Tutorial: Reservoir Computing

D. Brunner, M. Jacquot, L. Larger

FEMTO-ST / Optics Dept., CNRS Univ. Bourgogne-Franche-Comté,

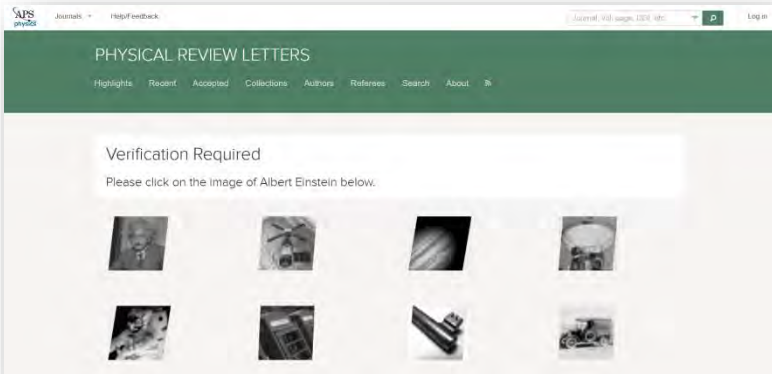
Besançon, France

Spintronics meets Neuromorphics, Spice, Mainz, Germany

October 11th, 2018

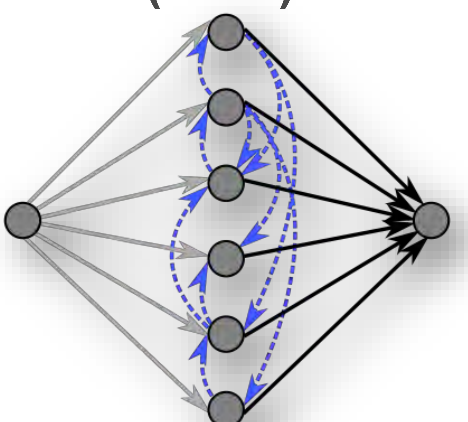
What is the problem: I. Classification

Sci-Hub: Where is Einstein?

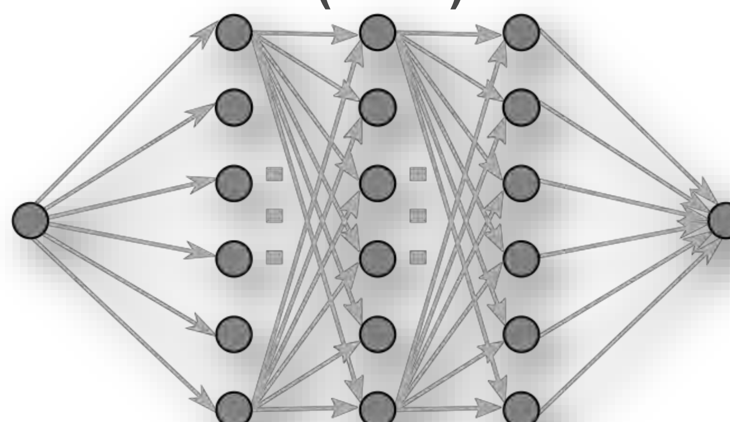


Postman: Cat /Dog ?

Recurrent Neural Network (RNN)



Deep Neural Network (DNN)

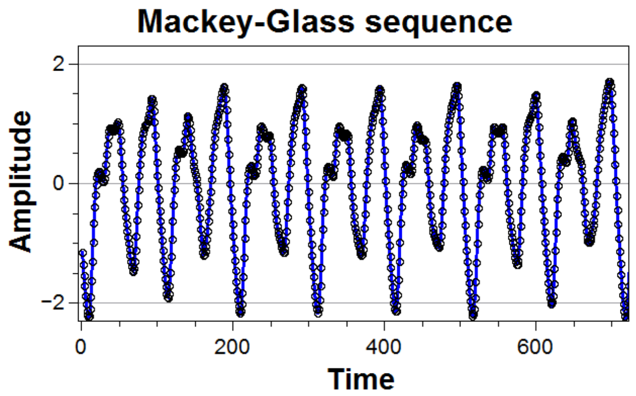


What is the problem: II. Prediction

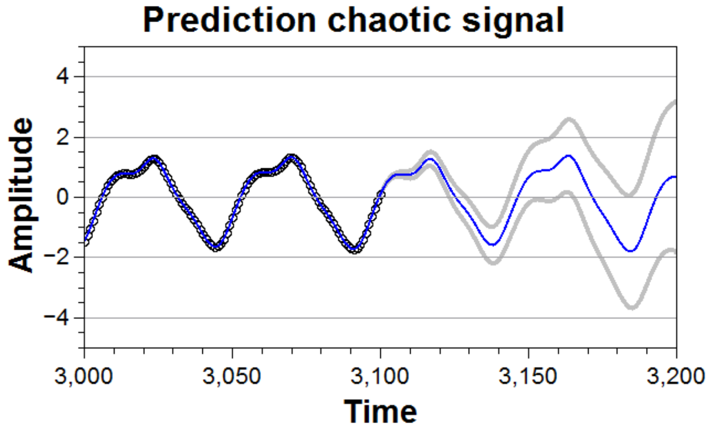
Algorithmic calculation:

$$y_{n+1} = y_n + \delta\left(\frac{0.2 y_t}{1 + {y_t}^{10}} - 0.1 y_t\right)$$

n=110000
Δt<0.1 s

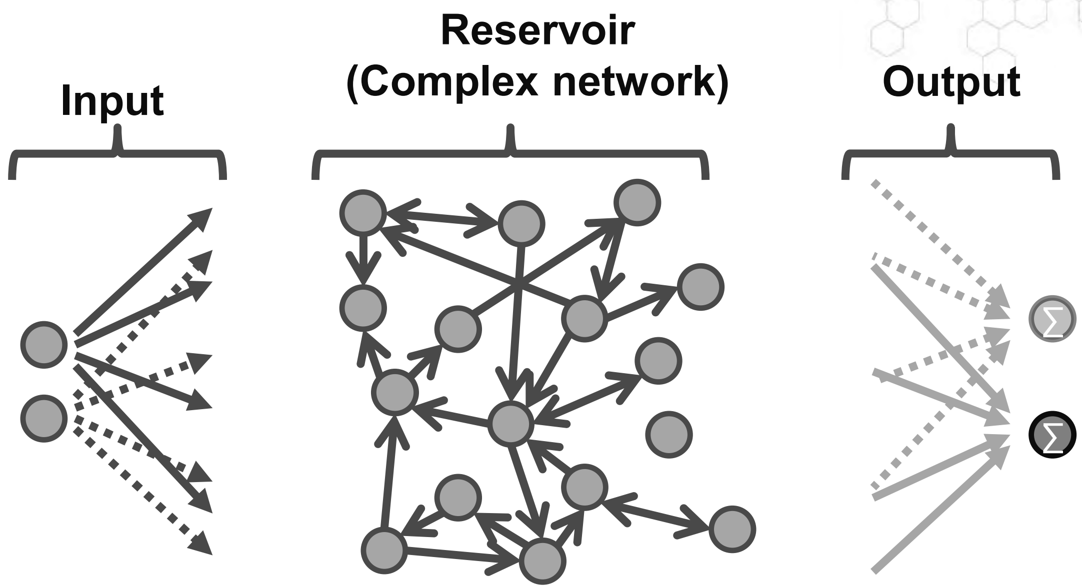


Complex, ambiguous behavior:



?

Reservoirs: Dynamical, Complex, Simple



Input:
Random injection

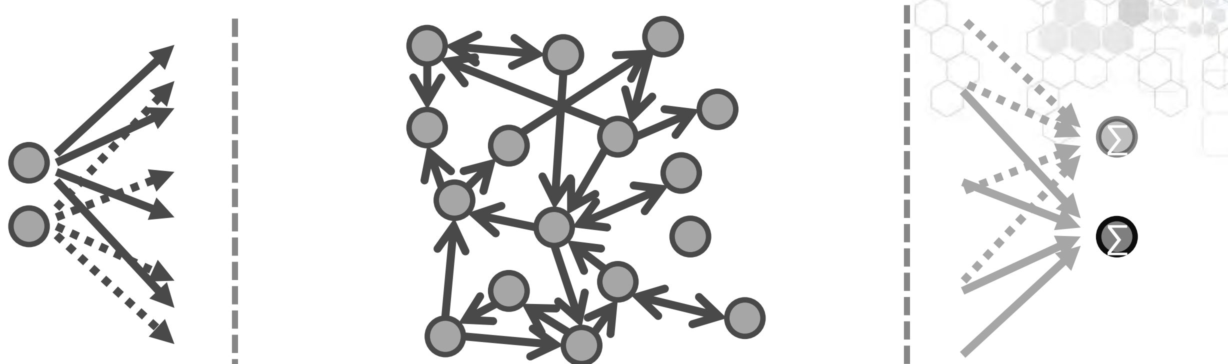
Reservoir:
Random connection
Resting state

Output:
Linear weights
Linear regression

- Simple: 2 random matrix multiplications
- Echo state property: working memory
- Excellent performance for prediction

Jaeger, et al. Science 2004

Operating a reservoir



Input:
$\boldsymbol{u}(n+1)$

Network state:
$\boldsymbol{x}(n+1) = \boldsymbol{f}[W\boldsymbol{x}(n) + W^{inj}\boldsymbol{u}(n+1) + \boldsymbol{b}]$

Output:
$\boldsymbol{y}(n+1) = W^{out}\boldsymbol{x}(n+1)$

1. Training data: set of inputs $\boldsymbol{u}(n)$ for which $\boldsymbol{y}(n)$ is known
2. Collect $\boldsymbol{x}(n)$ for $n \,\epsilon [1, \ldots, T]$
3. Cross-validate: randomly label instances of input data by $l \,\epsilon [1, \ldots, L]$
4. Select one $l_c \,\epsilon [1, \ldots, L]$
   - M_x: concatenated matrix of $\boldsymbol{x}(n)$ in response to $\boldsymbol{u}(n)$ for $l \neq l_c$
   - T^T: concatenated matrix of $\boldsymbol{y}(n)$ for $l \neq l_c$
   - Obtain $W^{out} = (M_x M_x^T + \lambda I)^{-1}(M_x T^T)$
5. Measure error for l_c

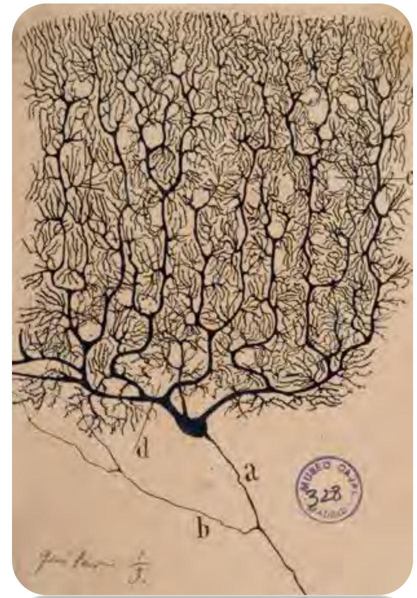
For hardware: Don't care about system ($\boldsymbol{x}(n+1) = \boldsymbol{f}[\boldsymbol{x}(n)]$)
- Good control of "distance" to criticality
- Good access to system state

Why novel hardware???

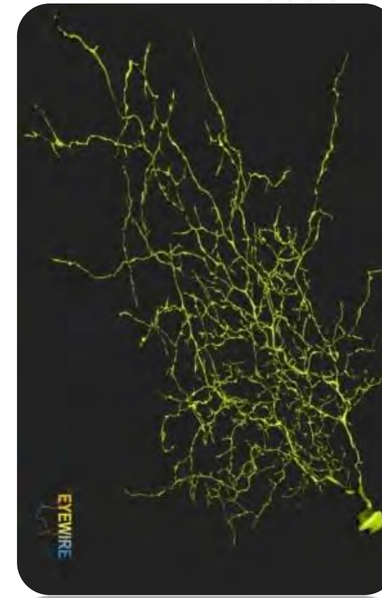
Big questions: What hardware

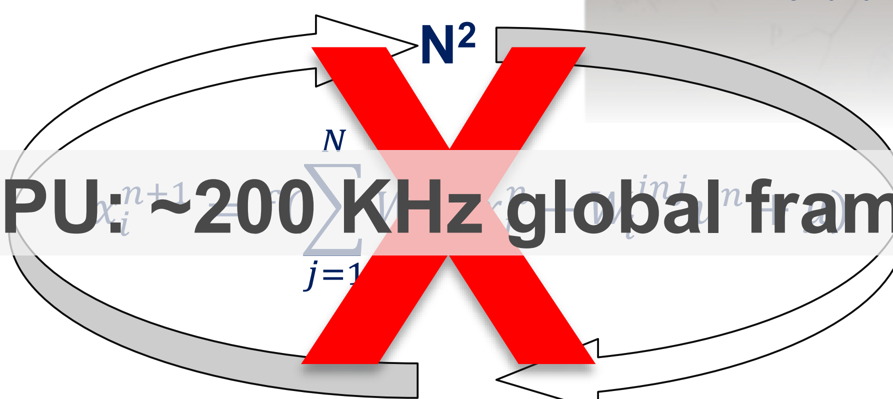
Neural Networks: What is special ???

