OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Inferring Nonlinear Neuronal Computation Based on Physiologically Plausible Inputs

James M. McFarland*, Yuwei Cui, Daniel A. Butts

Department of Biology and Program in Neuroscience and Cognitive Science, University of Maryland, College Park, Maryland, United States of America

Abstract

The computation represented by a sensory neuron's response to stimuli is constructed from an array of physiological processes both belonging to that neuron and inherited from its inputs. Although many of these physiological processes are known to be nonlinear, linear approximations are commonly used to describe the stimulus selectivity of sensory neurons (i.e., linear receptive fields). Here we present an approach for modeling sensory processing, termed the Nonlinear Input Model (NIM), which is based on the hypothesis that the dominant nonlinearities imposed by physiological mechanisms arise from rectification of a neuron's inputs. Incorporating such 'upstream nonlinearities' within the standard linear-nonlinear (LN) cascade modeling structure implicitly allows for the identification of multiple stimulus features driving a neuron's response, which become directly interpretable as either excitatory or inhibitory. Because its form is analogous to an integrate-and-fire neuron receiving excitatory and inhibitory inputs, model fitting can be guided by prior knowledge about the inputs to a given neuron, and elements of the resulting model can often result in specific physiological predictions. Furthermore, by providing an explicit probabilistic model with a relatively simple nonlinear structure, its parameters can be efficiently optimized and appropriately regularized. Parameter estimation is robust and efficient even with large numbers of model components and in the context of high-dimensional stimuli with complex statistical structure (e.g. natural stimuli). We describe detailed methods for estimating the model parameters, and illustrate the advantages of the NIM using a range of example sensory neurons in the visual and auditory systems. We thus present a modeling framework that can capture a broad range of nonlinear response functions while providing physiologically interpretable descriptions of neural computation.

Citation: McFarland JM, Cui Y, Butts DA (2013) Inferring Nonlinear Neuronal Computation Based on Physiologically Plausible Inputs. PLoS Comput Biol 9(7): e1003143. doi:10.1371/journal.pcbi.1003143

Editor: Matthias Bethge, University of Tübingen and Max Planck Institute for Biologial Cybernetics, Germany

Received October 3, 2012; **Accepted** June 1, 2013; **Published** July 18, 2013

Funding: All authors were supported by NSF IIS-0904430 (www.nsf.gov). 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: jmmcfarl@umd.edu

Introduction

Sensory perception in the visual and auditory systems involves the detection of elemental features such as luminance and sound intensity, and their subsequent processing into more abstract representations such as "objects" that comprise our perception. The neuronal computations performed during such sensory processing must be nonlinear in order to generate more complex stimulus selectivity, such as needed to encode the conjunction of multiple sensory features [1–3] as well as to develop invariance to irrelevant aspects of the raw sensory input [4,5]. While these computations can appear inscrutably complex, they are necessarily constructed from the underlying neural circuitry, which exhibits several well-known and relatively straightforward nonlinear properties.

Nevertheless, characterizations of sensory neurons still typically rely on the assumption of linear stimulus processing, which is often implicit in standard approaches such as spike-triggered averaging and – more recently – generalized linear models (GLMs) [6–8]. While such descriptions can often provide good predictions of the neuronal response [9–11], they necessarily leave out the nonlinear elements of neuronal processing that likely play a major role in building the sensory percept.

Unfortunately, the space of possible nonlinear models is not bounded. While one might be inclined to incorporate details of the system and circuitry in question, more complicated models require more data for parameter estimation, and often involve poorly behaved or intractable optimization problems. As a result, practical nonlinear modeling approaches must make assumptions that limit the space of functions considered by restricting to a defined set of nonlinear interactions.

Several different approaches have been developed in this regard. The most common is to identify a low dimensional "feature space" to which the neuron is sensitive, with the assumption that its firing rate depends on a nonlinear function applied only to these stimulus features. Prominent examples of this approach include spike-triggered covariance (STC) analysis [12,13], which uses the covariance of the stimuli that elicit spikes, and information-theoretic approaches such as maximally informative dimensions (MID) analysis [14] and iSTAC [15]. With the subspace determined, other methods can be used to estimate a nonlinear mapping between the projection of the stimulus onto this low dimensional feature space and the firing rate [13–19].

A second general approach is to assume the form of nonlinearities present, most commonly based on a second-order approximation of the nonlinear stimulus-response relationship, as with the Wiener-Volterra expansion [20–24], and more recent versions cast in a probabilistic context [25–28]. This category might also encompass neural network approaches, which characterize the

Author Summary

Sensory neurons are capable of representing a wide array of computations on sensory stimuli. Such complex computations are thought to arise in large part from the accumulation of relatively simple nonlinear operations across the sensory processing hierarchies. However, models of sensory processing typically rely on mathematical approximations of the overall relationship between stimulus and response, such as linear or quadratic expansions, which can overlook critical elements of sensory computation and miss opportunities to reveal how the underlying inputs contribute to a neuron's response. Here we present a physiologically inspired nonlinear modeling framework, the 'Nonlinear Input Model' (NIM), which instead assumes that neuronal computation can be approximated as a sum of excitatory and suppressive 'neuronal inputs'. We show that this structure is successful at explaining neuronal responses in a variety of sensory areas. Furthermore, model fitting can be guided by prior knowledge about the inputs to a given neuron, and its results can often suggest specific physiological predictions. We illustrate the advantages of the proposed model and demonstrate specific parameter estimation procedures using a range of example sensory neurons in both the visual and auditory systems.

stimulus-response relationship in terms of a set of fixed nonlinear basis functions, using either generic network elements [29,30] or more specific nonlinear models of upstream sensory processing [31,32].

A final commonly used approach assumes that relevant nonlinearities can be captured by directly augmenting the linear model to account for specific response properties, such as the addition of refractoriness to account for neural precision [7,8,33–35], feedback terms that account for adaptation to contrast [36–38], and other nonlinearities to capture response properties such as sensitivity to stimulus intensity and local context [25].

Here, we present a probabilistic modeling framework inspired by all of these approaches, the 'Nonlinear Input Model' (NIM), which limits the space of nonlinear functions by assuming that nonlinearities in sensory processing are dominated by spike generation, resulting in both rectification of the inputs to the neuron, as well as rectification of the neuron's output. By assuming a neuron's inputs are rectified, the NIM implicitly describes neuronal processing as a sum over excitatory and inhibitory inputs, which is increasingly being seen as an important factor in sensory processing [39–43]. The NIM expands directly on the GLM framework, and is able to utilize recent advances in the statistical modeling of neural responses [7,8,17,44,45], including the ability to model spike-refractoriness [7,8] and multi-neuron correlations [8,44].

As we show here, this results in a parsimonious nonlinear description of a range of neurons in both the visual and auditory systems, and has several advantages over previous approaches. Because of its relatively simple model structure, parameter estimation is well-behaved and makes efficient use of the data, even when the number of relevant inputs is large and/or the stimulus is high-dimensional. Importantly, because its form is based on an integrate-and-fire neuron, model selection and parameter estimation can be guided by specific knowledge about the inputs to a given neuron, and the elements of the resulting model can often be related to specific physiological predictions. The NIM thus provides a powerful and general approach for nonlinear modeling that complements other methods that rely on more abstract formulations of nonlinear computation.

Results

Nonlinear combination of multiple inputs: ON-OFF retinal ganglion cells

Perhaps the greatest success of linear models is in the retina, where it has been used primarily to describe the spike responses of retinal ganglion cells (RGCs) [10,11,16]. For a given RGC, estimating the components of the linear model typically involves measuring its spiking response to a noise stimulus, and then computing the average stimulus that preceded its spikes: the spike-triggered average (STA). The STA linear filter can produce very good response predictions for typical RGCs under stationary stimulus conditions, but clearly fails for ON-OFF cells (commonly found in rodents), which respond to both increases and decreases of light intensity [46–49]. This failure for ON-OFF cells occurs simply because the STA identifies only a single stimulus dimension, and averages out the opposing stimulus features that evoke ON and OFF responses.

To explore this situation, we construct a basic model of an ON-OFF RGC, which receives separate ON and OFF inputs (Fig. 1A). If these two inputs were to combine linearly, their effect would be identical to that of a single input generated by the sum of the two stimulus filters, i.e., $(\mathbf{s}\cdot\mathbf{k}_{\mathrm{ON}})+(\mathbf{s}\cdot\mathbf{k}_{\mathrm{OFF}}) = \mathbf{s}\cdot(\mathbf{k}_{\mathrm{ON}}+\mathbf{k}_{\mathrm{OFF}}) = \mathbf{s}\cdot\mathbf{k}_{\mathrm{SUM}}$. Here the stimulus $\mathbf{s}$ at a particular time is represented as a vector (which in general includes time-lagged elements to account for stimulus history) such that the operation of a linear filter $\mathbf{k}$ is given by a dot product. Because of the averaging implicit in linear processing, a nonlinear transformation must be applied to each input in order to enable the model ON-OFF neuron to respond to both types of stimuli: i.e., $f(\mathbf{s}\cdot\mathbf{k}_{\mathrm{ON}})+f(\mathbf{s}\cdot\mathbf{k}_{\mathrm{OFF}})$. These $f(.)$ are taken to be rectifying functions (Fig. 1A), as seen experimentally [50,51], and as modeled in [52–54]. As a result, the response of the neuron to increases or decreases of luminance is dominated by the ON or OFF pathways respectively (Fig. 1B), producing a response that is selective to both ON and OFF stimulus dimensions. As expected, the STA (Fig. 1C) for this neuron does not match either the ON or OFF stimulus filters, but rather reflects their average.

Thus, this is a clear example where nonlinear characterization is necessary to capture the RGC's stimulus selectivity. One such approach that has been applied to ON-OFF cells is spike-triggered covariance (STC) analysis [49,55], which identifies stimulus dimensions along which the variance of the spike-triggered ensemble is either increased or decreased relative to the stimulus distribution [12,13]. For the example neuron in Fig. 1, STC analysis identifies a stimulus dimension along which the variance of the spike-triggered ensemble is expanded (Figs. 1D, E). While neither the STA nor STC filters correspond to the true ON or OFF filters, together they define a stimulus subspace that contains the true filters (Fig. 1E).

Given the dimensionality reduction achieved in determining the STC subspace (or with other subspace identification methods), it is possible in principle to completely characterize the neural response function, i.e., $r = F[\mathbf{k}_1\cdot\mathbf{s}, \mathbf{k}_2\cdot\mathbf{s}]$. In two-dimensions, such as in this example, this nonlinear mapping from the subspace to a firing rate can be estimated non-parametrically [18,56] given enough data, and potentially approximated in higher dimensions [13,15,18,19,57].

However, even if accurate estimation of this nonlinear mapping were possible, such functions are difficult to interpret, even when arising from the conjunction of simpler components. For example, in our simulated ON-OFF RGC, neither the STA/STC filters

