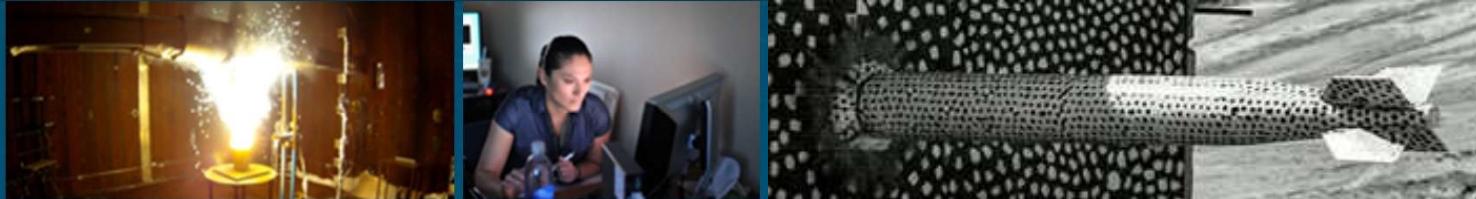


Georgia Institute of Technology
Center for Research into Novel Computing Hierarchies (CRNCH)
CRNCH Summit 2021

Reversible Computing as the Sustainable Path Forward for General Digital Computing



Friday, January 29th, 2021

Michael P. Frank, Center for Computing Research



Approved for public release, SAND2021-0849 C



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.

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 can 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 *recovers signal energies* and reuses them for multiple digital operations.
 - This is not a trivial change! It requires re-design and re-optimization of devices, circuits, architectures at multiple levels.
- It 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.
- 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.

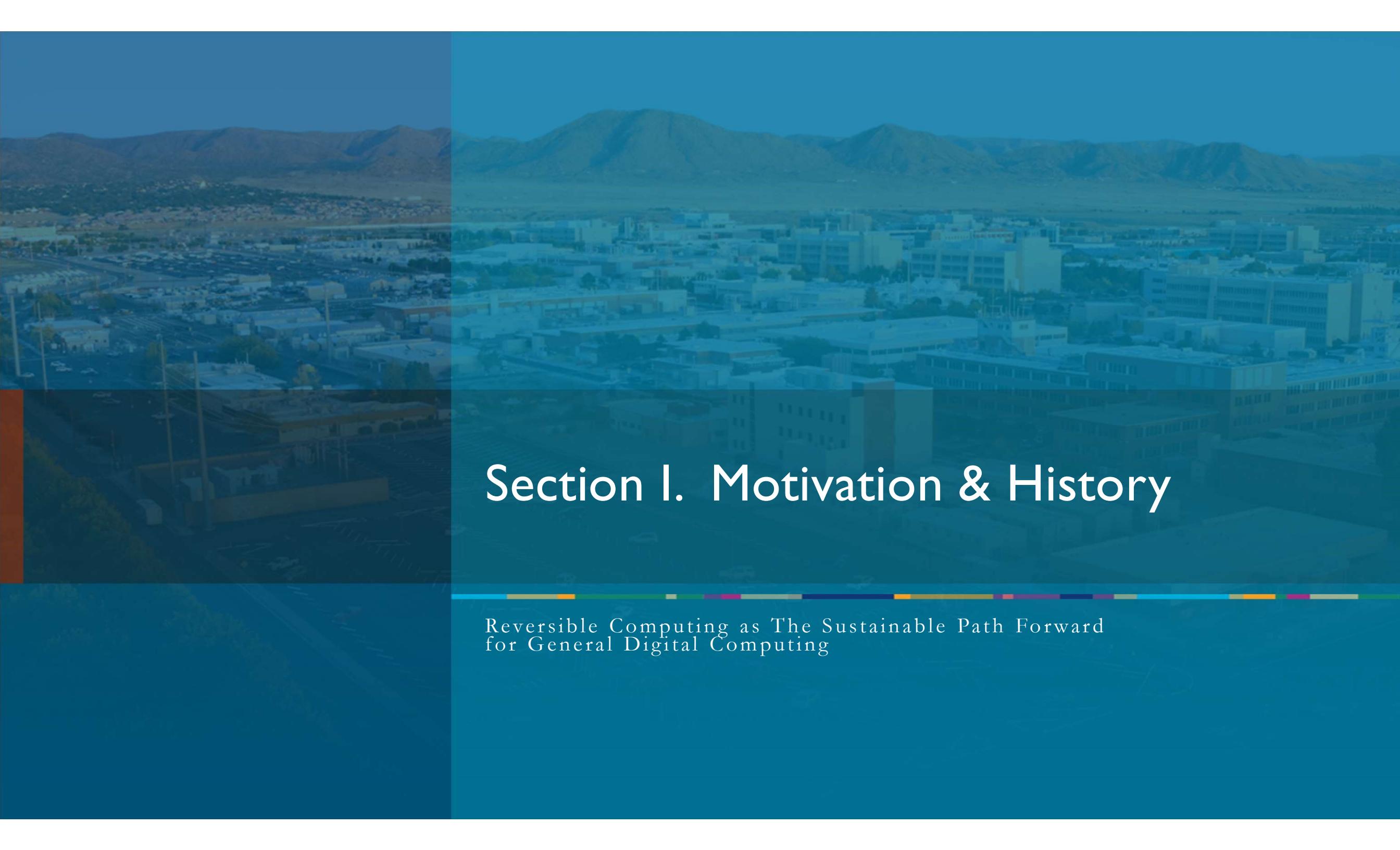


Outline of Talk

Reversible Computing as The Sustainable Path Forward 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.



The background of the slide features a wide-angle aerial photograph of a city, likely Salt Lake City, Utah, during the day. The city is nestled in a valley with several mountain ranges visible in the distance under a clear blue sky.

Section I. Motivation & History

Reversible Computing as The Sustainable Path Forward
for General Digital Computing

Motivation & Brief History

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 $E_{\text{diss}} \geq k_B T \ln 2$ energy to the environment for each bit's worth of known information that is lost.
- k_B is Boltzmann's constant $\approx 1.38 \times 10^{-23} \text{ J/K}$ = the natural logarithmic unit of entropy.
- NOTE:** T is *the temperature of the thermal environment into which the waste heat ends up getting ejected*.
- \therefore 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:

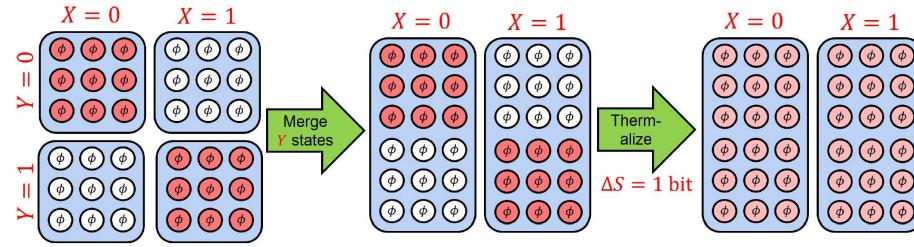
- RC means computing *without* oblivious erasure of known or correlated information.
 - In principle, energy dissipation per useful operation can be made *arbitrarily small* (can approach zero as technology improves).
 - \therefore Energy *efficiency* (operations per Joule) can theoretically approach *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:

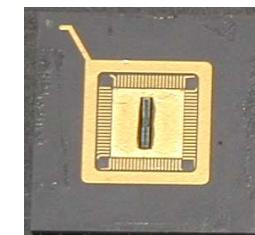
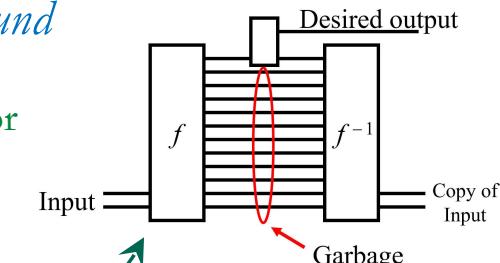
- RC was first shown *theoretically* coherent by Bennett, 1973 ([doi:10.1147/rd.176.0525](https://doi.org/10.1147/rd.176.0525)).
- First *engineering* implementation proposed by Likharev, 1977 ([doi:10.1109/TMAG.1977.1059351](https://doi.org/10.1109/TMAG.1977.1059351)).
- First fully-adiabatic sequential 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!

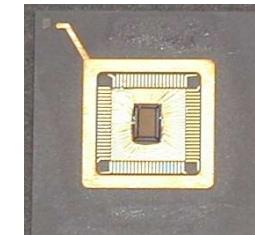
- I believe there is an opportunity right now to demonstrate some real breakthroughs.



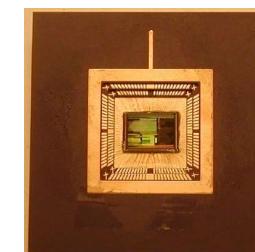
arXiv:
1901.
10327



Tick



FlatTop



Pendulum

Semiconductor Roadmap is Ending...

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

- This increases leakage, impairs practical device performance
- Thus, roadmap has minimum gate energy asymptoting to $\sim 2 \text{ eV}$

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

- Transistor width $10\text{-}20\times$ minimum width for fastest logic.
- Parasitic (junction, *etc.*) transistor capacitances ($\sim 2\times$).
- 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 ($\sim 2\times$).

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!

