Optimal population coding by noisy spiking neurons

Gašper Tkačik[a,2], Jason S. Prentice[a], Vijay Balasubramanian[a,1], and Elad Schneidman[b,1]

[a]Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104; and [b]Department of Neurobiology, Weizmann Institute of Science, 76100 Rehovot, Israel

Edited* by Curtis G. Callan, Princeton University, Princeton, NJ, and approved June 30, 2010 (received for review April 10, 2010)

In retina and in cortical slice the collective response of spiking neural populations is well described by "maximum-entropy" models in which only pairs of neurons interact. We asked, how should such interactions be organized to maximize the amount of information represented in population responses? To this end, we extended the linear-nonlinear-Poisson model of single neural response to include pairwise interactions, yielding a stimulus-dependent, pairwise maximum-entropy model. We found that as we varied the noise level in single neurons and the distribution of network inputs, the optimal pairwise interactions smoothly interpolated to achieve network functions that are usually regarded as discrete—stimulus decorrelation, error correction, and independent encoding. These functions reflected a trade-off between efficient consumption of finite neural bandwidth and the use of redundancy to mitigate noise. Spontaneous activity in the optimal network reflected stimulus-induced activity patterns, and single-neuron response variability overestimated network noise. Our analysis suggests that rather than having a single coding principle hardwired in their architecture, networks in the brain should adapt their function to changing noise and stimulus correlations.

adaptation | neural networks | Ising model | attractor states

Populations of sensory neurons encode information about stimuli into sequences of action potentials, or spikes (1). Experiments with pairs or small groups of neurons have observed many different coding strategies (2–6): (*i*) independence, where each neuron responds independently to the stimulus, (*ii*) decorrelation, where neurons interact to give a decorrelated representation of the stimulus, (*iii*) error correction, where neurons respond redundantly, in patterns, to combat noise, and (*iv*) synergistic coding, where population activity patterns carry information unavailable from separate neurons.

How *should* a network arrange its interactions to best represent an ensemble of stimuli? Theoretically, there has been controversy over what is the "correct" design principle for neural population codes (7–11). On the one hand, neurons have a limited repertoire of response patterns, and information is maximized by using each neuron to represent a different aspect of the stimulus. To achieve this, interactions in a network should be organized to remove correlations in network inputs and thus create a decorrelated network response. On the other hand, neurons are noisy, and noise is combatted via redundancy, where different patterns related by noise encode the same stimulus. To achieve this, interactions in a network should be organized to exploit existing correlations in neural inputs to compensate for noise-induced errors. Such a trade-off between decorrelation and noise reduction possibly accounts for the organization of several biological information processing systems, e.g., the adaptation of center-surround receptive fields to ambient light intensity (12–14), the structure of retinal ganglion cell mosaics (15–18), and the genetic regulatory network in a developing fruit fly (19, 20). In engineered systems, compression (to decorrelate incoming data stream), followed by reintroduction of error-correcting redundancy, is an established way of building efficient codes (21).

Here we study optimal coding by networks of noisy neurons with an architecture experimentally observed in retina, cortical culture, and cortical slice—i.e., pairwise functional interactions between cells that give rise to a joint response distribution resembling the "Ising model" of statistical physics (6, 22–26). We extended such models to make them stimulus dependent, thus constructing a simple model of stimulus-driven, pairwise-interacting, noisy, spiking neurons. When the interactions are weak, our model reduces to a set of conventional linear-nonlinear neurons, which are conditionally independent given the stimulus. We asked how internal connectivity within such a network should be tuned to the statistical structure of inputs, given noise in the system, in order to maximize represented information.

We found that as noise and stimulus correlations varied, an optimal pairwise-coupled network should choose continuously among independent coding, stimulus decorrelation, and redundant error correction, instead of having a single universal coding principle hardwired in the network architecture. In the high-noise regime, the resulting optimal codes have a rich structure organized around "attractor patterns," reminiscent of memories in a Hopfield network. The optimal code has the property that decoding can be achieved by observing a subset of the active neural population. As a corollary, noise measured in responses of single neurons can significantly overestimate network noise, by ignoring error-correcting redundancy. Our results suggest that networks in the brain should adapt their encoding strategies as stimulus correlations or noise levels change.

Ising Models for Networks of Neurons

In the analysis of experimental data from simultaneously recorded neurons, one discretizes spike trains $\sigma_i(t)$ for $i = 1,\ldots,N$ neurons into small time bins of duration Δt. Then $\sigma_i(t) = 1$ indicates that the neuron i has fired in time bin t, and $\sigma_i(t) = -1$ indicates silence. To describe network activity, we must consider the joint probability distribution over N-bit binary responses of the neurons, $\hat{P}(\{\sigma_i\})$, over the course of an experiment. Specifying a general distribution requires an exponential number of parameters, but for retina, cortical culture, and cortical slice, $\hat{P}(\{\sigma_i\})$ is well-approximated by the minimal model that accounts for the observed mean firing rates and covariances (6, 22–26). This minimal model is a *maximum-entropy* distribution (27) and can be written in the Ising form

$$\hat{P}(\{\sigma_i\}) = \frac{1}{Z(\{h_i, J_{ij}\})} \exp\left[\sum_i h_i\sigma_i + \frac{1}{2}\sum_{ij} J_{ij}\sigma_i\sigma_j\right]. \qquad [1]$$

Here the h_i describe intrinsic biases for neurons to fire, and $J_{ij} = J_{ji}$ are pairwise interaction terms, describing the effect of neuron i on neuron j and vice versa. We emphasize that the J_{ij} model functional dependencies, not physical connections. The denominator Z, or *partition function*, normalizes the probability distribution. The

Author contributions: G.T., J.S.P., V.B., and E.S. designed research; G.T., J.S.P., V.B., and E.S. performed research; G.T., J.S.P., V.B., and E.S. analyzed data; and G.T., V.B., and E.S. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]V.B. and E.S. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: gtkacik@sas.upenn.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1004906107/-/DCSupplemental.

model can be fit to data by finding couplings $\mathbf{g} = \{h_i, J_{ij}\}$, for which the mean firing rates $\langle\sigma_i\rangle$ and covariances $C_{ij} = \langle\sigma_i\sigma_j\rangle - \langle\sigma_i\rangle\langle\sigma_j\rangle$ over $\hat{P}(\{\sigma_i\})$ match the measured values (6, 11, 23, 24).

