Machine learning at the speed of light

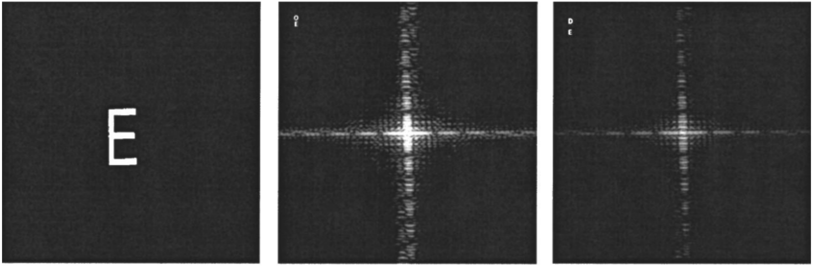
Piotr Antonik

April 7, 2022

CentraleSupélec

ULB

femto-st SCIENCES & TECHNOLOGIES



Li Y. *et al.*, Journal of Biomedical Optics **7** (2002)

Image source: Wikipedia

Zuse 3

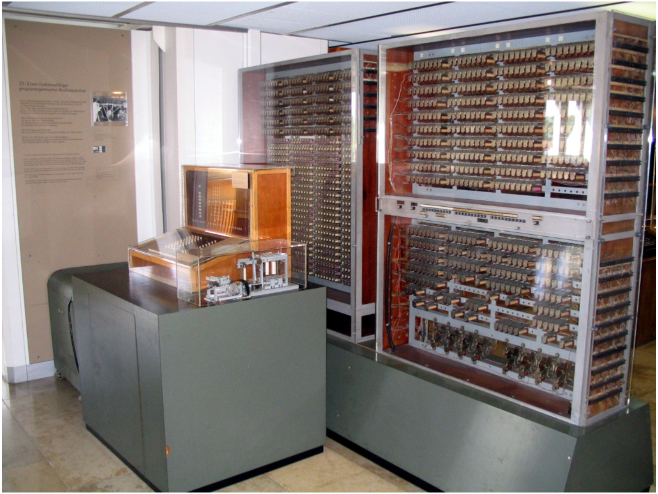


Image source: Wikipedia

	Electron	Photon
Spin	Fermion	Boson
Charge	Yes	No
Interaction	Yes	No
Nonlinear transformations	✓✓	X
Information transport	✓	✓✓

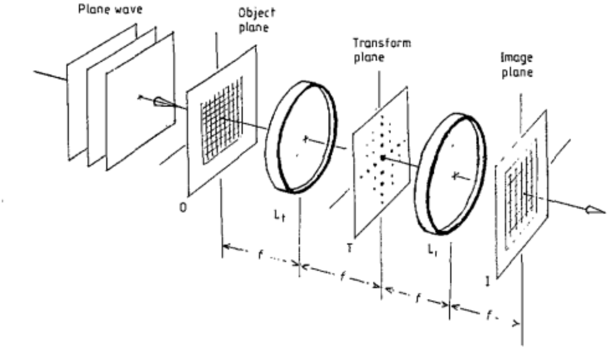
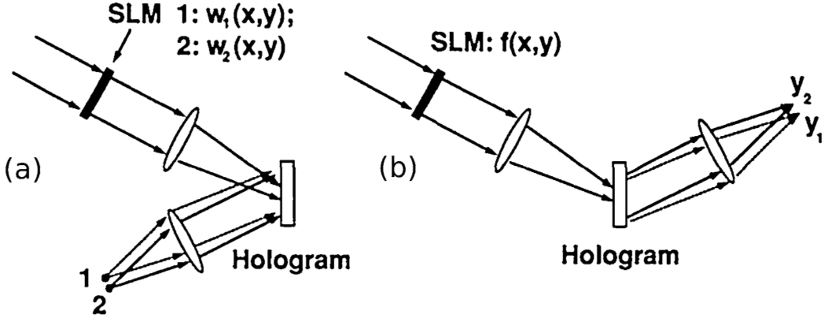


