Haute Ecole Spécialisée
de Suisse occidentale
Fachhochschule Westschweiz
University of Applied Sciences and Arts
Western Switzerland

Energy efficiency of internet traffic - models and bounds

Carbon Aware Networks Workshop - 2023

Sébastien Rumley

Reminder

- Kilo 10^{3}
- Mega 10^{6}
- Giga 10^{9}
- Tera 10^{12}
- Peta 10^{15}
- Exa 10^{18}
- Zetta 10^{21}

- 1B = 10 bits

- Milli 10^{-3}
- Micro 10^{-6}
- Nano 10^{-9}
- Pico 10^{-12}
- Femto 10^{-15}
- Atto 10^{-18}
- Zepto 10^{-21}

- 1 Joule = 1 Watt · 1 second
- 1 Watt = 1 Joule / 1 second

- 1 pJ/bit = 1 (pW · s)/bit
  = 1 pW/(bit/s)
  = 1 nW/Kbps
  = 1 uW/Mbps
  = 1mW/Gbps
  = 1W /Tbps
- 1 kWh/GB = 360 uJ/bit
- 1 Wh/GB = 360 nJ/bit

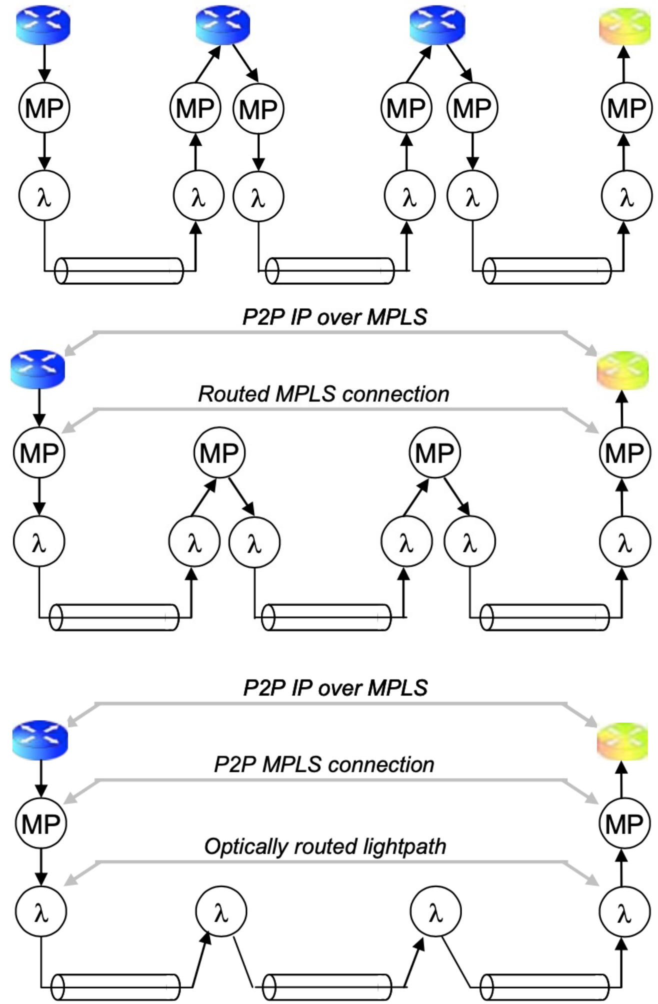
My background

- 2005 – 2012 : PhD, Long distance, backbone optical networking

Green design

Figures of merit (optics)

- Switching/grooming cost → **Switching energy**
- Lightpath cost → **Transmission energy**
- Fiber cost



Energy Saving Perspectives of IP over Wavelength Division Multiplexing

Sébastien Rumley, Christian Gaumier
Ecole Polytechnique Fédérale de Lausanne (EPFL)
TCOM – STI – EPFL, Station 11, 1015 Lausanne, Switzerland
E-mail: sebastien.rumley@a3.epfl.ch

S. Rumley, C. Gaumier, "Energy Saving Perspectives of IP over Wavelength Division Multiplexing", in Proceedings of the COST Action IC804, 2nd year, https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=1b811f0d0de148b1151792a2a5fc4afb9a9f57d5

More background

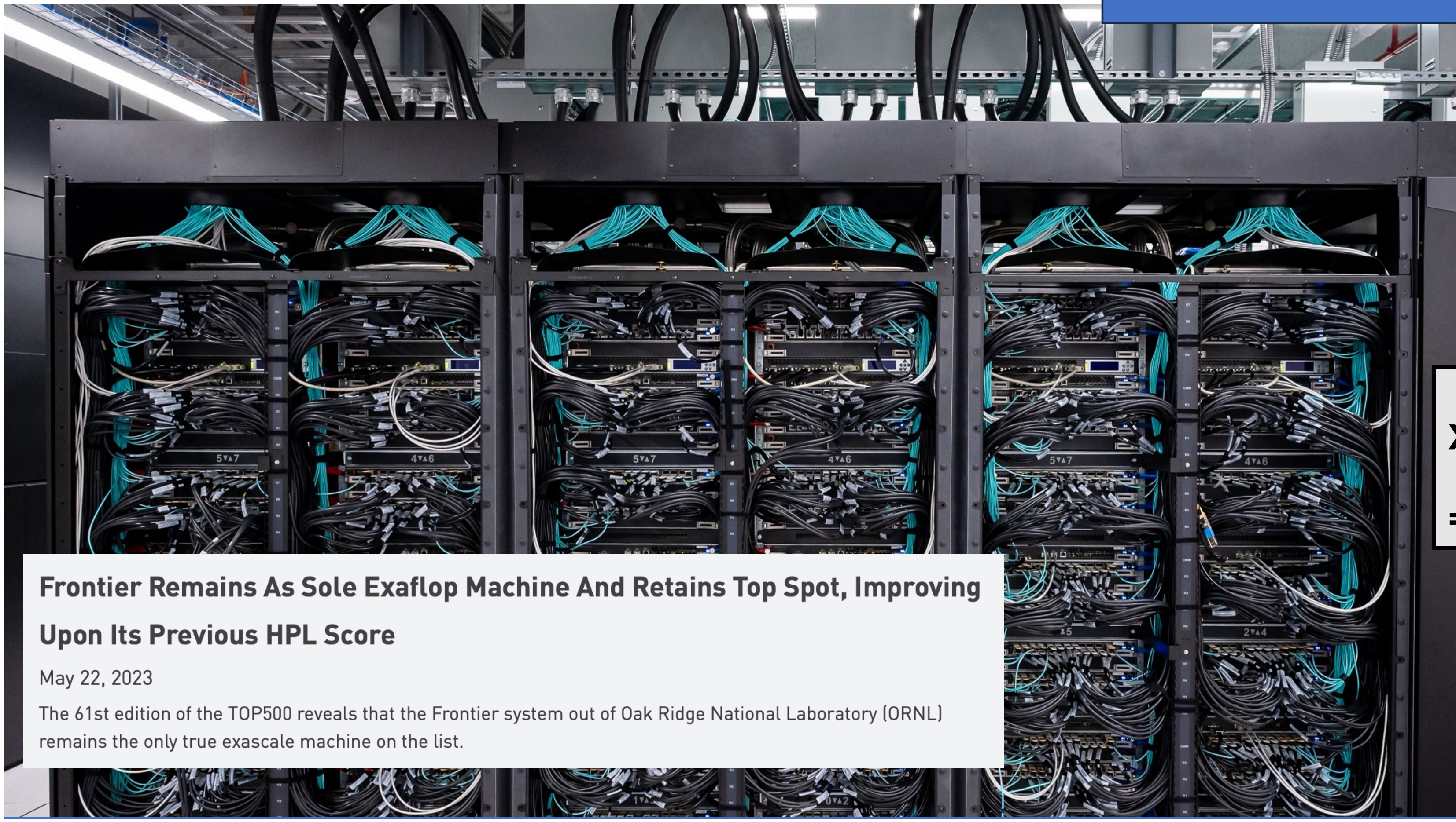
- 2012 – 2018: Post-doc, (optical) high performance interconnects

