nature neuroscience

Efficient codes and balanced networks

Sophie Denève[1] & Christian K Machens[2]

Recent years have seen a growing interest in inhibitory interneurons and their circuits. A striking property of cortical inhibition is how tightly it balances excitation. Inhibitory currents not only match excitatory currents on average, but track them on a millisecond time scale, whether they are caused by external stimuli or spontaneous fluctuations. We review, together with experimental evidence, recent theoretical approaches that investigate the advantages of such tight balance for coding and computation. These studies suggest a possible revision of the dominant view that neurons represent information with firing rates corrupted by Poisson noise. Instead, tight excitatory/inhibitory balance may be a signature of a highly cooperative code, orders of magnitude more precise than a Poisson rate code. Moreover, tight balance may provide a template that allows cortical neurons to construct high-dimensional population codes and learn complex functions of their inputs.

Since the first neural recordings by Adrian in 1926 (ref. 1), it has become accepted wisdom that neurons communicate information with their firing rates—that is, the number of spikes within a certain time window or across a population of neurons. Indeed, the information encoded by neurons is still usually extracted by studying their trial-averaged, time-varying firing rates. However, when neural recordings in cortical and other brain areas found that the spike trains of individual neurons are highly irregular, almost random, the firing rate hypothesis ran into two serious conundrums (**Fig. 1**).

First, firing spikes at random times, as in a Poisson point process, seems a particularly foolish idea when the goal is to convey information in spike counts (**Fig. 1a,b**). When discrete spike counts are used to represent continuous numbers, then the achievable precision is limited solely by the unavoidable discretization. For a given number of spikes M, the minimum error therefore scales with $1/M$. However, because of its unreliability, the error for a Poisson rate code scales with $1/\sqrt{M}$. In turn, neurons using a Poisson code need to fire a huge excess of spikes to reach a given level of precision (**Fig. 1c**). Consequently, the neural code chosen by the brain—firing rates to be inferred from unreliable spike trains—seems incongruous. Should evolution not have stumbled on a 'better' design?

Second, generating irregular spike trains within a network of recurrently connected neurons turns out to be a nontrivial problem. When random spike trains are integrated on a dendritic tree, then the recipient neuron will generally produce a regular output spike train[2]. So how can neurons ever produce irregular, Poisson-like spike trains?

Many solutions to the first problem, the 'coding problem', have been proposed[3–7], but mostly without specifying how the respective codes can be generated in recurrent neural networks. A simple solution to the second problem, the 'implementation problem', is to assume that the excitatory and inhibitory inputs to each neuron are balanced[8–12], a theoretical proposal that has been largely corroborated by various experimental observations[13–15]. However, while the theory of balanced networks solves the problem of how to generate networks that produce irregular spike trains and account for Poisson rate codes, it sidesteps the question of why neural systems would represent information so inefficiently. Indeed, implementing even simple functions in these networks requires thousands of neurons.

Here we briefly review the literature on balanced networks and then focus on several recent, theoretical studies that seek to reconcile the apparent randomness of spike trains with an efficient population code—that is, a code that scales with $1/M$ (refs. 16–21). These studies are based on networks in which the balance between excitatory and inhibitory inputs is temporally much tighter than in the original, 'loosely' balanced networks. We discuss the various theoretical benefits of these networks, which, besides the higher coding efficiency, include separate recurrent loops for coding and computation, and the ability to simultaneously represent almost as many variables as there are neurons in the network. We furthermore review several experimental studies that lend support to the notion of 'tight' balance. This recent body of work suggests that the irregularity and unreliability of spike trains at the single-neuron level coexist with a maximally efficient code at the population level.

Loosely balanced networks

We illustrate the relationship between the balance of excitatory and inhibitory input currents (E/I balance) and the variability of the output spike train in **Figure 1d,e**. Here an integrate-and-fire neuron is bombarded with noisy, Poisson-distributed spike trains from both excitatory and inhibitory sources. If excitation dominates, temporal averaging of the total input currents results in a mean positive drift of the membrane potential toward threshold, causing a relatively regular output spike train despite the high level of input noise (**Fig. 1d**). However, if excitation and inhibition cancel each other on a slower time scale yet are uncorrelated on a faster time scale, then the net input current will be dominated by these faster fluctuations, and the membrane potential will follow an uncorrelated random walk toward threshold, resulting in an output spike train with Poisson statistics (**Fig. 1e**). We will refer to this type of balance as 'loose' balance,

[1]Laboratoire de Neurosciences Cognitives, École Normale Supérieure, Paris, France. [2]Champalimaud Centre for the Unknown, Lisbon, Portugal. Correspondence should be addressed to S.D. (sophie.deneve@ens.fr) or C.K.M. (christian.machens@neuro.fchampalimaud.org).

Received 3 November 2015; accepted 13 January 2016; published online 23 February 2016; doi:10.1038/nn.4243

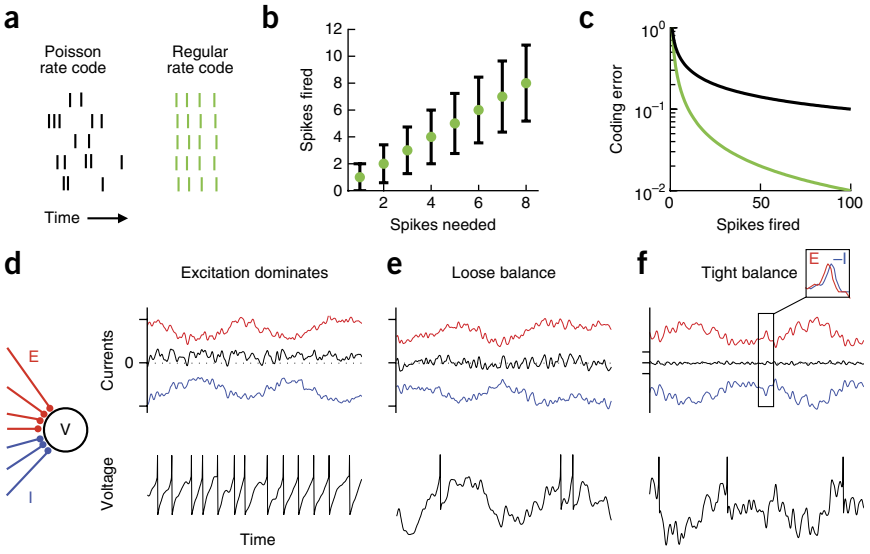

Figure 1 The conundrum of Poisson rate codes and E/I balance. (**a**) When seeking to produce an average rate of 4 spikes per second, the number of spikes in each trial will fluctuate for a Poisson rate code (left), yet remain constant for a perfect, regular rate code (right). (**b**) As the average rate or number of spikes per second needed grows, the variability of the Poisson rate code (s.d.; black bars) grows as well. (**c**) When encoding an analog number, any rate code will cause coding errors due to the unavoidable discretization of the spike count. While these coding errors drop as $1/M$ for M spikes (green), the Poisson rate coding errors drop only as $1/\sqrt{M}$ owing to the additional problems caused by unreliability (black). (**d**) In a neuron receiving both excitatory and inhibitory currents, even a small imbalance toward excitation will cause a spike train output that is far more regular than a Poisson spike train. (**e**) If excitatory and inhibitory currents are balanced on a slow time scale yet exhibit fast and uncorrelated (loose) fluctuations, then the net input current will depend on these faster fluctuations only, causing a random walk to threshold in the membrane potentials. (**f**) If the faster fluctuations are strongly correlated but shifted in time—for example, by inhibition trailing excitation (see inset)—then the system moves into a regime of tight E/I balance.

since the faster fluctuations in the excitatory and inhibitory inputs are uncorrelated.

