Energy-efficient neuromorphic computing with magnetic tunnel junctions

CNRS/Thales, France

Jacob Torrejon
Mathieu Riou
Flavio Abreu Araujo
Paolo Bortolotti
Vincent Cros
Julie Grollier

AIST, Japan

Sumito Tsunegi
Kay Yakushiji
Akio Fukushima
Hitoshi Kubota
Shinji Yuasa

C2N, France

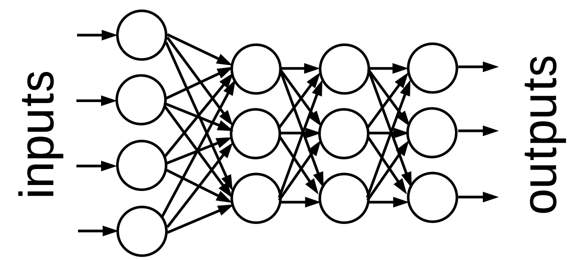
Damien Querlioz

NIST, USA

Guru Khalsa
Mark Stiles

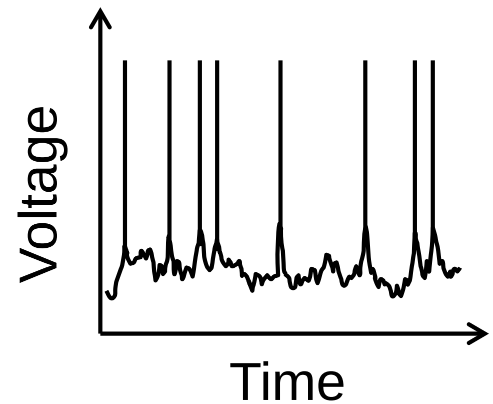
WHAT: Brain-inspired computing for cognitive tasks

- Neural networks.
  - Already in heavy commercial application.
  - Quasi-static.



- **Dynamic processing.**
  - **Make use of dynamical properties of devices.**

  → **Spikes, rates, oscillations, timing, …**



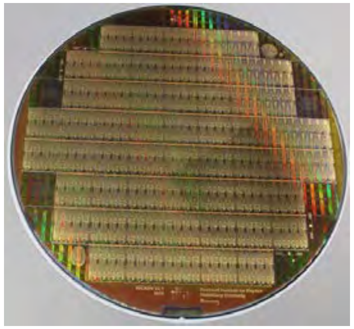
- Simulate brain function

Sophisticated CMOS-based neuromorphic chip development

Loihi (Intel)

TrueNorth (IBM)

BrainScales
(Human Brain Project)



- Brain-like features natural in CMOS
  - local energy source

    → Incoming signal does not power outgoing signal.
  - Connectivity

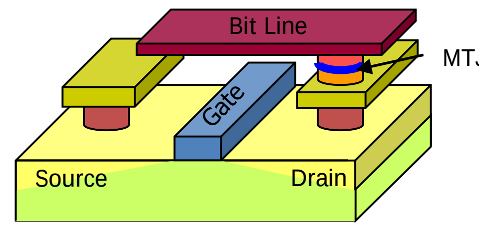
    → Digital spikes – shared communication channels

Energy efficiency and complexity is still far from the brain!

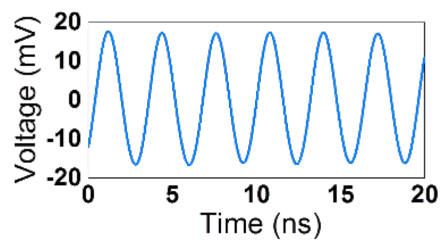
HOW: Augment CMOS with efficient neural devices (MTJs)

- Features for which CMOS may be inefficient (energy and/or device area)
  - Non-volatility
  - Plasticity (local learning)
  - Stochasticity
  - **Oscillators**

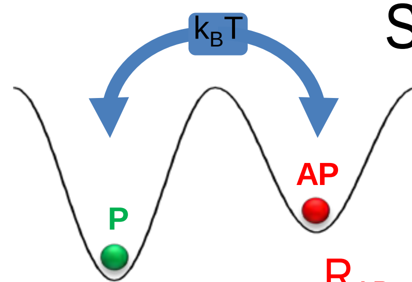
MRAM

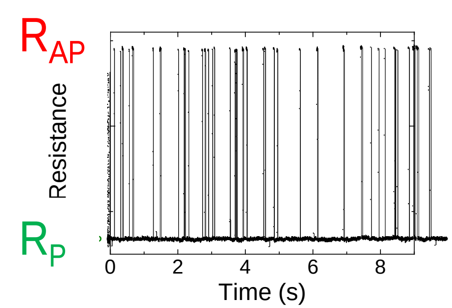


Nano-oscillators



Superparamagnetic tunnel junctions

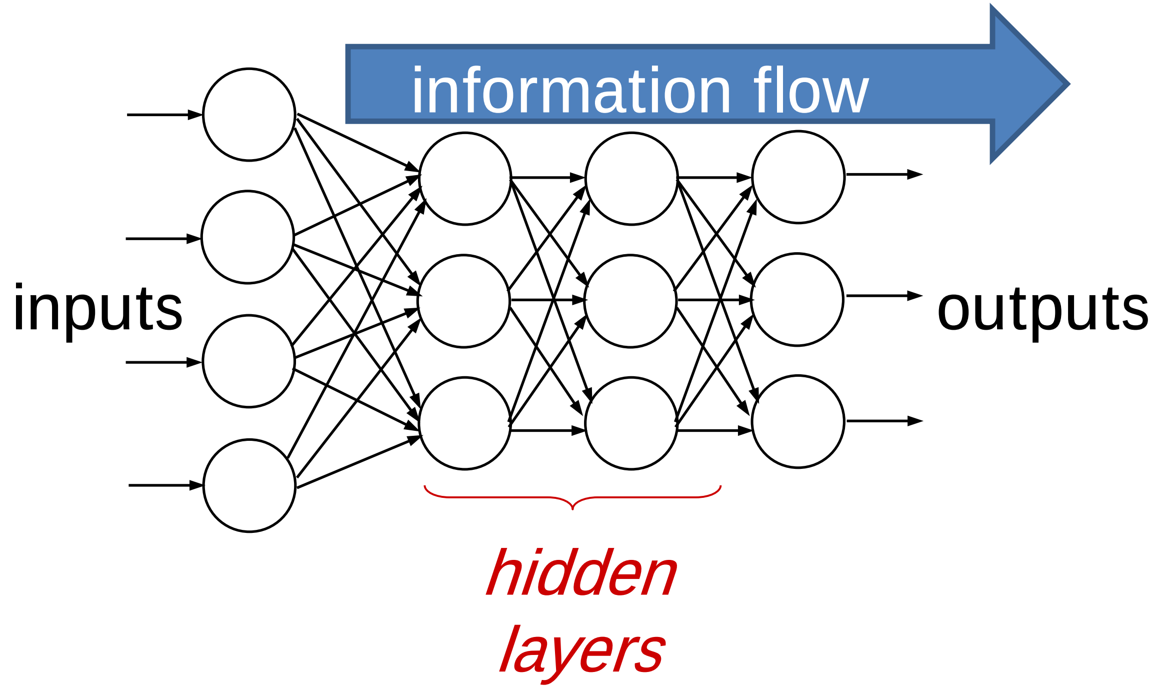




Single oscillator reservoir computing

- **Time multiplexed reservoir computing**
- Sine/Square Identification (Intrinsic memory)
- Spoken digit recognition (Non-linearity)
- Sine/Square Identification (Delayed feedback memory)

Feed-forward networks – one direction of information flow



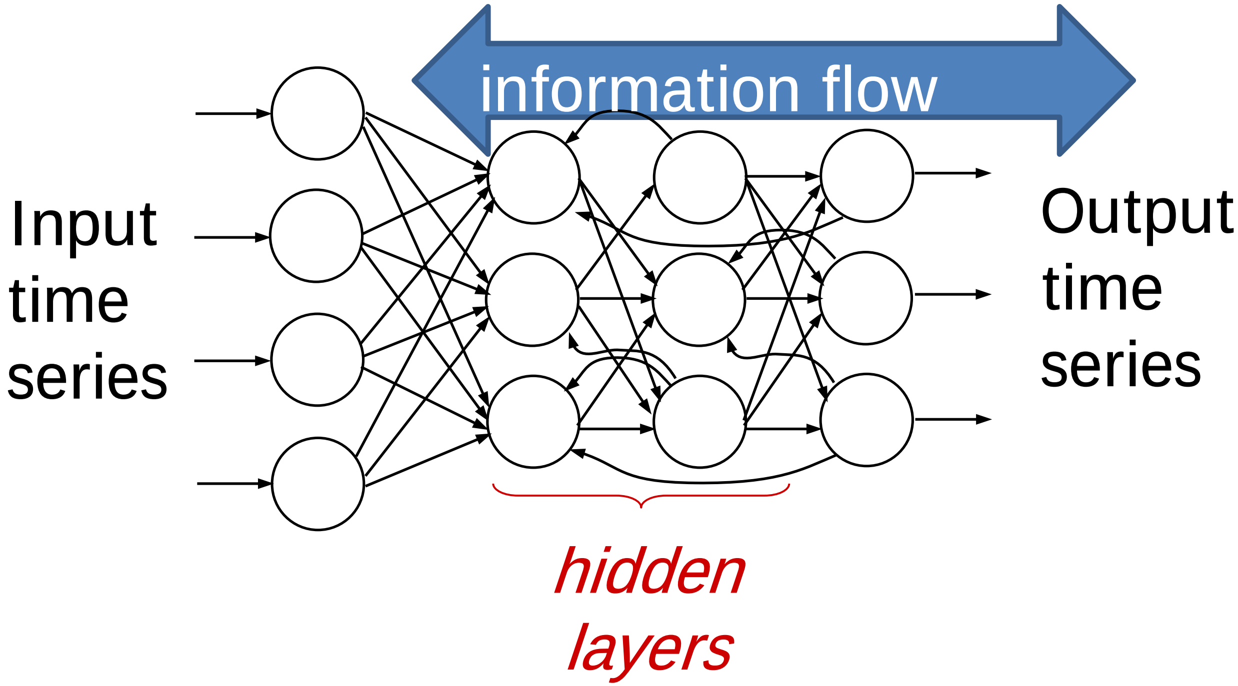
Non-linear nodes (neurons)

