Transmission of population coded information

Alfonso Renart, Mark C. W. van Rossum

Instituto de Neurociencias de Alicante. Universidad Miguel Hernandez - CSIC 03550 Sant Joan d’Alacant, Spain, Center for Molecular and Behavioral Neuroscience, Rutgers University 197 University Avenue, Newark, NJ 07102, USA, and Institute of Adaptive and Neural Computation, School of Informatics, University of Edinburgh, 10 Crichton Street, Edinburgh EH8 9AB, UK

Submitted to Proceedings of the National Academy of Sciences of the United States of America

As neuronal activity is transmitted through the nervous system, neuronal noise is believed to degrade the encoded information and limit performance. It is therefore important to know how loss of information can be prevented. Using a principled approach we show how information loss in a layered network depends on the connectivity between the layers and find that the optimal connection profile has a center-surround structure with a spatial extent closely matching the neurons’ tuning curves. In addition we show how the optimal connectivity depends on the correlation structure of the trial-to-trial variability in the neuronal responses. Our results suggest that the ubiquitous center-surround architectures in the nervous system provide an optimal way to communicate population coded information.

Abbreviations: FI, Fisher information

Neural information often passes through many different brain areas. Consider for instance a task where a subject is to reach for a visual target. Information about the target location will have to be transmitted across many stages: retina, LGN, visual, pre-motor, and motor cortices. Due a variety of noise sources, single neurons will be highly variable and unreliable at every stage [1, 2, 3]. It is believed that robustness is engendered by population codes, in which many neurons are activated by a given stimulus. Population coding has been studied extensively. In particular the role of the tuning curve shape and the neural noise model [4, 5], as well as the effect of correlations in the trial-to-trial variability of different neurons, have been studied in detail [6, 7, 8, 9]. Moreover, several algorithms have been proposed to ‘read out’ population codes [10, 11, 12].

Yet, although it is well-known how much information is available in a given population and how it can be read out, it is not known how information is preserved when it is transmitted between processing stages. A few studies have addressed the transmission of population codes [13, 14], but transmission quality has not been quantified. This is functionally important: in the example reaching task above, where each processing stage presumably uses a population code, suboptimal transmission would lead to errors in the perceived target location and thus a decreased performance.

In this study, we use the Fisher Information (FI) to quantify how information about a stimulus feature degrades as activity is transmitted between layers. Next, we use these results to find the synaptic weight profile that minimizes information loss and we find it to be center-surround (Mexican-hat), with a width approximately matched to the neurons’ tuning width. The generality of our results is tested with numerical simulations of integrate-and-fire neurons.

Results

As an concrete example for the propagation of information across a hierarchy of processing areas, we analyze the transmission of population coded information between two layers. We consider the network shown in Fig. 1A. Each of the two layers consists of a population of $N$ neurons encoding an angular variable labeled $\theta$, which could represent, for instance, orientation. The response of each neuron $i$ in the input layer depends on the difference between its preferred orientation $\phi_i$ and $\theta$, resulting in a population activity ‘bump’ on each trial (Fig. 1A top).

We denote the firing rate of neuron $i$ in the input layer as $r_i^{\text{in}}$. Its average across trials is given by $f(\theta - \phi_i)$, where $f$ specifies the tuning curve shape. Trial-to-trial response variability can in principle be correlated across neurons (so called ‘noise’ correlation, quantified by a covariance matrix $C_0$). The structure of the noise correlation can be dependent on the similarity between the tuning of different cells, i.e., $[C_0]_{ij} = C_0(\phi_i - \phi_j)$, but is assumed independent of the stimulus. Neurons in the output layer process the activity of the input layer through a set of synaptic weights $W$. The magnitude of the synaptic weight between an input and an output neuron is assumed to depend on the difference between their preferred angles only, $W_{ij} = W(|\phi_i - \phi_j|)$, so that activity in the input and output layers is aligned [15]. The firing rate of neuron $j$ in the output layer on a given trial is

$$r_j^{\text{out}} = g(I_j) = g\left(\frac{1}{N} \sum_{i=1}^{N} W_{ij} r_i^{\text{in}} + \eta_j\right)$$

where $I_j$ is the current to neuron $j$ and $g(\ldots)$ represents a non-linearity that maps the input current of a neuron into its firing rate. The term $\eta_j$, modeled as zero-mean Gaussian noise current with covariance matrix $C_1$, represents external noise unrelated to the stimulus and to the activity of the previous layer.

