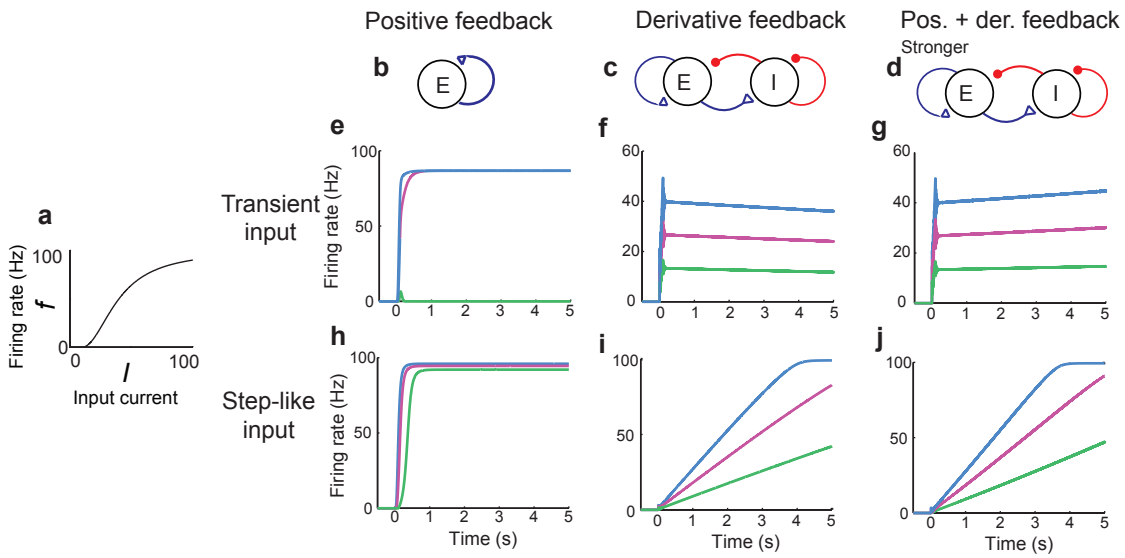
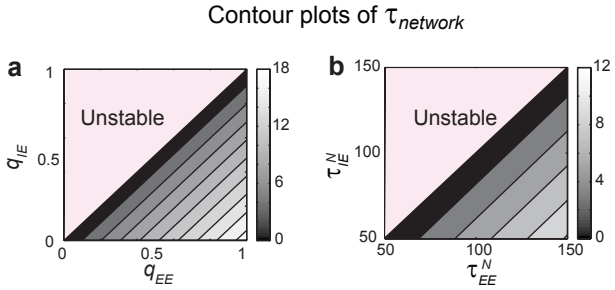
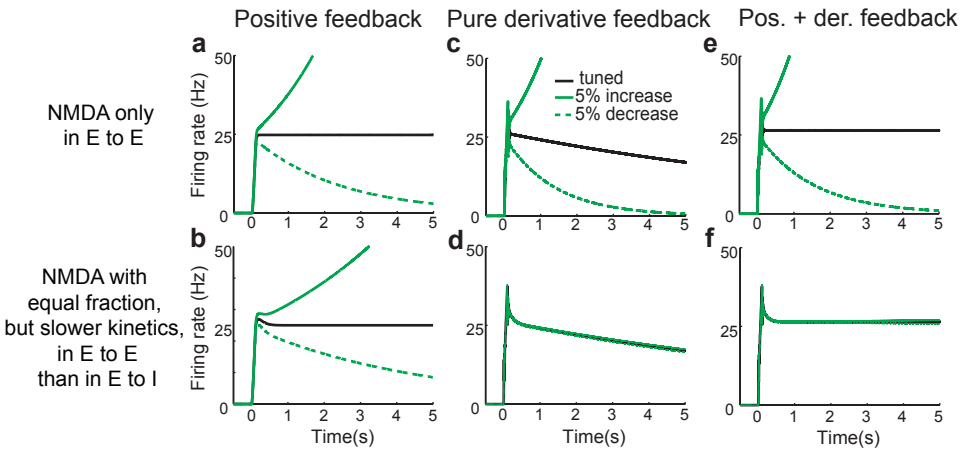


Supplementary Figure S1. Comparison of memory performance in networks with nonlinear neuronal input-output relations.

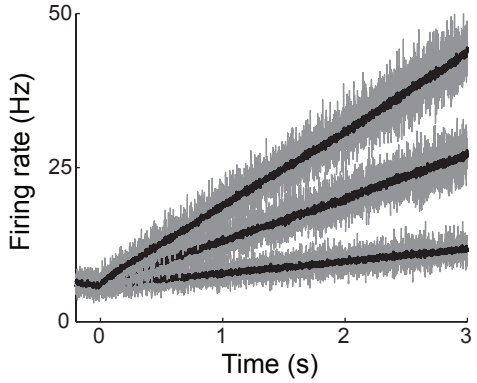
a, Nonlinear firing rate (*f*) vs. input current (*I*) relationship. **b-d,** Network structures of positive feedback **(b)**, negative derivative feedback **(c)**, and hybrid of positive and negative derivative feedback models **(d)**. **e-g,** Activity of the excitatory population in response to transient inputs with different strengths. **h-j,** Activity of the excitatory population in response to step-like inputs with different strengths.



Supplementary Figure S2. Negative derivative feedback networks with a mixture of NMDA and AMPA synapses in all excitatory pathways.
a-b, Time constant of decay of network activity $\tau_{network}$ as a function of the fractions of NMDA synapses for fixed NMDA time constant, τ^N_{EE} =τ^N_{IE} =100 ms (**a**) and as a function of the time constants of the NMDA synapses for fixed NMDA fractions, q_{EE}=q_{IE} =0.5 (**b**). The remaining parameters were J_{EE}=J_{IE} = 150, J_{EI} = J_{II} = 600, τ^A_{EE} =τ^A_{IE} =5 ms, and τ_{EI} =τ_{II} =10 ms

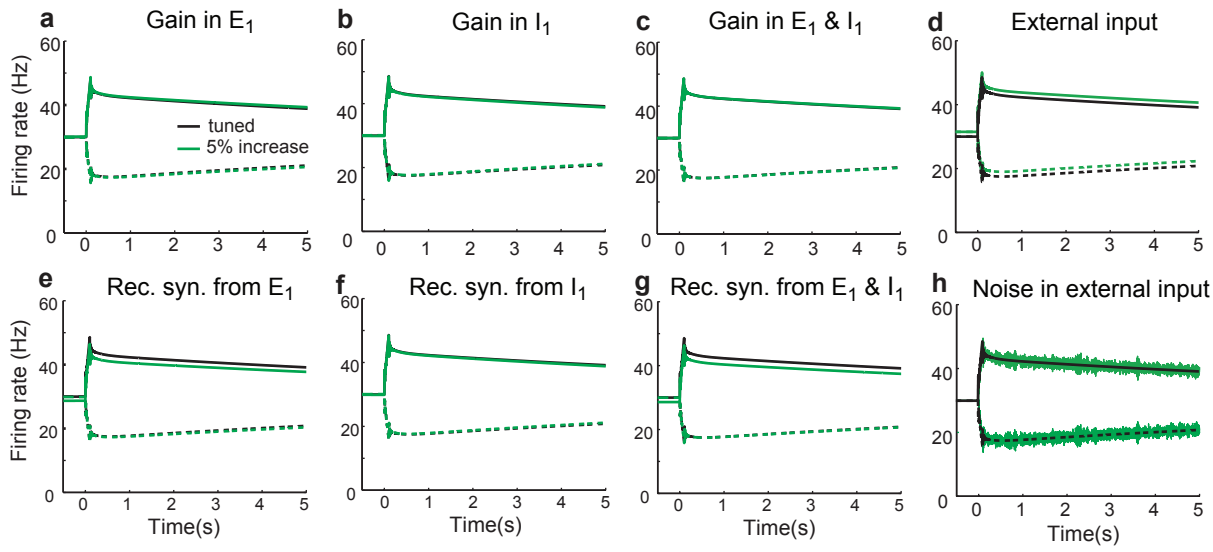


Supplementary Figure S3. Effect of perturbations in NMDA-type receptors.
a-b, Positive feedback networks with only NMDA-mediated excitatory currents (**a**) and with a mixture of NMDA- and AMPA-mediated excitatory currents (**b**). For consistency of the average time constant of excitatory synapses with that in Figure 4, τ_{EE}^{N} =100 ms with q_{EE}=1 in **a,** and τ_{EE}^{N} =150 ms and τ_{EE}^{A} =50 ms with q_{EE}=q_{IE} =0.5 in **b**. **c-f,** Purely derivative feedback networks (**c**,**d**) and hybrid of positive and negative derivative feedback networks (**e**,**f**) with only NMDA-mediated excitatory currents in E to E (**c**,**e**) or with a mixture of NMDA- and AMPA-mediated excitatory currents in E to E and E to I (**d**,**f**). When the fractions of NMDA-mediated currents are equal in E to E and in E to I connections, persistent activity is maintained following perturbations in the negative derivative feedback networks (**d**,**f**), unlike in positive feedback networks. In **c** and **e**, τ_{EE}^{N} =100 ms, τ_{IE}^{A} =25 ms, q_{EE}=1, and q_{IE} =0. In **d** and **f**, τ_{EE}^{N} =150 ms, τ_{EE}^{A} =50 ms, τ_{IE}^{N} =30 ms, τ_{IE}^{A} =20 ms, q_{EE}=q_{IE} =0.5, and the remaining parameters are the same as in Figure 4.



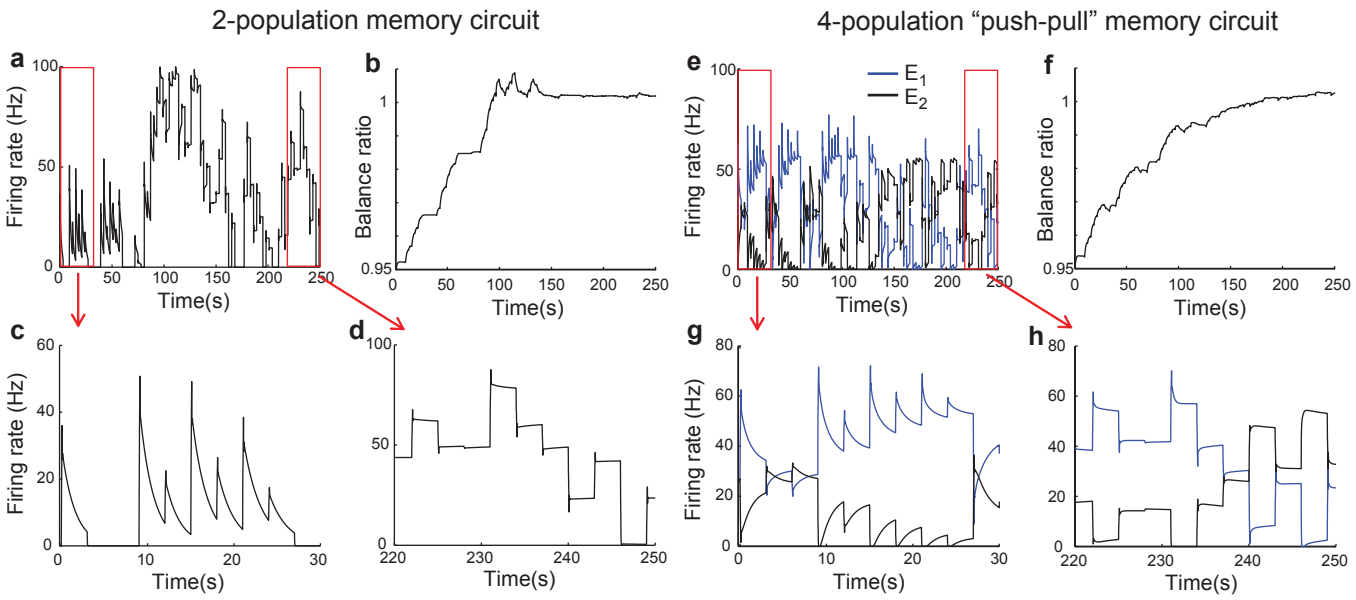
Supplementary Figure S4. Integration of inputs in spiking networks with negative derivative feedback.

Activity of the excitatory population in response to step-like inputs. The 3 traces show the responses to inputs with 3 different strengths. The network structure is the same as in Fig. 5. Instantaneous, population-averaged activity of the excitatory neurons was computed within time bins of 1 ms (gray) or 10 ms (black).



Supplementary Figure S5. Robust memory performance in networks of two competing populations with negative derivative feedback.

a-g, Firing rates of E_1 (solid) and E_2 (dashed) populations with 5% increase in intrinsic gains of E_1 (**a**), I_1 (**b**), or both E_1 and I_1 (**c**), and with 5% increase in the strengths of the external inputs to E_1 and E_2 (**d**), or the recurrent synapses from E_1 (**e**), I_1 (**f**), or both E_1 and I_1 (**g**). **h**, Firing rates of E_1 and E_2 populations with Gaussian white noise presented with the stimulus onset in the external inputs to E_1 and E_2.

