A Cost / Speed/ Reliability Trade-off in Erasing a Bit.

Manoj Gopalkrishnan
School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai, India.
(Dated: 7 October 2014)

We present a novel treatment of the fundamental problem of erasing a bit. We propose a cost function that extends to path space the principle due to Szilard and Landauer that erasing a bit requires at least $k_B T \log 2$ units of energy. We introduce notions of reliability of information storage via a reliability timescale $\tau_r$, and speed of erasing via an erasing timescale $\tau_e$. Our problem formulation captures the tradeoff between speed, reliability, and the action required to erase a bit.

We show that erasing a reliable bit fast costs at least $\log 2 - \frac{1}{2} \log \left( 1 - e^{-2\tau_r/\tau_e} \right) > \log 2$ units, which goes to $\frac{1}{2} \log \frac{2\tau_e}{\tau_r}$ units when $\tau_r >> \tau_e$.

I. MOTIVATION

Consider a device to store a single bit of information — let us call it a Brownian bit — with two symmetric positions “0” and “1.” We will describe a Brownian bit as a physical system in the non-equilibrium statistical mechanics sense. It will be modeled by a Langevin equation on state space $\mathcal{S}$, augmented by labeling some states by “0,” and others by “1.” For example, a symmetric bistable well with the left well labeled “0” and the right well labeled “1” is a Brownian bit. Another example is the symmetric two-state Markov chain with the two states labeled “0” and “1” respectively.

By erasing a Brownian bit, we mean the operation of resetting it to state “0” from a maximally unknown (i.e., equilibrium) initial state. What is being erased is randomness, just as when one erases a blackboard with some chalk marks on it, one resets the blackboard to the state of no chalk marks.

Szilard [2] and later Landauer [3] have argued from the second law of thermodynamics that erasing at temperature $T$ requires at least $k_B T \log 2$ units of energy, where $k_B$ is Boltzmann’s constant. More rigorous and general versions of this calculation are known, which also clarify why this is a lower bound [4,5].

For a Brownian bit to be useful in information processing, we ask that it satisfy two additional requirements.

- Erasing, being irreversible, requires the Brownian bit to be coupled to a heat bath. The heat bath causes randomization of information in the Brownian bit. We ask that our device must hold its information for a specified long time $\tau_e$. This is a reliability requirement. It may be achieved for the bistable well by increasing the height of the barrier between the wells.
- We ask that the designer of the Brownian bit provide a means for erasing the device fast, i.e. within a specified short time $\tau_r$. This is a speed requirement.

The requirement specifications for the Brownian bit consist of the operating temperature $T$, the required reliability timescale $\tau_r$, and the required timescale of erasing $\tau_e$. The designer of the device is free to choose any design compatible with the laws of physics. In particular, we do not a priori rule out systems that are perfectly isolated from a heat bath.

We ask: How much does it cost to erase a reliable Brownian bit fast? This question prompts the sub-question: What is the right cost function for this problem? Our main contributions are:

- We pose the question of a fundamental tradeoff between speed, reliability and cost in erasing a bit.
- By formally extending the Szilard-Landauer principle to path space, we obtain a relative entropy on path space as the cost function in Section III B. This choice turns out to be mathematically natural, and yields various statistical, informational, and thermodynamic interpretations which we point out in Section IV.
- We conjecture in Section III that the two-state Markov chain is the optimal Brownian bit. We solve the problem of optimal erasing for a two-state Markov chain in Section III C, obtaining a bound strictly greater than $\log 2$ that does not depend on technology limitations.

II. SZILARD-LANDAUER CORRESPONDENCE

Consider a physical system with state space $S$ and energy $E : S \rightarrow \mathbb{R}$. Define the Gibbs distribution $\pi$ at temperature $T$ by $\pi(i) = e^{-E_i/k_B T} / \left( \sum_{j \in S} e^{-E_j/k_B T} \right)$ for $i \in S$, and the free energy $F(p) := \sum_{i \in S} p_i E_i - k_B T \sum_{i \in S} p_i \log \frac{p_i}{\pi(i)}$, where $p$ is a probability distribution. (We work with finite state spaces for pedagogic reasons; the results hold more generally.)

Define the relative entropy $D(p||q) = \sum_{i \in S} p_i \log \frac{p_i}{q_i}$ with Euler’s constant for the base of the logarithm. Assuming that we have zero information about the system when it is at equilibrium, the relative entropy $D(p||\pi)$ quantifies the amount of information we know about the
system when it is in state $p$ [5].

