# Fundamental Energy Limits and Reversible Computing Revisited

Michael P. Frank
Center for Computing Research
Sandia National Laboratories

Invited talk at the Center for Nonlinear Studies,
Los Alamos National Laboratory, April 17, 2017

# Abstract

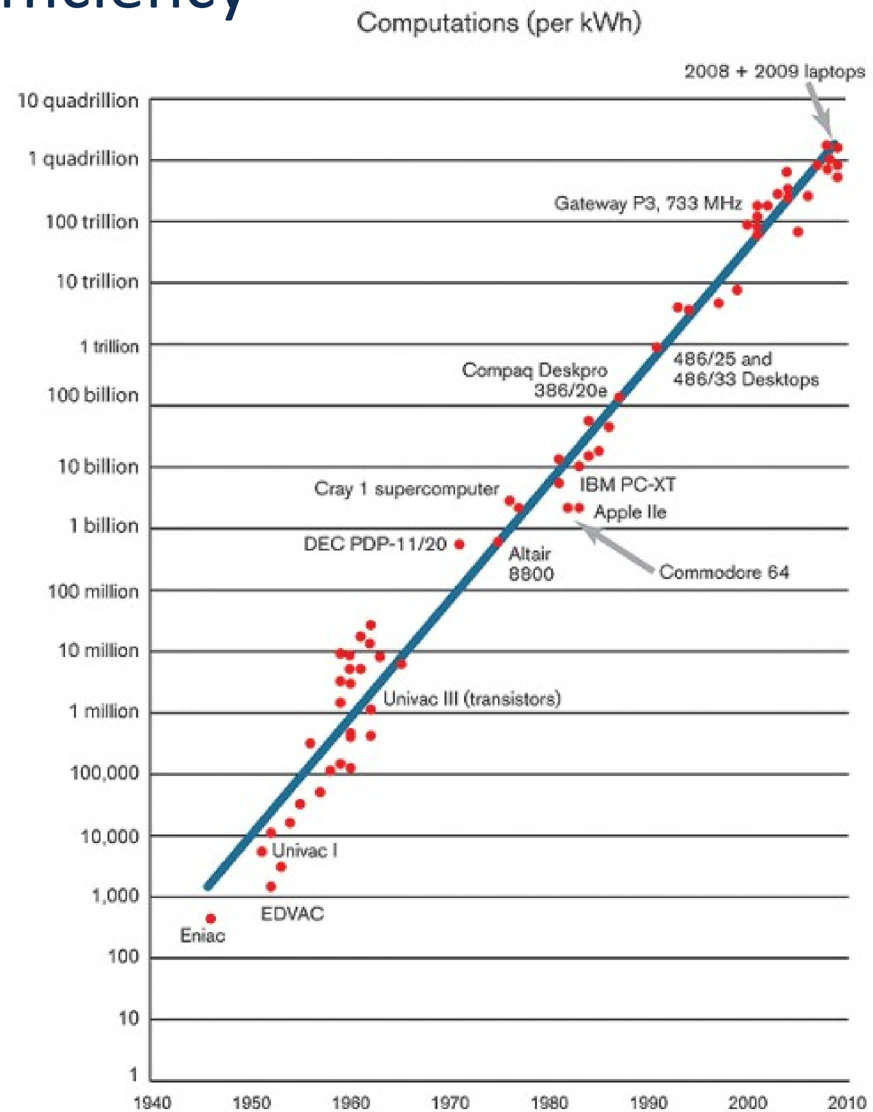
The fundamental thermodynamic limits of conventional computation are near enough to be an area of concern when contemplating future computing technologies. Several thermodynamic arguments imply lower limits on the energy required for computation, when conventionally construed. However, several of the known limits may be circumvented by using unconventional computing paradigms. Thermal noise limits on signal energies can potentially be circumvented in appropriately designed chaotic systems with sub-unity signal-to-noise ratios. And, limits on energy dissipation due to Landauer's Principle can be circumvented using reversible computing. We review some recent work in these areas, including a new general theoretical framework for reversible computing, and a framework for asynchronous reversible computation.

# Talk Outline

- Trends in Computational Energy Efficiency
- Fundamental Energy Limits on Computation
  - Thermodynamic limits of conventional computation
  - Quantum-mechanical limits on all computation
- Transcending the Thermodynamic Limits
  with Unconventional Computing Paradigms
  - (Generalized) Reversible Computing
  - Asynchronous Reversible Computing
  - Computation in Chaotic Systems?
- Conclusion

# The Power-Performance Trend
## and the importance of energy efficiency

- *Any* system (at any scale) scoped to have a fixed cost-of-ownership over its operational lifetime *must* implicitly carry some associated maximum budget for all energy-related costs.
  - These costs include things like:
    - In mobile devices, cost of batteries and inconvenience to user of charging
    - kWhr electricity costs for desktop owners
    - Cost to build and operate high-capacity machine room/datacenter cooling systems
    - Cost to build or lease a nearby power plant if required to supply an exascale machine
- We can't expect the cost of energy to ever decrease by orders of magnitude.
  - Essentially, energy *is* "nature's currency."
- Thus, fundamentally, *increasing affordable performance requires increasing computational energy efficiency*. (Useful ops done/Joule.)
  - And this has, indeed, been the historical trend, for >50 years.



(MIT Technology Review, Apr. 2012)

# Thermodynamic Limits on Computing

Sandia National Laboratories

- Landauer Limit, a.k.a. Landauer's Principle:
  - Rigorous theorem of mathematical physics!
  - Computational operations that eject entropy $\Delta S$ from the digital state imply energy dissipation $E_{diss} \geq T\Delta S$ to an environment at temp. $T$.
    - Special case: Erasing a uniformly-distributed bit ($\Delta S = k \ln 2$)
      - Energy dissipation $E_{diss} \geq kT \ln 2$
  - Landauer's Principle limits the number of *conventional* irreversible operations that can be done with a given total energy dissipation
    - However, (as we'll see) reversible operations circumvent the Landauer limit
- Thermal Noise Limit on Signal Energies:
  - Informal conventional wisdom…
    - Has never been formally stated and rigorously proven in any general way
  - Roughly stated (typically) as follows:
    - "Computing reliably with a probability of error $p_{err} = 1/R$ requires signal energies (or energy barriers) of magnitude $E_{sig} \geq kT \ln R$"
      - Informal argument based on the Boltzmann distribution
  - Note: Signal energy (or barrier height) need not be *dissipated*
  - Also: Conventional wisdom may be wrong! (See "Chaotic Logic" later)

# Information Loss = Entropy Increase

- All <u>fundamental</u> physical dynamics is (microscopically) *reversible*.
  - Any Hamiltonian dynamical system:
    - Let the time increments $\delta t$ be negative → Time-evolution runs in reverse.
  - Quantum mechanical time-evolution (generalized Schrödinger equation):
    - Any two quantum states that are initially mutually distinguishable (orthogonal) will always remain so, under any unitary time-evolution operator, $U(t) = \mathrm{e}^{-iHt/\hbar}$.
- ∴ *Detailed physical information can never, ever be destroyed!*
  - Only reversibly transformed, in place (locally)!
    - At most, we can only *lose track* (from a modeling perspective) of the (always-still-microscopically-reversible) transformations that have occurred.
      - Uncertainty increase → Effective randomization of the detailed state
  - If this were not true, the 2nd Law of Thermodynamics would not hold!
    - Effectively, entropy is simply that portion of the total physical information that happens to have already been randomized/scrambled beyond any hope of practically transforming it back into its original form.
      - ∴ If information could be destroyed, then entropy could simply vanish
- *To "irreversibly lose information"* <u>means</u> for that information to be (reversibly) transformed in any way that we cannot <u>*practically*</u> undo.
  - It's "lost" in the sense that its original form cannot be <u>practically</u> recovered.
  - "Irreversible information loss" <u>is exactly the same thing as</u> "entropy increase."