- Large scale parallel dynamics
- Highly connected
- **Direct, physical links (no address routing)**
  - ✓ Fully parallel
  - ✓ Stand alone / unclocked
  - ✓ Ideally energy efficient (no routing overhead)



© Ramón y Cajal

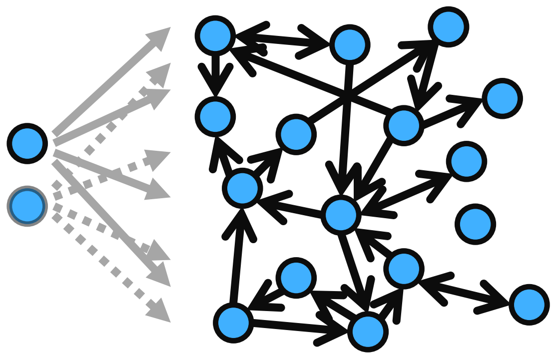




Even TPU: ~200 KHz global frame rate

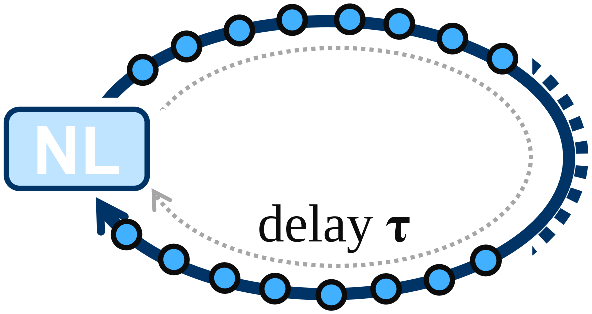
Photonic RC: delay networks

Spatially multiplexed reservoir



*Jaeger and Haas, Science **304,** 5667 (2004).*

Temporally multiplexed reservoir



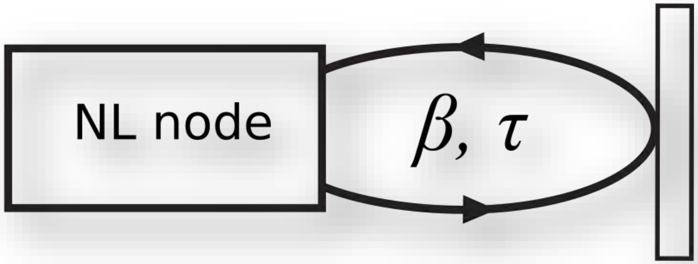
*L. Appeltant, et al., Nat. Comm. **2**, 468 (2011).*

- **Multiplexed in space:**
  - Injection weights constant in time
  - Data-rate: bandwidth of nodes
- **Hardware implementation challenging**
- **Flexible network-connectivity structure**

- **Multiplexed in time:**
  - Injection weights based on temporal modulation
  - Data-rate: bandwidth / N
- **Simplistic hardware implementation**
- **Potentially large memory**

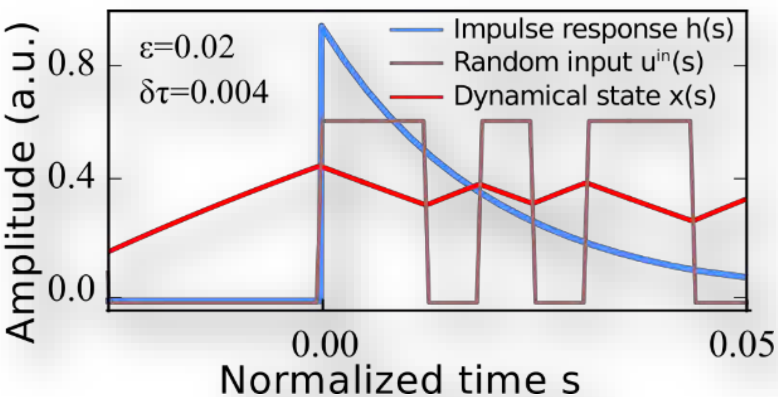
Neural Networks in Photonic Delay Systems

Delay systems 101

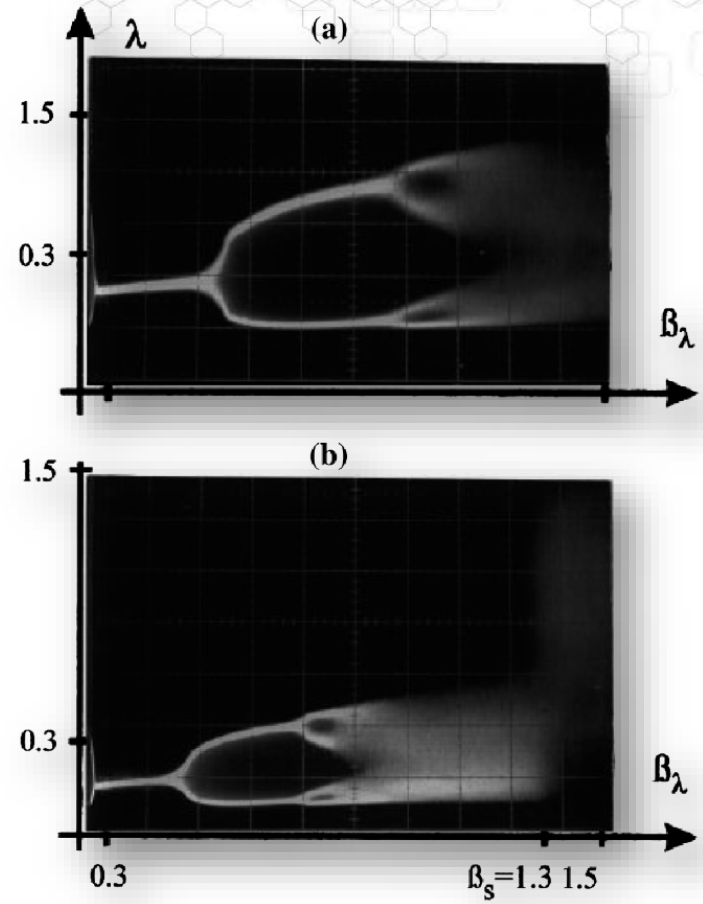


$$\varepsilon \dot{x}(s) + x(s) = f(\beta x(s-1), \boldsymbol{b})$$
$$s = t/\tau,\ \varepsilon = T_R/\tau$$

No feedback
Driven response



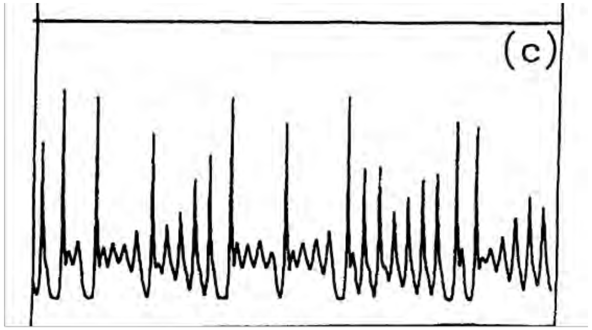
Feedback
No drive



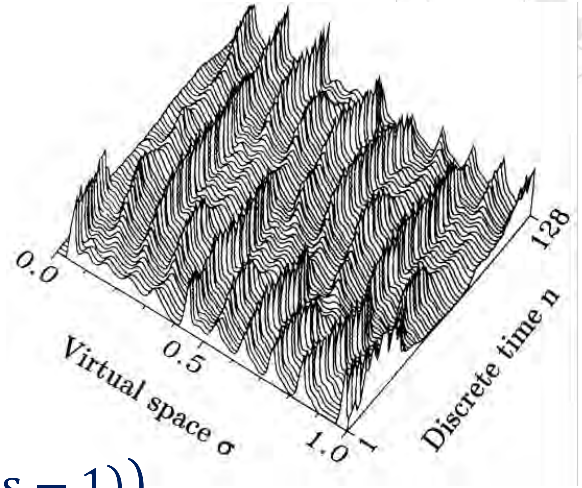
- **Highly simplistic system**
- **High dimensional**
- **Critical transition**

Soriano *et al.*, RMP **85**, 421 (2013).

Delay systems are ring networks



Arecchi *et al.*, Phys. Rev. A **45**, 4225 (1992).



$$\varepsilon\dot{x}(s) + x(s) = f\big(x(s-1)\big)$$

Solving via Green's function approach

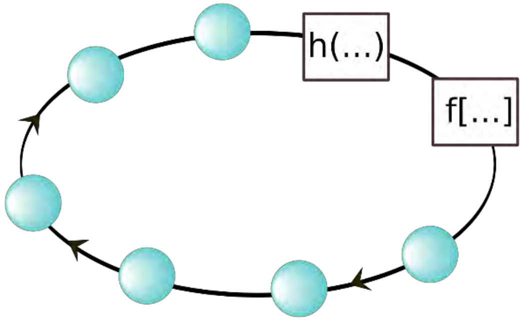
$x(s) = \int_{-\infty}^{s} h(s-\xi)\cdot f\big(x(\xi-1)\big)d\xi$, $h(s)$: Impulse response

Temporal reorganization

$$x_\sigma(n) = x(s), \sigma \in [0,1+\gamma], (1+\gamma)n + \sigma = s, n = 0,1,2,\ldots$$

n=time, σ=node

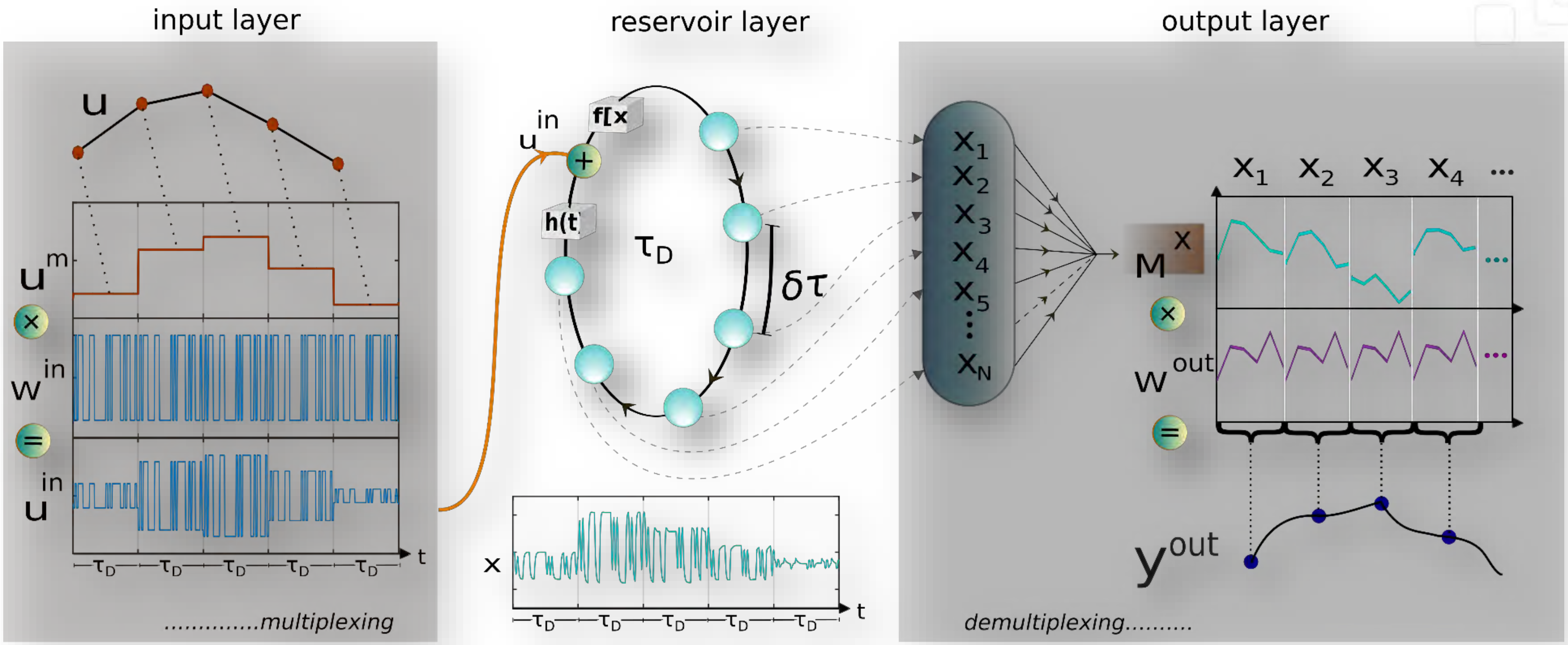
$$x_\sigma(n) \approx \int_{\sigma-1}^{\sigma} h(\sigma-\xi)\cdot f[x(\xi-1)]d\xi$$



Coupling = convolution with previous NL-transformed states

femto-st SCIENCES & TECHNOLOGIES

Delay reservoirs:

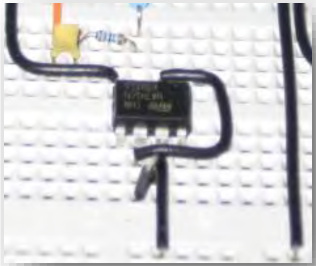


Input:
$W^{inj}\boldsymbol{u}(n+1)$
W^{inj} is temporal modulation

Network state:
$\varepsilon\dot{x}(s) + x(s) =$
$f\big(x(s-1), \boldsymbol{u}(s)\big)$

Output:
$W^{out}\boldsymbol{x}(n+1)$
W^{out} is temporal modulation

Tour de ville for delay substrates

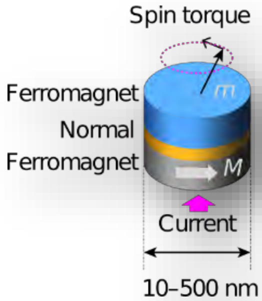


Electronic:
- Mackey-Glass nonlinearity

Appeltant *et al.*, Nat. Comm. 2, **468** (2011).