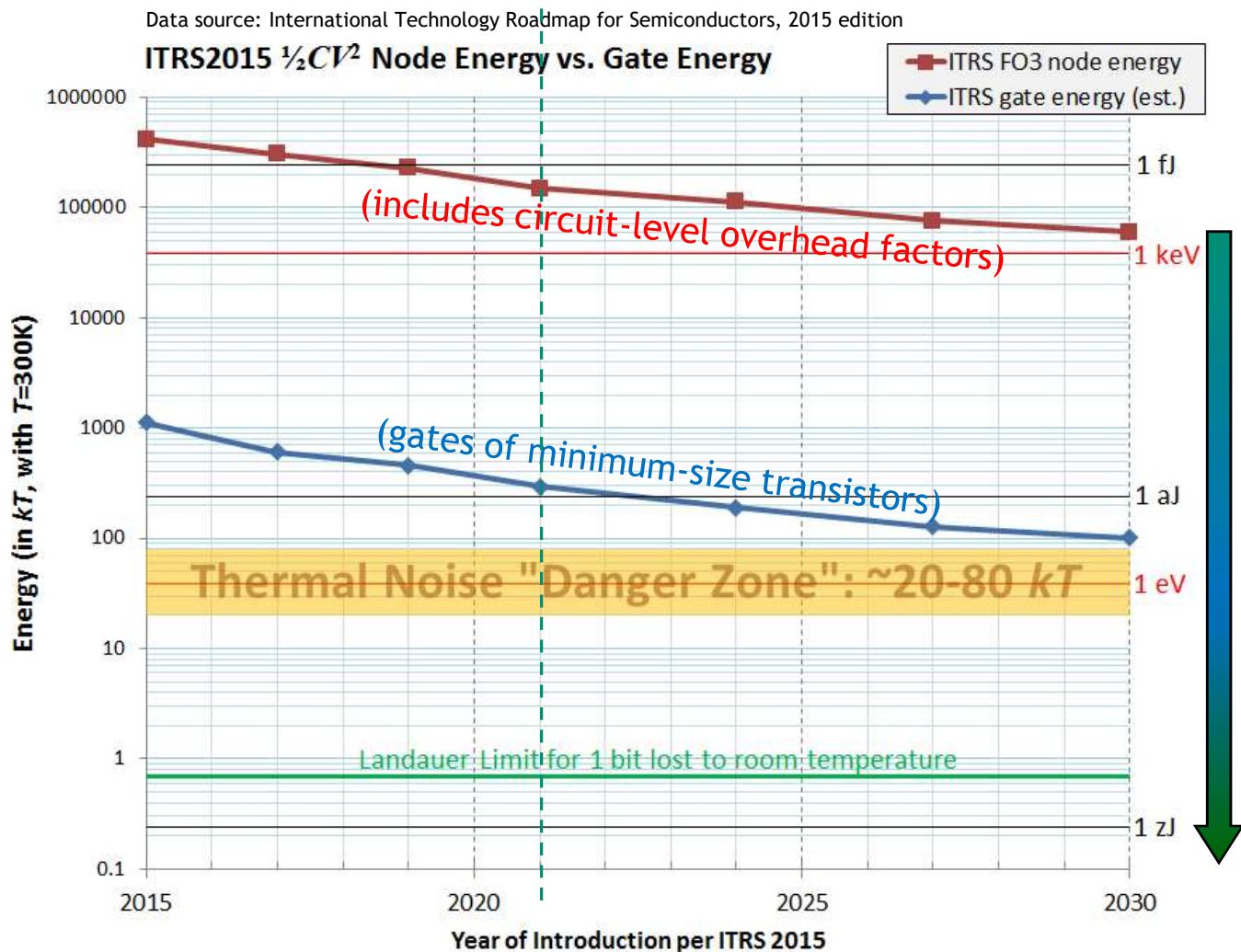
- $375\text{-}600\times$ (!) larger in ITRS'15.
- ∴ Practical bit energy for irreversible CMOS logic asymptotes to $\sim 1 \text{ keV}$!

Practical, real-world logic circuit designs can't just magically cross this $\sim 500\times$ 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 $\sim 1 \text{ keV}$ at the end of the CMOS roadmap, all the way down to $\ll kT$.

Why Reversible Computing Wins Despite Its Overheads!

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

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:

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

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.

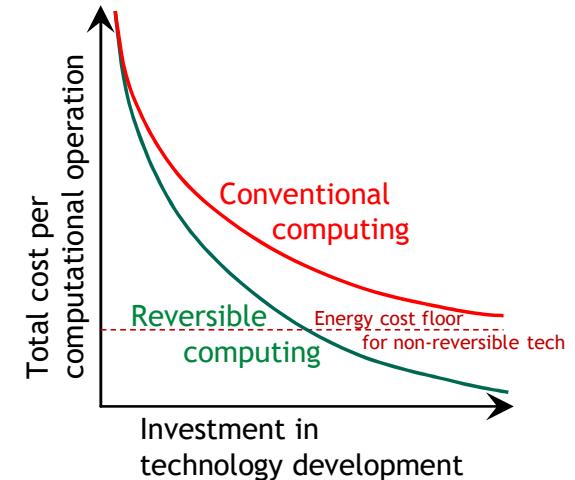
Two key observations:

- 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).

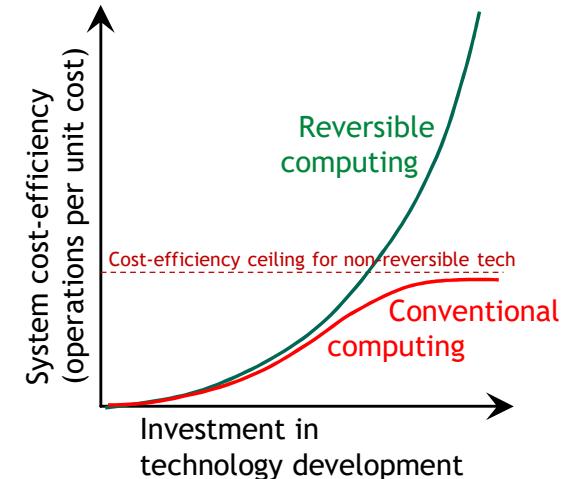
∴ Nothing prevents system-level cost efficiency of reversible machines from becoming *arbitrarily* larger than conventional ones, *even if* we have to scale t_{delay} and/or s_{elem} up as we scale $E_{\text{diss,op}}$ down!

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

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.

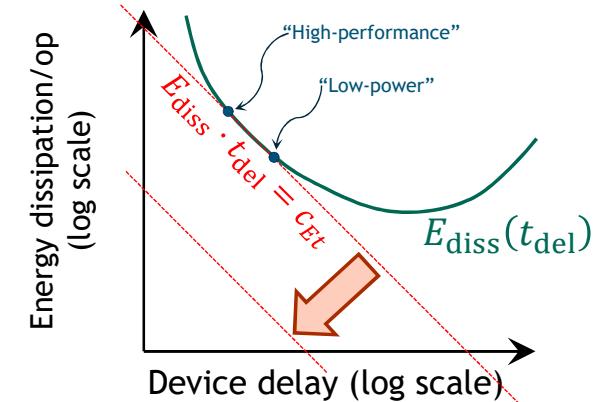
Can express total cost in terms of device parameters: $\$_{\text{tot}} = k_E E_{\text{diss}} + k_M t_{\text{del}}$

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:

