Homeostatic Plasticity for Single Node Delay-Coupled Reservoir Computing

Hazem Toutounji
hazem.toutounji@zi-mannheim.de
Neuroinformatics Department, Institute of Cognitive Science, University
of Osnabrück, 49069 Osnabrück, Germany, and Department of Theoretical
Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim
of Heidelberg University, 68159 Mannheim, Germany

Johannes Schumacher
joschuma@uni-osnabrueck.de
Gordon Pipa
gpipa@uni-osnabrueck.de
Neuroinformatics Department, Institute of Cognitive Science, University
of Osnabrück, 49069 Osnabrück, Germany

Supplementing a differential equation with delays results in an infinite-dimensional dynamical system. This property provides the basis for a reservoir computing architecture, where the recurrent neural network is replaced by a single nonlinear node, delay-coupled to itself. Instead of the spatial topology of a network, subunits in the delay-coupled reservoir are multiplexed in time along one delay span of the system. The computational power of the reservoir is contingent on this temporal multiplexing. Here, we learn optimal temporal multiplexing by means of a biologically inspired homeostatic plasticity mechanism. Plasticity acts locally and changes the distances between the subunits along the delay, depending on how responsive these subunits are to the input. After analytically deriving the learning mechanism, we illustrate its role in improving the reservoir's computational power. To this end, we investigate, first, the increase of the reservoir's memory capacity. Second, we predict a NARMA-10 time series, showing that plasticity reduces the normalized root-mean-square error by more than 20%. Third, we discuss plasticity's influence on the reservoir's input-information capacity, the coupling strength between subunits, and the distribution of the readout coefficients.

1 Introduction

Reservoir computing (RC) (Jaeger, 2001; Maass, Natschläger, & Markram, 2002; Buonomano & Maass, 2009; Lukoševičius & Jaeger, 2009) is a

computational paradigm that provides both a model for neural information processing (Häusler & Maass, 2007; Karmarkar & Buonomano, 2007; Yamazaki & Tanaka, 2007; Nikolić, Häusler, Singer, & Maass, 2009) and powerful tools for carrying out a variety of spatiotemporal computations. This includes time series forecasting (Jaeger & Haas, 2004), signal generation (Jaeger, Lukoševičius, Popovici, & Siewert, 2007), pattern recognition (Verstraeten, Schrauwen, & Stroobandt, 2006), and information storage (Pascanu & Jaeger, 2011). RC also affords a framework for advancing and refining our understanding of neuronal plasticity and self-organization in recurrent neural networks (Lazar, Pipa, & Triesch, 2007, 2009; Toutounji & Pipa, 2014).

This article presents a biologically inspired neuronal plasticity rule to boost the computational power of a novel RC architecture that is called a single node delay-coupled reservoir (DCR). The DCR realizes the same RC concepts using a single nonlinear node with delayed feedback (Appeltant et al., 2011). This simplicity makes the DCR particularly appealing for physical implementations, which has already been demonstrated on electronic (Appeltant et al., 2011), optoelectronic (Larger et al., 2012; Paquot et al., 2012), and all-optical hardware (Brunner, Soriano, Mirasso, & Fischer, 2013). The optoelectronic and all-optical implementations use a semiconductor laser diode as the nonlinear node and an optical fiber as a delay line, allowing them to maintain high sampling rates. They are also shown to compare in performance to standard RC architectures in benchmark computational tasks.

The DCR operates as follows. Different nonlinear transformations and mixing of stimuli from the past and the present are achieved by sampling the DCR's activity at virtual nodes (v-nodes), along the delay line. While neurons of a recurrent network are mixing stimuli via their synaptic coupling, which forms a network topology, the v-nodes of a DCR are mixing signals via their (nonlinear) temporal interdependence. Therefore, the v-nodes' temporal distances from one another, henceforth termed v-delays, are made shorter than the characteristic timescale of the nonlinear node. Thus, v-nodes become analogous to the connections of a recurrent network, providing the DCR with a certain network-like topology. In analogy to the spatial distribution of input in a classical reservoir, stimuli in a DCR are temporally multiplexed (see Figure 1). To process information, the external stimuli are applied to the dynamical system, thereby perturbing the reservoir dynamics. Here, we operate the DCR in an asymptotically stable fixed point regime. To render the response of the DCR transient (i.e., reflecting nonlinear combinations of past and present inputs), the reservoir dynamics must not converge to the fixed point, where it becomes dominated by the current stimulus. To ensure this, a random piecewise constant masking sequence is applied to the stimulus before injecting the latter to the reservoir (Appeltant et al., 2011). The positions where this mask may switch value match the positions of the v-nodes, which are initially chosen
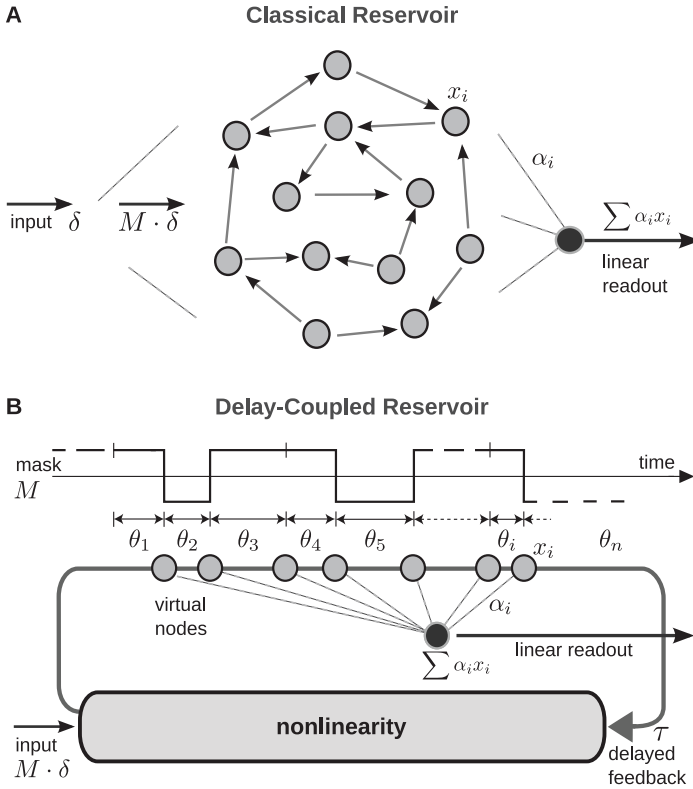
A   **Classical Reservoir**

x_i

α_i

input δ $M \cdot \delta$

$\sum \alpha_i x_i$

linear
readout

B   **Delay-Coupled Reservoir**

mask   time
M

θ_1 θ_2 θ_3 θ_4 θ_5 θ_i x_i θ_n

virtual
nodes α_i

linear readout

$\sum \alpha_i x_i$

nonlinearity

input τ
$M \cdot \delta$ delayed
                                                 feedback

Figure 1: Comparing classical and single node delay-coupled reservoir computing architectures. (A) A classical RC architecture. The input δ is spatially distributed by input weights M to a RNN of size n. The activity of the RNN is then linearly read out. (B) A single node delay-coupled reservoir. The input δ is temporally multiplexed across a delay line of length τ by using a random binary mask M of n bits. Each mask bit M_i is held constant for a short v-delay θ_i, such that the sum of these delays is the length of the delay line τ. The masked input is then nonlinearly transformed and mixed with past input by a nonlinear node with delayed feedback. At the end of each v-delay θ_i, there resides a v-node from which linear readouts learn to extract information and perform spatiotemporal computations through linear regression.

equidistant. However, given the fact that the v-delays directly influence the interdependence of the corresponding v-nodes states, and therefore the nonlinear mixing of the stimuli, it is immediately evident that v-delays are important parameters that may significantly influence the performance of the DCR.

To optimize the computational properties of the DCR, we employ neuroscientific principles using biologically inspired homeostatic plasticity (Davis & Goodman, 1998; Zhang & Linden, 2003; Turrigiano & Nelson, 2004) for adjusting the v-delays. Biologically speaking, homeostatic plasticity does not refer to a single particular process. It is, rather, a generic term for a family of adaptation mechanisms that regulate different components of the neural machinery, bringing these components to a functionally desirable operating regime. The choice of the operating regime depends on the functionality that a model of homeostatic plasticity aims to achieve. This resulted in many flavors of homeostatic plasticity for regulating recurrent neural networks in computational neuroscience (Somers, Nelson, & Sur, 1995; Soto-Treviño, Thoroughman, Marder, & Abbott, 2001; Renart, Song, & Wang, 2003; Lazar et al., 2007, 2009; Marković and Gros, 2012; Remme & Wadman, 2012; Naudé, Cessac, Berry, & Delord, 2013; Zheng, Dimitrakakis, & Triesch, 2013; Toutounji & Pipa, 2014), neurorobotics (Williams & Noble, 2007; Vargas, Moioli, Von Zuben, & Husbands, 2009; Hoinville, Siles, & Hénaff, 2011; Dasgupta, Wörgötter, & Manoonpong, 2013; Toutounji & Pasemann, 2014), and reservoir computing (Schrauwen, Wardermann, Verstraeten, Steil, & Stroobandt, 2008; Dasgupta, Wörgötter, & Manoonpong, 2013). Here, we use a homeostatic plasticity mechanism to regulate the v-delays so as to balance responsiveness to the input and its history on the one hand, against optimal expansion of its informational features into the high-dimensional phase space of the system, on the other hand. Furthermore, we show that this process can be understood as a competition between the v-nodes' sensitivity and their entropy, resulting in a functional specialization of the v-nodes. This leads to a high increase in the DCR's memory capacity and a significant improvement in its ability to carry out nonlinear spatiotemporal computations. We discuss the implications of the plasticity mechanism with respect to the DCR's entropy, as well as the virtual network topology, and the resulting regression coefficients.

