



## Optimization of nonequilibrium free energy harvesting illustrated on bacteriorhodopsin

Jordi Piñero <sup>1</sup>, Ricard Solé,<sup>1,2,3</sup> and Artemy Kolchinsky <sup>1,4,\*</sup>

<sup>1</sup>ICREA-Complex Systems Lab, Universitat Pompeu Fabra, 08003 Barcelona, Spain

<sup>2</sup>Institut de Biologia Evolutiva (CSIC-UPF), 08003 Barcelona, Spain

<sup>3</sup>Santa Fe Institute, Santa Fe, New Mexico 87501, United States

<sup>4</sup>Universal Biology Institute, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan



(Received 9 March 2023; revised 20 December 2023; accepted 12 February 2024; published 12 March 2024)

Harvesting free energy from the environment is essential for the operation of many biological and artificial systems. We use techniques from stochastic thermodynamics to investigate the maximum rate of harvesting achievable by optimizing a set of reactions in a Markovian system, possibly under various kinds of topological, kinetic, and thermodynamic constraints. This question is relevant for the optimal design of new harvesting devices as well as for quantifying the efficiency of existing systems. We first demonstrate that the maximum harvesting rate can be expressed as a constrained convex optimization problem. We illustrate it on bacteriorhodopsin, a light-driven proton pump from Archaea, which we find is close to optimal under realistic conditions. In our second result, we solve the optimization problem in closed-form in three physically meaningful limiting regimes. These closed-form solutions are illustrated on two idealized models of unicyclic harvesting systems.

DOI: [10.1103/PhysRevResearch.6.013275](https://doi.org/10.1103/PhysRevResearch.6.013275)

### I. INTRODUCTION

Many molecular systems, both biological and artificial, harvest free energy from their environments. Biological organisms require free energy to grow and replicate [1,2], and many species undergo selection for increased harvesting [3–6]. Artificial harvesting systems have also been constructed and optimized in the field of synthetic biology [7–14]. The optimization of free energy harvesting is thus a central problem both in biology and engineering.

As an example, consider a harvesting system such as a biological metabolic network that converts glucose to adenosine triphosphate (ATP) [15]. Suppose that the kinetic and thermodynamic parameters of one or more reactions can be optimized, either by natural selection or artificial design. What is the maximum rate of free energy harvesting that the network can achieve, and what are the kinetic and thermodynamic parameters that achieve it? These questions are relevant both for design of optimal harvesting devices and for quantifying the efficiency of existing systems.

In this paper, we use techniques from stochastic thermodynamics to derive bounds on maximum rate of free energy harvesting. We consider a harvesting system in nonequilibrium steady state which is coupled to an external source of free energy, an internal free energy reservoir, and a heat bath. The setup is well-suited for studying the kinds of small-scale

systems usually considered in stochastic thermodynamics [16], where assumptions of local detailed balance and steady state are justified. The steady-state assumption is reasonable in many molecular systems, where there is a separation of time scales between internal relaxation and environmental change.

We suppose that the system's dynamics can be separated into two kinds of processes, termed *baseline* and *control*. The baseline processes, which are held fixed, mediate the coupling to the external source of free energy. Control refers to all other processes which can be optimized for increasing the *harvesting rate* at which free energy flows into the internal reservoir. The particular separation of baseline/control generally depends on domain knowledge about the system and the scientific question at hand. For example, to study the efficiency of a particular reaction in a metabolic network, that reaction may be treated as control while the other reactions are baseline. The baseline/control separation is similar to the distinction in control theory between “plant” (a given system with fixed dynamical laws) and “controller” (the part that undergoes optimization) [17].

In our first set of results, we demonstrate that the optimization of the harvesting rate can be expressed as a convex optimization problem. The optimal solution of this problem determines both the maximum harvesting rate and the specific control processes that achieve that maximum. Importantly, the optimization can also account for various types of constraints on the possible network topology, kinetic timescales, and thermodynamic forces of the control processes.

We illustrate our results on bacteriorhodopsin (Fig. 1), a proton-pumping membrane protein. Bacteriorhodopsin is a photosynthetic system found in Archaea, with close relatives in bacteria [20,21]. It is also used in many artificial light-harvesting systems [7–9,11]. Using published thermodynamic and kinetic data, we develop a

\*artemyk@gmail.com

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

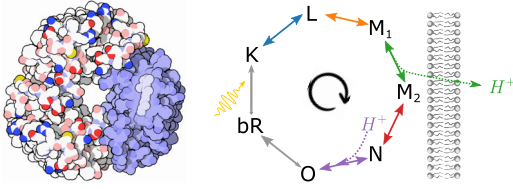


FIG. 1. Left: Bacteriorhodopsin is a biomolecular free energy harvesting machine [18], illustrated in its trimer configuration by Goodsell (CC-BY-4.0) [19]. Right: during each turn of the bacteriorhodopsin photocycle, the molecule absorbs a photon and pumps a proton against the cell's membrane potential.

thermodynamically consistent stochastic model of bacteriorhodopsin. We demonstrate that, under normal operating conditions, the bacteriorhodopsin system is remarkably efficient.

Our main result is formulated as a convex optimization problem which must be solved numerically in the general case. In the second part of this paper, we derive closed-form solutions of this problem for three physically meaningful regimes: the weakly driven linear response regime, the irreversible deterministic regime, and the intermediate near-deterministic regime. These solutions illustrate how optimal harvesting reflects the “alignment” between free energy input and relaxation dynamics. We illustrate these closed-form solutions on two unicyclic systems, which may be interpreted as idealized models of two types of nonequilibrium harvesting devices.

We finish our paper with a brief Discussion. There we relate our approach to previous work, including maximization of power output in steady-state engines and flux balance analysis. We also propose directions for future research.

## II. SETUP

We consider a system with  $n$  mesostates described by the distribution  $\mathbf{p} = (p_1, \dots, p_n) \in \mathbb{R}_+^n$ . The distribution evolves according to the master equation  $\dot{p}_i = \sum_j R_{ij} p_j$ , where  $R_{ij}$  is the transition rate  $i \rightarrow j$  ( $R_{ii} = -\sum_j R_{ji}$ ). Usually  $\mathbf{p}$  represents a probability distribution of a stochastic system with Markovian dynamics [22,23]. However, under an appropriate choice of units, it may also represent chemical concentrations in a deterministic chemical reaction network with pseudo-first-order reactions, such as an enzymatic cycle [24,25].

The system is coupled to a heat bath at inverse temperature  $\beta = 1/k_B T$ , an internal free energy reservoir, and a nonequilibrium environment that serves as an external source of free energy. For example, in a metabolic network, one may consider an internal reservoir of ATP and an external source of glucose. The system has nonequilibrium free energy

$$\mathcal{F}(\mathbf{p}) = \sum_i p_i f_i - \beta^{-1} S(\mathbf{p}), \quad (1)$$

where  $S(\mathbf{p}) := -\sum_i p_i \ln p_i$  is the Shannon entropy and  $f_i$  is the internal free energy of mesostate  $i$  [26].

As mentioned in the Introduction, we suppose that the dynamics of the system can be separated into *baseline* and *control* processes. We make one important assumption in our

analysis: the control processes are only coupled to the heat bath and internal free energy reservoir, but not directly to the external source of free energy. This means that control can only increase harvesting by interacting with the baseline, rather than directly increasing the inflow of free energy from the external source. For example, in a metabolic network, control processes cannot directly increase the import of glucose, but they can affect the rate at which glucose is converted into ATP by optimizing intermediate reactions and mechanisms. Control processes may be driven by the internal reservoir (e.g., coupled to ATP hydrolysis) or undriven (e.g., enzymes).

