Computational Challenges and Opportunities for a Bidirectional Artificial Retina

Nishal P. Shah 1,2,3
E.J. Chichilnisky 2,3,4
1. Department of Electrical Engineering, Stanford University
2. Hansen Experimental Physics Laboratory, Stanford University
3. Department of Neurosurgery, Stanford University
4. Department of Ophthalmology, Stanford University

Abstract

A future artificial retina that can restore high acuity vision in blind people will rely on the capability to both read (observe) and write (control) the spiking activity of neurons using an adaptive, bi-directional and high-resolution device. Translating the capabilities of such a device into more acute visual perception will also require substantial computational advances. Using high-density large-scale recording and stimulation in the primate retina with an ex vivo multielectrode array lab prototype, we frame several of the major computational problems, and describe current progress and future opportunities in solving them. First, we identify cell types and locations from spontaneous activity in the blind retina, and then efficiently estimate their visual response properties by using a low-dimensional manifold of inter-retina variability learned from a large experimental dataset. Second, we estimate retinal responses to a large collection of relevant electrical stimuli by passing current patterns through an electrode array, spike sorting the resulting recordings and using the results to develop a model of evoked responses. Third, we reproduce the desired responses for a given visual target by temporally dithering a diverse collection of electrical stimuli within the integration time of the visual system. Together, these novel approaches may substantially enhance artificial vision in a next-generation device.

1. Introduction

Neural interfaces are designed to restore or augment human capabilities, by either reading out or writing in the electrical activity of neurons. Read-out interfaces typically use neural activity to decode user intentions, for applications such as limb and cursor movements (Bacher et al., 2015; Jarosiewicz et al., 2015; Pandarinath et al., 2017) or speech (Anumanchipalli, Chartier and Chang, 2019; Stavisky et al., 2019). Write-in interfaces typically disrupt pathological neural activity for treating neurodegenerative diseases (Lozano and Lipsman, 2013) or stimulate sensory neurons to restore audition (Shannon, 2012), touch (Tyler, 2015) or vision (Goetz and Palanker, 2016). Future neural interfaces will likely enable applications such as controlling complex robotic devices, extending memory, or augmenting natural senses with artificial inputs (Marcus and Koch, 2014). These applications will likely require bi-directional capability, i.e. the ability to both read and write electrical activity in neural circuits. For example, read-write capability will be useful in providing sensory feedback about limb motion in motor prostheses, or
in optimizing electrical stimulation using the measured ongoing and evoked activity in sensory prostheses. Furthermore, to emulate and harness the natural resolution of signals in the nervous system, future devices will ideally function at single-cell, single-spike resolution. Neural interface functionality with this kind of adaptability and precision will require new approaches for accurate, efficient and fast information processing to match (or even exceed) the natural function of the neural circuit. Hence, important computational challenges must be addressed in order to realize the capabilities of such devices.

In this paper, we describe some of the key computational problems, and approaches to solving them, for a bi-directional, high-resolution artificial retina, which is designed to restore vision in people blinded by retinal degeneration. The retina is a thin layer of neural tissue at the back of the eye that performs the first steps in visual processing (Dowling, 1987; Wandell, 1995; Rodieck, 1998). Light from the visual scene is transduced by the photoreceptors into electrical signals, which are then processed by an extensive network consisting of many types of interneurons, and finally are encoded in the activity of each of ~20 functionally distinct retinal ganglion cell (RGC) types that project to diverse targets in the brain (Dacey, 2004). In certain conditions, notably age-related macular degeneration and retinitis pigmentosa, photoreceptor degeneration leads to incurable blindness. However, many ganglion cells remain excitable and connected to the brain, raising the possibility of artificially activating them to elicit visual perception. To accomplish this, retinal prostheses have been developed to electrically target various parts of the degenerated retina, either epiretinally at the RGC layer (Humayun et al., 2012), sub-retinally at the interneuron layer (Stingl et al., 2013; Palanker et al., 2020) or supra-choroidally outside the retina (Ayton et al., 2014). Although these devices have been able to restore some degree of vision, the resulting visual sensations are limited to rudimentary shapes such as blobs and arcs, and do not approach the high resolution, and functional value of natural vision (Goetz and Palanker, 2016). The performance of these devices is limited in part by the fact that they stimulate the diverse types of retinal neurons coarsely and non-selectively, and thus fail to reproduce the spatially and temporally precise, cell-type specific neural code of the retina.

We propose that a future artificial retina will more faithfully restore vision by replicating, as accurately as possible, the precise patterns of firing in the diverse RGC types in response to visual stimuli. To do so, the device will need to (a) use high-resolution electrical recordings of spontaneous activity to identify the cells and cell types in the region of the implant and their expected contribution to visual perception, (b) use electrical stimulation and recording to determine which cell(s) can be stimulated at what current level(s) and by which electrode(s) in the implant, and (c) use electrical stimulation to evoke the optimal achievable pattern of RGC activity for the incoming image, given the measured constraints of the neural interface.

Achieving these goals requires many advances in hardware as well as computation (Muratore and Chichilnisky, 2020). Multiple efforts are currently focused on building high-density implantable electrode arrays and interface electronics (Dragas et al., 2017; Jun et al., 2017; Muratore et al., 2019; Obaid et al., 2020). As these hardware development efforts proceed, we
can in parallel tackle the many accompanying computational problems that arise, using a large-scale high-resolution multi-electrode array recording and stimulation system as an ex vivo laboratory prototype of a future implanted device (Figure 1A). Using this experimental approach (Litke et al., 2004; Frechette et al., 2005; Hottowy et al., 2008, 2012; Sekirnjak et al., 2008), we identify and target different RGC types in the primate retina by electrical recording coupled with visual and electrical stimulation (Figure 1B). This approach features small electrodes (~10μm) that are densely packed (~30-60μm spacing), and brief (~0.1 ms) epiretinal stimulation at low current amplitudes (~1 μA) to activate RGCs with single-cell, single-spike resolution (Sekirnjak et al., 2007, 2008; Jepson et al., 2013; Jepson, Pawel Hottowy, et al., 2014). Based on experiments and analysis using this laboratory prototype, we propose that three major classes of computational problems must be solved to produce high-resolution artificial vision:

(1) How does the retina normally represent visual information? Decades of research have focused on developing an accurate encoding model \( R \sim g(S) \), that can predict the responses \( R \) of RGCs to any given visual stimulus \( S \) (Figure 1C) (Gollisch and Meister, 2010). This work has revealed that many different RGC types encode distinct features of the visual stimulus in their distinct spatial and temporal patterns of activity, providing a gold standard of performance for any artificial retina (Figure 1B). More recently, some studies have also developed decoding models \( S \sim h(R) \), to estimate the visual stimulus from recorded RGC responses, as a proxy for understanding how downstream neural circuitry in the brain could process the retinal signal (Rieke, 1996; Warland, Reinagel and Meister, 1997; Stanley, Li and Dan, 1999). These two modeling approaches have been largely separate in the computational neuroscience literature. Below we will argue that they find a meaningful meeting point in the neuroengineering application (Nirenberg and Pandarinath, 2012). Therefore, in what follows, we summarize encoding and decoding with a single utility function, \( f(S, R) \) (Figure 1C), which measures the degree of correspondence between a visual stimulus \( S \) and RGC firing \( R \).

