Optical Computing for Large Scale Artificial Intelligence

LightOn

Using Light to change the Future of Computing

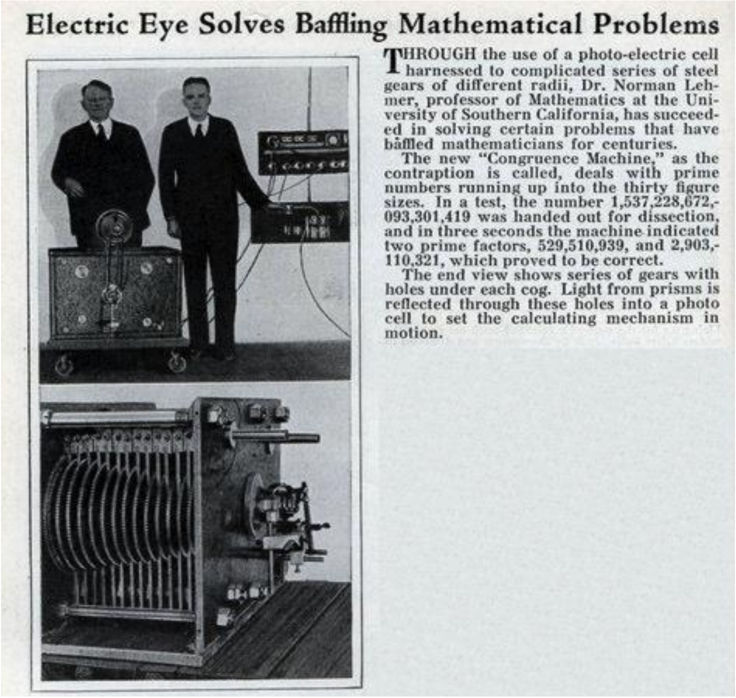
Igor Carron, CEO and Co-Founder

LightOn

Confidential & Proprietary

A short history of Optical Processing of Information

From Sieves ... to Fourier Transforms ... all the way to Neural Networks

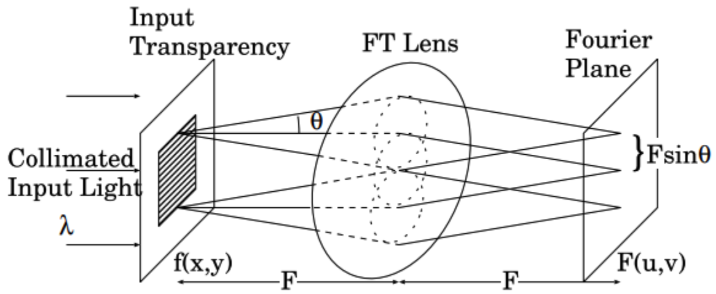


Electric Eye Solves Baffling Mathematical Problems

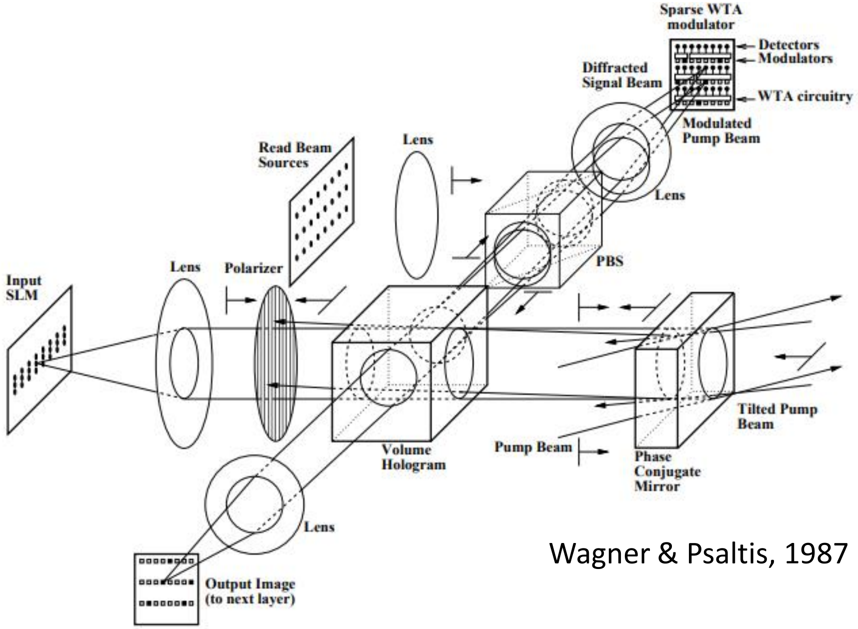
THROUGH the use of a photo-electric cell harnessed to complicated series of steel gears of different radii, Dr. Norman Lehmer, professor of Mathematics at the University of Southern California, has succeeded in solving certain problems that have baffled mathematicians for centuries.

The new "Congruence Machine," as the contraption is called, deals with prime numbers running up into the thirty figure sizes. In a test, the number 1,537,228,672,093,301,419 was handed out for dissection, and in three seconds the machine indicated two prime factors, 529,510,939, and 2,903,110,321, which proved to be correct.

The end view shows series of gears with holes under each cog. Light from prisms is reflected through these holes into a photo cell to set the calculating mechanism in motion.



$$F(u,v) = \int\int f(x,y) e^{i\frac{2\pi}{\lambda F}(xx'+yy')} dxdy$$



Wagner & Psaltis, 1987

1930's

1950's

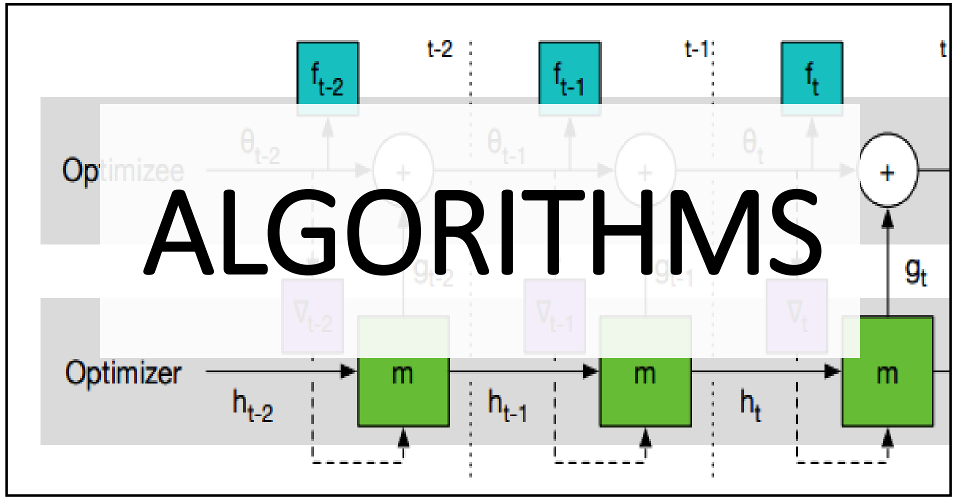
1980's

Confidential & Proprietary

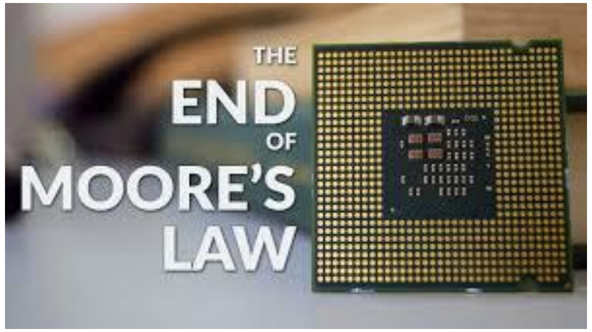
Then Came Winter

https://imgflip.com/memegenerator/18552174/Winter-Is-Coming

Rebooting Optical Computing: the AI era



SCALABLE ?



https://www.youtube.com/watch?v=Ak7HPuuJ1Ow

SUSTAINABLE ?

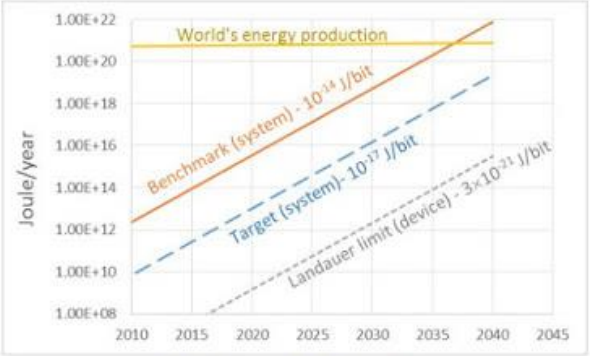


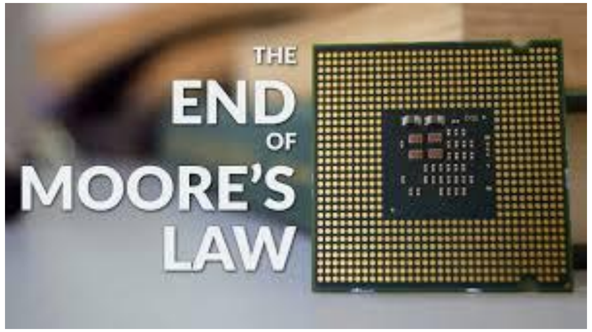
Fig. A8. Total energy of computing.

Guardian Environment Network

'Tsunami of data' could consume one fifth of global electricity by 2025

Rebooting Optical Computing: the AI era

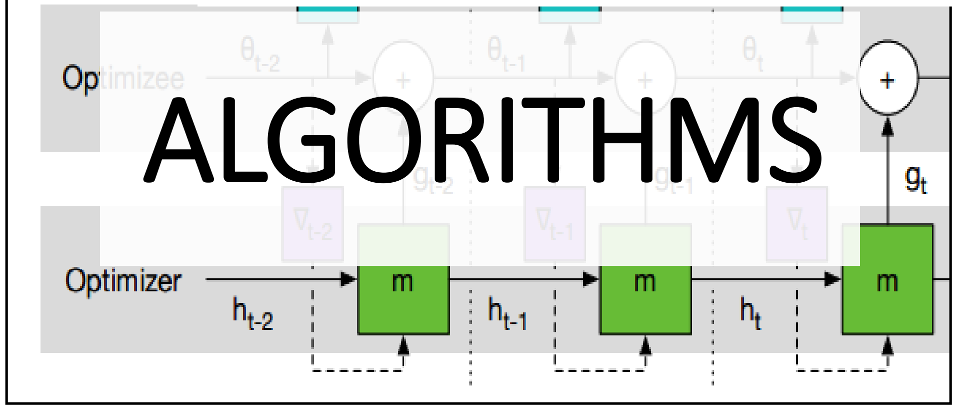
SCALABLE ?



https://www.youtube.com/watch?v=Ak7HPuuJ1Ow

Business

World is running out computer power, warns Microsoft boss Nadella



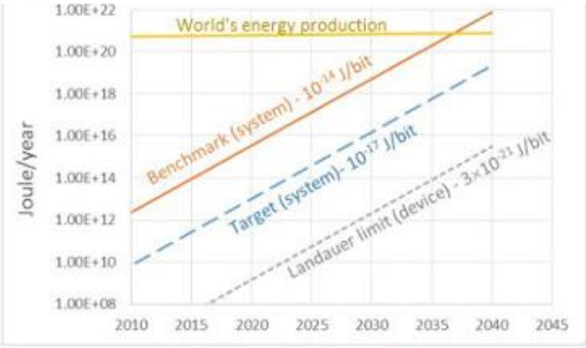


Fig. A8. Total energy of computing.