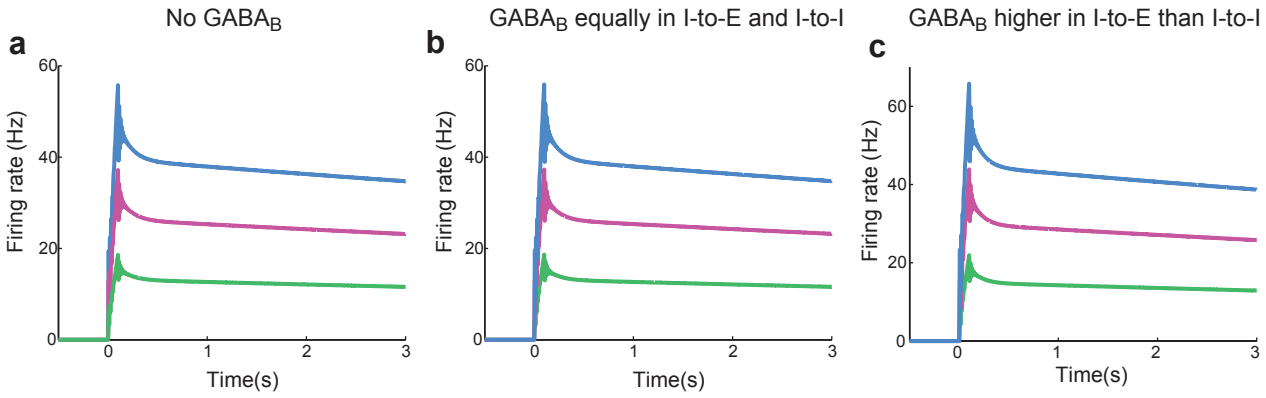


Supplementary Figure S6. Plasticity rule that recovers persistent activity and the balance condition in negative derivative feedback networks.

Illustration that the balance condition and persistent activity in negative derivative feedback networks can be obtained through a differential Hebbian learning rule in the recurrent synapses onto excitatory neurons. We consider a learning rule for stabilizing persistent activity adapted from that of Xie and Seung [53] and having the form, $\tau_{learning}\, dw_{ij}/dt = c_{ij}\, dr_i/dt * r_j$ for $|dr_i/dt| < K$ and $\tau_{learning}\, dw_{ij}/dt = c_{ij}\, K \mathrm{sgn}(dr_i/dt) * r_j$ for $|dr_i/dt| > K$, where $\mathrm{sgn}(x) = x/|x|$ gives the sign of x and K gives the maximum amplitude derivative that can be sensed by the learning mechanism. As shown in Xie and Seung [53], this form can be derived from a spike-timing dependent plasticity (STDP) rule in the limit that firing rates vary much more slowly than the width of the STDP window. Extending that work, we consider plasticity both in excitatory and inhibitory synapses onto the excitatory neurons, with anti-Hebbian plasticity in the excitatory synapses ($c_{EE} = -1$) and Hebbian plasticity in the inhibitory synapses ($c_{EI} = 1$). Plasticity in either E-to-E or I-to-E synapses alone produced similar results (data not shown).

(**a**-**h**) Recovery of persistent activity and the balance condition in circuits with the structure of the two-population memory circuit of Fig. 2 (**a**-**d**) or the four-population push-pull circuit of Fig. 6 (**e**-**h**). In each network, the initial strength of the E-to-E connections is decreased 5% from perfect tuning, resulting in a balance ratio of 0.95 (**b**,**f**) and activity decaying rapidly to a baseline (**c**,**g**). As the balance condition recovers to nearly perfect tuning (**b**,**f**), the time constant of activity decay gets longer (**a**,**e**) until persistent activity is maintained nearly perfectly (**d**,**h**). Simulations shown used $\tau_{learning}$=10s and K = 10Hz/s. The equations and parameters for the firing rate models were the same as those for Fig. 2 or Fig. 6, but with nonlinear firing rate vs. input current relationship as in Fig. 2c,d (bottom) to prevent negative firing rates. External inputs were presented every 3 seconds with strengths chosen independently and randomly from a uniform distribution.

Supplementary Figure S6



Supplementary Figure S7. Negative derivative feedback networks with or without slow GABA$_B$-type inhibitory currents.

a, Activity of the control network of Figures 3 and 4, in which neurons receive a mixture of NMDA- and AMPA-mediated excitatory currents, and only fast GABA$_A$-type inhibitory synaptic currents. The three different traces represent the response to three different amplitudes of transient input, as in Figure 2c. **b**-**c,** Activities of networks receiving the same mixture of NMDA- and AMPA-mediated excitatory currents, but with a mixture of fast GABA$_A$-type and slow GABA$_B$-type synaptic currents. Even in the presence of slow inhibitory current, when GABA$_B$-type synaptic currents are equally present in the I-to-E and I-to-I connections, the time constant of decay of network activity is unchanged. This is because the network decay time constant depends only upon the difference in the average time constants $\mathrm{aver}(\tau_{EI}) = q_{EI}\tau_{EI}^{GB} + (1-q_{EI})\tau_{EI}^{GA}$ and $\mathrm{aver}(\tau_{II}) = q_{II}\tau_{II}^{GB} + (1-q_{II})\tau_{II}^{GA}$, and this difference remains zero. Here, the superscripts GA and GB denote the fast (GABA$_A$) and slow (GABA$_B$) components and q_{EI} and q_{II} denote the proportion of GABA$_B$ currents (**a**,**b**). Even if the fraction of GABA$_B$-type synaptic currents is somewhat higher in the I-to-E connection (50% higher in panel **c**), negative derivative feedback still arises due to the slower and more NMDA-dominant composition of receptors in the E-to-E connection (**c**). For the simulations shown here, $J_{EE} = J_{IE} = 150$, $J_{EI} = J_{II} = 300$, and the time constants and fractions of NMDA-mediated synaptic currents were the same as in Figure 4. In **b** and **c**, the time constant of GABA$_B$ synaptic currents was 100ms and the fractions of GABA$_B$ were $q_{EI} = q_{II} = 0.2$ (**b**) and $q_{EI} = 0.3$, $q_{II} = 0.2$ (**c**).

Balanced cortical microcircuitry for maintaining short-term memory

Supplementary Modeling

Sukbin Lim, Mark S. Goldman

Table of Contents

1. Analytical description of firing rate model

In this section, we provide the analytical calculations underlying the results on the firing rate model described in the main text, and additionally provide simplified versions of the firing rate model that elucidate the core principles underlying negative derivative feedback networks. Using a control theoretic analysis, we find conditions on the network parameters for the network to generate persistent activity through derivative feedback control. We show that, unlike in previous models based on positive feedback, different temporal dynamics of recurrent excitatory and inhibitory inputs is critical to generating persistent activity through derivative feedback control. Furthermore, we show analytically that persistent firing in these networks is more robust against many natural perturbations than in traditional positive feedback based models.

The structure of this section is as follows. In Section 1.1, we first identify important features for derivative feedback control from a simple reduced-dimensionality firing rate model. In Section 1.2, we analyze the dynamics of the full-dimensional models of recurrently connected excitatory and inhibitory populations used in the main paper and find conditions on the network parameters for the models to generate persistent activity through positive feedback and/or negative-derivative feedback. Section 1.3 derives additional conditions assuring that the non-persistent modes of the derivative-feedback networks are stable. Section 1.4 describes the relationship between the rates of inhibitory and excitatory neurons during persistent firing, as well as the optimal input direction for driving maximal responses in derivative-feedback networks. In Section 1.5, we investigate the robustness of the maintenance of persistent activity against perturbations in the network connectivity parameters J_{ij}. In Section 1.6, we show how generation of persistent activity with negative derivative feedback control can be extended to networks whose neurons have a nonlinear firing rate versus input current relationship.

1.1. Simplified firing rate model illustrating negative derivative feedback

Here, we present a simplified network model that provides mathematical intuition for how derivative feedback can arise in a balanced network. Specifically, we show how derivative-like feedback arises from balanced positive feedback and negative feedback with different kinetics, and we relate the properties of this derivative-like feedback to the strengths and time scales of the positive and negative feedback pathways. The reader is referred to Sections 1.4 and 1.5 of this Supplement for rigorous derivations of the analogous properties in the full network of Fig. 2a.

Consider a population of neurons that receives excitatory and inhibitory recurrent inputs with equal strengths but with different filtering time constants:

$$\begin{aligned} \tau\dot{r} &= -r + Js_E - Js_I + J_O\delta(t) \\ \tau_E\dot{s}_E &= -s_E + r \\ \tau_I\dot{s}_I &= -s_I + r \end{aligned} \quad . \tag{A1}$$

Here, r denotes the firing rate of a neuron with time constant τ, s_E denotes recurrent excitatory synaptic input that conveys positive feedback with time constant τ_E, and s_I similarly denotes recurrent inhibitory synaptic input conveying negative feedback with time constant τ_I. Excitatory and inhibitory synaptic inputs are assigned equal strengths J. External input is modeled as a brief, delta function $\delta(t)$ pulse of input of strength J_O.

The key feature of the above model is that the synaptic inputs conveying positive and negative feedback, Js_E and Js_I, arrive with equal strengths but offset kinetics due to the different time constants τ_E and τ_I. We next show that, due to this balance in strength but difference in kinetics of the individual synaptic inputs, the total recurrent input approximates derivative feedback for the low-frequency responses characteristic of persistent activity. To show this, we recall that the Laplace transform of the time-derivative of a signal, dr(t)/dt, equals $uR(u)$ where $R(u)$ is the Laplace transform of $r(t)$ and u is the complex-valued frequency. Using that the synaptic functions s_E and s_I are exponentially filtered transformations of the firing rate $r(t)$, we obtain that the Laplace transform of the total recurrent input is proportional to

$$\mathcal{L}[s_E - s_I] = \frac{R(u)}{\tau_E u + 1} - \frac{R(u)}{\tau_I u + 1} = -\frac{(\tau_E - \tau_I)u}{(\tau_E u + 1)(\tau_I u + 1)} R(u), \tag{A2}$$

where $R(u)$ is the amplitude of the activity $r(t)$ at frequency u. For low frequencies u, $\mathcal{L}[s_E - s_I] \approx -(\tau_E - \tau_I)uR(u)$, which is a constant multiple of the Laplace transform of the derivative of the activity $uR(u)$. Thus, at the low frequencies characteristic of persistent activity, the difference between s_E and s_I is approximately proportional to the derivative of the activity $-W_{der}\frac{dr}{dt}$ in Eq. (1) in the main text:

$$Js_E - Js_I \approx -J(\tau_E - \tau_I)\frac{dr}{dt} \approx -W_{der}\frac{dr}{dt} \quad \text{for low-frequency } r. \tag{A3}$$

By contrast, at high frequencies (large u), $\mathcal{L}[s_E - s_I] \approx -(\tau_E - \tau_I)u/(\tau_E\tau_I u^2)R(u) \sim -R(u)/u$. This shows that high frequencies are suppressed, rather than differentiated, by the recurrent inputs. As noted in the main text, this may be a useful feature, because high frequencies are often associated with noise and would be amplified by an exact derivative feedback mechanism.

From the simple recurrent network defined in Eq. (A1), we can identify a few important features of negative derivative feedback. First, the time constant of network activity increases with the strength of the recurrent feedback and the difference between the time scales for excitatory and inhibitory feedback. From Eq. (1) in the main text and Eq. (A3), the time constant of decay of the activity is

$$\tau_{eff} \approx W_{der} \approx J(\tau_E - \tau_I) \text{ for large } J. \quad \text{(A4)}$$

Second, although the negative derivative feedback network is resistant against drift of activity in the absence of the external input, external input whose strength is comparable to that of the recurrent inputs results in a significant change of activity. For pulse-like input of strength J_O, as in Eq. (A1), the jump in activity is given by

$$\Delta r \approx J_O / \tau_{eff} \approx J_O / \{J(\tau_E - \tau_I)\}. \quad \text{(A5)}$$

From the equation above, we see that Δr does not approach zero even with large J if the strength of the external input J_O scales similarly to the strength of the recurrent inputs J. Indeed, since J represents the strength of the total (excitatory or inhibitory) recurrent synaptic connections averaged across the population, J should scale with the number of recurrent connections. Similarly, J_O scales with the number of connections onto the memory network from the population transmitting the information about the stimulus. Since the number of external connections scales with the network size in the same way as the number of recurrent connections, J_O should be of the same order as J. Thus, even with large derivative feedback, external inputs can produce large changes in the level of the persistent firing rate. In addition, for networks with separate excitatory and inhibitory populations, proper spatial arrangement of the external inputs can reduce the derivative feedback during stimulus presentation and thereby enhance the effect of the external inputs (see Section 1.4).

Finally, we note that this simple model is robust against uniform changes in neuronal gains or loss of a fraction of the neuronal population, because such changes maintain the balance of excitation and inhibition. However, this simple model is not robust against perturbations in excitatory or inhibitory synapses since these disrupt the balance between excitation and inhibition. This is a critical difference from the full-dimensional models described below, which exhibit robustness against perturbations in recurrent excitatory or inhibitory synapses (see Section 1.5).

1.2. Conditions for generation of persistent activity in full-dimensional models with linear dynamics

In this section and the remainder of Section 1, we analytically derive conditions for producing stable persistent activity in linear networks consisting of one excitatory and one

inhibitory population. Through this analysis, we separately identify parameter regimes for positive feedback control and derivative feedback control, and show that derivative feedback control requires recurrent excitation and inhibition to exhibit a close balance in strength but different temporal dynamics.

MATHEMATICAL CONDITIONS FOR GENERATION OF PERSISTENT ACTIVITY

To analyze the linear network, we use the eigenvector decomposition to decompose the coupled neuronal activities into non-interacting modes (eigenvectors) that can be considered independently [1]. For a linear network obeying the equation $d\vec{y}/dt = \overleftrightarrow{A}\vec{y}$, the right eigenvectors $\vec{q}_i^{\,r}$ and corresponding eigenvalues λ_i of the matrix $\overleftrightarrow{A}$ satisfy the equation $\overleftrightarrow{A}\vec{q}_i^{\,r} = \lambda_i \vec{q}_i^{\,r}$ for each i=1 to n, where n denotes the number of state variables. The decay of each mode is exponential with time constant $\tau_{i,eff} = -1/\lambda_i$.