(2) How does the retina respond to electrical stimulation? In principle, if we were able to selectively activate every target cell while avoiding other cells, arbitrary firing patterns in the diverse RGC types could be reproduced. However, existing devices instead stimulate many RGCs of all types simultaneously and non-selectively, providing only coarse control over neural activity and failing to provide the asynchronous, distinctive visual signals normally carried by the diverse RGC types to the brain. Although future high-resolution electrode arrays will allow more precise stimulation, perfect cellular selectivity will remain a challenge. Therefore, we present an approach to determine how to use the available electrical stimuli as effectively as possible in an implanted device. In what follows, we summarize the RGC responses \( R \) evoked by current passed through the electrodes \( C \) with an electrical response model \( M \), \( R \sim M(C) \) (Figure 1C).

(3) How do we optimally replicate RGC firing in a blind retina? Previous studies have proposed to optimize visual encoding while assuming high selectivity of stimulation (Nirenberg and Pandarinath, 2012) or while measuring selectivity using the elicited visual sensation reported by a blind person (Beyeler et al., 2019). In contrast, we use single-cell resolution measurements of the electrical response properties of RGCs to convert a target visual stimulus \( S \) into an
electrical stimulus (C) in real time. This is accomplished by identifying the electrical encoding of the incident image (P) that maximizes the expected visual utility function (f) between the target visual stimulus (S) and the probabilistic responses (R) generated via the electrical response function (M).

\[
C = P(S) = \arg\max_{E_{R-M(C)}} E_{R-M(C)} f(S, R)
\]  

(1)

This paper presents a detailed description of the above computational problems, progress towards addressing them, and possible future directions. Then, system design issues are discussed for implementing the solutions in distributed hardware. Finally, the caveats of the approach, and potential application of similar ideas in other neural interfaces, are considered.

Figure 1. Experimental prototype and computational methods for an adaptative, bi-directional, cellular-resolution artificial retina. (A) The experimental prototype consists of a multi-electrode array which is used to both record the electrical activity of RGCs and electrically stimulate them, along with a stimulus display for visual response characterization and cell type identification. (B) Functional properties of each cell are identified by correlating recorded spike trains with a stochastic visual stimulus, revealing the receptive fields of several hundred recorded cells (summarized by Gaussian fits; top). Distinct cell

()
types (side panels) are then identified by clustering in a low-dimensional space of receptive field parameters (central panel). The four numerically dominant cell types in the primate retina are shown, with their receptive fields forming uniform mosaics. Hence, the retina decomposes the visual scene into multiple distinct, orderly and complete visual representations that are transmitted to the brain. (C) Computational components of the artificial retina. A utility function $f(S, R)$ measures the degree of matching between the visual stimulus ($S$) and neural responses ($R$). This utility function can either measure the similarity in response space by encoding the visual stimulus with a response distribution ($g(S)$), or in the stimulus space by decoding the responses with a stimulus distribution ($h(R)$). The goal is to encode the stimulus ($S$) in the currents ($C$) passed through the electrodes in the array, using an image encoding function ($P$). The currents are optimized such that the electrically evoked responses ($R$) produced by the electrical encoding ($M$) maximize the utility function.

2. How does the retina normally represent visual information?

In this section, we focus on identifying the visual response properties of RGCs, captured by a utility function $f(S, R)$ which measures how closely the visual stimulus $S$ and neural responses $R$ correspond. Standard methods for identifying $f(.)$ are based on recording responses to visual stimuli, which is not possible in a blind retina. Instead, we first use light-evoked recordings from healthy retinas to model $f(.)$, and then adapt the information for a blind retina, in which only recordings of spontaneous activity are available. A major difficulty that arises is the observed variability of visual responses across different healthy retinas. We describe a method to systematically characterize this inter-retina variability and then to identify the optimal utility function in a blind person.

2.1 Identifying the utility function in a healthy retina

The utility function is most frequently examined in the response space, by transforming the visual stimulus into predicted responses, and then comparing this prediction to measured light responses in the retina. Specifically, the utility function is expressed as

$$f(S, R) = \rho(g(S), R),$$

where the encoder $g(.)$ produces a predicted neural response (or a distribution over possible responses) based on the visual stimulus, and $\rho(.)$ is a measure of similarity between responses (or distributions). Often, $\rho(.)$ is given as the negative log-likelihood of the data given the probabilistic prediction $g(.)$. Many previous studies have modeled $g(.)$ using a cascade of linear and nonlinear processing stages (see (Chichilnisky, 2001)), with extensions to include coupling between nearby cells (Mastronarde, 1983; Pillow et al., 2008; Greschner et al., 2011), adaptation (Smirnakis et al., 1997; Chander and Chichilnisky, 2001; Kim and Rieke, 2001; Carandini and Heeger, 2011; Kastner and Baccus, 2013), or spatial nonlinearity (Hochstein and Shapley, 1976; Demb et al., 1999, 2001; Freeman et al., 2015; Liu et al., 2017; Shah et al., 2020). Recent studies have also used machine learning to improve encoding models (Batty et al., 2016; L. T. McIntosh et al., 2016). Finally, instead of using negative-log likelihood or mean-squared error to measure the accuracy of model predictions, some studies have proposed similarity measures $\rho(.)$ that account for response variations unrelated to the stimulus.
Recently, machine learning methods have been developed to estimate a response similarity measure by minimizing the distance between pairs of responses generated by the same stimulus relative to the distance between pairs of responses generated by different stimuli (Shah et al., 2018).