Guardian Environment Network

'Tsunami of data' could consume one fifth of global electricity by 2025

Computing these days

"…. there is massively more information sent at shorter distances

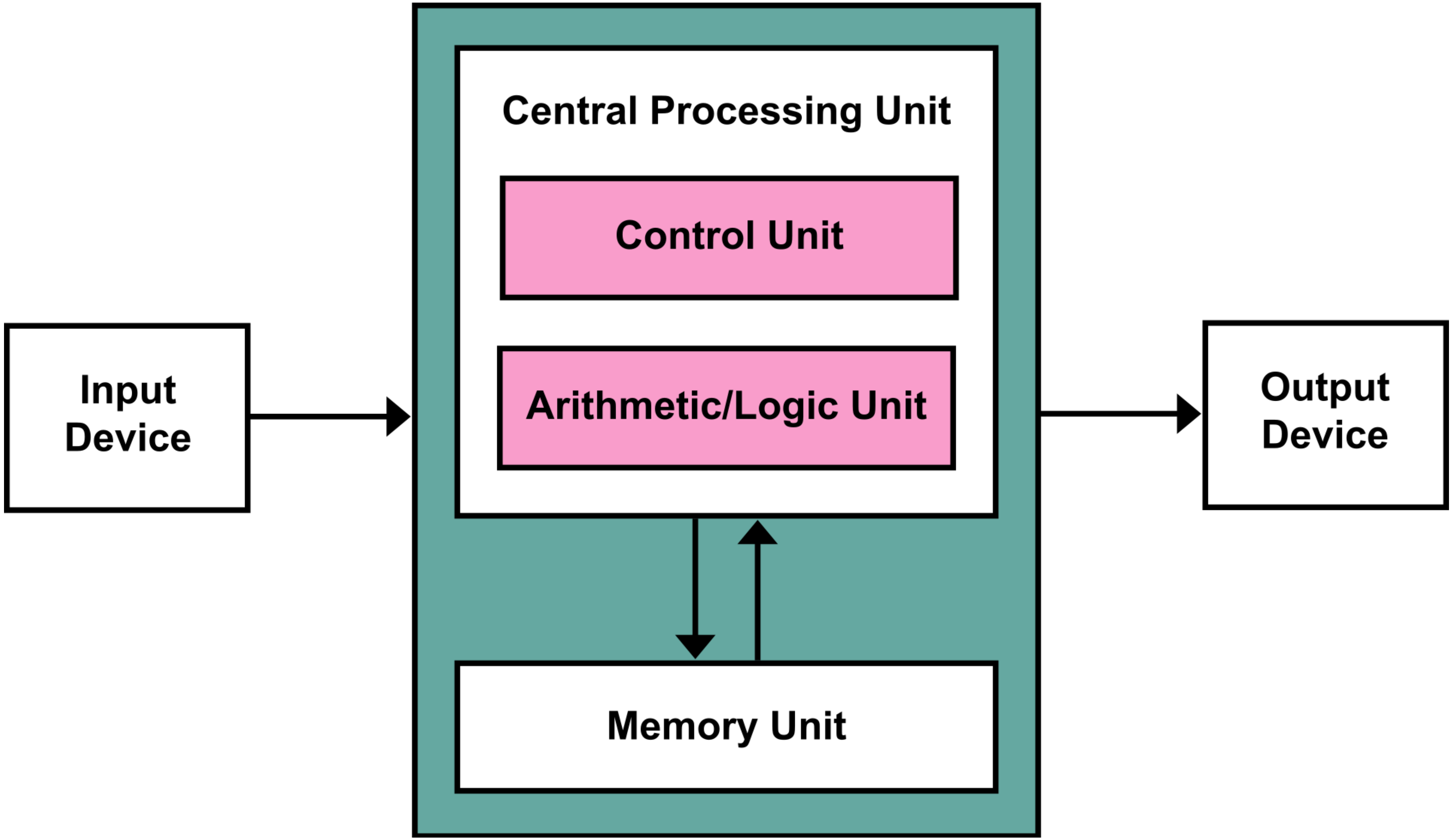
so much so that

most energy dissipation is in shorter links and in

<u>interconnects inside machines</u>…"

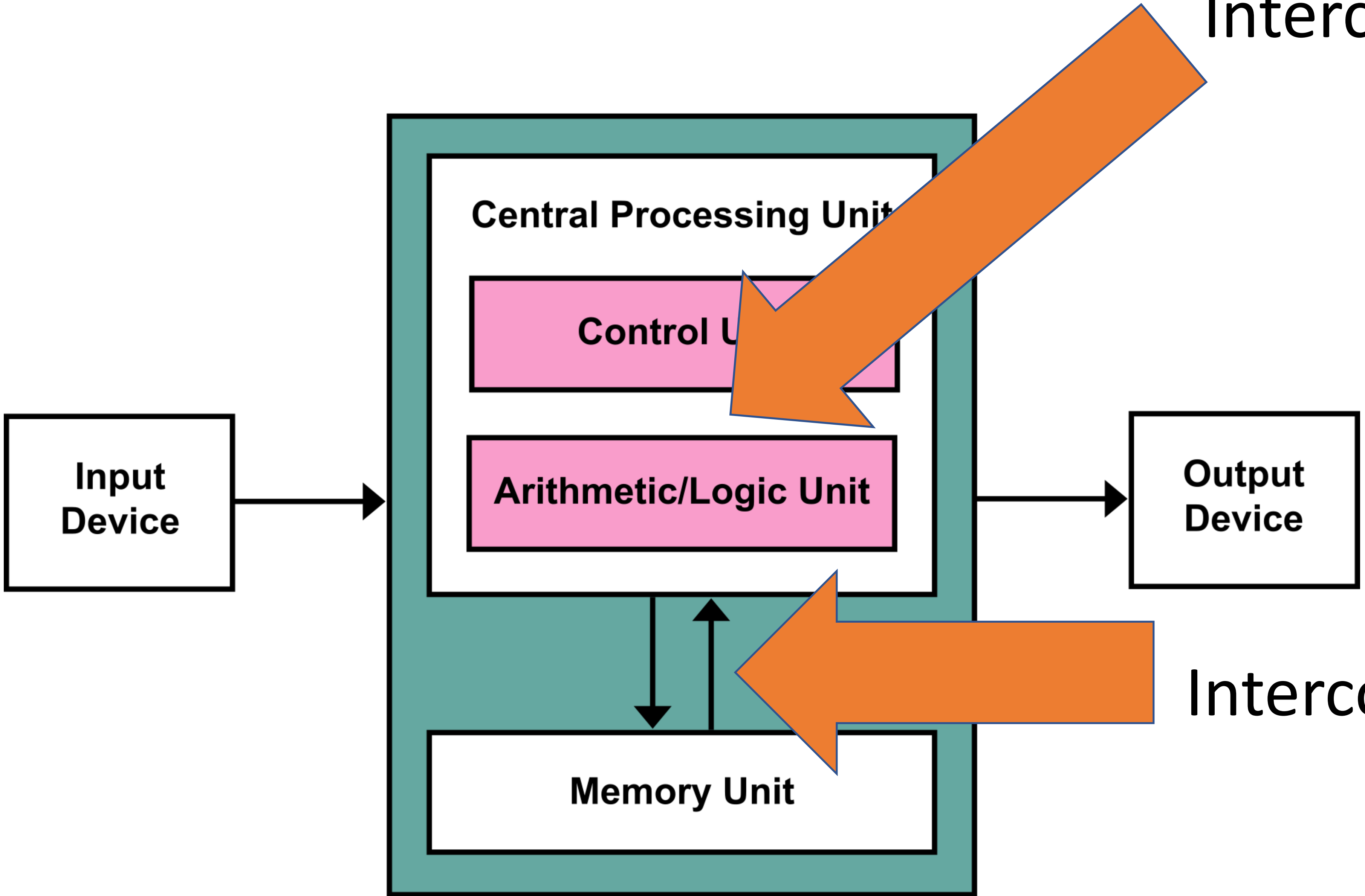
David Miller, Stanford EE

Von Neumann Architecture



Von Neumann Architecture

Interconnectsc

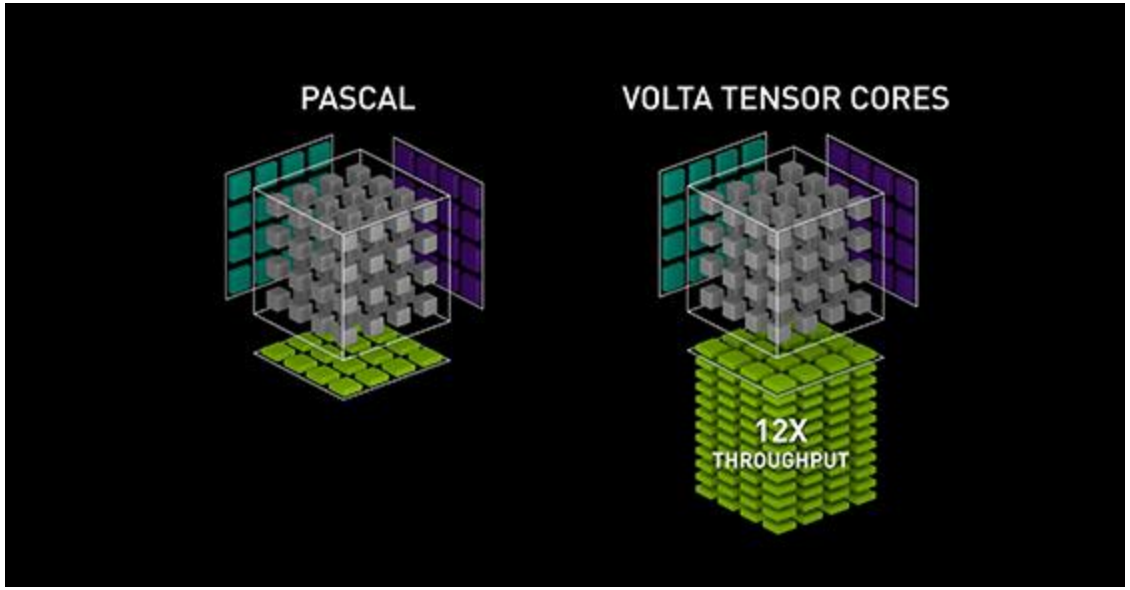


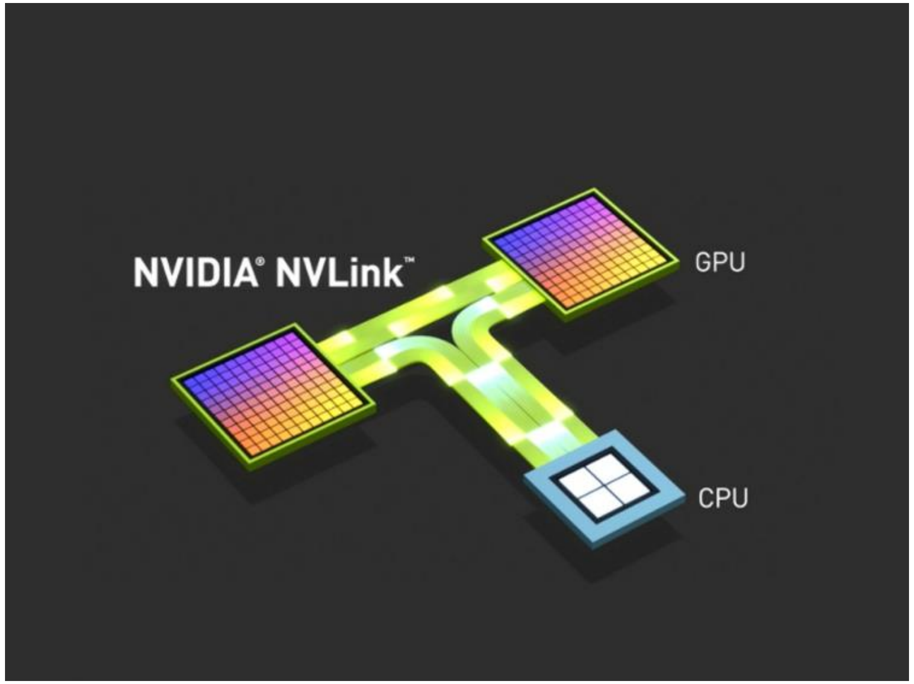
Interconnects

LightOn

Computing these days

It’s all about accessing the memory for your computations !