Opto-electronic:
- Ikeda nonlinearity

Larger *et al.*, Opt. Exp. **20**, (3) (2012).
Paquot *et al.*, Sci. Rep. **2**, 287 (2012).



Nano-electronics:
- Spin-torque oscillator

Torrejon, *et al.*, Nature **547,** 7664 (2017).

All optical:
•Semicondcutor devices

Brunner *et al.*, Nat. Commun. **4**, 1364 (2013).
Duport *et al.*, Opt. Express **20**, 22783 (2012).

Photonic RC: Van der Sande, et al., Nanophotonics **6**, 561 (2017)
Photonic delay RC: Brunner, et al., JAP, Special issue (2018)

Neural Networks in spatio-temporal photonic systems

Why Optics? Connections!

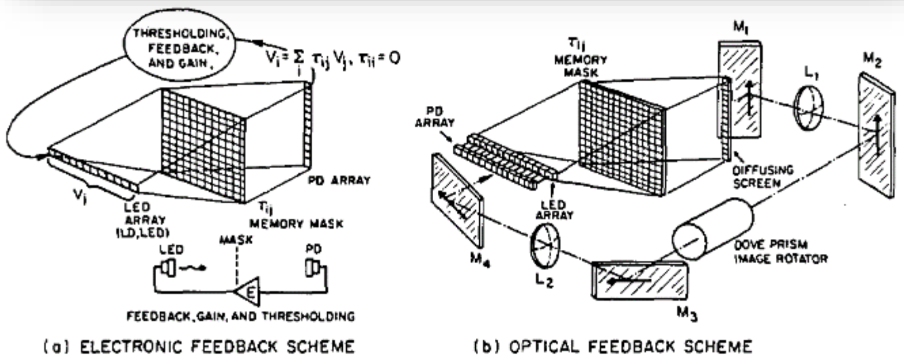
98 OPTICS LETTERS / Vol. 10, No. 2 / February 1985

Optical information processing based on an associative-memory model of neural nets with thresholding and feedback

Demetri Psaltis and Nabil Farhat*

Department of Electrical Engineering, California Institute of Technology, Pasadena, California 91125

Received July 9, 1984; accepted November 15, 1984

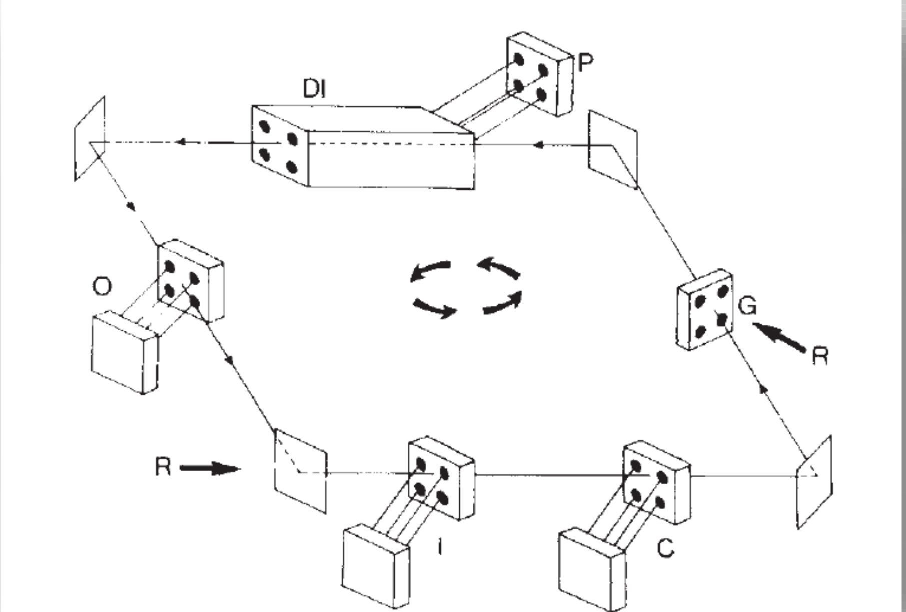


NATURE VOL. 316 25 JULY 1985 — LASERS REVIEW — 319

Lasers, nonlinear optics and optical computers

S. D. Smith

Department of Physics, Heriot-Watt University, Edinburgh EH14 4AS, UK



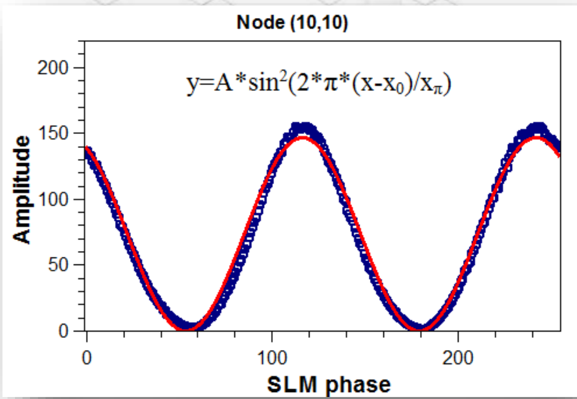
Why optical connections?

- Parallel
- 2D substrate not sufficient
- Negligible distance tradeoff
- Zero induction

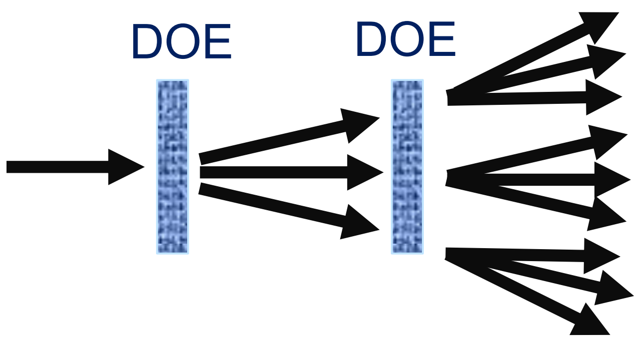
Photonic neural network

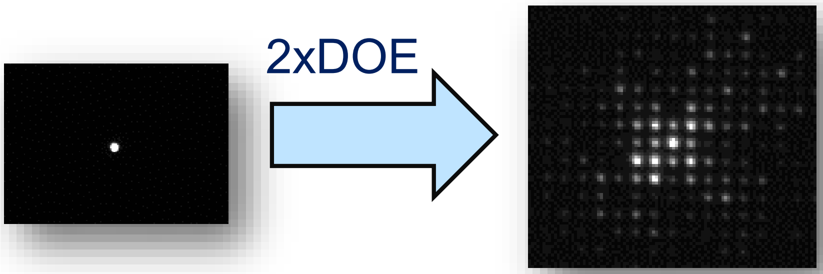
Neurons: liquid crystal pixels

Polarization filtering



Network: holographic fan-out



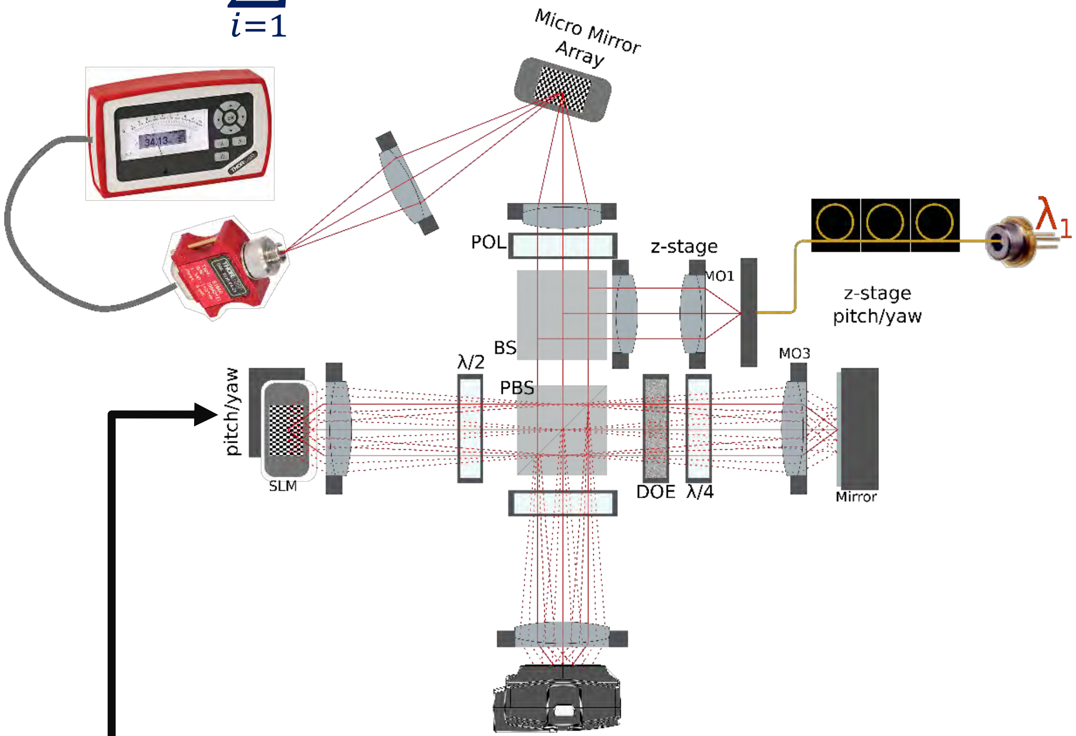


- Parallel state, analogue
- Single element

- DOE > 90.000 nodes
- SLM > 480.000 nodes

Recurrent coupling:

$$y^{n+1} \propto |\sum_{i=1}^{N} W_{i,k}^{DOE}\,(E_i^0 - E_i^{n+1})|^2$$



Matlab:

$$I_i^{n+1} = \sin^2(\beta|\sum_{j=1}^{N} \kappa_{i,j}^{DOE}\,E_j^n|^2 + \gamma\kappa_i^{inj}u^{n+1} + \Theta_0)$$

Electro-optical networks

- **Nodes: Pixels of SLM**
  - Density: 6400 / mm²
- **Coupling: DOE + imaging**
- **4f architecture:**
  - Self-coupling
- **2025 network nodes**
- **Dynamic evolution: iterative upda Camera -> SLM**

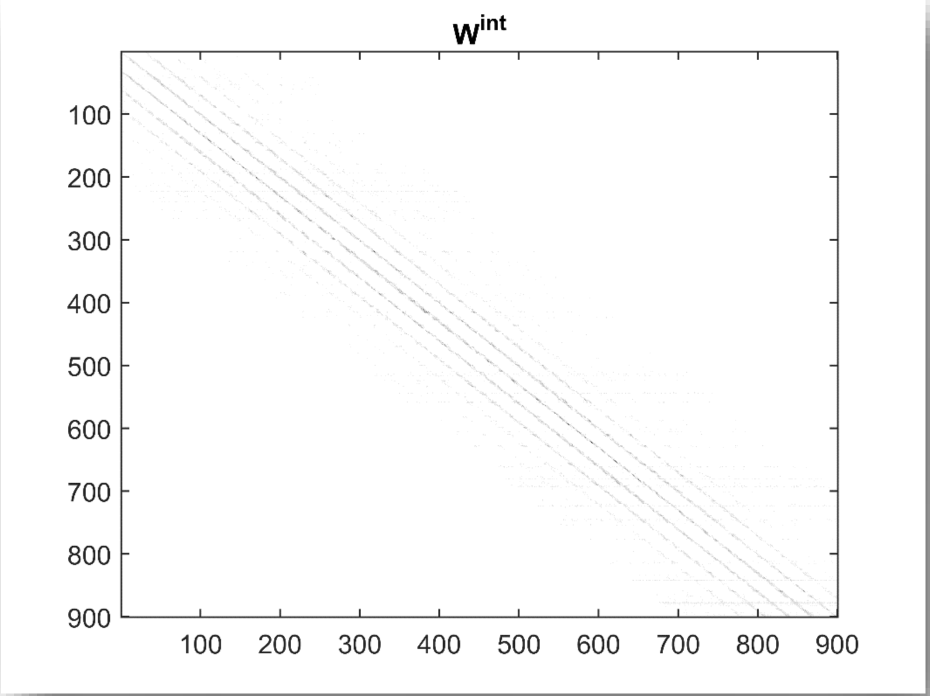
Readout weights:

