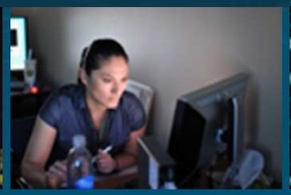
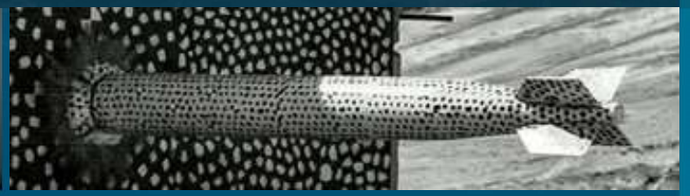
Laboratory for Physical Sciences (LPS)
Annual Colloquium

Reversible Computing: The Only Future for General Digital Computing





Wednesday, March 31st, 2021

Michael P. Frank, Center for Computing Research

with collaborators: Robert Brocato, Rupert Lewis, Nancy Missert & Brian Tierney (Sandia), Kevin Osborn & Liuqi Yu (LPS), Erik DeBenedictis (Zettaflops, LLC), Karpur Shukla (Brown), Rudro Biswas & Dewan Woods (Purdue), Tom Conte & Anirudh Jain (Georgia Tech).

Approved for public release, SAND2021-3907 PE

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

Contributors to the larger effort:

- Full group at Sandia:
  - Michael Frank (Cognitive & Emerging Computing)
  - Robert Brocato (RF MicroSystems)
  - David Henry (MESA Hetero-Integration)
  - Rupert Lewis (Quantum Phenomena)
  - Nancy Missert (Nanoscale Sciences)
    - Matt Wolak (now at Northrop-Grumman)
  - Brian Tierney (Rad Hard CMOS Technology

Thanks are due to Sandia's LDRD program, DOE's ASC program, and the DoD/ARO ACI (Advanced Computing Initiative) for their support of this line of research!

- Thanks are also due to the following colleagues & external collaborators:
  - Erik DeBenedictis
  - Kevin Osborn (LPS/JQI)
    - Liuqi Yu
  - Steve Kaplan
  - Rudro Biswas (Purdue)
    - Dewan Woods
  - Karpur Shukla (CMU/Brown U.)
    - w. Prof. Jingming "Jimmy" Xu
    - Also w. Victor Albert (CalTech)
  - David Guéry-Odelin (Toulouse U.)
  - FAMU-FSU College of Engineering:
    - Sastry Pamidi (ECE Chair)
    - Jerris Hooker (Instructor)
    - 2020-21 students:
      - Fadi Matloob, Frank Allen, Oscar L. Corces, James Hardy
    - 2021-22 students:
      - Marshal Nachreiner, Samuel Perlman, Donovan Sharp, Jesus Sosa
  - *Others may be forthcoming…*

Abstract (hide during talk)

The conventional, *non-reversible* approach to general digital computing is approaching hard limits due to fundamental thermodynamic constraints. Physics guarantees that every possible path forward that does not fundamentally change the computing paradigm will suffer from diminishing returns. However, we can still continue to improve the efficiency of general digital computing far beyond the limits of the conventional approach by transitioning to the *reversible computing* paradigm. This is clearly distinct from quantum computing, because its focus is on raw energy efficiency for *general* computing, not quantum coherence for special-purpose algorithms. The basic engineering principles of reversible computing have already been demonstrated in both semiconducting and superconducting technology platforms. In contrast to *all other approaches* to general digital computing, reversible computing offers a long-term path forward towards ever-greater levels of computing efficiency, with no clear limits in sight.

Abstract (bullet-point version)

All conventional (*i.e.*, *non-reversible*) digital computing is approaching a ***hard thermodynamic limit*** on its energy efficiency (and therefore also cost-efficiency, assuming only a cost floor for energy).

- Industry is already struggling to improve performance metrics; semiconductor roadmap ends in ~10 yrs.
- Digital computing (in the conventional paradigm) is faced with ***permanent technological stagnation***.

BUT! ***There is a solution*** (but only one!) that may allow us to sustain continuing improvements in energy & cost efficiency for digital technology far into the future: **Reversible Computing (RC).**

- Refers to computing in a way that *preserves signal energies* and reuses them over multiple digital operations.
  - This is not a trivial change! It requires re-design and re-optimization of devices, circuits, architectures at multiple levels.
  - However, demonstrations of RC already exist for both semiconducting and superconducting technology platforms.
- RC is ***distinct*** from quantum computing, although it may *also* leverage quantum phenomena & principles.
  - Focus of RC is on achieving far greater energy efficiency and practical performance for **ALL** digital computing, rather than quantum speedups on relatively few specialized applications.
  - Arguably, RC's eventual practical & economic impact can therefore be **much broader and greater** than that of QC.
    - By the end of the century, reversible computing could dominate the rest of the computing market by *many orders of magnitude*.
- Thus, reversible computing needs to be **front and center** in all high-level discussions about the long-term future of computing technology (or *at least* on a par with the more "fashionable" topics of neuro/quantum).
  - **Far** more attention should be being paid to it, as a viable technology development path.
  - A large-scale initiative is needed to help push this technology forward.

Outline of Talk

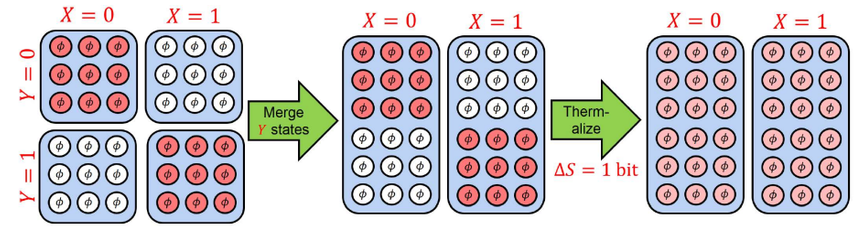
Reversible Computing: The Only Future for General Digital Computing

- I. Introduction: Motivation & History.
  - Landauer's Principle and Early Developments
  - The Fundamental Economics of Computing Cost Efficiency
  - The Dissipation-Delay Efficiency Metric & Trends
- II. Reversible Computing with Adiabatic CMOS.
  - Basic Principles of Adiabatic Switching.
  - Fully Adiabatic CMOS with 2LAL.
  - Fully Static, Fully Adiabatic CMOS with S2LAL.
- III. Reversible Superconducting Technologies.
  - Adiabatic Reversible Quantum Flux Parametron logic.
  - Ballistic Asynchronous Reversible Computing in Superconductors.
- IV. Fundamental Physical Limits of Reversible Computing.
  - Exponential Adiabaticity and Asymptotic Scaling.
- V. Future Work and Conclusion.

Section I. Motivation & History

Reversible Computing as The Sustainable Path Forward
for General Digital Computing

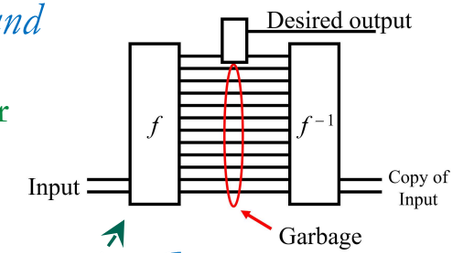
Motivation & Brief History



arXiv: 1901.10327

Landauer's Principle (1961):

- Elementary statistical physics and information theory together imply that there is a *fundamental upper bound* on energy efficiency for the conventional (*non-reversible*) computing paradigm.
  - *Oblivious* erasure of known/correlated information implies dissipation of $E_{\text{diss}} \geq k_{\text{B}} T \ln 2$ energy to the environment for each bit's worth of known information that is lost.
    - k_{B} is Boltzmann's constant $\simeq 1.38 \times 10^{-23}$ J/K = the natural logarithmic unit of entropy.
    - **NOTE:** T is ***the temperature of the thermal environment into which the waste heat ends up getting ejected.***
      - ∴ Simply lowering T *locally* cannot help *directly* to lower *system-level* E_{diss} if the *external* environment temperature is fixed.

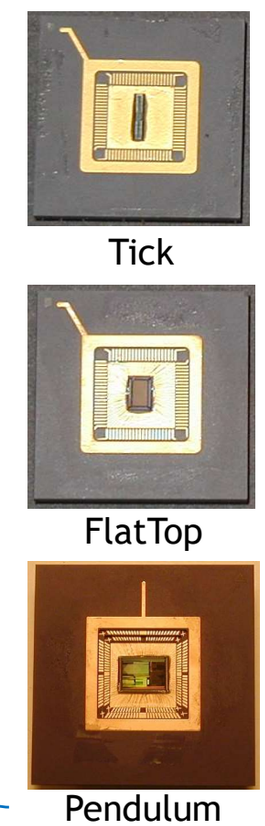


Reversible Computing (RC) provides a (theoretical, and eventually also practical!) solution:

- RC means computing *without* oblivious erasure of known or correlated information.
  - In principle, energy dissipation per useful operation can be made *arbitrarily small* (can approach zero as technology improves).
    - ∴ Energy *efficiency* (operations per Joule) can theoretically approach *infinity* (or at least, no limits to this are yet known).
  - This includes implications for avoiding differential power analysis (DPA) and related side-channel attacks.

Some early history of the reversible computing field:

- RC was first shown *theoretically* coherent by Bennett, 1973 (doi:10.1147/rd.176.0525).
- First *engineering* implementation proposed by Likharev, 1977 (doi:10.1109/TMAG.1977.1059351).
- First fully-adiabatic sequential CMOS logic style: Younis & Knight, 1993 (Proc. Int'l Symp. Res. Int. Sys.).
- First fabricated reversible processor chips! Frank, Knight, Love, Margolus, Rixner, Vieri (1996-1999).



The time is ripe for a resurgence!

- I believe there is an opportunity right now to demonstrate some real breakthroughs.

Semiconductor Roadmap is Ending…

Thermal noise on gate electrodes of minimum-width segments of FET gates leads to significant channel PES fluctuations if $E_g \lesssim$ 1-2 eV!

- This increases leakage, impairs practical device performance
  - Thus, roadmap has minimum gate energy asymptoting to ~2 eV

Further, real logic circuits incur many *compounding* overhead factors *multiplying* this raw transistor-level limit:

- Transistor width 10-20× minimum width for fastest logic.
- Parasitic (junction, *etc.*) transistor capacitances (~2×).
- Multiple (~2) transistors fed by each input to a given logic gate.
- Fan-out of each gate to a few (~3) downstream logic gates.
- Parasitic wire capacitance (~2×).

Due to all these overhead factors, the energy of each logic bit in real logic circuits is necessarily many times larger than the minimum-width gate energy!

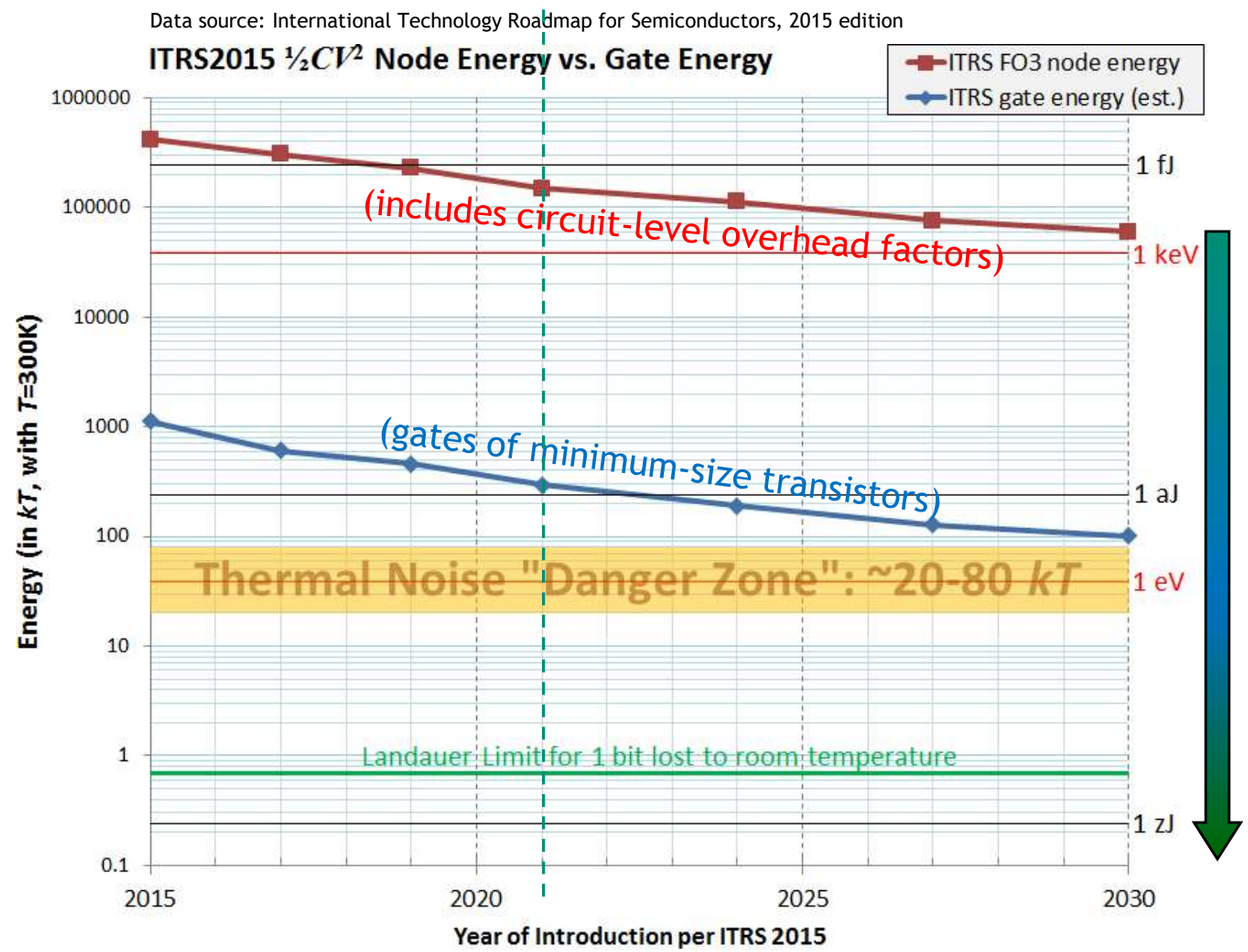
- 375-600× (!) larger in ITRS'15.
  - ∴ Practical bit energy for irreversible CMOS logic asymptotes to **~1 keV**!

Practical, real-world logic circuit designs can't just magically cross this ~500× architectural gap!

- ∴ Thermodynamic limits imply much *larger* practical limits!
  - The end is near!

This is Now!

Only about a decade left…



Only **reversible computing can take us from ~1 keV at the end of the CMOS roadmap, all the way down to ≪ *kT*.**

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} = total energy dissipated during a given period of operation.

$$C = C_{\text{tot}} = C_{\text{mfg}} + C_{\text{oper}}$$

(may be time-amortized)

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

- E_{op} = Energy dissipated due to *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.

$$\eta = \frac{1}{c_{\text{en}} \cdot E_{\text{op}} + c_{\text{dev},t} \cdot t_{\text{d}}} = \frac{1}{E_{\text{op}} t_{\text{d}} \left(\frac{c_{\text{en}}}{t_{\text{d}}} + \frac{c_{\text{dev},t}}{E_{\text{op}}} \right)}$$

Some observations from this equation.:

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

Why Reversible Computing Wins Despite Its Overheads!

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

$$C_{\text{tot}} = C_{\text{mfg}} + C_{\text{oper}}$$

Bumper-sticker slogan: "***Running Faster by Running Slower!***" (Wait, what?) More precisely:

- Reversible technology is so energy-efficient that we can overcome its overheads (including longer transition times!) by using much greater parallelism to increase overall performance within system power constraints.
  - This is borne out by a detailed economic/systems-engineering analysis.