NVIDIA®

LightOn: The Founding Team

Igor Carron

Laurent Daudet

Sylvain Gigan

Florent Krzakala

LightOn
Confidential & Proprietary

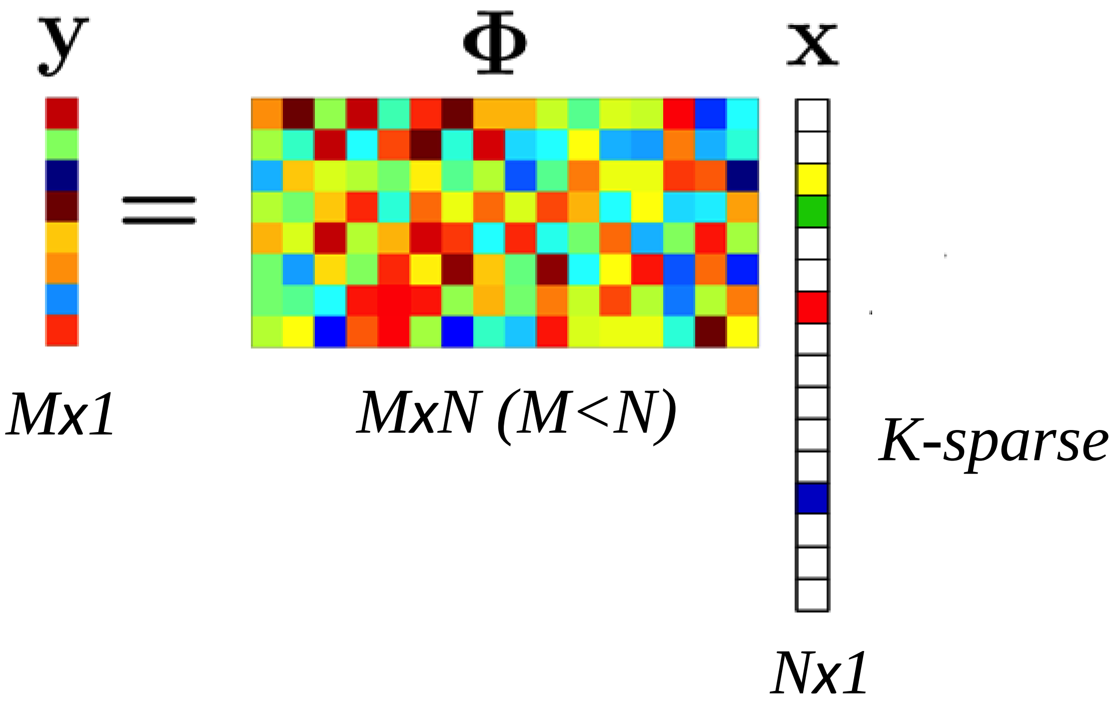
Information theory
Compressive Sensing

Igor

http://nuit-blanche.blogspot.com

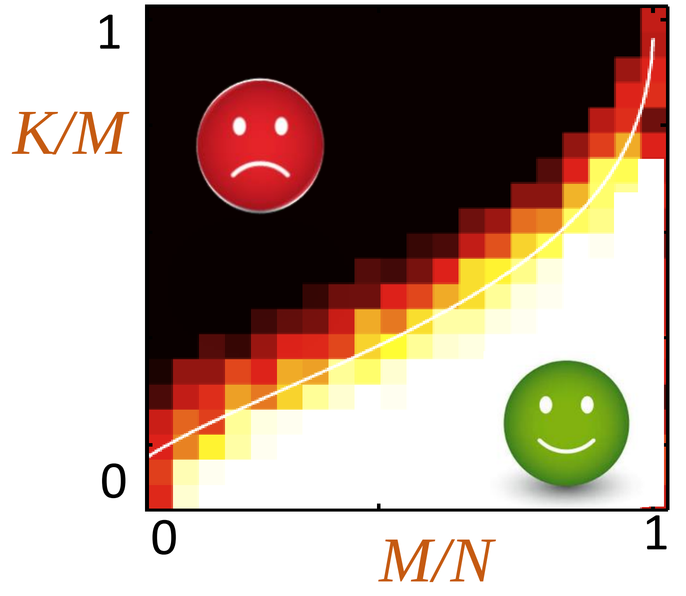
Laurent

Compressive Sensing



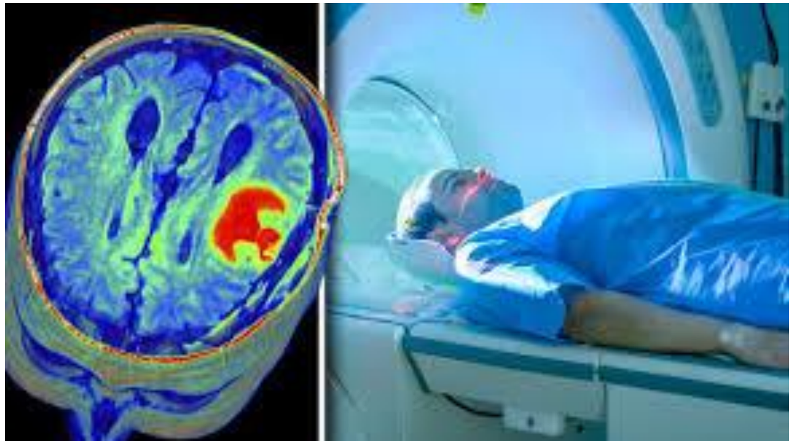
Can one recover x from y ?

YES with tractable algorithms for right values of *N*, *M*, *K*



Lessons from Compressive Sensing

- Signals can be sampled at the level of their information content
- Random Projections are very good for sensing at low data rate
- Strong theoretical background and large empirical evidence



Information theory
Compressive Sensing

Igor

Laurent

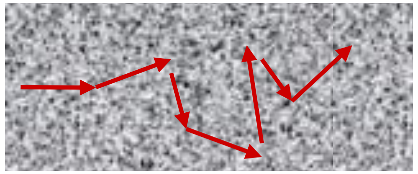
Light Transport in Diffusive Media

Sylvain

Light transport in diffusive media

Confidential & Proprietary

Origin: light is scattered by inhomogeneities



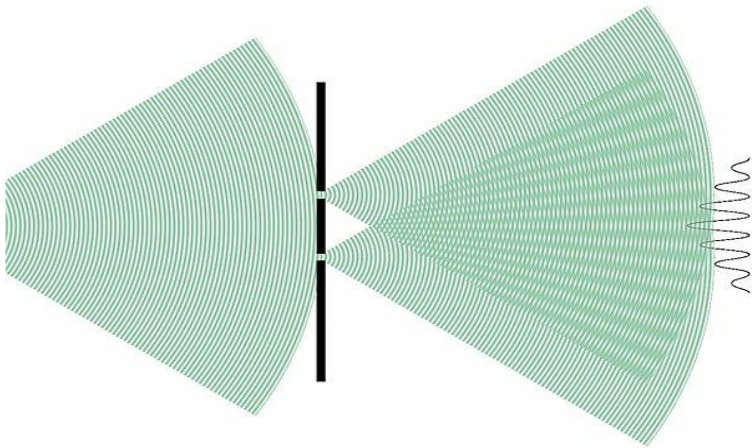
Light transport in diffusive media

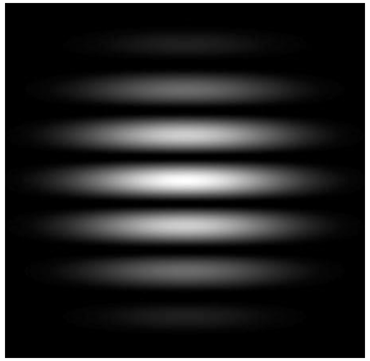
Light transport in diffusive media

Young's slit experiment:

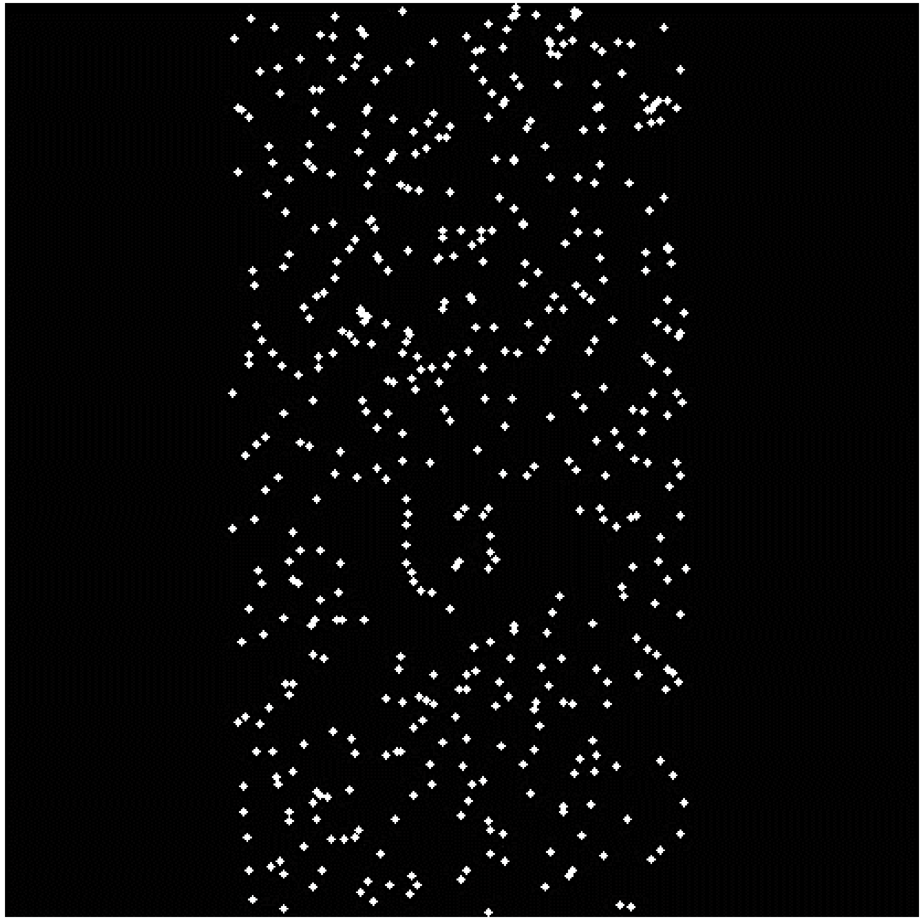
two wave interference

Fringes



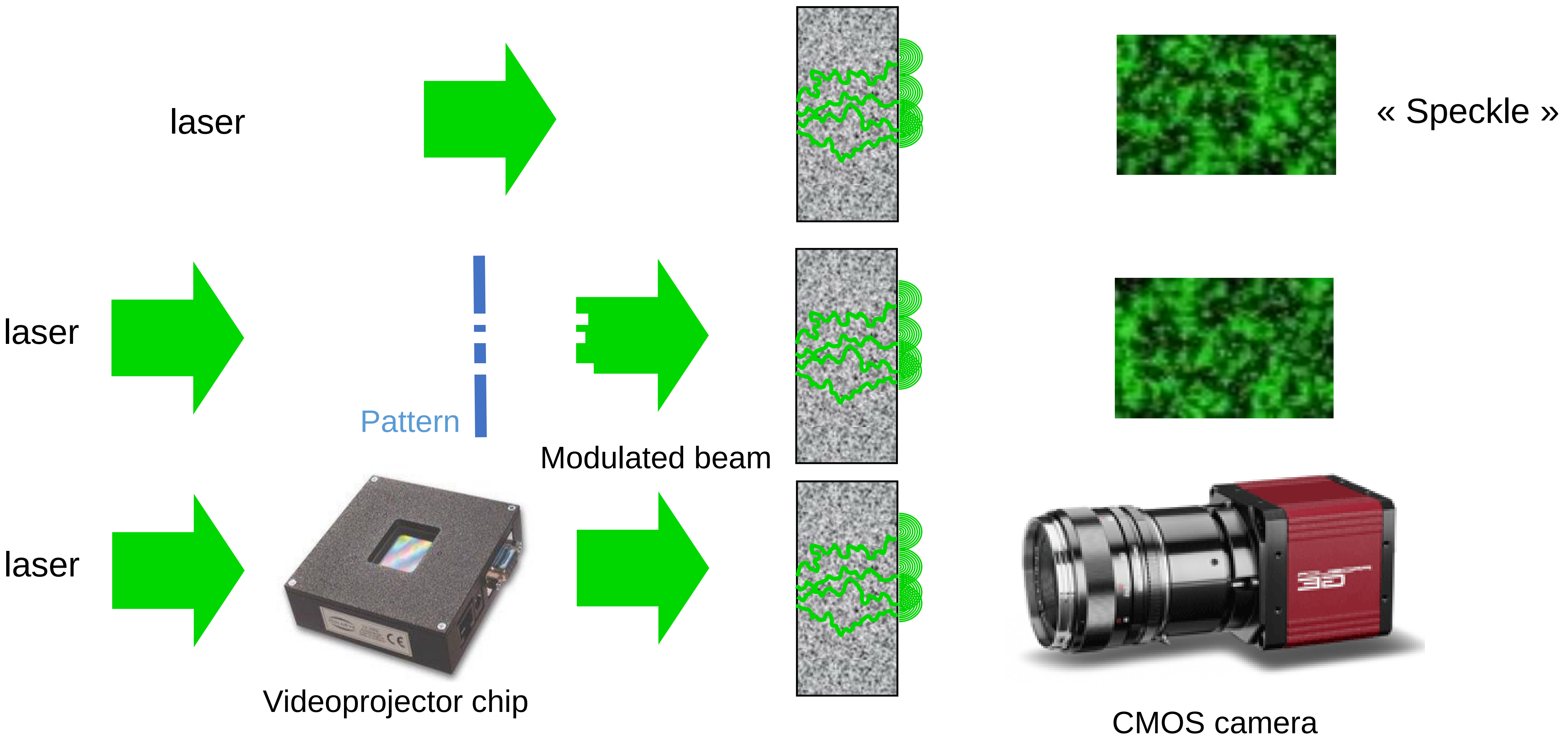


Light transport in diffusive media



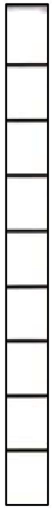
Credit: E. Bossy (UGA), SimSonic.