rearranges spaces to allow classification

Train off-line

adjust synaptic weights to optimize fit to test data

Recurrent networks – have intrinsic time scales

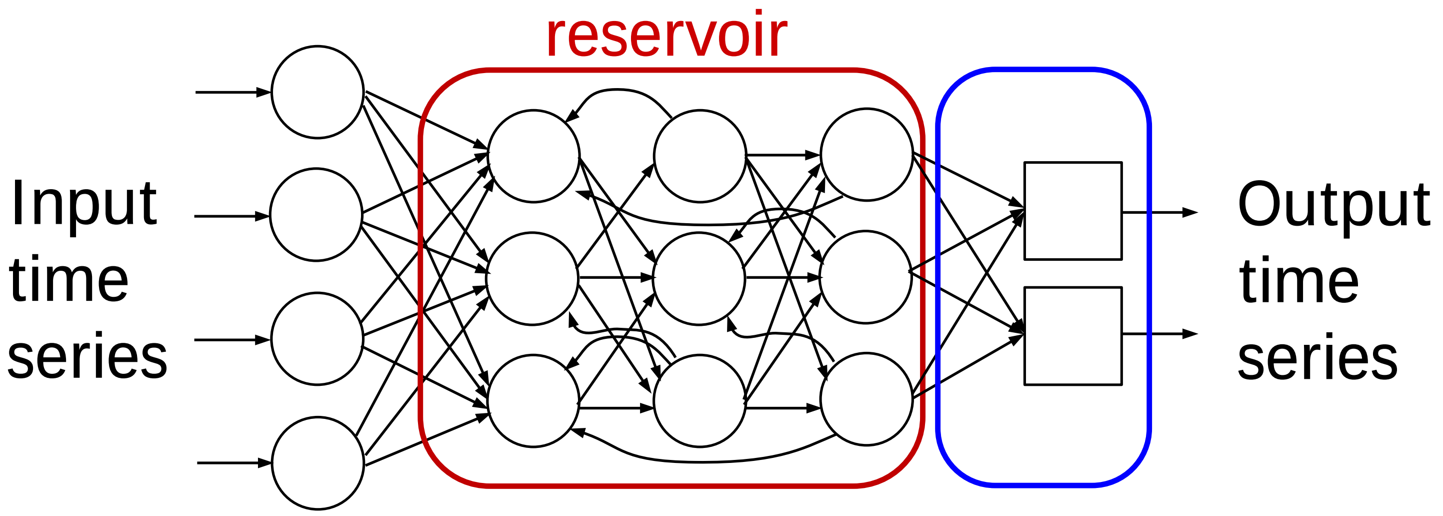


Output time dependent

input becomes time series

Training protocol not simple

Reservoir computing – a simply trainable recurrent network



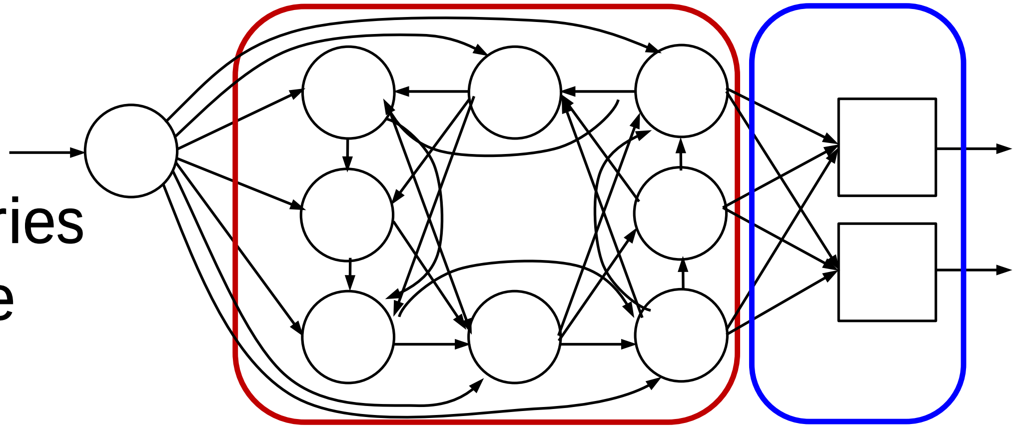
Non-linear,
Fixed synaptic
weights

Linear,
Trained synaptic
weights

Reservoir computing – ring geometry

reservoir

Input time series
to each device

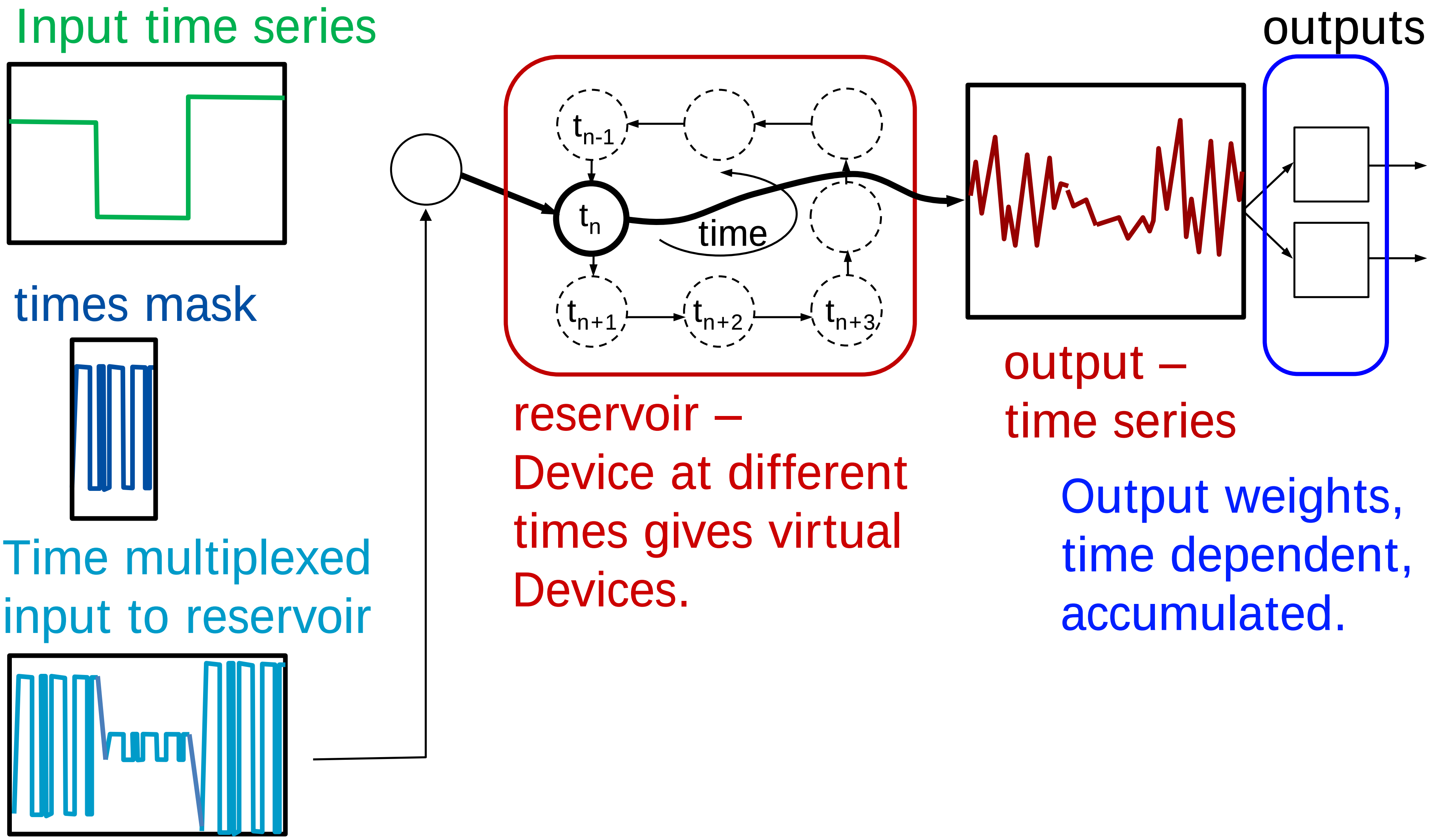


Output
time
series

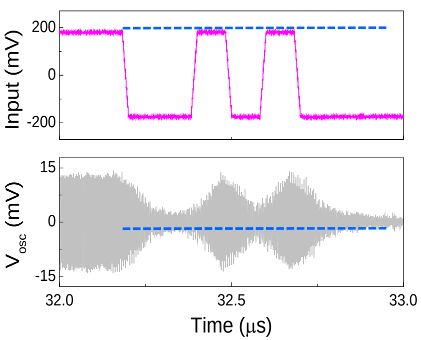
Non-linear,
Fixed synaptic
weights

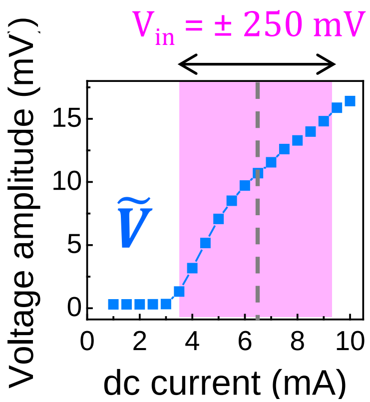
Linear,
Synaptic weights
trained off-line

Reservoir computing – time multiplexed single device



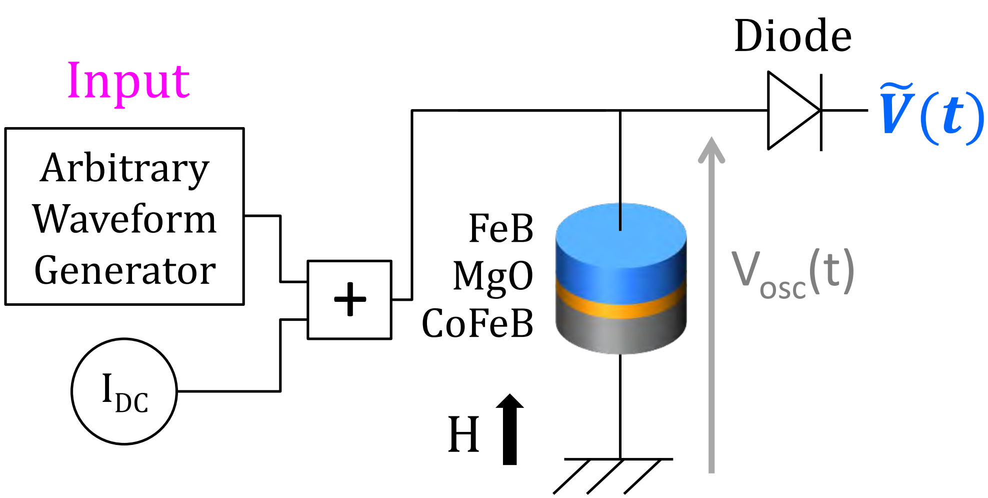
Spin-torque nano-oscillators: non-linear amplitude dynamics and memory

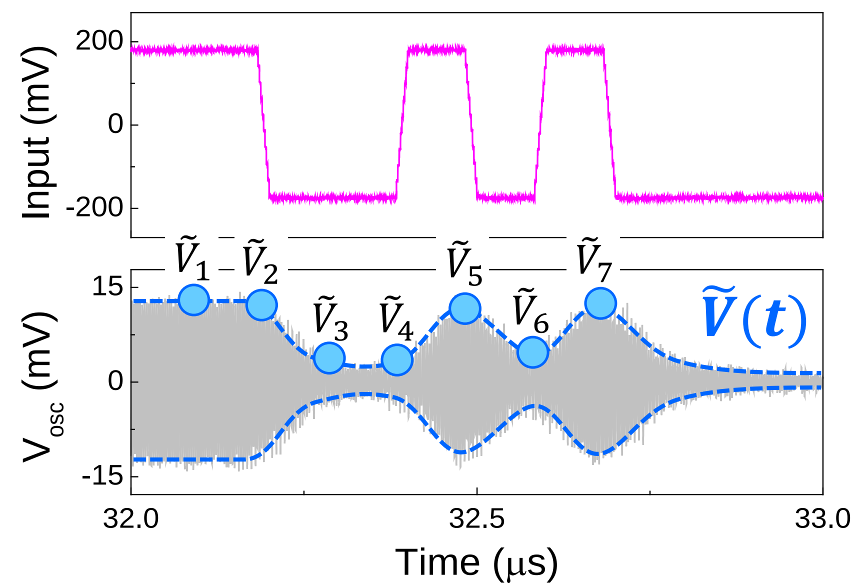




Torrejon *et al*., Nature 547, 428 (2017).