The proper balancing of input currents into each neuron might seem to require some type of synaptic fine-tuning. However, on the network level, such a balance can arise dynamically if two conditions are met. First, connections must be sparse and random, such that two neurons receive essentially uncorrelated inputs from the network. In this case, the average number of connections per neuron, K, is much smaller than the number of neurons in the network, N; that is, $K \ll N$. Second, inhibitory connections must be larger than excitatory connections[9,12,22]. In the balanced regime, the excitatory and inhibitory input currents into any given cell are so large that, were the inhibitory inputs to shut down, the cell would immediately be driven into saturation. However, the two currents balance on average, so that any mean increase in excitation recruits proportionally more mean inhibition, while the fast fluctuations in E and I currents remain uncorrelated. As a result, the neuron's membrane potentials undergo a random walk, the membrane potentials of different neurons are uncorrelated, and their spike trains are irregular and uncorrelated. On the network level, these irregular, asynchronous spiking dynamics are possible because the neurons' activity evolves on a chaotic[12] or quasi-chaotic[23] attractor.

One consequence of these chaotic dynamics is that the timing of individual spikes becomes strongly susceptible to even small amounts of noise[11]. Indeed, even tiny perturbations can lead to a complete reshuffling of future spikes. Since spike timing is not a reproducible property of these networks, it becomes impractical as a coding strategy[24]. Fortunately, the average firing rates of the neurons do carry information. Since the strength of the input fluctuations is coupled to the mean input signal, the output firing rate of a neuron grows linearly with the strength of the excitatory and inhibitory inputs[9,11]. Although the spike times are chaotic, the time-averaged firing rates of individual neurons remain relatively constant. Furthermore, chaotic balanced networks can precisely track any input signals, and the tracked signals can be read out by averaging spikes over the whole population[9]. This theoretical work therefore strongly supports the Poisson rate coding hypothesis of cortical populations: that is, that information about sensory and motor variables is represented in the firing rates of large neural populations and that spikes are simply random samples from the underlying rates[24].

Experimental evidence for E/I balance

The suggestion that E/I balance underlies the irregularity of cortical spike trains inspired researchers to test whether this balance exists *in vivo*. While the initial theories of loosely balanced networks ignored certain aspects of cortical architecture (for example, that connectivity is not random), they still made several key predictions that stood out.

Probably the first and foremost prediction is that inhibitory and excitatory current inputs should be balanced on a trial-by-trial basis. A key problem with the testability of that prediction is that one cannot simultaneously measure excitatory and inhibitory current inputs to the same neuron. However, experimenters can measure mean excitatory and inhibitory conductances (g_E and g_I) separately by maintaining the membrane voltage at different levels[25]. Since the ratio g_E/g_I determines the stationary membrane potential of the cell (at which point excitatory and inhibitory currents cancel), a signature of E/I balance is that this ratio will remain constant under different conditions (for example, for different stimuli or behavioral states). The precise value of the ratio, however, does not matter. Average conductance ratios have indeed been found to remain constant in many systems, which largely corroborates the ideas of E/I (current) balance[13,14,26,27]. Furthermore, cortical neurons do operate in a high-conductance regime[28].

The second prediction is that this average balance will be maintained no matter what external input is applied. This prediction was specifically tested in sensory systems where one can vary the stimuli and then observe whether and how that changes the E/I balance. In various sensory cortices and particularly in layer 4 thereof, stimuli that evoke the strongest mean excitation also evoke the strongest inhibition[29–38]. However, no consistent picture concerning the relative tuning of excitatory and inhibitory inputs has emerged. In some experimental preparations, the tuning curves of excitatory and inhibitory conductances are extremely similar[14,29,37,38] (**Fig. 2a**, left). In others, inhibition is more broadly tuned, yet it usually peaks at the same location as excitation[36,39–42] (**Fig. 2a**, right). Moreover, the weights of synapses from inhibitory to excitatory neurons are exquisitely tuned to match the excitation to each neuron, a fine balance that is homeostatically maintained[27]. Finally, in auditory cortex, the co-tuning of excitatory and inhibitory currents is actively enforced by an inhibitory

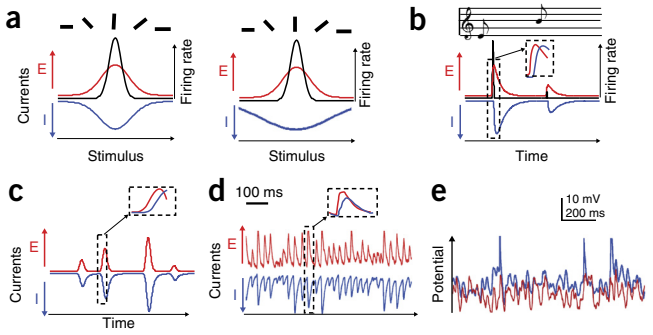

Figure 2 Schematic illustration of key experimental findings. (**a**) Co-tuning of excitatory (red) and inhibitory (blue) inputs into sensory neurons. Black represents the tuning curve of the neuron. We distinguish two types of observations. In one set of observations (left), excitatory and inhibitory inputs are balanced for all conditions or stimuli; firing occurs essentially because of temporal mismatch between excitation and inhibition (see **b**). In another set of observations (right), they are balanced only at the preferred stimuli of a neuron (right), and otherwise inhibition dominates, causing a sharpening of the tuning curve. (**b**) During stimulus presentations, inhibitory inputs match excitatory inputs in size and trail them by a few milliseconds only. (**c**,**d**) During spontaneous activity, inhibition still matches and trails excitation by a few milliseconds. Shown are the time course of inhibitory and excitatory currents and a magnification in time. (**c**) Switching between UP and DOWN states. (**d**) Gamma oscillations in the hippocampus. (**e**) Membrane and spike train correlations between nearby or similar pairs of neurons. Neurons with similar inputs usually exhibit strong co-fluctuations in the membrane voltage and much weaker or nonexistent correlations in their spike times (right). Panel **e** reproduced from ref. 54, Elsevier.

plasticity mechanism, a process that appears crucial for learning and behavioral performance[43–47].

