Efficient characterization of electrically evoked responses for neural interfaces

Nishal P. Shah
Department of Electrical Engineering
Stanford University
Stanford, CA 94301
nishalps@stanford.edu

Abstract

Future neural interfaces will read and write population neural activity at cellular resolution for diverse applications. For example, an artificial retina may restore vision to the blind by electrically stimulating retinal ganglion cells. Such devices must tune their function, based on stimulating and recording, to match the function of the circuit. However, existing methods for characterizing the neural interface scale poorly with the number of electrodes, limiting their practical applicability.

We test the idea that using prior information from previous experiments and closed-loop measurements may greatly increase the efficiency of characterizing the neural interface. Large-scale, high-density electrical recording and stimulation in primate retina were used as a lab prototype for an artificial retina. Three key calibration steps are optimized: spike sorting, response modeling, and adaptive stimulation. For spike sorting, exploiting the similarity of electrical artifact across electrodes and experiments substantially reduced the number of required measurements. For response modeling, a joint model that captures the inverse relationship between recorded spike amplitude and electrical stimulation threshold from previously recorded retinas resulted in greater consistency and efficiency. For adaptive stimulation, choosing which electrodes to stimulate next based on probability estimates from previous measurements also improved performance. Similar improvement in efficiency resulted from using either non-adaptive stimulation with a joint estimation model, or adaptive stimulation with a simpler independent model for each cell. Finally, image reconstruction revealed that these improvements may translate to improved performance of an artificial retina.

1 Introduction

Recent advances in large-scale electrical and optical recording have made it possible to record and stimulate neural circuits at unprecedented scale and resolution [Kipke et al., 2008, Kerr and Denk, 2008]. These advances suggest the possibility of using electronic devices to restore functions lost to disease or augment human capacities [Wilson et al., 1991, Schwartz, 2004]. One such application is an artificial retina, which can provide a treatment for incurable blindness by electrically stimulating retinal ganglion cells, the output neurons of the retina [Stingl et al., 2013, Humayun et al., 2012, Lorach et al., 2015]. A high-fidelity device must encode a visual scene by electrically stimulating retinal neurons in a way that produces accurate and useful visual perception (Figure 1A).

However, to achieve this goal, the device must match the precise, asynchronous patterns of activity transmitted by multiple ganglion cell types to the brain. This will require first identifying the location and type of individual ganglion cells in the patient's retina, then characterizing their responses to electrical stimulation. Previously, it has been shown that the location and types of individual retinal ganglion cells can be identified using recorded activity [Richard et al., 2015]. However, efficient characterization of electrical responses remains unsolved. Neural responses to electrical

Preprint. Under review.

stimulation are noisy and nonlinear, therefore in a device with thousands of electrodes, a complete probabilistic characterization of the neural interface is infeasible. We propose new methods to efficiently characterize the electrical stimulation properties of the retinal interface, in a manner that may extend to other neural systems.

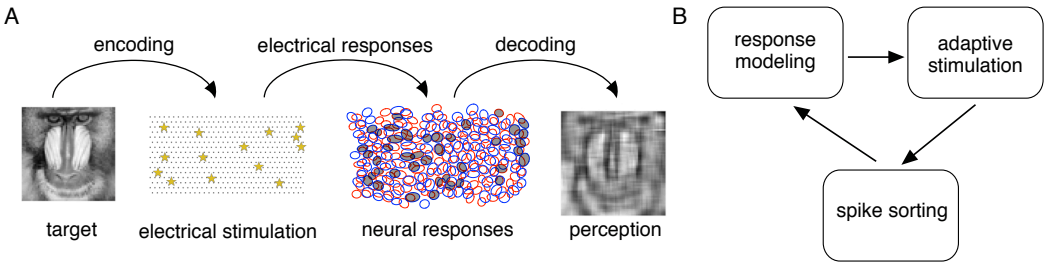

Figure 1: (A) Functional components of a retinal prosthesis. (B) Different steps in adaptive characterization of electrical response properties.

Data from *ex vivo* experiments with primate retina are used as a lab prototype for an artificial retina. Three major steps to calibrate the interface, based on the voltage recorded in response to electrical stimulation, are proposed (Figure 1B):

- **Spike sorting**: Identify spikes produced by target cells. We develop a novel approach to estimate electrical artifacts, a key hurdle in spike sorting [Mena et al., 2017], in a subspace identified from past experimental data.
- **Response modeling**: Develop a model of electrical response properties of the target cells. We use a model that incorporates as a prior the relationship between recorded spike amplitude and the threshold for electrical stimulation of a cell on a given electrode, from past experimental data.
- **Adaptive stimulation**: Choose the next electrical stimulation pattern to optimize estimation. Inspired by previous work [Lewi et al., 2009, Shababo et al., 2013], a method is developed to exploit the data already recorded from a retina to optimize the choice of stimulation patterns for the next batch of measurements.

Finally, to assess the performance of the approach, a modification is presented to minimize the variance in the *reconstructed* estimate of a spatial visual stimulus, a more meaningful indicator of performance for an artificial retina.

