Thermodynamic Costs of Turing Machines (Kolchinsky and Wolpert 2020)

Daniel Briseno

February 19, 2021

Context of the Paper

Prior work on Thermodynamics of Information Processing

- Landauer cost of erasing a bit: $kT \ln 2$ (1961)
- Logically reversible computations can be performed with no heat or entropy production (1973)
- Informal argument for minimum cost of $x \mapsto y$ (1989 - 2019)
- Development of non-equilibrium statistical physics
  - Trajectory-based and stochastic thermodynamics (2013-2015)
- Thermodynamic costs of specific implementations of Turing Machines (TM)(2015-2019)

Purpose of the Paper

Thermodynamic costs of computation

- Extends results to general class of TM
- Analyzes the thermodynamic costs of $f : \mathbb{N} \nrightarrow \mathbb{N}$ on a physical implementation of a TM M
- Logical properties of f and M impose constraints on thermodynamic costs.
- Result might generalize to any implementation of a TM

Background – Computer Science

Turing Machines

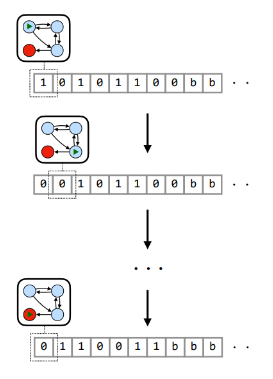


Figure: Graphical representation of a TM

Background – Computer Science

Additional Assumptions on TM M

1. $\Sigma = \{0, 1, b\}$
2. If and when M halts on an input, the tape will contain an output string $s \in \{0, 1\}^*$ followed by all blank symbols, and the pointer will be set to the start of the tape.

Assumptions do not affect the computational capabilities of M.

Background – Computer Science

Turing Machines as Partial Functions

Any computation performed by a TM M can be represented as

$$\phi_M : \{0,1\}^* \nrightarrow \{0,1\}^*$$

and $\phi_M(x) = y$ indicates that M started with input program x yields the output string y.

Universal TM

There exist Universal Turing Machines (UTM) such that given a UTM U and any TM M, there exists an interpreter program $\sigma_{U,M}$ such that

$$\phi_U(\sigma_{U,M}, x) = \phi_M(x)$$

Background – Computer Science

Computability

- Church Turing Thesis: A function can be calculated by a sequence of formal operations if and only if it is computable by a Turing Machine.
- Physical Church Turing Thesis: Any function implemented by a physical process can also be implemented by a Turing Machine

Background – Algorithmic Information Theory

Kolmogorov Complexity

The Kolmogorov complexity K_U of a bitstring x is the length of the shortest input program that when given to a UTM U can produce x as an output:

$$K_U(x) := \min_{z:\phi_U(z)=x} \ell(z)$$

- Measure of amount of information in x

Background – Algorithmic Information Theory

Kolmogorov Complexity of Bitstring x

$$K_U(x) := \min_{z:\phi_U(z)=x} \ell(z)$$

Kolmogorov Complexity of a Computable Function f

$$K_U(f) := \min_{M:\phi_M=f} \ell(\sigma_{U,M})$$

Conditional Kolmogorov Complexity of x Given Bitstring y

$$K_U(x|y) = \min_{z:\phi_U(z,y)=x} \ell(z)$$

Background – Algorithmic Information Theory

Invariance Theorem

For distinct UTM U, U':

$$K_{U'}(x) = K_U(x) + O(1)$$

Thus, U is usually omitted and we write $K(x)$ for Kolmogorov complexity of x

Background – Input Distributions

Coin-Flipping Distribution

- Input string x as random variable with probability distribution p_X
- Important example: coin flipping distribution of TM M

$$m_X^{\mathsf{coin}}(x) := \begin{cases} 2^{-\ell(x)} & \text{if } x \in \mathsf{dom}\ \phi_M \\ 0 & \text{otherwise} \end{cases}$$