For the system defined by Eq. (5) in the main text:

$$\begin{aligned} \tau_E \dot{r}_E &= -r_E + f_E(J_{EE}s_{EE} - J_{EI}s_{EI} + J_{EO}i(t)) \\ \tau_I \dot{r}_I &= -r_I + f_I(J_{IE}s_{IE} - J_{II}s_{II} + J_{IO}i(t)) \quad , \\ \tau_{ij}\dot{s}_{ij} &= -s_{ij} + r_j \qquad \text{for } i, j = E, \text{ or } I \end{aligned} \tag{A6}$$

where for the linear case $f_E(x)=f_I(x)=x$, $\vec{y} = (r_E, r_I, s_{EE}, s_{IE}, s_{EI}, s_{II})^{\mathrm{T}}$ and the matrix $\overleftrightarrow{A}$ is given by

$$\overleftrightarrow{A} = \begin{array}{c} \\ -1/\tau_E \\ 0 \\ 1/\tau_{EE} \\ 1/\tau_{IE} \\ 0 \\ 0 \end{array} \begin{array}{c} \begin{array}{cccccc} r_E & r_I & s_{EE} & s_{IE} & s_{EI} & s_{II} \end{array} \\ \begin{bmatrix} 0 & J_{EE}/\tau_E & 0 & -J_{EI}/\tau_E & 0 \\ -1/\tau_I & 0 & J_{IE}/\tau_I & 0 & -J_{II}/\tau_I \\ 0 & -1/\tau_{EE} & 0 & 0 & 0 \\ 0 & 0 & -1/\tau_{IE} & 0 & 0 \\ 1/\tau_{EI} & 0 & 0 & -1/\tau_{EI} & 0 \\ 1/\tau_{II} & 0 & 0 & 0 & -1/\tau_{II} \end{bmatrix} \end{array}. \tag{A7}$$

For persistent firing ($\tau_{i,eff}$ large), the system $\dot{\vec{y}} = \overleftrightarrow{A}\vec{y}$ defined by Eq. (A7) should have at least one eigenvector with its corresponding eigenvalue equal to or close to 0. Below we show two different manners by which one can obtain an eigenvalue equal to or close to 0 in networks of recurrently connected excitatory and inhibitory populations. One case corresponds to positive feedback based models, and the other corresponds to negative derivative feedback based models. In the former case, the recurrent connections in the network mediate positive feedback that precisely offsets the intrinsic leakiness of neurons [2-4], where this leakiness is represented mathematically by the decay terms $-r_E$ and $-r_I$ in equation (A6). In the latter case, the recurrent

feedback may not cancel the intrinsic leakiness precisely; instead, the recurrent connections mediate a balance between large positive and negative feedback that are offset in time, resulting in derivative-like feedback that opposes any drifts in activity. In the following, we identify these two parameter regimes in linear firing rate models.

To find the conditions on network parameters for which the network has an eigenvalue equal to or close to 0, we utilize the characteristic function of the linear system. The characteristic polynomial of a linear system is defined by $\mathrm{char}(x) = \det(\overleftrightarrow{A} - x\overleftrightarrow{I})$ where $\overleftrightarrow{I}$ is the *n*-by-*n* identity matrix. Eigenvalues λ of the system correspond to roots of the characteristic polynomial. In our 6-dimensional network model described by the matrix of Eq. (A7), the characteristic polynomial is given by

$$\begin{aligned}\mathrm{char}(x) &= \det\left(\overleftrightarrow{A} - x\overleftrightarrow{I}\right) \\ &= x^6 + a_5x^5 + a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_6),\end{aligned} \tag{A8}$$

where the coefficients a_i of char(x) are functions of the network parameters J_{ij} and τ_{ij}, with $i,j = E$ or I, and can be expressed in terms of the eigenvalues λ_i.

We examine the conditions for this characteristic polynomial to have roots whose values are 0 or close to 0. In particular, the constant term a_0 of the characteristic polynomial char(x) determines whether char(x) has a zero-valued root, since a_0 is the product of all eigenvalues of $\overleftrightarrow{A}$. However, this condition only determines the parameter sets having a precisely 0 eigenvalue. In the case that there is one eigenvalue λ_1 close to 0 and another eigenvalue whose magnitude is larger than the magnitude of $1/\lambda_1$, the product of all eigenvalues represented by a_0 can be finite.

Additionally, the ratio between the coefficient a_1 of the x term and the constant term a_0 can be used to identify a parameter set which allows the system to have eigenvalue close to 0. This can be shown using the expression for a_i in terms of the eigenvalues:

$$\begin{aligned}a_1 &= (-1)^{n-1}\prod_i \lambda_i\left(\sum_j 1/\lambda_j\right) \\ a_0 &= (-1)^n\prod_i \lambda_i \\ \Rightarrow a_1/a_0 &= -\sum_i 1/\lambda_i.\end{aligned} \tag{A9}$$

If an eigenvalue λ is close to zero, the magnitude of its reciprocal $1/\lambda$ will be large. Thus, if the magnitude of the ratio between a_1 and a_0 is large, there exists at least one eigenvalue close to zero. (Note that this condition is a sufficient but not necessary condition for the existence of an eigenvalue close to 0. In the case that there exist multiple eigenvalues close to zero having different signs, the reciprocal of each eigenvalue can be large but the sum can be finite due to cancellation.)

To find conditions on the network parameters for having an eigenvalue equal to or close to 0, we use the explicit expressions for a_1 and a_0 in terms of J_{ij} and τ_{ij}:

$$a_0 = \frac{J_{EI}J_{IE} - (J_{EE}-1)(J_{II}+1)}{\prod_{i=E,I}\tau_i \prod_{i,j=E,I}\tau_{ij}},$$

$$a_1 = \frac{J_{EI}J_{IE}(\tau_{EE}+\tau_{II}) - (J_{EE}-1)(J_{II}+1)(\tau_{IE}+\tau_{EI}) + (J_{II}+1)(\tau_E+\tau_{EE}) - (J_{EE}-1)(\tau_I+\tau_{II})}{\prod_{i=E,I}\tau_i \prod_{i,j=E,I}\tau_{ij}},$$

$$\begin{aligned} a_1/a_0 &= \frac{J_{EI}J_{IE}(\tau_{EE}+\tau_{II}) - (J_{EE}-1)(J_{II}+1)(\tau_{IE}+\tau_{EI}) + (J_{II}+1)(\tau_E+\tau_{EE}) - (J_{EE}-1)(\tau_I+\tau_{II})}{J_{EI}J_{IE} - (J_{EE}-1)(J_{II}+1)} \\ &= \frac{J_{EI}J_{IE}/(J_{II}+1)(\tau_{EE}+\tau_{II}) - (J_{EE}-1)(\tau_{IE}+\tau_{EI}) + (\tau_E+\tau_{EE}) - (J_{EE}-1)/(J_{II}+1)(\tau_I+\tau_{II})}{J_{EI}J_{IE}/(J_{II}+1) - (J_{EE}-1)}. \end{aligned} \quad \text{(A10)}$$

In the above expression, the ratio between a_1 and a_0 becomes large either when the denominator is small (corresponding to small a_0) or when the numerator is much larger than the denominator if the denominator is not close to zero. Below, we show that the former provides a condition for positive feedback networks, and the latter provides conditions for negative derivative feedback networks.

CONDITIONS FOR POSITIVE FEEDBACK NETWORKS

As described above, one condition that leads to an eigenvalue equal to 0 is to have the term a_0 of the characteristic polynomial of Eq. (A8) equal zero. From the set of equations above, this occurs when

$$J_{EI}J_{IE} - (J_{EE}-1)(J_{II}+1) = 0. \quad \text{(A11)}$$

Biologically, this condition corresponds to the precise cancellation of the intrinsic leakiness of the neurons by network-mediated positive feedback, a mechanism that has been suggested previously to underlie persistent firing [5]. To see how the above equation corresponds to such a mechanism, note that during persistent activity $\dot{\vec{y}} \approx 0$ so that the firing rate of the inhibitory population r_I in Eq. (A6) can be expressed in terms of the firing rate of the excitatory population r_E as $r_I = J_{IE}/(1+J_{II})r_E$. Then, in the equation for r_E, the inhibitory feedback strength through the inhibitory population becomes $J_{EI}J_{IE}/(1+J_{II})$ and the net recurrent feedback strength becomes the difference between the excitatory synaptic strength and the strength of the inhibitory feedback, $J_{EE} - J_{EI}J_{IE}/(1+J_{II})$. The amount of this net recurrent feedback precisely cancels the intrinsic leakiness if $J_{EE} - J_{EI}J_{IE}/(1+J_{II}) = 1$, which is the condition given in Eq.

(A11). Thus, Eq. (A11) corresponds to the condition used by traditional positive feedback models in which excess positive feedback is tuned to offset intrinsic neuronal leakiness.

CONDITIONS FOR NEGATIVE DERIVATIVE FEEDBACK NETWORKS

We next consider the alternative mathematical condition for having an eigenvalue close to 0, i.e. that the ratio between a_1 and a_0 becomes large (even if a_0 itself is not very close to zero). That is, even if $J_{EI}J_{IE}-(J_{EE}-1)(J_{II}+1)$ is not small, the network can have an eigenvalue close to zero if the numerator in Eq. (A10) is relatively large compared to the term in the denominator. Here, we show that this condition leads to two core requirements for negative derivative feedback control: first, a balance between positive and negative feedback in strength and, second, slower positive than negative feedback.

Networks can have an eigenvalue close to 0, that is, large a_1/a_0 in Eq. (A10) with finite $a_0 = J_{EI}J_{IE}-(J_{EE}-1)(J_{II}+1)$ in two ways: either having a large time constant τ (case 1) or, for finite τ, having large J's under special relations between the J's (case 2). In the first case, having long time constants of synapses obviously results in slow dynamics in the system and leads to slow decay of neural activity. Indeed, previous works have suggested that the use of long intrinsic or synaptic time constants may lessen the strictness of the tuning requirement that feedback connections must precisely offset intrinsic neuronal decay processes [6-8]. However, the slowest intrinsic time constant in most models is of order 100ms (e.g., the time constant of NMDA decay kinetics), much shorter than observed memory periods of many seconds.

In the second case, the network can have an eigenvalue close to 0 with finite $J_{EI}J_{IE}-(J_{EE}-1)(J_{II}+1)$ if the numerator is much larger than the denominator in Eq. (A10). As shown next, this can occur when the values of the J's are large. In this case, we can approximate the numerator and the denominator of Eq. (A10) with their leading terms in the J's, $J_{EI}J_{IE}\left(\tau_{EE}+\tau_{II}\right)-J_{EE}J_{II}\left(\tau_{IE}+\tau_{EI}\right)$ and $J_{EI}J_{IE}-J_{EE}J_{II}$, respectively. A sufficient condition for the ratio of these terms to be large is then that:

$$J_{EI}J_{IE}\left(\tau_{EE}+\tau_{II}\right)-J_{EE}J_{II}\left(\tau_{IE}+\tau_{EI}\right)\sim O(J^2),$$

$$J_{EI}J_{IE}-J_{EE}J_{II}\ll O(J^2),$$

or equivalently,

$$\left(\tau_{EE}+\tau_{II}\right)J_{EI}J_{IE}/J_{II}-\left(\tau_{IE}+\tau_{EI}\right)J_{EE}\sim O(J), \tag{A12}$$

$$J_{EI}J_{IE}/J_{II}-J_{EE}\ll O(J) \text{ i.e., } J_{EI}J_{IE}/\left(J_{EE}J_{II}\right)\sim 1. \tag{A13}$$

Networks with parameters satisfying the above conditions operate in a regime that corresponds to maintaining persistent firing through negative derivative feedback. To see this, recall from our discussion of the positive feedback mechanism that $J_{EI}J_{IE}/(1+J_{II})$ represents the strength of inhibitory feedback onto the excitatory population through the inhibitory population, and J_{EE} is the strength of recurrent excitatory feedback onto the excitatory population (Fig. 2a). When the J's are large, $J_{EI}J_{IE}/(1+J_{II}) \sim J_{EI}J_{IE}/J_{II}$; thus, Eq. (A13) implies that the strengths of the two feedbacks are similar, and we refer to this equation as the balance condition.

The second condition, given by Eq. (A12), constrains the time constants of the positive and negative feedback. In Eq. (A12), the time constants multiplying the feedback strengths correspond to the time scales of the positive and negative feedback, that is, $\tau_+ = \tau_{EE} + \tau_{II}$ and $\tau_- = \tau_{EI} + \tau_{IE}$; thus, from Eq. (A12), $\tau_+ \neq \tau_-$ (below, in separate stability analyses in Section 1.3, we will show that τ_+ must be greater than τ_-). Qualitatively, $\tau_{EI} + \tau_{IE}$ approximates the time for signals to traverse the negative feedback loop. Similarly, τ_{EE} is the time constant of the direct positive feedback onto the excitatory population and τ_{II} represents the time constant for indirect positive feedback onto the excitatory population by suppressing the inhibitory population (Fig. 2a).

Note that the conditions corresponding to positive feedback networks and negative derivative feedback networks are not mutually exclusive. If the J's are large, the condition for the positive feedback models given in Eq. (A11) becomes a subset of the balance condition for the negative derivative feedback models described in Eq. (A13). In particular, if the network satisfies both Eq. (A11) and Eq. (A12) for large J's, that is, the amount of large positive feedback is similar to, but slightly larger than that of negative feedback and the time scales of the two feedbacks are different, then the network receives large negative derivative feedback as well as additional positive feedback that precisely cancels off the intrinsic neuronal leakiness. This corresponds to the hybrid positive feedback and negative derivative feedback model of Fig. 1c.

CONNECTION TO PHENOMENOLOGICAL AND SIMPLIFIED FIRING RATE MODELS

We next show how the negative-derivative feedback models described above can be directly connected to the simpler phenomenological model of Eq. (1) of the main text that was defined by overall positive feedback and negative-derivative feedback strengths W_{pos} and W_{der}, respectively. Specifically, for the case that the J's are large as in the negative derivative feedback networks, we express W_{pos} and W_{der} in terms of the synaptic strengths J_{ij} and their time constants τ_{ij} and show that the amount of negative derivative feedback W_{der} is proportional to the product of the synaptic strength scale J and the difference between τ_+ and τ_-.

