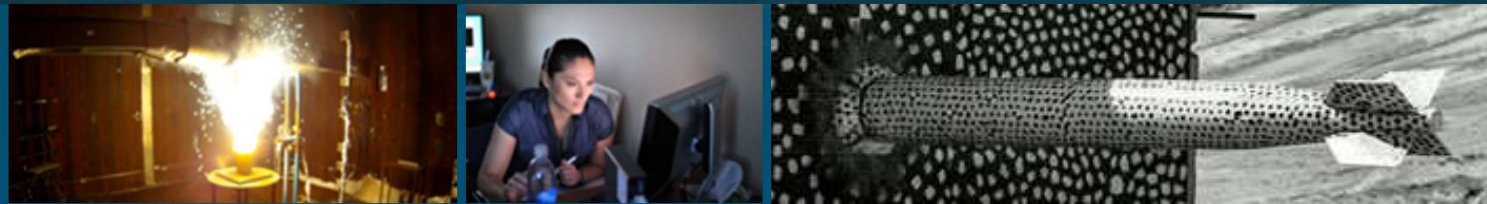


Novel Reversible Devices and Systems Implications



Presented at IEEE International Nanodevices and Computing Conference (INC), September 3, 2020

Michael P. Frank, Center for Computing Research

*with Karpur Shukla, Lab for Emerging Technologies, Brown University
and Rupert Lewis and Nancy Missert, Sandia National Laboratories*

Approved for public release, SAND2020-9249 C

Abstract text, for reference (*hide during talk*)



The principles of *reversible computing* offer a tantalizing prospect for an alternative path for continuing to improve the energy efficiency of *general digital computation* far beyond the physical limits that will soon cause the low-level energy efficiency of the conventional *non-reversible* computing paradigm to plateau. Microcircuit designs illustrating the physical and architectural principles of reversible computing already exist for both semiconducting and superconducting platforms. However, to maximize the practical applicability of the reversible approach, it is essential to explore ways to improve characteristics such as the *speed* (serial performance), and (even more broadly) cost-performance of the technology, simultaneously with its energy efficiency. This will likely require novel circuit, device, and even materials innovations. In this talk, we survey a range of new ideas for leveraging exotic quantum phenomena to help reduce energy dissipation *as a function of delay* in new classes of devices based on fundamentally novel physical mechanisms of operation, and discuss what the architectural and systems-level impacts of such technologies could be, if such concepts are eventually developed to the product level.

Topic: *Reversible computing* (RC) as a path for continuing to improve the energy efficiency of general digital computing far beyond the limits of the conventional (non-reversible) paradigm.

So far, there is still *no known* fundamental limit to the *ultimate* physical efficiency of this approach.

But, improvements to practical characteristics of known engineering implementations are still needed.

We can *already* build proof-of-concept reversible microcircuits in both semiconducting and superconducting technology platforms.

But, a key *practical* figure of merit is *computing performance per unit manufacturing cost* (“cost-performance”).

Cost-performance of any computing technology is significantly impacted by *speed* (here referring to serial performance).

Can characterize as *quickness* $q = 1/t_d$, where t_d is *time delay* (latency or initiation interval) for primitive digital device operations.

Improving these characteristics for RC while *simultaneously* improving energy efficiency will likely require novel circuit, device, and materials innovations, and exploitation of exotic physical phenomena.

The focus of this particular talk is to review:

Ideas for leveraging exotic quantum phenomena to reduce *dissipation as a function of delay*, $E_{\text{diss}}(t_d)$.

Accomplishing this could significantly increase the practical competitiveness of the reversible approach.

Architectural and systems-level impacts of improved device- and circuit-level reversible technologies.

Outline of Talk



Underlying economic motivation

Limits of conventional computation

Reversible computing theory

Existing implementation technologies

Semiconductor (CMOS) based implementations

Superconductor (JJ) based implementations

Future technology concepts

Limits of reversible computing?

Architectural and system-level impacts

Conclusion

Motivation from Economics / Systems Engineering



In general, *efficiency* η of any process can be defined as the amount P of some valued *product produced* by the process, divided by the amount C of *cost consumed* (in terms of resources, or dollars) by the process.

$$\eta = \frac{P}{C}$$

- For a computing system,
 - P can be amount of useful *information processing performed* (e.g., number of operations) by the system over its operating lifetime, and
 - C can be expressed the sum of manufacturing (& deployment) costs, plus operating costs over the system lifetime.
- We can also annualize the costs, in terms of, e.g. time-amortized manufacturing cost.
 - More sophisticated variations that account for net present value of future returns, depreciation curves, etc., not considered here.
- Operating costs largely amount to *energy-proportioned costs*: $C_{\text{oper}} = c_{\text{en}} \cdot E_{\text{oper}}$
 - c_{en} = operating cost per unit of energy dissipated; E_{oper} = energy dissipated during a given period of operation.

$$C = C_{\text{tot}} = C_{\text{mfg}} + C_{\text{oper}} \quad (\text{may be amortized})$$

We can thus reduce the efficiency formula $\eta = P/C_{\text{tot}}$ for computing to the form at right:

- E_{op} = Energy dissipated in one primitive device operation (or by one primitive device in time t_d).
- $c_{\text{dev},t}$ = Amortized manufacturing cost per primitive device per unit time t .

$$\begin{aligned} \eta &= \frac{1}{c_{\text{en}} \cdot E_{\text{op}} + c_{\text{dev},t} \cdot t_d} \\ &= \frac{1}{E_{\text{op}} t_d \left(\frac{c_{\text{en}}}{t_d} + \frac{c_{\text{dev},t}}{E_{\text{op}}} \right)} \end{aligned}$$