2 Spike sorting

The goal of spike sorting is to identify spikes fired in response to electrical stimulation and separate spikes produced by different neurons. Spike sorting in presence of electrical stimulation is difficult because the recorded spike voltages are corrupted by large stimulation artifacts, with magnitude and duration comparable to those of the spike waveforms. Hence artifacts must be subtracted before matching the residual traces to spike waveforms of previously identified cells. The novel aspect of our method is to estimate the artifact in a subspace identified from previous experiments.

Using the artifacts estimated by applying an existing [Mena et al., 2017] spike sorting algorithm, a basis matrix $A_{a,d(r,e)}$ is estimated for each stimulation amplitude (a) and distance $d(r,e)$ between the recording and stimulating electrodes. For recording electrode r and amplitude a, the estimated artifact is given by $A_{a,d(r,e)}\vec{b}_{a,r}$, with $\vec{b}_{a,r}$ representing the artifact in this subspace.

Let $\vec{y}_{a,r} \in \mathbb{R}^L$ be the recorded waveform, $\vec{x}_{c,a} \in \{0,1\}^L$ the spiking activity and $W_{c,r} \in \mathbb{R}^{L\times L}$ be the matrix consisting of shifted copies of a previously identified spike waveform recorded on electrode r. In neural recordings, the cell has at most one spike during this recording interval, and when it spikes the amplitude is exactly 1. This constraint is incorporated by parameterizing $x_{c,a}$ as a softmax function of real valued $\vec{z}_{c,a}$ with temperature τ : $\vec{x}_{c,a,t} = \frac{e^{z_{c,a,t}/\tau}}{\sum_t e^{z_{c,a,t}/\tau} + e^{q_{c,a}/\tau}}$, with $q_{c,a}$ a dummy parameter. Hence the reconstruction of neural recording given the spiking activity is given by

$\vec{y}_{a,r} = W_{c,r}\vec{x}_{c,a}$. Since only a few cells are activated in response to electrical stimulation, a sparsity enforcing L1 norm penalty is applied on $\vec{x}$.

The artifact parameters $\vec{b}$ and spike assignments $\vec{x}$ are estimated by minimizing the penalized reconstruction error ($\mathcal{L}_{\text{spike-sort}}$) for a particular stimulating electrode e, the recorded voltage traces on multiple recording electrodes and all the stimulating amplitudes simultaneously -

$$\mathcal{L}_{\text{spike-sort}} = \sum_a \sum_r \|\vec{y}_{a,r} - (A_{a,d(r,e)}\vec{b}_{a,r} + \sum_c W_{c,r}\vec{x}_{c,a})\|_2^2 + \lambda_{L1} \sum_c \|\vec{x}_{c,a}\|_1$$

Optimization is performed using Adam (learning rate = 0.01), with λ_{L1} chosen using cross-validation and temperature τ is reduced by a fraction of 0.8 every time the loss converges.

3 Response modeling

Given the spikes recorded in response to electrical stimulation, the goal of response modeling is to estimate the activation probability for each cell and electrode pair. We begin by presenting the standard approach, which involves estimating the response probability for each cell-electrode pair independently, followed by a joint model that incorporates priors from previous experiments.

For estimating these models, we are given N samples of electrical stimulus-response pairs $\{e_n, a_n, c_n\}_{n=1}^{n=N}$, with stimulating electrode $e_n \in \{1, \cdots, N_e\}$, activated cell $c_n \in \{1, \cdots, N_c\}$ and current level $a_n \in \{1, \cdots, N_a\}$.

3.1 Independent model

In this model, we assume that each cell-electrode pair behaves independently. For each sample, the the spiking probability is modeled as a Bernoulli distribution $P(R_n = 1) := \gamma_{e_n,a_n,c_n} = \frac{1}{1+e^{-(p_{e_n,c_n}(a_n - q_{e_n,c_n}))}}$, where p_{e_n,c_n}, q_{e_n,c_n} are the parameters of the sigmoidal activation curve for the stimulating electrode e_n and cell c_n. The parameters are inferred independently by maximizing the logistic log-likelihood for each cell electrode pair.

3.2 Joint model

Using prior data within the same retina along with data from previously recorded retinas could lead to more efficient characterization of activation probabilities. Previous work [Madugula et al., 2017] suggests that for a given electrode, the recorded spike amplitude and the stimulation threshold are inversely related, with a different curve for somatic stimulation and axonal stimulation. In this section, we present a model that jointly models this relationship across multiple cell-electrode pairs and describe a variational procedure to infer the global amplitude-threshold relationship within a retina, as well as infer the activation threshold for each cell-electrode pair.

The responses are denoted by $R_n \in \{0, 1\}$. For each cell c and electrode e, $E_{e,c}$ denotes the recorded spike amplitude and $T_{e,c}$ denotes whether the electrode e is recording from the soma or axon, as determined by the spike shape. The spike threshold $q_{e,c}$ is modeled as a Gaussian distribution with $q_{e,c} \sim \mathcal{N}(g(E_{e,c}, T_{e,c}), \nu^2)$, where $g(E, T) = x_T + y_T/E_i$ with a separate curve for electrodes that record from soma and axons. Further, the prior for $\{x, y\}$ is modeled with a two dimensional Gaussian $\{x_T, y_T\} \sim \mathcal{N}(\mu_T, \Sigma_T)$ for $T \in \{\text{soma}, \text{axon}\}$. The parameters for prior distribution (μ_T, Σ_T) are estimated from stimulated electrode and cell pairs in previous experiments (see Results).

