# Adiabatic, Reversible Computing for Ultra-Power-Efficient DSP

Dr. Michael P. Frank
**mpf@cise.ufl.edu**
CISE/ECE Depts., Univ. of Florida
→ ECE Dept., FSU/FAMU College of Engineering

Visit to Texas Instruments
Dallas, Texas
June 3, 2004

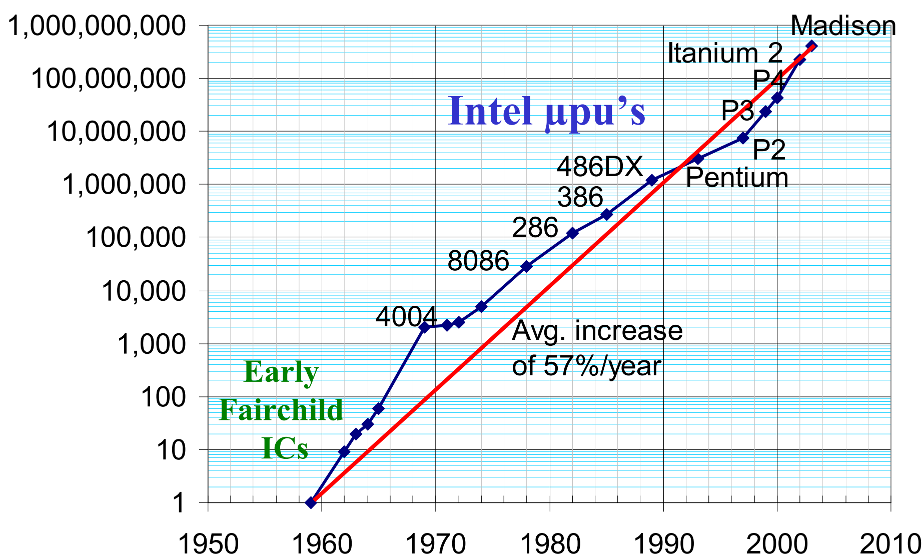
# Summary of Talk

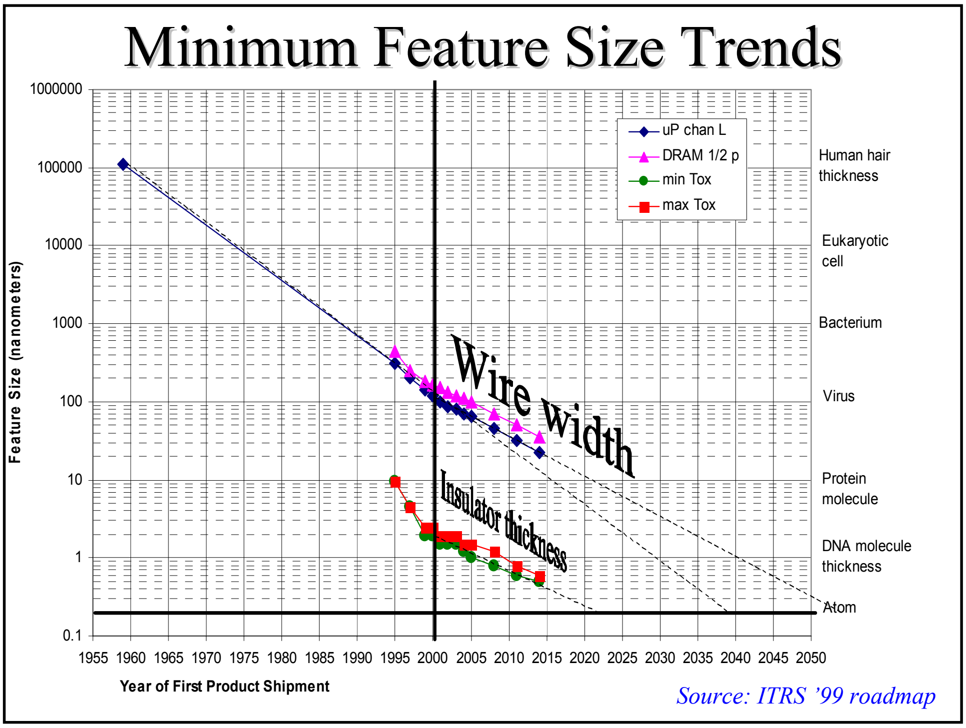
- The traditional *irreversible* digital computing paradigm is rapidly approaching firm limits on performance per unit power consumed.
  - Due to lower bounds on energy dissipated by irreversible switching events.
- The *only* hope for maintaining power-performance scaling beyond the near future is to explore the alternative: mostly *reversible* logic.
  - That is, substantially reversible both thermodynamically and logically.
- Highly *adiabatic* (mostly reversible) logic circuits have been built.
  - UF is presently exploring an integrated CMOS/MEMS technology for this.
- Results imply a prospect for near-term applications offering >10× power-performance boosts at ultra-low power levels, *e.g.*, for DSP.
  - With potential for many orders of magnitude further improvement, in various proposed alternative (but still reversible) technologies on the horizon.
- We invite TI to collaborate with UF & our collaborators on the development of mostly-reversible digital processor architectures.
  - Aimed towards products for ultra-low-power apps in the near term,
  - With potential migration to ultra-high-performance products down the road.

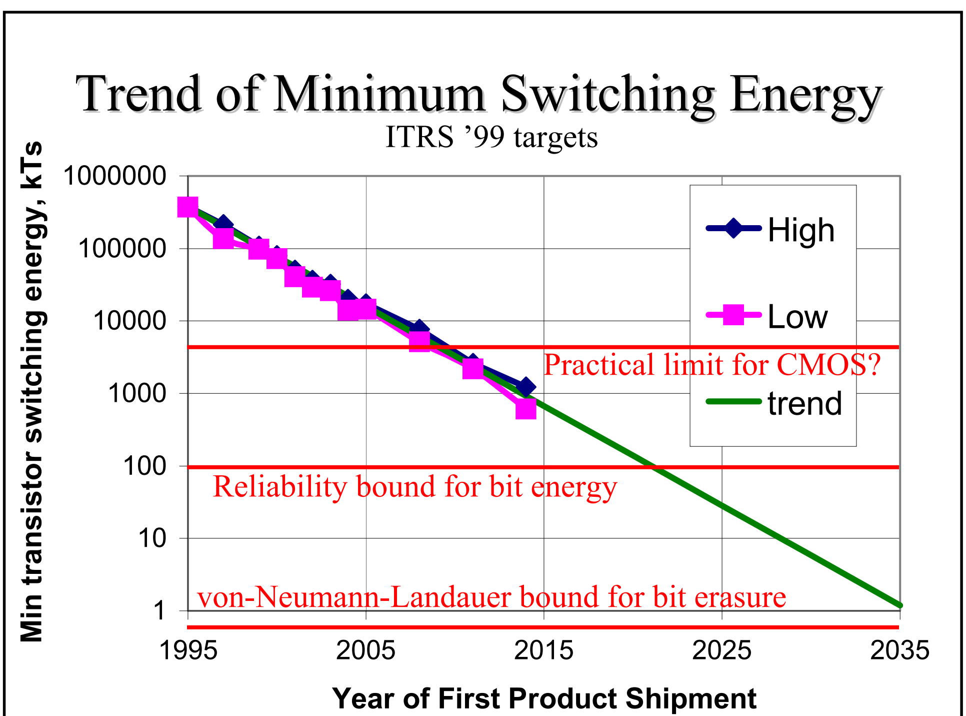
# Present Network of Collaborators

- **University of Florida**: Adiabatic CMOS/MEMS
  - Michael Frank (CISE), Huikai Xie (ECE)
    - with assistance from various TEC lab faculty
- **SUNY Stony Brook**: Reversible supercond. logic
  - Dmitri Averin, A. Semenov (Physics)
- **Notre Dame**: Reversible quantum-dot logic
  - Craig Lent, Peter Kogge
- **NASA**: Theory, spacecraft applications
  - Colin Williams (JPL), others from Ames, Goddard
- **Sandia nat'l lab**: Zettaflops applications, MEMS
  - Erik Debendictis, others.

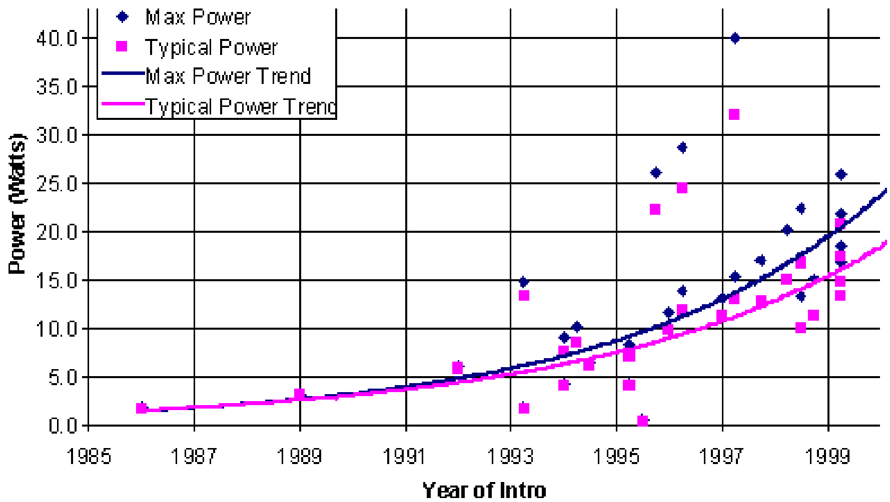
# Moore's Law – Devices per IC




Minimum Feature Size Trends
1000000
100000
10000
1000
100
10
1
0.1
Feature Size (nanometers)
uP chan L
DRAM 1/2 p
min Tox
max Tox
Human hair thickness
Eukaryotic cell
Bacterium
Virus
Protein molecule
DNA molecule thickness
Atom
Wire width
Insulator thickness
1955 1960 1965 1970 1975 1980 1985 1990 1995 2000 2005 2010 2015 2020 2025 2030 2035 2040 2045 2050
Year of First Product Shipment
Source: ITRS '99 roadmap


Trend of Minimum Switching Energy
ITRS '99 targets
Min transistor switching energy, kTs
1000000
100000
10000
1000
100
10
1
High
Low
trend
Practical limit for CMOS?
Reliability bound for bit energy
von-Neumann-Landauer bound for bit erasure
1995
2005
2015
2025
2035
Year of First Product Shipment

# Microprocessor Power Trends

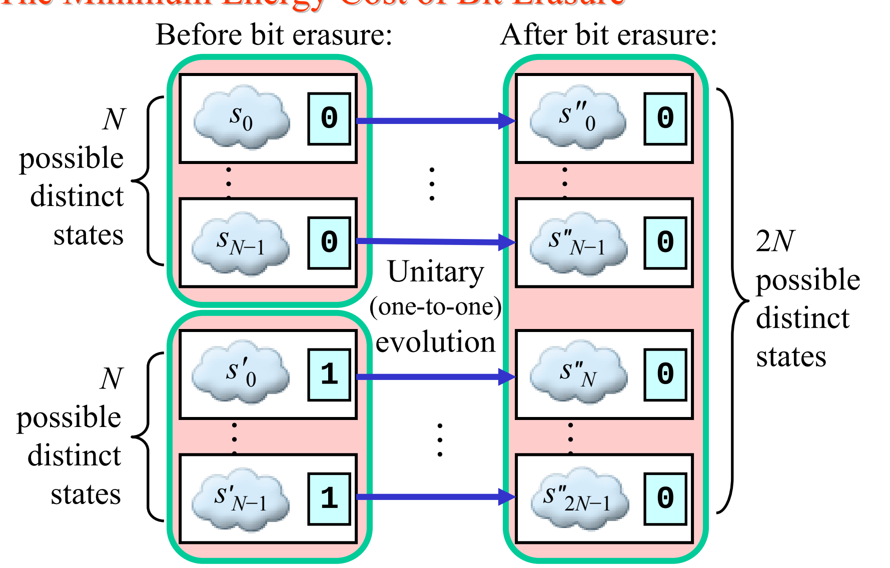



Gunther *et al.*, "Managing the Impact of Increasing Microprocessor Power Consumption," *Intel Technology Journal*, 1st quarter 2001.

