Minimal Models of Multidimensional Computations

Jeffrey D. Fitzgerald[1,2], Lawrence C. Sincich[3], Tatyana O. Sharpee[1,2]*

1 Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, California, United States of America, **2** Center for Theoretical Biological Physics and Department of Physics, University of California, San Diego, La Jolla, California, United States of America, **3** Beckman Vision Center, University of California, San Francisco, San Francisco, California, United States of America

Abstract

The multidimensional computations performed by many biological systems are often characterized with limited information about the correlations between inputs and outputs. Given this limitation, our approach is to construct the maximum noise entropy response function of the system, leading to a closed-form and minimally biased model consistent with a given set of constraints on the input/output moments; the result is equivalent to conditional random field models from machine learning. For systems with binary outputs, such as neurons encoding sensory stimuli, the maximum noise entropy models are logistic functions whose arguments depend on the constraints. A constraint on the average output turns the binary maximum noise entropy models into minimum mutual information models, allowing for the calculation of the information content of the constraints and an information theoretic characterization of the system's computations. We use this approach to analyze the nonlinear input/output functions in macaque retina and thalamus; although these systems have been previously shown to be responsive to two input dimensions, the functional form of the response function in this reduced space had not been unambiguously identified. A second order model based on the logistic function is found to be both necessary and sufficient to accurately describe the neural responses to naturalistic stimuli, accounting for an average of 93% of the mutual information with a small number of parameters. Thus, despite the fact that the stimulus is highly non-Gaussian, the vast majority of the information in the neural responses is related to first and second order correlations. Our results suggest a principled and unbiased way to model multidimensional computations and determine the statistics of the inputs that are being encoded in the outputs.

Citation: Fitzgerald JD, Sincich LC, Sharpee TO (2011) Minimal Models of Multidimensional Computations. PLoS Comput Biol 7(3): e1001111. doi:10.1371/journal.pcbi.1001111

Editor: Karl J. Friston, University College London, United Kingdom

Received August 12, 2010; **Accepted** February 17, 2011; **Published** March 24, 2011

Funding: This work was funded by NIH Grant EY019493; NSF Grants IIS-0712852 and PHY-0822283 and the Searle Funds; the Alfred P. Sloan Fellowship; the McKnight Scholarship; W.M. Keck Research Excellence Award; and the Ray Thomas Edwards Career Development Award in Biomedical Sciences. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: sharpee@salk.edu

Introduction

There is an emerging view that the primary function of many biological systems, from the molecular level to ecosystems, is to process information [1–4]. The nature of the computations these systems perform can be quite complex [5], often due to large numbers of components interacting over wide spatial and temporal scales, and to the amount of data necessary to fully characterize those interactions. Constructing a model of the system using limited knowledge of the correlations between inputs and outputs can impose implicit assumptions and biases leading to a mischaracterization of the computations. To minimize this type of bias, we maximize the noise entropy of the system subject to constraints on the input/output moments, resulting in the response function that agrees with our limited knowledge and is maximally uncommitted toward everything else. An equivalent approach in machine learning is known as conditional random fields [6]. We apply this idea to study neural coding, showing that logistic functions not only maximize the noise entropy for binary outputs, but are also special closed-form cases of the minimum mutual information (MinMI) solutions [7] when the average firing rate of a neuron is fixed. Recently, MinMI was used to assess the information content in constraints on the interactions between neurons in a network [8]. We use this idea to study single neuron coding to discover what statistics of the inputs are encoded in the outputs. In macaque retina and lateral geniculate nucleus, we find that the single neuron responses to naturalistic stimuli are well described with only first and second order moments constrained. Thus, the vast majority of the information encoded in the spiking of these cells is related only to the first and second order statistics of the inputs.

To begin, consider a system which at each moment in time receives a D-dimensional input $\mathbf{x}(t)=(x_1(t),...,x_D(t))$ from a known distribution $P(\mathbf{x})$, such as a neuron receiving a sensory stimulus or post-synaptic potentials. The system then performs some computation to determine the output $y(t)$ according to its response function $P(y|\mathbf{x})$. The complete input/output correlation structure, i.e. all moments involving y and $\mathbf{x}$, can be calculated from this function through the joint distribution $P(y,\mathbf{x})=P(y|\mathbf{x})P(\mathbf{x})$, e.g. $\langle yx_i\rangle=\int P(y,\mathbf{x})yx_i dyd\mathbf{x}$. Alternatively, the full list of such moments contains the same information about the computation as the response function itself, although such a list is infinite and experimentally impossible to obtain. However, a partial list is usually obtainable, and as a first step we can force the input/output correlations from the model to match those which are known from the data. The problem is then choosing from the infinite number of models that agree with those constraints. Following the argument of Jaynes [9,10], we seek the

Author Summary

Biological systems across many scales, from molecules to ecosystems, can all be considered information processors, detecting important events in their environment and transforming them into actions. Detecting events of interest in the presence of noise and other overlapping events often necessitates the use of nonlinear transformations of inputs. The nonlinear nature of the relationships between inputs and outputs makes it difficult to characterize them experimentally given the limitations imposed by data collection. Here we discuss how minimal models of the nonlinear input/output relationships of information processing systems can be constructed by maximizing a quantity called the noise entropy. The proposed approach can be used to "focus" the available data by determining which input/output correlations are important and creating the least-biased model consistent with those correlations. We hope that this method will aid the exploration of the computations carried out by complex biological systems and expand our understanding of basic phenomena in the biological world.

model which avails the most uncertainty about how the system will respond.

Information about the identity of the input can be obtained by observing the output, or vice versa, quantified by the mutual information $I(y;\mathbf{x}) = H_{\text{resp}} - H_{\text{noise}}$ [11,12]. The first term is the response entropy, $H_{\text{resp}} = -\int dy P(y)\log P(y)$, which captures the overall uncertainty in the output. The second term is the so-called noise entropy [13],

$$H_{\text{noise}} = -\int d\mathbf{x} P(\mathbf{x}) \int dy P(y|\mathbf{x}) \ln P(y|\mathbf{x}), \tag{1}$$

representing the uncertainty in y that remains if $\mathbf{x}$ is known. If the inputs completely determine the outputs, there is no noise and the mutual information reaches its highest possible value, $I = H_{\text{resp}}$. In many realistic situations however, repeated presentations of the same inputs produce variable outputs producing a nonzero noise entropy [14] and lowering the information transmitted.

