How Much Does NOT Cost?

Some Thermodynamics of Information Processing

Mikhael Semaan

PHY 256B Final Report
UC Davis | Spring 2018

After a brief motivational historical note, I employ the "information ratchet" framework to elucidate the fundamental trade-off between accuracy and energy for the reversible NOT operation. I find that enforcing detailed balance—the microscopic reversibility requirement—leads to a closed form expression for the necessary energy difference between the two internal states of the NOT dynamic, and that this energy difference varies linearly with temperature but logarithmically with accuracy.

Contents

1 History

The development of modern thermodynamics followed humanity's progress creating "heat engines"—objects which exchange energy between reservoirs through processes we now call heat and work. By the time Carnot published his *Reflections on the Motive Power of Heat* [5, 6], Hero had already described the aeolipile [9], Papin had already invented the "steam digester" [8], and Fulton had already commercialized a steam-powered boat [7]. In fact, in his *Reflections*' opening paragraph, Carnot writes "that [heat] possesses vast motive-power no one can doubt, in these days when the steam-engine is everywhere so well known."

A century later, Shannon published *A Mathematical Theory of Communication* [14, 13], founding information theory. Much like Carnot, he opened with motivation from already-existing applications: "the recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication."

The major role of both Carnot's and Shannon's papers was to explain why some schemes (thermal engines in Carnot's case; data compression methods in Shannon's) were better than others from a fundamental standpoint, and to establish general bounds on the behavior of such schemes. In neither case did the results lead immediately and directly to better real-world devices. In both cases, however, the establishment of those bounds launched entire fields of theoretical work and eventually guided design principles for things like the diesel engine and the compression schemes which make Netflix and YouTube possible.

Prior to the works of Szilard[1] [16, 15], Landaur [10], Penrose [12], Bennett [2], and others, information theory and thermodynamics seemed more or less distinct. As an example, in 1854—prior to Shannon and contemporary to Carnot—Lord Kelvin wrote that the heat-work interplay occurred "by means of forces acting between contiguous parts of bodies, or due to electric excitation; but in no other way known, or even conceivable, in the present state of science" [17]. But we now know that *information* is another means by which "mechanical action may be derived," a revelation inspired by attempts to understand Maxwell's demon and perhaps most famously made by Landauer's principle, which established a minimum energetic cost for the irreversibly erasing a bit of information.

[1] Szilard's work actually predated Shannon's.

2 Introduction

Recently there has been a wealth of progress exploring the thermodynamics of information and of various kinds of information processing [11]. As was the case with Carnot and Shannon, we already have working examples of information-processing machines: our various computers; our brains; animals, plants, and other biological organisms (including prokaryotes); and perhaps many others like collective behaviors and evolutionary dynamics.

I focus on one piece of this wealth of examples: the logical binary computer, ubiquitous in today's world. Landauer's principle is an appropriate starting point for working out the fundamental thermodynamic trade-offs in information processing, but the picture is not complete (in particular, see "Above and Beyond the Landauer Bound" [3]). Here, I borrow heavily from the *information ratchet* framework as used by Boyd et al. [4] to attempt to establish a different fundamental bound: that on the energy required to carry out a reversible, probabilistic NOT operation. More broadly, I am interested in the bounds on energetic requirements for various kinds of logical operation; this report is a first step in exploring that direction.

