

Optical Computing redux: Exploiting Light Scattering for Artificial Intelligence

Sylvain GIGAN

WICOM

June 11, 2025

ALUMNI

PhDs:

S.Popoff
M. Mounaix
D.Andreoli
B. Blochet
P.Bondareff
T. Sperber
T.Chaigne
J. Guilbert
H.Defienne
A. Boniface
S.Leedumrong-
J. Dong
L. Devaud
H. Wang
F. Mazzoncini

Postdocs:

P. Boucher
M. Dabrowski
D.Martina
C. Moretti
R. Savo
F. Soldevila
G. Volpe
L. Valzania
J.Bertolotti
J. Hu
O.Katz
C. Rimoli
T. Juffman
L. Valzania
R. Pandya
G. Iaccuci

PIs:

Pr. Sylvain GIGAN
Dr. Hilton BARBOSA DE
AGUIAR

PhD Students

Louis DELLOYE
Léa CHIBANI
Adrian MAKOWSKI
Baptiste COURME
Ziao WANG
Hankun YANG
Amol MAHURKAR
Malo JOLY

Postdocs

Dr. YoonSeok BAEK
Dr. Fei XIA
Dr. Clémence GENTNER
Dr. Vivek KUMAR
Dr. Xue DONG
Dr. Shupeng ZHAO
Dr. Andrew VOITIV
Dr. Raouf AMARA

Engineers

Dariusz NITA
Nassredine AMICHE

The Team



Our Goal :
Understand and Harness
light propagation in complex media

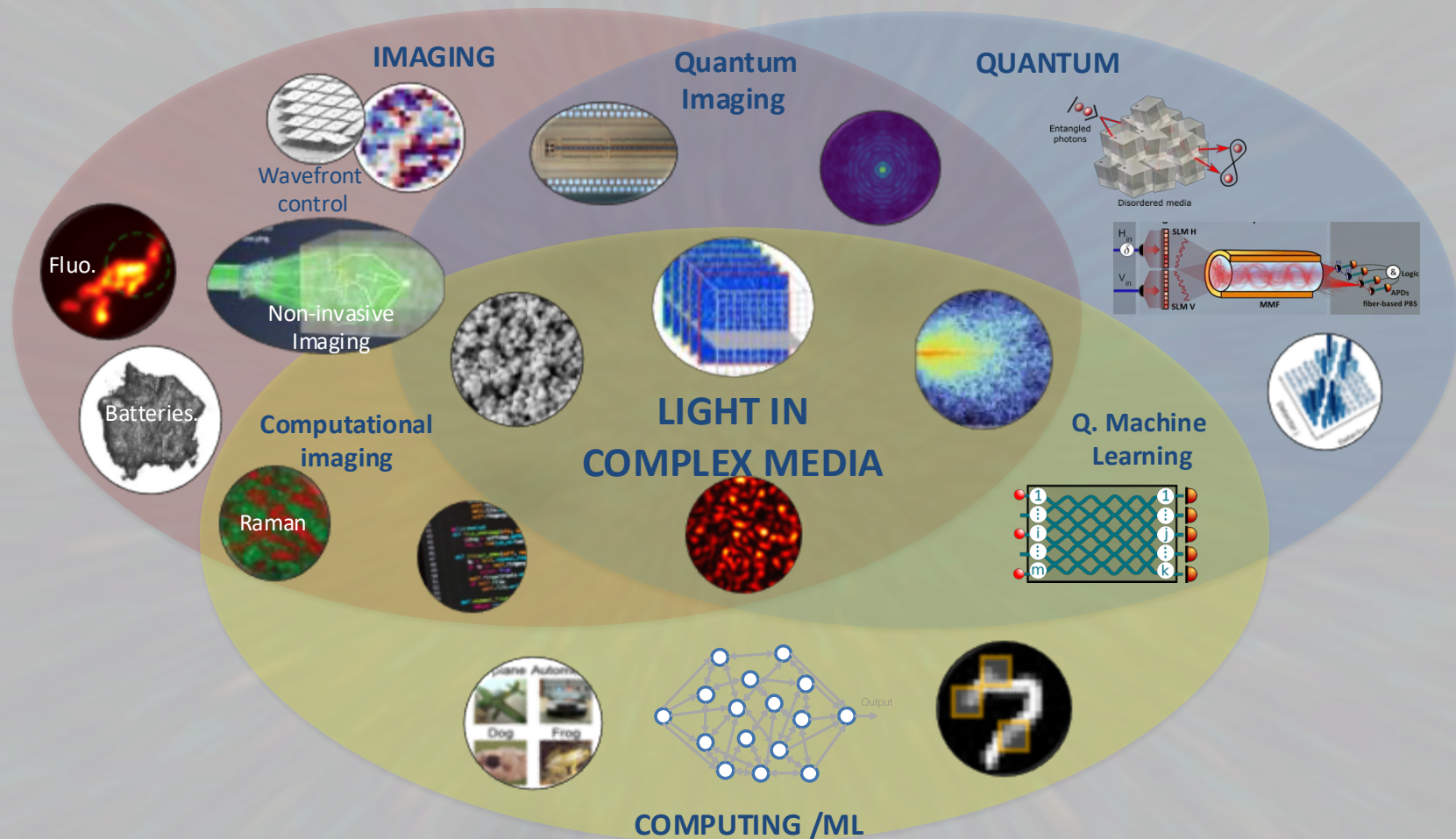
COLLABORATORS

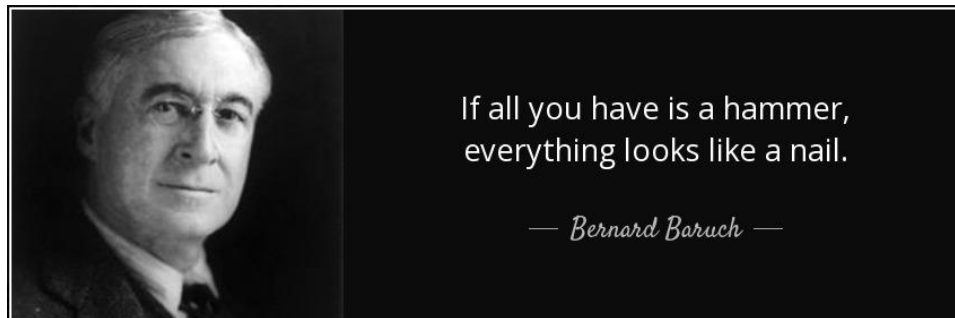
E. Charbon (EPFL)
P. Del Hougne (U. Rennes)
H. Defienne (CNRS)
H. Cao (Yale)
S. Rotter (TU Wien)
L. Bourdieu, C. Ventalon (IBENS)
F. Krzakala (EPFL)
S. Brasselet (Institut Fresnel)
C. Conti (U. Sapienza)
D. Brunner (Femto-ST)
R. Grange (ETHz)
P. Mounaix (CNRS)
R. Lapkiewicz (U. Warsaw)
R. Fickler (U. Tampere)
M. Paternostro (U. Palermo)
A. Vaziri (Rockefeller)
J.M Tarason (CdF)
P. Del Hougne (U. Rennes)
M. Guillon (U. Paris Cité)

Funding



Our Goal : Understand and Harness light propagation in complex media



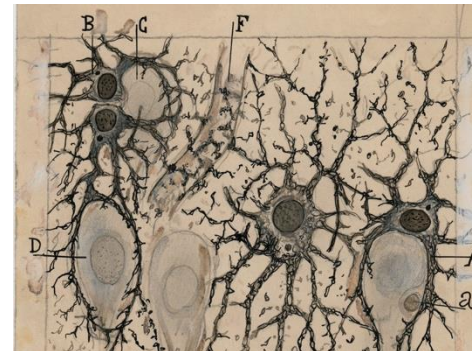
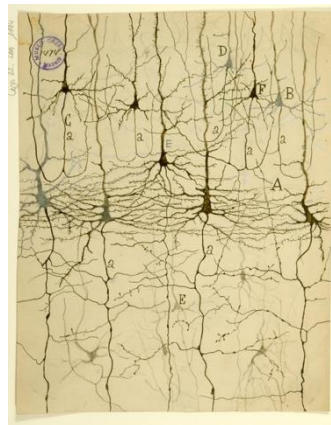


*« ...how hardware chooses which ideas succeed and which fail. »
(and vice-versa)*

advocates for « (...) joint collaboration between hardware,
software and machine learning communities. »

<https://hardwarelottery.github.io>
[arXiv:2009.06489](https://arxiv.org/abs/2009.06489)

The Human Brain

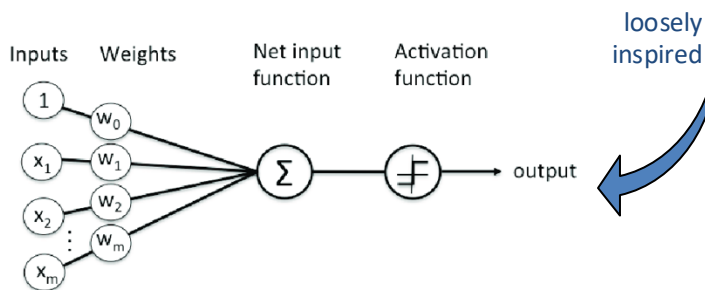
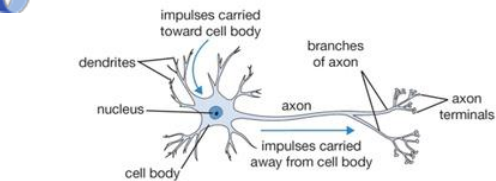


Ramon y Cajal (drawings)

- 10^{11} neurons
- 10^{15} synapses
- 10^4 synapses per neuron
- « Slow » millisecond timescale
 - ❖ 30 « Peta-Ops » per second
 - ❖ consume ~20 Watts
 - ❖ learns by itself



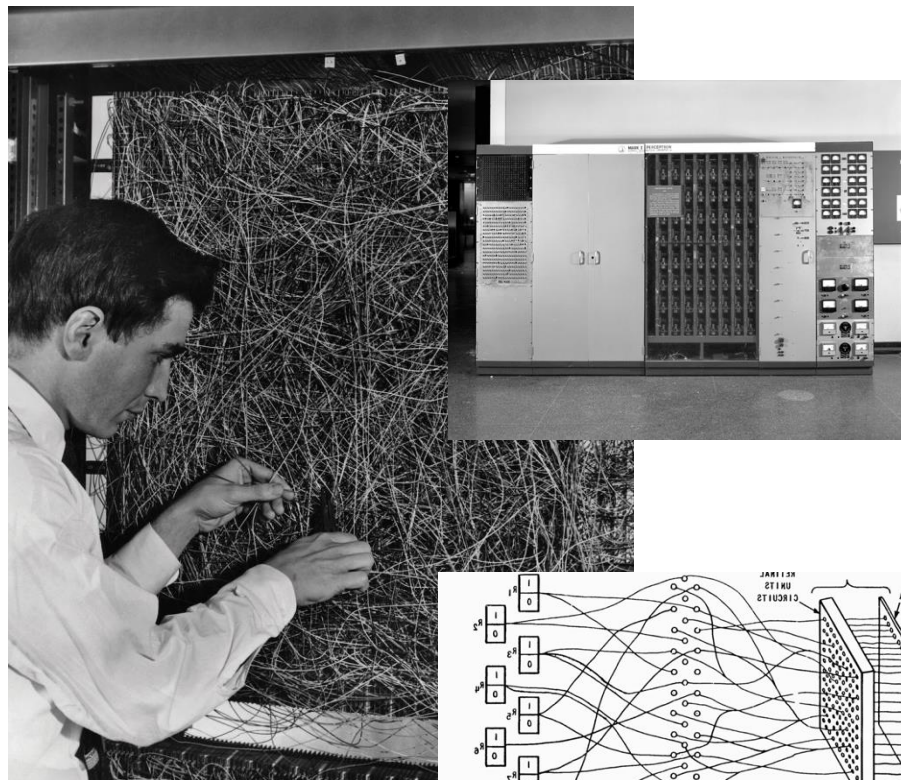
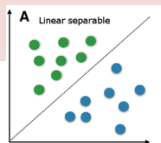
Video: E. Beaurepaire team (LOB)