This Ising-like model can be extended to incorporate the stimulus (s) dependence of neural responses by making the model parameters depend on s. We considered models where only the firing biases h_i depend on s:

$$P(\{\sigma_i\}|s) = \frac{\exp\left\{\beta\left(\sum_i (h_i^0 + h_i(s))\sigma_i + \frac{1}{2}\sum_{ij} J_{ij}\sigma_i\sigma_j\right)\right\}}{Z(\{h_i, J_{ij}\})}. \quad [2]$$

Here h_i^0 is a constant (stimulus-independent) firing bias, and $\vec{h}(s) \equiv \{h_i(s)\}$ is a stimulus-dependent firing bias. The parameter β, which we call "neural reliability," is reminiscent of the inverse temperature in statistical physics and reflects the signal-to-noise ratio in the model (9). Here, noise might arise from ion channel noise, unreliable synaptic transmission, and influences from unobserved parts of the network. As $\beta \to \infty$, neurons become deterministic and spike whenever the quantities $(h_i + h_i^0 + \sum_j J_{ij}\sigma_j)$ are positive and are silent otherwise. As $\beta \to 0$, neurons are completely noisy and respond randomly to inputs. Thus, β parametrizes the reliability of neurons in the model—larger β leads to more reliable responses, and lower β leads to less reliable, noisier responses.

The stimuli s are drawn from a distribution $P_s(s)$, which defines the stimulus ensemble. Our analysis will investigate how J_{ij} should vary with the statistics of the stimulus ensemble and neural reliability (β) in order to maximize information represented in neural responses. As such, J_{ij} will not depend on specific stimuli within an ensemble.

In the absence of pairwise couplings ($J_{ij} = 0$), the model describes stimulus-driven neural responses that are conditionally independent given the stimulus:

$$P(\{\sigma_i\}|s) = Z^{-1}\prod_i \exp[\beta(h_i^0 + h_i(s))\sigma_i], \quad [3]$$

$$\langle\sigma_i(s)\rangle = \tanh[\beta(h_i^0 + h_i(s))]. \quad [4]$$

Then, writing the stimulus-dependent drive $h(s)$ as a convolution of a stimulus sequence $s(t)$ with a linear filter (e.g., a kernel obtained using reverse correlation), Eq. **4** describes a conventional linear-nonlinear (LN) model for independent neurons with saturating nonlinearities given by tanh functions (shaped similarly to sigmoids). The bias of neurons is controlled by h_i^0, and the steepness of the nonlinearity by β. Thus, our model (Eq. **2**) can be regarded as the simplest extension of the classic LN model of neural response to pairwise interactions.

We will regard a given environment as being characterized by a stationary stimulus distribution $P_s(s)$. In our model, the stimulus makes its way into neuronal responses via the bias toward firing $h_i(s)$. Thus, for our purposes, a fixed environment can equally be characterized by the distribution of h_i, $P_h(\vec{h})$, implied by the distribution over s. So we will use the distribution $P_h(\vec{h})$ to characterize the stimulus ensemble from a fixed environment. The correlations in $P_h(\vec{h})$ can arise both from correlations in the external stimulus (s) as well as inputs shared between neurons in our network (28). We will show that given such a stimulus ensemble, and neural reliability characterized by β, information represented in network responses is maximized when the couplings $\mathbf{g} = \{h_i^0, J_{ij}\}$ are appropriately adapted to the stimulus statistics. In this way, the couplings effectively serve as an "internal representation" or "memory" of the environment, allowing the network to adjust its encoding strategy.

Maximizing Represented Information

Let N neurons probabilistically encode information about stimuli $\vec{h}$ in responses $\{\sigma_i\}$ distributed as Eq. **2** (see Fig. 1). The amount of information about $\vec{h}$ encoded in $\{\sigma_i\}$ is measured by the mutual information (29):

$$I(\{\sigma_i\};\vec{h}) = \int d\vec{h} P_h(\vec{h}) \sum_{\{\sigma_i\}} P(\{\sigma_i\}|\vec{h}) \log_2 \frac{P(\{\sigma_i\}|\vec{h})}{P(\{\sigma_i\})}, \quad [5]$$

where the conditional distribution of responses $P(\{\sigma_i\}|\vec{h})$ is given by Eq. **2** and the distribution of responses, $P(\{\sigma_i\})$, is given by $P(\{\sigma_i\}) = \int d\vec{h} P(\{\sigma_i\}|\vec{h}) P_h(\vec{h})$. This mutual information is an upper bound to how much downstream layers, receiving binary words $\{\sigma_i\}$, can learn about the world (1). Because of noise and ineffective decoding by neural "hardware," the actual amount of information used to guide behavior can be smaller, but not bigger, than Eq. **5**.

Eq. **5** is commonly rewritten as a difference between the entropy of the distribution over all patterns (sometimes called "output entropy") and the average entropy of the conditional distributions (sometimes called the "noise entropy"):

$$I(\{\sigma_i\};\vec{h}) = S[P(\{\sigma_i\})] - \langle S[P(\{\sigma_i\}|\vec{h})]\rangle_{P(\vec{h})}, \quad [6]$$

where the entropy of a distribution $S[P(x)] = -\sum_x P(x)\log_2 P(x)$ measures uncertainty about the value of x in bits.

