Sustainable Computing

by Art Scott and Ted G. Lewis

Editor's Introduction

Energy consumption by computers is expanding exponentially along with big data and AI processing. The trend can be broken by adopting alternate approaches to CPU and GPU design. Specifically, Adiabatic Reversible Logic (ARL) has been proposed as the solution. This essay surveys the technology of ARL and gives early examples of actual reversible machines.

Sustainable Computing

by Art Scott and Ted G. Lewis

The Challenge

Imagine shining a 100-watt heat ray on a postage stamp. Without cooling, the stamp eventually burns up. Now, imagine the stamp is a CMOS chip, dissipating dozens or hundreds of watts thrown off by its computing elements. We would need clever cooling technology such as liquid metal convection, portions of the chip turned off, and fans—lots of fans—to maintain thermal equilibrium. This is the conundrum of computing, today. As chips become more capable, they also generate more heat. And the trend is increasing along with performance.

Packing hundreds—even thousands—of stamp-sized computing elements into a relatively confined space, each chip dissipating dozens or even hundreds of watts, multiplies the challenge of power consumption versus performance. In fact, cooling becomes a larger problem than the problem of calculating. We are nearing the end of a long line of performance enhancements paid for by increasing the consumption of power.

Since power consumption of CMOS circuits increases with clock frequency and frequency determines computation speed, increases in performance come with increases in power consumption. Eventually, the amount of heat that must be shed becomes insurmountable and further advancement of computing elements become unsustainable unless the relationship between performance and energy dissipation is broken.

Koomey's law has trumped Moore's law. It says, "the number of computations per joule of energy dissipated doubles about every 1.57 years"[1]. It is validated in actual practice because performance in terms of GFLOPS/KWh rises exponentially over time and dissipated power rises linearly, because it is additive—more circuits and transistors means more dissipated energy. Both lines have been increasing mainly because more power is pumped into chips to energize more transistors. Even if the energy required to erase a bit and flip it from zero to one is made smaller, the demand for more performance grows even faster. Power consumption has been steadily increasing over the decades and will continue unless a new path is taken. Sustainability means curtailing this unbounded growth.

In fact, one can ask, "What is the least amount of power required to erase or flip a bit?" If the minimum is known, we could calculate the ultimate power requirements for all of computing. Using Koomey's law, we could keep on adding power to chips while ignoring heating problems for several more decades.

In fact, Landauer's limit dictates the minimum energy needed—kTln(2), or 285 x 10^{-23} Joules at room temperature[1]—to erase a bit of information, where *K* is Boltzmann constant and *T* is temperature [2]. What if we plugged that number into Koomey's law and limited heat dissipation so the postage stamp chip doesn't burn up? Can we extend the performance curve further without worrying about overheating?

"By the second law of thermodynamics and Landauer's principle, irreversible computing cannot continue to be made more energy efficient forever. As of 2011, computers have a computing efficiency of about 0.00001%. Assuming that the energy efficiency of computing will continue to double every 1.57 years, the Landauer bound will be reached in 2048. Thus, after about 2048, Koomey's law can no longer hold" [1].

Even if Landauer's limit is reached by 2048, stamp-sized chips risk burning up due to the heat thrown off by billions of processing elements. No amount of engineering can overcome this thermodynamic limit. But notice the principle applies to *irreversible* computing. What if computing became *reversible*? Landauer's limit no longer applies. This is known as adiabatic reversible computing, ARL, and is the only known way out [3].

What is "Sustainable Computing?"

Microprocessor circuits dissipate energy for a variety of reasons: resistance in wires, communication with the outside world, leakage, and parasitic capacitance. In addition, computer memory, displays, and wireless connectivity consume even more energy. All of these sources of energy dissipation need to be addressed. Where to begin?

We can begin at the heart of the matter: the CPU/GPU. The current path we are on is unsustainable—sooner or later, we either run out of power or we run out of ways to cool computing circuits. Either way, a new path must replace the current path if we are going to reach higher levels of performance. (This article ignores the detrimental impact on climate change, which is another issue for another day.)

Sustainable computing means different things to different people, but for the purposes of this paper, we define sustainable computing as, "diminishing power consumption as the number of operations per second increase." This definition departs from current practice because it achieves greater GFLOPS/KWh by decreasing KWh's faster than increasing GFLOPS. That is, sustainability means decreasing power consumption towards zero—if not to zero!