Bottom line: The computational *performance per unit budgetary cost* on parallelizable computing workloads can become as large as desired, given only that *both terms* in this expression for total *cost per operation* C_{op} can be made sufficiently small:

$$C_{\text{op}} = c_E \cdot E_{\text{diss,op}} + c_M\left(s_{\text{elem}} \cdot t_{\text{delay}}\right).$$

where:

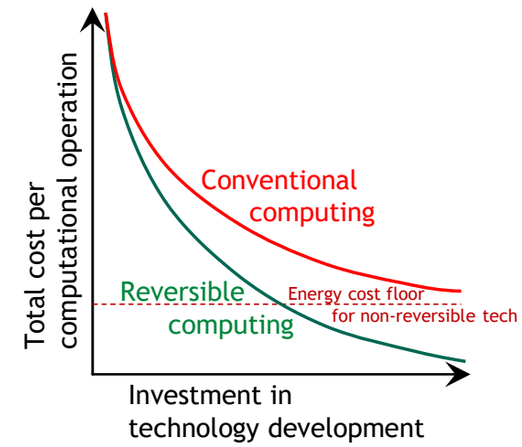
- c_E is the operating cost C_{oper} attributable to supplying power/cooling, divided by energy delivered.
- $E_{\text{diss,op}}$ is the system energy dissipation, divided by number of operations performed.
- c_M is the total cost C_{mfg} for system manufacturing & installation, *divided by* the number n_{elem} and physical size s_{elem} (in appropriate units) of individual computing elements, & the system's total useful lifetime t_{life}.
- t_{delay} is the average time delay between instances of re-use of each individual computing element.

Two key observations:

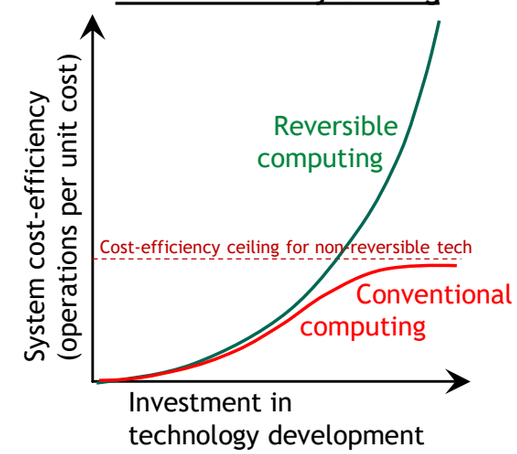
- The cost per operation of *all* conventional computing *approaches a hard floor* due to Landauer.
  - Assuming *only* that the economic cost of operation *per Joule delivered* cannot become arbitrarily small.
- But, there is *no* clear barrier to making the manufacturing cost coefficient c_M *ever smaller* as manufacturing processes are refined (and/or the deployed lifetime of the system increases).

∴ *Nothing prevents* system-level cost efficiency of reversible machines from becoming *arbitrarily* larger than conventional ones, *even* if we have to scale t_{delay} and/or s_{elem} up as we scale $E_{\text{diss,op}}$ down!

Amortized Cost Scaling



Cost-Efficiency Scaling

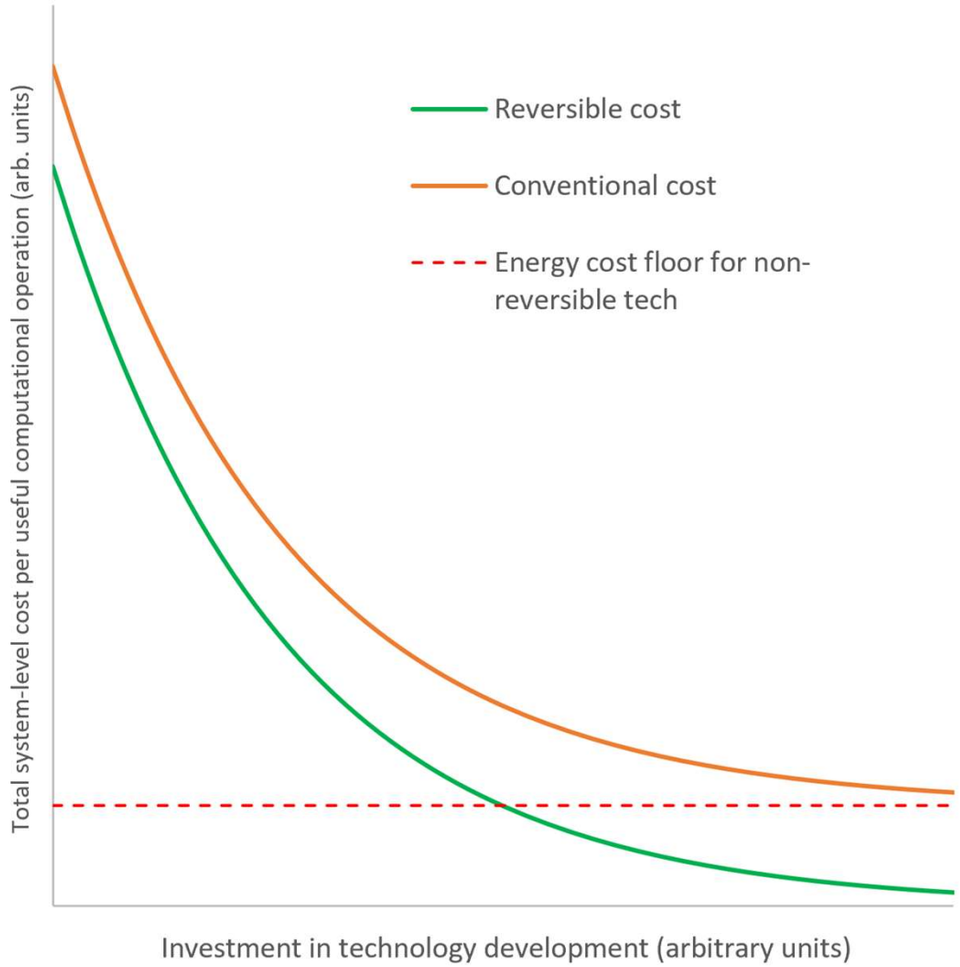


Economic Analysis at a Glance

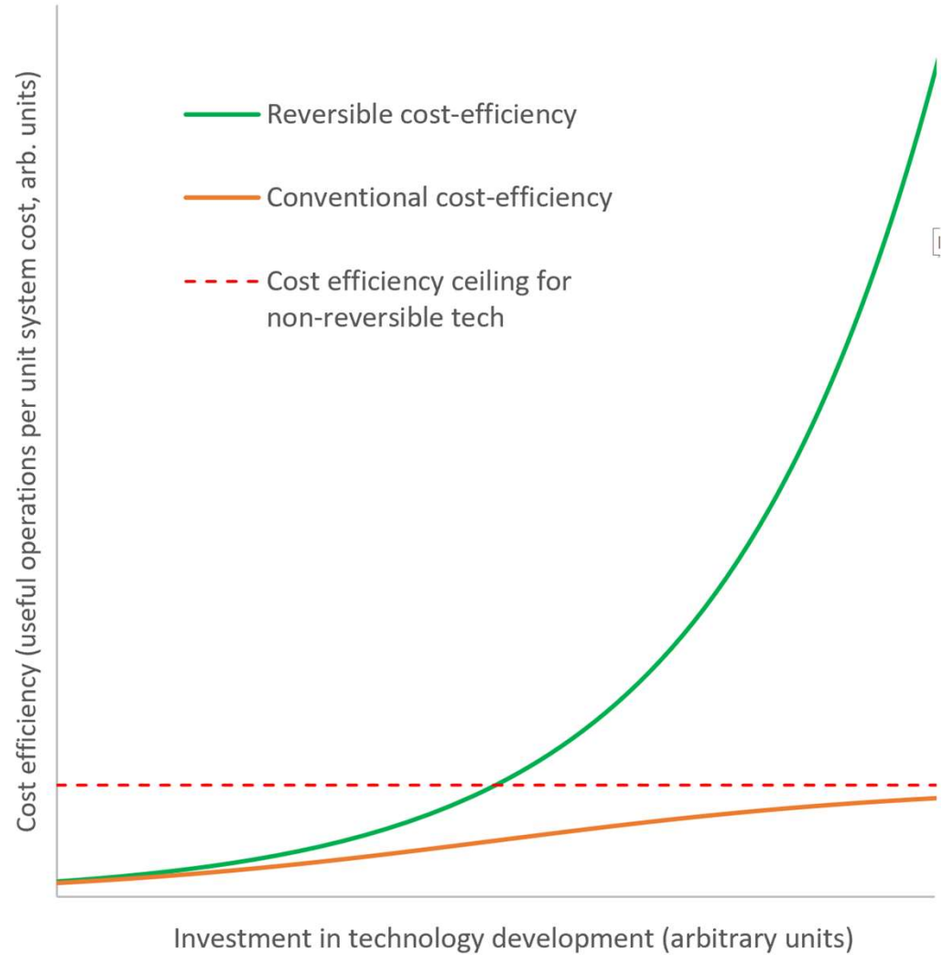
Same charts generated in Excel, using exponential decline in above-floor costs with investment.

- However, *any* rates of approach to 0 above-floor cost still lead to indefinitely-large long-term efficiency advantages for RC.

Amortized Cost Scaling



Cost-Efficiency Scaling



What is dissipation-delay efficiency, and why is it important?

Typically, the *total cost* $\$_{\text{tot}} = \$_E + \$_M$ to perform a computation is minimized when energy-related costs $\$_E$ and manufacturing-related costs $\$_M$ are roughly on the same order.

- Because, there are *diminishing returns* from individually reducing *either one* of these two cost components far below the other one.
  - And, doing so actually makes the total *larger*, if the other cost component gets *increased* as a result.

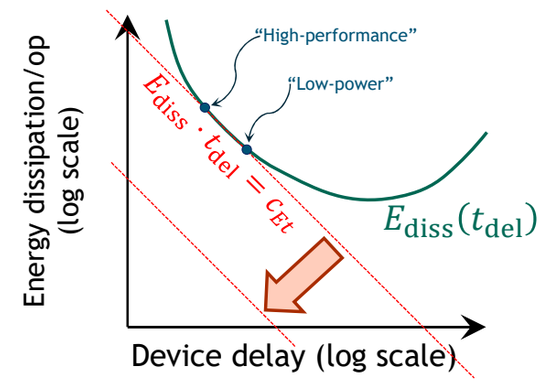
Can express total cost in terms of device parameters: $\$_{\text{tot}} = k_E E_{\text{diss}} + k_M t_{\text{del}}$

For *any* technology that permits tradeoffs between energy efficiency and serial performance, there will be *some* region of the energy-delay curve where the tangent line (on a log-log chart) has slope -1.

- In this region, the energy-delay *product* is roughly constant.
  - This is even true for voltage scaling in standard irreversible CMOS.
    - But, fully adiabatic techniques can extend this scaling region over a much wider range.
- Different operating points in this linear scaling region will be suitable for applications with different cost *coefficients* k_E, k_M that apply to energy vs. manufacturing cost.
  - *E.g.*, in spacecraft, the effective cost of energy vs. hardware is much greater than in grid-tied applications.

NOTE: If you can move to a new technology whose energy-delay frontier (curve) touches a min. energy-delay product line that is $N\times$ lower than before,

- Then it follows that *total cost* for some applications is reduced by at least $\sqrt{N}\times$!



Dissipation-delay product:

$$c_{Et} = E_{\text{diss}} \cdot t_{\text{del}}$$

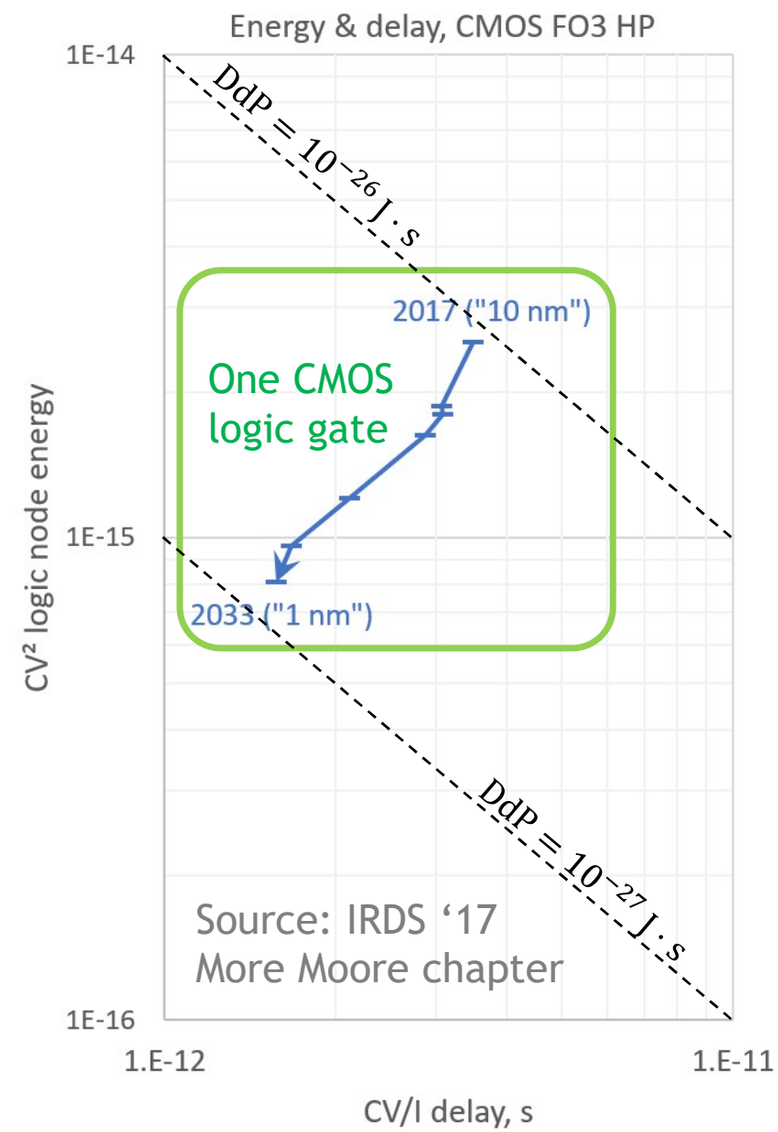
Dissipation-delay efficiency:

$$\eta_{Et} = \frac{1}{c_{Et}}$$

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 <~10× from now to the end of the roadmap (~2033).
  - Note the typical dissipation (per logic bit) at end-of-roadmap is projected to be ~0.8 fJ = 800 aJ = ~5,000 eV.
- Optimistically, let's suppose that ways might be found to lower dissipation by an additional 10× beyond even that point.
  - That still puts us at 80 aJ = ~500 eV per bit.
- We need at least ~1 eV ≈ 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×,
  - so, ~500 eV is *indeed* probably about the practical limit.
    - At least, this is a reasonable order-of-magnitude estimate.