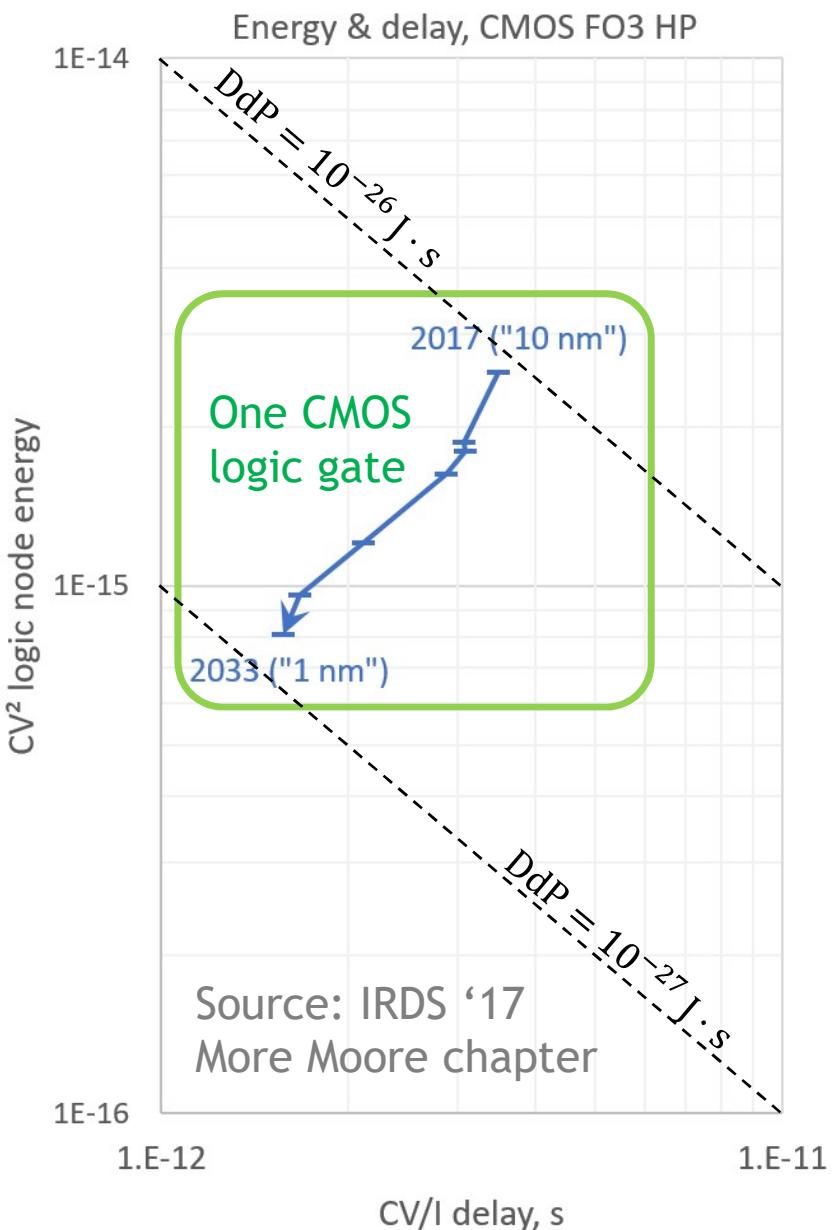
$$\eta_{Et} = \frac{1}{C_{Et}}$$

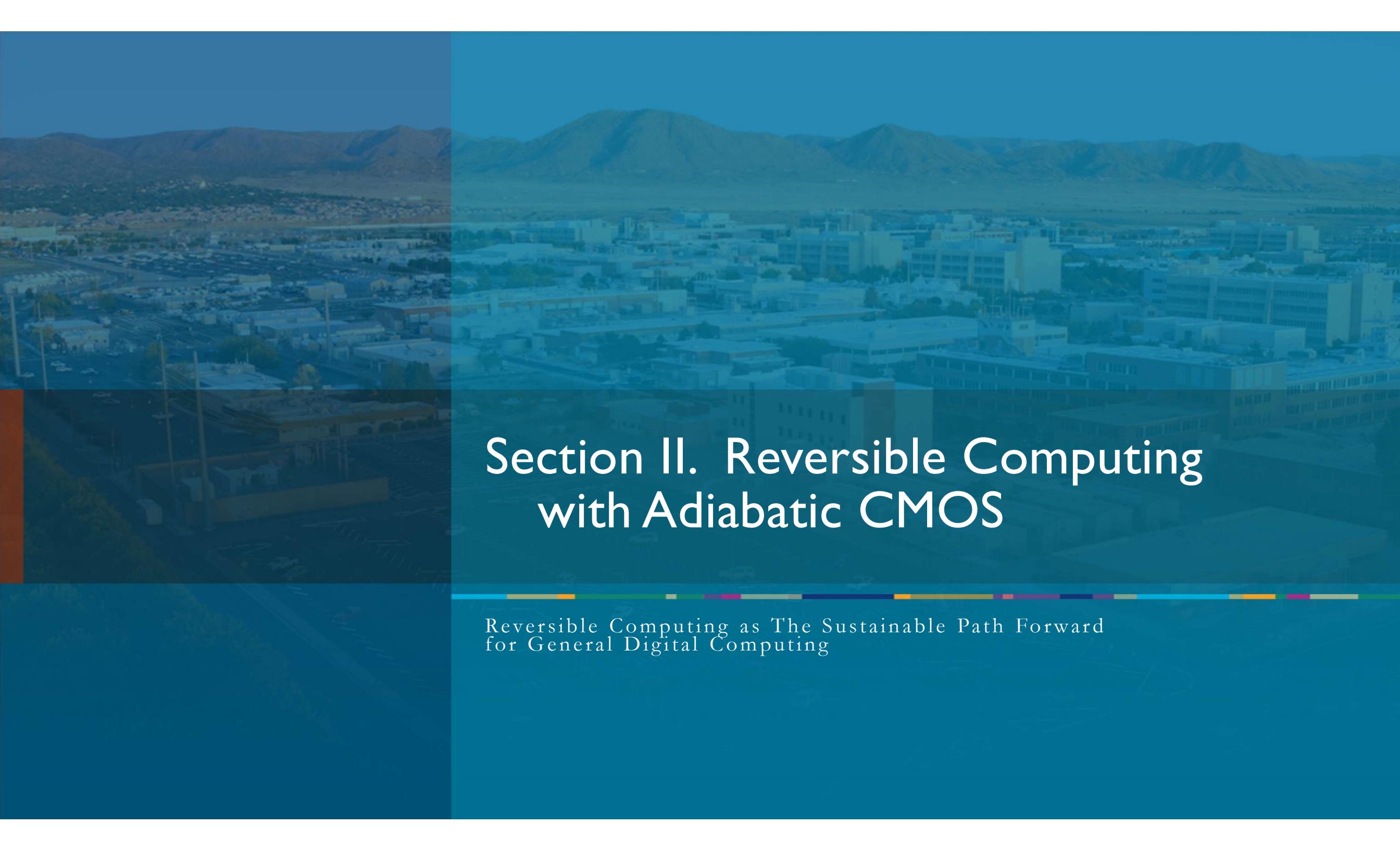


Existing Dissipation-Delay Products (DdP) —Non-reversible Semiconductor Circuits

Conventional (non-reversible) CMOS Technology:

- Recent roadmaps (e.g., IRDS '17) show Dissipation-delay Product (DdP) decreasing by only $<\sim 10\times$ 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 $\sim 1 \text{ eV} \approx 40 \text{ } kT$ 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.



The background of the slide features a wide-angle aerial photograph of a city, likely San Jose, California, viewed from above. In the foreground, there's a mix of industrial buildings, parking lots, and some green spaces. Behind the city, a range of mountains is visible under a clear blue sky.

Section II. Reversible Computing with Adiabatic CMOS

Reversible Computing as The Sustainable Path Forward
for General Digital Computing

Adiabatic Charging via MOSFETs



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

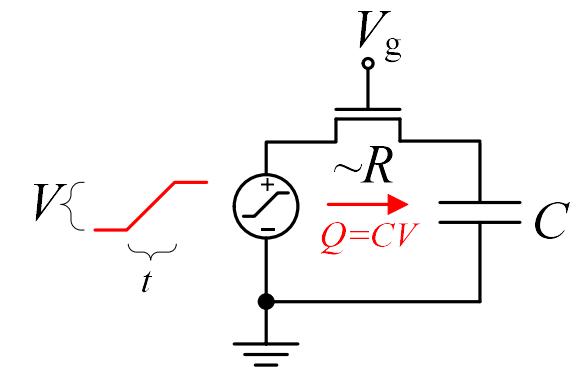
- Note that the load gets charged up *conditionally*, if the MOSFET is turned on (gate voltage $V_g \gtrsim V + V_t$) during ramp.
 - V_t is the transistor's threshold, typically $< \frac{1}{2}$ volt

Can discharge the load later using a similar ramp.

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

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

$$t \ll RC \Rightarrow E_{\text{diss}} \rightarrow \frac{1}{2}CV^2$$



Exact formula for linear ramps:

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

given speed fraction $s = RC/t$.

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

- 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](https://arxiv.org/abs/1806.10183) for the full GRC model.

Basic Requirements for Fully Adiabatic Operation



No diodes in charging paths!

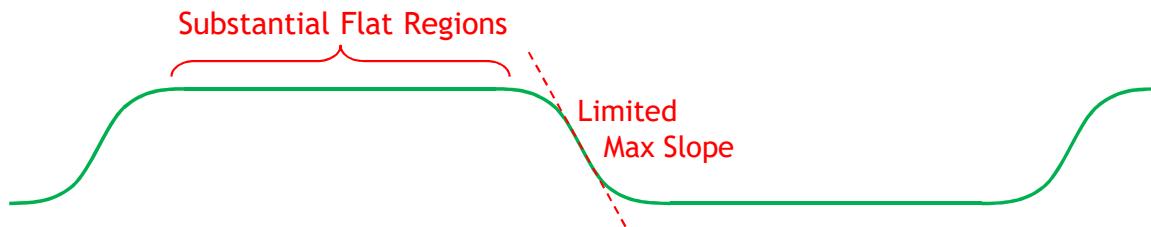
- All diodes have a built-in voltage drop for fundamental thermodynamic reasons.

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

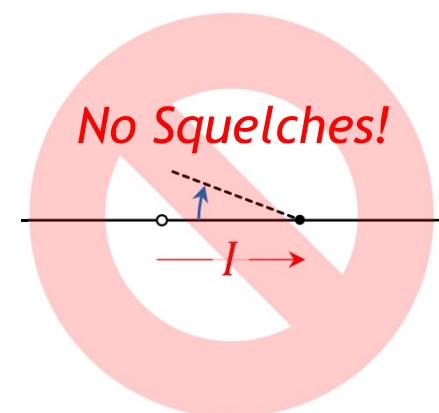
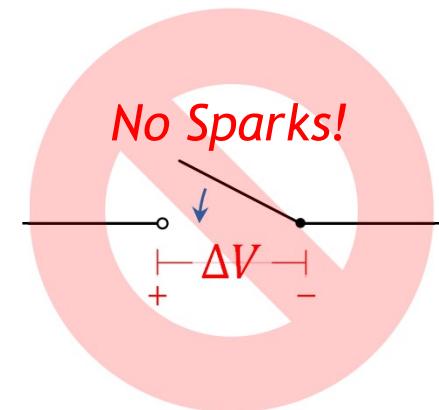
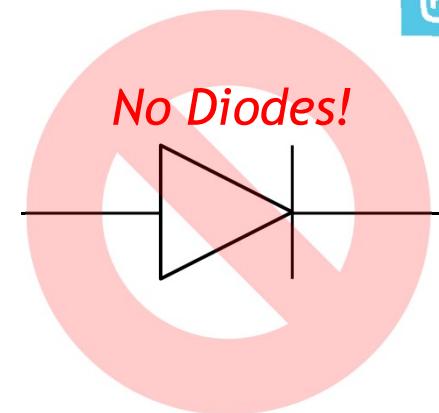
- Never turn on (close) a switch when there is a significant voltage difference $\Delta V \neq 0$ between its terminals.
 - Leads to a sudden, non-adiabatic flow of current.
 - More generally: No rapid voltage changes.
- Never turn off (open) a switch when there is a significant current flow $I \neq 0$ 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.



$$E_a = \xi_{tr} C_L V_{dd}^2 \frac{R C_L}{\tau_{tr}}$$





Review of 2LAL

2LAL is a simple variant of CRL, first described by M. Frank in lectures at the University of Florida in 2000.