Input: current
Output: amplitude of the oscillator’s voltage

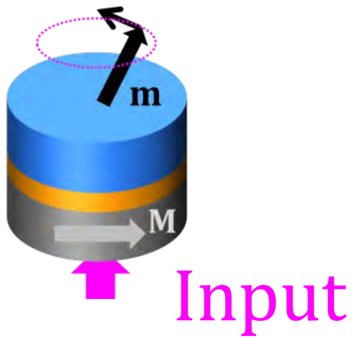


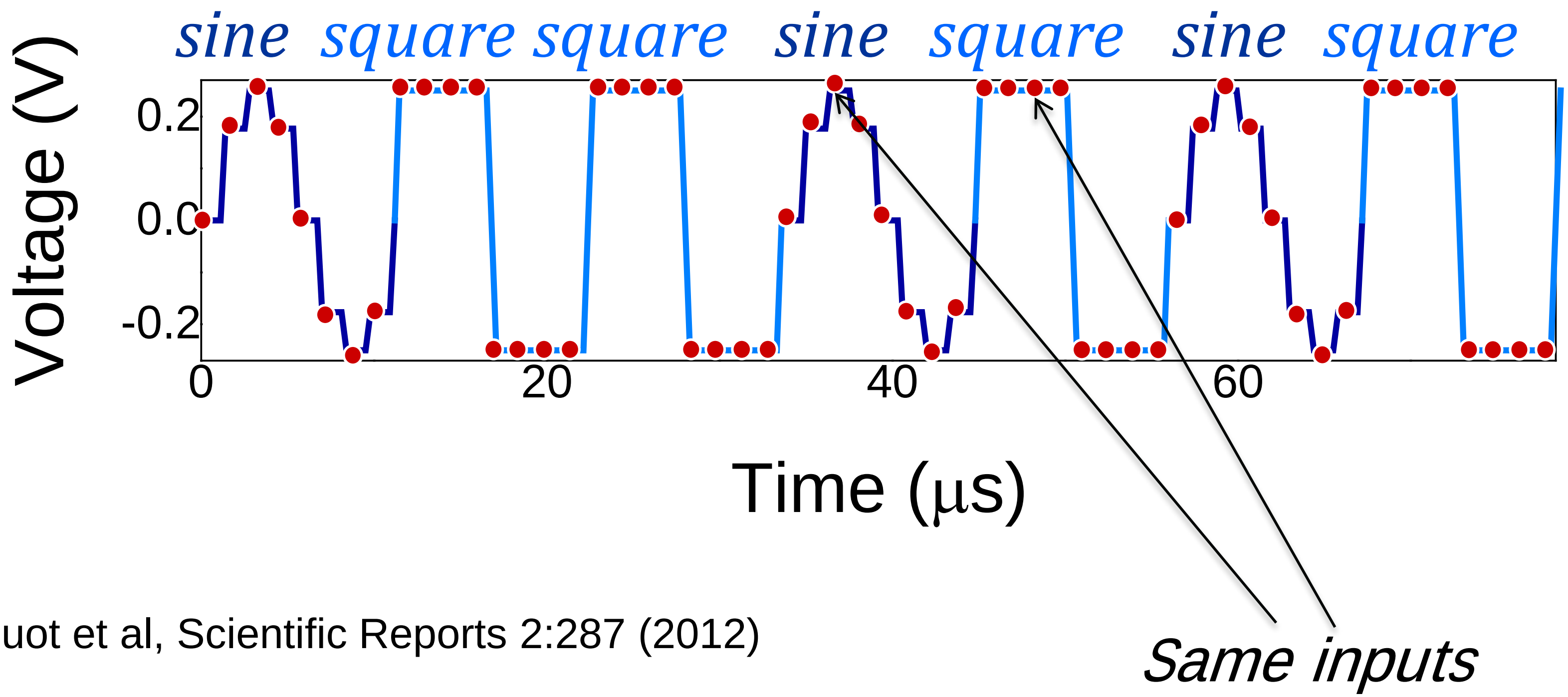


Single oscillator reservoir computing

- Time multiplexed reservoir computing
- **Sine/Square Identification (Intrinsic memory)**
- Spoken digit recognition (Non-linearity)
- Sine/Square Identification (Delayed feedback memory)

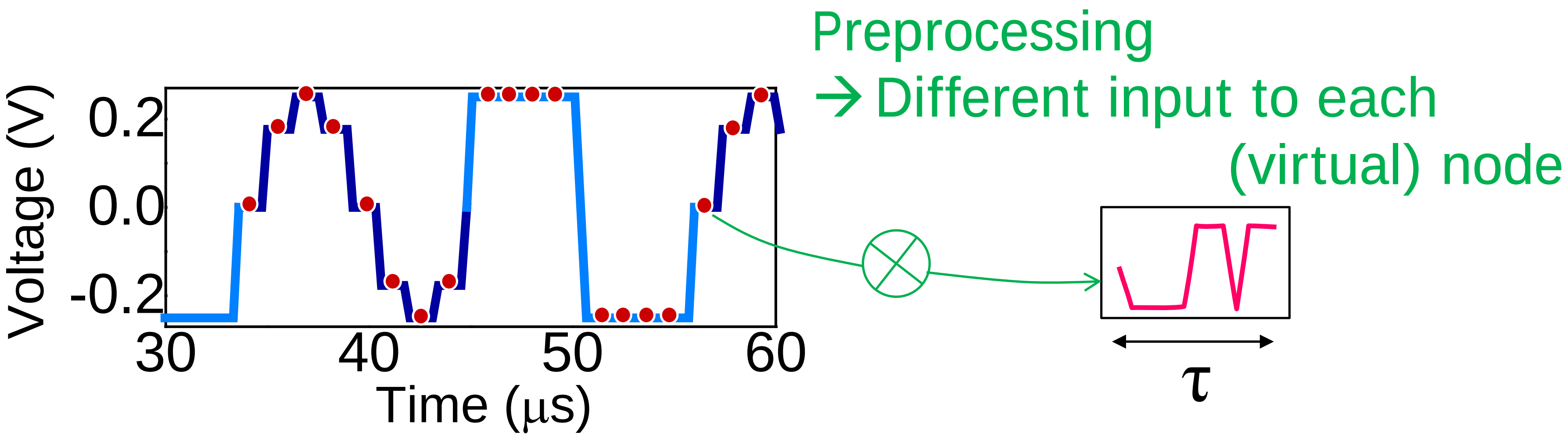
Task: recognizing sines from squares at each point in time with a single oscillator

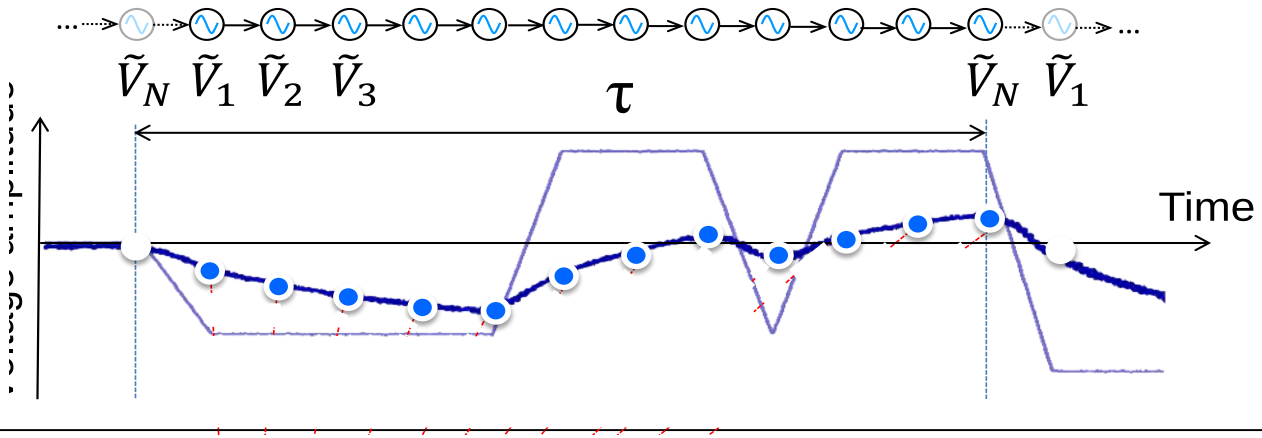




Paquot et al, Scientific Reports 2:287 (2012)

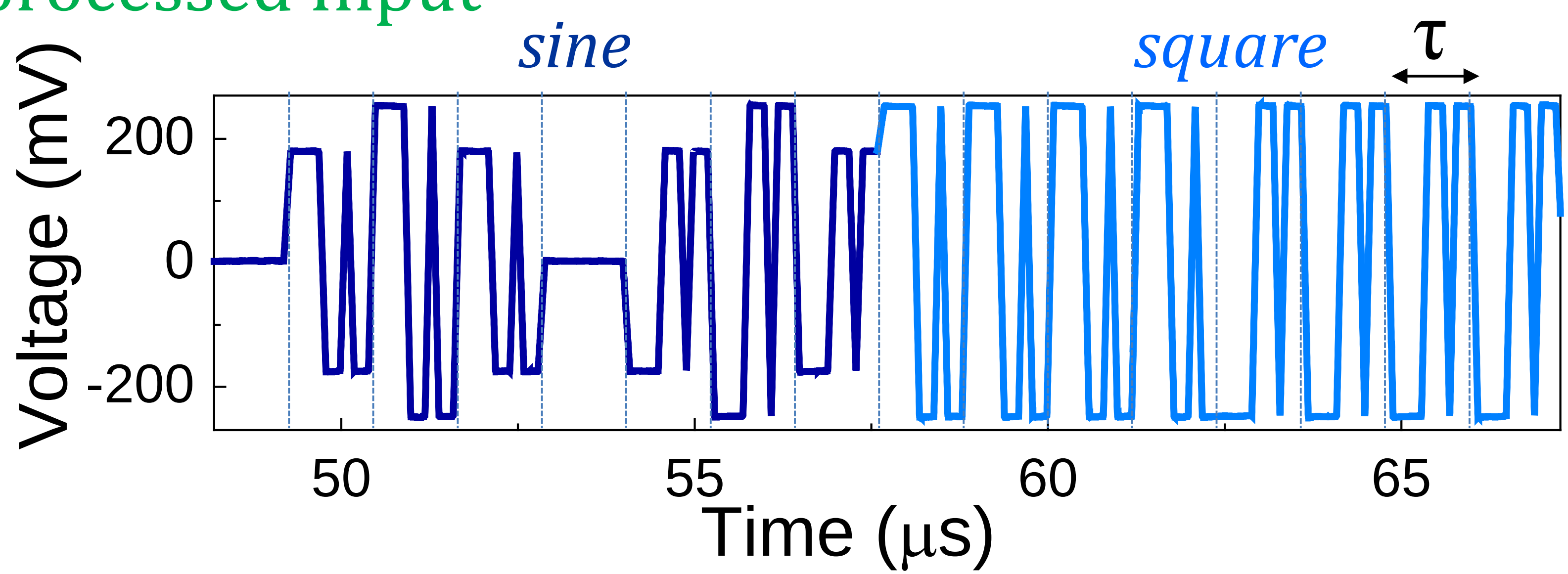
For extrinsic memory, each node should couple to a few other nodes.