Exaflop race

Our prediction in 2017

1.25 ExaFLOP
X 0.01 B/FLOP

= 125 Pb/s injection BW



Frontier Remains As Sole Exaflop Machine And Retains Top Spot, Improving Upon Its Previous HPL Score

May 22, 2023

The 61st edition of the TOP500 reveals that the Frontier system out of Oak Ridge National Laboratory (ORNL) remains the only true exascale machine on the list.

1.69 ExaFLOP
X 0.0018 B/FLOP

= 8 Pb/s injection BW

Power and power

- The DOE set a cap on supercomputer consumption very early on (~2008) [1]
- No more than 20MW for electronics!
  - Frontier (previous slide) was measured at 22MW during the Exascale benchmark
- Assume 10% of this goes to the interconnect
  - 2 Mwatt / **8 Pb/s** = **250 mW/Gb/s** = **250 pJ/bit end-to-end thru the interconnect**
  - **Around 60 pJ/bit/hop**
    - **Switching : 30 pJ/bit**
    - **Transmission : 10 pJ/bit (elec) + 20 pJ/bit (optical)**

[1] Bergman, Keren, et al. "Exascale computing study: Technology challenges in achieving exascale systems." *Defense Advanced Research Projects Agency Information Processing Techniques Office (DARPA IPTO), Tech. Rep* 15 (2008): 181.

[2] Sébastien Rumley, Meisam Bahadori, Robert Polster, Simon D. Hammond, David M. Calhoun, Ke Wen, Arun Rodrigues, Keren Bergman, Optical interconnects for extreme scale computing systems, Parallel Computing, Volume 64, 2017

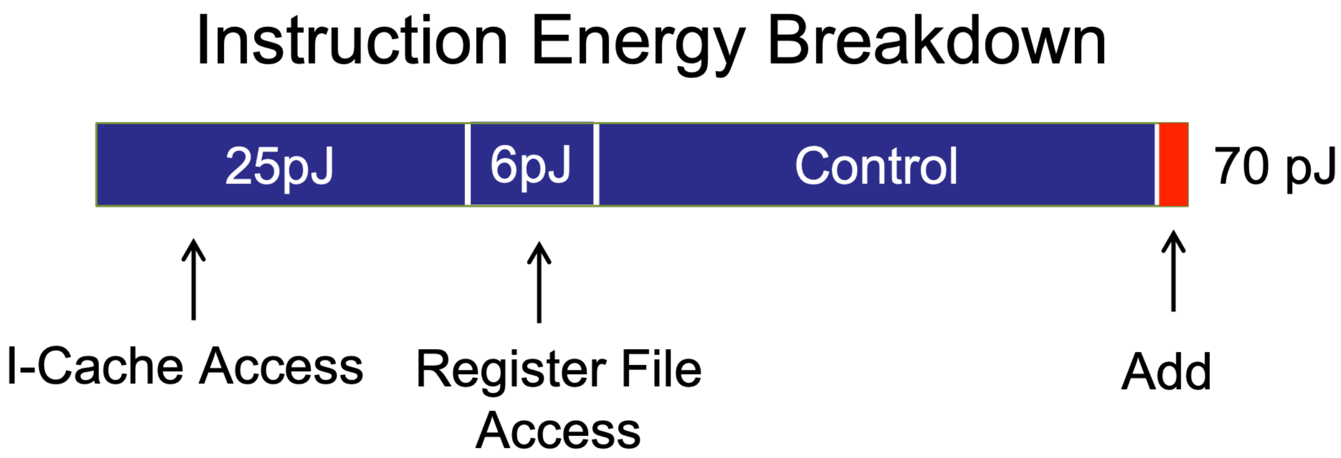
The pJ/bit obsession

Integer	
Add	
8 bit	0.03pJ
32 bit	0.1pJ
Mult	
8 bit	0.2pJ
32 bit	3 pJ

~3-4 fJ/bit

~25 fJ/bit

~100 fJ/bit



- Given that a 32bit instruction involves ~100 bit, the energy per bit (on chip) is 70 pJ / 100 bits = 0.7 pJ/bit
- Computer architects and chip designers are obsessed about the 1pJ/bit mark for off-chip interfaces

M. Horowitz, 1.1: Computing's Energy Problem: (and what we can do about it), International Solid-State Circuits Conference, 2014 IEEE

Import/export business

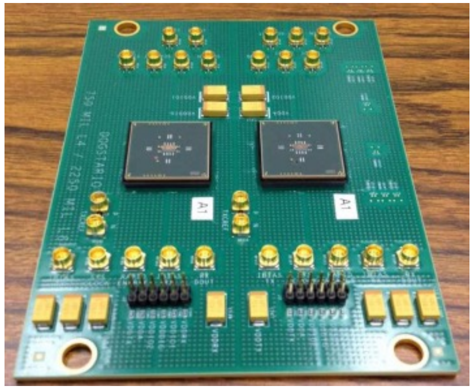
Memory	
Cache	(64bit)
8KB	10pJ
32KB	20pJ
1MB	100pJ
DRAM	1.3-2.6nJ

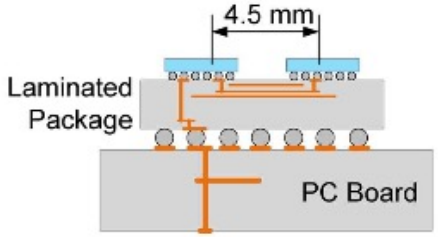
~150 fJ/bit → 8KB
~1.5 pJ/bit → 1MB
~30 pJ/bit → DRAM

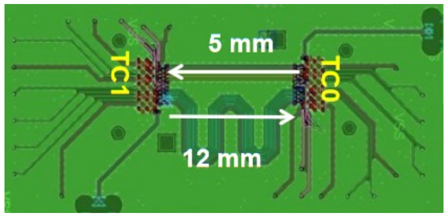
Numbers in literature
- Access SRAM O(10fJ/bit)
- Access DRAM cell O(1 pJ/bit)
- Movement to HBM/MCDRAM O(10 pJ/bit)
- Movement to DDR3 off-chip O(100 pJ/bit)

- IBM Watson, 2015 (Dickson et al.):
  - 6 Gb/s per pin
  - 1.4 pJ/bit over 19mm
- IBM Watson, 2015 (Dickson et al.):
  - 16 Gb/s per pin
  - 1.9 pJ/bit over 250mm
- Nvidia, 2013 (Poulton et al.):
  - 20 Gb/s per pin
  - 0.54 pJ/bit over 4.5mm
- Kandou bus, 2016 (Shokrollahi et al.):
  - 20.83 Gb/s per pin
  - 0.94 pJ/bit over 12mm

