles in a nonconvex boundary, *Phys. Rev. E* **91**, 012125 (2015).
- [41] H. Wioland, F. G. Woodhouse, J. Dunkel, J. O. Kessler, and R. E. Goldstein, Confinement stabilizes a bacterial suspension into a spiral vortex, *Phys. Rev. Lett.* **110**, 268102 (2013).
- [42] H.-J. Butt, K. Graf, and M. Kappl, *Physics and Chemistry of Interfaces* (John Wiley & Sons, Weinheim, 2013).
- [43] M. E. Cates, Diffusive transport without detailed balance in motile bacteria: Does microbiology need statistical physics? *Rep. Prog. Phys.* **75**, 042601 (2012).
- [44] R. Golestanian, Collective Behavior of Thermally Active Colloids, *Phys. Rev. Lett.* **108**, 038303 (2012).
- [45] E. Bertin, M. Droz, and G. Grégoire, Boltzmann and hydrodynamic description for self-propelled particles, *Phys. Rev. E* **74**, 022101 (2006).
- [46] M. Abramowitz, *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables* (Dover Publications, New York, 1965).
- [47] H. Risken, in *The Fokker-Planck Equation—Methods of Solution and Applications*, 2nd ed., edited by H. Haken (Springer-Verlag, Berlin, 1989).
- [48] V. Holubec, K. Kroy, and S. Steffenoni, Physically consistent numerical solver for time-dependent Fokker-Planck equations, *Phys. Rev. E* **99**, 032117 (2019).
- [49] A. J. Rodenburg, Thermodynamic variables for active brownian particles: Pressure, surface tension, and chemical potential, Ph.D. thesis, Utrecht University, 2020.
- [50] In the case of planar activity steps, the corresponding approximate moment equations, similar to Eqs. (17) and (18), could be mapped onto an exact one-dimensional run-and-tumble model of left- or right-moving particles [30]. This two-species model robustly captured the essential physics behind the formal results obtained for planar steps and justifies the large range of Péclet numbers over which the approximate moment equations [cf. Eqs. (17) and (18)] deliver accurate results. Since this equivalence between the two models is no longer present, deviations between approximate analytic and exact numerical results already occur for smaller Péclet numbers.
- [51]

$$\begin{aligned} & \int_0^{2\pi} d\psi \cos\psi \int_0^\psi d\tilde{\psi} \tilde{f}(r, \tilde{\psi}) \sin \tilde{\psi} \\ &= \int_0^{2\pi} d\tilde{\psi} \tilde{f}(r, \tilde{\psi}) \sin \tilde{\psi} \int_{\tilde{\psi}}^{2\pi} d\psi \cos \psi \end{aligned}$$



---

# Force-Free and Autonomous Active Brownian Ratchets

CONSTANTIN REIN<sup>1</sup>, MARTIN KOLÁŘ<sup>2</sup>, KLAUS KROY<sup>1</sup> and VIKTOR HOLUBEC<sup>2</sup>

<sup>1</sup> *Leipzig University, Faculty of Physics and Earth Sciences, Institute for Theoretical Physics, Brüderstraße 16, 04081 Leipzig* [rein@itp.uni-leipzig.de](mailto:rein@itp.uni-leipzig.de) [klaus.kroy@uni-leipzig.de](mailto:klaus.kroy@uni-leipzig.de)

<sup>2</sup> *Charles University, Faculty of Mathematics and Physics, Department of Macromolecular Physics, V Holešovičkách 2, CZ-180 00 Praha* [viktor.holubec@mff.cuni.cz](mailto:viktor.holubec@mff.cuni.cz)

PACS **nn.mm.xx** – First pacs description

PACS **nn.mm.xx** – Second pacs description

PACS **nn.mm.xx** – Third pacs description

**Abstract** – Autonomous active Brownian ratchets rectify active Brownian particle motion solely by means of a spatially modulated but stationary activity, without external forces. We argue that such ratcheting requires at least a two-dimensional geometry. The underlying principle is similar to the ratcheting induced by steric obstacles in microswimmer baths: suitably polarized swimmers get channeled, while the others get trapped in low-activity regions until they loose direction. The maximum current is generally reached in the limit of large propulsion speeds, in which the rectification efficiency vanishes. Maximum efficiency is attained at intermediate activities and numerically found to be on the order of a few percent, for ratchets with simple wedge-shaped low-activity regions.

---

**Introduction.** – Brownian ratchets are subtle microscale transport devices operating out of equilibrium [1, 2]. They combine two effects that individually do not promote directed transport, namely unbiased Brownian motion and spatially periodic asymmetric environments, such that a net directed particle current is produced [3–5]. Conventional designs with passive particles usually break the spatial symmetry by imposing an asymmetric potential. The non-equilibrium element is often represented by a time-dependent driving mechanism that, by itself, does not introduce any directionality [4]. Typical examples comprise the rocking (or “flashing”) of the potential or the overall temperature [3, 4, 6, 7]. Ratchets brought out of equilibrium by more complex stationary, i.e., time-independent, temperature fields have also been investigated [8–12].

The self-propulsion of an active Brownian particle (ABP) represents yet another non-equilibrium mechanism that one ought to be able to exploit for ratcheting. While it does transiently break the spatial and temporal symmetry of equilibrium Brownian motion [13], it does not give rise to a net macroscopic current by itself. One would however expect that one of the simplest realizations of an active Brownian ratchet should consist of an ABP exposed to a spatially asymmetric (periodic) activity landscape. Yet, even though a number of ratchet designs with active parti-

cles have been discussed in the literature [14–24], none of them was based solely on a stationary activity landscape. Instead, some relied on ABPs placed in a soft potential in one spatial dimension [17, 19], or in asymmetric hard potentials in two-dimensions [14–16, 18, 24]. The asymmetric potentials, so typical of conventional ratchets, can be relinquished entirely, though, if one exploits the tendency of ABPs to polarize towards low-activity regions and accumulate there [23, 25–28]. The standard flashing potential can then be replaced by a dynamic activity landscape. Examples include propagating optical activation pulses that induce aligned or anti-aligned drifts, depending on the persistence length of the ABP motion relative to the pulse width and propagation speed [20–22]. In general, traveling activity waves induce traveling density and orientation waves of the ABPs, and can thus plainly be employed to sort ABPs, e.g., by size [23].

To sum up, ratcheting has been demonstrated for active particles in spatially asymmetric potential landscapes or in space-and-time dependent activity landscapes. However, no fundamental symmetry prevents ABPs from ratcheting also in *stationary* spatially asymmetric activity landscapes. In the following, we show that such ratchets are indeed realizable and explore the maximum current and rectification efficiency of a class of simple shapes, numerically.

**Model.** – We consider the motion of an ABP in a unit-square arena (thus taking its size as the natural length unit) with periodic boundary conditions in two dimensions (see Fig. 1). The state at time  $t$  is fully characterized by the position  $\mathbf{r}(t) = [x(t), y(t)]$  and polarization  $\mathbf{n}(t) = [\cos \theta(t), \sin \theta(t)]$  of the ABP. The translational and rotational Brownian motion are represented by mutually independent and unbiased ( $\langle \eta_i \rangle = 0$ ) Gaussian white noises  $\eta_i(t)$ ,  $i, j = x, y, \theta$ , of unit strength,  $\langle \eta_i(t) \eta_j(t') \rangle = \delta_{ij} \delta(t - t')$ , and diffusion constants  $D_t$  and  $D_r$ , respectively. The stationary activity landscape enters via a superimposed deterministic speed field  $v[x, y]$ . The dynamical equations for the ABP read

$$\dot{x} = v(x, y) \cos(\theta) + \sqrt{2D_t} \eta_x, \quad (1a)$$

$$\dot{y} = v(x, y) \sin(\theta) + \sqrt{2D_t} \eta_y, \quad (1b)$$

$$\dot{\theta} = \sqrt{2D_r} \eta_\theta. \quad (1c)$$

We only consider activity fields symmetric in the  $y$ -direction,  $v(x, 1/2 + y) = v(x, 1/2 - y)$ , so that  $\langle \dot{y}(t) \rangle = 0$  and the steady-state current is a scalar  $I = \langle \dot{x}(t) \rangle$ . Whenever  $I$  is nonzero, the device exhibits ratcheting.

A few general observations about the dynamics are gleaned directly from the above equations. First, the essential stochastic ingredient of the model is the rotational diffusion. If  $D_r$  is taken to infinity, the ABP motion loses its persistence. The model then reduces to a passive gas locally equilibrated at a spatially modulated (effective) temperature  $T = D_t + v^2/2D_r$ , with Boltzmann's constant and the friction coefficient set to unity. While such a gas can move thermophoretically in the presence of a temperature gradient, it cannot maintain a steady current in a periodic temperature profile. (We comment on the more subtle limit of a Knudsen gas [29, 30], at the end of the paper.) The ratcheting effect must thus entirely result from a clever combination of the more or less persistent motion in the high- and low-activity regions, respectively.

Below, we show that, in one spatial dimension, one cannot achieve autonomous ratcheting by any stationary activity landscape. For conceptual purposes, in two dimensions, it is sufficient to consider piecewise constant spatially periodic landscapes  $v(x, y)$ , varying discontinuously between the minimum and maximum values of velocity 0 and  $v$  (see Fig. 1 for an example). The dynamics is anyway low-pass filtered by the translational diffusion process so that any small-scale details and discontinuities in  $v(x, y)$  will thereby effectively be washed out. Setting the maximum value of velocity  $v$  to a very large (formally infinite) value amounts to the idealization of strictly ballistic dynamics in the high-activity (or simply “active”) regions. Similarly, retaining a non-vanishing  $D_t > 0$  to avoid an absorbing state, the minimum value of  $v(x, y)$  can safely be set to zero in the low-activity (or simply “passive”) regions, without much loss of generality. This choice, which shall be adopted for the remainder, simply amounts to purely diffusive dynamics, inside the passive region.

In summary, translational diffusion acts as a regularization for discontinuous activity profiles, so that the archetypal activity landscape discretely jumps between 0 and some finite or possibly even infinite value  $v$ . In the latter case, the active region is traversed in no time, so that, the total dwell time  $\tau$  of the particle in the unit cell is equal to the time spent in the passive region. The latter is independent of  $v$  and, at first sight, of  $D_r$ . However,  $D_r$  limits the “take-off” of ABPs emerging from the passive region, and in fact also the whole particle distribution at the active-passive boundary. For example, the ABP cannot take off if it emerges with a swim direction pointing back into the passive region. Also it can “tunnel” through narrow edges of the passive region. One therefore generally still expects the current  $I \simeq \tau^{-1}$  (in our unit length setup) and the dwell time  $\tau$  to depend on  $D_t$  and  $D_r$ , even if one takes  $v \rightarrow \infty$ , in the active region. It is however plausible that for a given geometric shape of the passive region, one can often find an optimum choice of  $D_t \propto D_r$ . Brownian dynamics simulations indeed corroborate this (Fig. 2), with a geometric prefactor compatible with  $(\varepsilon(1 - 2\delta_x)/2\pi)^2/2 \simeq 0.003$ , as expected from  $4D_t\tau \simeq (\varepsilon(1 - 2\delta_x)/2)^2$  and  $2D_r\tau \simeq \pi^2$ . The corresponding optimum dwell time  $\tau(D_t, D_r) \rightarrow \tau(D_r)$  is proportional to the ABP's mean reorientation time  $D_r^{-1}$ , implying  $I \simeq D_r$ , with a purely geometric prefactor (Fig. 3). The latter can only depend on dimensionless features of the shape (such as the parameters  $\delta$  and  $\varepsilon$  in Fig. 1).

These general considerations based on an infinite step function  $v(x, y)$  may not always be practically useful, from an active-matter perspective. For instance, an experimental realization of our idealized ABP might possibly only allow for a maximum speed  $v$ , below the asymptotic regime alluded to above (in which the dwell time in the arena equals the trapping time in the passive region). This will clearly reduce the ratchet current from its maximum value, and the dwell time will depend both on  $D_t \simeq D_r$  and the maximum attainable value of  $v$ . This “attenuated” transport regime, with  $D_t \simeq D_r \simeq v$  may be of particular practical interest, if the active speed of the ABP is regarded as a costly input. The most desirable *modus operandi* of the ratchet will then not anymore be that of maximum current  $I \simeq D_r$ , obtained in the limit  $v \rightarrow \infty$ , because the ratio  $I/v$  vanishes in this limit. Instead, one will then typically be interested in conditions that optimize this ratio, which can be interpreted as the rectification efficiency of the active ratchet, very much in the spirit of ABP engines and bacterial motors [15, 16, 24]. The interested practitioner will then generally have to find the corresponding optimum parameter values  $D_t$ ,  $D_r$ , and  $v$  for a given ratchet geometry, numerically.

The remainder of the paper is dedicated to a more comprehensive analysis of the above general considerations. In particular, we first clarify why stationary active Brownian ratchets can only be realized in at least two space dimensions. We also estimate realistic values of the maximum dimensionless current  $I(v \rightarrow \infty)/D_r$  and rectification

efficiency  $I/v$ , for a simple wedge geometry, depicted in Fig. 1.

**One-dimensional activity patterns.** – Already in one spatial dimension, spatially varying activity profiles accommodate non-intuitive effects. For example, the mean first passage time may depend non-monotonically on the distance from a target and the target finding probability can increase if the activity increases towards the target [31]. This seemingly contradicts the known fact that active particles spend less time in regions of higher activity. However, while the latter is a steady-state property, the former relates to transient behavior. In fact, when an ABP is oriented along an activity gradient, it accelerates and thus increases its chance to reach a target before it loses its orientation. Similarly, an ABP placed in the middle of a one-dimensional domain with a linear activity gradient reaches the high-activity end faster and more often than the low-activity end [32]. Although these effects look promising with regard to designing autonomous active Brownian ratchets, e.g., with a sawtooth-shaped stationary activity landscape, there is a catch. In the cited experiments [31, 32], the particle is placed back in its initial position upon reaching the target or the boundary of the arena. For a genuine ratchet, such “*deus-ex-machina*” type outside interventions are clearly not a permissible option.

More formally, one can demonstrate the absence of ratcheting in one-dimensional activity landscapes, as follows. Activity landscape can sort and locally accumulate ABPs according to their orientation, but they do not reorient them. Crucially, and quite in contrast to potential landscapes, activity landscapes do not exert any forces or torques on the ABPs, which are a crucial mechanism underlying the ratcheting of ABPs in one-dimensional potential landscapes [17]. As all orientations are thus equally probable in an unbiased ensemble, the spatially integrated total polarization must vanish. Together with the continuity equation for particle number conservation [33], this entails that the net current vanishes, too. More concretely, one may evoke the continuity of the local polarization profile as a function of position, which holds even for piecewise continuous activity profiles [26–28]. From this one concludes that, for a vanishing total polarization, there must be at least one position  $x_0$  in the polarization profile at which the time-averaged orientation vanishes. The time averaged current  $I$  at this point is given by the time-integral over  $v[x(t) = x_0] \cos \theta(t)$ . Up to a constant factor, this is just the vanishing time-averaged orientation. And since, in one spatial dimension, the continuity condition implies that the steady state current is spatially constant,  $I$  vanishes everywhere if it vanishes locally, at  $x_0$ . We have corroborated this conclusion by extensive Brownian dynamics simulations and by numerical solution of the Fokker–Planck equation, associated with Eq. (1), using the method of Ref. [34].

**Two-dimensional activity patterns.** – Compared to one-dimensional activity landscapes, the situation is

much different in two and higher-dimensional activity landscapes. The main reason is that the inevitable zeros of the polarization do now no longer constrain the overall current to vanish, unless they cover a whole vertical line  $(x_0, \{y\})$ . The latter is by no means required by the condition on an overall vanishing polarization. Around an isolated point of vanishing current, the resulting systematic flow field (or, equivalently, polarization field) takes the form of a vortex, as seen in Fig. 1. The sorting and accumulation of ABPs according to their orientation along the  $x$ -direction, which is already possible in one-dimensional activity landscapes [26–28], and exploited in non-stationary active Brownian ratchets [20–23], is now modulated along the second spatial direction  $y$ . A particle moving along the  $y$ -direction therefore experiences an effectively time-modulated activity pattern along the transport direction  $x$ , which has a similar rectifying effect as a dynamical one-dimensional activity profile.

The stationary but spatially periodically modulated activity-landscape  $v(x, y)$  shown in Fig. 1 provides a proof-of-principle example and serves as an instructive illustration of a working ratchet. It features a piece-wise constant activity field with a wedge-shaped passive region, where  $v(x, y) = 0$ , in an otherwise moderately active unit square with constant  $v(x, y) = D_r$ . The landscape is asymmetric along the  $x$ -direction and mirror-symmetric along the  $y$ -direction. The dimensionless numbers  $\delta_x$ ,  $\delta_y$ , and  $w = \varepsilon(1 - 2\delta_x)$ , with  $\varepsilon \in [0, 1]$ , denote the distances of the edges from the periodic boundaries and the width of the wedge along its mirror-symmetry axis, respectively. The extreme geometries correspond to an infinitely thin passive region ( $\varepsilon = 0$ ) and a convex, triangular passive region ( $\varepsilon = 1$ ). Both yield sub-optimal ratchets.

While even this simple wedge model is not exactly solvable, its performance can qualitatively be understood, using simple physical arguments. First, the above-mentioned saturation of the ratchet current for infinite speed  $v \rightarrow \infty$  in the active region is simply due to the fact that the time spent by the ABP in the active region becomes negligible compared to the time  $\tau$  spent diffusing in the passive region. This limit is thus amenable to event-driven simulations. Below, we go one step further and exploit it to construct a simplified geometric toy model that can provide semi-analytical estimates for the ratcheting current. Unfortunately, as already pointed out above, the conceptually convenient large-speed limit is somewhat academic. The practitioner will be interested in more affordable, finite values of  $v$ . Therefore, one should also consider the rectification efficiency  $I/v$ , which is the current produced by the ratchet relative to that of a perfectly polarized ABP.

To understand the pertinence of the limits of infinite or vanishing diffusivities  $D_r$ ,  $D_t$ , recall that ratcheting is all about the geometric rectification of stochastic motion. In the limit  $D_r \rightarrow 0$  (perfect persistence), the initial orientation is however entirely conserved, while the limit  $D_r \rightarrow \infty$  (vanishing persistence) corresponds to thermophoresis within an effective temperature field. So both

limits do not correspond to genuine active ratcheting. Similarly, passive regions, with vanishing speed  $v = 0$ , would all become absorbing for  $D_t \rightarrow 0$ , while in active regions with a finite  $v < \infty$ ,  $D_t \rightarrow \infty$  would wipe out the persistent active motion. Again, both limits are irrelevant for the discussion of active ratcheting. And even though one could set  $D_t = 0$  without creating an absorbing state if a non-vanishing speed  $v > 0$  was maintained in the passive (or less active) region, this choice would be unnatural, as it requires passive regions with vanishing (or even “small”)  $v$  to be administratively forbidden. On the other hand, allowing for some finite  $D_t \lesssim D_r$  is not very consequential for the transport in the (more) active regions, where it merely partially degrades the persistence induced by the activity. This exposes  $D_t$  as a parameter of minor physical relevance except for its regularizing role in the passive regions. There are however two more reasons for including a non-vanishing  $D_t$ , in the discussion. Firstly, it will actually matter for the comparison to practical physical realizations of an ABP ratchet. And secondly, it also serves to regularize some fine-grained details of the ratchet geometry, thereby putting a limit on an otherwise potentially limitless ornamentation of the ratchet design that would in practice have to be cut off by a physical particle radius. In contrast to the indispensable finite rotational diffusivity  $D_r$ , the translational diffusivity  $D_t$  thus plays a rather technical role, as a model regularization parameter.

In conclusion, a pertinent discussion of a stationary ABP ratchet in two dimensions is best conducted for finite diffusivities  $D_r$  and  $D_t$ . While  $D_t^{-1}$  may at first suggest itself as the natural time unit of the ratchet (its dwell time), it turns out that its physical impact can, for a conceptual analysis, effectively be taken largely out of the game. The trick is to set it to an optimum value that maximizes the rectification efficiency  $I/v$ . Our numerical analysis (see Fig. 2) confirms the expectation that this “best” value is unique and on the order of  $D_r$ , for the simple geometry shown in Fig. 1. Its physical origin may be understood from the role played by  $D_t$  for controlling the ABP’s escape time from the passive region. As already pointed out, above, if  $D_t \gg D_r$ , the ABP will not have lost its polarization when it leaves the passive region, and therefore typically swim right back into it, unless that region is narrow enough to be traversed with a substantial (“tunneling”) probability. Additionally, the dominance of translational diffusion for  $D_t \gg D_r$  will unduly degrade the persistence in the active region beyond the inevitable minimum, set by  $D_r$ . In contrast, if  $D_t \ll D_r$ , the regularizing effect of the translational diffusion onto the absorbing state may become less than optimal, as the initial particle polarization will then have been lost long before the ABP reemerges from the passive region. Altogether, this suggests an optimum value of  $D_t$  on the order of  $D_r$ , as indeed numerically confirmed in Fig. 2.

To summarize, the natural length unit of the stationary active ratchet is set by the domain size, its natural time unit by the inverse rotational diffusion coefficient  $D_r^{-1}$ . And it is

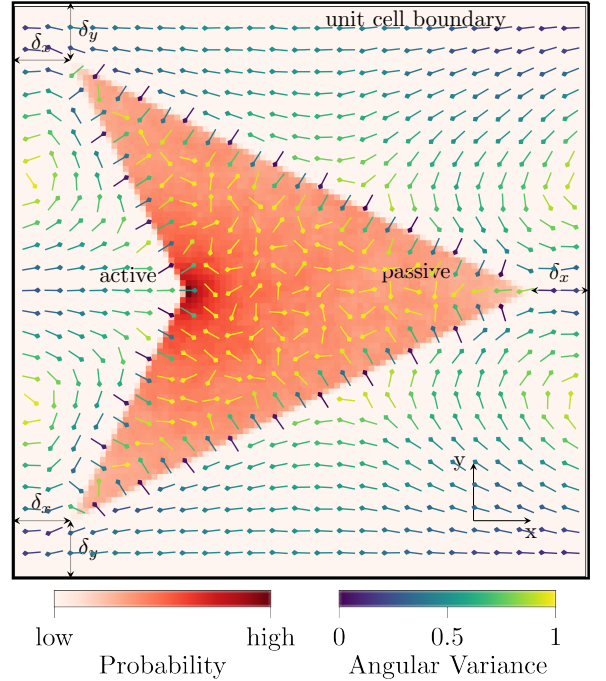


Fig. 1: Unit cell of a (unit width) two-dimensional square ratchet with  $\delta_x = \delta_y = 0.1$ ,  $\varepsilon = 0.75$ ,  $v = D_r$ , and  $D_t = 10^{-4}D_r$ . The background color encodes the probability density for the position of the ABP that predominantly dwells in the wedge-shaped passive region. Arrows show the mean orientation  $\langle \mathbf{n} \rangle$  of the ABP obtained from Brownian dynamics simulations, colors coding for the angular variance  $1 - (\langle n_x \rangle^2 + \langle n_y \rangle^2)^{1/2}$ ; small values indicate strong alignment and  $O(1)$ -values a random orientation.

conceptually convenient (if not generally highly advisable) to work with an optimized translational diffusivity  $D_t \simeq D_r$  of comparable magnitude. The natural scale for the maximum ratchet current  $I \simeq \tau^{-1} \simeq D_r$  is then  $D_r$  itself, while that of the natural efficiency  $I/v$  is  $(\tau v)^{-1} \simeq D_r/v$ . In practice, both quantities may be expected to be somewhat reduced by a dimensionless geometrical shape factor. The crucial message is then that determining the optimum current  $I/D_r$  and efficiency  $I/v$  boils down to an infinite dimensional geometric optimization problem intertwined with the “thermodynamic” optimizations of the parameters  $D_t$  and  $D_t, v/D_r$ , respectively.

**Numerical study.** – To provide a specific but instructive example, Fig. 1 illustrates the working principle of the active Brownian ratchet and its polarization field  $\langle \mathbf{n} \rangle$  for a wedge-shaped passive region in the unit square, with periodic boundary conditions. As already alluded to above, the orientation field is indeed seen to form vortices around the points with vanishing average orientation, which help to defy the no-go theorem for one-dimensional active ratchets. To create the figure, we solved Eq. (1) by a Brownian dynamics simulation with time-step  $dt = 10^{-4}/v$ . The central observable is the ratchet current  $I = x(T)/T$ , eval-

uated as the final traversed  $x$ -distance of the ABP divided by the total simulation time  $T = 10^7/v$ . We checked that the vertical current  $y(T)/T$  in the  $y$ -direction vanishes, as expected. As demonstrated in Refs. [26–28], along the active–passive boundary, the ABP points on average towards the passive region. This may seem surprising, since it seems to imply a net particle influx into the passive region. It is an illusion, however, since the swim pressure acting onto an active–passive boundary is not exerted across it [26]. Actually, the particle can therefore “escape” from the passive region, against this swim pressure. If it escapes along the tip-side (right in Fig. 1), it likely ends up in the indented concave part of the passive region (left in Fig. 1). On the other hand, if the ABP escapes in the vertical direction towards the horizontal active channels of width  $2\delta_y$  (top and bottom in Fig. 1), it can generate a net current from right to left. As a result, the passive region blocks particle paths to the right more than those to the left. Remarkably, active Brownian ratchets relying on potential forces acting like hard walls [14, 16, 18, 24, 35, 36] are based on the very same principle. The important difference here is that our setup does not involve any potential forces, and the ABP can thus freely pass back and forth between the passive and active region. With hard walls, the ABP would slide along the wedge until it gets trapped in the pocket or escapes into the channel, thereby generating a net ratchet current. In our force-free active ratchet, the sliding motion is replaced by the diffusive spreading inside the passive region.

For the setup illustrated in Fig. 1, we also investigated the rectification efficiency  $I/v$  for finite activity,  $v < \infty$ , as a function of the diffusivities  $D_r$  and  $D_t$ . In accord with our foregoing qualitative considerations, the numerical results shown in Fig. 2 feature a maximum around  $I/v \sim 0.014$  for  $D_r \sim 0.3v$  and  $D_t \sim 0.001v$ . These optimum values are specific for the chosen geometry and cannot be found without performing the numerical simulation.

A more challenging task is to find the most efficient ratchet geometry. Here, we restrict this infinite dimensional optimization problem to the class of wedge or arrowhead shapes illustrated in Fig. 1. We ask for the optimum depth of the concave indentation, which is parametrized by  $\varepsilon$ . For shallow indentations, the ABP spends more time in the passive region as needed to loose its polarization. This reduces the current and the rectification efficiency compared to a design with a stronger indentation. However, for very deep indentations, the passive region becomes too narrow to allow for a substantial reorientation of the traversing ABP, and the corresponding “tunneling” of the polarization eventually nullifies the ratcheting effect ( $I \propto \varepsilon \rightarrow 0$ ). In other words, there is necessarily a non-monotonic dependence of the rectification efficiency on  $\varepsilon$ . As illustrated in Fig. 3, this implies that the intermediate optimum value of  $\varepsilon$ , once again, needs to be found numerically. This result also nicely demonstrates the difference between our force-free active ratchet and its siblings operating with potential forces. In particular, for ratchets with hard walls around

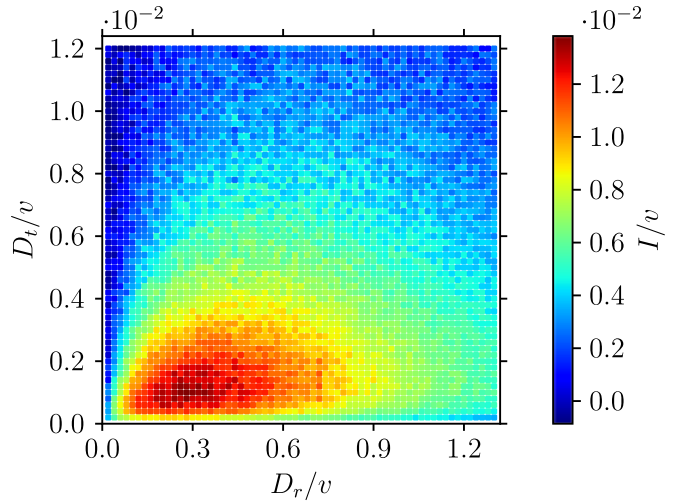


Fig. 2: Rectification efficiency  $I/v$  as function of the inverse Péclet numbers  $D_r/v$  and  $D_t/v$ , for the active Brownian ratchet with the same geometry as depicted in Fig. 1.

an exclusion zone of the same shape as our passive region, the ratcheting would always be maintained, regardless of the wall thickness. The figure also demonstrates that the non-monotonic dependence of the rectification strength on the indentation depth is robust against the fine tuning of the diffusivities, and that the optimization depends on the interplay between the geometry and the inverse Péclet numbers  $D_t/v$  and  $D_r/v$ .

Beyond the indentation depth, one can also consider the effect of the parameter  $\delta_y$  for the lateral width of the horizontal active channels. The current decreases both as  $\delta_y \rightarrow 0$ , when the channel width vanishes, and for  $\delta_y \rightarrow 1/2$ , when the passive volume becomes marginal relative to the overall domain width. Similarly, as for  $\varepsilon$ ,  $D_r/v$  and  $D_t/v$ , the rectification efficiency  $I/v$  thus also exhibits a maximum as a function of  $\delta_y$ . Finally, the remaining parameter  $\delta_x$  measures the overall width of the passive region in the  $x$ -direction. When  $\delta_x \rightarrow 1/2$ , the width of the passive region vanishes, and therefore also the current  $I$ , similarly as for  $\varepsilon \rightarrow 0$ . On the other hand, the current monotonically increases with decreasing  $\delta_x \rightarrow 0$ , until the passive region spans across the whole domain. This reveals that  $\delta_x$  is a non-essential parameter that can be set to 0 for conceptual purposes. Together, the shape parameters  $\delta_x$ ,  $\delta_y$ , and  $\varepsilon$  control how pointed and asymmetric the passive region may become. Generally speaking,  $I/v$  grows with increasing asymmetry.

**Geometric toy model.** – A more mechanistic insight into the effects of the ratchet geometry on the current can be obtained from a schematic, purely geometrical toy model. It is defined by the idealized rules that the particle moves with infinite speed  $v \rightarrow \infty$  in the active region and rotates and spreads sufficiently fast throughout the passive region to emerge from its surface with uniform

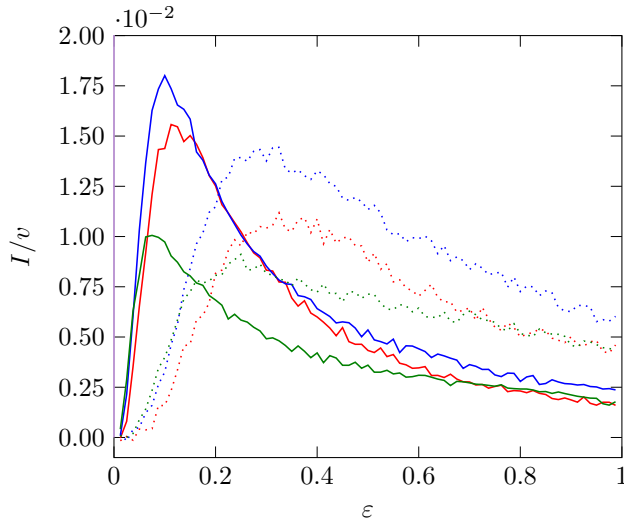


Fig. 3: Rectification efficiency  $I/v$  as a function of indentation depth, parametrized by  $1 - \varepsilon$  ( $\delta_x, \delta_y = 0.1$ ). Various combinations of  $D_r$  and  $D_t$  are shown, with colors coding for the value of  $D_r/v$ : 0.1 (red), 0.3 (blue) and 1 (green), and line style for  $D_t/v$ :  $10^{-4}$  (solid) and  $10^{-3}$  (dotted).

spatial and orientational distributions, after a dwell time  $\tau$ . The ensuing simplifications enable us to bypass the computationally expensive Brownian dynamics simulations for qualitative estimates. The path of the ABP in the active region is then uniquely determined by the ratchet geometry alone. Once the ABP leaves the passive region with randomized orientation and position, it immediately hits either another part of the same passive region or one of its periodic images, as sketched in Fig. 4.

One can therefore evaluate the probabilities  $P_{\leftarrow}$ ,  $P_{\rightarrow}$ , and  $P_{\downarrow}$  that the ABP leaving the passive region travels to the left, right, or merely vertically, respectively. If the dwell time  $\tau$  is approximated by the average reorientation time  $\tau = D_r^{-1}$  of the ABP, as would be the case for an optimum choice of  $D_t$ , one estimates the current as

$$I = D_r(P_{\leftarrow} - P_{\rightarrow}). \quad (2)$$

The resulting probabilities are shown in Fig. 4 as functions of the dimensionless horizontal width  $\varepsilon$  of the symmetry axis of the wedge-shaped passive region. One sees that  $P_{\leftarrow} > P_{\rightarrow}$  for all values of  $\varepsilon$ , so that the model always predicts a leftward current  $I$  that is numerically roughly comparable to the optimum currents  $I \simeq D_r$  obtained from the Brownian dynamics simulations. It naturally overestimates the current for extreme values of  $\varepsilon$ , corresponding to concave and vanishing passive volumes, respectively. The actual reorientation of the ABP is then much less efficient than assumed by the stylized model, so that the comparison further corroborates the primary role played by the optimized destruction of the particle polarization in the passive region, for the rectification efficiency of the ratchet.

An iterative evaluation of the toy model provides further

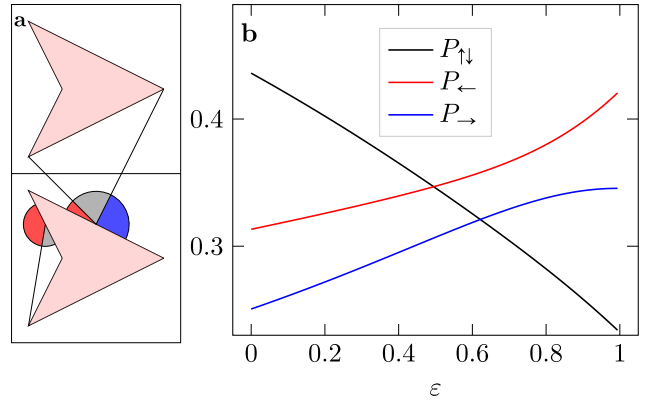


Fig. 4: Geometric toy model for the ratchet of Fig. 1. **a**: ABPs are emitted with random orientation from random positions on the circumference of the passive region. Those traveling to the left (red disk sectors) or right (blue disk sector) contribute to the horizontal current (2). The remaining ones contribute to the vanishing current in the vertical direction. **b**: The probabilities for transitions depicted in **a**.

insight into the role played by the active channels separating the passive image regions. One can find the distribution of positions where the uniformly distributed ABP ensemble leaving the active-passive boundary will become trapped on the boundary again. The resulting position distribution can be used as the initial condition for the next step, again assuming uniformly distributed orientations, for simplicity. After many iterations of this procedure, the position distribution no longer changes and one can consider it as an approximate stationary position distribution of the ABP. The resulting stationary distribution is similar to that obtained from the Brownian dynamics simulations, depicted in Fig. 1. It exhibits a maximum in the indentation pocket of the passive region and, for  $\delta_x = \delta_y = 0$ , also at the reverse indentations connecting the passive region with its periodic images. These particle accumulations would leak out into the horizontal active channels to constitute the ratchet current, for any  $\delta_y > 0$ .

Let us finally return to the similarities and differences between our toy model and gases in similar geometries. Dense gases or fluids, in which frequent mutual particle collisions can be relied on for establishing local equilibrium, should not exhibit ratcheting in spatially periodic setups, like ours. But in so-called rarefied or Knudsen [29] gases, without an efficient local equilibration mechanism, particles move ballistically in the space between boundaries, similarly as ABPs in the active region of our geometric toy model, so that the analysis of transport largely boils down to the problem of boundary conditions. This is then a more subtle issue [30, 37] that would deserve further study.

**Conclusion.** — Spatially inhomogeneous activity profiles can be used to sort active Brownian particles according to their orientations [26–28]. In one spatial dimension, the requirements for the overall system’s polarization to vanish,

together with particle conservation, prevent ratcheting in time-constant spatially-periodic activity landscapes. In two and more dimensions, such active ratcheting is possible. We analyzed a proof-of-principle realization of a wedge-shaped two-dimensional autonomous force-free active Brownian ratchet. It demonstrates that active ratcheting does not require a dynamic activity profile nor help from potential forces or walls.

Our study can be generalized in several ways. For example, it seems worthwhile to find out whether the wedge-shaped ratchet design maximizes the current or can be surpassed by more optimized geometries. One can also study how the ratcheting current would change in activity profiles moving with a constant velocity or in the presence of an external force interfering with the ratcheting. Another potentially interesting extension could be to ABPs with translational and/or orientational inertia [38]. And, eventually, it would be intriguing if the ratcheting currents in rarefied gases, hinted at by our toy model, could be experimentally demonstrated.

\*\*\*

**Acknowledgements.** – We acknowledge financial support from the pre-doc award program at Leipzig University, as well as from the Czech Science Foundation (project No. 20-02955J) and Charles University (project PRIMUS/22/SCI/009).

\*. – References

- <sup>1</sup>C. S. Peskin, G. M. Odell, and G. F. Oster, *Biophys. J.* **65**, 316–324 (1993).
- <sup>2</sup>S. I. Reed, *Nat. Rev. Mol. Cell Biol* **4**, 855–864 (2003).
- <sup>3</sup>R. D. Astumian and M. Bier, *Phys. Rev. Lett.* **72**, 1766–1769 (1994).
- <sup>4</sup>P. Reimann, *Phys. Rep.* **361**, 57–265 (2002).
- <sup>5</sup>J. M. R. Parrondo and B. J. de Cisneros, *Appl. Phys. A* **75**, 179–191 (2002).
- <sup>6</sup>M. O. Magnasco, *Phys. Rev. Lett.* **71**, 1477–1481 (1993).
- <sup>7</sup>J. Rousselet, L. Salome, A. Ajdari, and J. Prost, *Nature* **370**, 446–447 (1994).
- <sup>8</sup>M. v. Smoluchowski, *Physik. Zeitschr.* **13** (1912).
- <sup>9</sup>R. P. Feynman, R. B. Leighton, and M. Sands, Vol. 1 (Addison Wesley, Reading MA, 1963) Chap. 46.
- <sup>10</sup>M. Büttiker, *Zeitschrift für Physik B Condensed Matter* **68**, 161–167 (1987).
- <sup>11</sup>R. Landauer, *J. Stat. Phys.* **53**, 233–248 (1988).
- <sup>12</sup>A. Ryabov, V. Holubec, M. H. Yaghoubi, M. Varga, M. E. Foulaadvand, and P. Chvosta, *J. Stat. Mech. Theory Exp.* **2016**, 093202 (2016).
- <sup>13</sup>G. Falasco, R. Pfaller, A. P. Bregulla, F. Cichos, and K. Kroy, *Phys. Rev. E* **94**, 030602 (2016).
- <sup>14</sup>L. Angelani, R. Di Leonardo, and G. Ruocco, *Phys. Rev. Lett.* **102**, 048104 (2009).
- <sup>15</sup>R. D. Leonardo, L. Angelani, D. Dell’Arciprete, G. Ruocco, V. Iebba, S. Schippa, M. P. Conte, F. Mecarini, F. D. Angelis, and E. D. Fabrizio, *PNAS* **107**, 9541–9545 (2010).
- <sup>16</sup>L. Angelani and R. D. Leonardo, *New J. Phys.* **12**, 113017 (2010).
- <sup>17</sup>L. Angelani, A. Costanzo, and R. D. Leonardo, *EPL* **96**, 68002 (2011).
- <sup>18</sup>P. K. Ghosh, V. R. Misko, F. Marchesoni, and F. Nori, *Phys. Rev. Lett.* **110**, 268301 (2013).
- <sup>19</sup>B.-Q. Ai and F.-G. Li, **13**, 2536–2542 (2017).
- <sup>20</sup>A. Geiseler, P. Hänggi, and F. Marchesoni, *Entropy* **19** (2017).
- <sup>21</sup>A. Geiseler, P. Hänggi, and F. Marchesoni, *Sci. Rep.* **7**, 1–9 (2017).
- <sup>22</sup>A. Geiseler, P. Hänggi, F. Marchesoni, C. Mulhern, and S. Savel’ev, *Phys. Rev. E* **94**, 012613 (2016).
- <sup>23</sup>H. Merlitz, H. D. Vuijk, J. Brader, A. Sharma, and J.-U. Sommer, *J. Chem. Phys.* **148**, 194116 (2018).
- <sup>24</sup>P. Pietzonka, É. Fodor, C. Lohrmann, M. E. Cates, and U. Seifert, *Phys. Rev. X* **9**, 041032 (2019).
- <sup>25</sup>A. Sharma and J. M. Brader, *Phys. Rev. E* **96**, 032604 (2017).
- <sup>26</sup>S. Auschra, V. Holubec, N. A. Söker, F. Cichos, and K. Kroy, *Phys. Rev. E* **103**, 062601 (2021).
- <sup>27</sup>S. Auschra and V. Holubec, *Phys. Rev. E* **103**, 062604 (2021).
- <sup>28</sup>N. A. Söker, S. Auschra, V. Holubec, K. Kroy, and F. Cichos, *Phys. Rev. Lett.* **126**, 228001 (2021).
- <sup>29</sup>M. Knudsen, *Annalen der Physik* **336**, 205–229 (1909).
- <sup>30</sup>W. Steckelmacher, *Rep. Prog. Phys.* **49**, 1083 (1986).
- <sup>31</sup>H. D. Vuijk, A. Sharma, D. Mondal, J.-U. Sommer, and H. Merlitz, *Phys. Rev. E* **97**, 042612 (2018).
- <sup>32</sup>P. K. Ghosh, Y. Li, F. Marchesoni, and F. Nori, *Phys. Rev. E* **92** (2015).
- <sup>33</sup>S. Hermann and M. Schmidt, *Phys. Rev. Res.* **2**, 022003 (2020).
- <sup>34</sup>V. Holubec, K. Kroy, and S. Steffenoni, *Phys. Rev. E* **99**, 032117 (2019).
- <sup>35</sup>S. Hulme, W. DiLuzio, S. Shevkopyas, L. Turner, M. Mayer, H. Berg, and G. Whitesides, *Lab chip* **8**, 1888–95 (2008).
- <sup>36</sup>I. Berdakin, A. Silhanek, H. M. Cortéz, V. Marconi, and C. Condat, *Open Physics* **11**, 1653–1661 (2013).
- <sup>37</sup>Y. Qiao and Z. Shang, *Phys. A: Stat. Mech. Appl.* **596**, 127105 (2022).
- <sup>38</sup>H. Löwen, *J. Chem. Phys.* **152**, 040901 (2020).





PAPER • OPEN ACCESS

## Equilibrium stochastic delay processes

To cite this article: Viktor Holubec *et al* 2022 *New J. Phys.* **24** 023021

View the [article online](#) for updates and enhancements.

You may also like

- [Optimal Control Of Dynamic IS-LM Business Cycle Model With Two Time Delay](#)

Airin Nur Hidayati, Erna Apriliani and I Gst Ngr Rai Usadha

- [A Fast Algorithm of Direct Position Determination Using TDOA and FDOA](#)

Yuanyuan Song, Caiyong Hao, Mingbing Li et al.

- [Local Stability of AIDS Epidemic Model Through Treatment and Vertical Transmission with Time Delay](#)

Cascarilla Novi W and Dwi Lestari



## PAPER

## Equilibrium stochastic delay processes

## OPEN ACCESS

RECEIVED  
26 August 2021REVISED  
8 January 2022ACCEPTED FOR PUBLICATION  
14 January 2022PUBLISHED  
16 February 2022Viktor Holubec<sup>1,2,\*</sup> , Artem Ryabov<sup>1</sup> , Sarah A M Loos<sup>2,3</sup>  and Klaus Kroy<sup>2</sup> <sup>1</sup> Department of Macromolecular Physics, Faculty of Mathematics and Physics, Charles University, V Holešovičkách 2, CZ-180 00 Praha, Czech Republic<sup>2</sup> Institut für Theoretische Physik, Universität Leipzig, Postfach 100 920, D-04009 Leipzig, Germany<sup>3</sup> International Centre for Theoretical Physics, Str. Costiera 11, 34151 Trieste, Italy

\* Author to whom any correspondence should be addressed.

E-mail: [viktor.holubec@mff.cuni.cz](mailto:viktor.holubec@mff.cuni.cz)**Keywords:** time delay, stochastic processes, nonlinear, exact solutions, stochastic thermodynamics, fluctuation theorems, feedbackOriginal content from  
this work may be used  
under the terms of the  
[Creative Commons  
Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/).Any further distribution  
of this work must  
maintain attribution to  
the author(s) and the  
title of the work, journal  
citation and DOI.**Abstract**

Stochastic processes with temporal delay play an important role in science and engineering whenever finite speeds of signal transmission and processing occur. However, an exact mathematical analysis of their dynamics and thermodynamics is available for linear models only. We introduce a class of stochastic delay processes with nonlinear time-local forces and linear time-delayed forces that obey fluctuation theorems and converge to a Boltzmann equilibrium at long times. From the point of view of control theory, such ‘equilibrium stochastic delay processes’ are stable and energetically passive, by construction. Computationally, they provide diverse exact constraints on general nonlinear stochastic delay problems and can, in various situations, serve as a starting point for their perturbative analysis. Physically, they admit an interpretation in terms of an underdamped Brownian particle that is either subjected to a time-local force in a non-Markovian thermal bath or to a delayed feedback force in a Markovian thermal bath. We illustrate these properties numerically for a setup familiar from feedback cooling and point out experimental implications.

**1. Introduction**

Consider the stochastic delay differential equations (SDDEs)

$$\dot{x}(t) = v(t), \quad (1)$$

$$m\dot{v}(t) = F(t) + F_D(t - \tau) + \eta(t) \quad (2)$$

with a nonlinear time-local force  $F(t) = F(x, v, t)$  and a linear delay force ( $\tau > 0$ )

$$F_D(t - \tau) = -\kappa_\tau x(t - \tau) - \gamma_\tau v(t - \tau), \quad (3)$$

with constant coefficients  $\kappa_\tau$  and  $\gamma_\tau$ . The dynamics is randomly driven by a possibly non-Markovian, zero-mean Gaussian stochastic noise  $\eta(t)$ . Intuitively, one can think of equations (1) and (2) as describing the time evolution of the position  $x(t)$  and velocity  $v(t)$  of a Brownian particle with mass  $m$  and driven by the combined forces  $\eta$ ,  $F$ , and  $F_D$ . These forces can arise from various origins, e.g. from the environment and the experimental apparatus, including some specifically tailored feedback mechanisms. Further specifications and various interpretations will be provided below. Due to finite speeds of information transfer and processing and elements with slow response, such equations are ubiquitous in engineering [1], biology [2–4] and even economics [5–8]. Most frequently, they are applied in modelling of feedback loops [9–26], neural networks [27–30], population dynamics [31, 32], and epidemiology [33, 34].

Rising interest in SDDEs among physicists [35] is driven by recent experiments. In the so-called feedback cooling experiments with Brownian particles, one employs a feedback of the particle’s past velocity to achieve a more localised state [36–39]. In the surging field of active matter [40–42], inevitable time delays in the control of robotic swarms [18] led to investigations of the stability and localization of

many-body systems with delayed interactions [18, 21, 22, 25, 43]. In agreement with engineering practice [1, 10, 11, 44], it was found that delay generally introduces instabilities and oscillations into the dynamics [22, 43] and increases stability and localization only in special cases [41].

Similarly, inevitable instrumental and feedback delays in micro-manipulation experiments [22, 45] used to test stochastic thermodynamics [46, 47] has triggered investigation of the thermodynamic aspects of SDDs [13, 48–51]. There are interesting consequences of the acausality of time-reversed processes in delay systems due to the tracking (future) history for the time-reversal. If interpreted as feedback-driven systems with information inflow, their total entropy production rate,  $\dot{S}_{\text{tot}}$ , evaluated as a ratio of forward to backward path probabilities, is not just the sum,  $\dot{S}_S + \dot{S}_{\text{NM}}$ , of entropy fluxes into the system (S) and into the bath (B) [48, 49, 51, 52]. This means that the second law  $\dot{S}_{\text{tot}} \geq 0$  does not imply positivity of  $\dot{S}_S + \dot{S}_{\text{NM}}$ . These results are generic for the system (1) and (2) with a Gaussian white noise  $\eta(t) \propto \xi(t)$ ,  $\langle \xi(t) \rangle = 0$ ,  $\langle \xi(t)\xi(t') \rangle = \delta(t - t')$ . However, explicit expressions are currently only available for linear systems [48, 49, 51–53], which fail to describe a broad range of interesting effects observed in presence of nonlinear forces [48, 49, 51]. The same can be said about the probability densities for SDDs. They are available only for simple linear setups [43, 54–63], and nonlinear systems have been treated by various approximate techniques [18, 43, 60, 61, 64–66].

Even without time delay, an exact treatment of nonlinear systems is indeed difficult. However, their stationary and relaxation properties are known exactly in thermodynamic equilibrium. In this work, we extend this property to a certain class of SDDs. Our results can be of interest not only to the theory of delay processes but also in applied contexts, like in control theory.

## 2. Main results

As our main result, we identify a class of nonlinear delay processes that admit a standard thermodynamic description, including the second law inequality  $\dot{S}_{\text{tot}} = \dot{S}_S + \dot{S}_{\text{NM}} \geq 0$ . If not driven, they obey Boltzmann statistics in the steady state. We therefore characterize these processes as ‘equilibrium delay processes’. The key idea is to accompany the time-delayed feedback force applied to the system with a suitable colored noise  $\eta_{\text{FB}}$  and interpret the resulting overall system as a particle immersed in an equilibrium reservoir and controlled by time-local external forces. We further point out how to interpret equations (1) and (2) as a feedback-driven system and how to apply our results therein. Altogether, we provide three complementary interpretations for the same stochastic process: a special type of system with time-delayed forces (section 1), system with time-local forces and a heat bath with memory (section 2.1), and a feedback-driven system (section 2.3). They differ just in the interpretation of the individual forces on the right-hand side of equation (2). In table 1, we summarize relations between the three interpretations and the definitions of the corresponding forces. While the purpose of this paper is to primarily discuss equilibrium delay processes theoretically, appendix B offers some suggestions regarding possible experimental realizations for Markovian thermal reservoirs (described by white noise) using state-of-the-art experimental setups similar to those in [39, 67–69]. In the following, we take Boltzmann’s constant  $k_B$  as our unit of entropy.

### 2.1. Mapping to time-local control and non-Markovian heat bath (table 1(B))

In this section, we describe the reinterpretation of the delay system (table 1(A)) as an equilibrium system with memory (table 1(B)). Consider a delay system described by equations (1) and (2) with the time-local force

$$F(t) = F_E(x, v, t) + \kappa_\tau x(t) - \gamma_0 v(t). \quad (4)$$

The specific form of the terms proportional to the constants  $\gamma_0 > 0$  and  $\kappa_\tau$  facilitates the reinterpretation of the delay force  $F_D$  in equation (2) as part of a friction force, below. The remaining force in equation (4),

$$F_E(x, v, t) = -\partial_x U(x, t) + F_N(x, v, t) \quad (5)$$

is an arbitrary time-local force applied by external agents. It is composed of potential and non-potential components  $-\partial_x U(x, t)$  and  $F_N(x, v, t)$ . To distinguish the force  $F_E$  from the force applied via the feedback loop in the feedback interpretation of equations (1) and (2) (table 1(C)), we call it the ‘ordinary’ external force.

Equation (2) now assumes the form

$$m\dot{v}(t) = F_E(t) + F_F(t) + \eta(t) \quad (6)$$

with  $F_F(t) \equiv \kappa_\tau[x(t) - x(t - \tau)] - \gamma_0 v(t) - \gamma_\tau v(t - \tau)$ . It resembles the dynamical equation for the velocity of a particle subjected to an external force  $F_E$  and immersed in a viscoelastic solvent exerting on the

**Table 1.** Three interpretations of the delay Langevin equation (2) employed in this paper and the corresponding forces. The forces in the three interpretations yield the same change in the momentum  $m\dot{v}$  and thus the same stochastic process. The term  $\kappa_\tau x(t) - \gamma_0 v(t)$  in the force  $F$  in the interpretation A is introduced to facilitate the reinterpretation of the delayed force  $F_D$  as part of the friction force  $F_F$  in B. By ‘ordinary external forces’ in A we mean time-local forces arising from physical interactions and thus not applied via a feedback loop.

A. Delay system $m\dot{v} = F + F_D + \eta$
<i>Time-local systematic force:</i> $F = F_E + \kappa_\tau x(t) - \gamma_0 v(t)$
Ordinary (non-feedback) external force: $F_E = -\partial_x U(x, t) + F_N(x, v, t)$
Potential component of $F_E$ : $-\partial_x U(x, t)$
Non-potential component of $F_E$ : $F_N(x, v, t)$
<i>Time-delayed force:</i> $F_D = -\kappa_\tau x(t - \tau) - \gamma_\tau v(t - \tau)$
<i>Total coloured noise from the environment and experimental apparatus:</i> $\eta(t)$
B. System with non-Markovian heat bath and time-local control $m\dot{v} = F_E + F_F + \eta$
<i>Time-local external force:</i> $F_E$
<i>Time-delayed non-Markovian bath friction:</i> $F_F = \kappa_\tau x(t) - \gamma_0 v(t) - \kappa_\tau x(t - \tau) - \gamma_\tau v(t - \tau) = \kappa_\tau x(t) - \gamma_0 v(t) + F_D$
Total force from the non-Markovian heat bath at temperature $T$ : $F_F + \eta(t)$
<i>Coloured noise from the non-Markovian bath:</i> $\eta(t)$
Heat flux into the system from the non-Markovian bath: $\dot{Q}_{NM} = \langle (F_F + \eta)\dot{x} \rangle$
C. Feedback-driven system with Markovian heat bath $m\dot{v} = F_E + F_{FB} - \gamma_0 v(t) + \sqrt{2T_0\gamma_0}\xi(t)$
<i>Time-local external force:</i> $F_E$
<i>Feedback force:</i> $F_{FB} = \kappa_\tau x(t) - \kappa_\tau x(t - \tau) - \gamma_\tau v(t - \tau) = \kappa_\tau x(t) + F_D + \eta_{FB}(t) = F_F + \gamma_0 v(t)$
Non-Markovian noise exerted by the feedback loop: $\eta_{FB}(t) = \eta(t) - \sqrt{2T_0\gamma_0}\xi(t)$
Total force from the Markovian bath at temperature $T_0$ : $-\gamma_0 v(t) + \sqrt{2T_0\gamma_0}\xi(t)$
<i>Time-local friction from the Markovian bath:</i> $-\gamma_0 v(t)$
<i>White noise from the Markovian bath:</i> $\sqrt{2T_0\gamma_0}\xi(t)$
Heat flux into the system from the Markovian bath: $\dot{Q}_M = \langle (-\gamma_0 v + \eta)\dot{x} \rangle$

particle the overall force  $F_F + \eta$  with systematic component (friction)  $F_F$ , and stochastic component (noise)  $\eta$ .

Such noise and friction can be interpreted to arise from an ordinary equilibrium heat bath, i.e. a many-body system with infinite heat capacity in thermal equilibrium, with a somewhat peculiar memory that gives rise to an ‘echo’ in the noise and friction (table 1(B)). Notably, for an equilibrium heat bath with a friction force linear in the variables  $x$  and  $v$ , such as  $F_F$ , the time-reversal symmetry of the underlying microscopic dynamics implies that the friction and noise are interrelated by the so-called second fluctuation–dissipation theorem or fluctuation–dissipation relation (FDR) [70–73]

$$\langle \eta(t)\eta(t') \rangle = T\Gamma(|t - t'|). \quad (7)$$

Here  $T$  denotes the temperature and  $\Gamma(t)$  is the so called friction kernel defined by the integral

$$F_F(t) = -\int_{-\infty}^t dt' \Gamma(t - t')v(t'). \quad (8)$$

For a given friction  $F_F$ , the FDR (7) might imply that the noise must be complex valued. However, in order to admit its ordinary physical interpretation and realisability in a lab,  $\eta(t)$  is required to be a real-valued function. This condition implies that its power spectrum must be non-negative,

$$S(\omega) = \int_{-\infty}^{\infty} dt \langle \eta(t)\eta(0) \rangle \exp(-i\omega t) \geq 0. \quad (9)$$

For the system of equations (1)–(6), the conditions (7) and (9) can be satisfied for a certain range of model parameters only, see sections 3 and 4. In this range, equations (1) and (6) can be interpreted as describing a system with internal Hamiltonian  $H = U(x, t) + mv^2/2$  acted upon by a non-potential force  $F_N$  and coupled to a non-Markovian ‘equilibrium bath’ at temperature  $T$ . Let us now review some general properties of this system.

## 2.2. Properties of the mapping (table 1(B))

*Average thermodynamics.* If the above equilibrium mapping holds, the system’s thermodynamics obeys standard relations from classical [74] and stochastic [46, 47] thermodynamics. Namely, the average entropy flux into the non-Markovian heat bath at temperature  $T$  is given by the Clausius equality

$$\dot{S}_{NM} = -\dot{Q}_{NM}/T \quad (10)$$

where  $\dot{Q}_{\text{NM}} = \langle (F_{\text{F}} + \eta)\dot{x} \rangle$  is the average heat flux from the heat bath into the system. It can also be interpreted as the work done by the bath on the system per unit time. Here and below we employ Stratonovich calculus. The averages  $\langle \bullet \rangle$  should be performed over many realizations of the stochastic process.

The average heat flux is related via the first law,  $d\langle H \rangle/dt = \dot{Q}_{\text{NM}} + \dot{W}_{\text{E}}$ , to the average power input,  $\dot{W}_{\text{E}} = \langle \partial U/\partial t + F_{\text{N}}\dot{x} \rangle$ , of the system, due to external manipulations of the potential  $U$  and the non-potential force  $F_{\text{N}}$ . The sum of the rate of change of the system entropy,  $\dot{S}_{\text{S}}$ , and the entropy influx  $\dot{S}_{\text{NM}}$  in equation (10) is the total entropy production, which obeys the second law of thermodynamics [74]:

$$\dot{S}_{\text{tot}} = \dot{S}_{\text{S}} + \dot{S}_{\text{NM}} \geq 0. \quad (11)$$

*Dynamics.* Unlike a general delay system, which can exhibit over-damped, damped oscillatory, but also diverging behavior [10, 11, 13, 43, 61, 62], systems obeying the mapping of section 2.1 always eventually relax into a time-independent steady state for time independent parameters, confining potential  $U$ , and stationary non-potential forces  $F_{\text{N}}$ . If the latter vanishes in equation (5), the stationary probability density function (PDF) for position and velocity is given by the Gibbs canonical distribution,  $p(x, v; T) = p_x(x; T)p_v(v; T)$ , with

$$p_x(x; T) = \exp[-U(x)/T]/Z_x(T), \quad (12)$$

$$p_v(v; T) = \exp(-mv^2/(2T))/Z_v(T), \quad (13)$$

normalized by  $Z_x(T) = \int_{-\infty}^{\infty} dx \exp(-U(x)/T)$  and  $Z_v(T) = \int_{-\infty}^{\infty} dv \exp(-mv^2/(2T))$ . This is an equilibrium steady state and thus the corresponding entropy production rates  $\dot{S}_{\text{S}}$ ,  $\dot{S}_{\text{NM}}$ , and  $\dot{S}_{\text{tot}}$  vanish. For quasi-static variations of the potential, when the system PDF evolves through a set of such states, the total entropy change  $\Delta S_{\text{tot}} = \int_0^t \dot{S}_{\text{tot}}(t') dt'$  vanishes and the entropy change in the system,  $\Delta S_{\text{S}}$ , is exactly balanced by the entropy change in the bath,  $\Delta S_{\text{NM}}$ . The relaxation process to equilibrium is always accompanied by a decrease in the free energy of the system. This functional thus represents the Lyapunov function for the relaxation process that can easily be evaluated from stochastic trajectories of the system. Even stronger restrictions on the relaxation dynamics are imposed by the Evans–Searles fluctuation theorem [75, 76]. In contrast, besides a limited success [77], it is currently unknown if similar general restrictions also apply to relaxation towards non-equilibrium steady states.

The validity of these results for an arbitrary potential  $U(x)$  follows from general considerations of equilibrium statistical physics [74] and the FDR [70–73]. However, a closed dynamical equation, e.g. of Fokker–Planck type [78, 79], for the PDF of a nonlinear delay process is not known [13] making a general direct verification difficult. In section 5, we provide an explicit test for the specific potential  $U(x, t) = k_6 x^6/6 + k_3 x^3/3$  using Brownian dynamics (BD) simulations of equations (1) and (2). Besides, we tested the described results for various other polynomial potentials.

We stress that the described equilibrium-like properties of equilibrium delay processes do not trivialize their dynamics. As an example, consider a situation when the force  $F$  in equation (2) is linear in  $x$  and  $v$  and thus the system (1) and (2) is exactly solvable. For fixed initial conditions, one finds that the average position  $\langle x(t) \rangle$  and velocity  $\langle v(t) \rangle$  are identical for equilibrium ( $\eta_{\text{FB}}$  in table 1 determined by the conditions (7) and (9)) and standard ( $\eta_{\text{FB}} = 0$ ) delay processes. The four correlation functions  $\langle A(t)B(0) \rangle$  for  $A, B = x, v$  may then merely differ in the stationary distribution of the initial conditions.

*Fluctuation theorems.* From a stochastic-thermodynamics perspective, it is interesting to also consider a finite-speed protocol rendering the potential time-dependent. Specifically, in section 5, we test two fluctuation theorems for the stochastic work  $w = \int_0^t \partial U(x, t')/\partial t' dt'$  done on the system, if  $k_6 = k_6(t')$ ,  $t' \in (0, t)$  is varied non-quasi-statically, namely the Jarzynski equality [80]

$$\langle \exp(-w/T) \rangle = \exp(-\Delta F/T) \quad (14)$$

and the Crooks' fluctuation theorem [81]

$$\rho_{\text{F}}(w)/\rho_{\text{R}}(-w) = \exp[(w - \Delta F)/T]. \quad (15)$$

Here,  $\Delta F$  is the free energy difference between equilibrium states corresponding to the final and initial values of the potential,  $\rho_{\text{F}}$  is the probability distribution for work measured along the process when the potential changes from  $U(x, 0)$  to  $U(x, t)$ , and  $\rho_{\text{R}}$  is the probability distribution for work measured along the time-reversed process. For the both fluctuation theorems, the forward process departs from equilibrium. The validity of Jarzynski's equality requires the existence of initial and final Gibbs stationary states and Crooks' fluctuation theorem additionally requires the FDR and Gaussianity of the noise [82]. The described processes fulfill all these requirements and, indeed, our simulations confirm equations (14) and (15).

*Perturbative expansions.* Even though based on an ad hoc choice of the noise our results represent first exact analytical solutions to stationary PDFs for a nonlinear SDDE. As such, they might pave the way for studying steady states and thermodynamic properties of systems controlled by more general nonlinear SDDEs. We show in section 6 that linear-response theory [72] can be used to calculate time-dependent averages in perturbed (nonlinear) equilibrium delay systems. Besides such classical linear response, one can derive some explicit approximate formulas for specific perturbations on the level of moments calculated directly from the nonlinear system of SDDEs (1) and (2).

### 2.3. Equilibrium feedback (table 1(C))

The formal interpretation of dynamical equations (1) and (2) as a model for a system immersed in a non-Markovian equilibrium bath and driven by a time-local force  $F_E$ , in section 2.1, allowed us to utilize a wealth of known results. However, in practice, these equations usually describe feedback-driven systems in contact with a Markovian heat bath exerting a memoryless friction  $-\gamma_0 v$  and Gaussian white noise  $\sqrt{2T_0\gamma_0}\xi(t)$  with  $\langle \xi(t)\xi(t') \rangle = \delta(t-t')$ . Usually the system's environment provides such a bath. To investigate this 'more natural' interpretation, we rewrite the dynamical equation for the velocity as

$$m\dot{v}(t) = F_E(x, v, t) + F_{FB} - \gamma_0 v(t) + \sqrt{2T_0\gamma_0}\xi(t) \quad (16)$$

and interpret it as describing a system immersed in a standard, i.e. Gaussian and Markovian, heat bath at temperature  $T_0$ . This system is controlled by the time-local force  $F_E$  and the feedback force

$$F_{FB} = \kappa_\tau [x(t) - x(t - \tau)] - \gamma_\tau v(t - \tau) + \eta_{FB}(t) \quad (17)$$

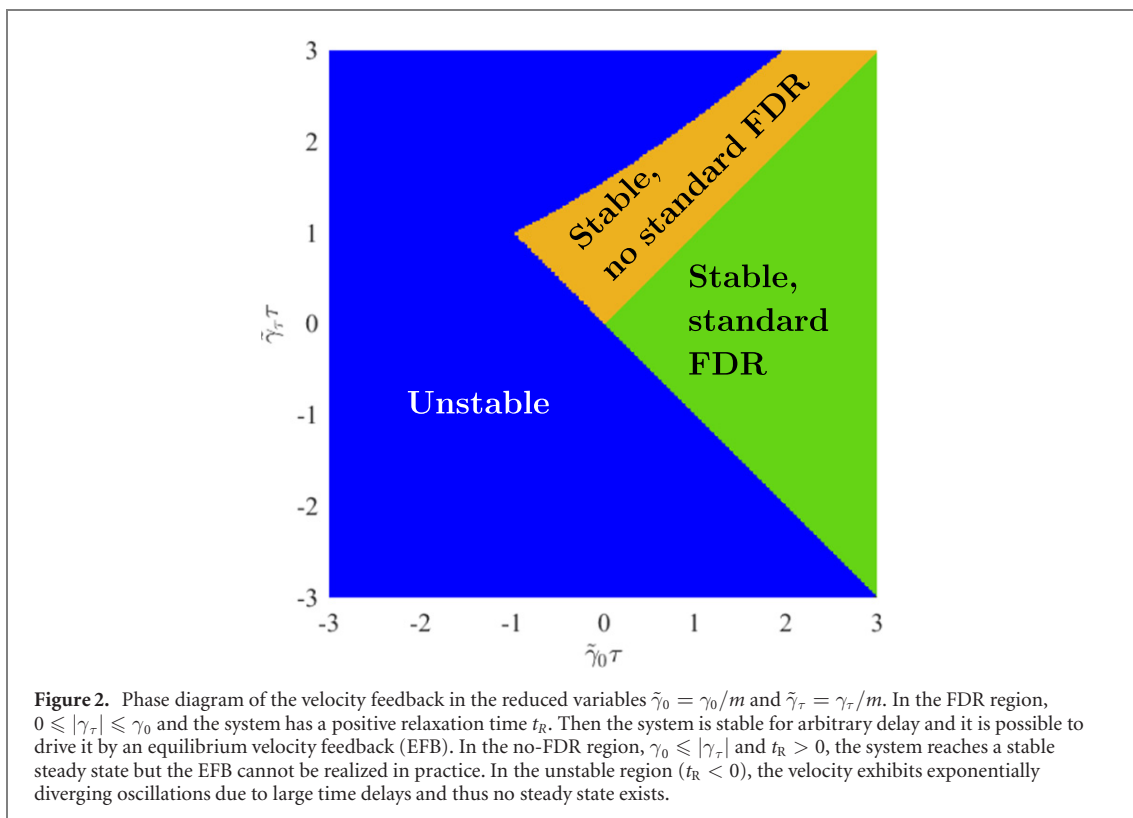
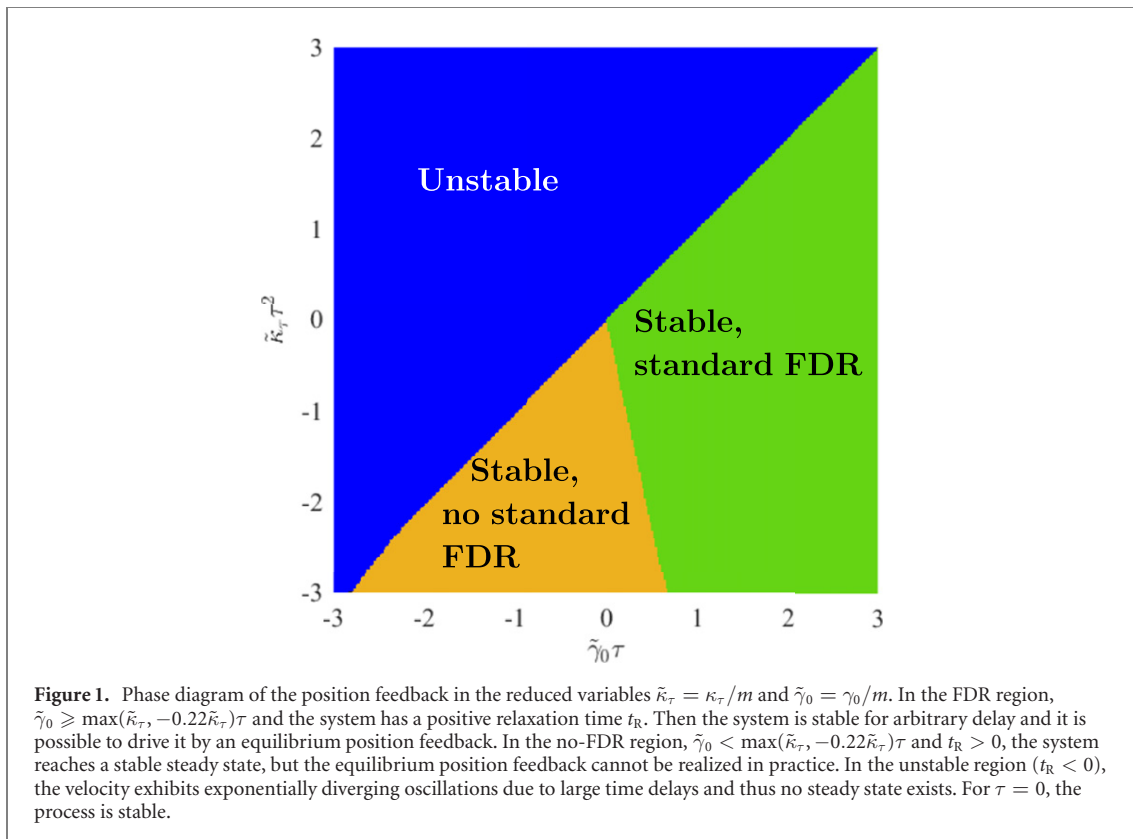
composed of the systematic delayed component  $F_F$  and the 'feedback' noise  $\eta_{FB}(t) \equiv \eta(t) - \sqrt{2T_0\gamma_0}\xi(t)$ , see table 1(C). Given that the conditions (7) and (9) are fulfilled, we call this process an equilibrium feedback (EFB) process.

Importantly, the formal results concerning the system dynamics, i.e. the stationary PDFs (12) and (13), are valid regardless of the interpretation, and thus they apply also for EFB. This means that the EFB is ideal from the point of view of passivity-based control [83], which is a branch of control theory that aims to balance the power delivered into the system with its dissipation. Generic feedback can lead to divergences and instabilities when the energy influx by the feedback gradually increases the internal energy of the system. However, EFB processes are always stable and passive in the sense that the resulting steady states are robust against perturbations and all the energy injected into the system is dissipated. In appendix A, we moreover show that, under realistic conditions, the temperature  $T$  corresponding to the Boltzmann PDF reached by the EFB is always larger than the ambient temperature  $T_0$ .

The thermodynamics of EFB has to be treated with care. In particular, the total entropy production is interpretation-dependent. But the stochastic work done on the system by varying the potential remains the same, and the fluctuation theorems (14) and (15) are still valid. Differences arise in the definitions of the remaining thermodynamic fluxes. With the present definition of the heat bath, the heat flux reads  $\dot{Q}_M = \langle (-\gamma_0 v + \sqrt{2T_0\gamma_0}\xi)\dot{x} \rangle$ . And, in addition to the average power  $\dot{W}_E$  delivered to the system by the potential and non-potential forces, one has to consider also the power  $\dot{W}_{FB} = \langle F_{FB}\dot{x} \rangle$  associated with the feedback force  $F_{FB}$ .

In a conventional feedback process, this power is accompanied by an information influx [48, 49, 51, 52] that, for example, allows the feedback to cool the system [36, 38]. The resulting (effective) temperature of the system is then smaller than the temperature of the ambient bath, implying a positive heat flux from the bath into the system,  $\dot{Q}_M > 0$ . In a steady state, the conventional feedback is thus able to cool the ambient bath by extracting the power  $-\dot{W}_{FB} = \dot{Q}_M > 0$  from it. However, for an arbitrary force  $F_E$ , the second law (11) together with the relation  $\dot{Q}_{NM} = \dot{Q}_M + \dot{W}_{FB} = -T\dot{S}_{NM}$  imposes an upper bound  $\dot{Q}_M \leq T\dot{S}_S - \dot{W}_{FB}$  on the heat delivered from the bath to the system via the EFB. And, in appendix A, we show that under equilibrium conditions,  $\partial U/\partial t = F_N = \dot{S}_S = 0$ , the EFB brings the system to an effective temperature,  $T$ , larger than the ambient temperature,  $T_0$ . Hence, the heat flux  $\dot{Q}_M$  is always negative, the EFB performs network on the system,  $\dot{W}_{FB} = -\dot{Q}_M > 0$ , and it eventually heats the ambient bath. This means that the EFB cannot be used for standard (zero non-potential force and time-independent potential) feedback cooling of the system [36, 38].

Sections 3 and 4 clarify when EFB can be realized with time-delayed forces depending on either the earlier position or velocity, i.e. when the corresponding feedback noise  $\eta_{FB}$  in table 1(C) can be constrained to be real valued. The technical details are given in appendix B. The resulting parameter regimes where the EFB can be realized in these two situations are depicted in phase diagrams (figure 1 and 2). EFB with time-delayed forces depending on both delayed position and velocity can be investigated in a similar manner, but the corresponding phase diagram becomes three-dimensional. In section 5, we verify the



validity of our theoretical results by a BD simulation of the equilibrium velocity feedback. In section 6, we discuss several perturbative expansions pushing the theory beyond the parameter regime of the equilibrium delay processes. We conclude in section 7.

### 3. Equilibrium position feedback

Let us now consider the situation of the position-dependent feedback force ( $\gamma_\tau = 0$  in equation (17))

$$F_{\text{FB}} = \kappa_\tau [x(t) - x(t - \tau)] + \eta_{\text{FB}}(t) = F_{\text{F}} + \gamma_0 v(t) + \eta_{\text{FB}}(t). \quad (18)$$

The generalized friction force  $F_{\text{F}} = \kappa_\tau [x(t) - x(t - \tau)] - \gamma_0 v(t)$  can be written using the friction kernel

$$\Gamma(t) = [2\gamma_0 \delta(t) - \kappa_\tau \Theta(\tau - t)] \Theta(t), \quad (19)$$

where  $\Theta(\cdot)$  denotes the Heaviside step function. This result can be verified by direct substitution into equation (8) and integrating the term including velocity  $v(t) = \dot{x}(t)$  by parts, cf equation (9.14) in reference [13].

The conditions (7) and (9) on the EFB imply that the friction kernel (19) and the total noise  $\eta(t) = \eta_{\text{FB}}(t) + \sqrt{2T_0\gamma_0}\xi(t)$  (see table 1(A)) must obey the FDR,

$$\langle \eta(t)\eta(t') \rangle / T = 2\gamma_0 \delta(t - t') - \kappa_\tau \Theta(\tau - |t - t'|), \quad (20)$$

and that the corresponding power spectrum must be non-negative,

$$S(\omega) = 2 \left[ \gamma_0 - \kappa_\tau \tau \frac{\sin(\omega\tau)}{\omega\tau} \right] \geq 0. \quad (21)$$

Using  $\max[\sin(x)/x] = 1$  and  $\min[\sin(x)/x] \approx -0.22$ , this implies the inequalities

$$0 \leq \max(\kappa_\tau, -0.22\kappa_\tau)\tau \leq \gamma_0, \quad (22)$$

which specify the parameter regime where the noise  $\eta(t)$  satisfying the FDR (20) can actually be realized in the lab (for details of noise realization, see appendix B). The inequalities require non-negative  $\gamma_0$  which is always fulfilled in the EFB interpretation, where  $\gamma_0$  measures the strength of the background friction. For  $\gamma_0 \geq 0$ , the inequalities (22) bounds the feedback strength  $\kappa_\tau$  as  $-\gamma_0/0.22 \leq \kappa_\tau \leq \gamma_0$ .

Under these conditions, the equilibrium position feedback fulfills all the properties described in section 2. In particular it eventually yields the stable equilibrium distribution (12) and (13) whenever  $\partial U/\partial t = F_{\text{N}} = 0$ . However, the time delay in a general feedback may yield diverging trajectories for certain parameter values. As an independent check that the parameter regime (22) allowing for equilibrium position feedback always leads to stable stationary solutions, we investigate the overall stability of position feedback described by equations (16) and (18) for the case  $F_{\text{E}} = 0$ , which can be inspected analytically.

Specifically, the process (16) eventually reaches a stable steady state if all the corresponding relaxation times,  $t_{\text{R}}$ , are positive. To calculate them, we substitute the feedback force (18),  $F_{\text{E}} = 0$ , and  $v = \dot{x}$  in equation (16), set  $\eta(t) = 0$ , and solve the resulting equation using the exponential ansatz  $x = \exp(-\lambda t/\tau)$ .<sup>4</sup> The obtained transcendental equation

$$m\lambda^2 = \gamma_0 \tau \lambda - \kappa_\tau \tau^2 [\exp(\lambda) - 1] \quad (23)$$

can in general only be solved numerically and has infinitely many solutions. As the relaxation time of the system,  $t_{\text{R}}$ , we identify the smallest  $\tau/\Re[\lambda]$  solving equation (23), where  $\Re[\bullet]$  denotes the real part. The system eventually relaxes into a stable steady state with  $\langle v(t) \rangle = 0$  if  $t_{\text{R}} > 0$ . An approximate explicit solution to equation (23) can be obtained in the limit of small delay. Expanding the friction  $F_{\text{F}}$  in equation (18) up to the first order in  $\tau$ , we get

$$m\dot{v}(t) \approx -(\gamma_0 - \kappa_\tau \tau)v(t) + \sqrt{2T_0\gamma_0}\xi(t). \quad (24)$$

The last two terms can be interpreted as a noise and friction from an equilibrium bath with friction coefficient  $\gamma_0 - \kappa_\tau \tau$ , which yields stable dynamics where all the energy injected into the system by the feedback is dissipated (passive dynamics) if

$$\kappa_\tau \tau \leq \gamma_0. \quad (25)$$

The conditions for the general case are depicted in figure 1. Indeed, the whole parameter regime where the EFB can be defined according to the FDR (20) (green) is found to be stable. This shows that, as expected, EFB is passive and stable. Nevertheless, the regime of stability,  $t_{\text{R}} > 0$ , is broader (orange). Noteworthy, the

<sup>4</sup> One can analogously treat systems with a potential  $U$ , by linearising it around a (local) minimum and absorbing the resulting linear time-local force into  $\kappa_\tau$ .



system can be stable even for  $\gamma_0 < 0$  if the feedback strength  $\kappa_\tau$  is also sufficiently negative. This could have been anticipated from the approximate condition for stability (25), which predicts the boundary between stable and unstable regimes for  $\tilde{\gamma}_0\tau \gtrsim -2$  remarkably well. The approximate dynamics allows to define an effective FDR with an effective temperature  $T_{\text{eff}} = T_0/(1 - \kappa_\tau\tau/\gamma_0)$ . However, beyond the small delay approximation, the existence of such effective FDR is not guaranteed. For smaller values of  $\tilde{\gamma}_0$ , higher order terms in the delay make the system more unstable than expected from the linear analysis. In the unstable regime (blue), the mean velocity exhibits exponentially increasing oscillations [43, 62].

#### 4. Equilibrium velocity feedback

Next, we perform the same analysis as in the previous section for the velocity-dependent feedback force ( $\kappa_\tau = 0$  in equation (17))

$$F_{\text{FB}} = -\gamma_\tau v(t - \tau) + \eta_{\text{FB}} = F_{\text{F}} + \gamma_0 v(t) + \eta_{\text{FB}}. \quad (26)$$

The friction  $F_{\text{F}} = -\gamma_\tau v(t - \tau) - \gamma_0 v(t)$  now corresponds to the friction kernel

$$\Gamma(t) = [2\gamma_0\delta(t) + \gamma_\tau\delta(t - \tau)] \Theta(t). \quad (27)$$

in equation (8). The FDR relation (7) for the total noise  $\eta(t) = \eta_{\text{FB}}(t) + \sqrt{2T_0\gamma_0}\xi(t)$  (see table 1(A)) now reads

$$\langle \eta(t)\eta(t') \rangle / T = 2\gamma_0\delta(t - t') + \gamma_\tau\delta(|t - t'| - \tau), \quad (28)$$

and thus the condition following from the positivity of the power spectrum (9) reads

$$S(\omega) = 2 [\gamma_0 + \gamma_\tau \cos(\omega\tau)] \geq 0. \quad (29)$$

The equilibrium velocity feedback thus can be realized if the inequality

$$0 \leq |\gamma_\tau| \leq \gamma_0, \quad (30)$$

holds (for detail of the realization, see appendix B). As we have seen for the equilibrium position feedback,  $\gamma_0$  must be non-negative, which is fulfilled in the EFB interpretation. For  $\gamma_0 \geq 0$ , the inequalities (30) impose that the amplitude  $\gamma_\tau$  of the delayed component of the friction cannot exceed that of the Markov component. Different from the corresponding inequality for  $\kappa_\tau$  in the position feedback, this condition is now symmetric with respect to  $\gamma_\tau = 0$ .

Similarly as in the case of the position feedback, we inspect the region of stability of the general linear velocity feedback for  $F_{\text{E}} = 0$  and compare it to the region (30) allowing to realize the stable EFB. To this end, we insert the feedback force (26) and  $F_{\text{E}} = 0$  in equation (16), set  $\eta(t) = 0$ , and solve the resulting equation using the exponential ansatz  $v(t) = \exp(-t/t_{\text{R}} + i\omega t)$ , with real parameters  $t_{\text{R}}$  and  $\omega$ . Solving the resulting algebraic equation for the relaxation time  $t_{\text{R}}$ , we find

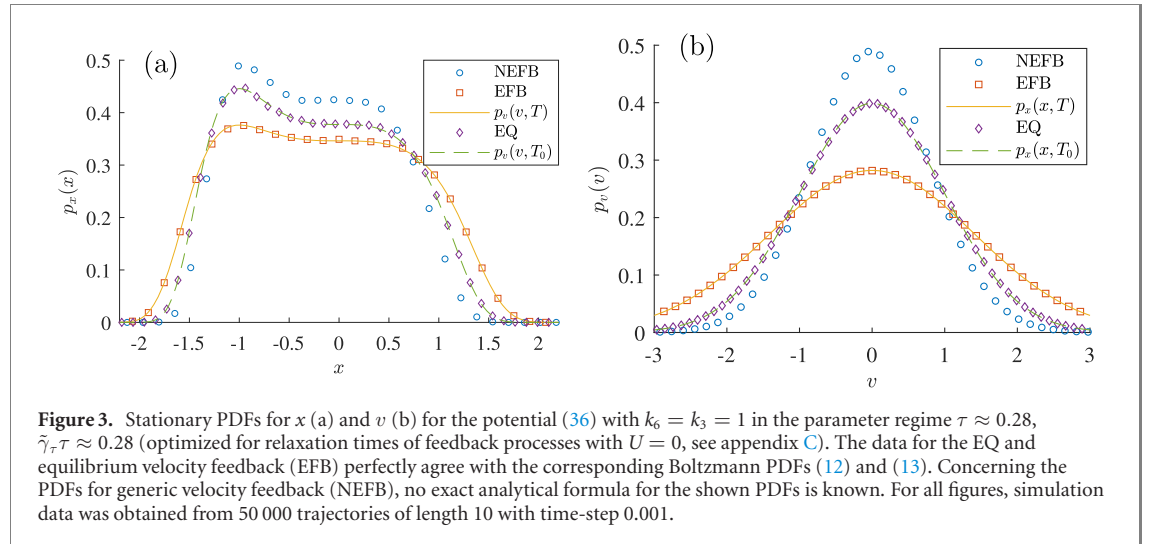
$$t_{\text{R}} = \frac{\tau}{\Re(\tilde{\gamma}_0\tau - W[-\tilde{\gamma}_\tau\tau \exp(\tilde{\gamma}_0\tau)])}, \quad (31)$$

where  $W[\cdot]$  stands for the Lambert W function,  $\Re(\cdot)$  denotes the real part, and  $\tilde{\gamma}_0 = \gamma_0/m$  and  $\tilde{\gamma}_\tau = \gamma_\tau/m$ . The Lambert W function is a multivalued function and, in order to assess stability of the system, we numerically determine the smallest  $t_{\text{R}}$  resulting from equation (31). In this case, the small-delay expansion of the friction  $F_{\text{F}}$  in equation (26) yields

$$m\dot{v}(t) \approx -(\gamma_0 + \gamma_\tau)v(t) + \gamma_\tau\tau\dot{v}(t) + \sqrt{2T_0\gamma_0}\xi(t) \quad (32)$$

and thus it suggest that the system will be stable for  $\gamma_\tau > -\gamma_0$  (positive effective friction coefficient) and  $\gamma_\tau\tau/m < 1$  (positive effective mass). It also allows to define an effective FDR with an effective temperature  $T_{\text{eff}} = T_0/(1 + \gamma_\tau/\gamma_0)$  valid for small delays.

These formulas correctly yield the bottom boundary between the unstable and stable regions in the phase diagram generated using the exact conditions (30) and (31) depicted in figure 2. As for the position feedback, the region where the FDR (20) and thus the equilibrium velocity feedback can be defined (green) is indeed stable. And the regime of stability,  $t_{\text{R}} > 0$ , is broader than the FDR regime and still extends to region of negative friction coefficients  $\gamma_0 < 0$  (orange). In the unstable regime (blue), the mean velocity again exhibits exponentially increasing oscillations [43, 62].



## 5. Demonstration of equilibrium velocity feedback

Let us now discuss a specific realization of the equilibrium velocity feedback and show that it indeed has all the properties described in section 2. As detailed in appendix B a possible (parsimonious) form of the total noise  $\eta(t) = \eta_{\text{FB}}(t) + \sqrt{2T_0\gamma_0}\xi(t)$ , which fulfills the FDR (28) for equilibrium velocity feedback, is obtained by setting  $\eta_{\text{FB}}(t) = \sqrt{\alpha_\tau}\xi(t - \tau)$ . The parameters of the corresponding feedback force  $F_{\text{FB}}(t)$  (26),

$$F_{\text{FB}}(t) = -\gamma_\tau v(t - \tau) + \sqrt{\alpha_\tau}\xi(t - \tau), \quad (33)$$

can be tuned to represent various equilibrium velocity delay process. As a benchmark for the EFB, we consider three processes distinguished by values of the coefficients  $\gamma_\tau$  and  $\alpha_\tau$  above: (i) equilibrium process (EQ) with  $\gamma_\tau = \alpha_\tau = 0$  and thus  $F_{\text{FB}}(t) = 0$ ; (ii) non-equilibrium (generic) velocity feedback (NEFB) with  $\gamma_\tau > 0$  and  $\alpha_\tau = 0$  and thus  $F_{\text{FB}}(t) = -\gamma_\tau v(t - \tau)$ ; and (iii) equilibrium velocity feedback (EFB) with  $\gamma_0 \geq \gamma_\tau > 0$  and  $\alpha_\tau > 0$  obeying equation (B.2). The last condition is compatible with an equilibrium state at arbitrary temperature  $T > T_0$ , if we set

$$\alpha_\tau = 2\gamma_0 (T - T_0), \quad (34)$$

$$\gamma_\tau = \pm 2\gamma_0 \sqrt{\frac{T_0}{T}} \sqrt{1 - \frac{T_0}{T}}, \quad (35)$$

where  $T_0/T \leq 1$ . Thus, in agreement with the discussion in appendix A, the additional noise present in the EFB always agitates or ‘heats’ the system above the ambient temperature  $T_0$ . Note that the above expressions do not depend on the delay  $\tau$ .

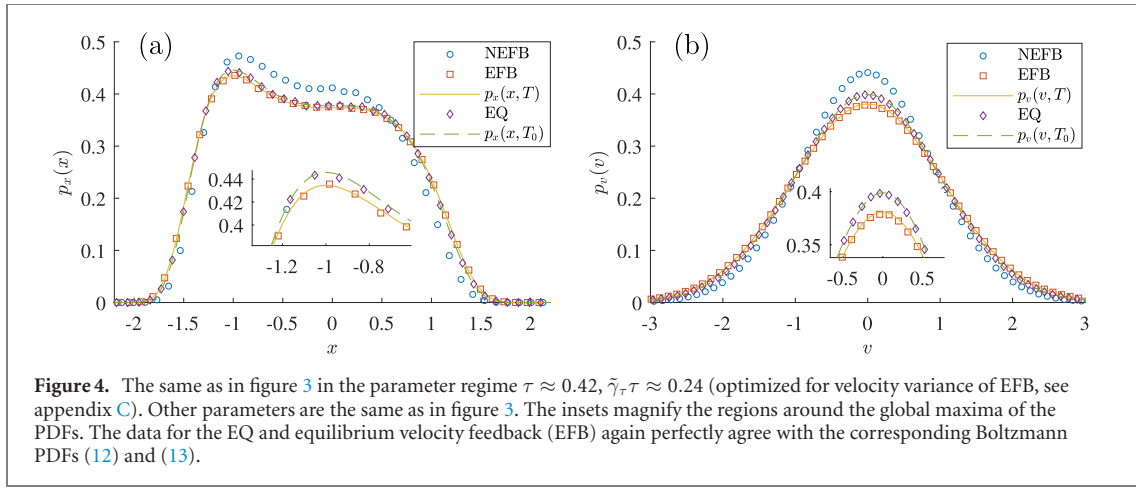
### 5.1. Dynamics

To gain intuition concerning the behavior of the equilibrium velocity feedback process, we now consider the specific system obeying equations (1) and (16) with the feedback force (33) and the force  $F_E = -\partial U/\partial x$  induced by the potential

$$U(x) = \frac{k_6}{6}x^6 + \frac{k_3}{3}x^3. \quad (36)$$

We solve the dynamical equations using BD simulations for the NEFB, EFB and EQ described above. In all our illustrations, we use  $1/\tilde{\gamma}_0$  as our time unit and  $\sqrt{T_0/m}$  as our length unit. Velocity is thus measured in units of  $\tilde{\gamma}_0\sqrt{T_0/m}$ . We show results from BD simulations for the two parameter sets  $(\tilde{\gamma}_0\tau, \tilde{\gamma}_\tau\tau) \approx (0.28, 0.28)$  and  $(\tilde{\gamma}_0\tau, \tilde{\gamma}_\tau\tau) \approx (0.42, 0.24)$ . The first one yields fast relaxation of both feedback processes for  $U = 0$ . The second one is optimised to provide small velocity variance for EFB for  $U = 0$ . For more details, see appendix C.

In figure 3 and 4 we show the stationary PDFs for  $x$  and  $v$  obtained for the first and second parameter set, respectively. In both figures, the simulated PDFs for EQ and EFB perfectly overlap with the corresponding analytical Boltzmann PDFs (12) and (13) providing numerical evidence for our claims in section 2. As expected, the position and velocity fluctuations are always smallest for the NEFB and largest for the EFB.



To compare the relaxation dynamics of the three processes, we show in figure 5 the corresponding mean values  $\langle x \rangle$  and  $\langle v \rangle$  and variances  $\sigma_x^2$  and  $\sigma_v^2$  as functions of time for the initial condition  $(v, x) = (1, 0)$  for  $t \leq 0$ . Interestingly, the first moments corresponding to the EFB (solid yellow line) relax faster than those for the NEFB (dot-dashed blue line) and much faster than those for the EQ process (broken green line). This is clearly a nonlinear effect because, for  $U = 0$ , the EFB and NEFB share the relaxation time (31). Especially for the EFB the relaxation is considerably faster for the first parameter set (panels (a)–(d)) than for the second choice (panels (e)–(h)). This suggests that at least some intuition gained from the linear regime  $U = 0$  also applies to the nonlinear dynamics. In accord with figures 1 and 2, the position and velocity fluctuations are always smallest for the NEFB and largest for the EFB. Smaller velocity but also position variance for the EFB is obtained for the second parameter set.

For the NEFB we were not able to analytically predict both the time evolution of the depicted variables and their asymptotic values. To solve the full transient dynamics for the EQ and EFB is also a difficult problem. However, figure 5 shows that the moments in question converge to the values calculated from the corresponding Boltzmann distributions (12) and (13) with temperatures  $T_0$  (EQ) and  $T > T_0$  (NEQ), which provides further numerical evidence for our claims in section 2.

## 5.2. Heat flux

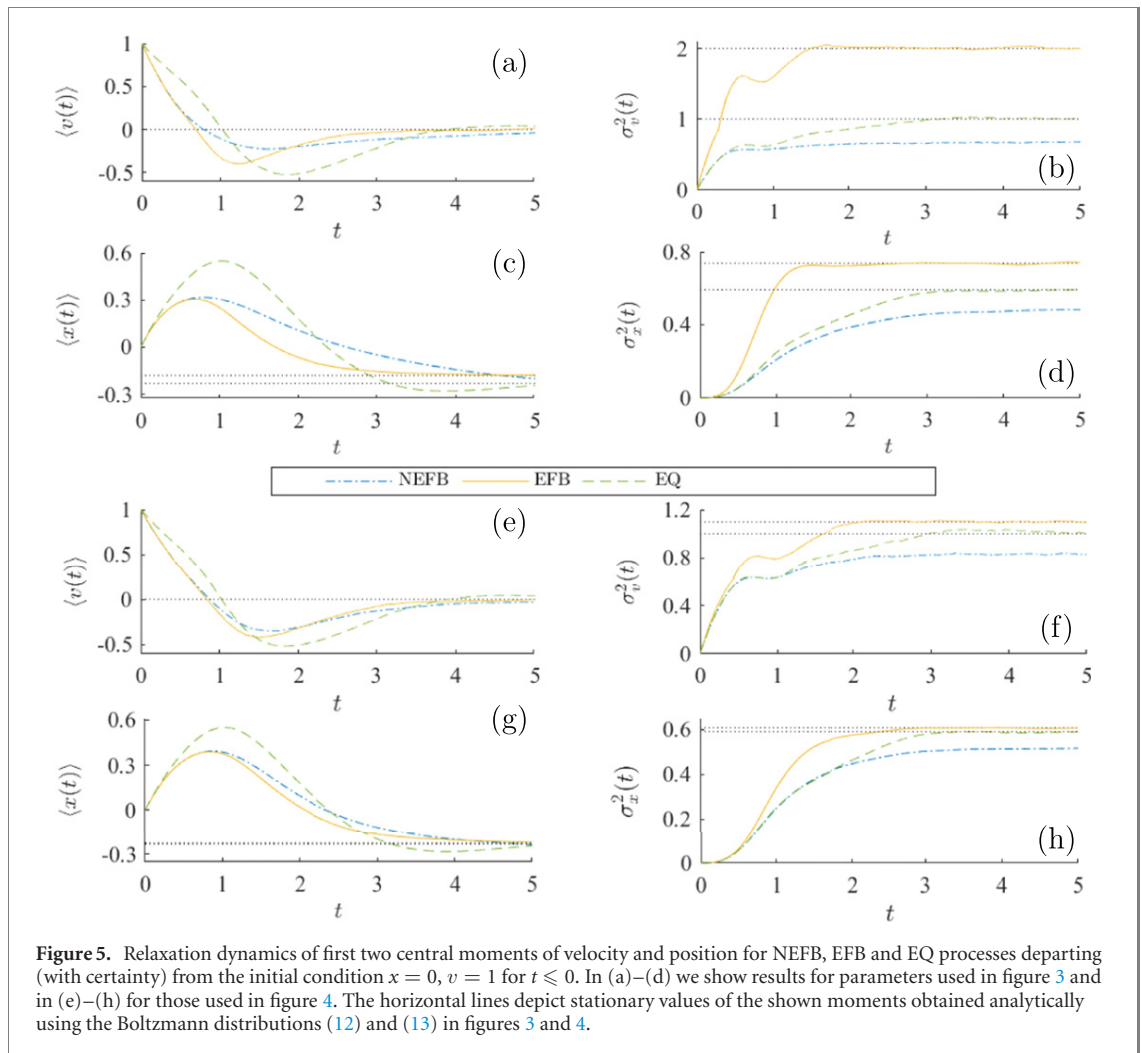
Let us now investigate the heat flux  $\dot{Q}_M = \langle (-\gamma_0 v + \sqrt{2T_0\gamma_0}\xi)\dot{x} \rangle$  from the ‘proper’ Markovian bath into the system due to the feedback, see table 1(C) and section 2.3. In appendix A, we show that for a general EFB with a positive delay time  $\tau$  the heat flux always reads

$$\dot{Q}_M^{\text{EFB}} = \gamma_0 ((\sigma_v^{\text{EQ}})^2 - (\sigma_v^{\text{EFB}})^2) = \frac{2\gamma_0}{m} (T_0 - T) < 0. \quad (37)$$

The EFB thus always performs work on the system, which is eventually dissipated in the bath. Figure 6 displays how the heat flux in the system evolves during the relaxation processes for EQ, EFB and NEFB discussed in figures 5(e)–(h). After the initial transient period, the heat flux for EFB converges to the negative value given by equation (37) and thus it heats both the system, as the corresponding stationary variances are larger than for the EQ process, and the proper bath. For EQ, the stationary heat flux is zero as imposed by the second law. For NEFB, the heat flux converges to a positive value. Thus the NEFB cools the system while absorbing heat from the proper bath.

The result (37) applies for arbitrarily small positive delay  $\tau$ . The specific form of the feedback force (33) allows us to also inspect what happens for vanishing delay. Then the system is still in the Boltzmann equilibrium state (12) and (13) with temperature  $T$ . However, the corresponding total noise  $\eta(t) = (\sqrt{2\gamma_0 T_0} + \sqrt{\alpha_\tau})\xi(t)$  and friction  $F_F = -(\gamma_0 + \gamma_\tau)v(t)$  can now be interpreted as a joint influence of the standard heat bath at temperature  $T_0$  and an additional ‘feedback heat bath’ at temperature  $T_F = \alpha_\tau/2\gamma_\tau = \sqrt{T/T_0} - 1/2$ . The laws of thermodynamics imply that heat flows from hot to cold and thus  $\dot{Q}_M$  is positive for  $T_F/T_0 > 1$  which occurs for  $T > 2T_0$ . Further, the heat flow is zero for  $T = 2T_0$ , where  $T_F = T_0$  and thus there is one global temperature only, and negative otherwise. Evaluating the heat flux  $\dot{Q}_M^{\text{EFB}}$  from equations (16) and (33) with  $\tau = 0$  using the approach of appendix A, we find the expression

$$\dot{Q}_M^{\text{EFB}} = \frac{\gamma_0 T_0}{m} \sqrt{\frac{T}{T_0} - 1} \left( 1 - \sqrt{\frac{T}{T_0} - 1} \right) \quad (38)$$



**Figure 5.** Relaxation dynamics of first two central moments of velocity and position for NEFB, EFB and EQ processes departing (with certainty) from the initial condition  $x = 0, v = 1$  for  $t \leq 0$ . In (a)–(d) we show results for parameters used in figure 3 and in (e)–(h) for those used in figure 4. The horizontal lines depict stationary values of the shown moments obtained analytically using the Boltzmann distributions (12) and (13) in figures 3 and 4.

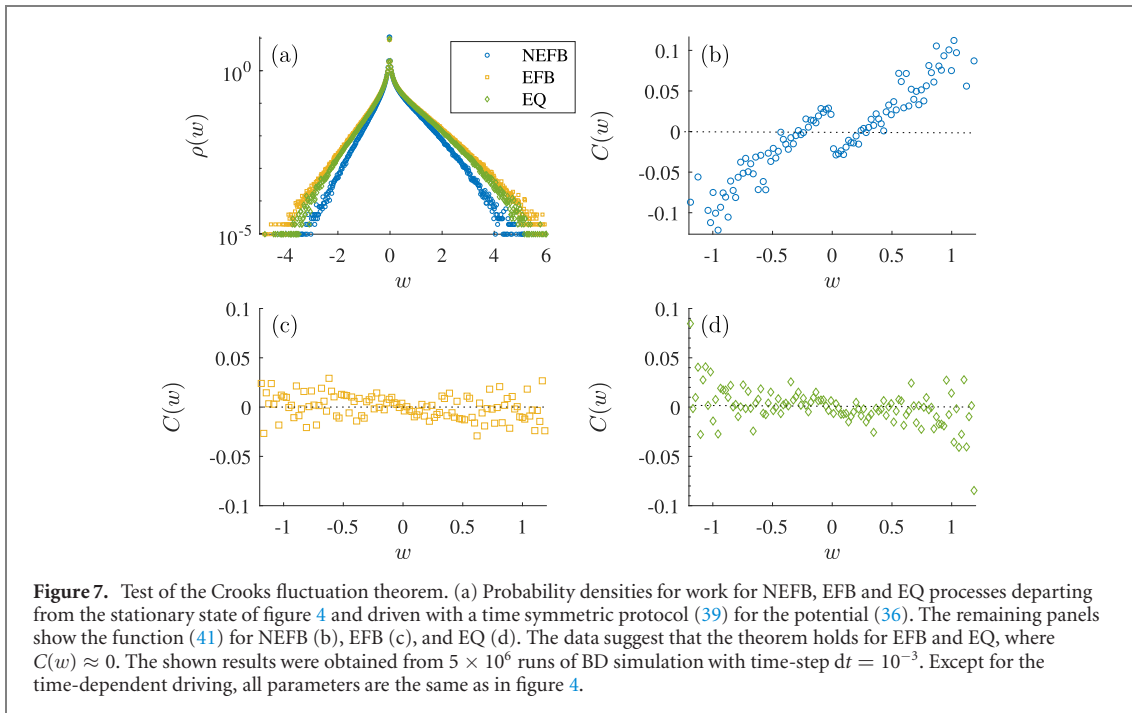
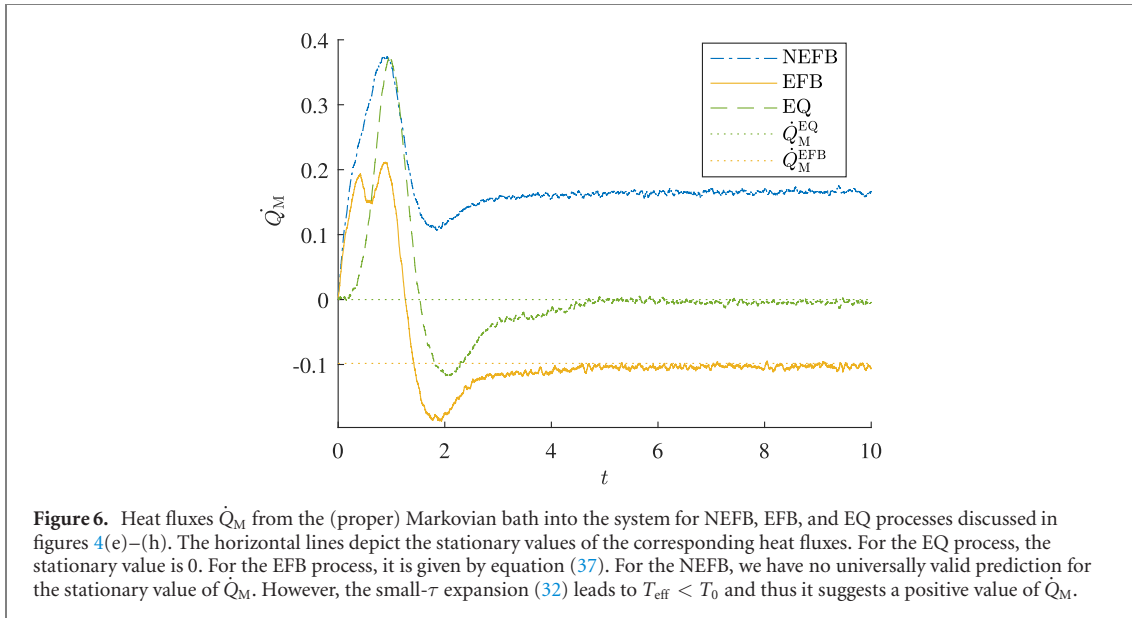
which indeed obeys the described properties.

Since  $\dot{Q}_M^{\text{EFB}}$  is strictly negative for  $\tau > 0$  and can be both positive and negative for  $\tau = 0$ , it exhibits a discontinuity at vanishing delay, in accord with the results described in reference [51]. Note that the presented situation with  $\tau = 0$  is physically weird since it seems impossible to record the noise and feed it back into the system without any delay. It also yields a strange behavior as the heat flux vanishes at the point where temperatures  $T_F$  and  $T_0$  are equal but  $T = 2T_0$ . This means that we constructed a bath at temperature  $T > T_0$  by using two strictly identical reservoirs at same the temperature  $T_0$  to which the system couples via different friction coefficients. The two baths provide the same realizations of the white noise  $\xi(t)$ , and thus the total noise intensity is given by the sum  $\sqrt{2\gamma_0 T_0} + \sqrt{\alpha\tau}$  of the intensities of the two noises. In contrast, connecting a system to two standard heat reservoirs always leads to equilibrium (vanishing heat flux) when the temperatures of the two baths are equal. The mathematical reason is that different reservoirs necessarily correspond to different noise realizations, regardless of their temperatures. As an example, consider heat baths A and B with friction and noise forces given by  $-\gamma_{A,B}v$ ,  $\sqrt{2\gamma_{A,B}T_0}\xi_{A,B}(t)$ , with independent Gaussian white noises  $\xi_{A,B}(t)$ . Then the joint action of these baths is described by the total friction  $-(\gamma_A + \gamma_B)v$  and noise  $\sqrt{2(\gamma_A + \gamma_B)T_0}\xi(t)$ , where  $\xi(t)$  is a unit variance Gaussian white noise (correlated with  $\xi_{A,B}(t)$ ). To sum up, the formal identification of the feedback force  $F_{\text{FB}}$  and noise  $\eta_{\text{FB}}$  for  $\tau = 0$  as effects of a standard heat bath correctly determines the sign of the heat flux  $\dot{Q}_M$  in equation (38), but it is physically problematic.

### 5.3. Fluctuation theorems

We conclude the numerical part of the paper by testing the work fluctuation theorems (14) and (15). To this end, we let the system relax into the steady state corresponding to the parameter regime of figures 4(e)–(h) and then we switch on the time-symmetric protocol

$$k_6 = 1 + 0.9 \sin(\pi t) \quad (39)$$



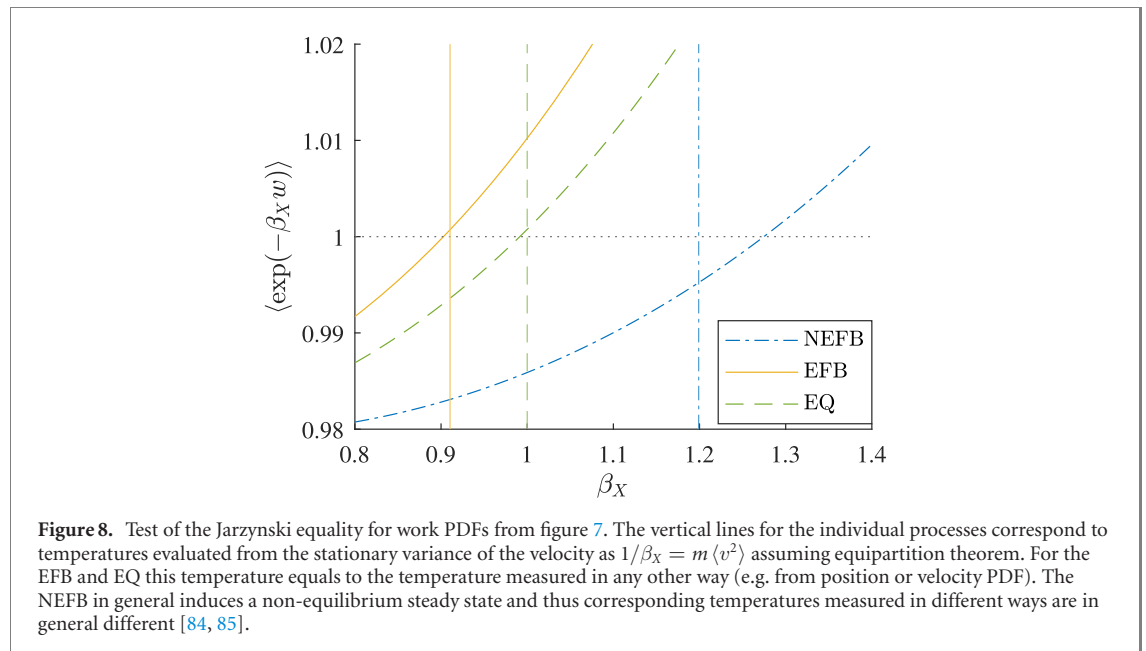
with  $t \in (0, 1)$  for the potential (36). During the time-dependent driving, we measure the stochastic work

$$w = \int_0^1 dt \frac{\partial U[x(t), t]}{\partial t} = \int_0^1 dt \dot{k}_6 x^6(t)/6 \quad (40)$$

and sample its PDF  $\rho(w)$ . Due to the symmetry of the protocol, the time-reversed process (R) and the forward process (F) in the fluctuation theorems (14) and (15) coincide and the free energy difference  $\Delta F$  vanishes. Validity of the Crooks fluctuation theorem (15) for the acquired PDFs thus implies that

$$C(w) \equiv \log \left[ \frac{\rho(w)}{\rho(-w)} \exp(w/T) \right] = 0. \quad (41)$$

In figure 7(a) we show the resulting PDFs for work obtained for the NEFB, EFB, and EQ processes. The panels (b)–(d) then show that from the three processes only the EFB (c) and EQ (d) yield  $C(w) = 0$  and thus fulfill the Crooks fluctuation theorem (41).



The validity of the Jarzynski equality is tested in figure 8, where we show values of averages  $\langle \exp(-\beta_X w) \rangle$  over the sample PDFs for work as functions of the parameter  $\beta_X$ . For EFB and EQ, we find that the average equals to one for  $\beta_X = 1/T$  and  $1/T_0$ , respectively, proving the validity of Jarzynski equality (14) with  $\Delta F = 0$  in these cases. For the NEFB, the system starts out of equilibrium so that it is not clear which (inverse) temperature should be used in equation (14). In the figure, we at least tested that choosing the temperature obtained from the variance of the velocity,  $1/\beta_X = 2\sigma_v^2$ , does not yield  $\langle \exp(-\beta_X w) \rangle = 1$ .

## 6. Beyond equilibrium feedback

In this section, we discuss possible analytical extensions of the equilibrium delay processes that might help to better understand general delay processes.

### 6.1. Classical linear response theory

Any Langevin equation where the friction and noise obey the FDR (7) can be thought of as a result of coarse-graining the full set of Hamiltonian equations for the system of interest and the corresponding bath over the bath degrees of freedom. This means that, the system with equilibrium delay can be regarded as a standard Hamiltonian system, which implies applicability of the classical linear response theory [70, 72, 73]. It states that the time-evolution of the mean value  $\langle A(t) \rangle$  induced by perturbations of the equilibrium system with Hamiltonian  $H = U + mv^2/2$  in the form  $H + \varepsilon f(t)B$  starting at time  $t = 0$  reads [72]

$$\langle A(t) \rangle_1 = \langle A(t) \rangle_0 + \frac{\varepsilon}{T} \int_0^t ds f(s) \langle A(t-s) \dot{B}(0) \rangle_0. \quad (42)$$

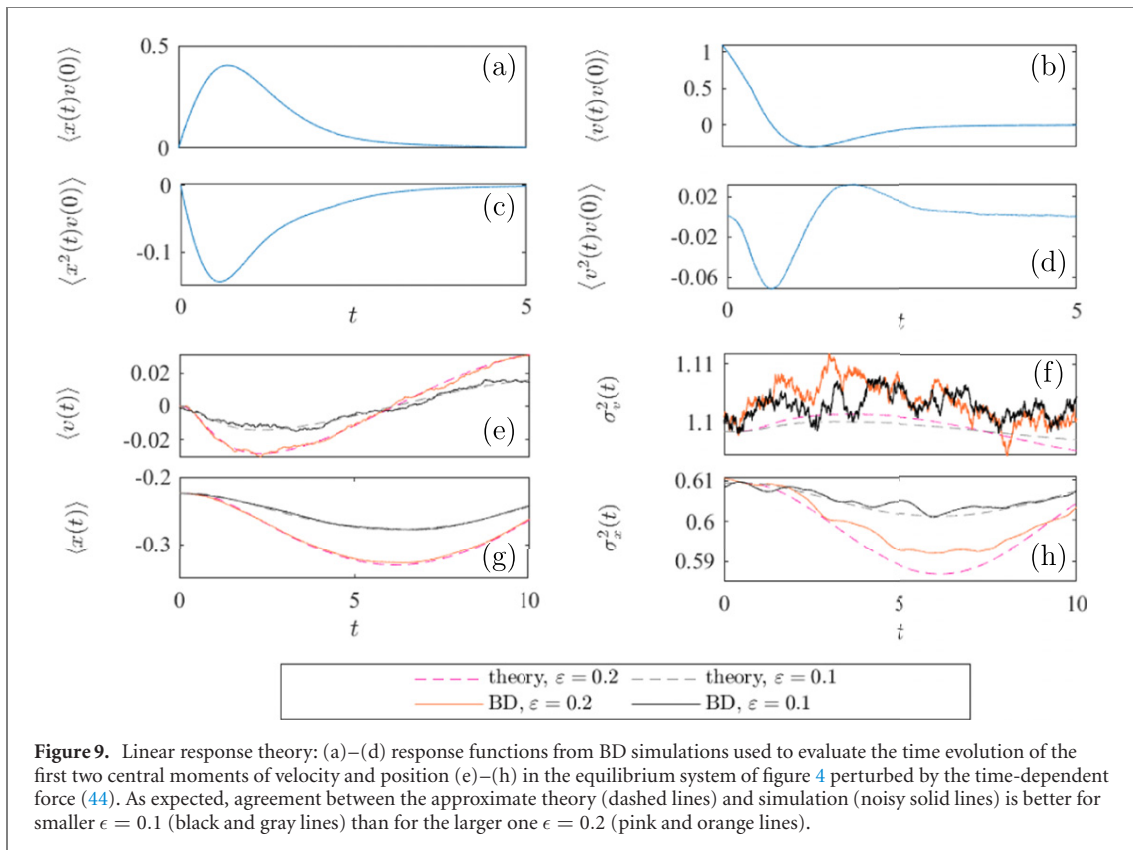
One assumes that the averages in the perturbed system can be expanded as  $\langle \dots \rangle = \langle \dots \rangle_0 + \varepsilon \langle \dots \rangle_1 + \dots$ , where the subscript 0 denotes average taken over the unperturbed Boltzmann PDF corresponding to Hamiltonian  $H$  (12) and (13), the subscript 1 denotes averages taken over the exact PDF up to the order  $\varepsilon$ , and so on.

