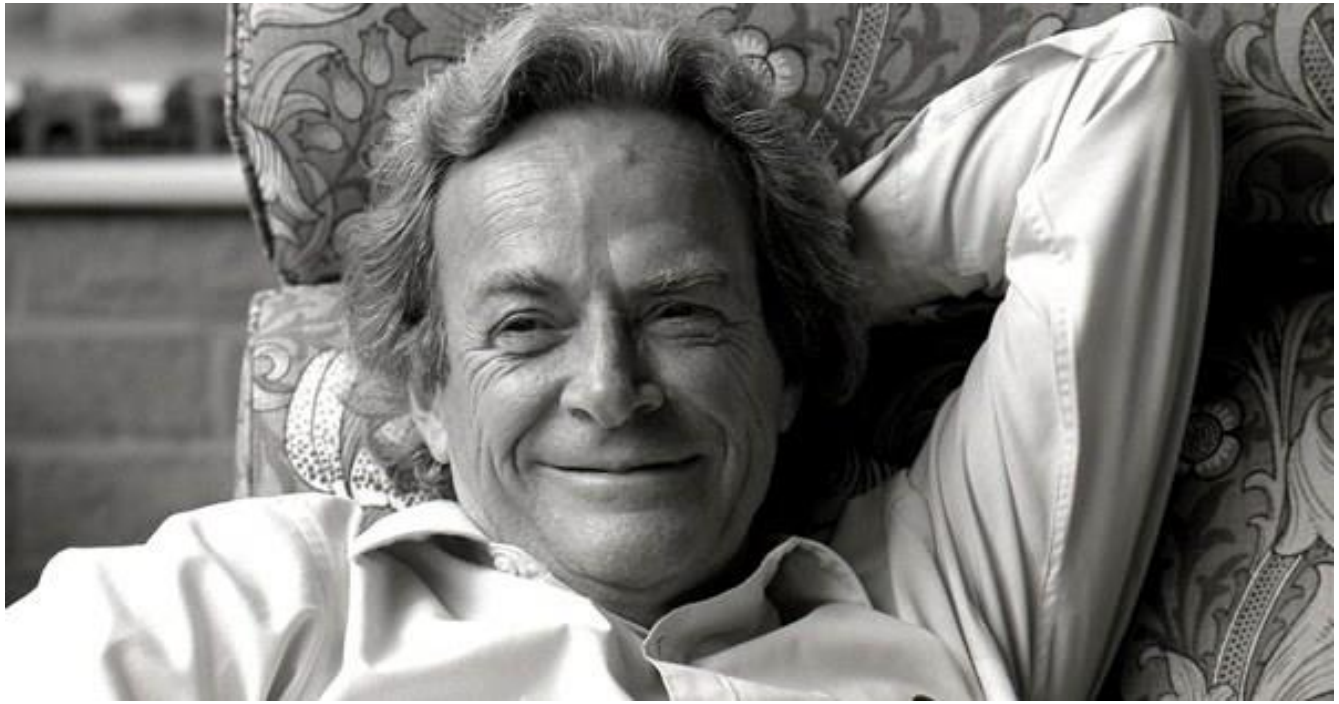
Where is the so called “plenty of room at the bottom”?

Pablo G. Tello, CERN

3rd ATTRACT Symposium

Technology Trends, Wishes and Dreams (TWD)

Tripoli, Greece, 31 May 2017 to 1 June 2017.



Once upon a time, concretely on December 29th, 1959,
Mr. Richard Feynman gave an interesting talk at an American Physical Society meeting at Caltech.

The title of his talk was "There's Plenty of Room at the Bottom"

And during it he said….

There is nothing that I can see in the physical laws that says the computer elements cannot be made enormously smaller than they are now.

The problem is that since Mr. Feynman was a vey clever guy he didn't exactly mention HOW….and since then we all have been struggling.

But he also said one important thing…

Who should do this and why should they do it? Well, I pointed out a few of the economic applications, but I know that the reason that you would do it might be just for fun. But have some fun!

As previous ones, this ATTRACT symposium carries the nickname of Technology TWD (Trends, Wishes and Dreams).