The following identity is easily verified:

$$F(p) - F(\pi) = k_B T D(p\|\pi).$$  \hspace{1cm} (1)

This identity supplies a dictionary between thermodynamics and information theory [5]. In particular, erasing a bit corresponds to increasing relative entropy which in turn corresponds — via the identity — to increasing available free energy $F(p) - F(\pi)$ by $k_B T \log 2$, recovering the classical result of Szilard as an alternative statement of the second law of thermodynamics. In the other direction, charging a battery corresponds to increasing available free energy which in turn corresponds — via the identity — to erasing of information. This relates the energy efficiency of charging a battery to the energy required to erase a bit.

III. ERASING THE TWO-STATE MARKOV CHAIN.

In this section, we obtain an analytic answer to the cost required to erase a reliable Brownian bit fast in the setting of the two-state Markov chain. For our choice of cost function, the two-state Markov chain should play to the theory of Brownian bits the same role as the Carnot engine plays to the theory of heat engines, in the following precise sense. We conjecture that, so long as the noise is Markovian, the cost for erasing the two-state Markov chain is a lower bound for all Brownian bits — for example for Langevin dynamics in a double-well potential.

A. The two-state Markov chain

1. As a model of a Brownian bit, consider a two-state continuous-time Markov chain with states 0 and 1 and the passive or uncontrolled dynamics given by transition rates $k_{01}$ from state 0 to state 1 and $k_{10}$ from state 1 to state 0.

Suppose the distribution at time $t$ is $(p_0(t), p_1(t))$ with $p_1(t) = 1 - p_0(t)$. Then the time evolution is given by the ODE

$$\dot{p}_0(t) = -k_{01} p_0(t) + k_{10} (1 - p_0(t)).$$  \hspace{1cm} (2)

Setting $\pi_0 = k_{10}/(k_{01} + k_{10})$ and $\tau_r := 1/(k_{01} + k_{10})$, this admits the solution

$$p_0(t) = \pi_0 + e^{-t/\tau_r} (p_0(0) - \pi_0)$$  \hspace{1cm} (3)

Here $\tau_r$ is the reliability timescale. The smaller the rates $k_{01}$ and $k_{10}$, the larger the value of $\tau_r$, and the slower the decay to equilibrium, so that the system remembers information for longer.

2. Fix a time $\tau_r$. Fix $p(0) = (1/2, 1/2)$. We want to control the dynamics with transition rates $u_{01}(t)$ and $u_{10}(t)$ to achieve $p(\tau_r) = (1, 0)$, where

$$\dot{p}_0(t) = -u_{01}(t) p_0(t) + u_{10}(t) (1 - p_0(t))$$  \hspace{1cm} (4)

We interpret this task as the erasing of a bit of reliability $\tau_r = 1/(k_{01} + k_{10})$ in time $\tau_r$. We want to find the cost of the optimal protocol $u_{01}^*(t)$ and $u_{10}^*(t)$ to achieve this objective, according to a cost function which we introduce next.

B. Cost function

3. Define the path space $\mathcal{P} := \{0, 1\}^{[0, \tau_r]}$ of the two-state Markov chain. This is the set of all paths in the time interval $[0, \tau_r]$ that jump between states 0 and 1 of the Markov chain.

4. Once the rates $u_{01}(t), u_{10}(t)$ and the initial distribution $p(0) = p$ for the Markov chain are fixed, there is a unique measure $\mu_{u,p}$ on path space which intuitively assigns to every path the probability of occurrence of that path according to the Markov chain evolution (Equation [4]) with initial conditions $p$. In detail, fix $i_0 \in \{0, 1\}$, and consider the set of paths $\mathcal{S}$ starting at $i_0$ with jumps occurring at times $t_1 < t_2 < \cdots < t_n$ within infinitesimal intervals $dt_1, dt_2, \ldots, dt_n$ and leading to the trajectory $(i_0, i_1, \ldots, i_n) \in \{0, 1\}^{n+1}$. Then setting $t_0 = 0$ and $t_{n+1} = \tau_r$ and $t_{n+1} = 1 \oplus i_n$:

$$\mu_{u,p}(\mathcal{S}) := p_{i_0} \prod_{j=0}^n e^{-\int_{t_j}^{t_{j+1}} u_{ij,ij+1}(s) ds} \prod_{j=1}^n u_{i_{j-1}i_j}(t_j) dt_j$$

where $p_{i_0}$ is the probability of starting at $i_0$, $e^{-\int_{t_j}^{t_{j+1}} u_{ij,ij+1}(s) ds}$ is the probability of staying in $i_0$ in the time interval $(0, t_1)$, $u_{i_{j-1}i_j}(t_j) dt_j$ is the probability of jumping from $i_0$ to $i_1$ in the interval $(t_1, t_1 + dt_1)$ and so on.

