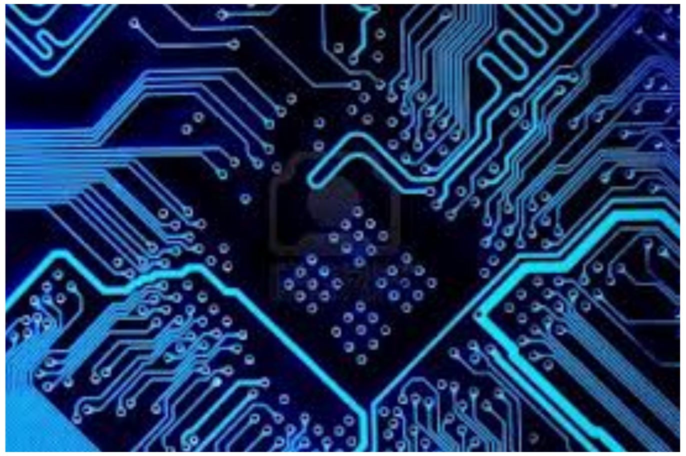
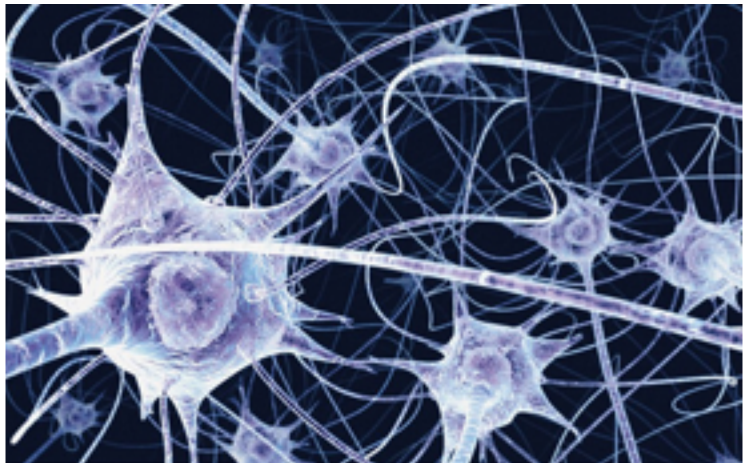
“Information Friction” and Minimum Power Consumed in Data Center and Big Data Networks

Pulkit Grover
CMU

Energy consumed in communication & computing

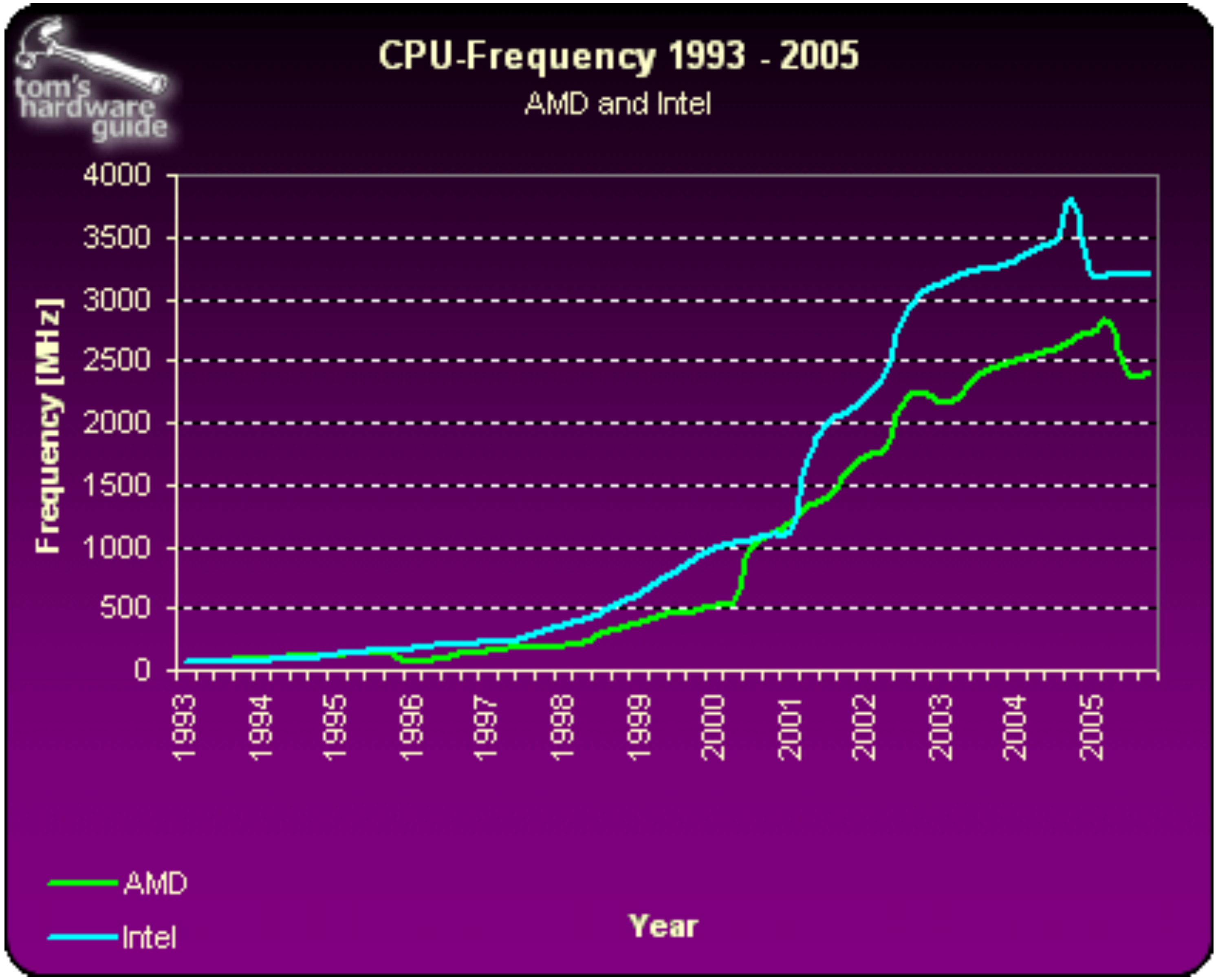
In the body

- 20% of total energy

In the world

- 8% of electricity in 2011
- Projection: 15% by 2020

Moore's law, and the difficulty of reducing computation power



Moore's law is saturating

Higher frequencies cost more power

But this is the era of BigData!

Solution to Moore's law saturation: Parallelize!

Use communication to compute

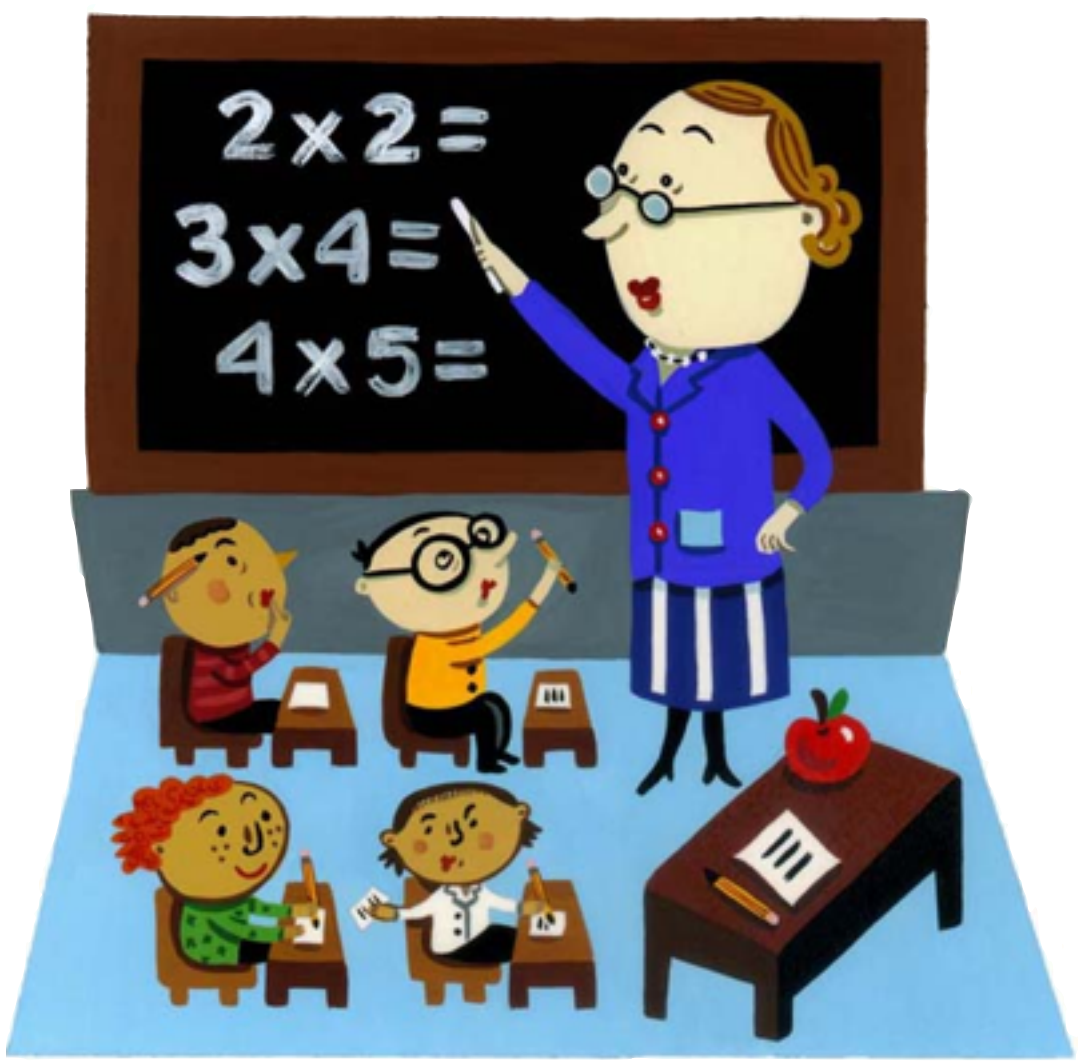
Communication has become a bottleneck

"We need a "Moore's Law" technology for networking" [Byrant, Katz, Lazowska '08]

How much power *must* communication consume?

