Art Scott and Michael Frank on Energy-Efficient Computing

by Ted G. Lewis

Editor's Introduction

Clock speeds of computing chips have leveled off dramatically since 2005, and putting more cores in systems on a chip (SoC) has produced more heat, adding a new ceiling to further advances. Leading-edge researchers, like Mike Frank, and dedicated technologists with a wealth of experience, like Art Scott, represent a new vanguard of the leap-forward beyond Dennard scaling and Landauer's limit. Art looks for ways to reduce energy consumption and Mike looks for ways to "architect" future chips according to principles of reversibility. Is the future in reversible, adiabatic computing and simpler architectures using posit arithmetic? My guests think so.

Art Scott and Michael Frank on Energy-Efficient Computing

by Ted G. Lewis

Ted Lewis: You have been in the computing industry for a long time, going all the way back to the beginning of Moore's law. Generally speaking, you have participated in making computers run faster and hotter, but since around 2005 Moore's law has hit a wall because of heat—the end of Dennard scaling has forced chip-makers to face the problem of heat dissipation. As a consequence, clock speeds have not improved for a decade, see Figure 1. Some say the answer is reversible computing. What is reversible computing and how did you become interested in it?

Art Scott: Energy efficient is the answer to Landauer's limit, the act of erasing a bit of information gives off an amount of heat related to the temperature and Boltzmann's constant—a total of 3×10^{-21} joules at room temperature. Reversible computing is energy efficient; therefore, reversible computing is the answer to Landauer's limit. Your readers can find out more online where Wikipedia describes the reason reversible computing is now center stage, "Probably the largest motivation for the study of technologies aimed at actually implementing reversible computing is that they offer what is predicted to be the only potential way to improve the computational energy efficiency of computers beyond the fundamental von Neumann-Landauer limit of $kT \ln 2$ energy dissipated per irreversible bit operation."¹ There, k is the [Boltzmann constant](#) (approximately 1.38×10^{-23} J/K), T is the absolute temperature of the environment, and $\ln 2$ is the [natural logarithm](#) of 2 (approximately 0.69315). Reversible means the computation can be reversed, in other words, the inputs can be obtained from the outputs, by running circuits backwards. However, the purpose of reversible design is not to run backwards, but to avoid heat generation associated with increased thermodynamic entropy,

¹ https://en.wikipedia.org/wiki/Reversible_computing

which rises whenever information is lost. Reversibility preserves the information (because it can be retrieved) and therefore, avoids unnecessary generation of thermodynamic entropy.

TL: There are several important ideas here that need to be parsed out. Landauer's limit goes all the way back to the early 1960s and places a lower limit on the amount of energy needed to destroy a bit of information. Reversible computing is a mechanism for avoiding the destruction of a bit of information to avoid the need to increase thermodynamic entropy. Why is Landauer's limit important to computing and how does it connect with reversible computing?

AS: Landauer's principle and limit are critical, vital information, knowledge to the IEEE rebooting computing initiative.² Based on the second law of thermodynamics. The first step to leaping beyond the Landauer limit is appreciating and understanding Landauer's work. The connection from Landauer's observations to "traditional" semi-partial reversible computing was made when Dr. Edward Fredkin invented the Fredkin gate thus initiating reversible computing. Reversible computing now has a number of computational logic circuits such as the Toffoli gate, Feynman gate, and Peres gate.

Landauer identified two possible sources of heat generation as, dissipation of heat due to irreversibility, and incomplete switching due to fast switching time. Although reversible logic reduces power dissipation, using reversible [complementary metal-oxide-semiconductor] CMOS gates alone in the design will not be sufficient. Adiabatic switching is an energy recovery design approach that helps to improve the performance through limiting power dissipation by progressively charging capacitance and recycling the energy of the charge at the end of every clock cycle. This technique helps to bring down the power dissipation well below CV^2 , the fundamental limit of conventional CMOS logic.

TL: Mike, your work has largely been in reducing energy loss due to adiabatic switching. Can you explain in more detail what adiabatic switching is, and some techniques for reducing energy loss due to adiabatic switching?