If neurons spiked deterministically ($\beta \to \infty$), the noise entropy in Eq. **6** would be zero, and maximizing mutual information between inputs and outputs would amount to maximizing the output entropy $S[P(\{\sigma_i\})]$. This special case of information maximization without noise is equivalent to all-order decorrelation of the outputs. It has been used for continuous transformations by Linsker (30) and Bell and Sejnowski (31), among others, to describe independent component analysis (ICA) as a general formulation for blind source separation and deconvolution. In contrast, here we examine a scenario where noise in the neural code cannot be neglected. In this setting, redundancy can serve a useful function in combating uncertainty due to noise (10). As we will see, information-maximizing networks in our scenario use interactions between neurons to minimize the effects of noise, at the cost of reducing the output entropy of the population.

Our problem can thus be compactly stated as follows. Given the distribution of inputs, $P_h(\vec{h})$, and the neural reliability β, find the parameters $\mathbf{g} = \{h_i^0 J_{ij}\}$ such that the mutual information $I(\{\sigma_i\};\vec{h})$ between the inputs and the binary output words is maximized.

Results

Two Coupled Neurons. We start with the simple case of two neurons, responding to inputs $\vec{h} = (h_1, h_2)$ drawn from two different

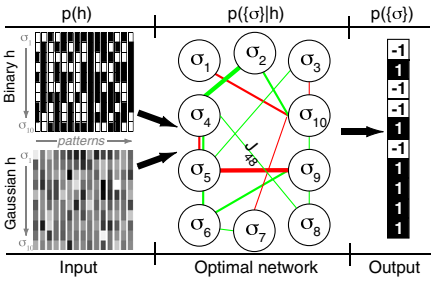

Fig. 1. Schematic diagram of information transmission by a spiking noisy neural population. Stimuli are drawn from one of two distributions $P_h(\vec{h})$: (*i*) "binary" distributions, where the input to each neuron is one of the $\mu = 1,\ldots,K$ patterns of $\{\pm 1\}$, i.e., $h_i^\mu = \pm 1$; (*ii*) Gaussian distributions with given covariance matrices. These h_i drive a neural response, parameterized by $\mathbf{g} = \{h_i^0, J_{ij}\}$ and neural reliability β, in Eq. **2**. Positive (negative) couplings between neurons J_{ij} are schematically represented as green (red) links, with thickness indicating interaction strength. Each input drawn from $P_h(\vec{h})$ is probabilistically mapped to binary words at the output, $\{\sigma_i\}$, allowing us to define the mutual information $I(\{\sigma_i\};\vec{h})$ in Eq. **5** and maximize it with respect to **g**.

distributions $P_h(\vec{h})$. The first is the *binary* distribution, where $h_{1,2}$ take one of two equally likely discrete values (± 1), with a covariance $\mathrm{Cov}(h_1,h_2) = \alpha$ (useful when the biological correlate of the input is the spiking of upstream neurons). In this case $P_h(-1,-1) = P_h(1,1) = (1+\alpha)/4$ and $P_h(-1,1) = P_h(1,-1) = (1-\alpha)/4$.

The second is a *Gaussian* distribution, where inputs take a continuum of values (useful when the input is a convolution of a stimulus with a receptive field). In this case, we also take the means to vanish ($\langle h_1 \rangle = \langle h_2 \rangle = 0$), unit standard deviations ($\sigma_{h_1} = \sigma_{h_2} = 1$), and covariance $\mathrm{Cov}(h_1,h_2) = \mathrm{Cov}(\vec{h}) = \alpha$. In both cases, α measures *input correlation* and ranges from −1 (perfectly anticorrelated) to 1 (perfectly correlated). We asked what interaction strength J between the two neurons (Fig. 2*A* and Eq. **2**) would maximize information, as the correlation in the input ensemble (parameterized by α) and the reliability of neurons (parameterized by β) were varied.

For the binary input distribution, the mutual information of Eq. **5** can be computed exactly as a function J, α, and β (see *SI Appendix*), and the optimal coupling $J^*(\alpha,\beta)$ is obtained by maximizing this quantity for each α and β (Fig. 2*B*). When β is small, the optimal coupling takes the same sign as the input covariance. In this case, interactions between the two neurons enhance the correlation present in the stimulus. The resulting redundancy helps counteract loss of information to noise. As reliability (β) increases, the optimal coupling J^* decreases in magnitude as compared to the input strength $|\vec{h}|$ (see *Discussion*). This is because, in the absence of noise, a pair of binary neurons has the capacity to carry complete information about a pair of binary inputs. Thus, in the noise-free limit the neurons should act as independent encoders ($J^* = 0$) of binary inputs.

For a Gaussian distribution of inputs, we maximized the mutual information in Eq. **5** numerically (Fig. 2 *C* and *D*). For small β, the optimal coupling J^* has the same sign as the input correlation, as in the binary input case, thus enhancing input correlations and using redundancy to counteract noise. However, for large β, the optimal coupling has a sign *opposite* to the input correlation.

