UNIVERSITÀ DI PISA

Intrinsic Plasticity in Echo State Networks

Seminar ID: 25

Computational Neuroscience 22/23

Irene Pisani - i.pisani1@studenti.unipi.it - 560104

M.Sc. Computer Science, Artificial Intelligence

Computer Science Department

September, 2023

Introduction to Intrinsic Plasticity [1][2]

Biological motivations and plausibility

Individual biological neurons change their intrinsic excitability through the modification of voltage gated channels.

- Intrinsic Plasticity (IP) is a form of plasticity that models a well-known phenomenon observed in a variety of biological neuron models called homeostatic plasticity: most of biological neurons tend to autonomously adapt to a fixed average firing rate for physiological reasons.
- Biological mechanisms are not yet known precisely: plausibly every single neuron tries to balance the conflicting requirements of maximizing its information transmission while obeying constraints on its energy expenditure/metabolic cost.
- The goal of IP mechanism is to obtain an approximately exponential distribution of the firing rate level that allow the neuron to transmit the maximum amount of information given a fixed level of metabolic cost/energy expenditure.

→ Investigate IP's formal computational counterpart in combination with standard artificial network learning algorithms.

IP formalizes these biological hypotheses by assuming 3 principles:

(1) **Information maximization.** The output of the neuron should contain as much information on the input as possible. This is achieved by maximizing the entropy of the output firing rates.

(2) **Constraints on the output distributions.** Constraints usually are limited output range of the neuron, but can be also the limited energy available.

(3) **Adapt the neurons intrinsic parameters.** A biological neuron is only able to adjust its internal excitability, not the individual synapses.

IP Goal: derive a gradient descent learning rule from these principles that adjust the parameter of a nonlinear transfer function to moves it close to the optimal nonlinearity and drive the neuron's firing distribution towards an exponential distribution, as observed in visual cortical neurons.

Derivation of IP rules [1][2]

For single continues activation model neurons

Consider a model neuron where x, input to the neuron's firing mechanism, is passed through a parametrized nonlinearity f to obtain the neuron's output

$y = f_{gen}(x) = f(ax + b)$. Given:

- $\tilde{p}(x)$, the actual probability density of a neuron's output activity,
- $p(x)$, a desired probability density function,

The goal is to adjust f_{gen} s.t. $\tilde{p}(y) \approx p(y)$, i.e., minimize the distance

$D_{KL}(\tilde{p}, p)$, **how D change if f_{gen} changes?**

f_{gen} depends on 2 parameters a and b. To construct the gradient rule to minimize D we must consider the partial derivatives of D wr.t. a and b.

The resulting **stochastic gradient descent rule for IP that minimize D** is given by:

$$a = a + \Delta a \qquad b = b + \Delta b$$

(1) **Fermi activation neuron** $y = sigmoid(x)$

- $p_{\exp} = \frac{1}{\mu}\exp(-\frac{y}{\mu})$ is the target exponential output distribution
- a and b are update with the following factors

$$\Delta a = \frac{\eta}{a} + \Delta b x \qquad \Delta b = \eta(1 - (2 + \frac{1}{\mu})y + \frac{1}{\mu}y^2)$$

(2) **Hyperbolic tangent activation neuron** $y = tanH(x)$

- $p_{norm} = \frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{(y-\mu)^2}{2\sigma^2})$

  is the target gaussian output distribution
- a and b are update with the following factors

$$\Delta a = \frac{\eta}{a} + \Delta b x \qquad \Delta b = -\eta(-\frac{\mu}{\sigma^2} + \frac{y}{\sigma^2}(2\sigma^2 + 1 - y^2 + \mu y))$$

Convergence of IP rules [1]

For bounded activation neurons and recurrent networks

IP learning rules are able to converge to the desired output distributions?

- Problem: due to the bounded nature of transfer function the output of the neurons cannot output arbitrary values.

  → Moments of the desired distribution will not be accurately approximated by a learning rule controlling a neuron with finite output range.

- Given the moments of the desired distribution p, it's possible to compute the effects of these bounds on the moments of the actual output distribution.

  → Mathematical derivation of the moments to which IP learning rule is supposed to converge.

IP learning is capable of driving the neurons to the desired exponential/Gaussian output distributions with the theoretically derived moments, for non recurrent single neuron.

IP rules are based on assuming independence of the neurons' input distributions; **Could IP manages to approximate the desired output distributions in large network in the presence of interfering recurrency?**

- Empirical results show that even in the case of using IP in a highly connected recurrent network, a very good approximation of the theoretical distributions it's reachable.