By maximizing the noise entropy, the model is forced to be consistent with the known stimulus/response relationships but is as uncertain as possible with respect to everything else. We show that this maximum noise entropy (MNE) response function for binary output systems with fixed average outputs is also a minimally informative one. This approach is a special closed-form case of the mutual information minimization technique [8], which has been used to address the information content of constraints on the interactions between neurons. Here we use the minimization of the mutual information to characterize the computations of single neurons and discover what about the stimulus is being encoded in their spiking behavior.

Results

Maximum noise entropy models

The starting point for constructing any maximum noise entropy model is the specification of a set of constraints $\{\langle A_j(y,\mathbf{x})\rangle\}$, where $\langle\ldots\rangle$ indicates an average over the joint distribution $P(y,\mathbf{x})$. These constraints reflect what is known about the system from experimental measurements, or a hypothesis about what is relevant for the information processing of the system. For neural coding, the constraints could be quantities such as the spike-triggered average [15–18] or covariance [19–22], equivalent to $\{\langle yx_i\rangle\}$ and $\{\langle yx_ix_j\rangle\}$, respectively. With each additional constraint, our knowledge of the true input/output relationship increases and the correlation structure of the model becomes more similar to that of the actual system.

Given the constraints, the general MNE response function is given by (see Methods)

$$P(y|\mathbf{x}) = \frac{1}{Z(\mathbf{x})}\exp\left(\sum_j \lambda_j A_j(y,\mathbf{x})\right), \tag{2}$$

where the $\mathbf{x}$-dependent partition function $Z(x) = \int dy \exp\left(\sum_j \lambda_j A_j(y,\mathbf{x})\right)$ ensures that the MNE response function is consistent with normalization, i.e. $\int dy P(y|\mathbf{x}) = 1$. The MNE response function in Eq. (2) has the form of a Boltzmann distribution [23] with a Lagrange multiplier λ_j for each constraint. The values of these parameters are found by matching the experimentally observed averages with the analytical averages obtained by from derivatives of $\log(Z)$ [23].

Binary responses and minimum mutual information

Many systems in biological settings produce binary outputs. For instance, the neural state y can be thought of as binary, with $y = 0$ for the silent state and $y = 1$ for the "spiking" state, during which an action potential is fired [13]. The inputs themselves could be a sensory stimulus or all of the synaptic activity impinging upon a neuron, both of which are typically high-dimensional [24]. Another example is gene regulation [25], where the inputs could be the concentrations of transcription factors and the binary output represents an on/off transcription state of the gene. For these systems, the constraints of interest are proportional to y. This is because any moments independent of y will cancel due to the partition function and any moments with higher powers are redundant, e.g. $y = y^2$ if $y = 0$ or 1. In this case, the set of constraints may be written more specifically as $\{ya_j(\mathbf{x})\}$ and the MNE response function becomes the well-known logistic function

$$P(y=1|\mathbf{x}) = \frac{1}{1+e^{-f(x)}}, \quad f(\mathbf{x}) = \sum_j \lambda_j a_j(\mathbf{x}), \tag{3}$$

with $P(y=0|\mathbf{x}) = 1 - P(y=1|\mathbf{x})$. Thus for all binary MNE models, the effect of the constraints is to perform a nonlinear transformation of the input variables, $f(\mathbf{x})$, to a space where the spike probability is given by the logistic function (inset, Fig. 1).

For neural coding, one of the most fundamental and easily measured quantities is the total number of spikes produced by a neuron over the course of an experiment, equivalent to the mean firing rate. By constraining this quantity, or more specifically its normalized version $P(y=1) = \langle y\rangle$, the MNE model is turned into a minimum information model. This holds because the response entropy H_{resp} is completely determined by the distribution $P(y)$, which is in turn constrained by $P(y=1)$ if the response is binary. With the response entropy constrained to match the experimentally observed system, maximizing the noise entropy is equivalent to minimizing information. Therefore, as was proposed in [7], any model that satisfies a given set of constraints will convey the information that is due only to those constraints. With each additional constraint our knowledge of the correlation structure increases along with the minimum possible information given that knowledge, which approaches the true value as illustrated schematically in Fig. 1.

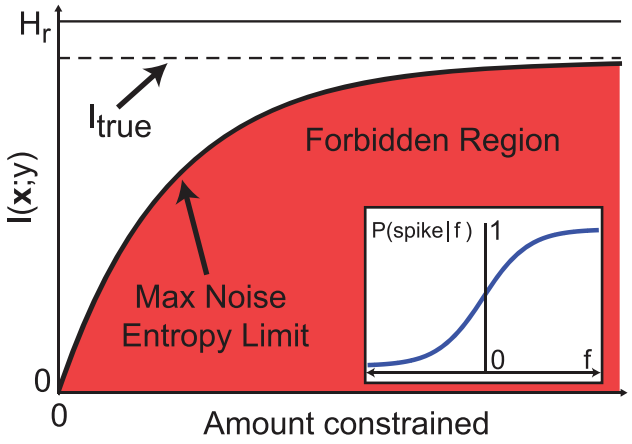

Figure 1. The maximum noise entropy (MNE) limit. This cartoon illustrates the consequences of a minimally informative, MNE response function. As knowledge of the correlation structure increases (which amounts to constraining more moments of the conditional output distribution), the least possible amount of information consistent with that knowledge increases along the solid line. Below the MNE limit is a forbidden region where a response function cannot be consistent with the given set of constraints. All models are bounded from above by the response entropy, corresponding to a noiseless system. Any response function above the MNE limit thus involves unknown and unconstrained moments which carry information. The information associated with the MNE response function increases toward the true value as the knowledge of the distribution tends to infinity. For a binary system, the response function is a logistic function (inset) in the transformed input space defined by $f(\mathbf{x})$, cf. Eq. (3).
doi:10.1371/journal.pcbi.1001111.g001

The simplest choice is a first order model (MNE_1) where the spikes are correlated with each input dimension separately. This model requires knowledge of the set of moments $\{\langle yx_i\rangle\}$, the spike-triggered average stimulus. For MNE_1, the transformation on the inputs is linear, $f(\mathbf{x})=\lambda_0+\sum_{j=1}^{D}\lambda_j x_j$, where the constant λ_0 is the Lagrange multiplier for the spike probability constraint. With knowledge of only first order correlations, we see that the model neuron is effectively one-dimensional, choosing a single dimension in the D-dimensional input space $\lambda=(\lambda_1,\ldots,\lambda_D)$ and disregarding all information about any other directions.

