Global Activity Shaping Strategies for a Retinal Implant

Martin Spencer^{1†} Tatiana Kameneva^{1,2}, David B. Grayden^{1,3}, Hamish Meffin^{4,5*}, Anthony N. Burkitt^{1,3*}

1. Department of Biomedical Engineering, The University of Melbourne, Parkville, Victoria, Australia

2. Faculty of Science, Engineering and Technology, Swinburne University of Technology, Australia

3. Centre for Neural Engineering, The University of Melbourne, Parkville, Victoria, Australia

4. Department of Optometry & Vision Science, The University of Melbourne, Parkville, Victoria, Australia

5. National Vision Research Institute, Australian College of Optometry, Carlton, Victoria, Australia

* Denotes joint senior authorship

† Email: martin.spencer@unimelb.edu.au

Abstract

Objective: Retinal prostheses provide visual perception via electrical stimulation of the retina using an implanted array of electrodes. The retinal activation resulting from each electrode is not point-like; instead each electrode introduces a spread of retinal activation that may overlap with activations from other electrodes. With most conventional stimulation strategies this overlap leads to image blur. Here we propose a ‘shaping’ algorithm that uses multiple electrodes to manipulate the current between electrodes in a desired way.

Approach: We assume a forward model for the conversion of electrode strengths to retinal activation. Three alternative global shaping algorithms are developed by calculating reverse models under different assumptions: linear inversion using singular value decomposition to

produce the pseudoinverse, a linearly constrained quadratic program, and a binary quadratic program to partition the target pattern. The algorithms were assessed using both the mean squared error between the resulting images and desired images, as well as their adherence to the maximum allowed electrode currents.

Main Results: Under wide activation spreads the linear inversion algorithm gave improved solutions but faced two limitations: under low-noise conditions the electrode amplitudes exceeded their set limit; the set of solutions did not include the possibility of using negative local currents to induce retinal activation. The linearly constrained quadratic program and binary quadratic program respectively addressed these problems, but required much greater computation time.

Significance: This provides a framework for improving the resolution of future retinal implants, especially those with high density electrode arrays.

Introduction

Retinal prostheses are devices that restore a sense of vision in people with profound retinal degeneration. They use electrode arrays to activate the neural cells of the retina to create a particular visual percept [1–6]. The activation of each electrode via a biphasic current or pulse induces a single spatially localized patch of retinal activation referred to as a phosphene. A positive electrode current refers to a cathodic first pulse, and a negative electrode current refers to an anodic first pulse. Loss of rod and cone cells is followed by remodelling and degeneration of the retinal circuitry [7]. Consequently, the ultimate target for electrical stimulation is the retinal ganglion cells. It is only the direct electrical activation of retinal ganglion cells that is modelled in this study. Electrical stimulation of retinal ganglion cells is caused by current entering through the membrane at the site of spike initiation, which is

usually the sodium channel band of the axon initial segment [8,9]. An important limitation in recreating a particular pattern of activation on the retina is the spread of electric currents in the retinal tissue. The extent of activation spread is related to electrode size, the electrical conductivity of the tissues and the distance between the electrode array and the neurons of the retina [10,11]. Under narrow activation spread conditions, stimulation of the retina takes the form of a matrix-like array of localized phosphenes. Under wide activation spread conditions, the influence of each electrode overlaps with neighboring electrodes and leads to a perceived blurred image. The wide activation spread conditions are found in any retinal prostheses with high electrode densities whether they are suprachoroidal, epi-retinal, or subretinal implants. Available literature indicates that most devices fall into the wide current spread regime. Future devices are expected to have higher electrode densities and are therefore more likely to have wide current spread relative to the electrode spacing.

A number of strategies have been proposed to achieve a desired target pattern of retinal activation. The conventional stimulation strategy sets electrode amplitudes to correspond to the desired target activations at their locations on the retina [12–19]. Under this strategy, the electrode activation pattern is simply a down-sampled (pixelated) version of the desired retinal activation pattern, with the amplitude typically scaled between the threshold and maximum allowed levels. The conventional approach is a one-to-one strategy in that each electrode is assumed to influence one region of the retinal activation pattern. It is rapid to compute and produces predictable and reproducible patterns of stimulation. Under narrow activation spread conditions, this conventional approach leads to a retinal activation pattern that is a matrix of phosphenes, with the position and spread of each determined by the electrode positions and properties. We show that in narrow spread conditions, this is the optimal outcome possible. Using the conventional algorithm, the accuracy of the retinal

activation pattern is limited by the intrinsic spread of current in the retinal tissue, and this in turn leads to an upper limit on the useful electrode density in the implanted array [10].

Multipolar strategies are used when the activation spreads from electrodes overlap, referred to here as wide activation spread conditions. Multipolar strategies attempt greater control over the spread and position of retinal activation by using multiple electrodes together to manipulate a single point of retinal activation, a many-to-one strategy. Multipolar stimulation uses a group of two or more electrodes activated at varying ratios to manipulate the position of a 'virtual' phosphene situated at a point between the electrodes [20]. If a single electrode's current is increased then the peak retinal activation will move its position closer to that electrode. Focused multipolar stimulation attenuates the spread of current resulting from a central electrode by using a group of surrounding electrodes of opposite polarity [21,22]. This results in more confined virtual phosphenes. These solutions are intended to improve upon the limitations of the conventional solution under wide activation spread conditions by providing the ability to manipulate the position and spread of each phosphene. However the result is still a phosphene-matrix, and each phosphene now requires multiple electrodes.

Under a global current steering strategy, each electrode current amplitude is jointly and simultaneously dependent on the whole target pattern of retinal activation. This is a many-to-many strategy. The problem is not framed in terms of phosphenes, and the result is not a matrix of phosphenes. Each electrode plays a role in creating the whole pattern of retinal activation; any point in the pattern is, in principle, influenced by the setting of every electrode. The strength of a given electrode can be influenced by remote features in the target retinal activation pattern. This global current steering for retinal activity shaping

strategy has been applied in the case of trans-cranial direct current stimulation [23], cochlear implants to restore hearing [24,25], and in retinal implants [26–28].

The model we adopt is simple enough to be useful in stimulation strategies, in contrast to volume conductor models, but accurate, as demonstrated experimentally [9,26,29–33]. The model includes a quantitative representation of the Electrical Receptive Field (ERF), which is the pattern of electrical stimulation on the electrodes that leads to activation of each neuron. These ERFs could be measured experimentally using the same array as is used for electrical stimulation. The *activation spread* can be thought of as the inverse equivalent of the ERF. It is the pattern of retinal activation that results from the current from a single electrode.

The present investigation adds to existing work by exploring the fact that both positive and negative current induces retinal activity, allowing a new set of solutions. It also explicitly considers the consequences of variation in the influence of each electrode on the neural tissue and noise in the measurement of these activation spreads. In addition we address the selection of eigenvectors to reduce over-fitting, and explicitly investigate the dependence of solution quality upon the density of electrodes.

Three alternative global current shaping strategies were explored and compared to the conventional strategy. We began with an assumed forward model, which describes the effect of each electrode on the retinal activation. This model was inverted to create an algorithm that gives a set of electrode activations intended to re-create a target pattern of retinal activity. The inversions were completed under different assumptions and with different requirements. Qualitative results were examined to discover any artifacts that may create a disproportionately large perceptual effect that could be underweighted by a purely quantitative approach.

Methods

Forward Model of Electrical Stimulation of the Retina

Accurate quantifiable analysis of alternative global shaping strategies initially requires a forward model to calculate retinal activations resulting from a particular pattern of electrode settings. The model will use scaled units with S – spikes, T – time, I – electric current, L – length.

