Automatica 123 (2021) 109366

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Maximal power output of a stochastic thermodynamic engine[☆]

Rui Fu[a], Amirhossein Taghvaei[a], Yongxin Chen[b], Tryphon T. Georgiou[a,*]

[a] Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA, United States of America
[b] School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, United States of America

ARTICLE INFO

Article history:
Received 29 January 2020
Received in revised form 7 October 2020
Accepted 18 October 2020
Available online 17 November 2020

Keywords:
Non-equilibrium thermodynamics
Optimal transportation
Mean-field optimal control

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.

1. 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 (Callen, 1998; Carnot, 1986) 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 non-vanishing power output in a thermodynamic process, this must take place in finite time, and thereby, away from *equilibrium* (Casas-Vázquez & Jou, 2003; De Groot & Mazur, 2013; Lebon, Jou, & Casas-Vázquez, 2008). To this end, the framework of stochastic thermodynamics (Brockett, 2017; Dechant, Kiesel, & Lutz, 2017; Parrondo, Horowitz, & Sagawa, 2015; Seifert, 2008, 2012; Sekimoto, 2010) 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 behavior 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 (*canonical*) 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* (Melbourne, Talukdar, & Salapaka, 2018; Talukdar, Bhaban, & Salapaka, 2017) and hence computation, i.e., a finite-time Landauer bound, as well

[☆] The research was supported in part by the NSF, United States of America under grants 1807664, 1839441, 1901599, 1942523, and the AFOSR, United States of America under FA9550-20-1-0029. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Valery Ugrinovskii under the direction of Editor Ian R. Petersen.

[*] Corresponding author.
E-mail addresses: rfu2@uci.edu (R. Fu), ataghvae@uci.edu (A. Taghvaei), yongchen@gatech.edu (Y. Chen), tryphon@uci.edu (T.T. Georgiou).

https://doi.org/10.1016/j.automatica.2020.109366
0005-1098/© 2020 Elsevier Ltd. All rights reserved.

as assessing the efficiency of thermodynamic engines operating at maximal power.

The question of efficiency at maximal power was studied independently by Chambadal (1957), Curzon and Ahlborn (1975) and Novikov (1958) 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, 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 (1989), based on differing sets of assumptions, arrived at different bounds. More recently Schmiedl and Seifert (2007), 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 (Seifert, 2012; Sekimoto, 2010). This is a *stochastic control problem*. Our analysis is based on an over-damped 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 output 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] (Eq. (49))

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

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 $\frac{M}{8}(\frac{T_h}{T_c} - 1)$ 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 $(\frac{T_h}{T_c} - 1)/(\frac{T_h}{T_c} + 1)$, which depends only on the ratio of temperatures of the two heat baths as well.

The exposition proceeds as follows. Section 2 details the stochastic model thermodynamic ensembles and the heat/energy exchange mechanism. Section 3 is a brief overview of optimal transport theory, on which the main results are based. Section 4 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 5 returns to the concept of a cyclically 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 6 contains the main results regarding seeking maximal power output. Specifically, Section 6.1 explains optimal scheduling times, Section 6.2 highlights questions that arise based on physical grounds for Gaussian thermodynamic states, Sections 6.3 and 6.4 discuss optimal thermodynamic states at the two ends of the Carnot-like cycle, and Sections 6.5 and 6.6 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.

2. 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.

2.1. Langevin dynamics

The (under-damped) Langevin equations

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

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

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, for temporal resolution $\Delta t \gg \frac{m}{\gamma}$ and small particle size, the dynamics reduce 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, Eq. (2) 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 Sekimoto (2010, 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

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

Remark 1. The under-damped Langevin equation (1) is the most common dynamical model for a particle immersed in a heat bath (Sekimoto, 2010); alternative models can be based on e.g., a

[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".

Table 1
Symbols and corresponding units.

Definition	Notation	Units
Time	t	s
Position of particle	X_t	m
Boltzmann constant	k_B	N m
Damping coefficient	γ	N s/m
Potential	$U(t,x)$	N m
Temperature	T	°K
Brownian motion	B_t	$\mathrm{s}^{\frac{1}{2}}$
Density in $\mathbb{R}^d$	$\rho(t,x)$	m^{-d}
Velocity field in $\mathbb{R}^d$	$v(t,x)$	m/s
Wasserstein metric, length	$\mathrm{W}_2(\cdot,\cdot)$, $\ell_{\rho[t_i,t_f]}$	m
Entropy	$\mathcal{S}(\rho)$	N m
Work (particle/ensemble)	$W, \mathcal{W}$	N m
Heat (particle/ensemble)	$Q, \mathcal{Q}$	N m
Energy (particle/ensemble)	$U, \mathcal{E}$	N m
Free energy	$\mathcal{F}$	N m
Bound in (49)	M	N m/s
Power	P	N m/s

Poisson process for the thermal excitation, space-dependent viscosity coefficient, and possibly nonlinear effects of an interaction potential. It has been used to model e.g., colloidal particles in a laser trap, enzymes and molecular motors in single molecule assays, and so on Seifert (2012). In the present work, we follow recent literature (Argun et al., 2017; Dechant et al., 2017; Gomez-Marin, Schmiedl, & Seifert, 2008; Park, Chun, & Noh, 2016; Schmiedl & Seifert, 2007) where, besides Brownian excitation, the focus on constant viscosity coefficient γ.

2.2. 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 is formally expressed as the product of force and displacement

$$(-\gamma\frac{\mathrm{d}X_t}{\mathrm{d}t}+\sqrt{2\gamma k_B T}\frac{\mathrm{d}B_t}{\mathrm{d}t})\circ \mathrm{d}X_t$$

in Stratonovich form. Using (2), formally,

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

which leads to the expression $\text{đ}Q=\nabla_x U(t,X_t)\circ \mathrm{d}X_t$ for the heat; see Sekimoto (2010, Chapter 4.1) for a more detailed exposition. Then, bringing in the Itô correction, we arrive at

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

Note that we use đ, as in the case of not-perfect differentials, to emphasize that $\int \text{đ}Q$ depends on the path and not just on end-point conditions.

The *work* transferred to the particle by a change in the actuating potential is taken as[3]

$$\text{đ}W=\frac{\partial U}{\partial t}(t,X_t)\mathrm{d}t. \tag{4}$$

Thence, since the internal energy is simply the value of the potential, the *first law of thermodynamics*, $\mathrm{d}U(t,X_t)=\text{đ}Q+\text{đ}W$, holds.

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

$$\text{đ}\mathcal{Q}=\left[\int_{\mathbb{R}^d}\left(-\frac{1}{\gamma}\|\nabla_x U\|^2+\Delta_x U\frac{k_B T}{\gamma}\right)\rho\,\mathrm{d}x\right]\mathrm{d}t \tag{5a}$$

$$\text{đ}\mathcal{W}=\left[\int_{\mathbb{R}^d}\frac{\partial U}{\partial t}\rho\,\mathrm{d}x\right]\mathrm{d}t, \tag{5b}$$

leading to the first law for the ensemble $\mathrm{d}\mathcal{E}(\rho,U)=\text{đ}\mathcal{Q}+\text{đ}\mathcal{W}$, where the internal energy is

$$\mathcal{E}(\rho,U)=\int_{\mathbb{R}^d}U\rho\,\mathrm{d}x, \tag{5c}$$

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

2.3. 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 $x,y\in\mathbb{R}^d$. Throughout the paper, the stochastic differential equations are stated in Itô form, unless the Stratonovich integration notation $\circ$ is used explicitly. The Gaussian distribution with mean m and covariance Σ is denoted by $N(m,\Sigma)$. For convenience we provide Table 1 of the various quantities, including the corresponding units in SI format: Newton (N), seconds (s), meter (m), absolute temperature in degrees Kelvin (°K).