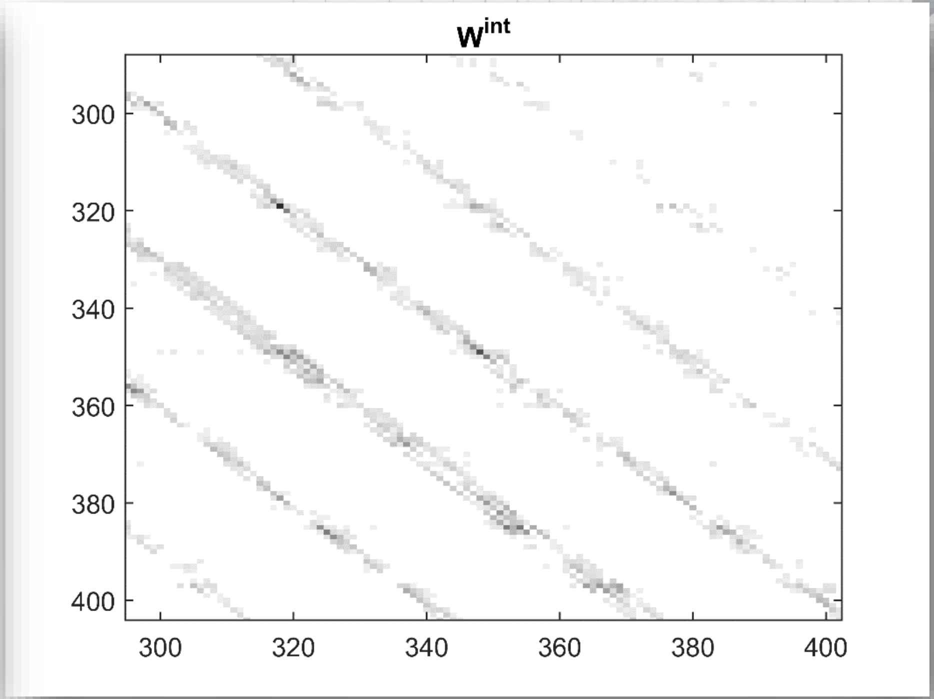
- Digital Micro-mirror Array
- Image SLM onto DMD
- Result: analog power meter

BUT

- Full state not known
- State detection non-linear
- **Matrix inversion not possible**

Coupling matrix



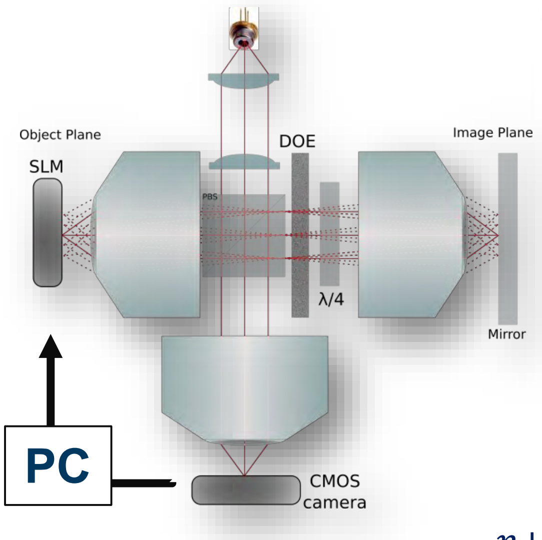


<u>Diffractive networks: coupling a neighborhood</u>

- **For 45x45 nodes: 100 KFlOP / per state**
- **For 300x300 nodes: 4.5 MFLOP / per state**
- **More complex / larger range trivial to create**

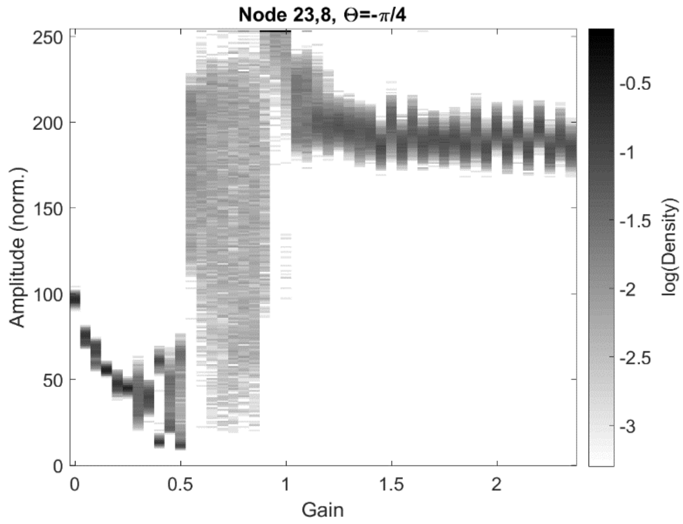
Bueno, et al., under review *Optica*.

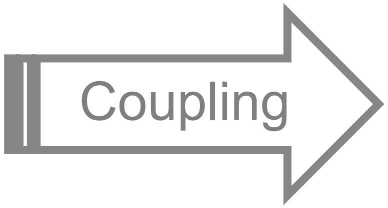
Autonomous dynamics

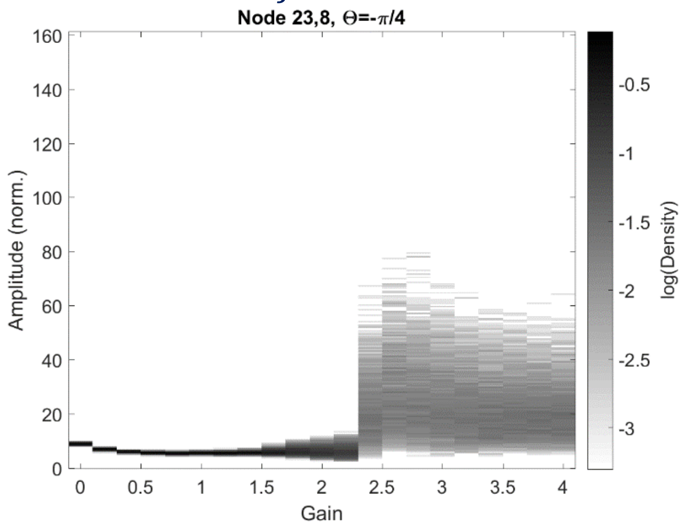


$$I_i^{n+1} = \sin^2(\beta I_j^n + \Theta_0)$$

$$I_i^{n+1} = \sin^2(\beta(\sum_{j=1}^{N} \kappa_{i,j}^{DOE} E_j^n)^2 + \Theta_0)$$

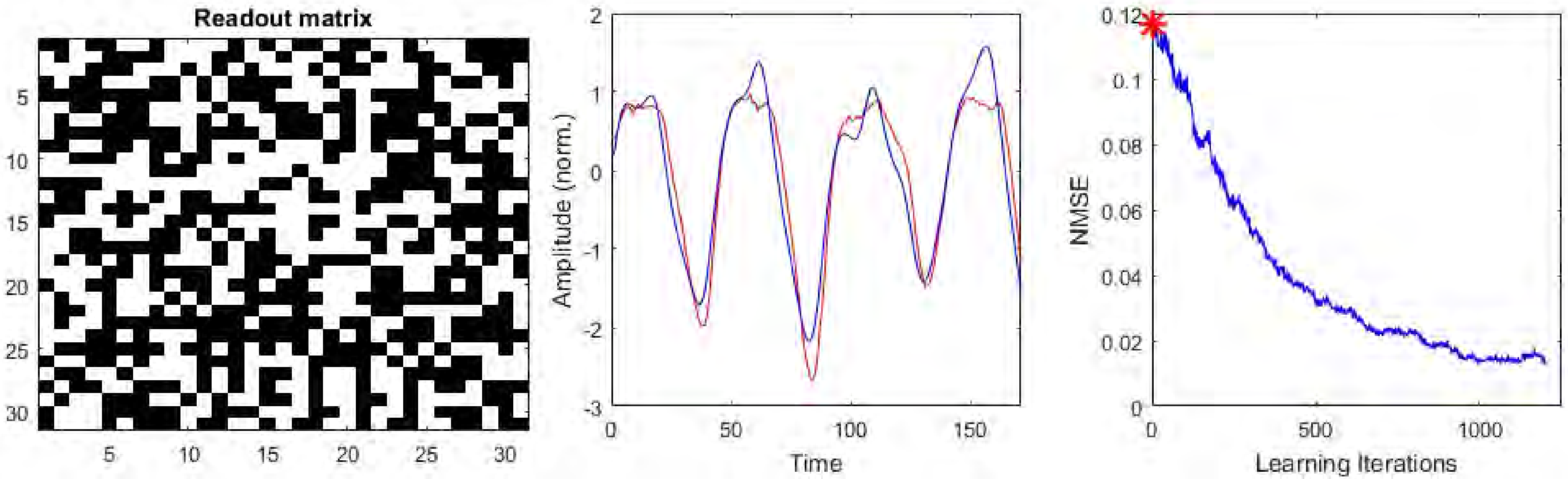






1. Training of Boolean readout

1. **Select mirror with bias to "not yet modified"**
2. **Flip mirror (1 → 0, 0 → 1)**
3. **Record output for 200 steps**
4. **Compare current with previous NMSE**



- **Error reduces efficiently**
- **Close to state of the art (remember: Boolean weights only)**

Bueno, et al., under review *Optica*.

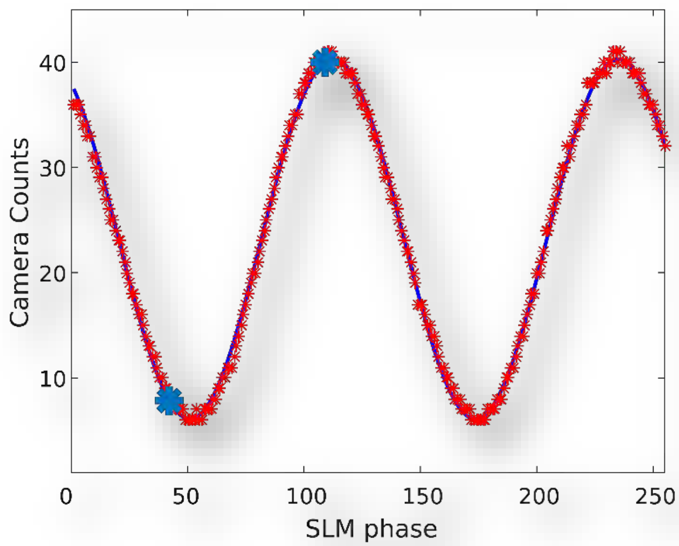
No negative values (uni-polarity)

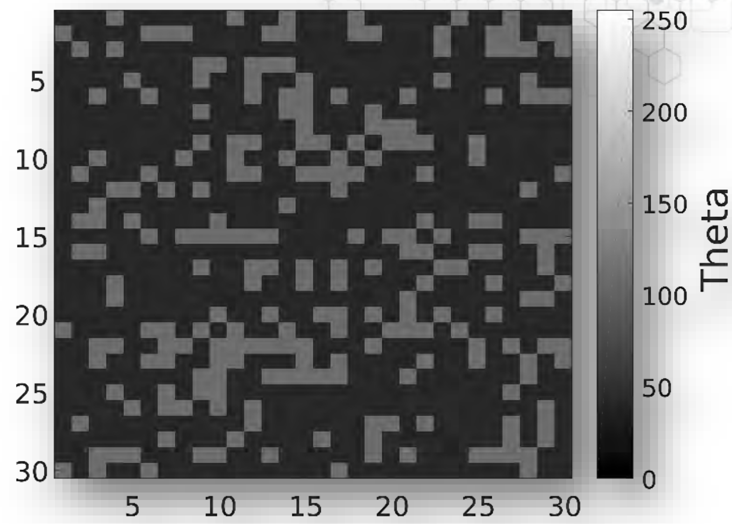
1. SLM: plane wave illumination
2. Polarization filtering: only constructive interaction
   - Only positive addition possible
3. Diffractive coupling: only constructive interaction
   - Only positive addition possible
4. Boolean readout weights: 0 / 1
   - Only positive addition possible

\+ No phase effects: stable

\- Connection weights always positive!