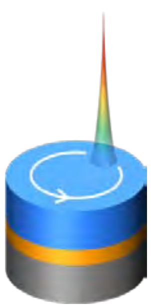


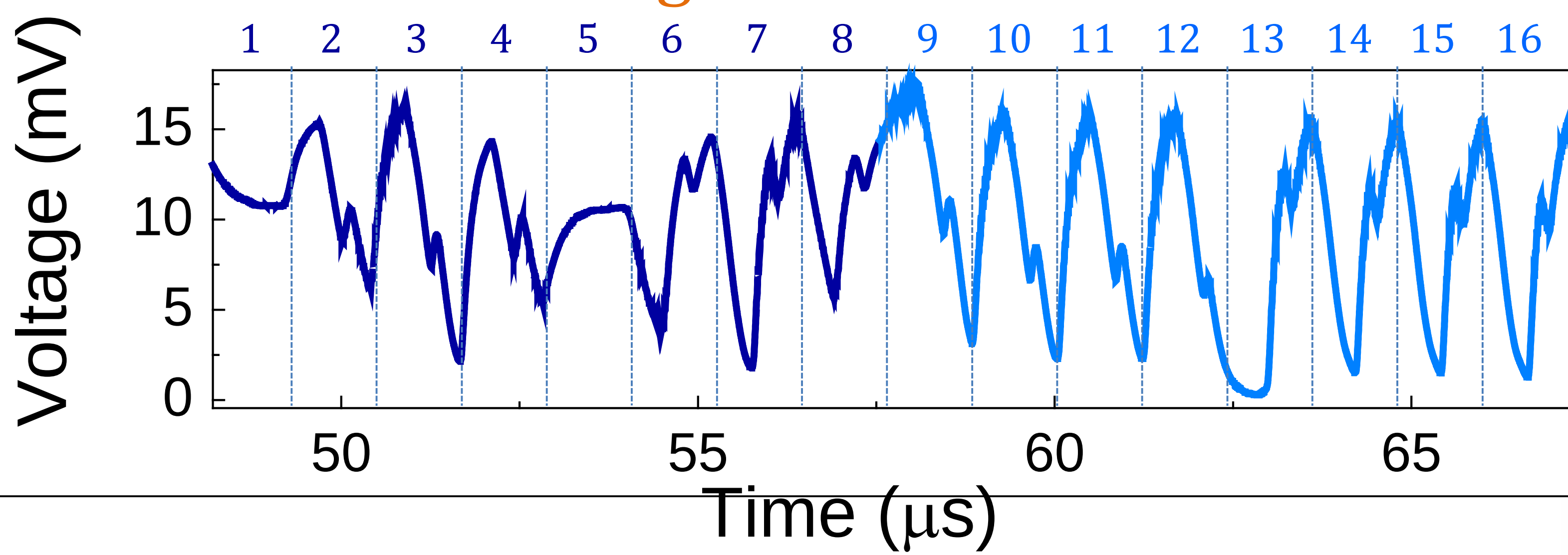
Experimental trajectories of oscillators' amplitude

Preprocessed Input



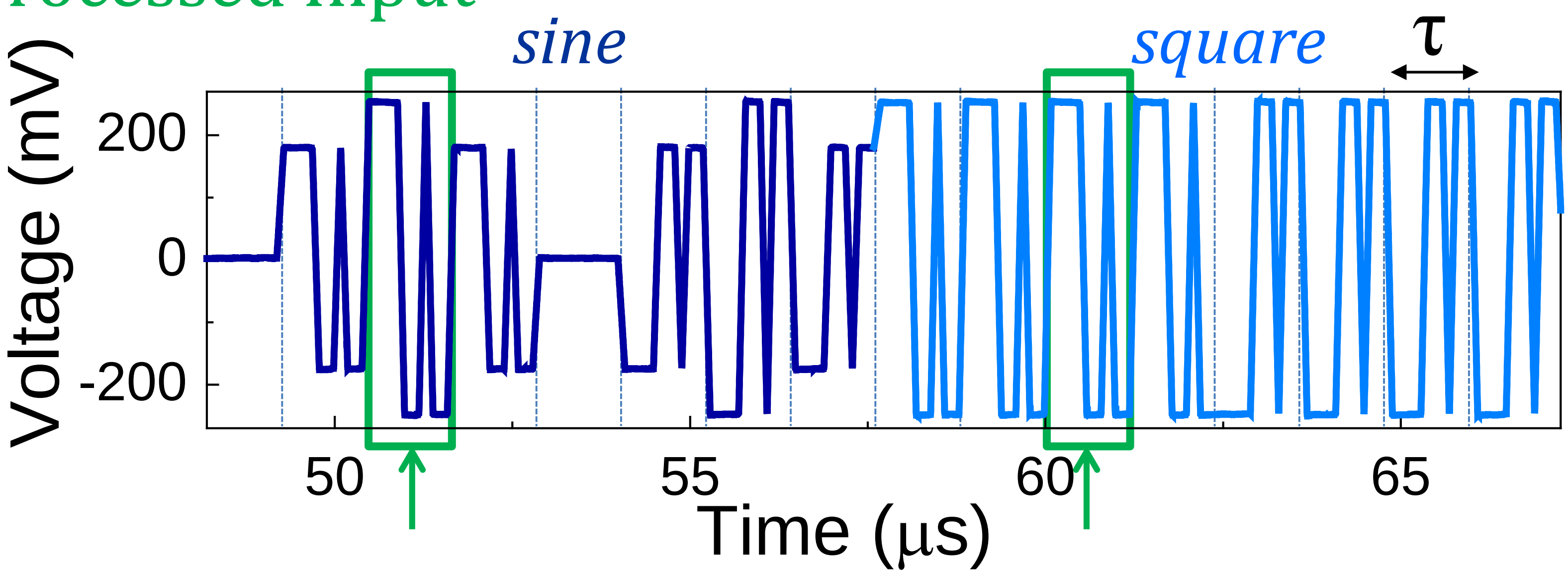
Oscillator's emitted voltage $\tilde{V}(t)$