Image source: PhysWiki, York University

Optical matrix multiplication



J. H. Hong *et al.*, Applied Optics **29** (1990)

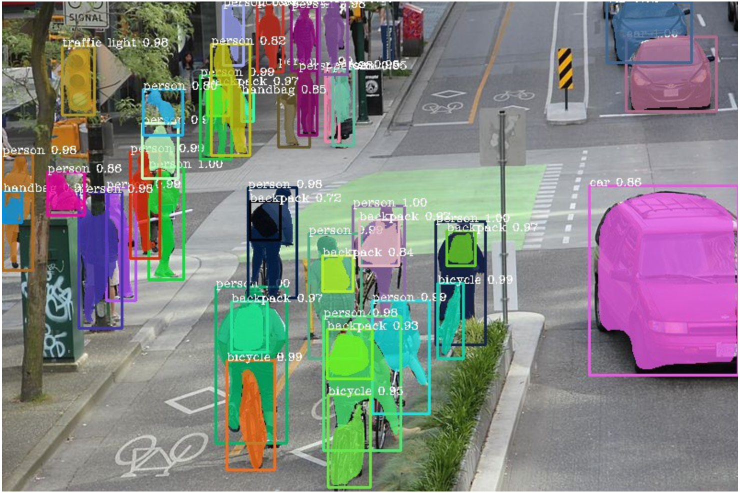
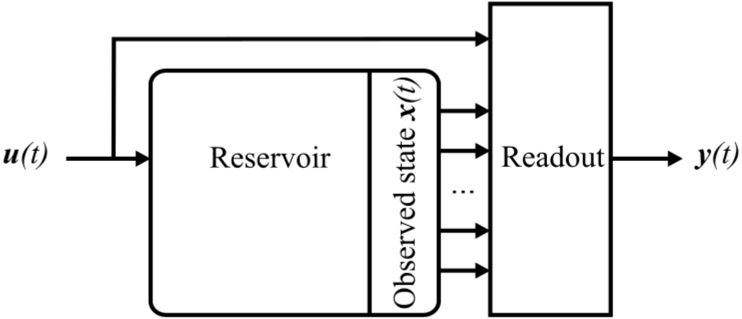
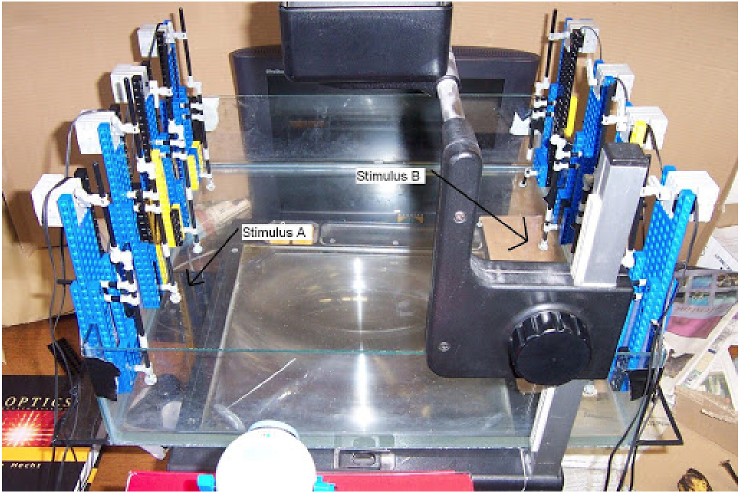


Image source: towardsdatascience.com

Summary

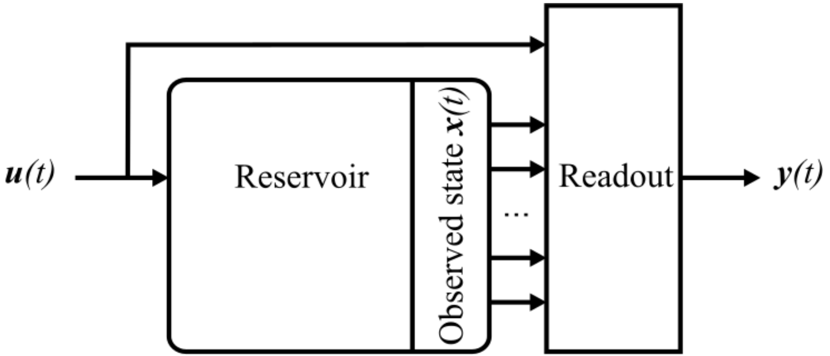


Photonic Reservoir Computing, De Gruyter (2019)



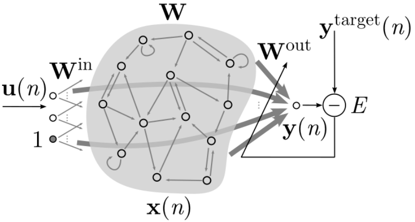
C. Fernando *et al.*, European conference on artificial life (2003)