To formalize the baseline/control distinction, we write the rate matrix as  $R = R^b + R^c$ , where  $R_{ji}^b$  and  $R_{ji}^c$  represent the transition rate of  $i \rightarrow j$  due to baseline and control. Given distribution  $\mathbf{p}$ , the increase of system free energy due to baseline processes is

$$\dot{\mathcal{F}}^b(\mathbf{p}) = \sum_{i,j} p_i R_{ji}^b (f_j + \beta^{-1} \ln p_j). \quad (2)$$

The increase due to control processes is defined analogously but using rate matrix  $R^c$ ,

$$\dot{\mathcal{F}}^c(\mathbf{p}) = \sum_{i,j} p_i R_{ji}^c (f_j + \beta^{-1} \ln p_j). \quad (3)$$

The *harvesting rate* is the rate at which free energy flows to the internal reservoir. The harvesting rate due to baseline processes is

$$\dot{\mathcal{G}}^b(\mathbf{p}) = \sum_{i,j} p_i R_{ji}^b g_{ji}^b + \sum_i p_i \dot{g}_i^b, \quad (4)$$

where  $g_{ji}^b$  is the free energy increase in the internal reservoir due to a single baseline transition  $i \rightarrow j$  and  $\dot{g}_i^b$  is the rate of free energy flow to the internal reservoir due to internal transitions within  $i$  (assuming  $i$  is a coarse-grained mesostate). Similarly, the harvesting rate due to control processes is

$$\dot{\mathcal{G}}^c(\mathbf{p}) = \sum_{i,j} p_i R_{ji}^c g_{ji}^c, \quad (5)$$

where  $g_{ji}^c$  is the free energy increase in the internal reservoir due to control transition  $i \rightarrow j$ . For simplicity, we assume that control cannot exchange free energy with the internal reservoir due to internal transitions within  $i$ . Negative values of  $g_{ji}^b, \dot{g}_i^b, g_{ji}^c$  indicate driving done on the system by the internal reservoir.

For a concrete example of how  $(R^b, R^c, \mathbf{f}, \mathbf{g}^b, \dot{\mathbf{g}}^b, \mathbf{g}^c)$  are defined for a real biomolecular system, see the bacteriorhodopsin example below and the Supplemental Material SM-II [27].

As standard in stochastic thermodynamics, we assume that control processes obey local detailed balance (LDB) [16,28],

$$\ln(R_{ji}^c/R_{ij}^c) = \beta(f_i - f_j - g_{ji}^c) \text{ for } R_{ji}^c > 0. \quad (6)$$

Equation (6) guarantees that the irreversibility of each control transition is determined by the amount of free energy dissipated by that transition [29]. Observe that the right side accounts for free energy changes of the system ( $f_i - f_j$ ) and the internal reservoir ( $g_{ji}^c$ ), but not the external source. This formalizes the assumption that control processes do not exchange free energy with the external source.

We do not require that the baseline rate matrix obeys LDB, although it will often do so for reasons of thermodynamic consistency.

### III. MAXIMUM HARVESTING RATE

Suppose that the combined rate matrix  $R = R^b + R^c$  has the steady-state distribution  $\boldsymbol{\pi}$ , which satisfies  $R^b \boldsymbol{\pi} + R^c \boldsymbol{\pi} = 0$ . The total steady-state harvesting rate due to baseline and control is

$$\dot{\mathcal{G}}^{\text{tot}} = \dot{\mathcal{G}}^b(\boldsymbol{\pi}) + \dot{\mathcal{G}}^c(\boldsymbol{\pi}). \quad (7)$$

We seek to maximize this harvesting rate by varying the parameters of the control processes ( $R^c, \mathbf{g}^c$ ) while holding the baseline parameters ( $\mathbf{f}, R^b, \mathbf{g}^b, \dot{\mathcal{G}}^b$ ) fixed. Finding this maximum would allow us to determine fundamental bounds on harvesting and to evaluate the efficiency of existing harvesting systems.

However,  $\dot{\mathcal{G}}^{\text{tot}}$  is not a concave function of the parameters ( $R^c, \mathbf{g}^c$ ), therefore maximization of Eq. (7) is not a convex optimization problem and is not generally intractable. In the following, we reformulate this maximization as a convex optimization with a physically interpretable objective. This allows us to solve the optimization numerically and, at least for some special cases, also in closed form.

To begin, we rewrite (7) as

$$\dot{\mathcal{G}}^{\text{tot}} = \dot{\mathcal{F}}^b(\boldsymbol{\pi}) + \dot{\mathcal{G}}^b(\boldsymbol{\pi}) - \dot{\Sigma}(\mathbf{J}^c)/\beta, \quad (8)$$

where we introduced the Schnackenberg formula for the entropy production rate (EPR) [22],

$$\dot{\Sigma}(\mathbf{J}^c) = \sum_{i \neq j} J_{ji}^c \ln(J_{ji}^c/J_{ij}^c) \geq 0, \quad (9)$$

where  $J_{ji}^c = \pi_i R_{ji}^c \geq 0$  is the one-way probability flux due to control transition  $i \rightarrow j$ .

Equation (8) has an intuitive physical interpretation: the total steady-state harvesting rate is the rate of free energy increase in the system and internal reservoir due to baseline, minus the rate of dissipation (EPR) due to the control fluxes. The derivation of this result proceeds in two steps (see SM-I A [27] for details). The first step is to show that  $\dot{\Sigma}(\mathbf{J}^c) = -\beta[\dot{\mathcal{F}}^c(\boldsymbol{\pi}) + \beta\dot{\mathcal{G}}^c(\boldsymbol{\pi})]$ , which follows by combining (9) with (3) and (6). This states that the EPR due to control is proportional to the free energy loss in the system and internal reservoir due to control. The second step is to show that  $\dot{\mathcal{F}}^b(\boldsymbol{\pi}) + \dot{\mathcal{F}}^c(\boldsymbol{\pi}) = 0$ , which follows because the left side is the overall derivative of the nonequilibrium free energy  $\mathcal{F}$ , as defined in (1), therefore it must vanish in steady state. The result (8) then follows by combining with Eq. (7) and rearranging.

Importantly, when expressed in the form (8), the harvesting rate is a concave function of the steady-state distribution  $\boldsymbol{\pi}$  and the control fluxes  $\mathbf{J}^c$  (see SM-I B [27]). To find the maximum harvesting rate, we optimize (8) with respect to  $\boldsymbol{\pi}$  and  $\mathbf{J}^c$ . Note that varying  $\boldsymbol{\pi}$  and  $\mathbf{J}^c$  is equivalent to varying the control rate matrix via  $R_{ji}^c = J_{ji}^c/\pi_i$  and control driving  $g_{ji}^c$  via Eq. (6). However, when performing this optimization, we must also ensure that  $\boldsymbol{\pi}$  is the steady-state distribution induced by the fluxes  $\mathbf{J}^c$ . This condition can be expressed as a linear constraint on  $\boldsymbol{\pi}$  and  $\mathbf{J}^c$  via  $R^b \boldsymbol{\pi} + \mathbb{B} \mathbf{J}^c = 0$ . Here  $\mathbf{J}^c$  is treated

as a vector in  $\mathbb{R}^{n^2}$  and  $\mathbb{B} \in \mathbb{R}^{n \times n^2}$  is the incidence matrix with entries  $\mathbb{B}_{k,ij} = \delta_{ki} - \delta_{kj}$ , which guarantees  $R^c \boldsymbol{\pi} = \mathbb{B} \mathbf{J}^c$ .