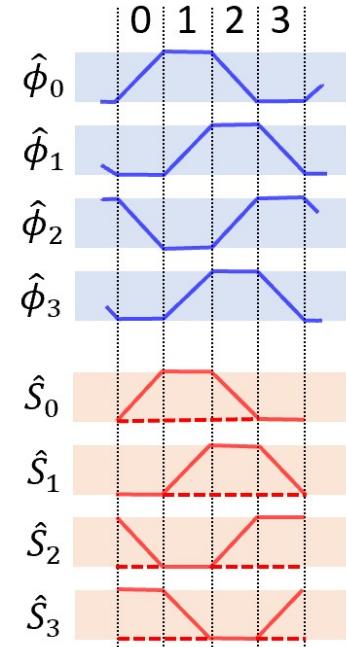
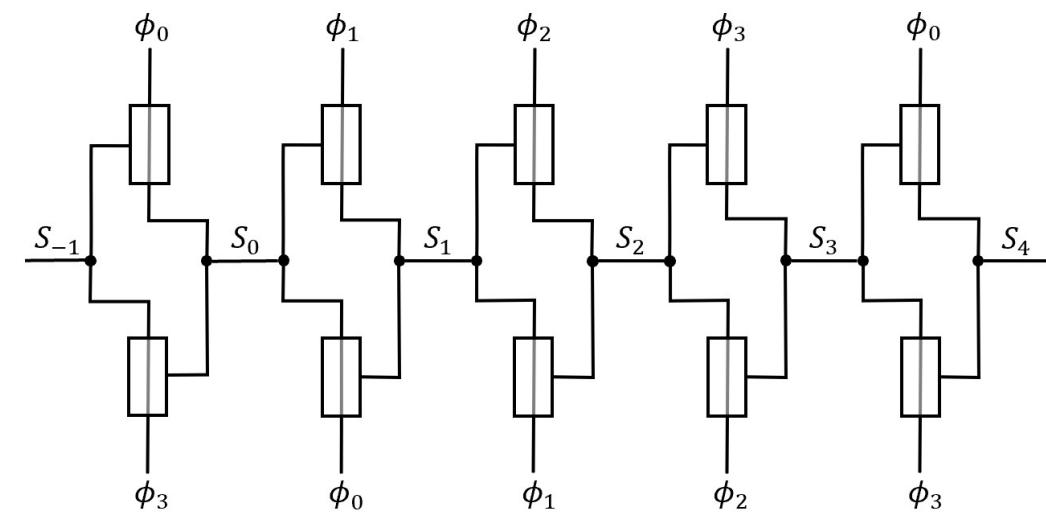
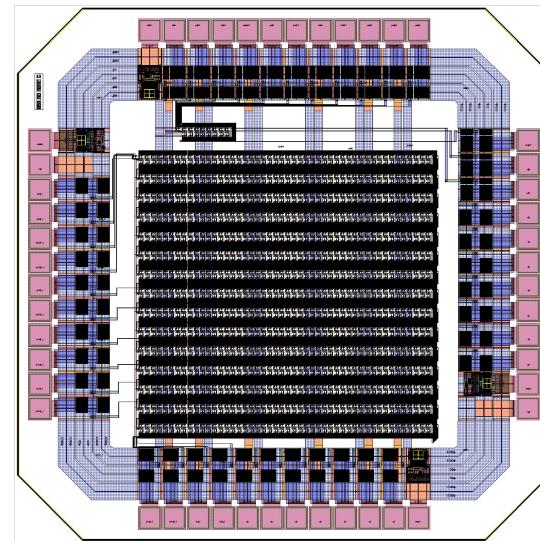
- Four clock phases, each active for one tick and inactive for one tick.
- A simple (one-symbol) shift register structure is shown.
- Series/parallel combinations of transmission gates can be used to do logic (not shown here).
 - 2LAL really only differs from CRL in terms of allowing more flexibility 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 $f = 1 \text{ MHz}$, $C_L = 10 \text{ fF}$:
 - Spectre simulation of MESA 350 nm, $W = 800 \text{ nm}$: $37 \text{ aJ} \approx 230 \text{ eV}$.
 - Spectre simulation of MESA 180 nm, $W = 480 \text{ nm}$: $6.9 \text{ aJ} \approx 43 \text{ eV}$. ← Comparable to a 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.
- $2 \times 2 \text{ mm}$ die.
- 8-stage & 720-stage shift registers.
- Goal: Verify function & dissipation.



S2LAL Reversible Pipeline Structure

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:

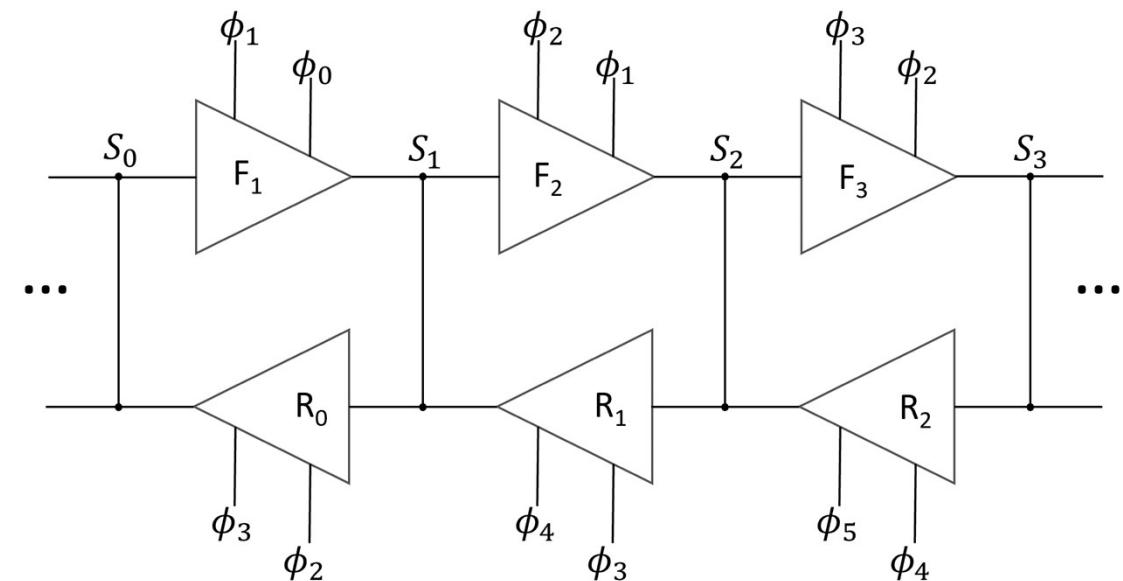
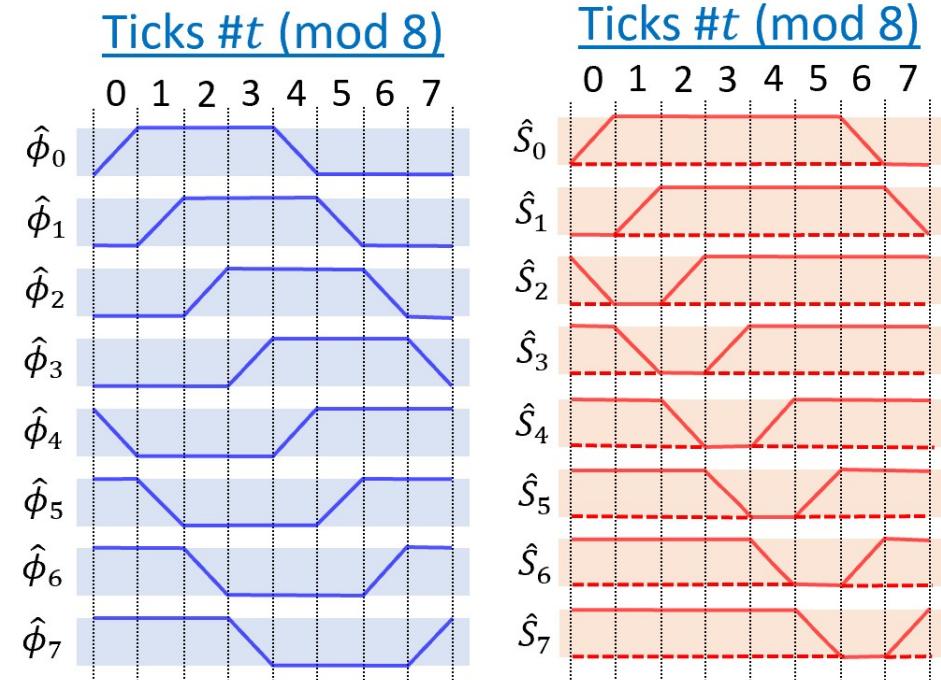
- Total cycle length = 8 ticks min.

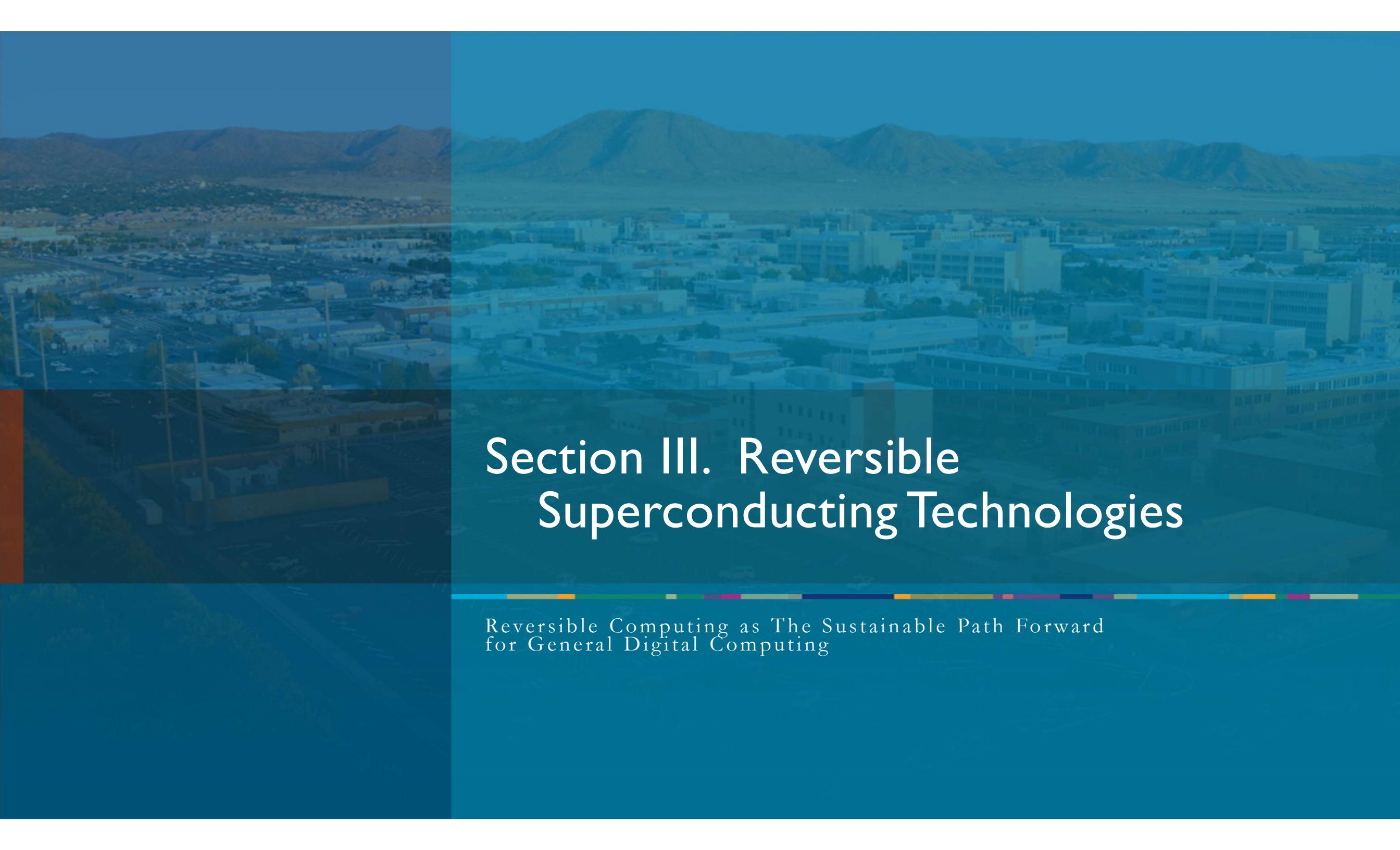
Note control of each signal S_i is handed off to forward stage F_i on ticks $\#i - 1$, and to reverse stage R_i on ticks $\#i + 3$.