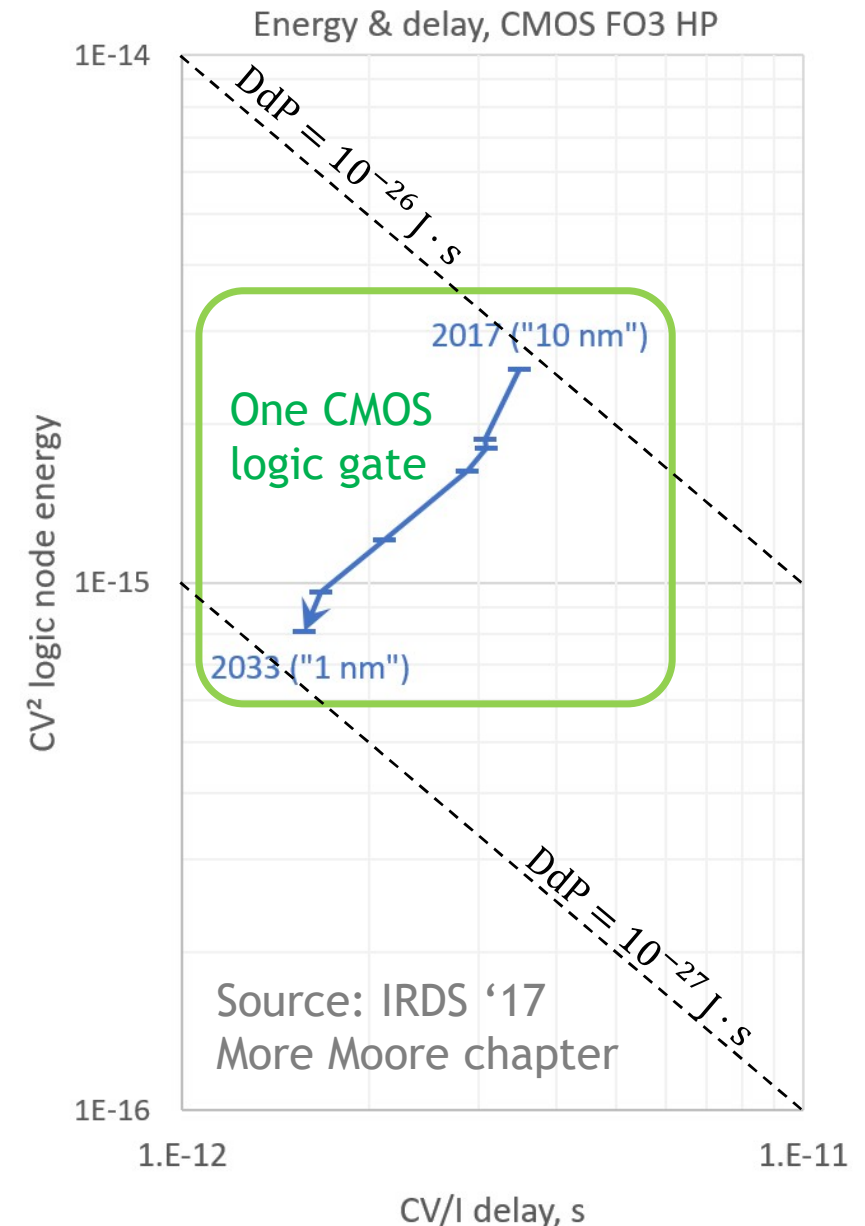
Some observations from this equation.:

- There are *diminishing* efficiency returns from decreasing either E_{op} or t_d in isolation
 - \therefore Continuing to push non-reversible technologies will ultimately reach a dead end!
- Note that if *both* E_{op} and t_d were decreased by $N\times$, efficiency would be increased by $N\times$. (All else being equal.)
- Decreasing $E_{\text{op}} \cdot t_d$ (dissipation-delay product, DdP) is *often* (but not always!) a win.
 - E.g., in scenarios where total lifetime cost of operation is very heavily energy-dominated, total cost can be reduced by lowering E_{op} , *even* in cases where $E_{\text{op}} t_d$ stays the same, or even increases somewhat!
- However, decreasing $E_{\text{op}}(t_d)$ (dissipation as a function of delay) at any given delay value t_d is *always a win*.
 - This will be our focus in future work.

6 Existing Dissipation-Delay Products (DdP) —Non-reversible Semiconductor Circuits

Conventional (non-reversible) CMOS Technology:

- Recent roadmaps (e.g., IRDS '17) show Dissipation-delay Product (DdP) decreasing by only $< \sim 10\times$ from now to the end of the roadmap (~ 2033).
 - Note the typical dissipation (per logic bit) at end-of-roadmap is projected to be $\sim 0.8 \text{ fJ} = 800 \text{ aJ} = \sim 5,000 \text{ eV}$.
- Optimistically, let's suppose that ways might be found to lower dissipation by an additional $10\times$ beyond even that point.
 - That still puts us at $80 \text{ aJ} = \sim 500 \text{ eV}$ per bit.
- We need at least $\sim 1 \text{ eV} \approx 40 kT$ electrostatic energy at a minimum-sized transistor gate to maintain reasonably low leakage despite thermal noise,
 - And, typical *structural* overhead factors *compounding* this within fast random logic circuits are roughly $500\times$,
 - so, $\sim 500 \text{ eV}$ is *indeed* probably about the practical limit.
 - At least, this is a reasonable order-of-magnitude estimate.