Michael Frank: Well, just to be clear—simply making gates reversible in terms of their logical function doesn't by itself necessarily save any energy at all. To actually save energy in practice, we also have to implement those gates in a physically reversible way—but this is not done at all

² https://en.wikipedia.org/wiki/IEEE_Rebooting_Computing

in traditional CMOS design. As Art was saying, the purpose of adiabatic switching is to recover and reuse the energy that is invested in charging the capacitances in a voltage-coded digital circuit, rather than dissipating that energy to heat; this allows us to approach the ideal of physical reversibility. In order for us to be able to recover almost all of the signal energy, all charging and discharging of circuit elements must be carried out almost adiabatically (from Greek *ἀδιάβατος*, “impassable”), which just means, in a way that holds back the energy in the circuit from becoming lost to the form of heat. To approach the ideal of adiabatic switching (even with ideal devices) requires gradual, closely-controlled transitions of all voltage levels in a circuit, and this requires that two switching rules must always be obeyed: The first rule, which is widely known, is that a switch (such as a field-effect transistor) must never be turned on when there is a significant voltage difference between its channel terminals. The second rule, which is less well known, is one that I discovered during my dissertation work, namely, that a switch must never be turned off when there is a significant electrical current that has been confined to flow through the channel of that switch. It turns out that in switch-based designs that always obey both of these rules, it’s completely impossible to erase voltage-coded digital information, and thus, such designs are necessarily logically reversible, which is what makes it possible, in principle, for Landauer’s principle to actually be circumvented within these circuits. However, in practice, these circuits’ energy efficiencies are limited by non-ideal characteristics of the transistors, such as, in particular, their tendency to leak charge (and dissipate energy) even when they are supposed to be turned off (non-conducting). To further improve the energy efficiency of adiabatic circuits in practice will thus require the design of new types of devices that are well optimized for low leakage and adiabatic operation, which mainstream leading-edge transistors are not. So, a new direction is sorely needed looking forward in terms of the low-level device technology.

I should also mention that there are a couple of additional classes of potential approaches to reversible computing besides the adiabatic one; I call these the ballistic and the chaotic modes of reversible operation. However, those two are less well developed than the adiabatic approach at present, so practical solutions involving either of those paradigms may be a little farther out. But they’re worth considering.

TL: How effective are these two: reversible logic and energy recovery design, and has anyone built actual reversible logic computers with efficient energy recovery?

AS: Researchers have demonstrated both. James et al. built an adder over a decade ago [1]. Thomsen reported on the design of a four-bit adder in 2012 [2], and Anantha Lakshmi and

Sudha designed a power efficient reversible floating point arithmetic unit for digital signal processing [3]. They appear to have demonstrated an order of magnitude improvement in energy consumption over conventional, Irreversible designs. Reviewing their reversible design for an IEEE 754 floating-point unit made me realize how really Byzantinely—complicated, involving a great deal of administrative detail-754 is. How any chip designer does 754 is a wonder! The latest work, “Foundations of Generalized Reversible Computing,” by Michael Frank puts it all together [4]. It is not all roses, however. Lukac et al. argue that interconnection wires still consume much more energy than savings obtained by reversibility [5]. More needs to be done to reduce interconnect heat dissipation.

MF: Actually, using adiabatic switching already scales down the dissipation of the energy associated with the capacitances in the wires, as well as the gates. In fact, there’s nothing about the concepts of reversible computing or adiabatic processes that’s inherently restricted to logic operations per se at all; the key point in reversible computing is that all transitions of digital states in the system must be carried out reversibly and adiabatically; this applies to transitions of the states of interconnects and memory elements, *i.e.*, to data movement and storage operations, as well as to transitions of the states of localized nodes encoding logic gate outputs, *i.e.*, logic operations. But, it’s not really all that difficult to maintain reversibility (and full adiabaticity) throughout all parts of a computer design if you know what you’re doing; even when we were students, my coworkers and I had already designed and prototyped complete CPU architectures to accomplish this in our DARPA-funded project at MIT back in the late 1990s.

TL: Cooler chips are certainly a big step forward, but is it enough to keep Moore’s law going? It seems to me Moore’s law does not completely account for performance enhancements over the past five decades, because it addresses the number of transistors one can put on a die, and not the architectural advancements that make good use of transistors. I know you have been working with sigmoid numbers— what John Gustafson [6] calls “posits”—to gain another order of magnitude improvement in performance. How do posits fit into your architectural ideas?

AS: Posit arithmetic: A new data type designed for direct drop-in replacement for IEEE Standard 754 floats. I am focused on ubiquitous energy efficient reversible posits; developing a semi, partial, [generalized reversible computing] GRC posit arithmetic unit (PAU) for ARM, x86, RISC-V, etc. drop-in replacement for IEEE 754. And for quantum architectures, sigmoid numbers are

a specialized subset of posit numbers for arithmetic. They are important to machine learning, not an IEEE 754 drop-in replacement.

TL: How do Gustafson’s posits fit into your architectural ideas?