To derive the expression for the amounts of negative derivative feedback and positive feedback in terms of network parameters, we examine the expression for the longest time constant of decay of network activity $\tau_{network}$. Since $\tau_{i,eff} = -1/\lambda_i$, the longest time constant of decay of network activity $\tau_{network}$ is the reciprocal of the eigenvalue closest to 0 whose expression is given by Eq. (A10). Denoting the balanced amount of positive feedback J_{EE} and negative feedback $J_{EI}J_{IE}/(J_{II}+1)$ as J_{der} and the difference between them, $J_{EE}-J_{EI}J_{IE}/(J_{II}+1)$, as J_{pos}, the approximate $\tau_{network}$ from Eq. (A10) for large J's is

$$\tau_{network} \approx \frac{J_{der}\left(\tau_{+}-\tau_{-}\right)+c_1}{1-J_{pos}}. \qquad \text{(A14)}$$

where c_1 is a constant of order 1. The above expression is analogous to the effective network time constant of the phenomenological model of the main text (Eq. (2) of the main text), $\tau_{eff} = \left(W_{der}+\tau\right)/\left(1-W_{pos}\right)$ where τ was the intrinsic (cellular or synaptic) time constant. Thus, we identify $W_{pos} = J_{pos}$ and $W_{der} \sim J_{der}(\tau_{+}-\tau_{-})$. Thus, the amount of negative derivative feedback increases linearly with the J's and the difference between the time constants of positive and negative feedback, similar to what was found for the simplified firing rate model network of Section 1.1 (Eq. (A4)).

In summary, in this section we found the conditions for persistent firing with positive feedback or derivative feedback control. The derivative feedback models are distinct from the previously studied positive feedback models: they require a close balance between excitation and inhibition (Eq. (3) in the main text) and different kinetics of excitation and inhibition (Eq. (4) in the main text). However, the positive feedback models and derivative feedback models are not mutually exclusive, and we show how a hybrid of the two models can be constructed.

1.3. Stability conditions for the derivative-feedback network

In the previous section, we discussed the conditions for the network parameters to have eigenvalue equal to or close to zero. We found a new parameter regime in which the network uses a derivative-like feedback mechanism to maintain persistent activity. Unlike previously models, the negative derivative feedback mechanism does not require perfect cancellation of intrinsic leakiness by positive feedback. Instead, it requires large positive and negative feedback inputs which balance each other but have different dynamics (Eqs. (A12) and (A13)). Here, we identify additional conditions on the network parameters for the networks to maintain persistent activity without unbounded growth of activity in the non-persistent modes. Specifically, the system requires that all eigenvalues except those close to 0 have a negative real part, and we refer to this as "the stability condition" for the network. In the following, we first show the necessary and sufficient stability conditions for a 4-dimensional reduced network in which the

intrinsic neuronal responses are assumed to be fast. Next, we show necessary stability conditions for the full 6-dimensional system.

STABILITY CONDITION FOR THE 4-DIMENSIONAL SYSTEM

To simplify the analytical calculation of the stability condition, here we assume the dynamics of the firing rates is rapid [9] so that the firing rates instantaneously follow their input in Eq. (A6). That is, τ_E and τ_I are considered small and the dynamics is reduced to the 4-dimensional system:

$$\begin{aligned} r_E &= J_{EE}s_{EE} - J_{EI}s_{EI} + J_{EO}i(t) \\ r_I &= J_{IE}s_{IE} - J_{II}s_{II} + J_{IO}i(t) \quad . \\ \tau_{ij}\dot{s}_{ij} &= -s_{ij} + r_j \text{ for } i, j = E, \text{ or } I \end{aligned} \tag{A15}$$

To determine the signs of the eigenvalues in this 4-dimensional system, we use the well-known stability test for linear systems, the Routh stability criterion [10]. In the Routh stability criterion, the number of positive eigenvalues is determined by examining functions of the coefficients of the characteristic polynomial through the use of a Routh table defined as follows:

$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0$$

$$\Rightarrow \quad \begin{array}{c|cccc} x^n & a_n & a_{n-2} & a_{n-4} & \cdots \\ x^{n-1} & a_{n-1} & a_{n-3} & a_{n-5} & \cdots \\ \cdot & b_1 & b_2 & b_3 & \cdots \\ \cdot & c_1 & c_2 & c_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

with

$$b_1 \equiv \frac{a_{n-1}a_{n-2} - a_n a_{n-3}}{a_{n-1}}, \quad b_2 \equiv \frac{a_{n-1}a_{n-4} - a_n a_{n-5}}{a_{n-1}}, \quad \cdots$$

$$c_1 \equiv \frac{b_1 a_{n-3} - a_{n-1}b_2}{b_1}, \quad c_2 \equiv \frac{b_1 a_{n-5} - a_{n-1}b_3}{b_1}, \quad \cdots.$$

In the table above, the number of roots with positive real parts is equal to the number of changes of sign of the elements of the first column of the Routh table.

The persistent activity networks considered here contain an eigenvalue close to 0 which can have either positive or negative real parts, so the system is marginally stable. Thus, before directly applying the Routh-Hurwitz criterion to the characteristic polynomial given in Eq.(A15), we factor out the root of the characteristic polynomial whose value is close to 0, denoted by λ_1, from the characteristic polynomial:

$$\begin{aligned} & x^4 + a_3 x^3 + a_2 x^2 + a_1 x + a_0 = 0 \\ & \Rightarrow (x - \lambda_1)\left\{x^3 + (a_3 + \lambda_1)x^2 + (a_2 + \lambda_1(a_3 + \lambda_1))x + (a_1 + \lambda_1(a_2 + \lambda_1(a_3 + \lambda_1)))\right\} \\ & \qquad + a_0 + \lambda_1(a_1 + \lambda_1(a_2 + \lambda_1(a_3 + \lambda_1))) = 0. \end{aligned}$$

Here, the remainder $a_0 + \lambda_1(a_1 + \lambda_1(a_2 + \lambda_1(a_3 + \lambda_1)))$ equals 0 since λ_1 is a root of the characteristic polynomial.

In particular, if the system has only one eigenvalue close to 0, from Eq. (A9), λ_1 can be approximated by $-a_0 / a_1$ and the quotient function becomes

$$Q(x) = x^3 + (a_3 - \frac{a_0}{a_1})x^2 + (a_2 - \frac{a_0}{a_1}(a_3 - \frac{a_0}{a_1}))x + (a_1 - \frac{a_0}{a_1}(a_2 - \frac{a_0}{a_1}(a_3 - \frac{a_0}{a_1}))). \qquad \text{(A16)}$$

Thus, we apply the Routh stability criterion to this third order quotient function. Moreover, using that the *J*'s are large, we further approximate the coefficients with the leading terms in *J* as follows:

$$\begin{aligned}
a_3 &= -\frac{J_{EE}-1}{\tau_{EE}} + \frac{1}{\tau_{IE}} + \frac{1}{\tau_{EI}} + \frac{J_{II}+1}{\tau_{II}} \sim -\frac{J_{EE}}{\tau_{EE}} + \frac{J_{II}}{\tau_{II}} \equiv a_{3,approx} \sim O(J) \\
a_2 &= -\frac{(J_{EE}-1)(J_{II}+1)}{\tau_{EE}\tau_{II}} + \frac{J_{IE}J_{EI}+1}{\tau_{IE}\tau_{EI}} + \left(-\frac{J_{EE}-1}{\tau_{EE}} + \frac{J_{II}+1}{\tau_{II}}\right)\cdot\left(\frac{1}{\tau_{IE}} + \frac{1}{\tau_{EI}}\right) \\
&\sim -\frac{J_{EE}J_{II}}{\tau_{EE}\tau_{II}} + \frac{J_{IE}J_{EI}}{\tau_{IE}\tau_{EI}} \equiv a_{2,approx} \sim O(J^2) \\
a_1 &= \left(J_{IE}J_{EI}\left(\tau_{EE}+\tau_{II}\right) - (J_{EE}-1)(J_{II}+1)\left(\tau_{IE}+\tau_{EI}\right) + \left(J_{II}+1\right)\tau_{EE} - \left(J_{EE}-1\right)\tau_{II}\right) / \prod_{i,j=E,I}\tau_{ij} \\
&\sim \left(J_{IE}J_{EI}\left(\tau_{EE}+\tau_{II}\right) - J_{EE}J_{II}\left(\tau_{IE}+\tau_{EI}\right)\right) / \prod_{i,j=E,I}\tau_{ij} \equiv a_{1,approx} \sim O(J^2) \\
a_0 &= (J_{IE}J_{EI} - (J_{EE}-1)(J_{II}+1)) / \prod_{i,j=E,I}\tau_{ij} = (J_{IE}J_{EI} - J_{EE}J_{II} - J_{EE} + J_{II} + 1) / \prod_{i,j=E,I}\tau_{ij} \sim O(J),
\end{aligned}$$

Note that a_0 is at most of order *J*, since in the balance condition in Eq. (A13), we additionally assume that the difference between the strengths of positive and negative feedbacks is of order 1, or equivalently, $J_{IE}J_{EI} - J_{EE}J_{II} \sim O(J)$. Applying the Routh-Hurwitz criterion to these asymptotic expressions for the coefficients of the quotient function *Q(x)* in Eq. (A16), we obtain the stability conditions

$$\begin{aligned}
-\frac{J_{EE}}{\tau_{EE}} + \frac{J_{II}}{\tau_{II}} > 0 \quad &\Leftrightarrow \quad \frac{J_{II}}{\tau_{II}} > \frac{J_{EE}}{\tau_{EE}} \\
-\frac{1}{\tau_{EE}\tau_{II}} + \frac{1}{\tau_{IE}\tau_{EI}} > 0 \quad &\Leftrightarrow \quad \tau_{EE}\tau_{II} > \tau_{IE}\tau_{EI} \quad . \\
\left(\tau_{EE}+\tau_{II}\right) - \left(\tau_{IE}+\tau_{EI}\right) > 0 \quad &\Leftrightarrow \quad \tau_{EE}+\tau_{II} > \tau_{IE}+\tau_{EI}
\end{aligned} \qquad \text{(A17)}$$

The last condition is similar to Eq. (A12), which showed that the time scales for the positive and negative feedback must be different to have stable persistent firing. The stability condition above additionally specifies that the positive feedback should be slower than the

negative feedback. The second condition above is similar to the last condition except that it constrains the product of the time constants. The first condition compares the magnitudes of recurrent excitation and recurrent inhibition; that is, for other non-persistent modes to be stable, the normalized strength of the inhibitory feedback must be larger than that of the excitatory feedback.

STABILITY CONDITION FOR THE 6-DIMENSIONAL SYSTEM

For the full 6-dimensional system given in Eq.(A6), the complete stability conditions also can be calculated by the Routh stability criterion. However, the stability conditions are far more complicated expressions in terms of the network parameters. Here, for ease of interpretation, we instead provide simpler, necessary conditions for stability. These necessary conditions are determined by the sign of the coefficients of the characteristic polynomial. To have a stable system, all eigenvalues must be negative and, correspondingly, the coefficients of the characteristic polynomial must all be positive. However, in our persistent activity network, the leading eigenvalue (the one close to 0) may be slightly positive (corresponding to very slow growth of activity in the persistent mode). Therefore, as in the 4-dimensional system, we factor out the eigenvalue close to 0, denoted by λ_1, and find conditions for all coefficients of the quotient function to be positive. To leading order in the J's, these conditions are given by

$$\begin{aligned}
\frac{J_{II}}{\tau_I \tau_{II}} &> \frac{J_{EE}}{\tau_E \tau_{EE}} \\
\frac{J_{II}}{\tau_I \tau_{II}}\left(\frac{1}{\tau_E}+\frac{1}{\tau_{EI}}+\frac{1}{\tau_{IE}}+\frac{1}{\tau_{EE}}\right) &> \frac{J_{EE}}{\tau_E \tau_{EE}}\left(\frac{1}{\tau_I}+\frac{1}{\tau_{EI}}+\frac{1}{\tau_{IE}}+\frac{1}{\tau_{II}}\right) \\
\tau_{EE}\tau_{II} &> \tau_{EI}\tau_{IE} \\
\tau_{EE}+\tau_{II} &> \tau_{EI}+\tau_{IE}.
\end{aligned} \tag{A18}$$

The last two conditions are the same as those obtained for the 4-dimensional system (Eq. (A17)). The first condition is similar to the 4-D case, but now the time constants of the population activity, τ_E and τ_I, contribute to the positive and negative feedback similarly to the synaptic time constants, τ_{EE} and τ_{II}, respectively. The second condition is similar to the first condition, but with extra terms containing the various τ's. Thus, having slower excitatory time constants than inhibitory ones is beneficial to stable persistent firing. The sufficient and necessary conditions obtained through the Routh stability criterion also follow these general rules but have much more complicated forms and thus are not shown here.

1.4. Activity patterns during persistent firing and the optimal input direction

In this section, we analytically obtain the activity patterns observed during persistent firing and the optimal input direction that maximizes the response of the network. We show below that the firing rates of the excitatory and inhibitory populations change proportionally for different levels of persistent firing, as has been observed experimentally [11]. On the other hand, we show that the response to external input is maximized when the external input excites the excitatory neurons and suppresses the inhibitory ones, as has been suggested to lead to a transient amplification of activity in sensory networks composed of excitatory and inhibitory populations [12].

To find the activity pattern for persistent firing and the best input direction, we decompose the network activity into its eigenvector components. In a linear system that is eigenvector-decomposable, the network activity in response to a transient input can be described by its eigenvalues and corresponding eigenvectors. In particular, when one eigenvalue has real part much larger than the remaining eigenvalues, the network activity can be expressed approximately in terms of this leading eigenvalue and its corresponding left and right eigenvectors [13]. Since the system with derivative feedback discussed in the previous sections has one eigenvalue close to 0 and the remaining eigenvalues have real parts strictly less than 0, the network activity with derivative feedback is well-described by

$$\vec{y} = \vec{y}_0 + e^{\overleftrightarrow{A}t}\vec{v} \simeq \vec{y}_0 + \mathrm{e}^{\lambda_1 t}(\vec{q}_1^{\,l}\cdot\vec{v})\vec{q}_1^{\,r}, \tag{A19}$$

where $\overleftrightarrow{A}$ is the matrix defined in Eq. (A7), $\vec{y}_0$ is the vector of states before the arrival of the transient input, and $\vec{v}$ is the external input vector.