We use the FI [16, 17] to quantify how much information about the stimulus is available in both layers. Under the above assumptions the FI is elegantly formulated in spatial frequency space [7, 8]. The FI in the input layer is

$$J_{\text{in}} = \sum_{n=0}^{N-1} \frac{\tilde{f}(n)^2}{C_0(n)}$$

where $n$ sums over the Fourier modes, and $\tilde{C}_0(n)$ and $\tilde{f}(n)$ are the $n$th Fourier components of the noise correlation matrix and of the derivative of the tuning curve with respect to $\theta$. This equation has an intuitive interpretation; it expresses the FI as a sum of signal-to-noise ratios $J(n) = |\tilde{f}(n)|^2 / \tilde{C}_0(n)$ at every spatial frequency. For a given spatial frequency, the signal is

Reserved for Publication Footnotes
the squared magnitude of the Fourier mode of the derivative of the tuning curve, while the noise is the corresponding Fourier mode of the covariance matrix. Because typical neuronal tuning curves are smooth and broad, only low spatial frequencies carry information. Note, however, that for homogeneous networks the zero mode of the signal (which equals the average derivative of the tuning curve) carries no information, as the total activity in the network is independent of $\theta$.

The currents to the neurons in the output layer have a Gaussian distribution with mean $I_i(\theta) = (1/N) \sum_j W_{ij} f_j(\theta)$ and covariance $C_{\theta ij} = (1/N^2) \sum_{kl} W_{ik} C_{\theta kl} W_{lj} + C_{\theta ij}$. Although the output firing rate distribution is not Gaussian, due to the non-linearity $g(\cdot)$, as long as this function is invertible, e.g. a sigmoid or a powerlaw with positive argument, the FI in the output rates equals the FI in the synaptic currents which is again a decoupled sum of Fourier components

$$J_{\text{out}} = \frac{N-1}{N} \sum_{n=0}^{N-1} J(n) \left[ \frac{|\tilde{W}(n)|^2}{|\tilde{W}(n)|^2 + T(n)} \right]$$

where $T(n) = \tilde{C}_1(n)/\tilde{C}_0(n)$ and $\tilde{W}(n)$ and $\tilde{C}_1(n)$ are the Fourier components of the weight profile and of the covariance matrix of the external noise. Compared to the FI in the input layer, Eq. 1, the output information at every frequency is attenuated by a factor between zero and one (the term in the square brackets). The output information is therefore always less or equal to the input information. The factor approaches one if the corresponding frequency mode of the synaptic weight is respectively much larger than the effective threshold $T(n)$. Only if there is no external noise in the output layer, $\tilde{C}_1(n) = 0$, the attenuation factors are exactly one. In that case no information is lost between the two layers and, save for the completely flat profile, any profile will suffice.

Next, we optimize the connectivity between input and output layer so that it minimizes the information loss between the input and the output layer. The loss is minimal if the attenuation factor in Eq. 2 approaches one for all frequencies. This can be trivially achieved if the weights strongly amplify all spatial frequencies, i.e. if $|\tilde{W}(n)|^2$ is sufficiently large. This is for instance the case if the weight matrix is a large multiple of the identity matrix. In that case the noise in the output layer becomes negligible compared to the synaptic input, and the FI in the output layer approaches the FI of the input layer. However, this solution implies unphysiologically large currents and high firing rates in the output layer. Clearly, a constraint in the magnitude of the weights has to be imposed. It is reasonable to constrain the magnitude of the weights by imposing that the spatial average of $|W(\theta_1 - \phi_1)|^2$ be of order unity, with the square preventing the constraint to be satisfied through a cancellation between positive (excitatory) and negative (inhibitory) weights. In spatial frequency space, this is equivalent to the constraint

$$\sum_{n=0}^{N-1} |\tilde{W}(n)|^2 = q$$

with $q$ of order unity.