3. A brief excursion into optimal mass transport

As it turns out, dissipation in Langevin models (2) is closely linked to the path that a thermodynamic ensemble traverses. This path is seen as a trajectory in the space probability distributions and its length, that quantifies dissipation, is metrized by the so-called Wasserstein metric. Thus, we now embark on a brief excursion into the basics of optimal mass transport so as to provide context for needed results in Wasserstein geometry—the pertinence of the Wasserstein metric to thermodynamics has been recognized in Aurell, Gawędzki, Mejía-Monasterio, Mohayaee, and Muratore-Ginanneschi (2012), Aurell, Mejía-Monasterio, and Muratore-Ginanneschi (2011), Chen, Georgiou, and Tannenbaum (2020), Dechant and Sakurai (2019), Jordan, Kinderlehrer, and Otto (1998) and Seifert (2012).

We denote by $\mathcal{P}_2(\mathbb{R}^d)$ the space of probability distributions with finite second-order moment. We utilize the notation $\mathrm{d}\mu$ to signify a probability measure while we write $\mathrm{d}\mu(x)=\rho(x)dx$ to signify that $\mathrm{d}\mu$ is absolutely continuous with respect to the Lebesgue measure $\mathrm{d}x$ with ρ the corresponding probability density.

For $\mathrm{d}\mu_0, \mathrm{d}\mu_1\in\mathcal{P}_2(\mathbb{R}^d)$, the 2-Wasserstein distance

$$\mathrm{W}_2(\mu_0,\mu_1):=\sqrt{\inf_{\pi\in\Pi(\mu_0,\mu_1)}\int_{\mathbb{R}^d\times\mathbb{R}^d}\|x-y\|^2\mathrm{d}\pi(x,y)}, \tag{6}$$

where $\Pi(\mu_0,\mu_1)$ denotes the set of probability measures on the product space $\mathbb{R}^d\times\mathbb{R}^d$ with μ_0,μ_1 as marginals, is a *bona fide* metric on the space of distributions. This metric, in fact, induces a Riemannian-like structure as we explain below.

The expression $\int\|x-y\|^2\mathrm{d}\pi(x,y)$ above is a relaxation of the transportation cost

$$\int_{\mathbb{R}^d}\|x-\Psi(x)\|^2\mathrm{d}\mu_0(x)$$

in *Monge's* problem (Villani, 2003) to be minimized over maps Ψ that transfer the "mass" distribution μ_0 into μ_1, i.e., such

[3] This particular formula for the work has been the subject of considerable debate (Horowitz & Jarzynski, 2008; Peliti, 2008a, 2008b; Vilar & Rubi, 2008).

that $\int_A d\mu_1 = \int_{\Psi^{-1}(A)} d\mu_0$ over measurable sets A. This relation is denoted by $\Psi\sharp\mu_0 = \mu_1$. In case the two measures admit densities, it can be expressed via the change of variables formula

$$\det(\nabla_x\Psi(x))\rho_1(\Psi(x)) = \rho_0(x),$$

for the respective ρ_i's ($i \in \{0, 1\}$). In fact, in this case where both measures admit densities, the support of the optimal π in (the convex problem) (6) coincides with the graph of the unique minimizing map $\Psi : \mathbb{R}^d \to \mathbb{R}^d$ for Monge's problem. Further, the optimal Ψ is the gradient of a convex function ψ on $\mathbb{R}^d$ (Villani, 2003, Ch. 5), i.e., $\Psi = \nabla_x\psi$. Interesting, being a gradient "vector field", Ψ is curl-free, which in itself characterizes optimality.

We now sketch how $\mathcal{P}_2(\mathbb{R}^d)$ can be equipped with a Riemannian-like structure, while we refer to Ambrosio, Gigli, and Savaré (2008) for a rigorous exposition. For brevity and for notational convenience, we are only concerned with distributions that admit densities and use the simplified notation $\rho \in \mathcal{P}_2(\mathbb{R}^d)$.

Consider an "infinitesimal" perturbation $\rho + \delta \in \mathcal{P}_2(\mathbb{R}^d)$ and the solution ϕ to the Poisson equation

$$\nabla_x \cdot (\rho\nabla_x\phi) = -\delta.$$

The map $\Psi = \mathrm{Id} + \nabla_x\phi$, where Id denotes the identity map, optimally transports ρ into $\rho + \delta$ as $(\mathrm{Id} + \nabla_x\phi)\sharp\rho \simeq \rho + \delta$. Alternatively, $v = \nabla_x\phi$ can be viewed as a velocity field effecting transport as $\rho - \nabla_x \cdot (\rho v) = \rho + \delta$.

Thus, the correspondence $\delta \to \nabla_x\phi$ identifies tangent directions δ on $\mathcal{P}_2(\mathbb{R}^d)$, i.e., rates of change $\frac{\partial\rho}{\partial t} = \delta$ about a given density ρ, with an (optimal) corresponding velocity field $\nabla_x\phi$. Hence, it is natural to consider the (twice) average "kinetic energy" to define a metric on tangent directions on $\mathcal{P}_2(\mathbb{R}^d)$. Specifically, if $\delta_i = \frac{\partial\rho}{\partial t_i}$, for $i \in \{1, 2\}$, represent two tangent directions at ρ, we define the inner product

$$\langle\frac{\partial\rho}{\partial t_1}, \frac{\partial\rho}{\partial t_2}\rangle_{\mathrm{W}} := \int_{\mathbb{R}^d}\langle\nabla_x\phi_1, \nabla_x\phi_2\rangle\rho\, \mathrm{d}x, \tag{7}$$

where the ϕ_i's solve $\nabla_x \cdot (\rho\nabla_x\phi_i) = -\frac{\partial\rho}{\partial t_i}$. The associated norm is

$$\|\frac{\partial\rho}{\partial t}\|_{\mathrm{W}} := \sqrt{\langle\frac{\partial\rho}{\partial t}, \frac{\partial\rho}{\partial t}\rangle_{\mathrm{W}}}.$$

Consider $\rho_{[t_i,t_f]} := \{\rho(t, \cdot) \in \mathcal{P}_2(\mathbb{R}^d) | t \in [t_i, t_f]\}$ as a curve (path) in $\mathcal{P}_2(\mathbb{R}^d)$. Two quantities of interest are its length,

$$\ell_{\rho_{[t_i,t_f]}} := \int_{t_i}^{t_f} \|\frac{\partial\rho}{\partial t}\|_{\mathrm{W}}\mathrm{d}t, \tag{8}$$

and the *kinetic energy* integral (*action*) along the path

$$\mathcal{A}_{\rho_{[t_i,t_f]}} := \int_{t_i}^{t_f} \|\frac{\partial\rho}{\partial t}\|_{\mathrm{W}}^2\mathrm{d}t \tag{9}$$

(modulo a factor of $\frac{1}{2}$). It can be seen that

$$\ell_{\rho_{[t_i,t_f]}} = \min\sqrt{(t_f - t_i)\mathcal{A}_{\rho_{[t_i,t_f]}}},$$

over time-parametrizations of the path, with the minimum corresponding to constant velocity. Moreover, the minimal path-length between two end-points ρ_{t_i} and ρ_{t_f} turns out to be precisely $\mathrm{W}_2(\rho_{t_i}, \rho_{t_f})$, and thus, $\mathcal{P}_2(\mathbb{R}^d)$ is a length space, Ambrosio et al. (2008) and Villani (2003, Chapter 8).

