Information processing with recurrent dynamical systems:

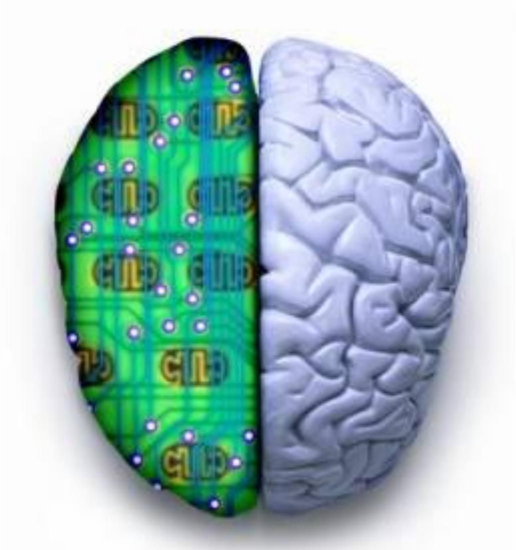
theory, characterisation and experiment.

S. Massar (ULB)
with
Y. Paquot, F. Duport, A. Smerieri, M. Haelterman (ULB)
and
J. Dambre, D. Verstraeten, B. Schrauwen (U. Gent)
and
M.Massar (New York)
and
L. Appeltant, G. Van der Sande, J. Danckaert (VUB)
and
M.C. Soriano,C.R. Mirasso, I. Fischer (IFISC, Palma de Mallorca)

My Background

- Theoretical Physicist
  - Quantum Gravity
  - Quantum Information Theory
- Experimental Physicist
  - Quantum and Non Linear Optics
- And recently:
  - Machine Learning and Reservoir Computing

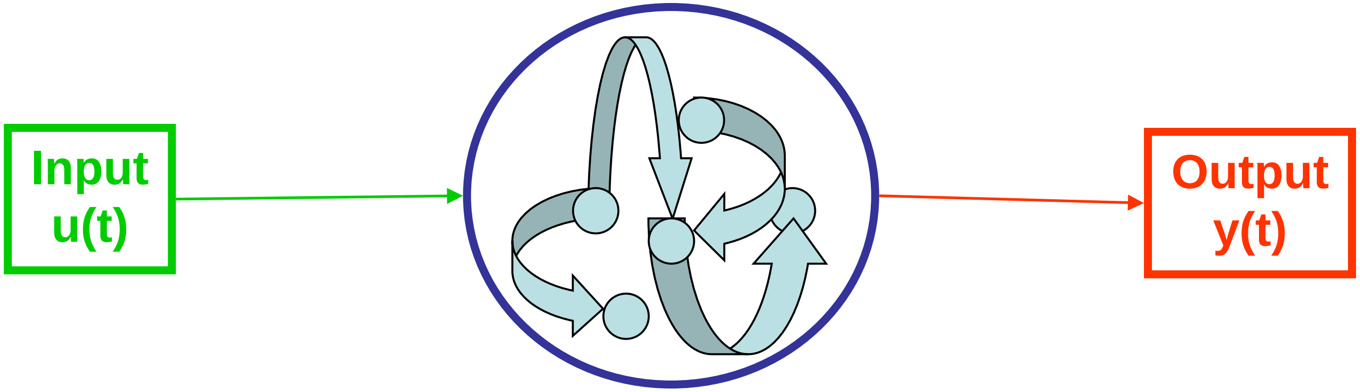
COMPUTERS VS BRAINS



- Low number of parallel units
- energy & space consumption
- Best at "hard" tasks
- Top-down architecture
- Designed
- Everything has a clear purpose

- High parallelism
- = energy and space consumption*10^5
- Good at "difficult to code" tasks
- Bottom-up architecture
- Evolved
- Doesn't even *have* to make sense

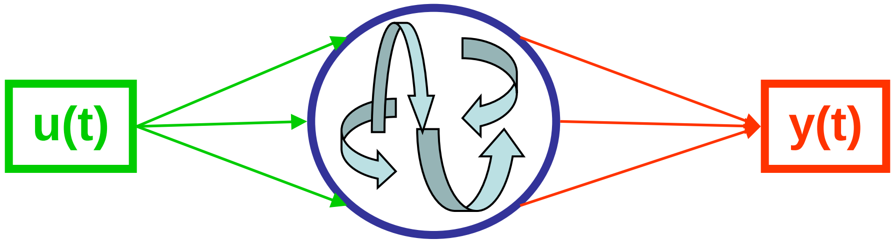
Training Recurrent Dynamical Systems



Recurrent Dynamical System

- Hard to train individual parameters using gradient descent.

Reservoir Computing
(Liquid State Machine)
(Echo State Network)



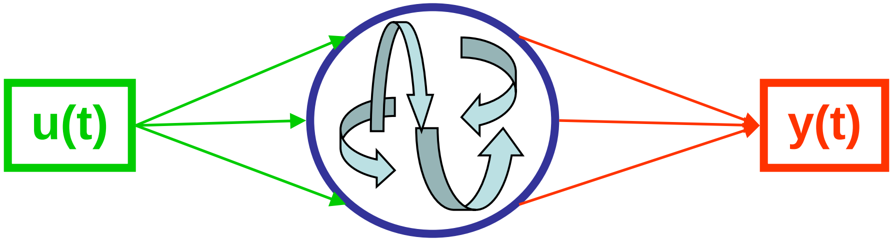
$$x_i(t+1) = \tanh\left(\sum_j \alpha_{ij} x_j(t) + \beta_i u(t)\right)$$

$$y(t) = \sum_i W_i x_i(t)$$

Linear Readout
Only these coefficients are trained.
→Easy optimisation

State of the art performance for:
- time series prediction
- speech recognition

Echo State Network



$$x_i(t+1) = \tanh\left(\sum_j \alpha_{ij} x_j(t) + \beta_i u(t)\right)$$

$$y(t) = \sum_i W_i x_i(t)$$

$N(0,\beta^2)$ Input Scaling

$N(0,\alpha^2)$ Feedback Strength

RANDOM COEFFICIENTS

!!ALMOST ALL RESERVOIRS ARE GOOD!!

→EXPERIMENT: Almost all experimental implementations will work.

→THEORY: Understand property of average reservoir

Plan

- Experimental reservoir computing
  - Opto-electronic reservoir computer with a single non linear node
- Characterise information processing of recurrent dynamical systems
  - Linear and Non Linear Memory Capacities
- Reservoir computing and statistical mechanics
  - Mean Field Theory of Echo State Networks

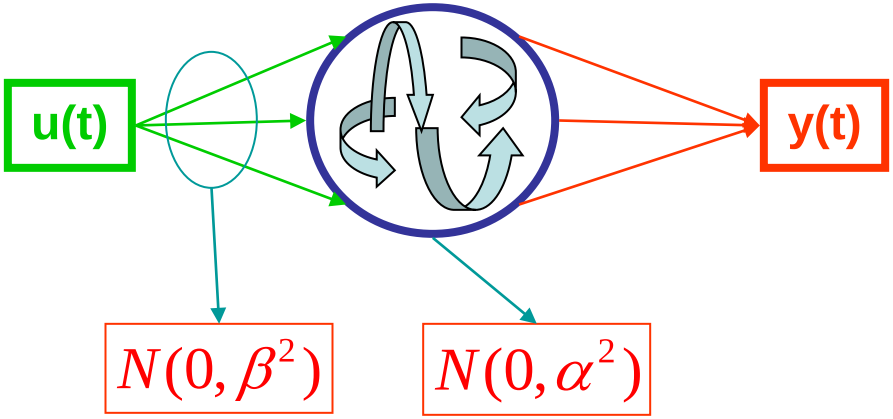
Experimental Reservoir Computer

- Experimental Demonstrations:
  - Water Bucket Fernando 2003
  - VLSI chip Schürmann 2005
  - Delay Line with a Single Non Linear Node
    - Electronic: Appeltant 2011 (in press in Nat. Comm.)
    - Opto-Electronic: Paquot 2011 (submitted)

- Why Experiments?
  - Gain new insights (e.g. noise resistance)
  - Possibility of applications:
    - Ultra fast / ultra low energy computation

- Why Optics?
  - Our labs expertise
  - Optics can in principle be super fast !!!

Echo State Network

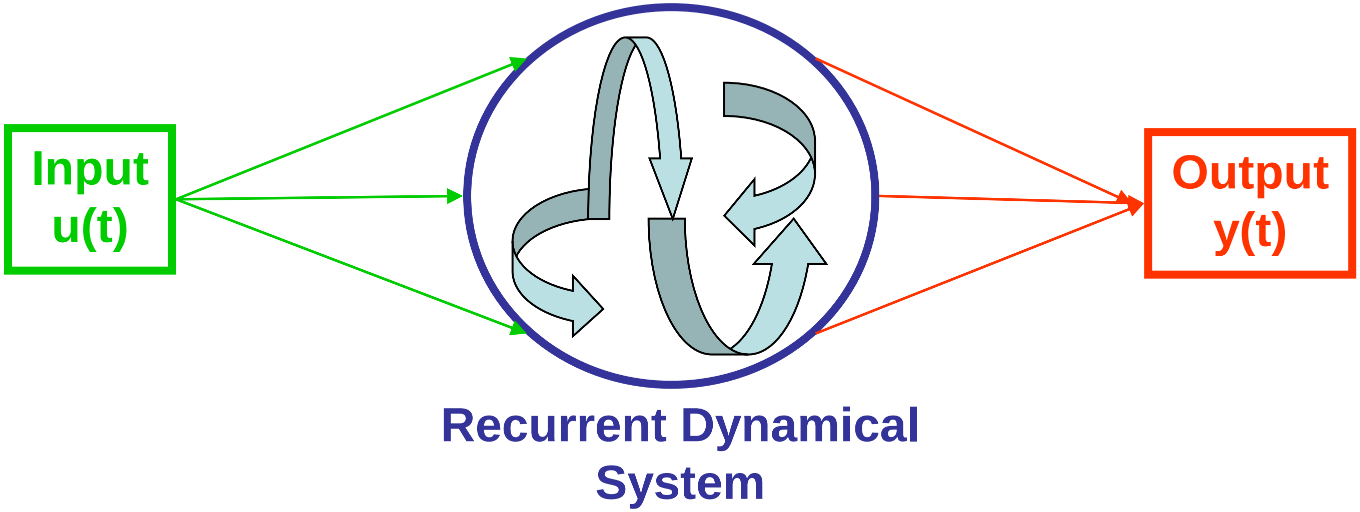


RANDOM COEFFICIENTS

!! Almost all experimental implementations will work !!