5. Specializing to $u_{01}(t) = k_{01}$ and $u_{10}(t) = k_{10}$, we obtain the probability measure $\mu_{u,k}$ induced on $\mathcal{P}$ by the passive dynamics (Eqn. [2]) with initial conditions $p$.

6. We declare the relative entropy $D(\mu_{u,p} \| \mu_{k,p})$ as the cost for implementing the control $u$. More generally, for a physical system with path space $\mathcal{P}$, passive dynamics corresponding to a measure $\nu$ on $\mathcal{P}$, and a controlled dynamics with a control corresponding to a measure $\mu$ on $\mathcal{P}$, we declare $D(\mu \| \nu)$ as the cost for implementing the control. We explore the thermodynamic, statistical, and informational significance of this cost function in Section [IV].
C. The Erasing Problem

7. Out of all controls $u(t)$ that start from $p(0) = \pi = (k_{10}/k_{01} + k_{10}), k_{01}/(k_{01} + k_{10})$, and achieve $p(\tau_e) = (1, 0)$, we want to find a control $u^*(t)$ that minimizes the relative entropy $D(\mu_{u^* \pi} \parallel \mu_{k, \pi})$.

8. Question 7 can be described within the framework of a well-studied problem in optimal control theory that has a closed-form solution [7][8]. Following Todorov [8], we introduce the optimal cost-to-go function $v(t) = (v_0(t), v_1(t))$. We intend $v_i(t)$ to denote the expected cumulative cost for starting at state $i$ at time $t < \tau_e$, and reaching a distribution close to $(1,0)$ at time $\tau_e$. To discourage the system from being in state 1 at time $\tau_e$, define $v_1(\tau_e) = +\infty$ and $v_0(\tau_e) = 0$.

10. Suppose the control performs actions $u_{01}(t)$ and $u_{10}(t)$ at time $t$. Fix a small time $h > 0$. Define the transition probability $u^h_{ij}(t)$ as the probability that a trajectory starting in state $i$ at time $t$ will be found in state $j$ at time $t + h$. When $i \neq j$, $u^h_{ij}(t) \approx h u_{ij}(t)$, whereas $u^h_{ii}(t) \approx 1 - u^h_{ij}(t)$ ignoring terms of size $O(h^2)$. We define $k^h_{ij}$ similarly.

11. To derive the law satisfied by the optimal cost-to-go $v(t)$, we approximate $v(t)$ by the backward recursion:

$$v_0(t) = \min_{u_{01}(t)} \mathbb{E} \left[ v_i(t + h) + \frac{u^h_{i0}(t)}{k^h_{0i}} \right]$$

$$v_1(t) = \min_{u_{10}(t)} \mathbb{E} \left[ v_i(t + h) + \frac{u^h_{1i}(t)}{k^h_{1i}} \right]$$

where $i$ \simlaw $(u^h_{0i}(t), u^h_{1i}(t))$ in the first equation, and $i$ \simlaw $(u^h_{10}(t), u^h_{1i}(t))$ in the second, and the approximation ignores terms of size $O(h^2)$. As $h \to 0$ the second terms $\mathbb{E} \log \frac{u^h_{1i}(t)}{k^h_{1i}}$ approach the relative entropy cost in path space over the time interval $(t, t + h)$.

12. In words, Eqn. 11 says that the cost-to-go from state 0 at time $t$ equals the cost of the control $u(t)$ plus the expected cost-to-go in the new state $i$ reached at time $t + h$. The cost of the control is measured by relative entropy of the control dynamics relative to the passive dynamics, over the time interval $(t, t + h)$.

