A time-resolved theory of information encoding in recurrent neural networks

Anonymous Author(s)
Affiliation
Address
email

Abstract

Mammalian brains process information by the collective dynamics of a deeply layered structure of recurrent networks. Information transmission in neural circuits depends on how well time-varying stimuli are encoded by neural populations. A dynamic balance of externally incoming currents by strong recurrent inhibition was previously proposed as a mechanism to accurately and robustly encode the information in a time-varying stimulus, but a full theory was missing. Here, we develop a non-stationary dynamic mean-field theory that transparently explains how tight balance of excitatory currents by recurrent inhibition improves information encoding. We demonstrate that the mutual information rate of a time-varying input increases linearly with the tightness of balance, both in the presence of additive noise and with recurrent chaotic network fluctuations. We corroborated our findings in numerical experiments and demonstrated that recurrent networks with positive firing rates trained to transmit a time-varying stimulus generically use recurrent inhibition to increase the information rate. We also found that networks trained to transmit multiple independent time-varying signals spontaneously form multiple local inhibitory clusters, one for each input channel. Our findings suggest that feedforward excitatory synaptic projections and local recurrent inhibition - as observed in many biological circuits - is a generic circuit motif for encoding and transmitting time-varying information in recurrent neural circuits.

1 Introduction

How fast and reliable time-varying incoming stimuli are encoded in the population activity of recurrent neural networks constrains information transmission between local circuits and inter-areal communication. While the brain has to respond rapidly and reliably to changes in the world, the time scales of synapses can be slow, single neurons exhibit Poisson-like temporally irregular dynamics and dynamic instability of recurrent neural circuit dynamics can be a source of chaotic variability [1, 2, 3, 4, 5].

Excitation and inhibition in biological neural networks is usually conveyed by different types of neurons with a predominance of recurrent inhibitory feedback, a property known as 'inhibition dominance' [6, 7, 8, 9]. Networks with very strong recurrent inhibition (where synaptic interactions scale $\mathcal{O}\left(1/\sqrt{N}\right)$ with network size N), a dynamic balance arises between excitatory currents and recurrent inhibitory currents that was originally proposed as a robust mechanism to describe the emergence of asynchronous irregular activity[1].

Binary networks in such a *balanced state* track time-varying signals [2]. While subsequent experimental work found evidence for a dynamic balance of input currents, the 'tightness' or 'looseness' of such a dynamic balance in cortical circuits, as well as the computational implications are a question

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

of active scientific debate [10]. 'Loosely' balanced networks, where excitatory and inhibitory currents are $\mathcal{O}(1)$ compared to the distance from reset to threshold, can respond in their population firing rate nonlinearly to external inputs, which was argued to be favorable for sensory processing [7].'Tight balance' refers to a more precise tracking of total excitatory and total inhibitory input currents in time. Such 'tight balance' was also studied in a series of works that arrived at spiking balanced networks from a normative predictive coding ansatz [11, 12, 13, 14]. It is important to investigate how such biological features such as 'tightness of balance' shape the reliability of population response to external stimuli. How the dynamic tracking of time-dependent inputs described in binary networks [1, 2] extends to firing rate models and how it affects the information encoding of time-varying stimuli in the presence of chaos and noise has not yet been addressed. Previous dynamic mean-field theory (DMFT) approaches to input-driven rate networks assumed that the mean of the external inputs across neurons does not depend on time, which facilitates DMFT [15, 16, 17, 14], but does not permit to investigate encoding of dynamic stimuli in the population firing rate.

To address this gap, we study how a time-varying stimulus is encoded in the population rate of an inhibition-dominated rate network under the influence of additive chaos and noise.

We show that the accuracy of neuronal population encoding improves in more tightly balanced recurrent networks because of a speedup of the effective time scale of the population mean dynamics. Conventional methods of dynamic mean-field theory [18, 19, 20, 17] are not adequate to capture the effects of time-varying common input. Therefore, we developed a dynamic mean-field theory that is non-stationary, meaning that the order parameters are time-dependent (Materials and Methods). Beyond a similar recent approach [21], our theory can treat both time-varying common and independent external input, which is crucial to analytically treat a signal-to-noise ratio that constrains population coding. This novel technique accurately captures the time-dependent mean, variance and two-point autocorrelation of the input-driven networks. Specifically, we calculate the cross-spectral density between input and output and the power spectral density of the population firing rate. Together with the knowledge of the input statistics, this allows us to calculate the mutual information between stimulus and population rate in the Gaussian channel approximation based on the spectral coherence.

We examine how the frequency-response and mutual information rate depends on tightness of balance, added noise, network chaos and statistics of the input stimulus, using both theory and simulation. All the analytic results match those from network simulations. We show that recurrent networks that are trained on tracking a time-dependent stimulus develop strong recurrent inhibition and strong, positive input weights, a fingerprint of the balanced state. Concomitantly, the mutual information rate between their network readout and the stimulus increases as predicted by our theory. This indicates that a more tightly balanced state is a generic solution to reliable transmitting a time-varying input in the presence of noise or chaos. Lastly, we find that networks trained on simultaneously transmitting multiple independent time-varying stimuli spontaneously break up into weakly connected subnetworks with strong local inhibition.

Our findings have important implications for information encoding in firing rate networks and for understanding how neural network architecture design shapes noise-robustness and information encoding.