With higher order constraints, the transformation is nonlinear and the model neuron is truly multidimensional. For instance, the next level of complexity is a second order model (MNE_2), in which spikes may also interact with pairs of inputs. This model is obtained by constraining $\{\langle yx_ix_j\rangle\}$, equivalent to knowing the spike-triggered covariance of the stimulus, resulting in the input transformation $f(\mathbf{x})=\lambda_0+\sum_j\lambda_j x_j+\sum_{i,j}\lambda_{ij}x_ix_j$. Any other MNE model can be constructed in the same fashion by choosing a different set of constraints, reflecting different amounts of knowledge.

The mutual information of the MNE model I_{MNE} is the information content of the constraints. The ratio of I_{MNE} to the empirical estimate I_{obs} of the true mutual information of the system is the percent of the information captured by the constraints. This quantity is always less than or equal to one, with equality being reached if and only if all of the relevant moments have been constrained. This suggests a procedure to identify the relevant constraints, described in Fig. 2A. First, a hypothesis is made about which constraints are important. Then the corresponding MNE model is constructed and the information calculated. If the information captured is too small, the constraints are modified until a sufficiently large percentage is reached. Any constraints beyond that are relatively unimportant for describing the computation of the neuron.

As an illustrative example of the MNE method, consider a binary neuron which itself receives binary inputs (i.e. a logic gate). If the neuron in question receives n binary inputs, we are guaranteed to capture 100% of the information with n^{th}-order statistics because all moments involving powers greater than one of either y or any x_i are redundant. However, different coding schemes may encode different statistics of the inputs. For instance, if the neuron receives only two inputs (Fig. 2B), the well-known AND and OR logic gate behaviors are completely described with only first order moments [26]. Correspondingly, the first order model MNE_1 captures 100% of the information. Such a neuron can be said to encode only first order statistics of the inputs, and the spike-triggered average stimulus contains all of the information necessary to fully understand the computation. On the other hand, the XOR gate (Fig. 2C, left) requires second order interactions. This is reflected by MNE_1 and MNE_2 accounting for 0% and 100% of the information, respectively. More complicated coding schemes may involve both first and second order interactions, such as for the gate shown in the right panel of Fig. 2C. Here, MNE_1 and MNE_2 account for 10% and 100% of the information, respectively, and correctly quantify the degree to which each order of interaction is relevant to this neuron.

Similar situations show up for neurons that receive three binary inputs. The top panel of Fig. 2D shows an example of a neuron which only requires second order interactions. The parameters of MNE_3 are exactly the same as MNE_2, with the third order coefficient $\lambda_{123}=0$. The bottom panel shows an example of a situation in which third order interactions are necessary. Correspondingly, MNE_3 increases the information explained over MNE_2 from 71% to 100%. These simulations demonstrate that despite the different coding schemes used by neurons, the information content of each order of interaction can be correctly identified using logistic MNE models.

Neural coding of naturalistic inputs

In their natural environment, neurons commonly encode high-dimensional analog inputs, such as a visual or auditory stimulus as a function of time. It is important to note that the non-binary nature of the inputs means that the ability to capture 100% of the information between y and the n inputs with n^{th}-order statistics is not guaranteed anymore. Often, the dimensionality of the inputs may be reduced because the neurons are driven by a smaller subspace of relevant dimensions (e.g. [27–33]). However, even in those cases we are often forced to use qualitative terms such as 'ring' or 'crescent' to describe the experimentally observed response functions. With no principled way of fitting empirical response functions, the details of the interactions between neural responses and reduced inputs have been difficult to quantify.

The MNE method provides a quantitative framework for characterizing neural response functions, which we now apply to 9 retinal ganglion cells (RGCs) and 9 cells in the lateral geniculate nucleus (LGN) of macaque monkeys, recorded *in vivo* (see Methods). The visual input was a time dependent sequence of luminance values synthesized to mimic the non-Gaussian statistics of light intensity fluctuations in the natural visual environment [34–36].

A 1s segment of the normalized light intensity $s(t)$ is shown in Fig. 3A. A previous study has shown that the responses of these neurons are correlated with the stimulus over an approximately 200 ms window preceding the response. When binned at 4 ms resolution, which ensures binary responses, the input is a vector in a 50 dimensional space. However, spikes are well predicted by