We first define a desired set of retinal activations, $\vec{r}_d^* [ST^{-1}]$, and a modeled set of retinal activations $\vec{r}_d [ST^{-1}]$ that result from a given set of electrode current amplitudes, $\vec{s} [I]$. The retinal activation vectors $\vec{r}_d^*$ and $\vec{r}_d$ represent the set of values associated with N_d discrete retinal locations. The current amplitude vector, $\vec{s}$, represents the values associated with N_e electrode locations. These sets of retinal activations and current amplitudes could be spatially configured in an arbitrary way, for example the set of electrodes could be configured as hexagonal arrays. However, in the present investigation, the retinal and electrode configurations take the form of square arrays of dimensions n_d , and n_e respectively, such that $N_d = n_d^2$, and $N_e = n_e^2$ and the spatial width of the array is $L_e [L]$ is set to be 1. A ground electrode is assumed to be distal, as is the case for most retinal implants. We also define the generator signal, $\vec{G} [I]$ which is an intermediate variable used to calculate retinal activations, with the same length, N_d , as the retinal activation vector. In the present investigation, the generator signal at any given location in the retinal tissue is the total local current at that location, which is the sum of the currents resulting from each electrode activation.

The forward model that converts electrode amplitudes, $\vec{s}$, to retinal activations, $\vec{r}_d$, is defined in two steps: First, a linear step that maps the vector of electrode amplitudes, $\vec{s}$, to a vector of generator signal values, $\vec{G}$,

$$\vec{G} = g(\vec{s}) , \quad (1)$$

And, second, a nonlinear step that acts on each element of the generator vector, $\vec{G}$, to give the retinal activation vector, $\vec{r}_d$,

$$r_{d,i} = f(G_i), \quad i = 1, \dots, N_d. \quad (2)$$

In the retina, the biophysical basis for the utility of a linear-nonlinear model and electrical receptive field have been demonstrated experimentally [9,26,29–33] and loosely correspond to the subthreshold membrane current, spike generation threshold, and field of network inputs, respectively. The functions f and g are defined below.

The function g in Eq. (1) is a transfer matrix:

$$\vec{G} = g(\vec{s}) = \mathbf{W}_d \vec{s} , \quad (3)$$

where, $\mathbf{W}_d$ is a matrix of dimensions $N_e \times N_d$, and gives the magnitude of the current at each retinal location resulting from the activation of each electrode. The rows of the matrix $\mathbf{W}_d$ are the ERFs for each retinal patch, and the columns contain the retinal activation spread for each stimulating electrode. Gaussian electrode activation patterns were chosen to approximate the activation spread resulting from each electrode [12], with each element, $W_{d,ij}$, given by:

$$W_{d,ij} = (1 + \rho \xi_{ij}) e^{\frac{-(x_i - x_j)^2 - (y_i - y_j)^2}{2\sigma^2}} , \quad (4)$$

where (x_i, y_i) $[L, L]$ is the spatial positions of the i th retinal patch, $\vec{r}_{d,i}$, and (x_j, y_j) $[L, L]$ the j th electrode, $\vec{s}_j$; σ $[L]$ is the spread of current, ρ is the magnitude of variation in the activation spread, and ξ_{ij} is a random number equally distributed between -1 and 1. The additive variation, $\rho\xi_{ij}$, represents deviations in the Gaussian shape due to heterogeneity in the retinal tissue.

In its general form, the function f (Eq. (2)) is the two-sided sigmoid function illustrated by the red line in Fig. 1(a) [9,28]. With appropriate limits placed on the range of retinal activations and by appropriate linear scaling, the function f can be approximated by a modulus $|G_i|$, as illustrated by the solid black line in Fig. 1(a). Using this approximation and Eq. (3), the final forward model described in Eq. (1) and Eq. (2) is now given by

$$\vec{r}_d = k \cdot |\mathbf{W}_d \vec{s}|, \quad (5)$$

where k $[ST^{-1}I^{-1}]$ is set to 1. This forward model is illustrated in one dimension in Fig. 1(b) – 1(e). However, in general, these spatial interactions take place in two-dimensions, giving the retinal activation pattern.

Figure 1: Schematic depictions of the forward model described by Eqs. (1)-(5). (a) Model of retinal activation as a function of the generator signal as used in the binary quadratic program, $r_{d,i} = |G_i|$ (2), (black line). This is compared to a hypothetical function in which the retinal activation saturates at higher values in the generator signal (red line). For the linear inversion and linearly constrained quadratic program, the relation $r_{d,i} = G_i$ is assumed (dashed line). (b) Example of five electrode amplitudes in one-dimension. (c) Activation spreads due to each individual electrode. (d) Generator signal. (e) Retinal activations (f) The overlap parameter γ as a function of the activation spread (σ) and electrode number (n_e).

The units of the elements of each the vectors, $\vec{r}_d$ and $\vec{s}$, and the matrix, $\mathbf{W}_d$, were scaled such that an electrode stimulus amplitude value of 1 [I] results in a peak generator signal value of

1 [I], which in turn results in a retinal activation of 1 [ST⁻¹] (Fig. 1(a)). In real implants, there is a limit on the maximum current $\vec{b}$ [I] deliverable by each electrode for both power and safety reasons.

There are two regimes in which the model can operate: the ‘narrow’ activation spread regime, in which the influence of neighboring electrodes is isolated and does not significantly overlap, and the ‘wide’ activation spread regime, in which the retinal activation at any given point is the result of significant contributions from multiple electrodes. In the present investigation, the overlap parameter, γ , was defined by the ratio between one standard deviation in activation spread and the distance between neighboring electrodes in the square array, $\gamma = \frac{\sigma}{L_e/n_e} = \frac{\sigma n_e}{L_e}$, and since we choose units with $L_e = 1$, then $\gamma = \sigma n_e$. This parameter was used to quantify the overlap, where $\gamma = 1$ approximately denotes the boundary between the narrow and wide activation spread regimes. The quantity, γ , and its dependence on σ , and n_e is illustrated in Fig. 1(f).

Global Shaping Model Development and Testing

The next stage in the development of the model requires derivation of an inverse model as illustrated in Fig. 2(a). From the forward model’s matrix $\mathbf{W}_d$ Eq. (5), (Fig. 2(b)) of size $N_d \times N_e$ we derive ‘measured’ activation spreads $\mathbf{W}_t$ (Fig. 2(c)) of dimension $N_t \times N_e$. This is a lower resolution and noisy version of $\mathbf{W}_d$. The matrix $\mathbf{W}_t$ is intended to represent an experimentally obtained ERF model and is the matrix used in the inverse model while the matrix $\mathbf{W}_d$ is intended to model the actual retinal activation that results from a given pattern of electrodes. Separating these two models is a conservative check on the results obtained by the methods proposed in this investigation. $\mathbf{W}_t$ was modeled by downsampling each of the 2D electrical receptive fields contained in $\mathbf{W}_d$ from N_d elements to N_t elements using the MATLAB 2017a

(Mathworks, MA, USA) function ‘interp2’, and then simulating additive noise by adding a quantity $\eta\xi_{ij}$ to each element of the resulting activation spreads, where η is the magnitude of the ‘estimation noise’ and ξ_{ij} is a random number between -1 and 1. This quantity represents errors in measuring the true activation spreads. This estimation noise is critical in assessing the quality of the algorithms because, unlike the conventional method, shaping methods are susceptible to over-fitting under certain conditions, which can lead to artifacts in the resulting retinal activation pattern.

Figure 2: Schematic depiction of the process used in the present investigation. (a) A desired retinal activation pattern $\vec{r}_d^*$ (defined at N_d spatial locations) is down-sampled to give a target retinal activation pattern $\vec{r}_t^*$ (defined at fewer; N_t spatial locations). An inverse model is used to calculate the electrode strengths $\vec{s}$ to induce this pattern. (b) The

error, ε_t , calculated using $\mathbf{W}_t$ is used to develop the inverse model in (a). (c) The error, ε_d , used to assess the outcome of each of the models calculated using $\mathbf{W}_d$.

All parameters are defined in table 1.

Table 1: List of parameters and their units of measurement. T – time, L – length, S – spikes, V – voltage, I - current.