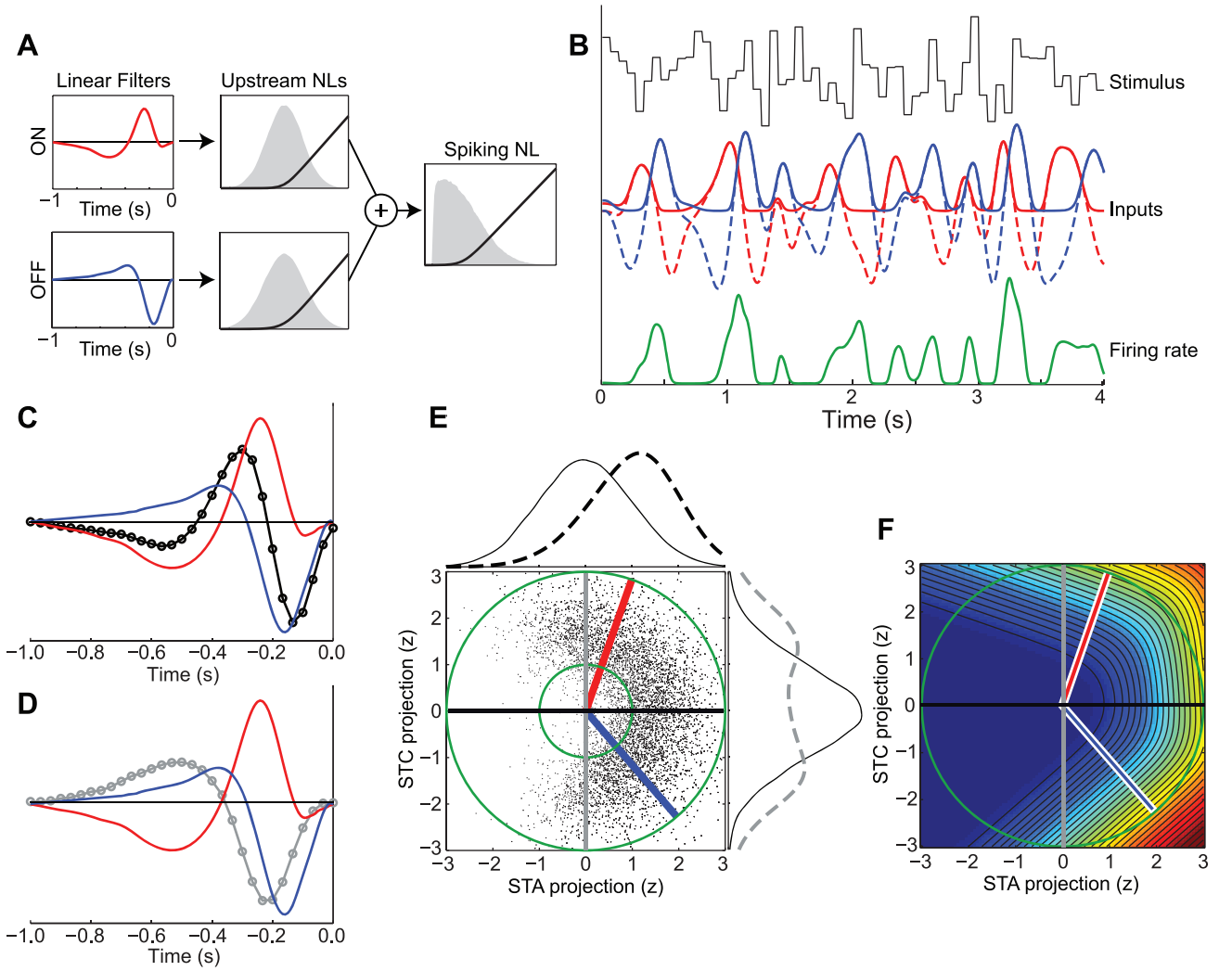

Figure 1. ON-OFF RGC simulation. A) Schematic showing the ON (top, red) and OFF (bottom, blue) inputs to the simulated ON-OFF RGC. The temporal filters (left) process the stimulus, and the upstream nonlinearities (black, middle) are then applied to the filter outputs. The sum of the two inputs is then passed through the spiking nonlinearity (black, right). The distributions of the stimulus filtered by the ON and OFF pathways, as well as the distribution of their summed input to the simulated neuron are shown as gray shaded regions. **B**) A simulation showing how the response to a 15 Hz (Gaussian) white noise stimulus is constructed. The stimulus (black) is filtered by the ON and OFF temporal kernels (dashed red and blue), and then transformed by the upstream nonlinearities (solid red and blue). The resulting instantaneous firing rate (green) is given by the sum of these inputs passed through the spiking nonlinearity. **C**) The STA (black) resembles the average of separate ON (red) and OFF (blue) filters of the generative model. **D**) Similar to panel C, the first STC filter (gray) resembles a mixture of the ON and OFF filters. **E**) Stimuli eliciting spikes (black dots) are projected onto the two-dimensional subspace spanned by the STA and first STC filter, shown in units of z-score. The distributions of stimuli corresponding to spikes (dashed lines on top and left) are compared to the marginal stimulus distributions (solid), demonstrating a systematic bias along the STA (horizontal) axis, and an increased variance along the STC (vertical) axis. The true ON and OFF filters (red and blue) are also contained in the STA/STC subspace, as indicated by the red and blue lines lying on the unit circle (green). The inner green circle has a radius of one standard deviation. **F**) The neuron's firing rate as a function of the stimulus projected into the 2-D STA/STC subspace (shaded color depicts firing rate: increasing from blue to red).
doi:10.1371/journal.pcbi.1003143.g001

themselves nor the measured nonlinear mapping make it clear that the response is generated from separate inputs with relatively straightforward nonlinearities.

This example thus motivates the modeling framework that we present here, the Nonlinear Input Model (NIM), which describes a neuron's stimulus processing as a sum of nonlinear inputs, following the structure of the generative model shown in Fig. 1. Below, we first present procedures for estimating the parameters of the NIM before demonstrating its ability to recover the inputs to the ON-OFF RGC, as well as its application to a range of other simulated and measured data from both visual and auditory brain areas.

Parameter estimation for the Nonlinear Input Model (NIM)

The computational challenges associated with parameter estimation are a significant barrier to the successful development and application of nonlinear models of sensory processing. In the standard linear-nonlinear (LN) model, the neuron's response is modeled by an initial stage of linear stimulus filtering, followed by a static nonlinear function ("spiking nonlinearity") that maps the output to a firing rate (Fig. 2A). The more recent adaptation of probabilistic models based on spike train likelihoods, such as in the Generalized Linear Model (GLM) [6–8], allows for integration of

other aspects of neuronal processing into the linear stimulus-processing framework, and can be used to model nonlinear stimulus processing through predefined nonlinear transformations [31,32,58]. Importantly, this approach also provides a foundation for parameter estimation for the NIM.

A principal motivation for the NIM structure is that if the neuronal output at one level is well described by an LN model, downstream neurons will receive inputs that are already rectified (or otherwise nonlinearly transformed). Thus, we use LN models to represent the inputs to the neuron in question, and the neuron's response is given by a summation over these LN inputs followed by the neuron's own spiking nonlinearity (Fig. 2B). Importantly, this allows us to account for the rectification of a neuron's inputs imposed by the spike-generation process. The NIM can thus be viewed as a 'second-order' generalization of the LN model, or an LNLN cascade [59,60]. Previous work from our lab [45] cast this model structure in a probabilistic form, and suggested several statistical innovations in order to fit the models using neural data [45,61,62]. Here, we present a general and detailed framework for NIM parameter estimation that greatly extends the applicability of the model. This model structure has also been suggested for applications outside of neuroscience in the form of projection pursuit regression [63], including generalizations to response variables with distributions from the exponential family [64].

The processing of the NIM is comprised of three stages (Fig. 2C): (a) the filters $\mathbf{k}_i$ that define the stimulus selectivity of each input; (b) the static 'upstream' nonlinearities $f_i(.)$ and corresponding linear weights w_i which determine how each input contributes to the overall response; and (c) the spiking nonlinearity $F[.]$ applied to the linear sum over the neuron's inputs. The predicted firing rate $r(t)$ is then given as:

$$r(t) = F\left[\sum_i \left(w_i f_i(\mathbf{k}_i \cdot \mathbf{s}(t))\right) + \mathbf{h} \cdot \mathbf{x}(t)\right], \quad (1)$$

where $\mathbf{s}(t)$ is the (vector-valued) stimulus at time t, $\mathbf{x}(t)$ represents any additional covariates (such as the neuron's own spike history), and $\mathbf{h}$ is a linear filter operating on $\mathbf{x}$. Note that equation (1) reduces to a GLM when the $f_i(.)$ are linear functions. The w_i can also be extended to include temporal convolution of the subunit contributions to model the time course of post-synaptic responses associated with individual inputs [45], as well as 'spatial' convolutions to account for multiple spatially distributed inputs with similar stimulus selectivity [65]. Since equivalent models can be produced by rescaling the w_i, and $f_i(.)$ (see Methods), we constrain the subunit weights w_i to be either $+/-1$. Because we generally assume the $f_i(.)$ are rectifying functions, the w_i thus specify whether each subunit will have an 'excitatory' or 'inhibitory' influence on the neuron.

Parameter estimation for the NIM is based on maximum likelihood (or maximum a posteriori) methods similar to those used with the GLM [6–8]. Assuming that the neuron's spikes are described in discrete time by a conditionally inhomogeneous Poisson count process with rate function $r(t)$, the log-likelihood (LL) of the model parameters given an observed set of spike counts $R_{\rm obs}(t)$ is given (up to an overall constant) by:

$$LL = \sum_t \left(R_{obs}(t) \log r(t) - r(t)\right). \quad (2)$$

To find the set of parameters that maximize the likelihood (eq. 2), we adapt methods that allow for efficient parameter optimization of the GLM [7]. First, we use a parametric spiking nonlinearity given by $F[x] = \alpha \log[1+\exp(\beta(x-\theta))]$, with scale α, shape β, and offset θ. Other functions can be used, so long as they satisfy conditions specified in [7]. This ensures that the likelihood surface will be concave with respect to linear parameters inside the spiking nonlinearity [7], and in practice will be well-behaved for other model parameters (see Fig. S1; Methods).

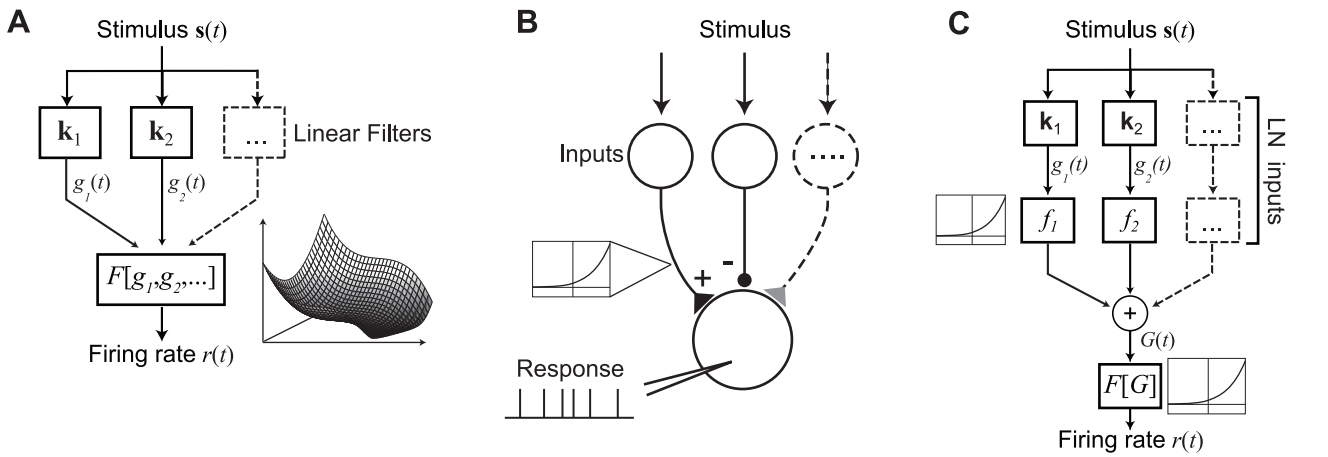

Figure 2. Schematic of LN and NIM structures. A) Schematic diagram of an LN model, with multiple filters (k_1, k_2, …) that define the linear stimulus subspace. The outputs of these linear filters (g_1, g_2, …) are then transformed into a firing rate prediction $r(t)$ by the static nonlinear function $F[g_1,g_2,\ldots]$, depicted at right for a two-dimensional subspace. Note that while the general LN model thus allows for a nonlinear dependence on multiple stimulus dimensions, estimation of the function $F[.]$ is typically only feasible for low (one- or two-) dimensional subspaces. B) Schematic illustration of a generic neuron that receives input from a set of 'upstream' neurons that are themselves driven by the stimulus s. Each of the upstream neurons provides input to the model neuron that is generally rectified due to spike generation (inset at left), and thus is either excitatory or inhibitory. The model neuron then integrates its inputs and produces a spiking output. C) Block diagram illustrating the structure of the NIM, based on (B). The set of inputs are represented as (one-dimensional) LN models, with a corresponding stimulus filter k_i, and "upstream nonlinearity" $f_i(.)$. These inputs are then linearly combined, with weights w_i, and fed into the spiking nonlinearity $F[.]$, resulting in the predicted firing rate $r(t)$. The NIM thus has a 'second-order LN' structure (or LNLN), with the neuron's own nonlinear processing shaped by the LN nature of its inputs.
doi:10.1371/journal.pcbi.1003143.g002

Because it is straightforward to estimate the linear term $\mathbf{h}$, and the w_i are constrained to be $+/-1$, the upstream nonlinearities $f_i(.)$ and the stimulus filters $\mathbf{k}_i$ are the key components that must be fit in the NIM. While it is typically not feasible to optimize the likelihood with respect to both sets of parameters simultaneously, an efficient strategy is to use block coordinate ascent [66], alternating between optimizing the $\mathbf{k}_i$ and $f_i(.)$, in each case holding the remaining set of parameters constant (see Methods). 'Linear' parameters, such as $\mathbf{h}$ and θ, can be optimized simultaneously during either (or both) optimization stages.

While the set of 'upstream nonlinearities' $f_i(.)$ can be represented as parametric functions such as rectified-linear or quadratic functions (see Methods), a powerful approach is to represent them as a linear combination of basis functions $\varphi_j(.)$ such as piecewise linear "tent" basis functions, i.e., $f_i(g) = \Sigma_j\, a_{ij}\varphi_j(g)$ [17,45]. In doing so, estimation of the upstream nonlinearities reduces to estimating linear parameters a_{ij} inside the spiking nonlinearity, with a single global optimum of the likelihood function for a given set of stimulus filters $\mathbf{k}_i$.

For a fixed set of upstream nonlinearities, the stimulus filters $\mathbf{k}_i$ can be similarly optimized, although the resulting likelihood surface will not in general be convex because the $\mathbf{k}_i$ operate inside the upstream nonlinearities. Nevertheless, we have found that in practice their optimization is well-behaved and that local minima can be avoided with appropriate optimization procedures (Fig. S1; see Methods). Furthermore, it is straightforward to evaluate the likelihood function and its gradient with respect to the $\mathbf{k}_i$ analytically (see Methods), allowing for efficient gradient-based optimization.

Thus, optimal parameter estimates for the NIM can be determined efficiently, even for models with large numbers of parameters (see examples below). The time required for filter estimation (typically the most time-consuming step) scales approximately linearly with the experiment duration, the dimensionality of the stimulus, and the number of model subunits (Fig. S2). This is very favorable compared with alternative nonlinear modeling approaches such as MID [14], which require using simulated annealing and quickly becomes intractable as the number of filters and/or stimulus dimensions is increased.

Furthermore, because the NIM provides an explicit probabilistic model for the neuronal spike response, regularization of the model components can be incorporated without adversely affecting the behavior of the optimization problem [7] (see Methods). This is particularly important when optimizing high-dimensional spatiotemporal filters and/or models with many inputs, which are both discussed further below. Likewise, as with other probabilistic modeling approaches – but not those relying on spike-triggered measurements [67] – the model can be optimized using data recorded with natural stimulus ensembles (containing complex correlation structure, and non-Gaussian distributions) without introducing biases into the parameter estimates.

The NIM thus provides a nonlinear modeling framework in which large numbers of parameters can be efficiently estimated using data recorded with arbitrarily complex stimulus ensembles. In addition to this flexibility, the NIM provides model fits that are more directly interpretable due to its physiologically motivated model structure. To illustrate these advantages, below we first apply the NIM to the example ON-OFF RGC from Fig. 1, and then demonstrate its wide applicability on recorded and simulated neurons in several different sensory areas.

Nonlinear models of the ON-OFF retinal ganglion cell example

Returning to the example ON-OFF RGC (Fig. 1), the NIM is a natural choice given that its structure matches that of the simulated neuron. Using the estimation procedures described above, the NIM is able to successfully capture the true stimulus selectivity of its individual inputs (Fig. 3A), including the 'upstream nonlinearities' associated with each input, as well as the form of the spiking nonlinearity (see Methods).