# Landauer's (1961) Principle:

## The Minimum Energy Cost of Bit Erasure

Foreshadowed by von Neumann '49




Increase in entropy: $\Delta S = \log 2 = k \ln 2$. Energy dissipated to heat: $T\Delta S = kT \ln 2$

# Reliability Bound on Bit Energy

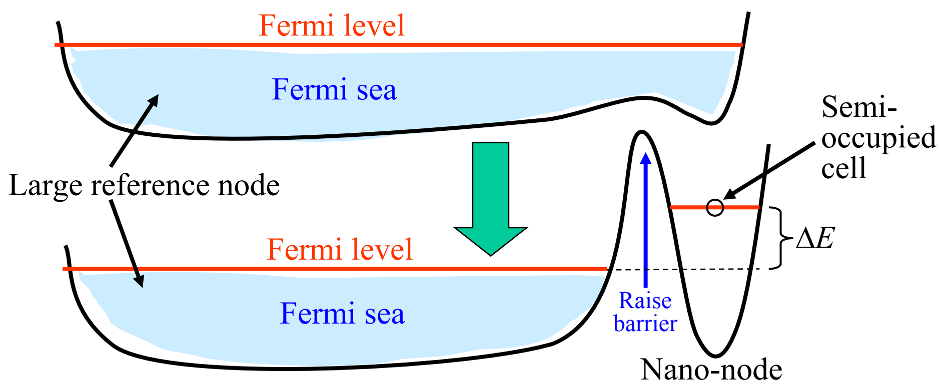
- To reliably store (latch) a bit of data with less than 1 error in $N$ repetitions requires that:
  - In the equilibrium microstate distribution, when latching, the number of accessible microstates leading to the correct stored bit value should be $N$ times the number leading to the incorrect bit value.
    - ∴ There should be $\Delta E \gtrsim k_B T \ln N$ energy difference between storage-cell states having correct and incorrect bit values, at time of latching, in a device at temperature $T$.
      - Follows directly from the Boltzmann distribution.
    - <u>If and when</u> energy of this magnitude later gets *dissipated* by the device, this would lead to an characteristic entropy increase of $\Delta S = \log N = k_B \ln N$.

# Reliability Bound: Numerical Example

- **Example:** Reliability factor of $N=10^{27}$ (1 error in a $10^9$ device processor running for ~3 years at 10 GHz)
  - Associated entropy:
    $\log 10^{27} = k_B \ln 10^{27} \approx 62k_B = 8.6\times10^{-22}$ J/K
  - Heat that must be output to a room-$T$ (300 K) environment:
    $k_B(300\text{ K})\ln 10^{27} = 2.6\times10^{-19}$ J (or 260 zJ, or 1.6 eV)
    - Sounds small, but…
  - If each device dumped this energy at a frequency of 10 GHz,
    - the total power dissipated by an entire $10^9$-device processor is 26 W.
    - Can have at most 4 such processors within a 100 W power budget.
  - Maximum performance: $4\times10^{20}$ device-cyles/sec.
    - or 4 PFLOPS, if processors do 100,000 device-ops per FLOP

## Reliability Bound: Physical Example

- Store a bit by raising an energy barrier to isolate electrons on a nano-island (w. discrete spectrum).
  - Prob. of trapping an extra electron in a state at $\Delta E$ is $1/(1+e^{\Delta E/kT}) \approx e^{-\Delta E/kT}$. (Fermi-Dirac distribution.)

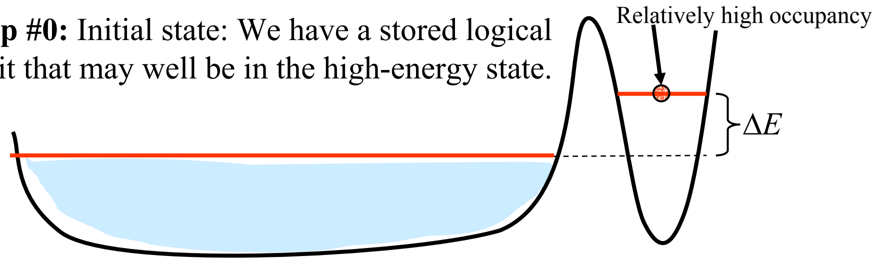



## What Exactly Does this Imply?

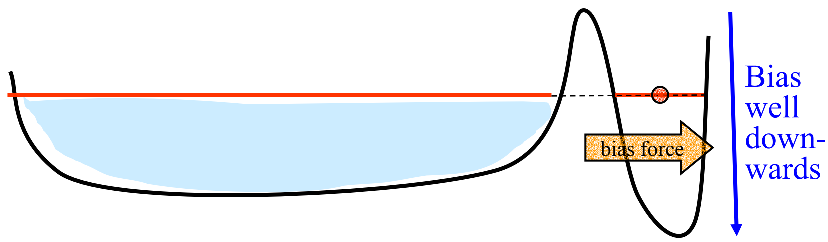
- The reliability bound indeed lower-bounds the energy difference between correct and incorrect bit states at the time that a bit is first stored,
  - And, the “simple, dumb” way to erase a bit is to just remove the potential barrier between states, which returns the stored energy to the heat bath…
- But, this is not the only possible way to erase a bit. The $kT \ln N$ energy need not be dissipated.
  - More cleverly designed erasure mechanisms can reduce the energy dissipation to approach the von Neumann-Landauer bound of $kT \ln 2$ arbitrarily closely, *without* sacrificing reliability!

# Cheap Bit Erasure

**Step #0:** Initial state: We have a stored logical bit that may well be in the high-energy state.

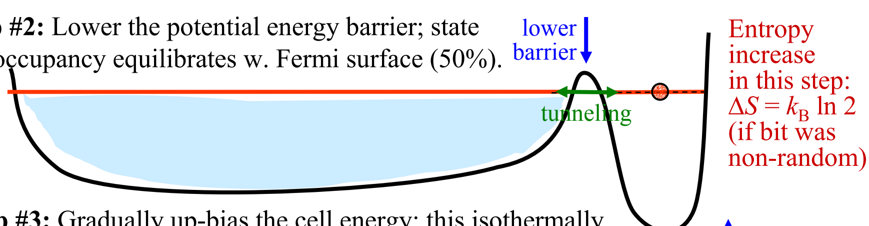



**Step #1:** Apply a bias force to the storage cell to lower the state's energy level to match the reference node's Fermi level.

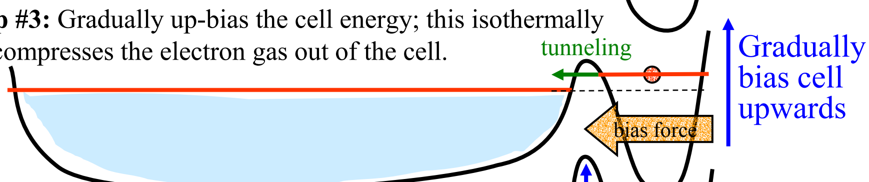



# Cheap Bit Erasure, continued…

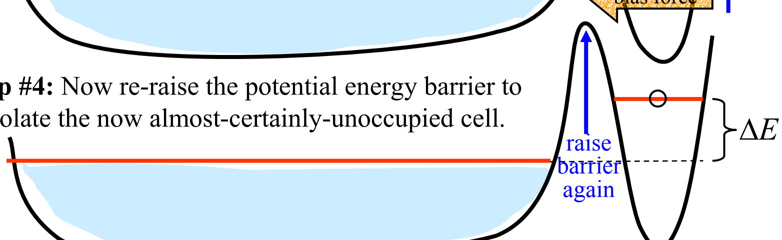
**Step #2:** Lower the potential energy barrier; state occupancy equilibrates w. Fermi surface (50%).




**Step #3:** Gradually up-bias the cell energy; this isothermally compresses the electron gas out of the cell.




**Step #4:** Now re-raise the potential energy barrier to isolate the now almost-certainly-unoccupied cell.




Total entropy generated: $\Delta S \rightarrow k_B \ln 2$

# Reversible Computing

- A *reversible* digital logic operation is:
  - Any operation that performs an invertible (one-to-one) transformation of the device's local digital state.
- Landauer's principle only limits the energy dissipation of ordinary *irreversible* (many-to-one) logic operations.
  - Reversible logic operations can dissipate much less energy,
    - Since they can be implemented in a thermodynamically reversible way.
- In 1973, Charles Bennett (IBM Research) showed how any desired computation can in fact be performed using *only* reversible operations (with no bit erasure).
  - This opened up the possibility of a vastly more energy-efficient alternative paradigm for digital computation.
- After 30 years of sporadic research, this idea is finally approaching the realm of practical implementability…
  - Making it happen is the goal of the RevComp project at UF.

# Adiabatic Circuits

- Reversible logic can be implemented using fairly ordinary voltage-coded CMOS VLSI circuits.
  - With a few changes to the logic-gate/circuit architecture.
- We avoid dissipating most of the circuit node energy when switching, by transferring charges in a nearly *adiabatic* (lit. "without flow of heat") fashion.
  - *I.e.*, asymptotically thermodynamically reversible.
- There are many designs for purported "adiabatic" circuits in the literature, but most of them contain fatal flaws.
  - Many past designers are unaware of (or accidentally failed to meet) all the requirements for true thermodynamic reversibility.

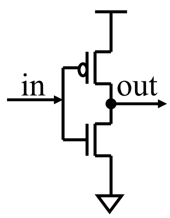
# Conventional Logic is Irreversible

Even a simple NOT gate, as traditionally implemented…

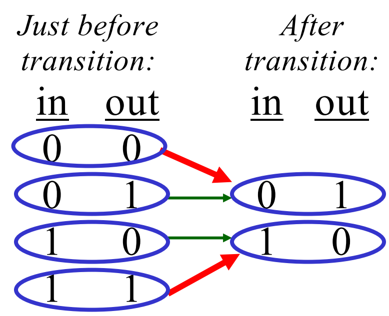
- Logic gate behavior (upon receiving new input):
  - Performs many-to-one transformation of local state!
  - ∴ required to dissipate $\gtrsim kT$, by Landauer principle
  - Incurs $\frac{1}{2}CV^2$ energy dissipation in 2 out of 4 cases.

**Example:**

Static CMOS Inverter:



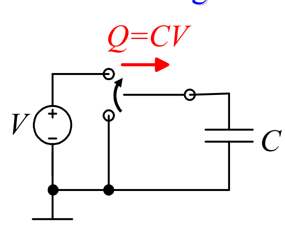

Transformation of local state:




# Conventional *vs.* Adiabatic Charging

For charging a capacitive load $C$ through a voltage swing $V$

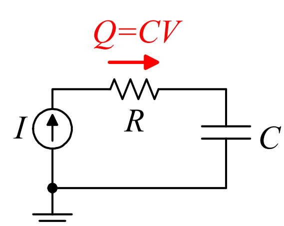
- Conventional charging:
  - Constant voltage source




  - Energy dissipated:

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

- Ideal adiabatic charging:
  - Constant current source



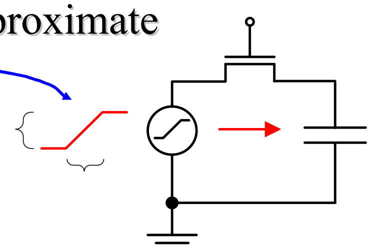

  - Energy dissipated:

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

**Note:** Adiabatic beats conventional by advantage factor $A = t/2RC$.

# Adiabatic Switching with MOSFETs

- Use a voltage ramp to approximate an ideal current source.
- Switch *conditionally*, if MOSFET gate voltage $V_g > V + V_T$ during ramp.



- Can discharge the load later using a similar ramp.
  - Either through the same path, or a different path.

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

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

Exact formula:
$E_{diss} = s\left[1 + s\left(e^{-1/s} - 1\right)\right]CV^2$
given *speed fraction* $s :\equiv RC/t$

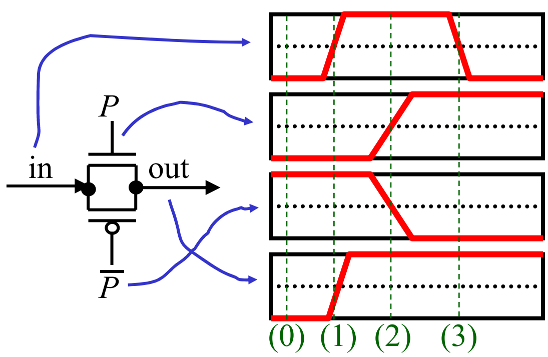
Athas '96, Tzartzanis '98

# Requirements for True Adiabatic Logic

- Avoid passing current through diodes.
  - Crossing the "diode drop" leads to irreducible dissipation.
- Follow a "dry switching" discipline (in the relay lingo):
  - Never turn on a transistor when $V_{DS} \neq 0$.
  - Never turn off a transistor when $I_{DS} \neq 0$. ← Important but often neglected!
- Together these rules imply:
  - The logic design must be logically reversible
    - There is no way to erase information under these rules!
  - Transitions must be driven by a quasi-trapezoidal waveform
    - It must be generated resonantly, with high $Q$
- Leakage power must also be kept manageable.
  - Because of this, the optimal design point will *not* necessarily use the smallest devices that can ever be manufactured!
    - The smallest devices may have insoluble problems with leakage.

