ANNUAL
REVIEWS

Annual Review of Vision Science

Data-Driven Approaches to Understanding Visual Neuron Activity

Daniel A. Butts

Department of Biology and Program in Neuroscience and Cognitive Science, University of Maryland, College Park, Maryland 20742, USA; email: dab@umd.edu

Annu. Rev. Vis. Sci. 2019. 5:451–77

First published as a Review in Advance on August 6, 2019

The *Annual Review of Vision Science* is online at vision.annualreviews.org

https://doi.org/10.1146/annurev-vision-091718-014731

Keywords

modeling, receptive field, neural coding, machine learning, neural networks

Abstract

With modern neurophysiological methods able to record neural activity throughout the visual pathway in the context of arbitrarily complex visual stimulation, our understanding of visual system function is becoming limited by the available models of visual neurons that can be directly related to such data. Different forms of statistical models are now being used to probe the cellular and circuit mechanisms shaping neural activity, understand how neural selectivity to complex visual features is computed, and derive the ways in which neurons contribute to systems-level visual processing. However, models that are able to more accurately reproduce observed neural activity often defy simple interpretations. As a result, rather than being used solely to connect with existing theories of visual processing, statistical modeling will increasingly drive the evolution of more sophisticated theories.

1. INTRODUCTION

Receptive field: a general term referring to the stimuli that drive a neuron's response; not used in this review due to its ambiguity

Visual processing is performed by billions of neurons spread across an array of visual areas. Understanding the function of such a system ultimately requires linking descriptions of visual processing with neural activity recorded across the visual system in appropriate visual contexts: where both the visual stimuli and neural responses are necessarily quite complex. As a result, links between function and recorded neuron activity necessarily require statistical models, which embody neural function in their mathematical form and are fit to the data such that their parameters correspond to specifics about each neuron's processing (such as its receptive field). However, the resulting functional descriptions are implicitly limited by the sophistication of the statistical models themselves, which is likewise constrained by both the experimental and computational approaches available.

The timing of this review coincides with a rapid expansion of statistical modeling in vision driven by technology development. On the experimental side, multi-electrode and imaging technologies enable simultaneous recording of increasing numbers of neurons (Jun et al. 2017, Stevenson & Körding 2011, Stringer et al. 2019), as well as longer recordings of individual neurons via chronic recording (McMahon et al. 2014, Oliver & Gallant 2017). On the theoretical side, breakthroughs associated with machine learning now allow optimization of increasingly sophisticated statistical models (Antolik et al. 2016, Batty et al. 2016, Cadena et al. 2019, Kindel et al. 2017, Klindt et al. 2017, Maheswaranathan et al. 2018b, Yamins & DiCarlo 2016), and computational resources allow the inclusion of many more parameters in such models, constrained by larger amounts of data (Paninski & Cunningham 2018).

In this review, I focus on statistical models describing visual processing, i.e., encoding models specifying what computation is performed on visual stimuli to result in the observed neural activity. While it is clear that visual neuron activity is driven by more than just the visual stimulus for functionally relevant purposes (Bondy et al. 2018, Cui et al. 2016a, Froudarakis et al. 2019, McFarland et al. 2016, Musall et al. 2018, Ni et al. 2018, Rabinowitz et al. 2015, Stringer et al. 2019, Vinck et al. 2015), modulation of stimulus processing by nonvisual elements in the awake animal is a rich subject beyond the scope of this review (Whiteway & Butts 2019). Thus, I begin the review (Section 2) by offering a definition of what is meant by visual computation and the resulting implications of such a definition for both experimental and model design. A second key consideration in statistical modeling is the methods used to fit the models—which have now largely converged on the statistical framework of maximum a posteriori (MAP) estimation (Paninski et al. 2007, Wu et al. 2006)—and I offer an intuitive description of this process (Section 3). The general adoption of this estimation framework has driven an explosion of new modeling approaches, which I describe in the context of two (often competing) goals: best predicting their response (Section 4) versus providing insight into biological mechanism and/or visual function (Section 5). The tension between model complexity and interpretability is driving the development of both new analytic tools and, ultimately, theories that both directly reflect and are possibly inspired by statistical modeling. In tandem, the recent machine learning–driven successes in computer vision (Kriegeskorte 2015, Krizhevsky et al. 2012, LeCun et al. 2015, Serre 2019) suggest a new range of possible approaches, as well as challenges, that we are only beginning to grapple with.

2. CHARACTERIZING VISUAL COMPUTATION WITH EXPERIMENTS

At the heart of any description of visual processing is the definition of the computation being performed, where computation refers to the transformation between the inputs and outputs. For vision, such a transformation can be thought of as a function that maps every visual stimulus $\mathbf{s}$ to a response r, such that $r = f(s)$ (**Figure 1**).

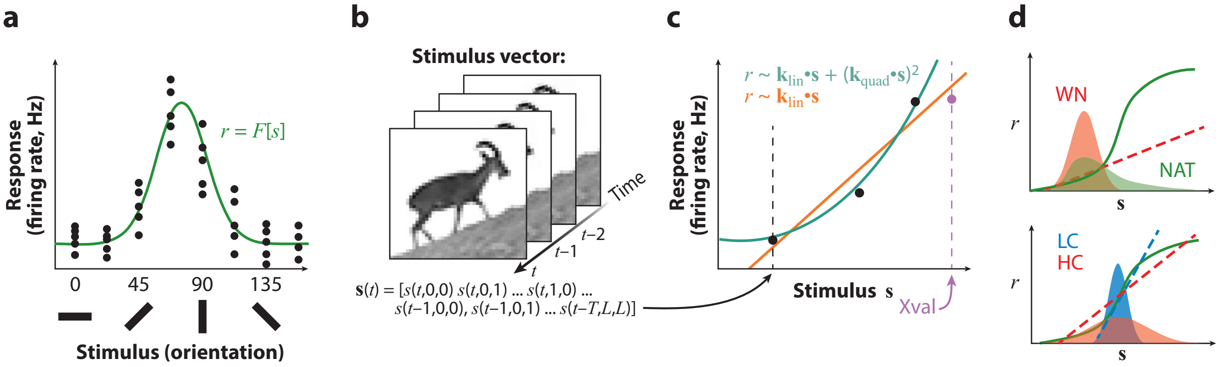

Figure 1

Neural computation: nonlinear functions in high-dimensional space. (*a*) A measured tuning curve for a V1 cell (inspired by Hubel & Wiesel 1962), showing data (*black points*) measured for multiple orientations of a flashed bar. Because the stimulus can be described by a single number (i.e., is one-dimensional), the experimenter can sample the full range of stimuli to get accurate measures of the neuron's computation $F[s]$ (*green*), which maps the stimulus to the neuron's response. (*b*) In comparison, natural visual stimuli are high-dimensional, extending across space and time, and must be represented as a vector $\mathbf{s}(t)$ that includes all luminance values across space (2,000 pixels in this picture, which is 50 × 40) as well as its recent history in time (time lags). With 10 time lags (for example), each stimulus $\mathbf{s}(t)$ at each time point would have 20,000 independently varying dimensions. (*c*) Thus, for general visual stimuli, experiments can only provide an extremely sparse sample of stimulus space [three data points shown, with a fourth point (*purple*) set aside for cross-validation (Xval)]. To describe the computation, it is necessary to assume a particular mathematical form (a linear and quadratic model are shown), with parameters fit to the existing data. In this example, both models fit the data points equivalently, although the simpler model (linear) cross-validates better. (*d*) The types of stimulus tested in a given experiment can affect the resulting models that emerge when the true computation is nonlinear (*green*). Different linear models are shown as dashed lines. (*Top*) A linear model might fit very well for stimulus distributions associated with the relatively narrow range of white noise (WN), compared with natural movies (NAT). (*Bottom*) Linear models can change depending on the size of the stimulus distribution [which is changed in high contrast (HC) versus low contrast (LC)], even if the true computation does not change.

Because most in vivo recordings of visual neurons are extracellular, the response considered is usually specified as a firing rate $r(t)$, defined as the spike count in a given window around a time point t. For other types of recordings, $r(t)$ can likewise be the membrane voltage, calcium fluorescence signal, and so on—but corresponds to a single value at each time point. While more sophisticated specification of the response is possible—such as distinguishing spike patterns over time (de Ruyter van Steveninck et al. 1997, Optican & Richmond 1987)—these details are often only apparent at high time resolutions generally considered to be unrelated to stimulus processing and are thus largely unaddressed in statistical modeling (although see Butts et al. 2007, 2011; Gollisch & Meister 2008b; VanRullen & Thorpe 2001).

Instead, challenges in describing visual computation are largely driven by the complexity of the stimulus itself. Experiments can be performed with very limited visual stimuli, such as in the work of Hubel & Wiesel (1962), where the orientation of an otherwise fixed bar was probed (**Figure 1*a***); the stimulus–response computation in this context can be completely determined using recorded data. However, outside of such reductionist experiments, visual stimuli typically require many variables to specify them (rather than a single variable such as orientation), and in such contexts, visual neurons are typically selective to complex combinations of luminances spread across space and time (**Figure 1*b***). Thus, descriptions of visual computation in these contexts specify the response of a given neuron across the stimulus space defined by a high-dimensional stimulus $\mathbf{s}$. [Throughout this review, I use boldface to represent a vector, which has many components, e.g., $\mathbf{s} = [s_1\ s_2\ s_3\ \ldots\ s_D]$ has a number for each independently varying element (**Figure 1*b***).] A key

Firing rate: the number of spikes in a given time window Δt (typically the time resolution of the experiment), divided by Δt

challenge in modeling visual neurons is that stimuli defined by more than a few independently varying components cannot be exhaustively presented in any reasonable experiment.

Natural stimuli: a wide class of complex visual stimuli usually generated based on photographs or videos, in contrast to parametric stimuli such as white noise

How does one then determine the mapping from stimulus to response in the context of such sparse sampling? A powerful approach is to assume a mathematical form to the function $f(s)$ that is able to approximate its true shape in high-dimensional space given parameters that can be fit to the data (**Figure 1*c***). There are a large number of mathematical forms that might be chosen, and each will have some amount of flexibility to fit the function in question, imparted by its particular parameters. As is explored in detail below, there are three main factors that influence the choice of mathematical form: (*a*) how easy it is to optimize the parameters (Section 3), (*b*) how well the resulting model can predict the data (Section 4), and (*c*) what insights about the visual neuron might be gained using the resulting model (Section 5).

From this perspective, the experimental data themselves will impact the choice and success of the resulting statistical model. First and foremost, the amount of data (i.e., length of the experiment) will dictate the number of model parameters that can be adequately constrained. Unfortunately, the relationship between the number of model parameters and the amount of data depends on many factors (see the sidebar titled How Much Data Is Required to Fit Statistical Models?), with more data allowing for more flexibility in model form.

Second, the choice of stimulus types presented also determines the degree to which the computation being fit will be constrained by the sampling of stimulus space (**Figure 1*d***). For example, relatively complex computations might appear to be simple when the experimental data only explore a localized region of stimulus space, such as when using white noise stimuli, compared with natural stimuli (David & Gallant 2005, Russ & Leopold 2015, Sinz et al. 2018, Talebi & Baker 2012). Similarly, approximate models might find different parameters when comparing responses in different stimulus contexts, such as in different contrasts (Cui et al. 2016b, Ozuysal & Baccus 2012, Shapley & Victor 1978) or with different spatial correlation structures (Coen-Cagli et al. 2015, Fournier et al. 2011).

The choice of stimulus distribution properties is especially important in the study of neurons in higher visual areas, which will often be selective to complex stimulus attributes that would be very improbable to come across by chance. In such cases, experiments based on noise-based paradigms or other artificially generated stimuli might not sufficiently sample the neuron's selectivity, and the resulting statistical models could misrepresent their function. Thus, the stimulus selection often implicitly contains assumptions about the function of the neurons being tested,

HOW MUCH DATA IS REQUIRED TO FIT STATISTICAL MODELS?

The amount of data recorded to fit a statistical model will directly affect how many parameters can be adequately constrained and, as a result, how sophisticated the model can be. However, the actual amount of data required to fit a given number of parameters will depend on many factors, including the form of the model itself, the reliability of the neuron, the complexity of the relevant stimulus space and how uniformly it is sampled (e.g., **Figure 1*d***), and the form and effectiveness of model regularization, with each of these factors potentially changing the amount of data required by orders of magnitude. For spiking data—as well as fluorescence imaging based on spiking—two key factors are the number of unique stimuli presented and the number of spikes recorded, both of which are related to but not determined by the total duration of the experiment. Intracellular data can in principle give much more information per time point to constrain models as long as the fluctuations in recorded voltage or current are meaningful with regard to stimulus processing. In general, the amount of data required to fit a given model must be empirically determined.

although this problem arguably can be mitigated by using natural visual stimuli (David et al. 2004, Sharpee 2013), which presumably contain the features that the visual system was designed to have. Alternatively, some approaches now iteratively optimize stimulus design to improve the resulting models (Bashivan et al. 2019, DiMattina & Zhang 2011, Lewi et al. 2011, Ponce et al. 2019), with the potential to converge to stimuli with the features that best drive the neuron and, by extension, describe its function.