Scattering: a coherent process



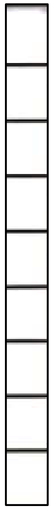
Scattering: a coherent process

Discrete input vector



$$\begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,N} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,N} \\ \vdots & & \ddots & \vdots \\ h_{M,1} & h_{M,2} & \cdots & h_{M,N} \end{pmatrix}$$

Random transmission matrix

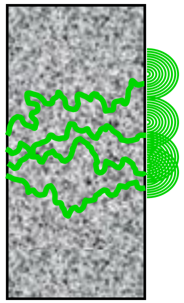


Discrete output vector (speckle intensity)

$$\mathrm{x} \qquad \mathrm{H} \qquad \mathrm{y} = |\, \mathrm{H}\, \mathrm{x} \,|^2$$

laser

Videoprojector chip

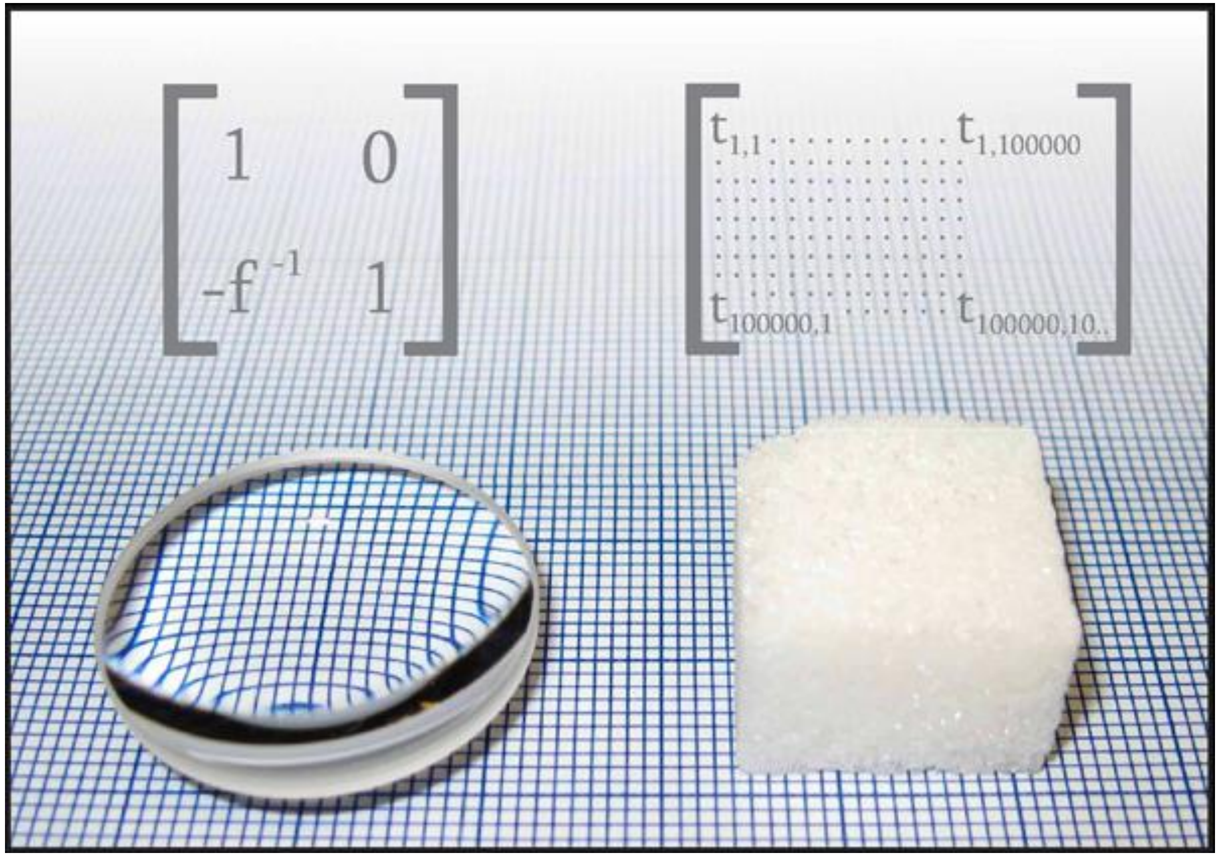


CMOS camera

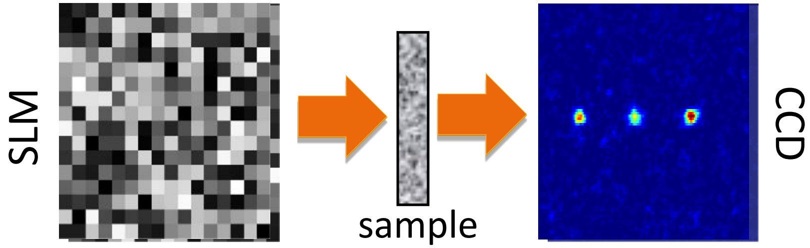
Popoff et al. *Nat. Commun.* 1:81 doi: 10.1038/ncomms1078 (2010)

The transmission matrix

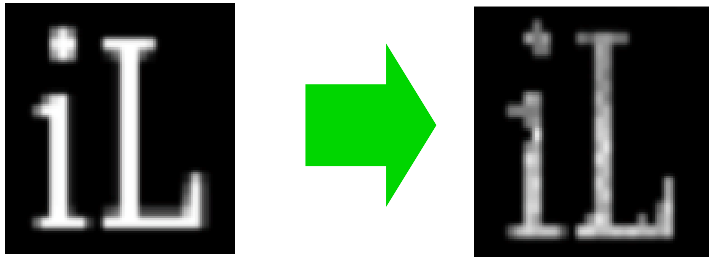
Scattering materials are « super-lenses »



Focusing ++



Imaging ++



Popoff et al. Phys. Rev. Lett. 104,100601 (2010) / Liutkus et al., Scientific Reports 4, 5552 (2014)

Lessons from Light Transport in Diffusing Media

- Scattering preserves the information content: it is possible to « see » through a thick layer of scattering material
- Scattering *optimally* mixes information, evenly spread on output pixels:
  - just like a Random Projection
  - just like in Compressive Sensing
- Matrix-vector multiplication, followed by non-linearity: sounds familiar ?

Information theory
Compressive Sensing

Igor

Machine Learning

Laurent

Light Transport in Diffusive Media

Sylvain

Florent

Random Projections in Machine Learning

- Random Projections act as distance-preserving point cloud embeddings

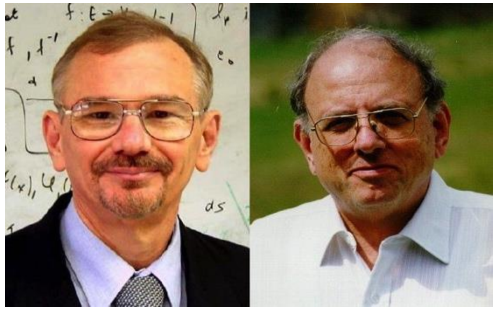
Johnson-Lindenstrauss Lemma (1984)

Lemma For any $0 < \epsilon < 1$ and any interger n let k be a po sitive interger such that

$$k \geq \frac{24}{3\epsilon^2 - 2\epsilon^3} \log n$$

then for any set A of n points $\in \Re^d$ there exists a map $f : \Re^d \rightarrow \Re^k$ such that for all $x_i, x_j \in A$

$$(1-\epsilon)||x_i - x_j||^2 \leq ||f(x_i) - f(x_j)||^2 \leq (1+\epsilon)||x_i - x_j||^2$$



- Supervised Learning
- Unsupervised Learning: Randomized PCA, etc…

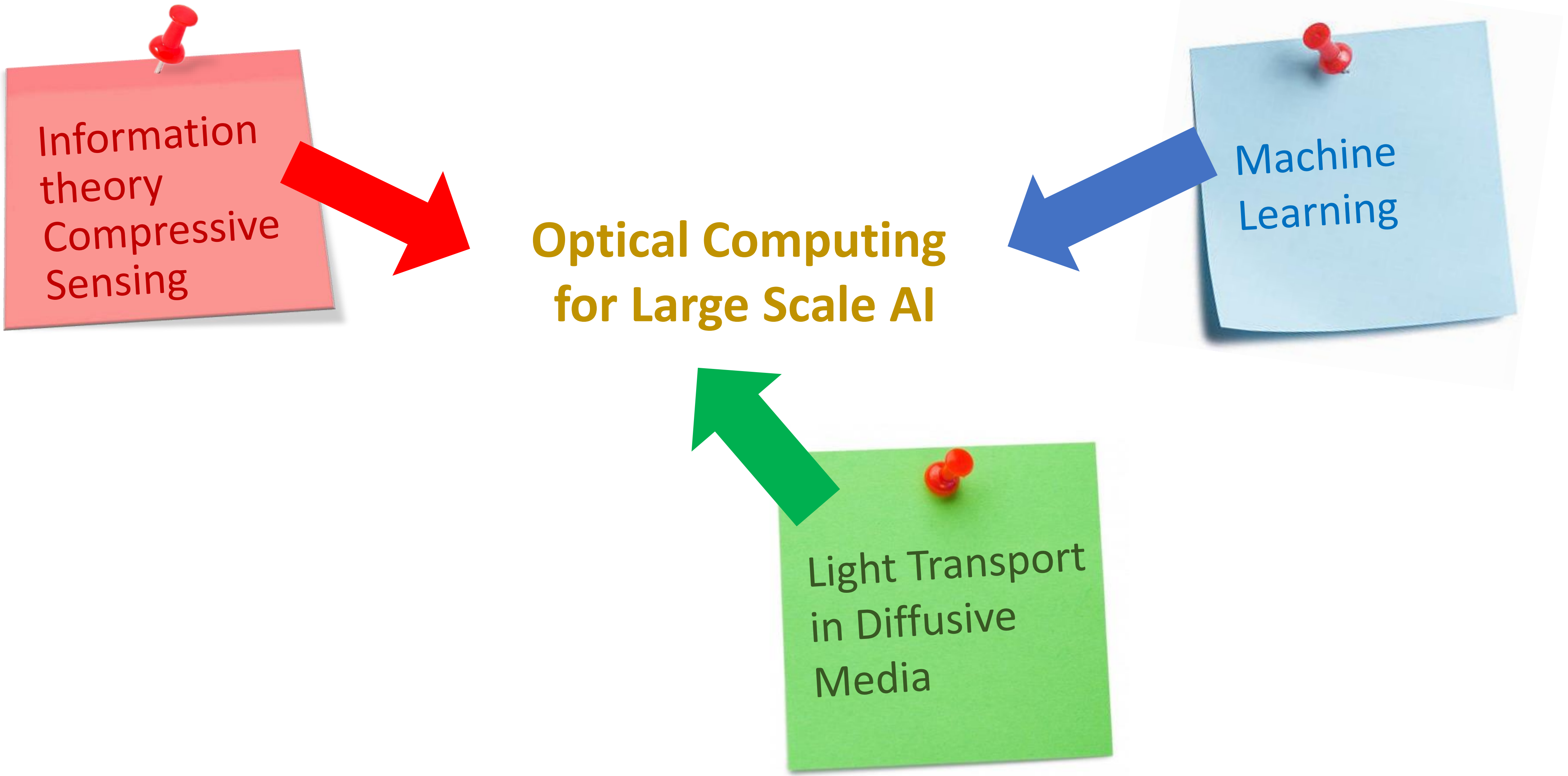
Lessons from Random Projections in Machine Learning

- Random projections act as dimensionality reduction or expansion
- They can also be seen as a dense layer in a Deep Learning model
- High dimensions reduce sensitivity to hyper-parameters
- Calibration-free

	Step 1	Random Proj.	Step 2 (Expansion)	Step 3 (WTA)
Fly olfaction	Antennae lobe 50 glomeruli	Sparse, binary Samples 6	Mushroom body 2000 Kenyon cells	APL neuron top 5%
Mouse olfaction	Olfactory bulb 1000 glomeruli	Dense, weak Samples all	Piriform cortex 100K semi-lunar cells	Layer 2A top 10%
Rat cerebellum	Pre-cerebellar nuclei	Sparse, binary Samples 4	Granule cell layer 100M granule cells	Golgi cells top 10–20%
Rat hippocampus	Entorhinal cortex 30K grid cells	Unknown	Dentate gyrus 1.2M granule cells	Hilar cells top 2%

The steps used in the fly olfactory circuit and their potential analogs in vertebrate brain regions.