AS: Posits are key to a beautiful arithmetic unit architecture, because they are fast and cool and have better closure under addition, multiplication and division than IEEE 754 floating point.³ Posits waste little or no resources on IEEE 754 not a number (NaN), overflow, and underflow, see Figure 2, and Table I. Posits eliminate underflow and overflow entirely, and essentially eliminate NaN, altogether. Plus, posits are faster. IEEE 754 is not only ugly and asymmetric, but a waste of silicon space, cycles time and energy heat.

There are two architecture design threads today, as of this interview: irreversible and reversible. Michael calls generalized reversible computing (GRC). I enjoy spending my 20W of wetware neural net cycles on getting as much energy efficiency as possible now with current irreversible computing posit, designing an irreversible classic computing posit AU is a step toward a reversible PAU. Its’ Verilog output may be used as input to a reversible logic design flow described in “[Design Automation and Design Space Exploration for Quantum Computers](#),” by Soeken, Roetteler, Wieb, and Micheli with EPFL and MSR [7]. Then hop, skip, jump, leap beyond Landauer’s limit with reversible computing, GRC, and posit, as soon as possible.

MF: Art is more well versed about posits than I am, so I don’t have much to add about them specifically. However, from what I have seen about them so far, they do appear to be an attractive alternative number representation. One of the projects I have been involved with at Sandia, led by Erik DeBenedictis and Jeanine Cook, has been investigating another approach to energy-efficient computer arithmetic called RRNS, for redundant residue number system, which was first investigated in the 1960s, but has made a bit of a comeback lately. The big selling point of RRNS is that the energy required to multiply n -bit numbers scales approximately linearly in n , rather than quadratically as in the usual binary number representations. Our student collaborators in Tom Conte’s group at Georgia Tech have made great strides in designing and simulating RRNS-based architectures [8]. So definitely, I agree there is still some useful progress to be made at the architectural level, even independently of reversible computing. However, I still like to remind people that, in the long run, solely architectural improvements

³ A set has closure under an operation if performance of that operation on members of the set always produces a member of the same set.

will inevitably run out of steam, once all of the “low-hanging fruit” has been picked, so to speak, and at that point, the only way to continue to make substantial progress will be to design reversible versions of our best architectures—regardless of whether they turn out to use IEEE 754, posits, RRNS, or something else. But, to engineer very high-quality reversible computers is actually extremely challenging, so, in my opinion, as a community, we really need to begin seriously cutting our teeth on reversible design principles sooner rather than later. While in parallel, still also pursuing work on nearer-term advances, of course.

AS: My architectural ideas design space is defined by energy efficient computing informed by Landauer’s Principle. The E in energy/space/time tradeoffs. So, in my architectural design space, informed by urgent IEEE Rebooting Computing now, posits are more energy efficient than IEEE 754 floating point, now.⁴ Even implemented irreversible. That’s a good energy efficient thing. I am encouraged to implement a GRC posit arithmetic unit by the reversible IEEE 754 work of Anantha Lakshmi and Gnanou Florence Sudha.

I am motivated by designing and implementing a clean, modern, post-irreversible era architecture energy efficient GRC PAU (generalized reversible computing posit arithmetic unit), focused on extreme energy efficient posits are all beautiful calming symmetry with no wasted energy or silicon, and it is correct, way more than 754 under closure! John Gustafson probably had his aha moment when he saw posits—wish I had been there.

TL: You seem to be pretty enthusiastic about posits.

AS: The irreversible computing silicon sand pile was perfected over 50 years. As Per Bak notes, it's over [9]. Time now—past time—for a new GRC computing silicon sand pile.

TL: What is the time frame for reversible computing to become mainstream? When can I buy a SoC (system on a chip) that is fully reversible?

MF: Ted, that’s a good question, and I wish I knew the answer, but unfortunately, it depends, to a large degree, on how soon people start investing significant resources into reversible computing related R&D. The problem is that to begin to really make reversible computing practical will likely require a fairly substantial overhaul of our technology base, from the micro-

⁴ <http://rebootingcomputing.ieee.org>

architectural and circuit levels all the way down to devices, materials, and fabrication processes, to really well-optimize the whole technology stack for the new reversible mode of operation. Over the decades, trillions of dollars have been invested in the development of the conventional irreversible technology base, and even though it's likely that a lot of the associated knowledge and tools can be reused to a great extent, I'd guess that at least some billions of dollars will still be needed to retool our technology foundations to most effectively support reversible operation—and that's assuming that the remaining research challenges can be solved. However, I like to remind people that Landauer's principle really does follow absolutely rigorously from fundamental physics, as I discuss in the GRC paper, so, if we want to keep the energy efficiency of computing from stalling out fairly soon, we really have no other choice than to begin pursuing reversible computing vigorously, under the optimistic assumption that it can indeed be made practical. We certainly can't succeed unless we try. Also, it's important to note that the potential upside from success here is almost infinite, in the sense that, as far as we know, there is no law of physics that prevents the amount of computation that we can accomplish using any given energy resources from continuing to increase without limit as technology is further refined—but this is only the case if reversible computing principles are used. It's simply much too big of an opportunity for us to continue to ignore it... I would argue that there really is no other logical choice, to advance the future of computing technology, than for us to seize this opportunity, and soon.

