Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

How Recurrent Neural Information Flow Can Generate a Setpoint

luciano fernandes

fernandeslg@gmail.com

Washington State University https://orcid.org/0000-0003-2845-1500

Research Article

Keywords: neural setpoint, information flow, data-rate theorem, recurrent networks, excitation–inhibition balance, inhibition-stabilized networks

Posted Date: October 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11308235/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

How Recurrent Neural Information Flow Can Generate a Setpoint

Luciano Gonçalves Fernandes[1]

[1]Integrative Physiology and Neuroscience Department, College of Veterinary Medicine, Washington State University, Pullman, WA, 99163, USA.

Abstract

What sets the level to which neural activity returns remains open. In an exact mean-field model of excitatory and inhibitory quadratic integrate-and-fire neurons with fixed connections and no plasticity, target-rate variable or added sensing and effector circuit, we prove that the balance of activity around the excitatory–inhibitory loop fixes a unique level, which is exponentially stable if and only if inhibition feeds the state back to excitatory cells with a gain inside an exact window (**7/16** to **1187/512** in a solvable network) and, at unit gain, a delay below a critical value. Applying the data-rate theorem to the nonlinear network, we prove that a feedback channel of capacity $\boldsymbol{C}$ makes the level stable only if $\boldsymbol{C} \geq \boldsymbol{h}$, where $\boldsymbol{h}$ is the rate at which the network's instability generates information about its state (28.9 bits/s in a reference network), and that state-independent signals must fail. In finite simulations of the reference network and random inhibition-stabilized networks, an efficient coder held the level just above $\boldsymbol{h}$, never below. Through a white Gaussian channel, with no other noise, the linearized network's received signal reveals the state at exactly $\boldsymbol{h}$ for every noise amplitude and nonzero gain; by Bode's integral, the same rate fixes the regulation budget against fluctuating input. Spiking networks (10,000 neurons) lost the level under state-independent and rate-limited feedback and gave a cross-spectral integral consistent with $\boldsymbol{h}$. The level resists transient perturbations but shifts under sustained input: it is a settling point, generated by the flow of information through the network.

Keywords: neural setpoint, information flow, data-rate theorem, recurrent networks, excitation–inhibition balance, inhibition-stabilized networks

1 Introduction

To explain a regulated biological level, it is necessary to distinguish the processes that restore a variable from the processes that determine the level toward which the variable returns. Physiological regulation often involves several interacting effectors and operating levels that depend on context, rather than one literal comparator for each regulated variable (Ramsay and Woods 2014). In neural systems, synaptic scaling and the recovery of firing after sensory deprivation give evidence of activity regulation, and the manipulation of mitochondria can change the firing level around which this compensation occurs (Turrigiano et al. 1998; Hengen et al. 2013; Styr et al. 2019). These findings raise a precise question: can an attracting level of neural activity arise from the flow of information through a recurrent network itself, without an explicit target-rate variable and without a separately added sensing or effector circuit? It is important to note that such a network is not without feedback, since recurrent inhibition,

which carries the state of the network back to the excitatory cells, performs a feedback function.

Many influential models explain how activity is maintained once a reference has been specified. Orchestrated excitatory and inhibitory plasticity can produce self-sustained activity using explicit rate setpoints (Soldado-Magraner et al. 2022), and a self-organizing recurrent network studied in the context of criticality includes a desired average activity in its plasticity rules (Del Papa et al. 2017). These accounts explain regulation toward a reference, but not where the reference comes from. On the other hand, self-consistent activity in recurrent spiking networks shows that collective dynamics can fix a firing level (Brunel 2000; van Vreeswijk and Sompolinsky 1996), and inhibition can stabilize an excitatory network that would be unstable alone (Tsodyks et al. 1997; Ozeki et al. 2009; Sanzeni et al. 2020; Ahmadian and Miller 2021). What has been missing is a precise statement of *what* recurrent inhibition must transmit for the level to be held, and of *how much* of it.

Control and information theory give the tools to state this. The data-rate theorem says that a feedback loop can hold an unstable system at an equilibrium only if its channel can carry information about the state at least as fast as the instability generates it (Tatikonda and Mitter 2004; Nair and Evans 2004; Nair et al. 2007), and local versions of it hold for nonlinear systems (Nair et al. 2004; Colonius and Kawan 2009; Colonius 2012; Kawan 2013). For a linear loop whose only noise is that of a Gaussian channel, this rate is the information that the loop actually transmits (Elia 2004; Ardestanizadeh and Franceschetti 2012), and Bode's sensitivity integral converts it into a budget for the attenuation of disturbances (Freudenberg and Looze 1985; Martins and Dahleh 2008). Related limits govern the entropy reduction achievable by feedback (Touchette and Lloyd 2000), the suppression of molecular fluctuations in cells (Lestas et al. 2010), the trade-off between robustness and efficiency in metabolic regulation (Chandra et al. 2011), the speed and accuracy of sensorimotor control through delayed, rate-limited nerves (Nakahira et al. 2021) and, more generally, the information that a controller must receive to hold a stochastic system at a target steady state (Leighton et al. 2026). None of these results is ours. Our contribution is to show that, in a recurrent neural network model without an explicit target-rate variable or a separately added sensing and effector circuit, they turn the expression "the setpoint arises from information flow" into exact and testable statements: a feedback window, a minimum channel capacity for holding the level, obtained by a direct proof of the nonlinear data-rate bound for the network and nearly attained by an efficient coder in simulation, a critical delay, an information identity for a noisy inhibitory pathway of the linearized network, and a regulation budget, each tested in the population model and, except for the critical delay, in spiking networks.

This view also separates the information that holds a level from other uses of the word. Changes in excitation and inhibition alter information capacity and transmission in cortical preparations (Shew et al. 2011), and criticality has been proposed as a computational setpoint of brain function (Hengen and Shew 2025; Cramer et al. 2020; Ma et al. 2019). However, an information-processing regime and a firing-rate setpoint are distinct observables, and pairwise predictability statistics such as transfer entropy (Schreiber 2000) do not, by themselves, show that information holds a level.

We use the exact population reduction of heterogeneous quadratic integrate-and-fire neurons (Montbrió et al. 2015) to prove, for networks with fixed connections and no added regulatory mechanism, that (i) the flow of activity around the excitatory–inhibitory loop fixes a unique positive level; (ii) the level is attracting (exponentially stable) if and only if the current inhibitory state is fed back with a gain inside an exact window and, at unit gain, with a delay below a critical value; (iii) any feedback under which the level is stable needs a channel whose capacity is at least an explicit rate h, and state-independent signals must fail; (iv) through a white Gaussian channel in the inhibitory pathway, with no other noise, the received signal of the linearized network carries information about the initial state at exactly h; and (v) the same rate bounds how well the network regulates a fluctuating input. We test these statements in finite-time simulations: with an efficient coder, finite-rate codes, delays and noisy channels in the population model and in random inhibition-stabilized networks, and with gain changes, finite-rate codes, noisy channels and sinusoidal drives in spiking networks of 10,000

neurons (all-to-all; sparse networks for the gain and sustained-input experiments). These simulations sample finitely many initial states over finite times, so they are consistent with the theorems but cannot prove them. Finally, we state precisely what such a level is not: under sustained input it shifts, as expected for a network without an integrating motif (Francis and Wonham 1976; Sontag 2003). Taken together, these results show that a recurrent neural model can generate an attracting operating level without an explicit target-rate variable, and that maintaining its stability imposes a quantitative requirement on information capacity; they do not establish a general biological origin of setpoints. Following the distinction between a computational construction and an empirically grounded mechanistic explanation, the mapping of these results to physiological setpoints remains a testable hypothesis (Kaplan 2011).

2 Methods

Each symbol is named where it first appears; Table 1, at the end of this section, lists all of them. Every proof is given in full in Appendix A, and the protocols, certificates and sensitivity analyses that are not needed to follow the argument are in the Supplementary Information.

2.1 Operational definitions

We define a *local neural setpoint* as a positive equilibrium firing level that is locally exponentially stable under the autonomous network dynamics. For a network state x obeying $\dot{x} = F(x; p)$, where F is the vector field and p the parameter vector, and for a population firing observable $r(x)$, this requires

$$\begin{aligned} F(x^*; p) = 0, \quad r(x^*) > 0, \\ \|x(t) - x^*\| \le \gamma e^{-\alpha t} \|x(0) - x^*\|. \end{aligned} \tag{1}$$

for some constants $\gamma, \alpha > 0$ and all initial states $x(0)$ close enough to the equilibrium x^*, with $\|\cdot\|$ a norm on the state space; γ bounds the transient amplification and α is the rate of return (Strogatz 2015). No desired rate or error signal occurs in F, and no separately added sensing or effector circuit acts on the network. The parameters, however, determine the equilibrium: the absence of a comparator is not the absence of structural specification. Throughout the paper, transient inputs, changes of initial state and sustained input changes are treated as different interventions.

Setpoint and settling point. In this paper, *setpoint* means a level that satisfies Eq. (1): it is determined by the network, locally attracting, and defended against small transient perturbations of the state and of the input. Throughout, *attracting* means attracting in the sense of Eq. (1), that is, at an exponential rate. Physiology often reserves the word *setpoint* for a level that is also invariant under sustained load, and calls a level that is attracting but moves under sustained input a *settling point* (Wirtshafter and Davis 1977; Speakman et al. 2011). The level studied here is a setpoint in the first sense and a settling point in the second (Section 3.9).

Information flow. By information flow we mean the rate, in bits per second, at which the pathway from inhibitory to excitatory cells can carry information about the state of the network, that is, its capacity in the sense of the data-rate theorem (Nair et al. 2007); for a noisy pathway, we also use the rate at which the received signal accumulates information about the state (Section 3.6). A signal *transmits the state* when the signal received by excitatory cells depends on the current state of the network, and not only on its equilibrium value.

2.2 Finite spiking networks

The network has an excitatory population E of $N_{\mathrm{E}} = 8000$ neurons and an inhibitory population I of $N_{\mathrm{I}} = 2000$ neurons, both with heterogeneous excitability. The phase θ_j^X of neuron j of population $X \in \{\mathrm{E}, \mathrm{I}\}$ follows the theta form of the quadratic integrate-and-fire neuron (Izhikevich 2000; Montbrió et al. 2015),

$$\tau \dot{\theta}_j^X = 1 - \cos\theta_j^X + (1 + \cos\theta_j^X)\, I_j^X(t), \tag{2}$$

with membrane time constant τ and total input current $I_j^X(t)$. A forward crossing of π registers a spike, and the phase is reset by subtracting 2π (the reset from $+\infty$ to $-\infty$ of the quadratic integrate-and-fire neuron). The input currents are

$$I_j^{\mathrm{E}} = \eta_j^{\mathrm{E}} + \tau(J_{\mathrm{EE}} s_{\mathrm{E}} - J_{\mathrm{EI}} s_{\mathrm{I}}) + u_{\mathrm{E}}(t), \tag{3a}$$

$$I_j^{\mathrm{I}} = \eta_j^{\mathrm{I}} + \tau(J_{\mathrm{IE}} s_{\mathrm{E}} - J_{\mathrm{II}} s_{\mathrm{I}}) + u_{\mathrm{I}}(t), \tag{3b}$$

where η_j^X is the intrinsic excitability of the neuron, $J_{XY} \geq 0$ is the magnitude of the coupling from population Y to population X (the negative signs implement inhibition), $u_X(t)$ is the afferent current to population X, and s_X is the synaptic variable of population X, filtered with synaptic time constant τ_s:

$$\tau_s \dot{s}_X = -s_X + \frac{1}{N_X} \sum_{j=1}^{N_X} \sum_k \delta(t - t_j^k), \qquad (4)$$

with t_j^k the time of the k-th spike of neuron j and δ the Dirac delta. Each spike increments s_X by $1/(N_X \tau_s)$, so that s_X is in s^{-1}; the synaptic variables are physical states, not stored desired rates.

The excitabilities follow a Lorentzian distribution with center $\bar{\eta}_X$ and half-width Δ, as the reduction requires (Montbrió et al. 2015). In the all-to-all network they are the deterministic midpoint quantiles

$$\eta_j^X = \bar{\eta}_X + \Delta \tan\left[\pi \left(\frac{j + 1/2}{N_X} - \frac{1}{2}\right)\right], \quad j = 0, \ldots, N_X - 1. \qquad (5)$$

To check that the results do not depend on all-to-all coupling or on quantile placement, we also used sparse networks in which every neuron receives exactly $K = 200$ randomly chosen inputs from each population, with weights J_{XY}/K, and excitabilities drawn at random from the Lorentzian distribution. The autonomous runs use $u_\mathrm{E} = u_\mathrm{I} = 0$. There is no synaptic plasticity, desired-rate variable, rate-error term, separately added sensing or effector circuit, or optimization toward an activity target in these dynamics.

2.3 Population equations

Under the Lorentzian ansatz, the network admits a closed evolution for the firing rate and the center of the voltage distribution of each population (Montbrió et al. 2015), related to the reduction of Ott and Antonsen (2008). With the dimensionless rate $R_X = \pi\tau r_X$ and synaptic variable $S_X = \pi\tau s_X$, where r_X is the population firing rate in s^{-1}, the state of the population model is

$$x = (R_\mathrm{E}, V_\mathrm{E}, R_\mathrm{I}, V_\mathrm{I}, S_\mathrm{E}, S_\mathrm{I}), \qquad (6)$$

with V_X the dimensionless center of the voltage distribution, and the equations are

$$\begin{aligned} \tau \dot{R}_\mathrm{E} &= \Delta + 2R_\mathrm{E}V_\mathrm{E}, \\ \tau \dot{V}_\mathrm{E} &= V_\mathrm{E}^2 + \bar{\eta}_\mathrm{E} + u_\mathrm{E} - R_\mathrm{E}^2 \\ &\quad + \frac{J_{\mathrm{EE}} S_\mathrm{E} - J_{\mathrm{EI}} S_\mathrm{I}}{\pi}, \end{aligned} \qquad (7a)$$

$$\begin{aligned} \tau \dot{R}_\mathrm{I} &= \Delta + 2R_\mathrm{I}V_\mathrm{I}, \\ \tau \dot{V}_\mathrm{I} &= V_\mathrm{I}^2 + \bar{\eta}_\mathrm{I} + u_\mathrm{I} - R_\mathrm{I}^2 \\ &\quad + \frac{J_{\mathrm{IE}} S_\mathrm{E} - J_{\mathrm{II}} S_\mathrm{I}}{\pi}, \end{aligned} \qquad (7b)$$

$$\begin{aligned} \tau_s \dot{S}_\mathrm{E} &= R_\mathrm{E} - S_\mathrm{E}, \\ \tau_s \dot{S}_\mathrm{I} &= R_\mathrm{I} - S_\mathrm{I}. \end{aligned} \qquad (7c)$$

The reduction is exact for the infinite, globally coupled Lorentzian network on its invariant manifold (Montbrió et al. 2015; Ott and Antonsen 2008); finite and sparse networks are separate validation problems.

At a positive autonomous equilibrium, marked by an asterisk, $S_X^* = R_X^*$ and $V_X^* = -\Delta/(2R_X^*)$, and the remaining stationary conditions are

$$0 = \frac{\Delta^2}{4(R_\mathrm{E}^*)^2} + \bar{\eta}_\mathrm{E} + \frac{J_{\mathrm{EE}} R_\mathrm{E}^* - J_{\mathrm{EI}} R_\mathrm{I}^*}{\pi} - (R_\mathrm{E}^*)^2, \qquad (8a)$$

$$0 = \frac{\Delta^2}{4(R_\mathrm{I}^*)^2} + \bar{\eta}_\mathrm{I} + \frac{J_{\mathrm{IE}} R_\mathrm{E}^* - J_{\mathrm{II}} R_\mathrm{I}^*}{\pi} - (R_\mathrm{I}^*)^2. \qquad (8b)$$

Each line states that the rate of a population is the self-consistent response to the activity that flows into it through the synapses: the level is a fixed point of the loop rate → synaptic signal → input → rate. We use two parameter sets (Supplementary Table S1): an exactly solvable construction, chosen so that the proofs are tractable (Eq. (13)), and a reference network used in the simulations ($\Delta = 1$, $\bar{\eta}_\mathrm{E} = 1$, $\bar{\eta}_\mathrm{I} = -1$, $(J_{\mathrm{EE}}, J_{\mathrm{EI}}, J_{\mathrm{IE}}, J_{\mathrm{II}}) = (20, 15, 10, 3)$, $\tau = 20$ ms, $\tau_s = 10$ ms). Neither is fitted to a biological preparation.

For the reference network under a sustained excitatory current, the positive equilibria were located through the inhibitory rate $q = R_\mathrm{I} > 0$: the inhibitory stationary condition gives R_E as a function of q, and the excitatory condition then

gives the current at which q is an equilibrium,

$$R_{\mathrm{E}}(q) = \frac{\pi}{J_{\mathrm{IE}}}\left[q^2 + \frac{J_{\mathrm{II}}}{\pi}q - \bar{\eta}_{\mathrm{I}} - \frac{\Delta^2}{4q^2}\right],$$
$$u_{\mathrm{E}}(q) = R_{\mathrm{E}}(q)^2 - \frac{\Delta^2}{4R_{\mathrm{E}}(q)^2} - \bar{\eta}_{\mathrm{E}} - \frac{J_{\mathrm{EE}}R_{\mathrm{E}}(q) - J_{\mathrm{EI}}q}{\pi}. \quad (9)$$

We scanned u_{E} from -4 to 4 and classified stability from the full six-dimensional Jacobian.

2.4 Interventions on the inhibitory signal received by excitatory cells

Every intervention acts only on the inhibitory signal received by excitatory cells, $S_{\mathrm{I}}^{(\mathrm{E})}$, which replaces S_{I} in the excitatory voltage equation (and s_{I} in I_j^{E}). The inhibitory cells, and every other interaction, are unchanged. Interventions that need the equilibrium value S_{I}^* use, in spiking networks, each network's own time-averaged inhibitory signal (36.03 Hz), so that finite-size offsets do not act as an input.

Feedback gain and delay. $S_{\mathrm{I}}^{(\mathrm{E})}(t) = S_{\mathrm{I}}^* + g\,[S_{\mathrm{I}}(t-d) - S_{\mathrm{I}}^*]$, with feedback gain g and delay d ($d = 0$ unless stated). The received signal equals S_{I}^* at the equilibrium for every g and d, so that the level is unchanged; what changes is how strongly, and how late, the current deviation reaches excitatory cells. The value $g = 1$, $d = 0$ is the intact network; $g = 0$ replaces inhibition by its equilibrium value. For $g > 0$, a noiseless received deviation is an invertible rescaling of the sent one, so that it carries the same information about the state for every $g > 0$: the gain manipulates the *strength* with which the transmitted state acts, not its information content. Information is manipulated by the channels below.

Constant and replayed signals. The received signal is replaced by the mean of a recorded baseline inhibitory signal, or by a circularly shifted replay of that record (Supplementary Table S3). Both preserve the mean, and the replay also preserves the sample distribution and the Fourier magnitudes, but neither depends on the current state.

Efficient coder. To test whether the information bound of Theorem 1 is not only necessary but also sufficient, we used the achievability construction of the data-rate theorem, a "zooming" coder (Brockett and Liberzon 2000; Tatikonda and Mitter 2004; Nair and Evans 2004), as an experimental probe on the inhibitory pathway of the population model. An encoder observes the network state. A decoder runs a copy $\hat{x}$ of the intact network with the same parameters, driven by its own output, and delivers $S_{\mathrm{I}}^{(\mathrm{E})} = \hat{x}_6$ (its estimate of S_{I}) to the excitatory cells. Because the network and the copy receive the same signal, the error $x - \hat{x}$ evolves, to first order, with the clamped Jacobian. Each unstable mode of this Jacobian (one real mode in the reference network; a real mode or a complex pair in the random networks) is tracked by an interval, or a parallelogram for a complex pair, that both sides propagate with the clamped dynamics. Every sampling period T_{s}, the encoder sends the index of the cell, among n_{q} cells per axis, that contains the current modal error, and the decoder moves its copy to the center of that cell. The bit rate is $\mathcal{R} = d_{\mathrm{m}} \log_2 n_{\mathrm{q}}/T_{\mathrm{s}}$ for a mode of dimension $d_{\mathrm{m}} \in \{1, 2\}$, and, for linear dynamics, the uncertainty contracts if and only if $n_{\mathrm{q}} > e^{\mathrm{Re}\,\lambda\, T_{\mathrm{s}}}$, that is, if and only if $\mathcal{R}$ exceeds the mode's share of h (Eq. (18)). An allowance of 10^{-3} of the initial cell is added at each sample for the nonlinear terms. The copy starts at the equilibrium x^*, which the decoder computes from the network's parameters, and the initial cell is centred there, as the achievability construction assumes a known bounded initial set; beyond this, the copy holds no reference other than the network's own parameters: its equilibrium is the level of the network, fixed by the same balance of flows (Eq. (8); Proposition 2 for the reference network). We stress that the coder is a probe of how much information the pathway must carry, and not a model of how inhibitory neurons encode. In the reference network, we tested 29 combinations of T_{s} (20 to 200 ms) and n_{q} (2 to 90), with $\mathcal{R}/h$ between 0.69 and 1.73 (Supplementary Table S4), each from 24 initial states with independent Gaussian perturbations of relative size $a_0 = 0.02\%$, 0.2% or 2% of every coordinate, for 40 s; the initial cell had a half-width of 4.5 standard deviations of the perturbation.

Static finite-rate codes. In these codes the received signal is generated from samples of S_{I} taken every T_{s} and transmitted with a fixed number of bits. In the *mid-tread code*, each sample is quantized uniformly to n_{b} bits on $[0, 2S_{\mathrm{I}}^*]$, with a reconstruction level at S_{I}^*, and held, so that

$\mathcal{R} = n_\mathrm{b}/T_\mathrm{s}$; the *mid-rise code* is the same quantizer with S_I^* on a decision boundary. In the *one-bit code*, each sample transmits only the sign of $S_\mathrm{I} - S_\mathrm{rec}$, and the receiver moves its estimate S_rec by a fixed step $\beta\, S_\mathrm{I}^*$ in that direction, so that $\mathcal{R} = 1/T_\mathrm{s}$; S_rec starts at the value of S_I when the channel is switched on (in the population model, the perturbed initial value of each start). These codes are engineering probes; the one-bit code is not a model of spikes. In the population model, we tested 176 settings ($n_\mathrm{b} \in \{1, 2, 3, 4, 6, 8\}$ for both quantizers with T_s from 0.5 to 64 ms, and the one-bit code with $\beta \in \{0.005, 0.01, 0.02, 0.05, 0.1\}$ and T_s from 2 to 64 ms), each from 24 initial states with 2% Gaussian perturbations, for 3 s. In the all-to-all spiking network, the one-bit code ($\beta = 0.02$, gain 1) was run at $\mathcal{R} \in \{50, 100, 125, 150, 175, 200, 417, 1000, 2500\}$ bits/s (each rate realized with an integer number of 0.2-ms steps per sample; 150 and 175 were realized as 151.5 and 172.4 bits/s) in three networks with independent initial conditions; an excitatory pulse of ± 0.5 lasting 50 ms was applied 0.2 s after the channel was switched on, and each run lasted 4 s.

Inhibitory messengers. In the all-to-all spiking network, only the spikes of a random subset of N_msg inhibitory neurons, drawn among those whose stationary rate at the population equilibrium lies between 5 and 150 Hz, reach the excitatory cells, rescaled so that their predicted mean contribution equals the population value S_I^*. The mean signal, and with it the level, is nearly preserved (with all 2,000 messengers the excitatory rate was 40.99 Hz, 0.9% below the intact network, and verdicts are taken relative to that run), while the received signal becomes noisier as N_msg falls; the inhibitory cells still receive the full signal. For $N_\mathrm{msg} \le 10$, four further random subsets were tested.

Noisy channel at fixed gain. To measure the information that reaches the excitatory cells without changing the gain or the mean of the signal, we added white Gaussian noise to the inhibitory signal received by excitatory cells,

$$S_\mathrm{I}^{(\mathrm{E})}(t) = S_\mathrm{I}(t) + \pi\tau\,\sigma\,\xi(t), \tag{10}$$

where ξ is unit white Gaussian noise ($\mathbb{E}[\xi(t)\xi(t')] = \delta(t - t')$, with $\mathbb{E}$ the expectation), independent of the network, and σ is the noise amplitude in rate units ($\mathrm{Hz\,s^{1/2}}$). The gain is $g = 1$, and the inhibitory cells receive the noiseless signal. In the population model, we integrated the equations with the Euler–Maruyama method (step $\Delta t = 0.025$ ms) for 80 s, discarding the first 4 s, in eight trials for each $\sigma \in \{0.01, 0.02, 0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 1.0, 1.5\}$; the same eight noise realizations, scaled by σ, were used at every σ. In the all-to-all spiking network, we ran 60 s per condition, discarding the first 2 s of each record for both the spectra and the rates, for three network seeds and $\sigma \in \{0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 1.0\}$.

Sinusoidal afferent input. To evaluate how well the network regulates its level against a continuously changing input, we drove the excitatory cells with $u_\mathrm{E}(t) = U \sin(2\pi f t)$, with amplitude U and frequency f. The only feedback is the network's own inhibition, scaled by g. Over whole cycles after the transients, we report the amplitude of the excitatory rate at the driving frequency, the shift of the mean rate, and the root-mean-square (RMS) deviation from the level. The population runs used $U \in \{0.25, 1, 2\}$, f from 0.5 to 80 Hz and $g \in \{0.5, 0.7, 1, 1.5\}$; the spiking runs (all-to-all, 4 s, last 3 s analyzed, 5-ms bins) used $U = 0.5$ and $f \in \{1, 2, 5, 10, 20, 40\}$ Hz at $g = 1$ and 0.6, and $U = 2$ at $g = 1$. The channel experiments used $f = 5$ Hz and $U = 0.5$ (12 starts with 2% perturbations per setting, last 2 s of 3 s); the delay experiments used the same drive in one run per delay, started at the level (3 s; whole cycles in the second half of the run). The linear predictions use the closed-loop transfer function from u_E to r_E of the reference network Jacobian.