Parameters			
Electrode Amplitudes (N_e)	$\vec{s}$ [I]	Electrode Array Dimensions	n_e
Maximum Electrode Amplitude	$s_{\max}$ [I]	Electrode Vector Length, n_e^2	N_e
Electrode Array Width, 1	$L_e = 1$ [L]	Neural Activation Constant	$k = 1$ [$ST^{-1}I^{-1}$]
‘True’ Transfer Matrix ($N_d \times N_e$)	$\mathbf{W}_d$ [$ST^{-1}I^{-1}$]	Electrode Locations	(x_j, y_j) [L, L]
Desired Retinal Array Dimensions	n_d	Retinal Locations	(x_i, y_i) [L, L]
Desired Retinal Vector Length, n_d^2	N_d	Linear Model	g
Desired Retinal Activations (N_d)	$\vec{r}_d^*$ [ST^{-1}]	Nonlinear model	f
Modelled Desired Retinal Activations (N_d)	$\vec{r}_d$ [ST^{-1}]	Random Value	$\xi \in (-1, 1)$
Mean Squared Error, $\ \vec{r}_d^* - \vec{r}_d\ ^2$	ε_d [S^2T^{-2}]	Generator Function Values (N_t)	$\vec{G}$ [I]
Current Spread	σ [L]	Current Overlap Parameter, $\sigma n_e/L_e$	γ
Variation Magnitude in $\mathbf{W}_d$	ρ	Linear Inversion Objective Function	Θ_{lin} [S^2T^{-2}]
Target Transfer Matrix ($N_t \times N_e$)	$\mathbf{W}_t$	Quadratic Program Objective Function	Θ_{quad} [S^2T^{-2}]
Target Retinal Array Dimensions	n_t	Binary Partition Objective Function	Θ_{bin} [S^2T^{-2}]
Target Retinal Vector Length, n_t^2	N_t	Eigenvalue Fraction	F_{eig}
Target Retinal Activations (N_t)	$\vec{r}_t^*$ [ST^{-1}]	Left Eigenvectors, retina ($F_{\text{eig}}N_e \times N_t$)	$\mathbf{P}$
Modelled Target Retinal Activations (N_t)	$\vec{r}_t$ [ST^{-1}]	Right Eigenvectors, electrode ($N_e \times F_{\text{eig}}N_e$)	$\mathbf{Q}$
Mean Squared Error, $\ \vec{r}_t^* - \vec{r}_t\ ^2$	ε_t [S^2T^{-2}]	Eigenvalues ($F_{\text{eig}}N_e \times F_{\text{eig}}N_e$)	$\mathbf{D}$
Estimation Noise Magnitude in $\mathbf{W}_t$	η	Pseudoinverse of $\mathbf{W}_t$ ($N_e \times N_t$)	$\mathbf{W}_t^+$
Binary Partitioning Vector (N_t)	$\vec{\varphi}$	Mean Error over 32 images	ε_{d32} [S^2T^{-2}]
Diagonal matrix form of $\vec{r}_t^*$ ($N_t \times N_t$)	$\mathbf{R}^*$ [ST^{-1}]	Partitioned retinal activity, $\mathbf{R}^* \vec{\varphi}$ (N_t)	$\vec{p}_t^*$ [ST^{-1}]
Maximum Allowed Electrode Amplitudes	$\vec{b}$ [I]		

The mean squared error, ε_t [S^2T^{-2}], (Fig. 2(b)) quantifies the magnitude of the error between a target retinal activation, $\vec{r}_t^*$, and the retinal activation resulting from electrical stimulation, $\vec{r}_t$:

$$\varepsilon_t = \|\vec{r}_t^* - \vec{r}_t\|^2 \quad (6)$$

This is used as an objective function, and each inverse model is developed by attempting to minimize this objective function under different assumptions.

As a test of the results of each inverse model, the electrode patterns are applied using Eq. (5), and the resulting pattern compared to the desired pattern to give ε_d [S^2T^{-2}]:

$$\varepsilon_d = \|\vec{r}_d^* - \vec{r}_d\|^2 \quad (7)$$

This process is illustrated in Fig. 2(c). The procedure is intended to replicate the real conditions by confirming that elements of the retina that are not included in $\mathbf{W}_t$ remain constrained by the inverse models that we develop.

In this investigation, 32 images were used as the basis for the desired retinal activation patterns. These images are shown in Fig. 3. The mean value of ε_d over over these 32 images was defined as ε_{d32} . The images were downsampled to give the desired retinal activation points N_d , and the target retinal activation points N_t . This was done using the MATLAB 2017a (MathWorks, MA, USA) function ‘imresize’ with the ‘antialiasing’ option, which uses a low-pass filter adapted to the resolution of the desired array dimensions.

Figure 3: The 32 images used during the investigation in the calculation of ε_{d32} .

The Conventional Algorithm

The conventional approach to calculating electrode activations in retinal implants is to simply downsample the desired activation pattern to give an array of retinal activation ‘pixels’ with the same dimensions as the electrode array. This is completed using the MATLAB function ‘imresize’ with the ‘antialiasing’ option. These electrode values, which are between 0 and 1 may lead to oversaturation of the retina, especially in the case of wide activation spread ($\gamma > 1$). To normalize the retinal activation, the initial electrode settings are provided to the forward model to calculate an initial set of retinal activations. The ratio of the resulting total mean retinal activation and the target total mean retinal activation is then used to scale the initial set of electrodes.

Linear Inversion Algorithm

In calculating the inverse model we make an approximation in which the modulus in Eq. (5) can be removed, and choose $k = 1$. The inverse model will then operate using the assumption that $r_i = G_i$, the dashed line in Fig. 1(a). The consequences of this approximation are discussed below. Replacing the true set of activation spreads, $\mathbf{W}_d$, with the measured set of activation spreads, $\mathbf{W}_t$, this approximation means that Eq. (5) becomes $\vec{r}_t = \mathbf{W}_t \vec{s}$. The error, ε_t Eq. (6), can then be re-written as an objective function, $\Theta_{\text{lin}}[S^2 T^{-2}]$:

$$\Theta_{\text{lin}} = \min_{\vec{s}_t \in \mathbb{R}} \|\vec{r}_t^* - \mathbf{W}_t \vec{s}\|^2. \quad (8)$$

The solution of this equation, $\vec{r}_t^* = \mathbf{W}_t \vec{s}$, can be arranged to give $\vec{s} = \mathbf{W}_t^{-1} \vec{r}_t^*$. However a true inverse of the matrix $\mathbf{W}_t$ is not possible because in general it is a non-square matrix, and we must use a pseudo-inverse, $\mathbf{W}_t^+$.

The pseudo-inverse of $\mathbf{W}_t$ is calculated using singular value decomposition (SVD). Under SVD, $\mathbf{W}_t = \mathbf{P}\mathbf{D}\mathbf{Q}^T$, where the matrices $\mathbf{P}$ and $\mathbf{Q}$ are unitary and contain the left and right eigenvectors of $\mathbf{W}_t$; these are associated with the generator signal pattern and the electrode pattern, respectively. $\mathbf{D}$ is a rectangular diagonal matrix and contains the eigenvalues associated with those eigenvectors. The pseudo-inverse of $\mathbf{W}_t$ can then be calculated as $\mathbf{W}_t^+ = \mathbf{Q}\mathbf{D}^{-1}\mathbf{P}^T$, and the electrode activation currents are given by

$$\vec{s} = \mathbf{W}_t^+ \vec{r}_t^*. \quad (9)$$

The number of eigenvectors used in the inversion can be varied, as explained in the results, and the fraction of eigenvectors used is defined as F_{eig} .

Quadratic Program

The linear inversion algorithm does not place constraints on electrode amplitudes. High electrode amplitudes can be avoided by using a quadratic program with linear constraints approach, which has the benefit of allowing the imposition of an explicit limit on the electrode current amplitudes.

