

A Simple Model for Low Variability in Neural Spike Trains

Ulisse Ferrari

ulisse.ferrari@gmail.com

*Sorbonne Université, INSERM, CNRS, Institut de la Vision, 17 rue Moreau,
F-75012 Paris, France*

Stéphane Deny

stephane.deny.pro@gmail.com

*Neural Dynamics and Computation Lab, Stanford University, Stanford,
CA 94305, U.S.A.*

Olivier Marre

olivier.marre@gmail.com

*Sorbonne Université, INSERM, CNRS, Institut de la Vision, 17 rue Moreau,
F-75012 Paris, France*

Thierry Mora

thierry.mora@gmail.com

*Laboratoire de physique statistique, CNRS, Sorbonne Université, Université
Paris-Diderot, and École Normale Supérieure (PSL University), 75005 Paris, France*

Neural noise sets a limit to information transmission in sensory systems. In several areas, the spiking response (to a repeated stimulus) has shown a higher degree of regularity than predicted by a Poisson process. However, a simple model to explain this low variability is still lacking. Here we introduce a new model, with a correction to Poisson statistics, that can accurately predict the regularity of neural spike trains in response to a repeated stimulus. The model has only two parameters but can reproduce the observed variability in retinal recordings in various conditions. We show analytically why this approximation can work. In a model of the spike-emitting process where a refractory period is assumed, we derive that our simple correction can well approximate the spike train statistics over a broad range of firing rates. Our model can be easily plugged to stimulus processing models, like a linear-nonlinear model or its generalizations, to replace the Poisson spike train hypothesis that is commonly assumed. It estimates the amount of information transmitted much more accurately than Poisson models in retinal recordings. Thanks to its simplicity, this model has the potential to explain low variability in other areas.

O.M. and T.M. contributed equally to this work.

1 Introduction

Neural variability imposes constraints on the way neurons transmit and process information (Movshon, 2000) and has been extensively studied in the mammalian visual system (Barlow & Levick, 1969; Heggelund & Albus, 1978; Tolhurst, Movshon, & Thompson, 1981; Kara, Reinagel, & Reid, 2000) and beyond (de Ruyter van Steveninck, Lewen, Strong, Koberle, & Bialek, 1997; Baddeley et al., 1997; DeWeese & Zador, 2003; Churchland et al., 2010). The Fano factor (the ratio of the spike count variance to the mean) is often used to characterize such variability. Many studies have reported various values for the Fano factor depending on brain area (Kara et al., 2000) or experimental condition (Churchland et al., 2010). A Fano factor lower than one suggests that neurons spike regularly, a condition sometimes termed underdispersion (Gur, Beylin, & Snodderly, 1997; Kara et al., 2000; Barberini, Horwitz, & Newsome, 2001; DeWeese & Zador, 2003; Maimon & Assad, 2009). A Fano factor above one is termed overdispersion (Baddeley et al., 1997; Churchland et al., 2010).

A common model for a neural spike count is to assume Poisson statistics, where the variance in the number of spikes emitted is equal to its mean and the Fano factor is thus one. Like the classical linear nonlinear poisson (LNP; Chichilnisky, 2001), many models describing how stimuli are processed by individual neurons rely on this assumption. New models are therefore needed to account for the deviations from Poisson statistics observed in the data. Several models have been proposed to account for overdispersion in the spike count distribution (Goris, Movshon, & Simoncelli, 2014; Scott & Pillow, 2012; Charles et al., 2017). For underdispersed spike count distribution, a few models have been proposed, but they come with specific constraints. One possibility is to use a spike history filter that allows past spikes to inhibit the probability of emitting a new spike at present (Pillow et al., 2008). However, this approach requires defining the probability of spiking over very small time bins (e.g., 1 ms), and consequently needs several parameters or strong regularization to describe the spike history filter. Other models have been proposed, but some have many parameters (e.g., approximately seven for each cell in Gao, Busing, Shenoy, & Cunningham, 2015), and may therefore require very large data sets to learn the model or make specific assumptions that may not always be verified in the data (Sellers, Borle, & Shmueli, 2012; Stevenson, 2016). Overall, it is unclear if a general yet simple model can account for the spike count distribution found in the data (Charles et al., 2017).

Here we present a simple model that can account for underdispersion in the spike count variability, with only two parameters. Our starting assumption is that the deviation from Poisson statistics comes from the refractory period—the fact that there is a minimal time interval between two spikes. We start from the analytic form of the spike count probability taken by a spike generation process composed of an absolute refractory period

followed by a Poisson process, and show that it can be approximated by a model with only one parameter to fit to the data. We further simplify the model to make it amenable to log-likelihood maximization. We then relax our assumption by allowing for a relative refractory period (Berry & Meister, 1998) and derive a more flexible model with two parameters that can accurately predict neural variability. The simple form of this model makes it possible to plug it on classical stimulus processing models (e.g., linear-nonlinear (LN) model; Pillow et al., 2008) or more complex cascade models (Deny et al., 2017), that can be fitted with log-likelihood maximization. We test our model on retinal data and find that it outperforms the Poisson model, as well as other statistical models proposed in the literature to account for underdispersion. Either our model performs better at describing the data or other models need more parameters for an equally accurate description. When combined with classical stimulus processing models, our model is able to predict the variance of the spike count over time, as well as the amount of information conveyed by single neurons, much better than classical models relying on Poisson processes. We thus propose a simple model for neural variability, with only two parameters, that can be used in combination with any stimulus processing model to account for sub-Poisson neural variability.

2 Sub-Poisson Behavior of Retinal Ganglion Cells

We used previously published data (Deny et al., 2017) where we recorded ganglion cells from the rat retina using multielectrode arrays (Marre et al., 2012; Yger et al., 2016). Two types of ganglion cells, ON and OFF, with receptive fields tiling the visual space, were isolated in these experiments ($n = 25$ and $n = 19$ cells, respectively). Cells were then stimulated with the video of two parallel horizontal bars performing a Brownian motion along the vertical direction (Deny et al., 2017). Some stimuli sequences were repeated and triggered reliable responses from ganglion cells (see Figures 1A and 1B, top panels).

We then binned the retinal ganglion cell (RGC) responses with 16.7 ms time windows ΔT locked to the frame update of the video stimulation (60 Hz). For each cell i in each time bin t and for each stimulus repetition rep , we associated an integer spike count $n_i^{rep}(t)$ equal to the number of emitted spikes in that time window. In order to analyze the cell reliability as a function of the firing rate, for each cell and time bin, we computed the mean ($\lambda_i(t)$) and variance ($V_i(t)$) of this spike count across repetitions of the same stimulus. In Figures 1A and 1B, (bottom panels), we show the firing rate of two example ON and OFF cells, together with its fluctuation across the repetitions (colored areas). For comparison, we also show the fluctuations that can be expected from a Poisson process, that is, with a variance equal to its mean. During transients of high activity, cells are more reliable than what can be expected.

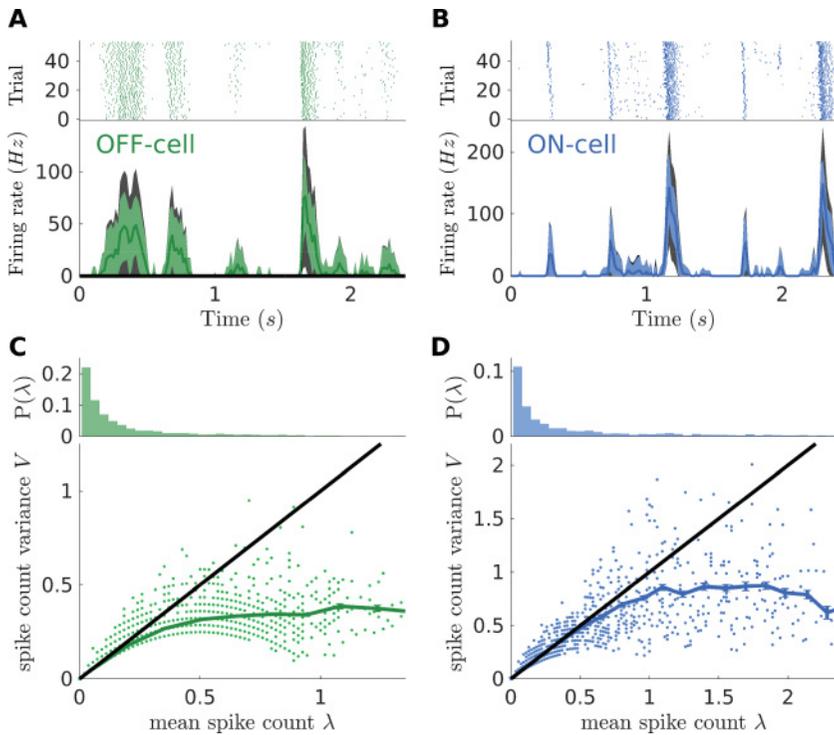


Figure 1: Sub-Poisson behavior of RGCs. (A) Top: Example raster plot for one OFF cell. Bottom: Firing rate behavior for the same cell. Colored (resp. black) area delimits the empirical (resp. Poisson) noise, as mean ± 1 SD. (B) Same as panel A but for an example ON cell. (C) Top: Histogram of the observed means of the spike count across stimulus repetitions (time bin of 16.7 ms) pooled across all cells of the same type. Time bins with zero mean have been excluded. Bottom: Mean of spike count across stimulus repetitions plotted against its variance. Each point corresponds to one cell in one time bin. Multiple points are superimposed. Green solid line: Average of points with similar mean. Black line: Prediction from a Poisson distributed spike count. (D) Same as in panel C but for ON cells. Note the increase of activity with respect to OFF cells.

