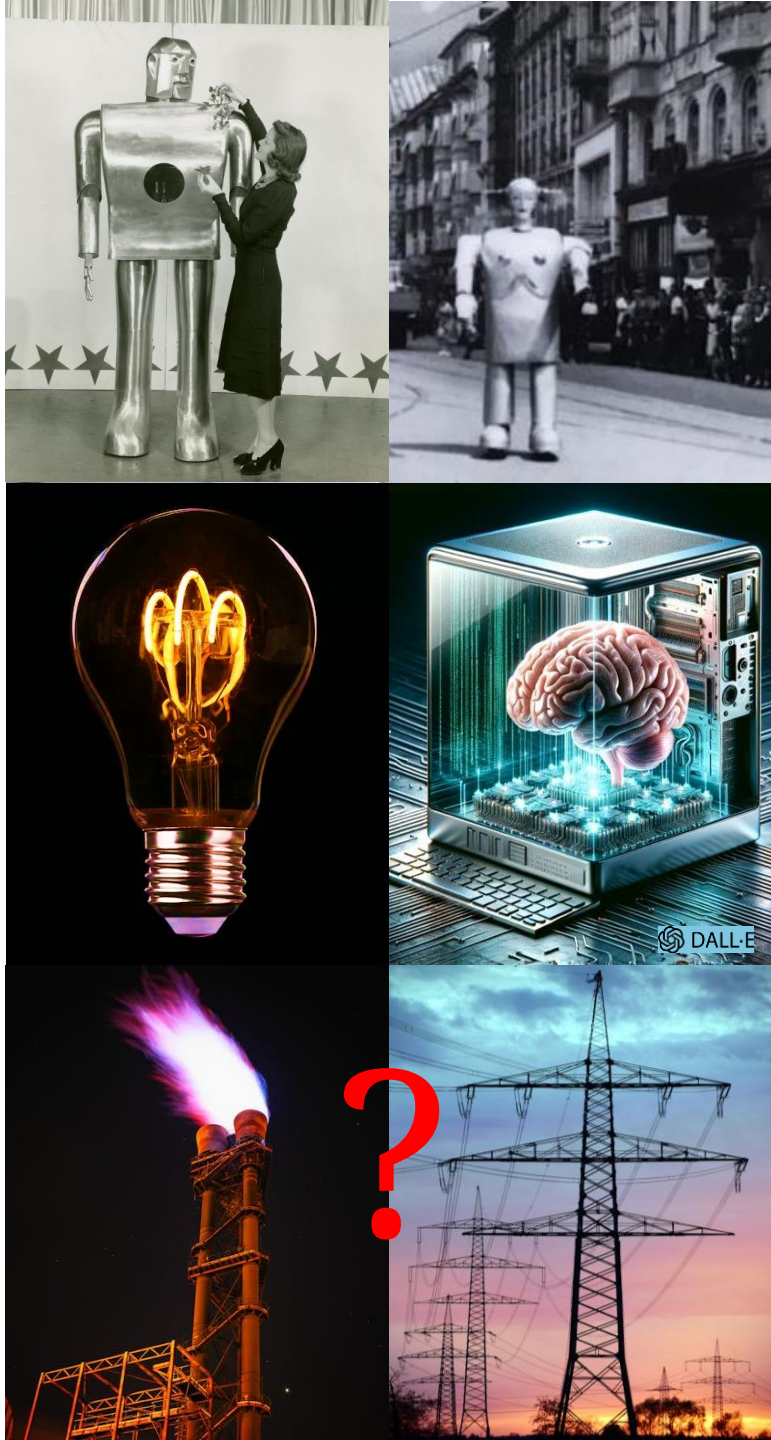
super computing systems

Light + AI 3.0 = NextGen DC

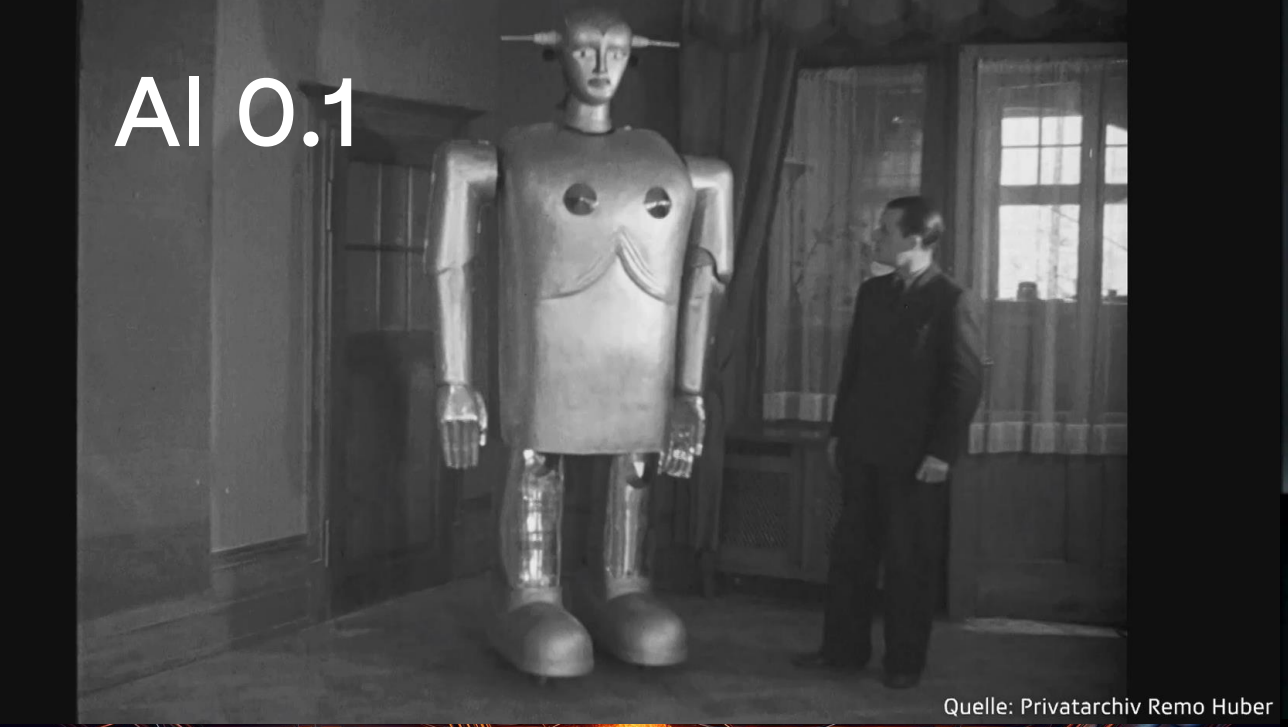


More Light and AI 3.0 Shape the Future of Data Centers

TEC Forum 2025

Christof Bühler, Supercomputing Systems AG

AI 0.1



Quelle: Privatarchiv Remo Huber

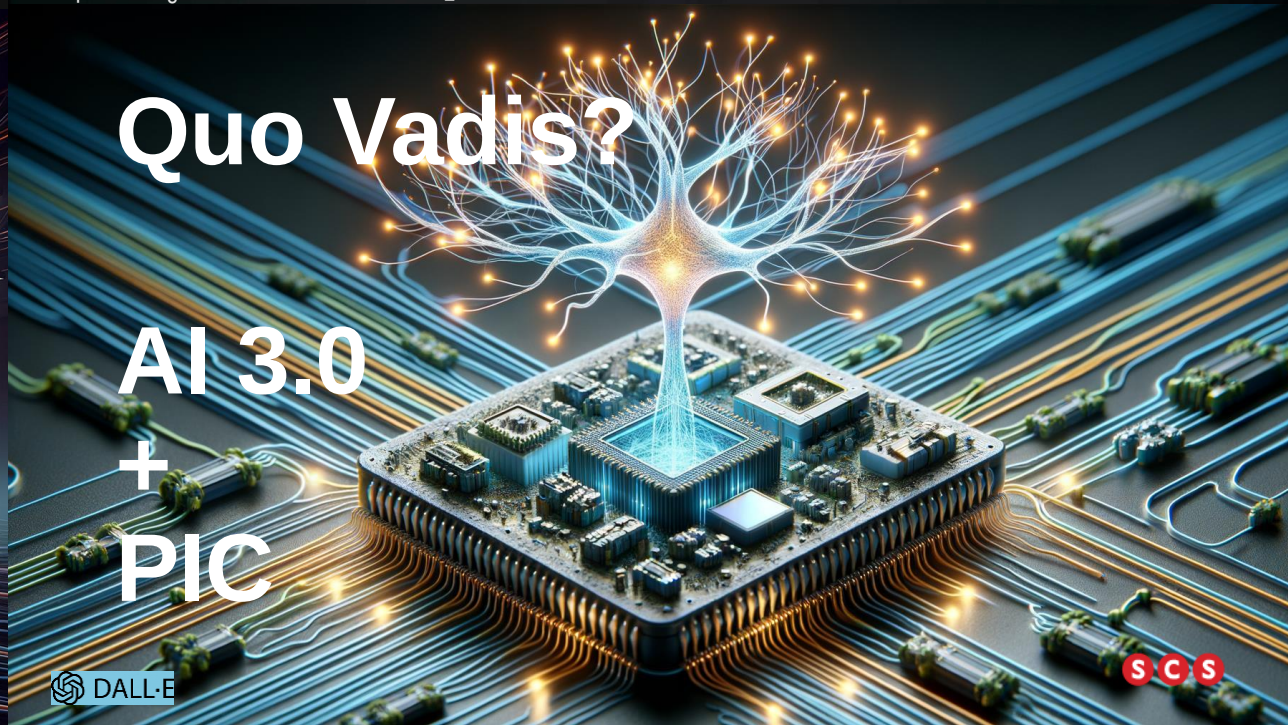
AI 2.0

https://www.youtube.com/watch?v=JzIsvFN_5HI

Dark Side of AI

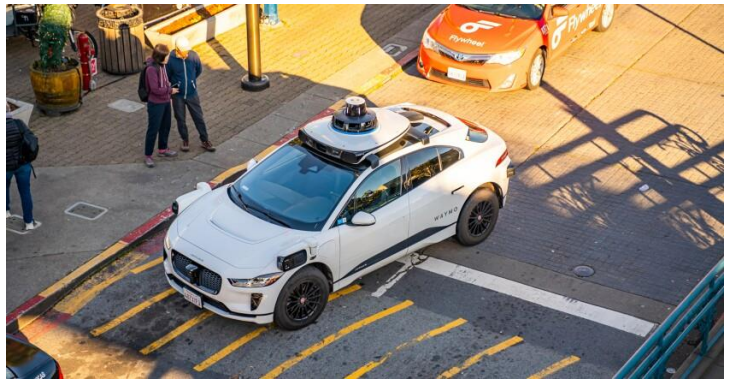
Quo Vadis?

