

Designing optimal stimuli to control neuronal spike timing

Yashar Ahmadian¹, Adam M. Packer², Rafael Yuste² and Liam Paninski¹

¹ Department of Statistics and Center for Theoretical Neuroscience, Columbia University,

² Howard Hughes Medical Institute, Department of Biological Sciences, Columbia University,
yashar@stat.columbia.edu, adampacker@columbia.edu, rmy5@columbia.edu, liam@stat.columbia.edu

April 5, 2011

Abstract

Recent advances in experimental stimulation methods have raised the following important computational question: how can we choose a stimulus that will drive a neuron to output a target spike train with optimal precision, given physiological constraints? Here we adopt an approach based on models which describe how a stimulating agent (such as an injected electrical current, or a laser light interacting with caged neurotransmitters or photosensitive ion channels) affect the spiking activity of neurons. Based on these models, we solve the reverse problem of finding the best time-dependent modulation of the input, subject to hardware limitations as well as physiologically inspired safety measures, that causes the neuron to emit a spike train which with highest probability will be close to a target spike train. We adopt fast convex constrained optimization methods to solve this problem. Our methods can potentially be implemented in real time and may also be generalized to the case of many cells, suitable for neural prosthesis applications. Using biologically sensible parameters and constraints, our method finds stimulation patterns that generate very precise spike trains in simulated experiments. We also tested the intracellular current injection method on pyramidal cells in mouse cortical slices, quantifying the dependence of spiking reliability and timing precision on constraints imposed on the applied currents.

1 Introduction

Controlling the activity of neurons is a valuable tool in both the study of neural circuits, and in medical applications. Direct electrical stimulation of intact brains, for example, has provided important insight into the function of various neural circuits in information processing and learning (Salzman *et al.*, 1990; Ditterich *et al.*, 2003). More recently, optical methods have been developed for highly precise and non-invasive control of spike timing in neuronal populations via optical uncaging of caged neurotransmitters (Callaway and Yuste, 2002; Nikolenko *et al.*, 2007; Matsuzaki *et al.*, 2008; Nikolenko *et al.*, 2008), or via photo-stimulation of light-sensitive ion channels (Boyden *et al.*, 2005; Han and Boyden, 2007; Mohanty *et al.*, 2008; Gunaydin *et al.*, 2010; Gradinaru *et al.*, 2010). In all of these methods, however, the computational task of designing the stimulation pattern so as to elicit a desired spike train in a single cell or a group of cells, as faithfully as the hardware allows, given safety constraints, can be complicated (Histed *et al.*, 2009).

In this paper we address the following computational problem: how can we artificially stimulate a neuron (e.g., electrically or optically) so that it will output a target spike train $\mathbf{r}_{\text{target}}$ with optimal precision? As an example, consider the case of photo-stimulation. Suppose

we have a good understanding of how a certain laser light interacts with the ion channels in a neuron and thus affects its membrane potential and spiking activity. Ours is the reverse question: how to find the best photo-stimulation pattern (best modulation of laser intensity with time), which according to our model will result in the neuron producing a spike train that will likely be close to the desired $\mathbf{r}_{\text{target}}$, while minimizing laser power. The problem is formally similar in the case of electrical stimulation.

This general method of optimal spike-train control can have several applications in neuroscience as well as in neuroprosthetics. As an example, we can use this framework to produce a set of desired spike trains (*in vivo* or *in vitro*) in a number of selected neurons out of some network, and observe the effect of the produced spikes on the subsequent activity of the neurons in the network or on behavior (Salzman *et al.*, 1990). In particular, this can be very useful in studying connectivity patterns in the network (Petreanu *et al.*, 2007; Wang *et al.*, 2007; Arenkiel *et al.*, 2007; Matsuzaki *et al.*, 2008). Another application can be in studying the role of different possible factors (inhibition, etc) in producing precisely timed spikes, given biophysical constraints (Mainen and Sejnowski, 1995). Model-based optimization methods have also been used in controlling firing rates in the subthalamic nucleus as a therapy for advanced Parkinson’s disease (Feng *et al.*, 2007a,b).

Another important application of this general problem is found in sensory neural prosthetics, such as cochlear (Loizou, 1998) or retinal implants (Weiland *et al.*, 2005). In such applications one needs to solve the following problem: given an auditory or visual sensory signal, how can we transduce this information by stimulating sensory neurons (e.g., in the cochlear or optic nerve) to fire in such a way that some percept of the sensory environment can be restored to the patient? More precisely, any sensory neural prosthetic must at least implicitly solve the following sequence of signal-processing problems in real time. First, sense the stimulus (e.g., via a microphone or a video camera). Then simulate the transduction of the sensory signal into a multineuronal spike train response \mathbf{r} (e.g., via a model of cochlear nerve or retinal ganglion cell response properties). Finally, stimulate real nerve cells to fire with the desired output pattern $\mathbf{r}_{\text{target}} = \mathbf{r}$. The final step corresponds to the problem that is the focus of the current paper.

Of course, the optimal control problem as stated above is ill-posed: for example, if we can inject any arbitrary current into a cell, we can simply make the cell fire with any desired pattern for a short period of time. But this may damage the cell or radically alter its biophysics. In addition, injecting too high a current from one extracellular electrode targeted at a particular cell will activate neighboring cells. Instead, we need to solve a constrained optimization problem, since there are limitations on the stimuli we can safely apply in any physiological preparation. Thus, as we will see, our task becomes a constrained optimization problem, and as we will show below, we can use fast and efficient optimization methods to solve it. We note that real-time convex optimization methods and their numerous applications, e.g., in control problems, constitute a major new theme in the engineering literature and have been reviewed recently in Mattingley and Boyd (2010).

In the following section we will formalize the general problem, and in the sections afterwards we will consider the more concrete cases of electrical, optical, and sensory stimulation.

2 Methods

2.1 General formulation

Let us return to the general problem. We assume that we have a model that accurately describes how an artificial stimulus \mathbf{x} (such as a time-varying injected current or laser that opens certain ion channels or pumps in the neuron) affects the output spike train of a neuron. Mathematically, this can be formalized by a conditional probability distribution $p(\mathbf{r}|\mathbf{x})$ which tells us how the neuron’s response (spike train), \mathbf{r} , varies as the applied stimulus, \mathbf{x} , changes. We also denote the constrained set of allowed stimuli by \mathcal{S} (a concrete example of \mathcal{S} , which we will use in Sec. 2.2 for the case of electrical stimulation, is the set of all current traces that never get larger in absolute value than some fixed maximum allowed current). As discussed in Sec. 1, such constraints have to be imposed in order to guard against unwanted physiological damage and other undesirable effects. Suppose we are given a measure of tolerance for how far the actually emitted spike train, \mathbf{r} , can be from the target spike train, $\mathbf{r}_{\text{target}}$. We can quantify this by a cost function, $C(\mathbf{r}, \mathbf{r}_{\text{target}})$, which becomes larger as \mathbf{r} deviates further from $\mathbf{r}_{\text{target}}$. Then we formalize our problem as finding the \mathbf{x} (subject to the constraint space \mathcal{S}) such that the average cost, given that we applied \mathbf{x} to the system, is minimized

$$\mathbf{x}_* = \arg \min_{\mathbf{x} \in \mathcal{S}} \langle C(\mathbf{r}, \mathbf{r}_{\text{target}}) \rangle_{p(\mathbf{r}|\mathbf{x})} = \arg \min_{\mathbf{x} \in \mathcal{S}} \sum_{\mathbf{r}} C(\mathbf{r}, \mathbf{r}_{\text{target}}) p(\mathbf{r}|\mathbf{x}), \quad (1)$$

where the sum is over all possible spike trains, \mathbf{r} . Here and below, “arg min” means the value of the optimized variable, \mathbf{x} , which yields the minimum value for the expression. Unfortunately, depending on the precise form of the cost function $C(\cdot, \cdot)$ and the details of the conditional distribution $p(\mathbf{r}|\mathbf{x})$, it may be quite challenging to solve this constrained optimization problem to choose the best stimulus \mathbf{x} fast enough, especially for online applications – although the general problem can in principle be dealt with if computational time is not a crucial factor. One approach for making the problem more tractable is to use the sharp cost function, $C(\mathbf{r}, \mathbf{r}_{\text{target}}) = -\delta(\mathbf{r}, \mathbf{r}_{\text{target}})$. The “delta function”, $\delta(\mathbf{r}, \mathbf{r}_{\text{target}})$, is defined so that it equals one when the two arguments are exactly the same and is zero otherwise (i.e. for any deviation of \mathbf{r} from $\mathbf{r}_{\text{target}}$). Using this cost function, the sum in (1) collapses and we obtain

$$\mathbf{x}_* = \arg \min_{\mathbf{x} \in \mathcal{S}} -p(\mathbf{r}_{\text{target}}|\mathbf{x}) = \arg \min_{\mathbf{x} \in \mathcal{S}} -\log p(\mathbf{r}_{\text{target}}|\mathbf{x}), \quad (2)$$

i.e., we need to optimize the likelihood function instead (in writing the last form we used the monotonicity of the logarithm). This will be tractable because we do not need to evaluate an integral involving the likelihood function as in Eq. (1), in each step of the minimization algorithm. Furthermore, in most of the models we consider the log-likelihood $\log p(\mathbf{r}_{\text{target}}|\mathbf{x})$ is concave (i.e., the cost function minimized in Eq. (2) is convex) so that fast convex optimization methods can be used for solving Eq. (2) (c.f. Appendix B). We mention that even though our convenient choice of the delta function as cost is singular (putting a relatively large cost on very small deviations from the target spike train), this will not be problematic in applications where the optimal stimulus, \mathbf{x}_* , does indeed elicit a spike train close to $\mathbf{r}_{\text{target}}$ with high accuracy (see, e.g., Sec. 3). However, when due to the shortcomings of the stimulation method or the circuit physiology, we are unable to precisely control all spike times, it will be preferable to devise a more practical cost function which is smoother and is less sensitive to small deviations from $\mathbf{r}_{\text{target}}$; e.g., one that does not significantly vary as long as the spikes in \mathbf{r} are close to the spikes

in $\mathbf{r}_{\text{target}}$, up to some finite time scale.¹

We will also generalize Eq. (1) to include soft constraints. By “soft constraints” we mean constraints that do not force the optimal \mathbf{x} to be strictly in a certain subset of all possibilities (the subset \mathcal{S}), but rather penalize some \mathbf{x} ’s more than others in a gradual way. This can be formalized by adding a penalty term, $R(\mathbf{x})$, to the cost function in Eq. (2),

$$\mathbf{x}_* = \arg \min_{\mathbf{x} \in \mathcal{S}} [-\log p(\mathbf{r}_{\text{target}}|\mathbf{x}) + R(\mathbf{x})]. \quad (3)$$

Equation (3) summarizes our general approach to optimal spike train control. We will discuss a few concrete examples of this approach in the following sections. However, first it is worth noting that the optimization problem (3) is formally identical to the problem of maximum *a posteriori* (MAP) based decoding of the stimulus \mathbf{x} given the observed spike train \mathbf{r} , as discussed, e.g., in Pillow *et al.* (2011). In a decoding problem the task is to estimate the stimulus, \mathbf{x} , that gave rise to an observed spike train, \mathbf{r} . According to Bayes’ rule, the posterior distribution of the stimulus given the observed spike train, $p(\mathbf{x}|\mathbf{r})$, is proportional to the product of the likelihood $p(\mathbf{r}|\mathbf{x})$ and the prior distribution $p(\mathbf{x})$. Taking the logarithm, we therefore have $\log p(\mathbf{x}|\mathbf{r}) = \log p(\mathbf{r}|\mathbf{x}) + \log p(\mathbf{x}) + \text{const.}$. In MAP-based decoding, we take the MAP stimulus as the posterior estimate, i.e. the \mathbf{x} which maximizes $p(\mathbf{x}|\mathbf{r})$, or equivalently

$$\mathbf{x}_{\text{MAP}} = \arg \min_{\mathbf{x}} [-\log p(\mathbf{r}|\mathbf{x}) - \log p(\mathbf{x})]. \quad (4)$$

If we take the prior $p(\mathbf{x})$ to be zero for $\mathbf{x} \notin \mathcal{S}$ (so that $-\log p(\mathbf{x}) = \infty$), and let $p(\mathbf{x}) \propto \exp(-R(\mathbf{x}))$ for $\mathbf{x} \in \mathcal{S}$, where $R(\mathbf{x})$ is smooth, then Eq. (4) becomes identical to Eq. (3). Thus the penalty term may be interpreted in this context as a log-prior. We note that the relationship between optimal control and inference and estimation problems has been noted before in the literature, in other settings (see, e.g., Todorov (2008)).

2.2 Optimal control of electrical stimulation

Now we will discuss the implementation of the optimization problem Eq. (3) in more specific details for concrete examples. First, let us consider direct electrical stimulation of the circuit, via a single stimulating electrode. Our goal is to devise a tractable model of the responses $p(\mathbf{r}|\mathbf{x})$, which we now write as $p(\mathbf{r}|\mathbf{I})$ where I_t is the injected current at time t (we use \mathbf{I} to denote the whole history of the injected current at all times). The model, which we will explicate below, describes the influence of \mathbf{I} on the response of the neuron, \mathbf{r} , in two stages. The first stage describes how \mathbf{I} influences the neuron’s membrane potential, and the second stage describes how subsequently the membrane potential affects the neuron’s spiking probability. We will start

¹The following potential compromise solution retains the computational advantage of using the delta function cost, yet can effectively approximate the usage of a cost function in Eq. (1), which is more robust with respect to spike train deviations from the target. Instead of minimizing $-\log p(\mathbf{r}_{\text{target}}|\mathbf{x})$ in Eq. (2), we minimize $-\sum_{i=1}^n w_i \log p(\mathbf{r}_i|\mathbf{x})$, where \mathbf{r}_i are n different “jittered” versions of $\mathbf{r}_{\text{target}}$, and w_i are positive weights (which can all be taken to be $1/n$ for simplicity). The idea is to choose the jittered spike trains \mathbf{r}_i to be close to $\mathbf{r}_{\text{target}}$, in the sense that the cost $C(\mathbf{r}_i, \mathbf{r}_{\text{target}})$ is not too large. This new objective emphasizes aspects of $\mathbf{r}_{\text{target}}$ that the cost $C(\cdot, \cdot)$ cares about (i.e., depends on more sensitively), and ignores other aspects. At the same time, we avoid the computationally expensive integration in Eq. (1), since the minimized objective $-\sum_{i=1}^n w_i \log p(\mathbf{r}_i|\mathbf{x})$ is still convex in \mathbf{x} , and therefore we can still exploit fast convex optimization methods. We have not yet pursued this approach in depth, however, and will therefore stick to the simpler framework of Eqs. (2) and (3) in the remainder of this paper.

by describing the latter mechanism, as this stage of the model is the same as in the other models we will analyze in the following sections in other settings for spike train control.