Section II. Reversible Computing with Adiabatic CMOS

Reversible Computing as The Sustainable Path Forward for General Digital Computing

Adiabatic Circuits in CMOS: A Brief History

A selection of some early papers:

Fredkin and Toffoli, 1978 (DOI:10.1007/978-1-4471-0129-1_2)

- Unfinished circuit concept based on idealized capacitors and inductors
  - How to control switches to do logic was left unspecified
  - Large design overhead—Roughly one inductor per gate

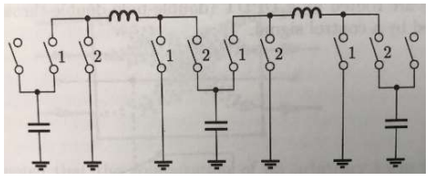


Figure reproduced with permission

Seitz *et al.*, 1985 (CaltechCSTR:1985.5177-tr-85)

- Realistic MOSFET switches; more compact integration (off-chip *L*)
- Not yet known to be general-purpose; required careful tuning

Koller and Athas, 1992 (DOI:10.1109/PHYCMP.1992.615554)

- Not yet fully-reversible technique; limited efficiency
- Combinational only; conjectured reversible *sequential* logic impossible

Hall, 1992; Merkle, 1992 (DOIs:10.1109/PHYCMP.1992.615549; 10.1109/PHYCMP.1992.615546)

- General-purpose reversible methods, but for combinational logic only

Younis & Knight, 1993 (http://dl.acm.org/citation.cfm?id=163468)

- First fully-reversible, fully-adiabatic *sequential* circuit technique (CRL)

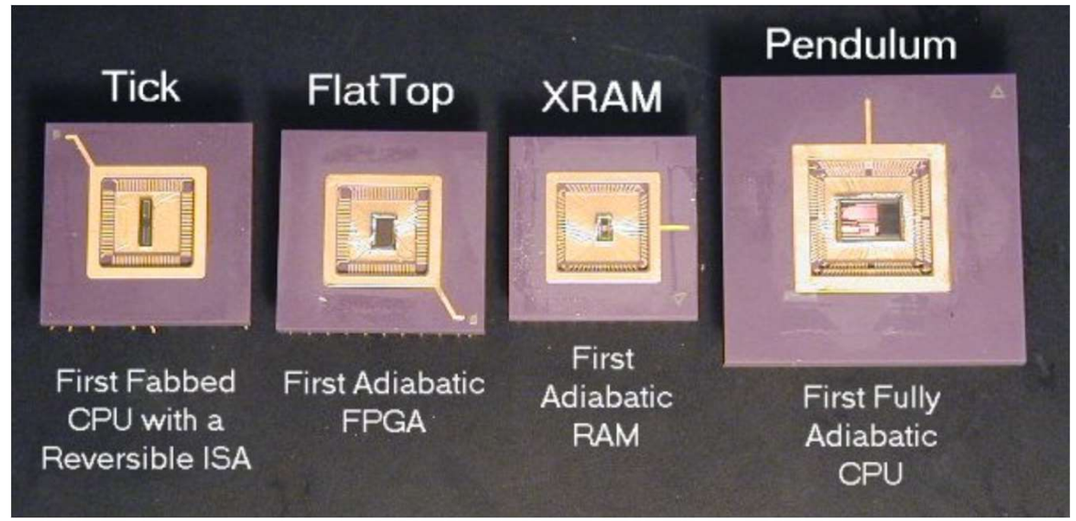
Adiabatic Circuits in CMOS: History, cont.

Younis & Knight, 1994

- Simplified 3-level adiabatic CMOS design family (SCRL)
  - However, the original version of SCRL contained a small non-adiabaticity bug which I discovered in 1997
    - This problem is easily fixed, however

Subsequent work at MIT, 1995-99

- Myself and fellow students
- Various chips designed using SCRL →
- Reversible processor architectures



Substantial literature throughout the late 90s / early 2000s…

- Too many different papers / groups to list them all here!
  - Most of the proposed schemes were not truly/fully adiabatic, though

Researchers recently active in adiabatic circuits include:

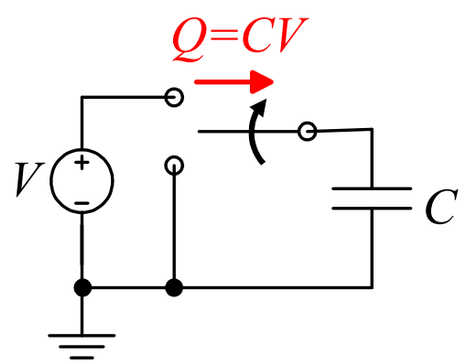
- A couple I know in the US:
  - Greg Snider (Notre Dame)
  - Himanshu Thapliyal (U. Kentucky)
- Also some groups in Europe, India, China, Japan…
- My group at Sandia (new work reported on slide #18)

Conventional *vs.* Adiabatic Charging

For charging a capacitive load C through a voltage swing V

Conventional charging:

- Constant *voltage* source

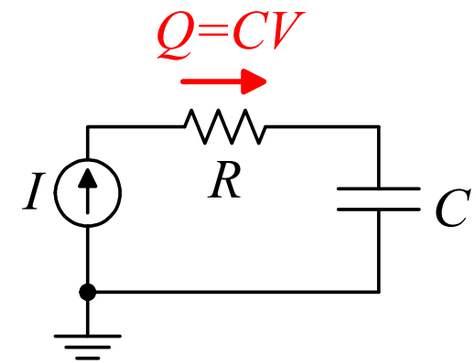


- Energy dissipated:

$$E_{\text{diss}}^{\text{conv}} = \frac{1}{2}CV^2$$

Ideal *adiabatic* charging:

- Constant *current* source



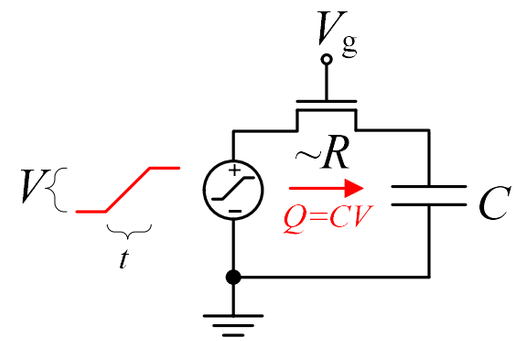
- Energy dissipated:

$$E_{\text{diss}}^{\text{adia}} = I^2Rt = \frac{Q^2R}{t} = CV^2\frac{RC}{t}$$

Note: Adiabatic charging beats the energy efficiency of conventional by advantage factor:

$$A = \frac{E_{\text{diss}}^{\text{conv}}}{E_{\text{diss}}^{\text{adia}}} = \frac{1}{2}\frac{t}{RC}$$

Adiabatic Charging via MOSFETs



A simple voltage ramp can *approximate* an ideal constant-current source.

- Note that the load gets charged up *conditionally*, if the MOSFET is turned on (gate voltage $V_g \gtrsim V + V_t$) during ramp.
  - V_t is the transistor's threshold, typically < ½ volt

Can discharge the load later using a similar ramp.

- Either through the same path, or a different path.

$$t \gg RC \Rightarrow E_{\text{diss}} \to CV^2 \frac{RC}{t}$$

$$t \ll RC \Rightarrow E_{\text{diss}} \to \tfrac{1}{2} CV^2$$

Exact formula for linear ramps:

$$E_{\text{diss}} = s\left[1 + s\left(e^{-1/s} - 1\right)\right]CV^2$$

given *speed fraction* $s = RC/t$.

The (ideal) operation of this circuit approaches *physical reversibility* ($E_{\text{diss}} \to 0$) in the limit $t \to \infty$, but *only* if a certain *precondition* on the initial state is met (namely, $V_g \gtrsim V_{\max} + V_t$)

- How does the possible physical reversibility of this circuit relate to its *computational* function, and to some *appropriate* concept of logical reversibility?
  - Traditional (Landauer/Fredkin/Toffoli) reversible computing theory does **not** adequately address this question, so, we need a more powerful theory!
  - The theory of ***Generalized Reversible Computing*** (GRC) meets this need.

See arxiv:1806.10183 for the full GRC model.

Basic Requirements for Fully Adiabatic Operation

No diodes in charging paths!

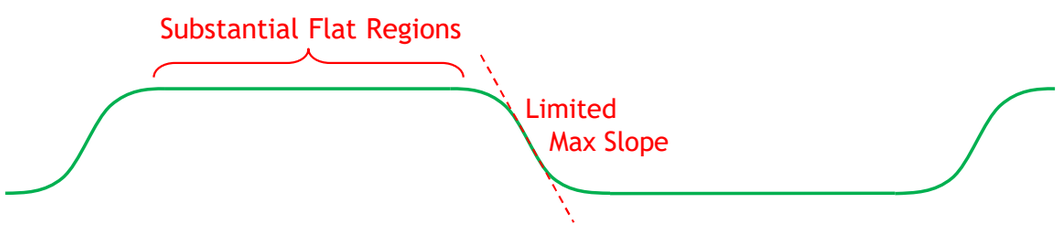
- All diodes have a built-in voltage drop for fundamental thermodynamic reasons.

Operate all switches (*e.g.*, FETs) with a "dry-switching" discipline:

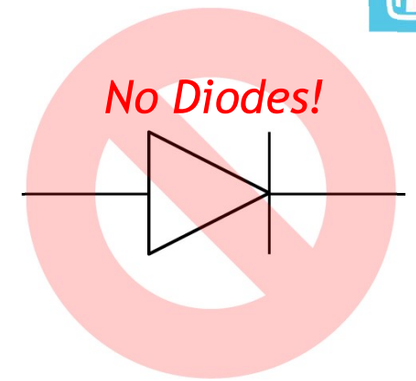
- Never turn on (close) a switch when there is a significant voltage difference $\Delta V \neq 0$ between its terminals.
  - Leads to a sudden, non-adiabatic flow of current.
  - More generally: No rapid voltage changes.
- Never turn off (open) a switch when there is a significant current flow $I \neq 0$ through the switch.
  - Leads to non-adiabatic losses as switch is (non-instantaneously) turning off.
    - Resistance through switch increases during turnoff → voltage drop increases → non-adiabatic loss across voltage drop.
  - Exception: If path has low inductance and there is an alternate path for the current.

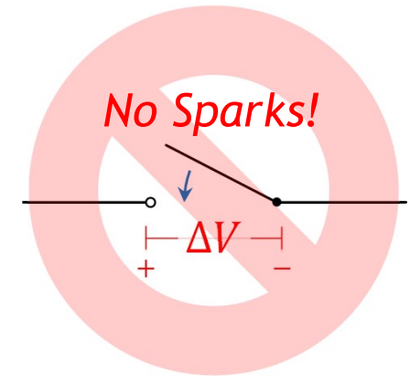
Use quasi-trapezoidal driving waveforms (no steep edges; flat tops and bottoms).

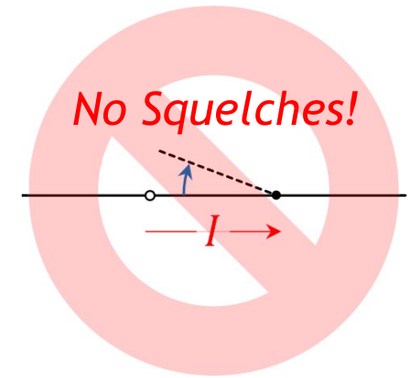
- This is necessary to obey the other rules.



$$E_{\mathrm{a}} = \xi_{\mathrm{tr}} C_{\mathrm{L}} V_{\mathrm{dd}}^2 \frac{R C_{\mathrm{L}}}{\tau_{\mathrm{tr}}}$$







Notations and Conventions Used *(slide 1 of 2)*

Two nominal voltage levels: 0 V (GND, "low") and $V_{\mathrm{dd}} \gtrsim 2|V_{\mathrm{t}}|$ ("high").

Divide time into equal, discrete intervals called *ticks*, each of duration $\bar{\tau}_{\mathrm{tr}}$, and numbered consecutively.

- Every *transition* between nominal levels is required to fit entirely within a tick,
  - so, the actual transition time τ_{tr} is upper-bounded by the tick length, $\tau_{\mathrm{tr}} \leq \bar{\tau}_{\mathrm{tr}}$.

The active energy dissipation from any given adiabatic transition is as follows:

$$E_{\mathrm{a}} = \xi_{\mathrm{tr}} C_{\mathrm{L}} V_{\mathrm{dd}}^2 \frac{R C_{\mathrm{L}}}{\tau_{\mathrm{tr}}},$$

where:

- ξ_{tr} is a constant *shape factor* that accounts for the departure of the ramp shape from the ideal;
- C_{L} is the capacitive load of the node that is transitioning;
- R is the effective resistance of the charging path.

The clock period τ_{p} is an integer number n of ticks, $\tau_{\mathrm{p}} = n\bar{\tau}_{\mathrm{tr}}$.

- Thus, the clock frequency is

$$f = (n\bar{\tau}_{\mathrm{tr}})^{-1}.$$

- Ticks within a cycle are numbered modulo n (*i.e.*, $0, \ldots, n-1$).

Notations and Conventions Used *(slide 2 of 2)*

In the logic styles we'll discuss, any given logic *symbol* L (e.g., 0 or 1) is represented by a complementary *signal pair*.

- Thus, for k-valued logic we require $2k$ signals.
- Normally we have just $k = 2$ symbols, $L \in \{0,1\}$.

Possible conditions for a given signal pair (when valid) are *active* or *inactive*.

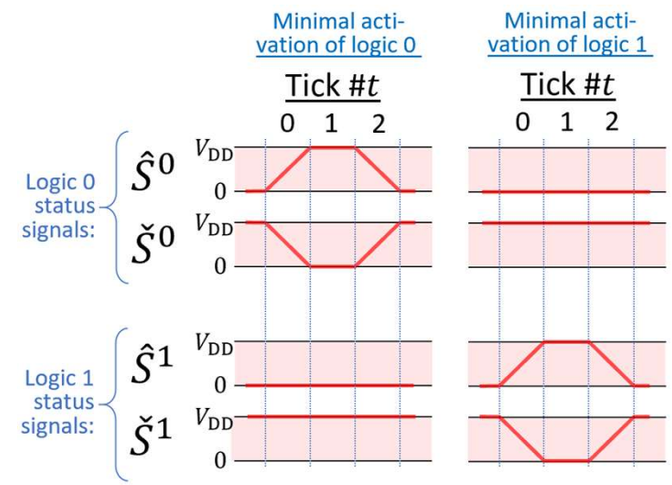
- One of the signals in each pair is *active-high*; the other is *active-low*.
  - When in the active state, we say the pair is *actively representing* the corresponding logic symbol L.
- The signal pair may feed the control terminals of a CMOS transmission gate.
  - The active-high signal controls the nFET, and the active-low signal controls the pFET.
  - Thus, the transmission gate is turned ON (conducting) when the signal pair is active.
  - The body terminals of the FETs should be separately biased (not tied to either channel terminal).
    - Can be used to increase device thresholds if desired.

The following notation is used for a signal pair:

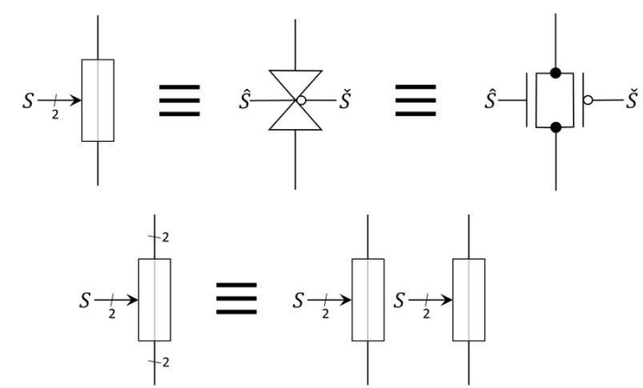
$$S^{L}_{t_\mathrm{b},t_\mathrm{e}} = (\hat{S}^{L}_{t_\mathrm{b},t_\mathrm{e}}, \check{S}^{L}_{t_\mathrm{b},t_\mathrm{e}})$$

where:

- $\hat{\ }$, $\check{\ }$ accents denote active-high and active-low signals, respectively.
  - No accent denotes the pair.
- L (if present) denotes the logic symbol the signal pair is representing.
- $t_\mathrm{b}, t_\mathrm{e}$ (if present) denote the transitional (*begin* and *end*) ticks of the active period.



Examples of minimal activations



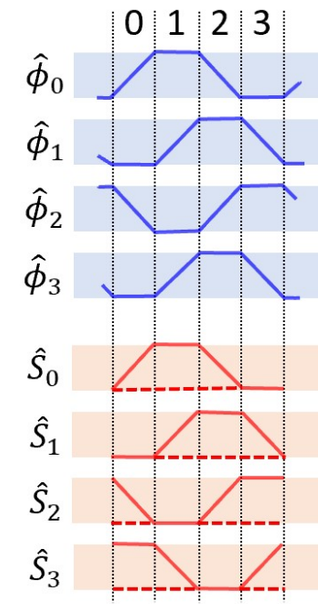
Transmission gate symbols

Review of 2LAL

See Frank *et al.* "Exploring the Ultimate Limits of Adiabatic CMOS", 38th IEEE Int'l Conf. on Computer Design (ICCD'20), 10.1109/ICCD50377.2020.00018

2LAL is a simple variant of CRL (Younis & Knight '93), which was rediscovered and described by M. Frank in lectures at the University of Florida in the year 2000.

