Analysis of synaptic scaling in combination with Hebbian plasticity in several simple networks

Christian Tetzlaff [1,2,3] *, Christoph Kolodziejski [2,3,4], Marc Timme [2,3,4] and Florentin Wörgötter [1,3]

[1] III. Institute of Physics – Biophysics, Georg–August–Universität Göttingen, Göttingen, Germany
[2] Network Dynamics Group, Max Planck Institute for Dynamics and Self-Organization, Göttingen, Germany
[3] Bernstein Center for Computational Neuroscience, Göttingen, Germany
[4] Institute for Non-linear Dynamics, Georg–August–Universität Göttingen, Göttingen, Germany

Conventional synaptic plasticity in combination with synaptic scaling is a biologically plausible plasticity rule that guides the development of synapses toward stability. Here we analyze the development of synaptic connections and the resulting activity patterns in different feed-forward and recurrent neural networks, with plasticity and scaling. We show under which constraints an external input given to a feed-forward network forms an input trace similar to a cell assembly (Hebb, 1949) by enhancing synaptic weights to larger stable values as compared to the rest of the network. For instance, a weak input creates a less strong representation in the network than a strong input which produces a trace along large parts of the network. These processes are strongly influenced by the underlying connectivity. For example, when embedding recurrent structures (excitatory rings, etc.) into a feed-forward network, the input trace is extended into more distant layers, while inhibition shortens it. These findings provide a better understanding of the dynamics of generic network structures where plasticity is combined with scaling. This makes it also possible to use this rule for constructing an artificial network with certain desired storage properties.

Keywords: plasticity, synaptic scaling, neural network, homeostasis, synapse, signal propagation

1. INTRODUCTION

Synaptic plasticity in neural systems needs to be regulated without which unwanted effects, like overly strong growth or shrinkage, might occur, destabilizing the network function. Little is known about the underlying biophysical mechanisms which control weight growth. Learning rules (plasticity rules) usually achieve this by weight regularization terms (Bienenstock et al., 1982; Oja, 1982; Miller and MacKay, 1994; Gerstner and Kistler, 2002). A possible alternative arises when considering so-called "synaptic scaling." This is a mechanism, discovered around 1998, by which network activity is homeostatically regulated (Turrigiano et al., 1998; Turrigiano and Nelson, 2000, 2004). Overly active networks will – on a time axis of hours up to days – down scale their activity and vice versa. This is achieved by altering the synaptic strengths usually across many neurons, which acts like a scaling mechanism (Turrigiano and Nelson, 2000). Thus, synaptic weights ω seem to be regulated by an activity dependent difference term. This term compares for every neuron – hence locally – output activity v against a terminal activity v_T such that $\frac{d\omega}{dt} \sim \gamma(v_T - v)$, where $\gamma \ll 1$ is a rate factor that strongly limits this effect and thereby defines a time scale much slower than that of conventional plasticity (e.g., Hebbian plasticity Hebb, 1949 or Spike-timing dependent plasticity, STDP; Bi and Poo, 1998), which is dominated by another rate factor $1 \gg \mu > \gamma$.

As a consequence, synaptic change needs to be described by a combination of a conventional plasticity rule together with this more slowly acting scaling rule: $\frac{d\omega}{dt} = \mu G(u, v) + \gamma H(v_T - v, \omega)$,

where G and H describe the specific instantiations of these rules (u is the presynaptic activity). For example, G would be different for a plain hebbian rule ($G = uv$) as compared to the BCM rule ($G = uv(v - \Theta)$).

In a previous study (Tetzlaff et al., 2011) we have shown that such a combination of Synaptic Plasticity and Synaptic Scaling (SPaSS) leads to a rule which is globally stable for a wide variety of conditions as soon as scaling depends quadratically on the weights ($H \sim \omega^2$). Several interesting properties were discussed. For example, a strong external input delivered to a neuron leads to large (stable) post-synaptic weights for this neuron and its direct as well as indirect target neurons. This way reliable propagation of the external signal along several stages becomes possible because all these neurons are connected with each other with strong synapses. Thus, an input trace is stored. This appears interesting as the SPaSS rule apparently allows the system to form such linked groups of neurons, which could be considered as a "cell assembly" (Hebb, 1949; Hahnloser et al., 2002; Harris et al., 2003). So far the dynamic creation and stabilization of cell assemblies has remained an enigma. While the SPaSS rule seem to achieve this (at least to some degree) so far it also remains unclear how these dynamics evolve. Thus, the goal of the current study is to analyze the dependency of size (number of neurons/stages) and strength (weights) of these input traces on the underlying connectivity and used plasticity parameters.

To this end we will use a combined rule with a hebbian plasticity term for G and the above mentioned quadratic weight dependence

for H (Tetzlaff et al., 2011) between neurons i and j:

$$\dot{\omega}_{ij} = \mu\, u_j\, v_i + \gamma(v_T - v_i)\omega_{ij}^2$$
$$= \mu\left(u_j\, v_i + \kappa^{-1}(v_T - v_i)\omega_{ij}^2\right) \quad (1)$$

with $\dot{\omega}_{ij} = \frac{d\omega_{ij}}{dt}$ and $\kappa = \mu/\gamma$. In contrast to (Tetzlaff et al., 2011), we will continue to use μ and κ and not μ and γ. This is because the fixed points of such systems are only influenced by the γ to μ ratio (i.e., κ) and μ as such merely alters the time scale of our model. Additionally we assume that the output of neuron i depends linearly on the product of input and weight: $v_i = \Sigma_j\, w_{ij} \cdot u_j$. Then we can adopt the results of a one synapse system from (Tetzlaff et al., 2011) and state the stable excitatory fixed point by using κ as:

$$\omega_{ij}^* = \frac{v_T}{2\,u_j} + \sqrt{\kappa\, u_j + \left(\frac{v_T}{2\,u_j}\right)^2} \quad (2)$$

or rewritten with $v_i = w_{ij} \cdot u_j$ as:

$$v_i^* = \frac{v_T}{2} + \sqrt{\kappa\, u_j^3 + \left(\frac{v_T}{2}\right)^2}. \quad (3)$$

This activity value v_i^* is the target activity of neuron i given input u_j. This target activity becomes equal to the terminal activity v_T if synaptic plasticity does not influence synaptic weights anymore (i.e., $\kappa = 0$).

Now, we want to reintroduce in more detail the results from (Tetzlaff et al., 2011) which are relevant for the current study: **Figure 1A** shows a schematic rendering of a network of 100 neurons that has been designed with three connections from each neuron to randomly chosen target neurons. All neurons are rate coded and we have used hebbian plasticity together with scaling as defined in equation (1). Three neurons have received strong, constant, external input, all other neurons only weak random input. This leads to two groups of neuronal descendent-lines in the network. Group 1: The three neurons receiving the strong

FIGURE 1 | Given a strong external input synaptic plasticity combined with synaptic scaling (equation 1) leads to an input trace similar to a cell assembly. (A) Schematic of post-synaptic connectivity of selected neurons up to stage five. The red neuron receives a strong external input (yellow arrow). Parts of the descendent network stages are highlighted to show the general connectivity structures analyzed in this study (c.f. **Figure 2**). **(B)** Neural activities (dashed lines) and weights (solid lines) after stabilization found for the first four stages. Weights and activities of the stages linked to the external input (black) are significantly larger compared to control neurons (gray) over the first three stages representing an input trace (cell assembly).

input (input neurons) project to 9 children, 27 grandchildren, 81 great-grandchildren, etc., where – due to the randomness of the connection patterns – loops can be formed, too. Group 2: The same descendent line arises for any randomly chosen other three neurons (control neurons), too. After network relaxation we have analyzed how the fixed points of activities and weights look like for these two groups of neurons. The black lines in **Figure 1B** show activities (solid) and weights (dashed) for the input-descendent group; the gray lines for the control descendent group. Neurons that descend from the input neurons represent the input along at least three connection stages by producing higher activities and higher, stable weights. Thus, this network was able to store an input trace and form a "cell assembly" (Hebb, 1949). For more details concerning this result see Tetzlaff et al., 2011.

In the current study we are going to investigate the sphere of influence of the given external input on the fixed points of weights and activities of the network dependent on the underlying connectivity. For this, we split the complex random network (c.f. **Figure 1A**) in smaller, more generalized parts (e.g., purple areas in **Figures 1A** and **2**) and analyze their influence on size (number of neurons/stages) and strength (weights) of the input trace (cell assembly).

First, we obtain analytical results for excitatory feed-forward networks (**Figure 2A**) with and without lateral connections and with feed-forward or lateral inhibition. In addition, we derive stability constraints for the maximally allowed input strength. An interesting result is that the combination of plasticity and scaling reduces lateral signal dispersion in feed-forward networks even without inhibition.
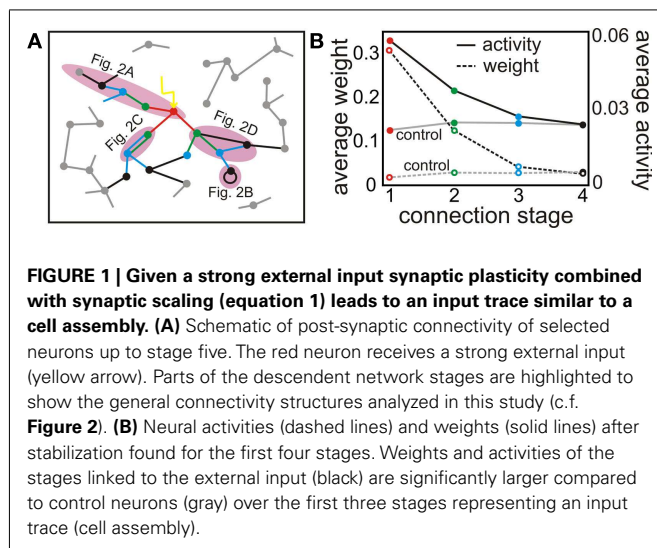
Second, we investigate recurrent network structures which are generally more difficult to handle and analytical results are hard to derive (here we mostly rely on numerics). However, we still provide analytical insights which lead to two observation: First, recurrences increase activity compared to feed-forward structures and, therefore, extend the input trace (cell assembly) within the network. Second, if we connect neurons so that they form a recurrent ring, the stability of rings of different size in a network is determined by the stability of the ring with the smallest number of neurons. All these aspects are consolidated by numerics.
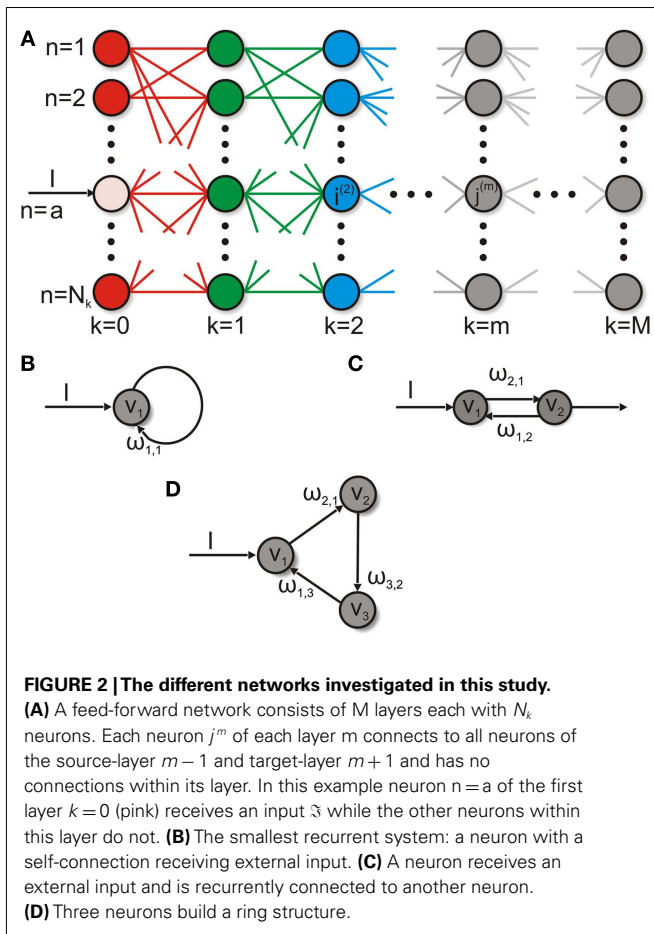
To help navigating through this diversity, we provide at the end of most subsections a short paragraph that summarizes the results from this subsection. As part of the Discussion section we, finally, summarize all main findings (also in **Table 1**) and based on the main findings of this study give an example of a topology which results in localized cell assemblies (**Figure 8**).