2 Population coding in recurrent networks

We study how well a time-varying scalar input $x(t)$ is encoded in the population firing rate $\phi(h(t))$ of a recurrent network of N nonlinear rate units ('neurons') that obey

$$\tau \frac{\mathrm{d}h_i}{\mathrm{d}t} = -h_i + \sum_{j=1}^{N} J_{ij}\phi\left(h_j\right) + bI(t) + \xi_i(t)\,, \tag{1}$$

with each entry of the coupling matrix $J_{ij} = -J_0/N + \tilde{J}_{ij}$ drawn from a Gaussian distribution with negative mean $-J_0/N$ and variance g^2/N, where g is a gain parameter that controls the heterogeneity of weights. The transfer function ϕ is set to a threshold-linear function $\phi(x) = \max(x, 0)$. The time-varying input signal $x(t)$ is decoded from the population firing rate $\phi(t)$ by a linear readout $\hat{x}(t) = 1/N \sum_i w_i^{\text{out}} \phi(h_i)$. The external input contains a signal component $I(t)$, which is identical across neurons and a noise term $\xi_i(t)$, which is independent across neurons. For concreteness, here we choose ξ_i to be independent additive white Gaussian noise processes (AWGN) with autocorrelation

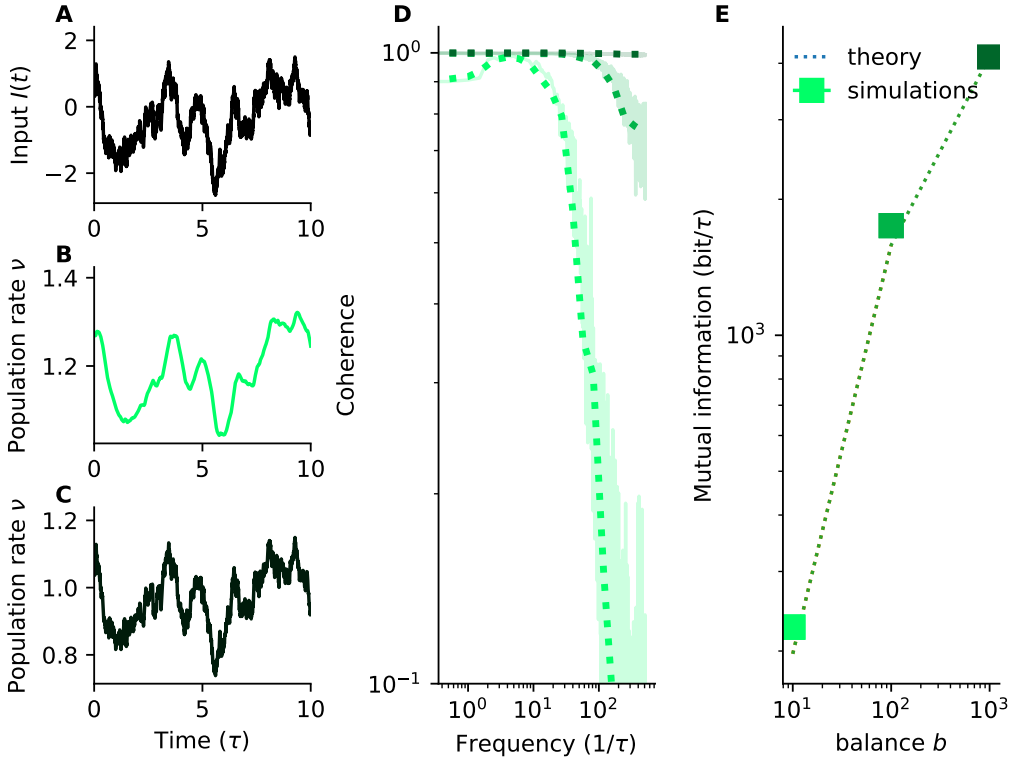

Figure 1: **Information rate between input and population response grows with depth of balance** **A)** Each neuron in the recurrent network receives an identical input stimulus I^{in} and additive Gaussian white noise $\xi(t)$ of strength σ that is independent across neurons. **B)** For small values of b ('loosely' balanced networks), the population response only tracks slow fluctuations of the stimulus. **C)** For large values of b ('tightly' balanced networks), the population response also tracks fast fluctuations of the stimulus. **D)** The dynamic gain for different values of balance b, direct numerical simulations (shaded line) and mean-field theory (dashed line) superimposed. (For color-code, see figure 1E). **E)** The Mutual information rate in Gaussian channel approximation based on spectral coherence, mean-field theory (dashed line) and direct numerical simulation (squares). Model parameters: $N = 4096$, $g = 2$, $\Delta t = 1/2^{10}/\tau$, $t_{\text{sim}} = 2^6\tau$, $I_0 = J_0 = 1$, $\tau_S = 1.0\tau$, $\sigma = 0.1$, $\sigma = 0.1$,$I_1 = 0.1$.

function $\langle \xi_i(t)\xi_i(t+t')\rangle = \tau\sigma_N^2\delta(t')$ and the signal $I(t)$ to be an Ornstein-Uhlenbeck process

$$\tau_S \frac{\mathrm{d}I}{\mathrm{d}t}(t) = -I(t) + I_0 + \xi_S(t) \tag{2}$$

The amplitude of the AWGN $\xi_S(t)$ is denoted by I_1. b is a parameter that multiplies both external input and recurrent coupling strength and thus regulates the tightness of balance. For firing-rate networks, the ability to track time-dependent input strongly depends on the strength of balance b (Fig 1), which determines how closely time-varying input is tracked by the population rate. To understand how this tracking arises in the model, it is useful to rewrite Eq 1 by decomposing $h_i = m + \tilde{h}_i$ into neuron-specific and neuron-nonspecific components and writing $J_{ij} = -J_0/b + \tilde{J}_{ij}$. This results in

$$\tau \frac{\mathrm{d}m}{\mathrm{d}t} = -m - bJ_0\nu(t) + bI(t)\,, \tag{3a}$$

$$\tau \frac{\mathrm{d}\tilde{h}_i}{\mathrm{d}t} = -\tilde{h}_i + \sum_j \tilde{J}_{ij}\phi\left(h_j\right) + \xi_i(t)\,. \tag{3b}$$

Here $\delta I(t)$ directly enters the expression for m, because it is identical across all neurons. It thus directly impacts the mean population rate $\nu(t) = \frac{1}{N}\sum_i \phi(h_i(t))$ and recruits, through the negative recurrent mean coupling $-J_0/b$, strong recurrent feedback $-bJ_0\nu(t)$ that is anticorrelated with the input and cancels most of the common external input $I(t)$. This cancellation can be seen by rewriting

Eq 5a as

$$\nu(t) = \frac{I(t)}{J_0} + \frac{1}{bJ_0}\left(-\tau\frac{\mathrm{d}m}{\mathrm{d}t} - m\right) . \tag{4}$$

This equation is commonly referred to as the 'balance equation' [2, 19, 20] in the absence of time-dependent input. On the other hand, the noise ξ_i affects only the residual fluctuations $\tilde{h}$.