Overall, experiment and model design together impact the quality of the ability to describe the data and provide insight into the function of the neuron.

3. FITTING STATISTICAL MODELS

Ideally, statistical models would be designed simply based on the assumed computations thought to be performed by the recorded neurons. Practically, the main consideration in statistical model design is often the ability to estimate parameters. The parameters of a given model can be represented as a vector $\boldsymbol{\Theta} = [\theta_1, \theta_2, \ldots \theta_\mathrm{P}]$ in a P-dimensional parameter space, and optimizing the parameters involves searching in this high-dimensional space for the parameters that yield the best model fit (**Figure 2*a***). Until relatively recently, nearly all statistical models of visual neurons used spike-triggered approaches for parameter estimation (Chichilnisky 2001, Simoncelli et al. 2004), thereby avoiding a direct search through this parameter space. However, modern computer power

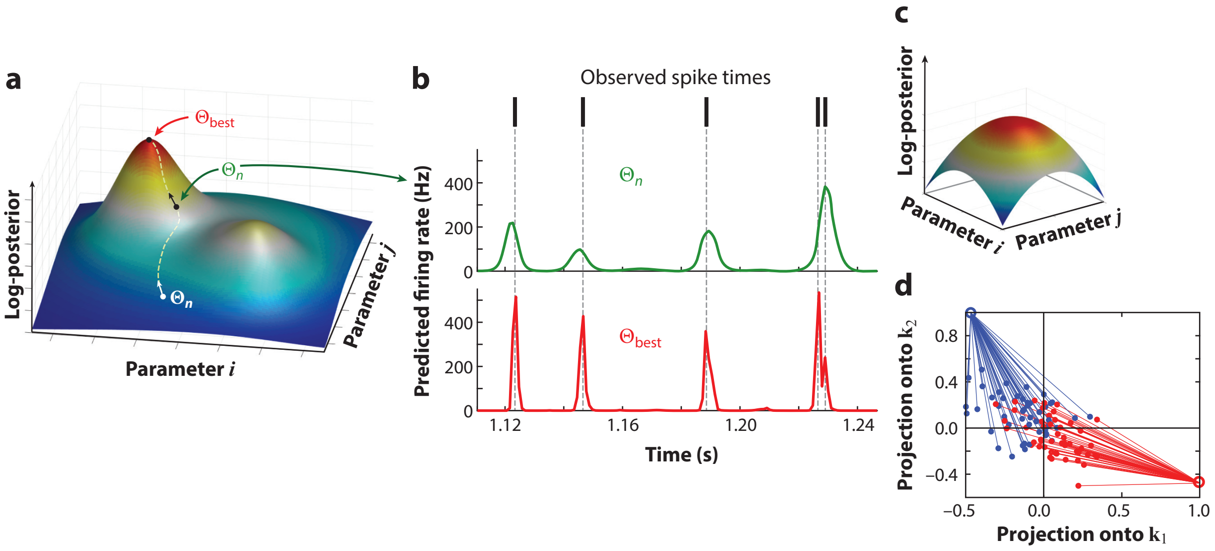

Figure 2

Parameter optimization and the log-posterior surface. (*a*) An illustration of how the log-posterior (LP) (vertical axis, also colored from *blue* to *red*) might depend on two parameters of a nonlinear model. An optimization path is shown, from the initial parameters $\boldsymbol{\Theta}_0$ and following the gradient (*black arrow*) to an LP-maximum at $\boldsymbol{\Theta}_\mathrm{best}$. This LP-surface pictured is relatively well-behaved, although gradient ascent could get stuck in the lower maximum (*right*) with some initial conditions. (*b*) A demonstration of how the model predictions (*green* and *red*) compare with the observed data (*black*) as the LP of the model improves. (*c*) A typical convex LP-surface, such as that associated with generalized linear models (GLMs). Such a surface only has a single local maximum, which can be found with any initial conditions. (*d*) A projection of the initial model filters [100 random initializations (*solid points*)] and final filters [after model optimization (*hollow points*)] in the filter space for an ON-OFF cell simulation (from McFarland et al. 2013, supplemental material). The optimization always converged to two positions in parameter space with equal LP, corresponding to the true filters in either order. Thus, while there is more than one local maximum of the LP, they are both correct answers, and the LP-surface is well-behaved.

has greatly expanded the capacity to fit models with a large number of parameters—and the correspondingly large amount of data required—and enabled a much larger range of statistical models based on direct searches in parameter space.

The choice of the mathematical form of the model has a critical effect on how easy it is to search in parameter space, as well as (of course) the overall number of parameters. Thus, an intuitive understanding of direct parameter estimation is necessary to understand the current breadth of statistical modeling in vision.

Noise model: assumed relationship between model predictions and observed data based on the type of noise in the system, such as Gaussian or Poisson

3.1. Maximum a Posteriori Estimation

Most frameworks for model parameter optimization in neuroscience fall under the category of MAP estimation (Wu et al. 2006). For a given mathematical form of the model that depends on parameters $\mathbf{\Theta}$, the best description of the data can be defined as the most probable choice of parameters that generated the data:

$$\begin{aligned}\mathbf{\Theta}_{\text{best}} &= \text{argmax}_{\mathbf{\Theta}} \Pr\left\{\mathbf{\Theta}|\text{data}\right\} = \text{argmax}_{\mathbf{\Theta}} \Pr\left\{\text{data}|\mathbf{\Theta}\right\} \times \Pr\left\{\mathbf{\Theta}\right\} / \Pr\left\{\text{data}\right\} \\ &= \text{argmax}_{\mathbf{\Theta}} \log \Pr\left\{\mathbf{\Theta}|\text{data}\right\} = \text{argmax}_{\mathbf{\Theta}} \left[\log \Pr\left\{\text{data}|\mathbf{\Theta}\right\} + \log \Pr\{\mathbf{\Theta}\}\right],\end{aligned} \qquad 1.$$

where Pr is shorthand for probability. The first line uses Bayes' rule to decompose the posterior probability Pr{$\mathbf{\Theta}$|data} into the probability of the parameters themselves Pr{$\mathbf{\Theta}$} and the probability that the data were generated by the model Pr{data|$\mathbf{\Theta}$}. Because the logarithm of the posterior [referred to as the log-posterior (LP)] will have maxima for the same parameters as the posterior itself, this optimization is performed over the LP (second line of Equation 1), allowing for the expression to involve a sum of terms, rather than a product. Note that because Pr{data} is independent of the model parameters $\mathbf{\Theta}$, it is dropped in the expression for the LP.

The first term of the LP, log Pr{data|$\mathbf{\Theta}$}, is the log-likelihood (LL). The LL represents how closely the predicted response matches the observed data (**Figure 2*b***) and depends on the noise model, which determines how likely a given observation $R_{\text{obs}}(t)$ is given the model prediction $r(t)$. For example, models that predict a firing rate can be linked to an observed spike count via a Poisson distribution (Paninski 2004) or a Bernoulli distribution at high temporal resolutions (Butts et al. 2016), whereas predictions of synaptic currents or fluorescence signals might best be captured by a Gaussian noise model.

The second term of the LP (Equation 1), Pr{$\mathbf{\Theta}$}, incorporates model regularization by placing assumptions on the parameters themselves. Given a limited amount of experimental data, different choices of model parameters may have nearly identical LLs, which means that any noise or other randomness in the recorded data can bias some parameter combinations over others. Models with a large number of parameters relative to the amount of data (to constrain them) will usually be able to achieve a better LL by fitting the noise on that particular data set, although such overfitting will result in worse performance on other cross-validation data sets (that do not have the same noise). As a result, specification of log-Pr{$\mathbf{\Theta}$} applies additional constraints on the parameters via a penalty term (subtracted from the LL). Different types of regularization penalties can codify assumptions about the model parameters, such as whether filters should be smoothly varying (e.g., McFarland et al. 2013), have few nonzero elements (e.g., Calabrese et al. 2011), or be space-time separable (Maheswaranathan et al. 2018a, Park & Pillow 2013, Shi et al. 2019).

The best model parameters will thereby maximize the LL while minimizing the regularization penalty. As with any Bayesian estimation, having fewer observed data, or lower-quality data with a shallower LL-surface, causes the optimal parameters to be more influenced by assumptions about the model structure, whereas in the limit of infinite (high-quality) data, the penalty term will have no effect. This balance is established through a weighting term corresponding to the width of the assumed distributions Pr{**Θ**}, which can be set using principled knowledge of the distributions (Sahani & Linden 2003) but are most often set empirically via nested cross-validation (Wu et al. 2006).

Nearly all statistical model estimation in neuroscience falls into this general framework of MAP estimation (Wu et al. 2006). First, as described above, MAP estimation is equivalent to maximum likelihood estimation if there are no assumptions made about the model parameters, i.e., Pr{**Θ**} is uniform and does not affect the optimal solution. Second, maximizing the LL is equivalent to maximizing the single-spike information between the model prediction and data (Kouh & Sharpee 2009, Rajan et al. 2013, Williamson et al. 2013), which is the basis for the Maximally Informative Dimension analysis (Sharpee et al. 2004) and iSTAC (Pillow & Simoncelli 2006). Finally, spike-triggered approaches (Simoncelli et al. 2004), such as spike-triggered averaging (Chichilnisky 2001, Reid et al. 1997) and spike-triggered covariance (STC) (Fairhall et al. 2006, Liu & Gollisch 2015, Schwartz et al. 2006, de Ruyter van Steveninck & Bialek 1988), are based on minimizing the mean-squared error between predicted and observed responses, which is equivalent to maximizing the LL under Gaussian assumptions for the noise distribution.

Model performance: a measure of how well the model predicts a cross-validation dataset; common measures include log-likelihood, fraction of explained variance (R-squared), and the correlation coefficient

3.2. Parameter Optimization Through Gradient Ascent

Thus, for a given data set and model form, every choice of model parameters **Θ** will have a corresponding LP, which might be thought of as a height at each point in parameter space (**Figure 2*a***). Parameter optimization can then be thought of as a search across this landscape for the highest location, corresponding to the maximum possible LP. Due to the large dimensionality of the parameter space, it is infeasible to search the whole space, and instead, efficient optimization strategies use a form of gradient ascent (Bishop 2006).

The gradient of the LP (with respect to the parameters **Θ**) provides the direction in parameter space of largest increase (slope) of the LP at any given point; thus, incrementally changing the parameters in the direction of the gradient will increase the LP until a local maximum is reached (**Figure 2*a***). Such a search is computationally efficient because it need only explore a single path in high-dimensional space, and also because such gradient computations can be efficiently implemented by increasingly available parallel computing hardware [e.g., graphical processing units (GPUs)].

The success of gradient ascent approaches ultimately relies on the LP-surface being well-behaved, meaning that the way in which the LP depends on the parameters **Θ** is smooth and has few peaks (**Figure 2*a***). Ideally, the LP-surface is convex (e.g., **Figure 2*c***), meaning that it has only a single maximum, in which case gradient ascent will locate this peak regardless of initial parameter choice. However, most statistical models will have many local maxima, and thus the model parameters found can depend on the initial condition (e.g., **Figure 2*d***). In such cases, gradient-based optimization is often performed using a random fraction of the available data on each iteration. Such stochastic gradient ascent (e.g., Kingma & Ba 2014, LeCun et al. 1989) effectively adds noise to the path, making it less likely to remain in smaller local maxima (Bishop 2006, Ruder 2016). Regardless, the mathematical form chosen can greatly affect how well-behaved the LP-surface is and subsequently how easy the model is to optimize.

4. STATISTICAL MODELS AS FUNCTION APPROXIMATORS

One overarching goal of statistical modeling is to find the best approximation to the computation being performed by the neuron, i.e., achieve the best model performance. This goal can be thought of in the framework of function approximation when describing the range of statistical models that are possible: More sophisticated models (with more parameters) will typically perform better in this respect, but likewise require more data (Section 2) and will be more difficult to fit (Section 3). The tradeoff between a model's simplicity and its ability to accurately predict neural responses will of course play out differently for different visual neurons, and model complexity is often explored systematically. Traditionally, the range of model complexity was explored by first starting with the linear (first-order) model—which can often offer meaningful approximations to more complex functions (e.g., **Figure 3*a***) while being easiest to fit—and adding higher-order terms via Wiener or Volterra expansions (Marmarelis 2004). A newer conception of model complexity—which I use below—instead starts the linear-nonlinear (LN) model and then increases model complexity by adding additional levels of LN processing, progressing to two-layer feature-space models and on to deeper neural networks.