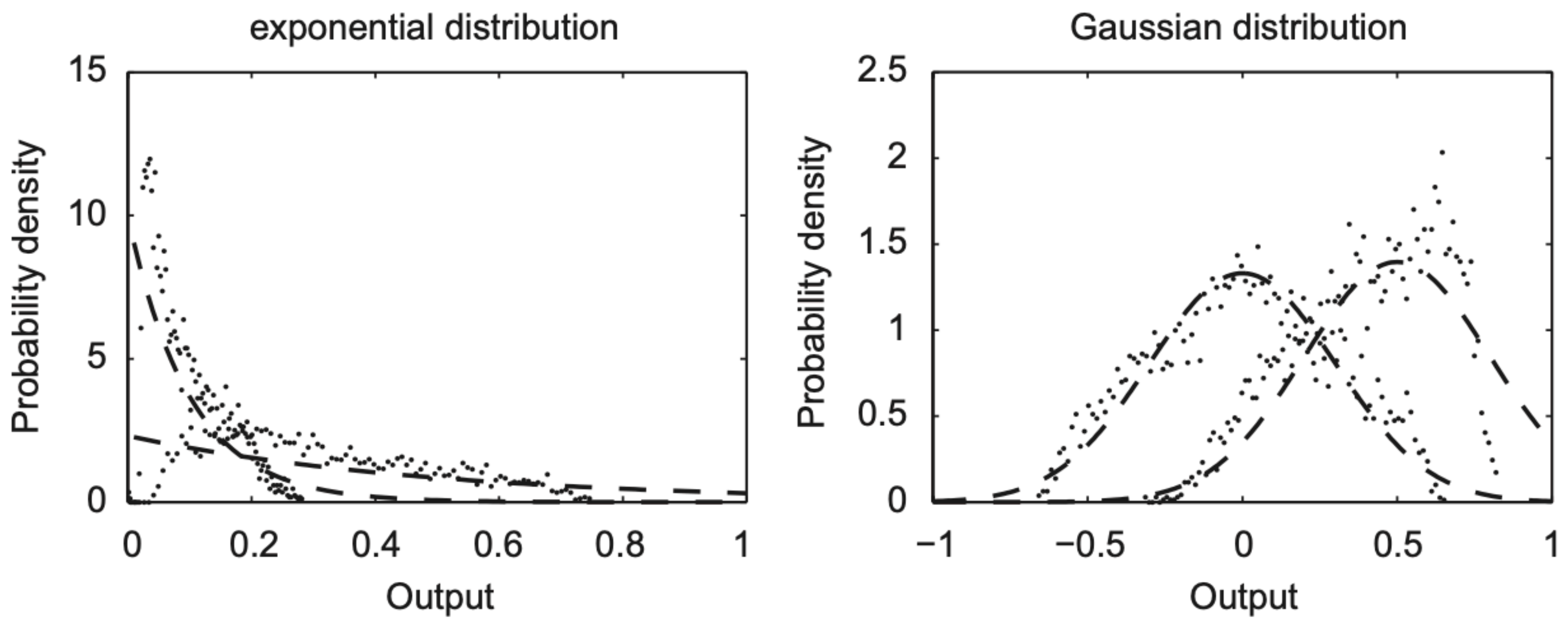


Fig. theoretical (dashed) and actual (dotted) probability densities for both the fermi/exponential and tanh/Gaussian case, for two different settings of the desired first moments and a reservoir of 100 neurons.

IP rule in Echo State Network (ESN) [1]

IP impact on reservoir optimization and ESN performance
Benchmarks: Narma 30, Memory Capacity, Digit recognition form speech

Optimization of ESN's reservoir for applications is typically based on experience, heuristics and on a brute-force search of the parameter space:

- Performance variance across different reservoirs with the same spectral radius is still substantial ➞ A simple computational way to adapt the reservoirs to the task would be welcome.

The performance of ESN networks with Fermi/tanH transfer functions can be improved by using IP rule:

- IP rule is local, input driven, and adapts the reservoir in an unsupervised way maximizing the information content of the reservoir states.
- It empowers the reservoirs by autonomously and robustly adapt their internal dynamics, irrespective of initial weight setting, input scaling or topology, to a dynamic regime which is suited for the given task.
- It results in exponential /normal distributed reservoir states.

- Performance improvements is evident when pre-adapting the reservoir using IP. **Spread on the performance when creating random reservoirs can be decreased using IP rule to pre-adapt the reservoir weights**.
- For tasks that mainly need linear responses (Memory and NARMA), the Gaussian distribution performs best, while on tasks that are non-linear (Speech), the exponential distribution performs best.

Best results for the three different benchmarks

	Fermi–specrad	Tanh–specrad	Fermi–exp. IP	tanh–Gauss. IP
MC	17.41 (1.48)	29.78 (1.87)	20.32 (0.77)	**31.31 (1.93)**
NARMA	0.77 (0.012)	0.52 (0.050)	0.74 (0.019)	**0.46 (0.042)**
Speech	0.070 (0.018)	0.069 (0.015)	**0.060 (0.012)**	0.069 (0.015)

IP is better than ranging the spectral radius, both in average performance as in standard deviation (except for one case), denoted between brackets. Some tasks perform better with a Gaussian distribution, others with an exponential distribution. Bold values denotes the best performance.

IP makes possible to use node types and topologies which normally perform very poorly as reservoir.

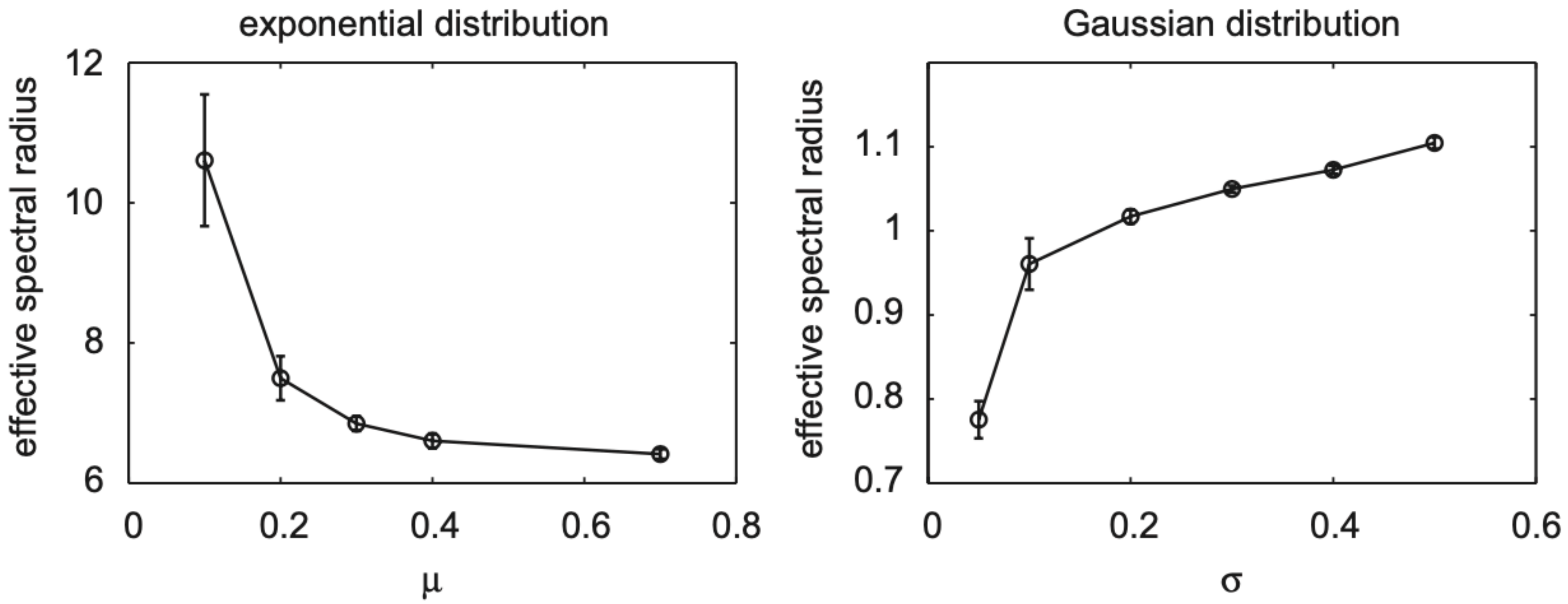
- Ring-topology, through the use of IP, can be used as a real reservoir.
- Consequences ➞ Implementing reservoirs in hardware an effective reservoir can be built using hardware friendly nearest neighbor connections.