Sustainability is achieved by adiabatic computing. That is, instead of dissipating energy when an information bit is flipped, we adiabatically reuse it. The degree of sustainability is equal to the percentage of energy reused. Theoretically, we can come arbitrarily close to 100 percent reuse

---

[1] 0.7377 Joules is the amount of energy used to move one pound a distance of one foot.

by employing reversible circuits in place of irreversible circuits for performing Boolean operations. The ideal CPU/GPU would only need to be powered up initially and run forever on recycled power. Of course, this is unrealistic, but it is a goal worth striving for. How close to net zero can we come?

An adiabatic reversible computer runs forward and backward and reuses the energy dissipated by reversible circuits. By shuttling energy back and forth between reversible circuits, and thermodynamically isolating circuits, it is possible to build an adiabatic reversible computer that produces little, if any heat, and therefore uses little, if any, power.

Reversible Circuits Are Real

Consider a very well insulated frictionless car as an analog of an adiabatic computer. Once you accelerate to a cruising speed, an adiabatic car can theoretically travel indefinitely on its momentum because it loses no energy to friction and radiates no energy into the environment. It can slow down using 100 percent efficient regenerative braking and then speed up again as long as momentum is conserved and no energy is dissipated by friction. Thus, reversing its acceleration recaptures all of the energy. Other than the initial energy needed to accelerate the car, no additional energy is consumed. The driver simply recycles the energy once it is put into the car. In practice no such car exists, but we drive cars that recover energy during breaking. We can expect to get closer to this ideal in the years ahead.

The NAND gate used in most switching circuits today is not reversible, meaning the circuit loses information every time it operates. For example, if a NAND circuit shows a 1 output, we cannot tell which three combinations of *a* and *b* (00, 01, and 10) gave that output. Every time a NAND gate switches, it generates a small amount of heat due to this loss of information. Even if the NAND could be implemented with superconducting materials that eliminate electrical resistance, the information loss would increase entropy and generate a little heat according to Landauer's principle.

In 1982 Edward Fredkin and Tommaso Toffoli of MIT devised a new kind of gate that is completely reversible [4]. Fredkin's reversible-NAND, for example, is reversible because it can be run backward to regenerate its original inputs. This requires that reversible gates have the same number of outputs as inputs, even though some inputs pass straight through unchanged; in particular the Fredkin reversible-NAND is reversible because given its outputs *a*, *b*, *c* we can tell exactly what inputs produced them. For more on reversible gates, see [3] and [5].

Reversibility means there is no loss of information because all inputs can be recovered from the output. Loss of information increases entropy, which is dissipated as heat. On the other hand, when there is no loss of information, entropy is unchanged, and there is no dissipation of heat. Thus, reversibility implies the absence of energy dissipated as heat during logic operations. (We

still have to worry about other sources of energy dissipation such as resistance in wires, displays, and communication.)

Adiabatic operation means that within a closed system, any heat gain or loss is nullified—an increase in one portion of the closed system is used to power another portion. Adiabatic and reversible logic (ARL) combines to minimize heat dissipation. For example, Sharmila and Bhanumathi describe a reversible MOS design that performs better than irreversible CMOS [6].

Unfortunately, there is a penalty associated with ARL implementation of digital circuits. They must run slower, because the amount of energy dissipated is proportional to clock frequency and the rise time of the clock signal. To overcome this problem, Frank et al. propose an architecture based on static two-level adiabatic logic (S2LAL)[7]: "S2LAL is, we think, the fastest possible such family (among fully pipelined sequential circuits), having a latency per logic stage of one "tick" (transition time), and a minimum clock period (initiation interval) of eight ticks. S2LAL requires eight phases of a trapezoidal power-clock waveform (plus constant power and ground references) to be supplied.[2] S2LAL should be capable of demonstrating a greater level of energy efficiency than any other semiconductor- based digital logic family known today."

Figure 1 compares a trapezoidal clock with a traditional clock. During the transition from one bit state to the other the amount of energy dissipated in the circuit scales inversely with the clock period and thereby can be made arbitrarily small. According to Frank, a sharply rising clock requires more energy than a gradually rising clock. By reducing the "angle of attack" the amount of dissipated energy is reduced. The trick is to find the right balance between rise-time and width of the trapezoidal clock pulse.