As the balance parameter b is increased, the effective timescale of the dynamics of m becomes shorter by a factor of b, which leads to faster and more precise tracking. This can be seen by rewriting

$$\frac{\tau}{b}\frac{\mathrm{d}m}{\mathrm{d}t} = -m - J_0\nu(t) + I(t) , \tag{5a}$$

$$\tag{5b}$$

In contrast, the timescale of the equation for $\tilde{h}_i$ is unchanged by b. Of course, the two equations for m and $\mathrm{d}\tilde{h}_i$ are interdependent via $\phi(h_i) = \phi(m + \tilde{h}_i)$, so for a complete treatment, the joint dynamics of m and $\tilde{h}_i$ has to be solved.

3 Non-stationary dynamic mean-field theory of temporal population coding

The basic idea of DMFT is that for large N, the distribution of the recurrent input for different neurons becomes Gaussian and pairwise uncorrelated, according to the central limit theorem. To this end, we characterize the distribution of the $\tilde{h}_i(t)$ by considering the (linear) stochastic dynamics:

$$\tau\frac{\mathrm{d}\tilde{h}}{\mathrm{d}t} = -\tilde{h} + \eta(t) , \tag{6}$$

where $\eta(t)$ is a Gaussian process with mean $\langle\eta(t)\rangle = 0$ and autocorrelation

$$q(t,s) = \langle\eta(t)\eta(s)\rangle = g^2\big\langle\phi(m(t) + \tilde{h}(t))\phi(m(s) + \tilde{h}(s))\big\rangle . \tag{7}$$

Here and in the following, angular brackets denote expectation values over the distribution of the stochastic process $\tilde{h}(t)$, which approximates population averages in the full network. The mean-field estimate for the mean $m(t)$ therefore evolves according to Eq 5a with $\nu(t) = \langle\phi(m(t) + \tilde{h}(t))\rangle$, the mean-field estimate of the mean population firing rate.

We obtain an expression for the time evolution of the two-time autocorrelation function $c(t,s) = \left\langle\tilde{h}(t)\tilde{h}(s)\right\rangle$, which explicitly depends on two time points. Together, the dynamic mean-field equations for $m(t)$, $c(t,s)$ and an auxiliary function $r(t,s) = \left\langle\tilde{h}(t)\eta(s)\right\rangle$ form a closed system of self-consistent dynamic equations and can be solved (For details of the nonstationary dynamic mean-field theory, see appendix A). In contrast to previous work [21], we consider here a network where neurons simultaneously receive a time-varying input component $I(t)$ ("signal") that is identical across neurons and a input component $\xi_i s$ ("noise") that is independent across neurons. This setup is necessary to meaningfully quantify an information rate for a continuous channel, that requires a signal-to-noise ratio. We note that we go beyond a previously published nonstationary dynamic mean-field theory [21], by introducing in the time-dependent mean-field theory a heuristic term that emulates fluctuating input into the m-equation due to finite network size N similar to [14] (For details of the heuristic finite size approximation, see appendix B). We note, that in contrast to [14], we do solve the self-consistent set of dynamic mean-field equations for m, c and r to calculate dynamic gain, mutual information rate and and mutual information rate density.

4 Frequency-dependent signal encoding in population response

We next systematically analyze how different input frequency are encoded in the population response and observe that the transmission of high frequencies stimuli is improved for large b ('tightly balanced' regime) (Figure 2A and C). We calculate dynamic gain $G(f) = |S_{xy}|/|S_{xy}|$ of the recurrent networks both in direct numerical simulations and using the nonstationary DMFT. This quantifies how in linear response a variation in the input signal $I(t)$ affects the population rate $\nu(t)$ and is a standard analysis

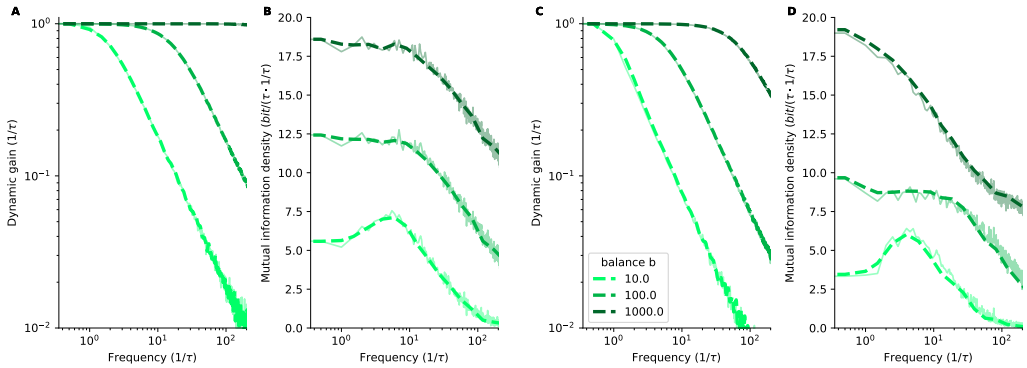

Figure 2: **Frequency-resolved information rate reveals different effects of chaos and white noise A)** The dynamic gain for different values of balance b, direct numerical simulations (shaded line) and mean-field theory (dashed line) superimposed for $g = 0$. Note that large values of b the dynamic gain shows an improved encoding bandwidth (For color-code, see figure 2C). **B)** The mutual information rate density in Gaussian channel approximation based on spectral coherence, mean-field theory (dashed line) and direct numerical simulation (transparent full lines) for $g = 0$. **C)** Same as **A)** but for $g = 2$. **D)** Same as **B)** but for $g = 2$. The recurrent residual fluctuation in $\tilde{h}_i$ arising from $g > 0$ reduce the information rate for low-frequency information. For larger values of b ('tight balance'), recurrent fluctuations have weaker effect on low-frequency information rate, because of tracking of network fluctuations. Model parameters: $N = 4096$,$g = 0$ for **A)** and **B)**. $g = 2$ for **C)** and **D)**, $\Delta t = 1/2^{10}/\tau$, $t_{\text{sim}} = 2^6\tau$, $I_0 = J_0 = 1$, $\tau_S = 1.0\tau$, $\sigma = 0.1$,$I_1 = 0.1$.

for characterizing the response properties of individual cells [22]. S_{xy} is the Fourier transformation of the input-output crosscorrelation function between $I(t)$ and $\nu(t)$. S_{xx} is the power spectral density of the OU-input.