In Eq. (A19), if $\vec{y}_0 = 0$, the right eigenvector $\vec{q}_1^{\,r}$ corresponds to the activity pattern that is maintained during persistent firing and the amplitude of this pattern is proportional to $\vec{q}_1^{\,l}\cdot\vec{v}$, that is, the projection of the input vector $\vec{v}$ onto the left eigenvector $\vec{q}_1^{\,l}$. Thus, the ratio between r_E and r_I is proportional to the ratio between the first and second elements of $\vec{q}_1^{\,r}$. In the derivative feedback networks, $\vec{q}_1^{\,r}$ can be found through its defining equation $\overleftrightarrow{A}\vec{q}_1^{\,r} = \lambda_1\vec{q}_1^{\,r} \approx 0$. Then, $\vec{q}_1^{\,r}$ is expressed in terms of the network parameters according to

$$\begin{array}{c}\left(-r_E + J_{EE}s_{EE} - J_{EI}s_{EI}\right)/\tau_E \approx 0\\ \left(-r_I + J_{IE}s_{IE} - J_{II}s_{II}\right)/\tau_I \approx 0\\ \left(-s_{EE} + r_E\right)/\tau_{EE} \approx 0\\ \left(-s_{IE} + r_E\right)/\tau_{IE} \approx 0\\ \left(-s_{EI} + r_I\right)/\tau_{EI} \approx 0\\ \left(-s_{II} + r_I\right)/\tau_{II} \approx 0\end{array} \quad \Rightarrow \quad \vec{q}_1^{\,r} \propto \begin{pmatrix} J_{II}\\ J_{IE}\\ J_{II}\\ J_{II}\\ J_{IE}\\ J_{IE}\end{pmatrix} \sim \frac{J_{II}}{J_{EI}}\begin{pmatrix} J_{EI}\\ J_{EE}\\ J_{EI}\\ J_{EI}\\ J_{EE}\\ J_{EE}\end{pmatrix} \quad \text{since } \frac{J_{EE}J_{II}}{J_{EI}J_{IE}} \sim 1,$$

and the ratio between r_E and r_I during persistent firing is $J_{II} / J_{IE} \sim J_{EI} / J_{EE}$. Since the J's are positive, this ratio is positive, that is, r_E and r_I positively covary for different levels of persistent firing.

The left eigenvector $\vec{q}_1^{\,l}$ can be computed similarly from its definition $\overleftrightarrow{A}^T \vec{q}_1^{\,l} \approx 0$. The expression for $\vec{q}_1^{\,l}$ in terms of the network parameters is found according to

$$\begin{aligned} -r_E / \tau_E + s_{EE} / \tau_{EE} + s_{IE} / \tau_{IE} &\approx 0 \\ -r_I / \tau_I + s_{EI} / \tau_{EI} + s_{II} / \tau_{II} &\approx 0 \\ J_{EE} r_E / \tau_E - s_{EE} / \tau_{EE} &\approx 0 \\ J_{IE} r_I / \tau_I - s_{IE} / \tau_{IE} &\approx 0 \\ -J_{EI} r_E / \tau_E - s_{EI} / \tau_{EI} &\approx 0 \\ -J_{II} r_I / \tau_I - s_{II} / \tau_{II} &\approx 0 \end{aligned} \quad \Rightarrow \quad \vec{q}_1^{\,l} \propto \begin{pmatrix} J_{IE}\tau_E \\ -J_{EE}\tau_I \\ J_{EE}J_{IE}\tau_{EE} \\ -J_{EE}J_{IE}\tau_{IE} \\ -J_{EI}J_{IE}\tau_{EI} \\ J_{EI}J_{IE}\tau_{II} \end{pmatrix}.$$

Notably, the first and second elements of $\vec{q}_1^{\,l}$ have different signs. Since the amplitude of persistent activity is proportional to $\vec{q}_1^{\,l} \cdot \vec{v}$, which is maximized when $\vec{v}$ is parallel to $\vec{q}_1^{\,l}$, the optimal external input to the excitatory and inhibitory populations should have different signs, that is, excite the excitatory populations and suppress the inhibitory populations. Note, by contrast, that the activities of the excitatory and inhibitory populations during persistent firing have the same sign. This difference between the persistent activity pattern and the optimal direction of the input, that is the difference between the left and right eigenvectors, arises from the asymmetry of the network connectivity. Thus, it is inherent in networks of excitatory and inhibitory populations [12].

In summary, in this section, we found the activity patterns during persistent firing and the external input pattern that maximizes the network response. We showed that the firing rates of the excitatory and inhibitory populations positively covary for different levels of persistent activity, while the maximal response is attained in response to input that excites the excitatory population and inhibits the inhibitory population.

1.5. Robustness against perturbations in the network connectivity

In this section, we study the effects of perturbations in the network connectivity on the ability to maintain persistent activity. We find that persistent firing in the derivative feedback network is robust against many commonly studied perturbations such as gain changes, changes of excitation or inhibition, and inactivation of a fraction of the excitatory or inhibitory populations (Fig. 4). To show this robustness, we check how the balance condition Eq. (A13) is affected under such perturbations.

We examine the types of perturbations of the network parameters under which the system still maintains persistent activity, that is, has an eigenvalue close to 0 and satisfies the stability conditions. In particular, we consider multiplicative scaling m_{ij} of the synaptic strengths, that is, the synaptic strengths become $m_{ij}J_{ij}$. Then, gain control in the entire population corresponds to a uniform increase or decrease of all m's and selective gain control in the excitatory or inhibitory population corresponds to a uniform increase in $m_{E,j}$ or $m_{I,j}$ for $j = E, I,$ and O. Similarly, inactivation (or loss) of a subpopulation of excitatory or inhibitory populations or presynaptic changes in transmission can be modeled by multiplicative changes in the strengths of excitatory or inhibitory synapses, corresponding to uniform increases in the m_{iE}'s and m_{iO}*'s* or m_{iI}'s, respectively, for $i = E$ or I.

Under this multiplicative change in the synaptic strengths, the balance condition for the existence of an eigenvalue close to 0 becomes $m_{EE}J_{EE}m_{II}J_{II} / \left(m_{IE}J_{IE}m_{EI}J_{EI}\right) \sim 1$, i.e. $m_{EE}m_{II} / \left(m_{EI}m_{IE}\right) \sim 1$. First, we note that for changes in intrinsic neuronal gains in the entire network, that is, uniform increase in m's, this condition is satisfied. This reflects that, since the positive and negative feedback change in the same manner, the net recurrent input continues to provide derivative feedback (Fig. 4i). Second, we see that multiplicative changes in the gain of the excitatory or inhibitory population or changes in excitatory (Fig. 4j) or inhibitory (Fig. 4k) synapses or inactivation of a subpopulation of excitatory or inhibitory populations similarly maintain the balance condition since presynaptic excitation, presynaptic inhibition, postsynaptic excitation, and postsynaptic inhibition are in both the numerator and denominator of the above expression.

The stability conditions given in Eq. (A17) also are satisfied under moderate perturbations in synaptic strengths. Only the first condition in Eq. (A17) depends on the synaptic strengths, requiring that $m_{EE}J_{EE}/\tau_{EE}$ should not exceed $m_{II}J_{II}/\tau_{II}$. In our models with biologically plausible parameters, τ_{EE} is an order of magnitude larger than τ_{II} and J_{EE} is of the same order as J_{II}. Thus, even in the presence of the perturbations which increase J_{EE} or decrease J_{II}, the system satisfies the stability conditions for a large range of perturbations. However, too much increase in the overall excitatory input to the system could break the stability condition and could make the system unstable.

We remark that the derivative feedback models are not robust against all forms of perturbations. For example, if the NMDA conductance is larger in excitatory to excitatory than in excitatory to inhibitory connections, perturbation specifically of NMDA-type synapses disrupts persistent firing since increasing m_{EE} more than m_{IE} breaks the balance condition $m_{EE}m_{II} / \left(m_{EI}m_{IE}\right) \sim 1$. However, the disruption resulting from this deviation in the balance condition is similar to that observed in positive feedback models: if only m_{EE} changes while all other m_{ij}=1, the time constant of the network activity in Eq. (A10) becomes

$$\begin{aligned}\tau_{network} &= \frac{J_{EI}J_{IE}/(J_{II}+1)\left(\tau_{EE}+\tau_{II}\right)-(m_{EE}J_{EE}-1)\left(\tau_{IE}+\tau_{EI}\right)+c}{J_{EI}J_{IE}/(J_{II}+1)-(m_{EE}J_{EE}-1)} \\ &\approx \frac{J_{EE}\left(\tau_{EE}+\tau_{II}-m_{EE}\left(\tau_{IE}+\tau_{EI}\right)\right)}{\left(1-m_{EE}\right)J_{EE}} \\ &\approx \frac{\tau_{EE}}{\left(1-m_{EE}\right)} \quad \text{if} \quad \tau_{EE} \gg \tau_{IE}, \tau_{EI}, \tau_{II}.\end{aligned} \tag{A20}$$

In the first approximation above, the balance condition Eq. (A13) is used to replace $J_{EI}J_{IE}/(J_{II}+1)$ by J_{EE}. The final expression above is similar to the time constant of decay $\tau/(1-W_{pos})$ in simple positive feedback networks (e.g. equation (1) of the main text when W_{der}=0 and when the dominant intrinsic cellular or synaptic time constant is τ_{EE}). When W_{pos} is perturbed by m from 1, the time constant becomes $\tau/(1-m)$, similar to Eq. (A20).

We note that negative derivative feedback models with NMDA-type synapses of approximately equal strength at all excitatory synapses, but with slower NMDA-synapses in the *E*-to-*E* pathway, can be far more robust against perturbations in NMDA-type synapses. To see this, we consider network models in which all excitatory connections are mediated by two different types of synaptic currents, NMDA-mediated currents and AMPA-mediated currents (Fig. 3 and online methods Eq. (7)). If we assume the ratios of NMDA- to AMPA-type synapses are the same in all excitatory pathways so that $q_{EE}=q_{IE}=q$, but the NMDA-type synapses in the *E*-to-*E* connection have slower kinetics, $\tau_{EE}^{N}>\tau_{IE}^{N}$, then perturbations in NMDA-type synapses by a fraction m maintain the balance condition as follows:

$$\begin{aligned}&J_{EI}\left(J_{IE}^{N}+J_{IE}^{A}\right)-\left(J_{EE}^{N}+J_{EE}^{A}\right)J_{II} \\ &= J_{EI}\left(qmJ_{IE}+(1-q)J_{IE}\right)-\left(qmJ_{EE}+(1-q)J_{EE}\right)J_{II} \\ &\approx (J_{EI}J_{IE}-J_{EE}J_{II})(qm+1-q) \ll O(J^{2}).\end{aligned}$$

Thus, the persistent activity is minimally affected by the perturbations in NMDA-type synapses, in contrast to the gross disruptions that occur in pure positive feedback models or in derivative feedback models with NMDA-type synapses only in *E*-to-*E* connections. Even in the case that $q_{EE}\neq q_{IE}$, the disruption of the persistent activity is less severe if the *E*-to-*E* NMDA-type synapses are relatively slow, $\tau_{EE}^{N}>\tau_{IE}^{N}$. By contrast, if all NMDA synapses have the same kinetics, $\tau_{EE}^{N}\approx\tau_{IE}^{N}$, and negative-derivative feedback is accomplished by having stronger NMDA conductance in *E*-to-*E* connections, $q_{EE}>q_{IE}$, then the network will exhibit the same disruption in persistent activity as in negative-derivative feedback models in which NMDA appears exclusively in *E*-to-*E* connections (calculations not shown). Thus, having slower NMDA-type

synapses in *E*-to-*E* than *E*-to-*I* connections [14-15] is advantageous in making the system more robust to disruptions of NMDA-type conductances.

1.6. Negative derivative feedback for networks of neurons with input-output nonlinearity

In this section, we consider a network model in which the individual neurons have a nonlinear firing rate versus input current relationship and show that the network implements derivative feedback control under conditions similar to the linear networks. In the presence of such nonlinearity, global analysis of the network dynamics through the eigenvector decomposition is not possible. Instead, we identify possible sets of steady states and check the local stability around those steady states.

Let us assume that there exists a steady state. To characterize this steady state, we linearize the system locally around it. For this steady state to belong to a continuous attractor, there should exist at least one eigenvalue equal to or close to 0 in the local linearization. If we denote this steady state as $(r_E^0, r_I^0, s_{EE}^0, s_{IE}^0, s_{EI}^0, s_{II}^0)^T$ and move the origin to the steady state, the linearization of Eq. (A6) becomes

$$\begin{aligned}
\tau_E \delta\dot{r}_E &= -\delta r_E + c_E(J_{EE}\delta s_{EE} - J_{EI}\delta s_{EI}) \quad \text{with } c_E = f_E'(J_{EE}s_{EE}^0 - J_{EI}s_{EI}^0) \\
\tau_I \delta\dot{r}_I &= -\delta r_I + c_I(J_{IE}\delta s_{IE} - J_{II}\delta s_{II}) \quad \text{with } c_I = f_I'(J_{IE}s_{IE}^0 - J_{II}s_{II}^0) \\
\tau_{EE}\delta\dot{s}_{EE} &= -\delta s_{EE} + \delta r_E \\
\tau_{IE}\delta\dot{s}_{IE} &= -\delta s_{IE} + \delta r_E \\
\tau_{EI}\delta\dot{s}_{EI} &= -\delta s_{EI} + \delta r_I \\
\tau_{II}\delta\dot{s}_{II} &= -\delta s_{II} + \delta r_I,
\end{aligned} \tag{A21}$$

where $f_i'(x_i)$ denotes the derivative of $f_i(x_i)$ evaluated at x_i, $\delta r_i = r_i - r_i^0$ and $\delta s_{ij} = s_{ij} - s_{ij}^0$. Eq. (A21) is the same as the system with linear input-output relationships in the previous sections, but with different slopes c_E and c_I. Thus, for the system to have an eigenvalue close to 0 through negative derivative feedback, we obtain similar conditions but with the replacement of each J_{ij} by $c_i J_{ij}$ for $i = E$, or I so that

$$c_E c_I \left\{ J_{EI} J_{IE} \left(\tau_{EE} + \tau_{II} \right) - J_{EE} J_{II} \left(\tau_{IE} + \tau_{EI} \right) \right\} \sim O(J^2),$$

$$c_E c_I \left\{ J_{EI} J_{IE} - J_{EE} J_{II} \right\} \ll O(J^2),$$

for large J's. If c_E and c_I are not too small, then the constants can be ignored and the above conditions are the same as the linear dynamics given by Eq. (A12) and Eq. (A13). Thus, the conditions for negative derivative feedback do not depend on the specific form of the input-output nonlinearity in the regime that the slopes of the input-output function are not small in

magnitude. Typical input-output nonlinearities such as sigmoid functions have a non-zero slope away from the threshold and the saturation. Thus, a continuum set of steady states corresponding to persistent activity will be located in such a regime.