GOAL: Build a SIMPLE system with GOOD performance

Opto-Electronic Reservoir based on a single non linear node and a delay line.



- Need to describe each part of the reservoir

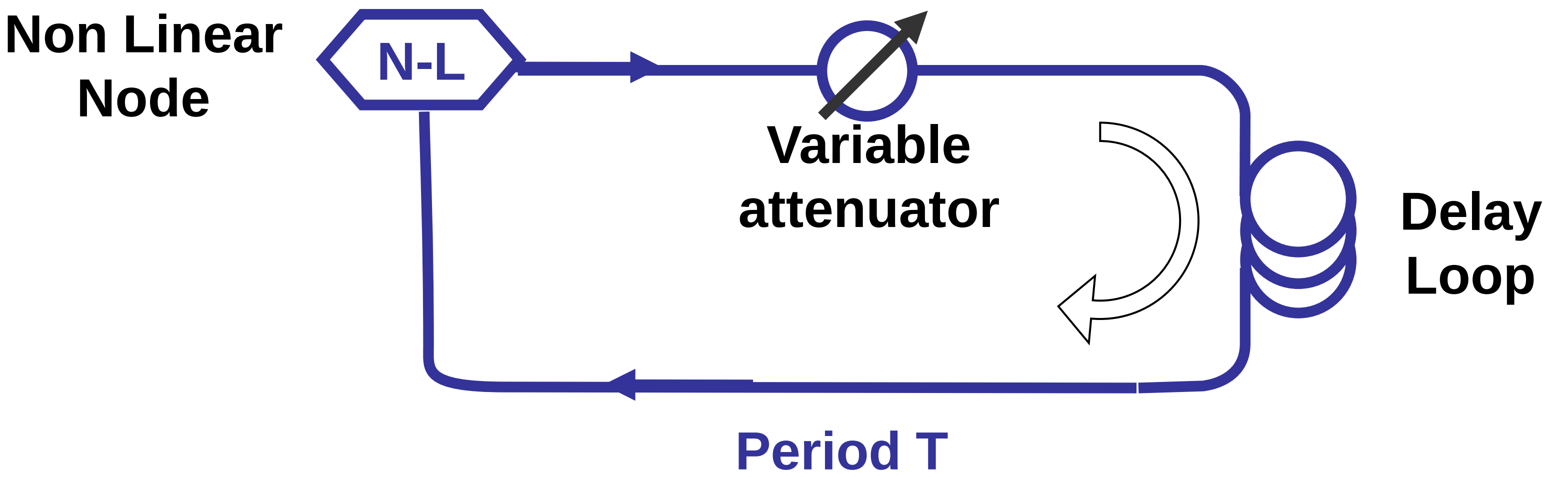
The Reservoir:
Single Non Linear Node
with Delayed Feedback

Delay systems:

✓Many internal variables

✓Rich dynamics with transition to chaos

✓Few components→Easy to build

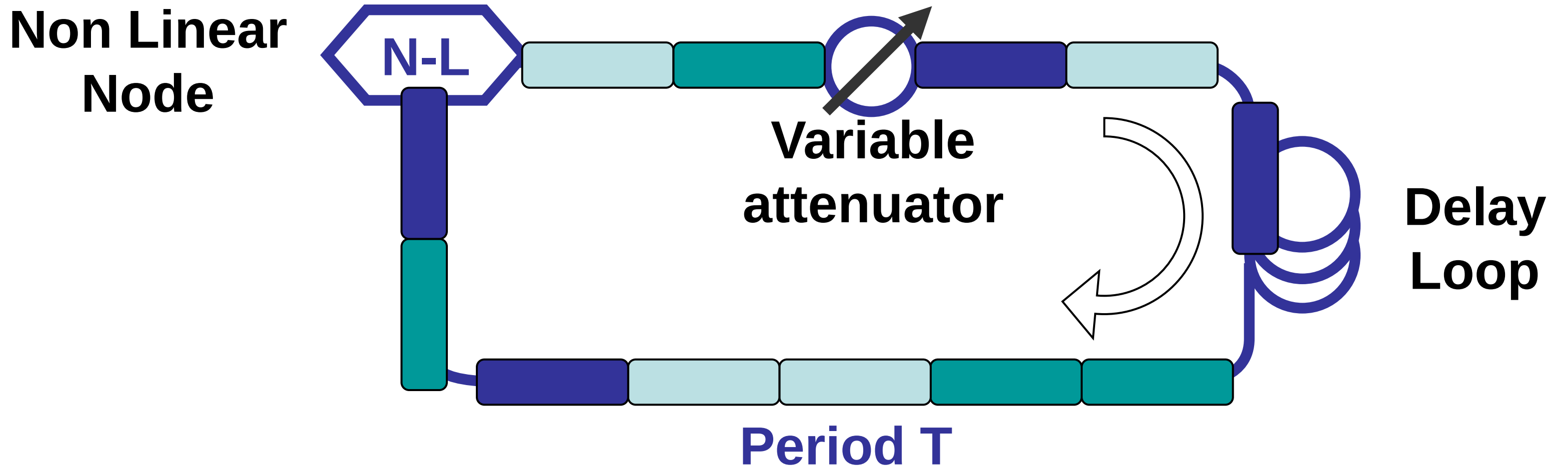


$$x(t+T)=F[\alpha x(t)]$$

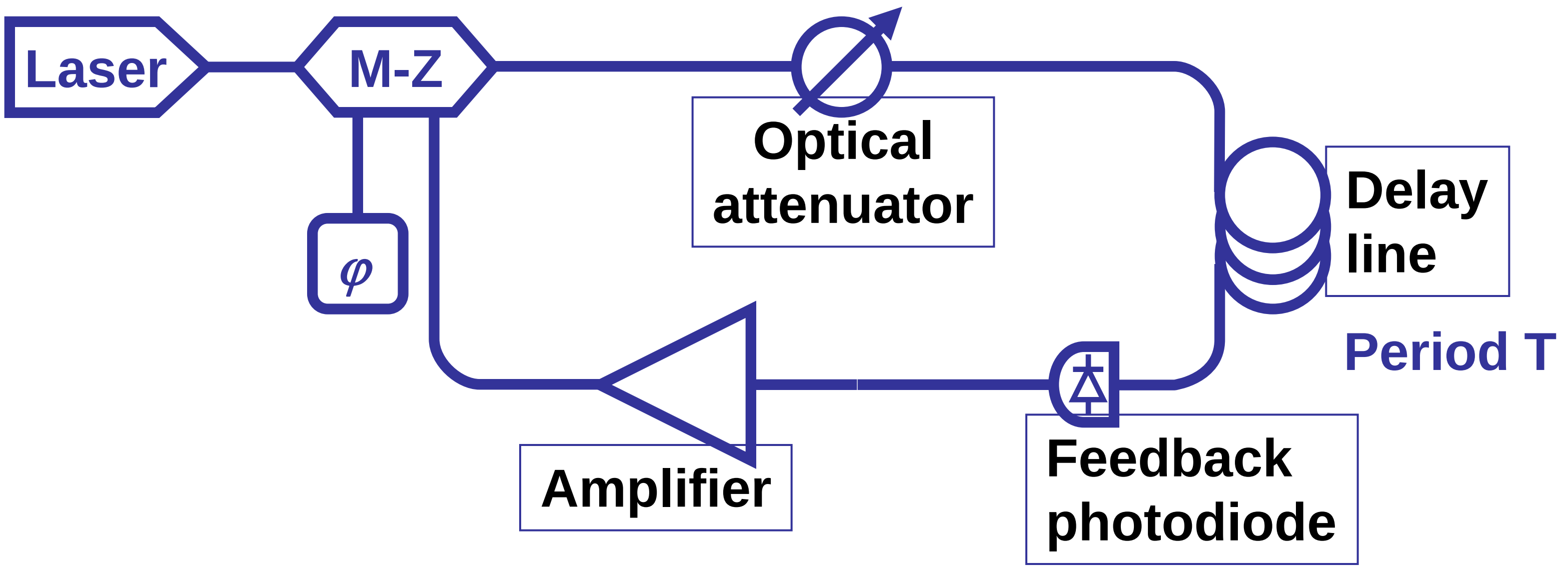
The Reservoir: Single Non Linear Node with Delayed Feedback

Dynamical variables=internal states inside delay loop

$x(t+T)=F[\alpha x(t)]$ → $x_i(n)=F[\alpha x_i(n-1)]$



Experimental Realisation



$$x(t) = \frac{I(t) - \langle I \rangle}{\langle I \rangle}$$

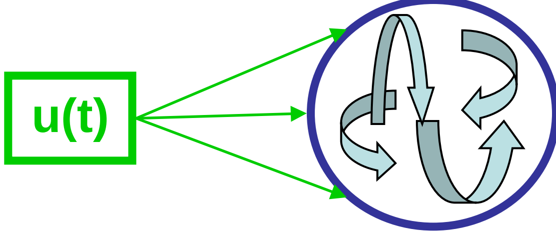
$$x(t+T) = \sin\left(\alpha\, \mathrm{x}(\mathrm{t}) + \varphi\right)$$