An abstract model of computation



Photonic Reservoir Computing, De Gruyter (2019)

- ESN: H. Jaeger and H. Harald Haas, Science **304** (2004)
- LSM: W. Maass *et al.*, Neural Comput. **14** (2002)



M. Lukoševičius, Neural Netw.: Tricks of the trade (2012)

$$\mathbf{x}(n) = \tanh\left[\mathbf{W}^{\text{in}}\mathbf{u}(n) + \mathbf{W}\mathbf{x}(n-1)\right]$$

$$\mathbf{y}(n) = \mathbf{W}^{\text{out}}\mathbf{x}(n)$$

How to train a reservoir computer

Linear regression layer:

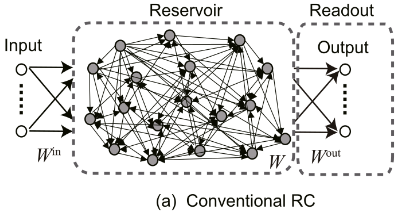
$$\mathsf{MSE}(\mathbf{y}, \mathbf{y}^{\mathsf{target}}) = \frac{1}{T} \sum_{n=1}^{T} \left(\mathbf{y}(n) - \mathbf{y}^{\mathsf{target}}(n) \right)^2$$

Minimise the MSE with:

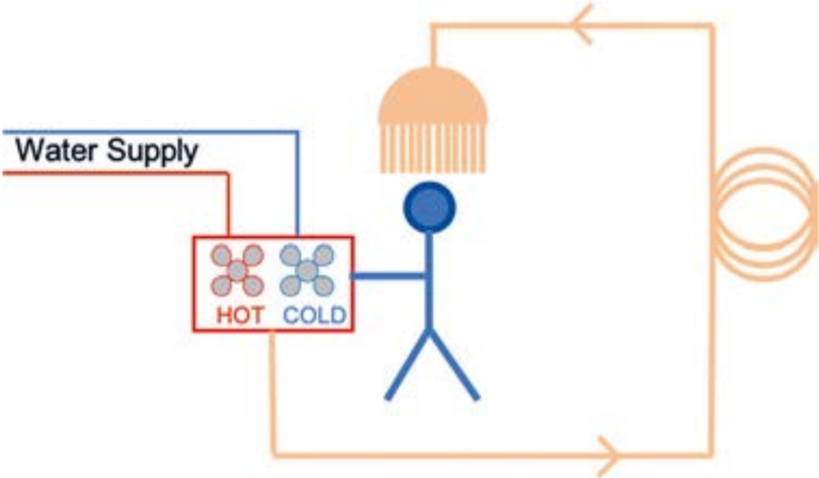
$$\mathbf{W}^{\mathsf{out}} = \left(\boldsymbol{X}^T \boldsymbol{X} \right)^{-1} \boldsymbol{X}^T \boldsymbol{y}$$

Ridge regression (regularisation):

$$\mathsf{MSE}_{\mathsf{ridge}} = \mathsf{MSE} + \lambda \left(\boldsymbol{W}^{\mathsf{out}} \right)^T \boldsymbol{W}^{\mathsf{out}}$$