I was asked to give an inspirational talk (which I find very intimidating in front of this audience).

But, since we are free to dream, I will talk about that…technology dreams.

…and in the way we maybe start figuring out HOW Mr. Feynman wanted to find “plenty of room at the bottom”.

- Moore's Law, as a key driver of the semiconductor industry, continues to have a transformative impact on society.
- However semiconductor technology will face significant physics-based scaling and performance limits in the (near) future.
- There are many ways to "predict" the end of Moore's law, (RC delays, leakage currents, interconnect parasitic capacitances….).
- I find one of them specially useful, which is to think about CMOS technology as binary switches and memory elements.
- With this perspective I will therefore start with the fundamental limits at the device level; then I will talk about extremely scaled artificial computing systems.
- And I will end up comparing those with Mother Nature.

Ultimate scaling limits of CMOS devices

- Two of the basic constituents for any information processing system are the binary switch and the memory element.
- These two constituents, regardless of the specific physics of their operation, can be represented by a generic controllable barrier model.
- The need for a barrier is fundamentally linked to the nature of information since it allows to distinguish physical states that translates into 0s and 1s.
- Let's therefore look into the ultimate device scalability limits from a barrier perspective.

- The barrier height, E_b, must be large enough to prevent spontaneous transitions (errors) that occur when the particle acquires thermal energy large enough to jump over the barrier.

- The corresponding probability for over-barrier transition p can be obtained from the Boltzmann distribution as:

$$p = exp\left(-\frac{E_b}{k_B T}\right)$$

- The minimum barrier height can be found from the distinguishability condition, requiring that the probability of errors $p < 0.5$, in which case the switch is being operated at the threshold of distinguishability.

$$E_{bmin} = k_B T ln2 \approx \sim 0.7 k_B T \sim k_B T$$

In a (more) practical device the error probability must be very low, and therefore $E_b > k_B T$.

In addition, the requirement that all N_{sw} switches in the logic system operate correctly raises E_b even higher. The probability p_{syst} of a correct operation of all N_{sw} switches in a circuit is:

$$p_{syst} = (1-p)^{N_{sw}}$$

For a given error probability of an individual device p, it is possible then to calculate the barrier height required for a reliable operation:

$$E_b = k_B T ln\left(1 - \sqrt[N_{sw}]{p_{syst}}\right)^{-1}$$

For a system with high reliability ($p_{syst} = 0.99$) and $N_{sw} = 10^9$ (~ today's microprocessor chips) $E_b \sim 25 k_B T$.

This corresponds to minimum operation voltage ($V = 0.65$ V) close to the projections by the International Technology Roadmap for Semiconductors (ITRS) for the end-of-scaling CMOS.

In ultimate scaling another class of errors that impose limits are quantum ones.

From Heisenberg principle we know that:

$$\Delta x \Delta p \sim h$$

Therefore, the minimum size of a scaled computational element or switch is:

$$L_{min} > \Delta x \sim \frac{h}{\Delta p} = \frac{h}{\sqrt{2mE_b}}$$

Using $E_b = 0.1\ eV$ and the effective mass of an electron in semiconductor $m^* = 0.19m_e$ (the transverse electron effective mass in Si), $L_{min} \sim 4\ nm$.

Corresponding to an approximate minimum channel length of the Si logic FET and again consistent with the ITRS projections.

The retention time of a memory cell depends on the barrier height and length.

In order to obtain a non-volatile memory cell, sufficiently high barriers are needed to retain the charge for a long period of time.

Using the same type of reasoning as in previous slides for retention $>10\ y$, the barrier height E_b must be more than $\sim 1.7\ eV\ (\sim 66\ k_B T)$, and length $L > 5\ nm$.

The corresponding practical minimum size of the floating gate cell is $\sim 10\ nm$.

Large barriers also result in high voltages required for memory operation: $\sim 5V$ for READ and $\sim 15V$ for WRITE.

Not my intention to offer detailed calculations of other key performance figures to the audience since excellent references already are available (see further slides). So, I borrow from them.

But it is interesting to look at key figures of merit reflecting ultimate scaling limits:

Keep this number in mind.