We conclude with an important inequality linking the Wasserstein metric to information functionals. Consider a reference probability distribution $\mathrm{d}\mathfrak{m} = e^{-V}\mathrm{d}x \in \mathcal{P}_2(\mathbb{R}^d)$, with $V(x)$ having Hessian $\nabla_x^2 V \geq \kappa I$ for $\kappa \in \mathbb{R}$, and $d\mu = \rho d\mathfrak{m}$ also in $\mathcal{P}_2(\mathbb{R}^d)$. The *relative entropy* and *Fisher information* functionals, respectively, are defined by

$$H(\mu|\mathfrak{m}) := \int_{\mathbb{R}^d} \rho\log(\rho)\,\mathrm{d}\mathfrak{m}, \tag{10a}$$

$$I(\mu|\mathfrak{m}) := \int_{\mathbb{R}^d} \|\nabla_x\log(\rho)\|^2\rho\,\mathrm{d}\mathfrak{m}. \tag{10b}$$

These are linked to the Wasserstein distance via the following HWI* inequality (Gentil, Léonard, Ripani, & Tamanini, 2019; Otto & Villani, 2000),

$$H(\mu_1|\mathfrak{m}) - H(\mu_2|\mathfrak{m}) \leq \mathrm{W}_2(\mu_1, \mu_2)\sqrt{I(\mu_1|\mathfrak{m})} - \frac{\kappa^2}{2}\mathrm{W}_2^2(\mu_1, \mu_2), \quad \forall\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d). \tag{11}$$

4. The second law, dissipation, and Wasserstein geometry

Next, we discuss the *second law of thermodynamics* in the context of an ensemble of particles obeying over-damped Langevin dynamics (2) for a heat bath with constant temperature $T(t) = T$. The classical formulation of the law amounts to the inequality

$$\mathcal{W} - \Delta\mathcal{F} \geq 0, \tag{12}$$

where $\mathcal{W} = \int_{t_i}^{t_f} đ\mathcal{W}$ is the work transferred to the ensemble over a time interval (t_i, t_f), and $\Delta\mathcal{F}$ is the change in the free energy[4]

$$\mathcal{F}(\rho, U) = \mathcal{E}(\rho, U) - T\mathcal{S}(\rho) \tag{13}$$

between the two end-point states, see Owen (2012) and Parrondo et al. (2015). Here,

$$\mathcal{S}(\rho) = -k_B\int_{\mathbb{R}^d}\log(\rho)\,\rho\,\mathrm{d}x \tag{14}$$

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

Inequality (12) becomes equality for quasi-static (reversible) thermodynamic transitions. In general, for irreversible transitions, the gap in (12) 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 (Jarzynski, 1997b) to hold for irreversible thermodynamic transitions between work and free energy, in the form,

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

where the expectation is taken over the probability law on paths, $W = \int đW$ represents the work along trajectories of individual particles, and $\Delta\mathcal{F}_{eq} = -\beta^{-1}\log(\frac{Z_{t_f}}{Z_{t_i}})$ signifies the difference of the equilibrium free energy $-\beta^{-1}\log(Z_t)$ at the two end-points in time $t \in \{t_i, t_f\}$. Here, $Z_t = \int_{\mathbb{R}^d} e^{-\beta U(t,x)}\mathrm{d}x$ where, as usual, $\beta = 1/k_BT$. In Jarzynski's original derivation (Jarzynski, 1997a, 1997b) of the Jarzynski equality, the notions of work and heat are in alignment with the ones used in this paper, though (Jarzynski, 1997a) considers more general stochastic dynamics satisfying one type of detailed balance condition (Jarzynski, 1997a, Section 1). Interestingly, the Jarzynski equality holds even for an alternative notion of work, see e.g., Kurchan (1998).

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

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

the ensemble. An alternative identity that quantifies explicitly the gap in (12) holds for irreversible thermodynamic transitions. This identity is (cf. Theorem 1)

$$\mathcal{W}-\Delta\mathcal{F}=\underbrace{\gamma\int_{t_i}^{t_f}\left\|\frac{\partial\rho}{\partial t}\right\|_{\mathrm{W}}^{2}\mathrm{d}t}_{\text{dissipation}},\tag{15}$$

which is γ times $\mathcal{A}_{\rho_{[t_i,t_f]}}$, the action integral along the time-parametrized path traversed. Thus, if the path is selected as a "constant speed" W_2-geodesic,

$$\mathcal{W}-\Delta\mathcal{F}=\frac{\gamma}{t_f-t_i}\mathrm{W}_2(\rho_{t_i},\rho_{t_f})^2\tag{16}$$

quantifies the least amount of work needed for transition between specified end-point thermodynamic states, or the maximal work that can be drawn. We recap the key points below.

Theorem 1. *Consider thermodynamic transitions between states ρ_{t_i}, ρ_{t_f}, under constant temperature T and a time-varying potential U for the overdamped Langevin model (2). Then,*

$$\mathcal{W}-\Delta\mathcal{F}\geq\frac{\gamma}{t_f-t_i}\ell^2_{\rho_{[t_i,t_f]}}.\tag{17}$$

Relation (17) holds with equality for a path of the thermodynamic ensemble chosen to be a constant speed W_2-geodesic, effected by a suitable potential, a choice that corresponds to minimal dissipation.

Proof. We first derive (15), cf. Graham (1978) and Pavon and Ticozzi (2006) for similar computations with time independent potential. Consider

$$\begin{aligned}\frac{\mathrm{d}\mathcal{F}}{\mathrm{d}t}(\rho,U)=&\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}(\rho,U)-T\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{S}(\rho)\\=&\int_{\mathbb{R}^d}\frac{\partial U}{\partial t}\rho dx+\int_{\mathbb{R}^d}(U+k_BT(1+\log\rho))\frac{\partial\rho}{\partial t}dx.\end{aligned}$$

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

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

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[5]

$$v:=-\frac{1}{\gamma}(\nabla_xU+k_BT\nabla_x\log\rho).\tag{18}$$

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

$$\Delta\mathcal{F}=\mathcal{W}-\gamma\int_{t_i}^{t_f}\int_{\mathbb{R}^d}\|v\|^2\rho\;\mathrm{d}x\,\mathrm{d}t,\tag{19}$$

where v is the gradient of $\phi=-\frac{1}{\gamma}(U+k_BT\log\rho)$ and satisfies the continuity equation $\nabla\cdot(\rho\nabla\phi)=\frac{\partial\rho}{\partial t}$ as claimed. This establishes (15).

The inequality (17) follows 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 (8). Specifically, provided $\int_{\mathbb{R}^d}\|v\|^2\rho\;\;\mathrm{d}x=\alpha^2$ remains constant along the path (i.e., for $t\in[t_i,t_f]$),

$$\alpha=\frac{1}{t_f-t_i}\ell_{\rho_{[t_i,t_f]}}.$$

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

$$\tilde{t}(t):=\frac{\ell_{\rho_{[t_i,t]}}}{\ell_{\rho_{[t_i,t_f]}}}(t_f-t_i)+t_i$$

will be traversed via a velocity field

$$\tilde{v}(\tilde{t}(t))=\frac{v(t)}{\|v(t)\|_\rho}\frac{\ell_{\rho_{[t_i,t_f]}}}{t_f-t_i}.$$

Knowing $\tilde{v}$, a new potential $\tilde{U}$ can be computed so that $\tilde{v}(\tilde{t},\cdot)=\nabla_x\tilde{U}(\tilde{t},\cdot)+k_BT\nabla_x\log(\rho(\tilde{t},\cdot))$. Finally, equality in (17) holds when taking $\rho_{[t_i,t_f]}$ to be a geodesic (Villani, 2008). □

Remark 2. Early work by Jordan et al. (1998), 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 fact was recognized in Aurell et al. (2012, 2011), Seifert (2012) and more recently developed in Chen et al. (2020) and Dechant and Sakurai (2019).

5. 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 corresponding actuation protocols.

5.1. Isothermal transition

We consider transition between states ρ_{t_i} and ρ_{t_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 (15) between work, free energy, and the dissipation, and the first law, we have the following identity relating thermodynamic quantities in isothermal transitions

$$\mathcal{W}=\Delta\mathcal{E}-T\Delta\mathcal{S}+\mathcal{W}_{\rm irr}\tag{20a}$$

$$\mathcal{Q}=T\Delta\mathcal{S}-\mathcal{W}_{\rm irr}\tag{20b}$$

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

$$\frac{\gamma}{t_f-t_i}\mathrm{W}_2(\rho_{t_i},\rho_{t_f})^2\tag{20c}$$

by the choice of actuation $\nabla_xU(t,\cdot)$ in (18) with v the optimal velocity field minimizing dissipation in (15) (item (iii) in Theorem 1).

