Physics of Information

Igor Neri - July 17, 2018

NiPS Summer School 2018
Energy aware transprecision computing
Limits to computation

- Minimum size of computing device
- Maximum computational speed of a self-contained system
- Information storage in a finite volume
- Energy consumption limit to:
  - computation
  - memory preservation
Minimum size of computing device
## Transistor count

<table>
<thead>
<tr>
<th>Processor</th>
<th>Transistor count</th>
<th>Date of introduction</th>
<th>Designer</th>
<th>Process</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMS 1000</td>
<td>8,000</td>
<td>1974</td>
<td>Texas Instruments</td>
<td>8,000 nm</td>
<td>11 mm²</td>
</tr>
<tr>
<td>Intel 4004</td>
<td>2,500</td>
<td>1971</td>
<td>Intel</td>
<td>10,000 nm</td>
<td>12 mm²</td>
</tr>
<tr>
<td>Intel 8008</td>
<td>3,500</td>
<td>1972</td>
<td>Intel</td>
<td>10,000 nm</td>
<td>14 mm²</td>
</tr>
<tr>
<td>MOS Technology 6502</td>
<td>3,510[4]</td>
<td>1975</td>
<td>MOS Technology</td>
<td>8,000 nm</td>
<td>21 mm²</td>
</tr>
<tr>
<td>Motorola 6800</td>
<td>4,100</td>
<td>1974</td>
<td>Motorola</td>
<td>6,000 nm</td>
<td>16 mm²</td>
</tr>
<tr>
<td>Intel 8080</td>
<td>4,500</td>
<td>1974</td>
<td>Intel</td>
<td>6,000 nm</td>
<td>20 mm²</td>
</tr>
<tr>
<td>RCA 1802</td>
<td>5,000</td>
<td>1974</td>
<td>RCA</td>
<td>5,000 nm</td>
<td>27 mm²</td>
</tr>
<tr>
<td>Intel 8085</td>
<td>6,500</td>
<td>1978</td>
<td>Intel</td>
<td>3,000 nm</td>
<td>20 mm²</td>
</tr>
<tr>
<td>61-core Xeon Phi</td>
<td>5,000,000,000</td>
<td>2012</td>
<td>Intel</td>
<td>22 nm</td>
<td>720 mm²</td>
</tr>
<tr>
<td>Xbox One main SoC</td>
<td>5,000,000,000</td>
<td>2013</td>
<td>Microsoft/AMD</td>
<td>28 nm</td>
<td>353 mm²</td>
</tr>
<tr>
<td>10-core Xeon Haswell-E5</td>
<td>5,580,000,000,000</td>
<td>2014</td>
<td>Intel</td>
<td>22 nm</td>
<td>661 mm²</td>
</tr>
<tr>
<td>IBM z14</td>
<td>6,100,000,000</td>
<td>2017</td>
<td>IBM</td>
<td>14 nm</td>
<td>898 mm²</td>
</tr>
<tr>
<td>Xbox One X (Project Scorpio) main SoC</td>
<td>7,000,000,000,000</td>
<td>2017</td>
<td>Microsoft/AMD</td>
<td>16 nm</td>
<td>560 mm²</td>
</tr>
<tr>
<td>IBM z13 Storage Controller</td>
<td>7,100,000,000</td>
<td>2015</td>
<td>IBM</td>
<td>22 nm</td>
<td>678 mm²</td>
</tr>
<tr>
<td>22-core Xeon Broadwell-E5</td>
<td>7,200,000,000,000</td>
<td>2016</td>
<td>IBM</td>
<td>14 nm</td>
<td>456 mm²</td>
</tr>
<tr>
<td>POWER9</td>
<td>8,000,000,000</td>
<td>2017</td>
<td>Intel</td>
<td>14 nm</td>
<td>883 mm²</td>
</tr>
<tr>
<td>72-core Xeon Phi</td>
<td>8,000,000,000</td>
<td>2018</td>
<td>Intel</td>
<td>14 nm</td>
<td>896 mm²</td>
</tr>
<tr>
<td>IBM z14 Storage Controller</td>
<td>9,700,000,000</td>
<td>2017</td>
<td>IBM</td>
<td>14 nm</td>
<td>896 mm²</td>
</tr>
<tr>
<td>32-core SPARC M7</td>
<td>10,000,000,000,000</td>
<td>2015</td>
<td>Oracle</td>
<td>20 nm</td>
<td>298 mm²</td>
</tr>
<tr>
<td>Centriq 2400</td>
<td>18,000,000,000,000</td>
<td>2017</td>
<td>Qualcomm</td>
<td>10 nm</td>
<td>768 mm² (4 x 192 mm²)</td>
</tr>
<tr>
<td>32-core AMD EPYC</td>
<td>19,200,000,000,000</td>
<td>2017</td>
<td>AMD</td>
<td>14 nm</td>
<td>768 mm² (4 x 192 mm²)</td>
</tr>
</tbody>
</table>
Moore’s law
Transistor size