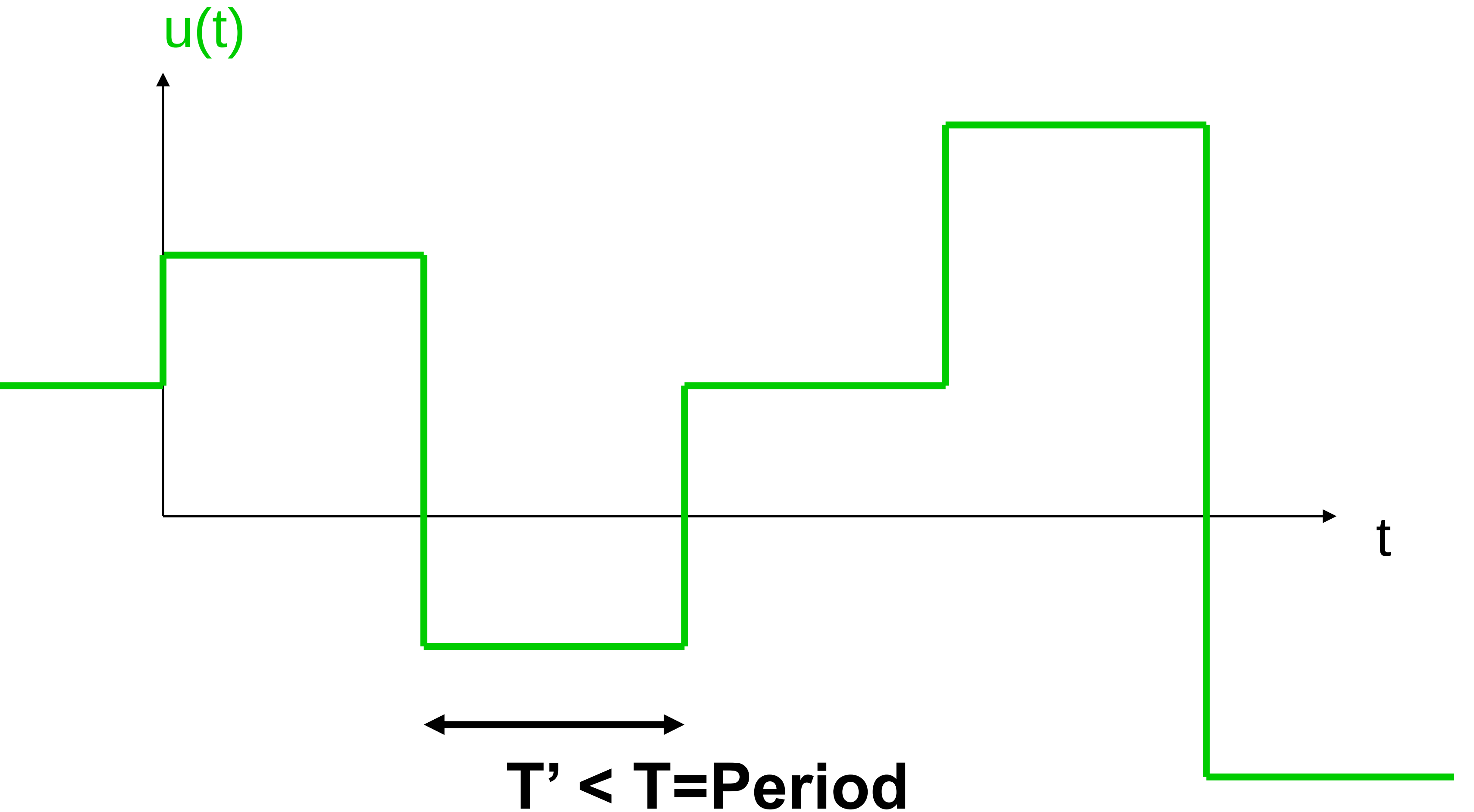
$$x_i(n) = \sin[\alpha x_i(n-1)]$$

Input

- Break symmetry
- Add richness to the reservoir dynamics

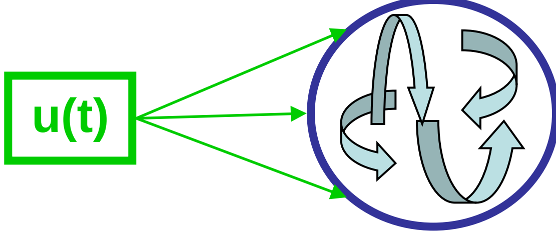
1) De-synchronyze





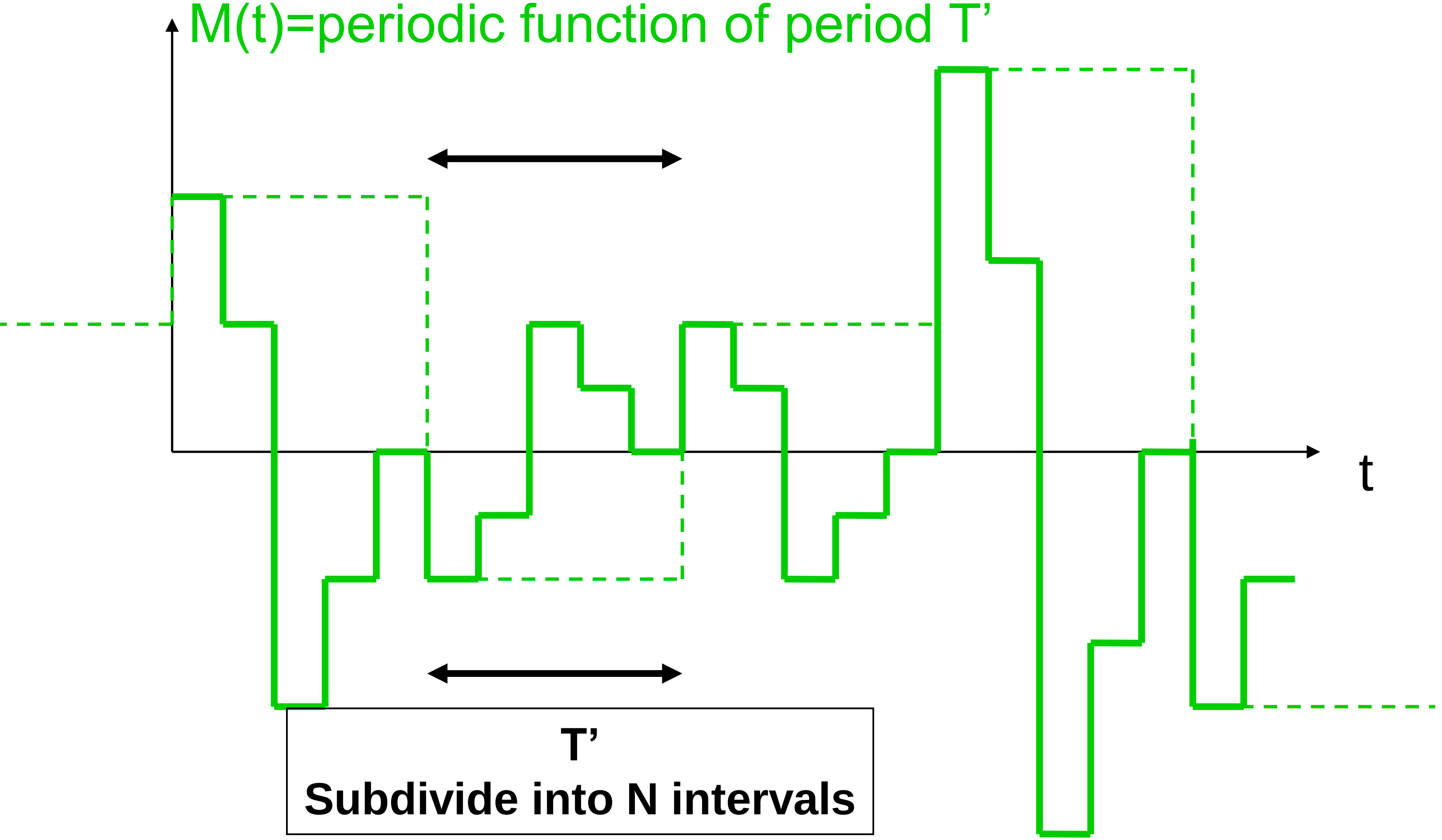
$$x_i(n)=\sin[\alpha x_{i-1}(n-1)+u(n)]$$