A soft-spoken teacher's dilemma

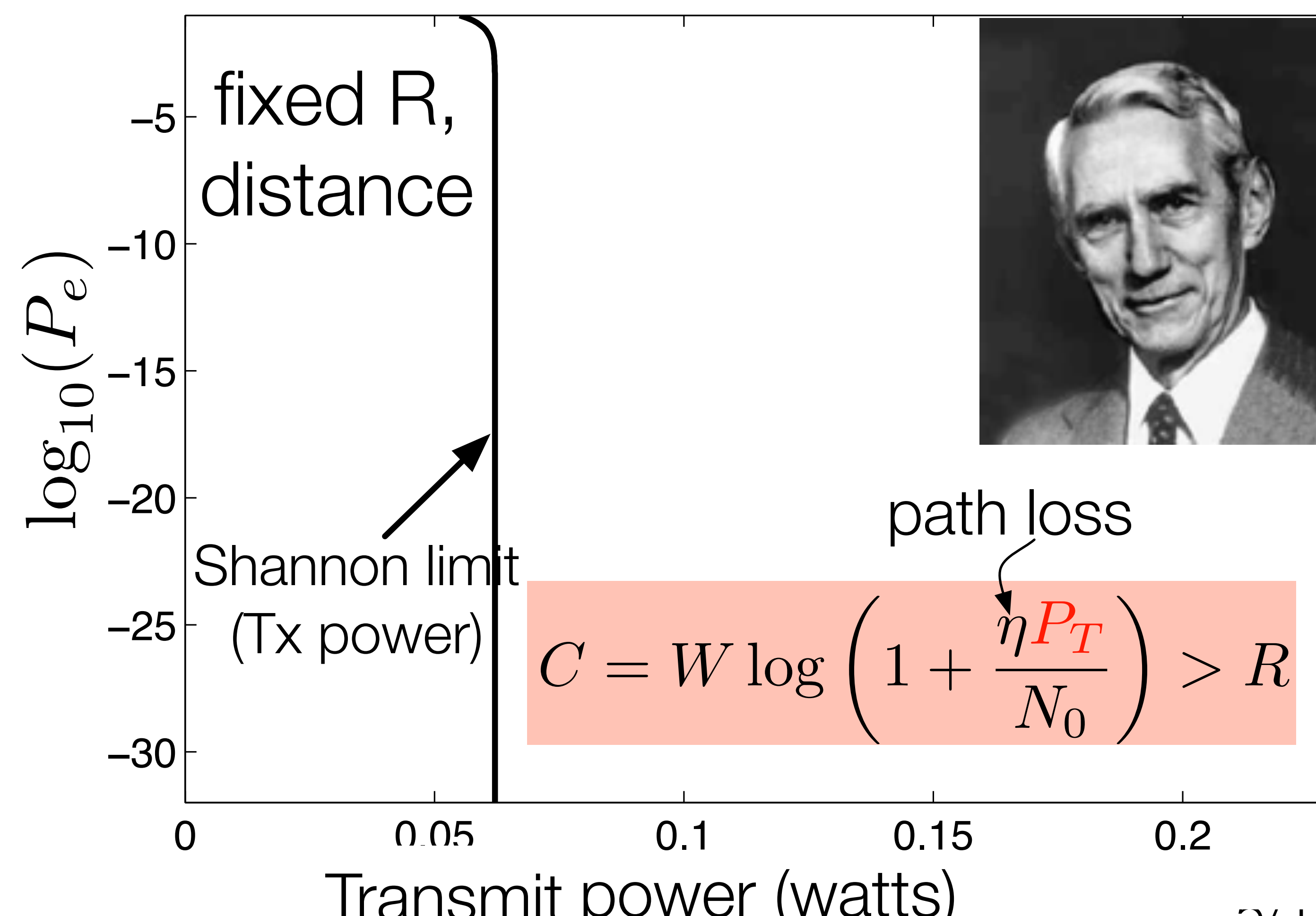
Speak slower? louder?



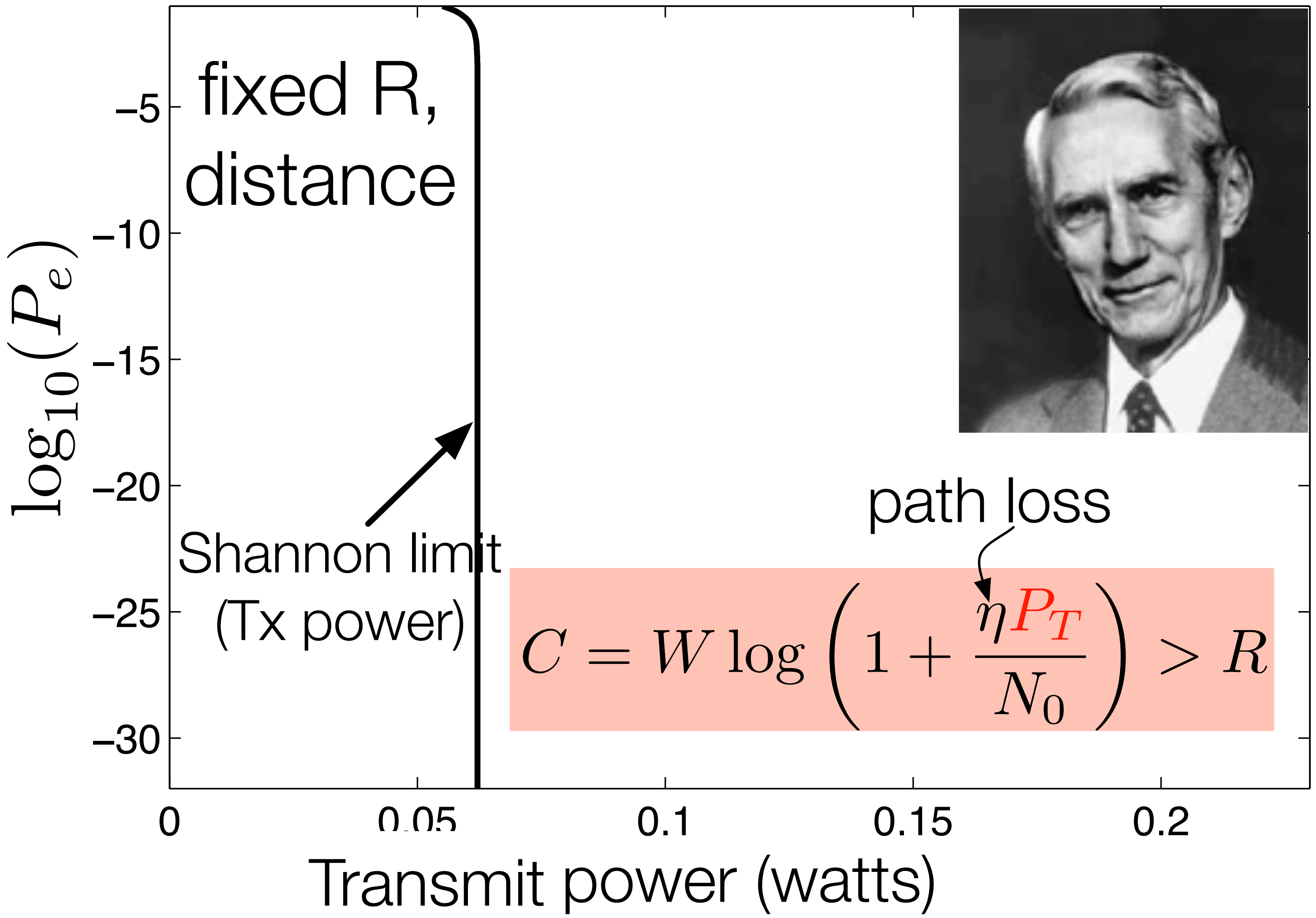
"Shannon's solution"

Neither!

Have a lot to say; "Mix" your information



How can we reduce communication power?