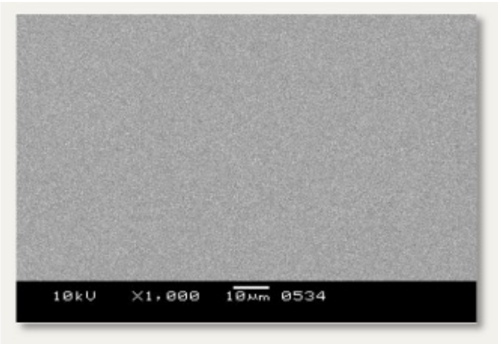
➔ <pJ/bit : immediate chip surrounding only

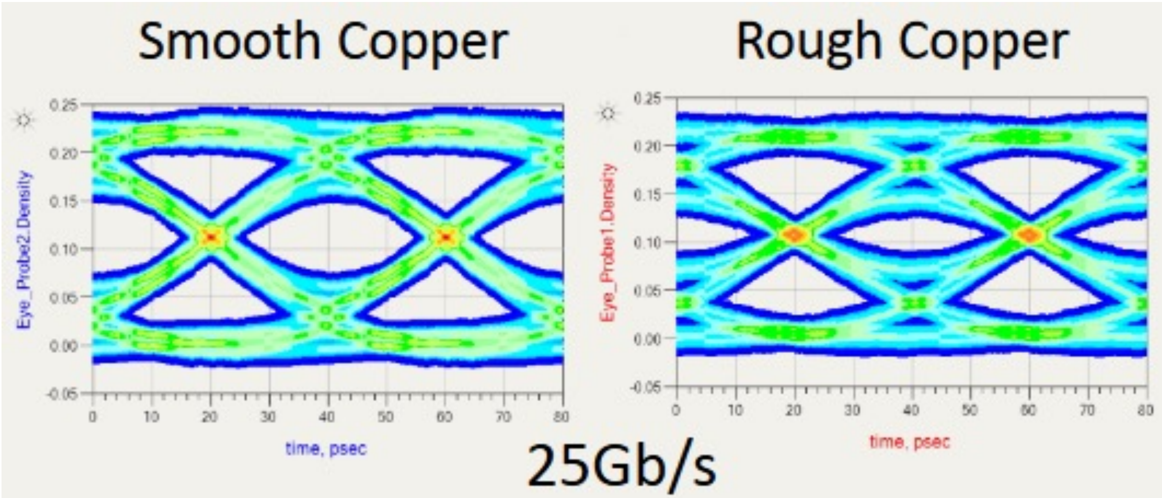


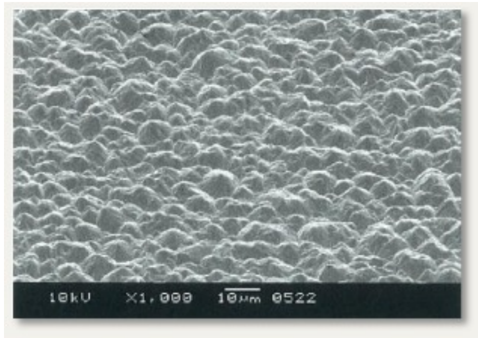




Matter matters



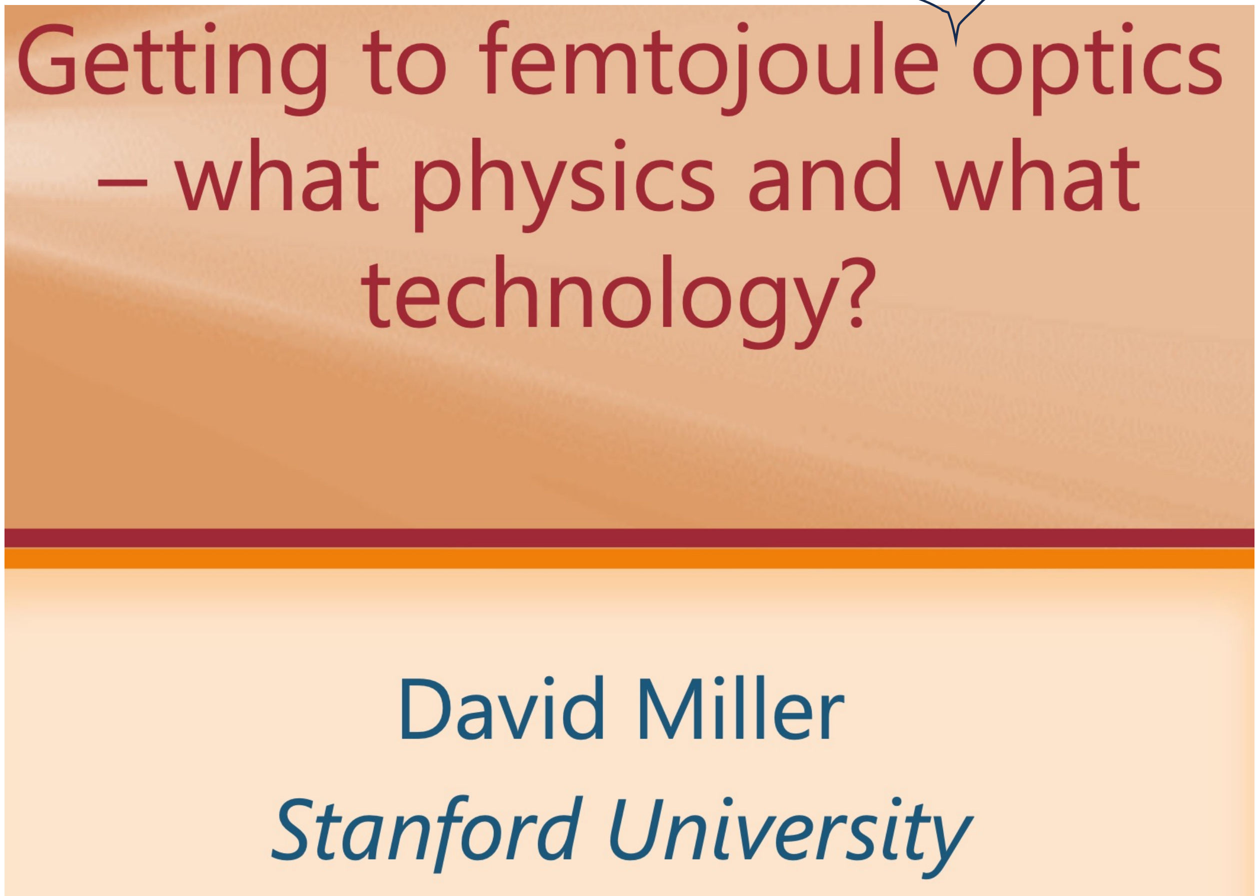




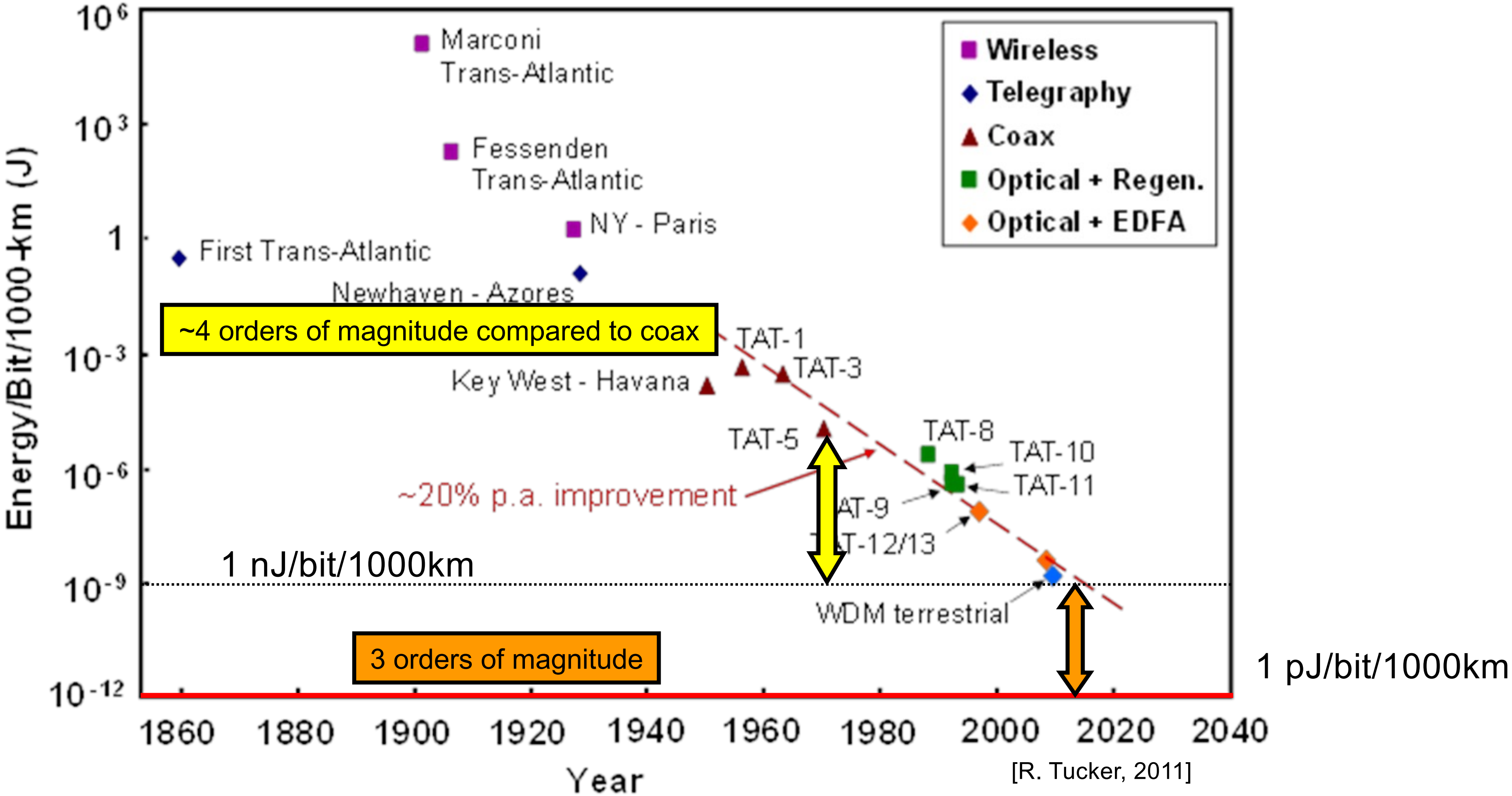
Bert Simonovich, Practical Method for Modeling Conductor Surface Roughness Using Close Packing of Equal Spheres, DesignCon 2015

University of Applied Sciences and Arts
Western Switzerland

per bit



Energy efficiency of optics



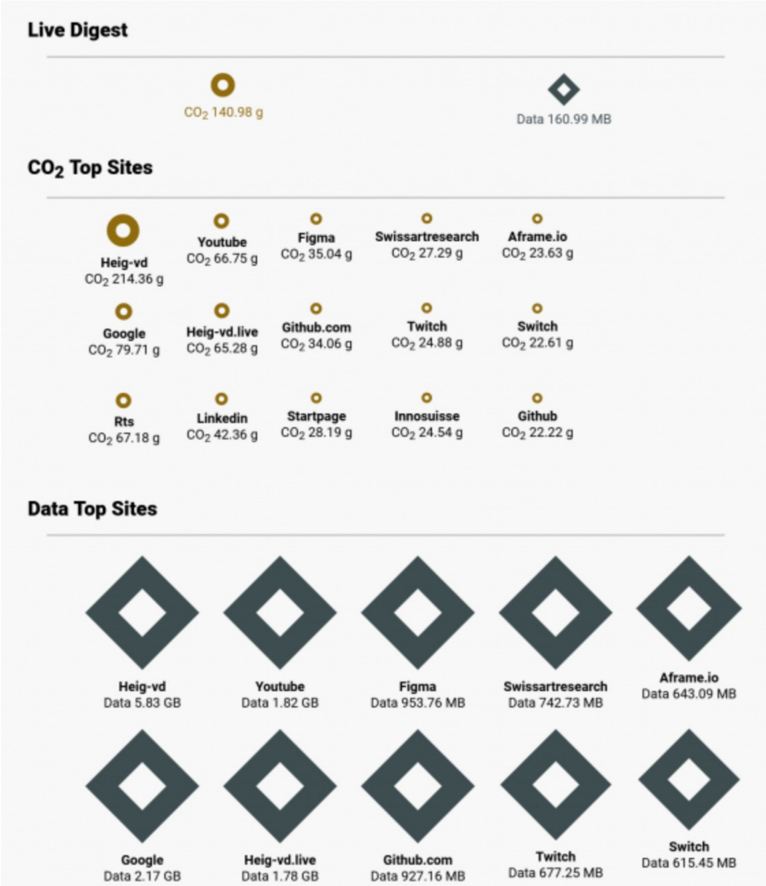
[R. Tucker, 2011]

Energy efficiency of optics

- Power efficient?
  - Photon energy at optical frequencies : ~0.1 attojoule (1 e-19 Joules)
    - Lower-bound: 0.1 aJ/bit
- However,…
  - Optical noise (spontaneous photon emissions) → 10 photons/bit
  - Electrical noise at (direct) receiver → ~10,000 photons/bit
    → 1 fJ/bit **at the receiver!**
  - 10dB of optical losses along the way → 10 fJ/bit **at the laser output**
  - 1% efficient light (laser) source? → 1 pJ/bit **at the laser power plug**
  - 10% efficient light (laser) source? → 0.1 pJ/bit
- **Not** taken into account here:
  - Drivers, **SERDES**, WDM mux/demux

