Rebooting Computing, computing, Moore's law, adiabatic computing, reversible computing, SCRL, 2LAL, retractile cascade, superconducting electronics

Editor: Erik P. DeBenedictis, Sandia National Laboratories; epdeben@sandia.gov

Rebooting Computing

The Opportunities and Controversies of Reversible Computing[1]

Erik P. DeBenedictis, Sandia National Laboratories
Jesse K. Mee, Air Force Research Laboratory
Michael P. Frank, Sandia National Laboratories

Industry's inability to reduce logic gates' energy consumption is slowing growth in an important part of the worldwide economy. Some scientists argue that alternative approaches could greatly reduce energy consumption. These approaches entail myriad technical and political issues.

For all the talk about computing's excessive energy costs, the community doesn't consistently work to lower them. Vacuum tubes were expensive and unreliable in the early days of computing, so Boolean logic circuits were designed to minimize the number of tubes. This design style hasn't changed over the years, even though we now put so many transistors on a chip we have to turn most of them off to avoid overheating.

The future path for reducing computers' energy consumption has an all-too-familiar story line. Theoretical physics says the minimum possible energy dissipation for a conventional logic operation is set by the thermodynamic energy level kT, or about 4.14 zeptojoules (a zeptojoule is 10^{-21} joules) at room temperature. However, industry's CMOS roadmaps show the limit to be thousands of times higher, as CMOS was never intended to reach the thermodynamic limit.

Some scientists have experimentally demonstrated the feasibility of attaining the kT dissipation level using an alternative approach called reversible computing, which avoids irreversibly discarding information and the energy use associated with doing so. And they argue that energy dissipation could reach even lower levels.

However, as has been the case with finding alternatives to the gasoline engine and supersonic flight, these scientists are underfunded and underappreciated because society focuses its attention and resources on mainstream approaches. Nonetheless, alternative approaches sometimes work and drive progress.

---

[1] Sandia National Laboratories approved for unlimited unclassified release SAND2017-8724 J
Published as DeBenedictis, Erik P., Jesse K. Mee, and Michael P. Frank. "The Opportunities and Controversies of Reversible Computing." *Computer* 50.6 (2017): 76-80. **DOI:** 10.1109/MC.2017.177

Different Computing Levels

To explain minimum energy consumption by computers in a consistent manner, we've divided energy-saving computing approaches into four levels, as Figure 1 shows. Each level, defined by the amount of energy used per gate operation, has its own technical and political issues.

Levels A and B represent the energy efficiency increase that occurred from the onset of the computing age with vacuum tubes to the present. Levels C and D are research ideas that yield even lower energy levels.

Level A: minimum signal energy

The concept of minimum energy level had a different meaning for the first computers in the 1940s, when the active components were vacuum tubes. The concept of minimum energy was reserved just for signals, originally electrical waves that transmitted voice over radio and later the ones and zeros transmitted in a computer. The century-old vacuum tube and the modern transistor can both process signals to energy levels as low as the kinetic energy in the electrons involved, due to their thermal motion, as Figure 1a shows. This energy is kT, k being Boltzmann's constant and T being the temperature in Kelvins. Modern electronic devices have a noise factor of about two, meaning the device acts as though noise was twice as large as it really is due to design imperfections. Thus, the circuits need twice as much energy.

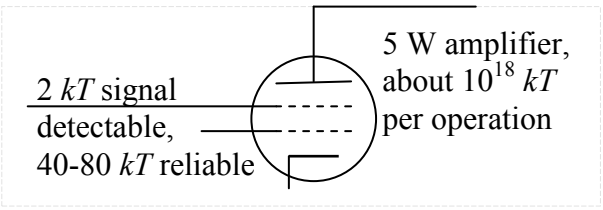
Level B: minimum CMOS gate energy

Advanced technology is bringing gate energy consumption closer to the minimum energy usage that the laws of physics permit. The vacuum tubes in the original EDVAC (Electronic Discrete Variable Automatic Computer) from the late 1940s consumed about 5 W and operated at 1 kHz, which represents about 10^{18} kT per gate operation. The typical gate in the modern CMOS technology shown in Figure 1b consumes about 30,000 kT per operation.[1] The gap between gate energy usage and the minimum possible signal energy has closed from 10^{18} to 10^{5}, but gates still consume much more than the minimum signal energy.