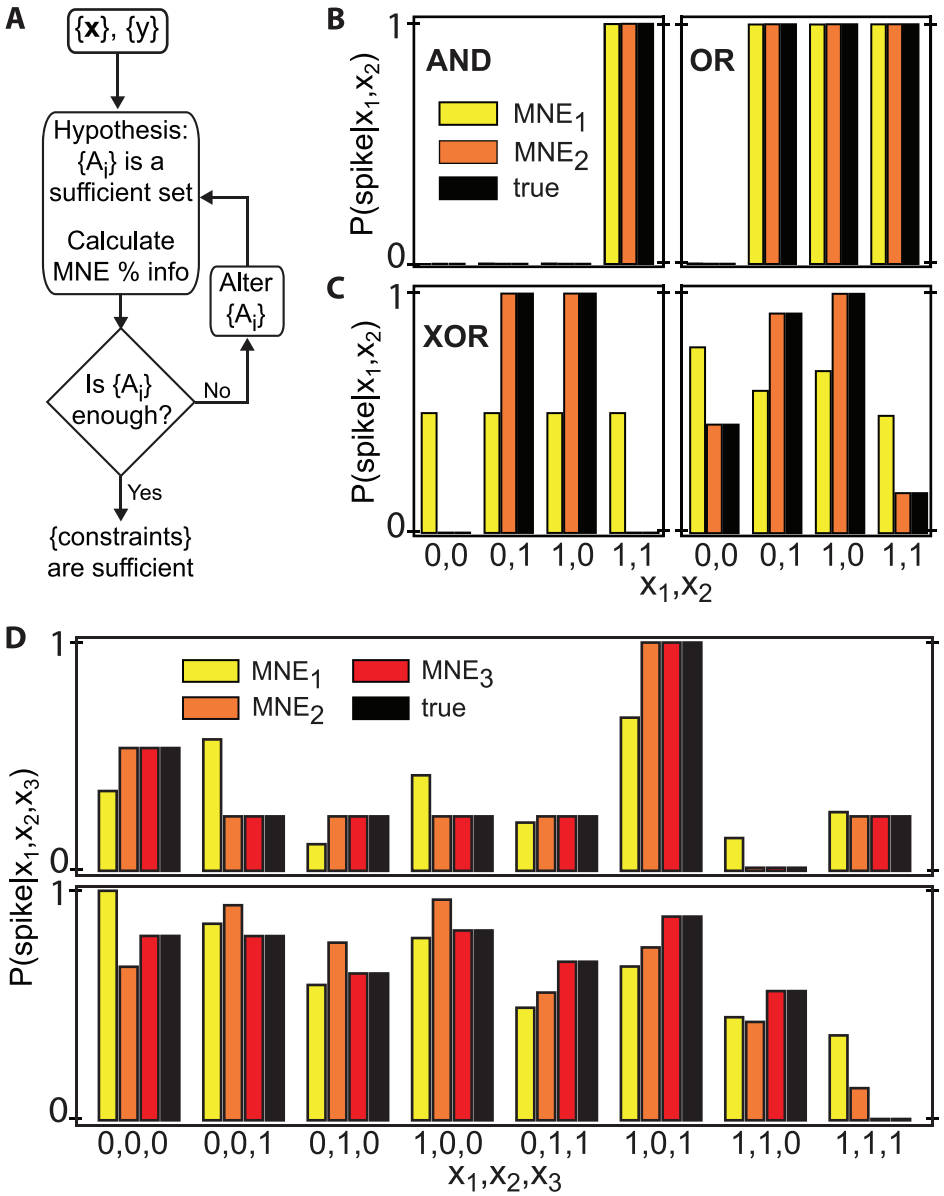

Figure 2. Using the MNE method for response functions to binary inputs. A) Flowchart representing how to determine the relevant constraints. The hypothesis that a minimal set of constraints is sufficient is tested by constructing the corresponding MNE model and calculating the information captured by the model. If the percent information is insufficient, the set of constraints is augmented. **B**) Response functions and MNE models for two binary inputs; the true system is shown in black, and first and second order MNE models (MNE_1 and MNE_2) in yellow and orange, respectively. The AND and OR gates use only first order interactions; both MNE models explain 100% of the information. **C**) The XOR gate (left) uses only second order interactions; MNE_1 explains 0% while MNE_2 explains 100% of the information. An example of a mixed response function (right), for which both first and second order interactions are used (10% and 100% respectively). **D**) Two examples of response functions with three binary inputs, with MNE_3 shown in red. Only second order interactions are necessary for the top gate, with 48%, 100% and 100% of the information captured by the first, second and third order MNE models. For the bottom gate, the models capture 39%, 71% and 100% of the information, indicating that third order constraints are necessary. In all cases, $I(y;\mathbf{x})$ was calculated assuming a uniform input distribution.
doi:10.1371/journal.pcbi.1001111.g002

using a 2 dimensional subspace [29] identified through the Maximally Informative Dimensions (MID) technique [37].

These two relevant dimensions, shown for a RGC in Fig. 3B, form a two dimensional receptive field which preserves the most information about the spikes in going from 50 to 2 dimensions. The two linear filters are convolved with the stimulus to produce reduced inputs $x_1(t)$ and $x_2(t)$, shown in Fig. 3A. The resulting input probability distribution in the reduced space is shown in Fig. 3C. The measured responses of the neuron then form a two-dimensional response function shown in Fig. 3D, where the color scale indicates the probability of a spike as a function of the two relevant input components.

To gain insight into the nature of this neuron's computational function and find the important interactions, we apply the MNE method starting with the first order MNE model shown in Fig. 3E. The first order model produces a response function which bears little resemblance to the empirical one and accounts for only 76% of the information. The next step is a second order MNE model

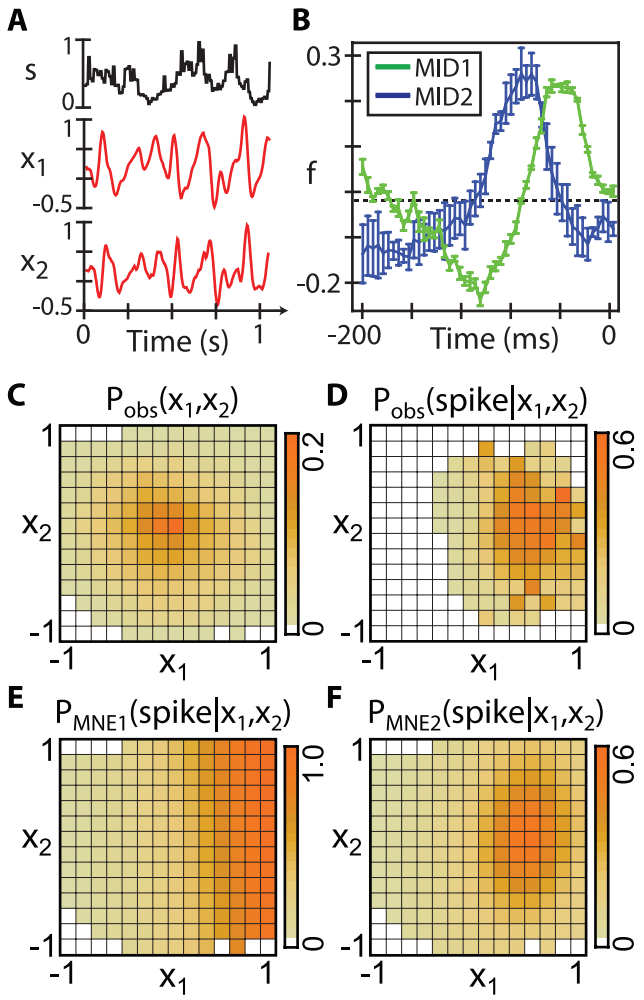

Figure 3. MNE models for a RGC. A) The normalized luminance $s(t)$ of the visual input, along with the two most informative reduced inputs, x_1 and x_2, shown for a section of the stimulus presented to neuron mn122R4_3_RGC. **B**) The two maximally informative dimensions (MID) for this neuron (error bars are standard error in the mean). Each dimension is a filter which spans 200 ms before the neural output. The convolution of these filters with the stimulus produce x_1 and x_2, which are normalized to lie in the range -1 to 1. In this 2-D reduced input space, the input distribution, **C**), and observed response function, **D**), are shown, discretized into 14 bins along each dimension. White squares in the input distribution indicate unsampled inputs, while white squares in the response function indicate no spikes were recorded. The first order, **E**), and second order, **F**), MNE response functions for this cell explain 76% and 98% of the information, respectively.
doi:10.1371/journal.pcbi.1001111.g003

(Fig. 3F), which produces a response function quite similar to the empirical one in both shape and amplitude, while accounting for 98% of the information. Thus, for this neuron, knowledge of second order moments is both necessary and sufficient to generate a highly accurate model of the neural responses.