- Four clock phases, each active for one tick and inactive for one tick.
- A simple (one-symbol) shift register structure is shown.
- Series/parallel combinations of transmission gates can be used to do logic (not shown here).
  - 2LAL really only differs from CRL in terms of allowing more flexibility in how internal nodes are handled

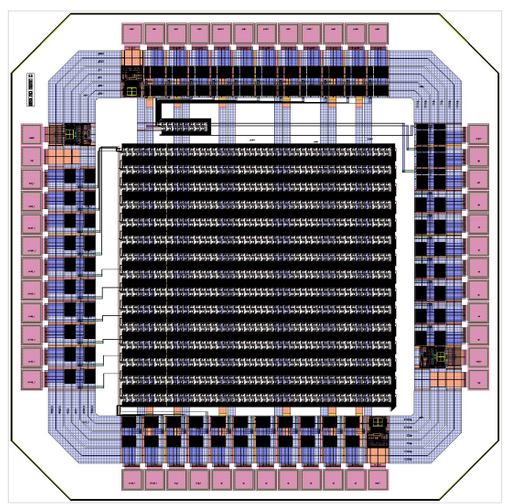


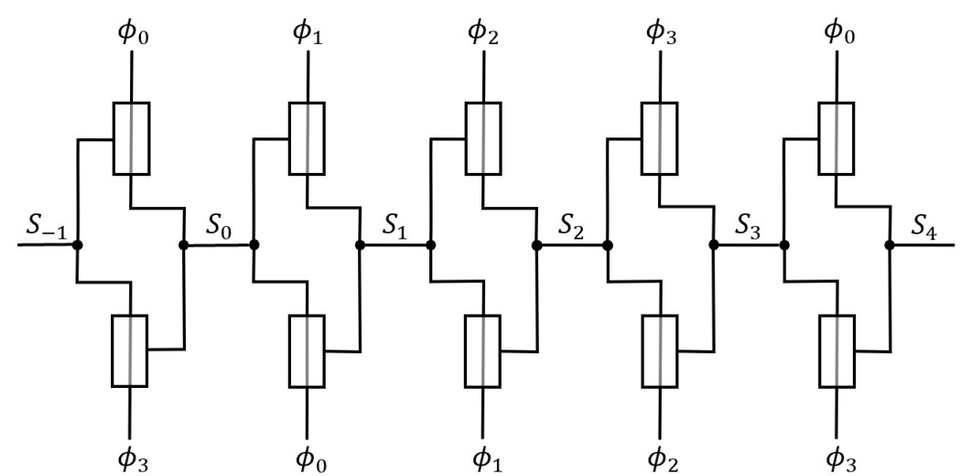
Simulation results for 2LAL obtained at Sandia in 2020:

- Energy dissipation per cycle per FET in shift register @50% activity factor at $f = 1$ MHz, $C_L = 10$ fF:
  - Spectre simulation of MESA 350 nm, $W = 800$ nm: **37 aJ ≈ 230 eV.**
  - Spectre simulation of MESA 180 nm, $W = 480$ nm: **6.9 aJ ≈ 43 eV.** ← Comparable to a UF data point for TSMC18 from 2004.
    - This beats end-of-roadmap standard CMOS substantially.

Test chip taped out in Aug. 2020:

- MESA 180 nm shuttle run.
- 2×2 mm die.
- 8-stage & 720-stage shift registers.
- Goal: Verify function & dissipation.

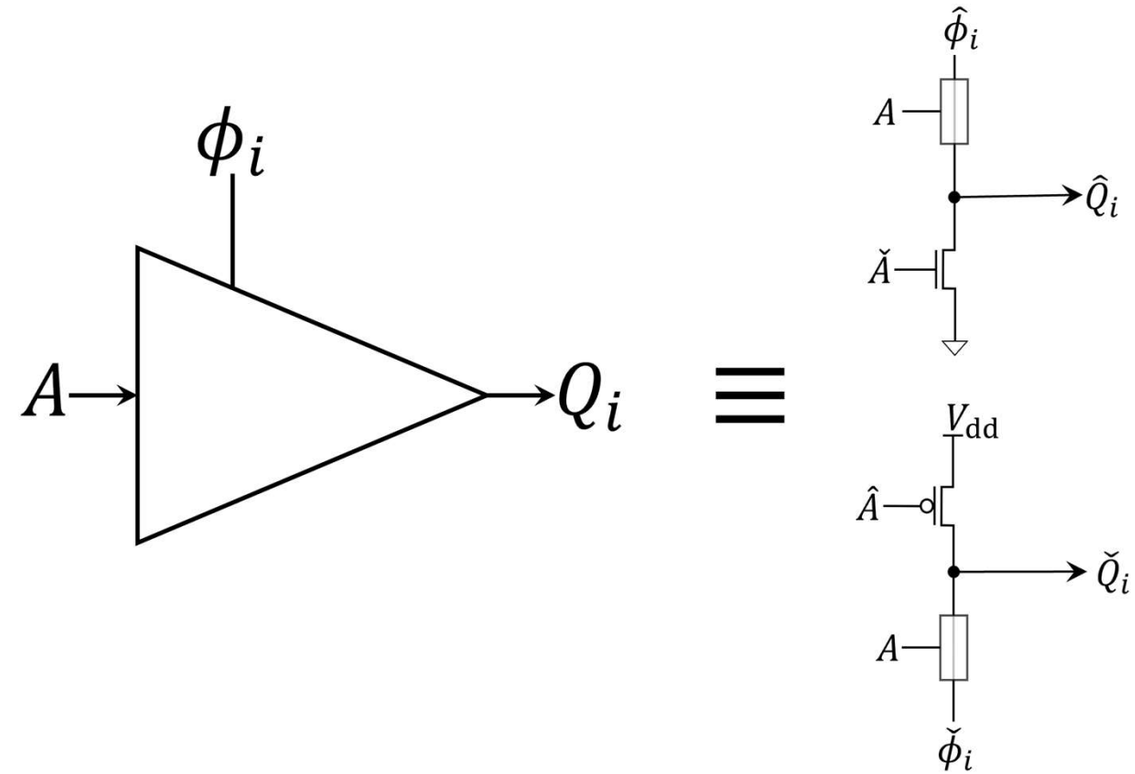


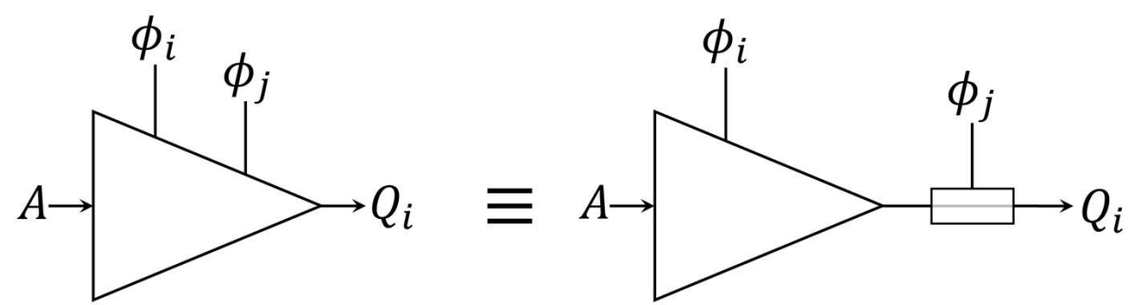


Basic Elements of S2LAL

Unlatched & Latching Static Adiabatic Buffers

- Unlatched version exchanges control of output between clock and fixed supply, depending on activity of input.
  - Handoff should only happen when levels match.
  - Athas '94 called this same element an *adiabatic amplifier*.
    - Athas, W.C., *et al.* "Low-power digital systems based on adiabatic-switching principles," *IEEE Trans. VLSI Sys.* **2**(4):398–407, 1994. doi:10.1109/92.335009
- *Latching* version uses an out-of-phase clock to latch (or unlatch!) the output.
  - NOTE: This requires additional higher-level structure to make the resulting circuit fully static!





S2LAL Reversible Pipeline Structure

Frank *et al.*, "Reversible Computing with Fast, Fully Static, Fully Adiabatic CMOS," IEEE Int'l Conf. on Rebooting Computing (ICRC '20), 10.1109/ICRC2020.2020.00014

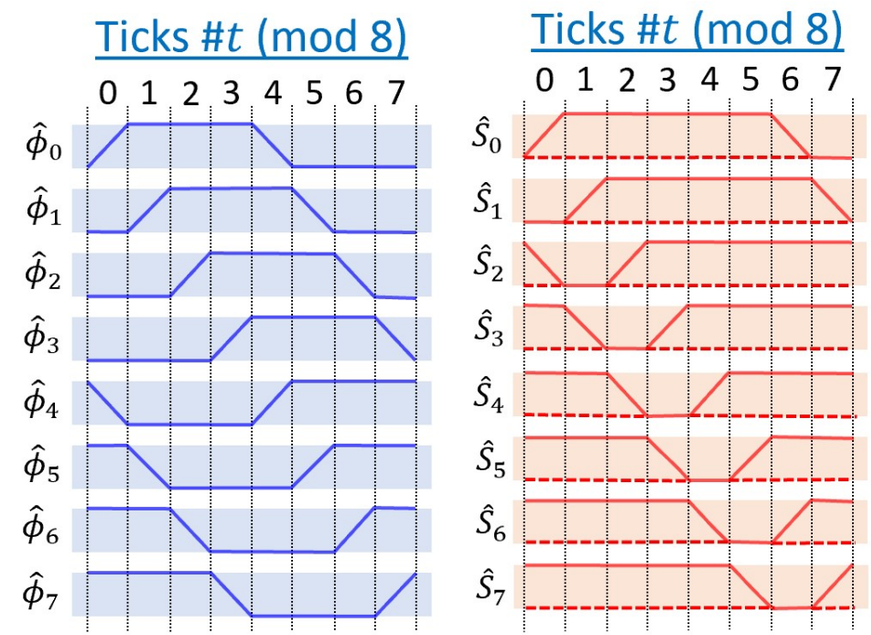
Paired forward and reverse stages:
- Forward stages activate to compute *later* signals from *earlier* ones.
- Reverse stages *de-activate* to *de-compute earlier* signals from *later* ones.

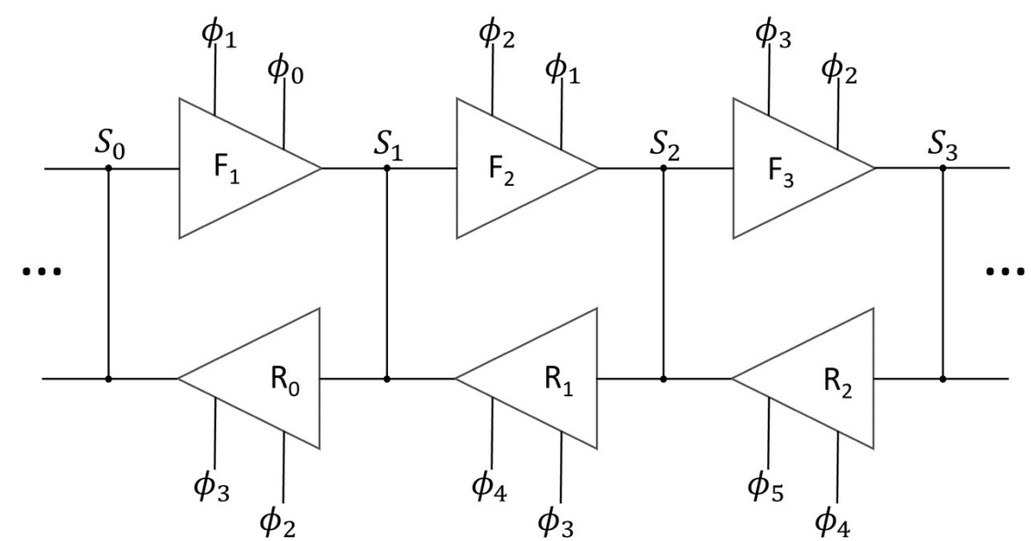
Every signal S_i must stay active for (at least) 5 ticks:
- Provides sufficient time for the following sequence of steps:
  - (1) Activate forwards stage F_{i+1}, (2) Activate reverse stage R_i, (3) Handoff control of S_i from F_i to R_i, (4) Deactivate forwards stage F_i, (5) Deactivate reverse stage R_{i-1}.

Add 3 ticks for transitions & inactive handoff:
- Total cycle length = **8 ticks** min.

Note control of each signal S_i is handed off to forward stage F_i on ticks $\#i-1$, and to reverse stage R_i on ticks $\#i+3$.
- Signal S_i goes valid on ticks $\#i$ and invalid (inactive) on ticks $\#i+6$.

For general logic, functions must be invertible.
- Optimizing whole pipeline gets into reversible algorithm design: Considered out of scope for this particular paper.





S2LAL Logic Gates

14-transistor AND gate, 16-transistor OR gate.

- Carefully designed to ensure that each internal node is always connected to either constant or variable source.
  - The structures shown are minimal, given the design constraints.

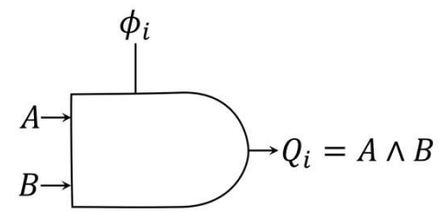
Inverting gates are done easily, by using signal pairs for complementary symbols:

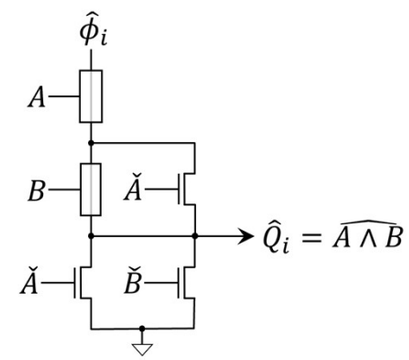
- $\mathrm{NOT}(A^1) = \mathrm{BUFFER}(A^0)$
- $\mathrm{NAND}(A^1, \mathrm{B}^1) = \mathrm{OR}(A^0, B^0)$
- $\mathrm{NOR}(A^1, \mathrm{B}^1) = \mathrm{AND}(A^0, B^0)$

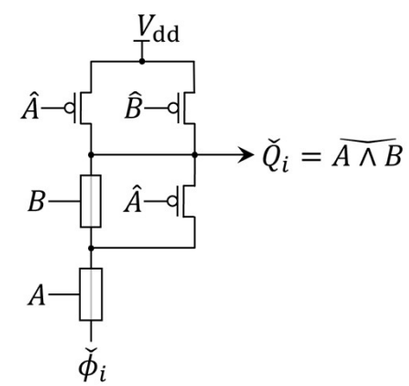
Also! Erik DeBenedictis invented an optimization to S2LAL that can compute the inverses as-needed, rather than keeping both the 0,1 signal pairs around:

- See https://zettaflops.org/zf004/ .