# Landauer's Principle—
## A Simplified Statement:

- For each bit's worth of local information that is irreversibly lost from (*e.g.*, obliviously "erased" by , or "destructively overwritten" by) any computational device encompassed by an external thermal environment at temperature $T$, no less than an amount

$$E_{\text{diss}} = k_B T \ln 2$$

  of free energy ("Landauer's limit") must eventually be dissipated as heat added to that thermal environment.
  - This is easily proven, as a theorem of applied mathematical physics.
- *Approachability hypothesis:*
  - Landauer's bound may be approached arbitrarily closely in a suitably-designed family of realistically-constructible physical mechanisms.
    - Abstract physical procedures described in the literature support this.

# Landauer's Principle—
## A Correct *General* Formulation:

- Consider any computational device $D$ that is designed to transform initial logical states $s_\mathrm{I} \in S_\mathrm{I} = \{s_{\mathrm{I}1}, s_{\mathrm{I}2}, \dots, s_{\mathrm{I}n}\}$ to final logical states $s_\mathrm{F} \in S_\mathrm{F} = \{s_{\mathrm{F}1}, s_{\mathrm{F}2}, \dots, s_{\mathrm{F}m}\}$ according to some (in general probabilistic) transition rule, $r_i(j) = \Pr\left[s_\mathrm{F} = s_{\mathrm{F}j} \middle| s_\mathrm{I} = s_{\mathrm{I}i}\right]$.
  - Now consider any given probability distribution over initial states, $p_\mathrm{I}(i) = \Pr[s_\mathrm{I} = s_{\mathrm{I}i}]$, defining a given statistical scenario in which $D$ is to be operated. (An "*operation context.*")
    - The entropy $H[p_\mathrm{I}]$ of this initial state distribution is:

$$H[p_\mathrm{I}] = \sum_{i=1}^{n} p_\mathrm{I}(i) \ln \frac{1}{p_\mathrm{I}(i)}.$$

    - And, after $D$ has operated, we can derive, from $p_\mathrm{I}$ and $r_i(j)$, the final state distribution $p_F$, which is

$$p_\mathrm{F}(j) = \Pr\left[s_\mathrm{F} = s_{\mathrm{F}j}\right] = \sum_{i=1}^{n} p_\mathrm{I}(i) \cdot r_i(j).$$

    - And the entropy $H[p_\mathrm{F}]$of the final state distribution is:

$$H[p_\mathrm{F}] = \sum_{j=1}^{m} p_\mathrm{F}(j) \ln \frac{1}{p_\mathrm{F}(j)}.$$

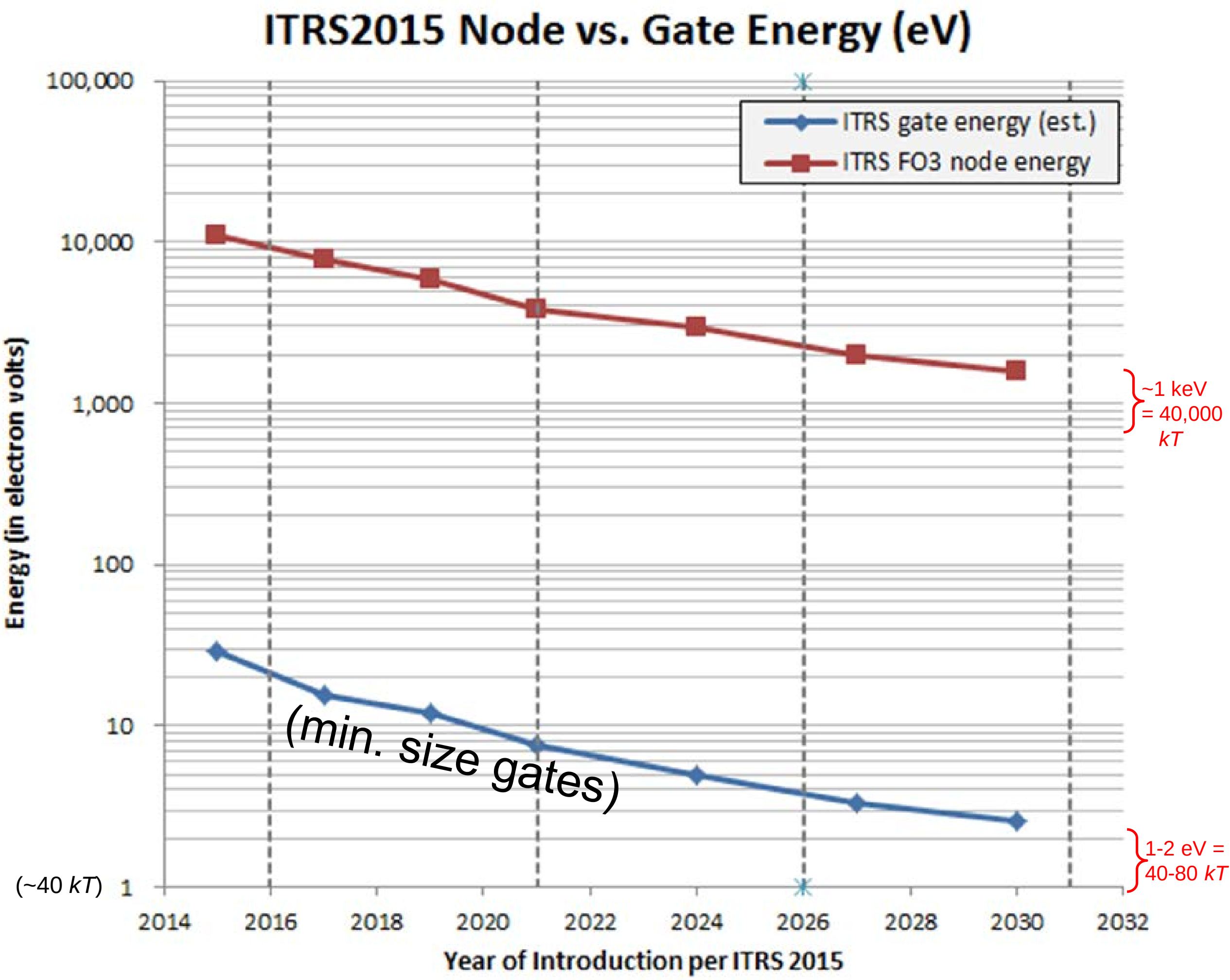
    - Then, the minimum entropy ejected from the device $D$ as a side-effect of its operation in context $p_\mathrm{I}$ must be:

$$\Delta H_D(p_\mathrm{I}) = H[p_\mathrm{I}] - H[p_\mathrm{F}],$$

      since total entropy cannot decrease (by fundamental reversibility/the 2nd law of thermodynamics).
- Therefore, device $D$, when operated in a statistical context $p_\mathrm{I}$, necessarily loses an amount of information (*i.e.*, ejects an amount of entropy) $\Delta H_D(p_\mathrm{I})$.
  - Suppose this entropy eventually ends up in some external thermal reservoir at temperature $T$.
  - Then, by the thermodynamic definition of temperature, we must add heat $\Delta Q = T\Delta H_D(p_\mathrm{I})$ to the reservoir.