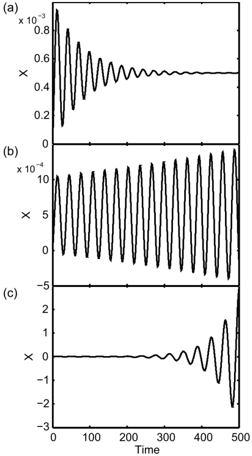


G. Tanaka *et al.*, Neural Netw. **115** (2019)

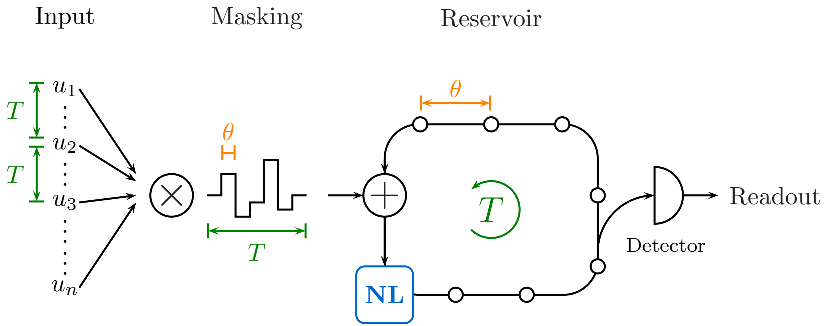


P. Antonik *et al.*, Photoniques 104 (2020)

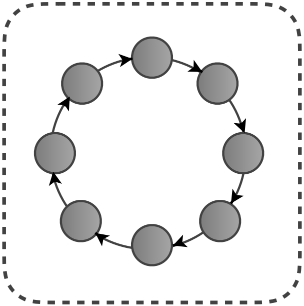
$$\dot{x}(t) = -\alpha x(t - \tau)$$



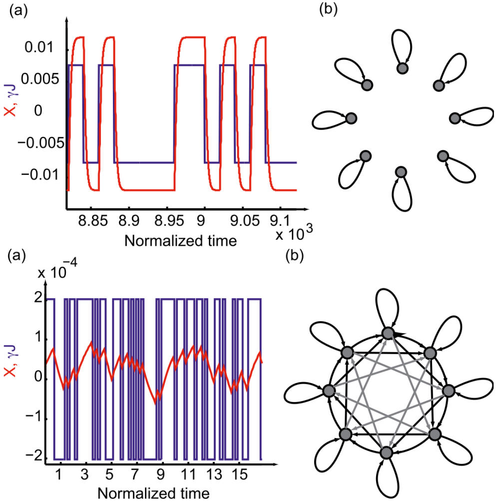
Photonic Reservoir Computing, De Gruyter (2019)



P. Antonik *et al.*, Photoniques 104 (2020)

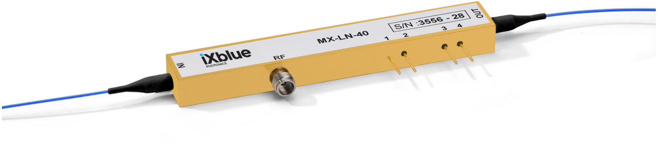


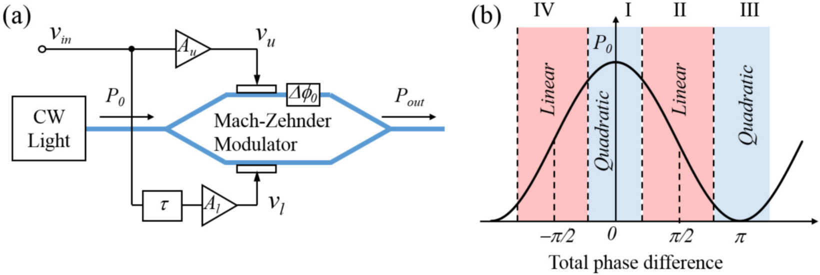
$$\begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$



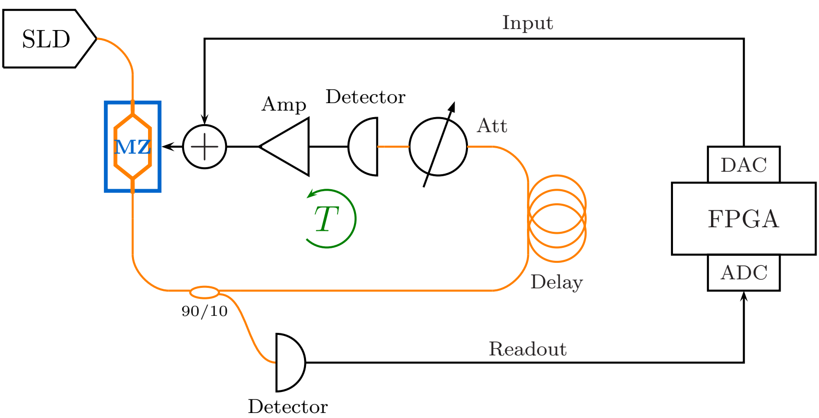
Photonic Reservoir Computing, De Gruyter (2019)