The stability conditions at each steady state (Eq. (A17)) do depend on c_E and c_I as $c_I J_{II}/\tau_{II} > c_E J_{EE}/\tau_{EE}$. However, for τ_{EE} 1-2 orders of magnitude larger than τ_{II}, this condition can hold for a wide range of c_E and c_I. Thus, in contrast to positive feedback networks (Supplementary Fig. S1e,h), the memory performance in derivative feedback networks (Supplementary Fig. S1f,i) or hybrid networks containing a large derivative feedback component (Supplementary Fig. S1g,j) is robust to adding an input-output nonlinearity.

2. Analysis of firing rate models of two competing populations

In the previous sections, we discussed a derivative feedback network model consisting of one excitatory population and one inhibitory population. In parametric working memory tasks [16] and decision making tasks such as two-alternative forced choice tasks [17], it has been suggested that there exist two competing populations whose firing rates vary in opposite directions as a function of the remembered stimulus parameter. In many traditional models, positive feedback within and between the populations has been utilized for the maintenance or integration of evidence toward one choice or another [16, 18]: When the two competing populations are connected through mutual inhibition or through a common inhibitory pool, this forms a disinhibitory positive feedback loop between the populations. Thus, in such models, both recurrent excitatory and recurrent inhibitory synaptic interactions provide positive feedback that prolongs the time constant of decay of network activity.

In contrast to these traditional models, we suggest a model of two competing populations based on negative derivative feedback. In Section 2.1, we show that previously suggested model architectures for competing populations cannot generate persistent firing through derivative feedback control. In Section 2.2, we construct a new network model of two competing populations. In the new network architecture, we find conditions on the network parameters for derivative feedback control and describe its dynamical features. In Section 2.3, we show that a network model with derivative feedback is robust against the same types of perturbations in network parameters considered previously.

2.1. Previous models with positive feedback

In this section, we analyze previously proposed short-term memory models with two competing populations (Figs. 6a,b). Network interactions in these previous models mediated positive feedback through recurrent excitation within a population and mutual inhibition between the two populations. Here, we show that these network architectures cannot contain large

derivative feedback in their synaptic interactions. The essence of the explanation is as follows: in previous models, inhibitory inputs are arranged as part of disinhibitory loops that contribute positive feedback to the system. Since the total amount of positive feedback to each neuron should be balanced with the intrinsic leakiness during persistent activity, the amount of excitatory and inhibitory inputs are bounded in such positive feedback models. Thus, the models cannot have the large balanced excitatory and inhibitory inputs required for strong derivative feedback. Below, we prove this mathematically.

First, we consider a positive feedback model with disynaptic mutual inhibition as shown in Fig. 6a. It consists of two populations, each of which consists of excitatory and inhibitory sub-populations. The inhibitory neurons receive inputs from excitatory neurons in the opposing population and inhibit the excitatory neurons in the same population. The system can be described by 12 state variables $\vec{y} = (r_{E_1}, r_{I_1}, s_{E_1E_1}, s_{I_1E_1}, s_{I_1I_1}, s_{E_2I_1}, r_{E_2}, r_{I_2}, s_{E_2E_2}, s_{I_2E_2}, s_{I_2I_2}, s_{E_1I_2})^{\mathrm{T}}$, where *E* and *I* stand for the excitatory and inhibitory populations and the subscript *1* or *2* is the index of the population. To see that large excitatory and inhibitory inputs are not allowed, we simplify the system by assuming that all variables except $s_{E_1E_1}$ and $s_{E_2E_2}$ have fast kinetics and approximate them as achieving their steady states instantaneously. Then the system is described by the following equations:

$$\begin{aligned}
r_{E_1} &= J_{E_1E_1}s_{E_1E_1} - J_{E_1I_1}s_{E_1I_1} + J_{E_1O}i(t) \\
r_{I_1} &= -J_{I_1I_1}s_{I_1I_1} + J_{I_1E_2}s_{I_1E_2} + J_{I_1O}i(t) \\
r_{E_2} &= +J_{E_2E_2}s_{E_2E_2} - J_{E_2I_2}s_{E_2I_2} + J_{E_2O}i(t) \\
r_{I_2} &= J_{I_2E_1}s_{I_2E_1} - J_{I_2I_2}s_{I_2I_2} + J_{I_2O}i(t) \\
s_{ij} &= r_j \quad \text{for } i, j = E_1, E_2, I_1, I_2 \text{ except when } i = j = E_1 \text{ or } i = j = E_2
\end{aligned} \tag{A22}$$

$$\begin{aligned}
\tau_{E_1E_1}\dot{s}_{E_1E_1} &= -s_{E_1E_1} + r_{E_1} \\
\tau_{E_2E_2}\dot{s}_{E_2E_2} &= -s_{E_2E_2} + r_{E_2}.
\end{aligned} \tag{A23}$$

In the absence of external input, when *i(t)*=0, the firing rates of the excitatory and inhibitory sub-populations can be expressed in terms of the slow variables, $s_{E_1E_1}$ and $s_{E_2E_2}$ by solving Eq.(A22). Using these expressions, we obtain two conditions for the system defined by Eq. (A23) to have one eigenvalue close to zero and one negative eigenvalue so that the system maintains persistent activity stably. The conditions are given by

$$\left(J_{E_1E_1} - 1\right)\left(J_{E_2E_2} - 1\right) = \left(J_{E_1I_1}J_{I_1E_2} / (1 + J_{I_1I_1})\right)\left(J_{E_2I_2}J_{I_2E_1} / (1 + J_{I_2I_2})\right), \tag{A24}$$

$$-2+\frac{J_{E_1E_1}+J_{E_2E_2}}{1-\left(J_{E_1I_1}J_{I_1E_2}/(1+J_{I_1I_1})\right)\left(J_{E_2I_2}J_{I_2E_1}/(1+J_{I_2I_2})\right)}\le 0$$
$$\Rightarrow \left(J_{E_1E_1}-1\right)+\left(J_{E_2E_2}-1\right)\le -2\left(J_{E_1I_1}J_{I_1E_2}/(1+J_{I_1I_1})\right)\left(J_{E_2I_2}J_{I_2E_1}/(1+J_{I_2I_2})\right). \quad \text{(A25)}$$

Assuming, for simplicity, symmetry in the connection strengths between the two populations, each $J_{EE}-1$ in Eq. (A25) should be negative since the right side of the inequality is less than 0. Furthermore, from Eq. (A24), $J_{E_1I_1}J_{I_1E_2}/(1+J_{I_1I_1})$ should be less than 1. Here, $J_{E_1E_1}r_{E_1}$ and $J_{E_1I_1}J_{I_1E_2}/(1+J_{I_1I_1})r_{E_2}$ represent the strengths of the excitatory input and the inhibitory input to r_{E_1}, respectively. Thus, the excitatory and inhibitory inputs cannot be large.

Similarly, two competing populations connected through a common inhibitory pool as in Fig. 6b cannot receive large excitatory and inhibitory inputs. The dynamics of this network can be described by 10 state variables $\vec{y}=(r_{E_1}, s_{E_1E_1}, s_{IE_1}, r_{E_2}, s_{E_2E_2}, s_{IE_2}, r_I, s_{E_1I}, s_{E_2I} s_{II})^{\mathrm{T}}$. As in the previous model, we assume that the dynamics of all variables except $s_{E_1E_1}$ and $s_{E_2E_2}$ is fast. Then, the system is given by

$$\begin{aligned} r_{E_1} &= J_{E_1E_1}s_{E_1E_1} - J_{E_1I}s_{E_1I} + J_{E_1O}i(t) \\ r_{E_2} &= J_{E_2E_2}s_{E_2E_2} - J_{E_2I}s_{E_2I} + J_{E_2O}i(t) \\ r_I &= J_{IE_1}s_{IE_1} + J_{IE_2}s_{IE_2} - J_{II}s_{II} + J_{IO}i(t) \end{aligned}$$

$$s_{ij} = r_j \quad \text{for } i,j = E_1, E_2, I \quad \text{except when } i=j=E_1 \text{ or } i=j=E_2$$

$$\begin{aligned} \tau_{E_1E_1}\dot{s}_{E_1E_1} &= -s_{E_1E_1}+r_{E_1} \\ \tau_{E_2E_2}\dot{s}_{E_2E_2} &= -s_{E_2E_2}+r_{E_2}. \end{aligned}$$

The conditions for persistent firing and stability are given by

$$\left(J_{E_1E_1}-1-\frac{J_{E_1I}J_{IE_1}}{1+J_{II}}\right)\left(J_{E_2E_2}-1-\frac{J_{E_2I}J_{IE_2}}{1+J_{II}}\right)=\frac{J_{E_1I}J_{IE_2}}{1+J_{II}}\frac{J_{E_2I}J_{IE_1}}{1+J_{II}} \quad \text{(A26)}$$

$$\frac{J_{E_1E_1}\left(1+J_{E_1I}J_{IE_1}/(1+J_{II})\right)+J_{E_2E_2}\left(1+J_{E_2I}J_{IE_2}/(1+J_{II})\right)}{\left(1+J_{E_1I}J_{IE_1}/(1+J_{II})\right)\left(1+J_{E_2I}J_{IE_2}/(1+J_{II})\right)-\left(J_{E_1I}J_{IE_2}/(1+J_{II})\right)\left(J_{E_2I}J_{IE_1}/(1+J_{II})\right)}-2\le 0.$$

To see that large excitatory and inhibitory inputs are not allowed in this system, we assume symmetry between the two populations by setting the strengths of the corresponding

recurrent synaptic variables in each population to be equal (e.g., $J_{E_1E_1} = J_{E_2E_2} \equiv J_{EE}$). Then, the stability condition simplifies to

$$\begin{aligned} &\frac{2J_{EE}\left(1+J_{EI}J_{IE}/(1+J_{II})\right)}{\left(1+J_{EI}J_{IE}/(1+J_{II})\right)^2-\left(J_{EI}J_{IE}/(1+J_{II})\right)^2}-2\le 0 \\ &\Rightarrow J_{EE}\left(1+J_{EI}J_{IE}/(1+J_{II})\right)\le 1+2J_{EI}J_{IE}/(1+J_{II}). \end{aligned} \tag{A27}$$

From Eq. (A27), J_{EE} is found to be less than 2, so that the excitatory input and the derivative feedback cannot be large.

2.2. Construction of two competing populations with negative derivative feedback

In the previous section, we showed that previously suggested network architectures for two competing populations generate persistent activity through positive feedback control. Here, we construct models of two competing populations that perform derivative feedback control and describe their dynamical features. We find that, despite their complicated algebraic forms, the core conditions for negative derivative feedback remain the same as in the recurrent network with one excitatory population and one inhibitory population. That is, a balance between positive and negative feedback in strength and slower positive feedback are necessary for negative derivative feedback control.

To construct a model of two competing populations based on negative derivative feedback, we assume that each population is composed of excitatory and inhibitory sub-populations as in the model of Section 1, and the two populations are interconnected through excitatory synapses (Fig. 6c; including inhibitory synapses between the populations leads to similar results, but with more complicated form). The dynamics of the system is written as

$$\begin{aligned} &\tau_i \dot{r}_i = -r_i + \sum_j J_{ij}s_{ij} + J_{iO}i(t) \\ &\tau_{ij}\dot{s}_{ij} = -s_{ij} + r_j \quad \text{for } i, j = E_1, I_1, E_2, I_2. \end{aligned} \tag{A28}$$

Here, we have 16 state variables with 4 variables for firing rates and 12 variables for recurrent synapses. *E* and *I* stand for the excitatory and inhibitory populations, respectively, and the subscript *1* or *2* denotes the index of the population. In the 12 recurrent synaptic variables, 8 synapses are for intra-connectivity (4 synapses in each population) and 4 synapses are for interconnectivity from an excitatory sub-population of one side to excitatory and inhibitory sub-populations on the other side. We assume that the time constants of the synaptic variables are the same for the same type of synapses, for example, $\tau_{E_1E_1} = \tau_{E_2E_2} = \tau_{EE}$ and $\tau_{E_1E_2} = \tau_{E_2E_1} = \tau_{EE,inter}$.

