Minimal Work Principle and its Limits for Classical Systems.

A.E. Allahverdyan1), Th.M. Nieuwenhuizen2)
1) Yerevan Physics Institute, Alikhanian Brothers St. 2, Yerevan 375036, Armenia,
2) Institute for Theoretical Physics, Valckenierstraat 65, 1018 XE Amsterdam, The Netherlands,
(Dated: March 23, 2022)

The minimal work principle asserts that work done on a thermally isolated equilibrium system, is minimal for the slowest (adiabatic) realization of a given process. This principle, one of the formulations of the second law, is operationally well-defined for any finite (few particle) Hamiltonian system. Within classical Hamiltonian mechanics, we show that the principle is valid for a system of which the observable of work is an ergodic function. For non-ergodic systems the principle may or may not hold, depending on additional conditions. Examples displaying the limits of the principle are presented and their direct experimental realizations are discussed.

PACS numbers: 05.70.Ln, 05.10.Ln

Thermodynamics originated in the nineteenth century as a science of macroscopic machines, constructed for transferring, applying or transmitting energy. The main results of this science are summarized in several formulations of the second law. Last fifty years witnessed progressive miniaturization of the components employed in the construction of devices and machines. This will open the way to new technologies in the fields of medicine, computation and renewable energy sources.

For microscopic machines and devices, we need to understand how the second law applies to small systems. There are two aspects of this program: i) Emergence of the second law, where one studies fluctuations of work or entropy, knowing that on the average they satisfy the second law. Important contributions to this topic were made by Smoluchowski and others, nearly hundred years ago. Recently, the activity in this field revived in the context of fluctuation theorems. ii) Limits of the second law, where the very formulation is studied from first principles. Here we continue on that and study (limits of) the minimal work principle based on classical mechanics. This formulation of the second law is operationally well-defined for finite systems, and it relates the energy cost of an operation to its speed. The principle was deduced from experience and postulated in thermodynamics, where it is equivalent to other formulations of the second law. Its derivation in statistical physics was formulated several times, on various levels of generality. Almost all these studies concentrate on macroscopic systems and confirm the validity of the principle. In Ref. the principle was derived for a finite quantum systems, and limits, related to energy level crossing, were indicated. However, the reduction level of quantum mechanics is not always needed, e.g., certain aspects of nanoscience are adequately understood already with classical ideas. In addition, it is not easy to design experiments in the quantum domain that would check the validity of formulations of the second law. Thus it is necessary to understand the principle in classical mechanics and this is our present purpose. Our results will apply to small systems, having one degree of freedom.

Consider a classical system with $N$ degrees of freedom and with a Hamiltonian $H(q,p,R_t)$, where $q = (q_1, ..., q_N)$ and $p = (p_1, ..., p_N)$ are, respectively, canonical coordinates and momenta. The interaction with external sources of work is realized via a time-dependent parameter $R_t = r(t/\tau_R)$ with time-scale $\tau_R$. The motion starts at the initial time $t_i = \tilde{t}_i\tau_R$, and we follow it till the final time $t_f = \tilde{t}_f\tau_R$. We shall denote $z \equiv (q,p)$, $dz \equiv dq dp$ and $H_k = H(z,R_k)$ with $k = i, f$, for the initial and final values, respectively.

The system starts its evolution from an equilibrium Gibbs state at temperature $T = 1/\beta > 0$: $P(z) = e^{-\beta H(z)}/Z$, where $Z = \int dz e^{-\beta H(z)}$. The work done on the system is the average energy difference:

$$W = \int dz [P_i(z)H_i(z) - P_f(z)H_f(z)] = \int_{t_i}^{t_f} ds R_s P(z,s) \partial_R H(z,R_s),$$

where $P_i(z)$ is the final distribution, and where the equivalence between and is established with help of the Liouville equation for $P$ and the standard boundary condition $P(z) = 0$ for $z \to \pm \infty$. Eq. justifies to call $w(z,R) \equiv \partial_R H(z,R)$ “the observable of work”. Consider processes with different $\tau_R$, but the same $r(\tilde{t})$ and $R_s = r(t_k)$ ($k = i, f$). The minimal work principle claims

$$\Delta W = W - \bar{W} \geq 0,$$

where $\bar{W}$ is the work for the adiabatically slow realization, where $\tau_R$ is much larger than the characteristic time $\tau_S$ of the system. Eq. is an optimality statement: the smallest amount of work to be put into the system ($W > 0$) is the adiabatic one, and the largest amount of work to be extracted from the system ($W < 0$) is again the adiabatic one. The condition of being slow is purely operational. If there exist several widely different characteristic times, there are several senses of being slow.