2. MATERIALS AND METHODS
2.1. GENERAL FIXED POINT ANALYSIS
The differential equations of weights and activities for different network topologies are derived in the main text and analyzed according to their fixed point structure. This analysis is based on standard methods for a given set of differential equations $\frac{d\omega}{dt} = \dot{\omega} = \Upsilon(\omega)$ determining the dynamics of ω and its fixed points ω^*, where $\dot{\omega} = 0$. To assess the stability of these fixed points, we analytically computed the Jacobian $J_{\Upsilon}(\omega)$ of Υ; a fixed point

FIGURE 2 | The different networks investigated in this study.
(A) A feed-forward network consists of M layers each with N_k neurons. Each neuron j^m of each layer m connects to all neurons of the source-layer $m-1$ and target-layer $m+1$ and has no connections within its layer. In this example neuron n = a of the first layer k = 0 (pink) receives an input $\Im$ while the other neurons within this layer do not. **(B)** The smallest recurrent system: a neuron with a self-connection receiving external input. **(C)** A neuron receives an external input and is recurrently connected to another neuron. **(D)** Three neurons build a ring structure.

Table 1 | Tabular summary of the influence of different network structures compared to a pure feed-forward network.

Network structure	v^{max}	Input interval	Length of inp. trace	Global fixed point (g)
Lateral excitation	⇓	⇓	⇑	⇓
Feed-forward inhibition	⇑	⇑	⇓	⇑
Feedback inhibition	⇑	⇑	⇓	⇑
Recurrent exci.(deeper layers)	–	–	⇑	–
Recurrent exci.(earlier layers)	–	⇓	⇑	–

Each network structure analyzed in this study influences different properties of the propagation mechanism in a feed-forward network with the SPaSS rule. Arrows (⇑, ⇓) show the change of these properties in comparison to the pure feed-forward network. Bar (–) means no difference. Properties: v^{max}: position of the upper boundary for bounded development; input interval: interval of inputs leading to bounded development; length of inp. trace: number of layers needed to reach the global fixed point; global fixed point (g): weight value ($\omega^{(g)}$) of the global fixed point. For more details see main text.

ω^* is stable if all eigenvalues at $\omega = \omega^*$ are smaller than zero, and unstable otherwise. Numerics are done by solving the differential equations with the Euler method.

2.2. FEED-FORWARD NETWORK
The feed-forward network (see **Figure 2A**) consists of M layers each with N_k neurons. Each neuron $n \in \{1, \ldots, N\}$ of each layer $k \in \{0, \ldots, M-1\}$ has an all-to-all connection to all neurons of the target-layer $k+1$ and no connection to neurons within its layer k if not stated differently. Therefore, neuron $n = i^{(m)}$ of layer m receives its inputs from all neurons $j^{(m-1)}$ of layer $m-1$ with their activities $v_j^{(m-1)}$ via synapses of strength $\omega_{i,j}^{(m)}$. Thus, the output activity $v_i^{(m)}$ of neuron $i^{(m)}$ is the function $F\left(\sum_j^{N_{m-1}} \omega_{i,j}^{(m)} \cdot v_j^{(m-1)}\right)$. It will be transmitted to all neurons of layer $m+1$. For simplicity, in this study the function F is chosen to be the identity which does not influence the results qualitatively (Tetzlaff et al., 2011).

3. RESULTS
3.1. FEED-FORWARD NETWORKS
3.1.1. Weights and activities in feed-forward networks
The simplified structure in **Figure 2A** depicts how an input signal to a random network is traveling along topological stages (layers) within this network and how the network "reacts" to the input by adapting weights and activities. For this we simulate the behavior of such a feed-forward network with one neuron per stage (c.f. **Figures 1A** and **3**) given a noisy input to layer one (activity of red neuron). As shown before (**Figure 1**), this activity leads to a strong post-synaptic synapse resulting in high (but not as high as red) activity at layer 1 (green neuron). This behavior is propagated until the activity vanishes in layer 4 (gray). This is the basic property of the above described input trace (cell assembly). Even if the input is reduced for a certain duration the network can re-adapt to its previous stable state very fast (c.f. **Figure 3**). Thus, even with a noisy external input, the network can construct a stable cell assembly along several stages.

In the following we analyze analytically the fixed points of weights and activities for each stage dependent on feed-forward connections and the input. We consider an input of average strength $\Im$ given to a neuron $j^{(0)} = a$ of the first layer $k = 0$ (pink neuron in **Figure 2A**), which leads then to the input activity $v_{j=a}^{(k=0)} = v_a^{(0)} = \Im$ of the first layer (the other neurons in layer $k = 0$ have $v_{j \neq a}^{(0)} = 0$). Next, we calculate the weights $\omega_{i,j}^{(1)}$ and activities $v_i^{(1)}$ of a neuron $i^{(1)}$ of the next layer $k = 1$ assuming that the dynamics have stabilized:

$$v_i^{(1)} = \sum_{j=1}^{N} \omega_{i,j}^{(1)} \cdot v_j^{(0)} = \omega_{i,a}^{(1)} \cdot v_a^{(0)} = \omega_{i,a}^{(1)} \cdot \Im. \quad (4)$$

Given this result, equation (1) and $u_j^{(m)} = v_j^{(m-1)}$ the dynamics for weights in the second layer write

$$\dot{\omega}_{i,a}^{(1)} = \mu \left(v_a^{(0)} \cdot v_i^{(1)} + \kappa^{-1} \left(v_T - v_i^{(1)} \right) \cdot \left(\omega_{i,a}^{(1)} \right)^2 \right)$$
$$= \mu \left(\Im^2 \cdot \omega_{i,a}^{(1)} + \kappa^{-1} \left(v_T - \Im \cdot \omega_{i,a}^{(1)} \right) \cdot \left(\omega_{i,a}^{(1)} \right)^2 \right) \quad (5)$$

This equation still has only one fixed point in the excitatory regime (i.e., $\omega > 0$):

$$\omega_{i,a}^{(1)} = \frac{v_T}{2\Im} + \sqrt{\kappa\,\Im + \frac{v_T^2}{4\,\Im^2}} \equiv \omega_{(\cdot),a}^{(1)} \qquad (6)$$

Substituting this solution into equation (4) gives

$$v_i^{(1)} = \frac{v_T}{2} + \sqrt{\kappa\,\Im^3 + \frac{v_T^2}{4}} \equiv v_{(\cdot)}^{(1)} \qquad (7)$$

We write $v_{(\cdot)}^{(m)}$ as this solution applies for all $i^{(1)} \in \{1, \ldots, N_1\}$ of layer $k = 1$ because neuron $j^{(0)} = a$ is connected to all neurons of layer $k = 1$.

All neurons of layer $k = 1$ have the same activity and since we use an all-to-all connectivity between each layer, neurons within each subsequent layer $k \geq 1$ will have equal activity, too. Hence, the (equal) activity of each neuron in layer $k = m$ and also the (equal) strength of each weight projecting to layer $k = m$ merely depends on the (equal) activity of each neuron in layer $k = m - 1$ and the number of neurons $N_{(m-1)}$ in this layer:

$$\omega_{i,(\cdot)}^{(m)} = \frac{v_T}{2\,N_{(m-1)}\,v_{(\cdot)}^{(m-1)}}$$
$$+ \sqrt{\kappa\,v_{(\cdot)}^{(m-1)} + \left(\frac{v_T}{2\,N_{(m-1)}\,v_{(\cdot)}^{(m-1)}}\right)^2} \equiv \omega_{(\cdot),(\cdot)}^{(m)} \qquad (8)$$

$$v_i^{(m)} = N_{(m-1)} \cdot \omega_{i,(\cdot)}^{(m)} \cdot v_{(\cdot)}^{(m-1)} = \frac{v_T}{2}$$
$$+ \sqrt{\kappa\,N_{(m-1)}^2 \left(v_{(\cdot)}^{(m-1)}\right)^3 + \left(\frac{v_T}{2}\right)^2} \equiv v_{(\cdot)}^{(m)} \qquad (9)$$

In the next subsection we will analyze and discuss the dynamics of these equations for different parameters given the same number of neurons in each layer ($N = N_k, \forall k \in \{0, \ldots, M-1\}$).