4.1. Linear Processing and Feature Detection

The simplest mathematical form of a model that responds to a high-dimensional stimulus $\mathbf{s}(t)$ is linear, where the neuron's response is derived as a weighted sum over the stimulus dimensions, i.e.,

$$g_{\text{lin}}(t) = \sum_i k_i s_i(t) = \mathbf{k} \cdot \mathbf{s}(t). \qquad 2.$$

How each dimension of the stimulus $s_i(t)$ affects the response is given by the weighting factor k_i. The weight vector (or linear filter) $\mathbf{k}$ will have the same dimensionality as the stimulus, and thus can be thought of as a stimulus feature itself (**Figure 3*a***). This single feature defines the selectivity of the neuron because g_{lin} will be highest for stimuli that most closely resemble this feature. Most stimuli, however, will be orthogonal to a given filter $\mathbf{k}$ (**Figure 3*b***), meaning that they will most likely have a pixel combination uncorrelated with the components of $\mathbf{k}$ and thus have a value of g_{lin} close to zero.

The linear model can be trivially extended by processing the output of its filter by a simple (one-dimensional) function $F[.]$, resulting in the LN model:

$$r(t) = F[\mathbf{k} \cdot \mathbf{s}] = F[g_{\text{lin}}(t)]. \qquad 3.$$

The nonlinearity $F[.]$ determines how the neuron's response scales given the presence of $\mathbf{k}$ in the stimulus but does not change the feature selectivity of the neuron itself; as a result, $F[.]$ is not considered to be part of the stimulus processing and is either estimated after the linear filter $\mathbf{k}$ (e.g., Chichilnisky 2001) or kept to be a fixed mathematical form (e.g., Paninski 2004). For describing spiking data [when $r(t)$ is a firing rate], $F[.]$ is referred to as the spiking nonlinearity and is typically a rectifying function (**Figure 3*a***) to capture the non-negativity of the firing rate.

Until relatively recently, LN models have dominated sensory neuroscience. First, as described below (Section 5), they offer straightforward interpretations of neural function in terms of both computation (e.g., feature detection) and biophysical mechanism. Second, they can be easily fit through spike-triggered averaging (Chichilnisky 2001, Simoncelli et al. 2004) or gradient ascent in the context of the generalized linear model (GLM). Indeed, the LP-surfaces of GLMs are convex (**Figure 2*c***) given reasonable choices for the form of the spiking nonlinearity (Paninski 2004, Truccolo et al. 2005). As a result, they have become a powerful tool to fit stimulus filters simultaneously with linear weights on other experimental observables (Weber & Pillow 2017), including the neuron's own past history of spiking (i.e., spike-refractoriness) (Paninski 2004, Truccolo et al. 2005), spike trains from other neurons (Butts et al. 2016, Okun et al. 2015, Pillow et al. 2008), the

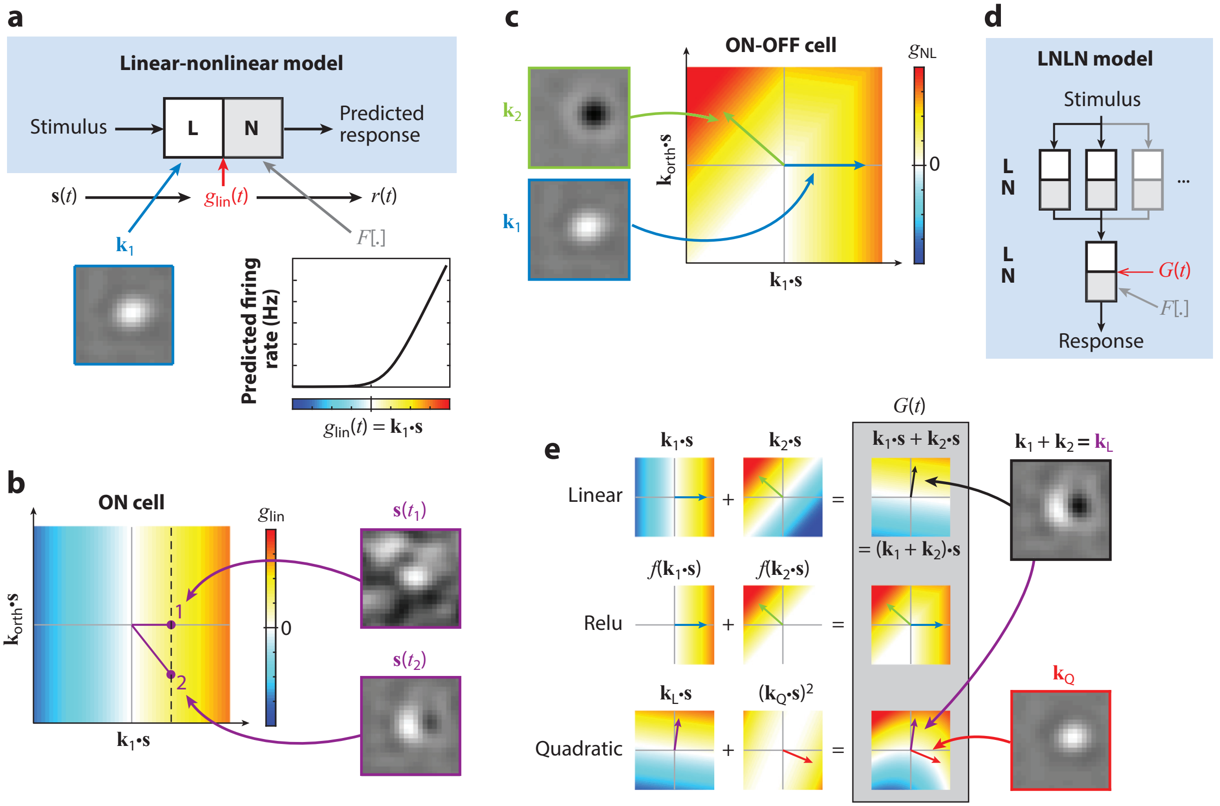

Figure 3

Nonlinear computation within feature spaces. (*a*) A schematic of stimulus space, with a different axis for each stimulus dimension, showing a linear filter $\mathbf{k}$ as a single direction in this space. Linear stimulus processing simply involves collapsing any stimulus in this entire space to this single axis $\mathbf{k}_1$, with all other axes not affecting the response. (*Bottom*) An example linear receptive field (an ON spatial filter of a mouse retinal ganglion cell) compared with two stimuli that result in the same linearly filtered output. (*b*) The dimensionality reduction achieved with linear filtering for a simulated ON-selective retinal ganglion cell, showing the linear output of the filter (*color*) along two dimensions of the stimulus space. The first (horizontal) axis corresponds to the filter of the ON cell ($\mathbf{k}_1$, shown in panel *a*), and the second axis is chosen to match that of other panels. Because linear processing is selective to only one stimulus direction (given by the filter), the color only depends on the direction given by $\mathbf{k}_1$, and stimuli 1 and 2 (shown projected into this plane) result in equivalent output. (*c*) In contrast, a simulated ON-OFF cell will have processing that depends on two directions in stimulus space, where the second dimension is given by an OFF filter $\mathbf{k}_2$, which is not exactly opposite $\mathbf{k}_1$ and so defines a second stimulus axis that the response depends on. (*d*) A model schematic for the LNLN cascade, composed of any number of feature-detecting linear-nonlinear (LN) components (subunits), whose output is linearly weighted and passed through a spiking nonlinearity. (*e*) Different mathematical forms of the model will generate different approximations to the ON-OFF cell response. (*Top*) A linear model of the ON-OFF processing with the two correct filters will only be able to generate selectivity in their average direction (*black*), i.e., $\mathbf{k}_1 \cdot \mathbf{s} + \mathbf{k}_2 \cdot \mathbf{s} = (\mathbf{k}_1 + \mathbf{k}_2) \cdot \mathbf{s}$. This filter (*right*) is identified by an LN model. (*Middle*) A rectified linear (relu)-based LNLN cascade will best approximate the true selectivity of the simulated neuron, in this case, because the simulation was generated by this form. (*Bottom*) An LNLN cascade with a linear and quadratic subunit can also offer an approximation of the true nonlinearity, although it generates different filters, corresponding to different directions within the true stimulus subspace. Additional examples of nonlinear function estimation are available in the **Supplemental Material**.

Supplemental Material >

local field potential (Cui et al. 2016a, Kelly et al. 2010, Rasch et al. 2008), and even task parameters such as locus of attention (Ni et al. 2018, Rabinowitz et al. 2015) and behavioral readouts such as pupil diameter (Musall et al. 2018, Stringer et al. 2019) and perception-based choice (Yates et al. 2017).

Finally, LN models have historically provided a good prediction of neural responses for some of the most heavily studied visual neurons (Shapley 2009), including neurons throughout the retina (Baccus & Meister 2002, Chichilnisky 2001), neurons in the lateral geniculate nucleus (LGN) (Carandini et al. 2005), and simple cells in the visual cortex (Carandini et al. 2005, Hubel & Wiesel 1962, Movshon et al. 1978b). However, the dominance of the LN model has been waning in the context of more sophisticated stimuli (David & Gallant 2005, Maheswaranathan et al. 2018b, Sharpee 2013) (**Figure 1*d***), the analysis of data at higher time resolution (Berry & Meister 1998; Butts et al. 2007, 2011; Keat et al. 2001; Uzzell & Chichilnisky 2004), increasing study of neurons deeper in the visual pathway (Movshon et al. 1978a, Serre 2015), and the availability of more powerful nonlinear statistical models (see below).

Nevertheless, linear modeling offers a first-order approximation to any neuron's computation and is also the first step of processing in nearly all more sophisticated models.

4.2. Feature Spaces and Multidimensional Nonlinearities

Linear stimulus processing is limited to capturing the dependence of a neuron's response to a single feature (**Figure 3*b***), and selectivity to multiple features requires nonlinear computation that combines the outputs of each separate feature detector. A general form of such computation is the generalized LN model, in which the response is a function of the separate linear processing of each feature, combined with a multidimensional spiking nonlinearity:

$$r(t) = F_N[\mathbf{k}_1 \cdot \mathbf{s}, \mathbf{k}_2 \cdot \mathbf{s}, \ldots \mathbf{k}_N \cdot \mathbf{s}]. \qquad 4.$$

This multidimensional function $F_N[.]$ returns a firing rate for each possible combination of N features: Each is a different projection of the stimulus onto the corresponding filter $\mathbf{k}_n$. Estimation of this feature space $\{\mathbf{k}_n\}$ is often a goal in itself, as it presumably contains the set of stimulus features that influence the response of the neuron (e.g., **Figure 3**).

The generalized LN model can thus simplify the problem of function approximation by mapping it into the lower-dimensional subspace spanned by the filters $\{\mathbf{k}_n\}$. For example, a multidimensional nonlinearity $F_N[.]$ of a model with only two features can be directly visualized (**Figure 3*c***), and even directly estimated given enough data (e.g., Cui et al. 2016b, Park et al. 2011, Rust et al. 2005, Touryan et al. 2002). In most cases, however, $F_N[.]$ will be higher dimensional (i.e., the neuron is selective to more than two features), and thus will still require a choice of parametric form.

4.2.1. Quadratic models.

Traditional system identification approaches in neuroscience were based on an explicit series expansion such as Wiener or Volterra series (Marmarelis 2004), which are multidimensional analogs to Taylor (power) series expansions of one-dimensional functions. These expansions typically do not go past the second-order term, the stimulus covariance, due to data limitations (although see Oliver & Gallant 2013). Cross-correlation of the stimulus covariance with the response can identify specific directions in stimulus space that modulate the response, resulting in a feature space (Fournier et al. 2014, Schwartz et al. 2006), which can be extracted by approaches such as STC analysis. Equivalent models are now typically fit via MAP estimation based on LNLN cascades (McFarland et al. 2013, Park & Pillow 2011, Rajan et al. 2013), described further below (**Figure 4**). As with standard mathematical series expansions, lower-order terms can in principle capture most of the neuron's computation in a localized region of stimulus space (e.g., **Figure 1*d***), which also explains why models that do not assume the functional form of subunit nonlinearity also find quadratic-like terms when fit with uncorrelated white noise stimuli (Rust et al. 2005; Touryan et al. 2002, 2005).