- With normalizing constant $\Omega_M := \sum_{x \in \mathsf{dom}\ \phi_M} 2^{-\ell(x)}$

$$p_X^{\mathsf{coin}}(x) = m_X^{\mathsf{coin}}(x)/\Omega_M$$

Background – Entropy

Shannon Entropy of Distribution p_X

$$S(p_X) = -\sum_{x \in X} p_X(x) \ln[p_X(x)]$$

- Measure of amount of information in p_X
- $-\ln[p_X(x)]$: "surprisal", how unexpected, and hence informative, is x?
- $p_X(x)$: how often do we receive surprise $\ln p_X$

Background – Entropy

Entropy Production (EP)

The expected EP, written $\Sigma(p_X)$ of a physical process with initial state distribution p_X and final state distribution p_Y is:

$$\Sigma(p_X) = S(p_Y) - S(p_X) + \langle Q \rangle_{p_X} / kT$$

Thermodynamically reversible processes have $\Sigma(p_X) = 0$. EP is always nonnegative.

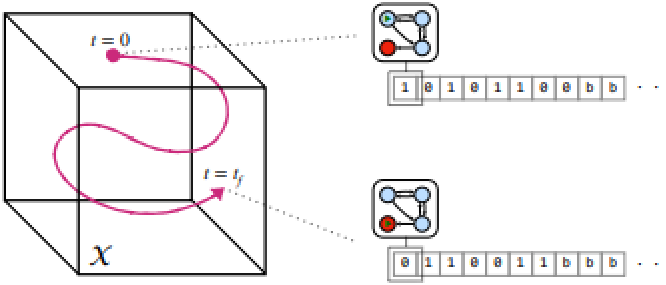
Physical Setup

Physical System Under Consideration

- Countable state-space $\mathcal{X}$
- Connected to a work reservoir
- Connected to heat bath at temperature T
  - Bath in a Boltzmann distribution
- System evolves according to driving protocol during time interval $[0, t_f]$

The heat function $Q(x)$ is the expected thermal energy transferred from the system to the heat bath.

Physical Setup



Physical Setup

System under consideration

The joint Hamiltonian of the system is

$$H_X^t(x) + H_B(b) + H_{\mathsf{int}}(x, b)$$

If $p_B(b)$ is the initial distribution of the bath and $p'_{B|X}$ is the final distribution, then:

$$Q(x) = \langle H_B \rangle_{p'_{B|x}} - \langle H_B \rangle_{p_B}$$

Realizations of a TM

Realization Formally Defined

- Let $p_{Y|X}$ be the conditional probability distribution of the system's final state y given initial state x
- A physical process is a **realization** of a partial function $f : \mathcal{X} \nrightarrow \mathcal{X}$ if

$$p_{Y|X}(y|x) = \delta(f(x), y)$$

Realization of a TM M Formally Defined

- TM M can be written as a partial function $\phi_M : \{0,1\} \nrightarrow \{0,1\}$
- A physical process is a realization of a TM M if it is a realization of ϕ_M
- A physical process is a **computable realization** if its heat map $Q(x)$ is computable

Proposition 1

Proposition Statement

Given a countable set $\mathcal{X}$ and partial functions $f : \mathcal{X} \nrightarrow \mathcal{X}$ and $G : \mathcal{X} \nrightarrow \mathbb{R}$, the following are equivalent:

1. For all p_X with supp $p_X \subseteq$ dom f

$$\langle G \rangle_{p_X} + S[p_{f(X)}] - S(p_X) \geq 0$$

2. For all $y \in$ img f

$$\sum_{x : f(x) = y} e^{-G(x)} \leq 1$$

3. There exists a realization of f coupled to a heat bath at temperature T whose heat function Q obeys

$$Q(x)/kT = G(x) \qquad \forall x \in \text{dom } f$$

Proposition 1

Recovering Landauer Cost From Proposition 1

Take $x \in \{0, 1\}$ to be a random bit determined by a coin toss, and f as the bit-erasing operation $f(x) = 0$. Then:

$$p_X(x) = \frac{1}{2}$$
$$p_{f(X)}(y) = \begin{cases} 0 & \text{if } y = 1 \\ 1 & \text{if } y = 0 \end{cases}$$

Then for any $G(x) = Q(x)/kT$, condition 1 implies:

$$\begin{aligned} \langle G \rangle_{p_X} + S[p_{f(X)}] - S(p_X) \geq 0 \\ \implies \langle G \rangle_{p_X} \geq S(p_X) \\ \implies \langle G \rangle_{p_X} \geq \ln 2 \end{aligned}$$

Proposition 1

Recovering Landauer Cost From Proposition 1 (cont.)