3.1.2. Development of weights and activities along layers

Equations (8) and (9) describe the dynamics of weights and activities in a feed-forward network that is stimulated by just one input. When we look at the propagation of the activity (after stabilization of the weights) for different parameter values, we will find two qualitatively different scenarios: Activities will increase (*divergent regime*) or decrease (*bounded regime*) from layer to layer. As divergence should be avoided and bounded development enforced, the following relation has to be fulfilled for all neurons $m > 1 : v_{(\cdot)}^{(m)} \leq v_{(\cdot)}^{(m-1)}$. We now calculate lower and upper bounds for the activity and the weights so that we stay in the bounded regime.
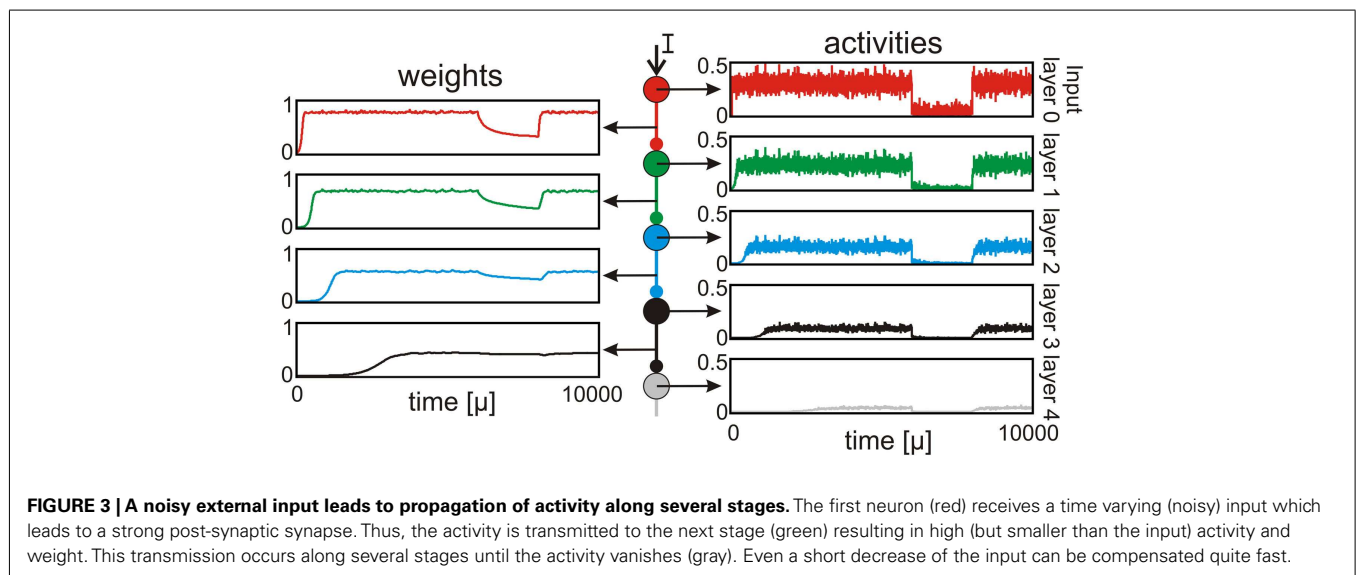
$$v_{(\cdot)}^{(m)} \overset{!}{\leq} v_{(\cdot)}^{(m-1)}$$

$$\frac{v_T}{2} + \sqrt{\kappa\,N^2 \left(v_{(\cdot)}^{(m-1)}\right)^3 + \left(\frac{v_T}{2}\right)^2} \leq v_{(\cdot)}^{(m-1)}$$

$$\left(v_{(\cdot)}^{(m-1)}\right)^3 - \frac{1}{\kappa\,N^2}\left(v_{(\cdot)}^{(m-1)}\right)^2 + \frac{v_T}{\kappa\,N^2}\,v_{(\cdot)}^{(m-1)} \leq 0$$

$$\Rightarrow \quad v_{(\cdot)}^{(m-1)} \leq 0$$

$$\vee \; \frac{1}{\kappa\,N^2}\left(\frac{1}{4\,\kappa\,N^2} - v_T\right) \geq \left(v_{(\cdot)}^{(m-1)} - \frac{1}{2\,\kappa\,N^2}\right)^2$$

The first inequality applies for negative firing rates. Thus, we ignore this result in the following. The second inequality defines an upper and lower bound on the activity or rather on the input $\Im$:

$$v_{(\cdot)}^{(m-1)} \leq \frac{1}{2\,\kappa\,N^2} + \sqrt{\frac{1}{\kappa\,N^2}\left(\frac{1}{4\,\kappa\,N^2} - v_T\right)} \equiv v^{max} \qquad (10)$$

$$v_{(\cdot)}^{(m-1)} \geq \frac{1}{2\,\kappa\,N^2} - \sqrt{\frac{1}{\kappa\,N^2}\left(\frac{1}{4\,\kappa\,N^2} - v_T\right)} \equiv v^{min} \qquad (11)$$

The two constraints equations (10) and (11), and the nullcline for the weights (equation A1 in Appendix) define the weight-activity phase space (cf. **Figures 4A–C**) as well as the development

FIGURE 3 | A noisy external input leads to propagation of activity along several stages. The first neuron (red) receives a time varying (noisy) input which leads to a strong post-synaptic synapse. Thus, the activity is transmitted to the next stage (green) resulting in high (but smaller than the input) activity and weight. This transmission occurs along several stages until the activity vanishes (gray). Even a short decrease of the input can be compensated quite fast.

of the fixed points of activity and weight along layers or rather stages m (cf. **Figures 4D–E**) in a feed-forward network. The trajectory depends on the terminal firing rate v_T and the number of neurons N (c.f. **Figure 4**).

The phase space consists of two qualitatively different curves. The first class of curves represents the fixed points of each layer (equation 8). These fixed points do not depend on layer number k directly but on the input to this layer which is the output of the preceding layer v_i^{m-1}. In **Figures 4A–C** these fixed points are represented by the solid line. The second class of curves deals with the layer to layer fixed points of our system, i.e., with the previously calculated two constraints equations (10) and (11) and, additionally, equation (A1) for the weights. These constraints define the limits of a region within which activity stays bounded from layer to layer. In **Figures 4A–C** these curves are dashed lines (vertical black lines are v^{min} and v^{max}; the purple line is the nullcline ω^{equ} for $\omega^m = \omega^{m-1}$). The constraints v^{min} and v^{max} divide the phase space into three qualitatively different regimes: if the activity of only a single layer is smaller than v^{min}, the activity of the next layer "jumps" to v^{min} and stays there for the descendant layers. If the activity is larger than v^{max}, activities increases from layer to layer; only if the activity stays between v^{min} and v^{max}, activities decrease from layer to layer and stay bounded (c.f. arrows in **Figure 4A**). Only v^{max} is significantly influenced by N leading to a smaller stable range as more neurons excite one neuron in the next stage.

Now we can understand how fixed points develop from layer to layer and, thus, how large a cell assembly, given input strength $\Im$, gets.

First, we apply an input from the bounded regime that causes activities and weights of the first layer to converge to their fixed points which lie on the blue line (e.g., $k = 0$ in **Figure 4A**). The resulting activity $v^{(0)*}$ is then transmitted to the next layer and causes in turn weights of the next layer to converge. Intuitively, this is visualized by a line going downward until it reaches the weight nullcline ω^{equ}. From here, weights together with the activity of the previous layer once again cause the activity of the next layer ($k = 1$) to converge, too. This is visualized by a line going leftwards to the blue fixed point curve.

This process repeats from layer to layer until the global fixed point at ($v^{(g)} = v^{min}, \omega^{(g)} = 1$) is reached for layer $k = g$. Activities and weights will not change for subsequent layers $k > g$. In other words, the external input cannot influence layers $k > g$.

Thus, in a pure feed-forward network any kind of information storage can only exists for the first k layers since the system reaches its global fixed point for layers beyond k independent of the initial input (as long as the input is in the bounded regime).

3.1.3. Influence of parameters

To assess the capacity of information storage (size of input trace), we need to determine the influence of network parameters on g. It turns out that most influence on g is caused by v_T. **Figures 4D,E** demonstrate activity and weight development for three different values of v_T. Note that although the development moves along discrete layers, we plot continuous lines.

Next, we estimate the "number of jumps" from the distance between weight nullcline w^{equ} and the blue fixed point curve (c.f.

arrows between blue and purple line in **Figure 4A**). A smaller distance (compare **Figures 4B,C**) leads to smaller "jumps" from layer to layer which in turn leads to more layers toward the global fixed point. However, at the same time the difference between v^{min} and v^{max} decreases which leads to a reduced total distance between the input value and v^{min}.

In **Figures 4A1–A6** we also show numerical results of activity and weight development when the input is chosen to be in one of each of the three different regimes; this confirms our analytical observations. For inputs greater than v^{max} the network diverges (**Figures 4A5** and **A6**). For inputs less than v^{min} the network converges quickly to its global fixed point ($v = v_{min}, \omega = 1$; **Figures 4A1** and **A2**). Only for inputs between v^{min} and v^{max} (**Figures 4A3** and **A4**) we notice distinctively different fixed points along several layers until at around layer eight the global fixed point is eventually reached. Thus, such a feed-forward network with inputs between v^{min} and v^{max} stores different values for activities and weights, here up to layer 7. The numerical results of **A3** and **A4** are qualitatively very similar to the analytical curves in **Figures 4D,E**.

We now define the theoretical limits of this system. On the one hand we see from **Figure 4D** that in order to maximize g the distance between v^{max} (equation 10) and v^{min} (equation 11) should be small and v_T should then be $\xi = (4\kappa N^2)^{-1}$:

$$\lim_{v_T \to \xi} v^{max} = \lim_{v_T \to \xi} v^{min} = \frac{1}{2\kappa N^2}$$

Given an input between v^{min} and v^{max}, the system would need $m \to \infty$ to reach v^{min}. However, with $v_T \to (4\kappa N^2)^{-1}$ the interval (v^{min}, v^{max}) from which inputs can be chosen becomes smaller – eventually zero – and the network is specialized to "store" only a few (or one) distinct inputs. Additionally, the difference between input representation (minimum fixed point weight ω^{min}) and global fixed point ($\omega^{(g)} = 1$) also becomes smaller (green curve in **Figures 4D,E**). So a noisy read-out of the stored inputs is complicated as an input difference could easily fall inside the noise level. On the other hand, to get a stable read-out and at the same time to "store" a large range of different inputs, the difference between v^{min} and v^{max} should be maximized. From equations (10) and (11) we find that v_T should be as negative as possible, however, in a pure feed-forward network $v_T < 0$ would lead to a negative firing rate for $v^{(g)} = v^{min} < 0$ which is mathematically possible but biologically implausible. Thus, v_T should minimally go to zero:

$$\lim_{v_T \to 0} v^{max} = \frac{1}{\kappa N^2}$$
$$\lim_{v_T \to 0} v^{min} = 0.$$

