UC Irvine

Title

Maximal power output of a stochastic thermodynamic engine

Permalink

Authors

Fu, Rui

Taghvaei, Amirhossein

Chen, Yongxin

et al.

Publication Date

DOI

Copyright Information

Maximal power output of a stochastic thermodynamic engine

Rui Fu, Amirhossein Taghvaei, Yongxin Chen, Tryphon T. Georgiou

Abstract

Classical thermodynamics aimed to quantify the efficiency of thermodynamic engines, by bounding the maximal amount of mechanical energy produced, compared to the amount of heat required. While this was accomplished early on, by Carnot and Clausius, the more practical problem to quantify limits of power that can be delivered, remained elusive due to the fact that quasistatic processes require infinitely slow cycling, resulting in a vanishing power output. Recent insights, drawn from stochastic models, appear to bridge the gap between theory and practice in that they lead to physically meaningful expressions for the dissipation cost in operating a thermodynamic engine over a finite time window. Indeed, the problem to optimize power can be expressed as a stochastic control problem. Building on this framework of *stochastic thermodynamics* we derive bounds on the maximal power that can be drawn by cycling an overdamped ensemble of particles via a time-varying potential while alternating contact with heat baths of different temperature (T_c cold, and T_h hot). Specifically, assuming a suitable bound M on the spatial gradient of the controlling potential, we show that the maximal achievable power is bounded by $\frac{M}{8}(\frac{T_h}{T_c} - 1)$. Moreover, we show that this bound can be reached to within a factor of $(\frac{T_h}{T_c} - 1)/(\frac{T_h}{T_c} + 1)$ by operating the cyclic thermodynamic process with a quadratic potential.

Keywords: Non-equilibrium Thermodynamics, Optimal Transportation, Mean-field optimal control

I. Introduction

Thermodynamics is the branch of physics which is concerned with the relation between heat and other forms of energy. Historically, it was born of the quest to quantify the maximal efficiency of heat engines, i.e., the maximal ratio of the total work output over the total heat input to a thermodynamic system. This was accomplished in the celebrated work of Carnot [9], [7] where, assuming that transitions take place infinitely slowly (*quasi-static* operation), it was shown that the maximal efficiency possible is $\eta_C = 1 - T_c/T_h$ (*Carnot efficiency*), where T_h and T_c are the absolute temperatures of two heat reservoirs, hot and cold respectively, with which the heat engine makes contact with during phases of a periodic operation known as *Carnot cycle*.

Carnot's result provides the absolute theoretical limit for the efficiency of a heat engine, but provides no insight on the amount of power output that can be achieved. Specifically, in order to reach Carnot efficiency, the period of the Carnot cycle must tend to infinity, resulting in quasi-static operation with vanishing total power output. Whereas, to achieve finite power output in a thermodynamic process, this must take place in finite time, and thereby, away from *equilibrium* [10], [23], [15].

To this end, the framework of stochastic thermodynamics [33], [35], [34], [27], [16], [5] has been developed in recent years, to allow quantifying work in non-equilibrium thermodynamic transitions. It is rooted in probabilistic models in the form of stochastic differential equations to specify the behaviour of particles in a thermodynamic ensemble. Manipulation of the ensemble is effected by a confining potential that serves as a *control input*. This potential, together with a heat reservoir in contact, couples the ensemble to the environment. Work and heat being transferred can then be computed at the level of individual particles and averaged over the ensemble. Important goals of the theory have been to assess the amount of work needed for *bit-erasure in finite time* [36], [24] and hence computation, i.e., a finite-time Landauer bound, as well as assessing the efficiency of thermodynamic engines operating at maximal power.

The question of efficiency at maximal power was studied independently by Chambadal [11], Novikov [26] and Curzon and Ahlborn [14] based on a certain "endoreversible" assumption to reflect finite-time heat transfer. They derived the bound $\eta_{CA} = 1 - \sqrt{T_c/T_h} = 1 - \sqrt{1 - \eta_C}$, where the T_h and T_c designate temperatures of a hot and cold heat reservoir,

R. Fu, A. Taghvaei, and T. T. Georgiou are with the Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA; rfu2@uci.edu, ataghvae@uci.edu, tryphon@uci.edu

Y. Chen is with the School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30332; yongchen@gatech.edu

respectively, at maximal power estimated to be $k(\sqrt{T_h} - \sqrt{T_c})^2$, with k being the heat conductance. Subsequent works, most notably by Chen and Yan [12], based on differing sets of assumptions, arrived at different bounds. More recently Schmiedl and Seifert [32], sought to improve, and reconcile these earlier results within the framework of stochastic thermodynamics, albeit for thermodynamic ensembles transitioning between Gaussian distributions. It is fair to say that there is no consensus on the firmness of these expressions, and that they serve as a guide to actual performance of thermodynamic engines.

The present work focuses on maximizing power in general, relaxing the Gaussian assumption, within the context of stochastic thermodynamics [35], [34]. This is a *stochastic control problem*. Our analysis is based on an overdamped Langevin model for thermodynamic processes (with damping coefficient γ), and explores advantages and pitfalls of selecting arbitrary control input, i.e., confining potential, for steering thermodynamic ensembles through cyclic operation while alternating contact between available heat reservoirs. It is noted that without physically motivated constraints on the actuation potential, the power ouput can become unbounded. The salient feature of actuation (time-varying potential $U(t,x)$, with t denoting time and $x \in \mathbb{R}^d$ the spacial coordinate) that draws increasing amounts of power is its ability to drive the thermodynamic ensemble to a state of very low entropy. Indeed, the magnitude of the spatial gradient of the potential $\nabla_x U(t,x)$ plays a key role. Thus, it is reasonable on physical grounds to suitably constrain this mode of "control" actuation, that is responsible for energy exchange between the ensemble and the environment. The present work puts forth and motivates the bound[1] (equation (47))

$$\frac{1}{\gamma} \int_{\mathbb{R}^d} \|\nabla_x U(t,x)\|^2 \rho(t,x)\, \mathrm{d}x \leq M,$$

where ρ denotes the thermodynamic state, as a suitable such constraint, and under this assumption it is shown that a maximal amount of power output that can be extracted by cyclic operation of a Carnot-like engine is

$$\frac{M}{8}\left(\frac{T_h}{T_c} - 1\right)\left(\frac{\frac{T_h}{T_c} - 1}{\frac{T_h}{T_c} + 1}\right) \leq P_{\max} \leq \frac{M}{8}\left(\frac{T_h}{T_c} - 1\right).$$

That is, the upper bound $\frac{M}{8}\left(\frac{T_h}{T_c} - 1\right)$ on power output only depends on M and the temperature of the two heat baths[2]. Moreover, this bound can be attained within a factor of $\left(\frac{T_h}{T_c} - 1\right)/\left(\frac{T_h}{T_c} + 1\right)$, which depends only on the ratio of temperatures of the two heat baths as well.

The exposition proceeds as follows. Section II details the stochastic model thermodynamic ensembles and the heat/energy exchange mechanism. Section III explores a connection between the second law of thermodynamics and the Wasserstein geometry of optimal mass transport that underlies the mechanism of energy dissipation in thermodynamic transitions. Section IV returns to the concept of a cyclicly operated thermodynamic engine and expresses the optimal efficiency and power output as functions of the operating protocol (solution of a stochastic control problem that dictates the choice of control time-varying potential), temperature of heat reservoirs, timing of the cyclic operation, and thermodynamic states at the end of phases of the Carnot-like cycle. Section V contains the main results regarding seeking maximal power output. Specifically, Section V-A explains optimal scheduling times, Section V-B highlights questions that arise based on physical grounds for Gaussian thermodynamic states, Sections V-C and V-D discuss optimal thermodynamic states at the two ends of the Carnot-like cycle, and Sections V-E and V-F derive bounds on maximal achievable power with or without constraint on the controlling potential. A concluding remarks section recaps and points to future research directions and open problems.

II. Stochastic thermodynamic models

We begin by describing the basic model for a *thermodynamic ensemble* used in this work. This consists of a large collection of Brownian particles that interact with a *heat bath* in the form of a stochastic excitation and driven under the influence of an *external (time varying) potential* between end-point states. The dynamics of individual particles are expressed in the form of stochastic differential equations.

---

[1]Interestingly, this can also be expressed in information theoretic terms, as a bound on the Fisher information of thermodynamic states.

[2]In general power output is an *extensive* quantity, as it depends on the size of the thermodynamic ensemble/engine. However, in our treatment, the ensemble is described by a probability distribution (normalized). Hence, the bounds appear as "intensive."

A. Langevin dynamics

The (under-damped) Langevin equations

$$\mathrm{d}X_t = \frac{p_t}{m}\mathrm{d}t \tag{1a}$$

$$\mathrm{d}p_t = -\nabla_x U(t, X_t)\mathrm{d}t - \gamma\frac{p_t}{m}\mathrm{d}t + \sqrt{2\gamma k_B T(t)}\mathrm{d}B_t, \tag{1b}$$

represent a standard model for molecular systems interacting with a thermal environment. Throughout, $X_t \in \mathbb{R}^d$ denotes the location of a particle and p_t denotes its momentum at time t, $U(t, x)$ denotes a time-varying potential for $x \in \mathbb{R}^d$, m is the mass of the particle, γ is the viscosity coefficient, k_B is the Boltzmann constant, $T(t)$ denotes the temperature of the heat bath at time t, and B_t denotes a standard $\mathbb{R}^d$-valued Brownian motion.

In this paper, we consider only the case where inertial effects in the Langevin equation (1b) are negligible for the time resolution of interest. Specifically, when the temporal resolution $\Delta t \gg \frac{m}{\gamma}$, averaging out the fast variable p_t leads to the *over-damped Langevin equation*

$$\mathrm{d}X_t = -\frac{1}{\gamma}\nabla_x U(t, X_t)\mathrm{d}t + \sqrt{\frac{2k_B T(t)}{\gamma}}\mathrm{d}B_t. \tag{2}$$

Intuitively, the over-damped Langevin equation is obtained from (1b) by setting $\mathrm{d}p_t = 0$ and replacing $\frac{p_t}{m}\mathrm{d}t = \mathrm{d}X_t$. For a more detailed explanation see [35, page 20].

Thus, we view $\{X_t\}_{t \geq 0}$ as a diffusion process. The state of the thermodynamic ensemble is identified with the probability density of X_t, denoted by $\rho(t, x)$, which satisfies the Fokker-Planck equation

$$\frac{\partial\rho}{\partial t} - \frac{1}{\gamma}\nabla_x \cdot [(\nabla_x U + k_B T\nabla_x \log\rho)\rho] = 0. \tag{3}$$

B. Heat, work, and the first law

The evolution of the thermodynamic ensemble under the influence of the time-varying thermal environment and the time-varying potential $U(t, x)$, leads to exchange of heat and work, respectively. Heat and work can be defined at the level of a single particle as explained below.

The energy exchange between an individual particle and the thermal environment represents *heat*. This exchange is effected by forces exerted on the particle due to viscosity $(-\gamma\frac{\mathrm{d}X_t}{\mathrm{d}t})$ and due to the random thermal excitation $(\sqrt{2\gamma k_B T}\frac{\mathrm{d}B_t}{\mathrm{d}t})$. It can be formally expressed as the product of force and displacement, in Stratonovich form, as $(-\gamma\frac{\mathrm{d}X_t}{\mathrm{d}t} + \sqrt{2\gamma k_B T}\frac{\mathrm{d}B_t}{\mathrm{d}t}) \circ \mathrm{d}X_t$, which, using (2), leads to

$$đQ = \nabla_x U(t, X_t) \circ \mathrm{d}X_t \tag{4}$$

$$= \nabla_x U(t, X_t) \circ \left(-\frac{1}{\gamma}\nabla_x U(t, X_t)\mathrm{d}t + \sqrt{\frac{2k_B T(t)}{\gamma}}\mathrm{d}B_t\right)$$

$$= -\frac{1}{\gamma}\|\nabla_x U(t, X_t)\|^2\mathrm{d}t + \Delta_x U(t, X_t)\frac{k_B T(t)}{\gamma}\mathrm{d}t$$

$$+ \nabla_x U(t, X_t)\sqrt{\frac{2k_B T(t)}{\gamma}}\mathrm{d}B_t,$$

where the last step includes the correction due to changing into the Itô form. Note that we use $đ$ to emphasize that $đQ$ is not a perfect differential in that the integral $\int đQ$ depends on the path taken and not just end-point conditions.