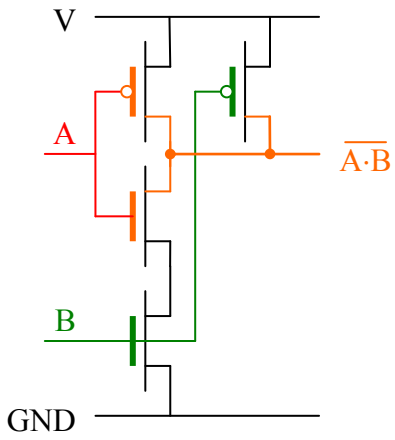
Level C: energy-efficient circuits using existing transistors

Decreasing gates' energy consumption is easy, but doing so increases system cost and slows performance. The International Technology Roadmap for Semiconductors created road maps for both the high-performance CMOS circuits used in consumer-grade electronics and low-power variants. Both types use the same lithographic-linewidth fab approach and circuit diagram for gates, but they differ in parameters such as doping levels, operating voltage, and transistor size.[1]

There are various types of energy-efficient, or adiabatic, gate circuit designs—such as SCRL (split-level charge recovery logic),[2] 2LAL (two-level adiabatic logic),[3] and retractile cascades[4]—that recycle energy to reduce net consumption. Figure 1c shows that the 2LAL circuit has more transistors per gate and more clock wires to lay out than the CMOS circuits with the same NAND-gate function shown in Figure 1b.



(a) Communications signals

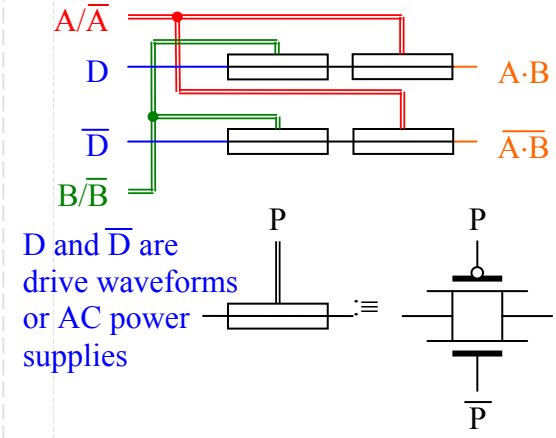


2,500 *kT* × 4 devices[1]

20,000 *kT* signal energy

Total 30,000 *kT*

(b) CMOS limits

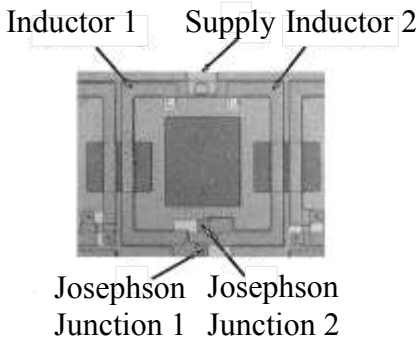


Energy recycled in 8 devices

Energy recycled in 2× wires

Recycling inefficiencies only

(c) Adiabatic circuits using familiar transistors



(d) Sub *kT* energy levels[6]

Note to artist: This figure was prepared using the default Microsoft Word grid for consistent spacing. The figure could be shrunk. Section (d) is an image from an IEEE publication. Could you please overlay lettering in the preferred font as indicated?

Figure 1. Logic operations' energy consumption by various technology level: (a) is a vacuum tube, (b) is two-input CMOS NAND gate, (c) is a 2LAL (two-level adiabatic logic) circuit, and (d) is a photomicrograph of an nSQUID, a two-junction SQUID (superconducting quantum interference device) that has negative mutual inductance between the two arms that can move signals around a system. Each level yields a characteristic minimum power-usage level for signals and for gate dissipation. Technological maturity, challenges, and business issues vary by level.

Both approaches moderately decrease energy consumption per operation when implemented with current devices but also use more chip area per function and have lower throughput. Thus, their economic viability depends on the relative cost of buying a host computer and powering the machine over its lifetime. These circuits could yield quite a bit more improvement with additional technology development in AC power supplies.