According to Frank, "The long term goal of the program is to build computing devices that go through their cycle of operations just coasting from one cycle to the next. In the long run, reversible computing is the only thing we can do to keep pushing performance limits" [7]. The S2LAL circuit captures the energy from the previous operation and provides it to the next operation, just like your frictionless car captures energy when braking so it can be used to accelerate.

Osborne et al. have demonstrated reversible gates in superconducting circuits employing adiabatic reversible logic circuits, achieving inverse scaling between gate time and energy cost [8]: "We are developing reversible fluxon logic (RFL) based on ballistic-reversible gates." Sometimes called the scattered billiard ball model, ballistic-reversibility conserves all of the potential energy of fluxons in logic operations. The authors claim, "Ballistic-reversible gates

---

[2] DeBenedictis, E. P., Inversion for S2LAL. Technical note ZF004 v1.01. Zettaflops, September 8, 2020.

moreover conserve a large fraction of the kinetic fluxon energy[3], thus yielding conservation of up to 97% of fluxon energy."

In 2020 a group at Yokohama University, Japan reportedly implemented a full microprocessor based on adiabatic principles [9]. The authors claim to "[d]emonstrate register file R[ead]/W[rite] access, ALU execution, and program branching performed at 100kHz under the cryogenic temperature of 4.2K. We also successfully demonstrated a high-speed chip of the microprocessor execution units up to 2.5GHz." Adiabatic computing operates near zero Kelvin and room temperature as long as circuits are reversible.

Torres et al. claim, "Novel nanotechnologies like quantum-dot Cellular Automata (QCA) allow for computations with very low energy dissipation and, hence, are promising candidates for breaking [the Landauer] limit. [We] simulated a QCA in a physics simulator enabling a precise consideration of how energy is dissipated in QCA designs. Our results provide strong evidence that QCA is indeed a suitable technology for near zero-energy computing"[10].

A commercial grade processor QCA running at room temperature has yet to be constructed, but an implementation based on quantum dot running at room temperature seems possible. Sarvaghad-Moghaddam et al. propose and simulate the design of a quantum dot cellular automata (QCA) to construct reversible Fredkin-Toffoli gates [11]. For a tutorial on electron spin implementation of quantum dots, see [12].

A 2019 Notre Dame thesis by Rene Celis-Cordova demonstrates the feasibility of constructing and simulating a 16-bit RISC processor using 90nm CMOS technology. Cordova reports, "The 16-bit adiabatic microprocessor is successfully implemented in 90 nm technology with an operating frequency of 0.5 GHz, which demonstrates the design of a real-life circuit using adiabatic reversible logic and shows a promising future for energy-efficient computing" [13].

As of 2020, there are no commercial grade ARL microprocessors on the market. But, various computing elements have been demonstrated with 10 times improvement in energy dissipation. The field is young, and development is slowed by low investment. And yet, we will never overcome unbounded power consumption by continuing to follow the current irreversible path. We must switch paths—adiabatic reversible logic may be the answer.

---

[3] In physics, a fluxon is a quantum of electromagnetic flux. A fluxon in a ballistic-reversible circuit contains the flux of a single flux quantum.
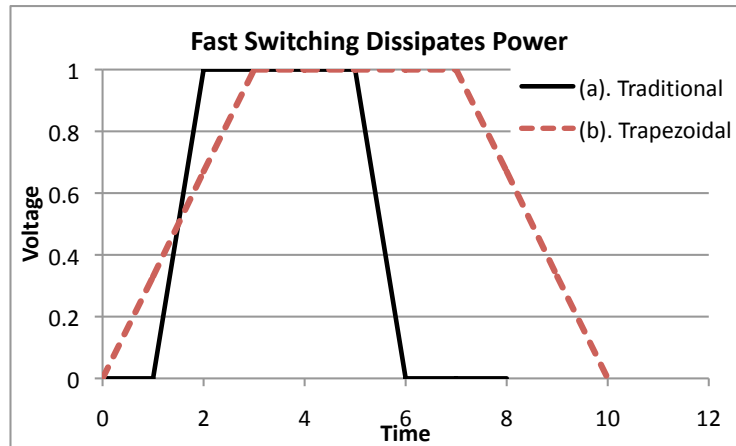
Figure 1. During the transition from one bit state to the other the energy dissipated in the circuit scales inversely with the clock period. (a) A traditional clock dissipates more energy because its clock period is relatively short. (b) The S2LAL clock period is just long enough to allow recycling of most of the dissipated energy.