The energy exchange between an individual particle and the external potential represents *work*. Specifically, the work transferred to the particle by a change in the actuating potential is

$$đW = \frac{\partial U}{\partial t}(t, X_t)\mathrm{d}t. \tag{5}$$

Naturally, the *first law of thermodynamics*,

$$\mathrm{d}U(t, X_t) = đQ + đW$$

Definition	Notation	Units
time	t	s
position of particle	X_t	m
Boltzmann constant	k_B	Nm
damping coefficient	γ	Ns/m
potential	$U(t,x)$	Nm
temperature	T	$^o\mathrm{K}$
Brownian motion	B_t	$s^{\frac{1}{2}}$
density in $\mathbb{R}^d$	$\rho(t,x)$	m^{-d}
velocity field in $\mathbb{R}^d$	$v(t,x)$	m/s
Wasserstein length	$\mathrm{length}_{W_2}(\cdot)$	m
entropy	$\mathcal{S}(\rho)$	Nm
work (particle/ensemble)	$W, \mathcal{W}$	Nm
heat (particle/ensemble)	$Q, \mathcal{Q}$	Nm
energy (particle/ensemble)	$U, \mathcal{E}$	Nm
free energy	$\mathcal{F}$	Nm
bound in (47)	M	Nm/s
power	P	Nm/s

TABLE I
SYMBOLS AND CORRESPONDING UNITS

holds, since the internal energy is simply the value of the potential.

Accordingly, for a thermodynamic ensemble at a state $\rho(t,x)$, the heat and work differentials are expressed as

$$d\!Q = \left[\int_{\mathbb{R}^d} \left(-\frac{1}{\gamma} \|\nabla_x U\|^2 + \Delta_x U \frac{k_B T}{\gamma} \right) \rho \, dx \right] dt \tag{6a}$$

$$d\!\mathcal{W} = \left[\int_{\mathbb{R}^d} \frac{\partial U}{\partial t} \rho \, dx \right] dt, \tag{6b}$$

leading to the first law for the ensemble

$$d\mathcal{E}(\rho, U) = d\!Q + d\!\mathcal{W},$$

where the internal energy is

$$\mathcal{E}(\rho, U) = \int_{\mathbb{R}^d} U \rho \, dx, \tag{6c}$$

and depends on ρ, U, whereas $\mathcal{Q}, \mathcal{W}$ depend on the path.

C. Summary notation

As usual, $\mathbb{R}^d$ denotes the d-dimensional Euclidean space, for $d \in \mathbb{N}$, with $\langle x, y \rangle$ and $\|x\| = \sqrt{\langle x, x \rangle}$ denoting the respective inner product and norm, for $x, y \in \mathbb{R}^d$. For two vector fields $\nabla_x \phi_1, \nabla_x \phi_2$, we denote $\langle \nabla_x \phi_1, \nabla_x \phi_2 \rangle_\rho = \int_{\mathbb{R}^d} \langle \nabla_x \phi_1, \nabla_x \phi_2 \rangle \rho dx$, and $\|\nabla_x \phi\|_\rho^2 := \langle \nabla_x \phi, \nabla_x \phi \rangle_\rho$. The Gaussian distribution with mean m and covariance Σ is denoted by $N(m, \Sigma)$. For convenience we provide Table II-C of the various quantities, including the corresponding units in SI format: Newton (N), seconds (s), meter (m), absolute temperature in degrees Kelvin ($^o\mathrm{K}$).

III. THE SECOND LAW, DISSIPATION, AND WASSERSTEIN GEOMETRY

We now discuss the *second law of thermodynamics* in the context of an ensemble of particles obeying over-damped Langevin dynamics (2), assuming that the temperature of the heat bath remains constant, i.e., $T(t) = T$. The classical formulation amounts to the inequality

$$\mathcal{W} - \Delta \mathcal{F} \geq 0 \tag{7}$$

where $\mathcal{W}$ is the work transferred to the ensemble over a time interval (t_i, t_f), namely,

$$\mathcal{W} = \int_{t_i}^{t_f} d\!\mathcal{W},$$

and $\Delta \mathcal{F}$ is the change in free energy

$$\mathcal{F}(\rho, U) = \mathcal{E}(\rho, U) - T\mathcal{S}(\rho) \tag{8}$$

between the two end-point states[3]. Here,

$$\mathcal{S}(\rho) = -k_B \int_{\mathbb{R}^d} \log(\rho)\, \rho\, \mathrm{d}x \tag{9}$$

denotes the entropy of the state ρ, and U denotes the potential as earlier.

Inequality (7) becomes equality for quasi-static (reversible) thermodynamic transitions. In general, for irreversible transitions, the gap in (7) quantifies dissipation. Interestingly, alternative formulations that shed light into irreversible transitions have recently been discovered. A most remarkable identity was discovered by Jarzynski in the late 90's [21] to hold for irreversible thermodynamic transitions between work and free energy, in the form,

$$\mathbb{E}\left\{e^{-\beta W}\right\} - e^{-\beta \Delta \mathcal{F}} = 0,$$

or, equivalently,

$$-\beta^{-1}\log \mathbb{E}\left\{e^{-\beta W}\right\} - \Delta \mathcal{F} = 0,$$

where $\beta^{-1} = k_B T$ and $\Delta \mathcal{F}$ denotes difference between equilibrium free energy, while the expectation is taken over the probability law on paths.

While the Jarzynski relation establishes equality between the above functional of the work and free energy differences, it does not allow quantifying the actual expected work performed on the ensemble. An alternative identity that quantifies explicitly the gap in (7) holds for irreversible thermodynamic transitions. This identity is

$$\mathcal{W} - \Delta \mathcal{F} = \underbrace{\gamma \int_{t_i}^{t_f} \left\|\nabla_x \phi(t,\cdot)\right\|^2_{\rho(t,\cdot)} \mathrm{d}t}_{\text{dissipation}}, \tag{10a}$$

where

$$\left\|\nabla_x \phi\right\|^2_\rho = \int_{\mathbb{R}^d} \left\|\nabla_x \phi\right\|^2 \rho\, \mathrm{d}x, \tag{10b}$$

$$\frac{\partial \rho}{\partial t} + \nabla_x \cdot (\rho \underbrace{\nabla_x \phi}_{v}) = 0. \tag{10c}$$

The dissipation has the form of an *action integral*, since $v = \nabla_x \phi$ is a velocity field that specifies the drift of the ensemble and, thereby, $\left\|\nabla_x \phi\right\|^2_\rho$ is the *averaged kinetic energy of the ensemble*. It is also seen that the dissipation depends solely on the *time-parametrized* path

$$\rho_{[t_i,t_f]} := \{\rho(t,\cdot) \mid t \in [t_i, t_f]\}.$$

Specifically, given any "tangent" $\frac{\partial \rho}{\partial t} = \delta$ at any point $\rho(t,\cdot)$ along the path $\rho_{[t_i,t_f]}$, the *Poisson equation* (10c) can be solved for $\phi(t,x)$ with $x \in \mathbb{R}^d$, giving rise to the time-varying, spatially irrotational vector field $v = \nabla_x \phi$ that transitions the thermodynamic system from the starting state $\rho(t,\cdot)$ to $\rho + \delta \mathrm{d}t$ over the time window $[t, t + \mathrm{d}t]$.

Interestingly, this recipe of identifying tangent perturbations δ (i.e., functions $\delta(x)$ such that $\int_{\mathbb{R}^d} \delta \mathrm{d}x = 0$) with an irrotational field $v = \nabla_x \phi$ via the Poisson equation, instills on the space of probability distributions a Riemannian-like structure via the quadratic form

$$\int_{\mathbb{R}^d} \langle \nabla_x \phi_1, \nabla_x \phi_2 \rangle \rho\, \mathrm{d}x. \tag{11}$$

Herein, probability distributions are assumed to have finite second-order moment and be positive. The space of probability distributions (or, measures [37], with finite second-order moments) with this inner-product structure is known as the Wasserstein manifold $\mathcal{P}_2(\mathbb{R}^d)$.

The geodesic distance, i.e., the minimum of

$$\text{length}_{W_2}\left(\rho_{[t_i,t_f]}\right) = \int_{t_i}^{t_f} \left\|\nabla_x \phi(t,\cdot)\right\|_{\rho(t,\cdot)} \mathrm{d}t, \tag{12}$$

---

[3]The free energy is the amount of energy that can be delivered at temperature T with fixed potential U. A rather revealing re-write of the free energy is as the relative entropy (KL-divergence) between the current state ρ and the Gibbs distribution $\rho_{\text{Gibbs}}(x) = e^{-\beta U(x)}/Z$, where $\beta = 1/k_B T$ and the normalizing factor $Z = \int_{\mathbb{R}^d} e^{-\beta U(x)} \mathrm{d}x$ is the partition function. Specifically, $\mathcal{F}(\rho, U) = \beta^{-1} \int_{\mathbb{R}^d} \log(\frac{\rho(x)}{\rho_{\text{Gibbs}}(x)}) \rho(x) \mathrm{d}x - \beta^{-1} \log(Z)$.

over paths $\rho_{[t_i,t_f]}$ connecting the two end-points ρ_{t_i}, ρ_{t_f}, turns out to be precisely the Wasserstein metric

$$W_2(\rho_{t_i}, \rho_{t_f}) := \sqrt{\inf_{\pi \in \Pi(\rho_{t_i}, \rho_{t_f})} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^2 \mathrm{d}\pi(x, y)} \tag{13}$$

of the Monge-Kantorovich theory of Optimal Mass Transport for quadratic transportation cost. Here, $\Pi(\rho_{t_i}, \rho_{t_f})$ denotes the probability measures on the product space $\mathbb{R}^d \times \mathbb{R}^d$ with $\rho_{t_i}(x)$ and $\rho_{t_f}(y)$ as marginals; see [37, Chapter 8]) and especially [1] for a detailed exposition of the differential structure of $\mathcal{P}_2(\mathbb{R}^d)$.

Returning to (10), provided the averaged kinetic energy (10b) during thermodynamic transitions remains constant over time (which can be ensured by a suitable scaling of time of the path $\rho_{[t_i,t_f]}$),

$$\mathcal{W} - \Delta\mathcal{F} = \frac{\gamma}{t_f - t_i} \left(\mathrm{length}_{W_2}\left(\rho_{[t_i,t_f]}\right)\right)^2, \tag{14}$$

while in general, the right hand side of (14) serves as a lower bound. If in addition to constancy of (10b) the path is selected as a W_2-geodesic, then

$$\mathcal{W} - \Delta\mathcal{F} = \frac{\gamma}{t_f - t_i} W_2(t_i, t_f)^2, \tag{15}$$

which quantifies the maximum amount of work that can be drawn by transitioning between specified end-point thermodynamic states. We recap the key points in the following statement.

Theorem 1. *Consider the overdamped model* (2) *for thermodynamic transitions between states* ρ_{t_i}, ρ_{t_f}, *under constant temperature* T *and a time-varying potential* U. *The following hold:*
i) In general,

$$\mathcal{W} - \Delta\mathcal{F} \geq \frac{\gamma}{t_f - t_i} \left(\mathrm{length}_{W_2}\left(\rho_{[t_i,t_f]}\right)\right)^2. \tag{16}$$

ii) Relation (16) *holds as the equality in* (14) *for a suitable time-reparametrization of the path* $\rho_{[t_i,t_f]}$ *of the thermodynamic ensemble, effected by a suitable choice of potential.*
iii) There is a unique path $\rho_{[t_i,t_f]}$ *(W_2-geodesic) for the thermodynamic transition that attains minimal dissipation and, in this case,* (15) *holds.*

Proof. We first derive (10): consider

$$\begin{aligned}
\frac{\mathrm{d}\mathcal{F}}{\mathrm{d}t}(\rho, U) &= \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}(\rho, U) - T\frac{\mathrm{d}}{\mathrm{d}t}S(\rho) \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d} U\rho\,\mathrm{d}x + k_B T \frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d} \rho\log\rho\,\mathrm{d}x \\
&= \int_{\mathbb{R}^d}\left(\frac{\partial U}{\partial t}\rho + U\frac{\partial\rho}{\partial t} + k_B T\frac{\partial\rho}{\partial t}\log\rho\right)\mathrm{d}x \\
&= \int_{\mathbb{R}^d}\frac{\partial U}{\partial t}\rho\,\mathrm{d}x + \int_{\mathbb{R}^d}(U + k_B T\log\rho)\frac{\partial\rho}{\partial t}\mathrm{d}x.
\end{aligned}$$

Using the Fokker-Planck equation (3), the second term

$$\begin{aligned}
&\int_{\mathbb{R}^d}(U + k_B T\log\rho)\frac{1}{\gamma}\nabla_x\cdot[(\nabla_x U + k_B T\nabla_x\log\rho)\rho]\,\mathrm{d}x \\
&= -\frac{1}{\gamma}\int_{\mathbb{R}^d}\|\nabla_x U + k_B T\nabla_x\log\rho\|^2\rho\,\mathrm{d}x \\
&= -\gamma\int_{\mathbb{R}^d}\|v\|^2\rho\,\mathrm{d}x,
\end{aligned}$$

where the first equality follows using integration by parts (under standard assumptions on the decay rate of ρ at infinity), while the second equality is a re-write using

$$v := -\frac{1}{\gamma}(\nabla_x U + k_B T\nabla_x\log\rho). \tag{17}$$

Thus,

$$\frac{\mathrm{d}\mathcal{F}}{\mathrm{d}t}(\rho, U) = \int_{\mathbb{R}^d}\frac{\partial U}{\partial t}\rho\,\mathrm{d}x - \gamma\int_{\mathbb{R}^d}\|v\|^2\rho\,\mathrm{d}x.$$

Integrating over $[t_i, t_f]$ yields

$$\Delta \mathcal{F} = \mathcal{W} - \gamma \int_{t_i}^{t_f} \int_{\mathbb{R}^d} \|v\|^2 \rho \, \mathrm{d}x \, \mathrm{d}t, \tag{18}$$

where v is the gradient of $\phi = -\frac{1}{\gamma}(U + k_B T \log \rho)$ and satisfies the continuity equation (10c) as claimed. This establishes (10).