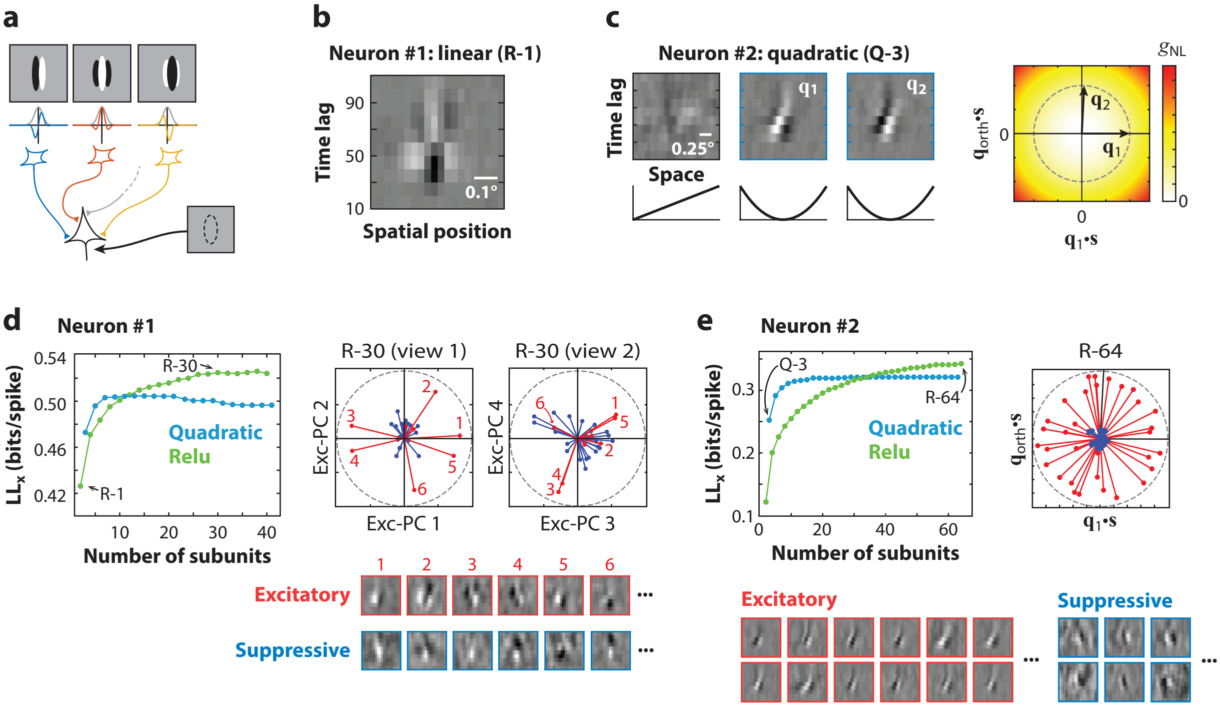

Figure 4

Understanding complex computations. (*a*) Early conceptual models of neurons in V1 include linear processing by simple cells and nonlinear processing by complex cells, following Hubel & Wiesel (1962): Phase-invariant processing by a complex cell (*black*) arises from inputs from simple cells (*colored*), each filtering the stimulus linearly with the same spatial frequency tuning and spatial position, but with different phases. Because of the (presumed) rectification before summation in the complex cell, the linear filter of the complex cell is largely canceled despite having a strong phase-invariant response to properly oriented gratings at that location. (*b*–*f*) Statistical models offer greater resolution in discerning the underlying computation performed and can thus interrogate theories of visual computation. This is demonstrated by two example V1 neurons and their associated models from a previous study (McFarland et al. 2014), neuron 1 (*b*,*d*), with a clear linear response, and neuron 2 (*c*,*e*), with almost no linear response. (*b*) The spatiotemporal filter of the linear-nonlinear (LN) model of neuron 1, which is depicted as a function of space (horizontal, scale bar = 0.1°) and time lag (vertical). Note that there is only one relevant spatial dimension probed by the random-bar experiment. (*c*) A quadratic LNLN cascade model fit to neuron 2 demonstrates clear complex cell tuning consistent with the energy model (Emerson et al. 1992), with two direction-selective (i.e., tilted in space-time) quadrature-pair filters (*left*). These filters span a two-dimensional subspace (*right*), with the combined filter output within this subspace with almost no phase dependence. (*d*, *left*) A range of LNLN cascades was fit to neuron 1, demonstrating that—despite its strong linear term (from panel *b*)—nonlinear models can offer significantly better descriptions of the neuron's computation. Indeed, most simple cells recorded in V1 typically reveal such complexity in the context of nonlinear models and/or detailed experimental studies (Fournier et al. 2014, Lochmann et al. 2013, Rust et al. 2005, Wilson et al. 2016). (*Right*) The best rectified linear (relu)-based model in this case has excitatory filters that span a higher-dimensional space, without clear relationships between them. The filters themselves represent a diverse array of features (*bottom*), and projecting them into the first four dimensions of the excitatory principle components shows that the excitatory filters fill out at least a four-dimensional space, which does not have any clear functional interpretation—other than suggesting something beyond current theories of V1 processing. All filters are represented by vectors of unit magnitude, and so would be on the dashed circle if completely within the shown two-dimensional subspace. (*e*, *left*) A range of LNLN cascades was fit to neuron 2, demonstrating that, while a relu-based model with a large number of subunits can provide the best description of its computation, it takes a large number of subunits to achieve better performance than the quadratic models. (*Bottom*) Unlike the excitatory filters of neuron 1, neuron 2 has filters that are very similar other than having spatially shifted features. Indeed, projecting the resulting subunit filters of the best-performing relu-based model demonstrates that excitatory filters ultimately project into the two-dimensional subspace found by the quadratic model (from panel *c*). Suppressive filters are clearly not explained by these filters, but the excitatory filters span all phases (see example filters at right), recapitulating the conceptual model in panel *a*. See the **Supplemental Material** for more information about these particular models.

Supplemental Material >

4.2.2. LNLN cascades.

A more general means of nonlinear function approximation is the LNLN cascade (**Figure 3*d***), comprised of LN elements (or subunits) that are each selective to a particular feature given by its linear filter. The neural response is generated by the sum of the subunit outputs followed by a spiking nonlinearity:

$$r(t) = F\left[\sum_n w_n f_n[\mathbf{k}_n \cdot \mathbf{s}(t)]\right] = F[g_{\mathrm{NL}}(t)]. \qquad 5.$$

Note, again, that the spiking nonlinearity typically does not influence stimulus processing, but simply scales $g_{\mathrm{NL}}(t)$ to the neural response. This functional form can be directly fit using standard numerical optimization methods (Section 3) and has thus largely replaced more limited forms based on series expansions.

Rectified linear (relu): a function that is linear for positive inputs and zero for negative inputs; this nonlinearity is now common in all forms of LN cascade models, including DNNs

Deep neural network (DNN): an artificial neural network comprised of feedforward stacks of LN components, which includes hierarchical convolutional neural networks

The key components determining the model parameters are the number of subunits and their associated nonlinearities $f_n(.)$, which can be directly fit (Butts et al. 2011, Maheswaranathan et al. 2018a, McFarland et al. 2013) but are typically fixed, with choices including rectified linear (relu) (McFarland et al. 2013), quadratic (Park & Pillow 2011, Park et al. 2013), sigmoid (Lau et al. 2002, Prenger et al. 2004), and cylindrical (Williamson et al. 2015) functions. Indeed, a general result from neural network approaches developed in the 1990s is that any nonlinear function can be represented by an LNLN cascade with appropriate subunit nonlinearities (Cybenko 1989, Hornik 1991, Kriegeskorte 2015).

LNLN cascades thus provide an approximation of the multidimensional nonlinearity $F_N[.]$ (Equation 4), with additional subunits (in principle) able to achieve increasingly accurate approximations. However, the number of subunits in the best-performing model will depend on how much each additional subunit increases the quality of the function approximation relative to the capacity for overfitting given its additional parameters. Thus, when estimating a complex nonlinear function, the amount of data available will typically limit the number of subunits that can be reliably fit (Lochmann et al. 2013). However, the dependence of the size of the model on the data also depends on how well the form of the model matches the computation of the neuron: A good choice of the subunit nonlinearity (e.g., quadratic, relu) can allow fewer subunits to more closely approximate the nonlinear function over the relevant space of the stimulus (e.g., **Figure 4**).

4.2.3. Deep neural networks.

Standard deep neural networks (DNNs) consist of stacks of LN units, with deeper layers receiving inputs from earlier units (**Figure 5**). The LN units comprising the first level of the DNN operate on the stimulus itself, and thus the first-level subunit filters comprise a feature space, much like the two-level LNLN cascades described above. In this sense, the remaining layers of the DNN provide another means of approximating the multidimensional nonlinearity $F_N[.]$ based on this feature space (Moskovitz et al. 2018). While LNLN cascades can in principle be used to describe any high-dimensional nonlinear function (Cybenko 1989, Hornik 1991, Kriegeskorte 2015), DNNs are often able to describe more complex nonlinear functions using many fewer parameters (Bengio 2009, Kriegeskorte 2015). Many recent advances in DNNs that replace simple LN-based layers with more sophisticated processing stages such as recurrent layers (Kriegeskorte 2015, Pascanu et al. 2013, Serre 2019) can further enhance the representations of the relevant nonlinear mappings and might be further guided by analogies between the structures of the DNNs and the visual system itself (**Figure 5**).

DNNs are also optimized using gradient ascent, although the expression for the gradient with respect to any one parameter calculation can be exceedingly complex and computationally intensive to compute. However, the mathematical expressions required for such function approximation are now automated in several machine learning tools, such as *TensorFlow* (Abadi et al. 2016) and

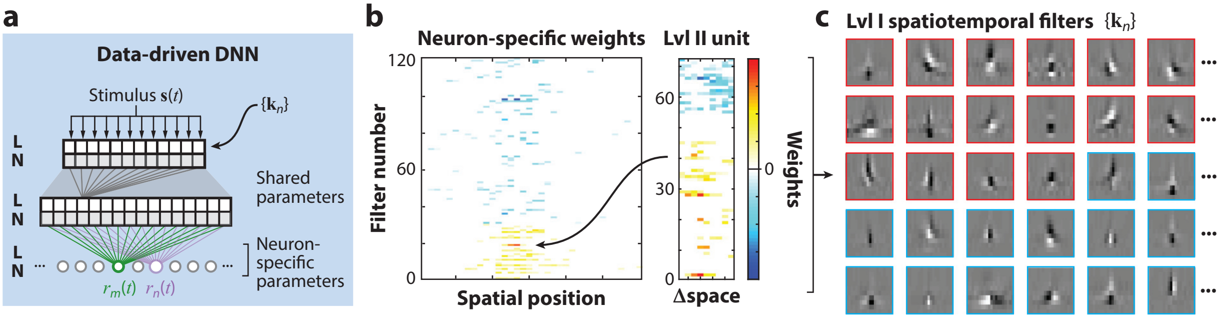

Figure 5

Data-driven deep neural networks (DNNs). (*a*) Schematic of a three-level DNN that was fit to a population of 101 V1 neurons responding to the random bar stimulus (same experiments as in **Figure 4**), adapted from Butts et al. (2018). This DNN yields average performance improvements across all V1 neurons of >20% above the best LNLN cascade models. (*b*, *left*) The predicted response $r_n(t)$ of an example V1 neuron is generated from its neuron-specific weights to the second level (Lvl II) of the DNN. Because Lvl II is convolutional, these weights act on specific filters at specific locations, and the neuron connects to a large number of filters at relatively localized spatial positions. (*Right*) The weights of a representative Lvl-II subunit demonstrate broad connectivity to Lvl I subunits. (*c*) The Lvl-I filters are spatiotemporal stimulus features and form a feature space from which all the neurons ultimately assemble their responses. However, due to the large interconnectivity from Lvl I to Lvl II to Lvl III, what a given neuron is selective to and how such selectivity is generated are essentially inscrutable.

PyTorch (Paszke et al. 2017), making such methods broadly accessible for using DNNs to fit a larger variety of data (LeCun et al. 2015). Furthermore, while the sheer number of parameters of such models would make it much more difficult to find a single global optimum, the number of dimensions in the space makes getting trapped in local maxima far from the optimum less likely because there are so many directions for movement (Kriegeskorte 2015). As a result, the LP-surface for DNNs can often be well-behaved (e.g., **Figure 2*a***), and the resulting parameters found through optimization tend to provide high performance, even if they are not globally optimal (e.g., **Figure 2*d***).

As a result, the main technical limitation of describing neural computation with DNNs is the amount of data that can be recorded in a given experiment (Yamins & DiCarlo 2016). While long chronic recordings might in principle yield enough data to fit reasonably sized DNNs to single neurons (Oliver & Gallant 2017, Zhang et al. 2019), most DNN approaches to date address such data limitations by fitting models of many neurons simultaneously (Antolik et al. 2016, Batty et al. 2016, Butts et al. 2018, Kindel et al. 2017, Klindt et al. 2017, Maheswaranathan et al. 2018b, Moskovitz et al. 2018), using a single model with shared components (**Figure 5**). Specifically, the population LP can be expressed as the sum of the individual neuron LPs, and in this sense, all the components of the separate models can be simultaneously optimized over a combined parameter space. Because any one neuron can draw on components from the shared network, it can have access to many more subunits, which are effectively constrained by data from other neurons. Such sharing naturally occurs between similarly tuned neurons and becomes increasingly likely in multilayer networks, particularly in early layers that represent simpler computations. Furthermore, as more neurons are added to the shared model, the number of parameters required for each additional neuron decreases, as most elements comprising their particular computation will already be present in the shared model. This permits much more sophisticated, parameter-intensive models to be fit to the same total duration of recording. As a result, recent DNN applications have outperformed shallower models in the retina (Batty et al. 2016, Maheswaranathan et al. 2018b, Moskovitz et al. 2018) and V1 (Butts et al. 2018, Cadena et al. 2019, Kindel et al. 2017, Klindt et al. 2017, Moskovitz et al. 2018, Oliver & Gallant 2017, Zhang et al. 2019) and are expected to be even more effective for higher visual areas (Mineault et al. 2014).