Combining, we arrive at the bound  $\dot{\mathcal{G}}^{\text{tot}} \leq \mathcal{G}$ , where

$$\mathcal{G} = \sup_{(\boldsymbol{p}, \mathbf{J}) \in \Lambda: R^b \boldsymbol{p} + \mathbb{B} \mathbf{J} = 0} \dot{\mathcal{F}}^b(\boldsymbol{p}) + \dot{\mathcal{G}}^b(\boldsymbol{p}) - \dot{\Sigma}(\mathbf{J})/\beta. \quad (10)$$

In this expression,  $\Lambda$  is the feasible set of distributions  $\boldsymbol{p}$  and control fluxes  $\mathbf{J}$ . At a minimum,  $\Lambda$  ensures the validity of the distribution  $\boldsymbol{p}$  and the fluxes  $\mathbf{J}$  via the linear constraints  $\sum p_i = 1$ ,  $p_i \geq 0$ , and  $J_{ji} \geq 0$ . We write sup instead of max because the set of allowed fluxes is potentially unbounded. Equation (10) implies a tradeoff between the total gain of free energy in the system and internal reservoir due to baseline (which depends only on  $\boldsymbol{p}$ ) and the dissipation due to control fluxes (which depends only on  $\mathbf{J}$ ).

Importantly, the feasible set  $\Lambda$  can include additional convex constraints, which may introduce topological, kinetic, thermodynamic, etc. restrictions on the control processes. Topological constraints restrict which transitions can be controlled; e.g.,  $J_{ji} = 0$  ensures that control does not use transition  $i \rightarrow j$ . Kinetic constraints restrict time scales of control processes, as might reflect underlying physical processes like diffusion; e.g., an upper bound on control transition rate  $R_{ji}^c = J_{ji}/p_i \leq \kappa$  can be enforced via  $J_{ji} \leq p_i \kappa$ . Thermodynamic constraints bound the affinity [22] of control transitions; e.g.,  $J_{ji} e^{-A} \leq J_{ij} \leq J_{ji} e^A$  ensures that  $|\ln(J_{ij}/J_{ji})| \leq A$ . The above examples all involve linear constraints. An example of a nonlinear, but still convex, constraint is an upper bound on the EPR incurred by control,  $\dot{\Sigma}(\mathbf{J}) \leq \dot{\Sigma}_{\text{max}}^c$ .

Equation (10) is our first main result. Importantly,  $\mathcal{G}$  is defined purely in terms of the thermodynamic and kinetic properties of the baseline process, along with desired constraints encoded in  $\Lambda$ . Thus,  $\mathcal{G}$  is the maximum steady-state harvesting rate that can be achieved by any allowed control processes, given a fixed baseline. In addition, Eq. (10) involves the maximization of a concave objective given convex constraints. This is equivalent to the minimization of a convex objective, thus Eq. (10) is a convex optimization problem that can be efficiently solved using standard numerical techniques [30]. The optimization also identifies an optimal steady-state distribution  $\boldsymbol{p}^*$  and control fluxes  $\mathbf{J}^*$  that achieve the maximum harvesting rate  $\mathcal{G}$  (or come arbitrarily close to achieving it). These fix the optimal control rate matrix via  $R_{ji}^{c*} = J_{ij}^*/p_i^*$ . Thus, our optimization specifies an upper bound on harvesting as well as the optimal control strategy that achieves this bound.

There is an important special case in which our optimization problem is simplified. Suppose that  $\Lambda$  does not enforce additional constraints on  $\boldsymbol{p}$  and  $\mathbf{J}$  (more generally, we permit topological constraints if the graph of allowed transitions is connected and contains all  $n$  states). Then, the objective is maximized in limit of fast control,  $\mathbf{J} \rightarrow \infty$  and  $\dot{\Sigma}(\mathbf{J}) \rightarrow 0$ . We can then write Eq. (10) as an optimization over steady-state distributions:

$$\mathcal{G} := \max_{\boldsymbol{p}: \sum p_i = 1, p_i \geq 0} \dot{\mathcal{F}}^b(\boldsymbol{p}) + \dot{\mathcal{G}}^b(\boldsymbol{p}). \quad (11)$$

The optimal  $\boldsymbol{p}^*$  is unique as long as the baseline rate matrix is irreducible. The optimal control rate matrix is very

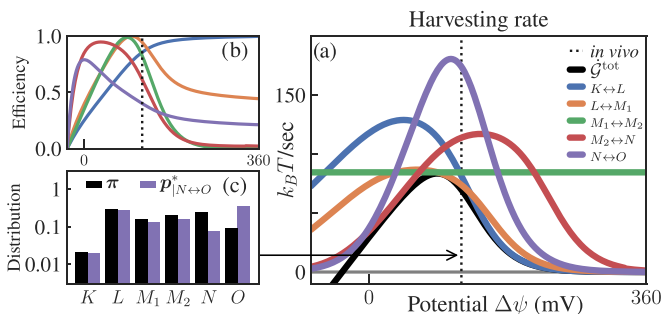


FIG. 2. (a) Comparison of the actual harvesting rate  $\dot{G}^{\text{tot}}$  at different electrical potentials  $\Delta\psi$ , versus maximum rate  $\mathcal{G}$  achieved by optimizing five intermediate transitions (color scheme from Fig. 1, right). (b) Efficiency  $\dot{G}^{\text{tot}}/\mathcal{G}$  computed while separately optimizing each transition, with colors as in (a). (c) The actual steady state  $\pi$  versus the optimal distribution  $p^*$  when optimizing the  $N \leftrightarrow O$  transition (at  $\Delta\psi = 120$  mV).

fast ( $R^{c*} \rightarrow \infty$ ) and obeys detailed balance for  $p^*$ ,  $R^{c*}_{ji} p^*_i = R^{c*}_{ij} p^*_j$ . For details, see the SM-I C and SM-I D [27].

#### IV. BACTERIORHODOPSIN

We illustrate our results using bacteriorhodopsin, a light-driven proton pump from Archaea [18].

We define a thermodynamically consistent model of the bacteriorhodopsin cycle using published thermodynamic [31] and kinetic [32] data (see SM-II [27]). The system operates in a cyclical manner, absorbing a photon and pumping a proton during each turn of the cycle (Fig. 1, right). Specifically, the transition  $M_1 \rightarrow M_2$  pumps a proton out of the cell. This stores free energy in the internal reservoir (the membrane electrochemical potential),

$$g_{M_2 M_1} = -g_{M_1 M_2} = e\Delta\psi - (\ln 10)\beta^{-1}\Delta\text{pH}, \quad (12)$$

where  $e$  is the elementary charge constant,  $\Delta\psi$  is the membrane electrical potential, and  $\Delta\text{pH}$  is the membrane pH difference. The other transitions in the cycle do not affect the free energy of the internal reservoir ( $g_{ij} = 0$  and  $\dot{g}_i = 0$ ).

During the transition  $\text{bR} \rightarrow K$ , the system leaves the ground state by absorbing a photon at 580 nm, thereby harvesting free energy from the external source and dissipating some heat to the environment at  $T = 293^\circ$  K. This transition is much faster (picoseconds) than the other transitions in the photocycle (micro- to milliseconds). As commonly done in photochemistry [33], we coarse-grain transitions  $O \rightarrow \text{bR}$  and  $\text{bR} \rightarrow K$  into a single effective transition  $O \rightarrow K$ .

To explore the performance of bacteriorhodopsin under different conditions, we vary the membrane electrical potential  $\Delta\psi$  between  $-75$  and  $350$  mV, while using a realistic fixed  $\Delta\text{pH} = -0.6$  [34]. We show the actual harvesting rate ( $\dot{G}^{\text{tot}}$  in units of  $k_B T/\text{sec}$ ) at different potentials as a black line in Fig. 2(a). At a plausible *in vivo*  $\Delta\psi = 120$  mV [34], the model exhibits a steady-state current of 11.5 protons/sec, each proton carrying  $6.1 k_B T$  of free energy, corresponding to  $\dot{G}^{\text{tot}} \approx 70 k_B T/\text{sec}$ . The largest output is achieved near the *in vivo* potential: at lower potentials, the cycle current saturates while the free energy per proton drops, and at higher potentials the pump stalls.