A complementary approach is to measure similarity in the visual stimulus space, instead of the response space, by modeling how the brain could decode the visual stimulus from retinal responses. In this case, the loss function is expressed as

\[ f(S, R) = \sigma(S, h(R)), \]

where the decoder \( h(.) \) estimates the visual stimulus (or a distribution over possible visual stimuli) from neural responses and \( \sigma(.) \) is a measure of perceptual similarity between visual stimuli (or distributions of stimuli). Since the neural decoding function implemented by the brain is unknown, simplifying assumptions are required. A reasonable assumption is that the brain optimally decodes the neural responses to estimate the visual stimulus, based on the statistics of retinal responses to natural stimuli. A simple version of this idea is a linear stimulus decoder (Warland, Reinagel and Meister, 1997; Stanley, Li and Dan, 1999; Golden et al., 2019; Shah, Madugula, et al., 2019) selected to minimize the mean-squared error between the true and decoded stimulus. Examples of more elaborate nonlinear decoders include explicitly inverting the encoding models using priors from natural images (Pillow et al., 2008; Brackbill et al., 2017), or using supervised machine learning (Parthasarathy et al., 2017; Zhang et al., 2020). Similarly, more accurate measures of decoding error could be used, based on perceptual error metrics that capture the features of images that humans can most readily discriminate (Wang et al., 2004). An advantage of the decoding approach is that, in principle, it allows performance to be measured in the domain of greatest interest: the similarity of the perceived visual image to the incident image.

A third alternative is to represent both the visual stimulus and neural responses in an intermediate latent space in which similarity can be measured with a simple function. In this formulation,

\[ f(S, R) = ||m(S) - n(R)||_2, \]

where the visual stimulus is transformed to the latent space by \( m(.) \), neural responses are transformed to the latent space by \( n(.) \), and \( ||.||_2 \) is the mean-squared error cost. This formulation is useful when the similarity in either response or stimulus space is unknown, but many visually-evoked responses are available to estimate both \( m \) and \( n \). For example, this approach might be useful when we want to identify a utility function for specific tasks, such as navigation or reading, for which the standard methods of measuring similarity in visual stimulus space are too general. This approach may also be suitable for other sensory modalities, such as olfaction,
in which it is difficult to measure the similarity between different stimuli (Su, Menuz and Carlson, 2009).

Irrespective of the exact implementation of the utility function f(), the differences between visual stimulus properties of different healthy retinas recorded using the experimental lab prototype, and future application in the retina of a blind person, must be considered. Most experimental studies on visual encoding in the retina have either used simple white noise stimuli, or highly structured stimuli, such as moving bars or gratings, that isolate specific computational mechanisms. However, in natural viewing, the statistics of stimuli are very different, and the eyes are constantly saccading to different parts of the visual scene, with microsaccades and jitter around the fixation location (Kowler, 2011). Hence a population of RGCs in a region of the retina sees natural textures and edges through a tiny moving window. Recently, some studies have begun to develop computational models to explain RGC responses to such stimuli (Segal et al., 2015; Batty et al., 2016; Heitman et al., 2016; L. McIntosh et al., 2016; Salisbury and Palmer, 2016; Turner and Rieke, 2016; Maheswaranathan et al., 2019; Turner et al., 2019; Shah et al., 2020). Visual stimuli can also be specialized to tasks that are important for a blind person, such as indoor navigation or reading, which probably involve different stimulus statistics.

In addition to the visual stimulus, differences in the distribution of the visually and electrically evoked responses must also be considered. For example, overlapping ON and OFF cells are frequently activated simultaneously in electrical stimulation, but this rarely happens with visual stimulation (but see (Turner and Rieke, 2016)). If the parameters of the utility function f() are estimated from visually evoked responses, the utility function may be inaccurate for electrically evoked responses that are rarely if ever evoked by visual stimulation. A possible solution may be to use simpler models of the utility function (e.g. linear models), because they may be more robust to this problem than more complex and richer models (e.g. machine learning models) (Ben-David et al., 2010).

Finally, multiple considerations suggest that the utility function f() may depend significantly on the location in the visual field (Wandell, 1995; Rodieck, 1998). Ideally, a retinal implant would be placed at the fovea, the region of the retina with the highest cell density and thus the highest visual acuity. However, many experiments have focused on the less dense peripheral retina, for technical reasons (Frechette et al., 2005; Gauthier et al., 2009; Turner and Rieke, 2016) (but see (Wachtler, Wehrhahn and Lee, 1996; Lee, Martin and Grünert, 2010; Sinha et al., 2017)). In addition to the smaller receptive field sizes, differences in temporal integration in the cone photoreceptors and dominance of ON and OFF midget cell types have been reported in the fovea relative to the periphery (Dacey, 2004; Sinha et al., 2017). Further work is needed to fully characterize the unique visual response properties of the foveal region.

2.2 Identifying cell locations and types in a blind retina

A final challenge is to adapt our understanding of neural responses in healthy retinas to a blind retina. The most obvious approach is to identify the RGC locations and types in the target blind
retina, and then use the extensive body of retinal research to infer their original, healthy visual responses properties.

However, traditional approaches to cell type and location identification rely on light-evoked responses of RGCs and thus cannot be used in a blind retina. Recent work shows that the cell locations and types in a given retina can often be identified from recordings of electrical activity without evaluating light responses (Richard, Goetz and Chichilnisky, 2015). Specifically, the electrophysiological image (EI) of a cell obtained with high-density recordings, which is measured by averaging the recorded voltage traces on each electrode across many spikes recorded from that cell, reveals structural and biophysical properties of the somatic, axonal and dendritic compartments of the cell (Litke et al., 2004). First, the location of the soma and dendritic arbor provide an estimate of the receptive field location. Second, the sizes of the somatic and dendritic compartment and the axon conduction velocity can be particularly helpful in distinguishing parasol and midget cells (Li et al., 2015), the most numerous classes of RGCs in the primate retina (Dacey, 2004). Furthermore, additional differences in functional response properties such as firing rate (Sekirnjak et al., 2011) and temporal structure of spike trains (Rhoades et al., 2019) can be used to distinguish cell types, notably ON and OFF types within the midget and parasol cell classes (Sekirnjak et al., 2011; Richard, Goetz and Chichilnisky, 2015). Together, these anatomical, biophysical and functional differences yield ~85% accuracy in distinguishing the five numerically dominant cell types in the primate retina (Richard, Goetz and Chichilnisky, 2015). Further work is needed to understand how cell-type identification methods developed for primate retina would translate to humans (Section 6).