A third prediction is that, when a neural system is driven by strong, dynamic external inputs, inhibition should closely track excitation. This has indeed been found in many experiments. In visual, somatosensory and auditory cortices, the sudden occurrence of a sensory stimulus causes a sharp increase in excitatory currents, followed a few milliseconds later by a sharp increase in inhibitory currents[14,31,33,34]. This short delay between excitation and inhibition, compatible with disynaptic inhibition, opens a brief 'window of opportunity' for the neuron to fire (**Fig. 2b**).

Experimental evidence for a tighter E/I balance

While the above results support the hypothesis that cortical networks are in a balanced state, further experimental observations suggest that this balance may be tighter than expected in the loosely balanced regime. One apparent mismatch between experimental observations and loosely balanced networks concerns spontaneous activity, and contrasts with the stimulus-driven scenario. Loose balance of networks predicts that spontaneous fluctuations in excitatory and inhibitory currents (in the absence of a time-varying external drive) should be uncorrelated. Since one cannot measure conductance ratio by averaging over 'trials' in this case, the prediction cannot be tested directly. However, adjacent neurons in cortex receive strongly correlated synaptic inputs[48,49], so that one can use pairs of nearby neurons to infer the correlations of E and I currents in each one of them. Excitatory and inhibitory events are strikingly correlated during global current fluctuations caused by transitions between the depolarized, high-conductance UP states and the hyperpolarized, low-conductance DOWN states in cortex[13,26,49], with inhibitory events trailing excitatory events by only a few milliseconds (**Fig. 2c**). While somewhat weaker, correlations between E and I currents are also observed within the UP states or in the absence of DOWN states[48,50,51]. Moreover, strong temporal correlations between excitatory and inhibitory currents are observed during oscillations in the gamma frequency range in the hippocampus[52] (**Fig. 2d**) or the beta frequency range in the piriform cortex[42]. In those examples, the phasic increase in inhibition immediately follows the increase in excitation with a delay of a few milliseconds. The amplitude of change in the two current fluctuations are even correlated on a cycle-by-cycle basis[52].

Another mismatch between experimental observations and loosely balanced networks concerns the correlations between the membrane potentials of similarly tuned neurons when they respond to their preferred stimuli. Loosely balanced networks predict that the membrane potentials of such neurons should be uncorrelated. Although spike train correlations range from weak to absent[53], the membrane potential fluctuations of similarly tuned neurons are often correlated. For example, the membrane potential of nearby neurons exhibit high levels of synchrony when they are responding to their preferred stimulus, despite the fact that their spike trains remain irregular and mainly asynchronous[54,55] (**Fig. 2e**).

These experimental observations suggest that we may need to modify a crucial aspect of the theory of loosely balanced networks. Spike generation may indeed not result from a random walk of the membrane potential to threshold, caused by uncorrelated fluctuations in E and I currents. Instead, E and I currents could be tightly locked, such that neurons are 'allowed' to fire only during the brief occasions when inhibition fails to track stimulus drive and spontaneous fluctuations in excitation. Since inhibition is provided by locally dense and strong connections to and from interneurons, spiking may in fact results from a highly interactive, competitive process between neurons in the same local population.

Tightly balanced networks

How does the brain achieve this regime wherein inhibition closely tracks excitation, even in the absence of external stimulation? A recent theoretical study showed that such temporally detailed balance can still be realized in randomly connected networks[56]. In the loosely balanced networks presented in the first section, neurons receive few connections K compared to the size of the network N, so that $K \ll N$. Renart and colleagues investigated a scenario in which connections are dense instead of sparse, so that K is of the same order as N, an assumption compatible with measurements in local microcircuits. In that scenario, excitatory and inhibitory currents received in the same neuron, as well as currents received by different neurons, are strongly correlated, and inhibition tracks excitation on a fine temporal scale[56] (**Fig. 1f**).

However, it is also possible to achieve a tight balance in sparsely connected networks, by using inhibitory plasticity rules that render inhibition more apt at tracking excitation on a fine temporal scale. For example, Vogels and colleagues showed that a rule that reinforces inhibitory connections if pre- and postsynaptic spiking occurs in close temporal proximity, and otherwise weakens them, could achieve a temporally tight balance[57]. This type of spike-time-dependent plasticity rule was later reported in auditory cortex[43].

The problem of coding efficiency revisited

While these theoretical studies show how tight balance can be achieved, a larger question remains that pertains to balance *per se*. Why would neural systems operate in this regime? Why would cortex "simultaneously push on the accelerator and on the brake?"[15]. Why not simply use smaller excitatory currents to achieve the same firing rates but with less variability? Surprisingly, recent theoretical studies