2.5 When is the level held?

We used prespecified criteria to classify whether the level was held, with a separate criterion for deterministic population runs without drive. For runs evaluated by window statistics, the excitatory rate in the final window must (i) have a mean within 5% of the level and (ii) fluctuate no more than ten times what the imposed input alone produces in the linearized intact loop. For deterministic population runs without drive, we chose a different criterion: r_E must stay within 5% of the level throughout the final window. This criterion implies (i), but it is not equivalent to (i)–(ii); deterministic oscillations around the level are possible in this model (for example, beyond

the upper boundary of the gain window), and the criterion accepts them only when their excursions stay within 5% of the level. For runs with noise, (ii) bounds the standard deviation (s.d., 10-ms bins in spiking networks) by ten times the s.d. expected in the linearized intact loop: $[0.54^2 + (20.6\,\sigma)^2]^{1/2}$ Hz in spiking networks, where 0.54 Hz is the finite-size s.d. of the intact network and $20.6\,\sigma$ the contribution of the channel noise, and $22.6\,\sigma$ in the population model, in both cases with a floor of 0.54 Hz. For runs with a periodic drive, (i) applies to the mean over whole cycles, and (ii) to the s.d. of the part of r_{E} that is not locked to the drive (after removing the first five harmonics), with the same floor of 0.54 Hz. The s.d. condition excludes fluctuations, including self-sustained oscillations, whose standard deviation exceeds the adopted limit; smaller oscillations can meet this criterion. The final window is the last second of the static-code runs and of the spiking gain experiments; the last 2 s of the spiking one-bit and messenger runs and of the population runs with drive and channels; for the other population runs with drive, the whole cycles within the last four periods of the drive (or within the second half of the run, when the feedback was delayed); the last 3 s of the spiking runs with drive; the last 10% of each efficient-coder run; the last quarter of each random-network run; and, for the noisy channel, the whole record after the initial transient (the first 4 s in the population model; in spiking networks, the first 2 s of each record, that is, $t > 3$ s counted from the start of the 1-s warm-up). In the gain experiments of four of the five all-to-all spiking networks and of four of the five sparse networks, and in the driven runs of four of the five spiking networks, only the mean rates (and, for the drive, the amplitudes) were stored, so that only (i) could be applied there; where traces were stored (44 runs: the 20 gain runs of the all-to-all network of seed 0, the 8 gain runs of the sparse network of seed 11 and 16 driven runs of seed 0, listed in Supplementary Table S3), adding (ii) changed no verdict; in the gain experiments, every loss of the level was a move to another state, which (i) detects, and in the driven runs that met (i) the amplitude agreed with the linear prediction (Section 3.5). The 5% band is an analysis tolerance, not a physiological claim or a desired rate supplied to the neurons. Supplementary Table S5 gives every verdict under tolerances of 5, 10 and 20%, with and without condition (ii).

2.6 Information estimates

Gaussian spectral estimate. Let $y(t) = S_{\mathrm{I}}^{(\mathrm{E})}(t)/(\pi\tau)$ be the received signal in rate units. For the stationary linearized loop of Theorem 4, in which the channel noise is the only noise, the rate at which the received signal accumulates information about the initial state equals $\frac{1}{4\pi}\int \ln[\Phi_y(\omega)/\sigma^2]\,\mathrm{d}\omega$, where Φ_y is the power spectral density of y at angular frequency ω; with intrinsic noise, the same expression is larger (Corollary 2). From a record, we computed the *Gaussian spectral estimate*

$$\widehat{I}_{\mathrm{G}} = \frac{1}{\ln 2}\int_0^{f_{\max}} \left[\ln \frac{\widehat{\Phi}_y(f)}{\sigma^2} - \big(\psi(n_{\mathrm{seg}}) - \ln n_{\mathrm{seg}}\big)\right] \mathrm{d}f. \tag{11}$$

in bits/s, where $\widehat{\Phi}_y(f)$ is the estimated two-sided spectral density at frequency f (Welch's method with a Hann window and n_{seg} non-overlapping segments of 1.64 s; Welch 1967), ψ is the digamma function, so that $\psi(n_{\mathrm{seg}}) - \ln n_{\mathrm{seg}}$ removes the bias of the logarithm of an average of n_{seg} periodograms, and $f_{\max} = 100$ Hz, which contains 99.92% of the theoretical integral for the reference network; the interval below the first frequency bin takes the value of that bin. On the linearized reference network loop, where the exact answer is $h = 28.91$ bits/s, the estimator gave values between 28.1 and 29.5 bits/s for $g \in \{0.5, 1, 1.5\}$ and $\sigma \in \{0.01, 0.1, 1\}$ (each value from the spectrum pooled over eight trials of 100 s). Outside the stationary linear Gaussian case (nonlinear dynamics, spiking networks, finite records, a truncated band), $\widehat{I}_{\mathrm{G}}$ is an estimate whose relation to the information rate is not certified; we therefore use it as a consistency check and report how it depends on $f_{\max}$ (Supplementary Table S6). Errors are empirical standard errors across trials (s.d. with $n-1$ in the denominator, divided by $\sqrt{n}$). In the population model with channel noise, $\widehat{I}_{\mathrm{G}}$ was computed for each trial, and its mean and standard error were taken over all eight trials at each σ, whatever the verdict of each trial.

Cross-spectral estimate of the sensitivity. Because the injected noise ξ is known (it was regenerated from its seed), the received signal can

be correlated directly with it. We computed the cross-spectral ratio

$$\widehat{S}(f) = \frac{\widehat{\Phi}_{y\xi}(f)}{\sigma\,\widehat{\Phi}_{\xi\xi}(f)}, \tag{12}$$

where $\widehat{\Phi}_{y\xi}$ is the estimated cross-spectral density of the received signal and the noise and $\widehat{\Phi}_{\xi\xi}$ the spectral density of the noise. If the received signal is written as $y = \sigma S_g \xi + y_{\mathrm{r}}$, with $S_g\xi$ the injected noise filtered by the sensitivity function of the linearized loop and y_{r} the residual, the spectral densities satisfy $\Phi_{y\xi}/(\sigma\Phi_{\xi\xi}) = S_g + \Phi_{y_{\mathrm{r}}\xi}/(\sigma\Phi_{\xi\xi})$ at each frequency; thus, $\widehat{S}$ is an estimator of S_g only if the residual is uncorrelated with the injected noise, a condition that knowing the realization of ξ does not, by itself, establish. In the linearized loop of Corollary 2, the residual is the response to the intrinsic noise $\mathcal{U}$, which is independent of ξ, so that $\int_0^{f_{\max}} \log_2 |\widehat{S}(f)|^2 \,\mathrm{d}f$ estimates $\int_0^{\infty} \log_2 |S_g|^2 \,\mathrm{d}f = h$ (Eq. (22)) up to the truncation at $f_{\max}$ and, in the limit of long records, whatever the intrinsic fluctuations; in a finite record, the intrinsic fluctuations increase the variance of $\widehat{S}$ and, through the logarithm, tend to bias the integral upward. In a nonlinear network, the residual also contains nonlinear distortions of the response to ξ and any difference between the linear response of the finite network and S_g, which are in general correlated with ξ; the ratio then describes the best linear approximation, for the given excitation, of the transfer from the injected noise to the received signal, and it depends on the amplitude and the spectrum of that noise (Schoukens et al. 2005; Pintelon and Schoukens 2012). For this reason, we use the integral computed from spiking data, which we call the *empirical cross-spectral integral*, as a test of consistency with the population prediction h, examining its dependence on $f_{\max}$ and on σ, and not as an independent measurement of the instability rate of the finite network.

Power and capacity. The power of the state-dependent part of the received signal, P_{s}, was computed as the variance of the sent signal $S_{\mathrm{I}}/(\pi\tau)$. A white Gaussian channel with this power and noise amplitude σ has capacity $C = P_{\mathrm{s}}/(2\sigma^2 \ln 2)$ bits/s (Cover and Thomas 2006).

2.7 Random networks

To verify that the results do not depend on the chosen parameters, we drew random population models with $\Delta = 1$, $\tau = 20$ ms and the other parameters uniform on $\bar{\eta}_{\mathrm{E}} \in [-1, 2]$, $\bar{\eta}_{\mathrm{I}} \in [-2, 1]$, $J_{\mathrm{EE}}, J_{\mathrm{EI}} \in [5, 30]$, $J_{\mathrm{IE}} \in [5, 20]$, $J_{\mathrm{II}} \in [0, 10]$ and $\tau_s \in [5, 15]$ ms. A network was retained if it had exactly one positive equilibrium (located with Eq. (9)), an excitatory rate between 1 and 100 Hz, a stable equilibrium with intact feedback, and an equilibrium that became unstable when the inhibitory signal received by excitatory cells was clamped ($g = 0$). With clamped inhibition the Jacobian is block triangular, and its eigenvalues are those of the excitatory block (coordinates $R_{\mathrm{E}}, V_{\mathrm{E}}, S_{\mathrm{E}}$) and of the inhibitory block. We analyzed the inhibition-stabilized networks, whose excitatory block is unstable, and computed h from the unstable eigenvalues of the excitatory block; networks whose instability came only from the inhibitory block were excluded. The draw stopped after 74 networks had passed the first selection.

For each network we computed the gain threshold g_c (the smallest g at which the leading eigenvalue of the Jacobian crosses zero), the critical delay d_c at $g = 1$ (from the characteristic quasi-polynomial $\chi_0(\lambda) + e^{-\lambda d}[\chi_1(\lambda) - \chi_0(\lambda)]$, exact because g enters the Jacobian through a single entry), and the derivative $\mathrm{d}r_{\mathrm{I}}^*/\mathrm{d}u_{\mathrm{I}}$ of the inhibitory rate with respect to a current injected into inhibitory cells, which is negative when the network shows the paradoxical response (Tsodyks et al. 1997; Sanzeni et al. 2020). We then ran three experiments. (1) The one-bit code ($\beta = 0.02$) at all 25 rates of a logarithmic grid from 5 to 5000 bits/s (adjacent rates differ by a factor of 1.33; each rate realized with an integer number of 0.05-ms steps per sample), from 12 initial states with 2% perturbations, for $\max(2\,\mathrm{s}, 10/\lambda_u)$, with λ_u the unstable eigenvalue of the clamped network (Proposition 5), so that the instability had time to act even in the slowest networks; we report $\mathcal{R}_{\mathrm{hold}}$, the lowest grid rate from which every higher grid rate held the level from all 12 starts. (2) A zero-information control with the received signal fixed at S_{I}^*, under the same protocol. (3) The efficient coder with $n_{\mathrm{q}} = 2$ and T_{s} chosen so that $\mathcal{R}/h \in \{0.8, 0.95, 1.05, 1.2, 1.5\}$, from 12 starts with perturbations of 0.02% and 2%, for $\max(4\,\mathrm{s}, 250\,T_{\mathrm{s}})$, with T_{s} the sampling period

at $\mathcal{R} = 0.95\,h$ (the same duration for every setting of a network; time step 0.25 ms). Correlations are Spearman's ρ, with 95% confidence intervals from 10,000 bootstrap resamples of the networks (Efron and Tibshirani 1993); the partial correlation removes the ranks of a third variable by linear regression.

2.8 Conditional information assay

Transfer entropy measures the information that a source history supplies about the next value of a target, conditional on the target's own history (Schreiber 2000). In the original reference network experiments we estimated its linear-Gaussian proxy (Barnett et al. 2009), $\max\{0, \frac{1}{2}\log_2(\widehat{\sigma}^2_{\mathrm{reduced}}/\widehat{\sigma}^2_{\mathrm{full}})\}$ bits per bin, from the residual variances of least-squares predictions of the next 10-ms bin without and with three source bins (three target bins, and the known current probes at the current and three prior bins, entered both predictions), and compared it with 99 circular shifts of the source (Vicente et al. 2011). The protocol is in Supplementary Table S3.

2.9 Numerical methods and certificates

The spiking simulations integrate Eq. (2) exactly within each 0.2-ms step, holding the synaptic input constant during the step (for input $I > 0$, $\tan(\theta/2)$ advances along $\sqrt{I}\tan(\sqrt{I}\,t/\tau + c)$, with c fixed by the phase at the start of the step; for $I < 0$, the tanh or coth form is used); spikes are added to the synaptic variables, which decay exactly. With this integrator, the all-to-all baseline was 41.36/36.02 Hz, in agreement with the original fourth-order Runge–Kutta integrator (41.36/36.03 Hz). Deterministic population trajectories use the Dormand–Prince method (LSODA for the runs beyond the upper gain boundary), split at input discontinuities (Dormand and Prince 1980; Virtanen et al. 2020), or fixed-step fourth-order Runge–Kutta where channels or delays are present (steps of 0.02 to 0.25 ms, listed in Supplementary Table S3); the gain–delay map of Fig. 2d integrates the linearized delay equation with the Euler method (step 0.02 ms). The first Lyapunov coefficients of the four Hopf points of the population equations without delay (Supplementary Table S8a) were computed with Kuznetsov's formula (Kuznetsov 2004); since the vector field is quadratic in the state, the bilinear term is exact and the cubic term vanishes. The symbolic results were obtained with SymPy (Meurer et al. 2017) and checked by direct numerical evaluation (`proofs_check.py`). The uniqueness of the reference network level (Proposition 2) uses a cover certified in interval arithmetic with 30 significant digits and outward rounding (Supplementary Table S7), and, at $u_{\mathrm{E}} = -3$, 0 and 3, interval calculations checked a contraction map enclosing a unique root and a quadratic Lyapunov inequality for the nonlinear equations (Khalil 2002; mpmath developers n.d.) (Supplementary Section S10). The basin assessment used 64 Latin hypercube points and three fixed starts at each input (SciPy developers n.d.). No parameter was retuned when a check failed. The code uses Python, NumPy, SciPy and Matplotlib (Harris et al. 2020; Virtanen et al. 2020; Hunter 2007), and the script `paper_numbers.py` recomputes every number quoted in the Results from the stored data.

Table 1 Notation. Each symbol is named once in the text, at the place given in the last column; symbols used only inside one proof are named in that proof (Appendix A). Rates written R are dimensionless; rates written r are in s^{-1}; information rates are in bits/s unless stated

Symbol	Meaning	Where
Network		
E, I; X, Y	excitatory and inhibitory populations; generic population labels	Sec. 2.2
N_X; K	number of neurons in population X; in-degree per population of the sparse network	Sec. 2.2
θ_j^X, I_j^X, η_j^X	phase, input current and excitability of neuron j	Eqs. (2)–(3)
τ; τ_s	membrane time constant; synaptic time constant	Eqs. (2), (4)
J_{XY}	coupling magnitude from population Y to population X	Eq. (3)
u_X	afferent current to population X	Eq. (3)
s_X, r_X	synaptic variable and firing rate of population X (s^{-1})	Eq. (4)
$\bar{\eta}_X$, Δ	center and half-width of the Lorentzian excitability distribution	Eq. (5)
R_X, V_X, S_X	dimensionless rate, voltage center and synaptic variable	Eq. (6)
x, x^*; F, p	population state and its equilibrium; vector field and parameters	Eqs. (1), (6)
γ, α	transient bound and rate of return in the definition of a setpoint	Eq. (1)
q	inhibitory equilibrium rate used as continuation coordinate	Eq. (9)
Interventions		
$S_{\mathrm{I}}^{(\mathrm{E})}$	inhibitory signal received by excitatory cells	Sec. 2.4
g; g_c, g_H	feedback gain; lower (gain threshold) and upper boundary of the stable range	Sec. 2.4, Prop. 4
d; d_c	feedback delay; critical delay	Sec. 2.4, Prop. 6
T_{s}, n_{q}, n_{b}	sampling period; cells per axis of the coder; bits per sample of a static code	Sec. 2.4
$\mathcal{R}$; $\mathcal{R}_{\mathrm{hold}}$	channel bit rate; lowest grid rate from which every higher rate held	Secs. 2.4, 2.7
β; a_0	step of the one-bit code (relative to S_{I}^*); relative size of the initial perturbation	Sec. 2.4
$\hat{x}$; d_{m}	decoder copy of the network state in the efficient coder; dimension of the coded mode	Sec. 2.4
S_{rec}	receiver estimate of the one-bit code	Sec. 2.4
N_{msg}	number of inhibitory messengers	Sec. 2.4
σ; ξ	noise amplitude of the channel ($\mathrm{Hz}\,\mathrm{s}^{1/2}$); unit white Gaussian noise	Eq. (10)
U, f	amplitude and frequency of the sinusoidal drive	Sec. 2.4
Δt	integration step	Sec. 2.4
Linearization and information		
$M(g)$, M_0; χ_g	dimensionless Jacobian at gain g, and at $g = 0$; its characteristic polynomial	Eq. (15)
$P(\lambda)$	characteristic polynomial of one uncoupled population	Eq. (15)
v_c, w_c; ς; Θ_1, Θ_2	right and left null vectors of the Jacobian at g_c; amplitude of the critical mode; coefficients of its reduced equation	Sec. 3.2
λ; λ_u	complex frequency (eigenvalue variable); unstable eigenvalue with clamped inhibition	Props. 3, 5
z, F_{E}, A, b, v	excitatory deviation, its vector field, Jacobian and input vector; received deviation	Eq. (17)
m, λ_i, E_u	number of unstable eigenvalues of A, the eigenvalues, their invariant subspace	Eq. (17)
M_{E}	dimensionless Jacobian of the excitatory population with clamped inhibition	Eq. (17), Appendix A.5
ε, ε_0	radii in the definition of Lyapunov stability	Thm. 1

Table 1, continued

Symbol	Meaning	Where
C, $N(t)$, c_0	channel capacity; number of distinguishable symbol sequences; constant of the channel model	Eq. (17)
h	minimum information rate	Eq. (18)
T_g, S_g, L_g	transfer function from u_{E} to δR_{E}; sensitivity function; loop function	Eqs. (19)–(20)
ω	angular frequency	Eq. (21)
ζ, w, $\mathcal{Y}$, W	state deviation, sent inhibitory deviation, cumulative received signal, Wiener process	Eq. (24)
b_6, e_k, Σ_0	input column, k-th unit vector, covariance of the initial state	Eq. (24)
$\bar{I}$; $\widehat{I}_{\mathrm{G}}$	information rate of Theorem 4; its Gaussian spectral estimate	Eqs. (25), (11)
Φ_y, Φ_v; $\widehat{S}$; y_{r}	spectral densities of received signals; cross-spectral ratio (an estimator of S_g only if y_{r} is uncorrelated with ξ); residual of the received signal	Eqs. (11), (12)
$f_{\max}$, n_{seg}, ψ	upper frequency, number of spectral segments, digamma function	Eq. (11)
P_{s}	stationary power of the state-dependent part of the received signal	Sec. 2.6
$\mathcal{U}$, Γ	intrinsic noise process and the matrix through which it enters	Cor. 2
$\mathsf{X}, \mathsf{Y}, \mathsf{Z}$; $\alpha_1, \beta_1, \beta_2$; ϵ_n, ν^2, μ	variables, coefficients, innovation and its variance, conditional variance of the observation model	Eq. (29)
$A_{\mathrm{o}}, B_{\mathrm{o}}, \Sigma_{\mathrm{o}}, C_{\mathrm{o}}, x_{\mathrm{o}}$	matrices and state of a generic linear system	Prop. 9

3 Results

Table 2 summarizes what is proved, what is tested, and where. The proofs are in Appendix A.

3.1 The flow of activity around the loop fixes a unique level

The framework is that of Montbrió et al. (2015) (Fig. 1a); the construction and the proofs are ours.

Proposition 1 (Unique positive equilibrium) *For* $\Delta = 1$, $\bar{\eta}_{\mathrm{E}} = \bar{\eta}_{\mathrm{I}} = -1/4$, $\tau_s = \tau/2$ *and*

$$(J_{\mathrm{EE}}, J_{\mathrm{EI}}, J_{\mathrm{IE}}, J_{\mathrm{II}}) = (3\pi, 2\pi, 2\pi, \pi), \tag{13}$$

the autonomous population equations (7) *have exactly one equilibrium with both rates positive,*

$$\begin{aligned} x^* &= (1, -1/2, 1, -1/2, 1, 1), \\ r^*_{\mathrm{E}} &= r^*_{\mathrm{I}} = \frac{1}{\pi\tau}. \end{aligned} \tag{14}$$

Proof: Appendix A.1.

For $\tau = 20$ ms, the level is $r^* = 1/(\pi\tau) \approx 15.9$ Hz. This value is an algebraic consequence of Eq. (14), not a fitted prediction, and it is produced by the recurrent flow and not by the cells alone: without coupling, the same population would satisfy $4R^4 + R^2 - 1 = 0$, whose positive root, $R = [(\sqrt{17} - 1)/8]^{1/2} \approx 0.625$, corresponds to 9.9 Hz. No state variable, input or parameter of the network is a reference against which the rate is compared. The reference network, used in all simulations, also has a unique level for every drive.

Proposition 2 (Unique level in the reference network) *For the reference network parameters and every excitatory drive* $u_{\mathrm{E}} \in \mathbb{R}$, *Eq.* (7) *has exactly one equilibrium with positive rates.*

Proof: Appendix A.2; the interval-arithmetic certificate is Supplementary Table S7.

At $u_{\mathrm{E}} = 0$, the reference network level is 41.25/36.07 Hz (excitatory/inhibitory).

3.2 Attraction requires that inhibition feed the current state back

Proposition 3 (Local exponential attraction) *The equilibrium of Proposition 1 is locally exponentially stable in the full six-dimensional population state; the eigenvalues of the linearization are* $-1/\tau$ *and* $(-3 \pm i\sqrt{15})/(2\tau)$, *each with algebraic multiplicity two.*

Proof: Appendix A.3.

This implies recovery from small displacements of the state and, by the continuity of solutions over finite times, recovery after any sufficiently weak input of finite duration; it does not, by itself, prove recovery from every large input or global attraction.

We now scale the state-dependent part of the inhibitory signal received by excitatory cells, $S_{\mathrm{I}}^{(\mathrm{E})} = S_{\mathrm{I}}^* + g\,(S_{\mathrm{I}} - S_{\mathrm{I}}^*)$, which leaves the level unchanged. Let $M(g)$ be the Jacobian of Eq. (7) at x^*, in the time unit τ, with $M_0 = M(0)$, and let $\chi_g(\lambda) = \det(\lambda I - M(g))$ be its characteristic polynomial in the complex frequency λ (units of $1/\tau$). For the exact construction,

$$\begin{aligned} \chi_g(\lambda) &= \big(P(\lambda) - 4\big)^2 - 64(1 - g), \\ P(\lambda) &= (\lambda + 2)\big[(\lambda + 1)^2 + 4\big]. \end{aligned} \tag{15}$$

where P is the characteristic polynomial of one uncoupled population (Appendix A.4).

Proposition 4 (Exact feedback window) *In the exact construction with* $g \geq 0$, *the equilibrium* x^* *is the same for every* g, *and it is locally exponentially stable if and only if*

$$\frac{7}{16} < g < \frac{1187}{512} \approx 2.318. \tag{16}$$

For $g < 7/16$ *the linearization has a positive real eigenvalue; at* $g = g_H = 1187/512$ *a pair of complex eigenvalues* $\pm i\,\omega_H$, *with* $\omega_H = 1.2247/\tau$, *lies on the imaginary axis.*

Proof: Appendix A.4.

Proposition 5 (Instability after constant inhibitory replacement) *If* S_{I} *is replaced by its equilibrium value only in the excitatory voltage equation of the exact construction* ($g = 0$), *the equilibrium* x^* *is unchanged but unstable, with a single eigenvalue in the closed right half-plane,* $\lambda_u \approx 0.20298/\tau$, *the positive root of* $\lambda^3 + 4\lambda^2 + 9\lambda - 2$.

Proof: Appendix A.5.

Propositions 4 and 5 separate the value of the level from its attraction. The value is set by

Table 2 Summary of the results. "Proved" refers to the exact construction unless stated; "tested" lists the simulations that probe each statement

Statement	Proved	Tested
The balance of activity flows fixes a unique level ($r^* = 1/(\pi\tau)$ in the exact construction; 41.25 Hz, computed, in the reference network)	Props. 1, 2	spiking baselines within 0.3% (Sec. 3.8)
Without delay, the level attracts (Eq. (1)) if and only if the current inhibitory state is fed back with $7/16 < g < 1187/512$	Props. 3–5; type of each boundary computed	population model beyond both boundaries; spiking gain experiments with $g \leq 1$ (lower threshold only), all-to-all and sparse (Sec. 3.2)
At $g = 1$, it attracts if and only if $d < d_c = 1.5675\,\tau$ (31.3 ms for $\tau = 20$ ms; 13.3 ms, computed, in the reference network)	Prop. 6	population simulations beyond d_c (Sec. 3.3)
Any feedback under which the level is stable needs a channel of capacity at least h; state-independent signals fail	Thm. 1 (nonlinear, local), Cor. 1, Thm. 2	finite-time runs consistent with it: no channel below h held in 176 code settings, 29 coder settings, and 397 tests in the 63 random networks with $h > 5$ bits/s; constant and replayed signals (Secs. 3.2, 3.4, 3.7)
An efficient coder nearly attains the bound	— (simulation only)	held at $1.02\,h$, failed at $0.99\,h$ (reference network, small perturbations); 97.5% of starts at $1.05\,h$, none at $0.95\,h$ (the 63 random networks whose clamped instability is confined to E)
The information rate fixes the regulation budget	Thm. 3	sinusoidal drives in population and spiking networks, through channels and with delays (Sec. 3.5)
Through a white Gaussian channel, with no other noise, the received signal of the linearized network reveals the state at exactly h, for every noise amplitude and nonzero gain	Thm. 4 (linearized network), Cor. 2	population model, with the linear loop as control; spiking networks: empirical cross-spectral integral, 29.4 ± 2.1 bits/s at weak noise, consistent with the population value of h but not an independent measurement of the network's instability rate, nor of the information received (Sec. 3.6)
Sustained input moves the level	Prop. 8	maintained input, all-to-all and sparse (Secs. 3.8, 3.9)
Pairwise Shannon measures do not see the level	Lemma 1, Prop. 9	probe assay (Sec. 3.11)

the stationary balance of flows, Eq. (8), and does not depend on g. The attraction depends only on how strongly the *current* inhibitory state acts on the excitatory cells: with too little feedback the excitatory loop carries the state away from the level, and with too much the loop overshoots into oscillation. The reference network behaves in the same way. Its intact spectrum has leading real part $-26.4\ \mathrm{s}^{-1}$; the clamped spectrum has one unstable eigenvalue, $\lambda_u = 20.04\ \mathrm{s}^{-1}$; and the level is stable for $g_c < g < g_H$, with gain threshold $g_c = 0.420$ and upper boundary $g_H = 2.303$ (Fig. 1c).