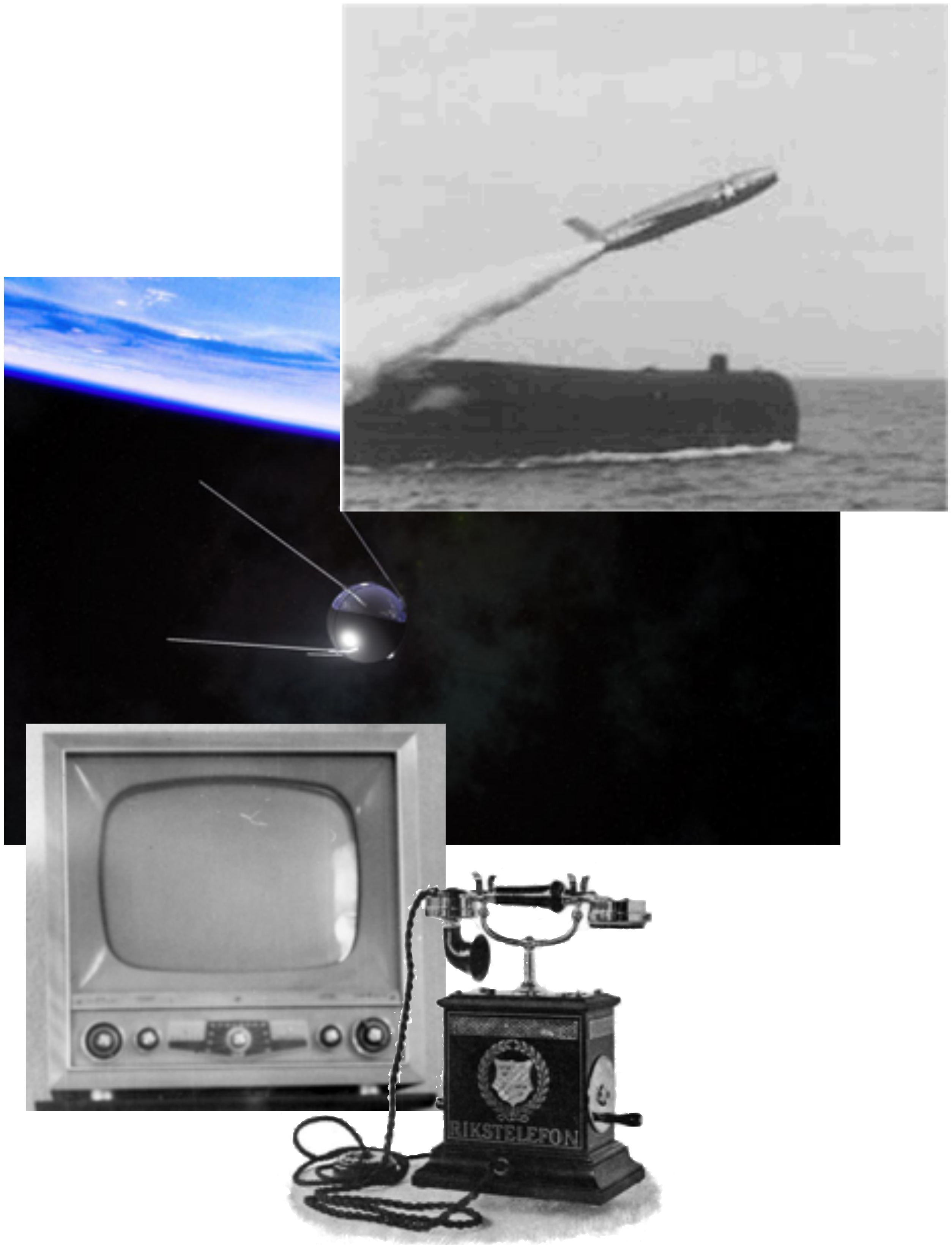


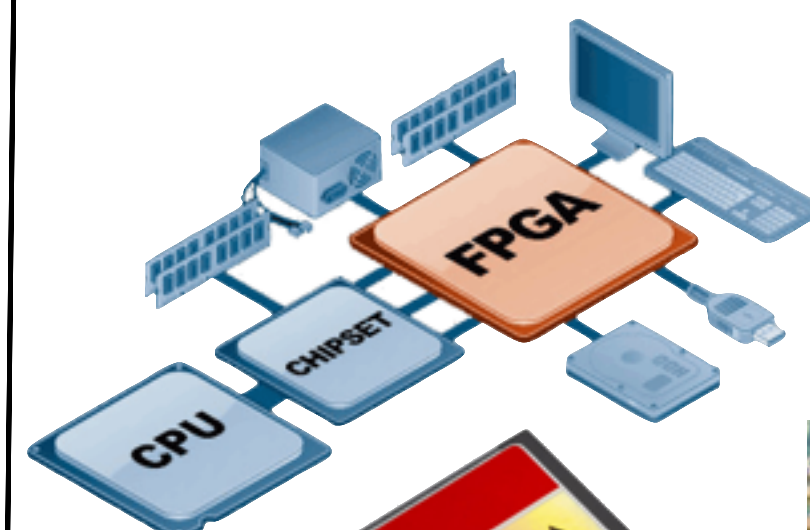
Shannon limit is already achieved

[Turbo Codes ’93] [LDPC codes ’98] [Polar Codes ’08]

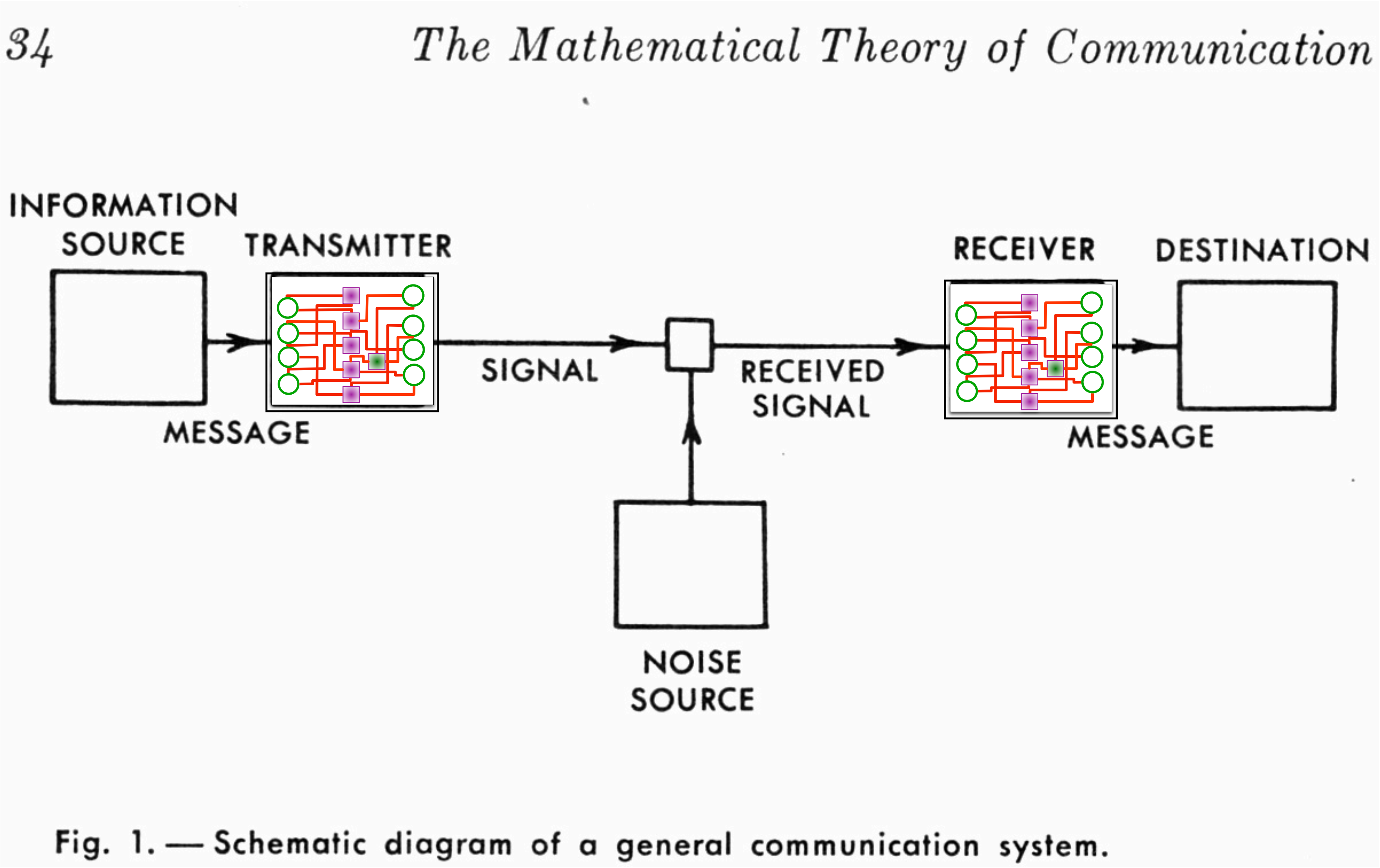
We are already at the theoretical minimum transmit power