Poisson assumption: modeling assumption that the observed spike count of the neuron is representing a firing rate, specifying the probability of an observed spike count

Nevertheless, it should be noted that all of the approaches to fitting data discussed above have a feedforward structure, meaning that stimulus processing processes serially through stages of processing—even where individual layers are capable of recurrent connectivity (Liao & Poggio 2016, Pascanu et al. 2013). While such a feedforward approach is capable of capturing any nonlinear function (as discussed above), it is clear from the visual system itself that additional capabilities will arise from feedback connectivity (Kafaligonul et al. 2015, Kar et al. 2019, Shou 2010). Capacities for fitting nonlinear systems without an implicit feedforward structure are recently coming online (Pandarinath et al. 2018, Whiteway & Butts 2019), although these models are not yet close to having the structured stimulus processing capabilities of current DNNs.

5. INTERPRETABLE STATISTICAL MODELS

While there can be utility in achieving better predictions of neural responses via black-box modeling approaches (Benjamin et al. 2018), a valuable consideration in model design is their ability to provide insight into aspects of the visual computation that produces these predictions. Following the categories suggested by Marr & Poggio (1976), such functional insights might include (*a*) the global goal(s) of processing within the visual system, (*b*) how the neuron's computation contributes as part of this goal, and (*c*) the underlying cellular and circuit mechanisms implementing the neuron's computation.

5.1. Relating Neural Computation to Mechanism

The form of the statistical model can be informed by the physiological details of the neurons being modeled at the level of the cell, synapse, and circuit. By explicitly shaping their mathematical form based on such mechanisms, statistical models can accomplish two goals. First, they can be used to understand how a particular mechanism might contribute to the computation being performed by the neuron. Second, direct mathematical modeling of cellular and circuit mechanisms might untangle their complex interactions into localized elements within the model, which otherwise would be distributed across multiple subunits in the context of more general function approximations.

Biophysically motivated statistical models, in fact, start with the LN model (Equation 3). To first order, dendritic integration is often considered to be linear, and in this sense, the weights of the linear filter can be thought of as a proxy for the strength of synaptic inputs, and the output of the linear filter as an estimate of synaptic currents or even membrane potential (Mohanty et al. 2012, Shapley 2009). The nonlinearity then translates this proxy for membrane potential into a probability of spiking, under the Poisson assumption. Making spike generation an explicit stage of the model following the spiking nonlinearity results in the linear-nonlinear-Poisson (LNP) cascade. Considering the LNP model as the basis for a biophysical approximation, more realistic statistical models can be generated by better approximations of dendritic integration, spike generation, and details of synaptic inputs at the circuit level.

More mechanistic models of spike generation often start with explicit models of spike refractoriness, which can be approximated using a linear term that acts on a neuron's previous history of spiking (Berry & Meister 1998). Such spike history can be naturally incorporated into a GLM (Paninski 2004, Truccolo et al. 2005), as well as any other more sophisticated cascade model (Butts et al. 2011, 2016; McFarland et al. 2013). With the ability to capture spike-history effects explicitly, the Poisson assumption applied in most statistical models of spiking neurons can be replaced by a MAP-based leaky integrate-and-fire model (Paninski et al. 2004, Pillow et al. 2005). Such detailed modeling of the spike-generation process is particularly useful in describing spike patterns at millisecond time resolution (Berry & Meister 1998, Butts et al. 2011, Cui et al. 2016b, Keat et al. 2001, Pillow et al. 2005).

Other types of statistical models address the linear approximation of synaptic integration. First, dendritic integration itself can be nonlinear and modeled explicitly as an LNLN cascade (Poirazi et al. 2003). Similarly, having spiking neurons as presynaptic inputs, or other reasons for rectifying inputs such as a threshold for synaptic vesicle release, will mean that separate inputs will not add linearly and has also been explicitly modeled using relu-based LNLN cascades (e.g., **Figure 3*d***) (McFarland et al. 2013). Importantly, nonlinear inputs can be separately distinguished in such statistical models and have been used (particularly in the retina) to infer different bipolar cell inputs to retinal ganglion cells (Freeman et al. 2015, Liu et al. 2017, Maheswaranathan et al. 2018a, Turner & Rieke 2016), to separate ON and OFF inputs (**Figure 3*e***) (Gollisch & Meister 2008a, Shi et al. 2019), and for the distinct tuning of excitatory and inhibitory inputs (Butts et al. 2011, McFarland et al. 2013).

Statistical models can also capture multiplicative and divisive interactions arising from presynaptic inhibition (Cui et al. 2016b, Franke et al. 2017, Olsen & Wilson 2008), synaptic depression (Jarsky et al. 2011, Ozuysal & Baccus 2012), and the more general properties of response normalization (Carandini & Heeger 2012, Chance et al. 2002). Such explicitly fit nonlinear mechanisms potentially offer explanations for computational properties of these cells such as adaptation to contrast (Cui et al. 2016b, Mante et al. 2008, Ozuysal & Baccus 2012, Shapley & Victor 1978), the generation of temporal precision in early visual circuits (Butts et al. 2011, Cui et al. 2016b, Levy et al. 2013), nonlinear integration by Y-cells (Freeman et al. 2015, Schwartz et al. 2012, Victor & Shapley 1979), and contextual modulation (Coen-Cagli et al. 2015, Williamson et al. 2016).

However, explicit modeling of mechanistic elements within statistical models can severely compromise the ability to fit these models by adding too many extra parameters that cannot be supported by the data or—especially—by having complex relationships between those parameters and the resulting LP-surface (i.e., making it not well-behaved). As a result, statistical models have largely been restricted to incorporating only one or a few explicitly modeled mechanistic components. By contrast, full biophysical simulations at the cellular (Izhikevich 2007) or network (Markram et al. 2015) level typically do not attempt gradient-based approaches for parameter estimation, and instead often constrain parameters based on targeted experimental measurements.

Studies using mechanistically inspired statistical models also highlight the potential danger of drawing conclusions simply based on improved model performance. Because neural responses are shaped by many mechanisms in tandem—often with overlapping effects—an increase in model performance through the addition of mechanism-inspired components does not constitute proof of that mechanism's involvement. For example, several aspects of retinal ganglion cell responses can be explained by a delayed suppressive input, which could arise from spike history (Berry & Meister 1998), direct inhibition (Butts et al. 2011, 2016), presynaptic inhibition (Cui et al. 2016b), synaptic depression (Ozuysal & Baccus 2012), or some combination of these. Models of any one mechanism will thus seem to provide confirmation of the mechanism tested, without further targeted experiments to distinguish between them (Cui et al. 2016b).

5.2. Interpreting Neural Computation in Terms of Feature Detection

The mathematical form of statistical models can also be chosen based on a given neuron's hypothesized computational role(s) within the system. For visual neurons, descriptions of function are largely based on the concept of feature detection, whereby a given neuron's response signals the presence of a particular feature in the visual stimulus. The intuition of visual neurons as feature detectors has its foundation in linear models of visual neurons, whose filter is a stimulus feature with its output signaling the degree to which that feature is present (e.g., **Figure 3*a***). In this case, using an LN model (or GLM) to fit a given neuron uses the statistical model to identify the feature to which it is selective and thus offers a direct measurement of the neuron's function.

However, the concept of feature detection extends to all levels of the visual pathway, although it necessarily becomes more abstract at higher levels of processing. Instead of a particular pattern of luminance, a feature at an intermediate stage of processing might correspond to the amount of power at particular spatiotemporal frequencies (Emerson et al. 1992), or to the binocular disparity (i.e., the comparison of the difference in the location of a pattern between the left and right eyes) (Cumming & DeAngelis 2001, Henriksen et al. 2016). At the highest levels, one can consider features to be objects that generate the visual scene, which might necessarily be inferred from the composition of intermediate-level features. From this perspective, the hierarchical organization of visual areas has been traditionally viewed as a means for neurons to acquire selectivity to more complex—and more abstract—features, progressing from linear feature selectivity in the retina, to V1 simple cells (Carandini et al. 2005), to selectivity identification of objects, faces, and other abstract visual categories (Serre 2015).

The foundational idea for how more complex processing arises from level-to-level transformations was suggested in Hubel & Wiesel's (1962) initial studies of V1 complex cells. Complex cells are selective to oriented gratings, but in a way that cannot be described by linear processing due to their invariance to spatial phase. Hubel & Wiesel hypothesized that such phase invariance arises from nonlinear integration over simple cell inputs (**Figure 4*a***) that have selectivity to the same orientation but different phases. Such proposed selectivity can be explicitly captured by LNLN cascades (Movshon et al. 1978a), which recapitulate both the resulting energy model computation (Adelson & Bergen 1985, Emerson et al. 1992) with quadratic subunits (**Figure 4*c***) (Lochmann et al. 2013; Rust et al. 2005; Touryan et al. 2002, 2005) and discrete relu-based subunits spread across spatial phase (**Figure 4*d***). More generally, such nonlinear integration is thought of as a model by which neurons at any level develop invariance to irrelevant features of the visual computation, leading to concurrent selectivity to more abstract elements (DiCarlo et al. 2012).

Likewise, the summation over nonlinear elements implicit in the LNLN cascade of V1 complex cells might be generalized to any level of the visual pathway (**Figure 6*a***):

$$r(t) = F\left[\sum_n w_n \phi[s(t);\theta_n]\right], \qquad 6.$$

where the basis function $\phi[\mathbf{s}(t);\boldsymbol{\theta}_n] \equiv \phi_n[\mathbf{s}(t)]$ represents models of the outputs of upstream neurons, each with its own set of parameters $\boldsymbol{\theta}_n$. This type of model can be particularly useful when the computational space of a neuron's inputs can be approximated with a small number of free parameters. For example, the primate dorsal stream largely derives motion selectivity from direction-selective neurons in area V1 (Movshon & Newsome 1996) that can be represented via the motion-energy model (Adelson & Bergen 1985, Emerson et al. 1992). Using this as a basis in spatiotemporal frequency space (Nishimoto & Gallant 2011, Rust et al. 2006) and as patterned velocity processing (Cui et al. 2013) revealed a variety of more complex motion processing of MT neurons (Nishimoto & Gallant 2011), including selectivity to pattern versus component motion (Rust et al. 2006) and selectivity to three-dimensional motion (i.e., including depth) (Cui et al. 2013). A similar model applied to the downstream area MST based on a constructed MT input basis suggested how MST derives selectivity to self-motion from nonlinear combinations of MT units (**Figure 6*b***) (Mineault et al. 2012). In all of these cases, these models thus revealed both novel

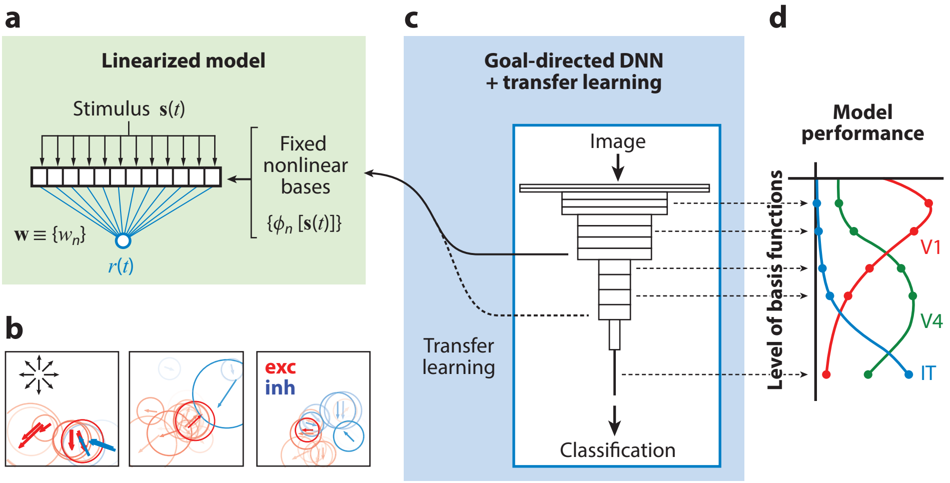

Figure 6

Hierarchical models and distributed computation. (*a*) A schematic for a linearized model, which is based on pregenerating a basis set of nonlinear elements {ϕ_n} that might explain a given neuron's response and fitting linear weights {w_n} to this basis. (*b*) A model of this form can describe the selectivity of MST neurons to visual motion generated from motion of the observer (self-motion) (Mineault et al. 2012), which is thought to be generated by combining MT neuron inputs with different direction selectivity and spatial positions. Such a model reveals how three MST neurons tuned for an expansion motion pattern generate their selectivity based on connectivity to idealized MT neurons, whose spatial selectivity is depicted by colored circles, direction selectivity by arrows, and weight by color and shading. Notably, the three diverse solutions suggest that there is more to MST selectivity than simple optic flow. (*c*) Linearized models can also be used in tandem with goal-directed deep learning to describe neurons in the context of global systems-level computation via transfer learning. This schematic demonstrates that a goal-directed deep neural network (DNN) (such as VGG-19; Cadena et al. 2019, Simonyan & Zisserman 2014) that is trained in visual classification tasks can be used to fit neural data via transfer learning. In this case, different layers of the network can be used as a nonlinear basis to fit neurons recorded in different visual areas. (*d*) A schematic depicting the results of several studies (Cadena et al. 2019, Kriegeskorte 2015, Yamins & DiCarlo 2016) demonstrating how linearized models fit to neuron responses from areas V1, V4, and IT will have the best performance based on bases from different levels of the goal-directed DNN, demonstrating a clear link between the distributed computations performed by goal-directed networks and the hierarchical organization of visual cortex.

aspects of feature selectivity within a given visual area and predictions for how such selectivity was derived.