Our derivation of the principle consists of three steps and assumes that $w(z,R)$ is an ergodic observable of the
dynamics with \( R = \text{const} \). Note that the known inequality \( W \geq F_t - F_i \), where \( F_t = -T \ln \int dz e^{-H_i(z)/T} \) are free energies referring to the initial temperature \( T \), does not in general provide any information about the principle, since for finite systems the adiabatic work is not equal to \( F_t - F_i \); see [11] for examples, and also below.

1. The initial distribution \( P_i(z) \) can be generated by sampling microcanonical distribution \( M \) with initial energy probability \( P_i(E) \),

\[
M(z, E, R_i) = \frac{1}{\omega_i(E)} \delta[E - H(z, R_i)],
\]

\[
P_i(E) = \frac{\omega_i(E)}{Z_i} e^{-\beta E}, \quad \omega_i(E) \equiv \int dz \delta[E - H_i(z)],
\]

\[
P_i(z) = \int dE P_i(E) M(z, E, R_i).
\]

Hamilton’s equations of motion imply \( \frac{d}{dt} H(z_t, R_t) = \frac{\partial}{\partial z} H(z_t, R_t) \). On times \( \tau_S \ll \tau \ll \tau_R \) we have for the energy change

\[
\Delta_r E = \frac{1}{\tau} [H(z_{t+\tau}, R_{t+\tau}) - H(z_t, R_t)] = \int_t^{t+\tau} d\tau \frac{dH}{dz}(z, R_s) = \frac{R_t}{\tau} \int_t^{t+\tau} d\tau \frac{\partial H}{\partial z}(z, R_t) + o(\tau).
\]

The last integral refers to the frozen-parameter dynamics with \( R_t = R \). Now recall the Liouville theorem \( dz = dz \) and energy conservation \( H(z_{t+\tau}) = H(z_t) = E_t \), so that

\[
\int dz w(z) M(z, E_t) = \frac{1}{\tau} \int_t^{t+\tau} d\tau \int dz w(z) M(z, E_t) = \int dz \frac{M(z, E_t)}{\tau} \int_t^{t+\tau} d\tau w(z)[z_s; z_t],
\]

where \( z[z; z_t] \) is the trajectory at time \( s \), that starts at \( z_t \) at time \( t \). If \( w(z, R) = \partial H(z, R) \) is an ergodic observable of the \( R = \text{const} \) dynamics, then for \( \tau \gg \tau_S \) the time-averaging in [8] does not depend on the initial condition \( z_t \) [12], so the integration over \( z_t \) in [8] is trivial, and we get from [10] that the time-average in [10] is equal to the microcanonical average at the energy \( E_t \):

\[
\Delta_r E = \frac{1}{\tau} \int_t^{t+\tau} d\tau \frac{\partial H}{\partial z}(z, R_t) + o(\tau).
\]

2. Let \( P(E|E') \) be the conditional probability of having energies \( E \) and \( E' \) at \( t = t_f \) and \( t = t_i \), respectively:

\[
P(E|E')P_i(E') = \int d\zeta d\zeta' \delta[E - H_i(\zeta)]\delta[E' - H_i(\zeta')]P(\zeta|\zeta'),
\]

where \( P(\zeta|\zeta') \) is the phase-space conditional probability

\[
P(\zeta|\zeta') = U \delta[p - p'] \delta[q - q'] \quad \text{for} \quad U \equiv \int dz \frac{P(E)}{\omega_i(E)} \int d\zeta E',
\]

and where \( \omega_i(E) \equiv \partial H(z, R) \partial H(z, R) \) is the Liouville operator, with \( \omega_i \) being the chronological exponent.

\[
P(E|E') = \int d\zeta d\zeta' \delta[E - H_i(\zeta)]\delta[E' - H_i(\zeta')]P(\zeta|\zeta'),
\]

in addition to the normalization \( \int P(E|E') \int dE' = 1 \). For the adiabatic situation we get from conservation of \( \Omega \):

\[
\Omega(E|E') = \delta[E - \phi_i(E')],
\]

3. Now recall [11] and write \( \Delta W \) as

\[
\Delta W = \int dz H_i(z)[P_i(z) - \tilde{P}_i(z)] = \int E dE \int dE' P_i(E')[P(E|E') - \tilde{P}(E|E')] = \int dE g_E.
\]