The two boundaries differ in nature. At the lower boundary a simple real eigenvalue crosses zero, while the other eigenvalues keep negative real parts and the level remains an equilibrium for every g. These facts alone do not determine the type of the bifurcation, and we therefore verified the nondegeneracy conditions of a transcritical bifurcation (Kuznetsov 2004). Let v_c and w_c be the right and left null vectors of the Jacobian at g_c, normalized so that the R_E component of v_c is 1 and $w_c \cdot v_c = 1$. On the center manifold, the amplitude ς of the critical mode obeys $\dot{\varsigma} = \Theta_1 (g - g_c)\,\varsigma + \Theta_2\,\varsigma^2 + \ldots$, with $\Theta_1 = w_c \cdot (\partial_g D_x F)\, v_c$, the rate at which the critical eigenvalue changes with the gain, and $\Theta_2 = \frac{1}{2}\, w_c \cdot D_x^2 F(v_c, v_c)$; the bifurcation is transcritical if both coefficients are nonzero. In the exact construction, computed in rational arithmetic, $\Theta_1 = -16/27$ and $\Theta_2 = -5/42$ (in units of $1/\tau$), and the other eigenvalues at $g_c = 7/16$ have real parts of at most $-0.82/\tau$; in the reference network, $\Theta_1 = -61.4\ \mathrm{s}^{-1}$ and $\Theta_2 = -19.4\ \mathrm{s}^{-1}$, with the other real parts at most $-19.5\ \mathrm{s}^{-1}$ (Supplementary Section S7). Both conditions therefore hold, and the slope of the second branch predicted by the reduced equation, $-\Theta_1/\Theta_2$ in R_E per unit of gain, that is, -3.17 in the reference network, or -50.5 Hz of $r_\mathrm{E} = R_\mathrm{E}/(\pi\tau)$ per unit of gain, agrees with the computed branch within 0.5% at $g = 0.41$ and 0.43. In the population model of the reference network, this branch is a stable state above the level for $g < g_c$ (59.8, 47.0 and 42.2 Hz at $g = 0$, 0.3 and 0.4) and, for $g > g_c$, an equilibrium below the level with exactly one eigenvalue of positive real part (40.7 and 37.0 Hz at $g = 0.43$ and 0.5); a low state of 2.7–3.9 Hz is stable for every g from 0 to 0.5 (Supplementary Table S8). Below g_c, the level itself has exactly one unstable eigenvalue, and small perturbations of opposite signs led to the upper state and to the low state. Above g_c, the stable manifold of the unstable equilibrium has codimension one and is a natural candidate for part of the boundary of the basin of the level; we did not compute this manifold, and it is important to note that no single firing rate characterizes the basin. At $g = 0.43$, where this equilibrium has an excitatory rate of 40.7 Hz, starts that differed from the level only in R_E returned to the level unless r_E was lowered below 31.4 Hz, whereas starts that differed from it only in the synaptic variable S_E went to the low state as soon as S_E was lowered below the value that corresponds to 40.8 Hz; at $g = 0.5$, the two values were 12.2 and 37.8 Hz, against 37.0 Hz for the equilibrium (`transcritical_check.py`). Two-dimensional sections of the basin, on grids of starts with rates from 0.5 to 120 Hz, are in Supplementary Fig. S1: with intact feedback every start returned to the level, whereas at $g = 0.5$ and 0.43 between 60 and 68% of the starts did and the remaining classified starts went to the low state, with two numerical failures at $g = 0.43$ left unresolved and the sampled boundary passing close to the level. With inhibition clamped at its equilibrium value ($g = 0$), a 1% upward perturbation of the excitatory rate sent the network to 59.8/45.7 Hz and a 1% downward perturbation to 2.7/7.5 Hz (Fig. 1b). The open-loop network thus amplifies arbitrarily small differences of its state into macroscopically different outcomes, at rate λ_u; in the language of information theory, it generates information about its initial condition at $\lambda_u / \ln 2$ bits/s. At the upper boundary a Hopf bifurcation occurs, and it is supercritical: the first Lyapunov coefficient (Kuznetsov 2004) is -0.235 in the exact construction and -0.053 in the reference network (where the crossing frequency is 23.5 Hz). Accordingly, just beyond the boundary a small stable oscillation replaces the level, and its amplitude grows as the square root of the distance: for gains 0.5, 1, 2 and 4% above g_H, the excitatory rate oscillated with amplitudes of 0.119, 0.168, 0.237 and 0.334 times the level in the exact construction, and of 5.5, 7.7, 10.9 and 15.4 Hz in the reference network (Supplementary Table S8).

Spiking networks.

In the all-to-all network of 10,000 neurons, the outcomes were consistent with the gain threshold of the population model, which these runs bracket but do not locate (Fig. 1d; Fig. 7c). After pulses

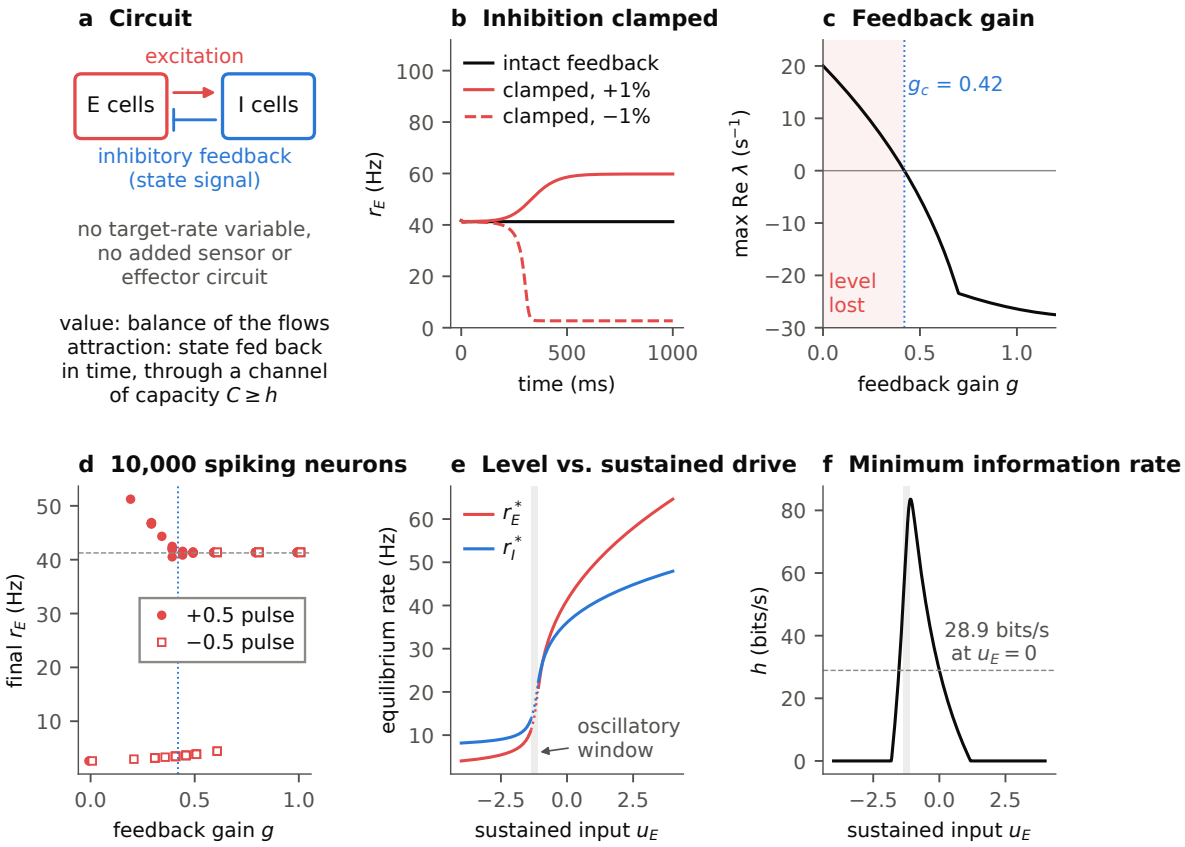


Fig. 1 The level and its attraction. (a) Excitatory (E) and inhibitory (I) populations with fixed connections; there is no target-rate variable and no added sensor or effector circuit, but recurrent inhibition nevertheless performs a feedback function: it is the network's own feedback. The value of the level is fixed by the balance of the activity that flows around the loop, and its attraction requires that inhibition feed the current state back to E cells. (b) Reference network population model. With intact feedback the excitatory rate stays at the level (41.25 Hz); with the inhibitory signal received by E cells clamped at its equilibrium value, a perturbation of +1% or −1% of the excitatory rate sends the network to 59.8 or 2.7 Hz. (c) Largest real part of the eigenvalues of the Jacobian against the feedback gain g; the level is unstable below $g_c = 0.42$ (the upper boundary, $g_H = 2.30$, is outside the axis). (d) All-to-all spiking networks of 10,000 neurons (five networks with independent initial conditions): mean excitatory rate in the last second after pulses of ±0.5 lasting 50 ms, against g. Dotted line, g_c of the population model; dashed line, the level. (e) Equilibrium rates of the reference network against a sustained excitatory input u_E; dotted segments, unstable equilibria in the oscillatory window ($-1.363 < u_E < -1.084$). (f) Minimum information rate h along the same branch: 28.9 bits/s at $u_E = 0$, and zero where the excitatory population is stable without inhibitory feedback

of ±0.5 lasting 50 ms, five networks with independent initial conditions lost the level in every run with $g \leq 0.35$ ($g = 0.3$ in all five networks; $g = 0$, 0.2 and 0.35 in one); at $g = 0.4$–0.5, on either side of the population threshold, the negative pulse still sent every network to the low state (3.5–3.9 Hz), whereas after the positive pulse the rate stayed within 5% of the level (40.5–42.5 Hz at $g = 0.4$, where the population model has its stable state at 42.2 Hz); at $g = 0.6$, 8 of 10 runs held the level; and at $g = 0.8$ and 1 every run returned to 41.35 ± 0.01 Hz (mean ± s.d. over networks). The population model shows the same asymmetry: after the same pulses, the negative pulse sent it to the low state for every g below 0.609 (a boundary located by bisection), whereas the positive pulse never did. We interpret this asymmetry through the branch of equilibria: just above g_c, an equilibrium with one unstable direction lies a few percent below the level and none lies above it, so that a

downward pulse can leave the basin of the level, whereas an upward pulse finds no other state above the level to reach; just below g_c, the stable state that replaces the level lies within 5% of it. Near g_c, therefore, positive pulses cannot locate the threshold under the rule of Section 2.5. In five sparse networks with random connectivity and excitabilities, $g = 0$ and $g = 0.3$ lost the level in all ten runs, $g = 0.6$ held it in 8 of 10, and $g = 1$ in all 10. The verdicts of every spiking run under each tolerance are in Supplementary Table S5.

Constant and replayed signals.

In the original reference network spiking experiment, in which all branches started from the same warmed state and received the same brief excitatory perturbation, live communication gave late means of 41.36/36.03 Hz, whereas the constant mean gave 59.58/45.50 Hz and the circular replay 59.58/45.49 Hz (Supplementary Table S12; Supplementary Fig. S3). Both state-independent signals lost the level in these finite runs, an observed outcome consistent with Corollary 1 below, which is stated for the population model, and both moved the network to the upper state of the clamped network (59.8/45.7 Hz in the population model). Since the replay preserved the mean and the variance of the original record, what matters is the dependence of the signal on the current state, and not its statistics.

3.3 Holding the level requires a minimum rate of information flow

We now treat the inhibitory signal received by excitatory cells as an input to the excitatory population. With $z = (\delta R_{\mathrm{E}}, \delta V_{\mathrm{E}}, \delta S_{\mathrm{E}})$ the deviation of the excitatory coordinates from the level and $v = S_{\mathrm{I}}^{(\mathrm{E})} - S_{\mathrm{I}}^{*}$ the deviation of the received signal, Eq. (7) reads

$$\begin{aligned} \dot{z} &= F_{\mathrm{E}}(z) + b\, v, \quad F_{\mathrm{E}}(0) = 0, \\ A &= DF_{\mathrm{E}}(0), \quad b = \big(0, -J_{\mathrm{EI}}/(\pi\tau), 0\big)^{\mathsf{T}}. \end{aligned} \tag{17}$$

where F_{E} is the excitatory vector field with clamped inhibition and A its Jacobian ($A = M_{\mathrm{E}}/\tau$, with M_{E} given in Appendix A.5 for the exact construction). The input enters additively because $S_{\mathrm{I}}^{(\mathrm{E})}$ enters the voltage equation linearly. The rest of the network, which observes z through S_{E} and produces v, plays the role of encoder, channel and decoder. Following the standard approach, we model a channel of capacity C bits/s as follows: over any interval $[0, t]$, the receiver distinguishes at most $N(t) \le 2^{Ct+c_0}$ symbol sequences, for some constant c_0, and v on $[0, t]$ is a function of the received sequence (the decoder may have memory). The encoder may use any information about the network. The encoder, channel and decoder are deterministic; if they use randomness independent of the network, the statement and its proof apply to each realization. Let $\lambda_1, \dots, \lambda_m$ be the eigenvalues of A with positive real part, and E_u their m-dimensional invariant subspace.

Theorem 1 (Minimum information rate) *Suppose that the level is stable in the sense of Lyapunov under the feedback: for every $\varepsilon > 0$ there is $\varepsilon_0 > 0$ such that every trajectory of Eq.* (17) *with $\|z(0)\| \le \varepsilon_0$ (all other variables of the network and of the coder starting from fixed values) satisfies $\|z(t)\| \le \varepsilon$ for all $t \ge 0$. This holds, in particular, if the level is locally exponentially stable. Then*

$$C \;\ge\; h := \frac{1}{\ln 2} \sum_{i=1}^{m} \operatorname{Re} \lambda_i. \tag{18}$$

For the linearized network ($F_{\mathrm{E}}(z) = Az$), it suffices that the trajectories starting in some ball remain in some bounded set.

Proof: Appendix A.6.

Theorem 1 is the data-rate theorem (Tatikonda and Mitter 2004; Nair and Evans 2004; Nair et al. 2007). Local versions of the bound for nonlinear systems are also known, in discrete time through topological feedback entropy (Nair et al. 2004) and in continuous time through invariance entropy and minimal bit rates for exponential stabilization (Colonius and Kawan 2009; Colonius 2012; Kawan 2013). The proof in Appendix A.6 is a direct continuous-time version of the volume argument for the network, by means of an invariant cone that controls the stable directions. The theorem concerns every initial state of a neighborhood and all future time. The simulations below sample finitely many initial states over finite runs: a run that fails below h is an observed finite-time outcome consistent with the theorem, but the theorem does not require that each sampled start fail, and a run

that holds above h is evidence consistent with sufficiency, not a proof of it. The intact network is not rate-limited, because its inhibitory signal is analog and noiseless in the population model; the bound becomes binding when the signal is carried by a limited channel.

Corollary 1 (State-independent signals cannot hold the level) *Suppose that $h > 0$, that is, that the excitatory population is unstable when inhibition is clamped ($m \geq 1$). If the inhibitory signal received by excitatory cells does not depend on the state of the network (a constant, a prerecorded replay, or any signal generated without access to the state), the level is not stable, for the linearized and for the nonlinear network.*

Proof: Appendix A.7. For the linearized network it follows from superposition alone; Theorem 1 gives the quantitative form, since such a signal is a channel with $N(t) = 1$, that is, $C = 0 < h$.

In the exact construction, $m = 1$ and $\lambda_u = 0.20298/\tau$ (Proposition 5), so that $h = 0.2928/\tau$ bits/s, or 14.6 bits/s for $\tau = 20$ ms. In the reference network, $h = 28.9$ bits/s. The bound depends on the operating regime: along the equilibrium branch of the reference network, h is positive only for drives between about -1.8 and 1.2, reaches 84 bits/s next to the oscillatory window, and is zero outside, where the excitatory loop is stable without inhibitory feedback (Fig. 1f). The regime with $h > 0$ is the inhibition-stabilized regime (Tsodyks et al. 1997; Ozeki et al. 2009; Sanzeni et al. 2020).

The information must also arrive in time. It is known that, for linear plants, an unstable pole limits the delay that any linear time-invariant feedback can tolerate, and that this limit becomes tighter as the pole grows (Middleton and Miller 2007); for the network, the limit is the following.

Proposition 6 (Critical delay) *In the exact construction with $g = 1$ and the inhibitory feedback to excitatory cells delayed by d, the level is locally exponentially stable if $0 \leq d < d_c$ and unstable if $d > d_c$, where $d_c \approx 1.5675\,\tau$ (31.3 ms for $\tau = 20$ ms); at $d = d_c$ a pair of characteristic roots lies on the imaginary axis and crosses it from left to right as d increases.*

Proof: Appendix A.8.

For other gains, the critical delay depends on g: in the exact construction, the characteristic roots first reach the imaginary axis at $d = 3.43\,\tau$ for $g = 0.5$ and at $d = 0.75\,\tau$ for $g = 1.5$. In the reference network, the critical delay at $g = 1$ is 13.3 ms (8.5 ms at $g = 1.5$), and the stable region in the gain–delay plane closes as either the strength or the timeliness of the state signal is reduced (Fig. 2d). Beyond the boundary, the level was replaced by an oscillation whose amplitude grew with the delay: zero at 12 and 13 ms, and 6.5, 10.0, 14.5 and 17.2 Hz at 13.6, 14, 15 and 16 ms; at 20 ms the rate swung between 3.5 and 87 Hz.

Together, these results give the conditions under which the level is attracting.

Theorem 2 (Conditions for an attracting level) *In the exact construction, with no plasticity, no target-rate variable and no added sensing or effector circuit: (i) the level $r^* = 1/(\pi\tau)$ is the unique self-consistent state of the flow of activity around the excitatory–inhibitory loop; (ii) without delay, the level is attracting, in the sense of Eq.* (1)*, if and only if the current inhibitory state is fed back to excitatory cells with a gain $g \in (7/16, 1187/512)$; at $g = 1$, it is attracting if and only if the delay is below $1.5675\,\tau$; (iii) every implementation of this feedback under which the level is stable in the sense of Theorem 1, whatever its code, needs a channel of capacity at least $0.2928/\tau$ bits/s, and every state-independent signal fails.*

Proof: Appendix A.9.

3.4 An efficient coder nearly attains the information bound

Theorem 1 is a necessary condition. To ask whether h is also enough, we replaced the inhibitory pathway of the reference network by the efficient coder of Section 2.4, at a fixed sampling period, and varied only the number of cells n_q, that is, the number of bits sent per sample. With small initial perturbations ($a_0 = 0.02\%$), every one of the 17 settings with $\mathcal{R} \geq 1.017\,h$ held the level from all 24 starts, and every one of the 12 settings with $\mathcal{R} \leq 0.988\,h$ failed from all 24 starts (Fig. 2a,b; Supplementary Table S4). Across all tested settings, therefore, the largest rate at which the level was lost was $0.988\,h$ and the smallest rate at which it was held was $1.017\,h$. At a single sampling period the bracket is wider: for example, at $T_s = 100$ ms the level was lost with $n_q = 7$ ($0.971\,h$) and held with $n_q = 8$ ($1.038\,h$), and at $T_s = 200$ ms it was lost with

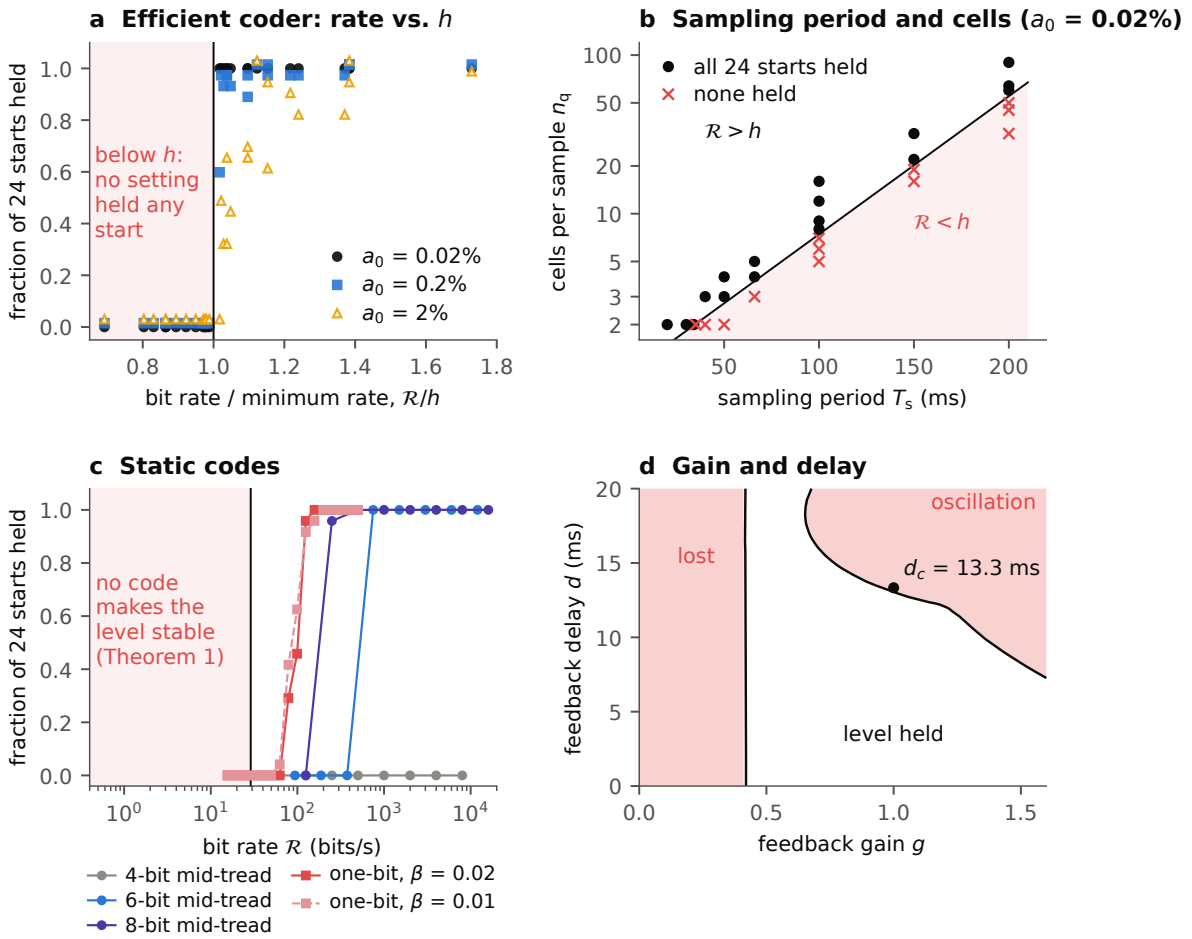


Fig. 2 Capacity requirement and efficient coding (population model of the reference network; finite-time runs). (a) Efficient coder in the inhibitory pathway: fraction of 24 starts in which the level was held, against the bit rate relative to h, for 29 combinations of the sampling period T_s (20–200 ms) and of the number of cells n_q, and for initial perturbations of relative size $a_0 = 0.02\%$, 0.2% and 2% (markers offset vertically for visibility). No setting below h held any start. (b) The same settings with $a_0 = 0.02\%$, in the plane of T_s and n_q: circles, all 24 starts held; crosses, none held. The line $n_q = e^{\lambda_u T_s}$ is $\mathcal{R} = h$. (c) Static codes: fraction of 24 starts held (2% perturbations, 3 s) against the bit rate, for the mid-tread quantizer with 4, 6 and 8 bits per sample and for the one-bit code with $\beta = 0.01$ and 0.02 (the other static settings, none of which held below h, are not shown). Shaded, $\mathcal{R} < h$, where, by Theorem 1, no code makes the level stable. (d) Linear stability of the level in the plane of feedback gain g and delay d (shaded, unstable); at $g = 1$ the critical delay is 13.3 ms

$n_q = 50$ ($0.976\,h$) and held with $n_q = 60$ ($1.022\,h$). The outcomes at the sampling periods from 20 to 200 ms are consistent with a common threshold near h. This is a manipulation of the number of bits at fixed timing, fixed mean and fixed gain, and, in these finite runs, the switch from failure to success lay close to h. With larger perturbations ($a_0 = 2\%$), a margin was needed: only 2 of the 17 settings above h held all 24 starts (at $1.12\,h$ and $1.38\,h$), settings between $1.15\,h$ and $1.73\,h$ still lost 1 to 10 starts, and no setting below h held any start. We attribute this dependence on a_0 to the nonlinear terms of the population equations, and not to initial errors outside the coder's cell, which did not occur (Supplementary Section S3).

Static codes.

Codes that do not track the network's dynamics need much more than h. Of the 176 static settings tested in the population model, none below h held the level from any start. The lowest rate at which a static code held any start was 62.5 bits/s, or 2.2 h (one-bit code with $\beta = 0.01$, 1 of 24 starts). With $\beta = 0.02$, the one-bit code held all 24 starts at 157.5 bits/s, 23 at 125, 11 at 99, 7 at 79 and none at 62.5 bits/s or below. The quantizers needed at least 250 bits/s with 8 bits per sample and 750 bits/s with 6 bits, and with 4 bits or fewer they failed at every rate up to 8000 bits/s (Fig. 2c). For static quantizers, holding is therefore not a function of the bit rate alone: a fixed quantization step confines the state to a residual set whose size depends on the resolution, not on the sampling rate (Delchamps 1990).

One-bit code in spiking networks.

To manipulate the information rate in the spiking network at gain 1 and with the mean preserved, we transmitted the inhibitory state through the one-bit code ($\beta = 0.02$), which also introduces a quantization and a lag that grow as the rate falls, in three networks with independent initial conditions. At 50 and 100 bits/s, the level was lost in all six runs: the excitatory rate oscillated with s.d. of 30–36 Hz. At 125 bits/s, 3 of 6 runs held the level; at 151.5 and 172.4 bits/s, 5 of 6; and from 200 bits/s on, all six, with s.d. of 1.0–1.3 Hz at 200 bits/s and of 0.5–0.6 Hz at 2500 bits/s, close to the 0.54 Hz of the intact network (Fig. 3c,d). The spiking network thus needs between 125 and 200 bits/s of this code, in agreement with the population model, and well above h, consistent with Theorem 1.

Inhibitory messengers.

Reducing the number of inhibitory neurons whose spikes reach excitatory cells, with the mean signal held fixed, makes the received signal noisier: it is a noise manipulation, of the kind analyzed in Section 3.6. With all 2,000 messengers, the excitatory rate fluctuated with s.d. 0.54 Hz (10-ms bins); with 300, 100, 30, 10, 5 and 3 messengers, the s.d. was 1.5, 2.2, 4.0, 7.4, 9.0 and 11.9 Hz, while in the subsets shown the mean stayed within 6.1% of the run with all 2,000 messengers (Fig. 3a,b). Under the rule of Section 2.5, the level was held down to 30 messengers, lost with 10, 5 and 3 messengers in every subset (by fluctuations larger than ten times the intact s.d.; with three messengers, two of the five subsets also shifted the mean by more than 5%), held in 2 of 5 subsets with two messengers, and lost with a single messenger in all five subsets (the network either fell to the low state, 2.8 Hz, or oscillated with s.d. near 21 Hz). The s.d. of the level grew steadily as the number of messengers fell from 2,000 to 3; with two messengers the outcome depended on the subset (held in 2 of 5, with s.d. of 2.8 and 4.4 Hz), and with one it was lost in every subset. These observed finite-time outcomes are consistent with Theorem 4, although they do not follow from it: for the linearized network, noise in the pathway does not by itself reduce below h the information that the received signal carries about the state; we did not compute this information for the messenger experiments.

3.5 The information rate fixes the regulation budget against fluctuating input

Physiological variables are regulated around a set-point against inputs that change continuously. We therefore drove the excitatory cells with $u_{\mathrm{E}}(t) = U\sin(2\pi f t)$ and asked how well the network's own inhibitory feedback, without any added regulator, holds the level.

Proposition 7 (Frequency response) *In the exact construction with gain g, the linear response of the excitatory rate to the afferent input, $\delta R_{\mathrm{E}} = T_g(\lambda)\, u_{\mathrm{E}}$, is*

$$T_g(\lambda) = \frac{2(\lambda+2)(\lambda^3+4\lambda^2+9\lambda+14)}{\chi_g(\lambda)}, \qquad T_g(0) = \frac{14}{16g-7}. \tag{19}$$

and $T_g = T_0\, S_g$, with the sensitivity function S_g and the loop function L_g

$$S_g(\lambda) = \frac{\chi_0(\lambda)}{\chi_g(\lambda)} = \frac{1}{1+L_g(\lambda)}, \qquad L_g(\lambda) = \frac{64g}{\chi_0(\lambda)}. \tag{20}$$

Proof: Appendix A.10.