Next, we quantify the maximum harvesting rate that can be achieved by optimizing the parameters of individual transitions. This analysis is relevant for understanding limits on increasing bacteriorhodopsin output, whether via natural selection or biosynthetic methods [35–38]. Interestingly, such transition-level optimization may be feasible in bacteriorhodopsin, as the properties of several transitions are known to be individually controlled by particular amino acid residues in the bacteriorhodopsin protein [35,39–41].

For each reversible transition in the cycle, for instance  $N \leftrightarrow O$ , we define the baseline as the rest of the cycle without that transition. We then optimize control under the topological constraint that only the relevant transition (e.g.,  $N \leftrightarrow O$ ) is allowed. The analysis is repeated for all transitions, except for the (coarse-grained) photon-absorbing transition  $O \leftrightarrow K$ , which is in accordance with our assumption that control cannot directly exchange free energy directly with the external source.

Figure 2(a) shows  $\mathcal{G}$ , the maximum  $\dot{G}^{\text{tot}}$  achievable by optimizing each reversible transition. In Fig. 2(b), we also show the efficiency  $\dot{G}^{\text{tot}}/\mathcal{G} \leq 1$  for each transition, that is the ratio of the actual and maximum harvesting rate.

Several transitions, such as  $K \leftrightarrow L$ ,  $L \leftrightarrow M_1$ ,  $M_1 \leftrightarrow M_2$ , are remarkably efficient ( $\geq 85\%$ ) near *in vivo* membrane potentials. The reprotonation step  $N \leftrightarrow O$  is the least efficient ( $\sim 40\%$ ) and also has the slowest kinetics of the five transitions studied in Fig. 2. This suggests that  $N \leftrightarrow O$  is a bottleneck whose optimization can have a big impact on the harvesting rate, while optimization of other nonbottleneck transitions has a more limited effect.

Observe that  $\mathcal{G}$  for  $M_1 \leftrightarrow M_2$  does not depend on  $\Delta\psi$ . This is because  $\mathcal{G}$  is a function of baseline properties, which do not depend on the membrane potential when  $M_1 \leftrightarrow M_2$  is treated as control. Conversely,  $M_1 \leftrightarrow M_2$  as control transition can be optimized by varying the membrane potential and/or scaling up the forward/backward rates. Our results show that this transition is very close to optimal at *in vivo* membrane potentials and kinetic time scales.

Optimal distributions  $p^*$  are also obtained, with a typical one shown in Fig. 2(c). We find a consistent shift toward state  $O$ , which accelerates the reset of the cycle and increases the flux across the photon-absorbing transition  $O \rightarrow K$ .

In the SM-II [27], we illustrate how the efficiency of bacteriorhodopsin transitions can be evaluated under other types of constraints, including constraints on thermodynamic affinity, dynamical activity, and kinetics.

As a final analysis, instead of optimizing individual existing transitions in the bacteriorhodopsin cycle, we ask to what extent the harvesting rate can be increased by any *additional* control processes. For example, this could involve an additional enzyme that shifts the cycle's steady state by permitting a new transition between distant states (e.g.,  $L \leftrightarrow N$ ), possibly yielding an increase of the proton pumping rate.

In this case, we treat the entire bacteriorhodopsin system as the baseline, and we do not introduce any additional constraints on the steady-state distribution or the control fluxes. Then, as shown in the SM-I D [27], the objective in Eq. (10) is achieved in the limit of fast control, and the maximum harvesting rate can be found by solving the simplified optimization problem (11).

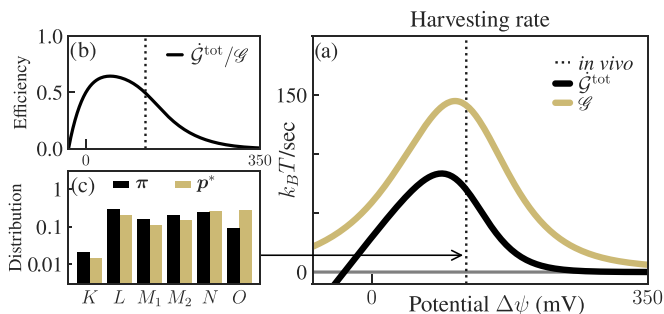


FIG. 3. (a) Comparison of the actual harvesting rate  $\dot{G}^{\text{tot}}$  at different electrical potentials  $\Delta\psi$ , versus maximum rate  $\mathcal{G}$  achieved by fixing the bacteriorhodopsin cycle as baseline and allowing any additional transitions as control. (b) Efficiency of the actual bacteriorhodopsin cycle with respect to the optimized cycle. (c) The actual steady state  $\pi$  and optimal distribution  $p^*$  (at  $\Delta\psi = 120$  mV).

For this setup, Fig. 3(a) shows the baseline (actual) harvesting rate  $\dot{G}^{\text{tot}}$  and the maximum harvesting rate  $\mathcal{G}$  at varying  $\Delta\psi$ . Interestingly, both peak at around the *in vivo* values of the membrane potential. In Fig. 3(b), we show that the actual bacteriorhodopsin cycle harvests approximately 50% of the fundamental bound given by  $\mathcal{G}$  (at *in vivo* values of  $\Delta\psi$ ). This suggests that bacteriorhodopsin is remarkably close to optimal, relative to improvements that could be achieved by introducing any additional control processes.

We also show the actual steady-state distribution and the optimal distribution  $p^*$  in Fig. 3(c). The optimal distribution increases the probability of state O, similar to the optimal distribution found by optimizing the  $N \leftrightarrow O$  transition, shown in Fig. 2(c). However, unlike Fig. 2(c), where most of the extra probability is taken from state N, in Fig. 3(c) the probability is drawn more uniformly from other states in the cycle, indicating the presence of distributed control.

## V. LIMITING REGIMES

Our results are stated via an optimization problem that generally does not have a closed-form solution. In our second set of results, we identify closed-form expressions in three physically meaningful regimes. For simplicity, here we focus on the simplified objective (11). See the SM-III [27] for detailed derivations, including analysis of the conditions under which each of these three approximation are valid.

For convenience, we first rewrite (11) as

$$\mathcal{G} = \max_{\mathbf{p}} -\dot{S}^b(\mathbf{p})/\beta + \sum_i p_i \phi_i, \quad (13)$$

where  $\dot{S}^b(\mathbf{p}) = -\sum_{i,j} R_{ij}^b p_j \ln p_i$  is the increase of the Shannon entropy of  $\mathbf{p}$  under  $R^b$  and for convenience we defined  $\phi_i := \dot{g}_i^b + \sum_j R_{ji}^b (f_j - f_i + g_{ji}^b)$ . The objective (13) contains a nonlinear term  $-\dot{S}^b(\mathbf{p})/\beta$  quantifying the decrease of information-theoretic entropy and a linear term  $\sum_i p_i \phi_i$  quantifying the flow of thermodynamic free energy.

Next, we consider three regimes.