7 Existing Dissipation-Delay Products (DdP)— Adiabatic Reversible Superconducting Circuits

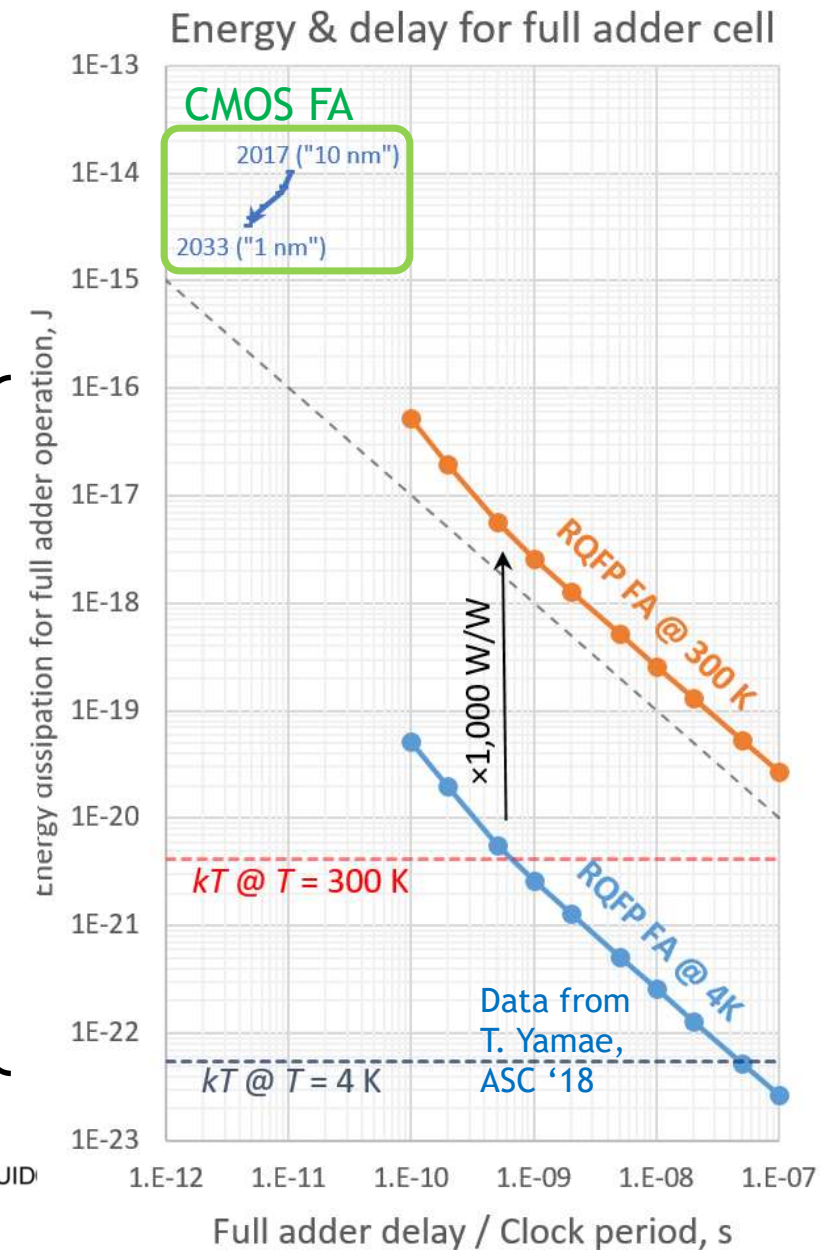
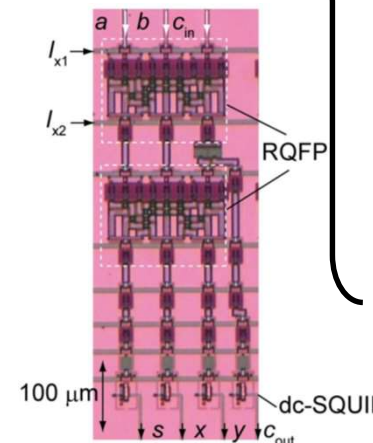
Reversible adiabatic superconductor logic:

- State-of-the-art is the **RQFP** (Reversible Quantum Flux Parametron) technology from Yokohama National University in Japan.
- Chips were fabricated, function validated.
- Circuit simulations predict DdP is $>1,000\times$ *lower* than even *end-of-roadmap* CMOS.
- Dissipation extends *far below* the 300K Landauer limit (and even below the Landauer limit at 4K).
- DdP is *still* better even after adjusting by a conservative factor for large-scale cooling overhead ($1,000\times$).

Question: Could some *other* reversible technology do even better than this?

- We have a project at Sandia exploring one possible superductor-based avenue for this...
- But, what are the *fundamental* (technology-independent) limits, if any?

RQFP =
Reversible
Quantum Flux
Parametron
(Yokohama U.)



Limits of Non-Reversible Computation



Practical limits on dissipation for end-of-roadmap *non-reversible* CMOS ultimately rest on the kT energy scale.

- This gives them a connection to the fundamental Landauer limit ($kT \ln 2$ energy dissipation per bit lost).
- In effect, many bits' worth of *physical* information are being lost (becoming entropy) for each logical bit that's erased in CMOS.

The Landauer limit, itself, is just a trivial consequence of basic statistical physics and information theory.

- It's completely impossible to *violate* the limit within the laws of physics.

∴ Avoiding thermodynamic limits on energy efficiency requires that we avoid losing information.

- This is why we *must* migrate to the reversible computing paradigm to bypass the present efficiency roadblocks.

Fundamental
theorem
of the
thermodynamics
of computing:

Computing System (\mathcal{G}),
total entropy $S(\Phi) = -\sum p \log p$

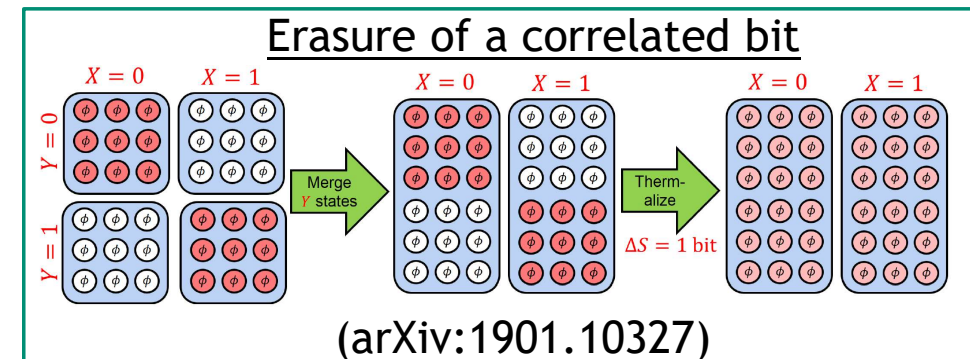
Non-Computational
Subsystem ($\mathcal{N}\mathcal{C}$)

non-comp. / cond. entropy
 $S_{nc} = S(\Phi|C) = S(\Phi) - H(C)$

Computational
Subsystem (\mathcal{C})

info. entropy $H(C) = -\sum P \log P$

Landauer's Principle



Basic Reversible Computing Theory

Fundamental theorem of traditional reversible computing:

- A deterministic computational operation is (unconditionally) non-entropy-ejecting if and only if it is *unconditionally* logically reversible (injective over its entire domain).