AI 3.0
+
PIC



AI ≠ Human Brain

Machine Learning



> 7 million test kilometers

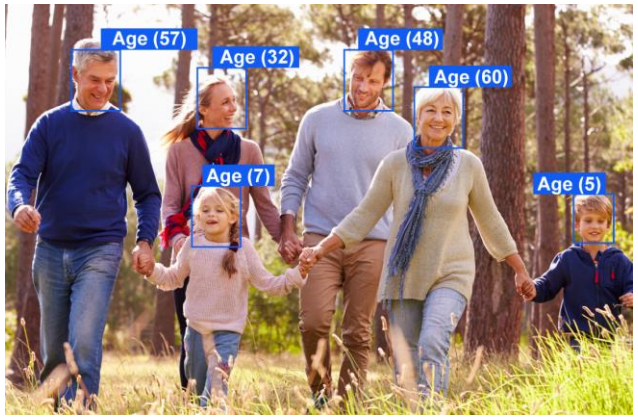
Millions of labeled photos

Many engineers

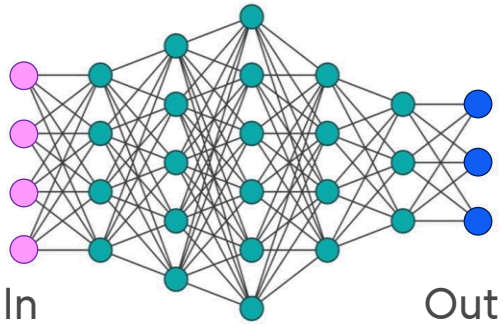
Human learning

20 - 30h driving lessons

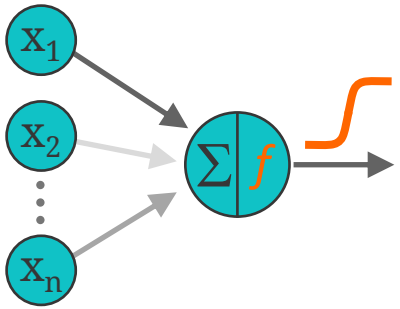
AI : Programming Becomes Training



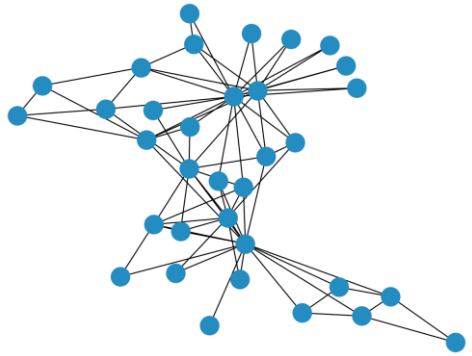
Learning Algorithm

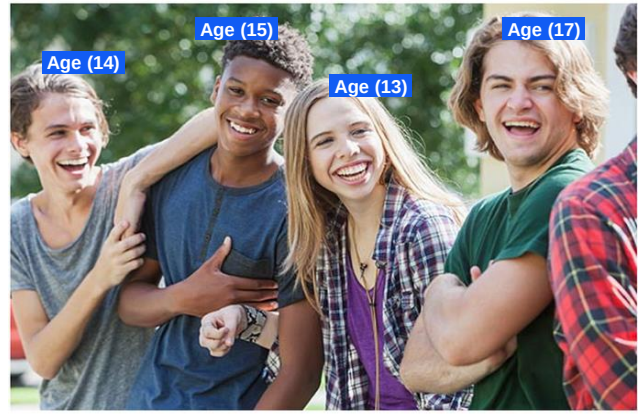


Artificial Neuron



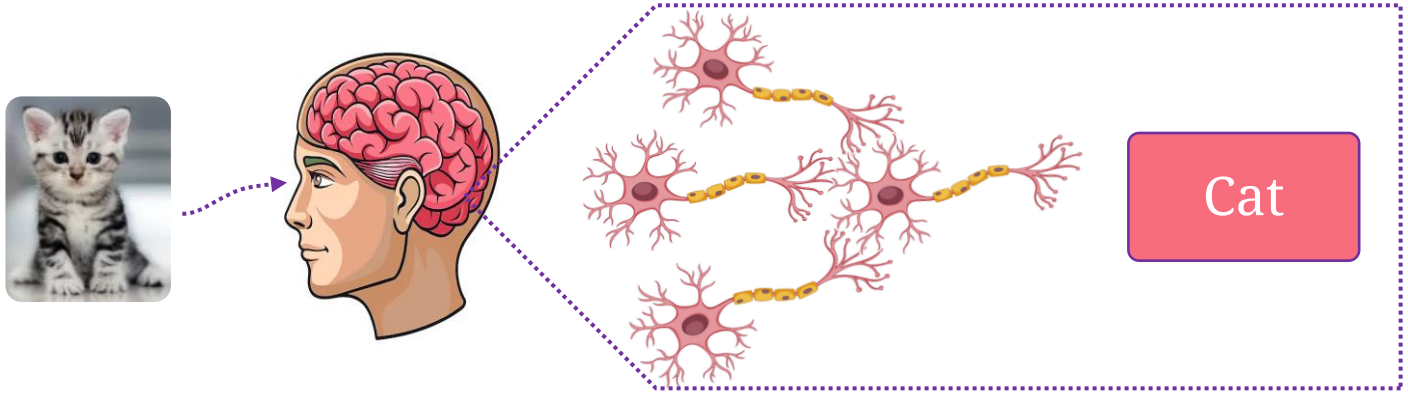
AI Model



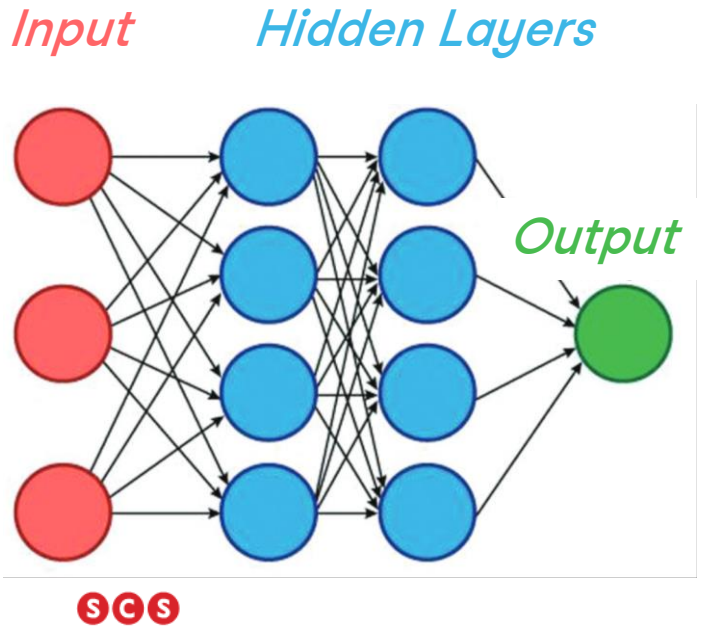


AI : Refresher in 60+ Seconds

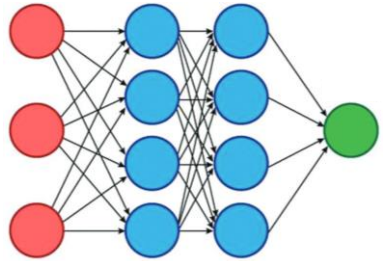
Biology



Technical



Training Z_t



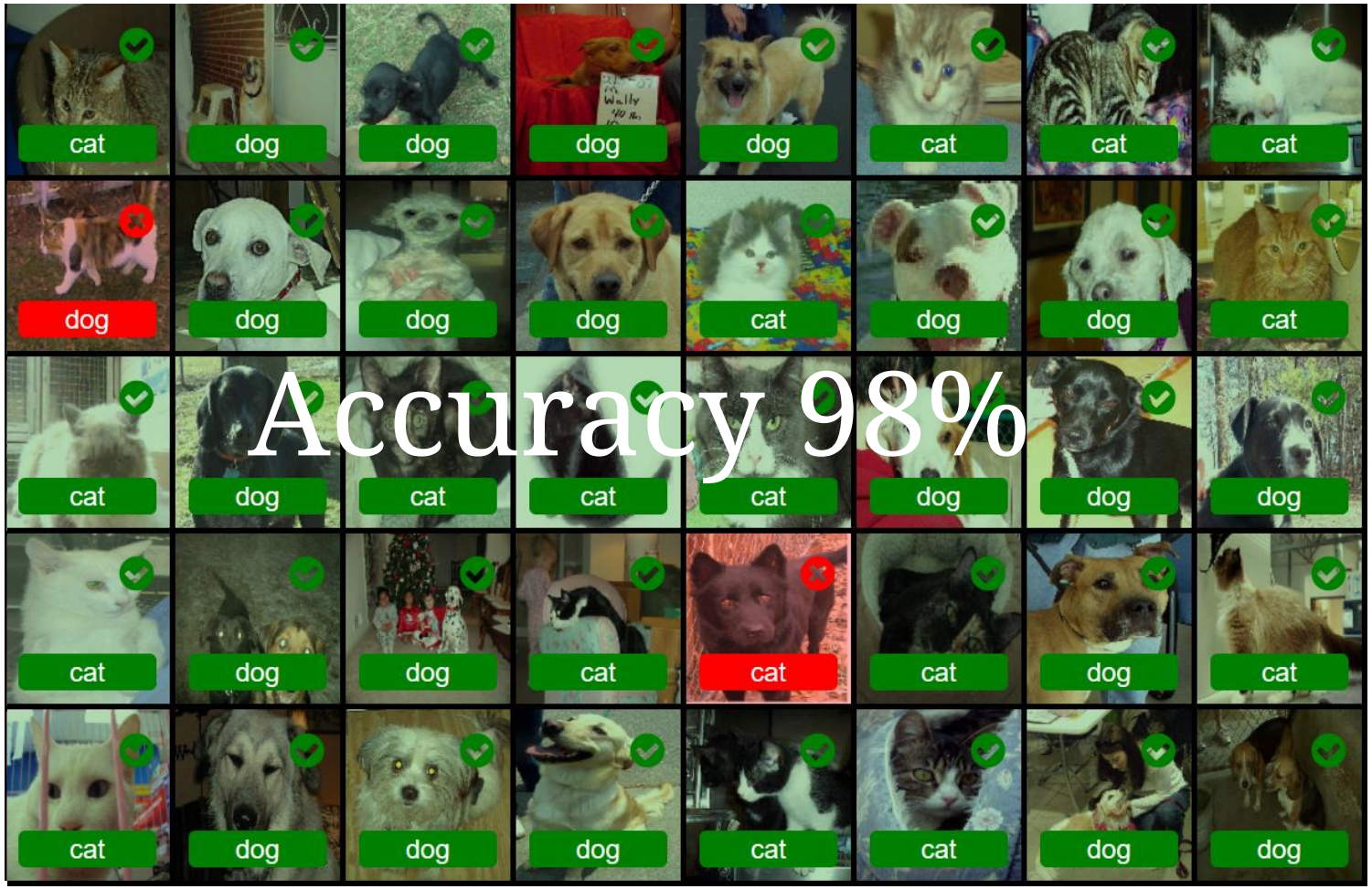
$G_w(Z_t)$

Iteration $_t$ D_t

Backprop: How much missing? $E = \|D_t - G_w(Z_t)\|^2$

Learning: adjust weights $W^{t+1} = W^t - \alpha \frac{\partial E}{\partial W}(W^t)$