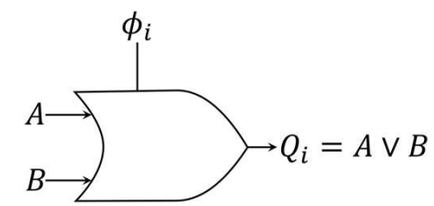
AND

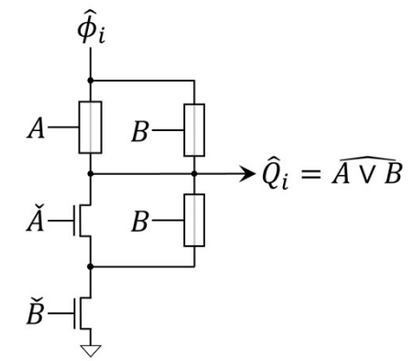


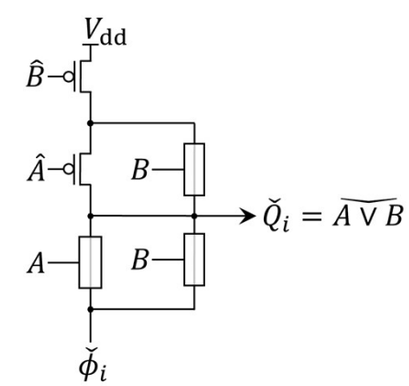




OR







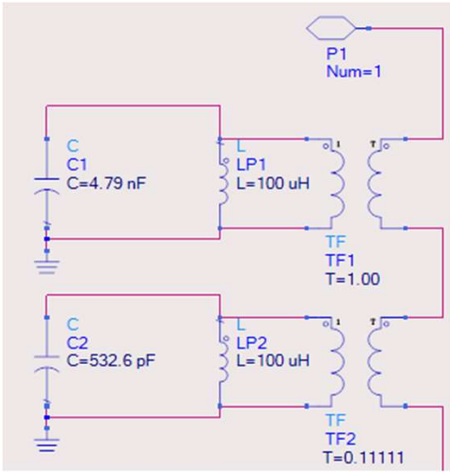
Resonator design effort, in progress…

Goal of this effort:

- Design & validate a high-efficiency resonant oscillator (for low-to-medium RF frequencies) that approximates a trapezoidal output voltage waveform.

Innovative design concept:

- **Transformer-coupled** assemblage of LC tank circuits with resonant frequencies corresponding to odd multiples of the fundamental frequency, excited in the right relative amplitudes to approximate the target wave shape

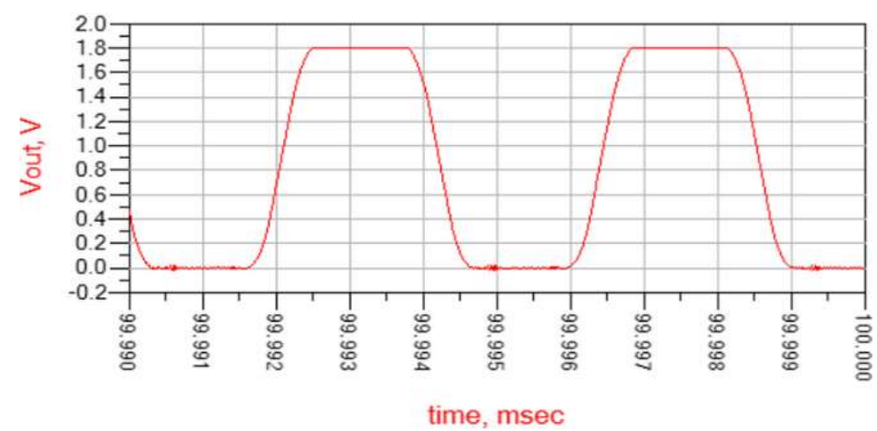


Some detailed requirement specifications:

- Initial target operating point: 230 kHz, 1.8V (optimal point for minimum dissipation in the UF study) **(MET.)**
  - However, our circuit technique should be adaptable over a wide range of frequencies and voltages.
- Tops and bottoms of trapezoidal wave should be within ≤5% of flatness throughout ¼ clock period. **(MET.)**
- The 10-90% rise/fall time should be between 75 & 100% of its nominal value (80% of 1/4 clock period) **(MET.)**
- Efficiency goals:
  - Quality factor of resonator during unpowered ring-down should be ≥1,000. (**MET.** Simulated value: ~3,000.)
  - Total energy dissipation per cycle during steady-state powered operation should be ≤1% of magnetically-stored energy in the resonator, when the oscillator is running in isolation. (Still needs validation.)
  - Total energy dissipation per cycle during steady-state powered operation should be ≤10% of the capacitively-stored energy on an appropriately-sized model (RC) load, when the oscillator is coupled to the load. (Needs validation.)

A number of significant design challenges that have been encountered so far:

- How to tune the relative amplitudes of the component resonant modes (Solved.)
- How to prevent phase drift and transfer of energy between modes (Solved.)
- Identifying/tailoring components to have precise-enough *L*, *C* values
- Designing a driver circuit that meets efficiency goals during steady-state operation
- Packaging & integration for a complete system including a resonator & a 2LAL die.



A provisional patent application has been filed on our resonator design.

Section III. Reversible Superconducting Technologies

Reversible Computing as The Sustainable Path Forward for General Digital Computing

Adiabatic Reversible Computing in Superconducting Circuits

Work along this general line has roots that go all the way back to Likharev, 1977.

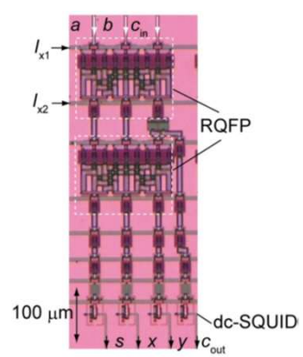
Most active group at present is Prof. Yoshikawa's group at Yokohama National University in Japan.

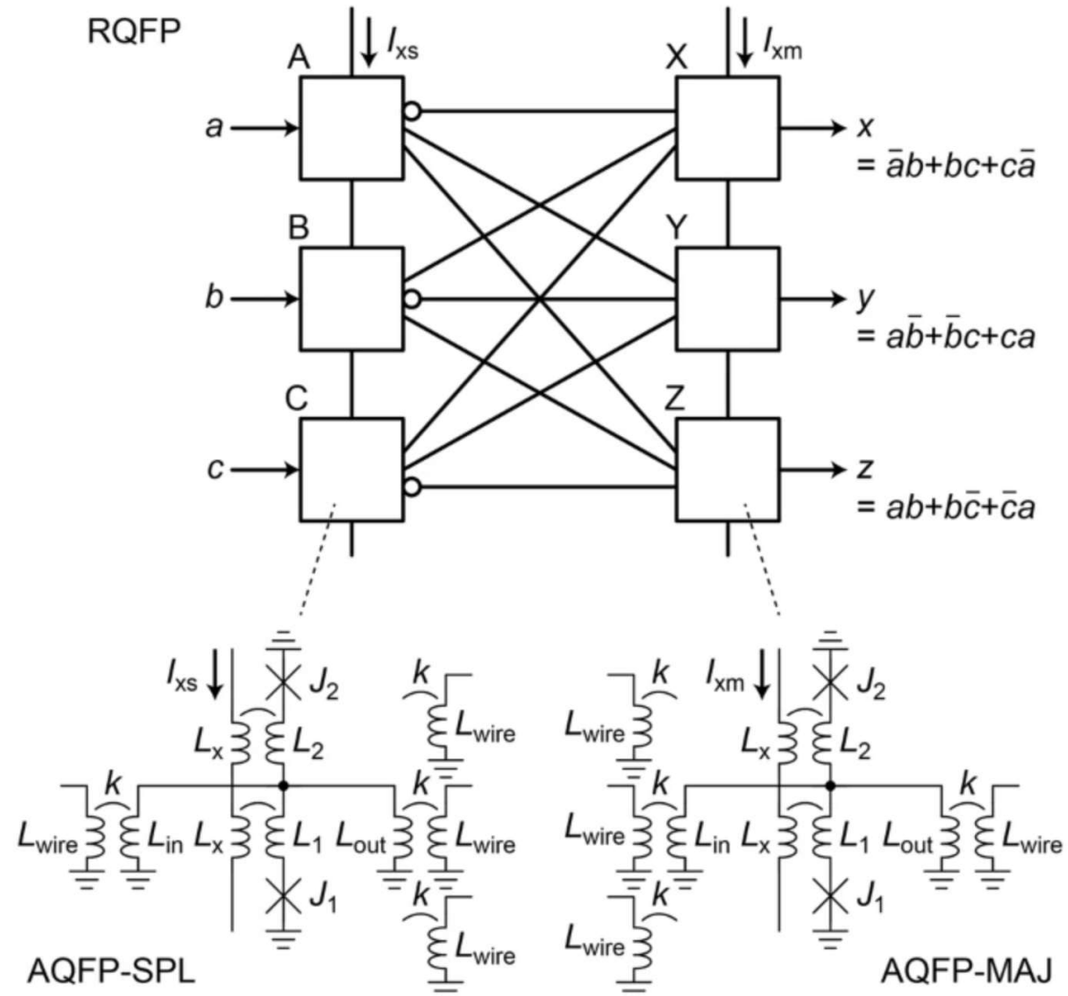
Logic style called *Reversible Quantum Flux Parametron* (RQFP).

Shown at right is a 3-output *reversible majority gate*.

Full adder circuits have also been built and tested.

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





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

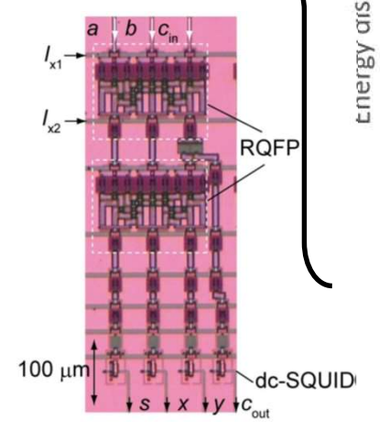
Reversible adiabatic superconductor logic:

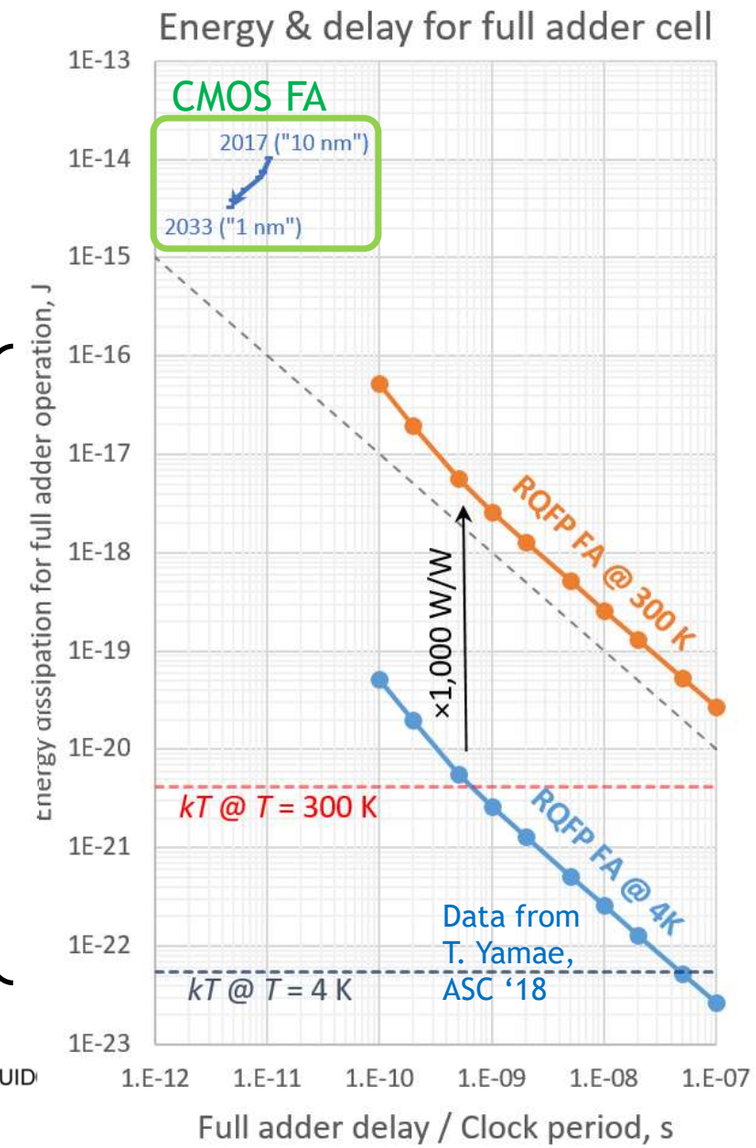
- State-of-the-art is the **RQFP** (Reversible Quantum Flux Parametron) technology from Yokohama National University in Japan.
  - Chips were fabricated, function validated.
- Circuit simulations predict DdP is >1,000× *lower* than even *end-of-roadmap* CMOS.
  - Dissipation extends *far below* the 300K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

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

- We have a project at Sandia exploring one possible superconductor-based approach for this (more later)…
  - But, what are the *fundamental* (technology-independent) limits, if any?

RQFP = Reversible Quantum Flux Parametron (Yokohama U.)



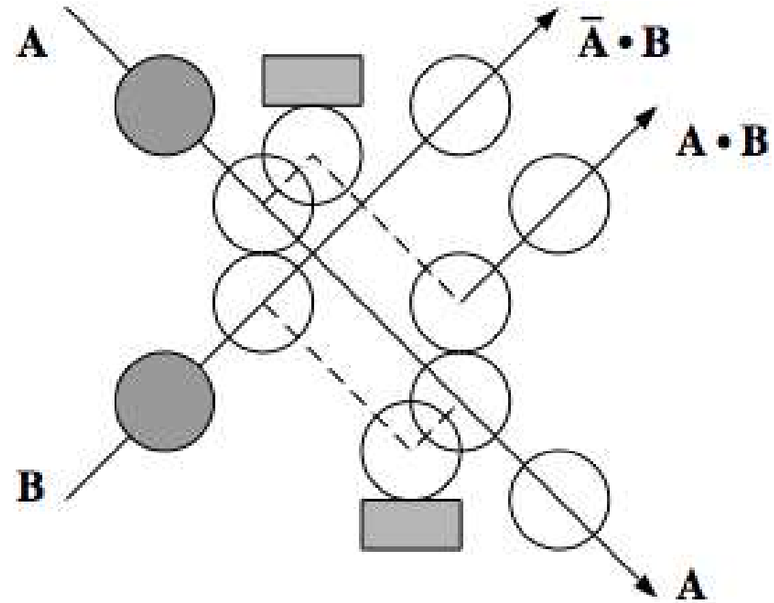


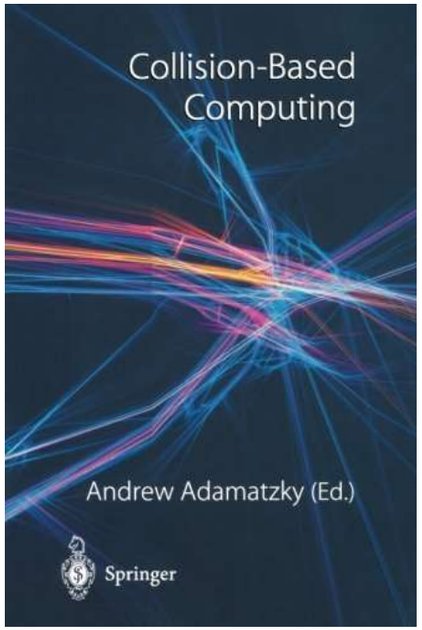
Ballistic Reversible Computing

Can we envision reversible computing as a *deterministic* elastic interaction process?

Historical origin of this concept:

- Fredkin & Toffoli's *Billard Ball Model* of computation ("Conservative Logic," IJTP 1982).
  - Based on elastic collisions between moving objects.
  - Spawned a subfield of "collision-based computing."
    - Using localized pulses/solitons in various media.





No power-clock driving signals needed!

- Devices operate when data signals arrive.
- The operation energy is carried by the signal itself.
  - Most of the signal energy is preserved in outgoing signals.

However, all (or almost all) of the existing design concepts for ballistic computing invoke implicitly *synchronized* arrivals of ballistically-propagating signals…

- Making that approach work in reality presents some serious difficulties, however:
  - Unrealistic in practice to assume precise alignment of signal arrival times.
    - Thermal fluctuations & quantum uncertainty, at minimum, are always present.
  - Any relative timing uncertainty leads to chaotic dynamics when signals interact.
    - Exponentially-increasing uncertainties in the dynamical trajectory.
  - Deliberate *re*synchronization of signals whose timing relationship has become uncertain incurs an inevitable energy cost.