In terms of size [of transistors] you can see that we're approaching the size of atoms which is a fundamental barrier, but it'll be two or three generations before we get that far - but that's as far out as we've ever been able to see. We have another 10 to 20 years before we reach a fundamental limit. By then they'll be able to make bigger chips and have transistor budgets in the billions. - G. Moore
Energy limits speed of computation
Energy limits speed of computation

• What limits the laws of physics place on the speed of computation?
Energy limits speed of computation

Heisenberg uncertainty principle

\[ \Delta E \Delta t \geq \frac{\hbar}{2} \]

**Wrong interpretation:** it takes time \( \Delta t \) to measure energy to an accuracy \( \Delta E \)

**Right interpretation:** a quantum state with spread in energy \( \Delta E \) takes time at least

\[ \Delta t = \frac{\pi \hbar}{2 \Delta E} \]

to evolve to an orthogonal (and hence distinguishable) state

Ultimate physical limits to computation - Seth Lloyd
Maximum computational speed of a self-contained system

- **Bremermann's Limit** is the maximum computational speed of a self-contained system in the material universe. It is derived from Einstein's mass-energy equivalency and the Heisenberg uncertainty principle, and is \( c^2/h \approx 1.36 \times 10^{50} \) bits per second per kilogram.

- The **Margolus–Levitin theorem** gives a fundamental limit on quantum computation (strictly speaking on all forms on computation). The processing rate cannot be higher than \( 6 \times 10^{33} \) operations per second per joule of energy.
Maximum computational speed of a laptop
Maximum computational speed of a laptop

• If the mass is $m$ then $E = mc^2$

• $m = 1$ Kg, $E = 1 \times (3 \times 10^8)^2 = \text{approx} \approx 10^{17}$ J

• $\Delta t = \text{approx} \approx 10^{-34}/10^{17} = 10^{-51}$ s
Comparison with existing computers

- Conventional laptops operate much more slowly than the ultimate laptop

- Two reasons for this inefficiency:
  - most of the energy is locked up in the mass of the particles of which the computer is constructed
  - a conventional computer employs many degrees of freedom for registering a single bit

Ultimate physical limits to computation - Seth Lloyd
Memory space limits
Memory space limits

- The amount of information that a physical system can store and process is related to the number of distinct physical states accessible to the system.

- A collection of $M$ two-state systems has $2^M$ accessible states and can register $M$ bits of information.

- A system with $N$ accessible states can register $\log_2 N$ bits of information.
Memory space limits

- The number of accessible state, $W$, of a physical system is related to its thermodynamic entropy by the formula: $S = k_B \log W$

- The amount of information that can be registered by a physical system is $I = \frac{S(E)}{k_B \log 2}$

- $S(E)$ is the thermodynamic entropy of a system with expectation value for the energy $E$
Information storage in a finite volume

- The **Bekenstein bound** limits the amount of information that can be contained within a given finite region of space which has a finite amount of energy:

\[
S \leq \frac{2\pi k R E}{\hbar c}
\]

\[
I \leq \frac{2\pi c R m}{\hbar \ln 2} \approx 2.577 \times 10^{43} mR
\]
Information storage in a finite volume

- Human brain

\[ I \leq \frac{2\pi c R m}{\hbar \ln 2} \approx 2.577 \times 10^{43} mR \]

- mass \( m = 1.5 \) kg

- volume of 1260 cm\(^3\)

- approximating volume to a sphere \( R = 6.7 \) cm

- \( I = 2.6 \times 10^{42} \) bits

- \( O = 2^I \) states of the human brain must be less than \( \approx 10^{7.8 \times 10^{41}} \)
Comparison with existing computers

- The amount of information that can be stored by the ultimate laptop $\approx 10^{31}$ bits

- Conventional laptops can store $\approx 10^{12}$ bits

- This is because conventional laptops use many degrees of freedom to store a bit where the ultimate laptop uses just one

- There are considerable advantages to using many degrees of freedom to store information, stability and controllability being perhaps the most important
Minimum energy consumption for computation
Information is physical
Maxwell’s demon

- Gas equilibrates to initial temperature
- Attach weight to piston
Maxwell’s demon

Cyclic process converts heat completely into work!