Statements i) and ii) follow from the fact that the W_2-length of the path $\rho_{[t_i, t_f]}$ (i.e., as a curve in $\mathcal{P}_2$), is given by (12). Specifically, provided $\int_{\mathbb{R}^d} \|v\|^2 \rho \, \mathrm{d}x = \alpha^2$ remains constant along the path (i.e., for $t \in [t_i, t_f]$),

$$\alpha = \frac{1}{t_f - t_i} \text{length}_{W_2} \left(\rho_{[t_i, t_f]} \right)$$

and (14) follows from (18). If on the other hand the kinetic energy varies with time, then the path $\rho(t, \cdot)$, time-reparametrized by

$$\tilde{t}(t) := \frac{\text{length}_{W_2} \left(\rho_{[t_i, t]} \right)}{\text{length}_{W_2} \left(\rho_{[t_i, t_f]} \right)} (t_f - t_i) + t_i$$

will be traversed via a velocity field

$$\tilde{v}(\tilde{t}(t)) = \frac{v(t)}{\|v(t)\|_\rho} \frac{\text{length}_{W_2} \left(\rho_{[t_i, t_f]} \right)}{t_f - t_i}.$$

Knowing $\tilde{v}$, a new potential $\tilde{U}$ can be computed so that $\tilde{v}(\cdot, \tilde{t}) = \nabla_x \tilde{U}(\cdot, \tilde{t}) + k_B T \nabla_x \rho(\cdot, \tilde{t})$.

Finally, statement iii) follows by taking $\rho_{[t_i, t_f]}$ to be a geodesic. $\qquad\square$

Remark 1. *Early work by Jordan etal. [22], pointing out that the gradient flow of the free energy in W_2 is the Fokker-Planck equation, set the stage for understanding the role of the Wasserstein geometry in quantifying dissipation. This was recognized in [3], [2], [34] and more recently developed in [13], [17].*

IV. Cyclic operation of engines

We consider two types of thermodynamic transitions, isothermal and adiabatic. The first corresponds to a situation where the system remains in contact with a heat bath of constant temperature T while a time-varying potential steers its thermodynamic state $\rho(t, .)$ from an *initial* $\rho(t_i, \cdot)$ to a *final* $\rho(t_f, \cdot)$. The adiabatic transition amounts to abrupt changes in both, the temperature of the heat bath as well as the shape of the potential, that are fast enough not to have any measurable effect on the state $\rho(t, .)$ and, as a consequence, to the entropy of the ensemble. We evaluate next the energy and work budgets in the correspoinding actuation protocols.

A. Isothermal transition

We consider transition between states ρ_i and ρ_f for the ensemble modeled by (2), over a time interval $[t_i, t_f]$, under the time-varying potential $U(t, X_t)$ and in contact with a heat bath of temperature T. Using the relationship (10a) between work, free energy, and the dissipation, and the first law, we have the following identity relating thermodynamic quantities in isothermal transitions

$$\mathcal{W} = \Delta \mathcal{E} - T \Delta \mathcal{S} + \mathcal{W}_{\text{irr}} \tag{19a}$$
$$\mathcal{Q} = T \Delta \mathcal{S} - \mathcal{W}_{\text{irr}} \tag{19b}$$

with the *irreversible* $\mathcal{W}_{\text{irr}}$ that represents dissipation attaining its minimal value

$$\frac{\gamma}{t_f - t_i} W_2(\rho_{t_i}, \rho_{t_f})^2 \tag{19c}$$

by the choice of actuation $\nabla_x U(t, \cdot)$ in (17) with v the optimal velocity field minimizing dissipation in (10) (item iii) in Theorem 1).

It is important to note that the minimizing v can be obtained by solving a convex reformulation of (10) in terms of the density $\rho(t, \cdot)$ and the momentum field $\mathbf{p}(t, \cdot) = v(t, \cdot)\rho(t, \cdot)$, in the form

$$\min_{\mathbf{p}(t,\cdot),\rho(t,\cdot)} \int_{t_i}^{t_f} \int_{\mathbb{R}^d} \frac{\|\mathbf{p}\|^2}{\rho} \mathrm{d}x\mathrm{d}t \tag{20a}$$

$$\text{subject to } \frac{\partial \rho}{\partial t} + \nabla_x \cdot \mathbf{p} = 0 \tag{20b}$$

$$\text{and } \rho(t_i, \cdot), \rho(t_f, \cdot) \text{ specified.} \tag{20c}$$

Then, $v = \mathbf{p}/\rho$, see [4, Section 4] and [37, p. 241].

B. Adiabatic transition

We now consider transition between ρ_i and ρ_f for the ensemble modeled by (2), over a time interval $[t_i, t_f]$, under abrupt changes in the potential $U(t, X_t)$ and the temperature T of the heat bath.

The transition takes place over an infinitesimally short time interval about time t (with t^-/t^+ indicating the left/right limits, respectively). Thus, the temperature T of the heat bath jumps between values $T(t^-)$ and $T(t^+)$ while, at the same time, the controlling potential switches from $U(t^-, \cdot)$ to $U(t^+, \cdot)$.

The energy budget of the transition no longer contains irreversible losses, as the right hand side of (10a) vanishes. Moreoverm, the entropy of the ensemble remains constant because $\rho(t^+, \cdot) = \rho(t^-, \cdot)$. Thus, the work input into the system equal to change in internal energy,

$$\mathcal{W} = \int_{\mathbb{R}^d} (U(t^+, x) - U(t^-, x))\rho(t, x)\mathrm{d}x = \Delta\mathcal{E}, \tag{21a}$$

and therefore no heat transfer takes place, and therefore,

$$\mathcal{Q} = 0. \tag{21b}$$

C. Finite-time Carnot cycle

We are now in position to consider a complete *Carnot-like thermodynamic cycle* where the ensemble is steered between two states ρ_a and ρ_b during isothermal expansion (from ρ_a to ρ_b) and contraction (from ρ_b to ρ_a) phases, separated by adiabatic transitions. Periodic operation about such a scheduling is sought as a means to extract work from a heat bath. A schematic in Figure 1 depicts the phases of the cyclic operation. These four phases are described in detail next.

1) Isothermal process in temperature T_h ("hot"): The first step is an isothermal expansion over the time interval $(0, t_1)$ in contact with a heat bath of temperature $T = T_h$. Change in the potential steers the ensemble from a starting state ρ_a to a terminal state ρ_b. As in (19),

$$\mathcal{W}^{(1)} = \Delta\mathcal{E}^{(1)} - T_h\Delta\mathcal{S}^{(1)} + \mathcal{W}_{\mathrm{irr}}^{(1)} \tag{22a}$$

$$\mathcal{Q}^{(1)} = T_h\Delta\mathcal{S}^{(1)} - \mathcal{W}_{\mathrm{irr}}^{(1)} \tag{22b}$$

where the superscript enumerates the phase in the cycle, and the minimal work loss $\mathcal{W}_{\mathrm{irr}}^{(1)}$ depends only on the end-point states as it equals

$$\mathcal{W}_{\mathrm{irr}}^{(1)} = \frac{\gamma}{t_1}W_2(\rho_a, \rho_b)^2. \tag{22c}$$

2) Adiabatic process: The second phase of the cycle is an adiabatic transition at time $t = t_1$, over an infinitesimal interval (of duration "$t_2 = 0$"), bringing the ensemble in contact with a heat bath of temperature T_c ("cold"). As in (21),

$$\mathcal{W}^{(2)} = \Delta\mathcal{E}^{(2)} \tag{23a}$$
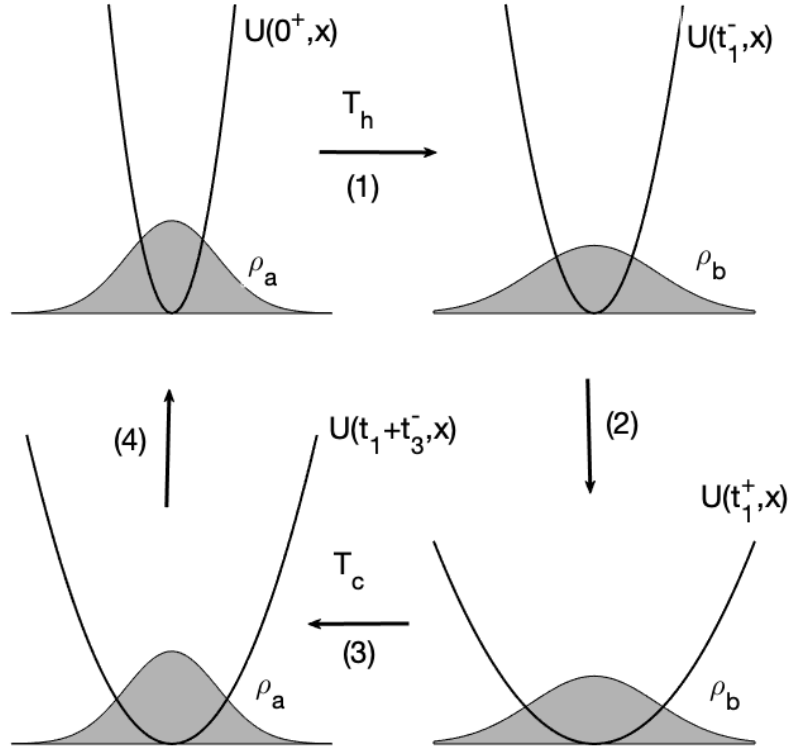
$$\mathcal{Q}^{(2)} = 0 \tag{23b}$$

Fig. 1. Carnot-like cycle of a stochastic model for a heat engine (with $d = 1$): the operation cycles clockwise through two isothermal transitions (1) and (3), and two adiabatic transitions (2) and (4). During the isothermal transitions having duration t_1 and t_3, the ensemble is in contact with a "hot" reservoir of temperature T_h, and a "cold" one of temperature T_c, respectively. The adiabatic transitions are considered to be instantaneous, i.e., $t_2 = t_4 = 0$. The marginal densities are ρ_a and ρ_b.

while the state remains at ρ_b.