2) PeriodicMask

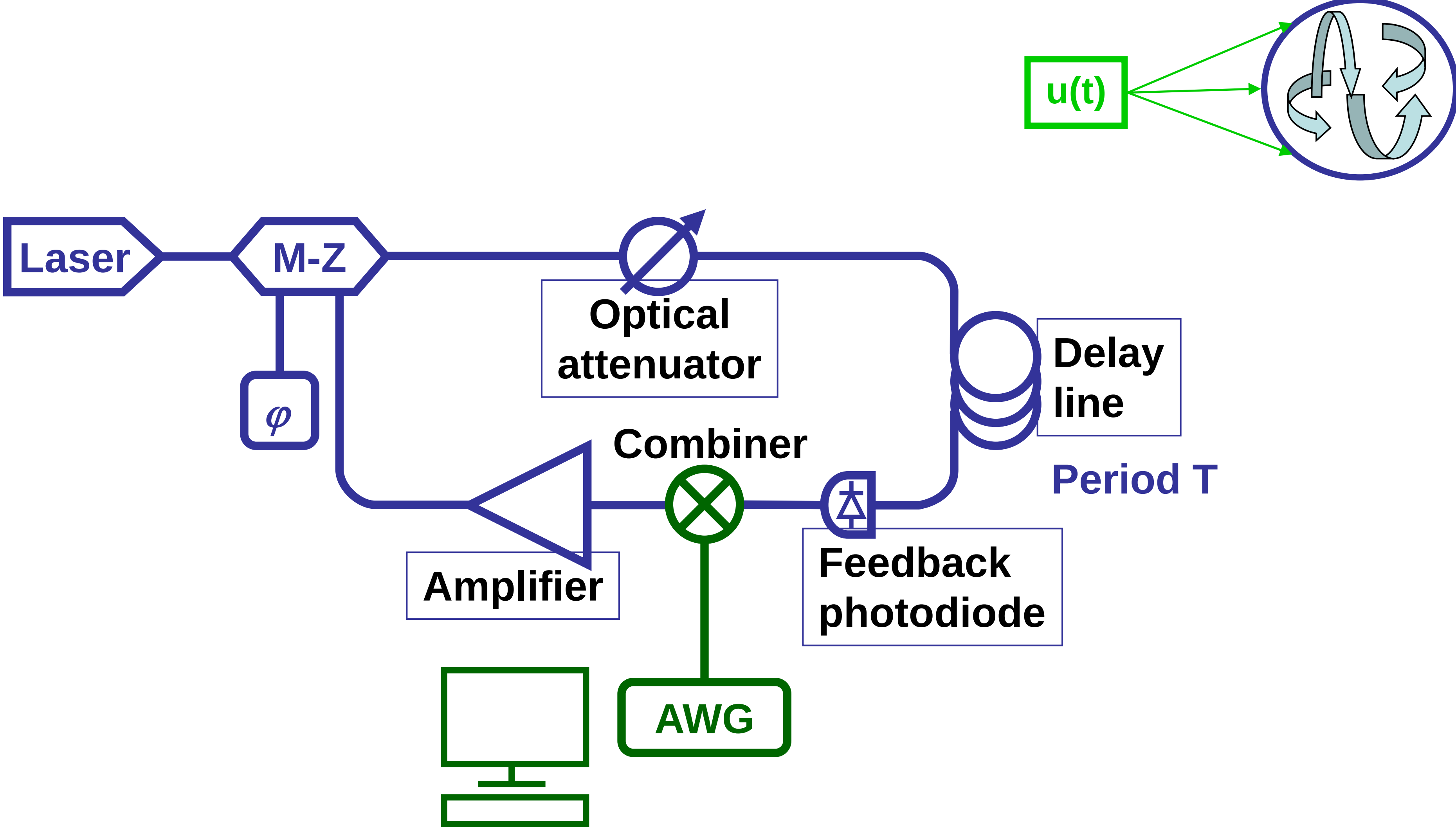


$J(t)=u(t)*M(t)$

$M(t)$=periodic function of period T'