# A Simple Reversible CMOS Latch

- Uses a single standard CMOS *transmission gate* (T-gate).
- Sequence of operation:
  - (0) input level initially matches latch 'contents' (output);
  - (1) input changes gradually → output follows closely;
  - (2) latch closes, charge is stored dynamically (node floats);
  - (3) input signal can now be removed



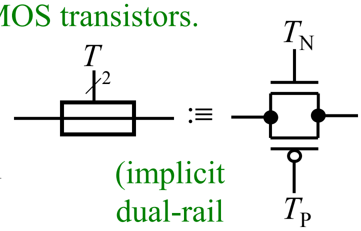

| *Before input:* | | *Input arrived:* | | *Input removed:* | |
|---|---|---|---|---|---|
| in | out | in | out | in | out |
| 0 | 0 | 0 | 0 | 0 | 0 |
| | | 1 | 1 | 0 | 1 |

- Later, we can reversibly "unlatch" the data with an exactly time-reversed sequence of steps.

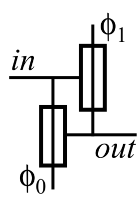
# 2LAL: 2-level Adiabatic Logic

A pipelined fully-adiabatic logic invented at UF (Spring 2000), implementable using ordinary CMOS transistors.

- Use simplified T-gate symbol:

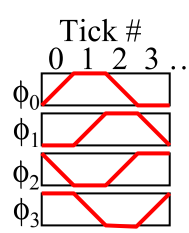



- Basic buffer element:
  - cross-coupled T-gates:
    - need 8 transistors to buffer 1 dual-rail signal

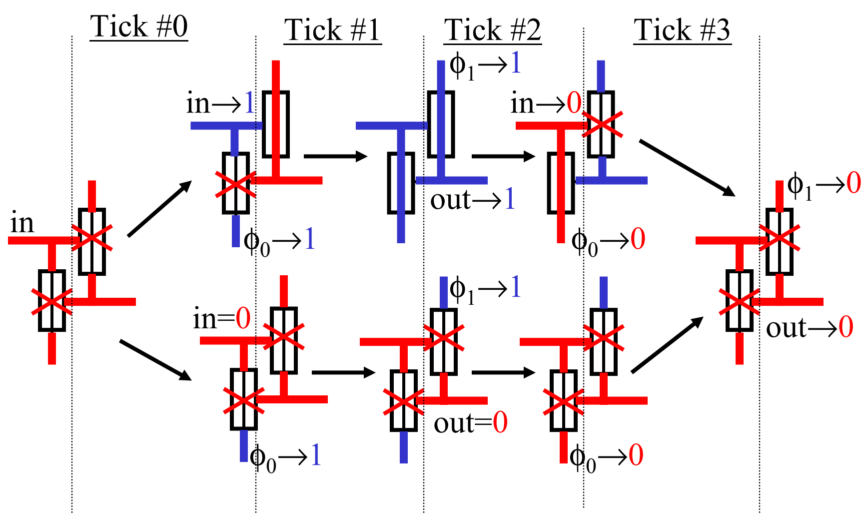



(implicit dual-rail encoding everywhere)

- Only 4 timing signals $\phi_{0\text{-}3}$ are needed. Only 4 ticks per cycle:
  - $\phi_i$ rises during ticks $t \equiv i$ (mod 4)
  - $\phi_i$ falls during ticks $t \equiv i+2$ (mod 4)



# 2LAL Cycle of Operation

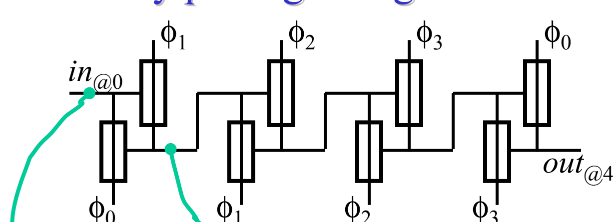



# 2LAL Shift Register Structure

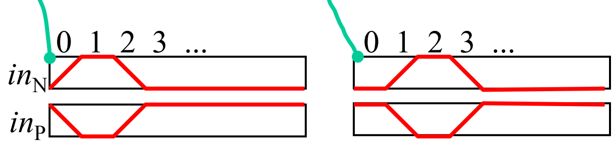
Animation:

2lal.swf

- 1-tick delay per logic stage:

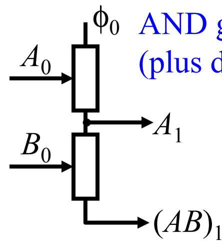



- Logic pulse timing and signal propagation:



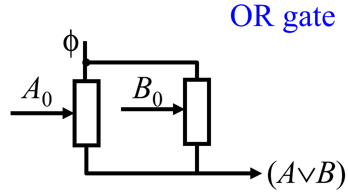
# More Complex Logic Functions

- Non-inverting multi-input Boolean functions:

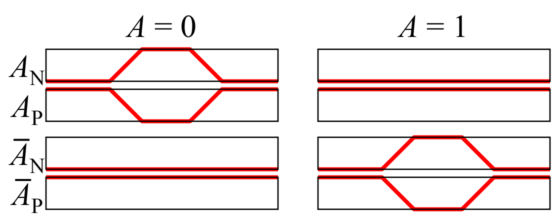



AND gate
(plus delayed *A*)

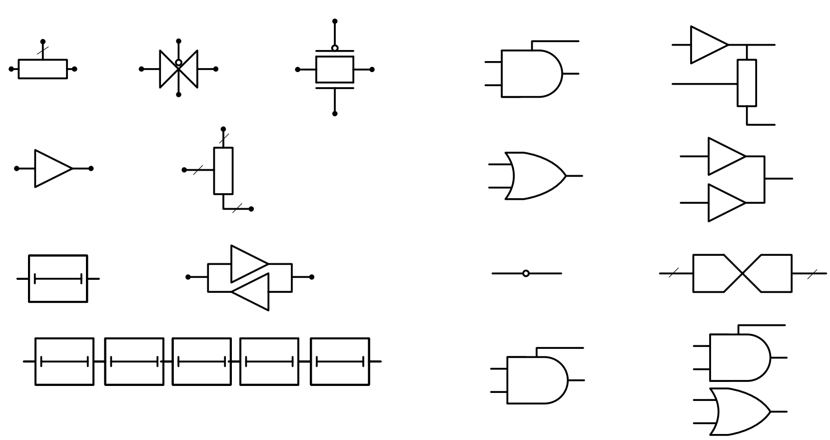
OR gate




- For inverting functions, we must use a quad-rail logic encoding:
  - To invert, just swap the rails!
    - Zero-transistor "inverters."

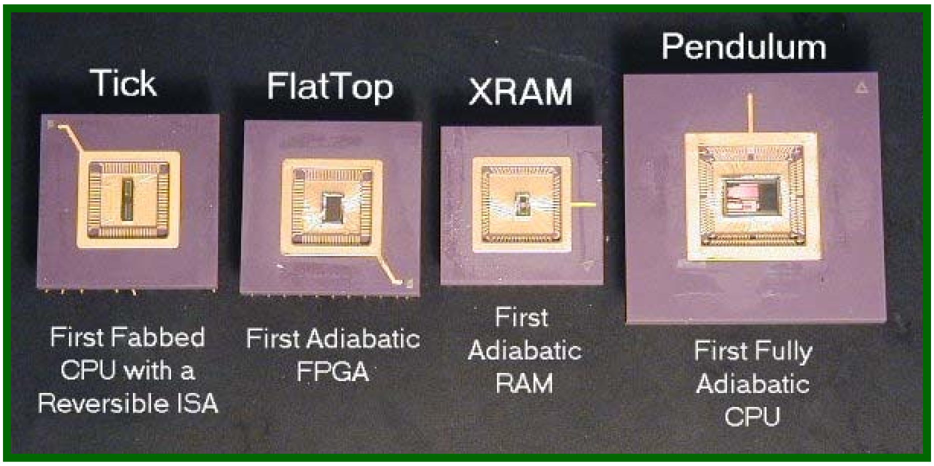



# A Graphical Notation for 2LAL

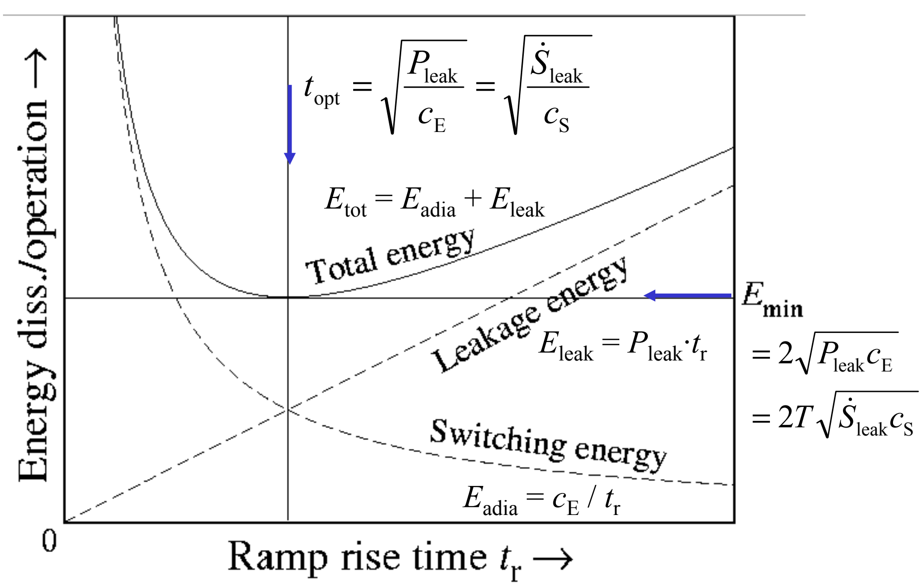


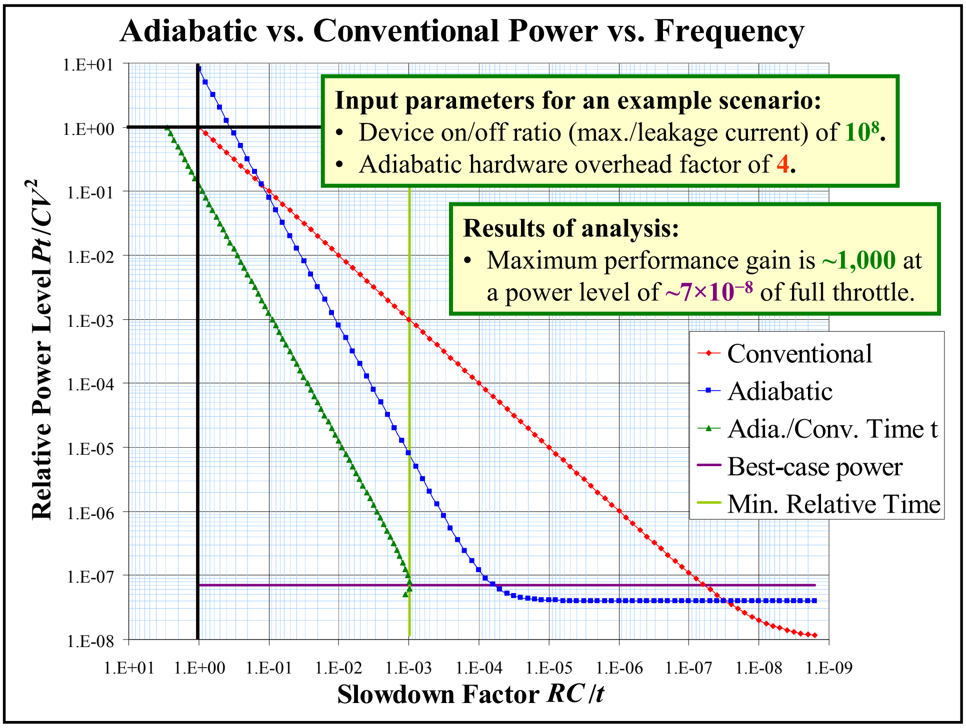
# Reversible and/or Adiabatic VLSI Chips Designed @ MIT, 1996-1999

By Frank and other then-students in the MIT Reversible Computing group, under CS/AI lab members Tom Knight and Norm Margolus.