→ Sub pJ/bit optical links unlikely to emerge in the next few years

Efforts in sustainable IT



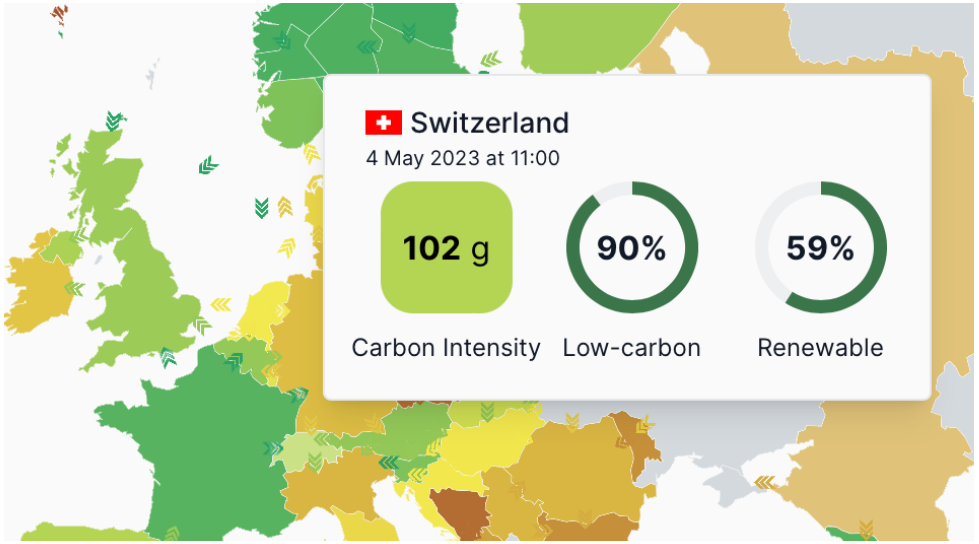
36 gCO2 / GB

Energy per internet GB ?

- 36 gCO2/GB / 102 gCO2/kWh

→ 0.35 kWh / GB = 126 uJ/bit

- Realistic ?

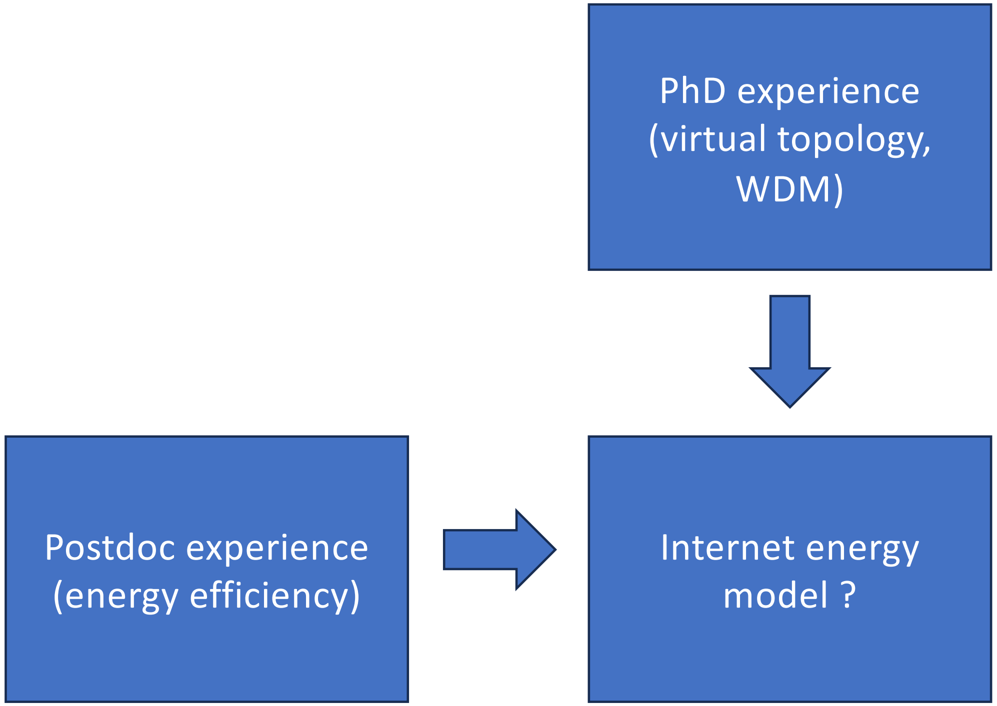


https://app.electricitymaps.com/map

The *energy intensity* of the Internet, expressed as energy consumed to transmit a given volume of data, is one of the most controversial issues. Existing studies of the Internet energy intensity give results ranging from 136 kWh/GB [11] down to 0.0064 kWh/GB [12], a factor of more than 20,000. Whether and to

Coroama, V.C., Schien, D., Preist, C., Hilty, L.M. (2015). The Energy Intensity of the Internet: Home and Access Networks. In: Hilty, L., Aebischer, B. (eds) ICT Innovations for Sustainability. Advances in Intelligent Systems and Computing, vol 310. Springer, Cham. https://doi.org/10.1007/978-3-319-09228-7_8

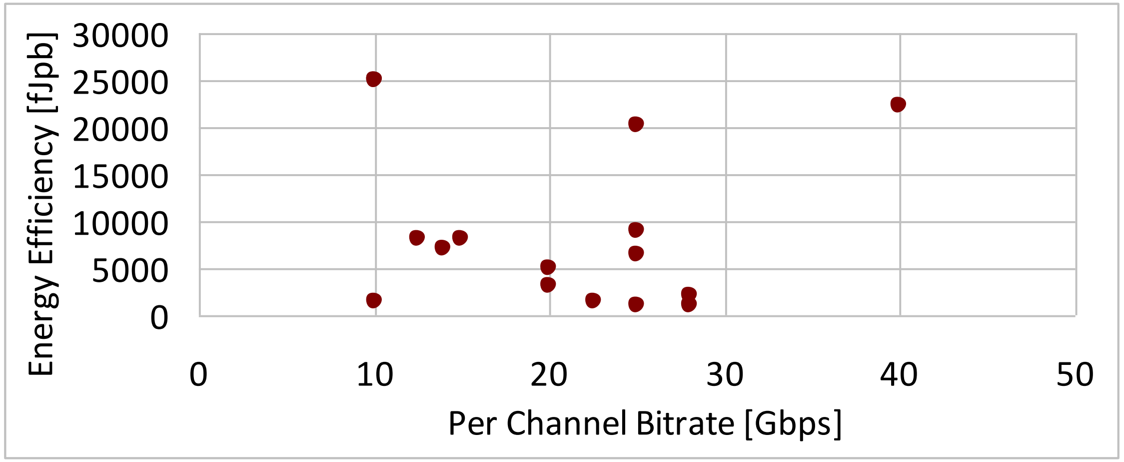
Can we approximate a lower bound?