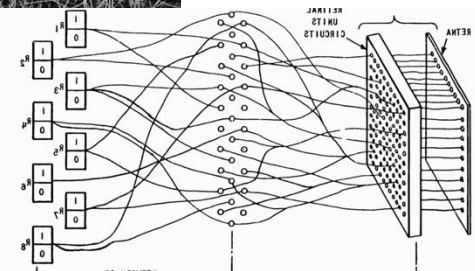
$$y = \varphi\left(\sum_{i=1}^n w_i x_i + b\right) = \varphi(\mathbf{w}^T \mathbf{x} + b)$$

- Multiply-and-Accumulate (MAC)
- Tunable weights
- Non-linear activation function

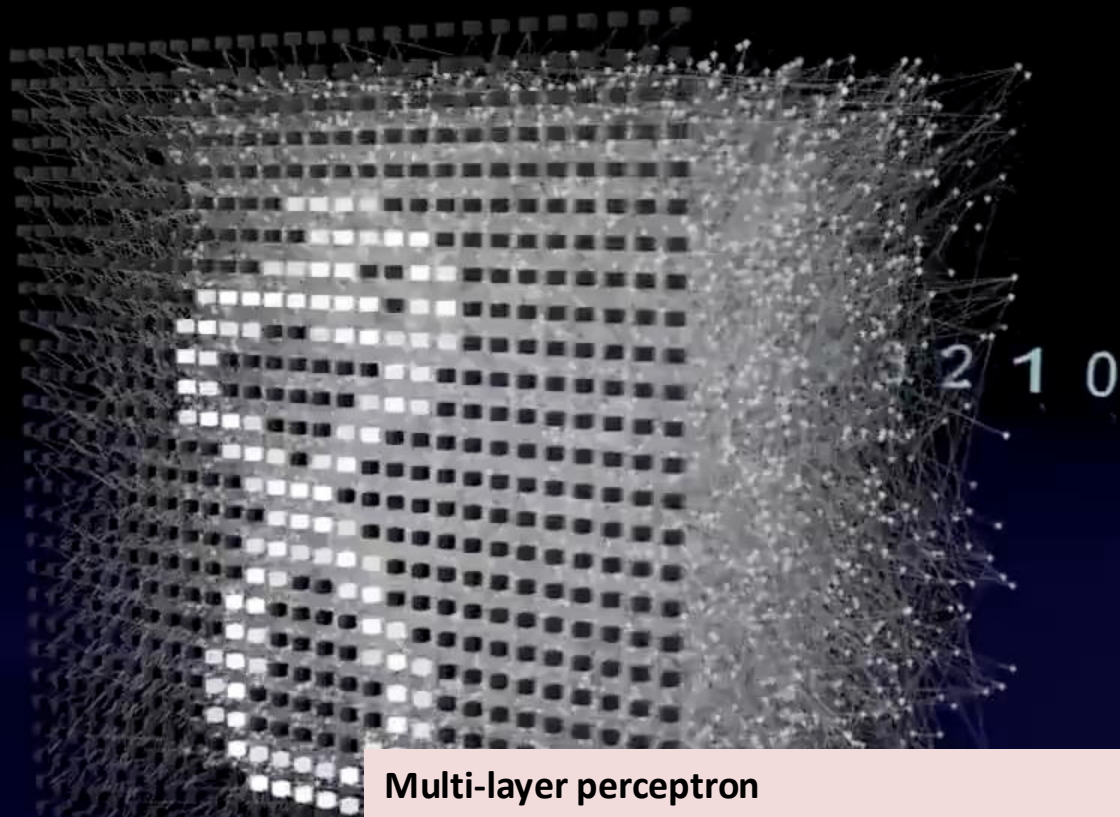
❖ Linear classifier



The Mark I perceptron
(Frank Rosenblatt – 1960')



Type: ML Perceptron
Data Set: MNIST
Hidden Layers: 3
Hidden Neurons: 10000
Synapses: 24664180
Synapses shown: 2%
Learning: BP



Multi-layer perceptron

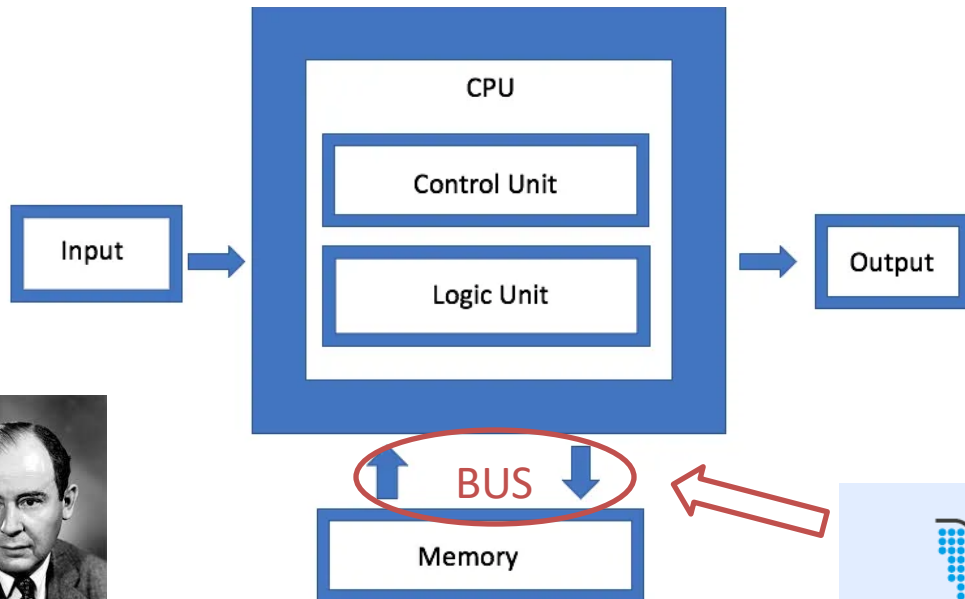
- All-to-all connectivity in each layer
- adjustable weights (matrix multiplication)
- Non-linear activation function at each (hidden) layer
- Non-linear classifier

PetaFlop/s.days





John VON NEUMANN
(1903-1957)



CPU
Desktop PC
~100 GigaFLOPS
300W



GPU
NVIDIA A100
~10 TeraFLOPS
300W



TPU (Google)
~4 T-Ops
75W

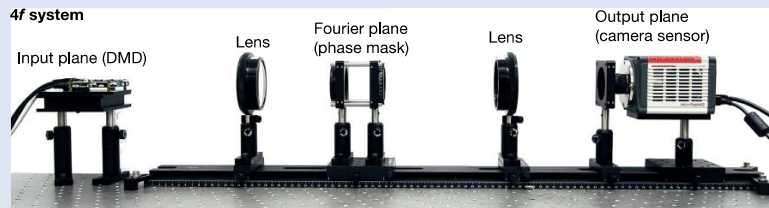
Versatile But not adapted for AI : slow, power inefficient

Optics has distinct advantages...

- Low energy consumption
- Easy interconnect – Multiplexing
- Low latency + blazing speed

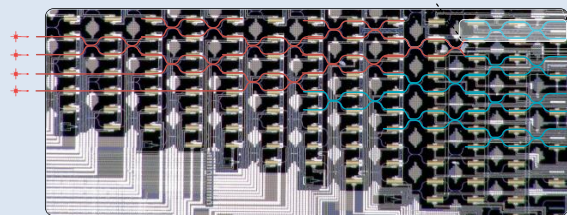
... and some disadvantages:

- Bulky
- Tricky non-linearities
- Storage



Sci. Rep. **8**, 12324 (2018).

Free space
VS
Integrated optics



Sci. Adv. **5**, eaay6946 (2019).

Nature Photonics 15, 10 (2021)

Photonics for artificial intelligence and neuromorphic computing

REVIEW ARTICLE | FOCUS
<https://doi.org/10.1038/s41566-020-00754-y>

Bhavin J. Shastri^{1,2,7}, Alexander N. Tait^{2,3,7}, T. Ferreira de Lima², Wolfram H. P. Pernice⁴, Harish Bhaskaran⁵, C. D. Wright⁶ and Paul R. Prucnal²

Perspective

Nature 588, 39 (2020)

Inference in artificial intelligence with deep optics and photonics

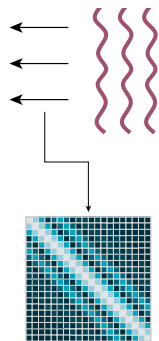
<https://doi.org/10.1038/s41586-020-2973-6>

Received: 28 November 2019

Gordon Wetzstein^{1,8}, Aydogan Ozcan², Sylvain Gigan³, Shanhui Fan¹, Dirk Englund⁴, Marin Soljačić⁴, Cornelia Denz², David A. B. Miller¹ & Demetri Psaltis⁴

Basic building blocks:

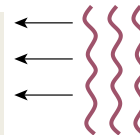
Free space



Thin mask



layered scatterer

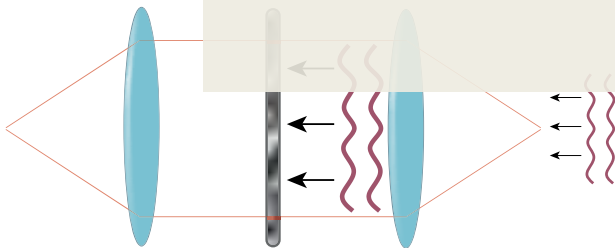


Good for **large scale** single-layer Neural Networks

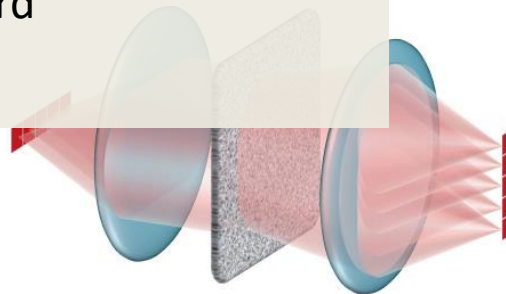
All-optical **deep** neural networks :
cascaded **non-linearities** remains **challenging**

Training is hard

Advanced functions

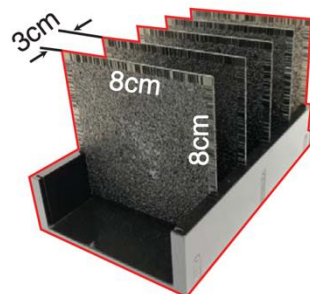
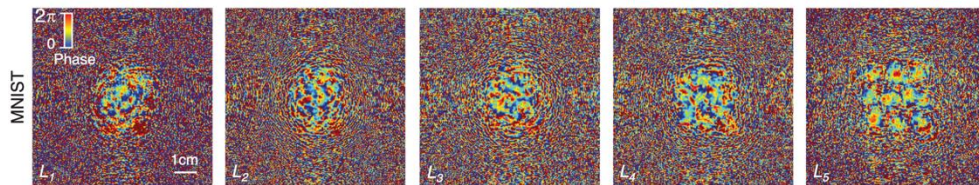
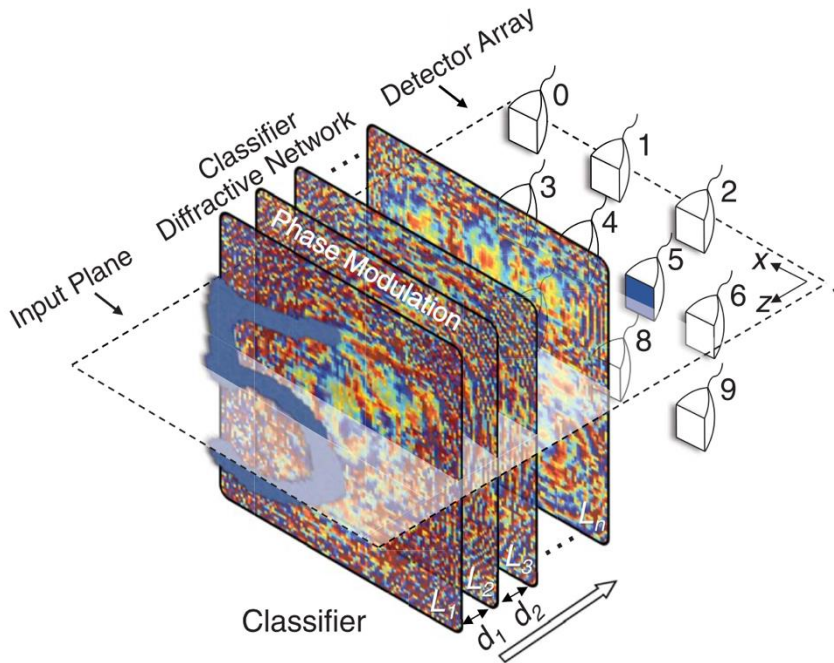