Mach-Zehnder intensity modulator

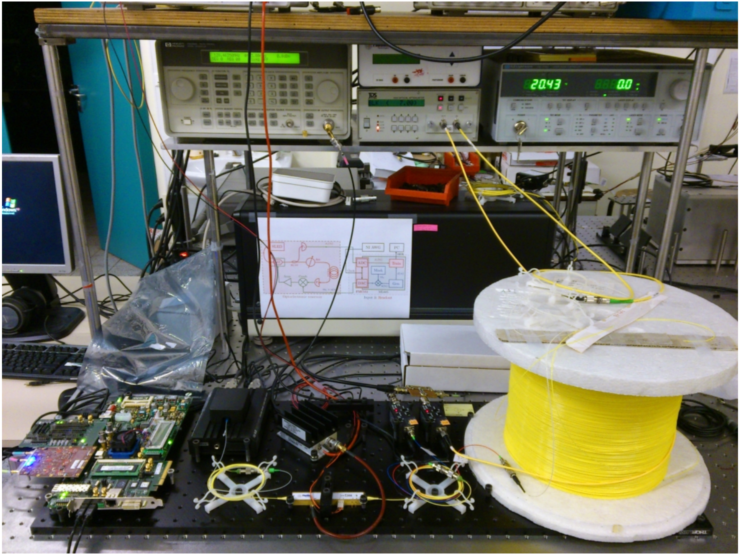




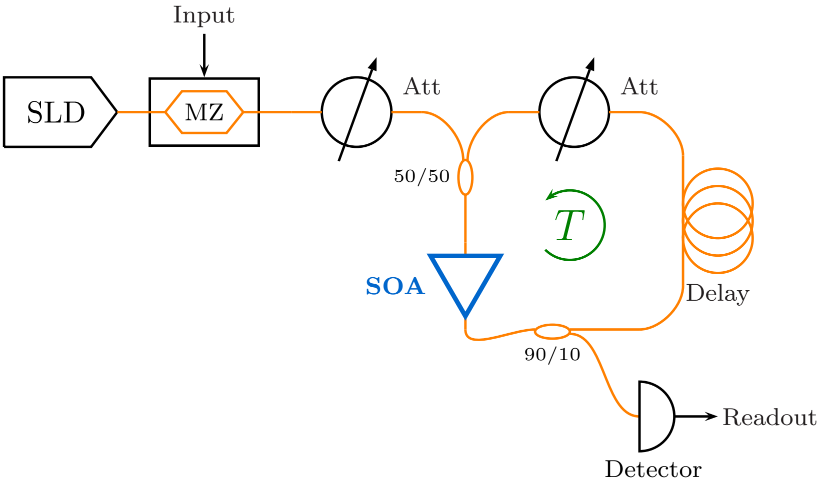
Jing Gao *et al.*, OpEx 24 (2016)



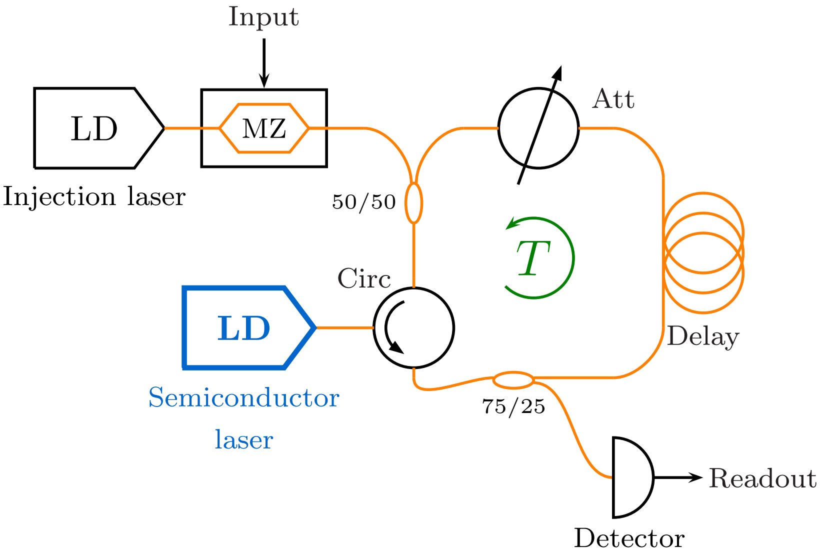
P. Antonik *et al.*, Photoniques 104 (2020)


20.43
0.0

SOA-based setup



P. Antonik *et al.*, Photoniques 104 (2020)



P. Antonik *et al.*, Photoniques 104 (2020)

Selected for a Viewpoint in *Physics*
PHYSICAL REVIEW X **7,** 011015 (2017)

High-Speed Photonic Reservoir Computing Using a Time-Delay-Based Architecture: Million Words per Second Classification