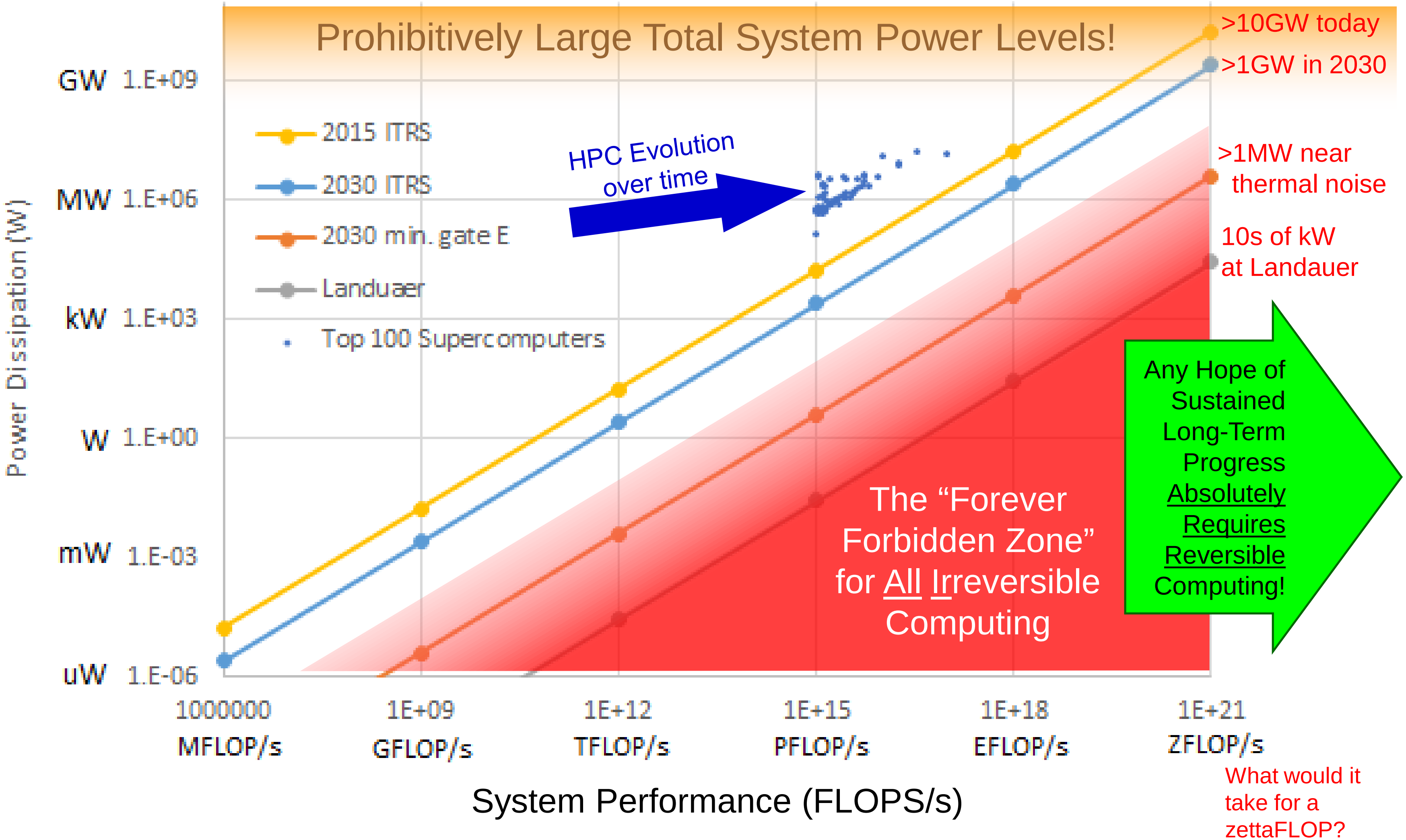
# Energy limits for conventional technology are not far away!

- Energy of min.-width FET gates affects channel fluctuations < ~1-2 eV
  - Impact on leakage
- Real gates are often wider (~ 20 × min.)
  - Also there is fanout, wire capacitance, etc.
- Note: ITRS is aware of thermal noise issue, and so has min. gate energy asymptoting to ~2 eV
  - Node energy *follows*, asymptoting to ~1 keV
- Practical circuit architectures can't just magically cross this gap!
  - ∴ Fundamental thermal limits translate to much *larger* practical limits!



# Implications for FLOPS & power

Note: The limits suggested by the diagonal lines do not even include power for interconnects, memory, or cooling!



# Quantum-Mechanical Energy Limits

- Quantum mechanics is not known to directly limit the energy that must be *dissipated* to carry out a computation,
  - But, it does appear to limit the energy that must be *invested* in a computation to attain a given performance
- The Margolus-Levitin bound (1996)
  - The maximum rate at which a system may transition between orthogonal states is given by $\nu_\perp \leq 4E/h$. (Example: 1 eV → ~1 PHz)
- Generalized in Frank 2005, "On the Interpretation of Energy as the Rate of Quantum Computation"
  - For any system, its Hamiltonian energy is exactly the rate at which it "exerts computational effort" by several measures
    - We can characterize a minimum effort or *difficulty* for given operations
  - Thus, any given computation requires a certain minimum Hamiltonian action (energy invested × time) to carry out.
- Note, however, as with the thermal noise limit, the energy *invested* in the computation need not be *dissipated*…
  - For maximizing total computation carried out given fixed free-energy resources, the energy dissipated and the Landauer bound are essential

# Transcending the Energy Limits

- We can transcend the traditional thermodynamic limits of computing using new computing paradigms:
  - ***Reversible Computing*** – Absolutely *required* to reuse signal energies and avoid Landauer's limit
    - ***Generalized Reversible Computing*** – Clarifies the precise requirements to avoid Landauer's limit. More general concept of logical reversibility
    - ***Asynchronous Reversible Computing*** – A ballistic computing scheme that avoids the clocking overheads of synchronous adiabatic approaches.
  - ***Computing with Chaos*** – A possible direction to reduce signal energies
    - Can also be viewed as a special case of reversible computing
  - ***Quantum Computing*** – Not the focus of this talk. About finding more efficient algorithms for some problems using coherent trajectories.
    - Doesn't address practical energy efficiency of general-purpose computation.