To find conditions for negative derivative feedback, we compare the coefficients of the first order term a_1 and the constant term a_0 of the characteristic polynomial of the system in Eq. (A28), as in Section 1.2. This leads to two conditions:

$$\begin{aligned}&\left(J_{E_1E_1}J_{I_1I_1}-J_{E_1I_1}J_{I_1E_1}\right)\left(J_{E_2E_2}J_{I_2I_2}-J_{E_2I_2}J_{I_2E_2}\right)\\&-J_{E_1E_2}J_{E_2E_1}J_{I_1I_1}J_{I_2I_2}-J_{E_1I_1}J_{I_1E_2}J_{E_2I_2}J_{I_2E_1}+J_{I_1I_1}J_{E_1E_2}J_{E_2I_2}J_{I_2E_1}+J_{I_2I_2}J_{E_2E_1}J_{E_1I_1}J_{I_1E_2}\ll O(J^4),\end{aligned} \tag{A29}$$

$$\begin{aligned}&\left(\begin{aligned}&\left(J_{E_1E_1}J_{I_1I_1}-J_{E_1I_1}J_{I_1E_1}\right)\left(J_{E_2E_2}J_{I_2I_2}-J_{E_2I_2}J_{I_2E_2}\right)\\&-J_{E_1E_2}J_{E_2E_1}J_{I_1I_1}J_{I_2I_2}-J_{E_1I_1}J_{I_1E_2}J_{E_2I_2}J_{I_2E_1}+J_{I_1I_1}J_{E_1E_2}J_{E_2I_2}J_{I_2E_1}+J_{I_2I_2}J_{E_2E_1}J_{E_1I_1}J_{I_1E_2}\end{aligned}\right)\left(\sum_{i,j=E,I}\tau_{ij}+\sum_{i=E}\tau_{iE,inter}\right)\\&+\left(J_{E_1E_1}J_{I_1I_1}-J_{E_1I_1}J_{I_1E_1}\right)\left(J_{E_2E_2}J_{I_2I_2}-J_{E_2I_2}J_{I_2E_2}\right)\left(\tau_{EE,inter}+\tau_{IE,inter}\right)\\&-\left(J_{E_1E_2}J_{E_2E_1}J_{I_1I_1}J_{I_2I_2}+J_{E_1I_1}J_{I_1E_2}J_{E_2I_2}J_{I_2E_1}-J_{I_1I_1}J_{E_1E_2}J_{E_2I_2}J_{I_2E_1}-J_{I_2I_2}J_{E_2E_1}J_{E_1I_1}J_{I_1E_2}\right)\left(\tau_{EE}+\tau_{IE}\right)\\&-\left(J_{E_1E_1}J_{I_1I_1}J_{E_2E_2}J_{I_2I_2}-J_{E_1I_1}J_{I_1E_1}J_{E_2I_2}J_{I_2E_2}\right)\left(\tau_{EE}+\tau_{II}-\tau_{IE}-\tau_{EI}\right)\\&+\left(J_{E_1E_2}J_{E_2E_1}J_{I_1I_1}J_{I_2I_2}-J_{E_1I_1}J_{I_1E_2}J_{E_2I_2}J_{I_2E_1}\right)\left(\tau_{EE,inter}+\tau_{II}-\tau_{IE,inter}-\tau_{EI}\right)\sim O(J^4).\end{aligned} \tag{A30}$$

As in the simpler networks of Section 1, the above conditions constrain the connectivity strengths and time scales of the synapses. Analogous to Eq. (A13), Eq. (A29) imposes a constraint on the connectivity strengths; the first term of the equation is the product of the net positive feedback (subtraction of the total negative feedback from the total positive feedback) within each population and the remaining 4 terms are the feedbacks through the opposite population. In these remaining 4 terms, the first two terms are the positive feedbacks through the mutually excitatory loop and through the indirect mutually inhibitory loop, and the third and fourth terms correspond to negative feedback through activating the opposite inhibitory population. Thus, Eq. (A29) states that a balance between net positive feedback and negative feedback is necessary for generating persistent activity in the coupled networks.

Eq. (A30) imposes a constraint on the time constants of the synapses analogous to Eq. (A12) for the single population. In Eq. (A30), the first term is negligibly small compared to the remaining terms due to Eq. (A29). Furthermore, if we assume the time constants within the population and across the populations are the same for the same synaptic types so that $\tau_{EE}=\tau_{EE,inter}, \tau_{IE}=\tau_{IE,inter}$, then the second and third terms of Eq. (A30) cancel each other and the remaining terms are simplified to one product term,

$$\left(J_{E_1E_1}J_{I_1I_1}J_{E_2E_2}J_{I_2I_2}-J_{E_1I_1}J_{I_1E_1}J_{E_2I_2}J_{I_2E_2}-J_{E_1E_2}J_{E_2E_1}J_{I_1I_1}J_{I_2I_2}+J_{E_1I_1}J_{I_1E_2}J_{E_2I_2}J_{I_2E_1}\right)\left(\tau_{EE}+\tau_{II}-\tau_{IE}-\tau_{EI}\right).$$

Thus, each factor of the above product should be nonzero and we obtain $\tau_{EE}+\tau_{II}\neq\tau_{IE}+\tau_{EI}$.

If we furthermore assume that the network is symmetric and denote $J_{E_1E_1}=J_{E_2E_2}=J_{EE}$ and $J_{E_1E_2}=J_{E_2E_1}=J_{EE,inter}$, then Eq. (A29) assumes the much simpler form

$$\left(J_{EE}J_{II}-J_{IE}J_{EI}\right)^2-\left(J_{EE,inter}J_{II}-J_{IE,inter}J_{EI}\right)^2 \ll O(J^4).$$

The first term above is the difference between positive and negative feedback within the same-side population, and the second term is the difference between positive and negative feedback through the opposite-side population. Thus, in this case, the required balance between positive and negative feedback can be expressed in a relatively simple and intuitive form.

Note that the conditions for derivative feedback, Eq. (A29) and Eq. (A30), do not guarantee competition between the two populations. Since each population in general can project onto either the excitatory or inhibitory neurons of the other population, the activities of the two populations may exhibit positive or negative correlations, unlike in traditional models that only contain functionally inhibitory inter-connections and exhibit negative correlations. For the two populations to exhibit negative correlations, the inhibitory interactions between the populations (which are mediated by the excitatory projections of one population onto the inhibitory neurons of the opposing population) should be stronger than the excitatory interactions (Fig. 6c, calculation not shown).

In summary, in this section, we found conditions for two competing populations to implement negative derivative feedback. The network requires a similar balance condition between excitation and inhibition as in the network of one excitatory population and one inhibitory population, but the positive and negative feedback now include the interactions between the two competing populations. A time delay in positive feedback can be satisfied with slow recurrent excitation as in the single population models.

2.3. Robustness against perturbations in the network connectivity

Next, we investigate whether the model of two competing populations with negative derivative feedback can maintain persistent firing when the network parameters are perturbed. To do so, as in Section 1.5 for the case of the single population, we check whether the balance condition given in Eq. (A29) is satisfied in the presence of various perturbations.

Persistent firing in two competing populations with negative derivative feedback is found to be robust against many natural perturbations, as in the single population. First, we consider changes in intrinsic neuronal gains in the entire network, which can be modeled as a uniform multiplication of all synaptic strengths by a factor *m*. In this case, the balance condition in Eq. (A29) is multiplied by m^4 and is still satisfied. Second, selective gain control in any individual excitatory or inhibitory population also maintains persistent firing. For example, gain control in excitatory population 1 can be represented by multiplying all $J_{E_1 j}$ by *m* for $j = E_1, E_2, I_1, I_2$ so that the balance condition becomes

$$\left(mJ_{E_1E_1}J_{I_1I_1} - mJ_{E_1I_1}J_{I_1E_1}\right)\left(J_{E_2E_2}J_{I_2I_2} - J_{E_2I_2}J_{I_2E_2}\right)$$
$$-mJ_{E_1E_2}J_{E_2E_1}J_{I_1I_1}J_{I_2I_2} + mJ_{E_1I_1}J_{I_1E_2}J_{E_2I_2}J_{I_2E_1} - J_{I_1I_1}mJ_{E_1E_2}J_{E_2I_2}J_{I_2E_1} - J_{I_2I_2}J_{E_2E_1}mJ_{E_1I_1}J_{I_1E_2} \ll O(J^4),$$

which holds from Eq. (A29).

Finally, we consider multiplicative changes in the strengths of excitatory or inhibitory synapses projecting from a given population that might, for example, correspond to changes in presynaptic transmission or loss of a fraction of a subpopulation of neurons. For example, if the strengths of all synapses from E_1 are uniformly increased by m, the balance condition becomes Eq. (A29) multiplied by m and is satisfied. However, note that, for example, if the excitatory receptors change only within the same population but not across to the other population, that is, J_{E_1,E_1} and J_{I_1,E_1} are multiplied by m but J_{I_2,E_1} remains the same, then the condition for the persistent activity is not maintained, suggesting that this is a perturbation to which the network would be sensitive.

3. Analysis of spiking network models

Here, we provide a simple analytical description of spiking network activity that enables us to find conditions on the network connectivity for balanced excitation and inhibition. Based on previous work [19], we describe the network activity through a mean-field approximation that takes into account the mean presynaptic input received by all neurons of a given population (E and I). Within this framework, we find a condition on synaptic strengths for persistent firing to be maintained that has the same form as the balance condition in the firing rate model of Section 1.

For simplicity, we assume that each connection consists of one type of synapse, slow excitatory synapses from E to E, fast excitatory synapses from E to I, and fast synapses for all inhibitory synapses. Then Eq. (9) and Eq. (10) in the online methods can be rewritten as

$$\begin{aligned}
\tau_E \frac{dV_E^l}{dt} &= -\left(V_E^l - V_L\right) + I_{EE}^l(t) - I_{EI}^l(t) + I_{EO}^l(t) \\
\tau_I \frac{dV_I^l}{dt} &= -\left(V_I^l - V_L\right) + I_{IE}^l(t) - I_{II}^l(t) + I_{IO}^l(t) \\
\tau_{ij} \frac{dI_{ij}^l}{dt} &= -I_{ij}^l + \tilde{J}_{ij} \sum_{m=1}^{N_j} p_{ij}^{lm} \sum_{t_j^m} \delta(t - t_j^m),
\end{aligned} \tag{A31}$$

where I_{ij}^l denotes the total synaptic current from population j to neuron l in population i.

To obtain the mean-field description of the average firing rates of the excitatory and inhibitory populations, we assume neurons exhibit independent, irregular firing with Poisson

statistics, as has been suggested to occur in sparsely connected balanced networks with strong synapses [20]. Note, however, that unlike previous works that used synapses with instantaneous dynamics, synapses in our model have time constants comparable to or much longer than the membrane time constant. This may result in temporal correlations in spike trains [21-22] and may violate the assumption of Poisson statistics of firing activity. Thus, although the mean-field theory description below was sufficient to produce a memory network with graded persistent firing, a more precise description of network activity may additionally require an analysis of higher order moments of firing activity (see [23] for such an analysis in networks with instantaneous synapses).

Using the independent, Poisson assumption, we can obtain simple expressions for the dynamics of the population-averaged synaptic currents. Taking the population average of the dynamics of I_{ij}^l for $i,j = E,I$ in Eq. (A31) gives

$$\tau_{ij}\frac{dI_{ij}}{dt} = -I_{ij} + \left\langle \tilde{J}_{ij}\sum_{m=1}^{N_j} p_{ij}^{lm}\sum_{t_m}\delta(t-t_m)\right\rangle, \tag{A32}$$

where < > denotes the average across the population and I_{ij} is $\left\langle I_{ij}^l\right\rangle$. Since firing rates of excitatory and inhibitory neurons do not change much during persistent activity, we assume that the spike train of each neuron is generated from a stationary Poisson process with average rates r_E or r_I for the excitatory or inhibitory population, respectively. Since p_{ij}^{lm} is a binary random variable with probability p, the average number of synaptic contacts from j to i is $N_j p$, so that the temporal mean and variance of the input are given by

$$\begin{aligned}
\text{mean}\left(\left\langle \tilde{J}_{ij}\sum_{m=1}^{N_j} p_{ij}^{lm}\sum_{t_m}\delta(t-t_m)\right\rangle\right) &\approx \tilde{J}_{ij}pN_j r_j \equiv \mu_{ij} \\
\text{var}\left(\left\langle \tilde{J}_{ij}\sum_{m=1}^{N_j} p_{ij}^{lm}\sum_{t_m}\delta(t-t_m)\right\rangle\right) &\approx \tilde{J}_{ij}^{\,2}pN_j r_j \equiv \sigma_{ij}^2
\end{aligned}. \tag{A33}$$

We note that the term $\tilde{J}_{ij}pN_j$ multiplying the presynaptic firing rate r_j in the above expression represents the strength of the mean synaptic input from population j to population i, and is analogous to the strength of synaptic input J_{ij} in the firing rate models. Thus, we denote $\tilde{J}_{ij}pN_j$ as J_{ij}.

