Physical Limits of Computing

Michael P. Frank

CCR Summer Seminar

July 25th, 2016

Approved for Unclassified Unlimited Release
SAND2016-xxx PE

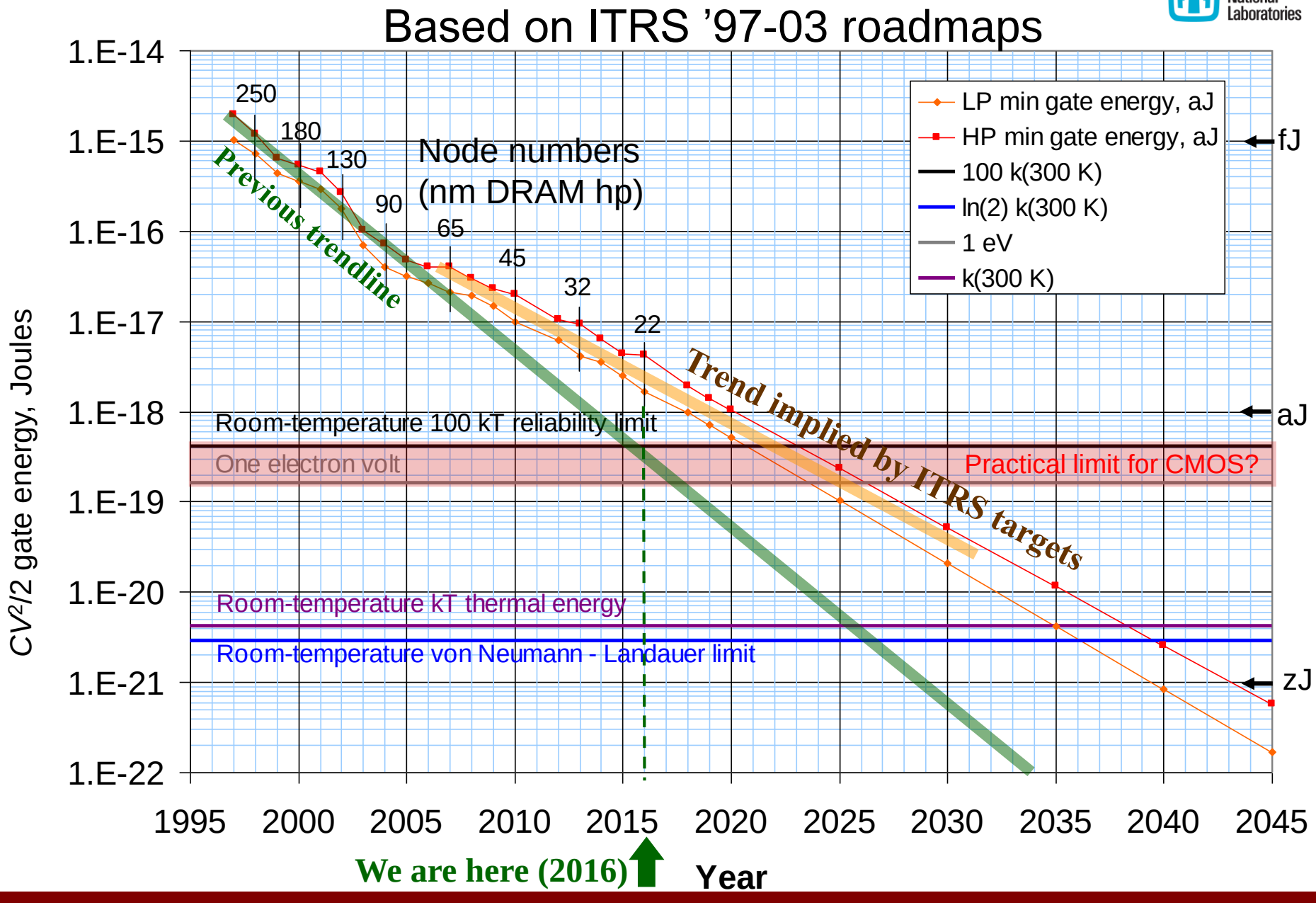
Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Abstract

- Computational performance per unit cost (including energy/cooling cost) has improved exponentially over the last > half-century
  - Enabled by transistor downscaling and associated energy reductions
  - But, physical limits to this scaling path are only ~10 years away
    - Intrinsic limitations of MOSFET technology
- Industry is still looking for the "next" technology that can replace CMOS and enable continued efficiency scaling…
  - Understanding fundamental (technology-independent) efficiency limits is necessary to help guide us in this search
- An important class of energy efficiency limits arises from fundamental quantum theory and thermodynamics
  - Performing optimally within these limits will require fundamentally new computing paradigms (not just "better transistors" for conventional logic)
- This talk will focus on new computing paradigms that may improve the efficiency of general-purpose digital computing
  - By leveraging principles such as reversibility, nondeterminism, and chaos
  - Full-blown Quantum Computing would likely confer even greater benefits, but only for more specialized types of applications. ← Not our focus here