Reversing Directions – A New Path

It's imperative that we reverse course. Instead of increasing power consumption endlessly, the new path to follow decreases power consumption per unit of computation, endlessly. This is Koomey's law in reverse: "the number of joules of energy dissipated per computation halves about every two years." This will get us to the future by about 2032—the first steps having been demonstrated in the laboratory in 2020.

According to Andrae, "If a traditional 180W processor chip using 5nm node … would process the anticipated operations in 2030 with current transistor technology, an absurd amount of electricity will be used for computing. The same chip using reversible computing would use only 0.08W" [14]. Clearly, burning an absurd amount of electricity is not going to be practical in 2030.

Instead, Andrae notes a steady decline in energy consumption per Figure 2. Measured in Landauer units of kT, this projection forecasts sub-Landauer consumption by 2030. It is a projection, only, and may not happen. We have to make it come true to make computing sustainable.

The impact on the semiconductor industry is rather large. It will be necessary to develop a standard cell library and semi-automatic tools to reduce the time and effort in design and verify ARL circuits. The whole tool chain must adapt to the reality that reduction of power consumption is critical. This will take time, but in the end, the industry will be re-invented.
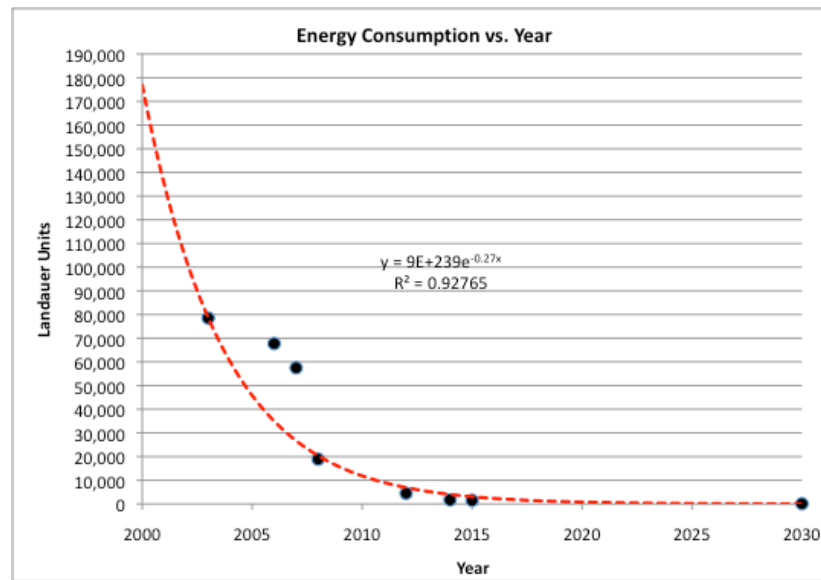
Figure 2. Forecast of sustainable energy demands of computing in terms of Landauer units = kT [14].

References

[1] Koomey, J., Berard, S, Sanchez, M. and Wong, H. Implications of historical trends in the electrical efficiency of computing. *IEEE Annals of the History of Computing* 33. 3 (2010), 46–54; doi:10.1109/MAHC.2010.28, ISSN 1058-6180.

[2] Landauer, R. Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development* 5, 3 (1961), 183–191; doi:10.1147/rd.53.0183.

[3] Denning, P. J. and Lewis, T. G. Computers that can run backwards: reversible computations—which can, in principle, be performed without giving off heat—may be the future of computing. *American Scientist* 105, 5 (2017), 270–274.

[4] Fredkin, E. and Toffoli, T. Conservative Logic *Int. J. Theor. Phys.* 21, 219 (1982).

[5] Lewis, T. G. Is Computing in reverse the next big thing? *Ubiquity* May 25, 2017.

[6] Devi, S. S., and Bhanumathi, V. Design of reversible logic based full adder in current-mode logic circuits. *Microprocessors and Microsystems* 76 (2020). https://doi.org/10.1016/j.micpro.2020.103100.