- 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.



The background of the slide features a wide-angle aerial photograph of a city, likely Salt Lake City, Utah, during the day. The city is nestled in a valley with several mountain ranges visible in the distance under a clear blue sky.

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.

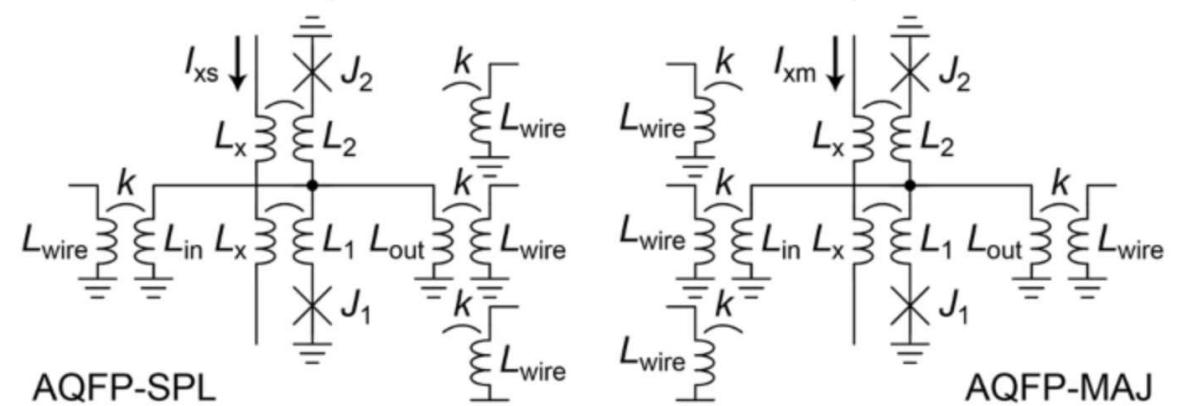
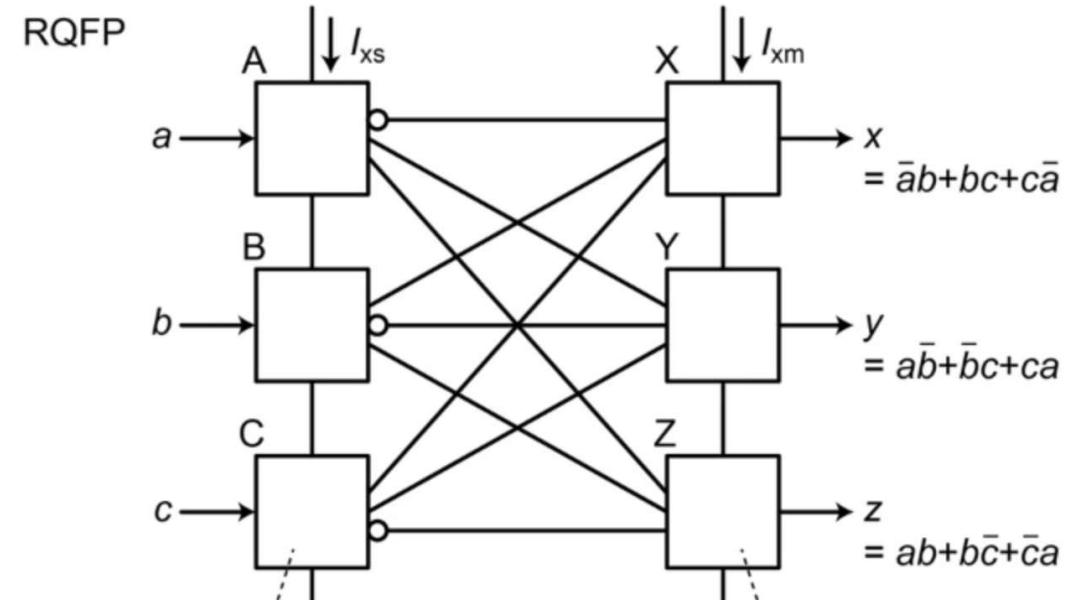
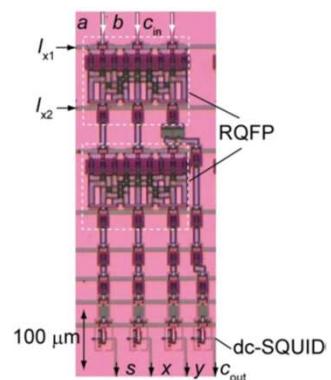
Most active group at present is Prof. Yoshikawa's group at Yokohama National University in Japan.

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 $T = 4K$, at speeds on the order of 10 MHz



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

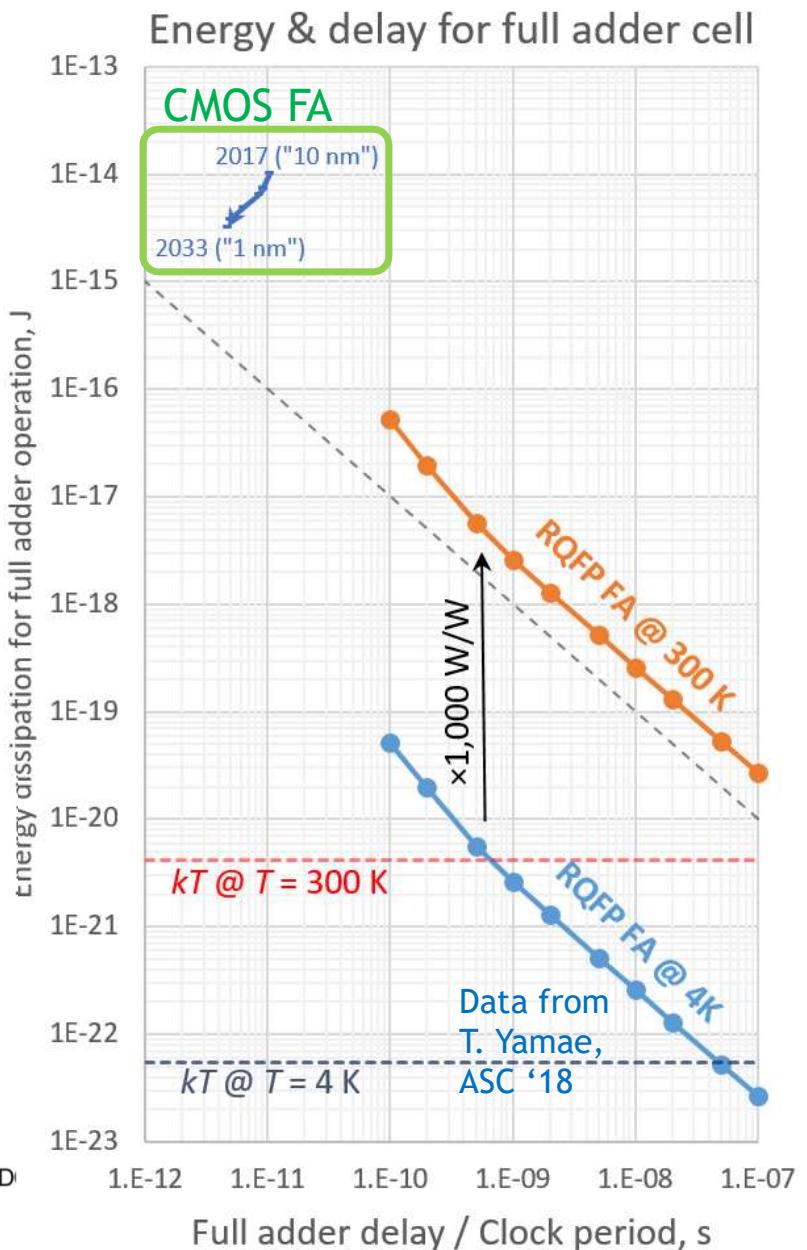
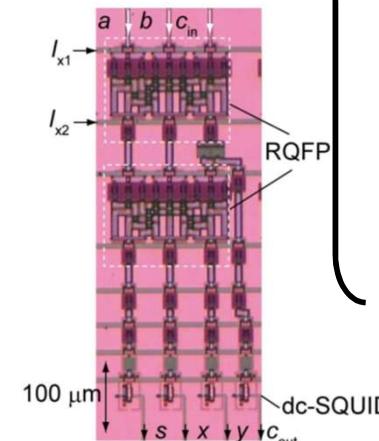
Reversible adiabatic superconductor logic:

- 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 $>1,000\times$ 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 than CMOS 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 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 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

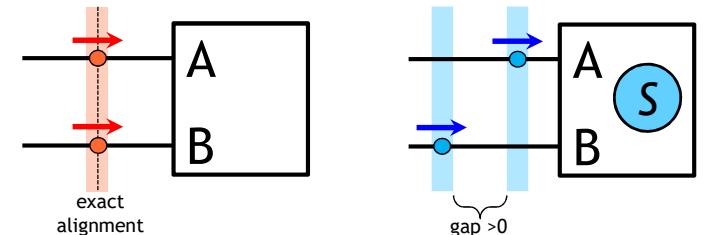
- Device’s dynamical trajectory then becomes *independent* of the precise (absolute *and* relative) pulse arrival times
 - As a result, timing uncertainty per logic stage can now accumulate only *linearly*, not exponentially!
 - Only relatively occasional re-synchronization will be needed
- For devices 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!