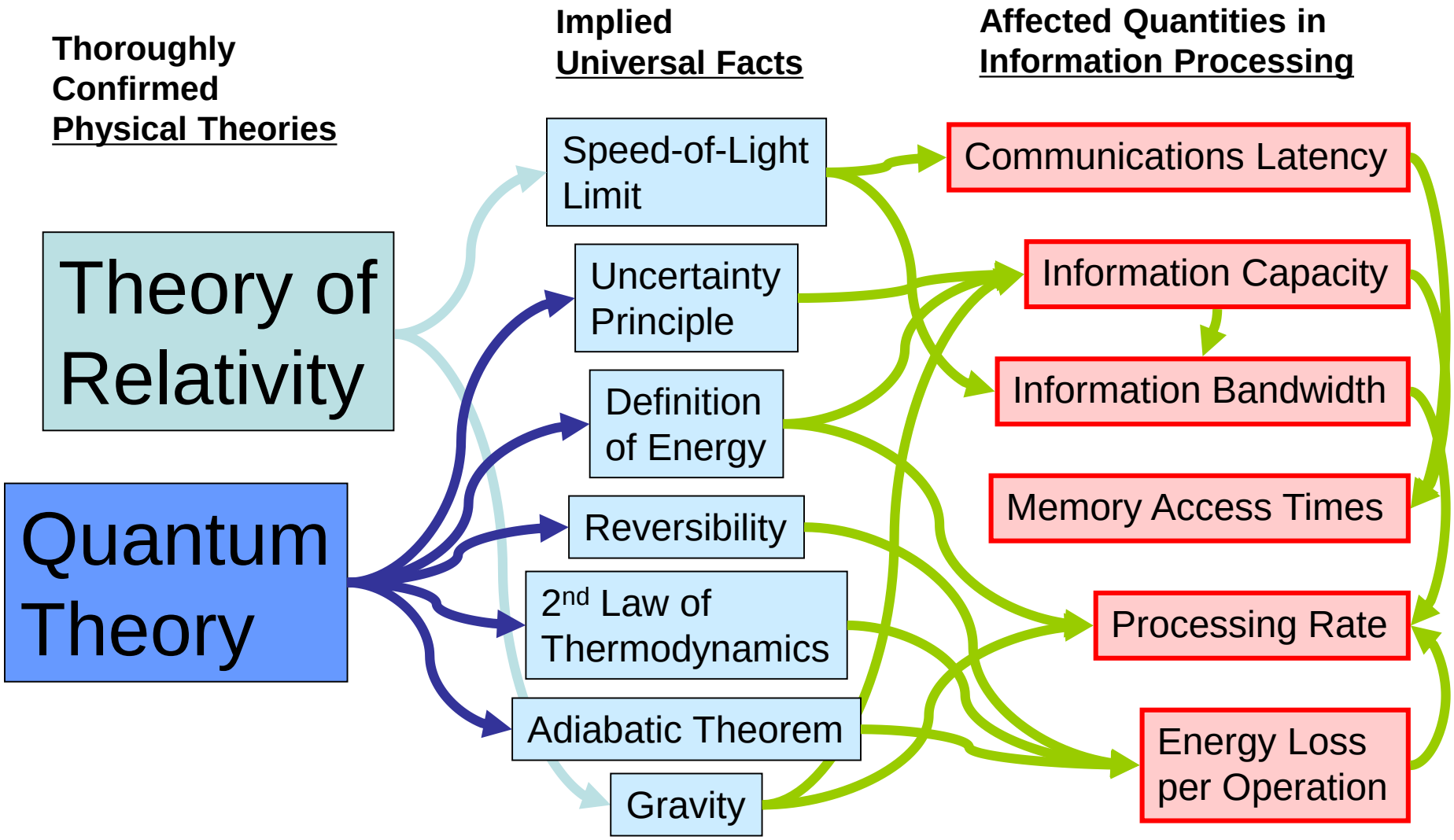
Trend of Min. Transistor Switching Energy

Sandia National Laboratories



Fundamental Physics Implies Various Firm Limits on Computing

Sandia National Laboratories



Physics Interpreted Computationally

Sandia National Laboratories

- Many (if not all) physical quantities can themselves be interpreted in terms of information processing concepts:
  - Entropy
    - The amount of unknown or incompressible information in a system.
  - Action
    - The amount of computational effort exerted in a given transformation.
  - Energy
    - Rate at which computational effort is being exerted in a given system.
  - (Generalized) Temperature
    - Rate of computational effort exerted per unit of information capacity.
  - Momentum
    - “Motional” computational effort exerted per unit distance traversed.
  - And so on…

Some Important Facts of Physics

Sandia National Laboratories

Bearing on the fundamental limits of computing:

- Information cannot propagate faster than light
  - Ignoring here the possibility of exotic spacetime configurations
- Only mutually *orthogonal* quantum states can be reliably distinguished from each other
  - Limits the information content of physical systems of given energy
- The complete microscopic (quantum) state of any physical system evolves *reversibly* (more specifically, unitarily).
  - No microscopic information loss → 2nd law of thermo.
  - "Erasure" of digital information → Entropy increase → Energy loss
- Energy itself is a measure of the rate of quantum state change
  - Limits the speed of even reversible computations as a function of their energy content

Quantum State Counting, Information, and Entropy

Sandia National Laboratories

- Suppose a given physical system (as defined in a given context) has *N* distinguishable (orthogonal) quantum states,
  - Then we can say its *physical information capacity* is $C = \log N$.
    - The base of the logarithm determines the information unit.
      - Base 2: Unit is 1 bit. Base *e*: Unit is 1 "nat" or k_B (Boltzmann's constant).
- Given a state of knowledge about the system expressed by a probability distribution p_i (where i indexes system states),
  - Then we say the system's *entropy* is its unknown information content,

$$S = \sum_{i=1}^{N} p_i \log \frac{1}{p_i} \leq C.$$

  - and its *known information* (a.k.a. negentropy) is the remainder,

$$K = C - S = \log N - \sum_{i=1}^{N} p_i \log \frac{1}{p_i}.$$

Quantum Time-Evolution & The 2nd Law of Thermodynamics

Sandia National Laboratories

- The complete evolution of any quantum system over a time *t* is expressed by some unitary transformation,

$$U(t) = e^{-\mathrm{i}Ht/\hbar}$$

  - where *H* is the system's Hamiltonian operator, an energy-valued hermitian (self-adjoint) linear operator.
- Unitary transformations are generalized rotations; they have the mathematical property that angles between vectors are preserved by the transformation.
  - Thus if $|\Psi_1\rangle$ and $|\Psi_2\rangle$ are mutually orthogonal, then so are $U|\Psi_1\rangle$ and $U|\Psi_2\rangle$. *States that start distinguishable, stay distinguishable.*
- Therefore, the physical information capacity *C* of a system is conserved by its quantum time evolution…
  - Moreover, barring measurement from outside, its entropy *S* can only subjectively increase, *e.g.* if we don't know *H* exactly and can't track the exact evolution. *It can never spontaneously decrease.* → 2nd law
- In other words, *physical information cannot be destroyed.*