Optimization of the FI in the output layer in the presence of the constraint effectively introduces competition between the different frequency modes of the weights. The optimal weights transmit only the most informative spatial frequencies; assigning power in the weakly informative spatial frequencies does not increase the transmitted information much, but, due to the constraint, takes power away from the informative modes. Full details of this multi-dimensional constrained optimization process, schematically illustrated in Fig. 1B, can be found in the Appendix. There, it is shown that the optimal weights are given by

$$|\tilde{W}(n)|^2 = \sqrt{T(n)} \left[ \frac{\sqrt{J(n)/\lambda} - \sqrt{T(n)}}{\sqrt{\lambda}} \right]_+$$

where $[x]_+ = \max(0,x)$. The constant $\lambda$ is given by

$$1 = \frac{\sum_{n'} |\tilde{W}(n')|^2 + T(n')}{\sum_{n'} \sqrt{T(n')}}$$

and the sum over $n'$ runs over all frequency modes where the weights are non-zero. Due to the thresholding non-linearity in Eq. 3, which appears because power can only be positive, the previous equations have to be solved self-consistently.

We consider first the simple case where the external noise and neurons in the input layer are independent, i.e. $C_{\theta 1 ij} = \sigma^2_{\theta ij} \delta_{ij}$. In this case, the optimal weights are simply a thresholded version of the signal, $|\tilde{W}(n)|^2 = |\alpha f(n)| - T_+$, with $\alpha$ and $T$ constants. Optimal weight profiles in this case are shown in Fig. 2A for three different widths of the input tuning curve: When the input tuning curves are very narrow, the intermediate frequencies are most informative, giving the optimal weight profile an oscillatory character. For wider tuning curves, the low frequencies modes are most informative, yielding a smoother weight profile. In general, the width of the optimal connectivity approximately matches that of the input tuning curve (Fig. 2B). The optimal weights depend on the noise as well. If, for instance, the post-synaptic noise is small, the optimal weights will transmit intermediate spatial frequencies. If, in contrast, the postsynaptic noise is large, the weights are concentrated around the large, low frequency modes, yielding broader weight profiles.

We next address the optimal transmission strategy in the presence of noise correlations. Noise correlations reduce the information of neural codes, as they limit the potential for averaging out noise by pooling over neurons [18, 19], which is reflected in a saturation of the FI in the presence of limited range correlations [7, 8]. Noise correlations affect the optimal weights through two different mechanisms. First, they substantially change the amount of information present at each spatial frequency in the input layer (Fig. 2D). In particular, limited range correlations reduce the signal-to-noise ratio predominantly at low spatial frequencies, which are the most informative. Second, they lower the effective thresholds $T(n)$, so that less power in the low frequency weights is required (Fig. 2D inset; Eq. 3). Thus in the presence of limited range correlations the optimal weights emphasize the intermediate frequencies more than for uncorrelated noise, leading to a narrower profile (Fig. 2E-F).

Although the precise shape of the optimal weight profile depends on detailed properties of the population firing statistics (Eq. 3), it has several general properties. Because there is no information about the stimulus in the total activity of the population, the optimal weights are ‘balanced’: their zero-frequency mode (the area under the weight profile) is zero. Secondly, if the tuning curves are not too narrow, most information is in the lowest spatial frequencies, which are amplified the most by the optimal profile. In summary, the optimal weights have most power in the low frequency modes and none at zero, which implies that they have a center-surround or ‘Mexican-hat’ structure.

In order to test the generality of the results we investigated a two layer network of spiking integrate and fire neurons (see Supplementary Information). The connectivity profile in the simulations is parametrized by the width of the excitatory center and the strength of the inhibition, while the gain in the connectivity is tuned so that the response amplitudes in the layers are identical. A long lasting stimulus is applied and the
spikes in the output layer of the network are collected in time bins and the stimulus is estimated, Fig. 3A. Although these simulation have the advantage of being realistic, measuring the true FI in simulations of spiking neurons is computationally prohibitive for non-Gaussian, correlated noise. Instead we use an optimal linear discriminator to measure the accuracy of the population code in the output layer (see Methods) [15, 20]. In Fig. 3B the trial-to-trial standard deviation in the stimulus angle (a lower bound on the inverse square root of the FI) is plotted as a function of the width of the connectivity profile and various levels of inhibition. The error can be compared to the error in the input layer (gray line). Using an excitatory-only weight profile, the error in the output layer (dotted curve) is substantially higher than in the error in the input. The error in the output is largest for wide weight profiles, as in this case the output activity is smeared, Fig. 3C(iii). Adding surround inhibition substantially improves the accuracy in the output layer. Using a balanced, center-surround profile (solid line) with a half-width approximately matching the width of the activity profile (arrow), the error is minimal. The output activity profile is in this case similar to the input, Fig. 3(ii).