Laurent Larger,[1] Antonio Baylón-Fuentes,[1] Romain Martinenghi,[1] Vladimir S. Udaltsov,[1,2] Yanne K. Chembo,[1] and Maxime Jacquot[1]

[1]*FEMTO-ST Institute/Optics Department, CNRS & University Bourgogne Franche-Comté, 15B avenue des Montboucons, 25030 Besançon Cedex, France*
[2]*Vavilov Optical State Institute, Saint-Petersburg, Russia*

(Received 30 January 2015; revised manuscript received 13 November 2016; published 6 February 2017)

Reservoir computing, originally referred to as an echo state network or a liquid state machine, is a brain-inspired paradigm for processing temporal information. It involves learning a “read-out” interpretation for nonlinear transients developed by high-dimensional dynamics when the latter is excited by the information signal to be processed. This novel computational paradigm is derived from recurrent neural network and machine learning techniques. It has recently been implemented in photonic hardware for a dynamical system, which opens the path to ultrafast brain-inspired computing. We report on a novel implementation involving an electro-optic phase-delay dynamics designed with off-the-shelf optoelectronic telecom devices, thus providing the targeted wide bandwidth. Computational efficiency is demonstrated experimentally with speech-recognition tasks. State-of-the-art speed performances reach one million words per second, with very low word error rate. Additionally, to record speed processing, our investigations have revealed computing-efficiency improvements through yet-unexplored temporal-information-processing techniques, such as simultaneous multisample injection and pitched sampling at the read-out compared to information “write-in”.

DOI: 10.1103/PhysRevX.7.011015 Subject Areas: Complex Systems, Nonlinear Dynamics, Photonics

Time-delay photonic reservoir computers:

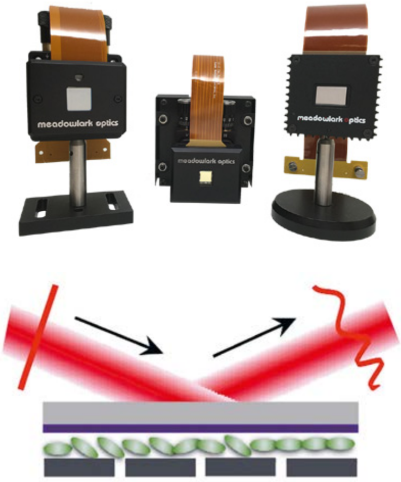
- were the first experimental demonstrations
- are bulky $\rightsquigarrow$ integration

UNIVERSITEIT GENT

imec

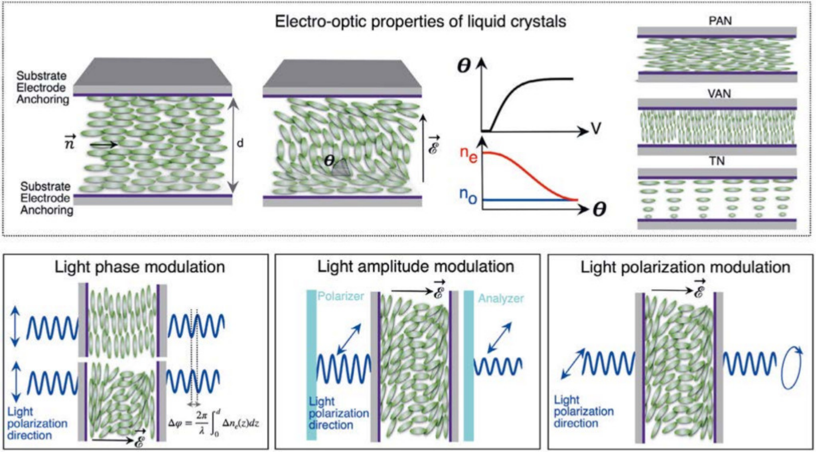
- are difficult to scale up $\rightsquigarrow$ parallel systems

Spatial Light Modulator

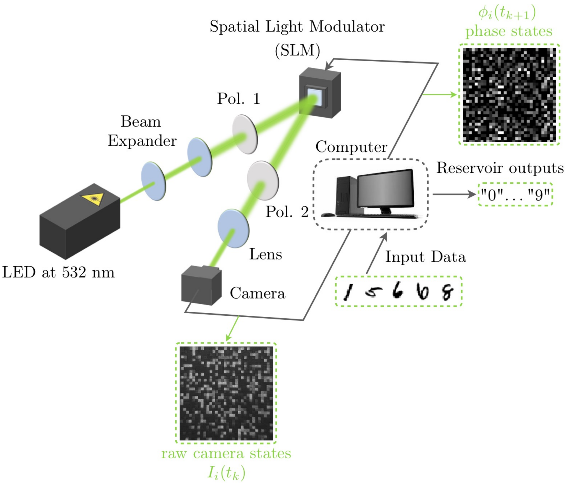


Jullien A., Photoniques 101 (2020)