Relationship between spectral radius and IP [1]

Benchmarks: Narma 30, Memory Capacity, Digit recognition form speech

Scale reservoir by ranging over spectral radius, and evaluate the use of IP for Fermi nodes with Exp. Distribution (ranging over mean) and tanH nodes with Gaussian distribution (mean = 0, ranging over variance).

- comparing optimal spectral radius and the effective spectral radius of the optimal s: where the s was optimal, the effective spectral radius was equal to the optimal spectral radius of a network without IP.
- small variance of the relation between moments and effective spectral radius. Imposing certain output distributions on the nodes of the reservoir is thus actually a precise way of controlling the dynamics in the reservoir.



Relation between the mean and standard deviation of the exponential and Gaussian distribution, respectively, and the effective spectral radius which is attained after pre-training the reservoir using IP

IP and spectral radius influence each other in two ways.

(1) Initial weight matrix size can alter the learning behavior of IP, because the adjustment factors of the intrinsic parameters by the rule depend on the amount of activity present in the network.

(2) Changing a of the transfer function corresponds to scaling all incoming weights, and therefore changing the spectral radius of the weight matrix. Thus, the effective spectral radius after applying IP will be different from the one the network was initialized to.

Synergies between Intrinsic Plasticity and Synaptic Plasticity [3]

Theoretical and experimental results for individual model neuron

What is the interaction between Hebbian Synaptic Plasticity (SP) with IP?

- Consider a sigmoidal model neuron $y = S_{ab}(x)$, with $x = w^T u$, the simple Hebbian rule with weights normalization will be:

$$\Delta w = \alpha u Y(u) = \alpha u S_{ab}(w^T u) \qquad w = w / \parallel w \parallel$$

- Assume, even if biological implausible, the limit of IP being faster than SP.
- S_{ab} has adapted to give an approximately exponential distribution of the firing rate Y before w can change much.

→ **Δw can be seen as a weighted sum of the pre-synaptic inputs u**, with the activities Y acting as weights that follow an approximately exponential distribution.

The combination of intrinsic with Hebbian plasticity can result in the neuron weight vector w correspond to a sparse/heavy-tailed direction in the input space.

Combination of IP and SP "Bars" problem:

- Input domain: $N \times N$ retina with horizontal and vertical bars Each bars show up independently with a probability $p = 1/N$. → Find the individual bars.
- Train a model neuron on input domain, setting IP to be slower than SP. Can the model still discover sparse directions in the input?
- w aligns with one of the individual bars as soon as IP pushed the neuron into a regime where its responses are sparse. → One of the individual bars has been discovered.