To sum up, there is a trade-off, which depends on v_T, between the number of layers representing an input and the stability of this representation against noise. On the one hand, a high v_T ensures an extended representation over many layers, but, on the other hand, this representation is now quite susceptible to noise during the read-out process as fixed points are close to each other, and vice versa for a small v_T.
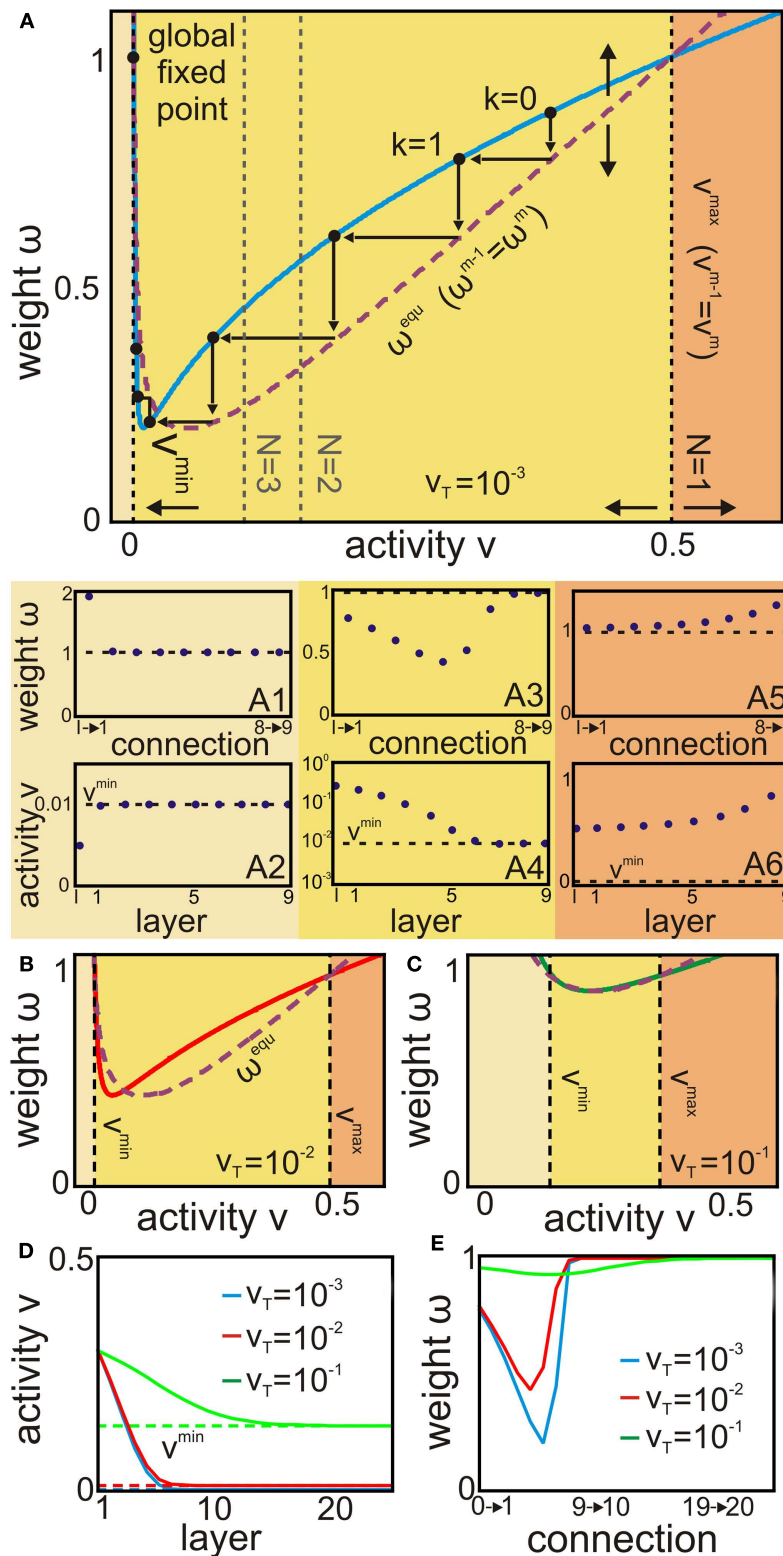
FIGURE 4 | The phase space of a feed-forward network and how the representation of a given input depends on the parameter v_T. (A–C) The constraints v^{min}, v^{max} (vertical black dashed lines), and ω^{equ} (purple dashed line) define the direction of change of weight and activity (arrows) from layer to layer on the fixed point curve (colored continuous line). **(A1–A6)** The sub-panels show the fixed point values of the weights (top) and activities (bottom) of the layers at different regimes (colored according to the phase space). **(D,E)** Resulting from these curves the number of layers m needed to reach the global fixed point (thus, representing an input) varies for different v_T. Dashed lines are the respective v^{min} lines. For more details see main text. Parameters: $k = 2$, $N = 1$; **(A)** $v_T = 10^{-3}$; **(B)** $v_T = 10^{-2}$; **(C)** $v_T = 10^{-1}$ **(D–E)** $\Im = 0.3$.

3.1.4. Signal dispersion in a feed-forward network with local connectivity

Signals and therefore information in feed-forward networks are transmitted from layer to layer. If we demand that there is still information of our initial signal even after many layers, the network has to maintain the spatio-temporal integrity of inputs. Ideally, inputs should not disperse randomly both in space and in time.

We therefore test the capability of a feed-forward network (**Figure 2C**) for signal representation and transportation. The synapses of the network are modified by the SPaSS rule. Constant weights represent the control (to mimic, for instance, weight hard-bounds). In the first of our two scenarios we present a spatially restricted signal (*transport signal*) to the first layer and then measure the dispersion along layers (**Figures 5A,B**). In the second scenario we inject as an input two adjacent signals with a small spatial gap between them (*resolution signal*; **Figures 5C,D**). With both scenarios we are able to analyze the network's capability to differentiate its inputs after a certain number of layers. Different from the previously analyzed network the feed-forward network is constructed as follows: each neuron projects and in turn receives synapses to and from a neuron at the same position and its two neighbors in the next and the former layer.

If the weights are set to a constant value ω_c, the input smears out with every subsequent layer. This is best visible in **Figure 5B** when comparing the activity profile of each layer with the dashed lines of the original signal. Additionally, the amplitude of the signal gets reduced with every subsequent layer. Although smearing is reduced by using smaller constant weights ω_c, this also leads to a stronger reduction of amplitude. The SPaSS rule, on the other hand, can reduce (after weight stabilization) both smearing and amplitude reduction (cf. **Figure 5A**) at the same time. It is interesting to note that only in the first layer the activity of the border neurons is significantly increased. In all subsequent layers, the signal amplitude of these border neurons does not increase but merely follows the general reduction of signal amplitude.

In the second scenario smearing reduces the gap between the two initially separate signals from layer to layer (**Figure 5D**). For constant weights ω_c the signals are difficult to distinguish after the third layer (**Figure 5D**). For plastic synapses the resolution decreases, too, but both signals are still distinguishable until the signal vanishes at layer five (**Figure 5C**). This effect has its origin in the scaling mechanism which tries to balance incoming signals. This leads to a suppression of later signals while at the same time the strong forward pathways are being maintained.

In summary, the SPaSS rule avoids a spreading or rather smearing of spatially restricted signals in deep layers of feed-forward networks. Therefore, the rule ensures that spatially distinct signals presented to a network are still distinguishable after passing several neuronal layers.

3.1.5. Feed-forward networks with excitatory and inhibitory connections

In this section we consider step by step our feed-forward network with -i- lateral excitatory connections within layers (top row in **Figure 6**), -ii- feed-forward inhibition (middle row in **Figure 6**), and -iii- feedback (self-)inhibition (bottom row in **Figure 6**) while all intra-layer connections stay constant. Under these constraints, analytical results can still be obtained and we state only the main equations here. All calculations – similar to those shown above – are provided in the Appendix. In principle we could also do analytics for mixtures of the structures presented in **Figure 6** but calculations get lengthy and opaque then, and the same is true for plastic intra-layer connections.

3.1.5.1. Lateral excitation.
First we introduce lateral excitatory connections with weights ω_R (top row in **Figure 6**). We consider a system in which each neuron $i^{(m)}$ of layer m still receives input from all neurons $j^{(m-1)}$ of the previous layer $m-1$ and, additionally, inputs from R neurons within its own layer m. As all neurons in a layer receive the same input, the activity $v_i^{(m)}$ is the same for all these neurons:
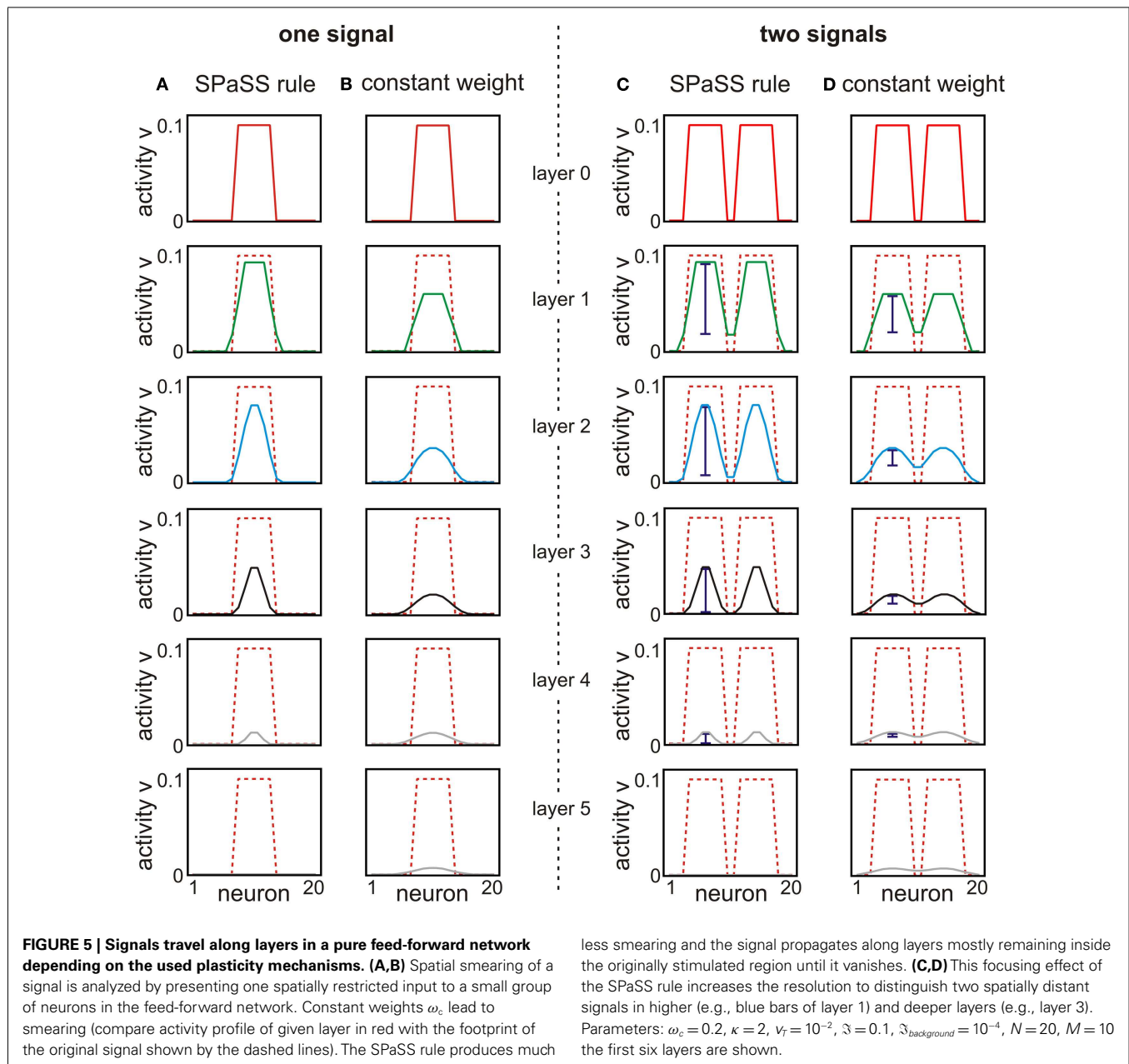
$$v_i^{(m)} = \sum_{j=1}^{N} \omega_{i,j}^{(m)} v_j^{(m-1)} + \sum_{l}^{R} \omega_R v_l^{(m)}$$
$$= N \omega_{(\cdot),(\cdot)}^{(m)} v_{(\cdot)}^{(m-1)} + R \omega_R v_{(\cdot)}^{(m)} \equiv v_{(\cdot)}^{(m)} \qquad (12)$$