After identifying the spatial properties of visual processing for each cell (the receptive field), the other parameters of light response must be also estimated to fully identify the utility function. Let us consider a simple utility function based on the Linear Nonlinear Poisson (LNP) encoding model, which predicts RGC light responses by passing the visual stimulus through a cascade consisting of a spatio-temporal linear filter followed by an instantaneous output nonlinearity and Poisson spike generation (Chichilnisky, 2001). The temporal filter and output nonlinearity shapes are very similar within a cell type; hence, these parameters can be estimated for all the cells of a given type simultaneously (Chichilnisky and Kalmar, 2002). To estimate these parameters in a blind retina, we can use cell-type specific parameters from previously recorded healthy retinas. However, there is substantial variation of these properties across different healthy retinas. Below, we describe an approach to address the problem of variability across retinas, not only for the LNP encoding model described above, but also for a broader class of encoding models.

2.3 Characterizing diversity in neural responses

Most studies of neural processing in the retina focus on the features that are similar across retinas, often ignoring variability or treating it as random noise. However, visual response properties measured in many primate retinas exhibit clear structure in their variation (Shah, Brackbill, et al., 2019). For simplicity, we describe our approach for studying this variability by modeling the utility function \( f(.) \) using a response encoding model \( g(S) \) and a log-likelihood
response similarity metric \( \rho(\cdot) \), as described in Eq (2). A multi-layered convolutional neural network (CNN) encoding model learned using data from 95 retinas failed to capture responses in all of the individual retinas simultaneously, because of inter-retina variability. To better capture this variability, a low-dimensional representation was learned with parameters to capture the individual properties of each retina in the collection, along with a single shared CNN with parameters common to all retinas (Figure 2). In order to capture the responses in a particular retina, the different CNN layers were scaled using weights determined by the entries of a vector representing the location of the retina in the low-dimensional manifold (Shah, Brackbill, et al., 2019). This manifold revealed a smooth variation in response properties such as the receptive field size, the dynamics of the temporal filter, the degree of output nonlinearity and the autocorrelation of responses, for two major cell types. The manifold also revealed surprising systematic differences in the visual response properties of RGCs in male versus female retinas. In addition to these variations, the manifold preserved the previously reported invariances across retinas, such as larger receptive field size and faster temporal integration of ON parasol cells compared to OFF parasol cells (Chichilnisky and Kalmar, 2002).

The low-dimensional characterization of variability makes it possible to leverage hundreds of healthy retina recordings to efficiently identify the original visual responses in the retina of a blind person. The appropriate RGC encoding model, and hence a utility function \( f(\cdot) \), can be identified by estimating an appropriate low-dimensional manifold location for the retina that the brain is expecting in a psychophysical discrimination task. At each step, the set of plausible manifold locations for characterizing visual behaviors is updated as follows: (i) sample a few of the plausible locations and generate corresponding retinal responses for a fixed visual stimulus, (ii) stimulate the artificial retina to reproduce these responses, (iii) query the subject about which response produces the most accurate sensation for a verbally described object, and (iv) update the set of plausible manifold locations. In simulations of this task, the set of plausible manifold locations converged to the location of the unknown retina in a small number of steps (Shah, Brackbill, et al., 2019). Note, however, that the process could be substantially longer in a more realistic clinical application due to differences between the simulations using healthy primate retinas and the degenerated human retina (see Section 6). In addition to the behavioral task, the manifold permits other approaches for estimating neural encoding in a previously unseen retina using any physiological measurements that may be available (Figure 2).
Figure 2. **Identifying the neural code using a low-dimensional manifold of inter-retina variability.** Points in the manifold (gray dots) reflect measured neural encoding by different retinas, whereas other manifolds represent neural encoding not experimentally observed. To estimate neural encoding in a new retina, its manifold location is determined, leveraging data from previous retinas in several possible ways: averaging (red): computing the mean location of all previously observed retinas; approximation (orange): computing the mean location only of retinas with features that are similar to the new retina; optimization (green): selecting the manifold location that maximizes response prediction accuracy obtained with limited data from the new retina.

3. **How does the retina respond to electrical stimulation?**

The second major computational challenge is to measure how cells respond to electrical stimulation, $R \sim M(C)$ (Figure 1). Ideally, the choice of electrical stimuli should be effective in reproducing the desired neural activity, the evoked responses should be accurately captured by a simple model, and the parameters of this model should be efficiently estimated using few measurements. To advance these goals, the problem of electrical response modeling can be broken down into several computational steps: choosing the electrical stimuli (Section 3.1), identifying neural responses from recorded voltage traces after electrical stimulation (Section 3.2, 3.3) and modeling the electrically evoked responses (Section 3.4).

3.1 **Choice of electrical stimuli**

Ideally, the electrical stimuli used should be effective in reproducing the desired neural activity. Electrical stimulus consists of arbitrary current waveforms and amplitudes passed through an arbitrary combination of electrodes. Attempting to probe all possible stimuli produces a combinatorial explosion, thus, we start with a smaller, yet diverse collection of stimulation patterns. The simplest stimulation pattern is one in which a brief pulse of current (typically,
charge-balanced, peak amplitude ~1 μA, duration ~0.1 ms; (Sekirnjak et al., 2007)) is passed through a single electrode. In our lab prototype, single-electrode stimulation using 512 electrodes with 40 distinct current amplitudes spans 40*512 current patterns, a small fraction of the total of 40^512 possible patterns, thus avoiding the combinatorial explosion.

However, this simple starting point may not reveal the full ability of the device to replicate the neural code. Although an exhaustive exploration of more complex electrical stimuli is infeasible, the biophysics of electrical stimulation provides a guide to choosing a somewhat richer set of stimuli. One approach is to focus on selectivity, i.e., the ability to stimulate targeted cell(s) while avoiding others (Grosberg et al., 2017). Several studies have shown that replacing the distant electrical ground in single-electrode stimulation with “local return” electrodes can reduce the spread of the electric field, enhancing selectivity (Dommelel et al., 2005; Palanker et al., 2005; Wong et al., 2009; Abramian et al., 2011; Joarder et al., 2011; Cicione et al., 2012; Habib et al., 2013; Palanker, 2014; Flores et al., 2016; Sinclair et al., 2016; Spencer, Fallon and Shivdasani, 2018), including in recent work performed at cellular resolution (Fan et al., 2019). Also, more general “current steering” methods aim to improve selectivity further by passing currents through multiple nearby electrodes in a way that systematically shifts the spatial focus of electrical activation (Townshend and White, 1987; Sweeney, Ksienski and Mortimer, 1990; Bonham and Litvak, 2008; Butson and McIntyre, 2008; Martens et al., 2011; Habib et al., 2013; Matteucci et al., 2013). Similarly, other methods modify the temporal properties of activation by trading off current amplitude with pulse duration (Boingrov, Loudin and Palanker, 2010; Rattay, Paredes and Leao, 2012; Lee, Eddington and Fried, 2013; Boinagrov et al., 2014; Weitz et al., 2015).