Figures 1C and 1D show that both ON and OFF cells have a sub-Poisson variability. Variances of spike counts are much lower than their means, and their ratio (the Fano factor, equal to 1 for Poisson distributions) is not only smaller than one but decreases with the mean. Note that here and for the rest of the letter, we pool together all the cells of the same type when estimating the parameters of the spike generators. We found that spike statistics were remarkably homogeneous across cells of the same type, and this gives us

more data to fit different models. However, the same process could have been applied on a single cell, with enough repetitions of the same stimulus.

These results are consistent with previous findings (Berry & Meister, 1998; Kara et al., 2000): ganglion cells emit spikes more reliably than a Poisson process. Consequently, a model for predicting the RGC response that accounts for noise with a Poisson generator—for example, the LNP model (Chichilnisky, 2001)—largely underestimates this spiking precision. We then search for a simple model to account for this relation between signal and noise.

3 Spike Count Statistics for a Refractory Neuron

Absolute or relative refractory periods are known to impose some regularity on the sequence of emitted spikes (Berry & Meister, 1998; Kara et al., 2000) and thus to decrease the neural variability. We aim at looking how refractoriness affect the spike count distribution of the spike train. We consider a model of refractory neurons where the instantaneous rate is inhibited by an absolute refractory period of duration τ . Such a neuron has the following interspike interval (ISI) distribution:

$$\rho_{\Theta}(t) = \Theta(t - \tau) r e^{-r(t-\tau)}, \quad (3.1)$$

where $\Theta(u)$ is the Heaviside unit step ($\Theta(u) = 1$ for $u > 0$ and zero otherwise) and r is the firing rate in absence of refractoriness. From the ISI distribution 3.1, we can estimate the probability distribution of the spike count n , that is, the probability distribution of the number of spikes emitted in a time bin Δt . For the particular case of $\tau = 0$, $\lambda \equiv \langle n \rangle = r \Delta t$, and n follows a Poisson distribution:

$$P_{\text{Pois}}(n|\lambda) = \frac{\lambda^n}{n!} e^{-\lambda}. \quad (3.2)$$

For $\tau > 0$, n follows a sub-Poisson distribution $P_{\Theta}(n|r, \Delta t, \tau)$, which admits an analytic (albeit complex) expression (Müller, 1974; see equation 10.1) in section 10, Methods, and the derivation in the supplementary information).

To test if this expression accounts well for the data, we need to adjust the value of the refractory period τ for both cell types. Because the complex expression of $P_{\Theta}(n|r, \Delta t, \tau)$ makes it hard to apply likelihood maximization, we perform the inference of τ by minimizing the mean square error (MSE) of the mean-variance relation, summed over all the cells and time bins of the training set (see section 10, Methods). We estimate τ for the two populations separately and found $\tau = 8.8$ ms for OFF cells and $\tau = 3.1$ ms for ON cells. In Figure 2 we compare the model prediction, with the mean-variance relation measured on the testing set. The good agreement of the predictions suggests that this simple model of refractory neurons accounts well for the

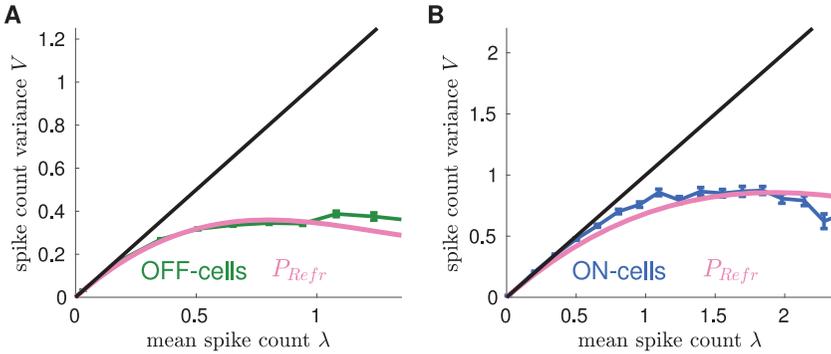


Figure 2: A simple neuron model with an absolute refractory period accounts for the observed spike count statistics. (A) Relation between mean and variance predicted by the model (pink) and measured for the OFF population (green; same data as Figure 1). (B) Same as panel A, but for the ON population.

neural variability, even when the firing rate is large. A simple model that takes into account the deviation from Poisson statistics due to the refractory period is thus able to predict the mean-variance relation observed in the data.

4 Simple Models for the Spike Count Statistics

We have shown how a rather simple model of refractory neurons accurately reproduces the mean-variance relation of the recorded cells. However, the model has a complicated analytic form, which is not easily amenable to a likelihood-maximization approach, and it cannot be considered an alternative to models based on Poisson generators such as the LNP model. To overcome this limitation, we propose a further simplification of the refractory neuron that is much more efficient and tractable.

For $\tau = 0$, the refractory distribution $P_{\Theta}(n | r, \Delta t, \tau)$ reduces to the Poisson distribution (see equation 3.2). The small values of the refractory periods relative to the bin size suggest that an expansion at small τ could capture most of the model's behavior. We expand $P_{\Theta}(n | r, \Delta t, \tau)$ around the Poisson distribution to the second order in the small parameter $f = \tau/\Delta T$ (see section A.3 for details):

$$P_{\Theta}^{(2\text{nd})}(n | \lambda, f) = \exp \left\{ \theta_{\lambda} n - (f - f^2)n^2 - \frac{f^2}{2}n^3 - \log n! - \log Z_{\lambda} \right\} \quad (4.1)$$

$$\theta_{\lambda} \quad \text{chosen such that} \quad \langle n \rangle_{P_{\Theta}^{(2\text{nd})}(n | \lambda, f)} = \lambda, \quad (4.2)$$

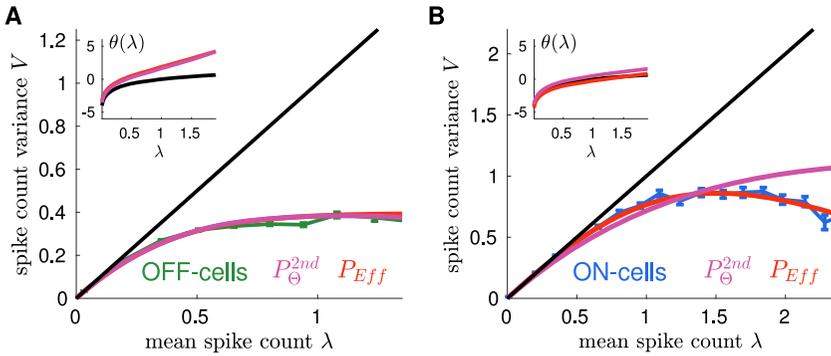


Figure 3: Second-Order and Effective models account well for observed spike count statistics. (A) The empirical behavior for the OFF population estimated from the repeated two bars stimulus (Green; same data as Figure 1) is compared with the prediction of Second-Order (purple) and effective (red) models. Lines superimpose as the two models take a very similar form for the OFF population. For comparison, Poisson prediction is shown in black (equality line) Inset: The function θ_{λ} of the two models (see the text), is compared with $\theta_{\lambda} = \log \lambda$ for the Poisson model (black line) (B) Same as panel A, but for the ON population. In all cases, models are learned on a separate training set.

where Z_{λ} is a normalization constant. The “second-order” model, equation 4.1, has only one free parameter, f or, equivalently, τ . θ_{λ} is a function of λ and has to be estimated numerically in order to reproduce the mean spike count λ once the refractory period τ has been fixed. For instance, for $\tau = 0$, $\theta_{\lambda} = \log \lambda$ and the second-order reduces to the Poisson distribution. Importantly, the coefficients of the n^2 and n^3 terms do not depend on the firing rate, r , insofar as the refractory period does not either. The exponential form of the model makes it easy to calculate its derivatives and to use with maximum likelihood methods.

We infer τ through a log-likelihood maximization (see Methods, section 10) using equation 4.1 for OFF and ON populations separately. From these values we can estimate the refractory period for the second-order model: $\tau = 10.8$ ms for OFF cells and $\tau = 3.0$ ms for ON cells. In Figure 3 we compare the model predictions for spike count variance V with the empirical values. Unlike in the previous section, model parameters were not optimized to best fit this curve, since it was fitted using log-likelihood maximization. Yet the second-order model shows a high performance for the OFF population, even at large firing rates. Despite the approximation, the model is still able to describe accurately the mean-variance relation and can now be fitted using maximum likelihood approaches. For ON cells, the model outperforms the Poisson model, but the mean-variance relation is not perfectly predicted. We will explore why this could be the case and

improve our model to arrive at a more flexible one that can be suited for a broader range of experimental cases.

5 A Simple, Effective Model to Describe Spike Train Statistics

The previous results hold for a very specific assumption about how refractoriness constrains firing, with just a single fitting parameter: the refractory period. We thus wondered whether more general rules of refractoriness could give rise to a broader class of sub-Poisson spiking distributions, allowing for better agreement with the data.