Violates second law of thermodynamics!

No process is possible whose sole result is the absorption of heat from a reservoir and the conversion of this heat into work.
Maxwell’s demon

- Gas equilibrates to initial temperature
- Heat has been converted to work
- Attach weight to piston
- Pressure lifts weight

$Q$
Information is physical
What happens when computation is logically irreversible?

- Minimum amount of energy required greater than zero
- Let assume the operation of bit reset
- # of initial states: 2
- # of final states: 1
Landauer principle

- Initial condition: two possible states
  \[ S = k_B \log W \]
  \[ Q \leq T \Delta S \]

- Final condition: one possible state
  \[ S_i = k_B \log 2 \]
  \[ S_f = k_B \log 1 \]
  \[ \Delta S = S_f - S_i = -k_B \log 2 \]

- Heat produced
  \[ Q \leq T \Delta S = -k_B T \log 2 \]
Landauer principle at room temperature

\[ Q \leq T \Delta S = -k_B T \log 2 \sim 10^{-21} J \]

Figure 3: Energy per logic operation

After Electronics Beyond Nano-scale CMOS, Shekhar Borkar
Landauer principle
experimental verification

Brownian particle in a double-well potential

Barut et al. Nature 2012

Giliberto ENS Lyon

Measured erasure cycle:

Landauer's original thought experiment

The physics of information: from Maxwell’s demon to Landauer - Eric Lutz - University of Erlangen-Nürnberg
Landauer principle
experimental verification

Even if you're not burning books, destroying information generates heat. - Sergio Cicliberto
Landauer principle
experimental verification

Generic two-state memory:

- initial configuration: two states with equal probability $1/2$
  → system can store 1 bit of information
  Shannon entropy: $S_i = - \sum_n p_n \ln p_n = \ln 2$

- final configuration: one state with probability 1
  → system can store 0 bit of information
  Shannon entropy: $S_f = - \sum_n p_n \ln p_n = 0$

→ original bit has been deleted: $\Delta S = - \ln 2$
Landauer principle
experimental verification

Second law of thermodynamics (for system and reservoir):

$$\Delta S = \Delta S_{sys} + \Delta S_{res} \geq 0$$

Reservoir always in equilibrium:

$$Q_{res} = T \Delta S_{res} \geq -T \Delta S_{sys}$$

Equivalence between entropies:

$$\Delta S_{sys} = k \Delta S = -k \ln 2$$

Heat produced in reservoir:

$$Q_{res} \geq kT \ln 2$$

→ connection between information theory and thermodynamics

($Q_{res} = kT \ln 2$ in quasistatic limit i.e. long cycle duration)
Landauer principle
experimental verification

Experimental results:

We measure work $W$ and deduce heat $Q = -\Delta U + W = W$

$\rightarrow$ Landauer can be bound approached but not exceeded

Note: $kT \ln 2 \approx 3 \times 10^{-21} J$ at room temperature
Reset on colloidal particles

Colloidal particle bit

Total energy landscape

\[ E(x, t) = U(x, t) - F_0 f(t)x \]

\[ U(x, t) = -\frac{a}{2} g(t) x^2 + \frac{b}{4} x^4 \]

- \( g(t) \) and \( f(t) \): dimensionless parameter in \([0, 1]\). Their value at time \( t \) depends on a given protocol.

Time-dependent study

For a fixed $\tau_{pr}$ with $Q(\tau_{pr}) \approx -T \Delta S(\tau_{pr})$, study $-T \Delta S(t)$, $Q(t)$, $W(t)$, $\Delta E(t)$.

$Q(t) = -T \Delta S(t), \quad \forall t \in [0, \tau_{pr}]$

Consistent with qualitative analysis.

Landauer principle
experimental verification

High-Precision Test of Landauer’s Principle in a Feedback Trap

Yonggun Jun, Momčilo Gavrilov, and John Bechhoefer
Department of Physics, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada
(Received 15 August 2014; published 4 November 2014)

We confirm Landauer’s 1961 hypothesis that reducing the number of possible macroscopic states in a system by a factor of 2 requires work of at least $kT \ln 2$. Our experiment uses a colloidal particle in a time-dependent, virtual potential created by a feedback trap to implement Landauer’s erasure operation. In a control experiment, similar manipulations that do not reduce the number of system states can be done reversibly. Erasing information thus requires work. In individual cycles, the work to erase can be below the Landauer limit, consistent with the Jarzynski equality.