A second inverse model can be developed by substituting $\vec{r}_t = \mathbf{W}_t \vec{s}$ in Eq. (6), and ignoring the constant term $\vec{r}_t^{*T} \vec{r}_t^*$ in the subsequent quadratic expansion. The error, ε_t Eq. (6), can then be re-written as an objective function, $\Theta_{\text{quad}}[S^2T^{-2}]$:

$$\Theta_{\text{quad}} = \min_{\vec{s}_t \in (-2,2)} \left(-2\vec{r}_t^{*T} \mathbf{W}_t \vec{s} + \vec{s}^T \mathbf{W}_t^T \mathbf{W}_t \vec{s} \right). \quad (10)$$

This equation fits the form of the standard quadratic program minimization, $\left(\min \left(c^T \vec{v} + \frac{1}{2} \vec{v}^T \mathbf{M} \vec{v} \right); \mathbf{A} \vec{v} < \vec{b} \right)$, with $\mathbf{M} = 2\mathbf{W}_t^T \mathbf{W}_t$ and $c^T = -2\vec{r}_t^{*T} \mathbf{W}_t$. The problem is convex because $2\mathbf{W}_t^T \mathbf{W}_t$ is positive semi-definite. Each element of the vector of electrode

strengths, $\vec{s}$, must be limited to conform to the maximum allowed electrode amplitudes. To achieve this two-sided boundary, $\mathbf{A}$ is set to be a pseudo identity matrix with values of 1 in the top left quadrant and values of -1 in the bottom right quadrant, and $\vec{b}$ is the vector of the limits on the electrode strength magnitudes. With these substitutions, the MATLAB 2017a (Mathworks, MA, USA) function 'quadprog' was used to solve Eq. (10). This function uses an interior point convex method [34,35].

It is important to note that removing the modulus from the forward model (Eq. (5)) in order to derive the two inverse models (the linear inversion model (Eq. (9)) and the quadratic programming model (Eq. (10))) prevents these inverse models from finding solutions where negative generator signal could have been utilized to induce retinal activation. A binary quadratic program approach restores the modulus to the inverse model allowing it to discover these negative generator signal solutions.

Binary Quadratic Program

The modulus in Eq. (5), which was removed in calculating the linear inversion and quadratic program inverse models, can be retained in the calculation of the inverse model. Using Eq. (5) with $k = 1$, and $\mathbf{W}_d$ replaced with $\mathbf{W}_t$, we wish to solve $\vec{r}_t^* = |\mathbf{W}_t \vec{s}|$ for $\vec{s}$. This is equivalent to:

$$\mathbf{R}^* \vec{\varphi} = \mathbf{W}_t \vec{s}, \quad (11)$$

where $\mathbf{R}^*$ is a diagonal matrix containing the vector, $\vec{r}_t^*$, and $\vec{\varphi} = \text{sgn}(\mathbf{W}_t \vec{s})$, which is here referred to as the binary partitioning vector. We use $\vec{p}_t$ [$S^1 T^{-1}$] as the notation for the partitioned retinal activation resulting from a set of electrode amplitudes, and its target value $\vec{p}_t^*$ [$S^1 T^{-1}$] is given by

$$\vec{p}_t^* = \mathbf{R}^* \vec{\varphi} = \vec{\varphi}^T \mathbf{R}^*. \quad (12)$$

The solution to Eq. (11) is $\vec{s} = \mathbf{W}_t^+ \mathbf{R}^* \vec{\varphi}$, and so the resulting partitioned retinal activation will be

$$\vec{p}_t = \mathbf{W}_t \mathbf{W}_t^+ \mathbf{R}^* \vec{\varphi} = \mathbf{P} \mathbf{D} \mathbf{Q}^T \mathbf{Q} \mathbf{D}^{-1} \mathbf{P}^T \mathbf{R}^* \vec{\varphi} = \mathbf{P} \mathbf{P}^T \mathbf{R}^* \vec{\varphi}. \quad (13)$$

Using the partitioned form of retinal activation the error can be defined as

$$\varepsilon_t = \|\vec{p}_t^* - \vec{p}_t\|^2. \quad (14)$$

Substituting Eq. (12) and Eq. (13) into Eq. (14) gives

$$\varepsilon_t = \|\vec{\varphi}^T \mathbf{R}^* - \mathbf{P} \mathbf{P}^T \mathbf{R}^* \vec{\varphi}\|^2 = \vec{\varphi}^T [\mathbf{R}^* (\mathbf{I} - \mathbf{P} \mathbf{P}^T) \mathbf{R}^*] \vec{\varphi}. \quad (15)$$

The minimization of this quantity is then performed over the partitioning vector $\vec{\varphi}$, by introducing the objective function, $\Theta_{\text{bin}}[S^2 T^{-2}]$:

$$\Theta_{\text{bin}} = \min_{\vec{\varphi}_i \in \{-1, 1\}} \vec{\varphi}^T [\mathbf{R}^* (\mathbf{I} - \mathbf{P} \mathbf{P}^T) \mathbf{R}^*] \vec{\varphi}. \quad (16)$$

Constant elements in this equation that do not vary with sign changes in elements of $\vec{\varphi}$ can be removed from the equation without affecting the minimization. Since $\vec{\varphi}_i \in \{-1, 1\}$, the diagonal elements of Θ_{bin} will contain $\vec{\varphi}_i^2 = 1$, a constant that will not influence the minimization over $\vec{\varphi}$. This allows us to remove all diagonal elements of Θ_{bin} ; i.e. the identity matrix $\mathbf{I}$ is removed, and $\mathbf{P} \mathbf{P}^T$ is replaced with $\mathbf{P} \mathbf{P}^T - \mathbf{diag}(\mathbf{P} \mathbf{P}^T)$ giving

$$\Theta_{\text{bin}} = \min_{\vec{\varphi}_i \in \{-1, 1\}} \vec{\varphi}^T [\mathbf{R}^* (\mathbf{diag}(\mathbf{P} \mathbf{P}^T) - \mathbf{P} \mathbf{P}^T) \mathbf{R}^*] \vec{\varphi}. \quad (17)$$

Since $\vec{\varphi}_i \in \{-1, 1\}$ this equation fits the form of the binary optimization problem, $\vec{v}^T \mathbf{M} \vec{v}$, with $\mathbf{M} = \mathbf{R}^* (\mathbf{diag}(\mathbf{P} \mathbf{P}^T) - \mathbf{P} \mathbf{P}^T) \mathbf{R}^*$ and $\vec{v} = \vec{\varphi}$. There are 2^{N_t} possible solutions to this equation.

The binary optimization problem is NP-hard even when $\mathbf{M}$ is positive definite, with solutions often falling into local minima and unable to ‘escape’ to a globally optimum solution.

Our approach to solving the local minima problem is to use a Hopfield network in which the weights between the neurons are given by the values contained in the matrix $\mathbf{M}$. The initial values of the vector $\vec{\varphi}$ are chosen to be random and then updated according to the rule, $sgn(\mathbf{M}\vec{\varphi}_k) \rightarrow \vec{\varphi}_{k+1}$. This process is iterated until the vector $\vec{\varphi}$ is static.

In initial investigations using the Hopfield network, it was observed that solutions frequently fell into local minima, with many unnecessary and detrimental partitions between positive and negative current regions resulting in high value of ε_d . To mitigate this problem, simulated annealing was applied. Elements of the vector $\vec{\varphi}$ were randomly ‘flipped’ on each cycle before that vector was used as the initial conditions of a new Hopfield network optimization. The number of annealing cycles was set to 12000, with an initial probability that any given element of $\vec{\varphi}$ would be flipped set at 0.5. This probability was linearly reduced so that it reaches $1/N_t^2$ as the iterations reach 12000. For each of the images this process was repeated 10 times and the solution with lowest value of $\tilde{\varepsilon}_t$ selected as the output.

Results

Linear Inversion Using SVD to Calculate the Pseudoinverse

The conventional and linear inversion algorithms were applied to calculate electrode strengths for a particular desired retinal activation pattern (Fig. 4). Then the forward model Eq. (5) was used to convert those electrode strengths into their retinal activation patterns (Fig. 4).

Figure 4: The results of the conventional model compared to the linear inversion model.

(a) The desired retinal activation pattern. (b) The stimulation patterns showing the electrode strengths calculated by the conventional and linear inversion algorithms. (c) The resulting retinal activation patterns. In producing these results: $n_d = 200, n_t = 40, n_e = 8, \sigma = 0.15, \rho = 0.05, \eta = 0.05, F_{\text{eig}} = 1$. Image © Khoury - used with permission.