2D convolutions



1D vector-matrix multiplier

IDEA : fabricate and stack masks



3D Printed D²NN
(Classifier)

THz realization

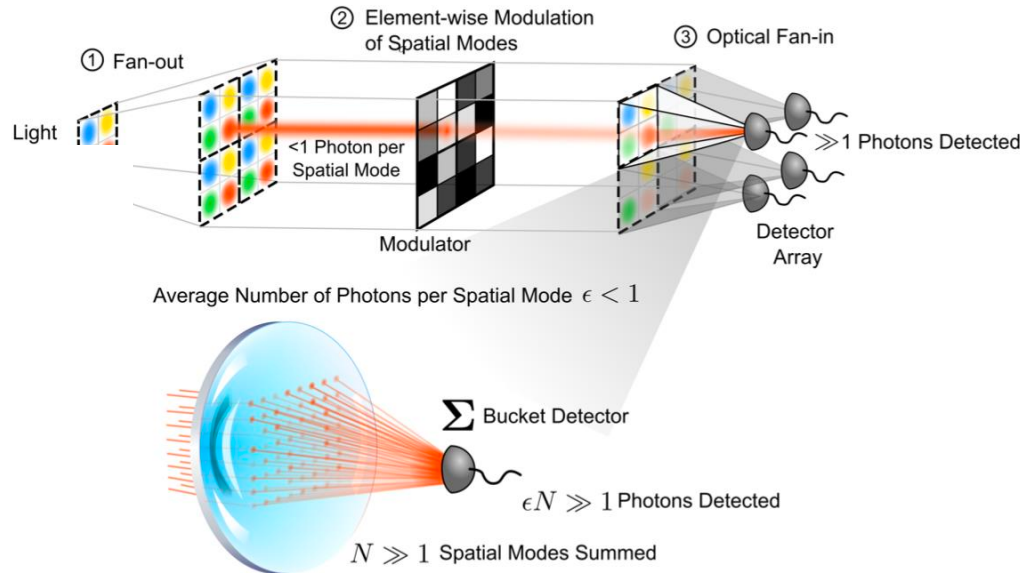
Multilayer : « **deep** »
no non-linearity between layers:
 It's a just a dense single NN layer!
 « **diffractive** »

Article | Open Access | Published: 10 January 2022

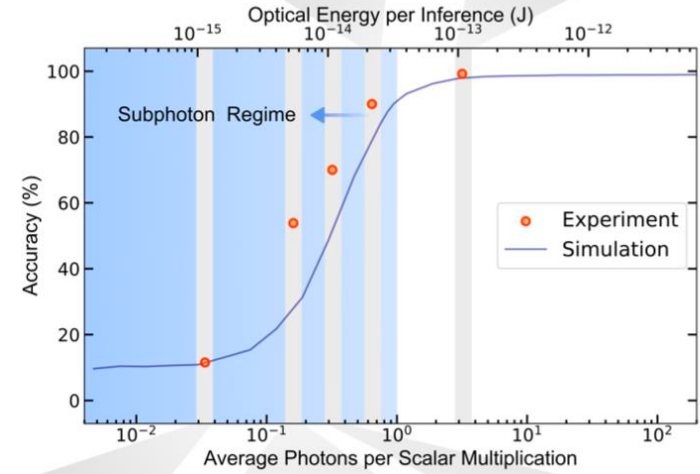
An optical neural network using less than 1 photon per multiplication

Tianyu Wang , Shi-Yuan Ma, Logan G. Wright, Tatsuhiro Onodera, Brian C. Richard & Peter L. McMahon 

Nature Communications **13**, Article number: 123 (2022) | [Cite this article](#)



Accuracy on MNIST:
90% with < 1 (detected) photon/MAC



September 2022

nature physics insight

Complex optics

Review Article | [Published: 08 September 2022](#)

Shaping the propagation of light in complex media

[Hui Cao](#) ✉, [Allard Pieter Mosk](#) & [Stefan Rotter](#)

[Nature Physics](#) **18**, 994–1007 (2022) | [Cite this article](#)

Review Article | [Published: 08 September 2022](#)

Physics of highly multimode nonlinear optical systems

[Logan G. Wright](#), [Fan O. Wu](#), [Demetrios N. Christodoulides](#) & [Frank W. Wise](#) ✉

Perspective | [Published: 08 September 2022](#)

Quantum light in complex media and its applications

[Ohad Lib](#) & [Yaron Bromberg](#) ✉

[Nature Physics](#) **18**, 986–993 (2022) | [Cite this article](#)

Comment | [Published: 08 September 2022](#)

Controlling random lasing action

[Riccardo Sapienza](#) ✉

[Nature Physics](#) **18**, 976–979 (2022) | [Cite this article](#)

Review Article | [Published: 08 September 2022](#)

Imaging in complex media

[Jacopo Bertolotti](#) ✉ & [Ori Katz](#) ✉

[Nature Physics](#) **18**, 1008–1017 (2022) | [Cite this article](#)

Perspective | [Published: 08 September 2022](#)

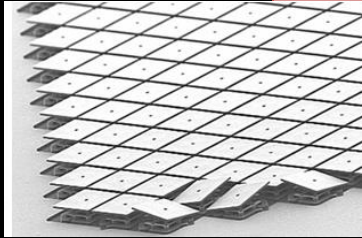
Imaging and computing with disorder

[Sylvain Gigan](#) ✉

[Nature Physics](#) **18**, 980–985 (2022) | [Cite this article](#)

LASER





MODULATORS

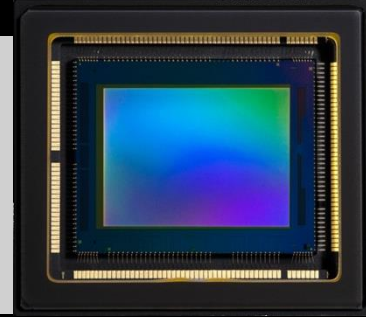


« Deep » multiple scattering regime :

- ✗ No more ballistic light
- ✗ Strong spatial and temporal perturbation
- ✓ Coherence is maintained



3D random Sample
« white paint »



DETECTORS

Complex Medium

- One layer neural network
- dense i.i.d random
- 10^{12} Complex coefficients
- Detection (intensity) non-linearity

Spatial
Light
Modulator (SLM)
= Input layer

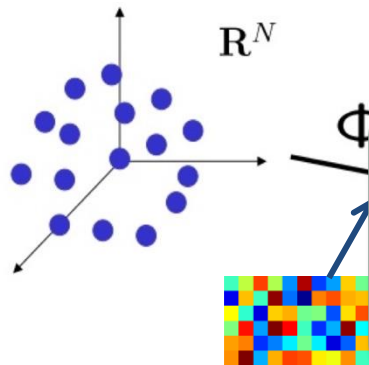
Medium
= weights

a.k.a. in signal processing : « **random
projections** »
A **universal** operation

Using disorder for computation in
Machine Learning?

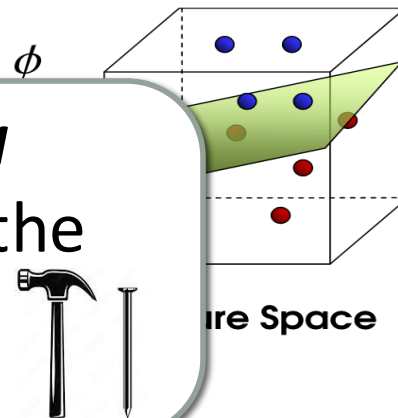
Dimensional Reduction

« Johnson & Lindenstrauss »



Dimensional expansion

The “kernel Trick” « Rahimi & Recht »



Optical pre-processing
 You don't need to know the
 matrix!
 (just know that it is random)

- Random matrix : reduction in dimension
- **conserve distances** even for $M \ll N$

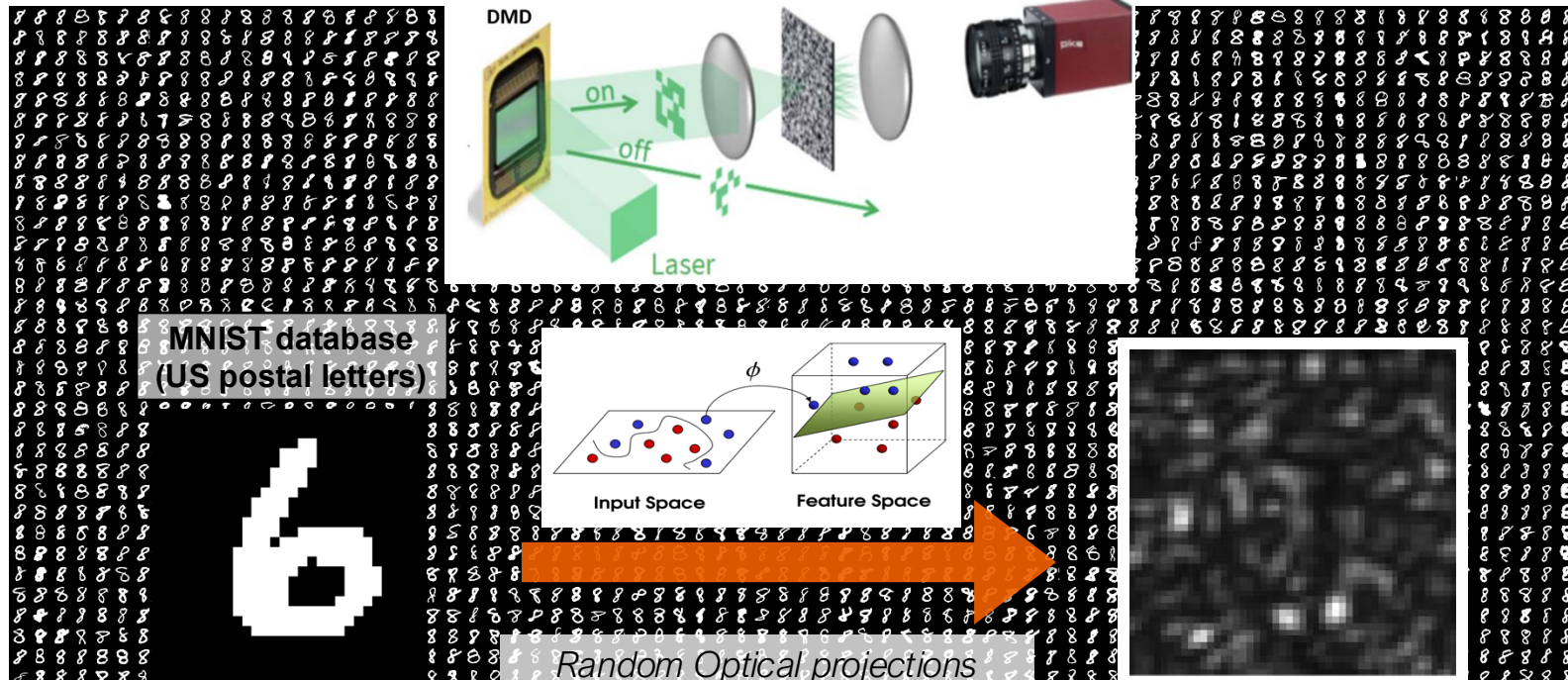
- Make a **non-linear** regression problem **linear**
- Random projections are **efficient** and **universal**

Johnson, W. B., & Lindenstrauss, J. Extensions of Lipschitz mappings into a Hilbert space. *Contemporary mathematics*, 26,189(1984)

cited >3000 times

Rahimi, A., & Recht, B. (2007). Random features for large-scale kernel machines. In *Advances in neural information processing systems* (pp. 1177-1184).

cited >4000 times



A. Saade, F. Caltagirone, I. Carron, L. Daudet, A. Drémeau, S. Gigan, F. Krzakala,
**Random projections through multiple optical scattering: approximating kernels
 at the speed of light, ICCASP (2016)**



EXTRA-LARGE

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

&

SUPER-FAST

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



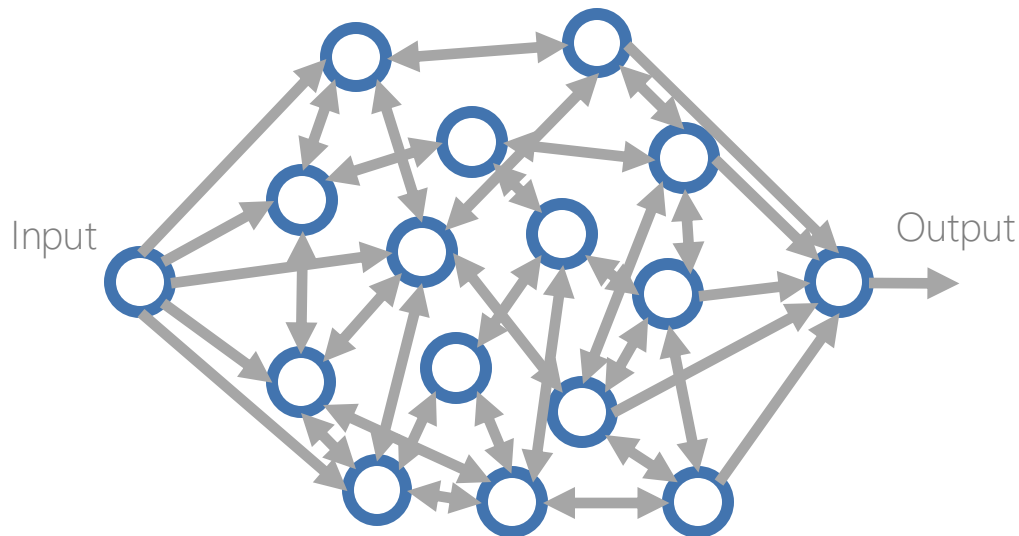
Equivalent 10^{15} operations / s : You would need a
Peta-scale computer to do the same !