We would like to characterize the cost of an arbitrary bit deletion, so taking G to be identical for inputs $\{0, 1\}$

$$G(x) \geq \ln 2$$

and using equivalent condition 3 from Proposition 1 we recover the Landauer cost of a bit deletion:

$$\begin{aligned} Q(x)/kT &\geq \ln 2 \\ \implies Q(x) &\geq kT \ln 2 \end{aligned}$$

Realizations of TM

Realizations of TM Used in Analysis

- **Coin-Flipping Realization**: thermodynamically reversible when inputs are sampled from coin-flipping distribution
- **Dominating Realization**: produces less heat than any computable realization of a TM

Coin-Flipping Realization

Input Distribution

$$m_X^{\mathsf{coin}}(x) := \begin{cases} 2^{-\ell(x)} & \text{if } x \in \text{dom } \phi_M \\ 0 & \text{otherwise} \end{cases}$$

$$p_X^{\mathsf{coin}}(x) = m_X^{\mathsf{coin}}(x)/\Omega_M$$

Output Distribution

$$m_Y^{\mathsf{coin}}(y) = \sum_{x:\phi_M(x)=y} 2^{-\ell(x)}$$

$$p_Y^{\mathsf{coin}}(y) = m_Y(y)/\Omega_M$$

Coin-Flipping Realization

Associated Heat Function of Coin-Flipping Realization for TM M

Can be shown that

$$G(x) = -\ln p_X^{\mathsf{coin}}(x) + \ln p_Y^{\mathsf{coin}}[\phi_M(x)]$$

Satisfies condition 2 of Prop 1. Thus, multiplying by kT and using definitions of p_X^{coin} and p_Y^{coin}:

$$\begin{aligned} Q_{\mathsf{coin}}(x) &= kT\{-\ln p_X^{\mathsf{coin}}(x) + \ln p_Y^{\mathsf{coin}}[\phi_M(x)]\} \\ &= kT \ln\{\ell(x) + \log_2 m_Y[\phi_M(x)]\} \end{aligned}$$

Coin-Flipping Realization

Zero Entropy Production

$$Q_{\mathsf{coin}}(x) = kT\{-\ln p_X^{\mathsf{coin}} + \ln p_Y^{\mathsf{coin}}[\phi_M(x)]\}$$

Using

$$\langle Q_{\mathsf{coin}} \rangle_{p_X} = \sum_{x \in X} p_X(x) Q(x)$$

We can verify that:

$$\begin{aligned} \langle Q_{\mathsf{coin}} \rangle_{p_X} &= kT\{S(p_X^{\mathsf{coin}}) - S(p_Y^{\mathsf{coin}})\} \\ \implies \Sigma(p_X^{\mathsf{coin}}) &= S(p_Y^{\mathsf{coin}}) - S(p_X^{\mathsf{coin}}) + S(p_X^{\mathsf{coin}}) - S(p_Y^{\mathsf{coin}}) = 0 \end{aligned}$$

Coin-Flipping Realization

Associated Heat Function of Coin-Flipping Realization for TM M

$$Q_{\mathsf{coin}}(x) = kT \ln\{\ell(x) + \log_2 m_Y[\phi_M(x)]\}$$

Recall definition of m_Y:

$$m_Y^{\mathsf{coin}}(y) = \sum_{x:\phi_M(x)=y} 2^{-\ell(x)}$$

- $\log_2 m_Y[\phi_M(x)]$ minimal for logically reversible ϕ_M.
- Q_{coin} minimal for short and logically reversible input programs.