We calculate $G(f)$ both using direct numerical network simulations and using dynamic mean field theory (For details of the spectral analysis, see appendix C). The Fourier transform of the autocorrelation of the input is the power spectral density of the input, according to the Wiener-Khinchin theorem. For an Ornstein-Uhlenbeck process, the power spectral density is $S_{xx}(f) = \frac{2\tau I_1^2}{1+(2\pi\tau f)^2}$

We observe that increasing b boosts the dynamic gain (Figure 2A and C for $g = 0$ and $g = 2$). This is a direct consequence of the speedup of the dynamics of m-dynamics as seen in Eq. 5. Thus, as the network becomes more tightly balanced, high-frequency signals can be encoded and transmitted more reliable. To quantify that improved information encoding in terms of bit per τ we turn to a information theoretic analysis in the next section.

5 Mutual information rate between time-varying stimulus and population response

With information theory we can treat the recurrent neural network as a noisy communication channel transforming a signal embedded in a noisy current into a population rate that can be readout by another population. The mutual information rate measures how much the uncertainty about the input $I(t)$ is reduced given the population rate $\nu(t)$ per unit time τ:

$$R(X,Y) = h(Y) - h(Y|X) = h(X) - h(X|Y) = \lim_{T\to\infty} \frac{1}{T} \int_X \int_Y p(x,y) \log_2 \left(\frac{p(x,y)}{p(x)p(y)} \right) \quad (8)$$

where $p(x,y)$ is the joint probability density function of $x(t)$ and $y(t)$, $h(X)$ and $h(Y)$ are their entropy rates and $h(X|Y)$ is the conditional entropy rate of $x(t)$ given $y(t)$. The continuous Gaussian channel gives a lower bound on the mutual information rate for a Gaussian input signal: [23, 24]:

$$R(X,Y) = h(X) - h(X|Y) \geq h(X) - h_{\text{Gaussian}}(X|Y) = R_{\text{lb}}(X,Y) \quad (9)$$

The inequality results from the property that a Gaussian process has the maximum entropy of all processes with fixed variance. Recently it was shown for spiking neurons that for moderate input

modulation in a fluctuation driven regime this lower bound is very close to the mutual information rate estimated from direct methods [25]. . This is convenient, because estimating the mutual information from empirical data is computationally costly, as the sample size has to be much larger than the size of the alphabet [26]. . Continuous processes usually have to be discretized resulting in very large alphabets. In our case a Gaussian channel approximation to the mutual information rate between input current and output population rate was estimated based on the spectral coherence between input current and output spike train, which is based purely on second-order statistics, which we can obtain through the non-stationary dynamic mean-field theory

$$R_{\mathrm{lb}}(X,Y) = -\int_0^{f_{\mathrm{cutoff}}} df \log_2\left(1 - C_{xy}(f)\right) \tag{10}$$

$C_{xy}(f)$ denotes the magnitude squared spectral coherence. The spectral coherence is the frequency-domain analog of correlation and measures the linear relationship between frequency components of input and output signal. Its magnitude square is

$$C_{xy}(f) = \frac{|S_{xy}(f)|^2}{|S_{xx}(f)||S_{yy}(f)|}. \tag{11}$$

It consists of $S_{xx}(f)$,which is the power spectrum of the input signal (a OU process), $S_{yy}(f)$ is the power spectrum of the population rate $\nu(t)$. $S_{yy}(f) = \lim_{T\to\infty} \frac{1}{T} \langle \tilde{y}\tilde{y}^* \rangle$, where $\tilde{y}(f)$ is the Fourier transform of the population firing rate ν. If and only if $x(t)$ and $y(t)$ are linearly scaled copies of each other, then $C_{xy}(f)$=1. If $x(t)$ and $y(t)$ are independent, then $C_{xy}(f)$=0, the reverse generally does not hold. Nonlinear effects and noise reduce the coherence. We calculated the spectral coherence in two independent ways: Firstly based on the nonstationary DMFT and secondly based on direct numerical simulations. The dependence of $C_{xy}(f)$ on the depth of balance b are depicted in Figure 1D. DMFT and direct numerical simulations with additive band-limited white noise $\xi(t)$ are in excellent agreement.

The mutual information rate scales approximately linear with b (See figure 1E). For large b, the mutual information rate saturates. The saturation level is determined by the band limit of the incoming signal (See figure in appendix C). We use the analytical high-frequency behavior and low-frequency limit to give an analytical approximation of the linear scaling of the mutual information rate (See appendix C). From the low-frequency limit and the high-frequency limit of the coherence, we can get an analytical estimate of the spectral coherence. The high-frequency behavior strongly depends on the depth of balance. We analytically approximate the transition points from the low-frequency response to the high-frequency response and stitched analytic high- and low-frequency estimates together at the transition points. Plugging into Eq. 10 results in a linear scaling of the mutual information rate with b (see appendix D for derivation).

We conclude that the mutual information rate scales approximately linearly with tightness of balance b for sufficiently high band limit. Where does the linear scaling of the mutual information rate with tightness of balance arise from? The reason is that the b determines the cutoff frequency, up to which the network transmits information faithfully. Contributions from frequencies beyond this cutoff are negligible to first order. This conclusion is not restricted to networks of rectified linear units, a similar line of argumentation holds also e.g. for the threshold-powerlaw and other models with non-negative firing rates (See e.g. [19, 20] for balanced rate networks with other input-output transfer functions).