Hence, the parameters of the model for electrically evoked responses are given by $\Theta = \{\{p_{e,c}, q_{e,c}\}_{e=1,c=1}^{e=N_e,c=N_c}; \{x_j, y_j\}_{j\in\{\text{soma,axon}\}}, \nu\}$ and the resulting model likelihood ($\mathcal{L}_{\text{model}}$) is

$$\Pi_n P(R_n|a_n; p_{e_n,c_n}, q_{e_n,c_n}) \Pi_{e,c} P(q_{e,c}|E_{e,c}; x_{T_{e,c}}, y_{T_{e,c}}, \nu_{T_{e,c}}) \Pi_{i\in\{soma,axon\}} P(x_i, y_i|\mu_i, \Sigma_i)$$

During inference, ν is treated as non-random and other parameters are estimated by variational approximation [Blei et al., 2017, Wainwright et al., 2008] of the $P(\Theta|\{R_n, e_n, a_n, c_n\}_{n=1}^{n=N})$. Let $z^{p_{e,c}}, z^{q_{e,c}}, z^{x_i}, z^{y_i}$ represent the variational parameters corresponding to parameters in Θ. We learn a mean-field variational approximation of the posterior $P(\Theta|\{R_n, e_n, a_n, c_n\}_{n=1}^{n=N}) \approx \Pi_{e,c} q(z^{p_{e,c}}) q(z^{q_{e,c}}) \Pi_{i\in\{soma,axon\}} q(z^{x_i}) q(z^{y_i})$. The parameters of the variational distribution (ϕ) are estimated by maximizing the evidence lower bound (ELBO) on the log-likelihood ($-\log \mathcal{L}_{\text{model}}$):

$$-\log \mathcal{L}_{\text{model}} \geq \mathbb{E}_{q(z)} \log P(R, Z) + H(q(z)) \tag{1}$$

As shown below, the first term of the joint probability corresponds to modeling electrically evoked spikes, the second term corresponds to modeling spike threshold from spike amplitudes, and third term corresponds to the relationship between spike threshold and spike amplitude for the all the cell-electrode pairs within a retina:

$$\mathbb{E}_{q(z)} \log P(R, Z) = \sum_{n=1}^{n=N} \mathbb{E}_{q_\phi(z^{p_{e_n,c_n}}), q_\phi(z^{q_{e_n,c_n}})} \log P(R_n, z^{p_{e_n,c_n}}, z^{q_{e_n,c_n}} | a_n)$$
$$+ \sum_{e,c} \mathbb{E}_{q_\phi(z^{q_{e,c}})} \log P(z^{q_{e,c}}, z^{x_{T_{e,c}}}, z^{y_{T_{e,c}}} | E_{e,c}, \nu_{T_{e,c}}) + \sum_i \mathbb{E}_{q_\phi(z^{x_i}), q_\phi(z^{y_i})} \log P(z^{x_i}, z^{y_i} | \mu_i, \sigma_i)$$

The variational distributions are parameterized as Gaussians: $q_\phi(z^{p_{e,c}}) = \mathcal{N}(\phi_\mu^{p_{e,c}}, \phi_{\sigma^2}^{p_{e,c}})$; $q_\phi(z^{q_{e,c}}) = \mathcal{N}(\phi_\mu^{q_{e,c}}, \phi_{\sigma^2}^{q_{e,c}})$; $q_\phi(z^{x_i}) = \mathcal{N}(\phi_\mu^{x_i}, \phi_{\sigma^2}^{x_i})$ and $q_\phi(z^{y_i}) = \mathcal{N}(\phi_\mu^{y_i}, \phi_{\sigma^2}^{y_i})$, with $H(q)$ being the sum of Gaussian entropy corresponding to each variational parameter.

For maximizing the ELBO, the variational parameters are sampled using the re-parametrization trick [Kingma and Welling, 2013]: $z = \phi_\mu + \phi_{\sigma^2}\epsilon, \;\; \epsilon \sim \mathcal{N}(0, I)$. The empirical approximation of ELBO is computed by averaging 10 samples for p, q and one sample for x, y. This approximation is maximized by stochastic gradient descent with norm clipping. At each step of the stochastic gradient descent, the objective function is evaluated over all samples, resulting in randomness only due to sampling of variational parameters. Finally, the posterior activation probabilties are estimated by averaging over multiple 1000 random samples of z^p, z^q.

4 Adaptive stimulation

In this section, the goal is develop an algorithm that uses responses from prior stimulation to choose subsequent stimulation patterns, in closed loop. Since real-time closed loop experiments are generally not feasible using existing hardware, we assume that the experiment would run in multiple phases, with the algorithm choosing the entire collection of current patterns to stimulate in next phase. The first phase is non-adaptive: each electrode and amplitude is stimulated T times, for a total of $N_e N_a T$. In subsequent phases, parameter estimates from earlier phases are used to allocate a total of $N_e N_a T$ stimuli unevenly across electrodes and amplitudes.