Coin-Flipping Realization

Levin's Coding Theorem for UTM

$$-\log_2 m_Y(y) = K(y) + O(1)$$

Heat Function for UTM

$$Q_{\mathsf{coin}}(x) = kT \ln 2\{\ell(x) - K[\phi_M(x)]\} + O(1)$$

- Q_{coin} achieves its minimum value when x is the shortest program capable of producing $\phi_U(x)$ (always true if ϕ_U is reversible).

$$\min_{x:\phi_U(x)=y} Q_{\mathsf{coin}}(x) = O(1)$$

Coin-Flipping Realization

Expected Heat of Coin-Flipping Distribution

Recall that

$$\langle Q_{\mathsf{coin}} \rangle_{p_X} = kT\{S(p_X^{\mathsf{coin}}) - S(p_Y^{\mathsf{coin}})\}$$

- Difference of entropies is infinite
- Implies infinite expected heat
- Implies infinite expected length of input programs and infinite expected runtime

Coin-Flipping Realization

Initial Distribution for Minimum Expected Heat

Input distribution can be varied to minimize $Q(x)$ in a UTM:

$$p_X^{\mathsf{min}}(x) = \delta(x_0, x)$$
$$Q_{\mathsf{coin}}(x_0) = \min_{x \in X} Q(x) = O(1)$$
$$\langle Q_{\mathsf{coin}} \rangle_{p_X^{\mathsf{min}}} = O(1)$$

But then EP is no longer 0:

$$\Sigma(p_X^{\mathsf{min}}) = S(p_Y^{\mathsf{min}}) - S(p_X^{\mathsf{min}}) + O(1) > 0$$

Dominating Realization

Heat Function for Dominating Realization of TM M

Can be shown that $G(x) = \ln(2) K[x|\phi_M(x)]$ satisfies condition 2 of Prop 1. Thus

$$Q_{\mathsf{dom}} = kT \ln(2) K[x|\phi_M(x)]$$

is the heat function for a realization, called the *dominating realization*, of TM M.

- Inputs generating a lot of heat are large and incompressible, and ϕ_M is non-invertible for that input
- Inputs generating little heat are those for which ϕ_M is invertible
    - For these inputs, $Q(x) = O(1)$

Dominating Realization

Non-Computability

$$Q_{\mathsf{dom}} = kT\ln(2)K[x|\phi_M(x)]$$

- Dominating realization is not computable
- It is upper semi-computable
  - Can be obtained in limit by sequence of increasingly efficient computable realizations $Q_n(x)$
  - Converges on $Q_{\mathsf{dom}}(x)$ from above

Dominating Realization

Efficiency of Dominating Realization

For any other *computable* realization with heat function $Q(X)$:

$$Q(x) \geq Q_{\mathsf{dom}} - kT\left[\ln(2)K(Q/kT) + K(\phi_M)\right] + O(1)$$

- Q_{dom} is minimal up to a negative constant.
  - For $Q(x) \leq Q_{\mathsf{dom}}$, ϕ_M has to have high complexity, or Q has to have high complexity
- The above inequality only holds for computable realizations.

Dominating Realization

Applicability of the Dominating Realization

$$Q(x) \geq Q_{\mathsf{dom}} - kT\left[\ln(2)K(Q/kT) + K(\phi_M)\right] + O(1)$$

- Above inequality only holds if LHS is Q for a *computable* realization
- Validity of Church-Turing Thesis extends applicability to all physical realizations

Dominating Realization

Comparison With Coin-Flipping Realization

- Coin-Flipping Realization uncomputable, thus, not necessarily true that $Q_{\text{coin}}(x) \geq Q_{\text{dom}}(x) + O(1)$. Can be shown that

$$Q_{\text{coin}}(x) \geq Q_{\text{dom}}(x) + O\{\log(K[\phi_U(x)])\}$$

- Minimal heat production for output y:
  - Q_{coin} minimized for $\ell(x) = K(\phi_M(x))$. Finding such a program is uncomputable.
  - Q_{dom} minimized if $\phi_M^{-1}(y) = x$. Finding such a program as simple as "print y".
- Coin-Flipping realization has 0 EP on coin-flipping distribution. Dominating realization has $EP > 0$ for all input distributions