If we compare fixed point solutions (equations A2 and A3 in Appendix) and constraints (equations A4–A6 in Appendix) with those of the pure feed-forward network without lateral excitatory connections (equations 8–11; equation A1 in Appendix), we find that only the number of input neurons N is rescaled by the number of lateral excitatory connections R so that we can replace

$$\frac{N}{1 - R \omega_R} = N_{Lat} \ (with \ R \omega_R < 1). \qquad (13)$$

A larger number of lateral excitatory connections R and/or stronger weights ω_R increase the effective number N_{Lat} of input neurons per layer to neuron $i^{(m)}$. Such an increase leads to a strong decrease in v^{max} while the change in v^{min} remains negligible (dashed lines in **Figure 6A**). Thus, for a stable trace the inputs are now restricted to a smaller region. The positions of fixed points within a layer (continuous red lines), however, do not change much compared to the pure feed-forward network (blue line). This finding lets us expect that global convergence is reached earlier. However, this does not apply (see **Figures 6B,C**) as layer to layer fixed points for ω shift (not shown), too, leading to almost the same activity value. Only the weight value of the global fixed point ω_{global} is reduced. This depends on the strength of the lateral excitatory connections ($\omega_{global} \approx 1 - R\omega_R$; see **Figure 6C**, constant parts of the curves). However, more layers are needed to reach this global fixed point (**Figure 6B**).

Thus, with respect to signal dispersion, lateral excitation in feed-forward networks leads to a reduced decay of activity from layer to layer compared to a pure feed-forward structure (**Figure 5**) without effecting the smearing effect much. Thus, as smearing remains limited, distinct signals

FIGURE 5 | Signals travel along layers in a pure feed-forward network depending on the used plasticity mechanisms. (A,B) Spatial smearing of a signal is analyzed by presenting one spatially restricted input to a small group of neurons in the feed-forward network. Constant weights ω_c lead to smearing (compare activity profile of given layer in red with the footprint of the original signal shown by the dashed lines). The SPaSS rule produces much

less smearing and the signal propagates along layers mostly remaining inside the originally stimulated region until it vanishes. **(C,D)** This focusing effect of the SPaSS rule increases the resolution to distinguish two spatially distant signals in higher (e.g., blue bars of layer 1) and deeper layers (e.g., layer 3). Parameters: $\omega_c = 0.2$, $\kappa = 2$, $v_T = 10^{-2}$, $\Im = 0.1$, $\Im_{background} = 10^{-4}$, $N = 20$, $M = 10$ the first six layers are shown.

are transported over more stages when lateral excitation is present.

3.1.5.2. Inhibition. Next, we introduce inhibitory lateral connections with weights ω_I. As we define ω_I as positive, we use a minus sign in all formulas. Inhibitory weights stay constant throughout this study and the derivative of ω_I is set to zero ($\omega_I = $ const.) because plasticity mechanisms for inhibitory synapses are still highly debated (Woodin et al., 2003; Haas et al., 2006; Caporale and Dan, 2008).

We investigate feed-forward and feedback inhibition; both types are equivalent to simpler structures (middle row in **Figure 6**, beneath the blue arrows) that are easier to analyze.

Feed-forward inhibition: Feed-forward inhibition (**Figure 6**, middle row) is an important structure in neuronal networks as, for instance, the ubiquitously present lateral inhibition can be approximated by multiple feed-forward structures.

Feed-forward inhibition (**Figure 6** middle, above blue arrow) consists of three neurons of which one sends signals to the other two neurons. One of these two neurons in turn inhibits the second neuron. This neuron – receiving excitatory and inhibitory inputs – is defined to be the output. We simplify this structure to two neurons (**Figure 6** middle, beneath blue arrow) which is identical to the original structure if the connection from the source neuron to the inhibitory neuron is set to the identity and neuronal transfer functions are, as before, linear. In the simplified structure the source neuron directly sends signals to the output neuron

FIGURE 6 | Inhibition and recurrent connections influence fixed points of activity and weights from layer to layer. The **(A,D,G)** show the phase space of the modified feed-forward networks (blue lines show the control from **Figure 4A**). Different structures results in different activity **(B,E,H)** and weight **(C,F,I)** development over stages m. **(A–C)** In the feed-forward network there are constant excitatory recurrent connections of weight ω_R within each layer. These recurrences shift the constraints v^{min} and v^{max} of the fixed points and narrow the regime of possible inputs. However, the length of a trace and its difference are larger for stronger recurrences ($R\omega_R$). **(D–F)** Fixed points with feed-forward inhibition **(G–I)**. As **(A–C)**, however, with inhibitory recurrent connections. Inhibition (feed-forward and feedback) enlarges the regime of inputs resulting in stable development over layers. However, length and difference of the traces are shortened. Parameters: $\kappa = 2$, $v_T = 10^{-2}$, $N = 1$, and $R\omega_{R/I} = 0, 0.1, \ldots, 0.5$. For **(B,C,E,F,H,I)** $\Im = 0.2$.

through a plastic excitatory synapse ω_E and an inhibitory synapse with constant weight ω_I^{ff}.

The activity of the output neuron is then simply the difference between both weights multiplied with the activity of the source neuron u:

$$
\begin{aligned}
v &= u\,\omega_E - u\,\omega_I^{ff} \\
&= u\left(\omega_E - \omega_I^{ff}\right).
\end{aligned}
\tag{14}
$$

The results for fixed points and constraints (equations A8–A12 in Appendix) show that feed-forward inhibition increases the region of a stable input trace defined by the difference between v^{max} and v^{min} (**Figure 6D**). The length of the input trace (cell assembly), however, decreases (**Figures 6E,F**). If we compare this behavior with our pure feed-forward network, we find that feed-forward inhibition influences the dynamics similar to a change in v_T toward smaller values (**Figure 4**). The global stable fixed point weight, however, increases with higher inhibitory weights ω_I for feed-forward inhibition (**Figure 6F**, constant parts of the curves) but not

when reducing v_T; then it always stays at a value of 1 (**Figure 4E**).

In summary, feed-forward inhibition in a feed-forward network effects the fixed point structure in a similar way as a reduced v_T in a pure feed-forward network.

Feedback inhibition: Conceptually different to feed-forward inhibition is feedback inhibition (**Figure 6**, bottom). Here, a neuron inhibits itself via a second neuron (**Figure 6** bottom, above blue arrow). The second neuron receives excitatory input from – and in turn inhibits – the first neuron which receives inputs from preceding layers and sends signals to subsequent layers. We simplify this structure, similar to the feed-forward inhibitory structure, by setting the excitatory weight to one and, as we do anyhow by (again) using linear transfer functions for all neurons (**Figure 6** bottom, beneath blue arrow). Thus, we have one neuron left which inhibits itself with constant weight ω_I^{fb}; only the weight ω_E is plastic. Therefore, the output activity v of the neuron is:

$$
v = u\,\omega_E - v\,\omega_I^{fb}
\tag{15}
$$

$$
\Leftrightarrow \quad v = \frac{u\,\omega_E}{1 + \omega_I^{fb}}.
\tag{16}
$$

The equations for fixed points and constraints (equations A16–A18 in Appendix) determine the phase space of such a system (**Figure 6G**). The influence of feedback inhibition mainly shifts ν^{max} to higher activity values – similar to feed-forward inhibition. This again enlarges the regime of inputs that results in a stable input trace. However, such a representation needs less layers (**Figure 6H**) than in a pure feed-forward network.

In summary, in a feed-forward network feedback inhibition has similar effects as feed-forward inhibition but already smaller inhibitory weights ω_I yield the same results. Another difference is that the fixed point curve (equation A18 in Appendix) shifts less than for feed-forward inhibition and the dynamics of a network with feedback inhibition are more similar to the dynamics of a pure feed-forward network.

3.2. RECURRENT NETWORK STRUCTURES

Until now we looked only at feed-forward structures but recurrent structures are as important, for instance, to understand the dynamics of randomly connected networks. For the rest of this study we will focus on recurrent network structures. We start with presumably the simplest recurrent system: a self-connected neuron (**Figure 2B**) and then move on to slightly more complex ring-like structures (**Figures 2C,D**).

Two main results emerge from this investigation. First, the resulting activities ν in different network structures after stabilization are ordered as follows: $self-connected > bi-directional > three-ring > N-ring > feed-forward$. Intuitively we need to consider that recurrent structures lead to a self-enhancement of activity but at the same time the number of neurons that are involved in such a recurrent ring decrease the activity if connections are on average below 1. As a consequence a second, more important observation arises: recurrent connections can be indeed stabilized by the SPaSS rule and if a recurrent structure with only a few number of neurons involved is stable, then all recurrent structures with more neurons involved will be stable, too.

