Theory of Thermodynamics of Computation

Ming Li*

Computer Science Department
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
mli@math.waterloo.edu

Paul Vitányi†

CWI
Kruislaan 413
1098 SJ Amsterdam
The Netherlands, paulv@cwi.nl

Abstract

We investigate a new research area: we are interested in the ultimate thermodynamic cost of computing from x to y. Other than its fundamental importance, such research has implications for future miniaturization of VLSI chips reducing the energy dissipation below kT (thermal noise), and the similarity distance problem in pattern recognition.

It turns out that the theory of thermodynamic cost of computation can be axiomatically developed. Our fundamental theorem connects physics to mathematics, providing the key that makes such a theory possible. It establishes optimal upper and lower bounds on the ultimate thermodynamic cost of computation.

By computing longer and longer, the amount of dissipated energy gets closer to these limits. In fact, one can trade time for energy: there is a provable time-energy trade-off hierarchy. The fundamental theorem also induces a thermodynamic distance metric. The topological properties of this metric show that neighborhoods are sparse, and get even sparser if they are centered on random elements. The proofs use Symmetry of Information in a basic way.

These notions also find an application in pattern recognition. People have been looking without success for an objective notion of cognitive distance to account for the intuitive notion of 'similarity' of pictures. Thermodynamic considerations lead to a recursively invariant notion of cognitive distances. It turns out that the thermodynamic distance is a universal cognitive distance which discovers all effective features used by any cognitive distance whatsoever.

*Supported in part by NSERC operating grant OGP-046506.

†Partially supported by NWO through NFI Project ALADDIN, and by NSERC International Scientific Exchange Award ISE0125663.

1 Introduction

This is a brief review of the approach, techniques, and results. The full paper will be published elsewhere and is available from the authors.

Computers can be regarded as engines that dissipate energy in order to process information. The ultimate limits of miniaturization of computing devices, and therefore the speed of computation, are governed by unavoidable heat increase through energy dissipation. Such limits have already been reached by current high density electronic chips [19]. Therefore, the question of how to reduce the energy dissipation of computation determines future advances in computing power. Extrapolations of current trends suggest that reduction of the energy dissipation per logic operation below kT (thermic noise) becomes a relevant issue within 20 years. This requires the use of reversible logic for fundamental thermodynamic reasons. In [21] two methods to implement such reversible computations using electronic switching devices in conventional technologies (like nMOS, CMOS, and Charge Coupled Devices) are proposed. We develop a mathematical framework for the theory of thermodynamics of computation, in particular for the ultimate limits on energy dissipation.

In the early fifties, J. von Neumann [24] thought that a computer operating at temperature T must dissipate at least $kT \ln 2$ Joule per elementary bit operation (about 3×10^{-21} J at room temperature), where k is Boltzmann's constant. Around 1960, R. Landauer [15] more thoroughly analyzed this question and concluded that it is only 'logically irreversible' operations that must dissipate energy. An operation is *logically reversible* if its inputs can always be deduced from the outputs. Erasure of information is not reversible. Erasing each bit costs $kT \ln 2$ energy, when computer operates at temperature T. Solidly based on principles of physics, we develop a mathematical

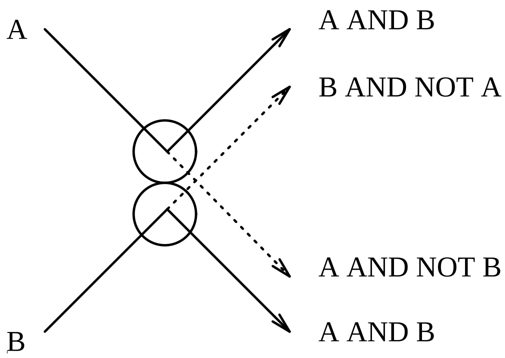

Figure 1: Implementing reversible AND gate and NOT gate

theory of the thermodynamic cost of computation.

2 Physical Background

Briefly, Landauer's line of reasoning ran as follows. Distinct logical states of a computer must be represented by distinct physical states of the computer hardware. Suppose n bits are erased, *i.e.*, reset to zeroes. Before the erasure operation, these n bits could be in any of the 2^n possible states. After the erasure, they are compressed to just one unique state. But, in order to compress the computer's logical state, one must in fact compress its physical state, hence lower the entropy of the hardware. According to the second law, such decrease of entropy of the hardware must dissipate energy.

As an example, consider an ideal computer using elastic frictionless billiard-balls (like molecules). The presence of a ball represents a 1 and no ball represents a 0. The ballistic computer contains mirrors to reflect the balls at some positions. All collisions are perfectly elastic. Between the collisions, the balls travel in straight lines with constant speed, by Newton's first law.