3) Isothermal process in temperature T_c ("cold"): The third step is an Isothermal contraction over the time interval $(t_1, t_1 + t_3)$ while in contact with a heat bath of temperature T_c. Actuation in the form of the time-varying potential causes the state of the ensemble to return to ρ_a back from starting at ρ_b. Once again, as in (19),

$$\mathcal{W}^{(3)} = \Delta\mathcal{E}^{(3)} - T_c \Delta\mathcal{S}^{(3)} + \mathcal{W}_{\text{irr}}^{(3)} \tag{24a}$$

$$\mathcal{Q}^{(3)} = T_c \Delta\mathcal{S}^{(3)} - \mathcal{W}_{\text{irr}}^{(3)} \tag{24b}$$

$$\mathcal{W}_{\text{irr}}^{(3)} = \frac{\gamma}{t_3} W_2(\rho_a, \rho_b)^2. \tag{24c}$$

4) Adiabatic process: Finally, an adiabatic transition over an interval of infinitesimal duration ("$t_4 = 0$") returns the ensemble to be in contact with a heat reservoir of temperature T_h for a total period of the cycle $t_{\text{period}} = t_1 + t_3$. The state of the ensemble remains at ρ_a, to begin the cycle again. As before, in (21),

$$\mathcal{W}^{(4)} = \Delta\mathcal{E}^{(4)} \tag{25a}$$

$$\mathcal{Q}^{(4)} = 0 \tag{25b}$$

D. Thermodynamic efficiency & power delivered

For a cyclic process the total change in internal energy

$$\sum_{i=1}^{4} \Delta \mathcal{E}^{(i)} = 0.$$

On the other hand, the entropy doesn't change during the adiabatic transitions

$$\Delta \mathcal{S}^{(i)} = 0, \text{ for } i = 2, 4,$$

while, since it depends only on the end-point states ρ_a, ρ_b,

$$\Delta S^{(1)} = -\Delta S^{(3)} = \mathcal{S}(\rho_b) - \mathcal{S}(\rho_a) =: \Delta \mathcal{S}.$$

As a result, the total work output (negative of the work input) is

$$-\mathcal{W} = -\left(\sum_{i=1}^{4} \Delta \mathcal{E}^{(i)} - \sum_{i=1}^{4} T_i \Delta \mathcal{S}^{(i)} + \sum_{i=1}^{4} \mathcal{W}_{irr}^{(i)} \right) \tag{26}$$
$$= (T_h - T_c) \Delta \mathcal{S} - \mathcal{W}_{irr}^{(1)} - \mathcal{W}_{irr}^{(3)}.$$

Thus, assuming optimality of the choice of the potential to minimize $\mathcal{W}_{irr}$ in each transition, we conclude that the total work output possible is

$$-\mathcal{W} = (T_h - T_c) \Delta \mathcal{S} - \gamma \left(\frac{1}{t_1} + \frac{1}{t_3} \right) W_2(\rho_a, \rho_b)^2. \tag{27}$$

Since $T_h > T_c$, naturally, a necessary condition for positive work output is that

$$\Delta \mathcal{S} := \mathcal{S}(\rho_b) - \mathcal{S}(\rho_a) > 0$$

which dictates that phase 1 is an isothermal expansion and phase 3, an isothermal contraction.[4]

The thermodynamic efficiency of an engine is the ratio of work extracted over the heat dissipated,

$$\eta = \frac{-\mathcal{W}}{\mathcal{Q}_h} \tag{28}$$

where the heat input during isothermal expansion, from (22b), is

$$\mathcal{Q}_h = \Delta \mathcal{Q}^{(1)} = T_h \Delta \mathcal{S} - \mathcal{W}_{irr}.$$

Once again assuming optimality ($\mathcal{W}_{irr} = \frac{\gamma}{t_1} W_2(\rho_a, \rho_b)^2$), the bound on the efficiency is seen to be

$$\eta = \frac{(T_h - T_c)\Delta \mathcal{S} - \gamma(\frac{1}{t_1} + \frac{1}{t_3})W_2(\rho_a, \rho_b)^2}{T_h \Delta \mathcal{S} - \gamma \frac{1}{t_1} W_2(\rho_a, \rho_b)^2} \tag{29}$$

When the period of the cyclic process tends to infinity (and hence, $t_1, t_3 \to \infty$), tends to the Carnot limit for quasistatic (infinitely slow) transitions

$$\eta_C = 1 - \frac{T_c}{T_h}. $$

Periodic operation, over a finite period $t_1 + t_3$ (since $t_2 = t_4 = 0$), delivers

$$P = -\mathcal{W}/(t_1 + t_3)$$
$$= \frac{(T_h - T_c)\Delta \mathcal{S} - \gamma(\frac{1}{t_1} + \frac{1}{t_3})W_2(\rho_a, \rho_b)^2}{t_1 + t_3} \tag{30}$$

units of power. Note that the power output is zero when Carnot efficiency is achieved, because the total duration $t_1 + t_3 \to \infty$. In the sequel, we focus on assessing bounds on available power.

---

[4]The opposite would be true if we sought to operate the cycle for refrigeration purposes.

V. Fundamental limits to power

Our main interest is in assessing the maximal amount of power that can be drawn by a thermodynamic engine operating between heat baths with temperatures T_h and $T_c < T_h$, i.e.,"hot" and "cold", respectively. In the present work we draw conclusions based on the basic model in (2) via analysis of the thermodynamic cycle that was presented in Section IV.

Consider the expression in (30) for the power that can be drawn via a cyclic operation as discussed. Preparation of the ensemble, and actuation during the cycle, allow a number of choices. Specifically, the power depends on the period $t_1 + t_3$, the times of the two isothermal phases t_1, t_3 individually, as well as the end-point states (distributions) ρ_a, ρ_b. The latter choice impacts both, the Wasserstein distance $W_2(\rho_a, \rho_b)$ as well as the change in entropy $\Delta\mathcal{S}$. We will explore systematically the various options.

A. Optimizing the time scheduling

Optimizing the maximal power delivered during cyclic operation

$$P = \frac{1}{t_1 + t_3}(T_h - T_c)\Delta\mathcal{S} - \frac{\gamma}{t_1 t_3}W_2(\rho_a, \rho_b)^2,$$

with respect to choices for t_1, t_3, with $W_2(\rho_a, \rho_b)$, T_h, T_c and $\Delta\mathcal{S}$ kept fixed, gives that

$$t_1 = t_3 = \frac{4\gamma W_2(\rho_a, \rho_b)^2}{(T_h - T_c)\Delta\mathcal{S}}, \tag{31}$$

and therefore that the period for the cycle is

$$
\begin{aligned}
t_{\text{cycle}} &:= t_1 + t_3 \\
&= \frac{8\gamma W_2(\rho_a, \rho_b)^2}{(T_h - T_c)\Delta\mathcal{S}}.
\end{aligned}
\tag{32}
$$

If instead we specify the period of the cycle t_{cycle}, and optimize with respect to the breakdown between t_1 and t_3, we once again obtain that the durations of the two phases are equal

$$t_1 = t_3 = \frac{t_{\text{cycle}}}{2}. \tag{33}$$

Remark 2 (Efficiency at maximum power). *The thermodynamic efficiency* (28) *of the engine, when it is operating at optimal transition times* (31) *that maximize the power, is equal to*

$$\eta_{SS} = \frac{2(T_h - T_c)}{3T_h + T_c} = \frac{\eta_C}{2 - \frac{\eta_C}{2}} \tag{34}$$

The result (34) *first appeared in [33], for special setting when the two marginal distributions ρ_a and ρ_b are Gaussians, and the potential $U(t, x)$ is a quadratic function of x. Our derivation verifies the result* (34) *in a general setting.*

Using the expression (32), the total power delivered

$$P = \frac{(T_h - T_c)^2}{16\gamma}\left(\frac{\Delta\mathcal{S}}{W_2(\rho_a, \rho_b)}\right)^2. \tag{35}$$

But as we will see in Section V-B, optimizing the power for ρ_a, ρ_b leads to the non-physical conclusion of a vanishingly small t_{cycle} .

B. The caveat of optimal t_{cycle}: Gaussian states ρ_a, ρ_b

The case where the two marginal distributions/states are Gaussian allows for closed-form expressions for $\Delta\mathcal{S}$ and their Wasserstein distance. Indeed, if ρ_a, ρ_b are Gaussian distributions with zero mean and variances Σ_a, Σ_b, respectively, then

$$W_2^2(\rho_a, \rho_b) = \text{trace}\left(\Sigma_a + \Sigma_b - 2(\Sigma_a^{1/2}\Sigma_b\Sigma_a^{1/2})^{1/2}\right), \tag{36a}$$

$$
\begin{aligned}
\Delta\mathcal{S} &= \mathcal{S}(\rho_b) - \mathcal{S}(\rho_a) \\
&= \frac{1}{2}k_B \log\det(\Sigma_b\Sigma_a^{-1}).
\end{aligned}
\tag{36b}
$$

Evidently, these allow deriving explicit expressions for the available power in terms of the respective variances.

Specializing to the case of scalar processes with σ_i $(i \in \{a, b\})$ the corresponding standard deviation, i.e., $\Sigma_i = \sigma_i^2$, and period t_{cycle} for the thermodynamic cycle as in (32), we obtain that the maximal power available, as a function of σ_a and σ_b, is given by

$$P(\sigma_a, \sigma_b) = \frac{k_B^2 (T_h - T_c)^2}{16\gamma} \left(\frac{\log \frac{\sigma_b}{\sigma_a}}{\sigma_b - \sigma_a} \right)^2. \tag{37}$$

The corresponding heat uptake from the hot reservoir and the work extracted during one cycle are computed as

$$\mathcal{Q}^{(1)} = \mathcal{Q}_h = \frac{1}{4} k_B (3T_h + T_c) \log \frac{\sigma_b}{\sigma_a}$$

and

$$-\mathcal{W} = \frac{1}{2} k_B (T_h - T_c) \log \frac{\sigma_b}{\sigma_a},$$

respectively.

The maximum of the power $P(\sigma_a, \sigma_b)$ over either σ_a, or σ_b, takes place when

$$\sigma_a = \sigma_b.$$

But at this limiting condition, although

$$\max_{\sigma_b} P(\sigma_a, \sigma_b) = \frac{k_B^2 (T_h - T_c)^2}{16\gamma \sigma_a^2} \tag{38a}$$

and the rate with which heat is drawn is

$$\lim_{\sigma_b \to \sigma_a} \frac{\mathcal{Q}_h}{t_{\text{cycle}}} = \frac{k_B^2 (3T_h + T_c)(T_h - T_c)}{32\gamma \sigma_a^2},$$

the limiting values of $-\Delta\mathcal{W}$, $\mathcal{Q}_h$ over a cycle vanish, as does the period t_{cycle} of the cycle. Thus we are led to a non-physical situation of a vanishingly small period for the thermodynamic cycle.

A similar issue in the context of power in quantum engines is brought up in [18]. In the setting herein, in addition, it is seen that taking

$$\sigma_a \to 0$$

and operating with a vanishingly small period for the cycle, leads to infinite power. Once again, bringing up a non-practical situation that is questionable on physical grounds. In the sequel we focus on t_{cycle} being finite.

C. Optimizing the thermodynamic state ρ_b

Henceforth we fix the period t_{cycle} as well as the duration of the isothermal phases according to (33). The power delivered, as a function of the ρ_i's $(i \in \{a, b\})$, is

$$\frac{(T_h - T_c)}{t_{\text{cycle}}} (\mathcal{S}(\rho_b) - \mathcal{S}(\rho_a)) - \frac{4\gamma}{t_{\text{cycle}}^2} W_2(\rho_a, \rho_b)^2. \tag{39}$$

We now consider the problem to maximize power over choice of ρ_b, with ρ_a specified. This problem reduces to finding a suitable minimizer of

$$\min_{\rho_b} \{ W_2(\rho_a, \rho_b)^2 - h\mathcal{S}(\rho_b), \} \tag{40}$$

for

$$h = \frac{t_{\text{cycle}}(T_h - T_c)}{4\gamma}.$$

Throughout we assume that states have finite second-order moments. As noted earlier, the space of probability distributions (measures, in general) with finite second-order moments $\mathcal{P}_2(\mathbb{R}^d)$ is metrized by the Wasserstein metric $W_2(\cdot, \cdot)$ and, as can easily be verified, the expression

$$W_2(\rho_a, \rho_b)^2 - h\mathcal{S}(\rho_b) \tag{41}$$

is strictly convex, which leads to the following statement.

Proposition 2. *Assuming that T_h, T_c as well as t_{cycle} and an initial state $\rho_a \in \mathcal{P}_2(\mathbb{R}^d)$ are specified, there exists a unique minimizer ρ_b of (40).*

Proof. Equation (40) is similar to one step in the so-called JKO-scheme (also, proximal projection) that displays the heat equation as the gradient flow of the Shannon entropy [22]. While $W_2(\rho_a, \rho_b)^2 - h\mathcal{S}(\rho_b)$ is strictly convex, it is not automatically bounded from below. Thus, a rather extensive and technical argument is needed to show existence and uniqueness of a minimizer. This is detailed in [22, Proposition 4.1]. $\qquad\square$

We conclude this section with two statements. The first establishes implicit conditions for optimality of ρ_b, in maximizing the expression in (39) (equivalently, minimizing (41)). For ease of referencing we view the expression in (39) as a function of ρ_b, namely,

$$f(\rho_b) := \frac{(T_h - T_c)}{t_{\text{cycle}}}(\mathcal{S}(\rho_b) - \mathcal{S}(\rho_a)) - \frac{4\gamma}{t_{\text{cycle}}^2}W_2(\rho_a, \rho_b)^2. \tag{42}$$

The following lemma provides stationarity conditions for $f(\rho_b)$ that, albeit, are implicit in that they involve the optimal transport map from ρ_a and ρ_b that minimizes quadratic transportation cost [37, Ch. 5].