The conventional algorithm stimulation pattern consisted of entirely positive values. This is expected given that it is simply a downsampled and intensity-scaled version of the original image. The linear inversion algorithm produced a stimulation pattern with relatively high electrode amplitudes of both polarities. These higher amplitude electrode strengths summate to produce a similar overall level of retinal activation as the conventional strategy, but with sharper contrast. This qualitative difference is the benefit of a global shaping strategy.

In the results shown in Fig. (4) the linear inversion algorithm utilized all eigenvectors associated with the pseudoinverse of $\mathbf{W}_t$ in Eq. (9). The use of different numbers of eigenvectors leads to different solutions (Fig. (5)).

Figure 5: Effect of varying the fraction of eigenvectors, F_{eig} . The value of F_{eig} and the associated number of eigenvectors is shown above each column. (a) The electrode pattern, $\vec{s}$. (b) the predicted retinal pattern, $\vec{r}_d$. The parameters used are otherwise the same as those used Fig. 3: $n_d = 200$, $n_t = 40$, $n_e = 8$, $\sigma = 0.15$, $\rho = 0.05$, $\eta = 0.05$.

It can be seen (Fig. 5(b)) that the use of fewer eigenvectors resulted in a retinal activation pattern of higher error ($\varepsilon_d = 0.031$) and lower spatial frequency, while the use of more eigenvectors created a retinal activation pattern of lower error ($\varepsilon_d = 0.015$) and with features of higher spatial frequency.

Figure 6 explores the effects on the solution of different numbers of eigenvectors. It shows the form of eigenvalues and eigenvectors under four different parameter combinations as a function of F_{eig} (Fig. 6(a) – 6(c)). The normalized value of ε_{d32} under 648 parameter combinations as a function of F_{eig} is also shown (Fig. 6(d)). These parameter combinations are intended to represent different device conditions as explained below.

Figure 6: Selection of eigenvectors associated with the SVD of W_t that appears in Eq. (9).

$n_d = 100, n_t = 40, n_e = 8$ (a) The contents of P for $F_{\text{eig}} = 0.016$ (1/64 eigenvectors), 0.156 (10/64), 0.297 (19/64), 0.438 (28/64), 0.578 (37/64), 0.719 (46/64), 0.859 (55/64), 1.000 (64/64). (b) The magnitude of each eigenvalue contained in the matrix D associated with each parameter combination. (c) The ‘right’ eigenvectors associated with the pattern of electrode activation. (d) Grey lines show different parameter combinations for σ, ρ , and η . Each are given six values over their range (216 conditions for each of the 3 electrode densities). The curves show the value of the ϵ_{d32} at each value of F_{eig} as a ratio with the smallest ϵ_{d32} for all F_{eig} at that parameter combination and electrode density. This is referred to as the ‘normalized ϵ_{d32} ’. The two black curves in each plot show the maximum and median normalized ϵ_{d32} for all parameter combinations. The point highlighted in the 3rd plot shows that if F_{eig} is chosen to be 0.6, then this gives a maximum

of 10% worse outcome for certain parameter combinations than if another value of F_{eig} was used for those combinations.

In principle, the greater the number of eigenvectors used in the calculation of the pseudoinverse the more accurate the result will be. However, in practice there are two limiting factors to the number of eigenvectors used. First, if the matrix $\mathbf{W}_t$ is affected by noise, then some eigenvectors will be corrupted, and their inclusion in the calculation of the pseudo-inverse will lead to overfitting. Second, some eigenvectors require very large changes in the amplitude of electrodes to make a significant change in retinal activation. Both these limitations apply to eigenvectors associated with small eigenvalues. Under SVD, the eigenvectors are ordered by the magnitude of their associated eigenvalues. When using a reduced number of eigenvectors we chose to remove those associated with the smallest eigenvalues first.

The left eigenvectors (Fig. 6(a)) are the retinal activation pattern that results from the electrode amplitude pattern contained in the associated right eigenvectors (Fig. 6(c)). The eigenvalue magnitudes (Fig. 6(b)) can be interpreted as the ratio of the amplitude of the associated pair of left and right eigenvectors. In other words, if an eigenvalue is relatively small then the associated electrode pattern (right eigenvector) will have to be of relatively higher amplitude if it is to have as much impact on amplitude of the generator signal pattern. It is also apparent that generator signal patterns with higher spatial frequency are associated with small eigenvalues (Fig. 6(a)), and as mentioned above, it is the small eigenvalues that are removed first. Given these observations, it is possible to predict that reducing the fraction of eigenvectors used, F_{eig} , will lead to solutions of lower spatial frequency and low electrode amplitudes, and this is exactly what was observed (Fig. 5). In a device with wide activation

spreads, $\sigma = 0.2$, no variation, $\rho = 0$, and no noise, $\eta = 0$ (parameter combination 4 in Figure 6(a)), the lowest magnitude eigenvalues have smaller values than with the other parameter combinations shown. Therefore, using the fact that lower eigenvalues predict higher electrode amplitudes, it can be predicted that linear inversion will create higher amplitude electrode patterns under these conditions. With narrower activation spreads, the smallest eigenvalues increase in magnitude (parameter combination 3 in Fig. 6(a)) which would reduce the resulting electrode pattern's amplitude. With more variation or measurement noise (parameter combinations 3, and 4 in Fig. 6(a)) the higher eigenvalues also increase in magnitude which again predicts lower electrode amplitudes.

In the presence of ERF variation or measurement noise, the eigenvectors associated with the smallest eigenvalues become spatially unstructured and noisy so that they can no longer contribute to improving the resolution of the evoked retinal activity pattern (Figure 6(a), parameter combination 1). This observation implies that over-fitting in the presence of noise could be avoided by excluding eigenvectors associated with eigenvalues of small magnitude in calculating the pseudoinverse. However, as seen in Fig. 5, the use of fewer eigenvectors can reduce the accuracy of the solution by placing an upper limit on the spatial frequencies of eigenvectors used. It was therefore necessary to test the effect of using a different fraction of eigenvalues under the full range of parameters (Fig. 6(d)).

Each of the parameters ρ , and η were given six values covering the ranges, $0.05 < \rho < 0.3$, and $0.05 < \eta < 0.3$, and σ was given values such that $1 < \gamma < 16$, giving 216 parameter combinations, and these were tested with three different electrode numbers $N_e = 2, 8$, and 32 making a total of 648 different combinations of the four dimensions. These 648 combinations can be thought of as representing potential devices.

For each of these 648 devices, the mean value of ε_d across 32 images, ε_{d32} , was calculated for each value of F_{eig} and then divided by the smallest ε_{d32} for all F_{eig} for that combination of parameters. This is referred to as the 'Normalized ε_{d32} '. Each parameter combination can be thought of as creating the conditions for a specific device. The normalization was undertaken because the aim was to find the best choice of F_{eig} for a given device, independent of the absolute value of ε_{d32} for that device for different values of F_{eig} .

It was found that the choice of $F_{\text{eig}} = 0.6$ led to an acceptable result under all devices explored (Fig. 6(d)). In the worst case device, it led to a 27% increase in the ε_{d32} compared to the best choice of F_{eig} for that device (Fig. 6(d)). This value of F_{eig} was used for all other calculations in the present investigation.

Figure 7 compares the linear inversion and conventional algorithms under different activation spread conditions by varying the values of σ and n_e . The conventional and linear inversion algorithms gave different results for different electrode overlaps. For narrow activation spreads (first row in Fig. 7), an increase in electrode density for both the conventional and linear inversion algorithms reduced error, ε_d , in the resulting retinal activation patterns. However, the same does not apply under wide activation spreads (last row, Fig. 7).

Figure 7: Qualitative comparison of the conventional and linear inversion algorithms under different activation spread conditions and electrode densities. $n_d = 200, n_t = 60, \rho = 0.05, \eta = 0.05, F_{\text{eig}} = 0.6$. (a) For $n_e = 2$ three different values of σ were used. For each value of σ the three plots show the outcome using the conventional approach (left) and linear inversion approach (middle), and a quantitative comparison of the mean retinal activation across the image (right) (b) $n_e = 8$. (c) $n_e = 32$.