A way around unipolar weights





$$I_i^{n+1} = \sin^2(\beta|\sum_{j=1}^{N} \kappa_{i,j}^{DOE} E_j^n|^2 + \gamma \kappa_i^{inj} u^{n+1} + \Theta_0)$$

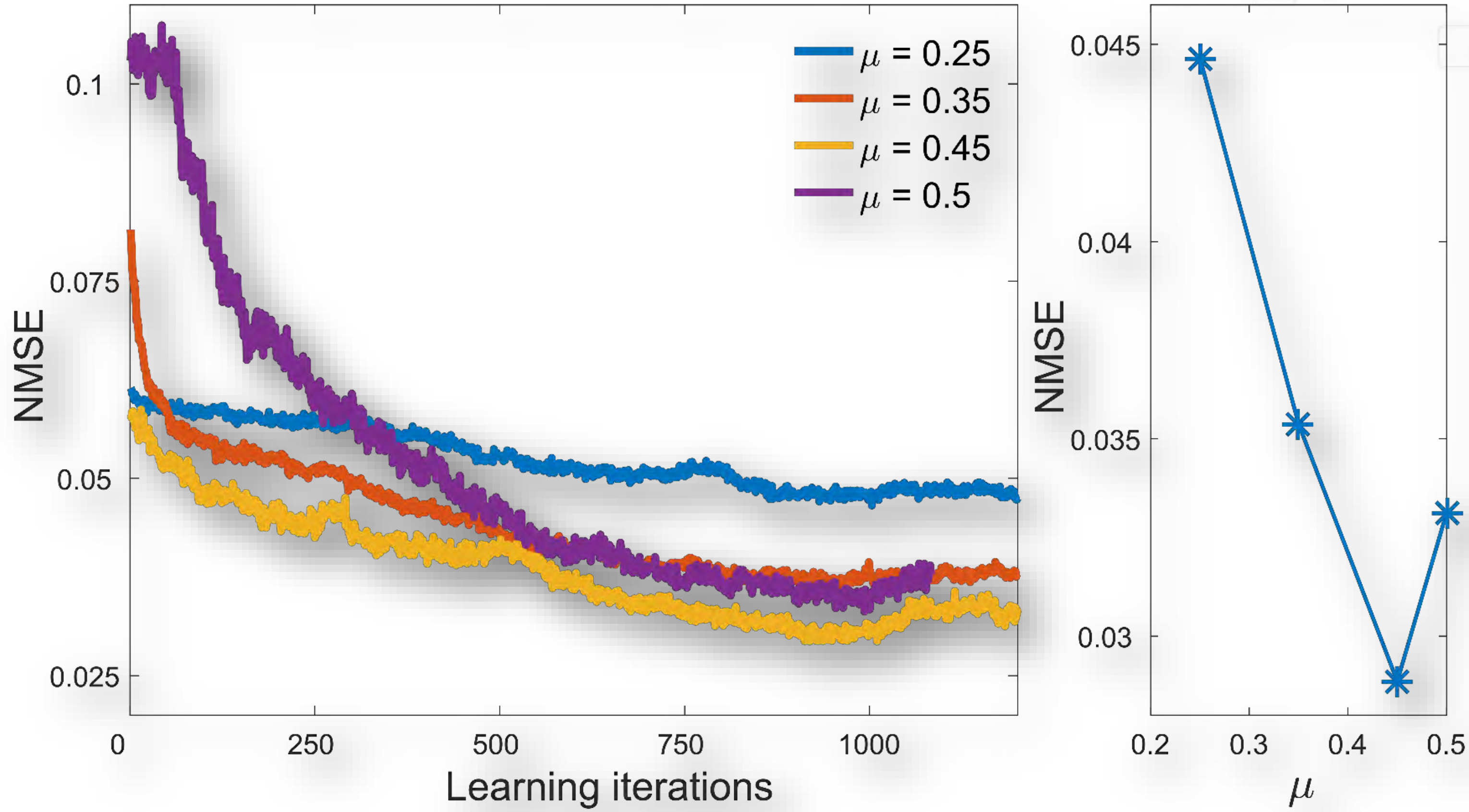
$\boldsymbol{\Theta_0}$ is a Matrix:

- $(1-\mu)$ – Values: $\Theta_0 = 42$
- (μ) – Values: $\Theta_0 = 109$

- Problem with unipolarity: no negative ‘slopes’
- Solution: harvest periodic nonlinearity

Bueno, *et al.*, *Optica* **5**, *756 (2018)*.

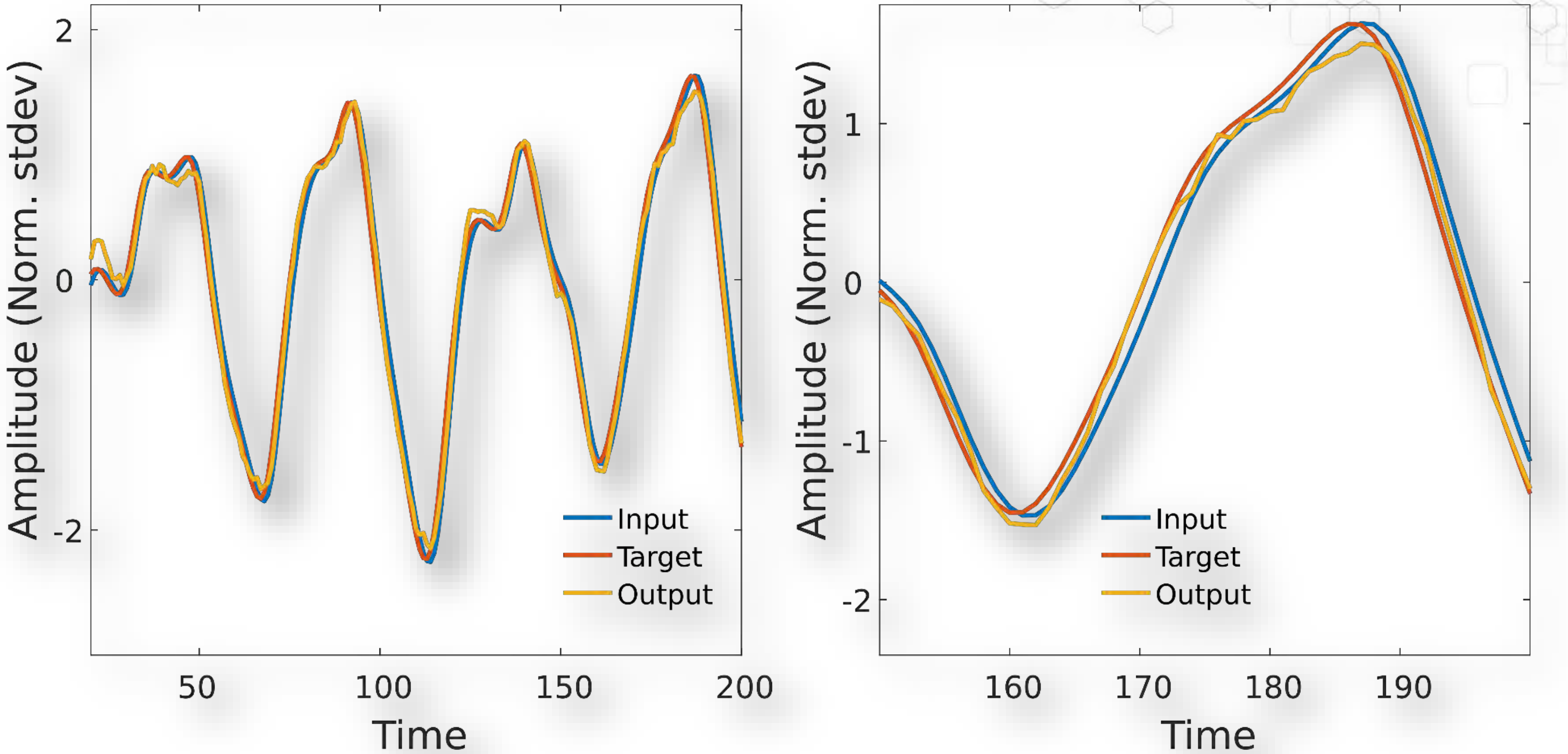
No negative weights



- Distribute operation points strongly aids performance
- Best points: close to 50/50 division

Bueno, et al., under review *Optica*

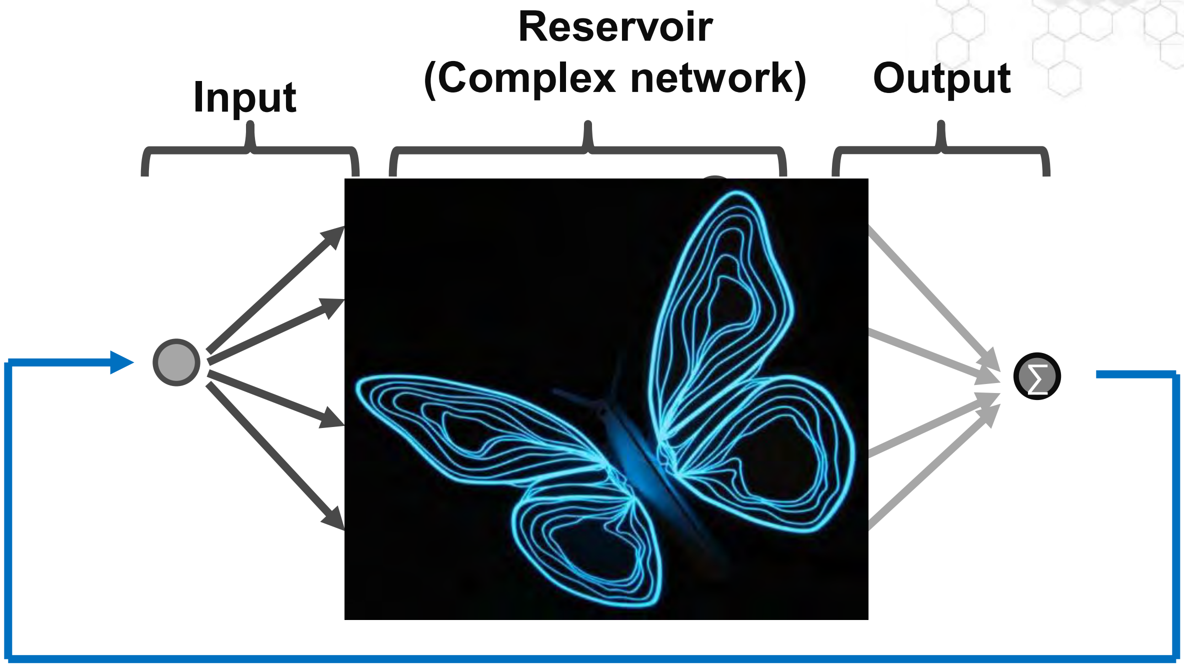
1. Training of Boolean readout



1. **Clear orientation toward "one-step leader"**
2. **Divergence largely close to local extrema**