The theory of optimal transport provides that, since ρ_a, ρ_b are densities (as opposed to measures), the support of Π in (13) coincides with the graph of a map

$$\Psi : \mathbb{R}^d \to \mathbb{R}^d \,:\, x \mapsto y,$$

which in fact is the gradient of a convex function ψ on $\mathbb{R}^d$ [37, Ch. 5], i.e., $\Psi = \nabla \psi$. This is the (unique) optimal transport map for the so-called *Monge problem* to minimize $\int_{\mathbb{R}^d} \|x - \Psi(x)\|^2 \rho_a(x)\mathrm{d}x$ over all maps that transfer mass from ρ_a to ρ_b. The transference of the "mass" distribution ρ_a into ρ_b is indicated by

$$\nabla \psi \sharp \rho_a = \rho_b,$$

which is a compact notation for the change of variables formula

$$\det(\nabla^2 \psi(x))\rho_b(\nabla \psi(x)) = \rho_a(x).$$

We first highlight stationarity conditions that characterize the minimizer of $f(\cdot)$ in (42).

Lemma 3. *Consider two probability densities ρ_a, ρ_b^* in $\mathcal{P}_2(\mathbb{R}^d)$, where ρ_b^* is the unique maximizer of $f(\rho_b)$, and let $\nabla \phi$, for a convex function ψ on $\mathbb{R}^d$, be such that $\nabla \psi \sharp \rho_a = \rho_b^*$. The following (stationarity) condition holds*

$$k_B(T_h - T_c)\nabla \log \rho_b^*(y) - \frac{8\gamma}{t_{\text{cycle}}}\left((\nabla \psi)^{-1} - \mathrm{Id}\right)(y) = 0, \tag{43}$$

where Id *denotes the identity map.*

Proof. The proof is given in Appendix A. $\qquad\square$

The lemma, which is of independent interest, is used in the proof of the following proposition which concludes the section. The proposition states that, for scalar distributions for simplicity, if ρ_a is Gaussian, then so is ρ_b. As a consequence the optimal actuation protocol is based on a time-varying potential $U(t, x)$ that is quadratic in x.

Proposition 4. *If ρ_a is a one-dimensional Gaussian distribution with zero mean and variance σ_a^2, then ρ_b^* is also Gaussian with zero mean and variance σ_b^2, where*

$$\sigma_b = \frac{1 + \sqrt{1 + c}}{2}\sigma_a, \tag{44}$$

and $c = \frac{k_B(T_h - T_c)t_{\text{cycle}}}{2\gamma \sigma_a^2}$.

Proof. The proof is given in Appendix B. $\qquad\square$

Remark 3. *In earlier works, it is commonly assumed that the marginal distributions ρ_a, ρ_b are Gaussian and the potential function $U(t, x)$ is quadratic in x. Proposition 4 justifies this assumption to some extent: if ρ_a is specified to be Gaussian, the optimal ρ_b and the optimal potential function that achieve the maximum power, are Gaussian and*

quadratic, respectively. However, as we will see in Section V-D, if instead ρ_b is specified as Gaussian distribution, the optimal ρ_a is not Gaussian. Gaussian distributions turn out instead to be local minimizers of the power under certain conditions (see discussion following Remark 4).

D. Optimizing the thermodynamic state ρ_a

We now consider the dependence of the maximal power on ρ_a, i.e., on the thermodynamic state at which the ensemble begins its expansive phase. As we will see, the situation is not symmetric to the conclusions drawn in Section V-C with regard to ρ_b and, without further assumptions, an optimal ρ_a does not exist. Interestingly, on closer inspection, the source of this conundrum is the unreasonably high demands on the magnitude of $\nabla_x U$ for the controlling potential $U(t, x)$. The insights gained lead to the framework for maximal power in the follow up section.

For simplicity, and without any loss of generality for the purposes of this section, we assume that ρ_b is specified to be a zero-mean Gaussian distribution with standard deviation σ_b. In view of (39), a choice of ρ_a that is close to a Dirac delta distribution allows arbitrarily large negative values for the entropy, i.e., $\mathcal{S}(\rho_a) \simeq -\infty$, and hence infinite power.

Thus, it is natural to impose a lower bound on the entropy of ρ_a, or simply fix

$$-\infty < s_a = \mathcal{S}(\rho_a) < \mathcal{S}(\rho_b).$$

But in this case, and once more in view of (39), maximal power would be drawn by minimizing $W_2(\rho_a, \rho_b)$ over probability densities ρ_a with entropy s_a. We claim that

$$\inf_{\rho_a}\{W_2(\rho_a, \rho_b) \mid \mathcal{S}(\rho_a) = s_a > -\infty\} = 0. \tag{45}$$

To see this note that

$$\inf_{\rho_a} W_2(\rho_a, \rho_b) = 0$$

by taking ρ_a to approximate an increasingly fine train of suitably scaled Dirac deltas, i.e.,

$$\rho_a(x) \approx \sum_{i \in \mathbb{Z}} \rho_i \delta_{x_i}(x)$$

where $\rho_i = \int_{x_i}^{x_{i+1}} \rho_b(x)\mathrm{d}x$ and x_i $(i \in \mathbb{Z})$ equispaced. The latter is a singular distribution which, however, can be approximated arbitrarily closely in W_2 by a probability density with any given entropy. Such a density can be produced by approximating Dirac deltas by a piecewise constant function with finite support.

The optimization problem (45) is inherently related to the continuity of the entropy functional with respect to the Wasserstein distance. For a rigorous treatment of the problem, see [29], where it is shown that unless certain regularity assumptions are in place for ρ_a and ρ_b, the infimum in (45) is zero.

Remark 4 (Gaussian is not optimal for ρ_a). *The preceeding arguments show that a Gaussian distribution is not the optimal choice for ρ_a with respect to maximizing power, even when ρ_b is Gaussian, unless additional constraints are introduced.*

Since the Gaussian distribution maximizes entropy when mean and variance are specified, it is natural to explore constraints on the mean and variance of ρ_a for the purposes of maximizing power. Without loss of generality, the mean can be assumed to be zero and the variance specified to be $\sigma_a^2 < \sigma_b^2$. First-order and second order optimality analysis for the power output (39), at $\rho_a = N(0, \sigma_a^2)$ is carried out. It turns out that, although $N(0, \sigma_a^2)$ satisfies the first-order optimality condition, it does not satisfy the second-order optimality condition. In fact, $N(0, \sigma_a^2)$ is a local minimizer when $\sigma_a < \sigma_b < k_B(T_h - T_c)t_{\text{cycle}}/(8\gamma\sigma_a)$. The analysis is given in Appendix C, and aims to highlight that the conjecture that a Gaussian ρ_a is optimal fails.

E. Maximum power with arbitrary potential

In this section, we show that the power output of a thermodynamic engine, under any choice of potential $U(t,x)$ cannot exceed a bound that involves the Fisher information of the marginal state ρ_a. The Fisher information is defined as

$$I(\rho) := \int_{\rho>0} \frac{\|\nabla\rho\|^2}{\rho} \, dx.$$

Proposition 5. *Under the standing assumptions on the Carnot-like cycle, the power output (39), is bounded by*

$$P \leq \frac{k_B^2(T_h - T_c)^2}{16\gamma} I(\rho_a). \tag{46}$$

Proof. This is based on the following HWI inequality (see [38], [19] for details),

$$\mathcal{S}(\rho_b) - \mathcal{S}(\rho_a) \leq k_B W_2(\rho_a, \rho_b) \sqrt{I(\rho_a)},$$

Using the formula for power (39), we have

$$
\begin{aligned}
P &= \frac{(T_h - T_c)\Delta\mathcal{S}}{t_{\text{cycle}}} - \frac{4\gamma}{t_{\text{cycle}}^2} W_2(\rho_a, \rho_b)^2 \\
&\leq \frac{(T_h - T_c)\Delta\mathcal{S}}{t_{\text{cycle}}} - \frac{4\gamma}{t_{\text{cycle}}^2} \frac{\Delta\mathcal{S}^2}{k_B^2 I(\rho_a)} \\
&= -\frac{4\gamma}{t_{\text{cycle}}^2} \frac{1}{k_B^2 I(\rho_a)} \left(\Delta S - \frac{t_{\text{cycle}} k_B^2 (T_h - T_c)}{8\gamma} I(\rho_a)\right)^2 \\
&\quad + \frac{k_B^2(T_h - T_c)^2}{16\gamma} I(\rho_a) \\
&\leq \frac{k_B^2(T_h - T_c)^2}{16\gamma} I(\rho_a),
\end{aligned}
$$

concluding the bound (46).   □

We point out that the bound (46) is achieved when t_{cycle} is given by its optimal value (32) and in the limit as $\rho_b \to \rho_a$. In particular, suppose ρ_a and ρ_b are Gaussian distributions $N(0, \sigma_a^2)$ and $N(0, \sigma_b^2)$, respectively, and that the cycle period t_{cycle} is equal to the optimal value (32), then as $\sigma_b \to \sigma_a$ the power output is given by (38a), which is equal to the bound (46), because $I(\rho_a) = \frac{1}{\sigma_a^2}$.

F. Maximum power under constrained potential

While a lower bound on $\mathcal{S}(\rho_a)$ readily implies an upper bound on the available power, achieving such a bound in general requires a cyclic operation involving an irregular and complicated potential function $U(t,x)$ to bring back the ensemble to ρ_a at end of each cycle. It is unreasonable to expect technological solutions to such demands, and therefore, a constraint on the complexity of the potential function seems meaningful. To this end, we propose the constraint

$$\frac{1}{\gamma} \int_{\mathbb{R}^d} \|\nabla_x U(t,x)\|^2 \rho(t,x) \, dx \leq M \tag{47}$$

for all $t \in (0, t_{\text{cycle}})$. Thus, we analyze the maximum power (39) that can be extracted from a thermodynamic engine, under the constraint (47).

Theorem 6. *Consider a thermodynamic ensemble, undergoing a Carnot cycle as described in Section IV, governed with the over-damped Langevin equation (2). Then, the maximum power P that can be extracted from the cycle, over all marginal probability distributions ρ_a and ρ_b, the cycle period t_{cycle}, and all potential functions $U(t,x)$ that respect the bound (47), satisfies*

$$\frac{M}{8}\left(\frac{T_h}{T_c} - 1\right) \frac{\frac{T_h}{T_c} - 1}{\frac{T_h}{T_c} + 1} \leq P_{\max} \leq \frac{M}{8}\left(\frac{T_h}{T_c} - 1\right) \tag{48}$$