At $g = 1$, $T_1(0) = 14/9$, which is the sensitivity to sustained input of Proposition 8: the settling-point property is the zero-frequency end of the frequency response. L_g is the loop through which the inhibitory state signal returns to excitatory

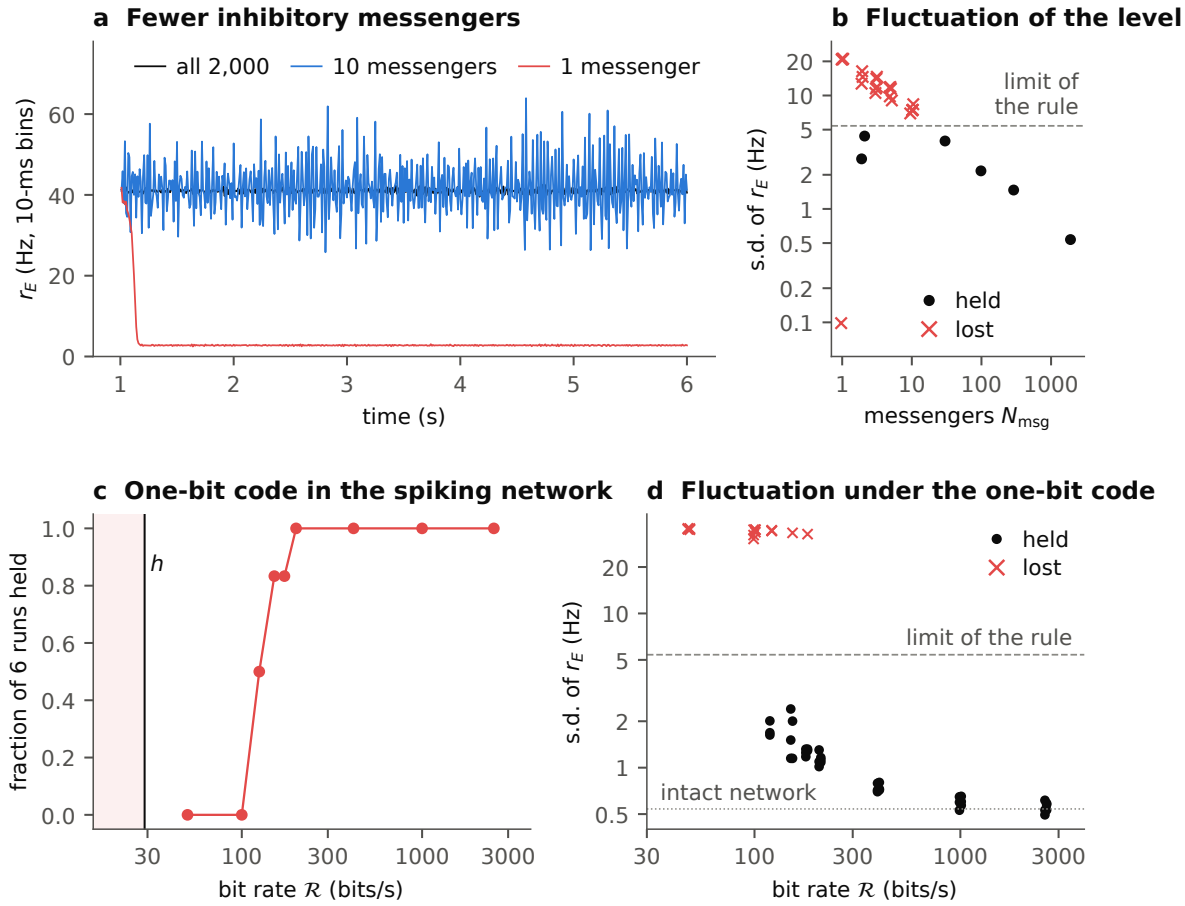


Fig. 3 State information in the spiking network (all-to-all, 10,000 neurons, gain 1, mean of the received signal preserved). (a) Excitatory rate when the inhibitory signal received by excitatory cells is carried by all 2,000 inhibitory neurons, by 10 or by a single one (messengers). (b) s.d. of the excitatory rate (last 2 s, 10-ms bins) against the number of messengers: one subset for $N_{\text{msg}} \geq 30$ and five subsets for $N_{\text{msg}} \leq 10$ (circles, held; crosses, lost). Dashed line, the limit of the rule of Section 2.5, ten times the s.d. of the intact network. The single-messenger runs fell to the low state or oscillated. (c) One-bit code ($\beta = 0.02$): fraction of six runs (three networks, pulses of ± 0.5) in which the level was held, against the realized bit rate; vertical line, h. (d) s.d. of the excitatory rate in each run of (c); dotted line, s.d. of the intact network (0.54 Hz)

cells, and S_g is the factor by which this returning signal changes the response relative to the response without it, T_0.

Theorem 3 (Sensitivity integral) *(i) For every g in the window* (16)*,*

$$\int_0^\infty \ln |S_g(i\omega)| \, \mathrm{d}\omega = \pi \lambda_u, \tag{21}$$

with the angular frequency ω and λ_u in units of $1/\tau$; in physical units, with frequency f in Hz,

$$\int_0^\infty \log_2 |S_g(2\pi i f \tau)| \, \mathrm{d}f = \frac{h}{2}. \tag{22}$$

(ii) For any two gains g_1, g_2 in the window,

$$\int_0^\infty \ln \left| \frac{T_{g_1}(i\omega)}{T_{g_2}(i\omega)} \right| \mathrm{d}\omega = 0. \tag{23}$$

Proof: Appendix A.11.

Part (i) is Bode's sensitivity integral for an open-loop unstable system (Freudenberg and Looze 1985), here derived for the network; its reading in terms of information rates is due to Martins and Dahleh (2008). It is a statement relative to T_0, the response without the state signal, which cannot be observed because that network is unstable. Part

(ii) is the observable form: between any two stable gains, the response of the network to the drive is redistributed across frequencies with zero net log-gain, so that stronger feedback buys attenuation of slow inputs with amplification of faster ones (Fig. 4a). No choice of feedback strength makes $|S_g| < 1$ at all frequencies; the excess, $h/2$, does not depend on g. For the reference network, numerical integration gave $\int_0^\infty \ln|S_g|\,\mathrm{d}\omega = 62.96\ \mathrm{s}^{-1}$ for $g = 0.5$, 0.7, 1 and 1.5, equal to $\pi\lambda_u$ to five digits, that is, $\int \log_2|S_g|\,\mathrm{d}f = 14.5$ bits/s $= h/2$.

Nonlinear population model.

The redistribution was visible in the full nonlinear dynamics (Fig. 4b, $U = 0.25$). With $U = 1$ (not shown), the intact feedback ($g = 1$) let the excitatory rate oscillate with amplitudes of 11.7–12.2 Hz for drives of 0.5 to 5 Hz and 7.3 Hz at 20 Hz. Raising the feedback to $g = 1.5$ halved the slow response (5.4–5.5 Hz at 0.5–2 Hz) and raised the response at 20 Hz to 11.0 Hz, the redistribution that Theorem 3 describes for the linearized network. Weakening the feedback had the opposite effect and, beyond the linear regime, lost the level: at $g = 0.7$, drives of $U = 1$ at 10 Hz or slower moved the mean away from the level (to 6.6–38.6 Hz), and at $g = 0.5$ every drive slower than 40 Hz sent the network to the low state near 4 Hz. With small drives ($U = 0.25$), the amplitudes followed the linear theory at $g \geq 1$. At 40–80 Hz the response was small for every g, because membrane and synaptic filtering attenuate the drive before feedback is needed. Large drives also shift the mean through the curvature of the rate response: at $g = 1$, the cycle mean fell by 5.1–5.2% for $U = 1$ at 2 and 5 Hz (just outside our 5% band, with the oscillation locked to the drive) and by 16–20% for $U = 2$ at 10 Hz or slower.

Spiking networks.

In five all-to-all networks of 10,000 neurons with intact feedback and $U = 0.5$, the oscillation amplitude of the excitatory rate was 5.09, 5.11, 5.14, 4.52, 3.52 and 1.53 Hz (s.d. across networks at most 0.04 Hz) at 1, 2, 5, 10, 20 and 40 Hz, against linear predictions, corrected for the 5-ms bins, of 5.00, 5.01, 5.03, 4.50, 3.58 and 1.54 Hz, and the mean stayed within 0.4 Hz of the level (Fig. 4c). With the inhibitory state signal reduced to $g = 0.6$, the slow drives (1 and 2 Hz) sent every network to the low state (4.5 Hz) within half a second (Fig. 4d); the 5-Hz drive lowered the mean to 36.5–37.2 Hz in four networks and to 20.8 Hz in the fifth, so that it was lost in all five; and the drives of 10, 20 and 40 Hz were held in every network, with amplitudes of 3.77, 2.74 and 1.82 Hz, against predictions of 3.75, 2.77 and 1.83 Hz. A large drive ($U = 2$) at $g = 1$ produced oscillations of 26 Hz at 2 Hz and 22 Hz at 10 Hz, locked to the drive, and lowered the mean by 6.5–6.9 Hz (16–17%), again through curvature.

Information rate and delay.

We then sent the inhibitory feedback through finite-rate codes while driving the population model at 5 Hz with $U = 0.5$ (Fig. 4e). With ideal feedback, the RMS deviation from the level was 3.72 Hz. With the one-bit code ($\beta = 0.02$), it was 3.91 Hz at 500 bits/s, 4.01 Hz at 354 bits/s, 4.15 Hz at 250 bits/s and 4.75 Hz at 177 bits/s (8 of 12 starts held), and at 125 bits/s and below the level was lost from every start (RMS deviation 34–40 Hz). With $\beta = 0.01$, the receiver could not follow the drive below 354 bits/s. The 8-bit mid-tread code gave 3.78, 3.90, 4.69 and 6.64 Hz at 16,000, 4,000, 1,000 and 500 bits/s, and lost the level at 250 bits/s. The regulation error rose as the bit rate fell, and, with this code, regulating a fluctuating input required a higher bit rate than holding a constant level (at least 177 bits/s here, against 79–158 bits/s without drive; $h = 28.9$ bits/s), in line with the rate requirements for disturbance attenuation (Martins and Dahleh 2008). Delay had the same effect (Fig. 4f): the RMS deviation from the level grew from 3.7 Hz without delay to 4.1, 5.1 and 6.6 Hz with delays of 2, 5 and 8 ms (amplitudes at the driving frequency 5.2, 5.8, 6.9 and 8.8 Hz), and from 11 ms on the level was lost (mean 30 Hz, RMS deviation 25–28 Hz), earlier than the linear limit of 13.3 ms; we interpret this as the drive carrying the state out of a basin that shrinks as the delay grows, which we did not compute.

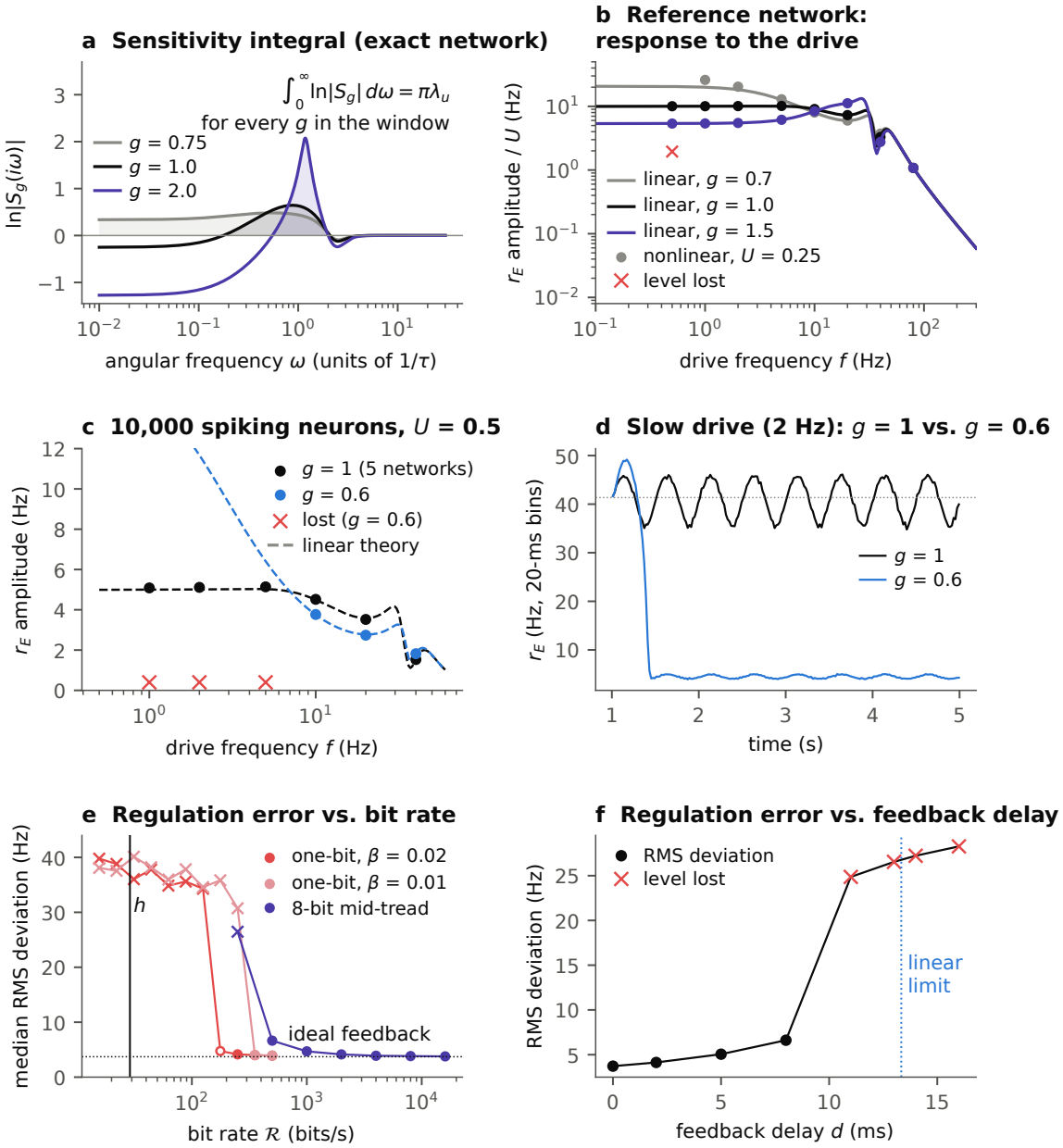


Fig. 4 Regulation against a fluctuating input. (a) Exact construction: $\ln|S_g(i\omega)|$ for $g = 0.75$, 1 and 2; for every g in the window, the area above zero exceeds the area below it by $\pi\lambda_u$ (Theorem 3). (b) Reference network: amplitude of the excitatory rate per unit drive amplitude against the drive frequency, from the linearization (lines) and from the nonlinear population model with $U = 0.25$ (circles; crosses, level lost). (c) All-to-all spiking networks with $U = 0.5$: amplitude at the driving frequency (mean ± s.d. over five networks) and linear predictions corrected for the 5-ms bins (dashed). At $g = 0.6$, the drives of 1, 2 and 5 Hz lost the level (crosses, drawn at the bottom). (d) Excitatory rate under a 2-Hz drive at $g = 1$ and $g = 0.6$ (one network). (e) Population model driven at 5 Hz ($U = 0.5$), with the inhibitory feedback sent through finite-rate codes: median RMS deviation from the level over 12 starts, against the bit rate (filled, all starts held; open, some; crosses, none). Dotted line, ideal feedback; vertical line, h. (f) The same drive with delayed feedback: RMS deviation against the delay (crosses, level lost); dotted line, linear critical delay

3.6 Through a noisy inhibitory pathway, the received signal of the linearized network reveals the state at exactly the rate the instability generates

The gain changes how strongly the inhibitory state acts, not what it carries. To measure how much information about the state reaches the excitatory cells, at fixed gain and fixed mean, we sent the inhibitory signal through a noisy channel (Eq. (10)). For the exact construction, the answer can be stated exactly, provided that the probabilistic setting is stated exactly; we do this first.

Setting. Time is in units of τ. Let $\zeta(t) \in \mathbb{R}^6$ be the deviation of the population state from x^*, with a Gaussian initial value $\zeta(0)$ of positive-definite covariance Σ_0, and let W be a standard Wiener process independent of $\zeta(0)$. The network sends the inhibitory deviation $w = e_6^{\mathsf{T}}\zeta$ (with e_k the k-th unit vector), and the excitatory cells receive the cumulative signal $\mathcal{Y}$, so that the linearized network with a white Gaussian channel of noise amplitude $\sigma > 0$ is

$$\begin{aligned} \mathrm{d}\zeta &= M_0\zeta\,\mathrm{d}t + b_6\mathrm{d}\mathcal{Y}, \\ \mathrm{d}\mathcal{Y} &= g\,w\,\mathrm{d}t + \sigma\mathrm{d}W, \quad w = e_6^{\mathsf{T}}\zeta. \end{aligned} \tag{24}$$

where $b_6 = -2e_2$ is the column through which the received signal enters the excitatory voltage equation, so that $M(g) = M_0 + g\,b_6 e_6^{\mathsf{T}}$. Formally, the received signal is $v = \mathrm{d}\mathcal{Y}/\mathrm{d}t = g\,w + \sigma\xi$. The quantity of interest is the rate at which the received signal accumulates information about the initial state,

$$\bar{I} := \lim_{t\to\infty} \frac{1}{t}\, I\big(\zeta(0); \mathcal{Y}_{[0,t]}\big), \tag{25}$$

where $\mathcal{Y}_{[0,t]}$ is the received path on $[0,t]$. Because the sent signal is a function of $\zeta(0)$ and of the past received signal, $I(\zeta(0); \mathcal{Y}_{[0,t]})$ is also the directed information from the sent to the received signal (Massey 1990; Kadota et al. 1971; Weissman et al. 2013), the measure appropriate for a channel inside a feedback loop (Kim 2010). Conditioning matters: with a different source of uncertainty, the quantity changes (Corollary 2).

Theorem 4 (Information through a noisy inhibitory pathway) *In the exact construction, with the setting of Eq.* (24)*:*

(i) for every $g \neq 0$ and every $\sigma > 0$, $\bar{I} = \lambda_u$ nats per unit time, that is, h bits/s in physical units;

(ii) if, in addition, g lies in the window (16)*, the received signal has, in the stationary regime, the spectral density $\Phi_v(\omega) = \sigma^2|S_g(i\omega)|^2$, and $\bar{I} = \frac{1}{4\pi}\int_{-\infty}^{\infty} \ln[\Phi_v(\omega)/\sigma^2]\,\mathrm{d}\omega$;*

(iii) for g in the window, in the stationary regime, the power of the state-dependent part of the received signal, $P_{\mathrm{s}} = \lim_{t\to\infty}\mathbb{E}[(g\,w(t))^2]$, satisfies

$$\frac{P_{\mathrm{s}}}{\sigma^2} = \frac{1}{\pi}\int_0^{\infty} \big(|S_g(i\omega)|^2 - 1\big)\,\mathrm{d}\omega \;>\; 2\lambda_u, \tag{26}$$

so that the capacity of the channel, $C = P_{\mathrm{s}}/(2\sigma^2)$ nats per unit time, exceeds $\bar{I}$: $C > h$ in bits/s.

Proof: Appendix A.12.

Remark 1 Part (i) holds for every $g \neq 0$, including gains outside the window, where the level is lost. The identity $\bar{I} = h$ is therefore a signature of the instability, not its cure: the unstable mode amplifies the uncertainty about the initial state at rate λ_u, and, because this mode is visible in the inhibitory signal, the received signal reveals the state at exactly that rate, whatever the noise. Its value lies in measurement: injecting noise into the inhibitory pathway and recording the received signal is, for the linearized network, an assay of h. What holds the level is stated by Theorem 1 and Proposition 4; part (iii) shows that the capacity used by the intact network exceeds h, the noisy-channel form of Theorem 1 (Braslavsky et al. 2007). For gains in the window, part (i) and parts (ii)–(iii) are the result of Elia (2004), who identified the Bode integral with the information transmitted by a linear loop through a Gaussian channel (see also Martins and Dahleh 2008; Ardestanizadeh and Franceschetti 2012); what the present proof adds is that part (i) holds for every $g \neq 0$, including the gains at which the loop is unstable.

In a finite spiking network, the inhibitory signal also fluctuates because of the finite number of neurons. The following corollary concerns a specified model of such fluctuations: the linearized network driven, in addition to the channel noise, by an independent additive intrinsic noise source, with the information counted about the enlarged source formed by the initial state and this noise. For this noise model, the rate of that information is at least λ_u; the corollary does not establish the corresponding statement for finite spiking networks,

whose intrinsic fluctuations need not be additive or independent of the state.

Corollary 2 (Intrinsic fluctuations) *Let the linearized network also be driven by an intrinsic noise process $\mathcal{U}$, independent of $\zeta(0)$ and W, through a matrix Γ: $\mathrm{d}\zeta = M_0\zeta\,\mathrm{d}t + b_6\,\mathrm{d}\mathcal{Y} + \Gamma\,\mathrm{d}\mathcal{U}$. Then, for every $g \neq 0$, $\liminf_{t\to\infty} \frac{1}{t} I\big(\zeta(0), \mathcal{U}_{[0,t]}; \mathcal{Y}_{[0,t]}\big) \geq \lambda_u$. If $\mathcal{U}$ is a Wiener process and g lies in the window, the Gaussian spectral expression splits as*

$$\frac{1}{4\pi}\int_{-\infty}^{\infty} \ln \frac{\Phi_v(\omega)}{\sigma^2}\,\mathrm{d}\omega = \lambda_u + \frac{1}{4\pi}\int_{-\infty}^{\infty} \ln\Big(1 + \frac{\Phi_{\mathcal{U}}(\omega)}{\sigma^2 |S_g(i\omega)|^2}\Big)\mathrm{d}\omega. \tag{27}$$

where $\Phi_{\mathcal{U}} \geq 0$ is the spectral density of the part of v driven by $\mathcal{U}$.

Proof: Appendix A.13.

When the channel noise is small relative to the intrinsic fluctuations, the second term of Eq. (27) is large; as the channel noise grows, it shrinks, and the expression approaches λ_u from above. For the exact construction at $g = 1$, numerical integration of Theorem 4(ii) gave $0.20298/\tau$ nats per unit time, equal to λ_u to eleven digits, and Eq. (26) gave $P_{\mathrm{s}}/\sigma^2 = 0.8007/\tau$, so that $C = 1.97\,h$. For the linearized reference network at $g = 0.5$, 0.7, 1 and 1.5, the information rate was 28.912 bits/s in every case, whereas $C = 75.9$, 53.0, 71.3 and 155.6 bits/s.

Population model.

The simulations agree with Theorem 4 where it applies and show where it ceases to apply (Fig. 5a,b). For $\sigma = 0.01$, 0.02 and 0.05 Hz $\mathrm{s}^{1/2}$, the level was held in all eight trials, and the Gaussian spectral estimate was 30.1 ± 0.4, 30.2 ± 0.5 and 30.6 ± 0.5 bits/s (mean $\pm$ s.e. of the eight single-trial estimates), against $h = 28.9$ bits/s. The excess of about one bit/s belongs to the eight noise realizations, not to the network: the linearized reference network loop driven by the same realizations gave 30.1 ± 0.4 bits/s, whereas 24 independent realizations gave 28.8, 29.3 and 29.3 bits/s in three sets of eight. Over all 32 realizations, the linear loop gave 29.4 ± 0.5 bits/s, within one standard error of h. At $\sigma = 0.01$, the capacity computed from the power of the sent signal was 72.0 ± 0.4 bits/s, against the linear prediction of 72.1 bits/s at this integration step (71.3 bits/s in continuous time). The s.d. of the excitatory rate was $22.7\,\sigma$, against the predicted $22.7\,\sigma$ at this integration step ($22.6\,\sigma$ in continuous time, the value used in the rule of Section 2.5), and the mean stayed within 0.02 Hz of the level. Noise in the inhibitory pathway thus neither moved the level nor reduced the Gaussian spectral estimate below h (an estimate that is exact only for the linearized network, which gave the same values with the same realizations); it only made the level fluctuate, in proportion to the noise. For larger noise, the linearization ceases to apply. At $\sigma = 0.1$, 0.2 and 0.3, the level was still held in every trial, but the estimate (mean of the eight trials) rose to 32.3, 38.6 and 48.5 bits/s and the mean fell by up to 0.8 Hz, which we attribute to the curvature of the rate equations. At $\sigma = 0.5$, the level was held in 2 of 8 trials under the 5% rule, and from $\sigma = 0.7$ on in none: the mean fell by 7–9.5%, with excursions above twice the level 230 to 1400 times per minute, while the s.d. stayed well within the bound of condition (ii). These verdicts depend on the tolerance, since with a band of 10% the level would have been classified as held in 8, 8, 6 and 8 of the eight trials at $\sigma = 0.5$, 0.7, 1 and 1.5; under large channel noise, therefore, the simulations show a displacement of the operating level whose classification depends on the criterion, and they do not, by themselves, demonstrate a loss of attraction. For the linearized network, Theorem 4 implies that channel noise does not reduce the information rate below h; in these nonlinear runs the estimate is not certified and is not a lower bound, so it cannot show the same. Our interpretation is that noise of this amplitude drives the network out of the region in which the linearization holds, so that the curvature of the rate equations lowers the mean rate; whether the level still attracts under such noise was not tested.

Spiking network.