Bueno, et al., under review *Optica*

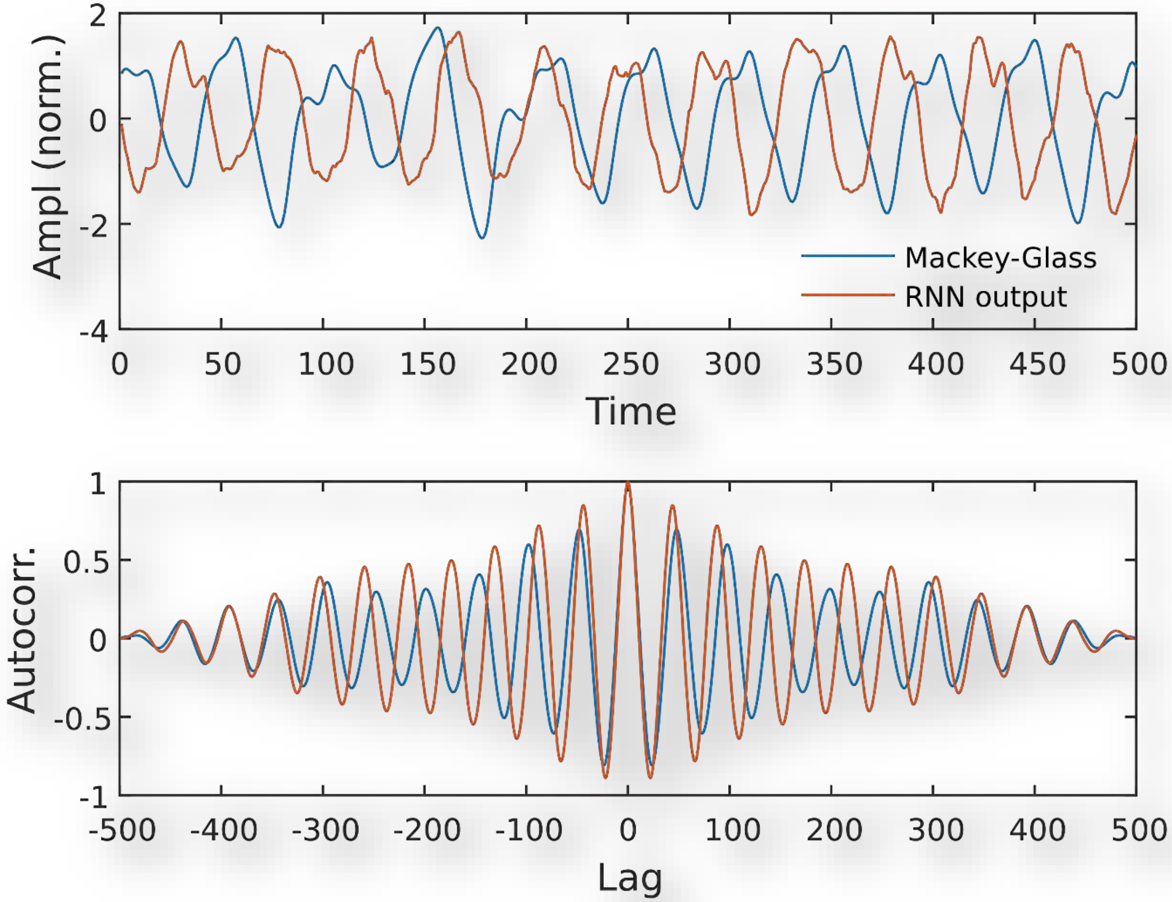
2. Feedback of RNN output



- **Prediction target: future state**
  - ➢ Output approximates future input
- **Output feedback: self consistent, autonomous system**
- **Morphing into target system (Neuronal Network → Butterfly effect)**
- **Important: principle of motor control**

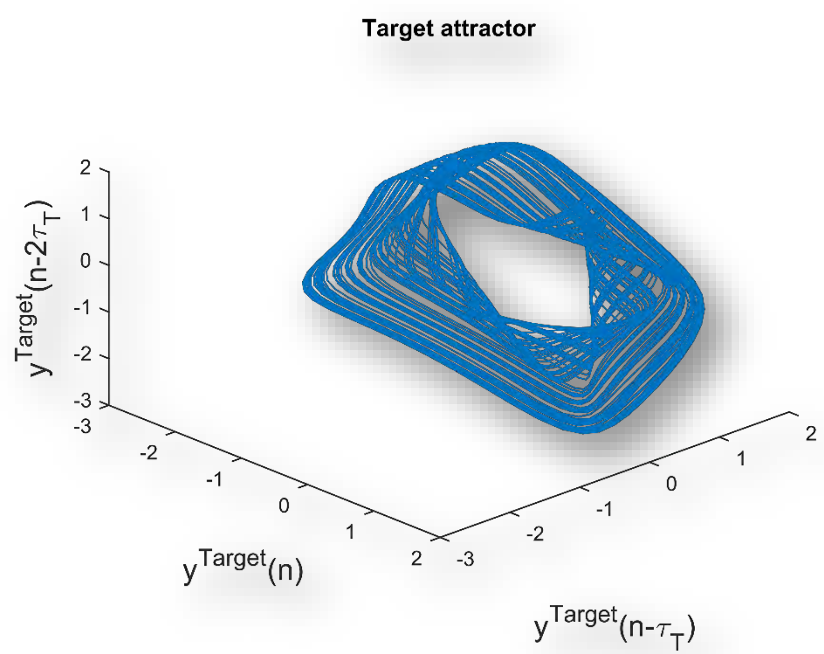
Image:
© http://testoil.com/did-you-know/the-butterfly-effect/

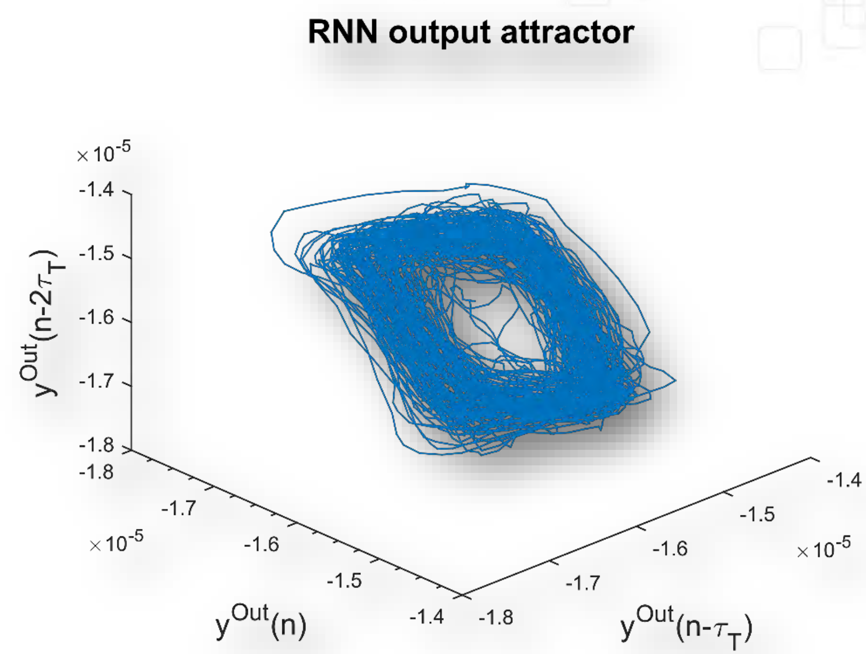
2. Feedback of readout result



- **RNN creates autonomous, nonlinear oscillator**
- **Period very close to target**

2. Feedback of RNN output





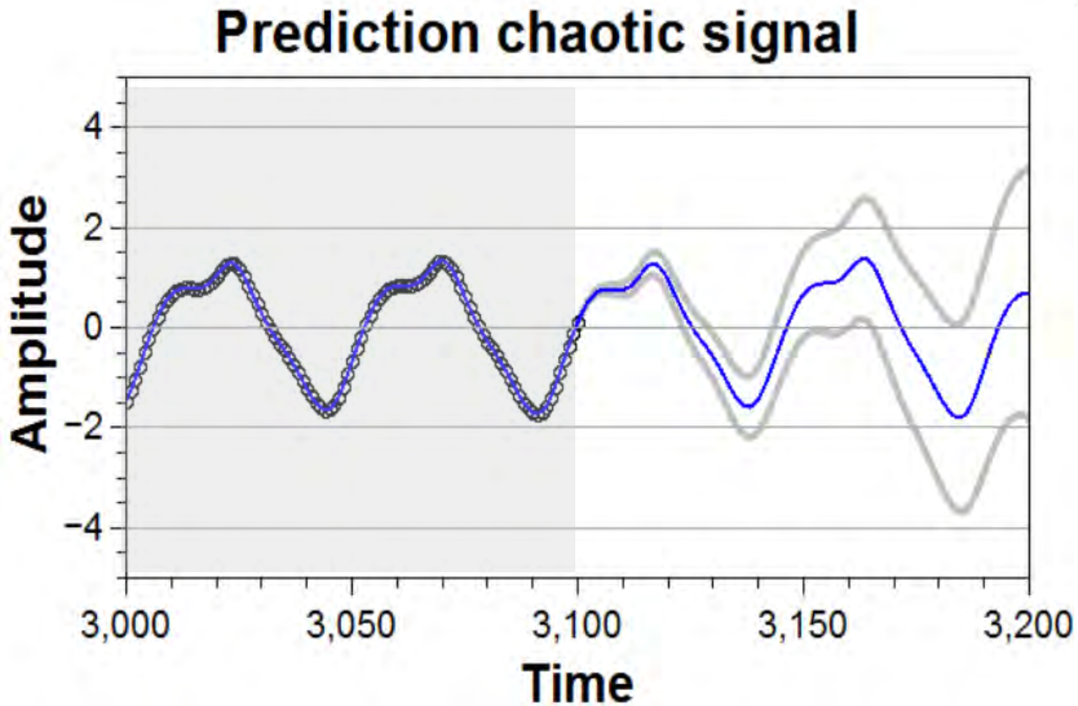
- **Output reassemble MG attractor**
- **Readout weights fully passive**
- **Readout weights have no bandwidth limitation**

How do Neural Networks predict (chaos)

What is chaos?

Mackey-Glass system

$$\frac{dy(t)}{dt} = \beta \frac{y(t-\tau_m)}{1+[y(t-\tau_m)]^\alpha} - \gamma y(t)$$



Chaotic systems are **sensitive to initial conditions:** point in phase space not suffficient

This sensitivity creates **unpredictability:** impossibility to determine future development for all times.

M. C. Mackey and L. Glass, *Science* **197**, 287 (1977).

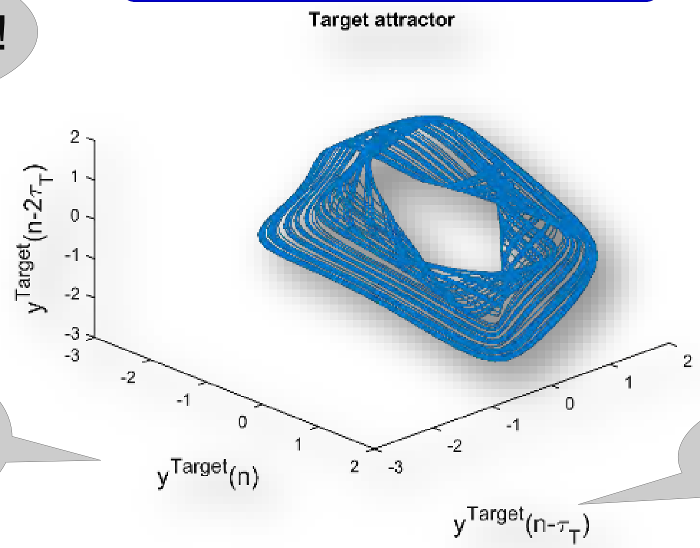
State space chaotic systems

Takens embedding theorem

Theorem 1. *The time delayed version of one time series suffices to reveal the structure of an attractor. Let us represent the data in M-dimensional space by the vectors* $\boldsymbol{x} = [y(t),\ y(t-\tau_0),\ \ldots,\ y(t-(M-1)\tau_0]^{\dagger}$. *Where* $(\cdot)^{\dagger}$ *as transpose matrix. The pair dimension-delay for the embedding* (M, τ_0) *contributes to reconstruct the right object in the state space.*

Mackey-Glass attractor

reconstructed!



original

reconstructed!

Temporal position of delay vectors:

- $\min|(AC(x(t))|$

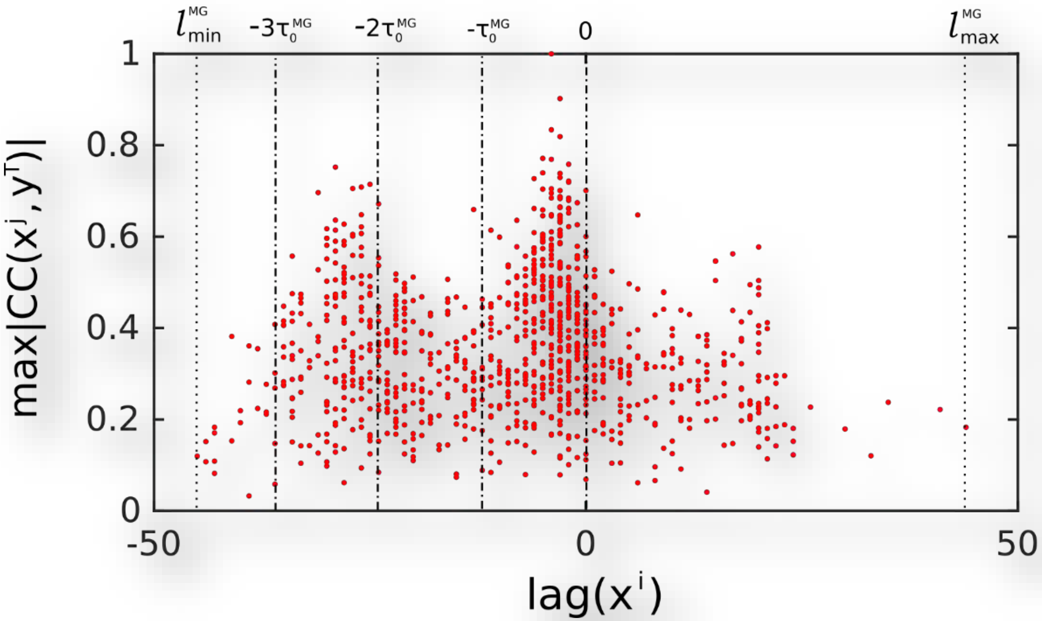
or

- $\min(MI(x(t), x(t-n))$

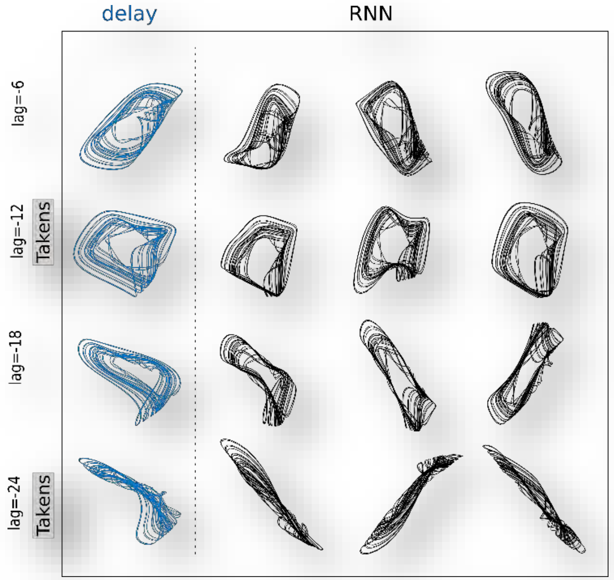
F. Takens, Detecting strange attractors in turbulence, Dynamical Systems and Turbulence, Lecture Notes in Mathematics, 1981.