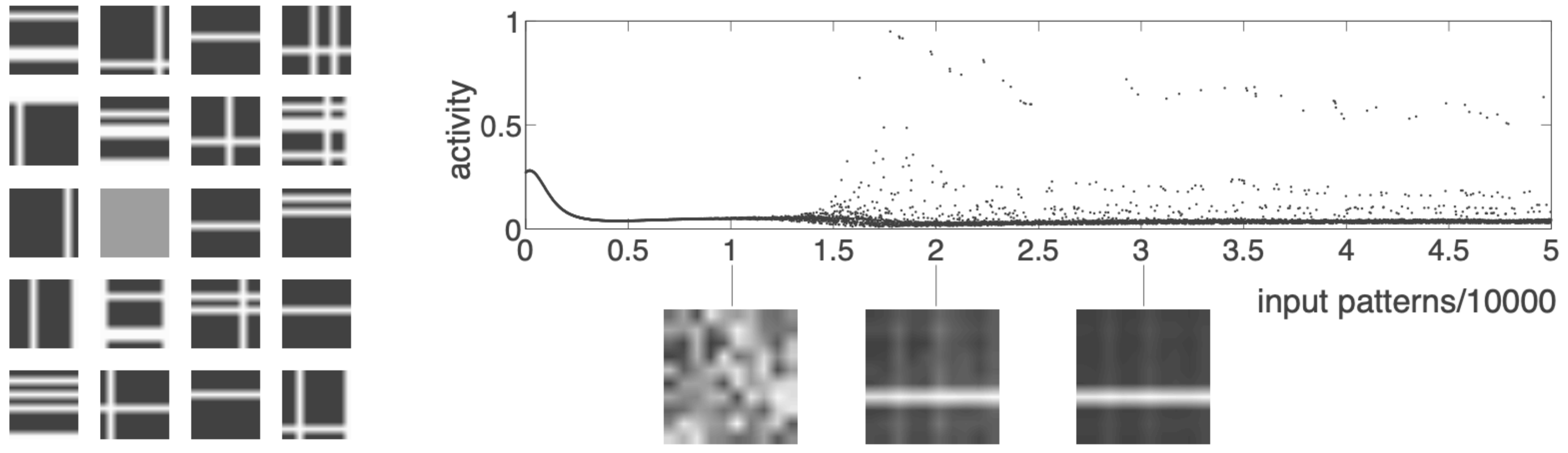
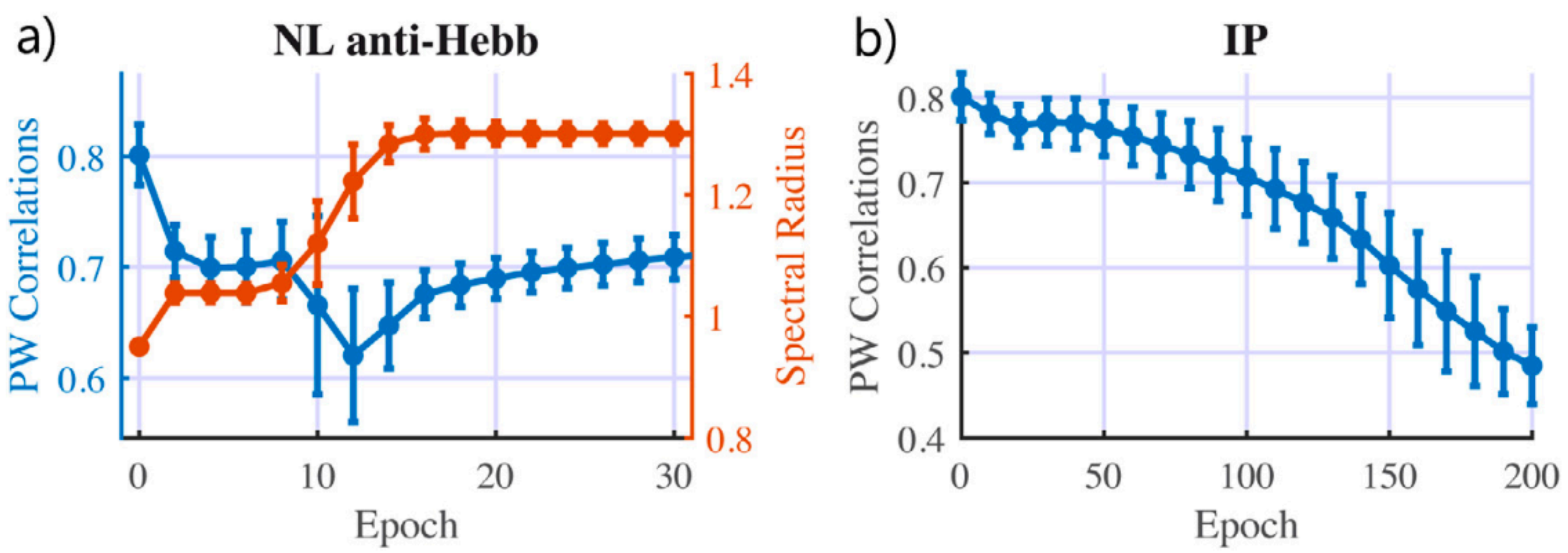


Figure 3: Left: example stimuli from the "bars" problem for a 10 by 10 pixel retina. Right: the activity record shows the unit's response to every 10th input pattern. Below, we show the learned weight vector after presentation of 10,000, 20,000, and 30,000 training patterns.

ESN: influence of plasticity rules on reservoir dynamics [4]

Synaptic and Intrinsic Plasticity: experimental results on MG-17 prediction task



At the network level, different quantities can be modified by the plasticity rules.

- For the non-local anti-Hebbian rule, the **sudden increase in the reservoir weight matrix spectral radius co-occurs with a sharp drop on the states correlations at consecutive time**.
- The optimal number of epochs occurred just before the transition to a periodic self-sustained dynamics inside the reservoir.
- **Continuous application of the IP rule also tends to decorrelate the states of the neurons within the reservoir**. An over-training of the IP rule eventually results in a loss of the consistency of the neural responses and the corresponding performance degradation.