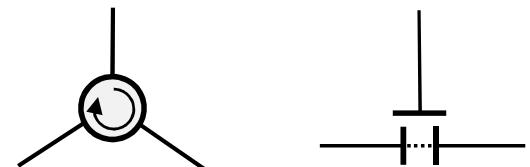
- Devices simply operate whenever data pulses arrive
- The operation energy is carried by the pulse itself
 - Most of the energy is preserved in outgoing pulses
 - Signal restoration can be carried out incrementally

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

(BARCS effort)

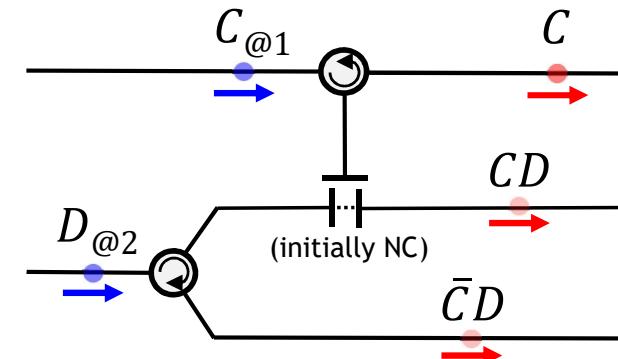


Synchronous Ballistic Asynchronous Ballistic



Rotary (Circulator) Toggled Barrier

Example BARC device functions



Example logic construction



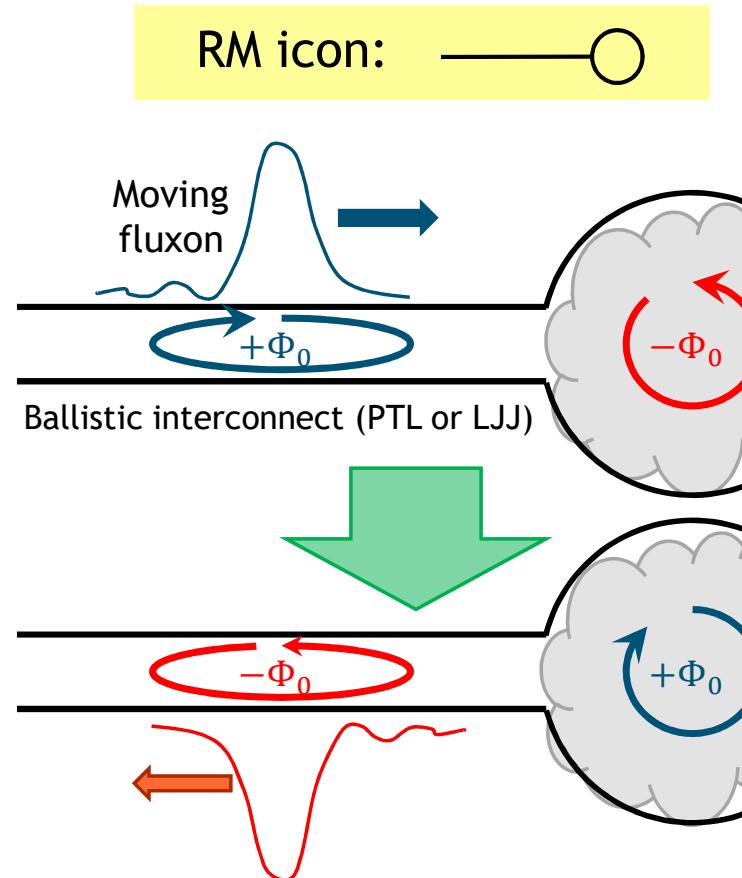
Simplest Fluxon-Based (bipolarized) BARC Function

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...



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 Transition Table

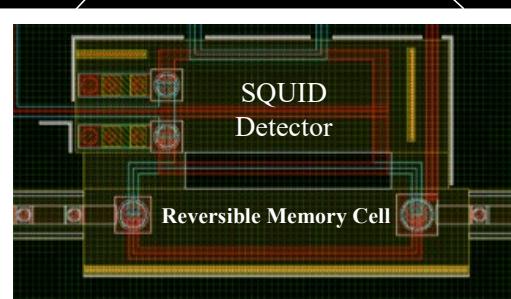
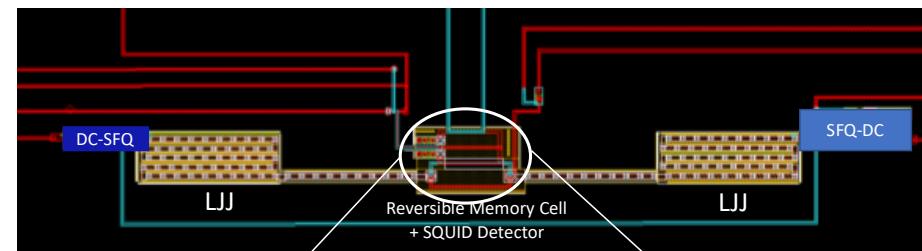
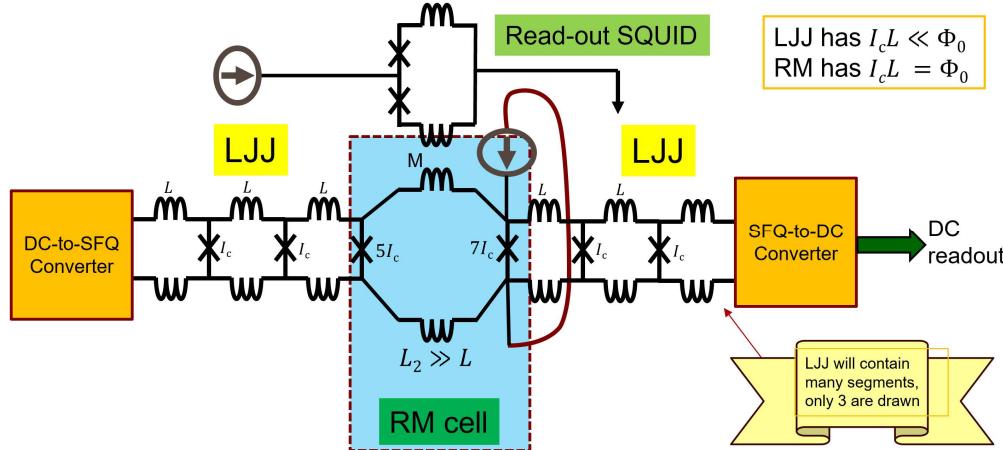
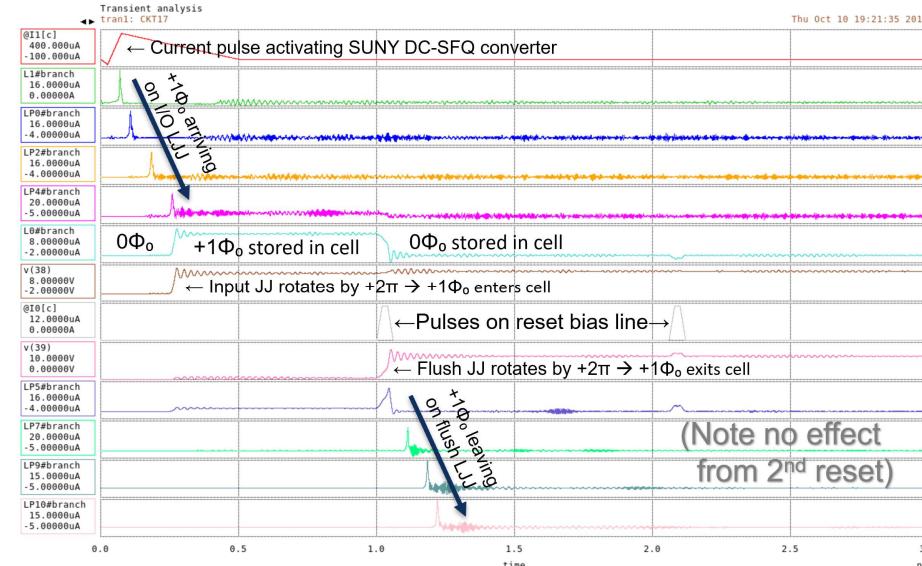
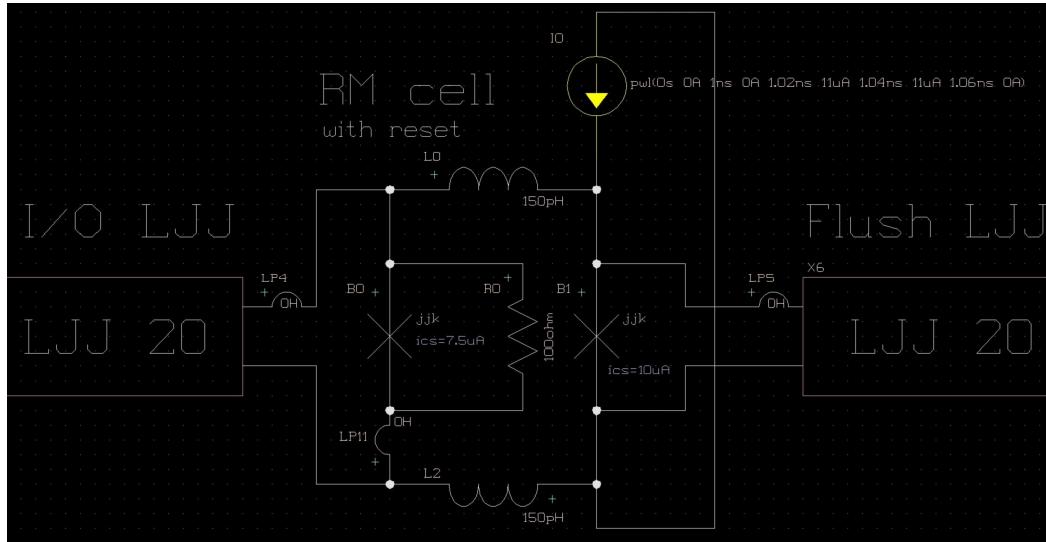
Input Syndrome		Output Syndrome
+1(+1)	→	(+1)+1
+1(-1)	→	(+1)-1
-1(+1)	→	(-1)+1
-1(-1)	→	(-1)-1



Resettable version of RM cell—Designed & Fabricated!