The number of stimuli for electrode and amplitude $T_{e,a} \in \mathbb{Z}_+$ are computed by minimizing a loss function U, that that depends on the accuracy of parameter estimates. In this paper, we choose the utility function as total variance in the estimate of response probability across electrodes and amplitudes $U = \sum_{e,a,c} var(\gamma_{e,a,c})$, where $var(\gamma_{e,a,c})$ denotes the variance in estimate of activation probability. This condition is identical to A-optimality in optimal design literature [Atkinson et al., 2007], departing from the commonly used information-theoretic Lewi et al. [2009], Paninski et al. [2007].

The variance in activation probability is computed by first estimating the Fisher information of the two parameters of the sigmoidal activation curve. Let $T'_{e,a}$ denote the previous number of measurements for electrode e and amplitude a. Define $X_a = [a, 1]$, $\theta_{e,c} = [p_{e,c}, -p_{e,c}q_{e,c}]$. Under this notation, the activation probability $\gamma_{e,a,c} = \frac{1}{1+exp(-\theta_{e,c}^T X_a)} = f(-\theta_{e,c}^T X_a)$. The Fisher information of $\theta_{e,c}$ after $(T_{e,a} + T'_{e,a})$ trials is computed using the chain rule as $I(\theta_{e,c}) = \sum_a (T_{e,a} + T'_{e,a})\gamma_{e,a,c}(1 - \gamma_{e,a,c})X_a X_a^T$. Variance can be approximated with the inverse of Fisher information $I(\theta_{e,c})^{-1}$. Finally, the first-order expansion of $\gamma_{e,a,c}$ gives the variance of individual parameter estimates as $var(\gamma_{e,a,c}) \approx (f')^2 var(\theta_{e,c}) = Q_{e,a,c}^T var(\theta_{e,c}) Q_{e,a,c}$, where $Q_{e,a,c} = \gamma_{e,a,c}(1 - \gamma_{e,a,c})X_a$. Normalizing the current levels a between -1 and 1 leads to better condition number for Fisher information and better approximation of the variance. After relaxing the integer constraint on $T_{e,a}$, the optimization problem is:

$$\begin{aligned} \underset{T_{e,a}}{\text{minimize}} \quad & \mathcal{L}_{\text{adapt-stim}} = \sum_{e,a,c} Q'_{e,a,c} [\sum_{a'} (T_{e,a'} + T'_{e,a'})\gamma_{e,a',c}(1 - \gamma_{e,a',c})X_{a'} X'_{a'}]^{-1} Q_{e,a,c} \\ \text{subject to} \quad & \sum_{e,a} T_{e,a} \leq N_e N_a T, \qquad T_{e,a} \geq 0 \;\; \forall e, a \end{aligned} \tag{2}$$

Note that the final formulation only uses probability estimates γ, which are either estimated by fitting the independent model, or performing variational inference on the joint model. Re-parametrization

of $T_{e,a}$ converts the constrained optimization problem into an unconstrained optimization problem. Specifically we use a soft-max representation of $T_{e,a} = (N_e N_a T)\frac{e^{t_{e,a}/\tau}}{\sum_{e',a'} e^{t_{e',a'}/\tau} + e^{\delta_{e'}/\tau}}$, where $\tau = 1000$ is the temperature parameter and δ_e is a dummy parameter to allow for loose constraints. After minimizing the unconstrained problem using Adam optimization [Kingma and Ba, 2014], the integer solution for $T_{e,a}$ is obtained by rounding.

5 Evaluation for artificial retina application

The objective function used for adaptively selecting the measurements can be modified based on the specific application of the estimated probabilities. For example, recent work [Shah et al., 2019] proposed that the contribution to perception of many stimulation patterns can be combined by sequential stimulation faster than visual integration times. In this framework, we can quantify how estimates of response probabilities from each cell-electrode pair contribute to the accuracy of perception. Specifically, let R_γ denote the cell response, where the spike with probability γ. When $\vec{d}_c$ denotes the linear reconstruction decoding filter associated with activation of cell c, the cell's contribution to perception is given by $\vec{d}_c R_\gamma$. The expected change in perception when the cell response is generated by estimated probability $\hat{\gamma}$ instead of the true probability γ is given by $\mathbb{E}_{\hat{\gamma},R}\|\vec{d}_c(R_{\hat{\gamma}} - R_\gamma)\|^2 = \|\vec{d}_c\|^2(var(R_{\hat{\gamma}}) + var(R_\gamma) + var(\hat{\gamma}))$. Since the adaptive stimulation choice only affects the variance in estimate of probabilities (last term), the objective of the adaptive stimulus selection algorithm can be modified to minimize the decoding error, averaged across different stimulation patterns: $\sum_e \sum_a \sum_c \|d_c\|_2^2 var(\gamma_{e,a,c})$.