6 Frequency-resolved mutual information rate analysis

Inspired by a recent work on the frequency-resolved information encoding rate in single cells [25], we analyze for the population rate of recurrent neural networks how how the encoding of different frequency components of the input signal depend on the tightness of balance b, the strength of added weight noise σ and the heterogeneity of the recurrent weights g.

For that, we considered the mutual information rate density $r(f)$, which gives a frequency-resolved quantification of the information encoding rate (See for details of definition [25]). $r_{\mathrm{lb}}(X,Y) = -\log_2\left(1 - C_{xy}(f)\right)$ Note that the integral over frequencies of $r_{\mathrm{lb}}(X,Y)$ gives the Gaussian channel approximation of the mutual information rate, we thus call it mutual information rate density and remark that it comes in units of $\frac{bit/\tau}{1/\tau}$.

We observe that more tightly balanced networks can transmit more high-frequency information (Figure 2B and D for $g = 0$ and $g = 2$). Overall, the information rate for more tightly balanced networks is increased because of the boost of the signal to noise ratio coming from the fact that the signal $I(t)$ is grows by b, while the noise is independent of b. The above described frequency dependent effect comes on top of that.

For recurrent networks with weight heterogeneity ($g > 0$), we observe a that the residual slow fluctuations $\tilde{h}$ become stronger which leads to dip in the mutual information density for low frequencies (Compare Figure 2B for $g = 0$ and D for $g = 2$).

This comes from the fact that $g > 0$ leads do additional recurrent low-frequency fluctuations that act as an additional source of noise. But because these fluctuations are low-frequency, they have a different effect than additive Gaussian white noise (See also [14]).

For more tightly balanced networks (larger b), the residual fluctuations do not impair information encoding (see darker green lines in figure 2B), because they are canceled by fast tracking. A related phenomenon was observed in densely connected recurrent binary networks ([27, 28]). That variance originating from weight heterogeneity ($g > 0$) is more effectively reduced by balanced networks as function of b compared to variance originating from added Gaussian white noise σ was already observed in rate networks with static input [14].

7 Training networks on auto-encoding results in strong inhibition

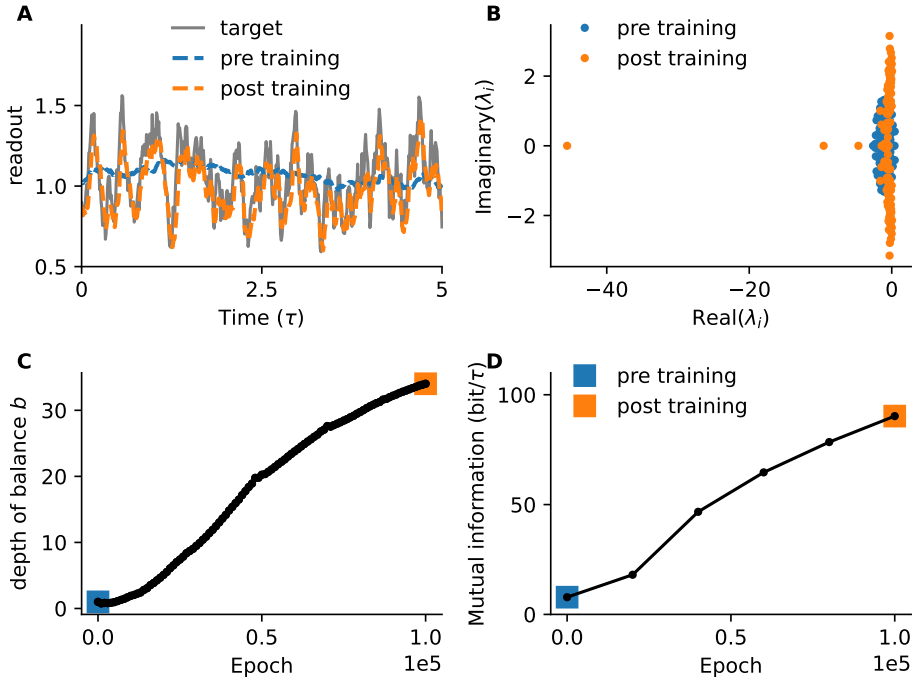

Figure 3: **RNNs trained on tracking time-varying input become more tightly balanced throughout training. A)** vanilla RNNs are trained to approximate a time-varying external input $I(t)$ by linear readout $\hat{x}(t) = 1/N \sum_i w_i^{\text{out}} \phi(h_i)$, by minimizing the mean squared loss $l = \int_t |y(t) - I(t)|^2 \, \mathrm{d}t$. **B)** The eigenvalue spectrum of the dynamics linearized at the fixed point indicates that during training, an eigenmode with strongly negative real part emerges. **C)** The networks become more balanced throughout training. The depth of balanced is quantified here by the magnitude of the mean recurrent coupling $b = |1/N \sum_i \sum_j J_{ij}|$. **D)** The mutual information between the time-varying external input and the linear readout as function of training epochs. Information rate increase throughout training. Model parameters at initialization: $N = 100$, $g = \sqrt{2}$, $\Delta t = 0.01/\tau$, $t_{\text{sim}} = 1000\tau$, $b = 1$, $I_0 = J_0 = 1$, $\tau_S = 0.1$, $\sigma = 1$, epochs$= 10^5$.

We corroborated the finding that the tightness of balance has a crucial role in information encoding in trained recurrent networks which demonstrates the generality of our findings. For that, we train recurrent networks on an auto-encoder task. We initialize small-sized recurrent networks ($N = 100$) that follow the dynamics of $Eq.\ 1$ in a loosely balanced state ($b = 1$), e.i., with Gaussian connectivity J_{ij} that has a negative mean of b/N and variance g^2/N with $g = sqrt2$, input weights w^{in} and output weights w^{out} are drawn from a Gaussian distribution with a positive mean. The network receives the time-varying input signal $I_i(t) = w_i^{\text{in}} I(t)$, so in contrast to our theoretical setup, in principle each neuron can receive the input signal $I(t)$ with a different scale w_i^{in}. We then trained input weights w^{in}, output weights w^{out} and recurrent weights J_{ij} using backpropagation-through-time with the ADAM optimizer with standard hyper parameters (see appendix G for additional details on training setup and additional controls of training results). We minimize the mean squared loss mean squared loss $l = \int_t |y(t) - I(t)|^2 \, \mathrm{d}t$ between time-varying input signal $I(t)$ and a linear readout $\hat{x}(t) = 1/N \sum_i w_i^{\text{out}} \phi(h_i)$.