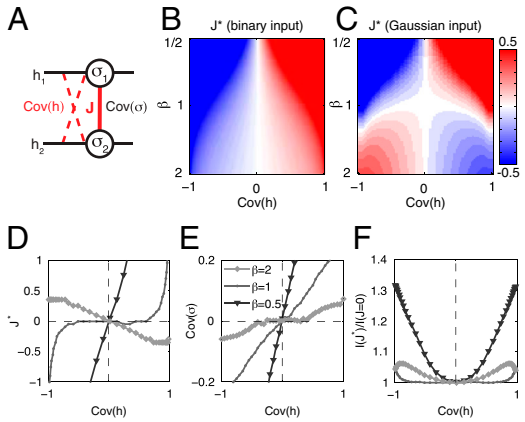

Fig. 2. Information transmission in a network of two neurons. (*A*) Schematic of a two-neuron network, $\{\sigma_1,\sigma_2\}$, coupled with strength J, receiving correlated binary or Gaussian inputs. $\alpha = \mathrm{Cov}(h)$ = input correlation; $\mathrm{Cov}(\sigma) = \langle \sigma_1\sigma_2 \rangle - \langle \sigma_1 \rangle \langle \sigma_2 \rangle$ = correlation between output spike trains. (*B*) Optimal J^* as a function of input correlation, $\mathrm{Cov}(h)$, and neural reliability β for binary inputs. (*C*) Optimal J^* as a function of input correlation and neural reliability for Gaussian inputs. (*D*) J^* as a function of input correlation for three values of reliability ($\beta = 0.5$, 1, 2, grayscale) and Gaussian inputs; these are three horizontal sections through the diagram in *C*. At high reliability the optimal J^* has an opposite sign to the input correlation; at low reliability it has the same sign. (*E*) Output correlation as a function of input correlation and reliability for Gaussian inputs. At high reliability ($\beta = 2$) the network decorrelates the inputs. At low reliability ($\beta = 1/2$) the input correlation is enhanced. (*F*) Fractional improvement in information transmission in optimal (J^*) vs. uncoupled ($J = 0$) networks.

Thus the neural output decorrelates its inputs (Fig. 2*E*). This occurs because binary neurons do not have the capacity to encode all the information in continuous inputs. Therefore, in the absence of noise, the best strategy is to decorrelate inputs to avoid redundant encoding of information. The crossover in strategies is at $\beta \sim 1$ and is driven by the balance of output and noise entropies in Eq. **6**, as shown in Fig. S1. In all regimes more information is conveyed with the optimal coupling (J^*) than by an independent ($J = 0$) network. The information gain produced by this interaction is larger for strongly correlated inputs (Fig. 2*F*).

For both binary and Gaussian stimulus ensembles, the biases toward firing (h_i^0) in the optimal network adjusted themselves so that individual neurons were active about half of the time (see *SI Appendix*). Adding a constraint on the mean firing rates would shift the values of h_i^0 in the optimal network, but would leave the results for the optimal coupling J^* qualitatively unchanged.

Thus, information represented by a pair of neurons is maximized if their interaction is adjusted to implement different functions (independence, decorrelation to remove redundancy, and averaging to reduce noise) depending on the input distribution and neural reliability.

Networks of Neurons. We then asked what would be the optimal interaction network for larger populations of neurons. First, we considered a network of N neurons responding to an input ensemble of K equiprobable N-bit binary patterns chosen randomly from the set of 2^N such patterns. For $N \lesssim 10$ it remained possible to numerically choose couplings h_i^0 and J_{ij} that maximized information about the input ensemble represented in network responses. We found qualitatively similar results to two neurons responding to a binary stimulus: For unreliable neurons (low β), the optimal network interactions matched the sign of input correlations, and for reliable neurons (high β), neurons became independent encoders. Input decorrelation was never an optimal strategy, and the capacity of the network to yield substantial improvements in information transmission was greatest when $K \sim N$ (see *SI Appendix*). Our results suggest that decorrelation will never appear as an optimal strategy if the input entropy is less than or equal to the maximum output entropy.

We then examined the optimal network encoding correlated Gaussian inputs drawn from a distribution with zero mean and a fixed covariance matrix. The covariance matrix was chosen at random from an ensemble of symmetric matrices with exponentially distributed eigenvalues (*SI Appendix*). As for the case of binary inputs, we numerically searched the space of **g** for a choice maximizing the information for $N = 10$ neurons and different values of neural reliability β. As β is changed, the optimal (J^*) and uncoupled ($J = 0$) networks behave very differently. In the uncoupled case (Fig. 3*A*), decreasing β increases both the output and noise entropies monotonically. In the optimal case (Fig. 3*B*), the noise entropy can be kept constant and low by the correct choice of couplings J^*, at the expense of losing some output entropy. The difference of these two entropies is the information, plotted in Fig. 3*C*. At low neural reliability β, the total information transmitted is low, but substantial relative increases (almost twofold) are possible by the optimal choice of couplings. The optimal couplings are positively correlated with their inputs, generating a redundant code to reduce the impact of noise (Fig. 3*D*). At high β, the total information transmitted is high, and optimal couplings yield smaller, but still significant, relative improvements (~10%). The couplings in this case are anticorrelated with the inputs, and the network performs input decorrelation.

For unreliable neurons our results give evidence that the network uses redundant coding to compensate for errors. But theoretically there are many different kinds of redundant error-correcting codes—e.g., codes with checksums vs. codes that distribute information widely over a population. Thus we sought