Note that although most of these electrical stimuli have been developed for improved selectivity, it is possible that aspects other than selectivity are more important for reproducing the desired visual sensations (see Section 4).

3.2 Spike sorting in the presence of electrical artifact

Determining how effectively the electrical stimulus activates each cell is accomplished by passing current, recording the resulting voltage traces on all the electrodes, and identifying any elicited spikes. Unfortunately, the recorded voltage traces are typically corrupted by electrical stimulation artifacts, the properties of which are a complex function of the hardware, tissue and the electrical stimulus itself. The goal of spike sorting in this setting is to estimate and subtract stimulation artifacts to obtain the true physiological signal, and then identify in this signal the waveforms of spikes produced by cells that were previously identified in the artifact-free spontaneous activity. Typically, manual spike sorting to identify the responses to a particular electrical stimulus over a range of amplitudes requires several minutes for a trained human. For thousands of electrodes and multiple cells that are potentially activated by each electrode, manual spike sorting each electrode-cell pair is extremely time-consuming. Hence, automation of this step is important, but it is challenging (Hashimoto, Elder and Vitek, 2002; Wagenaar and Potter, 2002; Heffer and Fallon, 2008; Erez et al., 2010). In one automated approach applied to single-electrode stimulation, a prior on artifact shape to enforce smooth changes across successive current amplitudes, and a fast binary pursuit spike matching approach, yielded
>90% accuracy in estimating the activation probability compared to manual analysis (Mena et al., 2017). Modification of this spike sorting algorithm for more complex stimulation patterns would require a better understanding of the artifact shape and how it varies for a wide range of electrical stimuli. An ideal alternative would be to find electrical stimuli that are effective at eliciting desired activity but also have much smaller stimulation artifacts, simplifying spike sorting (Brown et al., 2008; Wichmann and Devergnas, 2011; Hottowy et al., 2012).

3.3 Identifying and avoiding axonal activation of distant cells

The approach of electrically stimulating each cell to match its healthy visually evoked responses is not applicable when the location of the cell body, and thus the location of its original visual receptive field, cannot be determined. This happens with electrical stimuli that activate the axon of a RGC with a cell body that is away from the recording array and which thus cannot be located. Indeed, in existing clinical implants (Humayun et al., 2012), simultaneous activation of many cells via their axons results in large, arc-shaped percepts determined by the distant locations of their cell bodies (Nanduri, 2011). Detecting and avoiding this problem can be accomplished by exploiting the unique spatio-temporal footprint of axonal activation on a high-density electrode array: a voltage waveform originating at the stimulation site and propagating bi-directionally along the axon’s trajectory (Grosberg et al., 2017; Tandon et al., 2017). Note that this technique is particularly effective in the retina and other systems in which the prevalent direction of axons is known.

3.4 Modeling the response to electrical stimulation

To fully exploit the capabilities of a high-resolution artificial retina, it would be helpful to have a simple description of the relationship between electrical stimuli and spiking activity, in order to predict RGC responses to arbitrary stimulation patterns. For single-electrode stimulation, a sigmoidal relationship between current amplitude and the activation probability of a given cell has been effective (Sekirnjak et al., 2008; Fried et al., 2009; Jepson et al., 2013). However, both the threshold and sensitivity of this relationship can vary greatly at different locations near the cell body, axon, and dendrites (Figure 3), and for large current values, inactivation of cells at high current levels may require a non-monotonic functional form (Meng et al., 2018). As with the choice of electrical stimuli (see Section 3.1), modeling of electrical responses can benefit from biophysical insights, such as the different electrical properties of the axon and soma (Rattay, 1986, 1989; Rattay, Paredes and Leao, 2012). However, electrical responses also depend on properties that are not easily measured, such as the spatial relationship between the cell and electrode, and the anisotropic extracellular environment consisting of axon fibers, other cells, and the inner limiting membrane. Hence, the benefit of exact biophysical models for understanding extracellular stimulation is less obvious than it is for intracellular stimulation.

A possible alternative is to derive the mathematical form of the relationship between current and cellular activation using biophysical models, and then fit the unknown model parameters using data obtained by stimulating and recording. This approach has been useful for comparing somatic and axonal stimulation (Fohlmeister and Miller, 1997; Rattay, 1999) and for
characterizing the efficacy of stimulation with current passed through two nearby electrodes (Jepson, Pawel Hottowy, et al., 2014). In the latter study, biophysical evidence suggests that the RGC response probability can be modeled as a combination of sigmoids, each over a different linear combination of stimulation currents (Esler et al., 2018). Each sigmoid is hypothesized to be associated with a distinct spike initiation site on the cell, depending on which electrode dominates activation. Each spike initiation site is associated with a different linear tradeoff in the current passed through the two electrodes to reach activation threshold, resulting in a piecewise linear iso-response curve (Jepson, Pawel Hottowy, et al., 2014).

In the future, electrical response models could build upon methods that have been used to estimate visual response models. For example, electrical response models for bi-electrode stimulation can be fitted to experimental data using techniques developed for estimating visual receptive field “subunits” (Kaardal et al., 2013; McFarland, Cui and Butts, 2013; Freeman et al., 2015; Maheswaranathan et al., 2018; Shah et al., 2020). These subunits affect neural responses through a nonlinear combination of linear summations of inputs, broadly similar to the effect of multiple electrical activation sites, though the underlying mechanism is different (synaptic rectification due nonlinear biophysics of transmitter release). In addition to the model architecture, tailored stimuli for the efficient estimation of model parameters can also build on methods developed for visual response characterization. Recently, the method used for identifying an LNP visual response model using white noise stimuli (Chichilnisky, 2001) has been adapted for electrical stimuli (Halupka et al., 2017; Ho et al., 2018; Rathbun et al., 2018).

3.5 Efficient calibration

In all of the above approaches, efficiency is critical. For example, in the laboratory prototype experiments, measurement of only the responses to single-electrode stimulation (512 electrodes X 40 current amplitudes) requires ~1.5 hours of experimentation, and calibrating the evoked responses to a large fraction of the distinct possible multi-electrode stimulation patterns is not feasible. One approach to increasing efficiency is to adaptively select stimulation patterns based on previous measurements. In the case of single-electrode stimulation, it would be natural to focus on intermediate current levels that lead to activation probability near 0.5, to rapidly constrain an estimate of the sigmoidal activation curve that relates current to spike probability. However, doing this in closed-loop requires computationally expensive steps of spike sorting and response modeling with each measurement. An alternative approach with fewer computational requirements is to use electrical stimulation properties measured in previous recordings as priors. For spike sorting, the similarity of recorded spikes and artifacts across experiments can be exploited to reduce the dimensionality of the problem (Shah, Madugula, et al., 2019). For response modeling, the inverse relationship between recorded spike amplitude and activation threshold can be used to estimate activation curves with fewer trials (Madugula et al., 2019). Importantly, using these priors reduces the required number of measurements by an amount similar to the reduction produced by adaptive stimulation, without requiring closed-loop computation (Shah, Madugula, et al., 2019). Additional work will be
needed to understand the benefits of these priors in the presence of possible implant instability over time or greater diversity of responses across cells in a blind person.