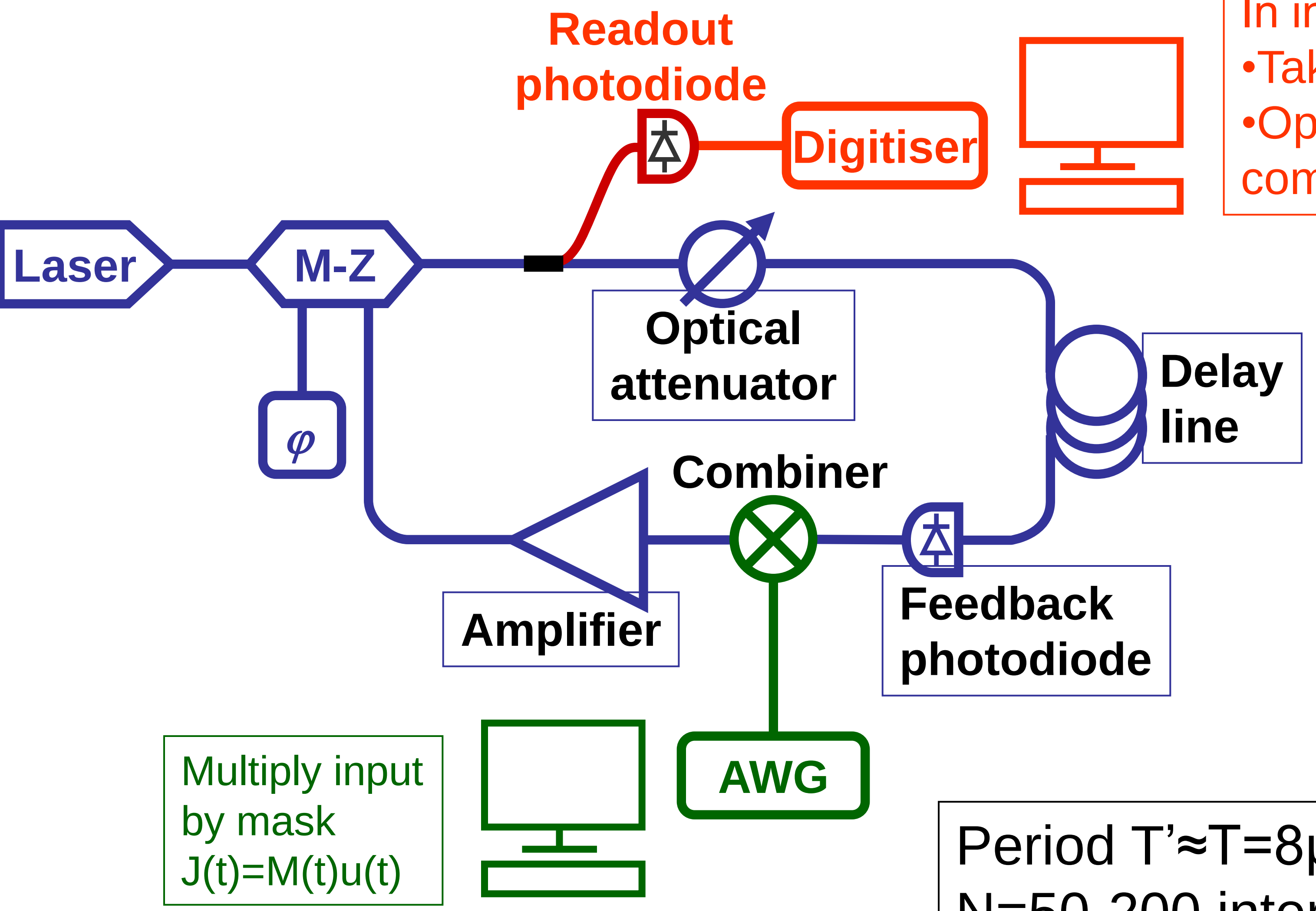


$$x_i(n)=\sin[\alpha x_{i-1}(n-1)+m_i u(n)]$$



$$x_i(n)=\sin[\alpha x_{i-1}(n-1)+m_i u(n)+\varphi]$$

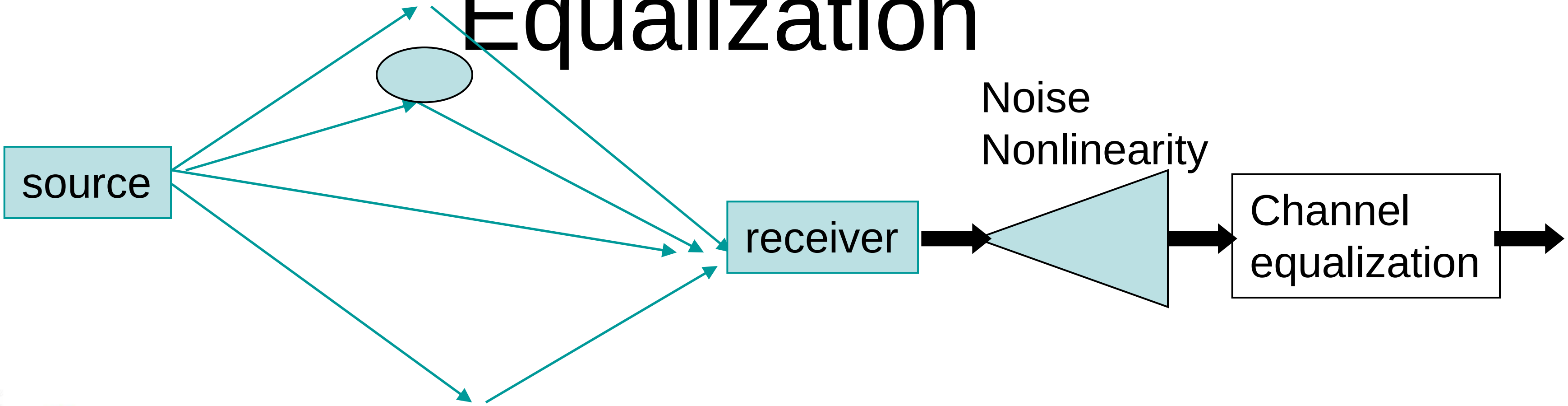
Opto-Electronic Reservoir

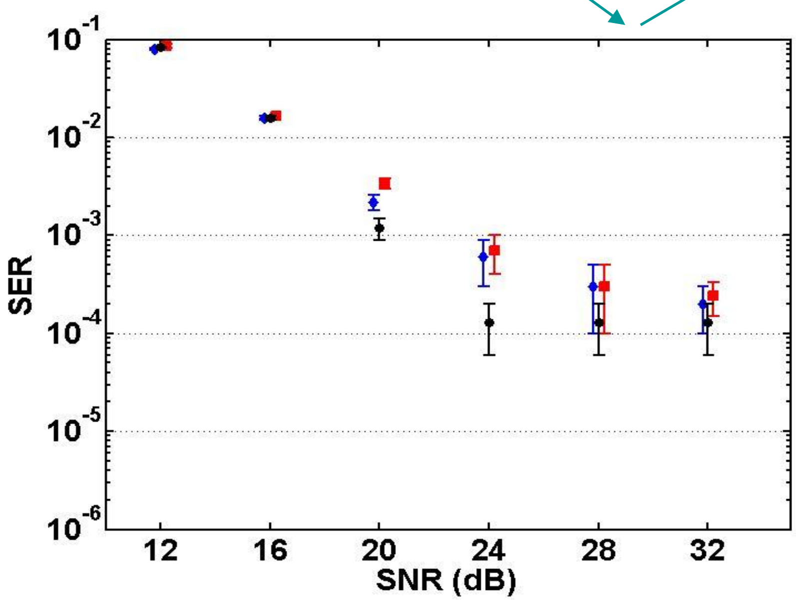


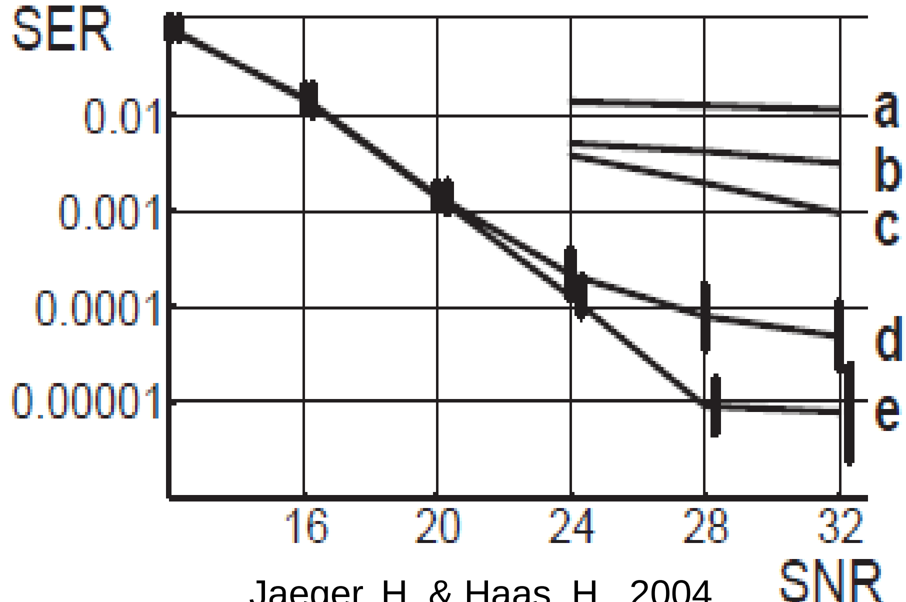
In interval T'
- Take N samples
- Optimize linear combination

Period T'≈T=8μs
N=50-200 internal nodes

Task 1: Non Linear Channel Equalization



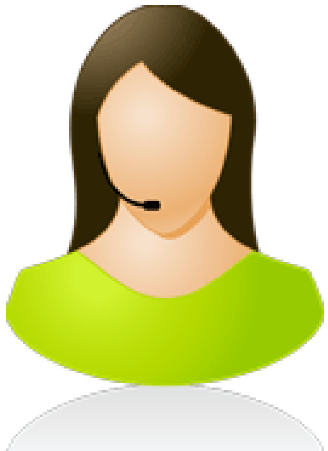




Jaeger, H. & Haas, H., 2004. *Science*

Our results (**two simulations** and **experiment**) comparable with average results (d) of Jaeger (same reservoir size).

Results:
2) Spoken Digit Recognition



•Five speakers pronounce several times digits from 0 to 9

•Goal: recognize the digits from a pre-processed version (cochlear model) of the audio files

- Word Error Rate of 0.4% – that's 2 errors on the whole 500 digits set

Method	WER
Opto-Electronic (Paquot 11)	0.4%
Electronic (Appeltant 11)	0.2%
First reservoir (Verstraeten 05)	4.3%
Verstraeten 06	0.2%
SPHINX 4	0.55%

Experiment: Summary

- First analog RC with performance comparable to digital RC
  - Very simple architecture: Non Linear node with delayed feedback

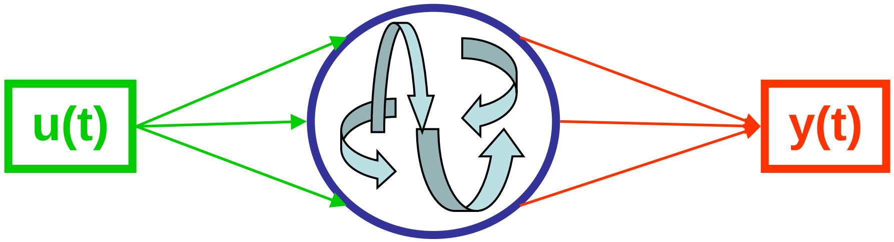
- Perspectives:
  - Remove electronic pre- and post-processing
  - Low energy / High speed

Recipe to build experimental reservoirs

- Non Linear Dynamical System
  - 20-100 internal variables
  - Driven by external input u(t)
- Adjustable parameters
  - Feedback strength → threshold of instability
  - Input gain → non linearity of reservoir
- Readout
  - Measure internal variables
  - Adjust weights of linear combination

- Controled feedback loops→autonomous behaviour

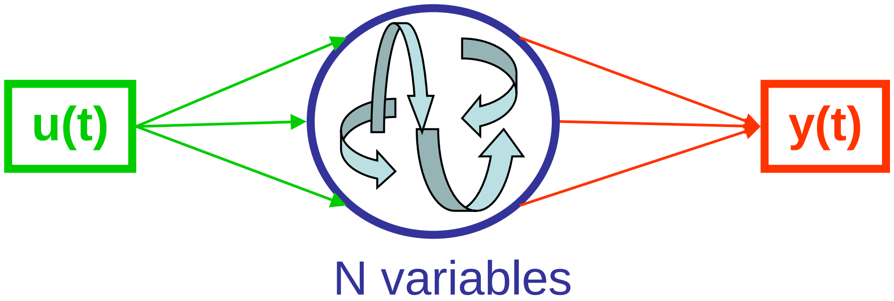
Linear and Non Linear Memory Capacity



- Characterise the information processing by recurrent dynamical systems?
  - Are some systems better than others?
  - Can we associate performance on tasks to properties of the system?

Linear Memory Capacity
(Jaeger 2002)

IID:Independent Identically Distributed



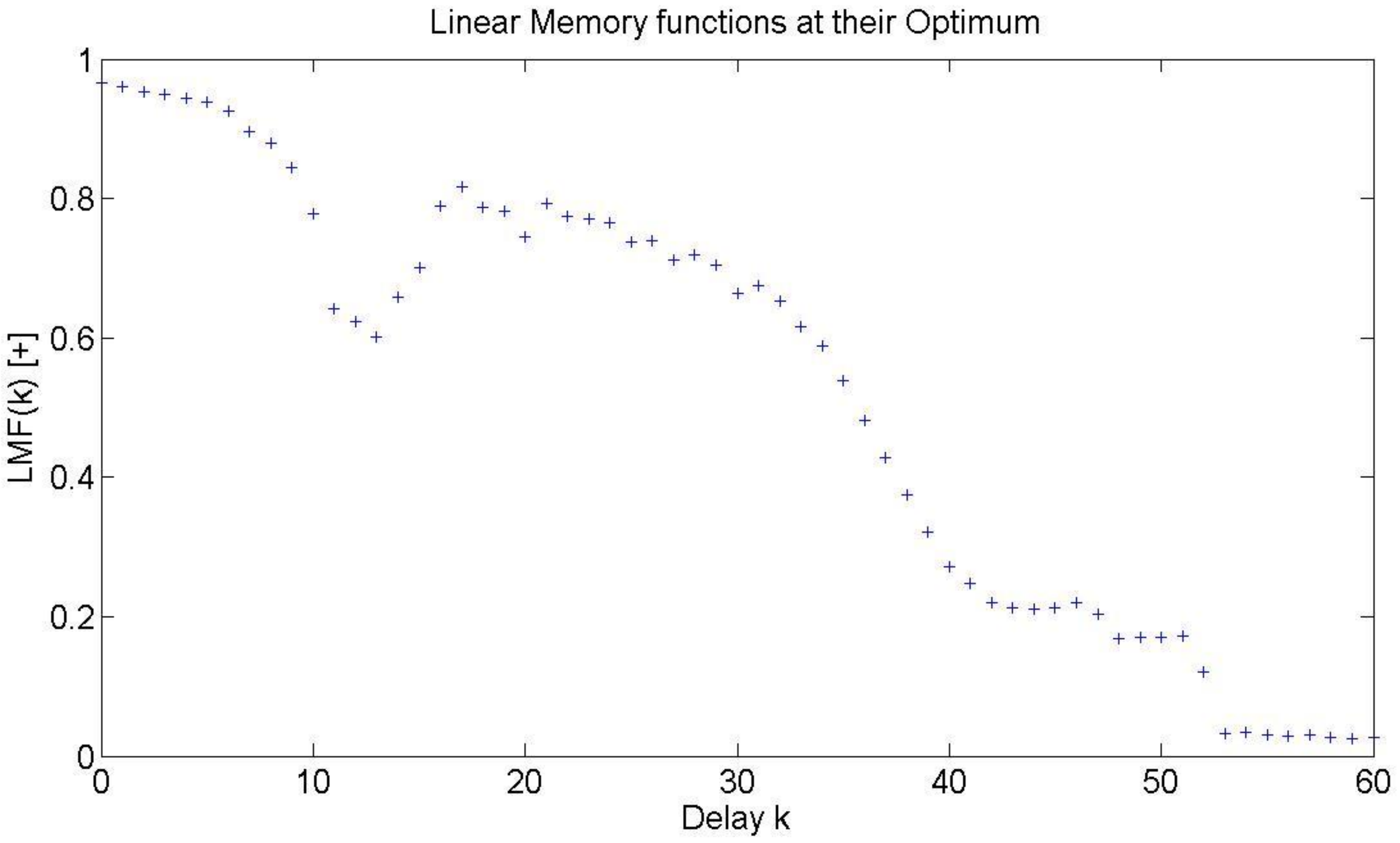
Goal: Reproduce Input k steps in the past:
y(t)≈u(t-k)

$C_L(k)$ = ability to reproduce input k steps in the past

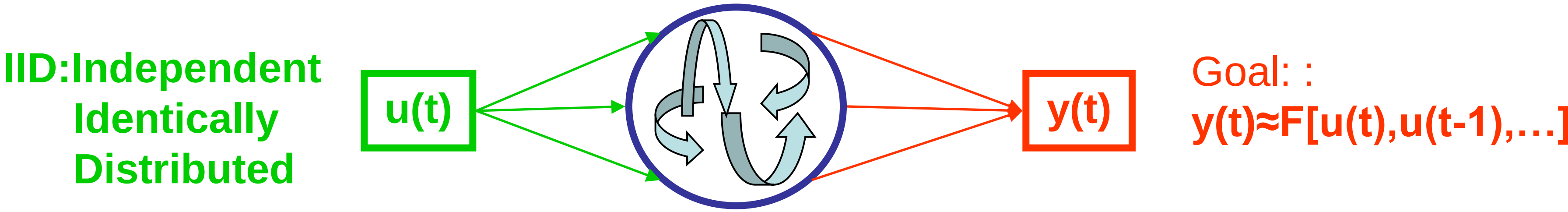
$0 \le C_L(k) \le 1$

$\sum_k C_L(k) \le N$ =size of reservoir (Equality for linear reservoirs and linearly independent internal states)

Linear Memory Function



Non Linear Memory Capacity



$C(F)$ = ability to reproduce F[u(t-1),u(t-2),...]