	Logic	*Memory*
F_{min} (nm) (critical feature size)	4.5	10
n_{2D} (cm^{-2}) (logic density)	5×10^{11} ($1/8F^2$)	2.5×10^{11} ($1/4F^2$)
V (volts)	0.65	5 (read)/ 15 (write)
E_{bit} (Energy/bit)	3×10^{-18} J ~ 1000 k_BT	~ 10^{-13} J ~ 10^7 k_BT (read) ~ 10^{-12} J~ 10^8 k_BT (write)
P_{leak} (W) (power consumption by leakage currents)	2×10^{-9}	low
n_{3D} (cm^{-3}) (logic density)	1.5×10^{17} ($1/72F^3$)	4.2×10^{26} ($1/24F^3$)

Scaling (information processing) systems to the fundamental limits

Let's suppose that we can make electronic devices, such as transistors, reliably very small, ultimately exhibiting feature sizes of ~5–10 nm nanometres.

However, information processing system miniaturization efforts are still typically on the centimetre scale.

What if we were targeting for a small microsystem ~1μm^3 volume?

WHY? Because ~1μm^3 is the typical volume of a biological cell.

Devices		
Molecules count		***Function***
DNA size	4.6×10^{6} bp=9.6 Mbit	Nonvolatile memory
Number of RNA/cell	222,000	Memory interface
Number of cytoplasmic proteins	1,000,000	1) Logic processor 2) Signal processor 3) Metabolic functions*
Number of ribosomal proteins	900,000	I/O interface
Number of logic 'devices' (Proteins and RNA)	>1,000,000*	1) Logic processor 2) Signal processor 3) Memory interafce 4) I/O interface
Number of all proteins	3,600,000	1) Logic processor 2) Signal processor 3) Metabolic functions* 4) Structural functions*
Dimensions		
Length	2 µm	
Diameter	0.8 µm	
Volume	1 $\mu m^3 = 10^{-12}$ cm^3	
Surface Area	$6\mu m1= 6\times10^{-8}$ cm^2	
Timing		
Time for cell replication	40min=2400s	
Energetics		
Total energy stored	$\sim2\times10^{-12}$ J	
Power dissipation	1.4×10^{-13} W	

*proteins with metabolic and structural functions, excluded from the device count

Let's take a look at *E. Coli* from the point of view of an information micro-processing system.

A staggering observation is that the power consumption of *E. Coli* is about $1.4\times10^{-13}\,W$.

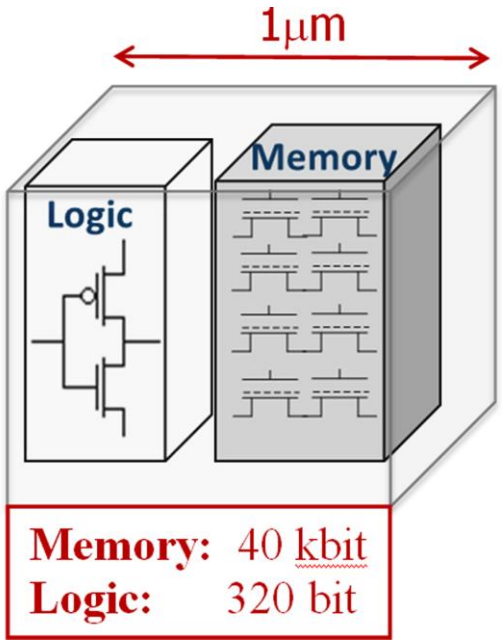
Given this, the energy per equivalent binary operation in the cell can be estimated as $\sim10^{-20}\,J/bit < 10k_BT$.

This estimated energy utilization per switching event is at least **two orders of magnitude** lower than the "Ultimate CMOS" information processor I showed you before!!!!

Now, let's dream about the best Si micro-processing information system we could achieve.

I mean, using CMOS technology as I showed you before scaled to its ultimate limits.

How well can be do?



Let's assume that our hypothetical $1\mu m^3$ Si computer or Si-μCell would contain *320* logic transistors and *40 kbit* of non-volatile memory.

(~ von Neumann minimum logic circuit complexity definition).