We considered a general model in which the instantaneous spike rate is now inhibited by a time-dependent factor, $\alpha(r, u)r$, where r is the spike rate in the absence of refractoriness and u is the time following the previous spike, with $\alpha(r, 0) = 0$ and $\alpha(r, u \geq \tau) = 1$. For instance, a pure refractory period is defined by $\alpha(r, u) = 0$ for $u < \tau$ and $\alpha(r, u) = 1$ for $u > \tau$. A soft refractory period, where the spike rate recovers linearly in time, would be given by $\alpha(r, u) = u/\tau$ for $u < \tau$ and $\alpha(r, u) = 1$ for $u > \tau$. Calculating the spike count distribution under this assumption is intractable analytically, but a second-order expansion such as the one performed in the previous section can still be performed, and yields the following expression at second-order (see section A.3):

$$P_{\alpha}^{(2\text{nd})}(n | \lambda, r, f, \Delta t) = \exp \{ \theta_{\lambda} n - \gamma_{\alpha} n^2 - \delta_{\alpha} n^3 - \log n! - \log Z_{\lambda} \} \quad (5.1)$$

$$\gamma_{\alpha} = f - f^2 + (2 + r\Delta t) \left(\frac{f^2}{2} - g \right)$$

$$\delta_{\alpha} = f^2 - g = \frac{f^2}{2} + \left(\frac{f^2}{2} - g \right)$$

$$\theta_{\lambda} \quad \text{chosen such that} \quad \langle n \rangle_{P_{\alpha}^{(2\text{nd})}(n | \lambda, r, f, \Delta t)} = \lambda, \quad (5.2)$$

where $f \equiv \int_0^{\infty} du (1 - \alpha(r, u))/\Delta t$ and $g \equiv \int_0^{\infty} du u (1 - \alpha(r, u))/\Delta t^2$. The special case of a pure refractory period, $\alpha(r, u) = 0$ for $0 \leq u \leq \tau$, gives back equation 4.3 and the previous definition of f . In this case, $f^2/2 - g$ vanishes. This quantity can thus be considered as an estimate of the deviation from the absolute refractoriness. Again, θ_{λ} should be adjusted to match the average number of spikes in each cell and time bin, $\lambda_i(t)$.

This analytic development shows that the coefficients of n^2 and n^3 , γ_{α} and δ_{α} , can have very different forms depending on the exact form of the refractoriness. We thus decided to relax the assumption of a strict dependence between γ_{α} and δ_{α} . We tested if a model with γ_{α} and δ_{α} that does not depend on r shows good agreement with the data. In the following, we thus

treat the coefficients of n^2 and n^3 in equation 5.1 as two parameters that are independent of each other and also independent of the firing rate.

The resulting Effective model is defined for any mean spike count λ as

$$P_{\text{Eff}}(n | \lambda, \gamma, \delta) = \exp \{ \theta_\lambda n - \gamma n^2 - \delta n^3 - \log n! - \log Z_\lambda \} \quad (5.3)$$

$$\theta_\lambda \quad \text{chosen such that} \quad \langle n \rangle_{P_{\text{Eff}}(n | \lambda, \gamma, \delta)} = \lambda, \quad (5.4)$$

where, as before, θ_λ is not a free parameter but a uniquely defined function of λ with parameters γ and δ . While λ is set to its time-dependent value λ_t , γ and δ are assumed to be constant parameters of the model. The probability distribution, equation 5.3, belongs to the class of weighted Poisson distributions, and its mathematical properties have been already studied elsewhere (del Castillo & Pérez-Casany, 2005).

We infer γ and δ through a log-likelihood maximization on the OFF and ON cell separately. For the OFF population we obtain similar values as the Second-Order model (see Figure 3A). By contrast, for the ON population, the Effective model takes advantage of the additional free parameter and uses it to improve its performance. We obtained the values $\gamma_{\text{ON}}^* = -0.52$ and $\delta_{\text{ON}}^* = 0.15$, while the equivalent parameter values in the Second-Order model are $f_{\text{ON}}^* - f_{\text{ON}}^{*2} = 0.15$ and $f_{\text{ON}}^{*2}/2 = 0.02$.

This Effective model is therefore a simple model able to describe accurately, with only two parameters, the mean-variance relation for different types of refractoriness.

6 Benchmark of Proposed Models

The Effective model outperforms a Poisson model at predicting the empirical spike count variance. In this section, we compare its performance with two other spike count models proposed in the literature. The Generalized Count (Gen.Count) model (Gao et al., 2015) can be seen as a generalization of our Effective model (with arbitrary dependence on n in the correction to Poisson, not limited to a linear combination of n^2 and n^3 ; see equation 10.13 in Methods, section 10) and thus offers larger flexibility to model the spike count statistics. This, however, comes at the price of introducing more parameters to fit and could potentially lead to overfitting. The Conway-Maxwell-Poisson (COMP) model (Sellers et al., 2012 see equation 10.14) has been proposed to account for both an under- and overdispersed mean-variance relation (Stevenson, 2016). COMP is a one-parameter extension of Poisson, which differs from our Effective model but is still a particular case of Gen.Count (see section 10, Methods, for details). For completeness, we also compare the Refractory model, P_Θ of equation 10.1 and its second-order expansion, the Second-Order model, $P_\Theta^{2\text{nd}}$ of equation 4.1.

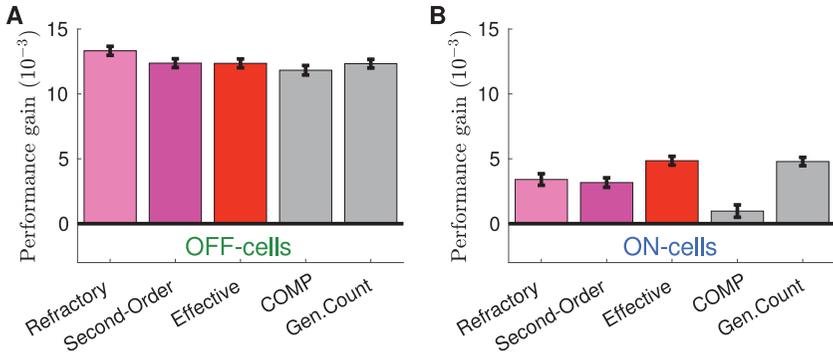


Figure 4: Second-Order (see equation 4.1) and Effective (see equation 5.3) model performance compared with known spike count models: COMP (see equation 10.14) and Generalized Count (see equation 10.13). Log-likelihood improvement over the Poisson for several models, all learned on a separate training set, for OFF cells (A) and ON cells (B).

We compare the performance of different models as an improvement over the Poisson log likelihood (see Figure 4). For the OFF population (see Figure 4A), all models outperform Poisson and have similar performance. This is probably because OFF cells show rather small firing rates and rarely emit more than two spikes in the same time bin. The addition of one parameter (with respect to Poisson) is thus enough for accurately modeling the spike count statistics. However, for the ON population (see Figure 4B), while all the considered models outperform Poisson, Effective and Gen.Count show the largest improvement. Remarkably, the one-parameter Refractory and Second-Order models show very high performance as well, despite the first being learned by fitting the mean-variance relation of the spike count rather than by maximizing the likelihood. Also, in this case, the larger flexibility of the Gen.Count model does not bring an improvement.

To test if these models can generalize to other stimulus statistics, after learning the parameters of the model on the responses to the moving bars stimuli, we test them on the responses of the same cells to a repeated sequence of checkerboard stimulus. We find that neurons also had sub-Poisson behavior in response to this stimulus, with spike count variances smaller than the corresponding means, very similar to Figures 1C and 1D. Figure 5 shows the log-likelihood improvement over the Poisson model. All models performed better than Poisson, except for the COMP model, which for the ON cells does not seem to generalize to other stimuli. In particular, for OFF cells we found that the Refractory model has the best performance, just a bit larger than the Second-Order and the Effective model (both have very similar performances; see the previous section). For ON cells, Effective and Gen.Count models show the best performance. Also in this case,

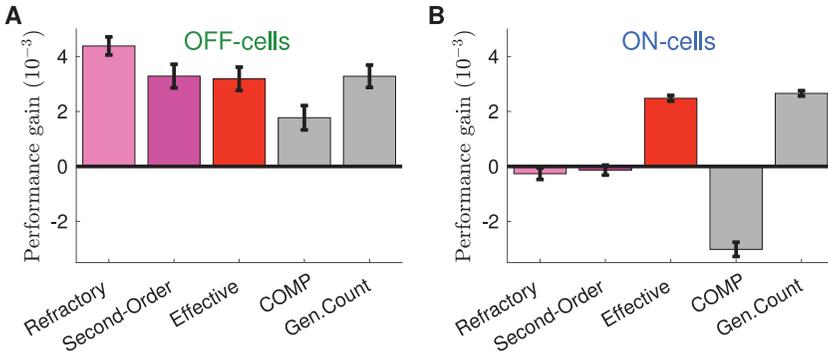


Figure 5: Performance of different models on response to checkerboard stimulation. Same as Figure 4, but the testing set was the response to a repeated checkerboard stimulation

the larger flexibility of Gen.Count does not bring a performance improvement. The Effective model is thus a simple model to describe deviation from Poisson statistics with only two parameters. It works as well as the most general model, Gen.Count, to describe the sub-Poisson variability but with fewer parameters. It also performs better than other models with a low number of parameters.