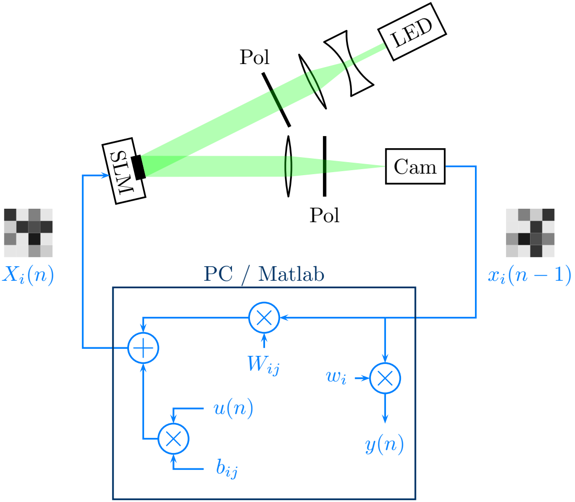
Spatial Light Modulator



Jullien A., Photoniques 101 (2020)



P. Antonik *et al.*, JSTQE 26 (2019)



P. Antonik *et al.*, Cogn. Comput. (2021)

Image source: Wikipedia

P. Antonik *et al.*, Nat. Mach. Intel. **1** (2019)

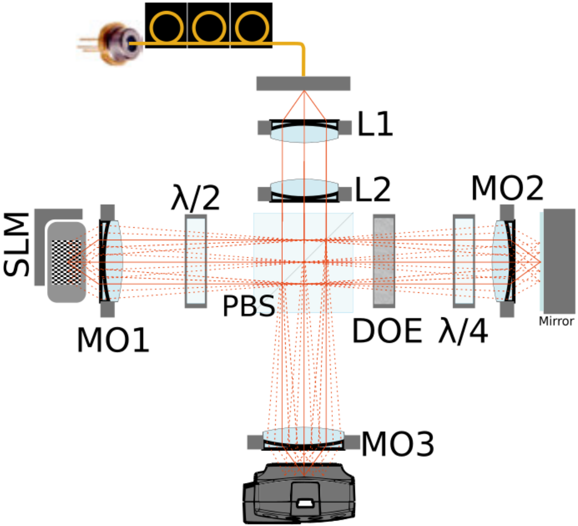
Human action recognition in videos

Table 1 | Performance of various state-of-the-art digital approaches compared to our best experimental result

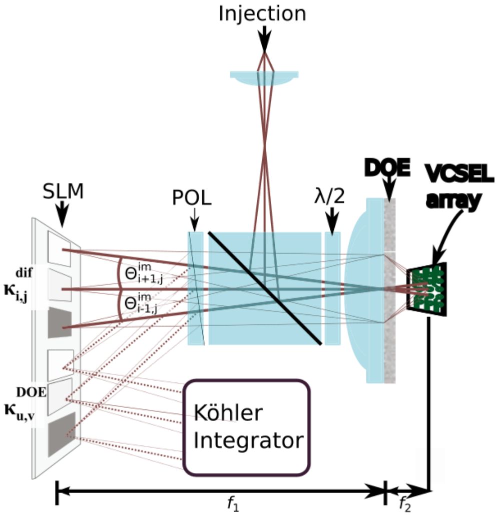
Authors	Method	Database split	Training time	Processing speed (f.p.s.)	Performance s1 scenario (%)	Performance Full database (%)
Yadav et al.[38]	IP + SVM	80%–20%	–	–	–	98.20
Shi et al.[39]	DTD, DNN	9–16	–	–	–	95.6
Kovashka et al.[40]	BoW + SVM	8–8–9	–	–	–	94.53
Gilbert et al.[33]	HCF + SVM	LOOCV	~5.6 h	24	–	94.5
Baccouche et al.[41]	CNN & RNN	16–9	–	–	–	94.39
Ali and Wang[42]	DBN & SVM	50%–20%–30%	–	–	–	94.3
Wang et al.[43]	DT + SVM	16–9	–	–	–	94.2
Liu et al.[44]	MMI + SVM	LOOCV	–	–	–	94.15
Sun et al.[45]	FT + SVM	Auto	–	–	–	94.0
Veeriah et al.[46]	Differential RNN	16–9	–	–	–	93.96
Shu et al.[47]	SNN	9–16	–	–	95.3	92.3
Laptev et al.[48]	FT + SVM	8–8–9	–	–	–	91.8
Jhuang[31]	StC_2 + SVM	16–9	–	0.4	96.0	91.6
Klaeser et al.[49]	3D Grad + SVM	8–8–9	–	–	–	91.4
This work	**Photonic RC**	**75%–25%**	**1.6–5.5 h**	**2–7**	**91.3**	–
Grushin et al.[32]	LSTM	16–9	1 day	12–15	–	90.7
Ji et al.[50]	3DCNN	8–8–9	–	–	–	90.02
Escobar et al.[51]	MT cells	16–9	–	–	74.63	–
Schuldt et al.[23]	FT + SVM	8–8–9	–	–	–	71.83