In the all-to-all network, the level was held in every run with $\sigma \leq 0.3$ (three network seeds), and, under the 5% rule, in one of three at $\sigma = 0.5$ and in none at $\sigma \geq 0.7$, where the mean fell by 7–9.5% (Fig. 5c); with a tolerance of 10%, all nine runs at $\sigma = 0.5$–1 would have been classified as held, including the s.d. condition (Supplementary Table S5), so that these verdicts record a displacement of the operating level whose classification depends on the criterion and do not, by themselves, demonstrate

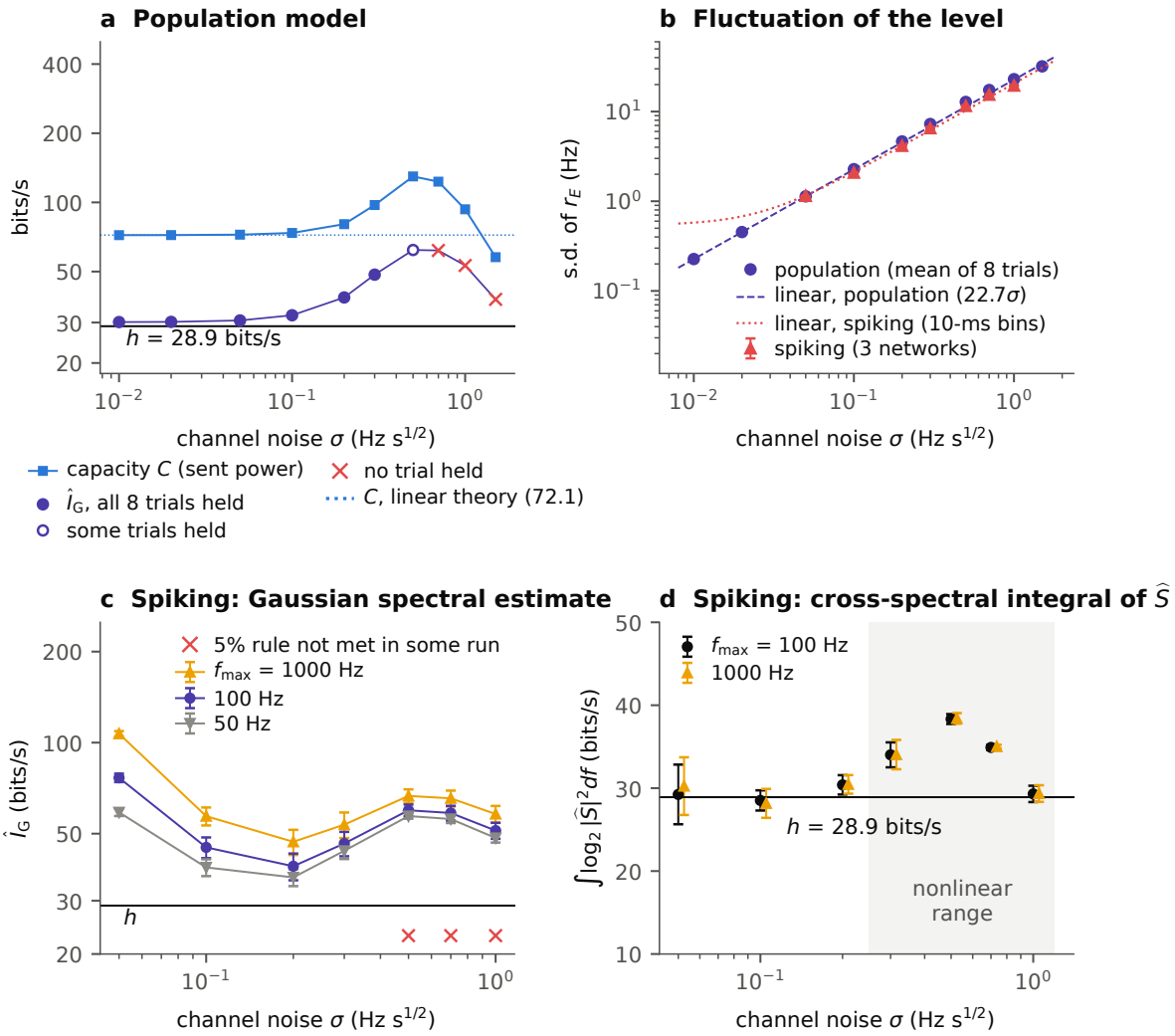


Fig. 5 Information through a noisy inhibitory pathway at fixed gain ($g = 1$) and fixed mean. (a) Population model (eight trials of 80 s for each σ): Gaussian spectral estimate $\widehat{I}_{\mathrm{G}}$ of the received signal computed for each trial and summarized over all eight trials at every σ, whatever the verdict of each trial (mean ± s.e. of the eight single-trial estimates; filled circles, $\sigma \leq 0.3$, all eight trials held; open circle, $\sigma = 0.5$, where 2 of 8 trials held, so that the mean pools held trials and trials that did not meet the 5% rule; crosses, $\sigma \geq 0.7$, where no trial met the rule, drawn at the mean of the eight trials without error bars) and capacity $C = P_{\mathrm{s}}/(2\sigma^2 \ln 2)$ computed from the power of the sent signal (squares, mean over all eight trials). Horizontal lines: h, and the linear prediction of C at the integration step used (72.1 bits/s). (b) s.d. of the excitatory rate against σ in the population model (mean of all eight trials) and in the all-to-all spiking network (mean ± s.d. over three networks, 10-ms bins), with the linear predictions. (c) Spiking network: $\widehat{I}_{\mathrm{G}}$ (mean ± s.d. over three networks) for upper frequencies $f_{\max}$ of 50, 100 and 1000 Hz; crosses, noise amplitudes at which at least one run did not meet the 5% rule of Section 2.5, although every run met it with a tolerance of 10% (Supplementary Table S5); means and s.d. include every run, whatever its verdict. (d) Spiking network: empirical cross-spectral integral $\int_0^{f_{\max}} \log_2 |\widehat{S}(f)|^2 \, \mathrm{d}f$, from the ratio of Eq. (12), for $f_{\max} = 100$ and 1000 Hz (mean ± s.d. over three networks), compared with the population value of h; for $\sigma \geq 0.3$ (shaded) the response is no longer linear, and the estimate depends on σ

a loss of attraction. The s.d. of the rate grew in proportion to the noise, between $19.2\,\sigma$ and $23.5\,\sigma$, against the predicted $20.6\,\sigma$ for 10-ms bins. The Gaussian spectral estimate was above h in every run: 76.5 ± 2.5 bits/s at $\sigma = 0.05$ (mean $\pm$ s.d. over seeds), 45.0 ± 3.5 at 0.1 and 39.0 ± 4.0 at 0.2, and between 46 and 60 bits/s for larger noise. The decrease from $\sigma = 0.05$ to 0.2 is consistent with Corollary 2, which, for the linearized network with additive intrinsic noise, predicts an estimate above h that decreases toward h as the channel noise grows; the rise at larger noise, where the mean also shifted, lies outside its linear assumptions; our interpretation, which we did not test separately, is that at weak channel noise the received signal is dominated by the intrinsic fluctuations of the 2,000 inhibitory neurons. The estimate at weak noise also grew with the upper frequency (from 59 to 109 bits/s as $f_{\max}$ went from 50 to 1000 Hz for one seed at $\sigma = 0.05$; Supplementary Table S6), because the intrinsic fluctuations are broadband, so that its value there is not a measurement of a single rate. The cross-spectral ratio, Eq. (12), retains only the part of the received signal that is linearly correlated with the injected noise and, in this nonlinear network, describes the best linear approximation of the loop for the given excitation (Section 2.6). For $\sigma \leq 0.2$, the empirical cross-spectral integral $\int_0^{100} \log_2 |\widehat{S}(f)|^2 \, \mathrm{d}f$ was 29.4±2.1 bits/s (nine runs, range 26.5–33.3), against $h = 28.9$ bits/s, and it varied by at most 3.3 bits/s over $f_{\max}$ from 50 to 1000 Hz (Fig. 5d). Its mean over the three networks was 29.2, 28.5 and 30.4 bits/s at $\sigma = 0.05$, 0.1 and 0.2, that is, it changed by less than 2 bits/s over a fourfold range of the noise amplitude, less than its s.d. across networks at $\sigma = 0.05$ (3.6 bits/s). This is consistent with a weak dependence of the best linear approximation on the level of excitation in this range, but it does not establish it, since a single integral does not characterize the whole frequency response. For larger noise, it depended clearly on the amplitude: it rose to 32–39 bits/s at $\sigma = 0.3$–0.7, where the response is no longer linear, and at $\sigma = 1$, where no run met the 5% rule, it returned to 28.3–30.3 bits/s; agreement with h at a single noise amplitude is therefore not, by itself, evidence that the estimate reflects the linearized loop. The spiking network therefore reproduces two statements: the level survives a moderately noisy inhibitory pathway at fixed gain and fixed mean, and, at weak noise, the empirical cross-spectral integral is consistent with the population prediction h, with little dependence on the band and on the noise amplitude. This is evidence of consistency with the prediction, not an independently validated measurement of the instability rate of the finite spiking network, and the integral does not, in any case, measure the information that the excitatory cells receive. The Gaussian spectral estimate exceeded h in every run, but outside the stationary linear Gaussian case with channel noise only it is not a lower bound on the information, so the spiking data do not show that the information reaching the excitatory cells is at least h.

3.7 The information requirement holds across random networks

Of 641 random parameter sets, 74 had a unique, stable level with an excitatory rate between 1 and 100 Hz that became unstable when inhibition was clamped; in 68 of them the excitatory block was unstable (inhibition-stabilized networks), and the other 6, whose instability came only from the inhibitory population, were excluded. In the 68 networks, the excitatory rate ranged from 6.7 to 63 Hz (median 26 Hz; 13 networks below 10 Hz), h from 0.23 to 101 bits/s (median 24.9 bits/s), g_c from 0.004 to 0.96 (median 0.39), and d_c from 0.17 to 60 ms (median 8.0 ms) (Fig. 6).

The efficient coder holds the level just above h.

In the 63 networks whose clamped instability was confined to the excitatory population (in the other five, the inhibitory population was also unstable on its own, and the coder had to track those modes as well), the efficient coder with small perturbations held none of the 756 starts at $0.8\,h$ or $0.95\,h$, 97.5% of them at $1.05\,h$ (all starts in 58 networks), 99.2% at $1.2\,h$ and all of them at $1.5\,h$ (Fig. 6a). The incomplete cases at $1.05\,h$ were five networks with h below 6.2 bits/s, three of which have, when inhibition is clamped, another stable state within 3% of the level; we interpret these failures as the residual uncertainty of the coder reaching that nearby state, and we report them as a limitation of this coder, not of the bound. With larger perturbations ($a_0 = 2\%$), none of the starts was held below h, and 45, 74 and 87% of them were held at 1.05, 1.2 and $1.5\,h$.

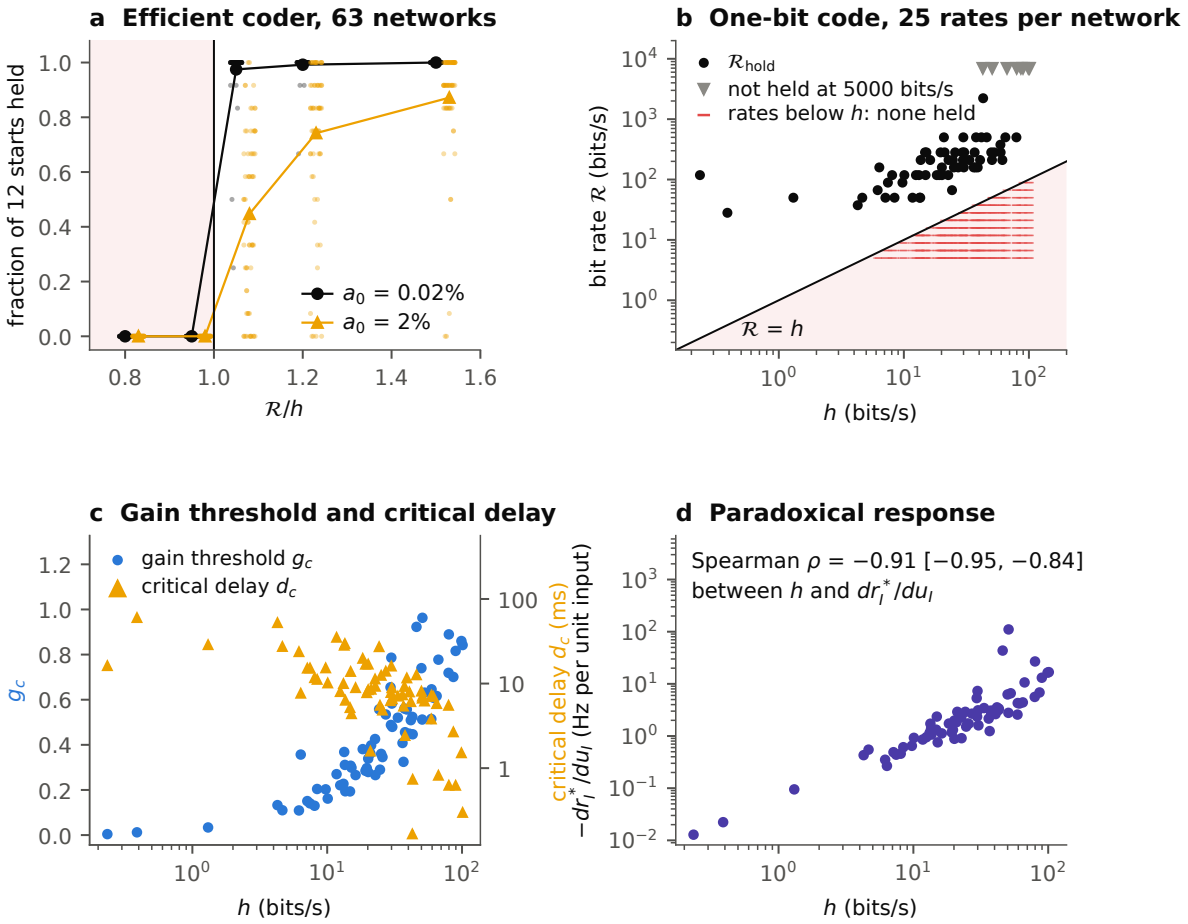


Fig. 6 The information requirement across 68 random inhibition-stabilized networks (population model). (a) Efficient coder, in the 63 networks whose clamped instability is confined to the excitatory population: fraction of 12 starts held at bit rates of 0.8, 0.95, 1.05, 1.2 and 1.5 h, pooled over networks (lines) and for each network (dots), for perturbations $a_0 = 0.02\%$ and 2%. (b) One-bit code at 25 rates from 5 to 5000 bits/s: $\mathcal{R}_{\text{hold}}$ of each network (circles), every tested rate below the network's own h (ticks; none held the level from any start), and networks not held at 5000 bits/s (triangles). Shaded, $\mathcal{R} < h$. (c) Gain threshold g_c (left axis) and critical delay d_c (right axis) against h. (d) Size of the paradoxical response, $-\mathrm{d}r_{\mathrm{I}}^*/\mathrm{d}u_{\mathrm{I}}$, against h

No one-bit channel below h held the level. In each of the 63 networks whose h exceeded the lowest grid rate (5 bits/s), every grid rate below h was tested (397 tests); none held the level from any start (Fig. 6b). The highest of these rates lay between 75% and 99.8% of the network's own h (median 87%; above 90% in 20 networks), so that the finite-time runs probed rates close to the bound. With the received signal fixed at S_{I}^* (the zero-information control), the level was lost in every network; in three networks with $h < 1.4$ bits/s, 4 to 6 of 12 starts ended within 5% of the level, because the clamped network has another stable state within 3% of it. Above h, the one-bit code needed much more than the efficient coder: every holding rate was at least $2.7\,h$, and $\mathcal{R}_{\text{hold}}$, defined in 60 networks, exceeded h by a factor between 2.7 and 506 (median 8.2), the inefficiency of a code that is not matched to each network. Holding was monotone in the rate in 57 of the 61 networks held at some rate. Seven networks were not held even at 5000 bits/s; all had a large h (43–101 bits/s) and a small stability margin (six had $d_c < 1.6$ ms, and the seventh $g_c = 0.96$), so that the lag and

quantization of the code plausibly exceed that margin.

The information rate ranks the other requirements.

Across the 68 networks, h was strongly rank-correlated with the gain threshold (Fig. 6c) (Spearman $\rho = 0.91$, 95% confidence interval 0.84 to 0.94), inversely with the critical delay ($\rho = -0.71$, -0.82 to -0.56), and with $\mathcal{R}_{\text{hold}}$ ($\rho = 0.76$, 0.62 to 0.85; 60 networks). Because $\mathcal{R}_{\text{hold}}$ was also strongly related to the critical delay ($\rho = -0.83$), we computed the partial correlation of h and $\mathcal{R}_{\text{hold}}$ given d_c, which was 0.57: h predicts the rate that the one-bit code needs beyond what the delay tolerance explains. Every network showed the paradoxical response, a fall of the inhibitory rate when inhibitory cells are excited, as expected of inhibition-stabilized networks (Tsodyks et al. 1997; Sanzeni et al. 2020), and h grew with its size ($\rho = -0.91$ between h and $\mathrm{d}r_{\mathrm{I}}^*/\mathrm{d}u_{\mathrm{I}}$, -0.95 to -0.84; Fig. 6d). Networks whose instability generates information about their state more rapidly thus require stronger, richer and more timely inhibitory feedback, in agreement with the delay limits known for linear plants (Middleton and Miller 2007).

3.8 Finite networks reproduce the population level

In the original reference network series, the spiking baselines were 41.36 and 36.03 Hz, within $+0.26\%$ and -0.11% of the population equilibrium (Supplementary Table S10). After pulses of ± 3 (0.2 s), the means recovered to within 0.017% of baseline (Supplementary Fig. S2). Halving the integration step changed the final means by at most 0.0015%, networks of 5,000 and 20,000 neurons changed them by at most 0.12%, and the excitatory means of three initializations differed by at most 0.016% (inhibitory, 0.033%; Supplementary Fig. S5); these are comparisons in one realization, not replicate statistics. Under a sustained input of $u_{\mathrm{E}} = -3$ for 6 s, the all-to-all network settled at 4.75/8.50 Hz (population: 4.79/8.62 Hz), and under $+3$ at 60.31/45.84 Hz (population: 60.30/45.94 Hz); after removal of the input, the excitatory mean returned to 41.357 Hz in both branches, within 0.003% of baseline (Fig. 7a,b). Five sparse networks, each with its own connectivity and excitabilities, held baselines of 41.8 ± 0.5 Hz, settled at 5.6 ± 0.7 Hz and 60.8 ± 0.2 Hz under sustained drives of -3 and $+3$, and every one returned to within 0.02 Hz of its own baseline when the drive was removed (Fig. 7d–f). The mechanism therefore does not depend on all-to-all coupling, quantile excitabilities or one random realization.

3.9 What the information flow does not do: sustained input moves the level

Under a sustained drive, the level moves and stays displaced, from 41.4 Hz to 4.7 or 60.3 Hz. This follows from the equations.

Proposition 8 (Sensitivity to sustained input) *(i) In the exact construction, a small sustained excitatory input moves the equilibrium according to*

$$\frac{\mathrm{d}R_{\mathrm{E}}^*}{\mathrm{d}u_{\mathrm{E}}} = \frac{14}{9}, \qquad \frac{\mathrm{d}R_{\mathrm{I}}^*}{\mathrm{d}u_{\mathrm{E}}} = \frac{8}{9}. \tag{28}$$

(ii) An uncoupled population with the same excitability width and the same rate ($\bar{\eta} = 3/4$) has $\mathrm{d}R/\mathrm{d}u = 2/5$, so that the recurrent loop amplifies the displacement by the factor $35/9 \approx 3.9$.

Proof: Appendix A.14.

In the reference network, the excitatory sensitivity is 9.98 Hz per unit of input, against 3.05 Hz for an uncoupled population at the same rate, a factor of 3.3. The return of the network after a transient perturbation therefore cannot be read as maintenance of the same rate under a permanent bias, and the recurrent loop, which makes the level attracting, also amplifies its sensitivity to sustained input. This statement concerns the networks studied here. A fixed network can reject a sustained input robustly if it contains an integrating motif, that is, a variable that accumulates the deviation of the regulated rate (Francis and Wonham 1976; Sontag 2003; Yi et al. 2000; Briat et al. 2016), and slow processes that integrate activity, such as calcium-dependent regulation of conductances, have been proposed to play this role in neurons (O'Leary et al. 2014). Such a motif would be a regulating mechanism of the kind that this paper excludes. The level that arises in these networks from the flow of activity and information is, then, a locally attracting level with an exact

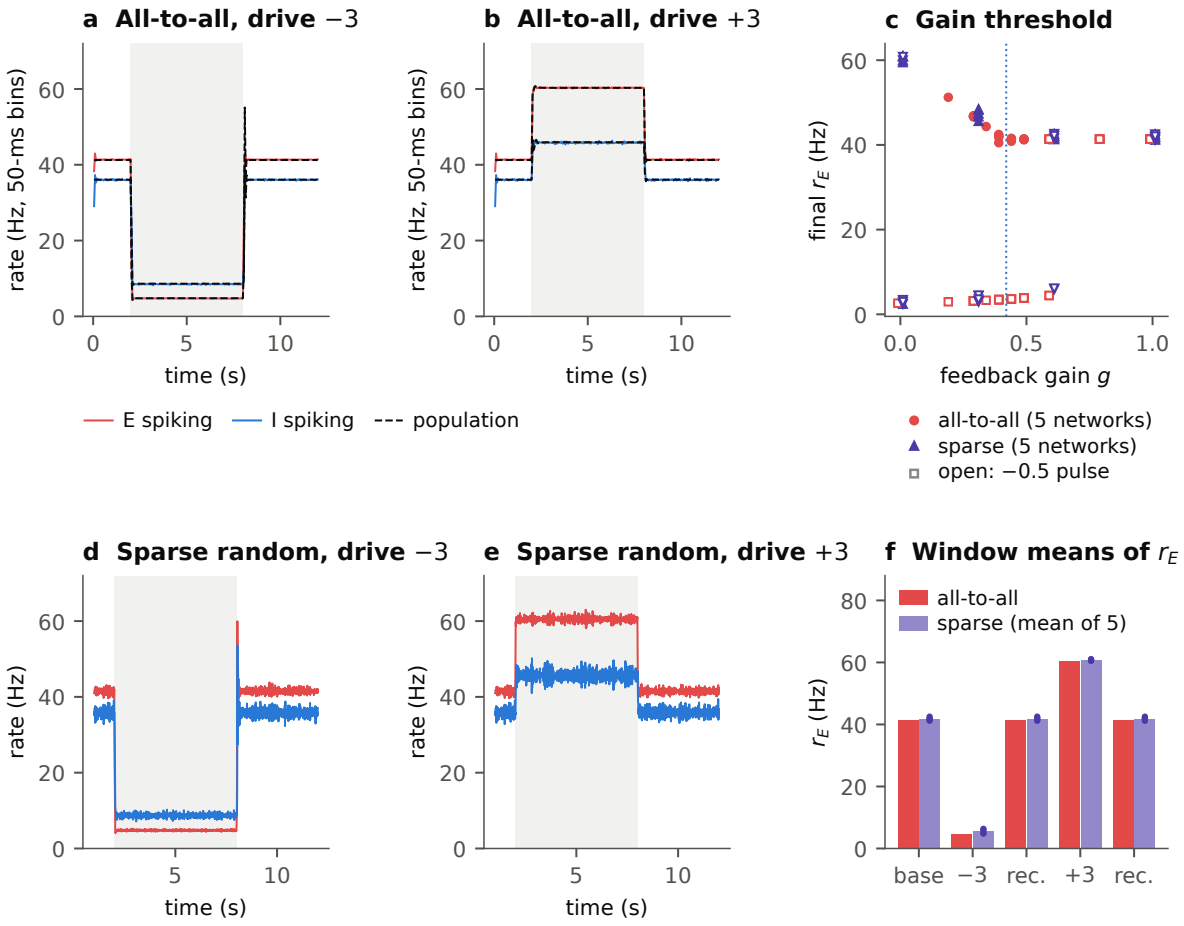


Fig. 7 Finite networks reproduce the population level. (a,b) All-to-all spiking network (10,000 neurons): excitatory and inhibitory rates (50-ms bins) under a sustained input $u_E = -3$ (a) or $+3$ (b) between 2 and 8 s (shaded), and the population model (dashed). (c) Mean excitatory rate in the last second after pulses of ± 0.5, against the feedback gain, in five all-to-all networks (circles, squares) and in five sparse networks with $K = 200$ inputs from each population and random excitabilities (triangles); filled, positive pulse; open, negative pulse; dotted line, g_c of the population model. (d,e) One sparse network under the same sustained inputs. (f) Mean excitatory rate at baseline, during the sustained input (-3 or $+3$) and after its removal (rec.), in the all-to-all network and in the five sparse networks (bars, mean; dots, individual networks)

defense against transient perturbations and a quantified shift under sustained input: a settling point, in the terminology of physiology (Wirtshafter and Davis 1977; Speakman et al. 2011).

3.10 Equilibrium branch and oscillatory window of the reference network

By Proposition 2, the reference network has exactly one positive equilibrium for every input. It is stable except for $-1.36324 < u_E < -1.08416$, where a complex pair has positive real part (Fig. 1e). Both Hopf points are supercritical, with first Lyapunov coefficients of -0.253 and -0.246 and frequencies of 4.57 and 4.50 Hz; inside the window the excitatory rate oscillates (for instance between 9.3 and 32.1 Hz at $u_E = -1.2$), with an amplitude that grew as the square root of the distance from each Hopf point (log–log slopes of 0.50 and 0.47). At $u_E = -3$, 0 and 3, all 67 sampled starts at each input approached the equilibrium, and the three interval certificates passed (Supplementary

Section S10); these establish local attraction in certified ellipsoids and give sampled evidence about the basin, but not global attraction.

3.11 Transmitted state information and predictive information are distinct

Theorem 1 measures information flow by the capacity of the channel that carries the state back. A pairwise statistic such as transfer entropy measures something else. Consider the local Gaussian observation model

$$\begin{aligned} \mathsf{Y}_{n+1} &= \alpha_1 \mathsf{Y}_n + \beta_1 \mathsf{X}_n + \beta_2^{\mathsf{T}} \mathsf{Z}_n + \epsilon_n, \\ \epsilon_n &\sim \mathcal{N}(0, \nu^2), \quad \nu^2 > 0. \end{aligned} \tag{29}$$

in which Y_n is a target, X_n a source and Z_n a vector of observed inputs at step n, jointly Gaussian; α_1, β_1 and the vector β_2 are coefficients; and the innovation ϵ_n is independent of $(\mathsf{X}_n, \mathsf{Y}_n, \mathsf{Z}_n)$, with variance ν^2.

Lemma 1 (Conditional information in a transmitted state) *Under Eq.* (29)*, let* $\mu = \mathrm{Var}(\mathsf{X}_n \mid \mathsf{Y}_n, \mathsf{Z}_n) > 0$*. Then*

$$\begin{aligned} &I(\mathsf{Y}_{n+1}; \mathsf{X}_n \mid \mathsf{Y}_n, \mathsf{Z}_n) \\ &\quad = \frac{1}{2} \log_2 \left(1 + \frac{\beta_1^2 \mu}{\nu^2} \right) \ \textit{bits.} \end{aligned} \tag{30}$$

which is positive if $\beta_1 \neq 0$ *and zero if* $\beta_1 = 0$*.*

Proof: Appendix A.15.

With X the fluctuation of the inhibitory synaptic state and Y that of the excitatory voltage, an Euler step Δt gives $\beta_1 = -\Delta t \, J_{\mathrm{EI}}/(\pi\tau)$: transfer entropy is positive whenever the coupling is nonzero, but it is a property of the fluctuations around an operating point, not of the point itself.

Proposition 9 (Shannon measures do not see a sustained mean shift) *Let* $\dot{x}_{\mathrm{o}} = A_{\mathrm{o}} x_{\mathrm{o}} + B_{\mathrm{o}} u + \Sigma_{\mathrm{o}} \xi$*, with* A_{o} *a Hurwitz matrix,* B_{o} *and* Σ_{o} *the matrices through which a constant input* u *and white Gaussian noise* ξ *enter. The mutual information, conditional mutual information and transfer entropy between any coordinates of the stationary process do not depend on* u*, whereas the stationary mean of an output* $C_{\mathrm{o}} x_{\mathrm{o}}$ *is* $-C_{\mathrm{o}} A_{\mathrm{o}}^{-1} B_{\mathrm{o}} u$*.*

Proof: Appendix A.16.

A network whose level moves with a sustained input and one whose level does not would be indistinguishable by these measures if their local dynamics were the same. A positive transfer entropy is consistent with, but does not establish, the channel capacity that Theorem 1 requires; a deterministic network sitting at its equilibrium transmits no fluctuations at all, although its feedback holds the level. The probe assay agrees (Supplementary Table S13; Supplementary Fig. S4): live communication gave proxies above the shift null in both directions (about 0.2 bits per 10-ms bin), whereas the constant and replayed signals gave small or null values; but in the probe assay, which used separate 20-s runs with random probe currents (Supplementary Table S3), both state-independent conditions ended in the low state (2.63/7.35 Hz, excitatory/inhibitory), not in the upper state reached after the brief perturbation of Section 3.2, so that the proxies compare different activity regimes as well as different signals (Barnett et al. 2009; Vicente et al. 2011). The quantity that holds the level is the capacity of the feedback channel (Theorem 1), whose threshold was located with rate-limited codes in Section 3.4; the noise experiments of Section 3.6 measure, under the conditions stated there, the rate at which the received signal reveals the state, a signature of h rather than the cause of holding.

