Minimum Energy for Computation, the Landauer Principle, and Adiabatic CMOS

Gregory Snider, Graham Boechler, Cameron Thorpe, Craig Lent, and Alexei Orlov
Department of Electrical Engineering
University of Notre Dame
What is the underlying cause of this power dissipation?
Power in Conventional Logic

Conventional CMOS

\[ P = N(CV^2 f + \text{Passive Dissipation}) \]

2\(E_{\text{Bit}}\)

How to reduce power?
- Reduce V
- Reduce C
- Reduce f (multi-core)
- Turn off parts of the circuit
- Reduce passive power

How low can you go?

If \(E_{\text{Bit}} = 100 \, k_B T\), \(f = 100 \, \text{GHz}\), \(N = 10^{11} \, \text{cm}^{-2}\)

\[ P = 4 \, \text{kW/cm}^2 \]
Fundamental Limits for Computation?

Is there a fundamental lower limit on energy dissipation per bit?

*i.e. is there a minimum amount of heat that must be generated to compute a bit?*
Minimum Energy for Computation

- Maxwell’s demon (1875) – by first measuring states, could perform reversible processes to lower entropy
- Szilard (1929), Brillouin (1962): measurement causes $k_B T \ln(2)$ dissipation per bit.
- Landauer (1961, 1970): only erasure of information must cause dissipation of $k_B T \ln(2)$ per bit (Landauer’s Principle)
- Bennett (1982): full computation can be done without erasure.
  - logical reversibility – physical reversibility

Still somewhat controversial.
The Debate


The connection between logical and thermodynamic irreversibility, James Ladyman, Stuart Presnell, Anthony J. Short, Berry Groisman, Studies in History and Philosophy of Modern Physics, 38, 58 (2007).
Analysis of Erasure Process

Helpful to contrast two cases:

- **Erasure without a copy**
  - Irreversible logical operation
  - Key feature:
    The system cannot be biased toward the state it is in, so there’s an uncontrolled step

- **Erasure with a copy**
  - Reversible logical operation
  - Key feature:
    The copy biases the system toward the state it is in
What About State Variables?

• Does the choice of a state variable affect dissipation?

• Doesn’t charge as the state variable bring fundamental limits?

• If we use a different state variable like spin, can a fundamental limit be avoided, or at least get a much lower dissipation than with charge?
The SRC Analysis

Beginning in 2003 Zhirnov, Cavin, and Hutchby from SRC have published a series of highly influential papers indicting charge as a state variable.


**Their conclusions:**

- At least $k_B T \ln(2)$ must be dissipated at each transition
- This result was generalized to all charge-based devices

This is true for conventional CMOS, but what about other charge based paradigms?

Fig. 1. Energy model for limiting device: $w =$ width of left-hand well (LHW) and right-hand well (RHW); $a =$ barrier width; $E =$ barrier energy
What About “Reversible” Computing?

Following Landauer, the idea is to avoid erasure of information.

A key technology in reversible computing is adiabatic charging and discharging of capacitors: recycle charge rather than throwing it to ground.

\[ E_R = \frac{1}{2} CV^2 \]

\[ E_R < \frac{1}{2} CV^2 \]

The SRC critique: Cavin’s Demon
Cavin’s Demon

Assertion 1. Energy must be dissipated to make logic transitions.

Energy spent by the demon must be $> k_B T \ln(2)$

Since there is no input this is really just creating a bit of information.

Figure 6: Adiabatic barrier transitions
What About Adiabatic Charging?

Assertion 2: Charging a capacitor requires at least $k_B T \ln(2)$ of energy

![Graphs showing charging processes](image)

**Figure 3**: Signal forms for “non-adiabatic” and “adiabatic” charging

**Figure 4**: Due to the presence of thermal noise, the linear ramp is corrupted

---

Note 1 on adiabatic charging: The energy dissipated in RC circuit by adiabatic charging cannot be smaller than $kT \ln 2$

What About Clocking?

Cavin’s Demon

Assertion 3:

This is a systems level assertion that depends on the signal generator. However, signal generators can scale differently than integrated circuits!

Worst case: Signal generator more easily cooled than an IC!

On-chip resonators for the clock can address the signal generator issue

SRC’s Conclusion: Charge is dead!
Are Spins or Magnets Better than Charge?

Recently, it has been shown that the minimum energy dissipated to switch a charge-based device like a transistor at a temperature $T$ is $\sim NkT\ln(1/p)$, where $N$ is the number of information carriers (electrons or holes) in the device and $p$ is the bit error probability.$^1$ On the other hand, the minimum energy dissipated to switch a single-domain nanomagnet (which is a collection of $M$ spins) can be only $\sim kT\ln(1/p)$, since the exchange interaction between spins makes $M$ spins rotate together in unison like a giant classical spin.$^{1,2}$ This gives the magnet an advantage over the transistor.