To start the computation, if an input bit is 1 we fire a ball, if an input bit is 0, we do not fire a ball. All input balls are fired simultaneously. Figure 1 implements an AND gate for input A and B. If we set B=1, then we also have a NOT gate for A (and setting A=1 gives a NOT gate for B).

We will also need the constructions in Figure 2 using mirrors to deflect a ball's path, shift a path, delay the ball's motion without changing its final direction, and allow two lines to cross.

It is possible to emulate any computation using the above gadgets. Soppose the setup let all the balls simultaneously reach the output end. After we observe the output, we can simply reflect back all the output balls, including the many 'garbage balls', to reverse the computation. The billiard balls will then come out of the ballistic computer exactly where we sent them in, with the same speed. Then the kinetic energy can be absorbed by the device that kicked the balls in. Then the device is ready for a next round of dissipationless action. A scheme for a ballistic ball computer is shown in Figure 3.

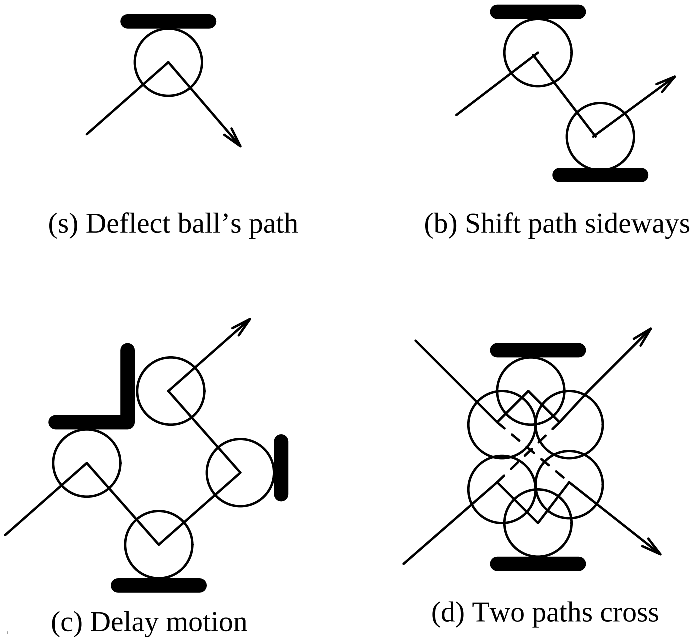

Figure 2: Controlling billiard balls' movements

Suppose we introduce a soft-pad in the device which stops incoming balls dead. If we funnel the garbage balls to the soft-pad, then the computation becomes irreversible because the information represented by the garbage balls is erased. This erasure causes energy dissipation by converting the kinetic energy of the balls to heat.

3 Previous Work

There is a large body of proposals for effective physical realization of (almost) energy free reversible computing. Among others, this has been analyzed with respect to bistable magnetic devises for reversible copying/canceling of records in [15] and Brownian computers [12], for Turing machines and Brownian enzymatic computers [3, 4, 6], with respect to reversible Boolean circuits by [9], for molecular (billiard ball) comput-

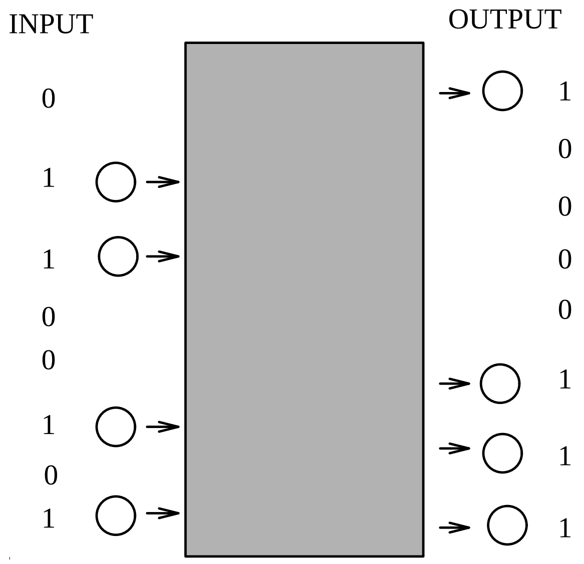

Figure 3: A billiard ball computer

ers by [23], Brownian computing using Josephson devices in [17], quantum mechanic computers in [1, 2, 18] and notably by R. Feynman [7, 8]. All these models seem mutually simulatable. For background information, see [5]. Implementations in current solid state technologies (nMOS, CMOS, CCD) of two methods of using switches to implement reversible computations are presented in [21]. We note that conventional approaches in circuits assume that dissipation occurs when a wire switches from one logic state to another. In [13] a theory based on this assumption is developed and design techniques are presented to reduce this type of dissipation.