We test the linear response theory using the specific equilibrium velocity feedback system discussed in section 5. We perturb the Hamiltonian by the term  $\varepsilon f(t)x$ . This term corresponds to a homogeneous time-dependent force  $-\varepsilon f(t)$  and thus the dynamical equation for velocity reads

$$m\dot{v}(t) = -\frac{\partial U}{\partial x} + F_{\text{FB}} - \gamma_0 v(t) + \sqrt{2T_0\gamma_0}\xi(t) - \varepsilon f(t). \quad (43)$$

The potential  $U$  is given by equation (36) and the feedback force by equation (33). We consider the parameter regime of figure 4 and the specific perturbation

$$\varepsilon f(t) = \varepsilon \sin(\pi t/10). \quad (44)$$



In figures 9(a)–(d) we show the time correlation functions  $\langle A(t)\dot{x}(0) \rangle_0 = \langle A(t)v(0) \rangle_0$ ,  $A = x, v, x^2, v^2$  obtained using BD simulations of this system with  $\epsilon = 0$ . The first two central moments of velocity and position obtained using equation (42) with force (44) via these time correlation functions are depicted in figures 9(e)–(h) together with the corresponding quantities obtained from BD simulation of the perturbed system. The figures show good agreement between simulations and linear response theory, which improves for smaller  $\epsilon$ .

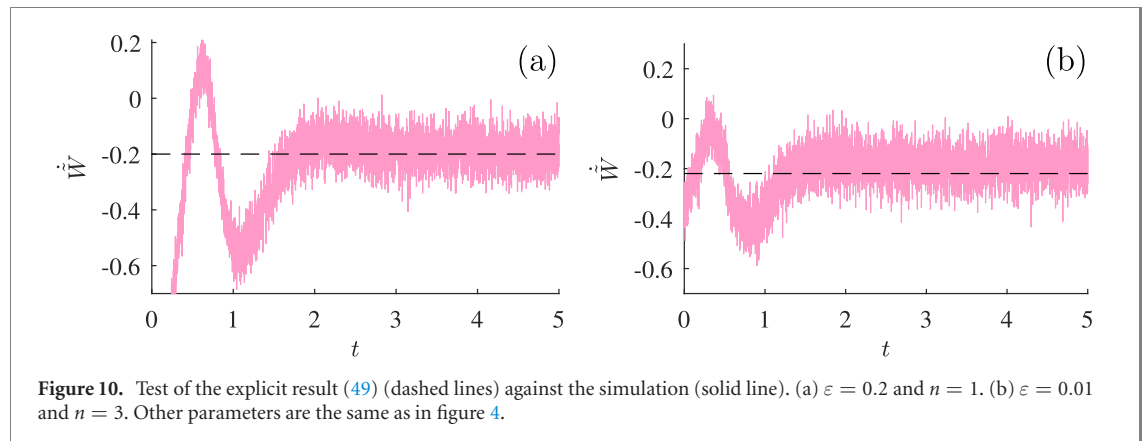
The validity of the linear response theory can be rationalized as follows. Even though we do not have an exact dynamical equation for the PDF for the equilibrium delay system, we know that the PDF for system and bath obeys a Liouville equation. The corresponding Liouville operator is composed of the system Hamiltonian  $H$ , the bath Hamiltonian, and the system-bath interaction energy. Even though it is hard to identify the latter two, one can rely on this Liouville equation as a starting point for perturbation theories around the parameter regime of the equilibrium delay. For example, one can derive equation (42) using the textbook approach of reference [72].

## 6.2. Langevin equation

The classical linear response theory (42) applies only for perturbations that can be subsumed into the Hamiltonian of the system. Other perturbations can be treated, e.g. on the footing of linear irreversible thermodynamics [74] or directly on the level of the Langevin equations (1) and (2). In order to present two simple examples of the latter type, we write these equations in the form of the equilibrium interpretation of table 1(B)

$$\dot{x} = v, \quad m\dot{v} = \left( -\frac{\partial U}{\partial x} + F_F + \eta \right) + \varepsilon g, \quad (45)$$

where the term proportional to  $\varepsilon$  is a perturbation. Perturbations dependent on time and/or time-delayed variables again require evaluation of time-correlation functions, which can rarely be obtained analytically. Perturbations that depend only on position can be absorbed into the potential, and the stationary PDF can be evaluated exactly. To obtain non-trivial analytical results, we will investigate two properties of steady states induced by perturbations of the form  $g = g[v(t)]$ , i.e. which depend solely on velocity. However, the obtained general restrictions (47) and (48) on the system dynamics apply for an arbitrary function  $g$ . In particular,  $g$  can, for these expressions, be a nonlinear function of position and velocity in the past. In such a case, the resulting process (45) is a truly nonlinear delay differential equation.



**Figure 10.** Test of the explicit result (49) (dashed lines) against the simulation (solid line). (a)  $\varepsilon = 0.2$  and  $n = 1$ . (b)  $\varepsilon = 0.01$  and  $n = 3$ . Other parameters are the same as in figure 4.

*Virial theorem.* The virial theorem states that twice the average kinetic energy of a system equals to the virial  $-\langle Fx \rangle$  [86]. For the system at hand, the total force  $F$  is given by the right-hand side (rhs) of equation (45). For the unperturbed system, this implies that

$$m \langle v^2 \rangle = - \left\langle \left( -\frac{\partial U}{\partial x} + F_F + \eta \right) x \right\rangle \quad (46)$$

as can be checked directly from the Langevin equations after multiplying equation (1) by  $x$ , equation (45) with  $\varepsilon = 0$  by  $v$ , summing the result, and taking the stationary ensemble average so that the time derivative of the cross-correlation  $\langle xv \rangle$  vanishes. Repeating this procedure for nonzero perturbation in equation (45), we find that if  $\langle g(v) \rangle_0 = 0$  the unperturbed virial theorem (46) remains valid up to first order in  $\varepsilon$  because  $\langle xg(v) \rangle_0 = \langle x \rangle_0 \langle g(v) \rangle_0$ . More generally, we find that

$$m \langle v^2 \rangle_n + \left\langle \left( -\frac{\partial U}{\partial x} + F_F + \eta \right) x \right\rangle_n = -\varepsilon \langle xg(v) \rangle_{n-1} \quad (47)$$

holds for all corrections of order  $n \geq 1$ . Even though this result cannot be evaluated explicitly for  $n > 1$  since  $\varepsilon^n \langle xg(v) \rangle_1$  is unknown, it can provide a stringent consistency check for simulations.

*Power.* The power  $\langle Fv \rangle$  exerted by the total force  $F$  on the rhs of equation (45) vanishes in the steady state of the unperturbed system since the time-derivative of  $\langle v^2 \rangle$  vanishes. With the perturbation switched on, we find

$$\left\langle \left( -\frac{\partial U}{\partial x} + F_F + \eta \right) v \right\rangle_n = -\varepsilon \langle vg(v) \rangle_{n-1}. \quad (48)$$

Besides providing another set of expressions useful as consistency checks in simulations, this equation provides an explicit non-trivial result for  $n = 1$ . Then the rhs  $-\varepsilon \langle vg(v) \rangle_0$  is in general non-zero and can be evaluated as average over the PDF (13). For example, for  $g(v) = v$ , we find

$$\dot{W} \equiv \left\langle \left( -\frac{\partial U}{\partial x} + F_F + \eta \right) v \right\rangle_n = -\varepsilon T. \quad (49)$$

Figure 10 verifies this approximate result by BD simulations for  $n = 1$  and  $n = 3$  and the same unperturbed dynamics as in section 6.1 above.

## 7. Conclusion

Using the second FDR, we have identified situations where one can interpret time-delayed feedback forces proportional to velocity or position in Langevin equations as friction forces imposed by an equilibrium bath. Our analysis reveals a previously unnoticed class of nonlinear SDDs, which can be treated analytically. They describe processes that obey standard thermodynamic constraints. In particular, their long-time distributions are of Gibbs canonical form and they obey standard fluctuation theorems. From the point of view of control theory, especially passivity-based control, the corresponding EFB processes are automatically stable and passive. However, their dynamics retains the full complexity of generic delay processes. One disadvantage is that the EFB always heats up the system and thus, unlike generic feedback protocols, cannot be used for standard feedback cooling.

As a practical demonstration, we have realized the equilibrium velocity feedback using BD simulations and shown that it exhibits all the formally derived properties. For so-called equilibrium velocity feedback,



not only the velocity at an earlier time but also the noise at that time are used to drive the system at present. As outlined in appendix B, we think it would be worthwhile to attempt to realize this feedback mechanism in actual laboratory experiments with Brownian particles [12, 14–16, 22, 24, 26, 39] or robots [18, 21, 25]. In particular, the so-called velocity damping protocols for feedback cooling of trapped microscopic particles record the particle velocity and later, after an experimental latency, apply to the particle a force proportional to that velocity [39, 69]. Assuming that the thermal bath is Markovian and position and velocity dependence of the systematic force in the dynamical equation for particle motion is known, measuring both the velocity and position at time  $t$  allows to determine the thermal noise, which can then be applied to the particle in the future, similarly as the velocity-dependent force. The potential drawback of this approach is that the measured noise will be affected by measurement uncertainties and finite measurement time resolution. As detailed in appendix B, large velocity measurement uncertainties might render the realization of the EFB impossible. On the other hand, the finite measurement time resolution just means that the obtained noise will effectively be integrated (low-pass filtered) over one measurement frame, which is also the case in our BD simulations. Other promising setups, where the EFB might be realised, are state-of-the-art bath engineering experiments [68] or experiments with feedback-driven overdamped Brownian particles [22]. Finally, EFB of the same type as in our computer simulations could be realised in experimental setups where artificial noise completely overshadows thermal noise. An example is shaken granular matter [67], where the noise is realised by shaking the granular system.

From a theoretical perspective, we believe that our results will shed further light on investigations of the dynamics and thermodynamics of nonlinear SDDs, which are known to be immensely resistant to analytical treatments. For example, the known stationary distributions for the equilibrium delay processes can serve as starting points of new perturbation theories valid for arbitrarily large delays. And, the known thermodynamics of the equilibrium delay processes can help to better understand the individual contributions to the total entropy production derived for nonlinear SDDs as studied in references [48, 49, 52].

## Acknowledgments

We acknowledge funding through a DFG-GACR cooperation by the Deutsche Forschungsgemeinschaft (DFG Project No. 432421051) and by the Czech Science Foundation (GACR Project No. 20-02955J). VH gratefully acknowledges support by Humboldt foundation.

## Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

## Appendix A. No feedback cooling with EQ feedback

Consider the heat flux  $\dot{Q}_M^{\text{EFB}}$  from the proper bath into the system in the steady state created by an EFB. The force applied by the proper bath on the system is  $-\gamma_0 v(t) + \sqrt{2T_0\gamma_0}\xi(t)$  and thus

$$\dot{Q}_M^{\text{EFB}} = -\gamma_0 \langle v^2 \rangle + \sqrt{2T_0\gamma_0} \langle \xi(t)v(t) \rangle \quad (\text{A.1})$$

with  $\langle v^2 \rangle = \lim_{t \rightarrow \infty} \langle v(t)^2 \rangle = \sigma_v^{\text{EFB}} = T/m$ . To calculate  $\langle \xi(t)v(t) \rangle$ , we use the formal solution

$$v(t) = \frac{1}{m} \int_0^t dt' \left[ F[x(t'), v(t'), t'] + F_D[x(t' - \tau), v(t' - \tau)] + \sqrt{2T_0\gamma_0}\xi(t') + \eta_{\text{FB}}(t') \right]. \quad (\text{A.2})$$

For a positive delay,  $\tau > 0$ , the causality implies that all terms on the rhs except for  $\xi(t')$  are independent of the white noise  $\xi(t)$  at time  $t$  (values of velocity and position at time  $t' \leq t$  were not affected by the white noise yet, and it is also reasonable to assume that the feedback noise  $\eta_{\text{FB}}(t')$  cannot be constructed in such a way that it would depend on  $\xi(t)$ ). In symbols we obtain

$$\langle \xi(t)v(t) \rangle = \frac{1}{m} \int_0^t dt' \sqrt{2T_0\gamma_0} \delta(t - t') = \frac{1}{m} \sqrt{\frac{T_0\gamma_0}{2}}. \quad (\text{A.3})$$

Using  $\sigma_v^{\text{EQ}} = T_0/m$  we get the stationary flux  $\dot{Q}_M$  induced by the EQ feedback in the form

$$\dot{Q}_M^{\text{EFB}} = \gamma_0 \left( (\sigma_v^{\text{EQ}})^2 - (\sigma_v^{\text{EFB}})^2 \right) = \frac{2\gamma_0}{m} (T_0 - T). \quad (\text{A.4})$$

As might have been anticipated from the beginning, the EQ feedback can cool the proper bath ( $\dot{Q}_M^{\text{EFB}} > 0$ ) only if it leads to a smaller velocity variance (effective temperature) than the proper bath. Let us now show that this can happen only if the feedback force contains also a time-local term in velocity.

Inserting the total noise  $\eta(t) = \sqrt{2T_0\gamma_0}\xi(t) + \eta_{\text{FB}}(t)$  (see table 1) into the FDR (7), we find

$$\langle \eta(t)\eta(t') \rangle / T = 2(\gamma_0 + \delta)\delta(t - t') + \dots = (2T_0\gamma_0 + \epsilon)/T\delta(t - t') + \dots \quad (\text{A.5})$$

The first line corresponds to the time-local component of the friction kernel. Specifically, the term proportional to  $\gamma_0$  corresponds to the background friction  $\gamma_0 v(t)$ . And the term proportional to  $\delta$  stems from the time-local component of the feedback force. The remaining terms abbreviated by  $\dots$  are determined by time non-local components of the friction kernel. The second line corresponds to the actual noise correlations. The term proportional to  $2T_0\gamma_0$  is obtained from the background noise and the term proportional to  $\epsilon\delta(t - t')$  originates from the time-local ( $t - t' = 0$ ) component of  $\langle \eta_{\text{FB}}(t)\eta_{\text{FB}}(t') \rangle$ . The remaining terms are given by the cross correlations  $\langle \xi(t)\eta_{\text{FB}}(t') \rangle$  and the  $t - t' \neq 0$  component of  $\langle \eta_{\text{FB}}(t)\eta_{\text{FB}}(t') \rangle$ .

Demanding that prefactors in front of the  $\delta$ -functions in equation (A.5) equal, we find

$$\frac{T_0}{T} = \frac{2(\gamma_0 + \delta) - \epsilon/T}{2\gamma_0}. \quad (\text{A.6})$$

Since  $\epsilon$  is the variance of the feedback noise  $\eta_{\text{FB}}(t)$ , it must be positive. The feedback can thus cool the system and the ambient bath ( $T < T_0$  and  $\dot{Q}_M^{\text{EFB}} > 0$ ) only if it contains a strong enough time-local component of the friction  $-\delta v(t)$ ,  $\delta > \epsilon/2T$ . Otherwise  $T \geq T_0$  and  $\dot{Q}_M^{\text{EFB}} \leq 0$  and thus the cooling by EFB is not possible. Realizing a feedback force containing a term proportional to a non-delayed velocity of the system seems technologically impossible and thus we conclude that, unlike generic feedback, the EFB cannot be used to cool the system in practical setups [36, 38].

## Appendix B. Noise generation in practice

An arbitrary Gaussian noise with a given positive power spectrum, and thus also any noise  $\eta(t)$  fulfilling the conditions (7) and (9), can be realized in practice using one of the standard procedures for generating a Gaussian process with given autocorrelation function [87–89].

Concerning the position delay of section 3 in the parameter regime (22), our brute-force implementation of the spectral method using equation (35) in reference [87] does not yield satisfactory results. However, we can recommend the method of references [88, 89] based on a discrete representation of the noise  $\eta(t)$  by  $M_t \mathbf{e}$ , where  $\mathbf{e}$  is a column vector of independent Gaussian white noises and the matrix  $M_t$  is given by a square root of a matrix describing the autocorrelation function (20).

A noise fulfilling the conditions (28) and (30) corresponding to the velocity delay of section 4 can be constructed analytically by introducing the ansatz

$$\eta(t) = \sqrt{\alpha_0}\xi(t) + \sqrt{\alpha_\tau}\xi(t - \tau). \quad (\text{B.1})$$

Here,  $\alpha_0 = 2T_0\gamma_0$ ,  $\xi(t)$  is the zero-mean, unit-variance, Gaussian white noise, i.e.  $\langle \xi(t) \rangle = 0$ ,  $\langle \xi(t)\xi(t') \rangle = \delta(t - t')$ . For  $\tau > 0$ , such  $\eta(t)$  obeys equation (28) if  $(\alpha_0 + \alpha_\tau)/T = 2\gamma_0$  and  $\sqrt{\alpha_0\alpha_\tau}/T = \gamma_\tau$ . Solving these equations with respect to  $\alpha_0$  and  $\alpha_\tau$ , we obtain the two roots

$$\alpha_0/T = \gamma_0 \pm \sqrt{\gamma_0^2 - \gamma_\tau^2}, \quad \alpha_\tau/T = \gamma_0 \mp \sqrt{\gamma_0^2 - \gamma_\tau^2}, \quad (\text{B.2})$$

which are equivalent due to the symmetry  $\alpha_0 \leftrightarrow \alpha_\tau$  of the noise leading to the same dynamics of  $v(t)$ . In agreement with our discussion of the power spectrum (29), this mapping breaks down if  $\gamma_0 < |\gamma_\tau|$  when  $\eta(t)$  becomes complex.

To realize (B.1) in practice, one needs to measure the (thermal) noise at time  $t - \tau$ , record it, and feed it back by some feedback mechanism at time  $t$ . However, from a practical perspective, the white noise  $\xi(t)$  represents an unreachable mathematical abstraction, as the trajectories are always measured with a finite time resolution. This applies not only to experiments but also to standard BD simulations, where the stochastic differential equation  $\dot{v}(t) = f(t, t - \tau) + \sigma\xi(t)$  is usually integrated over a time interval  $[t, t + dt]$ . This yields the update rule  $v(t + dt) = v(t) + f(t, t - \tau)dt + \sigma\sqrt{dt}L(t)$ , where the so-called Wiener process  $\sqrt{dt}L(t) = \int_t^{t+dt} dt'\xi(t')$  is a zero-mean Gaussian random variable with variance  $dt$ . The equilibrium velocity delay process (4) with the noise (B.1) has the form  $\dot{v}(t) = f(t, t - \tau) + \sqrt{\alpha_0}\xi(t) + \sqrt{\alpha_\tau}\xi(t - \tau)$ . And we simulate it using the update rule

$v(t + dt) = v(t) + f(t, t - \tau)dt + \sqrt{\alpha_0 dt} L(t) + \sqrt{\alpha_\tau dt} L(t - \tau)$ . To simulate the process, one thus has to keep track of the sequence  $L(t')$ ,  $t' \in [t - \tau, t]$  of the integrated noise.

An analogous procedure can in principle be applied to realize the required feedback noise in experiments with Markov heat bath described by white noise. Let us assume that we can measure the velocity  $v(t)$ , and we apply at time  $t$  a known force  $f(t, t - \tau)$  (this force can already contain the feedback noise). Subtracting  $f(t, t - \tau)dt$  from the measured change in the velocity  $v(t + dt) - v(t)$  per one measurement frame of length  $dt$  gives the integrated white noise  $\sqrt{\alpha_0 dt} L(t)$  at time  $t$ . This noise can be recorded and fed back into the system similarly as in the BD simulations. Let us now briefly discuss possible complications, which can arise in real-world experiments where the idealized mathematical abstraction might fail. (i) Computer simulations based on the model

$$\dot{x} = v, \quad m\dot{v} = F_E + F_{FB} - \gamma_0 v + \sqrt{2T_0\gamma_0}\xi(t), \quad (\text{B.3})$$

introduced in table 1(C), are in good agreement with measurements performed both in the underdamped and overdamped regime. An example of the former is the velocity damping experiments, where an optomechanically trapped particle diffusing in a dilute gas is cooled using velocity feedback [38, 39]. Overdamped experiments are usually performed with colloidal particles diffusing in water [22]. The only notable complication concerning equation (B.3) is that the friction and noise terms,  $-\gamma_0 v$  and  $\sqrt{2T_0\gamma_0}\xi(t)$ , in the experiments do not originate solely from the thermal bath. Instead, they arise as a combined effect of the bath, feedback apparatus, and trapping mechanism. However, the total noise and friction affect the particle dynamics in the same way as noise and friction induced by a Markov equilibrium bath described by an effective temperature  $T_0$  and friction coefficient  $\gamma_0$ . (ii) In the suggested experimental realization of the EFB, we assumed precise measurements of velocity. However, concerning underdamped setups, such measurements can nowadays be only performed in specialized apparatuses with charged particles [38]. Other underdamped experimental setups predict the particle velocity from its position, e.g. by using the so-called Wiener filter [39]. The result of such a measurement is instead of the real velocity  $v$  a velocity  $v + \delta v$ , where  $\delta v$  is the measurement error. This error has two effects. First, when a feedback force proportional to the measured velocity is applied to the system, the noise is not only increased by the error of the feedback loop, but also by the measurement error  $\delta v$ . The second effect, important for the experimental realization of the velocity feedback, is that the difference in the measured velocities does not equal to  $v(t + dt) - v(t)$  but to  $v(t + dt) - v(t) + \delta v(t + dt) - \delta v(t)$ . Therefore, the procedure for realization of the EFB suggested above is feasible only if the measurement noise  $\delta v(t + dt) - \delta v(t)$  is much smaller than the integrated 'thermal' noise  $\sqrt{2T_0\gamma_0} \int_t^{t+dt} dt' \xi(t')$ . (iii) Under overdamped conditions, equation (B.3) simplifies to

$$\gamma_0 v = F_E + F_{FB} - \gamma_0 v + \sqrt{2T_0\gamma_0}\xi(t), \quad (\text{B.4})$$

and the problems with velocity measurements affect only the magnitude of the effective temperature. This is because the integrated noise can now be calculated directly from the particle displacement  $x(t + dt) - x(t)$  during the interval  $dt$  as  $\sqrt{2T_0\gamma_0} \int_t^{t+dt} dt' \xi(t') = [x(t + dt) - x(t) - F_E(t) + F_{FB}(t)]/\gamma_0$ . The position measurements can be performed with sufficient precision to make this procedure feasible. To sum up, the suggested method for experimental realization of equilibrium velocity feedback should be realizable in a number of real-world applications with overdamped dynamics, while underdamped systems could be problematic if the velocity cannot be measured with sufficient precision. To close this section, we note that, even though the noise in equation (B.3) is often white, this is not valid generally. However situations with colored thermal noises are beyond the scope of this paper and will be covered elsewhere.

### Appendix C. Parameter sets used in figures 3 and 4

The two parameter sets considered in figures 3 and 4 are chosen as follows. The first one  $(\tilde{\gamma}_0\tau, \tilde{\gamma}_\tau\tau) \approx (0.28, 0.28)$  minimizes the ratio

$$\frac{t_R}{t_R^{\text{EQ}}} = \frac{\tilde{\gamma}_0\tau}{\Re[\tilde{\gamma}_0\tau - W(-e^{\tilde{\gamma}_0\tau}\tilde{\gamma}_\tau\tau)]} \quad (\text{C.1})$$

of the relaxation time  $t_R^{\text{EQ}} \equiv 1/\tilde{\gamma}_0$  for the EQ process and  $t_R$  (31) for the EFB and NEFB for  $U = 0$ . The second one  $(\tilde{\gamma}_0\tau, \tilde{\gamma}_\tau\tau) \approx (0.42, 0.24)$  minimizes the measure

$$\frac{t_R}{t_R^{\text{EQ}}} \left( \frac{\sigma_v^{\text{EFB}}}{\sigma_v^{\text{NEFB}}} \right)^2 \quad (\text{C.2})$$

of the trade-off between the relaxation time ratio (C.1) and the ratio of stationary velocity variance  $(\sigma_v)^2 = \langle v^2 \rangle$  for EFB and NEFB. For NEFB we take the variance for  $U = 0$ , when it can be calculated

analytically as

$$(\sigma_v^{\text{NEFB}})^2 \equiv \sigma_v^2 = \frac{T_0 \tilde{\gamma}_0 [\Omega + \tilde{\gamma}_\tau \sinh(\Omega\tau)]}{m \Omega [\tilde{\gamma}_0 + \tilde{\gamma}_\tau \cosh(\Omega\tau)],} \quad (\text{C.3})$$

where  $\Omega = \sqrt{\tilde{\gamma}_0^2 - \tilde{\gamma}_\tau^2}$ . For EFB, the variance  $(\sigma_v^{\text{EFB}})^2 = T/m$  follows for any  $U$  from equipartition.

In addition, we also tested the parameter set  $(\tilde{\gamma}_0\tau, \tilde{\gamma}_\tau\tau) \approx (0.34, 0.034)$  which minimizes the ratio  $\sigma_v^{\text{EFB}}/\sigma_v^{\text{NEFB}}$  but due to the small magnitude of the feedback force the numerical results for NEFB, EFB, and EQ are hardly distinguishable and we decided not to show them.

## Appendix D. Derivation of equation (C.3)

Consider the simple velocity process

$$\dot{v}(t) = -\tilde{\gamma}_0 v(t) - \tilde{\gamma}_\tau v(t - \tau) + \sqrt{\alpha} \xi(t). \quad (\text{D.1})$$

In the steady state,  $\langle v \rangle = 0$  and the variance  $\sigma_v^2 = \langle v^2 \rangle$  can be evaluated as follows.

The general solution to equation (D.1) reads [43]

$$v(t) = \lambda(t)v_0 - \tilde{\gamma}_\tau \int_{-\tau}^0 dt' \lambda(t - t' - \tau)v(t') + \sqrt{\alpha} \int_0^t dt' \lambda(t - t')\xi(t'), \quad (\text{D.2})$$

where  $v_0 = v(0)$ . The Green's function  $\lambda(t)$  solves equation (D.1) with vanishing noise term ( $\alpha = 0$ ), i.e.

$$\dot{\lambda}(t) = -\tilde{\gamma}_0 \lambda(t) - \tilde{\gamma}_\tau \lambda(t - \tau), \quad (\text{D.3})$$

and the initial condition  $\lambda(t < 0) = 0$  and  $\lambda(0) = 1$ . The most straightforward way for finding  $\lambda(t)$  is to employ a Laplace transformation in time. The result is [62]

$$\lambda(t) = \sum_{l=0}^{\infty} \frac{(-\tilde{\gamma}_\tau)^l}{l!} (t - l\tau)^l e^{-\tilde{\gamma}_0(t-l\tau)} \Theta(t - l\tau). \quad (\text{D.4})$$

In the stable regime,  $t_R > 0$  (cf equation (31)), where the Green's function eventually decays to zero,  $\lim_{t \rightarrow \infty} \lambda(t) = 0$ , the general solution (D.2) to equation (D.1) can be used for calculation of the time-correlation function,  $C(t) = \lim_{t_0 \rightarrow \infty} \langle v(t_0 + t)v(t_0) \rangle$ . The stationary variance  $\sigma_v^2 = C(0)$ , we are actually interested in, comes as a by-product of this calculation.

We find that for  $t > 0$  (see also reference [43])

$$C(t) = \alpha \lim_{t_0 \rightarrow \infty} \int_0^{t_0} dt' \lambda(t + t_0 - t')\lambda(t_0 - t'). \quad (\text{D.5})$$

This expression can already be used for plotting the time correlation function, however, it is not very suitable for inferring its properties. Following the approach in references [43, 61], we take the time derivative of equation (D.5) and use equation (D.3) for the Green's function to obtain the dynamical equation

$$\dot{C}(t) = -\tilde{\gamma}_\tau C(t - \tau) - \tilde{\gamma}_0 C(t) \quad (\text{D.6})$$

valid for  $t > 0$  due to the nonanalyticity of  $\lambda(t)$  at  $t = 0$ . The solution to this equation is given by equation (D.2):

$$C(t) = \lambda(t)C_0 - \tilde{\gamma}_\tau \int_{-\tau}^0 dt' \lambda(t - t' - \tau)C(t') \quad (\text{D.7})$$

and thus the decay time of the time-correlation function is given by the decay time (31) of the Green's function  $\lambda(t)$ . To evaluate the above expression, we need to find the stationary variance  $\alpha_0 \equiv C(0)$  and the delayed initial condition  $C(t)$  for  $t \in (-\tau, 0)$ . This can be done as follows. Employing the symmetry  $C(t) = C(-t)$  of the stationary time-correlation function, we rewrite equation (D.6) as

$$\dot{C}(t) = -\tilde{\gamma}_\tau C(\tau - t) - \tilde{\gamma}_0 C(t). \quad (\text{D.8})$$

For  $t \in (0, \tau)$ , we can differentiate this equation once again. The result is

$$\ddot{C}(t) = \Omega^2 C(t), \quad (\text{D.9})$$

where we used equation (D.6) and defined the (possibly imaginary) frequency  $\Omega = \sqrt{\tilde{\gamma}_0^2 - \tilde{\gamma}_\tau^2}$ . From equation (D.9), we find that for  $t \in [-\tau, \tau]$

$$C(t) = C_0 \cosh(\Omega t) + \dot{C}_0 \Omega^{-1} \sinh(\Omega |t|). \quad (\text{D.10})$$

Here,  $\dot{C}_0 = \lim_{t \rightarrow 0^+} \dot{C}(t)$  denotes the time-derivative of the time-correlation function, that is discontinuous at  $t = 0$  [61], for  $t > 0$  infinitesimally close to 0. From equation (D.5), we find that

$$\begin{aligned}\dot{C}_0 &= \alpha \lim_{t_0 \rightarrow \infty} \int_0^{t_0} dt' \dot{\lambda}(t_0 - t') \lambda(t_0 - t') = -0.5\alpha \lim_{t_0 \rightarrow \infty} \int_0^{t_0} dt' \frac{d}{dt'} \lambda^2(t_0 - t') \\ &= -0.5\alpha \lim_{t_0 \rightarrow \infty} [\lambda^2(0) - \lambda^2(t_0)] = -0.5\alpha.\end{aligned}$$

In order to evaluate  $C_0$ , we note that  $\dot{C}_0$  also follows from equation (D.8) with  $t = 0$ , yielding  $\dot{C}_0 = -0.5\alpha = -\tilde{\gamma}_\tau C(\tau) - \tilde{\gamma}_0 C_0$ . Using  $C(\tau) = C_0 \cosh(\Omega\tau) - 0.5\alpha\Omega^{-1} \sinh(\Omega\tau)$  given by equation (D.10) for  $t > 0$ , we finally get the desired result

$$C_0 = \sigma_v^2 = \frac{\alpha}{2} \frac{\Omega + \tilde{\gamma}_\tau \sinh(\Omega\tau)}{\Omega [\tilde{\gamma}_0 + \tilde{\gamma}_\tau \cosh(\Omega\tau)]}. \quad (\text{D.12})$$

For  $\tilde{\gamma}_0 = 0$ , we obtain  $\Omega = i\sqrt{\tilde{\gamma}_\tau}$ . Using the identities  $\sinh(ix) = i \sin x$  and  $\cosh(ix) = \cos x$ , the formula (D.12) can be written as

$$\sigma_v^2 = \frac{\alpha}{2\tilde{\gamma}_\tau} \frac{1 + \sin(\tilde{\gamma}_\tau\tau)}{\cos(\tilde{\gamma}_\tau\tau)}, \quad (\text{D.13})$$

which is the result derived in references [43, 61].

## ORCID iDs

Viktor Holubec  <https://orcid.org/0000-0002-6576-1316>  
 Artem Ryabov  <https://orcid.org/0000-0001-8593-9516>  
 Sarah A M Loos  <https://orcid.org/0000-0002-5946-5684>  
 Klaus Kroy  <https://orcid.org/0000-0002-6927-8647>

## References

- [1] Kyrychko Y N and Hogan S J 2010 *J. Vib. Control* **16** 943–60
- [2] Beuter A, Bélair J, Labrie C and Belair J 1993 *Bull. Math. Biol.* **55** 525–41
- [3] Chen Y, Ding M and Kelso J A S 1997 *Phys. Rev. Lett.* **79** 4501–4
- [4] Novák B and Tyson J J 2008 *Nat. Rev. Mol. Cell Biol.* **9** 981–91
- [5] Mackey M C 1989 *J. Econ. Theory* **48** 497–509
- [6] Voss H U and Kurths J 2002 *Analysis of Economic Delayed-Feedback Dynamics* (Berlin: Springer) pp 327–49
- [7] Stoica G 2005 *Proc. Am. Math. Soc.* **133** 1837–41
- [8] Gao Q and Ma J 2009 *Nonlinear Dyn.* **58** 209
- [9] Bechhoefer J 2005 *Rev. Mod. Phys.* **77** 783–836
- [10] Atay F 2010 *Complex Time-Delay Systems: Theory and Applications (Understanding Complex Systems)* (Berlin: Springer) <https://books.google.de/books?id=e2SGvHLpursC>
- [11] Lakshmanan M and Senthilkumar D 2011 *Dynamics of Nonlinear Time-Delay Systems (Springer Series in Synergetics)* (Berlin: Springer) <https://books.google.de/books?id=nV5F0kMsqa4C>
- [12] Gernert R, Loos S A M, Lichtner K and Klapp S H L 2016 Feedback control of colloidal transport *Control of Self-Organizing Nonlinear Systems* (Berlin: Springer) pp 375–92
- [13] Loos S 2020 Stochastic systems with time delay—probabilistic and thermodynamic descriptions of non-Markovian processes far from equilibrium *PhD Thesis* TU Berlin
- [14] Baraban L, Streubel R, Makarov D, Han L, Karnausenko D, Schmidt O G and Cuniberti G 2013 *ACS Nano* **7** 1360–7
- [15] Qian B, Montiel D, Bregulla A, Cichos F and Yang H 2013 *Chem. Sci.* **4** 1420–9
- [16] Bregulla A P, Yang H and Cichos F 2014 *ACS Nano* **8** 6542–50
- [17] Vásárhelyi G, Virágh C, Somorjai G, Tarcai N, Szörényi T, Nepusz T and Vicsek T 2014 Outdoor flocking and formation flight with autonomous aerial robots 2014 *IEEE/RSJ Int. Conf. Intelligent Robots and Systems* pp 3866–73
- [18] Mijalkov M, McDaniel A, Wehr J and Volpe G 2016 *Phys. Rev. X* **6** 011008
- [19] Zheng J, Dong J-G and Xie L 2017 *IEEE Trans. Autom. Control* **62** 5866–72
- [20] Zhang J, Liu Y-x, Wu R-B, Jacobs K and Nori F 2017 *Phys. Rep.* **679** 1–60
- [21] Leyman M, Ogemark F, Wehr J and Volpe G 2018 *Phys. Rev. E* **98** 052606
- [22] Muiños-Landin S, Fischer A, Holubec V and Cichos F 2021 *Sci. Robot.* **6** eabd9285
- [23] Khadka U, Holubec V, Yang H and Cichos F 2018 *Nat. Commun.* **9** 3864
- [24] Lavergne F A, Wendehenne H, Bäuerle T and Bechinger C 2019 *Science* **364** 70–4
- [25] Piwowarczyk R, Selin M, Ihle T and Volpe G 2019 *Phys. Rev. E* **100** 012607
- [26] Bäuerle T, Löffler R C and Bechinger C 2020 *Nat. Commun.* **11** 2547
- [27] Foss J, Longtin A, Mensour B and Milton J 1996 *Phys. Rev. Lett.* **76** 708–11
- [28] Marcus C M and Westervelt R M 1989 *Phys. Rev. A* **39** 347–59
- [29] Sompolinsky H, Golomb D and Kleinfeld D 1991 *Phys. Rev. A* **43** 6990–7011
- [30] Haken H 2007 *Brain Dynamics: An Introduction to Models and Simulations (Springer Series in Synergetics)* (Berlin: Springer) <https://books.google.de/books?id=H44amAlqgLGMC>
- [31] Gopalsamy K 2013 *Stability and Oscillations in Delay Differential Equations of Population Dynamics (Mathematics and Its Applications)* (Berlin: Springer) <https://books.google.de/books?id=0TfpCAAQBAJ>

- [32] Mao X, Yuan C and Zou J 2005 *J. Math. Anal. Appl.* **304** 296–320
- [33] Beretta E, Hara T, Ma W and Takeuchi Y 2001 *Nonlinear Anal. Theory Methods Appl.* **47** 4107–15
- [34] Rihan F A, Alsakaji H J and Rajivganthi C 2020 *Adv. Differ. Equ.* **2020** 502
- [35] Otto A, Just W and Radons G 2019 *Phil. Trans. R. Soc. A* **377** 20180389
- [36] Bushev P et al 2006 *Phys. Rev. Lett.* **96** 043003
- [37] Li T 2013 Millikelvin cooling of an optically trapped microsphere in vacuum *Fundamental Tests of Physics with Optically Trapped Microspheres* (Berlin: Springer) pp 81–110
- [38] Goldwater D, Stickler B A, Martinetz L, Northup T E, Hornberger K and Millen J 2019 *Quantum Sci. Technol.* **4** 024003
- [39] Penny T W, Pontin A and Barker P F 2021 *Phys. Rev. A* **104** 023502
- [40] Ramaswamy S 2010 *Annu. Rev. Condens. Matter Phys.* **1** 323–45
- [41] Bechinger C, Di Leonardo R, Löwen H, Reichhardt C, Volpe G and Volpe G 2016 *Rev. Mod. Phys.* **88** 045006
- [42] Gompper G et al 2020 *J. Phys.: Condens. Matter* **32** 193001
- [43] Geiss D, Kroy K and Holubec V 2019 *New J. Phys.* **21** 093014
- [44] Fridman E and Shaikhet L 2016 *Automatica* **74** 288–96
- [45] Jun Y and Bechhoefer J 2012 *Phys. Rev. E* **86** 061106
- [46] Sekimoto K 2010 *Stochastic Energetics (Lecture Notes in Physics)* (Berlin: Springer)  
<https://books.google.de/books?id=8Fq7BQAAQBAJ>
- [47] Seifert U 2012 *Rep. Prog. Phys.* **75** 126001
- [48] Rosinberg M L, Munakata T and Tarjus G 2015 *Phys. Rev. E* **91** 042114
- [49] Rosinberg M L, Tarjus G and Munakata T 2017 *Phys. Rev. E* **95** 022123
- [50] Van Vu T and Hasegawa Y 2019 *Phys. Rev. E* **100** 012134
- [51] Loos S A M and Klapp S H L 2019 *Sci. Rep.* **9** 2491
- [52] Munakata T and Rosinberg M L 2014 *Phys. Rev. Lett.* **112** 180601
- [53] Munakata T, Iwama S and Kimizuka M 2009 *Phys. Rev. E* **79** 031104
- [54] Adelman S A 1976 *J. Chem. Phys.* **64** 124–30
- [55] Fox R F 1977 *J. Math. Phys.* **18** 2331–5
- [56] Hänggi P 1978 *Z. Phys. B* **31** 407–16
- [57] Sancho J M, Miguel M S, Katz S L and Gunton J D 1982 *Phys. Rev. A* **26** 1589–609
- [58] Hernández-Machado A, Sancho J M, San Miguel M and Pesquera L 1983 *Z. Phys. B* **52** 335–43
- [59] Krüchler U and Mensch B 1992 *Stoch. Stoch. Rep.* **40** 23–42
- [60] Guillouez S, L'Heureux I and Longtin A 1999 *Phys. Rev. E* **59** 3970–82
- [61] Frank T D, Beek P J and Friedrich R 2003 *Phys. Rev. E* **68** 021912
- [62] McKetterick T J and Giuggioli L 2014 *Phys. Rev. E* **90** 042135
- [63] Giuggioli L, McKetterick T J, Kenkre V M and Chase M 2016 *J. Phys. A: Math. Theor.* **49** 384002
- [64] Frank T D 2005 *Phys. Rev. E* **71** 031106
- [65] Loos S A M and Klapp S H L 2017 *Phys. Rev. E* **96** 012106
- [66] Loos S A M and Klapp S H L 2019 *J. Stat. Phys.* **177** 95–118
- [67] D'Anna G, Mayor P, Barrat A, Loreto V and Nori F 2003 *Nature* **424** 909–12
- [68] Murch K W, Vool U, Zhou D, Weber S J, Girvin S M and Siddiqi I 2012 *Phys. Rev. Lett.* **109** 183602
- [69] Ferialdi L, Setter A, Toroš M, Timberlake C and Ulbricht H 2019 *New J. Phys.* **21** 073019
- [70] Kubo R 1966 *Rep. Prog. Phys.* **29** 255–84
- [71] Felderhof B U 1978 *J. Phys. A: Math. Gen.* **11** 921–7
- [72] Zwanzig R 2001 *Nonequilibrium Statistical Mechanics* (Oxford: Oxford University Press)
- [73] Kubo R, Toda M and Hashitsume N 2012 *Statistical Physics II: Nonequilibrium Statistical Mechanics (Springer Series in Solid-State Sciences)* (Berlin: Springer) <https://books.google.de/books?id=cF3wCAAQBAJ>
- [74] Callen H B 1985 *Thermodynamics and An Introduction to Thermostatistics* 2nd edn (New York: Wiley)
- [75] Evans D J, Searles D J and Williams S R 2009 *J. Stat. Mech.* **P07029**
- [76] Evans D J, Searles D J and Williams S R 2016 *Fundamentals of Classical Statistical Thermodynamics: Dissipation, Relaxation, and Fluctuation Theorems* (New York: Wiley)
- [77] Maes C, Netočný K and Wynants B 2011 *Phys. Rev. Lett.* **107** 010601
- [78] Risken H 1984 *Fokker–Planck Equation* (Berlin: Springer)
- [79] Van Kampen N 2011 *Stochastic Processes in Physics and Chemistry* (Amsterdam: Elsevier)  
<https://books.google.cz/books?id=N6II-6HIPxEC>
- [80] Jarzynski C 1997 *Phys. Rev. Lett.* **78** 2690–3
- [81] Crooks G E 1999 *Phys. Rev. E* **60** 2721–6
- [82] Speck T and Seifert U 2007 *J. Stat. Mech.* **L09002**
- [83] Ortega R, Perez J, Nicklasson P and Sira-Ramirez H 2013 *Passivity-based Control of Euler-Lagrange Systems: Mechanical, Electrical and Electromechanical Applications (Communications and Control Engineering)* (Berlin: Springer)  
<https://books.google.cz/books?id=jJzeBwAAQBAJ>
- [84] Cugliandolo L F 2011 *J. Phys. A: Math. Theor.* **44** 483001
- [85] Holubec V, Steffenoni S, Falasco G and Kroy K 2020 *Phys. Rev. Res.* **2** 043262
- [86] Goldstein H, Poole C and Safko J 2001 *Classical Mechanics* (Reading, MA: Addison-Wesley)
- [87] Shinozuka M and Deodatis G 1991 *Appl. Mech. Rev.* **44** 191–204
- [88] Pichot G 2016 Algorithms for stationary Gaussian random field generation *Technical Report RT-0484* (Paris: INRIA)  
<https://hal.inria.fr/hal-01414707>
- [89] Graham I G, Kuo F Y, Nuyens D, Scheichl R and Sloan I H 2018 *SIAM J. Numer. Anal.* **56** 1871–95

## ARTICLE

DOI: 10.1038/s41467-018-06445-1

OPEN

# Active particles bound by information flows

Utsab Khadka<sup>1</sup>, Viktor Holubec<sup>2,3</sup>, Haw Yang<sup>1</sup> & Frank Cichos<sup>4</sup>

Self-organization is the generation of order out of local interactions. It is deeply connected to many fields of science from physics, chemistry to biology, all based on physical interactions. The emergence of collective animal behavior is the result of self-organization processes as well, though they involve abstract interactions arising from sensory inputs, information processing, storage, and feedback. Resulting collective behaviors are found, for example, in crowds of people, flocks of birds, and swarms of bacteria. Here we introduce interactions between active microparticles which are based on the information about other particle positions. A real-time feedback of multiple active particle positions is the information source for the propulsion direction of these particles. The emerging structures require continuous information flows. They reveal frustrated geometries due to confinement to two dimensions and internal dynamical degrees of freedom that are reminiscent of physically bound systems, though they exist only as nonequilibrium structures.

<sup>1</sup>Department of Chemistry, Princeton University, Princeton, NJ 08544, USA. <sup>2</sup>Institute for Theoretical Physics, Universität Leipzig, 04103 Leipzig, Germany.

<sup>3</sup>Department of Macromolecular Physics, Faculty of Mathematics and Physics, Charles University, V Holešovičkách 2, CZ-180 00 Praha, Czech Republic.

<sup>4</sup>Peter Debye Institute for Soft Matter Physics, Universität Leipzig, 04103 Leipzig, Germany. Correspondence and requests for materials should be addressed to F.C. (email: [cichos@physik.uni-leipzig.de](mailto:cichos@physik.uni-leipzig.de))

Active particles serve as simple microscopic model systems for living objects such as birds, fish, or people and mimic in particular the propulsion of bacteria or cells without the complexity of physical properties and chemical networks in living objects<sup>1</sup>. They consume energy to propel persistently and as such they have given considerable insight into collective behaviors of active materials already<sup>2–6</sup>. With their bare function of self-propulsion they are, however, missing the important ingredients of sensing and feedback, which most living objects from cells up to whole organisms have in common. All of their living relatives have signaling inputs which they use to gain information about the environment. Employing external information, organisms interact such that birds and fish are able to self-organize into flocks or schools<sup>7–9</sup> and, on a microscopic level, cells may regulate gene expression<sup>10,11</sup>. Using sensing, information processing and feedback, living systems may go beyond what is prescribed by physical interactions such as the Coulomb, van der Waals or hydrophobic interactions. For cells/bacteria the information about cellular density (quorum sensing) is, for example, inferred from the concentration of signaling molecules released by the cells leading to a regulation of various physiological activities such as biofilm formation<sup>10</sup>. Concentration gradients of nutrients may serve as sensory input leading to a directed motion termed chemotaxis<sup>12</sup>. For birds, the information used for the formation of flocks is suggested to be the visual perception of the number of neighboring birds<sup>8</sup>. In this respect all information that is processed by the organism is linked to a physical representation<sup>13</sup> (e.g. concentration, number of objects, orientation,...), but the resulting structure and dynamics is disconnected from a direct physical interaction. In the particular examples mentioned above, though, the structure formation depends on the active response of the organism and its ability to steer based on its recognition of the environment<sup>7,14–16</sup>. While active particles do not have sensory inputs, information processing units and feedback mechanisms built in yet, suitable control mechanisms may introduce this complexity fostering the exploration of new emergent phenomena. An information exchange between active particles has not been tackled so far, but seems to be a natural step towards extending their functionality.

Like bacteria, active particles have to break the time symmetry of low Reynolds number hydrodynamics in order to propel. They have to provide asymmetries to generate directed motion. In many cases this asymmetry is built into the structure in form of two hemispheres with different chemical or physical properties, so called Janus particles<sup>1,17</sup>. As the propulsion direction is bound to the symmetry axis of the particle, rotational diffusion becomes a relevant process that limits the control of Janus particles<sup>18,19</sup>. The self-propulsion mechanism presented below relies on a scheme for generating self-thermophoresis<sup>20</sup>; unlike past approaches our new scheme utilizes the spatially controlled asymmetric input and release of energy around a symmetric particle. This scheme delivers a precise control of each individual particle in a larger ensemble, and allows us to demonstrate how active particles may form structures just by the exchange of information on other particle positions using a feedback control mechanism for steering the particles.

## Results

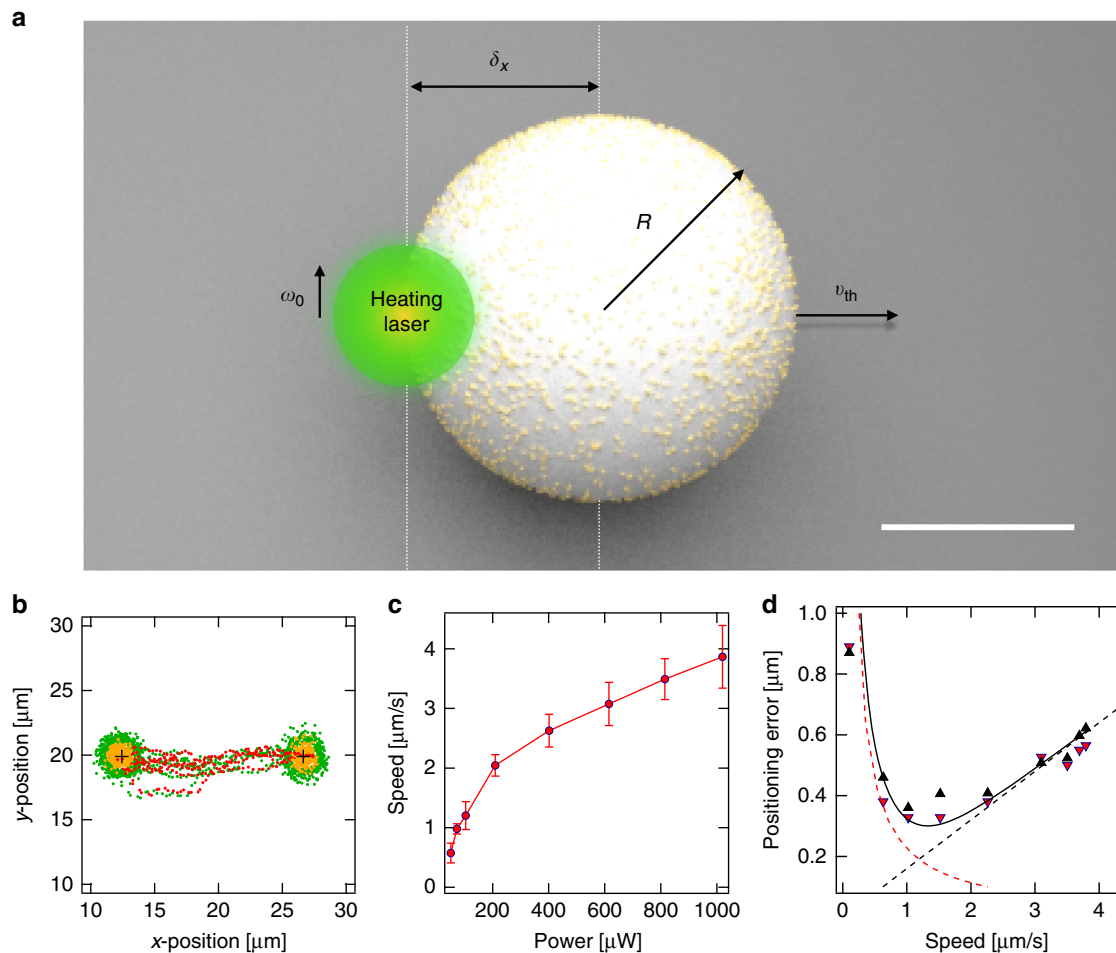
**Particle control.** Our active particle is constructed of a melamine resin sphere with 30% of the surface uniformly decorated by gold nanoparticles of about 10 nm diameter (Supplementary Fig. 1). Illuminating the particle asymmetrically by a deflectable focused laser beam generates an inhomogeneous surface temperature and results in the desired self-thermophoretic motion based on thermo-osmotic surface flows<sup>21,22</sup> (Fig. 1a). This design allows for

a new control scheme for active particle steering. To control the particle propulsion direction, we place the laser beam's focal spot near the circumference of the particle. The propulsion direction is then the vector from that heated circumferential spot to the particle center. Different from standard active particles, the timescale of the rotational diffusion of the particle is irrelevant due to the missing particle asymmetry (Supplementary Fig. 5).

We first evaluate properties of a single active particle and the control accuracy in an experiment driving the particle between two target positions and confining it for a certain time at the targets (Fig. 1b). We extract the dependence of the active particle propulsion velocity on the heating power from this experiment finding a non-linear scaling (Fig. 1c) as the particle slips away from the focus during the camera exposure time. Incorporating this slipping process in a model (Supplementary Note 1) reproduces this nonlinear increase qualitatively (Supplementary Fig. 4). The confinement at the target position can be characterized by a positioning error  $\sigma = \sqrt{\langle (\mathbf{r} - \mathbf{r}_t)^2 \rangle}$  – the standard deviation of the probability density of the particle–target distance in the steady state. Here,  $\mathbf{r}$  and  $\mathbf{r}_t$  are the coordinates of the swimmer and target, respectively. The positioning error obeys two regimes<sup>18,19,23,24</sup>. When the displacement of the particle due to the propulsion is smaller than the diffusive displacement during the exposure time  $\Delta t_{\text{exp}}$ , the positioning error is reflected by a simple sedimentation model. A constant particle speed  $v_{\text{th}}$  drives the particle radially towards the target position against the Brownian motion with a diffusion coefficient  $D_0$ . Hence, the density distribution in the steady state is exponential with a characteristic length scale (the sedimentation length in two dimensions)  $\rho = \sqrt{6}D_0/v_{\text{th}}$  as indicated by the red dashed curve in Fig. 1d<sup>19</sup>. When the active particle speed increases, an overshooting of the particle over the target position due to the finite sampling of the position of the particle with the camera exposure time  $\Delta t_{\text{exp}}$  defines the positioning error<sup>25</sup>. The overshooting distance equals the traveled distance within the time between two frames  $\Delta t_{\text{exp}}$  and increases linearly with the velocity of the active particle as shown by the black dashed line. The sum of both contributions determines the positioning error  $\sigma$  which is depicted for different particle speeds from the experiments (triangle markers, forward and backward motion in Fig. 1b) together with the predicted curve  $\sigma = \sqrt{6D_0^2/c^2v_{\text{th}}^2 + c^2v_{\text{th}}^2\Delta t_{\text{exp}}^2}$  with no free parameters (solid black curve). A minimum positioning error of  $\sigma = 370$  nm is found for the exposure time of  $\Delta t_{\text{exp}} = 80$  ms, a diffusion coefficient of  $D_0 = 0.23 \mu\text{m}^2 \text{s}^{-1}$  (experimentally found as compared to theory  $D_0 = 0.20 \mu\text{m}^2 \text{s}^{-1}$ ) and a velocity of  $v_{\text{th}} = 1.3 \mu\text{m s}^{-1}$ . A factor  $c = 2$  accounts for the fact that the particle travels twice the distance due to the feedback delay of one exposure time  $\Delta t_{\text{exp}}$ .

**Multiple active particles, swarms and structures.** Multiple particle control is introduced by illuminating multiple particles at suitable positions at their respective circumferences. In the current setup, an acousto-optic deflector multiplexes the focused heating laser spot between different particle positions within one exposure of duration  $\Delta t_{\text{exp}}$  (Methods). The incident heating power is therefore available for a time  $\Delta t_{\text{exp}}/N$  to each of the  $N$  particles and the average heating power per particle decreases when keeping the overall incident laser power constant. Figure 2a depicts the control of six individual active particles in a spatially fixed pattern of six target positions, arranged as the nodes of a symmetric hexagon (Supplementary Movie 2). The particles were initially distributed randomly in the field of view. Once the control was initiated, each of the particles was first driven towards its nearest target, after which it was confined there. The resulting





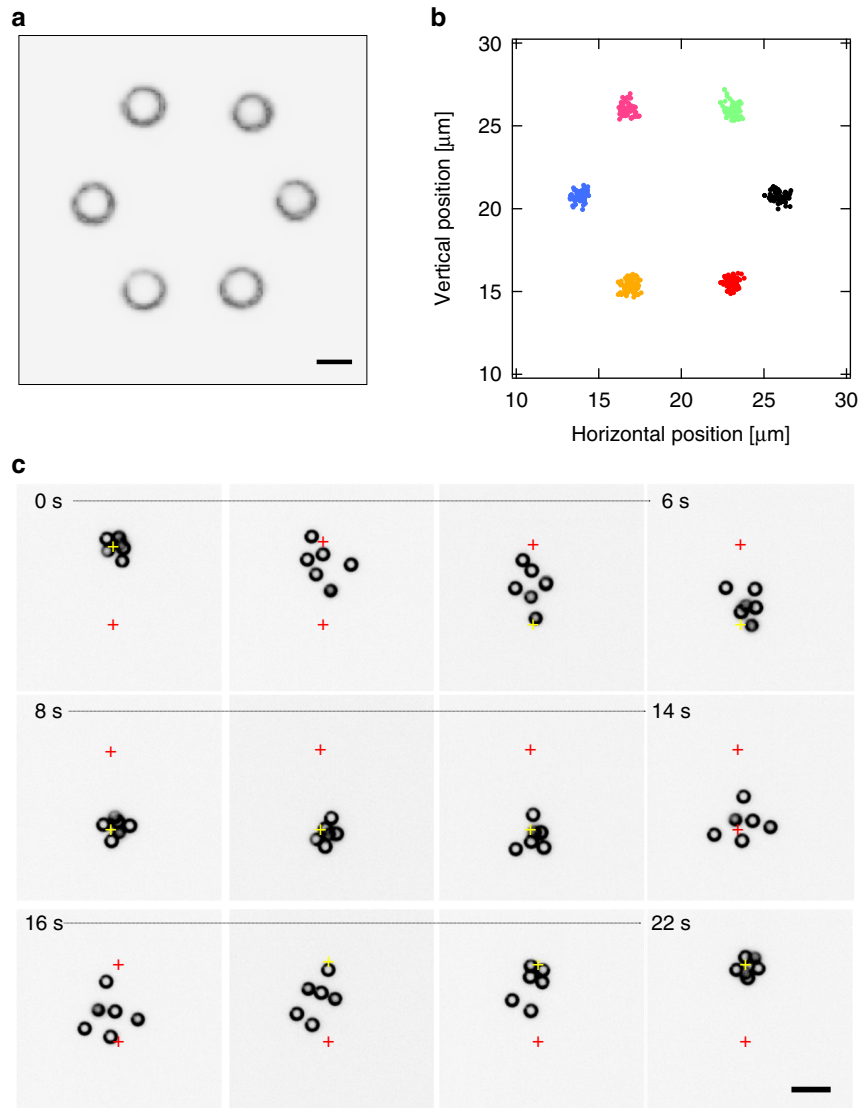
**Fig. 1** Symmetric self-thermophoretic active particle and its properties. **a** Scheme of the self-thermophoretic active particle composed of a melamine resin particle ( $R = 1.09 \mu\text{m}$ ) covered at 30% surface area by 10 nm gold nanoparticles. The nanoparticles can be heated by an incident laser beam (beam waist  $\omega_0$ ) to generate an inhomogeneous temperature profile along the particle surface. This profile causes a thermo-osmotic slip flow propelling the particle at  $v_{\text{th}}$  away from the laser. The velocity of the particle is determined by the displacement  $\delta x$  of the laser focus from the particle center. The scale bar has a length of  $1 \mu\text{m}$ . **b** The particle velocity and control accuracy is derived from an experiment driving the particle between two target positions and confining it for 100 frames at each position. Example trajectory points (laser positions (green), particle position during trapping (orange), particle positions during driving (red)) (see Supplementary Movie 1). **c** Extracted particle speed as a function of the heating power. The error bars reflect the standard deviations from at least driving periods as shown in Supplementary Fig. 5. The nonlinear dependence is analyzed in the Supplementary Information. **d** Control accuracy as a function of the particle speed determined from the experiment in Fig. 1b. The upwards and downwards triangles correspond to the data from the left and right target, respectively. The dashed lines are the contributions from the 2-dimensional sedimentation (red) and the particle overshooting (black). The solid black curve represents the sum of both contributions

steady state distribution of the particles overlaps with the distribution of the assigned target coordinates, as shown in Fig. 2b. The accuracy of the particle control follows the dependencies revealed earlier in Fig. 1. In contrast to optical tweezers, the applied scheme does not involve external forces, but just dissipative fluxes and is thus physically different from common trapping<sup>24</sup>. An active particle always dissipates energy even if it appears stationary, i.e., in a collision with a wall, while this is not the case for a particle driven by external forces through a liquid. Yet, the stationary position distribution of each particle around its target may be realized by external forces<sup>25,26</sup> and appears in the example of Fig. 2 to be similar to what has been achieved with the help of holographic optical tweezers<sup>27,28</sup>.

A collection of particles may be driven to a target either by first arranging them into a structure and subsequently translocating the structure, or by driving each particle directly to the target without prior structural arrangement, as demonstrated in the Supplementary Movies. While the former approach results in a

collective motion resembling the transport of an organized fleet of vehicles (see Supplementary Movie 2), in the latter case the unstructured collective driving results in a swarm-like motion determined by the active particles steric repulsion, Brownian motion and the propulsion towards the target (see Fig. 2c). In addition to the switching of propulsion directions, the local modulation of the propulsion speed may also lead to a well-controlled effective potential sculpting the particle probability density distributions according to  $p(\mathbf{r}) \propto 1/v_{\text{th}}(\mathbf{r})$  much like in the motility-induced phase transitions observed in active particle ensembles<sup>29,30</sup>.

**Self-organized active particle molecules.** The described multiple particle control scheme opens the realm of feedback-induced virtual interactions for active particles. Here, particles are allowed to exchange information via the real-time tracking feedback loop of the microscopy system. It is similar to a situation where birds



**Fig. 2** Controlling multiple self-thermophoretic active particles. **a** Darkfield microscopy snapshot (inverted grayscale) of six active particles ( $R = 1.09 \mu\text{m}$ ) arranged at the nodes of a symmetric hexagon ( $7.1 \mu\text{m}$  edge length) with the help of the particle control procedure described in the text. The incident laser power per particle is  $P = 0.2 \text{ mW}$ . The scale bar has a length of  $2 \mu\text{m}$ . **b** Corresponding trajectory points of the particles over a time period of 11 s with  $\Delta t_{\text{exp}} = 110 \text{ ms}$  exposure time/inverse frame rate. **c** Dark-field microscopy image series (inverted grayscale) of six active particles driven between two target positions separated by  $14.2 \mu\text{m}$ . Targets are colored for clarity. The incident heating power per particle is  $P = 0.2 \text{ mW}$  and the time resolution of the experiment is  $\Delta t_{\text{exp}} = 80 \text{ ms}$ . The scale bar has a length of  $7 \mu\text{m}$

or fish react to the action of their neighbors to form flocks and schools or to escape predators. It introduces a signaling channel between the particles, which can be tweaked almost arbitrarily to design virtual interactions and paves the way for a vast amount of studies from the self-organization of new structures to the information flow in flocks<sup>9</sup>, or the application of machine learning to study adaption and the emergence of collective patterns<sup>31</sup>. Moreover, it provides a minimal scalable robotic system with a simple propulsion and intrinsic noise due to Brownian motion. Here, we demonstrate the structure formation by defining a pairwise control, which intends to keep the active particles at a prescribed separation distance  $r_{\text{eq}}$  by just changing their propulsion direction, but not the speed. If the in-plane distance  $r_{ij}$  between two particles ( $i$  and  $j$ ) is below the separation distance  $r_{\text{eq}}$ , the particles are pushed away from each other, each with a speed  $v_{\text{th}}$ . In the case  $r_{ij} > r_{\text{eq}}$ , the particles are pushed towards each other with the same speed, which results in an

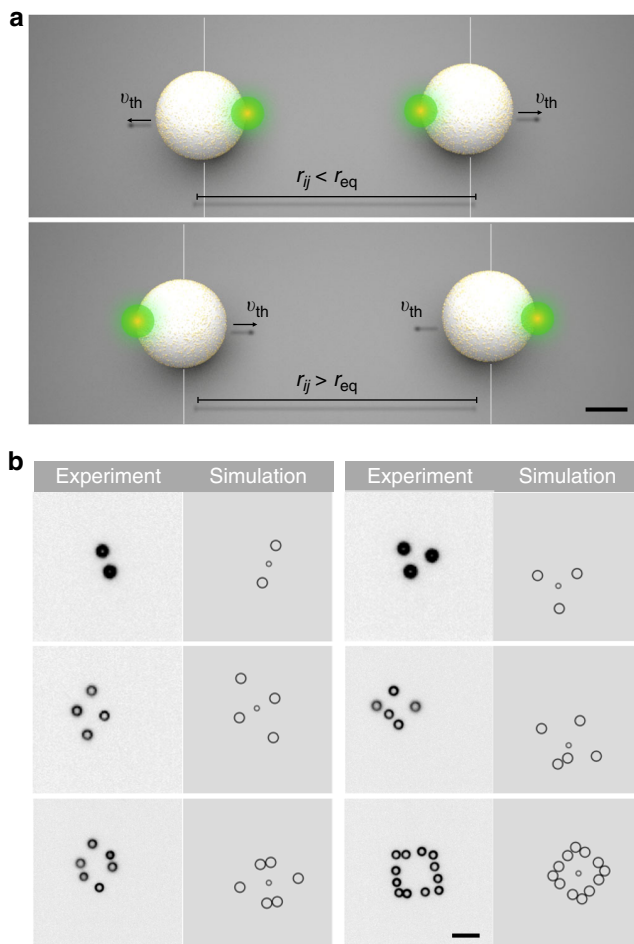
effective V-shaped interaction potential for a pair of active particles. For a number of  $N$  interacting particles, this feedback rule is represented for particle  $i$  by the velocity

$$\mathbf{v}_i(t) = -v_{\text{th}} \mathbf{e}_i(t), \quad (1)$$

where the propulsion direction is determined by

$$\mathbf{e}_i(t) = \frac{\sum_{j \neq i}^N \text{sign}(r_{ij}(t - \delta t) - r_{\text{eq}}) \mathbf{e}_{ij}}{\left| \sum_{j \neq i}^N \text{sign}(r_{ij}(t - \delta t) - r_{\text{eq}}) \mathbf{e}_{ij} \right|} \quad (2)$$

with  $r_{ij} = |\mathbf{r}_j(t - \delta t) - \mathbf{r}_i(t - \delta t)|$  and  $\mathbf{e}_{ij} = (\mathbf{r}_j(t - \delta t) - \mathbf{r}_i(t - \delta t))/r_{ij}$ . As stated by Eq. (1), the speed of motion is always  $v_{\text{th}}$ , while its direction  $\mathbf{e}_i(t)$  is defined by the positions of the other active



**Fig. 3** Feedback rule and self-organized active particle structures. **a** Pair interaction rule used to “bind” the active particles to each other. Particles are propelled with a speed  $v_{th}$ . If the distance is below  $r_{eq}$ , the particles are moved away from each other, if above, they are propelled towards each other. The scale bar has a length of  $1\ \mu\text{m}$ . **b** Example snapshots of 6 different “active particle molecules” that are bound by the interaction rule in **a**. The structures are highly dynamic (see Supplementary Movies 4–9). The scale bar has a length of  $7\ \mu\text{m}$ . The experimental images are compared to snapshots of corresponding numerical simulations involving Brownian motion and a simple delayed feedback as in the experiment (not drawn to scale, see Supplementary Note 2). The smaller central circle marks the center of mass. An effective potential description (not included) also reveals equivalent structures up to the pentamer

particles. The interaction rule is implemented in a real-time particle tracking loop (Methods) detecting the center positions of the active particles and steering the heating laser to the corresponding spots at their circumferences. Position measurements and action are separated by one exposure time  $\Delta t_{exp}$  introducing the feedback delay  $\delta t = \Delta t_{exp}$ .

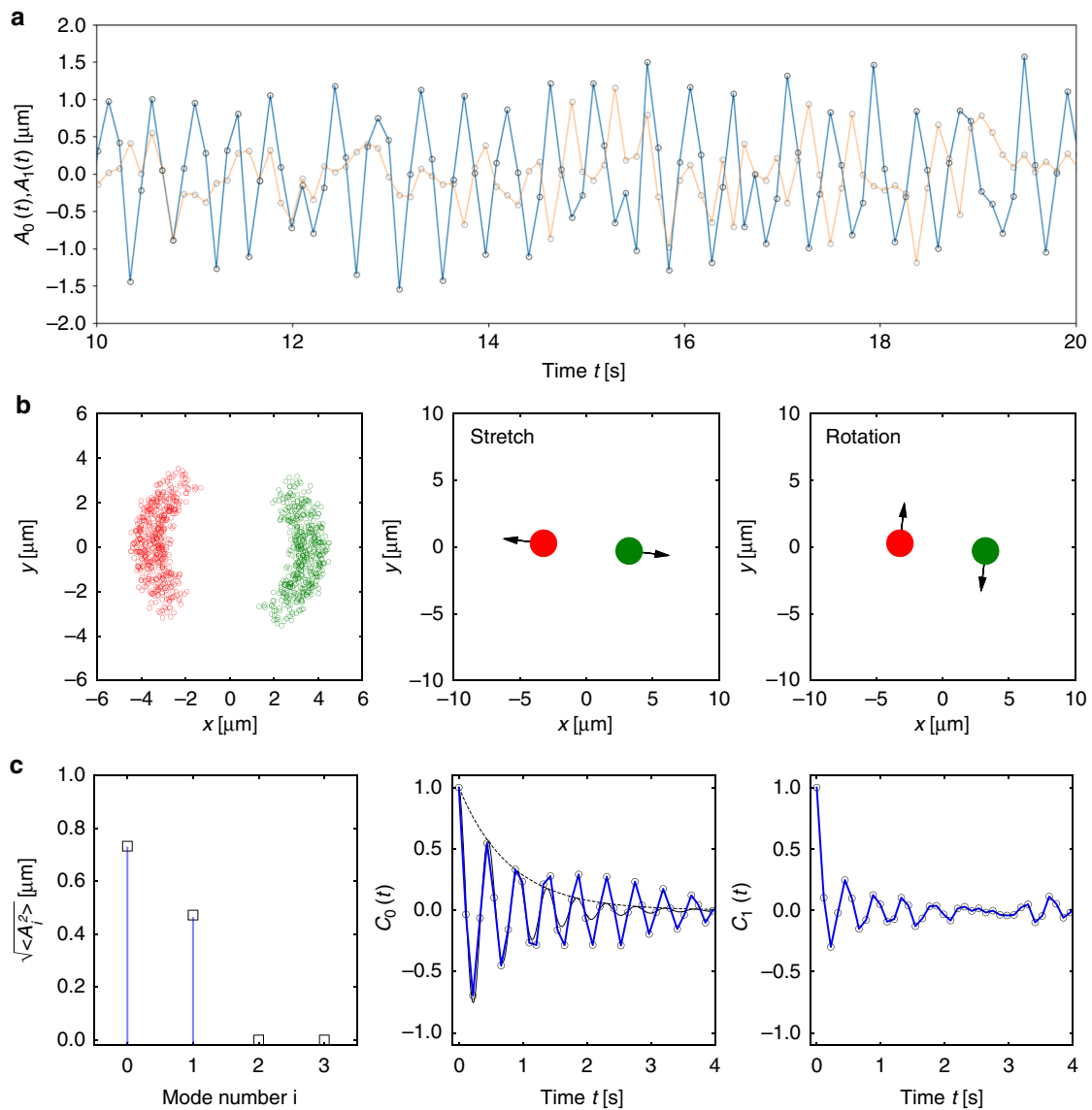
After a short self-organization phase, the particles arrange into freely diffusing dynamic structures that are reminiscent of simple molecules (see Supplementary Movies 4–9). They are not prescribed by a given potential energy landscape, but arrange due to the defined rules. Figure 3 depicts snapshots of 6 self-organized structures all with  $r_{eq} = 7.1\ \mu\text{m}$ . Active particle molecules with  $N = 2$  and  $N = 3$  particles (dimer, Fig. 4 and trimer, Fig. 5a) form structures where all average interparticle distances correspond to the adjusted value of  $r_{eq}$  (measured  $\langle r_{ij} \rangle = 7.23\ \mu\text{m}$ ). All clusters with a larger number of particles ( $N \geq 4$ )

are structurally frustrated due to the confinement in two dimensions. Four active particles cannot form a tetrahedral structure with equilateral triangular faces. Instead a 2-dimensional structure with two “isomers” is found (Fig. 5b, shown together with a timetrace of the isomerization process in Fig. 5c). Larger active particle molecules ( $N > 4$ ) form frustrated structures with interparticle separations of less than the defined value  $r_{eq}$ . Figure 3b highlights snapshots of a pentamer (middle right), a hexamer (bottom left) and a dodecamer (bottom right). The dodecamer, for example, forms a transient square structure out of 4 right angle subunits of 3 particles.

A detailed picture on the experimentally observed dynamics of the active particle structures is obtained from a principle component analysis (PCA)<sup>32</sup> of the displacement vectors of the individual particles between successive frames in the center of mass frame  $\Delta \vec{r}_i^{COM}$ . The identified eigenvectors correspond to a set of  $2N - 2$  orthogonal directions with the largest displacement variances in the COM frame. Figure 4b presents the two modes for the dimer structure (stretch and rotation,  $N = 2$ ) and Fig. 5a the four modes of the trimer (symmetric stretch, bending mode, asymmetric stretch, and rotation,  $N = 3$ ) where the modes agree well with the normal modes of an equilateral triangular structure in two dimensions. The appearance of a rotational mode is at first glance surprising as the feedback is designed to act along the connecting line between the bound particles. Due to the feedback delay and Brownian motion, the laser heating position along the circumference is fluctuating as well and introducing a coupling between the translational and rotational motion. The oscillatory motion as indicated for the two modes of the dimer (Fig. 4a) is a fundamental feature of a time-delayed negative feedback system and inherent to electronic oscillators, but also appearing at all levels of biological systems<sup>33</sup>. In our active particle assemblies the dynamics is controlled by three parameters, the feedback delay  $\delta t$ , the propulsion speed  $v_{th}$ , and the single particle diffusion coefficient  $D_0$ . Using these parameters and restricting the analysis to the dynamics of the dimer along the connecting line (stretch mode), we can model the stretch mode dynamics with an overdamped Langevin description,

$$\dot{r}_{12}(t) = -2v_{th} \text{sign}(r_{12}(t - \delta t) - r_{eq}) + \sqrt{4D_0}\eta_{12}(t). \quad (3)$$

Here  $\eta_{12}(t)$  is a zero-mean, unit-variance Gaussian white noise, i.e.,  $\langle \eta_{12}(t) \rangle = 0$  and  $\langle \eta(t)\eta(t') \rangle = \delta(t - t')$ , such that the variance of the noise term in Eq. (3) corresponds to  $4D_0$ . As the bond length involves the relative motion of two particles, the relative velocity  $2v_{th}$  and the relative diffusion coefficient  $2D_0$  enter the equation. Its solution yields the observed oscillatory motion with a triangular shape and an oscillation period of  $T = 4\delta t$  (Supplementary Note 3). Accordingly, the oscillatory dynamics vanishes for zero feedback delay and stiff structures appear. The amplitude of the motion linearly depends on the delay and the active particle velocity ( $\Delta r_{12} = 2v_{th}\delta t$ ). A feedback delay of  $\delta t = 0.11\ \text{s}$  and a propulsion velocity  $v_{th} = 3.4\ \mu\text{m s}^{-1}$  ( $P_{heat} = 0.75\ \text{mW}$  per particle) delivers  $T = 0.44\ \text{s}$  and  $\Delta r_{12} = 0.75\ \mu\text{m}$ , which compares well to the experimental data for the dimer  $T_{exp} = 0.44\ \text{s}$  and the amplitude of the stretch mode  $\sqrt{\langle A_0^2 \rangle} = 0.72\ \mu\text{m}$  obtained from the first eigenvalue of the PCA,  $A_0$ . The effect of the last term in Eq. (3) is to introduce phase and amplitude noise to the oscillatory motion. The oscillations are therefore losing coherence and the autocorrelation  $C_i(t) = \langle A_i(\tau) A_i(\tau + t) \rangle / \langle A_i(\tau)^2 \rangle$  of the oscillating modes  $A_i(t)$  decays. The timescale of this damping is the dephasing time, which is termed  $T_2$  in molecular spectroscopy. Using an approximate solution of Eq. (3) for the dimer (see Supplementary Note 3), we find a dephasing time  $T_2 \approx 32\Delta r_{12}^2\pi^{-4}D_0^{-1}$ , which scales inversely with the strength of the noise given by the diffusion coefficient (see



**Fig. 4** Dynamics of a self-organized active particle dimer. **a** Change of the dimer bond length parallel (blue) and perpendicular (orange) to the connecting line of the two particles. The dimer bond length is oscillating with a triangular shaped elongation with a period of  $T = 0.44$  s. The period corresponds to four times the feedback delay time  $\delta t = 0.11$  s. **b** Left: trajectory points of the two bound particles in the center of mass (COM) frame. Middle and right: principle components of the particle displacements in the center of mass frame as obtained from a principle component analysis (PCA). **c** Left: principle component amplitudes as calculated for the two modes in **b**. Middle and right: autocorrelation functions for the displacements of the particles in the center of mass frame of the two eigenvectors obtained from the PCA. The two modes reveal a damped oscillation due to the Brownian motion of the active particles. The middle graph shows in addition the theoretical prediction for the oscillation (black solid line) and the exponential decay due to the dephasing (black dashed line)

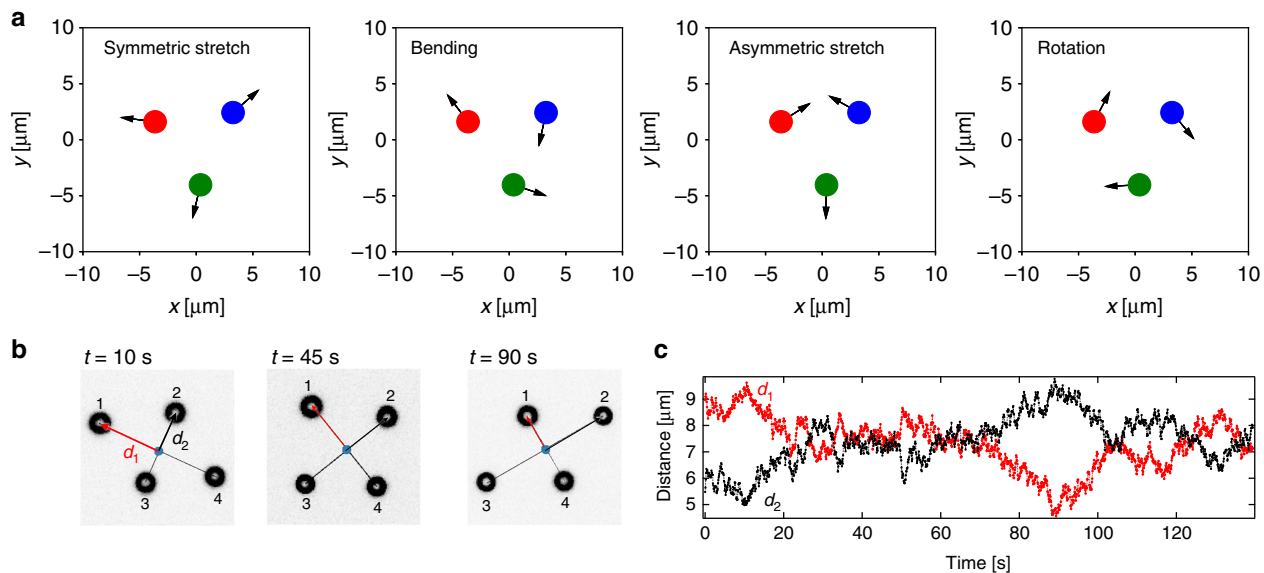
Supplementary Note 3). For the dimer bond length oscillation displayed in Fig. 4c we find a dephasing time of  $T_2 = 0.8$  s (dashed line). In larger structures, each particle contributes to the total noise such that coherent oscillations as observed for the dimer disappear quickly with growing size of the cluster.

## Discussion

The presented symmetric active particles and the introduced manipulation technique allow us to self-assemble structures by designed feedback controlled interaction rules. Similar structures of passive colloidal particles have been assembled in previous work with the help of external forces in optical tweezers setups<sup>27,28</sup>. The structures there are commonly based on a prescribed optical potential energy landscape in which the particles

occupy energetic minima and fluctuate according to Boltzmann statistics in equilibrium.

The assemblies created in our experiments are conceptually different as there is no external force acting on the particles<sup>17</sup> and more importantly, the structures do only exist in nonequilibrium. The particles need to be propelled continuously to form the assemblies much like living systems need nonequilibrium to maintain their shape and function. The fluctuations of the structures are nonequilibrium even for a vanishing feedback delay and thus they are not tied to a well-defined temperature or Boltzmann statistics. Moreover, as demonstrated for the oscillatory motion, non-zero delay in our setup leads to additional oscillations/fluctuations determined by the propulsion speed and feedback delay.



**Fig. 5** Principle components of the trimer and isomerization of the tetramer. **a** The principle components of the trimer as calculated from a principle component analysis (PCA) of the particle displacements in the center of mass coordinate system. Depicted are the symmetric stretch, bending motion, asymmetric stretch, and the rotation sorted by the magnitude of the eigenvalue found in the PCA. **b** Isomerization of the tetramer. The images represent the observed dark-field images of the 4 active particles at three different times ( $t = 10$  s,  $t = 45$  s,  $t = 90$  s). Particles have been labeled to keep track of their position. **c** Timetrace of the distance  $d_1$  of particle 1 and  $d_2$  of particle 2 from the center of mass indicating the isomerization. The distances  $d_1$  and  $d_2$  exchange in size after a time period of about 75 s. The isomerization is driven by Brownian motion

In artificial active particle systems several sources of structure formation and local density enhancement may exist. Active particles are observed, for example, to cluster based on chemical, composition, or temperature gradients mimicking tactic behavior<sup>34</sup>. In the case of chemotaxis of artificial microswimmers the propulsion speed is modulated as direct physical consequence of the local concentration of a substance. In fact, a modulation of the propulsion speed will lead to local enhancement or depletion of active particle concentrations<sup>35</sup> in steady states as it is observed also in motility induced phase transitions with a mutual locking of particles<sup>3,36</sup>.

With the feedback rule applied in our experiments, we now remove also the modulation of the speed as particles are constantly propelled with the same speed. Yet there is a structure forming contribution resulting from the feedback loop. It stems from the fact that we determine the particle positions in a measurement and decide on the required action, i.e., the velocity direction. It is this particular processing of the information of the particle positions which extracts entropy from the system to form the structures. Analyzing the entropy production in the system, we can identify several contributions (see Supplementary Fig. 6). On one side, there is the entropy production rate maintaining the temperature gradients. This “housekeeping” entropy production amounts to the fraction of the incident laser power absorbed by the gold nanoparticles  $P_a$  divided by the temperature of the surrounding liquid  $T$  to which this power is dissipated. It is required for the mobility although only a fraction of it is used during propulsion<sup>37</sup>. This entropy flux is still not sufficient for the structure formation. The propulsion speed and the dissipated power are constant over the trajectory of the particle no matter if it is driven on a random path or is bound in the active particle molecule. Consequently, it is not a spatial modulation of this dissipation which is causing the structure formation.

The feedback process is extracting entropy from the system by steering the propulsion direction and not modulating their speed. A theoretical analysis of the entropy fluxes can be obtained by relating the feedback process to the thermodynamics of resetting

processes as described by Fuchs et al.<sup>38</sup> (see Supplementary Note 4). According to that, it is the continuous loss of structure due to Brownian motion which one has to correct for with the feedback. To form a stable stationary structure, the entropy extracted per time unit in the feedback loop has to compensate at least the increase of entropy per time unit due to Brownian motion. The structure formation is thus the result of the information flow in the feedback loop only. Overall, the structure forming entropy production rate is very small as compared to all other entropy fluxes (see Supplementary Note 4) but sufficient to create well-defined active particle assemblies.

Returning to the previous example of chemotaxis, one may argue that the response of an artificial active particle to a chemical gradient may be seen as an information based interaction as well even if it is the direct physical consequence of osmotic pressure differences. The local chemical gradient is the information that is available to the particle and causes a response in form of a modulated propulsion speed. While this interpretation of physical interactions may also be valid, we restrict our definition of information based interactions to situations, where the physical interactions are irrelevant. In the well-known thought experiment of Maxwell, for example, a small daemon uses the information on the speed of particles to sort them into two boxes (slow and fast) just by actuating a shutter between the boxes. The physical interactions of the particles are irrelevant there for the appearing temperature difference between the two boxes. A correct physical description not violating the laws of thermodynamics, however, requires to include the entropy of information the daemon uses. In a very similar way, it is the information flow in the feedback loop that is required for a correct physical description of the assembled active particle molecules in our experiment.

In conclusion, propulsion speed, Brownian motion as well as the feedback related information flow shape the morphology and dynamics of these artificial self-organized active structures. They require nonequilibrium conditions to exist and contrary to structures assembled by optical tweezers they are not bound to Boltzmann statistics, equipartition or global detailed balance.

Nevertheless, the dynamics of the structures carries a number of features, which are found in equilibrium as well. Using the described method, almost any type of interaction can be designed to create large scale interacting assemblies or new self-organized shapes which may not be accessible by conventional interactions, i.e., which do not need to obey the action–reaction principle. Fundamental interaction and signaling rules for emergent complex behavior of cells up to animals may be explored with our model system in the same way as concepts of nonequilibrium thermodynamics. The applied technique provides a direct interface to machine learning algorithms including predictive information or reinforcement learning. The details of information flows in large ensembles may be studied easily and can be connected to different timescales of delayed information processing. Especially the latter type of application including coupled active feedback networks with different inherent timescales shall ignite a vast variety of research on emergent collective and crowd dynamics.

## Methods

**Sample preparation.** Samples consist of commercially available gold nanoparticle coated melamine resin particles of a diameter of 2.13  $\mu\text{m}$  (microParticles GmbH, Berlin, Germany). The gold nanoparticles are covering about 30% of the surface and are between 8 and 30 nm in diameter (Supplementary Fig. 1). Glass cover slips have been dipped into a 5% Pluronic F127 solution, rinsed with deionized water and dried with nitrogen. The Pluronic F127 coating prevents sticking of the particles to the glass cover slides. Two microliters of particle suspension are placed on the glass cover slides to spread about an area of 1 cm  $\times$  1 cm to form a 3  $\mu\text{m}$  thin water film. The edges of the sample have been sealed with silicone oil to prevent water evaporation.

**Microscopy setup.** Samples have been investigated in a custom-built inverted microscopy setup (Supplementary Note 5). The setup is based on an Olympus IX 71 microscopy stand. Optical heating of the active particles is carried out by a CW 532 nm laser. The laser intensity is controlled by a Conoptics 350–50 electro-optical modulator. An acousto-optic deflector (AOD) together with a 4-f system (two  $f = 20$  cm lenses) is used to steer the 532 nm wavelength laser focus in the sample plane. The AOD is controlled by an FPGA (National Instruments) via a LabVIEW program. The calibration of the AOD for precise laser positioning is carried out using a 2D projection method. A Leica 100x, infinity-corrected, NA 1.4–0.7 (set to 0.7), HCX PL APO objective lens is used for focusing the 532 nm laser to the sample plane as well as for imaging the active particles. Active particles are imaged under dark-field illumination using an oil immersion dark-field condenser. The scattered light from the sample is collected with the Leica objective lens and imaged with a  $f = 30$  cm tube lens to an emCCD camera (Cascade 650). A region of interest (ROI) of  $200 \times 200$  pixels is utilized for the real time imaging, analysis and recording of the particles, with an exposure time  $\Delta t_{\text{exp}}$  of 0.08 s or 0.110 s.

**Single particle tracking and real-time feedback loop.** The active particles appear as rings in dark-field microscopy images and are tracked in real time in a LabVIEW program. A Matlab node in the LabVIEW determines the centers of the particles using a Hough transform function of Matlab. The particle coordinates are used to calculate the position of the laser focus for each individual particle. During one camera exposure of  $\Delta t_{\text{exp}}$ , the laser is shared among the particles selected for feedback control. The switching is done using the AOD as mentioned above.

For the control experiments (Fig. 1), the controlled particle is driven back-and-forth between two prescribed target positions. Upon reaching a target, it is actively positioned there for 100 frames first, before being driven to the the target. This procedure is repeated several times (5–7) with different laser powers up to 1 mW. The particle positions around the targets are used to determine the localization error  $\sigma = \sqrt{\langle \delta r^2 \rangle}$ , where  $\delta r = r - r_t$  is the 2-d position vector from the target to the particle. Velocities are determined by projecting the particle displacement between two subsequent frames onto the unit vector given by the laser position and the particle center  $\langle v \rangle = \langle (r(t + \Delta t_{\text{exp}}) - r(t)) \cdot e_{\text{lp}} \rangle / \Delta t_{\text{exp}}$ , where  $e_{\text{lp}} = (r - r_{\text{laser}}) / |r - r_{\text{laser}}|$ .

**Principle component analysis of the dynamics.** The experiments sample the position vector  $r_i(t)$  for each of the  $N$  particles in  $N_{\text{steps}} = t_{\text{meas}} / \Delta t_{\text{exp}} + 1$  time instants. The particle coordinates are converted into the center of mass frame of the structure. In two dimensions, we obtain  $N_{\text{data}} = 2 \times N \times N_{\text{steps}}$  data points, which we analyze using the principal component analysis.

First, we construct the  $2N$  time-series of displacements of the individual degrees of freedom in our experiment (coordinates of the individual  $N$  particles). Second, we put all these displacements for a given  $t$  into a single vector:  $S(t) = (\Delta x_1(t), \Delta y_1(t), \dots, \Delta x_N(t), \Delta y_N(t))$ . Then, we construct the matrix  $X$  containing in its lines the vectors  $S(t)$  corresponding to the individual measurement times, so the element

$[i, j]$  of this matrix is given by  $X_{ij} = S_j((i - 1)\Delta t_{\text{exp}})$ . The matrix  $X$  thus has  $2N$  columns and  $N_{\text{steps}} - 1$  rows. The matrix  $X$  is used to calculate the covariance matrix  $M = X^T X$ . This matrix  $M$  is a symmetric  $2N \times 2N$  matrix with the elements  $M_{ij} = \sum_{k=1}^{N_{\text{steps}}-1} X_{ki} X_{kj} = \sum_{k=1}^{N_{\text{steps}}-1} S_i((k - 1)\Delta t_{\text{exp}}) S_j((k - 1)\Delta t_{\text{exp}})$ . The diagonal thus contains the variances corresponding to the degrees of freedom measured in the experiment, i.e., of the displacements of the positions of the individual particles.

We determine all  $2N - 2$  nonzero eigenvalues  $A_i$  and normalized eigenvectors  $V_i$  of the matrix  $M$ .

The individual eigenvectors determine new  $2N - 2$  collective degrees of freedom, which are mutually independent. For example for the dimer, the eigenvector with the largest eigenvalue determines the vibrational mode and the corresponding collective coordinate is proportional to the vector connecting the two particles.

The vector form of the time series corresponding to the mode given by the column eigenvector  $V_i$  can be obtained by projecting the eigenvector  $V_i$  onto the matrix  $X$ , i.e.,  $K_i = X \cdot V_i$ . The elements of the  $K_i$  represent the time series  $A_i(t)$  of the motion along the eigenvector  $V_i$ . To access the dynamics of the mode we calculate its autocorrelation  $C_i(t) = \langle A_i(\tau) A_i(\tau + t) \rangle / \langle A_i(\tau)^2 \rangle$ . The subscript  $\tau$  denotes that the correlation function is obtained from the time average (Supplementary Note 2).

## Data availability

All data is available from the corresponding author on request.

Received: 16 March 2018 Accepted: 6 September 2018

Published online: 21 September 2018

## References

1. Bechinger, C. et al. Active particles in complex and crowded environments. *Rev. Mod. Phys.* **88**, 045006 (2016).
2. Palacci, J., Sacanna, S., Steinberg, A. P., Pine, D. J. & Chaikin, P. M. Living crystals of light-activated colloidal surfers. *Science* **339**, 936–940 (2013).
3. Buttinoni, I. et al. Dynamical clustering and phase separation in suspensions of self-propelled colloidal particles. *Phys. Rev. Lett.* **110**, 238301 (2013).
4. Solon, A. P., Fily, Y., Baskaran, A., Cates, M. E. & Kafri, Y. Pressure is not a state function for generic active fluids. *Nat. Phys.* **11**, 673–678 (2015).
5. Theurkauff, I., Cottin-Bizonne, C., Palacci, J., Ybert, C. & Bocquet, L. Dynamic clustering in active colloidal suspensions with chemical signaling. *Phys. Rev. Lett.* **108**, 268303 (2012).
6. Fily, Y. & Marchetti, M. C. Athermal phase separation of self-propelled particles with no alignment. *Phys. Rev. Lett.* **108**, 235702 (2012).
7. Pearce, D. J. G., Miller, A. M., Rowlands, G. & Turner, M. S. Role of projection in the control of bird flocks. *Proc. Natl Acad. Sci. USA* **111**, 10422–10426 (2014).
8. Ballerini, M. et al. Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. *Proc. Natl Acad. Sci. USA* **105**, 1232–1237 (2008).
9. Attanasi, A. et al. Information transfer and behavioral inertia in starling flocks. *Nat. Phys.* **10**, 691–696 (2014).
10. Miller, M. B. & Bassler, B. L. Quorum sensing in bacteria. *Ann. Rev. Microbiol.* **55**, 165–199 (2001).
11. Tkacik, G., Callan, C. G. & Bialek, W. Information flow and optimization in transcriptional regulation. *Proc. Natl Acad. Sci. USA* **105**, 12265–12270 (2008).
12. Micali, G. & Endres, R. G. Bacterial chemotaxis: information processing, thermodynamics, and behavior. *Curr. Opin. Microbiol.* **30**, 8–15 (2016).
13. Landauer, R. The physical nature of information. *Phys. Lett. A* **217**, 188–193 (1996).
14. Katz, Y., Tunström, K., Ioannou, C. C., Huepe, C. & Couzin, I. D. Inferring the structure and dynamics of interactions in schooling fish. *Proc. Natl Acad. Sci. USA* **108**, 18720–18725 (2011).
15. Swain, D. T., Couzin, I. D. & Ehrlich Leonard, N. Real-time feedback-controlled robotic fish for behavioral experiments with fish schools. *Proc. IEEE* **100**, 150–163 (2012).
16. Berdahl, A., Torney, C. J., Ioannou, C. C., Faria, J. J. & Couzin, I. D. Emergent sensing of complex environments by mobile animal groups. *Science* **339**, 574–576 (2013).
17. Kroy, K., Chakraborty, D. & Cichos, F. Hot microswimmers. *Eur. Phys. J. Spec. Top.* **225**, 2207–2225 (2016).
18. Selmeke, M., Khadka, U., Bregulla, A. P., Cichos, F. & Yang, H. Theory for controlling individual self-propelled micro-swimmers by photon nudging I: directed transport. *Phys. Chem. Chem. Phys.* **20**, 10502–10520 (2018).
19. Selmeke, M., Khadka, U., Bregulla, A. P., Cichos, F. & Yang, H. Theory for controlling individual self-propelled micro-swimmers by photon nudging II: confinement. *Phys. Chem. Chem. Phys.* **4**, 1–12 (2018).

20. Jiang, H. R., Yoshinaga, N. & Sano, M. Active motion of a Janus particle by self-thermophoresis in a defocused laser beam. *Phys. Rev. Lett.* **105**, 268302 (2010).
21. Bregulla, A. P., Würger, A., Günther, K., Mertig, M. & Cichos, F. Thermodynamic flow in thin films. *Phys. Rev. Lett.* **116**, 188303 (2016).
22. Bickel, T., Majee, A. & Würger, A. Flow pattern in the vicinity of self-propelling hot Janus particles. *Phys. Rev. E* **88**, 012301 (2013).
23. Bregulla, A. P., Yang, H. & Cichos, F. Stochastic localization of microswimmers by photon nudging. *ACS Nano* **8**, 6542–6550 (2014).
24. Qian, B., Montiel, D., Bregulla, A., Cichos, F. & Yang, H. Harnessing thermal fluctuations for purposeful activities: the manipulation of single microswimmers by adaptive photon nudging. *Chem. Sci.* **4**, 1420–1429 (2013).
25. Jun, Y. & Bechhoefer, J. Virtual potentials for feedback traps. *Phys. Rev. E* **86**, 061106 (2012).
26. Braun, M., Bregulla, A. P., Günther, K., Mertig, M. & Cichos, F. Single molecules trapped by dynamic inhomogeneous temperature fields. *Nano Lett.* **15**, 5499–5505 (2015).
27. Curtis, J. E., Koss, B. A. & Grier, D. G. Dynamic holographic optical tweezers. *Opt. Commun.* **207**, 169–175 (2002).
28. Padgett, M. & Di Leonardo, R. Holographic optical tweezers and their relevance to lab on chip devices. *Lab Chip* **11**, 1196–1205 (2011).
29. Cates, M. E. & Tailleur, J. When are active Brownian particles and run-and-tumble particles equivalent? Consequences for motility-induced phase separation. *Eur. Phys. Lett.* **101**, 20010 (2013).
30. Cates, M. E. Diffusive transport without detailed balance in motile bacteria: does microbiology need statistical physics? *Rep. Prog. Phys.* **75**, 042601 (2012).
31. Palmer, G. & Yaida, S. Optimizing collective fieldtaxis of swarming agents through reinforcement learning. Preprint at <http://arXiv.org/abs/1709.02379> (2017).
32. Jolliffe, I. T. *Principal component analysis*. (Springer, New York, NY, 1986). Springer series in statistics.
33. Smith, H. *An introduction to delay differential equations with applications to the life sciences*. (Springer, New York, 2011). Texts in Applied Mathematics.
34. Baraban, L., Harazim, S. M., Sánchez, S. & Schmidt, O. G. Chemotactic behavior of catalytic motors in microfluidic channels. *Angew. Chem. Int. Ed.* **52**, 5552–5556 (2013).
35. Schnitzer, M. J. Theory of continuum random walks and application to chemotaxis. *Phys. Rev. E* **48**, 2553–2568 (1993).
36. Palacci, J., Cottin-Bizonne, C., Ybert, C. & Bocquet, L. Sedimentation and effective temperature of active colloidal suspensions. *Phys. Rev. Lett.* **105**, 088304 (2010).
37. Bregulla, A. P. & Cichos, F. Size dependent efficiency of photophoretic swimmers. *Faraday Discuss.* **184**, 381–391 (2015).
38. Fuchs, J., Goldt, S. & Seifert, U. Stochastic thermodynamics of resetting. *Eur. Phys. Lett.* **113**, 60009 (2016).

## Acknowledgements

Discussions with J. Shaevitz (Princeton University), K. Kroy (Universität Leipzig) and help with the sample preparations by D. Cichos (Berlin) are acknowledged. H.Y. and U. K. acknowledge support by the Betty and Gordon Moore foundation (grant # 4741). V.H. is supported by a Humboldt grant of the Alexander von Humboldt Foundation and by the Czech Science Foundation (project No. 17-06716S). F.C. is supported by grant CI 33/16-1 and the CRC TRR 102 “Polymers under multiple constraints” of the German Research Foundation (DFG).

## Author contributions

U.K. and F.C. designed the experiments. U.K. and F.C. performed the experiments and analyzed the data. V.H. contributed to the theoretical analysis. H.Y. provided the experimental equipment. U.K., V.H., H.Y., and F.C. wrote the manuscript. All authors discussed the results and commented on the manuscript.

## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-018-06445-1>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

# Supplementary Information - Active Particles Bound by Information Flows

Utsab Khadka<sup>1</sup>, Viktor Holubec<sup>2,3</sup>, Haw Yang<sup>1</sup>, Frank Cichos<sup>4\*</sup>

<sup>1</sup> *Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA.*

<sup>2</sup> *Institute for Theoretical Physics, Universität Leipzig, 04103 Leipzig, Germany.*

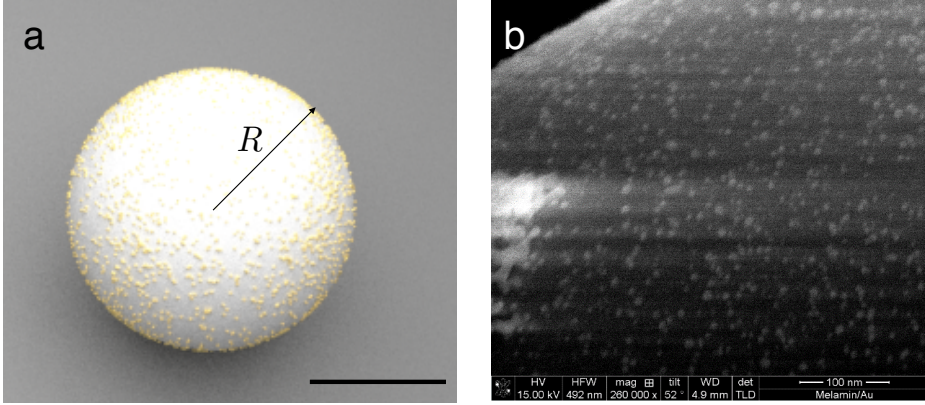
<sup>3</sup> *Charles University, Faculty of Mathematics and Physics, Department of Macromolecular Physics, V Holešovičkách 2, CZ-180 00 Praha, Czech Republic.*

<sup>4</sup> *Peter Debye Institute for Soft Matter Physics, Universität Leipzig, 04103 Leipzig, Germany. E-mail: cichos@physik.uni-leipzig.de*



## Supplementary Note 1: Symmetric Active Particle Velocity

The active particle used throughout the experiments is symmetric in structure, with 30 % of the melamine resin particle surface covered with gold nanoparticles (AuNP). The propulsion velocity  $v_{\text{th}}$  is the result of an asymmetric illumination with a highly focused laser at a wavelength of  $\lambda = 532 \text{ nm}$ , which heats the gold nanoparticles at the surface and thus creates a surface temperature gradient and corresponding thermo-osmotic creep flows. A sketch of the particle and a corresponding electron microscopy image is shown in Supplementary Figure 1.

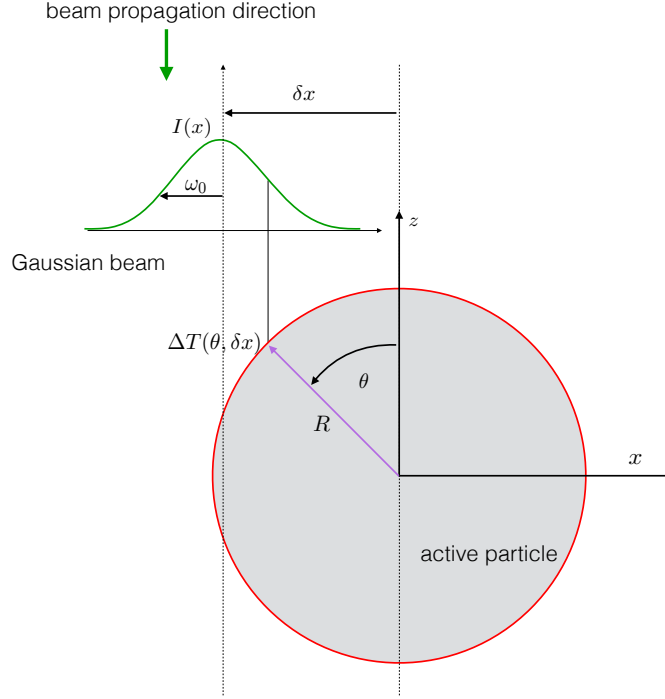


**Supplementary Figure 1** | (a) Sketch of the symmetric active particle of radius  $R$ . For the experiments a melamine particle of  $R = 1.09 \mu\text{m}$  covered with 10 nm gold nanoparticles at 30 % of its surface is heated with a focused laser. The scale bar corresponds to  $1 \mu\text{m}$ . (b) Electron microscopy image of the gold nanoparticle at the surface of the melamine resin particle (kindly provided by Santiago Muiños Landin, Molecular Nanophotonics Group).

**Active particle velocity and laser focus displacement.** As explained in the main text, the magnitude and the direction of the expected active particle velocity depends on the displacement of the heating laser focus from the center of the particle. The displacement of the laser focus from the particle center is also responsible for the nonlinear power dependence of the propulsion velocity. To model this nonlinear dependence, we simplify the 3-dimensional geometry and only consider the 2-dimensional situation sketched in Supplementary Figure 2. The  $z$ -direction denotes the direction perpendicular to the sample plane, while the  $x$ -direction lies in the sample plane. For heating the gold nanoparticles at the surface of the active particle, we assume a Gaussian heating beam with an intensity profile

$$I(x) = I_0 \exp(-(x - \delta x)^2 / 2\omega_0^2). \quad (1)$$

Here  $\omega_0$  is the beam radius and the center of the Gaussian is displaced by  $\delta x$  with respect to the particle center as indicated in Supplementary Figure 2. The beam radius is assumed to be constant in the sample region and thus independent of the  $z$ -position. To obtain an expression for the propulsion velocity as a function of the laser displacement



**Supplementary Figure 2** | Definition of symbols for modeling the nonlinear power dependence of the active particle velocity. An one dimensional intensity profile  $I(x)$  representing a Gaussian beam with a beam waist of  $\omega_0$  is assumed to propagate along the  $z$ -direction. The beam center is displaced by  $\delta x$  from the center of a symmetric swimmer of radius  $R$ . To calculate the resulting propulsion velocity we consider the intensity at an angle  $\theta$  on the particle surface.

$\delta x$ , we need to determine the temperature profile along the particle surface, i.e., the circle circumference in the described model geometry. To do so we project the Gaussian intensity profile to the circle of radius  $R$  by substituting  $x = -R \sin(\theta)$  into supplementary equation 1. Assuming that the temperature increment due to the heating is proportional to the incident laser intensity,  $\Delta T(\theta, \delta x) \propto I(\theta, \delta x)$  we obtain the temperature increase as compared to the ambient temperature  $T_0$  along the circumference of the circle as

$$\Delta T(\theta, \delta x) = T(\theta, \delta x) - T_0 = \Delta T_0 \exp(-(R \sin(\theta) - \delta x)^2 / 2\omega_0^2). \quad (2)$$

$\Delta T_0$  indicates the maximum temperature increase at the particle circumference. The tangential temperature gradient along the circumference is then

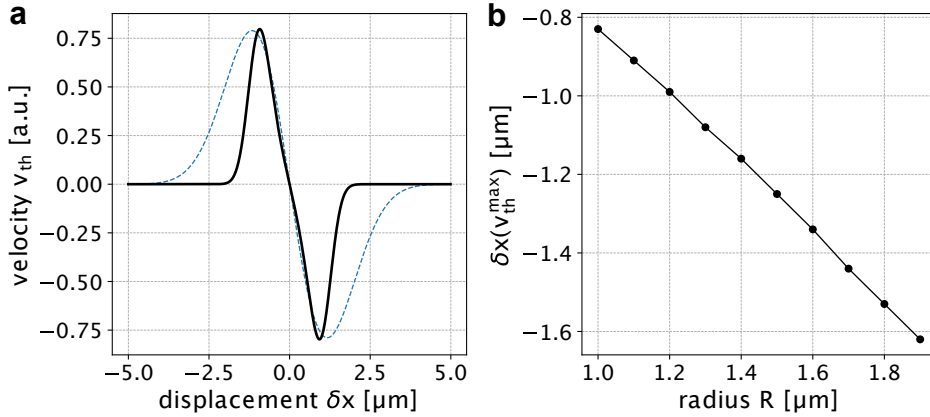
$$\nabla_{\parallel} T = \frac{\partial \Delta T(\theta, \delta x)}{\partial \theta} = -\Delta T(\theta, \delta x) \frac{R \cos(\theta)(R \sin(\theta) - \delta x)}{\omega_0^2}. \quad (3)$$

This tangential temperature gradient leads to a quasi thermo-osmotic slip velocity  $v_s = \chi \nabla_{\parallel} T$  setting the hydrodynamic boundary condition at the particle surface (with the thermo-osmotic mobility coefficient  $\chi$ )<sup>[1,2]</sup>. The propulsion velocity  $v_{\text{th}}$  of the active particle is then

the surface average (circumference average here) of the slip velocity vector (supplementary equation 4). For symmetry reasons all components perpendicular to the  $x$ -axis cancel out and only the components along the  $x$ -direction represented by the additional  $\cos(\theta)$  factor need to be considered. The propulsion velocity along the  $x$ -axis is therefore

$$v_{\text{th}}(\delta x) = \frac{\chi}{2\pi} \oint_{\theta} \Delta T(\theta, \delta x) \frac{R \cos^2(\theta)(R \sin(\theta) - \delta x)}{\omega_0^2} d\theta. \quad (4)$$

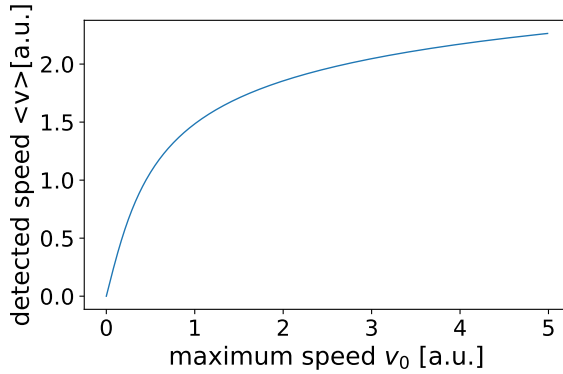
To obtain the dependence of the particle velocity on the laser displacement  $\delta x$ , we numerically integrate supplementary equation 4. Supplementary Figure 3 shows the obtained particle velocity as a function of the beam displacement  $\delta x$  for two different beam radii  $\omega_0$  in the left panel. The right panel is indicating the laser displacement for the maximum particle velocity as a function of the particle radius. Accordingly, to obtain the maximum velocity the laser has to be placed at about a distance corresponding to the particle radius.



**Supplementary Figure 3** | (a) Calculated active particle propulsion velocity as a function of the displacement  $\delta x$  of a Gaussian heating laser beam from the center of the particle according to supplementary equation 4 for  $R = 1.09 \mu\text{m}$  and a beam radius of  $\omega_0 = 0.3 \mu\text{m}$  (solid line) and  $\omega_0 = 1 \mu\text{m}$  (dashed). (b) Dependence of the displacement  $\delta x$  from the particle center for a maximum velocity  $v_{\text{th}}^{\text{max}}$  as a function of the particle radius  $R$ .

**Power dependence of the active particle velocity** The nonlinear power dependence of the velocity of the active particle is the result of the dependence of the velocity on the laser displacement  $\delta x$  and the finite exposure time  $\Delta t_{\text{exp}}$ . Within the exposure time, the laser is placed at the particle rim to cause the maximum velocity according to the previous section. The particle starts to move and the instantaneous velocity drops as the laser is spatially fixed. The average velocity observed is then

$$\langle v_{\text{th}} \rangle = \frac{1}{\Delta t_{\text{exp}}} \int_0^{\Delta t_{\text{exp}}} v_{\text{th}}(\delta x(t)) dt \quad (5)$$



**Supplementary Figure 4** | Numerical solution of supplementary equation 7 showing the average speed as a function of the maximum speed (proportional to the incident heating power) for  $\omega_0 = 1 \mu\text{m}$  and  $\delta x = 1 \mu\text{m}$ . The dependence captures the nonlinear increase of the detected active particle velocity as a function of the incident heating power. The nonlinearity is due to the particle moving out the laser beam during the exposure time.

The displacement  $\delta x$  increases with the instantaneous velocity  $v_{\text{th}}(\delta x)$  of the particle itself. We assume the following approximate function for the velocity

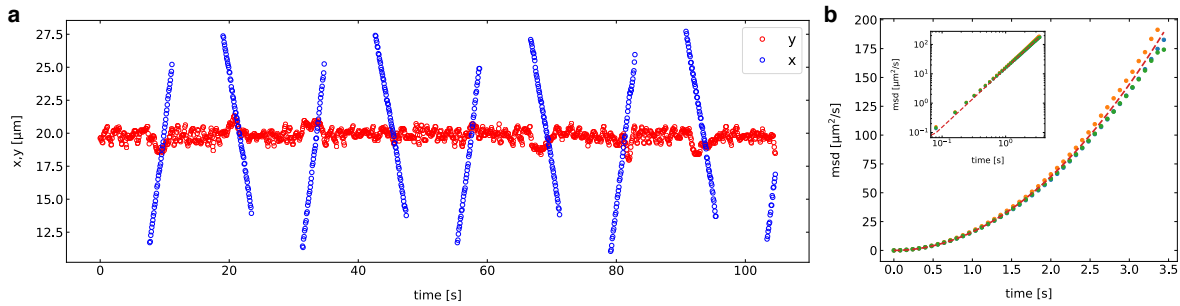
$$v_{\text{th}}(\delta x) = -v_0(P) \frac{\delta x}{[\mu\text{m}]} \cdot e^{-\frac{\delta x^2}{2\sigma^2}} \quad (6)$$

with a velocity amplitude  $v_0 \propto P_{\text{heat}}$ , which depends linearly on the incident heating power  $P_{\text{heat}}$ , a width  $\sigma$  and  $[\mu\text{m}]$  denoting that the unit of  $\delta x$  is removed. The total displacement within a time period  $\Delta t_{\text{exp}}$  is then obtained from integrating  $d\delta x = v_{\text{th}}(\delta x) dt$  resulting in

$$\int_{t=0}^{\Delta t_{\text{exp}}} dt = \Delta t_{\text{exp}} = - \int_{\delta x}^{\delta x + \Delta x} \frac{[\mu\text{m}]}{v_0 \delta x'} e^{\frac{\delta x'^2}{2\sigma^2}} d\delta x' = - \frac{[\mu\text{m}]}{2v_0} \left\{ \text{Ei} \left( \frac{(\delta x + \Delta x)^2}{2\sigma^2} \right) - \text{Ei} \left( \frac{\delta x^2}{2\sigma^2} \right) \right\} = \frac{\Delta x'}{\langle v_{\text{th}} \rangle}. \quad (7)$$

Supplementary equation 7 has to be solved to determine the total displacement  $\Delta x'$  of the active particle during the time period  $\Delta t_{\text{exp}}$  and thus for the average velocity. A solution is only available by numerical integration. The obtained velocity as a function of the maximum velocity  $v_0$  is displayed in Supplementary Figure 4 and captures the trend observed experimentally. Note that in the case of multiple particles being heated, the exposure time has to be replaced by the heating time  $\Delta t_{\text{heat}} = \Delta t_{\text{exp}}/N$  ( $N$  is the total number of particles interacting), which is due to the multiplexing of the heating beam. A more detailed description of the properties of the symmetric active particles will be published elsewhere.