13. Define the desirability $z_0(t) = e^{-v_0(t)}$ and $z_1(t) = e^{-v_1(t)}$. Define

$$G_0[z](t) = k^h_{00} z_0(t) + k^h_{01} z_1(t),$$

$$G_1[z](t) = k^h_{10} z_0(t) + k^h_{11} z_1(t).$$

We can rewrite as

$$\log z_0(t) = \log G_0[z](t + h) - \min_{u_{01}(t)} \mathbb{E} \left[ \log \frac{u^h_{0i}(t)G_0[z](t + h)}{k^h_{0i} z_i(t + h)} \right]$$

$$\log z_1(t) = \log G_1[z](t + h) - \min_{u_{10}(t)} \mathbb{E} \left[ \log \frac{u^h_{1i}(t)G_1[z](t + h)}{k^h_{1i} z_i(t + h)} \right]$$

14. Since the last term is the relative entropy of $(u^h_{0i}(t), u^h_{1i}(t))$ relative to the probability distribution $(k^h_{00} z_0(t + h)/G_0[z](t + h), k^h_{10} z_1(t + h)/G_1[z](t + h))$, its minimum value is 0, and is achieved by the protocol $u^*$ given by:

$$u^*_i(t) = \lim_{h \to 0} \frac{e^{-v_i(t + h)} - e^{-v_i(t)}}{h}$$

when $i \neq j$.

15. It remains to solve for $z(t)$ and the optimal cost. From Eqs. 13 and 14, at the optimal control $u^*$ the desirability $z(t)$ must satisfy the equation $-\log z(t) = -\log G[z](t + h)$, so that:

$$\left( \begin{array}{c} z_0(t) \\ z_1(t) \end{array} \right) = \left( \begin{array}{cc} 1 - k_{01} h & k_{01} h \\ k_{10} h & 1 - k_{10} h \end{array} \right) \left( \begin{array}{c} z_0(t + h) \\ z_1(t + h) \end{array} \right)$$

which simplifies to $\frac{dz}{dt} = -K z$ in the limit $h \to 0$, where $K = (\frac{k_{01}}{k_{10}} - \frac{k_{10}}{k_{01}})$ is the infinitesimal generator of the Markov chain. This has the formal solution $z(t - \tau_e) = e^{K t} z(\tau_e)$ where $z(\tau_e) = (\frac{1}{\tau_e})$ by 8.

16. In the symmetric case $k_{01} = k_{10}$,

$$z(t) = e^{H(t)} \left( \begin{array}{c} 1/2 \\ 1/2 \end{array} \right) + e^{-\tau_e/\tau_r} \left( \begin{array}{c} 1/2 \\ -1/2 \end{array} \right)$$

where $\tau_r = 1/(k_{01} + k_{10})$. Substituting $t = 0$ and taking logarithms, we find the cost-to-go function at time 0:

$$v(0) = \log 2 - \left( \log (1 + e^{-\tau_e/\tau_r}) + \log (1 - e^{-\tau_e/\tau_r}) \right)$$

17. When $k_{01} = k_{10}$, the cost $C_{\text{erase}}(\tau_r, \tau_e, T)$ required for erasing a bit of reliability $\tau_r = 1/(k_{01} + k_{10})$ in time $\tau_e$ at temperature $T$ is:

$$C_{\text{erase}}(\tau_r, \tau_e, T) \geq \log 2 - \frac{1}{2} \log \left( 1 - e^{-2\tau_r/\tau_e} \right)$$

Note that $C_{\text{erase}} \geq \log 2$ with equality when $\tau_e/\tau_r \to \infty$, since $1 - e^{-2\tau_r/\tau_e} \leq 1$.

18. From Eqn. 8 $C_{\text{erase}} \geq \frac{1}{2} \log \frac{2\tau_e}{\tau_r}$ when $\tau_r \gg \tau_e$.

IV. INTERPRETING THE PATH-SPACE RELATIVE ENTROPY

1. The relative entropy $D(\mu || \nu)$ counts the number of nats erased by the control in path space, relative to the passive dynamics. Since the Szilard-Landauer principle asserts that erasing one bit requires at least $k_B T \log 2$ units of energy, our cost function may be viewed as a formal extension of this principle to path space.
2. We wish to compare the cost $D(\mu\|\nu)$ with the usual thermodynamic expected work. Before doing so, we will find it convenient to define the time reversal Markov chain. Given a distribution $q$ at time $\tau_e$, the time reversal Markov chain of the Markov chain in Eqn. 4 evolves backward in time according to the time-reversed ODE:

$$q_0(t) = u_{01}(t)q_0(t) - u_{10}(t)(1 - q_0(t))$$

$$q(\tau_e) = q.$$

The measure $\mu_{\tau_e}^{\text{rev}}$ is the measure on path space described by Eqn. 7.