It is important to note that the minimizing v can be obtained by solving a convex reformulation of (15) in terms of the density

[5] We note that v is known as Nelson's current velocity (Chen, Georgiou, & Pavon, 2016).

$\rho(t,\cdot)$ and the momentum field $\mathbf{p}(t,\cdot)=v(t,\cdot)\rho(t,\cdot)$, in the form

$$\min_{\mathbf{p}(t,\cdot),\rho(t,\cdot)} \int_{t_i}^{t_f}\int_{\mathbb{R}^d} \frac{\|\mathbf{p}\|^2}{\rho}\mathrm{d}x\mathrm{d}t \tag{21a}$$

$$\text{subject to } \frac{\partial\rho}{\partial t}+\nabla_x\cdot\mathbf{p}=0 \tag{21b}$$

$$\text{and } \rho(t_i,\cdot),\ \rho(t_f,\cdot) \text{ specified.} \tag{21c}$$

Then, $v=\mathbf{p}/\rho$, see Benamou and Brenier (2000, Section 4) and Villani (2003, p. 241).

5.2. Adiabatic transition

We now consider transition between ρ_{t_i} and ρ_{t_f} for the ensemble modeled by (2), over a time interval $[t_i,t_f]$, under abrupt changes in the potential $U(t,\cdot)$ 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 (15) vanishes. Moreover, the entropy of the ensemble remains constant. Thus, the work input into the system equal to change in internal energy,

$$\mathcal{W}=\int_{\mathbb{R}^d}(U(t^+,x)-U(t^-,x))\rho(t,x)\mathrm{d}x=\Delta\mathcal{E}, \tag{22a}$$

and therefore no heat transfer takes place, and therefore,

$$\mathcal{Q}=0. \tag{22b}$$

5.3. 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 Fig. 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 (20),

$$\mathcal{W}^{(1)}=\Delta\mathcal{E}^{(1)}-T_h\Delta\mathcal{S}^{(1)}+\mathcal{W}_{\text{irr}}^{(1)} \tag{23a}$$

$$\mathcal{Q}^{(1)}=T_h\Delta\mathcal{S}^{(1)}-\mathcal{W}_{\text{irr}}^{(1)} \tag{23b}$$

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

$$\mathcal{W}_{\text{irr}}^{(1)}=\frac{\gamma}{t_1}\mathrm{W}_2(\rho_a,\rho_b)^2. \tag{23c}$$

(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 (22),

$$\mathcal{W}^{(2)}=\Delta\mathcal{E}^{(2)} \tag{24a}$$

$$\mathcal{Q}^{(2)}=0 \tag{24b}$$

while the state remains at ρ_b.

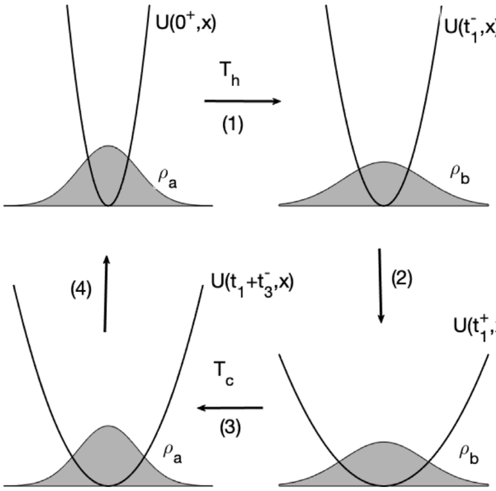


Fig. 1. Carnot-like cycle of a stochastic model for a heat engine (with $d=1$): the operation cycles clockwise through two isothermal transitions (1) and (3), and two adiabatic transitions (2) and (4). During the isothermal transitions having duration t_1 and t_3, the ensemble is in contact with a "hot" 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.

(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 (20),

$$\mathcal{W}^{(3)}=\Delta\mathcal{E}^{(3)}-T_c\Delta\mathcal{S}^{(3)}+\mathcal{W}_{\text{irr}}^{(3)} \tag{25a}$$

$$\mathcal{Q}^{(3)}=T_c\Delta\mathcal{S}^{(3)}-\mathcal{W}_{\text{irr}}^{(3)} \tag{25b}$$

$$\mathcal{W}_{\text{irr}}^{(3)}=\frac{\gamma}{t_3}\mathrm{W}_2(\rho_a,\rho_b)^2. \tag{25c}$$

(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 (22),

$$\mathcal{W}^{(4)}=\Delta\mathcal{E}^{(4)} \tag{26a}$$

$$\mathcal{Q}^{(4)}=0 \tag{26b}$$

5.4. 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 does not 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

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

As a result, the total work output is

$$\begin{aligned}-\mathcal{W}&=-\left(\sum_{i=1}^{4}\Delta\mathcal{E}^{(i)}-\sum_{i=1}^{4}T_i\Delta\mathcal{S}^{(i)}+\sum_{i=1}^{4}\mathcal{W}_{irr}^{(i)}\right)\\&=(T_h-T_c)\Delta\mathcal{S}-\mathcal{W}_{\text{irr}}^{(1)}-\mathcal{W}_{\text{irr}}^{(3)}.\end{aligned} \tag{27}$$

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

$$-\mathcal{W} = (T_h - T_c)\Delta\mathcal{S} - \gamma(\frac{1}{t_1} + \frac{1}{t_3})\mathrm{W}_2(\rho_a, \rho_b)^2. \tag{28}$$

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.[6]

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

$$\eta = \frac{-\mathcal{W}}{\mathcal{Q}_h} \tag{29}$$

where the heat input during isothermal expansion is

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

Once again assuming optimality ($\mathcal{W}_{\rm irr} = \frac{\gamma}{t_1}\mathrm{W}_2(\rho_a, \rho_b)^2$), the bound on the efficiency is seen to be

$$\eta = \frac{(T_h - T_c)\Delta\mathcal{S} - \gamma(\frac{1}{t_1} + \frac{1}{t_3})\mathrm{W}_2(\rho_a, \rho_b)^2}{T_h\Delta\mathcal{S} - \gamma\frac{1}{t_1}\mathrm{W}_2(\rho_a, \rho_b)^2}. \tag{30}$$

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

$$\begin{aligned} P &= -\mathcal{W}/(t_1 + t_3) \\ &= \frac{(T_h - T_c)\Delta\mathcal{S} - \gamma(\frac{1}{t_1} + \frac{1}{t_3})\mathrm{W}_2(\rho_a, \rho_b)^2}{t_1 + t_3} \end{aligned} \tag{31}$$

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.

6. 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 5.

Consider the expression in (31) 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 $\mathrm{W}_2(\rho_a, \rho_b)$ as well as the change in entropy $\Delta\mathcal{S}$. We will explore systematically the various options.