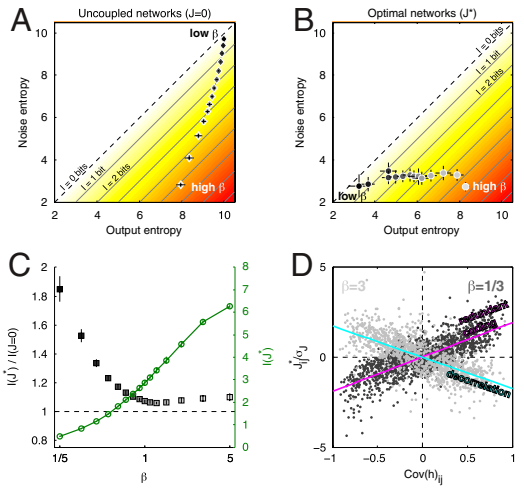

Fig. 3. Information transmission in networks with $N = 10$ neurons and Gaussian inputs. (*A* and *B*) Output and noise entropies in bits for uncoupled ($J = 0$) and optimal (J^*) networks, shown parametrically as neural reliability (β) changes from high ($\beta = 5$, bright symbols) to low ($\beta = 1/5$, dark symbols). Transmitted information I is the difference between output and noise entropies and is shown by the color gradient. Networks with low reliability transmit less information and thus lie close to the diagonal, while networks that achieve high information rates lie close to the lower right corner. The optimal network uses its couplings J^* to maintain a consistently low network noise entropy despite a 10-fold variation in neural reliability. Error bars are computed over 30 replicate optimizations for each β. (*C*) The information transmitted in optimal networks (green, right axis) in bits, and the relative increase in information with optimal vs. uncoupled networks (black, left axis), as a function of β. For low reliability β, large relative increases in information are possible with optimal coupling, but even at high β, where the baseline capacity is higher, the relative increase of ∼10% is statistically significant. (*D*) Scatter plot of optimal couplings J^*_{ij} against the correlation in the corresponding inputs, plotted for two values of β (light symbols, blue line, high β; dark symbols, magenta line, low β). Each point represents a single element of the input covariance matrix, $[\mathrm{Cov}(\vec{h})]_{ij}$, plotted against the corresponding coupling matrix element J^*_{ij} that has been normalized by the overall scale of the coupling matrix, $\sigma_J = \mathrm{std}(J^*)$. Results for 30 optimization runs are plotted in overlay for each β. At low β, the optimal couplings are positively correlated with the inputs, indicating that the network is implementing redundant coding, whereas for high β, the anticorrelation indicates that the network is decorrelating the inputs.

to characterize more precisely the structure of our optimal network codes.

The Structure of the Optimal Code, Ongoing Activity, and the Emergence of Metastable States. How does the optimal network match its code to the stimulus ensemble? Intuitively, the optimal network has "learned" something about its inputs by adjusting the couplings. Without an input, a signature of this learning should appear in correlated *spontaneous* network activity. Fig. 4 *A* and *B Top* shows the distributions of ongoing, stimulus-free activity patterns, $P(\{\sigma_i\}|h = 0)$, of the noninteracting network ($J = 0$) and those of a network that is optimally matched to stimuli (J^*). While the activity of the $J = 0$ network is uniform over all patterns, the ongoing activity of the optimized network echoes the responses to stimuli.

To make this intuition precise and interpret the structure of the optimal code, it is useful to carefully examine the coding patterns in the stimulus-free condition. We find that the ability of the optimal network to adjust the couplings J^*_{ij} to the stimulus ensemble makes certain response patterns a priori much more likely than others. Specifically, the couplings generate a probability landscape over response patterns that can be partitioned into basins of attraction (see *SI Appendix*). The basins are organized around patterns with locally maximal likelihood (ML). For these ML patterns, $\mathcal{G}_\mu = \{\sigma_i\}_\mu$, flipping any of the neurons (from spiking to silence or vice versa) results in a less likely pattern. For all

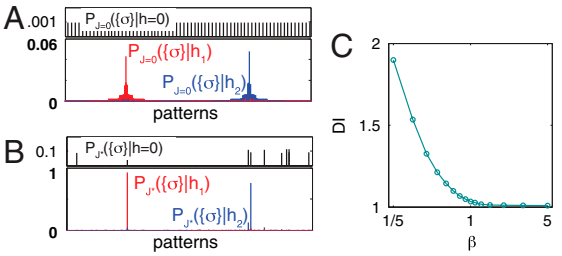

Fig. 4. Coding patterns in a network of $N = 10$ neurons exposed to Gaussian stimuli. (*A*) Uncoupled ($J = 0$) network. (*Top*) The probability distribution over response patterns $\{\sigma_i\}$ in the absence of a stimulus. Because $J = 0$, this probability is uniform (for clarity, not all 1,024 values are individually shown). (*Bottom*) The response distributions $P(\{\sigma_i\}|\vec{h})$, for two individual stimuli, red and blue; patterns on the x axis have been reordered by their $P(\{\sigma_i\}|\vec{h})$ to cluster around the red (blue) peak. The distributions have some spread and overlap—the response has significant variability. (*B*) Optimally coupled (J^*) network that has been tuned to the input distribution. Patterns ordered as in *A*. (*Top*) Because $J^* \neq 0$, the prior probability over the patterns is not uniform. The most probable patterns have similar likelihoods. Here, the network has "learned" the stimulus prior and has memorized it in the couplings J^* (see text). (*Bottom*) When either the blue or red stimulus is applied, the probability distribution collapses completely onto one of the two coding patterns that have a high prior likelihood. The sharp response leads to higher information transmission. (*C*) "Discriminability index" (DI) measures the separability of responses to pairs of inputs in an optimal vs. uncoupled network. To measure separability of responses to distinct inputs, we first compute the average Jensen–Shannon (JS) distance between response probabilities, $D = \langle D_{\mathrm{JS}}[P(\vec{\sigma}|\vec{h}_i)|P(\vec{\sigma}|\vec{h}_j)]\rangle_{ij}$, across pairs of inputs $\vec{h}_{i,j}$ drawn independently from $P_h(\vec{h})$. Discriminability index is $\mathrm{DI} = D(J^*)/D(J = 0)$, i.e., the ratio of the average response distance in optimal vs. uncoupled networks.

other patterns within the same basin, their neurons can be flipped such that successively more likely patterns are generated, until the corresponding ML pattern is reached.

In optimal networks, when no stimulus is applied, the ML patterns have likelihoods of comparable magnitude, but when a particular input $\vec{h}$ is chosen, it will bias the prior probability landscape, making one of these ML patterns the most likely response (Fig. 4*B Bottom*). This maps similar stimuli into the same ML basin, while increasing the separation between responses coding for very different stimuli. Overall this improves information transmission. We used the Jensen–Shannon distance to quantify discriminability of responses in an optimal network, compared to the uncoupled ($J = 0$) network, as a function of neural reliability β (Fig. 4*C*).* For high reliability, the independent and optimized networks had similarly separable responses, whereas at low reliability, the responses of the optimized network were much more discriminable from each other.

The appearance of ML patterns is reminiscent of the storage of memories in dynamical "basins of attraction" for the activity in a Hopfield network (32) (for a detailed comparison, see *SI Appendix*). We therefore considered the hypothesis that in the optimal network a given stimulus could be encoded not only by the ML pattern itself, but redundantly by all the patterns within a basin surrounding this ML pattern. Since ML patterns are local likelihood maxima, the noise alone is unlikely to induce a spontaneous transition from one basin to the next, making the basins of attraction potentially useful as stable and reliable representations of the stimuli.