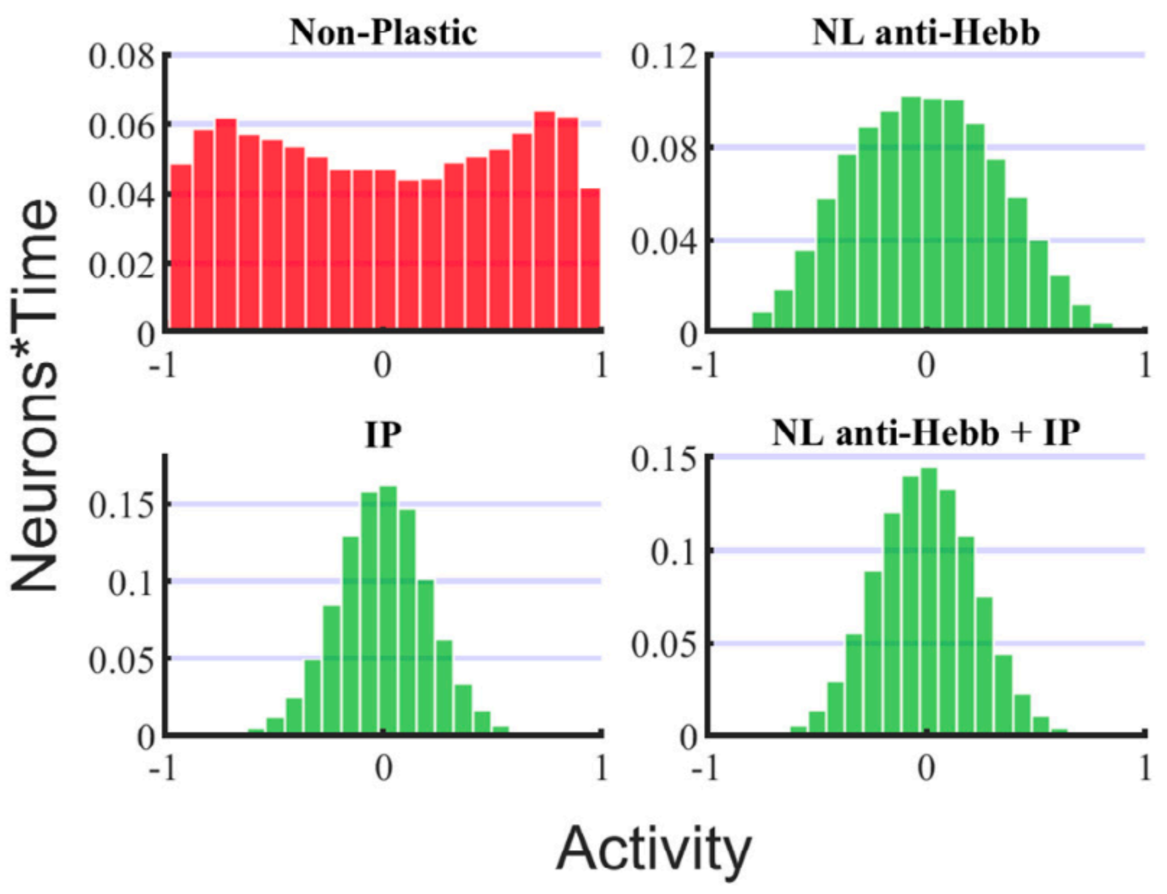


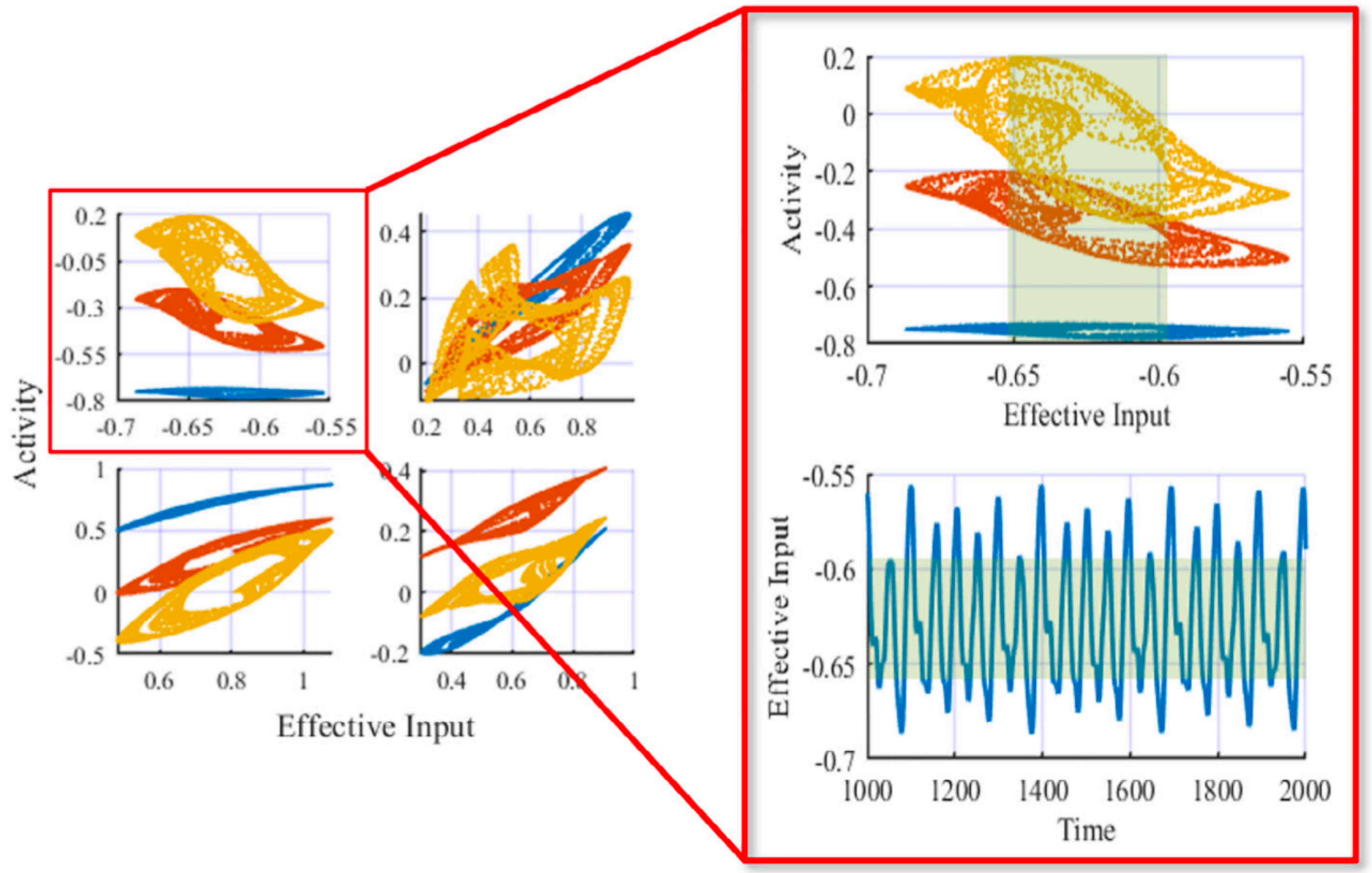
Fig. 5. Distribution of reservoir states after training for the explored ESN models. Histograms were constructed averaging over 20 different realizations of the corresponding ESNs.