LightOn

We bring Light to AI

(Col disclosure: S.G. acknowledges
financial interest in LightOn)

- many, many use cases (inference, training, linear algebra...) 😊
- already at scale for modern machine learning 😊
- 1st commercial optical co-processor 😊
- Recently moved away from optics ☹️



Recurrent Neural Networks are notoriously hard to train

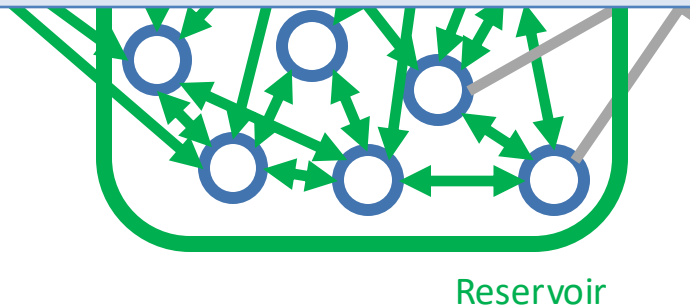
Particularly well suited for physical implementations

- Dedicated electronics
- Exotic architectures
- **Integrated & free space photonics**

Input

Van der Sande, Guy, Daniel Brunner, and Miguel C. Soriano.

"Advances in photonic reservoir computing." *Nanophotonics* 6.3 (2017): 561-576.