The conventional algorithm's error value does not change with the number of electrodes ($\epsilon_d = 0.043$). In contrast, the linear inversion algorithm can utilize additional electrodes to shape the total current ($\epsilon_d = 0.043 \rightarrow \epsilon_d = 0.030$). This apparent difference between the conventional and linear inversion algorithms was also explored quantitatively; the parameters σ , ρ , and η were given three values covering the previous ranges, giving 27 parameter combinations. The value of ϵ_{d32} for the conventional and linear inversion algorithms was calculated for electrode numbers (n_e) between 2 and 32 (Fig. 8(a) and 8(b)).

Figure 8: Quantitative comparison of the conventional and linear inversion algorithms under 27 different activation spread conditions and electrode densities. $n_d = 100$, $n_t = 40$, $n_e = 8$, $F_{eig} = 0.6$. (a) The ϵ_{d32} for the conventional algorithm and (b) the linear inversion algorithm. (c) The ϵ_{d32} of the linear inversion algorithm divided by the ϵ_{d32} of the conventional algorithm for each parameter combination. In each plot the dashed curves are associated with ‘narrow spread’ conditions ($\gamma < 1$). The solid curves show ‘wide spread’ conditions ($\gamma > 1$). Note that changes in either ρ or η do not influence the outcome of the conventional algorithm; this leads to degeneracy and the 27 simulations overlap, leaving the three lines associated with the three values of σ , two of these lie an indistinguishable distance from one another (a). This effect is not seen in the linear inversion algorithm, which is sensitive to variation, noise, and activation spread (b).

The accuracy of the conventional algorithm appears to improve with increasing number of electrodes until the stimulation overlap reaches a certain level and then the improvement ceases (Fig. 8(a)). In contrast, the linear inversion algorithm shows reducing improvement in ϵ_{d32} with increasing number of electrodes (Fig. 8(b)).

The ratio of the ϵ_{d32} for the conventional and linear inversion algorithms (Fig. 8(c)) demonstrates that, for $\gamma > 1$, the linear inversion algorithm provides lower ϵ_{d32} in most

circumstances. The exception is under the highest activation spread, $\sigma = 1$, and highest measurement noise, $\eta = 0.3$, parameter combination.

Quadratic Program with Linear Constraints on the Electrode Amplitudes

Figure 9 shows the results of linear inversion under a range of noise and current-spread conditions. In the case with no noise, $\eta = 0$, the linear inversion algorithm produces a solution in which the electrodes take high amplitudes. For illustration purposes, we select a value of 2 as the maximum value, $s_{\max}$, twice the value of 1, which produces the maximum retinal activation of 1. So each element of $\vec{b}$ is set to be 2. With this selection it can be seen that under certain parameter combinations the electrodes exceed this maximum (box in Fig. 9). This is because wide activation spread and low noise conditions lead to very small eigenvalues being used in the pseudoinverse (Fig. 5(a), parameter combination 4). Use of a quadratic program with linear constraints on the electrode amplitudes can eliminate the high electrode amplitudes that arise using linear inversion seen in figure 9.

Figure 9: The effects of wide activation spread in low estimation noise and low variation ($\rho = 0$) conditions. $n_d = 100, n_t = 60, n_e = 16, F_{eig} = 0.6$. The box indicates the case where the linear inversion algorithm produces electrode strengths higher than the maximum allowed amplitude of 2 (The dark grey and light grey values that exceed the +2 and -2 values, respectively).

Figure 10 compares the linear inversion and quadratic program solutions at both high noise, $\eta = 1$, and low noise, $\eta = 10^{-5}$, conditions.

Figure 10: Performance of the Linear Inversion and Quadratic Program algorithms under low estimation noise, and zero variability conditions. (a) Qualitative results showing the electrode activation pattern and the retinal activation pattern. The dark grey and light grey values that exceed the +2 and -2 values, respectively. The dark blue and dark red

colors represent electrodes equal to +2 and -2, respectively. (b) Quantitative results ε_{d32} and $s_{\max}$ of the Linear Inversion and Quadratic Program algorithms as a function of the estimation noise. The horizontal line shows the maximum allowable electrode amplitude, $s_{\max} = 2$. $n_d = 100$, $n_t = 40$, $n_e = 16$, $\rho = 0$, $\sigma = 0.2$, $F_{\text{eig}} = 0.6$.

In the case of the linear inversion algorithm, low estimation noise and wide activation spread led to the selection of high-amplitude electrodes (Fig. 10(a)). However under the same conditions, the quadratic program avoided high electrode amplitudes while the value of E increased only marginally, ($\varepsilon_d = 0.012 \rightarrow \varepsilon_d = 0.014$), (Fig. 10(b)). These results were examined quantitatively across the same range using the value ε_{d32} (Fig. 10(b)). It can be seen (dashed red line in Fig. 10(b)) that the quadratic program avoids the increase in electrode amplitude with reduced noise, $\eta < 0.05$. In both the qualitative (Fig. 10(a)) and quantitative (Fig. 10(c)) analyses it can be seen that increase in ε_{d32} that results from limiting the electrode amplitude is small. It can also be seen that outside these low-noise circumstances the quadratic program gives an identical outcome to the linear inversion algorithm by using the same restricted eigenvalue version of $\mathbf{W}_t$, with $F_{\text{eig}} = 0.6$. The quadratic program avoids overfitting in the case of high noise wide activation spread conditions and also places an explicit limit on the amplitudes of the electrodes, avoiding overstimulation in low noise, wide activation spread conditions.

The limitation of the quadratic program approach is computational cost. In MATLAB, the process of calculating the electrode pattern for a single image using the quadratic program inverse model was ~ 1 s, which compares to the linear inversion time of ~ 1 ms, and the conventional algorithm time of ~ 10 ms.

Binary Quadratic Program for Partitioning the Target Retinal Activation Pattern

In calculating the linear inversion and the quadratic program solutions, the modulus operation in Eq. (5) was removed to allow for simple linear inversion Eq. (9). This decision limited those two algorithms to use only positive generator signal values to induce retinal activation (positive values of G_i in Fig. 1A). Partitioning the target retinal activation pattern into positive and negative regions restores the modulus present in Eq. (5), allowing for retinal activation to be induced by either a positive or negative generator signal (Fig. 1(a)).

Figure 11(a)-(d) shows examples of target retinal activity patterns, $\vec{r}_t^*$, potential beneficial partitions of these patterns, $\vec{p}_t^*$, and the resulting retinal activation pattern using linear inversion, $\vec{r}_t$. Borders between positive and negative regions of the generator signal pattern appear in the retinal activation pattern as closed loops of zero retinal activity. This is because the generator signal values must pass through zero between regions of positive and negative values, for example see Figure 11(c).

Figure 11: Schematic illustration of the effects of partitioning of the target retinal activation pattern. The four columns show a target retinal activity patterns, $\vec{r}_t^*$, then a potential partitioned target pattern, $\vec{p}_t^*$, and the resulting retinal activation pattern for each case, $\vec{r}_t$. The resulting activity is after applying the modulus in Eq. (5), so the result is $\vec{r}_t$, not $\vec{p}_t$. In the partition, $\vec{p}_t^*$, red and blue indicate the polarity of the partitioned target generator signal. (a-b) Two beneficial partitions. (c-d) Potential beneficial partitioned target retinal activations that lead to artifacts or ambiguity in the solution.

The fractures in the solutions derived from a partitioned target can lead to lower ε_{d32} in cases where the original target possesses such features (Fig. 11(a) and 11(b)). However, there are two ways in which partitioning could create ambiguities and artefactual features in the solution. In the first case, a partition assists to more accurately create an unclosed line of zero retinal activity in the target activation pattern, but the partition border must be a closed loop, and this leads to an artifact by creating a line of zero retinal activity where none exists (Fig. 11(c)). In the second case, the configuration of a closed loop in the target image may allow

for a beneficial partition, but fails to do the same in an equivalent line (Fig. 11(d)). In target retinal activation patterns based on natural images, these two problems may not be encountered frequently. For example, in natural images, the algorithm may take advantage of ambiguities in the target patterns (Fig. 11(b)).