In the last three decades there have been many partial precursors and isolated results to the complete mathematical theory developed in this paper. However, it is the formulation of our *Fundamental Theorem* $E(x,y) \approx K(x|y) + K(y|x)$ that provides the key to the theory of thermodynamics of computation. Technically, this theorem is also nontrivially stronger than, and implies all, previous results on this issue which are comparable. Informally, $E(x,y)$ is the optimal thermodynamic cost of computing y from x, and $K(x|y)$ is the length of the shortest effective description of x given y.

At least in [5] Kolmogorov complexity was used in the analysis of reversible computing. In [20] a Kolmogorov complexity based metric for picture similarity was proposed (which is too complex), without clear justification or further results. One of us, stimulated by that paper, proposed (but did not publish) the proper definition $K(x|y) + K(y|x)$ of universal cognitive measure (presented in this paper) in 1988, but did not obtain any further results on it. With respect to cognitive distance, we are not aware of further comparable work: all previous work involves ad hoc approaches, and no objective measure has been proposed [25].

The closest in spirit, and most important stimulation, is the work of W. Zurek [27]. Since he does not provide a formal model, and charges costs in different ways in different places, we need to interpret his work in our model to obtain a proper comparison with our results. He established that the ultimate thermodynamic cost of erasure of a record x, provided the shortest program of length $K(x)$ for x is given, has an upper bound of $K(x)$ (units of $kT \ln 2$). Since we charge both for the provided bits and for the erased bits, this says $E(x,\epsilon) \leq 2K(x)$. Moreover, he gives a lower bound on the thermodynamic cost of computing y from x of $K(x|y)$. In our terminology this is $E(x,y) \geq K(x|y)$. He gives a thermodynamic distance assuming that $K(x|y) + K(y|x)$ bits are provided, which also have to be erased. This shows $E(x,y) \leq 2K(x|y) + 2K(y|x)$ (his information metric).

4 Summary of Results

• Firstly, the minimum thermodynamic cost of a computation is the sum of the energy involved in the providing (inverse erasure) the extra bits required in the course of a computation plus the destroying (erasure, [5]) of the generated garbage bits. This corresponds to the nonreversible part of the computation, and according to Landauer's principle only this nonreversible part of computation dissipates heat. We axiomatize this in terms of effective computations. Our "Fundamental Theorem" gives tight upper and lower bounds on the ultimate limits of the thermodynamic cost of effective computations, and makes a full theory of thermodynamics of computation possible. If we denote by $E(x,y)$ the minimum thermodynamic cost to transform an input x into an output y, then our *Fundamental Theorem* states that, up to a logarithmic additive term, $E(x,y) = K(x|y) + K(y|x)$. Here, $K(x|y)$ is the *Kolmogorov complexity* of x given y for free, [14, 16]. (The Kolmogorov complexity of x is the length of the shortest effective description of x.)

• It has been stated before on the evidence of physical analogies that slow computations may dissipate less energy—like slower moving billiard balls in water generate less friction, [5]. We mathematically prove this statement to be true: there is a proper time-energy trade-off hierarchy of diminishing energy costs using increasing time of computation. Essentially, like in real life, garbage (like disposable information) needs to be compressed before it is destroyed, and this costs time.

• An *effective* distance is a distance which can be computed by a Turing machine. To compute an effective distance we have to spend some minimal thermodynamic cost, the effective *thermodynamic distance*. Thermodynamic distance is symmetric and induces a distance metric. We analyze the topological properties of this metric. This topology is sparse: each d-ball contains at most 2^d elements. (Compare this with a 1-ball around each $x \in \{0,1\}^n$ contains n elements in $\{0,1\}^n$ for Hamming distance.) The more random an object is, the less elements of the same size there are in a d-ball around it; if it is completely random then this number of elements is about $2^{d/2}$. Finally, in each set of size d almost all pairs of elements have distance $2d$ (which is also the maximum if the set is recursively enumerable).

• Given two pictures, are they similar? Answering such question is the goal of pattern recognition. Whatever we mean by picture similarity or picture distance is the first fundamental question that must be dealt with in pattern recognition. For example, Hamming distance is way off between positive and negative prints of the same image: in this case, Hamming distance is the largest, while the pictures are *cognitively* close to human eyes.

