LightOn

We bring Light to AI

LightOn Optical Processing Unit

Scaling-up AI and HPC with a Non von Neumann co-processor

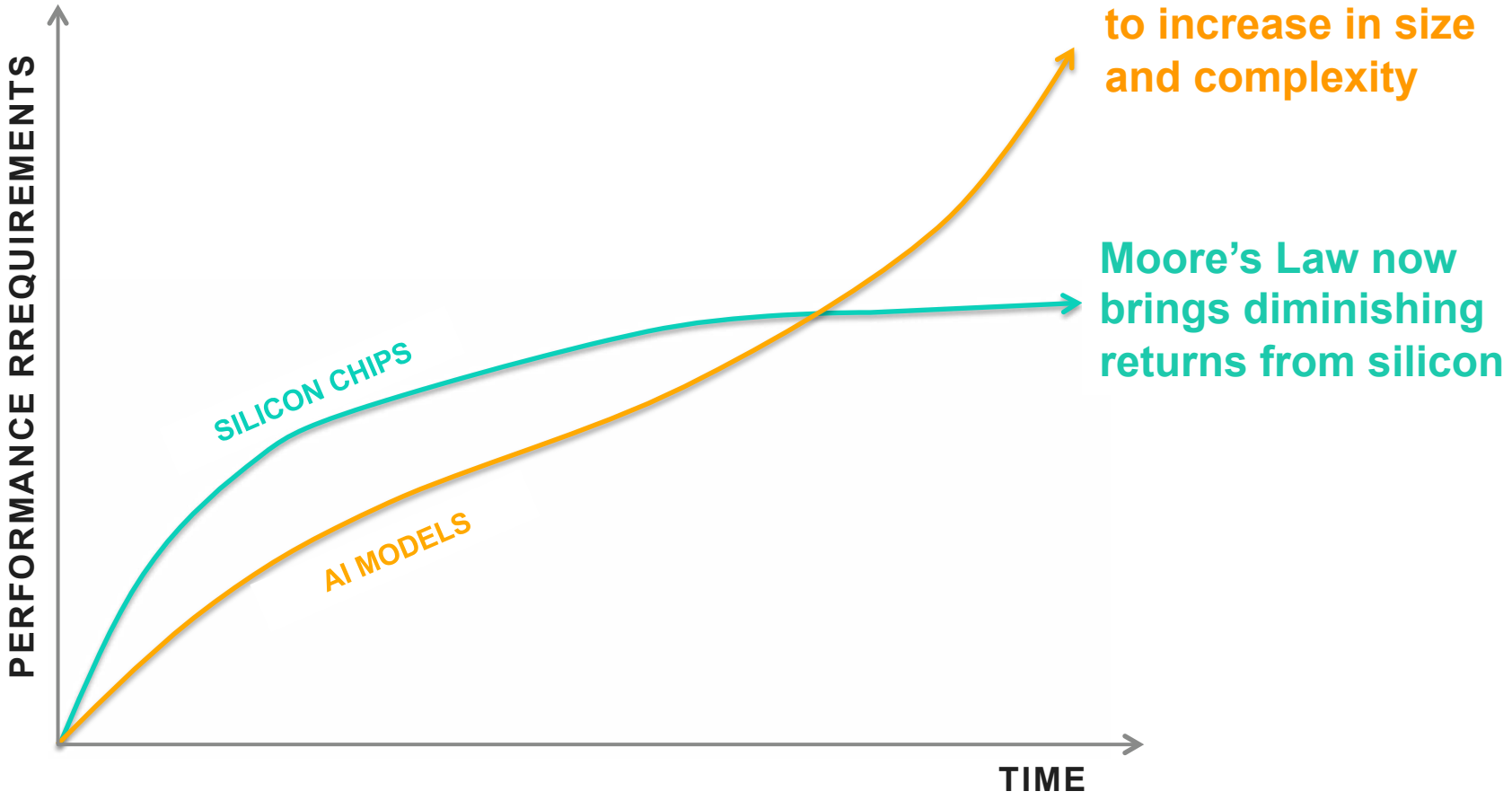
Charles BROSSOLLET, Alessandro CAPPELLI, Igor CARRON, Charidimos CHAINTOUTIS, Amélie CHATELAIN, Laurent DAUDET, Sylvain GIGAN, Daniel HESSLOW, Florent KRZAKALA, Julien LAUNAY, Safa MOKAADI, Fabien MOREAU, Kilian MULLER, Ruben OHANA, Gustave PARIENTE, Iacopo POLI, and Giuseppe L. TOMMASONE