The amount of inhibition required is not very critical as long as the width of the profile is matched to that of the tuning curve. A three-fold increase in the inhibition yields similar performance for a profile with an optimally tuned width as the width of the profile is matched to that of the tuning curve. A three-fold increase in the inhibition yields significantly different (Fig. 3C(i)). These simulation results are in line with the theoretical analysis above: in order to preserve information, the connectivity profile needs to amplify spatial frequencies that convey the most information, while filtering out the high and zero spatial frequencies.

### Discussion

We have considered a scenario where information is processed in stages and thus needs to be represented in multiple brain regions. Our algorithm maximizes the transmission of Fisher information, and can be seen as the analog of the water-filling algorithm used to maximize transmission of Shannon information [16]. The results show that the information encoded by a population of broadly tuned neurons is best transmitted through center surround connectivity profiles.

We believe that quantifying information transmission between populations is an important extension to reading out information in a given population. A read-out algorithm can be thought of as a map from the high dimensional population activity to a low-dimensional encoded quantity, such as stimulus orientation. But in the nervous system, although the encoded quantity is processed as activity flows through, say, the sensory-motor stream, subsequent stages do not read-out the previous one. Instead, the high-dimensional nature of the encoding process is preserved until the very last stages where a small set of muscles are driven by the corresponding motor-neurons.

Center surround profiles have been shown to be useful for minimizing redundancy at early stages of the visual processing stream, as they remove broad stimulus-induced correlations [21]. For the transmission of population-coded information considered here however, the center-surround solution has a different origin, i.e., the need to filter out the uninformative high- and zero spatial-frequency modes. While the current study only considers transmission of activity, the ultimate goal is to understand information processing. The findings here are also applicable to computing networks: In an earlier study using recurrent networks to implement an optimal read-out algorithm, the width of the connectivity profile had to be set approximately equal to the tuning curve width [12]. Another study found that spiking feed-forward networks employing a width matched center-surround connectivity accurately computed with population codes [22]. Finally, in man-made readout of population codes matching the connectivity to the activity profile was found to give good performance [23].

Our findings have implications for the interpretation of neural connectivity. In sensory systems center-surround has been interpreted as a contrast enhancement mechanism. However, it has been unclear why contrast enhancement would be preserved every stage. Our results suggests that the abundance of center-surround connectivity in many parts of the nervous system [24, 25, 26, 27], is important for accurate population code processing.

### Materials and Methods

Briefly, we considered two layers of \( N \) neurons with no lateral interactions (Fig. 1A). Neurons in the input layer encode the location of an angular variable through bell-shaped tuning curves, leading to a noisy population ’bump’ on each trial. Tuning curves are parametrized as

\[
f_i(\phi - \theta) = r_p \exp\left(\frac{\cos(2(\phi_i - \theta)) - 1}{a^2}\right) \equiv f_i(\theta) \tag{4}
\]

where \( \phi_i = (2i-1)/N \) and \( N \) is the preferred orientation of cell \( i \), \( \theta \) is the stimulus orientation and \( r_p \) and \( a \) are the peak firing rate of the cell and tuning curve width, respectively.

**Analytic study.** Trial-to-trial response variability in the input layer was assumed to be Gaussian, so that the response probability for the whole layer on a trial is given by an \( N \)-dimensional Gaussian distribution with mean vector \( \{ f_i(\theta) \} \) and a covariance matrix

\[
[C_{ij}] = C_0 (\phi_i - \phi_j) = (\sigma_0^2 - \sigma_0 \delta_{ij} + c \exp\left[-\frac{2|\phi_i - \phi_j|}{\rho_0}\right])
\]