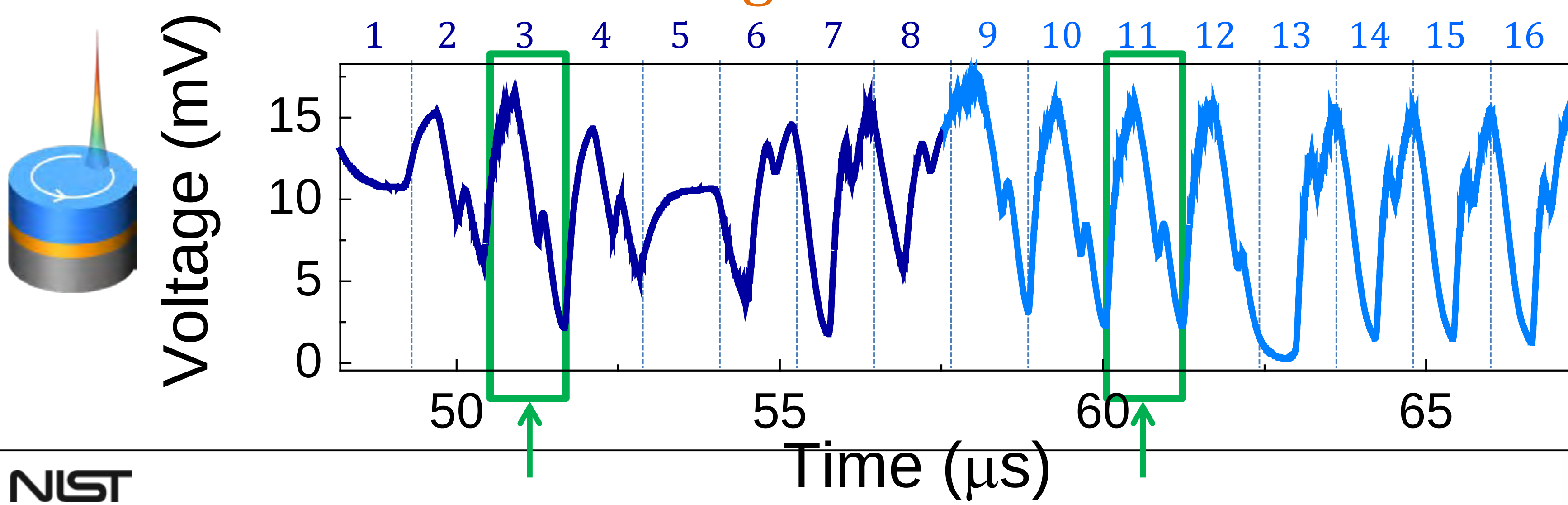


Experimental trajectories of oscillators' amplitude

Preprocessed Input

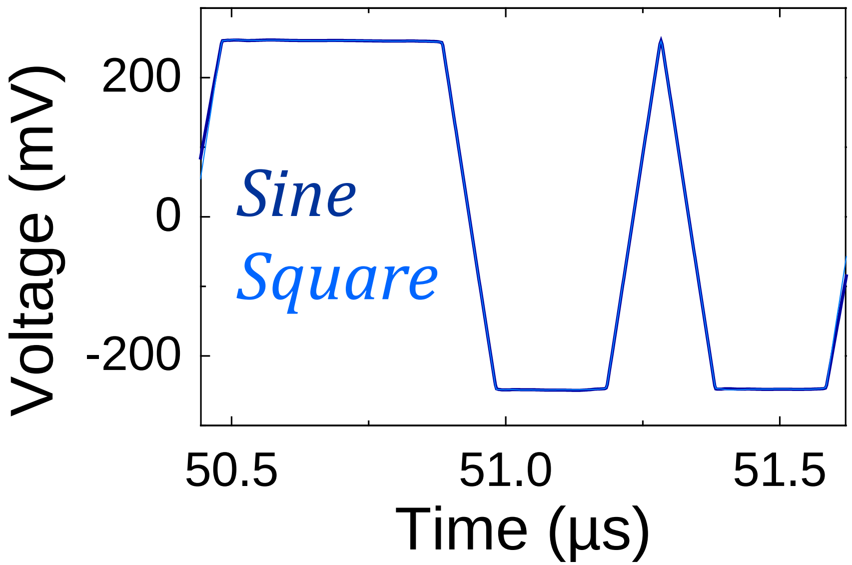


Oscillator's emitted voltage $\tilde{V}(t)$

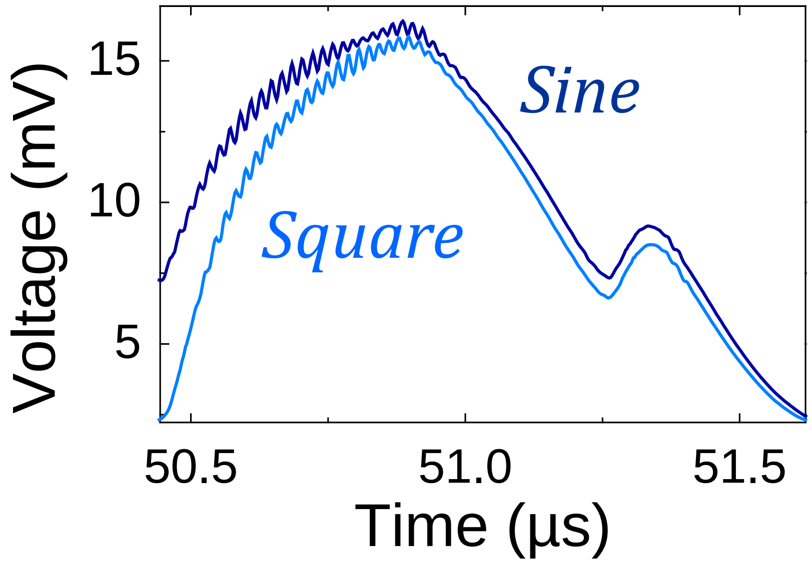


Different trajectories: data separation is achieved

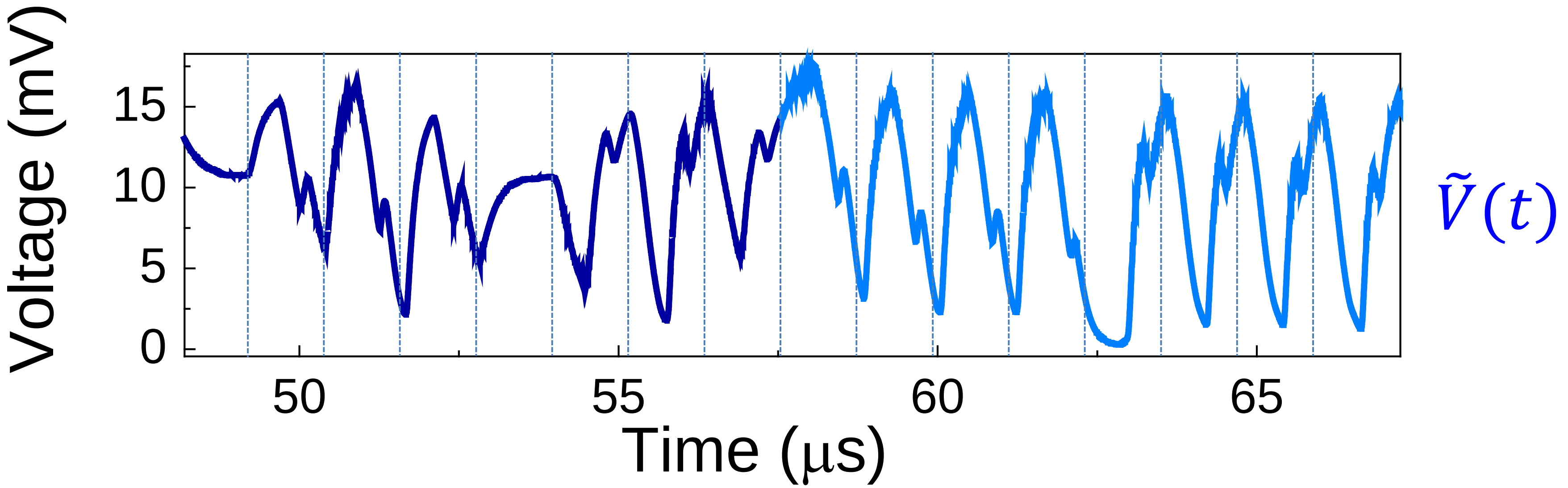
Same input for sine and square

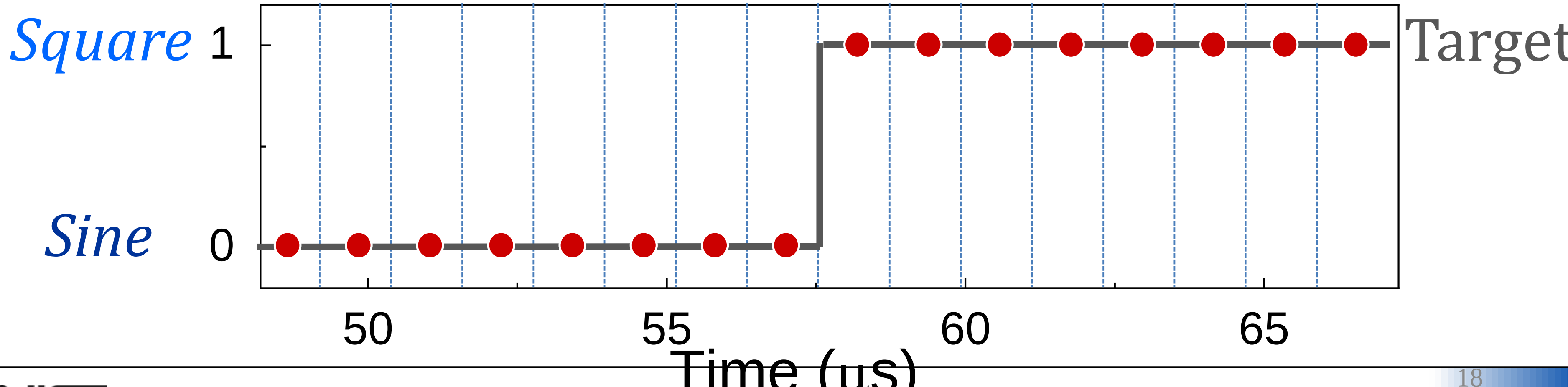


Different outputs $\tilde{V}(t)$



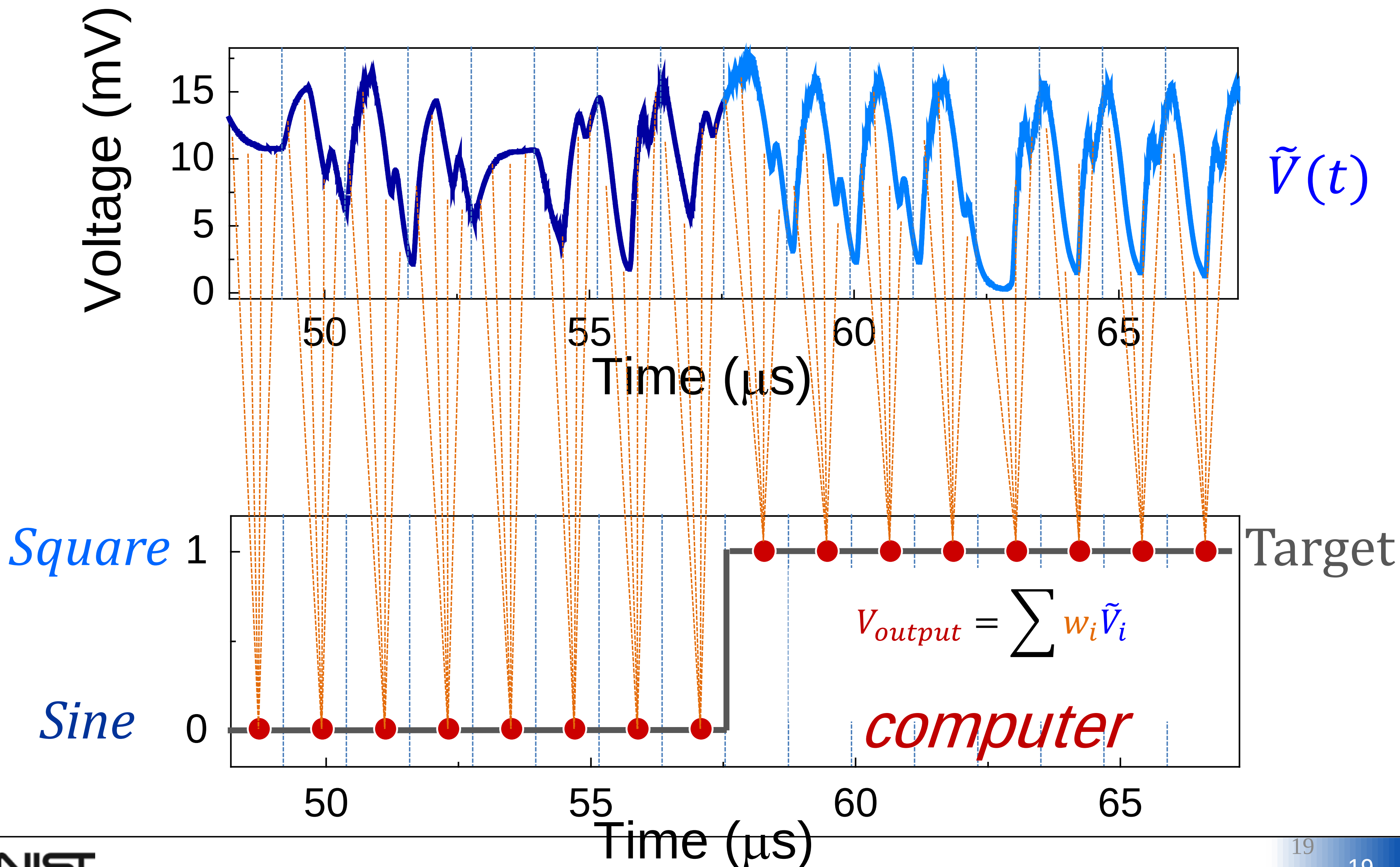
Trajectories need to be grouped to be classified in sines and squares





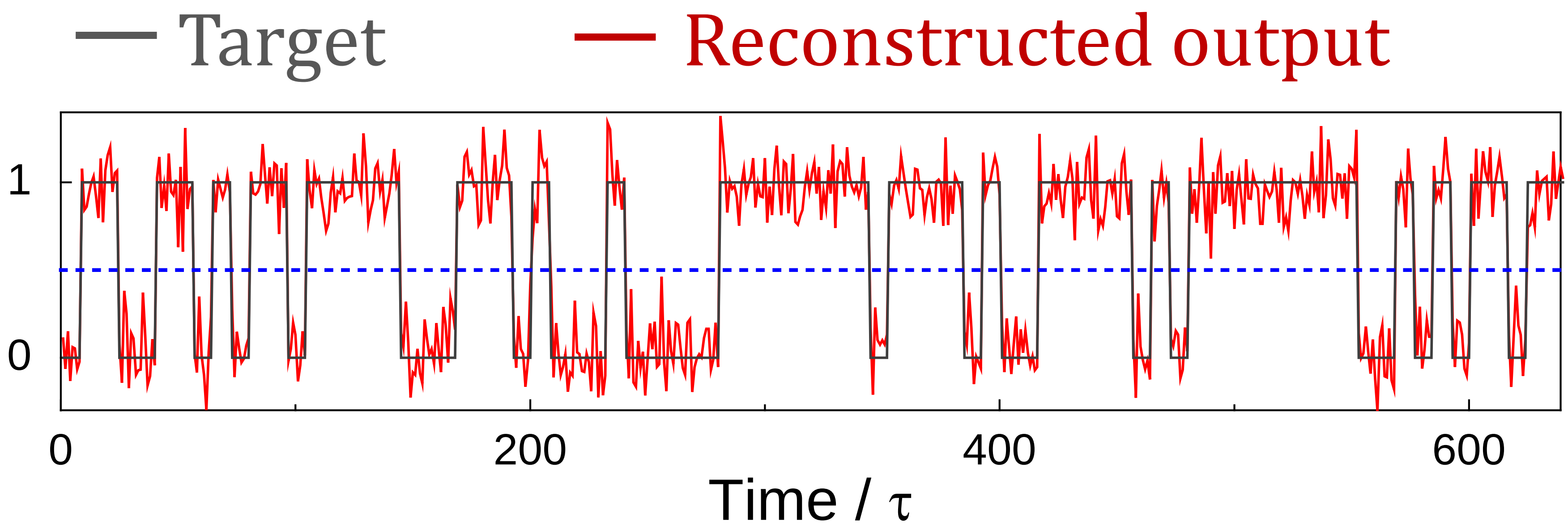
18

Classification: constructing the output
0 for sine, 1 for square



Experimental result : RMS = 10%
perfect classification of sines and squares

waveform with 80 randomly arranged sines and squares



H = 3800 Oe, I_{DC} = 6.4 mA
8 τ per period, 24 nodes, θ = 100 ns
640 first τ for training, 640 next τ for classification