For the rest of the presentation, O(1pJ/bit) must be understood as **no less than 1pJ/bit**

R(1pJ/bit) must be understood as **realistically around 1 pJ/bit**

Transmission



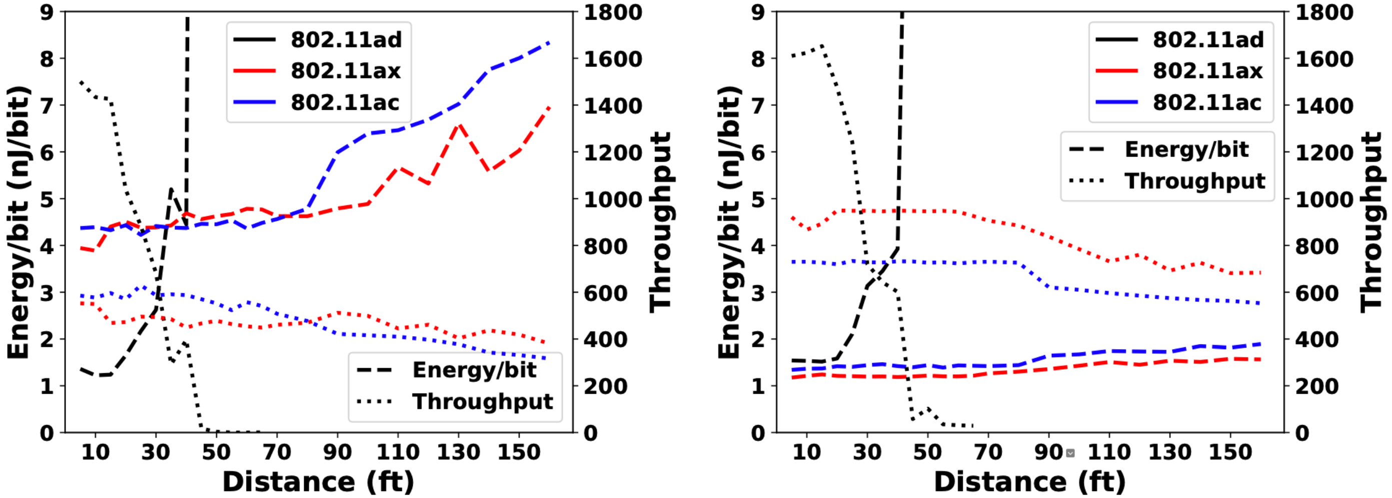
- Short distance optics
  - No evident correlation between bitrate and efficiency
  - We can assume short distance optical links to consume O(1pJ/bit)

- Long distance optics
  - More stringent requirements on signal quality (x30)
  - Presence of amplifiers along the way (x10-x100)

→ Max [O(1nJ/bit/1000km) , O(30pJ/bit)]
→ R(300 pJ/bit)

Access transmission (wireless, wifi)

- Wifi
  - O(3nJ/bit)
  - R(10nJ/bit)
- 4G/5G
  - 1kW base station
  - 1 Gbps
  - → R(1uJ/bit)



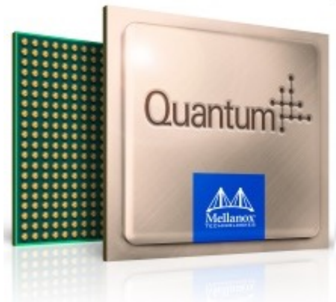
(a) Uplink. (b) Downlink.

Fig. 3: Energy-throughput-range tradeoff.

From : Aggarwal et al. 802.11ad in Smartphones: Energy Efficiency, Spatial Reuse, and Impact on Applications, 2021

Switching

- Infiniband / HPC switches



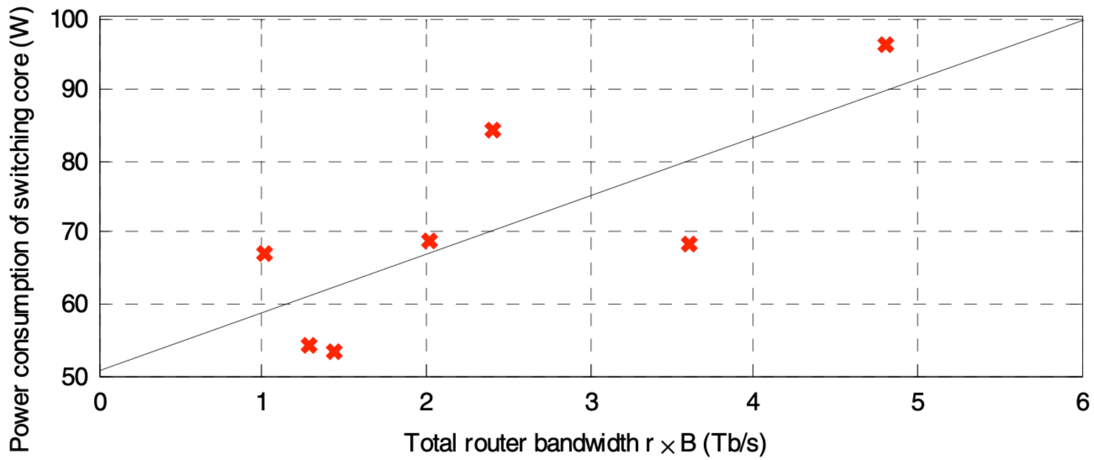


TABLE II. PROPERTIES OF COMMERCIAL ROUTERS (CHIP POWER = 70% OF TOTAL POWER)

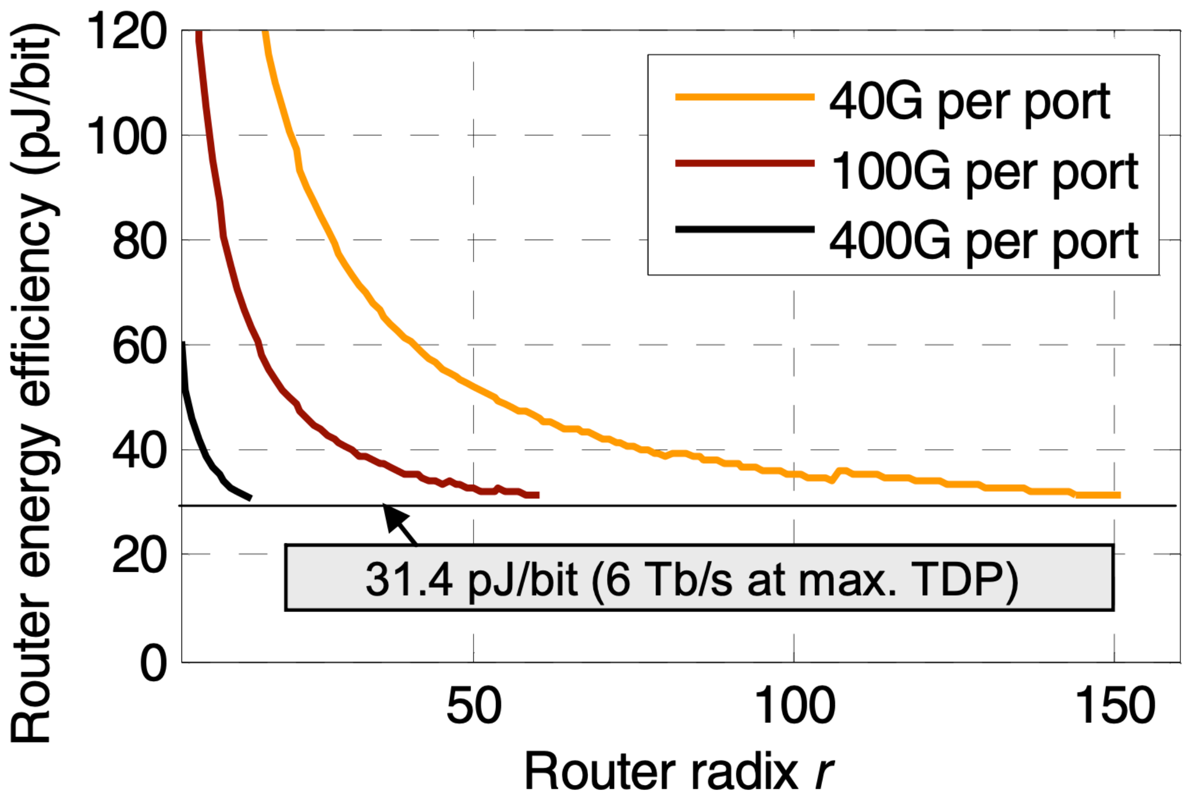
Prod.	Model	Ports	Line-rate (Gb/s)	Total BW (Tb/s)	Power (W)	Chip power (W)
Mellanox	M3601Q	32	40	1.28	85	59.5
Mellanox	SX6015	18	56	1.008	103	72.1
Mellanox	SX6025	36	56	2.016	113	79.1
Mellanox	SB7700	36	100	3.6	136	95.2
Intel	12200	36	40	1.44	85	59.5
Intel	Omni-P.	24	100	2.4	146	102.2
Intel	Omni-P.	48	100	4.8	189	132.3