Can we come up with a *new* ballistic model of reversible computing that avoids these problems?

Ballistic Asynchronous Reversible Computing (BARC)

Problem: Conservative (dissipationless) dynamical systems generally tend to exhibit chaotic behavior…

- This results from direct nonlinear *interactions* between multiple continuous dynamical degrees of freedom (DOFs), which amplify uncertainties, exponentially compounding them over time…
  - *E.g.*, positions/velocities of ballistically-propagating "balls"
    - Or more generally, any localized, cohesive, momentum-bearing entity: Particles, pulses, quasiparticles, solitons…

Core insight: In principle, we can greatly reduce or eliminate this tendency towards dynamical chaos…

- We can do this simply by *avoiding* any direct interaction between continuous DOFs of different ballistically-propagating entities

Require localized pulses to arrive *asynchronously*—and furthermore, at clearly distinct, *non-overlapping* times

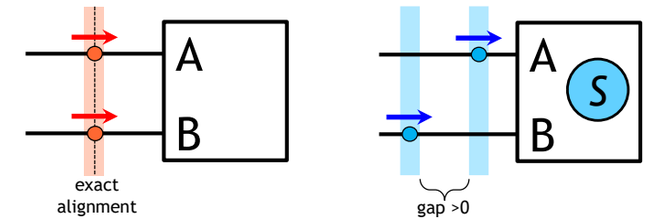
- Device's dynamical trajectory then becomes *independent* of the precise (absolute *and* relative) pulse arrival times
  - As a result, timing uncertainty per logic stage can now accumulate only *linearly*, not exponentially!
    - Only relatively occasional re-synchronization will be needed
- For devices to still be capable of doing logic, they must now maintain an internal discrete (digitally-precise) state variable—a stable (or at least metastable) stationary state, *e.g.*, a ground state of a well

No power-clock signals, unlike in adiabatic designs!

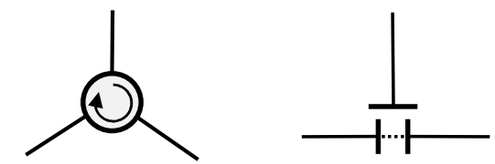
- Devices simply operate whenever data pulses arrive
- The operation energy is carried by the pulse itself
  - Most of the energy is preserved in outgoing pulses
    - Signal restoration can be carried out incrementally, or periodically

Goal of current effort at Sandia: Demonstrate BARC principles in an implementation based on fluxon dynamics in SuperConducting Electronics (SCE)

(BARCS effort)

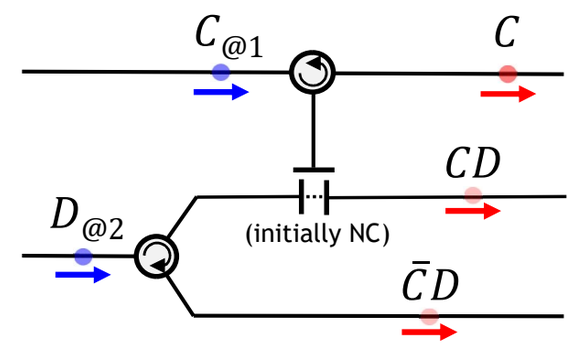


Synchronous Ballistic Asynchronous Ballistic



Rotary (Circulator) Toggled Barrier

Example BARC device functions



Example logic construction

Simplest Fluxon-Based (bipolarized) BARC Function

One of our early tasks: Characterize the simplest nontrivial BARC device functionalities, given a few simple design constraints applying to an SCE-based implementation, such as:

- (1) Bits encoded in fluxon polarity; (2) Bounded planar circuit conserving flux; (3) Physical symmetry.

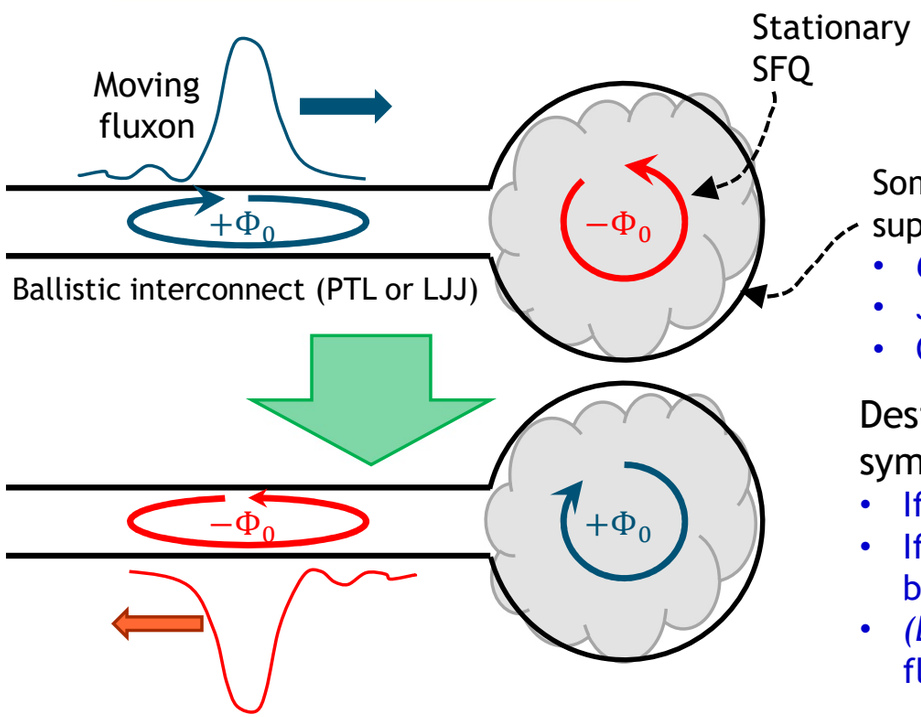
Determined through theoretical hand-analysis that the simplest such function is the ***1-Bit, 1-Port Reversible Memory Cell (RM):***

- Due to its simplicity, this was then the preferred target for our subsequent detailed circuit design efforts…

RM Transition Table

Input Syndrome		Output Syndrome
$+1(+1)$	$\rightarrow$	$(+1)+1$
$+1(-1)$	$\rightarrow$	$(+1)-1$
$-1(+1)$	$\rightarrow$	$(-1)+1$
$-1(-1)$	$\rightarrow$	$(-1)-1$

RM icon:



Some planar, unbiased, reactive SCE circuit w. a continuous superconducting boundary

- *Only* contains L's, M's, C's, and *unshunted* JJs
- Junctions should mostly be *subcritical* (avoids R_N)
- Conserves total flux, approximately nondissipative

Desired circuit behavior (NOTE: conserves flux, respects T symmetry & logical reversibility):

- If polarities are opposite, they are swapped (shown)
- If polarities are identical, input fluxon reflects back out with no change in polarity (not shown)
- *(Deterministic) elastic 'scattering'* type interaction: Input fluxon kinetic energy is (nearly) preserved in output fluxon

RM—First working (in simulation) implementation!

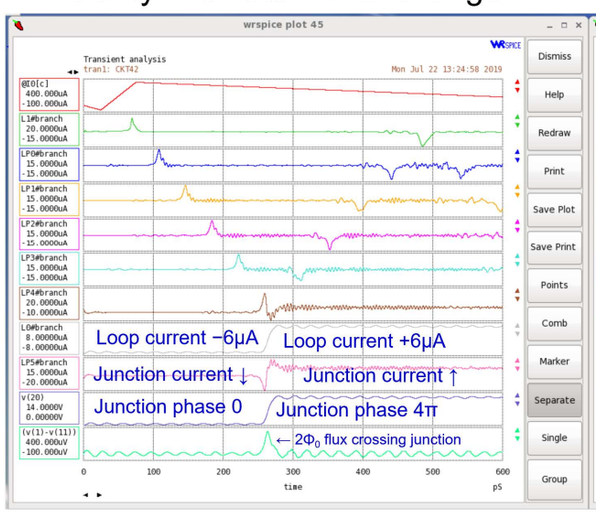
Erik DeBenedictis: "Try just strapping a JJ across that loop."
- This actually works!

"Entrance" JJ sized to = about 5 LJJ unit cells (~1/2 pulse width)
- I first tried it twice as large, & the fluxons annihilated instead…
  - "If a 15 μA JJ rotates by 2π, maybe ½ that will rotate by 4π" 🤔

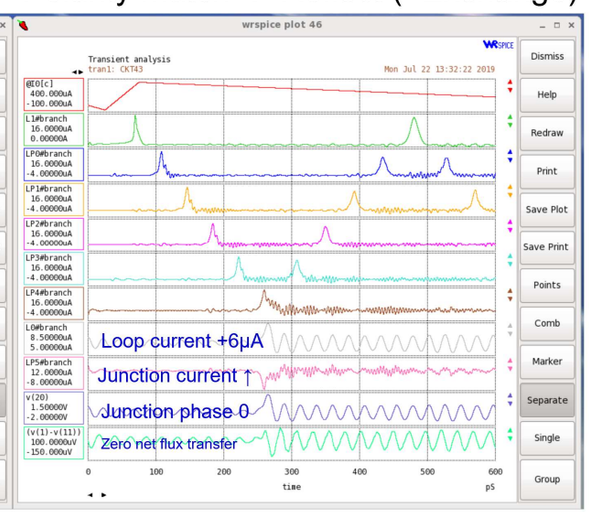
Loop inductor sized so ±1 SFQ will fit in the loop (but not ±2)
- JJ is sitting a bit below critical with ± 1

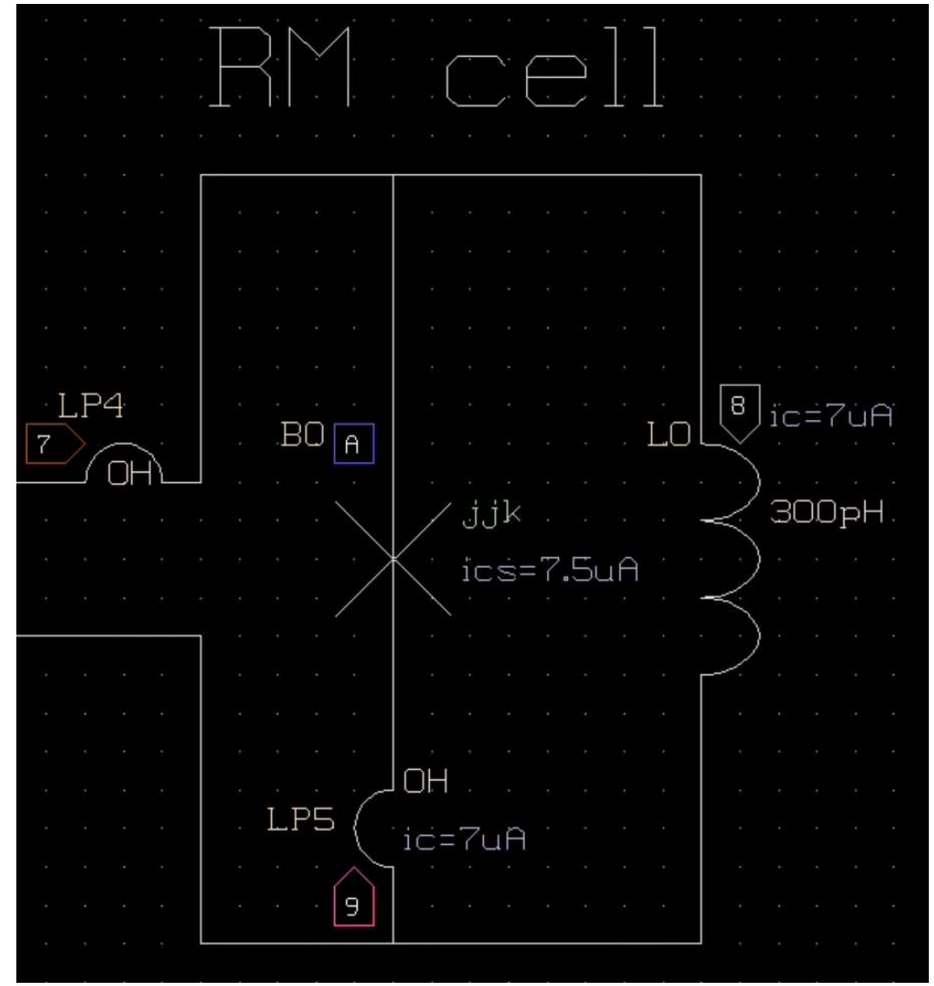
WRspice simulations with ±1 fluxon initially in the loop
- Uses `ic` parameter, & `uic` option to `.tran` command
  - Produces initial ringing due to overly-constricted initial flux
    - Can damp w. small shunt *G*

Polarity mismatch → Exchange



Polarity match → Reflect (=Exchange)

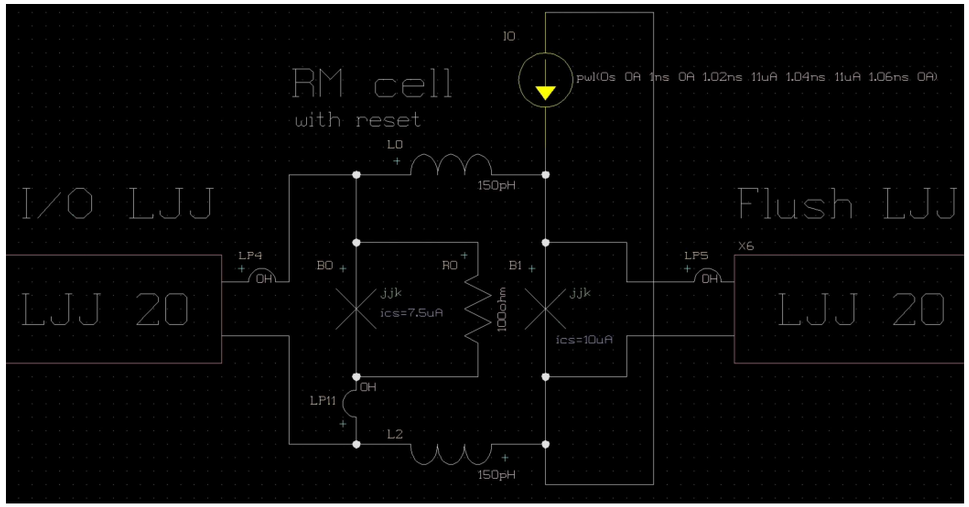


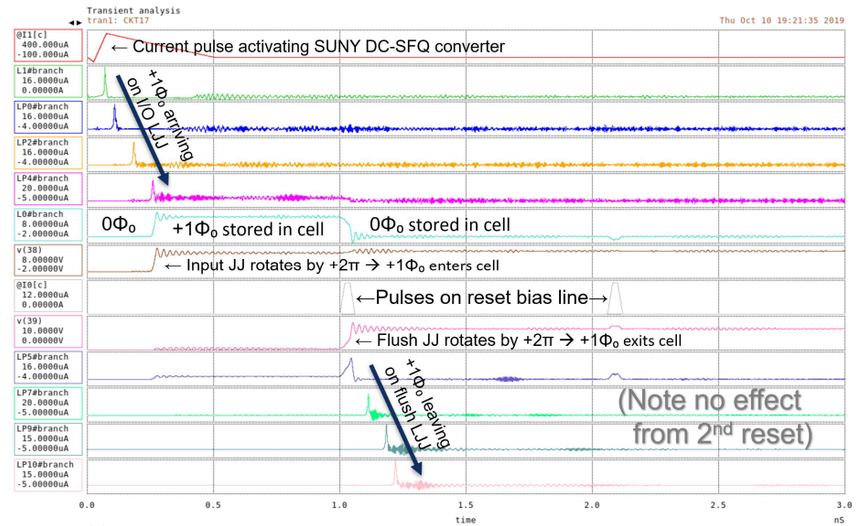


Resettable version of RM cell—Designed & Fabricated!