This example thus illustrates the core motivation behind the NIM of modeling a neuron's stimulus processing in terms of rectified neuronal inputs. While the structure of the simulated RGC neuron in this example may appear to be a convenient choice, its form is consistent with other models of ON-OFF processing [48,53], and with models of RGCs in general [54,68].

Thus, to understand the advantages and disadvantages of the NIM structure, it is useful to compare it with the dominant alternative approach for describing nonlinear stimulus processing: "quadratic models". Such models have recently been cast in an information-theoretic context [15,27,28], as well as in the form of an explicit probabilistic model [26] which has been referred to as the 'Generalized Quadratic Model' (GQM). The GQM can be viewed as a probabilistic generalization of STA/STC analysis [26] and of the second-order Wiener-Volterra expansion [20]. The GQM can also be written in the form of a NIM where the upstream nonlinearities $f_i(.)$ are fixed to be linear or squared functions:

$$r(t) = F\left[\mathbf{k}_L\mathbf{s} + \mathbf{s}^T\mathbf{C}\mathbf{s}\right] \approx F\left[\mathbf{k}_L\mathbf{s} + \sum_{i=1}^{M} w_i(\mathbf{k}_i\cdot\mathbf{s})^2\right], \qquad (3)$$

where $\mathbf{k}_L$ is a linear filter, and the M squared filters $\mathbf{k}_i$ generally provide a low-rank approximation to the quadratic component $\mathbf{C}$ [26]. In this sense, the probabilistic framework described here is easily extended to encompass quadratic models, providing a means for direct comparison between different nonlinear structures.

For the ON-OFF RGC, the GQM finds one linear and two quadratic filters, all of which are contained in the two-dimensional subspace identified by STC analysis, meaning that the GQM filters are also linear combinations of the true ON and OFF filters (Fig. 3B). Note that while two filters are sufficient to span the relevant stimulus subspace, the third GQM filter provides an additional degree of freedom to capture the best quadratic approximation to the underlying 'neural response function' (Fig. 3E).

Although in this example the resulting quadratic function cannot completely capture the form of the response function constructed from rectified inputs, we note that it still provides a good approximation, as shown by only modest reductions in model performance compared to the NIM (Fig. 3F). However, as expected from a second-order Taylor series expansion, such an approximation breaks down further from the "origin" of the subspace. Thus, the quadratic approximation will typically be less robust for stimuli with heavy-tailed distributions such as those associated with natural stimuli [69–71]. To illustrate this point we performed simulations of the same ON-OFF RGC presented with white noise stimuli having a Student's t-distribution, where the tail thickness was controlled by varying the number of degrees of freedom (Fig. 3G). The improved performance of the NIM over the GQM is indeed substantially enhanced for stimulus distributions with heavier tails (Fig. 3H). We also verified similar effects for a range of simulated neurons (data not shown).

We emphasize that one of the key advantages of the NIM over previously described methods is that it provides a more interpretable picture of stimulus processing as a sum of rectified neuronal inputs. As we demonstrate through several examples below in both the visual and auditory systems, it appears that sensory computation by neurons will often adhere to this general

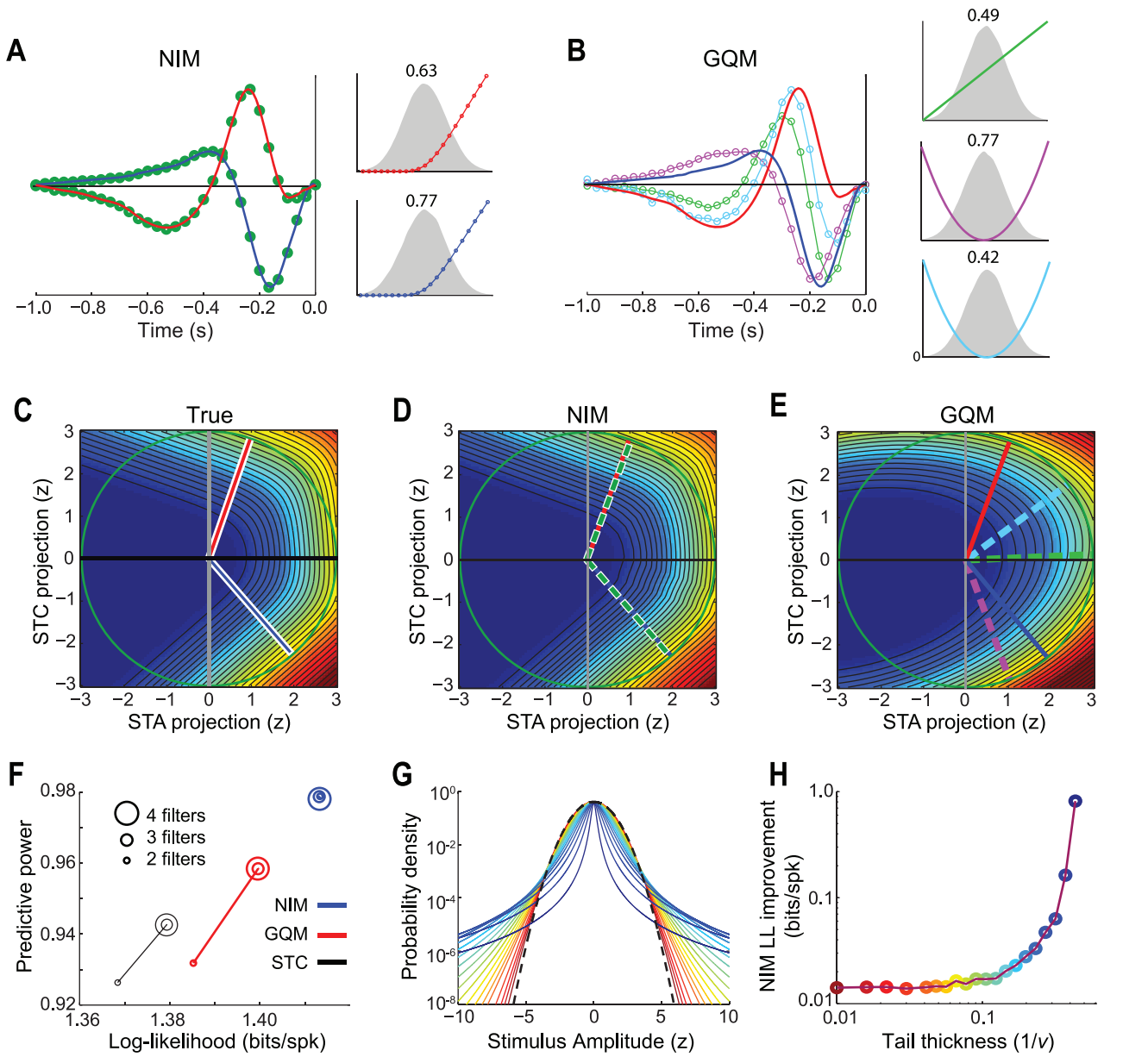

Figure 3. Comparison of NIM and quadratic model. A) The filters fit by the NIM (green dots) are able to capture the true underlying ON and OFF filters (red and blue), as well as the shape of the upstream nonlinearities (right), which are shown relative to the corresponding distributions of the filtered stimulus (gray shaded). The ranges of the y-axes for different subunits are indicated by the numbers above, for comparison of their relative magnitudes. These 'subunit weights' are scaled so that their squared magnitude is one. **B**) The filters fit by the GQM, consisting of two (excitatory) squared filters (magenta and light blue) and a linear filter (green trace), are different than the true filters (red and blue), but are in the same subspace, as demonstrated in (E). **C**) The simulated neuron's response function (shaded color depicts firing rate) and true filters (red and blue) projected into the STC subspace (identical to Fig. 1F). **D**) Response function predicted by the NIM. The filters identified by the NIM (dashed green) overlay onto the true filters. **E**) Same as (D) for the GQM, with colored lines corresponding to the filters in (B). **F**) Model performance is plotted for the STC, GQM, and NIM fit with different numbers of filters (indicated by different circle sizes). Log-likelihood (relative to the null model) is shown on the x-axis, and the 'predictive power' [99] is shown on the y-axis; both were evaluated on a simulated cross-validation data set. The NIM (blue) outperforms the GQM (red), both of which outperform a nonlinear model based on the STC filters (black, see Methods). The STC model and GQM achieve maximal performance with 3 filters, since this is sufficient for capturing the best-fit quadratic function in the relevant 2-D stimulus subspace, while the NIM achieves optimal performance with two filters, as expected. **G**) To determine how model performance depends on the stimulus distribution we simulated the same neuron's response to white noise luminance stimuli with Student's *t*-distributions, ranging from Gaussian (i.e., $\nu = \infty$, dashed black) to "heavy tailed" (decreasing ν from red to blue). **H**) The log-likelihood improvement of the NIM over the GQM increases as a function of the tail thickness (parameterized by $1/\nu$) of the stimulus distribution (which also determines the tail thickness of the filtered stimulus distributions). The GQM is able to provide a very close approximation for large values of ν (i.e., a more normally-distributed stimulus), but has lower performance compared to the NIM for more heavy-tailed stimuli.

doi:10.1371/journal.pcbi.1003143.g003

form, which is motivated primarily by physiological, rather than mathematical, considerations.

Inferring the interplay of excitation and inhibition in visual and auditory neurons

One of the main advantages of the NIM structure is the ability to specifically model the effects of inhibitory inputs, which are increasingly being shown to have a large impact on neuronal processing in many sensory areas [72–74]. Indeed, the NIM generates predictions of the functional tuning of excitation and inhibition, and provides insight into their role in sensory processing. To demonstrate this, we apply the NIM to example neurons from visual and auditory areas.

We first consider an example cat LGN neuron during the presentation of natural movies [45,75,76]. Accurate characterization of LGN processing poses substantial challenges for previous nonlinear approaches, due to the high temporal resolution of LGN responses in this context [77] combined with the large number of spatial dimensions of the stimulus. As a result, previous nonlinear applications have either utilized lower temporal resolutions [78,79] or parametric models of the spatial processing [38,45,80]. The methods described here allow for (appropriately regularized) spatiotemporal receptive fields (STRFs) of LGN neurons to be fit at sufficiently high resolution, using natural movies. We find that the response of the example LGN neuron consists of an excitatory receptive field that is delayed relative to the linear STRF (Fig. 4A), along with a second, more delayed 'suppressive' receptive field (Fig. 4B), corresponding to putative inhibitory input. Unlike in previous studies, the tractability of the fitting procedures used here allows for high spatial and temporal resolution of the putative inputs (Fig. 4B), as well as the application of sparseness and smoothness regularization (see Methods). By comparison, the GQM identifies similar STRFs, but has worse performance (Fig. 4C), as well as a different nonlinear structure and resulting physiological interpretation (Fig. S3).

Next we consider an example neuron from zebra finch area MLd, as the animal is presented with conspecific bird songs [81–83]. These neurons respond to specific frequencies of the song input, and hence their stimulus selectivity can be characterized by a linear spectrotemporal receptive field (STRF) [9], which can be recovered in an unbiased manner using maximum-likelihood estimation (Fig. 5A) [84] despite the presence of higher order correlations in the stimulus. Application of the NIM to this example neuron again recovers both an excitatory and a temporally delayed suppressive component (Fig. 5B). The description of the neuron's stimulus tuning provided by the NIM is closely related to that given by the linear model, but instead of identifying positive and negative domains of the linear STRF as excitatory and suppressive, these effects are segregated into different nonlinear processing subunits, each individually rectified. The separate excitatory and suppressive inputs provide a more accurate description of the underlying stimulus processing than a single linear STRF, as demonstrated by the significantly improved model performance of the NIM compared with the LN model (Fig. 5C). As with the LGN example, the GQM identifies similar excitatory and suppressive filters as the NIM, but again provides a less physiologically interpretable description of the underlying computation (Fig. S4), and has comparable, if slightly reduced, performance (Fig. 5C).

Modeling complex neural response functions in terms of rectified inputs

Thus far we have only considered cases where the neuron's response is described by a NIM with a small number of inputs,

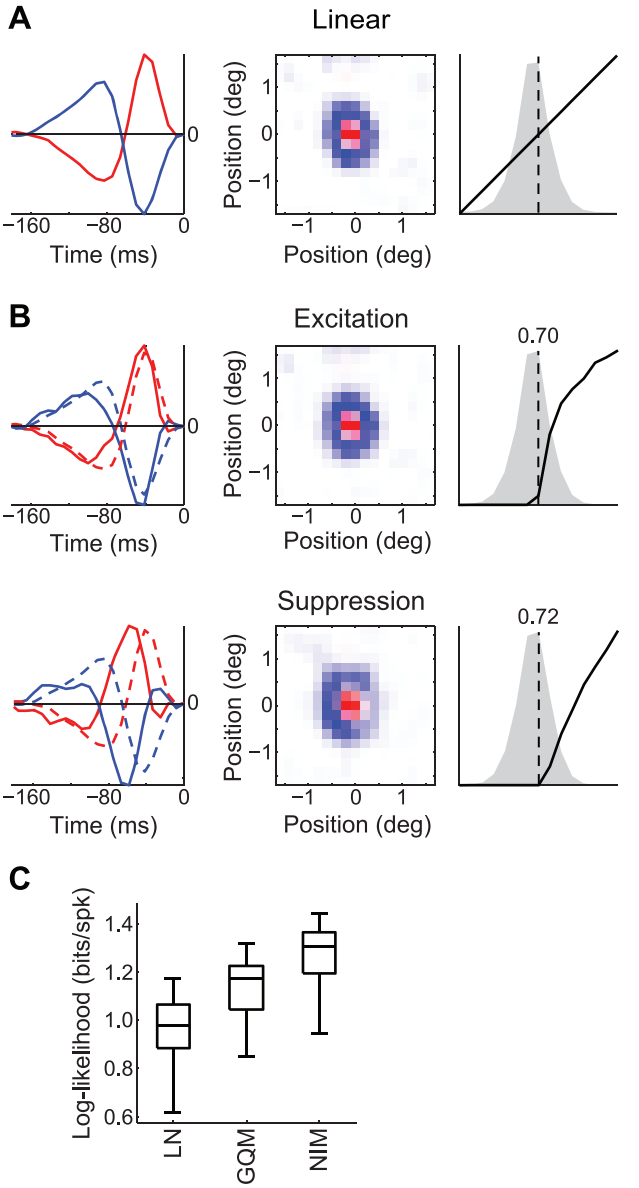

Figure 4. Spatiotemporal tuning of excitatory and suppressive inputs to an LGN neuron. A) The linear receptive field can be represented as the sum of two space-time separable components, corresponding to the receptive field center (red) and surround (blue). B) The NIM with excitatory (top) and suppressive (i.e., putative inhibitory, bottom) inputs. The excitatory and suppressive components (solid) both have slower, and less biphasic, temporal responses (left) compared with the linear model (dashed). The suppressive input is also delayed relative to the excitatory input. Both excitatory and suppressive inputs have roughly the same spatial profiles (middle), and both provide rectified input through the corresponding upstream nonlinearities (right). C) The NIM has significantly better performance, as measured by cross-validated log-likelihood, compared to the linear model ($p = 0.002$; $n = 10$ cross-validation sets; Wilcoxon signed rank test) and the GQM ($p = 0.002$).
doi:10.1371/journal.pcbi.1003143.g004

consistent with simpler stimulus processing in sub-cortical areas. In contrast, in the visual cortex, even V1 'simple cells' can exhibit selectivity to large numbers of stimulus dimensions [57,62]. Further, the dominant model of V1 'complex cells' is the nonlinear "Energy Model" [10,85–87], which posits quadratic