The Convergence



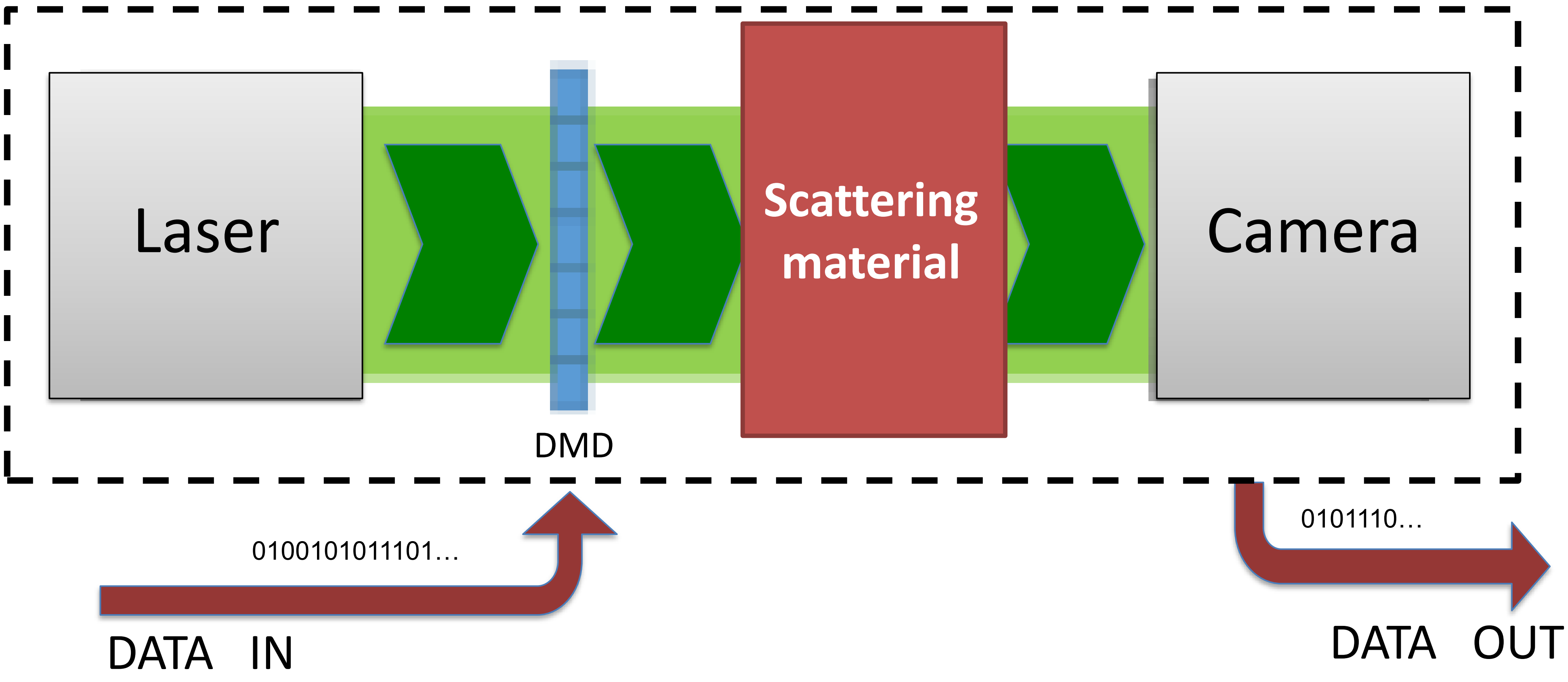
At LightOn , we bring Light to AI …

… using diffusive media as memories !



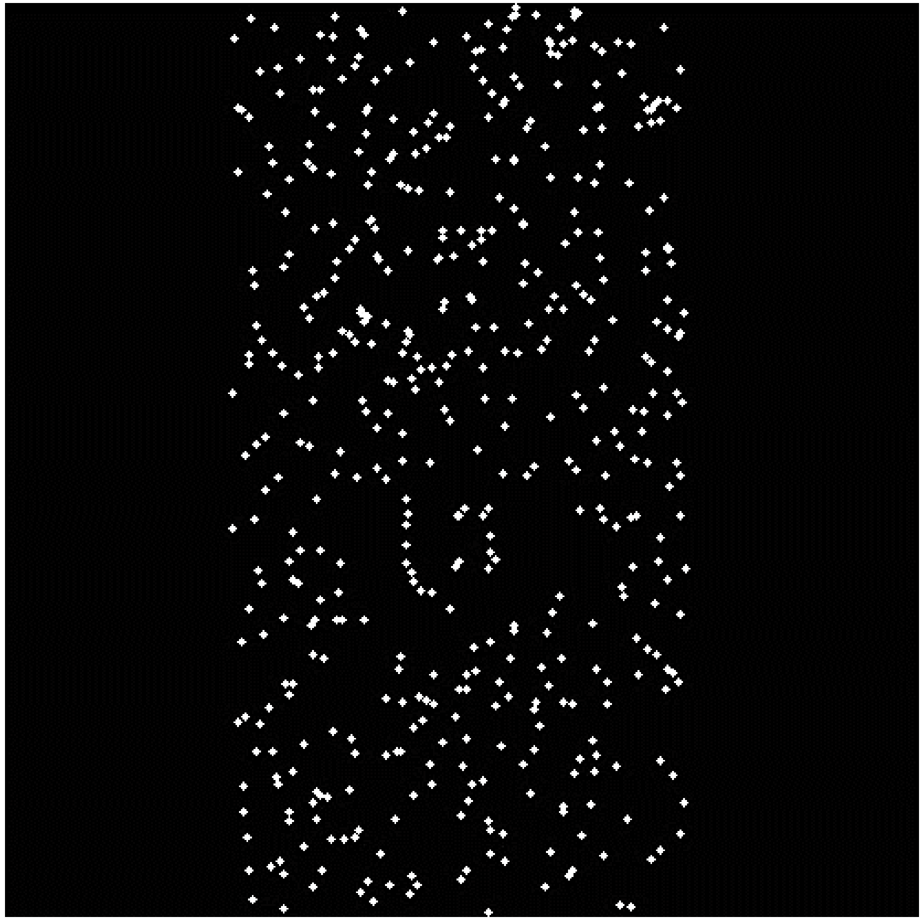
Optical Computing for Large Scale AI

LightOn
Confidential & Proprietary



Harvesting Universal Compression from Nature

Data in



Data out

Result: a **universal compression** function

Credit: Emmanuel Bossy (Université Grenoble Alpes), SimSonic Software

Optical Computing for Large Scale AI

LightOn
Confidential & Proprietary

This performs **Random Projections** in the analog domain

$$y = |Hx|^2$$

with H a complex random iid matrix

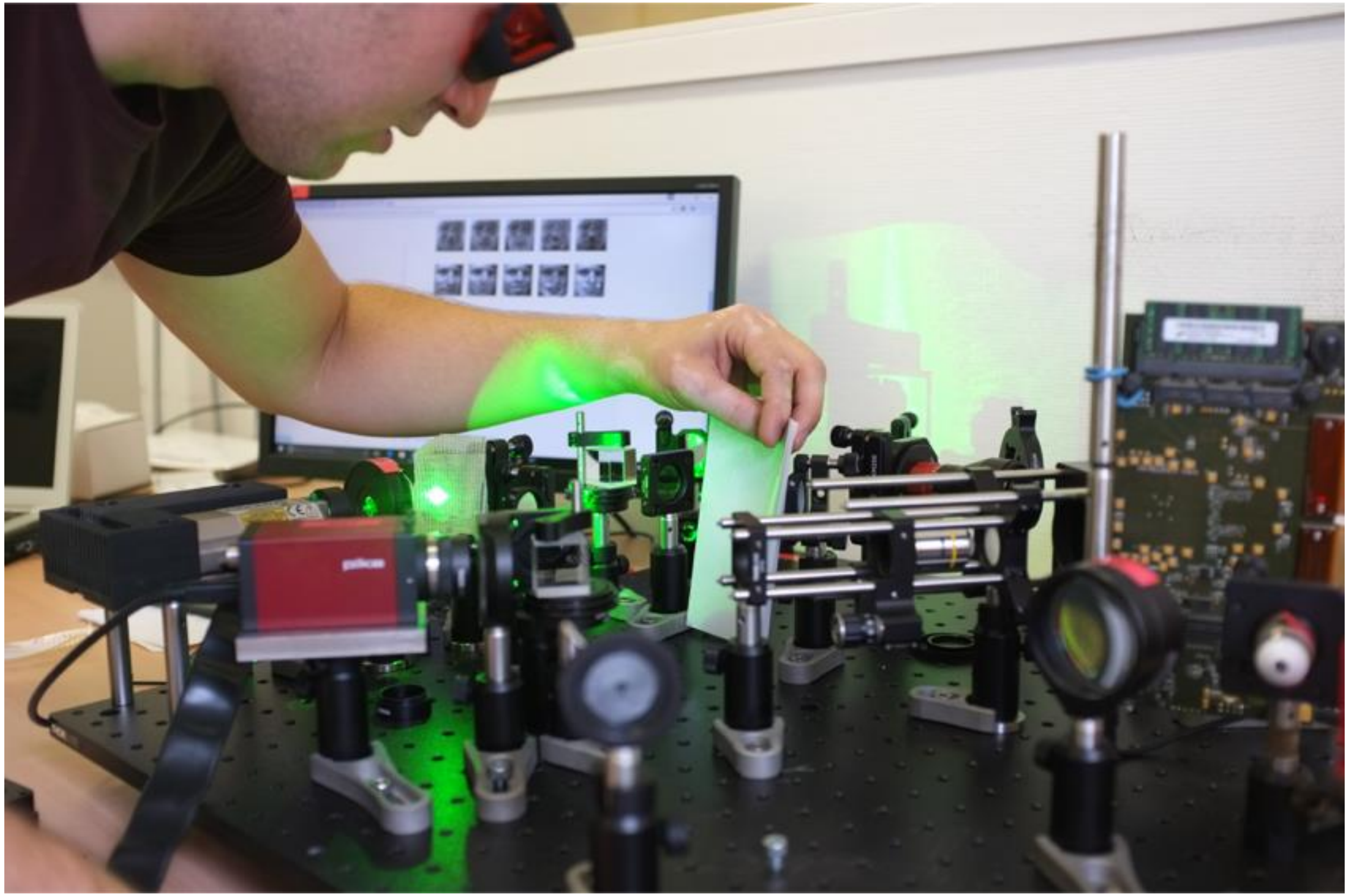
EXTRA-LARGE & SUPER-FAST

H of size higher than $10^6 \times 10^6$ (TBs of memory)