Communication systems: then and now



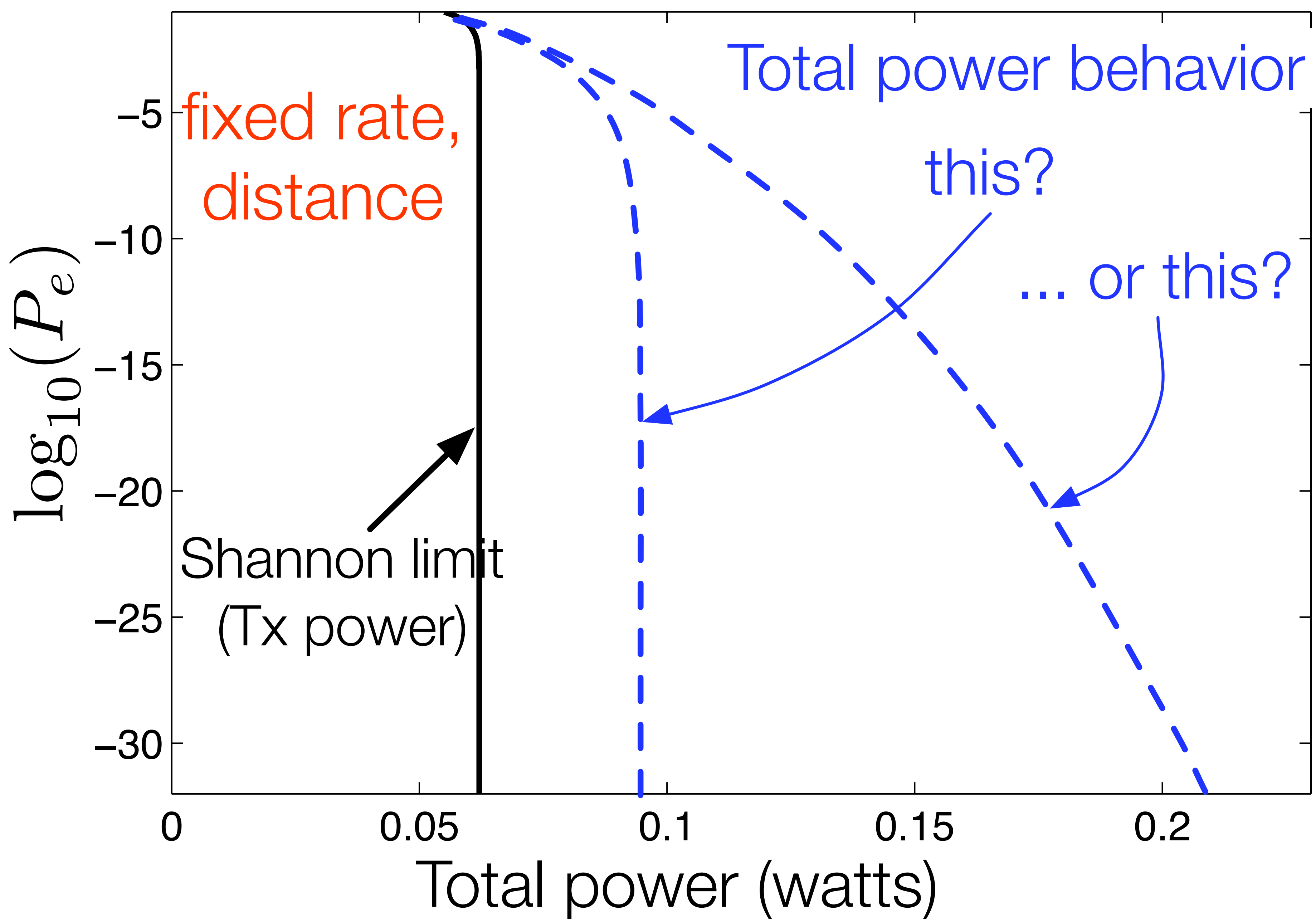


The problem has changed



We now want to minimize total (transmit + circuit) power

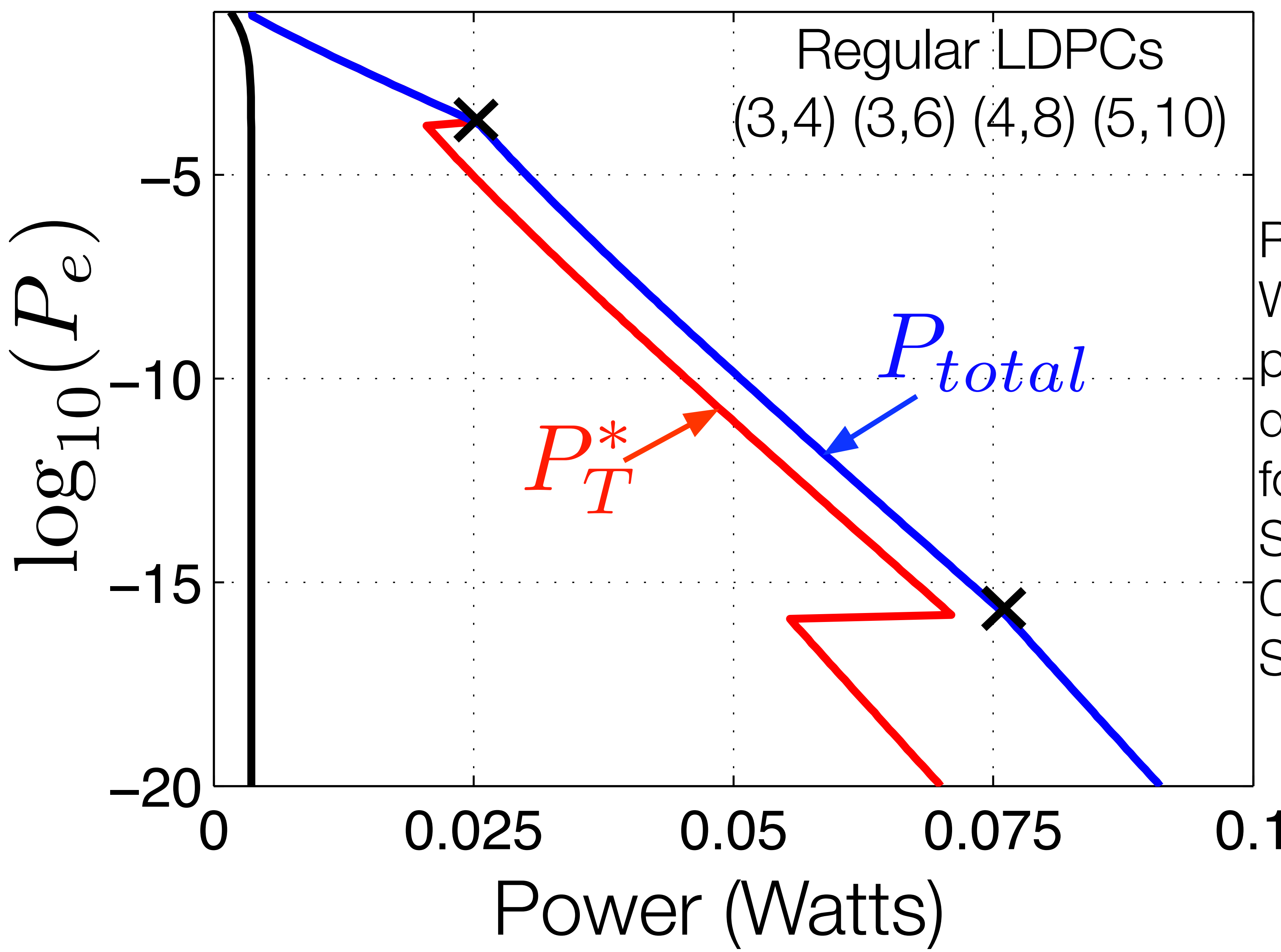
What could fundamental limits on total power look like?



Total power: Hints from experiments

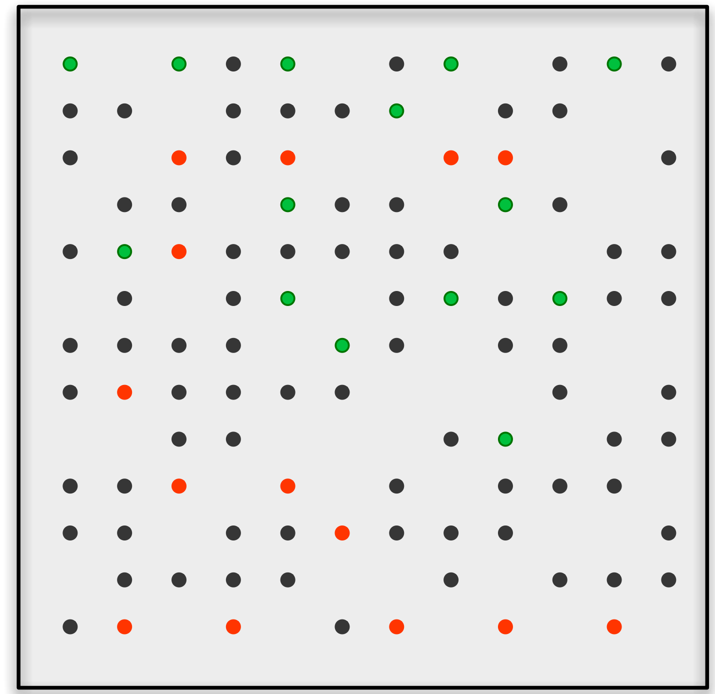
Circuit-simulation-based modeling of decoding power

[Ganesan, Grover, Rabaey SiPS'11][Ganesan, Wen, Grover, Rabaey'12]



R = 7 Gbps
W = 7 GHz
path-loss coefft = 3
d = 3 meters
fc = 60 GHz
STM CMOS 90 nm
Cadence Encnter (Layout)
Synopsys HSPICE (post-lyt)