where \( \sigma_0^2 \), \( c \), and \( \rho_0 \) are the firing rate variance of the neurons, peak correlation and correlation length, respectively. The external input was also assumed Gaussian, with zero-mean and a covariance matrix \( C_1 \) parameterized by \( \sigma_1^2 \), \( c_1 \) and \( \rho_1 \). We estimated the information about the stimulus orientation in the population activity using the FI which asymptotically provides a lower bound for the variance of any unbiased estimator [16, 17]. If \( p(\theta | f) \) is the probability distribution of activity in a given trial, then \( FI = -\langle \log p(\theta | f) \rangle \) which, for a Gaussian distribution with mean \( \langle f \rangle \) and stimulus-independent covariance \( C \) results in

\[
FI = 2 \langle f \rangle C^{-1} \langle f \rangle
\]

We use the following convention for the Fourier transform:

\[
\hat{f}_n = (1/N) \sum_{j=1}^{N} e^{-2 \pi i |n|/N} f_j
\]

Although we write the Fourier components of different quantities as ranging from \( n = 0, \ldots, N - 1 \), in the figures we displayed the highest frequencies as negative (\( n = N - m \rightarrow n = -m \)) to highlight the symmetry of the Fourier transforms.

**Numerical Study.** For the simulation study we use a layered network consisting of 128 current-based integrate-and-fire neurons per layer with periodic boundary conditions. In order to prevent synchronization, all neurons are injected with a Gaussian-distributed, low-pass filtered noisy background current. For simplicity neurons can make both excitatory and inhibitory synapses. Biophysical parameters of the neurons, synapses and background noise are as in [28].

The input layer is stimulated by a static profile of current injections with a Gaussian shape according to Eq. 4, with a full width at half maximum of 60° degrees. The strength of the synaptic connections between layers is parametrized by a rectified cosine with a uniform inhibitory surround. The connection strength between a neuron at position \( i \) and a neuron at position \( j \) in the next layer, equals \( W_{ij} = A \cos((\phi_j - \phi_i)/90°) \) whenever \( \phi_j \leq \phi_i \), while \( W_{ij} = -c \sum_i |W_{ij}| + \) otherwise. For a purely excitatory profile \( c = 0 \), while for a balanced profile \( c \) is adjusted such that \( \sum_i W_{ij} = 0 \). For every choice of connectivity profile the amplitude \( A \) is calibrated so that the peak firing rate in the input and output layer of the network are the same (50Hz).
We estimated the position of the stimulus using an optimal linear discriminator, which provides a lower bound on the Fisher information. On each trial, spike counts (25 ms time window) were used as the input to the discriminator. On half the trials the stimulus had an angle $+\Delta \phi$ ($T^+$), while on the other trial the angle was $-\Delta \phi$. A linear discriminator has to distinguish these two stimuli by calculating $\phi L(\pm \Delta \phi)$, where $\phi$ are the weights of the discriminator. Because of symmetry, the discriminator weights can be written as a discrete Fourier Series $v_l = \sum_{-K}^{K} a_k \sin(2\pi lk/N)$. The discriminator was optimized by numerically minimizing the SNR with respect to the Fourier coefficients $a_k$ (8 Fourier coefficients proved sufficient). This optimization was done independently for every setting of the connectivity. The signal to noise ratio of the discriminator is used to calculate the standard deviation in the angle estimate, which equals the inverse square root of $J$ [20, 9].

**Appendix: Optimal Connectivity Profile**

The optimal connectivity profile is found by maximizing Eq. 2, subject to the constraint $\sum_n |\bar{W}(n)|^2 = q$. Define the Lagrangian

$$L = \frac{1}{2} \sum_n \frac{J(n)|\bar{W}(n)|^2}{|W(n)|^2 + T(n)} - \lambda \left( \sum_n |\bar{W}(n)|^2 - q \right)$$

where $\lambda$ is the Lagrange multiplier. Imposing $\partial L/\partial |\bar{W}(n)| = 0$ results in

$$|\bar{W}(n)|^2 = \frac{\sqrt{T(n)} \sqrt{J(n)}/\lambda - \sqrt{T(n)}}{\lambda}$$

This new value is used to calculate the updated $|\bar{W}(n)|^2$, and this procedure is repeated until $\lambda$ no longer changes at which point $\sum_n |\bar{W}(n)|^2 = q$. Substituting this final value for $\lambda$ into Eq. 5 results in the solution for the optimal weights in Eq. 3. The final weight matrix is retrieved by transforming back into real space. Note that, formally, our solution only specifies the squared magnitude of the weights, i.e., their ‘power spectrum’. Thus, any solution obtained by multiplying the $|\bar{W}(n)|$ by any phase $e^{i\theta_n}$ is also a solution. However, because we require the weight profile to be an even function, the Fourier coefficients can only be real.