6 Results

512-electrode *ex vivo* extracellular stimulation of and recording from primate retinal ganglion cells were used to evaluate the algorithms [Litke et al., 2004, Frechette et al., 2005]. First, recorded voltages from 30 minutes of visual stimulation were spike sorted using a custom software. The expected spatio-temporal spike waveform for each cell was identified by averaging the recorded voltage waveforms obtained from thousands of spikes. Spike amplitude was measured on each electrode as the maximum negative voltage deviation, and spike shape was used to determine if the electrode was recording from the soma (biphasic waveform) or axon (triphasic waveform). Subsequently, electrical stimulation experiments were performed by passing current through each of electrode at 38 different current levels ($0.1 - 4\mu A$) with 25 repeats per current level.

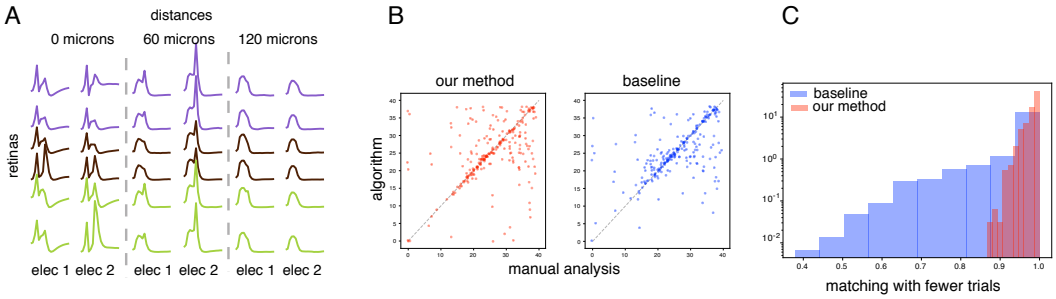

Figure 2: **Spike sorting** (A) Artifact recorded for a $0.68\mu A$ triphasic current pulse, on electrodes at different distances on a $60\mu m$ hexagonal grid. Lines in same column correspond to artifacts from different retinal preparations but the same electrode, lines with the same color indicate recordings from different pieces of retina from the same animal. (B) Comparison of estimated current index for 50% spiking probability for the algorithm (y-axis) and manual analysis (x-axis) for our method (red) and the simplified method from Mena et al. [2017] (baseline, blue). Each of the $\sim$ 200 dots represents a cell-electrode pair. (C) Comparison of spike sorting results on 5 trials, when the 5 trials were analyzed independently, and as part of a total of 24 trials. Histogram across multiple cell-electrode pairs for our method (red) and a simple form of a previous approach (blue) [Mena et al., 2017].

6.1 Spike sorting

The performance of spike sorting was evaluated with voltage traces recorded in response to repeated electrical stimulation. For each of the recorded traces, stimulation artifacts were estimated by applying

the simplified algorithm previously proposed in Mena et al. [2017]. Briefly, the artifact estimate is initialized to the results obtained with a lower current amplitude, and then is updated by iterating between greedy spike estimation and artifact estimation. The estimated artifacts for different relative locations of stimulating and recording electrodes, across multiple experiments, are shown in Figure 2A. For a given stimulation current and a fixed distance between stimulating and recording electrodes, the artifact was similar across different stimulating electrodes, and across recordings. A 9 dimensional space obtained by PCA explained most of the variance in the estimated artifacts.

Improved spike sorting was therefore implemented using the reduced space to regularize the artifact estimates. Performance was tested on the responses of one retina, with the artifact waveform basis estimated from 22 recordings from different retinas. Analysis of 24 repetitions for each electrode and amplitude revealed that the estimated activation threshold (current value that elicits a spike with probability 0.5) matched the value obtained with manual analysis by a human, in 241 cell-electrode pairs (Figure 2B). In some cases the algorithm produced higher thresholds than the human, and vice-versa, but no large overall bias was observed. To test the accuracy of the method, both algorithms were applied in two cases: using 5 trials, or 24 trials, per stimulation pattern. The new approach resulted in greater consistency: spike times identified using fewer trials matched spike times identified using more trials more closely (Figure 2C). This shows that incorporating priors on artifact waveforms across recordings can allow for more accurate spike detection.

6.2 Response modeling

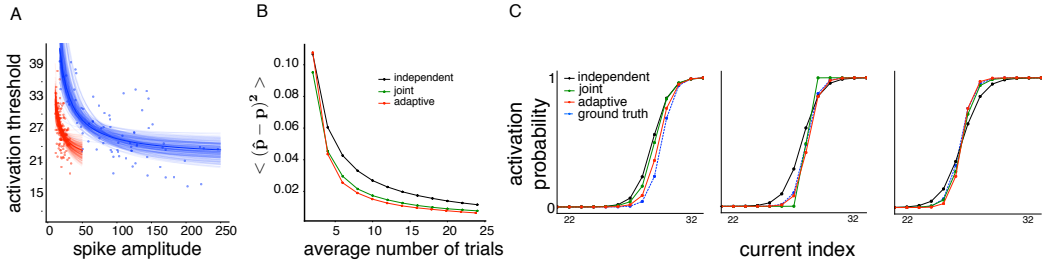

Figure 3: **Response modeling** (A) Relationship between observed spike amplitude (horizontal axis) and activation threshold (vertical axis) for electrodes recording from soma (blue) or axon (red). Each dot corresponds to a cell-electrode pair. Thin lines indicate different fits obtained by randomly sampling a subset of cells (B) Squared error in probability estimates, averaged over multiple cell-electrode pairs (vertical axis) with different number of measurements using different methods. Measurements are done in batches, with a total $2N_eN_a$ measurements per batch. (C) Estimated spike probability for a few cell-electrode pair using different methods.