6.1. Optimizing the time scheduling

Optimizing the maximal power delivered during cyclic operation

$$P = \frac{1}{t_1 + t_3}(T_h - T_c)\Delta\mathcal{S} - \frac{\gamma}{t_1 t_3}\mathrm{W}_2(\rho_a, \rho_b)^2,$$

with respect to choices for t_1, t_3, with $\mathrm{W}_2(\rho_a, \rho_b)$, T_h, T_c and $\Delta\mathcal{S}$ kept fixed, gives that

$$t_1 = t_3 = \frac{4\gamma\mathrm{W}_2(\rho_a, \rho_b)^2}{(T_h - T_c)\Delta\mathcal{S}}, \tag{32}$$

and therefore that the period for the cycle is

$$t_{\rm cycle} := t_1 + t_3 = \frac{8\gamma\mathrm{W}_2(\rho_a, \rho_b)^2}{(T_h - T_c)\Delta\mathcal{S}}. \tag{33}$$

If instead we specify the period of the cycle $t_{\rm cycle}$, 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

$$t_1 = t_3 = \frac{t_{\rm cycle}}{2}. \tag{34}$$

Remark 3 (*Efficiency at Maximum Power*)**.** The thermodynamic efficiency (29) of the engine, when it is operating at optimal transition times (32) that maximize the power, is equal to

$$\eta_{SS} = \frac{2(T_h - T_c)}{3T_h + T_c} = \frac{\eta_C}{2 - \frac{\eta_C}{2}} \tag{35}$$

This result appeared in Esposito, Kawai, Lindenberg, and Van den Broeck (2010a), and Schmiedl and Seifert (2007) for the case of Gaussian marginals ρ_a, ρ_b and potential $U(t, x)$ that is quadratic in x. Our derivation establishes (35) in a general setting.

Using the expression (33), the total power delivered

$$P = \frac{(T_h - T_c)^2}{16\gamma}\left(\frac{\Delta\mathcal{S}}{\mathrm{W}_2(\rho_a, \rho_b)}\right)^2. \tag{36}$$

But as we will see in Section 6.2, optimizing the power for ρ_a, ρ_b leads to the non-physical conclusion of a vanishingly small $t_{\rm cycle}$.

6.2. The caveat of optimal $t_{\rm cycle}$: 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

$$\mathrm{W}_2(\rho_a, \rho_b)^2 = \mathrm{trace}\big(\Sigma_a + \Sigma_b - 2(\Sigma_a^{1/2}\Sigma_b\Sigma_a^{1/2})^{1/2}\big) \tag{37a}$$

$$\Delta\mathcal{S} = \mathcal{S}(\rho_b) - \mathcal{S}(\rho_a) = \frac{1}{2}k_B \log\det(\Sigma_b\Sigma_a^{-1}). \tag{37b}$$

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 $t_{\rm cycle}$ for the thermodynamic cycle as in (33), we obtain that the maximal power available, as a function of σ_a and σ_b, is given by

$$P(\sigma_a, \sigma_b) = \frac{k_B^2(T_h - T_c)^2}{16\gamma}\left(\frac{\log\frac{\sigma_b}{\sigma_a}}{\sigma_b - \sigma_a}\right)^2. \tag{38}$$

The corresponding heat uptaken from the hot reservoir and the work extracted during one cycle are

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

and

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

respectively.

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

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

$$\max_{\sigma_b} P(\sigma_a, \sigma_b) = \frac{k_B^2(T_h - T_c)^2}{16\gamma\sigma_a^2} \tag{39a}$$

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 Esposito, Kawai, Lindenberg, and Van den Broeck (2010b). 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.

6.3. Optimizing the thermodynamic state ρ_b

Henceforth we fix the period t_{cycle} as well as the duration of the isothermal phases according to (34). The power delivered, as a function of the ρ_i's ($i \in \{a, b\}$), is

$$\frac{(T_h - T_c)}{t_{\text{cycle}}}(\mathcal{S}(\rho_b) - \mathcal{S}(\rho_a)) - \frac{4\gamma}{t_{\text{cycle}}^2}\mathrm{W}_2(\rho_a, \rho_b)^2. \tag{40}$$

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

$$\min_{\rho_b}\{\mathrm{W}_2(\rho_a, \rho_b)^2 - h\mathcal{S}(\rho_b)\} \tag{41}$$

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 $\mathrm{W}_2(\cdot, \cdot)$ and, as can easily be verified, the expression

$$\mathrm{W}_2(\rho_a, \rho_b)^2 - h\mathcal{S}(\rho_b) \tag{42}$$

is strictly convex, which leads to the following statement.

Proposition 1. *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 (41).*

Proof. Eq. (41) 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 (Jordan et al., 1998). While $\mathrm{W}_2(\rho_a, \rho_b)^2 - h\mathcal{S}(\rho_b)$ 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 Jordan et al. (1998, Proposition 4.1). □

We conclude this section with two statements. The first establishes implicit conditions for optimality of ρ_b, in maximizing the expression in (40) (equivalently, minimizing (42)). For ease of referencing we view the expression in (40) as a function of ρ_b, namely,

$$f(\rho_b) := \frac{(T_h - T_c)}{t_{\text{cycle}}}(\mathcal{S}(\rho_b) - \mathcal{S}(\rho_a)) - \frac{4\gamma}{t_{\text{cycle}}^2}\mathrm{W}_2(\rho_a, \rho_b)^2. \tag{43}$$

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 (Villani, 2003, Ch. 5).

We first highlight stationarity conditions that characterize the minimizer of $f(\cdot)$ in (43).

Lemma 2. *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 $\nabla_x\psi$, for a convex function ψ on $\mathbb{R}^d$, be such that $\nabla_x\psi\sharp\rho_a = \rho_b^*$. The following (stationarity) condition holds*

$$k_B(T_h - T_c)\nabla_x \log \rho_b^*(y) - \frac{8\gamma}{t_{\text{cycle}}}\left((\nabla_x\psi)^{-1} - \mathrm{Id}\right)(y) = 0, \tag{44}$$

where Id *denotes the identity map.*

Proof. The proof is given in Appendix A.1. □

The lemma, which is of independent interest, is used in the proof of the following proposition concluding 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 3. *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*

$$\sigma_b = \frac{1 + \sqrt{1 + c}}{2}\sigma_a, \tag{45}$$

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

Proof. The proof is given in Appendix A.2. □

Remark 4. 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 3 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 6.4, 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 5).

6.4. 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 6.3 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 (40), 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 (40), maximal power would be drawn by minimizing $\mathrm{W}_2(\rho_a, \rho_b)$ over probability densities ρ_a with entropy s_a. We claim that

$$\inf_{\rho_a}\{\mathrm{W}_2(\rho_a, \rho_b) \mid \mathcal{S}(\rho_a) = s_a > -\infty\} = 0. \tag{46}$$

To see this note that

$$\inf_{\rho_a} \mathrm{W}_2(\rho_a, \rho_b) = 0$$

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 (46) is inherently related to the continuity of the entropy functional with respect to the Wasserstein distance. For a rigorous treatment of the problem, see Polyanskiy and Wu (2016), where it is shown that unless certain regularity assumptions are in place for ρ_a and ρ_b, the infimum in (46) is zero.

Remark 5 (*Gaussian is not Optimal for ρ_a*)**.** The preceding 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 (40), at $\rho_a = N(0, \sigma_a^2)$ can be 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 only a local minimizer when $\sigma_a < \sigma_b < k_B(T_h - T_c)t_{\text{cycle}}/(8\gamma\sigma_a)$. The analysis, detailed in Appendix A.3 of the arXived preprint (Fu, Taghvaei, Chen, & Georgiou, 2020), aims to highlight that the conjecture of a Gaussian ρ_a being optimal fails. In hindsight, this is not surprising. Maximizing power over ρ_a is equivalent to minimizing the entropy of ρ_a. Minimizing entropy under a fixed variance constraint does not lead to a Gaussian distribution since two Dirac delta distributions with the desired mean and variance achieve negative infinity entropy.

6.5. 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.

Proposition 4. *Under the standing assumptions on the Carnot-like cycle, the power output (40), is bounded by*

$$P \leq \frac{k_B^2(T_h - T_c)^2}{16\gamma} I(\rho_a \mathrm{d}x|\mathrm{d}x). \tag{47}$$

Proof. It is a consequence of the HWI* inequality (11) (see supplemental proof in Appendix A.3) that

$$\mathcal{S}(\rho_b) - \mathcal{S}(\rho_a) \leq k_B \mathrm{W}_2(\rho_a, \rho_b)\sqrt{I(\rho_a \mathrm{d}x|\mathrm{d}x)}. \tag{48}$$

Using the formula for power (40), we have

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

concluding the bound in (47). □

We point out that the bound in (47) becomes tight when t_{cycle} takes the optimal value (33) and $\rho_b \to \rho_a$. Specifically, if $\rho_a = N(0, \sigma_a^2)$ and $\rho_b = N(0, \sigma_b^2)$ are Gaussian distributions and t_{cycle} takes the optimal value (33), then as $\sigma_b \to \sigma_a$ the power output is given by (39a), which coincides with (47), since $I(\rho_a \mathrm{d}x|\mathrm{d}x) = \frac{1}{\sigma_a^2}$.