“Information-friction” energy model



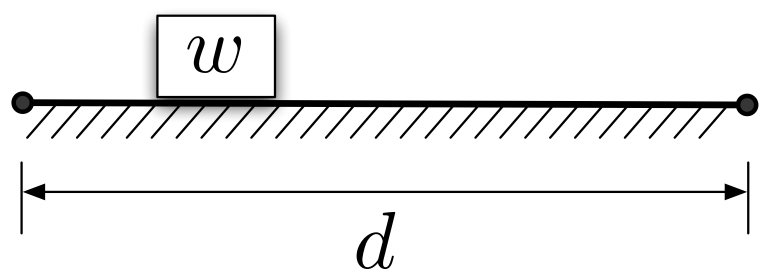
<u>Circuit implementation model:</u>

Computational nodes that can talk to one another over circuit links

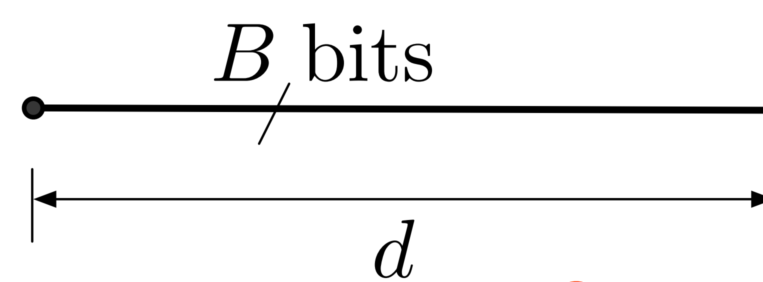
Min distance between nodes $= \lambda$

<u>“Info-friction” model of energy consumed in circuit-links:</u>

weight w



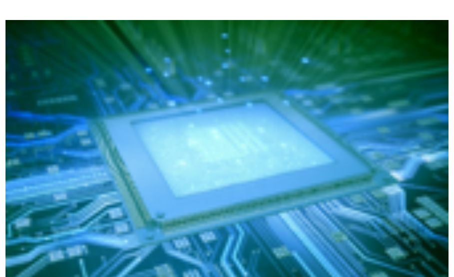
$$E_{\text{friction}} = \mu \; w \; d$$

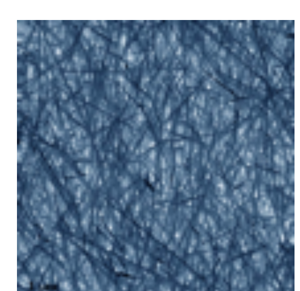


$$E_{\text{info-friction}} = \mu \; B \; d$$

Coefficient of informational friction

<u>Example applications:</u>



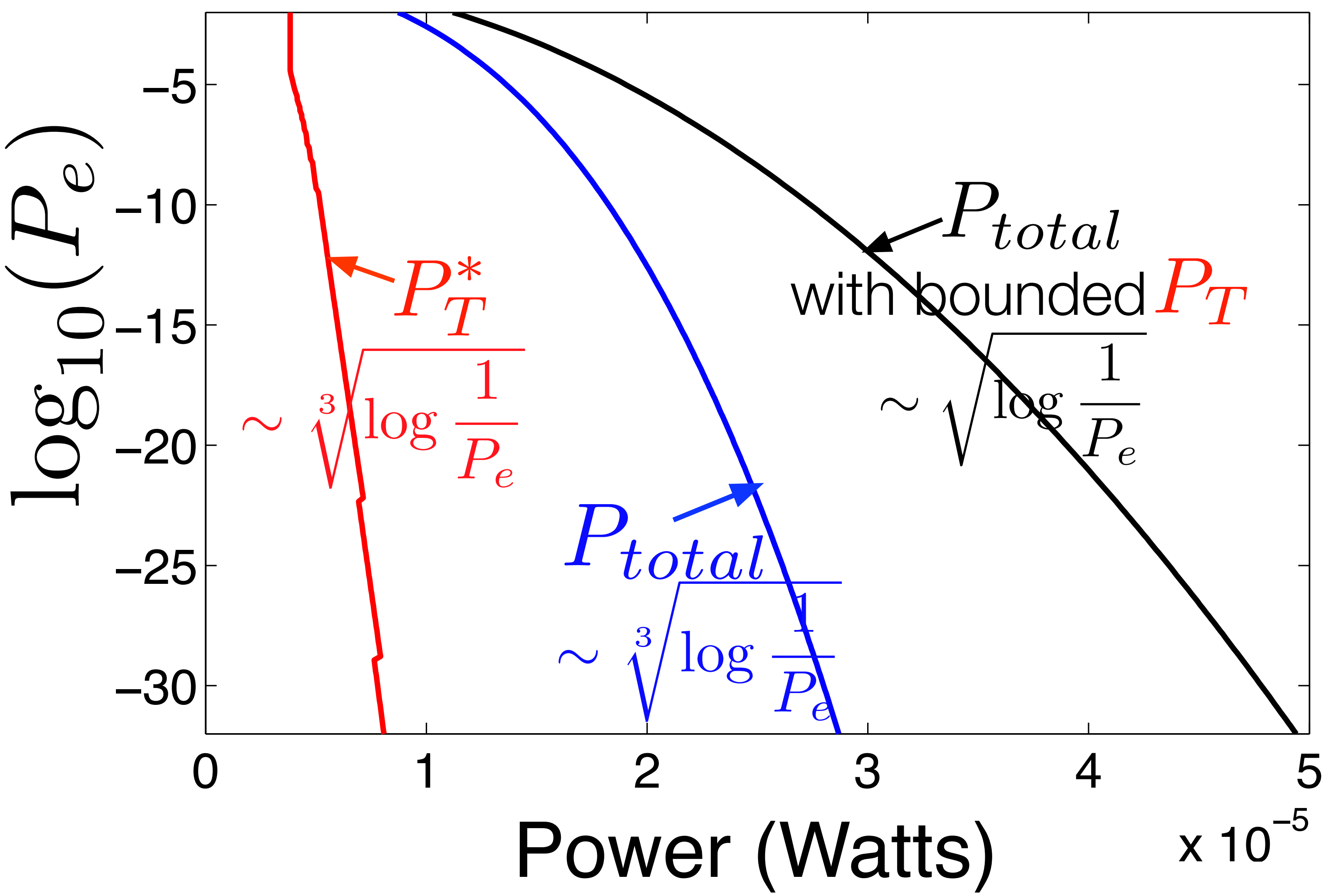


Info-friction losses in encoding/decoding

Theorem [Grover, ISIT '13]

$$E_{\text{enc,dec}} \geq \Omega\left(k\sqrt{\frac{\log\frac{1}{P_e}}{P_T}}\right)$$

for any code, and any encoding & decoding algorithm implemented in the circuit model



In comparison, uncoded transmission:

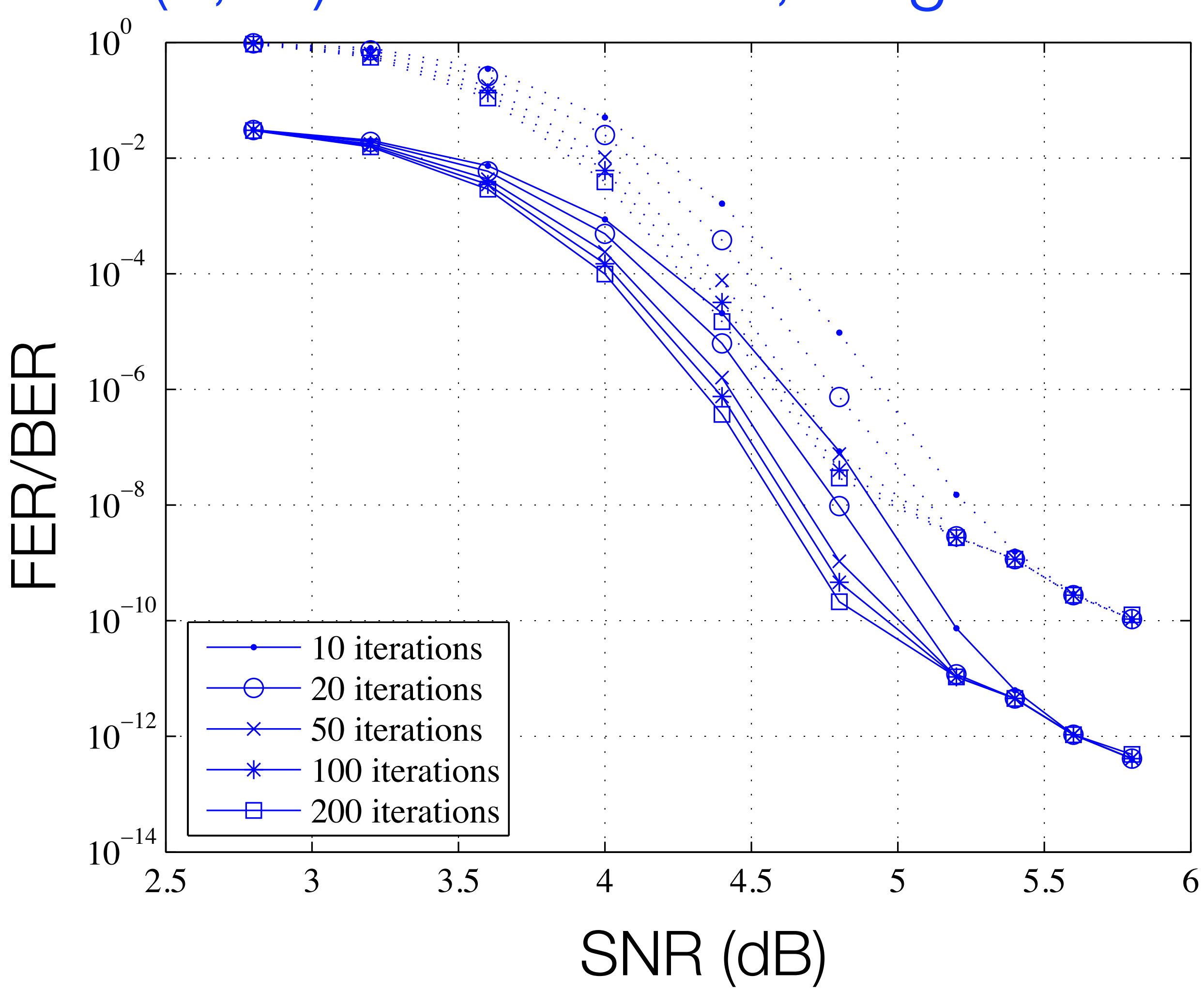
$$P_T \sim \log\frac{1}{P_e}$$

Experimental corroboration [Ganesan, Grover, Goldsmith, Rabaey '12]