To improve the efficiency of response model estimation, we examine the effect of imposing priors on electrical properties of cells, and test the impact of these priors separately from spike sorting. We simulate neural responses based on activation probabilities estimated by applying the existing spike sorting algorithm on a previously recorded retina. Two models were compared: (a) *independent* model, in which spike probabilities are estimated without priors; (b) *joint* model, in which spike probabilities are estimated with priors.

The key concept we use for priors is that electrical stimulation thresholds for a cell on different electrodes should bear some relationship to the magnitude of the spike recorded on those electrodes, because both threshold and spike amplitude should depend inversely on the physical distance to the spike initiation region of the cell. Thus, for the joint model, the relationship between threshold and spike amplitude was learned from a set of 199 cell-electrode pairs across 3 data sets, identified by applying the spike sorting algorithm in Mena et al. [2017]. As described in Section 3.2, this relationship can be described by a separate curve for somatic and axonal activation (Figure 3A). The mean and variance of the parameters of the inverse relationship were identified fitting curves by randomly sampled cell-electrode pairs.

The impact of using this prior was evaluated by measuring the responses from the simulated retina in batches, where an average of 2 measurements per electrode and amplitude were delivered in each batch. As the number of measurements increased, the joint model (Figure 3B, green) produced estimates of spike probability which were more accurate than estimates made by the independent model (black curve). Examples of the improved estimate of response probability as a function of

stimulation current using the joint approach, for a few cell-electrode pairs, is shown in Figure 3C. Thus, particularly in the limited data regime, priors on the relationship between spike amplitude and threshold can improve estimation of the response model.

6.3 Adaptive stimulation

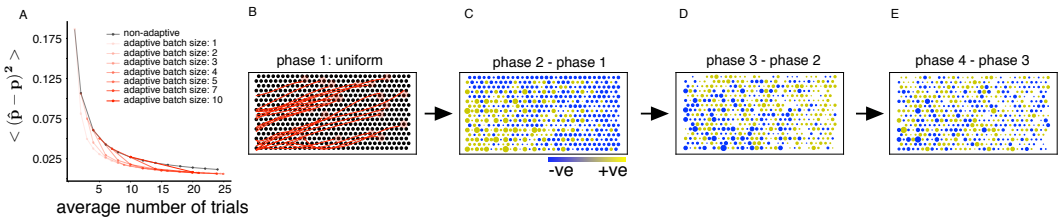

Figure 4: **Adaptive stimulation** (A) Adaptive stimulation with different batch sizes (red lines) and non-adaptive stimulation, both with the independent response model (black). (B)Spatial distribution of the electrodes selected by the adaptive algorithm. All cells and electrodes are selected uniformly (2 times each). Size of circle indicates the number of times an electrode is stimulated. Cell soma (red circle) and axon direction (red lines) are estimated from detected spike waveform across electrodes. (C) The increase (yellow) and decrease (blue) in number of measurements in second (first adaptive) phase compared to the first (non-adaptive) phase. (D, E) Successive difference in the number of measurements for each electrode.

An alternative to regularizing based on previous retinas is to use feedback from measurements already made in the target retina to select the next electrical stimulus, and thus potentially improve estimation in closed loop. This can also be considered as a prior based on immediately preceding measurements. To test the value of this approach, an adaptive algorithm was developed using two stimuli on average per electrode and amplitude for each batch ($2N_eN_a$ measurements in total). After the first non-adaptive phase, the adaptive algorithm divides all the available capacity in the next batch across stimulations to minimize estimation error. Adaptive stimulation, with the simpler independent model improved response model estimation to the non-adaptive method (Figure 3B, red). The adaptive method with the independent model and the non-adaptive method with the joint model exhibited similar performance, suggesting that, with sufficient data, the computationally expensive adaptive stimulation can be replaced with better priors in non-adaptive stimulation. Adaptive method with joint model had similar curve as the adaptive model with independent model (not shown), possibly because the contribution of prior from previous experiments is reduced with better stimulus selection. The estimated response probability as a function of stimulation current for a few cell-electrode pairs after three phases of each approach are shown in Figure 3C, again showing similar performance of the two approaches.

Reduction in the number of adaptive phases leads to a proportional decrease in the number of computational steps. We tested the algorithm with smaller number of adaptive phases, with a corresponding increase in stimulation capacity per phase. We observed that for the same number of measurements (ex. 10 trials on average), a smaller number of adaptive phases gives similar estimation accuracy (ex. two phases with batch size 5 each gives similar accuracy as 10 phases with batch size 1 each Figure 4A) leading to computational efficiency.