**Linear response (LR)** applies when the optimal distribution  $p^*$  is close to the steady-state distribution of the baseline rate matrix  $R^b$ . Suppose that  $R^b$  is irreducible and has a unique

steady state  $\pi^b$  with full support. We introduce the ‘‘additive reversibilization’’ of  $R^b$ ,

$$A_{ij} = (R_{ij}^b + R_{ji}^b \pi_i^b / \pi_j^b) / 2. \quad (14)$$

The rate matrix  $A$  obeys detailed balance (DB) for the steady-state distribution  $\pi^b$  and has the same dynamical activity [42] on all edges as  $R^b$ .  $A$  may be considered as a DB version of  $R^b$ , and it is equal to  $R^b$  when the latter obeys DB [43,44]. Let  $\mathbf{u}^\alpha$  indicate the  $\alpha$ th right eigenvector of  $A$  normalized as  $\sum_i (u_i^\alpha)^2 / \pi_i^b = 1$ , and  $\lambda_\alpha$  the corresponding real-valued eigenvalue ( $\lambda_1 = 0$ ). The LR solution for the maximum harvesting rate and the optimal distribution is

$$\mathcal{G} \approx \dot{G}^b(\pi^b) + \beta \sum_{\alpha>1} \frac{|\Omega_\alpha|^2}{-\lambda_\alpha}, \quad p^* \approx \pi^b + \beta \sum_{\alpha>1} \frac{\Omega_\alpha}{-\lambda_\alpha} \mathbf{u}^\alpha, \quad (15)$$

where  $\Omega_\alpha = (\phi + \beta^{-1} R^{bT} \ln \pi^b)^T \mathbf{u}^\alpha / 2$  quantifies the harvesting ‘‘amplitude’’ for mode  $\alpha$ .

Equation (15) has a simple interpretation. In addition to the baseline harvesting rate  $\dot{G}^b(\pi^b)$ ,  $\mathcal{G}$  contains contributions from the relaxation modes of  $A$ , with each mode weighed by its squared amplitude and relaxation time scale  $-1/\lambda_\alpha$ . All else being equal,  $\mathcal{G}$  is large when slow modes have large harvesting amplitudes. The optimal  $p^*$  shifts the baseline steady state  $\pi^b$  toward mode  $\alpha$  in proportion to that mode’s harvesting amplitude and relaxation timescale, thereby optimally balancing the tradeoff between harvesting and dissipation.

The **Deterministic (D)** regime applies when the nonlinear information-theoretic term in Eq. (13) is much smaller than the linear thermodynamic term. We can then ignore the former, turning Eq. (13) into a simple linear optimization. This gives the approximate solution

$$\mathcal{G} \approx \phi_{i^*}, \quad p_i^* \approx \delta_{i,i^*}, \quad (16)$$

where  $i^* = \arg \max_i \phi_i$  is the optimal mesostate. This solution concentrates probability on a single mesostate, effectively ignoring the cost of maintaining this low-entropy distribution.

The **Near-Deterministic (ND)** regime lies between Linear Response and Deterministic ones. By perturbing  $p^*$  around  $\delta_{i,i^*}$ , we can decouple the values of  $p_i$  in the objective function (11). The maximal harvesting rate and optimal distribution in this regime are then given by

$$\mathcal{G} \approx \phi_{i^*} + \beta^{-1} \sum_{i \neq i^*} R_{ii^*}^b (\ln p_i^* - 1),$$

$$p_i^* \approx \begin{cases} R_{ii^*}^b / [\beta(\phi_{i^*} - \phi_i) + R_{ii^*}^b] & i \neq i^* \\ 1 - \sum_{i:i \neq i^*} p_i^* & i = i^* \end{cases}. \quad (17)$$

The ND solution also has a simple interpretation. It perturbs the D solution by shifting probability towards states with high transition rates away from the optimal state (large  $R_{ii^*}^b$ ) and small decreases in harvesting ( $\phi_{i^*} - \phi_i$ ). This balances the benefit of harvesting against the cost of pumping probability against  $R_{ii^*}^b$ .

## VI. EXAMPLE: UNICYCLIC SYSTEMS

We illustrate our closed-form solutions using two simple models, both based on a unicyclic system with  $n$  states. The

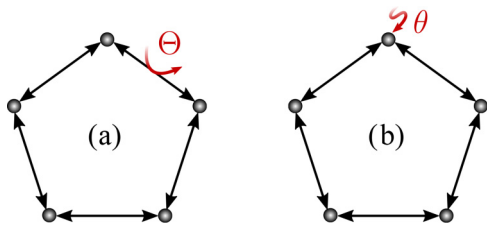


FIG. 4. (a) Unicyclic system where free energy  $\Theta$  is harvested by a single transition. (b) Unicyclic system where free energy per unit time  $\theta$  is harvested when the system is in a particular mesostate.

baseline dynamics involve diffusion across a one-dimensional ring, with left and right jump rates set to 1. The baseline steady state is a uniform distribution,  $\pi_i^b = 1/n$ , with no cyclic current. We assume a uniform free energy function,  $f_i = 0$  for all  $i$ .

We consider two different scenarios. In the first scenario, shown schematically in Fig. 4(a),  $\Theta$  of free energy is harvested each time the system carries out the transition  $1 \rightarrow 2$ , so

$$g_{21}^b = -g_{12}^b = \Theta,$$

and  $g_{ij}^b = g_i^b = 0$  otherwise. This scenario may be interpreted as an idealized model of a biomolecular harvesting cycle, such as a transporter. In the second scenario, shown schematically in Fig. 4(b), free energy is harvested at a rate of  $\theta$  per unit time when the system is located in one particular mesostate  $i^* = 1$ , so

$$g_1^b = \theta,$$

and  $g_{ij}^b = g_i^b = 0$  otherwise. This scenario may be interpreted as an idealized model of error correction or self-assembly, where free energy can only be harvested when the system is in some particular functional mesostate.

For both scenarios, we evaluate the maximum harvesting rate  $\mathcal{G}$  and the optimal distributions achievable by adding any possible control to the system, without constraints. We report exact values found by numerical optimization of Eq. (13), as well as the LR, ND, and D approximations described above. To calculate the LR values, we exploit the fact that the baseline unicyclic rate matrix is a circulant matrix with a simple eigendecomposition [45]. Full details of the derivations for the two scenarios are provided in the SM-IV [27].

We first report results for the first scenario from Fig. 4(a), where free energy is harvested during the transition  $2 \rightarrow 1$ . Observe that baseline harvesting rate vanishes,  $\dot{\mathcal{G}}^b(\boldsymbol{\pi}^b) = 0$ , since harvesting free energy requires a cyclic current. In Fig. 5(a), we plot the maximum harvesting rate  $\mathcal{G}$  and its approximations as a function of the supplied free energy  $\Theta$ . For small  $\Theta$ , LR applies and the maximum harvesting rate is

$$\mathcal{G} \approx \beta\Theta^2(n-1)/4n^2. \quad (19)$$

The optimal distribution in the LR regime, shown in Fig. 5(c), builds up in equal increments starting from  $i = i^* + 1$  until the optimal state  $i^* = 1$ , after which it drops sharply. For large  $\Theta$ , the D regime is relevant and the optimal distribution concentrates on the optimal state  $i^* = 1$ , so

$$\mathcal{G} \approx \Theta. \quad (20)$$

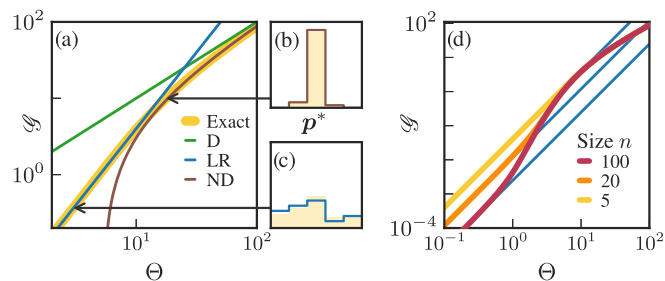


FIG. 5. (a) Maximum harvesting rate  $\mathcal{G}$  for the unicyclic system from Fig. 4(a), as a function of supplied free energy  $\Theta$ . Exact value is found numerically, LR, D, and ND are calculated using approximations described in the text. Exact and approximate optimal distributions  $\boldsymbol{p}^*$  in ND (b) and LR (c) regimes are shown, with the optimal state  $i^* = 1$  located in the middle of the histograms. (d)  $\mathcal{G}$  and its LR approximation for different  $\Theta$  and system sizes  $n$ .