3.2.1. Self-connected neurons

The simplest recurrent network is a single neuron with an input $\Im$ and a self-connecting synapse with weight ω (**Figure 2B**). The activity ν_t at time t of such a neuron is simply the sum of a constant external input $\Im$ and the weighted output of the neuron a time step before. (Instead of one time step a fixed delay term d could be used.)

$$\nu_t = \Im + \omega_t \cdot u_t = \Im + \omega_t \cdot \nu_{t-1} \qquad (17)$$

To obtain stability both variables ν and ω have to be constant over time, i.e., $\dot{\omega} = 0$ and $\nu_t = \nu_{t-1}$. This leads to the following dependency between activity and weight:

$$\nu = \frac{\Im}{1 - \omega}. \qquad (18)$$

We insert this equation into the differential equation for weight development:

$$\dot{\omega} = \mu \left(\frac{\Im^2}{(1-\omega)^2} + \kappa^{-1} \left(\nu_T - \frac{\Im}{1-\omega} \right) \cdot \omega^2 \right) = 0 \qquad (19)$$

and the solutions of equation (19) are

$$\omega^{*1} = \frac{1}{12} \left(-\sqrt{3}\,\Upsilon - \Lambda - 3b \right)$$

$$\omega^{*2} = \frac{1}{12} \left(-\sqrt{3}\,\Upsilon + \Lambda - 3b \right)$$

$$\omega^{*3} = \frac{1}{12} \left(\sqrt{3}\,\Upsilon - \Lambda - 3b \right)$$

$$\omega^{*4} = \frac{1}{12} \left(\sqrt{3}\,\Upsilon + \Lambda - 3b \right)$$

(20)

with

$$\Upsilon = \sqrt{3b^2 + 8ac + 2^{\frac{5}{2}} a \left(c^2 + 12ad\right) \Gamma^{-1} + 2^{\frac{4}{3}} a \Gamma},$$

$$\Lambda = \sqrt{6} \sqrt{\begin{array}{c} 3b^2 + 8ac - 2^{\frac{4}{3}} a \left(c^2 + 12ad\right)\Gamma^{-1} - 2^{\frac{2}{3}} a\Gamma \\ + 3^{\frac{3}{2}} \left(b^3 + 4abc\right) \Upsilon^{-1} \end{array}},$$

$$\Gamma = \left(-2c^3 + 27b^2 d + 72acd \right.$$
$$\left. + \sqrt{-4\left(c^2 + 12ad\right)^3 + \left(-2c^3 + 27b^2 d + 72acd\right)^2} \right)^{1/3},$$

$a = \nu_T$, $b = \Im - 2\nu_T$, $c = \Im - \nu_T$, and $d = \kappa\,\Im^2$.

These solutions have to be real (complex part equals zero) and the derivative of equation (19) at these solutions has to be smaller than zero in order to stabilize the dynamics, so parameters $\Im$, κ, and ν_T have to be restricted. The calculations for assessing these parameters are complex and closed-forms can not be derived anymore. Therefore, the stability for a given set of parameters has to be assessed by numerical calculations of the roots. To reduce the number of parameters only the ratio of the rates (e.g., equations 8–22) $\kappa = \mu/\gamma$ is considered and we plot the maximally allowed input $\Im^{max}$ that still leads to a stable weight. All smaller inputs $(0 < \Im < \Im^{max})$ lead to a stable weight ω, too, and inputs above this value $(\Im > \Im^{max})$ lead to divergent weight dynamics. Thus, we show the parameters which lead to stable dynamics in a κ-ν_T-plot with $\Im^{max}$ color coded (**Figure 7A**). The self-connected neuron stabilizes best if γ is as large as μ ($\kappa \approx 1$) but for the more realistic ratios of one order of magnitude difference ($\kappa \approx 10$) stabilization is still possible for a wide regime. Furthermore, stability requires that ν_T is small or even negative. Here we stress that (1) negative ν_T does not imply negative neuronal firing rates as these systems balance plasticity and scaling ($\nu_i^* \neq \nu_T$; see equation 3 and Tetzlaff et al., 2011) and (2) negative ν_T values can be avoided by simply adding inhibition to any of these recurrent networks (**Figures 2B–D**). As shown above, constant weight inhibition (**Figures 6D–I**) leads to a shift of the network properties toward larger allowed positive values. The general disk-like shape of the stability plot, however, does not change with inhibition added and we will thus continue to consider only excitatory recurrent networks (although ν_T might be negative).

We have confirmed our numerical calculations with simulations up to four places after the decimal point for weight ω and activity ν. (For example, for $\kappa = 2$, $\Im = 0.065$, and $\nu_T = 0.01$ the

weight and the activity in the simulation as well as in the numerics is $\omega = 0.5674$ and $\nu = 0.1503$ respectively).

We now compare the activity of a neuron with a self-connection ν_{sc} with the activity ν_{ff} of a neuron in a feed-forward structure receiving the same input $\Im$ if equation (19) is simplified in the following way (see also Appendix): as ω is smaller than one, weight-dependent terms of order three ($\Theta(\omega^3)$) and higher can be ignored. This leads to following neuronal activity:

$$\nu \approx \frac{\Im}{1 - \sqrt{\frac{\kappa \Im^2}{\Im - \nu_T}}} \quad (21)$$

which is larger than $\nu_{ff} = \Im$.

In summary, self-connected structures in a neuronal network will raise the risk of instability. As this instability depends on the input (or rather the activity given to the self-connected neuron) this risk will decrease if a self-connected structure is in deeper layers (because activity has already declined there). However, as the resulting fixed point activity of self-connected neurons is larger than the activity of a neuron in a feed-forward structure, self-connections will extend input trace into deeper layers.

3.2.2. Simple recurrent ring networks: bi-directional and three-ring

In recurrent networks there are also other types of recurrent structures than simple self-connected neurons. Thus, in the following we derive the influence of larger recurrences on the stability and activity of neuronal networks. Equations (A19) and (A20) in Appendix show that also for the bi-directional connection the activity is increased compared to the feed-forward networks. This effect could be expected as already constant recurrences enhance network activity, but the SPaSS rule sustains the enhancement and still guarantees stable weights. Interestingly, this effect does not depend on the position of the bi-directional structure within the feed-forward network. The second neuron can be part of the feed-forward network but it can also be outside of the layered structure.

Additionally, we numerically calculated the stable regime for different parameter values and plot the results similar to the self-connected neuron. This analysis (**Figure 7B**) confirms the result that stability is independent of the chosen parameters ν_T and κ. However, we observe that the maximally allowed input $\Im^{max}$ that

still leads to convergence *increases* compared to the self-connected neuron (mind the changed color scale bars in **Figures 7A,B**).

For the three-neuron ring an enhancement of these observations is observed. Its equations (equation A23 in Appendix) again show that for the three-ring the activity is increased, but less, as compared to the feed-forward networks.

Furthermore, we find that the stability range for the three-ring (**Figure 7C**) looks similar to the one discussed above (**Figures 7A,B**) and the maximally allowed input $\Im^{max}$ has further increased.

These considerations can be generalized in a straight-forward way for an N-ring, too (not shown).

To conclude, the results presented above confirm what we stated at the beginning of this section. First, if the smallest existing ring structure in a randomly wired recurrent network is stable, then every longer ring will be stable, too. Second, there is an ordering of the resulting activities (see Appendix):

w.l.o.g. $\nu_T \to 0$

$$\nu_{sc} \approx \frac{\Im}{1 - \sqrt{\kappa \Im}} > \nu_{bi} \approx \frac{\Im}{1 - \omega_{1,2}\sqrt{\kappa \Im}} >$$

$$\nu_{nr} \approx \frac{\Im}{1 - \omega_{1,N} \prod_{i=3}^{N} \omega_{i,i-1} \cdot \sqrt{\kappa \Im}} > \nu_{ff} = \Im,$$

with $\omega_{i,j} < 1 \, \forall i, j \in N$

($self-connected > bi-directional > N-ring > feed-forward$). Furthermore (similar to the direct self-connection discussed above) if one neuron connects recurrently *via a ring to itself*, its fixed point activity is enlarged. Thus, given a feed-forward network, this fact ensures that also all following (down stream) layers will have an increased activity and, therefore, more layers are needed to reach the global fixed point. This effect is in general even stronger with the SPaSS rule than with constant weights (c.f. section 5) as the denominator of the activity depends on the input $\Im$. Thus, the length of the input trace is increased with any excitatory recurrent structure.

4. DISCUSSION

Hebb (1949) proposed that a group of interlinked neurons forms a "cell assembly." In this study we show that the combination of synaptic scaling and hebbian plasticity can lead to such "cell assemblies."
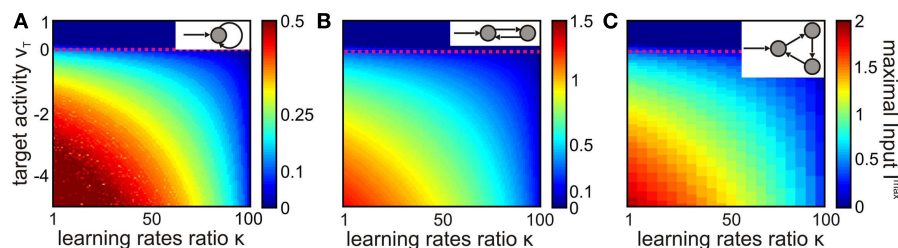
FIGURE 7 | Different recurrent structures are stabilized by SPaSS rule. For different parameter values κ and ν_T the maximal input $\Im^{max}$ leading to stable weights is calculated and shown in color code (note the different scale bars). **(A)** Self-connected neurons, **(B)** bi-directional structures, and **(C)** three-neuron rings are analyzed. If one structure is stabilized for a given parameter set (κ, ν_T) all other structures are stable, too. Only the maximal input decreases with more direct (less neurons) recurrences.

Synaptic scaling is a slow biological process which interacts with faster conventional synaptic plasticity (Abbott and Nelson, 2000; Turrigiano and Nelson, 2000). In a previous study (Tetzlaff et al., 2011) we showed that the (mathematical) combination of both processes leads to stable and reasonable weight growth. In the current study we have extended the analysis of this combined plasticity and scaling (SPaSS) rule to different types of feed-forward and recurrent networks. Our main goal was to try to understand why such networks are capable of forming an input trace in a stable way, with stably enhanced weights along several network stages as shown in **Figure 2B**. Thus, such networks are to some degree capable of *learning and storing* a "cell assembly" (Hebb, 1949) along the network layers.

Our main results are: (1) feed-forward networks indeed obtain a stable sequence of fixed points along several network stages, which can still be analytically calculated. This happens for a certain input range $v^{min} < \Im < v^{max}$. If the network has many layers eventually a final stable fixed point is reached for all layers $k > g$, after which no more useful information processing takes place. Lateral connections and inhibition (feed-forward and feedback) can modify this range of information processing, hence, the extend of the input trace (see **Table 1**). For the lower layers one finds that networks with a SPaSS rule reduce signal dispersion as compared to networks with constant connections (see **Figure 5**). (2) As soon as one introduces recurrent connections, calculations become more difficult and most results have to rely on a numerical analysis. Here

we find that small excitatory recurrent structures (self-connection, or small rings) are more "dangerous" than large rings as it is easier to get unstable dynamics from small as compared to large rings. Thus, if one stabilizes the smallest ring in a network, one will find that larger rings are stable, too. However, as the resulting activity from a ring is larger than in a feed-forward structure, recurrences enhance the length of an input trace and, thus, extend the influence of an external input into the deeper layers of the network. This way a larger cell assembly representing this input is formed.