We decompose the total membrane potential of the neuron as a sum of two contributions, $V_t + h_t$, where V_t is the contribution of the injected current (we will model the relationship between V_t and I_t below – see Eqs. (7)), and h_t accounts for refractoriness and depends only on the history of the spiking activity, \mathbf{r} ; given this history, h_t is independent of the injected current. We model the spiking activity as a point process with an instantaneous firing rate, λ_t , which depends on $V_t + h_t$ according to the simple model

$$\lambda_t = f(V_t + h_t). \quad (5)$$

In other words, the spiking probability in a small enough time bin of size dt is given by $\lambda_t dt$. The nonlinearity $f(\cdot)$ (which we take to be monotonically increasing, and approaching zero for large negative values of its argument) can be thought of as representing a soft spiking threshold as we will discuss further below. Using the standard point-process likelihood formula (Snyder and Miller, 1991), we thus have

$$-\log p(\mathbf{r}|\mathbf{V}) = \sum_t (\lambda_t dt - r_t \log \lambda_t) + \text{const.}, \quad \lambda_t = f(V_t + h_t), \quad (6)$$

where r_t is a binary vector representing the spike train (r_t is 1 if we have a spike in bin t , and 0 otherwise), and we use \mathbf{V} to denote the whole history of V_t at all times. The constant term in Eq. (6) does not depend on \mathbf{V} . We mention that this model is an instance of a generalized linear model (Brillinger, 1988; McCullagh and Nelder, 1989; Paninski, 2004; Truccolo *et al.*, 2005; Pillow *et al.*, 2008), and spike-response model (Gerstner and Kistler, 2002). Since we will be using the framework of Eq. (2), we will use the fixed target spike train, $\mathbf{r}_{\text{target}}$, in Eq. (6). The spike history contribution, h_t , is then formed by convolving a refractory kernel with the fixed, known target spike train $r_{\text{target}}(t)$;² thus we take h_t to be known and fixed (in particular, we will not need to average over history-dependent variability in h_t here – for a treatment of the latter problem see Toyozumi *et al.* (2009)).

We now turn to the first stage of the model which describes the influence of current on the membrane potential. For a type I neuron (Rinzel and Ermentrout, 1989; Tateno *et al.*, 2004), we model V_t as a simple noisy, leaky integrator

$$\frac{dV(t)}{dt} = -\frac{1}{\tau_V} V(t) + \frac{1}{C} I(t) + \sigma \varepsilon(t), \quad (7)$$

where τ_V is the membrane time constant, $\varepsilon(t)$ is standard Gaussian white noise, σ sets the strength of this noise, and C is the membrane capacitance divided by a gain factor which depends on the electrode and is the fraction of the measured current that actually enters the cells. Ideally, all these model parameters should be fit to data for each neuron.

We pause here to note an important relationship between the full model described by Eq. (7) together with Eq. (5), and the widely used integrate and fire model of neurons. In the integrate and fire model, spikes are activated by a sharp threshold mechanism: when the membrane potential reaches the spiking threshold, the neuron spikes, and the potential is immediately set to some reset value well below the threshold. In cases where the differential equation obeyed by

² That is, it can be written as $h_t = \sum_{t_\alpha < t} H(t - t_\alpha)$, where t_α are the target spike times, and the sum runs over all spike times preceding t . The contribution of different spikes have the same fixed shape, as described by the function $H(\cdot)$, being simply displaced in time according to the spike time.

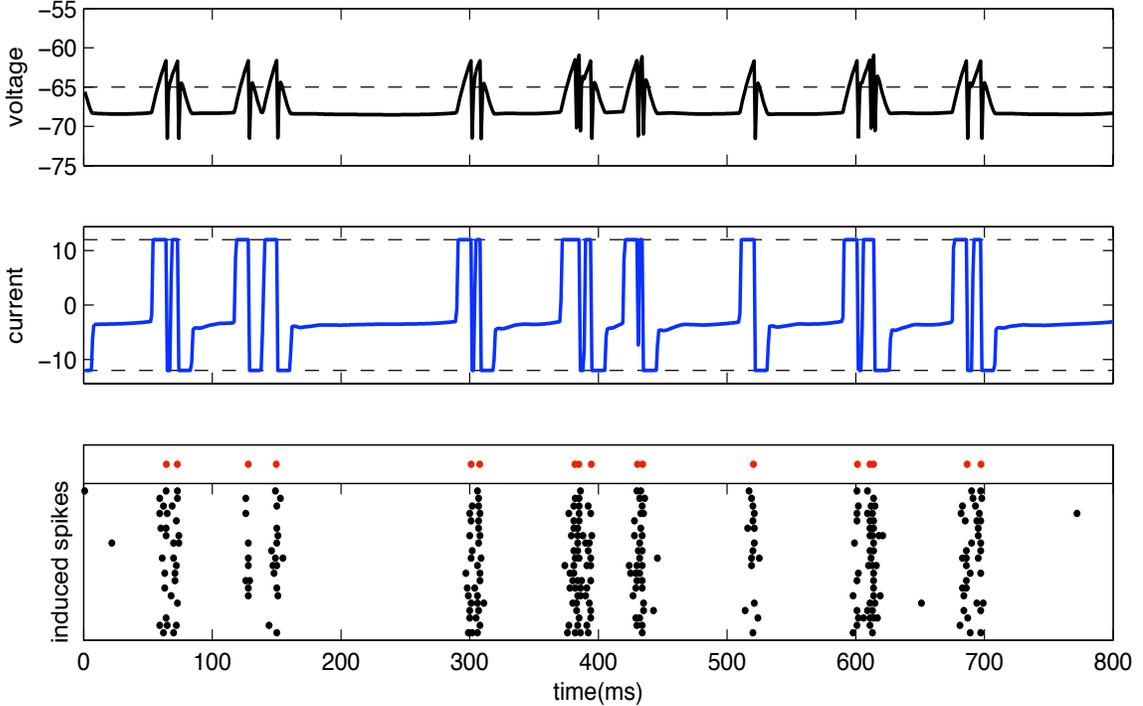


Figure 1: Example of an optimized control current. The top row of the bottom panel shows the target spike train in red. Top panel: predicted voltage trace created by the optimized current. Middle panel: the optimal current. Here, the dashed lines show upper and lower current bounds, $\pm I_{\max}$, where $I_{\max} = 12$ nA. The value of the other model and cost function parameters (see Eqs. (7), (10) and (11)) used in this example were $C = 20$ nF, $\tau_V = 20$ ms, $\sigma = 0.0007$ mV/ $\sqrt{\text{ms}}$, $c = 7 \times 10^{-5}$ ms/nA², and $\tau_J = 15$ ms. In this simulation, and in all other simulations and tests in this paper, we used an exponential function for the threshold nonlinearity of Eq. (5). The spike rasters (in black) in the bottom panel show twenty trial responses to the optimal current injection, to test the reliability of the resulting spike trains. This simulation shows that with physiologically reasonable parameters and constraints the algorithm is able to elicit relatively precise spike trains close to the target.

the sub-threshold membrane voltage is linear (as is the case with Eq. (7)), it is possible to write the full solution (i.e., with reset) for the membrane voltage in such a model as a sum of two parts: a part which obeys that differential equation *without* any reset mechanism, and a part which is given by a sum of negative reset kernels contributed at past spike times.³ This corresponds precisely to the manner in which we have decomposed the membrane potential in Eq. (5), with V_t integrating the current according to Eq. (7) without any reset mechanism; in this analogy, h_t corresponds to the voltage reset mechanism. To describe spiking, however, we have adopted the somewhat different model, Eq. (5), which in the light of this analogy can be thought of as a soft and stochastic threshold mechanism. The advantage of using a soft threshold is that this

³More concretely, for an integrate and fire model in which the total membrane potential, $\tilde{V}(t)$, obeys Eq. (7) *with* the reset mechanism at spike times, it is possible to write $\tilde{V}(t) = V(t) + h(t)$, where $V(t)$ obeys Eq. (7) *without* the reset mechanism, and $h(t)$ which accounts for the reset mechanism is given by $h(t) = \sum_{t_\alpha < t} H(t - t_\alpha)$ with $H(t) = -(V_{\text{th}} - V_{\text{res}}) \exp(-t/\tau_V)$ for $t \geq 0$ and $H(t) = 0$, otherwise. Here, V_{th} and V_{res} are the threshold and the reset voltages, respectively.

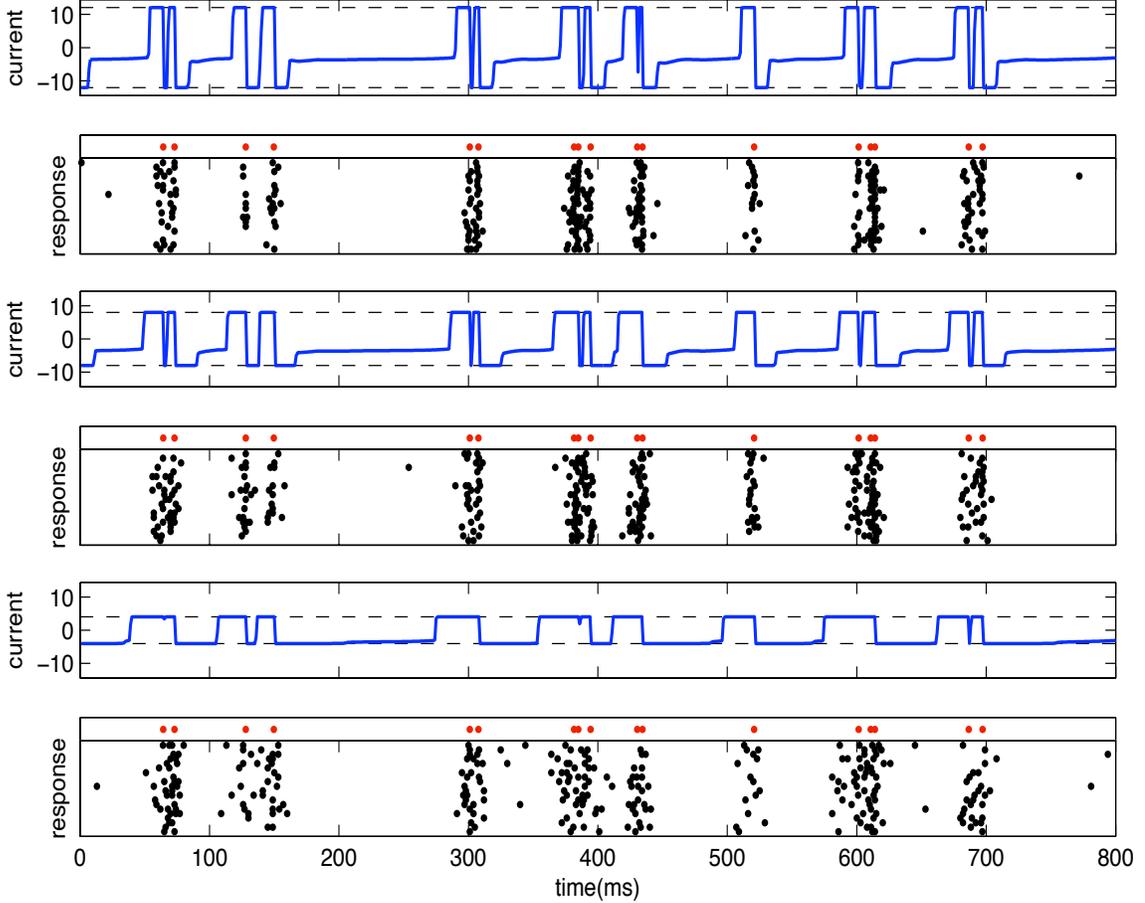


Figure 2: Dependence of optimal control on maximum current bound, I_{\max} . First, third and fifth panels (from top) show the optimized current injections for $I_{\max} = 12$ nA, 8 nA, and 4 nA, respectively (all other model and cost parameters and attributes were fixed in all three cases and were the same as those used in Fig. 1). Dashed lines show upper and lower current bounds, $\pm I_{\max}$. The second, fourth and sixth panels show the target spike trains on the top row (in red) and rasters of twenty trial responses, to test the reliability of the resulting spike train given the optimal currents for those values of I_{\max} . As expected, imposing more restrictive constraints on the current reduces the accuracy of the spike trains elicited by injecting the optimal current.

model is more robust to small changes in parameter settings; in addition, these soft-threshold models are easier to estimate from data (McCullagh and Nelder, 1989; Paninski, 2004; Truccolo *et al.*, 2005).

Going back to Eq. (7), we discretize time in steps of dt , and using the fact that the noise in Eq. (7) is Gaussian, we obtain for the likelihood

$$-\log p(\mathbf{V}|\mathbf{I}) = \frac{1}{2\sigma^2 dt} \sum_t \left(V_{t+dt} - V_t + \frac{dt}{\tau_V} V_t - \frac{dt}{C} I_t \right)^2 + \text{const}, \quad (8)$$

where the constant term is independent of \mathbf{I} and \mathbf{V} (see Appendix A for a detailed derivation of Eq. (8)). Now to choose the best stimulus current, I_t^* , we optimize the penalized log-likelihood

of $\mathbf{r}_{\text{target}}$ given I_t , under some physiologically imposed constraints:

$$\min_{(\mathbf{V}, \mathbf{I}): |I_t| < I_{\max}} -\log p(\mathbf{r}_{\text{target}}|\mathbf{V}) - \log p(\mathbf{V}|\mathbf{I}) + R(\mathbf{I}), \quad (9)$$

where $-\log p(\mathbf{r}|\mathbf{V})$ is given by Eq. (6), and the soft constraints are imposed by

$$R(\mathbf{I}) = c \sum_t J_t^2 dt. \quad (10)$$

Here, J_t is a low-pass filtered version of the injected current \mathbf{I}

$$\tau_J \frac{dJ_t}{dt} + J_t = I_t, \quad (11)$$

and c is a constant which sets the relative importance of the penalty term $R(\mathbf{I})$. J_t is proportional to the charge built up on an electrode with an RC time τ_J . The current is thus constrained to remain within an acceptable safety range $|I_t| < I_{\max}$, while the quadratic cost $R(\mathbf{I})$ ensures that no significant charge builds up on the electrode over a time scale τ_J (since high charge densities lead to cell damage; see, e.g., (Sekirnjak *et al.*, 2006) for further discussion). We note that it is also possible to impose a similar hard constraint on J_t as well, in addition to or instead of the quadratic cost Eq. (10). Figure 1 shows an example of joint optimization of V_t (top panel) and I_t (middle panel) according to Eq. (9). The main features of the optimized current are the sharp rises preceding the spike times followed by sharp inhibitory drops to the minimum allowed current (in this figure, -12 nA) immediately afterwards. Had there been no cost associated to the smoothed energy J_t^2 (see Eq. (10)), the optimal current would stay at the minimum bound throughout most of the silent region, but due to the nonzero cost, Eq. (10), the current rises up to a value closer to zero to minimize that cost.

The constraint parameters (I_{\max}, c) should be set according to physiological safety margins and determine, along with the $1/C$ (which determines the strength of the impact of I_t on V_t), the trade-off between the size of the applied current and the reliability of the output spike train: making I_{\max} larger, for example, leads to larger allowed currents and therefore more reliable spike outputs. Figure 2 shows the effect of constraints (in this case the bounds set by I_{\max}) on the performance of the optimal current. The spike rasters in this figure show the simulated performance of the neuron after being injected with the optimized current subject to three different values of I_{\max} . When I_{\max} is too low, the current can not become large enough such that one could induce a spike by a brief, large current pulse. Instead, the optimal current stays at the maximum allowed value for a relatively long time to compensate for the small magnitude. This in turn leads to a reduced temporal resolution in the timing of the induced spikes.