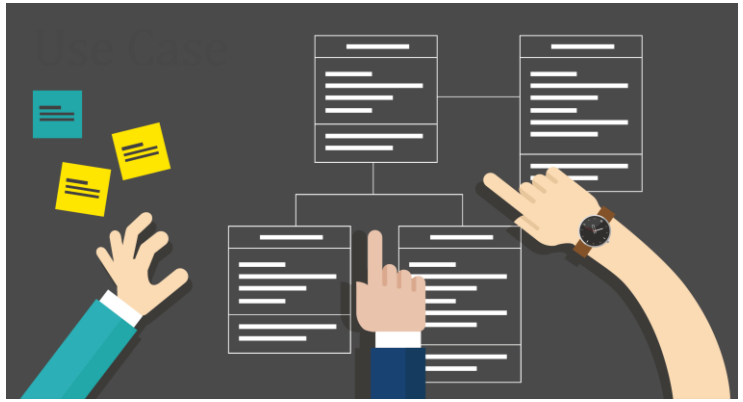
Have Fun with AI tools



http://ai-demo.scs.ch

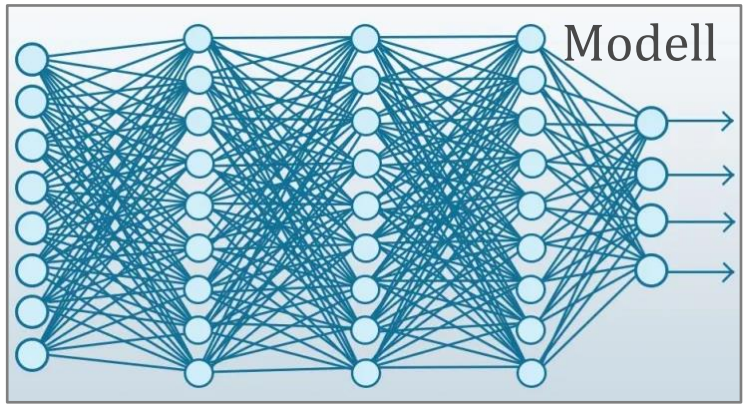
http://ai-demo.scs.ch/#/dashboard

AI / ML : Challanges



Must offer real added value

Enough & unbiased



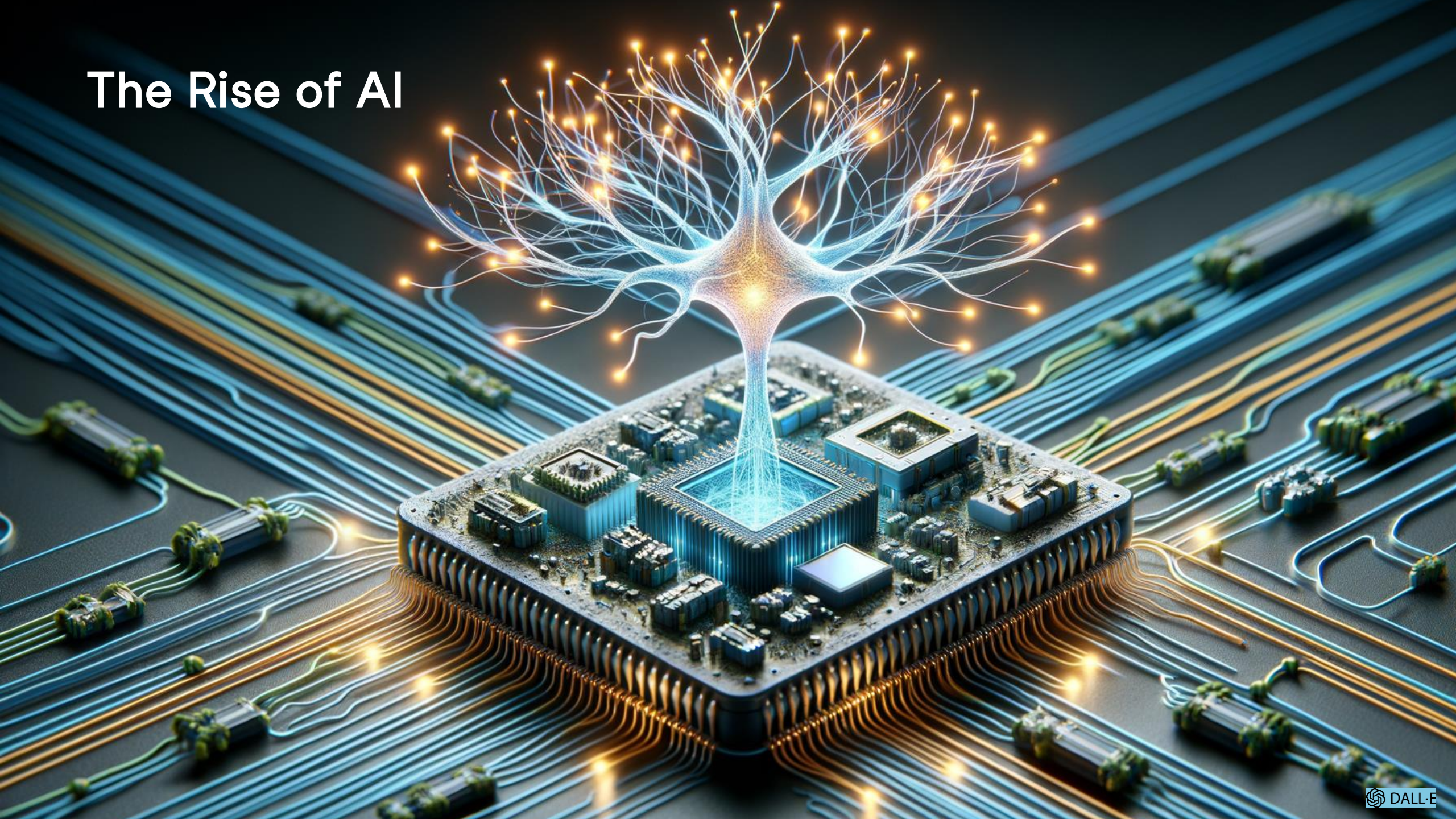
Finding the right complexity

Costs: often >> than expected

Planning ahead

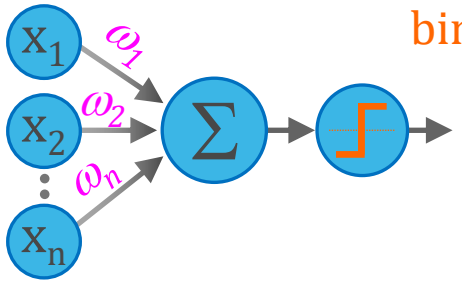
Expectations: often too high

The Rise of AI



The Rise of AI

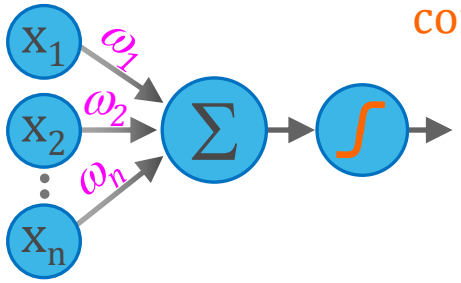
AI 1.0



binary threshold

= {0, 1}

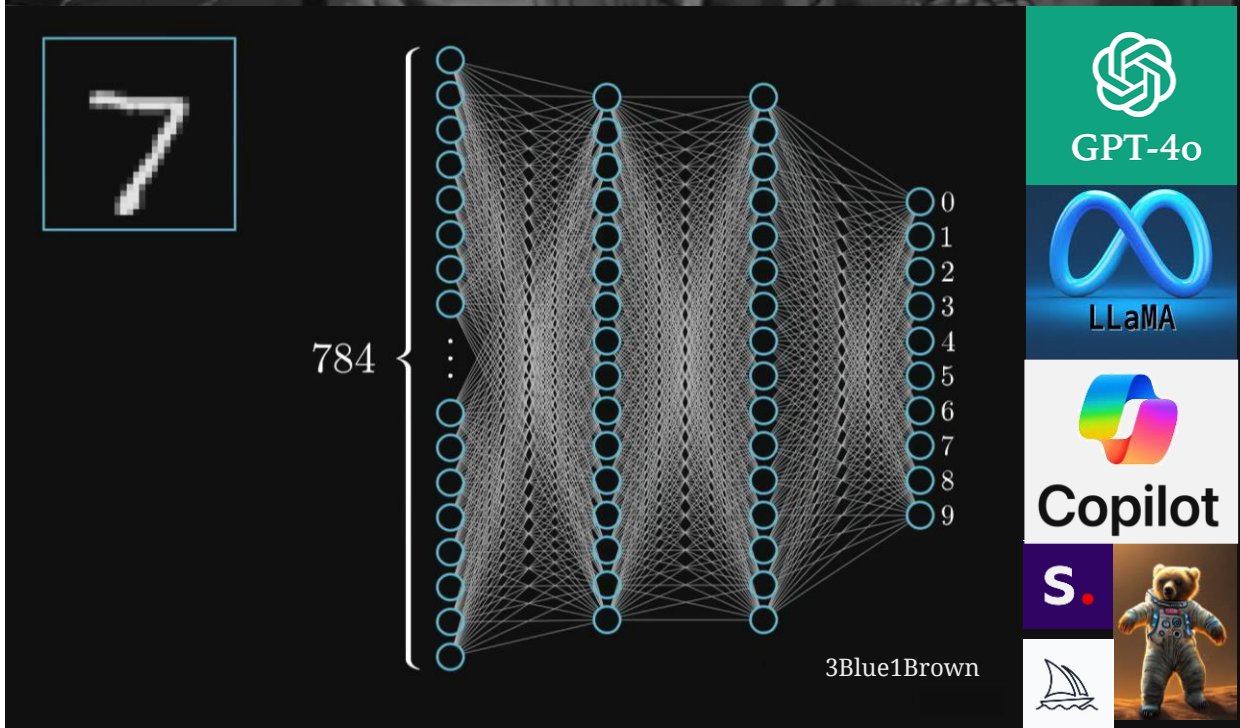
AI 2.0



cont' non-linear activation

= f(Σ)

∴ KI 2.0: x_n-in x_m-out ⇨ not intelligent yet useful



Issues and Light Awaits

Smartness

AI 1.0 << AI 2.0 << AI 3.0 << … << AGI* < Brain

AI «sucks» ⇨ energy, speed, smartness

Scaling

☞ Technical, economical and intellectual

AI 3.0: Neuromorphic Computing (NC)

- ⎍-based, sparse, async. communication
- highly parallel: 10^5 - 10^6 cores
- low-power, real time, adaptive learning

∴ KI 3.0: ⎍-in ⎍-out» … a leap towards AGI?!

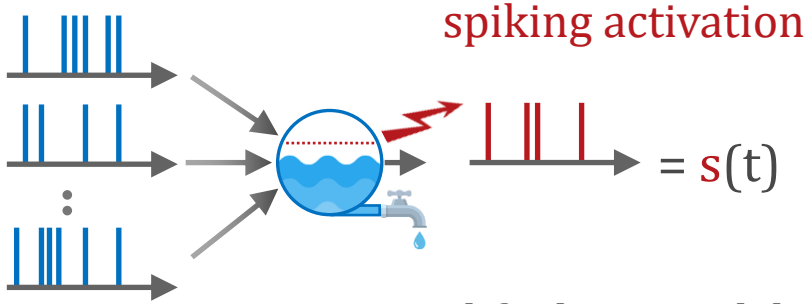
SCS

*AGI: Artificial General Intelligence

The Rise of AI: Light Awaits?

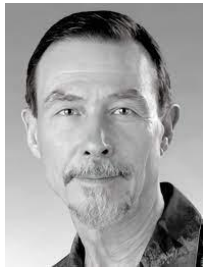
AI 3.0

neuromorphic
spiking activation



Simplified LIF model

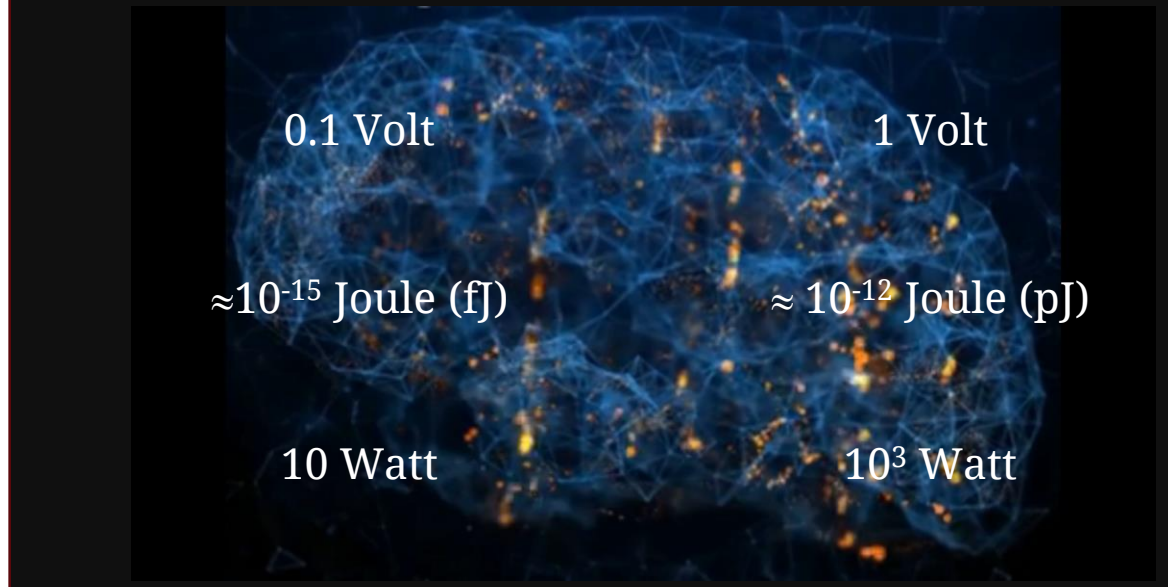
LIF: Leaky Integrate & Fire



Carver Mead (*1934 U.S.): The Father of Neuromorphic Computing (NC)

1988: NC has higher likelihood of neuronal dynamics …

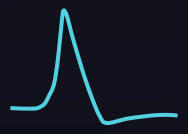
- Neuromorphic computing mimics nature: spike-based communication & sparsity



Detailed Biophys. Model

$$
\begin{aligned}
C_m \frac{dV}{dt} &= \bar{g}_{\mathrm{Na}} m^3 h (E_{\mathrm{Na}} - V) \\
&\quad + \bar{g}_{\mathrm{K}} n^4 (E_{\mathrm{K}} - V) + g_{\mathrm{L}} (E_{\mathrm{L}} - V) \\
\frac{dm}{dt} &= \alpha_m(V)(1-m) - \beta_m(V) m \\
\frac{dh}{dt} &= \alpha_h(V)(1-h) - \beta_h(V) h \\
\frac{dn}{dt} &= \alpha_n(V)(1-n) - \beta_n(V) n
\end{aligned}
$$