3. The usual thermodynamic expected work can be defined as follows. Run the control dynamics Eqn. 9 forwards from initial condition $p(0)$ up to time $\tau_e$ to obtain the distribution $p(\tau_e)$. Now consider the measure $\mu_{\tau_e}(p(\tau_e))$. By the First Law of Thermodynamics,

$$\Delta W = \Delta F + k_B T D(\mu_{\tau_e}||\mu_{\tau_e}(p(\tau_e)))$$

where the increase in free energy of the system $\Delta F = k_B T (D(p(\tau_e)||\pi) - D(p(0)||\pi))$ by Eqn. 7 and $D(\mu_{\tau_e}||\mu_{\tau_e}(p(\tau_e)))$ is the total entropy production during the time interval $[0, \tau_e]$.

4. It is instructive to compare our cost function with $\Delta W$. We first recognize the time reversal $\mu_{\tau_e}^{\text{rev}}$ as a specialization of $\mu_{\tau_e}^{\text{rev}}$. After some algebra, we obtain

$$D(\mu_{\tau_e}||\mu_{\tau_e}(p(\tau_e))) = \frac{\Delta F}{k_B T} + D(\mu_{\tau_e}||\mu_{\tau_e}(p(\tau_e)))$$

where $p(\tau_e)$ is — as in Eqn. 8 — the solution to the control dynamics Eqn. 4 at time $\tau_e$.

5. Comparing 8 and 9 we have replaced the total entropy production $D(\mu_{\tau_e}||\mu_{\tau_e}(p(\tau_e)))$ in 8 by $D(\mu_{\tau_e}||\mu_{\tau_e}(p(\tau_e)))$. It would be most satisfying if a thermodynamic interpretation for this difference could be obtained by drawing on the deep links between dissipation and irreversibility.

6. Our cost function $D(\mu\|\nu)$ also admits a large deviation interpretation which was, remarkably, already noted by Schrödinger in 1931 [10,13]. Motivated by quantum mechanics, Schrödinger asked: conditioned on a more or less astonishing observation of a system at two extremes of a time interval, what is the least astonishing way in which the dynamics in the interval could have proceeded? Specializing to our problem of erasing, suppose a two-state Markov chain with passive dynamics given by Eqn. 2 is observed at time 0 and time $\tau_e$ and found to be in state $\{1, 2\}$, and state $\{1, 0\}$ respectively. Conditioned on this rare event that the system spontaneously erased itself, what is the least unlikely measure $\mu^*$ on path space via which the process took place?

From a statistical treatment of multiple single particle trajectories, Schrödinger found that the likelihood of an empirical measure $\mu$ on path space falls exponentially fast with the relative entropy $D(\mu||\nu)$ where $\nu$ is the measure induced by the passive dynamics. In particular, the least unlikely measure $\mu^*$ is that measure which — among all $\mu$ whose marginals at time 0 and time $\tau_e$ respect the observations — minimizes $D(\mu||\nu)$. So for the problem of erasing, $\mu$ is any measure that has marginal $(1/2, 1/2)$ at time 0 and marginal $(1, 0)$ at time $\tau_e$, and $\mu^*$ is that measure among all such $\mu$ that minimizes $D(\mu||\mu^*(1/2, 1/2))$. Thus our optimal control produces in expectation the least surprising trajectory among all controls that perform rapid erasing.

7. Eqn. 3 is not accidental for this example, but is in fact a general feature when the cost function is relative entropy $D$. More abstractly, the Radon-Nikodym derivative (i.e., “probability density”) $\frac{d\mu^*}{d\nu}$ of the measure $\mu^*$ induced on path space by the optimal control $u^*$ with respect to the measure $\nu$ induced by the passive dynamics is a Gibbs measure, with the cost-to-go function $v(t)$ playing the role of an energy function. In other words, mathematically our problem is precisely the free energy minimization problem so familiar from statistical mechanics. There is also a possible physical interpretation of paths in $P$ as microstates, instead of points in phase space [12].

8. Posing the erasing problem with the usual thermodynamic work $\Delta W$ as the cost function leads to the Szilard-Landauer answer of $k_B T \log 2$. Neither the reliability timescale $\tau_e$ nor the erasing timescale $\tau_e$ appear in this answer. The protocol that achieves this is an adaptation of the infinite-time isothermal protocol that proceeds by raising the ‘1’ indefinitely, waiting for the system to equilibrate, and repeating. We can convert this into a finite-time protocol by letting the control rates $u_{01}(t)$ and $u_{10}(t)$ tend to $+\infty$ since $\Delta W$ depends only on the ratio $u_{01}(t)/u_{10}(t)$. We need to wait for an arbitrarily small amount of time to get arbitrarily close to equilibrium.