The choice whether to use low-energy circuits entails an economic tradeoff, as Figure 2 illustrates. Say the microprocessor in a desktop computer costs $75 and consumes an average of 25 W. Assuming a lifetime of three years and a typical cost of wall-plug power of $1 per watt per year, the combined capital and energy cost will be $150. Doubling the energy efficiency would double the purchase price. Thus, this approach's overall cost would be (2 × $75) + (0.5 × $75) = $187.50. This would be more expensive overall and thereby discourage the use of such circuits. However, future developments could change this.

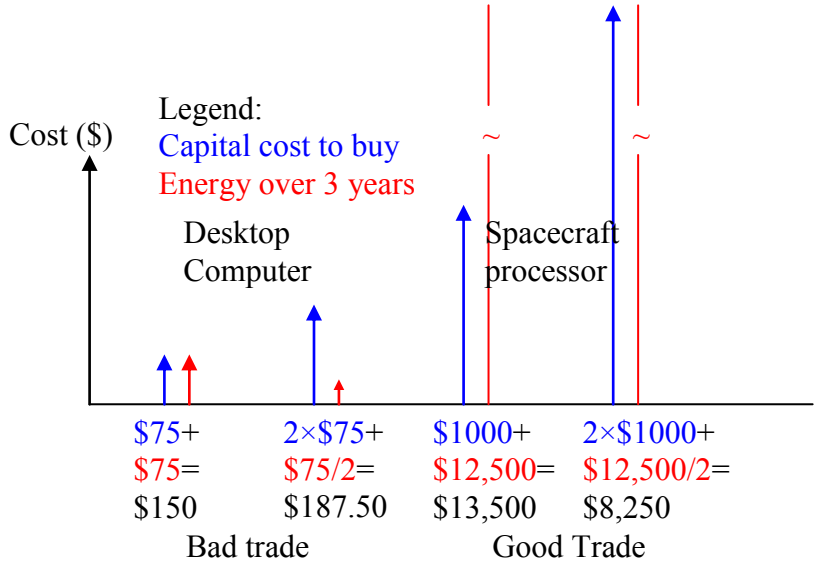


Figure 2. Saving energy via energy-efficient circuits. (a) Energy-efficient CMOS devices running on wall-plug power can be more expensive to own and operate than those that aren't energy efficient, except in (b) systems for which power costs are higher, such as spacecraft.

While energy-efficient circuits can be a losing proposition for desktop computers at wall-plug energy prices, these circuits might be viable today in systems where the power costs are higher and energy availability is more problematic.

Spacecraft are an example of a system for which energy costs are very high. A space-grade multijunction photovoltaic solar-to-electrical conversion assembly can cost more than $1,000 per watt. Furthermore, arrays must be sized to provide end-of-life power to the spacecraft bus and payload architecture. Consequently, the energy source for a 10-W processor could cost as much as $12,500. On the other hand, the initial cost to add energy-efficient processors is only about $1,000. Therefore, doubling a system's power efficiency makes economic sense, as Figure 2 illustrates.

Energy-efficient circuits could become more viable over time if manufacturing cost per

device continues to decline, even if energy costs stay the same. Devices' manufacturing costs have been declining exponentially due to Moore's law and continue to do so for 3D systems, principally memory. However, the 2D fin field-effect transistor (FinFET) chips used in some microprocessors cost almost the same as previous generations. If energy-efficient circuits could be manufactured in 3D, the technology would not only benefit from a declining manufacturing cost relative to energy consumption but also from reduced susceptibility to device overheating due to the lower-power circuits in the 3D stack. So energy-efficient circuits might represent a cost-effective alternative to FinFETs and other exotic computing devices.

Level D: reaching 1 kT, a long-term vision

Theorists developed an extensive set of literature in the 1960s, 1970s, and 1980s about the minimum energy required by computation, including the use of various gate sets and architectures, as noted in a December 1982 special issue of the International Journal of Theoretical Physics. While the results were thoroughly peer reviewed, the scientists' theories couldn't be tested because the energy dissipation from parasitic losses and manufacturing imperfections were far higher than the minimum energy levels the researchers had to measure.

A second body of literature emerged in the 1990s and 2000s, when engineers responsible for dramatic CMOS advances focused on the ultimate limits of CMOS scaling. By this time, researchers had controlled parasitic losses and improved manufacturing processes. A smaller but significant body of peer-reviewed research emerged that expressed concern that the earlier literature was fundamentally flawed.