# Minimum Losses w. Leakage






# Adiabatic Performance Boost

- Approx. performance gain factor of adiabatics, given power level is:

$$g_{\text{adia}} = \frac{\sqrt{P_{\text{g,full}}\left(\frac{P_{\text{g,max}}}{O_{\text{adia}}} - P_{\text{g,lk}}\right)}}{2(P_{\text{g,max}} - P_{\text{g,lk}})}$$

- Where:
  - $P_{\text{g,full}} = E_{\text{g,sw}} f_{\text{max}}$ = "Full throttle" switching power per logic gate,
    - $E_{\text{g,sw}} = CV^2$ switching energy per logic gate
    - $f_{\text{max}}$ = "Full throttle" switching frequency $1/RC$ of gates
  - $P_{\text{g,max}}$ = Maximum allowed power dissipation per gate, imposed by constraints on application's power and/or cooling system
  - $O_{\text{adia}}$ = Hardware overhead factor of adiabatic logic design
  - $P_{\text{g,lk}}$ = Leakage power dissipation per gate in given technology
- This is maximized when $P_{\text{g,max}} = P_{\text{g,lk}}(2O_{\text{adia}} - 1)$, in which case we have:
  - where $R_{\text{on/off}} = I_{\text{on}}/I_{\text{off}} = P_{\text{full}}/P_{\text{lk}}$

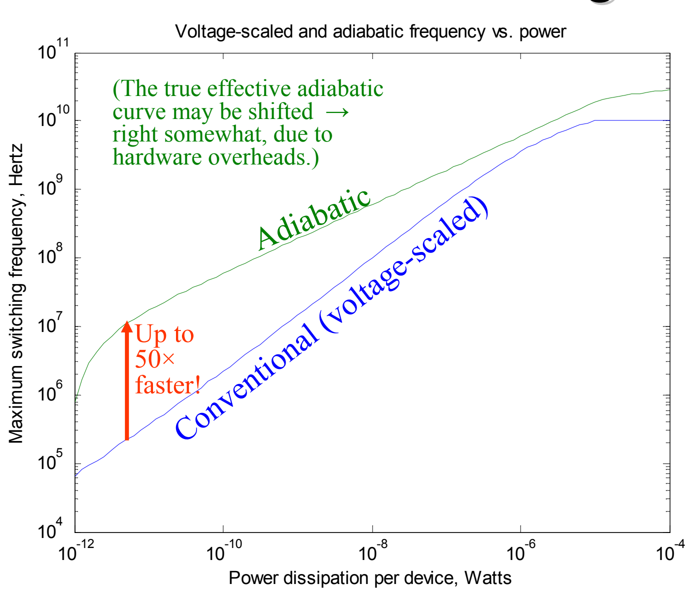
$$g_{\text{adia}} = \frac{1}{4}\sqrt{\frac{R_{\text{on/off}}}{O_{\text{adia}}(O_{\text{adia}} - 1)}}$$

- This is >1 when $16\, O_{\text{adia}}(O_{\text{adia}} - 1) < R_{\text{on/off}}$ of transistor technology

# Example Ultra-Low-Power Scenario

- Technology scenario:
  - ITRS hp65 (65 nm half-pitch) technology node (expect ~2007).
- Application scenario:
  - 1Mgate processor chip (*e.g.* TI's C6000 line of GHz DSPs)
  - Requirement for ≲ 100 µW processor power dissipation
- Leakage per NAND gate in hp65: ~13 pW.
  - ∴ chip would dissipate ~13 µW even at zero frequency!
- Irreversibly switching NAND gate's output takes 250 aJ.
  - 1 million *active* gates would dissipate 250 pJ per clock cycle
    - 87 µW switching power constraint → max. freq. ~350 kHz!
      - Max. NAND transition rate = 23 GHz, slowdown ~66,000 ×
- **Adiabatic solution:** Using overhead factor 4×
  - Run clock at 35 MHz instead of 350 kHz (**100×** faster!)
    - But note this is still 660 times slower than max transition frequency.
  - ∴ each switching op dissipates only ~1/660th the $CV^2$ energy
    - Or, ~1/160th even after the 4× logic overhead is included
  - Leakage power: ~50 µW, switching power: ~50 µW.

# Adiabatics vs. Voltage Scaling



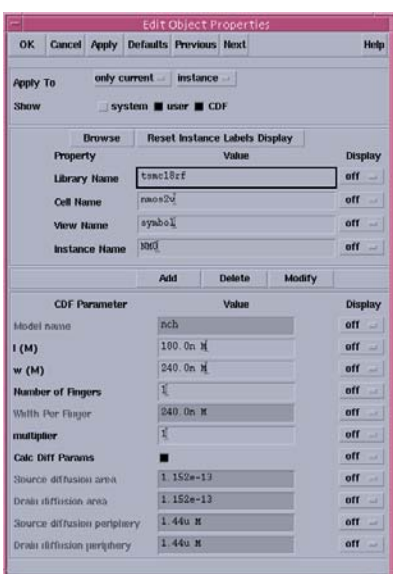

- Technology:
  - TSMC 0.18µm (a few years old)
- Optimum benefit obtained at:
  - 6.3 pW/device
  - 50× speedup!
- Adiabatic:
  - 12 MHz @ 1.8V
- Conventional:
  - 250 kHz @ .24V

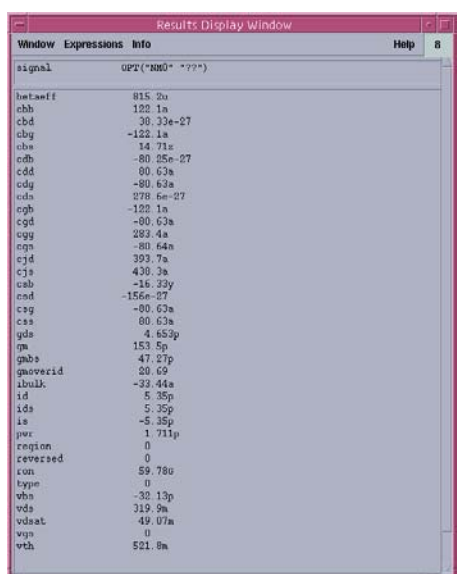
- Maximum speedup can be increased arbitrarily by using higher-threshold devices (and/or low operating temperatures).

# Cadence simulation results

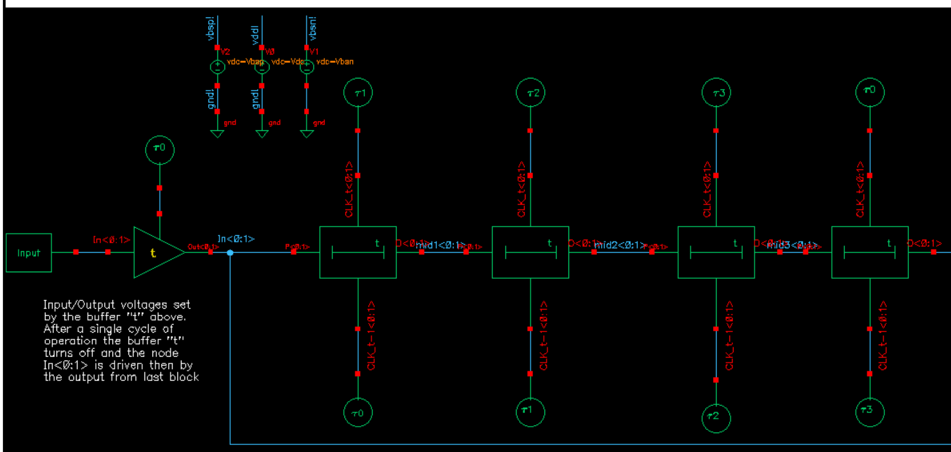
Work by AdiaMEMS project students:
Krishna Natarajan
Venkiteswaran Anantharam
(UF ECE Dept.)

# Cadence parameter windows



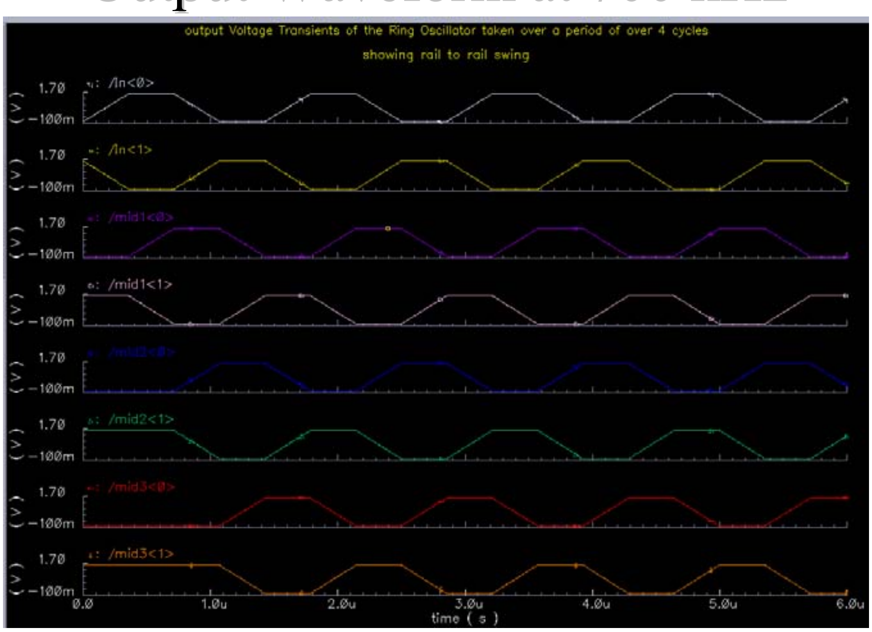



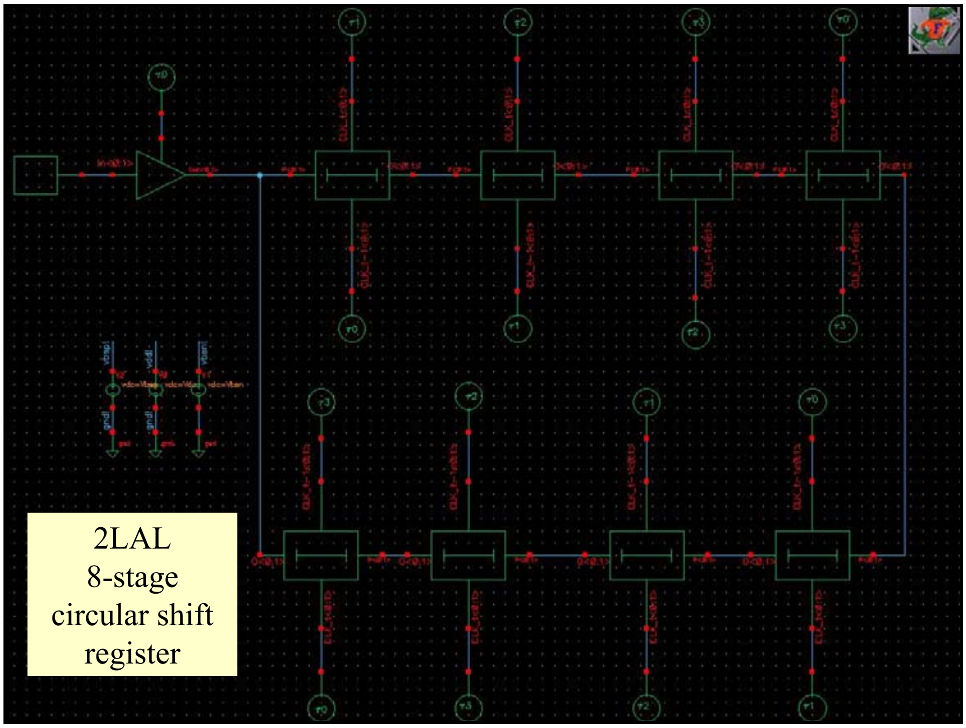
# Cadence schematic for the 4-stage 2LAL circular reversible shift register



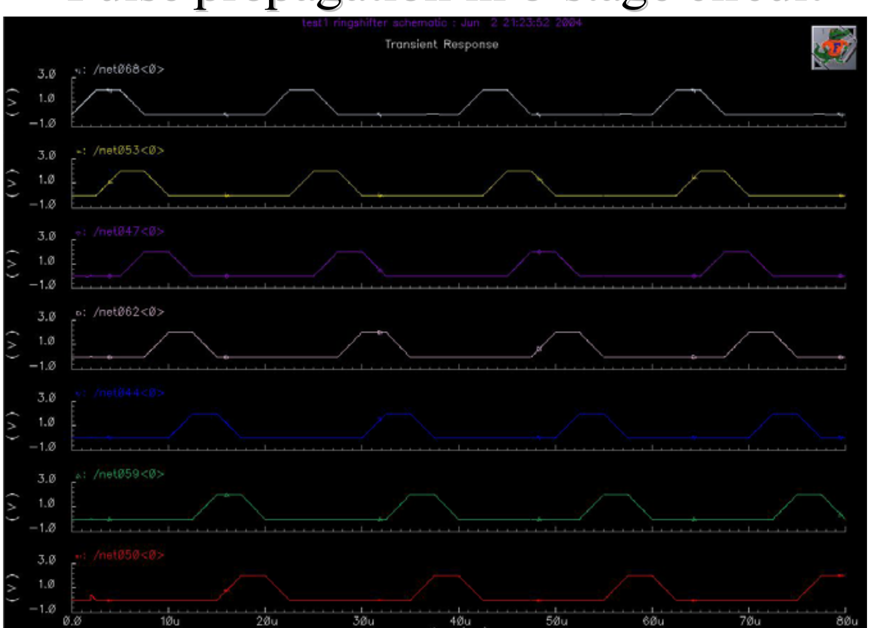

- Propagates a dual-rail voltage pulse circularly through 4 2LAL delay/buffer stages. Pulse propagates a full 4 stages in 1 clock cycle.
- Truly, fully adiabatic operation. All transitions take place over a full ¼ cycle.
- Totally symmetric, reversible design: Direction of pulse propagation would be reversed if clock phases were inverted.