The astute reader may have realized that in writing Eq. (9) we have apparently deviated from the proposed framework of Eq. (3); since in the present case the artificial stimulus, \mathbf{x} , over which we have control is the current, \mathbf{I} , and not the pair (\mathbf{V}, \mathbf{I}) , according to Eq. (3) we should have minimized $-\log p(\mathbf{r}|\mathbf{I}) + R(\mathbf{I})$, instead of Eq. (9). This would require us to evaluate the marginal probability

$$p(\mathbf{r}|\mathbf{I}) = \int p(\mathbf{r}|\mathbf{V})p(\mathbf{V}|\mathbf{I})d\mathbf{V}. \quad (12)$$

However, if the voltage noise parameter, σ , is small enough, the integrand in Eq. (12) (considered as a function of \mathbf{V}) will be sharply peaked around its maximum, allowing us to approximate the integral using the Laplace method (see, e.g., Kass *et al.* (1991) and Berger (1993)). To a first approximation, this will result in

$$\log p(\mathbf{r}|\mathbf{I}) \approx \max_{\mathbf{V}} [\log p(\mathbf{r}|\mathbf{V}) + \log p(\mathbf{V}|\mathbf{I})], \quad (13)$$

up to corrections that are systematically smaller (Paninski *et al.*, 2010). But using this approximation in Eq. (3) (with \mathbf{I} replacing \mathbf{x}) one arrives at Eq. (9), and the use of the latter optimization is justified. For truly noisy neurons with large σ (e.g., neurons receiving unknown inputs from many separate sources – so that their total input other than I_t can be assumed to be Gaussian), however, the approximation Eq. (13) fails and using Eq. (9) may lead to suboptimal results. In that case, one could potentially use other methods (such as particle filtering) to evaluate the integral in Eq. (12). We will not consider such cases in this paper.

Finally note that, from Eqs. (8), (6), and (10)–(11), the cost function of Eq. (9) is jointly convex in (\mathbf{V}, \mathbf{I}) when $f(\cdot)$ is convex and log-concave (Paninski, 2004). This means that we can carry out the minimization in Eq. (9) using efficient gradient-based approaches such as the well-known Newton-Raphson algorithm (Press *et al.*, 1992). Even so, this problem is still in general computationally expensive, as the computational cost of the matrix inversion involved in the Newton-Raphson step generically scales cubically with the temporal duration of the spike train. However, it turns out that if the Hessian matrix of the minimized cost function is banded, we can improve this scaling down to linear computational time. This is true of the Hessian of the cost function (9) (it is in fact block tri-diagonal with 2×2 blocks) and therefore we can carry out the corresponding minimization in computational time that scales only linearly with the duration of the spike train. For more details about the form of the Hessian and the fast optimization method see Appendix B.

Generalization to oscillatory neural dynamics

The framework of Eqs. (7)–(9) can also be extended to cover neurons which can not be effectively modeled as leaky integrators. As an example, certain neurons show oscillations and subthreshold resonance in their membrane potential. Such behavior has been observed in thalamic neurons (Jahnsen and Karnup, 1994; Puil *et al.*, 1994), neocortical neurons (Gutfreund *et al.*, 1995; Hutcheon *et al.*, 1996; Dickson *et al.*, 2000), and hippocampal neurons (Leung and Yu, 1998; Pike *et al.*, 2000). In the simplest case, the state variable of an oscillatory neuron can be modeled as a two dimensional vector \mathbf{W} which obeys a linear dynamics that gives rise to damped oscillations in its two components (Izhikevich, 2001; Badel *et al.*, 2008). Linearizations of the Hodgkin-Huxley model in certain parameter regimes also give rise to such dynamics (Koch, 1999). In this model, it is only the first component of \mathbf{W} which contributes to the membrane voltage and hence the firing rate. If we denote this component by $V(t)$, Eq. (5) is unchanged for this model. The only modification is that the dynamical equation Eq. (7) changes to

$$\frac{d\mathbf{W}(t)}{dt} = A\mathbf{W}(t) + \mathbf{B}I(t) + \boldsymbol{\varepsilon}(t), \quad (14)$$

where $\mathbf{W}(t) = \begin{pmatrix} V(t) \\ Y(t) \end{pmatrix}$ is the 2-dimensional state vector ($Y(t)$ is the second state variable), \mathbf{B} is a 2 dimensional vector representing the coupling strength of the current, $I(t)$, to the components of $\mathbf{W}(t)$, and A is some 2×2 dynamics matrix, which could give rise to damped oscillating modes. Also, normally we can assume that the injected current will only directly enter the equation for $V(t)$ and not $Y(t)$, and therefore we take $\mathbf{B} = \begin{pmatrix} B \\ 0 \end{pmatrix}$.

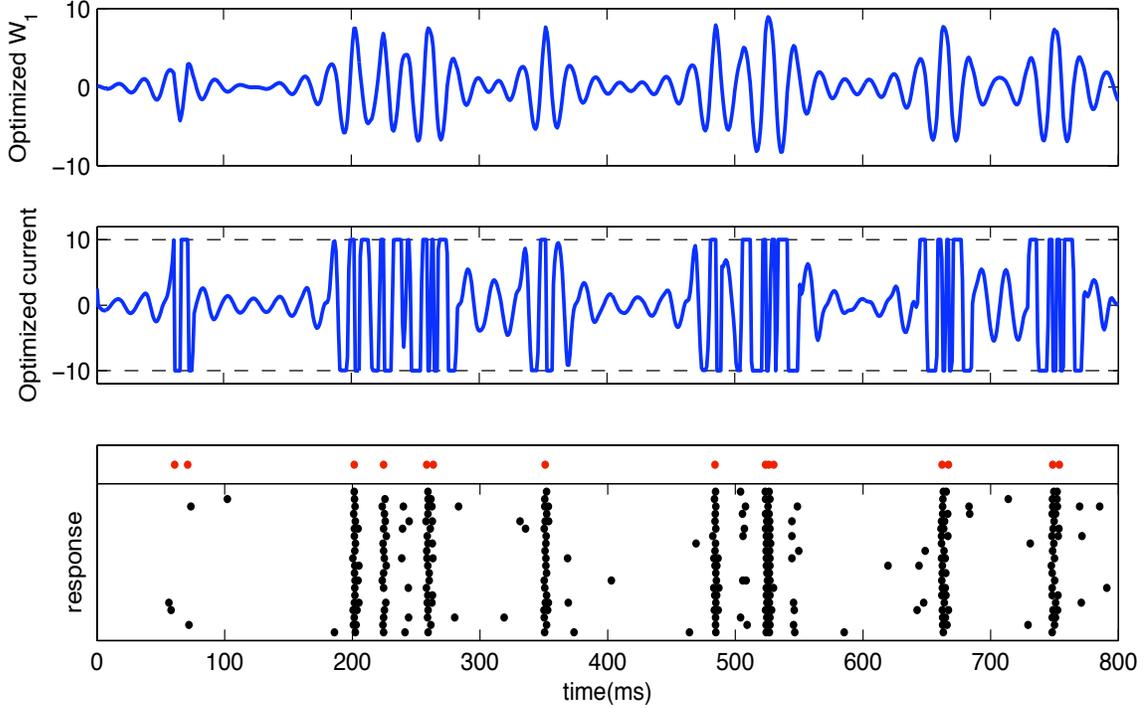


Figure 3: Example of an optimized control current for a neuron with oscillatory dynamics modeled by Eq. (14). In this example, we took the matrix A in Eq. (14) to have the form Eq. (18), with a decay time $\tau_V = 20$ ms and the oscillation period $T := 2\pi/\omega = 20$ ms. The noise covariance C_ϵ was taken to be proportional to the identity, with $\sigma_V^2 = 10^{-8}$ mV²/ms. The electrode time constant, τ_J , and the penalty strength, c , were 15 ms and 0.01 ms/nA², respectively. Top panel: predicted trace of $V(t)$ (first component of $\mathbf{W}(t)$) created by the optimized current. Middle panel: optimized current. The dashed lines show upper and lower current bounds, $\pm I_{\max}$, where $I_{\max} = 10$ nA (taking $B^{-1} = 20$ nF). Bottom panel: the top row shows the target spike train (in red) and the raster shows twenty trial responses to the optimal current injection, to test the reliability of the resulting spike trains.

Thus, the cost function, Eq. (10), and Eq. (6) remain unchanged, but Eq. (8) is modified to

$$-\log p(\mathbf{W}|\mathbf{I}) = \frac{1}{2dt} \sum_t (\mathbf{W}_{t+dt} - \mathbf{W}_t - dt(A\mathbf{W}_t + \mathbf{B}I_t))^T C_\epsilon^{-1} (\mathbf{W}_{t+dt} - \mathbf{W}_t - dt(A\mathbf{W}_t + \mathbf{B}I_t)) + \text{const.}, \quad (15)$$

where C_ϵ is the 2×2 covariance of the temporally white noise ϵ

$$C_\epsilon = \begin{bmatrix} \sigma_V^2 & \rho\sqrt{\sigma_V\sigma_W} \\ \rho\sqrt{\sigma_V\sigma_W} & \sigma_W^2 \end{bmatrix}, \quad (16)$$

and ρ is the correlation coefficient between the two components. The optimization is formally the same as in Sec. 2.2, except we have three variables, (\mathbf{W}_t, I_t) , per time step to optimize over instead of the former two (V_t, I_t) . The optimization problem is

$$\min_{(\mathbf{W}, \mathbf{I}): |I_t| < I_{\max}} -\log p(\mathbf{r}|\mathbf{V}) - \log p(\mathbf{W}|\mathbf{I}) + R(\mathbf{I}). \quad (17)$$

In particular, the fast decoding method referred to at the end of Sec. 2.2, and discussed in detail in Appendix B, can be applied in this case as well. Figure 3 shows a simulated example of optimizing the current injection for this model. In this example, the matrix A had the form

$$A = \begin{bmatrix} -\frac{1}{\tau_V} & -\omega \\ \omega & -\frac{1}{\tau_V} \end{bmatrix}. \quad (18)$$

With this form, and in the absence of the input $I(t)$ and the noise, according to Eq. (14) the components of \mathbf{V} will have damped harmonic oscillations with a decay time of τ_V , and a period of $2\pi/\omega$. Notice the marked difference between the optimized current in this case and those shown in Figs. 1–2, demonstrating the major difference between oscillatory and leaky dynamics. As expected the optimized current itself oscillates at the natural (resonance) frequency of the state vector, at which the latter has the highest susceptibility. The exact form of the optimal solution, however, is quite complicated in this case, and it is unlikely that simple, ad-hoc approaches for designing optimal stimuli will work acceptably here. One interesting aspect of the result is that in the spike rasters for test trials (testing the performance of the optimized current), the unwanted spikes that are fired are not entirely random. The cell’s state vector has a natural oscillation period of 20 ms, which leads to the same periodic tendency in its spike firing. This is the case even if the target spike train does not have a similar periodic pattern; the cell is more likely to fire (unwanted) spikes after about a period since its last firing. This is visible in the bottom panel of Fig. 3, in the timing of the unwanted spikes.

We mention that the problem of optimizing currents for eliciting spikes in various neural oscillator models have also been studied using analytical methods in Moehlis *et al.* (2006); we will discuss this point further in Sec. 4.

Generalization to many cells and electrodes

The formalism introduced above easily generalizes to the case of many cells stimulated by many extracellular electrodes. For simplicity, we will focus here on the case of the leaky integrator neurons described by Eq. (7), although the generalization to oscillatory neurons is straightforward. We introduce a $N \times M$ matrix of gain factors, R_{ij} , describing the influence of electrode j on neuron i (N is the number of cells and M is the number of electrodes). In this setting, Eqs. (5) and (7) generalize to

$$\lambda_i(t) = f(V_i(t) + h_i(t)), \quad (19)$$

$$\frac{dV_i(t)}{dt} = -\frac{1}{\tau_i}V_i(t) + \sum_{j=1}^M R_{ij}I_j(t) + \sigma_i\varepsilon(t), \quad (20)$$

with i ranging over all the cell numbers. We have allowed for cells to have different membrane time constants, τ_i , and noise input strengths, σ_i . The objective function of Eq. (9) is also modified accordingly; $-\log p(\mathbf{r}_{\text{target}}|\mathbf{V})$ and $-\log p(\mathbf{V}|\mathbf{I})$ are each replaced by a sum over the contributions of different cells, with each cell contributing a term similar to Eqs. (6) and (8), respectively, and $R(\mathbf{I})$ is replaced by a sum over all electrodes of terms similar to Eq. (10). In the case where there is a one to one correspondence between electrodes and cells ($M = N$), and furthermore the electrodes designated for each cell does not affect other cells (diagonal R_{ij} matrix), the computational cost of the optimization clearly scales linearly with the number of cells. This is because, in this case, the problem of finding the optimal current in each

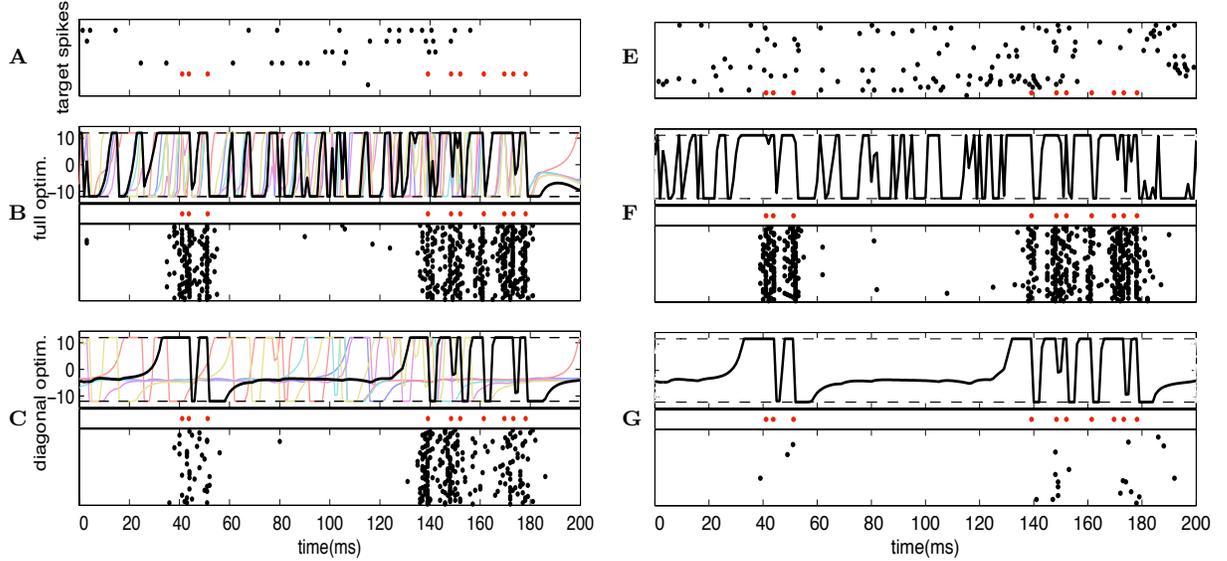


Figure 4: Examples of current optimization in the case of many electrodes stimulating many cells. Panels A–C (E–F) belong to a simulation with 6 (26) cells and 6 (26) electrodes. Panels A and E show the target spike trains, with different rows depicting the target spike train of a different cell. In panels B, C, F and G, the top parts plot the optimized currents, while the bottom parts are rasters of spike train response of one of the cells in the group (whose target spike train is shown in red on top of the raster, as well as in the top panels on each side) to the optimized current, shown in the same panel, in 60 test trials. The optimized currents for the electrode corresponding to the cell whose test spike trains are depicted in the raster are plotted with a thick black curve, while the currents for other electrodes are plotted in color, and are thinner and paler to improve visibility (in panels F and G, the currents in other electrodes are not plotted to avoid confusion). Panels B and F show these for the case of joint optimization of the current in all electrodes, without any approximation. Panels C and G show the optimized currents and the resulting rasters for the approximate optimization where only the diagonal of the cell-electrode coupling matrix is retained. For the exact form of this matrix used in these examples see the main text. The parameters used in these simulations were $B^{-1} = 20$ nF, $\tau_i = 20$ ms, $\sigma_i = 0$, $c = 7 \times 10^{-5}$ ms/nA², and $\tau_J = 20$ ms. The examples shown in this figure show that for a large enough number of electrodes, each coupled to more than one cell, the uncoupled approximation fails completely, yet the full optimization according to Eq. (19), by optimally adjusting the input into each cell from all electrodes, can achieve accurate spike trains.

electrode is independent of the others, and we only need to repeat the single-electrode process for all electrodes. However, if the influence of the control agent designated for one cell on the membrane voltage of other cells cannot be completely avoided (e.g., as in extracellular current stimulation (Histed *et al.*, 2009)), the computational time of a naive joint optimization will scale cubically with the number of cells (see Appendix B for a discussion of the scaling of the Newton-Raphson optimization), which can make the procedure unfeasible for a large number of cells. It may be possible to improve this undesirable scaling, by exploiting the sparse structure of these cross-talk terms (normally, each electrode will affect only a few cells out of all cells of

interest); this remains an important direction for future work.⁴ However, our simulations with groups of cells containing up to about thirty cells show that, even without exploiting sparsity and despite the adverse scaling, the joint optimization algorithm is still tractable for such cases, even for a general coupling matrix R_{ij} .