Apply current pulse of appropriate sign to flush the stored flux (the pulse here flushes out positive flux)

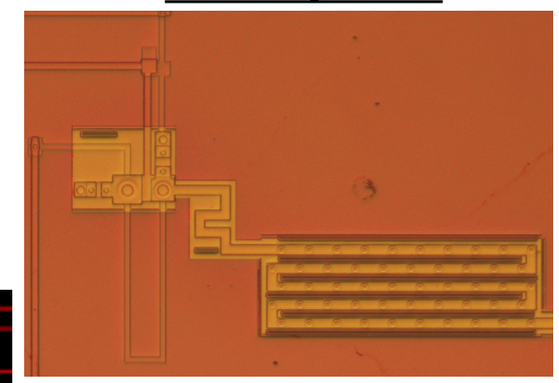
- To flush either polarity → Do both (±) resets in succession

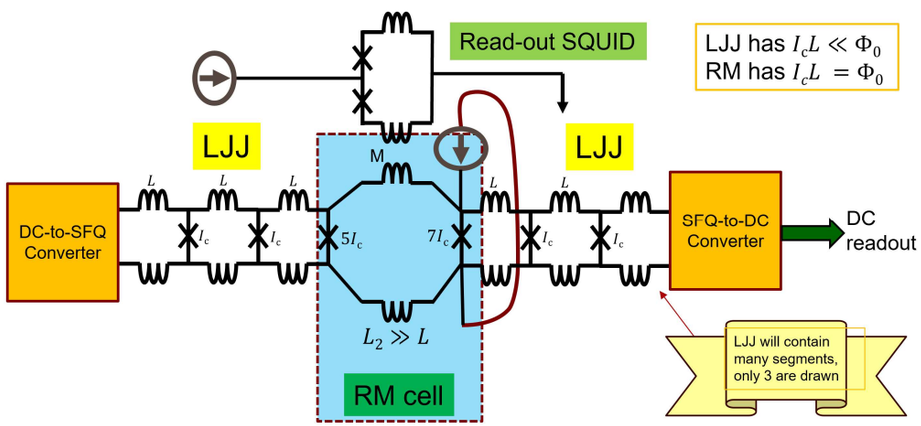


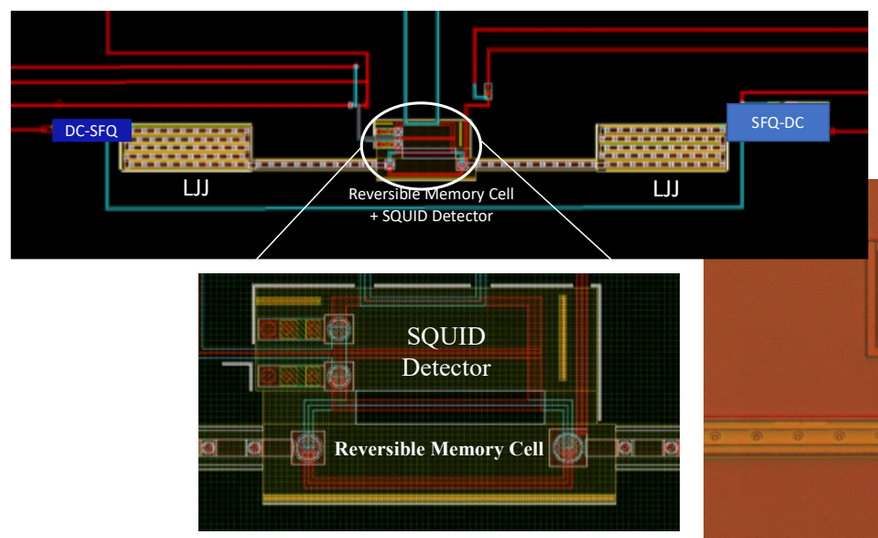


Fabrication at SeeQC with support from ACI

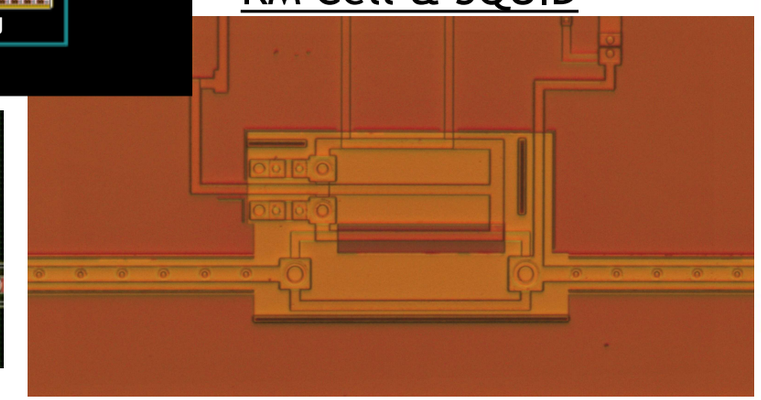
DC-SFQ & LJJ







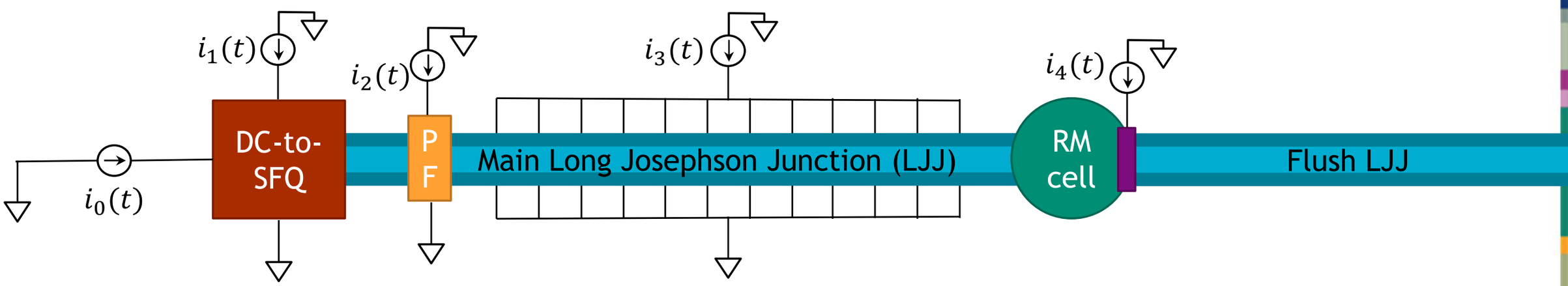
RM Cell & SQUID



Concept for energy dissipation measurements

Experimental protocol:

- Pulse the "flush/reset" JJ using i_4 bias ± to ensure there is 0 trapped flux in the RM cell initially.
- Use the DC-SFQ converter to inject a $+\Phi_0$ pulse through the polarity filter (PF) into RM cell and store it.
- Use the DC-SFQ converter to inject a $-\Phi_0$ pulse through the polarity filter (PF), and then immediately…
- Turn off the polarity filter (PF)—that is, reset it to 0 bias current (and ideally, tristate it).
- Initiate a *periodic* ± current bias waveform (symmetric square wave) on the LJJ (i_3).
  - Purpose of this: Alternate between accelerating $-\Phi_0$ pulses to the right and $+\Phi_0$ pulses to the left, vs. the opposite (after reflection off PF).
- At appropriate combinations of amplitude I & frequency f, the i_3 drive signal will hit a resonance.
  - Detect resonance by measuring reflected i_3 power with an RF network analyzer—at resonance there will be a dip in reflected power.
- From the resonance point I, f and the measured S_{11}, we can immediately calculate the following parameters:
  - Fluxon velocity
  - Total energy dissipation per cycle
- To infer what part of the energy dissipation is due to the RM cell:
  - Just do a similar test with a simple inverting reflector (*e.g.*, an open circuit) in place of the RM cell.

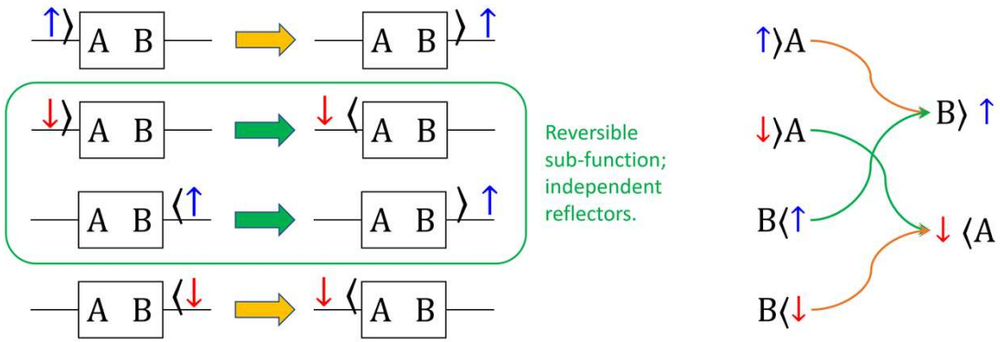


Some new ballistic asynchronous device concepts (in progress)

Directional Polarity Filter (DPF)

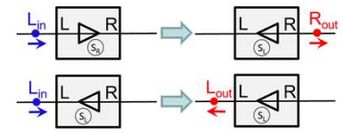
(not fully reversible, but simple & useful!)

- Passes "up" fluxons (↑) moving towards the right, and "down" (↓) fluxons moving towards the left (or vice-versa, depending on bias)

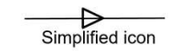


Flipping Diode (Notes from 2016)

- The only nontrivial two-state, two-terminal, time-reversal-symmetric A.R. device
  - The only other TRS two-state, two-terminal AR devices are just barriers or renamers with redundant states
- Later, we'll see that with some signal routing/renaming, this can act as a reversible SR flip-flop (reversible SRAM cell)
  - And if we also add a simple sequencing protocol, we can even make it into an asynchronous reversible AND gate!



Input Symbol	Initial State	Final State	Output Symbol
L	S_R	S_L	R
L	S_L	S_L	L
R	S_R	S_R	R
R	S_L	S_R	L



Reversible Polarity Filter (RPF)

Table 42. #2. Down-Pass Reversible Polarity Filter (DRPF). $\mathcal{R}_2$-symmetric. Violates $\mathcal{M}$, $\mathcal{P}$. (Requires bias or trapped flux.)

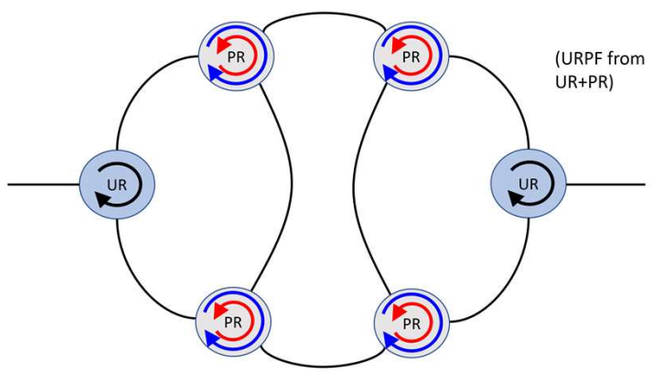
Input syndrome	Output syndrome
↑⟩A	↑⟨A
↓⟩A	↓⟨B
↑⟩B	↑⟨B
↓⟩B	↓⟨A

Table 47. #7. Up-Pass Reversible Polarity Filter (URPF). $\mathcal{R}_2$-symmetric. Violates $\mathcal{M}$, $\mathcal{P}$. $\mathcal{M}$-symmetric to #2.

Input syndrome	Output syndrome
↑⟩A	↑⟨B
↓⟩A	↓⟨A
↑⟩B	↑⟨A
↓⟩B	↓⟨B

Note it would be theoretically possible to construct this element if it's possible to implement a Universal Rotary (UR) and a Polarized Rotary (PR).

- Of course, there may also be simpler implementations.



(Detailed SCE circuit designs for the various functions shown here are in progress.)

Polarity-Dependent Flipping Diode (PFD)

I/O Syndrome		
Port	State	Fluxon
Left	↑↓	↑
Left	↑↓	↓
Left	↓↑	↑
Left	↓↑	↓
Right	↑↓	↑
Right	↑↓	↓
Right	↓↑	↑
Right	↓↑	↓

Section IV. Fundamental Physical Limits of Reversible Computing

Reversible Computing as The Sustainable Path Forward for General Digital Computing

Fundamental Physical Limits of Reversible Computing

This is a severely under-studied topic, but preliminary theoretical indications to date are that:

- For quantum-mechanical reversible devices that are well isolated from their thermal environment, there is a regime in which *exponential adiabaticity* (*i.e.*, Landau-Zener scaling) can substantially suppress dissipation even at relatively high speeds (Pidaparthi & Lent '21)
  - Pidaparthi, S.S., & Lent, C.S., "Energy dissipation during two-state switching for quantum-dot cellular automata," *J. Appl. Phys.*, **129**(2), 024304, 2021. doi:10.1063/5.0033633
  - This result could have **enormous practical implications** for the economic competitiveness of reversible computing.
- At slow speeds, dissipation asymptotically converges to the classic adiabatic scaling (1/delay). (Earley '20)
  - Earley, W., "Engines of Parsimony: Parts I-III," preprints, Jul. 2020–Jan. 2021. arXiv:{2007.03605, 2011.04054, 2012.05655}
- Implies that asymptotically, performance boost from RC scales up with $\sqrt{D}$ (D = depth/thickness) —This is a result that actually dates all the way back to Frank & Knight '97:
  - Frank, M. P., & Knight, Jr., T. F., "Ultimate theoretical models of nanocomputers," *Nanotechnology*, **9**(3):162–176, 1998. doi:10.1088/0957-4484/9/3/005; and Frank, M.P., "Reversibility for Efficient Computing," Ph.D. thesis, MIT, 1999.
- It appears likely that fundamental theoretical tools from the field of non-equilibrium quantum thermodynamics (NEQT) can be applied to make the above results more general and rigorous.
  - K. Shukla, "Fundamental Thermodynamic Limits of Classical Reversible Computing via Open Quantum Systems," position paper, CCC Workshop on Physics & Engineering Issues in Adiabatic/Reversible Classical Computing, Oct. 2020. https://cfwebprod.sandia.gov/cfdocs/CompResearch/docs/Shukla-etal-20-CCC-pos-paper.pdf
  - K. Shukla, "Foundations of the Lindbladian Approach to Adiabatic and Reversible Computing," plenary talk, CCC Workshop on Physics & Engineering Issues in Adiabatic/Reversible Classical Computing, Oct. 2020. https://cra.org/ccc/wp-content/uploads/sites/2/2020/10/Shukla-Fundamental-Thermodynamic-Limits-of-Classical-Reversible-Computing-via-Open-Quantum-Systems.pdf, https://www.youtube.com/watch?v=mzoL6m-2rrA&feature=emb_logo

Likharev's dissipation limits

Likharev '81 analyzed limits of dissipation for his reversible JJ-based Parametric Quantron (PQ) technology concept.

- Based on analyzing rates of crossing a potential energy barrier through thermal excitation and quantum tunneling.

International Journal of Theoretical Physics, Vol. 21, Nos. 3/4, 1982

Classical and Quantum Limitations on Energy Consumption in Computation

K. K. Likharev

Department of Physics, Moscow State University, Moscow 117234, U.S.S.R.

Received May 6, 1981

Main results:

- Limit due to classical thermal excitation over barrier (assuming underdamped junction):

$$W_C \approx \frac{k_B T}{\omega_c \tau} \ln \frac{1}{\omega_A \tau p}.$$

  - $\omega_c = \frac{k}{\eta} \approx \frac{2\Delta}{\hbar}$ with elasticity modulus $k = \frac{d^2U}{dx^2}$ and effective viscosity η; and 2Δ is the superconducting gap energy;
  - ω_A approximates to the JJ plasma frequency $\omega = \sqrt{k/\eta} = \sqrt{2q_e I_c/\hbar C}$, and τ is the cycle period;
  - p is the tolerable error probability per operation.