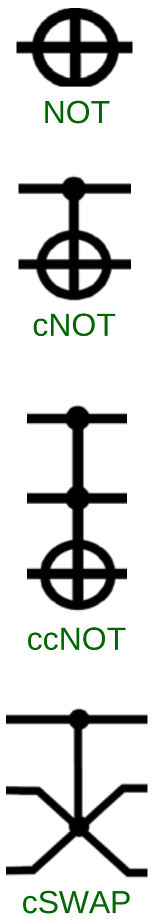
# Enter Reversible Computing…

- **Problem:** Landauer's Principle teaches us that losing computational information (merging computational states) implies unavoidable energy dissipation.
- **Solution:** *Compute without losing information!*
  - Don't ever try to erase bits / merge two distinct computational states.
  - Instead, *transform* computational states one-to-one into new states.
    - No decrease in computational entropy
    - No need to eject computational entropy to the physical state
  - This is what we mean by reversible computing.
- Bennett (1973) showed that reversible computations can still compute any function…
  - To get rid of temporary results that are no longer needed, you can always reversibly *decompute* them
    - instead of erasing/overwriting them

# Unconditionally Reversible (UR) Gates
## (These are only a special case!)

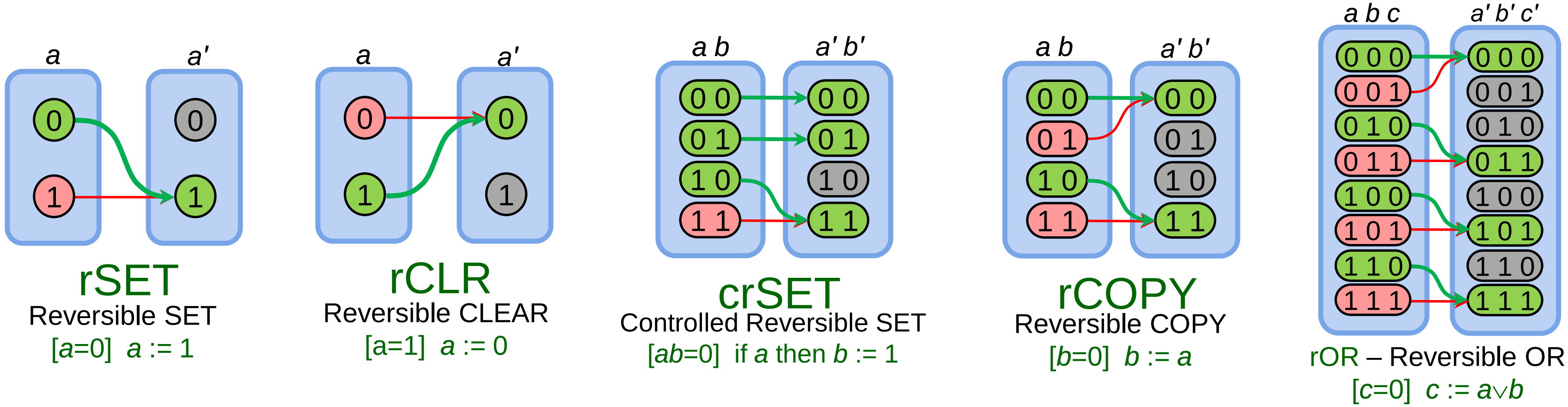
Sandia National Laboratories

- Any total, reversible, deterministic operation is simply a permutation (bijective transformation) of the state set.
- Some example UR operations (misleadingly called “gates”) on binary-encoded states:
  - NOT($a$) — $a := \neg a$ — In-place bit-flip
  - cNOT($a$,$b$) — if $a$=1 then $b := \neg b$ — Controlled NOT
  - ccNOT($a$,$b$,$c$) — if $ab$=1 then $c := \neg c$ — A.k.a. “Toffoli gate”
  - cSWAP($a$,$b$,$c$) — if $a$=1 then $b \leftrightarrow c$ — A.k.a. “Fredkin gate”
- ccNOT and cSWAP are each universal UR gates
  - The latter in the case of functions on dual-rail-encoded bit-strings
- No set of just 1- and 2-bit classical UR gates is universal
  - However, cNOT plus 1-bit quantum (unitary) gates comprise a universal set



# Generalized Reversible Computing (GRC) also includes Conditional Reversibility (CR)!

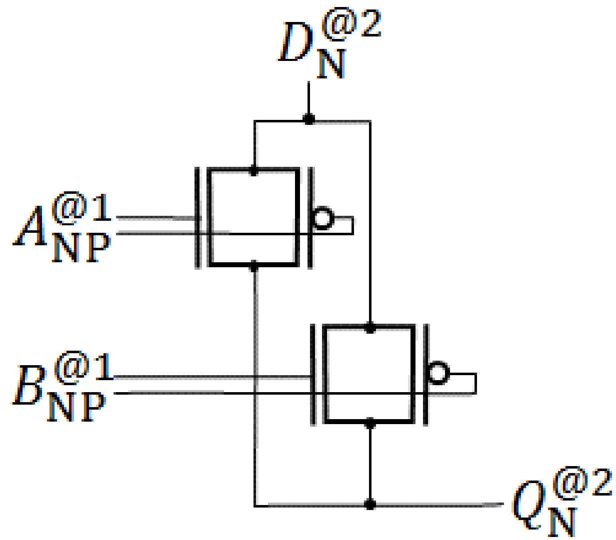
- Definition: A (deterministic) operation $O$ is *conditionally reversible under precondition* $P \subseteq S$ if and only if the restriction of $O$ to $P$ (as a partial operation) is an injective (one-to-one) operation.
  - Given any initial probability distribution $p$ over states in $S$ such that $p(x) = 0$ for all $x \notin P$, the application of the operation $O$ does not reduce the entropy of the computational state at all, and so incurs no minimum dissipation under Landauer's principle.
    - And, as all those $p(x) \to 0$, so does the minimum Landauer dissipation.
- Examples of some conditionally reversible operations:
  - Green denotes the restriction of the operation to the precondition
  - Red: States that would result in dissipation b/c precondition not met

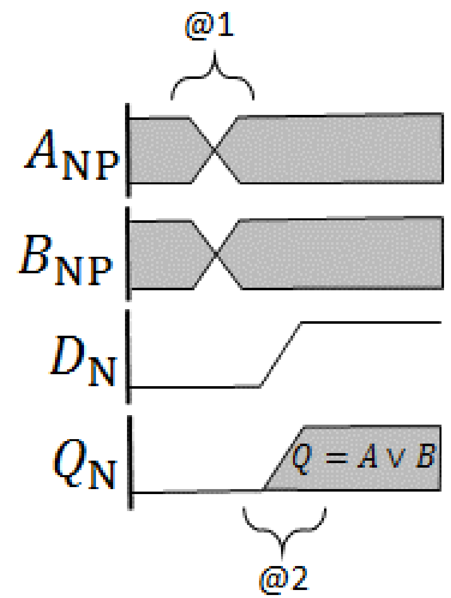


# Implementing Conditionally-Reversible Operations

Sandia National Laboratories

- Not very difficult!
  - Straightfoward to do with adiabatic switching
- E.g., this CMOS structure can be used to do/undo latched rOR operations
  - Example of 2LAL logic family
    - Based on CMOS transmission gates
    - Implicit dual-rail complementary signals (PN pairs) in this notation
- Computation sequence:
  1. Precondition: Output signal *Q* initially at logic 0
  2. Driving signal *D* is also initially logic 0
  3. At time 1 (@1), inputs *A*, *B* transition to new levels
     - Connecting *D* to *Q* if and only if *A* or *B* is logic 1
  4. At time 2 (@2), driver *D* transitions from 0 to 1
     - *Q* follows it to 1 if and only if *A* or *B* is logic 1
     - Now *Q* is the logical OR of inputs *A*,*B*
- Reversible things that we can do afterwards:
  - Restore *A*, *B* to 0 (latching *Q*), or, undo above steps