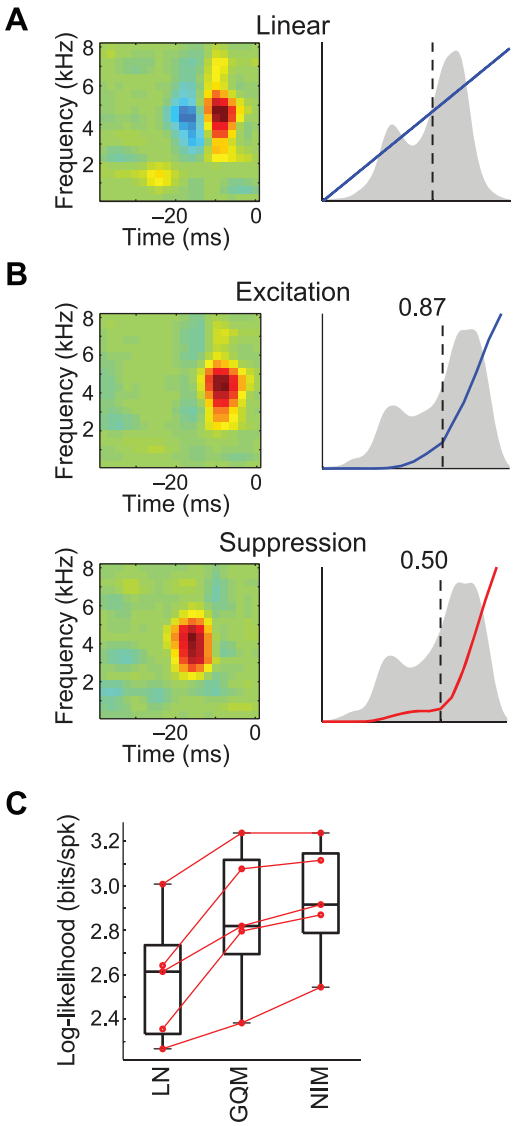

Figure 5. Spectrotemporal tuning of excitation and suppression in the songbird auditory midbrain. A) The linear spectrotemporal receptive field (STRF; left) contains two subfields of opposite sign. B) The excitatory (top) and suppressive (bottom) spectrotemporal filters identified by the NIM are similar to the positive and negative subfields of the linear STRF respectively. However, these inputs are both rectified by the upstream nonlinearities (right), resulting in different stimulus processing (see Fig. S4). C) Comparison of log-likelihoods of the LN model, GQM, and NIM. Red lines show the performance across models for each cross-validation set. Note that the duration of the recording, and the neuron's relatively low firing rate, limit the statistical power of model comparisons.
doi:10.1371/journal.pcbi.1003143.g005

stimulus processing that results in the response representing the amount of local, oriented, band-pass "stimulus energy". The Energy Model has been broadly tested [10,87], and is well supported by previous nonlinear modeling approaches [13,26,27,57,88]. While the Energy Model provides a functional description of stimulus processing for V1 complex cells, it is less clear how such stimulus selectivity is constructed, and how it is related to V1 simple cell processing. Here we demonstrate that the NIM can describe both simple and complex cell processing as a sum of rectified inputs, providing a basis for a unified description of visual cortical neuron computation [62].

We first consider two simulated V1 neurons in order to demonstrate the capacity for such a unified description, before applying the NIM to experimental data. We generate simulated data using a one-dimensional white-noise bar stimulus aligned with the simulated neurons' preferred spatial orientation (Fig. 6A), which is a common, relatively low-dimensional, stimulus used in nonlinear characterizations of V1 neurons [57,88,89]. The first simulated neuron's response is constructed as a sum of six rectified direction-selective inputs (Fig. 6B), consistent with the structure of the NIM, while the second neuron's response is constructed from four such inputs processed by a squaring nonlinearity, similar to the standard Energy Model of V1 complex cells [85].

For the neuron with rectified inputs, the NIM fitting procedure is indeed able to identify the true underlying stimulus filters and the form of the rectifying upstream nonlinearities (Fig. 6C). Additionally, while the optimal number of filters can be determined using the cross-validated model performance, the identified stimulus filters, and the resulting model performance itself, are relatively insensitive to specification of the precise number of model subunits (Fig. S5). This demonstrates the ability of the NIM to robustly identify even relatively complex stimulus processing, in cases where such processing arises from a sum of rectified inputs.

Furthermore, as with the ON-OFF RGC example above (Figs. 1 and 3), STC analysis of this simulated V1 neuron can identify the appropriate stimulus subspace, although not the true underlying filters (Fig. 6D). Because of the high dimensionality of the resulting subspace, however, it is more difficult to estimate the mapping from the subspace to the neuronal response compared with the ON-OFF example. The lack of alignment between the STA/STC filters and the true filters further complicates a straightforward interpretation of the estimated function.

By comparison, the GQM identifies filters with characteristics that more closely resemble those of the true input filters (e.g., more localized, fewer lobes). The improved performance of the GQM compared with an STC-based model (Fig. 6E) highlights the greater power and flexibility of a probabilistic modeling framework, particularly the importance of regularization. Nevertheless, the GQM filters still reflect non-trivial linear combinations of the true filters, as with the STC filters (Fig. 6E, *bottom*).

Of course, one would expect the NIM to outperform other models when the generative model is composed as a sum of rectified inputs. In a second simulated example, however, we illustrate the flexibility of the NIM in capturing other neural response functions. The second simulated neuron is constructed from four direction-selective inputs that are squared and summed together to generate a quadratic response function (Fig. 6F). The NIM is still able to identify the true generative model using pairs of rectified inputs with equal but opposite input filters to represent each quadratic filter (Fig. 6G). This representation is certainly not the most efficient in this case, as the GQM is able to identify the correct filters (Fig. 6I) using fewer parameters and a more straightforward estimation procedure.

These two simulated V1 examples thus illustrate the potential tradeoffs between the NIM and GQM. On the one hand, the NIM provides a more flexible framework that can capture a broader range of nonlinear stimulus processing. In fact, any response function can in principle be represented with this structure [90]. The NIM structure is also more appropriate for explicitly modeling neuronal inputs, and thus allows for more plausible physiological interpretation of its components. On the other hand, the GQM can capture the nonlinear mapping up to second order more efficiently, and identifies the relevant stimulus subspace robustly. This suggests the potential for combining these

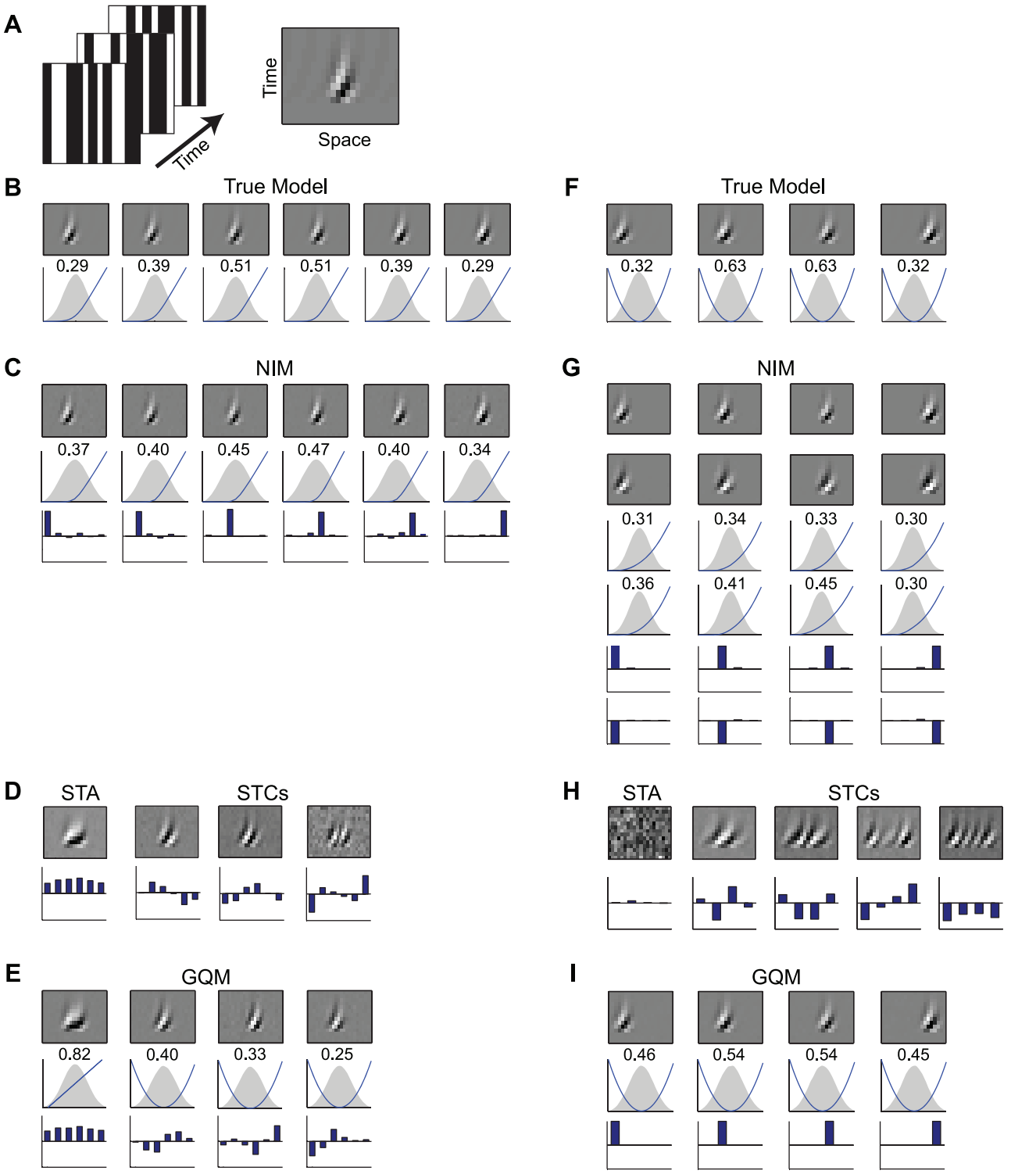

Figure 6. Modeling stimulus selectivity arising from many inputs. A) Simulated V1 neurons are presented with one-dimensional spatiotemporal white noise stimuli (left). Their stimulus processing is constructed from a set of spatiotemporal filters (example shown at right), depicted with one spatial dimension (x-axis) and time lag (y-axis). B) The first simulated neuron is constructed from six spatially overlapping direction-selective filters (top), similar to those observed experimentally for V1 neurons. Below, the corresponding filtered stimulus distributions are shown along with the respective upstream nonlinearities (blue). C) The NIM identifies the correct spatiotemporal filters (top), as well as the form of the upstream nonlinearities (middle). The projections of the NIM filters onto the true filters (bottom) illustrate that the NIM identifies the true filters. D) The STA for the simulated neuron (left), along with the three significant STC filters (right) are largely contained in the subspace spanned by the true filters, but reflect non-trivial linear combinations of these filters (bottom). E) The GQM is composed of a linear input (left) and three excitatory squared inputs (right). While the GQM filters are more similar to the true filters, they also represent non-trivial linear combinations of them (bottom). F) The second simulated neuron consists of four similar, but spatially shifted, inputs that are squared. G) The NIM represents each true (squared) input by an opposing pair of rectified inputs. H) The STA (left) does not show any structure because the neuron's response is, by construction, symmetric in the

stimulus. The four significant STC filters (right) represent distributed linear combinations of the four underlying filters. I) The GQM recovers the correct stimulus filters, given appropriate sparseness regularization.
doi:10.1371/journal.pcbi.1003143.g006

approaches when investigating complex neuronal processing, such as by using the GQM to identify the relevant stimulus subspace and provide initial estimates of the number and properties of NIM filters, followed by application of the NIM framework (see Methods; Figs. S1, S3).

Application of the NIM to recorded V1 neurons

While the simulated examples above allowed for model comparisons when the neurons' response functions were known, they also provide a foundation for understanding model fits to real V1 data. We first consider a V1 neuron recorded from an anesthetized macaque in the context of similar one-dimensional white noise stimuli [57]. While this neuron has a clear STA and is considered a simple cell by classical measures, STC analysis identifies two excitatory and six suppressive stimulus dimensions (based on inspection of the eigenvalue spectrum) in addition to the STA (Fig. 7A). In this case, the GQM identifies similar filters to STC analysis, although the application of smoothness and sparseness regularization allows it to resolve more realistic stimulus filters (Fig. 7B), and produce significantly improved model performance (Fig. 7D) compared to an STC-based model (see Methods).