7 Impact of Noise Distribution on Information Transfer

The ability of neurons to transmit information is limited by their variability (Movshon, 2000). If neurons are less reliable or, equivalently, have larger variance, their capacity to transmit information should be significantly decreased. To properly estimate the amount of information transmitted by a neuron, a model should reproduce such variability. To test for this, we quantify the amount of stimulus information encoded by emitted spikes as the mutual information *MI* between the spike count *n* and its mean λ ,

$$MI(n, \lambda) = H[P(n)] - \langle H[P(n|\lambda)] \rangle_{\lambda}, \tag{7.1}$$

where *H* is the entropy function and *P*(*n*) is the distribution of the spike count without conditioning on the mean over repetitions. $\langle \dots \rangle_{\lambda}$ is the average over the observed mean spike count, and *P*(*n*|\lambda) is the distribution of *n* at fixed λ , estimated either empirically from the repeated data or with the model. To avoid undersampling issues typical of time bins with low activity and due to the finite number of stimulus repetitions, we restrict the averages to all cells and time bins with $\lambda_i(t) > 0.1$.

In Figure 6, for both OFF and ON populations, Poisson largely underestimates mutual information, whereas the Effective model predicts well

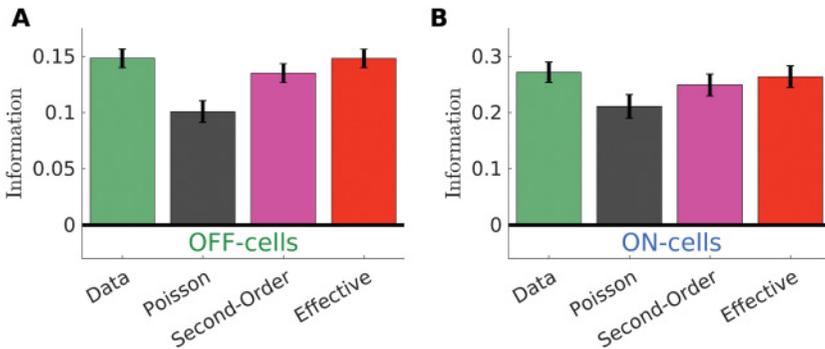


Figure 6: Second-Order and Effective model predicts empirical information transmission. Empirical estimation of the mutual information is compared with prediction for Poisson (black), Second-Order (pink), and Effective (red) models. Using Poisson as noise distribution for the spike count leads to a strong underestimation, biasing the prediction of information transfer of stimulus processing models. Error bars are standard deviation of the mean over λ (see equation 7.1).

the value of mutual information. Our model can thus be used to correctly estimate the mutual information thanks to its accurate prediction of the mean-variance relation.

Note that equation 7.1 is a mutual information for a single time bin and cannot be used to calculate information rates over spike trains because of its nonadditivity across time due to stimulus temporal correlations. (See Figure 8 for further details and explanations.)

8 Improving Stimulus Processing Models

The Effective model, equation 5.3, describes efficiently the relation between mean and variance and predicts well the amount of information transmitted by a neuron. Here we show how it can easily be plugged to any stimulus processing model like the Linear-Nonlinear (LN) model, instead of a Poisson process. To estimate the parameters of LN models, a classical approach is to assume a Poisson process for spike generation and then to maximize the likelihood of the spiking data with respect to the parameters of the model. The major advantage of this method is that the gradient of the likelihood has a very simple form that allows for iterative log-likelihood maximization. Here we show that using the Effective model also leads to a tractable form of the log-likelihood gradient, which can similarly be used for iterative optimization, but with the added advantage that the mean-variance relationship is accurately reproduced.

In general, a stimulus encoding model is defined by a series of computations—parameterized by parameters ψ —that takes the past

stimulus S_t as input and provides a prediction $\hat{\lambda}_\psi(S_t)$ for the spike count mean as a function of time t . Only at the last stage is a stochastic spike counter P introduced to predict the number of spikes $n(t)$ emitted in the time bin t :

$$S_t \rightarrow \hat{\lambda}_\psi(S_t) \rightarrow n(t) \sim P(n(t) | \hat{\lambda}_\psi(S_t)), \quad (8.1)$$

where, for example, P is a Poisson distribution. The classical example for this is the Linear-Nonlinear-Poisson (LNP) model (Chichilnisky, 2001), but many generalizations have been proposed, especially for the retina (McFarland, Cui, & Butts, 2013; McIntosh, Maheswaranathan, Nayebi, Ganguli, & Baccus, 2016; Deny et al., 2017).

One of the major advantages of using a Poisson spike counter P_{Pois} is that it allows for a straightforward optimization of the model parameters ψ (Chichilnisky, 2001; McFarland et al., 2013). Thanks to the explicit expression of the log-likelihood $\ell(\psi)$ (see section 10, Methods), the log-likelihood gradient $\nabla_\psi \ell(\psi) \equiv d\ell(\psi)/d\psi$ takes a very simple form:

$$\begin{aligned} \nabla_\psi \ell(\psi) &= \frac{1}{T} \sum_t \left. \frac{d \log P_{\text{Pois}}(n(t) | \lambda)}{d\lambda} \right|_{\lambda=\hat{\lambda}_\psi(S_t)} \nabla_\psi \hat{\lambda}_\psi(S_t) \\ &= \frac{1}{T} \sum_t \left(\frac{n(t) - \hat{\lambda}_\psi(S_t)}{\hat{\lambda}_\psi(S_t)} \right) \nabla_\psi \hat{\lambda}_\psi(S_t). \end{aligned} \quad (8.2)$$

Once $\nabla_\psi \hat{\lambda}_\psi(S_t)$ is evaluated, equation 8.2 allows for iterative log-likelihood maximization.

The particular structure of the Effective model, equation 5.3, allows for an easy estimation of the log-likelihood gradient (see section 10, Methods, for the calculation details):

$$\nabla_\psi \ell(\psi) = \frac{1}{T} \sum_t \left(\frac{n(t) - \hat{\lambda}_\psi(S_t)}{V_{\hat{\lambda}_\psi(S_t)}^{\text{Eff}}} \right) \nabla_\psi \hat{\lambda}_\psi(S_t). \quad (8.3)$$

With respect to the Poisson case, the inference of the parameters ψ requires only the variance V_λ^{Eff} as a function of λ , which depends only on the noise model and can be easily estimated from it before running the inference (see section 10, Methods, for more details).

As in Deny et al. (2017) and for both OFF and ON populations, we infer a Linear-Nonlinear-Linear-Nonlinear-Poisson (LN²-Pois) model to predict the average cell response to the two-bar stimulation. This is a stimulus processing model composed by a cascade of two layers of linear filtering and nonlinear-transformation followed by a Poisson spike counter. Similarly, we infer a two-layer cascade Effective model (LN²-Eff) to predict the

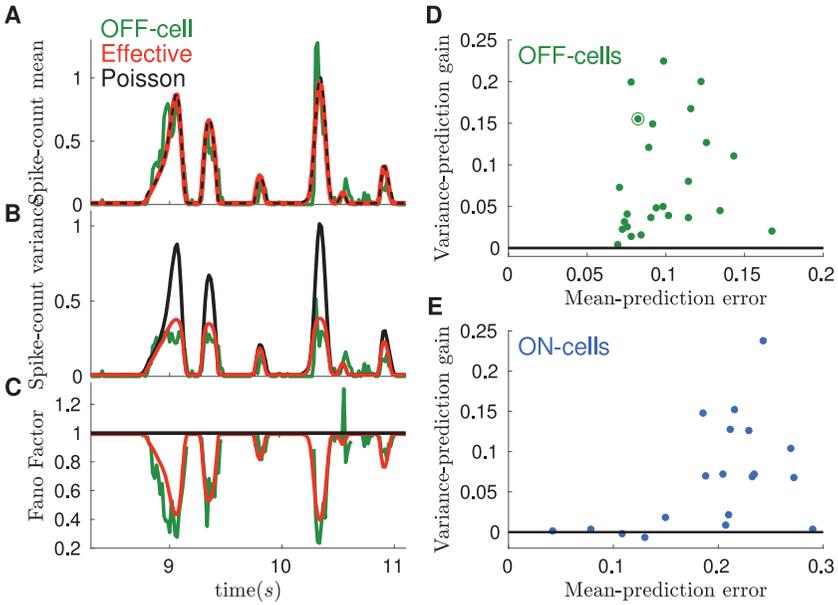


Figure 7: Effective model improves performance of stimulus processing models. (A) PSTH for an example OFF cell (green) is compared with predictions by the LN² models equipped with a Poisson (black) or Effective (red) spike counter. Note that the discrepancies in the spike rate predictions are those of the LN² model, not of the Poisson or Effective model, which inherit them by construction. (B) Same as panel A but for the spike count variance. The Poisson model largely overestimates the empirical variance during transients of high activity. (C) The Fano factor as a function of time. The Effective model (red) accounts for the empirical behavior (green). The Poisson prediction (black constant line) does not. (D) Improvement with respect to Poisson in the mean square error (MSE) between empirical and model variance, plotted as a function of the MSE between empirical and model mean spike count. Each point represents a cell of the OFF population. The circled point refers to the example cell in panels A to C. (E) Same as panel D but for the Effective model applied on the ON population.

average cell response of OFF and ON cells, respectively. LN²-Eff differs from LN²-Pois in the noise generator, either Effective (see equation 5.3) or either Poisson. The two models show very similar prediction for the mean spike activity (see Figure 7) while LN²-Pois largely overestimates the spike count variance (see Figure 7B). LN²-Eff also predicts well the Fano Factor estimated over time (see Figure 7C). For most of the OFF cells, replacing the Poisson by the Effective model leads to significant performance improvement when trying to predict the variance (see Figure 7D). For OFF cells, the Second-Order model, equation 4.1, performs as well as the

Effective model. The Effective model also predicts well the variance for ON cells (see Figure 7E). Our Effective model can therefore be plugged in encoding models to describe accurately the variance of the spike count over time.