Adaptive stimulation also revealed systematic spatial structure in the electrodes selected for stimulation. Compared to the uniform stimulation in the first phase, the electrodes on one side of the array are stimulated more frequently in the subsequent adaptive phase (Figure 4C). This could potentially be explained by the geometry of the axons in the recording, which cross the array in a particular direction as they head toward the optic nerve (Figure 4B). Because electrodes on the side of the array with more axons would potentially stimulate more of the cells over the array, the adaptive algorithm preferentially selects these electrodes to achieve a greater reduction in the average estimation error. However, in the third phase, the algorithm corrects itself and selects electrodes with fewer stimulated cells, due to higher residual estimation error (Figure 4D). For subsequent phases, there is no spatial structure in selected electrodes, since the residual estimation is now similar across electrodes (Figure 4E).

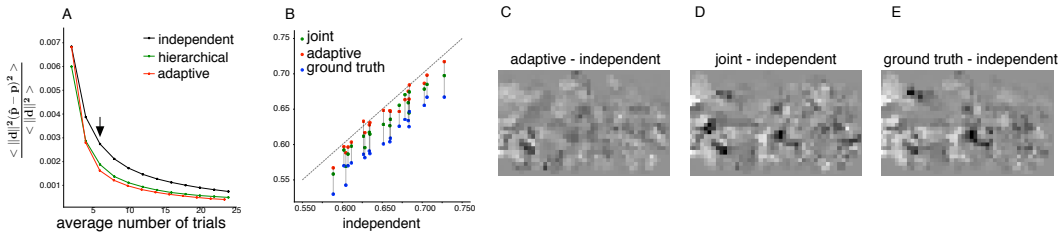

Figure 5: **Impact of electrical calibration on decoded stimulus** (A) Variance in estimated probabilities, weighted by decoder norm as a function of the average number of stimulation pulses (per electrode and amplitude) for the three methods presented in the text. Independent and joint model same as Figure 3, but stimuli for adaptive method chosen using the modified objective. Black arrow indicates the trials at which the estimates are compared in (B). (B) Expected mean squared error of decoded stimulus when different dictionaries are used for choosing the stimulation pattern. Each dot corresponds to a different white noise image. (C) The improvement in the expected linearly decoded stimulus using different probability estimates (over the independent model).

6.4 Performance for neural interface

While the above techniques improve spike identification, response model estimation, and stimulus selection, a larger issue is how effective these improvements are for the function of the neural interface. In the case of vision restoration, performance ultimately depends on how each targeted cell contributes to vision, and on how well an actual image can be represented in the collection of cells.

To test functional impact in a way that accounts for how cells contribute to perception, the adaptive method was used with a modified error metric. Previous work Shah et al. [2019] proposes that an artificial retina could linearly combine the expected perception from stimulation of different electrodes by temporally multiplexing within the integration time of the brain. Expected perception is inferred based on the idea that the brain performs optimal linear reconstruction of the stimulus from retinal inputs. When electrode e is stimulated at amplitude a, the change in perception due to error in the estimate of response probability is given by $\sum_c \|d_c\|^2 var(\hat{\gamma}_{e,ac})$, where d_c is the optimal linear reconstruction filter for cell c. Hence, the adaptive stimulation algorithm was modified to minimize the average estimation error across electrodes and amplitudes (see Methods). Applying the modified algorithm to simulated data revealed a faster decrease in the variance of the reconstructed stimulus compared to non-adaptive stimulation (Figure 5A). As before, the joint model with non-adaptive stimulation joint model also outperformed the independent model.

To test functional impact in a way that accounts for image structure, we exampled the spatial reconstruction of 20 different target images based on electrical stimulation. Optimizing (using method in [Shah et al., 2019]) the electrical stimulation using estimated response probabilities for both the adaptive and joint calibration algorithms resulted in more accurate stimulus reconstruction compared to the independent model (Figure 5B,C,D). Thus, the gains from efficient characterization of electrical response properties are likely to be translated into improved artificial vision.

7 Summary

This paper presents three novel methods to optimize the function of a neural interface, specifically, an artificial retina for treating blindness. Using large scale multi-electrode recordings from primate retina as a lab prototype, prior information and closed-loop approaches improved the accuracy and efficiency of spike sorting, response modeling and stimulus selection. Notably, the computationally expensive closed-loop stimulation approach exhibited similar performance to a much simpler non-adaptive approach that uses priors from previous experiments, highlighting the value of using large data sets to improve device function. Evaluation of image reconstruction showed that these approaches improved overall function in terms of the quality of the visual image transmitted to the brain.

Extensions of this work are possible by incorporating additional priors. For spike sorting, prior information about activation probabilities can be incorporated by an additional logistic loss, exploiting the soft-max parameterization. For response modeling, additional structure in the relationship between recording and stimulation using multi-electrode stimulation patterns could be valuable [Jepson et al., 2013, Fan et al., 2018].

In principle, similar approaches may be useful for different multi-electrode stimulation patterns, and for other neural systems (e.g. control of external devices using brain recordings) and interfaces (e.g. optical recording and stimulation [Shababo et al., 2013]). With advent of large-scale data sets, as well as availability of motor and visual prosthesis technologies in many subjects, the methods developed here may be helpful in capturing similarities and differences among individuals and experiments.