Hodgkin–Huxley 1952



The Rise of Hardware

From CPU to NPU

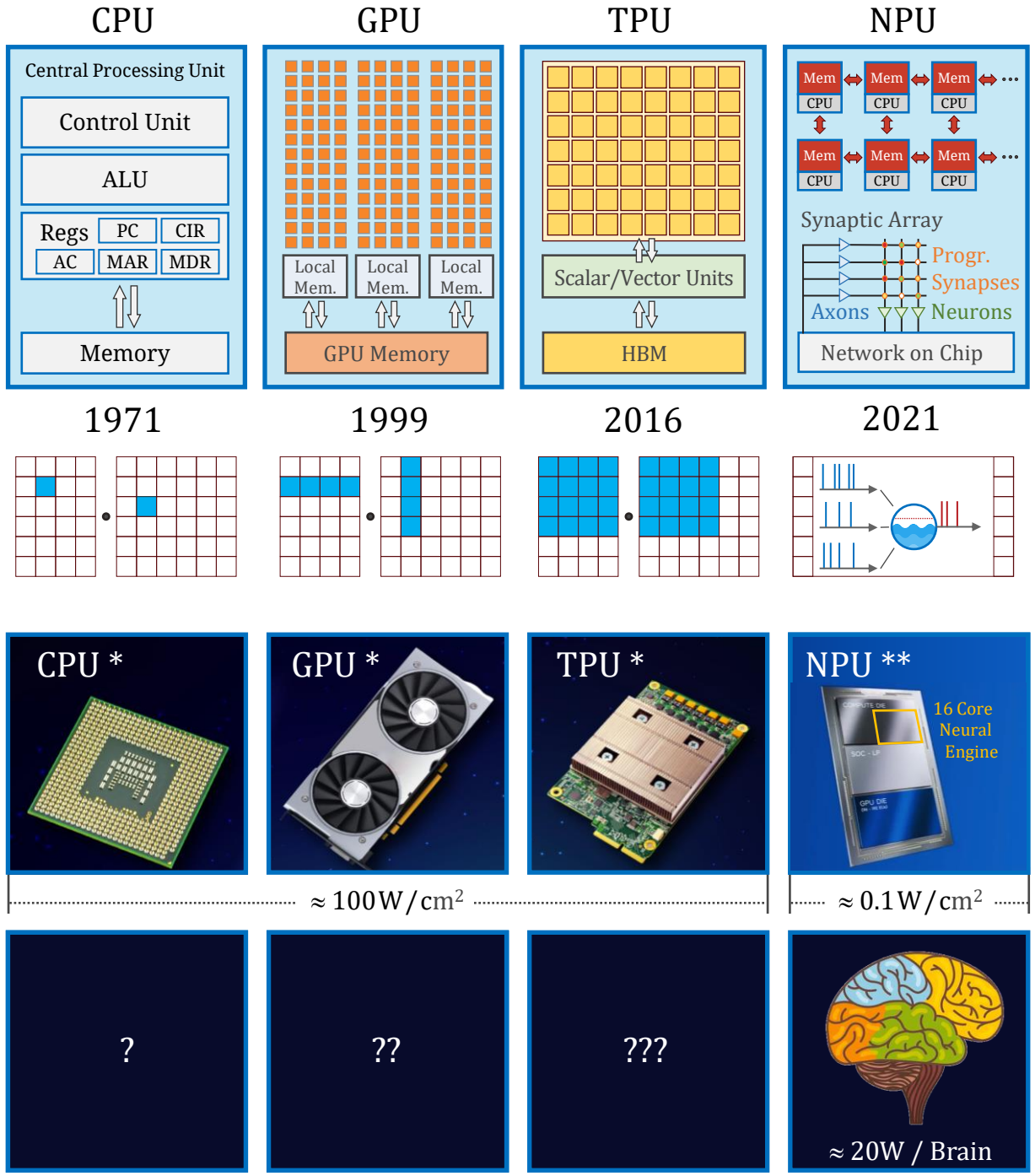
Or from numbers x_1 to spikes 

From NPU to Brain

Neuromorphic hardware

- relays, in part, on asynchronous communication
- uses spikes for power efficiency
- runs highly parallel: 10^5 - 10^6 cores
- targets low-power, real time, dynamic apps.

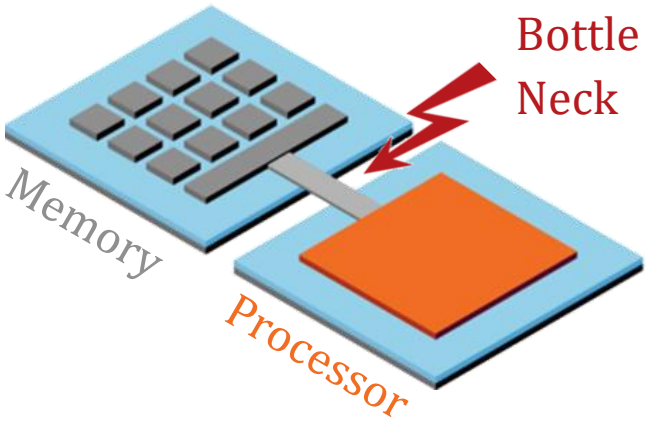
∴ The next step towards AGI is probably neuromorphic computing.



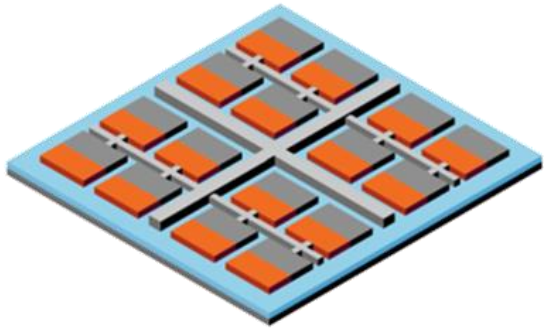
The Issue of Energy

Process Near / In Memory & Sparsity

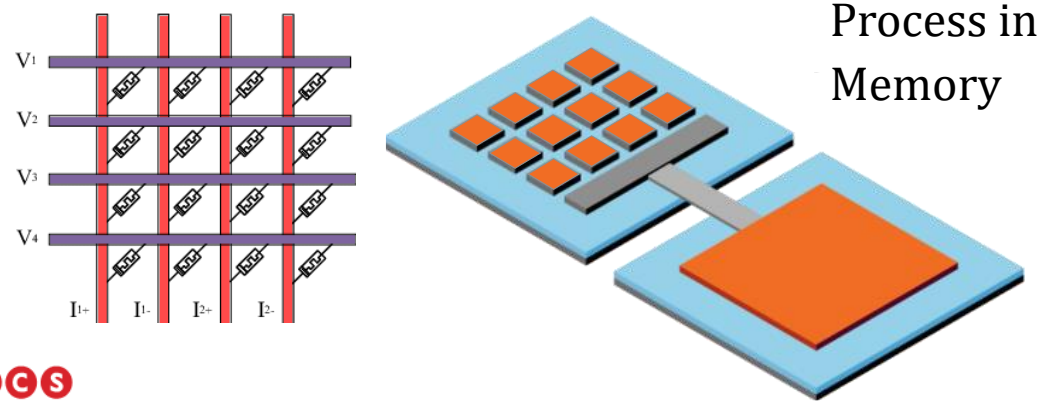
Von Neumann Architecture

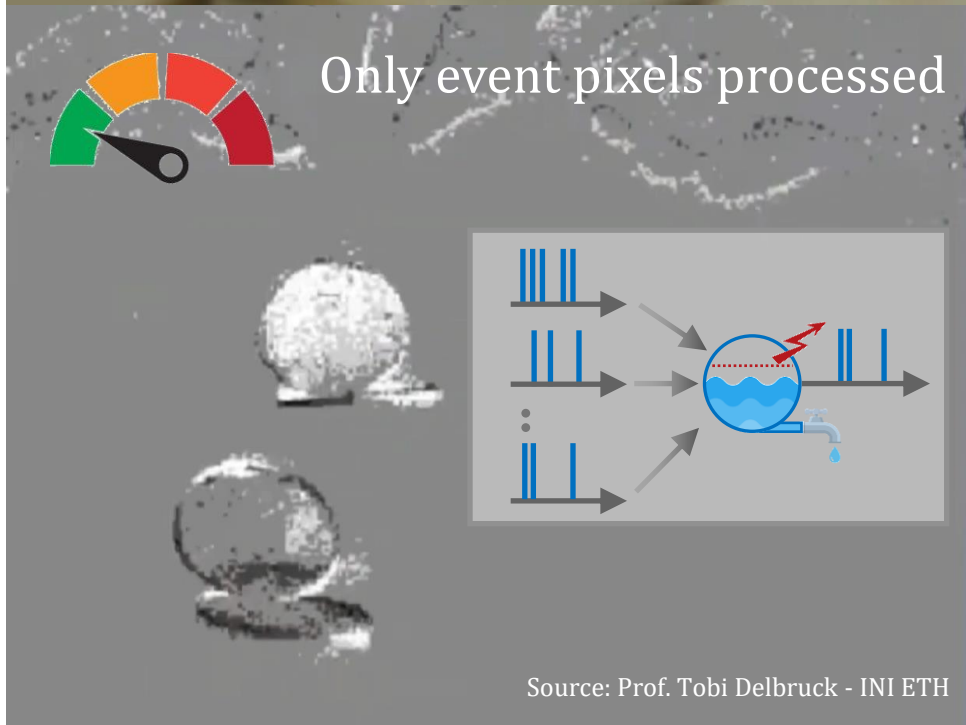


Near Memory Computing



Process in Memory





Neuromorphic Hardware

Many Approaches: Analog … Digital

Analog / Mixed

+ High bio-realism
+ Long time scales
+ Ultra-low power

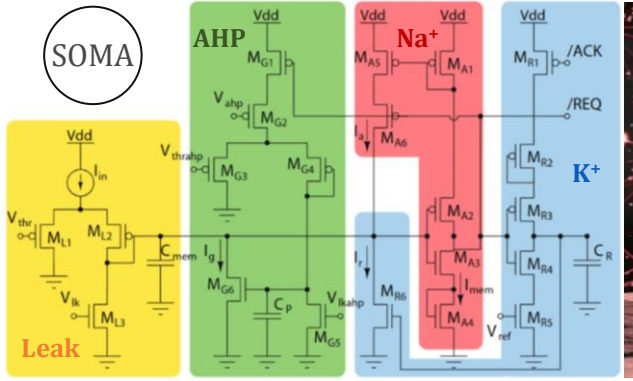
× Sensitive to noise
× Expensive (area, design)
× Few manufactures

Digital

+ High signal fidelity
+ High chip density
+ Faster design

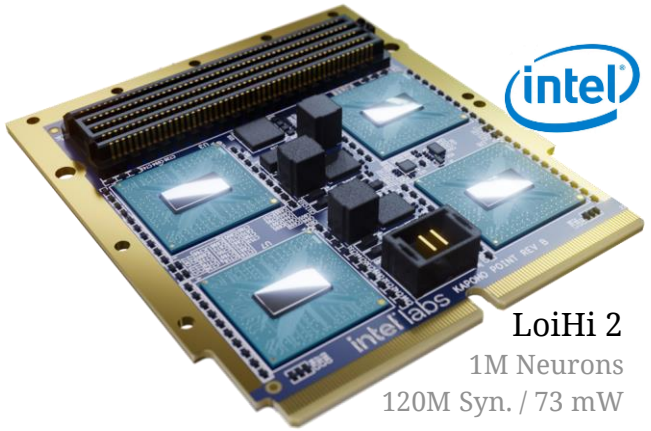
× Von Neuman Architecture
× Large on-chip memory

SNN: Power ⇘ up to 1000× / Speed ⇗ up to 10×



LIF Neuron: Fully Analog Circuit

ICNS, Western Sydney University



LoiHi 2
1M Neurons
120M Syn. / 73 mW

AIWIRE Kapoho Point Eval-Board

SynSense Speck™
Vision NSoC: DVD 128×128pix
320k Neurons / ≈5mW

brainchip Akida™

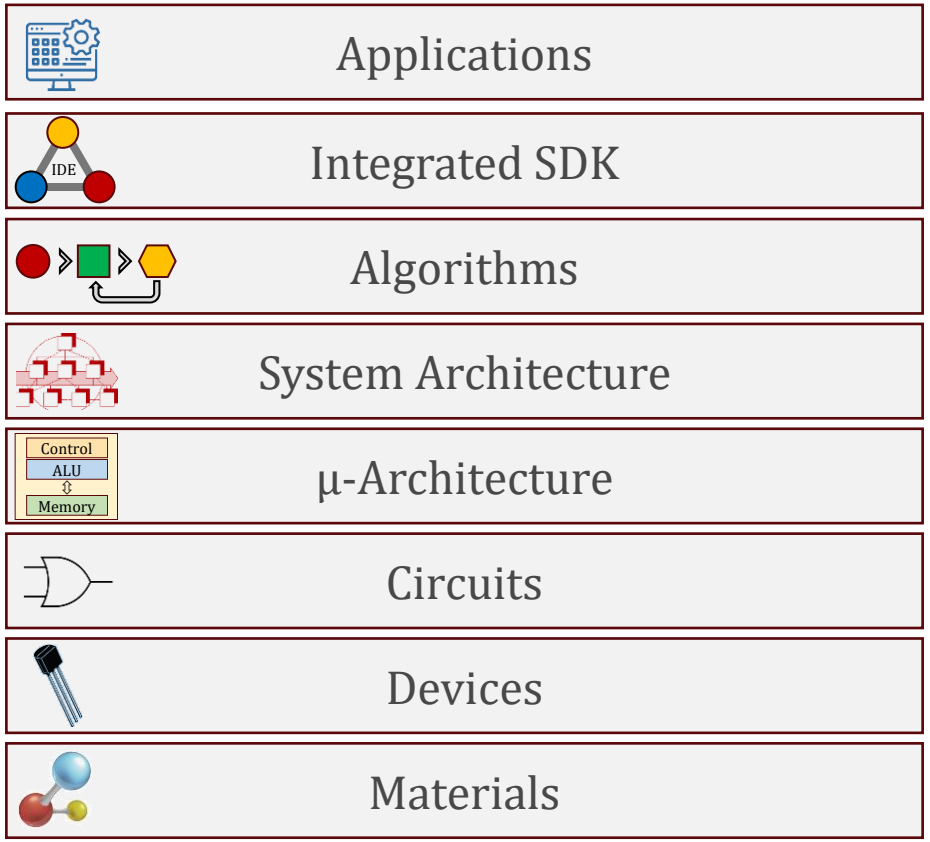
1.2M Nrns.
10b Syn.
2 mW

IBM NorthPole
1M Neurons
256M Synapses
73 mW / Chip

SEMICONDUCTOR ENGINEERING NorthPole scale-out assembly (Prototype)