To check this hypothesis, we quantified how much information about the stimulus was carried by the identity of the basins surrounding the ML patterns, as opposed to the detailed activity patterns of the network (Fig. 5*A*). To do this, we first mapped each neural response $\{\sigma_i\}$ to its associated basin of attraction

*Given distributions p and q, let $m(\alpha) = (p(\alpha) + q(\alpha))/2$. The Jensen–Shannon distance is $D_{\mathrm{JS}} = 0.5\int d\alpha p(\alpha)\log_2[p(\alpha)/m(\alpha)] + 0.5\int d\alpha q(\alpha)\log_2[q(\alpha)/m(\alpha)]$. $D_{\mathrm{JS}} = 0$ for identical, and $D_{\mathrm{JS}} = 1$ for distinct p,q.

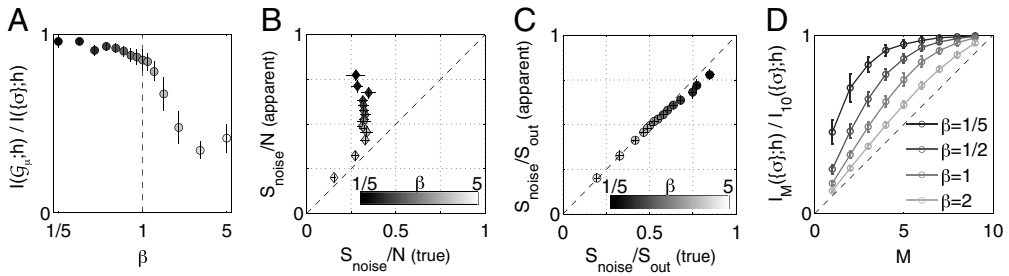

Fig. 5. The structure of the optimal code in a network of $N = 10$ neurons exposed to Gaussian stimuli. (*A*) Information about the stimulus carried by the identity of the maximum likelihood basin $\mathcal{G}_\mu$ (see text), divided by information carried by the full spiking pattern $\{\sigma_i\}$. At low reliability β, most of the information is contained in the identity of the basin; at high β, each neuron conveys some nonredundant information and the detailed activity pattern is important. (*B*) Apparent noise, estimated from single-neuron variability, overestimates noise in network responses. As β decreases (grayscale) single-neuron noise entropy increases monotonically, whereas network noise entropy saturates. (*C*) Nevertheless, in optimal networks the noise entropy as a fraction of output entropy can be estimated from single-neuron measurements. (*D*) Information carried, on average, by randomly chosen subsets of M neurons, normalized by information carried by the whole network of $N = 10$ neurons, as a function of β. The dashed line shows the case where each neuron carries $1/N$ of the total information. At low β, stimuli can be decoded from a small subset of responses.

indexed by the ML pattern $\mathcal{G}_\mu$ within it. In effect, this procedure "compresses" the response of neurons in the network to one number—the identity of the basin. Then we computed the mutual information between the identity of the response basins and the stimulus $I(\mathcal{G}_\mu;\vec{h})$. We found that at high neural reliability β, information is carried by the detailed structure of the response pattern, $\{\sigma_i\}$. But when neural reliability β is low, most of the information is already carried by the identity of the ML basin to which the network activity pattern belongs. This is a hallmark of error correction via the use of redundancy—at low β, all the coding patterns within a given ML basin indeed encode the same stimulus.

While noise in individual neurons will still result in response variability, at low β the optimal network uses its interactions to tighten the basin around the ML pattern within which a response is likely. Thus, noise in individual neural responses should overestimate network noise. To test this, we first measured network noise entropy per neuron: true $S_{\text{noise}}/N = \langle S[P(\{\sigma_i\}|\vec{h})]\rangle_{P(\vec{h})}/N$, which quantifies the variability in network responses to given stimuli. Then we measured apparent noise entropy per neuron: apparent $S_{\text{noise}}/N = (\sum_i \langle S[P(\sigma_i|\vec{h})]\rangle_{P(\vec{h})})/N$, which quantifies the average variability in individual neural responses to given stimuli. We found that apparent single-neuron noise could greatly overestimate network noise (Fig. 5*B*). Furthermore, as neural reliability β decreased, single-neuron noise entropy increased monotonically, whereas noise in the optimal network responses saturated. In contrast, the noise entropy as a fraction of the total output entropy was similar when measured on the population or the single-neuron level, regardless of the value of β (Fig. 5*C*). This surprising property of the optimal codes would therefore allow one to obtain an estimate of the coding efficiency of a complete optimal network from the average of many single-neuron coding efficiency measurements.

Finally, in an optimal network with unreliable neurons (low β) most of the network information can be read out by observing just a subset of the neurons (Fig. 5*D*). Meanwhile at high β, every neuron in a network of N neurons carries roughly $1/N$ of the total information, because in the high β regime the neural output tends toward independence. In no case did we find an optimal network with a synergistic code such that observing a subset of neurons would yield less-than-proportional fraction of the total information.

Discussion

Wiring neurons together in a network creates dependencies between responses and thus effectively reduces the repertoire of joint activity patterns with which neurons can encode their stimuli. This reduces the information the network can convey. On the other hand, connections between neurons can enable a population code to either mitigate the noisiness of each of the elements or decorrelate the network inputs. This would increase the information the network can convey. Here we studied the functional value of interactions by finding information-maximizing networks of pairwise-interacting, stimulus-dependent model neurons. We explored the coding properties of models containing three key features: (*i*) neurons are spiking (not continuous), (*ii*) neurons are noisy, and (*iii*) neurons can be functionally interacting, and recurrent connections can also be tuned to achieve optimal information transmission. We found that the optimal population code interpolates smoothly between redundant error-correcting codes, independent codes, and decorrelating codes, depending on the strength of stimulus correlations and neuronal reliability. In a related vein, other recent work has shown that efficient coding and discrimination of certain types of stimulus distributions favor nonzero interactions in a network (33, 34).