**Active particle motion** The experimental results shown in Figure 1 b–d of the main text have been obtained from a driving of the particle back and forth between two target positions (see Supplementary Fig. 5a) and a localization for 100 frames at each of the target positions.



**Supplementary Figure 5** | (a) Positions  $(x, y)$  of a single active particle driven between two target locations along the  $x$ -direction as presented in Figure 1b of the main text. The incident laser power is  $P = 1$  mW. (b) Mean squared displacement (msd) calculated from the driving periods of the experiments shown in (a). The graph displays the msd (inset: double log scale) for three different driving periods as a function of time for a single particle at an incident laser power of  $P = 1$  mW as well as the predicted parabolic time dependence according to the measured velocity of  $v_{\text{th}} = 3.9 \mu\text{m s}^{-1}$  (dashed line).

As the rotational diffusion of the particle is not influencing the motion, the mean squared displacement during these driving periods should be purely parabolic. Supplementary Figure 5 shows the mean squared displacement during three different driving periods taken from the experiments shown in Figure 1 b–d of the main text. The particle is propelled with an incident laser power of  $P = 1$  mW. The resulting velocity is  $v_{\text{th}} = 3.9 \mu\text{m s}^{-1}$ . The experimental mean squared displacement is purely parabolic over the time period of 3.5 seconds and agrees well with the expected one for a driven motion with constant velocity (dashed line).

## Supplementary Note 2: Simulated Structures

Brownian simulations of the active particle molecules have been carried out to obtain comparative snapshots of the formed structures (Figure 3 main text) not aiming at a quantitative comparison. Simulations use the Processing environment (<http://www.processing.org>). For each particle we define a velocity and a noise amplitude to take care of the propulsion and the Brownian motion of the particles. The speed of the active particle has been set to 6 pixel/frame. Gaussian distributed noise with a variance of 1 pixel has been added to the particle positions at each frame. The propulsion direction is set according to the rules defined in the main text with a defined value of  $r_{\text{eq}}$  corresponding to 80 pixels. The feedback delay is set to one frame, meaning that the positions in the previous frame are used to calculate the directions of propulsion. Similar to the experiments, particle self-organize into a dynamic structure from which snapshots are taken and depicted in Figure 3 of the main text.

## Supplementary Note 3: Analytical Description of the Particle Dynamics

**Dimer bond length oscillation** To evaluate the active particle molecule dynamics we consider the special case of a dimer. For two active particles at the positions  $\mathbf{r}_1(t)$  and  $\mathbf{r}_2(t)$  we may write down two Langevin equations in the overdamped limit to study their dynamics,

$$\dot{\mathbf{r}}_1(t) = \mathbf{v}_1(t) + \sqrt{2D_0}\eta_1, \quad (8)$$

$$\dot{\mathbf{r}}_2(t) = -\mathbf{v}_1(t) + \sqrt{2D_0}\eta_2. \quad (9)$$

The two equations can be combined to give the dynamics of the bond vector  $\mathbf{r}_{12}(t) = \mathbf{r}_1(t) - \mathbf{r}_2(t)$ :

$$\dot{\mathbf{r}}_{12}(t) = 2\mathbf{v}_1(t) + \sqrt{4D_0}\eta_{12} = -2v_{\text{th}}\text{sign}(|\mathbf{r}_{12}(t - \delta t)| - r_{\text{eq}})\mathbf{e}_{12}(t - \delta t) + \sqrt{4D_0}\eta_{12}, \quad (10)$$

where  $\eta_{12}$  is a zero-mean, unit-variance Gaussian white noise (vector), i.e., ( $\langle \eta_{12}(t) \rangle = 0$  and  $\langle \eta(t)\eta(t') \rangle = \delta(t - t')$ ) such that the variance of the noise term in supplementary equation 10 corresponds to  $4D_0$ . The velocity  $|\mathbf{v}_1| = v_{\text{th}}$  is the propulsion velocity and  $\delta t$  the feedback delay time. Note that for  $\delta t \neq 0$  it is not guaranteed that the vectors  $\mathbf{r}_{12}(t) = r_{12}\mathbf{e}_{12}(t)$  and  $\mathbf{e}_{12}(t - \delta t)$  are parallel. For vanishing thermal noise ( $D_0 = 0$ ), however, the situation simplifies because then the vectors  $\mathbf{r}_{12}(t) = r_{12}\mathbf{e}_{12}(t)$  and  $\mathbf{e}_{12}(t - \delta t)$  are always parallel and the motion of the dimer becomes effectively one-dimensional. Setting  $D_0 = 0$  in supplementary equation 10 and taking the scalar product of the result and  $\mathbf{e}_{12}(t) = \mathbf{e}_{12}(t - \delta t)$  we obtain the formula

$$\dot{r}_{12}(t) = -2v_{\text{th}}\text{sign}(|r_{12}(t - \delta t)| - r_{\text{eq}}). \quad (11)$$

This equation can be solved with the result

$$r_{12}(t) = r_{\text{eq}} + Ax_{\text{tr}}(t + \phi_0), \quad (12)$$

where  $A = 2v_{\text{th}}\delta t$  is the amplitude of oscillations and

$$x_{\text{tr}}(t) = \frac{8}{\pi^2} \sum_{k=0}^{\infty} (-1)^k \frac{\sin(2\pi(2k+1)ft)}{(2k+1)^2} \quad (13)$$

denotes a triangular wave with amplitude 1 and period  $T = 1/f = 4\delta t$ . The phase shift  $\phi_0$  is determined by the initial condition which is assumed to be drawn from the interval  $[r_{\text{eq}} - A, r_{\text{eq}} + A]$  attained by the solution 12 in the stationary state.

Scalar multiplication of supplementary equation 10 by the vector  $\mathbf{e}_{12}(t)$  yields

$$\dot{r}_{12}(t) = -2v_{\text{th}}\text{sign}(|r_{12}(t - \delta t)| - r_{\text{eq}})\cos\alpha + \sqrt{4D_0}\eta_{12}, \quad (14)$$

where  $\cos\alpha = \cos\alpha(t, t - \delta t) = \mathbf{e}_{12}(t) \cdot \mathbf{e}_{12}(t - \delta t)$ . For experimentally relevant parameters (the distance diffused in the direction perpendicular to  $\mathbf{e}_{12}(t - \delta t)$  per  $\delta t$  is small compared to the minimal inter-particle distance  $r_{\text{eq}} - A$ ), it is reasonable to assume that  $\cos\alpha \approx 1$ . Then we arrive at the formula

$$\dot{r}_{12}(t) = -2v_{\text{th}}\text{sign}(|r_{12}(t - \delta t)| - r_{\text{eq}}) + \sqrt{4D_0}\eta_{12}. \quad (15)$$

Numerical analysis of this equation reveals that the nonzero noise leads to damped averaged oscillations of the bond. Although the exact analytical calculation of the damping constant seems to be beyond our reach due to the non-analytical nature of the force, an approximate solution can be found.

**Time correlation function for the bond length** The normalized time correlation function  $C(t)$  of the dimer bond length fluctuations  $\Delta r_{12}(t) = r_{12}(t) - r_{\text{eq}}$  is defined as

$$C(t) = \frac{\int_0^{n/f} d\tau \Delta r_{12}(\tau) \Delta r_{12}(\tau + t)}{\int_0^{n/f} d\tau \Delta r_{12}^2(\tau)}. \quad (16)$$

Here, the time integration runs over  $n$  periods of the oscillations with frequency  $f$  measured in the experiment. To obtain an approximate time dependence of the correlation function including the effect of dephasing due to the Brownian motion of the particles we investigate the influence of the noise on the first term of the series 13. In this approximation the bond length is given by

$$\Delta r_{12}(t) = 8A/\pi^2 \sin(2\pi ft + \phi(t)), \quad (17)$$

where the effect of the noise is summarized in  $\phi(t)$  and solely caused by thermal fluctuations of the dimer (amplitude  $A$  fluctuations are neglected). We assume that  $\phi(t=0) = 0$  and thus the wave without noise evolves according to  $\Delta r_{12}^{D_0=0}(t) = 8A/\pi^2 \sin(2\pi ft)$ . The distance  $\Delta x(t)$  between the noise free and the noisy solution within a time  $t$  is then given by

$$\Delta x(t) = \Delta r_{12}^{D_0=0}(t) - \Delta r_{12}(t) = \frac{8A}{\pi^2} [\sin(2\pi ft) - \sin(2\pi ft + \phi(t))]. \quad (18)$$

Assuming that  $\phi(t)$  is small, one can expand the right-hand side with the result

$$\Delta x(t) = \Delta r_{12}^{D_0=0}(t) - \Delta r_{12}(t) \approx \frac{8A}{\pi^2} \left[ -\cos(2\pi ft)\phi(t) + \sin(2\pi ft)\frac{\phi(t)^2}{2} + \dots \right]. \quad (19)$$

Using the first order term we obtain

$$\phi(t) \approx \frac{\pi^2 \Delta x(t)}{8A} \quad (20)$$

assuming that the maximum phase shift is given when  $|\cos(2\pi ft)| = 1$ . If the displacements  $\Delta x(t)$  obey a Gaussian distribution

$$p(\Delta x) = \frac{1}{\sqrt{4\pi D \Delta t}} e^{-\frac{\Delta x^2}{4D \Delta t}} \quad (21)$$

with  $D = 2D_0$  due to the relative motion, where  $D_0$  is the diffusion coefficient of a free individual active particle, one can calculate the approximate damping constant as follows.

Inserting supplementary equation 17 into the time-correlation function 16 and assuming additivity of the phase shift  $\phi(t + \tau) = \phi(t) + \phi(\tau)$  gives  $C(t) = \cos(2\pi ft + \phi(t))$ . Averaging this expression over the ensemble of Brownian displacements 21 yields

$$C(t) = \int_{-\infty}^{\infty} p(\phi) \cos(2\pi ft + \phi) d\phi = \frac{\pi^2}{8A} \int_{-\infty}^{\infty} p(\Delta x) \cos\left(2\pi ft + \frac{\pi^2 \Delta x}{8A}\right) d\Delta x. \quad (22)$$

The oscillations with frequency  $f$  and phase shift  $\phi$  sum up to form an exponentially decaying oscillation

$$C(t) = e^{-\pi^4 D t / 64 A^2} \cos(2\pi f t). \quad (23)$$

The time constant of the exponential decay is therefore  $T_2 = 64 A^2 / \pi^4 2 D_0$ . It is named  $T_2$  due to its analogy with the dephasing time in optical spectroscopy or magnetic resonance. Inserting the amplitude  $A$  yields

$$T_2 = \frac{256 v_{\text{th}}^2 \delta t^2}{2 \pi^4 D_0}. \quad (24)$$

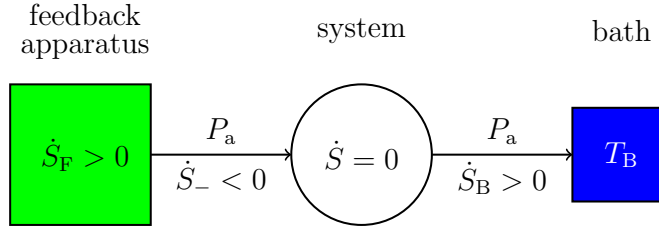
The dephasing time is therefore decreasing with increasing noise, i.e., increasing diffusion coefficient  $D_0$ .

## Supplementary Note 4: Entropy Fluxes

Without the feedback driving, the colloidal particles would start from an initial configuration gradually spreading out due to Brownian motion. During such a process, the entropy  $S_B$  of the water (bath) remains constant and the entropy of the bond lengths between the individual colloids (system)  $S$  increases. The diffusion of the center of mass of the colloidal molecule is insensitive to the feedback and the corresponding entropy always increases. To address the structure formation caused by the feedback, we thus consider the distribution of the bond lengths only. The total entropy production rate during the process reads  $\dot{S}_{\text{tot}} = \dot{S} > 0$ . A similar time evolution is observed if one switches on the laser and targets random locations at the circumference of the particles, without using the information about the relative positions of the particles. In that case, a part of the laser power  $P$  is absorbed by the particles and gradually dissipated to the bath, leading to  $\dot{S}_B = P_a / T > 0$ , where  $T$  is the bath temperature. The total entropy production rate during the process thus reads  $\dot{S}_{\text{tot}} = \dot{S} + \dot{S}_B > 0$ . In addition to that, the particles now spread faster than without the heating. Utilizing the information about particle positions according to the rules described in the main text in placing the laser implies a qualitatively different evolution of the system. The particles form localized structures which fluctuate due to the Brownian motion and oscillate due to the feedback. After long times, the system attains a non-equilibrium steady-state with a time-independent probability distribution of interparticle distances. The laser heating still produces nonzero entropy flux to the bath, but the system entropy is now time-independent. The total energy input into the system both in the situation without aiming the laser and with the aiming is the same, given by  $P_a$ . When comparing the random driving and the structure formed, the structure has a lower entropy than the random particle distribution which is due to the utilization of the information. It is the processing of this information which requires the additional energy input. More precisely, the information processing is necessarily accompanied with a positive entropy flux  $\dot{S}_F$ , which is bounded from below by the Landauer's principle. The total entropy production rate in the steady-state thus reads  $\dot{S}_{\text{tot}} = \dot{S}_B + \dot{S}_F > 0$ .

The feedback loop uses the information about the particle positions to balance the particle spreading due to the diffusion. Differently speaking, the feedback introduces a negative entropy influx  $\dot{S}_-$  into the system. The described entropy fluxes through the system in the non-equilibrium steady-state are depicted in Supplementary Figure 6.





**Supplementary Figure 6** | Thermodynamic diagram of the system.

In order to evaluate the negative entropy influx  $\dot{S}_-$ , it is possible to follow the analysis of reference<sup>[3]</sup> devoted to entropy production in stochastic systems with resetting. Let us for simplicity consider just the dimer case, which can be straightforwardly generalized to more particles. We denote as  $\rho(x, t)$  the probability distribution for the bond length  $x = r_{12}$  at time  $t$ . The rate of change of the Shannon entropy of this distribution reads

$$\dot{S} = -k_B \int_{-\infty}^{\infty} dx \ln \rho(x, t) \frac{\partial \rho(x, t)}{\partial t}. \quad (25)$$

Intuitively, this function is large after switching on the feedback in an unstructured system and decreases towards zero while the system relaxes to a time-independent non-equilibrium steady state, where  $\partial \rho(x)/\partial t = 0$ ,  $\rho(x) = \lim_{t \rightarrow \infty} \rho(x, t)$ .

The dynamical equation for  $\rho(x, t)$  can be written as

$$\frac{\partial \rho(x)}{\partial t} = 2D_0 \frac{\partial^2}{\partial x^2} \rho(x) + L[\rho(x, t)], \quad (26)$$

where the first term stands for the change of the density due to the diffusion and the second one due to the driving. Inserting supplementary equation 26 into equation 25 we obtain for the entropy change in the steady-state

$$\dot{S} = -2D_0 k_B \int_{-\infty}^{\infty} dx \frac{\partial^2}{\partial x^2} \rho(x) \ln \rho(x) + S_- = 0. \quad (27)$$

Regardless the specific form of the operator  $L[\rho]$  in supplementary equation 26, the negative entropy influx due to the feedback is determined by the stationary distribution  $\rho(x)$ :

$$\dot{S}_- = 2D_0 k_B \int_{-\infty}^{\infty} dx \ln \rho(x) \frac{\partial^2}{\partial x^2} \rho(x) = -2D_0 k_B \int_{-\infty}^{\infty} dx \frac{1}{\rho(x)} \left[ \frac{\partial}{\partial x} \rho(x) \right]^2 < 0. \quad (28)$$

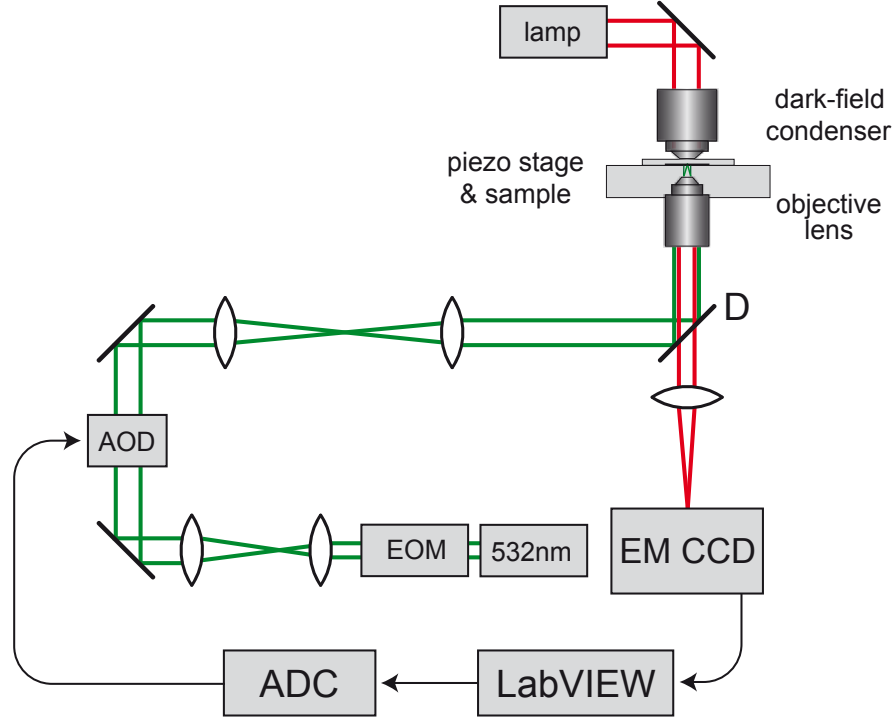
Let us now consider the case of vanishing time delay  $\delta t$ . Then the feedback creates the effective V-type potential  $U(x) = \gamma v_{\text{th}} ||x| - r_{\text{eq}}|$  for the bond length, where  $\gamma = 6\pi\eta R$  is the Stokes friction coefficient. The stationary distribution of the bond length thus reads  $\rho(x) = \exp(-\gamma v_{\text{th}} ||x| - r_{\text{eq}}| / k_B T) / Z$  and the formula (28) gives us

$$\dot{S}_- = -2D_0 k_B \beta^2 \gamma^2 v_{\text{th}}^2 = -2k_B \beta \gamma v_{\text{th}}^2 = -2 \frac{\gamma v_{\text{th}}^2}{T} = -2\eta \frac{P_a}{T}. \quad (29)$$

Here  $\eta P_a$  denotes the fraction of the absorbed light power (efficiency  $\eta$ , absorbed power  $P_a$ ), which is used for the particle propulsion,  $\beta = (k_B T)^{-1}$  and  $Z$  as the equilibrium partition

function. Quite intuitively, the entropy influx is stronger (more negative) for larger velocities. Its absolute increase with decreasing temperature is not so intuitive. Considering a more general potential of the form  $kx^n$  we find that  $\dot{S}_- \propto (k/T)^{2/n}T$ . Thus the increase of  $|\dot{S}_-|$  with decreasing  $T$  is slower for stronger potentials/more localized distributions and for  $n \geq 2$  the absolute entropy influx even increases with  $T$ .

## Supplementary Note 5: Experimental Setup



**Supplementary Figure 7** | Experimental setup for the feedback controlled active particles. The following abbreviations are used: AOD – acousto-optic-deflector, EOM – electro-optical modulator, emCCD – Electron Multiplying CCD, FPGA – Field Programmable Gate Array, D – Dichroic mirror. All the other components are lenses, mirrors and a standard microscopy lamp.

Samples have been investigated in a custom built inverted microscopy setup. The setup is based on an Olympus IX 71 microscopy stand (see Supplementary Figure 7). Optical heating of the active particles is carried out by a CW 532 nm wavelength laser. The laser intensity is controlled by a Conoptics 350-50 electro-optical modulator (EOM). An acousto-optic deflector (AOD) together with a 4-f system (two  $f = 20$  cm lenses) is used to steer the 532 nm wavelength laser focus in the sample plane. The AOD is controlled by a Field Programmable Gate Array (FPGA, National Instruments) via a LabView program. The calibration of the AOD for precise laser positioning is carried out using a 2D projection method developed in the lab A Leica 100x, infinity-corrected, NA 1.4 - 0.7 (set to 0.7), HCX

PL APO objective lens is used for focusing the 532 nm laser to the sample plane as well as for imaging the active particles. Active particles are imaged under dark field illumination. When the sample is placed under the microscope, an oil immersion dark field condenser (Olympus 1.2 NA) is approached from the top. The scattered light from the sample is collected with the Leica objective lens and imaged with a  $f = 30$  cm tube lens to an emCCD camera (Cascade 650). A region of interest (ROI) of 200 pixels x 200 pixels is utilized for the real-time imaging, analysis and recording of the particles, with an exposure time of 80 ms or 110 ms.

## Supplementary References

- [1] Bregulla, A. P., Würger, A., Günther, K., Mertig, M. & Cichos, F. Thermo-osmotic flow in thin films. *Phys. Rev. Lett.* **116**, 188303 (2016).
- [2] Bickel, T., Majee, A. & Würger, A. Flow pattern in the vicinity of self-propelling hot Janus particles. *Phys. Rev. E* **88**, 012301 (2013).
- [3] Fuchs, J., Goldt, S. & Seifert, U. Stochastic thermodynamics of resetting. *Eur. Phys. Lett.* **113**, 60009 (2016).



PAPER • OPEN ACCESS

## Brownian molecules formed by delayed harmonic interactions

To cite this article: Daniel Geiss *et al* 2019 *New J. Phys.* **21** 093014

View the [article online](#) for updates and enhancements.



**IOP | ebooks**<sup>TM</sup>

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.



## PAPER

## Brownian molecules formed by delayed harmonic interactions

## OPEN ACCESS

RECEIVED  
15 May 2019REVISED  
26 July 2019ACCEPTED FOR PUBLICATION  
21 August 2019PUBLISHED  
10 September 2019

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/4.0/).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Daniel Geiss<sup>1,2</sup>, Klaus Kroy<sup>2</sup> and Viktor Holubec<sup>2,3,4</sup><sup>1</sup> Max Planck Institute for Mathematics in the Sciences, Inselstr. 22, D-04103, Leipzig, Germany<sup>2</sup> Institut für Theoretische Physik, Universität Leipzig, Postfach 100 920, D-04009, Leipzig, Germany<sup>3</sup> Charles University, Faculty of Mathematics and Physics, Department of Macromolecular Physics, V Holešovičkách 2, CZ-180 00 Praha, Czech Republic<sup>4</sup> Author to whom any correspondence should be addressed.E-mail: [daniel.geiss@mis.mpg.de](mailto:daniel.geiss@mis.mpg.de), [klaus.kroy@uni-leipzig.de](mailto:klaus.kroy@uni-leipzig.de) and [viktor.holubec@mff.cuni.cz](mailto:viktor.holubec@mff.cuni.cz)**Keywords:** stochastic delay differential equation, active matter, feedback driving, non-Markovian dynamics, transition-state theory, Fokker–Planck equation, structure formation**Abstract**

A time-delayed response of individual living organisms to information exchanged within flocks or swarms leads to the emergence of complex collective behaviors. A recent experimental setup by (Khadka *et al* 2018 *Nat. Commun.* **9** 3864), employing synthetic microswimmers, allows to emulate and study such behavior in a controlled way, in the lab. Motivated by these experiments, we study a system of  $N$  Brownian particles interacting via a retarded harmonic interaction. For  $N \leq 3$ , we characterize its collective behavior analytically, by solving the pertinent stochastic delay-differential equations, and for  $N > 3$  by Brownian dynamics simulations. The particles form molecule-like non-equilibrium structures which become unstable with increasing number of particles, delay time, and interaction strength. We evaluate the entropy and information fluxes maintaining these structures and, to quantitatively characterize their stability, develop an approximate time-dependent transition-state theory to characterize transitions between different isomers of the molecules. For completeness, we include a comprehensive discussion of the analytical solution procedure for systems of linear stochastic delay differential equations in finite dimension, and new results for covariance and time-correlation matrices.

**1. Introduction****1.1. Feedback systems**

From the synchronized response of a flock of starlings [1] avoiding an attack of a predator to the formation of colonies of living bacteria [2, 3], the surging field of active matter provides a wide range of fascinating phenomena. Its ultimate aim is to develop a microscopic understanding of the behavior of large numbers of interacting, active and energy consuming ‘agents’ [4, 5], with a focus on emergent collective behavior [6]. Most of the quantitative models, such as the Vicsek model [7], neglect the finite speed of information transmission between the individual particles. However, recent studies [8–11] have shown that a time delay in the interaction may significantly affect the system dynamics. Moreover, experimental realizations mimicking natural interacting systems require implementing the non-physical interactions, such as a reaction of a bird to its environment, via a feedback loop [10–13]. Finite processing of the information in the feedback loop then inevitably introduces time delay into the system dynamics.

Current (mainly optical) micromanipulation techniques allow to realize such feedback systems on microscale [11, 13–19]. Many [13, 18, 19] of these techniques are based on spherical Janus particles [20, 21] with hemispheres coated with different materials in order to excite surface flows to propel them actively upon illumination or in presence of other energy sources (e.g. chemical fuel added to the solvent). In order to steer these particles, one usually has to wait until the rotational diffusion reorients them towards the desired location. This issue was resolved by the setup introduced by Khadka *et al* [11] based on Brownian particles, symmetrically decorated by gold nanoparticles, that thermophoretically self-propel in the direction determined by the position of the laser focus on their circumference. In the feedback experiment, the particles are tracked with a camera with finite exposure time and the position of the heating laser is determined by positions of the particles in the

previous frame. The setup allows to create arbitrary time-delayed interactions in the many-body system. In [11], an interaction leading to constant absolute values of velocities of the individual particles was considered.

In the present paper, we theoretically analyze a system similar to that considered by Khadka *et al* [11], but with harmonic interactions between the individual particles. Similarly to the case of [11], the two-dimensional  $N$ -particle system is described by a set of  $2N$  coupled nonlinear stochastic delay differential equations (SDDE). For small enough values of the delay, highly symmetric non-equilibrium molecular-like structures form after a transient period, which fluctuate due to thermal noise. The resulting structures strongly differ from the molecules created with the constant-velocity protocol studied in [11], which oscillated, even for vanishing noise amplitude, due to the nonzero delay. Another difference between the two realizations is that our setup leads, for large delays, to oscillations with amplitudes exponentially increasing in time, while, in the setup of Khadka *et al*, they are always bounded. The specific form of the interaction considered in our setup moreover allows us to linearize the underlying set of SDDE and to study many aspects of the model behavior analytically. For dimer ( $N = 2$ ) and trimer ( $N = 3$ ), we use the linearized model to calculate properties of the resulting Gaussian probability distributions for the bond length, namely its mean values, covariance matrix, and time-correlation matrix. We verify the validity of these results by Brownian dynamics (BD) simulations of the complete model. Moreover, we use the BD simulations to show that the behavior of larger molecules ( $N > 3$ ) is qualitatively the same as that of the dimer and the trimer, with the difference that the critical value of the delay, beyond which the molecules become unstable, decreases inversely in the particle number  $N$ . If we would scale the interaction strength by the particle number, as it is common in toy models of many-body systems, the critical value of the delay would thus be constant. If we label the individual particles, we can distinguish between several isomers of the respective molecules according to their ordering. In the course of time, the noise induces jumps of a given molecule between the individual isomers. We utilize our analytical results for the dimer and for the trimer to evaluate the corresponding transition rates using Kramers' theory [22, 23] and a more recent theory by Bullerjahn *et al* [24]. We compare the results with the rates calculated from our BD simulations and identify a useful formula for the transition rate that provides good predictions for small and moderate values of the delay.

## 1.2. Stochastic delay differential equations

In general, delay differential equations (DDE's) [25, 26] may generate rich dynamics [27]. Their solutions may converge to fixed points or limit cycles, behave chaotically, and exhibit multistability [28]. For systems affected by noise, the DDEs are generalized to SDDE [29], which exhibit non-Markovian dynamics. Due to delay-induced temporal correlations, the corresponding Fokker–Planck equation (FPE) cannot be written in a closed form [30–32]. Instead, one obtains an infinite hierarchy of coupled FPEs for the  $n$ -time joint probability densities for which generally no finite closure is known.

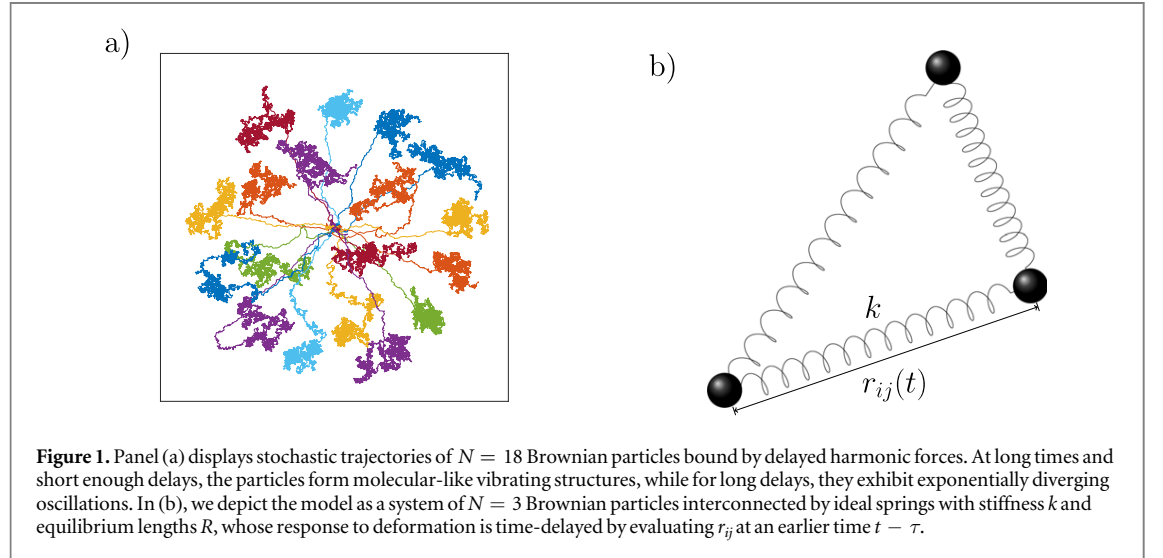
For nonlinear systems, there are three established approximate approaches how to tackle the infinite hierarchy: (i) the so called small delay approximation [30], which employs a Taylor expansion in the delay to make the equations time local; (ii) also, if the delayed terms in the SDDE are small so that the system dynamics is almost Markovian, a perturbation theory can be applied, leading to closed FPEs for the individual joint probability densities [33]; (iii) a closed equation for the 1-time probability density valid in the steady state can be obtained by linearization of all equations of the FPE hierarchy except for the first one [32].

So far, the only exactly solved problem is a one-dimensional linear stochastic delay equation with Gaussian white noise, whose  $n$ -time probability densities are given by multivariate Gaussians. Its stationary solution and the conditions for its existence were discussed in [30, 31, 34]. Recently, a full time-dependent solution for 1- and 2-time probability densities was found by Giuggioli *et al* [35]. Employing the so-called time-convolutionless transform introduced in the 1970s [36–40], these authors transformed the non-Markovian linear delayed Langevin equation (LE) into a time-local form. Afterwards, they utilized this result in a derivation of analytically solvable time-local FPEs for 1- and 2-time probability densities.

Even though an analytical treatment is thus rather complicated, there is a great interest in understanding DDE's and SDDE's, due to their broad range of applications. Prominent examples are found in population dynamics [41, 42], where the delay results from maturation times, economics [43–46] when the limited reaction times of the market participants matters, or engineering [47]. In biology, finite transition times can play a significant role in physiological systems [48–50] and neural networks [28, 51–53]. Recently [54–56], first efforts were also made to incorporate a time delay into the language of stochastic thermodynamics [57, 58] in order to evaluate energy and entropy fluxes in time-delayed stochastic system.

## 1.3. Outline

The rest of the paper is structured as follows. In section 2 we first introduce the general model and formulate the underlying equations of motion in terms of SDDEs. After appropriate linearization, we study them analytically,



considering both transient and stationary properties of the probability distributions for ‘bond’ lengths in ‘molecules’ self-assembling from two and three particles. Larger systems are studied in section 3. In section 4, we apply the obtained results for evaluation of the entropy outflux (or information influx) from the system due to the feedback maintaining the non-equilibrium stationary structures. In order to obtain a more quantitative characterization of the stability of the non-equilibrium molecules, we utilize our analytical description of the dimer and trimer for analytical and numerical investigation of the isomer transitions and back up the results by BD simulations in section 5. In order to assess the robustness of our findings, section 6 addresses the role of the functional form of the memory kernel considering negative delays and exponential memories. We summarize our findings and conclude in section 7. Most of the technical details are given in appendices A–C. In appendix A, we review the known results concerning the solution of systems of LDDEs. In appendix B, we show how to generalize these results for linear SDDEs. Finally, we apply the obtained results in appendix C for the calculation of the time-correlation matrix and the covariance matrix for systems of linear SDDEs.

## 2. Stochastic dynamics

We consider a two-dimensional system of  $N$  overdamped Brownian particles coupled via time-delayed harmonic pair interactions given by the potential

$$V_2[r_{ij}(t - \tau)] = \frac{k}{2}[r_{ij}(t - \tau) - R]^2, \quad (1)$$

depicted in figure 1 by springs connecting the individual particles. In equation (1),  $R > 0$  denotes the equilibrium spring length,  $k$  their stiffness, and  $r_{ij}(t - \tau) = |\mathbf{r}_i(t - \tau) - \mathbf{r}_j(t - \tau)|$  is the distance between the particles  $i$  and  $j$  located at positions  $\mathbf{r}_i$  and  $\mathbf{r}_j$  at an earlier time  $t - \tau$ . Clearly, the picture of linear springs can properly represent the time-delayed interactions only for a vanishing time delay  $\tau$ .

Altogether, the particles diffuse in the composit potential

$$V = \frac{1}{2} \sum_{(i,j)} V_2(r_{ij}), \quad (2)$$

where the summation runs over all pairs  $(i, j)$ , so that the  $i$ th particle is driven by the force  $\mathbf{F}_i = -\nabla_i V = (\partial_{x_i} V, \partial_{y_i} V)$ . Because at time  $t$  the particle feels the value of the potential corresponding to its position at time  $t - \tau$ , here  $x_i$  and  $y_i$  denote the Cartesian coordinates of the position vector  $\mathbf{r}_i$  in the past. In effect, the  $N$ -particle system therefore obeys the set of *nonlinear* delayed Langevin equations

$$\dot{\mathbf{r}}_i(t) = -\frac{k}{\gamma} \sum_{j \neq i} [r_{ij}(t - \tau) - R] \mathbf{e}_{ij}(t - \tau) + \sqrt{2D_0} \boldsymbol{\eta}_i(t), \quad i = 1, \dots, N. \quad (3)$$

The unit vector  $\mathbf{e}_{ij} = \mathbf{r}_{ij}/r_{ij}$  points from particle  $j$  to particle  $i$  and the diffusion coefficient  $D_0 = (\beta\gamma)^{-1}$  is related to the inverse temperature  $\beta = 1/k_B T$  and the friction coefficient  $\gamma$  via the Einstein relation ( $k_B$  denotes the Boltzmann constant). The vectors  $\boldsymbol{\eta}_i$  comprise independent Gaussian white noises satisfying the relations



$$\langle \boldsymbol{\eta}_i(t) \rangle = 0, \quad \langle \eta_i^\alpha(t_1) \eta_j^\beta(t_2) \rangle = \delta_{ij} \delta_{\alpha\beta} \delta_D(t_1 - t_2). \quad (4)$$

The numbers  $\alpha$  and  $\beta$  label the components of the vector  $\boldsymbol{\eta}_i$ , while  $i$  refers to the specific particle. Note that the noise and the friction in equation (3) are related by the fluctuation-dissipation theorem [59] for a vanishing delay  $\tau = 0$  only, and that the system is always out of thermodynamic equilibrium for  $\tau > 0$  [56]. In order to obtain a model that would obey the fluctuation-dissipation theorem, one should consider a different noise correlation function (4).

For small enough values of the time delay, the particles form, after an initial transient period, highly symmetric molecular-like structures, some of which are displayed in figure 4(a) in section 3. For  $N = 2$  (dimer) and  $N = 3$  (trimer) the steady-state structures occupy the global minimum of the potential  $V$ . For  $N > 3$  the global minimum becomes inaccessible due the chosen two-dimensional geometry and the resulting structures are thus frustrated in the sense that some of the springs do not reach their equilibrium length in the steady-state. The structures are dynamical, due to the Brownian motion of the particles, which persistently kicks the system out of the minimum of the potential energy (2). The effect of the delay is that the system may exhibit exponentially decaying oscillations on its return to the energy minimum. The decay rate of these oscillations decreases with increasing delay, and, for delays larger than a certain threshold, their amplitude exponentially increases. This is because large delays induce in the system a ‘swing effect’, when the repulsive force from one side of the potential propels the particle to a ‘higher’ position at its opposite side, and so on.

Within the equilibrium model that obeys the fluctuation-dissipation theorem, the stationary probability density function (PDF) for positions of the individual particles would simply be the Boltzmann distribution  $\exp(-\beta V)/Z$  with potential  $V$ , inverse temperature  $\beta$ , and normalization  $Z$ . However, the physical situation at hand, where the delay is interpreted as a result of a feedback control mechanism and thus is independent of the noise, requires the more involved description with equation (4) that leads to non-trivial non-equilibrium steady states. Consequently, the Boltzmann distribution can no longer be assumed. For the simplest case of a dimer with short delay time, we will find an approximately Gaussian distribution, corresponding to a (‘deformed’) Boltzmann factor at an effective temperature. For larger particle numbers and longer delay times, the situation becomes more complicated.

A similar system with a quasi-constant force between the particles (constant upto a change of sign at distance  $R$ ), i.e. obeying the set of Langevion equations

$$\dot{\mathbf{r}}_i(t) = -\frac{k}{\gamma} \frac{\sum_{j \neq i} \text{sign}[r_{ij}(t - \tau) - R] \mathbf{e}_{ij}(t - \tau)}{\left| \sum_{j \neq i} \text{sign}[r_{ij}(t - \tau) - R] \mathbf{e}_{ij}(t - \tau) \right|} + \sqrt{2D_0} \boldsymbol{\eta}_i(t), \quad (5)$$

$i = 1, \dots, N$ , with  $\text{sign}(x)$  denoting the signum function, was discussed earlier in [11]. The main difference from our setting (3) is that, in equation (5), the absolute value of the force does not depend on the interparticle distances and the particle number  $N$ . The main benefit of assuming the harmonic potential in equation (3) is that it allows much more complete analytical treatment. To allow for an easy comparison of the two models, we illustrate our results using parameters inspired by [11]. In the following, we first review some analytical results for stochastic dynamics of dimers and trimers. On this basis, we will return to the discussion of the emerging structures in section 3.

### 2.1. Center of mass

Similarly as for the dynamics considered in [11], the center of mass coordinate  $\mathbf{r}_c \equiv (\sum_{i=1}^N \mathbf{r}_i)/N$  of the system obeys the Langevin equation

$$\dot{\mathbf{r}}_c(t) = \sqrt{2D_c} \boldsymbol{\eta}_c(t), \quad (6)$$

where  $\boldsymbol{\eta}_c \equiv \sum_{i=1}^N \boldsymbol{\eta}_i / \sqrt{2}$  denotes Gaussian white noise satisfying equation (4) (with the labels  $i, j$  replaced by  $c$ ) and the diffusion coefficient  $D_c = D/N$ . Regardless of the interactions, the center of mass performs ordinary Brownian motion and, assuming the center of mass is at time  $t = 0$  located at the point  $\mathbf{r}_0$ , the PDF for  $\mathbf{r}_c$  reads

$$P_N^c(\mathbf{r}, t) = \sqrt{\frac{N}{4\pi D t}} \exp\left[-N \frac{(\mathbf{r} - \mathbf{r}_0)^2}{4D t}\right]. \quad (7)$$

### 2.2. Dimer

Let us now consider the simplest case of two interacting particles. For  $N = 2$ , equation (3) yields the system of four coupled equations of motion:

$$\dot{\mathbf{r}}_1(t) = -\frac{k}{\gamma} [r(t - \tau) - R] \mathbf{e}(t - \tau) + \sqrt{2D} \boldsymbol{\eta}_1(t), \quad (8)$$

$$\dot{\mathbf{r}}_2(t) = +\frac{k}{\gamma} [r(t - \tau) - R] \mathbf{e}(t - \tau) + \sqrt{2D} \boldsymbol{\eta}_2(t), \quad (9)$$

where we have used the abbreviations  $\mathbf{e}(t) \equiv \mathbf{e}_{12}(t)$  and  $r(t) \equiv r_{12}(t)$ . In the previous section, we have already resolved the dynamics of the center of mass coordinate for arbitrary  $N$ . Now, we consider only the dynamics of

the relative coordinate  $\mathbf{r} = \mathbf{r}_{12} = \mathbf{r}_1 - \mathbf{r}_2$  which obeys the equation of motion

$$\dot{\mathbf{r}}(t) = -\omega[r(t - \tau) - R]\mathbf{e}(t - \tau) + \sqrt{4D}\boldsymbol{\eta}_r(t) \quad (10)$$

with frequency

$$\omega \equiv 2k/\gamma \quad (11)$$

and  $\boldsymbol{\eta}_r \equiv (\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2)/\sqrt{2}$  describing Gaussian white noise satisfying equation (4) with the vector components  $i, j$  replaced by the label  $r$ .

Projecting equation (10) on the direction of the bond at time  $t$  (by multiplication with  $\mathbf{e}(t) = (\cos \varphi(t), \sin \varphi(t))$ ) and on the direction perpendicular to the bond (by multiplying it with  $\mathbf{e}_\varphi = (-\sin \varphi(t), \cos \varphi(t))$ ), we obtain the equations

$$\dot{r}(t) = -\omega[r(t - \tau) - R] \cos[\varphi(t, t - \tau)] + \sqrt{4D}\eta_r^r(t), \quad (12)$$

$$\dot{\varphi}(t) = \omega \frac{r(t - \tau) - R}{r(t)} \sin[\varphi(t, t - \tau)] + \sqrt{\frac{4D}{r^2(t)}}\eta_r^\varphi(t), \quad (13)$$

where  $\varphi(t, t - \tau) = \varphi(t) - \varphi(t - \tau)$  denotes the change of orientation of the vector  $\mathbf{e}(t - \tau)$  during time  $\tau$ .

Above, we used the formulas  $\mathbf{r} \equiv r\mathbf{e}$ ,  $\dot{\mathbf{r}} \equiv \dot{r}\mathbf{e} + r\dot{\varphi}\mathbf{e}_\varphi$  and  $\boldsymbol{\eta}_r \equiv \eta_r^r\mathbf{e} + \eta_r^\varphi\mathbf{e}_\varphi$ .

From symmetry considerations, it follows that the stationary PDF for the orientation must be constant in  $\varphi$ . To gain analytical insight into the dynamics and PDF of the bond-length  $r$ , we linearize the coupled Langevin equations (12) and (13). If the angle dependent stiffness  $2k/\gamma \cos[\varphi(t, t - \tau)]$  in equation (12) is strong enough such that the terms proportional to  $[r(t - \tau) - R]/R$  can safely be neglected independently of the noise strength, the formula (13) for the angle assumes the form<sup>5</sup>

$$\dot{\varphi}(t) = \sqrt{\frac{4D}{R^2}}\eta_r^\varphi(t). \quad (14)$$

The corresponding transition PDF (Green's function) for change of the orientation by  $\varphi(t, t - \tau) = \varphi(t) - \varphi(t - \tau)$  during the time interval  $\tau$  reads [60–62]  $p[\varphi(t, t - \tau), \tau] = (2\pi)^{-1} + \pi^{-1} \sum_{m=1}^{\infty} \cos[m\varphi(t, t - \tau)] \exp[-2m^2\tau D/R^2]$ . Using this function, we average equation (12) over  $\varphi(t, t - \tau)$  obtaining

$$\dot{r}(t) = -\omega_\tau[r(t - \tau) - R] + \sqrt{4D}\eta_r^r(t), \quad (15)$$

where  $\omega_\tau = \omega \exp(-2D\tau/R^2)$  is the natural relaxation rate. Note that the same formula with  $\omega_\tau$  substituted by  $\omega$  is obtained by simply assuming that the change of the orientation  $\varphi(t, t - \tau)$  of the bond per delay time  $\tau$  is small, i.e. for  $2\tau D/R^2 \ll 1$ . The main difference between the two approximations is that  $D/R^2$  does not necessarily have to be small in the first case. We will discuss the regime of validity of the equation (15) in more detail around equation (25) below.

Equation (15) is a linear SDDE which can be solved analytically for  $r \in (-\infty, \infty)$ . In our setting,  $r \geq 0$  and thus we should solve equation (15) with a reflecting boundary at the origin. However, since we have assumed that  $|r(t - \tau) - R|/r(t) \ll 1$ , we already work in the regime where  $r$  only seldom significantly deviates from  $R$  and thus the solution of equation (15) on the full real axis should approximate well the desired solution on the positive half-line. The solution of equation (15) for  $r \in (-\infty, \infty)$  and  $t \geq 0$  in terms of deviations of the bond length from its equilibrium length (which we call as *shifted bond length*),

$$x(t) = r(t) - R, \quad (16)$$

can be derived by several methods. We review two of them (time-convolutionless transform and Gaussian ansatz) for a general multidimensional linear SDDE in appendices A–B. Here, we present just the main formulas. Assuming that the system was initially in state  $x(0) = x_0$  and that  $x(t) = 0$  for  $t < 0$ , the formal solution of equation (15) for  $r \in (-\infty, \infty)$  and  $t \geq 0$  reads

$$x(t) = x_0\lambda(t) + \sqrt{4D} \int_0^t ds \lambda(t - s)\eta_r^r(s), \quad (17)$$

where the dimensionless Green's function

$$\lambda(t) = \sum_{k=0}^{\infty} \frac{(-\omega_\tau)^k}{k!} (t - k\tau)^k \theta(t - k\tau) \quad (18)$$

solves equation (15) with  $D = 0$ ,  $\lambda(t) = 0$  for  $t < 0$  and  $\lambda(0) = 1$ . The symbol  $\theta(x)$  in equation (18) stands for the Heaviside step function. For an arbitrary initial condition  $x(t), t < 0$  and  $x(0) = x_0$ , the expression  $x_0\lambda(t)$  in equation (17) must be substituted by  $x_0\lambda(t) - \omega_\tau \int_{-\tau}^0 ds \lambda(t - \tau - s)x(s)$ . Based on the value of the *reduced*

<sup>5</sup> In equation (13), we set  $r = [(r - R)/R + 1]R$ , expand it in  $(r - R)/R$ , and neglect all terms proportional to  $(r - R)/R$ .

delay  $\omega_\tau\tau$  which is a dimensionless measure for the relevance of the delay relative to the natural relaxation time, the Green's function  $\lambda(t)$  in equation (17) exhibits three different dynamical regimes discussed in detail in appendix A, in figure A1 and also below: (i) monotonic exponential decay to zero for short delays  $0 \leq \omega_\tau\tau \leq 1/e$ , (ii) oscillatory exponential decay to zero for intermediate ('resonant') delays  $1/e \leq \omega_\tau\tau \leq \pi/2$ , and (iii) oscillatory exponential divergence for long delays  $\omega_\tau\tau > \pi/2$ .

The stochastic process  $x(t)$  in equation (17) arises as a linear combination of white noises and thus the corresponding PDFs must be Gaussian. Indeed, we find that one- and two-time conditional PDFs for  $x(t)$  with the initial condition  $\delta(x)$  for  $t < 0$  and  $\delta(x - x_0)$  at  $t = 0$  read

$$P_1(x, t|x_0, 0) = \frac{1}{\sqrt{2\pi\nu(t)}} \exp\left\{-\frac{1}{2}\left(\frac{x - \mu(t)}{\sqrt{\nu(t)}}\right)^2\right\}, \quad (19)$$

$$P_2(x, t; x', t'|x_0, 0) = \frac{1}{\sqrt{4\pi^2\nu(t)\nu(t')[1 - w^2(t', t)]}} \times \exp\left\{\frac{1}{2[1 - w^2(t', t)]}\left[\left(\frac{x - \mu(t)}{\sqrt{\nu(t)}}\right)^2 + \left(\frac{x' - \mu(t')}{\sqrt{\nu(t')}}\right)^2 - 2w(t, t')\left(\frac{x - \mu(t)}{\sqrt{\nu(t)}}\right)\left(\frac{x' - \mu(t')}{\sqrt{\nu(t')}}\right)\right]\right\}, \quad (20)$$

where  $t \geq t' \geq 0$  and

$$\mu(t) \equiv \langle x(t) \rangle = x_0\lambda(t), \quad (21)$$

$$\nu(t) \equiv \langle x^2(t) \rangle - \mu^2(t) = 4D \int_0^t ds \lambda^2(s), \quad (22)$$

$$w(t, t') \equiv \frac{\langle x(t)x(t') \rangle - \mu(t)\mu(t')}{\sqrt{\nu(t)\nu(t')}} = \frac{4D}{\sqrt{\nu(t)\nu(t')}} \int_0^{t'} ds \lambda(t-s)\lambda(t'-s) \quad (23)$$

denote the mean of the shifted bond-length (16), its variance, and normalized time correlation, respectively.

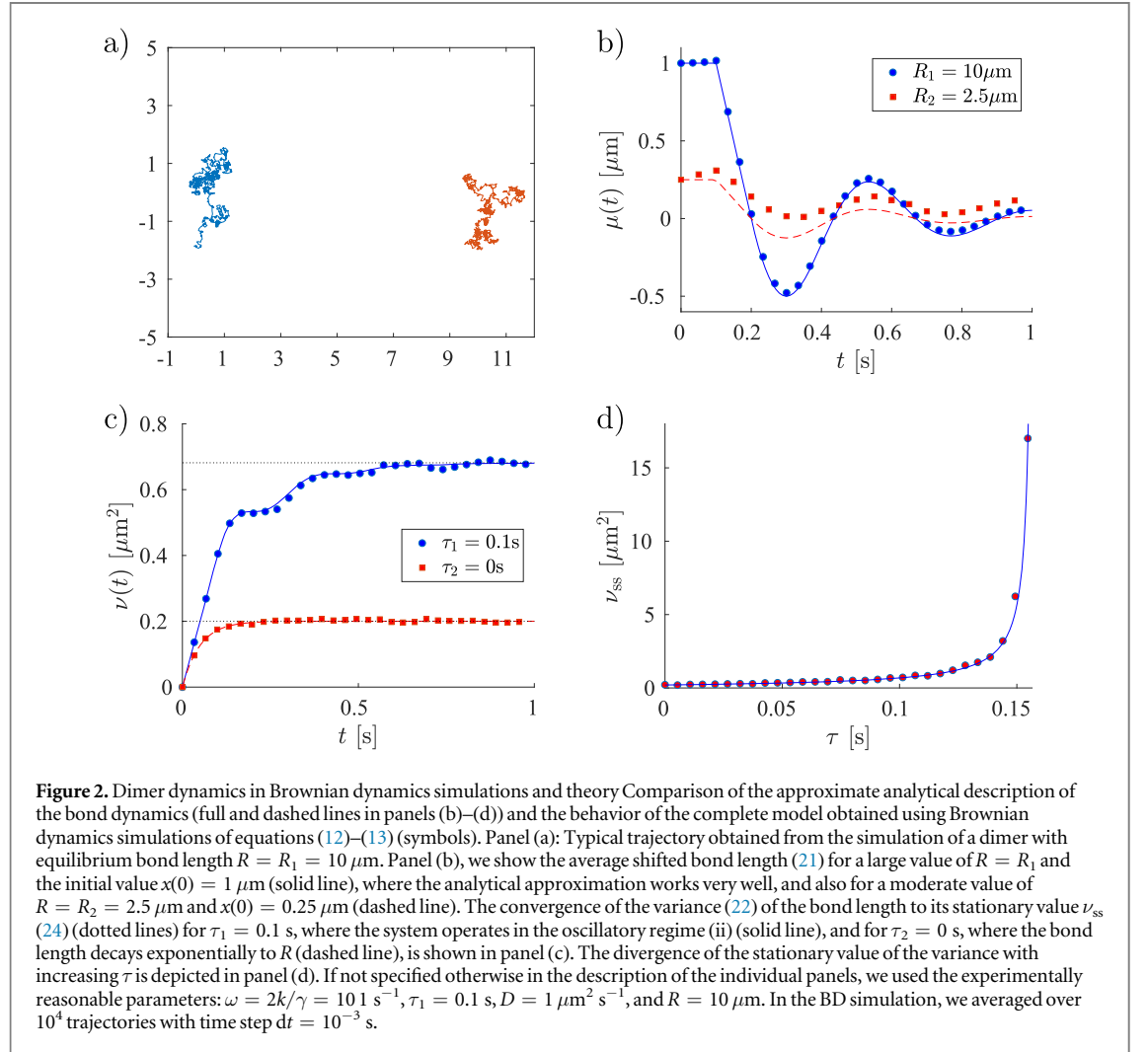
The function  $P_1(x, t|x_0, 0) dx$  stands for the probability that the system which departs with certainty from state  $x_0$  at time 0 is found at time  $t$  somewhere in the interval  $(x, x + dx)$ . Similarly,  $P_2(x, t; x', t'|x_0, 0) dx dx'$  denotes the probability that the (shifted) bond length is in the interval  $(x', x' + dx')$  at time  $t'$  and at a later time  $t$  in  $(x, x + dx)$  under the condition that it has started at time  $t = 0$  at  $x_0$ . The one-time PDF  $P_1$  possesses the same structure as the corresponding PDF for  $\tau = 0$  [63]. The non-Markov character of the process (15) with nonzero delay manifests itself in the fact that the two-time PDF  $P_2$  cannot be constructed from the one-time PDF  $P_1$ , while this is always possible for a Markov process.

The FPEs corresponding to the PDFs (19) and (20) are given by equations (B.6) and (B.7) in appendix B, respectively. Interestingly enough, the diffusion and drift terms in the FPEs are given by the natural scales  $2D$  and  $\omega_\tau$  only in the limit  $\tau \rightarrow 0$ . Moreover, both coefficients acquire a time dependence, determined by the function  $\lambda(t)$ . Specifically, the diffusion and drift coefficients in equation (B.6) for  $P_1$  read  $D_\tau(t) = D\lambda^2(t) d\left[\int_0^t ds \lambda^2(s)/\lambda^2(t)\right]/dt$  and  $\omega_\tau(t) = -\dot{\lambda}(t)/\lambda(t)$ , respectively<sup>6</sup>. While the drift coefficient in equation (B.7) for  $P_2$  is also given by  $\omega_\tau(t)$ , the diffusion coefficient reads  $2D_\tau(t) + 4D\lambda(t) \int_0^{t'} ds d[\lambda(t-s)\lambda(t'-s)/\lambda(t)]/dt$ . This difference in diffusion coefficients is the reason why the PDF  $P_2$  can be constructed from  $P_1$  in the standard way for Markov processes only for  $\tau = 0$ , where both diffusion coefficients coincide.

For  $\tau = 0$  the PDF  $P_1$  always eventually relaxes to a time independent stationary state which does not depend on the initial condition and which is described by the equilibrium Gibbs formula  $P_1(x, \infty; x_0, 0) \propto \exp(-\omega x^2/2D)$ . For a nonzero delay in the regimes (i) and (ii), i.e. when the noiseless solution (18) and thus the mean value  $\mu(t) = x_0\lambda(t)$  of  $x$  converges to 0 for  $t \rightarrow \infty$ , the system relaxes into a time-independent non-equilibrium steady state  $P_1(x, \infty; x_0, 0) \propto \exp[-\omega_\tau(\infty)x^2/D_\tau(\infty)]$  with  $\omega_\tau(\infty) \neq \omega$  and  $D_\tau(\infty) \neq 2D$ , see figure 8 in section 5.1. In these cases, our approximate model thus predicts that, in the long run, the delay merely 'deforms' the (approximate) Gaussian equilibrium distribution through a parameter renormalization. This state is characterized by a nonzero entropy production rate [56]. For long time delays, no stationary state exists. In comparison to the analogous setting with a piece-wise constant force discussed previously [11], this destabilization for long delay times  $\tau$  is a new feature, due to increasingly high systematic forces that may occur for long delays.

In the regimes (i) and (ii), the variance  $\nu(t) = \langle x^2(t) \rangle - \langle x(t) \rangle^2$  converges to the stationary value [30, 31, 34]

<sup>6</sup>The case of  $\lambda(t) = 0$ , where these coefficients diverge, is discussed in more detail in section 5.1.



$$\nu_{\text{ss}} = \lim_{t \rightarrow \infty} \nu(t) = \frac{2D}{\omega\tau} \frac{1 + \sin(\omega\tau)}{\cos(\omega\tau)}. \quad (24)$$

The derivation of this formula is given in appendix C, where we also derive an analytical expression for the stationary time correlation function  $C(t) = \lim_{s \rightarrow \infty} \langle x(s)x(s+t) \rangle$ . Note that the variance (24) diverges upon entering the unstable regime (iii) for  $\omega\tau \rightarrow \pi/2$ .

The formula (24) finally allows us to specify the regime  $\nu_{\text{ss}} \ll R^2$  where the approximation  $[r(t-\tau) - R]/r(t) \approx 0$  used in the derivation of equation (15) from equations (12) and (13) is reasonable because the PDF for  $r$  is relatively sharply peaked around the mean bond length  $R$ . As we already noted, equation (15) is also valid in the case  $2\tau D/R^2 \ll 1$  when the bond rotates only slightly in each delay interval and thus the angle  $\varphi(t, t-\tau)$  in equations (12) and (13) is small. However, also in this case we need to additionally assume that  $\nu_{\text{ss}} \ll R^2$  in order to ensure that the error caused by considering the wrong boundary condition at  $r = -R$  is negligible. Altogether, the used approximation is expected to give good results if the condition

$$\nu_{\text{ss}} \ll R^2 \quad (25)$$

is fulfilled.

An example of the stochastic evolution of the dimer obtained from BD simulations of the exact system (12) and (13) is depicted in figure 2(a). In figure 2(b), we compare the results obtained from BD simulations with the time evolution of the average shifted bond length (21) for different values of the equilibrium length  $R$ . As expected, the approximate analytical formula (21) describes well the exact result for large enough  $R$  satisfying the inequality (25). For larger values of  $\nu_{\text{ss}}/R^2$ , the analytical result underestimates the correct value. This is because the bond length in the BD simulation is obtained from equation (12) with the reflecting boundary at the origin, while we allow negative values of  $r(t)$  in the approximate analytical description. Similarly as for the mean value, the analytical formula (22) for the bond length variance  $\nu(t)$  approximates very well the value obtained from BD simulations for large enough  $R$ , as shown in figure 2(c). In figure 2(d), we depict the monotonous rapid

divergence of the stationary variance (24) as the time delay  $\tau$  reaches the unstable regime (iii). This means that the delay tends to delocalize structures. On the other hand, the variance can be optimized as a function of the frequency  $\omega_\tau$ . The best localized structure is obtained for the frequency fulfilling the formula  $\cos(\omega_\tau \tau) = \omega_\tau \tau$ , i.e.  $\omega_\tau \approx 0.74/\tau$  and thus  $2k\tau/\gamma \approx 0.74 \exp(2D\tau/R^2)$ . For the corresponding value 0.74 of the reduced delay  $\omega_\tau \tau$  the system is in the dynamical regime (ii) with exponentially decaying oscillations.

### 2.3. Trimer

Let us now consider the system composed of three particles. Then, equation (3) gives the system of six coupled equations of motion:

$$\begin{aligned}\dot{\mathbf{r}}_1(t) &= \frac{k}{\gamma} \{ [r_{12}(t-\tau) - R] \mathbf{e}_{21}(t-\tau) + [r_{13}(t-\tau) - R] \mathbf{e}_{31}(t-\tau) \} + \sqrt{2D} \boldsymbol{\eta}_1(t), \\ \dot{\mathbf{r}}_2(t) &= \frac{k}{\gamma} \{ [r_{12}(t-\tau) - R] \mathbf{e}_{12}(t-\tau) + [r_{32}(t-\tau) - R] \mathbf{e}_{32}(t-\tau) \} + \sqrt{2D} \boldsymbol{\eta}_2(t), \\ \dot{\mathbf{r}}_3(t) &= \frac{k}{\gamma} \{ [r_{32}(t-\tau) - R] \mathbf{e}_{23}(t-\tau) + [r_{31}(t-\tau) - R] \mathbf{e}_{13}(t-\tau) \} + \sqrt{2D} \boldsymbol{\eta}_3(t).\end{aligned}$$

For the relative coordinates  $\mathbf{r}_{12}(t) = \mathbf{r}_1(t) - \mathbf{r}_2(t)$  we obtain

$$\begin{aligned}\dot{\mathbf{r}}_{12}(t) &= -\omega \left\{ [r_{12}(t-\tau) - R] \mathbf{e}_{12}(t-\tau) + \frac{1}{2} [r_{13}(t-\tau) - R] \mathbf{e}_{13}(t-\tau) \right. \\ &\quad \left. + \frac{1}{2} [r_{32}(t-\tau) - R] \mathbf{e}_{32}(t-\tau) \right\} + \sqrt{2D} [\boldsymbol{\eta}_1(t) - \boldsymbol{\eta}_2(t)],\end{aligned}\quad (26)$$

where  $\omega = 2k/\gamma$  and similarly for  $\dot{\mathbf{r}}_3(t)$  and  $\dot{\mathbf{r}}_2(t)$ . To get analytical results for bond lengths  $r_{ij}(t) = |\mathbf{r}_{ij}(t)|$ , we multiply the formulas for  $\dot{\mathbf{r}}_{ij}(t)$  by the corresponding unit vectors  $\mathbf{e}_{ij}(t) = \mathbf{r}_{ij}(t)/|\mathbf{r}_{ij}(t)|$  and linearize the resulting equations. To this end, we need to deal with the expressions  $\mathbf{e}_\alpha(t-\tau) \cdot \mathbf{e}_\beta(t)$ ,  $\alpha, \beta = \text{I}, \dots, \text{III}$ , where we introduced roman numbers as a shorthand indexing  $\text{I} \equiv 12, \text{II} \equiv 13, \text{III} \equiv 32$ . For a vanishing delay  $\tau = 0$ ,  $\mathbf{e}_\alpha \cdot \mathbf{e}_\alpha = 1$  and the scalar products  $\mathbf{e}_\alpha \cdot \mathbf{e}_\beta$  describe the angles of the triangle formed by the three particles (see figure 1 in section 2). We find that up to the leading order in the equilibrium bond length  $R$  the triangle is equilateral and thus the internal angles are  $\pi/3$ , leading to the relations  $\mathbf{e}_\text{I} \cdot \mathbf{e}_\text{II} = \mathbf{e}_\text{I} \cdot \mathbf{e}_\text{III} = -\mathbf{e}_\text{II} \cdot \mathbf{e}_\text{III} \approx 1/2 + O[(r_\alpha - r_\beta)/R]$  with a correction that is on the order of  $(r_1 - r_{\text{II}})/R$  for  $\mathbf{e}_\text{I} \cdot \mathbf{e}_\text{II}$  and similarly for the other scalar products. The linearized equation for the relative coordinate  $x_\alpha \equiv r_\alpha - R$  thus reads

$$\begin{aligned}\dot{x}_\alpha(t) &= -\omega \left[ x_\alpha(t-\tau) + \frac{1}{4} x_{\alpha+\text{I}}(t-\tau) + \frac{1}{4} x_{\alpha+\text{II}}(t-\tau) \right] + \sqrt{2D} \left( \sum_{i=1}^3 \mathcal{A}_{\alpha i} \boldsymbol{\eta}_i(t) \right) \cdot \mathbf{e}_\alpha(t), \\ \mathcal{A} &= \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & 1 \end{pmatrix},\end{aligned}\quad (27)$$

where the lower index  $\alpha \equiv \alpha \bmod \text{III}$  is considered as periodic with the period 3, i.e.  $x_{\text{IV}} \equiv x_1$  and  $x_{\text{V}} \equiv x_{\text{II}}$ . Similarly as in the case of the dimer, equation (27) describes the dynamics of  $x_\alpha(t)$  well for large equilibrium bond lengths  $R$  and for time delays small compared to reorientation times of the unit vectors  $\mathbf{e}_\alpha$ .

For an analytical treatment, it is advantageous to rewrite the system (27) in the matrix form

$$\dot{\mathbf{x}}(t) = -\omega M \mathbf{x}(t-\tau) + \sqrt{2D} \boldsymbol{\xi}(t),\quad (28)$$

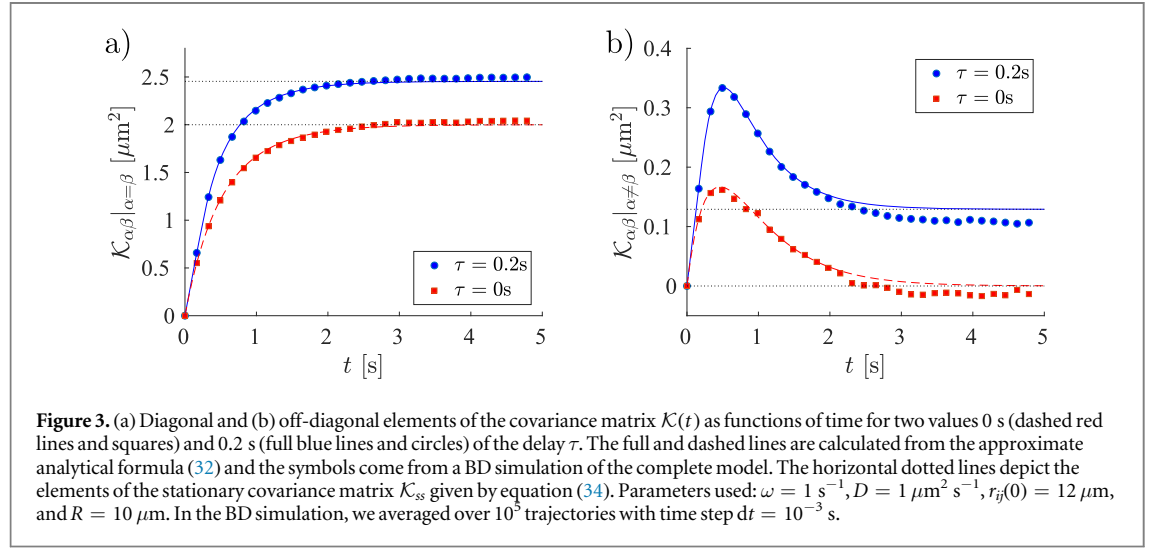
for the three-dimensional column vector  $\mathbf{x}(t) = [x_1(t), x_{\text{II}}(t), x_{\text{III}}(t)]^\top$ . In equation (28), the noise vector  $\boldsymbol{\xi}(t)$  is given by  $\boldsymbol{\xi}(t) \equiv A^1 \boldsymbol{\eta}^1(t) + A^2 \boldsymbol{\eta}^2(t)$  with the auxiliary noise vectors  $\boldsymbol{\eta}^j(t) \equiv [\eta_1^j, \eta_2^j, \eta_3^j]^\top$ ,  $j = 1, 2$ , containing the  $j$ th components of the original noises  $\boldsymbol{\eta}_i(t)$ ,  $i = 1, \dots, 3$ . From the system (27) follows that the matrices  $M, A^1$  and  $A^2$  read

$$M = \frac{1}{4} \begin{pmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{pmatrix}, A^1 = \begin{pmatrix} e_1^1 & -e_1^1 & 0 \\ e_{\text{II}}^1 & 0 & -e_{\text{II}}^1 \\ 0 & -e_{\text{III}}^1 & e_{\text{III}}^1 \end{pmatrix}, A^2 = \begin{pmatrix} e_1^2 & -e_1^2 & 0 \\ e_{\text{II}}^2 & 0 & -e_{\text{II}}^2 \\ 0 & -e_{\text{III}}^2 & e_{\text{III}}^2 \end{pmatrix},\quad (29)$$

where  $e_\alpha^j$  denote  $j$ th component of the two-dimensional unit vector  $\mathbf{e}_\alpha$ . The time-correlations between the three components of the noise vector  $\boldsymbol{\xi}(t)$  are not mutually independent and read

$$\langle \xi_\alpha(t) \xi_\beta(t') \rangle = (A^1 A^{1\top} + A^2 A^{2\top})_{\alpha\beta} \delta_D(t-t') = 2M_{\alpha\beta} \delta_D(t-t').\quad (30)$$

Due to the linearity of equation (28) and Gaussianity of the noise, the Green's function for the one-time PDF  $P_1(\mathbf{r}, t | \mathbf{r}_0, 0)$  is Gaussian [37], and determined by the mean value  $\boldsymbol{\mu}(t) = \langle \mathbf{x}(t) \rangle$  and the covariance matrix  $\mathcal{K}(t) = \langle \mathbf{x}(t) \mathbf{x}^\top(t) \rangle - \boldsymbol{\mu}(t) [\boldsymbol{\mu}(t)]^\top$ . We review in detail the derivation of these functions in appendices A and B.2.



For the initial condition  $\mathbf{x}(t) = 0$ ,  $t < 0$  and  $\mathbf{x}(0) = \mathbf{x}_0$  we get

$$\boldsymbol{\mu}(t) = \lambda(t)\mathbf{x}_0, \quad (31)$$

$$\mathcal{K}(t) = 4DM \int_0^t ds \lambda^2(s), \quad (32)$$

where  $\lambda(t)$  denotes the Green's function for equation (28) given by

$$\lambda(t) = \sum_{k=0}^{\infty} \frac{(-\omega M)^k}{k!} (t - k\tau)^k \theta(t - k\tau). \quad (33)$$

Multiplying  $\lambda(t)$  by the vector  $[1, 1, 1]^\top$ , using the formula  $M[1, 1, 1]^\top = 3/2[1, 1, 1]^\top$  and comparing the result to the one-dimensional Green's function (18), we find that  $\lambda(t)$  undergoes with increasing  $t$  (i) monotonous exponential decay to 0 for  $0 < 3\omega\tau/2 \leq 1/e$ , (ii) oscillatory exponential decay to 0 for  $1/e < 3\omega\tau/2 < \pi/2$  and (iii) oscillatory exponential divergence for  $\pi/2 < 3\omega\tau/2$ . In the regimes (i) and (ii) the stationary value of the covariance matrix reads

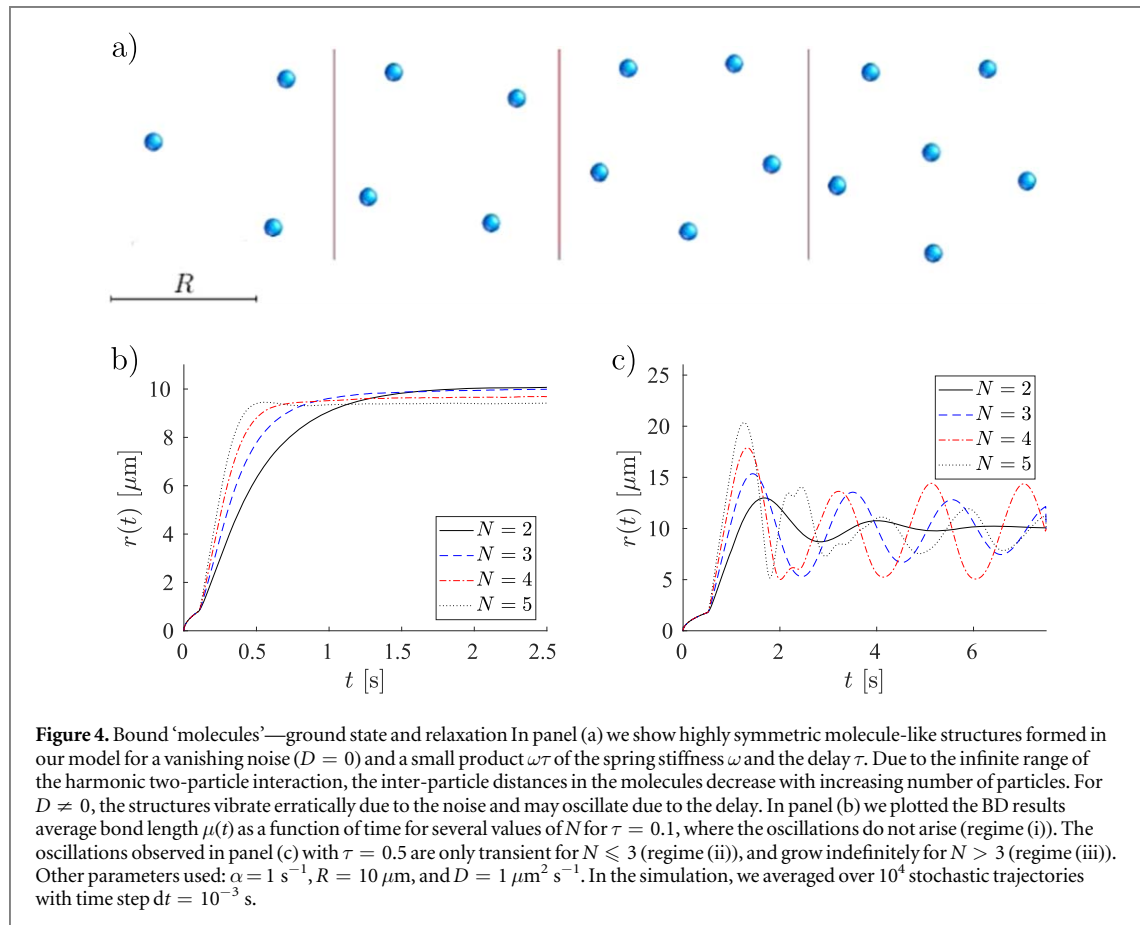
$$\mathcal{K}_{ss} = \lim_{t \rightarrow \infty} \mathcal{K}(t) = \frac{2D}{\omega} \frac{\mathcal{I} + \sin(\omega\tau M)}{\cos(\omega\tau M)}, \quad (34)$$

where  $\mathcal{I}$  denotes the identity matrix. This formula follows from the results of appendix C after substituting the matrices  $\omega$  and  $\sigma\sigma^\top$  from the formula (C.7) in the appendix by  $\omega M$  and  $4DM$ . In the appendix, we also derive an analytical expression for the stationary time correlation matrix  $C(t) = \lim_{s \rightarrow \infty} \langle \mathbf{x}(s)\mathbf{x}^\top(s+t) \rangle$ . The regime of stability  $3\omega\tau/2 < \pi/2$  of the trimer can also be determined from the condition that the matrix  $\cos(\omega\tau M)$  is not singular, i.e. its determinant  $\cos^2(3\omega\tau/2) \cos(3\omega\tau)$  is nonzero.

Due to the symmetry of the problem, all diagonal elements of the matrix  $\mathcal{K}_{ss}$  are identical and the same holds also for all its off-diagonal elements. The approximate analytical, time-dependent solution (32) for the covariance matrix is compared to the exact covariance matrix obtained by BD simulations of the complete model in figure 3. Given the approximations made, we find very good agreement. The analytical results only slightly underestimate the diagonal elements (probably for the same reason as for the dimer) and overestimate the off-diagonal elements. The behavior of the covariance matrix as a function of the frequency  $\omega$  and delay  $\tau$  is similar to the behavior of the variance (24) for the dimer. Specifically, the diagonal elements of  $\mathcal{K}_{ss}$  monotonically increase (the PDF for the bond lengths become broader) with  $\tau$  and exhibit a minimum as functions of  $\omega$ , opening a possibility to optimize the width of the bond length PDF. The off-diagonal elements of  $\mathcal{K}_{ss}$  monotonously increase (the individual bonds of the trimer become more correlated) both with the delay and with the natural relaxation frequency.

### 3. Structure formation

The approximate analytical study of the dimer and trimer revealed that both systems obey three dynamical regimes: (i) and (ii) a monotonous and an oscillatory exponential relaxation towards a steady state with the average bond length  $\mu(\infty) = R$ , respectively, and (iii) an exponential divergence. The performed BD simulations confirmed that for large  $R$ , when the model is well described by the approximate analytical formulas, these regimes can indeed be observed also in the complete model (3). Furthermore, the analytical study



predicted that the dimer is in the unstable regime (iii) for  $\omega\tau > \pi/2$  and the trimer for  $\omega\tau > \pi/3$ . Let us now discuss how general the presented findings are.

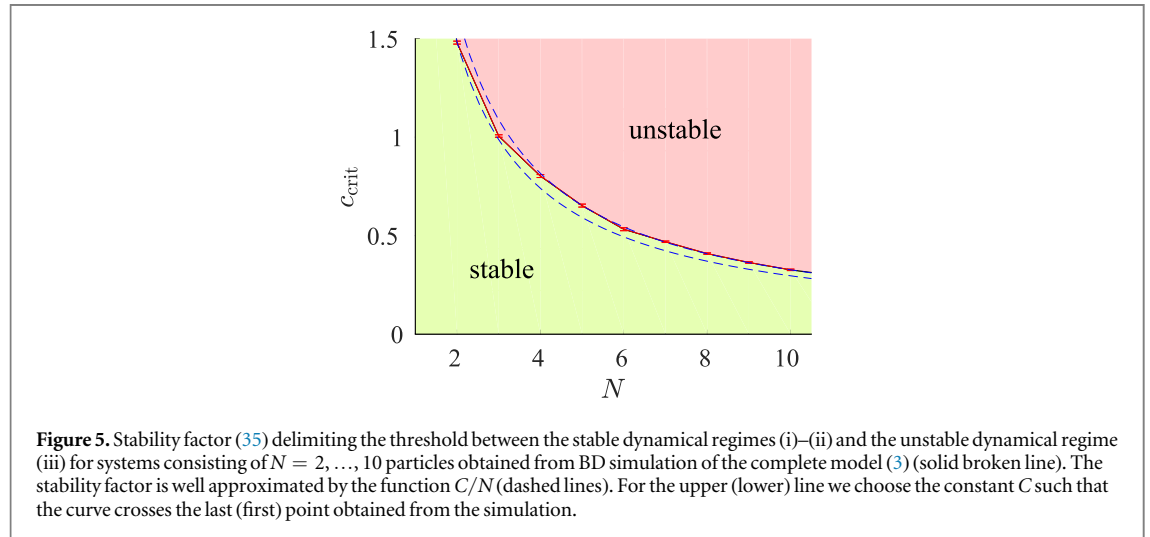
The stationary average bond length  $\mu$  can be determined by minimizing the potential energy  $V = \frac{1}{2} \sum_{(i,j)} V_2(r_{ij})$  with the two-particle potential  $V_2$  given by equation (1). Minimizing the potential in our two-dimensional geometry yields the highly symmetric molecule-like structures shown in figure 4(a). Due to the confinement to 2d, the global minimum of the potential corresponding to  $\mu = R$  is accessible only for the dimer ( $N = 2$ ) and the trimer ( $N = 3$ ). For larger molecules, the average bond length decreases as a result of the infinite range of the potential. The system asymptotically relaxes to the depicted structures if the noise  $D$  vanishes and the reduced delay time  $\omega\tau$  is small enough such that the system is in the dynamical regime (i) or (ii). Nonzero  $D$  leads to fluctuations around the asymptotic structures and large  $\omega\tau$  causes exponentially diverging oscillations.

We have solved the complete model using BD simulations and depict the behavior of the average bond length  $r(t)$  for several values of  $N$  in the dynamical regimes (i) and (ii)–(iii) in figures 4(b) and (c), respectively. In the regime (i), we observe that larger systems relax faster than those with smaller  $N$ . Furthermore, in figure 4(c), we see that larger systems oscillate with larger amplitudes and that the threshold between the regimes (ii) and (iii) is reached at smaller values of  $\omega\tau$ . More precisely, all the curves in figure 4(c) are plotted using the same parameters except for  $N$  and, while the curves for  $N \leq 3$  are in the regime (ii), the curves for  $N > 3$  correspond to the regime (iii). These observations are in accord with our analytical findings for the dimer and trimer.

By analyzing the mean bond length at late times, we have evaluated the *critical reduced delay*  $(\omega\tau)^*$  determining the threshold between the regimes (ii) and (iii) for  $N = 2, \dots, 10$ . In figure 5, we show its rescaled value

$$c_{\text{crit}} \equiv \frac{2}{\pi} (\omega\tau)^*, \quad (35)$$

where the coefficient  $2/\pi$  is introduced for the comparison to the approximate result for the dimer. We find that the stability factor is well described by  $C/N$  as suggested by the approximate analytical results for the dimer [ $(\omega\tau)^* = \pi/2$ ] and trimer [ $(\omega\tau)^* = \pi/3$ ]. However, the analytical results would imply that the constant  $C$  equals to 2, which is smaller than the value  $C \approx 3$  obtained from the complete model. The actual dimer and trimer are thus more stable than their linearized versions considered in our analytical study. The found scaling



$c_{\text{crit}} \approx C/N$  implies that the stability of a system with rescaled potential stiffness  $k \rightarrow k/N$  (to render the energy extensive [81]) would be (almost) independent of the particle number.

To better understand this behavior, let us first consider the approximate analytical model described in section 2.2. Imagine a particle in a harmonic potential centered around  $x = 0$ , which is initially located at  $x(0) = x_0 > 0$ . Assuming that  $x(t) = 0$  for  $t < 0$  and neglecting the noise, the particle does not feel any force in the time interval  $t \in [0, \tau]$  such that  $x(t) = x_0$  for all  $t \in [0, \tau]$ . In the subsequent time interval  $t \in [\tau, 2\tau]$ , the particle experiences the force  $F = -\omega_\tau x_0$  pushing it towards the opposite wall of the trap. For times  $t > 2\tau$  the force starts changing dynamically according to the earlier position at time  $t - \tau$ . The particle keeps its direction of motion until it reaches the position  $x_1$  where the force changes its sign. For large delays, the particle may stop significantly later than crossing the minimum, so that  $x_1 < 0$  and  $|x_1| > |x_0|$ . A similar process then repeats when the particle returns back, with the difference that now it reaches a maximum position  $x_2 > 0$ ,  $|x_2| > |x_1| > |x_0|$ , etc. The amplitude thus increases after each half-period of oscillation causing a diverging behavior.

In order to understand the difference between the approximate analytical and the complete (numerical) solutions of the model, it is helpful to project the latter to one dimension, where a particle moves in the double-well potential depicted figure 7 in section 5.1. We assume that the particle starts in the right well and oscillates with increasing amplitude as discussed above. After some time the amplitude becomes large enough that the particle crosses the barrier to the left well. Due to the presence of the additional well, the potential now contains much wider low-energy region compared to the purely harmonic case. The particle needs longer time to travel from one (unbounded) side of the potential to its other side, and hence also the (reduced) delay  $\omega\tau$  required for inducing diverging oscillations is larger than in the harmonic case. As a consequence, the complete model is seen to be more stable than foreseen by our analytical considerations. Moreover, our discussion reveals the existence of a fourth dynamical regime, preceding the unstable regime (iii), where the particle hops between the individual wells of the potential and the amplitude of the oscillations remains finite.

Compared to the quasi-constant velocity model investigated in [11], our analysis thus reveals two qualitative differences. First, the structures formed in the quasi-constant velocity model oscillate for arbitrary nonzero delay  $\tau$ , while in the harmonic model these oscillations appear only if  $\omega\tau$  is large enough. Second, the amplitude of the oscillations in the quasi-constant velocity model is always constant in time, while the oscillations in the harmonic model either vanish with time, if the system is in the regime (ii), or explode with time, if the system is in the regime (iii). The behavior observed in the harmonic model can be traced back to the increase of the force with the particle distance and thus we expect an analogous behavior also for other systems with time-delayed forces increasing with distance.

#### 4. Entropy fluxes

The investigated model, much as the model discussed in [11], is inspired by self-organized systems, where a feedback based on the information about the state of the system at a previous time leads to structure formation. Interpreting the delayed interactions in our model as a result of such feedback control, we can investigate the entropy flow out of the system caused by the feedback. Due to the non-analyticity of the model with quasi-constant forces considered in [11], the analysis of entropy flows in the supplementary material therein was



performed for vanishing delay only. Using the approximate Gaussian PDFs found in sections 2.2 and 2.3 for the dimer and the trimer, respectively, we can repeat this analysis with nonzero delay.

Without feedback, i.e. without the time-delayed harmonic interactions (2), the particles would spread diffusively and the system entropy would increase accordingly. The feedback control utilizes the information about the particle positions to drive the system into a non-equilibrium steady state with a time independent PDF  $P(\mathbf{x})$  and thus with a time constant configurational entropy  $\mathcal{S} = -k_B \int d\mathbf{x} P(\mathbf{x}) \log P(\mathbf{x})$ . The smaller the entropy of the non-equilibrium steady state, the more localized the steady-state structure and thus the better the result of the feedback control. Another measure of the performance of the feedback is the rate  $-\dot{\mathcal{S}}_-$  of entropy taken from the system per unit time that can also be interpreted as the amount of information pumped into the system per unit time by the feedback device. This entropy flow balances the diffusive spreading in the steady state and is thus moreover a measure of the useful ‘work’ (in units of  $\text{J K}^{-1}$ ) performed by the feedback device against thermal dispersal. Evaluating the stationary entropy production  $\dot{\mathcal{S}}_F$  due to the feedback control mechanism and the stationary entropy production  $\dot{\mathcal{S}}_D$  due to the breaking of the fluctuation-dissipation theorem in equations (3) and (4) for  $\tau > 0$ , one can define the feedback efficiency as the ratio  $\eta_F = -\dot{\mathcal{S}}_- / (\dot{\mathcal{S}}_F + \dot{\mathcal{S}}_D)$ . The entropy production  $\dot{\mathcal{S}}_D$  can be calculated along the lines of [56]. The entropy production  $\dot{\mathcal{S}}_F = q_H/T$  is the housekeeping heat flux  $q_H$  flowing to the bath at temperature  $T$ , due to the overall operation of the feedback device, divided by the bath temperature. It clearly depends on the specific technical realization of the feedback. In all known relevant realizations of the feedback in microscopic systems [11, 13, 16, 18, 19], the housekeeping heat flux is very large compared to the ‘functional’ energy fluxes in the controlled system, resulting in a large  $\dot{\mathcal{S}}_F$  compared to  $-\dot{\mathcal{S}}_-$ , so that the efficiency  $\eta_F$  of such devices is usually negligibly small.

To evaluate the entropy flow due to the feedback (the time-delayed harmonic interaction) in the present setup, we proceed along the similar lines as in [11, 64]. The center of mass coordinate of the system is not affected by the feedback and diffuses freely (see section 2.1). The structure formation due to the feedback thus occurs only on the level of the bonds. Let us now consider the time-dependent PDF  $P(\mathbf{x}, t)$  for the bonds that converges to a time-independent non-equilibrium steady state due to the competition between feedback and diffusion. The rate of change of its Shannon entropy  $\mathcal{S}(t) = -k_B \int d\mathbf{x} P(\mathbf{x}, t) \log P(\mathbf{x}, t)$  can formally be written as

$$\dot{\mathcal{S}}(t) = \dot{\mathcal{S}}_+(t) + \dot{\mathcal{S}}_-(t), \quad (36)$$

where  $\dot{\mathcal{S}}_+(t)$  stands for the positive entropy flowing into the system due to the diffusive spreading of the particles and  $\dot{\mathcal{S}}_-(t)$  corresponds to the outflow of entropy due to the feedback.

Assuming that the stochastic dynamics of the column vector  $\mathbf{x}(t)$  describing the bonds obeys the generalized Langevin equation (GLE)  $\dot{\mathbf{x}}(t) = \mathbf{F}[\mathbf{x}(t), \mathbf{x}(t - \tau)] + \sigma\eta(t)$ , where  $\eta$  denotes a zero mean Gaussian white noise with the covariance matrix  $\langle \eta_i(t) \eta_j(t') \rangle = \delta_{ij} \delta(t - t')$ , the dynamical equation for  $P(\mathbf{x}, t)$  can be written in the form [30, 65]

$$\frac{\partial}{\partial t} P(\mathbf{x}, t) = \frac{1}{2} \sum_{ij} (\sigma\sigma^\top)_{ij} \frac{\partial^2}{\partial x_i \partial x_j} P(\mathbf{x}, t) + \mathcal{L}[\mathbf{x}, t]. \quad (37)$$

In this equation, the first term on the right stands for the diffusive spreading of the PDF. The term  $\mathcal{L}[\mathbf{x}, t]$  corresponds to the time-delayed force  $\mathbf{F}[\mathbf{x}(t), \mathbf{x}(t - \tau)]$  in the Langevin equation and thus it describes the effect of the feedback. Its concrete form is not relevant for the discussion below and thus we refer to the works [30, 65] for more details about its structure.

Inserting equation (37) into the formal time derivative  $\dot{\mathcal{S}}(t) = -k_B \int d\mathbf{x} [\partial_t P(\mathbf{x}, t)] \log P(\mathbf{x}, t)$  of the system entropy  $\mathcal{S}(t)$ , we find that

$$\dot{\mathcal{S}}_+(t) = -\frac{k_B}{2} \int d\mathbf{x} \sum_{ij} (\sigma\sigma^\top)_{ij} \left( \frac{\partial^2}{\partial x_i \partial x_j} P \right) \log P = -k_B \int d\mathbf{x} (\nabla P)^\top \frac{\sigma\sigma^\top}{P} \nabla P, \quad (38)$$

$$\dot{\mathcal{S}}_-(t) = -k_B \int d\mathbf{x} \mathcal{L}[\mathbf{x}, t] \log P(\mathbf{x}, t) = \dot{\mathcal{S}}(t) - \dot{\mathcal{S}}_+(t). \quad (39)$$

The last equation allows us to calculate the amount of entropy taken out of the system due to the feedback per unit time,  $\dot{\mathcal{S}}_-(t)$ , from the PDF  $P(\mathbf{x}, t)$  without knowing the explicit form of the operator  $\mathcal{L}$ . It is interesting to adopt the Seifert’s idea of trajectory-dependent entropy [57, 66] and use equation (39) to define the stochastic (position dependent) entropy flux

$$\dot{s}_-(\mathbf{x}, t) = k_B \left[ (\nabla P)^\top \frac{\sigma\sigma^\top}{P^2} \nabla P - \frac{1}{2} \partial_t (\log P)^2 \right]. \quad (40)$$

The average flux (39) then follows as the average  $\dot{\mathcal{S}}_-(t) = \langle \dot{s}_-(\mathbf{x}, t) \rangle$  either over the PDF  $P(\mathbf{x}, t)$  or over the individual stochastic trajectories generated in a BD simulation. To the best of our knowledge, the statistics of the entropy flux (40) has not been investigated yet and thus it is not known whether its PDF fulfills some fluctuation

symmetries. Such an investigation would clearly be beyond the scope of the present paper and we leave it for a future work.

Let us now evaluate the three entropy fluxes (36), (38) and (39) for a general  $d$ -dimensional Gaussian PDF

$$P(\mathbf{x}, t) = \frac{1}{\sqrt{(2\pi)^d |\mathcal{K}(t)|}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mu(t))^\top \mathcal{K}^{-1}(t)(\mathbf{x} - \mu(t))\right], \quad (41)$$

where  $|\mathcal{K}(t)|$  denotes determinant of the covariance matrix  $\mathcal{K}(t)$ . The corresponding entropy  $\mathcal{S}(t)$  reads

$$\frac{\mathcal{S}(t)}{k_B} = \frac{d}{2} + \frac{1}{2} \log[(2\pi)^d |\mathcal{K}(t)|] \quad (42)$$

leading to the rate of change

$$\frac{\dot{\mathcal{S}}(t)}{k_B} = \frac{1}{2} \frac{d}{dt} \log|\mathcal{K}(t)|. \quad (43)$$

From equation (38) we then find that

$$\frac{\dot{\mathcal{S}}_{\pm}(t)}{k_B} = \frac{1}{2} \text{Tr}[\sigma\sigma^\top \mathcal{K}^{-1}(t)]. \quad (44)$$

For finite times, all the entropy fluxes depend on the initial conditions and can be determined from equations (39), (43) and (44).

Let us now focus on the specific setups considered in sections 2.2 and 2.3. For the dimer, we have investigated the PDF for the length of the single bond and thus  $d = 1$ . Using our analytical findings with  $\sigma\sigma^\top = 4D$  and  $\mathcal{K}(t) = \nu(t)$ , we get

$$\frac{\mathcal{S}(t)}{k_B} = 1 + \frac{1}{2} \log[(2\pi)^2 \nu(t)] \quad (45)$$

and

$$\frac{\dot{\mathcal{S}}(t)}{2Dk_B} = \frac{\lambda^2(t)}{\nu(t)}, \quad \frac{\dot{\mathcal{S}}_+(t)}{2Dk_B} = \frac{1}{\nu(t)}, \quad \frac{\dot{\mathcal{S}}_-(t)}{2Dk_B} = \frac{\lambda^2(t) - 1}{\nu(t)}, \quad (46)$$

where  $\lambda(t)$  is given by equation (18) and the variance reads  $\nu(t) = 4D \int_0^t ds \lambda^2(s)$ . Similarly, in our analytical investigation of the trimer, we have fixed the angles between the individual bonds and investigated the PDF for the three bond length only, implying that  $d = 3$ . Using  $\sigma\sigma^\top = 4DM$  and Jacobi's formula for the derivative of determinants, we obtain the expressions

$$\frac{\mathcal{S}(t)}{k_B} = \frac{3}{2} + \frac{1}{2} \log[(2\pi)^3 |\mathcal{K}(t)|] \quad (47)$$

and

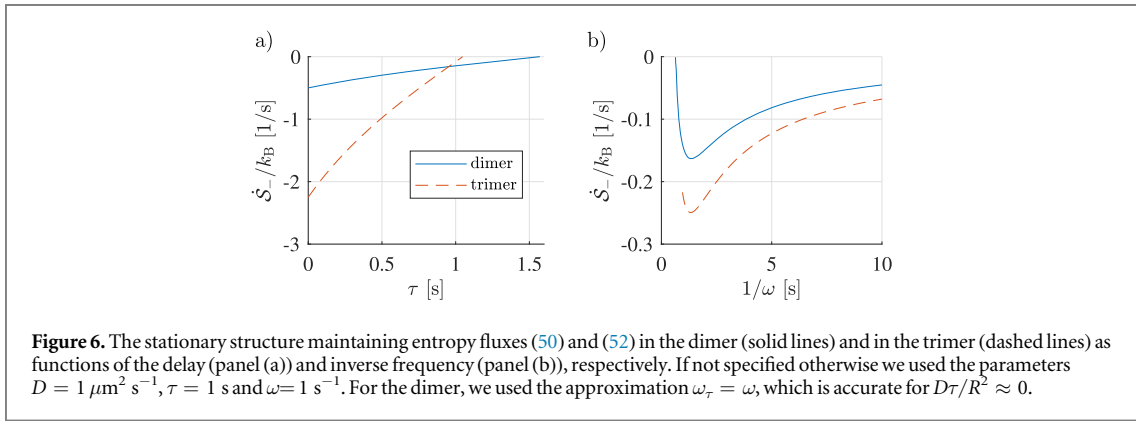
$$\frac{\dot{\mathcal{S}}(t)}{2Dk_B} = \text{Tr}\left[\frac{M\lambda^2(t)}{\mathcal{K}(t)}\right], \quad \frac{\dot{\mathcal{S}}_+(t)}{2Dk_B} = \text{Tr}\left[\frac{M}{\mathcal{K}(t)}\right], \quad \frac{\dot{\mathcal{S}}_-(t)}{2Dk_B} = \text{Tr}\left[M \frac{\lambda^2(t) - \mathcal{I}}{\mathcal{K}(t)}\right], \quad (48)$$

where  $\lambda(t)$  is given by equation (33) and the covariance matrix  $\mathcal{K}(t)$  reads  $4DM \int_0^t ds \lambda^2(s)$ .

The formulas (45)–(48) are valid both in the stable regimes (i) and (ii), where the system at long times relaxes to a stationary time-independent structured state, and in the unstable regime (iii). In the unstable regime, the variance  $\nu(t)$  and the covariance  $\mathcal{K}(t)$  diverge in time. As a result, the system entropy  $\mathcal{S}(t)$  diverges and the entropy flow  $\dot{\mathcal{S}}_+(t)$  decays to zero, because the variance of the PDF is so large that the diffusion can hardly further increase it. On the other hand, the rate of entropy change  $\dot{\mathcal{S}}(t)$  and thus also the entropy outflow  $\dot{\mathcal{S}}_-(t)$  remain finite oscillating functions, as can be seen for the dimer by using the exponential long-time approximation (A.7) for  $\lambda(t)$ , and similarly for the trimer.

For the purpose of structure formation, only the regimes (i) and (ii) are of interest, because only then the PDF reaches a time-independent non-equilibrium steady state at long times, i.e.  $\lim_{t \rightarrow \infty} \mathcal{S}(t) = S$ ,  $\lim_{t \rightarrow \infty} \dot{\mathcal{S}}(t) = 0$  and  $\dot{\mathcal{S}}_- \equiv \lim_{t \rightarrow \infty} \dot{\mathcal{S}}_-(t) = -\lim_{t \rightarrow \infty} \dot{\mathcal{S}}_+(t)$ . Let us therefore now evaluate the long-time stationary system entropies  $S$  and entropy fluxes  $\dot{\mathcal{S}}_{\pm}(t)$  maintaining the molecular-like structures formed in our model for the dimer and the trimer in these two regimes. Using the asymptotic formulas (24) and (34) for the dimer bond-length stationary variance  $\nu_{ss}$  and trimer covariance  $\mathcal{K}_{ss}$ , we find from equations (45)–(48)

$$\frac{S}{k_B} = 1 + \log(2\pi) + \frac{1}{2} \log\left[\frac{2D}{\omega_\tau} \frac{1 + \sin(\omega_\tau \tau)}{\cos(\omega_\tau \tau)}\right], \quad (49)$$



**Figure 6.** The stationary structure maintaining entropy fluxes (50) and (52) in the dimer (solid lines) and in the trimer (dashed lines) as functions of the delay (panel (a)) and inverse frequency (panel (b)), respectively. If not specified otherwise we used the parameters  $D = 1 \mu\text{m}^2 \text{s}^{-1}$ ,  $\tau = 1 \text{ s}$  and  $\omega = 1 \text{ s}^{-1}$ . For the dimer, we used the approximation  $\omega_\tau = \omega$ , which is accurate for  $D\tau/R^2 \approx 0$ .

$$\frac{\dot{S}_-}{k_B} = -\frac{\omega_\tau \cos(\omega_\tau \tau)}{1 + \sin(\omega_\tau \tau)}, \quad (50)$$

for the dimer and

$$\frac{S}{k_B} = \frac{3}{2} + \frac{3}{2} \log(2\pi) + \frac{1}{2} \log[(a-b)^2(a+2b)], \quad (51)$$

$$\frac{\dot{S}_-}{k_B} = -2D \text{Tr} \left[ \frac{M}{K_{ss}} \right] = -6D \frac{(a+b)}{(a-b)(a+2b)}, \quad (52)$$

for the trimer. In the last two formulas,  $a = \mathcal{K}_{ss}(1, 1)$  denotes diagonal and  $b = \mathcal{K}_{ss}(1, 2)$  off-diagonal elements of the covariance matrix. The system entropies (49) and (51) are determined by the width of the PDFs for the bonds. Therefore, they monotonously increase with temperature  $T = \gamma D/k_B$  and with the delay time  $\tau$ , and exhibit a minimum as functions of the frequencies  $\omega_\tau$  (dimer) and  $\omega$  (trimer), similarly as the variance  $\nu_{ss}$  and the diagonal matrix elements of the covariance matrix  $\mathcal{K}_{ss}$ . The quality of the steady-state structures is thus in our model always unfavorably influenced by the delay and, for a given delay, one can tune the frequency in order to minimize this (usually unwanted) effect.

The two entropy fluxes (50) and (52) are negative, highlighting that they correspond to entropy outflows from (or information inflows into) the system. Interestingly enough, the entropy fluxes do not depend on the temperature  $T$  (or noise strength  $D$ ) as already predicted in [11]. This means that the fluxes are discontinuous in the formal limit  $T \rightarrow 0$  because they must inevitably vanish for zero noise, where the PDF for the system evolution is a  $\delta$ -function for all times. We plot the stationary entropy fluxes (50) and (52) as functions of the delay  $\tau$  in figure 6(a) and frequency  $\omega$  in figure 6(b). Naturally, maintaining a stationary structure in a bigger system (trimer, dashed lines) requires a larger (more negative) entropy flux (or more information) than in the smaller one (dimer, solid lines). The maximum of the fluxes  $|\dot{S}_-|$ , depicted in figure 6(b), arises as a result of a competition between stronger confinement, corresponding to larger frequencies  $\omega$ , and gradual destabilization with increasing  $\omega\tau$ , when the system enters the unstable regime (iii). For the figure, we used for simplicity the approximation  $\omega_\tau \approx \omega$  for the dimer.

## 5. Transition rates for isomer transformations

The particles within the molecular structures depicted in figure 4(a) may exchange their positions. Assuming the particles to be distinguishable, different arrangements of the same structure may arise which can be interpreted as different isomers of the same molecule. Their study can provide further insight into the stability properties of our non-equilibrium molecules. In fact, we find that the study for molecules made out of only a few particles is informative also for the phenomenology observed for large particle numbers. While for the purely deterministic motion then isomer transitions only appear for time delays  $\tau$  in the unstable regime (iii), in a system affected by thermal fluctuations the transitions occur for arbitrary delays. The evaluation of the frequencies of such transitions, which measure the stability of the individual isomers, can thus provide insight into the overall energy landscape responsible for the non-equilibrium structure formation. It is the main topic of transition rate theory [23].

The transition rate  $\kappa_{A \rightarrow B}(t)$  for switching from a conformation  $A$  to a conformation  $B$  at time  $t$  can be (for arbitrary dynamics) found from the mean number of transitions  $N_{A \rightarrow B}(t)$  from  $A$  to  $B$  during an infinitesimal time interval  $\Delta t$  as  $\kappa_{A \rightarrow B}(t) = N_{A \rightarrow B}(t)/\Delta t$ . Alternatively, one can get it from the inverse mean first passage time for changing the two isomers, leading to the same results. In general, the deduced transition rates depend on the initial state of the system and on time and they can be calculated analytically only in few simple situations.

While they in principle can straightforwardly be evaluated in simulations, this can take a (forbiddingly) long time if the transition rates are small.

For an analytical treatment, it is more convenient to define the transition rate via the so-called survival probability  $S(t)$  that the system has not changed its initial isomer until time  $t$ . Our particular problem concerning the transitions between different isomers can be mapped to a particle moving in a high-dimensional energy landscape. We denote by  $S_{A \rightarrow B}(t)$  the survival probability that the system, starting in the conformation  $A$  with an absorbing boundary at the top of the barrier to conformation  $B$  (and reflecting barriers elsewhere), will remain in the configuration  $A$  until time  $t$ . The transition rate between the states  $A$  and  $B$  is then given by  $\kappa_{A \rightarrow B}(t) = \dot{S}_{A \rightarrow B}(t)/S_{A \rightarrow B}(t)$ . Hence, if the dynamical equation for the probability distribution for the state of the system with the correct boundary conditions is known, we can determine the transition rate numerically and, in some situations, even analytically.

Considering standard Markovian Langevin dynamics, the asymptotic form  $\lim_{t \rightarrow \infty} \kappa_{A \rightarrow B}(t)$  of the transition rate can (approximately) be calculated using Kramers' rate theory [22, 23] which was originally developed to describe chemical reaction rates. The approximation works best for a high energy barrier compared to the thermal energy. Kramers' theory was extended to reaction rates for GLEs describing non-Markovian systems. A crucial contribution in this direction came from Grote and Hynes [67] who derived a dynamical correction to Kramers' result. While their analysis was based on a parabolic barrier, Pollak [68] investigated the decay rate of an underdamped particle trapped in a symmetric cusp double well potential obeying the GLE with an arbitrary memory kernel satisfying the fluctuation-dissipation theorem. The time-dependent rate  $\kappa_{A \rightarrow B}(t)$  for driven overdamped systems can be calculated using the recent theory of Bullerjahn *et al* [24] for forcible molecular bond breaking.

To the best of our knowledge, the literature on the rate theory of time-delayed systems is scarce. The escape from a cubic metastable well under a time-delayed friction was investigated in [69]. Based on their small-delay approximation, Guillouzic *et al* [70] calculated the transition rate and the mean first passage time for an overdamped Brownian particle in a delayed quartic potential. From an experimental point of view, Curtin *et al* [71] studied transitions in a bistable system under time-delayed feedback.

Our model does not belong to any of the previously investigated classes of systems. However, for a vanishing delay one can use Kramers' theory, since the system obeys a Markovian overdamped Langevin equation. Moreover, for nonvanishing delays in the stable regimes (i) and (ii), the one-time PDFs for dimer and trimer can be described by standard (time-local) FPEs with time-dependent coefficients, where Bullerjahn's theory applies and where one can evaluate the transition rate numerically. Furthermore, after long times, the coefficients in these FPEs become time independent suggesting that Kramers' theory may be applied also to obtain the long-time form of the transition rates for a nonzero delay.

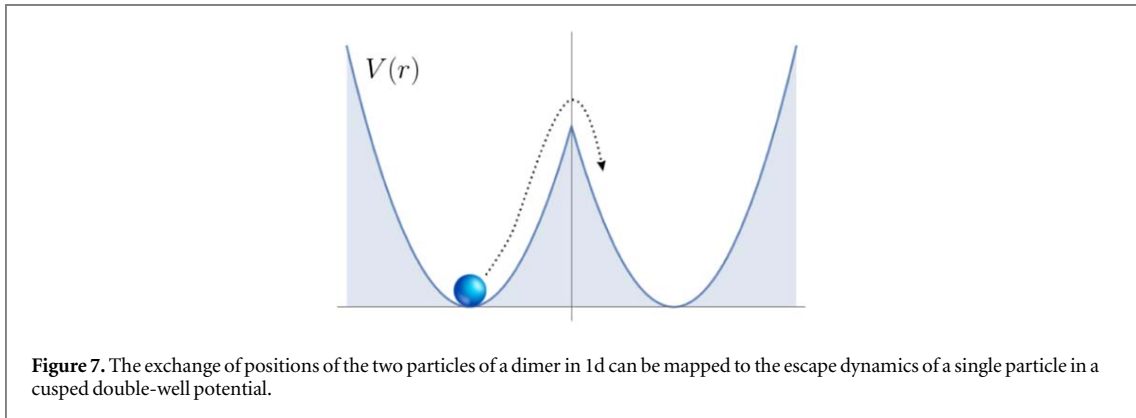
Although looking promising, all the techniques above are based on the time-local FPE. For non-zero delay, they share one drawback, which may limit their applicability to small delays: the time-local FPE is derived from solutions to the delayed Langevin equations without the absorbing boundary condition. While this represents no problem for diffusion dynamics without delay, it can cause problems in our delayed system. In the following sections, we compare predictions of Kramers' theory, Bullerjahn's theory and direct numerical evaluation of the transition rates from the time-local FPE against BD simulations of the transition rates for dimer and trimer, demonstrating that the rates obtained from the time-local FPE are indeed accurate for small and moderate delays only.

### 5.1. Dimer

To study transition rates, the simplest configuration of our model is the dimer with two distinguishable particles in one dimension. (Due to rotational symmetry, we cannot distinguish between dimer isomers in two dimensions.) The setting is described by the approximate Langevin equation (15) for the inter-particle distance  $r = |\mathbf{r}_1 - \mathbf{r}_2| > 0$ . A transition occurs when the two particles exchange their positions, and can be assigned to the moment when the bond length  $r$  vanishes. To illustrate the problem, it is useful to extend the domain of the distance variable  $r$  such that it is positive for one isomer and negative for the other. For vanishing delay, this redefined signed bond length  $\tilde{r}$  then diffuses in the cusped double-well potential  $V(\tilde{r}) = \gamma\omega(|\tilde{r} - R|^2)/2$ , depicted in figure 7, with the diffusion coefficient  $2D$ .

For a nonzero delay, based on the approximate solution (17) to the Langevin equation (15) assuming  $r \in (-\infty, \infty)$  and  $x(t) = 0$  for  $t < 0$ , we have found that the one-time PDF  $P_1 = P_1(x, t) = P_1(x, t|x_0, 0)$  obeys the FPE (B.6), which reads

$$\frac{\partial}{\partial t} P_1 = \frac{\partial}{\partial x} \left[ \omega_\tau(t)x + 2D_\tau(t) \frac{\partial}{\partial x} \right] P_1 \equiv \mathcal{L}(x, t) P_1. \quad (53)$$



This equation describes diffusion in the harmonic potential

$$V(\tilde{r}, t) = \gamma\omega_\tau(t)(|\tilde{r}| - R)^2/2, \quad (54)$$

with the time-dependent stiffness  $\gamma\omega_\tau(t)$  (given by (B.4)) and the time-dependent diffusion coefficient  $2D_\tau(t)$  (given by (B.8)).

The validity of equation (53) for  $P_1$  with natural boundary conditions suggests that one can further employ the analogy between the delayed dynamics and the (effective) Markovian model for calculating the transition rate  $\kappa(t)$  for switching between the isomers. In the Markovian case, the transition rate for surpassing the (effective) energy barrier  $\gamma\omega_\tau(t)R^2/2$  at  $r = 0$  to the other isomer can be calculated from equation (53) with an absorbing boundary at  $x = -R$  [63]. We now review several methods suitable for this task, and compare the results to BD simulations of the complete model with energy barrier  $\gamma\omega R^2/2$  and delayed dynamics. In order to study the transition rate between the isomers of the dimer analytically, it is enough to consider the dynamics of the system in one of the wells of the potential, i.e. for  $x = \tilde{r} - R > 0$ .

### 5.1.1. Numerical method

We first consider the situation when the system dwells in the state  $x(t) = 0$  for  $t \leq 0$  and then starts to diffuse in the time-dependent potential (54). Then, the time-dependent Markovian rate  $\kappa_M(t)$  can be determined from the equation

$$\partial_t \tilde{P}_a(x, t) = \mathcal{L}(x, t) \tilde{P}_a(x, t) + \kappa_M(t) \tilde{P}_a(x, t), \quad (55)$$

for the normalized PDF  $\tilde{P}_a(x, t) = P_a(x, t)/S_a(t)$  for the position of the particle surviving in the right well of the cusped potential [72]. Here,  $P_a(x, t)$  is the solution to the FPE (53) with *absorbing boundary* at  $x = -R$  and  $S_a(t) = \int_{-R}^{\infty} dx P_a(x, t)$  is the probability that the particle survives in the right well until time  $t$ . Equation (55) follows from equation (53) by inserting the definitions of the PDF  $\tilde{P}_a$  and of the transition rate

$$\kappa_M(t) = -\dot{S}_a(t)/S_a(t). \quad (56)$$

We solved it numerically using the method presented in [73].