2 Model

In this section, we describe the RC architecture that is based on a single nonlinear node with delayed feedback. We then formulate this architecture using concepts from neural networks.

2.1 Single Node Delay-Coupled Reservoir. Generally, RC comprises a set of models where a large dynamical system called a reservoir (e.g., a recurrent neural network) nonlinearly maps a set of varying stimuli to a high-dimensional space (Jaeger, 2001; Maass et al., 2002). The recurrency allows a damped trace of the stimuli to travel within the reservoir for a certain period of time. This phenomenon is termed *fading memory* (Boyd & Chua, 1985). Then, random nonlinear motifs within the reservoir nonlinearly mix

past and present inputs, allowing a desired output to be linearly combined from the activity of the reservoir, using a linear regression operation. As the desired output is usually a particular transformation of the temporal and spatial aspects of the stimuli, the operations that a RC architecture is trained to carry out are termed *spatiotemporal computations*.

In a classical RC architecture, past and present inputs $\delta \in \mathbb{R}^m$ undergo nonlinear mixing via injection into a recurrent neural network (RNN) of n nonlinear units. This spatial distribution of the input is a mapping $f : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$. The dynamics are modeled by a difference equation for discrete time,

$$x(t + 1) = f(x(t), \delta(t)), \tag{2.1}$$

or an ordinary differential equation (ODE) for continuous time,

$$\dot{x}(t) = f(x(t), \delta(t)), \tag{2.2}$$

where $x(t) \in \mathbb{R}^n$ is the network activity and $\dot{x}(t)$ the activity's time derivative.

In a single node delay-coupled reservoir (DCR), the recurrent neural network is replaced by a single nonlinear node with delayed feedback. Formally, the dynamics can be modeled by a forced (or driven) delay differential equation (DDE) of the form

$$\dot{x}(t) = -x(t) + f(x(t - \tau), \delta(t)) \tag{2.3}$$

where τ is the delay time and $x(t), x(t - \tau) \in \mathbb{R}$ are the current and delayed DCR activities. Figure 1 illustrates the DCR architecture and compares it to the standard RNN approach to reservoir computing.

Solving system 2.3 for $t \geq 0$ requires specifying an appropriate initial value function $\phi_0 : [-\tau, 0] \to \mathbb{R}$. As is already suggested by the initial conditions, the phase space of system 2.3 is a Banach space $C_{1,\tau} = C([-\tau, 0], \mathbb{R}) \ni \phi_0$ which is infinite dimensional (Guo & Wu, 2013). Using a DDE as a reservoir, this phase space thus provides a high-dimensional feature expansion for the input signal, which is usually achieved by using an RNN with more neurons than input channels.

To inject a signal into the reservoir, it is multiplexed in time. The DCR receives a single constant input $u(\bar{t}) \in \mathbb{R}^m$ in each reservoir time step $\bar{t} = \lceil \frac{t}{\tau} \rceil$, corresponding to one τ-cycle of the system. During each τ-cycle, the input is again linearly transformed by a mask $M \in [0, \tau]^m$ that is piecewise constant for short periods θ_i, representing the temporal spacing, or v-delays, between sampling points of $i = 1, \ldots, n$ virtual nodes, or v-nodes, along the delay line. Accordingly, the v-delays satisfy $\sum_{i=1}^{n} \theta_i = \tau$, where n is

the effective dimensionality of the DCR. Here, the mask M is chosen to be binary with random mask bits $M_i \in \{-\mu, +\mu\}^m$, so that the v-node i receives a weighted input $M_i u(\bar{t})$. In order to ensure that the DCR possesses fading memory of the input, system 2.3 is set to operate in a regime governed by a single fixed point in case the input is constant. Thus, the masking procedure effectively prevents the driven dynamics of the underlying system from saturating to the fixed point.

A sample is read out at the end of each θ_i, yielding n predictor variables (v-nodes) $x_i(\bar{t})$ per time step $\bar{t}$. Computations are performed on the predictors using a linear regression model for some scalar target time series y, given by $\hat{y}(\bar{t}) = \sum_{i=1}^{n} \alpha_i x_i(\bar{t})$ where x_i with $i = 1, \ldots, n$ denote the DCR's v-nodes (see equation 2.6) and α_i are the coefficients determined by regression, for example, using the least squares solution minimizing the sum of squared errors $\sum_{\bar{t}} (y(\bar{t}) - \hat{y}(\bar{t}))^2$.

In what follows, our model of choice for the DCR nonlinearity is an input-driven Mackey-Glass system (Glass & Mackey, 2010) that is operating, when not driven by input, at a fixed point regime,

$$\dot{x}(t) = -x(t) + \frac{\eta \big(x(t - \tau) + \gamma M \delta(t) \big)}{1 + \big(x(t - \tau) + \gamma M \delta(t) \big)}, \tag{2.4}$$

where γ and η are model parameters. In addition to favorable analytical properties that are to be stated in turn, the current choice of nonlinearity is motivated by the superior performance it achieves on spatiotemporal computations. It can also be approximated by electronic circuits (Appellant et al., 2011). Figure 2A shows the response of the DCR governed by equation 2.4 to a single-channel input.

2.2 The DCR as a Virtual Network. The goal is to optimize the computational properties of the DCR as a network given a vector of v-delays $\Theta = (\theta_1, \ldots, \theta_i, \ldots, \theta_n)$ of its n v-nodes. In the case of equidistant v-delays, approximate v-node equations were already derived by Appellant et al. (2011), who also conceptualized the DCR with equidistant v-delays as a network. We extend this result to account for arbitrary v-node spacings on which our plasticity rule can operate. To that end, we first need to define the activity $x(t)$ of the DCR given θ_i for $i = 1, \ldots, n$.

First, we solve the DDE, equation 2.3 by applying the method of steps (see appendix A for details on solving and simulating the DCR). If system 2.3 is evaluated at $(\nu - 1)\tau \leq t \leq \nu\tau$, where a continuous function $\phi_\nu \in C_{[(\nu-2)\tau, (\nu-1)\tau]}$ is the solution for $x(t)$ on the previous τ-interval, we can replace $x(t - \tau)$ by $\phi_\nu(t - \tau)$. Consequently, the solution to equation 2.3
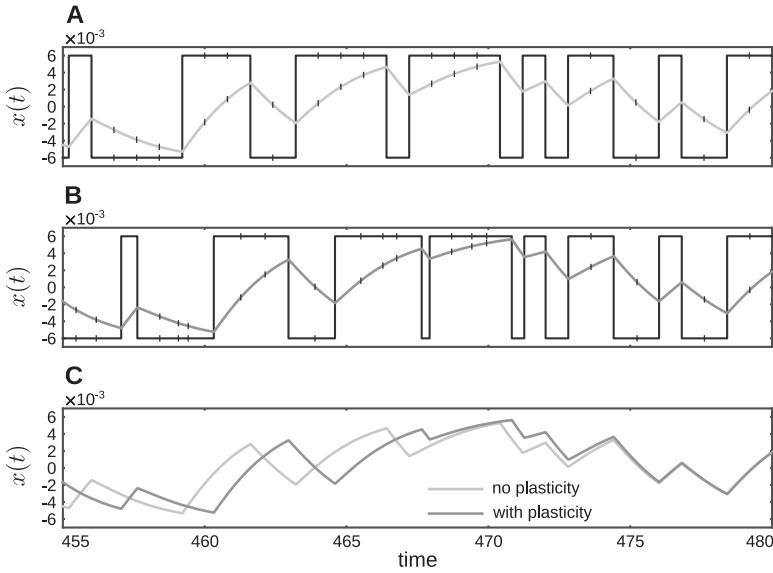
Figure 2: DCR activity superimposed on the corresponding mask (A) before and (B) after plasticity. (C) Comparison between the DCR's activity before and after plasticity.

subject to $x((v-1)\tau) = \phi_v((v-1)\tau)$ is given by

$$x(t) = \phi_v((v-1)\tau)e^{(v-1)\tau-t} + e^{(v-1)\tau-t}\int_{(v-2)\tau}^{t-\tau} f(\phi_v(s), \delta(s))e^{s-(v-2)\tau}ds.$$

$$(2.5)$$

Let the DCR activity at a particular v-node $x_i(\bar{t}) = x((v-1)\tau + \sum_{j=1}^{i}\theta_j)$, its nonlinearity $f_i(\bar{t}) = f(x_i(\bar{t}-1), M_i \cdot u(\bar{t}))$, and the DCR time step $\bar{t} = \lceil\frac{t}{\tau}\rceil = v$. As shown in appendix A, the solution mapping equation 2.5 to the DCR can be approximated by assuming $f(\cdot)$ to be piecewise constant at each θ_i. This is a valid approximation since $\theta_i \ll \tau$, and it yields the following expression of the DCR activity at a v-node i as a function of $\{\theta_1, \ldots, \theta_i\}$:

$$x_i(\bar{t}) = e^{-\sum_{j=1}^{i}\theta_j}x_n(\bar{t}-1) + \sum_{j=1}^{i}(1-e^{-\theta_j})e^{-\sum_{k=j+1}^{i}\theta_k} \cdot f_j(\bar{t}). \qquad (2.6)$$