The approach of precomputing the output of a basis set $\phi[\mathbf{s}(t);\theta_n]$ that captures difficult-to-fit nonlinearities and leaves model estimation to the linear weights {w_n} is known as linearized modeling (Wu et al. 2006) and has a number of more general applications. In particular, a nonlinear basis set representing inputs to a given neuron can be generated through many means and has recently been used, for example, to describe face-selective neurons from area IT based on simple linear transformations on extensively preprocessed (but fixed) images of faces (Chang & Tsao 2017). At another extreme, a large preprocessed basis set that included visual features, hierarchically transformed signals, and semantic categories was used to broadly classify selectivity of every voxel within both well-studied and novel areas of the visual cortex using functional magnetic resonance imaging (Agrawal et al. 2014, Naselaris et al. 2009). As discussed below, linearized models are also used in transfer learning (Yamins & DiCarlo 2016), which links recordings of neural activity to goal-directed deep networks.

Why not just fit the entire network at once rather than precompute a basis, as with the data-driven DNNs considered above? First, given the ability to appropriately select the basis functions $\{\phi_n\}$, linearized models can have substantially fewer parameters (i.e., the linear weights w_n only) that can be optimized as a GLM, which can greatly simplify model optimization and requires fewer data. Second, the computational properties of the basis functions are generally well understood (e.g., as feature detectors themselves). As a result, the model parameters (Equation 6) will specify how the recorded neurons' feature selectivity is constructed from the simpler selectivity in its inputs, thus implicitly isolating the relevant computation performed by that neuron. In contrast, the hidden layers of a DNN are not constrained to segregate computations between levels in the same way that the brain does, and the relevant computations of a given neuron will likely be spread across its many levels.

However, the success of these models depends on an appropriate model for the inputs $\phi_n[\mathbf{s}(t)]$, and—unlike fully data-driven approaches—will rely on the underlying assumptions generating the basis functions. While such assumptions can be specific to previous characterizations of inputs (as with models of neurons in areas MT and MST described above), models of this form can also be used to as a means to use recorded neural data to interrogate visual computation at the systems level.

5.3. Goal-Directed Deep Networks and Distributed Computation

At the global level, visual neuron computation might be understood in terms of the goals of the visual system as a whole. Goals, in this context, refer to solving tasks such as object and face recognition and have played a crucial role in driving the development of DNNs (LeCun et al. 2015, Schmidhuber 2015, Serre 2019). Such goal-driven DNNs (Cadena et al. 2019, Yamins & DiCarlo 2016)—in contrast to data-driven DNNs designed to reproduce neural data—have recently begun to achieve human-level performance, due in large part to the ability to train with extremely large curated image data sets. Despite the fact that such networks are not explicitly constrained to represent the same computations as neurons in the visual system, striking similarities emerge between their properties and those of the visual system (Hong et al. 2016, Kriegeskorte 2015, Yamins & DiCarlo 2016). This suggests that goal-directed DNNs provide an additional means to predict visual neuron responses directly via transfer learning (Yamins & DiCarlo 2016) and thus to understand their function in this larger context.

One can do so by explicitly simulating the responses of units throughout a trained goal-directed network and then fitting a linearized model (Equation 6) to the outputs (**Figure 6*c***). Such transfer learning is surprisingly effective (Yamins & DiCarlo 2016) and can result in model performances that rival those of data-driven networks (Cadena et al. 2019). Furthermore, the levels of the goal-directed DNNs that contribute the most to the predictions of each neuron typically have a clear correspondence to the area from which that neuron was recorded (e.g., **Figure 6*d***): IT neurons fit best using the deeper levels of the DNNs, and neurons from V1 and V4 to earlier levels (Cadieu et al. 2014, Khaligh-Razavi & Kriegeskorte 2014, Yamins et al. 2014). In addition to providing overall validation for the conception of hierarchal visual system organization, this also suggests that biological neuron selectivity might be explained by these same goals.

Along these lines, linearized models that linearly integrate (via their weights) over units within a given level of a network have the same functional form for regression-based decoders, which might solve a classification task by adjusting their weights (via linear or logistic regression) (Bishop 2006). While information about a task such as object classification is in principle available throughout the visual pathway, it can only be linearly decoded after a number of transformations are performed (DiCarlo et al. 2012). In this way, transfer learning is the converse of decoding studies using population neural activity and provides a complementary means to link neural activity to the systems-level computations performed across the visual hierarchy (Glaser et al. 2017).

At the same time, the insights that such an approach offers into the function of individual neuron classes—and even visual areas—are implicitly limited by this approach because of the current lack of biologically motivated design of goal-directed DNNs. Goal-directed DNNs have many configurations that can achieve similar performance (and performance comparable to that of human perception) and thus do not require the same visual areas, nor the specific anatomy and connectivity in the visual system. From this perspective, the visual system itself might be a large function approximator (Marblestone et al. 2016), with individual visual neurons only understandable as part of this global function (Robinson 1992). Indeed, unperturbed visual function in the context of brain damage or selective inactivation of certain visual areas (Jazayeri & Afraz 2017) might support such a view.

Furthermore, goal-directed DNNs require explicit specification of a goal, which the network is then trained to achieve. While tasks such as object and scene recognition might be explanatory for some aspects of neural responses, the same neurons might also contribute to other tasks not being probed (e.g., Russ & Leopold 2015). Furthermore, in animals other than primates (such as mice), it is not clear whether visual system goals might be referenced to object recognition at all, which might be reflected in observed differences in the higher areas of the visual cortex.

Nevertheless, goal-driven DNNs offer the ability to reference individual neuron responses in terms of global goals. Thus, they might serve as a foundation for incorporating more features of biological visual systems (see above) and aid in formalizing specific constraints on visual neuron function based on recorded visual neuron activity.

6. CONCLUSION

In the past decade, neurophysiology has entered the era of big data (Gao & Ganguli 2015, Paninski & Cunningham 2018, Stevenson & Körding 2011), driven by advances in experimental neurophysiology, computer technology, and the increasing sophistication of modeling frameworks to describe such increasingly complex data sets. As might be expected, the resulting descriptions of visual neuron function have been correspondingly complex. The use of simple stimuli leads to simple models of neural computation, and the combination of complex (e.g., natural) stimuli and flexible nonlinear modeling frameworks might be what is necessary to finally move beyond what Hubel & Wiesel's (1962) initial hand-mapping experiments from the past century suggested (Olshausen & Field 2005, Rust & Movshon 2005).

However, the success of increasingly sophisticated models has not necessarily been coupled with a greater understanding of visual system function. While achieving better predictions of visual neuron responses with black-box models has some intrinsic utility (Benjamin et al. 2018), the very elements that contribute to the improved performance of these models are not clearly consistent with more traditional conceptions of the role of neurons in visual processing. As described above, one hope is that such complexity might result from a hierarchically stacked set of canonical computations (Kouh & Poggio 2008, Orban 2008) and thus be decomposed into understandable elements with the appropriately tailored statistical model. Such computations might be uncovered, for example, by future applications involving DNNs tailored to specific elements of visual system components at the cellular and circuit levels.

An alternative possibility—suggested by parallels between units in goal-driven DNNs and visual neurons—is that the visual system itself behaves as a goal-directed network (Marblestone et al. 2016), with its function specified at the system level, and with no meaningful constraints that make function at the individual neuron level interpretable. In this sense, the computations performed by neurons within any given animal's visual system might be representative of one of many solutions for solving visual tasks required by that animal and influenced by its evolutionary history, including particular constraints based on the specifics of its brain architecture and mechanisms of synaptic plasticity and computations on the cellular and circuit levels. From this perspective, a natural question is how one can hope to back out how the system works based on recordings from a randomly sampled fraction of individual elements (Robinson 1992).

Likewise, while modern statistical models can now explain most of the visual neuron responses early in the visual pathway—predominantly in the retina (e.g., Cui et al. 2016b)—the best statistical models explain at best around half of the response variance in the cortex (e.g., Cadena et al. 2019, Cui et al. 2016a, Kindel et al. 2017, Klindt et al. 2017), raising the question of what they are missing about stimulus processing (Olshausen & Field 2005). Much of the unexplained variance, however, is likely the result of uncontrolled factors affecting neural activity (Whiteway & Butts 2019), including eye movements (McFarland et al. 2014, 2016); task context, including resulting top-down signals (Bondy et al. 2018, Ni et al. 2018, Nienborg et al. 2012, Rabinowitz et al. 2015); and behavioral outputs that have no clear function with regard to stimulus processing (Musall et al. 2018, Stringer et al. 2019, Vinck et al. 2015). The relatively large impact of such behavior on neural activity throughout the visual pathway likewise suggests its importance in incorporating broader (nonvisual) elements in descriptions of visual neuron function (Froudarakis et al. 2019), as well as the possibility of functional roles of visual neurons outside of stimulus processing, which might also be addressed via more general statistical models (Pandarinath et al. 2018, Paninski & Cunningham 2018, Whiteway & Butts 2019).

Thus, the rapid expansion of statistical modeling in vision described in this review has been driven by experimental and computational technology development in its relative infancy. As a result, the advances in statistical modeling detailed above are likely the beginning of a new era of data-based interrogation of theories of visual neuron function—in relevant behavioral contexts—which should result in a similarly dramatic expansion of these theories as this expanding field matures.

SUMMARY POINTS

1. Statistical models provide a powerful means to use recorded neural activity to interrogate the computations performed throughout the visual pathway.
2. Statistical models are defined by a mathematical form that can instantiate knowledge about the specific types of computations performed, with parameters that are fit to data. The quality of the resulting fits to the data demonstrates how good a description of the neural data the particular model provides.
3. The choice of the form of the model is shaped not only by the underlying knowledge of the neurons being modeled, but also the ability to fit its parameters. We are in the midst of a dramatic expansion of new approaches for fitting statistical models enabled by technological advances in computation.
4. Statistical models can be used to understand visual neuron function on all three levels suggested by Marr: (*a*) understanding the contribution of mechanism to neural computation, (*b*) testing hypotheses of what computations are being performed, and (*c*) relating neural computation to system- and organism-level function.

FUTURE ISSUES

1. Increasingly sophisticated statistical models are able to offer superior descriptions of the data, but often at the expense of interpretability of the resulting models.
2. Future progress in statistical modeling of visual neurons will involve linking models of individual neurons (the focus of this review) to computations performed at the systems level. This is increasingly possible through advances in large-scale recording techniques and in machine learning.
3. Interaction between computer vision applications such as goal-directed DNNs give the ability to understand neural function at the systems level, but further progress will require incorporation of more biologically realistic components into artificial neural networks. This could include details of cell types and circuitry, as well as model structures that include recurrency and feedback connections.
4. Much of the neural variability associated with studies of stimulus processing is often averaged away but is likely a critical component of understanding cortical function. A full understanding of visual neuron function thus likely requires incorporation of inputs that are not explicitly stimulus driven, including aspects of brain state and behavior.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The author would like to thank J. Antolik, F. Bartsch, B. Cumming, A. Ecker, V. Kowtha, and M. Whiteway for comments on the manuscript and would like to acknowledge support from the National Science Foundation grant IIS-1350990 and National Institutes of Health grant EY025403-02.

LITERATURE CITED

Abadi M, Barham P, Chen J, Chen Z, Davis A, et al. 2016. TensorFlow: a system for large-scale machine learning. In *OSDI '16: Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, pp. 265–83. Berkeley, CA: USENIX Assoc.