The panels on the left and right in Fig. 4 show two examples of current optimization in such a setting, for $N = M = 6$ and $N = M = 26$, respectively. In these two examples the matrix R_{ij} was composed of positive elements only, and had the following form. All the diagonal elements were taken to be equal with $R_{ii} = R$, while the off-diagonal elements were sampled independently from a uniform distribution on the interval $[0, R/5]$ (resulting in an average of $R/10$). The panels C and G of Fig. 4 show the result of an approximate solution, obtained by simply ignoring the off-diagonal elements of the R_{ij} matrix; when optimizing for the current from electrode i , the optimizer assumes this current will only affect cell i , and is blind to its effect on the firing of other cells. However, such effects do in fact exist, leading to the suboptimality of the result. In particular, if in the vicinity of the spike times of some cell, some of the currents intended for other cells become negative, the cell’s true current input (which normally peaks immediately before the spike times) can become considerably smaller than what the approximate optimizer assumes, leading to a poor success rate in inducing spikes. The effect is obviously more severe when there are more electrodes, as is visible on panel G of Fig. 4 where $M = N = 26$; here the approximate solution is useless. The full method based on joint optimization of all electrode currents, however, has no such problem; it takes into account all the cross-couplings and balances the electrode currents accordingly, achieving a high performance even in this complicated setting. Here, the optimized currents have no simple relationship with the spike trains, and simple, by-hand approaches for constructing control currents cannot compete with the optimal control method presented here.

2.3 Optimal control of photo-stimulation

Another setting of interest involves photo-stimulation of the circuit instead of direct electrical stimulation. Optical methods for perturbing neural circuits with high spatiotemporal precision have developed dramatically over the last decade (Callaway and Yuste, 2002): for example, neurons can be excited (inhibited) locally via optical uncaging of caged glutamate (GABA) compounds (Nikolenko *et al.*, 2007; Matsuzaki *et al.*, 2008; Nikolenko *et al.*, 2008), or via photo-stimulation following membrane expression of channel rhodopsin-2 (ChR2) or halorhodopsin pumps (NpHR) (Boyden *et al.*, 2005; Han and Boyden, 2007; Mohanty *et al.*, 2008; Gunaydin *et al.*, 2010).

To develop optimal stimulation protocols in this setting, we again need to start with a model of the neural response $p(\mathbf{r}|\mathbf{x})$, where now \mathbf{x} corresponds to various light intensity traces L_t^a , where a is an index distinguishing different light colors. Typically each color of light interacts mainly with one kind of ion channel or pump, although there is some nonzero cross-coupling. For example, ChR2 and NpHR are activated by a blue and yellow light, respectively. In the

⁴In particular, our investigations show that a coordinate descent (CD) approach can be useful in this setting. In this algorithm, instead of solving the full optimization problem jointly for all the electrodes, we adopt a greedy approach where at each step of the CD cycle, we minimize the full cost function (with the full, non-diagonal R_{ij}) only with respect to the current in one electrode. We repeat this step once for all electrodes in each full cycle of the algorithm. For a sparse R_{ij} , the computational cost of each cycle of the CD algorithm scales only linearly with the number of cells. Furthermore, our numerical experiments show that the required number of CD cycles does not show a significant scaling with the number of cells. This leads to a linear (as opposed to cubic) overall scaling of the computational cost with the number of cells.

model below we therefore use the superscript E (I) for both the ChR2 conductance (NpHR pump current) and the blue light (yellow light) that activates it. Again, we use a two stage model of the response, with the instantaneous firing rate depending instantaneously on the membrane voltage, which in turn is described by a conductance-based leaky integrator model:

$$\lambda_t = f(V_t + h_t), \quad (21)$$

$$\frac{dV_t}{dt} = -\frac{1}{\tau_V}V_t + g_t^E(V^E - V_t) - I_t^I + \sigma\epsilon_t, \quad (22)$$

where ϵ_t is Gaussian white noise. Here g_t^E models the conductance of the excitatory ChR2 channels with reversal potentials V^E , and $-I_t^I$ is the inhibitory NpHR pump current. To model the interaction of g_t^E and I_t^I with light, we begin here with a simple first-order model for the dynamics of conductances and pumps driven by photo-stimulation (although, again, more complicated models maybe handled using similar techniques):

$$\frac{dg_t^E}{dt} = -\frac{g_t^E}{\tau_E} + w_{EI}L_t^I + w_{EE}L_t^E, \quad (23)$$

$$\frac{dI_t^I}{dt} = -\frac{I_t^I}{\tau_I} + w_{II}L_t^I + w_{IE}L_t^E, \quad (24)$$

where L_t^E and L_t^I are the applied light intensities in the excitatory and inhibitory channels, respectively, and the weights w_{ab} (where $a, b \in \text{I, E}$), represent a 2×2 gain matrix summarizing how strongly the optical pulse L_t^b influences the conductance g_t^a ; this is a function of the optical qualities of the stimulating laser, and the overlap in the absorption spectra of the optical sensors, as well as the concentration of the caged compound in the neuropil or opsin channel expression in the membrane. Depending on the application, if saturation effects turn out to be important, then we should modify Eqs. (23)–(24) into nonlinear equations in which the conductances can not grow without bound if a large light intensity is used. We will, however, ignore this complication in this paper.

Now we may write down our optimization problem. Our major constraint in the electrical stimulation setting was that the applied current I_t was bounded in absolute value; here, the applied light intensities L_t^E and L_t^I are constrained to be nonnegative and bounded above by L_{\max}^E and L_{\max}^I , respectively. So we want to solve the problem

$$\min_{(\mathbf{V}, \mathbf{g}^E, \mathbf{I}^I): 0 \leq L_t^I \leq L_{\max}^I, 0 \leq L_t^E \leq L_{\max}^E} -\log p(\mathbf{r}|\mathbf{V}) - \log p(\mathbf{V}|\mathbf{g}^E, \mathbf{I}^I) + R(\mathbf{L}^E, \mathbf{L}^I), \quad (25)$$

where $\log p(\mathbf{r}|\mathbf{V})$ is again given by Eq. (6), and

$$-\log p(\mathbf{V}|\mathbf{g}^E, \mathbf{I}^I) = \frac{1}{2\sigma^2} \sum_t \left(V_{t+dt} - V_t - dt \left(-\frac{1}{\tau_V}V_t - I_t^I + g_t^E(V^E - V_t) \right) \right)^2, \quad (26)$$

$$R(\mathbf{L}^E, \mathbf{L}^I) = c_E \sum_t (L_t^E)^2 dt + c_I \sum_t (L_t^I)^2 dt, \quad (27)$$

where we obtained Eq. (26) by discretization of Eq. (22) (similar to Eq. (38)), using the fact that ϵ_t is Gaussian (cf. the derivation of Eq. (8)). In writing Eq. (25) as an instance of Eq. (3), we used $-\log p(\mathbf{r}|\mathbf{L}^E, \mathbf{L}^I) \approx \min_{\mathbf{V}} [-\log p(\mathbf{r}|\mathbf{V}) - \log p(\mathbf{V}|\mathbf{g}^E, \mathbf{I}^I)]$, similar to the approximation Eq. (13) that we used in writing Eq. (9) (cf. the discussion leading to Eq. (13)). The soft constraint $R(\mathbf{L}^E, \mathbf{L}^I)$ controls the total energy of the light pulses, which we need to limit in

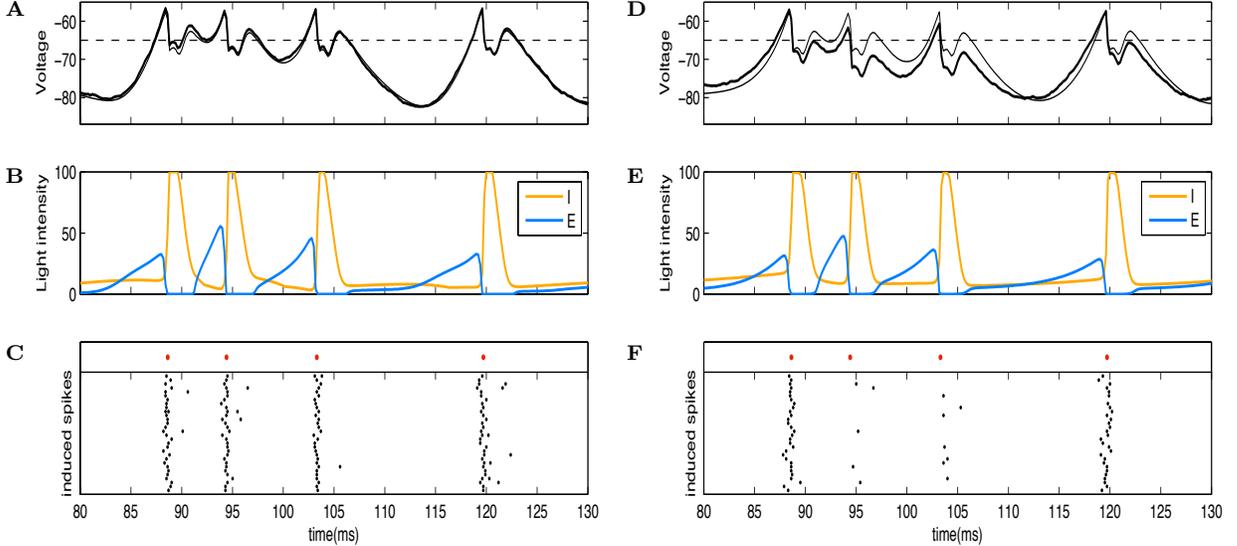


Figure 5: Spike train control by light intensity traces, optimized according to the scheme described after Eq. (30) (panels A-C on the left), and optimized only according to the auxiliary model, Eqs. (28)–(30) (panels D-F on the right). The target spike train is shown on the top row of panels C and F (red spikes). Panels A and D show the membrane potential voltage (the sum of the optimized V_t and the history contribution h_t from the target spike train), and Panels B and E show the optimized light intensity traces. The yellow light, L_t^I , opens inhibitory pumps, and the blue light, L_t^E , opens excitatory channels. The spike rasters (black) in panels C and D each show twenty trials of simulated responses generated by the full model, described by Eqs. (21)–(24), after it was exposed to the optimized light traces of panels B and R, respectively. In panel A (D), the smooth thin curve shows the voltage trace as optimized according to Eq. (25) (Eqs. (29)–(30)), while the slightly deviating thicker and more jagged curve is the actual voltage trace generated stochastically (since $\sigma \neq 0$) according to Eq. (22) in response to the optimized light intensities of panel B (E). The parameters used in this simulation were: $V^E = 20$ mV (in Eqs. (21)–(24), all voltages are measured from the reversal potential of the leak conductance which was taken to be -65 mV in this simulation), $\sigma = 0.36$ mV/ $\sqrt{\text{mS}}$, $\tau_V = 10$ ms, $\tau_E = 5$ ms, $\tau_I = 10$ ms, $L_*^E = L_*^I = 100$, $c_E = c_I = 3 \times 10^{-5}$, $w_{EE} = 0.2$, $w_{II} = 5$, and $w_{IE} = w_{EI} = 0$. We set the average voltage, \bar{V} , used in the auxiliary model, Eq. (28), equal to -70 mV.

order to avoid photodamage. It is implicit in Eq. (25) that when evaluating $R(\mathbf{L}^E, \mathbf{L}^I)$ (and checking the hard constraints on $(\mathbf{L}^E, \mathbf{L}^I)$) for some value of the optimized vector $(\mathbf{V}, \mathbf{g}^E, \mathbf{I}^I)$, we are using Eqs. (23)–(24) to calculate $(\mathbf{L}^E, \mathbf{L}^I)$ in terms of $(\mathbf{g}^E, \mathbf{I}^I)$; the relationship between $(\mathbf{L}^E, \mathbf{L}^I)$ and $(\mathbf{g}^E, \mathbf{I}^I)$ is one-to-one and deterministically fixed via Eqs. (23)–(24). For the same reason, one need not add $(\mathbf{L}^E, \mathbf{L}^I)$ to the list of optimized variables $(\mathbf{V}, \mathbf{g}^E, \mathbf{I}^I)$ in Eq. (25) (see Appendix B for more details on the choice of the set of variables with respect to which we optimize the objective). Once the optimization is finished, we can use Eqs. (23)–(24) again to compute the optimal $(\mathbf{L}^E, \mathbf{L}^I)$, which is what we set out to compute.