Both types of plasticity lead to unimodal distributions of states centered around zero.

- Optimal performance is achieved—for this type of temporal tasks—when most neurons distribute along the hyperbolic tangent (activation function) by avoiding the saturation regions.
- Similar results observed both in vivo and in artificial single neuron models, suggest that this form of states distribution helps to achieve optimal encoding of the inputs.

ESN: influence of plasticity rules on single reservoir neuron dynamics [4]

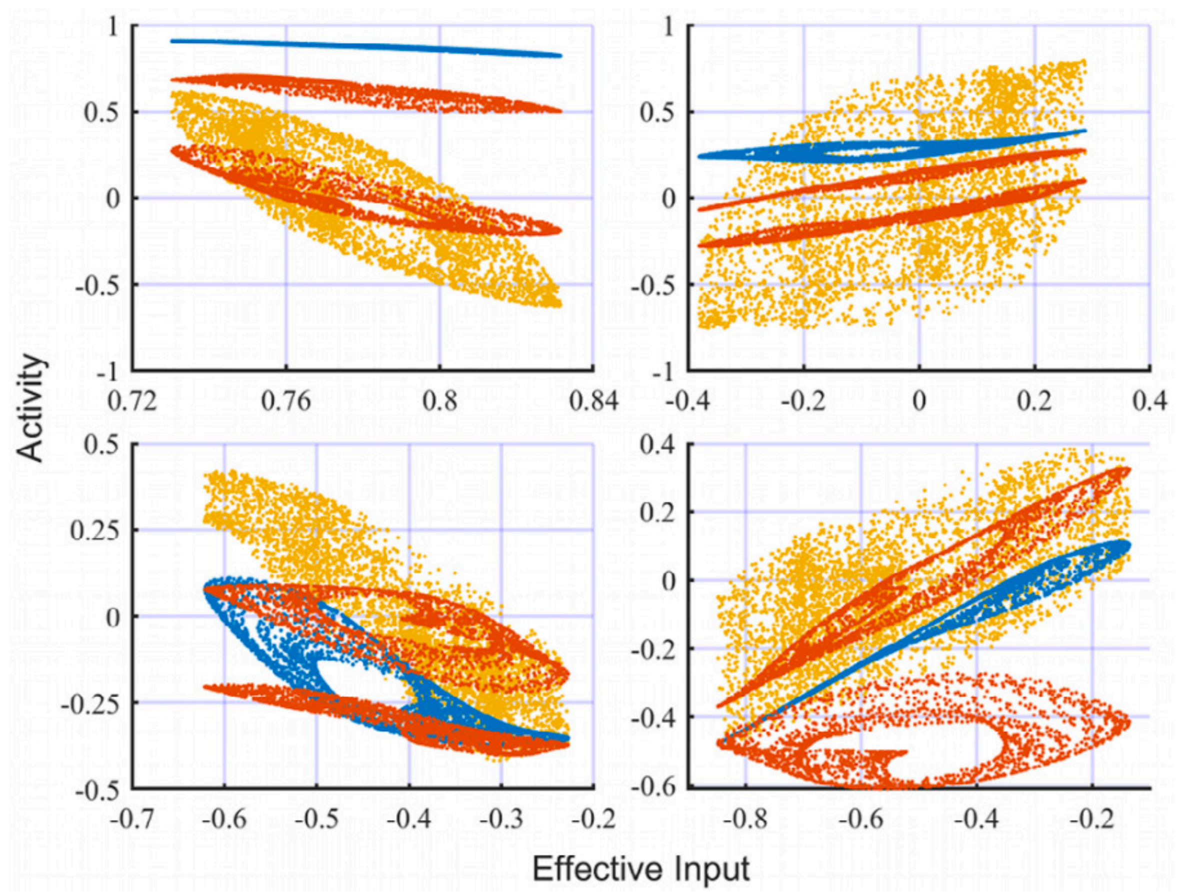
Experimental results on MG-17 prediction task



How individual neuron"reacts" to the input after implementation of the plastic rules?

Response of 4 neurons to the input before (blue dots) and after the implementation of the non-local anti-Hebbian (red) and IP rules (yellow) show that **plasticity rules expand each neuron activity space** (measured in terms of its area), adapting to the properties of the input and thus possibly enhancing the computational capability of the whole network.

- For the non-plastic case, when a neuron is mainly influenced by external input at each time-step the corresponding states are distributed in a narrow region around the activation function curve.
- A broadened phase space (founded after plasticity implementation) suggests a greater role of the interactions among past values of the reservoir units in determining the neuron state. When IP was implemented, a further displacement of their activity towards the center of the activation function is observed.



The regime of performance degradation found when over-training the plastic parameters is of different nature for the SP and IP rules.

- SP case (red): the phase space region occupied by the activity of each neuron splits in two disjoint regions, with the state jumping from one region to the other at consecutive time steps. Instability is associated with a self-sustained periodic dynamics.
- IP case (yellow): the de-correlation of the states and expansion of their phase space continues progressively, with different inputs eventually leading to similarly broad projections of the reservoir states in the activity phase space.