# Output Waveform at 700 kHz

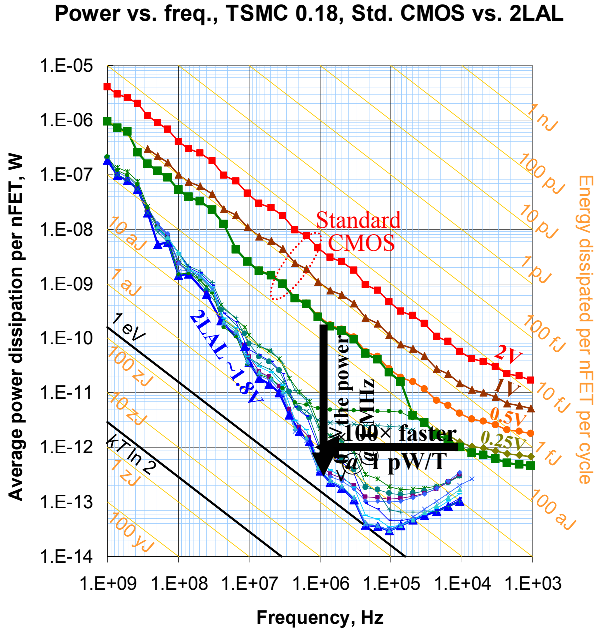





## Pulse propagation in 8-stage circuit



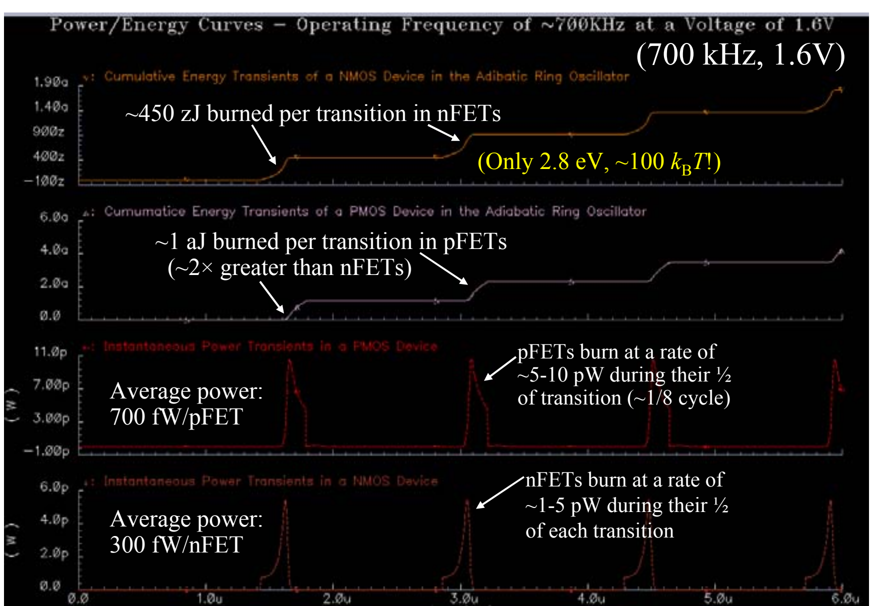
# Simulation Results from Cadence



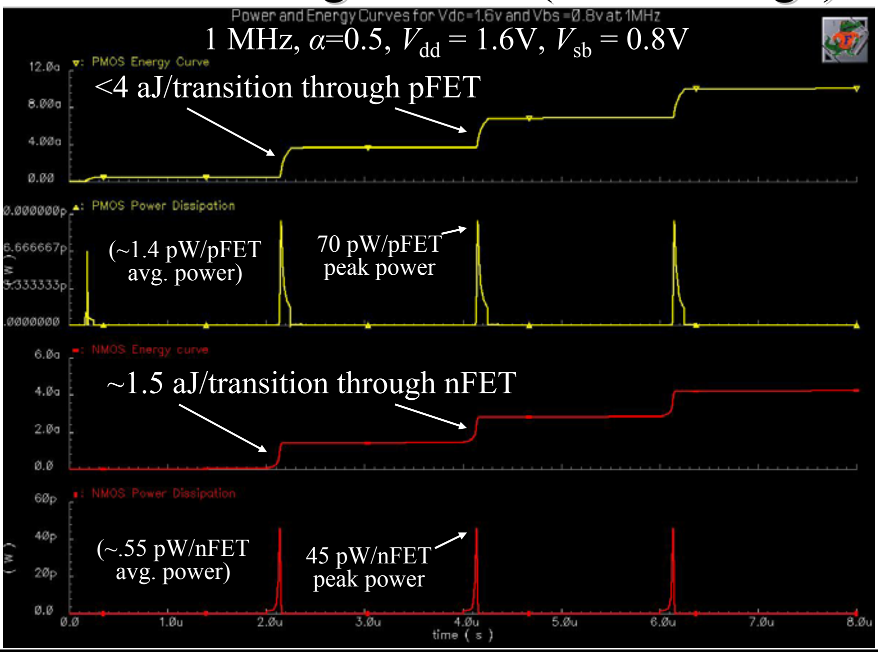

**Assumptions & caveats:**

- Assumes ideal trapezoidal power/clock waveform.
- Minimum-sized devices, 2λ×3λ
  - * .18 µm×.24 µm
- nFET data is shown
  - * pFETs data is very similar
- Various body biases tried
  - * Higher $V_{th}$ suppresses leakage
- Room temperature operation.
- Interconnect parasitics have not yet been included.
- Activity factor (transitions per device-cycle) is 1 for CMOS, 0.5 for 2LAL in this graph.
- Hardware overhead from fully-adiabatic design style is not yet reflected
  - * ≥4× transistor-tick hardware overhead in known reversible CMOS design styles

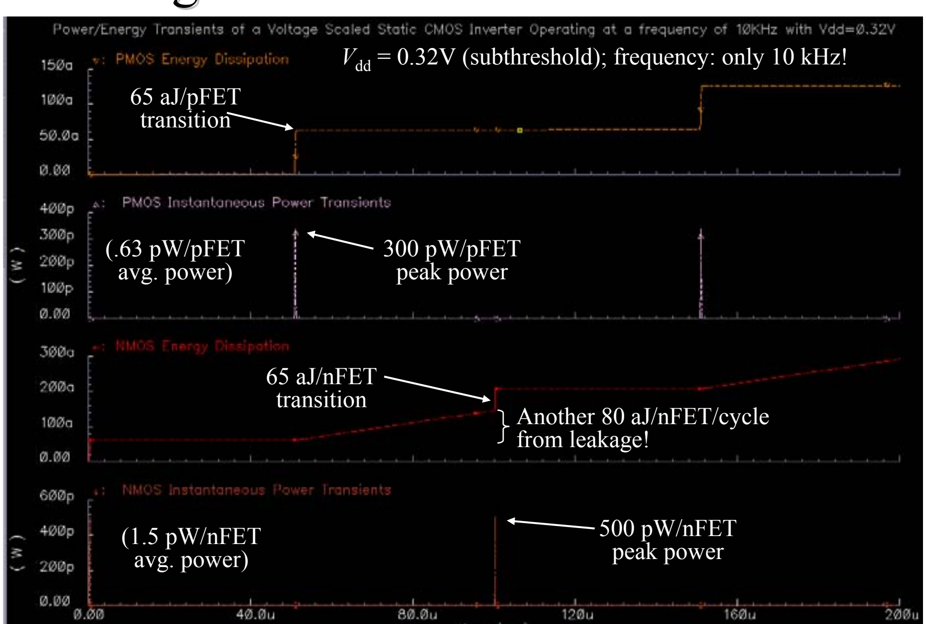
# Example energy & power transients



# Power in 8-stage circuit (incl. leakage)

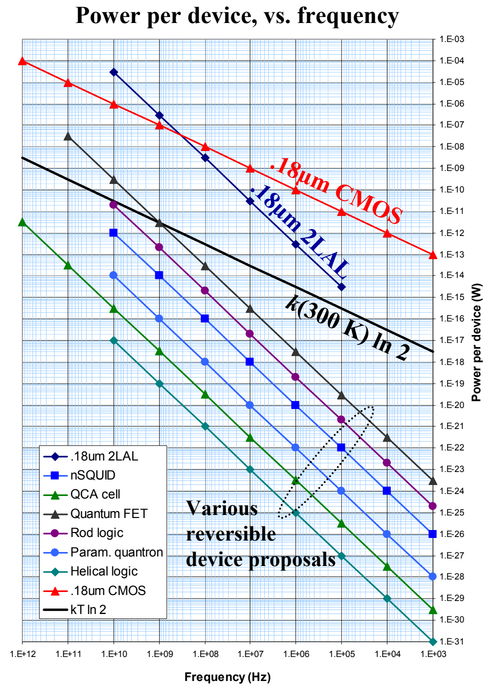



# Voltage-scaled static CMOS behavior



# Plenty of Room for Device Improvement

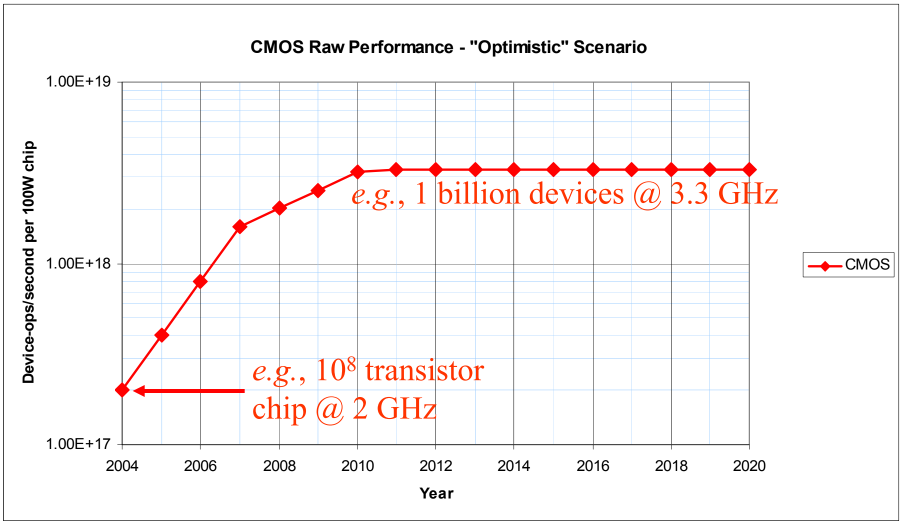
- Irreversible device technology has at most ~3-4 orders of magnitude of power-performance improvements remaining.
  - And then, the firm $kT \ln 2$ limit is encountered.
- But, a wide variety of proposed reversible device technologies have been analyzed by physicists.
  - With theoretical power-performance up to 10-12 orders of magnitude better than today's CMOS!
    - Ultimate limits are unclear.