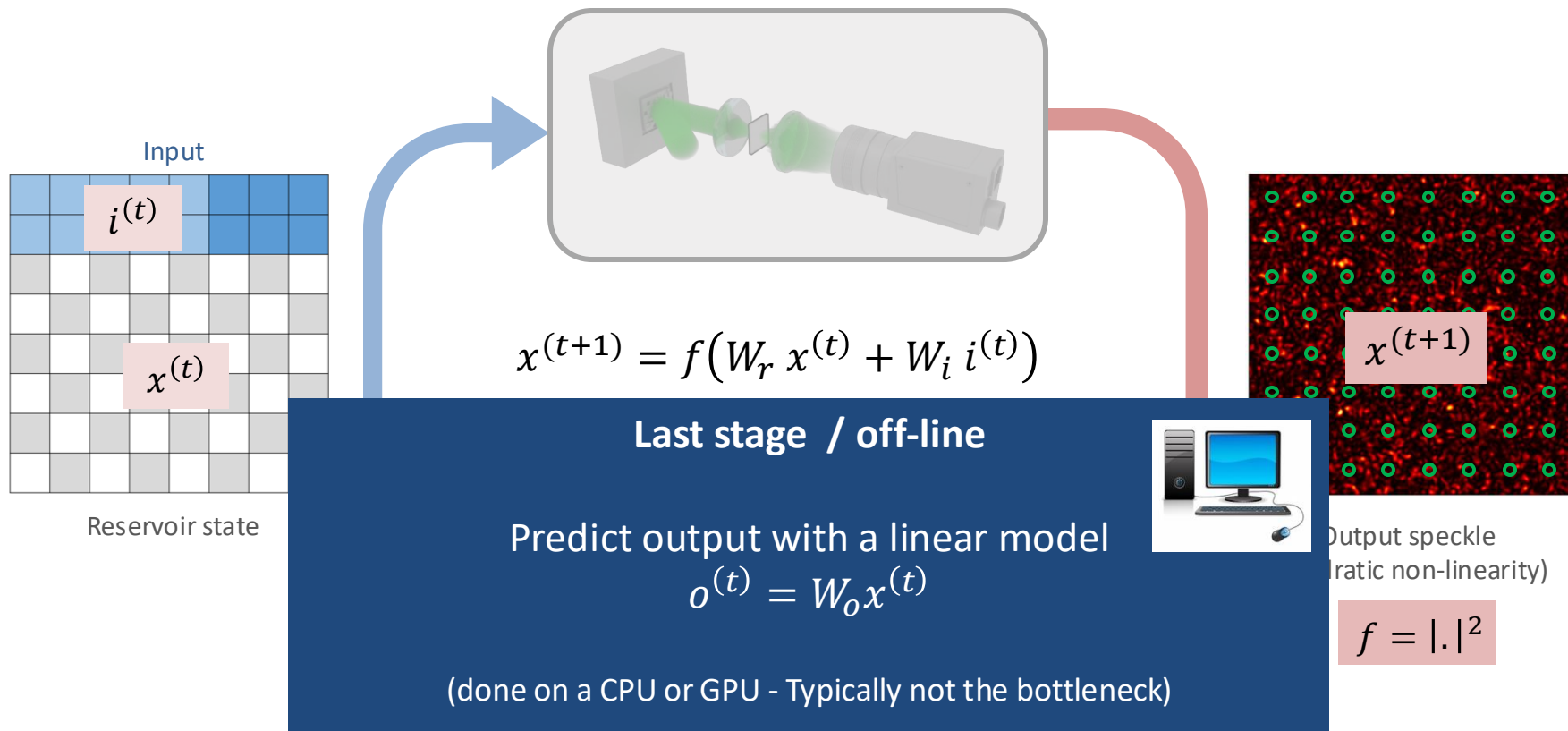


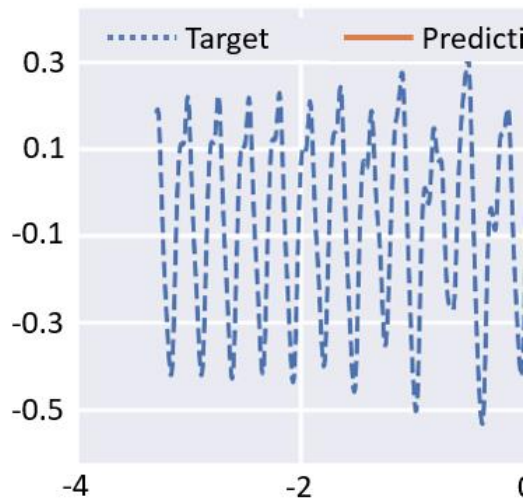
Only the **output weights** are trained

$$x^{(t+1)} = f(W_r x^{(t)} + W_i i^{(t)})$$

next reservoir current reservoir current input

Random matrices



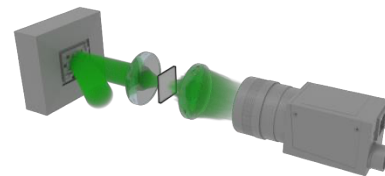


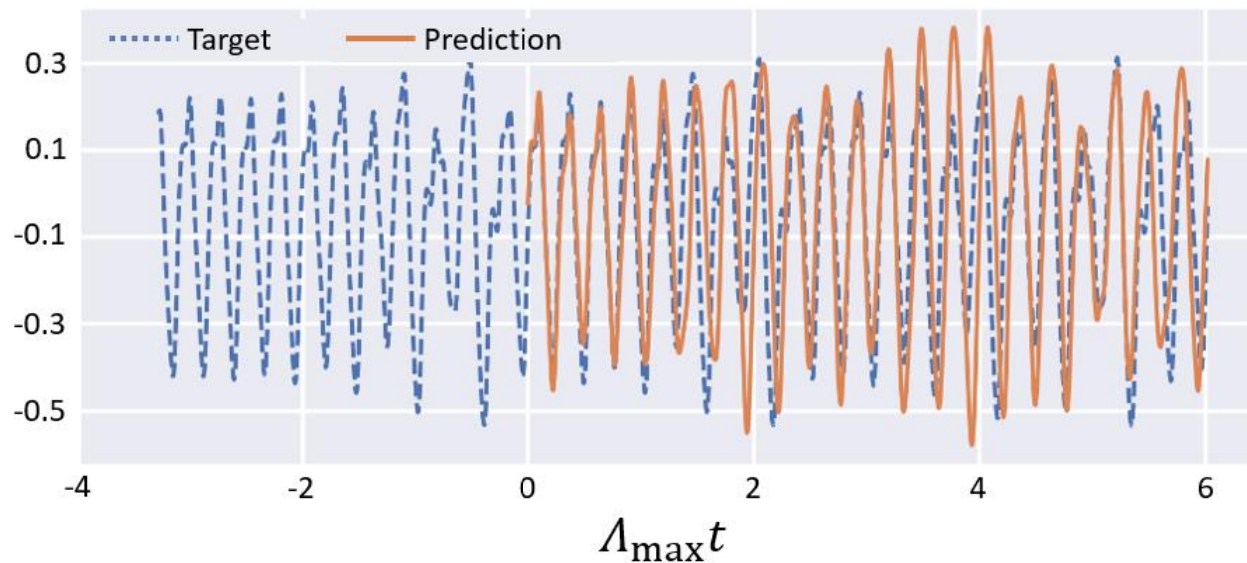
1. Compute the reservoir states

$$x^{(t+1)} = \frac{1}{\sqrt{N}} f(W_r x^{(t)} + W_i i^{(t)})$$

2. Output with a linear model

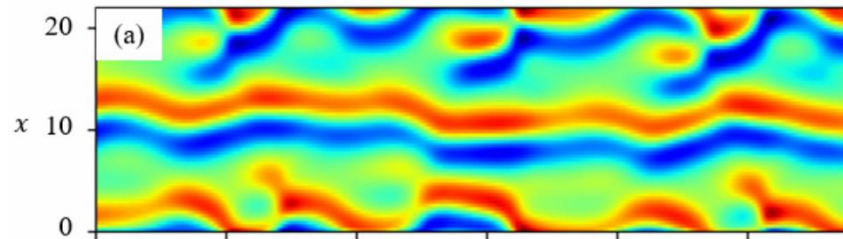
$$o^{(t)} = W_o x^{(t)}$$







Ground truth

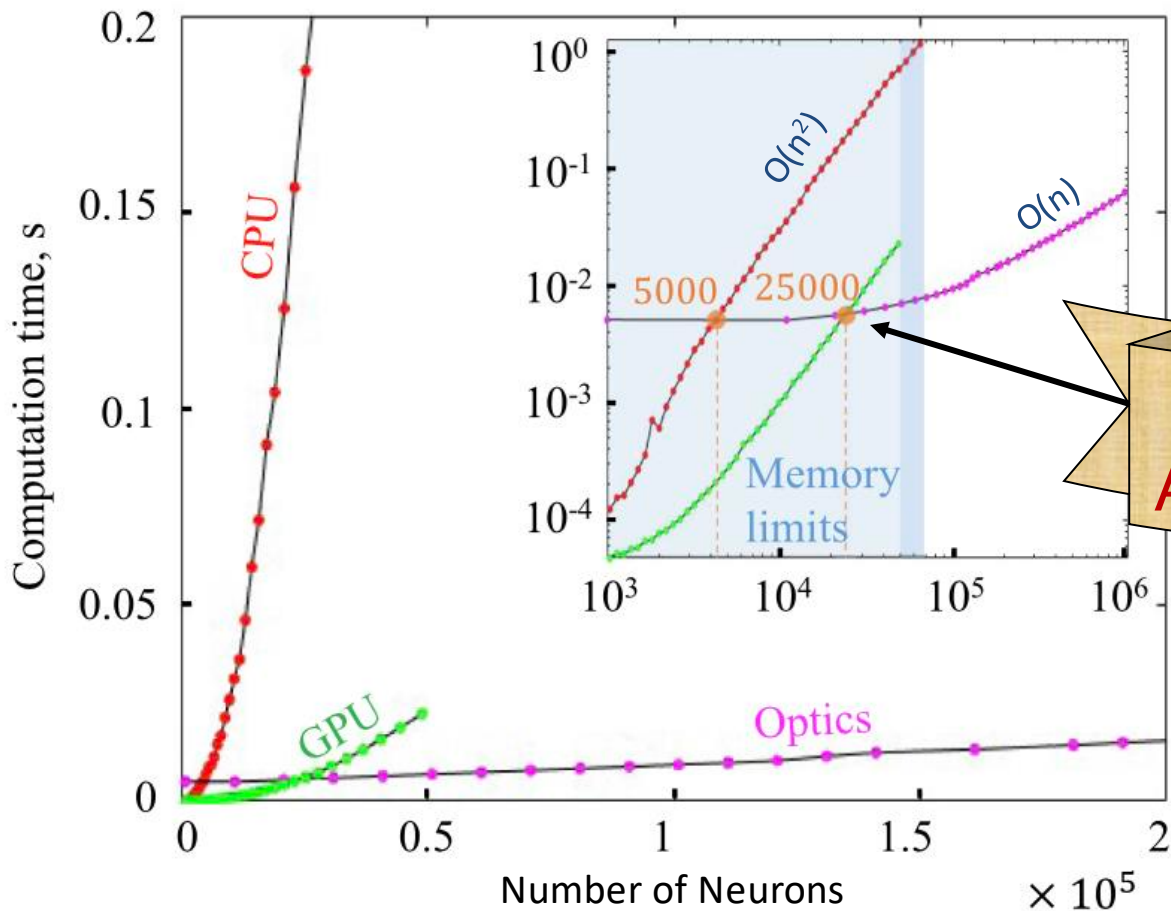


Speed

	Electronics	Optics
Speed	$O(n^2)$	$O(1)$
Energy efficiency	~150 W	~30 W
Dimensionality	Memory limit (~ GB)	Resolution limit (~ TB)

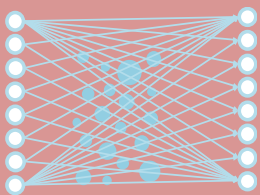
Energy efficiency

Dimensionality



**Optical
Advantage !**

Optical Random Projections



- Easy fabrication
- All-to-all connectivity
- Fixed weights
- Already at scale
- Low power consumption

➤ **Proof of principle:** classification, reservoir computing, Ising models ...

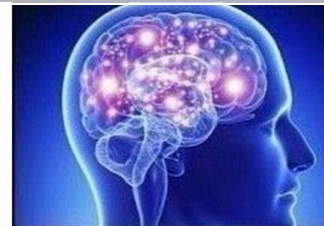
➤ Drawbacks:

- No possibility to train
- No non-linearity (single layer)

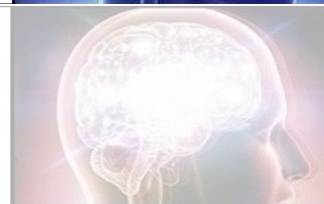
Random Projections



Trainable Random Projections



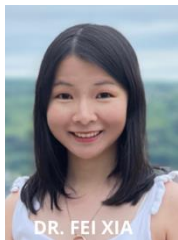
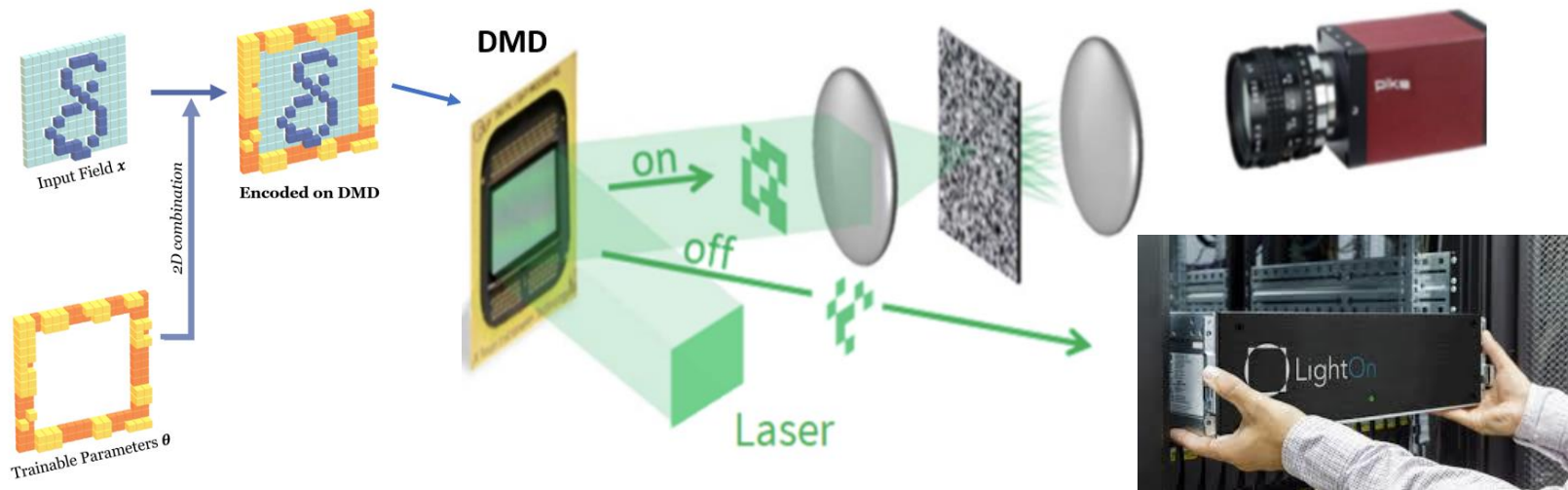
Non-Linear Random Projections



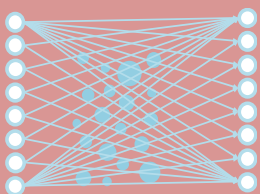
Passive Non-Linear Random Projections



- Tunable in weight by encoding on DMD



Optical Random Projections



- Easy fabrication
- All-to-all connectivity
- Fixed weights
- Already at scale
- Low power consumption

➤ **Proof of principle:** classification, reservoir computing, Ising models ...

Random Projections



Trainable Random Projections



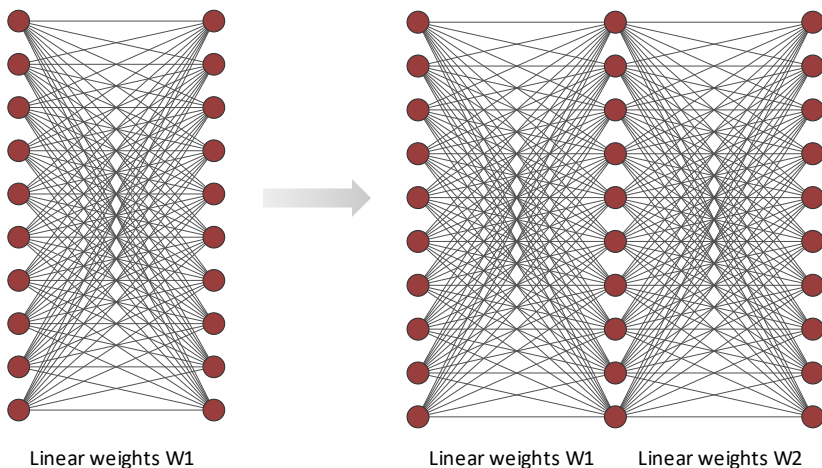
Non-Linear Random Projections



Passive Non-Linear Random Projections



Nonlinearity is the prerequisite for deep neural networks and essential for good performance



For instance:

Third-order nonlinearity: U. Tegin et al., Nat. Comput. Sci., 1, 542–549 (2021)

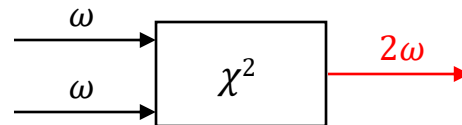
O-E nonlinearity: A. Farshid et al., Nature, 606, 501-506 (2022)

O-E-O nonlinearity: T. Wang et al., Nat. Photonics, 17, 408-415 (2023)

...

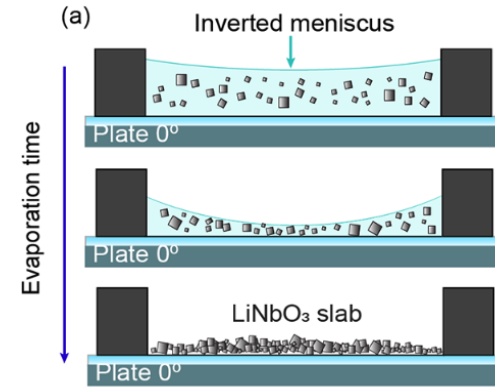
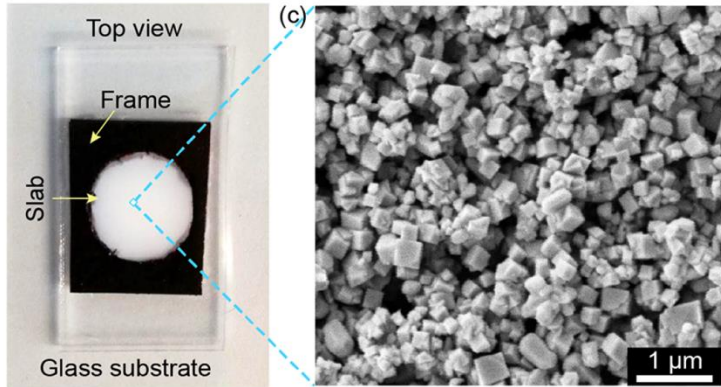
- How to introduce nonlinearity into the optical scattering process ?

Second-harmonic generation

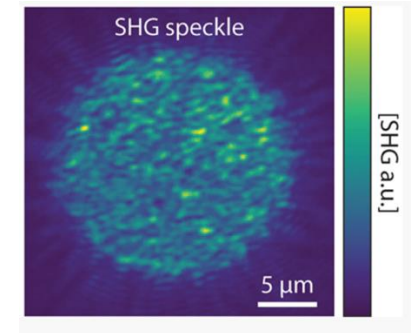


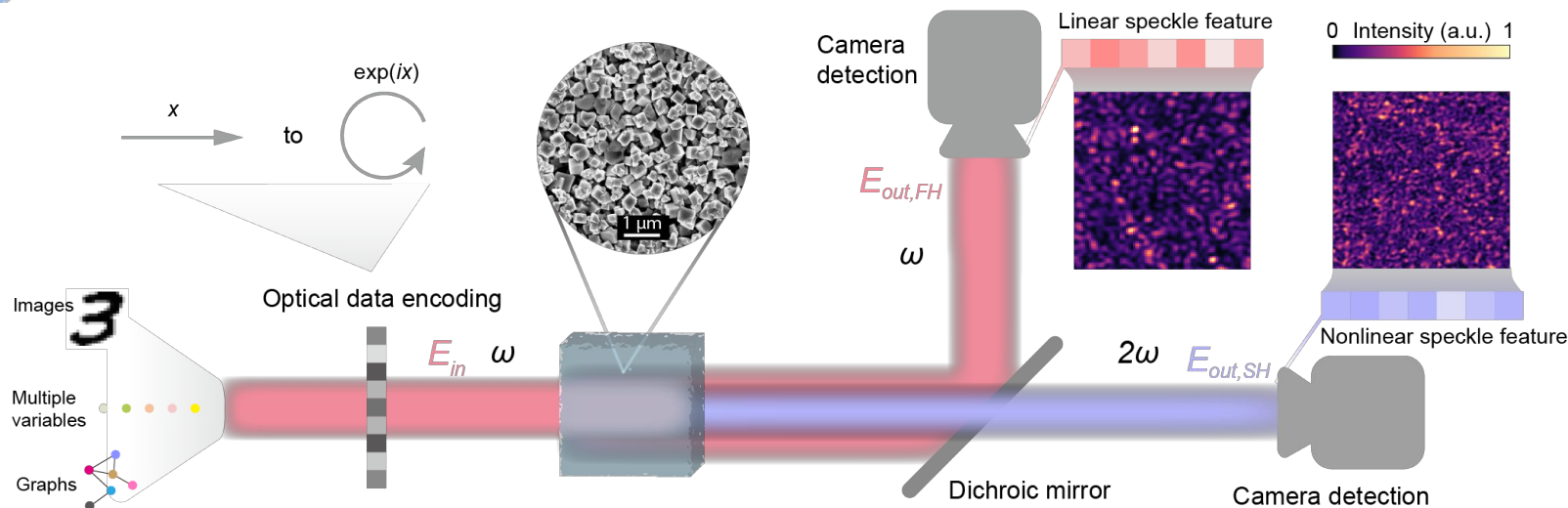
Other second or third-order nonlinear processes, photon-phonon processes (Raman, Brillouin), etc can be explored