**Acknowledgments.** We would like to thank Larry Abbott, David Barber, Mike Oram, Peggy Series, Gina Turrigiano, and Si Wu for discussions. AAR was supported by the EPSRC and HFSP. AR was supported by a start-up Ramon y Cajal grant from the Spanish Ministry of Education. AR would like to thank X.J. Wang and K.D. Harris for support during the initial and final stages of this project.

**Figure Legends.** Figure 1. The network architecture and a schematic of the procedure to find the weight that optimally preserve information. A, Architecture of the two-layer network. The stimulus orientation (vertical arrow) is encoded in the population activity of the input layer (top). Neurons in the output layer (bottom) pool the activity of the input layer through a set of frequency modes such that $|\bar{W}(n)|^2 > 0$. The rectification appears because $|\bar{W}(n)|^2 > 0$, and $\lambda$ has to be chosen so that the constraint is satisfied. We solve these equations self-consistently: given an initial arbitrary value of $\lambda$, we calculate $|\bar{W}(n)|^2$ using Eq. 5. Let us denote by $\{n'\}$ those frequency modes such that $|\bar{W}(n)|^2 > 0$. Imposing the constraint we obtain

$$\frac{1}{\lambda} = \frac{q + \sum_{n'} T(n')}{\sum_{n'} |\bar{W}(n')|^2 T(n')}$$

The rectification appears because $|\bar{W}(n)|^2 > 0$, and $\lambda$ has to be chosen so that the constraint is satisfied. We solve these equations self-consistently: given an initial arbitrary value of $\lambda$, we calculate $|\bar{W}(n)|^2$ using Eq. 5. Let us denote by $\{n'\}$ those frequency modes such that $|\bar{W}(n)|^2 > 0$. Imposing the constraint we obtain

$$\frac{1}{\lambda} = \frac{q + \sum_{n'} T(n')}{\sum_{n'} |\bar{W}(n')|^2 T(n')}$$

This new value is used to calculate the updated $|\bar{W}(n)|^2$, and this procedure is repeated until $\lambda$ no longer changes at which point $\sum_n |\bar{W}(n)|^2 = q$. Substituting this final value for $\lambda$ into Eq. 5 results in the solution for the optimal weights in Eq. 3. The final weight matrix is retrieved by transforming back into real space. Note that, formally, our solution only specifies the squared magnitude of the weights, i.e., their ‘power spectrum’. Thus, any solution obtained by multiplying the $|\bar{W}(n)|$ by any phase $e^{i\theta_n}$ is also a solution. However, because we require the weight profile to be an even function, the Fourier coefficients can only be real.

**Acknowledgments.** We would like to thank Larry Abbott, David Barber, Mike Oram, Peggy Series, Gina Turrigiano, and Si Wu for discussions. AAR was supported by the EPSRC and HFSP. AR was supported by a start-up Ramon y Cajal grant from the Spanish Ministry of Education. AR would like to thank X.J. Wang and K.D. Harris for support during the initial and final stages of this project.

**Figure Legends.** Figure 1. The network architecture and a schematic of the procedure to find the weight that optimally preserve information. A, Architecture of the two-layer network. The stimulus orientation (vertical arrow) is encoded in the population activity of the input layer (top). Neurons in the output layer (bottom) pool the activity of the input layer through a set of frequency modes such that $|\bar{W}(n)|^2 > 0$. The rectification appears because $|\bar{W}(n)|^2 > 0$, and $\lambda$ has to be chosen so that the constraint is satisfied. We solve these equations self-consistently: given an initial arbitrary value of $\lambda$, we calculate $|\bar{W}(n)|^2$ using Eq. 5. Let us denote by $\{n'\}$ those frequency modes such that $|\bar{W}(n)|^2 > 0$. Imposing the constraint we obtain

$$\frac{1}{\lambda} = \frac{q + \sum_{n'} T(n')}{\sum_{n'} |\bar{W}(n')|^2 T(n')}$$