LightOn, Paris, France

HOT CHIPS

IEEE Hot Chips 33 symposium, August 2021

Artificial Intelligence requires *much* more compute



AI models continue to increase in size and complexity

Moore's Law now brings diminishing returns from silicon

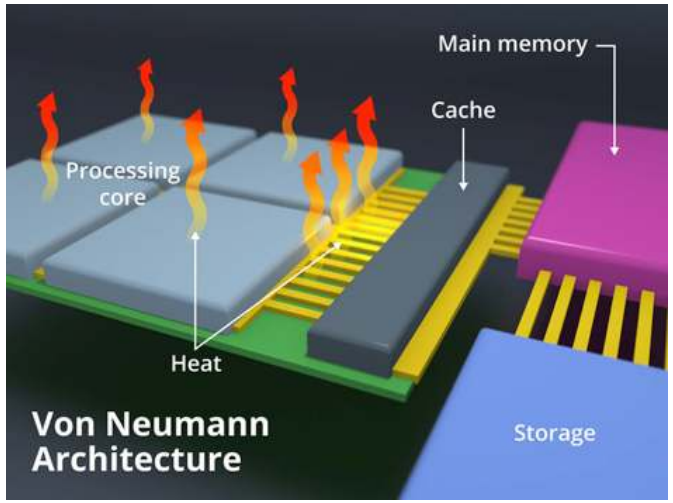
It now takes whole supercomputers to train AI models

Beyond pure Von Neumann processing

Scalability of AI / HPC models is limited by the

Von Neumann bottleneck

for accessing massive amounts of memory,
driving up power consumption.



http://www.rochester.edu/newscenter/microprocessors-computing-architecture-304252/vonneumann-architecture/

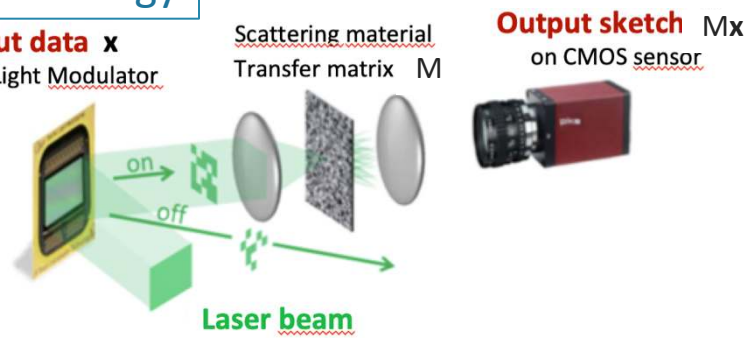
→ A new hybrid computing paradigm combining classical components and photonic Non Von Neumann accelerators