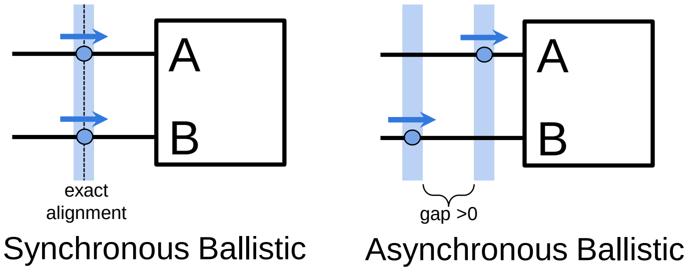


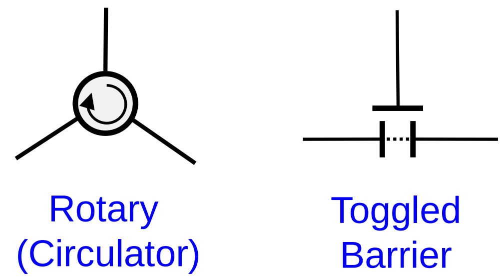



# Asynchronous Ballistic Reversible Computing

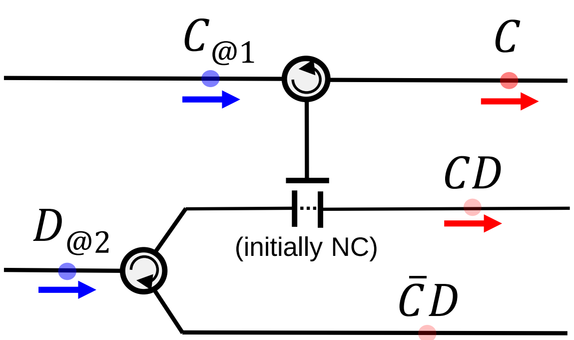
Sandia National Laboratories

- Some problems with all of the existing *adiabatic* schemes for reversible computing:
  - In general, numerous power/clock signals are needed to drive adiabatic logic transitions
  - Distributing these signals adds substantial complexity overheads and parasitic power losses
- Ballistic logic schemes can eliminate the clocks!
  - 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
- But, *synchronous* ballistic logic has some issues:
  - Unrealistically precise timing alignment required
  - Chaotic amplification of timing uncertainties when signals interact
- Benefits of asynchronous ballistic logic:
  - Much looser timing constraints
  - Linear instead of exponential increase in timing uncertainty per logic stage
  - Potentially simpler device designs
- New effort to investigate implementing ABRC in superconducting circuits (N&M LDRD idea)…



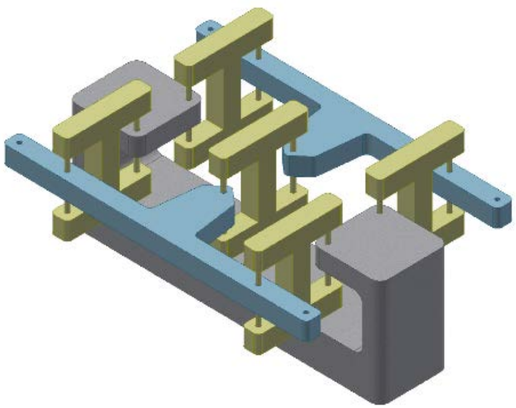

Example ABR device functions

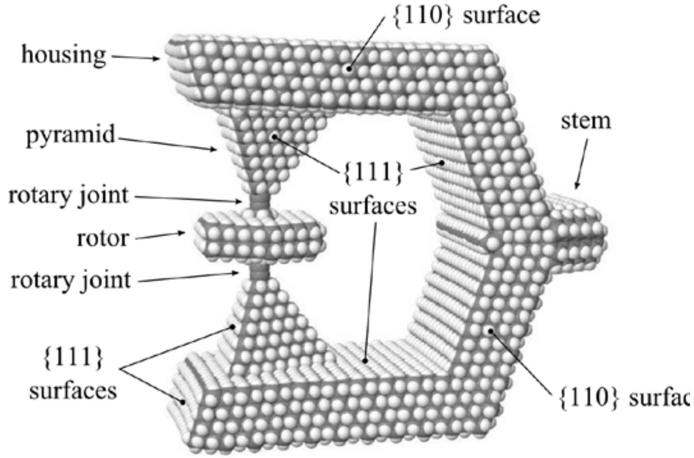


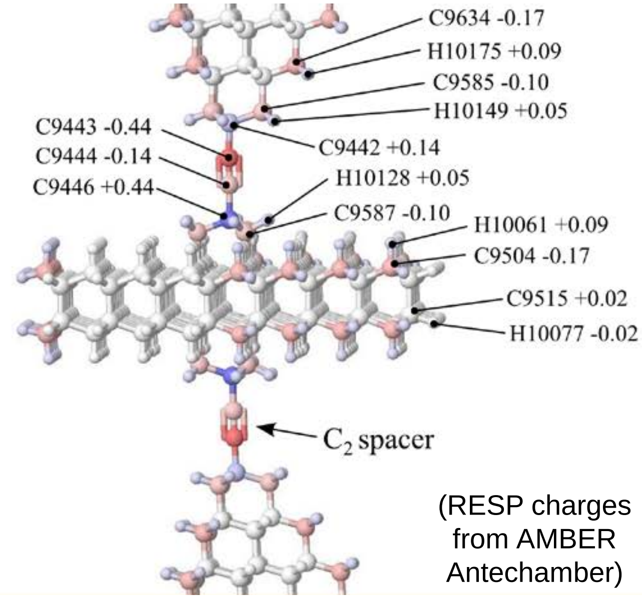

Example logic construction

# Nanomechanical Rotary Logic

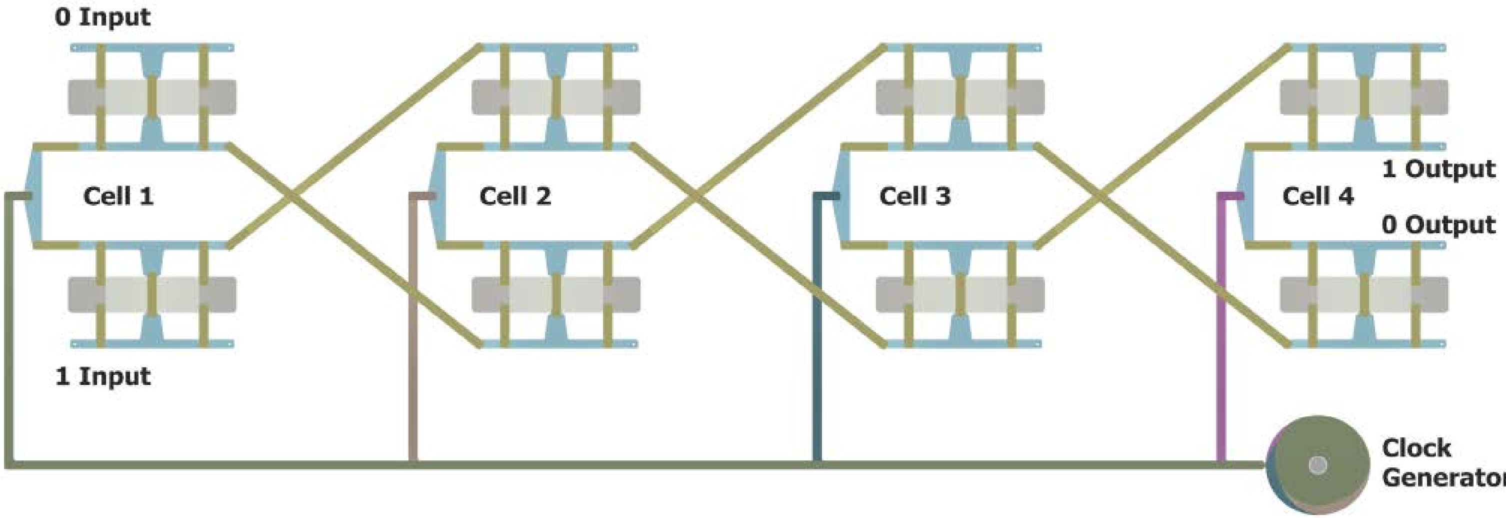
Merkle et al., IMM Report 46 and Hogg et al., arxiv:1701.08202 (reproduced with permission)