This would be the result

	Logic	*Memory*
F_{min} (nm)	4.5	10
N	320	40000
E_{bit} (J/bit)	3×10^{-18}	~ 10^{-15} (read)
E_{cycle}	~ 10^{-15}	~ 10^{-13} (read)
F_{clock} (MHz)	100	
P_{active} (W)	10^{-7}	10^{-5}
P_{leak} (W)	6.4×10^{-7}	assumed low
P_{total} (W)	~ 1.1×10^{-5}	
Q_{active} (W/cm2)	1.7	167
Q_{leak} (W/cm2)	11	assumed low
Q_{total} (W/cm2)	~ 180	

Impressive???

Well, let's compare some key figures with the ultimate REAL biological µ-Biological Cell.

Parameter	*"Ultimate Si μ-Cell"*		*Real μ-Biological cell*	
	Logic	Memory	Logic	Memory
Density	10^{17} cm^{-3}	10^{16} cm^{-3}	10^{18} cm^{-3}	10^{19} cm^{-3}
Energy/bit	10^{3} k_BT	10^{8} k_BT	<10 k_BT	<10^{4} k_BT
Power	10^{-7} W		10^{-13} W	
Heat Flux	1 W/cm^2		10^{-6} W/cm^2	
Energy per 10^{11} output bits	10^{-2} J		< 10^{-9} J	
Time to compute 10^{11} output bits	10^{5} s		10^{3} s	

Both processors are benchmarked against a task equivalent to providing an information output of 10^{11} *bits* (the amount of information estimated that needs to be generated by the cell processor to build a new cell).

It is assumed that the logic processor of the *Si μ-Cell* is implemented by a Minimal Turing Machine (MTM) consisting of 320 transistors and operating at *1 MHz* clock frequency (to stay within the *1 W/cm*^{2} heat limit for passive cooling systems).

About *~500* "raw" bit transitions in MTM are required per one output bit, thus *~5×10*^{13} binary transition are required to emulate the "replication" task, and this will require *~10*^{5} *s* to complete by the MTM.

This is 100x longer than the replication cycle of the E.coli cell!

Overall, is clear that Mother Nature is beating us badly on density of memory and logic elements, operational speed and operational energy.

I show this table again for us to really see how Mother Nature is beating us badly.

Parameter	*"Ultimate Si μ-Cell"*		*Real μ-Biological cell*	
	Logic	Memory	Logic	Memory
Density	10^{17} cm^{-3}	10^{16} cm^{-3}	10^{18} cm^{-3}	10^{19} cm^{-3}
Energy/bit	10^{3} k_BT	10^{8} k_BT	<10 k_BT	<10^{4} k_BT
Power	10^{-7} W		10^{-13} W	
Heat Flux	1 W/cm^2		10^{-6} W/cm^2	
Energy per 10^{11} output bits	10^{-2} J		< 10^{-9} J	
Time to compute 10^{11} output bits	10^{5} s		10^{3} s	
	I will call this a "trewish" (mixture of trends and wishes).		I will call this a dream.	

So, indeed as Mr. Feynman said, “there is plenty of room at the bottom”.

Apparently Mother Nature has found it.

Can we?

Yes, you may argue that she needed billions of years of evolution by trial and error to come up with HOW.

But these billions of years of evolution resulted in our brains, which are superior to any computer we have built until now…and probably we may build in some years to come.

So, again, can we? Perhaps some clever ATTRACT proposals start “siliconizing cells” or “biologizing silicon”…JUST FOR FUN!

Thanks for your attention

I hope it was fun and inspiring.

References

R. P. Feynman, “There is plenty of room at the bottom”, available @ https://www.pa.msu.edu/~yang/RFeynman_plentySpace.pdf

Victor V. Zhirnov et al., *Proceedings of the IEEE, vol. 91, No. 11, November 2003*.
(and references therein).

Laszlo B. Kish, *Physics Letters A 305 (2002) 144–149*.

Y. Li et al, *Fluctuation and Noise Letters, vol. 6, No. 2 (2006) L127–L131*.

Victor V. Zhirnov et al., *Solid-State Electronics 54 (2010) 810–817*.

Ralph K. Cavin III, et al., *Proceedings of the IEEE, vol. 100, May 2012*.