Disordered lithium niobate (LN) nanocrystals



- Nanocrystal size: 100 – 400 nm
- Slab thickness: 5 μm
- The LN slab is both nonlinear and random scattering
- SH speckle is generated by the $\chi^{(2)}$ nonlinearity of LN and random quasi-phase-matching

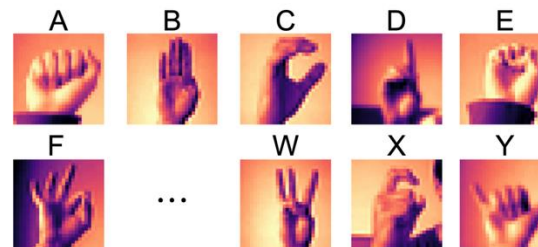




- Both fundamental-harmonic (FH, linear) and second-harmonic (SH, nonlinear) speckle features are generated
- Ridge regression for analytic readout layer (digital part)
- Focus on optical nonlinearity, with linear random projection as baseline

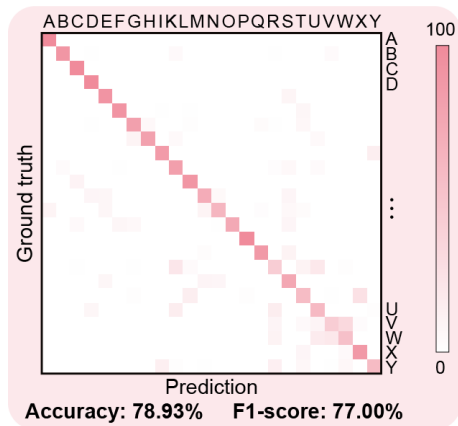
Classification of 24 sign language letters

- Training/ test size: 27,455/7,172
- Image size: 28×28

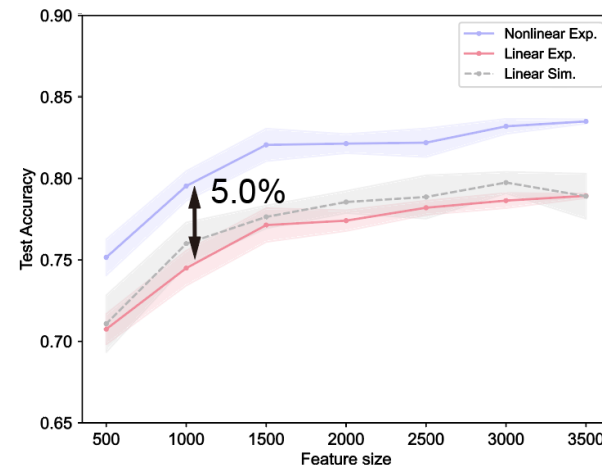
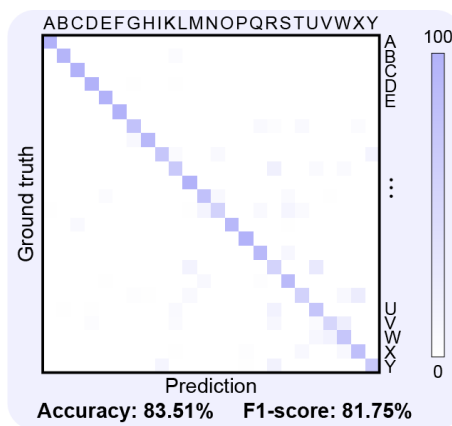


At the same feature size of 3,500:

Linear feature

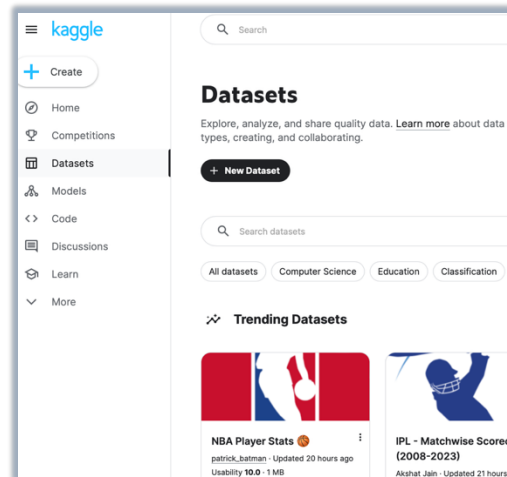


Nonlinear feature



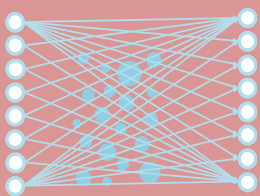
Dataset	Task description	Training/test size	Sample dimension	Macro-pixel size	Average optical nonlinearity gain
SLD	Image classification (10 categories)	1,649/413	64×64	5	5.1%
ASL	Image classification (24 categories)				
CIFAR-10	Image classification (10 categories)	50,000/10,000			
Sinc function	Interpolate the sinc function (univariate regression)	2			
UCI yacht hydrodynamics	Predict the residuary resistance of sailing yachts (multivariate regression)	2			
Concrete compressive strength	Predict the concrete compressive strength (multivariate regression)	927			
SBM-generated graph dataset	Graph classification (2 categories)	400/100			
Reddit-binary graph dataset	Graph classification (2 categories)	1,800/450			
MNIST	Image classification (10 categories)	60,000/10,000			
FashionMNIST	Image classification (10 categories)	60,000/10,000	28×28	10×10	-0.2% ~ 0.9%
STL-10	Image classification (10 categories)	8,000/5,000	96×96×3	2×2	2.4% ~ 3.5%

- Combination of the nonlinear conversion and light scattering
- Large dimension given by megapixel SLM and camera
- Gain compared to linear random projections



The nonlinear speckle features outperform their linear counterparts (and ideal linear random projection) for a large collection of datasets

Optical Random Projections



- Easy fabrication
- All-to-all connectivity
- Fixed weights
- Already at scale
- Low power consumption

➤ **Proof of principle:** classification, reservoir computing, Ising models ...

Random Projections



Trainable Random Projections



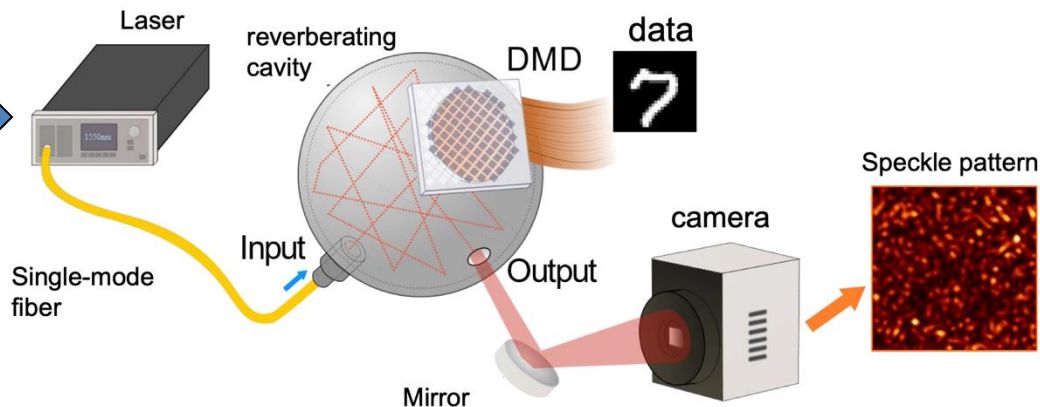
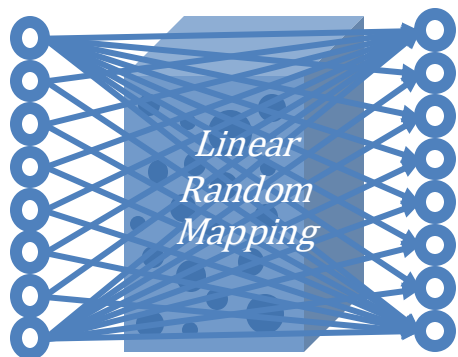
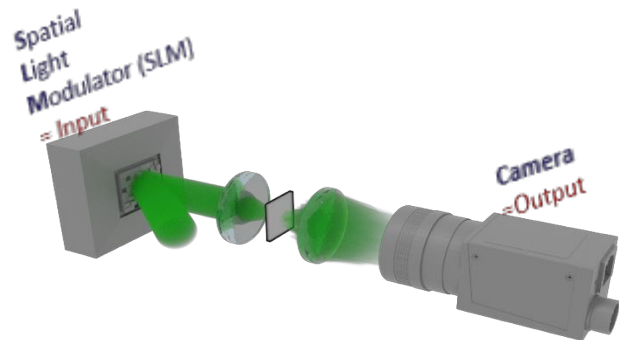
Non-Linear Random Projections



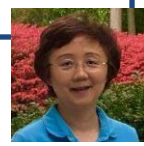
Passive Non-Linear Random Projections



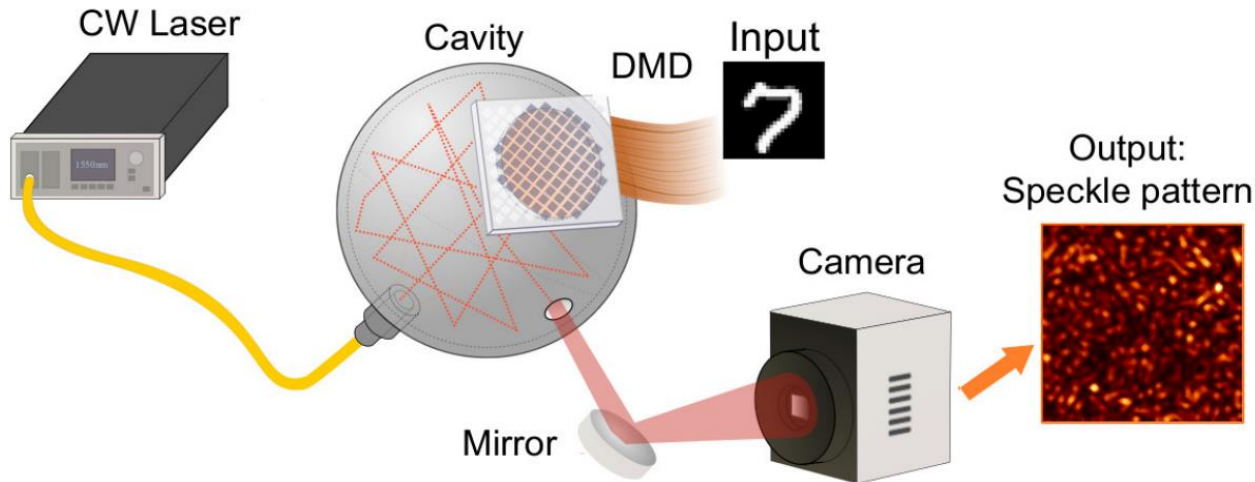
Idea : multiple bounces off the same spatial light modulator



F. Xia, K. Kim, Y. Eliezer, L. Shaughnessy, S. Gigan, **Hui Cao (Yale)** Nature Photonics (2024)
See also : Yildirim et al. Nature Photonics (2024) (Chris Moser & Demetri Psaltis, EPFL)
See also Waniura & Marquardt. Nature Physics (2024)



Key idea: encoding in scattering potential $\mathbf{V}$
 → a nonlinear data encoder in the optical domain