9 Discussion

We have shown that a simple model that takes into account refractoriness in the spiking process explains most of the deviation from Poisson that we observe in the spike count statistics. The model has only two parameters and can easily be plugged into any encoding model to be fitted to sensory recordings. It allows for an accurate estimation of the relation between the mean spike count and its variance, but also of the amount of information carried by individual neurons. The form of this model is inspired by the regularity imposed on spike trains by the refractory period. However, it can potentially work for data with other sources of regularity. In the retina, this model works for two different types of cells. Thanks to its simplicity and generality, this model could potentially be used to account for mean-variance relation in neurons recorded in other sensory areas (Gur et al., 1997; Kara et al., 2000; Barberini et al., 2001; DeWeese & Zador, 2003; Maimon & Assad, 2009), even if their neural variability is not solely determined by the refractory period in the spike generation of the neuron. Previous work (Movshon, 2000; Kara et al., 2000) suggested that the refractory period present in the phenomenological model may reflect refractoriness at any stage in the circuit and may not directly correspond to the refractoriness of the recorded cell.

We have compared our model with others already proposed in the literature: the Conway-Maxwell-Poisson (COMP; Stevenson, 2016) and the Generalized Count (Gen.Count; Gao et al., 2015). We found the COMP model was quite inefficient at fitting our data. The Gen.Count model can be considered as encompassing a much larger class of possible models and includes ours as a special case. However, this comes at the cost of having many more parameters to fit, which can lead to overfitting in some cases, as we have shown in the results. Moreover, in our data, these additional parameters did not allow improving the performance of the model. The model we propose has only two parameters, which makes it easy to fit and usable in many cases, even when the number of data are limited. It will be interesting to compare our effective models to the Gen.Count model in other sensory structures.

The relation between the mean and the variance of the spike count strongly depends on the bin size chosen to bin the cell response (Kara et al., 2000). We illustrate this fact on our data by plotting the Fano factor as a function of window size in Figure 8A. One can distinguish three regimes as a function of the window size relative to the other timescales of the system (see Figure 8B): the inverse of the spike rate r^{-1} , the timescale of spike rate

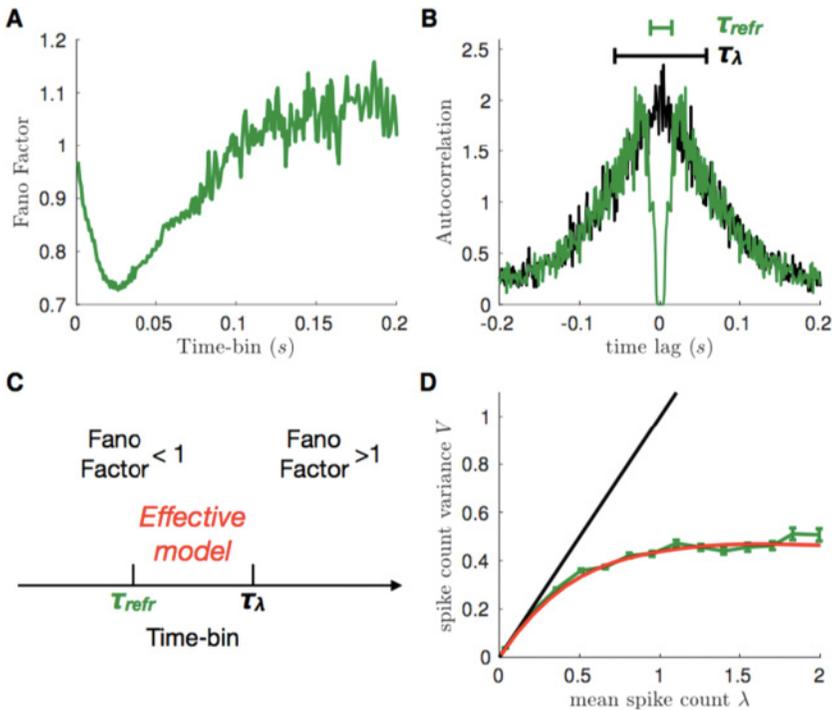


Figure 8: Range of applicability of the model. (A) The Fano factor as a function of the window size (time bin). For short windows, the Fano factor is smaller than one because of refractoriness, while for long windows, spike rate fluctuations yield Fano factors larger than one. (B) The autocorrelation function of the spiking activity (green) carries the footprint of both the refractory period (τ_{refr}) and the timescale of spike rate fluctuations (τ_λ). The autocorrelation computed from shuffled spike trains across repetitions of the stimulus (black) only reflects the latter, as expected. (C) The Effective model we have introduced is expected to be valid when the window size Δt is larger than the refractory period (τ_{refr}) but smaller than the typical timescale of rate variations τ_λ . (D) Mean-variance relation for a window of 33 ms. The green curve is data and the red line is the Effective model prediction.

variations τ_λ , and the refractory period τ_{refr} , the last two of which can be directly read off from the autocorrelation spiking activity (see Figure 8B). If the window size is of the same order compared to the inverse of the spike rate, then the mean spike count λ is small and the spiking process can be described by a Bernoulli binary variable, with variance $V = \lambda(1 - \lambda)$ and Fano factor $1 - \lambda \approx 1$. Further, if the window size is smaller or even of the order of the refractory period, then spiking events in subsequent windows

cannot be treated independently, and the modeling strategy of focusing on single time windows must be revised. In this case a common solution is to introduce a spike history filter (Pillow et al., 2008) that models the spike probability with a dependence on past activity. At the other extreme, if the window size is large compared to the typical timescale over which the spike rate varies, then each window will contain a mixture of spike rates, leading to a Fano factor larger than one (Kara et al., 2000). Several solutions have been proposed to model this overdispersion (Scott & Pillow, 2012; Goris et al., 2014; Charles et al., 2017).

We have chosen a bin size between these two extremes, such that the spike count is not a binary variable ($\Delta t \sim r^{-1}$), but the firing rate stays roughly constant within a single time bin ($\Delta t \ll \tau_\lambda$), and the dependence between spikes in consecutive time bins is relatively weak ($\Delta t \gg \tau_{\text{refr}}$). The advantage of this choice of bin size is the small number of parameters (only two) needed to describe the spike regularity. We also note that the bin size chosen corresponds to the timescale of the retinal code (Berry & Meister, 1998). To test the robustness of our modeling approach to the choice of window size, we have fitted an Effective model (see equation 5.3) to the distribution of spike counts taken in a window of 33 ms instead of 16 ms and still found excellent agreement (see Figure 8D). The interplay between the different timescales discussed here will vary across brain areas and conditions, and the right window size should be picked as a function of these considerations for our approach to be applicable.

Here we have focused on the spike count statistics of single cells. The activity of other cells in the network is also expected to affect spike count variability through noise correlations. Future work should encompass both sources of non-Poissonian variability through self- and cross-correlations in a single framework (Ferrari et al., 2018).

In this study, we have pooled cells of the same type (ON and OFF), because they are known to have equivalent response properties (Devries & Baylor, 1997). However, we checked that fitting the Effective model on single cells rather than on pooled data also gave excellent performance, with even an improvement for about half of the cells (14/25 for OFF cells, 7/19 for ON cells), with a $\approx 11\%$ and $\approx 16\%$ increase in log-likelihood gain relative to the Poisson model for OFF and ON cells, respectively. This demonstrates the potential applicability of our method to brain areas where cells cannot be pooled by types.

10 Methods

10.1 Equilibrium Poisson Process with Absolute Refractory Period.

A Poisson process with firing rate $\Theta(u - \tau)r$, where Θ is the Heaviside step function of the time from the previous spike, has an interspike interval distribution given by equation 3.1. For such a process, it is possible to compute the probability distribution of the number of spikes emitted in a time

window Δt , whose starting point is chosen as random (Müller, 1973, 1974), $P_{\Theta}(n|r, \tau, \Delta t) = P_{\Theta}(n|v, f)$, where $f = \tau/\Delta t$ and $v = r\Delta t$ (see section A.1 in the appendix for a derivation):

$$\begin{aligned}
 P_{\Theta}(n|v, f) = & \frac{1}{1 + vf} \left[\Phi(n) + \Theta(n^{\text{Max}} - 2 - n) \sum_{j=0}^n (n + 1 - j) \right. \\
 & \times \frac{v^j (1 - (n + 1)f)^j e^{-v(1 - (n + 1)f)}}{j!} \\
 & - 2\Theta(n^{\text{Max}} - n - 1) \sum_{j=0}^{n-1} (n - j) \frac{v^j (1 - nf)^j e^{-v(1 - nf)}}{j!} \\
 & + \Theta(n^{\text{Max}} - n) \sum_{j=0}^{n-2} (n - 1 - j) \\
 & \left. \times \frac{v^j (1 - (n - 1)f)^j e^{-v(1 - (n - 1)f)}}{j!} \right] \tag{10.1}
 \end{aligned}$$

$$\Phi(n) \equiv \begin{cases} 0 & , \quad n \leq n^{\text{Max}} - 2 \\ n^{\text{Max}}(1 + vf) - v & , \quad n = n^{\text{Max}} - 1 \\ v - (n^{\text{Max}} - 1)(1 + vf) & , \quad n = n^{\text{Max}} \end{cases} \tag{10.2}$$

where n^{Max} is the smallest integer larger than $\Delta t/\tau$ and we used the convention $\Theta(0) = 1$. The distribution 10.1 has an expected value given by

$$E_{\Theta}(n|v, f) = \frac{v}{1 + vf} \tag{10.3}$$

and exact variance (Müller, 1974):

$$\begin{aligned}
 V_{\Theta}(n|v, f) = & \frac{2 \sum_{n=0}^{n^{\text{Max}}-1} \left[v(1 - nf) - n + \sum_{j=0}^{n-1} (n - j) \frac{v^j (1 - nf)^j e^{-v(1 - nf)}}{j!} \right] - v - \frac{v^2}{1 + vf}}{1 + vf} \tag{10.4}
 \end{aligned}$$