Figure 3. Properties of electrically evoked responses. Top panels show the measured receptive fields of ON and OFF parasol cells (blue, red) represented by ellipses, in a particular retinal recording. Electrical stimulation (location shown in green) with a single electrode at a particular current level activates a small number of cells (shaded). Different electrodes and different current levels activate cells in various spatial arrangements (A,B,C). Bottom panels show activation curves of several cells (colors) obtained using three different electrodes (D,E,F), revealing the different degrees of selectivity at each electrode. Selectivity is revealed in the ability to activate a single cell at low current levels without activating others (D). Shaded region indicates the current range for which distant cells are stimulated at their axons. While in this representative sample, only OFF cells are activated without stimulating distant cells, ON cells are also frequently activated.

4. How do we optimally replicate RGC firing in a blind retina?

A third major computational problem for an advanced artificial retina is using the visual and electrical response properties determined above to optimize electrical stimulation, for a given target visual stimulus. Ideally, we aim to reproduce the desired neural activity cell-by-cell and spike-by-spike. However, even the most advanced neural interfaces are likely to have imperfect selectivity, and it is not a priori obvious how to measure and optimize performance when the desired neural responses cannot be replicated exactly. This is because different kinds of stimulation errors would be expected to influence visual perception in different ways. For example, due to the spatial correlations in natural images, it may often be acceptable to simultaneously activate two nearby cells of the same type with little loss of resolution, but activating cells of different types (e.g. ON and OFF cells) may be more problematic. Hence, it is
desirable to identify a utility function $f(S, R)$ that measures device performance in a way that is related to the accuracy of the elicited visual perception. In the above example, the utility function would typically not greatly penalize the simultaneous activation of the two nearby cells of the same type, but would penalize activation of different cell types more strongly.

Given an appropriate utility function, the next challenge is optimizing it in real time, under the constraints of the neural interface. This problem can be solved using standard methods if there exists an electrical response model $M$ that correctly predicts the neural responses evoked by arbitrary electrical stimuli. However, as described in Section 3, $M$ is difficult to estimate from measurements with a limited collection of electrical stimuli.

One approach is to exploit the relatively slow integration time of the visual system by rapidly presenting a diverse sequence of electrical stimuli from a limited, calibrated dictionary, so that the elicited responses are effectively summed in downstream neural circuits to produce accurate visual perception. Note that the calibrated dictionary allows both axonal and somatic activation, but avoids high current levels that produce off-array axonal activation, because in these cases the receptive location, and hence the desired visual response, cannot be identified. The temporal dithering approach exploits the fact that electrical stimulation hardware is fast ($\sim 10$ kHz) and evoked spike latencies in response to direct electrical stimulation are precise ($\sim 0.1$ ms), while synaptic integration of neural inputs downstream is comparatively slow (e.g. $\sim 10$ ms) (see Figure 4) (Sekirnjak et al., 2008; Jepson et al., 2013; Kandel, 2013; Shah, Madugula, et al., 2019). The underlying assumptions are that the evoked RGC responses are generated independently over time during the rapid stimulation sequence, and that the resulting visual perception depends only on the integrated neural activity downstream. With this assumption, the problem in Eq (1) reduces to finding a sequence of dictionary elements $(C_1, \ldots, C_n)$ such that the summation of corresponding probabilistically generated responses $(\sum_{i=1}^{n} R_i)$ maximizes the expected utility function $f(S)$ with the target visual stimulus $S$:

$$C_1, \ldots, C_n = \arg\max_{i=n} E_{R_i\sim M(C_i)} f(S, \sum_{i=1}^{n} R_i)$$

With this framework, optimization of the device comes down to selecting the appropriate stimulation sequence in real time. Perhaps the simplest approach is a greedy one, in which an electrical stimulus is chosen at each time step such that it maximizes the increment to the utility function. We refer to this approach as greedy temporal dithering (GTD).

The visual perception produced by the GTD approach was analyzed using the experimental lab prototype, with data from ON and OFF parasol cells in isolated (healthy) primate retina (Figure 4) (Shah, Madugula, et al., 2019). Visually evoked responses were used to estimate the parameters of the utility function $f(.)$, assuming as a cost the mean-squared error between the target visual stimulus and the predicted perception obtained by linearly decoding the RGC responses evoked by electrical stimulation. A coarse white noise stimulus, which has some of the spatial correlations in natural images due to large pixel size, was used to estimate the
optimal linear decoding filters. Electrically evoked responses were measured using a dictionary of single-electrode stimulation patterns: each dictionary element corresponded to current passed through one of the 512 electrodes at one of 40 amplitudes. For each dictionary element, the response probabilities for every cell were estimated by applying automated spike sorting (Mena et al., 2017) to recordings of multiple presentations of the electrical stimulus. Then, for a target visual image, a sequence of dictionary elements was selected in rapid succession, to greedily maximize the increment to the utility function. Simulated RGC responses to this sequence were generated using the calibrated (measured) activation probabilities associated with each electrical stimulus. The resulting visual perception was estimated by linearly decoding the image from these simulated RGC responses using the STA of each cell as the decoding filter. Across multiple target images and sampled responses, the GTD algorithm showed nearly optimal performance compared to the best possible temporal dithering sequence (Shah, Madugula, et al., 2019). This result implies that for the development of an implanted device, more accurate visual restoration performance does not rely on better algorithms, but instead requires a higher resolution interface with the retina to produce a richer stimulation dictionary. Further work is needed to experimentally validate the main assumptions behind the GTD approach, notably the independence and additivity of responses elicited by successive electrical stimuli.