9. In optimal control theory, $D(\mu\|\nu)$ has been found to be a natural choice in the context of “risk-sensitive” and $H_\infty$-robust control [7,8,15,16].

V. CONCLUDING REMARKS

Several groups [17,19] have recognized that rapid erasing requires entropy production which pushes up the cost of erasing beyond $k_B T \log 2$, and have obtained bounds for this problem. A grossly oversimplified sketch of these
various results is obtained by considering the energy cost of compressing the Szilard engine fast. Recall that the Szilard engine is a single molecule of an ideal gas in a cylindrical vessel [2]. Specializing a result from finite-time thermodynamics [20] to the case of the Szilard engine, one obtains an energy cost 
\[
\frac{k_B T \log 2}{\sigma - k_B \log 2}
\]
where \( \sigma \) is the coefficient of heat conductivity of the vessel. Our work differs from previous works in seeking a fundamental answer that is not limited by technological bounds, hence does not depend on \( \sigma \).

Though early workers like von Neumann [21] and Swanson [22] remarked on reliability, the notion of reliability has been rarely considered in modern works. This is because in a resource model where there is no cost to raising and lowering a barrier so long as the barrier is not populated, the reliability of bits can be arbitrarily modified. We may treat the raised and lowered states of the barrier as a Brownian bit. Since the barrier is required to have reliability at least as high as the bit it is protecting, raising and lowering the barrier fast is equivalent to switching a reliable bit fast. Especially if one is considering the manipulation of a single bit in isolation, with no amortization among many bits of the cost of switching the barrier, it is not at all clear that this can be done cheaply. Merely asserting that there is no cost provides no guide to synthesis.

Since charging a battery can be thought of as erasing a bit, our result may also hold insights into the design of batteries that must be rapidly charged, and must hold their energy for a long time.

We speculate that our cost function may have an interpretation as an action, so that our result points to a fundamental action/ reliability/ speed trade-off to erasing a bit.

---

[1] U. Seifert, Rep. Prog. Phys. 75, 126001 (2012).
[2] L. Szilard, Z Phys 53, 840 (1929).
[3] R. Landauer, IBM Journal of Research and Development 5, 183 (1961).
[4] M. Esposito and C. Van den Broeck, EPL (Europhysics Letters) 95, 40004 (2011).
[5] M. Gopalkrishnan, arXiv preprint arXiv:1311.3533 (2013).
[6] D. Reeb and M. M. Wolf, arXiv preprint arXiv:1306.4352 (2013).
[7] W. H. Fleming and S. K. Mitter, Stochastics: An International Journal of Probability and Stochastic Processes 8, 63 (1982).
[8] E. Todorov, Proceedings of the national academy of sciences 106, 11478 (2009).
[9] P. Dupuis and R. S. Ellis, A weak convergence approach to the theory of large deviations, Vol. 902 (John Wiley & Sons, 2011).
[10] E. Schrödinger, Akad. Wissen., Berlin Phys. Math 144 (1931).
[11] A. Beurling, Annals of Mathematics , 189 (1960).
[12] H. Föllmer, in École d’Été de Probabilités de Saint-Flour XV–XVII, 1985–87 (Springer, 1988) pp. 101–203.
[13] R. Aebi, Schrödinger diffusion processes (Springer, 1996).
[14] A. Wissner-Gross and C. Freer, Physical review letters 110, 168702 (2013).
[15] D. Hernandez-Hernandez and S. I. Marcus, Systems & Control Letters 29, 147 (1996).
[16] S. P. Coraluppi and S. I. Marcus, Automatica 35, 301 (1999).
[17] E. Aurell, K. Gawędzki, C. Mejía-Monasterio, R. Mohayaee, and P. Muratore-Ginanneschi, Journal of statistical physics 147, 487 (2012).
[18] G. Diana, G. B. Bagci, and M. Esposito, Physical Review E 87, 012111 (2013).
[19] P. R. Zulkowski and M. R. DeWeese, Physical Review E 89, 052140 (2014).
[20] P. Salamon and A. Nitzan, The Journal of Chemical Physics 74, 3546 (1981).
[21] J. von Neumann, “Theory of self-reproducing automata,” (University of Illinois Press, Urbana, 1966) p. 66, lecture delivered at University of Illinois in December, 1949.
[22] J. Swanson, IBM Journal of Research and Development 4, 305 (1960).