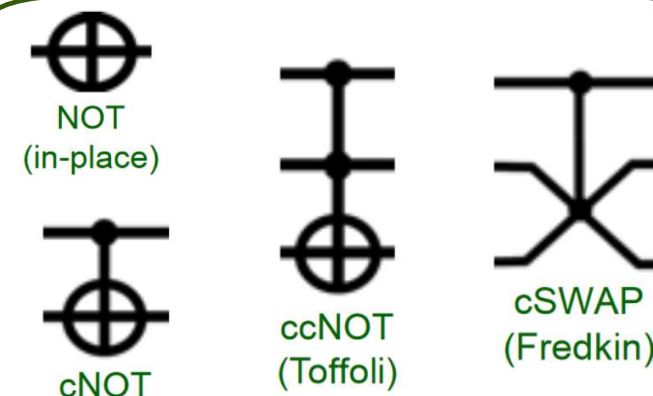
Fundamental theorem of generalized reversible computing:

- A *specific* (contextualized) deterministic computation is (specifically) non-entropy-ejecting if and only if it is *specifically* logically reversible (injective over the set of *nonzero-probability* initial states).
- Also, for any deterministic computational operation, which is conditionally reversible under some assumed precondition, then the entropy required to be ejected by that operation approaches 0 as the probability that the precondition is satisfied approaches 1.

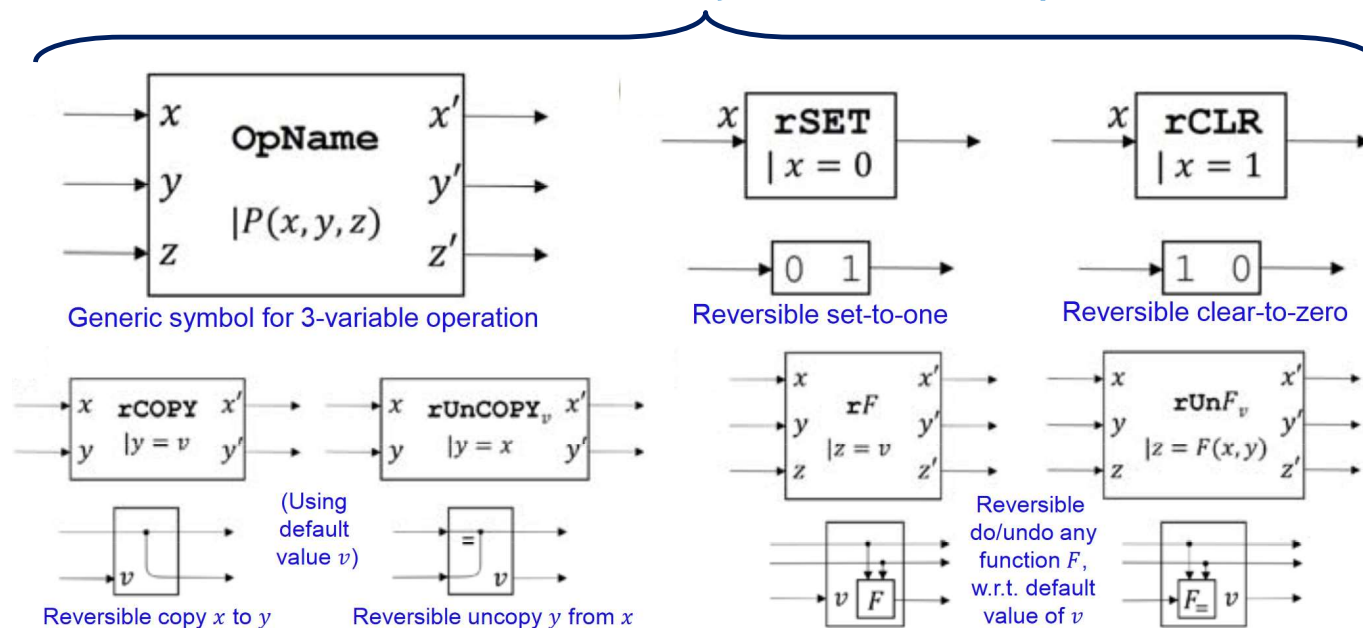
Bottom line: To avoid requiring Landauer cost, it is *sufficient just to have reversibility when some specified preconditions are satisfied*.

- Basis for practical engineering implementations.

Traditional Unconditionally Reversible “Gates” (Operations)

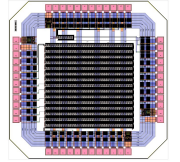


Generalized *Conditionally* Reversible Operations



Perfectly Adiabatic Reversible Computing in CMOS

2LAL test chip
taped out at
Sandia, Aug. '20



To approach reversible computing in CMOS...

We must aggressively eliminate *all* sources of non-adiabatic dissipation, including:

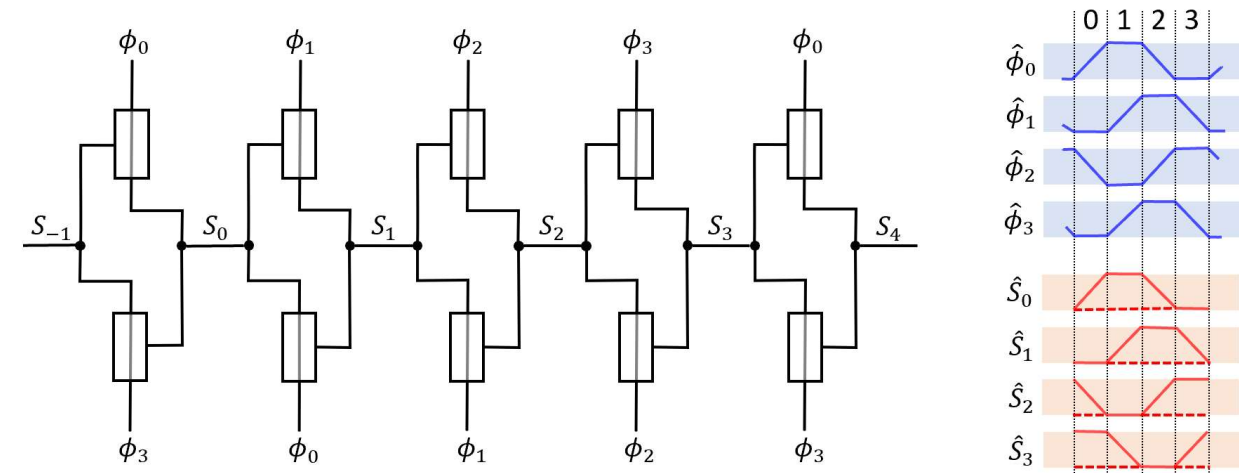
- Diodes in charging path, “sparking,” “squelching,”
 - Eliminated by “**truly, fully adiabatic**” design. (E.g., CRL, 2LAL).
 - Suffices to get to a few aJ (10s of eV) in 180 nm *before* voltage optimization.
- Voltage level mismatches that dynamically arise on floating nodes before reconnection.
 - Eliminated by static, “**perfectly adiabatic**” design. (E.g., S2LAL).