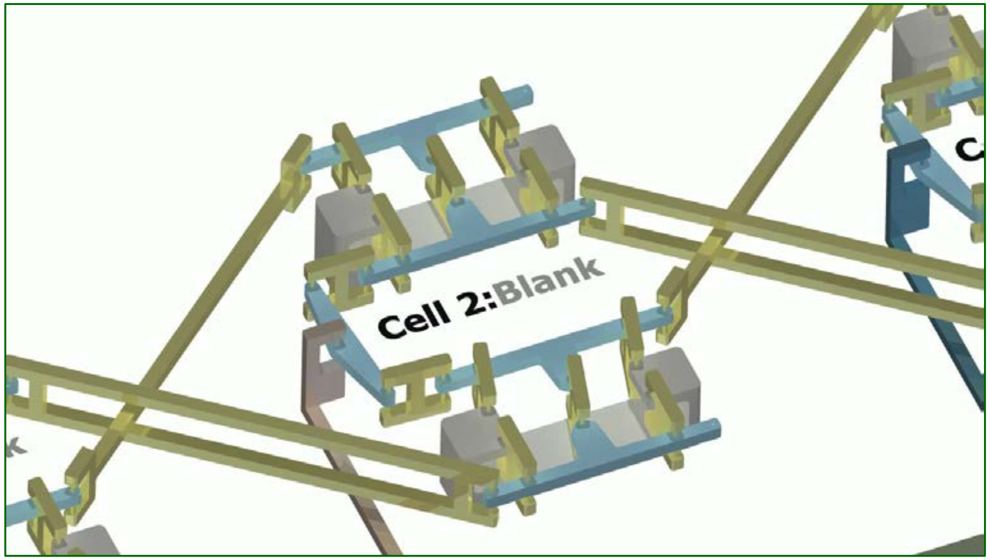



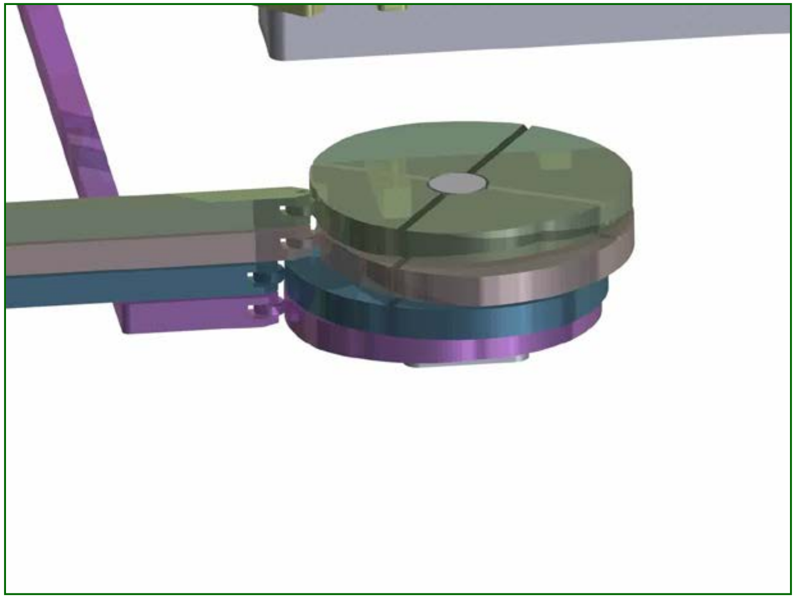

(RESP charges from AMBER Antechamber)



# Rotary Logic Lock Operation

- Videos animate schematic geometry of a pair of locks in a reversible shift register
- Molecular Dynamics modeling/simulation tools used for analysis include:
  - LAMMPS, GROMACS, AMBER Antechamber
- Simulated dissipation:
  - ~4 × $10^{-26}$ J/cycle at 100 MHz
    - 74,000 × below Landauer limit for irreversible ops!
- Speeds up into GHz range should also be achievable



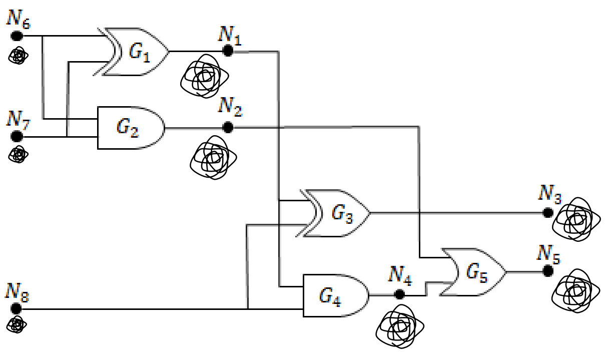



# Chaotic Logic – Summary

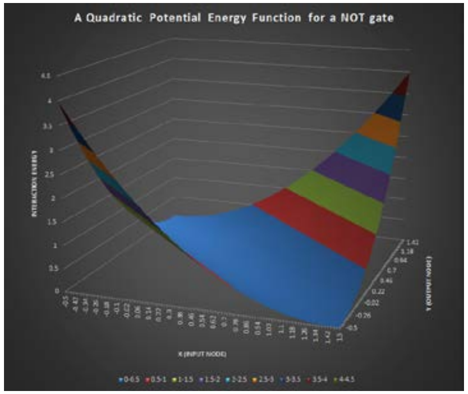
- Shannon teaches us that reliable communication is still possible with signal energies below the noise floor
  - Why not also reliable computation?
- Chaotic Network Model of logic:
  - Nodes are dynamic variables
  - Gates are Hamiltonian interaction terms
  - Node values chaotically fluctuate around a long-term average that encodes the result of the computation
- A simulator for this model was built…
  - cs.sandia.gov → Software → Dynamic
    - Page also links to a paper & a full talk