Single oscillator reservoir computing

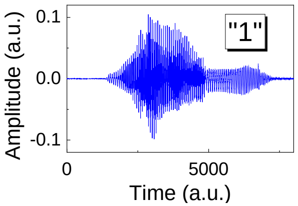
- Time multiplexed reservoir computing
- Sine/Square Identification (Intrinsic memory)
- **Spoken digit recognition (Non-linearity)**
- Sine/Square Identification (Delayed feedback memory)

Spoken digit recognition (NIST TI-46 corpus)

Oscillator

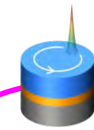
Input: audio file

Output



Spectrogram or Cochlear

Pre-processed input



Recorded trace

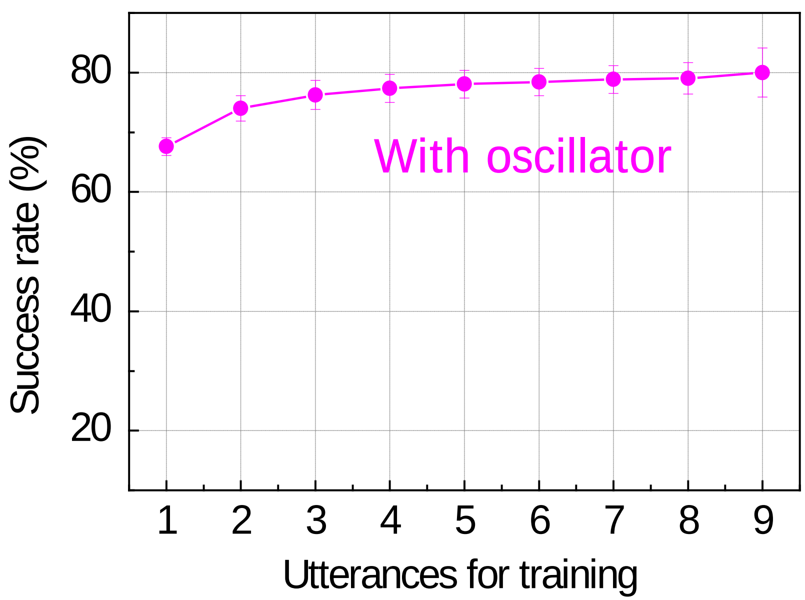
"1"

Acoustic features

Pre-processing

Computer

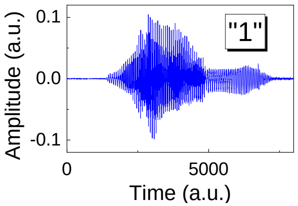
Spectrogram



Spoken digit recognition (NIST TI-46 corpus)

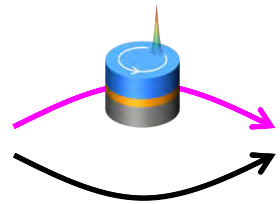
Oscillator

Input: audio file



Spectrogram or Cochlear

Pre-processed input



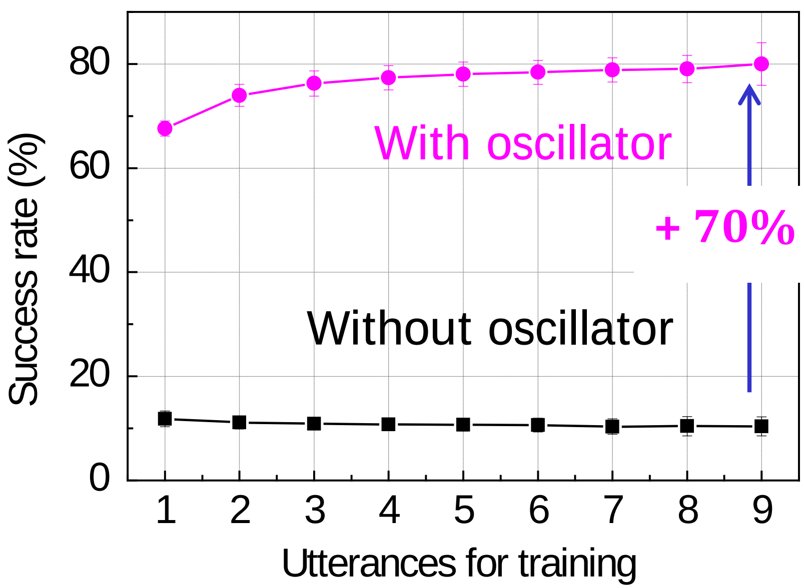
Recorded trace

Output

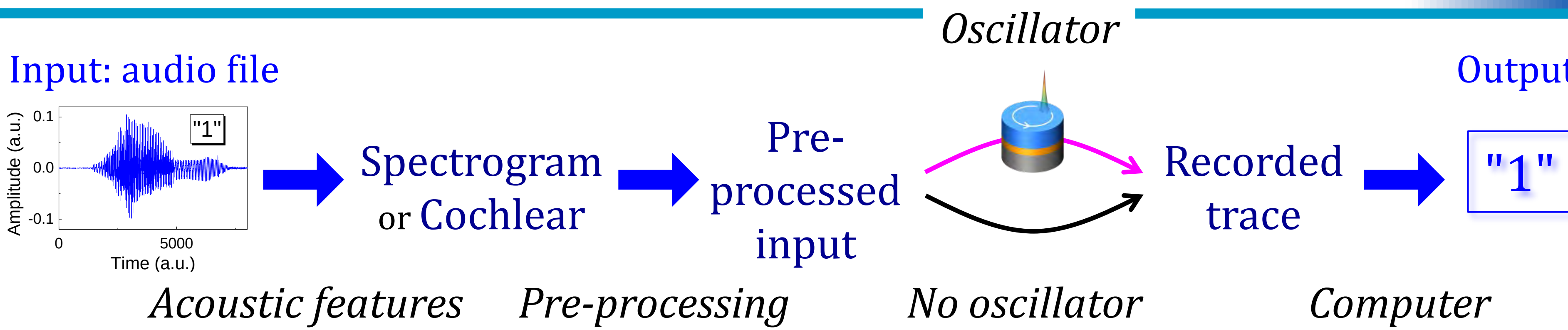
"1"

Acoustic features *Pre-processing* *No oscillator* *Computer*

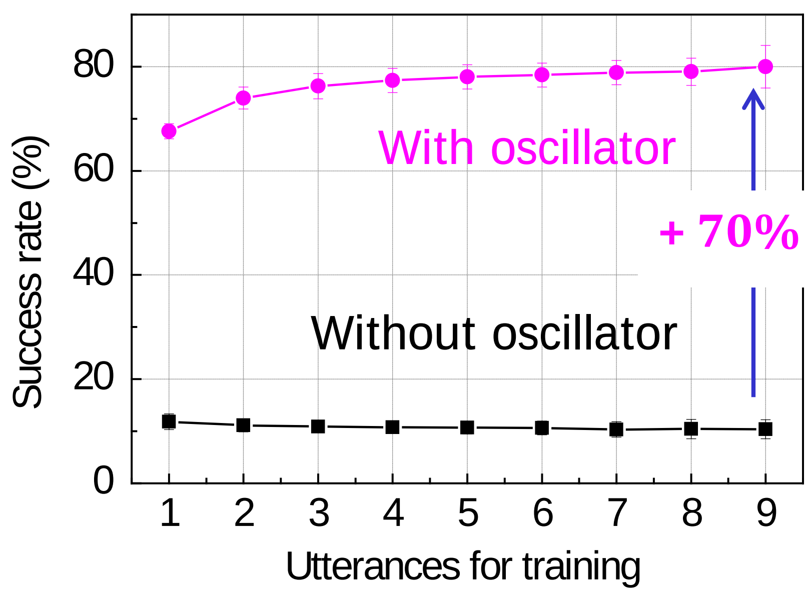
Spectrogram



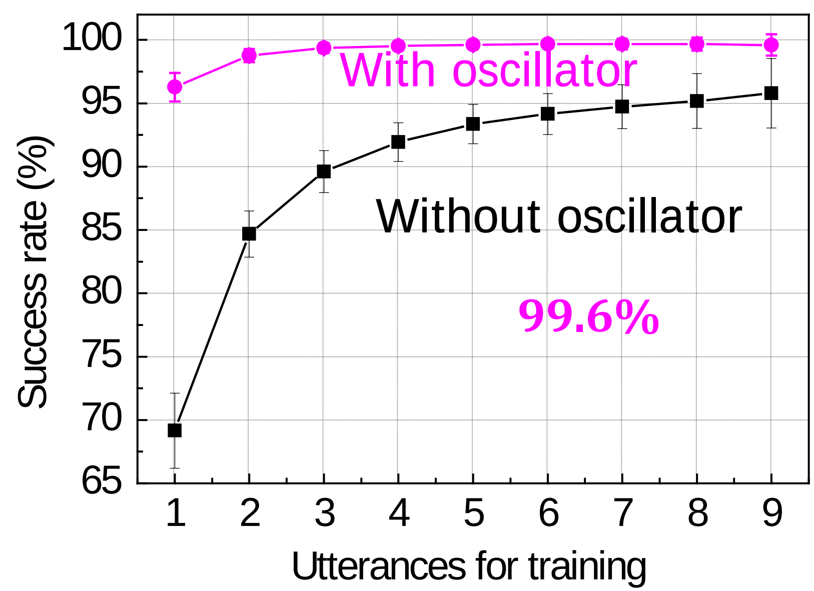
Spoken digit recognition (NIST TI-46 corpus)