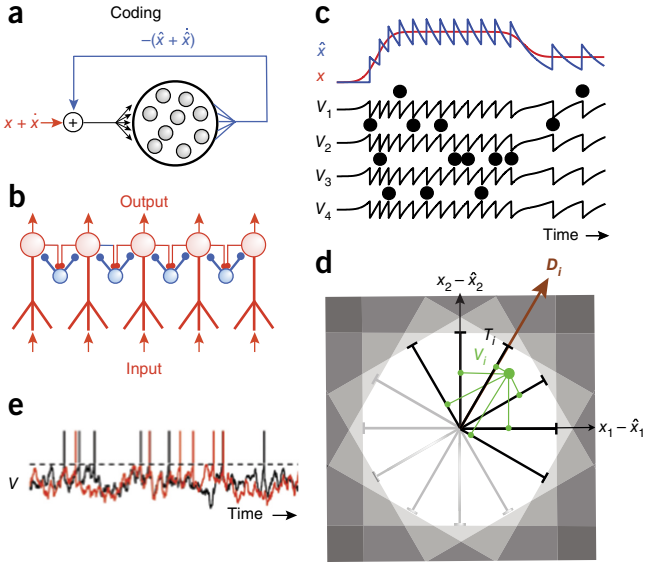

Figure 3 Coding with tightly balanced networks. (**a**) Using balance for coding. Here the network receives an excitatory input signal (red), which for technical reasons combines $x(t)$ and its derivative $\dot{x}(t) = dx/dt$ (refs. 16,18,19). The excitatory input signal in turn is balanced by an inhibitory feedback loop (blue). The feedback loop subtracts the network's representation of the input from the actual input, so that only the momentary representation or decoding error is fed into the network. In the tightly balanced regime, the inhibitory loop cancels the excitatory input, and therefore $\hat{x}(t) \approx x(t)$. (**b**) This schema can be implemented in an architecture of excitatory (red) and inhibitory (blue) neurons. Here the signal $x(t)$ is fed into the dendritic tree of the excitatory neurons. The spiking output of these neurons represents the readout $\hat{x}(t)$, which is fed back through an inhibitory loop to cancel the input signals. (**c**) Toy example with four identical neurons. Shown are the input signal $x(t)$, its estimate $\hat{x}(t)$, and the voltages V_i, and spikes (black dots) of the four neurons. The neuron's voltages track the error $x(t) - \hat{x}(t)$. The neurons take turns firing whenever this error exceeds a threshold. The fast mutual inhibition prevents more than one neuron from firing in each integration cycle. In turn, the input signal is well approximated by the population of four neurons. (**d**) Error space plot. In larger networks, the voltages of the neurons track projections (small green dots) of the reconstruction error (large green dot). Every neuron (index i) represents one specific direction in this error space, given by its feedforward weights $\boldsymbol{D}_i$. Whenever the error becomes too large in this direction, the neuron's voltage $\boldsymbol{V}_i$ exceeds its threshold $\boldsymbol{T}_i$ and the neuron spikes, thereby updating the readout $\hat{\boldsymbol{x}}$ and minimizing the reconstruction error $\boldsymbol{x} - \hat{\boldsymbol{x}}$. (**e**) In larger networks—that is, ones with hundreds or thousands of neurons that represent many variables—the voltages and spike trains no longer look as cartoonish as in **c**. Shown here are the membrane potentials of two neurons with similar tuning. The membrane potentials are strongly correlated, while the spike trains are only weakly correlated.

suggest that neural networks can achieve optimal accuracy and efficiency in their representation by ensuring that the balance is as tight as possible. Accordingly, just as loose balance is related to Poisson rate coding, for which coding errors decrease as $1/\sqrt{M}$ with the number of spikes M, these studies suggest that tight balance is related to efficient coding, for which coding errors decrease as $1/M$ (**Fig. 1c**)[16–18].

The problem of coding is revisited in **Fig. 3**. Here the initial objective is to design a network that represents a time-varying input with maximal efficiency; for example, with the minimum number of spikes for a given level of accuracy. Using this objective, one obtains a network of integrate-and-fire neurons with a specific, dense connectivity (**Fig. 3b**). The excitatory neurons receive the excitatory feedforward input signal and send connections to a population of inhibitory neurons, which in turn cancel (and thereby balance) the feedforward inputs into the excitatory neurons as accurately as possible. This scheme serves to rapidly communicate shared information through mutual inhibition, thus dividing the task of coding the time-varying inputs among the neurons. In turn, the resulting spike code is highly efficient at the population level. This type of network, which we call a 'spike-coding' network, can be interpreted as an auto-encoder that is driven by its own coding errors (**Fig. 3a**).

The essential workings of the spike-coding network can be illustrated in a toy example (**Fig. 3c**). Here four identical neurons seek to encode a time-varying input signal, $x(t)$, in their output spikes, such that the signal can be decoded from the neurons' spikes through linear filtering and summing, yielding an estimate $\hat{x}(t)$. By design, the voltage of each neuron can be reinterpreted to represent the decoding error, so that $V_i(t) \propto x(t) - \hat{x}(t)$ for $i = 1,...4$. Initially, no neuron fires and $\hat{x}(t) = 0$. As the excitatory input $x(t)$ grows, it depolarizes the membrane, which corresponds to an increase in the decoding error. As soon as one of the neurons hits thresholds and fires, the estimate $\hat{x}(t)$ is updated and the voltages or decoding errors of all neurons decrease. This instantaneous decrease of the voltages is mediated through fast inhibition from the spiking neuron (and its own reset). As time progresses, the voltages depolarize again and the game restarts. To put it briefly, the excitatory input signal causes an error in the population code, which is rapidly cancelled by recurrent inhibition as soon as one of the neurons spikes to correct the population code. In this view, tight E/I balance results through the quick resolution of coding errors, first created by excitation and then resolved through inhibition.

Three striking observations follow from this simple toy model. First, despite the practically deterministic tracking of the input signal, each spike train is irregular (or random). In consequence, each neuron exhibits high trial-to-trial variability when the same input signal is repeated[16]. The second observation is that the voltages are highly correlated (identical), but the spike trains are not. The third observation is that the network is robust against eliminating individual spikes[16,17] or against eliminating complete neurons[16,58].

Accordingly, efficient coding with spikes implies that neurons need to be tightly balanced. To what extent the reverse is true—that is, whether tightly balanced systems automatically provide better coding—remains to be fully investigated. The spike-coding network in **Figure 3** is not simply densely connected, but also has higher connection strength and more structured connections than other balanced networks. For instance, assuming a fixed average firing rate, the strength of connections in the network by Renart and colleagues decreases as $1/\sqrt{N}$ (ref. 56), while it remains independent of network size in the spike-coding network[16]. Moreover, neurons in the spike-coding network are not randomly connected, but share reciprocal, symmetrical connections. Finally, the membrane potentials in the spike-coding network are strongly correlated, while they are uncorrelated in the model of Renart *et al.*

Higher coding efficiency opens the door to multidimensional population coding

All of the balanced network models that we have discussed so far assume a population of neurons that share identical inputs. As such, these networks essentially represent a single, unidimensional signal. However, most cortical systems are thought to operate in high-dimensional sensory or motor spaces. Indeed, different neurons receive distinct, only partially overlapping sets of feedforward input signals, and they have different tuning curves as a result. How can a network remain balanced in such conditions?

Theoretically, a simple solution would be to represent each sensory variable by its own private E/I network, with no connections among different networks. In reality, however, sensory variables are multiplexed and each neuron receives multiple sensory inputs. When the efficient coding objective of the spike-coding network in **Figure 3** is extended to multidimensional inputs, it predicts that E and I currents are not only correlated and tightly balanced, but also similarly tuned in each neuron[16]. As before, the depolarization caused by the excitatory inputs corresponds to the accumulation of a coding error in the neurons' membrane voltages. However, since the coding error is now multidimensional, each neuron sees only a part of it, or, more specifically, sees only the projection of the error onto the neuron's feedforward weights in input space (**Fig. 3d**)[16,19]. If the error becomes large enough for one of the neurons to reach threshold, that neuron spikes, and its spike causes fast inhibition (via inhibitory interneurons) of all neurons that need to know that the error just changed. Even though the resulting spike trains are asynchronous and irregular, the networks track the multivariate input signals with an accurate, stepwise approximation (**Fig. 3c**)[16,19]. The resulting code reaches the maximal precision possible with a rate code by cleverly distributing spikes across the population. For large networks, the code outperforms a population of independent Poisson rate neurons by orders of magnitude.

If, in addition, one uses realistic synaptic dynamics rather than instantaneous updates, these networks reproduce the experimental results shown in **Figure 2b–e** (refs. 59,60). Inhibitory currents track excitatory currents very closely with a few milliseconds delay and have the same tuning curve. The membrane potentials are synchronous even if the spike trains are not (**Fig. 3e**). Furthermore, plausible spike-time-dependent plasticity rules can enforce this optimal network structure (ref. 18, W. Brendel, R. Bourdoukan, P. Vertechi, D. Barrett, C.K.M. *et al.*, unpublished observations).