Attractor reconstruction in rRNNs

Cross-correlation analysis for Mackey-Glass for μ=1.3



Attractor reconstruction from the rRNN



According to Cross-correlation analysis,

RNN does:

1. **Takens-like embedding**
2. **Increases sampling**

Characterize nearest neighbors:

Random Projections theory

Proposition *For any positive constant values ϵ_1, ϵ_2. Let V be a collection of S points $\{\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \ldots, \mathbf{y}^{(S)}\} \in \mathbb{R}^q$, with distances computed under L_2 norm. There is a map $\varphi : \mathbb{R}^q \to \mathbb{R}^h$, such that for all $\mathbf{y}^{(i)}, \mathbf{y}^{(j)} \in V$,*

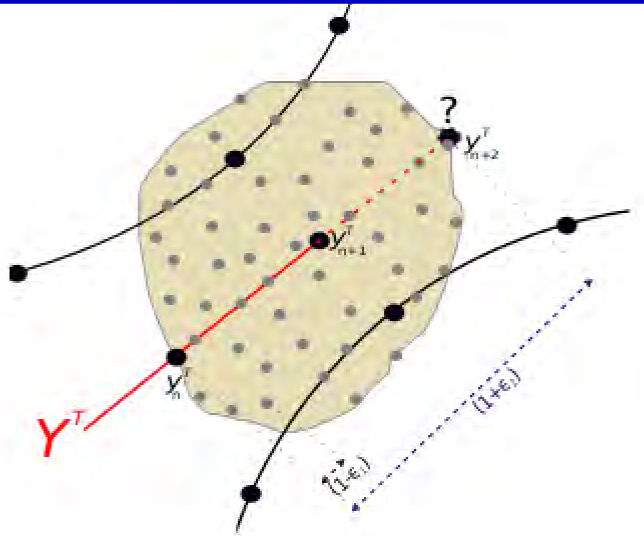
$$(1-\epsilon_1)\|\mathbf{y}^{(i)} - \mathbf{y}^{(j)}\| \leq \|\varphi(\mathbf{y}^{(i)}) - \varphi(\mathbf{y}^{(j)})\| \leq (1+\epsilon_2)\|\mathbf{y}^{(i)} - \mathbf{y}^{(j)}\|.$$

Example of a distribution of nearest neighbor states in rRNN

This proposition states that the **distances between two consecutive states** of the attractor are bound to the range

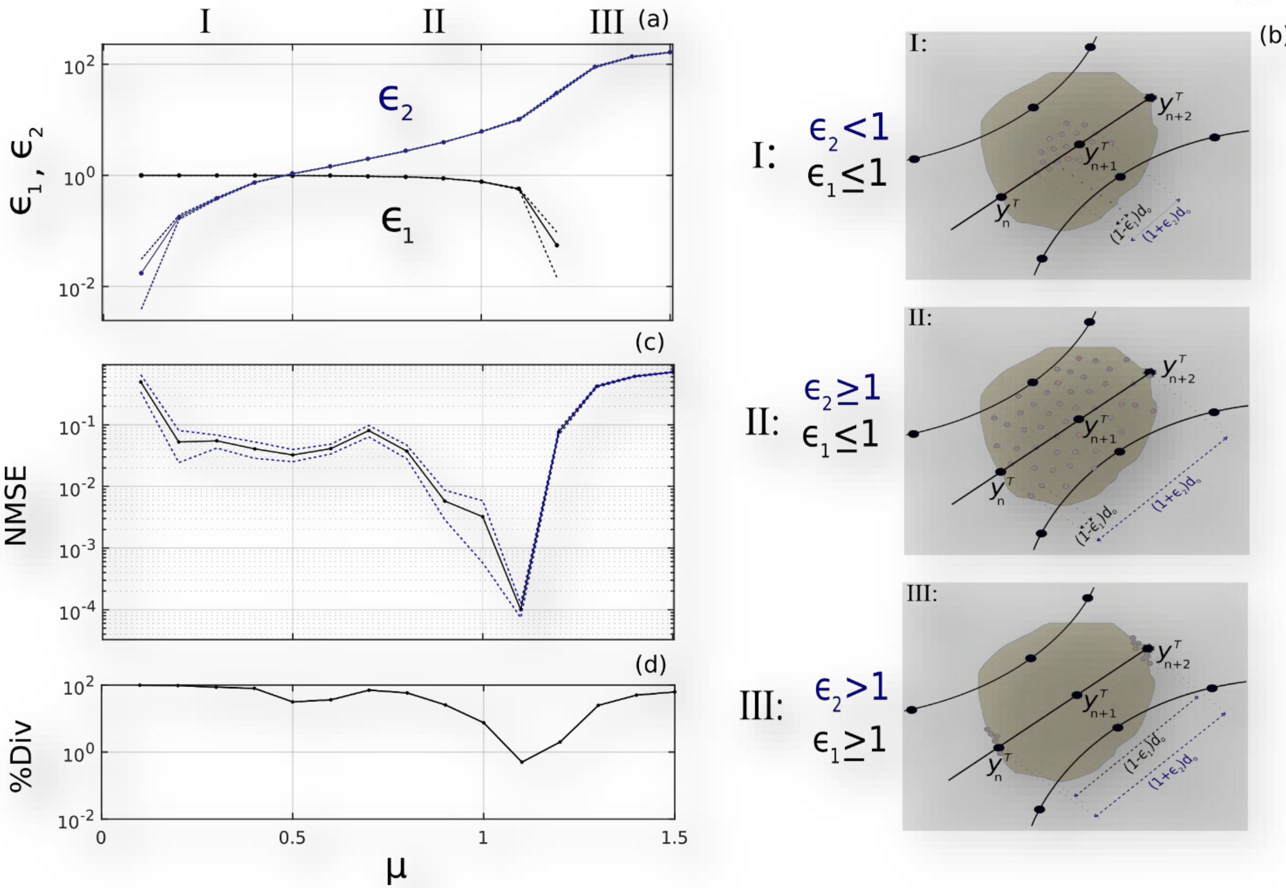
$$[(1-\epsilon_1), (1+\epsilon_2)]$$

where ϵ_1, ϵ_2 are arbitrary constant values.



NN limits according to Random Projection:

7



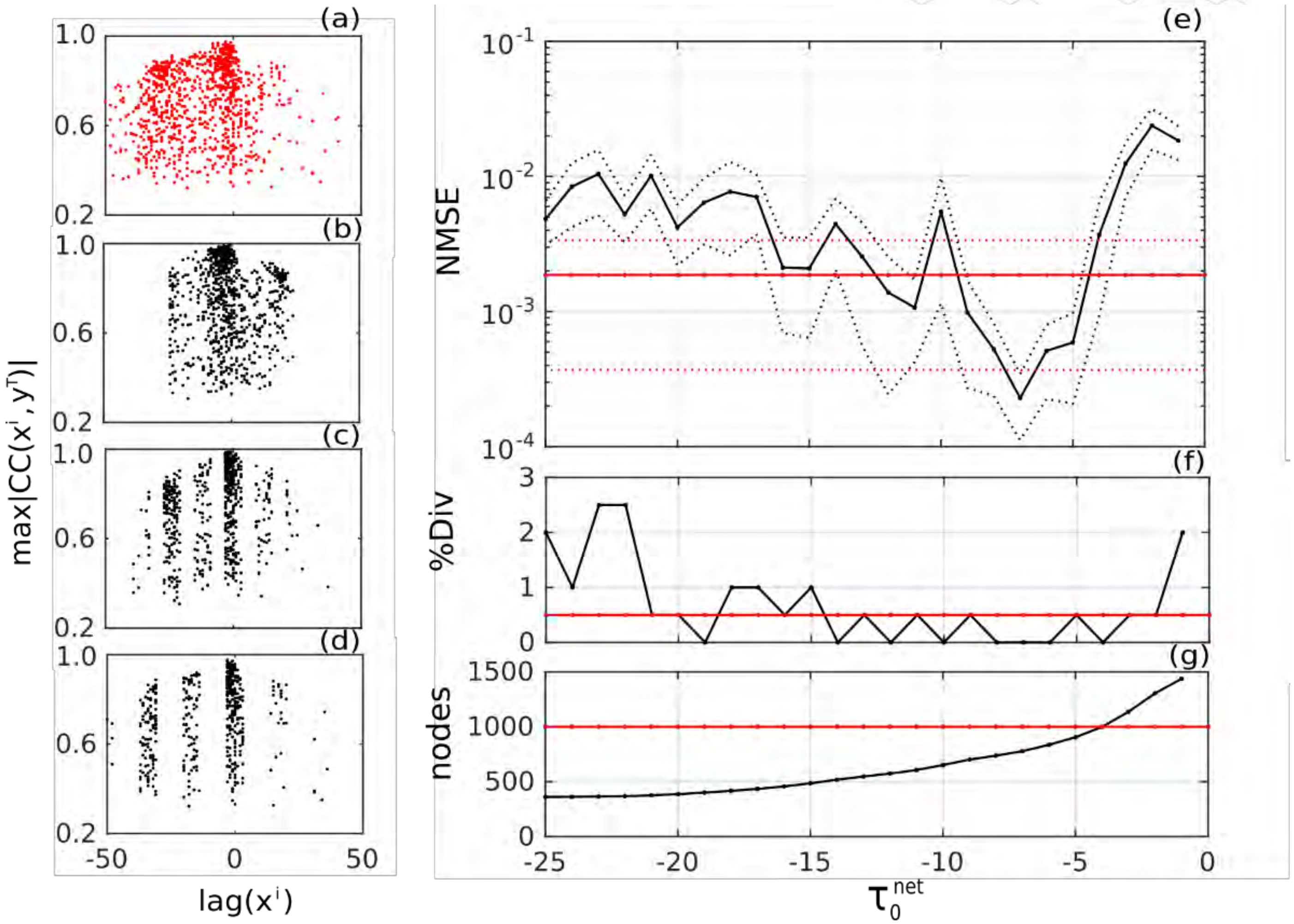
I. Sampling **dense** but **short-range**