Spectrogram



Cochlear

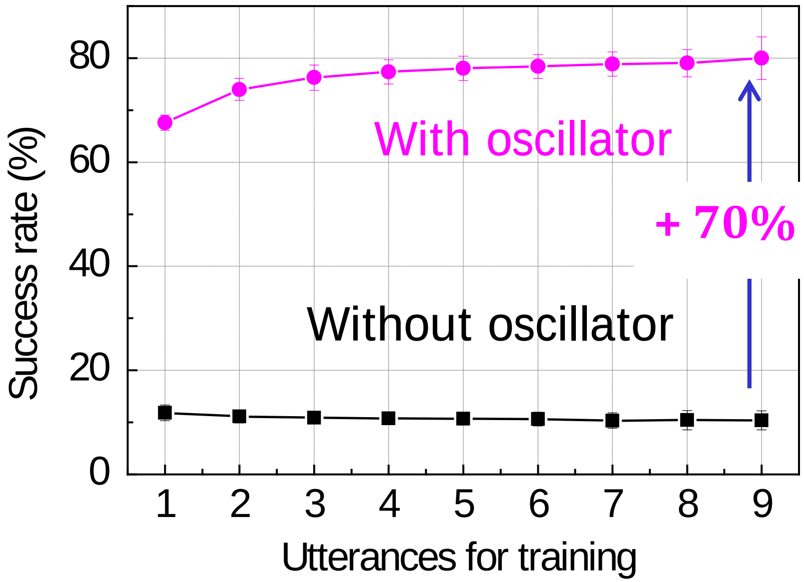


Spoken digit recognition (NIST TI-46 corpus)

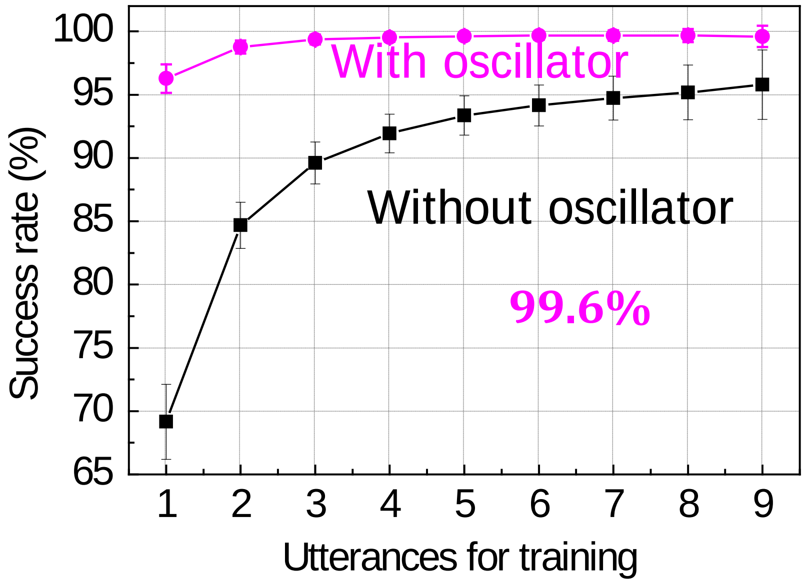
Jacob Torrejon-Diaz, Mathieu Riou, Flavio Abreu-Araujo, et al, Nature 547, 428-431 (2017)