Initially, the network does not track well the time-varying input signal as expected from our theoretical results. Over training epochs, we however find that the network learns to track the time-varying input (Figure 3A. Analyzing the eigenvalues of the trained network, we observe an outlier eigenvalue with very negative real part after training that indicates the emerging fast tracking (Figure 3B. This is confirmed by measuring the empirical level of balance b in the connectivity, which we quantify just by the absolute value of the mean of the sum of incoming recurrent weights into each neuron $b = |1/N \sum_i \sum_j J_{ij}|$. We find that the mean becomes progressively more balanced (Figure 3C over training epochs. We prematurely stop after 100000 training epochs to avoid numerical instabilities for very tightly balanced networks.

Consistent with our theoretical results, we find that the mutual information rate quantified in the Gaussian channel approximations grows over training epochs. Note that we did not train the network to maximize the mutual information rate, but to minimize the mean-squared error between target output and actual output. We note that this tracking does not rely on the threshold-linear input-output function, in fact we also successfully trained threshold-quadratic and sigmoidal input-output function ϕ. We note that training with a linear input-output function $\phi(x) = x$ fails, because the boost of high-frequency input through the tracking observed in tightly-balanced networks requires that the firing rate units have non-negative firing rates (see appendix G for additional controls of training results, e.i. training on sigmoid, threshold quadratic, and the role of additive noise and regularization of weights and firing rates).

We note that networks trained to simultaneously track multiple input Ornstein-Uhlenbeck signals by a linear readout exhibit a spontaneous symmetry breaking weakly connected subnetworks with strong local inhibition (see appendix H for results on training on multiple signals).

8 Limitations

For mathematical tractability, we considered firing rate networks here, but for more detailed neuron models, other biophysically features shape the information encoding rate. For example, in spiking neuron models, fluctuating background input can enhance the high-frequency encoding [29]. Moreover, the spike generation mechanism [30, 31, 32, 33, 34, 35],the synaptic dynamics and latent variables (e.g. adaptation [36]) and the synaptic timescale ([37, 38]) shape the frequency-response. It remains an important future work to extend the work here by such contributions. Moreover, moving beyond point-neurons, the shape of the dendritic tree can also affect the frequency response[35].

While we demonstrated that recurrent networks trained on a simple copy task can be explained by our non-stationary dynamic mean-field theory, it remains an important challenge to consider more complex tasks, e.g. involving memory, where the task dynamic might interfere with the balance.

Finally, our theory only rigorously describes the behavior of very large networks, and it is important to also consider finite-size effects. For finite network size, chaotic fluctuations also contribute to a fluctuation of the mean, which are also recurrently canceled. This was described previously in recurrent networks in a linear regime.

9 Discussion

We show how tight balance of excitatory currents by recurrent inhibition improves information encoding of a time-varying signal. We demonstrate that the mutual information rate of a time-varying signal increases linearly with the tightness of balance, both in the presence of additive noise and with recurrent chaotic network fluctuations. A non-stationary dynamic mean-field theory reveals a separation of time-scale between the mean currents which become linearly faster with tightness of balanced and enable reliable encoding of time-varying signals. In contrast, the time-scale of the chaotic dynamics in the residual fluctuations are largely unaffected by the tightness of balance. We find that networks become more robust to deteriorating effects of fluctuations from noise and chaos as the network becomes more tightly balanced.

Our study is relevant in the recent debate on the functional implications of how tightly excitatory currents are tracked by recurrent inhibition [27, 28, 7, 10, 13, 14].

Our work addresses this question by building a bridge from information theoretic measures of information encoding that were previously used in neuroscience mostly in sensory systems [39, 40, 41, 24, 23] to dynamic mean-field descriptions of recurrent networks dynamics that were previously used to describe the often chaotic dynamics of recurrent rate networks [18, 19, 20, 16, 42, 17, 43, 14].

Besides the implications on information encoding of more loosely or more tightly balanced networks, of course also biophysical, energetic and evolutionary constrains should be considered. Biological networks can for biophysical reasons not be arbitrarily tight balanced, which would come with arbitrary large synaptic currents. Too tight balance might also be questionable for energetic reasons, as was asked previously "Why should the cortex simultaneously push on the accelerator and on the brake?" [44]. However, such energetic and biophysical constrains might be better addressed in biophysical more detailed models.

We found that training networks on tracking a time-dependent signal by a linear readout by minimizing the squared error make them more tightly balanced, as reflected in more negative mean recurrent weights and more positive feedforward input weights. Moreover, the training also increased the mutual information rate between input signal and linear readout. This is consistent with our theoretical result. Moreover, the fact that training arrives at a tightly balanced solution seems to suggest that this is a typical solution for a network. (See for effect of different regularization of activity and weights and effect of noise appendix G)

Previous studies on temporal information encoding in rate networks were limited to independent inputs across neurons in the form of stochastic [15, 17] and sinusoidal [16] drive, but the networks were not balanced, and their connectivity had zero mean coupling. In these previous studies, the distribution of inputs across the population is time-independent [15, 16, 17] and stationary dynamic mean-field theory was sufficient to describe the results. However, the treatment of common input is only possible by the non-stationary dynamic mean-field approach introduced here.

The dynamic cancellation of time-varying input through recurrent inhibitory feedback has been previously studied in balanced networks with binary [2, 27, 28], and spiking neurons [45, 46]. Chaos in balanced firing-rate networks was studied previously [19, 20, 47, 17], but the dynamic cancellation of correlated input and its implications on information encoding in rate networks were not investigated, nor were the implications for training RNNs in a machine learning setup with backpropagation through time. It would be interesting to extend our mean-field analysis to rate networks with pre-existing low-rank structure on top of the random structure [48, 49, 50, 51].