This is a question about cognition. Up till now no objective measure for cognitive distance has been found [25]. Intuitively, the minimal cognitive distance between two objects corresponds to the minimal amount of work involved in transforming one object into the other—by brain or computer.

The thermodynamic cost measure readily induces a mathematical theory for a recursively invariant notion of *cognitive distance* that was seemingly undefinable for the long history of pattern recognition. We show that the class of cognitive distances contains a universal cognitive distance, which turns out to be the thermodynamic distance. This universal cognitive distance minorizes *all* cognitive distances: if two pictures are d-close under some cognitive distance, then they are $O(d)$-close under this universal cognitive distance. That is, it discovers all effective feature similarities between two objects.

References

[1] P.A. Benioff. Quantum mechanical Hamiltonian models of discrete processes that erase their histories: applications to Turing machines. *Int'l J. Theoret. Physics*, 21:177–202, 1982.

[2] P.A. Benioff. Quantum mechanical Hamiltonian models of computers. *Ann. New York Acad. Sci.*, 480:475–486, 1986.

[3] C.H. Bennett. Logical reversibility of computation. *IBM J. Res. Develop.*, 17:525–532, 1973.

[4] C.H. Bennett. Dissipation-error tradeoff in proofreading. *BioSystems*, 11:85–90, 1979.

[5] C.H. Bennett. The thermodynamics of computation—a review. *Int'l J. Theoret. Physics*, 21:905–940, 1982.

[6] C.H. Bennett and R. Landauer. The fundamental physical limits of computation. *Scientific American*, pages 48–56, July 1985.

[7] R.P. Feynman. Simulating physics with computers. *Int'l J. Theoret. Physics*, 21:467–488, 1982.

[8] R.P. Feynman. Quantum mechanical computers. *Optics News*, 11:11, 1985.

[9] E. Fredkin and T. Toffoli. Conservative logic. *Int'l J. Theoret. Physics*, 21(3/4):219–253, 1982.

[10] P. Gács. On the symmetry of algorithmic information. *Soviet Math. Doklady*, 15:1477–1480, 1974. Correction, Ibid., 15(1974), 1480.

[11] A. Goldberg and M. Sipser. Compression and ranking. In *Proc. 17th ACM Symp. Theory of Computing*, pages 440–448, 1985.

[12] R.W. Keyes and R. Landauer. Minimal energy dissipation in logic. *IBM J. Res. Develop.*, 14:152–157, 1970.

[13] G. Kissin. Upper and lower bounds on switching energy in VLSI. J. Assoc. Comput. Mach., 38(1): 222-254, 1991.

[14] A.N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems Inform. Transmission*, 1(1):1–7, 1965.

[15] R. Landauer. Irreversibility and heat generation in the computing process. *IBM J. Res. Develop.*, pages 183–191, July 1961.

[16] M. Li and P.M.B. Vitányi. Kolmogorov complexity and its applications. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, chapter 4, pages 187–254. Elsevier and MIT Press, 1990.

[17] K. Likharev. Classical and quantum limitations on energy consumption on computation. *Int'l J. Theoret. Physics*, 21:311–326, 1982.

[18] N. Margolus. Parallel quantum computation. In W.H. Zurek, editor, *Complexity, entropy and the physics of information*, pages 273–287. Addison-Wesley, 1991.

[19] C. Mead and L. Conway. *Introduction to VLSI Systems.* Addison-Wesley, 1980.

[20] D. Schweizer and Y. Abu-Mostafa. Kolmogorov metric spaces. Manuscript, Computer Science, 256-80, California Institue of Technology, Pasadena, CA 91125, 1988.

[21] R.C. Merkle. Reversible electronic logic using switches. Manuscript, Xerox PARC, Palo Alto, CA 94304, 1990.

[22] J. Storer. *Data Compression: Method and Theory.* Computer Science Press, Rockville, MD, 1988.

[23] T. Toffoli. Reversible computing. In *7th Int'l Colloquium Automata, Languages and Programming*, pages 632–644, 1980.

[24] J. von Neumann. *Collected Works*, volume 5. MacMillan, New York, 1963.

[25] S. Watanabe. *Pattern recognition.* John Wiley & Sons, 1985.

[26] J. Ziv and A. Lempel. Compression of individual sequences via variable-rate encoding. *IEEE Trans. Inform. Theory*, IT-24:530–536, 1978.

[27] W.H. Zurek. Thermodynamic cost of computation, algorithmic complexity and the information metric. *Nature*, 341:119–124, 1989.

[28] A.K. Zvonkin and L.A. Levin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Math. Surveys*, 25(6):83–124, 1970.