DOI: 10.1103/PhysRevLett.113.190601
PACS numbers: 05.70.Ln, 03.67.-a, 05.20.-y, 05.90.+m
Landauer principle
experimental verification

Feedback Trap

Landauer principle
experimental verification

Erasure protocol

Landauer principle
experimental verification

Work series for individual cycles

Beating the Landauer's limit by trading energy with uncertainty

\[ \Delta S = S_f - S_i = k_B (\ln(1) - \ln(2)) = -k_B \ln(2) \]

\[ S_f(P_e) = -k_B ((1 - P_e) \ln(1 - P_e) + P_e \ln(P_e)) \]

\[ Q(P_e) = -k_B T ((1 - P_e) \ln(1 - P_e) + P_e \ln(P_e)) + -k_B T \ln(2) \]

Beating the Landauer's limit by trading energy with uncertainty - L. Gammaitoni - arXiv:1111.2937 [cond-mat.mtrl-sci]
Beating the Landauer's limit by trading energy with uncertainty

\[ Q(P_e) = -k_B T((1 - P_e) \ln(1 - P_e) + P_e \ln(P_e)) + -k_B T \ln(2) \]

Beating the Landauer's limit by trading energy with uncertainty - L. Gammaitoni - arXiv:1111.2937 [cond-mat.mtrl-sci]
Micro-electromechanical memory bit based on magnetic repulsion

\[ U_{el} = \frac{1}{2} k x^2 \]

\[ F_m(r, m_1, m_2) = \frac{3\mu_0}{4\pi r^5} \left[ (m_1 \cdot r)m_2 + (m_2 \cdot r)m_1 + (m_1 \cdot m_2)r \right] - \frac{5(m_1 \cdot r)(m_2 \cdot r)}{r^2} \]

Micro-electromechanical memory bit based on magnetic repulsion
Micro-electromechanical memory bit based on magnetic repulsion,

Micro-electromechanical memory bit based on magnetic repulsion

Orders of magnitude above Landauer limit!
Solution: increase the temperature

\[ T_{\text{eff}} = 5 \times 10^7 \, \text{K} \]
Reset protocol

\[ Q = W - \Delta U \]

\[ Q(P_s) \geq k_B T \left[ \ln(2) + P_s \ln(P_s) + (1 - P_s) \ln(1 - P_s) \right] \]

Landauer reset with error

We shall call a device logically irreversible if the output of a device does not uniquely define the inputs. We believe that devices exhibiting logical irreversibility are essential to computing. Logical irreversibility, we believe, in turn implies physical irreversibility, and the latter is accompanied by dissipative effects.
Information is Physical

Rolf Landauer, 1961. Whenever we use a logically irreversible gate we dissipate energy into the environment.
Landauer has posed the question of whether logical irreversibility is an unavoidable feature of useful computers, arguing that it is, and has demonstrated the physical and philosophical importance of this question by showing that whenever a physical computer throws away information about its previous state it must generate a corresponding amount of entropy. Therefore, a computer must dissipate at least $k_B T \ln 2$ of energy (about $3 \times 10^{-21}$ Joule at room temperature) for each bit of information it erases or otherwise throws away.
Solution = Reversibility

• Charles Bennett, 1973: There are no unavoidable energy consumption requirements per step in a computer.

• Energy dissipation of reversible circuit, under ideal physical circumstances, is zero.
Reversible computation

- **Landauer/Bennett**: all operations required in computation could be performed in a reversible manner, thus dissipating no heat.

- The first condition for any deterministic device to be reversible is that its input and output be uniquely retrievable from each other, then it is called **logically reversible**.

- The second condition: a device can actually run backwards, then it is called **physically reversible**, and the second law of thermodynamics guarantees that it dissipates no heat.
Billiard ball computing

• Model of a reversible mechanical computer based on Newtonian dynamics

• Proposed in 1982 by Edward Fredkin and Tommaso Toffoli

• It relies on the motion of spherical billiard balls in a friction-free environment made of buffers against which the balls bounce perfectly
Billiard ball computing

Assume no friction, elastic collisions
Billiard ball computing

Use “mirrors” to implement “switching device”
This device is *reversible* because physics is
Billiard ball computing

• Using balls and mirrors, we can implement basic logic gates: AND, OR, NOT

• With a big enough billiard table, we could (in theory) implement a complete computer using a combination of these gates

• BUT…

  • billiard balls don't work in practice
Billiard ball computing

- **Thermal losses**
  - friction can't be ignored
  - Collisions aren't perfectly elastic

- **Chaotic motion**
  - Balls are actually conglomerates of many atoms in various states of vibration
  - Can't know their “initial state” perfectly
  - Small variations in initial conditional conditions can cause exponentially large differences in final state
Reversible computing

- The reasoning on connection between physical and logical reversibility applies only to systems that encode input and outputs on the system itself.