kHz operation $\rightarrow 10^3$ such multiplies / s

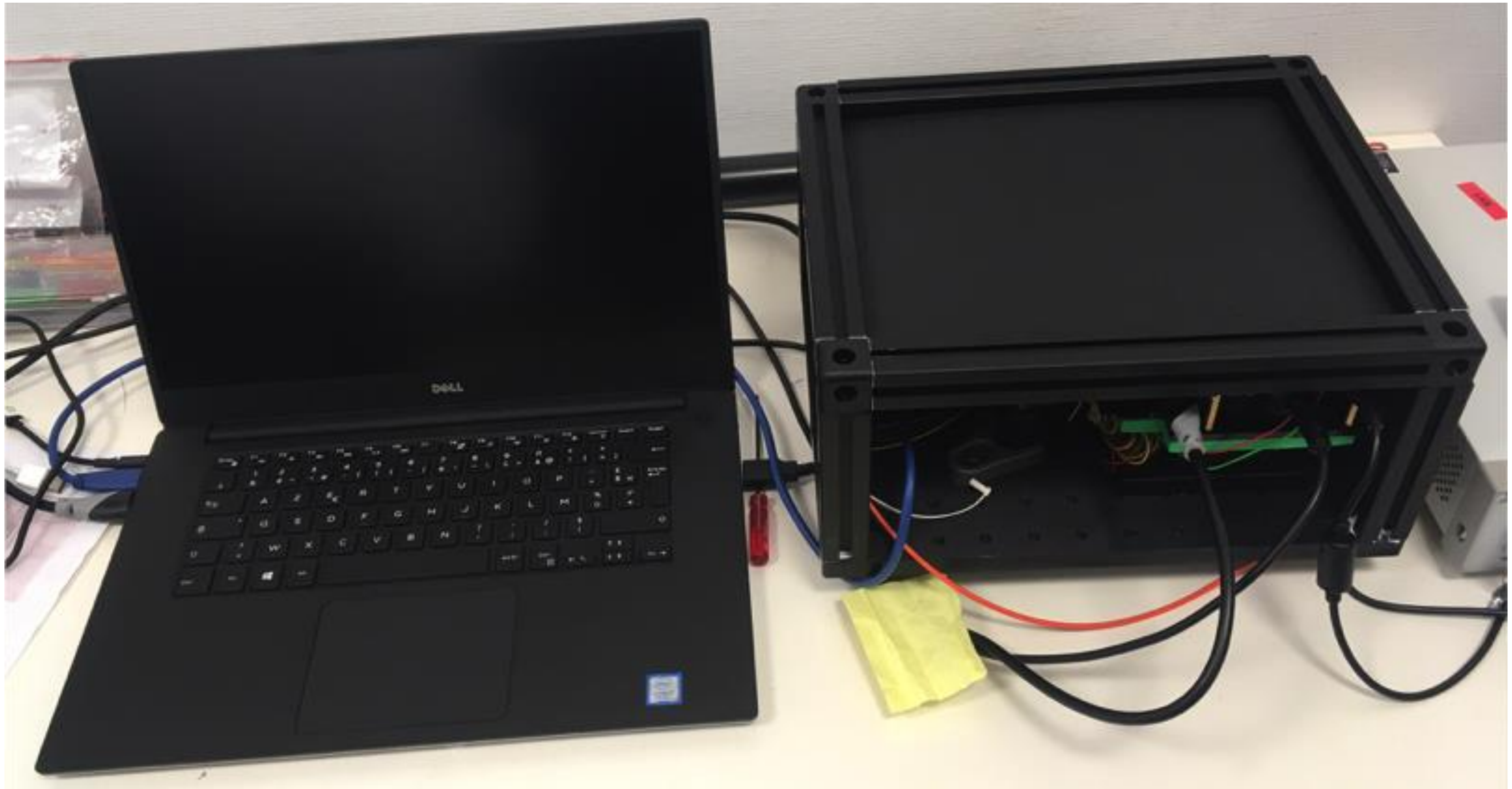
Equivalent 10^{15} operations / s : You would need a *Peta-scale* computer to do the same

Optical Computing for Large Scale AI



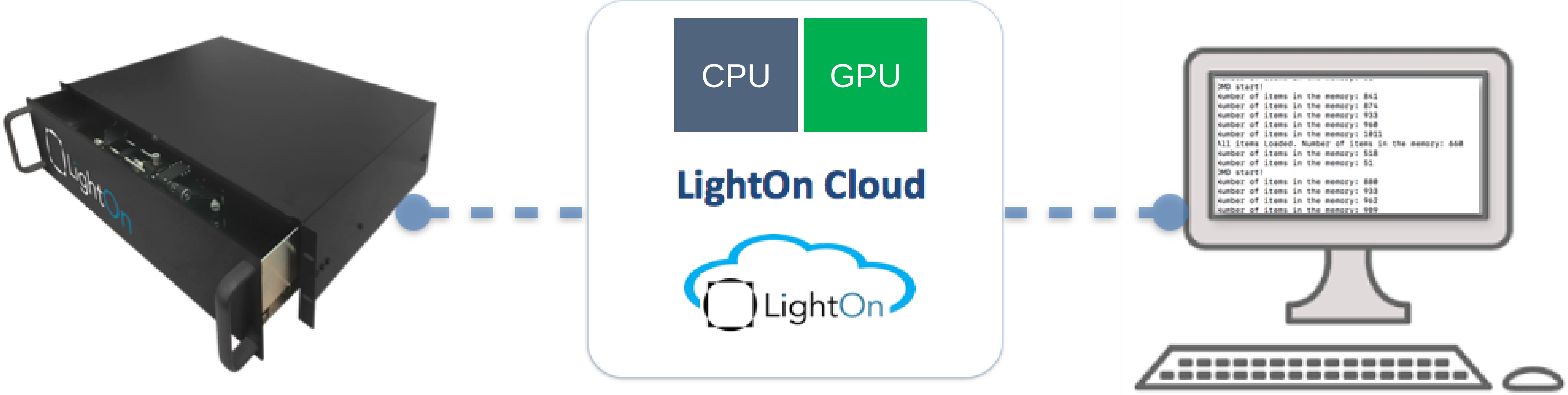
Optical Computing for Large Scale AI

LightOn
Confidential & Proprietary



Spring 2017 - The first « OPU »: Optical Processing Unit

Optical Computing for Large Scale AI

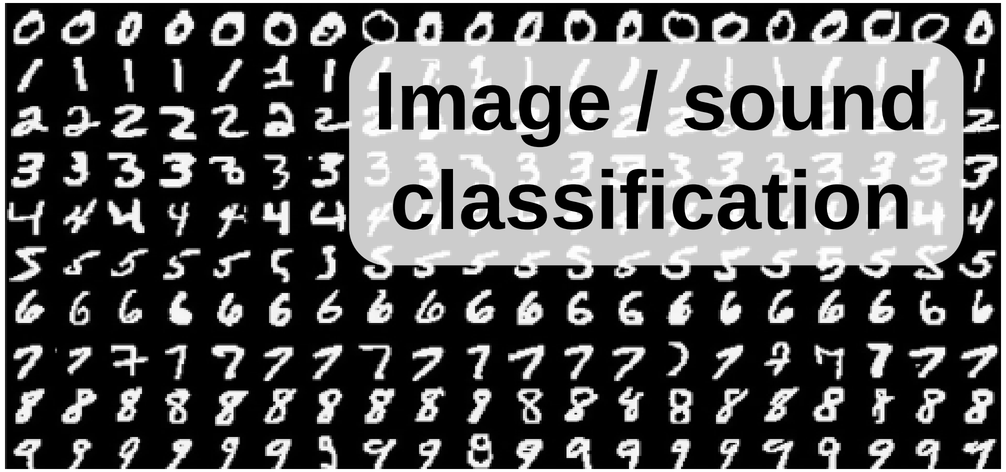


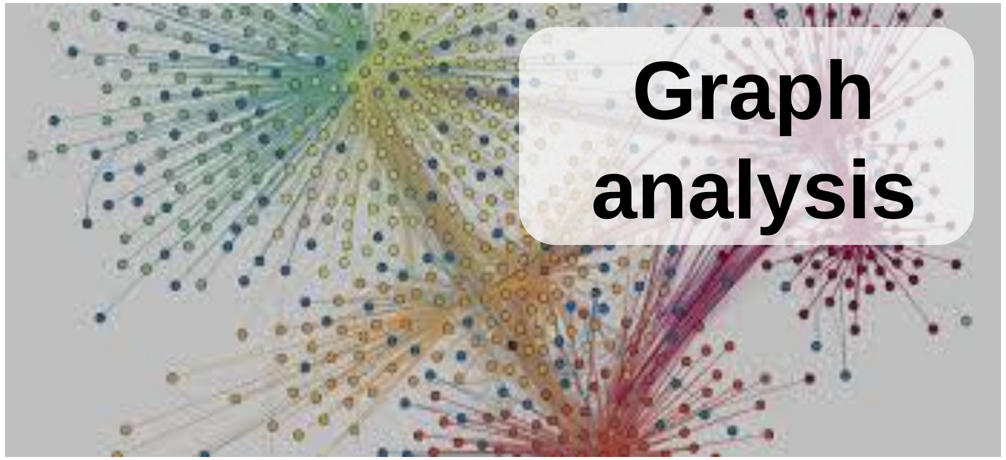
« Zeus » and « Vulcan »
OPU prototypes
30 W

OVH.com

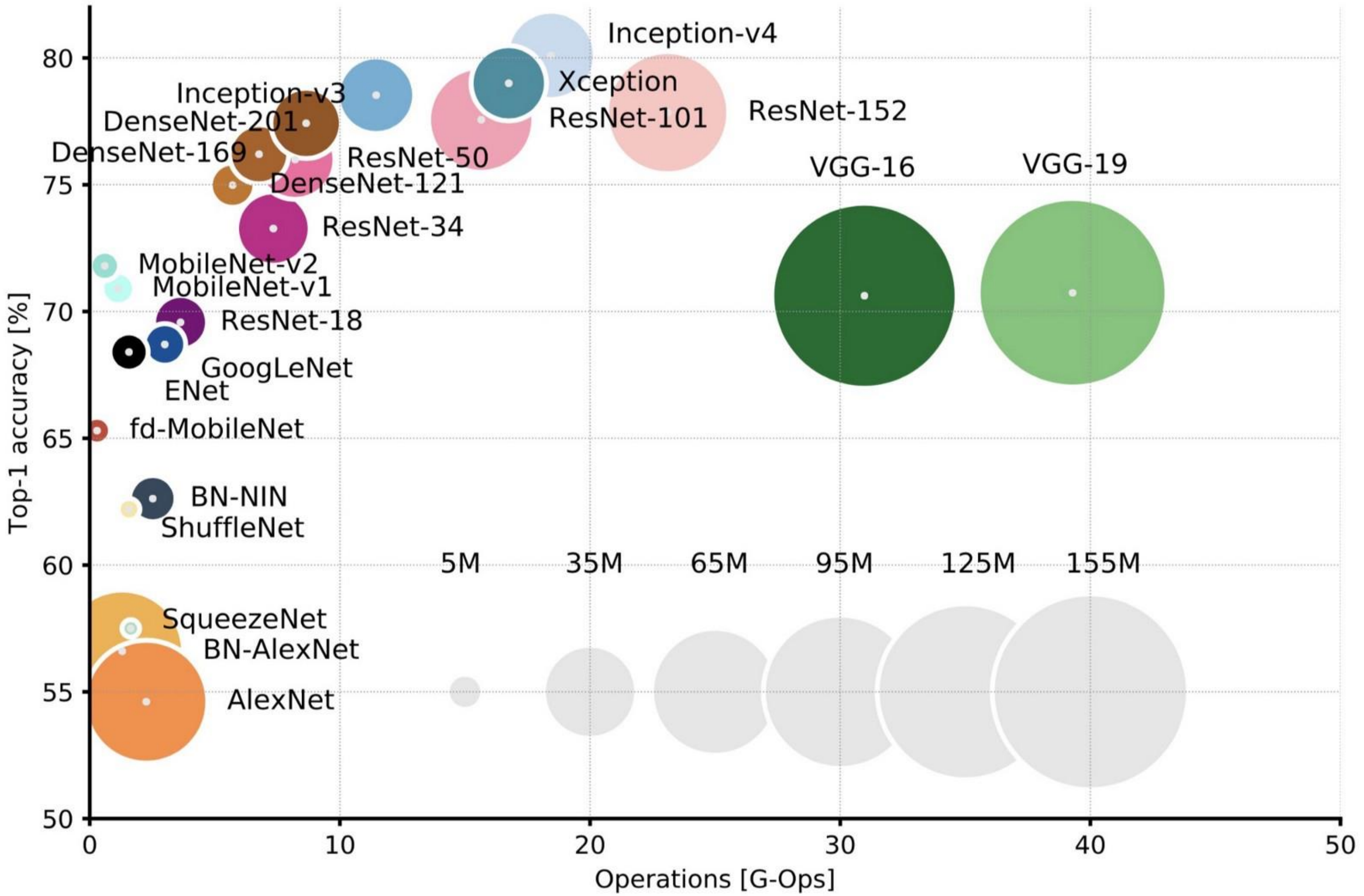
python
scikit learn
jupyter
openstack.