Landauer's Principle

- Due to the indestructibility of information, *all* physical operations are microscopically one-to-one (injective)…
  - Thus, any information that existed in the system before the operation, must still exist afterwards
  - Whenever we think that we have simply "erased" some information, we must have actually only transformed it to another form
    - *e.g.*, entropy in the environment
- If a computational operation reduces the entropy of the logical state by an amount ΔS, it must increase the entropy of some other part of the system or environment by at least ΔS.
  - If this entropy ends up in an environment at temperature T, this implies an amount of heat $\Delta Q = T\Delta S$ must have been added to the environment (by def'n of temperature).
- Erasing 1 bit ($k_B \ln 2$) of information →
  - $k_B T \ln 2 \approx 18\ \text{meV}$ energy dissipated to heat in thermal environment
    - Implies ≤ ~350 Eb erased per second (Exa=10^{18}) per Watt in room-*T* env.

Reversible Computing

Sandia National Laboratories

- A general definition of a *computational operation* on a state set S is a (possibly partial) mapping from *initial* states $x \in S$ to probability distributions over *final* states $y \in S$.
- Typically in computing, we wish to carry out *deterministic* operations in which the final state distributions are singular.
  - However, nondeterministic operations are also possible and useful.
- Call a computational operation *reversible* if and only if all of its final state distributions are non-overlapping.
  - Every possible final state has (at most) one predecessor.
- *Reversible computing* refers to computing with reversible operations.
  - Wherever reversible operations are used, the entropy of the logical state is not decreased, and so Landauer's principle does not require any minimum energy dissipation for those operations.