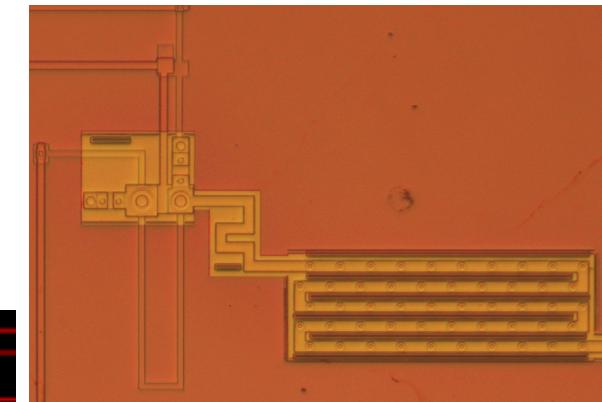
Apply current pulse of appropriate sign to flush the stored flux (the pulse here flushes out positive flux)

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

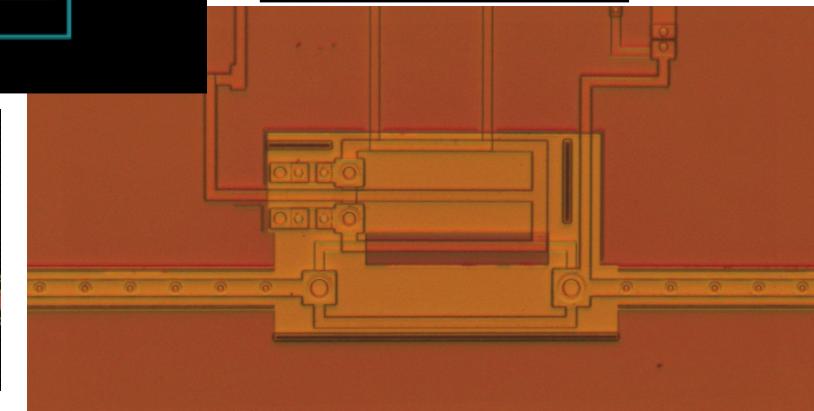


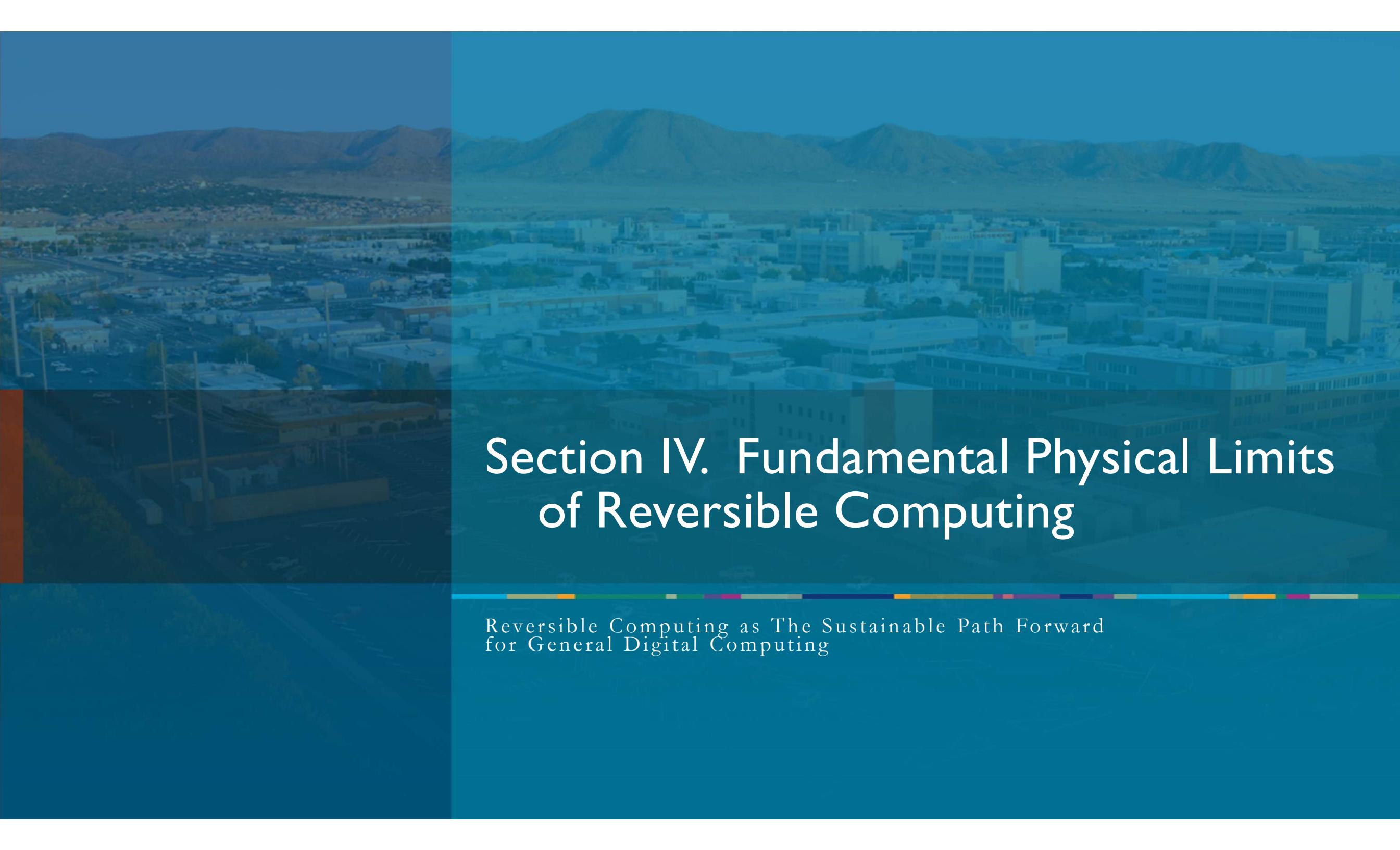
Fabrication at SeeQC
with support from ACI

DC-SFQ & LJJ



RM Cell & SQUID



The background of the slide features a wide-angle aerial photograph of a city, likely Salt Lake City, Utah, during the day. The city is nestled in a valley with several mountain ranges visible in the distance under a clear blue sky.

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](https://doi.org/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](https://arxiv.org/abs/2007.03605), [2011.04054](https://arxiv.org/abs/2011.04054), [2012.05655](https://arxiv.org/abs/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](https://doi.org/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

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?

J. Low Power Electron. Appl. **2018**, *8*(3), 30; <https://doi.org/10.3390/jlppea8030030>

Open Access Article

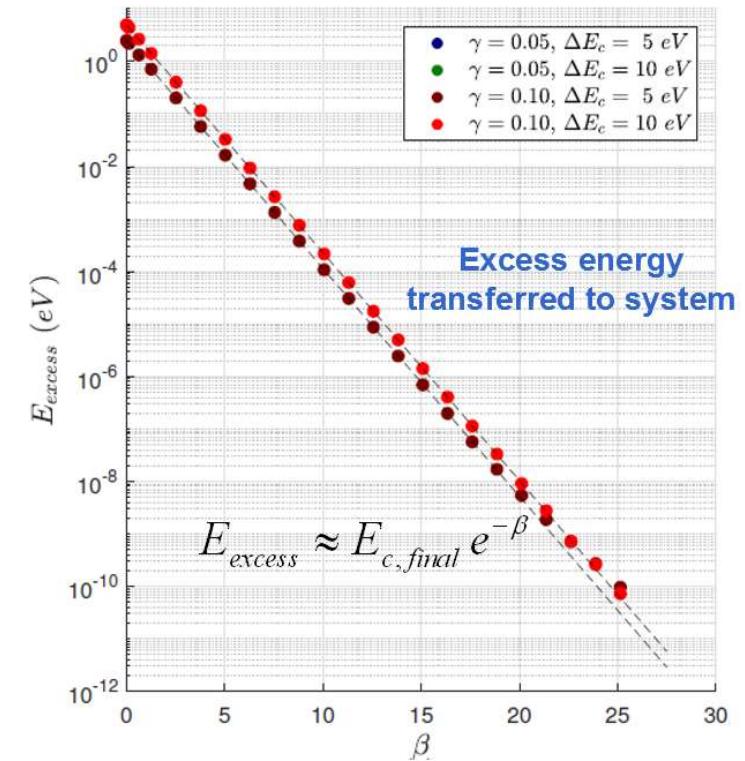
Exponentially Adiabatic Switching in Quantum-Dot Cellular Automata

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)