$$C = B\log_2\left(1 + \frac{S}{N}\right)$$

Channel capacity theorem




Full Adder dynamical network

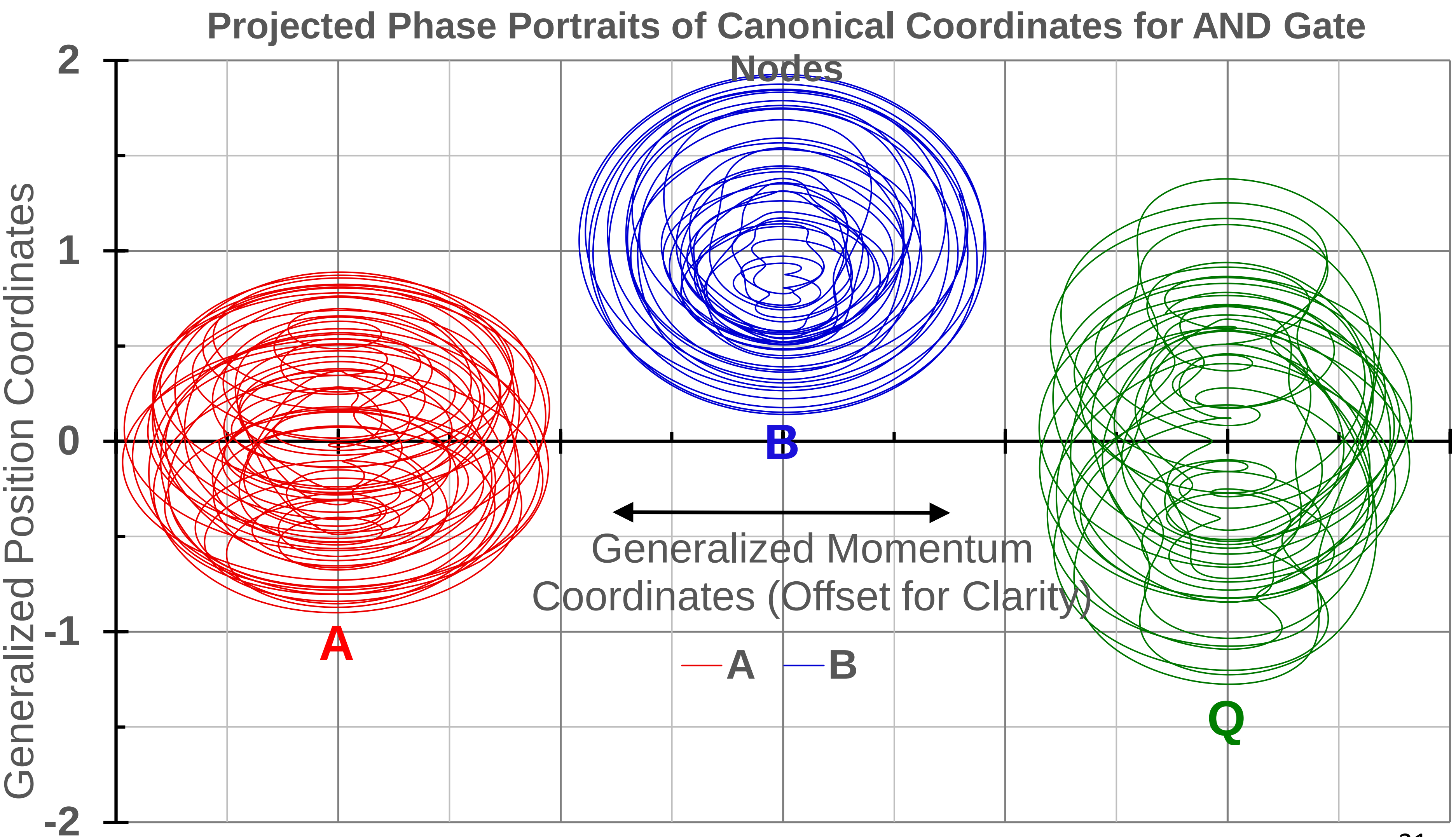



Logic gates implemented by potential energy surfaces

Frank & DeBenedictis '16, "A Novel Operational Paradigm for Thermodynamically Reversible Logic: Adiabatic Transformation of Chaotic Nonlinear Dynamical Circuits"

# Chaotic AND Gate

Mean A = -0.001531, B = 0.975105, Q = 0.001710



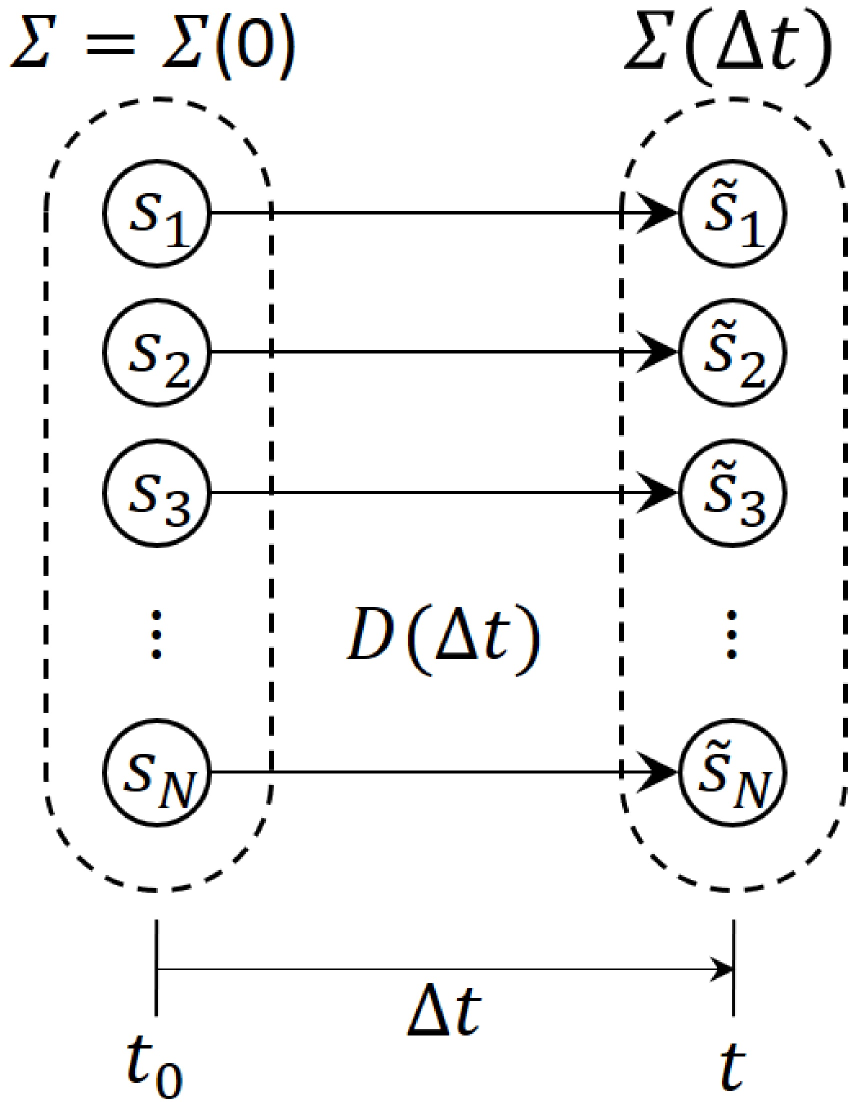
# Conclusion

- The increasing economic utility of computing has been enabled by steadily increasing *energy efficiency*
  - However, fundamental limits on energy efficiency threaten to prevent further general-purpose improvements in the relatively near term
- Transcending the practical and fundamental limits will necessarily require the increasing application of *reversible computing* principles…
  - The most general form of which is described by *Generalized Reversible Computing* theory (first paper to appear in RC'17)
  - A particularly efficient implementation of reversible computing is the new *Asynchronous Reversible Computing* (ARC) approach
    - In progress:  Paper, funding effort, possible patent application
  - There is another approach called *Chaotic Logic* which also avoids clocks, and can potentially use signal energies below thermal noise
- Further development of these research areas will be key to future computer performance and economic development