Outline

LightOn Appliance: large-scale NvN photonic accelerator

1st accelerator with LightOn OPU technology on the market

OPU technology



Input vector **x** → Output vector **y** = M**x** or **y**=|M**x**|^2

Dimension of **x** up to 1 million

Dimension of **y** up to 2 millions

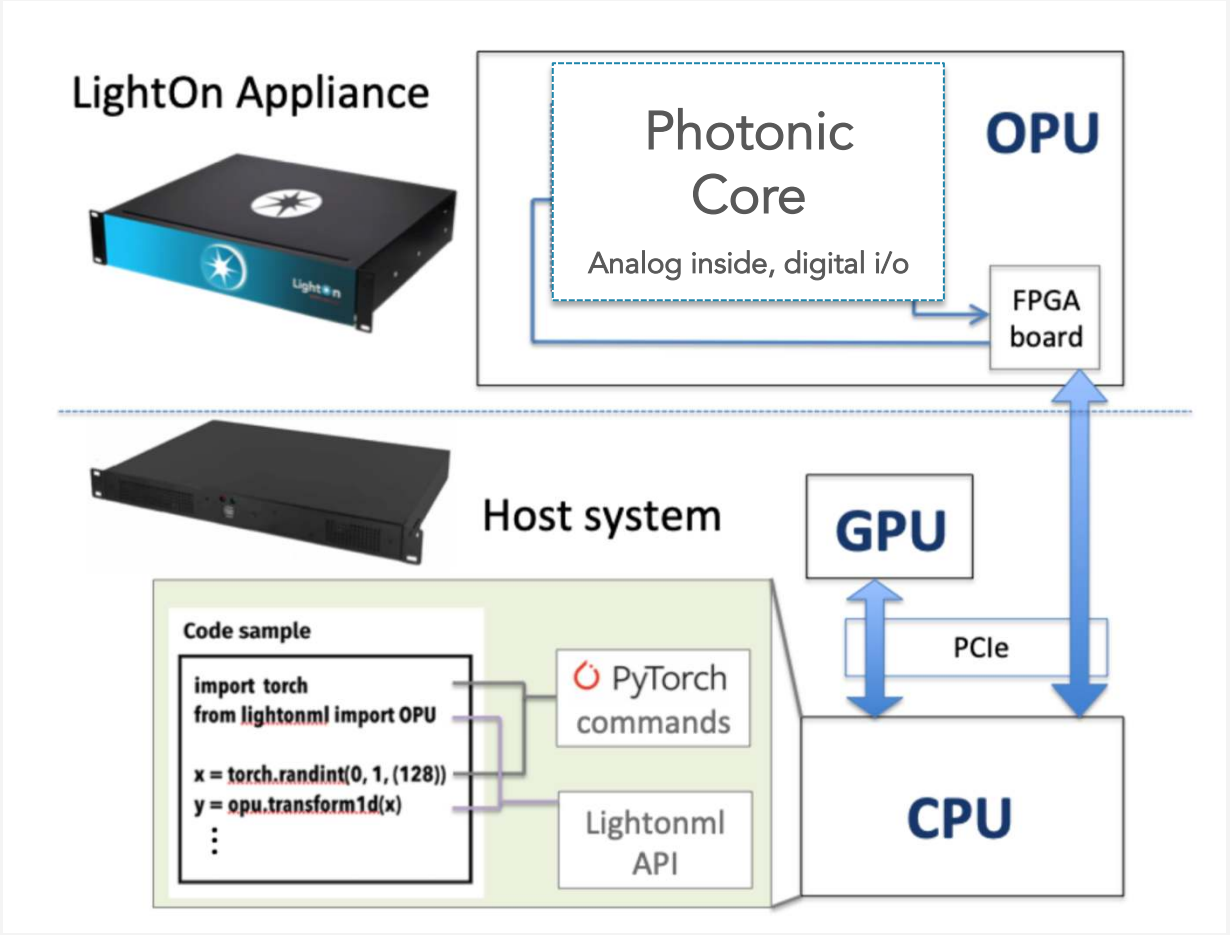
M coefficients matrix
fixed by design
2.10^{12} random i.i.d. gaussian numbers