3 A Modified Information Ratchet

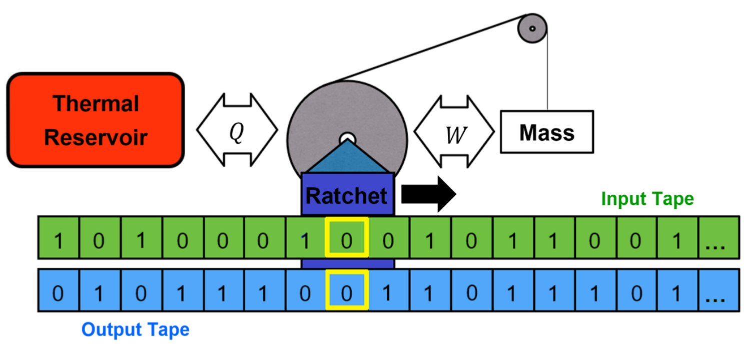

Figure 1: An information ratchet, which is at all times in contact with a thermal reservoir, a work reservoir, and an information reservoir (the input and output tapes). The ratchet is driven to the right at regular intervals, interacting with one bit at a time from each the input and output tapes; the random variables X_t and Y_t describe the states of those respective bits at time t. The information reservoir's state as a whole at time t is specified by the block random variables $X(t) \equiv X_{0:\infty}(t)$ for the input tape and $Y(t) \equiv Y_{0:\infty}(t)$ for the output tape. A temperature T gives the thermal reservoir's state; and we may specify the work reservoir's state by a distance h between a mass m and a massless, frictionless pulley from which the mass hangs in a gravitational field of strength mg.

An *information ratchet* is a device which mediates the interaction between *three* reservoirs: thermal, mechanical, and "information." Boyd et al. showed that in such a scheme the information reservoir can act as a source or sink of entropy, just as the familiar thermal and work reservoirs, and that this leads to various kinds of thermodynamic functionality [4].

Figure 1 depicts a ratchet modified from that in Ref. [4] in one major way: rather than moving along a single tape (in our case, a string of bits) and interacting with each bit one at a time, it moves along *two* tapes, interacting with *two* bits (one from each tape) at a time.

One is called the input tape and one the output tape, precluding the intention to *write* only to the output tape but *read* from both tapes. At some time $t = 0$, X_0 and Y_0 denote the random variable associated with the interacting bits' states. Henceforth, I refer to the interacting bits on the input and output tapes at a particular time as the "input bit" and "output bit." For simplicity, I make a few immediate assumptions about the operation of the ratchet:

1. Depending on the value of the input bit, the ratchet "selects" an internal dynamic which governs the interaction between it and the input and output bits.

2. The ratchet never alters the state of the input tape: a particular realization of the input tape $X = x$ remains constant.

3. The ratchet is driven to the right at unit time intervals.

4. Each transition of the internal dynamic in 1 takes a time $\tau < 1$.

5. It takes time $1 - \tau$ for the ratchet to move, read the new bits x_t and y_t, and select the appropriate dynamic via 1.

6. The internal transitions are not driven and thus occur due to thermal fluctuations—interactions with the thermal reservoir as exchanges of heat.

These assumptions lend themselves to Figure 2's schematic description of the ratchet. The advantage of such a schematic is that we may eventually be able to define each block in terms of *transducers* from computational mechanics [1]; this would describe the ratchet's operation constructively and explicitly. In this report, I focus on two of the blocks: Dynamic 1 and Dynamic 2, and I will describe them in terms of Markov chains whose states are elements of the joint state space of the ratchet, input, and output bits. They may be equivalently described by transducers, but I leave them as Markov chains for clarity when enforcing detailed balance later in this report.

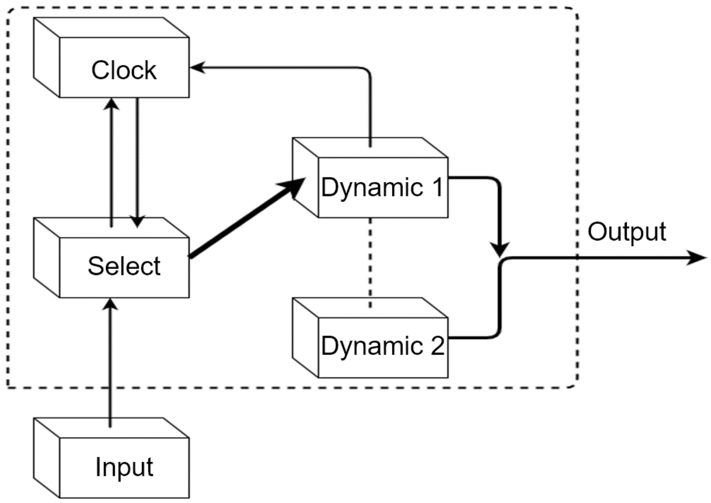

Figure 2: A schematic of the ratchet's operation. Some combination of the actual selection process and the dynamic(s) determines the time scale of the clock pulses such that assumptions 3–5 hold. Then, at regular time intervals, the ratchet reads the values of the input and output bits and selects a particular dynamic for interacting with those bits. A transition occurs, and the new values of the input and output tapes are the "output" of the ratchet. The dashed box encloses the blocks *internal* to the ratchet.