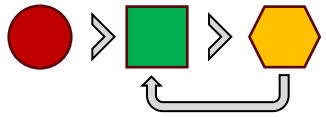
Eco System

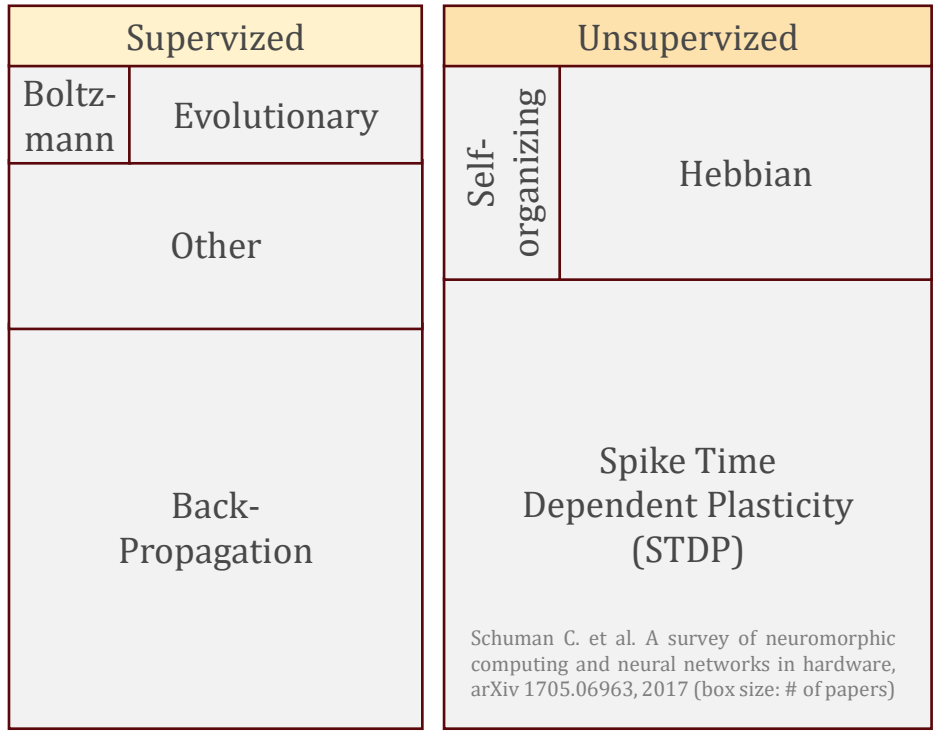
Neuromorphic Computing Stack



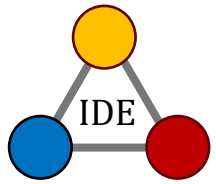
- Growing pallet of open-source SDKs

SNN Algorithms





SNN Integrated Development Tools (non-exhaustive list)



snnTorch www.snntorch.readthedocs.io/en/latest

Spinnaker www.spinnaker.io

nest:: www.nest-simulator.org

BRIAN2 www.briansimulator.org

BindsNET www.binds.cs.umass.edu/bindsnet.html

intel

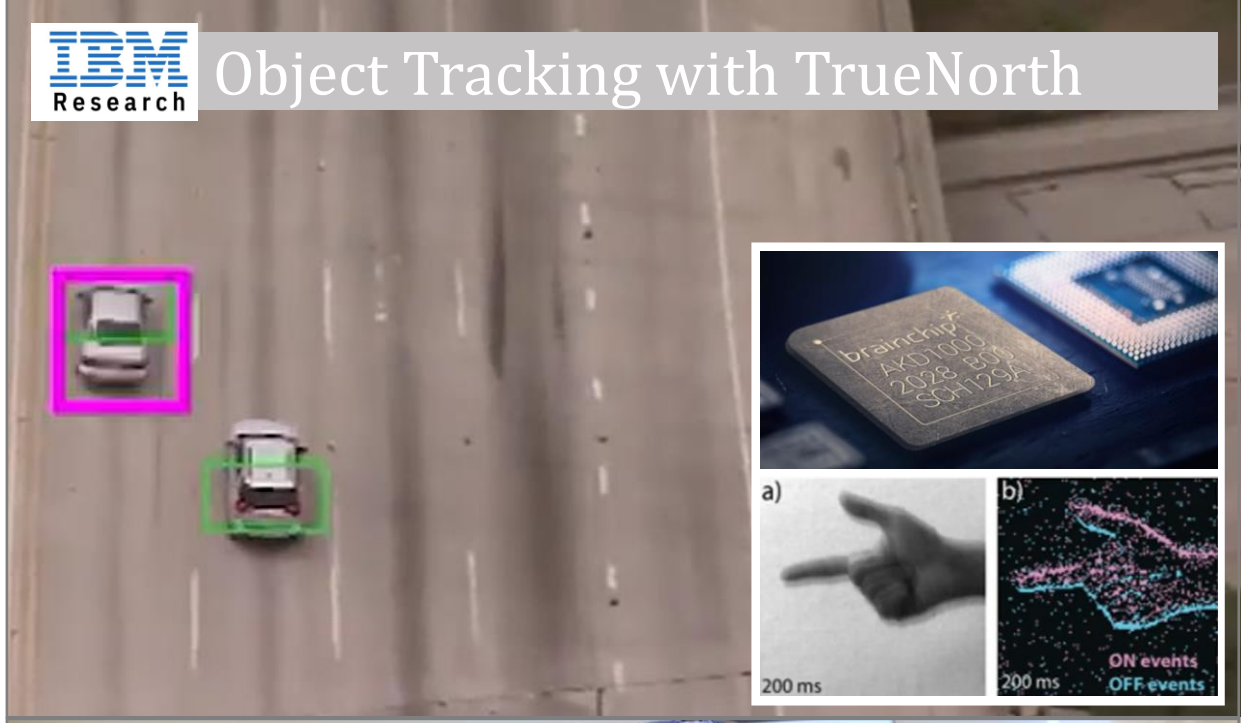
IBM COMPASS True North - IT Matters

Applications of SNNs

Broad range of suitable fields

- **AI:** brain-like processing, adaptive learning over time
- **Robotics:** real-time decisions in dynamic settings
- **IoT:** local, low-power data processing w/o cloud
- **Security / Surveillance:** real-time recognition of complex patterns and anomalies, e.g. in videos
- **Automotive**: Real-time processing of vast and dynamic sensor data for advanced driver-assistance
- **Healthcare:** efficient & adaptive processing of complex data, e.g. for personalized medicine
- ....

SNNs: Potential game-changer across many academic & industrial applications



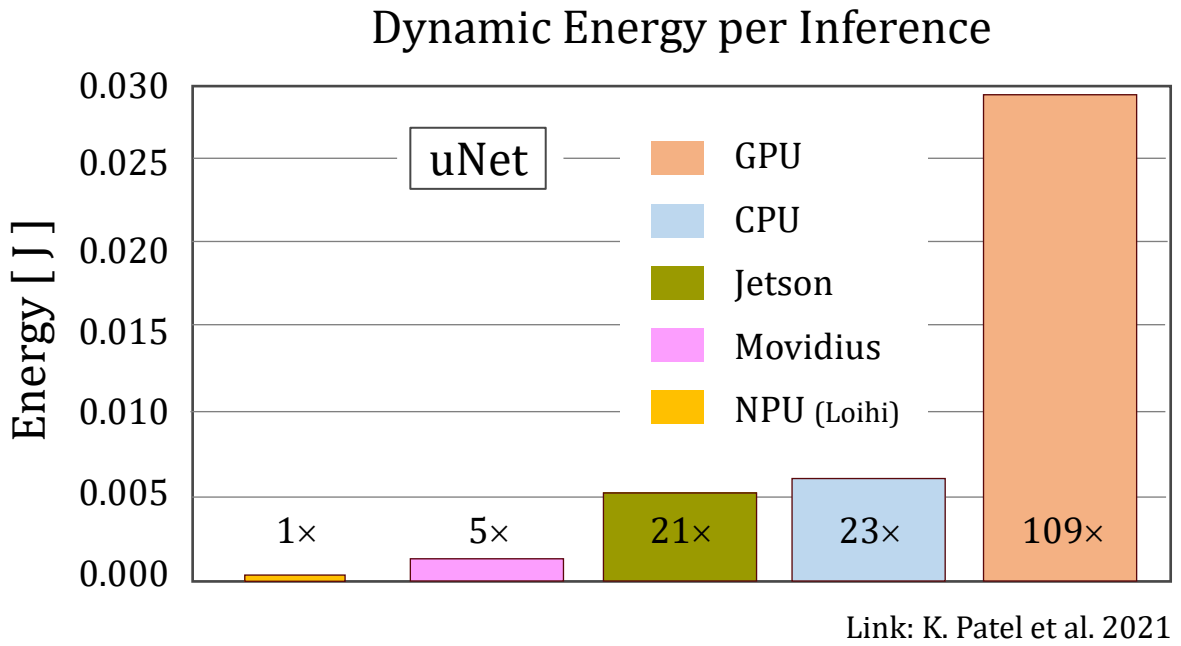
NC: Quo Vadis?

Pros and Cons

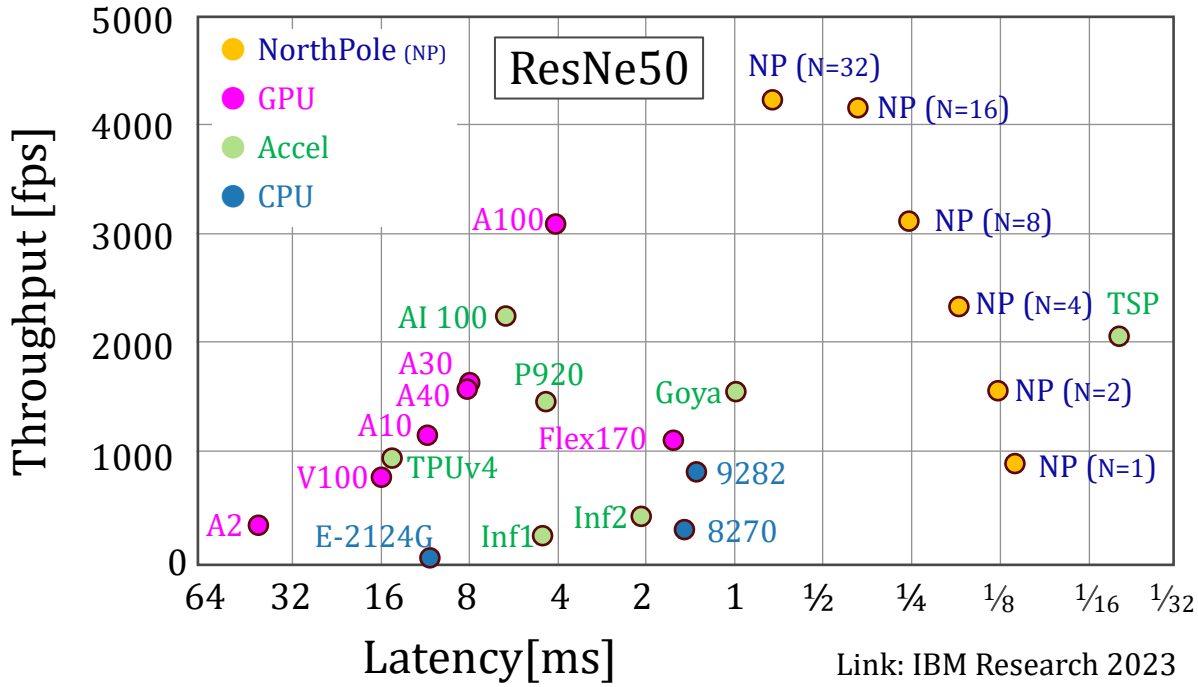
+ Energy ⇘, Speed ⇗, Scalability ✓
+ Parallel sparse / event-driven computing
+ Adaptive learning via neuron firing rules
+ Fault tolerance
+ Holds vast promise … ✗ yet faces many challenges

✗ Maturity of Technology of HW, SW & tools
✗ Compatibility issues with existing ecosystems
✗ Complex data processing w.r.t. std. ANNs
✗ Limited Vendor Options / Security concerns

SNNs: Intense research with emerging applications … ready to leave the lab!