TABLE III. POWER ESTIMATIONS FOR COMMERCIAL ROUTERS

Model	Chip power (W)	Chip eff. (pJ/bit)	IOBs. eff. (pJ/bit)	Core eff. (pJ/bit)	Core power (W)
M3601Q	59.5	46.48	4.25	42.23	54.06
SX6015	72.1	71.53	5.11	66.42	66.95
SX6025	79.1	39.24	5.11	34.13	68.80
SB7700	95.2	26.44	7.48	18.97	68.29
12200	59.5	41.32	4.25	37.07	53.38
Omni-P.	102.2	42.58	7.48	35.1	84.26
Omni-P.	132.3	27.56	7.48	20.09	96.42

S. Rumley, R. P. Polster, S. D. Hammond, A. F. Rodrigues and K. Bergman, "End-to-End Modeling and Optimization of Power Consumption in HPC Interconnects," *2016 45th International Conference on Parallel Processing Workshops (ICPPW)*, Philadelphia, PA, USA, 2016, pp. 133-140, doi: 10.1109/ICPPW.2016.33.

Routing efficiency



- MPLS networks O(30 pJ/bit)
- What about IP routers ?
  - Too many routes to fit in on-chip cache
  - More complex routing logic
  - Educated guess O(1nJ/bit)

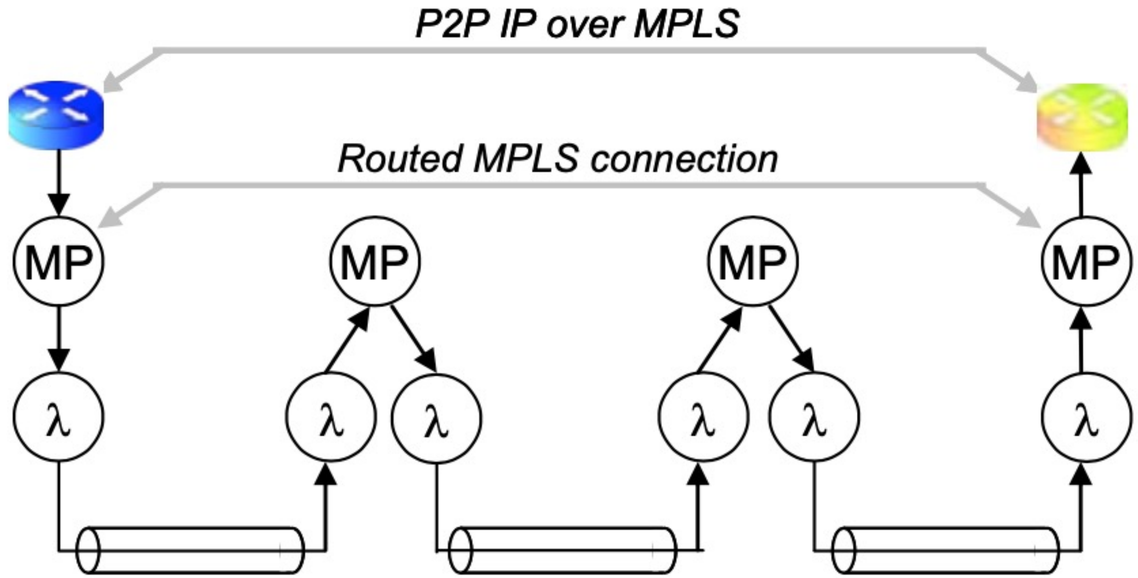
IP Topology

```
[rumley:~/$ traceroute www.google.com
traceroute to www.google.com (172.217.168.4), 64 hops max, 52 byte packets
 1  160.98.210.1 (160.98.210.1)  21.289 ms  10.337 ms  11.851 ms
 2  160.98.12.29 (160.98.12.29)  11.880 ms  10.208 ms  10.709 ms
 3  int.ngf-peri.hefr.ch (160.98.12.44)  11.836 ms  10.306 ms  10.649 ms
 4  gi-0-2.enro1.per80.hefr.ch (160.98.6.81)  12.979 ms  12.711 ms  11.150 ms
 5  160.98.7.221 (160.98.7.221)  13.087 ms  12.159 ms  11.952 ms
 6  swifr1-10ge-0-0-0-22.switch.ch (195.176.1.20)  14.114 ms  12.243 ms  12.986 ms
 7  swifr2-100ge-0-0-1-5.switch.ch (130.59.36.246)  12.692 ms  13.402 ms  11.743 ms
 8  swibe1-100ge-0-0-1-5.switch.ch (130.59.38.201)  15.066 ms  11.367 ms  11.759 ms
 9  swibf1-100ge-0-0-0-1.switch.ch (130.59.39.78)  12.009 ms  11.326 ms  12.376 ms
10  swiez3-b5.switch.ch (130.59.37.6)  15.459 ms  15.567 ms  14.685 ms
11  swiez2-b1.switch.ch (130.59.36.125)  17.387 ms  15.728 ms  17.285 ms
12  swiix3-100ge-0-0-0-1.switch.ch (130.59.36.178)  16.262 ms  14.658 ms  15.958 ms
13  72.14.195.4 (72.14.195.4)  15.224 ms  15.533 ms  15.016 ms
14  * * *
15  64.233.175.166 (64.233.175.166)  22.943 ms
    142.251.70.184 (142.251.70.184)  17.329 ms
    74.125.243.129 (74.125.243.129)  20.714 ms
16  zrh11s03-in-f4.1e100.net (172.217.168.4)  15.085 ms  15.834 ms
    74.125.243.135 (74.125.243.135)  18.167 ms
```

- Assume O(10 $hops_{IP}$)
- R(30 $hops_{IP}$) (not 10, not 100)

MPLS topology (between IP routers)

- At least one MPLS hop per IP hop $O(1hop_{MPLS}/hop_{IP})$



- Realistically, let's do an educated guess of $R(3hop_{MPLS}/hop_{IP})$

Model

	No less than	Realistically
Access	3 nJ/bit	10 – 1000 nJ/bit
Long distance trans.	30 pJ/bit	300 pJ/bit
MPLS switching	30 pJ/bit/hop_{MPLS}	100 pJ/bit/hop_{MPLS}
IP Switching	1 nJ/bit/hop_{IP}	
MPLS topology	1 hop_{MPLS}/hop_{IP}	3 hop_{MPLS}/hop_{IP}
IP topology	10 $hops_{IP}$	30 hop_{IP}

$$Energy_{bit} = Access + hops_{IP}\left(Switching_{IP} + hops_{MPLS/IP}(Switching_{MPLS} + transmission)\right)$$

- O(3 + 10 (1 + 1(0.03 + 0.03))) = O(13.6 nJ/bit) = O(0.035 Wh/GB)
- R(10 + 30 (1 + 3(0.1 + 0.3))) = R(76 nJ/bit) = R(0.2 Wh/GB)

The *energy intensity* of the Internet, expressed as energy consumed to transmit a given volume of data, is one of the most controversial issues. Existing studies of the Internet energy intensity give results ranging from 136 kWh/GB [11] down to 0.0064 kWh/GB [12], a factor of more than 20,000. Whether and to

Something is missing....