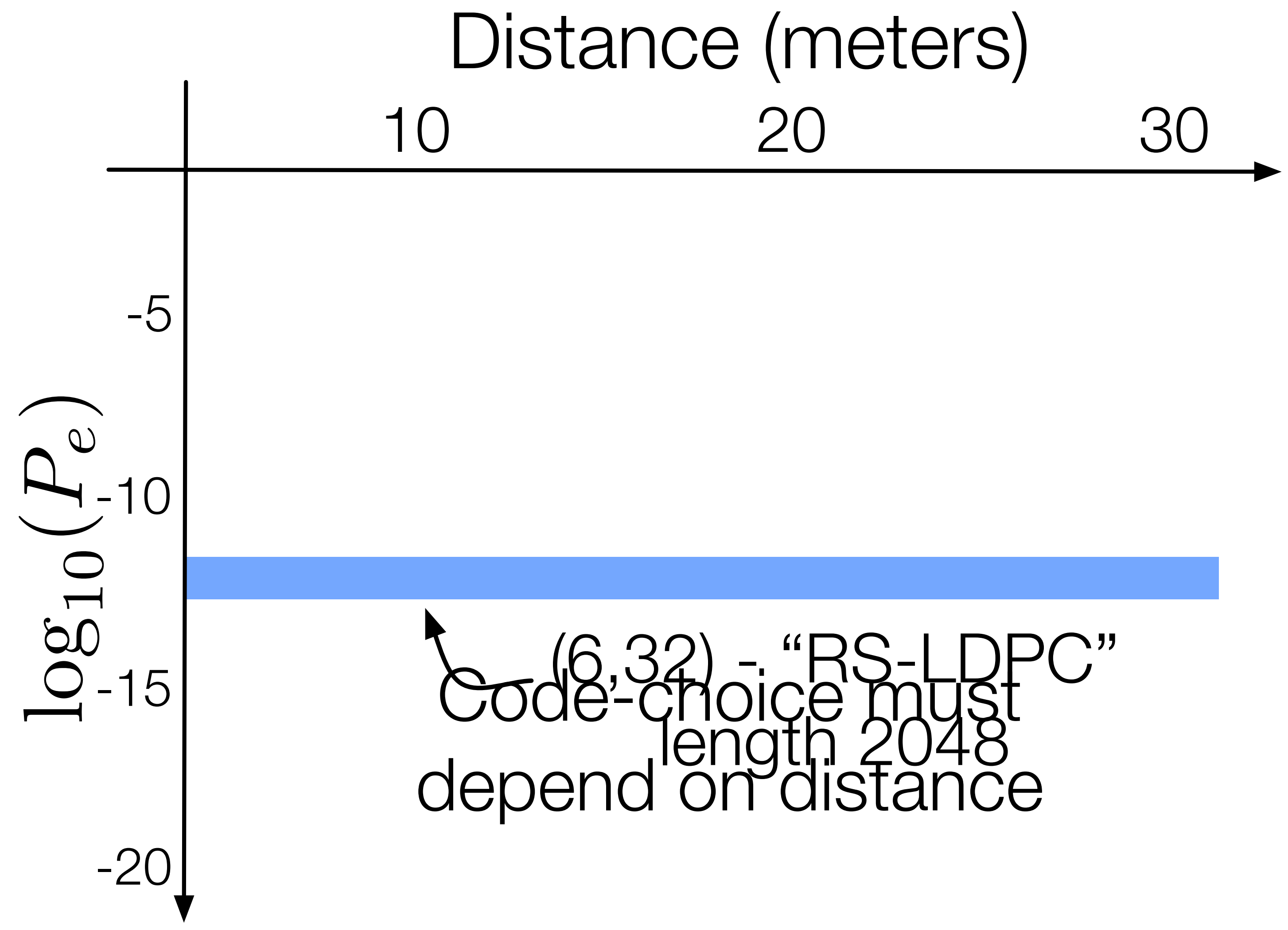
Why does this matter?

Traditional view

Code choice (10-Gbase-T):
(6,32) - “RS-LDPC”, length 2048

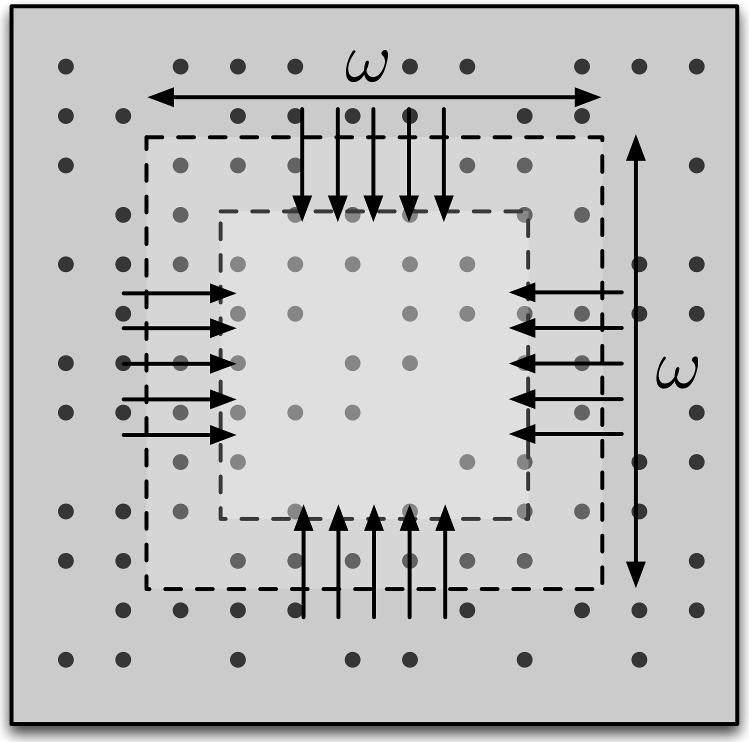


Total-power perspective



New codes that adapt and minimize total power [Mahzoon, Yang, Grover in prep]

Towards fundamental limits on energy of computing networks



Total computation energy $\geq$ on-chip energy + cable-energy

$$\geq \mu_{on\text{-}chip} B_{on\text{-}chip} d_{on\text{-}chip} + \mu_{cable} B_{cable} d_{cable}$$

Total energy for sorting, classification, …

Can noisy or approximate computing help?

The Shannon waterfall has inspired a study of noisy computing

[von Neumann'56]
[Pippenger'88]
[Hajek, Weller'91]
[Evans, Schulman'99]
… [Shanbhag, Kumar, Jones'10]

Theorem [Grover, ISIT '14 Sub.]

$$E_{total\ per\text{-}bit} \geq \Omega\left(\sqrt[3]{\log \frac{1}{P_e}}\right)$$

for any implementation with Gaussian noise in circuit elements

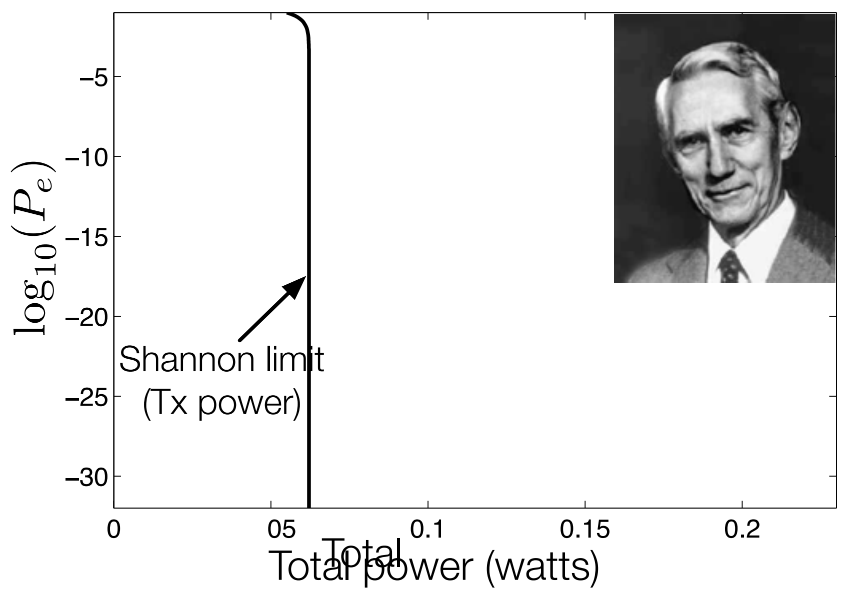
… i.e., the "Shannon-capacity" of noisy computing is zero!
[Grover, ISIT'14 Sub.]