Integrating by parts, denoting \( c_E(E') = \int^E d\nu P(u|E') \),

\[
ge_E = \int^E d\nu \int dE' P_i(E') \tilde{P}(u|E') - P(u|E')
\]

\[
= \int^E d\nu \int dE' P_i(E') - \int dE' P_i(E') c_E(E') = \int dE' P_i(E') \int^E d\nu \omega_i[E'] \int dE' c_E(E').
\]

We now employ [13] and then [12] to obtain

\[
Z_g \geq e^{-\beta \phi_i(E)} \int^{\phi_i(E)} dE' \omega_i[E'] \int dE' [1 - c_E(E')]
\]

\[
- e^{-\beta \phi_i(E)} \int^{\phi_i(E)} dE' \omega_i[E'] c_E(E')
\]

\[
= e^{-\beta \phi_i(E)} \int^{\phi_i(E)} dE' \omega_i[E'] - \int^E dE' \omega_i[E'].
\]

Recalling that \( \omega(E) = \Omega(E) \) and that \( \Omega(E_{\text{min}}) = 0 \) for the lowest energy \( E_{\text{min}} \), we get [15] \( = 0 \), i.e., \( g_E \geq 0 \). This means, from [13], that the principle is proven. Note that
the same proof applies for \( P_i(E)/\omega_i(E) \) being a monotonically decaying function of \( E \). Neither the proof, nor the principle itself, applies to the initial microcanonical situation, where \( P_i(E) = \delta[E - E_i] \).

For non-ergodic systems, where under driving the system can move from one ergodic component (of the \( R = \text{const} \) dynamics) to another, the above proof of the principle is endangered, since in general \( \Omega \) in \( (1) \) is not conserved. Indeed, the argument expressed by \( (14) \) may not apply, since now the time-average in \( (3) \) depends on the ergodic component to which the initial condition \( z_t \) belongs, and in general it cannot be substituted by the microcanonical average over the full phase-space. However, \( \partial_R H(z, R) \) may be ergodic, even though the system is not \( (12) \). For an example consider the symmetric double-well: \( H = \frac{1}{2}p^2 - R_xx^2 + g|x|^3 \), with \( g > 0 \). For \( E < 0 \), there are two ergodic components related by the inversion \( x \to -x \), but \( \partial_R H(z, R) = -2x^2 \) is degenerate with respect to them. Though the \( R_t = \text{const} \) motion on the separatrix \( E = 0 \) has an infinite period (due to unstable fixed point \( q = 0 \)), when the initial distribution is microcanonical, the fraction of particles trapped by the separatrix is negligible \( (13) \), so that for the ensemble \( \tau_S \) remains finite. Thus \( \Omega(E, R) \) (with the integration in \( (13) \) over the whole phase-space) is conserved \( (13) \). Thus for this case the above proof of the principle applies.

**Limits of the principle.** The above derivation of the principle assumes that the frozen-parameter dynamics supports the microcanonical distribution. This need not be always the case. Consider the basic model of the parametric oscillator: \( H = \frac{1}{2}p^2 + \frac{1}{2}R_qq^2 \). If \( R_q \) is always positive, the phase-space volume is conserved, and the above construction applies. But what happens if \( R_q \) touches zero at one instant? This is another non-ergodic example, since for \( R_q = 0 \) the frozen-parameter phase-space consists of two ergodic components with, respectively, \( p > 0 \) and \( p < 0 \). Though \( \partial_R H(z, R) = x^2/2 \) is degenerate with respect to them, the microcanonical distribution does not exist for \( R_q = 0 \).

To separate the effect of initial conditions, we write the solution of the equations of motion \( \hat{q} + R_q \hat{p} = 0 \) as

\[
p(t) = \theta_{pp}p_i + \theta_{pq}q_i, \quad q(t) = \theta_{qq}q_i + \theta_{qp}p_i,
\]

where \( \theta_{kl} = \theta_{kl}(t) \). With Gibbssian initial distribution \( P_i(p, q) \propto \exp[-\frac{1}{2} \theta_{pq}p_i - \frac{1}{2} \theta_{qq}q_i] \), the work reads from \( (1) \):

\[
W \frac{T}{T} = -1 + \frac{1}{2} \theta_{pp}^2 + \frac{1}{2} R_t \left( \theta_{pp}^2 + \frac{\theta_{pq}^2}{R_t} \right).
\]