The binary quadratic program, solved using a Hopfield network with simulated annealing, was applied to the same 32 images throughout this investigation. Figure 12(a) – 12(e) compares a sample of the lowest and highest ε_d outcomes of using partitions to the results without a partition. Figure 12(e) shows the histogram of the difference in ε_d for each of the 32 images used.

Figure 12: The results of using a binary quadratic program to partition the target retinal activation pattern before applying linear inversion. (a) Desired retinal activation pattern. (b) Target retinal activation pattern without partitioning. (c) Partitioned target retinal activation pattern. Red and blue indicated the polarities of the target generator signal (same as figure 10) (d) The model retinal activation pattern based on the unpartitioned

target. (e) The model retinal activation pattern based on the partitioned target. (f) The % change in E with the application of partitioned target retinal activation pattern. $n_d = 100, n_t = 20, n_e = 6, \rho = 0.05, \eta = 0.05, \sigma = 0.2, F_{\text{eig}} = 0.6$.

In 6 out of the 32 images the partition led to a detrimental outcome as measured by ε_d . In 15 out of the 32 images the best outcome resulted from an unpartitioned target pattern. However improvements in ε_d were found 11 out of 32 images (Fig. 12). Examination of the most detrimental partitions, as measured by the value of ε_d (Figure 12(a)-(e) bottom two rows), reveals that in qualitative terms, the result of partitioning is not necessarily worse than without.

The process of simulated annealing cycles of the Hopfield network for each image requires a large computational investment of ~ 100 s in MATLAB for each image.

Discussion

Using global shaping algorithms under wide activation spread conditions ($\gamma > 1$) it was found that the performance parameter, ε_{d32} , was substantially reduced over a range of device parameters when compared to the conventional algorithm. In contrast to the conventional algorithm, global shaping approaches improve the accuracy of the results even at high electrode densities. This result is significant because it demonstrates that, in principle, there is no intrinsic limit on the electrode spread relative to electrode spacing. Under the conventional algorithm, the resolution of the retinal activation pattern approaches the limit set by the current spread associated with each electrode (Fig. 8A) and, under global current-shaping, the limit is set by the spacing between electrodes (Fig. 8B). This means that the approach proposed in this paper allows calculation of the benefits of increasing electrode density. This information can then be balanced against the associated costs of large numbers of electrodes.

The linear inversion algorithm has a low computation requirement; calculation of the pseudoinverse is made for a given device, and subsequent image-by-image calculation is a simple matrix multiplication. The algorithm was made robust to noise by selection of the eigenvectors used in the pseudoinverse. However, linear inversion was susceptible to a low-noise condition in which it provided electrode settings of extremely high amplitude. The quadratic program could be used under these conditions to explicitly limit the electrode amplitudes, however the image-by-image calculation requires significantly more computation than linear inversion. Consequently, in the case with low-estimation noise and wide activation spread, the quadratic program should be used. However, if the estimation noise is high, linear

inversion combined with eigenvalue selection may be sufficient, and requires much less computational resources.

The binary quadratic program creates an initial partition of the target retinal activation pattern by an NP-hard optimization algorithm for each image. This is a very significant computational investment. The benefit of this choice is access to a new set of solutions that cannot be found by the linear inversion or quadratic program. These solutions are those that include high-contrast lines or fractures. It is likely that the value of this approach varies with the device parameters set, and this parameter exploration is a topic for future investigation. Our approach to solving this problem involved the use of a Hopfield network combined with a systematic cycle of simulated annealing. However, this problem may also be amenable to solution using an artificial neural network trained using a target retinal activation vector as input and the partitioning vector as output. The benefit of using an artificial neural network is that the final trained network would be able to partition images with higher computational speed.

The selection of eigenvectors influences the capacity of the pseudoinverse to fit electrodes to spatial frequencies in the target retinal activity patterns (Fig. 5). This influences the accuracy of some target patterns more than others. For example, with fewer eigenvectors, target patterns with high spatial frequencies will not be as easily achieved as those with low spatial frequencies. Future work on eigenvalue selection could examine how the statistics of the spatial frequencies in natural images compare to the statistics of the spatial frequencies in the eigenvectors in the SVD associated with different devices.

The number of eigenvectors used was selected to be a constant $F_{\text{eig}} = 0.6$ independent of the device parameters (Fig. 5D). This was done for simplicity in the present investigation. However, for further reductions in ε_{d32} the eigenvalue selection could be made device specific by developing a function $F_{\text{eig}}(n_e, \sigma, \eta, \varphi)$. Exploration of device-dependent eigenvalue selection is a useful topic for future investigation.

In the present investigation, natural image intensities were used as the basis for target retinal activation. This was sufficient for a proof of principle in the current investigation. It is observed that many subpopulations of retinal cells are not sensitive to the overall image intensity, but to edges in perceived images [36]. A potential model of this effect would use natural images passed through a whitening filter as the target retinal activation pattern. These kind of target images typically contain large regions of low retinal activity punctuated by filaments of high retinal activity. These patterns are associated with high spatial frequencies, and so may benefit from the inclusion of high spatial frequency eigenvectors in calculating the pseudoinverse. This process merits careful future exploration due to the fact that the present investigation suggests that eigenvectors containing high spatial frequencies are those that are corrupted by noise and also lead to high amplitude electrode solutions. In addition, the partitioning approach taken using the binary quadratic program seems unsuited to these kind of filamented target patterns. This is because that partitioning approach allows solutions that have thin sections of low retinal activity between wide areas of high-activity. However, these whitened images that emphasize edges may still be useful in conveying more useful information than unwhitened images.

This paper used a rate code model of retinal activation rather than a temporal code. However, electrical stimulation of the retina does not create a straightforward temporal step change in

retinal activation [37]. This is partly due to the fact that there is both direct electrical activation of retinal ganglion cells, as well as network effects that create temporal dynamics through subsequently generated activation of the retinal ganglion cells at a lower amplitude. These temporal effects may lead to the transfer matrix being a function of time, $\mathbf{W}_d(t)$. There are also differential effects on separate populations of neurons within the retina [38]. This could be modeled as a set of population-dependent and time-dependent transfer matrices and a set of population-dependent and time-dependent target patterns $\{(\mathbf{W}_{d,1}, \mathbf{W}_{d,2} \dots), (\vec{r}_{d,1}^*, \vec{r}_{d,2}^*, \dots)\}$. In the present investigation these time-dependent and population-dependent factors were put aside to investigate the global-current shaping aspects of the problem. This decision was justified because (a) a shaping model that does include these elements will depend upon the results of the current investigation; (b) it is expected that even ignoring these other factors, the results of the present investigation could be applied with some success.

The influence of each electrode on the retinal activation was modeled as circular 2D-Gaussians with small perturbations. In a more realistic model these functions could be more complex shapes and each electrode could be given a unique spread based on retinal activation measurements made by the stimulating array itself. This would be particularly useful to model the elongated generator signal patterns that may arise due to the fact that the retinal tissue is anisotropic [39]. The electrode locations too could be varied to reflect some designs in which hexagonal patterns are used. Though these variations were not part of the current investigation, Eq. (4) could be easily modified to include these effects in construction of the matrix $\mathbf{W}_d$.

An additional consideration was the fact that the use of Gaussians implies the activation spread from each electrode extends indefinitely. This is part of what leads to very small eigenvalues, and consequently the production of extremely high amplitude electrode settings under low-noise conditions when using linear inversion. An alternative model could avoid this if it is justifiable to accurately model the true influence of each electrode using spatially limited functions that terminate to 0 amplitude at some distance from the electrode.

In the forward model, it was assumed that the sigmoidal nonlinearity $f(x_i)$ (Fig. 1A) could be modeled as a rectified linear function. This over-estimates the effect of each electrode under high generator signal values where it results in unrealistically high retinal activation values. This apparently large discrepancy does not result in large errors because the target retinal activation pattern has a maximum value of 1, limiting the function $f(x_i)$ to the regime in which the difference between the sigmoidal curve and the rectified linear curve are small. Despite this, the rectified linear curve assumption does prevent the use of high generator signal values to induce values of retinal activation close to 1. This may mean that the solutions provided in this paper spend effort unnecessarily attenuating the generator signal towards 1, when much higher values would have been acceptable under the ‘true’ sigmoidal nonlinearity. In addition, in some cases, nonmonotonic activation functions have been observed [40]. These do not present a problem for the methods presented in this paper if, within a low generator function amplitude domain, the nonmonotonic response can be approximated by a rectified linear function.