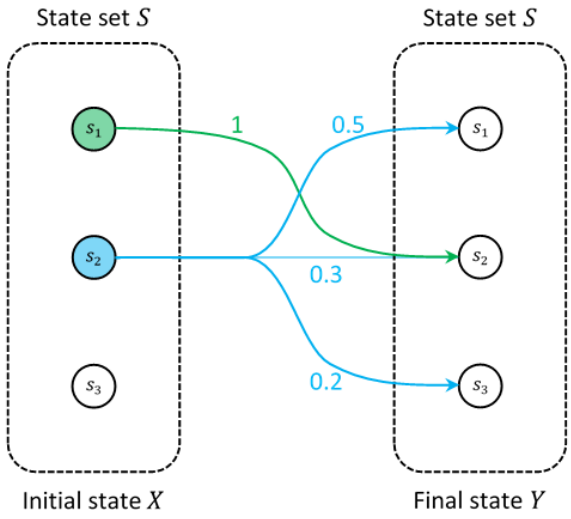
Types of Computational Operations

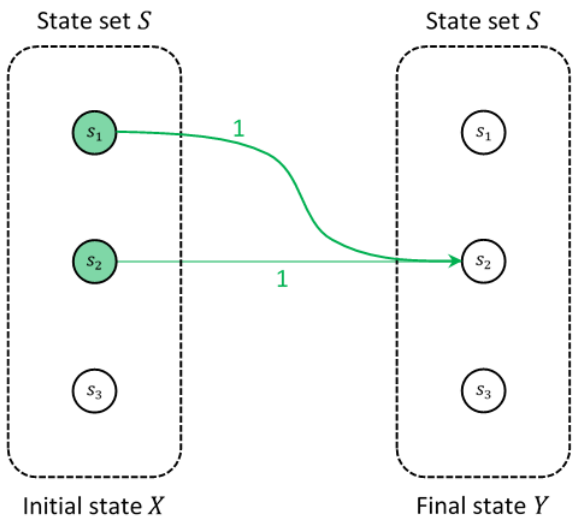
Sandia National Laboratories

Nondeterministic

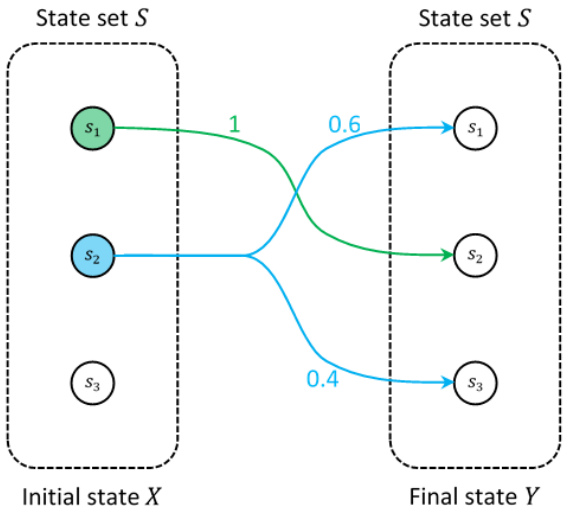
Deterministic

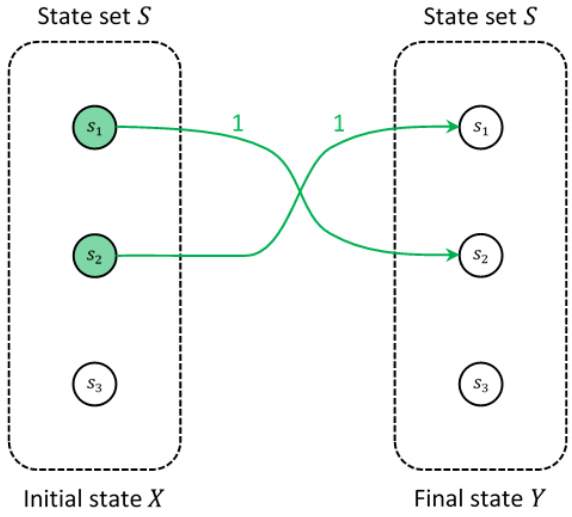
Irreversible





Reversible

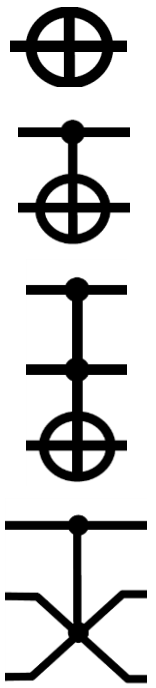




Unconditionally Reversible Gates

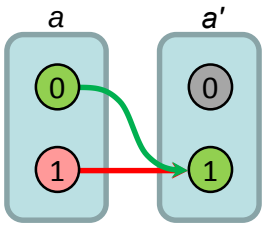
Sandia National Laboratories

- Any complete reversible, deterministic operation is simply a permutation (bijective transformation) of the state set.
- Some example reversible operations ("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 gates
  - The latter in the case of functions on dual-rail-encoded bit-strings
- No set of 1- and 2-bit reversible gates is universal
  - However, cNOT plus 1-bit quantum (unitary) gates is a universal set

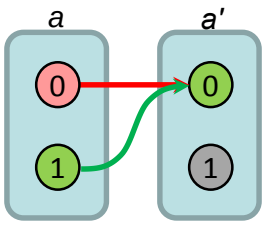


Conditional Reversibility (CR)

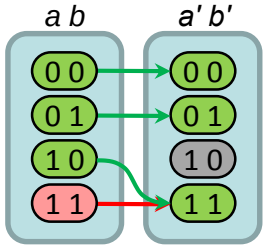
- Definition: An operation O is *conditionally reversible under precondition* $P \subseteq S$ if and only if the restriction of O to P is a reversible operation (as a partial function).
  - Given an initial probability distribution p over states in S such that $p(x) = 0$ for all $x \notin P$, the application of the operation *O* does not reduce the entropy of the computational state, and so incurs no minimum dissipation under Landauer's principle.
- 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



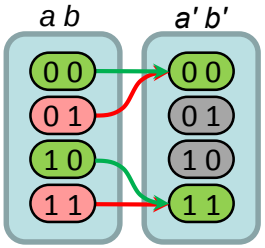
rSET
Reversible SET
(*a*=0) *a* := 1



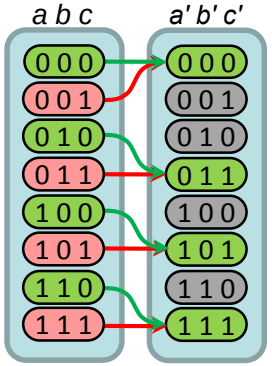
rCLR
Reversible CLEAR
(a=1) *a* := 0



crSET
Controlled reversible SET
(*ab*=0) if *a* then *b* := 1



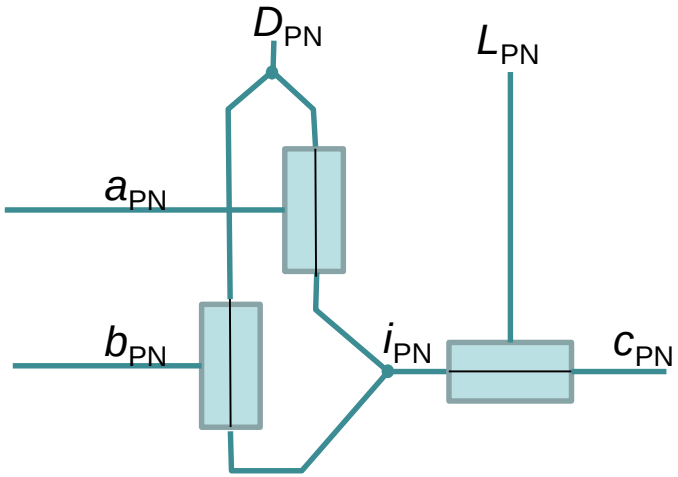
rCOPY
Reversible COPY
(*b*=0) *b* := *a*



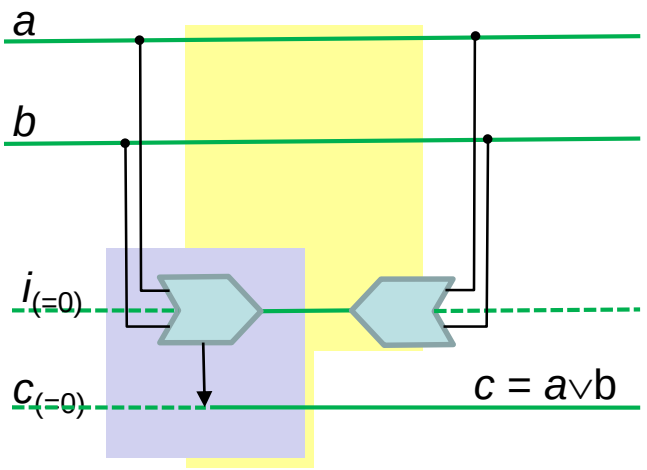
rOR – Reversible OR
(*c*=0) *c* := *a*∨*b*

Implementing CR Operations

- Not very difficult!
  - Easy to do with adiabatic switching
  - This structure can be used to do/undo rOR operations
  - Example of 2LAL logic family
    - Based on CMOS transmission gates
    - Implicit dual-rail complementary signals (PN pairs) in this notation
- Operation Sequence:
  - Initial state:
    - *a*, *b* are inputs, other nodes are logic 0
  - Latch control L_{PN} goes high (open)
    - *i* and *c* tied together (purple)
  - Gate drive D_{PN} goes high
    - *i* & *c* go high, dependent on $a \vee b$ (yellow)
  - Latch control L_{PN} goes low (closed)
    - *c* is now disconnected, independent
  - Gate drive D_{PN} reverts low
    - *i* returns to intermediate state