$0 \le C(F) \le 1$

$$\sum_{F \in \text{Complete Set}} C(F) \le N = \text{size of Reservoir (Equality for reservoirs with FADING MEMORY}$$

and linearly independent internal states)

Hilbert Space of Fading Memory Functions

$u(t)$ i.i.d.

One Input

$P(u)$ = probability distribution over u

F(u) = real valued function of u

Hilbert space: $\langle \mathrm{F,G} \rangle = \int du P(u) F(u) G(u)$

Many Inputs

$u^{-\infty} = u(t), u(t-1), u(t-2), \ldots$

$F[u^{-\infty}] = F[u(t), u(t-1), u(t-2), \ldots]$ = function all of previous inputs

H_{FM} =Hilbert space of Fading Memory Functions:

$$\langle \mathrm{F,G} \rangle = \int du(t) du(t-1) du(t-2) \ldots F[u(t), u(t-1), u(t-2), \ldots] G[u(t), u(t-1), u(t-2), \ldots]$$

$$= \int du^{-\infty} P(u^{-\infty}) F[u^{-\infty}] G[u^{-\infty}]$$

Reservoir with Fading Memory

Dynamical System has Fading Memory: $x_i(t) = x_i[u^{-\infty}] \in H_{FM}$

choose orthonormal basis:

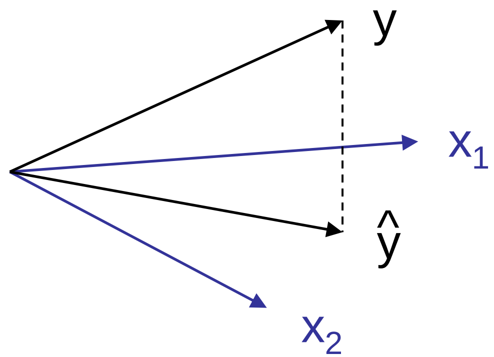
$$\tilde{\mathrm{x}}_{\mathrm{i}}[u^{-\infty}] = \sum_j c_{ij} x_j [u^{-\infty}]$$

$$\left\langle \tilde{\mathrm{x}}_{\mathrm{i}}, \tilde{\mathrm{x}}_{\mathrm{j}} \right\rangle = \delta_{ij}$$

Target function= $\mathrm{y}[u^{-\infty}]$

$\hat{y}[u^{-\infty}] = \sum_i W_i x_i [u^{-\infty}] =$ best linear approximation of $\mathrm{y}[u^{-\infty}]$

$\hat{y}[u^{-\infty}] = \sum_i \left\langle \tilde{\mathrm{x}}_{\mathrm{i}}, y \right\rangle x_i [u^{-\infty}] =$ projection of $y(t)$ onto space spanned by $x_i[u^{-\infty}]$



Non Linear Memory Capacity =capacity to reconstruct y[u$^{-\infty}$]

Target function= y[$u^{-\infty}$]

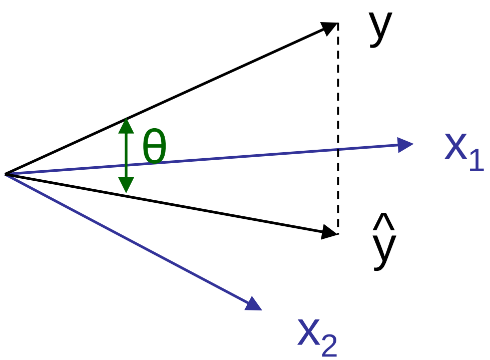
$$\hat{y}[u^{-\infty}] = \sum_i W_i x_i[u^{-\infty}] = \text{best linear approximation of y}[u^{-\infty}]$$

$$\hat{y}[u^{-\infty}] = \sum_i \langle \tilde{\mathrm{x}}_\mathrm{i}, y\rangle x_i[u^{-\infty}] = \text{ projection of } y(t) \text{ onto space spanned by } x_i[u^{-\infty}]$$

$$C[y] = \frac{\sum_i \langle \tilde{\mathrm{x}}_\mathrm{i}, y\rangle^2}{\langle y, y\rangle^2} = \frac{\text{norm square of best linear approximation of y}}{\text{norm square of y}}$$

$$0 \le C[y] \le 1$$

$\boxed{C[y]=\cos^2(\theta)}$



Completeness

$$C[y] = \frac{\sum_i \left\langle \tilde{\mathrm{x}}_{\mathrm{i}}, y \right\rangle^2}{\left\langle y, y \right\rangle^2}$$

$\{y_k\}$ = complete orthonormal basis of H_{FM}

$$\left\langle y_k, y_{k'} \right\rangle = \delta_{kk'}$$

$$\sum_{k \in Basis} C[y_k] = \sum_{k \in Basis} \sum_i \left\langle \tilde{\mathrm{x}}_{\mathrm{i}}, y_k \right\rangle = \sum_i \left(\sum_{k \in Basis} \left\langle \tilde{\mathrm{x}}_{\mathrm{i}}, y_k \right\rangle \right) = \sum_i 1 = N$$

In practice

$$u(t) \in [-1, +1]$$

$$\{y_k\} = \{\text{Legendre Polynomials and products of Legendre Polynomials}\}$$