Surprisingly, in this test the GTD algorithm tended to select electrodes that activated multiple cells simultaneously, rather one cell selectively (Shah, Madugula, et al., 2019). This observation is in contrast to previous efforts to optimize the selectivity of activation, i.e. activating one or a few target cells while not activating non-target cells (see Section 3 above) (Fried, Hsueh and Werblin, 2006; Wong et al., 2009; Freeman et al., 2010; Jepson et al., 2013; Jepson, Paweł Hottowy, et al., 2014; Lee and Im, 2019). Note that the preference for less selective electrodes was also observed using an alternative algorithm that provides a lower bound for the optimal algorithm; hence, the preference of non-selective stimulation is not merely a consequence of the greedy selection procedure. A possible interpretation is that when perfect selectivity is not possible, it is often beneficial to exploit the simultaneous activation of multiple nearby cells. For example, this could be helpful when the incoming image has a spatial structure which is well-matched to the collection of cells activated by a particular electrode. The idea that selectivity is not necessarily the most important goal could be further examined by applying the GTD approach using more selective electrical stimuli, e.g. those with local returns (see Section 3.1 above), and testing the impact on decoding. Because the GTD framework and lab prototype permit this kind of simulation of visual perception, the approach can also be used to study the robustness of electrically evoked responses to phenomena such as array movement, cell death and changes in stimulation efficacy over time.

Applying the GTD framework to data from the lab prototype also provides novel insights into the design of an implantable device. For example, restricting stimulation to the 50% of the most frequently used electrodes resulted in a negligible decrease in overall performance. This observation suggests an adaptive design, in which a high-density electrode array is implanted,
and half of the electrodes are subsequently turned off based on the measured electrical response properties, resulting in reduced power consumption and heat dissipation.

While the GTD framework is described above for a static visual target, the approach can be extended to natural vision. In natural viewing, the visual scene is scanned by eye movements that position the fovea at many different locations to obtain high-resolution visual information. Similarly, because the implant is small compared to the entire visual scene, it is expected that the implant user will scan the visual scene with eye movements (Sabbah et al., 2014). To simulate this process, the GTD framework can be applied in sequence at every saccade location with the corresponding region of visual scene as the target, and the overall perception of the scene can be simulated by summing the reconstructed image at these locations over space and time.

Figure 4. **Temporal dithering of electrical stimulation.** Electrical stimuli are delivered in rapid sequence (left column, circle size indicating the current amplitude), under the assumption that the stimulus interval is much smaller than visual integration time of the brain. Each stimulus generates a corresponding stochastic response (second column; see Figure 3). Each response contributes to perception according to the identified properties of the stimulated cell(s) (third column), assuming optimal linear stimulus decoding. Downstream neural circuitry is assumed to sum these contributions over time to produce perception (last column), capturing the spatial structure of the target visual stimulus (top right).

5. **System design for implementing the computational steps**

To translate these computational insights into greater acuity of artificial vision, the algorithms must be implemented within hardware constraints such as power, area and communication
bandwidth and latency. Figure 5 presents a system-level implementation consisting of three components: (i) an external processor with a video camera and eye position sensor, (ii) a data and power relay, mounted on the outside of the eye, and (iii) an electrode array interface implanted on the retina. Compared to the external processor, the implanted array has severely constrained area and power limits, for surgical and safety reasons (< 1mW / mm²) (Mathieson et al., 2012), as well as limited bandwidth and latency for wireless data and power transmission, because of the variable efficiency of the wireless link as the eye scans the visual field. The relay mitigates these limitations by providing power for intermediate computations and improving the efficiency of wireless transmission by being close to the processor and not moving relative to the implant. For visual and electrical response calibration, the recorded data are transmitted from the implant via the relay to the processor. Conversely, during runtime function, the target visual stimulus is captured by the external camera, and information about the stimulation sequence is sent from the processor via the relay to the interface.

A distributed implementation of the algorithms described earlier across different hardware components may enable continuous operation of the artificial retina within the constraints of power, processing, memory, and communication. For example, assuming a moderate communication bandwidth between the processor and the relay, and a lower bandwidth between the relay and implant, a possible distributed implementation of the GTD algorithm could be: (i) select a small subset of the electrical stimulation dictionary on the external processor, based on the target image and the measured eye position immediately after a saccade; (ii) send this reduced dictionary to the relay; (iii) use the reduced dictionary to compute the stimulation sequence on the relay; and (iv) send the stimulation sequence in real time to the interface. Further improvement is possible by precomputing the dictionary for various saccade locations.

If implementation of the proposed algorithms is difficult within the constraints of existing hardware, co-design of hardware and algorithms may permit a smoother trade-off between efficiency and accuracy. A potential example of hardware-algorithm co-design is based on the properties of eye movements that are crucial for normal visual perception. Recent studies have shown that the brain uses the statistics of microsaccadic eye movements, not their exact trajectory, to decode visual stimuli from RGC responses (Ratnam et al., 2017). This suggests the possibility of relaxed hardware requirements for the precision of measured eye movements, because adding a simulated (rather than measured) micro-saccade trajectory to the coarsely measured saccade trajectory may suffice for high-resolution artificial vision.

Another example of hardware-algorithm co-design is the development of compression for electrical recordings at the neural interface, to reduce data processing and communication bandwidth. Naive front-end lossy compression approaches could backfire by compromising subsequent computational steps, such as spike sorting and cell type classification. Instead, co-designing the hardware and algorithm using data from the lab prototype allows a balance between compression and performance of the downstream computational steps. Recently it has been shown that exploiting the sparsity of spike data by
projecting analog voltage traces non-uniquely across the rows and columns of the interface electrodes can produce ~40X compression with minimal loss of relevant spike data (Muratore et al., 2019). Similarly, spike sorting algorithms for electrical stimulation can benefit from hardware innovations resulting in reduced artifacts or better priors on the artifact waveform. Thus, system and hardware design may be improved insights obtained using the lab prototype.

Figure 5. System architecture for the artificial retina. The interface is implanted on the retina near the fovea, at the rear of the eye. A relay is mounted on the outside of the eye. An external processor, an image sensor and an eye sensor are housed in goggles. During calibration, information from recording is sent from the interface, via the relay, to the external processor. During runtime, information about stimulation is sent from the external processor, via the relay, to the interface.

6. Experimental Limitations and Caveats

The work described here builds on experimental data obtained from healthy isolated non-human primate retinas, using a laboratory prototype system for high-density recording and stimulation. Although this system provides major insights, it differs in important ways from the clinical goal: an artificial retina implanted in the eye of blind human subjects. Hence, future work must focus on computational approaches that are robust to these differences.
First, this work focused on macaque monkey retinas, because the gross structure, collection of cell types, and visual capabilities and behaviors of macaques are closer to those of humans than any other animal model. However, recent unpublished physiological work on human RGC types (Cowan et al., 2019; Reinhard and Münch, 2019) suggests that further investigation may be needed to understand whether the functional properties of the corresponding cell types in the two species are as similar as their morphology suggests (Dacey, 1993). Specifically, the computational goal would be to understand the normal function of a given human retina as a sample on the manifold obtained using many primate retinas (Section 2.3).