Proof. The proof for the upper-bound follows from bounding the entropy difference $S(\rho_b) - S(\rho_a)$ under the constraint (47). During the isothermal transition in contact with the cold bath with temperature T_c,

$$S(\rho_b) - S(\rho_a) = S(\rho(\frac{t_{\text{cycle}}}{2}, \cdot)) - S(\rho(t_{\text{cycle}}, \cdot))$$

$$= -\int_{\frac{t_{\text{cycle}}}{2}}^{t_{\text{cycle}}} \frac{\mathrm{d}}{\mathrm{dt}} S(\rho(t, \cdot)) \, \mathrm{dt}$$

$$= k_B \int_{\frac{t_{\text{cycle}}}{2}}^{t_{\text{cycle}}} \int_{\mathbb{R}^d} \log \rho(t, x) \frac{\partial \rho}{\partial t}(t, x) \, \mathrm{dx} \, \mathrm{dt}$$

$$= \frac{-k_B}{\gamma} \int_{\frac{t_{\text{cycle}}}{2}}^{t_{\text{cycle}}} \int_{\mathbb{R}^d} \langle \nabla_x \log \rho, \nabla_x U + k_B T_c \nabla_x \log \rho \rangle \rho \, \mathrm{dx} \mathrm{dt}$$

$$= \frac{-k_B}{\gamma} \int_{\frac{t_{\text{cycle}}}{2}}^{t_{\text{cycle}}} (\langle \nabla_x \log \rho, \nabla_x U \rangle_\rho + k_B T_c \|\nabla_x \log \rho\|_\rho^2) \mathrm{dt},$$

where with a slight abuse of notation we use $\langle \nabla_x f, \nabla_x g \rangle_\rho$ to also denote $\int_{\mathbb{R}^d} \langle \nabla_x f, \nabla_x g \rangle \rho \mathrm{dx}$. By the Cauchy-Schwartz inequality and constraint (47),

$$-\langle \nabla_x \log \rho, \nabla_x U \rangle_\rho \leq \|\nabla_x U\|_\rho \|\nabla_x \log \rho\|_\rho$$

$$\leq \sqrt{\gamma M} \|\nabla_x \log \rho\|_\rho.$$

Hence,

$$S(\rho_b) - S(\rho_a)$$

$$\leq \frac{k_B}{\gamma} \int_{\frac{t_{\text{cycle}}}{2}}^{t_{\text{cycle}}} \left(\sqrt{\gamma M} \|\nabla_x \log \rho\|_\rho - k_B T_c \|\nabla_x \log \rho\|_\rho^2 \right) \mathrm{dt}$$

$$\leq \frac{k_B}{\gamma} \int_{\frac{t_{\text{cycle}}}{2}}^{t_{\text{cycle}}} \frac{\gamma M}{4 k_B T_C} \mathrm{dt} = \frac{M}{8 T_C} t_{\text{cycle}}.$$

This concludes the bound $\Delta S \leq \frac{M}{T_C} \frac{t_{\text{cycle}}}{8}$ on the entropy difference, which yields to upper-bound on the power output:

$$P \leq \frac{(T_h - T_c)}{t_{\text{cycle}}} \Delta S - \frac{1}{t_{\text{cycle}}} W_{\text{irr}} \leq \frac{M(T_h - T_c)}{8 T_C} \tag{49}$$

where $W_{\text{irr}} \geq 0$ is used.

Next, we prove the lower-bound by describing a setting so that the power is equal to the lower bound. Assume the marginal distributions ρ_a and ρ_b are Gaussian $N(0, \sigma_a^2)$ and $N(0, \sigma_b^2)$ respectively, and the potential function $U(t, x) = \frac{1}{2} a_t x^2$ is a quadratic function. In this setting, the exact power output is equal to

$$P = \frac{1}{t_{\text{cycle}}} k_B (T_h - T_c) \log(\frac{\sigma_b}{\sigma_a})$$

$$- \frac{1}{\gamma t_{\text{cycle}}} \int_0^{t_{\text{cycle}}} (a_t - \frac{k_B T}{\sigma_t^2})^2 \sigma_t^2 \mathrm{dt}$$

with update law for the variance given by the Lyapunov equation:

$$\frac{\mathrm{d}\sigma_t^2}{\mathrm{dt}} = -2(\frac{a_t}{\gamma} - \frac{k_B T}{\gamma \sigma_t^2}) \sigma_t^2$$

with the constraint (47) given by

$$\frac{1}{\gamma} a_t^2 \sigma_t^2 \leq M$$

Then, in the limit as $t_{\text{cycle}} \to 0$, and $\sigma_a = \sigma_b = \sigma$, the power output is equal to

$$P = k_B (T_h - T_c) \frac{\lambda}{2} - \gamma \lambda^2 \sigma^2 \tag{50}$$

with the constraint

$$|\gamma \lambda + \frac{k_B T_C}{\sigma^2}| \leq \frac{\sqrt{\gamma M}}{\sigma}, \tag{51}$$

where we introduced a new variable $\lambda = \frac{a}{\gamma} - \frac{k_B T_c}{\gamma \sigma^2}$. It is shown in Appendix D, that the maximum of the expression (50) over all values of λ and σ that satisfy the constraint (51), is equal to

$$\frac{M}{8}\left(\frac{T_h}{T_c} - 1\right)\frac{\frac{T_h}{T_c} - 1}{\frac{T_h}{T_c} + 1}$$

concluding the lower-bound.   □

This final result is universal as it does not depend on the choice of ρ_a and ρ_b, unlike (46). Moreover, the bounds in this final result are especially appealing in that it the depend on the ratio T_h/T_c of the absolute temperatures of the two heat baths.

Remark 5. *It is noted that the upper bound in* (48) *on achievable power under the constraint* (47) *does not depend on* t_{cycle}, *whereas our construction for achieving the lower bound ensures that the bound is approached as* $t_{\text{cycle}} \to 0$.

Remark 6. *In the proof of Theorem 6, an operating point has been constructed to ensure that power equal the lower bound in* (48) *can be achieved. The parameters are given in equation* (65) *in the Appendix. For this operating point, which corresponds to maximal power constrained by* (47), *the efficiency turns out to be*

$$\eta = \frac{T_h - T_c}{T_h + T_c}.$$

It is interesting to note that

$$\eta_{SS} \leq \eta_{CA} \leq \eta \leq \eta_C,$$

where η_{SS} *is the efficiency obtained by Schmiedl and Seifert given in* (34), $\eta_{CA} = 1 - \sqrt{T_c/T_h}$ *is the Curzon-Ahlborn efficiency, and* $\eta_C = 1 - T_c/T_h$ *is the Carnot Efficiency. Furthermore,* η_{CA}, η *and* η_C *tend to 1 as* $T_c \to 0$, *while* $\eta_{SS} \to 2/3$.

VI. Concluding remarks

The present work focused on quantifying the maximal power that can be drawn by a Carnot-like heat engine operating by alternating contact with two heat reservoirs and modeled by stochastic overdamped Langevin dynamics driven by the time dependent potential. The framework that the work is based on is that of Stochastic Thermodynamics [33], [35], [34], [27], [16], which allows quantifying energy and heat exchange by individual particles in a thermodynamic ensemble, to be subsequently averaged, so as to quantify performance of the thermodynamic process as a whole. A physically reasonable bound is derived, which is shown to be reached within a specified factor, both depending on the ratio T_h/T_c of the absolute temperatures of the two heat baths, hot and cold, respectively. The present work is quite distinct from earlier results, within a similar framework, which is however restricted to Gaussian states. Conditions that suggest non-physical conclusions are highlighted, and a suitable constraint on the controlling potential is brought forth that underlies our analysis.

In the past few decades, there have been several attempts to quantify efficiency mainly, but also power, of thermodynamic processes operating in Carnot-like manner. It is fair to say that there has been neither a consensus on the type of assumptions that have been used by previous authors, and thereby, nor full consistency of the results. This is to be expected, since finite-period operation and finite-time thermodynamic transitions require substance/engine dependent assumptions to capture the complexity of heat transfer in non-equilibrium states. Thus, estimated bounds may never reach the "universality" of the celebrated Carnot efficiency. They are expected to provide physical insights and guidelines for engineering design. Thus, it will be imperative that these estimates be subject to experimental testing. The notable feature of our conclusions as compared to earlier works is that the expressions we derive are given in the form of ratio of absolute temperatures–a physically suggestive feature.

The present work follows a long line of contributions within the control field to draw links between thermodynamics and control, see e.g., [6], [28], [25], [31], [30], [39]. More recently, important insights have linked the Wasserstein distance of optimal mass transport, which itself is a solution to a stochastic control problem, to the dissipation mechanism in stochastic thermodynamics [3], [2], [34], [13], [17]. Indeed, the Wasserstein metric takes the form of an action integral and arises naturally in the energy balance of thermodynamic transitions. This fact has been explored and developed for the overdamped Langevin dynamics studied herein. Whether similar conclusions can be drawn for underdamped Langevin dynamics remains an open research direction at present. Furthermore, much work remains to reconcile and

compare alternative viewpoints and models for thermodynamic processes including those based on the Boltzmann equation.

Besides the potentially intrinsic value of the analysis and bounds that have been derived, it is hoped that the control-theoretic aspect of the problem to optimize Carnot-like cycling of thermodynamic process has been sufficiently highlighted, and that this work will serve to raise attention on this important and foundational topic to the control community.

<div align="center">APPENDIX</div>

A. Proof of Lemma 3

Consider an arbitrary smooth vector field with bounded support $\xi \in C_0^\infty(\mathbb{R}^d, \mathbb{R}^d)$. Let $\Psi_s : \mathbb{R}^d \to \mathbb{R}^d$ be the map generated by ξ defined according to

$$\frac{\partial}{\partial s}\Psi_s(x) = \xi(\Psi_s(x)), \quad \Psi_0 = \mathrm{Id},$$

for $x \in \mathbb{R}^d$ and $s \in \mathbb{R}^d$. Then, define

$$\rho_s := \Psi_s \sharp \rho_b^*.$$

We claim that

$$\lim_{s \to 0} \frac{1}{s}(f(\rho_s) - f(\rho_b^*)) \geq \int \langle D_f(x), \xi(x) \rangle \rho_b^*(x)\mathrm{d}x, \tag{52}$$

where, for $\Delta T := T_h - T_c$,

$$D_f(x) = -\frac{k_B \Delta T}{t_{\text{cycle}}}\nabla \log(\rho_b^*(x)) + \frac{8\gamma}{t_{\text{cycle}}^2}(\nabla \psi^{-1}(x) - x).$$

Assuming the claim is true (to be shown shortly), then, because ρ_b^* is the maximizer, $f(\rho_s) \leq f(\rho_b^*)$. Therefore

$$\int \langle D_f(x), \xi(x) \rangle \rho_b^*(x)\mathrm{d}x \leq \lim_{s \to 0} \frac{f(\rho_s) - f(\rho_b^*)}{s} \leq 0.$$

Hence, by symmetry $\xi \to -\xi$,

$$\int \langle D_f(x), \xi(x) \rangle \rho_b^*(x)\mathrm{d}x = 0. \tag{53}$$

This is true for all vector fields $\xi \in C_0^\infty(\mathbb{R}^d, \mathbb{R}^d)$. As a result, $D_f(x) = 0$, concluding (43) and the lemma.

It remains to prove (52). By definition, the difference $f(\rho_s) - f(\rho_b^*)$ is

$$f(\rho_s) - f(\rho_b^*) = \frac{\Delta T}{t_{\text{cycle}}}(\mathcal{S}(\rho_s) - \mathcal{S}(\rho_b^*))$$
$$- \frac{4\gamma}{t_{\text{cycle}}^2}(W_2(\rho_a, \rho_s)^2 - W_2(\rho_a, \rho_b^*)^2).$$

The entropy term

$$\mathcal{S}(\rho_s) = -k_B \int \log(\rho_s(x))\rho_s(x)\mathrm{d}x$$
$$= -k_B \int \log(\rho_s(\Psi_s(x)))\rho_b^*(x)\mathrm{d}x$$
$$= -k_B \int \log(\frac{\rho_b^*((x))}{\det(\nabla \Psi_s(x))})\rho_b^*(x)\mathrm{d}x$$
$$= \mathcal{S}(\rho_b^*) + k_B \int \log(\det(\nabla \Psi_s(x)))\rho_b^*(x)\mathrm{d}x.$$

Therefore

$$\lim_{s \to 0} \frac{1}{s}(S(\rho_s) - S(\rho_b^*))$$

$$= \lim_{s \to 0} \frac{k_B}{s} \int \log(\det(\nabla \Psi_s(x))) \rho_b^*(x) \mathrm{d}x$$

$$= k_B \int \nabla \cdot \xi(x) \rho_b^*(x) \mathrm{d}x$$

$$= -k_B \int \langle \xi(x), \nabla \log(\rho_b^*(x)) \rangle \rho_b^*(x) \mathrm{d}x.$$

The Wasserstein term

$$W_2(\rho_a, \rho_s)^2 - W_2(\rho_a, \rho_b^*)^2$$

$$\leq \int \|\nabla \psi^{-1}(x) - \Psi_s(x)\|^2 \rho_b^*(x) \mathrm{d}x$$

$$- \int \|\nabla \psi^{-1}(x) - x\|^2 \rho_b^*(x) \mathrm{d}x$$

$$= \int \langle x - \Psi_s(x), 2\nabla \psi^{-1}(x) - x - \Psi_s(x) \rangle \rho_b^*(x) \mathrm{d}x.$$

Therefore

$$\lim_{s \to 0} \frac{1}{s} \left[W_2(\rho_a, \rho_s)^2 - W_2(\rho_a, \rho_b^*)^2 \right]$$

$$\leq -2 \int \langle \xi(x), \nabla \psi^{-1}(x) - x \rangle \rho_b^*(x) \mathrm{d}x.$$

Using the two expressions, the one for derivative of the entropy and the other for the Wasserstein distance, the claim follows.