4 Discussion

4.1 What the results establish

The results establish a restricted, but substantive, possibility: in a recurrent network model with fixed connections, an attracting level of activity can arise from the flow of activity and information through the network, without an explicit target-rate variable, a separately added sensing or effector circuit, or plasticity. As shown in Sections 3.1 and 3.2, the value of the level and its attraction come from two different aspects of the same flow: the balance of activity around the loop fixes the value (Propositions 1 and 2), and the transmission of the current state through inhibition decides whether this value is attracting (Propositions 4 and 5). The information view states this transmission in numbers: a channel of capacity at least h (Theorem 1), arrival within a critical delay (Proposition 6) and dependence on the current state (Corollary 1). In finite

simulations, an efficient coder came close to the bound, in the reference network and across random networks (Sections 3.4 and 3.7). It is in this sense that, in the model, the setpoint arises from information flow: the network can hold its level only if its feedback can carry its own state at a rate of at least h (a channel of capacity $C \geq h$), and in time.

The mechanism, inhibition stabilization, is already established (Tsodyks et al. 1997; Ozeki et al. 2009; Sanzeni et al. 2020; Ahmadian and Miller 2021), and the tools that we use, namely the data-rate theorem, Bode's integral and its information identity, come from control theory, including the extension of the data-rate theorem to nonlinear systems (Tatikonda and Mitter 2004; Nair and Evans 2004; Nair et al. 2004; Colonius and Kawan 2009; Colonius 2012; Kawan 2013; Freudenberg and Looze 1985; Elia 2004; Martins and Dahleh 2008). What is new is the use of these tools to state, for a recurrent network model of this kind, exactly what inhibition must transmit and how much of it, in the statements (i)–(v) listed at the end of the Introduction. The bound depends on the regime: h is positive only where the excitatory subnetwork is unstable on its own, that is, in the inhibition-stabilized regime that is common in cortex (Sanzeni et al. 2020). Outside this regime, the excitatory dynamics hold the level by themselves, and no information from inhibition is required.

The sinusoidal experiments extend the result from the maintenance of a level to regulation around it. Against a fluctuating drive, the network's own inhibitory loop, which is not a dedicated regulator, obeys the same limits as any feedback loop: Theorem 3 shows that the excess $h/2$ of the log-sensitivity is spent across frequencies, so that stronger feedback attenuates slow inputs only by amplifying faster ones, and the channel experiments showed that the regulation error grows as information becomes scarce. Therefore, the same quantity, h, sets a necessary condition for the level to be an attractor and bounds how well it is defended.

To call this level emergent does not mean that it is uncaused or independent of the design of the model. The coupling magnitudes, the excitabilities and the time constants specify the vector field, and the coefficients of the exact construction were chosen for algebraic transparency. The absence of a comparator shows that one particular control architecture is unnecessary for local attraction; it does not explain how biological parameters are selected or why a particular level is advantageous (Kaplan 2011).

4.2 Is information the limiting quantity in a real circuit?

One may object that the intact inhibitory pathway, with 2,000 neurons firing near 36 Hz, can carry far more than 28.9 bits/s, so that the bound is always satisfied with a large margin. We agree with the premise. In the population model the intact signal is analog and noiseless, and, even through a noisy pathway, the capacity used by the network exceeds h (Theorem 4(iii); 71 bits/s against 28.9 bits/s in the reference network at $g = 1$). We do not claim that circuits in vivo operate close to the bound. Our claim is that a minimum capacity is necessary, that it has an exact value, and that, in our simulations, a matched code needed little more: from small perturbations, about 2% above h in the reference network. Three results support it. First, when the pathway was limited to a finite rate, the threshold of the finite runs lay close to h: with fixed timing, mean and gain, the largest rate at which the efficient coder failed was $0.99\,h$, and the smallest rate at which it held was $1.02\,h$, across all tested settings (Section 3.4); in this situation, information is the limiting quantity, and the outcomes at sampling periods from 20 to 200 ms were consistent with a common threshold. Second, codes that are not matched to the network need much more: static codes at least $2.2\,h$, the one-bit code between 125 and 200 bits/s in the spiking network, and a median of $8.2\,h$ across random networks. In practice, therefore, the code, its lag and its quantization set the rate that a circuit needs, and h is the floor below which no code can go. Third, across random networks, h ranked the gain threshold, the critical delay and the rate that the one-bit code needed, the last even after the delay tolerance was taken into account (Section 3.7). Information thus becomes limiting when the rate of the inhibitory pathway approaches h, that is, when the state reaches excitatory cells through a pathway that updates at a limited rate or with few distinguishable levels; noise and delay, on the other hand, degrade the level by other routes, presumably large excursions and loss of phase margin. For the linearized network without delay, channel

noise does not reduce below h the information that the received signal carries (Theorem 4).

4.3 Information flow and its measurement

In this work, *information flow* has a precise sense: the rate at which the feedback can carry the state of the network, that is, its capacity, which must be at least h for the level to be stable (Theorem 1), and, for a noisy pathway of the linearized network, the rate at which the received signal accumulates information about the initial state (Theorem 4). Understood in this manner, it is distinct from pairwise predictability statistics. Transfer entropy is a property of the fluctuations around an operating point (Lemma 1) and is blind to sustained shifts of that point (Proposition 9); the probe assay found conditional predictive structure under live communication, but the substituted conditions also changed the activity regime, so that its magnitudes cannot be read as stabilizing information. The finite-rate codes manipulate the information carried by the feedback at fixed mean, and the noisy channel supplies the model of noise that is needed to compute Shannon information exactly, for the linearized loop.

Theorem 4 must be read with care. The identity $\bar{I} = h$ holds for every nonzero gain, including gains at which the level is lost, and so it is a signature of the instability, not the reason why the level is held (Remark 1). Its value is practical: injecting noise into the inhibitory pathway and recording what the excitatory cells receive is, for the linearized network, an assay of h. The Gaussian spectral estimate $\widehat{I}_{\mathrm{G}}$ realizes this assay only where its assumptions hold, namely for a stationary, linear and Gaussian loop in which the injected channel noise is the only noise; even in the linearized network, independent additive white intrinsic noise makes it exceed h (Corollary 2). In the population model with weak noise, it agreed with h to within the variability of the noise realizations; in the spiking network, it exceeded h and depended on the frequency band, which is qualitatively consistent with Corollary 2, although the corollary concerns the linearized network with independent additive noise and does not imply this result for the nonlinear spiking estimate; outside the linear Gaussian case, moreover, the estimate is not a measurement of the information. The cross-spectral ratio, which uses the known injected noise, describes, in a nonlinear network, the best linear approximation of the loop for the given excitation, which is not necessarily its sensitivity function (Section 2.6). In the spiking network, its integral, 29.4 ± 2.1 bits/s at weak noise, agreed with the population prediction h and changed little with the frequency band and with the noise amplitude; this shows that the empirical integral is consistent with h, but it does not establish an independently validated measurement of the instability rate of the finite spiking network, which would require, for example, a comparison with the rate at which the same network runs away when its inhibition is clamped, a comparison that this study does not make. Moreover, the integral does not measure the information that the excitatory cells receive, and no spiking estimate in this paper shows that this information is at least h.

4.4 The limit of a fixed network

The level is defended against transient perturbations, but not against sustained ones, which move it (Proposition 8); as discussed in Section 3.9, rejecting a sustained input robustly would require an integrating motif, a regulating mechanism of the kind that this paper excludes. Firing-rate homeostasis after prolonged deprivation, which develops over hours to days (Hengen et al. 2013, 2016), therefore requires processes that our results do not model, such as activity-dependent regulation of conductances (O'Leary et al. 2014) or inhibitory plasticity (Vogels et al. 2011). What our results show is that, in a recurrent network model, an operating level of the kind around which such processes act can be fixed by the recurrent flow of activity, and its fast defense provided by the recurrent flow of information, without an explicit target-rate variable or a separately added regulating circuit. From this point of view, molecular mechanisms that change excitability or synaptic efficacy (Turrigiano et al. 1998; Styr et al. 2019) may act by changing the flow, and thereby the level, rather than by supplying the level; this is a hypothesis that our models suggest but do not test.

4.5 Contact with physiology

Two quantities can be compared, with caution, with published measurements. First, the critical delay of the reference network, 13.3 ms, exceeds

the lag of feed-forward inhibition behind monosynaptic excitation in hippocampal CA1, which is of a few milliseconds (Pouille and Scanziani 2001); across the random networks, however, the critical delay ranged from 0.17 to 60 ms (median 8.0 ms), and 10 of the 68 networks had a critical delay below 3 ms. Moreover, the delay of our model adds to synaptic and membrane filtering that are already modeled, so that it is not directly comparable with a feed-forward latency. Whether real circuits operate on the stable side of the delay limit cannot be decided from these unfitted models; it requires measuring the latency of recurrent inhibition in an identified inhibition-stabilized circuit. Second, signatures of the inhibition-stabilized regime, in which $h > 0$, have been reported in vivo, indirectly from conductance changes during surround suppression in cat visual cortex (Ozeki et al. 2009) and directly from paradoxical responses to the stimulation of inhibitory neurons in several areas of mouse cortex (Sanzeni et al. 2020). In the random networks, h grew with the size of the paradoxical response ($\rho = -0.91$ between h and $\mathrm{d}r_{\mathrm{I}}^*/\mathrm{d}u_{\mathrm{I}}$). This gives a prediction that can be tested with existing methods: circuits with a larger paradoxical response should tolerate less feedback delay and need stronger inhibitory feedback.

4.6 Limitations

The exact reduction assumes infinite populations, Lorentzian excitabilities and uniform coupling within each population (Montbrió et al. 2015; Ott and Antonsen 2008). The sparse networks show that the gain threshold, the loss of the level without state feedback and the shift under sustained input survive random connectivity and random excitabilities, but their weights scale as J_{XY}/K, which keeps the mean-field limit and not the strong fluctuations of balanced networks with weights of order $1/\sqrt{K}$ (van Vreeswijk and Sompolinsky 1996). The information manipulations were done in all-to-all networks only, and the efficient coder was run in the population model only; conductance-based synapses, distributed synaptic delays, adaptation and structured connectivity remain untested. The random networks were drawn from one family of models, and their inhibition-stabilized members had excitatory rates of 6.7 Hz or more, so that very low rates are underrepresented. The efficient coder needs a margin above h for larger perturbations (with 2% perturbations in the reference network, only 2 of 17 settings above h held every start, and settings up to $1.73\,h$ lost some), and in five random networks with very small h it did not hold every start at $1.05\,h$. The results on attraction are local: Theorem 1 concerns a neighborhood of the level, and our certificates and basin samples do not establish global attraction. The capacity requirement concerns every start in that neighborhood and all future time, whereas every simulation sampled finitely many starts over a finite time; the simulations are therefore consistent with the theorems, not proofs of them. Theorems 3 and 4 are statements about the linearized network; for large noise or large drives, the simulations departed from them, which we attribute to the curvature of the rate equations. In the spiking network, the cross-spectral ratio is a best linear approximation for the injected noise, and its integral, although consistent with h at weak noise, is not an independent measurement of the instability rate of the finite network. Finally, the decision whether a level was held uses a tolerance of 5%, an analysis choice whose effect is reported in Supplementary Table S5. The choice matters most under large channel noise: at $\sigma = 0.5$, 0.7 and 1, 1 of 9 spiking runs and 2 of 24 population trials met the 5% rule, against 9 of 9 and 22 of 24 with a tolerance of 10%, so that these verdicts describe a displacement of the operating level whose classification depends on the criterion, not, by themselves, a loss of attraction.

4.7 Experimental tests

The results suggest five tests in an inhibition-stabilized preparation, all of which preserve the mean of inhibition: the first, second and fifth manipulate the information that it carries, the third measures h, and the fourth manipulates its timing. First, replacing the inhibitory input onto excitatory cells by a signal of matched mean and variance that does not depend on the network state, for example by closed-loop optogenetic control (Newman et al. 2015) following a prerecorded pattern, should destabilize the operating level, whereas a closed-loop signal that tracks the state should not. Second, if the closed-loop signal is updated at a fixed period with a limited number of levels, the level should be lost when the number of bits per second falls below a threshold that, for

a well-matched code, approaches h. Third, h can be measured independently, either from the rate at which activity runs away when inhibition onto excitatory cells is clamped ($h = \lambda_u / \ln 2$ when, as in the reference network, a single real eigenvalue is unstable with inhibition clamped; with several unstable eigenvalues, h is the sum of their positive real parts divided by $\ln 2$, as in Theorem 1) or from the cross-spectral integral of the response to zero-mean noise injected into the inhibitory input (Section 3.6), provided that this integral changes little when the amplitude of the injected noise is varied, as expected where the loop responds nearly linearly; the two measurements should agree, which would provide the independent validation that our spiking simulations do not supply, and the injected noise should increase the fluctuations of the excitatory rate in proportion to its amplitude while the level is maintained. Fourth, a controllable delay in the closed loop should reveal a critical delay. Fifth, reducing the number of inhibitory neurons that contact excitatory cells, while their mean drive is compensated, should increase the fluctuations of the rate before the level is lost. It is important to note that matching the mean will not match every biophysical effect; the controls must therefore address timing, receptor recruitment and changes in the unobserved state.

5 Conclusion

In an exact mean-field model of recurrent excitatory and inhibitory spiking neurons, an attracting firing level arises from the flow of activity and information through the network, without an explicit target-rate variable and without a separately added sensing or effector circuit; the only feedback is the network's own inhibition. The balance of activity around the loop fixes the value of the level, and the transmission of the current state through inhibition makes it attracting when the feedback lies inside an exact window and arrives within a critical delay. Any feedback under which the level is stable must be carried by a channel of capacity at least h bits/s, a bound that follows from the data-rate theorem, proved here for the nonlinear network, and that an efficient coder approached in finite simulations, holding the level from $1.02\,h$ in the reference network for small perturbations. The same rate fixes the budget for regulation against fluctuating input, and, for the linearized network with a white Gaussian channel and no other noise, it is also exactly the rate at which the received signal reveals the state, so that it can be recovered from the response to noise injected into the inhibitory pathway; in spiking networks with weak injected noise, the empirical cross-spectral integral of this response was consistent with it. The level is defended against transient perturbations but moves under sustained input: it is a settling point, generated by the flow of information through the network. Whether the setpoints of biological neural circuits arise in the same way remains a hypothesis to be tested experimentally.

Statements and Declarations

Ethics approval.

Not applicable. This study is computational and involved no animals or human participants.

Consent to participate and consent for publication.

Not applicable.

Author contributions.

L.G.F. conceived the study, performed the analyses and simulations, and wrote the manuscript.

Use of generative AI.

ChatGPT (OpenAI) assisted with manuscript drafting and editing, code preparation, and checks of mathematical derivations and references. The author is responsible for the scientific content and its verification.

Data and code availability.

The accompanying reproducibility package contains the LaTeX and BibTeX sources, simulation and analysis code, stored results, figure scripts, and symbolic and interval checks. The code is available to reviewers through the submission materials.

A Proofs

Every result of the Results section is proved here in full. The statements are not repeated; each proof refers to the statement by its number, and the symbols follow the notation of Table 1. Symbols that appear only inside one proof are named where they are introduced and are local to that proof.

A.1 Proof of Proposition 1 (unique positive equilibrium)

With $\Delta = 1$ and $\bar{\eta}_{\mathrm{E}} = \bar{\eta}_{\mathrm{I}} = -1/4$, the terms of the stationary conditions (8) that do not involve the coupling are collected in the auxiliary function $\varphi(R) = 1/(4R^2) - 1/4 - R^2$, defined for $R > 0$, and the two conditions become

$$\varphi(R_{\mathrm{E}}) + 3R_{\mathrm{E}} - 2R_{\mathrm{I}} = 0, \qquad \varphi(R_{\mathrm{I}}) + 2R_{\mathrm{E}} - R_{\mathrm{I}} = 0, \tag{A1}$$

in which the linear coefficients are the ratios J_{XY}/π fixed by Eq. (13), with a negative sign for the inhibitory couplings. Subtracting the second condition from the first, we obtain $\varphi_1(R_{\mathrm{E}}) = \varphi_1(R_{\mathrm{I}})$ for the second auxiliary function $\varphi_1(R) = \varphi(R) + R$, whose derivative,

$$\varphi_1'(R) = 1 - 2R - \frac{1}{2R^3}, \tag{A2}$$

is negative for every $R > 0$: for $R \geq 1/2$, because $1 - 2R \leq 0$, and for $R < 1/2$, because $1/(2R^3) > 4$. Being strictly decreasing, φ_1 cannot take the same value at two distinct points, so that $R_{\mathrm{E}} = R_{\mathrm{I}} =: R$. Both conditions then reduce to $\varphi_1(R) = 0$, and multiplication by $4R^2$ yields

$$(R - 1)(4R^3 + R + 1) = 0. \tag{A3}$$

Since the cubic factor is positive for $R > 0$, the only positive solution is $R = 1$. The voltage and synaptic coordinates follow from $V_X^* = -\Delta/(2R_X^*)$ and $S_X^* = R_X^*$, which completes the proof. □

A.2 Proof of Proposition 2 (unique level in the reference network)

At a positive equilibrium, the inhibitory stationary equation determines R_{E} as the function $R_{\mathrm{E}}(q)$ of $q = R_{\mathrm{I}}$ in Eq. (9), and the excitatory equation then determines the drive $u_{\mathrm{E}}(q)$. The positive equilibria at drive u therefore correspond one to one to the solutions $q > 0$ of $u_{\mathrm{E}}(q) = u$ with $R_{\mathrm{E}}(q) > 0$. Since $R_{\mathrm{E}}'(q) = (\pi/J_{\mathrm{IE}})[2q + J_{\mathrm{II}}/\pi + \Delta^2/(2q^3)] > 0$, the function R_{E} is strictly increasing, and $R_{\mathrm{E}}(q) > 0$ exactly for $q > q_0 \approx 0.40216$. Differentiating $u_{\mathrm{E}}(q)$, we obtain

$$u_{\mathrm{E}}'(q) = \Big[2R_{\mathrm{E}} + \frac{\Delta^2}{2R_{\mathrm{E}}^3} - \frac{J_{\mathrm{EE}}}{\pi}\Big] R_{\mathrm{E}}'(q) + \frac{J_{\mathrm{EI}}}{\pi}, \tag{A4}$$

with $R_{\mathrm{E}} = R_{\mathrm{E}}(q)$. Because $R_{\mathrm{E}}' > 0$ and $J_{\mathrm{EI}}/\pi > 0$, we have $u_{\mathrm{E}}'(q) > 0$ wherever the bracket is nonnegative, which is the case when $R_{\mathrm{E}} \geq J_{\mathrm{EE}}/(2\pi)$ (owing to the first term) and when $R_{\mathrm{E}} \leq [\pi\Delta^2/(2J_{\mathrm{EE}})]^{1/3} \approx 0.428$ (owing to the second). It remains to examine the compact interval of q on which $0.428 < R_{\mathrm{E}}(q) < J_{\mathrm{EE}}/(2\pi)$, which lies inside $(q_0, q_b]$, where $R_{\mathrm{E}}(q_b) = J_{\mathrm{EE}}/(2\pi)$ and $q_b \approx 2.58806$. On this interval we evaluated Eq. (A4) in interval arithmetic with 30 significant digits (103-bit mantissa) and outward rounding, so that every computed interval encloses the exact range, with π and every decimal constant enclosed by intervals. The cover consists of 17 adjacent subintervals with decimal end points, from $q = 0.402159316$ to $q = 2.588055847$; we checked, also in interval arithmetic, that $R_{\mathrm{E}} < 0.428$ at the lower end and $R_{\mathrm{E}} > J_{\mathrm{EE}}/(2\pi)$ at the upper end, so that, R_{E} being increasing in q, the cover contains the whole interval; the certified lower bound of u_{E}' was positive on every subinterval, the smallest being 0.1097 (Supplementary Table S7 lists every subinterval and its bound). Hence u_{E} is strictly increasing on (q_0, ∞).

Moreover, $u_{\mathrm{E}} \to -\infty$ as $q \downarrow q_0$ (because $R_{\mathrm{E}} \to 0$) and $u_{\mathrm{E}} \sim R_{\mathrm{E}}^2 \to +\infty$ as $q \to \infty$. By continuity and strict monotonicity, $u_{\mathrm{E}}(q) = u$ has exactly one solution for each $u \in \mathbb{R}$, and so the network has exactly one positive equilibrium at each drive. □

A.3 Proof of Proposition 3 (local exponential stability)

In the time unit τ, the Jacobian $M = M(1)$ of Eq. (7) at x^* (with $R^* = 1$, $V^* = -1/2$ and $\tau/\tau_s = 2$) is

$$M = \begin{pmatrix} -1 & 2 & 0 & 0 & 0 & 0 \\ -2 & -1 & 0 & 0 & 3 & -2 \\ 0 & 0 & -1 & 2 & 0 & 0 \\ 0 & 0 & -2 & -1 & 2 & -1 \\ 2 & 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 2 & 0 & 0 & -2 \end{pmatrix}, \tag{A5}$$

with rows and columns in the state order $(R_{\mathrm{E}}, V_{\mathrm{E}}, R_{\mathrm{I}}, V_{\mathrm{I}}, S_{\mathrm{E}}, S_{\mathrm{I}})$. Reordering the state into the two-population vectors δR, δV and δS (deviations of the rates, voltage centers and synaptic variables of both populations), the Jacobian takes the block form

$$M = \begin{pmatrix} -I & 2I & 0 \\ -2I & -I & B \\ 2I & 0 & -2I \end{pmatrix}, \qquad B = \begin{pmatrix} 3 & -2 \\ 2 & -1 \end{pmatrix}, \tag{A6}$$

where I and 0 are the 2×2 identity and zero matrices and the coupling block B, with rows indexed by the receiving population and columns by the sending one, contains the ratios J_{XY}/π with a negative sign for inhibition. For an eigenvector with eigenvalue λ, the first and third block rows give $(\lambda+1)\delta R = 2\delta V$ and $(\lambda+2)\delta S = 2\delta R$; substituting both into the second row, $(\lambda+1)\delta V = -2\delta R + B\delta S$, and multiplying by $2(\lambda+2)$, we obtain, for $\lambda \neq -2$,

$$\big[P(\lambda)I - 4B\big]\delta R = 0, \qquad P(\lambda) = (\lambda+2)\big[(\lambda+1)^2 + 4\big] = \lambda^3 + 4\lambda^2 + 9\lambda + 10, \tag{A7}$$

so that P is the characteristic polynomial of the rate–voltage–synapse block of a single uncoupled population. By the Schur complement formula, the monic polynomials $\det(\lambda I - M)$ and $\det(P(\lambda)I - 4B)$, both of degree six, agree for every $\lambda \neq -2$, and hence everywhere. With $\operatorname{tr} B = 2$ and $\det B = 1$,

$$\det(\lambda I - M) = P^2 - 4\operatorname{tr}(b)\,P + 16\det B = (P-4)^2 = \big[(\lambda+1)(\lambda^2+3\lambda+6)\big]^2. \tag{A8}$$

The eigenvalues are therefore $-1/\tau$ and $(-3 \pm i\sqrt{15})/(2\tau)$, each with algebraic multiplicity two, all with negative real parts. Since the vector field is smooth near x^*, the linearization theorem gives local exponential stability (Strogatz 2015; Khalil 2002). We note, in passing, that $B - I \neq 0$ is nilpotent, $(B-I)^2 = 0$, so that M is not diagonalizable; this produces polynomial factors in the transients but does not affect the criterion. □

A.4 Proof of Proposition 4 (exact feedback window)

At x^*, the received signal equals S_{I}^* for every g, so that the stationary equations, and with them x^*, do not depend on g. The Jacobian $M(g)$ differs from Eq. (A5) only in the entry $(2,6)$, which becomes $-2g$; equivalently, B in Eq. (A6) is replaced by $B_g = \left(\begin{smallmatrix} 3 & -2g \\ 2 & -1 \end{smallmatrix}\right)$, with $\operatorname{tr} B_g = 2$ and $\det B_g = 4g - 3$. The elimination of Appendix A.3 is unchanged, and so

$$\chi_g(\lambda) = \det\big(PI - 4B_g\big) = P^2 - 8P + 16(4g-3) = (P-4)^2 - 64(1-g), \tag{A9}$$

which is Eq. (15). Expanding P,

$$\chi_g(\lambda) = \lambda^6 + 8\lambda^5 + 34\lambda^4 + 84\lambda^3 + 129\lambda^2 + 108\lambda + (64g - 28), \tag{A10}$$

so that g enters only the constant term. We write $\chi_g = \sum_{k=0}^{6} \kappa_k \lambda^{6-k}$, with $(\kappa_0, \ldots, \kappa_6) = (1, 8, 34, 84, 129, 108, 64g - 28)$. By the Routh–Hurwitz criterion, all roots of a real polynomial with $\kappa_0 > 0$ have negative real parts if and only if all leading principal minors D_k of the Hurwitz matrix $\mathcal{H} = (\kappa_{2j-i})_{i,j=1}^{6}$ (with $\kappa_k = 0$ outside $0 \le k \le 6$) are positive. Here

$$\mathcal{H} = \begin{pmatrix} 8 & 84 & 108 & 0 & 0 & 0 \\ 1 & 34 & 129 & 64g - 28 & 0 & 0 \\ 0 & 8 & 84 & 108 & 0 & 0 \\ 0 & 1 & 34 & 129 & 64g - 28 & 0 \\ 0 & 0 & 8 & 84 & 108 & 0 \\ 0 & 0 & 1 & 34 & 129 & 64g - 28 \end{pmatrix}, \tag{A11}$$

and direct evaluation gives

$$\begin{aligned} &D_1 = 8, \qquad D_2 = 8 \cdot 34 - 84 = 188, \qquad D_3 = 8400, \qquad D_4 = 128\,(752g + 3523), \\ &D_5 = -256\,(16g + 209)(512g - 1187), \\ &D_6 = (64g - 28)\,D_5 = -1024\,(16g - 7)(16g + 209)(512g - 1187). \end{aligned} \tag{A12}$$

The minors D_3 and D_4 follow by cofactor expansion of the leading 3×3 and 4×4 blocks, and $D_6 = \kappa_6 D_5$ because the last column of $\mathcal{H}$ has the single nonzero entry κ_6. For $g \ge 0$, D_1 to D_4 are positive and $16g + 209 > 0$; hence $D_5 > 0$ if and only if $g < 1187/512$, and $D_6 > 0$ if and only if, in addition, $16g > 7$. All minors are therefore positive exactly on the interval (16), which, by the linearization theorem, proves that x^* is locally exponentially stable exactly for these gains. For $0 \le g < 7/16$, $\chi_g(0) = 64g - 28 < 0$ while $\chi_g(\lambda) \to +\infty$ as $\lambda \to +\infty$, so that the intermediate value theorem gives a positive real eigenvalue. For $g \ge 1187/512$, $D_5 \le 0$, so that not every root lies in the open left half-plane. At $g_H = 1187/512$, the roots lie in the closed left half-plane, by continuity from the inside of the window; Orlando's formula, $D_5 = \prod_{i<j}(\lambda_i + \lambda_j)$ up to a nonzero factor, shows that $D_5 = 0$ forces two roots with $\lambda_i + \lambda_j = 0$; and since both lie in the closed left half-plane and $\lambda = 0$ is not a root ($\kappa_6 \ne 0$), they form a pair $\pm i\omega_H$ with $\omega_H \ne 0$ (numerically $\omega_H = 1.2247/\tau$). □

A.5 Proof of Proposition 5 (instability after constant replacement)

At x^*, the constant replacement equals the live inhibitory signal, so that every stationary equation is unchanged. With S_{I} removed from the excitatory voltage equation, the excitatory coordinates form an autonomous linearized block,

$$M_{\mathrm{E}} = \begin{pmatrix} -1 & 2 & 0 \\ -2 & -1 & 3 \\ 2 & 0 & -2 \end{pmatrix}, \qquad p_{\mathrm{E}}(\lambda) = \det(\lambda I - M_{\mathrm{E}}) = \lambda^3 + 4\lambda^2 + 9\lambda - 2, \tag{A13}$$

in the order $(R_{\mathrm{E}}, V_{\mathrm{E}}, S_{\mathrm{E}})$ and in the time unit τ. Since $p_{\mathrm{E}}(0) = -2$ and $p_{\mathrm{E}}(\lambda) \to +\infty$ as $\lambda \to +\infty$, there is a positive real root λ_u; since $p'_{\mathrm{E}}(\lambda) = 3\lambda^2 + 8\lambda + 9 > 0$, it is the only real root, and, because the three roots sum to -4, the complex pair has real part $-(4 + \lambda_u)/2 < 0$. Numerically, $\lambda_u = 0.20298$. The inhibitory coordinates are downstream of this block, so that the full Jacobian $M(0)$ is block triangular, with characteristic polynomial $\chi_0 = (P - 12)(P + 4) = p_{\mathrm{E}}(\lambda)\,(\lambda^3 + 4\lambda^2 + 9\lambda + 14)$ (Eq. (A9) with $g = 0$); its spectrum contains λ_u, and the equilibrium is unstable. The cubic $P + 4$ is Hurwitz, because all

its coefficients are positive and $4 \cdot 9 > 14$; hence λ_u is the only eigenvalue of $M(0)$ in the closed right half-plane. □

A.6 Proof of Theorem 1 (minimum information rate)

We write $n = 3$ for the dimension of z, A_u and A_c for the restrictions of A to E_u and to the complementary invariant subspace E_c (spanned by the generalized eigenvectors of the eigenvalues with nonpositive real part), and $z = z_u + z_c$ for the corresponding decomposition. The input enters Eq. (17) additively through the constant vector b, and this is the only structural property of the network that the proof uses.