The dynamics described by Eq. (21) are nonlinear in the state variables (V, L^E, L^I) , due to the multiplication of the conductance, g_t^E , and V_t , and therefore the resulting likelihood term, Eq. (26), and hence the optimization problem, Eq. (25), may be non-convex. This prevents us from readily using efficient convex optimization methods for solving Eq. (25). However, we can obtain a good initializer by replacing Eq. (22), with the following approximate current-based

system:

$$\frac{dV_t}{dt} = -\frac{1}{\tau_v}V_t + g_t^E(V^E - \bar{V}) - I_t^I + \sigma\epsilon_t, \quad (28)$$

where \bar{V} is an averaged, non-time-varying voltage (e.g., we can take it to be the membrane's rest potential), and the conditional intensity function λ_t and conductances g_t^a follow the same dynamics defined above. The optimization Eq. (25) is modified to

$$\min_{\substack{(\mathbf{V}, \mathbf{g}^E, \mathbf{I}^I): 0 \leq L_t^I \leq L_{\max}^I, 0 \leq L_t^E \leq L_{\max}^E, \\ V_t \leq V^E}} -\log p(\mathbf{r}|\mathbf{V}) - \log \tilde{p}(\mathbf{V}|\mathbf{g}^E, \mathbf{I}^I) + R(\mathbf{L}^E, \mathbf{L}^I) \quad (29)$$

where now the second term

$$-\log \tilde{p}(\mathbf{V}|\mathbf{g}^E, \mathbf{I}^I) = \frac{1}{2\sigma^2} \sum_t \left(V_{t+dt} - V_t - dt \left(-\frac{1}{\tau_v}V_t - I_t^I + g_t^E(V^E - \bar{V}) \right) \right)^2, \quad (30)$$

is convex. We are only using Eq. (30) as an auxiliary optimization for obtaining a good initialization for our non-convex problem (25). However, given that the reversal potential of the excitatory ChR2 channels are well above the sub-threshold values of the membrane potential, we can ignore the time variations of the difference $V^E - V_t$ in Eq. (22) in the first approximation; this corresponds to using Eq. (28). Thus in fact the solution of Eq. (29) gives a good approximation to the final solution of Eq. (25). Note that in Eq. (29) we have now also included an upper bound on V_t , since in the original conductance-based model, in the deterministic case (corresponding to $\sigma = 0$ in Eq. (22)), V_t will never rise above the reversal potential of the excitatory channels V^E , whereas this bound is not respected automatically in the current-based model.

The convex optimization problem, Eq. (29), can be solved with our usual fast barrier techniques once we write the Hessian in block-tridiagonal form, with each block of size 3×3 (one dimension each for V_t, g_t^I , and g_t^E). Again, generalizations to the multineuronal case are straightforward. Let us denote the solution to the auxiliary optimization problem (29) by $(\mathbf{V}_{(0)}, \mathbf{g}_{(0)}^E, \mathbf{I}_{(0)}^I)$. Once $(\mathbf{V}_{(0)}, \mathbf{g}_{(0)}^E, \mathbf{I}_{(0)}^I)$ is found, we can use it as an initialization for the original non-convex problem Eq. (25), which would now be effectively convex if $(\mathbf{V}_{(0)}, \mathbf{g}_{(0)}^E, \mathbf{I}_{(0)}^I)$ were close enough to the true minimum of Eq. (25). However, we saw that depending on the model and constraint parameters, sometimes $(\mathbf{V}_{(0)}, \mathbf{g}_{(0)}^E, \mathbf{I}_{(0)}^I)$ was not already in such a locally convex region close enough to the minimum of Eq. (25). On the other hand, our simulations indicate that for a wide range of parameters, this problem can be solved if we carry out the optimization of the full model in two steps. First, initializing by $(\mathbf{V}_{(0)}, \mathbf{g}_{(0)}^E, \mathbf{I}_{(0)}^I)$, we optimize the objective in Eq. (25) with respect to $(\mathbf{V}, \mathbf{I}^I)$, but keep \mathbf{g}^E fixed and equal to $\mathbf{g}_{(0)}^E$ throughout the optimization; note that even though the objective Eq. (25) is not jointly convex in $(\mathbf{V}, \mathbf{g}^E, \mathbf{I}^I)$, it is convex in $(\mathbf{V}, \mathbf{I}^I)$. We denote the resulting solution by $(\mathbf{V}_{(1)}, \mathbf{g}_{(0)}^E, \mathbf{I}_{(1)}^I)$. Finally, we use $(\mathbf{V}_{(1)}, \mathbf{g}_{(0)}^E, \mathbf{I}_{(1)}^I)$ to initialize the full, original optimization, Eq. (25), optimizing all three vectors to obtain the true optimal solution $(\mathbf{V}_*, \mathbf{g}_*^E, \mathbf{I}_*^I)$. As we noted above, we then use Eqs. (23)–(24) to solve for $(\mathbf{L}_*^E, \mathbf{L}_*^I)$ in terms of the optimized $(\mathbf{g}_*^E, \mathbf{I}_*^I)$.

Panels A-C of Fig. 5 show the final result for an example of such a three-stage optimization process. The main aspect of the optimized light traces is that for eliciting a spike, an excitatory light pulse is followed by a sharp inhibitory light pulse. The shapes of these pulses are quite stereotypical, except that for spikes that follow other spikes within $\lesssim 20$ ms, the optimal pulse shapes can depend in a nontrivial way on the duration of the corresponding inter-spike interval. One aspect of the result worth noting is that controlling inhibition is crucial for eliciting precisely

timed spikes; because the intrinsic time scale of the membrane potential, τ_V , is about 10-20 ms, without the sharp inhibitory light pulses, the voltage peaks generated by excitatory pulses alone would be too wide, staying near threshold on a time scale set by τ_V . This would lead to a (poor) spike timing precision of the same order. On the other hand, with inhibition included, spikes can be elicited with the temporal accuracy of 1-3 ms.

Panels D-F of Fig. 5 show the same setting, but with light intensity traces optimized only according to the auxiliary model, Eqs. (28)–(30). As opposed to the case of full optimization (Fig. 5), here the actual voltage deviates significantly from the optimized voltage trace, being systematically lower in a window of approximately 20 ms after a spike has been fired. As a result, the reliability of induced spikes which trail an initial spike within such a window are significantly degraded. This shows that when the model of the interaction of the stimulating agent and the membrane voltage used in the control optimization (in this case the auxiliary model Eq. (28)) miss important biophysical aspects (in this case the different effects of conductance vs. current input), the results can be highly suboptimal.

Before closing this section we note that, in the first approximation, we can also use the above model with minimal modification to model how optical stimulation using glutamate and GABA uncaging works. Since GABA activates inhibitory channels, as opposed to pumps, we have to model it similarly to the excitatory conductance in Eqs. (22)–(23). So we rewrite Eqs. (22) and (24) as

$$\frac{dV_t}{dt} = -\frac{1}{\tau_V}V_t + g_t^E(V^E - V_t) + g_t^I(V^I - V_t) + \sigma\epsilon_t, \quad (31)$$

$$\frac{dg_t^I}{dt} = -\frac{g_t^I}{\tau_I} + w_{\Pi}L_t^I + w_{IE}L_t^E, \quad (32)$$

and replace I_t^I in Eq. (26) and the auxiliary Eq. (30) with $g_t^I(V^I - V_t)$ and $g_t^I(V^I - \bar{V})$, respectively. Now we also need to impose the constraint $V_t \geq V^I$ when solving the modified auxiliary problem Eq. (29), for reasons similar to those given after Eq. (30). We have to mention, however, that the spatial diffusion of such uncaged neurotransmitters is potentially a significant factor which is ignored in the simplest treatment here, and will need to be considered in a more detailed study.

2.4 Optimal control of a sensory stimulus

Another application of the optimal control method, Eq. (3), can be found in cases where one aims to design or control a natural sensory stimulus, \mathbf{x} , for inducing a target spike train $\mathbf{r}_{\text{target}}$. In this case, the likelihood function $p(\mathbf{r}|\mathbf{x})$ can be interpreted as an encoding model that effectively describes the sensory processing in the neural circuit starting from the periphery to the neuron in which we seek to induce $\mathbf{r}_{\text{target}}$. The generalized linear model (GLM) framework provides an accurate and commonly used model for this purpose, especially for neurons in early sensory areas (Brillinger, 1988; McCullagh and Nelder, 1989; Paninski, 2004; Truccolo *et al.*, 2005; Pillow *et al.*, 2008).

We need only to minimally modify the formalism to cover this case. In particular, Eq. (6) remains valid. The modification enters in the relationship between V_t (which enters Eq. (5) for instantaneous firing rate) and the stimulus \mathbf{x} , which now replaces the current, I_t , in the last section. In the GLM framework, the relationship is provided by a linear filter \mathbf{k}_t , representing the cell’s “receptive field”,

$$V_t = \mathbf{k}_t \cdot \mathbf{x}. \quad (33)$$

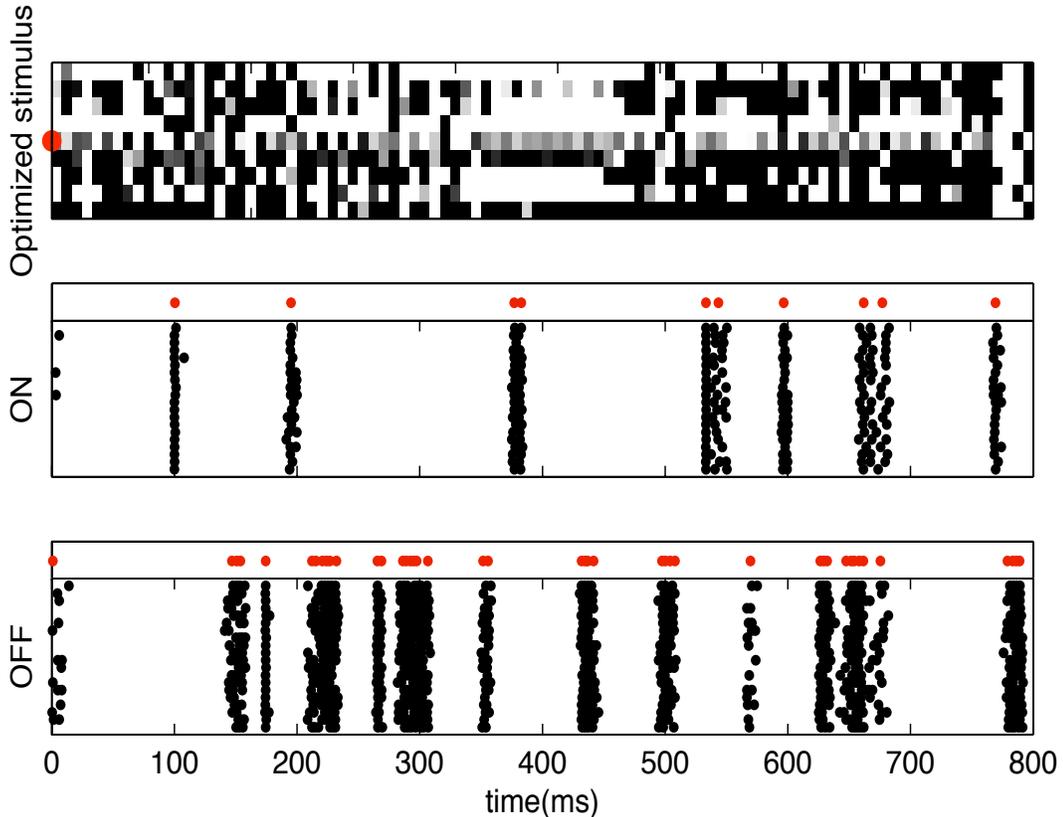


Figure 6: Optimizing a visual stimulus (a black and white checkered stimulus updated at 120 Hz) to induce target spike trains in a pair of ON and OFF retinal ganglion cells with more or less coinciding receptive fields. The top panel shows the stimulus, where the 3×3 square covered by the stimulus is shown by a nine pixel column for illustration purposes. The pixel covering the receptive field centers (of both cells) is marked by a red dot on the left side of this panel. The bottom two panels show the target spike trains (the top rows of each panel, in red) for the ON and the OFF cells, and rasters of 20 trial responses (black) of each cell to the optimized stimulus, respectively.

Since in this case, the relationship between V_t and the stimulus is deterministic (no noise enters Eq. (33)), the second term vanishes when modifying Eq. (9), and we instead should solve the optimization problem

$$\min_{\mathbf{x}: |\mathbf{x}| \leq x_{\max}} -\log p(\mathbf{r}_{\text{target}} | V_t = \mathbf{k}_t \cdot \mathbf{x}) + R(\mathbf{x}), \quad (34)$$

where the first term in the cost function is given by Eq. (6) with the proper substitutions. As before, $R(\mathbf{x})$ is a cost function similar to that of the last section, and depending on the problem at hand, can include various different terms which, e.g., may put soft constraints on the power of \mathbf{x} , or penalize sharp variations. Recall, as discussed above, that there is a close relationship between the optimal control setting Eq. (3) and maximum *a posteriori* (MAP) decoding in the GLM framework, as studied at moe length in Pillow *et al.* (2011).

Unlike that of Eq. (9), the Hessian of the likelihood term in Eq. (34), is not block triangular. However, due to the finite temporal length of the filter \mathbf{k} , the Hessian of this term is

still banded; the second derivative of the likelihood term in Eq. (34) with respect to \mathbf{x}_{t_1} and \mathbf{x}_{t_2} is zero when $|t_1 - t_2|$ is larger than twice the temporal length of \mathbf{k} . The set of linear equations at each Newton-Raphson iteration, Eq. (41), can be solved in $O(T)$ computational time, even when the equation matrix is not tridiagonal, as long as it is banded – see, e.g., Paninski *et al.* (2010) and Pillow *et al.* (2011) for further details. Thus, we retain the $O(T)$ scaling even in this case, as long as the cost function $R(\mathbf{x})$ also has a banded Hessian. This is the case for many commonly used cost functions, including the ones mentioned above.

As an example, in the case of controlling the spiking activity of a retinal ganglion cell, the filter \mathbf{k} is the spatiotemporal visual receptive field of the cell, and the stimulus \mathbf{x} can be a spatiotemporally varying contrast pattern on a monitor that is optically projected onto the retina. In this case, the sharp bound x_{\max} can represent the bound on the contrast that the monitor can produce. Figure 6 shows such an example of optimizing a visual stimulus for two simulated retinal neurons simultaneously. In this example, we modeled the cells using the GLM parameters fit to experimental recordings, as reported in Pillow *et al.* (2008). The stimulus had 9 pixels (arranged on a 3×3 square), covering the cells’ receptive fields center and most of the surround. We took the smooth cost function $R(\mathbf{I})$ in Eq. (34) to be zero, but we imposed the rigid bound $|\mathbf{x}| \leq x_{\max}$, for x_{\max} corresponding to a bound on the contrast of the stimulus. The optimized stimulus for the pixel covering the overlapping receptive field centers of both cells (marked by a red dot on the left side of top panel in the figure) shows a noticeable difference to that of the other pixels. Namely, unlike the optimized stimulus for the other pixels, it spends more time at intermediate values, away from the rigid bounds $\pm x_{\max}$. This is because spiking probability depends more sensitively on the value of the contrast in the center pixel compared to other pixels, as the receptive field filter is stronger in the center. Hence, the likelihood term in Eq. (34) is much more sharply concentrated as a function of the stimulus values for this pixel, compared to the other ones, and is thus able to confine the optimized \mathbf{x} away from $\pm x_{\max}$ for this dimension. By contrast, the log-likelihood is relatively flat along the dimensions corresponding to the other pixels, and thus the cost function of Eq. (34) tends to acquire its minimum on the boundaries $\pm x_{\max}$ for those dimensions (see Pillow *et al.* (2011) for further discussion of this effect).