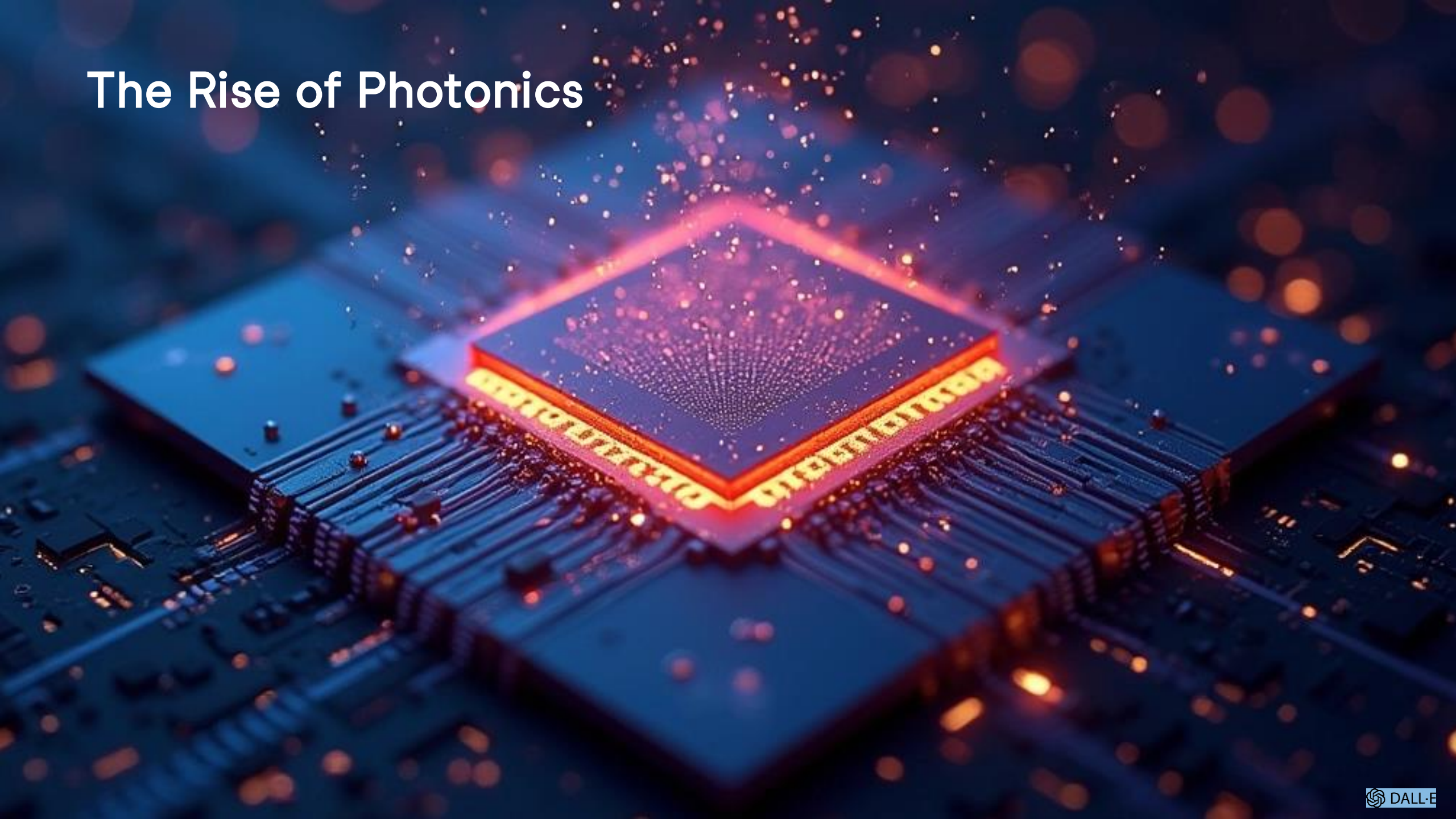


Link: K. Patel et al. 2021



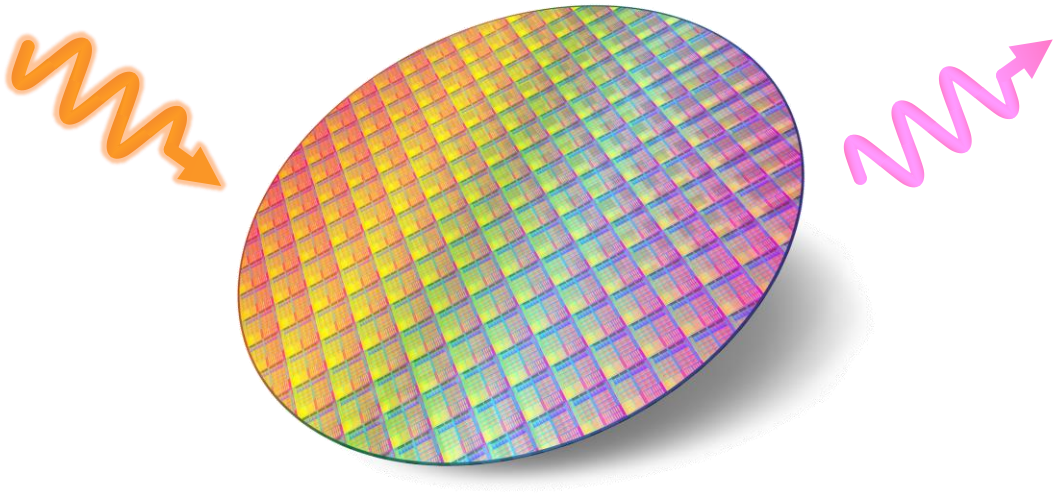
Link: IBM Research 2023

The Rise of Photonics



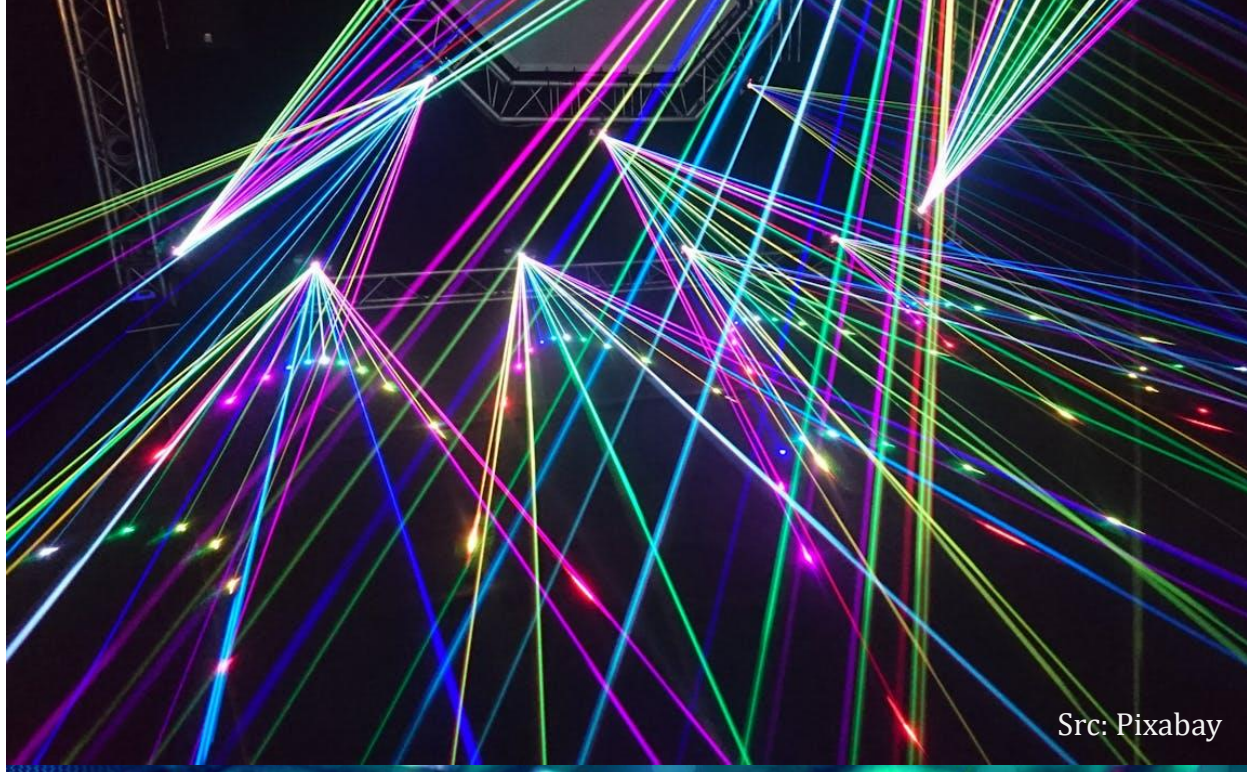
The Rise of Photonic ICs

Manipulating Light on Silicon Chips

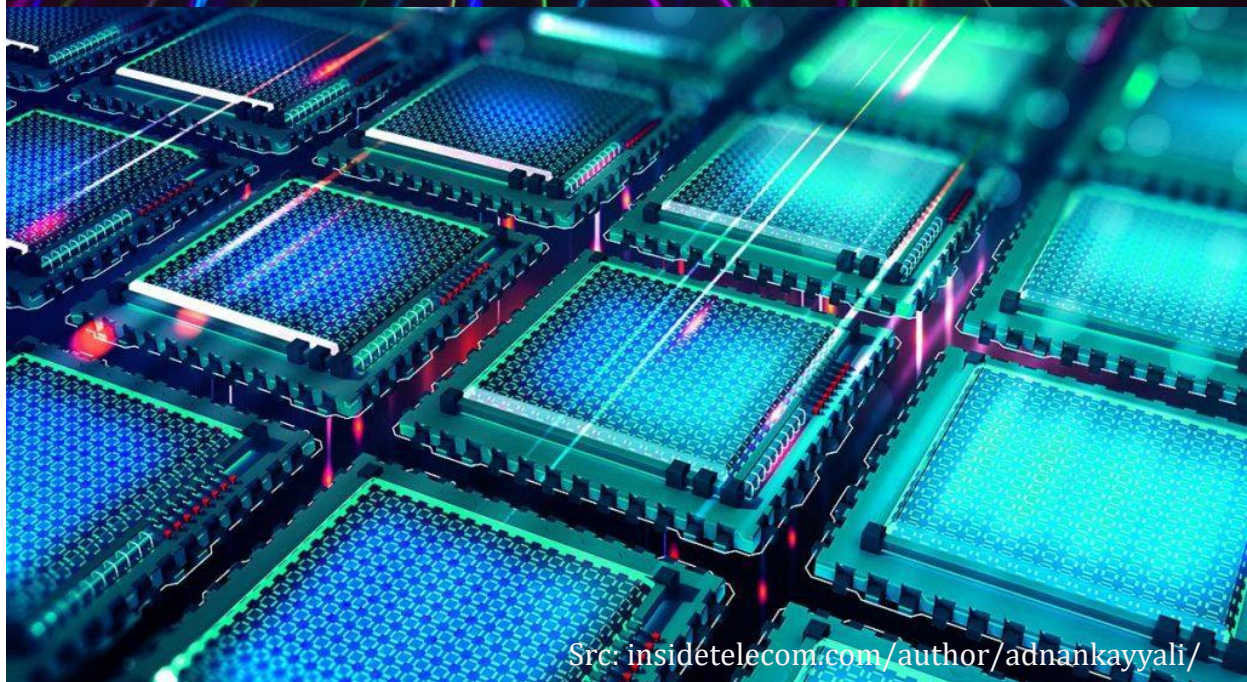


- \+ Energy efficiency
- \+ High bandwidth
- \+ Compactness
- × Complexity
- × Process requirements
- × Tech. stack ≠ standard

∴ Nano-scale CMOS Process: Si-Photonics ≙ MEMS



Src: Pixabay



Src: insidetelecom.com/author/adnankayyali/

The Rise of Photonic ICs

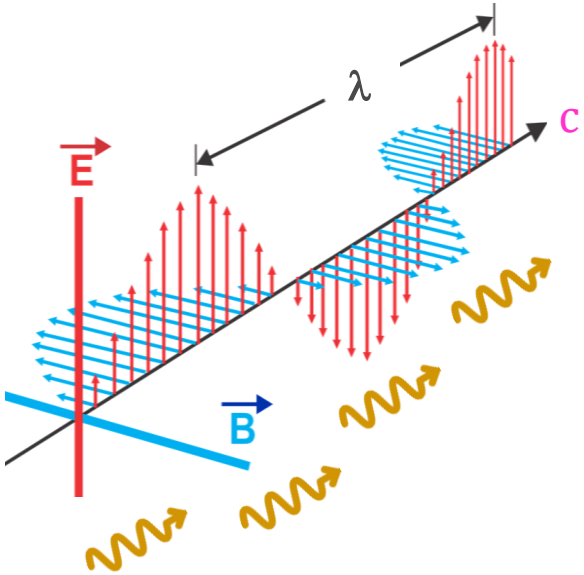
What is Light?