At intermediate  $\Theta$ , the ND regime applies, which gives

$$\mathcal{G} \approx \Theta - \beta^{-1} \{2 + \ln[2(\beta\Theta - 1)(\beta\Theta - 2)]\}. \quad (21)$$

The optimal distribution in the ND regime, shown in Fig. 5(b), allocates  $p_{i^*-1}^* = 1/(\beta\Theta - 2)$ ,  $p_{i^*+1}^* = 1/(2\beta\Theta - 2)$  and the rest to the optimal state  $p_{i^*}^*$ .

Next, we consider the second scenario from Fig. 4(b), where free energy is harvested when the system is in the optimal mesostate  $i^* = 1$ . Observe that the uniform baseline steady state assigns  $1/n$  probability to the optimal state, thus in this scenario the baseline harvesting rate is  $\dot{\mathcal{G}}^b(\boldsymbol{\pi}^b) = \theta/n$ . To facilitate comparison with the first scenario, we focus on the increase of the maximum harvesting rate relative to baseline,

$$\Delta\mathcal{G} := \mathcal{G} - \dot{\mathcal{G}}^b(\boldsymbol{\pi}^b) = \mathcal{G} - \theta/n. \quad (22)$$

In Fig. 6(a), we show  $\Delta\mathcal{G}$  and its approximations as a function of the free energy supply rate  $\theta$ . For small  $\theta$ , LR applies and the maximum harvesting rate is

$$\Delta\mathcal{G} \approx \beta\theta^2/48. \quad (23)$$

The optimal distribution in the LR regime, shown in Fig. 6(c), is symmetric about the optimal state  $i^* = 1$ . For large  $\theta$ , the

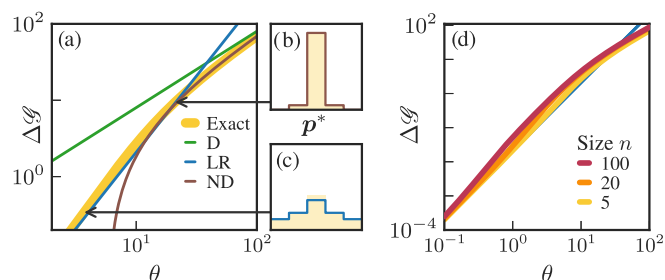


FIG. 6. (a) Maximum harvesting rate  $\mathcal{G}$  for the unicyclic system from Fig. 4(b), as a function of supplied free energy rate  $\theta$ . Exact value is found numerically, LR, D, and ND are calculated using approximations described in the text. Exact and approximate optimal distributions  $\boldsymbol{p}^*$  in ND (b) and LR (c) regimes are shown, with the optimal state  $i^* = 1$  located in the middle of the histograms. (d)  $\mathcal{G}$  and its LR approximation for different  $\theta$  and system sizes  $n$ .

D regime is relevant and the optimal distribution concentrates on the optimal state  $i^* = 1$ , so

$$\Delta\mathcal{G} \approx \theta - \theta/n. \quad (24)$$

At intermediate  $\theta$ , the ND regime applies, giving

$$\Delta\mathcal{G} \approx \theta - \theta/n - 2\beta^{-1}[1 + \ln(\beta\theta - 2)]. \quad (25)$$

The optimal distribution in the ND regime, shown in Fig. 6(b), allocates  $p_{i^*-1}^* = p_{i^*+1}^* = 1/(\beta\theta - 2)$  and the rest to  $p_{i^*}^*$ .

There are some similarities among the two harvesting scenarios. For both scenarios, in the LR regime, the increase of the harvesting rate scales quadratically in the supplied free energy ( $\Theta$  or  $\theta$ ) and linearly in inverse temperature  $\beta$ . This scaling reflects the fact that the optimal strategy has to balance harvesting ( $\Theta$  or  $\theta$  contributions) with the thermodynamic cost of maintaining a low entropy  $\mathbf{p}^*$  ( $\beta$  contributions). In the ND and D regimes,  $\mathcal{G}$  scales linearly in the supplied free energy and loses its linear dependence on  $\beta$ . Thus, outside of LR, the thermodynamic cost of maintaining a low entropy distribution has a minor effect on the optimal strategy.

There are also important differences between the two scenarios. For the first scenario, the optimal strategy maintains an asymmetric  $\mathbf{p}^*$ , thereby generating a net flux across the transition  $2 \rightarrow 1$ . In the LR regime, the cost of maintaining this asymmetric distribution grows with the system size  $n$ , therefore the maximum harvesting rate in Eq. (19) scales as  $\sim O(n^{-1})$ . This is shown in Fig. 5(d), where we plot  $\mathcal{G}$  and its LR approximation at various  $\Theta$  and  $n$ . For the second scenario, the optimal strategy maintains a peaked but symmetric  $\mathbf{p}^*$ . Remarkably, the cost of maintaining this distribution does not depend on system size  $n$ . This is shown in Fig. 6(d), where we plot  $\Delta\mathcal{G}$  at various  $\theta$  and  $n$ .

## VII. DISCUSSION

In this paper, we consider the problem of optimizing free energy harvesting in a nonequilibrium steady-state system. We demonstrate that this problem can be formulated as a constrained convex optimization problem, and we use this formulation to study optimal harvesting and efficiency in the bacteriorhodopsin proton pump. We also solve the convex optimization problem in closed-form for three limiting regimes, as illustrated on two unicyclic models discussed above.

A key step in our analysis is to separate the dynamics of the system into separate contributions from fixed baseline processes and optimizable control processes. We note that, in stochastic thermodynamics, the baseline/control separation has been previously used to study autonomous Maxwellian demons [46,47], counterdiabatic driving [48], and the cost of maintaining a nonequilibrium steady state [49,50].

We derive a simplified bound on the maximum harvesting rate in Eq. (13), which is achieved in the limit of fast dissipation-less control. Interestingly, this expression involves a tradeoff between two terms, one information-theoretic and one thermodynamic. At first glance, this resembles information/free-energy tradeoffs characteristic of Maxwellian demons and other ‘‘information engines’’ [51–58]. However, there are important differences. In a typical information engine, there is no external source of driving

and information serves as *fuel*, which can be converted into  $\beta^{-1} \ln 2$  of thermodynamic free energy per bit. In our case, there is an external source of free energy that in some cases can be harvested more effectively by reducing the system’s statistical entropy, e.g., by concentrating it on favorable states. Here, a bit of information can increase the harvesting rate by a very large amount (much larger than  $\beta^{-1} \ln 2/\text{bit}$ ), and information acts more like a *catalyst* than a fuel [59,60]. Loosely speaking, this is similar to how information encoded in the sequence of a metabolic enzyme is not itself fuel, but rather allows metabolism to harvest a large amount of fuel from elsewhere.

We finish by mentioning some connections to previous work and future directions. First, our approach may be related to prior work on optimizing power output and free energy transduction in steady-state molecular machines [29,61–66]. Here we consider the general problem of optimizing a set of control processes, given a fixed baseline and possible additional constraints on kinetics, topology, and thermodynamics. Previous work does not make the baseline/control distinction; instead, it is mostly concerned with the problem of optimizing system performance with respect to a small set of specific parameters or observables of interest, such as the location of free energy barriers [62,63,65,66], efficiency [66], and the size of fluctuations [61,64].