This result was typical across the population of cells, as illustrated in Fig. 4A by comparing the information captured by the first order versus second order models. The majority of the cells were well described by the second order model, accounting for over 90% of the information. When averaged across the population, the first order model captured 78% and the second order model captured 93% of I_{obs}. These results suggest that the inclusion of second order interactions are both necessary and sufficient to describe the responses of these neurons to naturalistic stimuli.

Since the MNE response function is a distribution of outputs given inputs, another way to check the effectiveness of any MNE model is to compare its moments with those obtained from experiments. The moments constrained to obtain the model will be identical to the experimental values by construction; it is the higher order moments, left unconstrained, that should be compared. In Fig. 4B we show two such comparisons for the correlation functions $C(y,x_1^3)=\langle(y-\langle y\rangle)(x_1^3-\langle x_1^3\rangle)\rangle$ and $C(y,x_1,x_2^2)=\langle(y-\langle y\rangle)(x_1-\langle x_1\rangle)(x_2^2-\langle x_2^2\rangle)\rangle$, which involve moments unconstrained in the MNE_1 and MNE_2 models. In both cases, the first order model predictions show more scatter than those of the second order model; the latter does a reasonable job of predicting the experimentally observed correlations. This result broadly demonstrates the sufficiency of second order interactions to model these neural responses, and shows that higher-order moments carry little to no additional information.

The two-dimensional second order MNE response functions have contours of constant probability which are conic sections. The parameter which governs the interaction between the two input dimensions, λ_{12}, is related to the degree to which the axes of symmetry of the conic sections are aligned with the two-dimensional basis. For example, if the contours are ellipses, then $\lambda_{12}=0$ if the semi-major and semi-minor axes are parallel to the axes chosen to describe the input space, and $\lambda_{12}\neq 0$ otherwise (see inset, Fig. 5). To assess the importance of this cross term, we compared the performance of second order MNE models with and without λ_{12}. This additional term can only improve the performance of the model; however, as shown in Fig. 5, the improvements across the population are small. Thus, the dimensions found using the MID method are naturally parallel to the axes of symmetry of the response functions; however, this does not imply that the response function is separable due to the $\mathbf{x}$ dependence of the normalization term $Z(\mathbf{x})$.

Discussion

For neural coding of naturalistic visual stimuli in early visual processing, we see that the bulk of what is being encoded is first order stimulus statistics. While the information gained by measuring the spike-triggered average is substantial, it is insufficient to accurately describe the neural responses. A second order model, which takes into account the spike-triggered input covariance, adds a sufficient amount of information. Thus the firing rates of these neurons have encoded the first and second order statistics of the inputs. Due to the fact that the natural inputs are non-binary and non-Gaussian, there exists a potential for very high-order interactions to be represented in the neural firing rate. It is known that higher order parameters of textures are perceptually salient [38–40], but it is unknown whether high order temporal statistics are also perceptually salient. Our results suggest that such temporal statistics are not encoded in the time-dependent firing rate, although they could be represented through populations of neurons or specific temporal sequences of spikes [41,42].

Jaynes' principle of maximum entropy [9,10] has a long and diverse history, with example applications in image restoration in astrophysics [43], extension of Wiener analysis to nonlinear stochastic transducers [44] and more recently in neuroscience [45–47]. In the latter studies, H_{resp} was maximized subject to constraints on the first and second order moments of the neural states $\{y_i\}$ and $\{y_iy_j\}$ for a set of neurons in a network. The resulting pairwise Ising model was shown to accurately describe

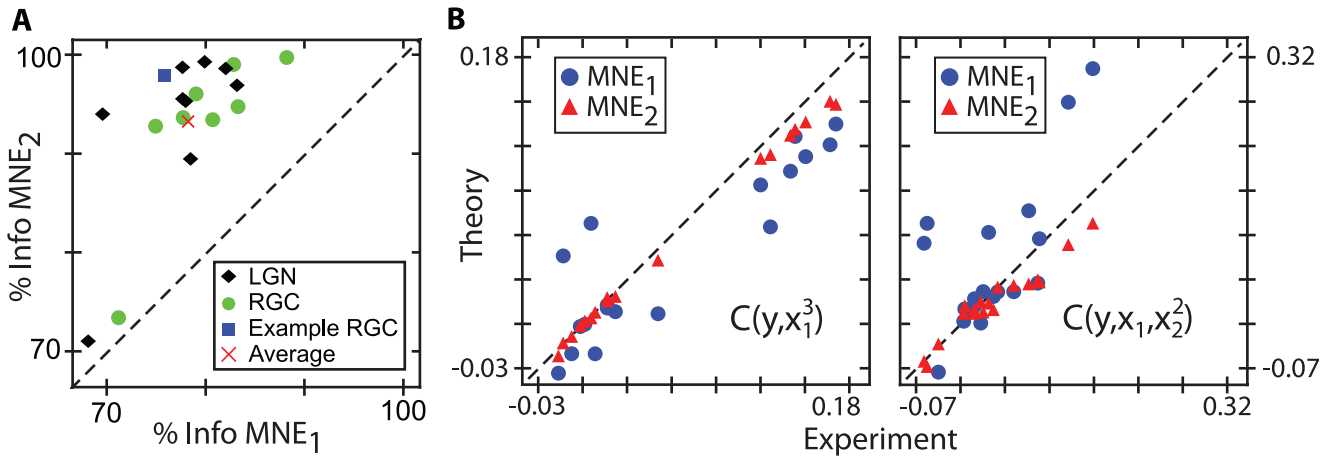

Figure 4. Second order MNE models are sufficient across the population. A) A direct comparison of the percent information captured by MNE_1 and MNE_2. No cells are sufficiently modeled with a first order model, but most are with the second order model. The average information captured is 78% for MNE_1 and 93% for MNE_2. **B**) Comparison of experimentally measured values to theoretical predictions for higher-order unconstrained moments. Predictions for $C(y,x_1^3)$, left, and $C(y,x_1,x_2^2)$, right, show a dramatic improvement when second order interactions are included in the model.
doi:10.1371/journal.pcbi.1001111.g004

the distribution of network states $P(\mathbf{y})$ of real neurons under various conditions. Since then the application of the Ising model to neuroscience has received much attention [48,49], and it is still a subject of debate if and how these results extrapolate to larger populations of neurons [50]. Temporal correlations have also been shown to be important in both cortical slices and networks of cultured neurons [47].