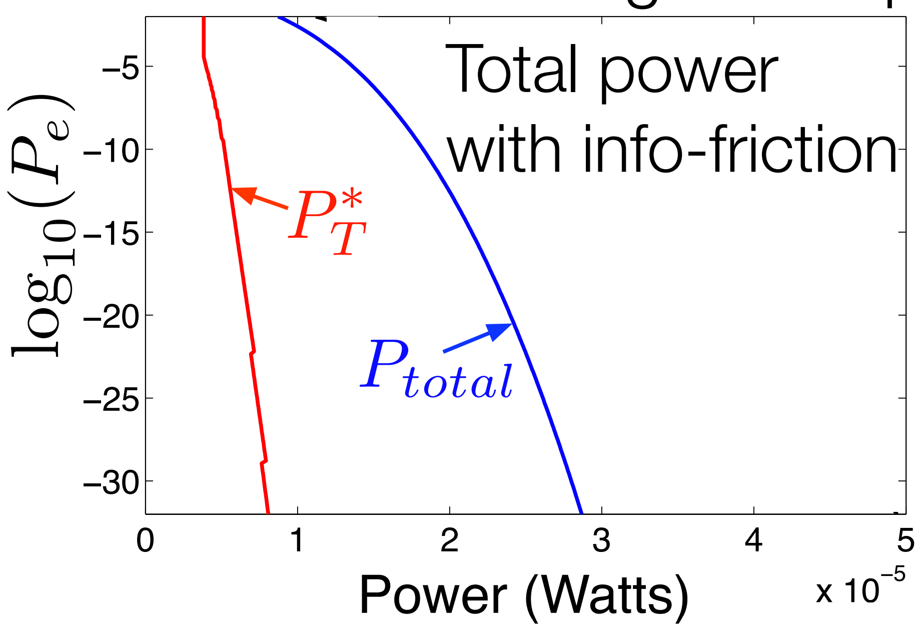
Nevertheless, noisy computing could help,
. . . but don't expect Shannon waterfall's "magic"

Summary

1) Traditional info theory ignores energy of encoding/decoding

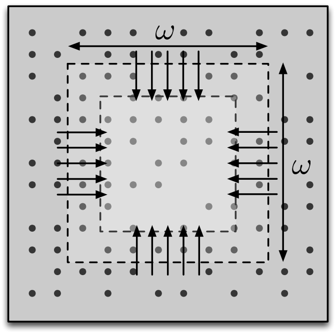
- useful in long distance communication
- misleading in computation





2) Info-friction model: fundamental limits on energy of various computations

- sorting/ranking, Fourier transforms, classification



3) Insights towards "good" computational network designs

4) Still need to understand noise in computing and its impact

Extra slides

How much power must communication consume?

The physicist

Minimal Energy Requirements in Communication

Rolf Landauer [1996]

The literature describing the energy needs for a communications channel has been dominated by analyses of linear electromagnetic transmission, often without awareness that this is a special case. This case leads to the conclusion that an amount of energy equal to $kT \ln 2$, where kT is the thermal noise per unit bandwidth, is needed to transmit a bit, and more if quantized channels are used with photon energies $h\nu > kT$. Alternative communication methods are proposed to show that there is no unavoidable minimal energy requirement per transmitted bit. These methods are invoked as part of an analysis of ultimate limits and not as practical procedures.

mmunication

path loss
(channel "frictional" loss)

$$C = W \log\left(1 + \frac{\eta P_T}{N_0}\right)$$

Noise

The engineer

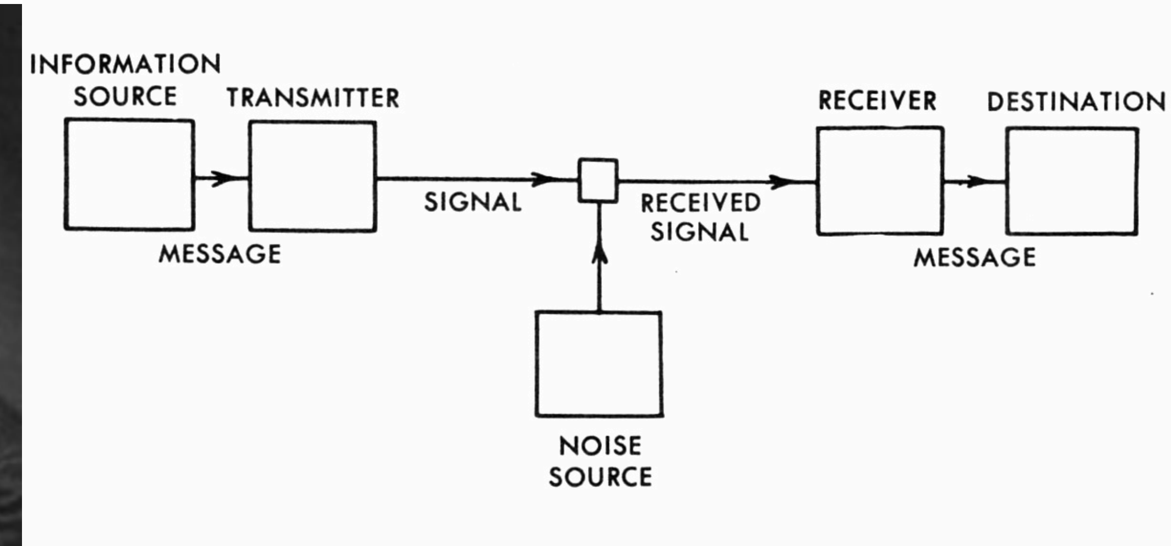


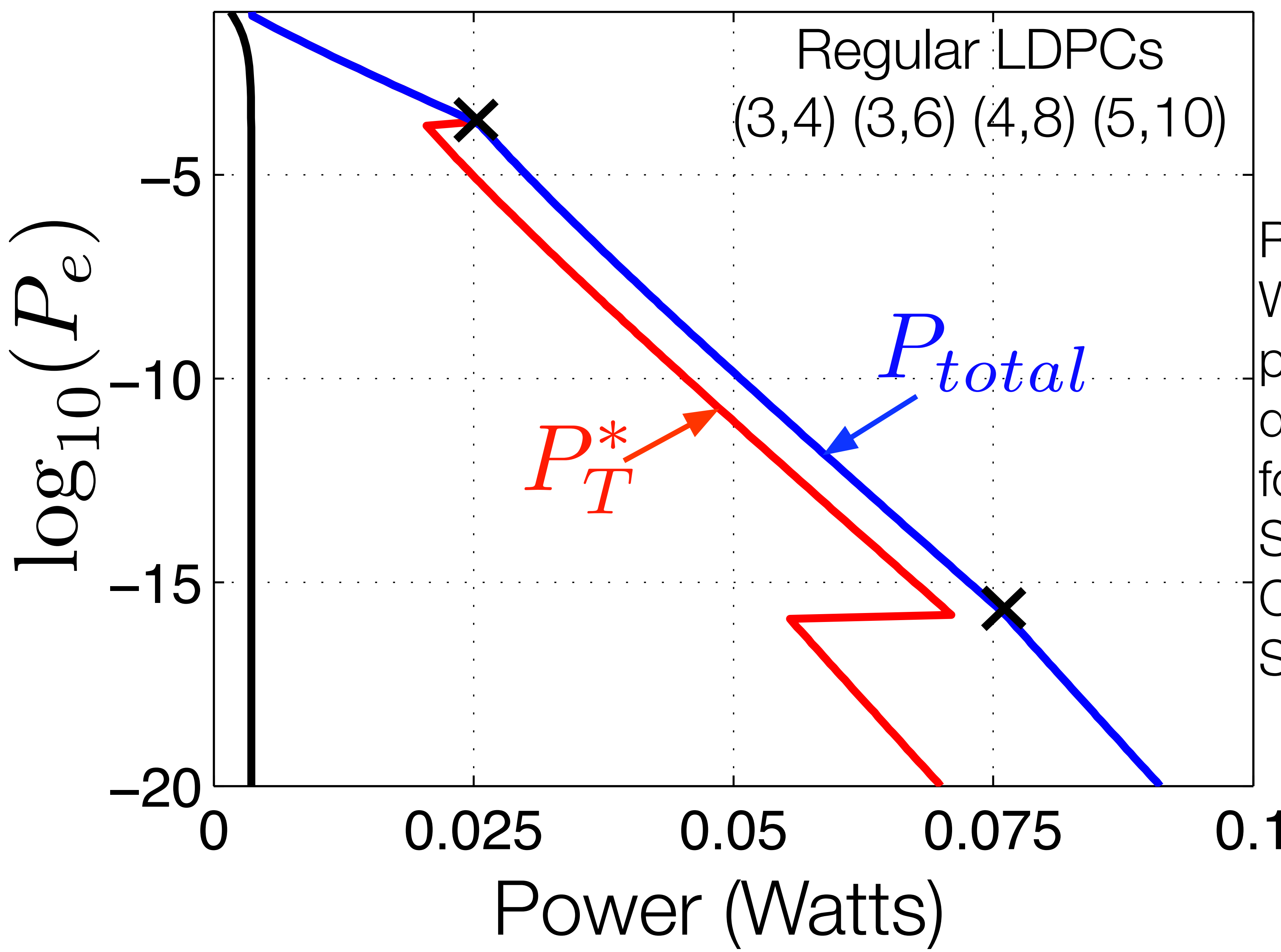
Fig. 1. — Schematic diagram of a general communication system.