We must also aggressively minimize standby power dissipation from leakage, including:

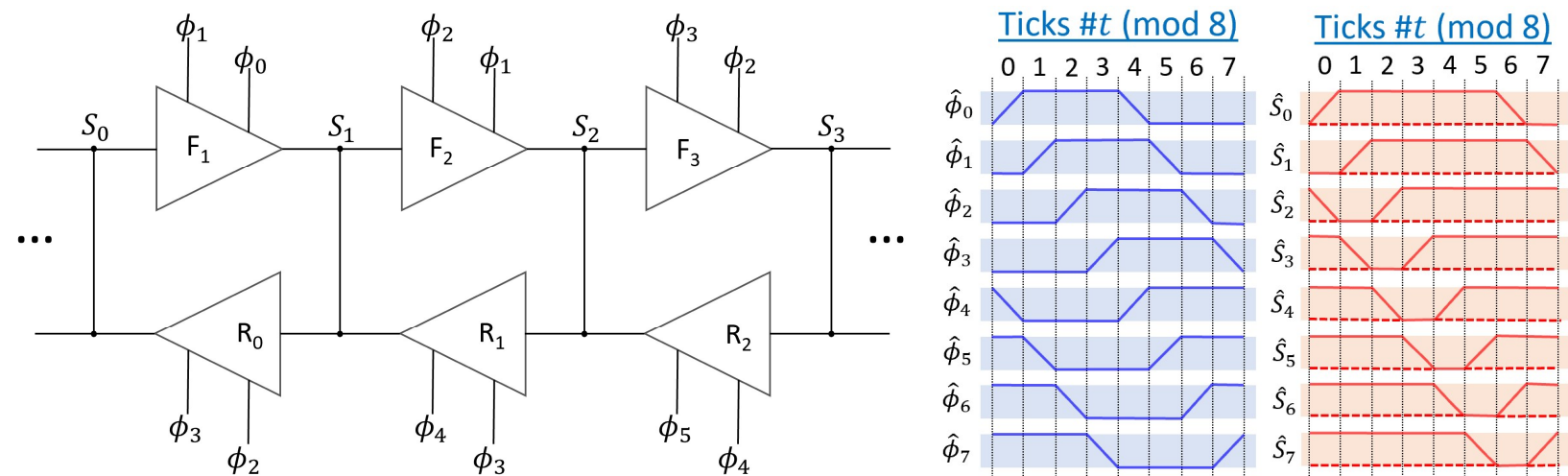
- Subthreshold channel currents
 - Low- T operation helps with this
- Gate oxide tunneling
 - Thicker gate oxides

Note: (Conditional) logical reversibility *follows from* perfect adiabaticity.

Shift Register Structure and Timing in 2LAL

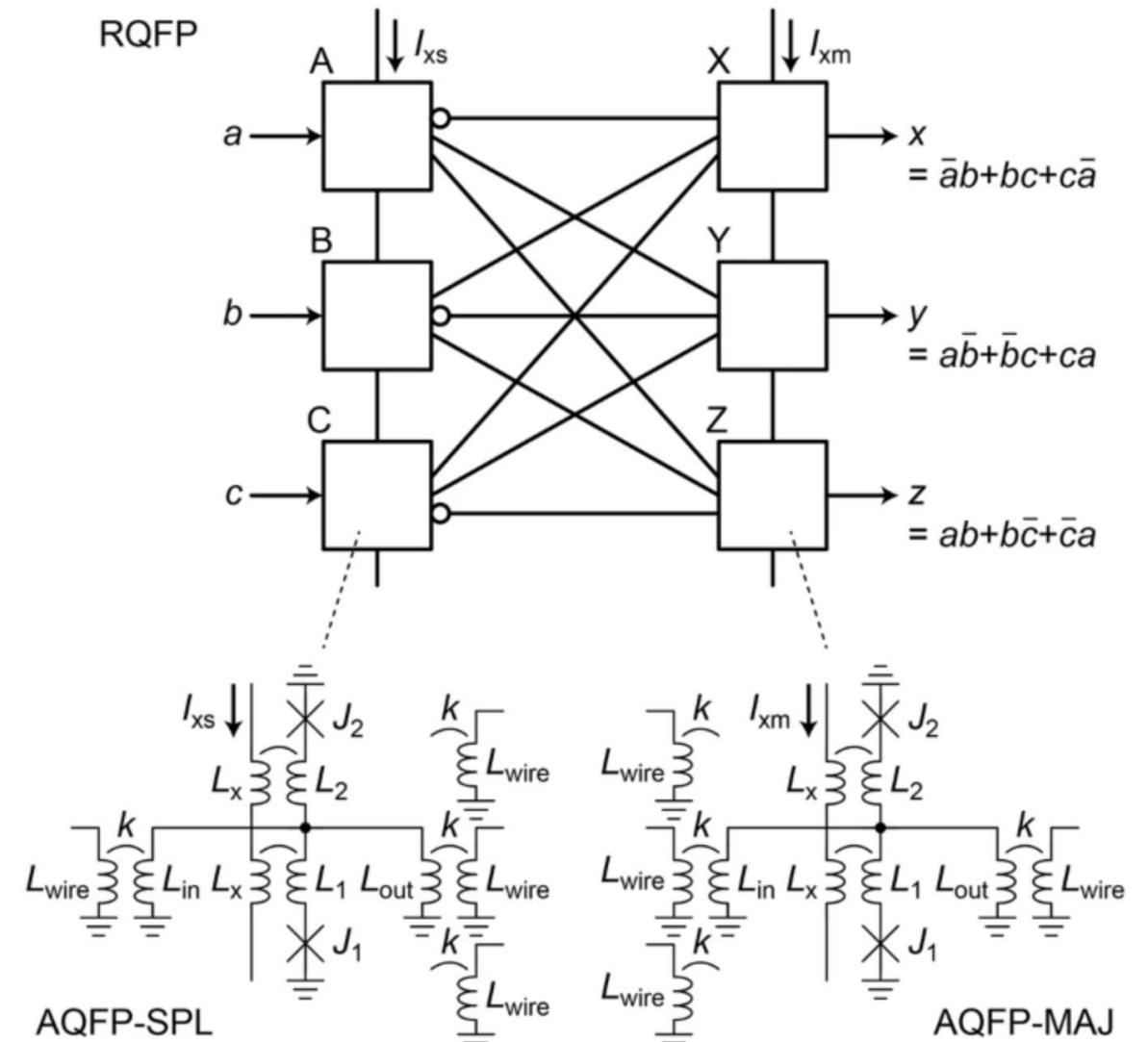
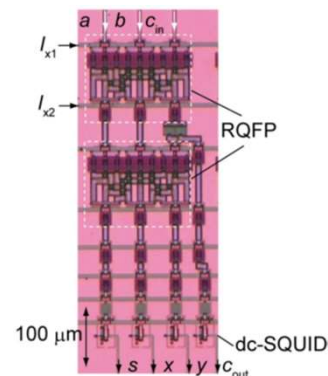