As a direct result of their higher efficiency, the representational capability of the spike-coding network is surprisingly high. Returning to **Figure 3d**, we notice that we could keep the error bounded with only $N = 4$ neurons if, for example, their feedforward weights are aligned with the axes and point in the four cardinal directions. More generally, spike-coding networks can reliably represent up to $N/2$ variables. In other words, a very tightly balanced network could work with an $N/2$-dimensional population code, far beyond what could be expected from a rate-based population code.

Computing without imbalancing

The above section suggests an explanation for why cortical neurons seem so tightly balanced and proposes that cortical networks represent information with high efficiency. While such 'autoencoder' systems can be used for feature extraction or causal inference[19,21], it is not immediately clear how to perform more general computations, such as, for instance, those that require some type of memory. The key problem is to reconcile the need for computation with the constraints imposed by balance. Many classical computational architectures seem to require some type of imbalance, either on the excitatory or on the inhibitory side. These include, among others, attractor networks based on positive feedback[61], Mexican-hat architectures for sensory processing[61] or reservoir computing schemes[62,63]. However, any increase in the excitatory drive will break the E/I balance and therefore lead to more regular firing[10,64] (**Fig. 1d**). As explained above, such imbalances on the excitatory side have not been observed. Rather, it seems that E/I balance is constantly maintained at the level of each neuron.

Three general solutions to this problem have been proposed. One possibility is to move the key nonlinearities used in attractor networks from the output side of a neuron to the input side. Indeed, when using synaptic and/or dendritic nonlinearities to perform computations, interferences with the balanced state can largely be avoided[65]. However, while loose balance can be maintained under such conditions, it is unclear whether one can also maintain tight balance and its efficient coding with spikes.

Another possibility is to use a more tightly balanced network and use the brief window when excitation and inhibition become unbalanced to perform computations[66]. In tightly balanced networks, if an input suddenly increases the activity of excitatory cells, the inhibitory cells quickly catch up, and the difference between excitatory and inhibitory responses decays rapidly back to zero. However, at the same time, the sum total of excitatory and inhibitory responses shows a strong transient amplification. If the network is tightly balanced—that is, if excitatory and inhibitory connections are strong and have a similar structure—this effect will selectively amplify the components of the feedforward inputs that match the patterns of recurrent connectivity, a mechanism that could explain why patterns resembling orientation columns are observed during spontaneous activity in cortex[66]. More recently, Hennequin and colleagues showed that a similar amplification effect can be used to provide a rich reservoir of dynamics in a tightly balanced network[67]. Combined with a simple linear readout, such selective balanced amplification can potentially explain complex population responses in motor cortex[67,68]. Again, however, it is unclear how these ideas, which are mostly worked out for rate networks, relate to the tight balance used for efficient coding with spikes.

Combining coding and computational benefits of tight balance

A third possibility to address the above concern is to separate the time scales on which the system encodes information from the time scale on which it computes. Indeed, the spike-coding framework (**Fig. 3**) can be generalized by introducing a second, slower feedback loop[16,17]. Neurons can then use the faster coding loop to balance both their feedforward input and the additional inputs from the slower feedback loop (**Fig. 4a**). In turn, this slower loop can be used to compute arbitrary linear[16,69] or nonlinear[20,70] differential equations.

The connectivity structure of these networks is shown in **Figure 4b**. Here the excitatory neurons receive three types of inputs: feedforward inputs, fast inhibitory inputs (for example, from parvalbumin cells in cortex) and slow recurrent inputs. The fast recurrent inputs perform the same function as in the spike-coding networks in **Figure 3**. However, the slow recurrent inputs reintroduce a delayed version of the represented variables into the neurons. If this delayed version is combined with dendritic nonlinearities, many kinds of nonlinear operations on the represented variables can be performed[20,71]. In practice, these slow inputs could arise both through recurrent connections from the excitatory population or mediated through additional inhibitory interneurons (not shown).

A particular example of these types of ideas is a 'neural integrator': that is, a network that takes an input $c(t)$ and provides as output

$$x(t) = \int_0^t \mathrm{d}t' c(t').$$

This computation can be implemented in the spike-coding network if the slow recurrent loop is excitatory and exactly matches the fast inhibitory loop in size[16,17,72,73]. Notice that the joint operation of the fast and slow loops effectively feeds the negative derivative $-\mathrm{d}x/\mathrm{d}t$ of the represented signal $x(t)$ back into the network. Since these two loops and the feedforward input, $c(t)$, are balanced, the resulting

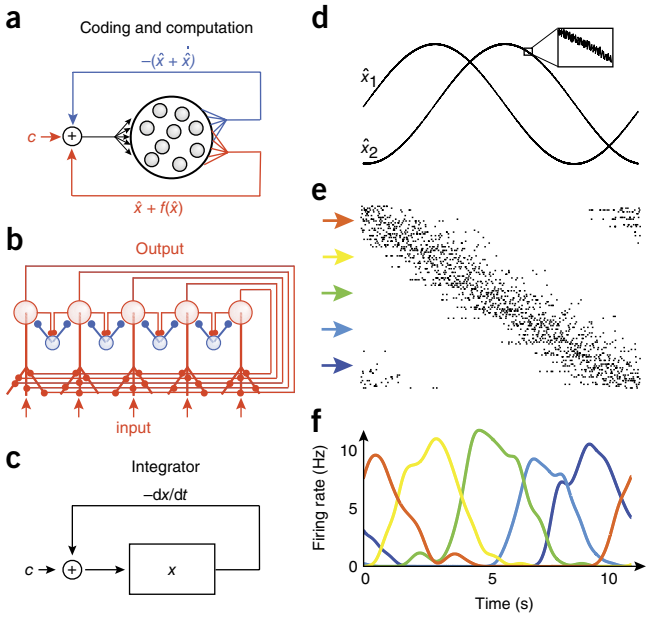

Figure 4 Computations with tightly balanced networks. (**a**) Using balance for computation. Here the network balances an excitatory input signal $c(t)$ and excitatory recurrent feedback $\hat{x}+f(\hat{x})$ with an inhibitory feedback loop (compare **Fig. 3a**). In the tightly balanced regime, the inhibitory and excitatory inputs cancel, and therefore $\dot{\hat{x}} \approx f(\hat{x})+c$, so that the network implements a particular differential equation. (**b**) This schema can likewise be implemented in a network of excitatory and inhibitory neurons. Just as in **Figure 3b**, some of the inhibitory neurons will mediate fast inhibition to construct the efficient population code. In addition to those found in **Figure 3b**, a second set of recurrent connections, which may include additional inhibitory or excitatory neurons (not shown), provides slower feedback and allows the system to perform computations. (**c**) A neural integrator as a specific implementation of this type of architecture. Here the balanced condition dictates that the recurrent input, $-dx/dt$, cancels the feedforward input c. The resulting computation, $dx/dt = c(t)$, implements an integrator. Negative derivative feedback is implemented through a fast inhibitory and a slower excitatory loop. (**d**) Implementation of a harmonic oscillator in a network of $N = 100$ neurons, using the network architecture in **b**, adapted from ref. 16. Shown are the self-generated oscillations of the readout variables $[\hat{x}_1]$ and $[\hat{x}_2]$. The inset highlights the discrete nature of the representation. (**e**) Spike raster of all 100 neurons in the network during a single oscillation phase. (**f**) Trial-averaged, time-varying firing rates of selected neurons.