4 NOT Ratchet-Tape Dynamics

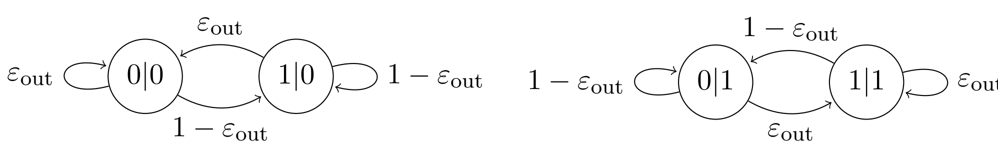

Figure 3: Dynamic 1 (left) and Dynamic 2 (right) Markov chain representations. The notation for the states is "output | input," referring to the values of the output and input bits. The parameter $\varepsilon_{\text{out}} < 1$ describes the chances of output error. Given our assumptions, where we reset by moving the ratchet after each transition, ε_{out} is precisely the per-bit probability of a "write error."

The two Markov chains in Figure 3 represent implementations of NOT, given a particular input. As mentioned in the Figure's caption, ε_{out} is a per-bit write error probability: for each bit we read from the input tape, there is a probability ε_{out} we will incorrectly write to the output tape, since our assumptions enforce driving the ratchet after every internal transition. The states in each Markov chain are elements of the of the ratchet, input bit, and output bit joint state space.

Let us focus on Dynamic 1, since as a Markov chain Dynamic 2 is equivalent up to state labeling. Let R be the random variable specifying the joint state of the input and output bits and the ratchet. We can specify the dynamic by transition matrix elements

$$\mathrm{T}_{ij} = \Pr(R_{t+\tau} = j | R_t = i), \tag{1}$$

where $i, j \in R$, and by a starting distribution $\langle \mu_0 |$ over the elements of R.

It is worth noting a couple of things. First, specifying R_t is equivalent to specifying Y_t given X_t and the dynamic (Dynamic 1 in this case). Also, because we drive the ratchet at unit time intervals and $\tau < 1$, $Y_{t+\tau}$ always refers to the same "slot" on the output tape as Y_t. Finally, as a result of this unit choice, we have that $t \in \mathbb{N}_0$, and that Y_t specifies the state of the output tape's tth bit *before* interaction with the ratchet, while $Y_{t+\tau}$ specifies the state of the very same bit *after* the interaction.

Dynamic 1 has transition matrix

$$\mathrm{T}_{\mathrm{D1}} = \begin{bmatrix} \varepsilon_{\mathrm{out}} & 1 - \varepsilon_{\mathrm{out}} \\ 1 - \varepsilon_{\mathrm{out}} & \varepsilon_{\mathrm{out}} \end{bmatrix}. \tag{2}$$

5 Detailed Balance

In trying to design a computer to do the logical operation specified by Figure 3, we might try to force $\varepsilon_{\mathrm{out}} = 0$, and so arrive at a perfectly deterministic logic gate. This would certainly not be attainable and thermodynamically reversible[2] due to its inability to satisfy *detailed balance*.

An arbitrary state distribution $\langle \pi |$ over R satisfies **detailed balance** with respect to T if $\langle \pi |_i \mathrm{T}_{ij} = \langle \pi |_j \mathrm{T}_{ji}$. The existence of such a $\langle \pi |$ implies that $\langle \pi |$ is a stationary distribution, so detailed balance implies the existence of a stationary distribution (but the converse is not true). Detailed balance also implies *reversibility*, which, physically, means there is no net entropy produced by the interactions.