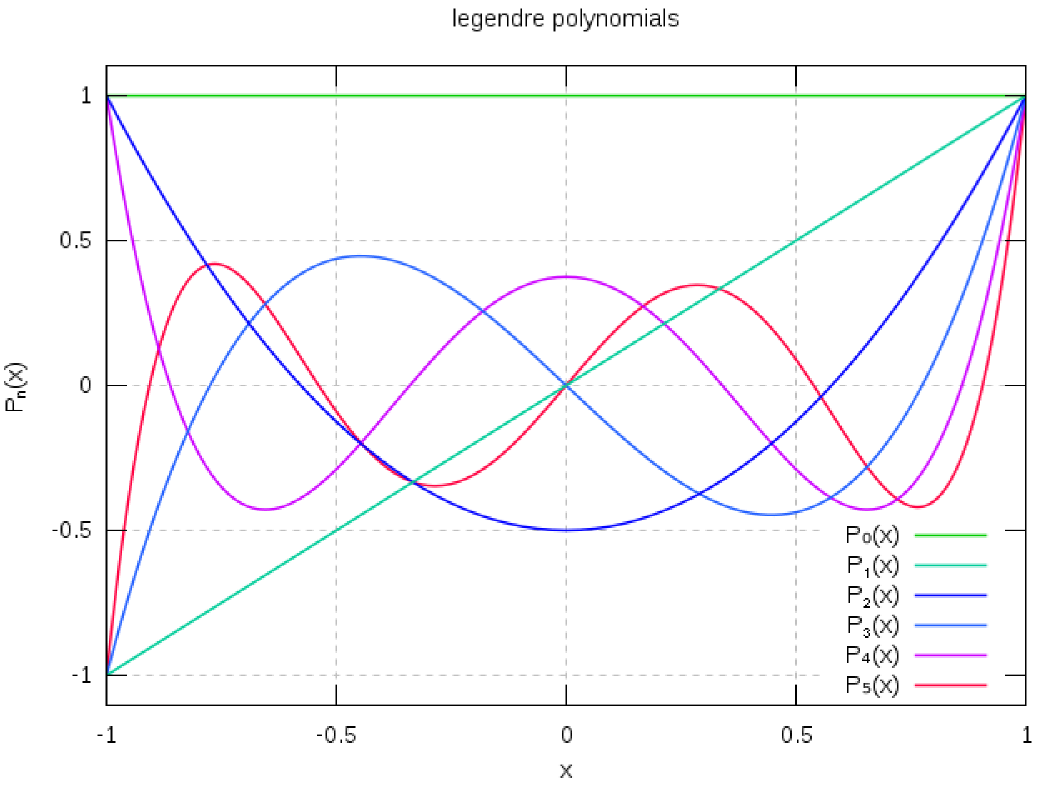
$$P_0 = 1$$

$$P_{1;k} = u(t-k)$$

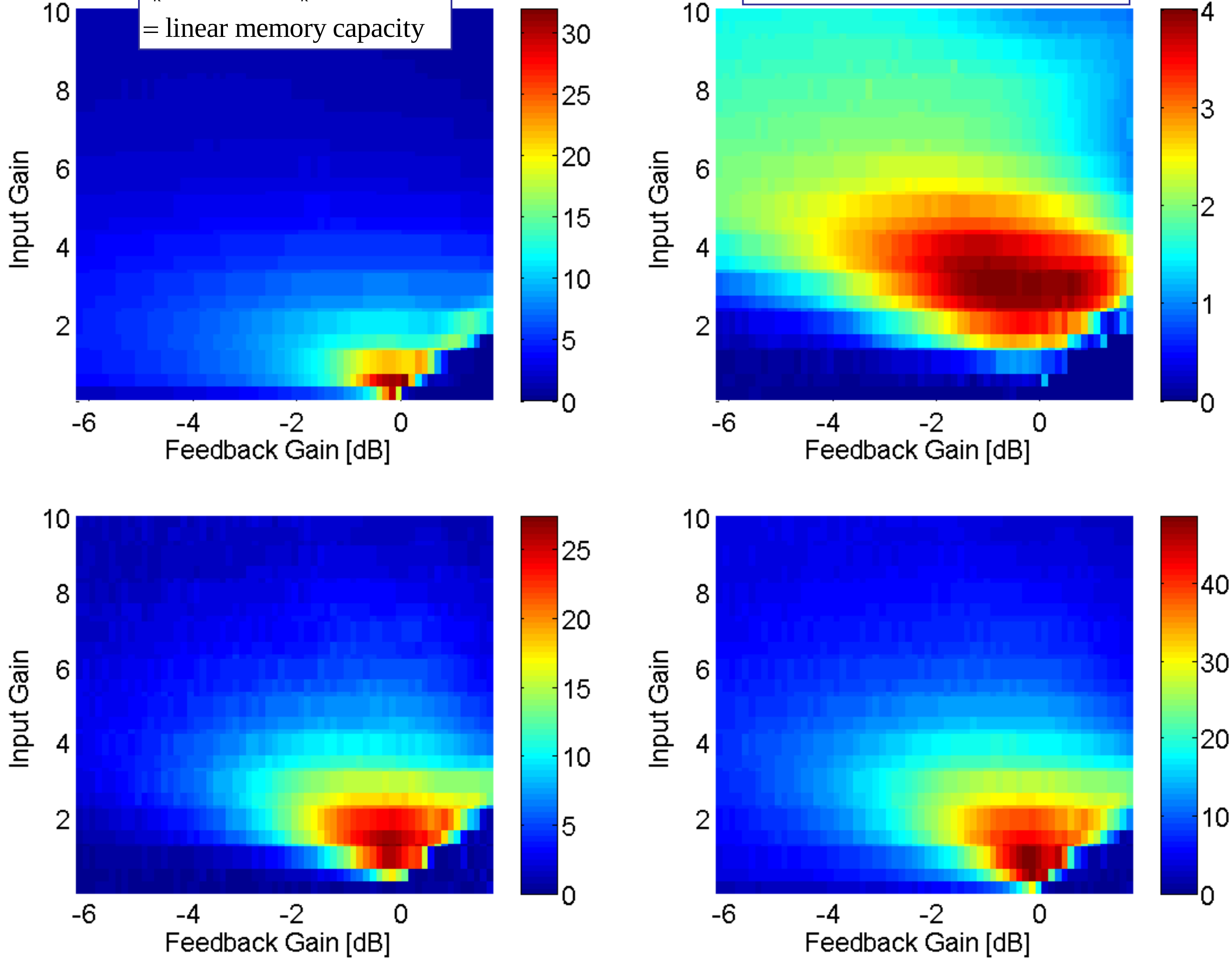
$$P_{2;k} = \frac{1}{2}\left(3u^2(t-k) - 1\right)$$

$$P_{11;kk'} = u(t-k)u(t-k')$$

$$etc...$$



In practice: Experimental Data (N=50)



$$\sum_k C\left[P_{1;k}\right] = \sum_k C\left[u(t-k)\right]$$
= linear memory capacity

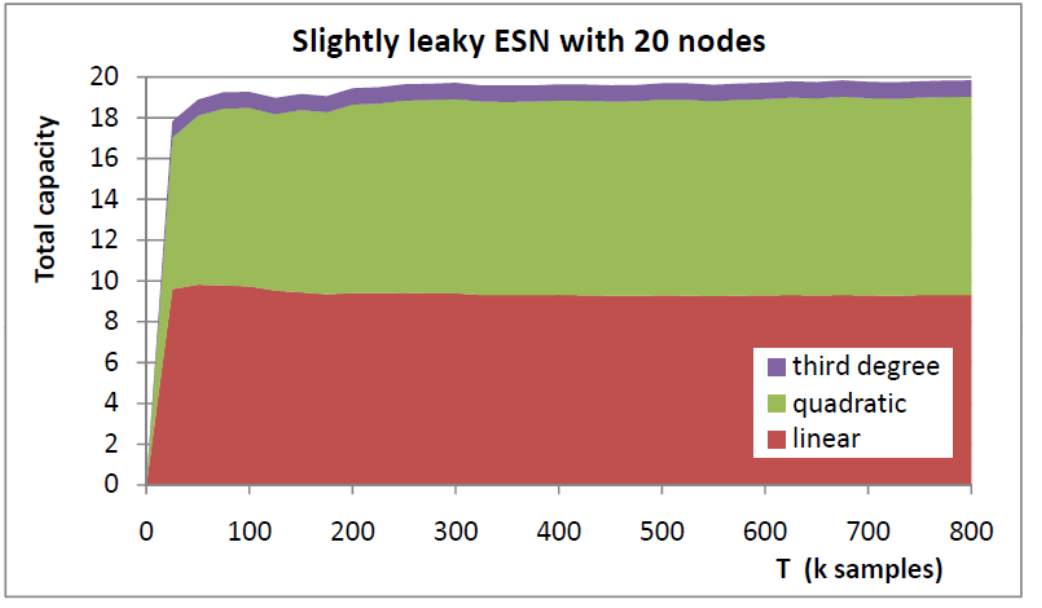
$$\sum_k C\left[P_{2;k}\right] = \sum_k C\left[\frac{3u^2(t-k)-1}{2}\right]$$

$$\sum_{kk'} C\left[P_{11;kk'}\right] = \sum_{kk'} C\left[u(t-k)u(t-k')\right]$$

Sum of capacities of order 1 and 2

Linear & Non Linear Memory Capacity

- New tool to characterise information processing by dynamical systems
  - !!!Convergence!!!



- Perspective: relate non linear memory capacity to performance on tasks

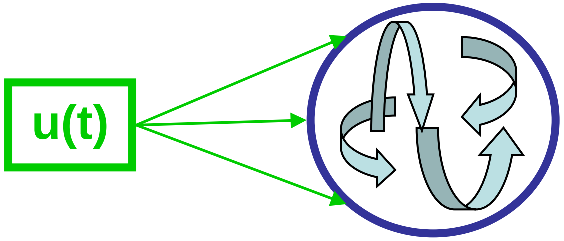
- Questions??

Mean Field Theory

- Why a Theory of Reservoir Computing?
  - Fundamental Curiosity: what enables computation?
  - Help in designing better reservoirs

- Aim: predict *ab initio* the memory capacities, the performance on tasks

Mean Field Theory

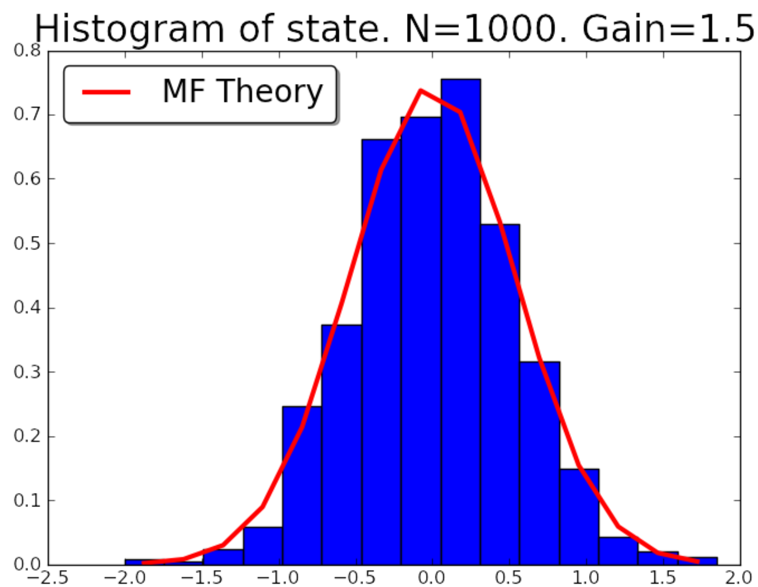


$$x_i(t+1) = \tanh\left(\sum_j \alpha_{ij} x_j(t) + \beta_i u(t)\right)$$

Key Insight:

Neglect correlations between $x_i(t)$

Then $\sum_j \alpha_{ij} x_j(t)$ = sum of independent variables = GAUSSIAN

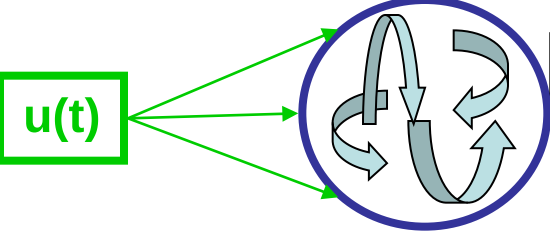


Mean Field Theory: No Source

$$\boxed{x_i\left(t+1\right) = \tanh\left(\sum_j \alpha_{ij} x_j(t)\right)}$$

$$a_i(t) = \sum_j \alpha_{ij} x_j(t) \quad \sim \quad N\left(0, \Sigma^2(t)\right)$$

$$x_i\left(t+1\right) = \tanh\left(a_i(t)\right) \qquad \rightarrow \text{ compute } \quad \mathrm{var}(x_i(t)) = \sigma^2(t)$$



Mean Field Theory

$$x_i(t+1) = \tanh\left(\sum_j \alpha_{ij} x_j(t) + \beta_i u(t)\right)$$

$$a_i(t) = \sum_j \alpha_{ij} x_j(t) \quad \square \quad N\left(0, \Sigma^2(t)\right)$$

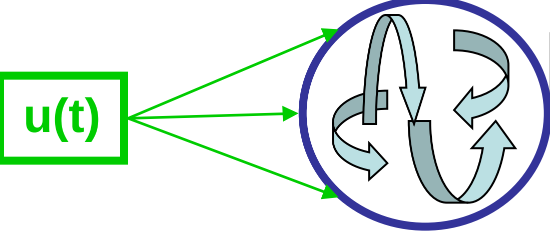
$$x_i(t+1) = \tanh\left(a_i(t)\right) \quad \to \text{ compute } \quad \operatorname{var}(x_i(t)) = \sigma^2(t)$$