# A Fairly Conventional "Optimistic" Technology Scenario for CMOS

- Suppose device lengths are cut in half every 3 years…
  - From 90 nm today down to 22 nm node in 2010 (then stop).
  - Node capacitances, gate delays also decrease accordingly…
- "Technology boosters" such as high-κ dielectrics & novel FET structures (FinFET, surround-gate, *etc.*) keep leakage power manageable, for a little while…
  - However, note the absolute minimum room-$T$ subthreshold slope for FETs will remain 60 mV/decade! $(= (kT/q) \log 10)$
    - Assume this point is also reached by around 2007.
- Voltages then reach a minimum of ~0.5V in 2007.
  - Can't go lower while keeping on/off ratio above $10^8$ level!
    - A minimum level chosen so as to keep leakage manageable
- Now, consider what all this implies about future chip performance, given a 100 W maximum power level…
  - Let max raw performance = 100 W / ($\frac{1}{2}CV^2$ gate energy)

# Not much life left for standard CMOS…



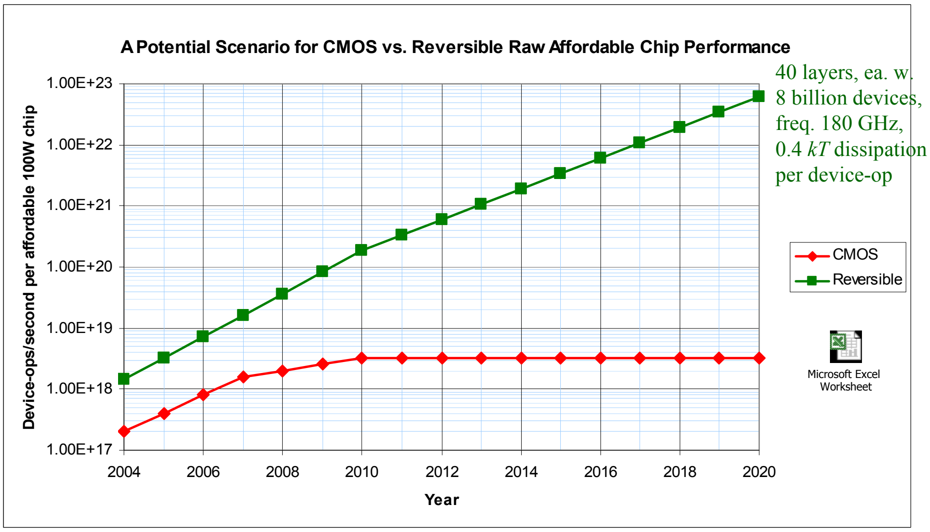

Even if the leakage problem were solved, the ~100 *kT* limit for reliable switching is only another factor of 70 beyond this point.

# A Potential Scaling Scenario for Reversible Computing Technology

Make same assumptions as previously, except:

- Assume *energy coefficient* (energy diss. / freq.) of reversible technology continues declining at historical rate of 16× / 3 years, through 2020.
  - For adiabatic CMOS, $c_E = CV^2RC = C^2V^2R$.
    - This has been going as $\sim\ell^4$ under constant-field scaling.
  - But, requires new devices after CMOS scaling stops.
    - However, many candidates are waiting in the wings…
- Assume number of affordable *layers* of active circuitry per chip (or per package, *e.g.*, stacked dies) doubles every 3 years, through 2020.
  - Competitive pressures will tend to ensure this will happen, esp. if device-size scaling stops, as assumed.

# Result of Scenario




Note that by 2020, there could be a factor of 20,000× difference in *raw* performance per 100W package. (*E.g.*, a 100× overhead factor from reversible design could be absorbed while still showing a 200× boost.)

**UF CONFIDENTIAL – PATENT PENDING**

# MEMS Resonator Concept

A potential approach for efficiently driving adiabatic logic transitions

# The Power Supply Problem

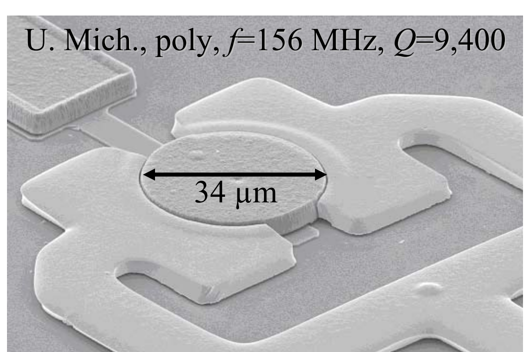
- In adiabatics, the factor of reduction in energy dissipated per switching event is limited to (at most) the *Q* factor of the clock/power supply.

$$Q_{\text{overall}} = (Q_{\text{logic}}^{-1} + Q_{\text{supply}}^{-1})^{-1}$$

- Electronic resonator designs typically have low *Q* factors, due to considerations such as:
  - Energy overhead of switching a clamping power MOSFET to limit the voltage swing of a sinusoidal *LC* oscillator.
  - Low coil count, substrate coupling in integrated inductors.
  - Unfavorable scaling of inductor *Q* with frequency.
- Our proposed solution:
  - Use electromechanical resonators instead!

# MEMS (& NEMS) Resonators

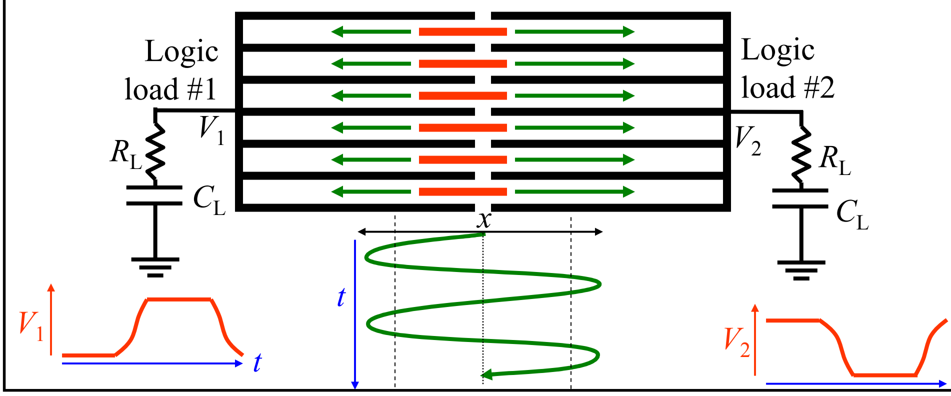
- State of the art of technology demonstrated in lab:
  - Frequencies up to the 100s of MHz, even GHz
  - *Q*'s >10,000 in vacuum, several thousand even in air!
- An important emerging technology being explored for use in RF filters, *etc.*, in communications SoCs, *e.g.* for cellphones.



**UF CONFIDENTIAL – PATENT PENDING**

# Original Concept

- Imagine a set of charged plates whose horizontal position oscillates between two sets of interdigitated fixed plates.
  - Structure forms a variable capacitor and voltage divider with the load.
- Capacitance changes substantially only when crossing border.
  - Produces nearly flat-topped (quasi-trapezoidal) output waveforms.
  - The two output signals have opposite phases (2 of the 4 φ's in 2LAL)

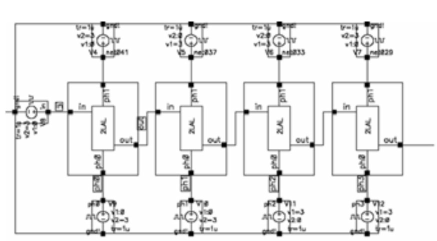



**UF CONFIDENTIAL – PATENT PENDING**

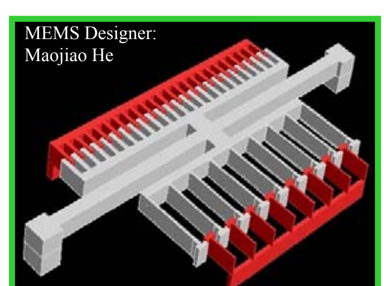
# MEMS Resonant Power Supply for Ultra-Low-Power Adiabatic Circuits

A.k.a. The "AdiaMEMS" Project

- Part of CISE's Reversible & Quantum Computing group
  - Collab. with Huikai Xie (MEMS, ECE dept.)
- **Goal:** Demonstrate orders-of-magnitude improvement in power-performance efficiency of digital CMOS circuits.
  - Based on reversible logic in adiabatic circuits powered by high-quality custom microelectromechanical resonators.
- **Funding:** $40K seed grant from SRC's Cross-Disciplinary Semiconductor Research (CSR) Program



VLSI designer: Krishna Natarajan

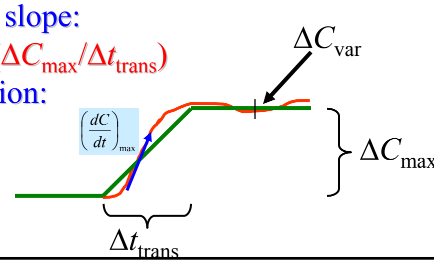


MEMS Designer:
Maojiao He

UF CONFIDENTIAL – PATENT PENDING

# Key Characteristics of Resonator

- **Goal:** Produce a near-ideal trapezoidal output voltage waveform resonantly, with high $Q$.
- To be optimized with logic: Resonant frequency $f$.
- Key resonator figures of merit:
  - Effective quality factor: $Q_{\text{eff}} = E_{\text{trans}}/E_{\text{diss}}$.
  - Area efficiency: $E_A = E_{\text{trans}}/A$.
- Key resonator figures of demerit:
  - Maximum relative transition slope:
    $$s_{\max} = (dC/dt)_{\max} \,/\, (\Delta C_{\max}/\Delta t_{\text{trans}})$$
  - Fractional capacitance variation:
    $$v_C = \Delta C_{\text{var}} \,/\, \Delta C_{\max}$$




UF CONFIDENTIAL – PATENT PENDING

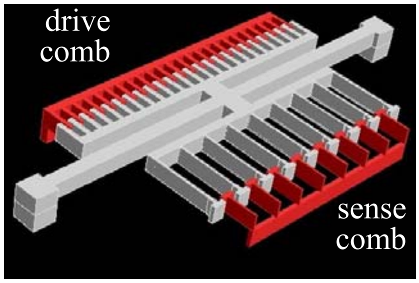
# First MEMS Technology Tried

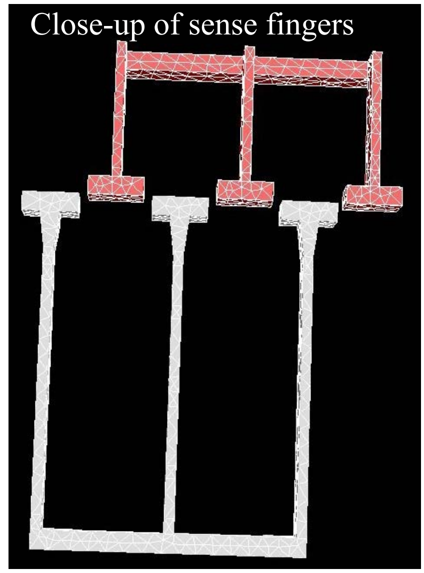
- MEMS process donated by Robert Bosch corp.
- It is a thin-film technology
  - Though a multi-layer, bulk single-crystal process can be expected to do better.
- Integrated CMOS/MEMS devices will eventually be available in this process.
  - However our initial design was dual-die
    - CMOS side was not mature yet in this process
- Minimum etched structure width: $\lambda = 0.5$ μm
- Minimum etched gap size: $d = 0.1$ μm