10.2 Inference of the Absolute Refractory Period Model. To infer the value of f for the absolute refractory model (see equation 10.1), we perform a mean square error (MSE) minimization of the mean-variance relation:

$$MSE \equiv \sum_{i,t} [V_i(t) - \text{Var}_{\Theta}(n|v_i(t), f)]^2, \tag{10.5}$$

$$v_i(t) = \frac{\lambda_i(t)}{1 - \lambda_i(t)f}, \quad (10.6)$$

where $\lambda_i(t)$ and $V_i(t)$ are the empirical mean and variance of n for the cell i in the time-bin t . We used the exact expression, equation 10.4, for $\text{Var}_{\Theta}(n|v_i(t), f)$, although an approximated results can be obtained by the simpler asymptotic expression (Müller, 1974):

$$V_{\Theta}(n|v, f) \approx \frac{v}{(1 + vf)^3}. \quad (10.7)$$

All predictions of the model were done using cross-validation. The data set was divided into two sets, each made from one-half of a 10 s film repeated 54 times: a training set, on which the fit was done, and a testing set, on which the model predictions were tested.

10.3 Inference of the Second-Order Refractory Model. To infer the value of f for the Second-Order refractory model (see equation 4.1), we perform log-likelihood maximization with a steepest descent algorithm. Thanks to the exponential form of $P_{\Theta}^{2\text{nd}}(n|\lambda, f)$, the derivative of the log likelihood (ℓ) with respect to f takes a simple form:

$$\begin{aligned} \ell &\equiv \sum_{i,t} \log P_{\Theta}^{2\text{nd}}(n|\lambda_i(t), f), \quad (10.8) \\ \frac{d\ell}{df} &= \sum_{i,t} \left[(2f - 1) (\langle n_i(t)^2 \rangle_{\text{data}} - \langle n_i(t)^2 \rangle_{\text{model}_i(t)}) \right. \\ &\quad \left. - f (\langle n_i(t)^3 \rangle_{\text{data}} - \langle n_i(t)^3 \rangle_{\text{model}_i(t)}) \right], \quad (10.9) \end{aligned}$$

where $\langle \cdot \rangle_{\text{model}_i(t)}$ means average with the model distribution $P_{\Theta}^{2\text{nd}}(n|\lambda = \lambda_i(t), f)$. At each iteration, we update the value of the parameter using the log-likelihood gradient and adjust the function θ_{λ} accordingly.

10.4 Inference of the Effective Model. To infer the value of γ and δ for the Effective model (see equation 5.3), we perform log-likelihood maximization with Newton's method, where at each iteration, we update the values of the parameters using the log-likelihood gradient:

$$\sum_{i,t} \frac{d}{d\gamma} \log P_{\text{Eff}}(n|\lambda_i(t), f) = \langle n_i(t)^2 \rangle_{\text{data}} - \langle n_i(t)^2 \rangle_{\text{model}_i(t)}, \quad (10.10)$$

$$\sum_{i,t} \frac{d}{d\delta} \log P_{\text{Eff}}(n|\lambda_i(t), f) = \langle n_i(t)^3 \rangle_{\text{data}} - \langle n_i(t)^3 \rangle_{\text{model}_i(t)}, \quad (10.11)$$

and the Hessian:

$$\mathcal{H} = - \begin{pmatrix} \langle n_i(t)^2 \rangle_{\text{model}_i(t)} - \lambda_i(t)^2 & \langle n_i(t)^3 \rangle_{\text{model}_i(t)} - \lambda_i(t) \langle n_i(t)^2 \rangle_{\text{model}_i(t)} \\ \langle n_i(t)^3 \rangle_{\text{model}_i(t)} - \lambda_i(t) \langle n_i(t)^2 \rangle_{\text{model}_i(t)} & \langle n_i(t)^4 \rangle_{\text{model}_i(t)} - \langle n_i(t)^2 \rangle_{\text{model}_i(t)}^2 \end{pmatrix}, \tag{10.12}$$

where $\langle \cdot \rangle_{\text{model}_i(t)}$ now means average with the model distribution $P_{\text{Eff}}(n | \lambda = \lambda_i(t), \gamma, \delta)$. After updating the parameters, we adjust the function θ_λ and iterate until convergence.

10.5 Generalized Count Model. Gao et al. (2015) define a Generalized Count (Gen.Count) distribution that in our notation and framework reads

$$P_{\text{Gen.Count}}(n | \lambda) = \exp \{ \theta_\lambda[G] n + G[n] - \log n! - \log Z_\lambda[G] \}, \tag{10.13}$$

where $G[n]$ is a generic real function defined on the nonnegative integers $n \in [0, \infty]$. To better characterize its λ dependence, we have rewritten Gen.Count introducing the proper $\theta_\lambda[G]$ function.

For $G[n] = -\gamma n^2 - \delta n^3$, the distribution, equation 10.13, reduces to the Effective model (see equation 5.3). The Gen.Count model is thus a generalization of our model and is potentially more flexible in modeling the spike count statistics. This, however, comes at the price of introducing more parameters to fit. In practice, one needs to define $G[n]$ for $n = 0, 1, \dots, n^{\text{Max}}$, where n^{Max} is the maximal value of n observed during the experiment. It can be shown that the model has $n^{\text{Max}} - 1$ free parameters (Gao et al., 2015). If $n^{\text{Max}} = 1, 2, 3$, then Gen.Count is equivalent to Poisson, Effective with $\delta = 0$, and Effective models, respectively. Otherwise it offers a potentially interesting way to generalize our cubic model at the price of inferring the parameters $G[n]$ for large values of n . Thanks to its exponential form, we easily infer the Gen.Count parameters with log-likelihood maximization and have set $n^{\text{Max}} = 4$ and 5 for, respectively, OFF and ON populations.

10.6 Conway-Maxwell-Poisson Model. In Stevenson (2016), the Conway-Maxwell-Poisson (COMP) model (Sellers et al., 2012) has been proposed to account for both under- and overdispersed mean-variance relation. In our notation, the COMP model reads

$$P_{\text{COMP}}(n | \lambda) = \exp \{ \theta_\lambda[\eta] n - \eta \log n! - \log Z_\lambda[\eta] \}, \tag{10.14}$$

where $Z_\lambda[\eta]$ is a normalization constant. For $\eta = 1$, the COMP reduces to the Poisson model, whereas for $G[n] = -(\eta - 1) \log n!$ the Generalized Count

reproduces the COMP. Also for the COMP, the exponential form allows for log-likelihood maximization.

10.7 Equipping Stimulus Processing Model with Second-Order or Effective Noise Term. In this section, we detail the calculation for computing the log-likelihood gradient of equation 8.3. For a general stimulus processing model (see equation 8.1) equipped with noise term P , the log-likelihood reads:

$$\ell(\psi) = \frac{1}{T} \sum_t \log P(n(t) | \hat{\lambda}_\psi(S_t)), \tag{10.15}$$

where the summation runs over the duration of the training set. Consequently, the log-likelihood gradient $\nabla_\psi \ell(\psi)$ reads

$$\nabla_\psi \ell(\psi) = \frac{1}{T} \sum_t \left. \frac{d \log P(n(t) | \lambda)}{d\lambda} \right|_{\lambda=\hat{\lambda}_\psi(S_t)} \nabla_\psi \hat{\lambda}_\psi(S_t). \tag{10.16}$$

To estimate $\nabla_\psi \ell(\psi)$, we thus need to compute the derivative of the log probability with respect to the mean spike count. If P is the Poisson distribution P_{Pois} , then

$$\frac{d}{d\lambda} \log P_{\text{Pois}}(n(t) | \lambda) = \frac{d}{d\lambda} (n(t) \log \lambda - \lambda - \log n(t)!) = \frac{n(t) - \lambda}{\lambda}. \tag{10.17}$$

If P is instead the Effective model,

$$\frac{d}{d\lambda} \log P_{\text{Eff}}(n(t) | \lambda) = \frac{d}{d\theta_\lambda} \left(\log P_{\text{Eff}}(n(t) | \lambda) \right) \frac{d}{d\lambda} \theta_\lambda \tag{10.18}$$

$$= (n(t) - \lambda) \left(\frac{d}{d\theta_\lambda} \lambda \right)^{-1} = \frac{n(t) - \lambda}{V_\lambda^{\text{Eff}}}. \tag{10.19}$$

The very same expression, with replaced V_λ^{Eff} by $V_\lambda^{2\text{nd}}$, holds for the Second-Order model. Note that equation 10.17 has the same form of equation 10.19 because for Poisson $V_\lambda^{\text{Pois}}(t) = \lambda$. Equation 8.3 is in fact a general result for spike counter P belonging to the exponential family. Note that V_λ^{Eff} and $V_\lambda^{2\text{nd}}$ are properties of the noise distribution (respectively, P_{Eff} and $P_\Theta^{2\text{nd}}$). These functions can be numerically estimated before running the inference of the stimulus processing model—for example, by computing their values for several values of λ and then interpolating with a cubic spline.