Some typical use cases



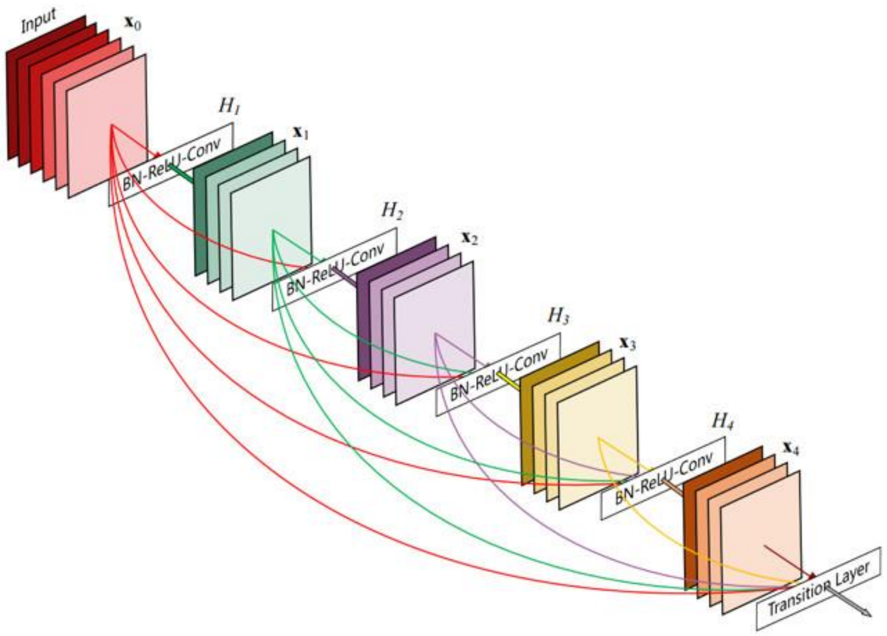


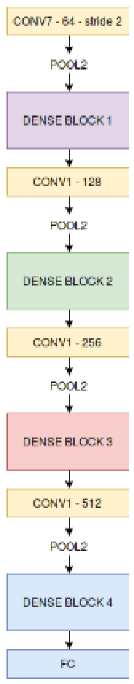
Fast Transfer Learning

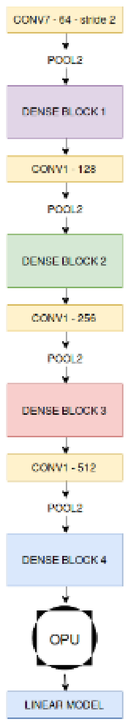


LightOn

Fast Transfer Learning







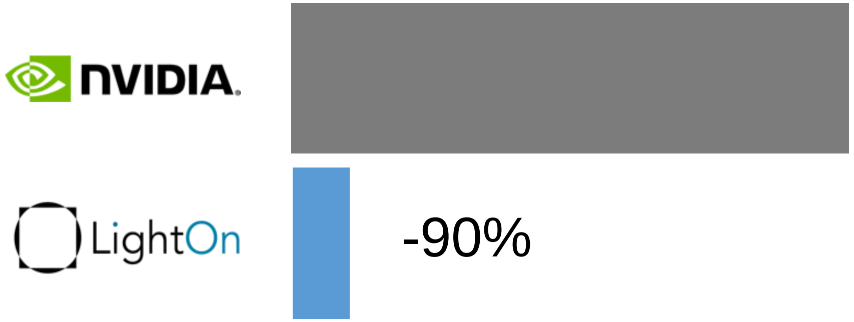
Fast Transfer Learning

From

to

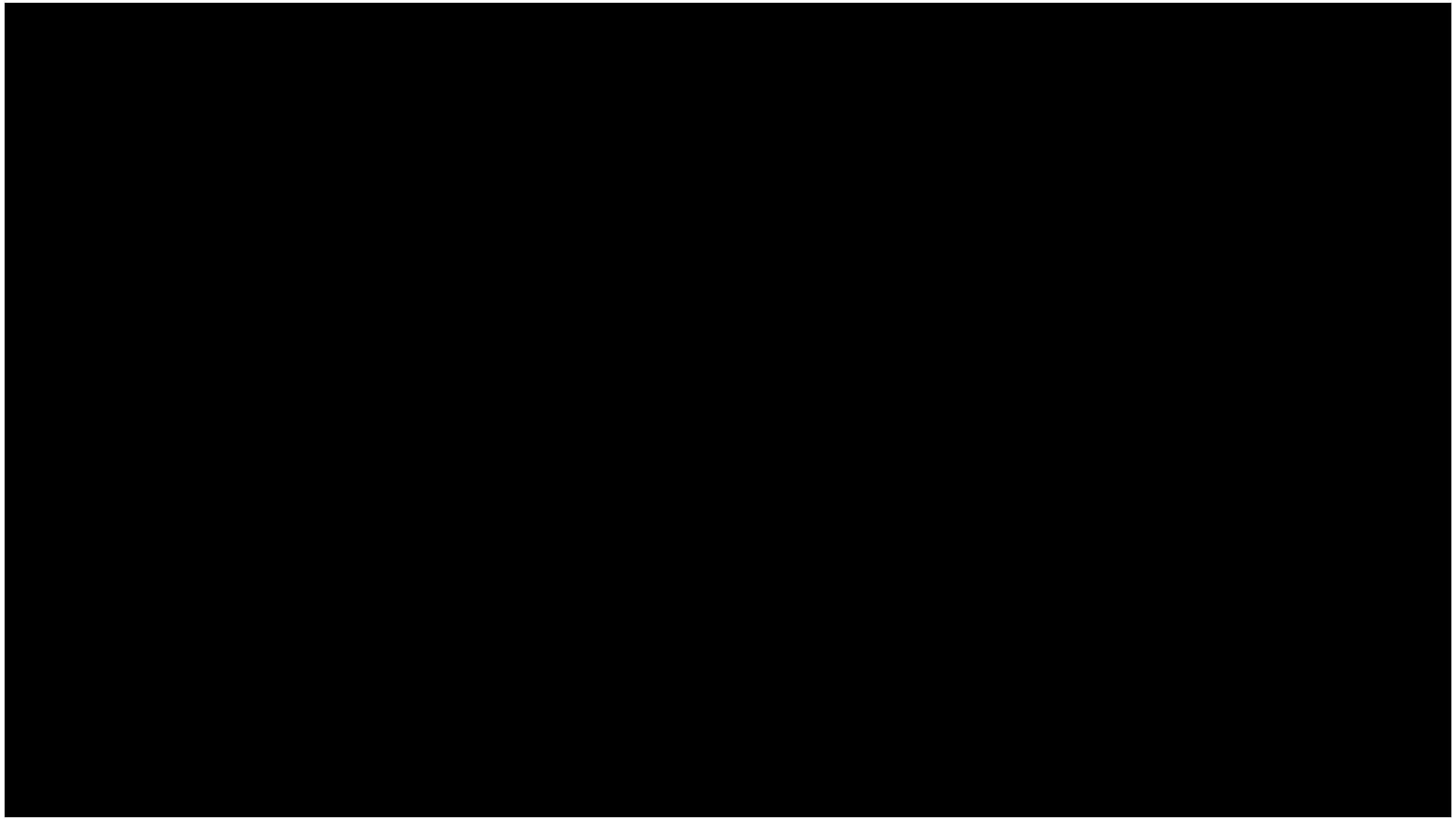
Speed

Power consumption

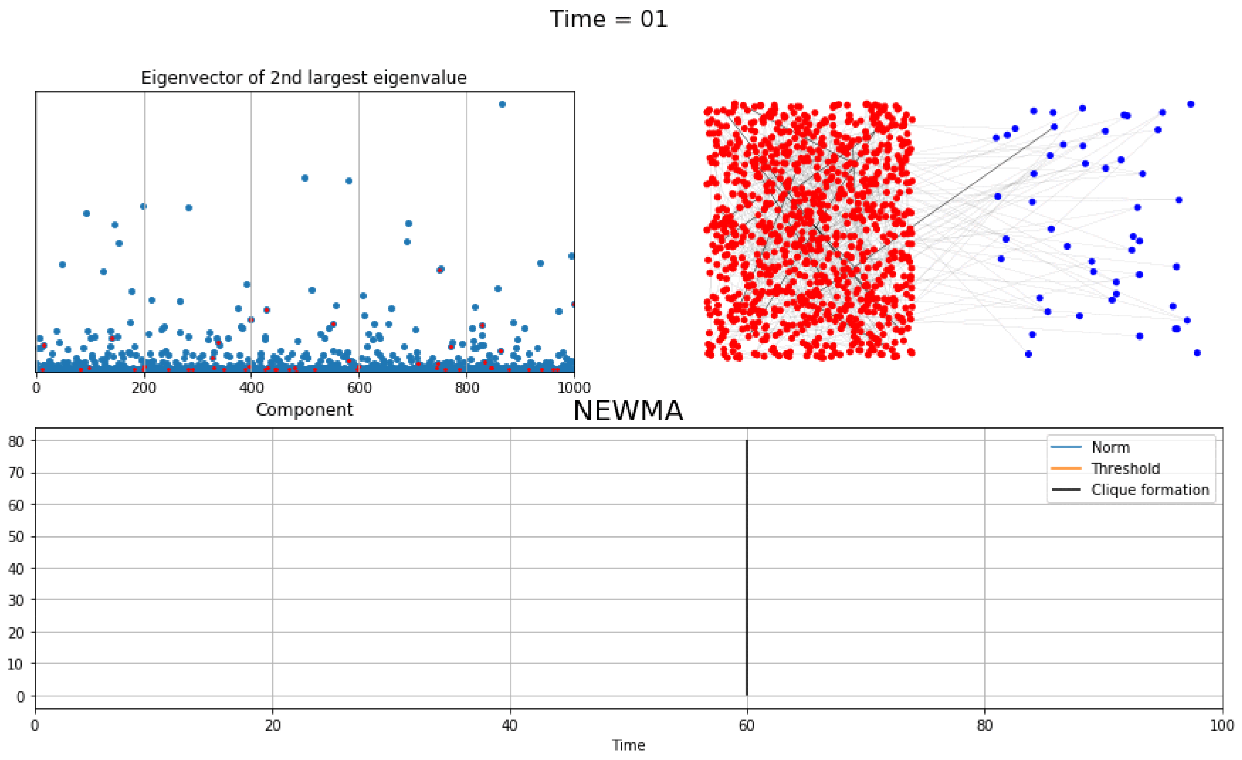


Large Scale Anomaly Detection

NEWMA: a new method for scalable model-free online change-point detection, Nicolas Keriven, Damien Garreau, Iacopo Poli, https://arxiv.org/abs/1805.08061



Clique detection on Graphs



Recommender Systems

Recommender Systems shape our lives at scale !