Equation 2.6 suggests that the activity of v-node i is a weighted sum of the nonlinear component of the preceding v-nodes' activity, down to the last v-node n in the cyclic network, the activity of which is carried over from
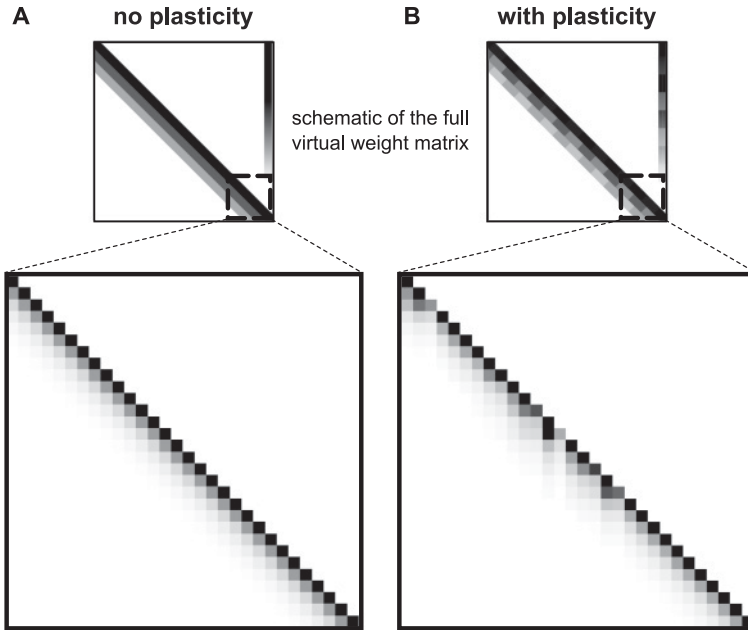
Figure 3: Virtual weight matrix of a DCR (A) before and (B) after plasticity. The magnified section corresponds roughly to connectivity within part of the delay span τ shown in Figure 2.

the previous reservoir time step. The resulting directed network topology is shown as a virtual weight matrix for equidistant v-nodes ($\theta_i = \tau/n$) in Figure 3A.

3 Plasticity

An important role of the randomly alternating mask M is to prevent the DCR dynamics from saturating, and thus losing history dependence and sensitivity to input. However, the random choice of the mask values and the equal v-delays do not guarantee an optimal choice of masking. A simple example that already illustrates this point is given by the occurrence of sequences of equal valued mask bits, as shown in Figure 2A, which leads to unwanted saturation. In general, many more factors exist that determine optimal computation in the reservoir and need balancing.

Our goal in this section therefore is to develop a plasticity mechanism that optimizes the resulting v-delays with respect to sensitivity, while retaining a suitable nonlinear feature expansion into the DCR's phase space. As we show in section 5.1, this results in a trade-off between sensitivity

and entropy of the v-nodes. Entropy and sensitivity counteract each other, thus forcing v-nodes to specialize. In a first step set out in section 3.1, we develop a partial plasticity mechanism that maximizes solely the sensitivity of individual v-nodes. In a second step in section 3.2, the mechanism is augmented by a counteracting regulatory term that tries to retain diverse feature expansion of the input. The delay τ, together with the number n of v-nodes, the mask M, and the parameters γ and η of the delayed nonlinearity, are given hyperparameters that are kept constant, and they determine the particular DCR that is subjected to the optimization process.

3.1 Sensitivity Maximization. We measure a v-node's sensitivity by the slope of its activity at the readout point (i.e., the end point of the θ_i interval), where a bigger slope corresponds to less saturation. The objective is to maximize the overall sensitivity of the DCR for all v-nodes simultaneously. First, we use the approximate solution mapping of a v-node's dynamics from equation 2.6 to derive a formula of a v-node's activity as a function of the v-delay θ_i from the previous v-node alone:

$$x_i(\bar{t}) = e^{-\theta_i} x_{i-1}(\bar{t}) + (1 - e^{-\theta_i}) f_i(\bar{t}), \qquad i = 2, \ldots, n, \tag{3.1}$$

$$x_1(\bar{t}) = e^{-\theta_1} x_n(\bar{t} - 1) + (1 - e^{-\theta_i}) f_n(\bar{t} - 1). \tag{3.2}$$

In addition, the dynamics of the DCR at a particular v-node i in units of reservoir time steps $\bar{t}$ is given by

$$\dot{x}_i(\bar{t}) = -x_i(\bar{t}) + f_i(\bar{t}). \tag{3.3}$$

Substituting equation 3.1 into 3.3 yields the following expression for the sensitivity of a v-node i as a function of θ_i:

$$S_i(\bar{t}) = \dot{x}_i(\bar{t}) = (-x_{i-1}(\bar{t}) + f_i(\bar{t}))e^{-\theta_i}. \tag{3.4}$$

From equation 3.4, we define a sensitivity vector $\mathbf{S} \in \mathbb{R}^n$. To optimize the overall sensitivity of the DCR, we maximize an objective function under the constraint that the sum of the v-delays stays equal to the overall delay τ,

$$\arg\max_{\Theta \geq 0}\{\|\mathbf{S}\|_2^2\} \quad \text{subject to} \quad \sum_{i=1}^{n} \theta_i = \tau. \tag{3.5}$$

where $\| \cdot \|_2$ is the Euclidean norm.

To find the vector Θ that solves the constrained optimization problem 3.5, we follow the direction of the steepest ascent, which is the gradient of the

objective function, and we project the outcome to the simplex $\sum_{i=1}^{n} \theta_i = \tau$. The element-wise gradient is given by

$$\nabla_i \|\mathbf{S}\|_2^2 = \frac{\partial \|\mathbf{S}\|_2^2}{\partial \theta_i}. \tag{3.6}$$

By iteratively inserting expression 3.1 into the sensitivity formula 3.4 and eliminating the iteration with equation 3.2, we can show that the sensitivity of a v-node i depends on the v-delays θ_j of all the preceding v-nodes $j \le i$,

$$S_i(\bar{t}) = e^{-\sum_{k=j+1}^{i} \theta_k} \cdot S_j(\bar{t}) + \Gamma(\theta_{j+1}, \cdots, \theta_i), \tag{3.7}$$

where $\Gamma(\cdot)$ is a term independent of θ_j. However, since the term $e^{-\sum_{k=j+1}^{i} \theta_k}$ decays exponentially the farther the v-node i is from the v-node j, one can ignore the contribution of θ_j to the sensitivity of the v-node i for $i > j$. This simplifies the element-wise gradient to

$$\nabla_i \|\mathbf{S}\|_2^2 = \frac{\partial S_i^2}{\partial \theta_i}$$
$$= -2(-x_{i-1}(\bar{t}) + f_i(\bar{t}))^2 e^{-2\theta_i}. \tag{3.8}$$

3.2 Homeostatic Plasticity. The optimization problem 3.5 maximizes the sensitivity of a v-node i by decreasing θ_i, its temporal distance from the previous v-node, as is suggested by the element-wise gradient, equation 3.8. As a result, the v-node becomes more sensitive to the input history delivered from its predecessor. This, however, leads to a loss of diversity in expanding informational features of the input, since the smaller the time allotted to a v-node is, the less excitable by input it becomes. In addition, the optimization objective prefers small θ_i, many of which may even go to 0, despite the constraint $\sum_{i=1}^{n} \theta_i = \tau$, which leads to a reduction of the reservoir's effective dimensionality.

We hypothesize that good spatiotemporal computational performance is achieved when diversity and sensitivity are balanced. To this end, we introduce a regulatory term into the sensitivity measure that punishes small v-delays, thus counteracting sensitivity by enforcing an increase in a v-node's distance from its predecessor. The choice of the regulatory term is motivated by favorable analytical properties (mentioned later in this section) and by allowing flexibility in the choice of regulation between diversity and sensitivity. As entropy is a natural measure of informational diversity, we later support the current intuitions behind our choice of the regulatory term by a rigorous mathematical argumentation. Namely, we show in section 5.1 how a plasticity mechanism that solely maximizes entropy of the

v-nodes leads to an unbounded increase of v-delays and therefore presents a proper counterforce to sensitivity.

The sensitivity measure with regulatory term has the form

$$S_{i_\rho}(\bar{t}) = \theta_i^\rho \cdot \dot{x}_i(\bar{t}) = \theta_i^\rho(-x_{i-1}(\bar{t}) + f_i(\bar{t}))e^{-\theta_i}, \tag{3.9}$$

where $\rho > 0$ is a regulating parameter that modulates the penalty afflicted on the decrease in θ_i. Lower ρ leads the objective to favor smaller v-delays and vice versa.