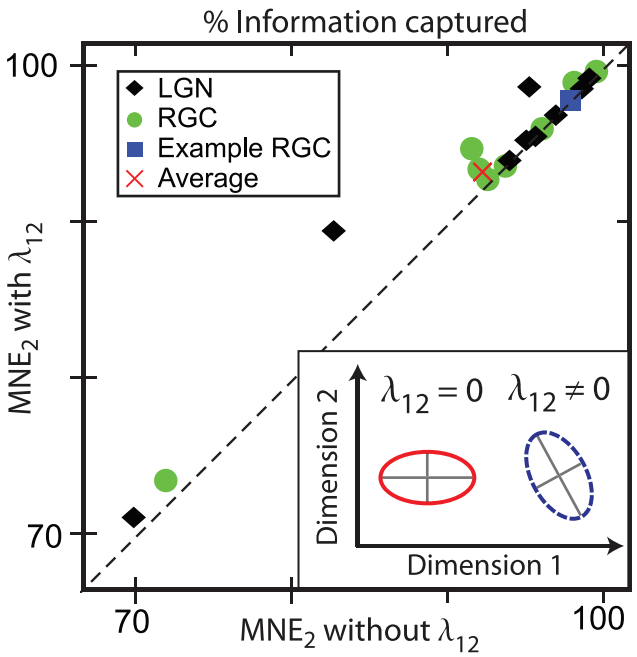

Figure 5. Importance of the mixed second order moments. A comparison of the percent of the information captured by a second order model (MNE_2) that constrains $\langle yx_1x_2\rangle$ (i.e. $\lambda_{12}\neq 0$) and a second order model with $\lambda_{12}\equiv 0$. For most cells, the information increases only slightly for $\lambda_{12}\neq 0$, indicating that little information is gained by constraining this moment. (Inset) The parameter λ_{12} determines the angle between the axes of symmetry of the response function and the basis of the input space.
doi:10.1371/journal.pcbi.1001111.g005

In contrast to maximum entropy models that deal with stationary or averaged distributions of states, the goal of maximizing the noise entropy is to find unbiased response functions. This approach is equivalent to conditional random field (CRF) models [6] in machine learning. The parameters of a CRF are fit by maximizing the likelihood using iterative or gradient ascent algorithms [51] and have been used, for example, in classification and segmentation tasks [52]. The parameters of MNE models may also be found using maximum likelihood, or as was done here, by solving a set of simultaneous constraint equations numerically. Another example of a maximum noise entropy distribution is the Fermi-Dirac distribution [23] from statistical physics, which is a logistic function governing the binary occupation of fermion energy levels. Thus, in the same way that the Boltzmann distribution was interpreted by Jaynes as the most random one consistent with measurements of the energy, the Fermi-Dirac distribution can be interpreted as the least biased binary response function consistent with an average energy. However, to our knowledge, this method has never been used in the context of neural coding to determine the input statistics which are being encoded by a neuron and create the corresponding unbiased models.

Previous work has applied the principle of minimum mutual information (MinMI) [7] to neural coding, thus identifying the relevant interactions between neurons [8]. We have shown that the closed-form MNE solutions for binary neurons constitute a special case of MinMI, since the response entropy is fixed if the average firing rate is constrained. In general, the MinMI principle results in a self-consistent solution that must be solved iteratively to obtain the response function. The reason why MNE models are closed-form is that the constraints are formulated in terms of moments of the output distribution instead of the output distribution itself. In addition to the case of binary responses, MNE models can become closed-form MinMI models for any input/output systems where the response entropy can be fixed in terms of the moments of the output variable. Examples include Poisson processes with fixed average response rate or Gaussian processes with fixed mean and variance of the response rate. The

framework for analyzing the interactions between inputs and outputs that we present here can thus be extended to a broad and diverse set of computational systems.

Our approach can be compared to other optimization techniques commonly used to study information processing. For example, rate-distortion theory [11,12,53,54] seeks minimum information transmission rate over a channel with a fixed level of signal distortion, e.g. lossy image or video compression. In that case, the best solution is the one which transmits minimal information because this determines the average length of the codewords. In our method, we also obtain minimally informative solutions, not because they are optimal for signal transmission, but because they are the most unbiased guess at a solution given limited knowledge of a complex system.

At the other end of the optimization spectrum is maximization of information [1,13,55]. The goal in that case is to study not how the neuron *does* compute, but how it *should* compute to get the most information, perhaps with limited resources. This strategy has been used to find neural response functions for single neurons [56,57], as well as networks [58,59]. When confronted with incomplete knowledge of the correlation structure, a maximum information approach would choose the values of the unconstrained moments such that they convey the most information possible, whereas the minimum information approach provides a lower bound to the true mutual information, and allows us to investigate how this lower bound increases as more moments are included. If the goal is to study the limits of neural coding, then maximizing the information may be the best procedure. If, however, the goal is to dissect the computational function of an observed neuron, we argue that the more agnostic approaches of maximizing the noise entropy or minimizing the mutual information are better-suited.

Methods

Ethics statement

Experimental data were collected as part of the previous study using procedures approved by the UCSF Institutional Animal Care and Use Committee, and in accordance with National Institutes of Health guidelines.