'Database split' indicates how the KTH database was split for training and testing of the system. Most studies choose to split by the number of subjects into either two groups (training and test, for example 16 subjects for training, 9 for the test) or three groups (training, validation and test, for example 8–8–9). LOOCV corresponds to 'leave-one-out cross validation': the system is trained on 24 subjects and tested on the remaining one. Training times and processing speeds are not discussed in most of the works, focusing on the classification performance. Some studies report specific results on the s1 scenario, as considered in this work. BoW, bag of words; CNN, convolutional neural network; DBN, deep belief network; DNN, deep neural network; DT, dense trajectories; DTD, deep trajectory descriptor; FT, features; HCF, hierarchical compound features; IP, interest points; LSTM, long short-term memory neural network; MMI, maximization of mutual information; MT, middle temporal area of the visual cortex; RNN, recurrent neural network; SNN, spiking neural network; StC_2, space-time oriented C_2 features; SVM, support vector machine.

P. Antonik *et al.*, Nat. Mach. Intel. **1** (2019)



Photonic Reservoir Computing, De Gruyter (2019)



Photonic Reservoir Computing, De Gruyter (2019)

Acknowledgments

The results were obtained with the help of these awesome people:

Serge Massar

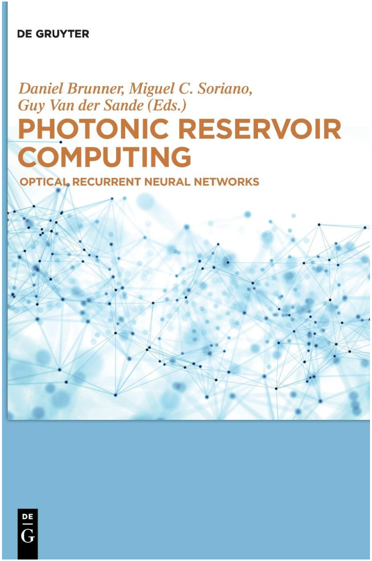
Daniel Brunner

Damien Rontani

Nicolas Marsal

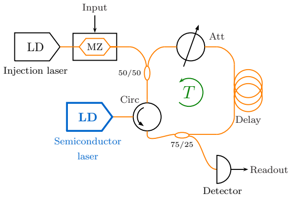
And these generous organisations:

For future reference

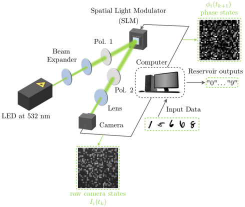


Why change the substrate?

- Photonics are getting better
- Moore's law doomed to fail
- New models for computation



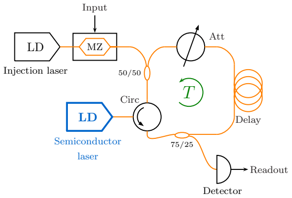
P. Antonik *et al.*, Photoniques 104 (2020)



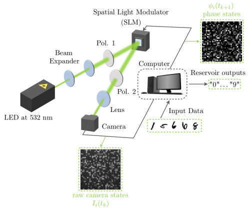
P. Antonik *et al.*, JSTQE 26 (2019)

Why change the substrate?

- Photonics are getting better
- Moore's law doomed to fail
- New models for computation



P. Antonik *et al.*, Photoniques 104 (2020)



P. Antonik *et al.*, JSTQE 26 (2019)

Thank you!