Second, during retinal degeneration, significant changes in network (Jones et al., 2016) and intrinsic RGC properties could affect the artificial visual signal transmitted to the brain (Spencer et al., 2018). Compared to stimulation methods that interface to the upstream circuitry, the direct epiretinal stimulation approach could be more robust to these changes, but challenges remain. Previous studies have reported oscillatory activity (Goo et al., 2015) and changes in spontaneous activity (Trenholm and Awatramani, 2015) in rodent models of retinal degeneration. Furthermore, in extreme cases of retinal degeneration, the excitable RGCs are found in patches, with only a small fraction of cells remaining in extreme cases (Ren et al., 2018; García-Ayuso and Di Pierdomenico, 2019)(García-Ayuso et al. 2019). The impact of these changes on the present methods is uncertain. Adding evoked spikes to spontaneous spikes could produce a visual signal that is contaminated by noise, however, it is possible that spontaneous firing will return to normal after extended ongoing electrical stimulation, and future approaches may make it possible to suppress firing with electrical stimulation. Degenerated patches of RGCs could create permanent islands of missing visual function, however, it is possible that eye movements will compensate by directing the image to intact locations. Resolving these concerns will require tests in primate models of degeneration.

Third, the methods used for obtaining high quality recordings using isolated retinas may not apply directly to the in vivo setting. For example, novel methods must be developed to establish close electrical contact between the implant and the retina. Additionally, if the implant or the retina are not stable over long durations (which cannot be tested with isolated retina), frequent re-calibration of electrical and visual response properties may be required.

Finally, results obtained in the peripheral retina (where most recordings are performed) should be interpreted with caution, because calibration of electrical responses will likely depend strongly on the region of the retina that is targeted and the electrical contact between the device and retina (see Section 2). Although the central retina is an attractive clinical target because the higher density of cells provides higher resolution natural vision, high cell density could also make it harder to selectively stimulate a targeted cell electrically while avoiding other nearby cells. These considerations raise a key question for device placement: does stimulation with low selectivity in the central retina produce more or less faithful vision restoration than stimulation with high selectivity in a more peripheral location? The raphe region of the retina (~2-5 mm from the fovea on the horizontal visual meridian; roughly 8-20° eccentricity), presents a possible compromise: in the raphe, cell density is relatively high, but RGCs do not form multiple layers as
they do in the fovea, leading to easier electrical access using a flat electrode array. Furthermore, because axons avoid the raphe region on their way from the peripheral retina to the optic disk, activation of distant cells via their axons is less of a problem (Grosberg et al., 2017). Investigation of the electrical and visual response properties of the raphe and surrounding regions is an important topic for future investigation (Gogliettino et al., 2019).

7. Broader implications and future work

Once a high-resolution, bidirectional artificial retina is built, the computational steps described here may guide new research into the mechanisms of visual processing. Examples include comparing the artificial vision produced by different RGC types and probing the perception resulting from the interaction between natural and artificial vision (Arens-Arad et al., 2020). Moreover, the presented framework also raises new questions about the relationship between advances in the basic science knowledge of visual processing and the engineering of a neural interface. What is the simplest visual utility function (f), that can be implemented efficiently in hardware, for a given electrical response model (M)? Conversely, what choices of electrical stimuli yield high performance for a variety of visual utility functions (f) and hardware implementation constraints?

The computational problems and methods introduced here are specific and concrete, in large part because the retina is a relatively well-understood neural circuit with well-defined function in vision. Related concepts may be applicable to other neural circuits, such as those mediating auditory, visual, or somatosensory capacities, and to other recording and stimulation technologies, such as optical, ultrasound or magnetic approaches. Adapting these ideas to different neural systems will depend on identifying the corresponding utility function (f), either as an encoding model, a decoding model or a joint embedding of stimuli and responses. Similarly, adapting to different recording and stimulation modalities may be possible by first measuring the corresponding response model (M) and compensating for possible non-selectivity and the high-dimensional space of neural stimulation patterns using temporal dithering.

Future work on the artificial retina and other neural interfaces may require revisiting some of the assumptions and the technical choices made in this work. First, we have proposed stimulation choices based on pre-calibration obtained with recording. An ideal high-performance closed-loop system (Krook-Magnuson et al., 2015) could further compensate for the stochasticity of electrical stimulation by passing current, recording the resulting voltages, identifying the activated cells, and modifying the subsequent electrical stimulus, all in real-time. The simpler approach presented here involves choosing electrical stimuli based on improvement in the expected utility function, averaged over possible patterns of evoked neural responses, based on a calibration performed in advance. While this approach is less computationally intensive and easier to implement in hardware, it is unsuitable if the expected utility function diverges significantly from evoked responses on individual trials. In the artificial retina application, because each RGC is activated independently of other RGCs and the “memory” of electrical stimulation is limited (sub-millisecond), the utility function on individual trials is close to the expected utility function. However, the use of the expected utility function
should be revisited for application in other neural circuits, where network-mediated recurrence could lead to increased inter-trial variability and interactions between stimulated neurons.

Second, we have assumed that the utility function is \emph{stationary}, i.e., visual response properties do not change over time. However, some degree of plasticity in visual response properties has been previously observed, for example, after retinal detachment and reattachment. Future work may need to address plasticity by adjusting to changes in the utility function. Furthermore, it may be possible to gradually drive plasticity with electrical stimulation, such that downstream processing adjusts to the properties of the implant. In other neural circuits, such as the cortex, plasticity may be more prominent and thus these issues may be more central. The simpler situation in the retina thus provides an important stepping stone in the design of bi-directional and high-resolution interfaces to other areas of the brain.

\textbf{Acknowledgements}

We thank Nora Brackbill, Sasidhar Madugula, Alex Gogliettino, Eric Wu, Raman Vilkhu, Sreela Kodali and Dante Muratore for helpful comments on the manuscript. We thank A. Litke, P. Hottowy, and S. Sher for technical development. We thank Liam Paninski, Yoram Singer, Jon Shlens, Subhasish Mitra, Boris Muramann, Marty Breidenbach and the entire Stanford Artificial Retina team for helpful discussions.

\textbf{References}


Smirnakis, S. M. *et al.* (1997) ‘Adaptation of retinal processing to image contrast and spatial
scale’, *Nature*, 386(6620), pp. 69–73.