6.6. 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

$$\frac{1}{\gamma}\int_{\mathbb{R}^d} \|\nabla_x U(t, x)\|^2 \rho(t, x)\,\mathrm{d}x \leq M \tag{49}$$

for all $t \in (0, t_{\text{cycle}})$. Thus, we analyze the maximum power (40) that can be extracted from a thermodynamic engine, under the constraint (49).

Theorem 2. *Consider a thermodynamic ensemble, undergoing a Carnot cycle as described in Section 5, 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 (49), satisfies*

$$\frac{M}{8}(\frac{T_h}{T_c} - 1)\frac{\frac{T_h}{T_c} - 1}{\frac{T_h}{T_c} + 1} \leq P_{\max} \leq \frac{M}{8}(\frac{T_h}{T_c} - 1) \tag{50}$$

Proof. The proof for the upper-bound follows from bounding the entropy difference $S(\rho_b) - S(\rho_a)$ under the constraint (49). During the isothermal transition in contact with the cold bath with temperature T_c,

$$\begin{aligned} \mathcal{S}(\rho_b) - \mathcal{S}(\rho_a) &= -\int_{\frac{t_{\text{cycle}}}{2}}^{t_{\text{cycle}}} \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{S}(\rho(t, \cdot))\,\mathrm{d}t \\ &= \frac{-k_B}{\gamma}\int_{\frac{t_{\text{cycle}}}{2}}^{t_{\text{cycle}}} (\langle \nabla_x \log\rho, \nabla_x U\rangle_\rho + k_B T_c \|\nabla_x \log\rho\|_\rho^2)\mathrm{d}t, \end{aligned}$$

where the notation $\langle \nabla_x f, \nabla_x g\rangle_\rho := \int_{\mathbb{R}^d} \langle \nabla_x f, \nabla_x g\rangle \rho \mathrm{d}x$ and $\|\nabla_x f\|_\rho = \sqrt{\langle \nabla_x f, \nabla_x f\rangle_\rho}$ is used. By the Cauchy–Schwartz inequality and constraint (49),

$$\begin{aligned} -\langle \nabla_x \log\rho, \nabla_x U\rangle_\rho &\leq \|\nabla_x U\|_\rho \|\nabla_x \log\rho\|_\rho \\ &\leq \sqrt{\gamma M}\|\nabla_x \log\rho\|_\rho. \end{aligned}$$

Hence,

$$\begin{aligned}&\mathcal{S}(\rho_b)-\mathcal{S}(\rho_a)\\ \leq&\frac{k_B}{\gamma}\int_{\frac{t_{\text{cycle}}}{2}}^{t_{\text{cycle}}}\left(\sqrt{\gamma M}\|\nabla_x\log\rho\|_\rho-k_BT_c\|\nabla_x\log\rho\|_\rho^2\right)\mathrm{d}t\\ \leq&\frac{k_B}{\gamma}\int_{\frac{t_{\text{cycle}}}{2}}^{t_{\text{cycle}}}\frac{\gamma M}{4k_BT_c}\mathrm{d}t=\frac{M}{8T_c}t_{\text{cycle}}.\end{aligned}$$

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

$$P\leq\frac{(T_h-T_c)}{t_{\text{cycle}}}\Delta\mathcal{S}-\frac{1}{t_{\text{cycle}}}W_{\text{irr}}\leq\frac{M(T_h-T_c)}{8T_c}\tag{51}$$

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 $U(t,x)=\frac{1}{2}a_tx^2$ is a quadratic function. In this setting, the exact power output is equal to

$$\begin{aligned}P=&\frac{1}{t_{\text{cycle}}}k_B(T_h-T_c)\log(\frac{\sigma_b}{\sigma_a})\\ &-\frac{1}{\gamma t_{\text{cycle}}}\int_0^{t_{\text{cycle}}}(a_t-\frac{k_BT}{\sigma_t^2})^2\sigma_t^2\mathrm{d}t\end{aligned}$$

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

$$\frac{\mathrm{d}\sigma_t^2}{\mathrm{d}t}=-2(\frac{a_t}{\gamma}-\frac{k_BT}{\gamma\sigma_t^2})\sigma_t^2$$

with the constraint (49) given by $\frac{1}{\gamma}a_t^2\sigma_t^2\leq M$. Then, in the limit as $t_{\text{cycle}}\to 0$, and $\sigma_b\to\sigma_a=\sigma$, the power output is equal to

$$P=k_B(T_h-T_c)\frac{\lambda}{2}-\gamma\lambda^2\sigma^2\tag{52}$$

with the constraint

$$|\gamma\lambda+\frac{k_BT_c}{\sigma^2}|\leq\frac{\sqrt{\gamma M}}{\sigma},\tag{53}$$

where we introduced a new variable $\lambda=\frac{a}{\gamma}-\frac{k_BT_c}{\gamma\sigma^2}$. It is shown in Appendix A.4, that the maximum of the expression (52) over all values of λ and σ that satisfy the constraint (53), is equal to the lower-bound. The lower-bound also holds in vector setting by extending this argument and considering a d-dimensional Gaussian distributions with independent components. □

This final result is universal as it does not depend on the choice of ρ_a and ρ_b, unlike (47). Moreover, the bounds in this final result are especially appealing in that it depend on the ratio T_h/T_c of the absolute temperatures of the two heat baths.

Remark 6. It is noted that the upper bound in (50) on achievable power under the constraint (49) 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 7. In the proof of Theorem 2, an operating point has been constructed to ensure that power equal the lower bound in (50) can be achieved. The parameters are given in Eq. (56) in the Appendix. For this operating point, which corresponds to maximal power constrained by (49), 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 given in (35), $\eta_{CA}=1-\sqrt{T_c/T_h}$ 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 $\eta_{SS}\to 2/3$. Interestingly, that η may be larger than η_{SS} is due to the fact it is obtained under an added constraint on the controlling potential, that seeks to maximize power, as compared to η_{SS}; the increase in efficiency is consistent with the inherent trade-off between power and efficiency.

7. 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 (Dechant et al., 2017; Parrondo et al., 2015; Seifert, 2008, 2012; Sekimoto, 2010), 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., Brockett and Willems (1979), Mitter and Newton (2005), Pavon (1989), Rajpurohit and Haddad (2017), Sandberg, Delvenne, Newton, and Mitter (2014) and Wallace (2014). 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 (Aurell et al., 2012, 2011; Chen et al., 2020; Dechant & Sakurai, 2019; Seifert, 2012). 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.

Appendix

A.1. Proof of Lemma 2

Consider an arbitrary smooth vector field ξ on $\mathbb{R}^d$ with bounded support, and $\Psi_s : \mathbb{R}^d \to \mathbb{R}^d$ defined by

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

for $x \in \mathbb{R}^d$ and $s \geq 0$. If $\rho_s := \Psi_s \sharp \rho_b^*$, we claim that

$$\lim_{s\to 0} \frac{1}{s}(f(\rho_s) - f(\rho_b^*)) \geq \int_{\mathbb{R}^d} \langle D_f(x), \xi(x)\rangle \rho_b^*(x)\mathrm{d}x, \tag{54}$$

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 $\xi \to -\xi$,

$$\int \langle D_f(x), \xi(x)\rangle \rho_b^*(x)\mathrm{d}x = 0. \tag{55}$$

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 (44) and the lemma.