[2] Very recently, I've started to see indications floating around some of the early papers—including Landauer's, which explicitly assumes irreversibility—that logical irreversibility implies thermodynamic irreversibility. If the converse (double-converse?) is true—that thermodynamic reversibility implies logical reversibility—then we have a very real, physical reason to look at cases where detailed balance does *not* hold. Since the meat of my project was enforcing detailed balance on a Markov chain's transitions, I will proceed undeterred, but I think this is worth more consideration.

5.1 Applying Detailed Balance to NOT

Assuming a Boltzmann distribution over the states implies, in terms of the marginal transition probabilities, that we can write the detailed balance requirement as

$$\frac{\mathrm{T}_{ij}}{\mathrm{T}_{ji}} = \mathrm{e}^{\Delta E/k_{\mathrm{B}}T}, \tag{3}$$

where $\Delta E \equiv E_i - E_j$ is the energy difference between the states i and j, T is the temperature of the thermal reservoir with which the ratchet maintains contact, and k_{B} is Boltzmann's constant.

Regardless of $\varepsilon_{\mathrm{out}}$, the allowed transitions T_{ii} and T_{jj} result in $\Delta E = 0$, which is reassuring: it means that the state i has the same energy as itself. We don't run into real trouble here with $\varepsilon_{\mathrm{out}} = 0$, either: although it produces an indeterminate left-hand side of Eq. 3, the numerator and denominator are identical, so $\lim_{\varepsilon_{\mathrm{out}} \to 0} = 1$.

However, suppose $i = 0|0$ and $j = 1|0$. Then Eq. 3 requires

$$\begin{aligned} \frac{1-\varepsilon_{\mathrm{out}}}{\varepsilon_{\mathrm{out}}} &= \mathrm{e}^{\Delta E/k_{\mathrm{B}}T} \\ \implies \Delta E &= k_{\mathrm{B}}T \ln\left(\frac{1-\varepsilon_{\mathrm{out}}}{\varepsilon_{\mathrm{out}}}\right). \end{aligned} \tag{4}$$

Eq. 4 sets exactly the required energy difference between the states $0|0$ and $1|0$ such that the dynamic is reversible. If we take $\varepsilon_{\mathrm{out}} \to 0$, that difference blows up (though only logarithmically, which is convenient).

Thus, forcing reversibility on the process not only rules out the possibility of a "perfect" NOT computer—it would require infinite energy to do any processing—but also gives an explicit closed form for the trade-off between error minimization and (one kind of) energetic cost. This requisite energy difference holds regardless of implementation and thus places a lower bound on the energy required to actually make a transition between the states (it assumes whatever scheme exists to make that transition is adiabatic).

6 More Considerations, and Future Work

Detailed balance must hold for any operations we expect to be physically reversible. This gives a prescription for coming up with energy-accuracy trade-offs for many other operations. One such notable one is NAND, but this already presents a few challenges: the classical NAND gate takes *two* inputs in parallel and produces one output. This is implementable by a ratchet similar to Figure 1's but with two input tapes rather than one, effectively increasing the number of dynamics from which to select. One may also consider a serial type of NAND processing, whereby the ratchet first reads two bits and then alters a third bit based on their values. The Markov dynamic for such a ratchet would necessarily include more states, because the ratchet would necessarily have more memory (it must "remember" the value of the previous two bits, or just the previous bit, if we include a separate output tape).

Another natural thing to consider is the possibility for *input* errors, where we accidentally alter the state of the input tape. Such a case also increases the state space of a given dynamic, and makes picturing a single large joint state space for both dynamics difficult; it also introduces at least one new constraint ε_{in}, with which we can consider the trade-offs imposed by detailed balance as above.

Towards both of these ends and for general usefulness, one piece of ongoing work is automating the process of constraining parameters in a Markov chain to satisfy detailed balance. Our case was simple, but for exploring accuracy-energy trade-offs for much larger Markov chains—as necessitated by more complicated operations—it can presumably range from tedious to intractable by hand.