computation, $c(t) - dx/dt \approx 0$, becomes an integration (**Fig. 4c**)[16,72]. In other words, the interplay of slow excitation and fast inhibition (**Fig. 4a**) can be seen as re-injecting the negative derivative of the population response into the network, thereby stabilizing the representation and getting rid of fast 'undesirable' fluctuations. Biophysically, this could occur if excitatory synapses have significantly slower dynamics than inhibitory synapses (for example, NMDA versus AMPA and $GABA_A$ for E-to-I and I-to-E synapses, respectively)[72,73]. The resulting networks are far more robust than alternative models relying on self-excitation only or on attractor dynamics with local excitation and global inhibition[73].

In **Figure 4d–f**, we show another example of a spike-coding network, this one having 100 neurons and generating the dynamics of a harmonic oscillator. Just as in the toy example in **Figure 3c**, the network creates a stepwise approximation to the desired dynamics (**Fig. 4d**). Individual neurons fire irregular and seemingly noisy spikes (**Fig. 4e**), yet this 'noise' is practically eliminated once the information is read out. Neurons are persistently active at different times during the oscillation cycle (**Fig. 4e,f**), which resembles the type of delay-period activity found in several higher order brain areas such as the striatum[74] or the parietal cortex[75]. Importantly, this network (or any other network following the scheme outlined in **Fig. 4a**) inherits the robustness, efficiency and stability of the autoencoder networks from **Figure 3**. Because of this increased efficiency, the network can achieve much longer integration times or short-term memory periods without suffering from the unstable nature of chaotic, loosely balanced networks.

Breaking the balance: untuned inhibition

Tight E/I balance is frequent, but not universal. There are cases when inhibition dominates or is significantly less tuned than excitation (**Fig. 2a**, right), such as orientation tuning in the visual cortex of mice[39] or odor coding in the olfactory cortex[42,76]. Moreover, excitatory and inhibitory currents have much more complex and variable interactions in deep layers of the visual cortex[30,77].

We propose that these mismatches to tight balance may in part be due to a confound between 'inhibition for coding' (the fast connections in **Fig. 4a,b**) and 'inhibition for computation' (the slow connections in **Fig. 4a,b**). Indeed, there is a fundamental asymmetry of E/I balancing: while inhibition tracks excitation when a neuron is stimulated, the reverse does not seem to hold; that is, excitation does not compensate for inhibition when the neuron is suppressed. This may explain, for example, why inhibition appears to dominate in the visual cortex of awake mice, with neurons spiking for less than 1% of stimulus presentations, while it is balanced with excitation in anesthetized mice, when neurons spike for around 8% of stimulus presentations[78]. Note that neurons in awake auditory cortex do seem to be balanced[79], suggesting that these issues require further investigation.

Interestingly, how well-tuned inhibition is appears to depend on whether the corresponding sensory feature is represented topographically[15]. For example, neighboring neurons have similar orientation selectivity in cat visual cortex[80] but not in rodents[81]. Accordingly, inhibition is similarly tuned to excitation in cats[29] but less so in mice[39]. A similar reasoning can be applied to odor coding in olfactory cortex[42,76]. In agreement with the observation that inhibitory tuning is limited to topographically represented features, inhibition has been reported to be extremely local and dense; that is, inhibitory neurons are connected to most local pyramidal cells[82–84] even if cells of very different selectivity connect to the same interneuron[85].

Loose or tight: a question for the future

In parallel with the increasing evidence for a finely tuned tight balance in cortical networks, we have shown that recent theoretical studies paint a new picture of E/I balance. Here balance does not simply generate noisy spike trains, but rather appears as a fundamental mechanism for the generation of efficient, high-dimensional population representations and computations with spikes.

While we hinted at the possibility of eventually reaching a coherent framework linking dynamic and computational approaches to tight E/I balance, many unanswered questions remain, both on the theoretical and experimental side. One of the most pressing questions concerns the relationship between the tightly balanced networks seen from a dynamic perspective, such as in the model by Renart and colleagues[56], and spike-coding networks derived from a computational objective[16]. They share some similarities in their phenomenology, but also major differences concerning the strength and structure of connections or the correlations in membrane potentials. It is unclear to what extent

all tightly balanced network inherit some of desirable properties of spike-coding networks.

Let us emphasize as well that experimental data remain ambiguous and do not uniquely support tightly balanced over loosely balanced networks, especially in awake, high conductance regimes. In fact, we might expect that cortical networks exist somewhere in a continuum between loosely balanced network, inefficient but cheap in terms of number of connections per neuron and structure, and very tightly balanced spike-coding networks, highly efficient but with extremely structured, dense connections that have to constantly be maintained by spike-timing-dependent plasticity rules.

What experiment could be performed to locate biological networks along this continuum? One promising direction would be to combine inactivation experiment (for example, using optogenetics) with recordings and behavior[58]. Indeed, spike-coding networks exhibit supreme robustness, not predicted for randomly connected networks. Another option would be to analyze the neural variability in more detail, using massive multiunit recordings (for example, calcium imaging or dense arrays). A large component of this variability should indeed be orthogonal to the stimulus representation (that is, it would not affect the decoding of the stimulus from the neural responses). Results from such studies could decide in the near future whether or not we need to revise widely shared ideas about the neural code.