Limit due to quantum-mechanical tunneling through the barrier:

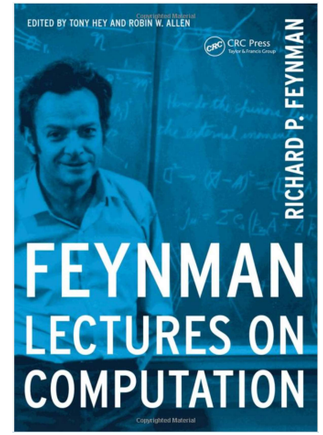
$$W_C \approx \frac{\hbar}{\tau} \ln \frac{1}{\omega_c \tau p}.$$

However! Likharev himself admits the limitations of this analysis:

- It is not a *fundamental*, technology-independent analysis.
- Alternative device concepts might do better!

> Our approach to the problem, of course, leaves open whether it is possible to invent some novel device providing lower power consumption. If we limit ourselves to the quasistatic devices, where the computation can be stopped at any moment, without inducing an error, one can hardly get away from the above estimates. In fact, the only role of the parametric quantron in our discussion has been to demonstrate how a flexible bistable potential well could be physically realized. (Of course, some numerical factors can appear in the estimates if peculiar well shapes are taken into account.)
>
> One can, however, argue that the above-mentioned condition of quasi-statics is by no means compulsory, and that the information can be processed by some "dynamical" devices, where the cycle period can be shorter than the relaxation time. This problem is left for further analysis.

Feynman's dissipation limits



An example we gave of reversible computing was that of the chemical process of copying DNA. This involved a machine (if you like) that progressed in fits and starts, going forward a bit, then backwards, but more one than the other because of some driving force, and so ended up doing some computation (in this case, copying). We can take this as a model for more general considerations and will use this "Brownian" concept to derive a formula for the energy dissipation in such processes. This will not be a general formula for energy dissipation during computation but it should show you how we go about calculating these things. However, we will precede this discussion by first giving the general formula[7], and then what follows can be viewed as illustration.

[7]This rule is pretty general, but there will be exceptions, requiring slight corrections. We will discuss one such, a "ballistic" computer, in §5.5. [RPF]

In lectures for his 1983-1986 CalTech course, "Potentialities and Limitations of Computing Machines," Feynman derived a limit on energy dissipation per step for *Brownian machines* (*e.g.*, DNA copying) driven by chemical potentials.

- He concludes that an approximate formula for this is:

$$energy\ loss/step = kT\,\frac{minimum\ time\ taken/step}{time/step\ actually\ taken}$$

However, he mentions in a footnote that a "slight correction" to this expression would be needed for *ballistic* machines, and later argues, quite informally, that in that case, the expression should be:

$$kT\,\frac{time\ to\ make\ collision}{speed\ [sic]\ at\ which\ it\ happens}$$

$$kT\,\frac{time\ to\ make\ collision}{speed\ at\ which\ it\ happens}. \qquad (5.37)$$

This expression has not been analyzed in any great detail for the billiard ball machine.

- An arguably very similar expression, *but*:
  - The whole argument in this part of the notes is *extremely* brief and informal ("hand-wavy")
  - The possible application of *e.g.* the Landau-Zener formula for quantum-mechanical scattering processes is not considered at all
  - Modern STA (Shortcuts to Adiabaticity) techniques had not even been developed yet, and so of course are also not considered
  - *Asynchronous* ballistic models (e.g. ABRC) which avoid chaotic instabilities had also not been invented yet

Thus, we must conclude that Feynman's analysis of this problem is *not definitive*, nor the final word.

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?

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

J. Low Power Electron. Appl. **2018**, *8*(3), 30; https://doi.org/10.3390/jlpea8030030 Open Access Article

Exponentially Adiabatic Switching in Quantum-Dot Cellular Automata

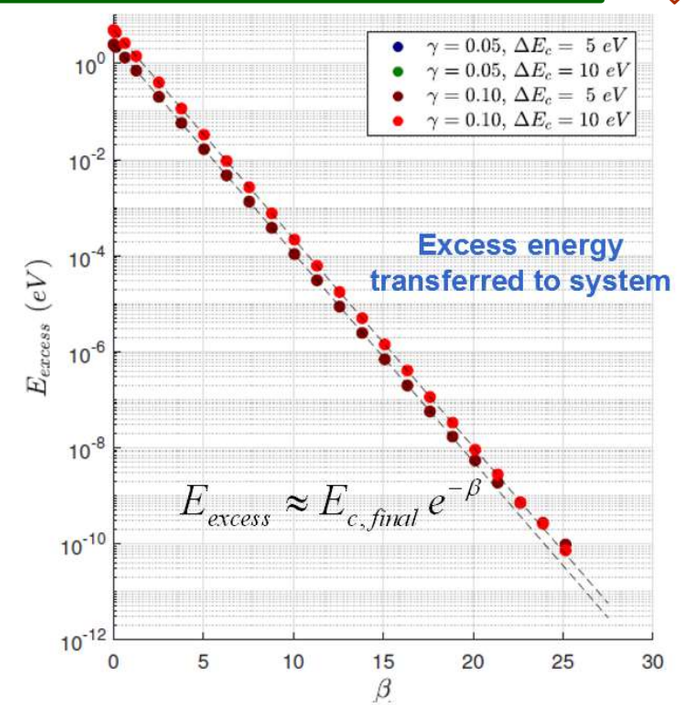
Subhash S. Pidaparthi and **Craig S. Lent** *

Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556, USA

* Author to whom correspondence should be addressed.

Received: 15 August 2018 / Revised: 5 September 2018 / Accepted: 5 September 2018 / Published: 7 September 2018

(This article belongs to the Special Issue Quantum-Dot Cellular Automata (QCA) and Low Power Application)



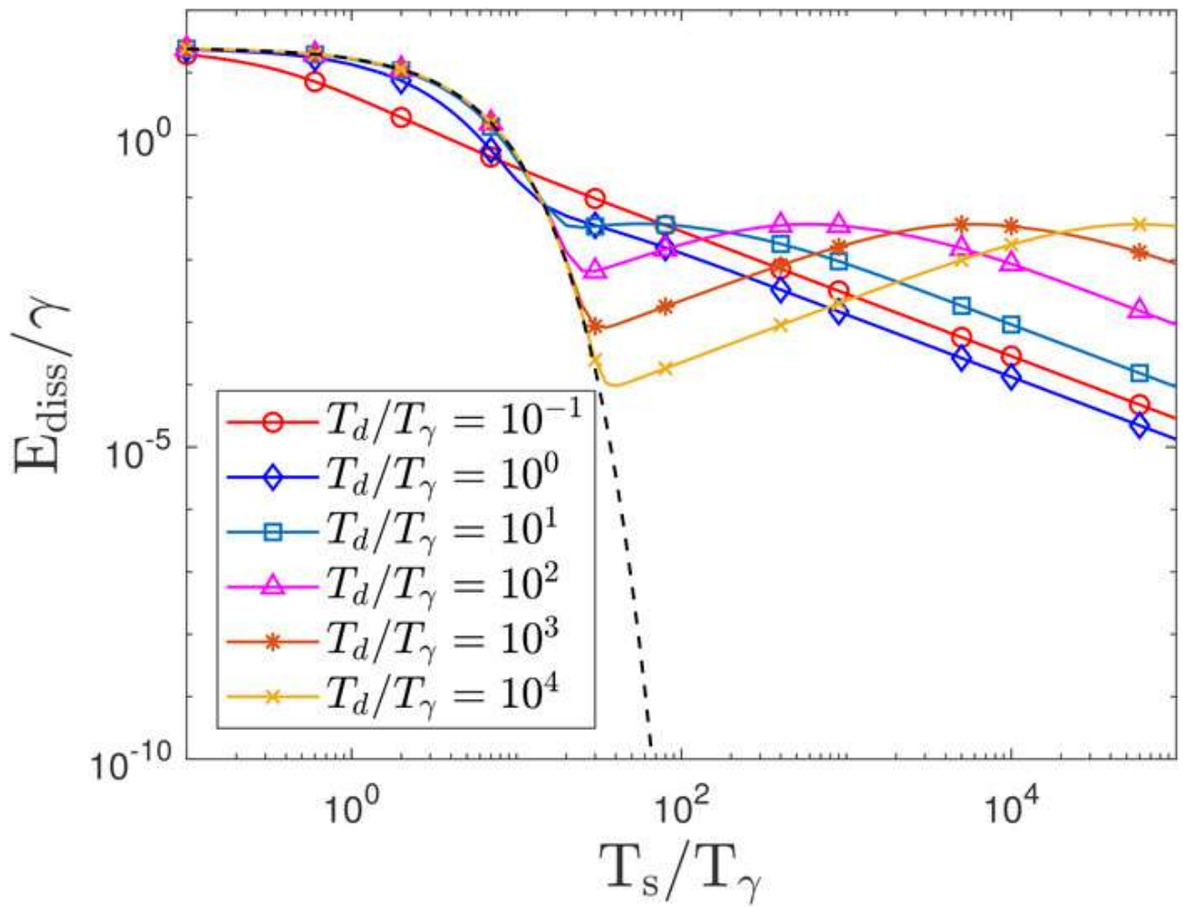


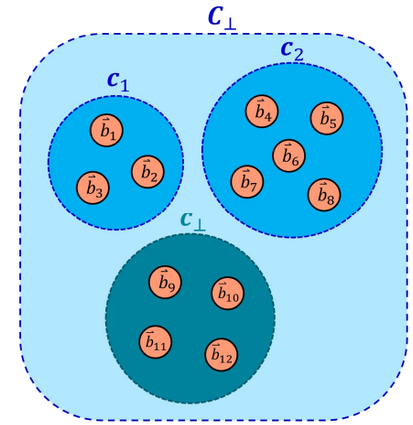
FIG. 10. Dissipated energy of an open system as a function of switching speed for different dissipation time constants. The dashed line is the excess energy of an isolated system. Here, the environmental temperature $k_BT / \gamma = 0.5$.

Published in: Subhash S. Pidaparthi; Craig S. Lent; *Journal of Applied Physics* **129,** 024304 (2021)
DOI: 10.1063/5.0033633
Copyright © 2021 Author(s). (Excerpted with permission.)

Fundamental Physics of Reversible Computing

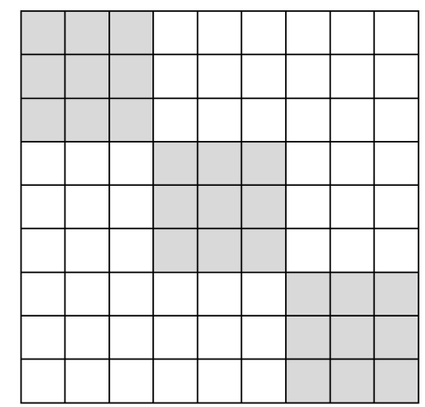
(Work with Karpur Shukla, Brown University)

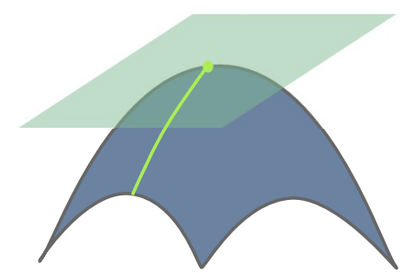
BROWN



$U_s^t(\mathfrak{S}, \mathcal{B}) \Vdash \mathcal{C}_s^t(O_s^t, \rho_s)$

- Goals of this effort:
  - Look for fundamental physical limits of reversible computing
    - *E.g.*, minimum entropy production per operation as a function of delay, temperature, etc.
  - Identify ways to harness exotic quantum phenomena if needed to saturate the limits
- Steps completed so far:
  - Identification of classical computational states with disjoint sets of orthonormal basis states in a (time-dependent, in general) *protocomputational basis* $\mathcal{B}$.
  - Formalization of what it *means* for a unitary *quantum* evolution U_s^t on a computational system $\mathfrak{S}$ (physical computer) to *implement* a given *classical* (and possibly reversible and/or stochastic) computational operation O_s^t between times s and t.
- Research strategy looking forward:
  - Computational states correspond to *decoherence-free subspace blocks* of overall Hilbert space.
  - Quantum Markov equation with multiple asymptotic states: Admits subspace dynamics 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…
    - Thermodynamic uncertainty relations, dissipation, and dissipation-delay product.





Section IV. Future Work & Conclusion

Reversible Computing as The Sustainable Path Forward
for General Digital Computing

Assessment of Architectural Implications

(Work with Tom Conte and Anirudh Jain, Georgia Tech)

Suppose the study of fundamental limits will be successful, and yield a better understanding of the limiting tradeoffs between dissipation, speed, *etc.*

- **Question:** What would be the *architectural* implications of attaining those limits?
  - **Note:** We can begin exploring this question even before the main study yields results!

Research plan for Sandia/GT collaboration:

- **Sandia** defines a common *generic* model of abstract reversible device technologies (including adiabatic and/or asynchronous variants), characterized by key parameters and their scaling, *e.g.*, $E_{\mathrm{diss}}(t_{\mathrm{d}})$, P_{leak}, *etc.*
- **Georgia Tech** designs a hierarchy of architectural components *composed* out of these generic reversible elements, leading towards a RISC style CPU architecture, including:
  - Multiplexers (32 bits wide, 2-to-1 and 4-to-1).
  - Comparators and Adders (32-bit-wide).
  - Integer Multipliers (32×32 bits, used for address arithmetic).
  - 32-bit ALU (Arithmetic-Logic Unit).
  - Canonical 5-stage pipelined RISC style processor including control unit.
- Meanwhile, **Sandia** supplies various special cases of the generic model reflecting interesting candidate (including hypothetical or preliminary) scaling relations emerging from main study.
  - **Georgia Tech** analyzes the effect of these particular model cases on the efficiency of architectural components
- **Georgia Tech** concludes by:
  - Conducting a study of the pareto optimal frontier of efficiency for *partially*-reversible architectures

Future Work

Some additional priority directions for future work in reversible computing technology include the following:

- Adiabatic CMOS:
  - Finish developing high-quality resonators, & integrate with fully adiabatic CMOS demonstration chips
  - Develop cell libraries and design tool enhancements to make adiabatic CMOS more accessible to designers
  - Design new FET geometries optimized for adiabatic operation at cryo temperatures
  - Develop commercializable adiabatic CMOS processors (both general- and special-purpose)
- Reversible superconducting technologies:
  - Continued development of the adiabatic reversible superconducting logic styles (AQFP/RQFP)
  - Continued development of the ballistic reversible superconducting logic styles (RFL/BARC)
- Invent/develop novel device technologies for RC
  - Harness topological invariants, quantum Zeno effects, other exotic phenomena?
- Continue firming up fundamental physical limits of RC
  - Derive a rigorous NEQT formulation of limits

Conclusion

Reversible computing will *definitely* be required in order for general digital computing to avoid hitting a plateau in gate-level energy efficiency, and beginning to stagnate in its development, within only the next decade or so…

- We had better begin working aggressively on it now for solutions to be ready in time!

Proof-of-concept implementations of reversible computing have already been constructed on top of both CMOS and superconducting technology platforms.

- Based on various concepts that have been under sporadic development since the 1970s.
- The technology is now ready for *much* more intensive practical development to start!

We have not even *begun* to approach the limits of what's possible to achieve if reversible computing technologies are developed aggressively…

- There is a potential to gain, over time, *vast* economically beneficial improvements in system-level power-performance and cost-performance figures of merit for *general* digital computing applications!
  - Potentially taking us ***orders of magnitude*** beyond ***any physically possible*** non-reversible technology!

There is an enormous opportunity here, that is just waiting for everyone to notice it!

- When the world finally realizes that reversible computing indeed offers a viable path forward that bypasses the roadblocks faced by conventional computing for general digital applications, it will be a watershed moment for the future of technology, and civilization in general.