[Shannon '48]

Total power: Hints from experiments

Circuit-simulation-based modeling of decoding power

[Ganesan, Grover, Rabaey SiPS'11][Ganesan, Wen, Grover, Rabaey'12]



R = 7 Gbps
W = 7 GHz
path-loss coefft = 3
d = 3 meters
fc = 60 GHz
STM CMOS 90 nm
Cadence Encnter (Layout)
Synopsys HSPICE (post-lyt)

Decoding complexity/power: Fundamental limits

Theorem*[Grover, Woyach, Sahai ISIT '08, JSAC '11]

# of clock-cycles, $\tau \gtrsim \frac{1}{\log(\alpha - 1)} \log\left(\frac{\log \frac{1}{P_e}}{(C(P_T) - R)^2 / V(P_T)}\right)$

... for **any** code and **any** decoding algorithm.

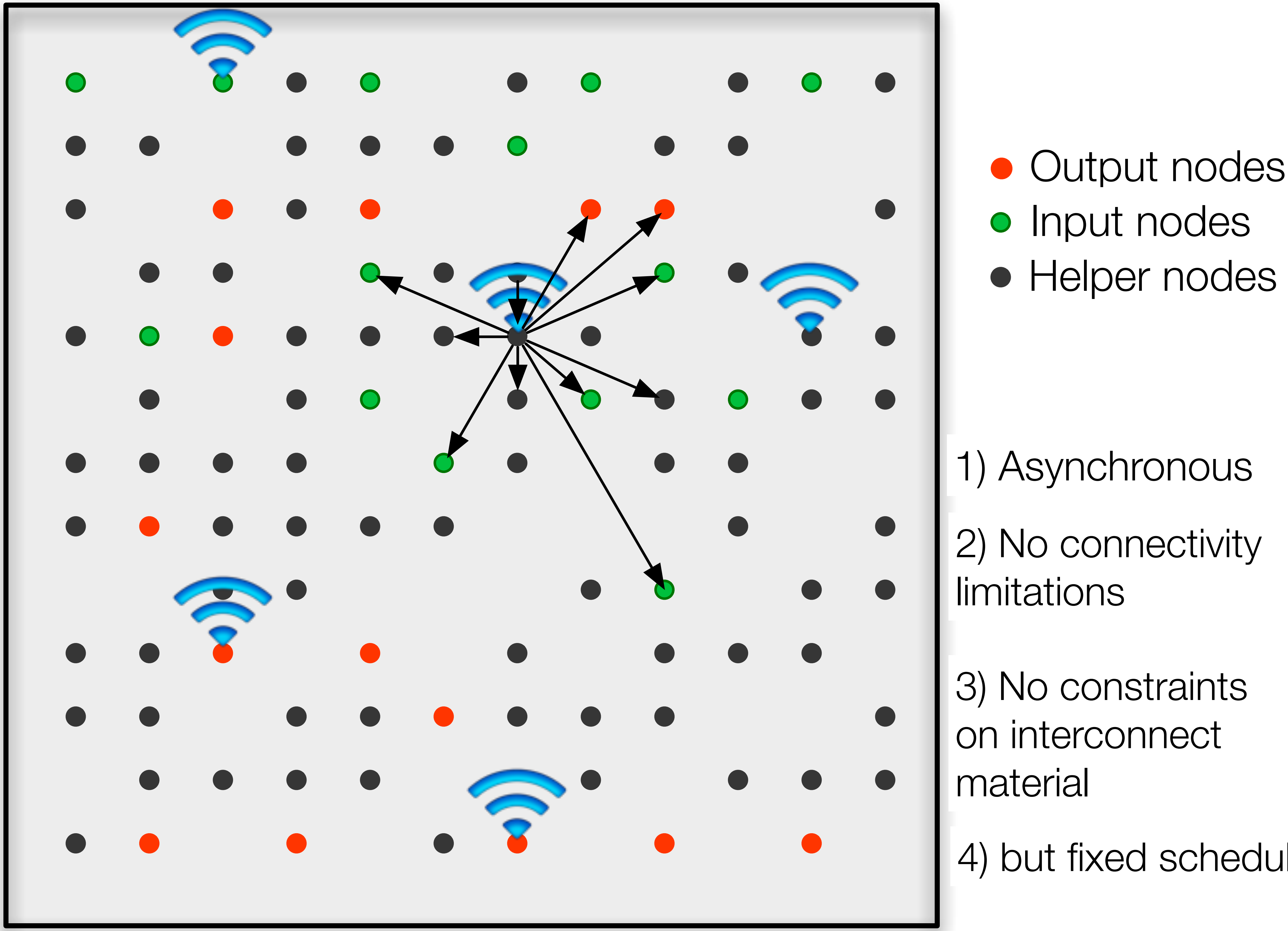
α : Max. node degree

$V(P_T)$: "channel dispersion"

Node mode: E_{node} joules per clock-cycle

* precise result for any P_e and any gap appears in [Grover, Woyach, Sahai '11]

Moving information: a model of implementation



1) Asynchronous

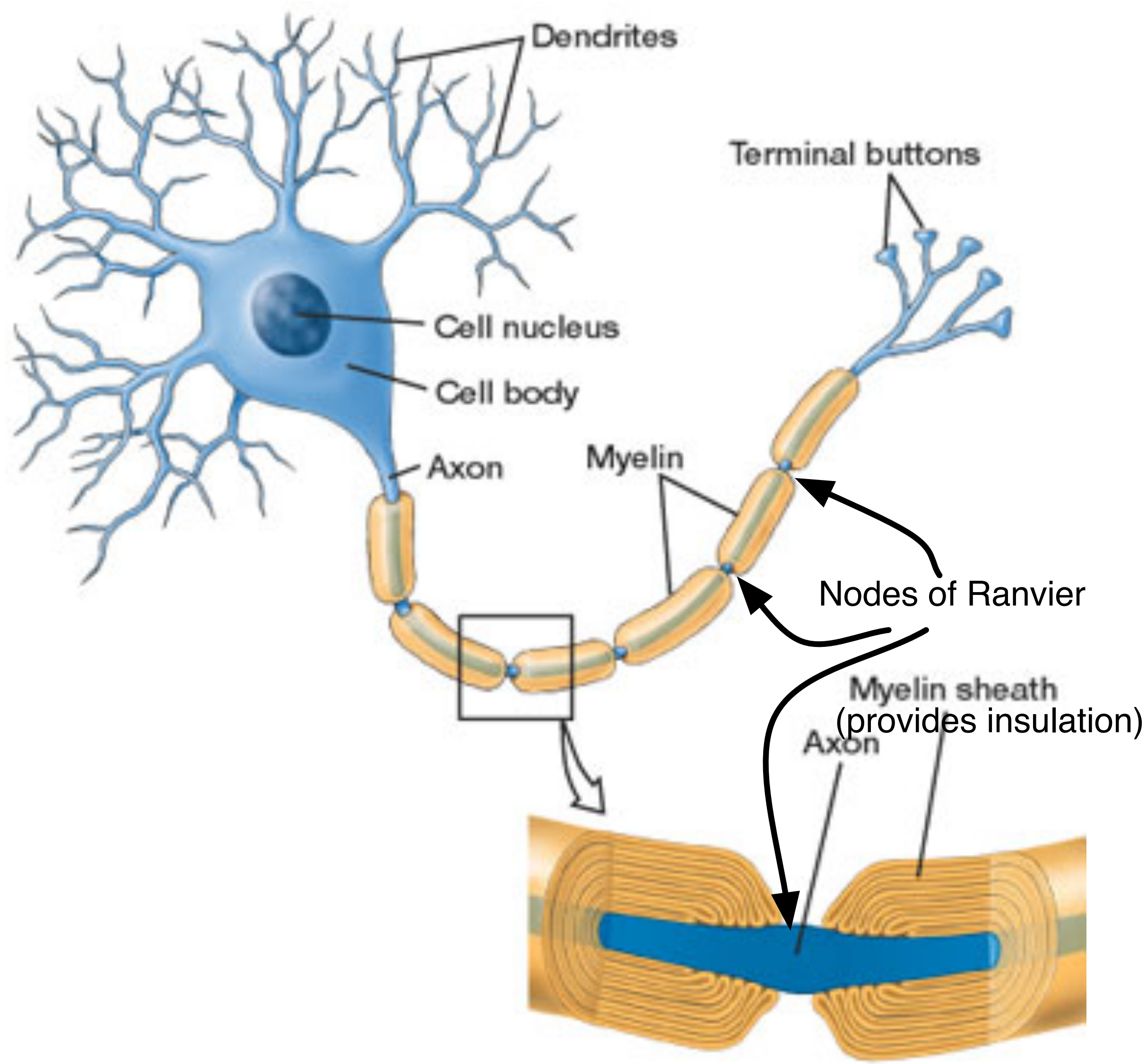
2) No connectivity limitations

3) No constraints on interconnect material

4) but fixed schedule

[Grover '13]

Looking forward: Neuronal and neuro-morphic processing



friction + noise in computing?