Shift Register Structure and Timing in S2LAL





Simulations indicate that RQFP circuits can dissipate $< kT \ln 2$ even at $T = 4\text{K}$, at speeds on the order of 10 MHz.

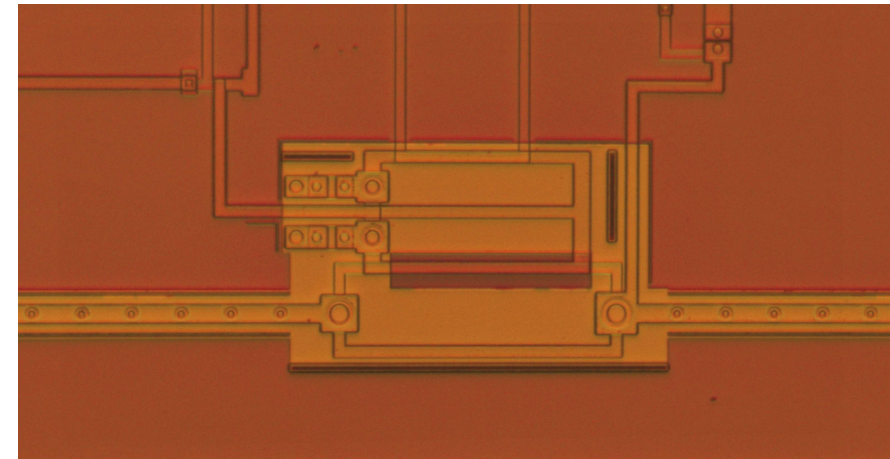


Future Technology Concepts for Reversible Computing



Ballistic Asynchronous Reversible Computing in Superconductors (BARCS) 🤖

- Based on a novel *Asynchronous Ballistic Reversible Computing* (ABRC) model of reversible computation.
- Utilizes ballistic propagation of flux solitons (*fluxons*) in long Josephson junctions.
 - Elastically interacting with stationary SFQ states in circuit elements, *e.g.* →
 - Asynchronous operation avoids chaotic instabilities.
- Current externally-funded project at Sandia exploring this approach.



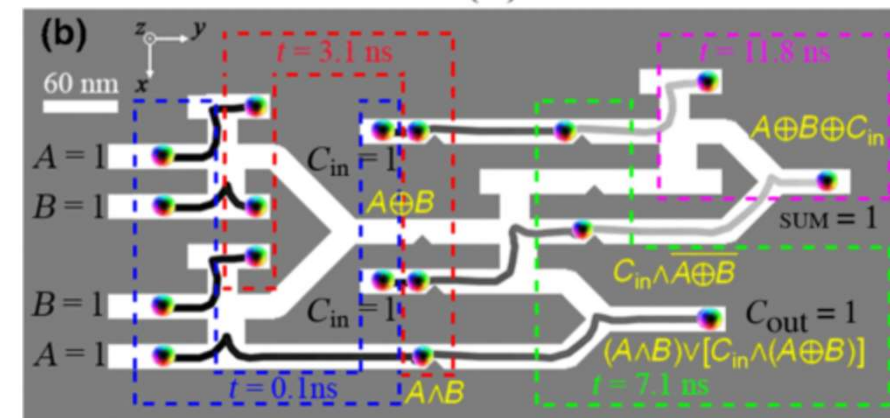
Reversible Memory (RM) Cell

Reversible Computing with Magnetic Skyrmions

- Joseph Friedman (U. Texas) & collaborators

Many other reversible device concepts have been, or could be explored:

- Nanomechanical rod logics (*e.g.*, Merkle *et al.*)
- Reversible computing using exotic topological quantum states
- Reversible computing using dynamic quantum Zeno stabilizer effects