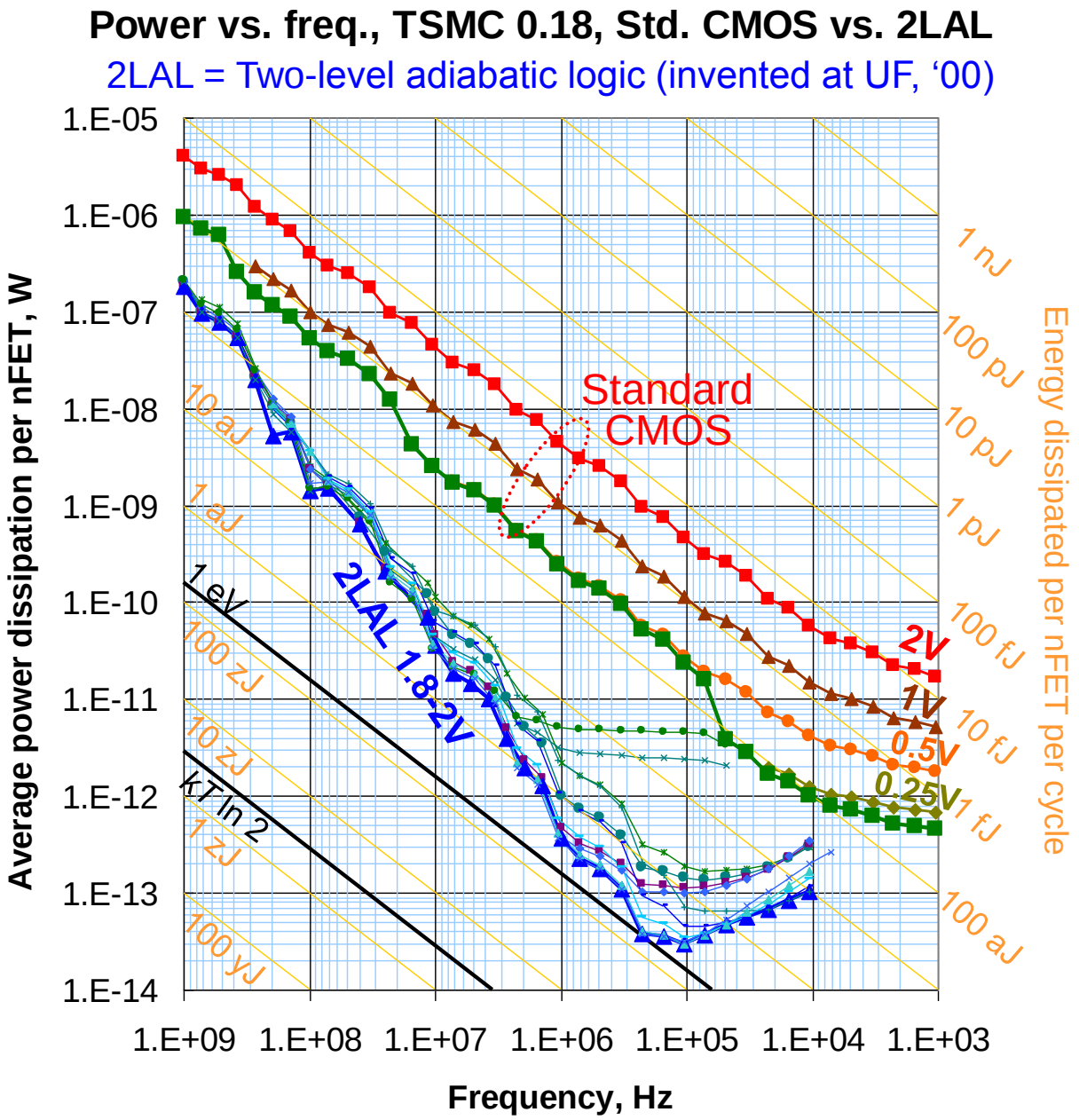


Hardware Schematic



Spacetime Diagram

Simulation Results (Cadence/Spectre)



- Graph shows power dissipation vs. frequency
  - in 8-stage shift register.
- At moderate frequencies (1 MHz),
  - Reversible uses < 1/100th the power of irreversible!
- At ultra-low power (1 pW/transistor)
  - Reversible is 100 × faster than irreversible!
- Minimum energy dissip. per nFET is < 1 eV!
  - 500 × lower than best irreversible!
    - 500 × higher computational energy efficiency!
- Energy transferred is still ~10 fJ (~100 keV)
  - So, energy recovery efficiency is 99.999%!
    - Not including losses in power supply, though

Some Possible Uses for Nondeterminism

Sandia National Laboratories

- Given appropriate device mechanisms, can be used to temporarily reduce entropy of environment (cooling it)
  - Entropy is moved from environment into computational bits
- Source of randomness for use in probabilistic (randomized) algorithms.
  - In some cases, such algorithms have computational complexity advantages over the best-known fully-deterministic algorithms
    - Can't prove the same results for pseudo-random number generators
- In cases where nondeterminism doesn't hurt, allowing it to occur permits us to use operations that are less reliable due to lower signal energies which may be contaminated by thermal noise. Doing so may improve energy efficiency

Boltzmann Distribution

- Derived from very general thermodynamic arguments about the interaction of a system at equilibrium with a much larger thermal environment at some temperature T.
  - I.e., independent of the technology used in the system
    - Apart from quantum-mechanical corrections for assemblages of fermions (Fermi-Dirac distribution) or bosons (Bose-Einstein distribution)
- The probability that the system will be found in any given state having energy E is proportional to $e^{-E/kT}$.
  - Thus, if we wish for the probability that a system at equilibrium is *not* in a certain desired state to be less than some small amount $p_\epsilon \ll 1$, then we must arrange for any non-desired state to have energy

$$E > kT \ln\left(\frac{1-p_\epsilon}{p_\epsilon}\right) \approx kT \ln\left(\frac{1}{p_\epsilon}\right),$$

    or even higher than this if there are multiple non-desired states.
- Less energy → Greater likelihood of thermally induced error