If neurons are unreliable (Fig. 6*A*), the optimal network "learns" the input distribution and uses this to perform error correction. This error correction is implemented in a distributed way, as opposed to using dedicated parity or check-bits that appear in engineered codes: The network "memorizes" different inputs using a set of patterns $\mathcal{G}_\mu$ that maximize the likelihood (ML) at zero input, $P_{J^*}(\{\sigma\}|\vec{h}=0)$. These ML memories are encoded in the optimal couplings $\{h_i^{0*} J_{ij}^*\}$. There are many such potential patterns, and the external input breaks the degeneracy among

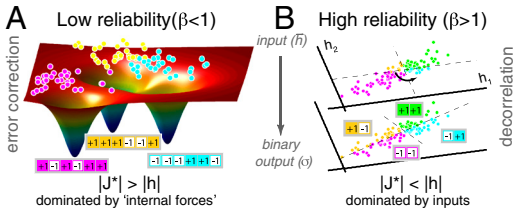

Fig. 6. A schematic representation of the function of optimal networks. (*A*) At low reliability ($\beta < 1$), optimal couplings J_{ij}^* give rise to a likelihood landscape of network patterns with multiple basins of attraction. The network maps different N-dimensional inputs $\{h_i\}$ (schematically shown as dots) into ML patterns $\mathcal{G}_\mu$—here, magenta, yellow, and cyan 6-bit words—that serve as stable representations of stimuli. The stimuli (dots) have been colored according to the $\mathcal{G}_\mu$ that they are mapped to. (*B*) At high reliability ($\beta > 1$), the network decorrelates Gaussian-distributed inputs. The plotted example has two neurons receiving correlated inputs $\{h_1,h_2\}$, shown as dots. For uncoupled networks (upper plane), the dashed lines partition the input space into four quadrants that show the possible responses of the two neurons: $\{-1,-1\}$ (magenta), $\{-1,1\}$ (cyan), $\{+1,+1\}$ (green), and $\{+1,-1\}$ (yellow). The decision boundaries must be parallel to the coordinate axes, resulting in underutilization of yellow and cyan, and overutilization of magenta and green. In contrast, optimally coupled networks (lower plane) rotate the decision axes, so that similar numbers of inputs are mapped to each quadrant. This increases the output entropy in Eq. **6** and increases the total information.

them by favoring one in particular. The information carried by just the identity of the basin around a ML pattern then approaches that carried by the microscopic state of the neurons, $I(\mathcal{G}_\mu;\vec{h}) \sim I(\{\sigma_i\};\vec{h})$. This mechanism is similar to one used by Hopfield networks, although in our case the memories, or ML patterns, emerge as a consequence of information maximization rather than being stored by hand into the coupling matrix (*SI Appendix*).

If neurons are reliable (Fig. 6*B*), the optimal network behavior and coding depend qualitatively on the distribution of inputs. For binary inputs, the single units simply become more independent encoders of information, and the performance of the optimal network does not differ much from that of the uncoupled network. In contrast, for Gaussian stimuli the optimal network starts decorrelating the inputs. The transition between the low- and the high-reliability regime happens close to $\beta \sim 1.2$. This represents the reliability level at which the spread in optimal couplings (standard deviation) is similar to the amplitude of the stimulus-dependent biases, h_i. Intuitively, this is the transition from a regime, in which the network is dominated by "internal forces" (low β, "couplings > inputs"), to a regime dominated by external inputs (high β, "inputs > couplings").

Independently of the noise, individually observed neurons in an optimal network appear to have more variability than expected from the noise entropy per neuron in the population. Interestingly, we found that the efficiency of the optimal code, or the ratio of noise entropy to output entropy, stays approximately constant. This occurs mainly because the per-neuron output entropy is also severely overestimated when only single neurons are observed. Our results also indicate that in an optimal network of size N, the amount of information about the stimulus can be larger than proportional to the size M of the observed subnetwork (i.e., $I_M > (M/N)I_N$). This means that the optimal codes for Ising-like models are not "combinatorial" in the sense that *all* output units need not be seen to decode properly. A full combinatorial code would be conceivable if the model allowed higher-than-pairwise couplings J.

All the encoding strategies we found have been observed in neural systems. Furthermore, as seen for our optimal networks, spontaneous activity patterns in real neural populations resemble responses to common stimuli (35, 36). One strategy—synergistic coding—that has been seen in some experiments (2–5) did not emerge from our optimization analyses. Perhaps synergy arises only as an optimal strategy for input statistics that we have not examined, or perhaps models with only pairwise interactions cannot access such codes. Alternatively, synergistic codes may not optimize information transmission—e.g., they are very susceptible to noise (10).

Our results could be construed as predicting adaptation of connection strengths to stimulus statistics (see, e.g., ref. 37). This prediction could be compared directly to data. To do this, we would select the $h_i(s)$ in our model (Eq. **2**) as the convolution of stimuli with the receptive fields of N simultaneously recorded neurons. Our methods would then predict the optimal connection strengths J_{ij} for encoding a particular stimulus ensemble. To compare to the actual connection strengths we would instead fit the model (Eq. **2**) directly to the recorded data (38, 39). Comparing the predicted and measured J_{ij} would provide a test of whether the network is an optimal, pairwise-interacting encoder for the given stimulus statistics. Testing the prediction of network adaptation would require changing the stimulus correlations and observing matched changes in the connection strengths.

ACKNOWLEDGMENTS. V.B. and G.T. thank the Weizmann Institute and the Aspen Center for Physics for hospitality. G.T., J.P., and V.B. were partly supported by National Science Foundation Grants IBN-0344678 and EF-0928048, National Institutes of Health (NIH) Grant R01 EY08124 and NIH Grant T32-07035. V.B. is grateful to the IAS, Princeton for support as the Helen and Martin Chooljian Member. E.S. was supported by the Israel Science Foundation (Grant 1525/08), the Center for Complexity Science, Minerva Foundation, the Clore Center for Biological Physics, and the Gruber Foundation.