### 5.1.2. Bullerjahn's method

Alternatively, one can determine the rate approximately using the analytical theory developed by Bullerjahn *et al* in [24]. Therein, the rate is constructed from the (Gaussian) solution  $P_1$  (20) of the FPE (53) with natural boundary conditions. Specifically, one approximates the probability current

$$j(-R, t) = \dot{S}_a(t) = -[\omega_\tau(t)x + 2D_\tau(t)\partial_x]P_a(x, t)|_{x=-R} \quad (57)$$

across the absorbing boundary by<sup>7</sup>

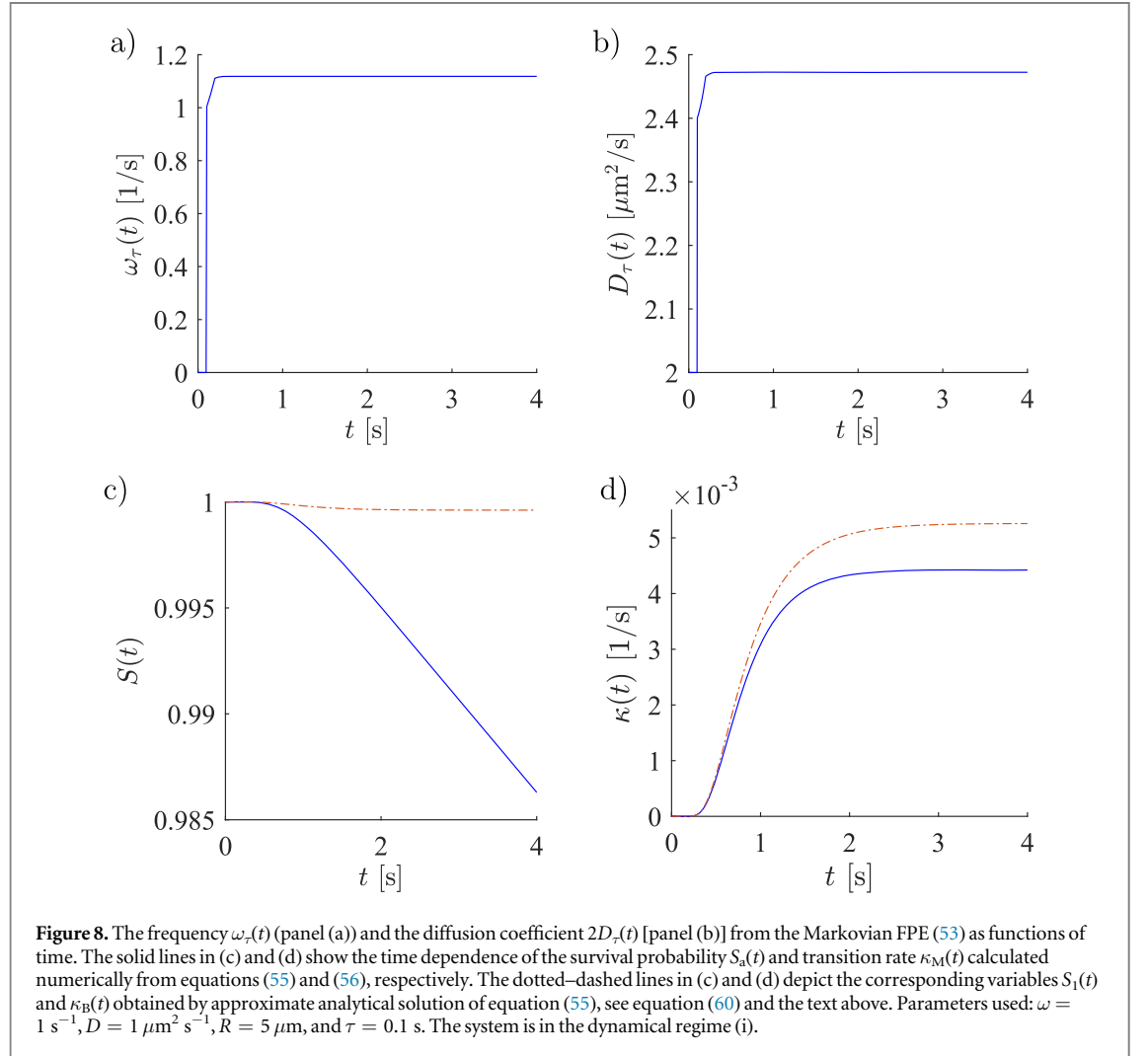
$$j^*(-R, t) \equiv -[\omega_\tau(t)x + 4D_\tau(t)\partial_x]P_1(x, t)|_{x=-R}, \quad (58)$$

and the survival probability  $S_a(t)$  by

$$S_1(t) = \int_{-R}^{\infty} dx P_1(x, t) = \frac{1}{2}[1 + \text{Erf}(R/\sqrt{2\nu(t)})]. \quad (59)$$

In the last expression, the symbol Erf denotes the error function and the variance  $\nu(t)$  is given by equation (22). The approximate Markovian transition rate is then given by

<sup>7</sup> Rescaling the diffusion coefficient by factor 2 corrects for the part of the diffusive flux that can not return to the system due to the absorbing boundary, see [24] for details.

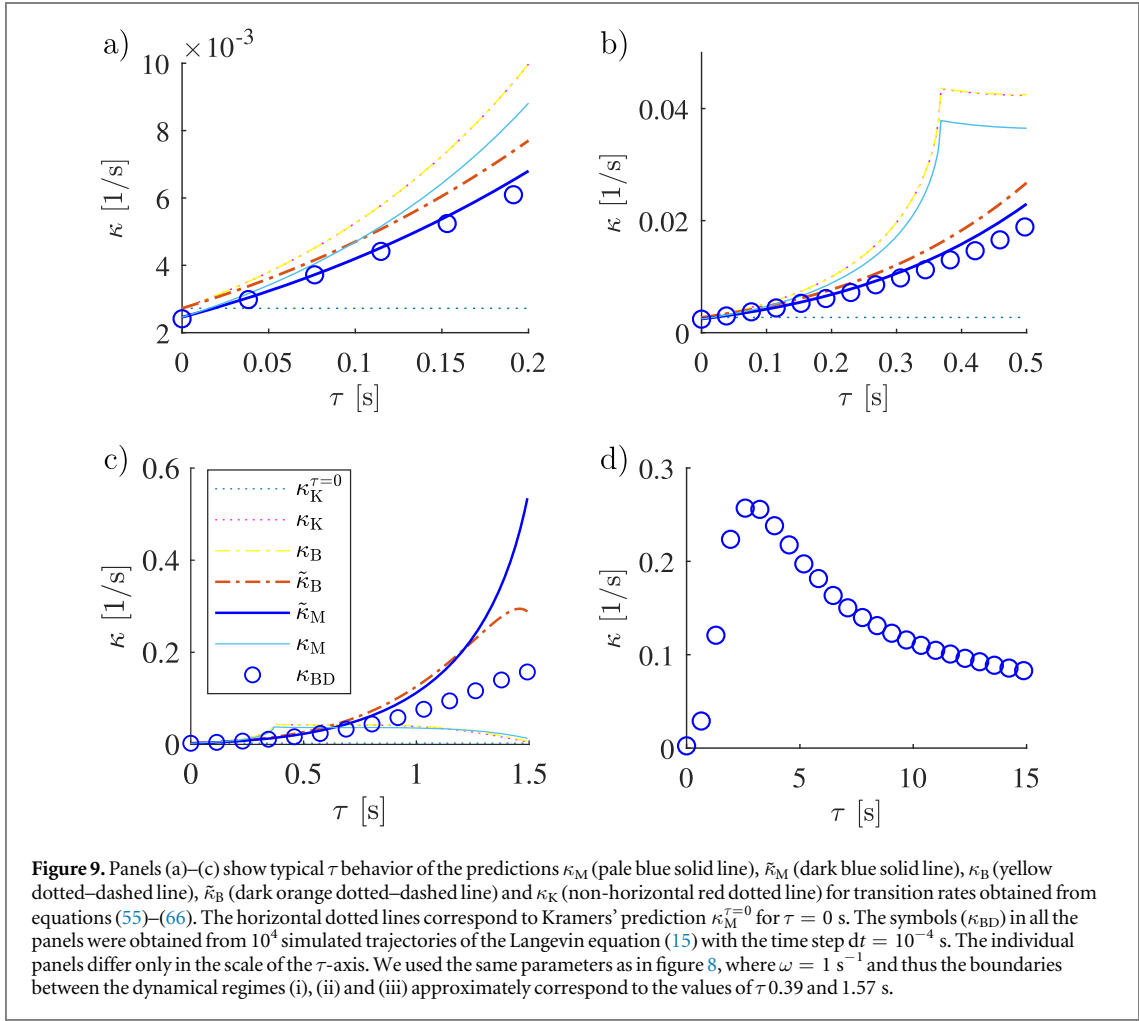


$$\kappa_B(t) = -j^*(-R, t)/S_1(t). \quad (60)$$

In figure 8, we show the frequency  $\omega_\tau(t)$ , the (effective) diffusion coefficient  $2D_\tau(t)$ , survival probabilities  $S_a(t)$  and  $S_1(t)$  and the transition rates  $\kappa_M(t)$  and  $\kappa_B(t)$  as functions of time  $t$  for parameters in the dynamical regime (i). One can observe that both the parameters  $\omega_\tau(t) = \dot{\lambda}(t)/\lambda(t)$  and  $D_\tau(t) = 2D\lambda(t)^2 + \omega_\tau(t)\nu(t)$  and the rates saturate with time. The relaxation time of  $\omega_\tau(t)$  is determined by the time in which the Green's function  $\lambda(t)$  approaches the long-time exponential representation (A.7), and the corresponding stationary value,  $\omega_\tau(\infty) = 1/t_R$ , is controlled by the relaxation time  $t_R$  for decay of  $\lambda(t)$  to 0, see also figure A1 in appendix A. The effective diffusion coefficient converges to the value  $D_\tau(\infty) = 2D[1 + \sin(\omega_\tau\tau)]/\cos(\omega_\tau\tau) \geq 2D$ , determined by the stationary variance  $\nu(\infty) = \nu_{ss}$ , see equation (24). The transition rates relax with the relaxation time  $t_R$ , similarly as the corresponding PDFs  $P_a$  and  $P_1$ . The analytical expressions for  $S_1(t)$  and  $\kappa_B(t)$  approximate the numerical results for  $S_a(t)$  and  $\kappa_M(t)$  best for short times  $t \ll t_R$ , where the PDFs  $P_a$  and  $P_1$  are still hardly affected by the different boundary conditions at  $x = -R$ . For long times and up to moderate values of time delay, the approximate analytical transition rate  $\kappa_B$  overestimates the corresponding exact rate  $\kappa_M$ , see also figures 9(a) and (b) below. For long delays, the (effective) barrier height over the (effective) thermal energy decreases so that the assumptions of the transition state theory are not valid any more, and  $\kappa_B < \kappa_M$ , see figure 9(c).

The situation of low barriers can be understood from the behavior at vanishing potential strength  $\omega \rightarrow 0$ , when  $D_\tau(t) = 2D$  and the finite time transition rates  $\kappa_M(t)$  and  $\kappa_B(t)$  can be calculated analytically. Namely, the PDFs  $\tilde{P}_a$  and  $P_1$  in the definitions (56) and (60) of the rates read  $\tilde{P}_a(x, t) = \{\exp(-x^2/4Dt) - \exp[-(x + 2R)^2/4Dt]\}/\sqrt{4\pi Dt}$  and  $P_1(x, t) = \exp(-x^2/4Dt)/\sqrt{4\pi Dt}$  [63] leading to the formulas

$$\kappa_M(t) = \frac{2R}{t} \frac{\exp(-R^2/4Dt)}{\int_{-R}^{\infty} dx \{\exp(-x^2/4Dt) - \exp[-(x + 2R)^2/4Dt]\}}, \quad (61)$$



$$\kappa_B(t) = \frac{2R}{t} \frac{\exp(-R^2/4Dt)}{\int_{-R}^{\infty} dx \exp(-x^2/4Dt)}. \quad (62)$$

Since the denominator in the expression for the rate  $\kappa_M(t)$  is smaller than that for  $\kappa_B(t)$ , we conclude that, for low energy barriers, the inequality  $\kappa_B(t) \leq \kappa_M(t)$  holds. On the other hand, for very high barriers, the PDFs  $\tilde{P}_a$  and  $P_1$  are (almost) identical because the absorbing boundary at  $x = -R$  is effectively inaccessible. In such a case, the probability current  $j(-R, t)$  is (almost) zero,  $S(t) \approx 1$ , and  $j^* \approx j - 2D\partial_x P_1|_{x=-R} < 0$  leads to the inequality  $\kappa_B(t) \geq \kappa_M(t)$ .

To gain a deeper insight into the behavior of the transition rates, let us consider the stationary regime,  $t \rightarrow \infty$ . In this regime,  $\tilde{P}_a(-R, \infty) = 0$ , due to the absorbing boundary condition at  $x = -R$ , and  $j_1(x, \infty) \equiv -[\omega_\tau(\infty)x + 2D_\tau(\infty)\partial_x]\tilde{P}_1(x, \infty)|_{x=-R} = 0$ , due to the conservation of probability  $\partial_t \tilde{P}_1(x, \infty) = \partial_t [P_1(x, \infty)/S_1(\infty)] = -\partial_x j_1(x, \infty)$ , where  $\partial_t \equiv \partial/\partial t$  and  $\partial_x \equiv \partial/\partial x$ . Then the transition rates  $\kappa_M(\infty) = -j_{Da}$  and  $\kappa_B(\infty) = -j_{D1}$  are determined solely by the diffusive fluxes  $j_{Da} \equiv 2D_\tau(\infty)\partial_x \tilde{P}_a(x, \infty)|_{x=-R}$  and  $j_{D1} \equiv 2D_\tau(\infty)\partial_x \tilde{P}_1(x, \infty)|_{x=-R}$ . The smaller the frequency  $\omega$ , the wider the PDFs  $\tilde{P}_1$  and  $\tilde{P}_a$ . For small  $\omega$ , the boundary at  $x = -R$  is in the region where the PDF  $\tilde{P}_1$  has its maximum and it is also close to the maximum of  $\tilde{P}_a$ . In such a case, the PDF  $\tilde{P}_a$ , which must vanish at  $x = -R$ , changes near the boundary faster than  $\tilde{P}_1$ , leading to  $|j_{D1}| < |j_{Da}|$  and  $\kappa_B(\infty) \leq \kappa_M(\infty)$ , in accord with the argument put forward in the previous paragraph. With increasing  $\omega$ , the boundary shifts away from the maxima of  $\tilde{P}_1$  and  $\tilde{P}_a$  towards their tails. Due to the trajectories trapped in the absorbing state [72, 74], the maximum of  $\tilde{P}_a$  is slightly farther away from the absorbing boundary than the maximum of  $\tilde{P}_1$ , and thus, with increasing  $\omega$ , the tail of  $\tilde{P}_a$ , with small derivative (small  $j_{Da}$ ), hits the boundary at  $x = -R$  before the corresponding tail of  $\tilde{P}_1$ . Hence, for large enough barrier height, the inequality between the rates crosses over to  $\kappa_B(\infty) \geq \kappa_M(\infty)$ . Finally, for very stiff traps ( $\omega \rightarrow \infty$ ), both  $j_{Da}$  and  $j_{D1}$  vanish and  $\kappa_M(\infty) = \kappa_B(\infty) = 0$ .

As shown in the figure 8(d), the transition rates converge with time to constant values in regime (i), where the limits  $\lim_{t \rightarrow \infty} \omega_\tau(t)$  and  $\lim_{t \rightarrow \infty} D_\tau(t)$  exist. Also in regime (ii), the PDF  $P_1$  assumes, after long times, the time-independent stationary form  $P_1 = \exp(-x^2/2\nu_{ss})/\sqrt{2\pi\nu_{ss}}$  with the variance  $\nu_{ss}$  given by equation (24). This

suggests that, also in this regime, the transition rate should saturate at long times. However, both the frequency  $\omega_\tau(t)$  and the diffusion coefficient  $2D_\tau(t)$  actually exhibit diverging oscillations caused by the oscillations in the Green's function  $\lambda(t)$  (18), in regime (ii). These divergences cause problems both in the FPE and in the approximate calculation of the rate using Bullerjahn's method. As a consequence, the (effective) Markov description can not be valid in the dynamical regime (ii). Nevertheless, let us now investigate to what extent the long-time transition rate  $\kappa = \kappa(\infty)$  obtained from BD simulations of the Langevin equation (15) is captured by the predictions (56) and (60) above.

### 5.1.3. Long-time behavior and Kramers' method

Assuming that at long times the PDF  $\tilde{P}$  is time-independent and the limits  $\lim_{t \rightarrow \infty} \omega_\tau(t)$  and  $\lim_{t \rightarrow \infty} D_\tau(t)$  exist, we can rewrite the formula (55) as the eigenvalue problem

$$\mathcal{L}(x, \infty) \tilde{P}_a(x, \infty) = \kappa_M \tilde{P}_a(x, \infty), \quad (63)$$

for the long time Markovian transition rate  $\kappa_M = \kappa_M(\infty)$ . We solve this formula numerically using the method described in [72, 73]. The steady state value of the transition rate predicted with Bullerjahn's method reads

$$\kappa_B = -j^*(\infty)/S_1(\infty) \quad (64)$$

with the survival probability  $S_1(\infty) = [1 + \text{Erf}(R/\sqrt{2}\nu_{ss})]/2$  and the probability current  $j^*(\infty) = \omega_\tau(\infty) R \exp(-R^2/2\nu_{ss})/\sqrt{2\pi\nu_{ss}}$ .

For high barriers, where  $S_1(\infty) \approx 1$ , the long time form of Bullerjahn's transition rate coincides with the classical prediction by Kramers [22, 23] for the transition rate for leaving one of the wells of a cusped potential with barrier height  $E_b$ ,

$$\kappa_K = \frac{2}{\sqrt{\pi} R^2} \frac{(\beta E_b)^{3/2}}{\beta \gamma} \exp(-\beta E_b) = \frac{\omega_\tau(\infty) R}{\sqrt{2\pi\nu_{ss}}} \exp\left(-\frac{R^2}{2\nu_{ss}}\right) = j^*(\infty). \quad (65)$$

In the penultimate equality, we used the appropriate inverse thermal energy  $\beta = 1/2\gamma D_\tau(\infty)$  and barrier height  $E_b = \gamma \omega_\tau(\infty) R^2/2$ , and also the asymptotic form of the diffusion coefficient  $D_\tau(\infty) = \omega_\tau(\infty) \nu_{ss}$ , which follows from the condition  $\lim_{t \rightarrow \infty} \lambda(t) = 0$ , valid in the dynamical regimes (i) and (ii).

### 5.1.4. Renormalized transition rates

Interestingly, the term  $\omega_\tau(t)$ , which causes divergences of the diffusion coefficient and the frequency of the potential in the FPE (53), does not enter the argument of the exponential in the rates  $\kappa_B$  and  $\kappa_K$ . This means that it just determines the kinetic prefactor, as can be also observed directly from the long-time form  $\partial_t P_1 = \omega_\tau(t) \partial_x (x + 2\nu_{ss} \partial_x) P_1$  of the FPE (53). In the dynamical regime (ii), the kinetic prefactor in equation (65) cannot be correct, due to the diverging oscillations in the time-dependent frequency  $\omega_\tau(t)$ . Nevertheless, the exponential term seems to be reasonable, and thus it is tempting to use in the prefactor of the transition rates simply  $\omega$ , instead of the problematic  $\omega_\tau(\infty)$ . This substitution gives the correct pre-exponential factor of the rate for vanishing delay  $\tau = 0$ , where  $\omega_\tau(\infty) = \omega$  and  $2D_\tau(\infty) = 2D$ . We denote the rates with the renormalized prefactor as

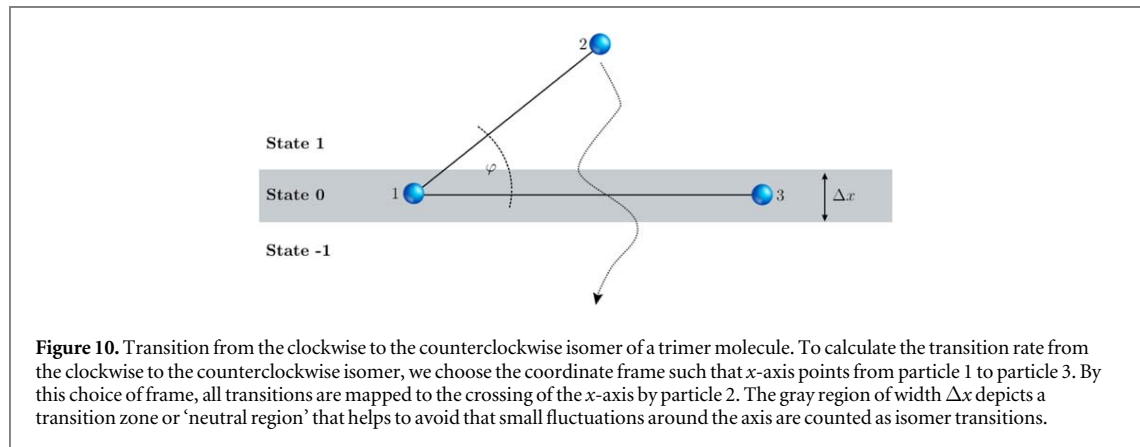
$$\tilde{\kappa}_x = \omega \kappa_x / \omega_\tau(\infty) \quad (66)$$

with  $x = M, B$  or  $K$  indicating Markov, Bullerjahn, or Kramers, respectively.

The necessity to change the kinetic prefactor in the rates stems from the fact that, although the absorbing boundary condition we used in equation (63) is correct for Markov dynamics ( $\tau = 0$ ), it can not be precisely valid for the time-delayed dynamics ( $\tau > 0$ ). To see this, it is enough to realize that the delayed system arriving at the boundary at a time  $t$  does not feel the energy barrier  $E_b = \gamma \omega R^2/2$ , but the barrier with energy  $\gamma \omega [r(t - \tau)]^2/2$ .

In figures 9(a)–(c), we compare the various analytical predictions  $\kappa_x$  and  $\tilde{\kappa}_x$ ,  $x = M, B$  or  $K$ , for the long-time transition rate (lines) with the asymptotic rate  $\kappa$  (symbols) calculated from BD simulations of equation (15) using the inverse first passage time for reaching the absorbing boundary at  $r = 0$ . For a broad range of parameters fulfilling  $k_B T \ll V(-R, 0)$ , we have found that the rate  $\kappa$  can be predicted reasonably well only for values of  $\tau$  in the dynamical regime (i) ( $\omega\tau < 1/e \approx 0.37$ ) and in the first part of the dynamical regime (ii) ( $\omega\tau < \pi/2 \approx 1.57$ ). The rate  $\kappa$  is best approximated by the expression  $\tilde{\kappa}_M$  obtained numerically from equation (63) with  $\omega$  substituted for  $\omega_\tau(\infty)$  in the operator  $\mathcal{L}$ . From the analytical expressions the re-scaled Bullerjahn expression  $\tilde{\kappa}_B$  (see equations (64) and (66)) works best. However, in the figure we have used parameters leading to a high barrier  $E_b$ , and thus Kramers' and Bullerjahn's predictions,  $\kappa_K$  and  $\kappa_B$ , almost coincide. As a consequence, the line for  $\tilde{\kappa}_K$  in the figure overlaps with  $\tilde{\kappa}_B$ , similarly as the line  $\kappa_K$  (suppressed in the figure for better readability) overlaps with  $\kappa_B$ .





### 5.1.5. Delay-dependence of $\kappa$

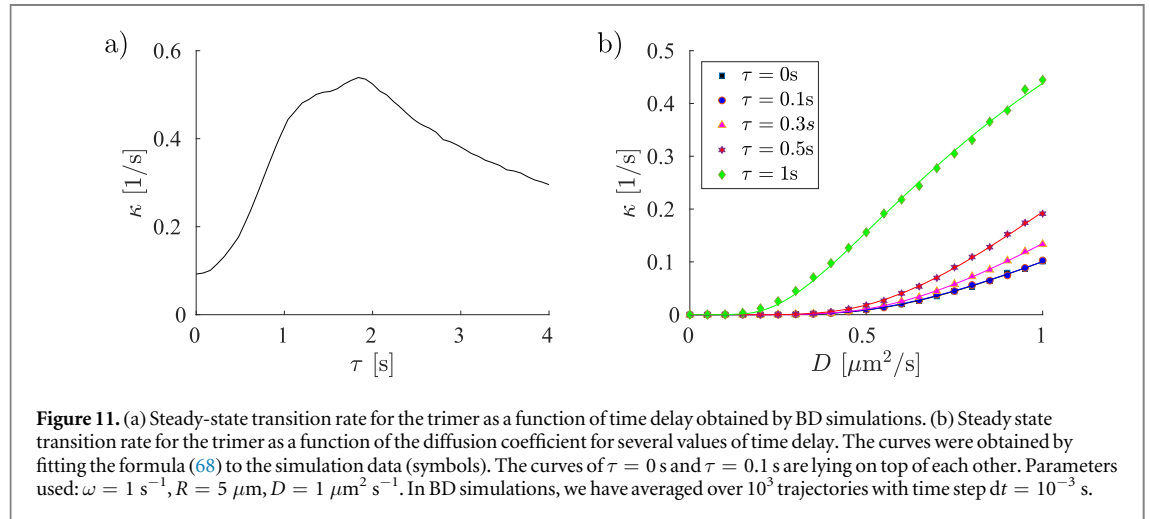
In figure 9(d), we also show the behavior of the rate  $\kappa$  in the parameter regime (iii) inaccessible to the analytical and numerical formulas due to the diverging oscillation. The simulated transition rate in figures 9(a)–(d) is first approximately exponential and thus convex in  $\tau$ , then its curvature changes to concave and it runs through a maximum and, finally, the rate starts to decrease. The value of  $\tau$  where the curvature changes sign coincides with the boundary  $\pi/2\omega\tau \approx 1.57$  between the dynamical regimes (ii) and (iii). Interestingly, no qualitative change of  $\kappa(\tau)$  is observed at the boundary  $\omega\tau = 1/e \approx 0.37$  between the regimes (i) and (ii). It is tempting to attribute, the (approximate) exponential increase of the rate with  $\tau$  in regimes (i) and (ii) to the increase of the steady-state variance  $\nu_{ss}$ , which is given as the ratio of the effective energy barrier  $V(-R, \infty)$  and the effective thermal energy  $\gamma/2D_\tau(\infty)$ . Although this explanation may work well for small delays, it breaks down for values of  $\tau$  in the second half of the regime (ii), where the actual rate  $\kappa$  is no longer well approximated by our analytical and numerical predictions. This means that the identification of the parameters  $\gamma\omega_\tau(t)$  and  $2D_\tau(t)$  with the effective potential stiffness and the effective diffusion coefficient, respectively, suggested by the effective Markov equation (53), is reasonable for relatively small values of  $\tau$  only. In the dynamical regime (iii), the particle undergoes oscillations with an amplitude that increases both with  $\tau$  and  $t$ . The corresponding transition rate, obtained from the BD simulations, thus decreases with the delay  $\tau$  as a result of oscillations leading away from the transition boundary at  $r = -R$ . Note that, in this regime, the stationary transition rate actually does not exist, since the amount of time spent distant from the boundary increases with  $t$  (so that the transition rate decreases with  $t$ ), where  $t$  is the duration of the simulation.

## 5.2. Trimer

Consider the trimer depicted in figure 10 with the three distinguishable particles labeled by the numbers 1–3. We can count the particles either clockwise or anticlockwise and thus two different isomers can form in two dimensions. As in the case of the dimer discussed above, transitions between the two isomers occur with a transition rate depending on the diffusion coefficient, the coupling strength, and the equilibrium spring length.

There are several ways how the clockwise isomer may turn into the anticlockwise one and vice versa. For example, the particles 1 and 2 can switch their positions, or the particle 2 can migrate from above the line connecting particles 1 and 3 to below that line, to name a few. The transition rate for hopping between the two isomers is then given as a sum of the transition rates for all the possible realizations of the transition. In order to make an analytical prediction for the transition rate, we choose the coordinate frame in such a way that the  $x$ -axis always points from particle 1 to particle 3 (see figure 10). Then all possible transitions between the two isomers boil down to a single event when particle 2 crosses the  $x$ -axis. In particular, this also includes the transitions due to exchanging particles 1 and 3. In this case, the direction of the  $x$ -axes changes and thus the particle 2 effectively moves to its other side. In the following, we estimate the long-time transition rate  $\kappa = \kappa(\infty)$  for the isomer transition in the steady state by means of Kramers’ theory and compare it to BD simulations.

In the BD simulations, we have evaluated the rate  $\kappa$  using the angle  $\varphi$  between the abscissas [12] and [13] and a neutral region of width  $\Delta x = \sqrt{3/16} R$  as exemplarily shown in figure 10. A neutral state  $|0\rangle$  is introduced to avoid over counting due to fluctuations of  $\varphi$  around 0 and  $\pi$ . It is occupied if the smallest height of the triangle formed by the three particles, is smaller  $\Delta x/2$ , i.e. either if  $|\varphi| < \phi_> \equiv \max[\arcsin(\Delta x/2r_{12}), \arcsin(\Delta x/2r_{13})]$  or  $|\varphi| > \phi_< \equiv \arccos(\Delta x/2r_{12}) + \arccos(\Delta x/2r_{13})$ . For  $\varphi \in [\phi_>, \phi_<]$  the system is said to be in the clockwise state  $|1\rangle$ , while for  $\varphi \in [-\phi_>, -\phi_<]$  it is in the counter-clockwise state  $|-1\rangle$ . To calculate the transition rate, we have counted the number of transitions between the states  $|\pm 1\rangle$  during a specific simulation time window, where the transition occurred if the system underwent the sequence of states  $|1\rangle \rightarrow |0\rangle \rightarrow |-1\rangle$  or  $|-1\rangle \rightarrow |0\rangle \rightarrow |1\rangle$ .



The resulting transition rate  $\kappa$  as a function of the time delay  $\tau$  is depicted in figure 11(a). Therein, one can observe the four dynamical regimes described in section 3. For small delays the rate is approximately exponential, then the curvature of  $\kappa(\tau)$  changes from positive to negative, after which the derivative of the rate starts to increase due to the appearance of the fourth dynamical regime where the particle hops between the individual minima of the potential, and for large delays in the unstable regime (iii) the rate drops, while the particle exhibits diverging fluctuations. The transition rate is qualitatively similar to that obtained for the dimer in figure 9. The only difference is that for the dimer we have not observed the fourth dynamical regime, because we have calculated the rate using the first-passage time method, which is insensitive to the potential shape beyond the boundary.

For an approximate analytical treatment, we map the situation to a one-dimensional transition problem, to which we apply Kramers' theory. Specifically, we focus on the distance  $r_b$  of the particle 2 from the  $x$ -axis and construct an appropriate effective energy barrier  $E_{b,\text{eff}}$  and diffusion coefficient  $D_{\text{eff}}$ . In the steady state, the three particles are most likely found at the vertices of an equilateral triangle. Using the notation of section 2.3, we express the vector  $\mathbf{r}_b$  from the particle 2 to the center of the abscissa |13| as  $\mathbf{r}_b = \mathbf{r}_{12} + \mathbf{r}_{31}/2$ . Hence the Gaussian white noise  $\boldsymbol{\eta}_b(t)$  corresponding to the coordinate  $\mathbf{r}_b$  is given by

$$\sqrt{4D}(\boldsymbol{\eta}_1(t) - \boldsymbol{\eta}_2(t)) + \frac{1}{2}\sqrt{4D}(\boldsymbol{\eta}_3(t) - \boldsymbol{\eta}_1(t)) \equiv \sqrt{2(3D)}\boldsymbol{\eta}_b(t), \quad (67)$$

where we have used the relations (4). Based on these considerations, we approximate the effective diffusion coefficient by  $D_{\text{eff}} = 3D$ .

For the corresponding effective energy barrier, we make the ansatz  $E_{b,\text{eff}} = C(\tau)\gamma\omega R^2/12$ . Namely, we first consider the minimum value  $E_b = kR^2/6 = \gamma\omega R^2/12$  of the energy difference between a configuration with all the three particles aligned ( $V = kR^2/6$ ) and the configuration where the particles form an equilateral triangle with side length  $R$  ( $V = 0$ ), where  $V$  is given by equation (2). The (possibly delay-dependent) unknown dimensionless prefactor  $C(\tau)$  accounts for the time delay and all additional ways the particle 2 may take in the multidimensional energy profile in order to pass from  $|1\rangle$  to  $|-1\rangle$ . The resulting Kramers' rate for the transition from  $|1\rangle$  to  $|-1\rangle$  thus reads

$$\kappa_{\text{TST}} = a(\tau)\exp\left\{-\frac{C(\tau)\omega R^2}{36D}\right\}, \quad (68)$$

where  $a(\tau)$  is a further unknown prefactor that may depend on all parameters of the model (see, for example, equation (65) for Kramers' rate in a cusped potential).

In order to test the formula (68), we have fitted it to the transition rate  $\kappa$  obtained from BD simulations as a function of the diffusion coefficient  $D$ . The results for different values of the time delay are shown in figure 11(b) and the corresponding values of the coefficients  $C(\tau)$  and  $a(\tau)$ , obtained from the fits, are given in table 1. The presented results prove that the transition rate exhibits exponential increase with the diffusion coefficient for a relatively broad range of values of the time delay, and thus the  $D$ -dependence of  $\kappa$  can be relatively well described by the Kramers-type ansatz (65).

Our attempts to include the effects of the delay in the effective diffusion coefficient and energy barrier, analogously to the approach described in section 5.1, where we made the replacements  $\omega \rightarrow \omega_{\tau,\text{ss}}$  and  $D \rightarrow D_{\tau,\text{ss}}$ , did not lead to a significant change of the value for  $C(\tau)$ . We thus conclude that the deviations of  $C(\tau)$  from unity are mainly caused by the assumption that the particle will almost in all cases cross the axis at the

**Table 1.** The phenomenological parameters  $C(\tau)$  and  $a(\tau)$  for five values of the time delay  $\tau$  corresponding to the three different dynamical regimes of the trimer isomerization. The presented values were obtained by fitting the formula (68) to the BD data shown in figure 11(b).

Regime	$\tau$ (s)	$C(\tau)$ (1)	$a(\tau)$ (s <sup>-1</sup> )
(i) Exponential decay	0	3.52 ± 0.08	1.16 ± 0.09
(i) Exponential decay	0.1	3.45 ± 0.01	1.10 ± 0.01
(i) Exponential decay	0.3	3.65 ± 0.07	1.69 ± 0.10
(ii) Damped oscillations	0.5	3.41 ± 0.07	2.07 ± 0.12
(iii) Exponential divergence	1.0	1.48 ± 0.05	1.23 ± 0.05

minimum of the potential energy. Further improvement would thus require a multidimensional transition rate theory, which is clearly beyond the scope of the present paper.

## 6. Extensions to other memory kernels

So far, we have considered only the interactions involving a given positive delay time. But how robust are our analytical findings? Do they critically hinge on (possibly artificial) model details and break down upon some minor variation of the model definition? It turns out that most of the presented results can be directly applied also to other models with delay or memory. To see this, note that the central equation (B.1) is equivalent to the GLE

$$\dot{x} = -\omega \int_0^t dt' \phi(t-t')x(t') + \sigma\eta(t) \quad (69)$$

with positive frequency  $\omega > 0$  and the memory kernel given by  $\phi(t) = \delta(t - \tau)$ ,  $\tau > 0$ . Deriving a time-local Langevin equation from the GLE (69) with arbitrary  $\phi(t)$  along the lines of appendix B.1, we obtain equation (B.3) with  $\lambda(t)$  being the Green's function for equation (69), i.e. solving (69) with the initial condition  $\lambda(0) = 1$  and  $\lambda(t) = 0$ ,  $t < 0$ . Therefore, all our results that do not depend on the specific form of the Green's function can be readily generalized to arbitrary  $\phi(t)$  after substituting the Green's function (A.5) for  $\phi(t) = \delta(t - \tau)$  by the Green's function corresponding to the chosen  $\phi(t)$ . In the rest of this section, we review some paradigmatic examples of memory kernels  $\phi(t)$ , to provide readers with a set of examples of the potential generalizations we have in mind.

The simplest generalization of systems with the memory kernel  $\phi(t) = \delta(t - \tau)$  are systems with multiple different time delays with the memory kernel

$$\phi(t) = \sum_{i=1}^N \omega_i \delta(t - \tau_i). \quad (70)$$

Properties of these systems are studied in [35].

A slightly unusual but interesting variation that makes sense in the context of active matter employs a negative delay. The individual active agents may react to a future state of their neighborhood which they predict in the basis of its present state. For idealized systems capable of a perfect prediction, the GLE (or, equivalently, the linear SDDE) contains the memory kernel  $\phi(t) = \delta(t + \tau)$ ,  $\tau > 0$  and it can be solved using the strategy described in appendix A. The resulting Green's function

$$\lambda(t) = \sum_{k=0}^{\infty} \frac{(-\omega)^k}{k!} (t + k\tau)^k \quad (71)$$

is the so called Bruwier series [75] that is convergent for  $|\omega| < |e\tau|$ . Alternatively, the series (71) can be written in the form [76]

$$\lambda(t) = \frac{e^{st}}{1 - \tau s}, \quad (72)$$

where  $s$  is the absolute value of the smallest root of the equation  $\rho = -\omega e^{\tau\rho}$ .

More realistic predictive systems might instead only have an imperfect knowledge of the future position and anticipate a position  $x_{\text{pre}}(t + \tau) \neq x(t + \tau)$ . Therefore, we reformulate the deterministic part of the GLE (69) as

$$\dot{x}(t) = -\omega x_{\text{pre}}(t + \tau). \quad (73)$$

One of the reasonable strategies for predicting  $x_{\text{pre}}$  is to use the linear extrapolation

$$x_{\text{pre}}(t + \tau) = x(t) + \tau\dot{x}(t), \quad (74)$$

which is identical to a small delay expansion of  $x(t + \tau)$ . The equation of motion (73) then assumes the time local form

$$(1 + \omega\tau)\dot{x}(t) = -\omega x(t) \quad (75)$$

which is solved by  $x(t) = x_0\lambda(t)$  with the exponentially decaying Green's function

$$\lambda(t) = \exp\left[-\frac{\omega}{1 + \omega\tau}(t - t_0)\right]. \quad (76)$$

The rescaled frequency  $\omega_r = \omega/(1 + \omega\tau)$  decreases with increasing delay and thus the resulting dynamics in general exhibits slower relaxation and larger fluctuations (variance) than a system with vanishing time delay. Note that for conventional time delays (now corresponding to  $\tau < 0$  in equation (73)), the presented first order approximation predicts dynamics with rescaled frequency  $\omega_r$  increasing with increasing time delay (for  $\omega\tau < -1$ ). However, such dynamics would lead to smaller variance  $\nu_{ss}$  for non-zero delays than for a vanishing delay, which contradicts our exact result (24), highlighting the limited practical significance of the naive small-delay expansion, as already pointed out in [30].

The most frequently used generic form of the memory kernel is the exponential

$$\phi(t) = b \exp(-bt) \quad (77)$$

which is obtained, for example, after integrating out the momentum in the Langevin equation for position of an underdamped harmonic oscillator. This memory kernel leads to the corresponding Green's function

$$\lambda(t) = \exp\left(-\frac{bt}{2}\right) \left[ \cos(\Omega t) + \frac{b}{2\Omega} \sin(\Omega t) \right], \quad (78)$$

where  $\Omega = \sqrt{\omega^2 - b^2/4}$ , which is reminiscent of the Green's function (A.7) for the system with conventional time delay [ $\phi(t) = \delta(t - \tau)$ ]. The difference is that, while the Green's function (78) always decays exponentially with time, the Green's function (A.7) allows also for negative relaxation times and thus an exponential increase with time. Properties of the Green's function (78), as well as those corresponding to a power-law memory, are discussed in more detail in [77].

## 7. Conclusion and outlook

Inspired by the surging interest in self-organized active matter and, more specifically, the experiments of Khadka *et al* [11], we considered  $N$  Brownian particles interacting via time-delayed harmonic interactions and confined to a plane, as depicted in figure 1 in section 2. The system is described by the set (3) of  $2N$  nonlinear delayed Langevin equations and hence its dynamics is non-Markovian. At long times, the particles form highly symmetric dynamical molecular-like structures, depicted in figure 4(a) in section 3, which become increasingly compact for large  $N$ .

We have analyzed small systems of  $N = 2$  (dimer) and  $N = 3$  (trimer) particles analytically finding molecules with nearest-neighbor distance given by the equilibrium spring length  $R$ . To this end, we linearized the corresponding Langevin equations around the zero-temperature steady-state configurations, or, equivalently, around the minimum of the potential energy (2). The linearized Langevin equations could be solved analytically, leading to Gaussian stationary probability densities with delay-dependent effective parameters. In the appendices, we provide analytical expressions for mean values, covariance matrix and time-correlation matrix for a multidimensional system of linear delayed Langevin equations. For the dimer and trimer, we have compared our analytical predictions with BD simulations of the complete model (3). We have found good quantitative agreement in the parameter regimes where the system evolves relatively close to its minimum energy configuration, and good qualitative agreement otherwise (see section 2).

Our analytical results for the dimer and trimer imply that these structures are stable only for small enough values of the product  $k\tau$ , where  $k$  denotes the stiffness of the potential. More precisely, these systems converge either exponentially or by exponentially-damped oscillations to corresponding steady states, or they exhibit exponentially diverging oscillations. Our analysis of systems with  $N \geq 2$  by BD simulations, described in section 3, reveals that these dynamical regimes are stable beyond the linearization approximation and for an arbitrary number of particles. Specifically, we have found that the stability actually extends to larger values of  $k\tau$  than predicted from the linearized equations, the critical value of the product  $k\tau$  decaying approximately as  $1/N$ . Therefore, larger systems are more unstable than smaller ones, and the dependence of the stability on the particle number almost vanishes after rescaling the potential stiffness as  $k \rightarrow k/N$ . We conjecture that these instabilities are induced by the chosen form of the interaction which has infinite range and diverges with increasing inter-particle distance. In contrast, the model with constant forces, considered in [11], did not lead to unstable behavior.

Interpreting the inter-particle interactions as an action of a feedback control mechanism, we have, in section 4, used our analytical results for the dimer and trimer to evaluate the amount of entropy extracted by the feedback from (or information injected to) the system in order to maintain the non-equilibrium structures. Interestingly

enough, the entropy fluxes do not depend on the noise amplitude  $D$  and hence they are discontinuous at  $D = 0$ , where the steady-state structures are stable without feedback and thus the entropy fluxes vanish.

Assuming the particles to be distinguishable, the steady-state structures (molecules) can form different isomers. Their transition dynamics can provide rich additional insight into the energy landscape underlying the non-equilibrium structure formation. For the dimer and trimer, we have investigated how and when the transitions between the individual isomers can be described by transition state theory. For the dimer, we have applied our analytical results, based on the time-convolutionless transform leading to the time-local FPE (53), to construct several analytical approximations for the transition rate using Kramers' theory [22, 23] and Bullerjahn's theory [24]. We have also calculated the transition rate from the FPE numerically. Finally, we have compared the obtained predictions to results of BD simulations of the full problem. While the FPE gives the exact value of the transition rate for vanishing delay, our results show that the obtained rates agree with the true ones for small and moderate values of the delay only. We conjecture that this is caused by the fact that the classical absorbing boundary used in our numerical and analytical evaluation of the transition rate can not be used for larger values of  $\tau$ . Concerning the analytical results, the best agreement with the true rates was obtained by the Bullerjahn's formula (64) with effective barrier height and diffusion coefficient taken from the time-local FPE (53) and the prefactor rescaled according to equation (66). In the case of the trimer, we have confirmed by BD simulations that the transition rate increases exponentially with the noise strength  $D$  even for longer delays and thus Kramers' or Bullerjahn's type predictions can be used also in this case. We plan to further investigate suitable absorbing boundary conditions for delayed systems to predict (at least numerically) transition rates also for large delays.

Finally in section 6, we have considered the robustness of our analytical results with respect to details of the realization of the delay. We demonstrated that most of the presented equations can be used also for systems with memory kernels different from that for discrete time delays, i.e.  $\phi(t) = \delta(t - \tau)$ . It is enough to substitute the Green's function  $\lambda(t)$  (A.5) corresponding the delayed Langevin equation by the Green's function corresponding to the memory kernel of interest. We reviewed some paradigmatic memory kernels and provided an outlook on the differences and similarities of the corresponding Green's functions. A more detailed study is left for future work.

As a further extension of our work, it would be interesting to consider physically more realistic interactions that vanish at large distances. Furthermore, we plan to investigate the reaction of the studied system to an external perturbation. Of particular interest could be the propagation and decay behavior of a local perturbation through the system, especially in case of large numbers of particles. Last but not least, we aim to investigate the behavior of the studied system under the action of an additional deterministic time-dependent driving and study the corresponding stochastic dynamics and thermodynamics.

## Acknowledgments

We acknowledge funding by Deutsche Forschungsgemeinschaft (DFG) via SPP 1726/1 and support from the German Research Foundation (DFG) and Universität Leipzig within the program of Open Access Publishing. VH gratefully acknowledges support by the Humboldt foundation and by the Czech Science Foundation (project No. 17-06716S). DG acknowledges funding by International Max Planck Research Schools (IMPRS). Furthermore, we thank Thomas Ihle and Sarah Loos for discussion.

## Appendix A. Solution of the Noiseless problem

In this appendix we solve the multi-dimensional linear delay differential equation (LDDE)

$$\dot{\mathbf{x}}(t) = -\omega \mathbf{x}(t - \tau), \quad (\text{A.1})$$

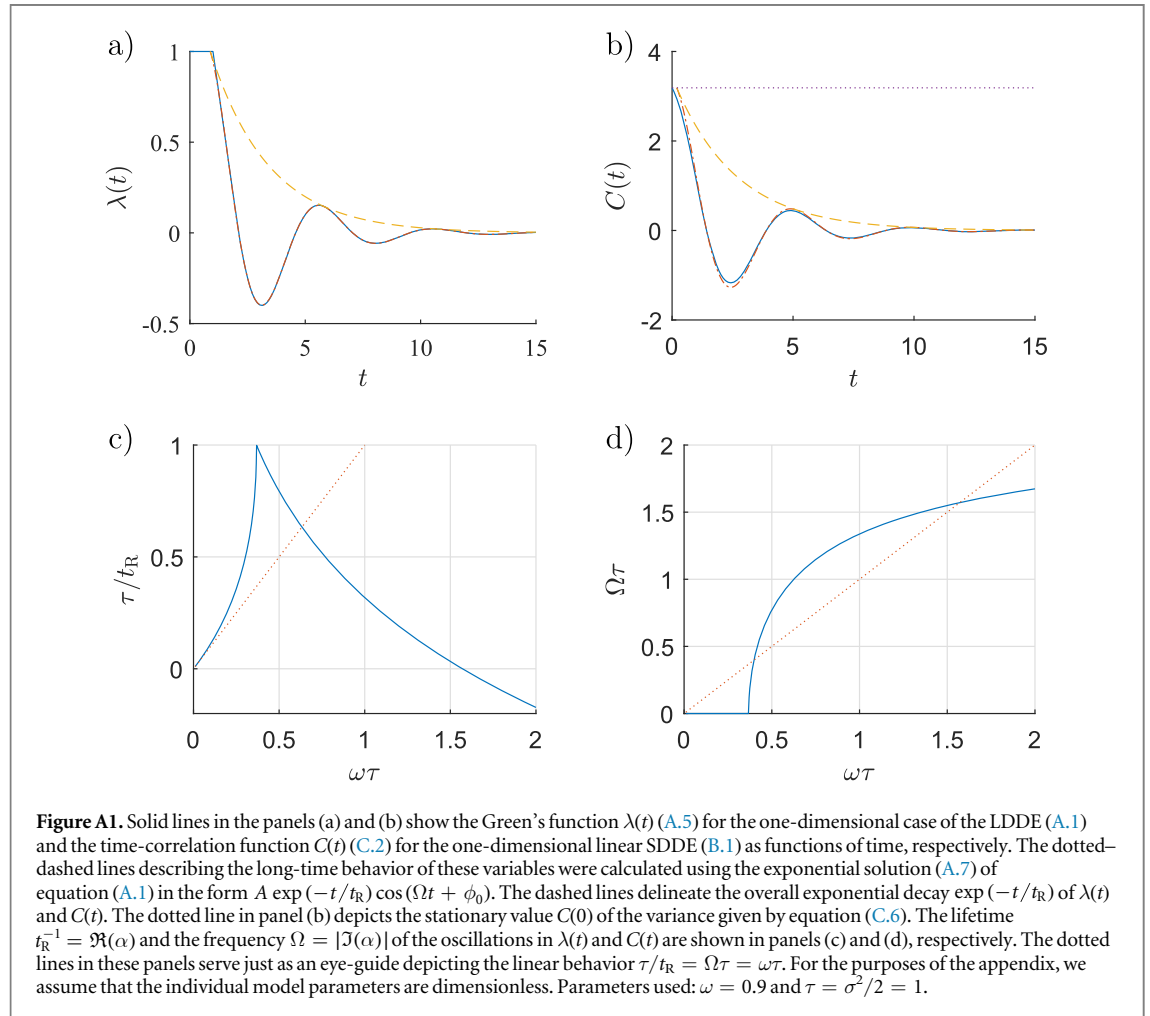
where  $\omega$  is a positive semi-definite matrix with real entries and  $\mathbf{x}(t)$  is a column vector. Laplace transformation of this equation leads to the formula

$$s\tilde{\mathbf{x}}(s) - \mathbf{x}_0 = -\omega e^{-s\tau} \left[ \tilde{\mathbf{x}}(s) + \int_{-\tau}^0 dt e^{-st} \mathbf{x}(t) \right] \quad (\text{A.2})$$

with  $\mathbf{x}_0 \equiv \mathbf{x}(0)$ . The solution of this equation for the Laplace transformed variable reads

$$\begin{aligned} \tilde{\mathbf{x}}(s) &= \int_0^\infty dt e^{-st} \mathbf{x}(t) = (s\mathcal{I} + \omega e^{-s\tau})^{-1} \left[ \mathbf{x}_0 - \omega e^{-s\tau} \int_{-\tau}^0 dt e^{-st} \mathbf{x}(t) \right] \\ &= \sum_{k=0}^{\infty} \frac{(-\omega)^k}{s^{k+1}} e^{-k s \tau} \left[ \mathbf{x}_0 - \omega e^{-s\tau} \int_{-\tau}^0 dt e^{-st} \mathbf{x}(t) \right], \end{aligned} \quad (\text{A.3})$$

where  $\mathcal{I}$  denotes the identity matrix. In the last step, we have expanded the inverse matrix using the Neumann series. The inverse Laplace transform of the ratio  $e^{-s\tau}/s^{k+1}$  is given by  $(t - \tau)^k \theta(t - \tau)/k!$  [78] and thus the



formula (A.3) can be inverted. Finally, the solution of equation (A.1) is given by

$$\mathbf{x}(t) = \lambda(t)\mathbf{x}_0 - \omega \int_{-\tau}^0 ds \lambda(t-s-\tau)\mathbf{x}(s), \quad (\text{A.4})$$

where

$$\lambda(t) = \sum_{k=0}^{\infty} \frac{(-\omega)^k}{k!} (t-k\tau)^k \theta(t-k\tau) \quad (\text{A.5})$$

is a matrix-valued function which solves equation (A.1) with the initial condition  $\mathbf{x}(t) = 0$  for all  $t < 0$  and  $\mathbf{x}(0) = \mathcal{I}$ . In the present paper, we always assume that the system is initialized at time  $t$  at position  $\mathbf{x}_0$  with a special history, namely  $\mathbf{x}(t) = 0$  for all  $t < 0$ , allowing us to simplify equation (A.4) to  $\mathbf{x}(t) = \lambda(t)\mathbf{x}_0$ .

The only fixed point of the DDE (A.1) is  $\mathbf{x}(t) = 0$ . In order to investigate its stability, it is useful to present an alternative solution of the LDDE (A.1) using the exponential ansatz  $\mathbf{x}(t) \propto \exp(-\alpha t)$ . Inserting this ansatz into equation (A.1) leads to the equation  $\alpha = \omega \exp(\alpha\tau)$  for the matrix  $\alpha$ . Except for some notable exceptions [79], the solution of this equation is given by

$$\alpha = -\frac{1}{\tau} W(-\tau\omega), \quad (\text{A.6})$$

where  $W$  denotes the matrix valued Lambert  $W$  function [80]. The Lambert  $W$  function is a multivalued complex function. The long-time behavior of solutions to equation (A.1) and thus also the stability of its fixed point are determined by the branch of  $W$  yielding the largest real parts of the eigenvalues of the matrix  $\alpha$ . The corresponding values of the Lambert  $W$  function strongly depend on the reduced delay  $\tau\omega$ .

For example, in one dimension, where  $\omega$  is a positive real number, the branch of the Lambert  $W$  function with the largest real part exhibits three qualitatively different regimes as a function of  $\tau\omega$  leading to three different dynamical regimes of solutions

$$x(t) \propto \Re[\exp(-\alpha t)] = \exp(-t/t_R) \cos(\Omega t), \quad (\text{A.7})$$

$1/t_R = \Re(\alpha)$ ,  $\Omega = |\Im(\alpha)|$ , to equation (A.1). The boundaries between these regimes can be determined analytically [25, 27, 30]: (i)  $0 < \tau\omega \leq 1/e$ , where  $\alpha$  is real and positive and  $x(t)$  decays exponentially to 0 with lifetime  $t_R$ ; (ii)  $1/e < \tau\omega < \pi/2$ , where  $\alpha$  is complex with a positive real part, producing exponentially damped oscillations of  $x(t)$  with frequency  $\Omega$  and lifetime  $t_R$ ; and (iii)  $\pi/2 < \tau\omega$ , where  $\alpha$  is complex with a negative real part, corresponding to exponentially diverging oscillations of  $x(t)$  with frequency  $\Omega$ . For  $\tau\omega = \pi/2$ ,  $1/t_R = 0$  and  $\lambda(t)$  oscillates with frequency  $\Omega$  without any decay.

In the panel (a) of figure A1, we show that for long times the Green's function  $\lambda(t)$  for equation (A.1) is well approximated by the exponential solution (A.7). The above described dynamical regimes are reflected in the behavior of the decay rate  $1/t_R$  and the frequency  $\Omega$  of oscillations, plotted in the panels (c) and (d), respectively. The panel (b) of the figure shows the steady auto correlation of  $x(t)$  calculated in appendix C, which is also well described by the formula (A.7).

## Appendix B. Solution with noise

### B.1. One dimension

Let us now solve equation (A.1) with an additional noise term. For simplicity, we present the detailed derivation first in one dimension. We thus want to solve the equation

$$\dot{x}(t) = -\omega x(t - \tau) + \sigma \eta(t), \quad (\text{B.1})$$

where  $\eta(t)$  is a white noise fulfilling  $\langle \eta(t) \rangle = 0$  and  $\langle \eta(t) \eta(t') \rangle = \delta(t - t')$ . Due to the time delay, this is a time-nonlocal (and consequently non-Markovian) Langevin equation, which, however, can be transformed into a time-local Langevin equation with a colored noise via the so-called time-convolutionless transform [35]. The time-local equation can then be used to derive the FPE for the PDFs for  $x(t)$ . In order to do that, we write the formal solution of equation (B.1) as

$$x(t) = x_0 \lambda(t) + \sigma \int_0^t ds \lambda(t-s) \eta(s) + \psi(t), \quad (\text{B.2})$$

where  $\lambda(t)$  is given by equation (A.5) and  $\psi(t) = -\omega \int_{-\tau}^0 ds \lambda(t-\tau-s) x(s)$  is determined by the initial condition  $x(t)$  for  $t < 0$ . As we show in the next section, the formula (B.2) can already be used for the calculation of the time-correlation function for  $x(t)$ . Here, we differentiate the solution (B.2) which yields the time-local Langevin equation

$$\dot{x}(t) = -\omega_\tau(t) x(t) + b(t) + \sigma \xi(t). \quad (\text{B.3})$$

Due to the time-nonlocal nature of equation (B.1), the potential in the time-local equation (B.3) possess the time-dependent stiffness

$$\omega_\tau(t) \equiv -\frac{\dot{\lambda}(t)}{\lambda(t)}, \quad (\text{B.4})$$

and the time-varying position of the minimum  $-b(t)/\omega_\tau(t)$ , with  $b(t) \equiv \omega_\tau(t) \psi(t) + \dot{\psi}(t)$  vanishing for the special initial condition  $x(t) = 0$  for all  $t < 0$ . Furthermore, equation (B.3) includes the Gaussian colored noise  $\xi(t) = \lambda(t) \frac{d}{dt} \int_0^t ds \frac{\lambda(t-s)}{\lambda(t)} \eta(s)$  which satisfies  $\langle \xi(t) \rangle = 0$  and

$$\langle \xi(t) \xi(t') \rangle = \lambda(t) \lambda(t') \frac{d}{dt'} \int_0^{t'} ds \frac{d}{dt} \frac{\lambda(t-s) \lambda(t'-s)}{\lambda(t) \lambda(t')}, \quad t' < t. \quad (\text{B.5})$$

While Markov processes are completely determined by the the transition probability density  $P_1(x, t|x_0, t_0)$  for going from the initial state  $x_0$  at time  $t_0$  to the final state  $x$  at time  $t$ , non-Markov processes in general require a full hierarchy of joint probability densities. Nevertheless, similarly to the Markovian case, the Gaussian non-Markov process (B.3) is completely determined by the joint probability distribution  $P_2(x, t; x', t'|x_0, 0)$  [35].

The FPEs for the one- and two-time probability distributions  $P_1(x, t|x_0, 0) = \langle \delta[x - x(t)] \rangle$  and  $P_2(x, t; x', t'|x_0, 0) = \langle \delta[x - x(t)] \delta[x' - x(t')] \rangle$ , where the averages are taken over all realizations of the process  $x(t)$  departing from state  $x_0$  at time 0, are found to be

$$\frac{\partial}{\partial t} P_1 = \frac{\partial}{\partial x} \left[ \omega_\tau(t) x - b(t) + 2D_\tau(t) \frac{\partial}{\partial x} \right] P_1, \quad (\text{B.6})$$

$$\frac{\partial}{\partial t} P_2 = \frac{\partial}{\partial x} \left[ \omega_\tau(t) x - b(t) + c(t, t') \frac{\partial}{\partial x} + 2D_\tau(t) \frac{\partial}{\partial x} \right] P_2. \quad (\text{B.7})$$

Similarly as the trap stiffness  $\omega_\tau(t)$ , also the effective diffusion coefficient, corresponding to the time-local description, is time dependent if  $\tau > 0$ . It reads

$$\begin{aligned}
2D_\tau(t) &\equiv \frac{\sigma^2 \lambda(t)^2}{2} \frac{d}{dt} \int_0^t ds \frac{\lambda(s)^2}{\lambda(t)^2} = \sigma^2 \left( \frac{\lambda(t)^2}{2} + \omega_\tau(t) \int_0^t ds \lambda(s)^2 \right) \\
&= \sigma^2 \lambda(t)^2 / 2 + \omega_\tau(t) \nu(t)
\end{aligned} \tag{B.8}$$

with the variance  $\nu(t) = \sigma^2 \int_0^t ds \lambda(s)^2$  and  $c(t, t') \equiv \sigma^2 \lambda(t) \frac{d}{dt} \int_0^{t'} ds \lambda(t-s) \lambda(t'-s) / \lambda(t)$ .

Because of the oscillatory nature of  $\lambda(t)$  in the dynamical regimes (ii) and (iii), the coefficients  $\omega_\tau$ ,  $b$ ,  $c$  and  $D_\tau$  in the FPEs (B.6) and (B.7) change their signs and they can even diverge. These divergences, however, always mutually balance each other such that the solutions of the FPEs, as given by the equations (19) and (20), are always reasonable [35].

## B.2. Higher dimensions

Let us now consider the problem

$$\dot{\mathbf{x}}(t) = -\omega \mathbf{x}(t - \tau) + \sigma \boldsymbol{\eta}(t), \tag{B.9}$$

with general matrices  $\omega$  and  $\sigma$  and the vector  $\boldsymbol{\eta}(t)$  of white noises fulfilling  $\langle \boldsymbol{\eta}(t) \rangle = 0$  and  $\langle \eta_i(t) \eta_j(t') \rangle = \delta_{ij} \delta(t - t')$ . Since this system of Langevin equations is linear, the one- and two-time probability distributions  $P_1(\mathbf{x}, t | \mathbf{x}_0, 0)$  and  $P_2(\mathbf{x}, t; \mathbf{x}', t' | \mathbf{x}_0, 0)$  for  $\mathbf{x}(t)$  defined in the preceding section must be Gaussian [63] as in the one-dimensional case and also the corresponding FPEs can be derived along similar lines as in one dimension. Instead of deriving these equations, we now provide a simpler alternative derivation of the properties of the Gaussian distribution

$$P_1(\mathbf{x}, t | \mathbf{x}_0, 0) = \frac{1}{\sqrt{(2\pi)^3 \det \mathcal{K}(t)}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}(t))^\top \cdot \mathcal{K}(t)^{-1} \cdot (\mathbf{x} - \boldsymbol{\mu}(t)) \right\} \tag{B.10}$$

based solely on the formal solution of the Langevin system (B.9)

$$\mathbf{x}(t) = \lambda(t) \mathbf{x}_0 + \int_0^t ds \lambda(t-s) \sigma \boldsymbol{\eta}(s) \tag{B.11}$$

with the initial condition  $\mathbf{x}(t) = 0$  for all  $t < 0$  and  $\mathbf{x}(0) = \mathbf{x}_0$  and with the Green's function  $\lambda(t)$  of the Langevin system given by equation (A.5).

The mean value  $\boldsymbol{\mu}(t) = \langle \mathbf{x}(t) \rangle$  and the elements  $\mathcal{K}_{ij}(t)$  of the covariance matrix  $\mathcal{K}(t) = \langle \mathbf{x}(t) \mathbf{x}^\top(t) \rangle - \langle \mathbf{x}(t) \rangle \langle \mathbf{x}^\top(t) \rangle$  defining the PDF (B.10) can be obtained by inserting  $\mathbf{x}_t(t)$  from equation (B.11) into the definitions and averaging over the noise  $\boldsymbol{\eta}(t)$ . The results are

$$\boldsymbol{\mu}(t) = \lambda(t) \mathbf{x}_0 \tag{B.12}$$

and

$$\mathcal{K}(t) = \int_0^t ds \lambda(t-s) \sigma \sigma^\top \lambda^\top(t-s). \tag{B.13}$$

These formulas can be generalized straightforwardly to arbitrary initial conditions, where  $\mathbf{x}(t)$  for  $t \leq 0$  is drawn from some probability distribution  $P[\mathbf{x}(t), t \leq 0]$ . Then the formal solution of the system (B.1) reads

$$\mathbf{x}(t) = \mathbf{y}(t) + \int_0^t ds \lambda(t-s) \sigma \boldsymbol{\eta}(s), \tag{B.14}$$

$$\mathbf{y}(t) = \lambda(t) \mathbf{x}(0) - \omega \int_{-\tau}^0 ds \lambda(t-s-\tau) \mathbf{x}(s). \tag{B.15}$$

The mean value  $\boldsymbol{\mu}(t)$  is given by  $\boldsymbol{\mu}(t) = \langle \mathbf{y}(t) \rangle_{\mathbf{x}(t \leq 0)}$ , and  $\mathcal{K}(t) = \langle \mathbf{y}(t) [\mathbf{y}(t)]^\top \rangle_{\mathbf{x}(t \leq 0)} + \int_0^t ds \lambda(t-s) \sigma \sigma^\top \lambda^\top(t-s)$ . The averages  $\langle \cdot \rangle_{\mathbf{x}(t \leq 0)}$  above are taken with respect to the PDF  $P[\mathbf{x}(t), t \leq 0]$ . The long-time behavior of the covariance matrix (B.13) is studied in the next section of this appendix.

## Appendix C. Time-correlation matrix and stationary covariance matrix

The coefficients in the Gaussian two-time PDF  $P_2(\mathbf{x}, t; \mathbf{x}', t' | \mathbf{x}_0, 0)$  can be obtained in a similar manner as in appendix B.2. Here, we calculate only the stationary space-time correlation matrix  $\mathcal{C}(t) = \lim_{s \rightarrow \infty} [\langle \mathbf{x}(s+t) \mathbf{x}^\top(s) \rangle - \langle \mathbf{x}(s+t) \rangle \langle \mathbf{x}^\top(s) \rangle] = \lim_{s \rightarrow \infty} \langle \mathbf{x}(s+t) \mathbf{x}^\top(s) \rangle$  which exists only if  $\lim_{s \rightarrow \infty} \langle \mathbf{y}(s) \rangle = \lim_{s \rightarrow \infty} \langle \mathbf{x}(s) \rangle = 0$ . Its matrix elements can be calculated in an analogous fashion as the elements of  $\mathcal{K}(t)$ . The result

$$\mathcal{C}(t) = \lim_{s \rightarrow \infty} \int_0^s ds' \lambda(s+t-s') \sigma \sigma^\top \lambda^\top(s-s') \tag{C.1}$$

can be evaluated numerically. It is possible to rewrite it in a simpler form. Taking the derivative of  $\mathcal{C}(t)$  with respect to  $t$  reveals that for  $t > 0$  the correlation matrix obeys the same DDE as the Green's function  $\lambda(t)$ , i.e.



$$\dot{\mathcal{C}}(t) = -\omega\mathcal{C}(t - \tau), \quad t > 0. \quad (\text{C.2})$$

The restriction  $t > 0$  for validity of this equation comes from discontinuity (and thus non-differentiability) of  $\lambda(t)$  at  $t = 0$ . The solution to equation (C.2) is given by equation (A.4) and hence the stationary space-time correlation matrix can be written as

$$\mathcal{C}(t) = \lambda(t)\mathcal{C}_0 - \omega \int_{-\tau}^0 ds \lambda(t - s - \tau)\mathcal{C}(s), \quad (\text{C.3})$$

where  $\mathcal{C}_0 = \mathcal{C}(0)\lim_{t \rightarrow \infty} \mathcal{K}(t)$  is given by the long time limit of the covariance matrix. The stationary correlation matrix is thus solely determined by the unknown initial condition  $\mathcal{C}(t)$ ,  $t \in [-\tau, 0]$ . Fortunately, this initial condition can be calculated using the approach of Frank *et al* [31] who calculated the time-correlation function  $\mathcal{C}(t)$  for  $t \in [0, \tau]$  in one-dimension (see also [32]). They utilized the symmetry  $\mathcal{C}(t) = \mathcal{C}(-t)$  following from equation (C.1) to rewrite the DDE (C.2) as  $\dot{\mathcal{C}}(t) = -\omega\mathcal{C}(\tau - t)$  for  $t \in (0, \tau)$ . Taking the derivative of this equation and using equation (C.2) yields the second order ordinary differential equation

$$\ddot{\mathcal{C}}(t) = -\omega^2\mathcal{C}(t), \quad t \in (0, \tau) \quad (\text{C.4})$$

for the initial condition  $\mathcal{C}(t) = \mathcal{C}(-t)$ ,  $t \in [-\tau, 0]$ . The solution of this equation reads  $\mathcal{C}(t) = \mathcal{C}_0 \cos(\omega t) + \dot{\mathcal{C}}_0 \omega^{-1} \sin(\omega t)$ , where we still need to determine the unknown coefficients  $\mathcal{C}_0 = \mathcal{C}(0)$  and  $\dot{\mathcal{C}}_0 = \lim_{t \rightarrow 0^+} \dot{\mathcal{C}}(t)$ . To this end, we need to evaluate independently  $\mathcal{C}(t)$  and/or  $\dot{\mathcal{C}}(t)$  for two times  $t \in (0, \tau)$ . Specifically, we show at the end of this appendix that  $\mathcal{C}(\tau) = 0.5\omega^{-1}\sigma\sigma^\top$  and  $\dot{\mathcal{C}}_0 = -0.5\sigma\sigma^\top$ . Using these results, the final expression for the correlation matrix for  $t \in [-\tau, \tau]$  reads

$$\mathcal{C}(t) = \mathcal{C}_0 \cos(\omega t) - 0.5\sigma\sigma^\top \omega^{-1} \sin(\omega|t|) \quad (\text{C.5})$$

with the initial value

$$\mathcal{C}(0) = \mathcal{C}_0 = \lim_{s \rightarrow \infty} \mathcal{K}(s) = \frac{1}{2}[\omega^{-1}\sigma\sigma^\top + \sigma\sigma^\top \omega^{-1} \sin(\omega\tau)] \cos^{-1}(\omega\tau) \quad (\text{C.6})$$

given by the stationary value of the covariance matrix. The whole time-dependence of  $\mathcal{C}(t)$  for  $t \geq 0$  is thus described by the formulas (C.3), (C.5), and (C.6). The correlation matrix for negative times then follows from the symmetry  $\mathcal{C}(t) = \mathcal{C}(-t)$ . Finally, let us note that in the one-dimensional case, where  $\omega$  and  $\sigma$  stand for real numbers and, more generally, if the matrices  $\omega^{-1}$  and  $\sigma\sigma^\top$  commute, we can rewrite equation (C.6) as

$$\mathcal{C}_0 = \frac{\sigma\sigma^\top}{2\omega} [\mathcal{I} + \sin(\omega\tau)] \cos^{-1}(\omega\tau), \quad (\text{C.7})$$

where  $\mathcal{I}$  denotes the identity matrix. An example of the time-correlation function for a one dimensional system is depicted in figure A1(b) in appendix A.

The expression for  $\mathcal{C}(\tau)$  can be obtained by multiplying equation (B.9) by  $\mathbf{x}^\top(s)$ , averaging the result over the noise, using the assumed stationarity of the process implying the formula  $\dot{\mathcal{C}}(0) = d[\lim_{s \rightarrow \infty} \langle \mathbf{x}(s)\mathbf{x}^\top(s) \rangle] / ds = 0$ , and applying the symmetry  $\mathcal{C}(t) = \mathcal{C}(-t)$ . The result is  $\mathcal{C}(\tau) = \omega^{-1}\sigma \langle \boldsymbol{\eta}(s)\mathbf{x}^\top(s) \rangle = 0.5\omega^{-1}\sigma\sigma^\top$ , where the last equality comes after inserting the formal solution (B.11) for  $\mathbf{x}(t)$  into the average, using the covariance of the noise, and noticing that in the resulting integral we integrate over half of the emerging  $\delta$ -function only. The expression for  $\dot{\mathcal{C}}_0$  then comes simply from equation (C.2), which is invalid for  $t = 0$ , but can be used for  $t$  arbitrarily close to 0 from the right, and using the result for  $\mathcal{C}(\tau)$ .

## References

- [1] Cavagna A, Cimarelli A, Giardina I, Parisi G, Santagati R, Stefanini F and Viale M 2010 *Proc. Natl Acad. Sci.* **107** 11865–70
- [2] Ben-Jacob E, Schochet O, Tenenbaum A, Cohen I, Czirok A and Vicsek T 1994 *Nature* **368** 46
- [3] Zhang H P, Be'er A, Florin E L and Swinney H L 2010 *Proc. Natl Acad. Sci.* **107** 13626–30
- [4] Elgeti J, Winkler R G and Gompper G 2015 *Rep. Prog. Phys.* **78** 056601
- [5] Ramaswamy S 2010 *Annu. Rev. Condens. Matter Phys.* **1** 323–45
- [6] Vicsek T and Zafeiris A 2012 *Phys. Rep.* **517** 71–140
- [7] Vicsek T, Czirók A, Ben-Jacob E, Cohen I and Shochet O 1995 *Phys. Rev. Lett.* **75** 1226
- [8] Attanasi A *et al* 2014 *Nat. Phys.* **10** 691
- [9] Piwowarczyk R, Selin M, Ihle T and Volpe G 2019 Influence of sensorial delay on clustering and swarming *Phys. Rev. E* **100** 012607
- [10] Mijalkov M, McDaniel A, Wehr J and Volpe G 2016 *Phys. Rev. X* **6** 011008
- [11] Khadka U, Holubec V, Yang H and Cichos F 2018 *Nat. Commun.* **9** 3864
- [12] Brambilla M, Ferrante E, Birattari M and Dorigo M 2013 *Swarm Intell.* **7** 1–41
- [13] Bäuerle T, Fischer A, Speck T and Bechinger C 2018 *Nat. Commun.* **9** 3232
- [14] Gibbs H M, Hopf F A, Kaplan D L and Shoemaker R L 1981 *Phys. Rev. Lett.* **46** 474
- [15] Arecchi F T, Giacomelli G, Lapucci A and Meucci R 1992 *Phys. Rev. A* **45** R4225
- [16] Masoller C 2002 *Phys. Rev. Lett.* **88** 034102
- [17] Kanter I, Aviad Y, Reidler I, Cohen E and Rosenbluh M 2010 *Nat. Photon.* **4** 58
- [18] Baraban L, Streubel R, Makarov D, Han L, Karnaushenko D, Schmidt O G and Cuniberti G 2013 *ACS Nano* **7** 1360–7
- [19] Qian B, Montiel D, Bregulla A, Cichos F and Yang H 2013 *Chem. Sci.* **4** 1420–9
- [20] Walther A and Müller A H E 2008 *Soft. Matter.* **4** 663–8

- [21] Jiang H R, Yoshinaga N and Sano M 2010 *Phys. Rev. Lett.* **105** 268302
- [22] Kramers H A 1940 *Physica* **7** 284–304
- [23] Hänggi P, Talkner P and Borkovec M 1990 *Rev. Mod. Phys.* **62** 251
- [24] Bullerjahn J T, Sturm S and Kroy K 2014 *Nat. Commun.* **5** 4463
- [25] Bellman R and Cooke K L 1963 *Delay Differential Equations* vol 6 (New York: Academic)
- [26] Atay F M 2010 *Complex Time-delay Systems: Theory and Applications* (Berlin: Springer)
- [27] Michiels W and Niculescu S I 2007 *Stability and Stabilization of Time-delay Systems: An Eigenvalue-based Approach* (Philadelphia, PA: SIAM)
- [28] Foss J, Longtin A, Mensour B and Milton J 1996 *Phys. Rev. Lett.* **76** 708
- [29] Longtin A 2009 *Stochastic delay-differential equations Complex Time-delay Systems* (Berlin: Springer) pp 177–95
- [30] Guillouzic S, L'Heureux I and Longtin A 1999 *Phys. Rev. E* **59** 3970
- [31] Frank T D, Beek P J and Friedrich R 2003 *Phys. Rev. E* **68** 021912
- [32] Loos S and Klapp S H L 2017 *Phys. Rev. E* **96** 012106
- [33] Frank T D 2005 *Phys. Rev. E* **71** 031106
- [34] Küchler U 1992 *Stoch. Int. J. Probab. Stoch. Process.* **40** 23
- [35] Giuggioli L, McKetterick T J, Kenkre V and Chase M 2016 *J. Phys. A: Math. Theor.* **49** 384002
- [36] Adelman S A 1976 *J. Chem. Phys.* **64** 124–30
- [37] Fox R F 1977 *J. Math. Phys.* **18** 2331–5
- [38] Hänggi P 1978 *Z. Phys. B* **31** 407–16
- [39] Sancho J M, San Miguel M, Katz S L and Gunton J D 1982 *Phys. Rev. A* **26** 1589
- [40] Hernández-Machado A, Sancho J M, San Miguel M and Pesquera L 1983 *Z. Phys. B* **52** 335–43
- [41] Gopalsamy K 2013 *Stability and Oscillations in Delay Differential Equations of Population Dynamics* vol 74 (Berlin: Springer)
- [42] Mao X, Yuan C and Zou J 2005 *J. Math. Anal. Appl.* **304** 296–320
- [43] Voss H U and Kurths J 2002 *Analysis of economic delayed-feedback dynamics Modelling and Forecasting Financial Data* (Berlin: Springer) pp 327–49
- [44] Stoica G 2005 *Proc. Am. Math. Soc.* **133** 1837–41
- [45] Mackey M C 1989 *J. Econ. Theory* **48** 497–509
- [46] Gao Q and Ma J 2009 *Nonlinear Dyn.* **58** 209
- [47] Kyrychko Y N and Hogan S J 2010 *J. Vib. Control* **16** 943–60
- [48] Beuter A, Bélair J and Labrie C 1993 *Bull. Math. Biol.* **55** 525–41
- [49] Chen Y, Ding M and Kelso J A S 1997 *Phys. Rev. Lett.* **79** 4501
- [50] Novák B and Tyson J J 2008 *Nat. Rev. Mol. Cell Biol.* **9** 981
- [51] Haken H 2007 *Brain Dynamics: An Introduction to Models and Simulations* (Berlin: Springer)
- [52] Marcus C M and Westervelt R M 1989 *Phys. Rev. A* **39** 347
- [53] Sompolinsky H, Golomb D and Kleinfeld D 1991 *Phys. Rev. A* **43** 6990
- [54] Rosinberg M L, Munakata T and Tarjus G 2015 *Phys. Rev. E* **91** 042114
- [55] Van Vu T and Hasegawa Y 2018 arXiv:1809.06610
- [56] Loos S A and Klapp S H 2019 *Sci. Rep.* **9** 2491
- [57] Seifert U 2012 *Rep. Prog. Phys.* **75** 126001
- [58] Sekimoto K 2010 *Stochastic Energetics* vol 799 (Berlin: Springer)
- [59] Kubo R 1966 *Rep. Prog. Phys.* **29** 255–84
- [60] Stephens M 1963 *Biometrika* **50** 385–90
- [61] Peruani F and Morelli L G 2007 *Phys. Rev. Lett.* **99** 010602
- [62] Selmke M, Khadka U, Bregulla A P, Cichos F and Yang H 2018 *Phys. Chem. Chem. Phys.* **20** 10502–20
- [63] Risken H and Frank T 1996 *Fokker–Planck Equation: Methods of Solution and Applications (Springer Series in Synergetics)* 2nd edn (Berlin: Springer)
- [64] Fuchs J, Goldt S and Seifert U 2016 *Europhys. Lett.* **113** 60009
- [65] Frank T 2016 *Phys. Lett. A* **380** 1341–51
- [66] Seifert U 2005 *Phys. Rev. Lett.* **95** 040602
- [67] Grote R F and Hynes J T 1980 *J. Chem. Phys.* **73** 2715–32
- [68] Pollak E 1990 *J. Chem. Phys.* **93** 1116–24
- [69] Grabert H and Linkwitz S 1988 *Phys. Rev. A* **37** 963
- [70] Guillouzic S, L'Heureux I and Longtin A 2000 *Phys. Rev. E* **61** 4906
- [71] Curtin D, Hegarty S P, Goulding D, Houlihan J, Busch T, Masoller C and Huyet G 2004 *Phys. Rev. E* **70** 031103
- [72] Ornigotti L, Ryabov A, Holubec V and Filip R 2018 *Phys. Rev. E* **97** 032127
- [73] Holubec V, Kroy K and Steffenoni S 2019 *Phys. Rev. E* **99** 032117
- [74] Šiler M, Ornigotti L, Brzobohatý O, Jákľ P, Ryabov A, Holubec V, Zemánek P and Filip R 2018 *Phys. Rev. Lett.* **121** 230601
- [75] Bruwier L 1930 Sur l'équation fonctionnelle  $y^{(n)}(x) + a_1 y^{(n-1)}(x + c) + \dots + a_{n-1} y'(x + \overline{n-1}c) + a_n y(x + nc) = 0$ . *C. R. Congrès Natl Sci.* **1930** 91–7
- [76] Perron O 1939 *Math. Z.* **45** 127–41
- [77] Chase M, McKetterick T J, Giuggioli L and Kenkre V 2016 *Eur. Phys. J. B* **89** 87
- [78] Olver F W J, Lozier D W, Boisvert R F and Clark C W 2010 *NIST Handbook of Mathematical Functions Hardback and CD-ROM* (Cambridge: Cambridge University Press)
- [79] Corless R M, Ding H, Higham N J and Jeffrey D J 2007 The solution of  $\text{sexp}(s) = a$  is not always the Lambert  $w$  function of  $a$ . *Proc. 2007 Int. Symp. Symbolic and Algebraic Computation, ISSAC '07 (New York, NY, USA)* (ACM) pp 116–21
- [80] Yi S, Nelson P W and Ulsoy A G 2010 *Time-Delay Systems: Analysis and Control Using the Lambert W Function* (Singapore: World Scientific)
- [81] Bouchet F, Gupta S and Mukamel D 2010 *Phys. A: Stat. Mech. Appl.* **389** 4389–405



# Spontaneous vortex formation by microswimmers with retarded attractions

Received: 2 August 2022

Accepted: 2 December 2022

Published online: 04 January 2023



Xiangzun Wang<sup>1</sup>, Pin-Chuan Chen<sup>2</sup>, Klaus Kroy<sup>2</sup>, Viktor Holubec<sup>3</sup> & Frank Cichos<sup>1</sup> ✉

Collective states of inanimate particles self-assemble through physical interactions and thermal motion. Despite some phenomenological resemblance, including signatures of criticality, the autonomous dynamics that binds motile agents into flocks, herds, or swarms allows for much richer behavior. Low-dimensional models have hinted at the crucial role played in this respect by perceived information, decision-making, and feedback, implying that the corresponding interactions are inevitably retarded. Here we present experiments on spherical Brownian microswimmers with delayed self-propulsion toward a spatially fixed target. We observe a spontaneous symmetry breaking to a transiently chiral dynamical state and concomitant critical behavior that do not rely on many-particle cooperativity. By comparison with the stochastic delay differential equation of motion of a single swimmer, we pinpoint the delay-induced effective synchronization of the swimmers with their own past as the key mechanism. Increasing numbers of swimmers self-organize into layers with pro- and retrograde orbital motion, synchronized and stabilized by steric, phoretic, and hydrodynamic interactions. Our results demonstrate how even most simple retarded interactions can foster emergent complex adaptive behavior in small active-particle ensembles.

Ordered dynamical phases of motile organisms are ubiquitous in nature across all scales<sup>1</sup>, from bacterial colonies to insect swarms, and bird flocks<sup>2</sup>. In particular, self-organization into vortex patterns is often observed and has been attributed to some local external attractor, e.g., light or nutrient concentration, together with behavioral rules like collision avoidance and mutual alignment<sup>3</sup>. The pertinent social interactions are commonly thought to be based on perception<sup>4–6</sup> and the ability to actively control the direction of motion<sup>3</sup>. They are also generally presumed to provide some benefits to the individual and to the collective, as in the case of collision avoidance or predator evasion<sup>7,8</sup>. However, since such interactions are usually derived only indirectly and approximately from observations<sup>9</sup>, it is arguably useful to coarse grain them, e.g., into simple alignment rules, in order to rationalize the collective effects with the help of simple mechanistic models, in particular with respect to their emerging universal

traits<sup>3,10–12</sup>. This strategy has been successful in physics and is also supported by the observation that biological collectives often appear highly susceptible to environmental influences and exhibit a dynamical finite-size scaling reminiscent of critical states in inanimate many-body assemblies<sup>13–16</sup>.

Importantly, the cascades of complex biochemical/biophysical processes<sup>17,18</sup> needed to transform signal perception into a navigational reaction inevitably result in retarded interactions upon coarse-graining<sup>19</sup> (cf. supplementary Table S1). This generic complication is often dismissed in the analysis, and dedicated models and experiments addressing the role of time delays in the active matter are still rare<sup>20–23</sup>, although these have occasionally been shown to fundamentally alter the collective dynamics<sup>21</sup> and to bring it closer to that found in nature<sup>24</sup>. To a first approximation, delay effects can resemble inertial corrections to an otherwise overdamped biological dynamics<sup>25</sup>. In particular,

<sup>1</sup>Peter Debye Institute for Soft Matter Physics, Leipzig University, 04103 Leipzig, Germany. <sup>2</sup>Institute for Theoretical Physics, Leipzig University, Postfach 100 920, 04009 Leipzig, Germany. <sup>3</sup>Department of Macromolecular Physics, Faculty of Mathematics and Physics, Charles University, 18000 Prague, Czech Republic. ✉ e-mail: [cichos@physik.uni-leipzig.de](mailto:cichos@physik.uni-leipzig.de)

both have the propensity to give rise to oscillations and inertia, moreover, to rotational motion around an attractive center, as familiar from planetary orbits.

Experiments that can assess or even deliberately control retarded interactions in living systems turn out to be difficult. But by imposing time delays onto synthetic active particles via computer-controlled laser activation we can create an ideal laboratory system to experimentally emulate such situations. Suitable feedback control techniques for active particles have recently become available through photon nudging<sup>26</sup>. The technique allows to adjust a particle's propulsion speed to acquire real-time information (positions and directions of motion) about the dynamical state of an ensemble. It has previously been employed to rectify the rotational Brownian motion for particle steering and trapping<sup>27</sup>, to explore orientation-density patterns in activity landscapes<sup>28</sup>, and to study information flow between active particles<sup>23</sup>, and their emerging critical states<sup>29,30</sup>. Beyond what related computer simulations accomplish<sup>31–33</sup>, these experiments additionally incorporate the full real-world complexity arising from actual physical interactions due to hydrodynamic, thermal, or concentration fields. In the following, we describe experiments with feedback-controlled active Brownian microswimmers aiming at a fixed target by a retarded thermophoretic self-propulsion. The systematic navigational errors resulting from the retardation are seen to cause a spontaneous symmetry breaking to a bi-stable dynamical state, in which the swimmers self-organize into a merry-go-round motion that switches transiently between degenerate chiralities.

## Results

### Single-particle retarded interaction

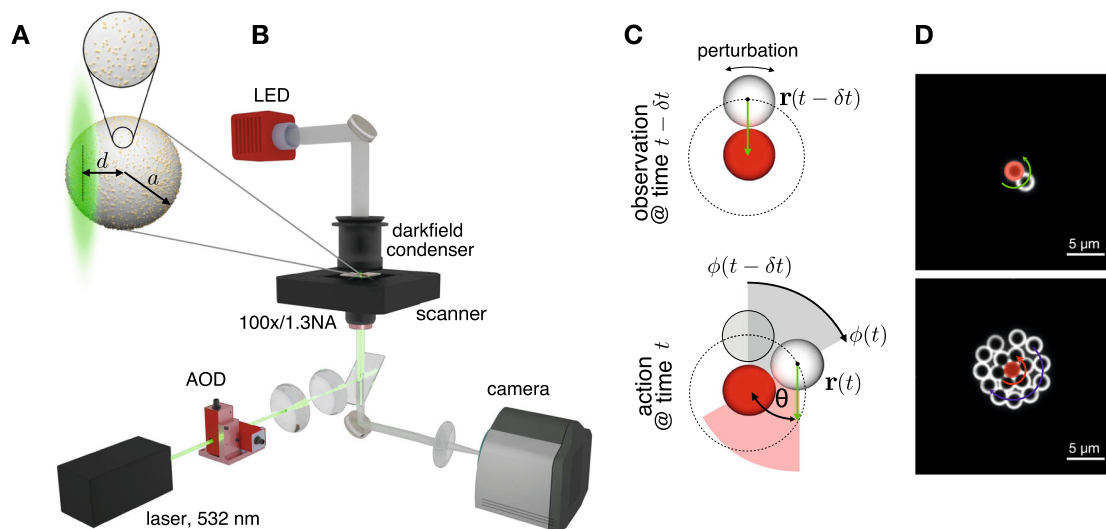
The elementary component of a swarm is a single active particle whose direction of motion depends dynamically on its environment. Even small fluctuations of the particle position and orientation render any prospective active motion based on the perception of the environment inaccurate, due to the inevitable finite perception–action delay. In the most symmetric setup, an active particle moves toward a target

position, which is occupied by an immobile particle of the same size in our experiments. Assuming that the active particle responds to the environment that was perceived a delay time  $\delta t$  earlier, its propulsion direction  $\hat{\mathbf{u}}(t)$  at time  $t$  is determined by its relative position to the target particle at time  $t - \delta t$  in the past, according to

$$\hat{\mathbf{u}}(t) = \frac{-\mathbf{r}(t - \delta t)}{|\mathbf{r}(t - \delta t)|}, \quad (1)$$

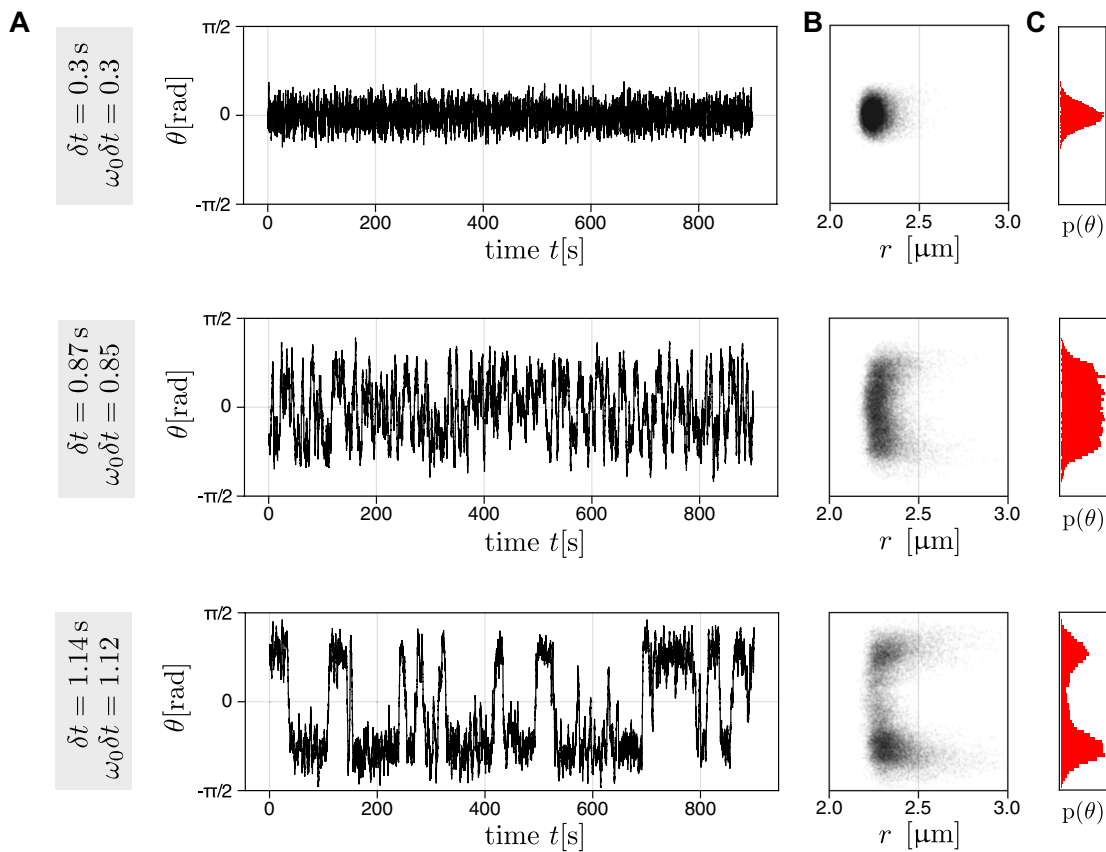
where  $\mathbf{r}$  is the location of the active particle with respect to the target particle's center. We implemented this interaction rule in an experimental feedback system that controls the active particles' self-propulsion. Our active particles are polymer spheres of radius  $a = 1.09 \mu\text{m}$ , decorated with gold nanoparticles and suspended in a thin film of water. Laser light with a wavelength of 532 nm is focused at distance  $d$  from the active particle center (Fig. 1A). The resulting excentric heating excites an osmotic flow that lets the particle swim with a speed  $v_0$  in the direction defined by Eq. (1)<sup>34</sup>. A darkfield microscopy setup is used to image the particles (Fig. 1B). A computer analyzes and records the positions of the particles and then controls the laser position accordingly via an acousto-optic deflector. We use a separate calibrator particle running on a quadratic trajectory as a reference for the speed  $v_0$  attained by a free swimmer. Further details are described in Sec. 2 of the Supplementary Information.

If  $\delta t = 0$  s, the active particle moves towards the target particle until it collides with it. Further motion of the active particle is then constrained by the presence of the fixed target sphere, resulting in a diffusive motion around it, at a fluctuating distance consistent with the barometer formula<sup>35,36</sup>. As the delay  $\delta t$  increases, the diffusive motion induces a stochastic “error” component due to the increasingly misaligned self-propulsion. Once a critical delay is reached, the particle begins to orbit around the target (see Supplementary Movies 1–3). We quantify this dynamics by the angle  $\theta$  between the direction of motion in Eq. (1) and the instantaneous negative radial direction  $-\mathbf{r}(t)$  (see Fig. 2A). The angle  $\theta$  itself or  $\sin(\theta)$  can serve as an indicator for



**Fig. 1 | Experimental realization.** **A** Particles used in the experiments consist of a melamine resin colloid ( $2.18 \mu\text{m}$  in diameter) with 8 nm gold nanoparticles scattered across the surface (covering up to 10% of the total surface area). A 532 nm laser focused at the edge of the particle at a distance  $d$  from its center induces a self-thermophoretic motion and allows for precise control of the propulsion direction. Importantly, optical forces are weak so the particles exhibit a truly self-phoretic autonomous motility, making them proper microswimmers. **B** Experimental setup used to image the particles by darkfield microscopy (LED, darkfield condenser, and camera) and guide their motion by sequential beam steering of the laser on the

sample plane with a two-axis acousto-optic deflector (AOD). All particles in the field of view are addressed during each exposure period of the camera. **C** The interaction rule for the delayed attraction of a single active particle (white sphere) towards a target (red sphere) is split into an observation made at a time  $t - \delta t$  that sets the direction of motion for the self-propulsion step exerted after a programmed delay time  $\delta t$ . The green arrows indicate the planned motion  $-\mathbf{r}(t - \delta t)$  and its actual realization at time  $t$ . **D** Examples of darkfield microscopy images where a single active particle (top) and 16 active particles (bottom) interact with one target particle (red).



**Fig. 2 | Propulsion angle at the different programmed delay.** **A** Trajectories of the propulsion angle  $\theta(t)$  of an active particle at three different delays (top:  $\delta t = 0.3$  s, middle:  $\delta t = 0.87$  s, and bottom:  $\delta t = 1.14$  s) for its attraction towards a target particle. The velocity of the active particle is  $v_0 = 2.16 \mu\text{m s}^{-1}$ . **B** Propulsion angle  $\theta(t)$

vs. the distance  $|\mathbf{r}(t)|$  of the particle from the target center. **C** Histograms of the propulsion angle over the whole trajectory. The delay for the individual panels in columns (**B**, **C**) is indicated on the left of the corresponding row.

deviations from the “intended” central orientation. Similarly, for many particles, numbered by the index  $i$ , it is useful to define the rotational order parameters  $o_{R_i} = (\hat{\mathbf{r}}_i \times \hat{\mathbf{u}}_i) \cdot \mathbf{e}_z = \sin(\theta_i)^{29,37}$ , where the hats denote vectors normalized to 1 and  $\mathbf{e}_z$  is a unit vector in the direction of  $z$  axis. Figure 2A shows the experimental trajectories of  $\theta$  for a single active particle with  $v_0 = 2.16 \mu\text{m s}^{-1}$  and three different delays. For short delays,  $\theta$  fluctuates with a small amplitude around zero (Fig. 2A top). The fluctuations increase with the delay and lead to a flat-top probability density of the propulsion angle for  $\delta t \approx 0.87$  s (Fig. 2A middle). At larger delays ( $\delta t = 1.14$  s), the propulsion angle fluctuates around a stable nonzero value that changes its sign intermittently (Fig. 2A bottom), corresponding to a bimodal probability density  $p(\theta)$  (Fig. 2C). The periods of consistent chirality increase in duration when the delay is increased further. At  $\delta t = 1.4$  s, the propulsion angle transiently fluctuates around  $\pm 80^\circ$ . Under these conditions, the cohesion of the particle to the target becomes marginal as the typical particle velocity is almost tangential to the target particle circumference. As a result, the distance  $|\mathbf{r}(t)|$  of the particle from the origin starts to fluctuate more strongly, as shown in the position histograms in Fig. 2B.

The net propulsion angle is the result of angular displacements  $\phi(t)$  of the particle position acquired due to the perception–action delay during the period  $[t - \delta t, t]$ :

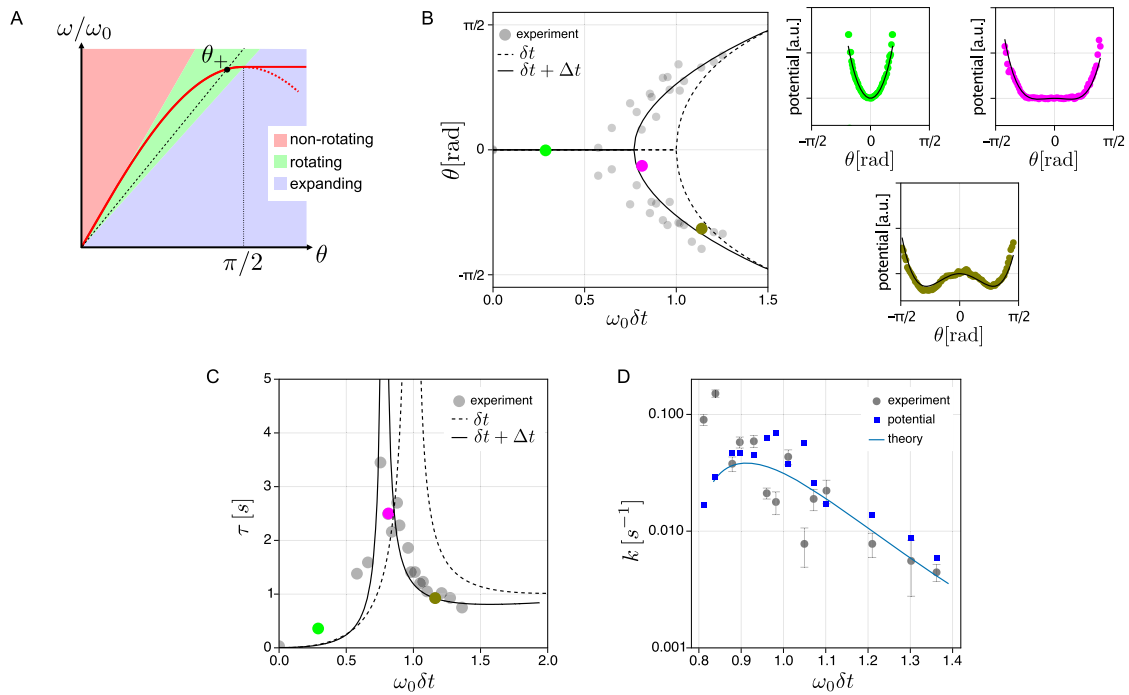
$$\theta(t) = \int_{t-\delta t}^t \omega(t') dt' = \phi(t) - \phi(t - \delta t) = \angle(\hat{\mathbf{u}}(t), -\mathbf{r}(t)). \quad (2)$$

Here,  $\phi(t)$  is the polar angle of the active particle in polar coordinates centered in the target particle, and we introduced  $\omega(t) = \dot{\phi}(t)$  as its corresponding angular velocity (Fig. 2C). The

observed dynamics can be understood by considering the active particle and the target particle in physical contact. Their distance is then constrained to be the sum of their radii ( $R = 2a = \langle |\mathbf{r}(t)| \rangle$ ) and the active particle slides around the target particle with an angular velocity  $\omega(t) = \omega_0 \sin(\theta(t))$ , where  $\omega_0 = v_0/R$  is the natural angular velocity for tangential propulsion with  $\theta = \pm\pi/2$ . As sketched in Fig. 3A, assuming a constant angular velocity  $\omega$  with  $\theta = \omega\delta t$ , the solutions to the equation for  $\theta$  are given by the intersections of a sine function and a linear function,

$$(\omega_0\delta t)^{-1}\theta = \sin(\theta). \quad (3)$$

For  $\omega_0\delta t < 1$ , there is a single intersection at  $\theta = 0$ , indicating a stable non-rotational state. For  $1 < \omega_0\delta t < \pi/2$ , the non-rotational state becomes unstable and two counter-rotational metastable solutions arise. For  $\omega_0\delta t > \pi/2$ , the rotating solutions correspond to  $|\theta| > \pi/2$ , and the radial component of propulsion becomes positive (repulsive), driving the active particle away from the target particle. As a result, the orbit “takes off” and its radius  $R$  increases until a new stable orbit with  $R = 2v_0\delta t/\pi > 2a$  and  $|\theta| = \pi/2$  is reached. For small particles ( $a \rightarrow 0$ ), the distance of the swimmer to the target position can thus, in principle, vanish ( $R \rightarrow 0$ ), and the rotating orbits can even occur at arbitrarily short programmed delays ( $\delta t \rightarrow 0$ ). Retarded attraction hence always leads to rotational orbital motion with a delay-dependent radius<sup>23</sup>. In the experiment, due to the presence of the fixed central particle, the smallest attainable orbit radius  $R = 2a$  is given by the particle diameter. Adding Brownian fluctuations to the deterministic Eq. (3) results in the nonlinear delayed stochastic differential equation  $\dot{\phi}(t) = \omega_0 \sin(\phi(t) - \phi(t - \delta t)) + \sqrt{2D_0/R^2} \eta(t)$ ,



**Fig. 3 | Transition to a rotational dynamical state for a single active particle.**  
**A** Graphical construction of condition (3) for a transition from a non-rotational state (red-shaded region) to a rotational state (green-shaded region). The red line ( $\sin \theta$ ) and the black dashed line with slope  $1/(\omega_0 \delta t)$  intersect at several  $\theta$ . The solution  $\theta = \theta_+$  in the green region and its chirally inverse image  $\theta_-$  in the third quadrant (not shown) correspond to co- and counter-clockwise rotation.  
**B** Experimentally measured propulsion angles (maxima of the histograms in Fig. 2C) as a function of  $\omega_0 \delta t$ , exhibiting a bifurcation at  $\omega_0 \delta t \approx 0.76$ . The dashed line corresponds to the analytical prediction of the theoretical model (5), neglecting the inevitable instrumental delay  $\Delta t$ . The solid line shows the solution of the refined theoretical model, which includes the instrumental delay  $\Delta t = 64$  ms of our setup in addition to the programmed delay  $\delta t$ . The colored dots indicate the control parameter values studied in Fig. 2 and the linked small color plots show the

corresponding potentials of mean force, determined from the propulsion angle histograms in Fig. 2C, together with a fit of the refined analytical model, including the instrumental delay  $\Delta t$  (see Sec. 2.2 and 3 of Supplementary Information). The only free parameter for fitting is the effective temperature of the system.  
**C** Relaxation time  $\tau$  of a single active particle as determined experimentally from the autocorrelation of the propulsion angle fluctuations (Eq. (8), data points). The solid lines correspond to the refined version of the theoretical prediction (Eq. (7)), including the instrumental delay  $\Delta t$  (see Sec. 2.2 of Supplementary Information for details). The colored dots have the same meaning as in panel (B). **D** Transition rates between the two rotational states obtained from the experiments (circles) plotted with the predictions from Kramers' theory, Eq. (9), with a global fit parameter  $D_\theta = 0.05 \text{ s}^{-1}$  (solid line) and  $D_\theta$  fitted to the probability distribution  $p(\theta)$  separately for each value  $\omega_0 \delta t$  (squares). Error bars represent the standard error.

where  $D_0 \approx 0.0642 \text{ } \mu\text{m}^2 \text{ s}^{-1}$  denotes the translational diffusion coefficient of the active particle and  $\eta(t)$  white noise. To solve this equation, we approximated  $\dot{\phi}(t)\delta t$  by  $\theta(t)$  and expanded the  $\sin(\phi(t) - \phi(t - \delta t))$  in a Taylor series around  $\delta t = 0$  up to the third order in  $\delta t$ . We dropped the term proportional to  $\ddot{\phi}(t)$  to secure the stability of the resulting equation<sup>38</sup> (for details, see Sec. 3 of Supplementary Information). The resulting noise term  $\sqrt{8D_0/(\omega_0 \delta t R^2)}$  turned out to be inaccurate compared to experimental and simulation data. We, therefore, introduce an effective diffusion coefficient  $D_\theta$  as a free parameter in the noise term in Eq. (4) to describe the rotation of the active particle around the target as the angular Brownian motion

$$\dot{\theta} = \frac{1}{3\delta t} [\theta_\pm^2 - \theta^2] \theta + \sqrt{2D_\theta} \eta \tag{4}$$

with

$$\theta_\pm = \pm \sqrt{6 \left( 1 - \frac{1}{\omega_0 \delta t} \right)}. \tag{5}$$

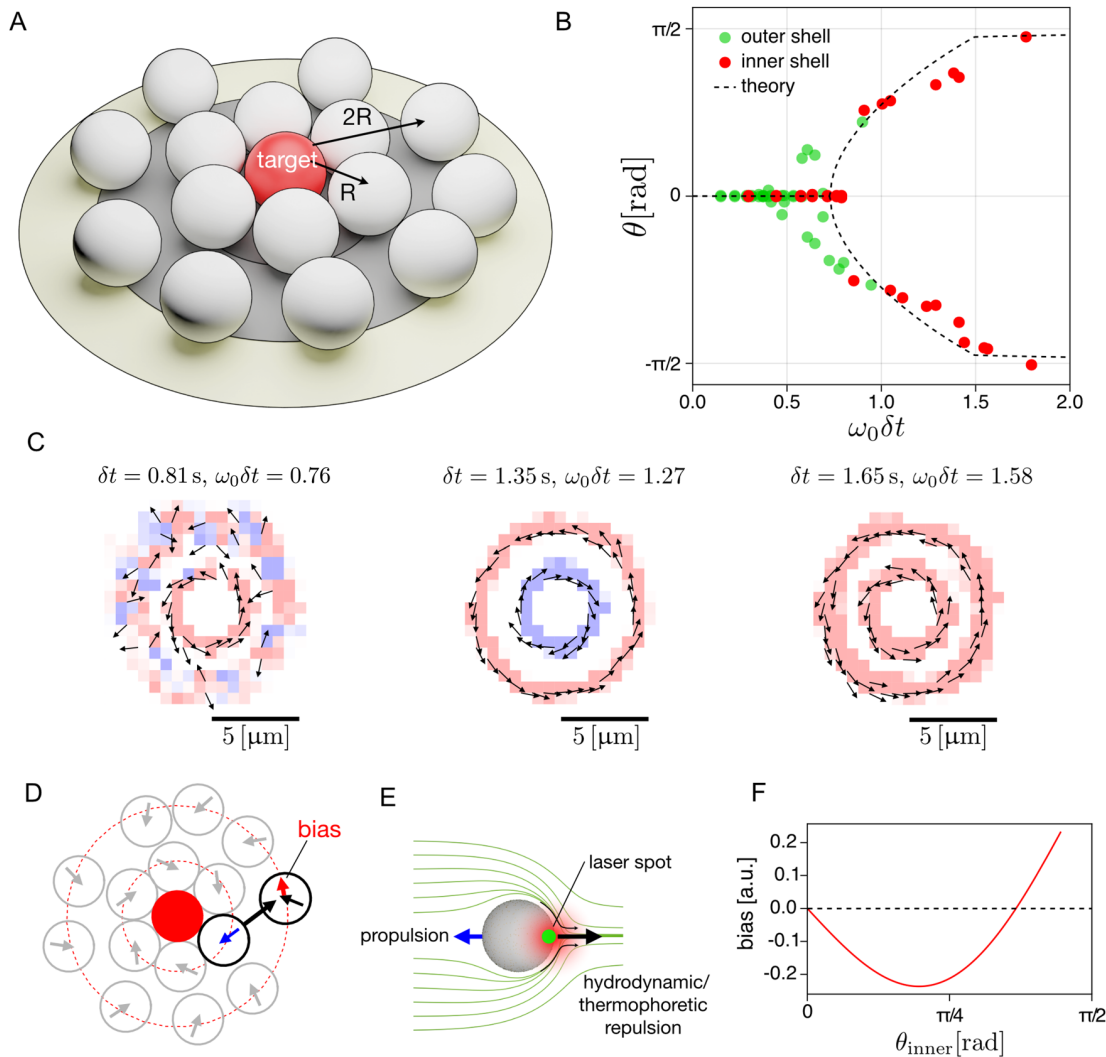
Eq. (4) yields the stationary solutions 0 and  $\theta_\pm$  with the bifurcation point  $\omega_0 \delta t = 1$ , for the transition from a non-rotational to a rotational state. The data points in Fig. 3B display the experimentally obtained maxima of the histograms  $p(\theta)$  of the propulsion angle (see Fig. 2C) as a function of  $\omega_0 \delta t$ . The transition

points in the experiments are located at lower values of the control parameter  $\omega_0 \delta t$ , due to the mentioned instrumental delay in the feedback loop of the experimental setup. This instrumental delay between the most recent exposure to the camera and the laser positioning affects the motion direction beyond the programmed delay  $\delta t$ <sup>34,39</sup>, causing an earlier onset of the transition to a stable rotation. The dashed line in Fig. 3B shows the theoretical prediction, which includes both the instrumental delay  $\Delta t$  and the programmed delay  $\delta t$ , as detailed in the Supplementary Information (Eq. (11)).

The Langevin equation (4) can be interpreted as a dynamical equation for the position  $\theta$  of an overdamped Brownian particle with diffusion coefficient  $D_\theta$  in a quartic potential (see derivation in Sec. 3 of Supplementary Information),

$$U(\theta) = \frac{1}{\delta t} \left[ \left( \frac{1}{\omega_0 \delta t} - 1 \right) \theta^2 + \frac{1}{12} \theta^4 \right], \tag{6}$$

which allows to classify the observed instability of the isotropic state as a normal supercritical pitchfork bifurcation<sup>40</sup>. The potential can also directly be extracted from the experimental data (Fig. 3B) by fitting the histogram  $p(\theta)$  with a (normalized) Boltzmann distribution  $\exp(-U(\theta)/D_\theta)/Z$  at the effective temperature  $D_\theta$ . The effective temperature thus links the measured potential of mean force  $-D_\theta \log p(\theta)$  to Eq. (6).



**Fig. 4 | Collective rotation of 15 particles attracted to a single target particle.** **A** Sketch of the shell structure and radii. **B** Bifurcation of the most probable propulsion angle as a function of the control parameter  $\omega_0 \delta t$  for a (calibrator) propulsion speed of  $v_0 = 2.06 \mu\text{m s}^{-1}$ . The red dots are obtained from the inner shell particles at a typical distance of  $R^{\text{in}} = 2.18 \mu\text{m}$ , while the green dots denote the outer shell particles at  $R^{\text{out}} = 4.47 \mu\text{m}$ . The dashed line corresponds to the theoretical single-particle prediction, including the instrumental delay  $\Delta t = 70 \text{ ms}$ . **C** Average velocity field of active particles at  $\delta t = 0.81 \text{ s}$  when the spontaneous rotation of the inner shell is constantly disrupted by the non-rotating outer shell, at  $\delta t = 1.35 \text{ s}$  when the two shells are counter-rotating,

and at  $\delta t = 1.65 \text{ s}$  when both shells are co-rotating. The arrows and colors denote the average direction of motion. **D** Snapshot of the active particles and their propulsion directions corresponding to **(C)** at  $\delta t = 1.35 \text{ s}$ . The repulsion induced by the flow and temperature fields of the inner shell causes a bias for the outer shell rotation. **E** Sketch of the flow and temperature fields induced by the laser (green dot) around an active particle, and the resulting repulsion. **F** Schematic sketch of the presumed magnitude of the bias caused by the temperature and flow fields on the rotation of the outer shell, as a function of the propulsion angle  $\theta^{\text{in}}$  of the inner shell particles (see Sec. 5 of Supplementary Information).

The latter resembles the Landau free energy at a second-order phase transition<sup>41</sup>. For readers familiar with this framework, this mathematical analogy allows to shortcut the following analysis, the details of which are given in Sec. 3 of the Supplementary Information. Note, however, that we are not discussing a thermodynamic phase transition but merely a dynamical bifurcation, here. The bifurcation and its potential energy landscape are not due to strong many-particle couplings, but to the interaction of the single active particle with its own past image. In Landau's theory, the control parameter  $1 - \omega_0 \delta t$  maps onto the dimensionless distance to the critical temperature. Both the activity  $\omega_0$  and the delay  $\delta t$  favor the transition to the symmetry-broken state. Hence, at high propulsion speeds, already short delays can give rise to rotating orbits. The inverse of the second derivative of  $U(\theta)$ , corresponding to the static susceptibility in Landau theory, gives

the time  $\tau$  (Eq. (7)) to relax in the (meta-)stable states,

$$\tau = \begin{cases} \frac{\delta t}{2} \left( \frac{1}{\omega_0 \delta t} - 1 \right)^{-1} & \omega_0 \delta t < 1 \\ -\frac{\delta t}{4} \left( \frac{1}{\omega_0 \delta t} - 1 \right)^{-1} & \omega_0 \delta t > 1. \end{cases} \quad (7)$$

We determine  $\tau$  experimentally via  $C(\tau)$  from the autocorrelation function,

$$C(t) = \frac{\langle \delta\theta(t' + t) \delta\theta(t') \rangle_{t'}}{\langle \delta\theta(t')^2 \rangle_{t'}} \quad (8)$$

of fluctuations of the propulsion angle  $\delta\theta(t) = \theta(t) - \langle \theta(t) \rangle$ , as  $C(\tau) = 1/e$  (Fig. 3C). The experimental data (circles) is compared to Eq. (7) (dashed line), and to an improved model prediction (solid line) that

also takes into account the inevitable instrumental delay  $\Delta t$ , as discussed in Sec. 3 of the Supplementary Information. The critical slowing down of the relaxation due to an increasingly flat potential close to the transition point at  $\omega_0\delta t = 1$ , corresponding to the potential plot in the middle of Fig. 3B, is thereby nicely confirmed, without any free parameter.

While the rotational orbits can be inferred from a purely deterministic model excluding Brownian motion, the observed spontaneous reversal of the chirality is driven by fluctuations in the propulsion angle and, thus, by the (non-equilibrium) noise in the system. It corresponds to transitions between the minima  $\pm\theta_*$  of the virtual potential, Eq. (6). We may thus apply Kramers' theory to estimate the corresponding transition rate as

$$k = \frac{\sqrt{2} |\omega_0\delta t - 1|}{\pi \omega_0\delta t^2} \exp\left[-\frac{3}{\delta t D_\theta} \left(\frac{1}{\omega_0\delta t} - 1\right)^2\right]. \quad (9)$$

The effective temperature  $D_\theta$  driving the fluctuations in the virtual potential is treated as a fit parameter. Figure 3D displays the experimentally measured transition rates, obtained from the observed mean residence times of  $\theta$  in the two potential wells. They are in good agreement with Eq. (9), despite the hybrid equilibrium/non-equilibrium origin of the noisy dynamics.

### Multiple particles

As demonstrated in the previous section, the rotation observed in our experiments results from a spontaneous symmetry breaking in the dynamics of a single active particle. It originates from the particle's retarded self-propulsion to a target, which differs from standard explanations of rotational dynamics in overdamped systems, which usually blame mutual ("social") interactions between multiple agents<sup>3,9,12,42</sup>. As we demonstrate in Fig. S9B, when adding up to five more active particles to the system, each of them exhibits the same rotation and bifurcation as a single swimmer. Steric, hydrodynamic, and thermophoretic interactions among the particles then synchronize and stabilize their motion, aligning their sense of rotation. So the system exhibits collective behavior, but the dynamical symmetry breaking to a chiral dynamical state is not primarily due to the mutual interactions.