Next we consider an exactly solvable situation \( R_t = t^2/\tau^2 \) (see below for generalizations). The equation of motion is solved by substitution \( q(t) = \sqrt{T}x(t) \):

\[
q(t) = c_1 \sqrt{\frac{1}{t}} J_{-1/4} \left( \frac{t^2}{2 \tau R} \right) + c_2 \frac{t}{\sqrt{T}} J_{1/4} \left( \frac{t^2}{2 \tau R} \right),
\]

where \( J_{\pm 1/4} \) are the Bessel functions, \( c_1 \) and \( c_2 \) are to be found from initial conditions, and where \( q(t) \) is written in a way that applies to \( t < 0 \): noting \( J_{-1/4}(x) = J_{1/4}(2|x|) \), we see that \( q(t) \approx c_1 + c_2 |t| \) near \( t = 0 \). Let us define \( t_i = -\sqrt{\tau R}/R_t \) and \( t_f = \tau R/\sqrt{R_t} \); in the slow limit \( \tau_R \gg 1 \) we need \( J_{1/4}(x) \approx \tau R \approx 4 \), as given by Eq. (23). Dashed line: \( t = 0.5 \), \( t_i = -\tau_R \) and \( t_f = \tau R \sqrt{0.5} \). Bold line: \( t = 0.5 \), \( t_i = -\tau_R \) and \( t_f = \tau R \). For the last two cases the adiabatic work (the limit \( \tau_R \to \infty \)) is, respectively, plus and minus infinity. The minimal work principle does not hold. Indeed, for a sudden change \( t \) to \( R_t = t^2/\tau^2 \) the adiabatic work (the limit \( \tau_R \to \infty \)) being a monotonically decaying function of \( E \) is sometimes smaller than (23), e.g., take \( R_t = R_t \), but

\[
W \frac{R_t}{2R_t} = \frac{1}{2} \sqrt{R_t - \frac{1}{2}} \).
\]
back (for $R_t > 0$), the work to be spent is larger for the slow case, since for a quick process the particle does not have time to move very far; see [24].

An extended setup $R_t = -b + t^2/\tau_R^2$, $b \geq 0$, can be worked out with help of hypergeometric functions. Fig. 4 illustrates that the principle is satisfied if $R_t$ decays monotonically: $t_1 < t_t \leq 0$. It is violated if the change of $R_t$ is non-monotonic: $t_1 < 0 < t_t$. The work does not saturate for $\tau_R \to \infty$ if $R_t$ becomes strictly negative. We checked numerically that $i)$ if $R_t$ is positive, but becomes very small, $R = \delta$, the system has two widely different characteristic times $\tau_R \sim 1$ and $\tau_\delta$ that diverges for $\delta \to 0$. Thus there are two senses of being slow: the principle is limited for $\tau_S \ll \tau_R \ll \tau_\delta$, while it is satisfied if $\tau_\delta \ll \tau_R$, since then the phase-space volume is conserved, as discussed above. $ii)$ The limits of the principle exist for any potential $U(q, R_t)$ that is globally confining ($U(q) \to \infty$ for $q \to \pm \infty$) but looses this feature for one value of $R_t$. These limits obviously exist for uncoupled particles. We expect that they extend to coupled particles put in a (de)confining potential.

Note that whenever the principle gets limited via the above scenario, the slowest process is irreversible. Recall that a process is reversible, if after supplementing it with its mirror reflection (the same process moved backwards with the same speed), the work done for the total cyclic process is zero [1]. As seen from (24), the work (equal to $2\mathcal{W}$) does not vanish for the cyclic adiabatic process with $R_t$ touching zero. Thus the process is irreversible. This fact contrasts the quantum limits of the minimal work principle found in (7). Those limits are related to energy level-crossing, where the adiabatic work is reversible (7).

Here is an experimentally realizable example that can demonstrate the above limits. The simplest LC circuit consists of capacitance $C$ and inductance $L$ (the resistance is either small or compensated) [10]. The Hamiltonian is $H = \frac{Q^2}{2C} + \frac{L}{2} \frac{d^2Q}{dt^2}$, where $Q$ (coordinate) is the charge, and where $\Phi = L \frac{dQ}{dt}$ (momentum) is the magnetic flux. The parametric oscillator with $R_t \to 0$ corresponds to a time-dependent $C$ (or $L$) becoming very large at some time. $R_t$ becoming negative at some time, can be achieved via a negative capacitance $C_n < 0$ given by a special active circuit [17]. If such a capacitance is sequentially added to a positive capacitance $C_a$, then the resulting inverse capacitance $C^{-1} = C^{-1}_a + C^{-1}_n$ can be made zero and then negative by tuning $C_a$. The same effect is obtained via a negative inductance [18] added in parallel to a normal one. As the negative inductance/capacitance emulators are widely applied in compensation of parasitic processes and for improving the radiation pattern in antennas, [16] [17] [18] they can serve to test our predictions.