Step 1: adapted norm. Let $a_u = \min_i \operatorname{Re}\lambda_i > 0$ and fix numbers a and b_0 with $\max\{0, \max \operatorname{Re}\operatorname{spec}(A_c)\} < b_0 < a < a_u$. The matrix $A_c - b_0 I$ is Hurwitz, so that the Lyapunov equation $(A_c - b_0 I)^{\mathsf{T}} Q_c + Q_c (A_c - b_0 I) = -I$ has a positive-definite solution Q_c; with the inner product $\langle y, y' \rangle_c = y^{\mathsf{T}} Q_c y'$ on E_c, we obtain $\langle y, (A_c - b_0 I) y \rangle_c = -\frac{1}{2}|y|^2 \le 0$, that is, $\langle y, A_c y \rangle_c \le b_0 \|y\|_c^2$. Likewise, $aI - A_u$ is Hurwitz, and the same construction gives an inner product on E_u with $\langle y, A_u y \rangle_u \ge a \|y\|_u^2$. We equip $\mathbb{R}^n$ with the Euclidean norm $\|z\|^2 = \|z_u\|_u^2 + \|z_c\|_c^2$, in which $E_u \perp E_c$ and the projections Π_u and Π_c onto the two subspaces have norm at most one; all balls and operator norms below refer to this norm (the hypothesis of the theorem does not depend on the norm, because all norms on $\mathbb{R}^n$ are equivalent).

Step 2: an invariant cone. Let $\kappa(\varepsilon) = \sup_{\|z\| \le \varepsilon} \|DF_{\mathrm{E}}(z) - A\|$, which tends to zero as $\varepsilon \to 0$ because DF_{E} is continuous and $DF_{\mathrm{E}}(0) = A$, and fix $\varepsilon > 0$ so small that $\kappa := \kappa(\varepsilon) < (a - b_0)/(2\sqrt{2})$. Let $y(t)$ solve $\dot y = (A + E(t))y$ on an interval, with E continuous and $\|E(t)\| \le \kappa$, and let $\phi = \|y_u\|^2 - \|y_c\|^2$. Using $\langle y_u, \Pi_u E y \rangle \le \kappa \|y_u\| \, \|y\|$ and the analogous bound on E_c,

$$\dot\phi \ge 2a\|y_u\|^2 - 2b_0\|y_c\|^2 - 2\kappa\|y\|\big(\|y_u\| + \|y_c\|\big).$$

At any time at which $\phi = 0$ and $y \ne 0$, we have $\|y_u\| = \|y_c\| =: \varrho > 0$ and $\|y\| = \sqrt{2}\varrho$, so that $\dot\phi \ge 2(a - b_0)\varrho^2 - 4\sqrt{2}\kappa\varrho^2 > 0$. Consequently, if $\phi(t_0) \ge 0$, then $\phi(t) \ge 0$ for all later t: otherwise, at the last time t_* before the first negative value at which $\phi = 0$, we would have $\dot\phi(t_*) \le 0$, a contradiction (a nonzero solution of a linear equation never vanishes). In other words, the cone $\mathcal{C} = \{y : \|y_c\| \le \|y_u\|\}$ is forward invariant. Inside $\mathcal{C}$, $\|y\| \le \sqrt{2}\|y_u\|$, and therefore $\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}\|y_u\|^2 \ge a\|y_u\|^2 - \sqrt{2}\kappa\|y_u\|^2$, which gives

$$\|y_u(t)\| \ge e^{(a - \sqrt{2}\kappa)(t - t_0)} \, \|y_u(t_0)\|, \qquad a - \sqrt{2}\kappa > 0. \tag{A14}$$

Step 3: volume growth inside each cell. Let $\varepsilon_0 > 0$ be the radius that the hypothesis associates with ε, and let $\mathcal{K} = \{z_0 \in E_u : \|z_0\| \le \varepsilon_0\}$, a set of positive m-dimensional volume $\mathrm{vol}(\mathcal{K})$. Fix $t > 0$. The received sequence on $[0, t]$ is a (measurable) function of the initial excitatory state, all other variables of the network and of the coder being started from fixed values, so that it partitions $\mathcal{K}$ into at most $N(t)$ cells $\mathcal{K}_\ell$, one for each received sequence ℓ. All initial states of a cell receive the same input v_ℓ on $[0, t]$, and therefore evolve under the same nonautonomous equation $\dot z = F_{\mathrm{E}}(z) + b\, v_\ell(s)$; we denote by Φ_ℓ its time-t map and by $f_\ell = \Pi_u \Phi_\ell$, restricted to $\mathcal{K}_\ell$, the component of the final state in E_u. Because v_ℓ enters additively, the derivative $Y(s) = D\Phi_\ell^{(s)}(z_0)$ solves the variational equation $\dot Y = DF_{\mathrm{E}}(z(s))\, Y$, $Y(0) = I$, which does not involve v_ℓ; and, because the trajectory from $z_0 \in \mathcal{K}_\ell$ stays in the ball of radius ε, this equation has the form $\dot Y = (A + E(s))Y$ with $\|E(s)\| \le \kappa$.

(a) *Determinant.* For $y_0 \in E_u$, the solution $Y(s)y_0$ starts in $\mathcal{C}$ and remains there (Step 2). Hence $Y_u(s) = \Pi_u Y(s)|_{E_u}$ is invertible (if $\Pi_u Y(s) y_0 = 0$, then $Y(s)y_0 = 0$ and $y_0 = 0$), and $Z(s) = \Pi_c Y(s)|_{E_u} \, Y_u(s)^{-1}$ has norm at most one. The E_u component of the variational equation reads $\dot Y_u = (A_u + E_{uu} + E_{uc} Z)\, Y_u$, with $E_{uu} = \Pi_u E|_{E_u}$ and $E_{uc} = \Pi_u E|_{E_c}$, both of norm at most κ. By Jacobi's formula, and because the trace of an operator on an m-dimensional space is bounded in modulus by m times its norm,

$$\frac{\mathrm{d}}{\mathrm{d}s} \ln\det Y_u = \operatorname{tr}\big(A_u + E_{uu} + E_{uc} Z\big) \ge \operatorname{tr} A_u - 2m\kappa.$$

With $\det Y_u(0) = 1$ and $\operatorname{tr} A_u = \sum_i \operatorname{Re}\lambda_i$ (complex eigenvalues occur in conjugate pairs), $\det Df_\ell(z_0) = \det Y_u(t) \geq \exp[(\sum_i \operatorname{Re}\lambda_i - 2m\kappa)\,t]$.

(b) *Injectivity.* Let $z_0 \neq z_0'$ be two points of $\mathcal{K}_\ell$, with trajectories z and z'. Their difference $e = z - z'$ satisfies $\dot e = F_{\mathrm{E}}(z) - F_{\mathrm{E}}(z') = \big[\int_0^1 DF_{\mathrm{E}}(z' + \theta(z - z'))\,\mathrm{d}\theta\big]e = (A + \bar E(s))e$, in which $\|\bar E(s)\| \leq \kappa$ because the segment between $z(s)$ and $z'(s)$ lies in the ball of radius ε. Since $e(0) \in E_u \setminus \{0\}$ lies in $\mathcal{C}$, Eq. (A14) gives $\|\Pi_u e(t)\| \geq e^{(a-\sqrt{2}\kappa)t}\|e(0)\| > 0$, that is, $f_\ell(z_0) \neq f_\ell(z_0')$.

(c) *Volume.* By (a), (b) and the change-of-variables formula for injective continuously differentiable maps, the image $f_\ell(\mathcal{K}_\ell) \subset E_u$ has volume $\int_{\mathcal{K}_\ell} |\det Df_\ell| \geq e^{(\sum_i \operatorname{Re}\lambda_i - 2m\kappa)t}\operatorname{vol}(\mathcal{K}_\ell)$.

Step 4: counting. Every image $f_\ell(\mathcal{K}_\ell)$ lies in the ball of radius ε of E_u, whose volume is $\varepsilon^m|\mathbb{B}^m|$, where $|\mathbb{B}^m|$ is the volume of the unit ball of E_u. Summing over the at most $N(t) \leq 2^{Ct+c_0}$ cells,

$$\operatorname{vol}(\mathcal{K}) = \sum_\ell \operatorname{vol}(\mathcal{K}_\ell) \leq 2^{Ct+c_0}\, e^{-(\sum_i \operatorname{Re}\lambda_i - 2m\kappa)t}\, \varepsilon^m|\mathbb{B}^m|.$$

The left-hand side is a fixed positive number; taking logarithms, dividing by t and letting $t \to \infty$, we obtain $C \ln 2 \geq \sum_i \operatorname{Re}\lambda_i - 2m\kappa(\varepsilon)$.

Step 5: conclusion. The hypothesis provides an ε_0 for every $\varepsilon > 0$, so that the last inequality holds for every sufficiently small ε; letting $\varepsilon \to 0$, $\kappa(\varepsilon) \to 0$ and $C \geq h$. For the linearized network, $F_{\mathrm{E}}(z) = Az$ and $\kappa \equiv 0$, so that the same argument requires only that the trajectories starting in some ball remain in some bounded set (with ε replaced by the radius of that set). □

The argument is the volume argument of Nair and Evans (2004) and Tatikonda and Mitter (2004), extended here to the nonlinear network by the cone of Step 2, which controls the stable directions; it uses neither the topological feedback entropy of Nair et al. (2004) nor a discretization of time.

A.7 Proof of Corollary 1 (state-independent signals)

If the received signal does not depend on the state of the network (for a random signal, we condition on its realization), every initial state receives the same input v. Take ε and the cone $\mathcal{C}$ of Appendix A.6, and suppose that the level were stable: there would be $\varepsilon_0 > 0$ such that all trajectories starting within ε_0 remain within ε. Take two distinct initial states $z_0, z_0' \in E_u$ with norms at most ε_0. Their difference obeys $\dot e = (A + \bar E(s))e$ with $\|\bar E\| \leq \kappa$, as in Step 3(b), and starts in $\mathcal{C}$, so that $\|e(t)\| \geq e^{(a-\sqrt{2}\kappa)t}\|z_0 - z_0'\|$ grows without bound, which contradicts $\|e(t)\| \leq 2\varepsilon$. Hence the level is not stable, and a fortiori not locally exponentially stable. For the linearized network the conclusion follows from superposition alone: $z(t) - z'(t) = e^{At}(z_0 - z_0')$, which diverges whenever $z_0 - z_0'$ has a nonzero component in E_u. No information theory is needed; Theorem 1 gives the quantitative form of the statement, because a state-independent signal is a channel with $N(t) = 1$, that is, $C = 0 < h$. □

A.8 Proof of Proposition 6 (critical delay)

With the delay, the entry $-2g$ of B_g becomes $-2g\,e^{-\lambda d/\tau}$ in the elimination of Appendix A.4, and the characteristic function becomes $\chi(\lambda, d) = Q(\lambda) + 64g\,e^{-\lambda d/\tau}$, with $Q = P^2 - 8P - 48$ and λ in units of $1/\tau$. For $d = 0$ and $g = 1$, all roots have negative real parts (Proposition 3). The roots of this retarded quasi-polynomial depend continuously on d, and they can enter the right half-plane only by crossing the imaginary axis; $\lambda = 0$ is never a root, since $\chi(0, d) = 36$. Because $|e^{-i\omega d/\tau}| = 1$, a root $\lambda = i\omega$ requires $|Q(i\omega)| = 64$. With $\varpi = \omega^2$, the condition $\Psi(\varpi) := |Q(i\sqrt{\varpi})|^2 - 64^2 = 0$ is the polynomial equation

$$\varpi^6 - 4\varpi^5 + 70\varpi^4 + 68\varpi^3 - 3407\varpi^2 + 18888\varpi - 3312 = 0,$$

which has exactly one positive root, $\varpi_c = 0.18125$, so that $\omega_c = 0.42573$. The smallest d satisfying $64e^{-i\omega_c d/\tau} = -Q(i\omega_c)$ is $d_c = 1.5675\,\tau$. For $0 \leq d < d_c$, no root lies on the imaginary axis, and all roots therefore remain in the left half-plane. Because the equation is of retarded type, only finitely many roots

lie in any half-plane $\operatorname{Re}\lambda > -c$, so that the largest real part of the roots is negative; by the principle of linearized stability for retarded functional differential equations (Hale and Verduyn Lunel 1993), the level is then locally exponentially stable, and it is unstable whenever a root has positive real part. Moreover, $\Psi'(\varpi_c) = 1.77 \times 10^4 > 0$, so that, by the criterion of Cooke and van den Driessche (1986), the pair of roots on the imaginary axis at $d = d_c$ crosses from left to right as d increases; because ϖ_c is the only positive root of Ψ, every later crossing, at $d = d_c + 2\pi k\tau/\omega_c$ for integer $k \geq 1$, is also from left to right. Hence the level is unstable for every $d > d_c$, which proves the equivalence. □

A.9 Proof of Theorem 2 (conditions for an attracting level)

Part (i) is Proposition 1; Eq. (8) expresses the level as the fixed point of the loop rate → synaptic signal → input → rate. Part (ii) is Proposition 4 for the gain without delay and Proposition 6 for the delay at $g = 1$. Part (iii) is Theorem 1, with $m = 1$ and $\lambda_u = 0.20298/\tau$ (Proposition 5), together with Corollary 1. □

A.10 Proof of Proposition 7 (frequency response)

With the input, the second block row of the elimination in Appendix A.3 becomes $(\lambda+1)\delta V = -2\delta R + B_g\delta S + u_{\mathrm{E}}e_1$, where $e_1 = (1,0)^{\mathsf{T}}$ selects the excitatory population. The same elimination gives $\big[P(\lambda)I - 4B_g\big]\delta R = 2(\lambda+2)\,u_{\mathrm{E}}\,e_1$, and Cramer's rule yields $\delta R_{\mathrm{E}} = 2(\lambda+2)\,u_{\mathrm{E}}\,[P - 4(B_g)_{22}]/\det(PI - 4B_g)$. Since $(B_g)_{22} = -1$ and $\det(PI - 4B_g) = \chi_g$, this is Eq. (19), with $P + 4 = \lambda^3 + 4\lambda^2 + 9\lambda + 14$; at $\lambda = 0$, $\chi_g(0) = 64g - 28$ gives $T_g(0) = 2 \cdot 2 \cdot 14/(64g - 28) = 14/(16g - 7)$. Finally, $\chi_g = \chi_0 + 64g$ (Eq. (15)), so that $\chi_g/\chi_0 = 1 + 64g/\chi_0$, which is Eq. (20), and $T_g/T_0 = \chi_0/\chi_g$ because the numerator of T_g does not depend on g. □

A.11 Proof of Theorem 3 (sensitivity integral)

Part (i). By Appendix A.5, $\chi_0 = p_{\mathrm{E}}\,(P+4)$ has exactly one root in the closed right half-plane, λ_u, so that $\chi_0(\lambda) = (\lambda - \lambda_u)\chi_s(\lambda)$ with χ_s a Hurwitz polynomial of degree five. Let $\mathcal{B}(\lambda) = (\lambda - \lambda_u)/(\lambda + \lambda_u)$ be the Blaschke factor of λ_u and

$$S_{\mathrm{mp}}(\lambda) = \frac{S_g(\lambda)}{\mathcal{B}(\lambda)} = \frac{(\lambda + \lambda_u)\,\chi_s(\lambda)}{\chi_g(\lambda)}$$

the minimum-phase part of S_g. For g in the window, χ_g is Hurwitz (Proposition 4), and so is χ_s, so that all the zeros and poles of S_{mp}, which are finitely many, lie in the open left half-plane. There is therefore an $\epsilon > 0$ such that S_{mp} is analytic and has no zeros in the half-plane $H_\epsilon = \{\operatorname{Re}\lambda > -\epsilon\}$; being a ratio of monic polynomials of equal degree, $S_{\mathrm{mp}} \to 1$ as $|\lambda| \to \infty$. We now construct the logarithm that the argument uses. Since H_ϵ is simply connected and S_{mp} is analytic and zero-free on it, S_{mp} has an analytic logarithm on H_ϵ, that is, an analytic function ℓ with $e^{\ell} = S_{\mathrm{mp}}$, determined up to an additive constant $2\pi i k$ with k an integer (Rudin 1987). Choose ρ_0 such that $|S_{\mathrm{mp}}(\lambda) - 1| < \frac{1}{2}$ for $|\lambda| \geq \rho_0$; there, the series $\operatorname{Log} S_{\mathrm{mp}} = \sum_{n\geq 1}(-1)^{n+1}(S_{\mathrm{mp}} - 1)^n/n$ converges and is another logarithm of S_{mp}. On the connected set $H_\epsilon \cap \{|\lambda| > \rho_0\}$, the difference $\ell - \operatorname{Log} S_{\mathrm{mp}}$ is continuous and takes values in $2\pi i\mathbb{Z}$, so it is constant; we fix the free constant of ℓ so that this difference vanishes. With this normalization, $\ell(\lambda) = \operatorname{Log} S_{\mathrm{mp}}(\lambda) \to 0$ as $|\lambda| \to \infty$ in H_ϵ, and $\operatorname{Re}\ell = \ln|S_{\mathrm{mp}}|$ on H_ϵ. Moreover, H_ϵ is symmetric under complex conjugation and the polynomials have real coefficients, so that $\lambda \mapsto \overline{\ell(\bar\lambda)}$ is also an analytic logarithm of S_{mp} on H_ϵ that vanishes at infinity; by uniqueness, $\ell(\bar\lambda) = \overline{\ell(\lambda)}$, and in particular $\operatorname{Re}\ell(i\omega)$ is even and $\operatorname{Im}\ell(i\omega)$ is odd in ω. On the imaginary axis $|\mathcal{B}(i\omega)| = 1$, so that $\ln|S_g(i\omega)| = \ln|S_{\mathrm{mp}}(i\omega)| = \operatorname{Re}\ell(i\omega)$. For large $|\lambda|$, $S_g = 1 - 64g/\chi_g = 1 + O(\lambda^{-6})$ and $1/\mathcal{B} = 1 + 2\lambda_u/\lambda + O(\lambda^{-2})$, so that $S_{\mathrm{mp}} - 1 = 2\lambda_u/\lambda + O(\lambda^{-2})$, and the series of Log gives

$$\ell(\lambda) = \frac{2\lambda_u}{\lambda} + O(\lambda^{-2}).$$

We integrate ℓ around the closed contour formed by the segment of the imaginary axis from $-i\Omega$ to $i\Omega$ and the semicircle $\lambda = \Omega e^{i\vartheta}$, ϑ from $\pi/2$ to $-\pi/2$. By Cauchy's theorem the integral vanishes. On the semicircle, $\mathrm{d}\lambda = i\Omega e^{i\vartheta}\mathrm{d}\vartheta$, so that the leading term contributes $\int_{\pi/2}^{-\pi/2} 2\lambda_u i\,\mathrm{d}\vartheta = -2\pi i\lambda_u$, and the remainder contributes $O(\Omega^{-1})$. Letting $\Omega \to \infty$, with the integral taken between symmetric limits, $i\int_{-\infty}^{\infty} \ell(i\omega)\,\mathrm{d}\omega = 2\pi i\lambda_u$, that is, $\int_{-\infty}^{\infty} \ell(i\omega)\,\mathrm{d}\omega = 2\pi\lambda_u$. Since $\mathrm{Re}\,\ell(i\omega) = \ln|S_{\mathrm{mp}}(i\omega)|$ is even and $\mathrm{Im}\,\ell(i\omega)$ is odd in ω, taking real parts gives $2\int_0^{\infty} \ln|S_{\mathrm{mp}}(i\omega)|\,\mathrm{d}\omega = 2\pi\lambda_u$, which, with $\ln|S_g| = \ln|S_{\mathrm{mp}}|$ on the axis, is Eq. (21). In physical time ($\omega_{\mathrm{phys}} = \omega/\tau = 2\pi f$ and $\lambda_u/\tau = h\ln 2$ by Theorem 1 with $m = 1$), Eq. (21) becomes $2\pi\ln 2\int_0^{\infty} \log_2|S_g|\,\mathrm{d}f = \pi h\ln 2$, which is Eq. (22).

Part (ii). For g_1, g_2 in the window, $T_{g_1}/T_{g_2} = \chi_{g_2}/\chi_{g_1} = 1 + 64(g_2 - g_1)/\chi_{g_1}$ (Appendix A.10). This function is analytic and has no zeros on a half-plane H_ϵ as in Part (i), since both polynomials are Hurwitz, and it tends to 1 as $|\lambda| \to \infty$. Its analytic logarithm on H_ϵ, normalized at infinity as in Part (i), equals $\mathrm{Log}(T_{g_1}/T_{g_2}) = O(\lambda^{-6})$ for large $|\lambda|$, has real part $\ln|T_{g_1}/T_{g_2}|$, and has an imaginary part that is odd on the imaginary axis. The contour argument of Part (i) then applies to this logarithm with no contribution from the semicircle, and taking real parts gives Eq. (23). □

A.12 Proof of Theorem 4 (noisy inhibitory pathway)

Time is measured in units of τ. We recall that $M(g) = M_0 + g\,b_6 e_6^{\mathsf{T}}$, with $b_6 = -2e_2$, so that Eq. (24) is the linear stochastic equation $\mathrm{d}\zeta = M(g)\zeta\,\mathrm{d}t + \sigma b_6\,\mathrm{d}W$.

Part (i). Step 1: removal of the feedback. The term $b_6\,\mathrm{d}\mathcal{Y}$ is known to an observer of $\mathcal{Y}$. Writing $\zeta(t) = e^{M_0 t}\zeta(0) + k(t)$, with $k(t) = \int_0^t e^{M_0(t-s)}b_6\,\mathrm{d}\mathcal{Y}(s)$ a functional of $\mathcal{Y}$ on $[0,t]$, we define $\tilde{\mathcal{Y}}(t) = \mathcal{Y}(t) - g\int_0^t e_6^{\mathsf{T}}k(s)\,\mathrm{d}s$, so that

$$\mathrm{d}\tilde{\mathcal{Y}} = g\,e_6^{\mathsf{T}}e^{M_0 t}\zeta(0)\,\mathrm{d}t + \sigma\,\mathrm{d}W. \tag{A15}$$

The map $\mathcal{Y} \mapsto \tilde{\mathcal{Y}}$ is causal and invertible: given $\tilde{\mathcal{Y}}$, k is the unique solution of the linear equation $\mathrm{d}k = M(g)k\,\mathrm{d}t + b_6\,\mathrm{d}\tilde{\mathcal{Y}}$, $k(0) = 0$, and $\mathcal{Y}$ is recovered from $\mathrm{d}\mathcal{Y} = \mathrm{d}\tilde{\mathcal{Y}} + g\,e_6^{\mathsf{T}}k\,\mathrm{d}t$. Hence $I(\zeta(0);\mathcal{Y}_{[0,t]}) = I(\zeta(0);\tilde{\mathcal{Y}}_{[0,t]})$.

Step 2: a Gaussian channel without feedback. Equation (A15) observes the Gaussian vector $\zeta(0)$ through the deterministic row $c(s)^{\mathsf{T}} = g\,e_6^{\mathsf{T}}e^{M_0 s}$ in white noise independent of $\zeta(0)$. By Girsanov's theorem, the likelihood of $\tilde{\mathcal{Y}}_{[0,t]}$ given $\zeta(0)$ is proportional to $\exp\{\sigma^{-2}[\zeta(0)^{\mathsf{T}}\int_0^t c\,\mathrm{d}\tilde{\mathcal{Y}} - \tfrac{1}{2}\zeta(0)^{\mathsf{T}}\mathcal{G}_t\zeta(0)]\}$, with the observability Gramian $\mathcal{G}_t = \int_0^t c(s)c(s)^{\mathsf{T}}\mathrm{d}s$, so that $\int_0^t c\,\mathrm{d}\tilde{\mathcal{Y}} = \mathcal{G}_t\zeta(0) + \sigma\int_0^t c\,\mathrm{d}W$ is a sufficient statistic (Liptser and Shiryaev 2001). Its noise is Gaussian with covariance $\sigma^2\mathcal{G}_t$ and independent of $\zeta(0)$, and the mutual information of a Gaussian vector observed in this way is

$$I(\zeta(0);\mathcal{Y}_{[0,t]}) = \tfrac{1}{2}\ln\det\left(I + \Sigma_0\mathcal{G}_t/\sigma^2\right) = \tfrac{1}{2}\ln\det\left(I + \Sigma_0^{1/2}\mathcal{G}_t\Sigma_0^{1/2}/\sigma^2\right), \tag{A16}$$

the second form following from Sylvester's determinant identity.