This new value is used to calculate the updated $|\bar{W}(n)|^2$, and this procedure is repeated until $\lambda$ no longer changes at which point $\sum_n |\bar{W}(n)|^2 = q$. Substituting this final value for $\lambda$ into Eq. 5 results in the solution for the optimal weights in Eq. 3. The final weight matrix is retrieved by transforming back into real space. Note that, formally, our solution only specifies the squared magnitude of the weights, i.e., their ‘power spectrum’. Thus, any solution obtained by multiplying the $|\bar{W}(n)|$ by any phase $e^{i\theta_n}$ is also a solution. However, because we require the weight profile to be an even function, the Fourier coefficients can only be real.

**Acknowledgments.** We would like to thank Larry Abbott, David Barber, Mike Oram, Peggy Series, Gina Turrigiano, and Si Wu for discussions. AAR was supported by the EPSRC and HFSP. AR was supported by a start-up Ramon y Cajal grant from the Spanish Ministry of Education. AR would like to thank X.J. Wang and K.D. Harris for support during the initial and final stages of this project.
weights (middle) and output synaptic current (bottom) for three different tuning curve widths: narrow \((a = 1/6, \text{left})\), medium \((a = 1/3, \text{center})\) and wide tuning \((a = 1, \text{right})\). Other parameters: \(q = 2, \ c_{0,1} = 0, \ \sigma_{0,1} = 2 \text{ Hz}^2, \ r_p = 20 \text{ Hz}, \ N = 501. \) B, Output width (half-width at half-height) of optimal synaptic weights (solid) and output currents (dashed) as a function of tuning width in the input layer. Output and input widths are approximately matched. C-F, The effect of noise correlations is to sharpen the optimal weights (parameters as in A, \(a = 0.85\)). Three cases were considered: No noise correlations (black), noise correlation in the input layer only and uncorrelated external noise (light gray, empty), and identical correlation structure in input layer and external noise (dark gray, filled). The correlations, when present, are given by \(c_{0,1} = 0.2\) and \(\rho_{0,1} = 1.\)

C, Tuning curve in the input layer (solid) and correlation structure (dotted). D, Correlations strongly reduce the Fisher Information in the input layer (dark and light gray, overlapping), compared to the uncorrelated case (black). Inset, Effective threshold for the synaptic weights at each frequency \(\tilde{C}_1(n)/\tilde{C}_0(n),\) see Equation 2; dark gray and black, overlapping. E, Fourier representation of the optimal weights. In the absence of correlations the weights amplify only the most informative, low frequency modes (black). The effective threshold across frequencies is the same when the input and external correlations are equal, as in the case where correlations are absent. But since correlations decrease the information at the lowest frequency modes (see panel D), more weight is devoted to higher frequencies (dark gray). Finally, when the external noise is uncorrelated, the low frequency thresholds decrease (D, inset). The information in the lowest frequencies can be now transmitted with smaller weights, allowing further emphasis of higher frequencies (light gray). F, After inverse Fourier transformation, the optimal weights are narrower when correlations are present.

Figure 3. Center surround connectivity of width matched to the tuning curve is optimal in networks of spiking integrate-and-fire neurons. A, Snapshot of one second of the activity in the input (top) and output (bottom) layers. B, Error of the optimal linear estimator (see Supplementary Information) of the stimulus as a function of the width of the connectivity profile based spike count in 25 ms bins. The error in the estimator in the input layer is shown by the gray line. Without inhibition estimation is suboptimal for all widths (dotted). With balanced inhibition (solid) the optimal width is similar to, but slightly larger than, the tuning curve width (arrow). Stronger inhibition (dashed) is also optimal if matched to the tuning curve width, but the error is larger when the profile is mis-tuned than in the case of balanced inhibition. C, Examples of the input and output 'bumps' of activity for three different situations: Narrow connectivity with strong inhibition (i), balanced inhibition at the optimal width (ii) and wide connectivity with no inhibition (iii). Input and output activity profiles are similar in the optimal case (using 200 ms bins for clarity).