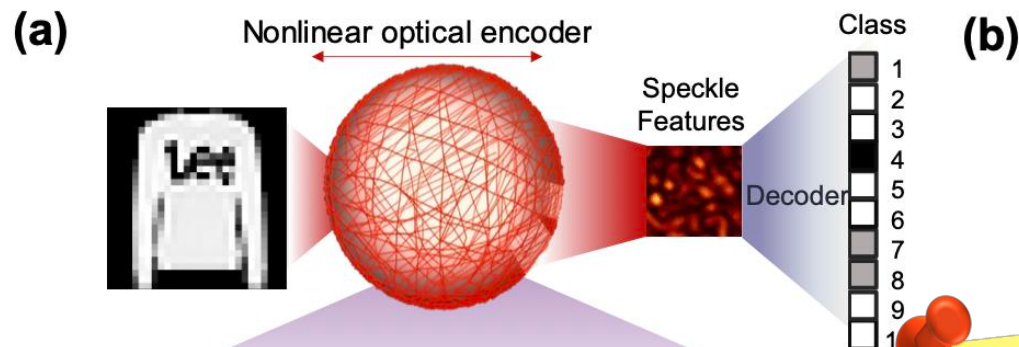


The input field E_{in} and output field E_{out} are related by:

$$E_{out} = \mathbf{W}E_{in} = (\mathbf{V} + \mathbf{V}(\mathbf{G}_o \mathbf{V}) + \mathbf{V}(\mathbf{G}_o \mathbf{V})^2 + \dots) E_{in}$$

Encoding in scattering potential $\mathbf{V}$

where $\mathbf{W}$ is a complex transmission matrix, $\mathbf{V}$ is scattering potential, $\mathbf{G}_o$ represents the free-space Green's function

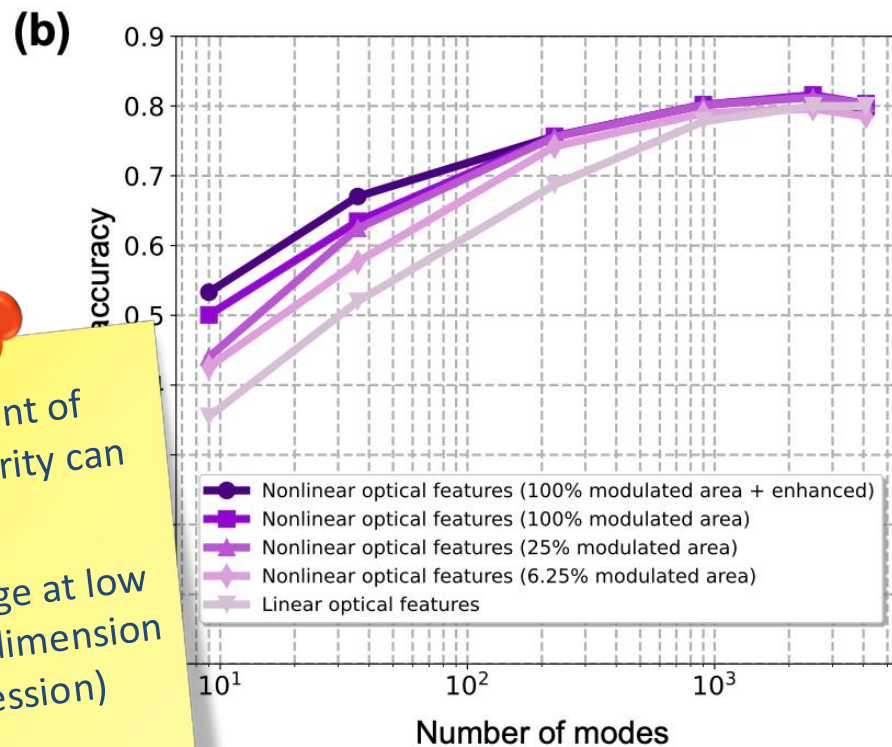


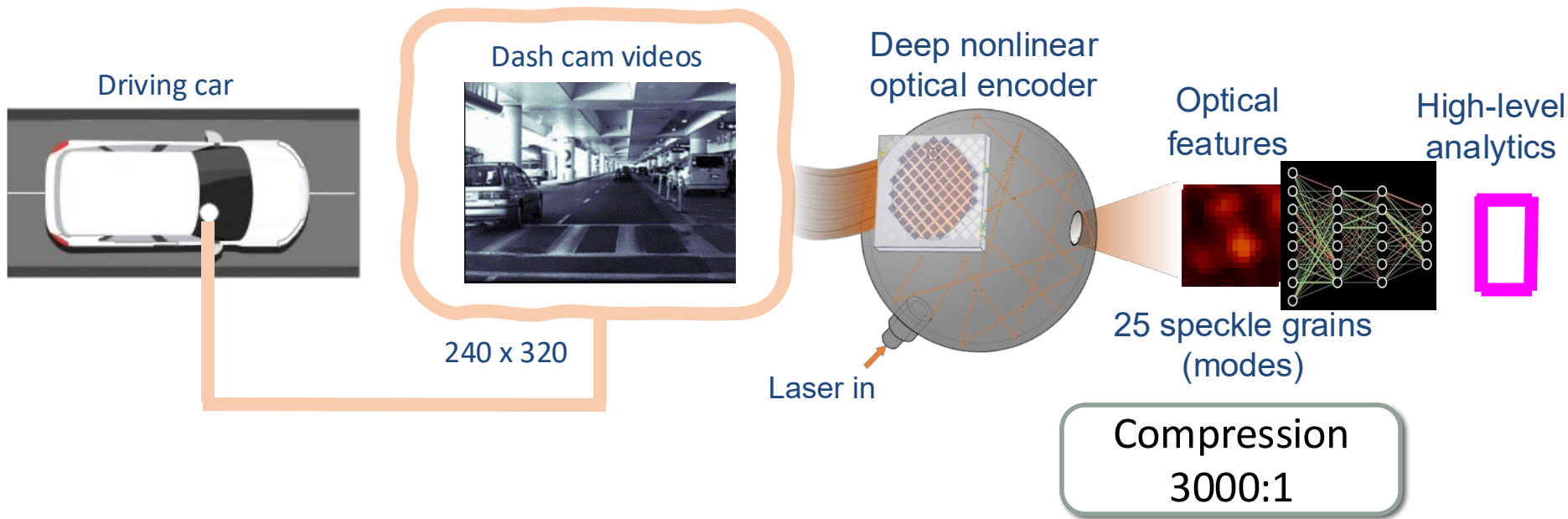
6.25% modulated area 25% modulated area 100% modulated area



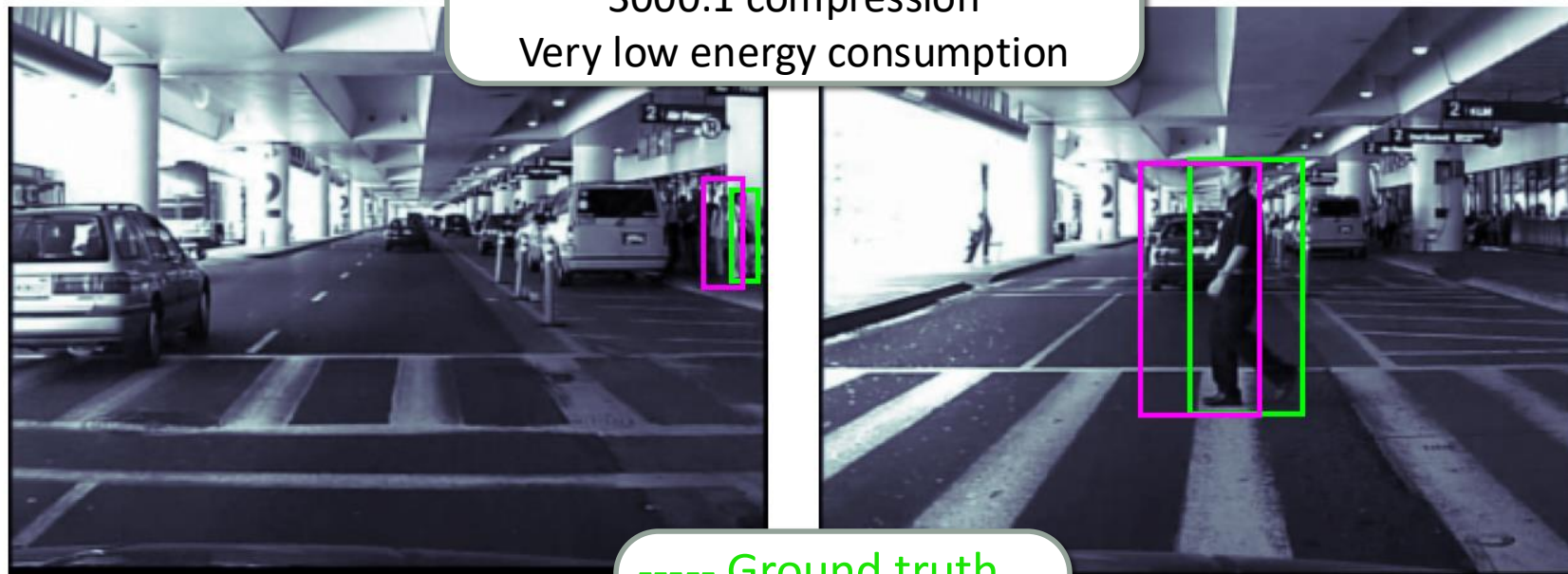
Tunable nonlinear mapping by adjusting the modulated area

- The amount of non-linearity can be tuned
- Advantage at low output dimension (compression)





25 output pixels
3000:1 compression
Very low energy consumption



----- Ground truth
----- speckle
<2 px sdev error

One more thing...

Can you use optics to train a (very) large digital model ?
Can we replace « backprop' » ?

Wang, Muller, et al., arXiv:2409.12965(2024)



FORWARD : inference



Training in progress...



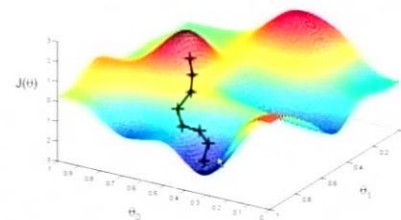
« **backprop'** »

- de-facto standard for training
- Power-intensive
- Constraints to a layered NN architecture