The underlying mechanisms of tracking of time-varying input we analyze here are not specific to fully-connected threshold-linear RNNs driven by Ornstein-Uhlenbeck signals with additive Gaussian white noise, which we merely chose for the sake of simplicity and analytical tractability. Rate networks with other input-output transfer function that are constrained to be non-negative, exhibit a qualitatively similar dynamic tracking in the tightly balanced regime. Moreover, the mechanisms of tracking described here is tightly related to coding balanced networks [13, 14].

References

[1] C. van Vreeswijk and H. Sompolinsky. Chaos in Neuronal Networks with Balanced Excitatory and Inhibitory Activity. *Science*, 274(5293):1724 –1726, December 1996.

[2] C. van Vreeswijk and H. Sompolinsky. Chaotic Balanced State in a Model of Cortical Circuits. *Neural Computation*, 10(6):1321–1371, 1998.

[3] Michael Monteforte and Fred Wolf. Dynamical Entropy Production in Spiking Neuron Networks in the Balanced State. *Physical Review Letters*, 105(26):268104, December 2010.

[4] Guillaume Lajoie, Kevin K. Lin, and Eric Shea-Brown. Chaos and reliability in balanced spiking networks with temporal drive. *Physical Review E*, 87(5):052901, May 2013.

[5] Rainer Engelken, Fred Wolf, and L. F. Abbott. Lyapunov spectra of chaotic recurrent neural networks. *arXiv:2006.02427 [nlin, q-bio]*, June 2020. arXiv: 2006.02427.

[6] Hirofumi Ozeki, Ian M. Finn, Evan S. Schaffer, Kenneth D. Miller, and David Ferster. Inhibitory Stabilization of the Cortical Network Underlies Visual Surround Suppression. *Neuron*, 62(4):578–592, May 2009.

[7] Yashar Ahmadian, Daniel B. Rubin, and Kenneth D. Miller. Analysis of the stabilized supralinear network. *Neural computation*, 25(8):1994–2037, August 2013.

[8] Fred Wolf, Rainer Engelken, Maximilian Puelma-Touzel, Juan Daniel Flórez Weidinger, and Andreas Neef. Dynamical models of cortical circuits. *Current Opinion in Neurobiology*, 25:228–236, April 2014.

[9] Alessandro Sanzeni, Bradley Akitake, Hannah C Goldbach, Caitlin E Leedy, Nicolas Brunel, and Mark H Histed. Inhibition stabilization is a widespread property of cortical networks. *eLife*, 9:e54875, June 2020.

[10] Yashar Ahmadian and Kenneth D. Miller. What is the dynamical regime of cerebral cortex? *Neuron*, 109(21):3373–3391, November 2021.

[11] David GT Barrett, Sophie Denève, and Christian K Machens. Optimal compensation for neuron loss. *eLife*, 5.

[12] Martin Boerlin, Christian K. Machens, and Sophie Denève. Predictive Coding of Dynamical Variables in Balanced Spiking Networks. *PLOS Comput Biol*, 9(11):e1003258, November 2013.

[13] Sophie Denève and Christian K. Machens. Efficient codes and balanced networks. *Nature Neuroscience*, 19(3):375–382, March 2016.

[14] J. Kadmon, J. Timcheck, and S. Ganguli. Predictive coding in balanced neural networks with noise, chaos and delays. *Advances in neural information processing systems*, 33, December 2020.

[15] L. Molgedey, J. Schuchhardt, and H. G. Schuster. Suppressing chaos in neural networks by noise. *Physical Review Letters*, 69(26):3717–3719, December 1992.

[16] Kanaka Rajan, L. F. Abbott, and Haim Sompolinsky. Stimulus-dependent suppression of chaos in recurrent neural networks. *Physical Review E*, 82(1):011903, July 2010.

[17] Jannis Schuecker, Sven Goedeke, and Moritz Helias. Optimal Sequence Memory in Driven Random Networks. *Physical Review X*, 8(4):041029, November 2018.

[18] H. Sompolinsky, A. Crisanti, and H. J. Sommers. Chaos in Random Neural Networks. *Physical Review Letters*, 61(3):259–262, July 1988.

[19] Jonathan Kadmon and Haim Sompolinsky. Transition to Chaos in Random Neuronal Networks. *Physical Review X*, 5(4):041030, November 2015.

[20] Omri Harish and David Hansel. Asynchronous Rate Chaos in Spiking Neuronal Circuits. *PLoS Comput Biol*, 11(7):e1004266, July 2015.

[21] Rainer Engelken, Alessandro Ingrosso, Ramin Khajeh, Sven Goedeke, and L. F. Abbott. Input correlations impede suppression of chaos and learning in balanced rate networks. *arXiv:2201.09916 [cond-mat, physics:nlin, q-bio]*, January 2022. arXiv: 2201.09916.

[22] Matthew H. Higgs and William J. Spain. Conditional Bursting Enhances Resonant Firing in Neocortical Layer 2–3 Pyramidal Neurons. *The Journal of Neuroscience*, 29(5):1285–1299, February 2009.

[23] W. Bialek, F. Rieke, RR de Ruyter van Steveninck, and D. Warland. Reading a neural code. *Science*, 252(5014):1854–1857, June 1991.

[24] Fred Rieke, David Warland, Rob de Ruyter van Steveninck, and William Bialek. *Spikes: Exploring the Neural Code*. A Bradford Book, reprint edition edition, June 1999.

[25] Davide Bernardi and Benjamin Lindner. A frequency-resolved mutual information rate and its application to neural systems. *Journal of Neurophysiology*, 113(5):1342–1357, March 2015.

[26] S. P. Strong, Roland Koberle, Rob R. de Ruyter van Steveninck, and William Bialek. Entropy and Information in Neural Spike Trains. *Physical Review Letters*, 80(1):197–200, January 1998.

[27] Alfonso Renart, Jaime de la Rocha, Peter Bartho, Liad Hollender, Néstor Parga, Alex Reyes, and Kenneth D. Harris. The Asynchronous State in Cortical Circuits. *Science*, 327(5965):587–590, January 2010.