From equation 3.9, we define a homeostatic sensitivity vector $\mathbf{S}_\rho \in \mathbb{R}^n$ and an optimization problem,

$$\arg\max_{\Theta \geq 0}\{\mathcal{O}(\Theta) = \|\mathbf{S}_\rho\|_2^2\} \quad \text{subject to} \quad \sum_{i=1}^n \theta_i = \tau, \tag{3.10}$$

and we maximize $\mathcal{O}$ by following the direction of the steepest ascent. Since the contribution of θ_j to the sensitivity of a v-node i for $i > j$ is ignorable, following the argumentation from equation 3.7, the element-wise gradient is simplified to

$$\nabla_i \mathcal{O}(\Theta) = \frac{\partial S_{i_\rho}^2}{\partial \theta_i}$$

$$= -2\theta_i^{2\rho-1}(\theta_i - \rho)(-x_{i-1}(\bar{t}) + f_i(\bar{t}))^2 e^{-2\theta_i}. \tag{3.11}$$

Defining a v-node's scaling factor $\sigma_i(\bar{t}) = (-x_{i-1}(\bar{t}) + f_i(\bar{t}))^2 \geq 0$, the maximized function $S_{i_\rho}^2 = \sigma_i\theta_i^{2\rho}e^{-2\theta_i}$ is unimodal, which entails the existence of a global maximum $\theta_i = \rho$, despite $S_{i_\rho}^2$ not being convex (For nonnegative v-delays, $S_{i_\rho}^2$ has one inflection point when $\rho \leq 0.5$ and two inflection points otherwise). This ensures convergence to the global maximum of the unconstrained optimization problem. The homeostatic plasticity learning rule for a single v-node i then reads

$$\theta_i(\bar{t} + 1) = \theta_i(\bar{t}) - 2\alpha\sigma_i(\bar{t})(\theta_i(\bar{t}) - \rho)\theta_i^{2\rho-1}(\bar{t})e^{-2\theta_i(\bar{t})}, \tag{3.12}$$

where the term $(\theta_i(\bar{t}) - \rho)$ homeostatically balances between the v-delay's increase and decrease, depending on the choice of the regulating term ρ.

Given the above, the update rule of the vector $\Theta = (\theta_1, \cdots, \theta_n)$ is given by

$$\Theta \leftarrow \pi_V(\Theta + \alpha \cdot \mathcal{J}_{\mathcal{O}}(\Theta)), \tag{3.13}$$

where α is a scalar learning rate, $\mathcal{J}_{\mathcal{O}} = \nabla \mathcal{O}(\Theta)$ is the Jacobian matrix of $\mathcal{O}$ with respect to Θ, and π_V is an orthogonal projection that ensures that Θ remains on the constraint simplex V defined by $\sum_{i=1}^{n} \theta_i = \tau$ (see appendix B for details of the constraint satisfaction). The global maximum belongs to V only when $\rho = \tau/n$, which leads to the convergence to equidistant v-nodes. Otherwise, the constrained gradient leads to the point on V closest to the global maximum.

4 Computational Performance

We next test the effect of the homeostatic plasticity mechanism, equation 3.12, on the performance of the DCR. Simulations are carried on 100 DCRs; the activity of each is sampled at 600 v-nodes that are initially equidistant with $\theta = 0.8$. Each DCR is completely distinguishable from the other by its binary mask M, and the 600 mask values are randomly chosen from the set $\{-0.1, +0.1\}$. Simulation starts with a short initial period for stabilizing the dynamics, followed by a plasticity phase of $n_p = 500$ time steps, each corresponding to one τ. The learning rate α is set to 0.01 and the regulating parameter ρ to 1.0. Afterward, readouts are trained on $n_t = 5000$ samples for both the original and modified v-delays θ_i and validated on another $n_v = 1000$ samples. The model parameters of the Mackey-Glass nonlinearity (see equation 2.4) are set to $\gamma = 0.05$ and $\eta = 0.4$. The DCR is subject to uniformly distributed scalar input $u(\bar{t}) \sim \mathcal{U}_{[0,0.5]}$. At this positive input range, the DCR dynamics resulting from the Mackey-Glass nonlinearity is saturating, as illustrated in Figure 2A. This condition ensures that the approximation, equation 2.6, is accurate enough that a decrease in a v-delay does increase a v-node's sensitivity.

Given a task-dependent target time series y and a linear regression estimate $\hat{y}(\bar{t}) = \sum_{i=1}^{n} \alpha_i x_i(\bar{t})$ (x_i being the DCR's v-nodes response to the input u), we measure the performance using the normalized root-mean-square error (nrmse):

$$\text{nrmse}(y, \hat{y}) = \sqrt{\frac{\sum_{n_v} (y - \hat{y})^2}{n_v \text{var}(y)}}. \tag{4.1}$$

4.1 Memory Capacity. The memory capacity of a reservoir is a measure of its ability to retain, in its activity, a trace of its input history. Optimal linear classifiers are trained for reconstructing the uniformly distributed scalar input $u(\bar{t}) \sim \mathcal{U}_{[0,0.5]}$ at different time lags ℓ. Figure 4 compares the memory capacity of DCRs before and after plasticity. For time lags $|\ell| > 5$, where the ability to reconstruct the input history starts to diverge from optimal (see Figure 4A), the increase of the DCR's memory capacity can reach up to 70%. The improvement is measured as the relative change in
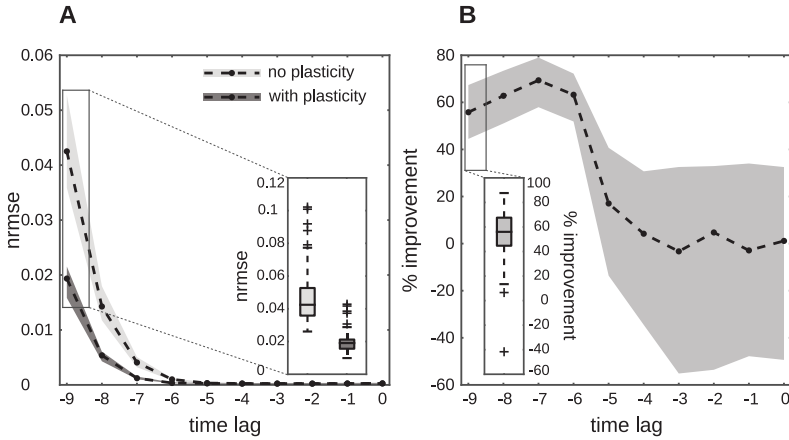
Figure 4: Memory capacity before and after plasticity. (A) Performance on memory construction before and after plasticity for different time lags. The inset shows performance on memory construction 10 time steps in the past ($\ell = -9$), before and after plasticity. (B) Relative improvement measured by the decrease in nrmse after applying homeostatic plasticity. The inset shows the improvement on memory construction 10 time steps in the past ($\ell = -9$). (A, B) The dotted lines are the medians of the corresponding plots, and the shaded areas mark the first and third quartiles. In addition to marking the quartiles, the insets show whiskers that extend to include data points within 1.5 times the interquartile range (the difference between the third and first quartiles). The crosses specify data points outside this range and correspond to outliers.

nrmse at each time lag due to plasticity. Only 1 of the 100 DCRs showed an approximately 20% deterioration in memory capacity after plasticity for the largest time lag (see the inset in Figure 4B).

4.2 Nonlinear Spatiotemporal Computations. A widely used benchmark in reservoir computing is the capacity to model a nonlinear autoregressive moving average (NARMA) system y in response to the uniformly distributed scalar input $u(\bar{t}) \sim \mathcal{U}_{[0,0.5]}$. The NARMA-10 task requires the DCR to compute, at each time step $\bar{t}$, a response

$$y(\bar{t}) = 0.3y(\bar{t} - 1) + 0.05y(\bar{t} - 1) \sum_{\bar{s}=1}^{10} y(\bar{t} - \bar{s})$$

$$+ 1.5u(\bar{t} - 1)u(\bar{t} - 10) + 0.1. \tag{4.2}$$

Thus, NARMA-10 requires modeling quadratic nonlinearities, and it shows a strong history dependence that challenges the DCR's memory
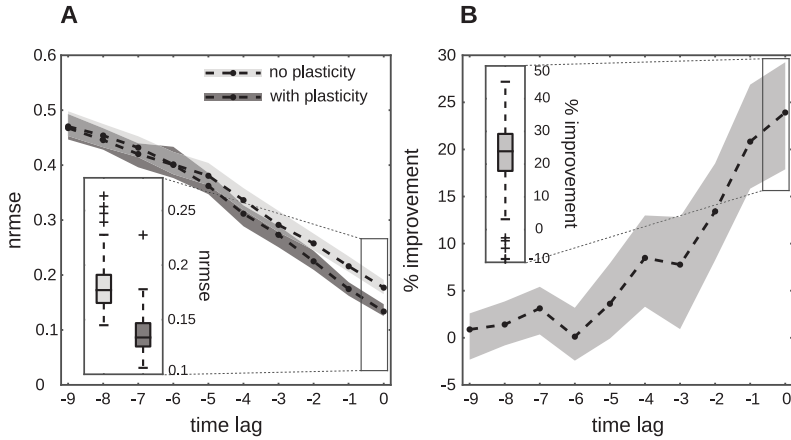
Figure 5: Spatiotemporal computational power before and after plasticity. (A) Performance on the NARMA-10 task before and after plasticity for different time lags. The inset shows performance at zero time lag before and after plasticity. (B) Relative improvement, measured by the decrease in nrmse, after applying homeostatic plasticity. The inset shows the improvement at zero time lag. (A, B) The dotted lines are the medians of the corresponding plots, while the shaded areas mark the first and third quartiles. In addition to marking the quartiles, the insets show whiskers that extend to include data points within 1.5 times the interquartile range. The crosses specify data points outside this range and correspond to outliers.

capacity. Figure 5 compares the performance in nrmse of DCRs before and after plasticity for different time lags. Even with no time lag $|\ell| = 0$, the task still requires the DCR to retain fading memory. This is in order to account for the dependence on inputs and outputs 10 time steps in the past. The plasticity mechanism achieves an approximately 22.8% improvement in performance on average, surpassing state-of-the-art values in both classical (Verstraeten et al., 2006) and delay-coupled reservoirs (Appeltant et al., 2011) with an average nrmse of 0.138 ± 0.02SD. Only in five trials did the performance deteriorate (see the inset in Figure 5B). The improvement decreases for larger time lags due to the deterioration in the DCR's memory capacity observed in Figure 4 but remains significant for $|\ell| < 5$.