Recommender Systems

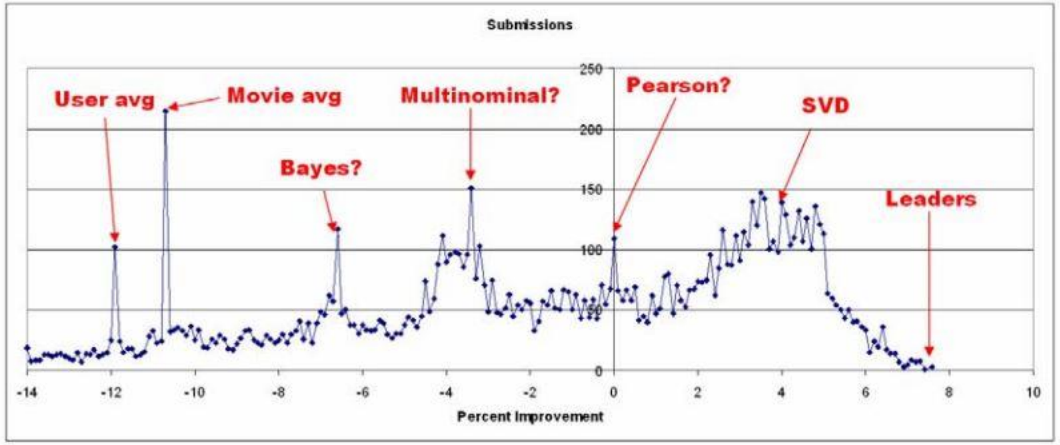
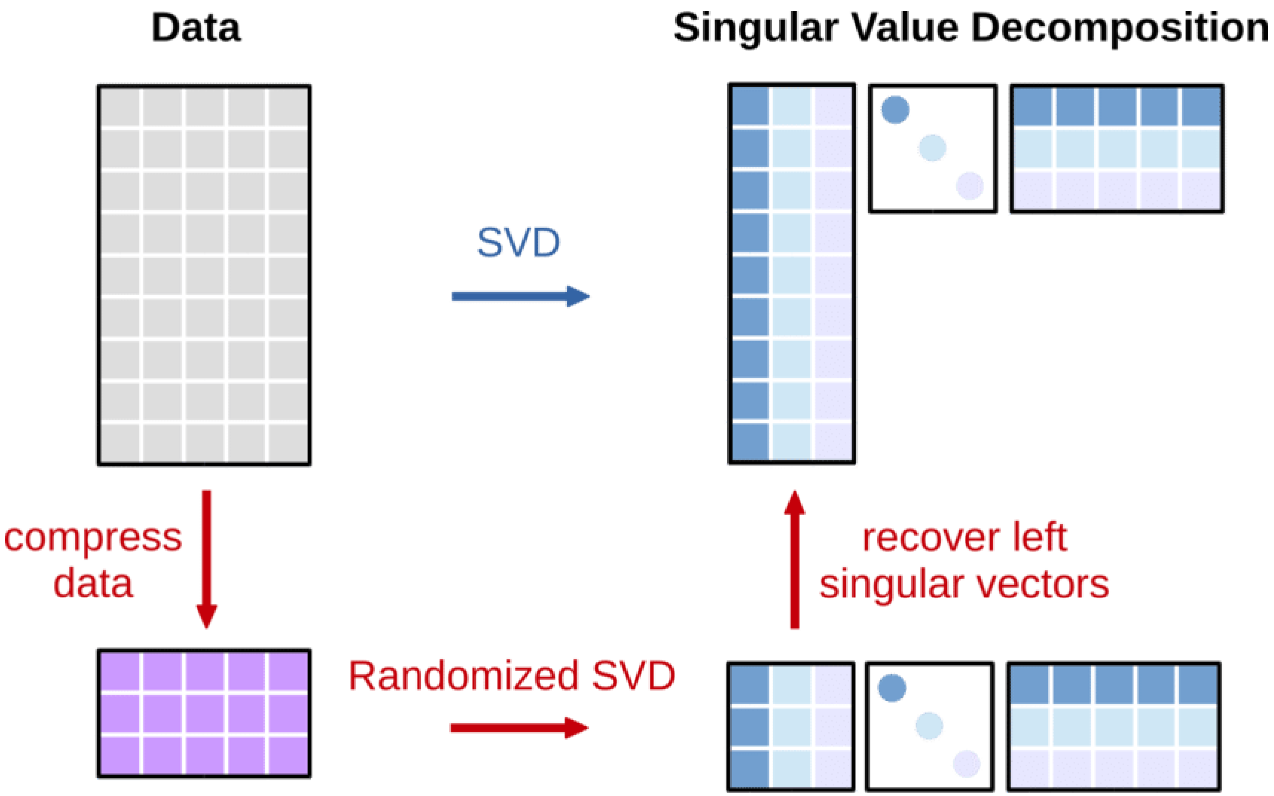
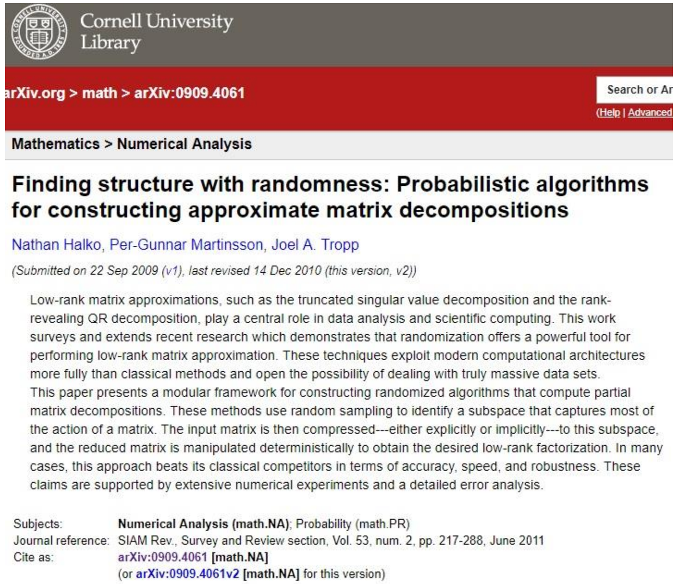


Figure 2: Detail of distribution of leading submissions indicating possible techniques

The Netflix Prize: Crowdsourcing to Improve DVD Recommendations, https://digit.hbs.org/submission/the-netflix-prize-crowdsourcing-to-improve-dvd-recommendations/

Recommender Systems




Cornell University Library

arXiv.org > math > arXiv:0909.4061

Search or Ar

(Help | Advanced

Mathematics > Numerical Analysis

Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions

Nathan Halko, Per-Gunnar Martinsson, Joel A. Tropp

(Submitted on 22 Sep 2009 (v1), last revised 14 Dec 2010 (this version, v2))

Low-rank matrix approximations, such as the truncated singular value decomposition and the rank-revealing QR decomposition, play a central role in data analysis and scientific computing. This work surveys and extends recent research which demonstrates that randomization offers a powerful tool for performing low-rank matrix approximation. These techniques exploit modern computational architectures more fully than classical methods and open the possibility of dealing with truly massive data sets.
This paper presents a modular framework for constructing randomized algorithms that compute partial matrix decompositions. These methods use random sampling to identify a subspace that captures most of the action of a matrix. The input matrix is then compressed---either explicitly or implicitly---to this subspace, and the reduced matrix is manipulated deterministically to obtain the desired low-rank factorization. In many cases, this approach beats its classical competitors in terms of accuracy, speed, and robustness. These claims are supported by extensive numerical experiments and a detailed error analysis.

Subjects: **Numerical Analysis (math.NA)**; Probability (math.PR)
Journal reference: SIAM Rev., Survey and Review section, Vol. 53, num. 2, pp. 217-288, June 2011
Cite as: arXiv:0909.4061 **[math.NA]**
(or **arXiv:0909.4061v2 [math.NA]** for this version)

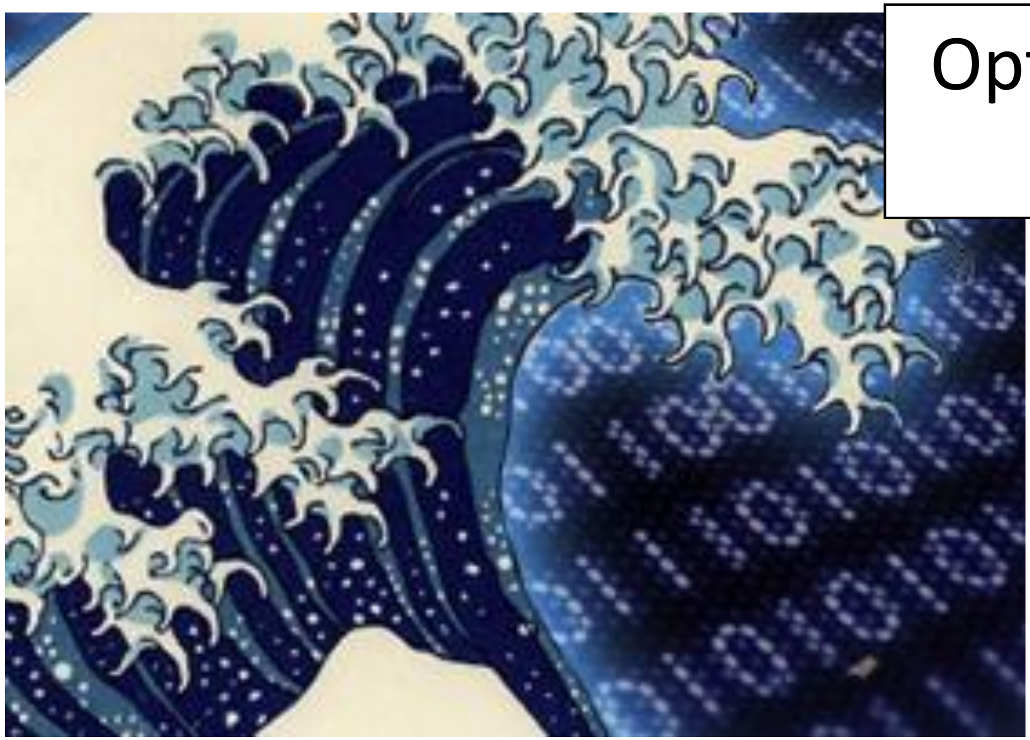
- Randomized Matrix Decompositions using R, Aug 2016, N. Benjamin Erichson, Sergey Voronin, Steven L. Brunton, J. Nathan Kutz

Recommender Systems

- Movielens with 20 millions records (size 26.000 x 140.000) with 0.5% non-zero entries
- At the moment, our OPU is competitive with Facebook-PCA approach (efficient randomized SVD CPU implementation)

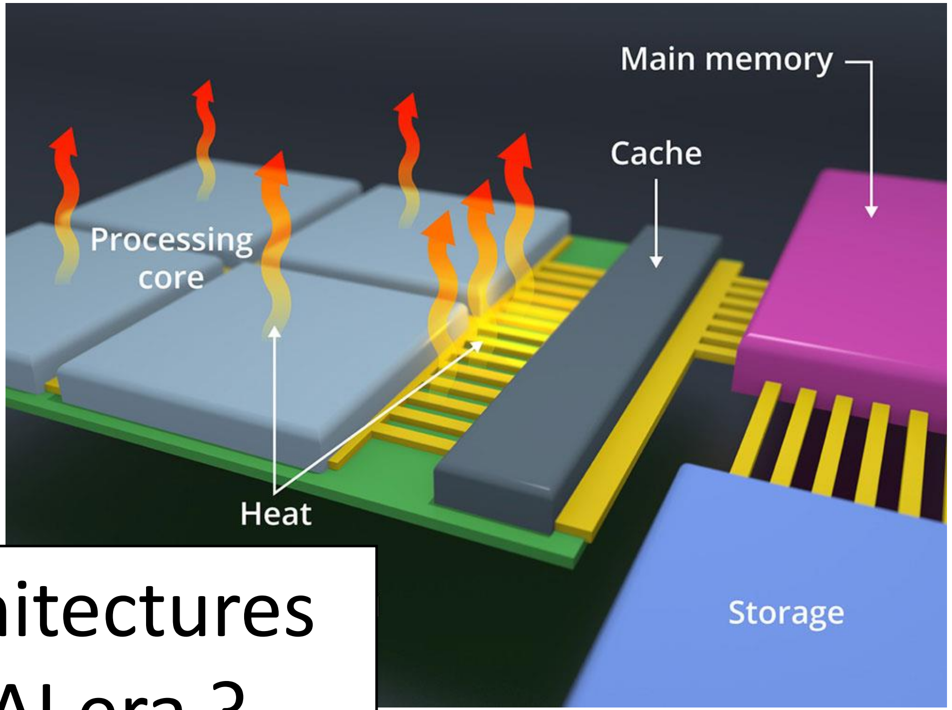
One statement and one question

LightOn
Confidential & Proprietary



https://blog.apterainc.com/the-data-tsunami-is-coming.-is-your-company-ready

Optical Computing is already **at scale** for the data tsunami



Will Von Neumann architectures stay prevalent in the AI era ?

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

Just try it !

LightOn Cloud

- Available for beta-users Q1 2019 (VMs via OpenStack)
- **Platform-as-a-Service** Integration within popular ML frameworks (Python-based: Scikit-Learn, TensorFlow to be supported ...)

Sign-up: lighton.io

Igor Carron

igor@lighton.io

http://Lighton.io

LightOn

Using Light to change the Future of Computing