To find the tuning condition on the network parameters, we approximate the arrival of presynaptic spikes as a continuous stochastic, white noise process [19]. This approximation is valid when the number of presynaptic contacts is large and the contribution of each presynaptic spike is small. Then Eq. (A32) becomes

$$\tau_{ij}\frac{dI_{ij}}{dt} = -I_{ij} + \mu_{ij} + \sigma_{ij}\eta_{ij}(t),$$

where μ_{ij} and σ_{ij} are defined in Eq. (A33) and $\eta_{ij}(t)$ is a white noise process with zero mean and unit variance. Thus, each synaptic current becomes an Ornstein-Uhlenbeck process [24] and, in steady state, the mean and variance of I_{ij} become

$$\begin{aligned}\mu_{I_{ij}} &= \mu_{ij} = \tilde{J}_{ij}\, pN_j r_j, \\ \sigma_{I_{ij}}{}^2 &= \frac{\sigma_{ij}{}^2}{2\tau_{ij}} = \frac{\tilde{J}_{ij}{}^2\, pN_j r_j}{2\tau_{ij}}.\end{aligned} \tag{A34}$$

Next, we derive a necessary condition on the network parameters to maintain persistent activity in the absence of external input. Under the diffusion approximation, the temporal mean of the total input to a neuron is expressed in terms of the average firing rates of the excitatory and inhibitory populations as

$$\begin{aligned}\text{mean}(I_{EE} - I_{EI}) &= \mu_{I_{EE}} - \mu_{I_{EI}} = \tilde{J}_{EE}\, pN_E r_E - \tilde{J}_{EI}\, pN_I r_I, \\ \text{mean}(I_{IE} - I_{II}) &= \mu_{I_{IE}} - \mu_{I_{II}} = \tilde{J}_{IE}\, pN_E r_E - \tilde{J}_{II}\, pN_I r_I.\end{aligned} \tag{A35}$$

Below, we denote the population average numbers pN_E and pN_I of synaptic contacts from E and I as K_E and K_I, respectively. Assuming $\tilde{J}_{ij}$ scales inversely proportionally to K_j as $\tilde{J}_{ij} = \hat{J}_{ij} / \sqrt{K_j}$, where $\hat{J}_{ij}$ is a constant, so that the variance of the input remains finite with increasing network size [20],

$$\begin{aligned}\text{mean}(I_{EE} - I_{EI}) &= \frac{\hat{J}_{EE}}{\sqrt{K_E}} K_E r_E - \frac{\hat{J}_{EI}}{\sqrt{K_I}} K_I r_I = \hat{J}_{EE}\sqrt{K_E}\, r_E - \hat{J}_{EI}\sqrt{K_I}\, r_I \\ \text{mean}(I_{IE} - I_{II}) &= \frac{\hat{J}_{IE}}{\sqrt{K_E}} K_E r_E - \frac{\hat{J}_{II}}{\sqrt{K_I}} K_I r_I = \hat{J}_{IE}\sqrt{K_E}\, r_E - \hat{J}_{II}\sqrt{K_I}\, r_I.\end{aligned} \tag{A36}$$

The mean excitatory and inhibitory recurrent inputs to each neuron increase with the network size, and therefore, with K_E and K_I. Thus, to have the mean input of the same order as the leak currents, that is, of order 1, the excitatory and inhibitory currents should nearly cancel each other as

$$\begin{matrix}\hat{J}_{EE}\sqrt{K_E}\, r_E - \hat{J}_{EI}\sqrt{K_I}\, r_I & \sim O(1) \\ \hat{J}_{IE}\sqrt{K_E}\, r_E - \hat{J}_{II}\sqrt{K_I}\, r_I & \sim O(1)\end{matrix} \Rightarrow \begin{matrix}\hat{J}_{EE}\sqrt{K_E / K_I}\, r_E - \hat{J}_{EI} r_I & \sim O\left(1/\sqrt{K_I}\right) \\ \hat{J}_{IE}\sqrt{K_E / K_I}\, r_E - \hat{J}_{II} r_I & \sim O\left(1/\sqrt{K_I}\right)\end{matrix}. \tag{A37}$$

For large K_E and K_I , nonzero r_E and r_I can occur when the strengths of the recurrent synapses satisfy

$$\hat{J}_{EE}\hat{J}_{II} / \hat{J}_{EI}\hat{J}_{IE} \sim 1, \tag{A38}$$

and with $\tilde{J}_{ij} pN_j = J_{ij}$, Eq. (A37) becomes

$$J_{EE}J_{II} / J_{EI}J_{IE} \sim 1 \text{ for large } J's. \tag{A39}$$

Thus, we find that the balance condition for the spiking networks has a similar form to that obtained previously in the firing rate models (Eq. (3) of the main text). We note that the balance condition above is different from that obtained in previous work [20] in which the recurrent inhibitory input is balanced by recurrent and external excitatory inputs, and the network is quiescent in the absence of external input. A similar tuning condition for persistent firing has been suggested in [23] but, since this previous work did not utilize different dynamics for excitation and inhibition, the system was not stable across a graded range of firing rates. Also note that even though Eqs. (A38) and (A39) impose a tight tuning condition on the average strength of connections between populations, the total amount of excitatory and inhibitory input onto an individual neuron can vary widely due to the random connectivity. Thus, the system can be stable even in the presence of heterogeneity of network connectivity onto individual neurons.

4. Parameters

4.1. Firing rate model of a single population

In all firing rate models, the intrinsic time constants of excitatory and inhibitory neurons are τ_E = 20ms and τ_I = 10ms. The decay time constants for recurrent synapses are τ_{EE} = 100ms, τ_{IE} = 25ms for excitatory synapses, and τ_{EI} = τ_{II} = 10ms for inhibitory synapses. For the currents driven by NMDA- and AMPA-type excitatory synapses in Figs. 3 and 4, the time constants onto the excitatory population were τ_{EE}^N = 150ms and τ_{EE}^A = 50ms, and those onto the inhibitory population were τ_{IE}^N = 45ms and τ_{IE}^A = 20ms. As noted in the Online Methods, we note that these time constants represent the observed decay times of NMDA-receptor *driven* and AMPA-receptor *driven* EPSP's observed experimentally at the soma of prefrontal cortical neurons following synaptic stimulation [15], and almost certainly represent a mixture of receptor kinetics plus local intrinsic processing that effectively extend the time scales of decay of the synaptic currents. The fractions of NMDA-type synapses onto excitatory and inhibitory synapses were q_{EE} = 0.5 and q_{IE} = 0.2, and these were chosen to approximately match the temporally integrated current observed experimentally for excitatory to excitatory and excitatory to inhibitory connections in the same prefrontal cortical slice experiments used to derive the time

constants of the synaptically driven currents [15]. External inputs $i(t)$ to the excitatory and inhibitory populations were given by square pulses of t_{window}=100 ms duration for pulse-like inputs or 5 sec for step inputs, filtered by an exponential filter of time constant τ_{ext} = 100 ms.

For the nonlinear function of Naka-Rushton form in Eq. (6), the maximal response M = 100, the half-activation parameter x_0 = 30, and the input threshold x_θ = 10 (Fig. 2c-d, bottom, Supplementary Fig. S1).

The connectivity strengths for each model are summarized in the following table.

Connectivity strength for firing rate model			
Population structure	Model type	Recurrent connection	Connection from external input
One exc. and one inh. population	Der. fdbk. network	$J_{EE} = J_{IE} = 150, J_{EI} = J_{II} = 300$ in Fig. 2 $J_{EE} = 150$, $J_{IE} = 150$, $J_{EI} = 300$, $J_{II} = 300$-1 for purely der. model, $J_{EE} = 151$, $J_{IE} = 150$, $J_{EI} = 300$, $J_{II} = 300$-1 for hybrid model, $J_{iE} = J_{IE} = 150, J_{EI} = J_{II} = 600$ in Fig. 3 and Supplementary Fig. S2	J_{EO} = 1500-4500, $J_{IO} = 0$ for pulse-like input J_{EO} = 100-300, $J_{IO} = 0$ for step-like input
	Pos. fdbk. network	$J_{EE} = 1$, $J_{IE} = J_{EI} = J_{II} = 0$	J_{EO} = 30-90, $J_{IO} = 0$ for pulse-like input J_{EO} = 20-60, $J_{IO} = 0$ for step-like input
Two competing populations	Der. fdbk. network	$J_{E_iE_i} = 300$, $J_{I_iE_i} = 450$, $J_{E_iI_i} = 900$, $J_{I_iI_i} = 900$, $J_{E_iE_j} = 150$, $J_{I_iE_j} = 300$	$J_{E_i,tonic} = 9000$ $J_{E_1O} = 3000, J_{E_2O}$ = 0-6000
	Pos. fdbk. with direct mutual inhibition	$J_{E_iE_i} = 0.5$, $J_{I_iE_j} = 1, J_{E_iI_i} = 0.5$	$J_{E_i,tonic} = 30$ $J_{E_1O} = 15$, J_{E_2O} = 0-30
	Pos. fdbk. with common inhibitory pool	$J_{E_iE_i} = 1$, $J_{IE_i} = 1$, $J_{E_iI} = 0.5$	$J_{E_i,tonic} = 30$ $J_{E_1O} = 30$, J_{E_2O} = 0-60

4.2 Spiking network model with leaky integrate-and-fire neurons

In the simulations, N_E = 16000, N_I = 4000, and N_O = 20000. The probability of connection p = 0.1 uniformly across the population. The time constants of excitatory and inhibitory neurons were 20ms and 10ms, respectively. The remaining parameters of the integrate-and-fire neuron, which were the same for both excitatory and inhibitory neurons, were V_L = -60mV, V_θ = -40mV, and V_{reset} = –52 mV with a refractory period τ_{ref} = 2 ms.

The time constants of the synaptic variables were τ_{EE}^{N}=150ms, τ_{EE}^{A} = 100ms, τ_{IE}^{N} = 45ms, τ_{IE}^{A} = 20ms, and $\tau_{EI} = \tau_{II}$ = 10ms. The fractions of the NMDA-type synapses were q_{EE} =0.5 and q_{IE} =0.2, as in the firing rate models. For the filtering of the external input, τ_{iO} = 100ms. For the synaptic strengths, $\tilde{J}_{iE} = \hat{J}_{iE} / \sqrt{N_E p} = 300 / \sqrt{1600} = 7.5$ and $\tilde{J}_{iI} = \hat{J}_{iI} / \sqrt{N_I p} = 400 / \sqrt{400} = 20$.

In Fig. 5, the transient input from the external population was present for a duration t_{window}=100ms starting at time 0. The columns from left to right correspond to input firing rates r_O = 50Hz, 100Hz, or 200Hz with Poisson statistics, respectively, and the synapses from the external input have strengths $\tilde{J}_{EO} = \hat{J}_{EO} / \sqrt{N_O p} = 20 / \sqrt{2000} \sim 0.45$ and $J_{IO} = 0$. In Supplementary Fig. S4, the step-like input was present from time 0, the input firing rates r_O = 10Hz, 20Hz or 30Hz, and the synaptic strengths $\tilde{J}_{EO} = \hat{J}_{EO} / \sqrt{N_O p} = 10 / \sqrt{2000} \sim 0.22$ and $J_{IO} = 0$.

References

1. Horn, R.A. and C.R. Johnson, *Matrix Analysis*. 1985, Cambridge, UK: Cambridge University Press.
2. Wang, X.J., *Synaptic reverberation underlying mnemonic persistent activity.* Trends Neurosci., 2001. **24**(8): p. 455-463.
3. Brody, C.D., R. Romo, and A. Kepecs, *Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors, and dynamic representations.* Curr. Opin. Neurobiol., 2003. **13**(2): p. 204-11.
4. Durstewitz, D., J.K. Seamans, and T.J. Sejnowski, *Neurocomputational models of working memory.* Nat Neurosci, 2000. **3**: p. 1184-1191.
5. Goldman, M.S., A. Compte, and X.J. Wang, *Neural integrator models*, in *Encyclopedia of Neuroscience*, L.R. Squire, Editor. 2009, Oxford: Academic Press. p. 165-178.
6. Marder, E., et al., *Memory from the dynamics of intrinsic membrane currents.* Proc Natl Acad Sci U S A, 1996. **93**(24): p. 13481-6.
7. Hempel, C.M., et al., *Multiple forms of short-term plasticity at excitatory synapses in rat medial prefrontal cortex.* J Neurophysiol, 2000. **83**(5): p. 3031-41.
8. Wang, Y., et al., *Heterogeneity in the pyramidal network of the medial prefrontal cortex.* Nat Neurosci, 2006. **9**(4): p. 534-42.
9. Ermentrout, B., *Neural networks as spatio-temporal pattern-forming systems.* Rep. Prog. Phys., 1998. **61**: p. 353-430.
10. Nise, N.S., *Control systems engineering*. 4 ed. 2004: John Wiley & Sons, Inc.
11. Constantinidis, C. and P.S. Goldman-Rakic, *Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex.* J Neurophysiol, 2002. **88**(6): p. 3487-97.
12. Murphy, B.K. and K.D. Miller, *Balanced amplification: a new mechanism of selective amplification of neural activity patterns.* Neuron, 2009. **61**(4): p. 635-48.

13. Lim, S. and M.S. Goldman, *Noise tolerance of attractor and feedforward memory models.* Neural Comput, 2012. **24**(2): p. 332-90.
14. Wang, H., et al., *A specialized NMDA receptor function in layer 5 recurrent microcircuitry of the adult rat prefrontal cortex.* Proc Natl Acad Sci U S A, 2008. **105**(43): p. 16791-6.
15. Rotaru, D.C., et al., *Glutamate receptor subtypes mediating synaptic activation of prefrontal cortex neurons: relevance for schizophrenia.* J Neurosci, 2011. **31**(1): p. 142-56.
16. Machens, C.K., R. Romo, and C.D. Brody, *Flexible control of mutual inhibition: a neural model of two-interval discrimination.* Science, 2005. **307**(5712): p. 1121-4.
17. Gold, J.I. and M.N. Shadlen, *The neural basis of decision making.* Annu Rev Neurosci, 2007. **30**: p. 535-74.
18. Wong, K.F. and X.J. Wang, *A recurrent network mechanism of time integration in perceptual decisions.* J. Neurosci., 2006. **26**(4): p. 1314-28.
19. Renart, A., N. Brunel, and X.J. Wang, *Mean-field theory of irregularly spiking neuronal populations and working memory in recurrent cortical networks*, in *Computational neuroscience: a comprehensive approach*, J. Feng, Editor. 2004, Chanpman & Hall/CRC. p. 425.
20. van Vreeswijk, C. and H. Sompolinsky, *Chaos in neuronal networks with balanced excitatory and inhibitory activity.* Science, 1996. **274**(5293): p. 1724-6.
21. Moreno-Bote, R. and N. Parga, *Auto- and crosscorrelograms for the spike response of leaky integrate-and-fire neurons with slow synapses.* Phys Rev Lett, 2006. **96**(2): p. 028101.
22. Moreno-Bote, R. and N. Parga, *Response of integrate-and-fire neurons to noisy inputs filtered by synapses with arbitrary timescales: firing rate and correlations.* Neural Comput, 2010. **22**(6): p. 1528-72.
23. Renart, A., et al., *Mean-driven and fluctuation-driven persistent activity in recurrent networks.* Neural Comput, 2007. **19**(1): p. 1-46.
24. Gardiner, C.W., *Handbook of stochastic methods*. 3rd ed. 2003: Springer.