It remains to prove (54). By definition,

$$f(\rho_s) - f(\rho_b^*) = \frac{\Delta T}{t_{\text{cycle}}}(\mathcal{S}(\rho_s) - \mathcal{S}(\rho_b^*))$$
$$- \frac{4\gamma}{t_{\text{cycle}}^2}(\mathrm{W}_2(\rho_a, \rho_s)^2 - \mathrm{W}_2(\rho_a, \rho_b^*)^2).$$

The entropy term

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

Therefore

$$\begin{aligned}&\lim_{s\to 0}\frac{1}{s}(S(\rho_s) - S(\rho_b^*)) \\ &= \lim_{s\to 0}\frac{k_B}{s}\int \log(\det(\nabla\Psi_s(x)))\rho_b^*(x)\mathrm{d}x \\ &= k_B \int \nabla\cdot\xi(x)\rho_b^*(x)\mathrm{d}x \\ &= -k_B \int \langle \xi(x), \nabla\log(\rho_b^*(x))\rangle\rho_b^*(x)\mathrm{d}x.\end{aligned}$$

The Wasserstein term

$$\begin{aligned}&\mathrm{W}_2(\rho_a, \rho_s)^2 - \mathrm{W}_2(\rho_a, \rho_b^*)^2 \\ &\leq \int \|\nabla\psi^{-1}(x) - \Psi_s(x)\|^2\rho_b^*(x)\mathrm{d}x \\ &- \int \|\nabla\psi^{-1}(x) - x\|^2\rho_b^*(x)\mathrm{d}x \\ &= \int \langle x - \Psi_s(x), 2\nabla\psi^{-1}(x) - x - \Psi_s(x)\rangle\rho_b^*(x)\mathrm{d}x.\end{aligned}$$

Therefore

$$\begin{aligned}&\lim_{s\to 0}\frac{1}{s}\left[\mathrm{W}_2(\rho_a, \rho_s)^2 - \mathrm{W}_2(\rho_a, \rho_b^*)^2\right] \\ &\leq -2\int \langle \xi(x), \nabla\psi^{-1}(x) - x\rangle\rho_b^*(x)\mathrm{d}x.\end{aligned}$$

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

A.2. Proof of Proposition 3

According to Proposition 1, 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 (45), satisfies the optimality condition (44). When ρ_a, ρ_b^* are Gaussian, $\nabla\psi^{-1}(y) = \frac{\sigma_a}{\sigma_b}y$. Hence, the optimality condition reads

$$\begin{aligned}&\frac{k_B t_{\text{cycle}}\Delta T}{8\gamma}\nabla\log\rho_b^*(y) - y + \nabla\psi^{-1}(y) \\ &= \frac{k_B t_{\text{cycle}}\Delta T}{8\gamma}\frac{y}{\sigma_b^2} - (1 - \frac{\sigma_a}{\sigma_b})y \\ &= (\frac{k_B t_{\text{cycle}}\Delta T}{8\gamma\sigma_b^2} - 1 + \frac{\sigma_a}{\sigma_b})y = 0, \quad \forall\, y \in \mathbb{R},\end{aligned}$$

which is satisfied when σ_b is according to (45).

A.3. Proof of Eq. (48)

The proof follows by expressing the HWI* inequality (11) for a Gaussian reference measure $\mathrm{d}\mathfrak{m}_g = (2\pi\sigma^2)^{-\frac{d}{2}}e^{-\frac{\|x\|^2}{2\sigma^2}}dx$ with constant $\kappa = \frac{1}{\sigma^2}$ and taking the limit as $\sigma \to \infty$. The relative entropy with respect to Gaussian measure is

$$\begin{aligned}H(\mu|\mathfrak{m}_g) &= \int \log(\frac{\mathrm{d}\mu}{\mathrm{d}x})\mathrm{d}\mu - \int \log(\frac{\mathrm{d}\mathfrak{m}_g}{\mathrm{d}x})\mathrm{d}\mu \\ &= -k_B^{-1}\mathcal{S}(\frac{\mathrm{d}\mu}{\mathrm{d}x}) + \frac{\sigma_{\mu_a}^2}{2\sigma^2} + \frac{d}{2}\log(2\pi\sigma^2).\end{aligned}$$

where $\sigma_\mu^2 := \int \|x\|^2\mathrm{d}\mu$. Therefore, the left hand side of (11)

$$H(\mu_a|\mathfrak{m}_g) - H(\mu_b|\mathfrak{m}_g) = k_B^{-1}(\mathcal{S}(\rho_b) - \mathcal{S}(\rho_a)) + \frac{\sigma_{\mu_a}^2 - \sigma_{\mu_b}^2}{2\sigma^2},$$

with $\rho_a = \mathrm{d}\mu_a/\mathrm{d}x$, $\rho_b = \mathrm{d}\mu_b/\mathrm{d}x$, and $\sigma_{\mu_a}^2, \sigma_{\mu_b}^2$ are the corresponding variances. On the right hand side, the Fisher information term becomes

$$\begin{aligned}I(\mu_a|\mathfrak{m}_g) &= \int \|\nabla\log(\frac{\mathrm{d}\mu_a}{\mathrm{d}\mathfrak{m}_g})\|^2\mathrm{d}\mu_a \\ &= \int \|\nabla\log(\frac{\mathrm{d}\mu_a}{\mathrm{d}x}) - \nabla\log(\frac{\mathrm{d}\mathfrak{m}_g}{\mathrm{d}x})\|^2\mathrm{d}\mu_a \\ &= \int \|\nabla\log(\frac{\mathrm{d}\mu_a}{\mathrm{d}x})\|^2\mathrm{d}\mu_a - 2\int \langle\nabla\log(\frac{\mathrm{d}\mu_a}{\mathrm{d}x}), \frac{-x}{\sigma^2}\rangle\mathrm{d}\mu_a \\ &+ \int \|\frac{-x}{\sigma^2}\|^2\mathrm{d}\mu_a = I(\mu_a|dx) - \frac{2d}{\sigma^2} + \frac{\sigma_{\mu_a}^2}{\sigma^4}\end{aligned}$$

Thus, taking the limit $\sigma \to \infty$, (48) follows.

A.4. Proof of the lower-bound in Theorem 2

The constraint (53) is expressed as:

$$0 \le \lambda \le \frac{\sqrt{\gamma M}}{\gamma\sigma} - \frac{k_BT_c}{\gamma\sigma^2}, \text{ for } \sigma \ge \frac{k_BT_c}{\sqrt{\gamma M}}.$$

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

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

for $\sigma_0 := k_BT_c/\sqrt{\gamma M}$, $\lambda_0 := M/k_BT_c$, and re-write (52) and the constraints,

$$P = Mf(x,y)$$
$$0 \le x \le g(y), \quad 0 < y \le 1$$

where $f(x,y) = \frac{\Delta T}{2T_c}x - \frac{x^2}{y^2}$, $g(y) = y - y^2$. As long as $y \le y_0$, 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 $x^*(y) \le g(y)$. When $y_0 < y \le 1$, the maximizer is at $x = g(y)$. Hence,

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

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

$$\sigma = \frac{k_BT_c}{\sqrt{\gamma M}}\frac{2(T_h+T_c)}{(T_h+3T_c)}, \ \lambda = \frac{M}{k_BT_c}\frac{(T_h+3T_c)(T_h-T_c)}{4(T_h+T_c)^2}. \tag{56}$$

References

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

Argun, Aykut, Soni, Jalpa, Dabelow, Lennart, Bo, Stefano, Pesce, Giuseppe, Eichhorn, Ralf, et al. (2017). Experimental realization of a minimal microscopic heat engine. *Physical Review E*, *96*(5), Article 052106.

Aurell, Erik, Gawędzki, Krzysztof, Mejía-Monasterio, Carlos, Mohayaee, Roya, & Muratore-Ginanneschi, Paolo (2012). Refined second law of thermodynamics for fast random processes. *Journal of Statistical Physics*, *147*(3), 487–505.