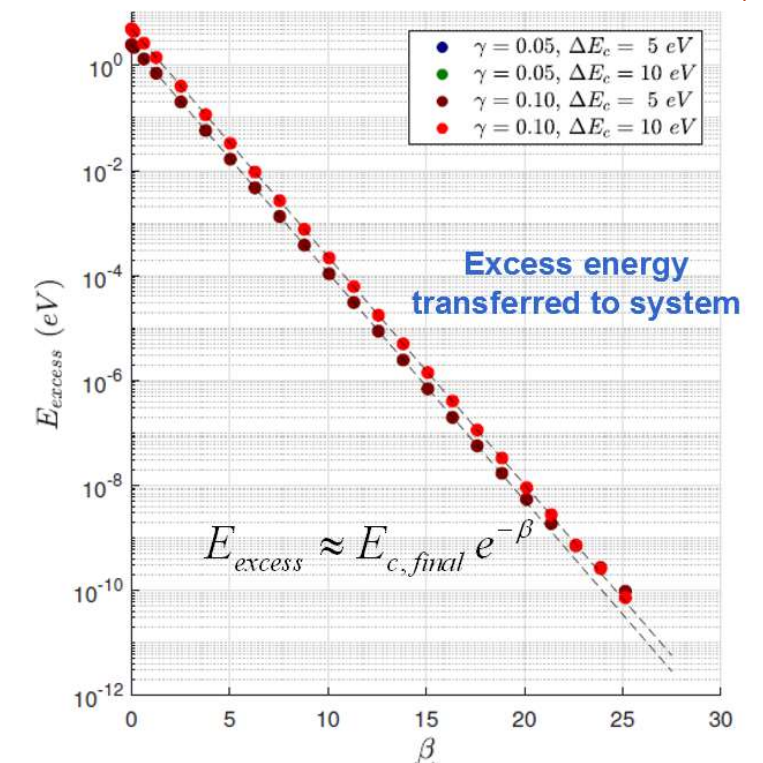
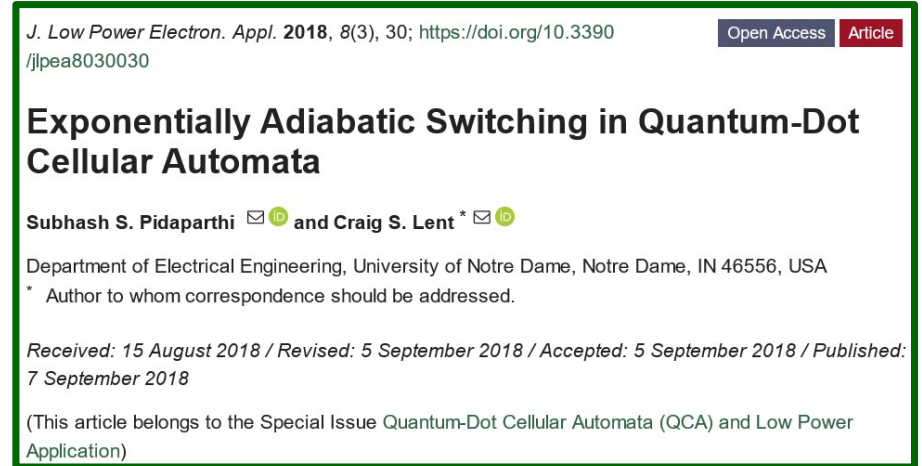
Magnetic Skyrmion Full Adder Concept

Can dissipation scale better than linearly with speed?

Some observations from Pidaparthi & Lent (2018) suggest Yes!

- Landau-Zener (1932) formula for quantum transitions in e.g. scattering processes with a missed level crossing...
 - Probability of exciting the high-energy state (which then decays dissipatively) scales down *exponentially* as a function of speed...
 - This scaling is commonly seen in many quantum systems!
- Thus, dissipation-delay *product* may have *no lower bound* for quantum adiabatic transitions—if this kind of scaling can actually be realized in practice.
 - I.e.*, in the context of a complete engineered system.
- Question:** Will unmodeled details (e.g., in the driving system) fundamentally prevent this, or not?

$$P_D = e^{-2\pi\Gamma}$$

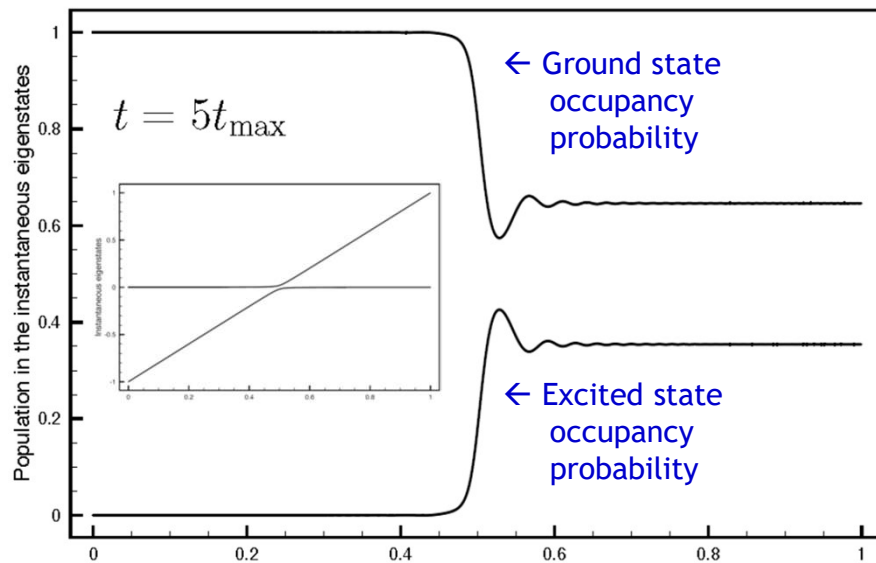


Shortcuts to Adiabaticity (STA)



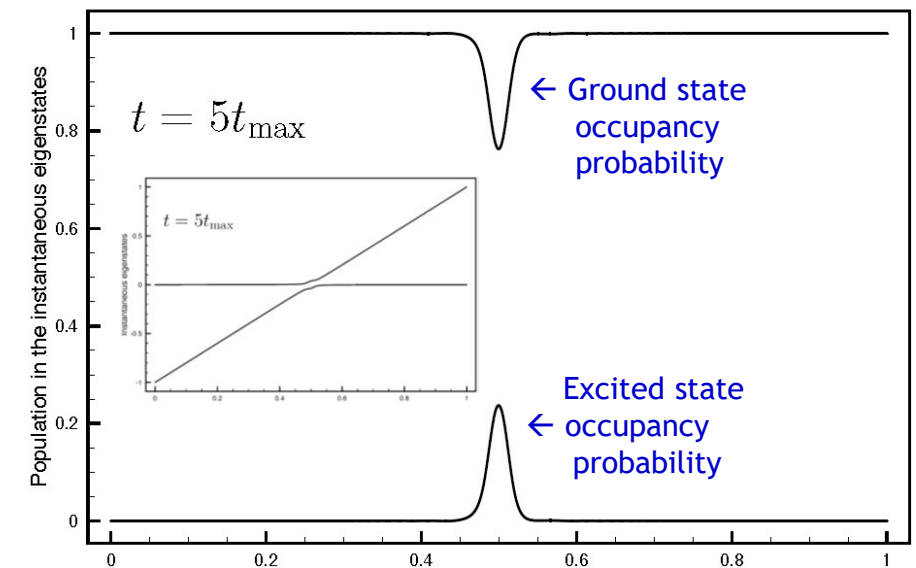
A line of theoretical physics research showing that, *in principle*, quantum state transformations can always be carried out with exactly zero dissipation, even at any given *finite* delay!