We also fit a NIM with six excitatory and six suppressive stimulus filters, where the number of filters was selected based on cross-validated model performance (Fig. 7C; see also Fig. S5). As expected, these 12 filters span a stimulus subspace that is largely overlapping with the subspace identified by the GQM. However, the additional stimulus filters, and the inferred upstream nonlinearities associated with each subunit, allow the NIM to capture additional aspects of the neural response function that significantly improve the cross-validated model performance relative to the quadratic models (Figs. 7D, E). We also note that the NIM appears to identify a more consistent set of stimulus filters than the quadratic models.

Similar comparisons also come to light in when applying the models to V1 complex cells, even in the most demanding stimulus contexts. To illustrate this, we consider an example V1 neuron recorded from an anesthetized cat presented with natural and naturalistic stimuli (Fig. 8A) [62,91]. Because the stimuli are sequences of two-dimensional images, the required spatiotemporal stimulus filters span two dimensions of space and one dimension of time (Fig. 8B), resulting in a very large number of parameters associated with each subunit. Nevertheless, the parameters of the GQM and NIM can be estimated directly utilizing appropriate regularization (see Methods).

The GQM estimated for this neuron is comprised of a pair of excitatory, direction-selective squared filters, as well as a weaker, non-direction-selective linear filter (Fig. 8C). This characterization reflects the neuron's spatial-phase invariance, and is thus consistent with an Energy Model description. While such selectivity suggests that this neuron would be ideally suited for a quadratic model, the NIM (Fig. 8D) significantly outperforms both the GQM and a whitened STC-based model [9,58,62] (Fig. 8E).

The NIM identifies four rectified excitatory inputs that share similar spatial tuning and direction selectivity, but with different spatial phases (Fig. 8D). This description is similar to that provided by the quadratic terms of the GQM, but by identifying the nonlinearities associated with each of these inputs individually, the NIM has additional flexibility that results in improved performance (Fig. 8E). This suggests that a description of complex cells using physiologically plausible inputs (in the form of the NIM) may be a viable alternative to the Energy Model. The improved performance of the NIM is also likely due, at least in part, to the heavy-tailed distribution associated with the naturalistic movie stimuli (as described above, Figs. 3G, H).

Thus, the application of the NIM to V1 neurons further illustrates the generality of the method, and specifically emphasizes its ability to capture substantially more complex stimulus processing, with large numbers of inputs. We note that because cortical neurons are several synapses removed from receptor neurons, a cascade model with a longer chain of upstream LN components might be more appropriate, although existing methods could not be used for parameter estimation with such a model. The ability of the NIM to capture a given neuron's stimulus processing thus relates to the extent to which the upstream neurons themselves can be approximated by LN models. In cases where this assumption is not appropriate, one can apply a fixed nonlinear transformation to the stimulus resembling the response properties of upstream neurons [31,32], thus allowing the problem to be cast into a more general NIM framework.

Conclusions

We have presented a physiologically inspired modeling framework, the NIM, which extends several recently developed probabilistic modeling approaches. Specifically, the NIM assumes a form analogous to an integrate-and-fire neuron, whereby a neuron receives a set of rectified excitatory and inhibitory inputs, each of which is assumed to process the stimulus linearly. The parameters can be estimated robustly and efficiently, and the resulting model structure is able to capture a broader range of neural responses than previously proposed probabilistic methods. Importantly, the physiologically inspired model structure of the NIM also allows for greater interpretability of the model fits, as the components of the model take the form of stimulus-driven excitatory and inhibitory inputs. The NIM thus provides a framework for connecting nonlinear models of sensory processing directly with the underlying physiology that can be applied in a range of sensory areas and experimental conditions.

Methods

Parameter estimation details

As described above, the key parameters in the NIM are the stimulus filters $\mathbf{k}_i$ and the set of coefficients a_{ij} representing the upstream nonlinearities $f_i(.)$. While these parameters cannot generally be optimized simultaneously, a powerful approach is to use block coordinate ascent [66] and alternate between optimizing the filters $\mathbf{k}_i$, and upstream nonlinearities $f_i(.)$, holding the remaining parameters fixed in each iteration. The parameters of the spiking nonlinearity function $F[x;\ \alpha,\beta,\theta] = \alpha\log[1+\exp(\beta(x-\theta))]$ can also be estimated iteratively, or as a final stage after convergence of the $\mathbf{k}_i$ and $f_i(.)$ (which we find is typically sufficient). Note that the parameter β is not generally identifiable in the model (being degenerate with the coefficients a_{ij} of the upstream nonlinearities), but joint estimation of α and β after the other model parameters are fixed allows for a more precise final fit to the spiking nonlinearity function.

Thus, at each stage of the fitting procedure we have the problem of maximizing a (penalized) log-likelihood function with respect to some subset of parameters, while holding a remaining set of

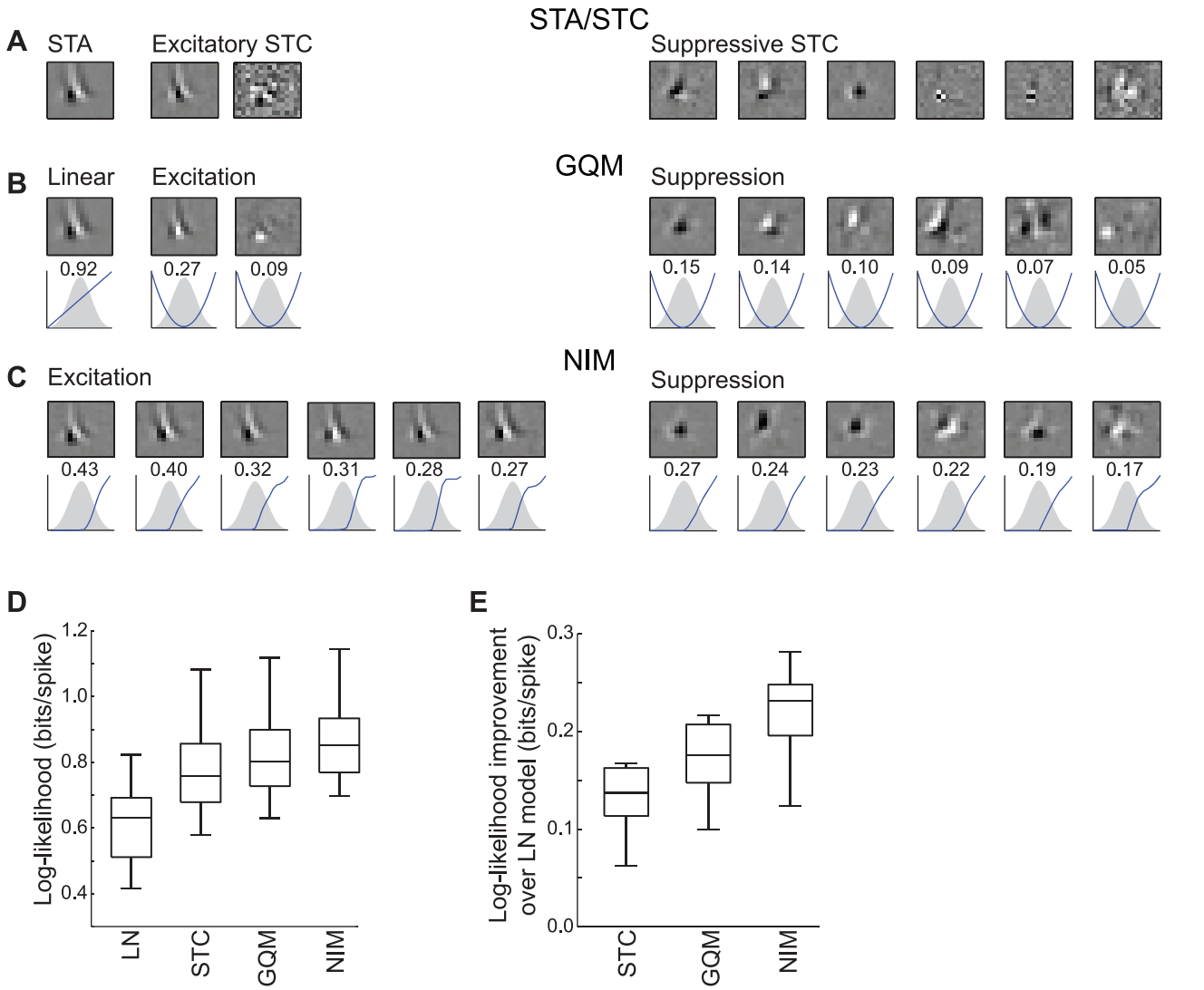

Figure 7. Models of multi-input stimulus processing in a V1 neuron. A) Standard spike-triggered characterization for this neuron reveals a 'complicated simple-cell' response [62], with a clear direction-selective STA (left), two excitatory STC filters (middle), and six suppressive STC filters (right). B) The GQM identifies a set of filters (one linear, two squared excitatory, and six squared suppressive) that are roughly similar to the STA/STC filters, but with smoother and sparser spatiotemporal structure (due to regularization). C) The NIM filters (top) and upstream nonlinearities (bottom) reveal a similar description of the stimulus processing, although with greater consistency among the (six) excitatory and (six) suppressive stimulus filters. D) Comparison of the cross-validated log-likelihood of the LN model (one linear filter), the 'STC model' given by fitting a GLM to the outputs of the STA/STC filters (see Methods), the GQM, and the NIM. Given the neuron's simple-cell-like response (i.e., large weight of the STA), a large fraction of the response can be captured with the linear filter alone (the LN model). Nevertheless, all three multi-filter models provide substantial improvements compared to the LN model. E) In order to compare the performance of the nonlinear (multi-filter) models directly, their improvement relative to the LN model is depicted. This shows that the GQM significantly outperforms the 'STC' model ($p = 0.002$; $n = 10$; Wilcoxon signed rank test), and that the NIM similarly outperforms the GQM ($p = 0.002$).
doi:10.1371/journal.pcbi.1003143.g007

parameters fixed. In all cases, we use a standard line search strategy to locate an optimum of the likelihood function given some initial values for the parameters. Because we are often optimizing very high-dimensional parameter vectors (specifically when optimizing the $\mathbf{k}_i$), we use a quasi-Newton method with a limited-memory BFGS approximation of the inverse Hessian matrix [92] to determine the search direction. This code is implemented in the *Matlab* function "minFunc", provided by Mark Schmidt (available at http://www.di.ens.fr/~mschmidt/). When using sparseness (L1) regularization we utilize the *Matlab* package "L1General", also provided by Mark Schmidt. When optimizing the coefficients a_{ij} of the upstream nonlinearities we additionally enforce a set of linear constraints (described below), and in such cases we utilize *Matlab*'s constrained optimization routine "fmincon". A *Matlab* implementation of the NIM parameter estimation routines described here is available from our website: (http://www.clfs.umd.edu/biology/ntlab/NIM/)

Optimizing the filters k_i

Optimization of the filters can be accomplished efficiently by analytic calculation of the log-likelihood gradient with respect to the $\mathbf{k}_i$, which is given by:

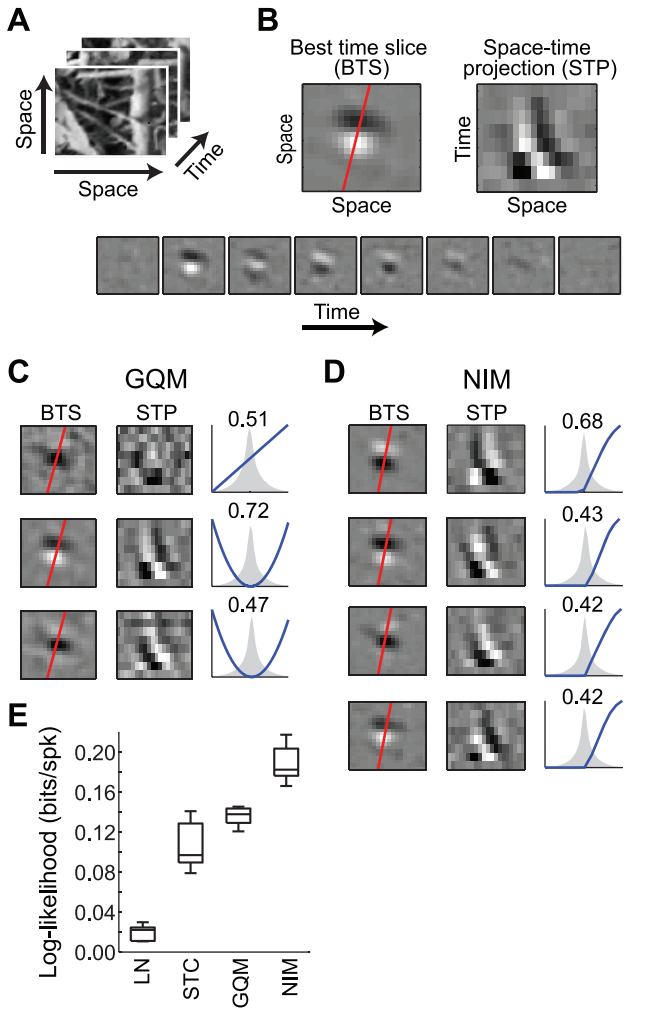

Figure 8. Models of a V1 neuron in the context of natural stimuli. A) The natural movie stimulus used here has two spatial and one temporal dimension. B) The neuron's response is characterized in terms of three-dimensional spatiotemporal filters. An example spatiotemporal filter is comprised of a spatial filter at each time step (at 20 ms resolution). To simplify the depiction of each filter, we take advantage of their stereotyped structure, and plot the spatial distribution at the best time slice (BTS, left), as well as the space-time projection (STP, right) along an axis orthogonal to the preferred orientation (red line; see Methods). C) The GQM for this neuron consists of one linear (top) and two excitatory squared filters (bottom). The BTS and STP for each filter are shown at left, and the distributions of the filtered stimulus, and associated nonlinearities, are shown at right. Note that the two squared filters roughly form a 'quadrature pair' of direction-selective Gabor filters. There is also a linear filter (top), which has less clear spatial structure, and is not direction-selective. D) The NIM consists of four excitatory filters (left) that are qualitatively similar to the quadrature pair of GQM filters. However, by identifying four inputs with inferred upstream nonlinearities (right), the NIM has greater flexibility in describing the neuron's computation. E) Comparison of model performance for the LN and STC-based models, as well as the GQM and NIM, showing that the NIM substantially outperformed other models for this neuron.
doi:10.1371/journal.pcbi.1003143.g008

$$\frac{\partial LL}{\partial k_{i,m}} = \sum_t \left(\frac{R_{obs}(t)}{r(t)} - 1\right) F'[G(t)] w_i f_i'(g_i(t)) s_m(t), \qquad (4)$$

where the 'internal generating function' $G(t) = \sum_i g_i(t) = \sum_i w_i f_i(\mathbf{k}_i \cdot \mathbf{s}(t))$, $F'[.]$ and $f_i'(.)$ are the derivatives of $F[.]$ and $f_i(.)$ with respect to their arguments, and $s_m(t)$ is the m^{th} element of the stimulus at time t. While the likelihood surface is not generally convex with respect to the $\mathbf{k}_i$, the optimization problem is well-behaved in practice. We note that while the derivatives of the $f_i(.)$ are discontinuous (piece-wise constant) when using the tent-basis representation (eq. 6 below), gradient-based optimization methods still provide robust results, in particular because we use regularization to enforce smooth $f_i(.)$ such that the contribution of the discontinuities to the overall log-likelihood gradient is negligible in practice.