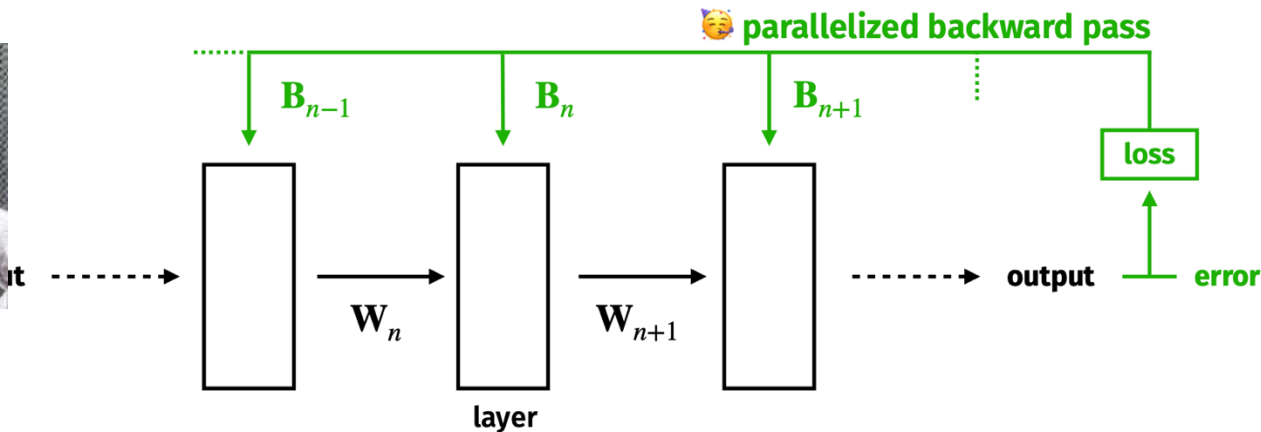


BACKWARD : Error propagation and weight-tuning based on local gradients

Gradient Descent

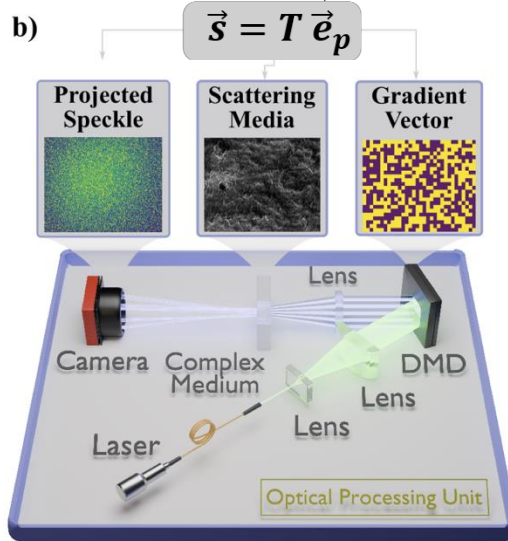
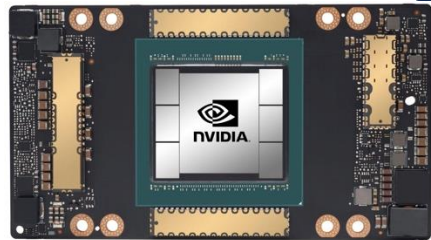
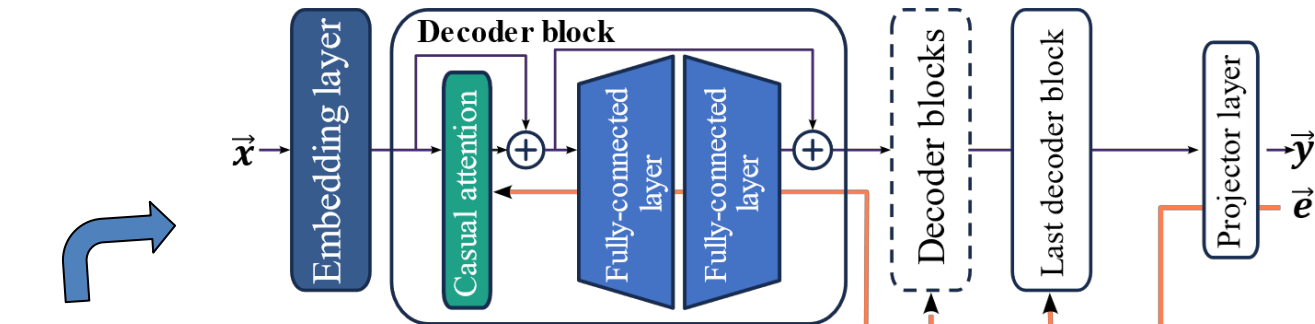


Use **random weights** to **directly** deliver feedbacks from the global loss.



$$\begin{array}{l} \text{gradient signal from layers above} \qquad \qquad \text{random projection of the loss gradient} \\ \delta \mathbf{W}_i^{BP} = - [(\mathbf{W}_{i+1}^T \delta \mathbf{a}_{i+1}) \odot f'_i(\mathbf{a}_i)] \mathbf{h}_{i-1}^T \longrightarrow \delta \mathbf{W}_i^{DFA} = - [(\mathbf{B}_i^T \mathbf{e}) \odot f'_i(\mathbf{a}_i)] \mathbf{h}_{i-1}^T \end{array}$$

General-purpose: agnostic to model architecture & scale well



- 1B parameters GPT-like transformer
- 400M parameters directly receive optical signal

Cornell Movie-Dialogs Corpus Example in the Dataset

- MILES: Back at you.
- JACK: Love you, man.
- MILES: Yeah.
- JACK: So I'll see you at the rehearsal.



Ziao Wang
(LKB)



Kilian Müller
(Lighton)

0% Training

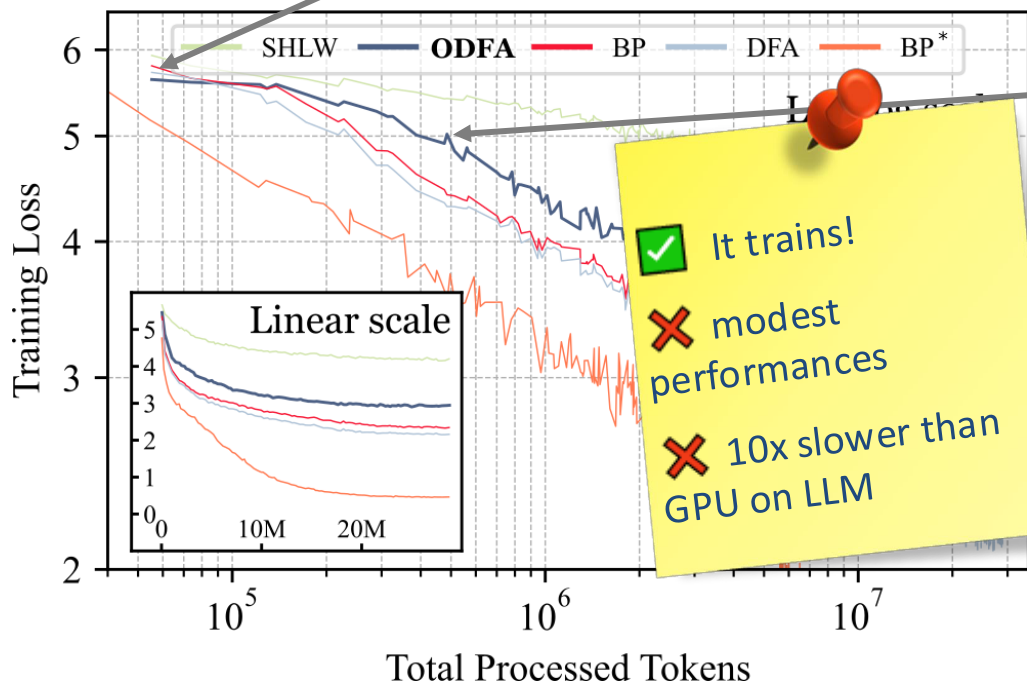
aQ o do wantQain do do twoain o
re[through show" happenQ start. 2lt
JULIANNE JULIANNE

20% Training

JACKIE : Snkups to the brags, but I
know that. ↵
CRAYC : It ' s just a concer. ↵

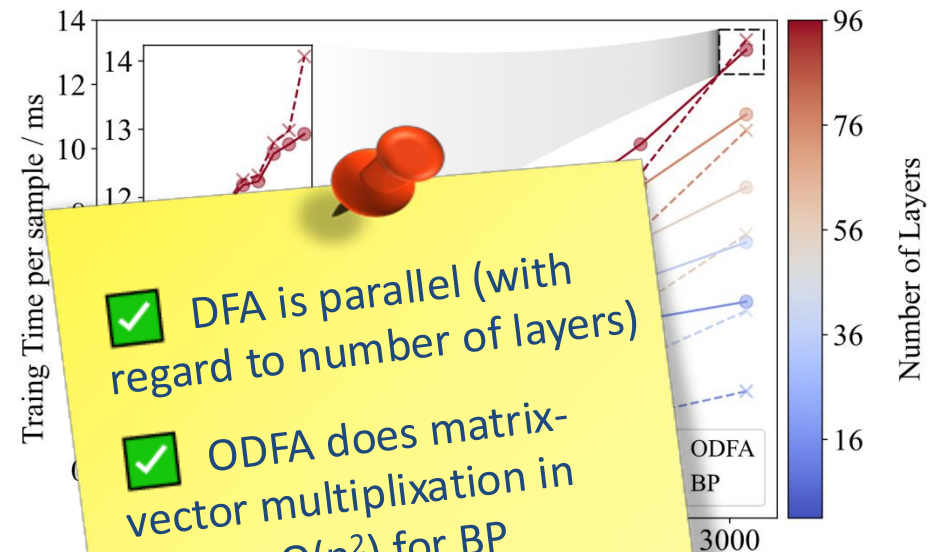
100% Training

JOE : So what are you talking about? ↵
ED : I mean I'm sorry, you know how
to know you are. ↵



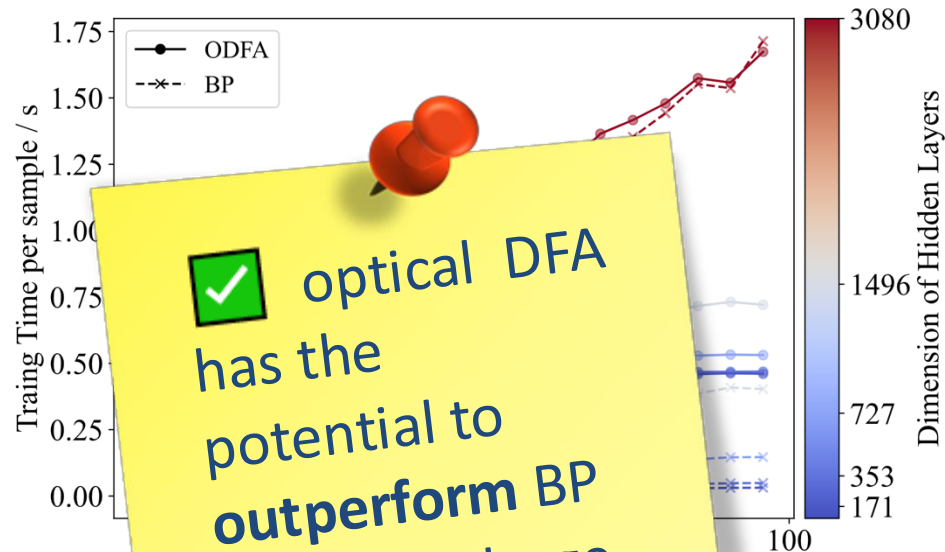
✓ It trains!
✗ modest performances
✗ 10x slower than GPU on LLM

Scaling along the dimension of hidden layers



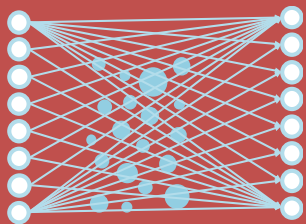
- ✓ DFA is parallel (with regard to number of layers)
- ✓ ODFA does matrix-vector multiplication in $O(1)$ vs $O(n^2)$ for BP

Scaling along the number of layers



- ✓ optical DFA has the potential to outperform BP for future large models !

Optical Random Projections



- Easy fabrication
- All-to-all connectivity
- Fixed weights
- at scale
- passive

Many proof of principles:

- Classification,
- Reservoir computing
- Image generation
- Ising models
- Quantum circuits
- ...

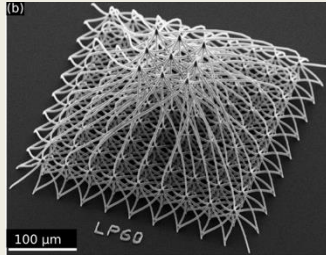
Beyond random projections

- Adding tunability and trainability
- Including low-power non-linearity
- « Reverberating » passive non-linearities

Ability to tackle relatively large-scale + advanced AI tasks

Convergence between free space and integrated optics?

“Best of both worlds”



Optica 7, 640-646 (2020) (D. Brunner team – FEMTO-ST)

Convergence between algorithms & optics?

(Not just deep learning)

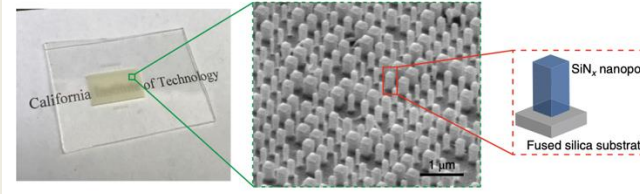
“the hardware lottery”



S. Hooker [arXiv:2009.06489](https://arxiv.org/abs/2009.06489)

Convergence between top-down and bottom-up

reproducibility / scalability / inverse design



Nature Photon 12, 84–90 (2018) (C. Yang and A. Faraon, Caltech)

Thanks to my team and collaborators
Special mention to : Ziao Wang, Fei Xia, Jianqi Hu, Hao Wang



Thank you for your attention !
Don't hesitate to ask for the slides

Mail : sylvain.gigan@lkb.ens.fr

If you are interested in the field :

REVIEWS OF MODERN PHYSICS

[Recent](#) [Accepted](#) [Authors](#) [Referees](#) [Search](#) [Press](#) [About](#)

Light fields in complex media: Mesoscopic scattering meets wave control

Stefan Rotter and Sylvain Gigan
Rev. Mod. Phys. **89**, 015005 – Published 2 March 2017

Perspective | Published: 02 December 2020

Inference in artificial intelligence with deep optics and photonics

Gordon Wetzstein , Aydogan Ozcan, Sylvain Gigan, Shanhui Fan, Dirk Englund, Marin Soljačić, Cornelia Denz, David A. B. Miller & Demetri Psaltis

Nature **588**, 39–47(2020) | [Cite this article](#)

Perspective | Published: 08 September 2022

Imaging and computing with disorder

[Sylvain Gigan](#)

Nature Physics **18**, 980–985 (2022) | [Cite this article](#)