Somewhat larger numbers of particles organize into multiple rotating shells. Figure 4 summarizes the key results obtained for an ensemble of 15 active particles attracted to the target particle with the same programmed and intrinsic delays  $\delta t$  and  $\Delta t$ , respectively. For the considered range of time delays, the active particles form two tightly packed shells around the target particle (Fig. 4A). The typical distance of the inner shell particles to the target is about half that of the outer shell,  $R^{\text{out}} \approx 2R^{\text{in}} = 4a$ . So based on the single-particle picture alone, the particles in the inner and outer shells swimming at the same speed would be expected to start rotating at different delays. However, in reality, the inter-particle interactions in the compact cluster strongly correlate with the particle motion and quantitatively change the picture. Compared to the theoretical prediction,  $\omega_0\delta t = 0.73$ , we observe that for  $v_0 = 2.06 \mu\text{m s}^{-1}$  the transition to the rotational phase of the inner shell is postponed to  $\omega_0^{\text{in}}\delta t \equiv v_0\delta t/R^{\text{in}} \approx 0.83$ , corresponding to  $\delta t = 0.9$  s (see the rightmost red data point lying on the horizontal axis in Fig. 4B). Slightly below the transition, the inner shell exhibits alternating periods of rotational and stationary states. Meanwhile, the stationary outer shell compresses the inner shell due to its inwards-pointing propulsion direction (Fig. 4C, left). Figure. 4C displays the velocity fields of the particles averaged over their trajectories with three different delays. The bifurcation for the outer shell is located at  $\omega_0^{\text{out}}\delta t \equiv v_0\delta t/R^{\text{out}} \approx 0.41$ , which corresponds to the same value  $\delta t = 0.9$  s of the delay at which the inner shell undergoes its bifurcation to the rotational state (see Fig. 4B and Supplementary Movies S4–S6). For delays slightly above the transition,  $0.9 \text{ s} < \delta t < 1.41 \text{ s}$ , the two shells rotate in opposite directions, as shown in the middle plot of Fig. 4C.

The simultaneous transition and the counter-rotation of the two shells suggest that the inner shell particles generate backflows opposite to their propulsion direction, thereby repelling the outer shell particles and facilitating their transition to the rotational state, as schematically depicted in Fig. 4D–F. These backflows are presumably caused by the directional hydrodynamic and thermophoretic interactions. The surface temperature gradient across each particle creates a thermosmotic surface flow that propels the particle<sup>43</sup>. If the particle motion is opposed by an external force, such as the steric force due to the immobilized target particle, the slowed-down particle acts as a pump, creating a hydrodynamic outflow at its hot side (Fig. 4D and Sec. 2.5 and 5.2 of Supplementary Information). Furthermore, thermophoretic interactions arise from temperature gradients across the surface of a particle caused by its neighbors<sup>33</sup>. These are commonly repulsive, as found, e.g., for Janus particles in external temperature gradients<sup>33</sup>. We have carried out finite element simulations of the flow field around a mobile and an immobile self-propelling swimmer (see Sec. 2.5 of Supplementary Information). The overall near-field hydrodynamic interactions are found to be quite complex, due to many interacting particles and the nearby substrate surface<sup>44–46</sup>. They also depend on the propulsion angle  $\theta$ . An increasing innershell propulsion angle results in a changing direction and magnitude of the rotational bias onto the outer shell, which presumably varies as sketched in Fig. 4F (see Sec. 5 of Supplementary Information). As a result, for  $\delta t \geq 1.41$  s, the two shells predominantly rotate in the same sense, as shown in Fig. 4C, right. The transition from counter- to co-rotation shells corresponds to the sign flip of the bias at  $\theta^{\text{in}} \approx 67^\circ$ . At even longer delays,  $\theta^{\text{in}}$  tends to reach  $90^\circ$ , and thus the inner shell tries to take off and expand against the compression exerted by the outer shell. These competing tendencies lead to particle exchange between the two shells. While we currently cannot separate thermophoretic and hydrodynamic effects in the experiment, hydrodynamic interactions may be expected to be more important here than for a single free particle in a temperature gradient: firstly, due to the collective character of the dynamics, and secondly, due to the pump effect caused by the partial blocking of the self-phoretic motion of the individual swimmers (see Sec. 2.5 and 5.2 of Supplementary Information). These features could provide a link between our experiments and the swarming observed in bacterial colonies<sup>47,48</sup>.

### Discussion

We have demonstrated above that the motion of an active particle induced by the delayed attraction to a target point can spontaneously undergo a transition from a diffuse isotropic "barometric" state to a dynamical chiral state, upon increasing the activity and/or the delay time. The transition is well described by a pitchfork bifurcation accompanied by a characteristic critical slowing down of the response<sup>40</sup>. Similar to certain mechanical analogs<sup>49</sup>, the single-particle dynamics thus already exhibit non-trivial features more commonly associated with (mean-field) phase transitions in strongly interacting passive many-body systems. This can be explained by noting that the deterministic part,  $\dot{\phi}(t) = \omega_0 \sin(\phi(t) - \phi(t - \delta t))$ , of our stochastic delay differential equation can also be understood as the dynamical equation for a single Kuramoto phase oscillator<sup>50,51</sup>, with vanishing eigenfrequency and coupling strength  $\omega_0$ , which is trying to synchronize with its own past state. In the chiral state, the particle orbits around the target point (the central obstacle is optional). The orbiting motion is stable against noise, but its sense of rotation is only transiently maintained. This should be contrasted with the chiral states resulting from non-reciprocal coupling in the time-local Kuramoto model (without delay), as discussed by ref. 52, which hinges on the stabilization by many-body cooperativity. Based on our results, we suggest that for the single retarded oscillator, the infinite number of relaxation modes encoded in the time-delayed equation of motion play a similar role<sup>53,54</sup>.



As we have shown, the nonlinear dynamics of our experimental system can be described by an approximate analytical model, which explains the emergence of a self-generated quartic virtual potential. While such potentials are frequently found in descriptions of phase transitions and collective effects in active-particle ensembles, following various behavioral rules<sup>29,30</sup>, we reiterate that the mechanism is a different one, here. Due to the activity and the (programmed) delay, it already occurs for a single active particle aiming at a spatially fixed target. In a whole swarm of particles that are all attracted to a common target, which might be its own perceived center of mass, the single-particle bifurcation is preserved. Inter-particle collisions merely synchronize, renormalize, and stabilize the rotational states of the individual particles. Upon close contact, hydrodynamic and thermophoretic interactions become important and help the swimmers to self-organize into co- and counter-rotating orbits. In biological motile ensembles, from bacteria to fish, similar hydrodynamic mechanisms may be at work, although precise details and scales may differ widely<sup>45,55–57</sup>. The corresponding many-body effects can be subtle and may elude coarse-grained simulations and theories. This underscores the importance of well-controlled experimental model systems that may act as “hybrid simulations”, combining computer-controlled active particles with real-world environments.

To conclude, while time delays are an unavoidable outcome of coarse-graining microscopic descriptions of the feedback processes in natural systems (cf. Table S1), they are often neglected in low-dimensional models of active particle collective effects<sup>5,10</sup>. In this respect, our model system provides a new perspective, as it takes the unavoidable systematic delays in the dynamics seriously and explores their generic effects. We find that, in overdamped systems, retardation plays a similar role as added inertia. Both effects lead to persistence and associated “aiming errors” in particle dynamics. In this sense, our analysis can provide a template for an entire class of motile ensembles exhibiting spontaneous rotational dynamics caused by aiming errors—as such, are associated with microswimmer navigation strategies employing “vision-cone”<sup>29,30</sup> or “acceptance-angle”<sup>27,36</sup> criteria. In fact, the effects of the time delay may be even richer<sup>20,24,54</sup>. While we considered only a positive delay, i.e., synchronization with the past, above, sophisticated biological organisms also possess predictive capabilities to extrapolate the current state into the future<sup>58,59</sup>. These can, to a first approximation, be incorporated in the form of a negative time delay. The inclusion of positive and negative delays may therefore provide a new, “more physical” perspective on phenomenologically extracted, rather sophisticated rules like collision avoidance and alignment interactions, commonly postulated as sources of emerging complex adaptive responses in living many-body systems.

## Methods

### Sample preparation

Samples were prepared using two glass coverslips (20 mm × 20 mm, 24 mm × 24 mm) to confine a thin liquid layer (3 μm thickness) in between. The edges of one coverslip are sealed with a thin layer of PDMS (polydimethylsiloxane) to prevent leakage and evaporation. The liquid film used in the sample is composed of 2.19-μm-diameter gold-coated melamine formaldehyde (MF) particles (microParticles GmbH) dispersed in 0.1% Pluronic F-127 solution. The latter prevents the cohesion of the particles and adsorption to the cover slide surface. The surface of the MF particles is speckled uniformly with gold nanoparticles of about 8 nm diameter with a total surface coverage of about 10% (Fig. S3A). SiO<sub>2</sub> particles (2.96 μm in diameter, microParticles GmbH) are added into the solution to keep the thickness of the liquid layer at about 3 μm. Finally, 0.3 μl of the mixed particle suspension is pipetted on one of the coverslips, for which the other serves as a lid.

### Experimental setup

The experimental setup (see Sec. 2 of Supplementary Information) consists of an inverted microscope (Olympus, IX71) with a mounted piezo translation stage (Physik Instrumente, P-733.3). The sample is illuminated with an oil-immersion darkfield condenser (Olympus, UDCW, NA 1.2–1.4) and a white-light LED (Thorlabs, SOLIS-3C). The scattered light is imaged by an objective lens (Olympus, UPlanApo × 100/1.35, Oil, Iris, NA 0.5–1.35) and a tube lens (250 mm) to an EMCCD (electron-multiplying charge-coupled device) camera (Andor, iXon DV885LC). The variable numerical aperture of the objective was set to a value below the minimum aperture of the darkfield condenser.

The microparticles are heated by a focused, continuous-wave laser at a wavelength of 532 nm (CNI, MGL-III-532). The beam diameter is increased by a beam expander and sent to an acousto-optic deflector (AA Opto-Electronic, DTSXY-400-532) and a lens system to steer the laser focus in the sample plane. The deflected beam is directed towards the sample by a dichroic beam splitter (D, Omega Optical, 560DRLP) and focused by an oil-immersion objective (Olympus, UPlanApo × 100/1.35, Oil, Iris, NA 0.5–1.35) to the sample plane ( $\omega_0 \approx 0.8 \mu\text{m}$  beam waist in the sample plane). A notch filter (Thorlabs, NF533-17) is used to block any remaining back reflections of the laser from the detection path. The acousto-optic deflector (AOD), as well as the piezo stage, are driven by an AD/DA (analog-digital/digital-analog) converter (Jäger Messtechnik, ADwin-Gold II). A LabVIEW program running on a desktop PC (Intel Core i7 2600 4 × 3.40 GHz CPU) is used to record and process the images as well as to control the AOD feedback via the AD/DA converter.

### Data availability

All data in support of this work is available in the manuscript or the Supplementary Information. Further data and materials are available from the corresponding author upon request.

### References

- Kauffman, S. *The Origins of Order: Self-organization and Selection in Evolution* (Oxford Univ. Press, 1993).
- Vicsek, T. & Zafeiris, A. Collective motion. *Phys. Rep.* **517**, 71–140 (2012).
- Delcourt, J., Bode, N. W. F. & Denoël, M. Collective vortex behaviors: diversity, proximate, and ultimate causes of circular animal group movements. *Q. Rev. Biol.* **91**, 1–24 (2016).
- Strandburg-Peshkin, A. et al. Visual sensory networks and effective information transfer in animal groups. *Curr. Biol.* **23**, R709–R711 (2013).
- Pearce, D. J. G., Miller, A. M., Rowlands, G. & Turner, M. S. Role of projection in the control of bird flocks. *Proc. Natl. Acad. Sci. USA* **111**, 10422–10426 (2014).
- Cremer, J. et al. Chemotaxis as a navigation strategy to boost range expansion. *Nature* **575**, 658–663 (2019).
- Couzin, I. D. & Krause, J. Self-organization and collective behavior in vertebrates. *Adv. Study Behav.* **32**, 1–75 (2003).
- Ioannou, C. C., Guttal, V. & Couzin, I. D. Predatory fish select for coordinated collective motion in virtual prey. *Science* **337**, 1212–1215 (2012).
- Berdahl, A. M. et al. Collective animal navigation and migratory culture: from theoretical models to empirical evidence. *Philos. Trans. R. Soc. B Biol. Sci.* **373**, 20170009 (2018).
- Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I. & Shochet, O. Novel type of phase transition in a system of self-driven particles. *Phys. Rev. Lett.* **75**, 1226–1229 (1995).
- Hemelrijk, C. K. & Hildenbrandt, H. Some causes of the variable shape of flocks of birds. *PLoS ONE* **6**, e22479 (2011).
- Costanzo, A. & Hemelrijk, C. K. Spontaneous emergence of milling (vortex state) in a Vicsek-like model. *J. Phys. D Appl. Phys.* **51**, 134004 (2018).

13. Cavagna, A. et al. Scale-free correlations in starling flocks. *Proc. Natl. Acad. Sci. USA* **107**, 11865–11870 (2010).
14. Mora, T. & Bialek, W. Are biological systems poised at criticality? *J. Stat. Phys.* **144**, 268–302 (2011).
15. Muñoz, M. A. Colloquium: criticality and dynamical scaling in living systems. *Rev. Mod. Phys.* **90**, 031001 (2018).
16. Cavagna, A. et al. Dynamic scaling in natural swarms. *Nat. Phys.* **13**, 914–918 (2017).
17. Kim, D. W., Hong, H. & Kim, J. K. Systematic inference identifies a major source of heterogeneity in cell signaling dynamics: the rate-limiting step number. *Sci. Adv.* **8**, eabl4598 (2022).
18. Zhang, J. & Zhou, T. Markovian approaches to modeling intracellular reaction processes with molecular memory. *Proc. Natl. Acad. Sci. USA* **116**, 23542–23550 (2019).
19. More, H. L. & Donelan, J. M. Scaling of sensorimotor delays in terrestrial mammals. *Proc. R. Soc. B Biol.* **285**, 20180613 (2018).
20. Mijalkov, M., McDaniel, A., Wehr, J. & Volpe, G. Engineering sensorial delay to control phototaxis and emergent collective behaviors. *Phys. Rev. X* **6**, 1–16 (2016).
21. Forgoston, E. & Schwartz, I. B. Delay-induced instabilities in self-propelling swarms. *Phys. Rev. E* **77**, 035203 (2008).
22. Piowarczyk, R., Selin, M., Ihle, T. & Volpe, G. Influence of sensorial delay on clustering and swarming. *Phys. Rev. E* **100**, 012607 (2019).
23. Khadka, U., Holubec, V., Yang, H. & Cichos, F. Active particles bound by information flows. *Nat. Commun.* **9**, 3864 (2018).
24. Holubec, V., Geiss, D., Loos, S. A. M., Kroy, K. & Cichos, F. Finite-size scaling at the edge of disorder in a time-delay vicsek model. *Phys. Rev. Lett.* **127**, 258001 (2021).
25. Attanasi, A. et al. Information transfer and behavioural inertia in starling flocks. *Nat. Phys.* **10**, 615–698 (2014).
26. Qian, B., Montiel, D., Bregulla, A., Cichos, F. & Yang, H. Harnessing thermal fluctuations for purposeful activities: The manipulation of single micro-swimmers by adaptive photon nudging. *Chem. Sci.* **4**, 1420–1429 (2013).
27. Bregulla, A. P., Yang, H. & Cichos, F. Stochastic localization of microswimmers by photon nudging. *ACS Nano* **8**, 6542–6550 (2014).
28. Söker, N. A., Auschra, S., Holubec, V., Kroy, K. & Cichos, F. How activity landscapes polarize microswimmers without alignment forces. *Phys. Rev. Lett.* **126**, 228001 (2021).
29. Baeuerle, T., Loeffler, R. C. & Bechinger, C. Formation of stable and responsive collective states in suspensions of active colloids. *Nat. Commun.* **11**, 2547 (2020).
30. Loeffler, R. C., Baeuerle, T., Kardar, M., Rohwer, C. M. & Bechinger, C. Behavior-dependent critical dynamics in collective states of active particles. *EPL* **134**, 64001 (2021).
31. Liebchen, B. & Löwen, H. Which interactions dominate in active colloids? *Chem. Phys.* **150**, 061102 (2019).
32. Stark, H. Artificial chemotaxis of self-phoretic active colloids: collective behavior. *Acc. Chem. Res.* **51**, 2681–2688 (2018).
33. Auschra, S., Bregulla, A., Kroy, K. & Cichos, F. Thermotaxis of Janus particles. *Eur. Phys. J. E* **44**, 90 (2021).
34. Fränzl, M., Muinos-Landin, S., Holubec, V. & Cichos, F. Fully steerable symmetric thermoplasmonic microswimmers. *ACS Nano* **15**, 3434–3440 (2021).
35. Selmke, M., Khadka, U., Bregulla, A. P., Cichos, F. & Yang, H. Theory for controlling individual self-propelled micro-swimmers by photon nudging I: directed transport. *Phys. Chem. Chem. Phys.* **20**, 10502–10520 (2018).
36. Selmke, M., Khadka, U., Bregulla, A. P., Cichos, F. & Yang, H. Theory for controlling individual self-propelled micro-swimmers by photon nudging II: confinement. *Phys. Chem. Chem. Phys.* **20**, 10521–10532 (2018).
37. Tunström, K. et al. Collective states, multistability and transitional behavior in schooling fish. *PLoS Comput. Biol.* **9**, e1002915 (2013).
38. Insperger, T. On the approximation of delayed systems by Taylor series expansion. *J. Comput. Nonlinear Dyn.* **10**, 024503 (2015).
39. Muinos-Landin, S., Fischer, A., Holubec, V. & Cichos, F. Reinforcement learning with artificial microswimmers. *Sci. Robot.* **6**, eabd9285 (2021).
40. Strogatz, S. H. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering* (Perseus Books, 1994).
41. Goldenfeld, N. *Lectures on Phase Transitions and the Renormalization Group* 1st edn (CRC Press, 1992).
42. Vollmer, J., Vegh, A. G., Lange, C. & Eckhardt, B. Vortex formation by active agents as a model for Daphnia swarming. *Phys. Rev. E* **73**, 061924 (2006).
43. Bregulla, A. P., Würger, A., Günther, K., Mertig, M. & Cichos, F. Thermo-osmotic flow in thin films. *Phys. Rev. Lett.* **116**, 188303 (2016).
44. Popescu, M. N., Uspal, W. E. & Dietrich, S. Chemically active colloids near osmotic-responsive walls with surface-chemistry gradients. *J. Phys. Condens. Matter* **29**, 134001 (2017).
45. Spagnolie, S. E. & Lauga, E. Hydrodynamics of self-propulsion near a boundary: predictions and accuracy of far-field approximations. *J. Fluid Mech.* **700**, 105–147 (2012).
46. Lintuvuori, J. S., Würger, A. & Stratford, K. Hydrodynamics defines the stable swimming direction of spherical squirmers in a nematic liquid crystal. *Phys. Rev. Lett.* **119**, 068001 (2017).
47. Wioland, H., Woodhouse, F. G., Dunkel, J. & Goldstein, R. E. Ferromagnetic and antiferromagnetic order in bacterial vortex lattices. *Nat. Phys.* **12**, 341–345 (2016).
48. Nishiguchi, D., Aranson, I. S., Snezhko, A. & Sokolov, A. Engineering bacterial vortex lattice via direct laser lithography. *Nat. Commun.* **9**, 4486 (2018).
49. Fletcher, G. A mechanical analog of first- and second-order phase transitions. *Am. J. Phys.* **65**, 74–81 (1997).
50. Kuramoto, Y. International symposium on mathematical problems in theoretical physics. *Lect. Notes Phys.* **30**, 420 (1975).
51. O’Keeffe, K. P., Hong, H. & Strogatz, S. H. Oscillators that sync and swarm. *Nat. Commun.* **8**, 1504 (2017).
52. Fruchart, M., Hanai, R., Littlewood, P. B. & Vitelli, V. Non-reciprocal phase transitions. *Nature* **592**, 363–369 (2021).
53. Loos, S. A. M. & Klapp, S. H. L. Fokker–planck equations for time-delayed systems via markovian embedding. *J. Stat. Phys.* **177**, 95–118 (2019).
54. Geiss, D., Kroy, K. & Holubec, V. Brownian molecules formed by delayed harmonic interactions. *New J. Phys.* **21**, 093014 (2019).
55. Drescher, K., Dunkel, J., Cisneros, L. H., Ganguly, S. & Goldstein, R. E. Fluid dynamics and noise in bacterial cell-cell and cell-surface scattering. *Proc. Natl. Acad. Sci. USA* **108**, 10940–10945 (2011).
56. Lauder, G. V. & Drucker, E. G. Forces, fishes, and fluids: hydrodynamic mechanisms of aquatic locomotion. *Physiology* **17**, 235–240 (2002).
57. Verma, S., Novati, G. & Koumoutsakos, P. Efficient collective swimming by harnessing vortices through deep reinforcement learning. *Proc. Natl. Acad. Sci. USA* **115**, 5849–5854 (2018).
58. Morin, A., Caussin, J.-B., Eloy, C. & Bartolo, D. Collective motion with anticipation: flocking, spinning, and swarming. *Phys. Rev. E* **91**, 012134 (2015).
59. Palmer, S. E., Marre, O., Berry, M. J. & Bialek, W. Predictive information in a sensory population. *Proc. Natl. Acad. Sci. USA* **112**, 6908–6913 (2015).

## Acknowledgements

We thank Daniel Geiss for assembling the data in Table S1 and acknowledge funding through DFG-GACR cooperation by the Deutsche Forschungsgemeinschaft (DFG Project No 432421051) and by the Czech Science Foundation (GACR Project No 20-02955J). V.H. was supported

by the Humboldt Foundation. We thank Andrea Kramer for proofreading the manuscript. We also acknowledge financial support by the Open Access Publishing Fund of Leipzig University supported by the German Research Foundation within the program Open Access Publication Funding.

### Author contributions

X.W. and F.C. conceived the experiment. X.W. carried out the experiment. X.W., F.C., P.-C.C., and V.H. analyzed the data. P.-C.C., K.K., and V.H. developed the theory. X.W., P.-C.C., and F.C. carried out simulations. X.W., P.-C.C., K.K., V.H., and F.C. discussed the results and wrote the manuscript.

### Funding

Open Access funding enabled and organized by Projekt DEAL.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-35427-7>.

**Correspondence** and requests for materials should be addressed to Frank Cichos.

**Peer review information** *Nature Communications* thanks Anton Souslov, and the other, anonymous, reviewer for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

# Supplementary Information

## Spontaneous vortex formation by microswimmers with retarded attractions

Xiangzun Wang<sup>1</sup>, Pin-Chuan Chen<sup>2</sup>, Klaus Kroy<sup>2</sup>, Viktor Holubec<sup>3</sup> and Frank Cichos<sup>1,\*</sup>

<sup>1</sup>*Peter Debye Institute for Soft Matter Physics, Molecular Nanophotonics Group, Universität Leipzig, 04103 Leipzig, Germany.*

<sup>2</sup>*Institute for Theoretical Physics, Leipzig University, Postfach 100 902, 04009 Leipzig, Germany.*

<sup>3</sup>*Department of Macromolecular Physics, Faculty of Mathematics and Physics, Charles University, 18000 Prague, Czech Republic*

\* *cichos@physik.uni-leipzig.de*

### Contents

<b>1</b>	<b>Delay times in swarming animals</b>	<b>2</b>
<b>2</b>	<b>Experimental Details</b>	<b>2</b>
2.1	Experimental setup . . . . .	2
2.2	Feedback system . . . . .	3
2.3	Active particle control . . . . .	3
2.4	Data analysis . . . . .	4
2.5	Thermophoresis and hydrodynamic flow . . . . .	4
<b>3</b>	<b>Single particle theory</b>	<b>5</b>
3.1	Deterministic dynamics with instrumental delay and bifurcation diagram . . . . .	5
3.2	Stochastic dynamics: time-local approximation . . . . .	6
3.3	Stochastic dynamics: potential and relaxation times . . . . .	8
3.4	Stochastic dynamics: transition rates and effective temperatures . . . . .	9
<b>4</b>	<b>Brownian dynamics simulations</b>	<b>10</b>
<b>5</b>	<b>Collective rotation</b>	<b>11</b>
5.1	Steric interactions . . . . .	11
5.2	Hydrodynamic interactions . . . . .	12
<b>6</b>	<b>Extraction of relaxation times from experimental and simulation data</b>	<b>15</b>

# 1 Delay times in swarming animals

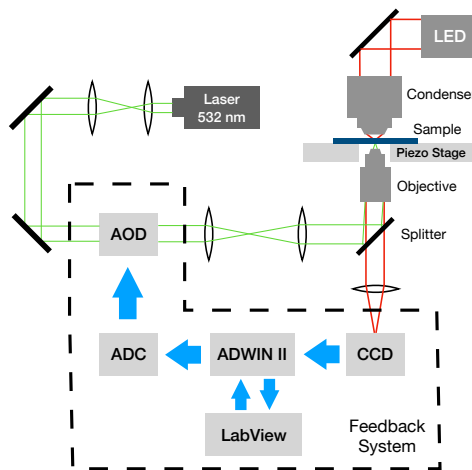
**Tab. S1:** Mean reaction times measured between a stimulus and the corresponding discrete response strongly vary among species and the type of stimulus. Delay times comparable to the characteristic time scale of the stimulus may be expected to trigger qualitatively new effects in the dynamical response, similar to those analyzed in the present work. Specifically, bacteria (such as *E. coli*) are of similar size to our particles and also operate in an aqueous thermal environment.

Animal	Stimulus/Response	Reaction Time [ms]	References
Human	auditory	140 – 160	1
	visual	180 – 200	1
	touch	~ 155	1
Fruit fly	roll perturbation	~ 5	2
	pitch perturbation	~ 12	3
	yaw perturbation	10 – 25	4
Starling	startling sound stimuli	64 – 80	5
	startling light stimuli	38 – 76	5
Teleost fish	startle response	5 – 10	6,7
Calanoida	stirring water	< 2.5	8
<i>E. coli</i>	chemical stimuli	~ 10 <sup>3</sup> – 10 <sup>4</sup>	9

## 2 Experimental Details

### 2.1 Experimental setup

Fig. S1 sketches the experimental setup including the feedback system and the signal flow.



**Fig. S1:** Sketch of the experimental setup.

The laser of 532 nm wavelength from the module (CNI Laser, MGL-III-532) is expanded in the beam diameter by two lenses (19 mm, 100 mm focal lengths) and projected to the acousto-optic deflector (AOD, AA Opto-Electronic, DTSXY-400-532). The two perpendicularly aligned TeO<sub>2</sub> crystals in the AOD diffract the incident laser in horizontal and vertical directions. The deflection angle and the output laser intensity are controlled by the frequency and amplitude of the voltage on the TeO<sub>2</sub> which is controlled by a real-time board Adwin-Gold II (Jäger Messtechnik) including digital analog converters. The Adwin board is further controlled by a LabView program developed in the group.

Through two lenses (500 mm, 300 mm focal lengths), the laser is guided to the dark field microscope (Olympus, IX71). Reflected by a dichroic beam splitter (Omega Optical, 560DRLP), the beam is then focused by an objective (100x, Olympus, UPlanApo x 100/1.35, Oil, Iris, NA 0.5 – 1.35) on the sample. The full width of the

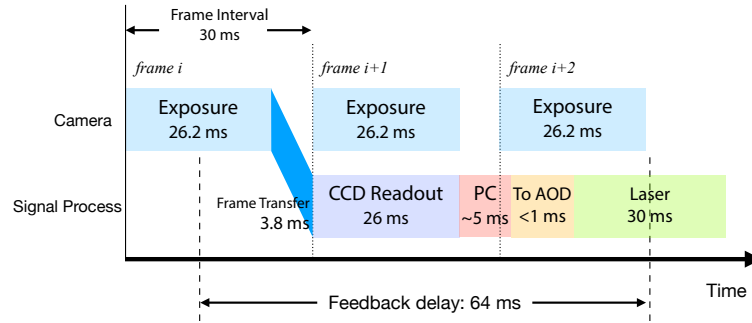
focus at half maximum of intensity is about  $1 \mu\text{m}$ . The two lenses and the objective after the AOD convert the deflection angles to the position of laser focus on the sample. About 20 mrad range of the deflection angle corresponds to the  $58 \times 58 \mu\text{m}^2$  area in the field of view in the microscope.

If multiple positions on the sample need to be shot by the laser, for instance to simultaneously propel multiple swimmers, the AOD keeps scanning the shoot positions circularly. Every position is exposed for  $320 \mu\text{s}$ , and the position switching takes about  $10 \mu\text{s}$ . The scanning order is randomized for each camera frame. Before every measurement, the AOD is calibrated with a sample of thin layer of Nile blue fluorescent dye that show the position of the laser focus on the sample. By scanning the laser in the field of view, the conversion between laser position and AOD voltage input is determined by linear interpolation.

White light from a LED lamp (Thorlabs, SOLIS-3C) is shined on the sample through an oil-immersion dark field condenser (Olympus, U-DCW, NA 1.2 – 1.4) with a glancing angle and illuminates the rims of swimmers. Hence, the swimmers can be observed as bright rings, the positions of which are evaluated in LabVIEW. The image of sample is projected through the objective and a lens (250 mm focal length) to the EMCCD camera (Andor, iXon DV885LC). The resolution of the image reaches  $0.058 \mu\text{m}/\text{px}$ . A notch filter (Thorlabs NF533-17) in front of the camera blocks the residual laser reflected by the sample.

## 2.2 Feedback system

In the experiments, the time for signal transfer and processing in the feedback system causes an inevitable instrumental delay  $\Delta t$  on swimmer control. This delay is a property of the experimental setup and, unlike the delay  $\delta t$  in the interaction rule (see Sec. 3.1), cannot be easily controlled. Influence of both  $\Delta t$  and  $\delta t$  on the system dynamics is discussed in detail in Sec. 3.



**Fig. S2:** Time diagram of signal processing in the feedback system for one active swimmer experiment.

Fig. S2 shows the time diagram of the signal processing in the feedback system. In the “Frame Transfer Mode”, the camera is exposed while exporting the image of the previous frame. The image is read out and transferred via the Adwin-Gold II (Jäger Messtechnik) to a desktop PC (Intel Core i7 2600 4 x 3.40 GHz CPU). The LabVIEW program (v. 2019) on the PC analyzes the image and evaluates the positions of the swimmers by their bright rims. The program stores a short history of the locations of the swimmers, and tracks the swimmers by comparing their locations in previous frames. The corresponding data are recorded to a hard drive by another CPU thread. The propelling directions at time  $t$  are determined from the swimmer locations at time  $t - \delta t$  in the past. The laser position for propelling is determined based on the latest measured swimmer location.

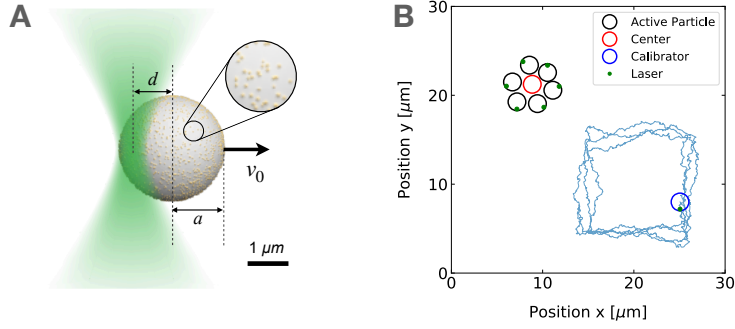
With the information from the Nile blue calibration, the laser positions are converted to voltage signals for the AOD. The signals are sent to the Adwin-Gold II, converted to analogue voltages, and transmitted to the AOD oscillator module.

The measurements with one active swimmer (Fig. 2 in the main text) used a  $512 \times 512$  pixels field of view and a 30 ms frame interval. The corresponding  $\Delta t$  was about 64 ms. The experiments with 15 swimmers (Fig. 4 in the main text) used a larger  $592 \times 592$  pixels field of view and, hence, a longer time for camera read-out resulting in a longer  $\Delta t$  of about 70 ms.

## 2.3 Active particle control

As self-propelled swimmers the experiment uses melamine formaldehyde (MF) particles of  $2.19 \mu\text{m}$  diameter (microParticles GmbH) decorated with gold particles of about 0.8 nm diameter on about 10 % of their surface.

To propel the MF particle, the laser is focused at its edge at a distance of  $d = 0.812 \mu\text{m}$  from the center, as sketched in Fig. S3A.



**Fig. S3:** (A) Sketch of the MF particle and the focused laser beam. The laser to particle distance is  $d = 0.812 \mu\text{m}$  in the experiment. Figure is reprinted from<sup>10</sup>. (B) The active swimmers (black circles), the target (red circle), and the trajectory (blue line) of the calibrator (blue circle) in a measurement.

The MF particle is transparent to the laser, which heats the gold particles and generates an asymmetric temperature gradient on the MF surface. Due to the thermophoretic effect, the swimmer is propelled away from the laser focus. For a detailed description of the propelling mechanism, see<sup>10,11</sup>.

During the sample preparation, the swimmers can be adsorbed on the glass coverslip and become immobilized. This effect is significantly weakened by adding 0.1 % Pluronic F-127 in the sample. In the experiments, we used the remaining adsorbed swimmers as the target particles.

To keep the swimmer velocities constant during the experiment, we use a “calibrator” particle driven to patrol in the field of view far enough from the other swimmers to be independent. The calibrator is driven to sequentially change its swimming direction to follow an approximately square trajectory, as plotted in Fig. S3B. Its speed averaged over the square loop is measured in real time and the laser power is tuned by LabVIEW and the AOD to keep it constant.

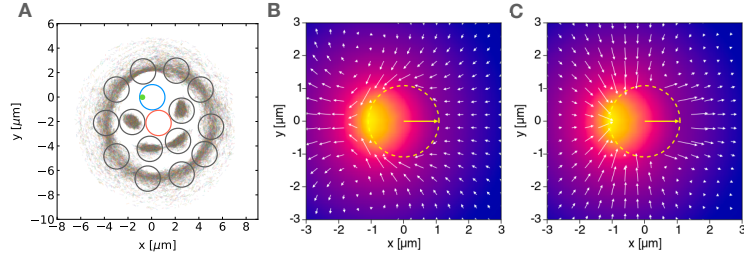
## 2.4 Data analysis

From the recorded positions of particles in the experiments, the trajectories of the propulsion angles  $\theta(t)$  between the position vectors at time  $t$  and  $t - \delta t$  are determined. Note that the definition of the propulsion angle as the angle between the propulsion direction  $\hat{\mathbf{u}}(t)$  and the negative position vector at time  $t$ , given on the right-hand side of Eq. (2) in the main text, is valid only for a vanishing instrumental delay  $\Delta t$ , and we do not use it when analyzing the data. The obtained trajectories  $\theta(t)$  are used to construct the  $\theta$  histograms with 10 mrad bin size shown in Fig. 2C in the main text. The velocity  $v_0$  of the active particles is determined by the average velocity of the calibrator introduced in Section 2.3. The orbit radius  $R$  is calculated as the average distance of the active particle to the target in one measurement. For the measurements with two shells of particles (Fig. 4 in the main text), the traces with  $2 \mu\text{m} < R < 3.09 \mu\text{m}$  are identified as particles in the inner shell, and  $3.5 \mu\text{m} < R < 6.18 \mu\text{m}$  as those in the outer shell.

## 2.5 Thermophoresis and hydrodynamic flow

The swimmer is propelled due to thermophoretic effects caused by the laser-induced temperature gradient. The propulsion mechanism generates a flow of the liquid around the particle, which influences the motion of nearby particles in the measurements with multiple swimmers (e.g., see Fig. S4A). Figs. S4B and C demonstrate the hydrodynamic flow field around a fixed and a freely propelling swimmer from a COMSOL simulation. The fixed swimmer generates a flow opposite to its heading motion at the “tail”, while the moving swimmer causes the flow in the direction of its motion. These hydrodynamic and thermophoretic interactions between the particles are causing the many-body effects (co- and counter-rotating shells) depicted in Fig. 4 in the main text. Note that the swimmers in Fig. 4 are neither fully fixed nor freely movable. Depending on their propulsion angles, the corresponding flow fields are thus between those of the free and fixed swimmer.

From the trajectories of particles in a cluster, the influence of hydrodynamic flow can be deduced from the velocity distribution of the particles relative to each other (Fig. S4). The flow at the “tail” of a swimmer

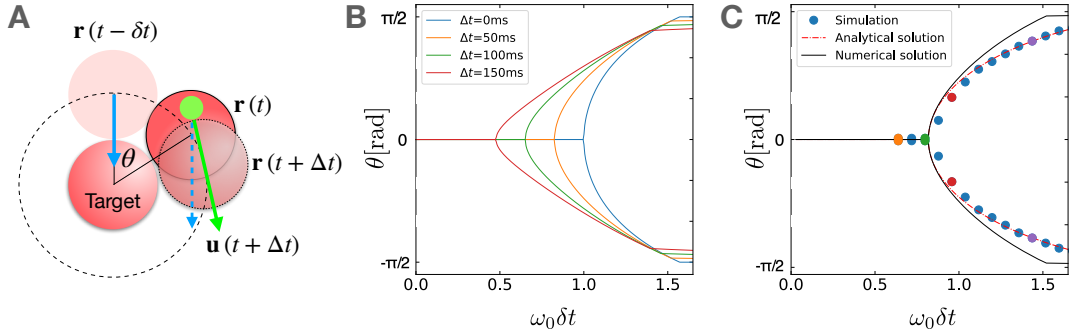


**Fig. S4:** (A) The distribution of 16 swimmers relative to one of them (blue circle) in a measurement. The red circle denotes the fixed target particle, and the green dot the laser.  $v_0 = 2 \mu\text{m/s}$ . Panels (B) and (C) show the temperature and (induced) hydrodynamic flow fields around a fixed swimmer and a freely moving swimmer, respectively. The data were obtained using a finite element method calculation with COMSOL.

repulses other swimmers behind it. This repulsion causes the opposite rotations of different swimmer shells, as introduced in the main text (for more details, see Sec. 5). Note that Fig. S4 is calculated by the trajectories of the self-propelling active particles that obey their own rule of interaction, hence cannot be considered equivalent as the real liquid flow.

### 3 Single particle theory

#### 3.1 Deterministic dynamics with instrumental delay and bifurcation diagram



**Fig. S5:** (A) Sketch of the effect of the instrumental delay  $\Delta t$  on the motion of the swimmer.  $\mathbf{r}(t)$  is the measured particle location detected by the camera, according to which the laser is focused at time  $t + \Delta t$  when the position of the swimmer is  $\mathbf{r}(t + \Delta t)$ . (B) The bifurcation of the stable angle  $\theta$  with different  $\Delta t$ 's obtained numerically from Eq. (2).  $\theta$  becomes constant after  $\omega_0 \delta t = \pi/2$ , when  $\angle(\hat{\mathbf{u}}, -\hat{\mathbf{r}}) = 90^\circ$  and the swimmer leaves the target's surface. (C) The bifurcation diagrams obtained with  $\Delta t = 64$  ms from Brownian dynamics simulations (circles), numerically (solid line), and using the approximate analytical solution (12) to Eq. (2) (dashed line). The data points with different colors correspond to the potentials in Fig. S7A. In B and C, we set  $v_0 = 2.03 \mu\text{m/s}$ ,  $a = 1.095 \mu\text{m}$  and  $d = 0.8 \mu\text{m}$ . In the simulations, we used the diffusion coefficient  $D_0 = 0.0642 \mu\text{m}^2/\text{s}$ .

Consider the dynamics of a single swimmer attracted by a fixed target with the delay  $\delta t$  and the instrumental delay  $\Delta t$ , as depicted in Fig. S5A. The laser position is determined from the swimmer location  $\mathbf{r}(t)$  detected by the camera at time  $t$ . However, due to the instrumental delay  $\Delta t$ , the actual location of the swimmer is  $\mathbf{r}(t + \Delta t)$  when the laser is updated. Neglecting the noise, the swimmer displacement  $\mathbf{r}(t + \Delta t) - \mathbf{r}(t)$  during  $\Delta t$  causes a change of its propulsion direction  $\hat{\mathbf{u}} = -\hat{\mathbf{r}}(t - \delta t)$  from Eq. (1) in the main text to

$$\hat{\mathbf{u}}(t + \Delta t) = \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t) - d\hat{\mathbf{r}}(t - \delta t)}{|\mathbf{r}(t + \Delta t) - \mathbf{r}(t) - d\hat{\mathbf{r}}(t - \delta t)|}, \quad (1)$$

where  $d$  is the programmed distance from the laser to the particle as plotted in Fig. S3A. The form  $\hat{\mathbf{u}}(t) = -\hat{\mathbf{r}}(t - \delta t)$  is recovered in the limit  $\Delta t \rightarrow 0$ . The instrumental delay  $\Delta t$  makes the swimmer motion considerably



complex. It is known to amplify the effects of noise by an amount roughly proportional to  $1/d^{12}$ . We will now show that it influences the swimmer's dynamics even if the Brownian motion is neglected.

Let us stick to the experimentally relevant situation depicted in Fig. S5 when the motile particle slides along the fixed particle and thus cannot move in the direction  $-\hat{\mathbf{r}}(t + \Delta t)$ . Then the true direction of motion of the motile particle is given by the projection  $(1 - \hat{\mathbf{r}}(t + \Delta t) \hat{\mathbf{r}}(t + \Delta t))$  of the propulsion direction  $\hat{\mathbf{u}}(t + \Delta t)$  in Eq. (1) to the direction  $(-\sin \phi(t + \Delta t), \cos \phi(t + \Delta t))$  perpendicular to  $\hat{\mathbf{r}}(t + \Delta t) = (\cos \phi(t + \Delta t), \sin \phi(t + \Delta t))$ . The propulsion thus creates a rotation around the fixed particle with diameter of rotation  $2a$  and angular velocity  $\omega(t + \Delta t) = v_0 \hat{\mathbf{u}}(t + \Delta t) \cdot (-\sin \phi(t + \Delta t), \cos \phi(t + \Delta t)) / (2a)$  given by

$$\frac{\omega(t)}{\omega_0} = \frac{2a \sin \theta(t, \Delta t) + d \sin \theta(t, \delta t + \Delta t)}{\sqrt{d^2 + 8a^2 - 8a^2 \cos \theta(t, \Delta t) + 4ad(\cos \theta(t - \Delta t, \delta t) - \cos \theta(t, \delta t + \Delta t))}}. \quad (2)$$

Here,  $\theta(t, t') = \int_{t-t'}^t dt'' \omega(t'')$  generalizes the propulsion angle from the main text. The maximum possible angular velocity of the swimmer is denoted by  $\omega_0 = v_0 / (2a)$ . When  $\Delta t = 0$ , Eq. (2) simplifies to Eq. (3) in the main text.

If we assume a stable rotation with a constant angular velocity  $\omega(t) = \omega \neq 0$ , and neglect Brownian motion and the dependency of the swimmer velocity  $v_0$  on the laser to the particle distance  $d^{10}$ , Eq. (2) can be used for generation of the bifurcation diagram. As shown in Fig. 3A in the main text, the bifurcation appears when derivatives of both sides of Eq. (2) with respect to  $\omega$  at  $\omega = 0$  are equal. This leads to the condition

$$\omega_0 \delta t = 1 - \omega_0 \Delta t \frac{2a + d}{d}. \quad (3)$$

Compared to the situation with  $\Delta t = 0$ , discussed in the main text, the instrumental delay shifts the critical point  $\Theta_0 \equiv \omega_0 \delta t$  to a lower value proportionally to  $\omega_0 \Delta t$ . For  $\Delta t \geq d / (\omega_0 (2a + d))$  when the right hand side of Eq. (3) becomes negative, a stable rotation can form even with  $\delta t = 0$ .

The right hand side of Eq. (3) depends on  $\omega_0$ , which enters the control parameter  $\omega_0 \delta t$ . Actually, Eq. (3) can also be written as

$$\omega_0 \delta t = \left( 1 + \frac{\Delta t}{\delta t} \frac{2a + d}{d} \right)^{-1} \equiv \Theta_B, \quad (4)$$

where the right-hand side depends on  $\delta t$ . These results suggest that  $\omega_0 \delta t$  is for  $\Delta t > 0$  no longer a good control parameter and one should redefine it, e.g., by introducing the effective delay time  $\delta t_{\text{eff}} = \delta t / \Theta_B$  so that the condition in Eq. (4) simplifies to  $\omega_0 \delta t_{\text{eff}} = 1$ . Nevertheless, we keep the control parameter  $\omega_0 \delta t$  for sake of consistency with the main text.

The bifurcation diagram in Fig. S5B shows the numerical solutions to Eq. (2) as function of  $\omega_0 \delta t$  for four different values of instrumental delay  $\Delta t$ . In the next section, we show that the bifurcation can also be characterized analytically using a Taylor expansion. The resulting approximate formula (Eq. (12)) is compared to the numerical solution in Fig. S5C. This panel also shows the bifurcation diagram obtained from Brownian dynamics simulations of the system (for details of the simulations, see Sec. 4). The simulations reveal that the Brownian motion neglected in the above analysis causes deviations from our predictions due to fluctuations of the propulsion direction (Eq. (1)) and swimmer to target distance  $R$ . Bifurcation diagrams obtained from simulations with a much smaller diffusion coefficient than that used in the figure coincide perfectly with the numerical solution (data not shown).

### 3.2 Stochastic dynamics: time-local approximation

Taking into account the noise, the dynamics of the motile particle can be described by the Langevin equation<sup>11</sup>

$$\dot{\mathbf{r}}(t) = v_0 \hat{\mathbf{u}}(t) + \mathbf{F}_{\text{int}} / \gamma_0 + \sqrt{2D_0} \boldsymbol{\eta}(t), \quad (5)$$

where  $D_0 = k_B T / \gamma_0$  ( $\sim 0.0642 \mu\text{m}^2/\text{s}$ ) with friction coefficient  $\gamma_0$ , Boltzmann constant  $k_B$ , and temperature  $T$ , denotes translational diffusion coefficient, and  $\boldsymbol{\eta}(t) = (\eta_x(t), \eta_y(t))^T$  is a column vector of independent Gaussian white noises. The force  $\mathbf{F}_{\text{int}}$  describes the hard-core interaction between the motile and fixed particle. Because the particle position  $\mathbf{r}(t)$  enters the expression Eq. (1) for the propulsion direction  $\hat{\mathbf{u}}(t)$  at time  $t$  with a delayed time argument, the Langevin equation is a non-linear stochastic delay differential equation<sup>13</sup>. In general, such equations are notoriously difficult to solve analytically, and one has to resort to approximations.

Similarly, as in the previous section, we will now assume that the distance of the motile particle from the origin is determined by twice the radius of the fixed particle,  $|\mathbf{r}(t)| \sim 2a$ , and we will focus only on the component of Eq. (5) that is perpendicular to the position vector  $\mathbf{r}(t)$ . This assumption allows us to neglect the force  $\mathbf{F}_{\text{int}}$  in the equation. The motion along the position vector has been investigated in a slightly different context in Ref.<sup>11</sup>.

Projecting Eq. (5) to the direction perpendicular to  $\mathbf{r}(t)$  is most easily achieved by scalar multiplying the equation by the vector  $(-\sin \phi(t), \cos \phi(t))$ . Using the expressions  $\mathbf{r}(t) = 2a(\cos \phi(t), \sin \phi(t))$  and  $\dot{\mathbf{r}}(t) = 2a(-\sin \phi(t), \cos \phi(t))\omega(t)$ , where  $\omega(t) = \dot{\phi}(t)$ , we obtain

$$\omega(t) = \omega_0 F(t) + \sqrt{2D} \eta(t) \quad (6)$$

with  $D \equiv D_0/(2a)^2$ , and  $\eta(t)$  the white noise. The factor  $F(t) = \hat{\mathbf{u}}(t) \cdot (-\sin \phi(t), \cos \phi(t))$  reduces the maximum angular velocity  $\omega_0$ . It is given by Eq. (2), which simplifies to  $\sin\left(\int_{t-\delta t}^t \omega(t') dt'\right)$  for a vanishing instrumental delay. Eq. (6) is still a non-linear stochastic delay differential equation, and it is virtually impossible to solve it exactly using the available techniques. However, qualitative analytical insights into the dynamics of the system can be obtained by expanding the factor  $F$  in a third-order Taylor series with respect to the delays  $\delta t$  and  $\Delta t$ . The resulting equation has the form

$$\omega(t) = c_0 \omega(t) + c_1 \omega^3(t) + c' \dot{\omega}(t) + c'' \ddot{\omega}(t) + \sqrt{2D} \eta(t) \quad (7)$$

with  $c' < 0$  and  $c'' > 0$ . This equation is local in time, and, by rearranging the terms, it can be interpreted as a Langevin equation for an underdamped Brownian particle with a mass  $-c''$  and friction  $-c'$  trapped in a quartic potential.

However, such a system would be unstable due to the negative mass term  $-c''$ , which stems from expanding the delay frequency up to second order in  $\delta t$ . That such higher-order expansions can lead to unstable solutions is well known in the theory of delay systems<sup>14</sup>, and thus we set  $c'' = 0$  in Eq. (7). Furthermore, the angular velocity process in Eq. (7) is already proportional to white noise  $\eta(t)$ , and thus the derivative  $\dot{\omega}(t)$  does not exist in a strict mathematical sense. To avoid this problem, we will from now on understand  $\omega(t)\delta t$  as an approximation for the propulsion angle  $\theta(t) = \int_{t-\delta t}^t dt' \omega(t')$ , which works reasonably well when  $\omega$  is approximately constant. Finally, we obtain the approximate time-local stochastic differential equation

$$\dot{\theta}(t) = - \left. \frac{\partial}{\partial \theta} U(\theta) \right|_{\theta=\theta(t)} + \sqrt{2D_T} \eta(t) \quad (8)$$

for the propulsion angle. It has the form of a Langevin equation describing a dimensionless position  $\theta$  of an overdamped Brownian particle with diffusion coefficient

$$D_T = \frac{(2 + \alpha)^2 \Theta_B^4}{(2\Theta_B^2 + \alpha)^2 \Theta_0^2} \frac{D_0}{a^2} \quad (9)$$

with  $\alpha \equiv d/a$ , diffusing in a dimensionless quartic potential

$$U(\theta) = \frac{1}{4} f_0 \theta^2 (\theta^2 - 2\theta_{\pm}^2). \quad (10)$$

This analogy will prove immensely useful in the following discussion. The parameters in the potential read

$$f_0 = \frac{c}{(2 + \alpha)(2\Theta_B^2 + \alpha)\Theta_B} \frac{1}{3\delta t}, \quad \theta_{\pm} = \frac{(2 + \alpha)\Theta_B}{\sqrt{c}} \sqrt{\frac{6}{\Theta_0} (\Theta_0 - \Theta_B)}, \quad (11)$$

where we introduced the shorthand  $c = 12 - 4(3 - \Theta_B)\Theta_B^2 + \alpha(6 + \alpha - 2\Theta_B^3)$  and, as in the main text, denoted the control parameter  $\omega_0 \delta t$  by  $\Theta_0$ . For a vanishing instrumental delay,  $\Theta_B = 1$ ,  $c = (2 + \alpha)^2$ , and the potential and effective diffusion coefficient simplify to the expressions given in the main text with

$$f_0 = \frac{1}{3\delta t}, \quad \theta_{\pm} = \sqrt{\frac{6}{\Theta_0} (\Theta_0 - 1)}. \quad (12)$$

As stated in the main text, the quartic potential (Eq. (10)) in Eq. (8) directly maps to the Landau theory of phase transitions. Further, the dynamical equation (8) with the potential (Eq. (10)) and zero noise ( $D_T = 0$ ) represents the normal form of the supercritical pitchfork bifurcation. Both these mappings yield expressions for the bifurcation diagram and the relaxation times. However, here we will employ a different approach based on the mentioned mapping of Eq. (8) to the overdamped diffusion, and show that, while it allows us to derive the results from the Landau theory, it also allows for characterizing the effect of the noise on the particle dynamics.

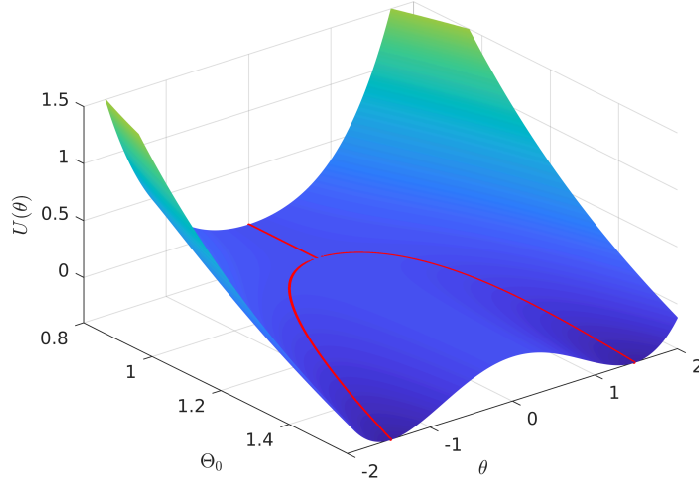
### 3.3 Stochastic dynamics: potential and relaxation times

The (approximate) bifurcation diagram can be deduced from the force

$$-\frac{\partial}{\partial\theta}U(\theta) = f_0\theta(\theta^2 - \theta_{\pm}^2). \quad (13)$$

For  $\Theta_0 < \Theta_B$  it vanishes only for  $\theta = 0$ . On the other hand, for  $\Theta_0 > \Theta_B$  the force vanishes for  $\theta = 0$  and  $\theta = \pm\theta_{\pm}$ . The bifurcation diagram shows the stationary solutions to the differential equation which thus must be stable to small perturbations. We introduce two ways how to determine the stability.

The simpler one is based on the potential: stable stationary solutions correspond to its local minima and unstable solutions to its local maxima. The potential  $U(\theta)$  as function of the characteristic parameter  $\Theta_0 = \omega_0\delta t$  is depicted in Fig. S6. From the figure, it follows that  $\theta = 0$  is a stable solution only for  $\Theta < \Theta_B$ , where it corresponds to a single minimum of the potential. For  $\Theta > \Theta_B$ , the potential develops two local minima and the system allows for two symmetric stable solutions  $\theta = \pm\theta_{\pm}$ . In accord with the discussion from the preceding section, the bifurcation appears at  $\Theta = \Theta_B$  when a single stable solution changes to two stable solutions.



**Fig. S6:** The potential  $U(\theta)$  as function of the control parameter  $\Theta_0 \equiv \omega_0\delta t$  according to Eq. (10) with  $\Delta t = 0$ , corresponding to  $\Theta_B = 1$  according to Eq. (4). With increasing  $\Theta_0$ ,  $U(\theta)$  transforms from a single well to a double well potential. The red curve denotes the local minima of  $U(\theta)$  and thus the stable angles  $\theta_{\pm}$  in the bifurcation diagram in Figs. S5B and C.

Another way to identify the stable solutions, which, moreover, yields approximate relaxation times of small perturbations from the stable points, is the linear response theory. Let us first investigate the stability of the solution  $\theta = 0$ . Assuming small perturbations around this point, i.e., taking  $\theta = \delta\theta$  in Eq. (8) and expanding the result up to the first order in  $\delta\theta$ , we find the equation

$$\frac{d}{dt}\delta\theta(t) = f_0\theta_{\pm}^2\delta\theta(t), \quad (14)$$

which describes an exponential relaxation  $\delta\theta \propto \exp(-t/\tau)$  with the relaxation time

$$\tau = -\frac{1}{f_0\theta_{\pm}^2} = \frac{2\Theta_B^2 + \alpha}{2 + \alpha} \frac{\delta t}{2\left(\frac{1}{\Theta_0} - \frac{1}{\Theta_B}\right)}. \quad (15)$$

This relaxation time is positive for  $\Theta_0 < \Theta_B$  and negative for  $\Theta_0 > \Theta_B$ . At the bifurcation,  $\Theta_0 \rightarrow \Theta_B$ , the relaxation time diverges, which is the manifestation of critical slowing down. Stationary solutions are stable if the corresponding relaxation time is positive and thus the solution  $\theta = 0$  is stable for  $\Theta_0 < \Theta_B$ .

Repeating this procedure for small perturbations around  $\theta = \pm\theta_{\pm}$ , we find that the corresponding relaxation time is given by

$$\tau = -\frac{2\Theta_B^2 + \alpha}{2 + \alpha} \frac{\delta t}{4\left(\frac{1}{\Theta_0} - \frac{1}{\Theta_B}\right)}. \quad (16)$$

This relaxation time is negative for  $\Theta_0 < \Theta_B$ , positive for  $\Theta_0 > \Theta_B$ , and diverges for  $\Theta_0 \rightarrow \Theta_B$ . The solutions  $\theta = \pm\theta_{\pm}$  are thus stable for  $\Theta_0 > \Theta_B$ . The linear stability analysis hence leads to the same bifurcation diagram as predicted from the potential in Fig. S6. The positive relaxation times constitute the system discussed in the main text. For vanishing instrumental delay  $\Delta t = 0$ , the formulas (15) and (16) simplify to Eq. (7) in the main text.

The theoretical predictions (15) and (16) are compared to relaxation times obtained from experiments and Brownian dynamics simulations in Fig. S11C in Sec. 6, where we discuss two complementary methods for extracting the relaxation times from the data.

### 3.4 Stochastic dynamics: transition rates and effective temperatures

In the regime when the potential becomes bi-stable, the noise induces transitions between the two minima. Employing the overdamped interpretation of Eq. (8), it seems natural to use the Kramers' rate theory<sup>15</sup> to describe these transitions. For an overdamped Brownian particle diffusing in a double well potential  $U$  with minima at  $\pm\theta_{\pm}$ , the Kramers theory predicts that transitions between the two wells appear with the transition rate

$$k = \frac{1}{2\pi} \sqrt{-U''(0)U''(\theta_{\pm})} \exp\left(\frac{-E_b}{D_{\theta}}\right). \quad (17)$$

Here,  $E_b = U(0) - U(\theta_{\pm})$  is the height of the energetic barrier between the minima, and  $D_{\theta}$  denotes the thermal energy, which is, in our units with unit friction coefficient, given by the diffusion coefficient. Using the potential and the diffusion coefficient from Eq. (8) thus leads to the prediction

$$k = \frac{f_0\theta_{\pm}^2}{\sqrt{2\pi}} \exp\left(-\frac{f_0\theta_{\pm}^4}{4D_{\theta}}\right), \quad (18)$$

with  $D_{\theta} = D_T$ . The red solid line and the circles in Fig. S7C show that this prediction unfortunately does not agree with transition rates evaluated from Brownian dynamics simulations of Eq. (5). This disagreement is caused by the uncontrolled approximations used in our derivation. Actually, there is no guarantee that a delay stochastic differential equation can be approximated by a time-local stochastic differential equation. The only case where such a mapping can be derived rigorously are linear stochastic differential equations, which can be shown to be equivalent to time-local stochastic differential equations with highly nontrivial time-dependent coefficients and a colored noise<sup>13</sup>. But even there the transition rates differ from those predicted using the simple overdamped description<sup>13</sup>.

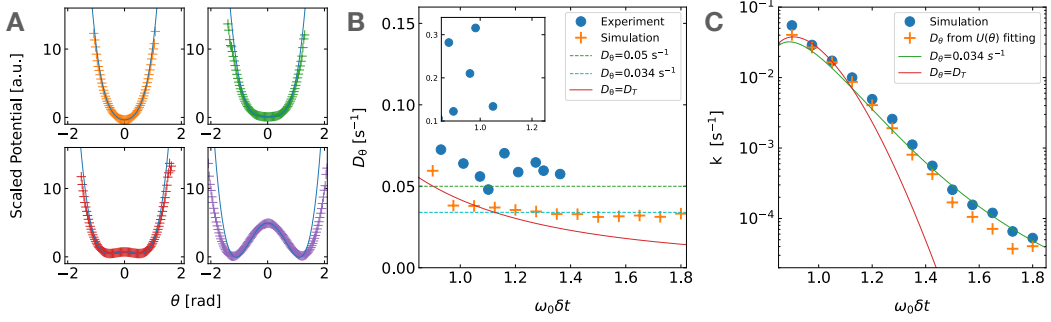
In the present situation, the agreement of the theoretically predicted bifurcation diagram and relaxation times with the experimental and simulation data shown in Figs. S5 and S11, and Fig. 3 in the main text gives us confidence that the potential Eq. (10) approximates well the systematic part of the delay force in Eq. (5). To improve the prediction for the transition rate, we thus decided to fit the diffusion coefficient  $D_{\theta}$  in Eq. (18) to the data. However, to further test the validity of the suggested mapping of Eq. (5) to the time-local overdamped Eq. (8), we have not determined  $D_{\theta}$  by fitting the transition rates. Instead, we fit the Boltzmann distribution

$$p(\theta) = \frac{\exp(-U(\theta)/(D_{\theta}))}{\int_{-\infty}^{\infty} \exp(-U(\theta)/(D_{\theta}))} \quad (19)$$

to the distributions for  $\theta$  determined from experiments and simulations. Examples of such fits to  $-\log p(\theta)$  obtained from simulations and experiments are depicted in Fig. S7A and Fig. 2C in the main text, respectively. The corresponding diffusion coefficients  $D_{\theta}$  are shown in Fig. S7B together with the prediction (9). Both  $D_{\theta}$  obtained from experiments and simulations converge to (different) constant values depicted in the figure by horizontal lines. The large fluctuations in  $D_{\theta}$  obtained from experimental data for small values of  $\omega_0\delta t$ , shown in the inset, are caused by insufficient statistics around the central peak of  $-\log p(\theta)$ , rendering the fitting results unreliable.

The transition rates calculated using Eq. (18) with the fitted  $D_{\theta}$  from Fig. S7B are depicted by symbols in Fig. S7C for the simulations and Fig. 3D in the main text for the experiments. Besides using the fitted  $\omega_0\delta t$  dependent diffusion coefficient, we also plotted the transition rates using the constant plateau values  $D_{\theta} = 0.034 \text{ s}^{-1}$  and  $D_{\theta} = 0.05$ , respectively for the simulation and the experiment. The remarkable agreement of the resulting transition rates with rates directly measured from simulation and experiments reveals that, at least

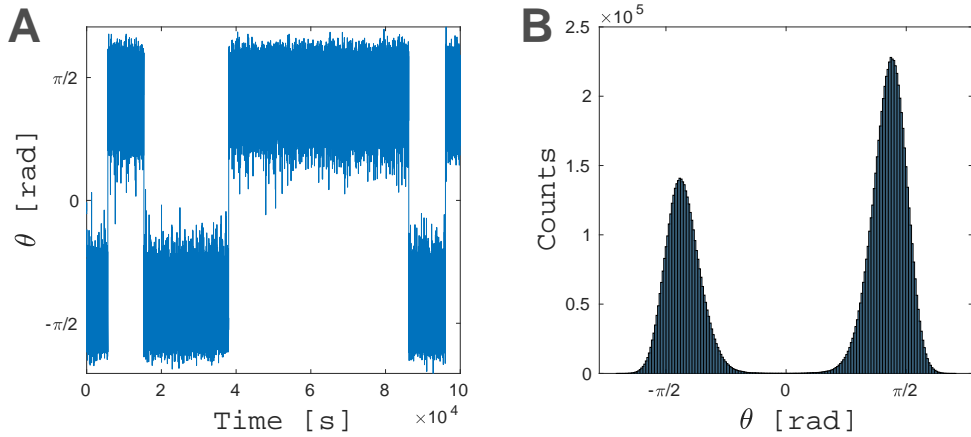
from the perspective of transition rates, the delay stochastic differential Eq. (5) can be well approximated by the time-local overdamped Langevin Eq. (8) with the potential Eq. (10) and white noise with diffusion coefficient  $D_\theta$  instead of  $D_T$  from Eq.(9).



**Fig. S7:** (A) Fits of the Boltzmann distribution (Eq. (19)) with the potential (Eq. (10)) (lines) to  $-\log p(\theta)$  from simulations (+ symbols) for  $\omega_0 \delta t = 0.60, 0.75, 0.90,$  and  $1.35$ , denoted by the same colors as the corresponding data points in Fig. S5C. (B) The diffusion coefficient  $D_\theta$  determined by the fitting of experimental and simulated distributions for  $\theta$  together with the theoretical prediction (Eq. (9)). (C) Transition rates  $k$  for jumps of  $\theta$  between the two wells of the potential, measured from the simulations (circles) and the predictions from Kramers' theory (Eq. (18)) with different diffusion coefficients  $D_\theta$ . The other parameters are the same as in Fig. S5C.

## 4 Brownian dynamics simulations

To test the approximate theoretical results beyond the parameter regimes accessible in experiments, we simulated the non-approximate Langevin equation (5) with the timestep  $dt = 10$  ms. To take into account finite sizes of the motile and the fixed particle, we assumed that  $\mathbf{F}_{\text{int}}$  describes a hard core repulsion, which restricts  $\mathbf{r}(t)$  to  $(2a, \infty)$ .



**Fig. S8:** Sample data from Brownian dynamics simulation for  $v_0 = 2.16 \mu\text{m/s}$  and  $\delta t = 1.62$  s. (A) The trajectory of the angle  $\theta(t)$  fluctuates around the stable values  $\theta_\pm$ . (B) The corresponding histogram for  $\theta$ . The other parameters are the same as in Fig. S5C.

Most of our simulations take into account all key ingredients of the experiment except for hydrodynamic interactions. Therefore, the simulated data show excellent agreement with single particle measurements, cf the example simulated trajectory of the delay angle and the corresponding histogram shown in Fig. S8A (Fig. S8B) to the experimental data in Fig. 2A (Fig. 2C) of the main text. However, our simulations of many-body systems exhibit significant disagreement from data from multiple particle experiments, especially if the particles form

several co- or counter-rotating layers. This disagreement allows us to conclude that the effects observed in many-body experiments, like the shift of the bifurcation to smaller values of control parameter and counter-rotating particle layers, are solely induced by hydrodynamic interactions among the active particles. As shown in Fig. S9 and discussed in the Sec. 5.1, the only significant many-body effect that is present in the simulations without hydrodynamic interactions is stabilization of the system (decrease of transition rate between clockwise and counter-clockwise rotation) with increasing number of particles.

We have adapted phenomenological backflow-induced forces in some simulations to study the role of hydrodynamic (and phoretic) interactions without reproducing the whole experiment *in silico*. We constructed them using three leading terms (monopole, dipole, and polar contributions) in the expansions of the hydrodynamic velocity field from Refs. <sup>16</sup> and <sup>17</sup>. We switched the monopole field decaying as  $(1/\text{distance})$  only when the target particle blocked the propelled particle. In the formula for the force exerted by the backflow, we additionally neglected the “diffusive” term proportional to the second derivative of the velocity field. We have chosen the coefficients in the three contributions so that the phenomenology observed in simulations and experiments agrees. In this sense, our simulation can also phenomenologically incorporate some phoretic effects described by the same decay laws as the considered hydrodynamic contributions. In Fig. S10 and Sec. 5.2 below, we demonstrate that these simulations reproduce the co- and counter-rotating shells observed in experiments.

## 5 Collective rotation

As described in the main text, considering an ensemble of particles each of which is driven towards the same target by the same retarded interaction leads to interesting collective effects. Due to steric, hydrodynamic and thermophoretic interactions, the particles rotate around the target in shells. In each of the shells, the inter-particle interactions synchronize and stabilize the rotation of the particles, as shown in Fig. S4A. The transition rate for switching between clockwise and counter-clockwise rotation thus decreases with the number of particles in a shell. Otherwise, the bifurcation in a given shell occurs approximately for the same parameters as for a single particle. The inter-particle interactions also couple dynamics of the neighbouring shells, which can either stably co-rotate or counter-rotate.

### 5.1 Steric interactions

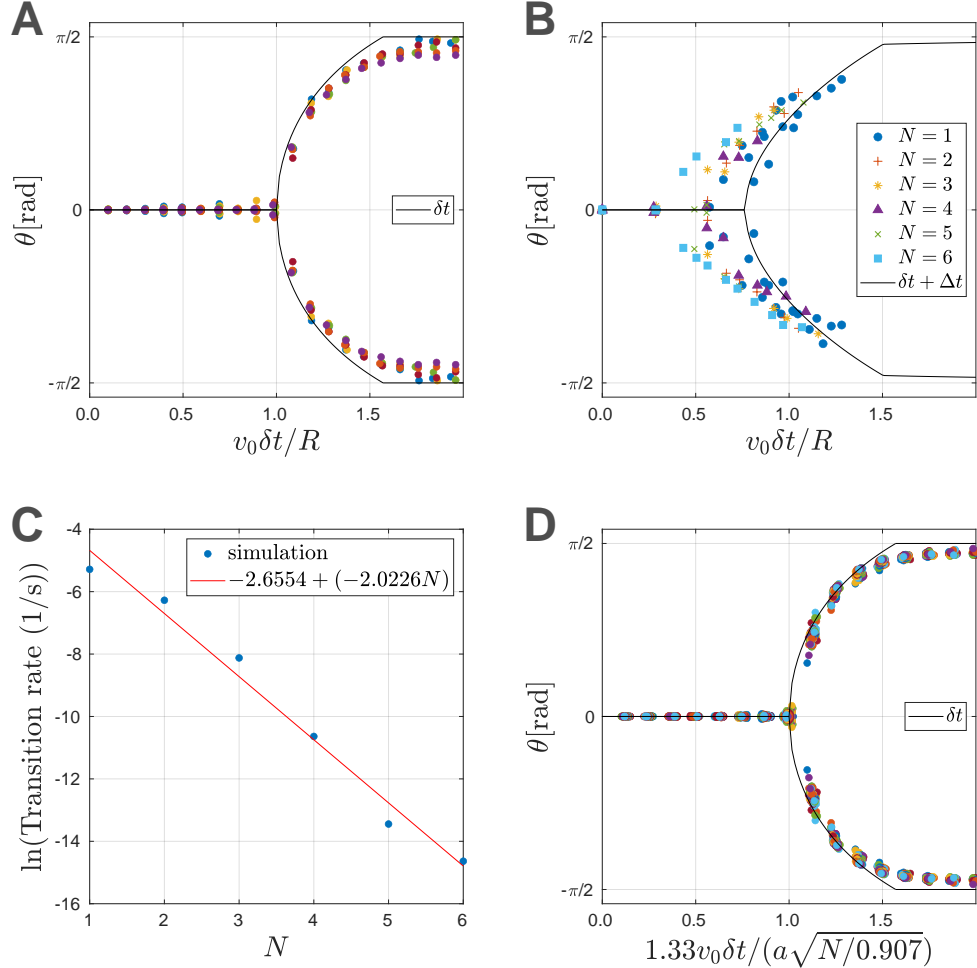
To understand the effect of steric interactions alone in the multi-particle rotation, we performed BD simulations without hydrodynamic forces and instrumental delay for  $N$  ranging from 1 to 69. For  $N = 6$  and  $\delta t < \pi a/v_0$ , the swimming particles form a closed hexagonal shell around the target particle of radius  $R \simeq 2a$ . For larger delays, the shell is no longer closely packed as its diameter increases to  $R \sim 2v_0\delta t/\pi$ . For  $N > 6$ , the innermost shell is pressed towards the central particle and its diameter returns to  $R \simeq 2a$ . The other shells form an optimal hexagonal lattice with  $6k$  particles in the  $k$ th shell. Assuming that the particles eventually fill a circle, the radius of the outermost shell can be estimated as  $a\sqrt{N}/0.907$ , where 0.907 is the density of the optimal circular packing in two dimensions.

Up to  $N = 6$  particles, the steric interactions just synchronize and stabilize the rotation. The bifurcation diagrams for  $1 \leq N \leq 6$  in Fig. S9A thus collapse to a single curve described by the single-particle theory (Eq. (5) in the main text). In Fig. S9B, we show that the corresponding bifurcation diagrams obtained from the experiments also approximately collapse for  $N \leq 6$ . We attribute the discrepancies in the collapse to hydrodynamic and phoretic interactions, neglected in the simulation, and measurement uncertainties. More detailed studies are subject of future work.

The stabilizing effect of the steric inter-particle interactions is best visualized by the exponential decrease of the transition rate for jumps between the two rotating states depicted in Fig. S9C. This linear scaling of the exponent in the transition rate with  $N$  can be intuitively understood as follows. Before the whole set of particles in a given shell changes its sense of rotation, it forms a “train of compartments” which can be thought of as a single quasiparticle with an effective diffusion coefficient that scales as  $1/N$  due to the central limit theorem. As the exponent in the transition rate is proportional to the inverse of the diffusion coefficient, this effect yields the observed linear scaling of the logarithm of the transition rate.

Finally, in Fig. S9D, we show bifurcation diagrams for  $N \geq 14$  obtained from simulations without hydrodynamic interactions by averaging the propulsion angles of all particles in the ensemble. After rescaling the control parameter  $v_0\delta t$  by the radius  $R \sim a\sqrt{N}/0.907$  of the outermost shell and a constant factor of 1.33, all these diagrams fall onto a single master curve described by the single-particle theory. For short delays, all particles

move with the same angular velocity as the outer shells propel toward the center of the flock, thus effectively imposing their angular velocity on the inner shells. With the increasing delay, the propulsion direction of the particles becomes more and more tangential, and the strength of the blocking drops. The outer shells lack behind the inner ones as the maximum angular velocity of the particles  $v_0\delta t/R$  decreases with the distance from the center. In both these regimes, the particles thus move on average with a smaller angular velocity than a single particle. It turns out that this “slowing down” is nicely described by the constant  $1.33 \approx 4/3$  we used to upscale the speed in the bifurcation diagram. Unfortunately, we cannot compare these simulation results to experimental data as control of such large ensembles of particles is beyond our experimental setup’s current capabilities.



**Fig. S9:** Collective behavior in simulations with only steric interactions and experiments. **(A)** The bifurcation diagrams for  $1 \leq N \leq 6$  obtained from simulations. The data collapse on the solid line corresponding to the single particle theory (Eq. (5) in the main text). **(B)** The bifurcation diagram for  $1 \leq N \leq 6$  extracted from experimental data. The solid line is the theoretical prediction (11) for the bifurcation curve that includes both the programmed delay  $\delta t$  and the instrumental delay  $\Delta t$ . In (A) and (B),  $R \approx 2a$  denotes the average distance of the particles from the center. **(C)** The logarithm of the transition rate for switching between clockwise and counter-clockwise rotation for  $1 \leq N \leq 6$ , obtained from simulations. **(D)** The bifurcation diagram for  $14 \leq N \leq 69$  obtained from simulations. After rescaling, the data collapse on the solid line corresponding to the single particle prediction.

## 5.2 Hydrodynamic interactions

now focus in detail on the the behavior in the two-shell scenario depicted in Fig. 4 in the main text. The maximum angular velocity in the inner shell,  $\omega^{in} = v_0/(2a)$ , is two times larger than that in the outer shell,

$\omega^{\text{out}} = v_0/(4a)$ . Therefore, one might conclude that the inner shell can rotate and the outer shell moves randomly. However, rotation in the inner shell always induces at least a weak rotation in the outer shell. As the control parameter  $\omega^{\text{out}}\delta t$  increases, the outer and inner shells start to counter-rotate. Finally, an even stronger increase in  $\omega^{\text{out}}\delta t$  leads to co-rotation of the two shells. This complicated behavior, caused by the hydrodynamic and thermophoretic interactions between particles in the two layers, can be understood using a simple phenomenological model. Let the particles in the inner shell stably rotate with a propulsion angle  $\theta^{\text{in}}$ . Then the dynamics of the propulsion angle  $\theta^{\text{out}}$  of particles in the outer shell can be qualitatively described by the formula

$$\theta^{\text{out}}(t) = \omega^{\text{out}}\delta t [\sin \theta^{\text{out}}(t - \delta t) - B \sin \theta^{\text{in}} \cos \theta^{\text{in}} + C \sin \theta^{\text{in}}] + \sqrt{2D_0/R^2}\xi(t). \quad (20)$$

The first term on the right-hand side comes from the delayed attraction to the center and leads to the single-particle potential in Eqs. (10) and (12) and the bifurcation diagram in Fig. (S5). The second term has the opposite sign than  $\theta^{\text{in}}$  and thus causes the counter-rotation of the two shells. The third term has the same sign as  $\theta^{\text{in}}$  and causes the co-rotation. The last term stands for the thermal noise. The dependency of the second and third terms on  $\theta^{\text{in}}$  can be understood as follows.

Addressing a particle by the laser causes flows of water which induce swimming of the particle. These flows are strongest in the direction opposite to the particle motion. When a particle is propelled with delay angle  $\theta^{\text{in}}$ , the backflow is thus strongest in the direction  $(-\sin \theta^{\text{in}}, \cos \theta^{\text{in}})$  in the coordinate system defined by the vector parallel to the particle's circular trajectory, and the position vector  $\mathbf{r}(t)$ , see Fig. S5A. If the backflow from the inner shell penetrates into the outer shell, it can cause the co- and counter-rotations (see Fig. 4 in the main text). The co-rotation can be induced if the delay angle  $\theta^{\text{in}}$  is so large that the backflow hits the particle in the inner shell and is reflected into the outer shell with opposite direction. We assume that this effect can be modeled by the term proportional to  $\sin \theta^{\text{in}}$ , i.e., to the component of the backflow tangential to the inner layer. The counter rotation is then induced by the amount of backflow which penetrates into the second layer. This portion of the backflow is proportional to  $\cos \theta^{\text{in}}$ . However, the leaked backflow can only cause rotation if it has a nonzero component within the second layer. This component is again proportional to  $\sin \theta^{\text{in}}$ . Altogether, we assume that the counter-rotation is caused by the term proportional to  $\sin \theta^{\text{in}} \cos \theta^{\text{in}}$ . Finally, we assume that the proportionality factor  $C > 0$  for the reflected backflow is smaller than  $B > 0$ , describing the direct effect of the backflow.

The phenomenological bias  $\omega^{\text{out}}\delta t(-B \sin \theta^{\text{in}} \cos \theta^{\text{in}} + C \sin \theta^{\text{in}})$  from the inner shell tilts the potential (10) and thus stabilizes one of the rotating states. For  $\Theta_0 < 1$ , there is no rotation in the inner shell, and thus  $\theta^{\text{in}} = 0$  and the bias vanishes. For  $\Theta_0 > 1$ ,  $\theta^{\text{in}}$  is in the present approximation given by Eq. (12). For  $0 < \theta^{\text{in}} < \arccos(C/B)$ , the term  $-B \sin \theta^{\text{in}} \cos \theta^{\text{in}}$  dominates and the bias favors  $\theta^{\text{out}}$  with opposite sign than  $\theta^{\text{in}}$ , and the other way round for  $\arccos(C/B) < \theta^{\text{in}}$ . Inserting  $\theta^{\text{in}} = \sqrt{6(1 - 1/\Theta_0)}$  from Eq. (12) in the condition  $\theta^{\text{in}} = \arccos(C/B)$ , we find that the transition from counter- to co-rotation occurs for

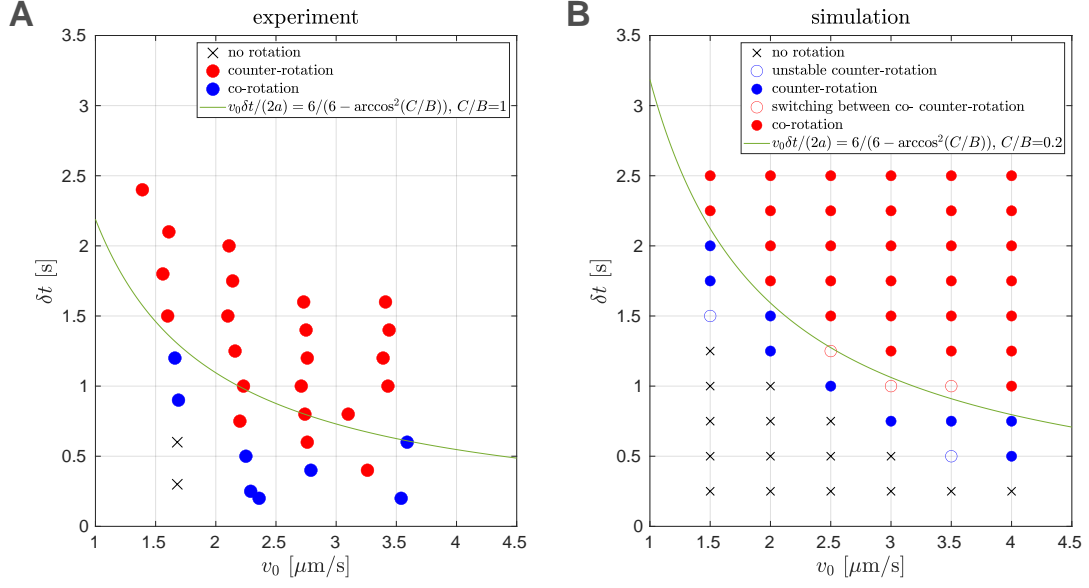
$$\Theta_0 = \frac{6}{6 - \arccos^2(C/B)}. \quad (21)$$

Depending on the value of the ratio of the phenomenological parameters  $B$  and  $C$ , the expression on the right-hand side interpolates between the values  $-24/(\pi^2 - 24) \sim 1.7$  for  $C/B = 0$  and 1 for  $C/B = 1$ . Translating this result to the language of time-delay  $\delta t$ , velocity  $v_0$ , and radius  $2a$  of the inner-shell, we find

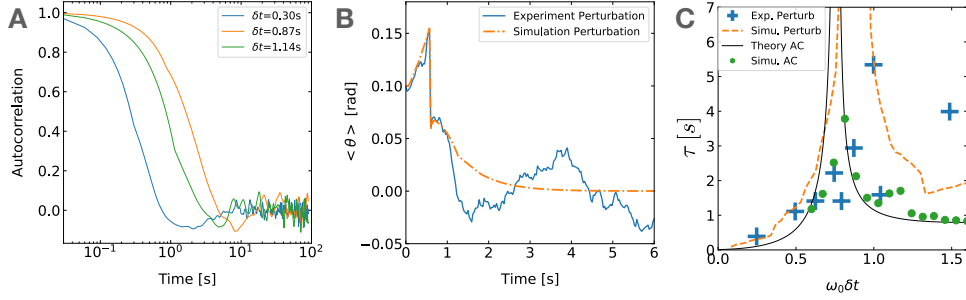
$$v_0\delta t = 12a/[6 - \arccos^2(C/B)]. \quad (22)$$

In Fig. S10B, we show the phase diagram for co and counter-rotating phases obtained from the simulation with steric repulsion, phenomenological hydrodynamic forces, and the instrumental delay of  $\Delta t = 0.064$  s (see Sec. 4 for details). The corresponding diagram obtained from experiments is shown in Fig. S10A. Eq. (22) describes the phase boundary in both these diagrams. However, the corresponding values of  $B/C$  for the simulation ( $B/C \sim 0.2$ ) and experiments ( $B/C \sim 1.0$ ) are much different. We attribute this discrepancy to the rough implementation of hydrodynamic and phoretic interactions in the simulation and the analytical model. The analytical model further does not incorporate steric interactions, which play an important role in the observed synchronization. Noteworthy, all these different implementations of the backflow yield the same phenomenology, which points to its robustness against details of the interactions' shape.





**Fig. S10:** Phase diagram of counter/co-rotating shells for  $N = 15$  obtained from experiments (A) and simulations (B). The simulations include the instrumental delay  $\Delta t = 0.064$  s and effective hydrodynamic interactions. The phase boundaries are described by Eq. (22) with  $C/B \simeq 1$  for experiments and 0.2 for simulations, corresponding to  $v_0\delta t/(2a) = 1$  and  $v_0\delta t/(2a) = 1.45$ , respectively. In addition to the two rotational phases observed in the experiments, we observed unstable/switching modes on their boundary in simulations.



**Fig. S11:** (A) Auto-correlation (AC) function of  $\theta(t)$  corresponding to the three colored experimental data points in Fig. 2 of the main text. (B) An example of the relaxation of average  $\theta(t)$  after a 0.1 rad perturbation of the stable non-rotating state. The solid line was measured in experiments and the dashed line was obtained numerically from Eq. (5). (C) The relaxation time  $\tau$  from theory (Eqs. (15) and (16), solid line), AC functions determined from Brownian dynamics simulations (circles), and relaxation of the perturbation exemplified in (B) in experiments (crosses), and numerics (dashed line). We used  $v_0 = 1.7$   $\mu\text{m/s}$  (B),  $\delta t \approx 0.6$  s ((B) and (C)),  $a = 1.095$   $\mu\text{m}$  and  $d = 0.8$   $\mu\text{m}$ .

## 6 Extraction of relaxation times from experimental and simulation data

There are two complementary ways to determine the relaxation times in the two stable states from experimental and simulation data. The first one is closely related to the idea of the linear stability analysis: one induces a small perturbation of the system from its stable state and measures how long it takes for this excitation to decay to  $1/e$  its initial value. If one cannot easily induce such a perturbation in the system, it is still possible to obtain the relaxation time as the decay time of the stationary time-correlation function calculated from trajectories trapped in the stable state.

The complementarity of the two approaches is based on the linearized Langevin equation

$$\frac{d}{dt}\delta\theta(t) = -\frac{1}{\tau}\delta\theta(t) + \eta(t) \quad (23)$$

approximately describing time evolution of small perturbations  $\delta\theta(t) = \theta(t) - \theta_S$  of a stable state  $\theta_S$ . Both the average solution to this equation,  $\langle\delta\theta(t)\rangle$  and its stationary normalized auto-correlation function,

$$C(t) = \frac{\langle\delta\theta(t+t')\delta\theta(t')\rangle_{t'}}{\langle\delta\theta(t')^2\rangle_{t'}} \equiv \lim_{t_m \rightarrow \infty} \frac{\int_0^{t_m} dt \delta\theta(t+t')\delta\theta(t')}{\int_0^{t_m} dt (\delta\theta(t'))^2}, \quad (24)$$

decay exponentially with the relaxation time  $\tau$ . In the definition of the time correlation function,  $t_m$  denotes the final time of the measurement and it is assumed that the process  $\theta(t)$  is initialized in a distant time in the past,  $t_i \rightarrow -\infty$ , so that it is already stationary at  $t = 0$ .

For an overdamped system with a harmonic potential, the two methods yield exactly the same relaxation time. For a nonlinear delay stochastic differential equation such as Eq. (5), the correspondence is not necessarily exact. However, also linearized stochastic delay differential equations yield average solutions with the same decay rate as the stationary correlation functions<sup>13</sup>. The two approaches can thus be expected to yield the same results whenever the system evolves most of the time close to the minimum of the potential, which is well approximated by a parabola. Such conditions are achieved whenever the thermal energy  $k_B T$  is small compared to a characteristic energy scale of the potential. That the two approaches yield similar results also in our situation is demonstrated in Fig. S11C and in Fig. 3C in the main text, where we compared the theoretical results Eqs. (15) and (16) and relaxation times  $\tau$  obtained from experiments and simulations using the two methods described above. More details are given in the following two paragraphs.

The time correlation functions Eq. (24) determined from experimental trajectories of  $\theta(t)$  trapped in the individual stable states ( $\theta(t)$  in the  $\pm 1$  rad range of the theoretical values  $0$  or  $\theta_{\pm}$ ) are shown in Fig. S11A. The corresponding decay times were determined as times needed for the correlation function to decay to  $1/e$  of its initial value. The resulting decay times averaged over data obtained from all trajectories trapped in a given state are shown in Fig. 3C in the main text. Similar time correlation functions also follow from the Brownian dynamics simulations. The corresponding decay times are depicted by solid circles in Fig. S11C. These decay times were evaluated in the same way as in experiments with the difference that we employed the formula

$$\int_0^{\infty} dt \sin\left(\frac{t}{\tau}\right) \frac{C(t)}{t} = \frac{\pi}{4} \quad (25)$$

designed to extract the most relevant timescale of a decaying function<sup>18</sup> to extract the decay time from the individual correlation functions. Both the experimental and the simulation results are in good agreement with the theoretical predictions Eqs. (15) and (16) according to the linear stability analysis (solid lines). We also verified that the simulation data agree with theoretical predictions better for smaller temperatures (data not shown).

An example of the time evolution of a perturbation  $\delta\theta = 0.1$  applied to the stationary non-rotating state  $\theta_S = 0$  at  $t = 0$ , corresponding to  $\omega(t) = \delta\theta(t)$  proportional to the  $\delta$ -function for  $t \leq 0$ , is shown in Fig. S11B. The dashed line results from numerical integration of Eq. (5) averaged over the noise. The solid line was measured in the experiment, where we averaged over 150 perturbation-relaxation cycles. During the time interval  $(0, \delta t)$ , the average propulsion angle  $\theta(t) = \int_{t-\delta t}^t dt' \omega(t')$  is given by the perturbation and increases as  $\omega(t) = \omega_0 \sin \theta(t)$  is positive. At time  $t = \delta t$ ,  $\theta(t)$  suddenly drops by  $\delta\theta$  as the perturbation no longer influences the angular velocity. This change then causes a next drop at  $t = 2\delta t$  etc. Such a non-smooth time evolution is for delay differential equations typical<sup>13</sup>.

For rotating stable states, the decay of a finite perturbation depends on its sign due to a slight asymmetry of the corresponding effective potential. Therefore, we evaluated the relaxation trajectories using the same

way as for  $\theta_S = 0$  but we in addition averaged over the perturbations  $\delta\theta = \pm 0.1$ . The decay times  $\tau$  of the resulting averaged relaxation trajectories, determined as  $\theta(\tau) = \theta(0)/e$ , are plotted as dashed lines (numerics) and crosses (experiments) in Fig. S11C. The non-smoothness of the resulting functions is related to the non-smooth evolution of the relaxation trajectories. The large fluctuations of the experimental results close to the bifurcation point, where the relaxation time diverges, arise from finite measurement time (6s) per relaxation event.

## References

- [1] R. J. Kosinski, A literature review on reaction time, *Clemson University* **10**, 337–344 (2008).
- [2] T. Beatus, J. M. Guckenheimer, and I. Cohen, Controlling roll perturbations in fruit flies, *Journal of The Royal Society Interface* **12**, 20150075 (2015).
- [3] L. Ristroph, G. Ristroph, S. Morozova, A. J. Bergou, S. Chang *et al.*, Active and passive stabilization of body pitch in insect flight, *Journal of The Royal Society Interface* **10**, 20130237 (2013).
- [4] L. Ristroph, A. J. Bergou, G. Ristroph, K. Coumes, G. J. Berman *et al.*, Discovering the flight autostabilizer of fruit flies by inducing aerial stumbles, *Proceedings of the National Academy of Sciences* **107**, 4820–4824 (2010).
- [5] H. Pomeroy and F. Heppner, Laboratory determination of startle reaction time of the starling (*Sturnus vulgaris*), *Animal Behaviour* **25**, 720–725 (1977).
- [6] R. C. Eaton, R. A. Bombardieri, and D. L. Meyer, The Mauthner-initiated startle response in teleost fish, *Journal of Experimental Biology* **66**, 65–81 (1977).
- [7] R. C. Eaton, *Neural mechanisms of startle behavior* (Springer Science & Business Media, 1984).
- [8] P. Lenz and D. Hartline, Reaction times and force production during escape behavior of a calanoid copepod, *Undinula vulgaris*, *Marine Biology* **133**, 249–258 (1999).
- [9] J. E. Segall, S. M. Block, and H. C. Berg, Temporal comparisons in bacterial chemotaxis, *Proceedings of the National Academy of Sciences of the United States of America* **83**, 3024160[pmid], 8987–8991 (1986).
- [10] M. Fränzl, S. Muiños-Landin, V. Holubec, and F. Cichos, Fully Steerable Symmetric Thermoplasmonic Microswimmers, *ACS Nano* **15**, 3434–3440 (2021).
- [11] U. Khadka, V. Holubec, H. Yang, and F. Cichos, Active Particles Bound by Information Flows, *Nat. Commun.* **9**, 3864 (2018).
- [12] S. Muiños-Landin, A. Fischer, V. Holubec, and F. Cichos, Reinforcement learning with artificial microswimmers, *Science Robotics* **6** (2021).
- [13] D. Geiss, K. Kroy, and V. Holubec, Brownian molecules formed by delayed harmonic interactions, *New Journal of Physics* **21**, 093014 (2019).
- [14] T. Insperger, On the Approximation of Delayed Systems by Taylor Series Expansion, *Journal of Computational and Nonlinear Dynamics* **10**, 024503 (2015).
- [15] P. Hänggi, P. Talkner, and M. Borkovec, Reaction-rate theory: fifty years after Kramers, *Rev. Mod. Phys.* **62**, 251–341 (1990).
- [16] I. Llopis and I. Pagonabarraga, Hydrodynamic interactions in squirmer motion: Swimming with a neighbour and close to a wall, en, *Journal of Non-Newtonian Fluid Mechanics* **165**, 946–952 (2010).
- [17] A. I. Campbell, S. J. Ebbens, P. Illien, and R. Golestanian, Experimental observation of flow fields around active Janus spheres, en, *Nature Communications* **10**, 3952 (2019).
- [18] B. I. Halperin and P. C. Hohenberg, Scaling Laws for Dynamic Critical Phenomena, *Phys. Rev.* **177**, 952–971 (1969).



---

# Active particles with delayed attractions form quaking crystallites

PIN-CHUAN CHEN<sup>1</sup>, KLAUS KROY<sup>1</sup>, FRANK CICHOS<sup>2</sup>, XIANGZUN WANG<sup>2</sup> and VIKTOR HOLUBEC<sup>3</sup>

<sup>1</sup> *Institute for Theoretical Physics - Universität Leipzig, 04103 Leipzig, Germany [chen@itp.uni-leipzig.de](mailto:chen@itp.uni-leipzig.de)  
[klaus.kroy@uni-leipzig.de](mailto:klaus.kroy@uni-leipzig.de)*

<sup>2</sup> *Molecular Nanophotonics Group, Peter Debye Institute for Soft Matter Physics - Universität Leipzig, 04103 Leipzig, Germany [cichos@physik.uni-leipzig.de](mailto:cichos@physik.uni-leipzig.de)*

<sup>3</sup> *Department of Macromolecular Physics, Faculty of Mathematics and Physics, Charles University - 18000 Prague, Czech Republic [viktor.holubec@mff.cuni.cz](mailto:viktor.holubec@mff.cuni.cz)*

**Abstract** – Perception-reaction delays have experimentally been found to cause a spontaneous circling of microswimmers around a targeted center. Here we investigate the many-body version of this experiment with Brownian-dynamics simulations of active particles in a plane. For short delays, the soft spherical discs form a hexagonal colloidal crystallite around a fixed target particle. Upon increasing the delay time, we observe a bifurcation to a chiral dynamical state that we can map onto that found for a single active particle. The different angular velocities at different distances from the target induce shear stresses that grow with increasing delay. As a result, tangential and, later, also radial shear bands intermittently break the rotating crystallite. Eventually, for long delays, the discs detach from the target particle to circle around it near the preferred single-particle orbit, while spinning and trembling from tidal quakes.

---

**Introduction.** – Recent experiments with synthetic microswimmers steered toward a fixed target have revealed a spontaneous vortex formation caused by a perception-reaction delay [1]. The observed phenomenology can be attributed to the delay-induced aiming errors, akin to those associated with microswimmer navigation strategies employing “vision-cone” [2, 3] or “acceptance-angle” [4, 5] criteria. The experiment thereby established a simple paradigmatic model system for swarm forming ensembles with delayed interactions. Notably, the response of all living creatures and artificial devices to external stimuli is delayed by the time required to transfer and process information and realize the required response. All these systems can be classified as feedback-driven systems [6], which are well-studied in control theory, an engineering branch of dynamical-systems theory. In physics, objects capable of active reactions to perceived stimuli, such as animals or robots, are commonly studied within the field of active matter [7]. Even though the models of active matter usually neglect perception-reaction delays, it was shown in several pioneering studies that delays can significantly impact stability, dynamical phases, and even finite-size scaling in active matter systems [8–16].

In this Letter, we extend the experimental model system of Ref. [1] to system sizes that are currently inaccessible to the experimental techniques employed in [17]. Using

Brownian dynamics simulations, we find that the average angular velocity of the system still exhibits the bifurcation described in [1], but that the many-body dynamics undergoes a surprisingly rich series of delay-induced dynamical phase transitions. For short delays, the system forms a densely packed crystallite around the target, which can be interpreted as a variant of motility-induced phase separation [18], with a strongly depleted gas phase. As the delay increases, the crystallite is intermittently broken up by delay-induced shear bands.

Even for experimentally realistic noise intensities, the phenomenology observed in our simulations resembles the behavior of sheared low-temperature colloidal suspensions or athermal granular materials [19, 20]. An important feature of densely packed crystalline and amorphous particle assemblies is that they can only be sheared if the packing structure is somewhat dilated to allow the particles to escape from their nearest-neighbor cages and move around each other. A typical defect structures observed under such conditions are therefore shear bands [21, 22]. In the field of granular rheology, one also speaks of the dilatancy effect. It is responsible for normal stresses and the non-affine response to shear. In everyday life, you may experience it in the form of drained halos around your feet when you step on wet sand. In contrast to common granular and colloidal rheological setups, the shear stresses in our

active-Brownian-particle ensembles are however not induced by a moving background solvent or a system boundary or immersed probe particle, but solely by the individual particles' activity, itself. This entails some counter-intuitive consequences. Most importantly, the time delay only entails relevant navigational aiming errors if the particles are actually moving, but not if they are jammed up in a dense cluster. This somewhat unconventional property distinguishes our setup from the myriad of superficially related rheological problems documented in the literature. It also impedes attempts to provide a complete mechanistic interpretation of the unique succession of dynamical phases and phase transitions, described in the following.

**Model.** – We consider a two-dimensional system of  $N$  overdamped active Brownian discs, interacting via soft steric interactions. One particle is held fixed at the origin. The  $N$  mobile particles aim to swim toward it with a constant speed  $v_0$ . As shown in Fig. 1a, they cannot react instantaneously to the detected target position, but only after a certain delay time  $\delta t$ . Since the particles keep moving during this time, the resulting retarded attraction to the central target acquires important aiming errors.

We fix length and time scales by setting the particle diameter and the swim speed to unity. The dimensionless position vector  $\mathbf{r}_i$  of the  $i$ th Brownian particle obeys the Langevin equation

$$\dot{\mathbf{r}}_i(t) = \mathbf{F}_i(t) + k \sum_{j \neq i} \mathbf{r}_{ij}(t) \Theta[1 - |\mathbf{r}_{ij}(t)|] + \sqrt{2D} \boldsymbol{\eta}_i(t), \quad (1)$$

where  $\mathbf{F}_i(t) = -\mathbf{r}_i(t - \delta t)/|\mathbf{r}_i(t - \delta t)|$  are the intended (or nominal) velocities of the individual particles, and  $t$  is the dimensionless time. The soft steric repulsion has a strength of given by the dimensionless stiffness  $k$  and a range cutoff at  $|\mathbf{r}_{ij}(t)| = 1$ , imposed by the Heaviside  $\Theta$  function. The diffusivity  $D$  controls the intensity of mutually independent Gaussian white noise vectors  $\boldsymbol{\eta}_i$ ,  $i = 1, \dots, N$ , with zero mean,  $\langle \boldsymbol{\eta}_i(t) \rangle = \mathbf{0}$ , and covariance  $\langle [\boldsymbol{\eta}_i(t)]_x [\boldsymbol{\eta}_j(t')]_y \rangle = \delta_{ij} \delta_{xy} \delta(t - t')$ .

We studied the model for particle numbers  $N = 15 \dots 1000$  that are neither analytically tractable nor currently realizable in experiments. The dynamical equations are solved by Brownian dynamics simulations with time step  $dt = 0.001$ ,  $k = 101.4$ , and  $D = 0.0136$ . These parameters are motivated by typical experimental conditions in aqueous solvents at room temperature, if one identifies the particle diameter with  $2.19 \times 10^{-6}$  m and the propulsion speed with  $2.16 \times 10^{-6}$  m/s [1]. We initialized the particles randomly around the origin, let them diffuse for a time  $t = \delta t$ , and simulated long enough such that the system relaxed to a steady state (see the supplementary videos SM). Afterward, we continued the simulation and collected the data. Varying  $k$  and  $D$  in the dynamical equations (1) within an experimentally reasonable range does not change the qualitative results. Hence, the relevant control parameters are the time de-

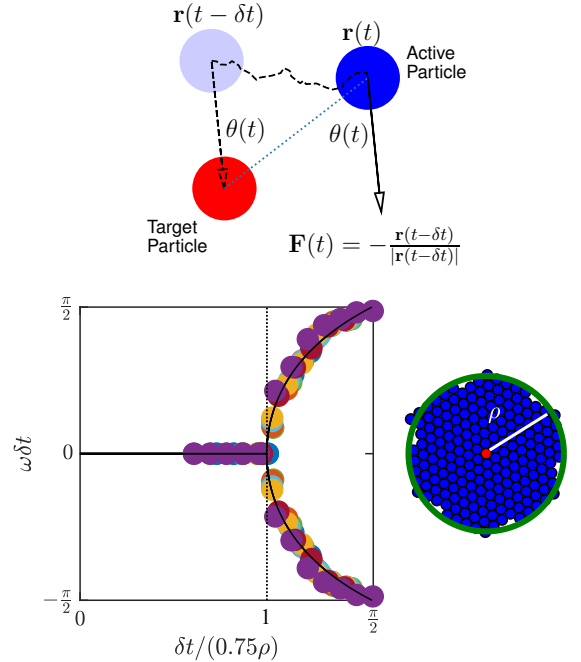


Fig. 1: Active Brownian particles (blue), swimming at constant speed, aim at a central target (red) with a perception-reaction delay  $\delta t$ . a) That the actual swim direction at time  $t$  is determined by position  $\mathbf{r}(t - \delta t)$  at the earlier time  $t - \delta t$ , gives rise to aiming errors and ensuing dynamical phases. b) The bifurcation diagram shows the average angular displacement  $\omega \delta t$  per delay time  $\delta t$ . Upon increasing delay, the isotropic static phase (I) gives way to radially symmetric chiral phases (II-IV). For the yin-yang and blob phases (V and VI),  $\omega \delta t \approx \pi/2$ . The colors code for various particle numbers  $N = 30 \dots 1000$ . The diagram obtained for athermal motion (diffusivity  $D = 0$ ) remains unchanged for an experimentally realistic noise intensity ( $D = 0.0136$ ). c) Close-packed crystallites of  $N \gtrsim 30$  particles have radius  $\rho = \sqrt{(N + 1)/3.62}$  in units of the (soft-)particle diameter.

lay  $\delta t$  and particle number  $N$ , or the corresponding radius  $\rho(N) = \sqrt{(N + 1)/3.62}$  of a close packed hexagonal crystallite (see Fig. 1c).

**Bifurcation.** – As shown in Ref. [1], for  $N = 1$ , the average angular velocity  $\omega$  of the single active Brownian particle around the fixed target is determined by a transcendental self-consistency equation. If the active-particle and target diameters are set to unity, it takes the form of the “sine map”  $\omega = \sin(\omega \delta t)$ . It exhibits a bifurcation from  $\omega = 0$  to  $\omega \neq 0$  at  $\delta t = 1$  (or, in the dimensional units of Ref. [1],  $v_0 \delta t = 2a$ ). For  $1 < \delta t < \pi/2$ , the single active Brownian particle “slides” around the target, and thus its dimensionless orbit radius is close to 1. When  $\delta t > \pi/2$ , swimmer and target particle lose touch and the circular orbit “takes off”. Its radius  $R = 2\delta t/\pi$  is now determined by the condition that the angular displacement of the particle

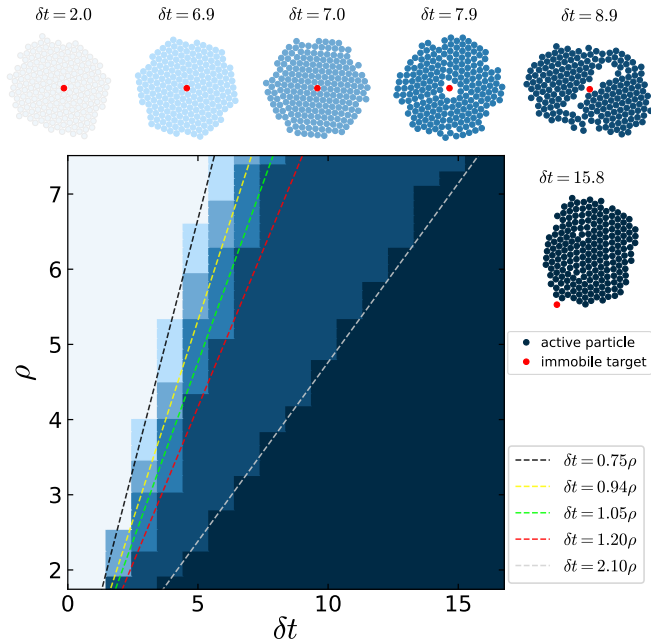


Fig. 2: Dynamic phase diagram. Like the preferred single-particle orbit  $R$ , the (binned) crystallite radius  $\rho$  grows with increasing delay time  $\delta t$ . We distinguish phases with a (I) static, (II) spinning, and (III) quaking crystallite, and a (IV) ring, (V) yin-yang/blobs, and (VI) satellite, respectively. Notice the appearance of predominantly concentric (III), radial (IV) and criss-crossing (V-VI) shear bands that intermittently break the crystallite, giving rise to a staircase-like increase of the shear strain  $\Gamma(t)$  (3rd row of Fig. 3), for all but the first two phases.

per one delay time,  $\omega\delta t$ , is  $\pi/2$ . In other words, for large delay times, the particle always propels tangentially (at a right angle) to the target, corresponding to a self-selected circular orbit.

Though not accessible experimentally, Ref [1] already demonstrated by Brownian dynamics simulations that the single-particle bifurcation diagram stays meaningful for many particles up to  $N = 100$ . The increased particle number actually stabilizes the spontaneously chosen sense of rotation against Brownian fluctuations, rendering the transient chiral symmetry breaking quasi permanent. More importantly, the increase in particle number merely renormalizes the bifurcation diagram. As shown in Fig. 2b, the average angular particle displacement  $\omega\delta t$  around the fixed target particle for  $N$  ranging from 30 to 1000 indeed falls on a single master curve, if plotted against  $\delta t/(0.75\rho)$ , corresponding to the renormalized sine map  $0.75\rho\omega = \sin(\omega\delta t)$ . The bifurcation curve coincides with that of a single large quasi-particle of radius  $(0.75\rho - 0.5)$ , rotating around the target particle of radius 0.5. In other words, the minimum radius for active rotation (originally given by the particle diameter) equals  $0.75\rho$ , in the many-body case. One can speculate that the effective radius  $0.75\rho$  could coincide with the crystallite’s

radius of gyration  $\int_0^\rho dx 2\pi x^2 / (\pi\rho^2) \approx 0.67\rho$ . This is indeed not far off, although the data is more suggestive of a matching of the radius  $R = 2\delta t/\pi$  of the optimal single-particle orbit with  $\rho/2$ . This could suggest that spinning starts when the preferred nominal velocity components of the particles inside and outside the optimal orbit cancel out. The difficulty with such interpretation is that the actually measured nominal velocity field created by the highly frustrated active particles in the bulk of a solid crystallite is, for the relevant delays, still purely central.

As shown in Fig. 2 for particle numbers  $N = 15 \dots 200$ , when the delay time  $\delta t$  is increased, the particle ensemble experiences a series of abrupt dynamical changes, thereby evolving from a static hexagonal crystallite to a continuously breaking elliptic satellite droplet, circling around the target on an orbit close to that preferred by a single active particle. Intriguingly, the average angular velocity in all these phases obeys the effective single-particle theory well. In fact, the single-particle theory can be used as a starting point for understanding most of the features of the various dynamical phases of the many-body model.

**Order parameters.** – To distinguish between the six dynamical phases in Fig. 2, we introduce the following three order parameters.

- the radial distribution  $p(r)$  (the probability density to find an active particle at distance  $r$  from the targeted center), normalized as  $2\pi \int_0^\infty dr r p(r) = 1$ .
- the (absolute) angular velocity  $\omega(r)$  of concentric shells of width 0.14, given by  $r^2\omega \equiv |\langle \mathbf{r}_i \times \dot{\mathbf{r}}_i \rangle|_{|\mathbf{r}_i| \approx r}$ .
- the cumulative shear strain  $\Gamma(t) = \int_0^t dt' |\dot{\Gamma}(t')|$  around a representative bulk particle at time  $t$ . Formally, the shear rate is defined as  $\dot{\Gamma} = (\partial v_x / \partial y + \partial v_y / \partial x) / 2$ , where  $v_{x,y}$  denote Cartesian components of the velocity field. As a proxy for our particulate system, we use

$$\dot{\Gamma}(t) = \frac{1}{2} \sum_j \left( \frac{\dot{x}_i(t) - \dot{y}_j(t)}{y_i(t) - y_j(t)} + \frac{\dot{y}_i(t) - \dot{x}_j(t)}{x_i(t) - x_j(t)} \right) \quad (2)$$

The sum runs over nearest-neighbor shell particles  $j$  that are less than  $\sqrt{2}$  away from a selected bulk particle  $i$ . To obtain the time derivatives of the components  $x_i(t)$  and  $y_i(t)$  of the position vector  $\mathbf{r}_i(t)$ , we average Eq. (1) over 200 simulation time steps. Spurious coordinate singularities are regularized by discarding terms with denominators smaller than 0.05.

**Dynamical phases.** – As shown in Fig. 3, each of the dynamical phases differs from the other five in the qualitative behavior of at least one of the characteristics  $p(r)$ ,  $\omega(r)$ , and  $\Gamma(t)$ . The figure also shows the average radial and tangential projections of the nominal velocities (or “forces”)  $\mathbf{F}_i - \dot{\mathbf{R}}_0$  of particles in the co-moving frame at a given distance from the center of mass  $\mathbf{R}_0 = \sum_{i=1}^N \mathbf{r}_i / N$  of the system. The average radial projection  $F_r(r) \equiv \langle (\mathbf{F}_i - \dot{\mathbf{R}}_0) \cdot (\mathbf{r}_i - \mathbf{R}_0) / r \rangle_{|\mathbf{r}_i - \mathbf{R}_0| \approx r}$  can be interpreted as a “shell pressure”. In the depicted average tangential component  $F_\phi(r) \equiv \langle |(\mathbf{F}_i - \dot{\mathbf{R}}_0) \times (\mathbf{r}_i - \mathbf{R}_0) / r \rangle_{|\mathbf{r}_i - \mathbf{R}_0| \approx r} - |\omega|r$

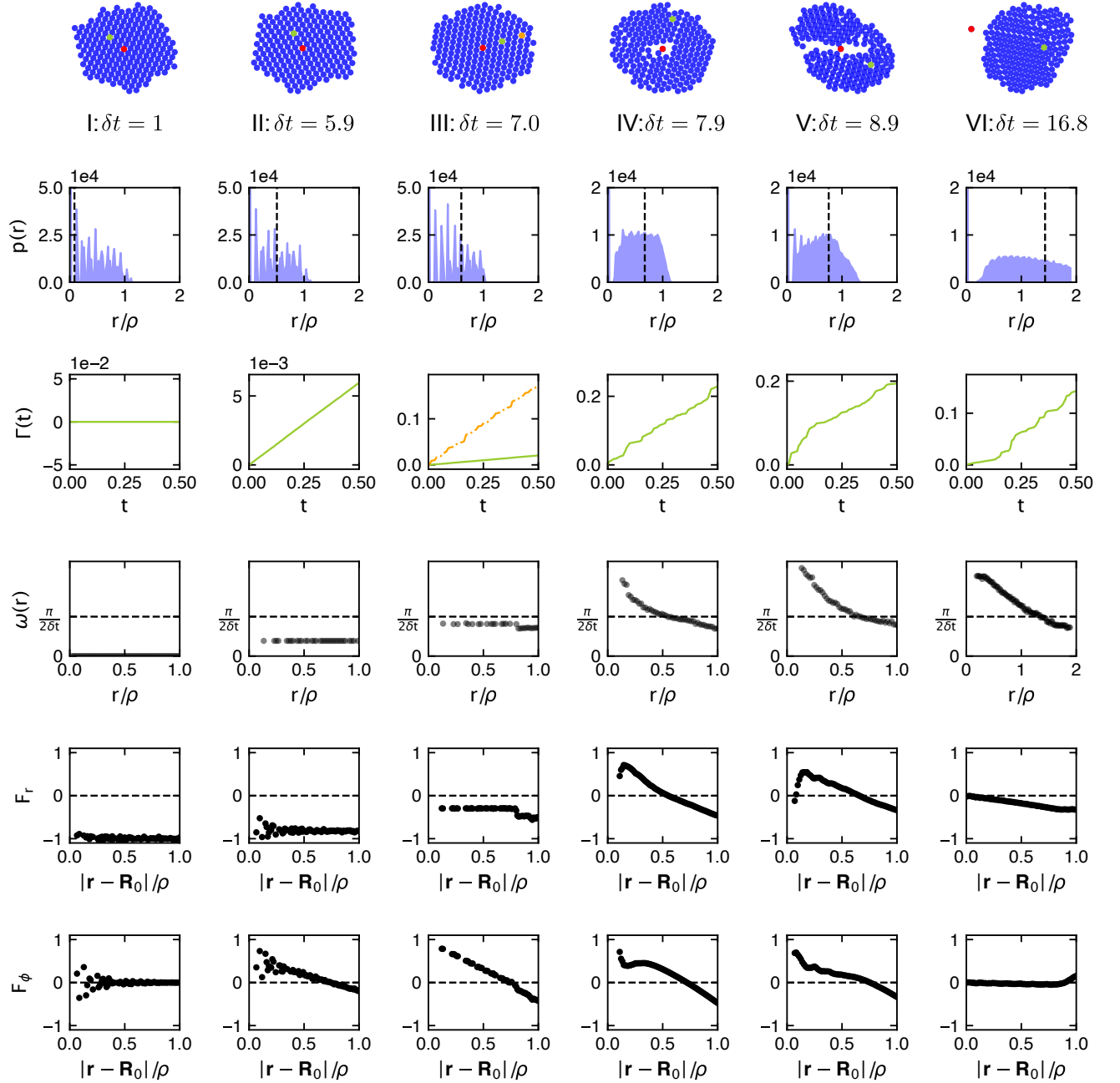


Fig. 3: Crystallite configurations and their shell angular velocities  $\omega(r)$  and the accumulated shear strains  $\Gamma(t)$  for selected bulk particles (green and orange), as caused by the radial and tangential “forces” or nominal swim velocities  $F_r$ ,  $F_\phi$  in the co-moving, co-spinning frame. The corresponding force fields are shown in SM Fig. S1. The dynamical phases I to VI of Fig. 2 were simulated for vanishing thermal noise  $D = 0$  and  $N = 199$  particles (corresponding to  $\rho \approx 7.43$  if close-packed). Vertical dashed lines in the radial distribution functions  $p(r)$  indicate the preferred single particle orbit radius  $R = 2\delta t/\pi$ .

we also subtracted the part responsible for the crystallite’s overall solid body rotation to improve the visibility of what can then be interpreted as a tangential shear stress. While the compression of the cluster by  $F_r$  mostly helps to maintain its crystalline structure, the tangential stress  $F_\phi$

imposes unequal torques on the concentric particle shells, thereby inducing the tangential shear bands and breakup events seen in the phases III-VI. It is noteworthy that, due to the normalization of the nominal velocities  $|\mathbf{F}_i| = 1$ , these two competing tendencies are not independent in our



system. Upon increasing the delay time  $\delta t$ , the nominal velocities increasingly tilt away from the central direction, meaning that the pressure decreases while the shear increases, at the same time, aggravating the destabilization. Also, for the non compact crystallites forming for longer delays, the radial forces themselves may cause radial shear bands and additionally contribute to the breaking of the crystalline configuration.