ACKNOWLEDGMENTS

We thank A. Renart and B. Atallah for discussions.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Adrian, E.D. & Zotterman, Y. The impulses produced by sensory nerve-endings: Part II. The response of a single end-organ. *J. Physiol. (Lond.)* **61**, 151–171 (1926).
2. Softky, W.R. & Koch, C. The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J. Neurosci.* **13**, 334–350 (1993).
3. Perkel, D.H. & Bullock, T.H. Neural coding. *Neurosci. Res. Program Bull.* **6**, 220–349 (1968).
4. Abeles, M. Role of the cortical neuron: integrator or coincidence detector? *Isr. J. Med. Sci.* **18**, 83–92 (1982).
5. Hopfield, J.J. Pattern recognition computation using action potential timing for stimulus representation. *Nature* **376**, 33–36 (1995).
6. Rieke, F., Warland, D., de Ruyter van Steveninck, R. & Bialek, W. *Spikes: Exploring the Neural Code* (MIT Press, 1997).
7. Fiser, J., Berkes, P., Orbán, G. & Lengyel, M. Statistically optimal perception and learning: from behavior to neural representations. *Trends Cogn. Sci.* **14**, 119–130 (2010).
8. Shadlen, M.N. & Newsome, W.T. Noise, neural codes and cortical organization. *Curr. Opin. Neurobiol.* **4**, 569–579 (1994).
9. van Vreeswijk, C. & Sompolinsky, H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* **274**, 1724–1726 (1996).
10. Amit, D.J. & Brunel, N. Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb. Cortex* **7**, 237–252 (1997).
11. Shadlen, M.N. & Newsome, W.T. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.* **18**, 3870–3896 (1998).
12. Brunel, N. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci.* **8**, 183–208 (2000).
13. Shu, Y., Hasenstaub, A. & McCormick, D.A. Turning on and off recurrent balanced cortical activity. *Nature* **423**, 288–293 (2003).
14. Wehr, M. & Zador, A.M. Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature* **426**, 442–446 (2003).
15. Isaacson, J.S. & Scanziani, M. How inhibition shapes cortical activity. *Neuron* **72**, 231–243 (2011).
16. Boerlin, M., Machens, C.K. & Denève, S. Predictive coding of dynamical variables in balanced spiking networks. *PLoS Comput. Biol.* **9**, e1003258 (2013).
17. Boerlin, M. & Denève, S. Spike-based population coding and working memory. *PLoS Comput. Biol.* **7**, e1001080 (2011).
18. Bourdoukan, R., Barrett, D., Deneve, S. & Machens, C.K. Learning optimal spike-based representations. In *Advances in Neural Information Processing Systems* (eds. Pereira, F., Burges, C.J.C., Bottou, L. & Weinberger, K.Q.) 2285–2293 (MIT Press, 2012).
19. Barrett, D.G., Deneve, S. & Machens, C.K. Firing rate predictions in optimal balanced networks. In *Advances in Neural Information Processing Systems* (eds. Pereira, F., Burges, C.J.C., Bottou, L. & Weinberger, K.Q.) 1538–1546 (MIT Press, 2013).
20. Thalmeier, D., Uhlmann, M., Kappen, H.J. & Memmesheimer, R.-M. Learning universal computations with spikes. Preprint at http://arXiv.org/abs/1505.07866 (2015).
21. Moreno-Bote, R. & Drugowitsch, J. Causal inference and explaining away in a spiking network. *Sci. Rep.* **5**, 17531 (2015).
22. van Vreeswijk, C. & Sompolinsky, H. Chaotic balanced state in a model of cortical circuits. *Neural Comput.* **10**, 1321–1371 (1998).
23. Monteforte, M. & Wolf, F. Dynamical entropy production in spiking neuron networks in the balanced state. *Phys. Rev. Lett.* **105**, 268104 (2010).
24. London, M., Roth, A., Beeren, L., Häusser, M. & Latham, P.E. Sensitivity to perturbations *in vivo* implies high noise and suggests rate coding in cortex. *Nature* **466**, 123–127 (2010).
25. Monier, C., Fournier, J. & Frégnac, Y. *In vitro* and *in vivo* measures of evoked excitatory and inhibitory conductance dynamics in sensory cortices. *J. Neurosci. Methods* **169**, 323–365 (2008).
26. Haider, B., Duque, A., Hasenstaub, A.R. & McCormick, D.A. Neocortical network activity *in vivo* is generated through a dynamic balance of excitation and inhibition. *J. Neurosci.* **26**, 4535–4545 (2006).
27. Xue, M., Atallah, B.V. & Scanziani, M. Equalizing excitation-inhibition ratios across visual cortical neurons. *Nature* **511**, 596–600 (2014).
28. Destexhe, A., Rudolph, M. & Paré, D. The high-conductance state of neocortical neurons *in vivo*. *Nat. Rev. Neurosci.* **4**, 739–751 (2003).
29. Anderson, J.S., Carandini, M. & Ferster, D. Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex. *J. Neurophysiol.* **84**, 909–926 (2000).
30. Martinez, L.M., Alonso, J.M., Reid, R.C. & Hirsch, J.A. Laminar processing of stimulus orientation in cat visual cortex. *J. Physiol. (Lond.)* **540**, 321–333 (2002).
31. Tan, A.Y., Zhang, L.I., Merzenich, M.M. & Schreiner, C.E. Tone-evoked excitatory and inhibitory synaptic conductances of primary auditory cortex neurons. *J. Neurophysiol.* **92**, 630–643 (2004).
32. Tan, A.Y., Brown, B.D., Scholl, B., Mohanty, D. & Priebe, N.J. Orientation selectivity of synaptic input to neurons in mouse and cat primary visual cortex. *J. Neurosci.* **31**, 12339–12350 (2011).
33. Tan, A.Y. & Wehr, M. Balanced tone-evoked synaptic excitation and inhibition in mouse auditory cortex. *Neuroscience* **163**, 1302–1315 (2009).
34. Wilent, W.B. & Contreras, D. Dynamics of excitation and inhibition underlying stimulus selectivity in rat somatosensory cortex. *Nat. Neurosci.* **8**, 1364–1370 (2005).
35. Liu, B.H. *et al.* Broad inhibition sharpens orientation selectivity by expanding input dynamic range in mouse simple cells. *Neuron* **71**, 542–554 (2011).
36. Wu, G.K., Arbuckle, R., Liu, B.H., Tao, H.W. & Zhang, L.I. Lateral sharpening of cortical frequency tuning by approximately balanced inhibition. *Neuron* **58**, 132–143 (2008).
37. Cardin, J.A., Palmer, L.A. & Contreras, D. Stimulus feature selectivity in excitatory and inhibitory neurons in primary visual cortex. *J. Neurosci.* **27**, 10333–10344 (2007).
38. Runyan, C.A. *et al.* Response features of parvalbumin-expressing interneurons suggest precise roles for subtypes of inhibition in visual cortex. *Neuron* **67**, 847–857 (2010).
39. Liu, B.H. *et al.* Visual receptive field structure of cortical inhibitory neurons revealed by two-photon imaging guided recording. *J. Neurosci.* **29**, 10520–10532 (2009).
40. Kerlin, A.M., Andermann, M.L., Berezovskii, V.K. & Reid, R.C. Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex. *Neuron* **67**, 858–871 (2010).
41. Niell, C.M. & Stryker, M.P. Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron* **65**, 472–479 (2010).
42. Poo, C. & Isaacson, J.S. Odor representations in olfactory cortex: "sparse" coding, global inhibition, and oscillations. *Neuron* **62**, 850–861 (2009).
43. D'amour, J.A. & Froemke, R.C. Inhibitory and excitatory spike-timing-dependent plasticity in the auditory cortex. *Neuron* **86**, 514–528 (2015).
44. Dorrn, A.L., Yuan, K., Barker, A.J., Schreiner, C.E. & Froemke, R.C. Developmental sensory experience balances cortical excitation and inhibition. *Nature* **465**, 932–936 (2010).
45. Froemke, R.C., Merzenich, M.M. & Schreiner, C.E. A synaptic memory trace for cortical receptive field plasticity. *Nature* **450**, 425–429 (2007).
46. Marlin, B.J., Mitre, M., D'amour, J.A., Chao, M.V. & Froemke, R.C. Oxytocin enables maternal behaviour by balancing cortical inhibition. *Nature* **520**, 499–504 (2015).
47. Froemke, R.C. *et al.* Long-term modification of cortical synapses improves sensory perception. *Nat. Neurosci.* **16**, 79–88 (2013).
48. Graupner, M. & Reyes, A.D. Synaptic input correlations leading to membrane potential decorrelation of spontaneous activity in cortex. *J. Neurosci.* **33**, 15075–15085 (2013).
49. Okun, M. & Lampl, I. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nat. Neurosci.* **11**, 535–537 (2008).
50. Tan, A.Y.Y., Andoni, S. & Priebe, N.J. A spontaneous state of weakly correlated synaptic excitation and inhibition in visual cortex. *Neuroscience* **247**, 364–375 (2013).
51. Salkoff, D.B., Zagha, E., Yüzgeç, Ö. & McCormick, D.A. Synaptic mechanisms of tight spike synchrony at gamma frequency in cerebral cortex. *J. Neurosci.* **35**, 10236–10251 (2015).