1. Rieke F, Warland D, de Ruyter van Steveninck RR, Bialek W (1997) *Spikes: Exploring the Neural Code* (MIT Press, Cambridge, MA).
2. Gawne TJ, Richmond BJ (1993) How independent are the messages carried by adjacent inferior temporal cortical neurons? *J Neurosci* 13:2758–2771.
3. Puchalla JL, Schneidman E, Harris RA, Berry MJ, II (2005) Redundancy in the population code of the retina. *Neuron* 46:493–504.
4. Narayanan NS, Kimchi EY, Laubach M (2005) Redundancy and synergy of neuronal ensembles in motor cortex. *J Neurosci* 25:4207–4216.
5. Chechik G, et al. (2006) Reduction of information redundancy in the ascending auditory pathway. *Neuron* 51:359–368.
6. Schneidman E, Berry MJ, II, Segev R, Bialek W (2006) Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440:1007–1012.
7. Barlow HB (1959) Sensory mechanisms, the reduction of redundancy, and intelligence. *Proceedings of the Symposium on the Mechanization of Thought Process* (National Physical Laboratory, HMSO, London).
8. Abeles M (1991) *Corticonics: Neural Circuits of the Cerebral Cortex* (Cambridge Univ Press, Cambridge, UK).
9. Amit DJ (1989) *Modeling Brain Function: The World of Attractor Neural Networks* (Cambridge Univ Press, Cambridge, UK).
10. Barlow H (2001) Redundancy reduction revisited. *Network Comput Neural Syst* 12:241–253.
11. Schneidman E, Still S, Berry MJ, 2nd, Bialek W (2003) Network information and connected correlations. *Phys Rev Lett* 91:238701.
12. Atick JJ, Redlich AN (1990) Towards a theory of early visual processing. *Neural Comput* 2:308–320.
13. Srinivasan MV, Laughlin SB, Dubs A (1982) Predictive coding: A fresh view of inhibition in the retina. *Proc R Soc London Ser B* 216:427–459.
14. van Hateren JH (1992) A theory of maximizing sensory information. *Biol Cybern* 68:23–29.
15. Devries SH, Baylor DA (1997) Mosaic arrangement of ganglion cell receptive fields in rabbit retina. *J Neurophysiol* 78:2048–2060.
16. Borghuis BG, Ratliff CP, Smith RG, Sterling P, Balasubramanian V (2008) Design of a neuronal array. *J Neurosci* 28:3178–3189.
17. Balasubramanian V, Sterling P (2009) Receptive fields and the functional architecture in the retina. *J Physiol* 587:2753–2767.
18. Liu YS, Stevens CF, Sharpee TO (2009) Predictable irregularities in retinal receptive fields. *Proc Natl Acad Sci USA* 106:16499–16504.
19. Tkačik G, Walczak AM, Bialek W (2009) Optimizing information flow in small genetic networks. *Phys Rev E* 80:031920.
20. Walczak AM, Tkačik G, Bialek W (2010) Optimizing information flow in small genetic networks. II. Feed-forward interactions. *Phys Rev E* 81:041905.
21. MacKay DJC (2004) *Information Theory, Inference and Learning Algorithms* (Cambridge Univ Press, Cambridge, UK).
22. Shlens J, et al. (2006) The structure of multi-neuron firing patterns in primate retina. *J Neurosci* 26:8254–8266.
23. Tkačik G, Schneidman E, Berry MJ, II, Bialek W (2006) Ising models for networks of real neurons. arXiv.org:q-bio.NC/0611072.
24. Tkačik G, Schneidman E, Berry MJ, II, Bialek W (2010) Spin-glass model for a network of real neurons. arXiv.org:0912.5500.
25. Tang S, et al. (2008) A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro. *J Neurosci* 28:505–518.
26. Shlens J, et al. (2009) The structure of large-scale synchronized firing in primate retina. *J Neurosci* 29:5022–5031.
27. Jaynes ET (1957) Information theory and statistical mechanics. *Phys Rev* 106:620–630.
28. Brivanlou IH, Warland DK, Meister M (1998) Mechanisms of concerted firing among retinal ganglion cells. *Neuron* 20:527–539.
29. Cover TM, Thomas JA (1991) *Elements of Information Theory* (Wiley, New York).
30. Linsker R (1989) An application of the principle of maximum information preservation to linear systems. *Advances in Neural Information Processing Systems*, ed DS Touretzky (Morgan Kaufmann, San Francisco, CA), Vol 1.
31. Bell AJ, Sejnowski TJ (1995) An information-maximization approach to blind separation and blind deconvolution. *Neural Comput* 7:1129–1159.
32. Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci USA* 79:2554–2558.
33. Tkačik G, Prentice J, Schneidman E, Balasubramanian V (2009) Optimal correlation codes in populations of noisy spiking neurons. *BMC Neurosci* 10:O13.
34. Fitzgerald JD, Sharpee TO (2009) Maximally informative pairwise interactions in networks. *Phys Rev E* 80:031914.
35. Kenet T, Bibitchkov D, Tsodyks M, Grinvald A, Arieli A (2003) Spontaneously emerging cortical representations of visual attributes. *Nature* 425:954–956.
36. Fiser J, Chiu C, Weliky M (2004) Small modulation of ongoing cortical dynamics by sensory input during natural vision. *Nature* 431:573–578.
37. Hosoya T, Baccus SA, Meister M (2005) Dynamic predictive coding by the retina. *Nature* 436:71–77.
38. Tkačik G (2007) Information flow in biological networks. PhD thesis (Princeton University, Princeton, NJ).
39. Granot-Atdegi E, Tkačik G, Segev R, Schneidman E (2010) A stimulus-dependent maximum entropy model of the retinal population neural code. *Frontiers in Neuroscience. Conference abstract COSYNE 2010.*