[28] Tom Tetzlaff, Moritz Helias, Gaute T. Einevoll, and Markus Diesmann. Decorrelation of Neural-Network Activity by Inhibitory Feedback. *PLOS Comput Biol*, 8(8):e1002596, August 2012.

[29] Bruce W. Knight. Dynamics of Encoding in a Population of Neurons. *The Journal of General Physiology*, 59(6):734–766, June 1972.

[30] Nicolas Fourcaud-Trocmé, David Hansel, Carl van Vreeswijk, and Nicolas Brunel. How Spike Generation Mechanisms Determine the Neuronal Response to Fluctuating Inputs. *The Journal of Neuroscience*, 23(37):11628–11640, December 2003.

[31] B. Naundorf, T. Geisel, and F. Wolf. Action Potential Onset Dynamics and the Response Speed of Neuronal Populations. *Journal of Computational Neuroscience*, 18(3):297–309, June 2005.

[32] Wei Wei and Fred Wolf. Spike Onset Dynamics and Response Speed in Neuronal Populations. *Physical Review Letters*, 106(8):088102, February 2011.

[33] Tatjana Tchumatchenko and Fred Wolf. Representation of Dynamical Stimuli in Populations of Threshold Neurons. *PLoS Comput Biol*, 7(10):e1002239, October 2011.

[34] Elinor Lazarov, Melanie Dannemeyer, Barbara Feulner, Jörg Enderlein, Michael J. Gutnick, Fred Wolf, and Andreas Neef. An axon initial segment is required for temporal precision in action potential encoding by neuronal populations. *Science Advances*, 4(11):eaau8621, November 2018.

[35] Chenfei Zhang, David Hofmann, Andreas Neef, and Fred Wolf. Ultrafast population coding and axo-somatic compartmentalization. *PLOS Computational Biology*, 18(1):e1009775, January 2022.

[36] Maximilian Puelma Touzel and Fred Wolf. Complete Firing-Rate Response of Neurons with Complex Intrinsic Dynamics. *PLOS Computational Biology*, 11(12):e1004636, December 2015.

[37] Nicolas Brunel, Frances S. Chance, Nicolas Fourcaud, and L. F. Abbott. Effects of Synaptic Noise and Filtering on the Frequency Response of Spiking Neurons. *Physical Review Letters*, 86(10):2186–2189, March 2001.

[38] Tatjana Tchumatchenko, Aleksey Malyshev, Fred Wolf, and Maxim Volgushev. Ultrafast Population Encoding by Cortical Neurons. *The Journal of Neuroscience*, 31(34):12171 –12179, 2011.

[39] J. H van Hateren and H. P Snippe. Information theoretical evaluation of parametric models of gain control in blowfly photoreceptor cells. *Vision Research*, 41(14):1851–1865, June 2001.

[40] J. H. van Hateren. Processing of natural time series of intensities by the visual system of the blowfly. *Vision Research*, 37(23):3407–3416, December 1997.

[41] J. H. van Hateren, L. Rüttiger, H. Sun, and B. B. Lee. Processing of Natural Temporal Stimuli by Macaque Retinal Ganglion Cells. *The Journal of Neuroscience*, 22(22):9945–9960, November 2002.

[42] M. Stern, H. Sompolinsky, and L. F. Abbott. Dynamics of random neural networks with bistable units. *Physical Review E*, 90(6):062710, December 2014.

[43] Samuel P. Muscinelli, Wulfram Gerstner, and Tilo Schwalger. How single neuron properties shape chaotic dynamics and signal transmission in random neural networks. *PLOS Computational Biology*, 15(6):e1007122, June 2019.

[44] Jeffry S. Isaacson and Massimo Scanziani. How Inhibition Shapes Cortical Activity. *Neuron*, 72(2):231–243, October 2011.

[45] Ran Darshan, William E. Wood, Susan Peters, Arthur Leblois, and David Hansel. A canonical neural mechanism for behavioral variability. *Nature Communications*, 8:15415, May 2017.

[46] Robert Rosenbaum, Matthew A. Smith, Adam Kohn, Jonathan E. Rubin, and Brent Doiron. The spatial structure of correlated neuronal variability. *Nature neuroscience*, 20(1):107–114, January 2017.

[47] Francesca Mastrogiuseppe and Srdjan Ostojic. Intrinsically-generated fluctuating activity in excitatory-inhibitory networks. *PLOS Computational Biology*, 13(4):e1005498, April 2017.

[48] Johnatan Aljadeff, Merav Stern, and Tatyana Sharpee. Transition to Chaos in Random Networks with Cell-Type-Specific Connectivity. *Physical Review Letters*, 114(8):088101, February 2015.

[49] Johnatan Aljadeff, David Renfrew, Marina Vegué, and Tatyana O. Sharpee. Low-dimensional dynamics of structured random networks. *Physical Review E*, 93(2):022302, February 2016.

[50] Francesca Mastrogiuseppe and Srdjan Ostojic. Linking Connectivity, Dynamics, and Computations in Low-Rank Recurrent Neural Networks. *Neuron*, 99(3):609–623.e29, August 2018.

[51] Itamar D. Landau and Haim Sompolinsky. Macroscopic fluctuations emerge in balanced networks with incomplete recurrent alignment. *Physical Review Research*, 3(2):023171, June 2021.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] See Section 8
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] , not applicable
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes] , see result section and appendix.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] See supplement
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See training section and supplement
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] Error bars for figure 1E are smaller than symbol size, figures on mutual information density and gain included comparison to theory as check of error
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] See appendix C
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] Does not apply
   (b) Did you mention the license of the assets? [Yes] Does not apply
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] Does not apply
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [Yes] Does not apply
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes] Does not apply
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [Yes] Does not apply
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [Yes] Does not apply
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [Yes] Does not apply

A Nonstationary dynamic mean-field theory with both independent and common input

B Heuristic finite size correction

C Details on numerical and analytical spectral analysis

D Details on analytical approximation of mutual information rate

E Finite-size correction for spectral coherence quantification

F Mixed excitatory-inhibitory circuits

G Details ond training setup and additional numerical training experiments

H Details on training networks to track multiple time-varying signals