# EXTRA SLIDES

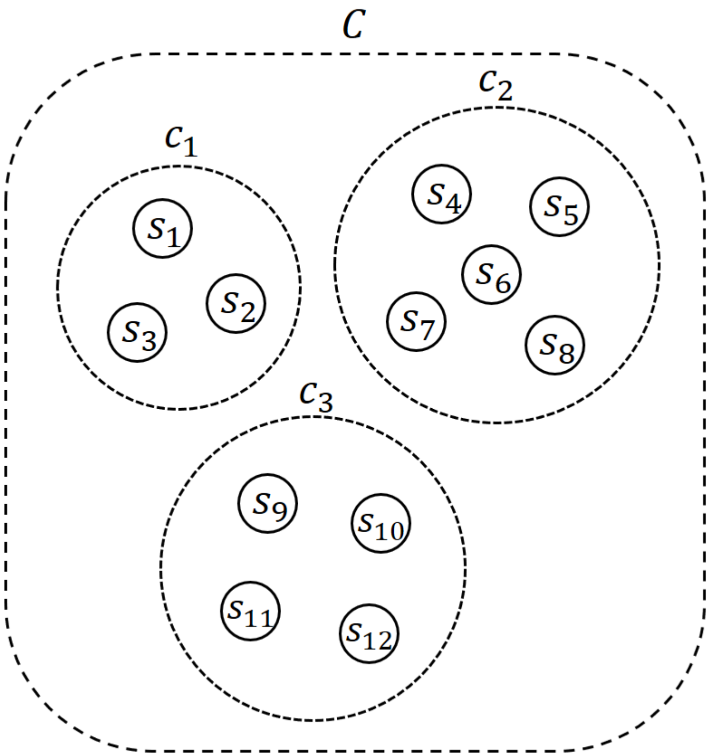
# Landauer's Principle in a Nutshell (1 of 4)

- 1. Fact: Fundamental Physics is Reversible
  - Dynamical evolution over time transforms old sets of distinguishable physical microstates (orthogonal quantum states) one-to-one to new distinguishable sets of physical microstates
    - Follows from unitarity of gen. Schrödinger equation
    - If it wasn't true, the Second Law of Thermodynamics would not hold!



# Landauer's Principle in a Nutshell (2 of 4)

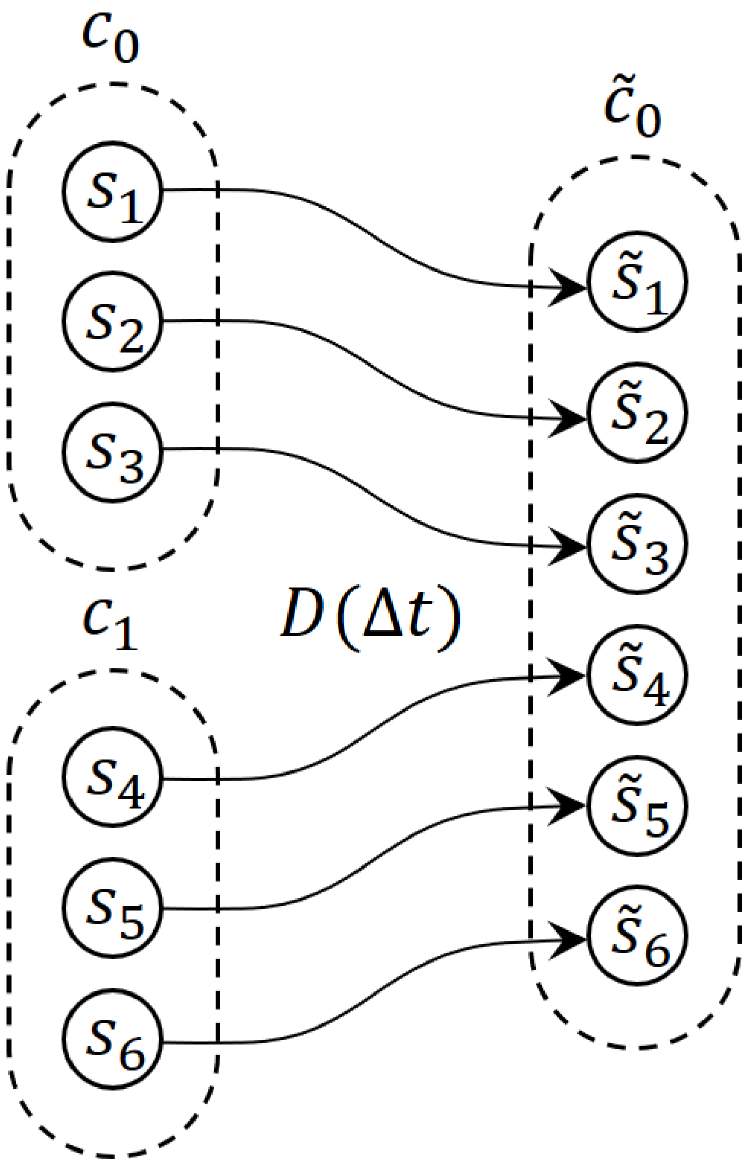
- 2. A *computational state* is just an equivalence class of distinct physical microstates that we interpret alike for computational purposes.
  - *E.g.* any state of a circuit node in which its average voltage $V$ is in some range, $V_{1L} < V < V_{1H}$, may represent a logic "1"
    - But, there are many detailed physical microstates consistent with this!
      - E.g., at nonzero temperature, many electron states near Fermi level may or may not be occupied




Example of a computational state space $C$ consisting of 3 distinct computational states $c_1, c_2, c_3$, each defined as a set of equivalent physical states.

# Landauer's Principle in a Nutshell (3 of 4)

- 3. When we "erase information" in a computer (merge computational states), the underlying physical microstates remain distinct
  - Before the erasure, the entropy of the detailed state *s*, conditioned on the computational state, is given by…
    - $H(s \mid c) = H(s) - H(c)$
  - After the erasure, there is no more entropy in the computational state, so
    - $H(s \mid c) = H(s)$
  - The physical entropy (from the user's perspective) has increased by $H(c)$!
- **Losing computational information increases physical entropy!**



# Landauer's Principle in a Nutshell (4 of 4)

- 4. Entropy/information is measured in logarithmic units.
  - Two equiprobable computational states → Entropy/information content of computational state is one factor-of-two logarithmic unit

$$H(c) = [\log 2] = [\log \mathrm{e}] \log_{\mathrm{e}} 2 = k_{\mathrm{B}} \ln 2$$

- 5. If entropy $\Delta S = H(c)$ ends up in a thermal environment at temperature $T$, this requires adding heat $\Delta Q = T\Delta S$ to the heat bath, by the definition of thermodynamic temperature:

$$\frac{1}{T} = \frac{\partial S}{\partial Q}$$

- ∴ Merging two equally-likely computational states implies that we must dissipate this amount of energy to the heat bath:

$$\Delta E_{\mathrm{diss}} = k_{\mathrm{B}} T \ln 2$$ ← Landauer limit