5 Discussion: Effects of Plasticity

In order to explain the observed results, we analyze and discuss the effects of the homeostatic plasticity mechanism, equation 3.12, on the system's entropy $\mathcal{H}(x)$, virtual network topology, and the readout weights distribution $p(\alpha)$. We also discuss the role of the regulating parameter ρ.

5.1 Entropy. In section 3.2, we stated that expanding the informational features of present input requires a mechanism that counteracts the reduction of a v-delay due to the maximization of the v-node's sensitivity. To prove this hypothesis, we derive a learning mechanism that explicitly maximizes the mutual information between the DCR's response and its present input. Again, we assume the v-nodes are independent, and for a particular v-node i, we maximize the quantity

$$\mathcal{I}(x_i; u) = \mathcal{H}(x_i) - \mathcal{H}(x_i|u), \tag{5.1}$$

where $\mathcal{H}(x_i)$ is the entropy of the v-node's response, while $\mathcal{H}(x_i|u)$ is the entropy of the v-node's response conditioned on the input. In other words, $\mathcal{H}(x_i|u)$ is the entropy of the response that does not result from the input. Bell and Sejnowski (1995) argued that maximizing equation 5.1 with respect to some parameter θ is equivalent to maximizing $\mathcal{H}(x_i)$, since the conditional entropy $\mathcal{H}(x_i|u)$ does not depend on θ; that is, maximizing a v-node's input-information capacity is equivalent to maximizing its self-information capacity or entropy.

The entropy of x_i is given by $\mathcal{H}(x_i) = -\mathrm{E}\left[\ln p_x(x_i)\right]$, where $p_x(x_i)$ is the probability density function (PDF) of the v-node's response. Since x_i is an invertible function of the Mackey-Glass nonlinearity f_i (see equation 3.1), which is itself an invertible function of the input u (if the nonlinearity is chosen appropriately, such as in equation 2.4), the PDF of x_i can be written as a function of the PDF of f_i:

$$p_x(x_i) = \frac{p_f(f_i)}{\left|\frac{\partial x_i}{\partial f_i}\right|}. \tag{5.2}$$

The entropy of the v-node's response is then given by

$$\mathcal{H}(x_i) = \mathrm{E}\left[\ln\left|\frac{\partial x_i}{\partial f_i}\right|\right] - \mathrm{E}\left[\ln p_f(f_i)\right]. \tag{5.3}$$

The term $-\mathrm{E}\left[\ln p_f(f_i)\right]$ measures the entropy of the nonlinearity f_i and is independent of θ_i. From equation 5.3, and taking into account equation 3.1, we can derive a learning rule that maximizes the entropy of the response by applying stochastic gradient ascent:

$$\Delta\theta_i \propto \frac{\partial \mathcal{H}(x_i)}{\partial \theta_i} = \frac{\partial}{\partial \theta_i}\left(\ln\left|\frac{\partial x_i}{\partial f_i}\right|\right) = \left(\frac{\partial x_i}{\partial f_i}\right)^{-1}\frac{\partial}{\partial \theta_i}\left(\frac{\partial x_i}{\partial f_i}\right)$$

$$= (1 - e^{-\theta_i})^{-1}\frac{\partial}{\partial \theta_i}(1 - e^{-\theta_i}). \tag{5.4}$$

This leads to the following learning rule:

$$\Delta \theta_i = \alpha \frac{e^{-\theta_i}}{1 - e^{-\theta_i}}, \tag{5.5}$$

where $0 < \alpha \ll 1$ is a learning rate.

The update term, equation 5.5, is a strictly positive monotonic function of the v-delay θ_i. This entails that, when unconstrained, maximizing a v-node's informational feature expansion results in an unbounded increase in its v-delay: $\theta_i \to +\infty$. On the other hand, the plasticity rule, equation 3.12, can be rewritten as

$$\Delta \theta_i = \alpha \varsigma_i \rho - \alpha \varsigma_i \theta_i, \tag{5.6}$$

where $\varsigma_i = 2\sigma_i \theta_i^{2\rho-1} e^{-2\theta_i} > 0$. The term $\alpha \varsigma_i \rho$ in the plasticity mechanism, equation 5.6, is also positive. This entails that it results, similar to equation 5.5, in an unbounded increase in the v-delay and, as a corollary, in an increase in the v-node's informational feature expansion.

Given the above, the homeostatic plasticity mechanism, equation 3.12, for a particular DCR with delay τ, improves spatiotemporal computations by leading v-nodes to specialize in function. This is mediated by a competition between the v-nodes' sensitivity and their entropy. Some v-nodes become more sensitive to small fluctuations in input history, while others are brought closer to saturation where their entropy is higher and, as such, their ability for expanding informational features.

5.2 Virtual Network Topology. The effects of the homeostatic plasticity mechanism, equation 3.12, on the DCR's network topology can be deduced from equation 2.6, according to which self-weights are given by $w_{ii} = (1 - e^{-\theta_i})$, and the weights the v-node i receives from the preceding v-node $j = i - 1$ are $w_{ij} = (1 - e^{-\theta_j})e^{-\theta_i}$.

When θ_i decreases, so does the v-node's self-excitation w_{ii}, which is consistent with less saturation of the v-node's activity. In addition, the choice of the regulating parameter ρ describes the tendency of the v-node i to converge toward a particular self-excitation level $w_{ii} = (1 - e^{-\rho})$. This entails that for higher ρ, the v-node's target activity level increases, which also corresponds to higher entropy, as discussed in section 5.1.

The decrease in θ_i also leads the corresponding v-node's afferent w_{ij} to increase. This in turn increases the v-node j's influence on the activity of the v-node i, which results in a higher correlation between the two (or higher anti-correlation, depending on the signs of the corresponding mask values M_j and M_i). The increase of correlation is in agreement with simulation results and in accord with the decrease of the v-node's entropy as its v-delay decreases. This is the case since the influence of the current input is
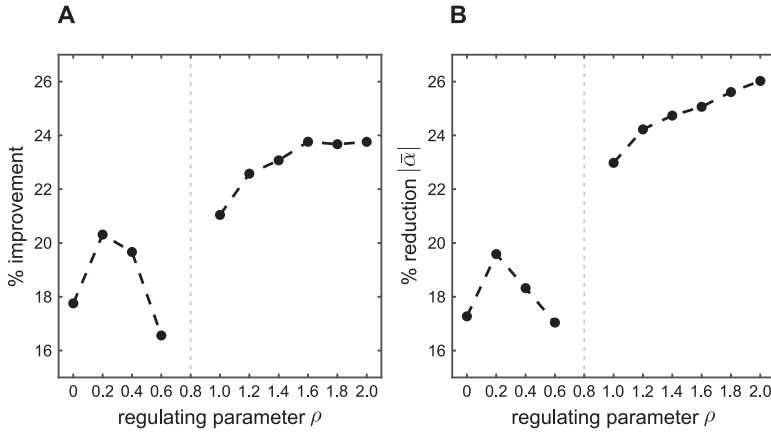
A

B

Figure 6: (A) Average improvement in performance and (B) reduction in average absolute values of the readout coefficients $|\bar{\alpha}|$ for different values of the regulating parameter ρ in comparison to the equidistant v-nodes case $\rho = 0.8$.

overshadowed by information from the input history that is delivered from the preceding v-node j, which now drives the v-node i. Figure 3B shows an exemplary virtual weight matrix following plasticity, which illustrates these changes in network topology due to the repositioning of v-nodes on the delay line.

5.3 Homeostatic Regulation Level. Introducing the parameter ρ is necessary for regulating the trade-off between sensitivity and entropy, that is, increasing and decreasing θ_i, as discussed analytically in section 5.1. It is also the defining factor in the v-node's tendency to collapse, as is evident from the form of the plasticity function, equation 5.6. The collapse of v-nodes is tantamount to a reduction in the DCR's dimensionality, which may be unfavorable with regard to the DCR's computational performance.