There is also an interesting relation between our work and flux balance analysis (FBA) [67–69]. The goal of FBA is to identify deterministic fluxes in biological metabolic systems that optimize biomass production, or other similar metrics of performance. This can be formulated as a linear program, which may include linear constraints that enforce thermodynamically favored reaction directions [69] (interestingly, in Ref. [70], the authors propose a version of FBA that also accounts for the entropy production rate). Our setting and optimization are different from FBA and its variants. We seek to optimize free energy harvesting at the stochastic level, and our objective involves nonlinear information-theoretic contributions to free energy. In addition, our optimization involves both the steady-state distribution  $\mathbf{p}$  and fluxes  $\mathbf{J}$ , which allows us to optimize harvesting due to internal transitions within coarse-grained mesostates, as in Fig. 4(b). Nonetheless, investigating the relationship between our approach and FBA is an interesting direction for future work.

Another interesting direction for future work is to consider stochastic fluctuations of free energy harvesting. In particular, the thermodynamic uncertainty relation may be used to study tradeoffs between the entropy production rate, the average harvesting rate (the quantity  $\dot{Q}^{\text{tot}}$  considered here), the fluctuations in the amount of harvested free energy [64,71]. For biomolecular systems, large fluctuations in harvesting can lead to starvation, suggesting that minimizing fluctuations may be of significant biological importance.

Finally, an interesting direction for future work is to consider free energy harvesting in a system embedded in a fluctuating environment. For example, one may imagine a harvesting system in an environment with fluctuating sugar sources, or with a fluctuating amount of available light. In this setting, it is natural to optimize the harvesting rate under the topological constraint that control fluxes cannot directly change the state of the environment, for instance using bipartite

models of Markovian dynamics [72]. It would be interesting to investigate how, under the optimal harvesting strategy, the information flow from the environment to the system varies with the abundance of free energy and complexity of the environment.

### ACKNOWLEDGMENTS

We thank members of the Complex Systems Lab, B. Corominas-Murtra, L. Seoane, D. Wolpert, and D. Sowiński

for useful discussions. A.K. also thanks Sosuke Ito for support and encouragement. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 101068029. J.P. was supported by the María de Maeztú fellowship MDM-2014-0370-17-2 and Grant No. 62417 from the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

- 
- [1] T. L. Hill, *Free Energy Transduction in Biology: The Steady-State Kinetic and Thermodynamic Formalism* (Academic Press, New York, 1977).
- [2] H. Morowitz, *Foundations of Bioenergetics* (Elsevier, Amsterdam, 2012).
- [3] A. J. Lotka, Contribution to the energetics of evolution, *Proc. Natl. Acad. Sci. USA* **8**, 147 (1922).
- [4] J. H. Brown, P. A. Marquet, and M. L. Taper, Evolution of body size: Consequences of an energetic definition of fitness, *Am. Nat.* **142**, 573 (1993).
- [5] W. B. Watt, Bioenergetics and evolutionary genetics: Opportunities for new synthesis, *Am. Nat.* **125**, 118 (1985).
- [6] O. P. Judson, The energy expansions of evolution, *Nat. Ecol. Evol.* **1**, 0138 (2017).
- [7] H.-J. Choi and C. D. Montemagno, Artificial organelle: ATP synthesis from cellular mimetic polymersomes, *Nano Lett.* **5**, 2538 (2005).
- [8] K. Y. Lee, S.-J. Park, K. A. Lee, S.-H. Kim, H. Kim, Y. Meroz, L. Mahadevan, K.-H. Jung, T. K. Ahn, K. K. Parker *et al.*, Photosynthetic artificial organelles sustain and control atp-dependent reactions in a protocellular system, *Nat. Biotechnol.* **36**, 530 (2018).
- [9] S. Berhanu, T. Ueda, and Y. Kuruma, Artificial photosynthetic cell producing energy for protein synthesis, *Nat. Commun.* **10**, 1325 (2019).
- [10] K. Villa and M. Pumera, Fuel-free light-driven micro/nanomachines: Artificial active matter mimicking nature, *Chem. Soc. Rev.* **48**, 4966 (2019).
- [11] C. Kleineberg, C. Wölfer, A. Abbasnia, D. Pischel, C. Bednarz, I. Ivanov, T. Heitkamp, M. Börsch, K. Sundmacher, and T. Vidaković-Koch, Light-driven atp regeneration in diblock-grafted hybrid vesicles, *ChemBioChem.* **21**, 2149 (2020).
- [12] T. E. Miller, T. Beneyton, T. Schwander, C. Diehl, M. Girault, R. McLean, T. Chotel, P. Claus, N. S. Cortina, J.-C. Baret *et al.*, Light-powered CO<sub>2</sub> fixation in a chloroplast mimic with natural and synthetic parts, *Science* **368**, 649 (2020).
- [13] C. Guindani, L. C. da Silva, S. Cao, T. Ivanov, and K. Landfester, Synthetic cells: From simple bio-inspired modules to sophisticated integrated systems, *Angew. Chem.* **134**, e202110855 (2022).
- [14] P. Albanese, F. Mavelli, and E. Altamura, Light energy transduction in liposome-based artificial cells, *Front. Bioeng. Biotechnol.* **11**, 1161730 (2023).
- [15] N. S. Chandel, Glycolysis, *Cold Spring Harb. Perspect. Biol.* **13**, a040535 (2021).
- [16] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, *Rep. Prog. Phys.* **75**, 126001 (2012).
- [17] J. C. Doyle, B. A. Francis, and A. R. Tannenbaum, *Feedback Control Theory* (Courier Corporation, Chelmsford, MA, 2013).
- [18] J. K. Lanyi, Bacteriorhodopsin, *Annu. Rev. Physiol.* **66**, 665 (2004).
- [19] D. Goodsell, Molecule of the Month: Bacteriorhodopsin, Protein data Bank 3 (2002), doi:10.2210/rcsb\_pdb/mom\_2002\_3.
- [20] O. Béja, L. Aravind, E. V. Koonin, M. T. Suzuki, A. Hadd, L. P. Nguyen, S. B. Jovanovich, C. M. Gates, R. A. Feldman, J. L. Spudich *et al.*, Bacterial rhodopsin: evidence for a new type of phototrophy in the sea, *Science* **289**, 1902 (2000).
- [21] L. Gómez-Consarnau, J. A. Raven, N. M. Levine, L. S. Cutter, D. Wang, B. Seegers, J. Arístegui, J. A. Fuhrman, J. M. Gasol, and S. A. Sañudo-Wilhelmy, Microbial rhodopsins are major contributors to the solar energy captured in the sea, *Sci. Adv.* **5**, eaaw8855 (2019).
- [22] J. Schnakenberg, Network theory of microscopic and macroscopic behavior of master equation systems, *Rev. Mod. Phys.* **48**, 571 (1976).
- [23] M. Esposito and C. Van den Broeck, Three faces of the second law. I. Master equation formulation, *Phys. Rev. E* **82**, 011143 (2010).
- [24] A. Wachtel, R. Rao, and M. Esposito, Thermodynamically consistent coarse graining of biocatalysts beyond Michaelis-Menten, *New J. Phys.* **20**, 042002 (2018).
- [25] A. Cornish-Bowden, *Fundamentals of Enzyme Kinetics* (Wiley, New York, 2013).
- [26] J. M. Horowitz, T. Sagawa, and J. M. R. Parrondo, Imitating chemical motors with optimal information motors, *Phys. Rev. Lett.* **111**, 010602 (2013).
- [27] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevResearch.6.013275> for a derivation of our main results, a full description of the bacteriorhodopsin model and the details for the analyses of limiting regimes and unicyclic examples.
- [28] C. Maes, Local detailed balance, *SciPost Phys. Lect. Notes* **32** (2021).
- [29] A. I. Brown and D. A. Sivak, Theory of nonequilibrium free energy transduction by molecular machines, *Chem. Rev.* **120**, 434 (2020).
- [30] S. P. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2004).
- [31] G. Varo and J. K. Lanyi, Thermodynamics and energy coupling in the bacteriorhodopsin photocycle, *Biochemistry* **30**, 5016 (1991).
- [32] V. A. Lórenz-Fonfría and H. Kandori, Spectroscopic and kinetic evidence on how bacteriorhodopsin accomplishes vectorial