II. Sampling **dense** and **long range**

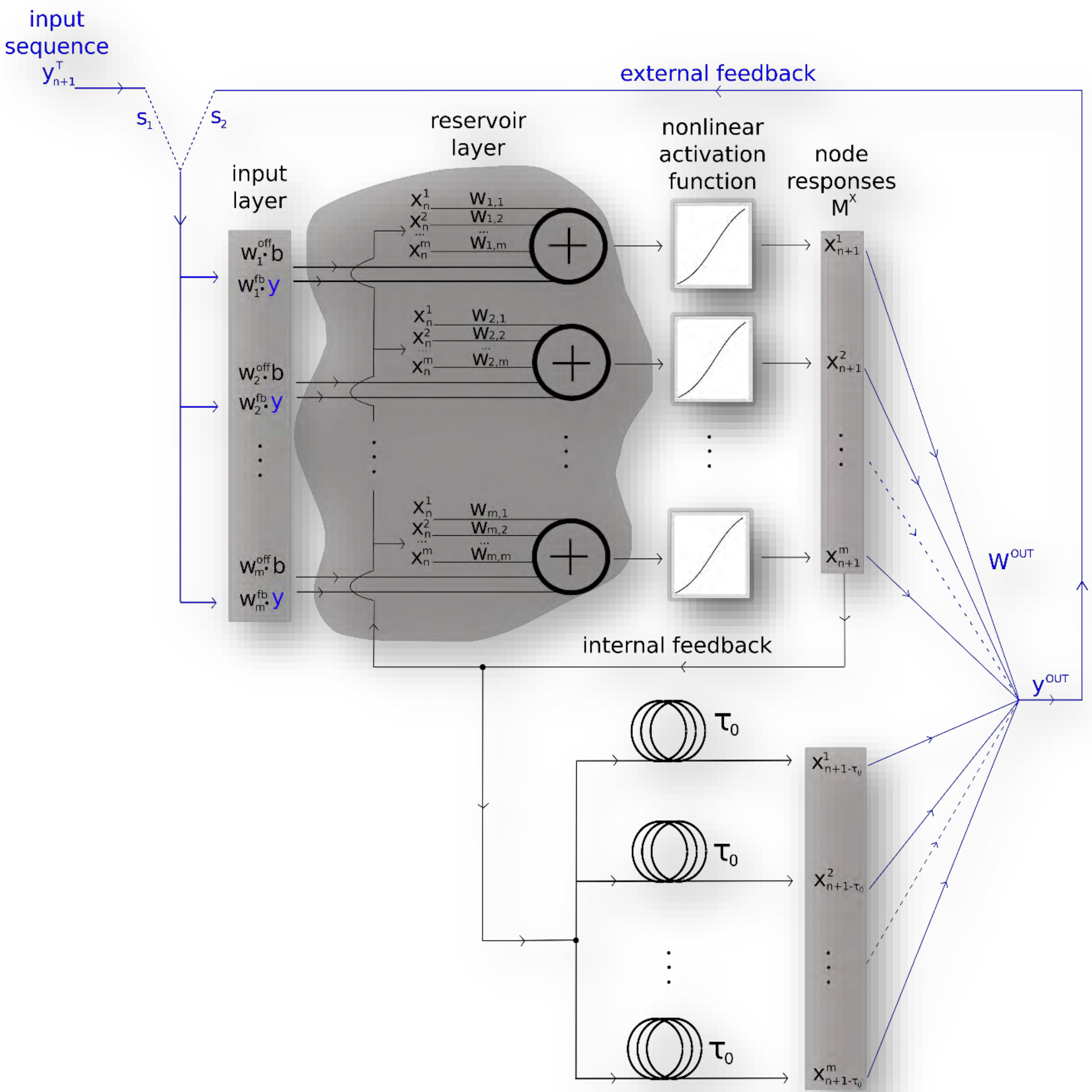
III. Sampling **not dense** but **long-range**

Limits of nearest neighbour-distances bind good prediction conditions

NN limits according to Random Projection:



New scheme: Taken RNN



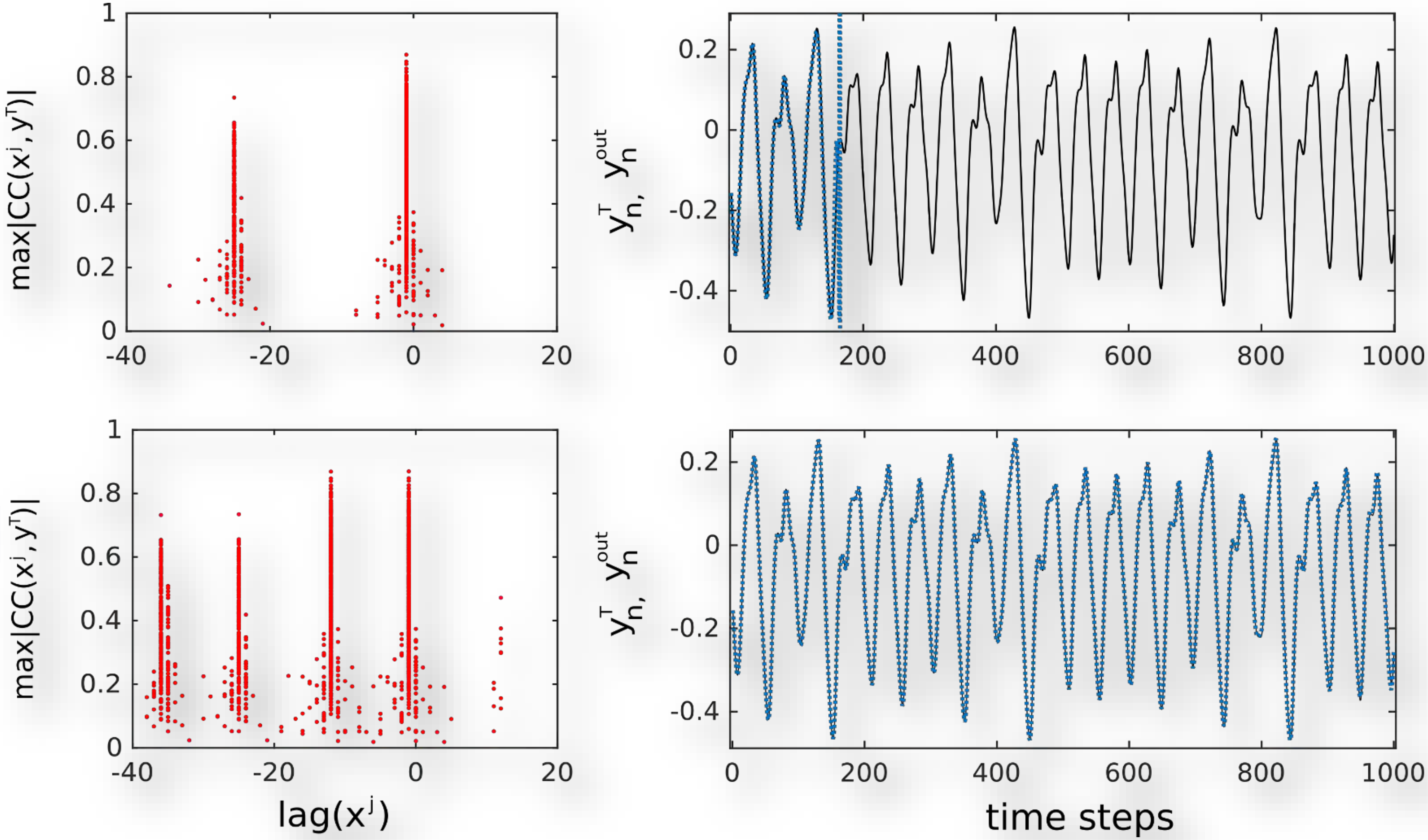
Introduce first-in first-out memory specifically with depth of Taken embedding delay

Operation at BAD parameters

Cross-correlation analysis for Mackey-Glass for μ=0.2

rRNN

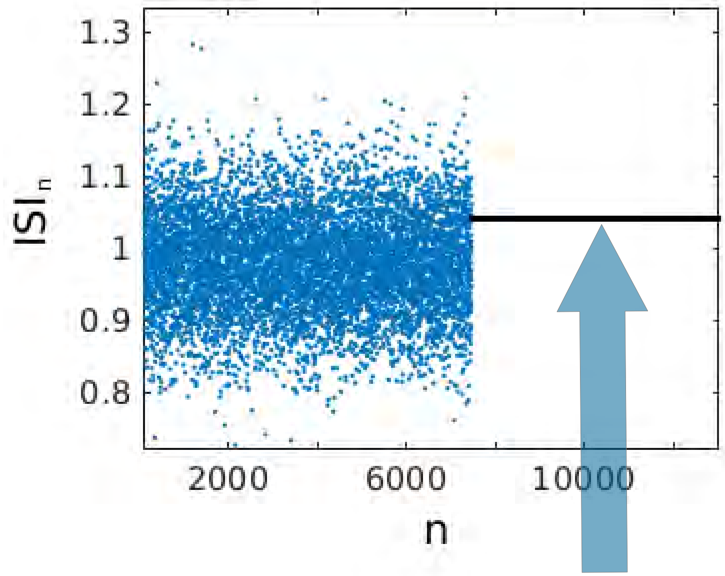
TrRNN

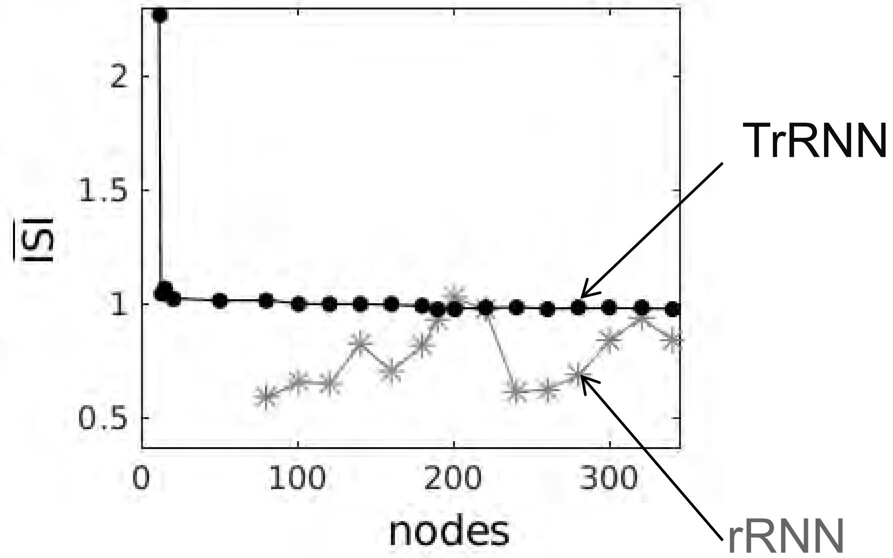


Applied to model for cardiac arrest

Interspike intervals (ISI) of an arrhythmic excitable system comparable to a heart

Comparison between the stabilized mean of the TrRNN and the classical rRNN





Stabilization of the system based on our TrRNN

The TrRNN requires 15 times less nodes, simultaneously achieving superior performance.

A. Garfinkel, M. L. Spano, W. L. Ditto, J. N. Weiss, *Science* **257**, 1230 (1992).

Marquez, et al., IEEE *ICRC* (2017).

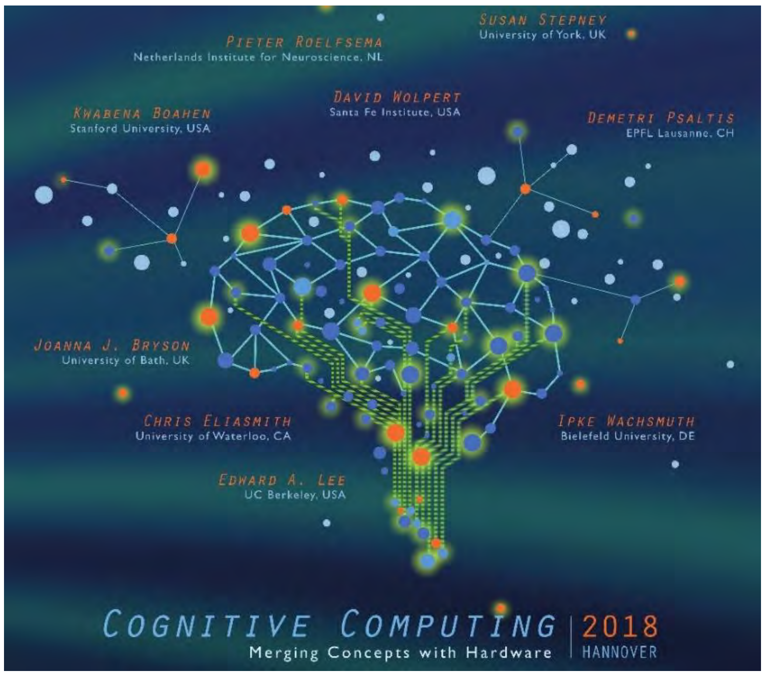
International conference on:

Cognitive Computing: Merging Concepts and Hardware

http://www.cognitive-comp.org/

18^{th} - 20^{th} of December 2018

at Herrenhausen Castle, Hannover, Germany



Topical Sessions:

- Cognitive neurosciences: how they may guide novel computing technologies
- Cognitive applications of current systems
- Theoretical concepts and mathematical foundations
- Towards neuronal hardware networks
- Novel substrates

Organizing Committee:
- Dr. Daniel Brunner
- Prof. Herbert Jaeger
- Prof. Gordon Pipa
- Prof. Stuart Parkin

VolkswagenStiftung

- Kwabena Boahen
- Joanna J. Bryson
- Chris Eliasmith
- Edward A. Lee
- Demetri Psaltis
- Pieter Roelfsema
- Susan Stepney
- Ipke Wachsmuth
- David Wolpert

Spin Phenomena Interdisciplinary Center, Mainz University

Summary

- 2025 EO network nodes, much larger to be expected
- Learning / analog, passive readout fully implemented
- Stability (?), noise (?)
- Approaching understanding of prediction in ANN

http://neuroqnet.com/

VolkswagenStiftung

NEUROQNET

Labex Action
Integrated smart systems

femto-st SCIENCES & TECHNOLOGIES