Maximum noise entropy model

A maximum noise entropy model is a response function $P(y|\mathbf{x})$ which agrees with a set of constraints and is maximally unbiased toward everything else. The constraints are experimentally observed moments involving the response y and stimulus $\mathbf{x}$, $\{\langle A_j(y,\mathbf{x})\rangle\}$, where $\langle A_j\rangle = \int\int d\mathbf{x}dy A_j(\mathbf{x},y)P(\mathbf{x})P(y|\mathbf{x})$, which must be reproduced by the model. The set of C constraints, including the normalization of $P(\mathbf{x},y)$, are then added to the noise entropy to form the functional

$$H_c = H_{\text{noise}} + \sum_{j=1}^{C}\lambda_j \int\int d\mathbf{x}dy A_j(\mathbf{x},y)P(\mathbf{x})P(y|\mathbf{x}), \qquad (4)$$

with a Lagrange multiplier λ_j for each constraint. Setting $\frac{\delta H c}{\delta P(y|\mathbf{x})} = 0$ and enforcing normalization yields Eq. 1. For a binary system, $y=0$ or 1, all the constraints take the form $\{ya_j(\mathbf{x})\}$, and the partition function is $Z(\mathbf{x}) = 1 + e^{f(\mathbf{x})}$, where $f(\mathbf{x}) = \sum_j \lambda_j a_j(y,\mathbf{x})$.

The values of the Lagrange multipliers are found such that the set of equations

$$\left\{\langle A_j\rangle = \int d\mathbf{x}P(\mathbf{x})\frac{\partial}{\partial\lambda_j}\log Z(\mathbf{x})\right\}, \qquad (5)$$

is satisfied, with the analytical averages on the right-hand side obtained from derivatives of the free energy $\log Z$ [23]. Simultaneously solving this set of equations has previously been shown to be equivalent to maximizing the log-likelihood [51].

Physiology experiment

The neural data analyzed here were collected in a previous study [29] and the details are found therein. Briefly, the stimulus was a spot of light covering a cell's receptive field center, flickering with non-Gaussian statistics that mimic those of light intensity fluctuations found in natural environments [35,36]. The values of light intensities were updated every 12.5ms (update rate 80Hz). The spikes were recorded extracellularly in the LGN with high signal-to-noise, allowing for excitatory post-synaptic potentials generated by the RGC inputs to be recorded. From such data, the complete spike trains of both RGCs and LGN neurons could be reconstructed [60].

Dimensionality reduction

The neural spike trains were binned at 4 ms resolution, ensuring that the response was binary. The stimulus was re-binned at 250 Hz to match the bin size of the spike analysis. The neurons were uncorrelated with light fluctuations beyond 200 ms before a spike, and the stimulus vector $\mathbf{s}(t)$ was taken to be the 200 ms window (50 time points) of the stimulus preceding t. Just two projections of this 50-dimensional input are sufficient to capture a large fraction of the information between the light intensity fluctuations and the neural responses (84% for the example neuron mn122R4_3_RGC, and 85% on average across the population). The two most relevant features of each neuron were found by searching the space of all linear combinations of two input dimensions for those which accounted for maximal information in the measured neural responses [37], subject to cross-validation to avoid overfitting. Each of the two features, $\mathbf{f}_1$ and $\mathbf{f}_2$, is a 50-dimensional vector which converts the input into a reduced input, calculated by taking the dot product, i.e. $x_1(t) = \mathbf{f}_1 \cdot \mathbf{s}(t)$. The algorithm for searching for maximally informative dimensions is available online at http://cnl-t.salk.edu.

Acknowledgments

We thank Jonathan C. Horton for sharing the data collected in his laboratory and the CNL-T group for helpful conversations.

Author Contributions

Conceived and designed the experiments: JDF LCS TOS. Performed the experiments: LCS. Analyzed the data: JDF TOS. Wrote the paper: JDF LCS TOS.

References

1. Haken H (1988) Information and Self-Organization. Berlin: Springer.
2. Bray D (1995) Protein molecules as computational elements in living cells. Nature 376: 307–312.
3. Körding K (2007) Decision theory: What "should" the nervous system do? Science 318: 606–610.
4. Sanfey AG (2007) Social decision-making: Insights from game theory and neuroscience. Science 318: 598–602.
5. Grenfell BT, Williams CS, Björnstad ON, Banavar JR (2006) Simplifying biological complexity. Nat Phys 2: 212–214.
6. Lafferty J, McCallum A, Pereira F (2001) Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: Proceedings of the Eighteenth International Conference on Machine Learning. pp 282–289.
7. Globerson A, Tishby N (2004) The minimum information principle for discriminative learning. In: Proceedings of the 20th conference on Uncertainty