To diagnose the presence of undesirable local maxima, and to identify the global optimum of the likelihood function, we use repeated random initializations of our optimization routine (Fig. S1). In some cases, such as the ON-OFF RGC example (Figs. 1, 3), this approach reveals that the choice of initial values for $\mathbf{k}_i$ does not affect the identified local optimum. In other cases, the likelihood surface will contain more than one distinct local maximum, although usually only a small number. For example, when optimizing the filters for the example MLd neuron (Fig. 5) we found two distinct local optima of the likelihood function. For models with large numbers of subunits, the filter optimization remains well-behaved, generally identifying a relatively small number of local optima that correspond to similar models (Fig. S1).

This procedure can be greatly sped up by initially optimizing the filters in a low-dimensional stimulus subspace, rather than in the full stimulus space. Such subspace optimization has been previous used in conjunction with STC analysis to identify the relevant stimulus subspace [15,62,93]; however the GQM provides a means of generalizing the robust subspace identification properties of STC analysis to arbitrary non-Gaussian stimuli, and in cases where regularization is important. With a low-dimensional subspace identified the filters of a NIM can be rapidly optimized, and many filter initializations can be tested.

Fitting the upstream nonlinearities $f_i(.)$

We begin the NIM fitting with its upstream nonlinearities $f_i(.)$ initialized to be threshold-linear functions:

$$f(x) = \left\{ \begin{array}{ll} 0 & \text{if } x \leq 0 \\ x & \text{otherwise} \end{array} \right\}, \qquad (5)$$

and perform initial estimation of the filters. While other rectifying functions can be used, the use of scale-invariant functions such as this one has the advantage that the effect of the upstream nonlinearity is independent of the scale of the filter.

After estimating the $\mathbf{k}_i$, we then estimate the $f_i(.)$ nonparametrically, as a linear combination of a set of piecewise linear basis functions $f_i(g) = \Sigma_j a_{ij} \varphi_j(g)$ [17,45], while holding the $\mathbf{k}_i$ fixed. These basis functions are given by:

$$\varphi_k(x) = \left\{ \begin{array}{ll} \frac{x - x_{k-1}}{x_k - x_{k-1}} & \text{if } x \in [x_{k-1}, x_k] \\ \frac{x_{k+1} - x}{x_{k+1} - x_k} & \text{if } x \in [x_k, x_{k+1}] \\ 0 & \text{otherwise} \end{array} \right\}. \qquad (6)$$

These piecewise linear functions are particularly useful as they provide a set of localized basis functions, requiring only that we choose a set of 'grid points' x_k. These points can be selected by

referencing the distribution of the argument of $f_i(.)$, i.e., $p(g_i)$ where $g_i(t) = \mathbf{k}_i \bullet \mathbf{s}(t)$, either at n-quantiles of $p(g_i)$, or at uniformly spaced points across the support of $p(g_i)$. In order to encourage interpretability of the model subunits as 'neural inputs', we constrain the $f_i(.)$ to be monotonically increasing functions by using a system of linear constraints on the a_{ij} during optimization. Because the model is invariant to shifts in the 'y-offset' of the $f_i(.)$ (which can be absorbed into the spiking nonlinearity function), we add the additional set of constraints that $f_i(0) = 0$ to eliminate this degeneracy. Furthermore, changes in the upstream nonlinearities can influence the effective regularization of the $\mathbf{k}_i$, by altering how each $\mathbf{k}_i$ contributes to the model prediction. As a result, the coefficients a_{ij} are rescaled after each iteration so that the standard deviation of each subunit's output is conserved. This ensures that the upstream nonlinearities do not absorb the scale of the $\mathbf{k}_i$.

Regularization

An important advantage of explicit probabilistic models such as the NIM is the ability to incorporate prior knowledge about the parameters via regularization. Because each of the filters $\mathbf{k}_i$ often contains a large number of parameters, regularization of the filters is of particular importance, as discussed elsewhere in the context of the GLM [9,58,84,94–97], as well as other nonlinear models [26]. Such regularization can impose prior knowledge about the smoothness [9,26,94], sparseness [58,84,94–96], and localization [62,97] of filters in space, frequency and time.

We consider several different forms of regularization in the examples shown, to encourage the detection of smooth filters with sparse coefficients. Specifically, we add a general penalty term of the form:

$$\sum_i \lambda_i^{Ls} \|\mathbf{L}^s \mathbf{k}_i\|_2 + \lambda_i^{Lt} \|\mathbf{L}^t \mathbf{k}_i\|_2 + \lambda_i^s \|\mathbf{k}_i\|_1 \qquad (7)$$

to the equation for the log-likelihood (equation 2), where $\mathbf{L^s}$ and $\mathbf{L^t}$ are the discrete Laplacian operators with respect to spatial (or spectral) and temporal dimensions respectively, and λ_i^{Ls}, λ_i^{Lt} and λ_i^s are hyperparameters which determine the strength of spatial and temporal smoothness, and sparseness regularization, respectively. Other types of regularization, such as those that encourage localized filters [62,97], as well as approximate Bayesian techniques for inferring hyperparameters [26,58,94,96] could be incorporated as well, although we do not do so here.

Because we also expect the upstream nonlinearities $f_i(.)$ to be smooth functions, we incorporate penalty terms when estimating the parameters of the $f_i(.)$. Because we represent the $f_i(.)$ as linear combinations of localized tent basis functions: $f_i(.) = a_{ij}\varphi_j(.)$, we can encourage smooth $f_i(.)$ by applying a penalty of the form: $\lambda_i^L \|\mathbf{L}a_{ij}\|_2$ to the set of coefficients a_{ij} corresponding to a given $f_i(.)$, where $\mathbf{L}$ is again the one-dimensional discrete Laplacian operator.

In general, the hyperparameters can be inferred from the data using Bayesian techniques [94], or estimated using a (separate) cross-validation data set. Both methods can be time-consuming, however, and in practice we find that similar results can be achieved by 'manually' tuning the hyperparameters to produce filters $\mathbf{k}_i$ and upstream nonlinearities $f_i(.)$ with the expected degree of smoothness/sparseness. To demonstrate that our results were not overly sensitive to the selection of hyperparameters, we compare the NIM and GQM fit to the example V1 neuron from Fig. 8 using a range of regularization strengths (Fig. S6).

Evaluating model performance

To evaluate model performance, we use k-fold cross-validation, in general taking the log-likelihood as a performance metric. The likelihood has the advantage over related measures such as R^2 in that it does not require repeated stimulus presentations to estimate, and thus can be applied to most data sets. It can also capture goodness-of-fit when spike history terms are incorporated [35]. Subtracting the log-likelihood of the null model (that predicts a constant firing rate, independent of the stimulus) provides a measure of the information carried by the spike train about the stimulus, in units of bits per spike [45,98]. This measure is also directly related to the more traditional measure of deviance, which compares the log-likelihood of the estimated model to that of the 'saturated' model. In order to provide a more direct connection to standard measures of model performance based on repeated presentations of a stimulus, we also computed the 'predictive power' of the models for the simulated ON-OFF RGC (Fig. 3F), which is defined as the fraction of 'explainable' variance accounted for by the model [99]. Due to the lack of sufficient repeat trial data for our recorded data examples we could not compute this measure in those cases, however qualitatively similar results would be expected.

Model selection

While selection of the optimal number of excitatory and suppressive subunits can be performed using standard model selection techniques, such as nested cross-validation, this choice can also often be guided by the specific application. Importantly, we find that the subunits identified by the NIM, as well as its performance, are generally robust towards precise specification of the number of excitatory and suppressive subunits, with 'nearby' models typically providing a very similar characterization of the neurons' stimulus processing (Fig. S5). This robustness is further aided by the incorporation of sparseness regularization on the filters, where the filters of extraneous subunits tend to be driven to zero. The procedure of testing a series of NIMs with different subunit compositions can again be substantially facilitated by optimizing the filters in a low-dimensional stimulus subspace, such as identified by STC or GQM analysis (Fig. S5).

RGC simulation details

In order to simulate the response of an ON-OFF RGC, we generated a Gaussian white noise process sampled at 15 Hz (such as a luminance-modulated spot stimulus), which was then filtered using separate ON- and OFF-like filters (Fig. 1A). These filter outputs were then rectified using functions of the form $f(x) = \log(1+\exp(b_1 x))$, summed together and the resulting signal was passed through a spiking nonlinearity of the form $F[x] = a\log(1+\exp(b_2(x-c)))$. This conditional intensity function was then used to generate a set of spike times. To generate heavy-tailed stimulus distributions (Figs. 3G, H), we sampled white noise from a Student's t-distribution with a range of values for the degrees of freedom to control the tail thickness.

The data were simulated at a temporal resolution of 8.3 ms, and model filters were represented at a lower resolution of 33 ms, with a length of 1 s. For the GQM and NIM we incorporated smoothness regularization on the filters, and for the NIM we also incorporated smoothness regularization on the upstream nonlinearity coefficients a_{ij}.

To identify the STA/STC subspace depicted in Figs. 1 and 3, we performed STC analysis after projecting out the STA. For comparison with the NIM, we also created a simple model based on the STA and STC filters, using a GLM-based optimization of linear coefficients on the outputs of the STA filter and the squared

outputs of the STC filters, similar to previous work [58]. Note that in order to maximize performance when estimating STC-based models, we did not project out the STA before computing the STC filters.

LGN neuron model fit details

Data for the LGN example were recorded extracellularly from an anaesthetized and paralyzed cat by the Alonso Lab [45,75,76]. The stimulus consisted of 800 seconds of a 32×32 pixel natural movie, refreshed at 60 Hz, which was recorded from a camera mounted on top of a cat's head [100]. A 17×17 pixel patch of the movie was cropped around the receptive field, detected via STA at the optimal latency, and the movie was up-sampled by a factor of six, to produce a temporal resolution of 2.8 ms. Ten-fold cross-validation was used for evaluating model performance.

Each filter was represented by space-time separable center and surround components, and thus consisted of two sets of spatial and temporal filters [101]. Temporal filters were represented with 30 equally spaced tent basis functions, with grid points ranging from 0 to −240 ms. For the LN model, the spatial filters were initialized as Gaussian functions with the same center as the STA and different widths (1 pixel for the center and 6 pixels for the surround). For the GQM and NIM, both excitatory and suppressive filters were initialized to be the same as the optimal linear filters. In the filter optimization stage, the spatial and temporal filters were optimized alternately until convergence of the log-likelihood. Both the GQM and the NIM were fit using smoothness regularization for the spatial and temporal kernels, and sparseness regularization for the spatial kernels. For the NIM, we also used smoothness regularization on the a_{ij} when estimating the upstream nonlinearities.

Songbird auditory midbrain model fit details

Data for the songbird auditory midbrain example were provided by the Theunissen lab through the CRCNS database [83], and details of experimental methods can be found in [81,82]. The example neuron was recorded extracellularly from the zebra finch mesencephalicus lateralis dorsalis (MLd). Stimuli consisted of 20 different conspecific bird songs, each lasting 2–4 sec, and each presented 10 times. These 20 songs were then divided into 5 equal groups for five-fold cross-validation. The raw sound waveforms were preprocessed by computing the spectrogram using a short-time Fourier transform to produce a stimulus matrix $X(t,f)$, representing the power of the audio signal at frequency f and time t. We used a time resolution of 2 ms, and 20 uniformly spaced frequency bins, ranging from 250 Hz to 8 kHz. For estimating spectrotemporal filters, we used 20 time lags. Thus, each filter was represented by 400 parameters. Filter estimates were regularized using sparseness and smoothness penalties, where the smoothness penalty utilized the spectrotemporal Laplacian (with equal weighting in the frequency and time dimensions).

V1 simulation details

The simulated V1 neurons shown in Fig. 6 were constructed as LNLN models (Fig. 2C). The stimulus filters were spatial Gabor functions that were amplitude- and phase-modulated in time (i.e., direction-selective). Stimulus filters were identical up to a spatial translation, and were weighted by a spatial Gaussian envelope. The filter outputs were then passed through a set of static nonlinear functions (either x^2, or $\log(1+\exp(b_1 x))$), before being summed together, and passed through the spiking nonlinearity (again, of the form $a\log(1+\exp(b_2(x-c)))$ to generate a conditional intensity function. Spike times were simulated in response to binary random bar stimuli [57], using a time resolution of 10 ms, and 24 bar positions.

Both the GQM and NIM were fit using a sparseness penalty on the filters. For the NIM, we also used smoothness regularization on the a_{ij} when estimating the upstream nonlinearities. To measure how well the estimated model filters matched the true filters, we represented the model filters as linear combinations of the true filters.