State of the art: 96 to 99.8 %

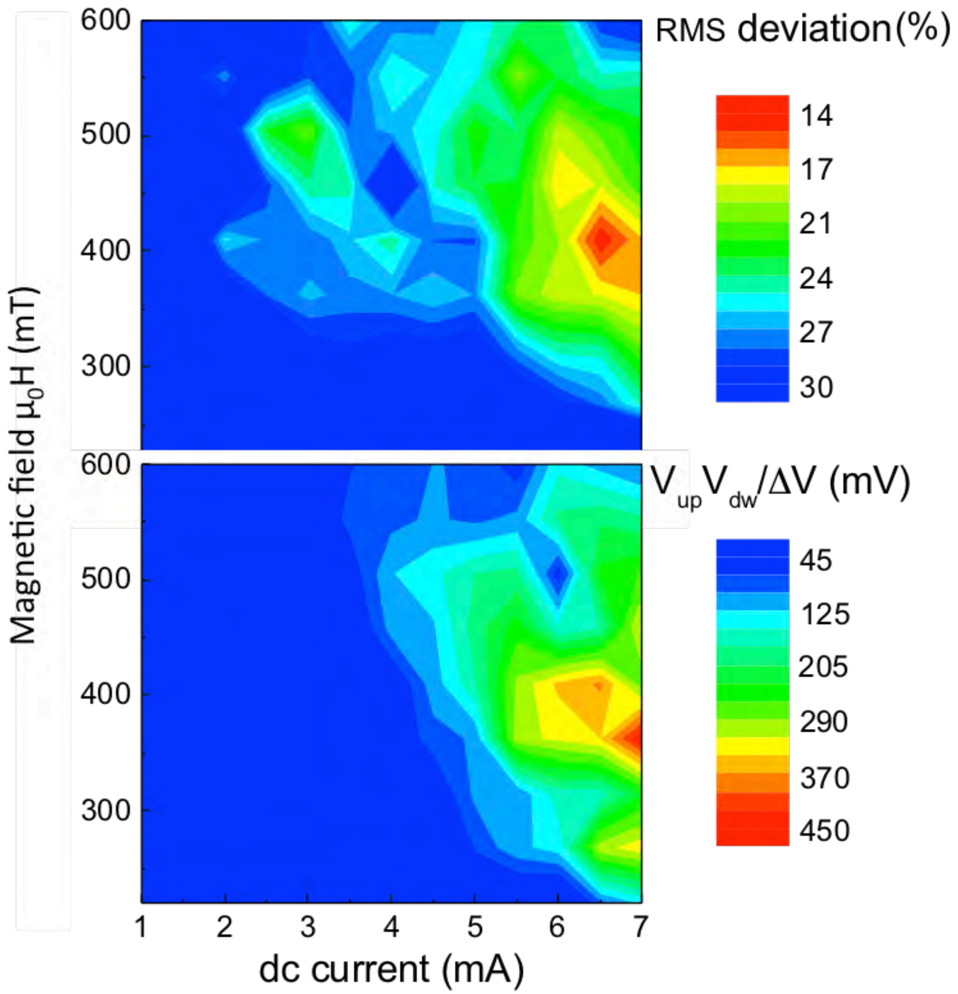
Spectrogram

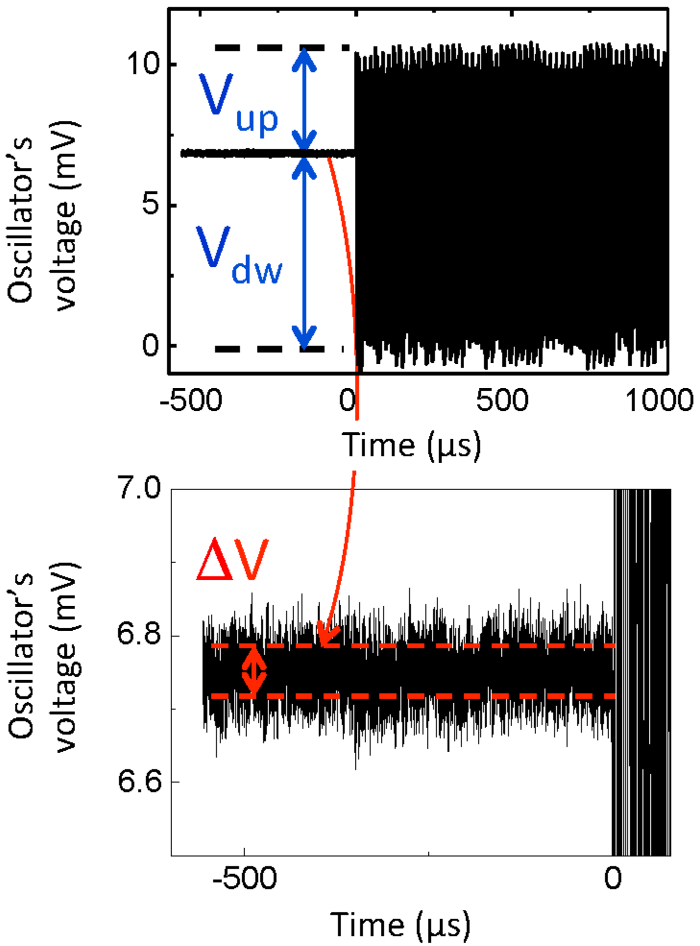


Cochlear



Recognition results are sensitive to the noise, an optimal bias area is found

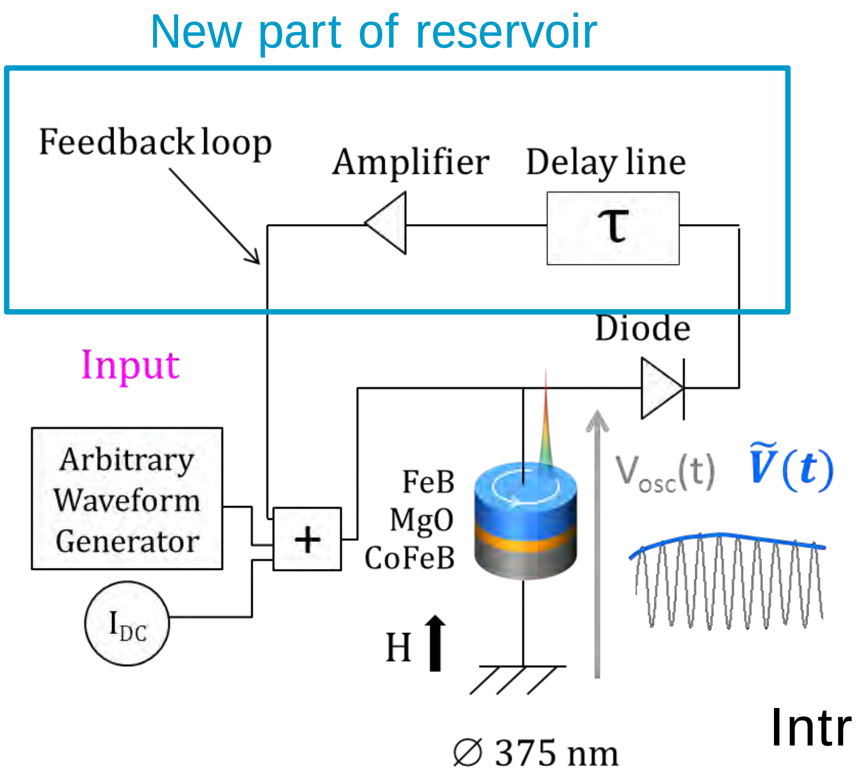


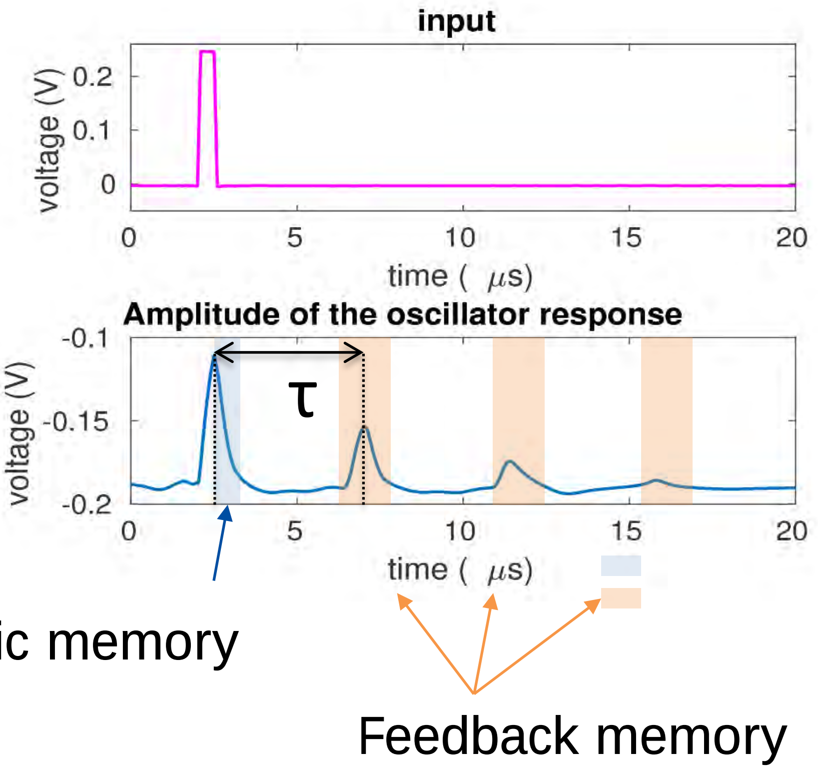


Single oscillator reservoir computing

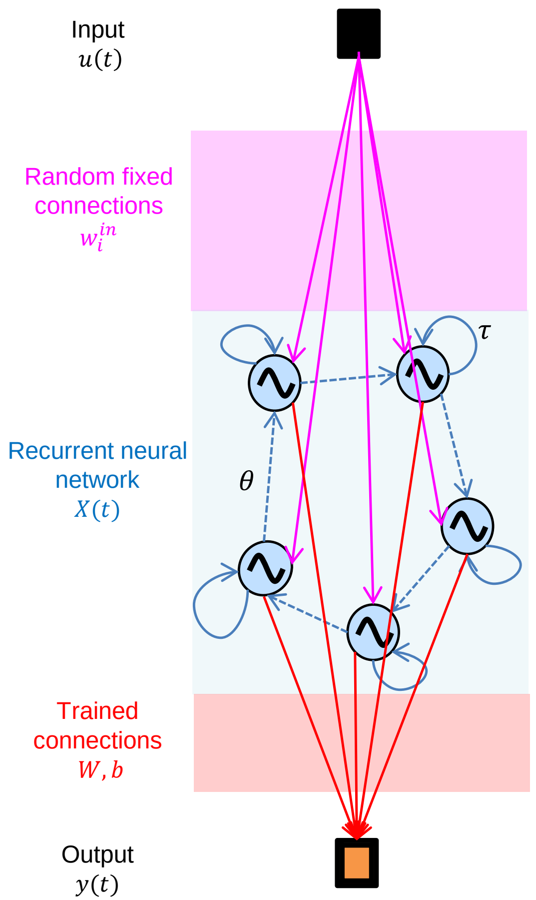
- Time multiplexed reservoir computing
- Sine/Square Identification (Intrinsic memory)
- Spoken digit recognition (Non-linearity)
- **Sine/Square Identification (Delayed feedback memory)**

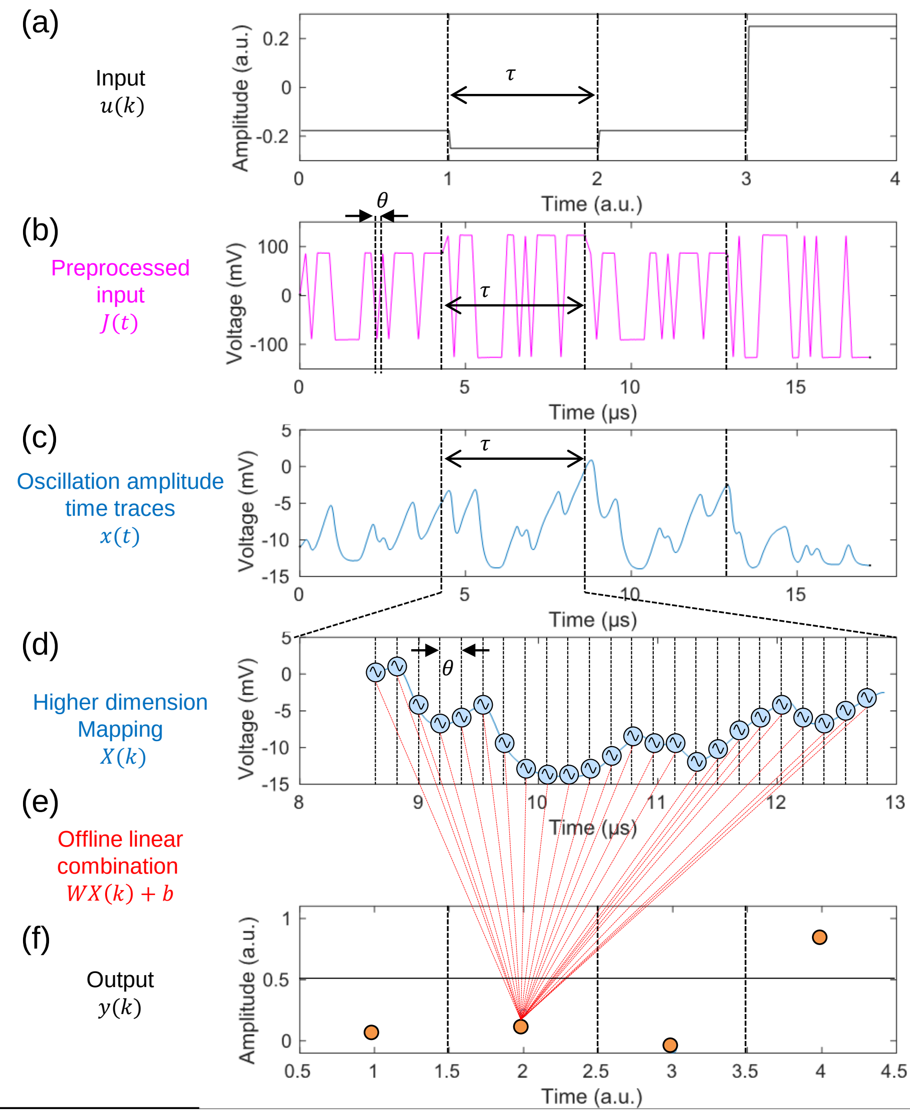
Replace intrinsic memory with delayed feedback



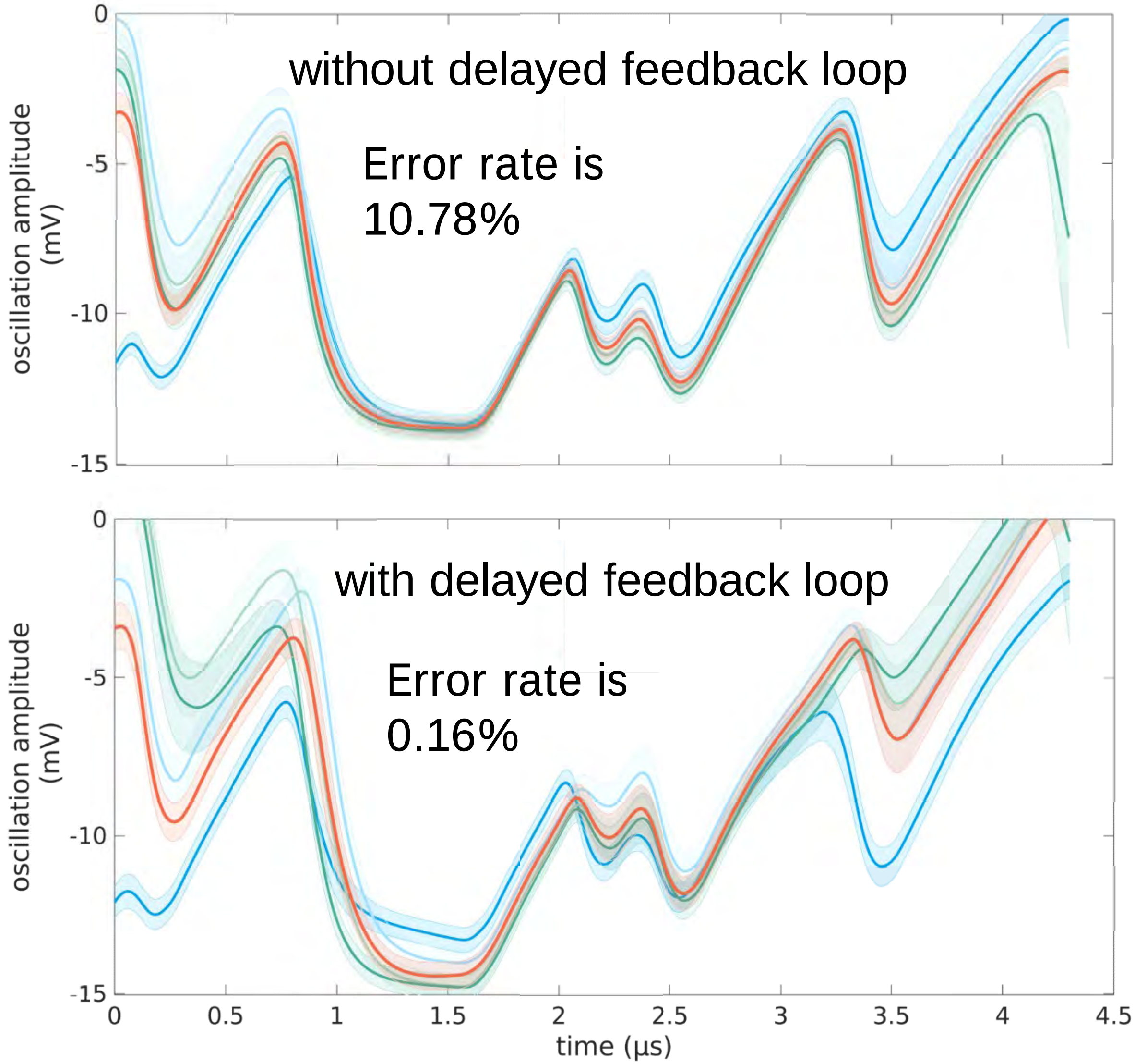


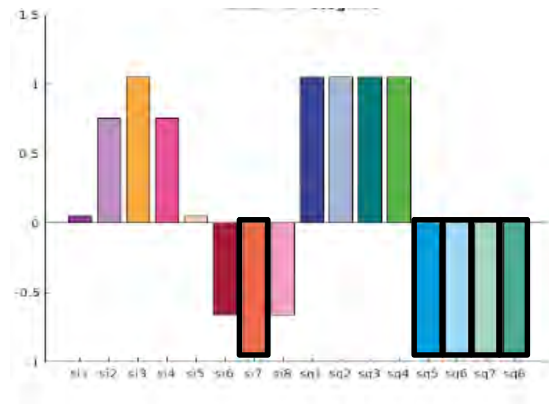
To focus on feedback, nodes only couple to past states of themselves





Feedback allows separation of similar inputs





Single oscillator reservoir computing

- Augmenting CMOS based neuromorphic circuits with energy-efficient spintronic devices.
  - MTJs already accessible in stat-of-the-art BEOL CMOS
- Time multiplexed reservoir computing.
  - Memory (intrinsic or delayed feedback) allows context dependent discrimination.
  - Single oscillator achieves state of the art at spoken digit recognition – non-linearity.
- Where to?
  - Small low power oscillators.
  - Efficient coupling.
  - Appropriate algorithms