Step 3: growth of the Gramian. By Appendix A.5, M_0 has the simple real eigenvalue $\lambda_u > 0$, and all its other eigenvalues have real parts at most $-\mu_0 < 0$. Let v_u and ℓ_u^{T} be right and left eigenvectors of λ_u, with $\ell_u^{\mathsf{T}}v_u = 1$. Then $e^{M_0 s} = e^{\lambda_u s}v_u\ell_u^{\mathsf{T}} + E_s(s)$, with $\|E_s(s)\| \le K_0 e^{-\mu_0 s/2}$ for some constant K_0. Moreover, $c_u := e_6^{\mathsf{T}}v_u \neq 0$: otherwise $M(g)v_u = M_0 v_u + g\,b_6(e_6^{\mathsf{T}}v_u) = \lambda_u v_u$ for every g, contradicting the stability of $M(1)$ (Proposition 3). (The unstable mode is visible in the inhibitory signal, as it must be if inhibition is to stabilize it.)

Lower bound. For a positive-semidefinite matrix X, $\det(I + X) \ge 1 + \lambda_{\max}(X)$, and the Rayleigh quotient with the vector $\Sigma_0^{-1/2}v_u$ gives $\lambda_{\max}(\Sigma_0^{1/2}\mathcal{G}_t\Sigma_0^{1/2}) \ge v_u^{\mathsf{T}}\mathcal{G}_t v_u/(v_u^{\mathsf{T}}\Sigma_0^{-1}v_u)$. Since $c(s)^{\mathsf{T}}v_u = g\,c_u\,e^{\lambda_u s}$, we have $v_u^{\mathsf{T}}\mathcal{G}_t v_u = g^2c_u^2(e^{2\lambda_u t} - 1)/(2\lambda_u)$, and therefore $I(\zeta(0);\mathcal{Y}_{[0,t]}) \ge \tfrac{1}{2}\ln[1 + k_1(e^{2\lambda_u t} - 1)]$ with $k_1 > 0$.

Upper bound. For every vector y, $(c(s)^{\mathsf{T}}y)^2 \le 2g^2c_u^2e^{2\lambda_u s}(\ell_u^{\mathsf{T}}y)^2 + 2g^2K_0^2e^{-\mu_0 s}\|y\|^2$, and integration gives $\mathcal{G}_t \preceq k_2e^{2\lambda_u t}\ell_u\ell_u^{\mathsf{T}} + k_3 I$, with $k_2 = g^2c_u^2/\lambda_u$ and $k_3 = 2g^2K_0^2/\mu_0$. The determinant in Eq. (A16) is

monotone in this order, and, with $D = I + k_3\Sigma_0/\sigma^2$ and the matrix determinant lemma,

$$\det\left(I + \Sigma_0^{1/2}\mathcal{G}_t\Sigma_0^{1/2}/\sigma^2\right) \le \det D\left(1 + \frac{k_2 e^{2\lambda_u t}}{\sigma^2}\,\ell_u^{\mathsf{T}}\Sigma_0^{1/2}D^{-1}\Sigma_0^{1/2}\ell_u\right) \le \det D\left(1 + k_4 e^{2\lambda_u t}\right).$$

Together, the two bounds give $I(\zeta(0); \mathcal{Y}_{[0,t]}) = \lambda_u t + O(1)$, so that $\bar{I} = \lambda_u$ nats per unit time. The constants depend on g, σ and Σ_0, but the rate does not, and neither bound requires $M(g)$ to be stable; hence the conclusion holds for every $g \neq 0$ and $\sigma > 0$. In physical time, $\bar{I} = \lambda_u/\tau$ nats/s, that is, h bits/s.

Part (ii). For g in the window, $M(g)$ is Hurwitz, so that ζ and v converge to stationary processes. The matrix determinant lemma gives the return-difference identity $\chi_g = \chi_0(1 - gG)$, with $G(\lambda) = e_6^{\mathsf{T}}(\lambda I - M_0)^{-1}b_6$ the transfer function from the received to the sent inhibitory signal; by Eq. (24), $\mathrm{d}\zeta = M(g)\zeta\,\mathrm{d}t + \sigma b_6\,\mathrm{d}W$ and $v = g\,e_6^{\mathsf{T}}\zeta + \sigma\xi$, so that in the stationary regime v is white noise passed through the filter $\sigma[1 + g\,e_6^{\mathsf{T}}(\lambda I - M(g))^{-1}b_6]$, which is stable because $M(g)$ is Hurwitz. By the Sherman–Morrison identity, $1 + g\,e_6^{\mathsf{T}}(\lambda I - M(g))^{-1}b_6 = 1/(1 - gG) = S_g$ as rational functions, with $S_g = \chi_0/\chi_g$ as in Eq. (20); the unstable open-loop G enters only through this identity, so that $v = \sigma S_g\,\xi$. A stable filter applied to white noise of intensity σ^2 yields the spectral density $\Phi_v = \sigma^2|S_g(i\omega)|^2$, and Theorem 3 gives $\frac{1}{4\pi}\int_{-\infty}^{\infty}\ln(\Phi_v/\sigma^2)\,\mathrm{d}\omega = \frac{1}{\pi}\int_0^{\infty}\ln|S_g(i\omega)|\,\mathrm{d}\omega = \lambda_u$.

Part (iii). In the stationary regime, the state-dependent part of the received signal is $gw = v - \sigma\xi = \sigma(S_g - 1)\xi$, with power $P_{\mathrm{s}} = \lim_{t\to\infty}\mathbb{E}[(g\,w(t))^2] = \frac{\sigma^2}{2\pi}\int_{-\infty}^{\infty}|S_g(i\omega) - 1|^2\mathrm{d}\omega$. We expand $|S_g - 1|^2 = |S_g|^2 - 1 - 2\,\mathrm{Re}(S_g - 1)$. Since $S_g - 1 = -64g/\chi_g$ is analytic in the closed right half-plane and decays as $|\lambda|^{-6}$, the contour argument of Appendix A.11, now with no contribution from the semicircle, gives $\int_{-\infty}^{\infty}(S_g(i\omega) - 1)\,\mathrm{d}\omega = 0$, and so $P_{\mathrm{s}}/\sigma^2 = \frac{1}{\pi}\int_0^{\infty}(|S_g|^2 - 1)\,\mathrm{d}\omega$. For every $y > 0$, $\ln y \le y - 1$, with equality only at $y = 1$; with $y = |S_g(i\omega)|^2$, $|S_g|^2 - 1 \ge 2\ln|S_g|$, so that $P_{\mathrm{s}}/\sigma^2 \ge \frac{2}{\pi}\int_0^{\infty}\ln|S_g|\,\mathrm{d}\omega = 2\lambda_u$. Equality would require $|S_g(i\omega)| = 1$ for almost every ω, which would make the integral of $\ln|S_g|$ vanish, contradicting Theorem 3; the inequality is therefore strict. The white Gaussian channel with power P_{s} and noise intensity σ^2 (one-sided spectral density $2\sigma^2$) has capacity $C = P_{\mathrm{s}}/(2\sigma^2)$ nats per unit time (Cover and Thomas 2006), which therefore exceeds $\lambda_u = \bar{I}$. □

A.13 Proof of Corollary 2 (intrinsic fluctuations)

By the chain rule of mutual information, $I(\zeta(0), \mathcal{U}_{[0,t]}; \mathcal{Y}_{[0,t]}) = I(\mathcal{U}_{[0,t]}; \mathcal{Y}_{[0,t]}) + I(\zeta(0); \mathcal{Y}_{[0,t]} \mid \mathcal{U}_{[0,t]}) \ge I(\zeta(0); \mathcal{Y}_{[0,t]} \mid \mathcal{U}_{[0,t]})$. Conditional on a realization of $\mathcal{U}$, which is independent of $\zeta(0)$ and W, the term $\Gamma\,\mathrm{d}\mathcal{U}$ is a known input, and Step 1 of Appendix A.12, applied with $k(t)$ now including $\int_0^t e^{M_0(t-s)}\Gamma\,\mathrm{d}\mathcal{U}(s)$, reduces the problem to Eq. (A15). Therefore $I(\zeta(0); \mathcal{Y}_{[0,t]} \mid \mathcal{U}_{[0,t]})$ is given by Eq. (A16) and equals $\lambda_u t + O(1)$, which proves the inequality. For the Gaussian spectral formula, the independence of $\mathcal{U}$ and W gives $\Phi_v = \sigma^2|S_g|^2 + \Phi_{\mathcal{U}}$, where $\Phi_{\mathcal{U}}$ is the spectral density of the part of v driven by $\mathcal{U}$, and the identity $\ln(\Phi_v/\sigma^2) = \ln|S_g|^2 + \ln\left(1 + \Phi_{\mathcal{U}}/(\sigma^2|S_g|^2)\right)$, integrated with Theorem 3, gives Eq. (27). □

A.14 Proof of Proposition 8 (sensitivity to sustained input)

Part (i). With the input, the stationary equations of the construction are $G_{\mathrm{E}}(R_{\mathrm{E}}, R_{\mathrm{I}}) + u_{\mathrm{E}} = 0$ and $G_{\mathrm{I}}(R_{\mathrm{E}}, R_{\mathrm{I}}) = 0$, with $G_{\mathrm{E}} = \varphi(R_{\mathrm{E}}) + 3R_{\mathrm{E}} - 2R_{\mathrm{I}}$ and $G_{\mathrm{I}} = \varphi(R_{\mathrm{I}}) + 2R_{\mathrm{E}} - R_{\mathrm{I}}$ (φ as in Appendix A.1). At $R_{\mathrm{E}} = R_{\mathrm{I}} = 1$, $\varphi'(1) = -1/2 - 2 = -5/2$, so that the Jacobian of $(G_{\mathrm{E}}, G_{\mathrm{I}})$ with respect to $(R_{\mathrm{E}}, R_{\mathrm{I}})$ is $\left(\begin{smallmatrix} 1/2 & -2 \\ 2 & -7/2 \end{smallmatrix}\right)$, with determinant $9/4 \neq 0$. By the implicit function theorem,

$$\begin{pmatrix} 1/2 & -2 \\ 2 & -7/2 \end{pmatrix}\begin{pmatrix} \mathrm{d}R_{\mathrm{E}}^*/\mathrm{d}u_{\mathrm{E}} \\ \mathrm{d}R_{\mathrm{I}}^*/\mathrm{d}u_{\mathrm{E}} \end{pmatrix} = -\begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

and Cramer's rule gives $\mathrm{d}R_{\mathrm{E}}^*/\mathrm{d}u_{\mathrm{E}} = (7/2)/(9/4) = 14/9$ and $\mathrm{d}R_{\mathrm{I}}^*/\mathrm{d}u_{\mathrm{E}} = 2/(9/4) = 8/9$.

Part (ii). An uncoupled population with the same $\Delta = 1$ and with median excitability $\bar{\eta} = 3/4$ has the stationary condition $1/(4R^2) + \bar{\eta} + u - R^2 = 0$, whose solution at $u = 0$ is $R = 1$. Implicit

differentiation gives $\mathrm{d}R/\mathrm{d}u = [1/(2R^3) + 2R]^{-1} = 2/5$ at $R = 1$. The ratio of the two sensitivities is $(14/9)/(2/5) = 35/9$. □

A.15 Proof of Lemma 1 (conditional information)

Conditional on $(\mathsf{Y}_n, \mathsf{Z}_n)$, the variable Y_{n+1} in Eq. (29) is Gaussian with variance $\beta_1^2\mu + \nu^2$, because the only remaining sources of its uncertainty are the term $\beta_1\mathsf{X}_n$, whose conditional variance is $\beta_1^2\mu$, and the independent innovation; conditional additionally on X_n, its variance is ν^2. The conditional entropy of a scalar Gaussian variable with variance s^2 is $\frac{1}{2}\log_2(2\pi e s^2)$, and the conditional mutual information is the difference of the two conditional entropies, which gives Eq. (30) (Cover and Thomas 2006). □

A.16 Proof of Proposition 9 (blindness to a sustained mean shift)

Since A_o is Hurwitz, the stationary solution is $x_\mathrm{o}(t) = -A_\mathrm{o}^{-1}B_\mathrm{o}u + \tilde{x}(t)$, in which the fluctuating part $\tilde{x}$ solves $\dot{\tilde{x}} = A_\mathrm{o}\tilde{x} + \Sigma_\mathrm{o}\xi$ and does not involve u. Thus u shifts only the mean of a Gaussian process whose covariance and autocovariance functions depend on $(A_\mathrm{o}, \Sigma_\mathrm{o})$ alone. All the listed information quantities are differences of Gaussian entropies, which are functions of covariance matrices only, and are therefore independent of u (Cover and Thomas 2006; Barnett et al. 2009); the stationary mean of the output $C_\mathrm{o}x_\mathrm{o}$ is $-C_\mathrm{o}A_\mathrm{o}^{-1}B_\mathrm{o}u$. □

References

Ahmadian Y, Miller KD (2021) What is the dynamical regime of cerebral cortex? Neuron 109(21):3373–3391. https://doi.org/10.1016/j.neuron.2021.07.031

Ardestanizadeh E, Franceschetti M (2012) Control-theoretic approach to communication with feedback. IEEE Transactions on Automatic Control 57(10):2576–2587. https://doi.org/10.1109/TAC.2012.2191047

Barnett L, Barrett AB, Seth AK (2009) Granger causality and transfer entropy are equivalent for gaussian variables. Physical Review Letters 103:238701. https://doi.org/10.1103/PhysRevLett.103.238701

Braslavsky JH, Middleton RH, Freudenberg JS (2007) Feedback stabilization over signal-to-noise ratio constrained channels. IEEE Transactions on Automatic Control 52(8):1391–1403. https://doi.org/10.1109/TAC.2007.902739

Briat C, Gupta A, Khammash M (2016) Antithetic integral feedback ensures robust perfect adaptation in noisy biomolecular networks. Cell Systems 2(1):15–26. https://doi.org/10.1016/j.cels.2016.01.004

Brockett RW, Liberzon D (2000) Quantized feedback stabilization of linear systems. IEEE Transactions on Automatic Control 45(7):1279–1289. https://doi.org/10.1109/9.867021

Brunel N (2000) Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. Journal of Computational Neuroscience 8(3):183–208. https://doi.org/10.1023/A:1008925309027

Chandra FA, Buzi G, Doyle JC (2011) Glycolytic oscillations and limits on robust efficiency. Science 333:187–192. https://doi.org/10.1126/science.1200705

Colonius F (2012) Minimal bit rates and entropy for exponential stabilization. SIAM Journal on Control and Optimization 50(5):2988–3010. https://doi.org/10.1137/110829271

Colonius F, Kawan C (2009) Invariance entropy for control systems. SIAM Journal on Control and Optimization 48(3):1701–1721

Cooke KL, van den Driessche P (1986) On zeroes of some transcendental equations. Funkcialaj Ekvacioj 29:77–90

Cover TM, Thomas JA (2006) Elements of Information Theory, 2nd edn. John Wiley & Sons, Hoboken, New Jersey, https://doi.org/10.1002/047174882X

Cramer B, Stöckel D, Kreft M, et al (2020) Control of criticality and computation in spiking neuromorphic networks with plasticity. Nature Communications 11:2853. https://doi.org/10.1038/s41467-020-16548-3

Del Papa B, Priesemann V, Triesch J (2017) Criticality meets learning: Criticality signatures in a self-organizing recurrent neural network. PLOS ONE 12(5):e0178683. https://doi.org/10.1371/journal.pone.0178683

Delchamps DF (1990) Stabilizing a linear system with quantized state feedback. IEEE Transactions on Automatic Control 35(8):916–924. https://doi.org/10.1109/9.58500

Dormand JR, Prince PJ (1980) A family of embedded Runge–Kutta formulae. Journal of Computational and Applied Mathematics 6(1):19–26. https://doi.org/10.1016/0771-050X(80)90013-3

Efron B, Tibshirani RJ (1993) An Introduction to the Bootstrap. Chapman and Hall, New York, https://doi.org/10.1201/9780429246593

Elia N (2004) When Bode meets Shannon: control-oriented feedback communication schemes. IEEE Transactions on Automatic Control 49(9):1477–1488. https://doi.org/10.1109/TAC.2004.834119

Francis BA, Wonham WM (1976) The internal model principle of control theory. Automatica 12(5):457–465. https://doi.org/10.1016/0005-1098(76)90006-6

Freudenberg JS, Looze DP (1985) Right half plane poles and zeros and design tradeoffs in feedback systems. IEEE Transactions on Automatic Control 30(6):555–565. https://doi.org/10.1109/TAC.1985.1104004

Hale JK, Verduyn Lunel SM (1993) Introduction to Functional Differential Equations, Applied Mathematical Sciences, vol 99. Springer, New York, https://doi.org/10.1007/978-1-4612-4342-7

Harris CR, Millman KJ, van der Walt SJ, et al (2020) Array programming with NumPy. Nature 585(7825):357–362. https://doi.org/10.1038/s41586-020-2649-2

Hengen KB, Shew WL (2025) Is criticality a unified setpoint of brain function? Neuron 113(16):2582–2598.e2. https://doi.org/10.1016/j.neuron.2025.05.020

Hengen KB, Lambo ME, Van Hooser SD, et al (2013) Firing rate homeostasis in visual cortex of freely behaving rodents. Neuron 80(2):335–342. https://doi.org/10.1016/j.neuron.2013.08.038

Hengen KB, Torrado Pacheco A, McGregor JN, et al (2016) Neuronal firing rate homeostasis is inhibited by sleep and promoted by wake. Cell 165(1):180–191. https://doi.org/10.1016/j.cell.2016.01.046

Hunter JD (2007) Matplotlib: A 2D graphics environment. Computing in Science & Engineering 9(3):90–95. https://doi.org/10.1109/MCSE.2007.55

Izhikevich EM (2000) Neural excitability, spiking and bursting. International Journal of Bifurcation and Chaos 10(6):1171–1266. https://doi.org/10.1142/S0218127400000840

Kadota TT, Zakai M, Ziv J (1971) Mutual information of the white Gaussian channel with and without feedback. IEEE Transactions on Information Theory 17(4):368–371. https://doi.org/10.1109/TIT.1971.1054664

Kaplan DM (2011) Explanation and description in computational neuroscience. Synthese 183:339–373. https://doi.org/10.1007/s11229-011-9970-0

Kawan C (2013) Invariance Entropy for Deterministic Control Systems: An Introduction, Lecture Notes in Mathematics, vol 2089. Springer, Cham, https://doi.org/10.1007/978-3-319-01288-9

Khalil HK (2002) Nonlinear Systems, 3rd edn. Prentice Hall, Upper Saddle River, NJ

Kim YH (2010) Feedback capacity of stationary Gaussian channels. IEEE Transactions on Information Theory 56(1):57–85

Kuznetsov YA (2004) Elements of Applied Bifurcation Theory, 3rd edn. Springer, New York, https://doi.org/10.1007/978-1-4757-3978-7

Leighton MP, Betancourt JM, Emonet T, et al (2026) On the information required for feedback control. arXiv preprint arXiv:260716639 https://doi.org/10.48550/arXiv.2607.16639

Lestas I, Vinnicombe G, Paulsson J (2010) Fundamental limits on the suppression of molecular fluctuations. Nature 467(7312):174–178. https://doi.org/10.1038/nature09333

Liptser RS, Shiryaev AN (2001) Statistics of Random Processes I: General Theory, 2nd edn. Springer, Berlin, https://doi.org/10.1007/978-3-662-13043-8

Ma Z, Turrigiano GG, Wessel R, et al (2019) Cortical circuit dynamics are homeostatically tuned to criticality in vivo. Neuron 104(4):655–664.e4. https://doi.org/10.1016/j.neuron.2019.08.031

Martins NC, Dahleh MA (2008) Feedback control in the presence of noisy channels: "Bode-like" fundamental limitations of performance. IEEE Transactions on Automatic Control 53(7):1604–1615. https://doi.org/10.1109/TAC.2008.929361

Massey JL (1990) Causality, feedback and directed information. In: Proceedings of the International Symposium on Information Theory and its Applications (ISITA), pp 303–305

Meurer A, Smith CP, Paprocki M, et al (2017) SymPy: symbolic computing in Python. PeerJ

Computer Science 3:e103. https://doi.org/10.7717/peerj-cs.103

Middleton RH, Miller DE (2007) On the achievable delay margin using LTI control for unstable plants. IEEE Transactions on Automatic Control 52(7):1194–1207. https://doi.org/10.1109/TAC.2007.900824

Montbrió E, Pazó D, Roxin A (2015) Macroscopic description for networks of spiking neurons. Physical Review X 5:021028. https://doi.org/10.1103/PhysRevX.5.021028

mpmath developers (n.d.) Contexts: Arbitrary precision interval arithmetic. Official mpmath documentation, URL https://mpmath.readthedocs.io/en/latest/contexts.html, accessed 2026-10-03

Nair GN, Evans RJ (2004) Stabilizability of stochastic linear systems with finite feedback data rates. SIAM Journal on Control and Optimization 43(2):413–436. https://doi.org/10.1137/S0363012902402116

Nair GN, Evans RJ, Mareels IMY, et al (2004) Topological feedback entropy and nonlinear stabilization. IEEE Transactions on Automatic Control 49(9):1585–1597. https://doi.org/10.1109/TAC.2004.834105

Nair GN, Fagnani F, Zampieri S, et al (2007) Feedback control under data rate constraints: An overview. Proceedings of the IEEE 95(1):108–137. https://doi.org/10.1109/JPROC.2006.887294

Nakahira Y, Liu Q, Sejnowski TJ, et al (2021) Diversity-enabled sweet spots in layered architectures and speed–accuracy trade-offs in sensorimotor control. Proceedings of the National Academy of Sciences 118(22):e1916367118. https://doi.org/10.1073/pnas.1916367118

Newman JP, Fong Mf, Millard DC, et al (2015) Optogenetic feedback control of neural activity. eLife 4:e07192. https://doi.org/10.7554/eLife.07192

O'Leary T, Williams AH, Franci A, et al (2014) Cell types, network homeostasis, and pathological compensation from a biologically plausible ion channel expression model. Neuron 82(4):809–821. https://doi.org/10.1016/j.neuron.2014.04.002

Ott E, Antonsen TM (2008) Low dimensional behavior of large systems of globally coupled oscillators. Chaos 18(3):037113. https://doi.org/10.1063/1.2930766

Ozeki H, Finn IM, Schaffer ES, et al (2009) Inhibitory stabilization of the cortical network underlies visual surround suppression. Neuron 62(4):578–592. https://doi.org/10.1016/j.neuron.2009.03.028

Pintelon R, Schoukens J (2012) System Identification: A Frequency Domain Approach, 2nd edn. Wiley-IEEE Press, Hoboken, New Jersey, https://doi.org/10.1002/9781118287422

Pouille F, Scanziani M (2001) Enforcement of temporal fidelity in pyramidal cells by somatic feed-forward inhibition. Science 293(5532):1159–1163. https://doi.org/10.1126/science.1060342

Ramsay DS, Woods SC (2014) Clarifying the roles of homeostasis and allostasis in physiological regulation. Psychological Review 121(2):225–247. https://doi.org/10.1037/a0035942

Rudin W (1987) Real and Complex Analysis, 3rd edn. McGraw-Hill, New York

Sanzeni A, Akitake B, Goldbach HC, et al (2020) Inhibition stabilization is a widespread property of cortical networks. eLife 9:e54875. https://doi.org/10.7554/eLife.54875

Schoukens J, Pintelon R, Dobrowiecki T, et al (2005) Identification of linear systems with nonlinear distortions. Automatica 41(3):491–504. https://doi.org/10.1016/j.automatica.2004.10.004

Schreiber T (2000) Measuring information transfer. Physical Review Letters 85(2):461–464. https://doi.org/10.1103/PhysRevLett.85.461

SciPy developers (n.d.) scipy.stats.qmc.LatinHypercube. Official SciPy API documentation, URL https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.qmc.LatinHypercube.html, accessed 2026-10-03

Shew WL, Yang H, Yu S, et al (2011) Information capacity and transmission are maximized in balanced cortical networks with neuronal avalanches. Journal of Neuroscience 31(1):55–63. https://doi.org/10.1523/JNEUROSCI.4637-10.2011

Soldado-Magraner S, Seay MJ, Laje R, et al (2022) Paradoxical self-sustained dynamics emerge from orchestrated excitatory and inhibitory homeostatic plasticity rules. Proceedings of the National Academy of Sciences 119(43):e2200621119. https://doi.org/10.1073/pnas.2200621119

Sontag ED (2003) Adaptation and regulation with signal detection implies internal model. Systems & Control Letters 50(2):119–126. https://doi.org/10.1016/S0167-6911(03)00136-1

Speakman JR, Levitsky DA, Allison DB, et al (2011) Set points, settling points and some alternative models: theoretical options to understand how genes and environments combine to regulate body adiposity. Disease Models & Mechanisms 4(6):733–745. https://doi.org/10.1242/dmm.008698

Strogatz SH (2015) Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering, 2nd edn. Westview Press, Boulder, CO

Styr B, Gonen N, Zarhin D, et al (2019) Mitochondrial regulation of the hippocampal firing rate set point and seizure susceptibility. Neuron 102(5):1009–1024.e8. https://doi.org/10.1016/j.neuron.2019.03.045

Tatikonda S, Mitter S (2004) Control under communication constraints. IEEE Transactions on Automatic Control 49(7):1056–1068. https://doi.org/10.1109/TAC.2004.831187

Touchette H, Lloyd S (2000) Information-theoretic limits of control. Physical Review Letters 84(6):1156–1159. https://doi.org/10.1103/PhysRevLett.84.1156

Tsodyks MV, Skaggs WE, Sejnowski TJ, et al (1997) Paradoxical effects of external modulation of inhibitory interneurons. Journal of Neuroscience 17(11):4382–4388. https://doi.org/10.1523/JNEUROSCI.17-11-04382.1997

Turrigiano GG, Leslie KR, Desai NS, et al (1998) Activity-dependent scaling of quantal amplitude in neocortical neurons. Nature 391:892–896. https://doi.org/10.1038/36103

Vicente R, Wibral M, Lindner M, et al (2011) Transfer entropy—a model-free measure of effective connectivity for the neurosciences. Journal of Computational Neuroscience 30(1):45–67. https://doi.org/10.1007/s10827-010-0262-3

Virtanen P, Gommers R, Oliphant TE, et al (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods 17:261–272. https://doi.org/10.1038/s41592-019-0686-2

Vogels TP, Sprekeler H, Zenke F, et al (2011) Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. Science 334(6062):1569–1573. https://doi.org/10.1126/science.1211095

van Vreeswijk C, Sompolinsky H (1996) Chaos in neuronal networks with balanced excitatory and inhibitory activity. Science 274(5293):1724–1726. https://doi.org/10.1126/science.274.5293.1724

Weissman T, Kim YH, Permuter HH (2013) Directed information, causal estimation, and communication in continuous time. IEEE Transactions on Information Theory 59(3):1271–1287. https://doi.org/10.1109/TIT.2012.2227677

Welch PD (1967) The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Transactions on Audio and Electroacoustics 15(2):70–73. https://doi.org/10.1109/TAU.1967.1161901

Wirtshafter D, Davis JD (1977) Set points, settling points, and the control of body weight. Physiology & Behavior 19(1):75–78. https://doi.org/10.1016/0031-9384(77)90162-7

Yi TM, Huang Y, Simon MI, et al (2000) Robust perfect adaptation in bacterial chemotaxis through integral feedback control. Proceedings of the National Academy of Sciences 97(9):4649–4653. https://doi.org/10.1073/pnas.97.9.4649

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementary.pdf