In the following, we characterize the individual dynamical phases in greater detail. The best intuitive insight into their dynamic nature is gathered from the corresponding videos in the supplementary material (SM).

(I) *Static crystallite*:  $\delta t \lesssim 0.75\rho$ . For short delays  $\delta t$ , the active Brownian particles are propelled exactly toward the target by their nominal velocities, as shown in last two lines of Figs. 3 and S1. Due to the steric repulsion, they form a non-rotating densely packed hexagonal crystallite (with small Brownian fluctuations). Its radial distribution function  $p(r)$  resembles that of close-packed hard discs, while the dynamical order parameters  $\omega$  and  $\dot{\Gamma}$  vanish.

(II) *Spinning crystallite*:  $0.75\rho < \delta t \lesssim 0.94\rho$ . Upon increasing the delay time  $\delta t$  beyond the threshold  $0.75\rho$ , the crystallite exhibits solid body rotation around the target particle. The order parameters thus remain the same as in phase I, with the exception that  $\omega(r) = \omega$  is given by a nonzero constant that is accurately predicted by the single-particle theory. However, as the particles' propulsion speed is fixed to 1, the particles closer to the target would individually prefer to move with larger angular velocities than those further away, while a constant  $\omega(r)$  is enforced by the steric interactions and the radial pressure exerted by the particles in the periphery, which still predominantly aim at the center. These features are nicely reflected in the radial and tangential projections of the nominal velocity in Fig. 3 and the nominal velocity field in Fig. S1 of the SM. Notice that the nominal tangential velocities of particles near the target/periphery are larger/smaller than  $\omega r$ , which induces the tangential shear stresses that attempt to break up the crystallite.

(III) *Quaking crystallite*:  $0.94\rho \lesssim \delta t \lesssim 1.05\rho$ . The tangential shear stresses caused by inhomogeneous angular velocity  $\omega(r)$  grow with increasing time delay. At  $\delta t \approx 0.94\rho$  they overcome the compressive forces and create shear bands. As shown in Fig. 3, the inner particles rotate (almost) at the optimal single-particle angular velocity  $\pi/(2\delta t)$ . The periphery lags behind, intermittently detaching and sliding around the rotating core (see the snapshots of the system in Figs. 3 and 4, and SM video 2). These stick-slip events cause a staircase-like increase of the shear strain  $\Gamma(t)$  (not observed around bulk particles that are not part of a shear band), and can be interpreted as quakes of the outer shell.

The last two rows of Fig. 3 and Fig. S1 of the SM moreover indicate that the nominal velocities of particles along radial rays from the center are no longer parallel. Closer to the center they have larger tangential components than

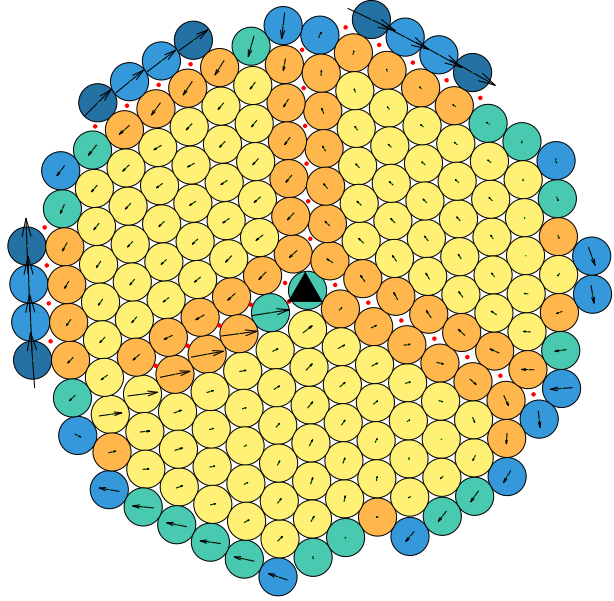


Fig. 4: Snapshot of the SM video S3 ( $\delta t = 7.1$ ,  $N = 199$ , phase III). Particle color codes for the number of nearest neighbors (from 2 to 6: steel blue, sky blue, aquamarine, orange and yellow). Red dots mark shear bands. The arrows show the actual velocities of the particles in the co-moving, co-rotating frame. The black triangle depicts the system's center of mass, which here overlaps with the central target particle. The meanings of the symbols in all videos of the SM are the same.

at the periphery, creating a pressure imbalance in the system. One can interpret this as a result of the tendency of the particles to propel toward the optimal single-particle orbit, which expands with increasing  $\delta t$ , as indicated in the  $p(r)$  panels. Upon increasing the delay somewhat beyond the value  $\delta t \approx 0.94\rho$ , for which the tangential shear bands appear (e.g., from  $\delta t = 7$  to  $\delta t = 7.1$  for  $N = 200$ ), the corresponding pressure imbalance eventually also causes the formation of system-spanning radial shear bands. Once a single radial band is formed, it destabilizes the next neighbor shell around the immobile target particle, which nucleates two more bands by the dilatancy effect, as shown in Fig. 4. The angles between the three bands are  $2\pi/3$ , corresponding to three equally sized fragments. Along the shear bands, particles slide in opposite directions (see the videos in the SM).

(IV) *Ring*:  $1.05\rho \lesssim \delta t \lesssim 1.20\rho$ . The single particle theory predicts that the outermost layer of the crystallite would start to rotate by itself (i.e., even if the rest of the crystallite was fixed), at  $\delta t = \rho$ . Some of the pressure onto the crystalline core is thereupon released, which facilitates its “breathing” due to the dilatancy effect. The core particles can then follow more freely their tendency to approach the optimal single particle orbit, thereby creating an outward pressure (Fig. S1). As a result, the crystal-

lite detaches from the central target particle and forms a ring, in which particles inside and outside the optimal orbit converge toward it. The crystalline structure is then no longer compressed only from the outside but also from the inside. The corresponding stresses increase the frequency of quakes and tangential and radial shear-band formation throughout the ring, as witnessed by  $\Gamma(t)$  (Fig. 3). The associated repeated breaking and healing effectively melts the crystalline structure as is reflected in the monotonic decay of the angular velocity  $\omega(r)$  with increasing distance  $r$  from the target and the loss of structure in the radial distribution function. Both effects are somewhat moderated within the fragments forming after the permanent breakup of the ring into the yin-yang shape, described next.

(V) *Yin-yang/blobs*:  $1.2\rho \lesssim \delta t \lesssim 2.1\rho$ . The effective contractile force due to the inward-outward pressures described above for the ring structure destabilize the ring in a manner similar to the capillary forces in a Plateau-Rayleigh instability [23]. It therefore tends to break up into  $2\pi l/2l = \pi \approx 3$  equally sized fragments, where  $2l$  is the ring width. Due to the (essentially) athermal conditions, the exact features of the breakup depend on initial conditions, as seen in videos 8-10 of the SM. The contractile forces towards the optimal orbit also causes larger clusters to orbit more slowly than smaller ones. They contain particles further away from the optimal radius, pointing less along the orbital direction. This slows down larger fragments compared to smaller ones, so that smaller fragments will chase the larger ones, thereby giving rise to some coarsening.

One might therefore conclude that the many-body system would ultimately form a giant quasi-particle, centered on the optimal orbit. However, as long as the radius of the closely packed crystallite  $\rho$  is larger than the optimal orbit radius  $R = 2\delta t/\pi$ , such quasi-particle would constantly be damaged by the fixed target particle and therefore actually cannot form. As a result, coarsening is interrupted and the system instead forms a highly dynamical yin-yang structure where the yin part continuously “steals” particles from the yang part, and *vice-versa*. For larger delays, the yin (or the yang) component outgrows its partner until it hits the target particle. The steady state ultimately consists of a single cluster in contact with the target particle, surrounded by several sub-clusters traveling close to the optimal single-particle orbit. Also note that, due to their fixed speed, the particles in the fragments move with larger angular velocities the closer they are to the center (Fig. 3). Together with their tendency to propel towards the optimal orbit, this causes a retrograde spinning of the fragments around their own centers of mass. With respect to the order parameters depicted in Fig. 3, the yin-yang phase exhibits the same phenomenology as the ring phase.

To quantify the phase boundaries, we again resort to the bifurcation diagram in Fig. 1. It shows that the average angular displacement during one delay time,  $\omega\delta t$ , monotonically increases with  $\delta t$  up to  $\delta t = 0.75\rho\pi/2 \approx 1.18\rho$ ,

when it saturates at the value  $\omega\delta t = \pi/2$ . This is when a single active particle would detach from the fixed target particle, as its optimal orbit of radius  $R$  takes off. This suggests that the tendency to break the ring and form a single eccentric crystallite, centered on the optimal orbit, would start at  $\delta t > 1.18\rho$ , which is indeed close to the observed value  $1.2\rho$ , and would eventually succeed once the optimal orbit radius  $R$  exceeds  $\rho$ . At this point a spherical crystallite would no longer interfere with fixed target particle at the center. Why this estimate fails to provide the correct condition for the transition to the last dynamical phase is explained in the next paragraph.

(VI) *Satellite*:  $\delta t \gtrsim 2.1\rho$ . As pointed out in the preceding paragraph, one would expect to find a single compact satellite orbiting the target particle (roughly) on the optimal single-particle orbit, when  $R \approx \rho$ , hence  $\delta = \pi\rho/2$ , which is actually not the case. The discrepancy is caused by the fact that the satellite is actually not circular but somewhat elongated along an axis that is slightly tilted relative to the radial direction. The reason is that the pressure exerted by the individual particles is no longer radially symmetric (see Fig. S1).

The stick-slip motion of particles along the shear bands in this phase is somewhat reminiscent of an extreme version of the terrestrial tides caused by the motion of the Moon around Earth. The major difference is that the tidal forces correspond to an attraction rather than a repulsion relative to the satellite center. As a result, the quake dynamics is approximately out of phase by  $\pi/2$ , with respect to the Moon-Earth system (see video S5). Moreover, the attraction does not act toward the satellite center but toward the optimal single-particle orbit. And finally, the elongation of the crystallite is not perfectly aligned with the direction to the center, giving rise to another phase shift that depends on the precise model parameters.

Concerning the order parameters depicted in Fig. 3, the satellite phase again exhibits almost the same phenomenology as the ring state. The only difference is the radial distribution of particles, which is now much broader than in the other five phases. This is indicative of the destructive effect of the tidal quakes, which dynamically melt the crystallite into an effectively liquid droplet.

**Discussion and conclusions.** – We have numerically studied a two-dimensional ensemble of soft active Brownian particles steered toward a target particle with a time delay. The particles form a closely packed hexagonal crystallite around the target for small delay times and experimentally relevant noise intensities. However, with increasing delay, a much richer behavior is observed. The average angular velocity around the target exhibits a bifurcation, which can be mapped to the one found recently for a single active particle [1]. An interesting “plastic” deformation of the hexagonal crystalline structure ensues. The tangential and radial shear stresses grow with time delay, eventually creating shear bands and breaking the crystallite. Its overall shape changes with increasing delay

from a disc over a ring around the target to a yin-yang and eventually an elongated retrograde spinning satellite orbiting the target.

Our study demonstrates that simple time-delayed interactions can induce very complex dynamical behaviors in many body systems, even in the case of delayed attractions to a common fixed target. As time delays are omnipresent in interacting active matter systems in nature, this observation should be taken into account when interpreting experimental data. To this end, it would be interesting to realize the studied many-body system experimentally. In this case, hydrodynamic interactions between the active particles would play an important role and potentially give rise to somewhat different results as obtained above, for the idealized active Brownian particle system. Our essentially athermal dynamics might thereupon become more ergodic and fluid-like [21, 22, 24].

Our results could be extended in several other directions. First, one may consider attraction to a fixed position in space rather than to a fixed target particle. Our preliminary results with this setup reveal two major differences. Firstly, the minimum radius of rotation is determined by the noise, not by the particle diameters. Secondly, omitting the target particle increases the accessible state space. For example, the dynamics in the yin-yang/blobs phase (V) becomes much richer without the central particle, allowing for the appearance of a state with an almost deterministic periodically switching chirality. Of more practical interest might be the extension of our setup to an all-to-all attraction between the particles. Our preliminary results show that the phenomenology essentially remains unchanged, for short delay times. Differences appear for longer delays, where the emerging patterns are more symmetric compared to what we found above, and would deserve further study.

\* \* \*

**Acknowledgements.** – We gratefully acknowledge funding through DFG-GACR cooperation by the Deutsche Forschungsgemeinschaft (DFG Project No 432421051) and by the Czech Science Foundation (GACR Project No 20-02955J). VH additionally acknowledges the support of Charles University through project PRIMUS/22/SCI/009.

## REFERENCES

- [1] WANG X., CHEN P.-C., KROY K., HOLUBEC V. and CICHOS F., *Nature Communications*, **14** (2023) 56.
- [2] BÄUERLE T., LÖFFLER R. C. and BECHINGER C., *Nature Communications*, **11** (2020) 2547.
- [3] LÖFFLER R. C., BÄUERLE T., KARDAR M., ROHWER C. M. and BECHINGER C., *Europhysics Letters*, **134** (2021) 64001.
- [4] BREGULLA A. P., YANG H. and CICHOS F., *ACS Nano*, **8** (2014) 6542.

- [5] SELMKE M., KHADKA U., BREGULLA A. P., CICHOS F. and YANG H., *Phys. Chem. Chem. Phys.*, **20** (2018) 10521.
- [6] BECHHOEFER J., *Rev. Mod. Phys.*, **77** (2005) 783.
- [7] RAMASWAMY S., *Journal of Statistical Mechanics: Theory and Experiment*, **2017** (2017) 054002.
- [8] SUN Y., LIN W. and ERBAN R., *Physical Review E*, **90** (2014) 062708.
- [9] GIUGGIOLI L., MCKETTERICK T. J. and HOLDERIED M., *PLOS Computational Biology*, **11** (2015) e1004089.
- [10] MIJALKOV M., MCDANIEL A., WEHR J. and VOLPE G., *Physical Review X*, **6** (2016) 011008.
- [11] LEYMAN M., OGE MARK F., WEHR J. and VOLPE G., *Physical Review E*, **98** (2018) 052606.
- [12] SCHOLZ C., JAHANSHAHI S., LDOV A. and LÖWEN H., *Nature Communications*, **9** (2018) 1.
- [13] KHADKA U., HOLUBEC V., YANG H. and CICHOS F., *Nature Communications*, **9** (2018) 1.
- [14] PIWOWARCZYK R., SELIN M., IHLE T. and VOLPE G., *Physical Review E*, **100** (2019) 012607.
- [15] MUIÑOS-LANDIN S., FISCHER A., HOLUBEC V. and CICHOS F., *Science Robotics*, **6** (2021) .
- [16] HOLUBEC V., GEISS D., LOOS S. A. M., KROY K. and CICHOS F., *Physical Review Letters*, **127** (2021) 258001.
- [17] FRÄNZL M., MUIÑOS-LANDIN S., HOLUBEC V. and CICHOS F., *ACS Nano*, **15** (2021) 3434.
- [18] BIALKÉ J., LÖWEN H. and SPECK T., *Europhysics Letters*, **103** (2013) 30008.
- [19] LHERMINIER S., PLANET R., VEHEL V. L. D., SIMON G., VANEL L., MÅLØY K. J. and RAMOS O., *Phys. Rev. Lett.*, **122** (2019) 218501.
- [20] TSAI J.-C. J., HUANG G.-H. and TSAI C.-E., *Phys. Rev. Lett.*, **126** (2021) 128001.
- [21] SCHALL P. and VAN HECKE M., *Annual Review of Fluid Mechanics*, **42** (2010) 67.
- [22] WU Y. L., DERKS D., VAN BLAADEREN A. and IMHOF A., *Proceedings of the National Academy of Sciences*, **106** (2009) 10564.
- [23] MEHRABIAN H. and FENG J. J., *Journal of Fluid Mechanics*, **717** (2013) 281–292.
- [24] STEVENS M. J., ROBBINS M. O. and BELAK J. F., *Phys. Rev. Lett.*, **66** (1991) 3004.

# Supplementary Material

## Active particles with delayed attractions form quaking crystallites

Pin-Chuan Chen<sup>1</sup>, Klaus Kroy<sup>1</sup>, Frank Cichos<sup>2</sup>, Xiangzun Wang<sup>2</sup> and Viktor Holubec<sup>3</sup>

<sup>1</sup> *Institute for Theoretical Physics, Leipzig University, Postfach 100 902, 04009 Leipzig, Germany.*  
chen@itp.uni-leipzig.de, klaus.kroy@uni-leipzig.de

<sup>2</sup> *Peter Debye Institute for Soft Matter Physics, Molecular Nanophotonics Group, Universität Leipzig, 04103 Leipzig, Germany.*

cichos@physik.uni-leipzig.de, wangxiangzun@gmail.com

<sup>3</sup> *Department of Macromolecular Physics, Faculty of Mathematics and Physics, Charles University, 18000 Prague, Czech Republic*  
viktor.holubec@mff.cuni.cz

### Abstract

The supplementary information contains a figure showing the nominal velocity field for the individual dynamical phases and the description of the supplementary videos 1-10.

### 1 Nominal velocity fields

In the first row of Fig. S1, we show the individual particles' nominal velocities  $\mathbf{F}_i$  in the lab frame. The second row of the figure depicts projections of  $\mathbf{F}_i$  in the comoving, corotating frame to the radial direction from the system's center of mass:

$$\mathbf{F}_r^i = (\mathbf{F}_i - \dot{\mathbf{R}}_0) \cdot \frac{\mathbf{r}_i - \mathbf{R}_0}{|\mathbf{r}_i - \mathbf{R}_0|}, \quad (1)$$

where  $\mathbf{R}_0 = \sum_{i=1}^N \mathbf{r}_i / N$ . The third row of Fig. S1 presents the tangential projections corresponding to the radial ones in the second row minus the average rotation of the system. They were calculated as

$$\mathbf{F}_\phi^i = \mathbf{F}_\parallel^i - \omega |\mathbf{r}_i - \mathbf{R}_0| \frac{\mathbf{F}_\parallel^i}{|\mathbf{F}_\parallel^i|}, \quad (2)$$

where  $\mathbf{F}_\parallel^i = (\mathbf{F}_i - \dot{\mathbf{R}}_0) \cdot \mathbf{F}_r^i$ .

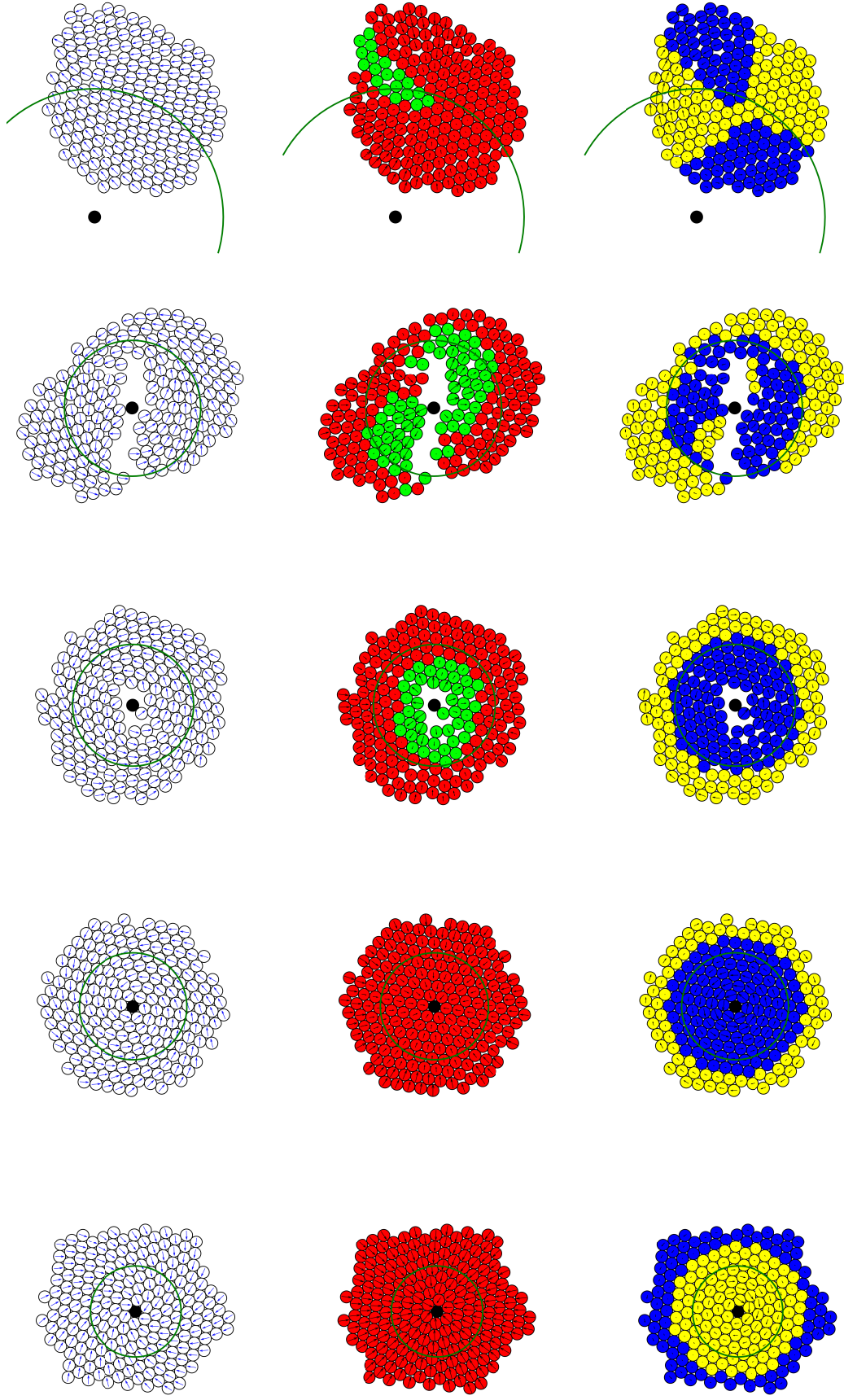


Figure S1: 1st row: nominal (intended) velocities  $\mathbf{F}_i(t)$  of the individual particles in the lab frame. The whole cluster rotates in the direction of the arrows. 2nd and 3rd row: nominal velocities in the co-moving, co-spinning frame projected on the radial and tangential directions, respectively. The colors mark the directions of the arrows (red-radial from the center of mass (COM), green-radial towards the COM, yellow-clockwise rotation around COM, and blue-counter-clockwise rotation around COM). The green circles depict the optimal single particle radius  $2\delta t/\pi$ . The black disc indicates the fixed target particle. Delay times  $\delta t$  corresponding to the individual columns are 5.9, 7, 7.9, 8.9, and 16.8, respectively.  $N = 200$ . The averaged values of  $F_r$  and  $F_{\theta}$  (2nd and 3rd row) as a function of distance to the COM are shown in the last two rows in the main text Fig. 3. In the non-rotating phase, which is not shown, the nominal velocities of all particles point to the center, as in the 1st panel of the second row.

## 2 Supplementary videos

The particle colors in the videos code for the number of their nearest neighbors (from 0 to 6: deep blue, purple, steel blue, sky blue, aquamarine, orange, and yellow). The shear bands are marked with red dots. The arrows indicate the actual velocities of the particles in the co-moving, co-rotating frame. The black triangle depicts the center of mass of the system.

To make the shear bands better visible, videos 1-6 and 8-9 were made with zero noise ( $D = 0$ ). Videos 7 and 10 show that nonzero noise ( $D = 0.0136$ ) makes the dynamics of the system more erratic without changing its qualitative features. Videos 1-7 were recorded after the system reached a steady state. Videos 8-10 show the whole time evolution of the system from the initial condition. In all the videos, we show  $N = 199$  particles, corresponding to  $\rho \approx 7.43$ . Except for the last three videos, all videos are sped up 3 times.

1.  $\delta t = 5.9$ , phase II: the spinning crystallite ( $D = 0$ ).
2.  $\delta t = 7$ , phase III: the quaking crystallite with tangential shear bands ( $D = 0$ ).
3.  $\delta t = 7.1$ , phase III: the quaking crystallite with tangential and radial shear bands ( $D = 0$ ).
4.  $\delta t = 7.9$ , phase IV: the ring ( $D = 0$ ).
5.  $\delta t = 8.9$ , phase V: the yin-yang/blobs ( $D = 0$ ).
6.  $\delta t = 16.8$ , phase VI: the satellite ( $D = 0$ ).
7.  $\delta t = 8.9$ , phase V: the yin-yang/blobs ( $D = 0.0136$ ).
8.  $\delta t = 8.9$ , phase V: the yin-yang/blobs. Typical relaxation trajectory to the yin-yang phase from a random initial condition with  $D = 0$ . The video is sped up 30 times.
9.  $\delta t = 8.9$ , phase V: the yin-yang/blobs. Another possible relaxation path to the yin-yang phase from a random initial condition with  $D = 0$ . The video is sped up 30 times.
10.  $\delta t = 8.9$ , phase V: the yin-yang/blobs. Typical relaxation path to the yin-yang phase from a random initial condition with nonzero noise intensity  $D = 0.0136$ . The video is sped up 30 times.

## ARTIFICIAL INTELLIGENCE

## Reinforcement learning with artificial microswimmers

S. Muiños-Landin<sup>1,2</sup>, A. Fischer<sup>1</sup>, V. Holubec<sup>3,4</sup>, F. Cichos<sup>1\*</sup>

Artificial microswimmers that can replicate the complex behavior of active matter are often designed to mimic the self-propulsion of microscopic living organisms. However, compared with their living counterparts, artificial microswimmers have a limited ability to adapt to environmental signals or to retain a physical memory to yield optimized emergent behavior. Different from macroscopic living systems and robots, both microscopic living organisms and artificial microswimmers are subject to Brownian motion, which randomizes their position and propulsion direction. Here, we combine real-world artificial active particles with machine learning algorithms to explore their adaptive behavior in a noisy environment with reinforcement learning. We use a real-time control of self-thermophoretic active particles to demonstrate the solution of a simple standard navigation problem under the inevitable influence of Brownian motion at these length scales. We show that, with external control, collective learning is possible. Concerning the learning under noise, we find that noise decreases the learning speed, modifies the optimal behavior, and also increases the strength of the decisions made. As a consequence of time delay in the feedback loop controlling the particles, an optimum velocity, reminiscent of optimal run-and-tumble times of bacteria, is found for the system, which is conjectured to be a universal property of systems exhibiting delayed response in a noisy environment.

## INTRODUCTION

Living organisms adapt their behavior according to their environment to achieve a particular goal. Information about the state of the environment is sensed, processed, and encoded in biochemical processes in the organism to provide appropriate actions or properties. These learning or adaptive processes occur within the lifetime of a generation, over multiple generations, or over evolutionarily relevant time scales. They lead to specific behaviors of individuals and collectives. Swarms of fish or flocks of birds have developed collective strategies adapted to the existence of predators (1), and collective hunting may represent a more efficient foraging tactic (2). Birds learn how to use convective air flows (3). Sperm have evolved complex swimming patterns to explore chemical gradients in chemotaxis (4), and bacteria express specific shapes to follow gravity (5).

Inspired by these optimization processes, learning strategies that reduce the complexity of the physical and chemical processes in living matter to a mathematical procedure have been developed (6). Many of these learning strategies have been implemented into robotic systems (7–9). One particular framework is reinforcement learning (RL), in which an agent gains experience by interacting with its environment (10). The value of this experience relates to rewards (or penalties) connected to the states that the agent can occupy. The learning process then maximizes the cumulative reward for a chain of actions to obtain the so-called policy. This policy advises the agent which action to take. Recent computational studies, for example, reveal that RL can provide optimal strategies for the navigation of active particles through flows (11–13), the swarming of robots (14–16), the soaring of birds (3), or the development of collective motion (17). The ability of how fish can harness the vortices in the

flow field of others for energy-efficient swimming has been explored (18). Strategies of how to optimally steer active particles in a potential energy landscape (19) have been explored in simulations, and deep Q-learning approaches have been suggested to navigate colloidal robots in an unknown environment (20).

Artificial microswimmers are a class of active materials that integrate the fundamental functionality of persistent directed motion, common to their biological counterparts, into a user-designed microscopic object (21). Their motility has already revealed insights into a number of fundamental processes, including collective phenomena (22–24), and they are explored for drug delivery (25) and environmental purposes (26). However, the integration of energy supply, sensing, signal processing, memory, and propulsion into a micrometer-sized artificial swimmer remains a technological challenge (27). Hence, external control strategies have been applied to introduce sensing and signal processing, yet only schemes with rigid rules simulating specific behaviors have been developed (28–31). Combining elements of machine learning and real-world artificial microswimmers would considerably extend the current computational studies into real-world applications for the future development of smart artificial microswimmers (32).

Here, we incorporate algorithms of RL with external control strategies into the motion of artificial microswimmers in an aqueous solution. While the learning algorithm is running on a computer, we control a real agent acting in a real world subjected to thermal fluctuations, hydrodynamic and steric interactions, and many other influences. In this way, it is possible to include real-world objects in a simulation, which will help to close the so-called reality gap, i.e., the difference of pure *in silico* learning and real-world machine learning even at microscopic length scales (27). Our experimental investigation thus goes beyond previous purely computational studies (3, 11–13, 20). It allows us to observe the whole learning process optimizing parameters, which are not accessible in studies of biological species, to identify the most important ingredients of the real dynamics and to set up more realistic, but still simple, models based on this information. It also provides a glimpse of the challenges of RL for objects at those length scales for future developments.

<sup>1</sup>Molecular Nanophotonics Group, Peter Debye Institute for Soft Matter Physics, Universität Leipzig, 04103 Leipzig, Germany. <sup>2</sup>AIMEN Technology Centre, Smart Systems and Smart Manufacturing–Artificial Intelligence and Data Analytics Laboratory, Pl. Cataboi, 36418 Pontevedra, Spain. <sup>3</sup>Institute for Theoretical Physics, Universität Leipzig, 04103 Leipzig, Germany. <sup>4</sup>Department of Macromolecular Physics, Faculty of Mathematics and Physics, Charles University, 18000 Prague, Czech Republic.

\*Corresponding author. Email: cichos@physik.uni-leipzig.de

## RESULTS

## Self-thermophoretic microswimmer

To couple machine learning with microswimmers, we used a light-controlled self-thermophoretic microswimmer with surface-attached gold nanoparticles (Fig. 1A and see the Supplementary Materials). For self-propulsion, the swimmer has to break the time symmetry of low Reynolds number hydrodynamics (33). This is achieved by an asymmetric illumination of the particle with laser light of 532-nm wavelength. It is absorbed by the gold nanoparticles and generates a temperature gradient along their surface, inducing thermo-osmotic surface flows and lastly resulting in a self-propulsion of the microswimmer suspended in water. The direction of propulsion is set by the vector pointing from the laser position to the center of the particle. The asymmetric illumination is maintained during the particle motion by following the swimmer's position in real time and steering the heating laser (see the Methods section below). As compared with other types of swimmers (28, 34, 35), this symmetric swimmer removes the time scale of rotational diffusion from the swimmer's motion and provides an enhanced steering accuracy (36, 37) (see the Supplementary Materials).

## Gridworld

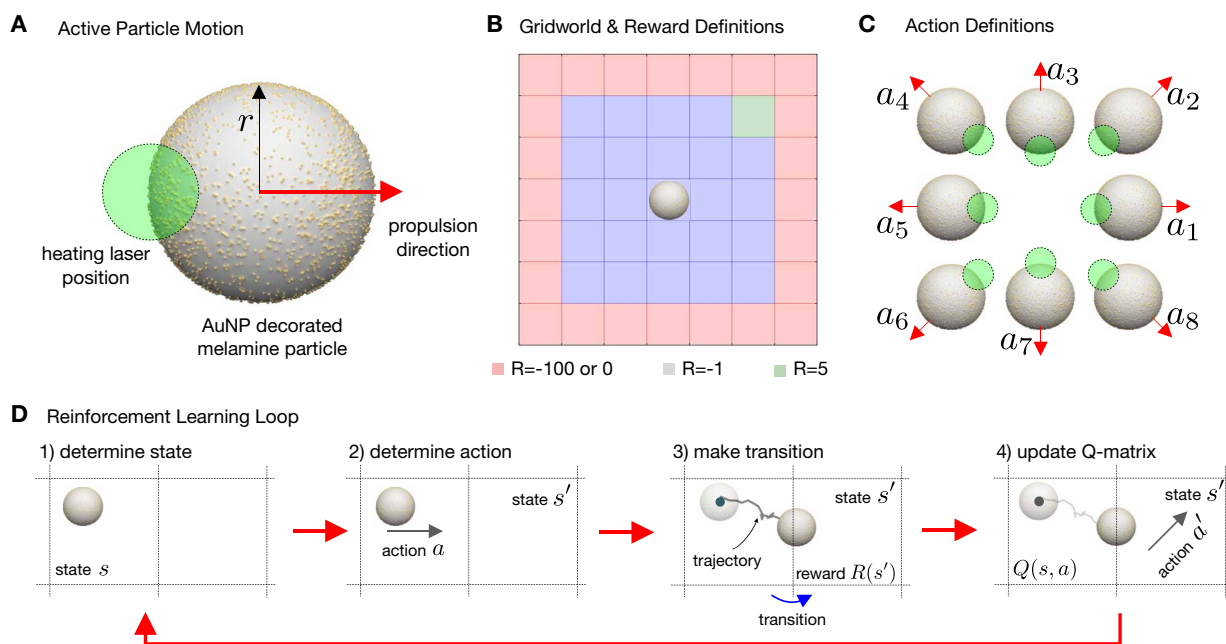
To show RL with a real-world microscopic agent, we refer to the standard problem of RL, the gridworld. The gridworld problem allows us to have an experimental demonstration while being able to access the problem numerically. We coarse grain a sample region of  $30\ \mu\text{m}$  by  $30\ \mu\text{m}$  into a gridworld of 25 states ( $5 \times 5$ ), each state

having a dimension of  $6\ \mu\text{m}$  by  $6\ \mu\text{m}$  (Fig. 1B). One of the states is defined as the target state (goal), which the swimmer is learning to reach. The gridworld is surrounded by 24 boundary states according to Fig. 1B. The obtained real-time swimmer position is used to identify the state  $s$  in which the swimmer currently resides. To move between states, we define eight actions  $a$ . The actions are carried out by placing the heating laser at the corresponding position on the circumference of the particle (see Fig. 1C). A sequence of actions defines an episode in the gridworld, which ends when the swimmer either leaves the gridworld to a boundary state or reaches the target state. During an episode, rewards or penalties are given. Specifically, the microswimmer gets a reward once it reaches the target state and a penalty in other cases (see the Supplementary Materials for details on the reward definitions). The reward function  $R$  thus only depends on the state  $s$ , i.e.,  $R = R(s)$ .

## RL implementation

We have implemented the model-free Q-learning algorithm to find the optimal policy that solves the navigation problem (38). The gained experience of the agent is stored in the Q-matrix (10), which tracks the utilities of the different actions  $a$  in each state  $s$ . When the swimmer transitions between two states  $s$  and  $s'$  (see the Supplementary Materials for details on the choice of the next state), the Q-matrix is updated according to

$$Q_{t+\Delta t}(s, a) = Q_t(s, a) + \alpha [R(s') + \gamma \max_{a'} Q_t(s', a') - Q_t(s, a)] \quad (1)$$



**Fig. 1. Gold nanoparticle–decorated microswimmer, states, and actions.** (A) Sketch of the self-thermophoretic symmetric microswimmer. The particles used have an average radius of  $r = 1.09\ \mu\text{m}$  and were covered on 30% of their surface with gold nanoparticles of about 10 nm diameter. A heating laser illuminates the colloid asymmetrically (at a distance  $d$  from the center), and the swimmer acquires a well-defined thermophoretic velocity  $\mathbf{v}$ . (B) The gridworld contains 25 inner states (blue) with one goal at the top right corner (green). A set of 24 boundary states (red) is defined for the study of the noise influence. (C) In each of the states, we consider eight possible actions in which the particle is thermophoretically propelled along the indicated directions by positioning the laser focus accordingly. (D) The RL loop starts with measuring the position of the active particle and determining the state. For this state, a specific action is determined with the  $\epsilon$  greedy procedure (see the Supplementary Materials for details). Afterward, a transition is made, the new state is determined, and a reward for the transition is given. On the basis of this reward, the Q-matrix is updated, and the procedure starts from step 1 until an episode ends by reaching the goal or exiting the gridworld to a boundary state.

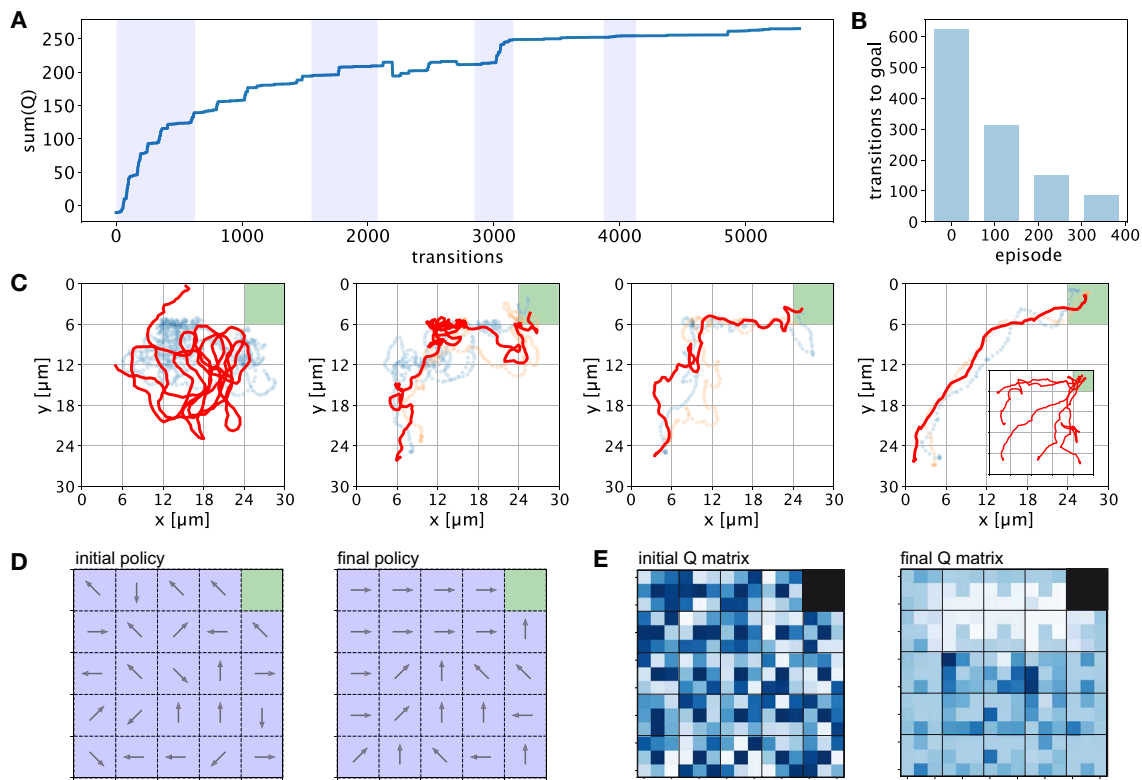


taking into account the reward  $R(s')$  of the next state, the utility of the next state  $Q_t(s', a')$  after taking the best action  $a'$ , and the current utility  $Q_t(s, a)$ . The influence of these values is controlled by two factors, the learning rate  $\alpha$  and the discount factor  $\gamma$ . The learning rate defines the fraction at which new information is incorporated into the Q-matrix, and the discount factor determines the value of future events into the learning process. The reward function is the only feedback signal that the system receives to figure out what it should learn. The result of this RL procedure is the optimal policy function  $\pi^*(s) \rightarrow a$ , which represents the learned knowledge of the system,  $\pi^*(s) = \operatorname{argmax}_a Q(s, a)$ ,  $Q(s, a) = \lim_{t \rightarrow \infty} Q_t(s, a)$ . Figure 1D highlights the experimental procedure of actuating the swimmer and updating the Q-matrix. As compared with computer models solving the gridworld with deterministic agents, there are four important differences to note. (i) The swimmer can occupy all positions within each state of  $6 \mu\text{m}$  by  $6 \mu\text{m}$  size. It can be arbitrarily close to the boundary. (ii) The swimmer moves in several steps through each state before making a transition. A swimmer velocity of  $v = 3 \mu\text{m s}^{-1}$  leads to a displacement of about  $6 \mu\text{m}$  within 2 s, corresponding to about 11 frames at an inverse frame rate  $\Delta t_{\text{exp}} = 180 \text{ ms}$  until a

transition to the next state is made. (iii) The new state after a transition does not have to be the state that was targeted by the actions. The microswimmers are subject to Brownian motion with a measured diffusion coefficient of  $D = 0.1 \mu\text{m}^2 \text{ s}^{-1}$ . The trajectory is therefore partially nondeterministic. With this respect, the system we consider captures a very important feature of active matter on small length scales that is inherent to all microscopic biological systems, where active processes have been optimized to yield robust functions in a noisy background. (iv) Due to a time delay in the feedback loop controlling the active particles, the action applied to the swimmer is not determined from its present position but from its position in the past, which is a common feature for all living and nonliving responsive systems.

### Learning process

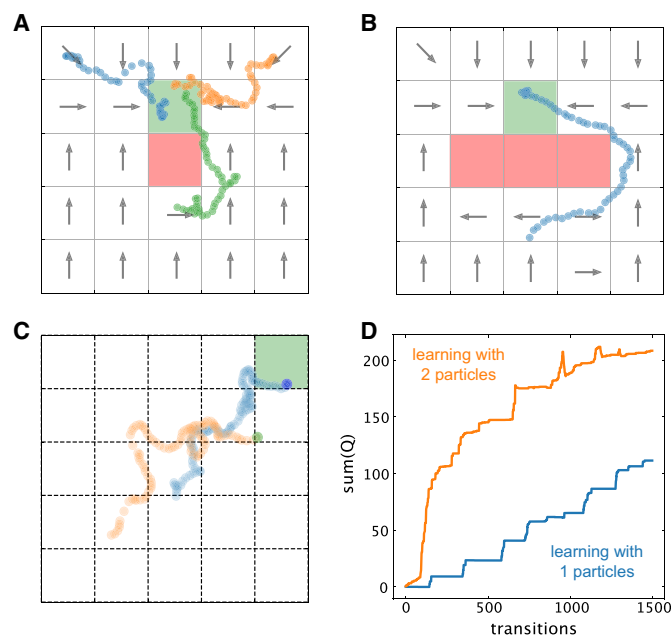
Figure 2 summarizes the learning process of our microswimmer for boundary states with  $R = 0$  and a velocity of  $v_{\parallel} = 3.0 \mu\text{m s}^{-1}$ ,  $v_{\parallel} = \langle \Delta \mathbf{r} \cdot \mathbf{e}_{\parallel} \rangle / \Delta t_{\text{exp}}$  where  $\langle \Delta \mathbf{r} \cdot \mathbf{e}_{\parallel} \rangle$  is the mean projected displacement of the swimmer along the direction of the action  $\mathbf{e}_{\parallel}$ . Over the course of more than 5000 transitions (more than 400 episodes, about 7 hours



**Fig. 2. Single microswimmer learning.** (A) Learning progress for a single microswimmer in a gridworld at a velocity of  $v_{\parallel} = 3.0 \mu\text{m s}^{-1}$ . The progress is quantified by the sum of all Q-matrix elements at each transition of the learning process. The Q-matrix was initialized randomly. The shaded regions denote a set of 25 episodes in the learning process, where the starting point is randomly chosen. (B) Mean number of steps required to reach the target when starting at the lower left corner as the number of the learning episodes increases. (C) Different examples of the behavior of a single microswimmer at different stages of the learning process. The first example corresponds to a swimmer starting at the beginning of the learning process at an arbitrary position in the gridworld. The trajectory is characterized by a large number of loops. With an increasing number of learning episodes, the trajectories become more persistent in their motion toward the goal. This is also reflected by the decreasing average number of steps taken to reach the goal [see (B)]. The inset in the rightmost graph reveals trajectories from different starting positions. (D) Policies  $\pi(s) = \operatorname{argmax}_a Q_t(s, a)$  defined by the Q-matrix before ( $Q_t(s, a) = Q_0(s, a)$ ) and after ( $Q_t(s, a) = Q(s, a)$ ) the convergence of the learning process. (E) Color representation of the initial and the final Q-matrix for the learning process. The small squares in each state represent the utility of the corresponding action (same order as in Fig. 1C) given by its Q-matrix entry, except for the central square. Darker colors show smaller utility, and brighter colors show a better utility of the corresponding action.

of experiment), the sum of all Q-matrix entries converges (Fig. 2A). During this time, the mean number of transitions to reach the goal state decreases from about 600 transitions to less than 100 transitions (Fig. 2B). Accordingly, the trajectories of the swimmer become more deterministic, and the swimmer reaches the goal state independent of the initial state (Fig. 2C and inset). As a result of the learning process, the initial random policy is changing into a policy driving the swimmer toward the goal state. In this respect, the final policy provides an effective drift field with an absorbing boundary at the goal state (Fig. 2D). During this process, which correlates the actions of neighboring cells, the average projected velocity  $v_{\parallel}$  causing the drift toward the goal also increases. Although the obtained policy is reflecting the best actions only, the Q-matrix shown in Fig. 2E provides the cumulative information that the swimmer obtained on the environment. It delivers, for example, also information on how much better the best action in a state has been as compared with the other possible actions. The representation in Fig. 2E encodes the Q-matrix value in the brightness of eight squares at the boundary of each state (center square has no meaning). Brighter colors thereby denote larger Q-matrix value.

Because our gridworld is overlaid to the real-world sample, we may also define arbitrary obstacles by providing penalties in certain regions. Figure 3 (A and B) shows examples for trajectories and policies where the particles have been trained to reach a goal state close to a virtual obstacle. Similarly, real-world obstacles can be inserted into the sample to prevent the particle from accessing specific regions and thus realizing certain actions. More complex applications



**Fig. 3. Learning with obstacles and shared information.** (A) Example trajectories for a learning process with a virtual obstacle (red square,  $R = -100$ ) next to the goal state ( $R = 5$ ) in the center of the gridworld. (B) Example trajectory for an active particle that has learned to reach a goal state ( $R = 5$ ) behind a large virtual obstacle (red rectangle,  $R = -100$ ). (C) Example trajectories for two particles sharing information during the learning process. The same rewards as in Fig. 2 have been used. (D) Sum of all Q-matrix elements at each transition comparing the learning speed with two particles sharing the information. In all the panels, the active particle speed during the learning process has been  $v_{\parallel} = 3.0 \mu\text{m s}^{-1}$ .

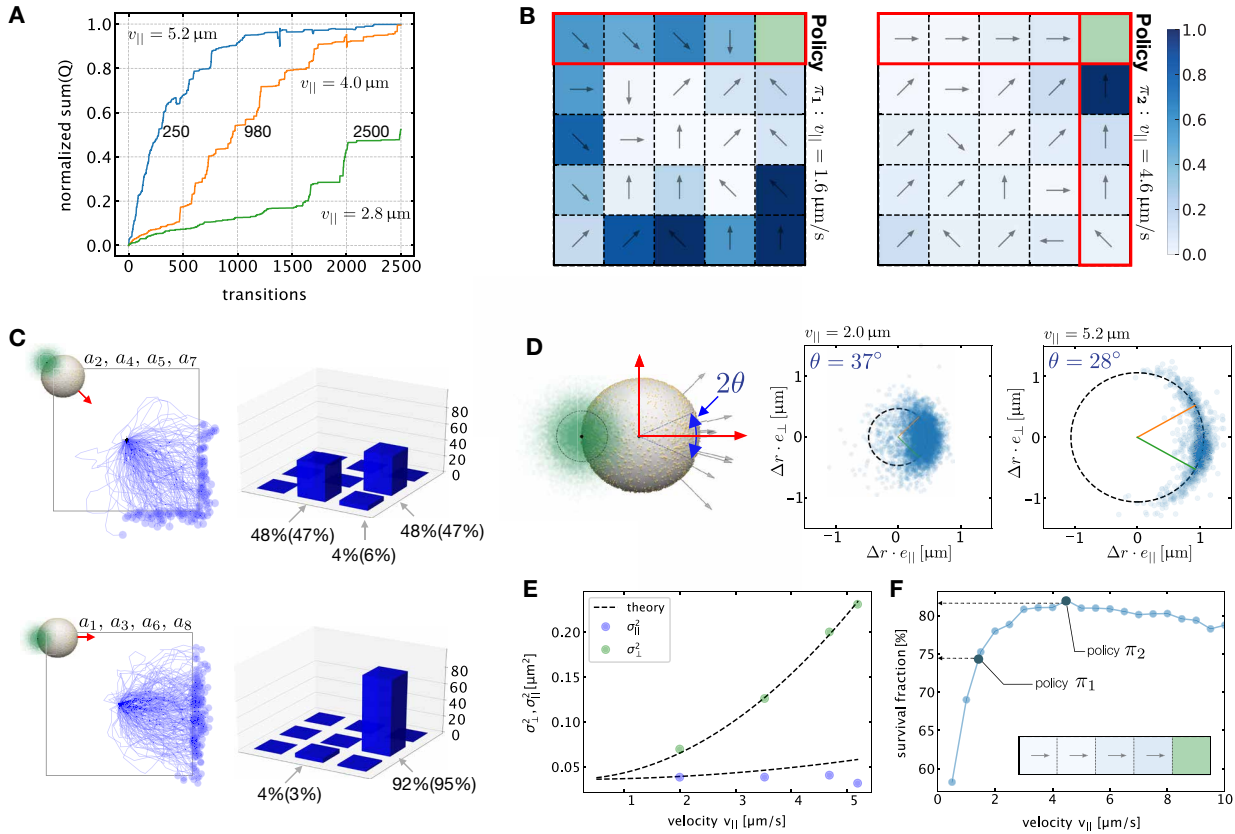
can involve the emergence of collective behavior, where the motion of multiple agents is controlled simultaneously (30). Different levels of collective and cooperative learning may be addressed (14, 39). A true collective learning is carried out when the swimmer is taking an action to maximize the reward of the collective, not only its individual one. Swimmers may also learn to act as a collective when positive rewards are given if an agent behaves like others in an ensemble (17). This mimics the process of developing swarming behavior implicated, for example, by the Vicsek model (40). Our control mechanism is capable of addressing multiple swimmers separately such that they may also cooperatively explore the environment. Instead of a true collective strategy, we are considering a low density of swimmers (number of swimmers  $\ll$  number of states), which share the information gathered during the learning process by drawing their actions from and updating the same Q-matrix. The swimmers are exploring the same gridworld in different spatial regions, and thus, a speedup of the learning is expected. Figure 3C displays the trajectories of two particles sharing the same Q-matrix, which is updated in each learning step. As a result, the learning speed is enhanced (Fig. 3D). The proposed particle control therefore provides the possibility to explore a collective learning or the optimization of collective behavior and thus delivers an ideal model system with real physical interactions.

### Influence of thermal fluctuations on the learning process

A notable difference between macroscopic agents, like robots, and microscopic active particles is the Brownian motion of microswimmers. There is an intrinsic positional noise present in the case of active particles, which is also of relevance for small living organisms like bacteria, cells, and all active processes on microscopic length scales. The advantage of the presented model system, however, is that the influence of the strength of the noise can be explored for the adaption process and the final behavior, whereas this is difficult to achieve in biological systems.

The importance of the noise in Brownian systems is commonly measured by the Peclet number,  $Pe = rv/2D$ , comparing the product of particle radius  $r$  and the deterministic particle displacement  $v\Delta t$  to the corresponding square displacements by Brownian motion  $2D\Delta t$ . To explore the influence of the noise strength, we change the speed of the active particle  $v$ , whereas the strength of the noise is given by the constant diffusion coefficient  $D$ . We further introduce a penalty in the boundary states  $R = -100$  to modify the environment in a way that the influence of noise can introduce quantitative consequences for the transitions.

When varying the speed  $v_{\parallel}$  between 2 and  $5 \mu\text{m s}^{-1}$ , we make four general observations. (i) Due to time delay in the feedback loop controlling the particles, the noise influence depends on the particle speed nonmonotonously (Fig. 4E and the Supplementary Materials). As a result, we find an optimal particle speed for which the noise is least important, as discussed in more detail in the following section. For the parameters used in the experiment, the optimal velocity is close to the maximum speed available. When increasing the speed in the limited interval of the experiment, the importance of the noise thus decreases. (ii) The Q-matrix converges considerably faster for higher particle speeds corresponding to a lower relative strength of the noise. This effect is intuitive because the stronger the noise, the lower the correlation between action and desired outcome. Figure 4A shows the convergence of the sum of the Q-matrix elements (summed over all entries for a given transition) for different



**Fig. 4. Influence of Brownian motion on the learning process.** (A) Sum of the Q-matrix elements as a function of the total number of transitions during the learning process. The different curves were obtained for learning with three different microswimmer speeds. (B) Policy obtained from learning processes at high noise (low velocity) ( $\pi_1 : v_{||} = 1.6 \mu\text{m s}^{-1}$ ) and low noise (high velocity) ( $\pi_2 : v_{||} = 4.6 \mu\text{m s}^{-1}$ ). The coloring of the states corresponds to the contrast between the value of the best action and the average of all other actions (Eq. 2). (C) Transition probabilities used in Bellman's Eq. 3 for diagonal and nondiagonal actions as determined from experiments with 500 trajectories for a velocity of 1.6 and  $4.6 \mu\text{m s}^{-1}$ . The blue lines indicate example experimental trajectories, which yield equivalent results for actions  $a_2, a_4, a_5, a_7$  (top) and  $a_1, a_3, a_6, a_8$  (bottom). The blue dots mark the first point outside the grid cell. The histograms to the right show the percentage arriving in the corresponding neighboring states. The numbers below denote the percentages for the two velocities (value in parentheses for higher velocity). (D) Origin of directional uncertainty. The green dots indicate the possible laser position due to the Brownian motion of the particle within the delay time  $\delta t$ . The two graphs to the right display the experimental particle displacements of a single microswimmer within the delay time  $\delta t = \Delta t_{\text{exp}} = 180 \text{ ms}$ , when starting at the origin for two different particle velocities. (E) Variances of the point clouds in (D) parallel and perpendicular to the intended direction of motion. The dashed lines correspond to the theoretical prediction according to Eq. 4 for the perpendicular motion ( $\sigma_{\perp}^2$ ) and  $\sigma_{||}^2 = 2D\delta t + (\cosh(\sigma_0^2) - 1)v_{||}^2\delta t^2$  for the tangential motion with  $\sigma_0^2 \approx 0.23 \text{ rad}^2, D = 0.1 \mu\text{m}^2 \text{ s}^{-1}$ , and  $t = \delta t = 180 \text{ ms}$ . (F) Survival fraction of particles moving in the upper states at the boundary toward the goal state in policy  $\pi_2$  indicated in the inset. The survival has been determined from simulations for the same parameters as in (E).

microswimmer speeds ( $v_{||} = 2.8 \mu\text{m s}^{-1}$ ,  $v_{||} = 4.0 \mu\text{m s}^{-1}$ , and  $v_{||} = 5.1 \mu\text{m s}^{-1}$ ). Although the sum reaches 50% after 250 transitions for the highest velocity, this requires almost 10 times more transitions at about half the speed. (iii) The resulting optimal policy depends on the noise strength. In Fig. 4B, we show the policies obtained for two different velocities ( $v_{||} = 1.6 \mu\text{m s}^{-1}$  and  $v_{||} = 4.6 \mu\text{m s}^{-1}$ ). Differences in the two policies are, in particular, visible in the states close to the boundary. Most of the actions at the top and right edge of the low-velocity policy point inward, whereas actions parallel to the edge are preferred at the higher velocity (see highlighted regions in Fig. 4, B and C). (iv) The contrast between the best action and the average of the other actions, which we take as a measure of the decision strength, is enhanced upon increasing importance of the noise. This contrast for a given state  $s_k$  is measured by

$$\Delta G(s_k) = \frac{1}{\Psi} \left\{ Q(s_k, a^b) - \langle Q(s_k, a_i) \rangle_i \right\} \quad (2)$$

where  $a^b$  denotes the best action for the state and  $\langle Q(s_k, a_i) \rangle_i = \sum_{i=1}^8 Q(s_k, a_i) / 8$ . The result is normalized by a factor  $\Psi$  to make the largest contrast encoded in the color of the states in Fig. 4B equal to one.

## DISCUSSION

Because the environment (gridworld with its rewards) stays constant for all learning processes at different velocities, all our above observations for varying particle speed are related to the importance of the noise strength. According to Bellman's equation (10)

$$Q(s, a) = \sum_{s'} P(s' | s, a) [R(s') + \gamma \max_{a'} Q(s', a')] \quad (3)$$

the influence of the noise on the learning process is encoded in the transition probabilities  $P(s' | s, a)$ , i.e., the probabilities that an action  $a$  in the state  $s$  leads to a transition to the state  $s'$ . This equation

couples the element  $Q(s, a)$  of the optimized Q-matrix, corresponding to a state  $s$  and action  $a$ , with the discounted elements  $\pi^*(s') = \max_a Q(s', a)$  of the optimal policy in the future states  $s'$  and the corresponding future rewards  $R(s')$ , weighted by transition probabilities  $P(s' | a, s)$ . Using this equation, one can obtain the Q-matrix and the optimal policy by a Q-matrix value iteration procedure if the transition probabilities are known. The transition probabilities thus contain the physics of the motion of the active particle, including the noise, and decide how different penalties or rewards of the neighboring states influence the value of  $Q$ .

We have measured the transition function for the two types of transitions (diagonal and nondiagonal) using 500 trajectories in a single grid cell. To obtain the transition function, we set the starting position of all the trajectories to the center of the grid cell, carried out the specific action, and determined the state in which the particle trajectory ended up. The results are shown in Fig. 4C with exemplary trajectories and a histogram to the right. The numbers below the histograms show the corresponding transition probabilities to the neighboring state in percent for a velocity of  $v_{\parallel} = 1.6 \mu\text{m s}^{-1}$  ( $v_{\parallel} = 4.6 \mu\text{m s}^{-1}$  for the values in parentheses). The two velocities show only weak changes in the transition probabilities for the nondiagonal actions, which appear to be responsible for the changes in the policies in Fig. 4B. Carrying out a Q-matrix value iteration confirms the changes in the policy in the marked regions for the measured transition probability range (see the Supplementary Materials).

The advantage of our experimental system is that we can explore the detailed physical behavior of each microswimmer in dedicated experiments. To this end, we find two distinct influences of the Brownian motion as the only noise source on the microswimmers' motion. Figure 4D shows the distribution of microswimmer displacement vectors within a time  $\Delta t_{\text{exp}} = 180 \text{ ms}$  for two different velocities. Each displacement starts at the origin, and the point cloud reflects the corresponding end points of the displacement vectors. With increasing velocity, the particles increase their step length in the desired horizontal direction. The mean distance corresponds to the speed of the particle, and the end points are located close to a circle. At the same time, a directional uncertainty is observed where the angular variance  $\sigma_{\theta}^2$  is nearly constant for all speeds (see the Supplementary Materials for details). This directional noise is the result of a delayed action in the experiments (30, 41), i.e., a time separation between sensing (imaging the position of the particle) and action on the particle position (placing the laser for propulsion). Both are separated by a delay time  $\delta t$ , which is the intrinsic delay of the feedback loop ( $\delta t = \Delta t_{\text{exp}} = 180 \text{ ms}$  in our experiments). A delayed response is a very generic feature of all active responsive systems, including biological species. In the present case of a constant propulsion speed, it leads to an anisotropic noise. In the direction perpendicular to the intended action, the Brownian noise gets an additional component that is increasing nonlinearly with the particle speed, whereas the noise along the intended direction of motion is almost constant (Fig. 4E).

The increase in the variance perpendicular to the direction of motion can be analyzed with a simple model (see the Supplementary Materials for details), which yields

$$\sigma_{\perp}^2 = v_{\parallel}^2 \delta t \sinh(\sigma_{\theta}^2) t + 2Dt \quad (4)$$

and corresponds well with experimental data (Fig. 4E) for  $\sigma_{\theta}^2 \approx 0.23 \text{ rad}^2$  and fixed time  $t = \delta t$ . In particular, it captures the nonlinear increase of  $\sigma_{\perp}^2$  with the particle speed  $v$ .

The increase has important consequences. When considering the motion in the top four states of policy  $\pi_2$  (Fig. 4B), the particle would move horizontally toward the goal starting at an arbitrary position in the leftmost state. From all trajectories that started, only a fraction will arrive at the goal state before leaving these states through the upper, lower, or left boundaries of those four states. This survival fraction has been determined from simulations (also see the Supplementary Materials for an approximate theoretical description). Overall, a change between the two policies  $\pi_1$  and  $\pi_2$  is induced by an increase of the survival by less than 10% when going from  $v_{\parallel} = 1.6 \mu\text{m s}^{-1}$  to  $v_{\parallel} = 4.6 \mu\text{m s}^{-1}$ . When further increasing the velocity, we find in simulations that an optimal velocity for maximum survival exists. This maximum corresponds to the minimum

$$v_{\parallel \text{opt}} = \sqrt{\frac{2D}{\sinh(\sigma_{\theta}^2) \delta t}} \quad (5)$$

in the variance (Eq. 4) for a fixed traveled distance  $a = v_{\parallel} t$ , which only depends on the diffusion coefficient  $D$ , the angular variance  $\sigma_{\theta}^2$ , and the sensorial delay  $\delta t$  (see the Supplementary Materials for details). In the limit of instantaneous actions ( $\delta t = 0$ ), an infinitely fast motion would yield the best results. Any nonzero delay will introduce a "speed limit" at which a maximum survival is ensured. We expect that the optimal policy for very high velocities should yield a similar policy as for low velocities. An experimental verification of this conjecture is currently out of reach, as Fig. 4F shows the results of the simulations.

The observed behavior of the survival probability, which exhibits a maximum for a certain particle velocity, implies that the probability to reach the target is maximal for the same optimal velocity. Moreover, because the underlying analysis is solely based on the competition of two noises omnipresent in (Brownian) active matter, namely the diffusion and the uncertainty in choosing the right direction, we conjecture that the observed type of behavior is universal. The precision of reaching the target (long time variance of the distance from the target) by the run-and-tumble motion of bacteria exhibits a minimum as a function of the run-and-tumble times (42, 43) reminiscent of our results. These results also demonstrate that the combination of machine learning algorithms with real-world microscopic agents can help to uncover physical phenomena (such as time delay in the present work), which play important roles in the microscopic motion of biological species.

Concluding, we have demonstrated RL with a self-thermophoretic microswimmer carrying out actions in a real-world environment with its information processing and sensing capabilities externalized to a computer and a microscopy setup. Already with this hybrid solution, one obtains a model system, where strategies in a noisy environment with virtual obstacles or collective learning can be explored. Although our simple realization of a gridworld is based on a global position detection defining the state of the swimmer, future applications will consider local information, e.g., the response to a temporal sequence of local physical or chemical signals, to allow for navigation in unknown environments. As compared with a computer simulation, our system contains a nonideal control limited by the finite reaction time of the feedback loop, presence of liquid flows, imperfections of the swimmers or sample container, hydrodynamic interactions, or other uncontrolled parameters that naturally influence the learning process. In this way, it resembles a new form of computer simulation using real-world agents. An important advantage is that

the physics of the agent can be explored experimentally in detail to understand the learned strategies, and the real-world interactions in more complex environments can be used to adapt the microswimmer's behavior. In that sense, even the inverse problem of using the learned strategy to reveal the details of these uncontrolled influences may be addressed as a new form of environmental sensing. Similarly, the control of active particles by machine learning algorithms may be used in evolutionary robotics (8, 44), where the interaction of multiple particles may be optimized to yield higher-order functional structures based on environmental interactions. Although the implementation of signaling and feedback by physical or chemical processes into a single artificial microswimmer is still a distant goal, the current hybrid solution opens a whole branch of new possibilities for understanding adaptive behavior of single microswimmers in noisy environments and the emergence of collective behavior of large ensembles of active systems.

## MATERIALS AND METHODS

### Materials

Samples consisted of commercially available gold nanoparticle-coated melamine resin particles of a diameter of 2.19  $\mu\text{m}$  (microParticles GmbH, Berlin, Germany). The gold nanoparticles were covering about 30% of the surface area and were between 8 and 30 nm in diameter (see the Supplementary Materials for details.) Microscopy glass cover slides were dipped into a 5% Pluronic F127 solution, rinsed with deionized water, and dried with nitrogen. The Pluronic F127 coating prevented sticking of the particles to the glass cover slides. Two microliters of particle suspension was placed on the cover slides to spread about an area of 1 cm by 1 cm, forming a 3- $\mu\text{m}$ -thin water film. The edges of the sample were sealed with silicone oil (polydimethylsiloxane) to prevent water evaporation.

### Methods

Samples were investigated in a custom-built inverted dark-field microscopy setup based on an Olympus IX-71 microscopy stand. The sample was held by a Piezo stage (Physik Instrumente) that was mounted on a custom-built stepper stage for coarse control. The sample was illuminated by a halogen lamp (Olympus) using a dark-field oil-immersion condenser [Olympus, numerical aperture (NA), 1.2]. The scattered light was collected by an oil-immersion objective lens (Olympus, 100 $\times$ , NA 1.35 to 0.6) with the NA set to 0.6 and captured with an Andor iXon emCCD camera. A  $\lambda = 532$  nm laser was focused by the imaging objective into the sample plane to serve as a heating laser for the swimmers. Its position in the sample plane was steered by an acousto-optic deflector (AOD; AA Opto-Electronic) together with a 4- $f$  system (two  $f = 20$  cm lenses). The AOD was controlled by an ADwin realtime board (ADwin-Gold, Jäger Messtechnik) exchanging data with a custom LabVIEW program. A region of interest of 512 pixels by 512 pixels (30  $\mu\text{m}$  by 30  $\mu\text{m}$ ) was used for the real-time imaging, analysis, and recording of the particles, with an exposure time of  $\Delta t_{\text{exp}} = 180$  ms. The details of integrating the RL procedure are contained in the Supplementary Materials.

## SUPPLEMENTARY MATERIALS

robotics.sciencemag.org/cgi/content/full/6/52/eabd9285/DC1

Fig. S1. Symmetric swimmer structure.

Fig. S2. Swimmer speed as a function of laser power.

Fig. S3. Directional noise as function of the swimming velocity measured in the experiment.

Fig. S4. Directional noise model.

Fig. S5. Results of the analytical model of the influence of the noise.

Fig. S6. Q-matrix value iteration result.

Movie S1. Single-swimmer free navigation toward a target during learning.

Movie S2. Single-swimmer free navigation toward a target after learning.

Movie S3. Navigation toward a target with virtual obstacles.

Movie S4. Multiple-swimmer free navigation toward a target.

## REFERENCES AND NOTES

- J. K. Parrish, W. M. Hamner, *Animal Groups in Three Dimensions* (Cambridge Univ. Press, 1997).
- Y. Lin, N. Abaid, Collective behavior and predation success in a predator-prey model inspired by hunting bats. *Phys. Rev. E* **88**, 062724 (2013).
- G. Reddy, A. Celani, T. J. Sejnowski, M. Vergassola, Learning to soar in turbulent environments. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E4877–E4884 (2016).
- J. A. Kromer, S. Märcker, S. Lange, C. Baier, B. M. Friedrich, Decision making improves sperm chemotaxis in the presence of noise. *PLOS Comput. Biol.* **14**, e1006109 (2018).
- B. ten Hagen, F. Kümmel, R. Wittkowski, D. Takagim, H. Löwen, C. Bechinger, Gravitaxis of asymmetric self-propelled colloidal particles. *Nat. Commun.* **5**, 4829 (2014).
- L. P. Kaelbling, M. L. Littman, A. W. Moore, Reinforcement learning: A survey. *J. Artif. Intell. Res.* **4**, 237–285 (1996).
- J. Kober, J. Peters, Reinforcement learning in robotics: A survey, in *Learning Motor Skills* (Springer Tracts in Advanced Robotics, 2014), vol. 97, pp. 9–67.
- S. Doncieux, N. Bredeche, J. B. Mouret, A. E. Eiben, Evolutionary robotics: What, why, and where to. *Front. Robot. AI* **2**, 4 (2015).
- M. Wiering, M. v. Otterlo, Reinforcement Learning, in *Adaptation, Learning, and Optimization* (Springer Berlin Heidelberg, 2012), vol. 12.
- R. S. Sutton, A. G. Barto, *Reinforcement Learning: An Introduction* (MIT Press, 1998).
- S. Colabrese, K. Gustavsson, A. Celani, L. Biferale, Flow navigation by smart microswimmers via reinforcement learning. *Phys. Rev. Lett.* **118**, 158004 (2017).
- K. Gustavsson, L. Biferale, A. Celani, S. Colabrese, Finding efficient swimming strategies in a three-dimensional chaotic flow by reinforcement learning. *Eur. Phys. J. E* **40**, 110 (2017).
- J. K. Alageshan, A. K. Verma, J. Bec, R. Pandit, Machine learning strategies for path-planning microswimmers in turbulent flows. *Phys. Rev. E* **101**, 043110 (2020).
- H. M. La, R. Lim, W. Sheng, Multirobot cooperative learning for predator avoidance. *IEEE Trans. Control Syst. Technol.* **23**, 52–63 (2015).
- M. Birattari, A. Ligot, D. Bozhinoski, M. Brambilla, G. Francesca, L. Garattoni, D. Garzón Ramos, K. Hasselmann, M. Kegeleirs, J. Kuckling, F. Pagnozzi, A. Roli, M. Salman, T. Stützle, Automatic off-line design of robot swarms: A manifesto. *Front. Robot. AI* **16**, 59 (2019).
- L. Pitonakova, R. Crowder, S. Bullock, Information flow principles for plasticity in foraging robot swarms. *Swarm Intell.* **10**, 33–63 (2016).
- K. Ried, T. Müller, H. J. Briegel, Modelling collective motion based on the principle of agency: General framework and the case of marching locusts. *PLOS ONE* **14**, e0212044 (2019).
- S. Verma, G. Novati, P. Koumoutsakos, Efficient collective swimming by harnessing vortices through deep reinforcement learning. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 5849–5854 (2018).
- E. Schneider, H. Stark, Optimal steering of a smart active particle. *EPL* **127**, 64003 (2019).
- Y. Yang, M. A. Bevan, B. Li, Efficient navigation of colloidal robots in an unknown environment via deep reinforcement learning. *Adv. Intell. Syst.* **2**, 1900106 (2020).
- C. Bechinger, R. Di Leonardo, H. Löwen, V. Volpe, V. Volpe, Active particles in complex and crowded environments. *Rev. Mod. Phys.* **88**, 045006 (2016).
- J. Palacci, S. Sacanna, A. P. Steinberg, D. J. Pine, P. M. Chaikin, Living crystals of light-activated colloidal surfers. *Science* **339**, 936–940 (2013).
- I. Buttinoni, J. Bialké, F. Kümmel, H. Löwen, C. Bechinger, T. Speck, Dynamical clustering and phase separation in suspensions of self-propelled colloidal particles. *Phys. Rev. Lett.* **110**, 238301 (2013).
- A. Aubret, M. Youssef, S. Sacanna, J. Palacci, Targeted assembly and synchronization of self-spinning microgears. *Nat. Phys.* **14**, 1114–1118 (2018).
- Y. Wu, X. Lin, Z. Wu, H. Möhwald, Q. He, Self-propelled polymer multilayer Janus capsules for effective drug delivery and light-triggered release. *ACS Appl. Mater. Interfaces* **6**, 10476–10481 (2014).
- M. Safdar, J. Simmchen, J. Jänis, Light-driven micro- and nanomotors for environmental remediation. *Environ. Sci. Nano* **4**, 1602–1616 (2017).
- F. Cichos, K. Gustavsson, B. Mehlig, G. Volpe, Machine learning for active matter. *Nat. Mach. Intell.* **2**, 94–103 (2020).
- B. Qian, D. Montiel, A. Bregulla, F. Cichos, H. Yang, Harnessing thermal fluctuations for purposeful activities: The manipulation of single micro-swimmers by adaptive photon nudging. *Chem. Sci.* **4**, 1420–1429 (2013).
- A. P. Bregulla, H. Yang, F. Cichos, Stochastic localization of microswimmers by photon nudging. *ACS Nano* **8**, 6542–6550 (2014).

30. U. Khadka, V. Holubec, H. Yang, F. Cichos, Active particles bound by information flows. *Nat. Commun.* **9**, 3864 (2018).
31. F. A. Lavergne, H. Wendehenne, T. Bäuerle, C. Bechinger, Group formation and cohesion of active particles with visual perception-dependent motility. *Science* **364**, 70–74 (2019).
32. A. C. H. Tsang, E. Demir, Y. Ding, O. S. Pak, Roads to smart artificial microswimmers. *Adv. Intell. Syst.* **2**, 1900137 (2020).
33. E. M. Purcell, Life at low Reynolds number. *Am. J. Phys.* **45**, 3–11 (1977).
34. H.-R. Jiang, N. Yoshinaga, M. Sano, Active motion of a Janus particle by self-thermophoresis in a defocused laser beam. *Phys. Rev. Lett.* **105**, 268302 (2010).
35. I. Buttinoni, G. Volpe, F. Kümmel, G. Volpe, C. Bechinger, Active Brownian motion tunable by light. *J. Phys. Condens. Matter* **24**, 284129 (2012).
36. M. Selmke, U. Khadka, A. P. Bregulla, F. Cichos, H. Yang, Theory for controlling individual self-propelled micro-swimmers by photon nudging I: Directed transport. *Phys. Chem. Chem. Phys.* **20**, 10502–10520 (2018).
37. M. Selmke, U. Khadka, A. P. Bregulla, F. Cichos, H. Yang, Theory for controlling individual self-propelled micro-swimmers by photon nudging II: Confinement. *Phys. Chem. Chem. Phys.* **20**, 10521–10532 (2018).
38. J. C. H. Watkins, thesis, King's College, Cambridge (1989).
39. L. Busoniu, R. Babuška, B. De Schutter, Multi-agent reinforcement learning: A survey, in *Proceedings of the 9th International Conference on Control, Automation, Robotics and Vision (ICARCV 2006)* (Singapore, 2006), pp. 527–532.
40. T. Vicsek, A. Czirók, E. Behn-Jakob, I. Cohen, O. Shochet, Novel type of phase transition in a system of self-driven particles. *Phys. Rev. Lett.* **75**, 1226–1229 (1995).
41. D. Geiss, K. Kroy, V. Holubec, Brownian molecules formed by delayed harmonic interactions. *New J. Phys.* **21**, 093014 (2019).
42. P. Romanczuk, G. Salbreux, Optimal chemotaxis in intermittent migration of animal cells. *Phys. Rev. E* **91**, 042720 (2015).
43. A. Diz-Muñoz, P. Romanczuk, W. Yu, B. Bergert, K. Ivanovitch, G. Salbreux, C.-P. Heisenberg, E. K. Paluch, Steering cell migration by alternating blebs and actin-rich protrusions. *BMC Biol.* **14**, 74 (2016).
44. S. Jones, A. F. Winfield, S. Hauert, M. Studley, Onboard evolution of understandable swarm behaviors. *Adv. Intell. Syst.* **1**, 1900031 (2019).

**Acknowledgments:** Helpful discussion with P. Romanczuk is acknowledged in pointing out observations of directional noise for biological systems. Fruitful discussion and help with extrapolating the theory to the experiments by K. Ghazi-Zahedi are acknowledged. We thank A. Kramer for helping to revise the manuscript. **Funding:** The authors acknowledge financial support by the DFG Priority Program 1726 “Microswimmers” through project 237143019. F.C. is supported by the DFG grant 432421051. V.H. is supported by a Humboldt grant of the Alexander von Humboldt Foundation and by the Czech Science Foundation (project no. 20-02955J). **Author contributions:** F.C. conceived the research. S.M.-L. and F.C. designed the experiments. S.M.-L. implemented the system, and S.M.-L. and A.F. performed the experiments. S.M.-L., V.H., and F.C. analyzed and discussed the data. F.C., V.H., and S.M.-L. wrote the manuscript. **Competing interests:** The authors declare that they have no competing financial interests. **Data and materials availability:** All data needed to evaluate the conclusions are available in the paper or in the Supplementary Materials. Additional data and materials are available upon request.

Submitted 21 July 2020

Accepted 26 February 2021

Published 24 March 2021

10.1126/scirobotics.abd9285

**Citation:** S. Muiños-Landin, A. Fischer, V. Holubec, F. Cichos, Reinforcement learning with artificial microswimmers. *Sci Robot.* **6**, eabd9285 (2021).

## Reinforcement learning with artificial microswimmers

S. Muiños-Landin, A. Fischer, V. Holubec and F. Cichos

*Sci. Robotics* **6**, eabd9285.  
DOI: 10.1126/scirobotics.abd9285

ARTICLE TOOLS	<a href="http://robotics.sciencemag.org/content/6/52/eabd9285">http://robotics.sciencemag.org/content/6/52/eabd9285</a>
SUPPLEMENTARY MATERIALS	<a href="http://robotics.sciencemag.org/content/suppl/2021/03/22/6.52.eabd9285.DC1">http://robotics.sciencemag.org/content/suppl/2021/03/22/6.52.eabd9285.DC1</a>
RELATED CONTENT	<a href="http://robotics.sciencemag.org/content/robotics/6/52/eabh1977.full">http://robotics.sciencemag.org/content/robotics/6/52/eabh1977.full</a>
REFERENCES	This article cites 38 articles, 4 of which you can access for free <a href="http://robotics.sciencemag.org/content/6/52/eabd9285#BIBL">http://robotics.sciencemag.org/content/6/52/eabd9285#BIBL</a>
PERMISSIONS	<a href="http://www.sciencemag.org/help/reprints-and-permissions">http://www.sciencemag.org/help/reprints-and-permissions</a>

Use of this article is subject to the [Terms of Service](#)

---

*Science Robotics* (ISSN 2470-9476) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Robotics* is a registered trademark of AAAS.

Copyright © 2021 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works

# Supplementary Information. Reinforcement learning of artificial microswimmers

Santiago Muiños-Landin<sup>1</sup>, Viktor Holubec<sup>2,3</sup>, and Frank Cichos<sup>1</sup>

<sup>1</sup>Molecular Nanophotonics, Peter Debye Institute for Soft Matter Physics, Linnestraße 5, 04103 Leipzig, Germany

<sup>2</sup>Soft Condensed Matter Theory, Institute of Theoretical Physics,Brüderstraße, 04103 Leipzig, Germany

<sup>3</sup>Charles University, Faculty of Mathematics and Physics, Department of Macromolecular Physics, V Holešovičkách 2, CZ-180 00 Praha, Czech Republic

October 26, 2019

## Contents

<b>1 Self-thermophoretic Swimmer</b>	<b>2</b>
<b>2 Reinforcement Learning Procedure</b>	<b>2</b>
<b>3 Directional Noise in the Experiment</b>	<b>3</b>
<b>4 Directional Noise Model</b>	<b>4</b>
<b>5 Noise Influence Model</b>	<b>5</b>
<b>6 Q-matrix Value Iteration Results</b>	<b>5</b>
<b>7 Video Description</b>	<b>6</b>
7.1 Single swimmer free navigation towards a target example . . . . .	6
7.2 Navigation towards a target with virtual obstacles . . . . .	7
7.3 Multiple swimmer free navigation towards a target . . . . .	7



# 1 Self-thermophoretic Swimmer

In the experiments, we used symmetric active particles with 30% of their melamine resin surface covered with gold nanoparticles (AuNP). The propulsion velocity  $v$  is the result of an asymmetric illumination with a highly focused laser at a wavelength of  $\lambda = 532$  nm, which heats the gold nanoparticles at the surface, creating a surface temperature gradient and corresponding thermo-osmotic creep flows. A sketch of the particle and a corresponding electron microscopy image are shown in Supplementary Figure 1.

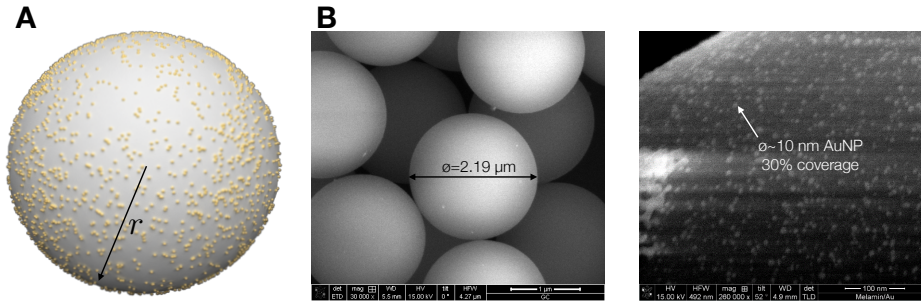


Figure 1: (a) Sketch of the symmetric active particle of radius  $R$ . For the experiments a melamine particle of  $R = 1.09 \mu\text{m}$  covered with 10 nm gold nanoparticles at 30% of its surface is heated with a focused laser. The scale bar corresponds to  $1 \mu\text{m}$ . (b) Electron microscopy image of the gold nanoparticle at the surface of the melamine resin particle.

The active particle has a propulsion velocity, which can be controlled with the help of the laser power. Figure 4 shows the particle velocity as a function of the laser power for various independent measurements. The velocity has been extracted from the particle displacements for an exposure time of  $\delta t_{\text{exp}} = 180$  ms with a laser focused to a spot size of 500 nm.

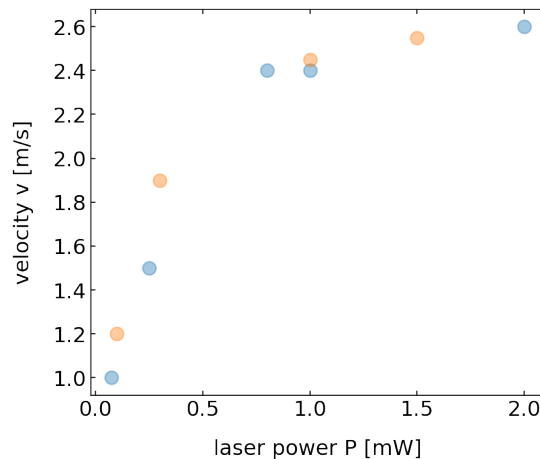


Figure 2: Dependence

The power dependence of the velocity is nonlinear since the particle is moving during the exposure while the laser spot is fixed for the exposure time. This yields a nonlinear dependence, where the maximum velocity is limited by the exposure time. Higher velocities are obtained for lower exposure times. A more detailed analysis is contained in reference and its supplementary information [1].

## 2 Reinforcement Learning Procedure

The reinforcement learning procedure is implemented in LabView inside MatLab nodes. The LabView program is using the NI Vision modules to perform real-time localization of the particles. The coordinates

are provided to the MatLab nodes to analyze the state. When the state has been identified, the Q-matrix values of the state are analyzed to select an appropriate action. To decide which action to take, an *epsilon* greedy selection procedure is applied. A random number  $0 \leq \eta \leq 1$  is drawn from an equal distribution and compared to  $\epsilon = 0.2$ . For  $\eta > \epsilon$ , the best action is selected according to the maximum value of the Q-matrix for the state. In all other cases, random action is selected. The selected action is applied until the swimmer transitions to one of the neighboring states. Note that image recording and action are separated by the time delay of the feedback loop which corresponds to one frame  $\delta t = \Delta t_{\text{exp}} = 180$  ms. When the swimmer transitions to a new state  $s'$ , the Q matrix of the previous state ( $Q(s, a)$ ) is updated according to

$$Q_{t+\Delta t}(s, a) = Q_t(s, a) + \alpha [R(s') + \gamma \max_{a'} Q_t(s', a') - Q_t(s, a)]. \quad (1)$$

The element  $Q(s, a)$  of the Q-matrix is updated by adding the amount inside the brackets multiplied by the learning rate, which has been set to  $\alpha = 0.5$  in all experiments. The bracket contains the reward  $R(s')$  for the transition to the next state  $s'$ , the discounted Q-matrix element for the state  $s'$  and the action  $a'$ . The discount factor  $\gamma$  determines the type of the learning process (finite horizon, infinite horizon, or average reward) [2]. In this work, we use the infinite horizon model, where future values of the Q-matrix are discounted with  $\gamma = 0.8$ . Further the the current Q-matrix element  $Q_t(s, a)$  is subtracted.

This procedure is repeated until the swimmer enters a boundary state or arrives at the goal. At this point an episode ends and the swimmer is placed at an arbitrary position on the grid world to start the next episode. The learning procedure is stopped manually when the sum of all Q matrix entries does not show any notable change.

### 3 Directional Noise in the Experiment

The directional noise (see Figure 4 of the main text) has been evaluated in the experiments by recording the motion of the swimmers and the corresponding actions during the learning process of the microswimmer at different swimmer velocities. The swimmer displacements  $\Delta \vec{r}$  between subsequent frames  $\Delta t = \Delta t_{\text{exp}} = 180$ ms have been projected to the direction of the action  $\mathbf{e}_{\parallel}$  and perpendicular to it ( $\mathbf{e}_{\perp}$ ). The results are shown in Figure 3 for different microswimmer speeds.

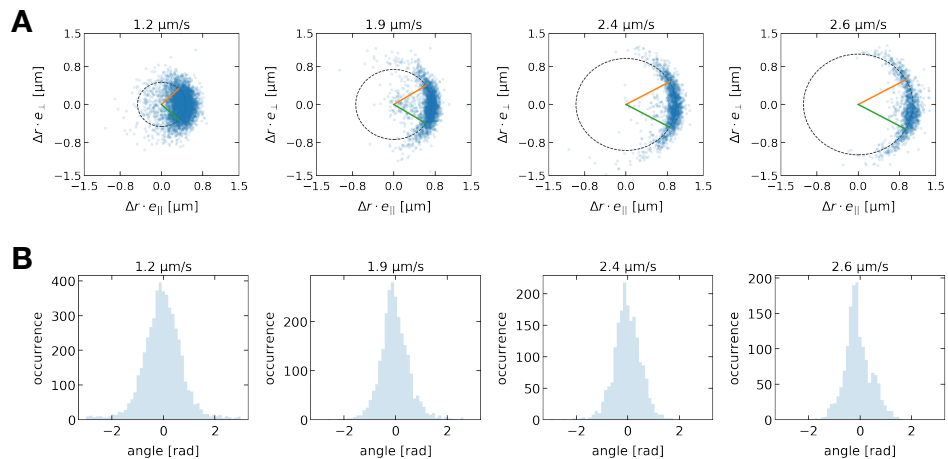


Figure 3: Caption

The results in Fig. 3A show, that the average displacement in the direction of the action increases due to the increase of the swimmer speed. From the displacement statistics, we have determined the mean velocities as denoted on the top of the graph. All measurements show a directional noise, i.e. an uncertainty in the direction of the velocity, while the mean direction follows the direction of the action. The angular distribution of the velocity directions is depicted in 3B. The measured half opening angles according to the variance of the angular distribution are  $44^\circ$ ,  $32^\circ$ ,  $30^\circ$  and  $30^\circ$  from left to right. The angular spread therefore stays approximately constant with changing microswimmer velocity.

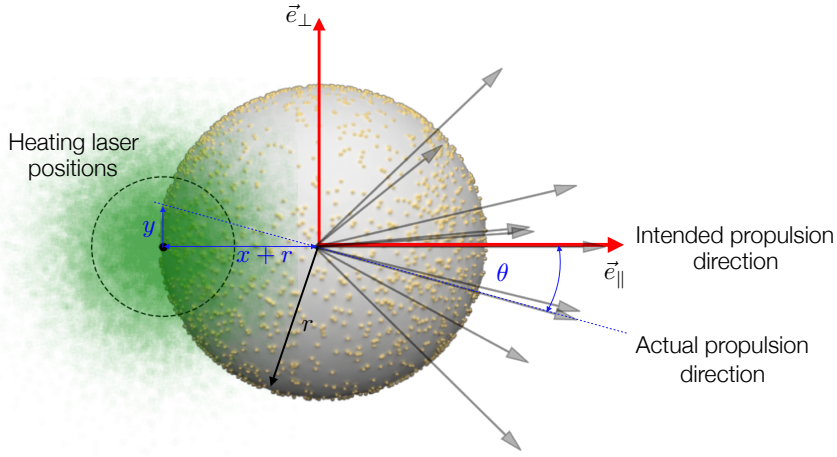


Figure 4: The angle  $\theta$  between the intended and the actual propulsion direction is a result of the diffusion of the microswimmer from the position  $(0,0)$  to the position  $(x,y)$  during the time interval  $\delta t$  between the measurement and switching on the laser.

## 4 Directional Noise Model

Consider the situation depicted in Fig. 4, and let us assume that we want to propel the microswimmer to the right in the direction of the red arrow (along  $\vec{e}_{\parallel}$ ). Then we need to focus the laser to the point at the opposite end of the particle, corresponding to the blue dot at  $(-r, 0)$  in the space fixed  $\vec{e}_{\parallel}, \vec{e}_{\perp}$  coordinate system. However, during the time interval of length  $\delta t$  between the measurement (determination of the target point for the laser) and switching on the laser, the particle diffuses from the position  $(0, 0)$  to  $(x, -y)$ , and thus the actual coordinates of the target point for the laser are  $[-(x+r), y]$  instead of  $(-r, 0)$ . The actual direction of the particle is thus declined by the angle  $\theta = \arctan y/(x+R)$  from the intended direction.

Due to the stochasticity of the Brownian motion, the angle  $\theta$  is also random. Its probability density can be calculated as

$$p(\theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx dy \delta \left[ \theta - \tan \left( \frac{y}{x+r} \right) \right] \frac{1}{4\pi D\delta t} \exp \left[ -\frac{x^2 + y^2}{4D\delta t} \right], \quad (2)$$

where we integrate over the Gaussian distribution of the displacements  $(x, y)$  during time  $\delta t$ . The integration can be performed without any approximation, but the result is not very enlightening. An intuitive formula can be obtained if we assume that the variances  $\langle y^2 \rangle = \langle x^2 \rangle = 2D\delta t$  are much smaller than  $r^2$ , i.e. that that the particle diffuses during the time  $\delta t$  by much less than  $r$ . Then the angle  $\theta$  is small and we can approximate  $\tan [y/(x+r)]$  as  $y/(x+r) \approx y/r$  yielding the Gaussian probability density

$$p(\theta) \approx \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx dy \delta \left( \theta - \frac{y}{r} \right) \frac{1}{4\pi D\delta t} \exp \left[ -\frac{x^2 + y^2}{4D\delta t} \right] = \frac{1}{\sqrt{2\sigma_{\theta}^2}} \exp \left( -\frac{\theta^2}{2\sigma_{\theta}^2} \right), \quad (3)$$

with the variance

$$\sigma_{\theta}^2 = \frac{2D\delta t}{R^2}. \quad (4)$$

The probability density (eq. 3) is normalized for  $\theta \in (-\infty, \infty)$  which might look strange since  $\theta$  is an angle. However, according to our assumption that  $\theta$  is small, the Gaussian 3 decays very fast and the normalization, caused by the used approximation, does not bring any practical problems as we see by comparing predictions based on Eq. 3 to simulated and measured data.

Let us note that due to the diffusion of the particle from  $(0, 0)$  to  $(x, y)$ , also the particle velocity imposed by the laser fluctuates. However, this effect can be described by an effective average velocity and, different from the distribution (eq. 3), it does not cause any qualitatively new behavior, compared to the situation with a deterministic velocity.

## 5 Noise Influence Model

Let us assume that the particle moves due to the drift velocity  $v_x$  in the direction of the action (say the  $x$ -direction) a certain distance  $a$ . The time required for the particle to arrive is then  $t = a/v_x$ . On the other hand, the particle position also spreads in the  $y$ -direction. Due to the action supplied after a delay time  $\delta t$ , the particle in the  $y$ -direction not only diffuses, but it also acquires a drift velocity  $v_y$ . This drift velocity fluctuates with each step such that the motion of the particle during a single step can be described by the following overdamped Langevin equation

$$\dot{x} = v \cos(\theta) + \sqrt{2D}\eta, \quad (5)$$

$$\dot{y} = v \sin(\theta) + \sqrt{2D}\eta. \quad (6)$$

Hereby, we assume that the angle  $\theta$  between the intended direction ( $x$ -axis) and the actual direction is normal distributed due to the delay with a distribution function given by Eq. 3. Therefore, the first terms on the right side of the Langevin equation have the following properties

$$v_x = \langle v \cos(\theta) \rangle = v \exp\left(-\frac{\sigma_\theta^2}{2}\right), \quad (7)$$

$$v_y = \langle v \sin(\theta) \rangle = 0, \quad (8)$$

$$\langle v^2 \sin^2(\theta) \rangle = v^2 \sigma_{y\theta}^2, \quad (9)$$

with  $v$  denoting the thermophoretic velocity of the particle, and

$$\sigma_{y\theta}^2 = \frac{1}{2} \left(1 - e^{-2\sigma_\theta^2}\right). \quad (10)$$

Therefore the position  $y$  also obeys

$$\langle y \rangle - \langle y_0 \rangle = 0 \quad (11)$$

$$\langle y^2 \rangle - \langle y \rangle^2 = \sigma_{y\theta}^2 v^2 \delta t^2 + 2D\delta t \quad (12)$$

$$(13)$$

after one time step  $\delta t$ . If the particle is now doing multiple steps in time to reach the goal, the variance of the distribution in the direction of  $y$  is

$$\sigma_\perp^2 = \langle y^2 \rangle_t - \langle y \rangle_t^2 = \sigma_{y\theta}^2 v^2 \delta t^2 \frac{t}{\delta t} + 2Dt, \quad (14)$$

where we used the fact that the individual steps are statistically independent. Accordingly the particles have traveled on average in the  $y$ -direction a distance

$$\sigma_\perp = \exp\left(\frac{\sigma_\theta^2}{4}\right) \sqrt{\sigma_{y\theta}^2 v a \delta t + 2D \frac{a}{v}} \quad (15)$$

The plot below shows that at small velocities the width of the distribution is actually decreasing as the drift starts to take over. At larger velocities the transverse drift becomes important and the uncertainty in the particle direction increases the width almost linearly in time. The width  $\sigma_\perp$  reveals an optimal for the velocity

$$v_{\text{opt}} = \sqrt{D \frac{2}{\sigma_{y\theta}^2 \delta t}}. \quad (16)$$

Thus a delayed action provides besides the positional noise a directional noise, which intrinsically yields a speed limit for the motion towards a dedicated target region. At this speed limit a maximum probability of arrival at the target region will be observed.

## 6 Q-matrix Value Iteration Results

We have calculated the optimal Q-matrix and policy with the help of a Q-matrix value iteration procedure. According to that the Q-matrix value of state  $s$  for an action  $a$  is given by the Bellman equation

$$Q^\pi(s, a) = \sum_{s'} P(s'|s, a) \left[ R(s') + \gamma \max_{a'} Q^\pi(s', a') \right], \quad (17)$$

with  $P(s'|s, a)$  being the transition probability from state  $s$  by the action  $a$  to state  $s'$ ,  $R(s')$  is the reward for the state  $s'$ ,  $\gamma$  the discount factor. Equation 17 defines the Q-matrix for a certain policy  $\pi$ . Wehn initializing the Q-matrix randomly, one obtains an iterative procedure for updating the Q-matrix until it converged and the optimal policy  $\pi^*$  is obtained. For our iterative calculation we use a discount factor of  $\gamma = 0.8$  as in the experiments. The transition probabilities were also adapted from the experimental results. Figure 5 A shows the definition of the used parameters for the transition probabilities for the two types of actions (non-diagonal and diagonal). Other transition probabilities than indicated are set to zero. For the non-diagonal actions we varied the parameter  $p_1$  between 0.9 and 1.0 reflecting the change of the transition probability with the speed of the swimmer in the experiments. For the diagonal actions we chose  $p_2 = 0.06$ , which is kept fixed for all calculations. The iterations were carried out until no further change in the sum of all Q-matrix values was observed, which typically happened after 10 to 20 iterations.

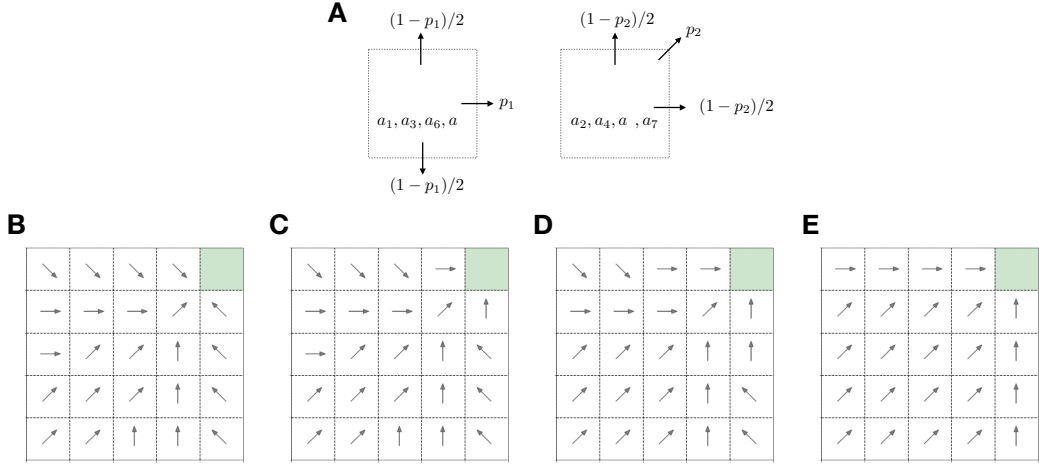


Figure 5: **Q-matrix value iteration result** **A**) Definition of the transition probabilities for the Q-matrix value iteration procedure. The left sketch is valid for the indicated actions and parameter  $p_1$  is varied between 0.9 and 1.0 according to the experimental observation. The right sketch is valid for all diagonal action as indicated. The parameter  $p_2$  if kept constant at a value of 0.06 during all simulation. **B** Optimal policy obtained for  $p_1 = 0.94$ . **C** Optimal policy obtained for  $p_1 = 0.96$ . **D** Optimal policy obtained for  $p_1 = 0.98$ . **E** Optimal policy obtained for  $p_1 = 1.0$ .

Figure 5B-E is showing the optimal policies obtained for different transition probabilities for the non-diagonal actions. The main observation in the calculation is a change in the actions at boundary states at the top grid cell row and the and right grid cell column. There the actions point for higher noise influence (low microswimmer speed) diagonal to the inside of the grid world, while they are pointing along the row/column for lower noise influence (higher particle speeds). Note that only small changes in the transition probability  $p_1$  are required to obtain the observed change. While the obtained differences are similar to the experiment, the iteration procedure neglects the position dependent transition probability given in the experiment.

## 7 Video Description

### 7.1 Single swimmer free navigation towards a target example

**Video 1** shows the particle during the learning process after different amount of episodes. The starting point of the swimmer is in the lower left cell and the absorption state (target) is placed as described in Fig. 1 of the main text in the top right corner. The video is 5X accelerated. The swimmer starts with a random policy and random motion until it reaches the absorption state and the episode end. After more and more episodes the system starts to find an optimal path to the absorption state. Measurements were carried out at a swimmer velocity of  $v = 1.5 \mu\text{m s}^{-1}$  and an inverse framerate of  $\Delta t_{\text{exp}} = 180 \text{ ms}$ .

**Video 2** shows directed motion of the particle from different starting locations in the grid world after convergence of the Q-matrix.

## 7.2 Navigation towards a target with virtual obstacles

**Video 3** displays the swimmer exploring the gridworld for a target inside the gridworld with different sized obstacles (large penalty areas) right next to it. The video shows how the system has learned the actions to be executed from different initial states. Measurements were carried out at a swimmer velocity of  $v = 1.5 \mu\text{m s}^{-1}$  and an inverse framerate of  $\Delta t_{\text{exp}} = 180 \text{ ms}$ . Video is 5X accelerated. If the swimmer is placed behind the wall it also finds the target after exploring. Even increasing the size of the virtual wall the system finds a solution of the exploration problem by finding a direct path to the target. The policies obtained by the reinforcement learning from these experiments are shown in Figure 3 of the main text.

## 7.3 Multiple swimmer free navigation towards a target

**Video 4** shows different episodes where two swimmers are searching the absorption state (target) placed again at the up right corner. The episodes finishes after one swimmer reaches the target.  $Pe = 80$  and framerate 180 ms.

## References

- [1] U. Khadka, V. Holubec, H. Yang, F. Cichos, Active particles bound by information flows. *Nature Communications* **9**, 1–9 (2018).
- [2] Kaelbling, L. P., M. L. Littman, A. W. Moore, Reinforcement learning: a survey. *J. Artif. Intell. Res* **4**, 237–285 (1996).

## Finite-Size Scaling at the Edge of Disorder in a Time-Delay Vicsek Model

Viktor Holubec<sup>1,2,\*</sup>, Daniel Geiss<sup>1,3</sup>, Sarah A. M. Loos<sup>1,5</sup>, Klaus Kroy,<sup>1</sup> and Frank Cichos<sup>4</sup>


<sup>1</sup>*Institut für Theoretische Physik, Universität Leipzig, Postfach 100 920, D-04009 Leipzig, Germany*

<sup>2</sup>*Charles University, Faculty of Mathematics and Physics, Department of Macromolecular Physics, V Holešovičkách 2, CZ-180 00 Praha, Czech Republic*

<sup>3</sup>*Max Planck Institute for Mathematics in the Sciences, D-04103 Leipzig, Germany*

<sup>4</sup>*Peter Debye Institute for Soft Matter Physics, Universität Leipzig, 04103 Leipzig, Germany*

<sup>5</sup>*ICTP — International Centre for Theoretical Physics, Strada Costiera 11, 34151, Trieste, Italy*

 (Received 13 July 2021; revised 29 September 2021; accepted 16 November 2021; published 14 December 2021)

Living many-body systems often exhibit scale-free collective behavior reminiscent of thermal critical phenomena. But their mutual interactions are inevitably retarded due to information processing and delayed actuation. We numerically investigate the consequences for the finite-size scaling in the Vicsek model of motile active matter. A growing delay time initially facilitates but ultimately impedes collective ordering and turns the dynamical scaling from diffusive to ballistic. It provides an alternative explanation of swarm traits previously attributed to inertia.

DOI: [10.1103/PhysRevLett.127.258001](https://doi.org/10.1103/PhysRevLett.127.258001)

Interacting assemblies of active elements ranging from neural networks in the brain to forest fires and bird flocks can exhibit scale-free behavior [1–4]. This might be indicative of an underlying powerful physical ordering principle overwriting their inherent complexity. Finite-size scaling theory [5] associates such behavior with a correlation length exceeding the system size and conjectured to arise from a mechanism called self-organized criticality [6]. It is indeed an appealing idea that simple interaction rules, when, e.g., individuals replicate actions of their neighbors, can drive a nonequilibrium ensemble toward criticality. Even though it does not generally seem to apply to both natural systems [7,8] and their models [9–11], studying the emergent finite-size scaling in natural assemblies is vital for their prospective modeling in the spirit of nonequilibrium many-body systems, as successful models should be required to reproduce the observed scaling [10] and correlations [12]. In this vein, the inertia spin model [12] was proposed to overcome known deficiencies of the classical Vicsek model (VM) [13] in comparison with empirical data for natural swarms and flocks. Inspired by observations of birds and insects [4], which cannot turn instantaneously, it adds inertia to the navigation rules for the individual motile spins and predicts dynamical scaling with exponent  $z = 1.5$  for a small-particle-velocity (“underdamped equilibrium”) regime and  $z = 1.3$  for a large-velocity (“underdamped off-equilibrium”) regime [14]. This brings the VM, with classical exponents  $z = 2$  and  $z = 1.7$ , respectively, closer to the dynamical scaling and time-correlation functions found in natural swarms of moderate size (dynamical exponent  $z \approx 1.1$ ) [15,16].

However, for motile ensembles, physical inertia can have quite similar effects as delayed reactions due to finite speeds of information transfer processing, and actuation

[17–19]. Such traits are indeed ubiquitously found in nature, from insects to birds, in various robotic systems [20–22] and are also thought to cause traffic jams [23]. Recent experiments [24,25] with feedback-driven artificial microswimmers [26] have moreover established their role in the naturally overdamped microscopic world of active Brownian particles such as bacteria, for which inertial effects are negligible. Beyond oscillatory behavior, which is also a common trait of inertial motion, time-delayed interactions can give rise to multistability, instabilities, and even chaos [19,27–29]. Conversely, intermediate time delays may facilitate clustering compared with the classical VM [28] and flocking in the Cucker–Smale model [30]. And recent indications that delay-dependent optimizations play a role in artificial microswimmer assemblies [25] seem reminiscent of the optimum run-and-tumble times of bacteria [31,32] or the improved localization achieved with feedback cooling [33,34] or feedback-driving of robots [35].

In this Letter, we demonstrate that the classical VM [13] with retarded reactions exhibits the same finite-size scaling and time correlations near the ordering transition as the inertia spin model [4]. This suggests that scaling and correlations similar to natural swarms can be expected for a wealth of systems with time-delayed interactions, including overdamped Brownian particle assemblies. We can also corroborate the observation that increasing delay times may have a nonmonotonic effect on the stability of coherent collective motion [28].

*Model.*—The (classical) VM [13] arguably is the simplest model for motile active-particle assemblies, ranging from bacteria to birds, and a central paradigm in the field of motile active matter [36–39]. In each discrete time step, all particles advance with the same constant speed  $v_0$ . And they instantaneously adapt their orientations to the previous

average orientation of their neighbors within an interaction sphere of radius  $R$ , up to some random error contributed by a local noise term. The orientation is thus clearly an overdamped variable, as it bears no inertia.

In the delay VM, depicted in Fig. S1 of the Supplemental Material [40], particle  $i$  adapts at time  $t$  to the mean orientation of all particles that had distance less than  $R$  from its previous position, at time  $t - 1 - \tau$ , with an integer-valued time delay  $\tau \geq 0$ . The discrete time step and the interaction radius  $R = 1$  serve as units of time and length, respectively. The dynamics of the standard VM is recovered for  $\tau = 0$ . The velocity  $\mathbf{v}_i$  and position  $\mathbf{r}_i$  of particle  $i$  in three spatial dimensions (3D) thus obey the set of equations [13]

$$\mathbf{v}_i(t+1) = v_0 \mathcal{R}_\alpha \Theta \left[ \mathbf{v}_i(t) + \sum_j n_{ij}(t-\tau) \mathbf{v}_j(t-\tau) \right], \quad (1)$$

$$\mathbf{r}_i(t+1) = \mathbf{r}_i(t) + \mathbf{v}_i(t+1). \quad (2)$$

The noise operator  $\mathcal{R}_\alpha X$  randomly rotates its argument  $X$  within a uniformly distributed solid angle  $4\pi\alpha$  centered around  $X$ , and  $\Theta(\mathbf{v}) \equiv \mathbf{v}/|\mathbf{v}|$  normalizes its argument. We assume geometric interactions corresponding to the connectivity matrix elements  $n_{ij}(t) = 1$  for  $i \neq j$  if  $r_{ij}(t) = |\mathbf{r}_i(t) - \mathbf{r}_j(t)| < R$ , and  $n_{ij}(t) = 0$  otherwise.

We simulated the delay VM with fixed speed  $v_0 = 0.05$  and noise strength  $\alpha = 0.45$  inside a cube with size  $L^3$  and periodic boundary conditions for six values of the particle number  $N = 2^n$ ,  $n = 6, \dots, 11$ . In this setting, we repeated the analysis performed in Refs. [15,49] for the static and dynamic scaling and the correlation functions of the standard VM, operating in its overdamped equilibrium regime [14], for delay times  $\tau = 0, \dots, 20$ . As control parameter, we prescribed the average nearest-neighbor distance  $r_1$  between the individual particles by varying  $L$ . Here, we present the main simulation results. Further data, technical details, and some analytical discussion can be found in the Supplemental Material [40].

The central object for our data analysis is the Fourier transformed spatiotemporal correlation function (CF)

$$C(k, t) = \left\langle \frac{1}{N} \sum_{i,j} \frac{\sin[kr_{ij}(t, t_0)]}{kr_{ij}(t, t_0)} \delta \hat{\mathbf{v}}_i(t_0) \cdot \delta \hat{\mathbf{v}}_j(t_0 + t) \right\rangle \quad (3)$$

of the normalized velocity fluctuations [4,15,40]

$$\delta \hat{\mathbf{v}}_i = \frac{\delta \mathbf{v}_i}{\sqrt{N^{-1} \sum_k \delta \mathbf{v}_k \cdot \delta \mathbf{v}_k}}, \quad (4)$$

where  $\delta \mathbf{v}_i = \mathbf{v}_i - \sum_k \mathbf{v}_k / N$  is the deviation of the velocity of particle  $i$  from the average velocity, and  $r_{ij}(t_0, t) = |\mathbf{r}_i(t_0) - \mathbf{r}_j(t)|$  is the distance between particles  $i$  and  $j$  at times  $t$  and  $t_0 < t$ . The average  $\langle \dots \rangle$  is taken over  $t_0$  [40].

*Static scaling.*—At  $t = 0$ ,  $C(k, 0)$  exhibits a global maximum at  $k = k^* \sim 1/\xi$ , where  $\xi$  corresponds to the correlation length. Assuming proportionality between

fluctuation and response, this value of the CF is interpreted as a susceptibility  $\chi \equiv C(k^*, 0)$  [4,15,40].

For given delay time  $\tau$  and particle number  $N$ , the susceptibility  $\chi$  exhibits a maximum  $\chi^* = \chi^*(\tau, N)$  as a function of the nearest-neighbour distance at  $r_1^* = r_1^*(\tau, N)$ . The system is found to be ordered (large average velocity) for  $r_1 < r_1^*$  and disordered (small average velocity) otherwise. For a given  $N$ , the susceptibility  $\chi^*$  at the transition decreases monotonically with growing  $\tau$  and eventually saturates [see Figs. 1(a) and 1(b) and, in the Supplemental Material, Figs. S2 and S3 [40]]. The equal-time orientation correlations are thus generally reduced for retarded as opposed to instantaneous interactions, which suggests that the sensitivity to external perturbations decreases accordingly. For sufficiently large  $\tau$  and  $N$ , the derivative of the susceptibility with respect to  $r_1$  abruptly increases at some  $r_1 < r_1^*$ ; see the vertical dotted line at  $r_1 \approx 0.4$  in Fig. 1(b). No such kink is observed for small  $\tau$  [40]. For a given  $\tau$  and large enough  $N$ , the susceptibility in the vicinity of the ordering transition [49] exhibits finite-size scaling according to Ref. [50]:

$$r_1^* \sim r_C + N^{-1/(3\nu)}, \quad (5)$$

$$\chi \sim N^{\gamma/(3\nu)}. \quad (6)$$

In other words, for any given  $\tau$ , the limiting location  $r_C = r_1^*(\tau, \infty)$  of the transition for large (infinite) particle numbers and the critical exponents  $\gamma$  and  $\nu$  of the susceptibility  $\chi \sim (r_1^* - r_C)^{-\gamma}$  and the correlation length  $\xi \sim (r_1^* - r_C)^{-\nu}$ , respectively, can all be extrapolated from a data collapse of the susceptibilities for different  $N$ . The procedure is illustrated in Figs. 1(c) and 1(d). The resulting exponents and  $r_C$  exhibit strong dependencies on  $\tau$ , which saturate as  $\tau v_0 / R \approx 1/2$ , when the advance during one delay time becomes comparable to the interaction radius [Figs. 1(e)–1(h)]. An analytical argument corroborates that a further increase of  $\tau$  should not significantly alter the qualitative physical picture [40]. For a particle of characteristic size 2.5 mm traveling with velocity 1 meter per second with an interaction radius of  $4 \times$  body length (10 mm), the condition  $\tau v_0 / R = 1/2$  implies a time delay of 5 ms, which are numbers roughly in accord with data available for fruit flies [51–54].