V1 neuron modeling details

The V1 neuron shown in Fig. 7 was recorded from an anesthetized macaque [57]. The stimuli (refreshed at 100 Hz) consisted of random arrays of black and white bars covering the neuron's classical receptive field, and oriented along its preferred orientation. Full experimental details can be found in [57]. Spatiotemporal filters were represented by 16 'pixels' and 14 time lags. For model evaluation we used ten-fold cross-validation. Model fitting was analogous to that described for the V1 simulations (Fig. 6). STC-based models were constructed as described above for the simulated ON-OFF RGC.

The V1 neuron shown in Fig. 8 was recorded from an anesthetized cat [91]. The stimuli consisted of natural and naturalistic movies at various contrasts, including noise processes with pink spatial and white temporal statistics, pink temporal and white spatial statistics, pink temporal and pink spatial statistics, and natural movies recorded with a 'cat cam' [100]. The mean luminance across all stimuli (15 different stimuli, each lasting 2 minutes) was the same. The raw movies were 64×64 pixels and were sampled at 50 Hz. These raw movies were spatially down-sampled and cropped to produce 20×20 pixel patches that were individually mean-subtracted. Model performance was evaluated using five-fold cross-validation, with cross-validation sets constructed by taking 20% of the data from each stimulus type.

For all analysis we used 8 time lags to construct spatiotemporal filters (each described by $8\times20\times20 = 3200$ parameters). For STA/STC analysis we first whitened the stimulus by rotating into the principal component axes and normalizing each dimension to have unit standard deviation [9]. Because the stimulus covariance matrix for natural stimuli has many eigenvalues close to zero, we avoided amplifying noise associated with these low-variance dimensions by using a pseudoinverse of the covariance matrix, effectively discarding the n lowest variance dimensions of the stimulus [9,95]. In addition to removing biases due to pairwise correlations in the stimulus, this method effectively imparts a prior favoring spatiotemporally smooth filters, since the lowest variance dimensions of natural stimuli have high spatial and temporal frequencies. We retained 500/3200 of the stimulus dimensions for STA/STC analysis.

To estimate filters of the LN model, GQM and NIM, we used sparseness regularization, as well as penalties on the (two-dimensional) spatial Laplacian at each time lag. To display the three-dimensional spatiotemporal filters we plot the time slice of each filter containing the most variance across pixels ('best time slice'), as well as the projection of the filter onto a spatial axis orthogonal to the neuron's preferred orientation ('space-time projection') [62]. The preferred orientation was determined by fitting a two-dimensional Gabor function to the best time slice for each filter, and taking the (circular) average of the individual Gabor orientations across all filters.

Supporting Information

Figure S1 Robustness of filter estimation. A) For the simulated ON-OFF RGC in Figs. 1 and 3, the likelihood function

with respect to the NIM stimulus filters shows only a single global optimum (up to an interchange of the filters) over a broad range of parameter space. To illustrate this, 100 iterations of the optimization were performed with random initializations of the filters, and in all cases the correct filters were identified. The initial filters are projected onto the true ON and OFF filters (inset), and are plotted along with the resulting optimized filter projections. Each iteration of the optimization is thus represented by a pair of optimized filters (large blue and red circles), along with a pair of initial filters (small blue and red circles, color coded based on the resulting filter estimates). **B**) For the example MLd neuron in Fig. 5, we found two distinct local maxima when optimizing the NIM stimulus filters, corresponding to the two clusters of the maximized log-likelihood across many repetitions of optimizing the filters with random initial conditions. The global optimum (right) corresponds to the set of filters shown in Fig. 5, while a locally optimum solution (left) corresponds to the excitatory filter matching the STA. **C**) For the simulated V1 neuron shown in Fig. 6A, optimization of the NIM is again well-behaved. In this case there are potentially several spurious local maxima, illustrated by the distribution of maximized log-likelihood values. However, these local maxima correspond to models that are very similar to the identified global maximum, as shown by the similar log-likelihood values, as well as the similarity of the identified filters (example models shown at left and right).
(EPS)

Figure S2 NIM parameter optimization scales approximately linearly. A) The time required to estimate the filters (black) and upstream nonlinearities (red) scales linearly as a function of data duration for the ON-OFF RGC simulation (with two subunits) shown in Figs. 1 and 3. The error bars show $+/-1$ standard deviation around the mean across multiple repetitions of the parameter estimation (with random initialization). Estimation was performed on a machine running Mac OS X 10.6 with two 2.26 GHz quad-core Intel Xeon processors and 16 GB of RAM. **B**) To measure parameter estimation time as a function of the number of stimulus dimensions, we simulated a V1 neuron (similar to that shown in Fig. 6A) receiving two rectified inputs (data duration of 10^5 time samples). We then varied the number of time lags used to represent the stimulus and measured the time required for parameter estimation. Estimation of the stimulus filters scales roughly linearly with the number of stimulus dimensions, while estimation of the upstream nonlinearities is largely independent of the number of stimulus dimensions. **C**) Parameter estimation time for the filters and upstream nonlinearities also scales approximately linearly as a function of the number of subunits. Here we again used a simulated V1 neuron similar to that shown in Fig. 6A, although with 10 rectified inputs (200 stimulus dimensions and data duration of 10^5 time samples). Note that the additional step of estimating the upstream nonlinearities adds relatively little to the overall parameter estimation time, especially for more complex models.
(EPS)

Figure S3 Comparison of the NIM and GQM for the example LGN neuron. The linear model (**A**), NIM (**B**), and GQM (**C**) fit to the example LGN neuron from Fig. 4 are shown for comparison. Here (A) and (B) are reproduced from Figs. 4A and B respectively. Note that the spatial and temporal profiles of the linear and squared (suppressive) GQM filters are largely similar to the (rectified) excitatory and suppressive filters identified by the NIM. Despite the similarity of the identified filters, however, the NIM and GQM imply a different picture of the neuron's stimulus processing, as illustrated in Fig. S4.
(EPS)

Figure S4 Different predictions of the GQM and NIM with excitation and delayed suppression. The GLM (**A**), NIM (**B**), and GQM (**C**) fit to the example MLd neuron in Fig. 5 (A and B here are reproduced from Figs. 5A and B). The NIM and GQM identify similar excitatory and suppressive filters, but the GQM assumes linear and squared upstream nonlinearities for these inputs respectively, while the NIM infers the rectified form of these functions. Despite the similarities in the identified filters, the different upstream nonlinearities in these models imply distinct interactions between the excitatory and suppressive inputs. To illustrate this, we consider how these different models process two stimuli in (D) and (E), which highlight these differences. **D**) First, we consider a negative impulse (left) presented at the preferred frequency (horizontal black lines in A–C). The outputs of the excitatory (blue) and suppressive (red) subunits are shown for the linear model (top), GQM (middle), and NIM (bottom). The combined outputs of these subunits are then transformed by the spiking nonlinearity into the corresponding predicted firing rates at right. In this case, only the linear model responds to the stimulus, since the GQM is strongly suppressed, and the NIM is largely unaffected due to the rectification of the negatively driven inputs. **E**) Similar to (D), we consider a biphasic stimulus (left), also presented at the neuron's preferred frequency. This stimulus drives different responses in all three models. The response predicted by the GQM is by far the weakest because the (squared) suppression driven by the initial negative phase of the stimulus coincides with the excitation driven by the positive phase of the stimulus, causing them to partially cancel each other out. For the NIM, the negative phase of the stimulus does not drive the suppression, due to rectification, and the excitation is able to elicit a much larger response. The response predicted by the linear model is even larger since this is essentially the optimal stimulus for driving the linear filter. This suggests targeted stimuli that might be able to distinguish the computations being performed by MLd neurons.
(EPS)

Figure S5 Selecting the number of model subunits. A) To illustrate the robustness of NIM parameter estimation to specification of the precise number of subunits, we first consider the simulated V1 neuron from Fig. 6A, which was constructed from six rectified excitatory subunits. Fitting a sequence of NIMs (blue) and GQMs (red) with increasing numbers of (excitatory) subunits reveals that the log-likelihood (evaluated on a simulated cross-validation data set) initially improves dramatically, but becomes nearly saturated for models with four or more subunits. Here we plot log-likelihood relative to that of the best model, and error bars show one std. dev. about the mean. While it is possible in this case to identify the true number of underlying subunits (six) from the cross-validated model performance of the NIM, the model performance is relatively insensitive to specification of the precise number of subunits. **B**) Stimulus filters from example NIM fits from (A), with four, six, and eight filters. Note that the identified filters are nearly identical across these different models, and when more than the true number (six) of subunits are included in the model, sparseness regularization on the filters tends to drive the extra filters to zero, yielding effectively identical models. **C**) To illustrate the procedure of selecting the number of model subunits with real data, we consider fitting a series of models to the example MLd

neuron from Fig. 5. In this case there are both excitatory and suppressive stimulus dimensions, so we independently vary the number of each. Average (+/−1 std. error) cross-validated model performance is depicted for each subunit composition for models with up to three subunits (the color indicates the number of suppressive subunits). While we do not have sufficient data to identify statistically significant differences, a two-filter model with one excitatory and one suppressive filter appears to achieve optimal performance. **D**) Three example NIM fits for one, two, and three-filter models corresponding to the Roman numerals in (C). (i) Model with one excitatory subunit. (ii) Model with one excitatory and one suppressive subunit. (iii) Model with two excitatory and one suppressive subunits. Note that the excitatory and suppressive filters from the two-filter model are also present in the three-filter model, and that the addition of a second excitatory subunit (resembling the linear filter) provides little, if any, additional predictive power. **E**) Similar to (C–D), we consider fitting models with different numbers of excitatory and suppressive subunits to the example macaque V1 neuron from Fig. 7. In this case, the neuron is selective to a large number of stimulus dimensions (Fig. 7A), and thus there are a large number of possible excitatory/suppressive subunit compositions to consider. To greatly speed this process (and illustrate a procedure for rapid model characterization), we fit the NIM filters in a reduced stimulus subspace (see Methods) that is identified by a GQM with four excitatory and six suppressive dimensions. The number of subunits in the GQM was selected in order to ensure that all filters with discernible structure were included. The figure then shows the average (+/−1 std. error) cross-validated log-likelihood (relative to a model with two excitatory and one suppressive filters) for NIMs with varying numbers of excitatory and suppressive subunits. Note that the model performance increases initially, but tends to saturate for models with more than about four excitatory and four suppressive subunits. For comparison, the cross-validated log-likelihood of the GQM (green line) – which established the stimulus subspace – is below most of the NIM solutions. While fitting the stimulus filters in the full stimulus space provides slightly different (though qualitatively very similar) results, limiting the NIM to the subspace provides a tractable way to fully explore the nonlinear structure of computation, and can then serve as an initial guess for a more computationally-intensive search in the full stimulus space. **F–G**) Two example NIM fits from those depicted in (E). A NIM with four excitatory and four suppressive subunits (F) is compared to a NIM with six excitatory and six suppressive subunits (G), the latter providing only a slight improvement relative to the former. Both models provide a qualitatively similar depiction of the neuron's stimulus processing, identifying largely similar sets of excitatory and suppressive inputs.
(EPS)

Figure S6 Selection of regularization parameters. To illustrate how the performance of our models depends on selection of the regularization hyperparameters, we fit a series of models to the example V1 neuron from Fig. 8. For this example neuron regularization of the stimulus filters is particularly important, given the large number (3200) of parameters associated with each filter. As described in the Methods section, we use both smoothness (L2 penalty on the spatial Laplacian) and sparseness regularization on the filters, each of which is governed by a hyperparameter. While in principle we could independently optimize these regularization parameters for each filter, we consider here only the case where all filters are subject to the same regularization penalties. Further, we consider optimizing the smoothness and sparseness penalties independently, which will not in general identify the optimal set of hyperparameters. **A**) We first set the sparseness regularization penalty to zero, and systematically vary the strength of the smoothness penalty. The cross-validated log-likelihood is plotted for the NIM (blue trace) and GQM (red trace), showing that the NIM outperforms the GQM over a range of smoothness regularization strengths. **B–D**) Representative filters are shown from model fits at several regularization strengths, as indicated by the black circles in (A). The filters are depicted as the 'best-time slice' (BTS) and the 'space-time projection' (STP), as in Fig. 8. **E**) Similar to (A), we next consider varying the strength of sparseness regularization given fixed values for the smoothness regularization (set to the value indicated by the vertical dashed line in A). Note that the performance of the NIM again remains significantly better than the GQM across a range of regularization strengths. **F–H**) Representative filters at several sparseness regularization strengths, as indicated in (E). Note that (F) is identical to (C), reproduced for ease of comparison.
(EPS)

Acknowledgments

The authors thank J.-M. Alonso for contributing the LGN data, the Theunissen lab and CRCNS database for providing the songbird auditory midbrain data, N. Rust and T. Movshon for contributing the macaque V1 data, and T. Blanche for contributing the cat V1 data. We also thank N. Lesica, and S. David for comments on the manuscript.

Author Contributions

Analyzed the data: JMM YC DAB. Wrote the paper: JMM DAB. Developed the model: JMM YC DAB.

References

1. Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. Nat Neurosci 2: 1019–1025.
2. Brincat SL, Connor CE (2004) Underlying principles of visual shape selectivity in posterior inferotemporal cortex. Nat Neurosci 7: 880–886.
3. Mineault PJ, Khawaja FA, Butts DA, Pack CC (2012) Hierarchical processing of complex motion along the primate dorsal visual pathway. Proc Natl Acad Sci U S A 109: E972–E980.
4. DiCarlo JJ, Cox DD (2007) Untangling invariant object recognition. Trends Cogn Sci 11: 333–341.
5. Rust NC, Stocker AA (2010) Ambiguity and invariance: two fundamental challenges for visual processing. Curr Opin Neurobiol 20: 382–388.
6. Brillinger DR (1988) Maximum likelihood analysis of spike trains of interacting nerve cells. Biol Cybern 59: 189–200.
7. Paninski L (2004) Maximum likelihood estimation of cascade point-process neural encoding models. Network 15: 243–262.
8. Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown EN (2005) A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. J Neurophysiol 93: 1074–1089.
9. Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, et al. (2001) Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. Network 12: 289–316.
10. Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, et al. (2005) Do we know what the early visual system does? J Neurosci 25: 10577–10597.
11. Shapley R (2009) Linear and nonlinear systems analysis of the visual system: why does it seem so linear? A review dedicated to the memory of Henk Spekreijse. Vision Res 49: 907–921.
12. de Ruyter van Steveninck R, Bialek W (1988) Coding and information transfer in short spike sequences. P Roy Soc Lond B Bio 234: 379–414.
13. Schwartz O, Pillow JW, Rust NC, Simoncelli EP (2006) Spike-triggered neural characterization. J Vis 6: 484–507.
14. Sharpee T, Rust NC, Bialek W (2004) Analyzing neural responses to natural signals: maximally informative dimensions. Neural Comput 16: 223–250.
15. Pillow JW, Simoncelli EP (2006) Dimensionality reduction in neural models: an information-theoretic generalization of spike-triggered average and covariance analysis. J Vis 6: 414–428.