Adelson EH, Bergen JR. 1985. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* 2(2):284–99

Agrawal P, Stansbury D, Malik J, Gallant JL. 2014. Pixels to voxels: modeling visual representation in the human brain. arXiv:1407.5104 [q-bio.NC]

Antolik J, Hofer SB, Bednar JA, Mrsic-Flogel TD. 2016. Model constrained by visual hierarchy improves prediction of neural responses to natural scenes. *PLOS Comput. Biol.* 12(6):e1004927

Baccus SA, Meister M. 2002. Fast and slow contrast adaptation in retinal circuitry. *Neuron* 36(5):909–19

Bashivan P, Kar K, DiCarlo JJ. 2019. Neural population control via deep image synthesis. *Science* 364(6439):eaav9436

Batty E, Merel J, Brackbill N, Heitman A, Sher A, et al. 2016. *Multilayer recurrent network models of primate retinal ganglion cell responses*. Paper presented at the International Conference on Learning Representations, Toulon, France, April 24–26

Bengio Y. 2009. Learning deep architectures for AI. *Found. Trends Mach. Learn.* 2(1):1–127

Benjamin AS, Fernandes HL, Tomlinson T, Ramkumar P, VerSteeg C, et al. 2018. Modern machine learning as a benchmark for fitting neural responses. *Front. Comput. Neurosci.* 12:56

Berry MJ II, Meister M. 1998. Refractoriness and neural precision. *J. Neurosci.* 18(6):2200–11

Bishop CM. 2006. *Pattern Recognition and Machine Learning*. Berlin: Springer

Bondy AG, Haefner RM, Cumming BG. 2018. Feedback determines the structure of correlated variability in primary visual cortex. *Nat. Neurosci.* 21(4):598–606

Butts DA, Bartsch F, Whiteway MR, Cumming BG. 2018. *Characterizing hierarchical computation in primary visual cortex*. Prog. 141.28, Neurosci. Meet. Plan., Soc. Neurosci. Online, Washington, DC

Butts DA, Cui Y, Casti ARR. 2016. Nonlinear computations shaping temporal processing of precortical vision. *J. Neurophysiol.* 116(3):1344–57

Butts DA, Weng C, Jin JZ, Alonso J-M, Paninski L. 2011. Temporal precision in the visual pathway through the interplay of excitation and stimulus-driven suppression. *J. Neurosci.* 31(31):11313–27

Butts DA, Weng C, Jin JZ, Yeh C-I, Lesica NA, et al. 2007. Temporal precision in the neural code and the timescales of natural vision. *Nature* 449(7158):92–95

Cadena SA, Denfield GH, Walker EY, Gatys LA, Tolias AS, et al. 2019. Deep convolutional models improve predictions of macaque V1 responses to natural images. *PLOS Comput. Biol.* 15(4):e1006897

Cadieu CF, Hong H, Yamins DLK, Pinto N, Ardila D, et al. 2014. Deep neural networks rival the representation of primate IT cortex for core visual object recognition. *PLOS Comput. Biol.* 10(12):e1003963

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(1):e16104

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(46):10577–97

Carandini M, Heeger DJ. 2012. Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* 13(1):51–62

Chance FS, Abbott LF, Reyes AD. 2002. Gain modulation from background synaptic input. *Neuron* 35(4):773–82

Chang L, Tsao DY. 2017. The code for facial identity in the primate brain. *Cell* 169(6):1013–14

Chichilnisky EJ. 2001. A simple white noise analysis of neuronal light responses. *Network* 12(2):199–213

Coen-Cagli R, Kohn A, Schwartz O. 2015. Flexible gating of contextual influences in natural vision. *Nat. Neurosci.* 18(11):1648–55

Cui Y, Liu LD, Khawaja FA, Pack CC, Butts DA. 2013. Diverse suppressive influences in area MT and selectivity to complex motion features. *J. Neurosci.* 33(42):16715–28

Cui Y, Liu LD, McFarland JM, Pack CC, Butts DA. 2016a. Inferring cortical variability from local field potentials. *J. Neurosci.* 36(14):4121–35

Cui Y, Wang YV, Park SJH, Demb JB, Butts DA. 2016b. Divisive suppression explains high-precision firing and contrast adaptation in retinal ganglion cells. *eLife* 5:e19460

Cumming BG, DeAngelis GC. 2001. The physiology of stereopsis. *Annu. Rev. Neurosci*. 24:203–38

Cybenko G. 1989. Approximation by superpositions of a sigmoidal function. *Math. Control Signal Syst.* 2(4):303–14

David SV, Gallant JL. 2005. Predicting neuronal responses during natural vision. *Network* 16(2–3):239–60

David SV, Vinje WE, Gallant JL. 2004. Natural stimulus statistics alter the receptive field structure of V1 neurons. *J. Neurosci.* 24(31):6991–7006

de Ruyter van Steveninck R, Bialek W. 1988. Real-time performance of a movement-sensitive neuron in the blowfly visual system: coding and information transfer in short spike sequences. *Proc. R. Soc. B* 234(1277):379–414

de Ruyter van Steveninck RR, Lewen GD, Strong SP, Koberle R, Bialek W. 1997. Reproducibility and variability in neural spike trains. *Science* 275(5307):1805–8

DiCarlo JJ, Zoccolan D, Rust NC. 2012. How does the brain solve visual object recognition? *Neuron* 73(3):415–34

DiMattina C, Zhang K. 2011. Active data collection for efficient estimation and comparison of nonlinear neural models. *Neural Comput*. 23(9):2242–88

Emerson RC, Bergen JR, Adelson EH. 1992. Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vis. Res.* 32(2):203–18

Fairhall AL, Burlingame CA, Narasimhan R, Harris RA, Puchalla JL, Berry MJ II. 2006. Selectivity for multiple stimulus features in retinal ganglion cells. *J. Neurophysiol.* 96(5):2724–38

Fournier J, Monier C, Levy M, Marre O, Sári K, et al. 2014. Hidden complexity of synaptic receptive fields in cat V1. *J. Neurosci.* 34(16):5515–28

Fournier J, Monier C, Pananceau M, Frégnac Y. 2011. Adaptation of the simple or complex nature of V1 receptive fields to visual statistics. *Nat. Neurosci.* 14(8):1053–60

Franke K, Berens P, Schubert T, Bethge M, Euler T, Baden T. 2017. Inhibition decorrelates visual feature representations in the inner retina. *Nature* 542(7642):439–44

Freeman J, Field GD, Li PH, Greschner M, Gunning DE. 2015. Mapping nonlinear receptive field structure in primate retina at single cone resolution. *eLife* 4:e05241

Froudarakis E, Fahey PG, Reimer J, Smirnakis SM, Tehovnik EJ, Tolias AS. 2019. The visual cortex in context. *Annu. Rev. Vis. Sci.* 5:317–39

Gao P, Ganguli S. 2015. On simplicity and complexity in the brave new world of large-scale neuroscience. *Curr. Opin. Neurobiol.* 32:148–55

Glaser JI, Chowdhury RH, Perich MG, Miller LE, Körding KP. 2017. Machine learning for neural decoding. arXiv:1708.00909 [q-bio.NC]

Gollisch T, Meister M. 2008a. Modeling convergent ON and OFF pathways in the early visual system. *Biol. Cybern.* 99(4–5):263–78

Gollisch T, Meister M. 2008b. Rapid neural coding in the retina with relative spike latencies. *Science* 319(5866):1108–11

Henriksen S, Tanabe S, Cumming BG. 2016. Disparity processing in primary visual cortex. *Philos. Trans. R. Soc. Lond. B* 371(1697):20150255

Hong H, Yamins DLK, Majaj NJ, DiCarlo JJ. 2016. Explicit information for category-orthogonal object properties increases along the ventral stream. *Nat. Neurosci.* 19(4):613–22

Hornik K. 1991. Approximation capabilities of multilayer feedforward networks. *Neural Netw*. 4(2):251–57

Hubel DH, Wiesel TN. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160:106–54

Izhikevich EM. 2007. *Dynamical Systems in Neuroscience*. Cambridge, MA: MIT Press

Jarsky T, Cembrowski MS, Logan SM, Kath WL, Riecke H, et al. 2011. A synaptic mechanism for retinal adaptation to luminance and contrast. *J. Neurosci.* 31(30):11003–15

Jazayeri M, Afraz A. 2017. Navigating the neural space in search of the neural code. *Neuron* 93(5):1003–14

Jun JJ, Steinmetz NA, Siegle JH, Denman DJ, Bauza M, et al. 2017. Fully integrated silicon probes for high-density recording of neural activity. *Nature* 551(7679):232–36

Kafaligonul H, Breitmeyer BG, Öğmen H. 2015. Feedforward and feedback processes in vision. *Front. Psychol.* 6:279

Kar K, Kubilius J, Schmidt K, Issa EB, DiCarlo JJ. 2019. Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nat. Neurosci.* 22:974–83

Keat J, Reinagel P, Reid RC, Meister M. 2001. Predicting every spike: a model for the responses of visual neurons. *Neuron* 30(3):803–17

Kelly RC, Kass RE, Smith MA, Lee TS. 2010. Accounting for network effects in neuronal responses using L1 regularized point process models. *Adv. Neural Inf. Process. Syst.* 23(2):1099–107

Khaligh-Razavi S-M, Kriegeskorte N. 2014. Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLOS Comput. Biol.* 10(11):e1003915

Kindel WF, Christensen ED, Zylberberg J. 2017. Using deep learning to reveal the neural code for images in primary visual cortex. arXiv:1706.06208 [q-bio.NC]

Kingma DP, Ba J. 2014. Adam: a method for stochastic optimization. arXiv:1412.6980 [cs.LG]

Klindt DA, Ecker AS, Euler T, Bethge M. 2017. Neural system identification for large populations separating "what" and "where." *Adv. Neural Inf. Process.* 31:3509–19

Kouh M, Poggio TA. 2008. A canonical neural circuit for cortical nonlinear operations. *Neural Comput.* 20(6):1427–51

Kouh M, Sharpee TO. 2009. Estimating linear-nonlinear models using Renyi divergences. *Network* 20(2):49–68

Kriegeskorte N. 2015. Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annu. Rev. Vis. Sci.* 1:417–46

Krizhevsky A, Sutskever I, Hinton GE. 2012. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 25:1097–105

Lau B, Stanley GB, Dan Y. 2002. Computational subunits of visual cortical neurons revealed by artificial neural networks. *PNAS* 99(13):8974–79

LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. *Nature* 521(7553):436–44

LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, et al. 1989. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* 1(4):541–51

Levy M, Fournier J, Frégnac Y. 2013. The role of delayed suppression in slow and fast contrast adaptation in V1 simple cells. *J. Neurosci.* 33(15):6388–400

Lewi J, Schneider DM, Woolley SMN, Paninski L. 2011. Automating the design of informative sequences of sensory stimuli. *J. Comput. Neurosci.* 30(1):181–200

Liao Q, Poggio TA. 2016. Bridging the gaps between residual learning, recurrent neural networks and visual cortex. arXiv:1604.03640 [cs.LG]

Liu JK, Gollisch T. 2015. Spike-triggered covariance analysis reveals phenomenological diversity of contrast adaptation in the retina. *PLOS Comput. Biol.* 11(7):e1004425

Liu JK, Schreyer HM, Onken A, Rozenblit F, Khani MH, et al. 2017. Inference of neuronal functional circuitry with spike-triggered non-negative matrix factorization. *Nat. Commun.* 8:149

Lochmann T, Blanche TJ, Butts DA. 2013. Construction of direction selectivity through local energy computations in primary visual cortex. *PLOS ONE* 8(3):e58666

Maheswaranathan N, Kastner DB, Baccus SA, Ganguli S. 2018a. Inferring hidden structure in multilayered neural circuits. *PLOS Comput. Biol.* 14(8):e1006291

Maheswaranathan N, McIntosh LT, Kastner DB, Melander J, Brezovec L, et al. 2018b. Deep learning models reveal internal structure and diverse computations in the retina under natural scenes. bioRxiv 340943

Mante V, Bonin V, Carandini M. 2008. Functional mechanisms shaping lateral geniculate responses to artificial and natural stimuli. *Neuron* 58(4):625–38

Marblestone AH, Wayne G, Körding KP. 2016. Toward an integration of deep learning and neuroscience. *Front. Comput. Neurosci.* 10:94

Markram H, Muller E, Ramaswamy S, Reimann MW, Abdellah M, et al. 2015. Reconstruction and simulation of neocortical microcircuitry. *Cell* 163(2):456–92

Marmarelis VZ. 2004. *Nonlinear Dynamic Modeling of Physiological Systems*. Hoboken, NJ: Wiley

Marr D, Poggio TA. 1976. *From understanding computation to understanding neural circuitry*. AI Memo 357, Artif. Intell. Lab. Mass. Inst. Technol., Cambridge, MA

McFarland JM, Bondy AG, Cumming BG, Butts DA. 2014. High-resolution eye tracking using V1 neuron activity. *Nat. Commun.* 5:4605

McFarland JM, Cui Y, Butts DA. 2013. Inferring nonlinear neuronal computation based on physiologically plausible inputs. *PLOS Comput. Biol.* 9(7):e1003143

McFarland JM, Cumming BG, Butts DA. 2016. Variability and correlations in primary visual cortical neurons driven by fixational eye movements. *J. Neurosci.* 36(23):6225–41

McMahon DBT, Jones AP, Bondar IV, Leopold DA. 2014. Face-selective neurons maintain consistent visual responses across months. *PNAS* 111(22):8251–56