Quantum Speed Limit

Sandia National Laboratories

- The energy E of any quantum system (above its ground state) determines the rate at which it exerts a certain quantum-theoretic measure $\mathcal{F}$ of *computational effort*.
  - Average angular distance traversed by the state's complex coefficients = Twice the total complex-plane area swept out = Imag. trajectory length.

$$\frac{d\mathcal{F}}{dt} = \frac{E(t) - E_0}{\hbar}.$$

(here dimensioned in angular velocity units)

  - *E.g.*: An excitation of 1 eV corresponds to 1.52×10^{15} rad/s.
- A minimum effort of $\mathcal{F} \geq \pi/2$ (rad), applied appropriately, is required to flip a bit, and we need $\mathcal{F} \geq \pi$ to progress one step along a non-repeating sequence of distinguishable states.
  - Any specific computational task has a minimum worst-case effort or *difficulty*, which determines the minimum energy-time investment required to carry out that computation on worst-case input states.

Temperature as "Clock Speed"

Sandia National Laboratories

- Thermodynamically, temperature T is defined by

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

  - Partial derivative of the system's entropy S (at equilibrium) with respect to the thermal energy E of the system.
- For simple systems with heat capacity proportional to temperature (*e.g.* an ideal Fermi gas), we find that $T \propto E/S$.
  - With a constant factor determined by a constant of integration.
- Equilibrium (maximum) entropy is just information capacity.
- Temperature is thus ($\propto$) energy per unit information capacity…
  - Computational effort per bit.
  - Note this quantity is well-defined even for non-equilibrium states!
- Example: Room-temperature Fermi gas →
  - Rate of effort sufficient for at most 4.33×10^{12} transitions / sec. / bit

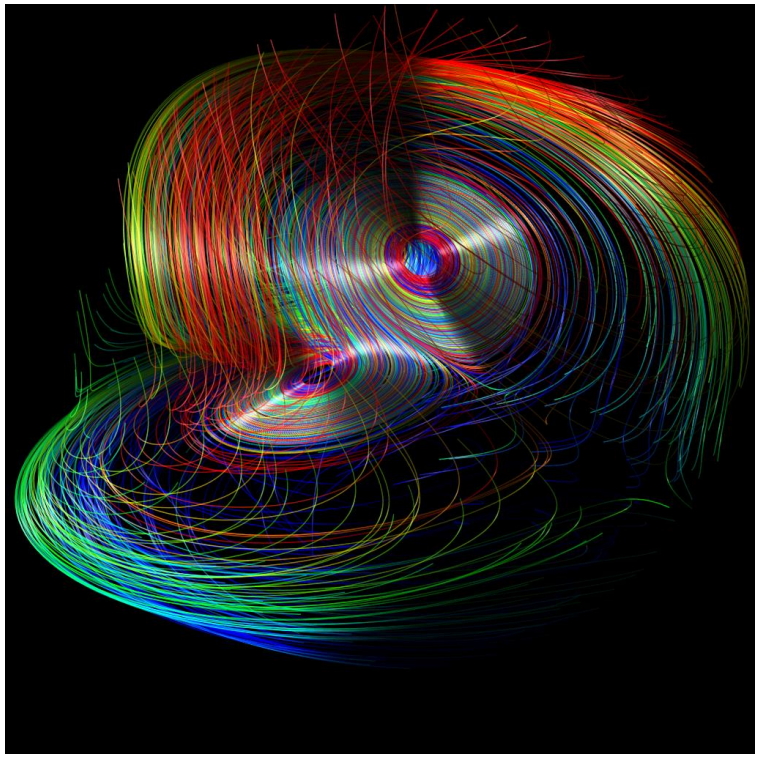
Chaos

- Typically, the dynamical behavior of real-world physical systems exhibits *chaos*
  - Extreme sensitivity to initial conditions
    - When the microstate is not known precisely, the long-term evolution cannot be accurately predicted even when the macroscopic state is known fairly accurately. → System behavior appears nondeterministic.
  - This feature persists despite the underlying determinism and reversibility of the microscopic quantum-mechanical dynamics!
    - It's simply too hard to know the parameters of a system's Hamiltonian precisely enough to predict its macroscopic dynamics exactly
    - Also, imperfect isolation of a system means that unavoidable interactions with its unknown environment will cause decoherence of its quantum state, effectively increasing its entropy
- Given that some degree of chaos appears unavoidable, can we harness it for computation, rather than be harmed by it?

Chaotic Computing?

Sandia National Laboratories

- What are some potential advantages of utilizing chaotic dynamical systems for computation?
  - In a conservative chaotic system, the strange attractor to which the dynamics converges represents a thermodynamic equilibrium state
    - Once converged onto the attractor, there is no further energy dissipation
    - This remains true even if the system is interacting with an external thermal environment once the system and environment temperatures equilibrate, due to the fluctuation-dissipation theorem
  - The identity of the attractor reflects information about the initial state and the time-series of external forcings being applied to the system
    - Automatically computes a function of these inputs (possibly a useful one)
    - Cheaply maps a simple input into a much higher-dimensional space of trajectories
      - This can be useful for learning, as in reservoir computing



Computing Below the Noise Floor

Sandia National Laboratories

- Shannon teaches us that reliable communication remains possible when SNR << 1, just at a low bit rate