Mathematical Appendix

A.1 Stimulus. Our stimulus was composed of one or two black bars moving randomly on a gray background. Each bar was animated by a Brownian motion, with additional feedback force to stay above the array, and repulsive forces so that they do not overlap. The trajectories of the bars x_1 and x_2 are described by the following equations (Mora, Deny, & Marre, 2015):

$$\begin{aligned} \frac{dv_1}{dt} = & -\frac{v_1}{\tau} + \text{sign}(x_1 - x_2) \left(\frac{R}{|x_1 - x_2|} \right)^6 \\ & - \omega_0^2(x_1 - \mu_1) + \sigma W_1(t) \end{aligned} \quad (\text{A.1})$$

$$\begin{aligned} \frac{dv_2}{dt} = & -\frac{v_2}{\tau} + \text{sign}(x_2 - x_1) \left(\frac{R}{|x_2 - x_1|} \right)^6 \\ & - \omega_0^2(x_2 - \mu_2) + \sigma W_2(t) \end{aligned} \quad (\text{A.2})$$

where $W_1(t)$ and $W_2(t)$ are two gaussian white noises of unit amplitude, $\mu_2 - \mu_1 = 600 \mu\text{m}$ is the shift between the means, $\omega_0 = 1.04 \text{ Hz}$, $\tau = 16.7 \text{ ms}$, $R = 655 \mu\text{m}$, and $\sigma = 21.2 \mu\text{m} \cdot \text{s}^{-3/2}$. The width of one bar is $100 \mu\text{m}$ and the length 2 mm . For receptive field mapping, a random binary checkerboard was displayed for 1 hour at 50 Hz (check size: $60 \mu\text{m}$).

A.2 Equilibrium Poisson Process with Absolute Refractory Period. Here we provide a sketch of the derivation to obtain the complete expression 10.1. We are interested in computing the probability distribution $P_\Theta(n|r, \tau, \Delta t)$ of the number of spikes emitted in a time window Δt , whose starting point is chosen at random when the interspike interval distribution is

$$\rho_\Theta(t) = \Theta(t - \tau) r e^{-r(t-\tau)}. \quad (\text{A.3})$$

P_Θ can be expressed as the difference between its cumulative distribution:

$$C(n|\Delta t) \equiv \sum_{k=n}^{\infty} P_\Theta(k|\Delta t, r, \tau), \quad (\text{A.4})$$

$$P_\Theta(n|\Delta t, r, \tau) = C(n|\Delta t) - C(n+1|\Delta t). \quad (\text{A.5})$$

Because $C(n|\Delta t)$ is the probability of having at least n spikes in the time bin Δt , it can be computed as

$$C(n|\Delta t) = \int_0^{\Delta t} dt [\rho_\Theta^E \star \rho_\Theta^{*(n-1)}](t), \quad (\text{A.6})$$

where \star is the convolution symbol and $(\cdot)^{\star n}$ means n -times self-convolution. ρ_{Θ}^E is the distribution of the first spike when the beginning of the time bin is chosen at random (equilibrium process) and its distribution can be computed as (Müller, 1973):

$$\begin{aligned} \rho_{\Theta}^E(t) &= \frac{\int_t^{\infty} dt' \rho_{\Theta}(t')}{\int_0^{\infty} dt' t' \rho_{\Theta}(t')} = r \frac{e^{-r \max(t-\tau, 0)}}{1+r\tau} \\ &= r \frac{\Theta(t)\Theta(\tau-t) + \Theta(t-\tau)e^{-r(t-\tau)}}{1+r\tau} = \frac{1}{1+r\tau} (r\mathcal{I}_{[0,\tau]}(t) + \rho_{\Theta}(t)), \end{aligned} \tag{A.7}$$

where $\mathcal{I}_{[a,b]}(t) = 1$ if $t \in [a, b]$ and zero otherwise. Thanks to the explicit decomposition of ρ_{Θ}^E , we have

$$\rho_{\Theta}^E \star \rho_{\Theta}^{\star(n-1)} = \frac{1}{1+r\tau} (r\mathcal{I}_{[0,\tau]} \star \rho_{\Theta}^{\star(n-1)} + \rho_{\Theta}^{\star n}). \tag{A.8}$$

In order to estimate the $\rho_{\Theta}^{\star n}$, we introduce the Laplace transform, which for a generic function $h(t)$ reads

$$\mathcal{L}[h(t)](s) = \int_0^{\infty} dt e^{-st} h(t), \tag{A.9}$$

and use it to get rid of the multiple convolutions:

$$\begin{aligned} \rho_{\Theta}^{\star n} &= \mathcal{L}^{-1} [\mathcal{L}[\rho_{\Theta}^{\star n}]] = \mathcal{L}^{-1} [\mathcal{L}[\rho_{\Theta}]^n] = \mathcal{L}^{-1} \left[\left(\frac{r}{r+s} e^{-s\tau} \right)^n \right] \\ &= \frac{r^n (t-n\tau)^{(n-1)} e^{-r(t-n\tau)}}{(n-1)!} \Theta(t-n\tau) \equiv \gamma[n, r](t-n\tau), \end{aligned} \tag{A.10}$$

where we have introduced the gamma distribution $\gamma[n, r](t) \equiv \Theta(t)r^n t^{(n-1)} e^{-rt} / (n-1)!$. From equation A.6, it follows that

$$\begin{aligned} C(n, \Delta t) &= \frac{1}{1+r\tau} \left(\int_0^{\Delta t} dt \int dt' \mathcal{I}_{[0,\tau]}(t-t') \gamma[n-1, r](t' - (n-1)\tau) \right. \\ &\quad \left. + \int_0^{\Delta t} dt \gamma[n, r](t-n\tau) \right). \end{aligned} \tag{A.11}$$

$C(n|\Delta t)$ can be computed by integrating several times the gamma distribution using the following relation:

$$\frac{1}{(n-1)!} \int_0^a dt t^{(n-1)} e^{-t} \Theta(t) = \left(1 - e^{-a} \sum_{k=0}^{n-1} \frac{a^k}{k!} \right) \Theta(a). \tag{A.12}$$

After some algebra (Müller, 1973, 1974) this calculation provides equation 10.1.

A.3 Small f Expansion of the Refractory Neuron Model. Here we derive the result, equation 5.1, for which equation 4.1 is a particular case. We consider a general model in which the instantaneous spike rate is modulated by a time-dependent factor, $\alpha(u)r$, where r is the spike rate in the absence of refractoriness, with $\alpha(u \geq \tau) = 1$, and u is the time following the previous spike. In this case, the interspike interval distribution is

$$\rho_\alpha(t) = \nu \alpha(t) \exp \left\{ -\nu \int_0^t \alpha(t) \right\} \Theta(t), \tag{A.13}$$

and our goal is to expand, for small τ , $P_\alpha(n|\Delta t, r, \alpha)$, the probability of having n spikes within the time bin Δt . First, we introduce two useful quantities:

$$f \equiv \frac{1}{\Delta t} \int_0^\infty dt (1 - \alpha(t)), \tag{A.14}$$

$$g \equiv \frac{1}{\Delta t} \int_0^\infty dt t (1 - \alpha(t)), \tag{A.15}$$

such that if $\alpha(t) = 1$ for all $t > 0$, then $f = g = 0$, and we expect to recover the Poisson case, and if $\alpha(t) = 0$ for all $t < \tau$, then $f = \tau/\Delta t$ and $g = \tau^2/\Delta t^2/2$, and we expect to recover the absolute refractory case, equation 4.1.

Much like the absolute refractory case, we consider the cumulative distribution of $P_\alpha(n|r, \alpha, \Delta t,)$,

$$C_\alpha(n|\Delta t) = \int_0^{\Delta t} [\rho_\alpha^E \star \rho_\alpha^{*(n-1)}](t), \tag{A.16}$$

where, as before, ρ_α^E is the distribution of the first spike for an equilibrium process and $\rho_\alpha^{*(n-1)}$ is ρ_α self-convoluted $n - 1$ times. To perform the expansion in small τ , we decompose ρ_α^E and ρ_α around an exponential distribution:

$$\rho_\alpha^E(t) = (1 + v^2 g) \rho(t) + \delta \rho_\alpha^E(t), \tag{A.17}$$

$$\rho_\alpha(t) = e^{r(1-A)\tau} \rho(t) + \delta \rho_\alpha(t). \tag{A.18}$$

In the following, we first perform the computation with $\rho(t)$ instead of ρ_α^E , leaving for the end the corrections due to the factor $1 + v^2g$ and term $\delta\rho_\alpha^E$. This is equivalent to consider a shifted process (Müller, 1973) where the time bin starts after the end of the last refractory period, instead of the equilibrium process we are considering here. As for $\rho_\alpha^{*n}(t)$ (see equation A.10), we can use the Laplace transform to compute $\rho^{*n}(t) = \gamma[n, r](t)$. This allows us to perform the following expansion:

$$[\rho \star \rho_\alpha^{*(n-1)}](t) = \rho \star \sum_{i=0}^{n-1} \binom{n-1}{i} e^{(n-i-1)vf} [\rho^{*(n-1-i)} \star \delta\rho_\alpha^{*i}](t) \quad (\text{A.19})$$

$$\begin{aligned} &\approx e^{(n-1)vf} \gamma[n, r](t) \\ &\quad + (n-1)e^{(n-2)vf} [\gamma[n-1, r] \star \delta\rho_\alpha](t) \\ &\quad + \binom{n-1}{2} e^{(n-3)vf} [\gamma[n-2, r] \star \delta\rho_\alpha^{*2}](t). \end{aligned} \quad (\text{A.20})$$