B. Proof of Proposition 4

According to Proposition 2, the maximizer is unique. Therefore, it is sufficient to show that the Gaussian distribution $N(0, \sigma_b^2)$, where σ_b^2 is given by (44), satisfies the optimality condition (43). When ρ_a, ρ_b^* are Gaussian, $\nabla \psi^{-1}(y) = \frac{\sigma_a}{\sigma_b} y$. Hence, the optimality condition reads

$$\frac{k_B t_{\mathrm{cycle}} \Delta T}{8\gamma} \nabla \log \rho_b^*(y) - y + \nabla \psi^{-1}(y)$$

$$= \frac{k_B t_{\mathrm{cycle}} \Delta T}{8\gamma} \frac{y}{\sigma_b^2} - (1 - \frac{\sigma_a}{\sigma_b})y$$

$$= (\frac{k_B t_{\mathrm{cycle}} \Delta T}{8\gamma \sigma_b^2} - 1 + \frac{\sigma_a}{\sigma_b})y = 0, \quad \forall y \in \mathbb{R},$$

which is satisfied when σ_b is according to (44).

C. Proof of statements following Remark 4

Let $\mathcal{A}_{0,\sigma^2}$ denote the set of absolutely continuous distributions with mean 0 and variance σ^2 :

$$\mathcal{A}_{0,\sigma^2} := \left\{ \rho \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d); \right.$$

$$\left. \int x \rho(x) \mathrm{d}x = 0, \int x^2 \rho(x) \mathrm{d}x = \sigma^2 \right\},$$

and consider the functional

$$g(\rho_a) = -\frac{T_h - T_c}{t_{\mathrm{cycle}}} S(\rho_a) - \frac{4\gamma}{t_{\mathrm{cycle}}^2} W_2(\rho_a, \rho_b)^2 \tag{54}$$

that represents the portion of power given in (39) that depends on ρ_a.

(i) We first show that

$$\max_{\rho_a \in \mathcal{A}_{0,\sigma_a^2}} g(\rho_a)$$

is unbounded, and hence that the maximizer does not exist. Consider a sequence of density functions $\{\mu_n\}_{n\in\mathbb{N}}$ according to

$$\mu_n = \frac{1}{2}\mu_n^{(1)} + \frac{1}{2}\mu_n^{(2)},$$

where

$$\mu_n^{(1)} = N(\sqrt{\sigma_a^2 - \frac{1}{n^2}}, \frac{1}{n^2})$$

$$\mu_n^{(2)} = N(-\sqrt{\sigma_a^2 - \frac{1}{n^2}}, \frac{1}{n^2})$$

(55)

It is easy to verify that $\mu_n \in \mathcal{A}_{0,\sigma_a^2}$. The goal is to show that $\lim_{n\to\infty} g(\mu_n) = \infty$. The entropy $\mathcal{S}(\mu_n)$ is bounded from above as follows,

$$\mathcal{S}(\mu_n) \leq \frac{1}{2}(k_B \log(2) + \mathcal{S}(\mu_n^{(1)})) + \frac{1}{2}(k_B \log(2) + \mathcal{S}(\mu_n^{(2)}))$$

$$= k_B \log 2 + \frac{k_B}{2} \log(2\pi e \frac{1}{n^2})$$

$$= k_B \log 2\sqrt{2\pi e} - k_B \log n.$$

The first inequality follows from a respective bound on the entropy of Gaussian mixtures [20, Theorem 3]. The Wasserstein distance $W_2(\mu_n, \rho_b)^2$ is bounded by

$$W_2(\mu_n, \rho_b)^2 \leq \frac{1}{2}W_2(\mu_n^{(1)}, \rho_b)^2 + \frac{1}{2}W_2(\mu_n^{(2)}, \rho_b)^2$$

$$= W_2(\mu_n^{(1)}, \rho_b)^2$$

$$= (\sqrt{\sigma_a^2 - \frac{1}{n^2}} - 0)^2 + (\frac{1}{n} - \sigma_b)^2$$

$$= \sigma_a^2 + \sigma_b^2 - 2\frac{\sigma_b}{n},$$

(56)

where the convexity of the functional $W_2^2(\cdot, \rho_b)$ is used [8, Eq.(2.12)]. Combining the two bounds for the entropy and Wasserstein distance yields

$$g(\mu_n) = -\frac{\Delta T}{t_{\text{cycle}}}\mathcal{S}(\mu_n) - \frac{4}{t_{\text{cycle}}^2}W_2(\mu_n, \rho_b)^2$$

$$\geq \frac{k_B \Delta T}{t_{\text{cycle}}}\left(-\log 2\sqrt{2\pi e} + \log n\right)$$

$$- \frac{4}{t_{\text{cycle}}^2}\left(\sigma_a^2 + \sigma_b^2 - 2\frac{\sigma_b}{n}\right).$$

Taking the limit $n \to \infty$ proves $\lim_{n\to\infty} g(\mu_n)$ is unbounded, and hence that there is no maximizer. Next, in (ii) and (iii) we show that the Gaussian distribution is instead a local minimizer, under certain conditions, and hence that it is the opposite of distributions that we seek.

(ii) We now perform first-order optimality analysis for the problem

$$\max_{\rho_a \in \mathcal{A}_{0,\sigma_a^2}} g(\rho_a)$$

at $\rho_a = N(0, \sigma_a^2)$. Consider a smooth vector field with bounded support $\xi \in C_0^\infty(\mathbb{R}, \mathbb{R})$ such that

$$\int_{\mathbb{R}^d} \xi(x)\rho_a(x)dx = 0.$$

(57)

Then, as before, define the flow $\Psi_s : \mathbb{R} \to \mathbb{R}$ generated by ξ according to

$$\frac{\partial}{\partial s}\Psi_s(x) = \xi(\Psi_s(x)), \quad \Psi_0 = \text{Id},$$

for $x \in \mathbb{R}$ and $s \in \mathbb{R}$. Now define

$$\tilde{\rho}_s := \Psi_s \sharp \rho_a.$$

Because of (57), the mean of the distribution $\tilde{\rho}_s$ remains constant at 0. However, the variance changes from σ_a^2, and $\tilde{\rho}_s$ is not inside $\mathcal{A}_{0,\sigma_a^2}$.

In order to keep the variance constant at σ_a^2, we project $\tilde{\rho}_s$ into $\mathcal{A}_{0,\sigma_a^2}$ by setting

$$\rho_s = G_s \sharp \tilde{\rho}_s,$$

where $G_s(x) := r(s)x$ and

$$r(s) = \frac{\sigma_a}{\sqrt{\int_{\mathbb{R}^d} |y|^2 \tilde{\rho}_s(y) \mathrm{d}y}} = \frac{\sigma_a}{\sqrt{\int_{\mathbb{R}^d} |\Psi_s(x)|^2 \rho_a(x) \mathrm{d}x}}.$$

By definition of G_s, the variance of ρ_s is equal to σ_a^2, hence $\rho_s \in \mathcal{A}_{0,\sigma_a^2}$.

Following the same procedure as in the proof of Lemma 3, the first-order optimality condition is

$$\int \langle D_g(x), v(x) \rangle \rho_a(x) \mathrm{d}x = 0, \tag{58}$$

where

$$\begin{aligned} v(x) &= \frac{\mathrm{d}}{\mathrm{d}s} G_s(\Psi_s(x))|_{s=0} \\ &= \xi(x) - \frac{x}{\sigma_a^2} \int \xi(z) z \rho_a(z) \mathrm{d}z, \end{aligned} \tag{59}$$

and

$$D_g(x) = \frac{k_B \Delta T}{t_{\text{cycle}}} \nabla \log(\rho_a^*(x)) + \frac{8\gamma}{t_{\text{cycle}}^2} (\nabla \psi(x) - x)$$

where $\nabla \psi$ is the optimal transport map from ρ_a to ρ_b. For the setting where ρ_a and ρ_b are $N(0,\sigma_a^2)$ and $N(0,\sigma_b^2)$, respectively, $\nabla \psi(x) = \frac{\sigma_b}{\sigma_a} x$ and

$$D_g(x) = -\frac{k_B \Delta T}{t_{\text{cycle}}} x + \frac{8\gamma}{t_{\text{cycle}}^2} \left(\frac{\sigma_b}{\sigma_a} - 1 \right) x. \tag{60}$$

Letting $\alpha := -\frac{k_B \Delta T}{t_{\text{cycle}}} + \frac{8\gamma}{t_{\text{cycle}}^2} \left(\frac{\sigma_b}{\sigma_a} - 1 \right)$ and inserting (60) into (58) yields

$$\alpha \int xv(x) \rho_a(x) \mathrm{d}x = 0,$$

which is satisfied because of (59). Hence, ρ_a being $N(0,\sigma_a^2)$ satisfies the first-order optimality condition.

(iii) We follow up by carrying out second-order analysis. The objective is to show that the limit

$$\begin{aligned} \lim_{s \to 0} \frac{1}{s^2} \left(g(\rho_s) + g(\rho_{-s}) - 2g(\rho_a) \right) = \\ -\frac{\Delta T}{t_{\text{cycle}}} \lim_{s \to 0} \frac{\mathcal{S}(\rho_s) + \mathcal{S}(\rho_{-s}) - 2\mathcal{S}(\rho_a)}{s^2} \\ -\frac{4\gamma}{t_{\text{cycle}}^2} \lim_{s \to 0} \frac{W_2(\rho_s, \rho_b)^2 + W_2(\rho_{-s}, \rho_b)^2 - 2W_2(\rho_a, \rho_b)^2}{s^2} \end{aligned}$$

can be strictly positive. Assume $\xi(x) = \nabla \eta(x)$ for some η, and define

$$\zeta(x) = \eta(x) - \frac{x^2}{2\sigma_a^2} \int_{\mathbb{R}} z \nabla \eta(z) \rho_a(z) \mathrm{d}z,$$

For the second order derivative of the entropy, we use the existing results from [8, Eq. (2.30), Eq.(3.37)], where it is shown that

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2} \mathcal{S}(\rho_s) \bigg|_{s=0} = -k_B \int \left(\|\nabla^2 \zeta\|_F^2 + \frac{1}{\sigma_a^2} \|\nabla \zeta\|^2 \right) \rho_a \mathrm{d}x \tag{61}$$

Next, we consider the second order derivative of the Wasserstein distance. Since $\nabla \psi \sharp \rho_a = \rho_b$ and $(G_s \circ \Psi_s) \sharp \rho_a = \rho_s$, we have

$$W_2(\rho_b, \rho_s)^2 \le \int_{\mathbb{R}} |\nabla \psi(x) - G_s(\Psi_s(x))|^2 \rho_a(x) \mathrm{d}x.$$

As a result

$$\lim_{s\to 0} \frac{W_2(\rho_s, \rho_b)^2 + W_2(\rho_{-s}, \rho_b)^2 - 2W_2(\rho_a, \rho_b)^2}{s^2}$$

$$\leq \lim_{s\to 0} \frac{1}{s^2} \left[\int |\nabla\psi(x) - G_s(\Psi_s(x))|^2 \rho_a(x)\mathrm{d}x + \right.$$

$$+ \int |\nabla\psi(x) - G_{-s}(\Psi_{-s}(x))|^2 \rho_a(x)\mathrm{d}x$$

$$\left. - 2\int |\nabla\psi(x) - x|^2 \rho_a(x)\mathrm{d}x \right]$$

$$= \lim_{s\to 0} \frac{1}{s^2} \left[\int \left(|G_s(\Psi_s(x))|^2 + |G_{-s}(\Psi_{-s}(x))|^2 \right) \rho_a(x)\mathrm{d}x \right.$$

$$\left. - 2\int \left(|x|^2 + \langle \Omega_s(x), \nabla\psi(x) \rangle \right) \rho_a(x)\mathrm{d}x \right],$$

where $\Omega_s(x) = G_s(\Psi_s(x)) + G_{-s}(\Psi_{-s}(x)) - 2x$. The first three terms cancel out, because the variance is constant. Therefore, the limit simplifies to

$$\lim_{s\to 0} \frac{W_2(\rho_s, \rho_b)^2 + W_2(\rho_{-s}, \rho_b)^2 - 2W_2(\rho_a, \rho_b)^2}{s^2}$$

$$\leq -2\int_{\mathbb{R}^d} \lim_{s\to 0} \frac{\Omega_s(x)}{s^2} \cdot \nabla\psi(x)\rho_a(x)\mathrm{d}x$$

$$= -2\int_{\mathbb{R}^d} \left.\frac{\partial^2 G_s(\Psi_s(x))}{\partial s^2}\right|_{s=0} \nabla\psi(x)\rho_a(x)\mathrm{d}x \tag{62}$$

$$= -2\frac{\sigma_b}{\sigma_a} \int_{\mathbb{R}^d} \left.\frac{\partial^2 G_s(\Psi_s(x))}{\partial s^2}(x)\right|_{s=0} x\rho_a(x)\mathrm{d}x,$$

where $\nabla\psi(x) = \frac{\sigma_b}{\sigma_a}x$ is used in the last step. Next we compute $\frac{\partial^2 G_s(\Psi_s(x))}{\partial s^2}(x)|_{s=0}$. Differentiating once gives

$$\frac{\partial G_s(\Psi_s(x))}{\partial s} = \frac{\partial}{\partial s}(r(s)\Psi_s(x))$$

$$= r(s)\nabla\eta\left(\Psi_s(x)\right) + \dot{r}(s)\Psi_s(x).$$

Differentiating twice and evaluating at $s = 0$ gives

$$\left.\frac{\partial^2 G_s(\Psi_s(x))}{\partial s^2}\right|_{s=0}$$

$$= 2\dot{r}(0)\nabla\eta(x) + r(0)\nabla^2\eta(x)\nabla\eta(x) + \ddot{r}(0)x.$$