Synergies between SP and IP in ESN [5]

Synergistic plasticity rule and experimental results

- In neurobiology, the **interplay between SP and IP** contributes to the adaptation of the nervous system to different synaptic input signals. Most computational models consider these plasticity mechanisms separately ⇒ biologically implausible.
- A synergistic plasticity learning rule, proposed to adapt the reservoir connections in ESNs, **simultaneously consider the regulation of the synaptic weights using the SP rule and the adjustment of neuronal intrinsic excitability using the IP rule.**

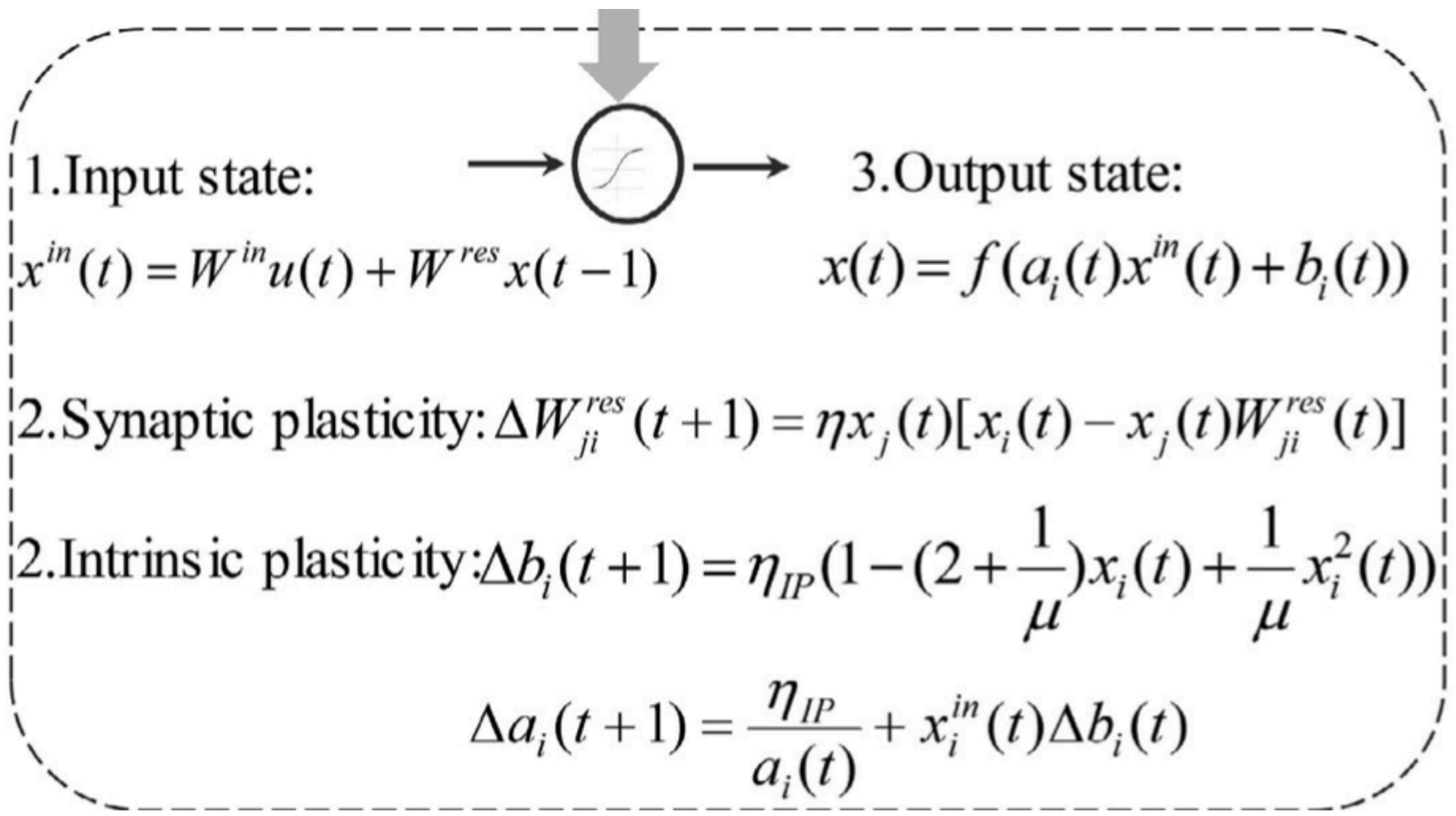
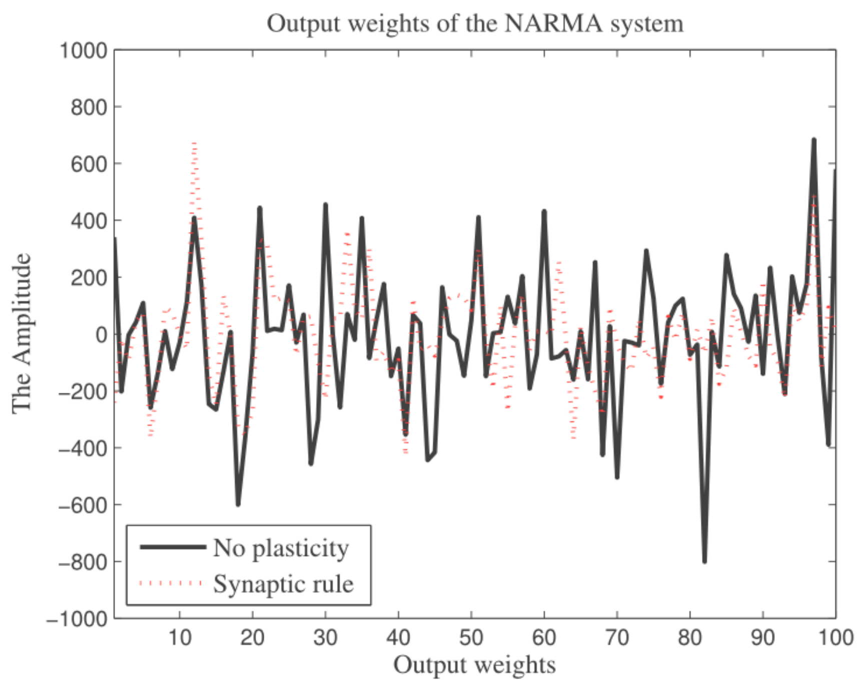