We test the latter hypothesis and the choice of the regulating parameter by running 1000 NARMA-10 trials for different ρ values that range between 0 and 2. Each trial shares the same mask M and the same NARMA-10 time series. As shown in Figure 6, the average improvement in performance in comparison to the reference equidistant case $\rho = \tau/n = 0.8$ increases for smaller ρ values but drops again for $\rho = 0$. An increase in $\rho > 0.8$ also increases the improvement of performance, but this increase saturates at $\rho = 1.6$. This is the case since the increase in v-delays favored by high ρ values makes the collapse of other v-delays inevitable in order to preserve the DCR's constant delay τ.

In a more detailed analysis, for each of the 1000 trials, we ranked different ρ values according to the resulting improvement of performance
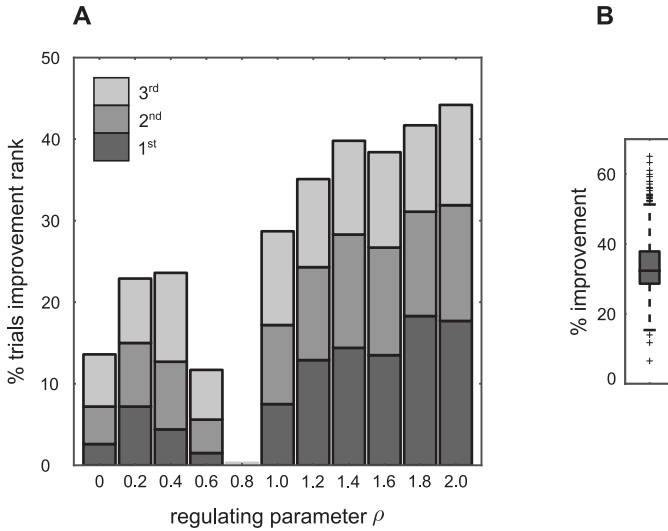
Figure 7: Performance of 1000 NARMA-10 trials for regulating parameter ρ values between 0 and 2. (A) Percentage of trials that achieved the first, second, and third highest improvement in performance for each ρ value. (B) Relative improvement measured by the decrease in nrmse, after applying homeostatic plasticity, given for each trial the best choice in ρ. The box plot marks the median, as well as the first and third quartiles. Whiskers extend to include data points within 1.5 times the interquartile range. The crosses specify data points outside this range and correspond to outliers.

in reference to the equidistant case $\rho = \tau/n = 0.8$. We then calculated the percentage of trials that achieved the highest improvement in performance (first rank) for some ρ value, compared to all other ρ values. We carried out the same procedure for the second and third ranks as well. Figure 7 confirms the previous results, as it shows that for $\rho = 0$, it is still possible to achieve the best improvement in performance, but it is less likely than other values. Figure 7 also illustrates a striking result. For none of the trials was the equidistant case, where no plasticity took place, the best choice regarding the computational power of the DCR. Only in 0.3% of the trials did the non-plastic equidistant case rank third. As a result, for a given DCR setup, there always exists a choice of ρ that results in nonequidistant v-nodes where spatiotemporal computations are enhanced. This is also summarized in Figure 7B, which shows the improvement in performance given the best choice in the regulating parameter ρ for each trial. The nrmse is reduced by approximately 33.7%, with an average performance that reaches an unprecedented value of nrmse approximately equal to 0.117 ± 0.01SD.

We point out that the homeostatic plasticity mechanism, equation 3.12, also reduces the average absolute values of the readout coefficients $|\bar{\alpha}|$ (see Figure 6B), which is similar in effect to an L_2-regularized model fit. This is not only advantageous with respect to numerical stability, but L_2-regularization also allows for a lower mean-square error on the validation set, as compared to an unregularized fit (Hoerl & Kennard, 1970).

We now briefly discuss the effects of the homeostatic regulation level ρ on the virtual network topology. As expected, and due to the simplex constraint, both smaller and larger values of ρ lead to a more uniform distribution of v-delays. However, most of the distribution's mass remains concentrated at $\theta_i = \tau/n$, that is, most v-delays remain unchanged or change only slightly. This has no effect on the qualitative features of the virtual network topology as outlined in section 5.2, but quantitatively, more weights approach the extremes of the range $[0, 1]$.

6 Commentary on Physical Realizability

We demonstrated that the suggested plasticity mechanism, equation 3.12, leads to spatiotemporal computational performances that surpass those of state-of-the-art results. An intuitive alternative to the plasticity mechanism would be to increase the number n of v-nodes within the constant full delay τ of the DCR. This solution, however, suffers from major drawbacks, particularly in regard to its realizability on physical hardware. Namely, there exists a physical constraint on the sampling rate of the DCR's activity, below which the speed and the feasibility of physical implementation is jeopardized. This imposes a minimal admissible v-delay within the full delay line and thus represents an upper bound on the number of equidistant v-nodes. This constraint is accounted for in the current approach by restricting the updates of v-delays due to plasticity to discrete step sizes κ. The parameter κ then corresponds to the minimal admissible v-delay (different from 0, which results in pruning the DCR). This is the case since κ is chosen such that $\tau/(\kappa n)$ is an integer, where this integer refers to the number of minimal v-delays that fit in one τ/n, the v-delay in the equidistant v-nodes case. In the current results, κ was chosen such that $\tau/(\kappa n) = 100$ in order for the discretization to present a good approximation of continuous v-delay values. Nevertheless, simulations show that improved computational power persists even for $\tau/(\kappa n) = 8$, which corresponds to κ, that is, an order of magnitude larger than the minimal experimentally viable v-delay. A stringent comparison between the results for different values of κ is problematic, since rougher quantization of v-delays, resulting from higher κ values, leads to less predictable effects on the behavior of the optimization problem, equation 3.10, particularly of how the discretized v-delay grid relates to the global maximum, which itself depends on the choice of the regulating term ρ. Nevertheless, the persistent improvement in performance stands in favor of the method's applicability in physical realizations.

Furthermore, increasing the number of v-nodes poses a practical limitation, even when v-delays remain within the constraints of physical implementation. As expected, the average computational performance does increase for larger numbers of v-nodes, but it saturates at some point. The plasticity mechanism improves the computational performance and, most important, reaches the saturation point of performance with a smaller number of v-nodes than the equidistant case. Beyond the performance saturation point, plasticity becomes ineffective on average, that is, it leads some trials to an increase and others to deterioration in computational performance. However, the redundancy resulting from increasing the number of v-nodes, even within the constraint of physical implementation, is disadvantageous with regard to the computational resources of the DCR: the linear readout mechanism remains a bottleneck, since increasing the number of regressors by sampling more v-nodes demands storing and inverting larger matrices, a serious challenge for both simulation and physical implementations. Again, the comparison between the results for different numbers of v-nodes is problematic, since changing the number of v-nodes modifies the statistics of the mask pattern, which may affect the proper choice of the regulating term ρ. With these considerations in mind, the plasticity mechanism is suitable for physical realization, since it saves resources by keeping the number of v-nodes smaller (and possibly pruning by leading some v-delays to collapse) and is computationally beneficial within the constraint of physical implementation, since it approaches the saturation point of computational performance using a smaller number of virtual nodes. Nevertheless, further detailed investigation remains necessary for addressing boundary conditions and applicability on physical implementation of the suggested plasticity mechanism.

7 Conclusion

We have introduced a plasticity mechanism for improving the computational capabilities of a DCR, a novel RC architecture where a single nonlinear node is delay-coupled to itself. The homeostatic nature of the derived plasticity mechanism, equation 3.12, relates directly to the information processing properties of the DCR in that it balances between sensitivity and informational expansion of input (see section 5.1). While the role of homeostasis in information processing and computation has only been discussed more recently, its function as a stabilization process of neural dynamics has acquired earlier attention (von der Malsburg, 1973; Bienenstock, Cooper, & Munro, 1982). From the perspective of the nervous system, pure Hebbian potentiation or anti-Hebbian depression would lead to destabilization of synaptic efficacies by generating amplifying feedback loops (Miller, 1996; Song, Miller, & Abbott, 2000), necessitating a homeostatic mechanism for stabilization (Davis & Goodman, 1998; Zhang & Linden, 2003; Turrigiano

& Nelson, 2004). Similarly, as suggested by the effects of the plasticity mechanism (see equation 5.6) on the virtual network topology (see section 5.2), the facilitating sensitivity term $-\alpha_\varsigma \theta$ is counteracted by the depressive entropy term $+\alpha_\varsigma \rho$, which prevents synaptic efficacies from overpotentiating or collapsing.

In addition, rewriting equation 3.12 as $\Delta\theta \sim (\theta - \rho)$ strongly relates the derived plasticity mechanism to normalization models of neural homeostatic plasticity. Normalization models consider plasticity rules that regulate the activity of the neuron toward a target firing rate. They are usually of the form $\Delta q \sim (r - r_{\text{tr}})$, where q is some quantity of relevance for learning, such as synaptic weights or the neuron's intrinsic excitability; r is an estimate of the neuron's output firing rate; and r_{tr} is the target firing rate (Kempter, Gerstner, & Van Hemmen, 2001; Renart et al., 2003; Vogels, Sprekeler, Zenke, Clopath, & Gerstner, 2011; Lazar et al., 2007, 2009; Zheng et al., 2013; Toutounji & Pipa, 2014). In analogy, the v-delay estimates the v-node's activity, since a larger θ_i results in higher self-excitation w_{ii}, while ρ defines the target activity of the v-node (see section 5.2). Furthermore, entropy of a neuron's output increases with its firing rate. As such, the increase of the v-delay θ, in response to the higher regulatory term ρ, also increases the v-node's entropy, as confirmed analytically in section 5.1.