Additionally, I would like to consider the effects of modularity dissipation, which Boyd et al. found places bounds on the dissipation due to modularity but can be overcome by designing ratchet states with larger internal memories (and thus larger Markov state spaces) [3].

Finally, the Markov dynamics presented here have equivalent representations in terms of transducers [1]; such a description is advantageous because of its smaller size but also because of its potential to allow for composition of transducers, which would bring clarity to the connections between the various blocks of Figure 2.

7 Acknowledgements

I would like to thank Alec B. Boyd, James P. Crutchfield, Gregory (from the group), and Ryan G. James for guidance and useful discussion. I'd also like to thank my fellow classmates for their camaraderie throughout the year, and the UC Davis Physics Department for hosting my presence here and allowing me to pursue my academic passions. (I feel like I get paid to learn the secrets of the Universe. What a day job.)

References

[1] Nix Barnett and James P. Crutchfield. "Computational Mechanics of Input–Output Processes: Structured Transformations and the ϵ-Transducer". In: *Journal of Statistical Physics* 161.2 (Oct. 1, 2015), pp. 404–451. arXiv: 1412.2690.

[2] Charles H. Bennet. "The thermodynamics of computation—a review". In: *International Journal of Theoretical Physics* 21.12 (Dec. 1, 1982), pp. 905–940.

[3] Alexander B. Boyd, Dibyendu Mandal, and James P. Crutchfield. "Above and Beyond the Landauer Bound: Thermodynamics of Modularity". In: *ArXiv e-prints* (Aug. 9, 2017). arXiv: 1708.03030.

[4] Alexander B. Boyd, Dibyendu Mandal, and James P. Crutchfield. "Identifying functional thermodynamics in autonomous Maxwellian ratchets". In: *New Journal of Physics* 18.2 (Feb. 22, 2016), p. 023049. arXiv: 1507.01537.

[5] Sadi Carnot. *Réflexions sur la puissance motrice du feu et sur les machines propres à développer atte puissance*. French. Paris: Bachelier Libraire, 1824.

[6] Sadi Carnot. *Reflections on the Motive Power of Heat*. English. Ed. and trans. by R. H. Thurston. New York: J. Wiley, 1890.

[7] Wikipedia contributors. "North River Steam Boat". In: *Wikipedia, The Free Encyclopedia* (2018). (Visited on 06/13/2018).

[8] The Editors of Encyclopaedia Britannica. "Denis Papin". In: *Encyclopædia Britannica* (Apr. 18, 2017). (Visited on 06/13/2018).

[9] The Editors of Encyclopaedia Britannica. "Aeolipile". In: *Encyclopædia Britannica* (June 6, 2016). (Visited on 06/13/2018).

[10] Rolf Landauer. "Irreversibility and Heat Generation in the Computing Process". In: *IBM Journal of Research and Development* 5.3 (July 1961), pp. 183–191.

[11] Juan M. R. Parrondo, Jordan M. Horowitz, and Takahiro Sagawa. "Thermodynamics of information". In: *Nature Physics* 11 (Feb. 3, 2015), pp. 131–139.

[12] Oliver Penrose. *Foundations of Statistical Mechanics. A Deductive Treatment*. Vol. 22. International Series in Natural Philosophy. Oxford: Pergamon Press (Elsevier Ltd.), 1970.

[13] Claude Shannon. "A mathematical theory of communication". In: *The Bell System Technical Journal* 27.4 (Oct. 1948), pp. 623–656.

[14] Claude Shannon. "A mathematical theory of communication". In: *The Bell System Technical Journal* 27.3 (July 1948), pp. 379–423.

[15] Leo Szilard. "On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings". English. In: *Behavioral Science* 9.4 (1964), pp. 301–310.

[16] Leo Szilard. "über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen". German. In: *Zeitschrift für Physik* 53.11 (Nov. 1929), pp. 840–856.

[17] William Thomson. "On the Dynamical Theory of Heat. Part V. Thermo-electric Currents." In: *Transactions of the Royal Society of Edinburgh* 21.9 (May 1854), pp. 123–172.