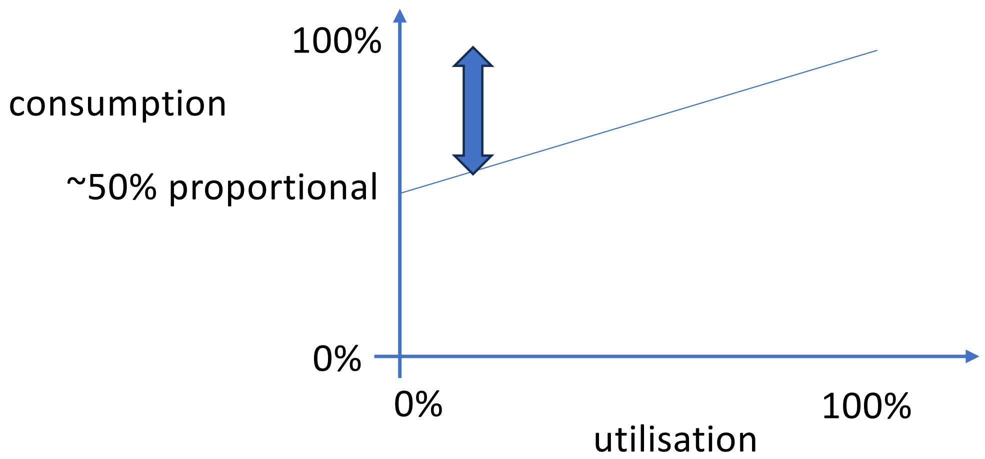
- Utilisation
- Energy proportionality

- In any network industry, you pay (money, energy) for two things
  - To use the network (once)
  - To have the network ready to be used (anytime)

- Networks are generally over-dimensioned
  - Probably fair to assume **at most 10% utilisation**

Energy proportionality

- Does 10% utilisation means 10% consumption ?
  - Generally not... We assume the proportionality to be **50% at most**



- This means we have to multiply the consumption by 10 * (55%) = 5.5

Our educated guess

O(74.8 nJ/bit) → O(0.2 Wh/GB)

Lower bound (can't be below) as of today

R(418 nJ/bit) → R(1.1 Wh/GB)

Is realistically above this mark

Validation

SWITCH

- Switch numbers:
  - Transmission : 250MWh / 345’638TB = 260 nJ/bit = 0.72 Wh/GB
  - Switching : 453 MWh / 345’638TB = 472 nJ/bit = 1.31 Wh/GB

732 nJ/bit → 2 Wh/GB

Switch numbers doesn’t include access and only partial number of hops

And then there is the grey energy…

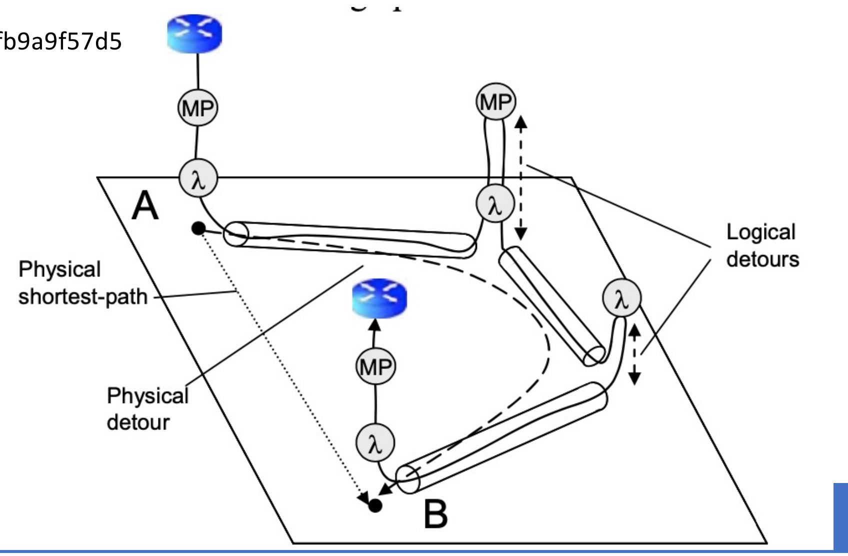
- https://db.resilio.tech
- Grey energy of my MacBook Pro (~1000kWh) represents 3.2 years of power consumption at 10% activity.
- Probably fair to double at least the figure
- What about grey energy of duct, cabinets, antennas…?

Where to improve tomorrow?

	No less than	Realistically
Access	3 nJ/bit	10 – 1000 nJ/bit
Long distance trans.	30 pJ/bit	300 pJ/bit
MPLS switching	30 pJ/bit/hop_{MPLS}	100 pJ/bit/hop_{MPLS}
IP Switching	1 nJ/bit/hop_{IP}	
MPLS topology	1 hop_{MPLS}/hop_{IP}	3 hop_{MPLS}/hop_{IP}
IP topology	10 $hops_{IP}$	30 hop_{IP}

S. Rumley, C. Gaumier, “Energy Saving Perspectives of IP over Wavelength Division Multiplexing”, in Proceedings of the COST Action IC804, 2nd year, https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=1b811f0d0de148b1151792a2a5fc4afb9a9f57d5

			Consumption scaling with	Domain
Links	Amplification		# of WDM systems Distance	Optical domain
Nodes	Optical processing (mux/demux, OXC)		# of systems	Optical domain
Nodes	Transmission (lasers)		# of lightpaths	Optical domain
Nodes	Processing		# of nodes, node bandwidth	Electrical domain
Nodes	Switching		# of nodes, node bandwidth	Electrical domain
Nodes	Switching	Lookup, classifying	Adress space complexity	Electrical domain
Nodes	Switching	Buffering	Latency	Electrical domain



Final calculation

- Assume world internet traffic of 3 Zettabytes/year
- At 1microJ/bit, this amounts to 8.3 TWh
- This is 16% of Switzerland's electricity consumption
- This is 0.05% of the world's electricity – not a big deal
- But if we multiply traffic by 100, then it is 5%...
  - With a CAGR of 24%, this is in 2043

Conclusion

- Take aways
  - Proposal to develop a model to estimate "physical" energy per bit
  - Model version 0.001 shows energy dominated by
    - IP routing
    - Access
  - Model shows actual consumption of internet not too concerning
    - But (as usual) further growth might hurt
- Research questions?
  - How to improve network energy efficiency?
  - How far can network energy efficiency be improved?
  - How to measure network energy efficiency?
  - How to foster network energy efficiency improvement?

Thank you