- Requires the introduction of a finely-tuned “counterdiabatic” perturbation to the system’s time-dependent Hamiltonian.
- **Again, we ask:** Is this idealized prediction *actually achievable*, if fundamental thermodynamic limits that apply to the complete system are accounted for?



Normal quantum adiabatic process:
Substantial excitation/dissipation

Figure credit:
Collaborator
David Guéry-Odelin
(Université de Toulouse)



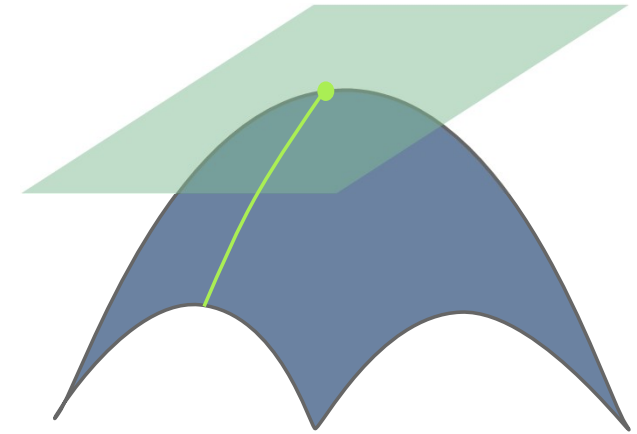
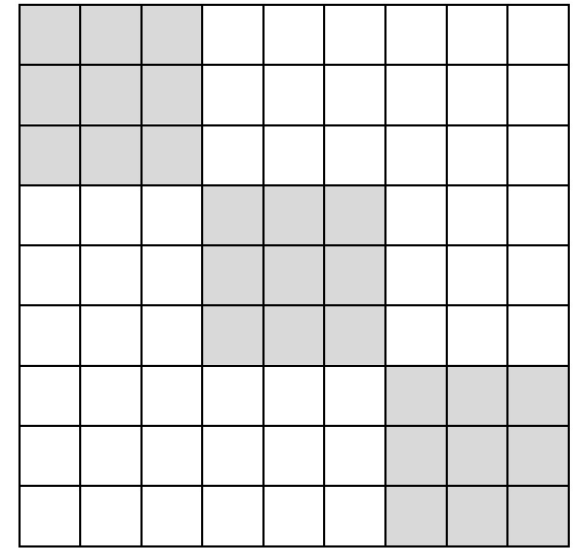
Using counterdiabatic protocol:
Zero net excitation/dissipation

Limits to Reversible Computing?

—An approach from the theory of Open Quantum Systems

(Work with Karpur Shukla, Brown University, and Victor V. Albert, CalTech)

- Computational states modelled as *decoherence-free subspace blocks* (DFSB) of overall Hilbert space.
- Quantum Markov equation with multiple asymptotic states: admits subspace dynamics (including DFSB structures) for open systems under Markov evolution.
 - Induces geometric tensor for *manifold of asymptotic states*.
 - Similar to quantum geometric tensor / Berry curvature for closed systems.
- Current work: use multiple asymptotic state framework to derive thermodynamic quantities...
 - Uncertainty relations, dissipation and dissipation-delay product.



Architectural and system-level impacts of new reversible tech.

Collaboration with Tom Conte & Anirudh Jain, Georgia Tech



It is true that reversible computing introduces cost-efficiency overheads at a number of levels:

- Impacts of dissipation-delay scaling (*e.g.* reducing clock speed to gain energy efficiency).
- Constant-factor overheads at the gate level (*e.g.*, for dual-rail or quad-rail operation).
- Architectural/algorithmic overheads imposed by the requirement for substantial logical reversibility.

However! In general, we can say:

- As long as the (time-amortized) cost per-device $C_{\text{dev},t}$ continues going down, *larger and larger overheads* from reversible operation can continue to be absorbed, while *still increasing* overall system-level efficiency η .
 - In contrast, the non-reversible approach has *no prospect* to continue scaling system efficiency, if cost is measured in energy units.

Thus: Over a long-enough future time horizon, the reversible approach *has to become a huge win*.

- Eventually becoming *far* more cost-efficient (when including energy costs over the system lifetime) than *any non-reversible general computing technology can possibly exist within the laws of physics*.

Insofar as the overall economy continues to increasingly rest on digital information processing,

- Developing reversible computing enables *increasing the total future economic value of civilization* by possibly *many orders of magnitude*, compared to *not* developing it (given any available energy resources).
 - New initiatives at the level of a Manhattan Project or a moonshot would therefore be a very wise investment.

Conclusion



For the efficiency of general digital computing to advance *far* beyond the limits of CMOS will *require* the development of advanced reversible computing technologies.

- This follows from irrefutable facts of fundamental physics.

The fundamental limits of dissipation-delay scaling in reversible technologies are *just beginning* to be investigated. Almost no serious work has been done on this yet!

- New reversible technologies with greatly improved scaling characteristics may yet be discovered.

Although reversible computing does impose various overheads on hardware complexity, these *do not prevent it from nevertheless improving overall system cost-efficiency*, when considering the system's total lifetime cost of operation, which *includes* energy-related costs.

- Within the reversible paradigm, achievable overall system cost efficiency can continue improving for as long as the lifetime-amortized per-device manufacturing cost keeps decreasing, with no clear end in sight.

Reversible computing R&D is an investment in the future value of civilization, with an almost unlimited potential to yield highly positive future returns for the legacy of humankind.