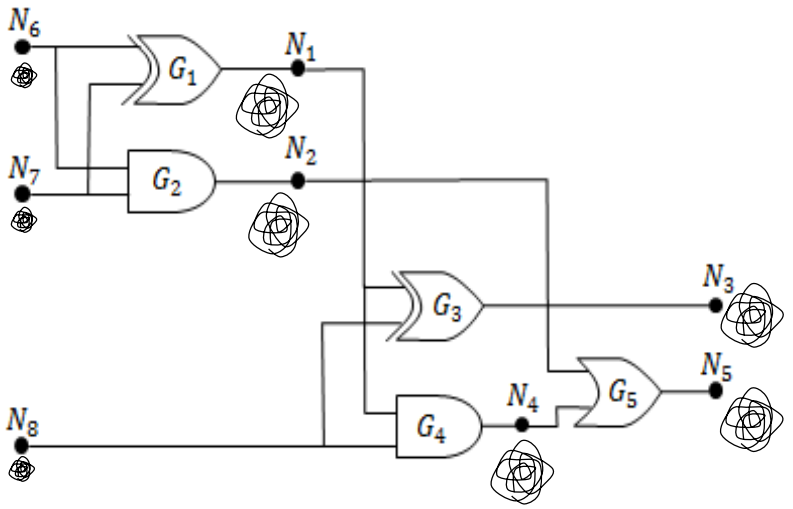
$$C = B \log_2 \left(1 + \frac{S}{N}\right)$$

- A computational dataflow can be considered as just a special case of a communication channel that happens to transform the data in transit!

- Therefore, it ought to be possible to carry out reliable computations as well using signals that have less than the thermal energy, just at a correspondingly slow rate

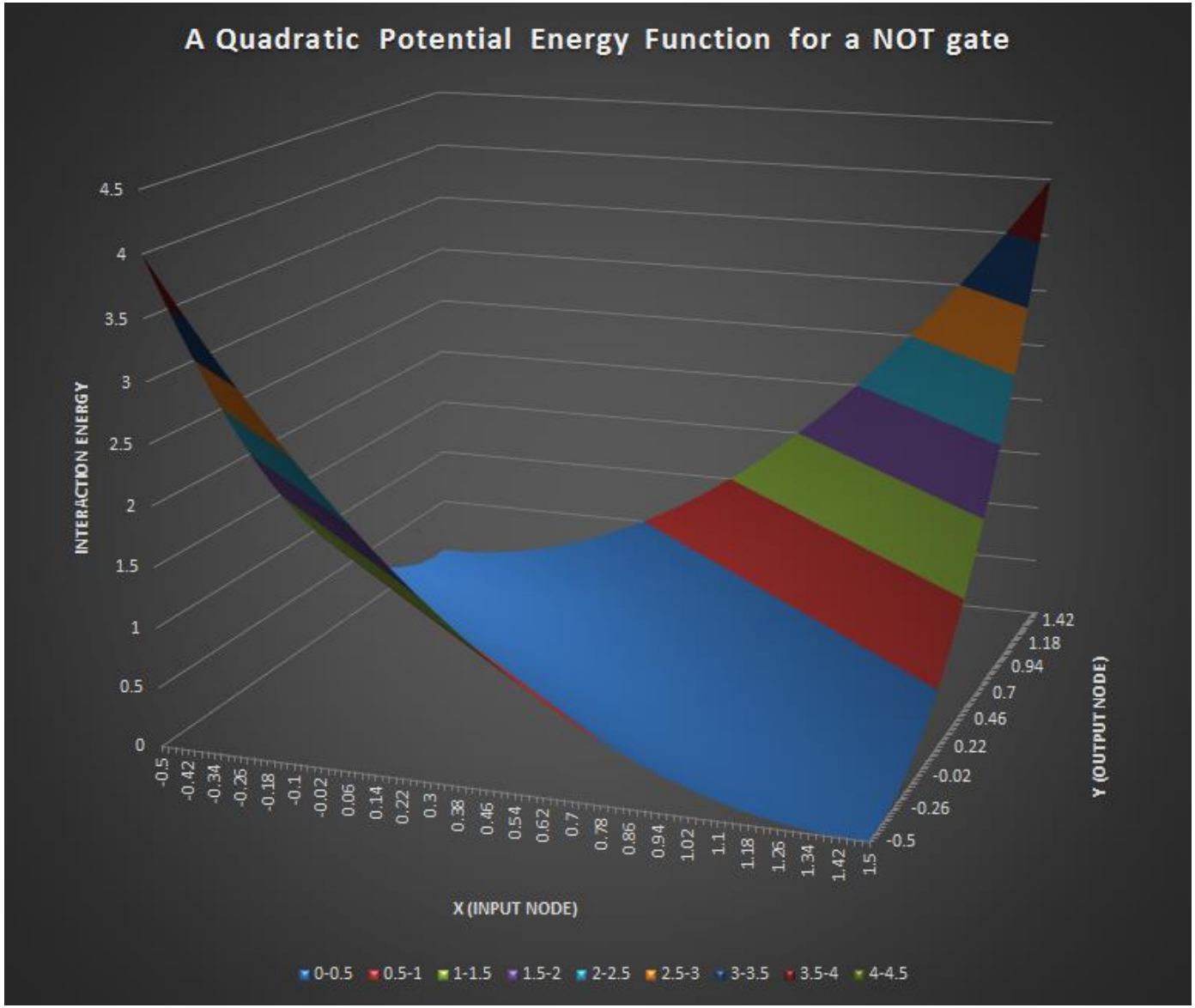
Chaotic Network Model of Logic

- Combinational logic via nonlinearly interacting degrees of freedom in classical conservative dynamical systems
  - Let each "node" be a generalized position coordinate q_i, continuous range incl. 0,1
    - There is a corresponding momentum p_i
    - Init. w. a thermally distributed kinetic energy
  - Logic "gates" become terms in a classical Hamiltonian energy function
    - Coupling neighboring degrees of freedom
    - Potential energy minimized
  - Each node traverses a complex (generally chaotic) trajectory in its phase space
    - Overall configuration is thermally distributed around the global ground state configuration
  - Network inputs can be tightly constrained
    - Deep potential well – low error probability
  - Although outputs fluctuate randomly,
    - Long-term average statistical behavior still conveys information about ideal result



- Adiabatic updating in one step:
  - Gradually transition inputs 0 ↔ 1
  - System remains close to a thermally distributed equilibrium state throughout the transition
    - Asymptotically zero heating of the system → no energy dissipation
  - Measure final state over a long period → learn result