Since the static critical exponents in the standard VM are known to depend strongly on the density, speed, and interaction radius [56], their absolute values are of limited interest. Rather, their trends and dependencies are revealing. The critical nearest-neighbor distance  $r_C$  in Fig. 1(h), proportional to the critical density of the system, exhibits a pronounced maximum at  $\tau v_0 / R \approx 0.2$ , indicating that a system with an intermediate delay time favors order already at lower densities as compared with the system without delay. This somewhat counterintuitive result is in agreement with findings of Refs. [28,35] that intermediate delays stabilize collective motion. For larger delay times,



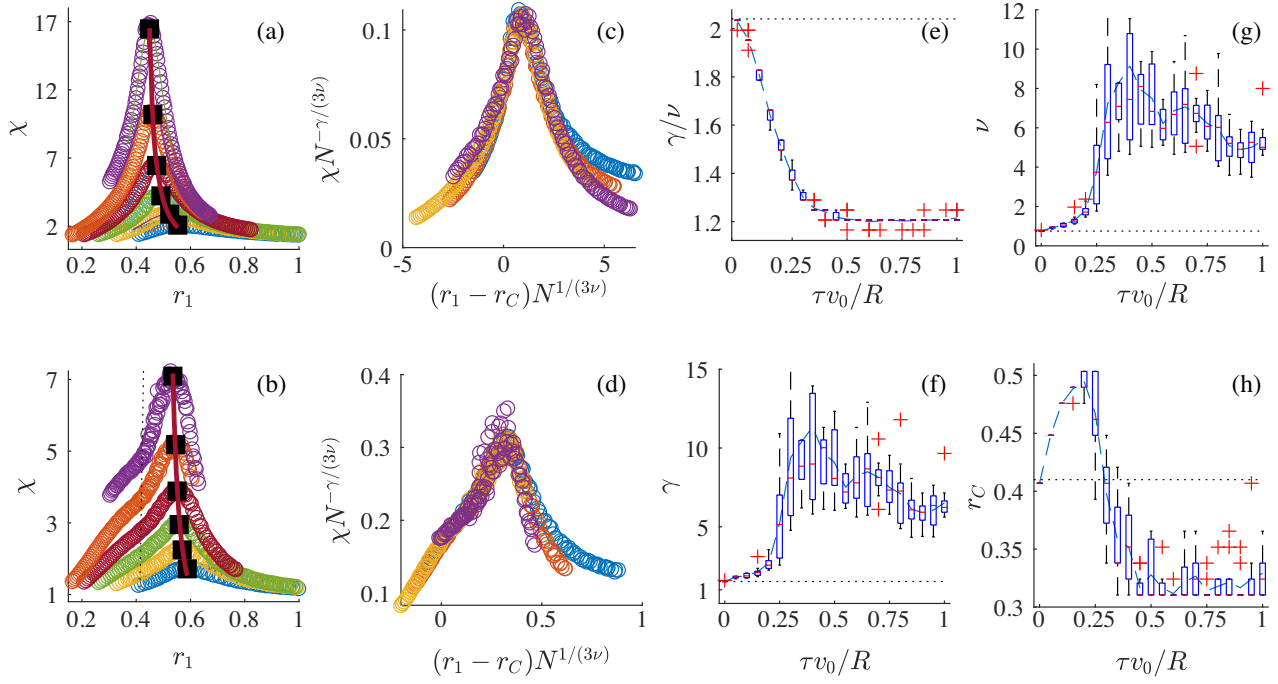


FIG. 1. Static finite-size scaling in the 3D delay VM. Each simulation departed from a random initial state and was evolved for a transient period of 1500 time steps before measurements started. (a),(b) The susceptibilities  $\chi \equiv \max_k C(k, 0)$  averaged over 11 trajectories of  $10^4$  time steps for  $N = 64, 128, 256, 512, 1024, 2048$  particles (from bottom to top) and delay times  $\tau = 0$  [(a), standard VM] and  $\tau = 15$  [(b),  $\tau v_0/R = 0.75$ ], respectively, over the mean nearest-neighbour distance  $r_1$ . (c),(d) The data collapse achieved for  $N \geq 256$ . (e)–(h) Box plots [55] of the exponent ratio  $\gamma/\nu$ , the individual exponents, and the extrapolated critical parameter  $r_C$  for  $N \rightarrow \infty$ , respectively, all for delay times  $\tau$  from 0 to 20. Broken dashed lines mark averages over the 11 realizations. The horizontal dotted lines depict the values  $\nu \approx 0.75$ ,  $\gamma \approx 1.53$ , and  $r_C \approx 0.41$  obtained for  $\tau = 0$ , where the model reduces to the standard VM, consistent with the data in Ref. [49]. In (a) and (b), black squares mark susceptibility maxima  $\chi(r_1^*)$  and solid lines are computed using the finite-size scaling relation  $\chi \sim (r_1^* - r_C)^{-\gamma}$  with parameters from (e)–(h). The vertical dotted line in (b) marks the abrupt changes in the slope of  $\chi$ .

the critical nearest-neighbor distance  $r_C$  drops sharply to a value below that for the standard VM. A possible explanation can be based on the decrease of the maximum susceptibility with delay time, shown in Sec. S2 in the Supplemental Material [40]. The susceptibility measures both the sensitivity to destabilizing perturbations and the ability to align in the flocking regime. The local maximum in  $r_C$  could thus originate from a trade-off between these tendencies, the increased resistance to fluctuations dominating at small  $\tau$  and the waning alignment at large  $\tau$ .

The exponents  $\nu$  and  $\gamma$  also display local maxima, but at somewhat larger  $\tau$ . Their saturation values are much higher than the respective critical exponents in all known universality classes, including the standard VM. To attribute the observed finite-size scaling to a critical point in the infinite-size limit according to the conventional scaling hypothesis [57] would require an extraordinarily sharp divergence of the correlation length and the susceptibility at criticality, which is approached extraordinarily slowly with increasing particle number  $N$  [see the solid lines in Figs. 1(a) and 1(b)]. As detailed in Sec. S1 of the Supplemental Material [40], we expect the observed scaling to hold whenever the density is approximately homogeneous, as it is the case

for intermediate  $N$ . Then, also, the standard VM shows the truly critical scaling of its incompressible variant [16,37], while, for very large  $N$ , it exhibits large density fluctuations leading to a discontinuous phase transition with microphase separation.

*Time correlation functions.*—The time dependence of the CFs [Eq. (3)] for the delay VM at the transition, quantifying the temporal loss of orientational correlations [4,15], is strongly influenced by the delay. Figure 2(a) shows that the normalized CFs  $\tilde{C}(t) \equiv C(k^*, t)/C(k^*, 0)$  acquire oscillations with period  $(\tau + 1)$  and an amplitude increasing with  $\tau$ . They reveal the transmission of orientational correlations over discrete time steps  $\tau + 1$  and can be understood analytically by a spin wave theory that accounts for the delay [40]. In Fig. 2(d), we show that logarithms of CFs for  $N = 2^n$ ,  $n = 8, \dots, 11$ , and  $\tau = 0$  collapse onto the master curve  $-t/\tau_R$  upon rescaling time by the relaxation times  $\tau_R$  obtained from Eq. (S14) in the Supplemental Material [40]. Figure 2(g) shows the corresponding increasingly negative time derivatives for  $t \rightarrow 0$ , indicating the exponential loss of correlation in the standard VM [15]. Due to the delay-induced superimposed oscillations, the initial slope of the CFs always steepens with increasing

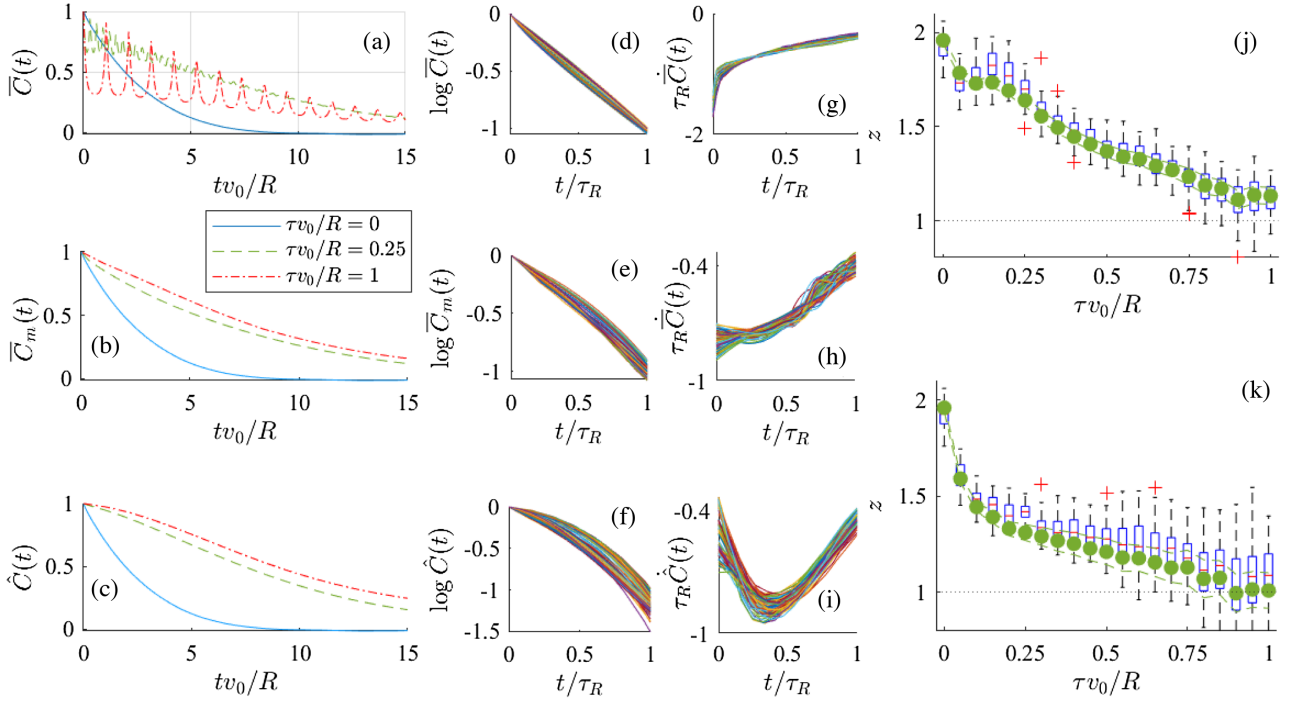


FIG. 2. Dynamical scaling of the orientational correlations at the susceptibility maximum displayed in Fig. 1. (a) Normalized time-correlation functions  $\bar{C}(t)$  for delays  $\tau = 0, 5, 20$  and  $N = 2048$ . (b) The upper envelopes  $\bar{C}_m$  of the curves in (a). (c) The correlation functions  $\hat{C}(t)$  calculated from trajectories (under)sampled with frequency  $1/(\tau + 1)$  for  $N = 2048$ . (d)–(f) The normalized CFs  $\bar{C}$ ,  $\bar{C}_m$ , and  $\hat{C}$  collapse upon measuring time in units of the relaxation time  $\tau_R$  for  $\tau = 0$  (d) and  $\tau = 20$  (e) and (f) and system sizes  $N = 2^n$ ,  $n = 8, \dots, 11$  (50 simulation runs for each system size). (g)–(i) The corresponding collapsed slopes. Box-plots [55] in (j) and (k) show the dynamical exponent  $z$  obtained from collapsing the CFs  $\bar{C}_m$  and  $\hat{C}$  by rescaling time as  $t\xi^z$ . Examples of the corresponding collapses are shown in Figs. S13 and S14 in the Supplemental Material [40]. The green circles are from linear fits to  $\log \tau_R(N) = -z \log k^*(N)$  for all 50 datasets with 95% confidence intervals of the fits (dashed).

delay time  $\tau$ . In contrast, the *overall* decay flattens with  $\tau$ , as revealed by the upper envelopes  $\bar{C}_m(t)$  of  $\bar{C}(t)$ , in Fig. 2(b). As intuitively expected, the delayed interactions thus tend to increase the memory in the VM. The data collapse of  $\bar{C}_m(t/\tau_R)$  in Fig. 2 confirms the nonexponential relaxation, while its slope in Fig. 2(h) still increases for  $t \rightarrow 0$  (see Sec. S5 in the Supplemental Material [40] for an approximate analytical result).

However, we note that the sampling rates used in practical measurements may not always be sufficient to resolve delay induced oscillations [40], which might moreover have a tendency to be washed out by a natural dispersion of the delay times. To account for undersampling, Fig. 2(c) shows normalized CFs  $\hat{C}(t)$  calculated from particle positions that were (under-)sampled with frequency  $1/(\tau + 1)$ , i.e., we calculated the corresponding velocities as  $\mathbf{v}_i(t) = [\mathbf{r}_i(t) - \mathbf{r}_i(t - \tau - 1)]/(\tau + 1)$ . In Fig. S10 of the Supplemental Material [40], we show that the under-sampled CFs are independent of the sampling rate as long as it is comparable to or smaller than  $1/\tau$ . The resulting CFs are shown in Fig. 2(c). The exponential initial decay for the vanishing delay time  $\tau = 0$  is seen to increasingly flatten with growing  $\tau$ , as also corroborated by the data collapse in

Fig. 2(f). For  $v_0\tau/R \gtrsim 1/2$ , the absolute slope  $|\dot{\hat{C}}(t)|$  starts to decrease for  $t \rightarrow 0$  as shown in Fig. 2(i) and in Figs. S7 and S9 in the Supplemental Material [40]. The under-sampled delay VM yields qualitatively the same relaxation of orientational correlations as observed for natural swarms [15]. It thus provides an alternative explanation for the data, which were so far interpreted within the inertia spin model [58]. The dynamics induced by a discrete time delay appears to be more prone to developing oscillatory patterns, such as the ones illustrated in Fig. 2(a), though.

*Dynamical scaling.*—The dynamical scaling hypothesis [57] states that the relaxation time  $\tau_R$  diverges with the correlation length as  $\tau_R \sim \xi^z \sim (k^*)^{-z}$ , at a critical point. Directly fitting the relation

$$\log \tau_R(N) = -z \log k^*(N) \quad (7)$$

for the relaxation times of  $\bar{C}_m$  and  $\hat{C}$  as functions of time delay  $\tau$  yields the dynamical exponent  $z$  as depicted by circles in Figs. 2(j) and 2(k), respectively. The figures also show box plots [55] resulting from the best data collapse of CFs  $\bar{C}_m$  and  $\hat{C}$  using  $(k^*)^{-z}$  with  $z$  as a free parameter in place

of the relaxation time  $\tau_R$ , itself. The exponents  $z$  obtained from these two approaches nicely agree for each CF. While the values for  $z$  obtained from the two alternative CFs differ, they exhibit the same robust trend: a crossover from the overdamped equilibrium to the underdamped off-equilibrium signature [14,40], i.e., from  $z \approx 2$  for  $\tau = 0$  to about  $z \approx 1.1$  for  $\tau v_0/R \approx 1$  [59]. Our analysis thus suggests that increasing the delay time drives the system further from equilibrium as if one had effectively increased the particle speed  $v_0$ . Analytical arguments corroborate this, at least for large  $\tau$  (Sec. S6 in the Supplemental Material [40]). Our results can therefore reconcile the standard VM predictions with the observations for natural swarms.

*Conclusion.*—We analyzed the VM with retarded interactions, as they are expected from natural delays between sensing and reaction. It provides an alternative to the rotational inertia hypothesis for reconciling the discrepancies in the dynamical scaling and relaxation between the standard VM and natural swarms [4,58]. While the navigation of insects and other flying species is certainly influenced both by inertia and time delay, our focus on delay could help to better understand their relation to feedback-driven robotic [35] or microparticle [24,25,60] swarms. Especially in the latter, current experimental techniques [26] allow inertial effects to be suppressed so that only the unavoidable time delay remains. While our analysis proves that the delay VM is compatible with finite-size scaling for the investigated system sizes, it raises many questions. Will larger systems exhibit a discontinuous transition with a phase separation, as in the standard VM? Will it mask the finite-size signatures of a continuous phase transition for practically relevant particle numbers? How does the phase diagram depend on the delay time? We hope to address some of these questions in the future.

We acknowledge funding through a DFG-GACR cooperation by the Deutsche Forschungsgemeinschaft (DFG Project No. 432421051) and by the Czech Science Foundation (GACR Project No. 20-02955J). V.H. was supported by the Humboldt Foundation. D.G. acknowledges funding by International Max Planck Research Schools (IMPRS). The time-correlation functions were computed at the GPU-Cluster Clara at the Scientific Computing Center in Leipzig. We are grateful to Lokrshi Dadhichi for his helpful remarks concerning the final manuscript and to Kiril Panayotov Blagoev for support during a preliminary stage of the project.

\* viktor.holubec@mff.cuni.cz

- [1] T. Mora and W. Bialek, Are biological systems poised at criticality?, *J. Stat. Phys.* **144**, 268 (2011).
- [2] A.-L. Barabási and E. Bonabeau, Scale-free networks, *Sci. Am.* **288**, 60 (2003).
- [3] M. A. Munoz, Colloquium: Criticality and dynamical scaling in living systems, *Rev. Mod. Phys.* **90**, 031001 (2018).

- [4] A. Cavagna, I. Giardina, and T. S. Grigera, The physics of flocking: Correlation as a compass from experiments to theory, *Phys. Rep.* **728**, 1 (2018).
- [5] M. Plischke and B. Bergersen, *Equilibrium Statistical Physics (3rd Edition)* (World Scientific Publishing Company, Singapore, 2006).
- [6] P. Bak, C. Tang, and K. Wiesenfeld, Self-organized criticality, *Phys. Rev. A* **38**, 364 (1988).
- [7] A. Clauset, C. R. Shalizi, and M. E. J. Newman, Power-law distributions in empirical data, *SIAM Rev.* **51**, 661 (2009).
- [8] J. Jhawar, R. G. Morris, U. R. Amith-Kumar, M. Danny Raj, T. Rogers, H. Rajendran, and V. Guttal, Noise-induced schooling of fish, *Nat. Phys.* **16**, 488 (2020).
- [9] G. Pruessner and H. J. Jensen, Broken scaling in the forest-fire model, *Phys. Rev. E* **65**, 056707 (2002).
- [10] L. Palmieri and H. J. Jensen, The forest fire model: The subtleties of criticality and scale invariance, *Front. Phys.* **8**, 257 (2020).
- [11] D. Martin, H. Chaté, C. Nardini, A. Solon, J. Tailleur, and F. Van Wijland, Fluctuation-Induced Phase Separation in Metric and Topological Models of Collective Motion, *Phys. Rev. Lett.* **126**, 148001 (2021).
- [12] A. Cavagna, L. Di Carlo, I. Giardina, L. Grandinetti, T. S. Grigera, and G. Pisegna, Dynamical Renormalization Group Approach to the Collective Behavior of Swarms, *Phys. Rev. Lett.* **123**, 268001 (2019).
- [13] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, Novel Type of Phase Transition in a System of Self-Driven Particles, *Phys. Rev. Lett.* **75**, 1226 (1995).
- [14] A. Cavagna, L. D. Carlo, I. Giardina, T. S. Grigera, S. Melillo, L. Parisi, G. Pisegna, and M. Scandolo, Natural swarms in 3.99 dimensions, [arXiv:2107.04432](https://arxiv.org/abs/2107.04432).
- [15] A. Cavagna, D. Conti, C. Creato, L. Del Castello, I. Giardina, T. S. Grigera, S. Melillo, L. Parisi, and M. Viale, Dynamic scaling in natural swarms, *Nat. Phys.* **13**, 914 (2017).
- [16] A. Cavagna, L. Di Carlo, I. Giardina, T. S. Grigera, and G. Pisegna, Equilibrium to off-equilibrium crossover in homogeneous active matter, *Phys. Rev. Research* **3**, 013210 (2021).
- [17] M. Nagy, Z. Ákos, D. Biro, and T. Vicsek, Hierarchical group dynamics in pigeon flocks, *Nature (London)* **464**, 890 (2010).
- [18] A. Attanasi, A. Cavagna, L. Del Castello, I. Giardina, T. S. Grigera, A. Jelić, S. Melillo, L. Parisi, O. Pohl, E. Shen, and M. Viale, Information transfer and behavioural inertia in starling flocks, *Nat. Phys.* **10**, 691 (2014).
- [19] D. Geiss, K. Kroy, and V. Holubec, Brownian molecules formed by delayed harmonic interactions, *New J. Phys.* **21**, 093014 (2019).
- [20] G. Vásárhelyi, C. Virágh, G. Somorjai, N. Tarcai, T. Szörényi, T. Nepusz, and T. Vicsek, Outdoor flocking and formation flight with autonomous aerial robots, in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IEEE, New York, 2014)*, pp. 3866–3873, <https://dx.doi.org/10.1109/IROS.2014.6943105>.
- [21] C. Virágh, G. Vásárhelyi, N. Tarcai, T. Szörényi, G. Somorjai, T. Nepusz, and T. Vicsek, Flocking algorithm for autonomous flying robots, *Bioinspiration Biomimetics* **9**, 025012 (2014).
- [22] G. Vásárhelyi, C. Virágh, G. Somorjai, T. Nepusz, A. E. Eiben, and T. Vicsek, Optimized flocking of autonomous drones in confined environments, *Sci. Rob.* **3**, eaat3536 (2018).

- [23] L. Davis, Modifications of the optimal velocity traffic model to include delay due to driver reaction time, *Physica (Amsterdam)* **319A**, 557 (2003).
- [24] U. Khadka, V. Holubec, H. Yang, and F. Cichos, Active particles bound by information flows, *Nat. Commun.* **9**, 3864 (2018).
- [25] S. Muiños-Landin, A. Fischer, V. Holubec, and F. Cichos, Reinforcement learning with artificial microswimmers, *Sci. Rob.* **6**, eabd9285 (2021).
- [26] M. Fränzl, S. Muiños-Landin, V. Holubec, and F. Cichos, Fully steerable symmetric thermoplasmonic microswimmers, *ACS Nano* **15**, 3434 (2021).
- [27] E. Forgoston and I. B. Schwartz, Delay-induced instabilities in self-propelling swarms, *Phys. Rev. E* **77**, 035203(R) (2008).
- [28] R. Piwowarczyk, M. Selin, T. Ihle, and G. Volpe, Influence of sensorial delay on clustering and swarming, *Phys. Rev. E* **100**, 012607 (2019).
- [29] S. A. M. Loos and S. H. L. Klapp, Heat flow due to time-delayed feedback, *Sci. Rep.* **9**, 2491 (2019).
- [30] R. Erban, J. Hakovec, and Y. Sun, A cucker-smale model with noise and delay, *SIAM J. Appl. Math.* **76**, 1535 (2016).
- [31] P. Romanczuk and G. Salbreux, Optimal chemotaxis in intermittent migration of animal cells, *Phys. Rev. E* **91**, 042720 (2015).
- [32] A. Diz-Muñoz, P. Romanczuk, W. Yu, M. Bergert, K. Ivanovitch, G. Salbreux, C.-P. Heisenberg, and E. K. Paluch, Steering cell migration by alternating blebs and actin-rich protrusions, *BMC Biol.* **14**, 74 (2016).
- [33] P. Bushev, D. Rotter, A. Wilson, F. Dubin, C. Becher, J. Eschner, R. Blatt, V. Steixner, P. Rabl, and P. Zoller, Feedback Cooling of a Single Trapped Ion, *Phys. Rev. Lett.* **96**, 043003 (2006).
- [34] D. Goldwater, B. A. Stickler, L. Martinetz, T. E. Northup, K. Hornberger, and J. Millen, Levitated electromechanics: All-electrical cooling of charged nano- and micro-particles, *Quantum Sci. Technol.* **4**, 024003 (2019).
- [35] M. Mijalkov, A. McDaniel, J. Wehr, and G. Volpe, Engineering Sensorial Delay to Control Phototaxis and Emergent Collective Behaviors, *Phys. Rev. X* **6**, 011008 (2016).
- [36] F. Ginelli, The physics of the vicsek model, *Eur. Phys. J. Spec. Top.* **225**, 2099 (2016).
- [37] L. Chen, J. Toner, and C. F. Lee, Critical phenomenon of the order-disorder transition in incompressible active fluids, *New J. Phys.* **17**, 042002 (2015).
- [38] L. Chen, C. F. Lee, and J. Toner, Incompressible polar active fluids in the moving phase in dimensions  $d > 2$ , *New J. Phys.* **20**, 113035 (2018).
- [39] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe, Active particles in complex and crowded environments, *Rev. Mod. Phys.* **88**, 045006 (2016).
- [40] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.127.258001> containing additional Refs. [41–48], for additional numerical and theoretical results.
- [41] D. R. J. Laming, *Information Theory of Choice-Reaction Times* (Academic Press, New York, 1968).
- [42] P. C. Hohenberg and B. I. Halperin, Theory of dynamic critical phenomena, *Rev. Mod. Phys.* **49**, 435 (1977).
- [43] A. T. Welford, *Reaction Times* (Academic Press, New York, 1980).
- [44] R. C. Eaton, R. A. Bombardieri, and D. L. Meyer, The mauthner-initiated startle response in teleost fish, *J. Exp. Biol.* **66**, 65 (1977).
- [45] R. C. Eaton, *Neural Mechanisms of Startle Behavior* (Springer Science & Business Media, 1984).
- [46] P. Lenz and D. Hartline, Reaction times and force production during escape behavior of a calanoid copepod, *undinula vulgaris*, *Mar. Biol.* **133**, 249 (1999).
- [47] H. Pomeroy and F. Heppner, Laboratory determination of startle reaction time of the starling (*sturnus vulgaris*), *Anim. Behav.* **25**, 720 (1977).
- [48] T. D. Frank, P. J. Beek, and R. Friedrich, Fokker-planck perspective on stochastic delay systems: Exact solutions and data analysis of biological systems, *Phys. Rev. E* **68**, 021912 (2003).
- [49] A. Attanasi, A. Cavagna, L. Del Castello, I. Giardina, S. Melillo, L. Parisi, O. Pohl, B. Rossaro, E. Shen, E. Silvestri, and M. Viale, Finite-Size Scaling as a Way to Probe Near-Criticality in Natural Swarms, *Phys. Rev. Lett.* **113**, 238102 (2014).
- [50] M. E. Fisher and M. N. Barber, Scaling Theory for Finite-Size Effects in the Critical Region, *Phys. Rev. Lett.* **28**, 1516 (1972).
- [51] S. Vogel, Flight in drosophila: I. Flight performance of tethered flies, *J. Exp. Biol.* **44**, 567 (1966).
- [52] T. Beatus, J. M. Guckenheimer, and I. Cohen, Controlling roll perturbations in fruit flies, *J. R. Soc. Interface* **12**, 20150075 (2015).
- [53] L. Ristroph, A. J. Bergou, G. Ristroph, K. Coumes, G. J. Berman, J. Guckenheimer, Z. J. Wang, and I. Cohen, Discovering the flight autostabilizer of fruit flies by inducing aerial stumbles, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4820 (2010).
- [54] L. Ristroph, G. Ristroph, S. Morozova, A. J. Bergou, S. Chang, J. Guckenheimer, Z. J. Wang, and I. Cohen, Active and passive stabilization of body pitch in insect flight, *J. R. Soc. Interface* **10**, 20130237 (2013).
- [55] Box-plot depicts statistics on data sets. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the '+' symbol. (Taken from MATLAB documentation.)
- [56] D. S. Cambui, A. S. de Arruda, and M. Godoy, Critical exponents of a self-propelled particles system, *Physica (Amsterdam)* **444A**, 582 (2016).
- [57] B. Halperin and P. Hohenberg, Scaling laws for dynamic critical phenomena, *Phys. Rev.* **177**, 952 (1969).
- [58] A. Cavagna, D. Conti, I. Giardina, T. S. Grigera, S. Melillo, and M. Viale, Spatio-temporal correlations in models of collective motion ruled by different dynamical laws, *Phys. Biol.* **13**, 065001 (2016).
- [59] The saturation of  $z(\tau)$  for large  $\tau$  follows from that of  $\xi(\tau)$  (implied by the saturation of the finite-size scaling parameters in Fig. 1) and  $\tau_R(\tau)$  (Fig. S12 [40]).
- [60] F. A. Lavergne, H. Wendehenne, T. Bäuerle, and C. Bechinger, Group formation and cohesion of active particles with visual perception-dependent motility, *Science* **364**, 70 (2019).

# Supplementary Information: Finite-size scaling at the edge of disorder in a time-delay Vicsek model

Viktor Holubec,<sup>1,2,\*</sup> Daniel Geiss,<sup>1</sup> Sarah A. M. Loos,<sup>1</sup> Klaus Kroy,<sup>1</sup> and Frank Cichos<sup>3</sup>

<sup>1</sup>*Institut für Theoretische Physik, Universität Leipzig, Postfach 100 920, D-04009 Leipzig, Germany*

<sup>2</sup>*Charles University, Faculty of Mathematics and Physics,*

*Department of Macromolecular Physics, V Holešovičkách 2, CZ-180 00 Praha, Czech Republic*

<sup>3</sup>*Peter Debye Institute for Soft Matter Physics, Universität Leipzig, 04103 Leipzig, Germany.*

(Dated: March 17, 2022)

Here we provide more theoretical background for the discussion in the main text as well as a full set of simulation results. This includes a review of critical behavior of Vicsek-type models in Sec. S1, simulation details in Sec. S2, definitions of correlation functions and their relation to susceptibility in Sec. S3, and details of the finite-size static and dynamical scaling in the delay Vicsek model (Secs. S4 and S5). In Sec. S6, we provide an analytical evidence for the saturation of the behavior of the delay Vicsek model for large delay times. The final section presents exact results obtained using an analytically solvable linearised version of the delay Vicsek model.

PACS numbers: 05.20.-y, 05.70.Ln, 07.20.Pe

## S1. CRITICALITY AND SCALING IN VICSEK-TYPE MODELS

Simulations of very large systems show that neither the variant of the Vicsek model (VM) with topological interaction nor the metric one is truly critical [1]. The situation is reminiscent of the forest-fire model, which was long believed to be critical [2, 3]. It turns out that fluctuations in the VM induce a density-dependent shift of the onset of order predicted from models without noise, which changes the nature of the order–disorder transition from continuous (second-order) to a discontinuous (first-order) phase-separation scenario. Concerning the VM with rotational inertia (inertia spin model - ISM)

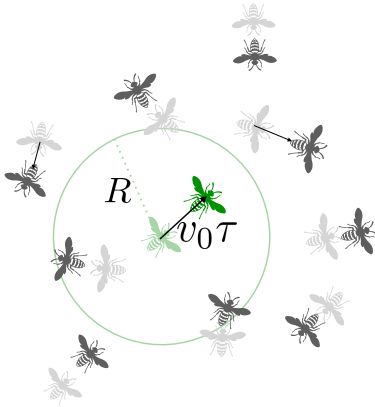


FIG. S1. Sketch of the delayed Vicsek model defined by Eqs. (1) and (2) in the main text (also Eqs. (S19) and (S20) here). The conventional alignment rule with interaction range  $R$  is executed with a time delay  $\tau$ .

\* viktor.holubec@gmail.com

and the delay VM introduced in the main text, in-depth investigations concerning the existence and type of the phase transition have not been performed yet. However, it is reasonable to assume that the density fluctuations will, for very large systems, again lead to a similar phenomenology. Further, it is plausible that, as for the standard VM, the discontinuous transition could be avoided and criticality restored if density fluctuations were suppressed [4, 5].

Computer simulations can currently address large enough systems that boundary effects can be safely neglected. But reaching systems sizes that allow for a detailed study of the mentioned large-scale density fluctuations remains challenging. Systems that are not large enough to admit large density fluctuations are therefore likely to exhibit finite-size scaling indicative of a second-order transition [6, 7]. The corresponding ‘critical’ exponents can be determined either by directly applying the finite-size scaling predictions to data from computer simulations of such systems or from a renormalization group analysis of their incompressible variants. This way it was found that the dynamical critical exponent  $z$  as a function of agent speed  $v_0$  and inertia in the VM generalized to the ISM exhibits an interesting phenomenology:

1. Low speed  $v_0$  and overdamped dynamics (negligible inertia) lead to  $z = 2$ . This exponent is thus observed in the standard VM if the speed is too low to break free from the equilibrium Heisenberg universality class. In the main text, we recover this limit for vanishing delay time  $\tau$ .
2. High speed  $v_0$  and overdamped dynamics lead to  $z = 1.7$ . This is the regime reached in the standard VM when driven far from equilibrium and dominated by large frictional losses. It cannot be observed for our choice of parameters.
3. Low speed  $v_0$  and underdamped dynamics lead to

$z = 1.5$ . This regime corresponds to a ballistic motion with negligible friction. It pertains to the equilibrium Heisenberg model with spin inertia and extends to the ISM with agents moving at low speed. We do not observe it in our simulations.

4. Large speed  $v_0$  and underdamped dynamics lead to  $z = 1.3$ . In this regime, frictional losses are significant and the system is far from equilibrium. This behavior occurs for the ISM with fast agents and also in the delay VM, introduced in the main text, for large delay times  $\tau$ .

In accord with intuition from previous studies [8], time delay induces a similar behavior as inertia and thus the shift from overdamped to underdamped behavior with an increasing delay could have been expected. An intuitive explanation of why an increase of the time delay also drives the system further from equilibrium, corresponding to an effective increase in  $v_0$ , is given in Sec. S6.

## S2. SIMULATION DETAILS

We simulated the delay VM, sketched in Fig. S1, inside a three-dimensional cube of size  $L^3$  and periodic boundary conditions. In the simulations, the initial positions and orientations of the  $N$  individuals were drawn from a uniform distribution and we let the system evolve for an ‘equilibration period’ of 1500 time steps. Then, we collected data for further processing for another  $T = 10^4$  time-steps. The static exponents were obtained by averaging over 50 runs and the dynamical exponent by averaging over 11 runs of the simulation for each delay and particle number.

For all simulations, we fixed the angular noise strength  $\alpha = 0.45$  rad and the speed  $v_0 = 0.05$ , and used the interaction radius as our length unit, i.e., we set  $R = 1$ . To get a direct comparison with results obtained in Ref. [9], we used as control parameter the mean nearest-neighbor distance

$$r_1 = \left\langle \min_j |\mathbf{r}_i(t) - \mathbf{r}_j(t)| \right\rangle_{i,t} \\ = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \min_j |\mathbf{r}_i(t) - \mathbf{r}_j(t)|. \quad (\text{S1})$$

adjusted by changing the linear system size  $L$ .

In this setting, we first detected the position  $r_1^*$  of the order-disorder transition for each of the six values  $N = 2^n$ ,  $n = 6, \dots, 11$  of particle number and twenty-one values  $\tau = 0, \dots, 21$  of delay time. We based the evaluation of the exponents on the data collapse (as predicted by finite-size scaling, see e.g. Sec. S4 B), rather than a direct fitting of the exponents to scaling relations, as it was done in Refs. [9–11]. The reason is that the range of particle numbers considered here leads to large fitting errors in the latter approach, even though the overall trends

are the same as using the data collapse. Due to the immense computational costs, simulating also the systems of  $2^{12}$  and  $2^{13}$  particles for all the delay times is currently beyond our reach. To obtain the presented data, we already used several months of CPU and GPU time at the university’s computing cluster.

## S3. CORRELATION FUNCTION $C(k, t)$

Following the analysis in Ref. [11], the main theoretical tool used in our discussion is the correlation function (CF)  $C(k, t)$  defined in Eq. (3) in the main text. Roughly speaking, it quantifies to which extent particles at different locations and different times behave similarly. These correlations are induced by interactions between the individual particles and thereby by an exchange of information. To ensure that the correlations are not dominated by the average velocity of the system  $\frac{1}{N} \sum_k \mathbf{v}_k$ , which is for finite particle numbers  $N$  at the point of phase transition nonzero, we consider the correlations of the dimensionless velocity fluctuations

$$\delta \hat{\mathbf{v}}_i = \frac{\delta \mathbf{v}_i}{\sqrt{1/N \sum_k \delta \mathbf{v}_k \cdot \delta \mathbf{v}_k}}, \quad (\text{S2})$$

with

$$\delta \mathbf{v}_i = \mathbf{v}_i - \frac{1}{N} \sum_k \mathbf{v}_k. \quad (\text{S3})$$

The spatio-temporal CF is defined by

$$C(r, t) = \frac{1}{N4\pi r^2 \rho} \left\langle \sum_{i,j} \delta \hat{\mathbf{v}}_i(t_0) \cdot \delta \hat{\mathbf{v}}_j(t_0 + t) \right. \\ \left. \times \delta_D(r - r_{ij}(t_0, t)) \right\rangle \quad (\text{S4})$$

and measures correlations of velocity fluctuations at different points in space and time. Here,  $\rho = N/L^3$  is the average particle density, and  $r_{ij}(t_0, t) = |\mathbf{r}_i(t_0) - \mathbf{r}_j(t_0 + t)|$  is the distance between particles  $i$  and  $j$  at different times. The average is taken with respect to the earlier time,  $t_0$ , according to

$$\langle f(t_0, t_0 + t) \rangle \equiv \frac{1}{T - t} \sum_{t_0=1}^{T-t} f(t_0, t_0 + t). \quad (\text{S5})$$

Due to the normalization  $N4\pi r^2 \rho$ , representing the average number of particles in the neighborhood of radius  $r$ , the CF (S4) is dimensionless.

For our purposes, it is advantageous to use instead of  $C(r, t)$  its Fourier transform  $C(k, t) \equiv \rho \int d\mathbf{r} e^{i\mathbf{k}\mathbf{r}} C(r, t)$ . In 3D, it is given by Eq. (3) in the main text, which reads

$$C(k, t) = \left\langle \frac{1}{N} \sum_{i,j} \frac{\sin[kr_{ij}(t, t_0)]}{kr_{ij}(t, t_0)} \delta \hat{\mathbf{v}}_i(t_0) \cdot \delta \hat{\mathbf{v}}_j(t_0 + t) \right\rangle. \quad (\text{S6})$$

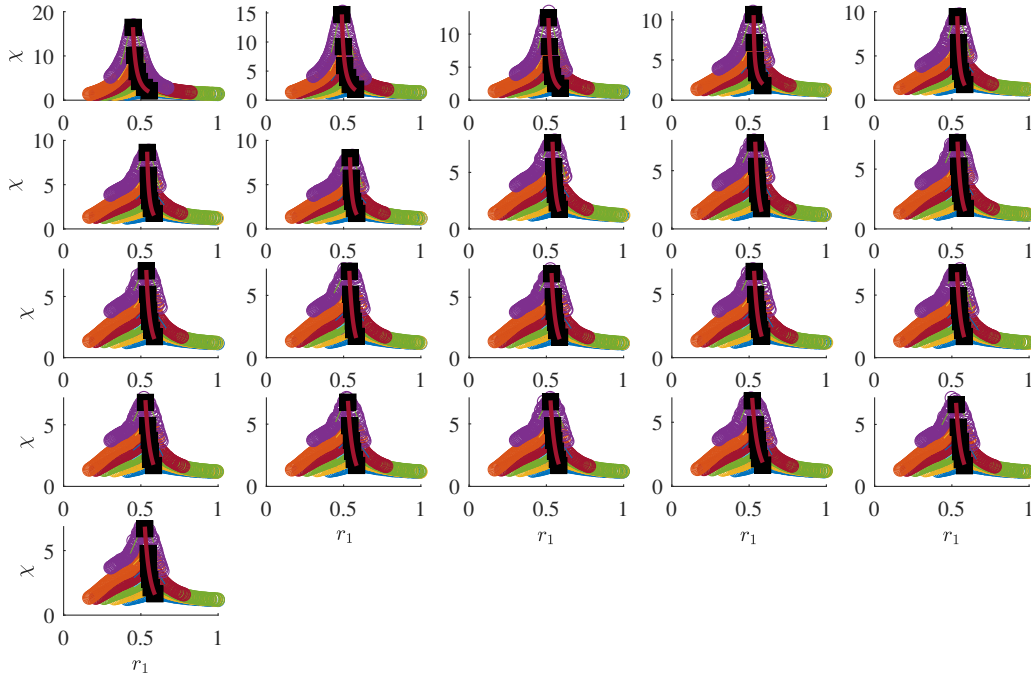


FIG. S2. Susceptibilities  $\chi$  as functions of the control parameter  $r_1$  for systems with different delay times  $\tau = 0, \dots, 20$  (from top-left to bottom-right panel) and particle numbers  $N = 64, 128, 256, 512, 1024$ , and  $2048$  (from bottom to top data set in each panel). Each shown data set was obtained by averaging over 11 simulation runs with the same parameters. The maximum susceptibilities are highlighted by black squares interconnected by solid lines. Their positions  $r_1^*$  mark the order-disorder transition of the VM. For given particle number, the susceptibility decreases with increasing delay time up to  $\tau \approx 10$ , when the susceptibilities saturate.

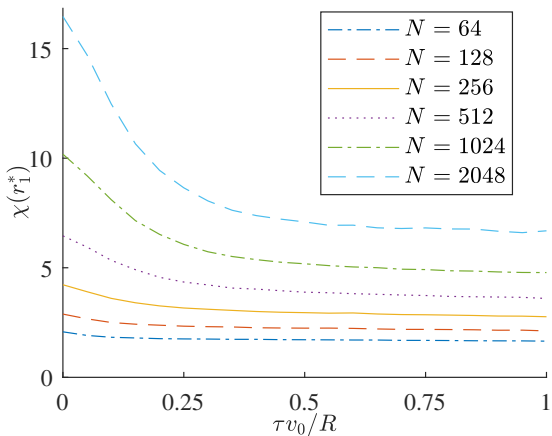


FIG. S3. Maximum susceptibilities from Fig. S2 as functions of the delay for all simulated system sizes. Their decline shows that the strength of correlations in the system decreases with increasing delay up to  $\tau v_0/R \approx 0.5$ , when it saturates.

#### S4. STATIC CORRELATIONS

In this section, we briefly review definitions of the correlation length, susceptibility, and the finite-size scaling theory given in Ref. [10]. Furthermore, we provide Figs. S2–S4 showing all susceptibilities obtained for the considered particle numbers  $N$  and delay times  $\tau$ .

##### A. Static correlation functions and susceptibility

The correlation length  $\xi$  measures the decay of correlations in the system with increasing distance and it is often given by an exponential decay rate. However, for scale-free systems, there is no such characteristic length scale and it is reasonable to define  $\xi$  as the smallest length for which the equal time CF  $C(r, 0)$  vanishes. Particles farther apart than  $\xi$  can be both correlated and anti-correlated. The integral  $\int_0^r dr' C(r', 0)$  over the CF thus reaches its maximum for  $r = \xi$ . This maximum measures the total correlation in the system and is, in the spirit of equilibrium fluctuation-dissipation relations, tentatively identified as its susceptibility,  $\chi$  [11].

The static CF in the Fourier domain,  $C(k, 0)$ , follows

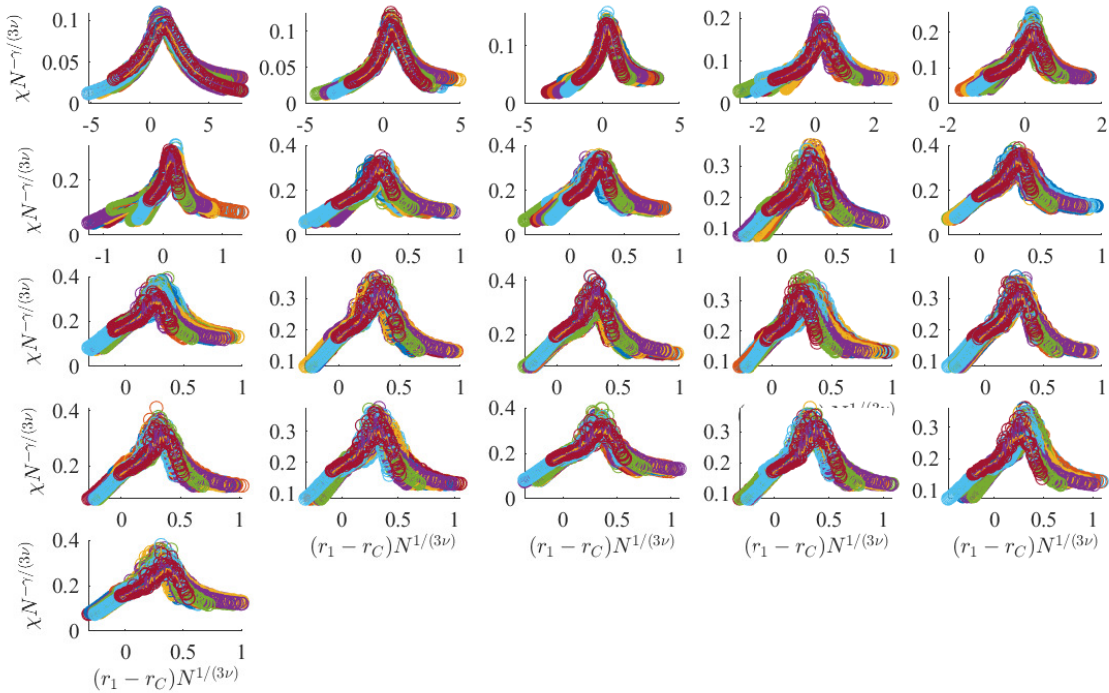


FIG. S4. Data collapse according to Eqs. (S8) and (S9) of all susceptibilities obtained from the 11 simulation runs for particle numbers  $N = 256, 512, 1024,$  and  $2048$  (different data sets in the individual panels) and delay times  $\tau = 0, \dots, 20$  (from top-left to bottom-right panel). These data collapses were used for determination of the exponents  $\gamma$  and  $\nu$  and the critical value of the order parameter  $r_C$  shown in Fig. 1 in the main text.

from the behavior of  $C(r, 0)$  after realising that  $k \sim 1/r$ . For large wave vectors  $k$  only the terms with  $i = j$  contribute to the sum (S6) and thus  $\lim_{k \rightarrow \infty} C(k, 0) = 1$ . For small  $k$ , a large number of uncorrelated terms contribute to the sum and thus  $\lim_{k \rightarrow 0} C(k, 0) = 0$ . The CF  $C(k, 0)$  attains its global maximum for  $k = k^* \sim 1/\xi$ , which corresponds to the integration of  $C(r, 0)$  up to  $r = \xi$  (for a plot of  $C(k, 0)$ , see Ref. [12]). The height of the maximum is thus used as a measure for the susceptibility  $\chi$ . To evaluate the correlation length and susceptibility from the position and height of global maximum of  $C(k, 0)$  is technically easier than from  $C(r, 0)$ . In the following, we use the definition

$$\chi \equiv \max_k C(k, 0). \quad (\text{S7})$$

Susceptibilities as functions of the nearest neighbour distance obtained for the considered system sizes  $N$  and delay times  $\tau$  are shown in Fig. S2. Positions of the maxima in the susceptibilities,  $r_1^*$ , mark the phase transition between the ordered (large mean velocity of the system) and disordered (small mean velocity) phase. With increasing delay, the particles are less correlated as witnessed by the decrease in  $\chi(r_1^*)$ , plotted in Fig. S3. The delay dependence of susceptibilities saturates for  $\tau \gtrsim 10$ .

## B. Finite-size scaling

The finite-size scaling hypothesis [13] predicts that systems with correlation length,  $\xi$ , on the order of the system size,  $L$ , exhibit a specific scaling of susceptibility and correlation length with the particle number. For all considered delay times  $\tau$  and particle numbers  $N$ , the delay VM exhibits large correlations at the point of the phase transition,  $r_1^*$ . The corresponding finite-size scaling reads

$$r_1^* \sim r_C + N^{-1/(3\nu)}, \quad (\text{S8})$$

$$\chi \sim N^{\gamma/(3\nu)}, \quad (\text{S9})$$

where  $r_C$  is the position of the putative transition in the thermodynamic limit of infinite  $N$ , and  $\nu$  and  $\gamma$  denote the associated critical exponents. The scaling relations (S8) and (S9) define the data collapse of the susceptibilities obtained for different  $N$  and the same  $\tau$ , shown in Fig S4. The values for  $r_C$ ,  $\nu$ , and  $\gamma$  shown in Fig 1 of the main text are those that yield the best data collapse.

For systems with second order phase transition in the thermodynamic limit, the exponents correspond to the scaling of the correlation length,

$$\xi \sim (r_1 - r_C)^{-\nu}, \quad (\text{S10})$$



and the susceptibility ,

$$\chi \sim (r_1 - r_C)^{-\gamma}, \quad (\text{S11})$$

as the control parameter  $r_1$  reaches its critical value  $r_C$ . However, as reviewed in Sec. S1, both the metric and the topological VM exhibit a discontinuous transition, where the scaling (S8) and (S9) eventually break down for very large  $N$ . In the thermodynamic limit, both variants of the VM exhibit a discontinuous phase separation into periodically altering ordered and disordered regions [14], unless the system is made incompressible [4, 5]. The size of these regions imposes an upper bound on the correlation length and thus on the validity of the relation (S10). In any case, for our finite systems, one can still understand the exponents  $\nu$  and  $\gamma$  as scaling exponents for correlation length and susceptibility if  $r_1$  in Eqs. (S10) and (S11) is substituted by  $r_1^* = r_1^*(N)$  and the scaling relations are understood as finite-size scaling relations of  $\xi$  and  $\chi$  with the particle number  $N$  at the transition. The scaling relation for the maximum susceptibility and different delay times is depicted by solid lines in Fig. S2. At first glance, the curves are steeper for larger delays, which results in the large values of the exponent  $\gamma$  for large  $\tau$  shown in Fig. 1 in the main text.

## S5. DYNAMICAL CORRELATIONS

In this section, we provide all the results of our analysis of the time dependence of the CF (S6) at the phase transition. We also provide details on the definition of their relaxation time  $\tau_R$  and dynamical scaling. The discussed results are depicted in Figs. S5–S11.

### A. Dynamical correlation functions

In the main text, we introduced three variants of the CF (S6),  $\bar{C}(t)$ ,  $\bar{C}_m(t)$ , and  $\hat{C}(t)$ , and discussed their decay with time  $t$  and dependence on the delay time  $\tau$ . All these CFs were calculated at the transition and for  $k^* \sim 1/\xi$ , yielding their maximum value for  $t = 0$ . The CF  $\bar{C}(t)$  is defined as

$$\bar{C}(t) \equiv C(k^*, t)/C(k^*, 0), \quad (\text{S12})$$

and  $\bar{C}_m(t)$  represents its upper envelope. For the CF  $\hat{C}(t)$ , we sampled the simulated trajectories of the delay VM with sampling frequency  $1/\Delta t_m < 1$ ,  $\Delta t_m = \tau + 1$ , and from the obtained under-sampled trajectories calculated the corresponding velocities as

$$\mathbf{v}_i(t) = \frac{|\mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t_m)|}{\Delta t_m}. \quad (\text{S13})$$

Using these under-sampled trajectories and velocities,  $\hat{C}(t)$  was calculated in the same way as  $\bar{C}(t)$  (S12). Throughout the text, we reserve the symbol  $\hat{C}(t)$  for the

specific value of  $\Delta t_m = \tau + 1$  corresponding to the distance between the spikes of CFs in Fig. S5. For other values of the sampling time, we denote this CF as  $\hat{C}_{\Delta t_m}(t)$ .

Representative plots of these CFs as functions of time  $t$  for various values of delay time  $\tau$ , and their time derivatives at  $t = 0$  are depicted in Fig. 2 in the main text and discussed below the figure. Here, we give these figures for all the investigated values of delay time. Namely, in Figs. S5, S6, and S7 we show the obtained CFs  $\bar{C}(t)$ ,  $\bar{C}_m(t)$ , and  $\hat{C}(t)$ , respectively. Time derivatives  $\bar{C}_m(t)$  and  $\hat{C}(t)$  are given in Figs. S8 and S9. The relaxation times  $\tau_R$  used to scale the time axis in these two figures are defined in Eq. (S17).

The reason behind the definition of the CF  $\hat{C}(t)$  is that in the field experiments, it can happen that the sampling frequency (e.g. of the camera) is lower than the delay times in the interactions. For example, in Ref. [10] the snapshots, used to detect the positions of the individual insects, were taken with the sampling time  $\Delta t_m \sim 6$ ms which is comparable to reaction times of the smallest insects, reviewed in Tab. S1. Figures S5–S9 show that large enough under-sampling washes away the delay-induced oscillation of the CF  $\bar{C}(t)$  and also makes it initially flat. To evaluate the robustness of these results against the chosen sampling frequency, we give in Fig. S10 CFs  $\hat{C}_{\Delta t_m}(t)$  obtained for various sampling times  $\Delta t_m$ . The figure shows that the oscillations in  $\hat{C}_{\Delta t_m}(t)$  gradually disappear with increasing sampling time, and its time derivative at  $t = 0$  flattens. The shape of the CFs saturates for  $\Delta t_m \approx \tau$ , i.e. the CFs obtained for larger sampling times fall on top of each other.

While models with a single delay time perfectly describe experiments with feedback-driven Brownian particles [15–17], where there is a single sharp value of instrumental delay time, the situation in nature is more complicated. Each individual may have its own (possibly time-dependent) reaction time, which leads to a distribution of delay times. A similar effect may also arise from sequential instead of parallel sampling of the environment as found for certain fish [18]. Furthermore, the delay times interfere with inertia of the motion. The combination of these ingredients can probably lead to similar effects on the CF  $\bar{C}(t)$  as the under sampling. To give a more precise answer, we plan to investigate effects of the distributed delay and sequential sampling in our future work.

### B. Relaxation time

In agreement with Ref. [12], we define the relaxation time  $\tau_R$  of the CFs  $K(t)$ ,  $K = \bar{C}, \bar{C}_m$ , and  $\hat{C}$  using the formula

$$\int_0^\infty dt \frac{K(t)}{t} \sin\left(\frac{t}{\tau_R}\right) = \frac{\pi}{4}. \quad (\text{S14})$$

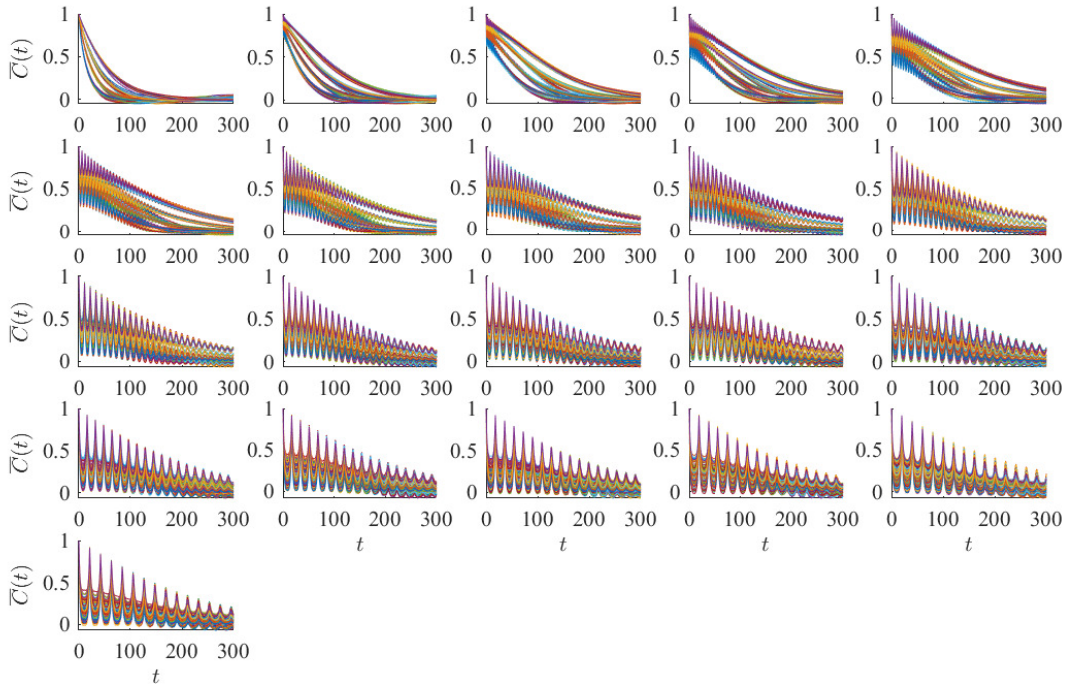


FIG. S5. Time-correlation functions (S12) for delay times  $\tau = 0, \dots, 20$  (from top-left to bottom-right panel) and system sizes  $N = 256, 512, 1024$ , and  $2048$  (bottom to top curves in each panel) as functions of time. The data were obtained from 50 simulation runs for each delay time and system size. The correlation functions are discussed in the main text. In Sec. S7 C, we describe an exactly solvable linearised version of the delay VM model, which provides qualitatively similar correlation functions as depicted.

Animal	Stimulus/Response	Reaction Time [ms]
Human	auditory	140 – 160 [19, 20]
	visual	180 – 200 [19, 20]
	touch	140 – 160 [19, 20]
Fruit fly	roll perturbation	$\sim 5$ [21]
	pitch perturbation	$\sim 12$ [22]
	yaw perturbation	10 – 25 [23]
Starling	startling sound stimuli	64 – 80 [24]
	startling light stimuli	38 – 76 [24]
Teleost fish	startle response	5 – 10 [25, 26]
Calanoida	stirring water	$< 2.5$ [27]

TABLE S1. Reaction times measured as the time between a certain stimulus and the corresponding response for various animals and stimuli.

This expression is equivalent [28] to the formula

$$\int_{-1/\tau_R}^{1/\tau_R} \tilde{K}(\omega) \frac{d\omega}{2\pi} = \frac{1}{2} K(0), \quad (\text{S15})$$

for the Fourier transform

$$\begin{aligned} \tilde{K}(\omega) &= \int_{-\infty}^{\infty} dt \bar{K}(t) \exp(i\omega t) \\ &= 2 \int_0^{\infty} dt \bar{K}(t) \cos(\omega t). \end{aligned} \quad (\text{S16})$$

The relaxation time defined by Eqs. (S14) and (S15) thus corresponds to the characteristic frequency  $1/\tau_R$  for which half of spectral contributions to the static correlation function results from the interval  $[-1/\tau_R, 1/\tau_R]$ . To actually evaluate the relaxation time using Eq. (S14) from the obtained data, we numerically solved the equation [10]

$$\sum_{t_i=1}^N \frac{K(t_i)}{t_i} \sin\left(\frac{t_i}{\tau_R}\right) = \frac{\pi}{4}. \quad (\text{S17})$$

Noteworthy, the dynamical scaling relations described in the next section are independent of the definition of the relaxation time as long as it measures the characteristic decay of the correlation function. Hence, the relaxation time can alternatively be defined as the time at which the CF  $K(t)$  decayed to a specific value, say  $K(\tau_R) = 1/2$ .

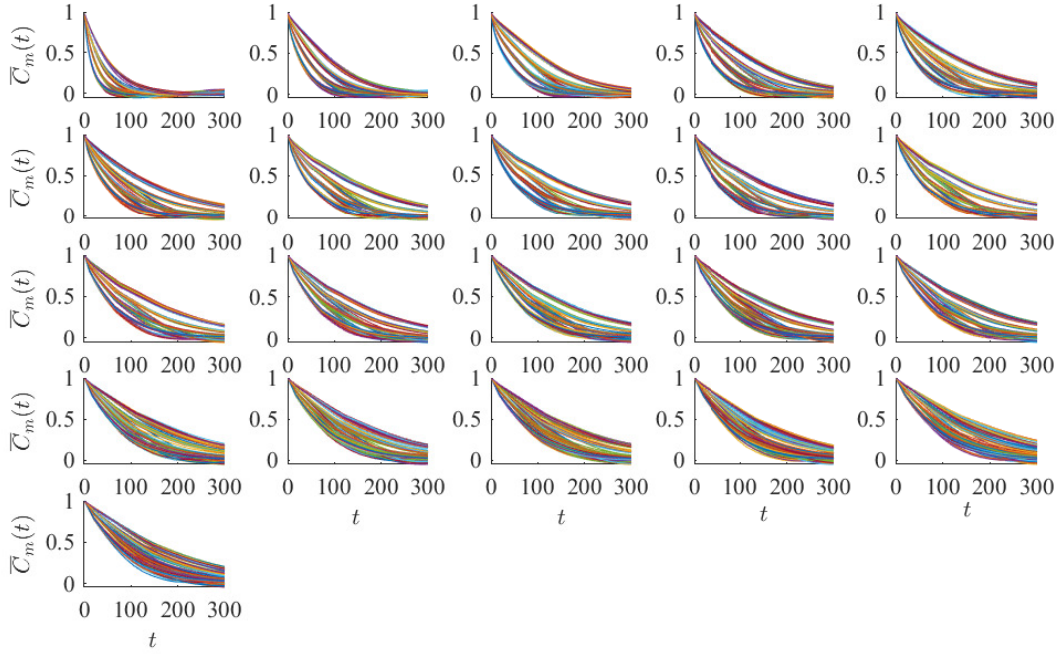


FIG. S6. The upper envelopes of time-correlation functions depicted in Fig. S5.

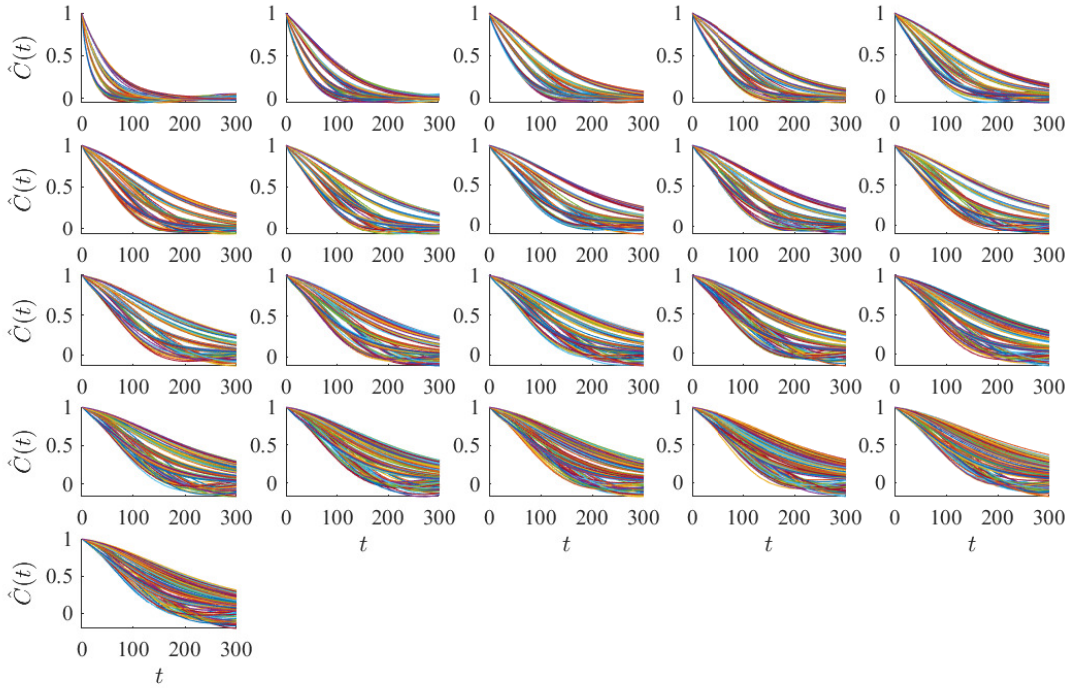


FIG. S7. Time-correlation functions (S12) calculated from under-sampled trajectories with sampling with frequency  $1/(\tau + 1)$  as functions of time. The individual curves correspond to those depicted in Fig. S5, where the sampling frequency is 1. The shape of the depicted curves is discussed in the main text.

We have tested that this definition leads to similar results as the definition (S17).

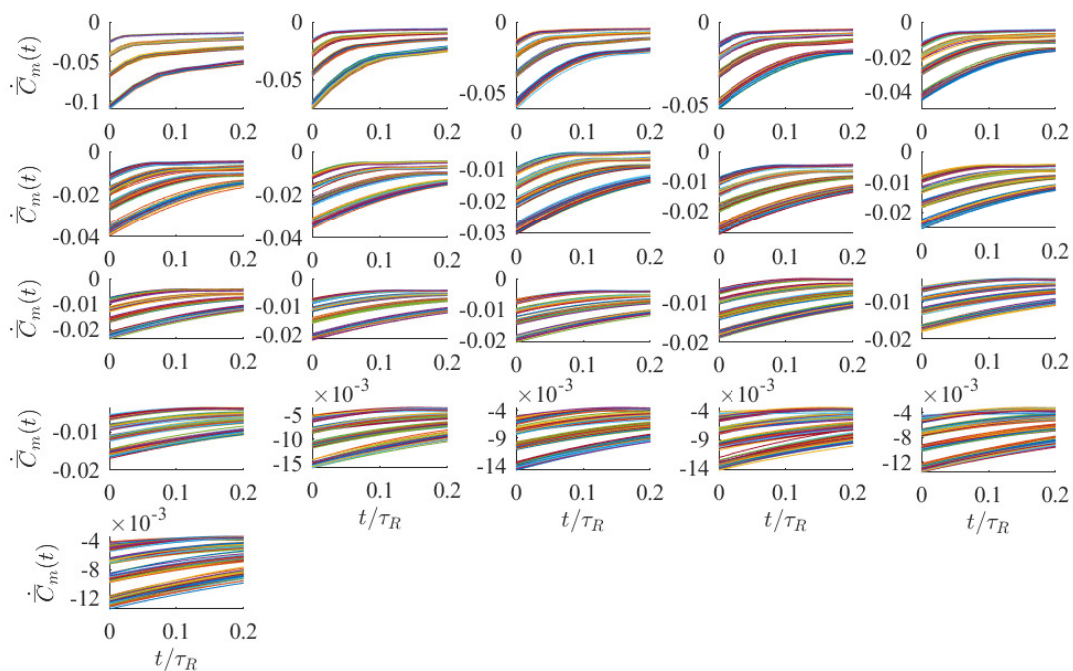


FIG. S8. Time derivatives of the time correlation functions  $\bar{C}_m(t)$  shown in Fig. S6.

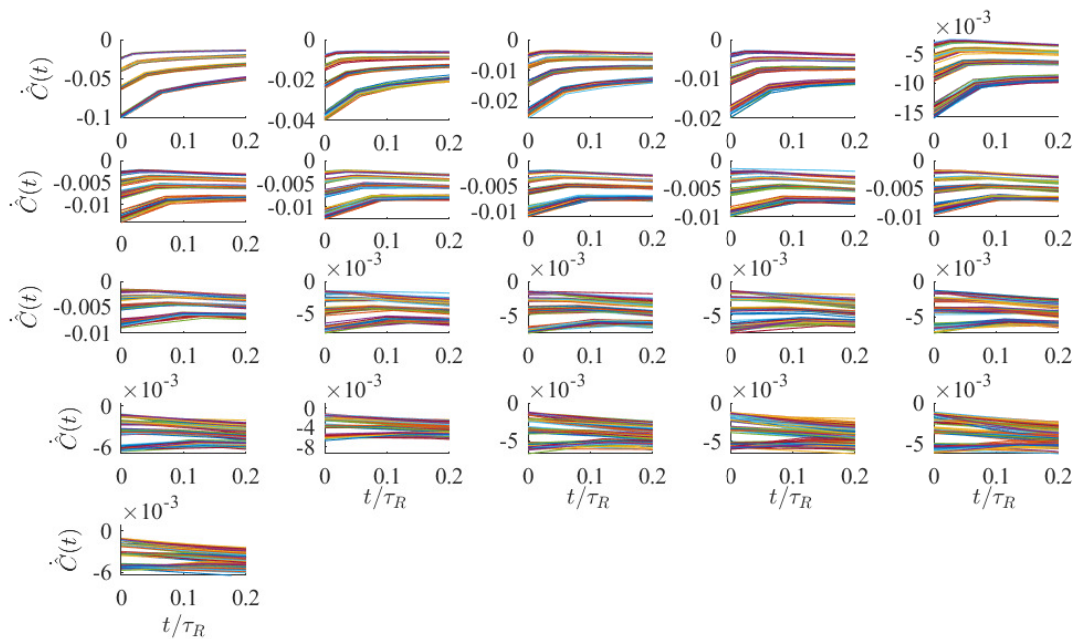


FIG. S9. Time derivatives of the time correlation functions  $\hat{C}(t)$  shown in Fig. S7.

### C. Dynamical scaling

dependent correlations in scale free systems obey a cer-

According to the dynamical scaling hypothesis [28], not only the static correlations (susceptibility) but also time

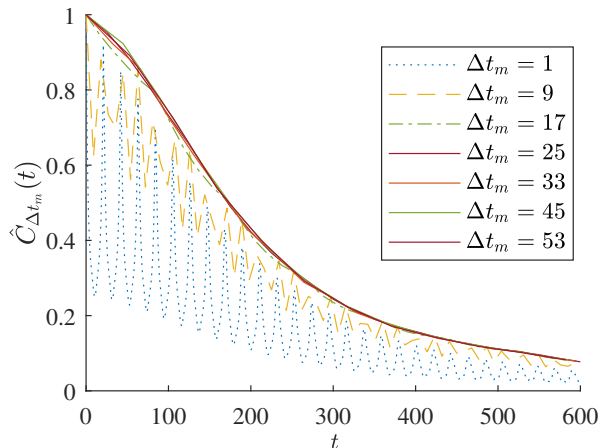


FIG. S10. Convergence of the under-sampled correlation function with decreasing sampling frequency  $1/\Delta t_m$  for the delay time  $\tau = 20$ . The oscillations disappear for  $\Delta t_m \approx \tau$ . A further decrease of  $1/\Delta t_m$  leaves  $\hat{C}_{\Delta t_m}(t)$  unchanged.

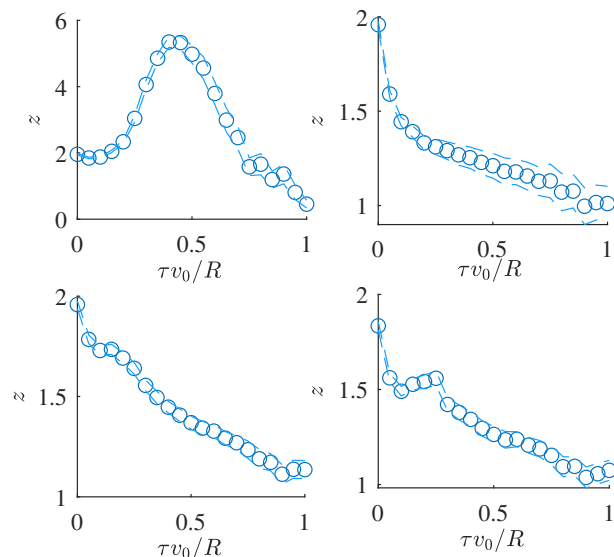


FIG. S11. Dynamical exponent  $z$  obtained by fitting the relation (S18) with  $\tau_R$  obtained using Eq. (S17) from the CF  $\bar{C}$  (top left),  $\bar{C}_m$  (top right), and  $\hat{C}$  (bottom left). The exponent  $z$  shown in bottom right panel was calculated from the CF  $\bar{C}_m$  using the same procedure but with the relaxation time defined by  $\bar{C}(\tau_R) = 1/2$ . Dashed lines mark the 95% confidence intervals of the fits.

tain scaling with particle number.

Specifically, the CFs  $K(t)$ ,  $K = \bar{C}, \bar{C}_m$ , and  $\hat{C}$ , for different particle numbers at a phase transition are conjectured to collapse onto a single master curve after rescaling the time by the appropriate relaxation times  $\tau_R(N)$ . We verified that this conjecture is valid for the delay VM for all the considered values of  $\tau$  and  $N$ . In Fig. 2 of

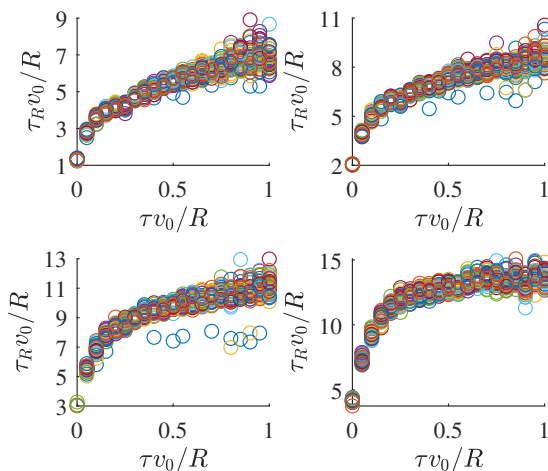


FIG. S12. Relaxation times  $\tau_R$  for correlation functions  $\bar{C}_m(t)$  for  $N = 256, 512, 1024, 2048$  as functions of the delay time, respectively. The data were obtained from all 50 runs of the simulation. Relaxation times obtained from  $\hat{C}(t)$  are qualitatively the same.

the main text, we provide representative examples of the obtained data collapse.

Furthermore, the relaxation time at the transition for given  $N$  is conjectured to be related to the corresponding correlation length  $\xi$ , and, through Eqs. (S8) and (S10), also to  $N$ , according to the relation

$$\tau_R \sim \xi^z \sim (k^*)^{-z} \sim N^{z/3}, \quad (\text{S18})$$

with the dynamical exponent  $z$ . For all the considered delay times and particle numbers, we verified that the delay VM conforms also with this relation.

The former two results provide two strategies for evaluation of the exponent  $z$ . The first one is to seek for the best data collapse of the CFs by rescaling time as  $t/\xi^z$  with parameter  $z$ . The second one is to create for each delay time the parametric plot  $\log k^*(N)$  against  $\log \tau_R(N)$  and determine  $z$  as a slope of linear fit to the plotted data.

The exponents as functions of delay time obtained using the latter procedure from CFs  $\bar{C}$ ,  $\bar{C}_m$ , and  $\hat{C}$  with relaxation time defined by Eq. (S17) are depicted in the first three panels of Fig. S11. In the last panel, we additionally give  $z$  obtained from  $\bar{C}_m$  using the same procedure but with relaxation time defined by  $\bar{C}_m(\tau_R) = 1/2$ . All the obtained results except for the one from the CF  $\bar{C}$  are qualitatively the same: with increasing delay time the dynamic exponent decays from  $z \approx 2$  for  $\tau = 0$  to  $z \approx 1.1$  for large  $\tau$ . In Fig. 2 in the main text, we show that the dynamical exponents obtained from the CFs  $\bar{C}_m$  and  $\hat{C}$  using the first method also exhibit the described behavior. Examples of the corresponding data collapses of CFs  $\bar{C}_m$  and  $\hat{C}$  are given in Figs. S13 and S14, respectively. The exponent obtained from  $\bar{C}$  in Fig. S11

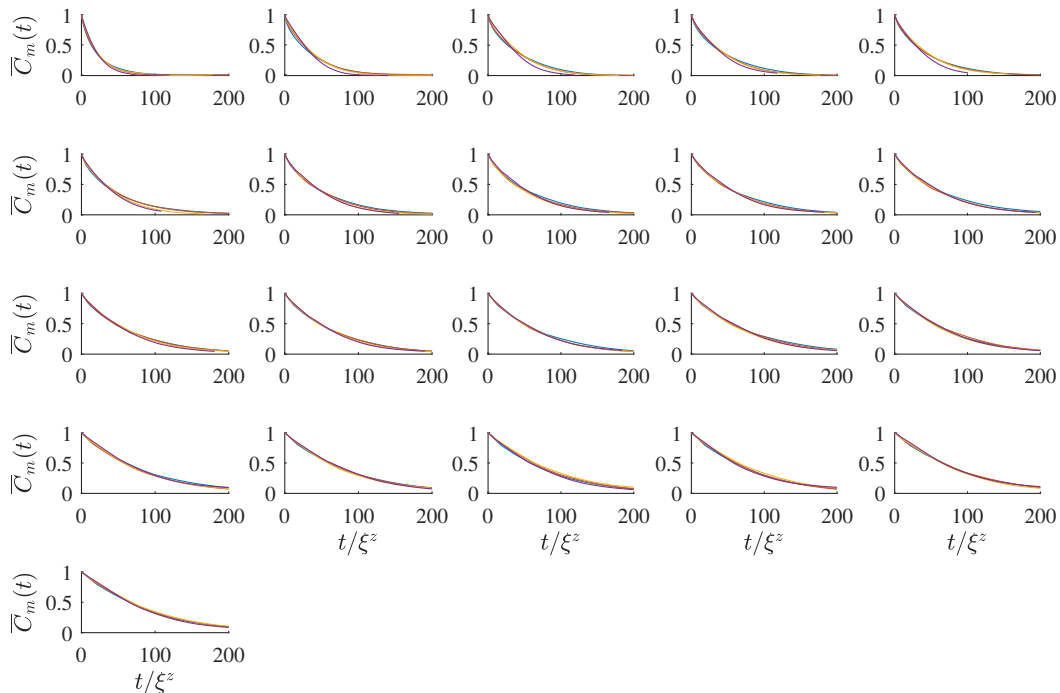


FIG. S13. Data collapse of CFs  $\overline{C}_m$  used for determination of the dynamical exponent  $z$  obtained using one of the 50 runs of the simulation. The box-plots showing  $z$  in Fig. 3j of the main text follow from statistical analysis of data collapses obtained for all 50 runs.

exhibits a maximum for  $v_0\tau \approx 0.4$  and then decays to zero. Due to the disagreement of this result with  $z$  obtained using all other methods, we consider it as wrong and attribute the error to the unsuitable use of Eq. (S17) for calculation of relaxation times for rapidly oscillating functions such as  $\overline{C}$ .

To further support our conclusion about convergence of  $z$  with increasing delay time made on the basis of Fig. S11, we show in Fig. S12 the corresponding convergence of relaxation times. The relaxation time converges with delay faster for larger particle numbers. In particular for the largest particle number  $N = 2048$  considered,  $\tau_R$  seems to relax for delays about  $\tau v_0/R = 1$ . The dynamical critical exponent  $z$  is determined by correlation length  $\xi$  and  $\tau_R$ . Convergence of  $\xi$  follows from our analysis of the static scaling, namely from the convergence of the susceptibility  $S_2$  with delay time. The provided strong numerical evidence for convergence of both  $\tau_R$  and  $\xi$  thus also supports convergence of  $z$ . In the next section, we provide further analytical evidence for this result.

## S6. LONG DELAY LIMIT

Let us now investigate how the dynamical equations for the delay VM change for  $\tau$  much larger than the unit discrete update time. The dynamical equations read

$$\mathbf{v}_i(t+1) = v_0 \mathcal{R}_\alpha \Theta \left[ \mathbf{v}_i(t) + \sum_{j \neq i} n_{ij}(t-\tau) \mathbf{v}_j(t-\tau) \right], \quad (\text{S19})$$

$$\mathbf{r}_i(t+1) = \mathbf{r}_i(t) + \mathbf{v}_i(t+1). \quad (\text{S20})$$

Re-scaling time as  $\tilde{t} = t/\tau$ , space as  $\mathbf{x}_i = \mathbf{r}_i/\tau$ , and defining quantities in the scaled time as  $\tilde{f}(\tilde{t}) = f(\tau\tilde{t})$ , with  $f = \mathbf{x}_i$  and  $\mathbf{v}_i$ , we find

$$\frac{\tilde{\mathbf{v}}_i(\tilde{t}+1/\tau) - \tilde{\mathbf{v}}_i(\tilde{t})}{1/\tau} = v_0 \tau R_i(\tilde{t}, \tilde{t}-1) \quad (\text{S21})$$

$$\frac{\tilde{\mathbf{x}}_i(\tilde{t}+1/\tau) - \tilde{\mathbf{x}}_i(\tilde{t})}{1/\tau} = \tilde{\mathbf{v}}_i(\tilde{t}+1/\tau), \quad (\text{S22})$$

where the functions

$$R_i = \mathcal{R}_\alpha \Theta \left[ \tilde{\mathbf{v}}_i(\tilde{t}) + \sum_j \tilde{n}_{ij}(\tilde{t}-1) \tilde{\mathbf{v}}_j(\tilde{t}-1) \right] - \frac{\tilde{\mathbf{v}}_i(\tilde{t})}{v_0}$$

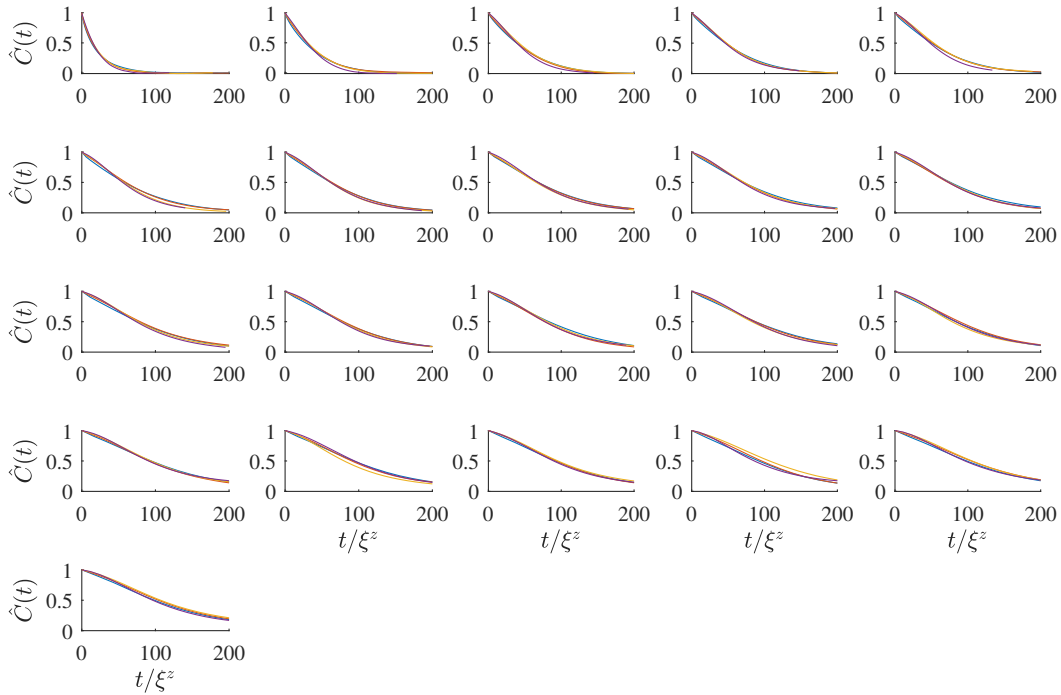


FIG. S14. Data collapse of CFs  $\hat{C}_m$  used for determination of the dynamical exponent  $z$  obtained using one of the 50 runs of the simulation. The box-plots showing  $z$  in Fig. 3k of the main text follow from statistical analysis of data collapses obtained for all 50 runs.

do not depend on the delay. Taking now the limit  $\tau \gg 1$  on the left-hand sides of Eqs. (S21) and (S22), we find a continuous version of the delay VM

$$\dot{\tilde{\mathbf{v}}}_i(\tilde{t}) = v_0 \tau R_i(\tilde{t}, \tilde{t} - 1) \quad (\text{S23})$$

$$\dot{\tilde{\mathbf{x}}}_i(\tilde{t}) = \tilde{\mathbf{v}}_i(\tilde{t}), \quad (\text{S24})$$

where the delay time appears on the same footing with the speed  $v_0$ . This suggests that the critical behavior of the delay VM should for large delays be independent of the specific value of  $\tau$ , similarly as for  $v_0$  [29]. The values of the exponents in this regime are given by the plateau values of exponents reached for  $v_0 \tau / R \rightarrow 1$  in the main text.

## S7. EXACTLY SOLVABLE LINEARIZED DELAY VICSEK MODEL

Time correlations in the delay VM can be investigated analytically in the so-called spin-wave expansion around the perfectly ordered state. In this section, we formulate the approximation, solve the corresponding dynamical equations, and show that the resulting CFs are qualitatively similar to the CFs obtained from our simulations of the complete delay VM, depicted Fig. S5.

### A. Continuous-time delay Vicsek model

In order to treat the delay VM analytically, we adopt its simpler time-continuous version

$$\eta \dot{\mathbf{v}}_i(t) = \left[ J \sum_j n_{ij} \mathbf{v}_j(t - \tau) + \sqrt{2D} \boldsymbol{\xi}_i(t) \right]_{\perp}^i, \quad (\text{S25})$$

$$\dot{\mathbf{x}}_i(t) = \mathbf{v}_i(t), \quad (\text{S26})$$

based on the time-continuous version of the standard VM described in detail in Refs. [11, 12]. Above,  $\eta$  sets the timescale,  $J$  measures the strength of the alignment interaction,  $n_{ij}$  is the connectivity matrix from the time-discrete delay VM (S19)–(S20), and  $D$  controls the intensity of the standard Gaussian white noise  $\boldsymbol{\xi}_i$  defined by  $\langle \boldsymbol{\xi}_i(t) \rangle = 0$  and  $\langle \boldsymbol{\xi}_i(t) \boldsymbol{\xi}_j(t') \rangle = d \delta_{ij} \delta(t - t')$ , where  $d$  is the spatial dimension, i.e.,  $d = 3$  in our case. The operation

$$[\mathbf{w}]_{\perp}^i \equiv \mathbf{w} - \left( \frac{\mathbf{w} \cdot \mathbf{v}_i(t)}{v_0} \right) \frac{\mathbf{v}_i(t)}{v_0} \quad (\text{S27})$$

keeps the magnitude of the velocity  $\mathbf{v}_i(t)$  constant by projecting the right-hand side of Eq. (S25) onto its perpendicular direction.

### B. Spin-wave expansion of the Vicsek equations

The continuous-time version (S25)–(S26) of the delay VM can be treated analytically using the so-called spin-wave expansion [11, 12] around the completely ordered state. To be specific, we assume that velocities  $\mathbf{v}_i(t)$  of the individual particles in the system point, up to small fluctuations, in the direction of the  $x$ -axis. In symbols,

$$\mathbf{v}_i(t) = v_0 \left( 1 - \frac{1}{2} \frac{\pi_i^2}{v_0^2} \right) \mathbf{e}_x + \boldsymbol{\pi}_i(t), \quad (\text{S28})$$

where  $\boldsymbol{\pi}_i(t) = [0, \varphi_i^y(t), \varphi_i^z(t)]$  measures the (small) fluctuations around the average direction of the system,  $\mathbf{e}_x = (1, 0, 0)$ , and  $\pi_i^2 = \boldsymbol{\pi}_i(t) \cdot \boldsymbol{\pi}_i(t)$ . Inserting the velocities (S28) into Eq. (S25) and keeping only the leading order terms in the fluctuations  $\varphi_y(t)$ ,  $\varphi_z(t)$ , we obtain the linear delay Langevin equation

$$\eta \dot{\varphi}_i(t) = J \sum_j n_{ij} [\varphi_j(t - \tau) - \varphi_i(t)] + \sqrt{2\tilde{D}} \xi(t). \quad (\text{S29})$$

Here,  $\varphi$  stands either for  $\varphi_i^y$  or  $\varphi_i^z$  and  $\xi$  is a standard Gaussian white noise described by  $\langle \xi \rangle = 0$  and  $\langle \xi(t) \xi(t') \rangle = \delta(t - t')$ .

For vanishing delay time ( $\tau = 0$ ), this equation is derived in Ref. [11]. Following further the discussion therein, we apply the so-called fixed network approximation, i.e., we assume that the connectivity matrix  $n_{ij}$  is constant in time and rewrite Eq. (S29) as

$$\begin{aligned} \eta \dot{\varphi}_i(t) &= -J \sum_j \Lambda_{ij} \varphi_j(t - \tau) \\ &+ J \sum_j n_{ij} [\varphi_i(t - \tau) - \varphi_i(t)] + \sqrt{2\tilde{D}} \xi(t). \end{aligned} \quad (\text{S30})$$

Here, we introduced the matrix  $\Lambda_{ij} = -n_{ij} + \delta_{ij} \sum_k n_{ik}$ , which defines a discrete version of the Laplace operator. In the limit of continuous space variable,  $\varphi_i = \varphi(\mathbf{r})$ ,  $\sum_j \Lambda_{ij} = -n_c a^2 \Delta$ , and the above equation transforms to the analytically manageable form

$$\begin{aligned} \eta \dot{\varphi}(\mathbf{r}, t) &= J n_c a^2 \Delta \varphi(\mathbf{r}, t - \tau) + J n_c [\varphi(\mathbf{r}, t - \tau) - \varphi(\mathbf{r}, t)] \\ &+ \sqrt{2\tilde{D} a^3} \xi(\mathbf{r}, t). \end{aligned} \quad (\text{S31})$$

Above,  $a$  is the lattice constant,  $n_c = \sum_{j \neq i} n_{ij}$  is the number of nearest neighbours of a lattice point,  $\mathbf{r} = (x, y, z)$  denotes the three dimensional position vector, and the Gaussian white noise  $\xi$  fulfills  $\langle \xi(\mathbf{r}, t) \rangle = 0$  and  $\langle \xi(\mathbf{r}, t) \xi(\mathbf{r}', t') \rangle = \delta^{(3)}(\mathbf{r} - \mathbf{r}') \delta(t - t')$  with  $\delta^{(3)}(\mathbf{r} - \mathbf{r}') = \delta(x - x') \delta(y - y') \delta(z - z')$ .

### C. Green's function

The Fourier-transformed Eq. (S31) reads

$$\begin{aligned} \dot{\varphi}(\mathbf{k}, t) &= \tau^{-1} \tilde{J} (1 - a^2 k^2) \varphi(\mathbf{k}, t - \tau) - \tau^{-1} \tilde{J} \varphi(\mathbf{k}, t) \\ &+ \sqrt{2\tilde{D}} \xi(\mathbf{k}, t), \end{aligned} \quad (\text{S32})$$

with the Gaussian white noise fulfilling  $\langle \xi(\mathbf{k}, t) \rangle = 0$ ,  $\langle \xi(\mathbf{k}, t) \xi(\mathbf{k}', t') \rangle = (2\pi)^3 \delta(t - t') \delta^{(3)}(\mathbf{k} + \mathbf{k}')$ ,  $\tilde{J} = J n_c \tau$  and  $\tilde{D} = D a^3 / \eta^2$ . The general solution to this equation is given by [8]

$$\begin{aligned} \varphi(\mathbf{k}, t) &= \lambda(\mathbf{k}, t) \varphi_0 + \tau^{-1} \tilde{J} b \int_{-\tau}^0 dt' \lambda(\mathbf{k}, t - t' - \tau) \varphi(\mathbf{k}, t') \\ &+ \sqrt{2\tilde{D}} \int_0^t dt' \lambda(\mathbf{k}, t - t') \xi(\mathbf{k}, t'), \end{aligned} \quad (\text{S33})$$

where  $\varphi_0 = \varphi(\mathbf{k}, t = 0)$  denotes the value of  $\varphi(\mathbf{k}, t)$  for  $t = 0$ , and

$$b = 1 - a^2 k^2. \quad (\text{S34})$$

The Green's function  $\lambda(\mathbf{k}, t)$  solves Eq. (S32) with vanishing noise term ( $\tilde{D} = 0$ ) and the initial condition  $\lambda(\mathbf{k}, t) = 0$  for  $t < 0$  and  $\lambda(\mathbf{k}, 0) = 1$ , i.e.

$$\dot{\lambda}(\mathbf{k}, t) = \tau^{-1} \tilde{J} b \lambda(\mathbf{k}, t - \tau) - \tau^{-1} \tilde{J} \lambda(\mathbf{k}, t). \quad (\text{S35})$$

The most straightforward way to solve this equation with the above initial conditions is to employ the Laplace transform in time. Multiplying Eq. (S35) by  $\exp(-st)$  and integrating over time, we obtain

$$\begin{aligned} s \lambda(\mathbf{k}, s) - \lambda_0 &= -\tau^{-1} \tilde{J} b \exp(-s\tau) \lambda(\mathbf{k}, s) - \tau^{-1} \tilde{J} x(\mathbf{k}, s) \\ &- \tau^{-1} \tilde{J} b \int_{-\tau}^0 dt \exp(-s\tau) \exp(-st) \lambda(\mathbf{k}, t), \end{aligned} \quad (\text{S36})$$

where  $\lambda_0$  is the value of  $\lambda(\mathbf{k}, t)$  for  $t = 0$ . The Laplace transform of  $\lambda(\mathbf{k}, t)$ ,

$$\lambda(\mathbf{k}, s) = \int_0^\infty dt \exp(-st) \lambda(\mathbf{k}, t), \quad (\text{S37})$$

thus reads

$$\lambda(\mathbf{k}, s) = \frac{1}{s - \tau^{-1} \tilde{J} b \exp(-s\tau) + \tau^{-1} \tilde{J}}, \quad (\text{S38})$$

where we used the initial conditions  $\lambda_0 = 1$  and  $\lambda(\mathbf{k}, t) = 0$  for  $t < 0$ . Expanding the denominator of this expression, we obtain the series

$$\lambda(\mathbf{k}, s) = \sum_{l=0}^{\infty} \frac{(\tau^{-1} \tilde{J} b)^l}{(s + \tau^{-1} \tilde{J})^{l+1}} \exp(-ls\tau). \quad (\text{S39})$$

The inverse Laplace transform of  $\exp(-ls\tau)/(s + J)^{l+1}$  is given by  $(t - l\tau)^l \exp[-\tau^{-1} \tilde{J}(t - l\tau)] \Theta(t - l\tau) / l!$ , where  $\Theta$  denotes the Heaviside theta function. The inverse Laplace transform of Eq. (S39) thus yields the Green's function

$$\lambda(\mathbf{k}, t) = \sum_{l=0}^{\infty} \frac{(\tilde{J} b)^l}{l!} \left( \frac{t}{\tau} - l \right)^l e^{-\tilde{J} \left( \frac{t}{\tau} - l \right)} \Theta \left[ \frac{t}{\tau} - l \right]. \quad (\text{S40})$$



Except for the exponential factor  $\exp(-\tilde{J}t/\tau)$ , this Green's function is of the same form as the one for Eq. (S35) without the term proportional to  $\lambda(\mathbf{k}, t)$  on the right-hand side, which was derived in Ref. [8]. Pulling it out of the sum yields the expression  $\exp(-\tilde{J}t/\tau)\tilde{\lambda}(t)$ , where  $\tilde{\lambda}(t)$  is the Greens function obtained in Ref. [8] for

$$\dot{\tilde{\lambda}}(t) = \tilde{J}b \exp(\tilde{J})\tilde{\lambda}(t - \tau). \quad (\text{S41})$$

Properties of the Green's function (S40) can be inferred from inserting the exponential ansatz  $\lambda_{\text{E}}(\mathbf{k}, t) = \exp(-\kappa t)$  in the Eq. (S35). One finds that it solves the equation for ( $W$  denotes the Lambert  $W$  function)

$$\kappa = \frac{1}{\tau} \left\{ \tilde{J} - W \left[ b\tilde{J} \exp(\tilde{J}) \right] \right\}, \quad (\text{S42})$$

which corresponds to the dispersion relation  $\omega(\mathbf{k}) = i^{-1}\kappa$  between the frequency  $\omega$  and the wave vector  $k$ , revealing which plane waves can propagate through the system [11]. The real part of the ansatz

$$\Re[\lambda_{\text{E}}(\mathbf{k}, t)] = \exp(-t/t_{\text{R}}) \cos(\nu t) \quad (\text{S43})$$

describes well the long-time behavior of the Green's function  $\lambda(\mathbf{k}, t)$ , allowing us to comfortably determine its relaxation time  $t_{\text{R}}$  and the frequency  $\nu$  as

$$t_{\text{R}} = 1/\Re(\kappa), \quad (\text{S44})$$

$$\nu = \Im(\kappa). \quad (\text{S45})$$

For  $t_{\text{R}} > 0$  we find an exponential decay, while for  $t_{\text{R}} < 0$  the system explodes/diverges. For  $\nu \neq 0$  the solutions exhibit damped or exploding oscillations. However, initial oscillations that eventually die out over time can also be present for  $\nu = 0$ .

For the VM, the magnitude of the wave vector  $k$  is bounded from below by the linear system size as  $1/L$  and from above by the lattice constant as  $1/a$ . Therefore, the term  $ak$  in the parameter  $b$  (S34) is always smaller than one and thus  $b > 0$ . In Fig. S15, we show various examples of the Green's function (S40) in this parameter regime. The spikes in the figures are separated by one delay time and physically originate in the tendency to align with the average orientation of the neighbors before one delay time. Mathematically, they result from the summation with theta-functions in the Green's function. For  $b > 0$ , each summand is positive, largest at  $t = n\tau$ , and decays exponentially with increasing time. Thus, the sum in Eq. (S40) suddenly increases at instants  $t = n\tau$ , which induce the spikes. These spikes start to vanish when the factorials in the denominator of the summands become much larger than the numerator. After that, only the behavior captured by the exponential ansatz (S43) survives. This happens roughly after  $k \approx \tilde{J}b$  steps in the sum, corresponding to  $t \approx \tilde{J}b\tau$ .

#### D. Correlation functions from spin-wave expansion

In the stable regime  $t_{\text{R}} > 0$ , where the Green's function eventually decays to zero,  $\lim_{t \rightarrow \infty} \lambda(\mathbf{k}, t) = 0$ , the

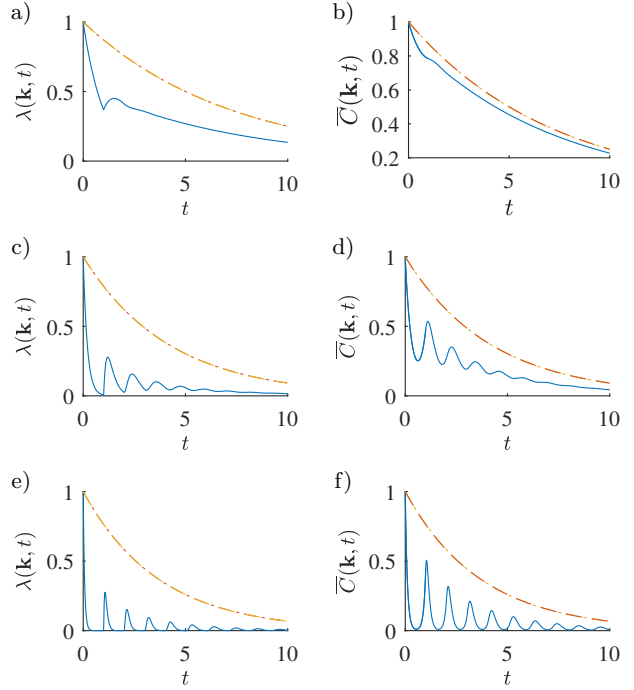


FIG. S15. Panels a), c), and e) show the Green's function (S40) as function of time for  $ak = 0.5$ ,  $\tau = 1$ , and  $\tilde{J} = 1, 5, 15$ , respectively. Panels b), d), and f) show the corresponding time-correlation functions (S48). The shape of these correlation functions is qualitatively similar to the shape of the correlation functions obtained from the complete delay Vicsek, e.g., see Fig. S16. The dashed and dotted overlapping lines represent the exponential envelope  $\exp(-t/t_{\text{R}})$  and the full expression (S43), respectively.

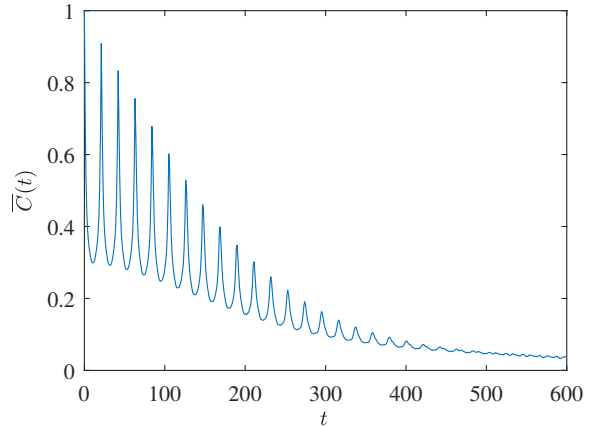


FIG. S16. Correlation function  $\bar{C}(t)$  as function of time obtained from the simulation of the complete delay VM with  $N = 2048$  and  $\tau = 20$ .

general solution (S33) to Eq. (S32) can be used for calculation of the time-correlation function in  $\mathbf{k}$ -space,

$C(\mathbf{k}, t) = \lim_{t_0 \rightarrow \infty} \langle \varphi(\mathbf{k}, t_0 + t) \varphi(-\mathbf{k}, t_0) \rangle$ . We find that for  $t > 0$  (see also Ref. [8])

$$C(\mathbf{k}, t) = 2\tilde{D} \lim_{t_0 \rightarrow \infty} \int_0^{t_0} dt' \lambda(\mathbf{k}, t + t_0 - t') \lambda(\mathbf{k}, t_0 - t'). \quad (\text{S46})$$

This expression can be used for plotting the time correlation function (see Fig. S15 for a comparison between the Green's function and the correlation function), however, it is not very suitable for mathematical analysis. Following the approach in Refs. [8, 30], we take the time derivative of Eq. (S46) and use Eq. (S35) for the Green's function. In this way, we obtain for  $C(\mathbf{k}, t)$  the dynamical equation

$$\dot{C}(\mathbf{k}, t) = \tau^{-1} \tilde{J} b C(\mathbf{k}, t - \tau) - \tau^{-1} \tilde{J} C(\mathbf{k}, t), \quad (\text{S47})$$

valid for  $t > 0$ , due to the nonanalyticity of  $\lambda(\mathbf{k}, t)$  at  $t = 0$ . The solution to this equation is given by Eq. (S33),

$$C(\mathbf{k}, t) = \lambda(\mathbf{k}, t) C_0 + \tau^{-1} \tilde{J} b \int_{-\tau}^0 dt' \lambda(\mathbf{k}, t - t' - \tau) C(\mathbf{k}, t'), \quad (\text{S48})$$

and thus the decay time of the time-correlation function is given by the decay time (S44) of the Green's function  $\lambda(\mathbf{k}, t)$ .

To evaluate the above expression, we need to find the static correlation function  $C_0 \equiv C(\mathbf{k}, 0)$  and the initial condition  $C(\mathbf{k}, t)$  for  $t \in (-\tau, 0)$ . This can be done as follows. Employing the symmetry  $C(\mathbf{k}, t) = C(\mathbf{k}, -t)$  of the stationary correlation function, we rewrite Eq. (S47) as

$$\dot{C}(\mathbf{k}, t) = \tau^{-1} \tilde{J} b C(\mathbf{k}, \tau - t) - \tau^{-1} \tilde{J} C(\mathbf{k}, t). \quad (\text{S49})$$

For  $t \in (0, \tau)$ , we can differentiate this equation once again. The result is

$$\ddot{C}(\mathbf{k}, t) = -\Omega^2 C(\mathbf{k}, t), \quad (\text{S50})$$

where we used Eq. (S47) and defined the (possibly imaginary) frequency  $\Omega = \tau^{-1} \tilde{J} \sqrt{b^2 - 1}$ . From Eq. (S50), we find that for  $t \in [-\tau, \tau]$

$$C(\mathbf{k}, t) = C_0 \cos(\Omega t) + \dot{C}_0 \Omega^{-1} \sin(\Omega |t|), \quad (\text{S51})$$

with  $\dot{C}_0 = \lim_{t \rightarrow 0^+} \dot{C}(\mathbf{k}, t)$  denoting the time-derivative of the correlation function infinitesimally close to 0 from the right (its time-derivative) is discontinuous at  $t = 0$  [30]). Since the derivative  $\dot{C}_0$  is evaluated at  $t > 0$  (even though infinitesimally close to  $t = 0$ ), it can be obtained from Eq. (S46). We find that

$$\begin{aligned} \dot{C}_0 &= 2\tilde{D} \lim_{t_0 \rightarrow \infty} \int_0^{t_0} dt' \dot{\lambda}(\mathbf{k}, t_0 - t') \lambda(\mathbf{k}, t_0 - t') \\ &= -\tilde{D} \lim_{t_0 \rightarrow \infty} \int_0^{t_0} dt' \frac{d}{dt'} \lambda^2(\mathbf{k}, t_0 - t') \\ &= -\tilde{D} \lim_{t_0 \rightarrow \infty} [\lambda^2(\mathbf{k}, 0) - \lambda^2(\mathbf{k}, t_0)] = -\tilde{D}. \end{aligned} \quad (\text{S52})$$

In order to evaluate  $C_0$ , we note that  $\dot{C}_0$  also follows from Eq. (S49) with  $t = 0$ , yielding  $\dot{C}_0 = -\tilde{D} = \tau^{-1} \tilde{J} b C(\mathbf{k}, \tau) - \tau^{-1} \tilde{J} C_0$ . Using  $C(\mathbf{k}, \tau) = C_0 \cos(\Omega \tau) - \tilde{D} \Omega^{-1} \sin(\Omega \tau)$  given by Eq. (S51) for  $t > 0$ , we find

$$C_0 = \frac{\tilde{D} \tau}{\tilde{J}} \left[ 1 - \frac{b \sin(\Omega \tau)}{\sqrt{b^2 - 1}} \right] \frac{1}{1 - b \cos(\Omega \tau)}. \quad (\text{S53})$$

For  $b > 0$ ,  $C_0$  monotonously decreases with the delay time in agreement with the behavior of susceptibilities in Figs. S2 and S3.

Since the involved integration runs just from  $-\tau$  to 0 instead of from 0 to  $\infty$ , the expressions (S48, S51–S53) provide a much more comfortable (and accurate) way to evaluate the correlation function than Eq. (S46). Moreover, the formula (S51) describes explicitly the short-time behavior of  $C(\mathbf{k}, t)$ . Its initial slope is determined solely by the noise strength  $\tilde{D} = D a^3$ . On the other hand, the initial slope of the normalized correlation function  $\bar{C}(\mathbf{k}, t) = C(\mathbf{k}, t)/C_0$ , usually evaluated in the experiments, reads

$$\begin{aligned} \dot{\bar{C}}_0 &= \frac{\dot{C}_0}{C_0} = \frac{\tilde{J} \sqrt{b^2 - 1} [1 - b \cos(\Omega \tau)]}{\tau \sqrt{b^2 - 1} - b \sin(\Omega \tau)} \\ &= \frac{\tilde{J} \sqrt{b^2 - 1}}{\tau} \frac{1 - b \cos(\tilde{J} \sqrt{b^2 - 1})}{\sqrt{b^2 - 1} - b \sin(\tilde{J} \sqrt{b^2 - 1})} \end{aligned} \quad (\text{S54})$$

which is noise-independent. Note that both  $\bar{C}_0$  and  $\dot{\bar{C}}_0$  are real valued even for  $b^2 - 1 < 0$ .

The correlation function in the presented approximate model is thus in general not initially flat ( $\dot{\bar{C}}_0 \neq 0$ ). This might seem surprising in the light of results of Cavagna et al. [10–12]. They showed that the time-derivative of the correlation function  $C(\mathbf{k}, t)$  at  $t \rightarrow 0$  is nonzero if the Fourier transform of the corresponding Green's function has a single pole in the positive plane, and that  $\lim_{t \rightarrow 0} \dot{C}(\mathbf{k}, t) = 0$  for two and more poles in the positive semi-plane. In general, the pole structure determines the dispersion polynomial [31]. Multiple poles with non-zero real parts point to propagation of spin-waves through the system, while a single pole means that the waves are overdamped and the relaxation is purely exponential.

The Fourier transform of the Green's function (S40) follows from Eq. (S38) after the substitution  $s \rightarrow -i\omega$ . The resulting Green's function  $\lambda(\mathbf{k}, \omega)$  contains in the denominator a polynomial of infinite order in  $\omega$  and thus possesses infinitely many poles in the positive semi-plane. However, the time-derivative of the time correlation function  $C(\mathbf{k}, t)$  at  $t \rightarrow 0$  (S52) is in general nonzero and thus the argument of Cavagna et al. [10–12] does not apply in our situation. The reason is that the contour integration used in their derivation can be performed only if all the poles are located in a finite region in the complex plane.

Nevertheless, our analysis of the full delay VM in the main paper and in the first part of this Supplemental Information strongly suggests that the delay VM is indeed very similar to a standard model with finite number

of positive poles – namely, the corresponding correlation

functions becomes initially flat if low-pass filtered, e.g. by a suitable undersampling of the trajectories.

- 
- [1] D. Martin, H. Chaté, C. Nardini, A. Solon, J. Tailleur, and F. Van Wijland, Fluctuation-induced phase separation in metric and topological models of collective motion, *Physical Review Letters* **126**, 148001 (2021).
- [2] G. Pruessner and H. Jeldtoft Jensen, Broken scaling in the forest-fire model, *Physical Review E* **65**, 056707 (2002).
- [3] L. Palmieri and H. J. Jensen, The forest fire model: The subtleties of criticality and scale invariance, *Frontiers in Physics* **8**, 257 (2020).
- [4] L. Chen, J. Toner, and C. F. Lee, Critical phenomenon of the order–disorder transition in incompressible active fluids, *New Journal of Physics* **17**, 042002 (2015).
- [5] L. Chen, C. F. Lee, and J. Toner, Incompressible polar active fluids in the moving phase in dimensions  $d > 2$ , *New Journal of Physics* **20**, 113035 (2018).
- [6] A. Cavagna, L. Di Carlo, I. Giardina, T. S. Grigera, and G. Piseгна, Equilibrium to off-equilibrium crossover in homogeneous active matter, *Physical Review Research* **3**, 013210 (2021).
- [7] A. Cavagna, L. D. Carlo, I. Giardina, T. S. Grigera, S. Melillo, L. Parisi, G. Piseгна, and M. Scandolo, Natural swarms in 3.99 dimensions (2021), arXiv:2107.04432 [cond-mat.stat-mech].
- [8] D. Geiss, K. Kroy, and V. Holubec, Brownian molecules formed by delayed harmonic interactions, *New Journal of Physics* **21**, 093014 (2019).
- [9] A. Attanasi, A. Cavagna, L. Del Castello, I. Giardina, S. Melillo, L. Parisi, O. Pohl, B. Rossaro, E. Shen, E. Silvestri, *et al.*, Finite-size scaling as a way to probe near-criticality in natural swarms, *Physical Review Letters* **113**, 238102 (2014).
- [10] A. Cavagna, D. Conti, C. Creato, L. Del Castello, I. Giardina, T. S. Grigera, S. Melillo, L. Parisi, and M. Viale, Dynamic scaling in natural swarms, *Nature Physics* **13**, 914 (2017).
- [11] A. Cavagna, I. Giardina, and T. S. Grigera, The physics of flocking: Correlation as a compass from experiments to theory, *Physics Reports* **728**, 1 (2018).
- [12] A. Cavagna, D. Conti, I. Giardina, T. S. Grigera, S. Melillo, and M. Viale, Spatio-temporal correlations in models of collective motion ruled by different dynamical laws, *Physical Biology* **13**, 065001 (2016).
- [13] M. E. Fisher and M. N. Barber, Scaling theory for finite-size effects in the critical region, *Physical Review Letters* **28**, 1516 (1972).
- [14] F. Ginelli, The physics of the vicsek model, *The European Physical Journal Special Topics* **225**, 2099 (2016).
- [15] U. Khadka, V. Holubec, H. Yang, and F. Cichos, Active particles bound by information flows, *Nature communications* **9**, 1 (2018).
- [16] M. Fränz, S. Muiños-Landin, V. Holubec, and F. Cichos, Fully steerable symmetric thermoplasmonic microswimmers, *ACS Nano* **15**, 3434 (2021).
- [17] S. Muiños-Landin, A. Fischer, V. Holubec, and F. Cichos, Reinforcement learning with artificial microswimmers, *Science Robotics* **6**, 10.1126/scirobotics.abd9285 (2021).
- [18] J. Jhawar, R. G. Morris, U. R. Amith-Kumar, M. Danny Raj, T. Rogers, H. Rajendran, and V. Guttal, Noise-induced schooling of fish, *Nature Physics* **16**, 488 (2020).
- [19] D. R. J. Laming, *Information theory of choice-reaction times* (Academic Press, 1968).
- [20] W. Welford, J. M. Brebner, and N. Kirby, *Reaction times* (Stanford University, 1980).
- [21] T. Beatus, J. M. Guckenheimer, and I. Cohen, Controlling roll perturbations in fruit flies, *Journal of The Royal Society Interface* **12**, 20150075 (2015).
- [22] L. Ristroph, G. Ristroph, S. Morozova, A. J. Bergou, S. Chang, J. Guckenheimer, Z. J. Wang, and I. Cohen, Active and passive stabilization of body pitch in insect flight, *Journal of The Royal Society Interface* **10**, 20130237 (2013).
- [23] L. Ristroph, A. J. Bergou, G. Ristroph, K. Coumes, G. J. Berman, J. Guckenheimer, Z. J. Wang, and I. Cohen, Discovering the flight autostabilizer of fruit flies by inducing aerial stumbles, *Proceedings of the National Academy of Sciences* **107**, 4820 (2010).
- [24] H. Pomeroy and F. Heppner, Laboratory determination of startle reaction time of the starling (*sturnus vulgaris*), *Animal Behaviour* **25**, 720 (1977).
- [25] R. C. Eaton, R. A. Bombardieri, and D. L. Meyer, The mauthner-initiated startle response in teleost fish, *Journal of Experimental Biology* **66**, 65 (1977).
- [26] R. C. Eaton, *Neural mechanisms of startle behavior* (Springer Science & Business Media, 1984).
- [27] P. Lenz and D. Hartline, Reaction times and force production during escape behavior of a calanoid copepod, *undinula vulgaris*, *Marine Biology* **133**, 249 (1999).
- [28] B. Halperin and P. Hohenberg, Scaling laws for dynamic critical phenomena, *Physical Review* **177**, 952 (1969).
- [29] D. S. Cambui, A. S. de Arruda, and M. Godoy, Critical exponents of a self-propelled particles system, *Physica A: Statistical Mechanics and its Applications* **444**, 582 (2016).
- [30] T. D. Frank, P. J. Beek, and R. Friedrich, Fokker-planck perspective on stochastic delay systems: Exact solutions and data analysis of biological systems, *Physical Review E* **68**, 021912 (2003).
- [31] P. C. Hohenberg and B. I. Halperin, Theory of dynamic critical phenomena, *Reviews of Modern Physics* **49**, 435 (1977).



## Information conduction and convection in noiseless Vicsek flocks

Daniel Geiß<sup>1,2</sup>, Klaus Kroy,<sup>1</sup> and Viktor Holubec<sup>3,\*</sup>

<sup>1</sup>*Institut für Theoretische Physik, Universität Leipzig, Postfach 100 920, D-04009 Leipzig, Germany*

<sup>2</sup>*Max Planck Institute for Mathematics in the Sciences, D-04103 Leipzig, Germany*

<sup>3</sup>*Charles University, Faculty of Mathematics and Physics, Department of Macromolecular Physics, V Holešovičkách 2, CZ-180 00 Praha, Czech Republic*



(Received 23 March 2022; accepted 30 June 2022; published 21 July 2022)

Physical interactions generally respect certain symmetries, such as reciprocity and energy conservation, which survive in coarse-grained isothermal descriptions. Active many-body systems usually break such symmetries intrinsically, on the particle level, so that their collective behavior is often more naturally interpreted as a result of information exchange. Here we study numerically how information spreads from a “leader” particle through an initially aligned flock, described by the Vicsek model without noise. In the low-speed limit of a static spin lattice, we find purely conductive spreading, reminiscent of heat transfer. Swarm motility and heterogeneity can break reciprocity and spin conservation. But what seems more consequential for the swarm response is that the dispersion relation acquires a significant convective contribution along the leader’s direction of motion.