UF CONFIDENTIAL – PATENT PENDING

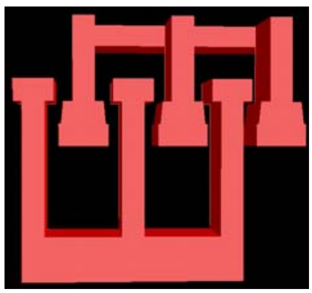
# Some Early Resonator Designs

By Ph.D. student Maojiao He, under supervision of Huikai Xie




Another finger design

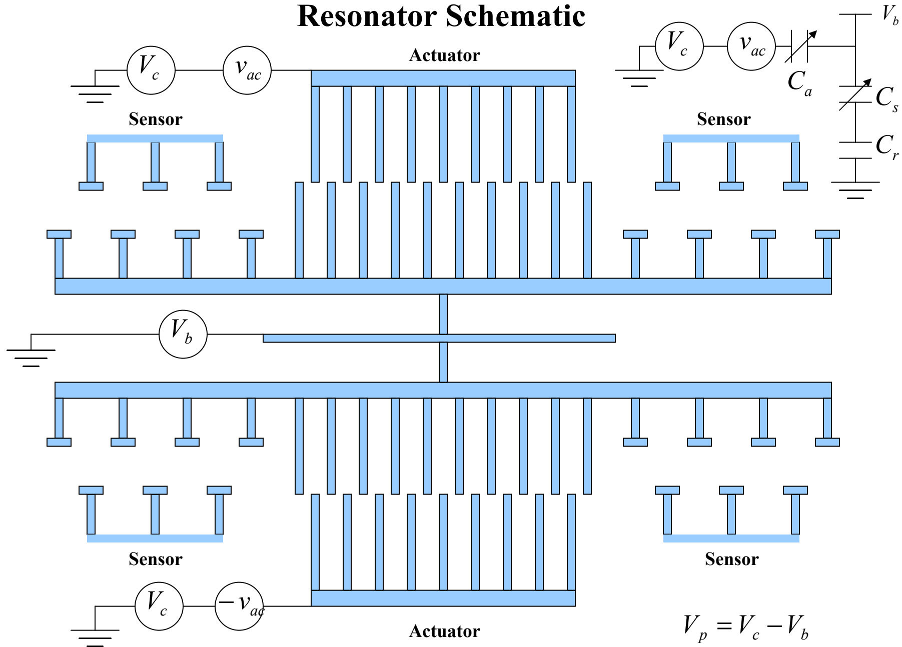


UF CONFIDENTIAL – PATENT PENDING

# Analysis of initial MEMS design

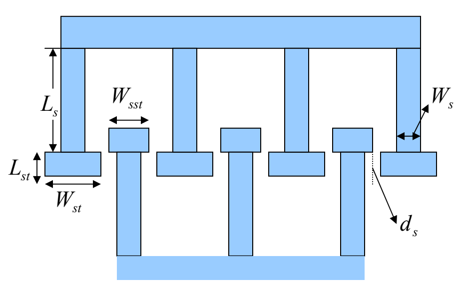
## UF CONFIDENTIAL – PATENT PENDING

### Resonator Schematic




## UF CONFIDENTIAL – PATENT PENDING

### Sensor Design




$d_s = d \qquad W_s = \lambda$

$L_{st} = \lambda \qquad X = 8L_{st}$

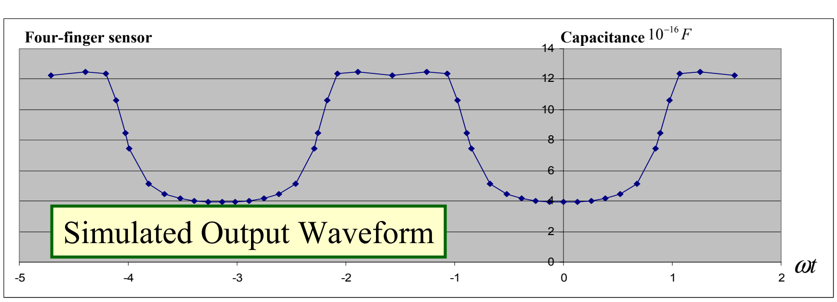
$W_{st} = W_s + 4d$

$W_{sst} = W_s + 2d$

$L_s >> d\ (L_s = 20d)$

(Early design w. thin fingers)

$4C_{sf} \approx 8 \times 10^{-16}\ F$



# Dissipation in Resonator

Ways to minimize some major sources of dissipation:

- Air damping:
  - Vacuum packaging, small size, or optimize airflow
- Clamping losses to the substrate:
  - Locate support at a nodal point of vibration mode
  - Use impedance-mismatched supports to reflect energy back
- Thermoelastic dissipation (heat flow resulting from nonuniform strain):
  - Small size
  - Use stiff, high thermal conductivity materials (Si, diamond?)
  - Utilize modes with uniform compression/expansion
- Surface loss mechanisms:
  - Avoid layered structures (thin-film interfaces) at surfaces
- Intrinsic material losses:
  - Prefer single-crystal materials

# Status / Plans for Near Future

- A small prototype resonator design was taped out in a post-CMOS MEMS process (TSMC .35)
  - Will be tested this summer.
- Improved resonator designs afforded by a suitably modified post-CMOS process flow are being developed.
  - I will briefly review some aspects.
- Process donation has been obtained from MOSIS for fabricating a integrated CMOS/MEMS test chip (~$20k).
  - Resonator driving a simple 2LAL shift register or adder pipeline
  - Tape-out scheduled for July 26.
- Test the various parts separately, & together.
  - Characterize power dissipation using sensitive calorimetry techniques.

# CMOS-MEMS Process

Huikai Xie

Department of Electrical and Computer Engineering
University of Florida
Gainesville, FL 32611
Email: hkxie@ece.ufl.edu

# Outline

- ❑ Introduction
- ❑ CMOS-MEMS Process
- ❑ 3-D Sensing and Actuation
- ❑ CMOS-MEMS Inertial Sensors
- ❑ Summary

# Why DRIE CMOS-MEMS?

***Thin-film micromachining technology***

- ✓ **On-chip electronics integration**
- ✓ **Multiple axis integration**
- ✓ **Size limitation due to residual stress**
- **ADI, Bosch, Carnegie Mellon, Samsung, Sandia, UC-Berkeley**

***Bulk micromachining technology***

- ✓ **Large mass**
- ✓ **No integrated interface circuitry**
- ✓ **Wafer-to-wafer bonding, two-side alignment**
- **Bosch, Draper, JPL, Murata , Samsung**

**Our approach:**

***DRIE CMOS-MEMS process***

- ✓ **On-chip electronics**
- ✓ **Large mass**

# Why CMOS-MEMS?

- ✓ "Smart" on-chip CMOS circuitry
- ✓ Multi-vendor accessibility
- ✓ Scalability
- ✓ Compact size
- ✓ More functions
- ✓ Low cost

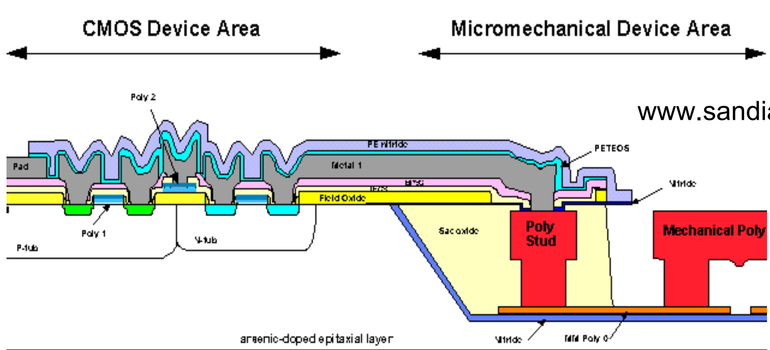
– MEMS structures can be made

- Before CMOS processes ("pre-CMOS")
- In-between CMOS processes ("intermediate-CMOS")
- After CMOS processes ("post-CMOS")

# CMOS-MEMS Processes

| | MEMS planarity | Vendor accessibility | Contami-nation | Temperature budget | |
|---|---|---|---|---|---|
| Pre-CMOS | Best | Limited | Yes | No | Sandia National Lab |
| Intermediate-CMOS | Good | Very limited | Yes | Yes | Analog Devices, Inc. |
| Post-CMOS | Varies | Good | No | Yes | Berkeley<br>CMU<br>UF<br>ETH |

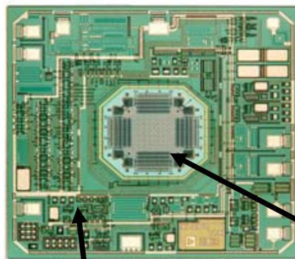
## Sandia National Laboratories iMEMS




www.sandia.gov

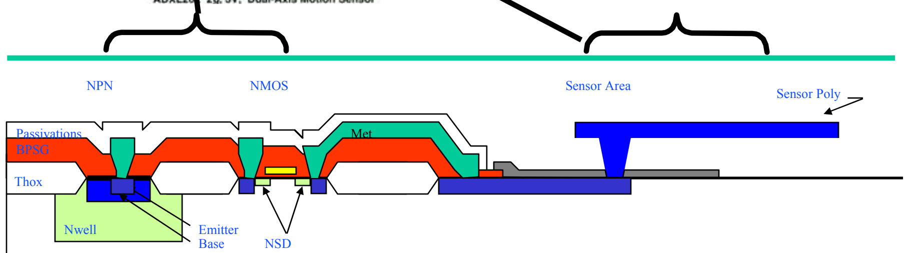
- Pre-etched trench to house MEMS structures
- CMP to planarize the wafer for regular CMOS processing
- Wet etch to release MEMS structures
- Need a dedicated production line

## Analog Devices, Inc. BiMEMS



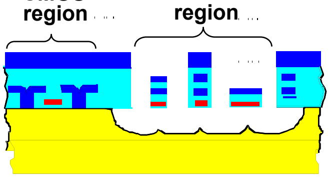
ADXL20 2g, 3V, Dual-Axis Motion Sensor

- Form transistors on bare wafers first
- Then deposit and anneal MEMS structural materials
- No CMP needed
- Only one interconnect metal layer
- Wet etch to release MEMS structures
- Need a dedicated production line




Courtesy of Mr. John Geen of Analog Devices, Inc.

## Post CMOS-MEMS Process (thin-film)




G. Fedder *et al.*, Sensors & Actuators A, v.57, no.2, 1996

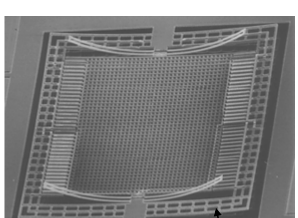
✓No lithography needed

✓Integrated CMOS circuitry

✓Low parasitic capacitance to substrate

✓High wiring flexibility

✓Curling can be matched



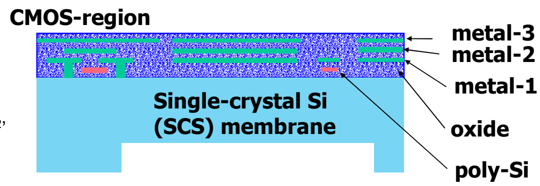
Curl matching frame

- Curling is still an issue
  - Size limitation
  - Temperature performance
- No bottom electrode for vertical capacitive sensing

# Post CMOS-MEMS Process (DRIE)

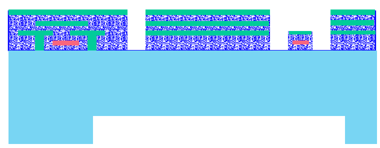
**(a) Backside etch**

**STS**: 12-sec etching
130-sccm $SF_6$, 13-sccm $O_2$,
23 mT, 600 W coil power,
12 W platen power;
8-sec passivation
85-sccm $C_4F_8$, 12 mT, 600 W
coil power, 0 platen power.

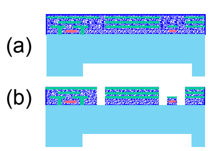



**(b) Oxide etch**

**PlasmaTherm-790:**
22.5-sccm $CHF_3$, 16-sccm
$O_2$, 100 W, 125 mT for 125
minutes and then 100 mT for
10 minutes.

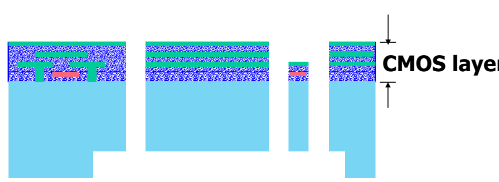


# Post CMOS-MEMS Process (DRIE)



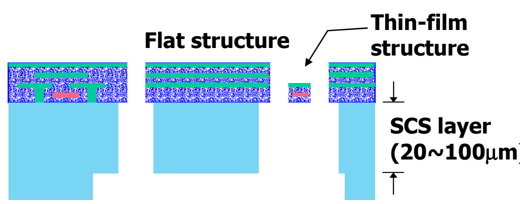

**(c) Deep Si etch**

**STS**: same as Step (a).




**(d) Si undercut**

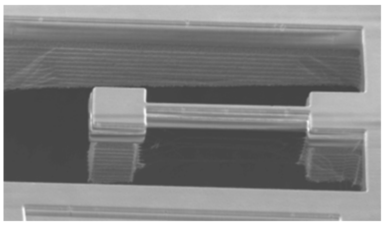
**STS**: 130-sccm $SF_6$,
13-sccm $O_2$, 23 mT,
600 W coil power,
and 0 platen power.