1) Electromagnetic wave

- Speed c = 300'000 km/h
- Wavelength: λ
- Frequency: ν

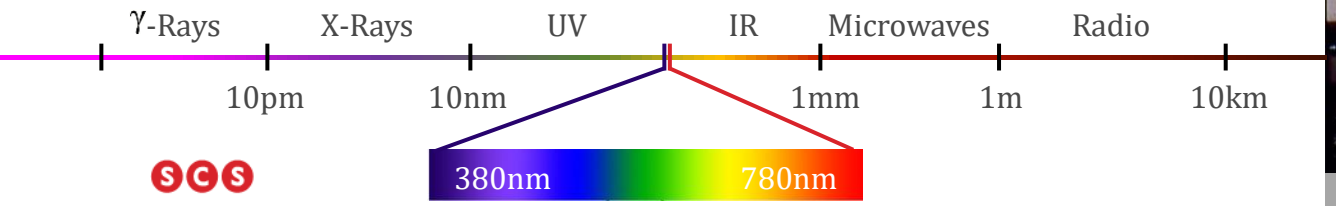
$c = \nu \times \lambda$

- Electrical Field: E
- Magnetic Field: B



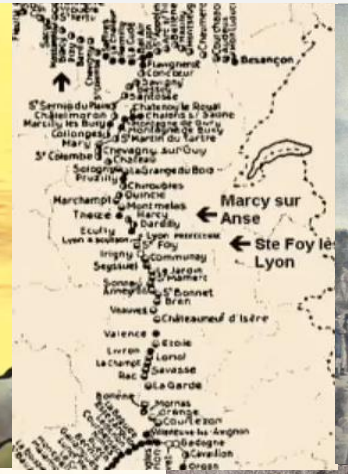
2) Flux of Photons

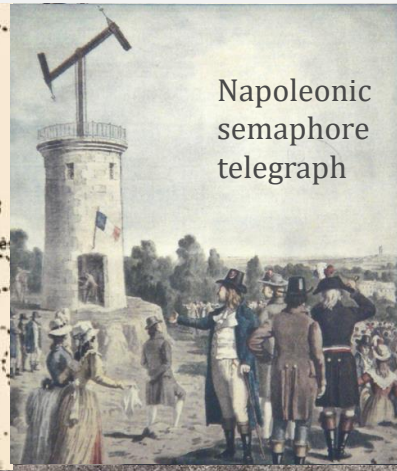
- Energy: h × ν



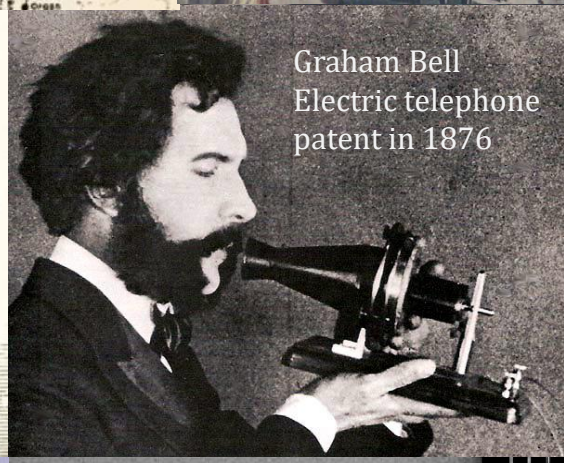
Optical Communication is Not New

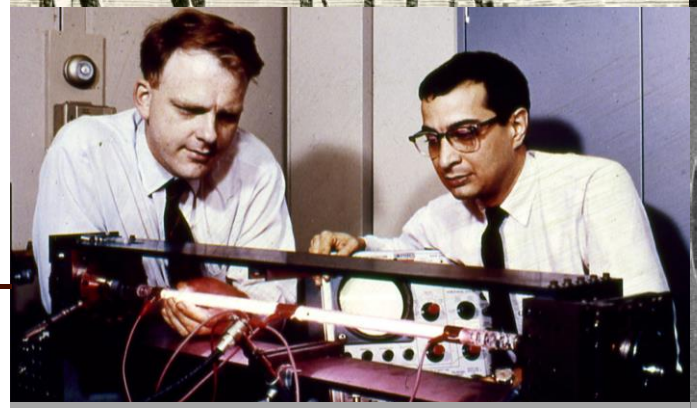
Src: Sean D'Souza

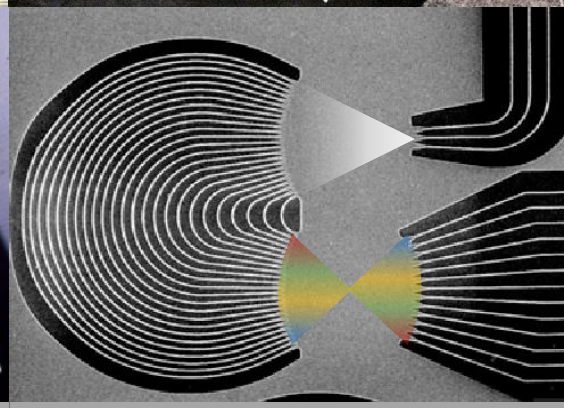



Napoleonic semaphore telegraph


Telegraph 1837

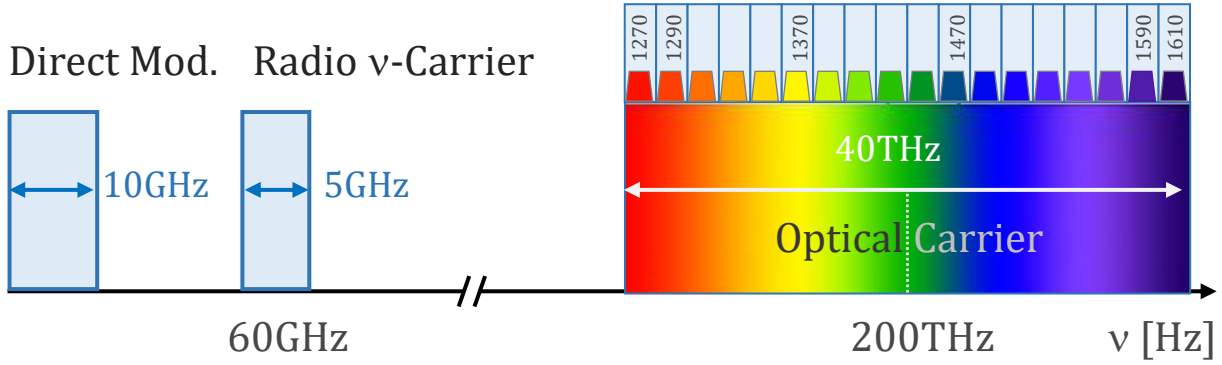

Graham Bell Electric telephone patent in 1876


Rubin Impulse laser, λ=694nm, 1960


1st PIC, AWG, Meint Smit, 1980

More .. Less

Key Advantages



More channel capacity (Shannon-Hartley)

$C = B \bullet \mathrm{Log}_2 (1 + S/N)$

C: CH-capacity [bps]

B: used bandwidth [Hz]