Figures and Tables

Figure 1. Clock frequency in megahertz increased exponentially in the 1990s through 2005, but then leveled off due to heat dissipation limitations.

Figure 2. Closure plots for addition, division, and multiplication. IEEE 754 floating-point performance is shown first followed by posits arithmetic performance.

(a). Addition closure: IEEE 754 on the top and posit closure on the bottom.

(b). Division closure: IEEE 754 on top and posit closure on the bottom.

(c). Multiplication closure: IEEE 754 on top and posit closure on the bottom.

Table I. Summary of IEEE 754 Floating Point Closure versus Posit Closure

	Exactness	NaN	Overflow	Underflow
IEEE 754	22.3%	10.7%	12.5%	3.3%
Posits	18.1%	.003%	-	-

References

- [1] James Rekha K., T. K. Shahana, K. Poulose Jacob, and Sreela Sasi. A New Look at Reversible Logic Implementation of Decimal Adder. *International Symposium on System-on-Chip*. Nov 20-21, 2007.
- [2] Thomsen, Michael Kirkedal. [Design of Reversible Logic Circuits using Standard Cells: Standard Cells and Functional Programming](#). Technical Report no. 2012-03ISSN: 0107-8283. University of Copenhagen. 2012.
- [3] Anantha Lakshmi A. V. and Gnanou Florence Sudha. A novel power efficient 0.64-GFlops fused 32-bit reversible floating point arithmetic unit architecture for digital signal processing applications. *Microprocessors and Microsystems*. Elsevier. 2017; <https://doi.org/10.1016/j.micpro.2017.01.002>
- [4] Frank, Michael P. [Foundations of Generalized Reversible Computing](#). In *International Conference on Reversible Computation*. Springer, Cham, 2017, 19-34.
- [5] Lukac, Martin, G. W. Dueck, M. Kameyama, and A. Pathak (2017). [Building a Completely Reversible Computer](#).
- [6] Gustafson, John. [Stanford Seminar: Beyond Floating Point: Next Generation Computer Arithmetic](#). YouTube. 2017.
- [7] Soeken, Roetteler, Wieb, and Micheli. [Design Automation and Design Space Exploration for Quantum Computers](#).
- [8] Deng, B., Srikanth, S., Hein, E. R., Rabbat, P. G., Conte, T. M., DeBenedictis, E., and Cook, J. Computationally-redundant energy-efficient processing for y'all (CREEPY). In *Rebooting Computing (ICRC), IEEE International Conference on* (pp. 1-8). IEEE, 2016.
- [9] Lewis, Ted G. *Bak's Sand Pile: Strategies for a Catastrophic World*, Agile Press (2011).

Further Reading

Michael P. Frank. [Reversible Computing: A Cross-Disciplinary Introduction](#). Invited talk presented to the Beyond-Moore Computing Research Challenge meeting, Sandia National Laboratories, Albuquerque, NM, March 10th, 2014.

Biographies

Art Scott is the founder of EETALL: Energy Efficient The Answer to Landauer Limit. Actively “Rebooting Computing”, transferring energy-efficient computing technologies—posit arithmetic, semi/partial reversible computing—to leap the Landauer Limit wall. Art is many faceted; creates outside the box; serial entrepreneur-intrapreneur; change agent, 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 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.

Michael P. Frank is a senior-level member of the technical staff at Sandia National Laboratories, in the Non-conventional Computing Technologies Department within the Extreme-Scale Computing Group at the Center for Computing Research. He previously held faculty positions in the CISE Department at the University of Florida, and the ECE Department at the FAMU-FSU College of Engineering. He received his B.Sci. in symbolic systems from Stanford University in 1991, and his M.Sci. and Ph.D. degrees in EECS from MIT in 1994 and 1999 respectively. His primary research interests since 1995 have been in the areas of future computing technologies, the physical limits of computing, and reversible computing.

DOI: 10.1145/3140589