52. Atallah, B.V. & Scanziani, M. Instantaneous modulation of gamma oscillation frequency by balancing excitation with inhibition. *Neuron* **62**, 566–577 (2009).
53. Cohen, M.R. & Kohn, A. Measuring and interpreting neuronal correlations. *Nat. Neurosci.* **14**, 811–819 (2011).
54. Yu, J. & Ferster, D. Membrane potential synchrony in primary visual cortex during sensory stimulation. *Neuron* **68**, 1187–1201 (2010).
55. Gentet, L.J., Avermann, M., Matyas, F., Staiger, J.F. & Petersen, C.C. Membrane potential dynamics of GABAergic neurons in the barrel cortex of behaving mice. *Neuron* **65**, 422–435 (2010).
56. Renart, A. *et al.* The asynchronous state in cortical circuits. *Science* **327**, 587–590 (2010).
57. Vogels, T.P., Sprekeler, H., Zenke, F., Clopath, C. & Gerstner, W. Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science* **334**, 1569–1573 (2011).
58. Barrett, D.G.T., Deneve, S. & Machens, C.K. Optimal compensation for neuron death. Preprint at http://biorxiv.org/content/early/2015/10/20/029512 (2015).
59. Schwemmer, M.A., Fairhall, A.L., Denéve, S. & Shea-Brown, E.T. Constructing precisely computing networks with biophysical spiking neurons. *J. Neurosci.* **35**, 10112–10134 (2015).
60. Chalk, M., Gutkin, B. & Deneve, S. Neural oscillations as a signature of efficient coding in the presence of synaptic delays. Preprint at doi:10.1101/034736 (2015).
61. Dayan, P. & Abbott, L.F. *Theoretical Neuroscience vol. 806* (MIT Press, 2001).
62. Jaeger, H. The echo state approach to analysing and training recurrent neural networks–with an erratum note. GMD Technical Report 148:34 (German National Research Center for Information Technology, Bonn, 2001).
63. Maass, W., Natschläger, T. & Markram, H. Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput.* **14**, 2531–2560 (2002).
64. Compte, A., Brunel, N., Goldman-Rakic, P.S. & Wang, X.J. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex* **10**, 910–923 (2000).
65. Hansel, D. & Mato, G. Short-term plasticity explains irregular persistent activity in working memory tasks. *J. Neurosci.* **33**, 133–149 (2013).
66. Murphy, B.K. & Miller, K.D. Balanced amplification: a new mechanism of selective amplification of neural activity patterns. *Neuron* **61**, 635–648 (2009).
67. Hennequin, G., Vogels, T.P. & Gerstner, W. Optimal control of transient dynamics in balanced networks supports generation of complex movements. *Neuron* **82**, 1394–1406 (2014).
68. Churchland, M.M. *et al.* Neural population dynamics during reaching. *Nature* **487**, 51–56 (2012).
69. Bourdoukan, R. & Deneve, S. Enforcing balance allows local supervised learning in spiking recurrent networks. In *Advances in Neural Information Processing Systems* (in the press).
70. Abbott, L.F., DePasquale, B. & Memmesheimer, R.-M. Building functional networks of spiking model. *Nat. Neurosci.* **19**, 350–355 (2016).
71. Poirazi, P., Brannon, T. & Mel, B.W. Pyramidal neuron as two-layer neural network. *Neuron* **37**, 989–999 (2003).
72. Lim, S. & Goldman, M.S. Balanced cortical microcircuitry for spatial working memory based on corrective feedback control. *J. Neurosci.* **34**, 6790–6806 (2014).
73. Lim, S. & Goldman, M.S. Balanced cortical microcircuitry for maintaining information in working memory. *Nat. Neurosci.* **16**, 1306–1314 (2013).
74. Mello, G.B., Soares, S. & Paton, J.J. A scalable population code for time in the striatum. *Curr. Biol.* **25**, 1113–1122 (2015).
75. Harvey, C.D., Coen, P. & Tank, D.W. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature* **484**, 62–68 (2012).
76. Stettler, D.D. & Axel, R. Representations of odor in the piriform cortex. *Neuron* **63**, 854–864 (2009).
77. Monier, C., Chavane, F., Baudot, P., Graham, L.J. & Frégnac, Y. Orientation and direction selectivity of synaptic inputs in visual cortical neurons: a diversity of combinations produces spike tuning. *Neuron* **37**, 663–680 (2003).
78. Haider, B., Häusser, M. & Carandini, M. Inhibition dominates sensory responses in the awake cortex. *Nature* **493**, 97–100 (2013).
79. Zhou, M. *et al.* Scaling down of balanced excitation and inhibition by active behavioral states in auditory cortex. *Nat. Neurosci.* **17**, 841–850 (2014).
80. Hubel, D.H. & Wiesel, T.N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. (Lond.)* **160**, 106–154 (1962).
81. Ohki, K., Chung, S., Ch'ng, Y.H., Kara, P. & Reid, R.C. Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature* **433**, 597–603 (2005).
82. Fino, E. & Yuste, R. Dense inhibitory connectivity in neocortex. *Neuron* **69**, 1188–1203 (2011).
83. Packer, A.M., McConnell, D.J., Fino, E. & Yuste, R. Axo-dendritic overlap and laminar projection can explain interneuron connectivity to pyramidal cells. *Cereb. Cortex* **23**, 2790–2802 (2013).
84. Ren, M., Yoshimura, Y., Takada, N., Horibe, S. & Komatsu, Y. Specialized inhibitory synaptic actions between nearby neocortical pyramidal neurons. *Science* **316**, 758–761 (2007).
85. Bock, D.D. *et al.* Network anatomy and *in vivo* physiology of visual cortical neurons. *Nature* **471**, 177–182 (2011).