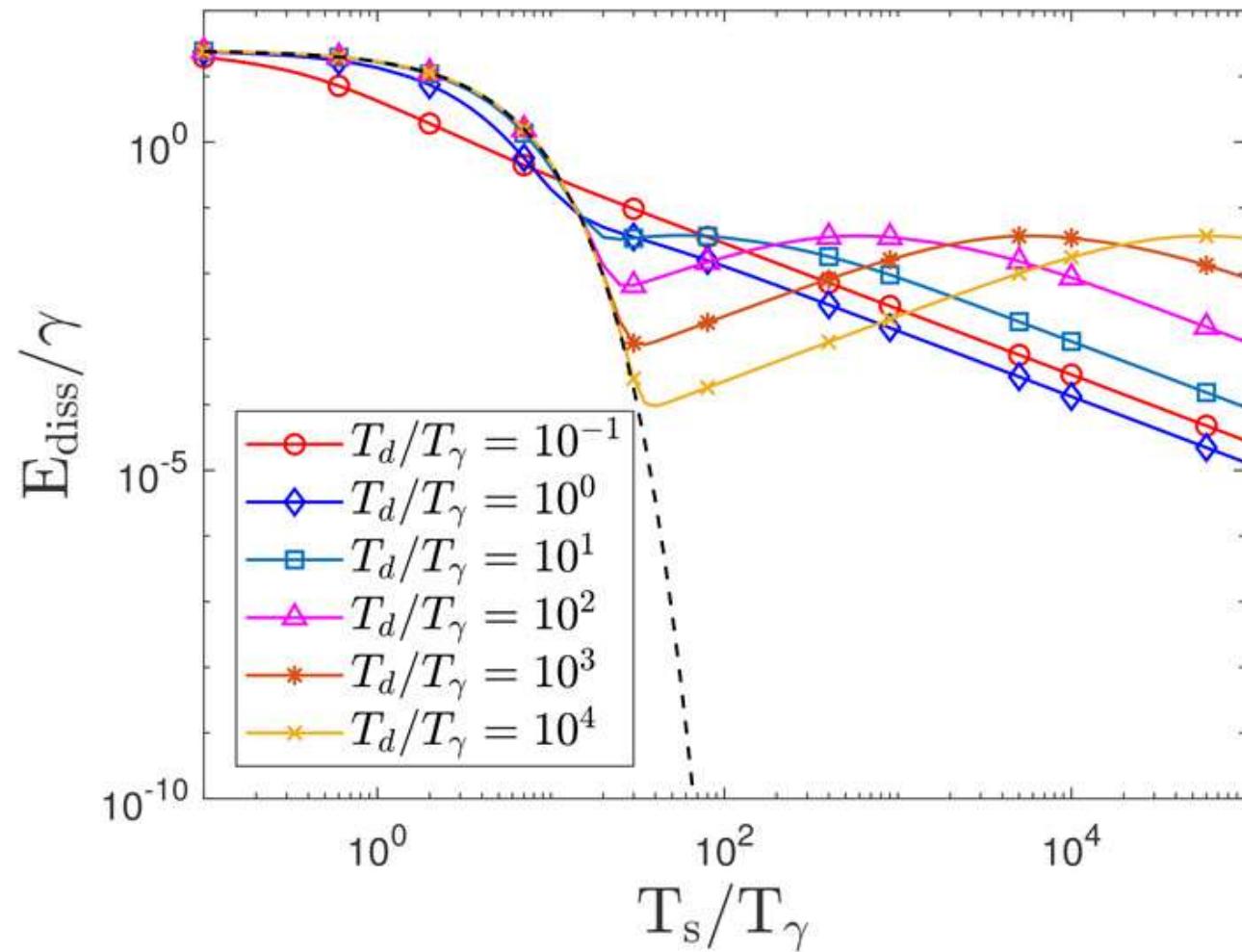


FIG. 10. Dissipated energy of an open system as a function of switching speed for different dissipation time constants. The dashed line is the excess energy of an isolated system. Here, the environmental temperature $k_B T / \gamma = 0.5$.

Published in: Subhash S. Pidaparthi; Craig S. Lent; *Journal of Applied Physics* **129**, 024304 (2021)

DOI: 10.1063/5.0033633

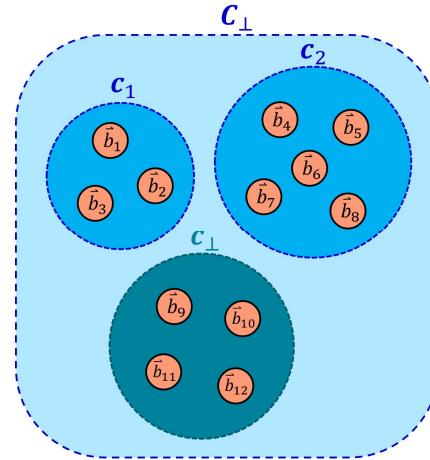
Copyright © 2021 Author(s). (Excerpted with permission.)

Fundamental Physics of Reversible Computing

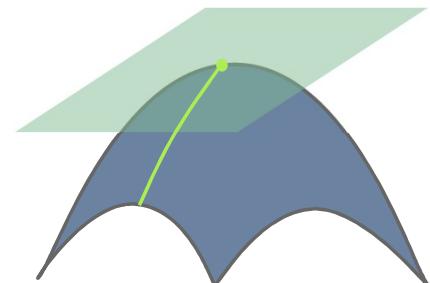
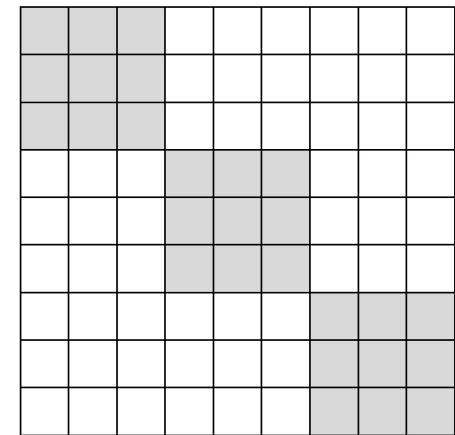
(Work with Karpur Shukla, Brown University)

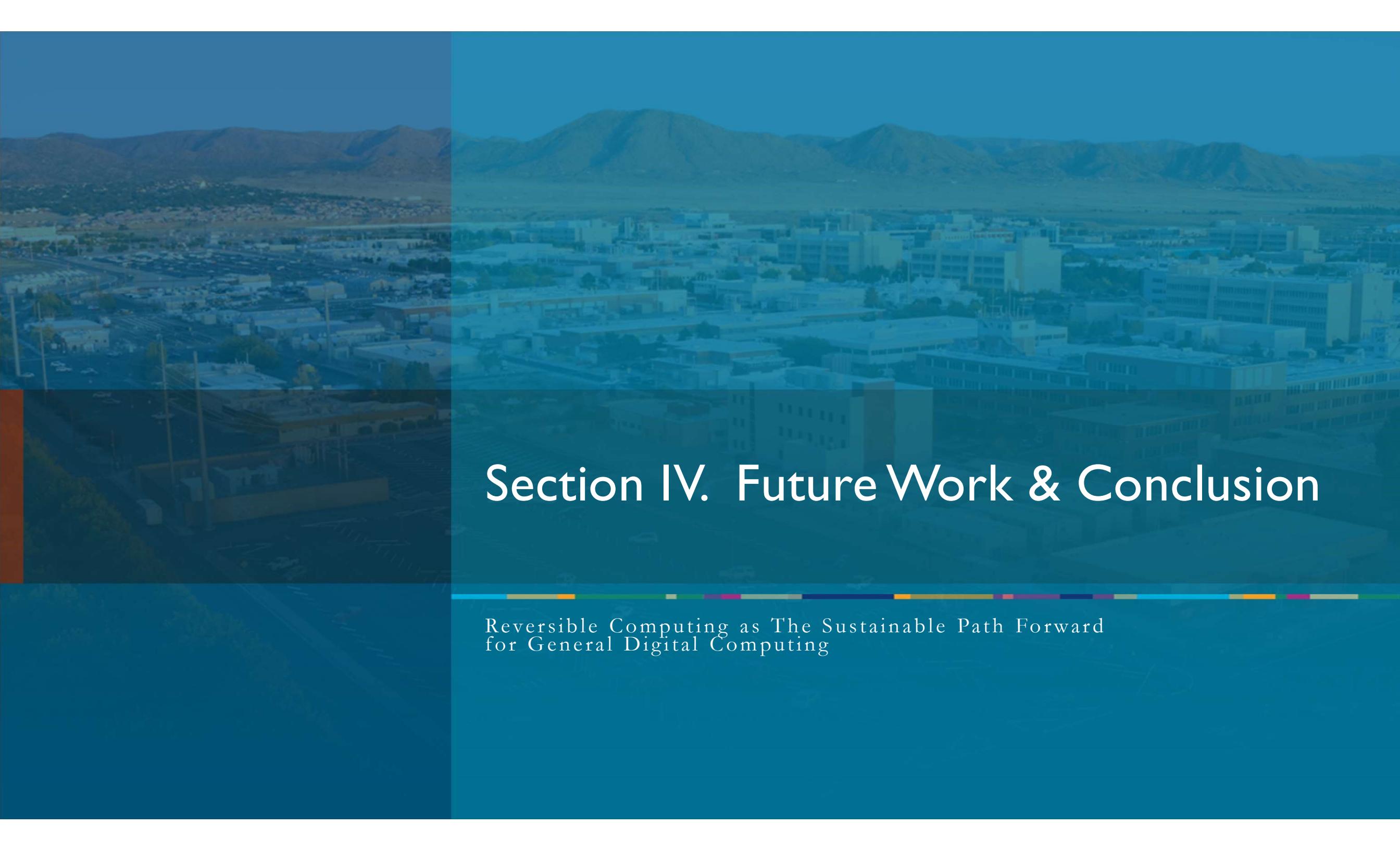


- Goals of this effort:
 - Look for fundamental physical limits of reversible computing
 - *E.g.*, minimum entropy production per operation as a function of delay, temperature, etc.
 - Identify ways to harness exotic quantum phenomena if needed to saturate the limits
- Steps completed so far:
 - Identification 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...
 - Uncertainty relations, dissipation and dissipation-delay product.



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



The background of the slide features a wide-angle aerial photograph of a city, likely Salt Lake City, Utah, during the day. The city is nestled in a valley with several mountain ranges visible in the distance under a clear blue sky.

Section IV. Future Work & Conclusion

Reversible Computing as The Sustainable Path Forward
for General Digital Computing

Assessment of Architectural Implications

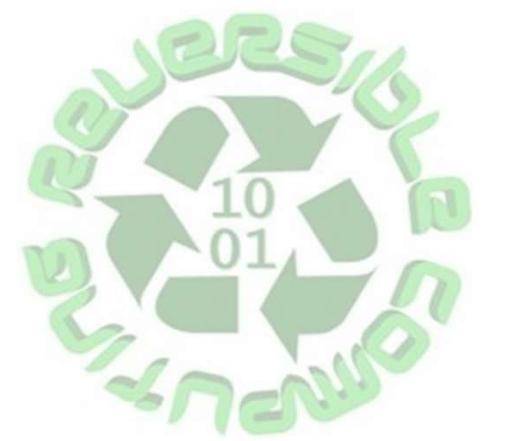
(Work with Tom Conte and Anirudh Jain, Georgia Tech)

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., $E_{\text{diss}}(t_d)$, 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, it will be a watershed moment for the future of technology, and civilization in general.