Aurell, Erik, Mejía-Monasterio, Carlos, & Muratore-Ginanneschi, Paolo (2011). Optimal protocols and optimal transport in stochastic thermodynamics. *Physical Review Letters*, *106*(25), Article 250601.

Benamou, Jean-David, & Brenier, Yann (2000). A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, *84*(3), 375–393.

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

Brockett, Roger W., & Willems, Jan C. (1979). Stochastic control and the second law of thermodynamics. In *1978 IEEE conference on decision and control including the 17th symposium on adaptive processes* (pp. 1007–1011). IEEE.

Callen, Herbert B. (1998). *Thermodynamics and an introduction to thermostatistics*. AAPT.

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

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

Chambadal, Paul (1957). Les centrales nucléaires. *Manuales Ingenieria*, *4*, 1–58.

Chen, Yongxin, Georgiou, Tryphon T., & Pavon, Michele (2016). On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint. *Journal of Optimization Theory and Applications*, *169*(2), 671–691.

Chen, Yongxin, Georgiou, Tryphon, & Tannenbaum, Allen (2020). Stochastic control and non-equilibrium thermodynamics: fundamental limits. *IEEE Transactions on Automatic Control*, *65*(7), 2979–2991.

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

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

De Groot, Sybren R., & Mazur, Peter (2013). *Non-equilibrium thermodynamics*. Courier Corporation.

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

Dechant, Andreas, & Sakurai, Yohei (2019). Thermodynamic interpretation of wasserstein distance. arXiv preprint arXiv:1912.08405.

Esposito, Massimiliano, Kawai, Ryoichi, Lindenberg, Katja, & Van den Broeck, Christian (2010a). Efficiency at maximum power of low-dissipation Carnot engines. *Physical Review Letters*, *105*(15), Article 150603.

Esposito, Massimiliano, Kawai, Ryoichi, Lindenberg, Katja, & Van den Broeck, Christian (2010b). Quantum-dot carnot engine at maximum power. *Physical Review E*, *81*(4), Article 041106.

Fu, Rui, Taghvaei, Amirhossein, Chen, Yongxin, & Georgiou, Tryphon T (2020). Maximal power output of a stochastic thermodynamic engine. arXiv preprint arXiv:2001.00979.

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

Gomez-Marin, Alex, Schmiedl, Tim, & Seifert, Udo (2008). Optimal protocols for minimal work processes in underdamped stochastic thermodynamics. *Journal of Chemical Physics*, *129*(2), Article 024114.

Graham, R. (1978). Path-integral methods in nonequilibrium thermodynamics and statistics. In *Stochastic processes in nonequilibrium systems* (pp. 82–138). Springer.

Horowitz, J., & Jarzynski, C. (2008). Comment on "Failure of the work-Hamiltonian connection for free-energy calculations". *Physical Review Letters*, *101*(9), Article 098901.

Jarzynski, Christopher (1997a). Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Physical Review E*, *56*(5), 5018.

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

Jordan, Richard, Kinderlehrer, David, & Otto, Felix (1998). The variational formulation of the Fokker–Planck equation. *SIAM Journal on Mathematical Analysis*, *29*(1), 1–17.

Kurchan, Jorge (1998). Fluctuation theorem for stochastic dynamics. *Journal of Physics A: Mathematical and General*, *31*(16), 3719.

Lebon, Georgy, Jou, David, & Casas-Vázquez, José (2008). *Understanding non-equilibrium thermodynamics, Vol. 295*. Springer.

Melbourne, James, Talukdar, Saurav, & Salapaka, Murti V. (2018). Realizing information erasure in finite time. In *2018 IEEE conference on decision and control (CDC)* (pp. 4135–4140). IEEE.

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

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

Otto, Felix, & Villani, Cédric (2000). Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *Journal of Functional Analysis*, *173*(2), 361–400.

Owen, David R. (2012). *A first course in the mathematical foundations of thermodynamics*. Springer Science & Business Media.

Park, Jong-Min, Chun, Hyun-Myung, & Noh, Jae Dong (2016). Efficiency at maximum power and efficiency fluctuations in a linear Brownian heat-engine model. *Physical Review E*, *94*(1), Article 012127.

Parrondo, Juan M. R., Horowitz, Jordan M., & Sagawa, Takahiro (2015). Thermodynamics of information. *Nature Physics*, *11*(2), 131.

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

Pavon, Michele, & Ticozzi, Francesco (2006). On entropy production for controlled Markovian evolution. *Journal of Mathematical Physics*, *47*(6), Article 063301.

Peliti, Luca (2008a). Comment on "Failure of the work-Hamiltonian connection for free-energy calculations". *Physical Review Letters*, *101*(9), Article 098903.

Peliti, Luca (2008b). On the work–Hamiltonian connection in manipulated systems. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(05), P05002.

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

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

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

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

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

Seifert, Udo (2012). Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on Progress in Physics*, *75*(12), Article 126001.

Sekimoto, Ken (2010). *Stochastic energetics, Vol. 799*. Springer.

Talukdar, Saurav, Bhaban, Shreyas, & Salapaka, Murti V. (2017). Memory erasure using time-multiplexed potentials. *Physical Review E*, *95*(6), Article 062121.

Vilar, Jose M. G., & Rubi, J. Miguel (2008). Failure of the work-Hamiltonian connection for free-energy calculations. *Physical Review Letters*, *100*(2), Article 020601.

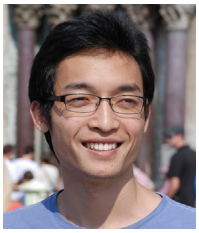
Villani, Cédric (2003). *Graduate studies in mathematics*: *vol. 58*, *Topics in optimal transportation*. American Mathematical Soc..

Villani, Cédric (2008). *Optimal transport: old and new, Vol. 338*. Springer Science & Business Media.

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

Rui Fu obtained her Bachelor's degree and Master's degree both in Applied Mathematics from Xuchang University and Northwestern Polytechnical University, China, respectively, and second Master's degree in Mechanical and Aerospace Engineering from University of California, Irvine. She is currently working towards the Ph.D. degree in the Department of Mechanical and Aerospace Engineering, University of California, Irvine.

Amirhossein Taghvaei is a Postdoctoral Researcher in the Department of Mechanical and Aerospace Engineering at University of California Irvine. He obtained his Ph.D. in Mechanical Science and Engineering, and M.S in Mathematics from University of Illinois at Urbana-Champaign. He is currently working in the area of Control theory and Machine learning.



Yongxin Chen received his B.Sc. from Shanghai Jiao Tong University in 2011 and Ph.D. from University of Minnesota in 2016, both in Mechanical Engineering. He is currently an Assistant Professor in the School of Aerospace Engineering at Georgia Institute of Technology. He has served on the faculty at Iowa State University (2017–2018). He received the George S. Axelby Best Paper Award in 2017 for his joint work with Tryphon Georgiou and Michele Pavon. His current research focuses on the intersection of control theory, machine learning, robotics and optimization.

Tryphon T. Georgiou (M'79–SM'99–F'00) received the Diploma in Mechanical and Electrical Engineering from the National Technical University of Athens, Greece, in 1979 and the Ph.D. degree from the University of Florida, Gainesville, FL, USA, in 1983. He is currently a Distinguished Professor with the Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA, USA. Earlier, he served on the faculty of Florida Atlantic University from 1983 to 1986, Iowa State University from 1986 to 1989, and the University of Minnesota from 1989 to 2016. He is a recipient of the George S. Axelby Outstanding Paper award of the IEEE Control Systems Society for the years 1992, 1999, 2003, and 2017, a Fellow of the Institute of Electrical and Electronic Engineers (IEEE) and the International Federation of Automatic Control (IFAC), and a Foreign Member of the Royal Swedish Academy of Engineering Sciences (IVA).