DOI: [10.1103/PhysRevE.106.014609](https://doi.org/10.1103/PhysRevE.106.014609)

### I. INTRODUCTION

Transfer of information, energy, or mass through complex interacting networks is of ubiquitous interest in many scientific disciplines. As examples think of the World Wide Web and social media [1,2], epidemics [3–6], or heat conduction and diffusion [7–9]. In particular, information, rather than the elementary physical interactions transmitting it, is key to groups of motile living agents, such as bird flocks [10,11] or bacterial colonies [12,13]. To understand the behavior of such far-from-equilibrium many-body systems is a main task of the surging field of active matter [14–16]. Many new interesting phenomena have been uncovered, including motility-induced phase separation (MIPS) [17] and related forms of self-organization [18,19] and pattern formation [20,21]. Such studies could eventually lead to the development of novel types of “smart (meta-)materials” [22,23]. Yet systematic studies of the mechanisms of information spreading through active matter systems are still scarce.

In this work, we therefore analyze the information spreading in a two-dimensional Vicsek model (VM) [24], which is a paradigmatic model of dry active matter (without momentum conservation in the solvent) [25,26]. It provides a minimalistic description of active collective phenomena such as the formation of bird flocks or insect swarms. The VM resembles a magnet consisting of  $N$  spins, which describe the orientations of the self-propelled particles. Their positions advance at constant speed, while their orientations are subject to mutual alignment interactions with their neighbors. Compared to the limit of interacting lattice spins or also to the case of digital information transport through disordered static networks (frequently studied in network theory) the VM is ca-

pable of more complex behavior. Its neighbor configurations are neither regular nor static but constitute a dynamical graph [27,28]. As a consequence, information in the VM spreads not only by conduction but also by convection, hitchhiking with the motile particles [6]. Moreover, the information about particle positions and orientations is continuous, not digital.

In the following, we try to disentangle the various complications, by first studying information spreading on a static square lattice. For vanishing noise, this limit allows for an exact solution, which simplifies the analysis and provides good insight. Then we investigate the full deterministic (no noise) VM with nonzero velocity. For both cases, we study the information spreading for a scenario known as *flooding* in network theory [28–31]: Starting in an orientationally ordered state with a single “leader” particle that deviates from the rest, we investigate how its perturbing effect spreads to the others. So far, flooding dynamics was mostly studied for static graphs; but see Ref. [28] for a more general approach. To assess the spatiotemporal information spreading in the VM, we numerically determine the corresponding dispersion relation. Naturally, the convective flooding due to particle motion is found to dominate over conduction at higher speeds and over long distances. But it also gives rise to a considerable forward-backward symmetry breaking, rendering the dispersion relation spatially highly nonisotropic.

The paper is structured as follows: In Sec. II we introduce the VM. The zero-speed limit of the VM is discussed in Sec. III, which introduces the two flooding scenarios considered in this work: The *firm leader* with constrained spin orientation, which eventually guides the flock into a new direction; and the *lax leader*, which delivers an initial impulse but afterwards relaxes freely, like all other spins. Finally, in Sec. IV we consider the general case of nonvanishing particle speeds, where the dispersion relation becomes ambiguous, before we conclude in Sec. V.

\* viktor.holubec@mff.cuni.cz

## II. VICSEK MODEL

Since its introduction in 1995, many modifications of the original VM have been discussed in the literature [32]. Here we consider the deterministic discrete-time variant describing  $N$  particles self-propelling with constant speed  $v_0$  in two dimensions with topological alignment interactions. The position  $\mathbf{r}_i(t)$  and velocity  $\mathbf{v}_i(t)$  of  $i$ th particle obey the dynamical equations

$$\mathbf{v}_i(t+1) = v_0 \Theta[\mathbf{v}_i(t) + \sum_{j \neq i} n_{ij}(t) \mathbf{v}_j(t)], \quad (1)$$

$$\mathbf{r}_i(t+1) = \mathbf{r}_i(t) + \mathbf{v}_i(t+1), \quad (2)$$

where  $\Theta(\mathbf{v}) \equiv \mathbf{v}/|\mathbf{v}|$  normalizes the velocity. The connectivity matrix  $n_{ij}(t)$  defines the interaction network. We assume topological interactions: Each particle interacts with its  $N_{\text{int}}$  nearest neighbors at time  $t$ . For these  $n_{ij}(t) = 1$ , while it vanishes otherwise. We have tested that metric interactions, where each particle interacts with all neighbors within a given spatial distance, leads to qualitatively the same results (data not shown). In contrast to the standard VM, we neglect the noise.

Instead of using the particle velocities  $\mathbf{v}_i(t)$  to characterize the system state, one can equivalently describe it by the angular variables  $\theta_i(t)$ , defined by  $\mathbf{v}_i(t) = v_0(\cos \theta_i, \sin \theta_i)$ . In this language, Eq. (1) assumes the form [33]

$$\theta_i(t+1) = \theta_i(t) + \frac{1}{N_i(t)} \sum_{j \neq i} n_{ij}(t) \sin[\theta_j(t) - \theta_i(t)], \quad (3)$$

where  $N_i(t) \equiv v_0^{-1} |\mathbf{v}_i(t) + \sum_{j \neq i} n_{ij}(t) \mathbf{v}_j(t)|$  stems from the normalization in Eq. (1).

We consider the situation where one of the particles (the leader) in a completely polarized system suddenly changes its direction and initiates a collective maneuver [10,34–36], due to the spreading of information about its flight direction through the flock [37,38]. To analyze the spreading of information in different directions with respect to the leader's velocity, it is useful to position it initially, at time  $t = 0$ , in the center of the flock. In the next section, we investigate the information transfer for the static spin lattice ( $v_0 = 0$ ), where the information spreads only by conduction. The interplay of conduction and convection, appearing for nonzero particle velocity, is then addressed in Sec. IV.

## III. ZERO-VELOCITY LIMIT OF THE VM

### A. Linearized lattice VM and its continuous limits

To make contact with classical spin models, the particles are placed on grid points  $\mathbf{r}_i$  of a two-dimensional square lattice and interact only with their direct neighbors. In this limit, the dynamics of the well-known XY model is restored. In the following, we label the orientations  $\theta_k$  of the individual spins (or particles) by their positions  $ij$  in the lattice. If all spins on the square lattice are well aligned, Eq. (3) can be expanded in fluctuations around the aligned state as

$$\theta_{ij}(t+1) = \frac{1}{5} [\theta_{ij}(t) + \theta_{i-1j}(t) + \theta_{i+1j}(t) + \theta_{ij-1}(t) + \theta_{ij+1}(t)], \quad (4)$$

where we assumed that the average orientation of the system is 0 and  $\theta_{ij} \ll 1$ . In this limit, the periodic boundary conditions

for  $\theta_{ij}$  do not need to be taken into account. An analogous linear formulation of the low-velocity VM has recently been employed [39] to calculate the total number of particles in a Vicsek flock from the orientational diffusion coefficient of a single particle.

Noteworthy, the same equation describes occupation probabilities of the individual grid-points for a symmetric random walk on a two-dimensional square lattice with equal probabilities to stay at a given point or to jump to a neighboring point. Unlike the standard Vicsek model, it thus conserves the total amount of ‘‘information’’  $\sum_{ij} \theta(t)$  unless some of the lattice points serve as sources or sinks of information. Information conservation would also be lost for less symmetric lattices, breaking reciprocity of the interactions (for details, see the Appendix).

Besides being exactly solvable, the importance of this simplified lattice model for understanding of information transfer in the VM is its similarity to other physical models such as lattice models of ferromagnetism, where  $\theta_{ij}(t)$  describes spin of the given grid point [40–44], the Google Search PageRank algorithm [45,46], measuring the importance of a web page by counting all links to it and weighting them by their quality, the majority vote model, and, most importantly, lattice models of heat conduction [47,48].

A central finding from the latter is that the heat flux is well described by Fourier's law implying that the local temperature  $\theta$  obeys the parabolic (diffusion)

$$\partial_t \theta = D \nabla^2 \theta \quad (5)$$

with the diffusion coefficient  $D$ . However, this equation leads to unphysical infinite propagation speed of heat [47,49], in the sense that a change in the temperature at the origin leads to infinitesimal changes in temperature far from the origin after an infinitesimally short time. Another issue is that Eq. (5) in general cannot describe the propagation of second sound, i.e., the thermal wave [8] encountered in low-temperature physics [50]. The most popular and simplest generalization of Eq. (5) which can describe both diffusive and wave-like transfer is the hyperbolic equation

$$\partial_t \theta + \frac{\tau}{2} \partial_t^2 \theta = \frac{\tau}{2} c^2 \nabla^2 \theta, \quad (6)$$

with maximum heat transfer velocity  $c$  and a characteristic time  $\tau$ . A standard derivation of this equation is based on Cattaneo's generalization of Fourier's law [51,52].

Interestingly, it turns out that both these equations are special limiting case of Eq. (4) [47,48]. Specifically, introducing a lattice constant  $\ell$  and the time  $\tau$  the signal needs to travel between two lattice points, it can be rewritten as

$$\begin{aligned} \theta(x, y, t + \tau) = & \frac{1}{5} [\theta(x, y, t) + \theta(x + \ell, y, t) \\ & + \theta(x - \ell, y, t) + \theta(x, y + \ell, t) \\ & + \theta(x, y - \ell, t)]. \end{aligned} \quad (7)$$

Now, taking the continuum limit  $\tau \rightarrow 0$  and  $\ell \rightarrow 0$ , while keeping constant the ratio  $5D \equiv \ell^2/\tau$  yields in the zeroth order in  $\tau$  the diffusion equation (5). On the other hand, taking the limit while keeping constant the velocity  $c\sqrt{5}/2 \equiv \ell/\tau$  leads in the first order in  $\tau$  to the hyperbolic equation (6). These nonstandard definitions of speed and diffusion coefficient result from the term  $\theta(x, y, t)$  on the right-hand side

of Eq. (7), which is not present in a standard random walk. The speed  $c$ , denoting the maximum speed of propagation in Eq. (6), is smaller than the maximum speed of propagation in the lattice model,  $v = l/\tau$ . Identifying  $\tau c^2/2$  in Eq. (6) with  $D$ , one can consider the parabolic equation (5) as a limit of infinitely fast ( $\tau = 0$ ) signal transmission between the neighboring lattice points. While this limit is often a good approximation for heat conduction [8] it might not be appropriate for biological agents with finite response time. A general statement about which of the two continuum limits fits better the description of the VM is not possible. It can heavily depend on the quantity of interest and the chosen parameters. Nonetheless, from our analysis below, it follows that the information spreading in the VM is approximately diffusive for small speeds  $v_0$  and increasingly nondiffusive as  $v_0$  grows.

### B. Firm and lax leaders

We now consider the following two specific flooding scenarios for the static VM (4). (1) In the firm leader scenario, the leader's orientation is held fixed. Measuring the angular variables  $\theta_{ij} \ll 1$  in units of the initial orientation of the leader, we set  $\theta_{00}(t) = 1$  for all times. This amounts to a steady information influx into the system. (2) In the lax leader scenario, the orientation of the leader is set to 1 at time 0 but then evolves according to Eq. (4). In both scenarios, all other particles are initially aligned with the  $x$  axis,  $\theta_{ij}(0) = 0$  for  $ij \neq 00$ . While (1) can be interpreted as a flock following a leader, (2) might describe a flock reacting to a sudden perturbation.

In the firm-leader scenario, the dynamical equation (4) is most easily written and solved using the matrix form  $\boldsymbol{\theta}(t+1) = M_0 \boldsymbol{\theta}(t) + \boldsymbol{\theta}(0)$ , where the vector  $\boldsymbol{\theta}(t)$  contains the values of orientations at all grid points at time  $t$ ,  $\theta_{ij}(t)$ , and  $M_0$  incorporates the interactions. It has vanishing entries for the feedback onto the leader's orientation, which is set by  $\boldsymbol{\theta}(0)$ , which has vanishing entries for all other particles. The solution is  $\boldsymbol{\theta}(t) = \sum_{i=0}^t M_0^i \boldsymbol{\theta}(0)$ . In the lax leader scenario, the dynamical equation is  $\boldsymbol{\theta}(t+1) = M \boldsymbol{\theta}(t)$ , and  $M$  incorporates the interactions between all the grid points, as described by Eq. (4), including the feedback onto the leader. The solution is  $\boldsymbol{\theta}(t) = M^t \boldsymbol{\theta}(0)$ . Both solutions nicely demonstrate that due to the linearity of the dynamics, the transmission of the information obeys the principle of superposition: The impact onto  $\theta_{ij}(t)$  depends on the number of possible paths of length  $t$  the signal may take from  $(0,0)$  to  $(i,j)$ , namely, the summation induced by the matrix multiplication in  $M_0^t$ . And it decays with time and distance due to the conservation enforced by the repeated normalization via the prefactor  $(1/5)^t$  in  $M_0^t$ .

In Fig. 1 we depict the information spreading in the linearized lattice VM for both scenarios. As expected, the information spreading quickly becomes isotropic, since discretizing the diffusion equation on a square lattice destroys the radial symmetry only for short paths and affects only the initial stage of the dynamics. The spreading for the firm leader scenario, with a fixed source at the origin, eventually aligns all particles to the leader. The rate of this approach decreases with growing distance of the grid points from the leader, and the saturation curves exhibit maximum slopes at intermediate times.

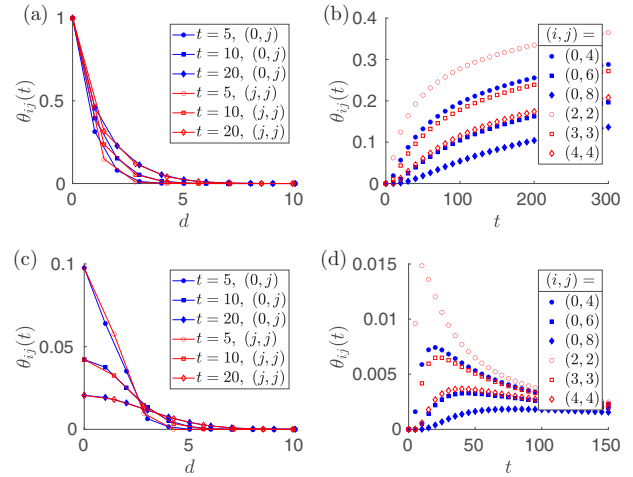


FIG. 1. Information spreading in the firm (a), (b) and lax (c), (d) leader scenarios. (a), (c) The spatial spreading of the orientation  $\theta_{ij}(t)$  at different times over the distance  $d = \sqrt{i^2 + j^2}$ , transverse (filled blue markers) and diagonal (open red markers) relative to the leader. (b), (d) The time evolution of the orientation for different grid points.

### C. Signal speed

In general, there is no unique definition of the speed of information spreading in the linearized lattice VM. The most obvious definition  $v = l/\tau = 1$  refers to the signal transfer between neighboring lattice points [cf. Eqs. (4)–(7)]. It provides the time  $d/v$  after which a grid point at distance  $d$  from the leader starts to receive the information. Yet it is of limited use because the strength of the received information is negligible if the grid point is far from the leader and there are only a few paths for the signal between the leader and the grid point. For example, in the case of a single path the signal strength received at time  $d/v$  is proportional to  $(1/5)^d$ .

A more informative definition is obtained from the time  $T_{\max}(d)$  when the change of orientation induced by the leader at distance  $d$  becomes maximal. The rate of change of orientation of the grid points is measured by the time derivatives  $\dot{\theta}_{ij}(t)$ , which exhibit a clear maximum [cf. Figs. 1(b) and 1(d)]. One may thus identify  $T_{\max}(d)$  with the time when  $\dot{\theta}_{ij}(t)$  with  $\sqrt{i^2 + j^2} = d$  is maximal. In Fig. 2(a) we show the resulting dispersion relation  $d(T_{\max})$  obtained from evaluating

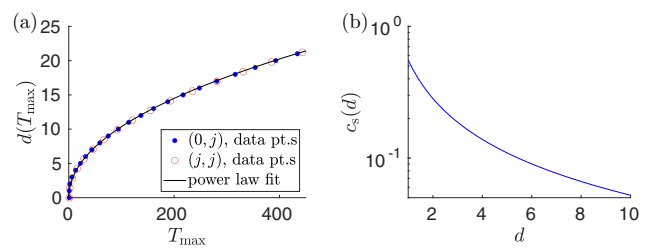


FIG. 2. (a) The dispersion relation for the linearized lattice VM transverse (blue filled circles) and diagonal (red open circles) relative to the leader. The solid line shows a fit  $d(t) = at^m$  with exponent  $m = 0.48$  and  $a \approx 1.1$ . (b) The corresponding signal speed  $c_s = \dot{d}$  as function of the distance  $d$  to the leader.

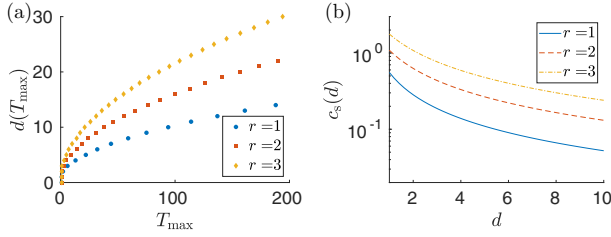


FIG. 3. (a) The dispersion relation for a generalized lattice VM where each grid point interacts with all neighbors at distances up to  $r$  lattice edges for three values of  $r$ . (b) The corresponding signal speeds  $c_s(d) = \dot{d}$ , respectively.

the signal propagation on the horizontal and on the diagonal axis with respect to the leader in the firm leader scenario. As expected, the found information spreading is well described by the diffusion relation  $d(T_{\max}) = \sqrt{4D_{\text{eff}}T_{\max}}$ . However, the diffusion coefficient  $D_{\text{eff}} = a^2/4 \approx 0.3$ , obtained by fitting the data, is much different than the diffusion coefficient  $D = 0.2$ , predicted from the limiting process leading to Eq. (5). In Fig. 2(b) we show the corresponding signal speed  $c_s = \dot{d}(t) \propto 1/\sqrt{t} \propto 1/d$ . The results obtained for the lax leader scenario are qualitatively the same (data not shown).

As an aside, we note that, while evaluating the evolution of the maxima of  $\dot{\theta}_i(t)$  is a reasonable approach for studying the signal spreading in the two flooding scenarios considered here, it is not suitable for more complex situations. A more universally applicable proxy for signal speed can be obtained by evaluating the connected acceleration correlations [10]. For our specific setting with a single leader and aligned initial state, the two approaches lead to the same results.

To close this section, we investigate the information spreading in a direct generalization of the linearized lattice VM (4), where the individual grid points interact not only with their nearest neighbors, but also with all grid points up to a distance of  $r$  lattice edges from the leader. Consequently, each grid point interacts with its  $N_{\text{int}} = 2r(r+1)$  nearest neighbors. The maximum speed of information propagation,  $v$ , is determined just by distances between particles at the circumference of the interaction zone, and thus it increases linearly with  $r$ . On the other hand, the  $r$  dependence of the speed  $c_s$ , shown in Fig. 3, is sublinear as the signal maximum is “slowed down” by the particles inside the interaction radius. Interestingly, the curves for different  $r$  values cannot be collapsed into a single master curve by multiplying each of them by a constant factor. Our analysis suggests that such a collapse is possible for long times only, with numerically obtained scaling factors 1.565 and 2.105 yielding the best asymptotic collapse of the curves for  $r = 1$  to those for  $r = 2$  and  $r = 3$ , respectively. These factors are close to the factors 1.64 ( $r = 1 \rightarrow 2$ ) and 2.28 ( $r = 1 \rightarrow 3$ ) obtained from the diffusion limit (5) of the individual lattice models as  $\sqrt{D_r/D_1}$  with  $D_r$  denoting the diffusion coefficient obtained for the individual values of  $r$ . Even though the diffusive scaling  $d = \sqrt{4D_r t}$ , predicted from Eq. (5), does not describe the data perfectly (in particular the prefactor  $4D_r$  is wrong), we take this as an indication that the formula  $d = \sqrt{4D_{\text{eff}}t}$  with  $D_{\text{eff}} \sim D_r$  is a

reasonable qualitative model for the spreading of information over long time and large length scales.

To sum up, in the linearized static spin model, the information spreads essentially diffusively. We next investigate how the situation changes when we allow particles to translate along their orientations.

## IV. THE MOTILE CASE

### A. The role of convection

Compared to equilibrium systems, active matter breaks certain local symmetries such as momentum and energy conservation. It is not *a priori* obvious whether this fundamental difference will lead to important effects on the information spreading and swarm behavior or if it is largely irrelevant, in practice. Differently from the lattice VM, the standard VM does not place the particles onto a regular lattice, and if so, this order would not be maintained for long. Notice that this breaks two important symmetries, namely, reciprocity and information conservation (see the Appendix). While the disorder itself does not affect the diffusive information spreading, the evolution of the neighborhood relations for  $v_0 > 0$  additionally allows for information convection. This situation is thus very similar to a moving heat source with the main difference that particles addressed by the leader tend to follow it, while heated passive particles generally do not induce a comparable flux.

Let us now derive a rough estimate for the particle speed  $v_0$  at which convection becomes important. The maximum conduction speed is given by the speed with which the signal spreads due to the interactions, i.e.,  $\ell_{\text{int}}/\Delta t$ . Here  $\Delta t = 1$  is the discrete update time in the VM and  $\ell_{\text{int}} = \sqrt{N_{\text{int}}/(\pi\rho)} = \sqrt{N_{\text{int}}/N}$  is the average interaction radius, assuming a more or less homogeneous density  $\rho = N/\pi$  after initiation inside a unit circle. The speed of convection is given by the relative speed of the individual particles on the order of  $v_0$ . Conduction and convection should thus compete when  $v_0 \approx \sqrt{N_{\text{int}}/N}$ . Alternatively, as in Sec. III C, we could measure the speed of signal propagation by the ratio  $d_i/T_i$ , where  $T_i$  is the time when the signal sent at time 0 causes a maximum change  $\dot{\theta}_i(t)$  of orientation at distance  $d_i$ , i.e.,  $\dot{\theta}_i(T_i) \equiv \max_t \dot{\theta}_i(t)$ . As we find below, the latter approach, which predicts a significantly lower conduction speed, is more appropriate to describe the data, yielding a correspondingly lower threshold velocity for the onset of convective transport. (For our choice of parameters, convection plays role already for velocities of about  $v_0 = 0.01$  while  $\sqrt{N_{\text{int}}/N} \approx 0.16$ .)

Besides inducing convection, motility further complicates the definition of a signal speed. Due to the relative motions of the particles there is no *a priori* choice of the distance  $d_i$  traveled by a signal. For this reason, we analyzed the speed of information propagation using two different definitions of  $d_i$ . First, the (average) initial distance  $|\mathbf{r}_i(0) - \mathbf{r}_L(0)|$  between the particles and the leader, which is the initial position of particle  $i$  at time 0, also encoded in the initial density  $\rho$ . Second, the distance  $|\mathbf{r}_i(T_i) - 1/N \sum_j \mathbf{r}_j(T_i)|$  between the particle  $i$  and the position of the center (of mass) of the flock at the characteristic “interaction time”  $T_i$  for conductive transport. We have performed the analysis below for both these definitions of the



distance and found no qualitative differences. Therefore, we show only the results obtained for the former, in the following.

### B. Numerical procedure

In the simulations, we place the leader always into the center of a unit circle. Positions of all  $N = 1000$  other particles are picked randomly inside the circle. All particles interact with their  $N_{\text{int}} = 24$  nearest neighbors, corresponding to  $r = 3$  in Fig. 3. Small density fluctuations in the initial condition are found to induce strong noise in the measured functions  $\theta_i(t)$  and  $\dot{\theta}_i(t)$ . To be able to determine the overall trend from these measurements, we averaged the resulting curves over  $N_{\text{runs}}$  runs with different initial conditions. We also employed two different smoothening procedures:

The average  $\langle \cdot \rangle_{nm}$  is calculated as follows. First, we collect the data  $\{d_i, \theta_i(t), \dot{\theta}_i(t)\}_{i=1, \dots, N}$  from  $N_{\text{runs}} = 100$  runs of the simulation. Then we sort the data according to the distance  $d_i$  to the origin at time 0. Finally, we calculate the smoothed variables  $\langle d_i \rangle_{nm}$ ,  $\langle \theta_i(t) \rangle_{nm}$ , and  $\langle \dot{\theta}_i(t) \rangle_{nm}$  by averaging  $d_i$ ,  $\theta_i(t)$ , and  $\dot{\theta}_i(t)$  over  $N_{\text{av}}$  neighbors of the particle  $i$ , i.e., over the particles  $j$  with  $N_{\text{av}}$  smallest distances  $|d_i - d_j|$ . Since the dispersion relation is a strictly monotonous function of time, one may alternatively perform the averaging with respect to nearest neighbors in time  $T_i$ , according to when the maximum signal has arrived at particle  $i$ . In other words one can average over the particles  $j$  with  $N_{\text{av}}$  smallest distances  $|T_i - T_j|$ . We have tested that both averaging procedures lead to qualitatively the same results. In the following, we show only those obtained using the averaging  $\langle \cdot \rangle_{nm}$  over  $N_{\text{av}}$  spatially nearest neighbors.

### C. Firm leader scenario

We now consider the firm leader scenario of Sec. III B, where the leader's orientation is fixed to  $\varphi$  at all times and all other particles are initially aligned with the perpendicular  $x$  axis and subsequently obey the dynamical equations (1) and (2). Note that this condition implies that reciprocity between the leader and the flock is maximally broken.

In Figs. 4(a) and 4(b) we show the resulting averaged orientations,  $\langle \theta_i(t) \rangle_{nm}$ , and the averaged changes in the orientation,  $\langle \dot{\theta}_i(t) \rangle_{nm}$ , as functions of the averaged distance  $\langle d_i \rangle_{nm}$ . To investigate the directional dependence of the information spreading, we distinguish between two directions of signal propagation. As the leader's orientation points into the positive half-plane, we identify the particles with positive  $y$  coordinates at time 0 as lying in the ‘‘positive direction’’ with respect to the leader. The remaining particles are lying in the ‘‘negative direction.’’ The results for  $\langle \theta_i(t) \rangle_{nm}$  and  $\langle \dot{\theta}_i(t) \rangle_{nm}$  for the positive and negative directions are given in Figs. 4(c) and 4(d) and Figs. 4(e) and 4(f), respectively. As the leader carries the source of information with it, particles lying in the positive direction show a significantly larger change of orientation than those in the negative direction. Furthermore, the leader affects nearby particles more than more distant ones. This leads to correspondingly stronger average direction changes  $\langle \dot{\theta}_i(t) \rangle_{nm}$  in its vicinity. Consequently, upon traversing the flock, the leader seems to drag around a cloud of ‘‘followers.’’ However, since the interaction rule allows only imperfect alignments, particles begin to realign with the less informed surroundings

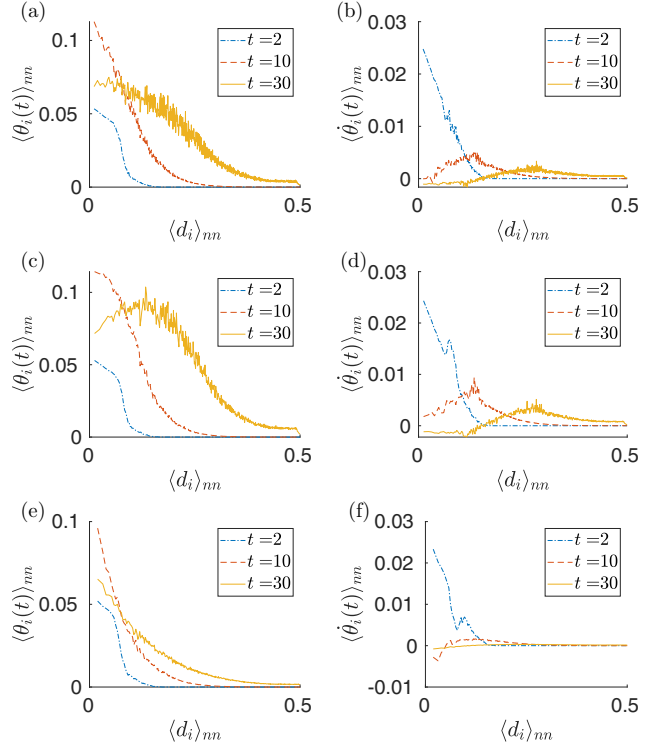


FIG. 4. Firm leader scenario. The average orientation,  $\langle \theta_i(t) \rangle_{nm}$  and change in orientation  $\langle \dot{\theta}_i(t) \rangle_{nm}$ , as functions of the average distance  $\langle d_i \rangle_{nm}$  at three different times: For the whole swarm (a), (b), along the leader direction (c), (d), and along the negative leader direction (e), (f). Parameters used:  $N = 1000$ ,  $v_0 = 0.01$ ,  $N_{\text{int}} = 24$ ,  $\varphi = \pi/4$ ,  $N_{\text{runs}} = 100$ ,  $N_{\text{av}} = N_{\text{runs}}$ .

after the leader has left their neighborhood. This is depicted by the moving maxima of  $\langle \theta_i(t) \rangle_{nm}$  and  $\langle \dot{\theta}_i(t) \rangle_{nm}$  in Figs. 4(c) and 4(d). In the negative direction, where the information propagates by pure conduction, no such structure is visible. The response of the swarm as a whole is dominated by the dynamics in positive direction. Repeating the described analysis for  $v_0 = 0$ , we found the same behavior as for the linearized lattice VM in Sec. III.

In Fig. 5(a) we show the time evolution of the change of orientation  $\langle \dot{\theta}_i(t) \rangle_{N, N_{\text{runs}}}$  averaged over all particles in the chosen particle set (total system, positive direction, and negative

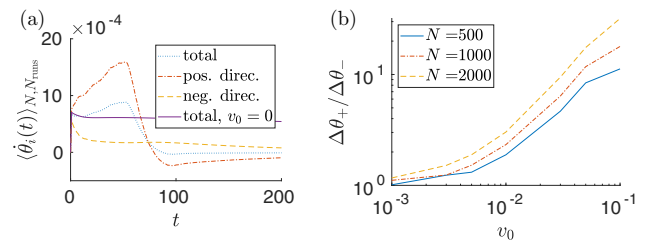


FIG. 5. Firm leader scenario. (a) Time evolutions of the change of direction averaged over the total system and the positive and negative directions for  $v_0 = 0.01$  and  $v_0 = 0$ , respectively. (b) The corresponding ratio (8) as function of the particle speed. Other parameters as in Fig. 4.

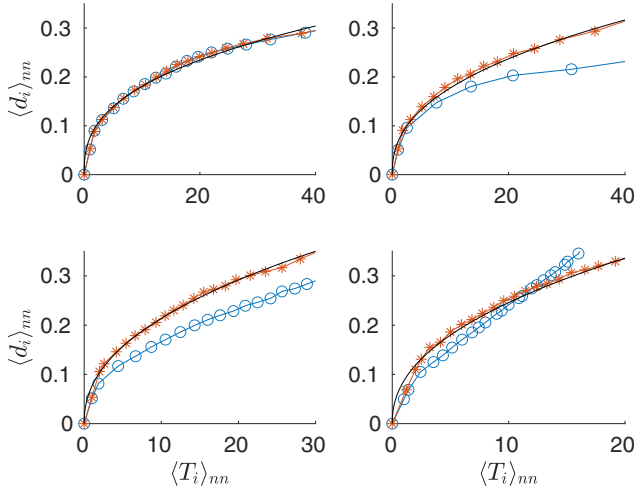


FIG. 6. Dispersion relation for the firm-leader scenario for positive (blue circles) and negative (red stars) directions and speeds  $v_0 = 0, 0.001, 0.01, 0.03$  increasing from the upper left to the bottom right panel. The black solid lines are fits of  $a(T_i)_{nm}^m$  to the data for the positive direction with  $m = 0.376, 0.417, 0.449, 0.451$  corresponding to the individual speeds. The slope of the data for speed  $v_0 = 0.03$  for the positive direction is  $\approx 0.018$  while the corresponding speed of the leader projected to the positive direction is  $v_0 \sin \varphi \approx 0.021$ . Parameters as in Fig. 4, except for  $N_{Av} = 10N_{runs}$ .

direction) and all simulations for  $v_0 > 0$  and  $v_0 = 0$ . For  $v_0 > 0$ , the average signal strength for the total system and particularly in the positive direction continuously increases until the leader approaches the edge of the system. This can be understood as follows. The change of orientation of the individual particles is largest when their orientation differs most from the average orientation of their neighbors. A moving leader constantly meets on its way in the positive direction particles (almost) aligned with the  $x$  axis, leading to a steady increase of the corresponding signal. Subsequently, at times  $t \gtrsim 50$  when the leader has left the flock,  $\langle \dot{\theta}_i(t) \rangle_{N, N_{runs}}$  rapidly decreases, and eventually it also changes sign. Since the leader mainly affects nearby particles, more distant particles are much less aligned with its orientation when it leaves. Therefore, particles that aligned with the leader during its passage through the flock begin to realign with the less affected particles after the leader has left. In the negative direction, the signal strength monotonously decreases, similarly as for  $v_0 = 0$  and the linearized lattice VM.

To quantify the asymmetry between the positive and negative direction, we integrate the positive areas

$$\Delta\theta_{\pm} = \int_{t_0}^{\infty} \langle \dot{\theta}_i(t) \rangle_{N, N_{runs}}^{\pm} \Theta(\langle \dot{\theta}_i(t) \rangle_{N, N_{runs}}^{\pm}), \quad (8)$$

beneath the corresponding curves in Fig. 5(a). Here + (−) corresponds to the positive (negative) direction and  $\Theta(\cdot)$  denotes the unit step function. The ratio  $\Delta\theta_{+}/\Delta\theta_{-}$  is shown in Fig. 5(b). As expected, it monotonously increases with the particle speed  $v_0$  and particle density  $N/\pi$ .

The main result of this section are the dispersion relations for four different velocities shown in Fig. 6. Regardless of  $v_0$ , the information initially spreads conductively, hence similarly

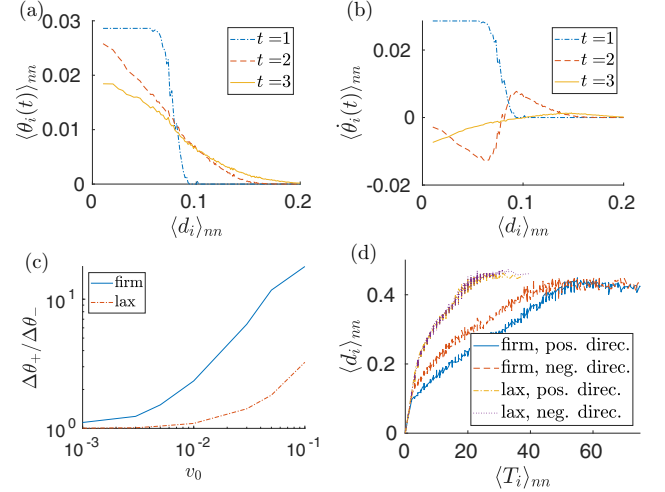


FIG. 7. Lax leader scenario. (a), (b) The orientation,  $\langle \theta_i(t) \rangle_{nm}$ , and the averaged change in the orientation,  $\langle \dot{\theta}_i(t) \rangle_{nm}$ , as functions of the averaged distance  $\langle d_i \rangle_{nm}$  for three different times averaged over the overall system. (c) The ratio (8) of the integrated positive changes in the orientation in the positive and negative directions as functions of the particle speed for the firm and lax leader scenario. (d) The dispersion relation for the positive and negative directions for the firm and lax leader scenario. We used the same parameters as in Fig. 4.

as in the linearized lattice VM, as diffusion beats convection over short times and distances. With increasing velocity  $v_0 > 0$ , the spreading in the positive direction becomes gradually more convective at late times. The slope of the dispersion relation converges to the velocity of the leader projected to the positive direction,  $v_0 \sin \varphi$ . In the negative direction, the spreading stays conductive regardless of  $v_0$ . Even though the particles in the negative direction are less affected by the turning event induced by the leader, the dispersion relation shows that the information reaches them faster than those in the positive direction. This counterintuitive effect is a consequence of the employed definition of  $T_i$ : The dispersion relation follows from determining times maximizing  $\dot{\theta}_i(t)$ . As the leader moves away from the particles behind it, the rates of change  $\dot{\theta}_i(t)$  of their orientations peak sooner than those in the positive direction. This somewhat counterintuitive behavior is reminiscent of observations of faster speeds for smaller pulses [8] or propagation of second sound against the heat flow [53].

#### D. Lax leader scenario

The information spreading is somewhat different in the lax leader scenario. For the same initial condition as in the preceding section, the leader now adapts dynamically according to Eqs. (1) and (2) to its neighbors for  $t > 0$ . It thereby virtually loses the information passed on to them. The interaction with the neighbors is thus more reciprocal than for the unwavering firm leader, yet not entirely so, since the topological notion of next neighbors is not necessarily fully reciprocal (see the Appendix).

In Figs. 7(a) and 7(b) we depict again the average orientation  $\langle \theta_i(t) \rangle_{nm}$  and the averaged change in the orientation

$\langle \dot{\theta}_i(t) \rangle_{mn}$  for the lax leader scenario. The parameters are the same as for the firm leader scenario in Fig. 4. Comparing the results for the two scenarios, we find the following differences: (1) the amplitudes of both  $\langle \theta_i(t) \rangle_{mn}$  and  $\langle \dot{\theta}_i(t) \rangle_{mn}$  are much smaller and the time derivative in the averaged angle converges to zero much faster, since the leader realigns with the rest of the flock in the lax leader scenario. (2) The time derivative  $\langle \dot{\theta}_i(t) \rangle_{mn}$  exhibits an excursion to negative values at small distances to the leader, due to the feedback from the flock, which requires a realignment of the leader and its neighborhood with the “winning majority” of other particles in the flock. At larger distances, the derivative returns to positive values, as expected for a moderate realignment of the merely slightly disturbed more distant particles. (3) There is again a directional dependence of the response, as in the firm-leader scenario. However, it is now much weaker, due to the mutual information exchange.

In fact, for the parameters used in the figure, it is not worthwhile to show the corresponding spatial distributions, as they would be hardly discernible from those for the total system in Fig. 7. The directional dependence of the information spreading in the lax leader scenario becomes noteworthy only for substantially larger speeds  $v_0$ , as demonstrated in Fig. 7(c). There we compare the ratio  $\Delta\theta_+/\Delta\theta_-$  of responses (8) integrated in the positive and negative directions between the lax and firm leader scenarios. Because of weaker total signal strength in the lax leader scenario, the changes of orientation  $\dot{\theta}_i(t)$  of the individual particles peak sooner, leading to steeper (but directionally barely distinguishable) dispersion relations; cf. Fig. 7(d).

## V. CONCLUSION

We studied transport of information about orientation of a leader in the Vicsek model (VM) with topological interactions. The two main mechanisms for propagation of information are conduction and convection. We have shown that the conductive aspect in the VM can well be understood using a simplified, exactly solvable variant of the model, where the individual particles are fixed at grid points of a regular lattice. This static spin lattice model allows for an analogy with heat transfer, which ceases to hold in the full dynamic VM. Nonlinearity and heterogeneity then break the underlying symmetries such as reciprocity and spin conservation. Yet this has no major practical consequences by itself. The visible changes between the dynamic model and the spin lattice are entirely dominated by the convective dynamics.

We considered two scenarios of information spreading from a single misaligned leader particle. While the diffusive or conductive spreading prevails over short times and distances, the spreading over longer times and distances gradually acquires a convective contribution, as the particle speed increases. We quantified this intuitive conclusion by measuring the dispersion relation. It was formulated for the timescale at which the signal induces its largest change in orientation of the particles at a given distance. The analysis revealed a strong directional dependence of the information transfer for the firm-leader scenario, in which the reciprocity of the mutual information exchange is maximally violated. A significant effect of convective information spreading is observed only

in the direction of the leader motion. In the wake zone behind the leader, the spreading remains diffusive, regardless of the speed.

While measuring the dispersion relation for zero speed of the particles is a relatively straightforward task, the definition of the distance over which the signal has propagated becomes ambiguous for the motile swarm. Nevertheless, we found that different length definitions lead to qualitatively close results.

Besides this ambiguity in the definition of the dispersion relation, which might deserve further analysis, our findings raise several questions. First, while some preliminary runs seemed to confirm the expectation that the inclusion of noise in the VM would yield qualitatively similar results, one could wish to study this issue more extensively, in particular with regard to the stability of the flocking transition; i.e., under which conditions can a leader move induce an ordering transition or the breakup of a flock? Further, many natural interaction networks are more heterogeneous than our flocks, containing, e.g., certain hierarchical structures [54,55] or distance- and density-weighted interactions [24]. Moreover, it might be interesting to consider a more realistic modification of the standard VM where the orientation of a particle under consideration would have a stronger weighting than the average orientation of its neighbors. This would yield a more persistent motion and might impact the information propagation. We took first steps in this direction in a follow-up study to the present work [56]. There we investigate information spreading in a 2D VM with time-delayed metric interactions [57] and also address the pertinence of the notion of linear response [58–60] and its relation to information propagation. Next, it might be interesting to connect information spreading in active matter with corresponding results in other research fields, such as network theory or epidemiology. Especially in the latter, the effects of network heterogeneity on the spreading of diseases is a widely studied aspect [6,61–63]. Finally, it would seem interesting to pursue the question how the interaction rules in the VM can be optimized to facilitate information transfer.

## ACKNOWLEDGMENTS

We acknowledge funding through a DFG-GACR cooperation by the Deutsche Forschungsgemeinschaft (DFG-Code KR 3381/6-1) and by the Czech Science Foundation (GACR Project No. 20-02955J). V.H. was supported by the Humboldt Foundation. D.G. acknowledges funding by International Max Planck Research Schools (IMPRS) as well as by the Deutscher Akademischer Austauschdienst (DAAD).

## APPENDIX: BREAKING OF INFORMATION CONSERVATION IN THE LINEARIZED VM

As discussed in Sec. III, the interaction rule in the linearized lattice VM,

$$\theta_{ij}(t+1) = \frac{1}{N_{ij}(t)} \sum_{(ij),(kl)} \theta_{kl}(t), \quad (\text{A1})$$

where the sum goes over all neighbors  $(kl)$  of  $(ij)$  including  $(ij)$  itself, and  $N_{ij} = \sum_{(ij),(kl)}$  denotes the number of

neighbors, conserves the total information content

$$\theta_{\text{tot}}(t) \equiv \sum_{ij} \theta_{ij}(t). \quad (\text{A2})$$

While the linearized Vicsek interaction rule yields reciprocal interactions if the interaction network is regular, it can render nonreciprocal interparticle interactions for irregular interaction networks, e.g., if the particle density of the system is inhomogeneous in space. The conservation condition  $\dot{\theta}_{\text{tot}}(t) = 0$  will then be broken. For a closed system, the reverse also holds: If the conservation is broken, this indicates the presence of some nonreciprocal interactions. As an illustrative example, consider the following closed interaction network consisting of three particles. Particle 1 and 3 interact solely with particle 2, which interacts with both 1 and 3. Assuming the initial condition  $\theta_1(0) = 1$  and  $\theta_{2,3}(0) = 0$ , then  $\theta_1(1) = 1/2$ ,  $\theta_2(1) = 1/3$ ,  $\theta_3(1) = 0$ . We thus see that  $\theta_{\text{tot}}(0) = 1 > \theta_{\text{tot}}(1) = 5/6$ . If we instead consider the initial

condition  $\theta_2(0) = 1$  and  $\theta_{1,3}(0) = 0$ , we find  $\theta_{\text{tot}}(0) = 1 < \theta_{\text{tot}}(1) = 3/2$ . These examples manifest a more general finding that  $\theta_{\text{tot}}(t)$  decreases when the information flows from a less to a more connected region, and vice versa. In this case, reciprocity is broken since the normalization  $N_{ij}$  of neighboring particles varies. Beyond the linear regime, the situation is more complicated as the normalization  $N_{ij}$  depends on the angular variables.

Similarly, also for topological interactions, the information content is not conserved. While each particle interacts with exactly the same number of neighbors (i.e.,  $N_{ij} = \text{const}$ ), density gradients may induce unilateral interactions. As an example, consider a closed system of four particles with topological interactions with two nearest neighbors. Let the particles 1, 2, and 3 reciprocally communicate with each other, while the distant particle 4 adjusts its direction to that of particles 2 and 3 without influencing them. Assuming the initial condition  $\theta_2(0) = 1$  and  $\theta_{1,3,4}(0) = 0$ , we find  $\theta_{1,\dots,4}(1) = 1/3$  and thus  $\theta_{\text{tot}}(0) = 1 < \theta_{\text{tot}}(1) = 4/3$ .

- 
- [1] K. Lerman and R. Ghosh, Information contagion: An empirical study of the spread of news on Digg and Twitter social networks, in *Proceedings of the 4th International Conference on Weblogs and Social Media (ICWSM)* (The AAAI Press, Menlo Park, CA, 2010), p. 90.
- [2] Y. Moreno, M. Nekovee, and A. F. Pacheco, Dynamics of rumor spreading in complex networks, *Phys. Rev. E* **69**, 066130 (2004).
- [3] R. Pastor-Satorras and A. Vespignani, Epidemic Spreading in Scale-Free Networks, *Phys. Rev. Lett.* **86**, 3200 (2001).
- [4] M. Boguñá and R. Pastor-Satorras, Epidemic spreading in correlated complex networks, *Phys. Rev. E* **66**, 047104 (2002).
- [5] H. W. Hethcote, The mathematics of infectious diseases, *SIAM Rev.* **42**, 599 (2000).
- [6] D. Levis, A. Diaz-Guilera, I. Pagonabarraga, and M. Starnini, Flocking-enhanced social contagion, *Phys. Rev. Research* **2**, 032056(R) (2020).
- [7] M. N. Özışik, *Heat Conduction* (John Wiley & Sons, New York, 1993).
- [8] D. D. Joseph and L. Preziosi, Heat waves, *Rev. Mod. Phys.* **61**, 41 (1989).
- [9] R. S. Brodkey and H. C. Hershey, *Transport Phenomena: A Unified Approach* (Brodkey Publishing, Columbus, 2003).
- [10] A. Cavagna, I. Giardina, and T. S. Grigera, The physics of flocking: Correlation as a compass from experiments to theory, *Phys. Rep.* **728**, 1 (2018).
- [11] A. Procaccini, A. Orlandi, A. Cavagna, I. Giardina, F. Zoratto, D. Santucci, F. Chiarotti, C. K. Hemelrijk, E. Alleva, G. Parisi, *et al.*, Propagating waves in starling, *Sturnus vulgaris*, flocks under predation, *Anim. Behav.* **82**, 759 (2011).
- [12] H.-P. Zhang, A. Beãžer, E.-L. Florin, and H. L. Swinney, Collective motion and density fluctuations in bacterial colonies, *Proc. Natl. Acad. Sci. USA* **107**, 13626 (2010).
- [13] E. Ben-Jacob, O. Schochet, A. Tenenbaum, I. Cohen, A. Czirok, and T. Vicsek, Generic modelling of cooperative growth patterns in bacterial colonies, *Nature (London)* **368**, 46 (1994).
- [14] G. Gompper, R. G. Winkler, T. Speck, A. Solon, C. Nardini, F. Peruani, H. Löwen, R. Golestanian, U. B. Kaupp, L. Alvarez *et al.*, The 2020 motile active matter roadmap, *J. Phys.: Condens. Matter* **32**, 193001 (2020).
- [15] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe, Active particles in complex and crowded environments, *Rev. Mod. Phys.* **88**, 045006 (2016).
- [16] S. Ramaswamy, The mechanics and statistics of active matter, *Annu. Rev. Condens. Matter Phys.* **1**, 323 (2010).
- [17] M. E. Cates and J. Tailleur, Motility-induced phase separation, *Annu. Rev. Condens. Matter Phys.* **6**, 219 (2015).
- [18] M. F. Hagan and A. Baskaran, Emergent self-organization in active materials, *Curr. Opin. Cell Biol.* **38**, 74 (2016).
- [19] T. Bäuerle, A. Fischer, T. Speck, and C. Bechinger, Self-organization of active particles by quorum sensing rules, *Nat. Commun.* **9**, 1 (2018).
- [20] F. D. C. Farrell, M. C. Marchetti, D. Marenduzzo, and J. Tailleur, Pattern Formation in Self-Propelled Particles with Density-Dependent Motility, *Phys. Rev. Lett.* **108**, 248101 (2012).
- [21] A. P. Solon, J.-B. Caussin, D. Bartolo, H. Chaté, and J. Tailleur, Pattern formation in flocking models: A hydrodynamic description, *Phys. Rev. E* **92**, 062111 (2015).
- [22] M. Brambilla, E. Ferrante, M. Birattari, and M. Dorigo, Swarm robotics: A review from the swarm engineering perspective, *Swarm Intelligence* **7**, 1 (2013).
- [23] F. J. Vernerey, E. Benet, L. Blue, A. Fajrial, S. L. Sridhar, J. Lum, G. Shakya, K. Song, A. Thomas, and M. Borden, Biological active matter aggregates: Inspiration for smart colloidal materials, *Adv. Colloid Interface Sci.* **263**, 38 (2019).
- [24] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, Novel Type of Phase Transition in a System of Self-Driven Particles, *Phys. Rev. Lett.* **75**, 1226 (1995).
- [25] F. Ginelli, The physics of the Vicsek model, *Eur. Phys. J.: Spec. Top.* **225**, 2099 (2016).

- [26] T. Vicsek and A. Zafeiris, Collective motion, *Phys. Rep.* **517**, 71 (2012).
- [27] F. Kuhn and R. Oshman, Dynamic networks: Models and algorithms, *ACM SIGACT News* **42**, 82 (2011).
- [28] A. Clementi, R. Silvestri, and L. Trevisan, Information spreading in dynamic graphs, *Distrib. Comput.* **28**, 55 (2015).
- [29] A. E. Clementi, F. Pasquale, A. Monti, and R. Silvestri, Information spreading in stationary Markovian evolving graphs, in *2009 IEEE International Symposium on Parallel & Distributed Processing* (IEEE Computer Society, Los Alamitos, CA, 2009), pp. 1–12.
- [30] R. Van Der Hofstad, G. Hooghiemstra, and P. Van Mieghem, The flooding time in random graphs, *Extremes* **5**, 111 (2002).
- [31] S. Melnik, H. Garcia-Molina, and E. Rahm, Similarity flooding: A versatile graph matching algorithm and its application to schema matching, in *Proceedings 18th International Conference on Data Engineering* (IEEE Computer Society, Los Alamitos, CA, 2002), pp. 117–128.
- [32] H. Chaté, F. Ginelli, G. Grégoire, F. Peruani, and F. Raynaud, Modeling collective motion: Variations on the Vicsek model, *Eur. Phys. J. B* **64**, 451 (2008).
- [33] P. Degond, G. Dimarco, and T. B. N. Mac, Hydrodynamics of the Kuramoto–Vicsek model of rotating self-propelled particles, *Math. Models Methods Appl. Sci.* **24**, 277 (2014).
- [34] A. Attanasi, A. Cavagna, L. Del Castello, I. Giardina, T. S. Grigera, A. Jelić, S. Melillo, L. Parisi, O. Pohl, E. Shen *et al.*, Information transfer and behavioural inertia in starling flocks, *Nat. Phys.* **10**, 691 (2014).
- [35] A. Cavagna, L. Del Castello, I. Giardina, T. Grigera, A. Jelic, S. Melillo, T. Mora, L. Parisi, E. Silvestri, M. Viale *et al.*, Flocking and turning: A new model for self-organized collective motion, *J. Stat. Phys.* **158**, 601 (2015).
- [36] A. Cavagna, I. Giardina, T. S. Grigera, A. Jelic, D. Levine, S. Ramaswamy, and M. Viale, Silent Flocks: Constraints on Signal Propagation across Biological Groups, *Phys. Rev. Lett.* **114**, 218101 (2015).
- [37] J. K. Parrish and W. M. Hamner, *Animal Groups in Three Dimensions: How Species Aggregate* (Cambridge University Press, Cambridge, 1997).
- [38] J. E. Treherne and W. A. Foster, Group transmission of predator avoidance behaviour in a marine insect: The Trafalgar effect, *Anim. Behav.* **29**, 911 (1981).
- [39] P. De Lellis and M. Porfiri, Inferring the size of a collective of self-propelled Vicsek particles from the random motion of a single unit, *Commun. Phys.* **5**, 86 (2022).
- [40] G. Joyce, Classical Heisenberg model, *Phys. Rev.* **155**, 478 (1967).
- [41] J. Kosterlitz, The critical properties of the two-dimensional XY model, *J. Phys. C* **7**, 1046 (1974).
- [42] M. Sentef, M. Kollar, and A. P. Kampf, Spin transport in Heisenberg antiferromagnets in two and three dimensions, *Phys. Rev. B* **75**, 214403 (2007).
- [43] L. Lima and A. Pires, Dynamics of the anisotropic two-dimensional XY model, *Eur. Phys. J. B* **70**, 335 (2009).
- [44] E. Sonin, Spin currents and spin superfluidity, *Adv. Phys.* **59**, 181 (2010).
- [45] L. Page, S. Brin, R. Motwani, and T. Winograd, The PageRank citation ranking: Bringing order to the web, Technical report, Stanford InfoLab (1999).
- [46] K. Avrachenkov and N. Litvak, The effect of new links on Google PageRank, *Stochastic Models* **22**, 319 (2006).
- [47] S. L. Sobolev, Transport processes and traveling waves in systems with local nonequilibrium, *Sov. Phys. Usp.* **34**, 217 (1991).
- [48] S. L. Sobolev, Discrete space-time model for heat conduction: Application to size-dependent thermal conductivity in nano-films, *Int. J. Heat Mass Transf.* **108**, 933 (2017).
- [49] G. Lebon, D. Jou, and J. Casas-Vázquez, *Understanding Non-equilibrium Thermodynamics* (Springer, Berlin, Heidelberg, 2008), Vol. 295.
- [50] L. Landau, Theory of the superfluidity of helium II, *Phys. Rev.* **60**, 356 (1941).
- [51] S. Goldstein, On diffusion by discontinuous movements, and on the telegraph equation, *Q. J. Mech. Appl. Math.* **4**, 129 (1951).
- [52] C. Cattaneo, Sur une forme de l'équation de la chaleur éliminant la paradoxique d'une propagation instantanée, *Comptes Rendus de l'Académie des Sci.* **247**, 431 (1958).
- [53] B. D. Coleman and D. R. Owen, On the nonequilibrium behavior of solids that transport heat by second sound, *Comput. Math. Appl.* **9**, 527 (1983).
- [54] I. D. Chase, Models of hierarchy formation in animal societies, *Behav. Sci.* **19**, 374 (1974).
- [55] M. Nagy, Z. Ákos, D. Biro, and T. Vicsek, Hierarchical group dynamics in pigeon flocks, *Nature (London)* **464**, 890 (2010).
- [56] D. Geiß, K. Kroy, and V. Holubec, Signal propagation and linear response in the delay Vicsek model, *arXiv:2205.12069* (2022).
- [57] V. Holubec, D. Geiss, S. A. M. Loos, K. Kroy, and F. Cichos, Finite-Size Scaling at the Edge of Disorder in a Time-Delay Vicsek Model, *Phys. Rev. Lett.* **127**, 258001 (2021).
- [58] A. Czirók, H. E. Stanley, and T. Vicsek, Spontaneously ordered motion of self-propelled particles, *J. Phys. A: Math. Gen.* **30**, 1375 (1997).
- [59] H. Chaté, F. Ginelli, G. Grégoire, and F. Raynaud, Collective motion of self-propelled particles interacting without cohesion, *Phys. Rev. E* **77**, 046113 (2008).
- [60] D. J. G. Pearce and L. Giomi, Linear response to leadership, effective temperature, and decision making in flocks, *Phys. Rev. E* **94**, 022612 (2016).
- [61] R. Olinky and L. Stone, Unexpected epidemic thresholds in heterogeneous networks: The role of disease transmission, *Phys. Rev. E* **70**, 030902(R) (2004).
- [62] L. A. Meyers, B. Pourbohloul, M. E. Newman, D. M. Skowronski, and R. C. Brunham, Network theory and SARS: Predicting outbreak diversity, *J. Theor. Biol.* **232**, 71 (2005).
- [63] G. Grossmann, M. Backenkoehler, and V. Wolf, Why ODE models for COVID-19 fail: Heterogeneity shapes epidemic dynamics, *medRxiv* (2021), doi:10.1101/2021.03.25.21254292.



## Information conduction and convection in noiseless Vicsek flocks

Daniel Geiß<sup>1,2</sup>, Klaus Kroy,<sup>1</sup> and Viktor Holubec<sup>3,\*</sup>

<sup>1</sup>*Institut für Theoretische Physik, Universität Leipzig, Postfach 100 920, D-04009 Leipzig, Germany*

<sup>2</sup>*Max Planck Institute for Mathematics in the Sciences, D-04103 Leipzig, Germany*

<sup>3</sup>*Charles University, Faculty of Mathematics and Physics, Department of Macromolecular Physics, V Holešovičkách 2, CZ-180 00 Praha, Czech Republic*



(Received 23 March 2022; accepted 30 June 2022; published 21 July 2022)

Physical interactions generally respect certain symmetries, such as reciprocity and energy conservation, which survive in coarse-grained isothermal descriptions. Active many-body systems usually break such symmetries intrinsically, on the particle level, so that their collective behavior is often more naturally interpreted as a result of information exchange. Here we study numerically how information spreads from a “leader” particle through an initially aligned flock, described by the Vicsek model without noise. In the low-speed limit of a static spin lattice, we find purely conductive spreading, reminiscent of heat transfer. Swarm motility and heterogeneity can break reciprocity and spin conservation. But what seems more consequential for the swarm response is that the dispersion relation acquires a significant convective contribution along the leader’s direction of motion.

DOI: [10.1103/PhysRevE.106.014609](https://doi.org/10.1103/PhysRevE.106.014609)

### I. INTRODUCTION

Transfer of information, energy, or mass through complex interacting networks is of ubiquitous interest in many scientific disciplines. As examples think of the World Wide Web and social media [1,2], epidemics [3–6], or heat conduction and diffusion [7–9]. In particular, information, rather than the elementary physical interactions transmitting it, is key to groups of motile living agents, such as bird flocks [10,11] or bacterial colonies [12,13]. To understand the behavior of such far-from-equilibrium many-body systems is a main task of the surging field of active matter [14–16]. Many new interesting phenomena have been uncovered, including motility-induced phase separation (MIPS) [17] and related forms of self-organization [18,19] and pattern formation [20,21]. Such studies could eventually lead to the development of novel types of “smart (meta-)materials” [22,23]. Yet systematic studies of the mechanisms of information spreading through active matter systems are still scarce.

In this work, we therefore analyze the information spreading in a two-dimensional Vicsek model (VM) [24], which is a paradigmatic model of dry active matter (without momentum conservation in the solvent) [25,26]. It provides a minimalistic description of active collective phenomena such as the formation of bird flocks or insect swarms. The VM resembles a magnet consisting of  $N$  spins, which describe the orientations of the self-propelled particles. Their positions advance at constant speed, while their orientations are subject to mutual alignment interactions with their neighbors. Compared to the limit of interacting lattice spins or also to the case of digital information transport through disordered static networks (frequently studied in network theory) the VM is ca-

pable of more complex behavior. Its neighbor configurations are neither regular nor static but constitute a dynamical graph [27,28]. As a consequence, information in the VM spreads not only by conduction but also by convection, hitchhiking with the motile particles [6]. Moreover, the information about particle positions and orientations is continuous, not digital.

In the following, we try to disentangle the various complications, by first studying information spreading on a static square lattice. For vanishing noise, this limit allows for an exact solution, which simplifies the analysis and provides good insight. Then we investigate the full deterministic (no noise) VM with nonzero velocity. For both cases, we study the information spreading for a scenario known as *flooding* in network theory [28–31]: Starting in an orientationally ordered state with a single “leader” particle that deviates from the rest, we investigate how its perturbing effect spreads to the others. So far, flooding dynamics was mostly studied for static graphs; but see Ref. [28] for a more general approach. To assess the spatiotemporal information spreading in the VM, we numerically determine the corresponding dispersion relation. Naturally, the convective flooding due to particle motion is found to dominate over conduction at higher speeds and over long distances. But it also gives rise to a considerable forward-backward symmetry breaking, rendering the dispersion relation spatially highly nonisotropic.

The paper is structured as follows: In Sec. II we introduce the VM. The zero-speed limit of the VM is discussed in Sec. III, which introduces the two flooding scenarios considered in this work: The *firm leader* with constrained spin orientation, which eventually guides the flock into a new direction; and the *lax leader*, which delivers an initial impulse but afterwards relaxes freely, like all other spins. Finally, in Sec. IV we consider the general case of nonvanishing particle speeds, where the dispersion relation becomes ambiguous, before we conclude in Sec. V.

\* viktor.holubec@mff.cuni.cz

## II. VICSEK MODEL

Since its introduction in 1995, many modifications of the original VM have been discussed in the literature [32]. Here we consider the deterministic discrete-time variant describing  $N$  particles self-propelling with constant speed  $v_0$  in two dimensions with topological alignment interactions. The position  $\mathbf{r}_i(t)$  and velocity  $\mathbf{v}_i(t)$  of  $i$ th particle obey the dynamical equations

$$\mathbf{v}_i(t+1) = v_0 \Theta[\mathbf{v}_i(t) + \sum_{j \neq i} n_{ij}(t) \mathbf{v}_j(t)], \quad (1)$$

$$\mathbf{r}_i(t+1) = \mathbf{r}_i(t) + \mathbf{v}_i(t+1), \quad (2)$$

where  $\Theta(\mathbf{v}) \equiv \mathbf{v}/|\mathbf{v}|$  normalizes the velocity. The connectivity matrix  $n_{ij}(t)$  defines the interaction network. We assume topological interactions: Each particle interacts with its  $N_{\text{int}}$  nearest neighbors at time  $t$ . For these  $n_{ij}(t) = 1$ , while it vanishes otherwise. We have tested that metric interactions, where each particle interacts with all neighbors within a given spatial distance, leads to qualitatively the same results (data not shown). In contrast to the standard VM, we neglect the noise.

Instead of using the particle velocities  $\mathbf{v}_i(t)$  to characterize the system state, one can equivalently describe it by the angular variables  $\theta_i(t)$ , defined by  $\mathbf{v}_i(t) = v_0(\cos \theta_i, \sin \theta_i)$ . In this language, Eq. (1) assumes the form [33]

$$\theta_i(t+1) = \theta_i(t) + \frac{1}{N_i(t)} \sum_{j \neq i} n_{ij}(t) \sin[\theta_j(t) - \theta_i(t)], \quad (3)$$

where  $N_i(t) \equiv v_0^{-1} |\mathbf{v}_i(t) + \sum_{j \neq i} n_{ij}(t) \mathbf{v}_j(t)|$  stems from the normalization in Eq. (1).

We consider the situation where one of the particles (the leader) in a completely polarized system suddenly changes its direction and initiates a collective maneuver [10,34–36], due to the spreading of information about its flight direction through the flock [37,38]. To analyze the spreading of information in different directions with respect to the leader's velocity, it is useful to position it initially, at time  $t = 0$ , in the center of the flock. In the next section, we investigate the information transfer for the static spin lattice ( $v_0 = 0$ ), where the information spreads only by conduction. The interplay of conduction and convection, appearing for nonzero particle velocity, is then addressed in Sec. IV.

## III. ZERO-VELOCITY LIMIT OF THE VM

### A. Linearized lattice VM and its continuous limits

To make contact with classical spin models, the particles are placed on grid points  $\mathbf{r}_i$  of a two-dimensional square lattice and interact only with their direct neighbors. In this limit, the dynamics of the well-known XY model is restored. In the following, we label the orientations  $\theta_k$  of the individual spins (or particles) by their positions  $ij$  in the lattice. If all spins on the square lattice are well aligned, Eq. (3) can be expanded in fluctuations around the aligned state as

$$\theta_{ij}(t+1) = \frac{1}{5} [\theta_{ij}(t) + \theta_{i-1j}(t) + \theta_{i+1j}(t) + \theta_{ij-1}(t) + \theta_{ij+1}(t)], \quad (4)$$

where we assumed that the average orientation of the system is 0 and  $\theta_{ij} \ll 1$ . In this limit, the periodic boundary conditions

for  $\theta_{ij}$  do not need to be taken into account. An analogous linear formulation of the low-velocity VM has recently been employed [39] to calculate the total number of particles in a Vicsek flock from the orientational diffusion coefficient of a single particle.

Noteworthy, the same equation describes occupation probabilities of the individual grid-points for a symmetric random walk on a two-dimensional square lattice with equal probabilities to stay at a given point or to jump to a neighboring point. Unlike the standard Vicsek model, it thus conserves the total amount of ‘‘information’’  $\sum_{ij} \theta(t)$  unless some of the lattice points serve as sources or sinks of information. Information conservation would also be lost for less symmetric lattices, breaking reciprocity of the interactions (for details, see the Appendix).

Besides being exactly solvable, the importance of this simplified lattice model for understanding of information transfer in the VM is its similarity to other physical models such as lattice models of ferromagnetism, where  $\theta_{ij}(t)$  describes spin of the given grid point [40–44], the Google Search PageRank algorithm [45,46], measuring the importance of a web page by counting all links to it and weighting them by their quality, the majority vote model, and, most importantly, lattice models of heat conduction [47,48].

A central finding from the latter is that the heat flux is well described by Fourier's law implying that the local temperature  $\theta$  obeys the parabolic (diffusion)

$$\partial_t \theta = D \nabla^2 \theta \quad (5)$$

with the diffusion coefficient  $D$ . However, this equation leads to unphysical infinite propagation speed of heat [47,49], in the sense that a change in the temperature at the origin leads to infinitesimal changes in temperature far from the origin after an infinitesimally short time. Another issue is that Eq. (5) in general cannot describe the propagation of second sound, i.e., the thermal wave [8] encountered in low-temperature physics [50]. The most popular and simplest generalization of Eq. (5) which can describe both diffusive and wave-like transfer is the hyperbolic equation

$$\partial_t \theta + \frac{\tau}{2} \partial_t^2 \theta = \frac{\tau}{2} c^2 \nabla^2 \theta, \quad (6)$$

with maximum heat transfer velocity  $c$  and a characteristic time  $\tau$ . A standard derivation of this equation is based on Cattaneo's generalization of Fourier's law [51,52].

Interestingly, it turns out that both these equations are special limiting case of Eq. (4) [47,48]. Specifically, introducing a lattice constant  $\ell$  and the time  $\tau$  the signal needs to travel between two lattice points, it can be rewritten as

$$\begin{aligned} \theta(x, y, t + \tau) = & \frac{1}{5} [\theta(x, y, t) + \theta(x + \ell, y, t) \\ & + \theta(x - \ell, y, t) + \theta(x, y + \ell, t) \\ & + \theta(x, y - \ell, t)]. \end{aligned} \quad (7)$$

Now, taking the continuum limit  $\tau \rightarrow 0$  and  $\ell \rightarrow 0$ , while keeping constant the ratio  $5D \equiv \ell^2/\tau$  yields in the zeroth order in  $\tau$  the diffusion equation (5). On the other hand, taking the limit while keeping constant the velocity  $c\sqrt{5}/2 \equiv \ell/\tau$  leads in the first order in  $\tau$  to the hyperbolic equation (6). These nonstandard definitions of speed and diffusion coefficient result from the term  $\theta(x, y, t)$  on the right-hand side



of Eq. (7), which is not present in a standard random walk. The speed  $c$ , denoting the maximum speed of propagation in Eq. (6), is smaller than the maximum speed of propagation in the lattice model,  $v = l/\tau$ . Identifying  $\tau c^2/2$  in Eq. (6) with  $D$ , one can consider the parabolic equation (5) as a limit of infinitely fast ( $\tau = 0$ ) signal transmission between the neighboring lattice points. While this limit is often a good approximation for heat conduction [8] it might not be appropriate for biological agents with finite response time. A general statement about which of the two continuum limits fits better the description of the VM is not possible. It can heavily depend on the quantity of interest and the chosen parameters. Nonetheless, from our analysis below, it follows that the information spreading in the VM is approximately diffusive for small speeds  $v_0$  and increasingly nondiffusive as  $v_0$  grows.

### B. Firm and lax leaders

We now consider the following two specific flooding scenarios for the static VM (4). (1) In the firm leader scenario, the leader's orientation is held fixed. Measuring the angular variables  $\theta_{ij} \ll 1$  in units of the initial orientation of the leader, we set  $\theta_{00}(t) = 1$  for all times. This amounts to a steady information influx into the system. (2) In the lax leader scenario, the orientation of the leader is set to 1 at time 0 but then evolves according to Eq. (4). In both scenarios, all other particles are initially aligned with the  $x$  axis,  $\theta_{ij}(0) = 0$  for  $ij \neq 00$ . While (1) can be interpreted as a flock following a leader, (2) might describe a flock reacting to a sudden perturbation.

In the firm-leader scenario, the dynamical equation (4) is most easily written and solved using the matrix form  $\boldsymbol{\theta}(t+1) = M_0 \boldsymbol{\theta}(t) + \boldsymbol{\theta}(0)$ , where the vector  $\boldsymbol{\theta}(t)$  contains the values of orientations at all grid points at time  $t$ ,  $\theta_{ij}(t)$ , and  $M_0$  incorporates the interactions. It has vanishing entries for the feedback onto the leader's orientation, which is set by  $\boldsymbol{\theta}(0)$ , which has vanishing entries for all other particles. The solution is  $\boldsymbol{\theta}(t) = \sum_{i=0}^t M_0^i \boldsymbol{\theta}(0)$ . In the lax leader scenario, the dynamical equation is  $\boldsymbol{\theta}(t+1) = M \boldsymbol{\theta}(t)$ , and  $M$  incorporates the interactions between all the grid points, as described by Eq. (4), including the feedback onto the leader. The solution is  $\boldsymbol{\theta}(t) = M^t \boldsymbol{\theta}(0)$ . Both solutions nicely demonstrate that due to the linearity of the dynamics, the transmission of the information obeys the principle of superposition: The impact onto  $\theta_{ij}(t)$  depends on the number of possible paths of length  $t$  the signal may take from  $(0,0)$  to  $(i,j)$ , namely, the summation induced by the matrix multiplication in  $M_0^t$ . And it decays with time and distance due to the conservation enforced by the repeated normalization via the prefactor  $(1/5)^t$  in  $M_0^t$ .

In Fig. 1 we depict the information spreading in the linearized lattice VM for both scenarios. As expected, the information spreading quickly becomes isotropic, since discretizing the diffusion equation on a square lattice destroys the radial symmetry only for short paths and affects only the initial stage of the dynamics. The spreading for the firm leader scenario, with a fixed source at the origin, eventually aligns all particles to the leader. The rate of this approach decreases with growing distance of the grid points from the leader, and the saturation curves exhibit maximum slopes at intermediate times.

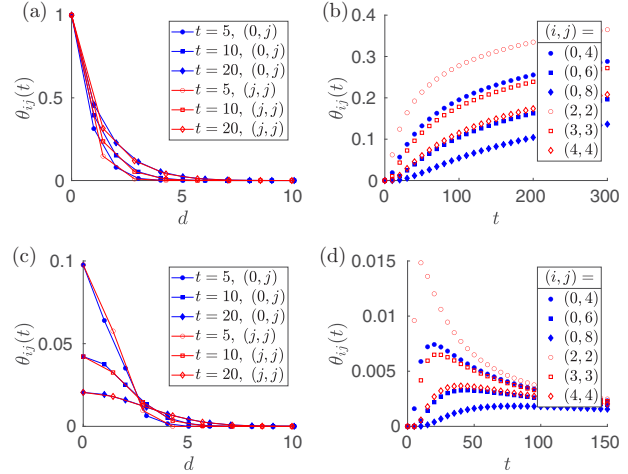


FIG. 1. Information spreading in the firm (a), (b) and lax (c), (d) leader scenarios. (a), (c) The spatial spreading of the orientation  $\theta_{ij}(t)$  at different times over the distance  $d = \sqrt{i^2 + j^2}$ , transverse (filled blue markers) and diagonal (open red markers) relative to the leader. (b), (d) The time evolution of the orientation for different grid points.

### C. Signal speed

In general, there is no unique definition of the speed of information spreading in the linearized lattice VM. The most obvious definition  $v = l/\tau = 1$  refers to the signal transfer between neighboring lattice points [cf. Eqs. (4)–(7)]. It provides the time  $d/v$  after which a grid point at distance  $d$  from the leader starts to receive the information. Yet it is of limited use because the strength of the received information is negligible if the grid point is far from the leader and there are only a few paths for the signal between the leader and the grid point. For example, in the case of a single path the signal strength received at time  $d/v$  is proportional to  $(1/5)^d$ .

A more informative definition is obtained from the time  $T_{\max}(d)$  when the change of orientation induced by the leader at distance  $d$  becomes maximal. The rate of change of orientation of the grid points is measured by the time derivatives  $\dot{\theta}_{ij}(t)$ , which exhibit a clear maximum [cf. Figs. 1(b) and 1(d)]. One may thus identify  $T_{\max}(d)$  with the time when  $\dot{\theta}_{ij}(t)$  with  $\sqrt{i^2 + j^2} = d$  is maximal. In Fig. 2(a) we show the resulting dispersion relation  $d(T_{\max})$  obtained from evaluating

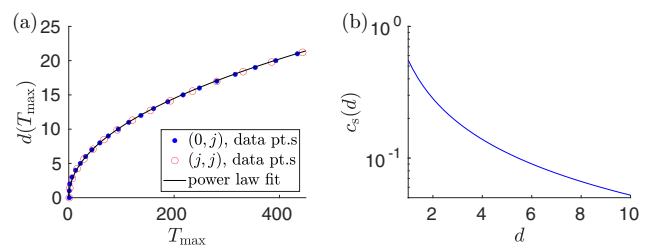


FIG. 2. (a) The dispersion relation for the linearized lattice VM transverse (blue filled circles) and diagonal (red open circles) relative to the leader. The solid line shows a fit  $d(t) = at^m$  with exponent  $m = 0.48$  and  $a \approx 1.1$ . (b) The corresponding signal speed  $c_s = \dot{d}$  as function of the distance  $d$  to the leader.

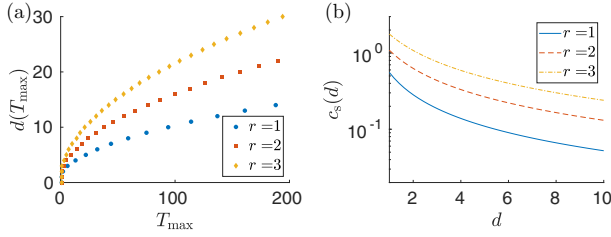


FIG. 3. (a) The dispersion relation for a generalized lattice VM where each grid point interacts with all neighbors at distances up to  $r$  lattice edges for three values of  $r$ . (b) The corresponding signal speeds  $c_s(d) = \dot{d}$ , respectively.

the signal propagation on the horizontal and on the diagonal axis with respect to the leader in the firm leader scenario. As expected, the found information spreading is well described by the diffusion relation  $d(T_{\max}) = \sqrt{4D_{\text{eff}}T_{\max}}$ . However, the diffusion coefficient  $D_{\text{eff}} = a^2/4 \approx 0.3$ , obtained by fitting the data, is much different than the diffusion coefficient  $D = 0.2$ , predicted from the limiting process leading to Eq. (5). In Fig. 2(b) we show the corresponding signal speed  $c_s = \dot{d}(t) \propto 1/\sqrt{t} \propto 1/d$ . The results obtained for the lax leader scenario are qualitatively the same (data not shown).

As an aside, we note that, while evaluating the evolution of the maxima of  $\dot{\theta}_i(t)$  is a reasonable approach for studying the signal spreading in the two flooding scenarios considered here, it is not suitable for more complex situations. A more universally applicable proxy for signal speed can be obtained by evaluating the connected acceleration correlations [10]. For our specific setting with a single leader and aligned initial state, the two approaches lead to the same results.

To close this section, we investigate the information spreading in a direct generalization of the linearized lattice VM (4), where the individual grid points interact not only with their nearest neighbors, but also with all grid points up to a distance of  $r$  lattice edges from the leader. Consequently, each grid point interacts with its  $N_{\text{int}} = 2r(r+1)$  nearest neighbors. The maximum speed of information propagation,  $v$ , is determined just by distances between particles at the circumference of the interaction zone, and thus it increases linearly with  $r$ . On the other hand, the  $r$  dependence of the speed  $c_s$ , shown in Fig. 3, is sublinear as the signal maximum is “slowed down” by the particles inside the interaction radius. Interestingly, the curves for different  $r$  values cannot be collapsed into a single master curve by multiplying each of them by a constant factor. Our analysis suggests that such a collapse is possible for long times only, with numerically obtained scaling factors 1.565 and 2.105 yielding the best asymptotic collapse of the curves for  $r = 1$  to those for  $r = 2$  and  $r = 3$ , respectively. These factors are close to the factors 1.64 ( $r = 1 \rightarrow 2$ ) and 2.28 ( $r = 1 \rightarrow 3$ ) obtained from the diffusion limit (5) of the individual lattice models as  $\sqrt{D_r/D_1}$  with  $D_r$  denoting the diffusion coefficient obtained for the individual values of  $r$ . Even though the diffusive scaling  $d = \sqrt{4D_r t}$ , predicted from Eq. (5), does not describe the data perfectly (in particular the prefactor  $4D_r$  is wrong), we take this as an indication that the formula  $d = \sqrt{4D_{\text{eff}}t}$  with  $D_{\text{eff}} \sim D_r$  is a

reasonable qualitative model for the spreading of information over long time and large length scales.

To sum up, in the linearized static spin model, the information spreads essentially diffusively. We next investigate how the situation changes when we allow particles to translate along their orientations.

## IV. THE MOTILE CASE

### A. The role of convection

Compared to equilibrium systems, active matter breaks certain local symmetries such as momentum and energy conservation. It is not *a priori* obvious whether this fundamental difference will lead to important effects on the information spreading and swarm behavior or if it is largely irrelevant, in practice. Differently from the lattice VM, the standard VM does not place the particles onto a regular lattice, and if so, this order would not be maintained for long. Notice that this breaks two important symmetries, namely, reciprocity and information conservation (see the Appendix). While the disorder itself does not affect the diffusive information spreading, the evolution of the neighborhood relations for  $v_0 > 0$  additionally allows for information convection. This situation is thus very similar to a moving heat source with the main difference that particles addressed by the leader tend to follow it, while heated passive particles generally do not induce a comparable flux.

Let us now derive a rough estimate for the particle speed  $v_0$  at which convection becomes important. The maximum conduction speed is given by the speed with which the signal spreads due to the interactions, i.e.,  $\ell_{\text{int}}/\Delta t$ . Here  $\Delta t = 1$  is the discrete update time in the VM and  $\ell_{\text{int}} = \sqrt{N_{\text{int}}/(\pi\rho)} = \sqrt{N_{\text{int}}/N}$  is the average interaction radius, assuming a more or less homogeneous density  $\rho = N/\pi$  after initiation inside a unit circle. The speed of convection is given by the relative speed of the individual particles on the order of  $v_0$ . Conduction and convection should thus compete when  $v_0 \approx \sqrt{N_{\text{int}}/N}$ . Alternatively, as in Sec. III C, we could measure the speed of signal propagation by the ratio  $d_i/T_i$ , where  $T_i$  is the time when the signal sent at time 0 causes a maximum change  $\dot{\theta}_i(t)$  of orientation at distance  $d_i$ , i.e.,  $\dot{\theta}_i(T_i) \equiv \max_t \dot{\theta}_i(t)$ . As we find below, the latter approach, which predicts a significantly lower conduction speed, is more appropriate to describe the data, yielding a correspondingly lower threshold velocity for the onset of convective transport. (For our choice of parameters, convection plays role already for velocities of about  $v_0 = 0.01$  while  $\sqrt{N_{\text{int}}/N} \approx 0.16$ .)

Besides inducing convection, motility further complicates the definition of a signal speed. Due to the relative motions of the particles there is no *a priori* choice of the distance  $d_i$  traveled by a signal. For this reason, we analyzed the speed of information propagation using two different definitions of  $d_i$ . First, the (average) initial distance  $|\mathbf{r}_i(0) - \mathbf{r}_L(0)|$  between the particles and the leader, which is the initial position of particle  $i$  at time 0, also encoded in the initial density  $\rho$ . Second, the distance  $|\mathbf{r}_i(T_i) - 1/N \sum_j \mathbf{r}_j(T_i)|$  between the particle  $i$  and the position of the center (of mass) of the flock at the characteristic “interaction time”  $T_i$  for conductive transport. We have performed the analysis below for both these definitions of the

distance and found no qualitative differences. Therefore, we show only the results obtained for the former, in the following.

### B. Numerical procedure

In the simulations, we place the leader always into the center of a unit circle. Positions of all  $N = 1000$  other particles are picked randomly inside the circle. All particles interact with their  $N_{\text{int}} = 24$  nearest neighbors, corresponding to  $r = 3$  in Fig. 3. Small density fluctuations in the initial condition are found to induce strong noise in the measured functions  $\theta_i(t)$  and  $\dot{\theta}_i(t)$ . To be able to determine the overall trend from these measurements, we averaged the resulting curves over  $N_{\text{runs}}$  runs with different initial conditions. We also employed two different smoothening procedures:

The average  $\langle \cdot \rangle_{nm}$  is calculated as follows. First, we collect the data  $\{d_i, \theta_i(t), \dot{\theta}_i(t)\}_{i=1, \dots, N}$  from  $N_{\text{runs}} = 100$  runs of the simulation. Then we sort the data according to the distance  $d_i$  to the origin at time 0. Finally, we calculate the smoothed variables  $\langle d_i \rangle_{nm}$ ,  $\langle \theta_i(t) \rangle_{nm}$ , and  $\langle \dot{\theta}_i(t) \rangle_{nm}$  by averaging  $d_i$ ,  $\theta_i(t)$ , and  $\dot{\theta}_i(t)$  over  $N_{\text{av}}$  neighbors of the particle  $i$ , i.e., over the particles  $j$  with  $N_{\text{av}}$  smallest distances  $|d_i - d_j|$ . Since the dispersion relation is a strictly monotonous function of time, one may alternatively perform the averaging with respect to nearest neighbors in time  $T_i$ , according to when the maximum signal has arrived at particle  $i$ . In other words one can average over the particles  $j$  with  $N_{\text{av}}$  smallest distances  $|T_i - T_j|$ . We have tested that both averaging procedures lead to qualitatively the same results. In the following, we show only those obtained using the averaging  $\langle \cdot \rangle_{nm}$  over  $N_{\text{av}}$  spatially nearest neighbors.

### C. Firm leader scenario

We now consider the firm leader scenario of Sec. III B, where the leader's orientation is fixed to  $\varphi$  at all times and all other particles are initially aligned with the perpendicular  $x$  axis and subsequently obey the dynamical equations (1) and (2). Note that this condition implies that reciprocity between the leader and the flock is maximally broken.

In Figs. 4(a) and 4(b) we show the resulting averaged orientations,  $\langle \theta_i(t) \rangle_{nm}$ , and the averaged changes in the orientation,  $\langle \dot{\theta}_i(t) \rangle_{nm}$ , as functions of the averaged distance  $\langle d_i \rangle_{nm}$ . To investigate the directional dependence of the information spreading, we distinguish between two directions of signal propagation. As the leader's orientation points into the positive half-plane, we identify the particles with positive  $y$  coordinates at time 0 as lying in the ‘‘positive direction’’ with respect to the leader. The remaining particles are lying in the ‘‘negative direction.’’ The results for  $\langle \theta_i(t) \rangle_{nm}$  and  $\langle \dot{\theta}_i(t) \rangle_{nm}$  for the positive and negative directions are given in Figs. 4(c) and 4(d) and Figs. 4(e) and 4(f), respectively. As the leader carries the source of information with it, particles lying in the positive direction show a significantly larger change of orientation than those in the negative direction. Furthermore, the leader affects nearby particles more than more distant ones. This leads to correspondingly stronger average direction changes  $\langle \dot{\theta}_i(t) \rangle_{nm}$  in its vicinity. Consequently, upon traversing the flock, the leader seems to drag around a cloud of ‘‘followers.’’ However, since the interaction rule allows only imperfect alignments, particles begin to realign with the less informed surroundings

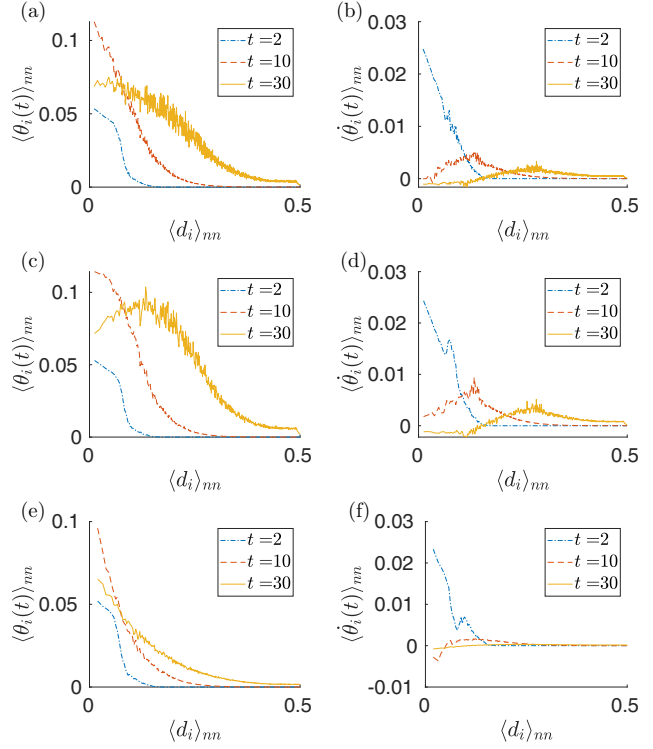


FIG. 4. Firm leader scenario. The average orientation,  $\langle \theta_i(t) \rangle_{nm}$  and change in orientation  $\langle \dot{\theta}_i(t) \rangle_{nm}$ , as functions of the average distance  $\langle d_i \rangle_{nm}$  at three different times: For the whole swarm (a), (b), along the leader direction (c), (d), and along the negative leader direction (e), (f). Parameters used:  $N = 1000$ ,  $v_0 = 0.01$ ,  $N_{\text{int}} = 24$ ,  $\varphi = \pi/4$ ,  $N_{\text{runs}} = 100$ ,  $N_{\text{av}} = N_{\text{runs}}$ .

after the leader has left their neighborhood. This is depicted by the moving maxima of  $\langle \theta_i(t) \rangle_{nm}$  and  $\langle \dot{\theta}_i(t) \rangle_{nm}$  in Figs. 4(c) and 4(d). In the negative direction, where the information propagates by pure conduction, no such structure is visible. The response of the swarm as a whole is dominated by the dynamics in positive direction. Repeating the described analysis for  $v_0 = 0$ , we found the same behavior as for the linearized lattice VM in Sec. III.

In Fig. 5(a) we show the time evolution of the change of orientation  $\langle \dot{\theta}_i(t) \rangle_{N, N_{\text{runs}}}$  averaged over all particles in the chosen particle set (total system, positive direction, and negative

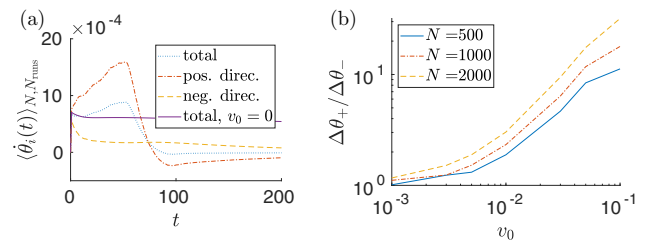


FIG. 5. Firm leader scenario. (a) Time evolutions of the change of direction averaged over the total system and the positive and negative directions for  $v_0 = 0.01$  and  $v_0 = 0$ , respectively. (b) The corresponding ratio (8) as function of the particle speed. Other parameters as in Fig. 4.

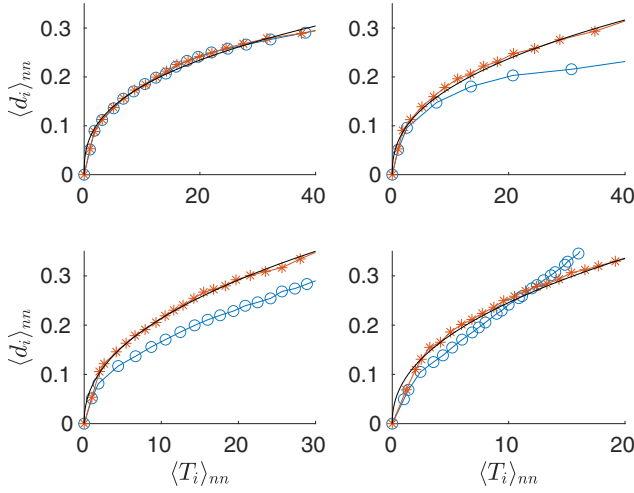


FIG. 6. Dispersion relation for the firm-leader scenario for positive (blue circles) and negative (red stars) directions and speeds  $v_0 = 0, 0.001, 0.01, 0.03$  increasing from the upper left to the bottom right panel. The black solid lines are fits of  $a(T_i)_{nm}^m$  to the data for the positive direction with  $m = 0.376, 0.417, 0.449, 0.451$  corresponding to the individual speeds. The slope of the data for speed  $v_0 = 0.03$  for the positive direction is  $\approx 0.018$  while the corresponding speed of the leader projected to the positive direction is  $v_0 \sin \varphi \approx 0.021$ . Parameters as in Fig. 4, except for  $N_{Av} = 10N_{runs}$ .

direction) and all simulations for  $v_0 > 0$  and  $v_0 = 0$ . For  $v_0 > 0$ , the average signal strength for the total system and particularly in the positive direction continuously increases until the leader approaches the edge of the system. This can be understood as follows. The change of orientation of the individual particles is largest when their orientation differs most from the average orientation of their neighbors. A moving leader constantly meets on its way in the positive direction particles (almost) aligned with the  $x$  axis, leading to a steady increase of the corresponding signal. Subsequently, at times  $t \gtrsim 50$  when the leader has left the flock,  $\langle \dot{\theta}_i(t) \rangle_{N, N_{runs}}$  rapidly decreases, and eventually it also changes sign. Since the leader mainly affects nearby particles, more distant particles are much less aligned with its orientation when it leaves. Therefore, particles that aligned with the leader during its passage through the flock begin to realign with the less affected particles after the leader has left. In the negative direction, the signal strength monotonously decreases, similarly as for  $v_0 = 0$  and the linearized lattice VM.

To quantify the asymmetry between the positive and negative direction, we integrate the positive areas

$$\Delta\theta_{\pm} = \int_{t_0}^{\infty} \langle \dot{\theta}_i(t) \rangle_{N, N_{runs}}^{\pm} \Theta(\langle \dot{\theta}_i(t) \rangle_{N, N_{runs}}^{\pm}), \quad (8)$$

beneath the corresponding curves in Fig. 5(a). Here + (−) corresponds to the positive (negative) direction and  $\Theta(\cdot)$  denotes the unit step function. The ratio  $\Delta\theta_{+}/\Delta\theta_{-}$  is shown in Fig. 5(b). As expected, it monotonously increases with the particle speed  $v_0$  and particle density  $N/\pi$ .

The main result of this section are the dispersion relations for four different velocities shown in Fig. 6. Regardless of  $v_0$ , the information initially spreads conductively, hence similarly

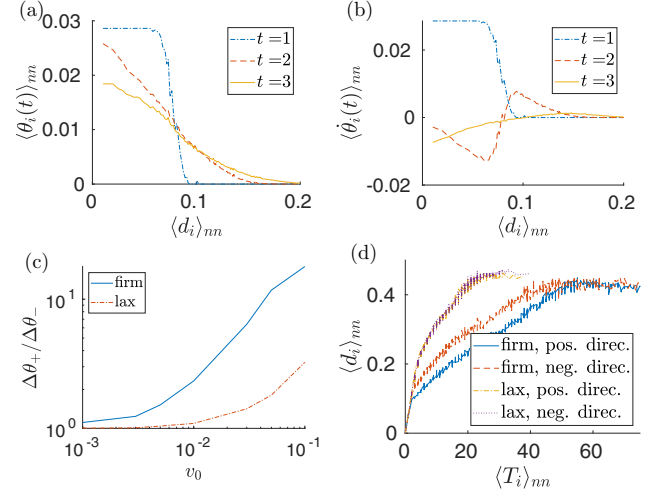


FIG. 7. Lax leader scenario. (a), (b) The orientation,  $\langle \theta_i(t) \rangle_{nm}$ , and the averaged change in the orientation,  $\langle \dot{\theta}_i(t) \rangle_{nm}$ , as functions of the averaged distance  $\langle d_i \rangle_{nm}$  for three different times averaged over the overall system. (c) The ratio (8) of the integrated positive changes in the orientation in the positive and negative directions as functions of the particle speed for the firm and lax leader scenario. (d) The dispersion relation for the positive and negative directions for the firm and lax leader scenario. We used the same parameters as in Fig. 4.

as in the linearized lattice VM, as diffusion beats convection over short times and distances. With increasing velocity  $v_0 > 0$ , the spreading in the positive direction becomes gradually more convective at late times. The slope of the dispersion relation converges to the velocity of the leader projected to the positive direction,  $v_0 \sin \varphi$ . In the negative direction, the spreading stays conductive regardless of  $v_0$ . Even though the particles in the negative direction are less affected by the turning event induced by the leader, the dispersion relation shows that the information reaches them faster than those in the positive direction. This counterintuitive effect is a consequence of the employed definition of  $T_i$ : The dispersion relation follows from determining times maximizing  $\dot{\theta}_i(t)$ . As the leader moves away from the particles behind it, the rates of change  $\dot{\theta}_i(t)$  of their orientations peak sooner than those in the positive direction. This somewhat counterintuitive behavior is reminiscent of observations of faster speeds for smaller pulses [8] or propagation of second sound against the heat flow [53].

#### D. Lax leader scenario

The information spreading is somewhat different in the lax leader scenario. For the same initial condition as in the preceding section, the leader now adapts dynamically according to Eqs. (1) and (2) to its neighbors for  $t > 0$ . It thereby virtually loses the information passed on to them. The interaction with the neighbors is thus more reciprocal than for the unwavering firm leader, yet not entirely so, since the topological notion of next neighbors is not necessarily fully reciprocal (see the Appendix).

In Figs. 7(a) and 7(b) we depict again the average orientation  $\langle \theta_i(t) \rangle_{nm}$  and the averaged change in the orientation

$\langle \dot{\theta}_i(t) \rangle_{mn}$  for the lax leader scenario. The parameters are the same as for the firm leader scenario in Fig. 4. Comparing the results for the two scenarios, we find the following differences: (1) the amplitudes of both  $\langle \theta_i(t) \rangle_{mn}$  and  $\langle \dot{\theta}_i(t) \rangle_{mn}$  are much smaller and the time derivative in the averaged angle converges to zero much faster, since the leader realigns with the rest of the flock in the lax leader scenario. (2) The time derivative  $\langle \dot{\theta}_i(t) \rangle_{mn}$  exhibits an excursion to negative values at small distances to the leader, due to the feedback from the flock, which requires a realignment of the leader and its neighborhood with the “winning majority” of other particles in the flock. At larger distances, the derivative returns to positive values, as expected for a moderate realignment of the merely slightly disturbed more distant particles. (3) There is again a directional dependence of the response, as in the firm-leader scenario. However, it is now much weaker, due to the mutual information exchange.

In fact, for the parameters used in the figure, it is not worthwhile to show the corresponding spatial distributions, as they would be hardly discernible from those for the total system in Fig. 7. The directional dependence of the information spreading in the lax leader scenario becomes noteworthy only for substantially larger speeds  $v_0$ , as demonstrated in Fig. 7(c). There we compare the ratio  $\Delta\theta_+/\Delta\theta_-$  of responses (8) integrated in the positive and negative directions between the lax and firm leader scenarios. Because of weaker total signal strength in the lax leader scenario, the changes of orientation  $\dot{\theta}_i(t)$  of the individual particles peak sooner, leading to steeper (but directionally barely distinguishable) dispersion relations; cf. Fig. 7(d).

## V. CONCLUSION

We studied transport of information about orientation of a leader in the Vicsek model (VM) with topological interactions. The two main mechanisms for propagation of information are conduction and convection. We have shown that the conductive aspect in the VM can well be understood using a simplified, exactly solvable variant of the model, where the individual particles are fixed at grid points of a regular lattice. This static spin lattice model allows for an analogy with heat transfer, which ceases to hold in the full dynamic VM. Nonlinearity and heterogeneity then break the underlying symmetries such as reciprocity and spin conservation. Yet this has no major practical consequences by itself. The visible changes between the dynamic model and the spin lattice are entirely dominated by the convective dynamics.

We considered two scenarios of information spreading from a single misaligned leader particle. While the diffusive or conductive spreading prevails over short times and distances, the spreading over longer times and distances gradually acquires a convective contribution, as the particle speed increases. We quantified this intuitive conclusion by measuring the dispersion relation. It was formulated for the timescale at which the signal induces its largest change in orientation of the particles at a given distance. The analysis revealed a strong directional dependence of the information transfer for the firm-leader scenario, in which the reciprocity of the mutual information exchange is maximally violated. A significant effect of convective information spreading is observed only

in the direction of the leader motion. In the wake zone behind the leader, the spreading remains diffusive, regardless of the speed.

While measuring the dispersion relation for zero speed of the particles is a relatively straightforward task, the definition of the distance over which the signal has propagated becomes ambiguous for the motile swarm. Nevertheless, we found that different length definitions lead to qualitatively close results.

Besides this ambiguity in the definition of the dispersion relation, which might deserve further analysis, our findings raise several questions. First, while some preliminary runs seemed to confirm the expectation that the inclusion of noise in the VM would yield qualitatively similar results, one could wish to study this issue more extensively, in particular with regard to the stability of the flocking transition; i.e., under which conditions can a leader move induce an ordering transition or the breakup of a flock? Further, many natural interaction networks are more heterogeneous than our flocks, containing, e.g., certain hierarchical structures [54,55] or distance- and density-weighted interactions [24]. Moreover, it might be interesting to consider a more realistic modification of the standard VM where the orientation of a particle under consideration would have a stronger weighting than the average orientation of its neighbors. This would yield a more persistent motion and might impact the information propagation. We took first steps in this direction in a follow-up study to the present work [56]. There we investigate information spreading in a 2D VM with time-delayed metric interactions [57] and also address the pertinence of the notion of linear response [58–60] and its relation to information propagation. Next, it might be interesting to connect information spreading in active matter with corresponding results in other research fields, such as network theory or epidemiology. Especially in the latter, the effects of network heterogeneity on the spreading of diseases is a widely studied aspect [6,61–63]. Finally, it would seem interesting to pursue the question how the interaction rules in the VM can be optimized to facilitate information transfer.

## ACKNOWLEDGMENTS

We acknowledge funding through a DFG-GACR cooperation by the Deutsche Forschungsgemeinschaft (DFG-Code KR 3381/6-1) and by the Czech Science Foundation (GACR Project No. 20-02955J). V.H. was supported by the Humboldt Foundation. D.G. acknowledges funding by International Max Planck Research Schools (IMPRS) as well as by the Deutscher Akademischer Austauschdienst (DAAD).

## APPENDIX: BREAKING OF INFORMATION CONSERVATION IN THE LINEARIZED VM

As discussed in Sec. III, the interaction rule in the linearized lattice VM,

$$\theta_{ij}(t+1) = \frac{1}{N_{ij}(t)} \sum_{(ij),(kl)} \theta_{kl}(t), \quad (\text{A1})$$

where the sum goes over all neighbors  $(kl)$  of  $(ij)$  including  $(ij)$  itself, and  $N_{ij} = \sum_{(ij),(kl)}$  denotes the number of

neighbors, conserves the total information content

$$\theta_{\text{tot}}(t) \equiv \sum_{ij} \theta_{ij}(t). \quad (\text{A2})$$

While the linearized Vicsek interaction rule yields reciprocal interactions if the interaction network is regular, it can render nonreciprocal interparticle interactions for irregular interaction networks, e.g., if the particle density of the system is inhomogeneous in space. The conservation condition  $\dot{\theta}_{\text{tot}}(t) = 0$  will then be broken. For a closed system, the reverse also holds: If the conservation is broken, this indicates the presence of some nonreciprocal interactions. As an illustrative example, consider the following closed interaction network consisting of three particles. Particle 1 and 3 interact solely with particle 2, which interacts with both 1 and 3. Assuming the initial condition  $\theta_1(0) = 1$  and  $\theta_{2,3}(0) = 0$ , then  $\theta_1(1) = 1/2$ ,  $\theta_2(1) = 1/3$ ,  $\theta_3(1) = 0$ . We thus see that  $\theta_{\text{tot}}(0) = 1 > \theta_{\text{tot}}(1) = 5/6$ . If we instead consider the initial

condition  $\theta_2(0) = 1$  and  $\theta_{1,3}(0) = 0$ , we find  $\theta_{\text{tot}}(0) = 1 < \theta_{\text{tot}}(1) = 3/2$ . These examples manifest a more general finding that  $\theta_{\text{tot}}(t)$  decreases when the information flows from a less to a more connected region, and vice versa. In this case, reciprocity is broken since the normalization  $N_{ij}$  of neighboring particles varies. Beyond the linear regime, the situation is more complicated as the normalization  $N_{ij}$  depends on the angular variables.

Similarly, also for topological interactions, the information content is not conserved. While each particle interacts with exactly the same number of neighbors (i.e.,  $N_{ij} = \text{const}$ ), density gradients may induce unilateral interactions. As an example, consider a closed system of four particles with topological interactions with two nearest neighbors. Let the particles 1, 2, and 3 reciprocally communicate with each other, while the distant particle 4 adjusts its direction to that of particles 2 and 3 without influencing them. Assuming the initial condition  $\theta_2(0) = 1$  and  $\theta_{1,3,4}(0) = 0$ , we find  $\theta_{1,\dots,4}(1) = 1/3$  and thus  $\theta_{\text{tot}}(0) = 1 < \theta_{\text{tot}}(1) = 4/3$ .

- 
- [1] K. Lerman and R. Ghosh, Information contagion: An empirical study of the spread of news on Digg and Twitter social networks, in *Proceedings of the 4th International Conference on Weblogs and Social Media (ICWSM)* (The AAAI Press, Menlo Park, CA, 2010), p. 90.
- [2] Y. Moreno, M. Nekovee, and A. F. Pacheco, Dynamics of rumor spreading in complex networks, *Phys. Rev. E* **69**, 066130 (2004).
- [3] R. Pastor-Satorras and A. Vespignani, Epidemic Spreading in Scale-Free Networks, *Phys. Rev. Lett.* **86**, 3200 (2001).
- [4] M. Boguñá and R. Pastor-Satorras, Epidemic spreading in correlated complex networks, *Phys. Rev. E* **66**, 047104 (2002).
- [5] H. W. Hethcote, The mathematics of infectious diseases, *SIAM Rev.* **42**, 599 (2000).
- [6] D. Levis, A. Diaz-Guilera, I. Pagonabarraga, and M. Starnini, Flocking-enhanced social contagion, *Phys. Rev. Research* **2**, 032056(R) (2020).
- [7] M. N. Özışik, *Heat Conduction* (John Wiley & Sons, New York, 1993).
- [8] D. D. Joseph and L. Preziosi, Heat waves, *Rev. Mod. Phys.* **61**, 41 (1989).
- [9] R. S. Brodkey and H. C. Hershey, *Transport Phenomena: A Unified Approach* (Brodkey Publishing, Columbus, 2003).
- [10] A. Cavagna, I. Giardina, and T. S. Grigera, The physics of flocking: Correlation as a compass from experiments to theory, *Phys. Rep.* **728**, 1 (2018).
- [11] A. Procaccini, A. Orlandi, A. Cavagna, I. Giardina, F. Zoratto, D. Santucci, F. Chiarotti, C. K. Hemelrijk, E. Alleva, G. Parisi, *et al.*, Propagating waves in starling, *Sturnus vulgaris*, flocks under predation, *Anim. Behav.* **82**, 759 (2011).
- [12] H.-P. Zhang, A. Beãžer, E.-L. Florin, and H. L. Swinney, Collective motion and density fluctuations in bacterial colonies, *Proc. Natl. Acad. Sci. USA* **107**, 13626 (2010).
- [13] E. Ben-Jacob, O. Schochet, A. Tenenbaum, I. Cohen, A. Czirok, and T. Vicsek, Generic modelling of cooperative growth patterns in bacterial colonies, *Nature (London)* **368**, 46 (1994).
- [14] G. Gompper, R. G. Winkler, T. Speck, A. Solon, C. Nardini, F. Peruani, H. Löwen, R. Golestanian, U. B. Kaupp, L. Alvarez *et al.*, The 2020 motile active matter roadmap, *J. Phys.: Condens. Matter* **32**, 193001 (2020).
- [15] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe, Active particles in complex and crowded environments, *Rev. Mod. Phys.* **88**, 045006 (2016).
- [16] S. Ramaswamy, The mechanics and statistics of active matter, *Annu. Rev. Condens. Matter Phys.* **1**, 323 (2010).
- [17] M. E. Cates and J. Tailleur, Motility-induced phase separation, *Annu. Rev. Condens. Matter Phys.* **6**, 219 (2015).
- [18] M. F. Hagan and A. Baskaran, Emergent self-organization in active materials, *Curr. Opin. Cell Biol.* **38**, 74 (2016).
- [19] T. Bäuerle, A. Fischer, T. Speck, and C. Bechinger, Self-organization of active particles by quorum sensing rules, *Nat. Commun.* **9**, 1 (2018).
- [20] F. D. C. Farrell, M. C. Marchetti, D. Marenduzzo, and J. Tailleur, Pattern Formation in Self-Propelled Particles with Density-Dependent Motility, *Phys. Rev. Lett.* **108**, 248101 (2012).
- [21] A. P. Solon, J.-B. Caussin, D. Bartolo, H. Chaté, and J. Tailleur, Pattern formation in flocking models: A hydrodynamic description, *Phys. Rev. E* **92**, 062111 (2015).
- [22] M. Brambilla, E. Ferrante, M. Birattari, and M. Dorigo, Swarm robotics: A review from the swarm engineering perspective, *Swarm Intelligence* **7**, 1 (2013).
- [23] F. J. Vernerey, E. Benet, L. Blue, A. Fajrial, S. L. Sridhar, J. Lum, G. Shakya, K. Song, A. Thomas, and M. Borden, Biological active matter aggregates: Inspiration for smart colloidal materials, *Adv. Colloid Interface Sci.* **263**, 38 (2019).
- [24] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, Novel Type of Phase Transition in a System of Self-Driven Particles, *Phys. Rev. Lett.* **75**, 1226 (1995).
- [25] F. Ginelli, The physics of the Vicsek model, *Eur. Phys. J.: Spec. Top.* **225**, 2099 (2016).

- [26] T. Vicsek and A. Zafeiris, Collective motion, *Phys. Rep.* **517**, 71 (2012).
- [27] F. Kuhn and R. Oshman, Dynamic networks: Models and algorithms, *ACM SIGACT News* **42**, 82 (2011).
- [28] A. Clementi, R. Silvestri, and L. Trevisan, Information spreading in dynamic graphs, *Distrib. Comput.* **28**, 55 (2015).
- [29] A. E. Clementi, F. Pasquale, A. Monti, and R. Silvestri, Information spreading in stationary Markovian evolving graphs, in *2009 IEEE International Symposium on Parallel & Distributed Processing* (IEEE Computer Society, Los Alamitos, CA, 2009), pp. 1–12.
- [30] R. Van Der Hofstad, G. Hooghiemstra, and P. Van Mieghem, The flooding time in random graphs, *Extremes* **5**, 111 (2002).
- [31] S. Melnik, H. Garcia-Molina, and E. Rahm, Similarity flooding: A versatile graph matching algorithm and its application to schema matching, in *Proceedings 18th International Conference on Data Engineering* (IEEE Computer Society, Los Alamitos, CA, 2002), pp. 117–128.
- [32] H. Chaté, F. Ginelli, G. Grégoire, F. Peruani, and F. Raynaud, Modeling collective motion: Variations on the Vicsek model, *Eur. Phys. J. B* **64**, 451 (2008).
- [33] P. Degond, G. Dimarco, and T. B. N. Mac, Hydrodynamics of the Kuramoto–Vicsek model of rotating self-propelled particles, *Math. Models Methods Appl. Sci.* **24**, 277 (2014).
- [34] A. Attanasi, A. Cavagna, L. Del Castello, I. Giardina, T. S. Grigera, A. Jelić, S. Melillo, L. Parisi, O. Pohl, E. Shen *et al.*, Information transfer and behavioural inertia in starling flocks, *Nat. Phys.* **10**, 691 (2014).
- [35] A. Cavagna, L. Del Castello, I. Giardina, T. Grigera, A. Jelic, S. Melillo, T. Mora, L. Parisi, E. Silvestri, M. Viale *et al.*, Flocking and turning: A new model for self-organized collective motion, *J. Stat. Phys.* **158**, 601 (2015).
- [36] A. Cavagna, I. Giardina, T. S. Grigera, A. Jelic, D. Levine, S. Ramaswamy, and M. Viale, Silent Flocks: Constraints on Signal Propagation across Biological Groups, *Phys. Rev. Lett.* **114**, 218101 (2015).
- [37] J. K. Parrish and W. M. Hamner, *Animal Groups in Three Dimensions: How Species Aggregate* (Cambridge University Press, Cambridge, 1997).
- [38] J. E. Treherne and W. A. Foster, Group transmission of predator avoidance behaviour in a marine insect: The Trafalgar effect, *Anim. Behav.* **29**, 911 (1981).
- [39] P. De Lellis and M. Porfiri, Inferring the size of a collective of self-propelled Vicsek particles from the random motion of a single unit, *Commun. Phys.* **5**, 86 (2022).
- [40] G. Joyce, Classical Heisenberg model, *Phys. Rev.* **155**, 478 (1967).
- [41] J. Kosterlitz, The critical properties of the two-dimensional XY model, *J. Phys. C* **7**, 1046 (1974).
- [42] M. Sentef, M. Kollar, and A. P. Kampf, Spin transport in Heisenberg antiferromagnets in two and three dimensions, *Phys. Rev. B* **75**, 214403 (2007).
- [43] L. Lima and A. Pires, Dynamics of the anisotropic two-dimensional XY model, *Eur. Phys. J. B* **70**, 335 (2009).
- [44] E. Sonin, Spin currents and spin superfluidity, *Adv. Phys.* **59**, 181 (2010).
- [45] L. Page, S. Brin, R. Motwani, and T. Winograd, The PageRank citation ranking: Bringing order to the web, Technical report, Stanford InfoLab (1999).
- [46] K. Avrachenkov and N. Litvak, The effect of new links on Google PageRank, *Stochastic Models* **22**, 319 (2006).
- [47] S. L. Sobolev, Transport processes and traveling waves in systems with local nonequilibrium, *Sov. Phys. Usp.* **34**, 217 (1991).
- [48] S. L. Sobolev, Discrete space-time model for heat conduction: Application to size-dependent thermal conductivity in nano-films, *Int. J. Heat Mass Transf.* **108**, 933 (2017).
- [49] G. Lebon, D. Jou, and J. Casas-Vázquez, *Understanding Non-equilibrium Thermodynamics* (Springer, Berlin, Heidelberg, 2008), Vol. 295.
- [50] L. Landau, Theory of the superfluidity of helium II, *Phys. Rev.* **60**, 356 (1941).
- [51] S. Goldstein, On diffusion by discontinuous movements, and on the telegraph equation, *Q. J. Mech. Appl. Math.* **4**, 129 (1951).
- [52] C. Cattaneo, Sur une forme de l'équation de la chaleur éliminant la paradoxie d'une propagation instantanée, *Comptes Rendus de l'Académie des Sci.* **247**, 431 (1958).
- [53] B. D. Coleman and D. R. Owen, On the nonequilibrium behavior of solids that transport heat by second sound, *Comput. Math. Appl.* **9**, 527 (1983).
- [54] I. D. Chase, Models of hierarchy formation in animal societies, *Behav. Sci.* **19**, 374 (1974).
- [55] M. Nagy, Z. Ákos, D. Biro, and T. Vicsek, Hierarchical group dynamics in pigeon flocks, *Nature (London)* **464**, 890 (2010).
- [56] D. Geiß, K. Kroy, and V. Holubec, Signal propagation and linear response in the delay Vicsek model, [arXiv:2205.12069](https://arxiv.org/abs/2205.12069) (2022).
- [57] V. Holubec, D. Geiss, S. A. M. Loos, K. Kroy, and F. Cichos, Finite-Size Scaling at the Edge of Disorder in a Time-Delay Vicsek Model, *Phys. Rev. Lett.* **127**, 258001 (2021).
- [58] A. Czirók, H. E. Stanley, and T. Vicsek, Spontaneously ordered motion of self-propelled particles, *J. Phys. A: Math. Gen.* **30**, 1375 (1997).
- [59] H. Chaté, F. Ginelli, G. Grégoire, and F. Raynaud, Collective motion of self-propelled particles interacting without cohesion, *Phys. Rev. E* **77**, 046113 (2008).
- [60] D. J. G. Pearce and L. Giomi, Linear response to leadership, effective temperature, and decision making in flocks, *Phys. Rev. E* **94**, 022612 (2016).
- [61] R. Olinky and L. Stone, Unexpected epidemic thresholds in heterogeneous networks: The role of disease transmission, *Phys. Rev. E* **70**, 030902(R) (2004).
- [62] L. A. Meyers, B. Pourbohloul, M. E. Newman, D. M. Skowronski, and R. C. Brunham, Network theory and SARS: Predicting outbreak diversity, *J. Theor. Biol.* **232**, 71 (2005).
- [63] G. Grossmann, M. Backenkoehler, and V. Wolf, Why ODE models for COVID-19 fail: Heterogeneity shapes epidemic dynamics, [medRxiv](https://medrxiv.org/abs/2021.03.25.21254292) (2021), doi:10.1101/2021.03.25.21254292.