2.5 Online implementation

For many practical applications, the above set up has to be implemented in an online manner. For example, in a sensory neuroprosthesis application where we need to artificially stimulate some sensory neurons to emit the appropriate spike trains in response to the sensory stimulus, the following sequence has to be performed repeatedly. As incoming sensory information keeps arriving, the putative response of the sensory neurons to that stimulus is simulated according to some encoding model for some time ahead; this constitutes the target spike train for the optimal control algorithm. Then the portion of the optimal artificial stimulus, \mathbf{x} , corresponding to this near future time-span has to be computed and “injected” in order to cause the sensory neurons to actually emit a spike train as close to the predicted one as possible. Another conceivable application is suppressing spike firing in neural circuits that exhibit rhythmic (i.e., predictable) spontaneous activity; if a rhythmic series of spikes is detected, one could potentially in real time compute and deliver the best stimulus to disrupt the unwanted rhythm. In both applications, the difference with the approach of the previous sections is that the earlier portions of the artificial stimulus are optimized earlier, and only with information about the target spike train a finite time into the future, rather than with full knowledge of the spike train for the entire span of the

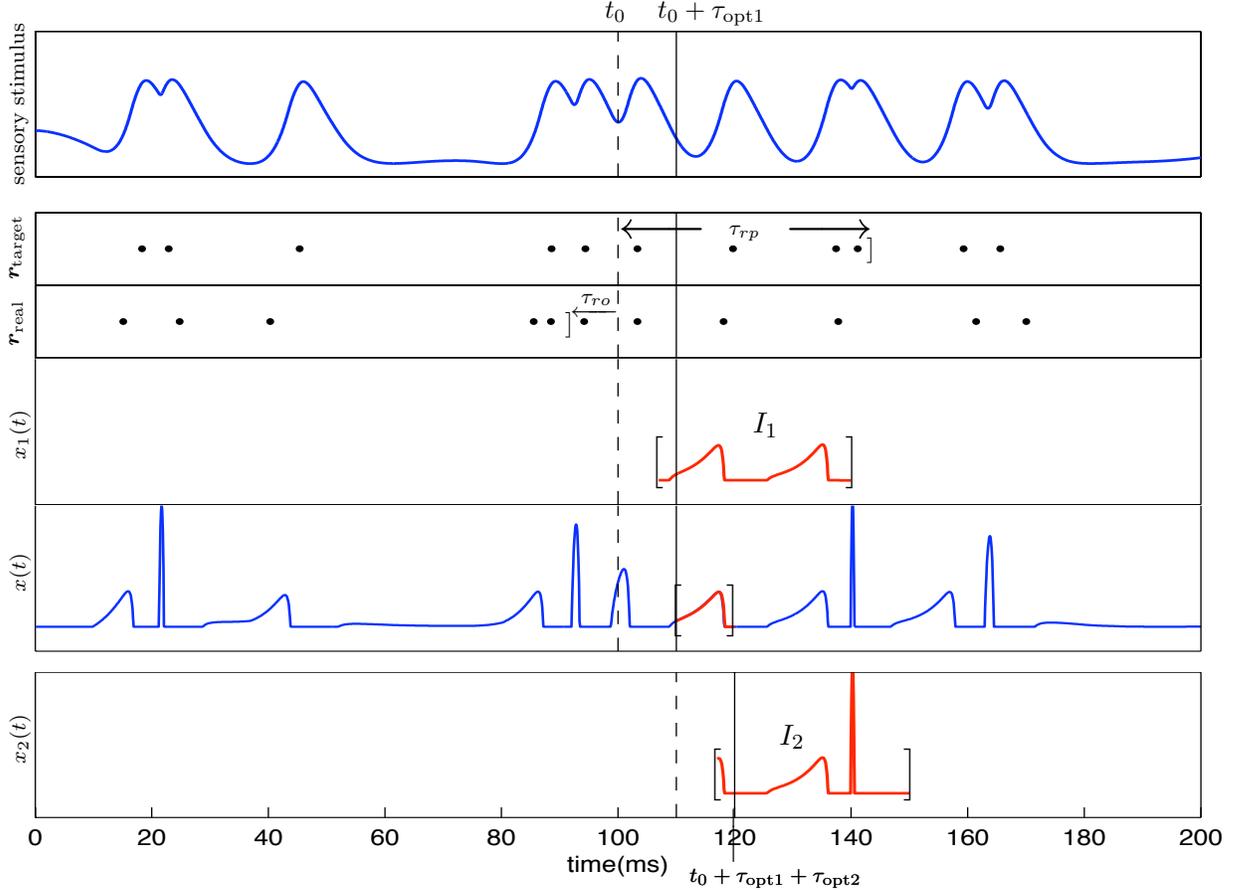


Figure 7: The sketch of different timings in the online implementation of optimal control. The top panel shows the sensory stimulus for all times. The second and the third panels (from top) show the predicted and emitted spike trains for all times. The fifth panel shows the actually injected stimulus for all times. The dashed (solid) vertical line crossing the top five panels shows the real time before the first optimization has started (ended). At the start of the first optimization, i.e. at $t = t_0$, the putative spike train response is predicted up to $t_0 + \tau_{rp}$ (show in second panel from top). Also, the feedback observation mechanism has updated the past spike times to set them to those of the actually emitted spikes before $t - \tau_{ro}$ (shown in the third panel from top). In the first optimization step, the optimization algorithm is carried out with this data for the interval I_1 . The fourth panel (from top) shows this interval and the obtained optimal solution $x_1(t)$ on it. When the algorithm finishes, we reach the vertical line (in the top five panels), at which point we inject the obtained solution. The actually injected part of this solution is shown as the red part of the curve in the fifth panel. At the same time, we start the second optimization process, with updated data, for the interval I_2 (bottom plot). By the time the second optimal solution, $x_2(t)$ is ready, we stop injecting the first solution $x_1(t)$ (therefore the red part of the curve in the fifth panel does not include the whole solution shown in the fourth panel), and use $x_2(t)$ instead.

application.

As a first solution, we will carry this out as follows. Suppose we are at real time $t = t_0$. We have sensed the sensory stimulus up to that time, and based on that observation (possibly taking into account correlations in the stimulus itself, which allow for its probabilistic prediction

further into the future) we are able to estimate a prediction for the spike train up to some later time $t_0 + \tau_{rp}$. Thus at this point, we have access to $r_{\text{target}}(t)$ for $t \in (-\infty, t_0 + \tau_{rp}]$. We can also assume that there is a feedback mechanism that has measured the actually emitted spike times before $t_0 - \tau_{ro}$ (for some fixed $\tau_{ro} > 0$, which is a characteristic of the observation mechanism), allowing us to update $r_{\text{target}}(t)$ for $t \in (-\infty, t_0 - \tau_{ro}]$ such that it agrees with the real (as opposed to predicted) spikes for those times.

At this point, we will run the optimization algorithm to find the artificial stimulus $x(t)$ for t in some sub-interval of $[t_0, t_0 + \tau_{rp}]$. Let us denote this sub-interval by $I_1 = [t_0 + \tau_i, t_0 + \tau_f]$, where

$$0 < \tau_i < \tau_f < \tau_{rp}. \quad (35)$$

Suppose that the optimization program takes τ_{opt1} to run, and at time $t_0 + \tau_{\text{opt1}}$ we have obtained the optimal solution $x_1(t)$ on I_1 . At this point we should start the optimization process anew, with the updated $\mathbf{r}_{\text{target}}$ (which is now predicted up to $t_0 + \tau_{\text{opt1}} + \tau_{rp}$) and for the interval $I_2 = [t_0 + \tau_{\text{opt1}} + \tau_i, t_0 + \tau_{\text{opt1}} + \tau_f]$, and simultaneously “inject” the optimized $x_1(t)$ during the time that the second optimization is running (taking τ_{opt2}), i.e., on $[t_0 + \tau_{\text{opt1}}, t_0 + \tau_{\text{opt1}} + \tau_{\text{opt2}}]$. Clearly, the latter part is only possible if I_1 covers $[t_0 + \tau_{\text{opt1}}, t_0 + \tau_{\text{opt1}} + \tau_{\text{opt2}}]$, hence we must have

$$\tau_i < \tau_{\text{opt1}}, \quad (36)$$

$$\tau_{\text{opt1}} + \tau_{\text{opt2}} < \tau_f < \tau_{rp}. \quad (37)$$

These are important bounds on τ_i and τ_f . Finally, at time $t_0 + \tau_{\text{opt1}} + \tau_{\text{opt2}}$ we will start injecting the new solution $x_2(t)$ (found on I_2), instead of the (possible) remainder of $x_1(t)$. This cycle is continued for the entire span of the application. The optimization times $\tau_{\text{opt}i}$ depend on the computer and amplifier hardware and details of software implementation and can vary from case to case. Figure 7 shows these various time intervals.

3 Experimental Validation

We tested the offline optimal current injection method of Sec. 2.2 in one simple setting using intracellular injections into patch-clamped pyramidal neurons in mouse cortical slices. All animal handling and experimentation was according to the National Institute of Health and local Institutional Animal Care and Use Committee guidelines. Coronal slices 350 μm thick were prepared from C57BL/6 mice at age P14 using a Leica VT1000-S vibratome. Pipette solution contained 130 mM K-methylsulfate, 2 mM MgCl_2 , 0.6 mM EGTA, 10 mM HEPES, 4 mM ATP-Mg, and 0.3 mM GTP-Tris, pH 7.2 (295 mOsm). All recordings were made at 33-36C from Layer 5 pyramidal neurons in the primary somatosensory cortex using the Multiclamp 700B amplifier (Molecular Devices, Sunnyvale, CA) digitized with a National Instruments 6259 multichannel card and recorded using custom software written using the LabView platform (National Instruments, Austin, TX).

To apply the method, we first need to estimate a model for $p(\mathbf{r}|\mathbf{I})$. To do this, we injected 5 minutes of white noise current (with a sampling rate of 1 kHz) into the cell in current clamp mode, and recorded the emitted spike times (extracted from the recorded voltage trace). Based on this data we fit the model parameters. In fitting the model to recorded data, we took a slightly different approach from that of Sec. 2.2. The model of Eq. (7), in the absence of noise (i.e., for $\sigma = 0$), can be solved explicitly for V_t . The result can be written as the convolution of

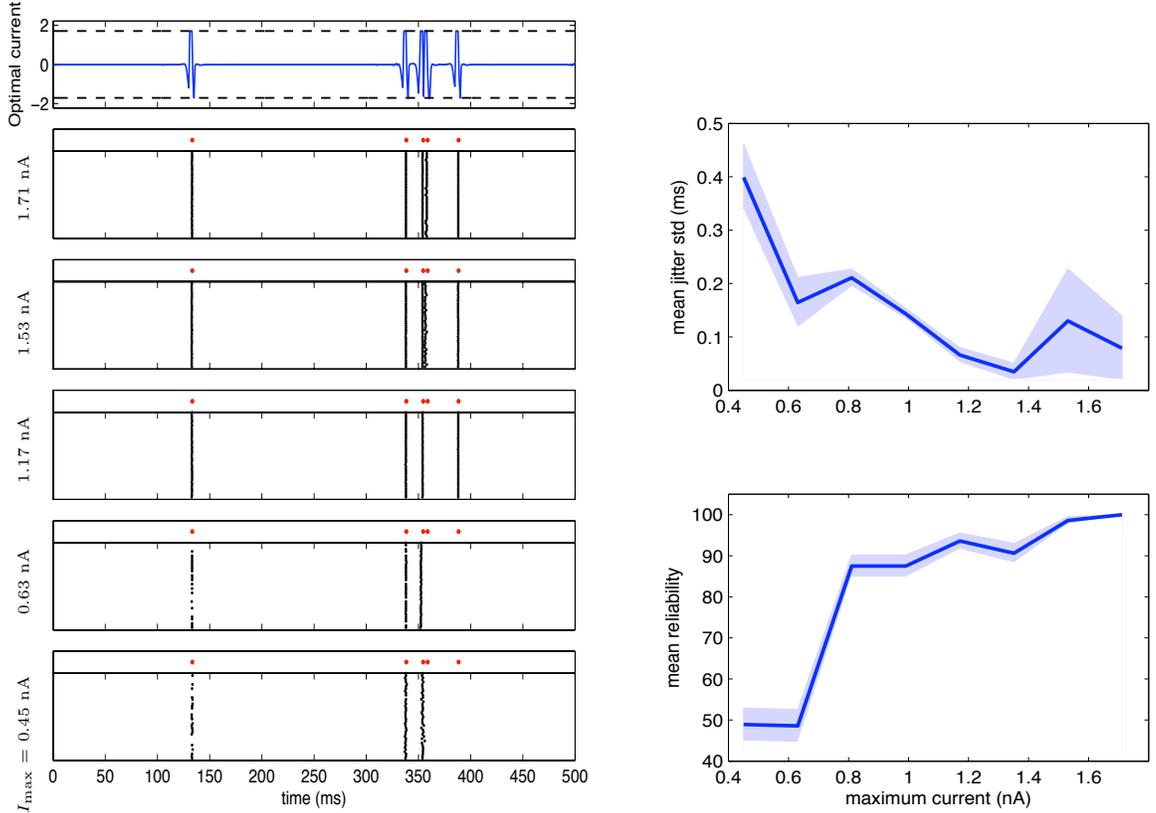


Figure 8: Experimental test of the current optimization method using *in vitro* intracellular injection into a mouse pyramidal neuron in layer 5 of the primary somatosensory cortex (in a cortical slice). Left figure shows five sets of rasters (each based on 40 trials) for five different values of the maximum current magnitude I_{\max} (see Eq. (9)). The top row of each raster panel shows the (same) target spike train (red spikes). The timing precision and reliability of the elicited spikes is visibly reduced as we constrain the range of current more (i.e., decrease I_{\max}). The plots on the right side show this more quantitatively based on averages over all the trials of all the tested target spike trains. The top and the bottom plots show the timing precision (standard deviation of the trial to trial timing jitter of each spike), and the reliability (percentage of trials in which the target spike was indeed elicited), respectively, as a function of the maximum allowed current. The method is able to achieve sub-millisecond precision and very high reliability with a biologically safe range of currents.

an exponential filter with the current: $V_t = K * I_t$, where $K(t) = \frac{1}{C} \exp(-t/\tau_V)$. This means that for $\sigma = 0$ this model can be rewritten in the GLM form of Eq. (33), where now the filter has the particular exponential form and is parametrized by τ_V and R . However, instead of fitting these two parameters to the recorded data, for computational purposes we parametrized the kernel as a weighted sum of exponential filters with different, yet known, time constants. I.e., we wrote $K(t) = \sum_i k_i \exp(-t/\tau_i)$ where we fixed the τ_i beforehand, choosing them to be in the range of 1-100 ms, and we maximized the model likelihood in terms of k_i to find the best fit.⁵ The advantage of this approach is that the parametrization of the log-likelihood in

⁵To estimate the spike history kernels discussed after Eq. (6) and in footnote 2, a similar parametrization in terms of exponential basis functions was also used. For more details about efficient ways of fitting GLM

terms of k_i is concave, whereas it need not be so in terms of τ_V , and therefore the likelihood optimization for fitting could be done more rapidly and reliably. With this parametrization, and limiting ourselves to the $\sigma = 0$ case (justified because intracellular noise is small), we used the GLM framework of Eq. (34) (with \mathbf{x} replaced by I_t and \mathbf{k} replaced by the estimated filter) to carry out the optimal control as well. We then ran our optimization program for six different target spike train patterns and for nine different values of the current magnitude bound I_{\max} (see Eq. (9)), in order to study the effect of this constraint on the achieved spike timing precisions and reliabilities. Then we injected these optimized currents in a series of 40 trials (separated by 0.5 second long silent intervals) to measure the performance. The top panel of Fig. 8 shows the optimized current for the highest value of I_{\max} we used. The curve is qualitatively similar to Fig. 1, and differences correspond to differences in model parameters and the penalty strengths. The dependence on I_{\max} (not shown in Fig. 1) is also similar to Fig. 2; as I_{\max} decreases, the peaks of the optimal current broaden, leading to lower spike timing precision.