In conclusion, we studied the second law in its minimal work formulation for classical Hamiltonian systems. It was shown to hold under the assumption that the observable of work (i.e., the derivative of the Hamiltonian w.r.t. the driven parameter) is an ergodic function. The result applies to small systems. There are, however, numerous examples of non-ergodicity both for finite and macroscopic systems. For such systems we explored several possibilities met in the one degree of freedom situation. The minimal work principle applies if the observable of work is degenerate over ergodic components and if the microcanonical equilibrium exists for all values of the driven parameters. If the latter condition is not met, the principle gets limits. The simplest example of the latter is provided by a parametrically driven harmonic oscillator whose frequency passes through zero. As we saw, this situation can be realized experimentally in LC electrical circuits. Multi-dimensional systems provide more complex examples of non-ergodicity. The understanding of the second law for such systems still deserves to be deepened, in view of the importance of non-ergodicity in processes of measurement and information storage [19].

A.E. A. was supported by CRDF grant ARP2-2647-YE-05 and partially by FOM/NWO.

[1] R. Balian, From Microphysics to Macrophysics, volume I (Springer, 1992). G. Lindblad, Non-Equilibrium Entropy and Irreversibility (D. Reidel, Dordrecht, 1983).
[2] V. Balzani, A. Credi and M. Venturi, Molecular Devices and Machines (Wiley-VCH, Weinheim, 2003).
[3] P.S. Epstein, Textbook of Thermodynamics, (Wiley & Sons, New York, 1937).
[4] G.N. Bochkov and Yu.E. Kuzovlev, Sov. Phys. JETP, 45, 125 (1977), C. Jarzynski, Phys. Rev. Lett. 78, 2690 (1997). For a review see J. Kurshan, cond-mat/0511073.
[5] H. Narnhofer and W. Thirring, Phys. Rev. A, 26, 3646 (1982), L.P. Kadanoff and P.C. Martin, Ann. Phys., 24, 419 (1963). D. Forster, Hydrodynamics, Broken Symmetry, and Correlation Functions, (Benjamin, New York, 1983). R. Fukuda, Prog. Theor. Phys. 77, 825 (1987). K. Sekimoto and S. Sasa, J. Phys. Soc. Jpn. 66, 3326 (1997).
[6] H. Tasaki, cond-mat/0009206 C. Maes and H. Tasaki, cond-mat/0511419.
[7] A. E. Allahverdyan and Th.M. Nieuwenhuizen, Phys. Rev. E 71, 046107 (2005).
[8] K. Sato et al., Phys. Rev. E 66, 016119 (2002).
[9] P. Hertz, Ann. Phys. (Leipzig) 33, 225 (1910); ibid. 33, 537. T. Kasuga, Proc. Jpn. Acad. 37, 366 (1961).
[10] R. Becker, Theory of Heat (Springer, New York, 1967). V. L. Berdichevsky, Thermodynamics of Chaos and Order, (Addison Wesley Longman, Essex, England, 1997). S. Sasa and T.S. Komatsu, Prog. Theor. Phys. 103, 1 (2000). H.H. Rugh, Phys. Rev. E 64, 055101 (2001).
[11] E. Ott, Phys. Rev. Lett. 42, 1628 (1979). C. Jarzynski, Phys. Rev. Lett. 71, 839 (1993).
[12] N. G. van Kampen, Physica 53, 98 (1971).
[13] R.W.B. Best, Physica 40, 182 (1968).
[14] F.W.J. Olver, Asymptotics and Special Functions (Aademic Press, New York, 1974).
[15] A.E. Allahverdyan and Th.M. Nieuwenhuizen, Physica A
[16] L. Chua, C.A. Desoer, and E.S. Kuh, *Linear and Nonlinear circuits* (McGraw-Hill, New-York, 1987).
[17] A.G. Butkovskii and Yu.I. Samoilenko, *Control of Quantum Mechanical Processes* (Springer, Berlin, 1991).
[18] K.L. Su, IEEE Journal of Solid State Circuits, **2**, 22 (1967). H. Funato et al., IEEE Trans. Power Electronics, **12**, 589 (1997).
[19] A.E. Allahverdyan, R. Balian and Th.M. Nieuwenhuizen, Europhys. Lett., **61** 452, (2003). J. Parrondo, Chaos, **11**, 725 (2001). S. Ishioka and N. Fuchikami, *ibid*. 734.