Roy, Bandyopadhyay, and Atulasimha, APL 99, 063108, 2011

Nanomagnet based computers dissipate $k_B T \ln 2$, while charge based computers must dissipate $Nk_B T \ln 2$, where $N \geq 10^4$
Are the Indictments of Charge Influential?

ITRS Roadmap

NRI Centers


What is Correct?
Let’s start with $Nk_B T \ln 2$

**Interacting systems for self-correcting low power switching**

Sayeef Salahuddin and Supriyo Datta
School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana 47907 and NSF Network for Computational Nanotechnology (NCN), Purdue University, West Lafayette, Indiana 47907

This letter first shows that dynamic switching schemes can be used to reduce energy dissipation below the thermodynamic minimum of $NkT \ln r$ ($N$=number of information carriers and $1/r$ =error probability), but only at the expense of the error immunity inherent in thermodynamic processes for which the final state is insensitive to the switching dynamics. It is further shown that, for a system which has internal feedback, e.g., nanomagnets, such that all $N$ spins act in concert, it should be possible to switch with an energy dissipation of the order of $kT \ln r$ (considerably less than the thermodynamic limit of $NkT \ln r$), while retaining an error immunity comparable to thermodynamic switching. © 2007 American Institute of Physics. [DOI: 10.1063/1.2709640]

Spins are in concert, electrons are not
How is this derived?

Claim: “a theoretical minimum dissipation of $N k_B T \ln r$ or $N k_B T \ln 2$”

$V_{min} = (k_B T/q) \ln 2$ (a constant for a given $T$) was simply substituted in the equation $C V^2 = N q V$ as a variable thus making this equation look like:

$$C V^2 = N q V = E_{min}$$

Watch my hands! We substitute $V_{min} = (k_B T/q) \ln 2$ for $V$:

$$N q (k_B T/q) \ln 2 = E_{min}$$

$$N k_B T \ln 2 = E_{min}$$

No other proof for this statement could be found in any of the cited literature
Is Cavin’s Demon Real?

Let’s do an experiment!

The challenge: $k_B T \ln(2)$ at $RT = 3 \text{ zJ}$!
The results

Experimental verification of Landauer’s principle linking information and thermodynamics

Antoine Bérut, Artak Arakelyan, Artyom Petrosyan, Sergio Ciliberto, Raoul Dillenschneider & Eric Lutz

Nature 483, p189, 2012
The Landauer Principle

The SRC group rejects the Landauer Principle, but can it be tested?

COPY BIT TO M

ERASE BIT FROM M
The Landauer Principle

Room temperature operations on a $30 \ k_B T$ bit of information

Dissipation was measured as low as $0.01 \ k_B T$, confirming the Landauer Principle.
Experimental Summary

![Bar chart showing energy measurements for COPY and ERASE WITH COPY scenarios, with different energy levels for Bit Energy and Dissipated Energy at various temperatures (T in μsec). The chart includes a line equation E = k_B T ln(2).]
Are We Cheating by Averaging?

• No! We are merely making an accurate measurement of the real dissipation (the energy you buy to run the logic operation).

• Only the thermal noise energy is averaged away.

The averaging has nothing to do with the bit of information. $30 k_B T$ is strong enough to see without averaging.
What about the small bit floating on a sea with waves the size of $k_B T$?
What about Interconnects?

Interconnects consume a significant portion of the power in a modern chip.

The issue is that due to the large capacitance of the interconnect wire $E_{Bit}$ is large.

Must this energy be dissipated to heat?

No! The energy can be recovered, depending on the RC time constant.
Future of Computation?

• No Matter what your state variable, you have to worry about $E_{\text{bit}}$
• Does that mean you have to go fully reversible?
• Are new devices necessary?

What are the trade-offs?

• Speed
• Complexity
Adiabatic CMOS

Power dissipation in an adiabatic CMOS system

\[ P_{Total} = N\left\{ CV_{DD}^2 f \left[ \alpha \frac{f}{f_o} + (1 - \alpha) \right] + A \exp\left( -\frac{qV_{DD}}{4\eta kT} \right) \right\} \]

Active power is not broken into two parts

- Adiabatic part dependent on \( f^2/f_o \)
- Non-adiabatic part dependent on \( f \)

- Passive power is exponentially dependent on \( V_{DD} \), while active power is only quadratically dependent.

- In adiabatic CMOS this difference can be exploited to reduce power.

\[ V_{Th} = \frac{V_{DD}}{4} \]
When the operating frequency is below $f_0$ the active power can be far lower than standard CMOS, and $V_{DD}$ can be raised without giving back much of the power savings.
Conclusions

• Future progress requires energy recycling
• Charge is a viable state variable for beyond the roadmap devices
• There is no fundamental lower limit on the energy needed for computation – only practical ones
• Charge as a state variable does not require exotic devices to achieve power savings
• Adiabatic CMOS can break the link between active and static power, and enable reductions in passive power without significant increases in active power.
• The key is to trade speed for power, a trade-off that is already being made.