Currently, and similar to the target firing rate r_{tr}, which is usually chosen according to biological constraints, the regulating parameter ρ is left as a free parameter, and its optimal choice for a particular DCR configuration is decided by brute force (see section 5.3). However, the statistics in Figures 6 and 7 conclusively show that any choice of ρ within the tested range leads to average and dominant improvement in computational performance in comparison to the equidistant case $\rho = \tau/n = 0.8$. Nevertheless, it is reasonable to assume that heuristics exist for the optimal choice of ρ, given a particular mask structure M, since the alterations in the mask values influence a v-node's sensitivity and entropy. A possible heuristic may relate the value of ρ to properties of maximum length sequences, by which Appellant, Van der Sande, Danckaert, & Fischer (2014) constructed mask sequences with equidistant v-nodes. Similarly, we speculate that the direction and amplitude of a v-delay's change, which are computationally advantageous, depend on the corresponding and preceding v-node's mask values M_j for $j \leq i$. The main difficulty arises from the fact that within the current formulation of the DCR in equations 2.5 and 2.6, no terms exist for relating different mask values to one another and to corresponding v-delays. This is also the main obstacle facing the derivation of plasticity mechanisms for updating the mask M beyond the binary pattern $\pm\mu$. The appropriate choice of ρ is complicated further by its dependence on the demands of the executed task in terms of memory, nonlinear computations, and entropy. Finding criteria that connect these aspects to the optimal choice of ρ requires extensive research, which is a subject of our current endeavors.

Enhancing the temporal multiplexing of input to the nonlinear node was the main goal of this article. We speculate that similar multiplexing may suggest a further important functionality of the extensive dendritic trees in some neuron types. On the one hand, Izhikevich (2006) discussed the infinite dimensionality that dendritic propagation delays offer to recurrent neural networks. On the other hand, several studies investigated the computational role of the spatial distribution of active dendrites (Rumsey & Abbott, 2006; Gollo, Kinouchi, & Copelli, 2009; Graupner & Brunel, 2012). In this article, we advocate a unified computational account that may integrate both the temporal and spatial aspects of dendritic computations. In particular, the spatial location of dendritic arbors may be optimized to achieve computationally favorable temporal multiplexing of the soma's input, in the fashion suggested by the DCR architecture. Consolidating this speculation will be the subject of future studies.

Appendix A: Solving and Simulating the DCR

In this appendix, we derive equations 2.5 and 2.6. We would like to solve system 2.3 for $x(t)$, with $(v - 1)\tau \leq t \leq v\tau$. Due to the recurrent dependency $x(t - \tau)$, this is not possible right away. However, if we assume a continuous function $\phi_v \in C_{[(v-2)\tau, (v-1)\tau]}$ is the solution for $x(t)$ on the previous τ-interval, we can replace $x(t - \tau)$ by $\phi_v(t - \tau)$. After the substitution, system 2.3 becomes solvable by the elementary method of variation of constants (Heuser, 2003). The latter provides a solution to an equation of type $\dot{x}(t) = a(t)x(t) + b(t)$ with initial condition $x(t_0) = c$. The general solution on the interval I to the inhomogeneous equation is then given by

$$x(t) = x_h(t) \left(c + \int_{t_0}^{t} x_h(t)^{-1} b(t) dt \right), \quad t \in I,$$

where

$$x_h(t) = \exp \left(\int_{t_0}^{t} a(t) dt \right)$$

denotes a solution to the associated homogeneous differential equation. Consequently, for $a(t) = -1$ and $b(t) = f(\phi_v(t - \tau), \delta(t))$, the solution to

$$\dot{x}(t) = -x(t) + f(\phi_v(t - \tau), \delta(t)),$$

subject to $x((v - 1)\tau) = \phi_v((v - 1)\tau)$, is given by

$$x(t) = e^{(v-1)\tau - t} \left(\phi_v((v - 1)\tau) + \int_{(v-2)\tau}^{t-\tau} f(\phi_v(s), \delta(s)) e^{s - (v-2)\tau} ds \right). \quad \text{(A.1)}$$

This expression can be used right away in a numerical solution scheme, where the integral is solved using the cumulative trapezoidal rule. The resulting simulation of the DCR has been shown to be comparable in its accuracy and computational capabilities to adaptive numerical solutions, while saving considerable computation time (Schumacher, Toutounji, & Pipa, 2013).

Recall that $t_i = (v-1)\tau + \sum_{j=1}^{i} \theta_j$, with θ_j being the temporal distances between consecutive virtual nodes. To arrive at a manageable analytical expression of the above solution for the sampling point t_i of virtual node i during the vth τ-cycle, we make the following approximation.

Let the DCR activity at a particular v-node $x_i(\bar{t}) = x(t_i)$, its nonlinearity $f_i(\bar{t}) = f(x_i(\bar{t}-1), M_i \cdot u(\bar{t}))$, and the DCR time step $\bar{t} = \lceil \frac{t}{\tau} \rceil = v$. If we assume that $f(\cdot)$ is piecewise constant at each θ_i, which is a valid approximation since $\theta_i \ll \tau$, expression 6.1 simplifies further to

$$x_i(\bar{t}) = e^{-\sum_{j=1}^{i}\theta_j}\left(x_n(\bar{t}-1) + \sum_{j=1}^{i} f_i(\bar{t})\int_0^{\theta_j} e^s ds\right)$$

$$= e^{-\sum_{j=1}^{i}\theta_j}\left(x_n(\bar{t}-1) + \sum_{j=1}^{i} f_i(\bar{t})\left(e^{\theta_j}-1\right)\right)$$

$$= e^{-\sum_{j=1}^{i}\theta_j}x_n(\bar{t}-1) + \sum_{j=1}^{i}(1 - e^{-\theta_j})e^{-\sum_{k=j+1}^{i}\theta_k} \cdot f_j(\bar{t}).$$

Appendix B: Constraint Satisfaction

The sensitivity update rule of the virtual node distances θ_j has to satisfy the constraint $\sum_j \theta_j = \tau$. This describes a constraint manifold for valid virtual node distance vectors $\Theta \in \mathbb{R}^n$ during learning. The manifold has the structure of a simplex,

$$V := \left\{x | x = (x_1, \dots, x_n)^T, \sum_{i=1}^{n} x_i = \tau\right\},$$

with $\dim V = n - 1$ and simplex corners given by τe_i $(i = 1, \dots, n)$, where $(e_i)_{i=1}^n$ is the standard orthonormal basis of $\mathbb{R}^n$. We implemented the constraint optimization problem by first computing an unconstrained update for Θ, followed by an orthogonal projection onto V. Due to the simple linear structure of V, this strategy will converge onto the constrained optimum for Θ.

Denote by $n_V = \frac{\tau}{n}\sum_i e_i$ the central point of the constraint simplex, and let $(v_i)_{i=1}^{n-1}$, $v_i \in \mathbb{R}^n$, be an orthonormal basis for V. The latter is computed

from an orthogonal basis, which can be constructed by simple geometrical considerations from the simplex corner point vectors as

$$
\begin{pmatrix} -\tau & \cdots & -\tau \\ & \tau I_{n-1} & \end{pmatrix} \in \mathbb{R}^{n \times n-1},
\tag{B.1}
$$

where I_{n-1} denotes the $(n-1)$-dimensional unit matrix. It is easily verified that this basis spans V and is indeed orthogonal. In conjunction with the inhomogeneity n_V, a normal vector with respect to V, any point on V can be expressed via the v_i. For some $x \in \mathbb{R}^n$ being the result of an unconstrained sensitivity update step, the constraint can be met by projecting x orthogonally onto V via the mapping

$$
\begin{aligned}
\pi_V(x) &= n_V + \sum_{i=1}^{n-1} \left((x - n_V)^T v_i \right) v_i \\
&= n_V + \sum_{i=1}^{n-1} \left(v_i^T (x - n_V) \right) v_i \\
&= n_V + \underbrace{\left(\sum_{i=1}^{n-1} v_i v_i^T \right)}_{:= M \in \mathbb{R}^{n \times n}} (x - n_V) \\
&= n_V + M(x - n_V).
\end{aligned}
\tag{B.2}
$$

The addition and subtraction of n_V take care of the fact that V, being a hyperplane, is translated out of the origin by the inhomogeneity n_V. If the V-plane was centered in the origin, $\underbrace{(v_i^T x)}_{\in \mathbb{R}} v_i$ would denote the orthogonal projection of x onto the ith orthonormal basis vector. Accordingly, the linear combination of these projections yields the representation of $\pi_V(x)$ with respect to the basis $(v_i)_{i=1}^{n-1}$.

Acknowledgments

References

Appeltant, L., Soriano, M. C., Van der Sande, G., Danckaert, J., Massar, S., Dambre, J., . . . Fischer, I. (2011). Information processing using a single dynamical node as complex system. *Nat. Commun.*, *2*.