in artificial intelligence. Arlington, Virginia: AUAI Press, UAI '04, 193–200, Available: http://portal.acm.org/citation.cfm?id=1036843.1036867.
8. Globerson A, Stark E, Vaadia E, Tishby N (2009) The minimum information principle and its application to neural code analysis. Proc Natl Acad Sci USA 106: 3490–3495.
9. Jaynes ET (1957) Information theory and statistical mechanics. Phys Rev 106: 620–630.
10. Jaynes ET (1957) Information theory and statistical mechanics ii. Phys Rev 108: 171–190.
11. Shannon C (1949) Communication in the presence of noise. Proc of the IRE 37: 10–21.
12. Cover TM, Thomas JA (1991) Information theory. New York: John Wiley & Sons, Inc.
13. Rieke F, Warland D, de Ruyter van Steveninck RR, Bialek W (1997) Spikes: Exploring the neural code. Cambridge: MIT Press.
14. Strong SP, Koberle R, de Ruyter van Steveninck RR, Bialek W (1998) Entropy and information in neural spike trains. Phys Rev Lett 80: 197–200.
15. de Boer E, Kuyper P (1968) Triggered correlation. IEEE Trans Biomed Eng 15: 169–179.
16. Victor J, Shapley R (1980) A method of nonlinear analysis in the frequency domain. Biophys J 29: 459–483.
17. Meister M, Berry MJ (1999) The neural code of the retina. Neuron 22: 435–450.
18. Chichilnisky EJ (2001) A simple white noise analysis of neuronal light responses. Network: Comput Neural Syst 12: 199–213.
19. de Ruyter van Steveninck RR, Bialek W (1988) Real-time performance of a movement-sensitive neuron in the blowfly visual system: coding and information transfer in short spike sequences. Proc R Soc Lond B 265: 259–265.
20. Bialek W, de Ruyter van Steveninck RR (2005) Features and dimensions: Motion estimation in fly vision. Q-bio/ 0505003.
21. Pillow J, Simoncelli EP (2006) Dimensionality reduction in neural models: An information-theoretic generalization of spike-triggered average and covariance analysis. J Vis 6: 414–428.
22. Schwartz O, Pillow J, Rust N, Simoncelli EP (2006) Spike-triggered neural characterization. J Vis 176: 484–507.
23. Landau LD, Lifshitz EM (1959) Statistical Physics. Oxford: Pergamon Press.
24. Dayan P, Abbott LF (2001) Theoretical neuroscience: computational and mathematical modeling of neural systems. Cambridge: MIT Press.
25. Kauffman SA (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. J Theoret Biol 22: 437–467.
26. Schneidman E, Still S, Berry MJ, Bialek W (2003) Network information and connected correlations. Phys Rev Lett 91: 238701.
27. Fairhall AL, Burlingame CA, Narasimhan R, Harris RA, Puchalla JL, et al. (2006) Selectivity for multiple stimulus features in retinal ganglion cells. J Neurophysiol 96: 2724–2738.
28. Rust NC, Schwartz O, Movshon JA, Simoncelli EP (2005) Spatiotemporal elements of macaque v1 receptive fields. Neuron 46: 945–956.
29. Sincich LC, Horton JC, Sharpee TO (2009) Preserving information in neural transmission. J Neurosci 29: 6207–6216.
30. Chen X, Han F, Poo MM, Dan Y (2007) Excitatory and suppressive receptive field subunits in awake monkey primary visual cortex (v1). Proc Natl Acad Sci USA 104: 19120–5.
31. Atencio CA, Sharpee TO, Schreiner CE (2008) Cooperative nonlinearities in auditory cortical neurons. Neuron 58: 956–966.
32. Maravall M, Petersen RS, Fairhall A, Arabzadeh E, Diamond M (2007) Shifts in coding properties and maintenance of information transmission during adaptation in barrel cortex. PLoS Biol 5: e19.
33. Hong S, Arcas BA, Fairhall AL (2007) Single neuron computation: from dynamical system to feature detector. Neural Comput 112: 3133–3172.
34. Ruderman DL, Bialek W (1994) Statistics of natural images: scaling in the woods. Phys Rev Lett 73: 814–817.
35. van Hateren JH (1997) Processing of natural time series of intensities by the visual system of the blowfly. Vision Res 37: 3407–3416.
36. Simoncelli EP, Olshausen BA (2001) Natural image statistics and neural representation. Annu Rev Neurosci 24: 1193–1216.
37. Sharpee T, Rust N, Bialek W (2004) Analyzing neural responses to natural signals: Maximally informative dimensions. Neural Computation 16: 223–250.
38. Chubb C, Econopouly J, Landy MS (1994) Histogram contrast analysis and the visual segregation of iid textures. J Opt Soc Am A 11: 2350–2374.
39. Chubb C, Landy MS, Econopouly J (2004) A visual mechanism tuned to black. Vision Research 44: 3223–3232.
40. Tkačik G, Prentice JS, Victor JD, Balasubramanian V (2010) Local statistics in natural scenes predict the saliency of synthetic textures. Proc Natl Acad Sci USA 107: 18149–18154.
41. Theunissen FE, Miller JP (1995) Temporal encoding in nervous systems: a rigorous definition. J Comput Neurosci 2: 149–162.
42. Brenner N, Strong SP, Koberle R, Bialek W, de Ruyter van Steveninck RR (2000) Synergy in a neural code. Neural Computation 12: 1531–1552.
43. Narayan R, Nityananda R (1986) Maximum entropy image restoration in astronomy. Ann Rev Astron Astrophys 24: 127–170.
44. Victor JD, Johannesma P (1986) Maximum-entropy approximations of stochastic nonlinear transductions: an extension of the wiener theory. Biol Cybern 54: 289–300.
45. Schneidman E, Berry MJ, Segev R, Bialek W (2006) Weak pairwise correlations imply strongly correlated network states in a neural population. Nature 440: 1007–1012.
46. Shlens J, Field GD, Gauthier JL, Grivich MI, Petrusca D, et al. (2006) The structure of multi-neuron firing patterns in primate retina. J Neurosci 26: 8254–8266.
47. Tang A, Jackson D, Hobbs J, Chen W, Smith JL, et al. (2008) A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro. J Neurosci 28: 505–518.
48. Roudi Y, Tyrcha J, Hertz J (2009) Ising model for neural data: model quality and approximate methods for extracting functional connectivity. Phys Rev E 79: 051915.
49. Roudi Y, Aurell E, Hertz J (2009) Statistical physics of pairwise probability models. Front in Comp Neuro 3: 22.
50. Roudi Y, Nirenberg S, Latham PE (2009) Pairwise maximum entropy models for studying large biological systems: When they can work and when they can't. PLoS Comput Biol 5: e1000380.
51. Malouf R (2002) A comparison of algorithms for maximum entropy parameter estimation. In: Proceedings of the Conference on Natural Language Learning. pp 49–55.
52. Berger AL, Pietra SAD, Pietra VJD (1996) A maximum entropy approach to natural language processing. Computational Linguistics 22: 39–71.
53. Berger T (1971) Rate distortion theory: Mathematical basis for data compression. New Jersey: Prentice Hall.
54. Tishby N, Pereira FC, Bialek W (1999) The information bottleneck method. In: Proceedings of the 37-th Annual Allerton Conference on Communication, Control and Computing. pp 368–377.
55. Li Z (2006) Theoretical understanding of the early visual processes by data compression and data selection. Network: Computation in neural systems 17: 301–334.
56. Laughlin SB (1981) A simple coding procedure enhances a neuron's information capacity. Z Naturf 36c: 910–912.
57. Sharpee TO, Bialek W (2007) Neural decision boundaries for maximal information transmission. PLoS One 2: e646.
58. Fitzgerald JD, Sharpee TO (2009) Maximally informative pairwise interactions in networks. Phys Rev E 80: 031914.
59. Nikitin AP, Stocks NG, Morse RP, McDonnell MD (2009) Neural population coding is optimized by discrete tuning curves. Phys Rev Lett 103: 138101.
60. Sincich LC, Adams DL, Economides JR, Horton JC (2007) Transmission of spike trains at the retinogeniculate synapse. J Neurosci 27: 2683–2692.