Because the timing precision was high, association of trial spikes to those in the target spike train was unambiguous. We then defined reliability (success rate for eliciting a desired spike) as the percentage of the trials in which the corresponding spike was indeed elicited within a 3 ms window around the target spike train. The timing precision was defined as the standard deviation, among the trials in which the corresponding trial spike was emitted, of the timing difference of the corresponding trial spike to that in the target spike train. The resulting rasters for one of the spike patterns are shown in Fig. 8. A noteworthy aspect of the results is the fact that spikes trailing other spikes within a short time window are the first that fail to be elicited as we lower I_{\max} . Due to refractory and short-term adaptation effects, the cell is less excitable for a few tens of milliseconds after firing a few spikes, requiring a high current to be driven to spiking, which may not be feasible if I_{\max} is too low. The right panel in Fig. 8 shows the dependence of the obtained timing precision and reliability, averaged over all spikes in all target patterns, on the bound I_{\max} . For this simplest of settings the method is capable of achieving sub-millisecond timing precision and very high reliability with biologically safe input levels. The data collection for fitting the model, and the likelihood optimizations for the fit and for optimal control overall do not take more than 2-3 minutes, thus allowing the method to be used as a tool in electrophysiology experiments.

4 Discussion

In this manuscript we provide a general formulation for the problem of optimal control of neural spiking. We tailored this general framework for a few specific cases of interest involving electric, optical, or sensory stimulation of neurons using relatively simple models which describe the input-output mechanism of the neuron. Importantly, we showed that the method based on these models is amenable to fast convex optimization methods. This property is key for online implementation of these methods (discussed in Sec. 2.5) which is necessary in most applications. In addition, convexity guarantees that the optimizer will not be caught in a non-global, local optimum; it will always find the globally optimal solution. We also noted the close relationship between our optimal control framework and the problem of maximum *a posteriori* decoding of spike trains. In many cases of interest, the optimal stimuli we obtain using our methods are quite non-intuitive (see, e.g., the examples in Figs. 3, 4 and 6). Even when the optimal stimuli have an intuitive form, the dependence of important details on the model choice and parameters could

parameters see Paninski (2004); Pillow *et al.* (2008).

be complicated; cf. the discussion of Fig. 5. It is unlikely that in such cases, simple, by-hand approaches for designing stimuli can perform as satisfactorily as a systematic method taking into account the detailed properties of the neurons and the stimulating agents. Furthermore, the example of Fig. 5 shows that when the employed neuronal models miss important qualitative aspects of the biophysics, the stimulus optimized according to them can in fact be considerably suboptimal. At the same time, our investigations show that the optimal stimulus found according to the models used in this paper are reasonably robust with respect to reasonable errors in model parameters; as long as the parameters used for optimizing the stimulus do not deviate by a very large amount from the true parameters, the performance of the resulting stimuli is not noticeably compromised. The conclusion we can draw is that as long as the employed model captures the qualitative biophysics correctly, the performance of the resulting optimized stimuli is in most cases quite robust with respect to reasonable quantitative errors in model parameters.

There are several directions for improvement and extension of our framework and the specific models it relies on, which we intend to address in future work. The model considered in Sec. 2.3 for the interaction of the laser with cellular pumps and channel conductances may prove to be too simplistic in certain applications, as it ignores saturation effects. These can be accounted for by adding appropriate nonlinear corrections to Eqs. (23)–(24). The drawback, however, is that these add new sources of non-convexity to the cost function and hence reduce the efficiency and accuracy of the optimization algorithm (for preliminary work on enhancing the modeling of photo-sensitive channels along these lines without severely compromising the efficiency of the optimization, see Grosenick *et al.* (2011)). Also, in the case of uncaging of caged neurotransmitters (discussed at the end of Sec. 2.3) a more accurate model should account for relevant spatial effects in the diffusion of uncaged compounds – ignored by the simple, non-spatial models described by Eqs. (21)–(23) and Eqs. (31)–(32). We also plan to improve the online implementation discussed in Sec. 2.5 so as to minimize redundancies. We noted in Sec. 3, that due to refractory and short-term adaptation effects, spikes that trail other spikes by less than about 50 ms, are the least robustly reproducible ones (this is also visible, in a different setting, in the comparison made in Fig. 5 – albeit for shorter inter-spike intervals). Another direction for future research is to investigate whether improving the employed neuronal models such that they account better for short-term adaptation could make the control of such spikes more robust.

In addition to the numerical methods that we considered here, the stimulus optimization problem can also be studied by adopting an analytical, continuous time approach. In this approach, the optimal solution is found by solving the Euler-Lagrange equations corresponding to the minimized cost function. As we mentioned in Sec. 2.2, in Moehlis *et al.* (2006), the authors adopted such an approach to find analytical optimal solutions for input currents to a certain class of neural oscillator models (similar methods have been applied to other related problems in theoretical neuroscience – see, e.g., Badel *et al.* (2005); Paninski (2006); Badel *et al.* (2008); Cocco *et al.* (2009)). Such approaches to stimulus optimization are quite fast by virtue of being analytic, but they only apply to rather specific problem settings. Our numerical approach, on the other hand, is much more general. However, there could be ways of combining the two methods to gain from the advantages of both. In our future work, we intend to use the mentioned analytic method to study the stimulus optimization in simple settings, such as the case of a target spike train containing a single, isolated spike. It may then be possible, in certain cases, to obtain a satisfactory, approximately optimal solution for a general target spike train, by a simple superposition of the analytically obtained solution for single spikes. Even if not very close to the optimal solution, this can provide a very good initialization for the numerical

optimization algorithm, and hence improve its speed.

Many interesting applications of optimal spiking control could be carried out *in vivo*. This creates a host of complications. For example, in the case of *in vivo* extracellular current injections, the relative position of the electrodes to various controlled neurons, as well as the intrinsic biophysics of those cells, are all subject to time-dependent variations. This means that model parameters (electrode couplings to various cells, membrane time constants, etc) change with time, and therefore a single estimation of these before the main application (as we did in our experimental test in Sec. 3) will not be sufficient. Furthermore, in many situations we may not be able to observe the spiking output our stimuli are creating (or we might only have crude information, such as behavioral feedback in a sensory prosthetic setting), so we might have to use a crude encoding model. It will thus be crucial to explore ways in which we can make our methods more robust to model mis-specification in the future. As a first step towards this goal, we plan to test a simple approach which is based on a small modification of the framework of Eq. (3), used in this paper. The spirit of this modification is similar to the method described in footnote 1 of Sec. 2.1, for generalizing beyond the sharp delta function cost function (see the discussion after Eq. (1)) without losing computational tractability. Let us show the model parameters collectively by θ . It is implicit in Eq. (3) that the distribution $p(\mathbf{r}|\mathbf{x})$ is conditioned on a single set of model parameters, $\hat{\theta}$, estimated from the relevant observed data; more explicitly, the distribution used in Eq. (3) is $p(\mathbf{r}|\mathbf{x}, \hat{\theta})$. Suppose we represent our uncertainty about the model parameters by a posterior distribution conditional on the observed data, and we have n samples, θ_i ($i = 1, \dots, n$), from this distribution. Now, to take into account the model mis-specification, we modify the minimized objective in Eq. (3) replace the overly confident $\log p(\mathbf{r}|\mathbf{x}, \hat{\theta})$ in Eq. (3) by its average, $1/n \sum_{i=1}^n \log p(\mathbf{r}|\mathbf{x}, \theta_i)$, over the samples from the parameter posterior. The important point is that with this modification of Eq. (3), the minimized objective (as a function of the stimulating agent, \mathbf{x}) is still convex, and therefore we can still use fast convex optimization methods for minimizing it.

Finally, we note that one could test our algorithms in other experimental settings including extracellular current injection and optical uncaging of caged glutamate and GABA compounds, to control the spiking activity of several neurons. The experimental test presented in this paper is only a simple first step; spike control using the above methods is more complicated than the intracellular current injection presented in Sec. 3, and is of more interest from the point of possible applications.

Acknowledgements

YA is supported by the Robert Leet and Clara Guthrie Patterson Trust Postdoctoral Fellowship, Bank of America, Trustee. RY’s laboratory is supported by NIH EY11787 and the Kavli Institute for Brain Studies. LP and RY share a CRCNS award, NSF IIS-0904353. LP is in addition supported by an NSF CAREER award and a McKnight Scholar award.

Appendix A: derivation of the likelihood expressions

In this appendix we derive the formula Eq. (8) for the log-likelihood $\log p(\mathbf{V}|\mathbf{I})$. First, we rewrite Eq. (7) in discrete time (with step size dt) as

$$V_{t+dt} = V_t - \frac{dt}{\tau_v} V_t + \frac{dt}{C} I_t + \sqrt{dt} \sigma \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, 1), \quad (38)$$

where $\mathcal{N}(0, 1)$ denotes the distribution $e^{-\frac{1}{2}\epsilon_t^2}/\sqrt{2\pi}$, i.e., the standard Gaussian distribution with mean 0 and variance 1. The Gaussian noise ϵ_t is scaled by \sqrt{dt} in order to ensure a proper continuous time limit. ϵ_t 's at different time steps are sampled independently from $\mathcal{N}(0, 1)$. Since the voltage is determined by the current, up to the Gaussian noise ϵ , which is integrated linearly, the joint conditional probability distribution $p(\mathbf{V}|\mathbf{I})$ takes the Gaussian form. In fact, the joint probability distribution of \mathbf{V} for all times is, up to normalization, given by the joint distribution of ϵ_t 's at all times. Since the latter are independent at different times, their joint distribution is simply the product of their distributions at each time step. We thus have

$$p(\mathbf{V}|\mathbf{I}) \propto \prod_t e^{-\frac{1}{2}\epsilon_t^2}, \quad (39)$$

where \propto means equality up to a constant multiplicative factor. Using Eq. (38) to solve for ϵ_t in terms of \mathbf{V} and \mathbf{I} , we obtain

$$p(\mathbf{V}|\mathbf{I}) \propto \prod_t e^{-\frac{1}{2\sigma^2 dt} \left(V_{t+dt} - V_t + \frac{dt}{\tau_V} V_t - \frac{dt}{C} I_t \right)^2}. \quad (40)$$

Taking the logarithm of both sides we obtain Eq. (8).

Appendix B: $O(T)$ optimization

Here we discuss in detail the fast optimization methods referred to in Sec. 2.2. As mentioned at the end of that section, we use the approximation (13) in carrying out the optimization in (3) (with \mathbf{I} replacing \mathbf{x}). We thus need to minimize the convex cost function of Eq. (9) in the joint state vector (\mathbf{V}, \mathbf{I}) . Let us schematically denote the optimized variables (\mathbf{V}, \mathbf{I}) by the vector \mathbf{x} and denote the function minimized in Eq. (9) by $F(\mathbf{x})$. At every step of the Newton-Raphson algorithm, when the optimized variable has the current value \mathbf{x} , one uses a second order Taylor expansion of $F(\mathbf{x})$ around \mathbf{x} to approximate this function by a quadratic function. Then one moves the optimized variable to the point, $\mathbf{x} + \mathbf{y}$, that minimizes this quadratic function. Minimizing the latter yields (see, e.g., Press *et al.* (1992)) the following set of linear equations for the displacement, \mathbf{y} :

$$H|_{\mathbf{x}} \mathbf{y} = -\nabla F|_{\mathbf{x}}, \quad (41)$$

where $H|_{\mathbf{x}}$ and $\nabla F|_{\mathbf{x}}$ are, respectively, the Hessian matrix of $F(\mathbf{x})$ (the matrix of its second order partial derivatives with respect to the variables (\mathbf{V}_t, I_t)) and its gradient at the current point, \mathbf{x} . To summarize, at each step of the Newton-Raphson algorithm one moves the current value \mathbf{x} , by an amount \mathbf{y} which is found by solving Eq. (41) (in practice, to guarantee convergence, we move the state vector \mathbf{x} by $\alpha\mathbf{y}$, $0 < \alpha \leq 1$, where α is found by a simple, stable backtracking procedure (Press *et al.*, 1992)). This step is repeated until we converge to some local minimum. Since the objective is convex, this minimum is guaranteed to be the global minimum.

In general, solving a set of d linear equations such as Eq. (41) requires $O(d^3)$ elementary equations, and hence the computational time spent in every Newton-Raphson step should scale cubically with the duration, T , of the the interval in which we intend to control the spike train - since in our case the dimension of \mathbf{x} is proportional to the number of time steps. However, in our case the Hessian matrix of the first two log-likelihood terms in the objective Eq. (9) is

block tri-diagonal with blocks of size 2×2 (one dimension each for V_t and I_t).⁶ This is because due to locality in time (the fact that each summand of the temporal sum in Eqs. (8)–(6) only depends on state variables at the time step t or adjacent time steps), the partial second derivate of each log-likelihood term with respect to a variable (\mathbf{V} or \mathbf{I}) at time t_1 and a second variable at time t_2 is zero, unless $t_2 = t_1$, or $t_2 = t_1 \pm 1$. When the matrix of the linear set of equations is block-tridiagonal (or, more generally, banded with a fixed band-width), the scaling of the computational time is improved to $O(d)$. In our case, this means that the computational time scales as $O(T)$, i.e., only linearly with the duration of the control interval (for a review of the application of fast optimization methods to problems in neuroscience, see Paninski *et al.* (2010)).

The only complication for using the Newton-Raphson method is that we have imposed a hard constraint on the current. This is because the Newton-Raphson method, Eq. (41), assumes a smooth objective function. To overcome this, we use the barrier or the interior-point method (Boyd and Vandenberghe, 2004; Paninski *et al.*, 2010), in which instead of directly enforcing the hard constraint, $|I_t| < I_{\max}$, we add a soft logarithmic barrier function of the form $\eta \sum_t [\log(I_{\max} - I_t) + \log(I_{\max} + I_t)]$ to the objective, and optimize the latter in this augmented form, but without the explicit bound $|I_t| < I_{\max}$ enforced. By iterating this procedure for progressively sharper barrier functions (corresponding to smaller η) we approach the solution to the hard constraint problem with any desired accuracy. Since the Hessian of the barrier function turns out to be diagonal, the total Hessian of the objective with respect to (\mathbf{V}, \mathbf{I}) remains block-tridiagonal, and we can still apply fast inversion algorithms in each iteration and hence the computational cost still scales as $O(T)$.