Appeltant, L., Van der Sande, G., Danckaert, J., & Fischer, I. (2014). Constructing optimized binary masks for reservoir computing with delay systems. *Sci. Rep.*, *4*.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.*, *7*(6), 1129–1159.

Bienenstock, E. L., Cooper, L. N., & Munro, P. W. (1982). Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *J. Neurosci.*, *2*(1), 32–48.

Boyd, S., & Chua, L. O. (1985). Fading memory and the problem of approximating nonlinear operators with Volterra series. *IEEE Trans. Circuits Syst.*, *32*(11), 1150–1161.

Brunner, D., Soriano, M. C., Mirasso, C. R., & Fischer, I. (2013). Parallel photonic information processing at gigabyte per second data rates using transient states. *Nature Commun.*, *4*, 1364.

Buonomano, D. V., & Maass, W. (2009). State-dependent computations: spatiotemporal processing in cortical networks. *Nat. Rev. Neurosci.*, *10*(2), 113–125.

Dasgupta, S., Wörgötter, F., & Manoonpong, P. (2013). Information dynamics based self-adaptive reservoir for delay temporal memory tasks. *Evolving Systems*, *4*(4), 235–249.

Davis, G. W., & Goodman, C. S. (1998). Synapse-specific control of synaptic efficacy at the terminals of a single neuron. *Nature*, *392*(6671), 82–86.

Glass, L., & Mackey, M. (2010). Mackey-Glass equation. *Scholarpedia*, *5*(3), 6908.

Gollo, L. L., Kinouchi, O., & Copelli, M. (2009). Active dendrites enhance neuronal dynamic range. *PLoS Comput. Biol.*, *5*(6), e1000402.

Graupner, M., & Brunel, N. (2012). Calcium-based plasticity model explains sensitivity of synaptic changes to spike pattern, rate, and dendritic location. *Proc. Natl. Acad. Sci. U.S.A.*, *109*(10), 3991–3996.

Guo, S., & Wu, J. (2013). *Bifurcation theory of functional differential equations*. New York: Springer.

Häusler, S., & Maass, W. (2007). A statistical analysis of information-processing properties of lamina-specific cortical microcircuit models. *Cereb. Cortex*, *17*(1), 149–162.

Heuser, H. (2003). *Lehrbuch der analysis*. Stuttgart: Teubner.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67.

Hoinville, T., Siles, C. T., & Hénaff, P. (2011). Flexible and multistable pattern generation by evolving constrained plastic neurocontrollers. *Adapt. Behav.*, *19*(3), 187–207.

Izhikevich, E. M. (2006). Polychronization: computation with spikes. *Neural Comput.*, *18*(2), 245–282.

Jaeger, H. (2001). *The "echo state" approach to analysing and training recurrent neural networks* (Tech. Rep. GMD 148). Bremen: German National Research Center for Information Technology.

Jaeger, H., & Haas, H. (2004). Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, *304*(5667), 78–80.

Jaeger, H., Lukoševičius, M., Popovici, D., & Siewert, U. (2007). Optimization and applications of echo state networks with leaky-integrator neurons. *Neural Netw.*, *20*(3), 335–352.

Karmarkar, U. R., & Buonomano, D. V. (2007). Timing in the absence of clocks: Encoding time in neural network states. *Neuron*, *53*(3), 427–438.

Kempter, R., Gerstner, W., & Van Hemmen, J. L. (2001). Intrinsic stabilization of output rates by spike-based Hebbian learning. *Neural Comput.*, *13*(12), 2709–2741.

Larger, L., Soriano, M., Brunner, D., Appeltant, L., Gutiérrez, J. M., Pesquera, L., . . . Fischer, I. (2012). Photonic information processing beyond Turing: An optoelectronic implementation of reservoir computing. *Opt. Express*, *20*(3), 3241–3249.

Lazar, A., Pipa, G., & Triesch, J. (2007). Fading memory and time series prediction in recurrent networks with different forms of plasticity. *Neural Netw.*, *20*(3), 312–322.

Lazar, A., Pipa, G., & Triesch, J. (2009). SORN: A self-organizing recurrent neural network. *Front. Comput. Neurosci.*, *3*(23).

Lukoševičius, M., & Jaeger, H. (2009). Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, *3*(3), 127–149.

Maass, W., Natschläger, T., & Markram, H. (2002). Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Comput.*, *14*(11), 2531–2560.

Marković, D., & Gros, C. (2012). Intrinsic adaptation in autonomous recurrent neural networks. *Neural Comput.*, *24*(2), 523–540.

Miller, K. D. (1996). Synaptic economics: Competition and cooperation in synaptic plasticity. *Neuron*, *17*(3), 371–374.

Naudé, J., Cessac, B., Berry, H., & Delord, B. (2013). Effects of cellular homeostatic intrinsic plasticity on dynamical and computational properties of biological recurrent neural networks. *J. Neurosci.*, *33*(38), 15032–15043.

Nikolić, D., Häusler, S., Singer, W., & Maass, W. (2009). Distributed fading memory for stimulus properties in the primary visual cortex. *PLoS Biol.*, *7*(12), e1000260.

Paquot, Y., Duport, F., Smerieri, A., Dambre, J., Schrauwen, B., Haelterman, M., & Massar, S. (2012). Optoelectronic reservoir computing. *Sci. Rep.*, *2*.

Pascanu, R., & Jaeger, H. (2011). A neurodynamical model for working memory. *Neural Netw.*, *24*(2), 199–207.

Remme, M. W., & Wadman, W. J. (2012). Homeostatic scaling of excitability in recurrent neural networks. *PLoS Comput. Biol.*, *8*(5), e1002494.

Renart, A., Song, P., & Wang, X.-J. (2003). Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. *Neuron*, *38*(3):473–485.

Rumsey, C. C., & Abbott, L. F. (2006). Synaptic democracy in active dendrites. *J. Neurophysiol.*, *96*(5), 2307–2318.

Schrauwen, B., Wardermann, M., Verstraeten, D., Steil, J. J., & Stroobandt, D. (2008). Improving reservoirs using intrinsic plasticity. *Neurocomputing*, *71*(7), 1159–1171.

Schumacher, J., Toutounji, H., & Pipa, G. (2013). An analytical approach to single node delay-coupled reservoir computing. In P. Mladenov, V. Koprinkova-Hristova, G. Palm, A.E.P. Villa, B. Appollini, & N. Kasabov (Eds.), *Lecture Notes in Computer Science*, *Vol. 8131*: *Artificial Neural Networks and Machine Learning–ICANN 2013* (pp. 26–33). New York: Springer.

Somers, D. C., Nelson, S. B., & Sur, M. (1995). An emergent model of orientation selectivity in cat visual cortical simple cells. *J. Neurosci.*, *15*(8), 5448–5465.

Song, S., Miller, K. D., & Abbott, L. F. (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nat. Neurosci.*, *3*(9), 919–926.

Soto-Treviño, C., Thoroughman, K. A., Marder, E., & Abbott, L. (2001). Activity-dependent modification of inhibitory synapses in models of rhythmic neural networks. *Nat. Neurosci.*, *4*(3), 297–303.

Toutounji, H., & Pasemann, F. (2014). Behavior control in the sensorimotor loop with short-term synaptic dynamics induced by self-regulating neurons. *Front. Neurorobot.*, *8*, 19.

Toutounji, H., & Pipa, G. (2014). Spatiotemporal computations of an excitable and plastic brain: Neuronal plasticity leads to noise-robust and noise-constructive computations. *PLoS Comput. Biol.*, *10*(3), e1003512.

Turrigiano, G. G., & Nelson, S. B. (2004). Homeostatic plasticity in the developing nervous system. *Nat. Rev. Neurosci.*, *5*(2), 97–107.

Vargas, P. A., Moioli, R. C., Von Zuben, F. J., & Husbands, P. (2009). Homeostasis and evolution together dealing with novelties and managing disruptions. *Int. J. Intelligent Computing and Cybernetics*, *2*(3), 435–454.

Verstraeten, D., Schrauwen, B., & Stroobandt, D. (2006). Reservoir-based techniques for speech recognition. In *International Joint Conference on Neural Networks* (pp. 1050–1053). Piscataway, NJ: IEEE.

Vogels, T. P., Sprekeler, H., Zenke, F., Clopath, C., & Gerstner, W. (2011). Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science*, *334*(6062), 1569–1573.

von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, *14*(2), 85–100.

Williams, H., & Noble, J. (2007). Homeostatic plasticity improves signal propagation in continuous-time recurrent neural networks. *Biosystems*, *87*(2), 252–259.

Yamazaki, T., & Tanaka, S. (2007). The cerebellum as a liquid state machine. *Neural Netw.*, *20*(3), 290–297.

Zhang, W., & Linden, D. J. (2003). The other side of the engram: Experience-driven changes in neuronal intrinsic excitability. *Nat. Rev. Neurosci.*, *4*(11), 885–900.

Zheng, P., Dimitrakakis, C., & Triesch, J. (2013). Network self-organization explains the statistics and dynamics of synaptic connection strengths in cortex. *PLoS Comput. Biol.*, *9*(1), e1002848.