S/N: Signal-to-Noise Ratio

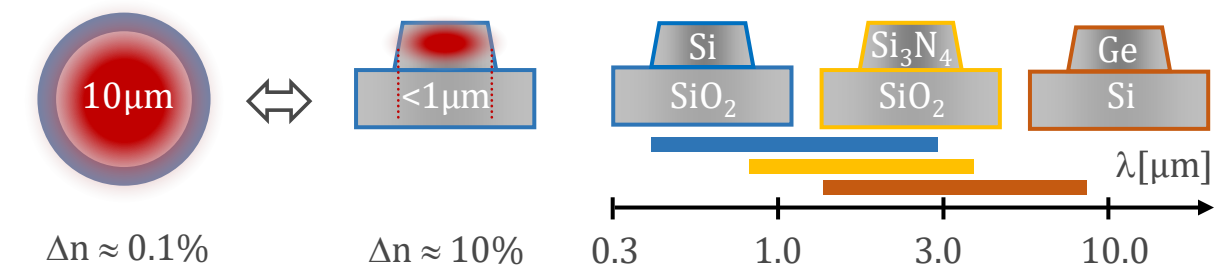
e.g. 125GHz ≈ Δ1nm @ 1.5µm

Better scaling & miniaturization

- Sub-µm waveguides: high n-contrast
- CMOS process compatible

Less energy

- Electrical : ≈ Nanojoule per bit flip
- Optical : ≈ Picojoule per bit flip



Photonic Integration

Key Ingrediencies

① Light source

② Light coupling

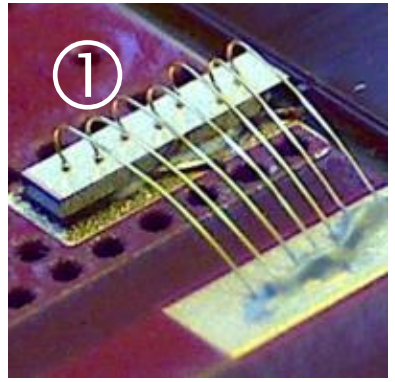
③ Light transport

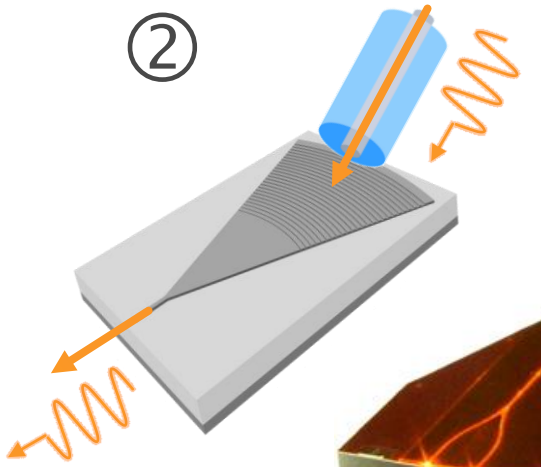
④ λ-Filtering / λ-multiplexing

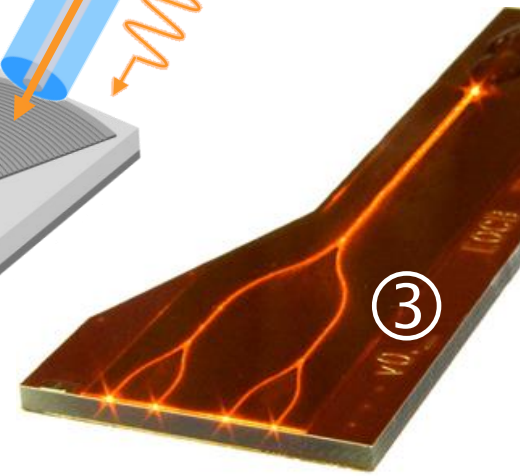
⑤ Signal modulation

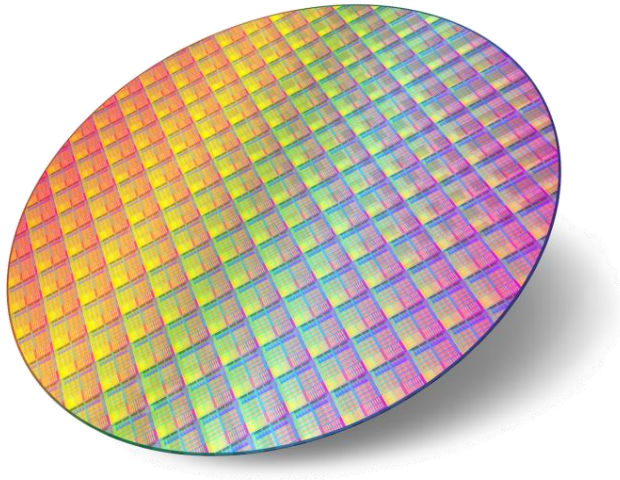
⑥ Signal detection

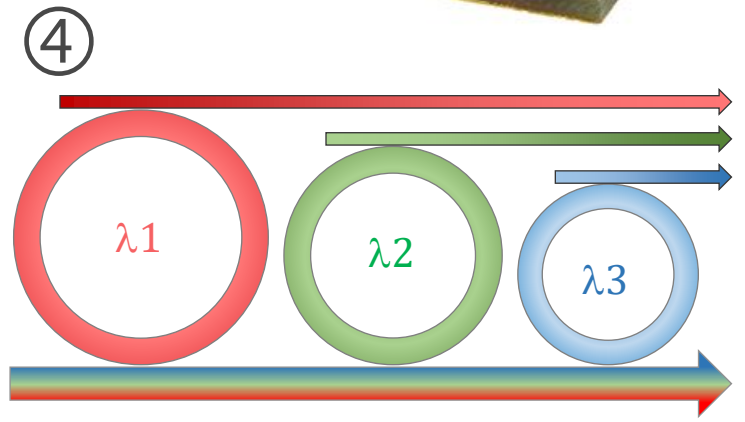
⑦ Supporting electronics



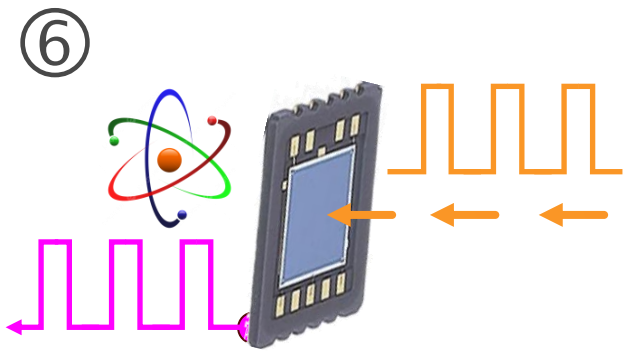


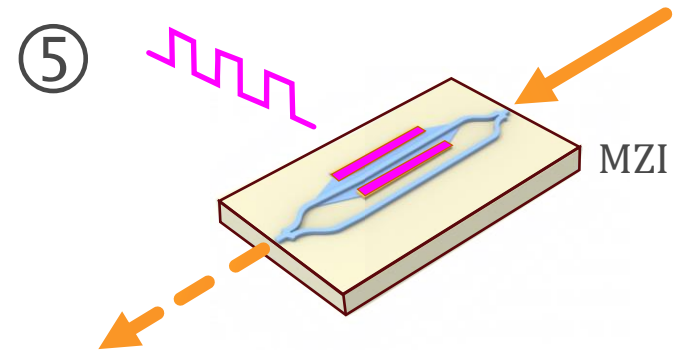






μ-Resonators





Challenges: Laser integration: Si ≠ III-V
CMOS precision: ≈1nm

Applications

Performance Scaling with Si-Photonics

Today	Tomorrow	Future
100-800 Gbps	1.6-6.4 Tbps	4-64 Tbps
25 pJ/bit	<15 pJ/bit	< 3 pJ/bit
SFPs	PIC / 3D MCP	PIC & EIC

Increasing optical interconnects

- Rack-to-Rack ⇨ Chip-2-Chip ⇨ On-Chip

Trends: Classic SFP ⇨ co-packaged PICs
Chip-2-chip tracks ⇨ PICs

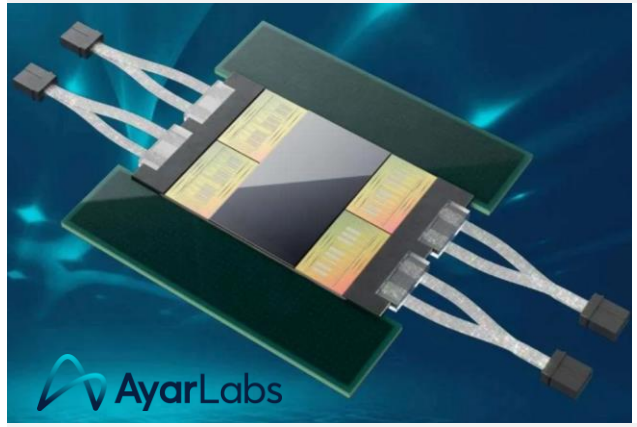
Industry 1st Ethernet switch with CPO: Barefoot Tofino 2

4 co-packaged optic ports
- 64× 100G ports

Tofino 2 vs. std. SFP
- BW density ⇗: 40x
- PWR ⇘: 30%

Apps ⇨ I/O
- Rack-2-Rack

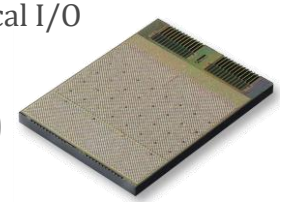
1st UCIe compatible optical interconnect: TeraPHY



16 co-packaged optic ports
- 8 Tbps BW

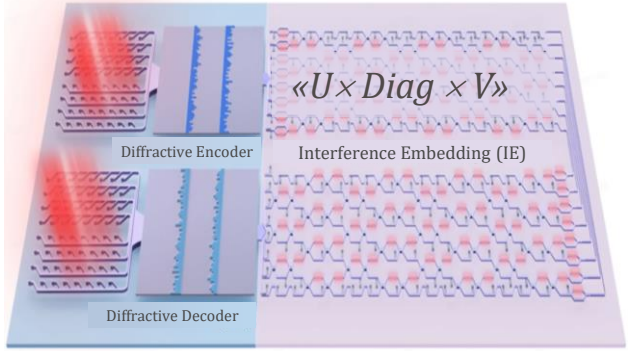
AyarLabs

TeraPHY vs. electrical I/O
- Data rate ⇗: 5×



Apps ⇨ I/O (mm - km)
- Chip-to-Chip
- Rack-to-Rack

Large-scale high-throughput photonic AI chiplet: TaiChi



Optical neural net (ONN)
- 14mio neurons / chip

清華大學 Tsinghua University

GPU H100 vs. TaiChi
- Energy ⇘: 1000×

Apps ⇨ AI @ Chip
- Scene recognition ≤ 1k classes
- Content generation AGI

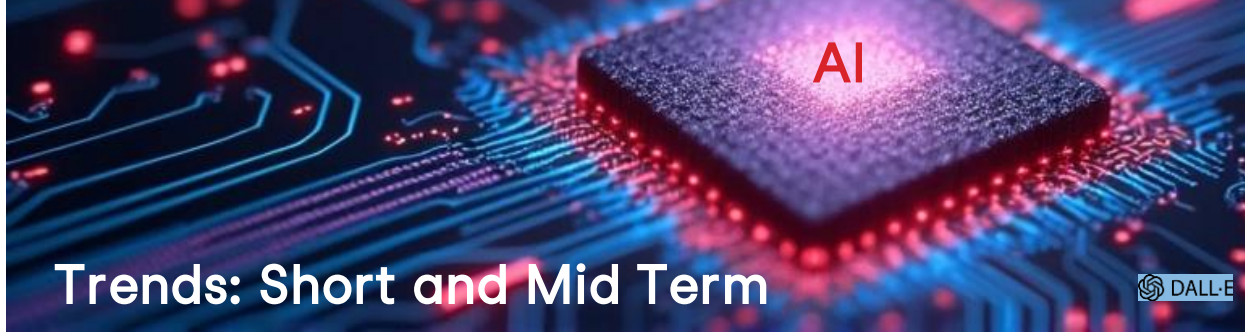
Src: www.doi.org/10.1126/science.adl1203

Si-Photonics: Quo Vadis?

Pros and Cons

- \+ High integration @ waver scale
- \+ High bandwidth: I/O bottleneck↘
- \+ Cost-effective manufacturing
- \+ Energy efficiency

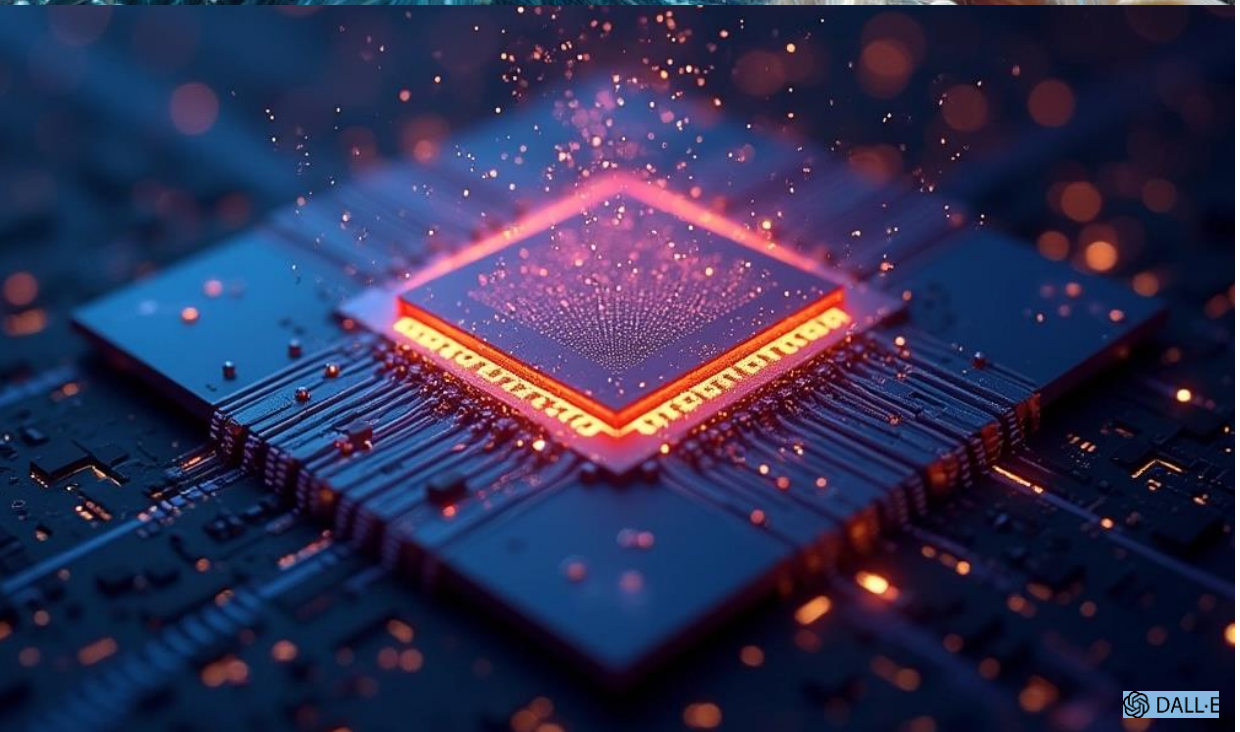
- ✕ Material limitations
- ✕ Complex packaging
- ✕ Optical losses & thermal sensitivity
- ✕ Extreme CMOS process accuracy: ≈1nm



- PICs can enhance DC-performance with high-speed, energy-efficient interconnects:
  - Optical transceivers: BW⇗, distance⇗, space↘
  - Server interconnects: copper cables ⇨ SiPh links
  - On-board CPO for optical chip-2-chip links

- PIC with optical artificial neural nets may
  - Fuel scalable AI advancements
  - Extend Moore's Law trajectory

BW: Bandwidth CPO: Co-Packaged Optics DC: Data Center PIC: Photonic Integrated Circuit SiPh: Silicon Photonics



Take away

Neuromorphic Computing is

- -based, sparse, async. communication
- low-power, real time, adaptive learning
- ready to leave the lab!

Silicon Photonics is

- ready to enhances DC-performance with high-speed, energy-efficient interconnects

... and may

- boost scalable AI applications
- extend Moore's Law trajectory!

Where do you see the greatest potential for you?

super computing systems

Supercomputing Systems

Christof Bühler
Senior Advisor
+41 43 456 16 59
www.scs.ch
christof.buehler@scs.ch

Wer sind wir

SCS besteht aus 140 hochqualifizierten Ingenieuren, Informatikern & Physikern, die sich für die Lösung technologischer Herausforderungen begeistern.

Unser gesamtes Team ist in Zürich.

Wir haben mehr als 400 Kunden in 31 Jahren - wir wissen, wie man Projekte durchführt.

Alle IPs gehen an unsere Kunden.

1995
GigaBooster

SCS
super computing systems