- proton transport under functional conditions, *J. Am. Chem. Soc.* **131**, 5891 (2009).
- [33] E. Penocchio, R. Rao, and M. Esposito, Nonequilibrium thermodynamics of light-induced reactions, *J. Chem. Phys.* **155**, 114101 (2021).
- [34] J. K. Lanyi, Light energy conversion in *Halobacterium halobium*, *Microbiol. Rev.* **42**, 682 (1978).
- [35] A. Miller and D. Oesterhelt, Kinetic optimization of bacteriorhodopsin by aspartic acid 96 as an internal proton donor, *Biochim. Biophys. Acta-Bioenergetics* **1020**, 57 (1990).
- [36] A. Seitz and N. Hampp, Kinetic optimization of bacteriorhodopsin films for holographic interferometry, *J. Phys. Chem. B* **104**, 7183 (2000).
- [37] K. J. Wise, N. B. Gillespie, J. A. Stuart, M. P. Krebs, and R. R. Birge, Optimization of bacteriorhodopsin for bioelectronic devices, *Trends Biotechnol.* **20**, 387 (2002).
- [38] J. R. Hillebrecht, K. J. Wise, J. F. Kosciulecki, and R. R. Birge, Directed evolution of bacteriorhodopsin for device applications, in *Methods in Enzymology* (Elsevier, Amsterdam, 2004), Vol. 388, pp. 333–347.
- [39] J. Tittor, M. Wahl, U. Schweiger, and D. Oesterhelt, Specific acceleration of de- and reprotonation steps by azide in mutated bacteriorhodopsins, *Biochim. Biophys. Acta-Bioenergetics* **1187**, 191 (1994).
- [40] S. P. Balashov, M. Lu, E. S. Imasheva, R. Govindjee, T. G. Ebrey, B. Othersen, Y. Chen, R. K. Crouch, and D. R. Menick, The proton release group of bacteriorhodopsin controls the rate of the final step of its photocycle at low Ph, *Biochemistry* **38**, 2026 (1999).
- [41] Q. Li, S. Bressler, D. Ovrutsky, M. Ottolenghi, N. Friedman, and M. Sheves, On the protein residues that control the yield and kinetics of O<sub>630</sub> in the photocycle of bacteriorhodopsin, *Biophys. J.* **78**, 354 (2000).
- [42] The dynamical activity refers to the overall rate of back-and-forth jumps across a transition,  $R_{ij}^b \pi_j^b + R_{ji}^b \pi_i^b$ .
- [43] A. Kolchinsky, N. Ohga, and S. Ito, Thermodynamic bound on spectral perturbations, with applications to oscillations and relaxation dynamics, *Phys. Rev. Res.* **6**, 013082 (2024).
- [44] J. A. Fill, Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process, *Ann. Appl. Probab.* **1**, 62 (1991).
- [45] R. M. Gray *et al.*, Toeplitz and circulant matrices: A review, *FNT in Commun. Info. Theory* **2**, 155 (2006).
- [46] N. Shiraishi, S. Ito, K. Kawaguchi, and T. Sagawa, Role of measurement-feedback separation in autonomous Maxwell's demons, *New J. Phys.* **17**, 045012 (2015).
- [47] N. Shiraishi, T. Matsumoto, and T. Sagawa, Measurement-feedback formalism meets information reservoirs, *New J. Phys.* **18**, 013044 (2016).
- [48] K. Takahashi, K. Fujii, Y. Hino, and H. Hayakawa, Nonadiabatic control of geometric pumping, *Phys. Rev. Lett.* **124**, 150602 (2020).
- [49] J. M. Horowitz and J. L. England, Information-theoretic bound on the entropy production to maintain a classical nonequilibrium distribution using ancillary control, *Entropy* **19**, 333 (2017).
- [50] J. M. Horowitz, K. Zhou, and J. L. England, Minimum energetic cost to maintain a target nonequilibrium state, *Phys. Rev. E* **95**, 042102 (2017).
- [51] J. M. Parrondo, J. M. Horowitz, and T. Sagawa, Thermodynamics of information, *Nat. Phys.* **11**, 131 (2015).
- [52] T. Sagawa and M. Ueda, Minimal energy cost for thermodynamic information processing: Measurement and information erasure, *Phys. Rev. Lett.* **102**, 250602 (2009).
- [53] D. Abreu and U. Seifert, Thermodynamics of genuine nonequilibrium states under feedback control, *Phys. Rev. Lett.* **108**, 030601 (2012).
- [54] F. J. Cao and M. Feito, Thermodynamics of feedback controlled systems, *Phys. Rev. E* **79**, 041118 (2009).
- [55] S. Ito and T. Sagawa, Information thermodynamics on causal networks, *Phys. Rev. Lett.* **111**, 180603 (2013).
- [56] D. Hartich, A. C. Barato, and U. Seifert, Stochastic thermodynamics of bipartite systems: Transfer entropy inequalities and a Maxwell's demon interpretation, *J. Stat. Mech.* (2014) P02016.
- [57] A. C. Barato, D. Hartich, and U. Seifert, Efficiency of cellular information processing, *New J. Phys.* **16**, 103024 (2014).
- [58] P. Sartori, L. Granger, C. F. Lee, and J. M. Horowitz, Thermodynamic costs of information processing in sensory adaptation, *PLoS Comput. Biol.* **10**, e1003974 (2014).
- [59] J. Barham, A dynamical model of the meaning of information, *Biosystems* **38**, 235 (1996).
- [60] A. Kolchinsky and D. H. Wolpert, Semantic information, autonomous agency and non-equilibrium statistical physics, *Interface Focus* **8**, 20180041 (2018).
- [61] P. Pietzonka, A. C. Barato, and U. Seifert, Universal bound on the efficiency of molecular motors, *J. Stat. Mech.* (2016) 124004.
- [62] A. I. Brown and D. A. Sivak, Allocating dissipation across a molecular machine cycle to maximize flux, *Proc. Natl. Acad. Sci. USA* **114**, 11057 (2017).
- [63] A. I. Brown and D. A. Sivak, Allocating and splitting free energy to maximize molecular machine flux, *J. Phys. Chem. B* **122**, 1387 (2018).
- [64] P. Pietzonka and U. Seifert, Universal trade-off between power, efficiency, and constancy in steady-State heat engines, *Phys. Rev. Lett.* **120**, 190602 (2018).
- [65] J. A. Wagoner and K. A. Dill, Mechanisms for achieving high speed and efficiency in biomolecular machines, *Proc. Natl. Acad. Sci. USA* **116**, 5902 (2019).
- [66] T. Schmiedl and U. Seifert, Efficiency of molecular motors at maximum power, *Europhys. Lett.* **83**, 30005 (2008).
- [67] D. A. Beard, S.-D. Liang, and H. Qian, Energy balance for analysis of complex metabolic networks, *Biophys. J.* **83**, 79 (2002).
- [68] K. J. Kauffman, P. Prakash, and J. S. Edwards, Advances in flux balance analysis, *Curr. Opin. Biotechnol.* **14**, 491 (2003).
- [69] M. Kschischo, A gentle introduction to the thermodynamics of biochemical stoichiometric networks in steady state, *Eur. Phys. J. Spec. Top.* **187**, 255 (2010).
- [70] R. M. Fleming, C. M. Maes, M. A. Saunders, Y. Ye, and B. Ø. Palsson, A variational principle for computing nonequilibrium fluxes and potentials in genome-scale biochemical networks, *J. Theor. Biol.* **292**, 71 (2012).
- [71] 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).
- [72] J. M. Horowitz and M. Esposito, Thermodynamics with continuous information flow, *Phys. Rev. X* **4**, 031015 (2014).