References

A. Atkinson, A. Donev, and R. Tobias. *Optimum Experimental Designs, with SAS*. 01 2007. ISBN 978-0-19-929659-0.

D. M. Blei, A. Kucukelbir, and J. D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017.

V. H. Fan, L. E. Grosberg, S. S. Madugula, P. Hottowy, W. Dabrowski, A. Sher, A. M. Litke, and E. Chichilnisky. Epiretinal stimulation with local returns enhances selectivity at cellular resolution. *Journal of neural engineering*, 2018.

E. S. Frechette, A. Sher, M. I. Grivich, D. Petrusca, A. M. Litke, and E. Chichilnisky. Fidelity of the ensemble code for visual motion in primate retina. *Journal of neurophysiology*, 94(1):119–135, 2005.

M. S. Humayun, J. D. Dorn, L. Da Cruz, G. Dagnelie, J.-A. Sahel, P. E. Stanga, A. V. Cideciyan, J. L. Duncan, D. Eliott, E. Filley, et al. Interim results from the international trial of second sight's visual prosthesis. *Ophthalmology*, 119(4):779–788, 2012.

L. H. Jepson, P. Hottowy, K. Mathieson, D. E. Gunning, W. Dąbrowski, A. M. Litke, and E. Chichilnisky. Focal electrical stimulation of major ganglion cell types in the primate retina for the design of visual prostheses. *Journal of Neuroscience*, 33(17):7194–7205, 2013.

J. N. Kerr and W. Denk. Imaging in vivo: watching the brain in action. *Nature Reviews Neuroscience*, 9(3):195, 2008.

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

D. P. Kingma and M. Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

D. R. Kipke, W. Shain, G. Buzsáki, E. Fetz, J. M. Henderson, J. F. Hetke, and G. Schalk. Advanced neurotechnologies for chronic neural interfaces: new horizons and clinical opportunities. *Journal of Neuroscience*, 28(46):11830–11838, 2008.

J. Lewi, R. Butera, and L. Paninski. Sequential optimal design of neurophysiology experiments. *Neural Computation*, 21(3):619–687, 2009.

A. Litke, N. Bezayiff, E. Chichilnisky, W. Cunningham, W. Dabrowski, A. Grillo, M. Grivich, P. Grybos, P. Hottowy, S. Kachiguine, et al. What does the eye tell the brain?: Development of a system for the large-scale recording of retinal output activity. *IEEE Transactions on Nuclear Science*, 51(4):1434–1440, 2004.

H. Lorach, G. Goetz, R. Smith, X. Lei, Y. Mandel, T. Kamins, K. Mathieson, P. Huie, J. Harris, A. Sher, et al. Photovoltaic restoration of sight with high visual acuity. *Nature medicine*, 21(5): 476, 2015.

S. Madugula, G. E. Mena, L. E. Grosberg, V. H. Fan, P. Hottowy, W. Dabrowski, A. Sher, A. Litke, L. Paninski, and E. Chichilnisky. Large-scale analysis of patterned epiretinal stimulation for prosthesis design. *TEATC*, 2017.

G. E. Mena, L. E. Grosberg, S. Madugula, P. Hottowy, A. Litke, J. Cunningham, E. Chichilnisky, and L. Paninski. Electrical stimulus artifact cancellation and neural spike detection on large multi-electrode arrays. *PLoS computational biology*, 13(11):e1005842, 2017.

L. Paninski, J. Pillow, and J. Lewi. Statistical models for neural encoding, decoding, and optimal stimulus design. *Progress in brain research*, 165:493–507, 2007.

E. Richard, G. A. Goetz, and E. Chichilnisky. Recognizing retinal ganglion cells in the dark. In *Advances in Neural Information Processing Systems*, pages 2476–2484, 2015.

A. B. Schwartz. Cortical neural prosthetics. *Annu. Rev. Neurosci.*, 27:487–507, 2004.

B. Shababo, B. Paige, A. Pakman, and L. Paninski. Bayesian inference and online experimental design for mapping neural microcircuits. In *Advances in Neural Information Processing Systems*, pages 1304–1312, 2013.

N. P. Shah, S. Madugula, G. Mena, P. Tandon, P. Hottowy, A. Sher, A. Litke, S. Mitra, and E. Chichilnisky. Optimization of electrical stimulation for a high-fidelity artificial retina. *International IEEE EMBS conference on Neuroengineering*, 2019.

K. Stingl, K. U. Bartz-Schmidt, D. Besch, A. Braun, A. Bruckmann, F. Gekeler, U. Greppmaier, S. Hipp, G. Hörtdörfer, C. Kernstock, et al. Artificial vision with wirelessly powered subretinal electronic implant alpha-ims. *Proceedings of the Royal Society B: Biological Sciences*, 280(1757): 20130077, 2013.

M. J. Wainwright, M. I. Jordan, et al. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305, 2008.

B. S. Wilson, C. C. Finley, D. T. Lawson, R. D. Wolford, D. K. Eddington, and W. M. Rabinowitz. Better speech recognition with cochlear implants. *Nature*, 352(6332):236, 1991.