Furthermore, we can take the results obtained in this study and conclude how big a cell assembly representing an external input would be and how to change its size by parameters and connectivity. For instance, a feed-forward network with one neuron per layer (also named transmission line) and recurrent (here constant) connections between layers (comparable to section 5 and 2) would lead to an input trace with stronger weights compared to a pure feed-forward network. If we introduce additionally long-range feed-forward inhibition the input trace can be "cut" after a few stages. Thus, short-range excitatory feedback together with long-range feed-forward inhibition lead to localized cell assemblies "around" the input neuron (compare black curve in **Figure 8** with gray).

Thus, the SPaSS rule naturally leads to spatially restricted formation of input representations. Such a representation might be indicative of a memory process, which is here reflected by the still very simple process of storing an input trace and this way forming a cell assembly. So far it has proven to be difficult to achieve stable representations of spatial-temporal patterns in time-continuous systems (c.f. Morrison et al., 2007 for STDP). In this study we analyzed mainly the spatial dispersion of such an input representation and how this dispersion is influenced by different connectivities. After stabilization removal of the input leads to slow forgetting, which however can be reversed quite fast by presenting the input again (see **Figure 3**). This process of forgetting and re-learning and, furthermore, the temporal aspect of the input representations has to be analyzed in future studies. However, the analysis here suggests that the combination of synaptic plasticity and synaptic scaling might be a good candidate to allow the formation of short term memories (Dudai, 2004) also in larger networks. The aspect of stable learning in attractor neural networks remains largely unresolved (Mongillo et al., 2008; Cutsuridis and Wennekers, 2009; Lansner, 2009) and we would hope that the results presented here will lead onward helping to solve this problem and leading to a better understanding of (short term) memory formation processes.

FIGURE 8 | Connectivity shapes the size of input traces. In a pure feed-forward network (gray) an input leads to an input trace along several stages. The weights (dashed) and activities (solid) decrease from layer to layer (as already explained in section 2). On the one hand, by introducing recurrences the weights have significantly larger fixed points (see black line of the first layers), thus, the cell assembly is more stable against perturbation and more prominent in the network. Long-range inhibition, on the other hand, decreases activities and weights in a way that the cell assembly is "cut" from layer to the next (here, for instance, between 2 and 3). This an example of using the results of this study to define a connectivity which leads to strongly localized cell assemblies. Parameters: $\omega_R = \omega_I^{ff} = 0.1$; others see **Figure 3**.

REFERENCES

Abbott, L. F., and Nelson, S. B. (2000). Synaptic plasticity: taming the beast. *Nat. Neurosci.* 3(Suppl.), 1178–1183.

Bi, G. Q., and Poo, M. M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* 18, 10464–10472.

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

Caporale, N., and Dan, Y. (2008). Spike timing-dependent plasticity: a hebbian learning rule. *Annu. Rev. Neurosci.* 31, 25–46.

Cutsuridis, V., and Wennekers, T. (2009). Hippocampus, microcircuits and associative memory. *Neural Netw.* 22, 1120–1128.

Dudai, Y. (2004). The neurobiology of consolidation, or, how stable is the engram? *Annu. Rev. Psychol.* 55, 51–86.

Gerstner, W., and Kistler, W. M. (2002). Mathematical formulations of Hebbian learning. *Biol. Cybern.* 87, 404–415.

Haas, J., Nowotny, T., and Abarbenel, H. D. I. (2006). Spike-timing-dependent plasticity of inhibitory synapses in the entorhinal cortex. *J. Neurophysiol.* 96, 3305–3313.

Hahnloser, R., Kozhevnikov, A., and Fee, M. (2002). An ultra-sparse code underlies the generation fo neural sequences in a songbird. *Nature* 419, 65–70.

Harris, K., Csiscvari, J., Hirase, H., Dragoi, G., and Buzsaki, G. (2003). Organization of cell assemblies in the hippocampus. *Nature* 424, 552–556.

Hebb, D. O. (1949). *The Organization of Behaviour.* New York: Wiley.

Lansner, A. (2009). Associative memory models: from the cell-assembly theory to biophysically detailed cortex simulations. *Trends Neurosci.* 32, 178–186.

Miller, K. D., and MacKay, D. J. C. (1994). The role of constraints in Hebbian learning. *Neural Comput.* 6, 100–126.

Mongillo, G., Barak, O., and Tsodyks, M. (2008). Synaptic theory of working memory. *Science* 319, 1543–1546.

Morrison, A., Aertsen, A., and Diesmann, M. (2007). Spike-timing-dependent plasticity in balanced random networks. *Neural Comput.* 19, 1437–1467.

Oja, E. (1982). A simplified neuron model as a principal component analyzer. *J. Math. Biol.* 15, 267–273.

Tetzlaff, C., Kolodziejski, C., Timme, M., and Wörgötter, F. (2011). Synaptic scaling in combination with many generic plasticity mechanisms stabilizes circuit connectivity. *Front Comput. Neurosci.* 5:47. doi:10.3389/fncom.2011.00047

Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C., and Nelson, S. B. (1998). Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature* 391, 892–896.

Turrigiano, G. G., and Nelson, S. B. (2000). Hebb and homeostasis in neuronal plasticity. *Curr. Opin. Neurobiol.* 10, 258–364.

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

Woodin, M., Ganguly, K., and Poo, M. M. (2003). Coincident pre- and postsynaptic activity modifies GABAergic synapses by postsynaptic changes in Cl⁻ transporter activity. *Neuron* 39, 807–820.

APPENDIX

NULLCLINE OF WEIGHT DEVELOPMENT OVER STAGES IN FEED-FORWARD NETWORKS:

$$\omega_{(\cdot),(\cdot)}^{(m)} \stackrel{!}{=} \omega_{(\cdot),(\cdot)}^{(m+1)}$$

$$\omega_{(\cdot),(\cdot)}^{(m)} = \frac{v_T}{2\,N\,v_{(\cdot)}^{(m)}} + \sqrt{\kappa\,v_{(\cdot)}^{(m)} + \left(\frac{v_T}{2\,N\,v_{(\cdot)}^{(m)}}\right)^2}$$

$$\omega_{(\cdot),(\cdot)}^{(m)} = \frac{v_T}{2\,N\,v_{(\cdot)}^{(m-1)}\cdot\omega_{(\cdot),(\cdot)}^{(m)}} + \sqrt{\kappa\,v_{(\cdot)}^{(m-1)}\cdot\omega_{(\cdot),(\cdot)}^{(m)} + \left(\frac{v_T}{2\,N\,v_{(\cdot)}^{(m-1)}\cdot\omega_{(\cdot),(\cdot)}^{(m)}}\right)^2}$$

which simplifies to the following nullcline for weights $\omega_{(\cdot),(\cdot)}^{(m)}$:

$$\omega_{(\cdot),(\cdot)}^{(m)} = \frac{\kappa\,N}{2}\,v_{(\cdot)}^{(m-1)} + \sqrt{\frac{v_T}{N^2\,v_{(\cdot)}^{(m-1)}} + \left(\frac{\kappa\,N}{2}\,v_{(\cdot)}^{(m-1)}\right)^2} \equiv \omega^{\mathrm{equ}} \tag{A1}$$

LATERAL EXCITATION IN FEED-FORWARD NETWORKS

As defined in the main text, the activity of neuron $i^{(m)}$ is

$$v_i^{(m)} = \sum_{j=1}^{N} \omega_{i,j}^{(m)}\,v_j^{(m-1)} + \sum_{l}^{R} \omega_R\,v_l^{(m)} = N\,\omega_{(\cdot),(\cdot)}^{(m)}\,v_{(\cdot)}^{(m-1)} + R\,\omega_R\,v_{(\cdot)}^{(m)} \equiv v_{(\cdot)}^{(m)}$$

leading to the stability assumption

$$v_{(\cdot)}^{(m)} = \frac{N}{1 - R\,\omega_R}\,v_{(\cdot)}^{(m-1)}\,\omega_{(\cdot),(\cdot)}^{(m)}.$$

Thus, the differential equation of the weight becomes:

$$\dot\omega_{(\cdot),(\cdot)}^{(m)} = \mu \left(\frac{\left(v_{(\cdot)}^{(m-1)}\right)^2}{1 - R\,\omega_R}\,\omega_{(\cdot),(\cdot)}^{(m)} + \kappa^{-1}\left(v_T - \frac{v_{(\cdot)}^{(m-1)}}{1 - R\,\omega_R}\,\omega_{(\cdot),(\cdot)}^{(m)}\right)\cdot\left(\omega_{(\cdot),(\cdot)}^{(m)}\right)^2 \right) \stackrel{!}{=} 0.$$

The resulting positive stable weight is

$$\omega_{(\cdot),(\cdot)}^{(m)} = \frac{v_T}{2\,N\,v_{(\cdot)}^{(m-1)}}\,(1 - R\,\omega_R) + \sqrt{\kappa\,v_{(\cdot)}^{(m-1)} + \left(\frac{v_T}{2\,N\,v_{(\cdot)}^{(m-1)}}\,(1 - R\,\omega_R)\right)^2} \tag{A2}$$

with associated activity

$$v_{(\cdot)}^{(m)} = \frac{v_T}{2} + \sqrt{\frac{\kappa\,N^2}{(1 - R\,\omega_R)^2}\left(v_{(\cdot)}^{(m-1)}\right)^3 + \left(\frac{v_T}{2}\right)^2}. \tag{A3}$$

However, to guarantee bounded development for activities $(v_{(\cdot)}^{k=m-1} \geq v_{(\cdot)}^{k=m})$ of the network two constraints, comparable to the pure feed-forward network (equations 10 and 11), have to be maintained

$$v_{(\cdot)}^{(m-1)} \leq \frac{(1 - R\,\omega_R)^2}{2\,\kappa\,N^2} + \sqrt{\frac{(1 - R\,\omega_R)^2}{\kappa\,N^2}\left(\frac{(1 - R\,\omega_R)^2}{4\,\kappa\,N^2} - v_T\right)} \equiv v_{rec}^{max} \tag{A4}$$

$$v_{(\cdot)}^{(m-1)} \geq \frac{(1 - R\,\omega_R)^2}{2\,\kappa\,N^2} - \sqrt{\frac{(1 - R\,\omega_R)^2}{\kappa\,N^2}\left(\frac{(1 - R\,\omega_R)^2}{4\,\kappa\,N^2} - v_T\right)} \equiv v_{rec}^{min}. \tag{A5}$$

Furthermore, the ω-nullcline is:

$$\omega^{(m)}_{(\cdot),(\cdot)} = \frac{\kappa \, v^{(m-1)}_{(\cdot)}}{2} \, \frac{N}{1 - R\,\omega_R} + \sqrt{\frac{v_T}{N^2 \, v^{(m-1)}_{(\cdot)}} \, (1 - R\,\omega_R)^2 + \left(\frac{\kappa \, v^{(m-1)}_{(\cdot)}}{2} \, \frac{N}{1 - R\,\omega_R} \right)^2} \equiv \omega^{equ}_{rec}. \tag{A6}$$