16. Chichilnisky EJ (2001) A simple white noise analysis of neuronal light responses. Network 12: 199–213.
17. Ahrens MB, Paninski L, Sahani M (2008) Inferring input nonlinearities in neural encoding models. Network 19: 35–67.
18. Park M, Horwitz G, Pillow JW (2011) Active learning of neural response functions with Gaussian processes. Adv in Neural Inf Process Syst (NIPS) 24: 2043–2051.
19. Truccolo W, Donoghue JP (2007) Nonparametric modeling of neural point processes via stochastic gradient boosting regression. Neural Comput 19: 672–705.
20. Marmarelis PZ, Marmarelis VZ (1978) Analysis of physiological systems: The white-noise approach. New York: Plenum Press.
21. Schetzen M (1989) The Volterra and Wiener theories of nonlinear systems. Malabar, Florida: Krieger.
22. Eggermont JJ (1993) Wiener and Volterra analyses applied to the auditory system. Hear Res 66: 177–201.
23. DeWeese MR (1995) Optimization principles for the neural code: Princeton.
24. Marmarelis VZ (2004) Nonlinear Dynamic Modeling of Physiological Systems. Hoboken, NJ: Wiley Interscience.
25. Ahrens MB, Linden JF, Sahani M (2008) Nonlinearities and contextual influences in auditory cortical responses modeled with multilinear spectrotemporal methods. J Neurosci 28: 1929–1942.
26. Park IM, Pillow JW (2011) Bayesian spike-triggered covariance analysis. Adv Neural Inf Process Syst (NIPS) 24: 1692–1700.
27. Fitzgerald J, D., Rowekamp R, J., Sincich L, C., Sharpee T, O. (2011) Second order dimensionality reduction using minimum and maximum mutual information models. PLoS Comput Biol 7: e1002249.
28. Rajan K, Bialek W (2012) Maximally informative "stimulus energies" in the analysis of neural responses to natural signals. arXiv:12010321 [q-bioNC].
29. Lau B, Stanley GB, Dan Y (2002) Computational subunits of visual cortical neurons revealed by artificial neural networks. Proc Natl Acad Sci U S A 99: 8974–8979.
30. Prenger R, Wu MC, David SV, Gallant JL (2004) Nonlinear V1 responses to natural scenes revealed by neural network analysis. Neural Netw 17: 663–679.
31. Nishimoto S, Gallant JL, Nishimoto S, Gallant JL (2011) A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. J Neurosci 31: 14551–14564.
32. Mineault PJ, Khawaja FA, Butts DA, Pack CC (2012) Hierarchical processing of complex motion along the primate dorsal visual pathway. Proc Natl Acad Sci U S A.
33. Berry MJ, Meister M (1998) Refractoriness and neural precision. J Neurosci 18: 2200–2211.
34. Keat J, Reinagel P, Reid RC, Meister M (2001) Predicting every spike: a model for the responses of visual neurons. Neuron 30: 803–817.
35. Pillow JW, Paninski L, Uzzell VJ, Simoncelli EP, Chichilnisky EJ (2005) Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. J Neurosci 25: 11003–11013.
36. Shapley RM, Victor JD (1978) The effect of contrast on the transfer properties of cat retinal ganglion cells. J Physiol 285: 275–298.
37. Meister M, Berry MJ (1999) The neural code of the retina. Neuron 22: 435–450.
38. Mante V, Bonin V, Carandini M (2008) Functional mechanisms shaping lateral geniculate responses to artificial and natural stimuli. Neuron 58: 625–638.
39. Wehr M, Zador AM (2003) Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. Nature 426: 442–446.
40. Murphy BK, Miller KD (2009) Balanced amplification: a new mechanism of selective amplification of neural activity patterns. Neuron 61: 635–648.
41. Vogels TP, Abbott LF (2009) Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. Nat Neurosci 12: 483–491.
42. Sun YJ, Wu GK, Liu BH, Li P, Zhou M, et al. (2010) Fine-tuning of pre-balanced excitation and inhibition during auditory cortical development. Nature 465: 927–931.
43. Dorrn AL, Yuan K, Barker AJ, Schreiner CE, Froemke RC (2010) Developmental sensory experience balances cortical excitation and inhibition. Nature 465: 932–936.
44. Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, et al. (2008) Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature 454: 995–999.
45. Butts DA, Weng C, Jin J, Alonso JM, Paninski L (2011) Temporal precision in the visual pathway through the interplay of excitation and stimulus-driven suppression. J Neurosci 31: 11313–11327.
46. Carcieri SM, Jacobs AL, Nirenberg S (2003) Classification of retinal ganglion cells: a statistical approach. J Neurophysiol 90: 1704–1713.
47. Farrow K, Masland RH (2011) Physiological clustering of visual channels in the mouse retina. J Neurophysiol 105: 1516–1530.
48. Zhang Y, Kim IJ, Sanes JR, Meister M (2012) The most numerous ganglion cell type of the mouse retina is a selective feature detector. Proc Natl Acad Sci U S A 109: E2391–2398.
49. Cantrell DR, Cang J, Troy JB, Liu X (2010) Non-centered spike-triggered covariance analysis reveals neurotrophin-3 as a developmental regulator of receptive field properties of ON-OFF retinal ganglion cells. PLoS Comput Biol 6: e1000967.
50. Demb JB, Zaghloul K, Haarsma L, Sterling P (2001) Bipolar cells contribute to nonlinear spatial summation in the brisk-transient (Y) ganglion cell in mammalian retina. J Neurosci 21: 7447–7454.
51. Kim KJ, Rieke F (2001) Temporal contrast adaptation in the input and output signals of salamander retinal ganglion cells. J Neurosci 21: 287–299.
52. Geffen MN, de Vries SEJ, Meister M (2007) Retinal ganglion cells can rapidly change polarity from Off to On. PLoS Biol 5: e65.
53. Gollisch T, Meister M (2008) Modeling convergent ON and OFF pathways in the early visual system. Biol Cybern 99: 263–278.
54. Schwartz GW, Okawa H, Dunn FA, Morgan JL, Kerschensteiner D, et al. (2012) The spatial structure of a nonlinear receptive field. Nat Neurosci 15: 1572–1580.
55. 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.
56. Rad KR, Paninski L (2010) Efficient, adaptive estimation of two-dimensional firing rate surfaces via Gaussian process methods. Network 21: 142–168.
57. Rust NC, Schwartz O, Movshon JA, Simoncelli EP (2005) Spatiotemporal elements of macaque v1 receptive fields. Neuron 46: 945–956.
58. Gerwinn S, Macke JH, Seeger M, Bethge M (2008) Bayesian inference for spiking neuron models with a sparsity prior. Adv in Neural Inf Proces Syst (NIPS) 20: 529–536.
59. Narendra K, Gallman P (1966) An iterative method for the identification of nonlinear systems using a Hammerstein model. IEEE T Automat Contr 11: 546–550.
60. Hunter IW, Korenberg MJ (1986) The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. Biol Cybern 55: 135–144.
61. Schinkel-Bielefeld N, David SV, Shamma SA, Butts DA (2012) Inferring the role of inhibition in auditory processing of complex natural stimuli. J Neurophysiol 107: 3296–3307.
62. Lochmann T, Blanche TJ, Butts DA (2013) Construction of direction selectivity through local energy computations in primary visual cortex. PLoS ONE 8: e58666.
63. Friedman JH, Stuetzle W (1981) Projection pursuit regression. J Am Stat Assoc 76: 817–823.
64. Lingjærde OC, Liestøl K (1998) Generalized projection pursuit regression. SIAM J Sci Comput 20: 844–857.
65. Vintch B, Zaharia A, Movshon JA, Simoncelli EP (2012) Efficient and direct estimation of a neural subunit model for sensory coding. Adv in Neural Information Processing Systems (NIPS) 25: 3113–3121.
66. Bertsekas DP (1999) Nonlinear Programming. Belmont, MA: Athena Scientific.
67. Paninski L (2003) Convergence properties of three spike-triggered analysis techniques. Network 14: 437–464.
68. J KM, M SH, K N (1989) Dissection of the neuron network in the catfish inner retina. III. Interpretation of spike kernels. J Neurophysiol 61: 1110–1120.
69. Field DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. J Opt Soc Am 4: 2379–2394.
70. Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381: 607–609.
71. Lewicki MS (2002) Efficient coding of natural sounds. Nat Neurosci 5: 356–363.
72. Gabernet L, Jadhav SP, Feldman DE, Carandini M, Scanziani M (2005) Somatosensory integration controlled by dynamic thalamocortical feed-forward inhibition. Neuron 48: 315–327.
73. Okun M, Lampl I (2008) Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. Nat Neurosci 11: 535–537.
74. Wu GK, Arbuckle R, Liu BH, Tao HW, Zhang LI (2008) Lateral sharpening of cortical frequency tuning by approximately balanced inhibition. Neuron 58: 132–143.
75. Lesica NA, Weng C, Jin J, Yeh CI, Alonso JM, et al. (2006) Dynamic encoding of natural luminance sequences by LGN bursts. PLoS Biol 4: e209.
76. Butts DA, Desbordes G, Weng C, Jin J, Alonso JM, et al. (2010) The episodic nature of spike trains in the early visual pathway. J Neurophysiol 104: 3371–3387.
77. Butts DA, Weng C, Jin J, Yeh CI, Lesica NA, et al. (2007) Temporal precision in the neural code and the timescales of natural vision. Nature 449: 92–95.
78. Sincich LC, Horton JC, Sharpee TO (2009) Preserving information in neural transmission. J Neurosci 29: 6207–6216.
79. Wang X, Vaingankar V, Sanchez CS, Sommer FT, Hirsch JA (2011) Thalamic interneurons and relay cells use complementary synaptic mechanisms for visual processing. Nat Neurosci 14: 224–231.
80. Kaplan E, Marcus S, So Y (1979) Effects of dark adaptation on spatial and temporal properties of receptive fields in cat lateral geniculate nucleus. J Physiol 294: 561–580.
81. Hsu A, Woolley SM, Fremouw TE, Theunissen FE (2004) Modulation power and phase spectrum of natural sounds enhance neural encoding performed by single auditory neurons. J Neurosci 24: 9201–9211.
82. Gill P, Zhang J, Wooley SMN, Fremouw T, Theunissen F, E. (2006) Sound representation methods for spectro-temporal receptive field estimation. J Comput Neurosci 21: 5–20.
83. Teeters JL, Harris KD, Millman KJ, Olshausen BA, Sommer FT (2008) Data sharing for computational neuroscience. Neuroinformatics 6: 47–55.

84. Calabrese A, Schumacher JW, Schneider DM, Paninski L, Woolley SMN (2011) A Generalized Linear Model for Estimating Spectrotemporal Receptive Fields from Responses to Natural Sounds. PLoS ONE 6: e16104.
85. Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. J Opt Soc Am A 2: 284–299.
86. Emerson RC, Bergen JR, Adelson EH (1992) Directionally selective complex cells and the computation of motion energy in cat visual cortex. Vision Res 32: 203–218.
87. Mante V, Carandini M (2005) Mapping of stimulus energy in primary visual cortex. J Neurophysiol 94: 788–798.
88. Touryan J, Lau B, Dan Y (2002) Isolation of relevant visual features from random stimuli for cortical complex cells. J Neurosci 22: 10811–10818.
89. Tanabe S, Haefner RM, Cumming BG (2011) Suppressive mechanisms in monkey V1 help to solve the stereo correspondence problem. J Neurosci 31: 8295–8305.
90. Diaconis P, Shahshahani M (1984) On nonlinear functions of linear combinations. SIAM J Sci and Stat Comput 5: 175–191.
91. Blanche TJ, Spacek MA, Hetke JF, Swindale NV (2005) Polytrodes: high-density silicon electrode arrays for large-scale multiunit recording. J Neurophysiol 93: 2987–3000.
92. Nocedal J (1980) Updating quasi-Newton matrices with limited storage. Math Comput 35: 773–782.
93. Saleem AB, Krapp HG, Shultz SR (2008) Receptive field characterization by spike-triggered independent component analysis. J Vis 8: 2.1–2.16.
94. Sahani M, Linden J (2003) Evidence optimization techniques for estimating stimulus-response functions. In: Becker S, Thrun S, Obermayer K, editors. Adv in Neural Information Processing Systems. Cambridge, MA: The MIT Press. pp. 317–324.
95. David SV, Mesgarani N, Shamma SA (2007) Estimating sparse spectro-temporal receptive fields with natural stimuli. Network 18: 191–212.
96. Gerwinn S, Macke JH, Bethge M (2010) Bayesian inference for generalized linear models for spiking neurons. Front Comput Neurosci 4: 12.
97. Park M, Pillow JW (2011) Receptive field inference with localized priors. PLoS Comput Biol 7: e1002219.
98. Kouh M, Sharpee TO (2009) Estimating linear-nonlinear models using Renyi divergences. Network 20: 49–68.
99. Sahani M, Linden J (2003) How linear are auditory cortical responses? In: Becker S, Thrun S, Obermayer K, editors. Advances in neural information processing systems. Cambridge: MIT.
100. Kayser C, Salazar RF, Konig P (2003) Responses to natural scenes in cat V1. J Neurophysiol 90: 1910–1920.
101. Cai D, Deangelis GC, Freeman RD (1997) Spatiotemporal receptive field organization in the lateral geniculate nucleus of cats and kittens. J Neurophysiol 78: 1045–1061.