There's no dispute that analyzing circuits like the one shown in Figure 1a can be valid even at 1 kT energy-consumption levels and even when energy recycling allowed net energy dissipation below this level, yet every circuit analyzed in detail showed a higher limit. The concern was that an unappreciated physics principle would doom all future circuits to have limits well above 1 kT.

However, several peer-reviewed experiments on ultra-low energy control use light tweezers to move microscopic plastic balls, measuring force and energy consumption. The results show that the system can control the balls using about 1 kT of energy.

In addition, an experimental logic gate using superconducting electronics,[5] shown in Figure 1d, consumed less than 1 kT of energy.[5] However, the laboratory-based structure isn't currently suitable for logic products. One of this article's authors visited the lab and found that the limitations' causes, such as design bugs, could be fixed by another few experiments.

Level E: Ultimate limits

Is there is an improvement path like Moore's law for circuits' energy consumption? Theory suggests that it's possible to compute with "arbitrarily low energy."

The communications theory represented in Figure 1a is, in fact, correct. This means the zeros and ones in a computer must contain 40–80 kT of energy to avoid errors that thermal noise causes. However, the energy per computation isn't the energy that the signals themselves contain but instead reflects the loss when the signals are moved from gate to gate and when they interact with one another. These losses are proportional to the amount of energy moved—like friction.

The loss is analogous to the effort needed to move 40–80 bricks from your driveway to your backyard. You could carry the bricks, or reduce the energy expended by rolling them in a wheelbarrow, or decrease it even more by using a steel-wheeled railcar that runs on low-friction rails. The bricks weigh the same in all these situations, but the energy loss goes down due to lower friction. In contrast, the dimensional scaling of Moore's law effectively shrinks your yard and lightens the bricks, allowing you to move the bricks more efficiently with the same apparatus.

Reversible computing's improvement path would reduce frictional losses from one generation to the next, albeit not to the extent that Moore's law calls for chip performance to increase. To create an improvement path, the fraction of energy recycled from one generation to the next would have to rise to 90 percent, then 99 percent, then 99.9 percent, and so on. Reducing friction this way would require improved designs and more precise devices.

Reversible computing and its variants could reduce a broad range of computing systems' energy consumption, with the benefit being independent of and complementary to reductions yielded by architectures, applications, and devices. The early reversible-computing adopters will likely be applications for which energy costs are high, such as spacecraft, systems using nonrechargeable batteries, systems that operate at cryogenic temperatures, and implantable medical electronics. Once we demonstrate and refine the technology for a few generations, we might learn enough to also use reversible computing in mainstream applications.

Acknowledgments

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the US Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

References

1. Nikonov, Dmitri E., and I.A. Young, "Benchmarking of Beyond-CMOS Exploratory Devices for Logic Integrated Circuits," IEEE J. Exploratory Solid-State Computational Devices and Circuits, vol. 1, 2015, pp. 3–11.
2. Younis, Saed G. Asymptotically Zero Energy Computing Using Split-Level Charge Recovery Logic," PhD dissertation, Dept. Electrical Eng. and Computer Science, MIT, 1994.
3. Anantharam, Venkiteswaran, et al., "Driving Fully-Adiabatic Logic Circuits Using Custom High-Q MEMS Resonators," Proc. 2004 Int'l Conf. Embedded Systems and Applications (ESA 04), 2004, pp. 5–11.
4. Campos-Aguillón et al., "A Mini-MIPS Microprocessor for Adiabatic Comput i ng," IEEE Int'l Conf. Rebooting Computing (ICRC 16), 2016, pp. 1–7.
5. Ren, Jie, and V.K. Semenov, "Progress with Physically and Logically Reversible Superconducting Digital Circuits," IEEE Trans. Applied Superconductivity, vol. 21, no. 3, 2011, pp. 780–786.

Erik P. DeBenedictis is a technical staff member at Sandia National Laboratories' Center for Computing Research. Contact him at epdeben@sandia.gov.

Jesse Mee is the deputy program manager of the Space Electronics Technologies Program at Air Force Research Laboratory's Space Vehicles Directorate. Contact him at jesse.mee@us.af.mil.

Michael P. Frank is a technical staff member at Sandia National Laboratories' Center for Computing Research. Contact him at mpfrank@sandia.gov.