Another small complication arises because the Hessian of the cost function Eq. (10) with respect to I_t is not banded. However, from Eq. (11) we can write $\mathbf{I} = A\mathbf{J}$, where A is a bi-diagonal Toeplitz matrix with zero entries everywhere except on the main diagonal (where they are all equal to $1/\tau_J - 1/dt$) and the first diagonal above it (where they are all equal to $1/dt$). Given that \mathbf{J} and \mathbf{I} are related by a deterministic linear transformation, we can change variables and carry out the minimization in Eq. (9) in terms of (\mathbf{V}, \mathbf{J}) instead of (\mathbf{V}, \mathbf{I}) . Once the optimal (\mathbf{V}, \mathbf{J}) are found, we can multiply the optimized \mathbf{J} by A to obtain the desired optimal solution for \mathbf{I} . But the Hessians of all three terms in the objective function in Eq. (9) with respect to (\mathbf{V}, \mathbf{J}) are banded, and thus the minimization in terms of the new variables can be done in $O(T)$ fashion. To see the bandedness of the corresponding Hessians note that the Hessian of the cost function term Eq. (10) with respect to \mathbf{J} is proportional to the identity matrix. The bandedness of the Hessians of the first two terms in Eq. (9) is also preserved after the variable change (note that they were block-tridiagonal with respect to (\mathbf{V}, \mathbf{I})). This is because, by the chain rule for derivatives, these Hessians are only modified by multiplications by A^T on the left side or by A on the right side or both (depending on the number and order of derivatives with respect to \mathbf{I} in the original variables), and multiplication by banded matrices preserves bandedness.

The situation is quite similar in the optimization for optical stimulation, Eq. (25). Here, the natural choice for the optimized variables would have been $(\mathbf{V}, \mathbf{L}^E, \mathbf{L}^I)$, and not $(\mathbf{V}, \mathbf{g}^E, \mathbf{I}^I)$. But the Hessian of the second term in the cost function is not banded with respect to $(\mathbf{L}^E, \mathbf{L}^I)$, whereas it is clearly block tri-diagonal with respect to $(\mathbf{g}^E, \mathbf{I}^I)$. Furthermore, if we define $\tilde{\mathbf{L}}^E = w_{EI}\mathbf{L}^I + w_{EE}\mathbf{L}^E$, and $\tilde{\mathbf{L}}^I = w_{II}\mathbf{L}^I + w_{IE}\mathbf{L}^E$, then Eqs. (23)–(24) take the same form as Eq. (11)

⁶Accordingly, for the case of the oscillatory neuron considered at the end of Sec. 2.2, the Hessian matrices of the first two terms in Eq. (17) with respect to (\mathbf{W}, \mathbf{I}) are block-triagonal with blocks of size 3×3 instead. This is because the state vector \mathbf{W}_t has two components per time step. But except for the change in block size, everything else is similar to case of optimization in (\mathbf{V}, \mathbf{I}) , and the discussion of the $O(T)$ optimization presented here generalizes to the oscillatory dynamics model, Eq. (15), as well.

(after discretization). Therefore, we have $\tilde{\mathbf{L}}^E = A^E \mathbf{g}^E$ and $\tilde{\mathbf{L}}^I = A^I \mathbf{I}^I$, where the bi-diagonal Toeplitz matrices A^E and A^I have the same form as the matrix A defined in the last paragraph, with τ_E and τ_I replacing τ_J , respectively. Using the chain rule for differentiation, we see that by the same argument as in the last paragraph, the Hessian of the third term $R(\mathbf{L}^E, \mathbf{L}^I)$ in the cost function Eq. (25) is also banded with respect to $(\mathbf{g}^E, \mathbf{I}^I)$. Thus the optimization in Eq. (25) can be carried out using the $O(T)$ method.

Finally, we found that in many cases where the Hessian matrix, $H|_{\mathbf{x}}$, is very ill-conditioned, we could improve the convergence of the Newton-Raphson algorithm by regularizing the Hessian matrix, as follows. Instead of solving Eq. (41) at each step of the algorithm, we solved

$$(H|_{\mathbf{x}} + \eta^2 I) \mathbf{y} = -\nabla F|_{\mathbf{x}}, \quad (42)$$

for a small η , where I denotes the identity matrix. Mathematically, convergence according to this rule (i.e., reaching a point \mathbf{x} where the displacement found by solving Eq. (42) once more is too small) still requires reaching a point where $\nabla F|_{\mathbf{x}}$ vanishes (or is very small). But convexity and smoothness guarantee that a stationary point with zero gradient is indeed the global minimum. As usual, Eq. (42) can be solved efficiently, since if H is banded then $H + \eta^2 I$ is also banded.

References

- Arenkiel, B. R., Peca, J., Davison, I. G., Feliciano, C., Deisseroth, K., Augustine, G. J., Ehlers, M. D., and Feng, G. (2007). In vivo light-induced activation of neural circuitry in transgenic mice expressing channelrhodopsin-2. *Neuron*, **54**(2), 205 – 218.
- Badel, L., Richardson, M., and Gerstner, W. (2005). Dependence of the spike-triggered average voltage on membrane response properties. *Neurocomputing*, **69**, 1062–1065.
- Badel, L., Gerstner, W., and Richardson, J. E. (2008). Spike-triggered averages for passive and resonant neurons receiving filtered excitatory and inhibitory synaptic drive. *Physical Review E*, **78**(1).
- Berger, J. (1993). *Statistical Decision Theory and Bayesian Analysis*. Springer.
- Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Oxford University Press.
- Boyden, E., Zhang, F., Bamberg, E., Nagel, G., and Deisseroth, K. (2005). Millisecond-timescale, genetically targeted optical control of neural activity. *Nat. Neurosci.*, **8**, 1263–1268.
- Brillinger, D. (1988). Maximum likelihood analysis of spike trains of interacting nerve cells. *Biological Cybernetics*, **59**, 189–200.
- Callaway, E. M. and Yuste, R. (2002). Stimulating neurons with light. *Current Opinion in Neurobiology*, **12**(5), 587 – 592.
- Cocco, S., Leibler, S., and Monasson, R. (2009). Neuronal couplings between retinal ganglion cells inferred by efficient inverse statistical physics methods. *Proceedings of the National Academy of Sciences*, **106**(33), 14058–14062.
- Dickson, C. T., Magistretti, J., Shalinsky, M. H., Fransén, E., Hasselmo, M. E., and Alonso, A. (2000). Properties and role of I_h in the pacing of subthreshold oscillations in entorhinal cortex layer ii neurons. *Journal of Neurophysiology*, **83**(5), 2562–2579.

- Ditterich, J., Mazurek, M., and Shadlen, M. (2003). Microstimulation of visual cortex affects the speed of perceptual decisions. *Nature Neuroscience*, **6**(8), 891–898.
- Feng, X.-J., Shea-Brown, E., Greenwald, B., Kosut, R., and Rabitz, H. (2007a). Optimal deep brain stimulation of the subthalamic nucleus – a computational study. *Journal of Computational Neuroscience*, **23**, 265–282.
- Feng, X.-J., Shea-Brown, E., Greenwald, B., Kosut, R., and Rabitz, H. (2007b). Toward closed-loop optimization of deep brain stimulation for Parkinson’s disease: Concepts and lessons from a computational model. *Journal of Neuroengineering*, **4**, L14–L21.
- Gerstner, W. and Kistler, W. (2002). *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge University Press.
- Gradinaru, V., Zhang, F., Ramakrishnan, C., Mattis, J., Prakash, R., Diester, I., Goshen, I., Thompson, K. R., and Deisseroth, K. (2010). Molecular and cellular approaches for diversifying and extending optogenetics. *Cell*, **141**, 154–165.
- Grosenick, L., Mattis, J., Broxton, M., Yizhar, O., Levoy, M., and Deisseroth, K. (2011). Towards closed-loop system identification and control of neural circuits with optogenetics. *COSYNE 11 Workshops*.
- Gunaydin, L., Yizhar, O., Berndt, A., Sohal, V., Deisseroth, K., and Hegemann, P. (2010). Ultrafast optogenetic control. *Nat. Neurosci.*, **13**(3), 387–392.
- Gutfreund, Y., Yarom, Y., and Segev, I. (1995). Subthreshold oscillations and resonant frequency in guinea-pig cortical neurons: physiology and modelling. *The Journal of Physiology*, **483**(Pt 3), 621–640.
- Han, X. and Boyden, E. (2007). Multiple-color optical activation, silencing, and desynchronization of neural activity, with single-spike temporal resolution. *PLoS ONE*, **2**(3), e299.
- Histed, M. H., Bonin, V., and Reid, R. C. (2009). Direct activation of sparse, distributed populations of cortical neurons by electrical microstimulation. *Neuron*, **63**, 508–522.
- Hutcheon, B., Miura, R. M., and Puil, E. (1996). Subthreshold membrane resonance in neocortical neurons. *Journal of Neurophysiology*, **76**(2), 683–697.
- Izhikevich, E. (2001). Resonate-and-fire neurons. *Neural Networks*, **14**, 883–894.
- Jahnsen, H. and Karnup, S. (1994). A spectral analysis of the integration of artificial synaptic potentials in mammalian central neurons. *Brain Res*, **666**(1), 9–20.
- Jolivet, R., Lewis, T., and Gerstner, W. (2003). The spike response model: a framework to predict neuronal spike trains. *Springer Lecture notes in computer science*, **2714**, 846–853.
- Kass, R., Tierney, L., and Raftery, A. (1991). Laplace’s method in Bayesian analysis. In N. Flournoy and R. Tsutakawa, editors, *Statistical Multiple Integration*, pages 89–99, Providence. Springer.
- Koch, C. (1999). *Biophysics of Computation*. Oxford University Press.

- Leung, L. S. and Yu, H.-W. (1998). Theta-frequency resonance in hippocampal cal neurons in vitro demonstrated by sinusoidal current injection. *Journal of Neurophysiology*, **79**(3), 1592–1596.
- Loizou, P. (1998). Mimicking the human ear: an introduction to cochlear implants. *IEEE Signal Processing Magazine*, **15**, 101–130.
- Mainen, Z. and Sejnowski, T. (1995). Reliability of spike timing in neocortical neurons. *Science*, **268**, 1503–1506.
- Matsuzaki, M., Ellis-Davies, G. C. R., and Kasai, H. (2008). Three-Dimensional Mapping of Unitary Synaptic Connections by Two-Photon Macro Photolysis of Caged Glutamate. *J Neurophysiol*, **99**(3), 1535–1544.
- Mattingley, J. and Boyd, S. (2010). Real-time convex optimization in signal processing. *IEEE Signal Processing Magazine*., **under review**.
- McCullagh, P. and Nelder, J. (1989). *Generalized linear models*. Chapman and Hall, London.
- Moehlis, J., Shea-Brown, E., and Rabitz, H. (2006). Optimal inputs for phase models of spiking neurons. *Journal of Computational and Nonlinear Dynamics*, **1**(4), 358–367.
- Mohanty, S. K., Reinscheid, R. K., Liu, X., Okamura, N., Krasieva, T. B., and Berns, M. W. (2008). In-depth activation of channelrhodopsin 2-sensitized excitable cells with high spatial resolution using two-photon excitation with a near-infrared laser microbeam. *Biophys. J.*, **95**(8), 3916–3926.
- Nikolenko, V., Poskanzer, K., and Yuste, R. (2007). Two-photon photostimulation and imaging of neural circuits. *Nature Methods*, **4**, 943–950.
- Nikolenko, V., Watson, B., Araya, R., Woodruff, A., Peterka, D., and Yuste, R. (2008). Slm microscopy: scanless two-photon imaging and photostimulation using spatial light modulators. *Frontiers in Neural Circuits*, **2**, 5.
- Paninski, L. (2004). Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems*, **15**, 243–262.
- Paninski, L. (2006). The most likely voltage path and large deviations approximations for integrate-and-fire neurons. *Journal of Computational Neuroscience*, **21**, 71–87.
- Paninski, L., Ahmadian, Y., Ferreira, D., Koyama, S., Rahnama Rad, K., Vidne, M., Vogelstein, J., and Wu, W. (2010). A new look at state-space models for neural data. *Journal of Computational Neuroscience*, **29**(1), 107–126.
- Petreaanu, L., Huber, D., Sobczyk, A., and Svoboda, K. (2007). Channelrhodopsin-2-assisted circuit mapping of long-range callosal projections. *Nat Neurosci*, **10**(5), 663–668.
- Pike, F. G., Goddard, R. S., Suckling, J. M., Ganter, P., Kasthuri, N., and Paulsen, O. (2000). Distinct frequency preferences of different types of rat hippocampal neurones in response to oscillatory input currents. *The Journal of Physiology*, **529**(1), 205–213.
- Pillow, J., Shlens, J., Paninski, L., Sher, A., Litke, A., Chichilnisky, E., and Simoncelli, E. (2008). Spatiotemporal correlations and visual signaling in a complete neuronal population. *Nature*.

- Pillow, J. W., Ahmadian, Y., and Paninski, L. (2011). Model-based decoding, information estimation, and change-point detection techniques for multineuron spike trains. *Neural Comput.*, **23**(1), 1–45.
- Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (1992). *Numerical recipes in C*. Cambridge University Press.
- Puil, E., Meiri, H., and Yarom, Y. (1994). Resonant behavior and frequency preferences of thalamic neurons. *Journal of Neurophysiology*, **71**(2), 575–582.
- Rinzel, J. and Ermentrout, G. B. (1989). *Analysis of neural excitability and oscillations*, pages 135–169. MIT Press.
- Salzman, C., Britten, K., and Newsome, W. (1990). Cortical microstimulation influences perceptual judgments of motion direction. *Nature*, **346**, 174–177.
- Sekirnjak, C., Hottowy, P., Sher, A., Dabrowski, W., Litke, A. M., and Chichilnisky, E. J. (2006). Electrical stimulation of mammalian retinal ganglion cells with multielectrode arrays. *J Neurophysiol*, **95**(6), 3311–3327.
- Snyder, D. and Miller, M. (1991). *Random Point Processes in Time and Space*. Springer-Verlag.
- Tateno, T., Harsch, A., and Robinson, H. P. C. (2004). Threshold firing frequency-current relationships of neurons in rat somatosensory cortex: Type 1 and type 2 dynamics. *Journal of Neurophysiology*, **92**(4), 2283–2294.
- Todorov, E. (9-11 Dec. 2008). General duality between optimal control and estimation. *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, pages 4286–4292.
- Toyoizumi, T., Rahnama Rad, K., and Paninski, L. (2009). Mean-field approximations for coupled populations of generalized linear model spiking neurons with markov refractoriness. *Neural Comput.*, **21**(5), 1203–1243.
- Truccolo, W., Eden, U., Fellows, M., Donoghue, J., and Brown, E. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble and extrinsic covariate effects. *Journal of Neurophysiology*, **93**, 1074–1089.
- Wang, H., Peca, J., Matsuzaki, M., Matsuzaki, K., Noguchi, J., Qiu, L., Wang, D., Zhang, F., Boyden, E., Deisseroth, K., Kasai, H., Hall, W. C., Feng, G., and Augustine, G. J. (2007). High-speed mapping of synaptic connectivity using photostimulation in channelrhodopsin-2 transgenic mice. *Proceedings of the National Academy of Sciences*, **104**(19), 8143–8148.
- Weiland, J., Liu, W., and Humayun, M. (2005). Retinal prosthesis. *Annual Review of Biomedical Engineering*, **7**, 361–401.