Inserting this expression into (62) gives

$$\lim_{s\to 0} \frac{W_2(\rho_s, \rho_b)^2 + W_2(\rho_{-s}, \rho_b)^2 - 2W_2(\rho_a, \rho_b)^2}{s^2}$$

$$\leq -2\frac{\sigma_b}{\sigma_a} \left[2\dot{r}(0)\int x\nabla\eta(x)\rho_a(x)\mathrm{d}x \right. \tag{63}$$

$$\left. + \int x\nabla^2\eta(x)\nabla\eta(x)\rho_a(x)\mathrm{d}x + \sigma_a^2\ddot{r}(0) \right].$$

Inserting the derivatives of $r(s)$,

$$\dot{r}(0) = -\frac{1}{\sigma_a^2} \int \nabla\eta(x)x\rho_a(x)\mathrm{d}x,$$

$$\ddot{r}(0) = \frac{3}{\sigma_a^4} \left(\int \nabla\eta(x)x\rho_a(x)\mathrm{d}x \right)^2$$

$$- \frac{1}{\sigma_a^2} \int (|\nabla\eta(x)|^2 + x\nabla^2\eta(x)\nabla\eta(x))\rho_a(x)\mathrm{d}x,$$

gives

$$\lim_{s \to 0} \frac{W_2(\rho_s, \rho_b)^2 + W_2(\rho_{-s}, \rho_b)^2 - 2W_2(\rho_a, \rho_b)^2}{s^2}$$

$$\leq -\frac{2\sigma_b}{\sigma_a} \left[\frac{1}{\sigma_a^2} \left(\int x \nabla \eta(x) \rho_a(x) \mathrm{d}x \right)^2 - \int |\nabla \eta|^2 \rho_a \mathrm{d}x \right]$$

$$= 2\frac{\sigma_b}{\sigma_a} \int |\nabla \zeta(x)|^2 \rho_a(x) \mathrm{d}x. \tag{64}$$

Using (61) and (64), we conclude that

$$\lim_{s \to 0} \frac{1}{s^2} \left(g(\rho_s) + g(\rho_{-s}) - 2g(\rho_a) \right)$$

$$\geq \left(\frac{k_B \Delta T}{t_{\text{cycle}} \sigma_a^2} - \frac{8\gamma \sigma_b}{t_{\text{cycle}}^2 \sigma_a} \right) \int \|\nabla \zeta\|^2 \rho_a \mathrm{d}x.$$

Hence, when $\sigma_b \in (\sigma_a, \frac{k_B \Delta T t_{\text{cycle}}}{8\gamma \sigma_a}]$, the second-order variation is positive and $\rho_a = N(0, \sigma_a^2)$ is a local minimizer.

D. Proof of the lower-bound in Theorem 6

The constraint (51) is expressed as:

$$0 \leq \lambda \leq \frac{\sqrt{\gamma M}}{\gamma \sigma} - \frac{k_B T_c}{\gamma \sigma^2}, \quad \text{for } \sigma \geq \frac{k_B T_c}{\sqrt{\gamma M}}.$$

The inequality $\lambda \geq 0$ ensures that the power is non-negative, whereas $\sigma \geq \frac{k_B T_c}{\sqrt{\gamma M}}$ ensures that the upper bound is positive. We utilize dimensionless variables

$$x := \frac{\lambda}{\lambda_0}, \quad y := \frac{\sigma_0}{\sigma}$$

for $\sigma_0 := k_B T_c / \sqrt{\gamma M}$, $\lambda_0 := M / k_B T_c$, and re-write (50) and the constraints,

$$P = M f(x, y)$$

$$0 \leq x \leq g(y), \quad 0 < y \leq 1$$

where $f(x, y) = \frac{\Delta T}{2T_C} x - \frac{x^2}{y^2}$, $g(y) = y - y^2$. As long as $y \leq y_0$, where $y_0 = \frac{1}{1 + \frac{\Delta T}{4T_C}}$, the unconstrained maximizer

$$x^*(y) = \underset{x}{\operatorname{argmax}} \ f(x, y) = \frac{\Delta T}{4T_C} y^2$$

satisfies the constraint $x^*(y) \leq g(y)$. When $y_0 < y \leq 1$, the maximizer is at $x = g(y)$. Hence,

$$\max_{x \leq y - y^2} f(x, y) = \begin{cases} \frac{(\Delta T)^2}{16 T_C^2} y^2, & 0 < y \leq y_0 \\ \frac{\Delta T}{2T_C} (y - y^2) - (1 - y)^2, & y_0 \leq y \leq 1 \end{cases}.$$

Maximizing the expressions in the two cases over y gives

$$\max \left\{ \left(\frac{\Delta T}{3T_C + T_H} \right)^2, \frac{(\Delta T)^2}{8T_C(T_C + T_H)} \right\} = \frac{(\Delta T)^2}{8T_C(T_C + T_H)}.$$

This is achieved for

$$\sigma = \frac{k_B T_c}{\sqrt{\gamma M}} \frac{2(T_h + T_c)}{(T_h + 3T_c)}, \quad \lambda = \frac{M}{k_B T_c} \frac{(T_h + 3T_c)(T_h - T_c)}{4(T_h + T_c)^2}. \tag{65}$$

E. Acknowledgments

References

[1] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures.* Springer Science & Business Media, 2008.

[2] Erik Aurell, Krzysztof Gawędzki, Carlos Mejía-Monasterio, Roya Mohayaee, and Paolo Muratore-Ginanneschi. Refined second law of thermodynamics for fast random processes. *Journal of statistical physics*, 147(3):487–505, 2012.

[3] Erik Aurell, Carlos Mejía-Monasterio, and Paolo Muratore-Ginanneschi. Optimal protocols and optimal transport in stochastic thermodynamics. *Physical review letters*, 106(25):250601, 2011.

[4] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[5] Roger W Brockett. Thermodynamics with time: Exergy and passivity. *Systems & Control Letters*, 101:44–49, 2017.

[6] Roger W. Brockett and Jan C. Willems. Stochastic control and the second law of thermodynamics. In *1978 IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes*, pages 1007–1011. IEEE, 1979.

[7] Herbert B Callen. Thermodynamics and an introduction to thermostatistics, 1998.

[8] Eric A. Carlen and Wilfrid Gangbo. Constrained deepest descent in the 2-Wasserstein metric. *Annals of Mathematics*, pages 807–846, 2003.

[9] Sadi Carnot. *Reflexions on the motive power of fire: a critical edition with the surviving scientific manuscripts.* Manchester University Press, 1986.

[10] J Casas-Vázquez and D Jou. Temperature in non-equilibrium states: a review of open problems and current proposals. *Reports on Progress in Physics*, 66(11):1937, 2003.

[11] Paul Chambadal. Les centrales nucléaires. *Manuales/Ingenieria*, 4:1–58, 1957.

[12] Lixuan Chen and Zijun Yan. The effect of heat-transfer law on performance of a two-heat-source endoreversible cycle. *The Journal of Chemical Physics*, 90(7):3740–3743, 1989.

[13] Yongxin Chen, Tryphon Georgiou, and Allen Tannenbaum. Stochastic control and non-equilibrium thermodynamics: fundamental limits. *IEEE Transactions on Automatic Control*, 2019.

[14] F. L. Curzon and B. Ahlborn. Efficiency of a Carnot engine at maximum power output. *American Journal of Physics*, 43(1):22–24, 1975.

[15] Sybren Ruurds De Groot and Peter Mazur. *Non-equilibrium thermodynamics.* Courier Corporation, 2013.

[16] Andreas Dechant, Nikolai Kiesel, and Eric Lutz. Underdamped stochastic heat engine at maximum efficiency. *Europhysics Letters*, 119(5):50003, 2017.

[17] Andreas Dechant and Yohei Sakurai. Thermodynamic interpretation of Wasserstein distance. *arXiv preprint arXiv:1912.08405*, 2019.

[18] Massimiliano Esposito, Ryoichi Kawai, Katja Lindenberg, and Christian Van den Broeck. Quantum-dot Carnot engine at maximum power. *Physical review E*, 81(4):041106, 2010.

[19] Ivan Gentil, Christian Léonard, Luigia Ripani, and Luca Tamanini. An entropic interpolation proof of the HWI inequality. *Stochastic Processes and their Applications*, 2019.

[20] Marco F Huber, Tim Bailey, Hugh Durrant-Whyte, and Uwe D Hanebeck. On entropy approximation for Gaussian mixture random vectors. In *2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 181–188. IEEE, 2008.

[21] Christopher Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78(14):2690, 1997.

[22] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker–Planck equation. *SIAM journal on mathematical analysis*, 29(1):1–17, 1998.

[23] Georgy Lebon, David Jou, and José Casas-Vázquez. *Understanding non-equilibrium thermodynamics*, volume 295. Springer, 2008.

[24] James Melbourne, Saurav Talukdar, and Murti V Salapaka. Realizing information erasure in finite time. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 4135–4140. IEEE, 2018.

[25] Sanjoy K Mitter and Nigel J Newton. Information and entropy flow in the Kalman–Bucy filter. *Journal of Statistical Physics*, 118(1-2):145–176, 2005.

[26] I. I. Novikov. The efficiency of atomic power stations (a review). *Journal of Nuclear Energy (1954)*, 7(1-2):125–128, 1958.

[27] Juan M. R. Parrondo, Jordan M. Horowitz, and Takahiro Sagawa. Thermodynamics of information. *Nature physics*, 11(2):131, 2015.

[28] Michele Pavon. Stochastic control and nonequilibrium thermodynamical systems. *Applied Mathematics and Optimization*, 19(1):187–202, 1989.

[29] Yury Polyanskiy and Yihong Wu. Wasserstein continuity of entropy and outer bounds for interference channels. *IEEE Transactions on Information Theory*, 62(7):3992–4002, 2016.

[30] Tanmay Rajpurohit and Wassim Haddad. Stochastic thermodynamics: A dynamical systems approach. *Entropy*, 19(12):693, 2017.

[31] Henrik Sandberg, Jean-Charles Delvenne, Nigel J Newton, and Sanjoy K Mitter. Maximum work extraction and implementation costs for nonequilibrium Maxwell's demons. *Physical Review E*, 90(4):042119, 2014.

[32] Tim Schmiedl and Udo Seifert. Efficiency at maximum power: An analytically solvable model for stochastic heat engines. *EPL (Europhysics Letters)*, 81(2):20003, 2007.

[33] Udo Seifert. Stochastic thermodynamics: principles and perspectives. *The European Physical Journal B*, 64(3-4):423–431, 2008.

[34] Udo Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on progress in physics*, 75(12):126001, 2012.

[35] Ken Sekimoto. *Stochastic energetics*, volume 799. Springer, 2010.

[36] Saurav Talukdar, Shreyas Bhaban, and Murti V Salapaka. Memory erasure using time-multiplexed potentials. *Physical Review E*, 95(6):062121, 2017.

[37] Cédric Villani. *Topics in optimal transportation.* Number 58. American Mathematical Soc., 2003.

[38] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[39] David Wallace. Thermodynamics as control theory. *Entropy*, 16(2):699–725, 2014.