Dominating Realization

Heat VS. Complexity Trade-off

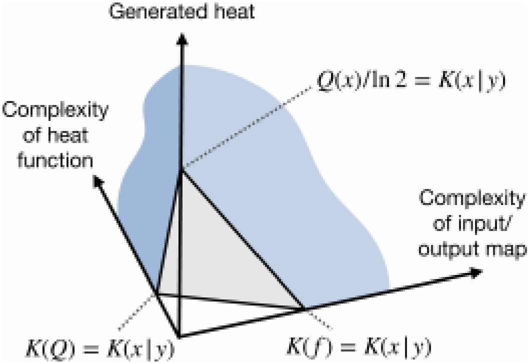
$$Q(x) \geq Q_{\mathsf{dom}} - kT[\ln 2K(Q/kT) + K(\phi_M)] + O(1)$$

Using $Q_{\mathsf{dom}} = kT \ln 2K[x|\phi_M(x)]$ and re-arraigning gives:

$$Q(x)/\ln 2 + K(Q) + K(f) \geq K(x|y) + O(1)$$

- Every computation mapping x to y comes with a "cost" of $K(x|y)$
- Cost can be paid by generating heat, having a high complexity heat function, or having a high complexity mapping f

Heat Vs. Complexity Trade-off



Heat VS. Complexity Trade-off

Example: Erasing a Bitstring

$$Q(x)/\ln 2 + K(Q) + K(f) \geq K(x|y) + O(1)$$

- Consider the an example where f erases a long and incompressible bitstring x.
- $x \mapsto y$ comes with an intrinsic cost of $K(x|y) = K(x) \approx \ell(x)$

Heat VS. Complexity Trade-off

Generate a Lot of Heat

Take f to be

$$f(x') = `000...000' \; \forall x'$$

- f has low complexity
- Using dominating implementation, $Q(x)/\ln 2 = K(x|y) = K(x) \approx \ell(x)$
  - Heat function has low complexity
  - x long and incompressible implies high heat generation

Heat Vs. Complexity Trade-off

Have a High Complexity Heat Function

Can be shown that the following heat function satisfies conditions of Prop.1 for dominating realization of $f(x') = \text{‘}000...000'$

$$Q(x') := \begin{cases} Q_{\mathsf{dom}}(x') & x' \notin \{x, \text{‘}000...000'\} \\ Q_{\mathsf{dom}}(\text{‘}000...000') & x' = x \\ Q_{\mathsf{dom}}(x) & x' = \text{‘}000...000' \end{cases}$$

- Generates little heat
- Low complexity f
- x hard-coded into Q implies high complexity heat function

Heat Vs. Complexity Trade-off

Have a High Complexity Mapping

Consider the logically reversible map:

$$f(x') := \begin{cases} x' & x \notin \{x, \text{`}000...000'\} \\ \text{`}000...000' & x = x' \\ x & x' = \text{`}000...000' \end{cases}$$

- Logically reversible maps can be carried out with 0 heat generation
- 0 heat generation would imply minimally complex heat map
- x hard-coded into f implies high complexity mapping

Physical Church-Turing Thesis

Significance of Physical Church Turing Thesis

- Current conclusions only apply to computable realizations
- In principle, non-computable realizations of TM could exist
- Validity of Church-Turing Thesis would imply any physical realization of a TM must follow thermodynamic constraints shown in paper

Conclusion

Final Remarks

- Proposition 1 allows us to relate logical properties of a TM to its thermodynamic properties.
- Coin-flipping realization gives a highly thermodynamically reversible case
  - Infinite expected heat for zero EP input distribution
  - Heat minimizing input distribution implies nonzero EP
- Dominating realization gives lower bound on heat production for any computable realization
  - Upper semicomputable
  - The inequality $Q(x)/\ln 2 + K(Q) + K(f) \geq K(x|y) + O(1)$ allows us to decompose intrinsic cost of mapping $x \mapsto y$ into complexity of heat function, complexity of mapping, and heat production.

References

[1] A. Kolchinsky and D. H. Wolpert, "Thermodynamic costs of turing machines", Physical Review Research **2**, 10.1103/physrevresearch.2.033312 (2020).