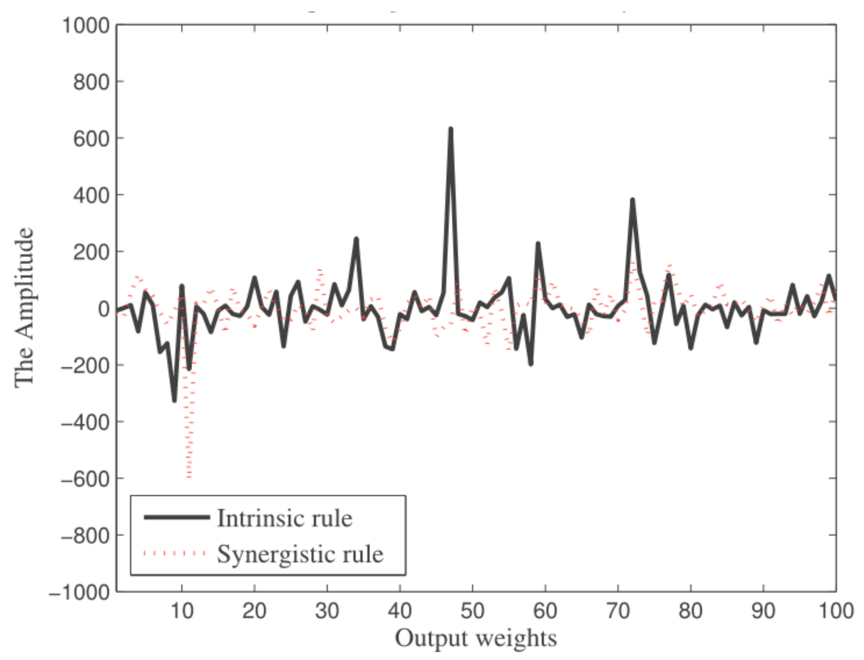
Fig. Regulation process of ESN with synergistic learning rule.

- IP and synergistic plasticity rule can substantially reduce the amplitude of the output weights (compare to SP and original ESN) , confirming that the IP can maintain the neuronal firing level in a desirable region and maximize information transmission. **small output weights are conductive to improving generalization capabilities of the reservoir.**
- The connection weights within the reservoir can be regulated by the SP rule (change of the spectral radius in Tab.), while IP rules does not affect the spectral radius of the reservoir.
- By comparing the learning performance of the ESN variants averaged over 30 independent trials, results demonstrate that **ESN with the synergistic plasticity learning rule performs the best**.

The comparison results of the NARMA system.

Approach	Testing NRMSE (Mean)	Testing NRMSE (Std.)	Number of neurons	Spectral radius
Synergistic rule	0.2463	2.5596e−03	100	0.9324
Synaptic rule	0.4352	3.4150e−03	100	0.9081
Intrinsic rule	0.3004	4.5924e−03	100	0.8000
GESN	0.8289	3.4566e−04	100	0.8000
PESN	0.5957	1.6702e−03	100	0.8000
LESN	0.6377	1.2991e−03	100	0.8000
GRU	0.5530	2.4498e−03	100	0.8000
ESN	0.8306	7.8666e−04	100	0.8000





Conclusions and frontier in neuromorphic computing

Intrinsic Plasticity in ESN: advantages and future research areas

- **Idea of autonomously self-regulated robustness of reservoir dynamics.**
  The IP rule allows the reservoirs to autonomously perceive and adapt their dynamics to a specific regime that leads to good performance for a given task, irrespective of disturbances, initial weights or input scaling.
- **Extend the idea of information maximization to different output distributions**, more parameters that can be trained and other node types.
  i.e., LSM-type reservoirs are currently difficult to tune using IP can be controlled with a single parameter.
- **Deeper understanding of the underlying similarities between SP and IP rules** could drive new research avenues.
- **Sytematic combination of multiple form of neural plasticity** (observed in biological neurons) may improve the learning performance of the ESN.
  i.e, neuromodulation and some plasticity-related proteins for tuning the weights of the reservoir connections).

Recent development in Neuromorphic computing: Simultaneous emulation of IP and SP using memrestive synapse. [6]

- Memristive neurosynaptic device mimics SP and IP concomitantly in a single cell.
- Volatile threshold switch (TS) layers and non-volatile phase change memory (PCM) layers are merged in TS-PCM device.
- IP is introduce through a bottom TS layer, which resembles the modulation of firing frequency in biological neuron.
- SP is also introduced through the nonvolatile switching of top PCM layer.
- IP and SP are simultaneously emulated in a single cell to establish the positive feedback between them. A positive feedback learning loop which mimics the retraining process in biological system is implemented in TS-PCM array for accelerated training.
- TS-PCM provides a concomitant solution for the hardware realization of an artificial synapse with synergistic interactions between IP and SP, presenting high similarity with learning mechanism of biological system.

Thank you for the attention!

References

[1] Schrauwen, Benjamin, et al. *Improving reservoirs using intrinsic plasticity.* Neurocomputing 71.7-9 (2008): 1159-1171.

[2] Triesch, Jochen. *A gradient rule for the plasticity of a neuron's intrinsic excitability.* International Conference on Artificial Neural Networks. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005.

[3] Triesch, Jochen. *Synergies between intrinsic and synaptic plasticity in individual model neurons.* Advances in neural information processing systems 17 (2004).

[4] Morales, Mirasso, and Soriano. *Unveiling the role of plasticity rules in reservoir computing.* Neurocomputing 461 (2021): 705-715.

[5] Wang, Jin hao. *Synergies between synaptic and intrinsic plasticity in echo state networks.* Neurocomputing 432 (2021): 32-34.

[6] Sung, Sang Hyun, et al. *Simultaneous emulation of synaptic and intrinsic plasticity using a memristive synapse.* Nature Communications 13.1 (2022): 2811.