FEED-FORWARD INHIBITION IN FEED-FORWARD NETWORK

Here, the activity of the output neuron is the difference between both weights multiplied with the activity of the source neuron u:

$$\begin{aligned} v &= u\,\omega_E - u\,\omega_I \\ &= u\,(\omega_E - \omega_I). \end{aligned} \tag{A7}$$

As mentioned in the main text, only the excitatory weight has a non-zero derivative

$$\dot{\omega}_E = \mu \left(u^2 \, (\omega_E - \omega_I) + \kappa^{-1} \, (v_T - u\,(\omega_E - \omega_I)) \cdot \omega_E^2 \right).$$

This equation has only a positive stable fixed point at

$$\omega_E = \frac{v_T + u\,\omega_I}{2\,u} + \sqrt{\kappa \, u + \left(\frac{v_T + u\,\omega_I}{2\,u} \right)^2 - \kappa \, u \, \frac{\omega_I}{\omega_E}}.$$

The term $\frac{\omega_I}{\omega_E}$ is approximately zero as ω_I is defined to be smaller than ω_E:

$$\omega_E = \frac{v_T + u\,\omega_I}{2\,u} + \sqrt{\kappa \, u + \left(\frac{v_T + u\,\omega_I}{2\,u} \right)^2}. \tag{A8}$$

This leads to an output activity of

$$v = \frac{v_T - u\,\omega_I}{2} + \sqrt{\kappa \, u^3 + \left(\frac{v_T + u\,\omega_I}{2} \right)^2} \tag{A9}$$

Also for this system there exists constraints defining if the output v is smaller than the input u:

$$u \leq \frac{1 + \omega_I}{2\,\kappa} + \sqrt{\frac{1 + \omega_I}{\kappa} \left(\frac{1 + \omega_I}{4\,\kappa} - v_T \right)} \equiv v^{max}_{ffi} \tag{A10}$$

$$u \geq \frac{1 + \omega_I}{2\,\kappa} - \sqrt{\frac{1 + \omega_I}{\kappa} \left(\frac{1 + \omega_I}{4\,\kappa} - v_T \right)} \equiv v^{min}_{ffi}. \tag{A11}$$

Additionally, nullclines can be calculated for weights if the system consists of feed-forward inhibition motifs linked together in a feed-forward structure:

$$\omega_E = \frac{\kappa \, u}{2} + \omega_I + \sqrt{\frac{v_T}{u} + \left(\frac{\kappa \, u}{2} \right)^2 - \kappa \, u \, \omega_I} \equiv \omega^{equ}_{ffi}. \tag{A12}$$

Feedback inhibition

The output activity v of the neuron is

$$\begin{aligned} v &= u\,\omega_E - v\,\omega_I \\ \Leftrightarrow v &= \frac{u\,\omega_E}{1 + \omega_I} \end{aligned} \tag{A13}$$

and, thus, the dynamics of the excitatory weight is

$$\dot{\omega}_E = \mu \left(\frac{u^2 \, \omega_E}{1 + \omega_I} + \frac{1}{\kappa} \left(v_T - \frac{u\,\omega_E}{1 + \omega_I} \right) \cdot \omega_E^2 \right). \tag{A14}$$

The positive stable fixed point of this equation is

$$\omega_E = \frac{v_T}{2\,u}\,(1 + \omega_I) + \sqrt{\kappa\,u + \frac{v_T^2}{4\,u^2}(1 + \omega_I)^2} \tag{A15}$$

and the related output activity of this structure is

$$v = \frac{v_T}{2} + \sqrt{\kappa\,\frac{u^3}{(1 + \omega_I)^2} + \frac{v_T^2}{4}}. \tag{A16}$$

The constraints of this system for having a smaller output v than input u are

$$u \le \frac{(1 + \omega_I)^2}{2\,\kappa} + \sqrt{\frac{(1 + \omega_I)^2}{\kappa}\left(\frac{(1 + \omega_I)^2}{4\,\kappa} - v_T\right)} \equiv v_{fbi}^{max} \tag{A17}$$

$$u \ge \frac{(1 + \omega_I)^2}{2\,\kappa} - \sqrt{\frac{(1 + \omega_I)^2}{\kappa}\left(\frac{(1 + \omega_I)^2}{4\,\kappa} - v_T\right)} \equiv v_{fbi}^{min}. \tag{A18}$$

Additionally, for the weight the nullcline is

$$\omega_E = \frac{\kappa\,u}{2\,(1 + \omega_I)} + \sqrt{\frac{v_T}{u}\,(1 + \omega_I)^2 + \left(\frac{\kappa\,u}{2\,(1 + \omega_I)}\right)^2} \equiv \omega_{fbi}^{equ}. \tag{A19}$$

CALCULATIONS FOR BI-DIRECTIONAL AND N-RING RECURRENCES

A bi-directional system can be written as

$$\dot{\omega}_{1,2} = \mu\left(v_1\,v_2 + \kappa^{-1}\,(v_T - v_1)\cdot\omega_{1,2}^2\right)$$
$$\dot{\omega}_{2,1} = \mu\left(v_1\,v_2 + \kappa^{-1}\,(v_T - v_2)\cdot\omega_{2,1}^2\right)$$
$$v_1 = \Im + \omega_{1,2}\cdot v_2$$
$$v_2 = \omega_{2,1}\cdot v_1.$$

For the stability conditions, v_2 is inserted into v_1 and v_1 into v_2 resulting in

$$v_1 = \frac{\Im}{1 - \omega_{1,2}\cdot\omega_{2,1}} \tag{A20}$$

$$v_2 = \frac{\omega_{2,1}\cdot\Im}{1 - \omega_{1,2}\cdot\omega_{2,1}}. \tag{A21}$$

This leads to the following weight equations

$$\dot{\omega}_{1,2} = \mu\left(\frac{\Im^2\cdot\omega_{2,1}}{\left(1 - \omega_{1,2}\cdot\omega_{2,1}\right)^2} + \kappa^{-1}\left(v_T - \frac{\Im}{1 - \omega_{1,2}\cdot\omega_{2,1}}\right)\cdot\omega_{1,2}^2\right) \overset{!}{=} 0$$

$$\dot{\omega}_{2,1} = \mu\left(\frac{\Im^2\cdot\omega_{2,1}}{\left(1 - \omega_{1,2}\cdot\omega_{2,1}\right)^2} + \kappa^{-1}\left(v_T - \frac{\Im\cdot\omega_{2,1}}{1 - \omega_{1,2}\cdot\omega_{2,1}}\right)\cdot\omega_{2,1}^2\right) \overset{!}{=} 0.$$

Although these equations can be solved, the solutions are complex expressions for $\omega_{1,2}$ and $\omega_{2,1}$ which are hard to interpret. However, the activity $v1$ of neuron 1 is larger than in a feed-forward network with the same input. On the other hand, given parameters $\mu = 0.01$, $\gamma = 0.005$, $\Im = 0.065$, and $v_T = 0.01$ the activity $v_1 = 0.0746 > \Im$ is still smaller than for the self-connected neuron ($v = 0.1503$). The activity of neuron 2 for these parameters is also larger than the activity of the second neuron in the feed-forward network ($v_2 = 0.0343 > 0.0290 = v^{k=1}$). This can also be shown by an approximation of $\dot{\omega}_{2,1} = 0$, thereby neglecting all terms of order $\theta(\omega_i^3)$ and higher as they are small ($\omega_i < 1$) anyhow:

$$\frac{\kappa\,\Im^2\cdot\omega_{2,1}}{\left(1 - \omega_{1,2}\cdot\omega_{2,1}\right)^2} + \left(v_T - \frac{\Im\cdot\omega_{2,1}}{1 - \omega_{1,2}\cdot\omega_{2,1}}\right)\cdot\omega_{2,1}^2 = 0$$

$$\Leftrightarrow (v_T\,\omega_{1,2} + \Im)\cdot\omega_{1,2}\,\omega_{2,1}^3 - (2\,v_T\,\omega_{1,2} + \Im)\cdot\omega_{2,1}^2 + v_T\,\omega_{2,1} + \kappa\,\Im^2 = 0$$

$$\text{without } \theta(\omega_i^3) \Rightarrow -\Im\,\omega_{2,1}^2 + v_T\,\omega_{2,1} + \kappa\,\Im^2 = 0.$$

The only stable weight from this approximation that is positive is equal to the feed-forward weight ωff (equation 1) with input $\Im$ and $N = 1$ (equation 8). Thus, the activity v_2 of neuron 2 is approximately

$$v_2 = \frac{\omega_{2,1} \cdot \Im}{1 - \omega_{1,2} \cdot \omega_{2,1}} \approx \frac{\omega^{ff} \cdot \Im}{1 - \omega_{1,2} \cdot \omega^{ff}} = \frac{v^{k=1}}{1 - \omega_{1,2} \cdot \omega^{ff}} > v^{k=1}, \quad \text{as } \omega_{1,2}, \, \omega^{ff} < 1. \tag{A22}$$

Furthermore, the activity of the first neuron receiving external input is approximately

$$v_1 = v_{bi} = \frac{\Im}{1 - \omega_{1,2} \, \omega_{2,1}} \approx \frac{1}{1 - \omega_{1,2} \, \omega^{ff}}. \tag{A23}$$

Therefore, bi-directional connections increase the activity compared to feed-forward networks. As stated in the main text, this fact does not depend on the position of the bi-directional structure within the feed-forward network. The second neuron can be part of the feed-forward network, but it can also be outside of the layered structure.

For an N-ring with N neurons the stability conditions for the activities are

$$v_1^N = \frac{\Im}{1 - \omega_{1,N} \, \Pi_{i=2}^{N} \omega_{i,i-1}}$$

$$v_2^N = \frac{\omega_{2,1} \, \Im}{1 - \omega_{1,N} \, \Pi_{i=2}^{N} \omega_{i,i-1}}$$

$$...$$

$$v_N^N = \frac{\Pi_{i=2}^{N} \omega_{i,i-1} \, \Im}{1 - \omega_{1,N} \, \Pi_{i=2}^{N} \omega_{i,i-1}}.$$

The weight $\omega_{2,1}$ can be again approximated by ω^{ff} if weight-dependencies of order three or even higher are neglected. Thus, the activity of neuron 1 is approximately

$$v_1^N = v_{nr} \approx \frac{\Im}{1 - \omega_{1,N} \, \Pi_{i=3}^{N} \omega_{i,i-1} \cdot \omega^{ff}}. \tag{A24}$$

As all weights should be smaller than one, v_{nr} is larger than the activity in feed-forward networks. This relation holds independent of the length (number of neurons N) of the ring structure.