1.500 TOPS
in a single photonic core

max. 30 W power consumption

Hybrid computing pipeline

The Appliance is connected to the host server through external PCIe



Seamless python™ programming with API compliant with PyTorch scikit learn

AI use cases: OPU makes CPUs and GPUs more efficient

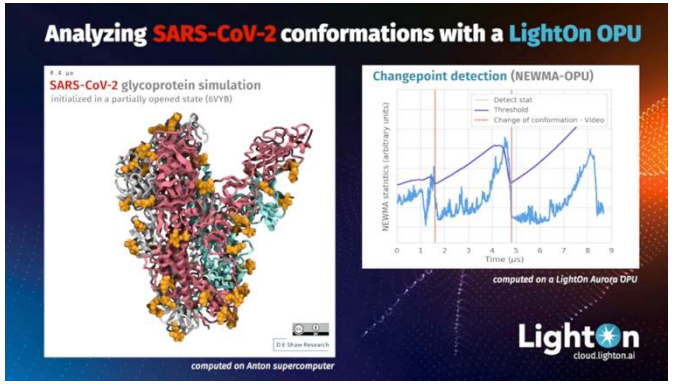
A set of examples of OPU use in AI algorithms (1/2)

TRANSFER LEARNING
In convolutional Deep NN

Fast adaptation of a pre-trained image or video DNN classifier

TIME SERIES ANALYSIS

Using the OPU to detect change points in million-dimension time series



And many more can be found on our blog: www.lighton.ai/blog

AI use cases: OPU makes CPUs and GPUs more efficient

A set of examples of OPU use in AI algorithms (2/2)

ADVERSARIAL ROBUSTNESS
by design

collaboration with

criteo

DISTRIBUTED TRAINING
for DNN through
Direct feedback Alignment

DIFFERENTIAL PRIVACY
by design

collaboration with

facebook AI Research

And many more can be found on our blog: www.lighton.ai/blog

HPC Use case: Accelerated Scientific Computing

Randomized Numerical Linear Algebra

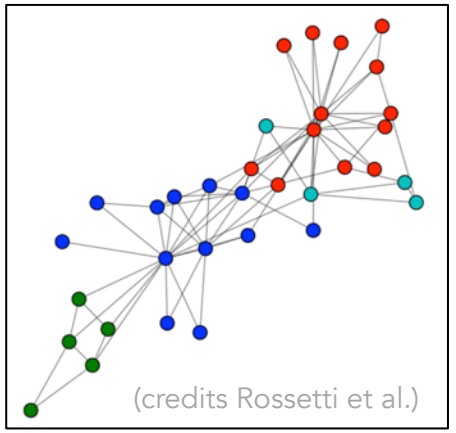
DOE RASC report (Jan 2021): randomized algorithms are "essential to the future of computational science and AI for Science."

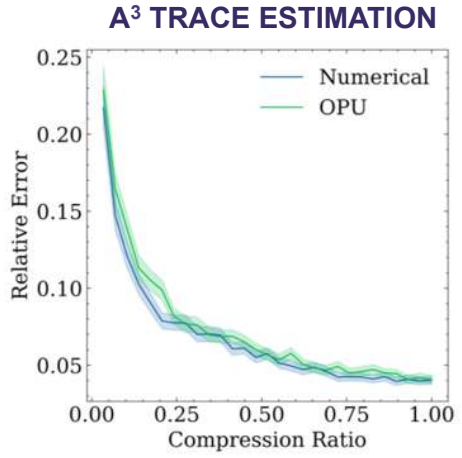
Use the OPU to optimally compress data for

- Approximate matrix multiplications
- Randomized SVD
- Randomized Inverse Problems...

See joint HC33 presentation by D. Hesslow *et al.* *Photonic co-processors in HPC: Using LightOn OPUs for Randomized Numerical Linear Algebra*

Example: Community detection in large networks
→ Triangle counting on graphs





Conclusion

- Not all coefficients need to be updated → hybrid schemes with NvN
- Model-free optimal coefficient distributions are random gaussian
  → Universal data compression / sketching
- Free-space photonics provide such computations at trillion scale
- Now on the market, seamlessly integrated into hybrid pipelines.

→ a unique pathway to “optical advantage”

LightOn

We bring Light to AI

Laurent Daudet

CTO

contact@LightOn.ai