Recurrence for the Variances:

$$\Sigma^2(t) = g\sigma^2(t) + \beta^2 u^2(t) \qquad \text{g=N*var}(\alpha_{\text{ij}}) \qquad \beta^2 = \operatorname{var}(\beta_i)$$

$$\sigma^2(t+1) = F\left(\Sigma^2(t)\right)$$

$$F\left(\Sigma^2\right) = \int da \tanh^2(a) \frac{\exp\left[-a^2 / 2\Sigma^2\right]}{\sqrt{2\pi\Sigma}}$$



Mean Field Theory

$$x_i(t+1) = \tanh\left(\sum_j \alpha_{ij} x_j(t) + \beta_i u(t)\right)$$

Recurrence for the Variances:

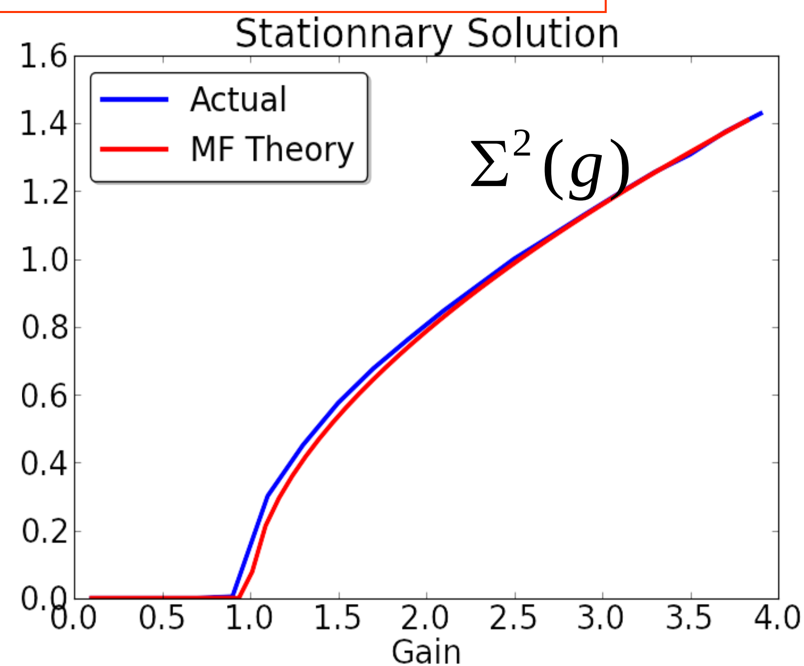
$$\Sigma^2(t) = g\sigma^2(t) + \beta^2 u^2(t) \qquad \text{g=N*var}(\alpha_{ij}) \qquad \beta^2 = \text{var}(\beta_i)$$

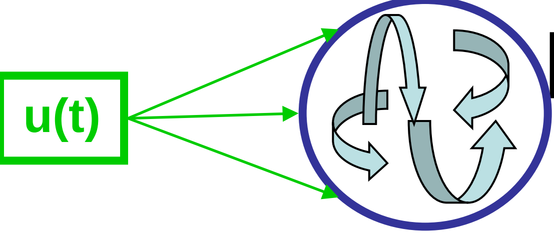
$$\sigma^2(t+1) = F\left(\Sigma^2(t)\right)$$

$$F\left(\Sigma^2\right) = \int da \tanh^2(a) \frac{\exp\left[-a^2/2\Sigma^2\right]}{\sqrt{2\pi\Sigma}}$$

Σ^2

No Source





Mean Field Theory

$$x_i(t+1) = \tanh\left(\sum_j \alpha_{ij} x_j(t) + \beta_i u(t)\right)$$

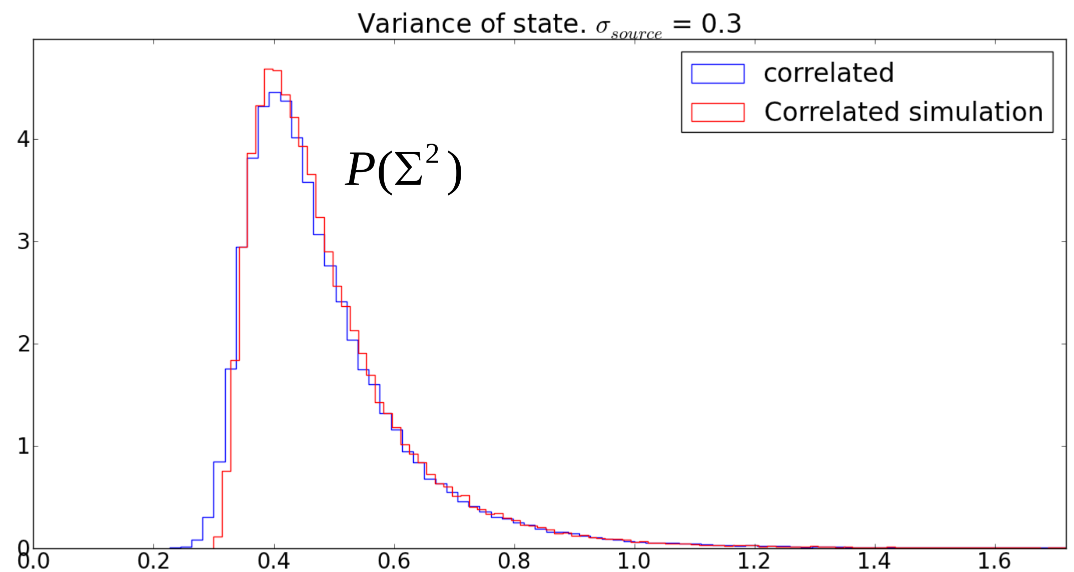
Recurrence for the Variances:

$$\Sigma^2(t) = g\sigma^2(t) + \beta^2 u^2(t) \qquad g = N^*\mathrm{var}(\alpha_{ij}) \qquad \beta^2 = \mathrm{var}(\beta_i)$$

$$\sigma^2(t+1) = F\left(\Sigma^2(t)\right)$$

$$F\left(\Sigma^2\right) = \int da \tanh^2(a) \frac{\exp\left[-a^2/2\Sigma^2\right]}{\sqrt{2\pi}\Sigma}$$

Correlated Source
→u(t) changes in time
→Probability distribution for $\sigma^2(t)$



Mean Field Theory: Computational Power

$$a_i(t) = \sum_j \alpha_{ij} x_j(t)$$

$$x_i(t+1) = \tanh\left(a_i(t) + \beta_i u(t)\right)$$

Consider two different inputs $u(t)$ and $u'(t)$

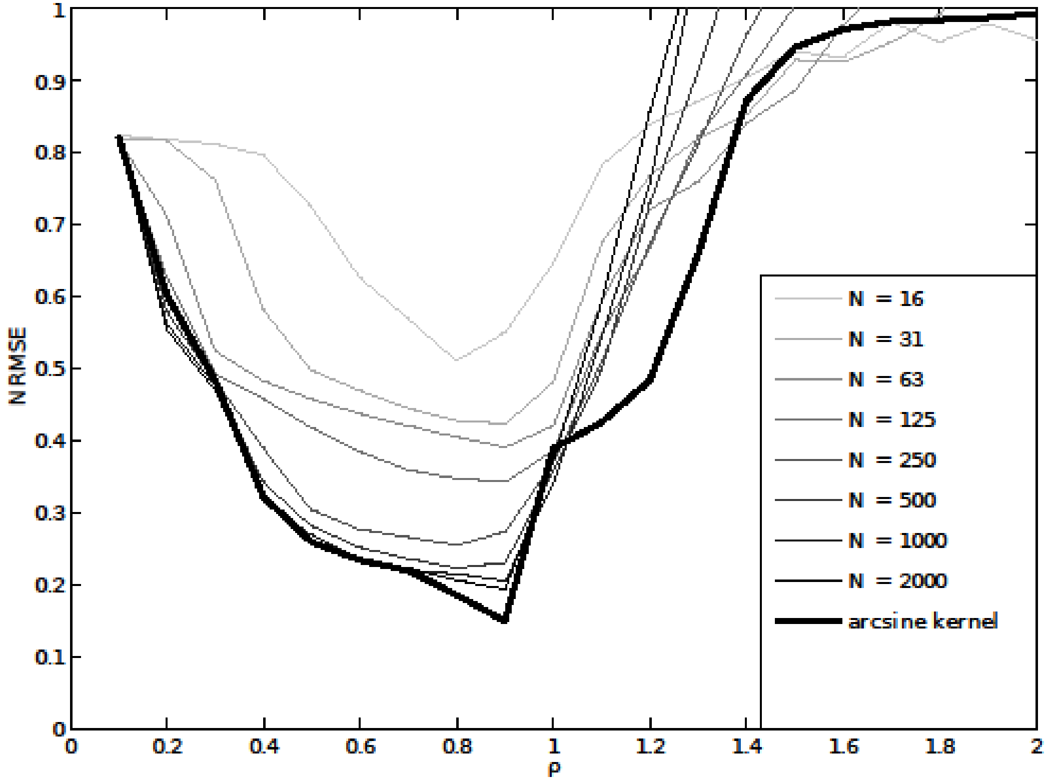
Mean field theory enables computation of the scalar product between $x_i[\mathrm{u}]$ and $x_i[u']$:

$$\sum_i \langle x_i[\mathrm{u}]\ , x_i[u']\ \rangle$$

From the scalar product, derive the performance on computational tasks

Herman and Schrauwen (2011)

Mean Field Theory: Computational Power



Performance on NARMA10 for reservoirs of different sizes versus spectral radius
Bold=mean field theory

Herman and Schrauwen (2011)

Summary

- First Experimental Reservoirs with performance comparable to digital implementation
  - Simple architecture: NL node + delay line
- Memory Capacities
  - New tool for characterising information processing in dynamical systems.
- Mean Field Theory = Initial steps towards understanding the dynamics of reservoirs