[7] Frank, M. P., Brocato, R. W., Tierney, B. D., Missert, N. A., Hsia, A. H. Reversible computing with fast, fully static, fully adiabatic CMOS. arXiv:2009.00448v2 [cs.AR]. Sept. 2, 2020.

8

[8] Osborn, K. D. and Wustmann, W. Reversible fluxon logic with optimized CNOT gate components. arXiv:2010.15991v1 [quant-ph]. Oct. 29, 2020.

[9] Ayala, C. L. , Tanaka, T. , Saito, R., Nozoe, M., Takeuchi, N. and Yoshikawa, N. MANA: A monolithic adiabatic integration architecture microprocessor using 1.4zJ/op superconductor Josephson junction devices. In *2020 IEEE Symposium on VLSI Circuits* (Honolulu, 2020). IEEE, New York, 2020, 1–2; doi: 10.1109/VLSICircuits18222.2020.9162792.

 [10] Torres, F. S. et al. Near zero-energy computation using quantum-dot cellular automata. *ACM Journal on Emerging Technologies in Computing Systems (JETC)* 16, 1 (2019), 1–16.

[11] Sarvaghad-Moghaddam, M., Orouji, A. A., Ramezani, Z., Amiri, I. S., and Nejad, A. M. Reversible gates in emerging quantum-dot cellular automata technology: An innovative approach to design and simulation. arXiv:1803.11017v2 [cs.ET]. Dec. 9, 2019.

[12] Meyer, C. Quantum computing with semiconductor quantum dots. Forschungszentrum Jülich. 2009.

[13] Celis-Cordova, R. Reversible computing: The design of an adiabatic microprocessor. Master's thesis. University of Notre Dame. 2019.

[14] Andrae, A. S. G. Prediction studies of electricity use of global computing in 2030. *International Journal of Science and Engineering Investigations* 8, 86 (2019).

Biographies

Ted G. Lewis is an author, speaker, and consultant with expertise in applied complexity theory, homeland security, infrastructure systems, and early-stage startup strategies. He has served in both government, industry, and academe over a long career, including, Executive Director and Professor of Computer Science, Center for Homeland Defense and Security, Naval Postgraduate School, Monterey, CA. 93943; Senior Vice President of Eastman Kodak, President and CEO of DaimlerChrysler Research and Technology, North America, Inc., and Professor of Computer Science at Oregon State University, Corvallis, OR. In addition, he has served as the Editor-in-Chief of a number of periodicals: *IEEE Computer Magazine*, *IEEE Software Magazine*, as a member of the IEEE Computer Society Board of Governors, and is currently Advisory Board Member of ACM Ubiquity and Cosmos+Taxis Journal (The Sociology of Hayek). He has published more than 35 books, most recently including *The Signal: A History of Signal Processing*, *Book of Extremes: The Complexity of Everyday Things*, *Bak's Sand Pile: Strategies for a Catastrophic World, Network Science: Theory and Practice*, and *Critical Infrastructure Protection in Homeland Security: Defending a Networked Nation,* 3rd ed. Lewis has authored or co-authored numerous scholarly articles in cross-disciplinary journals such as *Cognitive Systems Research*, *Homeland*

Security Affairs Journal, Journal of Risk Finance, Journal of Information Warfare, IEEE Parallel & Distributed Technology, Communications of the ACM, and *American Scientist.* Lewis resides with his wife and corgi dogs, in Monterey, California.

Art Scott is the founder of EETALL: Energy Efficient The Answer to Landauer's Limit. Actively "Rebooting Computing", transferring energy-efficient computing technologies—open perfectly adiabatic static 2LAL CMOS—to leap the Landauer Limit wall. Art is many faceted; he creates outside the box; he is a serial entrepreneur-intrapreneur and change agent, and is flexible-adaptable-situational. He has worked for Silicon Valley icons such as SRI, Applicon, Computer Sciences Corporation, Informatics, Atari R&D, Interactive Research Corp., Samsung Information Systems America. And he has participated in several startups: Digital Video Inc. (1985-2001), Ravisent Technologies (1998-1999), and recently, EETALL. Art lives the "Aloha Spirit" way with his wife in Menlo Park, California.