Example of a Nonlinear Interaction



Example Interaction Functions

Sandia National Laboratories

- Here are some simple quadratic interactions:
  - NOT gate coupling input x_j to output x_i:

$$E_i = \frac{1}{2}\, b\, kT\left(x_i + x_j - 1\right)^2$$

  - AND gate coupling inputs x_j, x_k to output x_i:

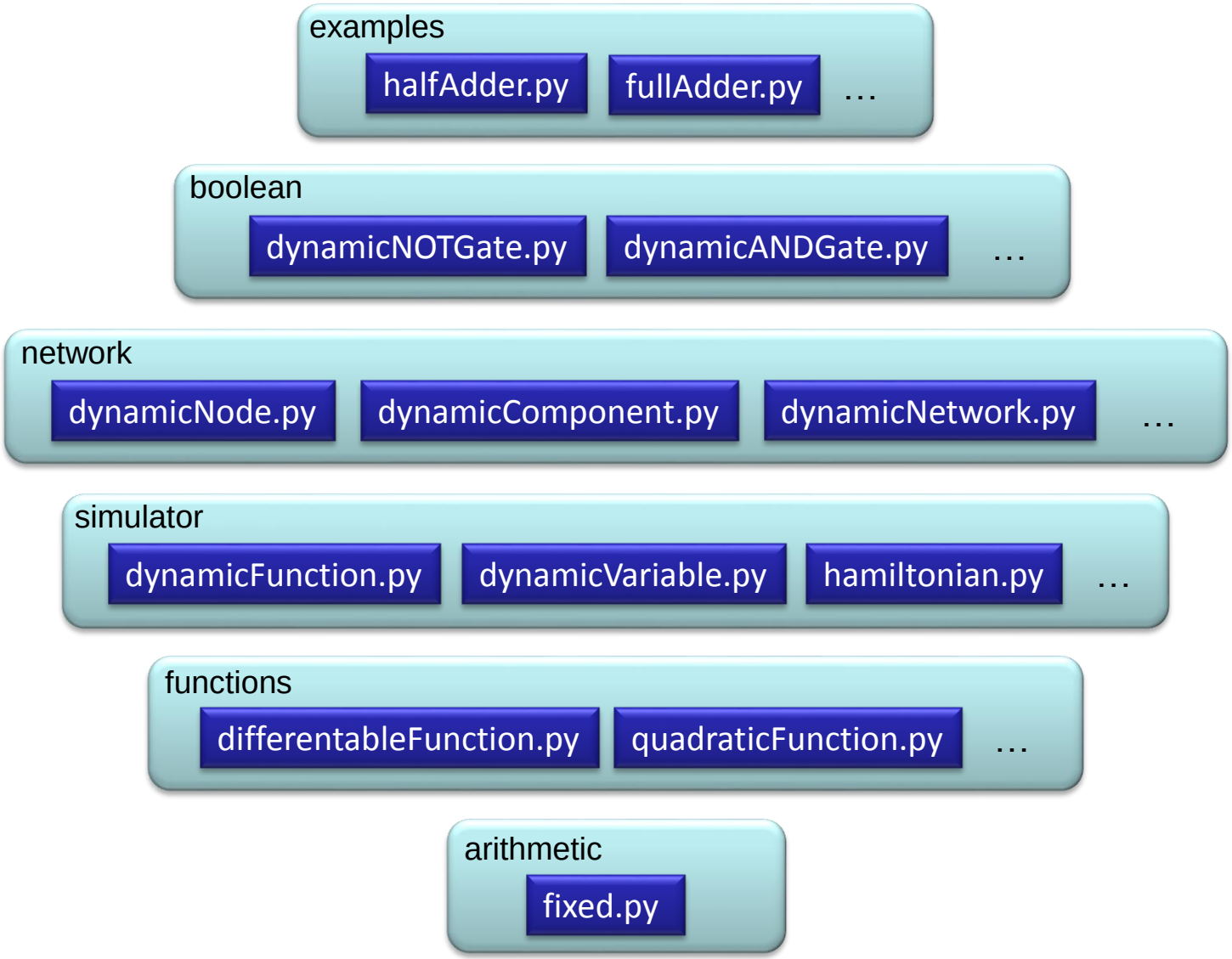
$$E_i = \frac{1}{2}\, a\, kT\left(x_i - x_j x_k\right)^2$$

DYNAMIC simulator

Sandia National Laboratories

- I am currently prototyping (in Python) a simple simulator called DYNAMIC for these types of dynamical networks.
  - Nodes interacting via arbitrary Hamiltonian interactions
  - Centered-difference leapfrog-style updates of fixed-point coordinates
    - Ensures reversibility of simulation (no entropy loss)
- Plan is to simulate chaotic dynamical network model with this simulator to validate that it can be used to do logic
  - Visualizations
    - phase portraits, equilibrium distributions
- Current status:
  - Core simulation framework is working
  - Testing on complex networks is still needed
  - Visualizations still needed

DYNAMIC Software Architecture



Some Possible Next Steps

Sandia National Laboratories

- Add an external thermal environment to the model
- Parallelize simulator so that simulating very large networks becomes feasible
- Explore possible implementation technologies
  - Superconducting circuits
  - Other?

Conclusions

Sandia National Laboratories

- Certain physical limits of computing are *fundamental.*
  - Independent of implementation technology!
  - Reflect fundamental aspects of the computing paradigm used.
  - Performing as well as possible requires new computing paradigms!
    - Not simply “better devices.”
- We saw some examples of fundamental limits:
  - Energy dissipation limit from Landauer’s Principle
    - Can be avoided by (at least conditionally-) reversible computing
  - Quantum-mechanical limit on parallel step rate per unit temperature
    - On the order of 15 GHz / degree Kelvin
    - Unavoidable, but still fairly far away
- We must maintain awareness of the above factors when developing future computing technologies
  - It seems likely that the best new technologies will be those that closely reflect the computational structure of physics itself