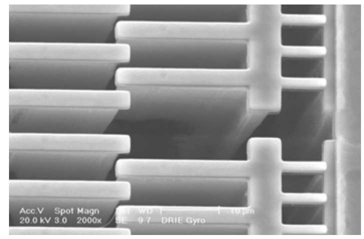

H. Xie et al, J. MEMS, Vol.11, no.2, 2002

## Electrical Isolation of Silicon

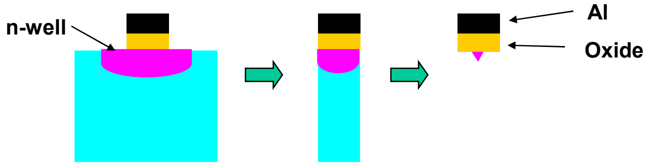
- Electrically isolated silicon island



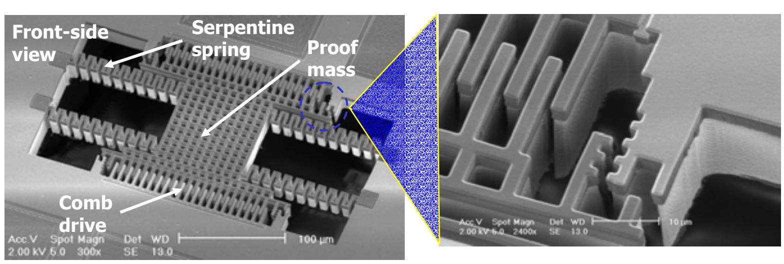
- Electrically isolated comb fingers

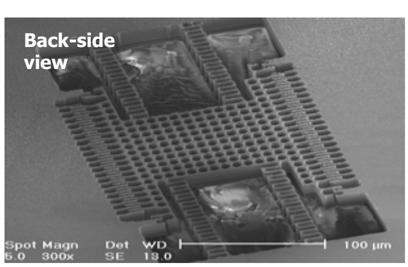


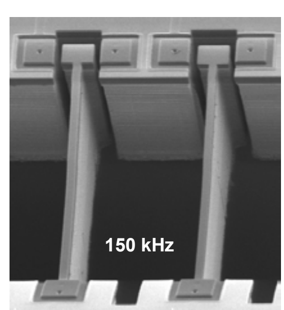
- Using n-well to improve undercut yield




## DRIE CMOS-MEMS Resonators



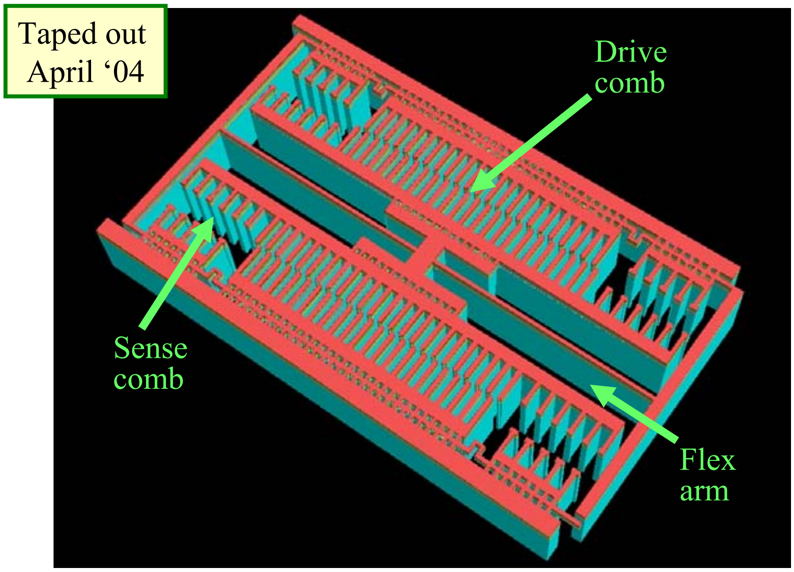




Resonators

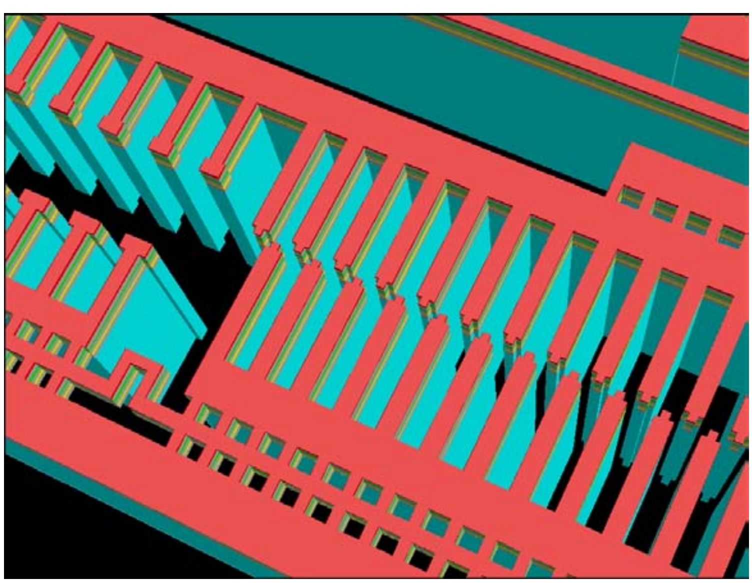
UF CONFIDENTIAL – PATENT PENDING

## Post-TSMC35 AdiaMEMS Resonator



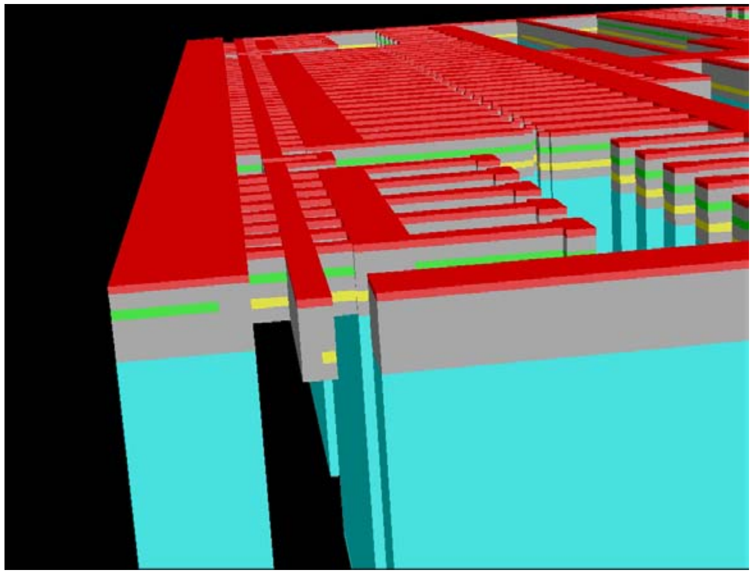

UF CONFIDENTIAL – PATENT PENDING

## Close-Up View, Drive/Sense Combs



UF CONFIDENTIAL – PATENT PENDING

# Side View, Showing Si Undercut



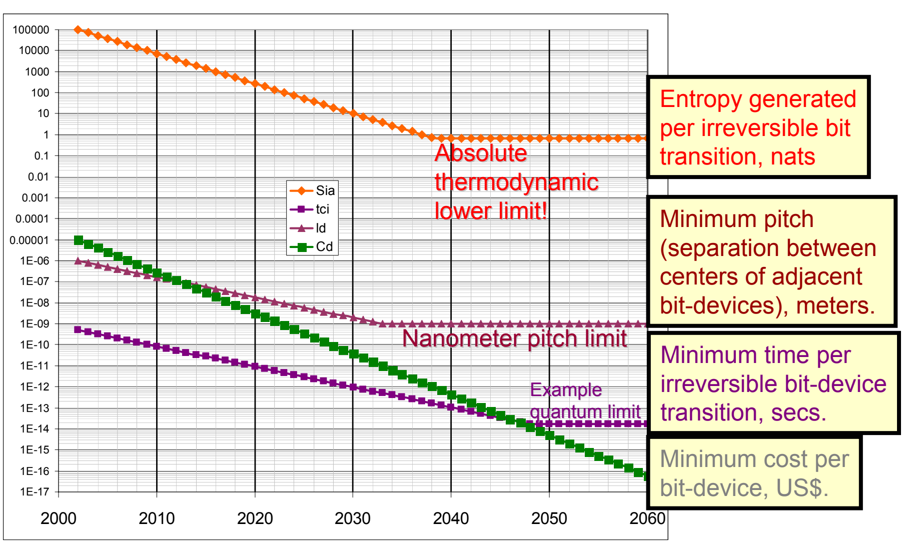
# Long-Term Projections

For future computational cost-efficiency improvements potentially available via reversible computing

# The Future of Reversible Computing

- What if we model how the hardware algorithm overheads for reversible computing scale?
  - Worst case: Increases with roughly $Q^{1.6}$
- Can reversible computing become practical for general-purpose, high-performance computing?
  - And not just for ‘niche’ ultra-low-power apps?
- What happens if present technological trends continue until fundamental limits are reached?
  - And, what happens after that?
- We performed an analysis that addresses these questions…

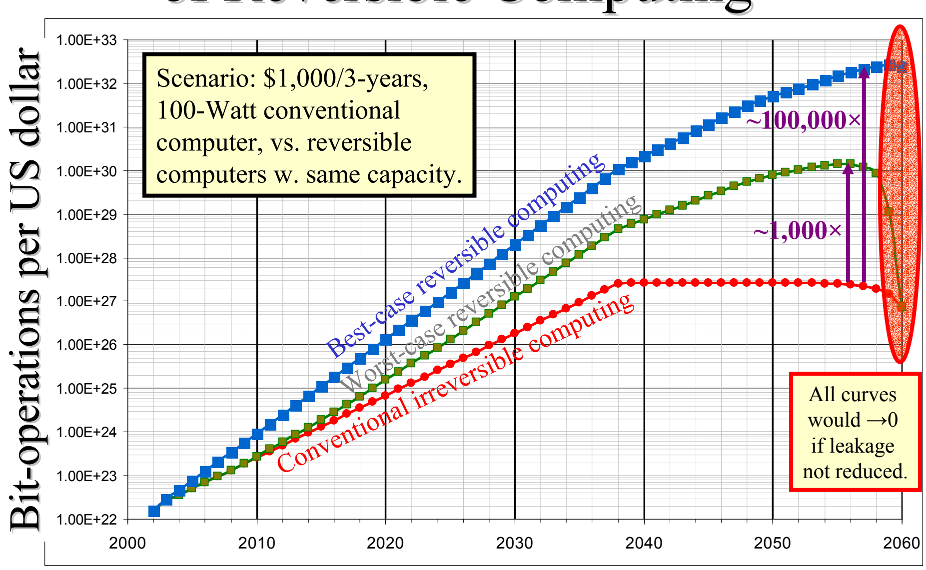
# Technological Trend Assumptions



# Fixed Technology Assumptions

- Total cost of manufacture: US$1,000.00
  - User will pay this for a high-performance desktop CPU.
- Expected lifetime of hardware: 3 years
  - After which obsolescence sets in due to price drops.
- Total power limit: 100 Watts
  - Practical limit for a laptop much quieter than a hair-dryer!
- Power flux limit: 100 Watts per square centimeter
  - Approximate limit of conduction/air-cooling capabilities
- Standby entropy generation rate: 1,000 nat/s/device
  - Arbitrarily chosen, but achievable *e.g.* by today's DRAMs.

# Cost-Efficiency Benefits of Reversible Computing



# Next Steps

- An industry partner in chip design is needed, to help convince funding agencies (NASA, DOD) that real products can result from this work.
  - We offer TI the chance to be our partner in developing these techniques towards DSP products.
- This partner would also join us in preparing various upcoming proposals for gov't funding.
  - *E.g.*, NASA "Code T" program (external call)
- We would like to work closely with a team of 1-3 serious architects who are willing to learn and try out a rather nontraditional logic framework.
  - Keeping the long-term benefits in mind.
- Work on the MEMS-based power supply is crucial, and ongoing…
  - Sandia lab may help us with this.

# Conclusions

- Standard CMOS is approaching imminent power-performance limits.
  - Due to various lower bounds on the energy dissipated by conventional irreversible switching.
- *Only* a mostly-reversible logic architecture has the potential to bypass all of the known energy limits!
  - Through migration to an increasingly adiabatic, ballistic mode of operation, and an increasingly reversible logic design.
    - With increasingly high-$Q$ energy transfers during logic.
- UF's AdiaMEMS project offers key techniques for near-term reversible implementation in CMOS/MEMS.
  - Potentially viable technology for ultra-low-power products.
- Architectures designed today in a mostly-reversible style will be the *only* ones that can be easily ported to future ultra-high-performance reversible logic-device technologies.