To perform the integration of equation A.20, we use the following approximation:

$$\int_0^\infty dt \delta\rho_\alpha(t) \approx -vf - \frac{1}{2}v^2f^2 + O(\tau^3), \quad (\text{A.21})$$

$$\int_0^\infty dt \delta\rho_\alpha^{*2}(t) \approx v^2f^2 + O(\tau^3), \quad (\text{A.22})$$

$$\int_0^\infty dt t \delta\rho_\alpha(t) \approx -vg\Delta t + O(\tau^3), \quad (\text{A.23})$$

which, after some algebra, we obtain for $P_\alpha^S(n|v)$, the distribution for the shifted process:

$$\begin{aligned} P_\alpha^S(n|v) &= \int_0^{\Delta t} dt \left[\rho \star \left(\rho_\alpha^{*(n-1)} - \rho_\alpha^{*n} \right) \right](t) \propto \\ &\quad \times \exp \left\{ n \log v - \log n! + c_1n + c_2n^2 + c_3n^3 \right\} \end{aligned} \quad (\text{A.24})$$

with

$$c_1 = (1 + v) \left(f - \frac{f^2}{2} \right) - \left(\frac{f^2}{2} - g \right), \quad (\text{A.25})$$

$$c_2 = -f + f^2 - (2 + v) \left(\frac{f^2}{2} - g \right), \quad (\text{A.26})$$

$$c_3 = -f^2 + g. \quad (\text{A.27})$$

We need now to account for the full $\rho_\alpha^E(t)$. Because

$$\int_0^\infty dt \delta \rho_\ominus^E(t) \approx -v^2 g \quad (\text{A.28})$$

is of the order τ^2 , we can estimate $P_\alpha(n|\Delta t, r) = P_\alpha(n|v)$ as

$$\begin{aligned} P_\alpha(n|v) &= (1 + v^2 g) P_\alpha^S(n|v) + P_{\text{Pois}}(n-1|v) \int_0^\infty dt \delta \rho_0^E(t) \\ &\propto \exp \{ n \log v - \log n! + (c_1 - v g) n + c_2 n^2 + c_3 n^3 \}. \end{aligned} \quad (\text{A.29})$$

which is equivalent to equation 5.1.

Acknowledgments

We thank R. Brette, M. Chalk, C. Gardella, B. Telenczuk, and M. di Volo for useful discussions. This work was supported by ANR TRAJECTORY, the French state program Investissements d’Avenir managed by the Agence Nationale de la Recherche [LIFESENSES: ANR-10-LABX-65], an EC grant from the Human Brain Project (FP7-785907), NIH grant U01NS090501, and AVIESAN-UNADEV grant to O.M.

Code availability. Code is available online at github.com/UFerrari/Effective_Model.

References

- Baddeley, R., Abbott, L. F., Booth, M. C., Sengpiel, F., Freeman, T., Wakeman, E. A., & Rolls, E. T. (1997). Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proceedings of the Royal Society of London B: Biological Sciences*, 264(1389), 1775–1783.
- Barberini, C. L., Horwitz, G. D., & Newsome, W. T. (2001). A comparison of spiking statistics in motion sensing neurons of flies and monkeys. In J. M. Zanker & J. Zeil (Eds.), *Computational, neural and ecological constraints of visual motion processing* (pp. 307–320). Berlin: Springer.
- Barlow, H., & Levick, W. (1969). Three factors limiting the reliable detection of light by retinal ganglion cells of the cat. *Journal of Physiology*, 200(1), 1–24.
- Berry, M. J., & Meister, M. (1998). Refractoriness and neural precision. *Journal of Neuroscience*, 18(6), 2200–2211.
- Charles, A. S., Park, M., Weller, J. P., Horwitz, G. D., & Pillow, J. W. (2017). Dethroning the Fano factor: A flexible, model-based approach to partitioning neural variability. *bioRxiv*, 165670.
- Chichilnisky, E. (2001). A simple white noise analysis of neuronal light responses. *Network: Computation in Neural Systems*, 12(2), 199–213.
- Churchland, M. M., Byron, M. Y., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S., . . . Shenoy, K. V. (2010). Stimulus onset quenches neural variability: A widespread cortical phenomenon. *Nature neuroscience*, 13(3):369–378.

- de Ruyter van Steveninck, R. R., Lewen, G. D., Strong, S. P., Koberle, R., & Bialek, W. (1997). Reproducibility and variability in neural spike trains. *Science*, 275(5307), 1805–1808.
- del Castillo, J., & Pérez-Casany, M. (2005). Overdispersed and underdispersed Poisson generalizations. *Journal of Statistical Planning and Inference*, 134(2), 486–500.
- Deny, S., Ferrari, U., Mace, E., Yger, P., Caplette, R., Picaud, S., . . . Marre, O. (2017). Multiplexed computations in retinal ganglion cells of a single type. *Nature Communications*, 8(1), 1964.
- Devries, S. H., & Baylor, D. A. (1997). Mosaic arrangement of ganglion cell receptive fields in rabbit retina. *Journal of Neurophysiology*, 78(4), 2048–2060.
- DeWeese, M. R., & Zador, A. M. (2003). Binary coding in auditory cortex. In M. R. DeWeese & A. M. Zador (Eds.), *Advances in neural information processing systems* (pp. 117–124). Cambridge, MA: MIT Press.
- Ferrari, U., Deny, S., Chalk, M., Tkačik, G., Marre, O., & Mora, T. (2018). Separating intrinsic interactions from extrinsic correlations in a network of sensory neurons. *Phys. Rev. E*, 98, 042410.
- Gao, Y., Busing, L., Shenoy, K. V., & Cunningham, J. P. (2015). High-dimensional neural spike train analysis with generalized count linear dynamical systems. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in neural information processing systems*, 28 (pp. 2044–2052). Red Hook, NY: Curran.
- Goris, R., Movshon, J., & Simoncelli, E. (2014). Partitioning neuronal variability. *Nature Neuroscience*, 17(6), 858–865.
- Gur, M., Beylin, A., & Snodderly, D. M. (1997). Response variability of neurons in primary visual cortex (V1) of alert monkeys. *Journal of Neuroscience*, 17(8), 2914–2920.
- Heggelund, P., & Albus, K. (1978). Response variability and orientation discrimination of single cells in striate cortex of cat. *Experimental Brain Research*, 32(2), 197–211.
- Kara, P., Reinagel, P., & Reid, R. C. (2000). Low response variability in simultaneously recorded retinal, thalamic, and cortical neurons. *Neuron*, 27(3), 635–646.
- Maimon, G., & Assad, J. A. (2009). Beyond Poisson: Increased spike-time regularity across primate parietal cortex. *Neuron*, 62(3), 426–440.
- Marre, O., Amodei, D., Deshmukh, N., Sadeghi, K., Soo, F., Holy, T., & Berry, M. (2012). Recording of a large and complete population in the retina. *Journal of Neuroscience*, 32(43), 1485973.
- McFarland, J., Cui, Y., & Butts, D. (2013). Inferring nonlinear neuronal computation based on physiologically plausible inputs. *PLoS Comput. Biol.*, 9(7), e1003143.
- McIntosh, L., Maheswaranathan, N., Nayebe, A., Ganguli, S., & Baccus, S. (2016). Deep learning models of the retinal response to natural scenes. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in neural information processing systems*, 29 (pp. 1361–1369). Red Hook, NY: Curran.
- Mora, T., Deny, S., & Marre, O. (2015). Dynamical criticality in the collective activity of a population of retinal neurons. *Phys. Rev. Lett.*, 114(7), 1–5.
- Movshon, J. (2000). Reliability of neuronal responses. *Neuron*, 27(3), 412.
- Müller, J. W. (1973). Dead-time problems. *Nuclear Instruments and Methods*, 112(1–2), 47–57.

- Müller, J. W. (1974). Some formulae for a dead-time-distorted Poisson process: To André Allisy on the completion of his first half century. *Nuclear Instruments and Methods*, 117(2), 401–404.
- Pillow, J., Shlens, J., Paninski, L., Sher, A., Litke, A., Chichilnisky, E. J., & Simoncelli, E. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454, 995–999.
- Scott, J., & Pillow, J. W. (2012). Fully Bayesian inference for neural models with negative-binomial spiking. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems*, 25 (pp. 1898–1906). Red Hook, NY: Curran.
- Sellers, K. F., Borle, S., & Shmueli, G. (2012). The Com-Poisson model for count data: a survey of methods and applications. *Applied Stochastic Models in Business and Industry*, 28(2), 104–116.
- Stevenson, I. H. (2016). Flexible models for spike count data with both over-and under-dispersion. *Journal of Computational Neuroscience*, 41(1), 29–43.
- Tolhurst, D., Movshon, J., & Thompson, I. (1981). The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Experimental Brain Research*, 41(3–4), 414–419.
- Yger, P., Spampinato, G. L. B., Esposito, E., Lefebvre, B., Deny, S., Gardella, C., . . . Marre, O. (2016). *Fast and accurate spike sorting in vitro and in vivo for up to thousands of electrodes*. bioRxiv.

Received February 15, 2018; accepted June 14, 2018.