Mineault PJ, Khawaja FA, Butts DA, Pack CC. 2012. Hierarchical processing of complex motion along the primate dorsal visual pathway. *PNAS* 109(16):E972–80

Mineault PJ, Zanos TP, Pack CC. 2014. *Converging encoding strategies in dorsal and ventral visual streams*. Prog. 236.19, Neurosci. Meet. Plan., Soc. Neurosci. Online, Washington, DC

Mohanty D, Scholl B, Priebe NJ. 2012. The accuracy of membrane potential reconstruction based on spiking receptive fields. *J. Neurophysiol.* 107(8):2143–53

Moskovitz TH, Roy NA, Pillow JW. 2018. A comparison of deep learning and linear-nonlinear cascade approaches to neural encoding. bioRxiv 463422

Movshon JA, Newsome WT. 1996. Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. *J. Neurosci.* 16(23):7733–41

Movshon JA, Thompson ID, Tolhurst DJ. 1978a. Receptive field organization of complex cells in the cat's striate cortex. *J. Physiol.* 283:79–99

Movshon JA, Thompson ID, Tolhurst DJ. 1978b. Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *J. Physiol.* 283:53–77

Musall S, Kaufman MT, Gluf S, Churchland A. 2018. Movement-related activity dominates cortex during sensory-guided decision making. bioRxiv 308288

Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL. 2009. Bayesian reconstruction of natural images from human brain activity. *Neuron* 63(6):902–15

Ni AM, Ruff DA, Alberts JJ, Symmonds J, Cohen MR. 2018. Learning and attention reveal a general relationship between population activity and behavior. *Science* 359(6374):463–65

Nienborg H, Cohen MR, Cumming BG. 2012. Decision-related activity in sensory neurons: correlations among neurons and with behavior. *Annu. Rev. Neurosci.* 35:463–83

Nishimoto S, Gallant JL. 2011. A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. *J. Neurosci.* 31(41):14551–64

Okun M, Steinmetz NA, Cossell L, Iacaruso MF, Ko H, et al. 2015. Diverse coupling of neurons to populations in sensory cortex. *Nature* 521(7553):511–15

Oliver MD, Gallant JL. 2013. *High-order Volterra models of area V4 capture complex selectivity*. Prog. 406.02, Neurosci. Meet. Plan., Soc. Neurosci. Online, Washington, DC

Oliver MD, Gallant JL. 2017. *A deep convolutional energy model of ventral stream areas V1, V2 and V4*. Prog. 312.06, Neurosci. Meet. Plan., Soc. Neurosci. Online, Washington, DC

Olsen SR, Wilson RI. 2008. Lateral presynaptic inhibition mediates gain control in an olfactory circuit. *Nature* 452(7190):956–60

Olshausen BA, Field DJ. 2005. How close are we to understanding V1? *Neural Comput*. 17(8):1665–99

Optican LM, Richmond BJ. 1987. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. *J. Neurophysiol.* 57(1):162–78

Orban GA. 2008. Higher order visual processing in macaque extrastriate cortex. *Physiol. Rev.* 88(1):59–89

Ozuysal Y, Baccus SA. 2012. Linking the computational structure of variance adaptation to biophysical mechanisms. *Neuron* 73(5):1002–15

Pandarinath C, O'Shea DJ, Collins J, Jozefowicz R, Stavisky SD, et al. 2018. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nat. Methods* 15(10):805–15

Paninski L. 2004. Maximum likelihood estimation of cascade point-process neural encoding models. *Network* 15(4):243–62

Paninski L, Cunningham JP. 2018. Neural data science: accelerating the experiment-analysis-theory cycle in large-scale neuroscience. *Curr. Opin. Neurobiol.* 50:232–41

Paninski L, Pillow JW, Lewi J. 2007. Statistical models for neural encoding, decoding, and optimal stimulus design. *Prog. Brain Res.* 165:493–507

Paninski L, Pillow JW, Simoncelli EP. 2004. Maximum likelihood estimation of a stochastic integrate-and-fire neural encoding model. *Neural Comput*. 16(12):2533–61

Park IM, Archer EW, Priebe NJ, Pillow JW. 2013. Spectral methods for neural characterization using generalized quadratic models. *Adv. Neural Inf. Process.* 26:2454–62

Park IM, Pillow JW. 2011. Bayesian spike-triggered covariance analysis. *Adv. Neural Inf. Process.* 24:1692–700

Park M, Horwitz G, Pillow JW. 2011. Active learning of neural response functions with Gaussian processes. *Adv. Neural Inf. Process.* 26:2043–51

Park M, Pillow JW. 2013. Bayesian inference for low rank spatiotemporal neural receptive fields. *Adv. Neural Inf. Process.* 26:2688–96

Pascanu R, Gulcehre C, Cho K, Bengio Y. 2013. How to construct deep recurrent neural networks. arXiv:1312.6026 [cs.NE]

Paszke A, Gross S, Chintala S, Chanan G, Yang E, et al. 2017. *Automatic differentiation in PyTorch*. Paper presented at the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA

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(47):11003–13

Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, et al. 2008. Spatio-temporal correlations and visual signaling in a complete neuronal population. *Nature* 454(7207):995–99

Pillow JW, Simoncelli EP. 2006. Dimensionality reduction in neural models: an information-theoretic generalization of spike-triggered average and covariance analysis. *J. Vis.* 6(4):414–28

Poirazi P, Brannon T, Mel BW. 2003. Pyramidal neuron as two-layer neural network. *Neuron* 37(6):989–99

Ponce CR, Xiao W, Schade PF, Hartmann TS, Kreiman G, Livingstone MS. 2019. Evolving images for visual neurons using a deep generative network reveals coding principles and neuronal preferences. *Cell* 177(4):999–1009.e10

Prenger R, Wu MCK, David SV, Gallant JL. 2004. Nonlinear V1 responses to natural scenes revealed by neural network analysis. *Neural Netw*. 17(5–6):663–79

Rabinowitz NC, Goris RLT, Cohen MR, Simoncelli EP. 2015. Attention stabilizes the shared gain of V4 populations. *eLife* 4:e08998

Rajan K, Marre O, Tkačik G. 2013. Learning quadratic receptive fields from neural responses to natural stimuli. *Neural Comput*. 25(7):1661–92

Rasch MJ, Gretton A, Murayama Y, Maass W, Logothetis NK. 2008. Inferring spike trains from local field potentials. *J. Neurophysiol.* 99(3):1461–76

Reid RC, Victor JD, Shapley RM. 1997. The use of m-sequences in the analysis of visual neurons: linear receptive field properties. *Vis. Neurosci.* 14(6):1015–27

Robinson DA. 1992. Implications of neural networks for how we think about brain function. *Brain Behav. Sci.* 15(4):644–55

Ruder S. 2016. An overview of gradient descent optimization algorithms. arXiv:1609.04747 [cs.CV]

Russ BE, Leopold DA. 2015. Functional MRI mapping of dynamic visual features during natural viewing in the macaque. *NeuroImage* 109:84–94

Rust NC, Mante V, Simoncelli EP, Movshon JA. 2006. How MT cells analyze the motion of visual patterns. *Nat. Neurosci.* 9(11):1421–31

Rust NC, Movshon JA. 2005. In praise of artifice. *Nat. Neurosci.* 8(12):1647–50

Rust NC, Schwartz O, Movshon JA, Simoncelli EP. 2005. Spatiotemporal elements of macaque V1 receptive fields. *Neuron* 46(6):945–56

Sahani M, Linden JF. 2003. Evidence optimization techniques for estimating stimulus-response functions. *Adv. Neural Inf. Process. Syst.* 15:317–24

Schmidhuber J. 2015. Deep learning in neural networks: an overview. *Neural Netw*. 61:85–117

Schwartz GW, Okawa H, Dunn FA, Morgan JL, Kerschensteiner D, et al. 2012. The spatial structure of a nonlinear receptive field. *Nat. Neurosci.* 15(11):1572–80

Schwartz O, Pillow JW, Rust NC, Simoncelli EP. 2006. Spike-triggered neural characterization. *J. Vis.* 6(4):484–507

Serre T. 2015. Hierarchical models of the visual system. In *Encyclopedia of Computational Neuroscience*, ed. D Jaeger, R Jung, pp. 1309–18. Berlin: Springer

Serre T. 2019. Deep learning: the good, the bad, and the ugly. *Annu. Rev. Vis. Sci.* 5:399–426

Shapley RM. 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. *Vis. Res.* 49(9):907–21

Shapley RM, Victor JD. 1978. The effect of contrast on the transfer properties of cat retinal ganglion cells. *J. Physiol.* 285:275–98

Sharpee TO. 2013. Computational identification of receptive fields. *Annu. Rev. Neurosci.* 36:103–20

Sharpee TO, Rust NC, Bialek W. 2004. Analyzing neural responses to natural signals: maximally informative dimensions. *Neural Comput*. 16(2):223–50

Shi Q, Gupta P, Boukhvalova A, Singer JH, Butts DA. 2019. Functional characterization of retinal ganglion cells using tailored nonlinear modeling. *Sci. Rep.* 9:8713

Shou T-D. 2010. The functional roles of feedback projections in the visual system. *Neurosci. Bull.* 26(5):401–10

Simoncelli EP, Pillow JW, Paninski L, Schwartz O. 2004. Characterization of neural responses with stochastic stimuli. In *The New Cognitive Neurosciences*, ed. M Gazzaniga, pp. 327–38. Cambridge, MA: MIT Press

Simonyan K, Zisserman A. 2014. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 [cs.CV]

Sinz FH, Ecker AS, Fahey PG, Erick C, Froudarakis E, et al. 2018. Stimulus domain transfer in recurrent models for large scale cortical population prediction on video. *Adv. Neural Inf. Process. Syst.* 31:7199–210

Stevenson IH, Körding KP. 2011. How advances in neural recording affect data analysis. *Nat. Neurosci.* 14(2):139–42

Stringer C, Pachitariu M, Steinmetz N, Reddy CB, Carandini M, Harris KD. 2019. Spontaneous behaviors drive multidimensional, brainwide activity. *Science* 364(6437):eaav7893

Talebi V, Baker CL. 2012. Natural versus synthetic stimuli for estimating receptive field models: a comparison of predictive robustness. *J. Neurosci.* 32(5):1560–76

Touryan J, Felsen G, Dan Y. 2005. Spatial structure of complex cell receptive fields measured with natural images. *Neuron* 45(5):781–91

Touryan J, Lau B, Dan Y. 2002. Isolation of relevant visual features from random stimuli for cortical complex cells. *J. Neurosci.* 22(24):10811–18

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(2):1074–89

Turner MH, Rieke F. 2016. Synaptic rectification controls nonlinear spatial integration of natural visual inputs. *Neuron* 90(6):1257–71

Uzzell VJ, Chichilnisky EJ. 2004. Precision of spike trains in primate retinal ganglion cells. *J. Neurophysiol.* 92(2):780–89

VanRullen R, Thorpe SJ. 2001. Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. *Neural Comput.* 13(6):1255–83

Victor JD, Shapley RM. 1979. The nonlinear pathway of Y ganglion cells in the cat retina. *J. Gen. Physiol.* 74(6):671–89

Vinck M, Batista-Brito R, Knoblich U, Cardin JA. 2015. Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. *Neuron* 86(3):740–54

Weber AI, Pillow JW. 2017. Capturing the dynamical repertoire of single neurons with generalized linear models. *Neural Comput.* 29(12):3260–89

Whiteway MR, Butts DA. 2019. The quest for interpretable models of neural population activity. *Curr. Opin. Neurobiol.* 58:86–93

Williamson RS, Ahrens MB, Linden JF, Sahani M. 2016. Input-specific gain modulation by local sensory context shapes cortical and thalamic responses to complex sounds. *Neuron* 91(2):467–81

Williamson RS, Sahani M, Pillow JW. 2013. Equating information-theoretic and likelihood-based methods for neural dimensionality reduction. arXiv:1308.3542 [q-bio.NC]

Williamson RS, Sahani M, Pillow JW. 2015. The equivalence of information-theoretic and likelihood-based methods for neural dimensionality reduction. *PLOS Comput. Biol.* 11(4):e1004141

Wilson DE, Whitney DE, Scholl B, Fitzpatrick D. 2016. Orientation selectivity and the functional clustering of synaptic inputs in primary visual cortex. *Nat. Neurosci.* 19(8):1003–9

Wu MCK, David SV, Gallant JL. 2006. Complete functional characterization of sensory neurons by system identification. *Annu. Rev. Neurosci.* 29:477–505

Yamins DLK, DiCarlo JJ. 2016. Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* 19(3):356–65

Yamins DLK, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ. 2014. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *PNAS* 111(23):8619–24

Yates JL, Park IM, Katz LN, Pillow JW, Huk AC. 2017. Functional dissection of signal and noise in MT and LIP during decision-making. *Nat. Neurosci.* 20(9):1285–92

Zhang Y, Lee TS, Li M, Liu F, Tang S. 2019. Convolutional neural network models of V1 responses to complex patterns. *J. Comput. Neurosci.* 46(1):33–54