- If the input and output are not part of the computing system (like in transistor based logic gates) there is no connection between physical and logical reversibility.
Back to the real world….

OR gate

<table>
<thead>
<tr>
<th>INGRESSO</th>
<th>USCITA</th>
<th>A OR B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Back to the real world....

OR gate

<table>
<thead>
<tr>
<th>INGRESSO</th>
<th>USCITA</th>
<th>A OR B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Back to the real world….

OR gate

www.randomwraith.com
The experimental setup
The experimental setup
The experimental setup

![Graph showing the Landauer limit](image)
XOR gate
Full adder
Minimum energy consumption for memory preservation
The refresh procedure

Symbol to store is known with no error
E. g. 1

Repeat $\frac{t}{t_R}$ times

Wait $t_R$

Symbol is 1 with probability $(1 - P_0)$

$P_0$ is a function of $t_R$

Refresh:
1) Read the current symbol
2) Write the read symbol

After Refresh

Symbol is 1 with probability $(1 - P_0)$

Diffusions of the physical quantity that encodes information are removed

After $\frac{t}{t_R}$ repetitions, the probability that there are undesired transitions to 0

$P_E = 1 - (1 - P_0)^{\frac{t}{t_R}}$
To evaluate the energy cost of the refresh procedure we need:

- A physical description of the memory
- A characterisation of $P_0$ as function of refresh time $t_R$
- A physical description of the refresh procedure
- A characterisation of total error probability $P_E$ as function of refresh time $t_R$ after a fixed time
Physical description of the memory

\[ U(x) \quad P(x, t) \]

symbol 0 \quad symbol 1

\[ P_0 \quad k_B T \]
Characterisation of $P_0$ as function of refresh time

\[
\frac{\partial}{\partial t} p(x, t) = \frac{\partial}{\partial x} \left( \frac{\partial U}{\partial x} p(x, t) \right) + \frac{k_B T}{\Delta U} \frac{\partial^2}{\partial x^2} p(x, t),
\]

\[
P_0(t) = \int_{-\infty}^{0} p(x, t) \, dx
\]
Physical description of the refresh procedure

\[ U(x) \]

- Narrow: \( P_0 \) is the same, Peaks sharpen
- Normal: \( P_0 \) is the same, Peaks are the same
Characterisation of $P_E$ as function of refresh time

![Graph showing the characterisation of $P_E$ as a function of refresh time ($t/\tau_k$). The graph uses a logarithmic scale for both axes, ranging from $10^{-7}$ to $10^{2}$ for $P_E$ and from $10^{0}$ to $10^{2}$ for $t/\tau_k$. The contour lines represent different values of $P_E$.](image-url)
What is the fundamental cost for preserving a memory for a fixed time with a given probability of error?
Study of the energy cost of refresh procedure

\[ (t_R) = r^2 w + \exp(-t_R/\tau_w) (\sigma_i^2 - \sigma_w^2) \text{ in } \sigma_i \]
Minimum energy required to preserve a memory over a fixed time with a given error probability

\[ Q_m = -NT\Delta S = \frac{\bar{t}}{t_R} k_B T \ln \left( \sqrt{\sigma_w^2 + e^{-\frac{t_R}{\tau_w}} \left( \frac{\sigma_i^2 - \sigma_w^2}{\sigma_i} \right)} \right) \]
Minimum energy required to preserve a memory over a fixed time with a given error probability

\[ P_E = 1 \times 10^{-6} \quad P_E = 1 \times 10^{-4} \quad P_E = 1 \times 10^{-2} \]
Minimum energy required to preserve a memory over a fixed time with a given error probability

\[ P_E = 1 \times 10^{-6} \quad P_E = 1 \times 10^{-4} \quad P_E = 1 \times 10^{-2} \]
Limits to computation

- Minimum size of computing device
- Maximum computational speed of a self-contained system
- Information storage in a finite volume
- Energy consumption limit to:
  - computation
  - memory preservation
References


Thank you for your attention!

igor.neri@nipslab.org