In some experimental measurements the final nonlinearity $f(x_i)$ shown in Fig. 1A is not symmetrical; negative currents induce a different retinal activation than positive total currents of the same magnitude [31]. This consideration does create a problem in the case of

the binary quadratic program, since this partition approach does assume symmetry. It may be that the asymmetry is small in which case the partition approach can continue to be used; otherwise, an alternative form of ‘quantitative partition’ or filter could be used. For example a quadratic program could be used to calculate both the sign and magnitude of the filter to apply to the target before applying linear inversion.

Global shaping intrinsically requires simultaneous stimulation by all electrodes to achieve the shaping effect. In addition, under wide activation spread conditions the amplitudes of the individual electrode activations is typically higher than those associated with the conventional algorithm. These two qualities require high power throughput to the device; if a global shaping strategy is used then this cost will have to be balanced against the perceptual benefits provided by the algorithm.

This investigation used the RMS error ε_d as the basis for assessment of the success of the outcomes. The methods themselves were developed using another RMS error function ε_t . In some circumstances this approach can produce distracting artifacts (Fig. 10(c) and 10(d)). An alternative approach would use a measure of visual saliency in an attempt to capture the impact of particular features of the target retinal activation pattern that are present in the modeled outcome and vice versa.

Acknowledgements

This project was supported by the Australian Research Council’s Discovery Projects funding scheme (Project DP140104533). HM acknowledges funding from the Australian Research Council Centre of Excellence for Integrative Brain Function (Project CE140100007). HM, TK, DBG, and ANB acknowledge partial support of NHMRC Project grant 1106390.

References

- [1] Weiland J D, Liu W and Humayun M S 2005 Retinal Prosthesis *Annu. Rev. Biomed. Eng.* **7** 361–401
- [2] Ayton L N, Blamey P J, Guymer R H, Luu C D, Nayagam D A X, Sinclair N C, Shivdasani M N, Yeoh J, McCombe M F, Briggs R J, Opie N L, Villalobos J, Dimitrov P N, Varsamidis M, Petoe M A, McCarthy C D, Walker J G, Barnes N, Burkitt A N, Williams C E, Shepherd R K, Allen P J and Consortium for the B V A R 2014 First-in-Human Trial of a Novel Suprachoroidal Retinal Prosthesis *PLOS ONE* **9** e115239
- [3] Zrenner E, Bartz-Schmidt K U, Benav H, Besch D, Bruckmann A, Gabel V-P, Gekeler F, Greppmaier U, Harscher A, Kibbel S, Koch J, Kusnyerik A, Peters T, Stingl K, Sachs H, Stett A, Szurman P, Wilhelm B and Wilke R 2010 Subretinal electronic chips allow blind patients to read letters and combine them to words *Proc. R. Soc. Lond. B Biol. Sci.* rspb20101747
- [4] Rizzo J F, Wyatt J, Loewenstein J, Kelly S and Shire D 2003 Methods and Perceptual Thresholds for Short-Term Electrical Stimulation of Human Retina with Microelectrode Arrays *Invest. Ophthalmol. Vis. Sci.* **44** 5355–61
- [5] Fujikado T, Kamei M, Sakaguchi H, Kanda H, Morimoto T, Ikuno Y, Nishida K, Kishima H, Maruo T, Konoma K, Ozawa M and Nishida K 2011 Testing of Semichronically Implanted Retinal Prosthesis by Suprachoroidal-Transretinal Stimulation in Patients with Retinitis Pigmentosa *Invest. Ophthalmol. Vis. Sci.* **52** 4726–33
- [6] Lorach H, Goetz G, Smith R, Lei X, Mandel Y, Kamins T, Mathieson K, Huie P, Harris J, Sher A and Palanker D 2015 Photovoltaic restoration of sight with high visual acuity *Nat. Med.* **21** 476–82
- [7] Jones B W, Watt C B and Marc R E 2005 Retinal remodelling *Clin. Exp. Optom.* **88** 282–91
- [8] Rattay F 1999 The basic mechanism for the electrical stimulation of the nervous system *Neuroscience* **89** 335–46
- [9] Esler T B, Maturana M I, Kerr R R, Grayden D B, Burkitt A N and Meffin H 2018 Biophysical basis of the linear electrical receptive fields of retinal ganglion cells *J. Neural Eng.* **15** 055001
- [10] Wilke R G H, Moghadam G K, Lovell N H, Suanning G J and Dokos S 2011 Electric crosstalk impairs spatial resolution of multi-electrode arrays in retinal implants *J. Neural Eng.* **8** 046016
- [11] Tahayori B, Meffin H, Sergeev E N, Mareels I M Y, Burkitt A N and Grayden D B 2014 Modelling extracellular electrical stimulation: part 4. Effect of the cellular composition of neural tissue on its spatio-temporal filtering properties *J. Neural Eng.* **11** 065005

- [12] Chen S C, Suaning G J, Morley J W and Lovell N H 2009 Simulating prosthetic vision: I. Visual models of phosphenes *Vision Res.* **49** 1493–506
- [13] Chen S C, Suaning G J, Morley J W and Lovell N H 2009 Simulating prosthetic vision: II. Measuring functional capacity *Vision Res.* **49** 2329–43
- [14] Chai X, Yu W, Wang J, Zhao Y, Cai C and Ren Q 2007 Recognition of Pixelized Chinese Characters Using Simulated Prosthetic Vision *Artif. Organs* **31** 175–82
- [15] Buffoni L-X, Coulombe J and Sawan M 2005 Image Processing Strategies Dedicated to Visual Cortical Stimulators: A Survey *Artif. Organs* **29** 658–64
- [16] Boyle J, Maeder A and Boles W 2003 Scene specific imaging for bionic vision implants *3rd International Symposium on Image and Signal Processing and Analysis, 2003. ISPA 2003. Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis, 2003.* **1** 423–427
- [17] Rheede J J van, Kennard C and Hicks S L 2010 Simulating prosthetic vision: Optimizing the information content of a limited visual display *J. Vis.* **10** 32–32
- [18] Wang J, Li H, Fu W, Chen Y, Li L, Lyu Q, Han T and Chai X 2016 Image Processing Strategies Based on a Visual Saliency Model for Object Recognition Under Simulated Prosthetic Vision *Artif. Organs* **40** 94–100
- [19] Rathbun D L, Ghorbani N, Shabani H, Zrenner E and Hosseinzadeh Z 2018 Spike-triggered average electrical stimuli as input filters for bionic vision – a perspective. *J. Neural Eng.*
- [20] Dumm G, Fallon J B, Williams C E and Shivdasani M N 2014 Virtual electrodes by current steering in retinal prostheses *Invest. Ophthalmol. Vis. Sci.* **55** 8077–85
- [21] Spencer T C, Fallon J B, Thien P C and Shivdasani M N 2016 Spatial restriction of neural activation using focused multipolar stimulation with a retinal prosthesis *Invest. Ophthalmol. Vis. Sci.* **57** 3181–3191
- [22] Spencer T C, Fallon J B and Shivdasani M N 2018 Creating virtual electrodes with 2D current steering *J. Neural Eng.* **15** 035002
- [23] Dmochowski J P, Datta A, Bikson M, Su Y and Parra L C 2011 Optimized multi-electrode stimulation increases focality and intensity at target *J. Neural Eng.* **8** 046011
- [24] Townshend B and White R L 1987 Reduction of Electrical Interaction in Auditory Prostheses *IEEE Trans. Biomed. Eng.* **BME-34** 891–7
- [25] Bonham B H and Litvak L M 2008 Current focusing and steering: modeling, physiology, and psychophysics *Hear. Res.* **242** 141–53
- [26] Jepson L H, Hottowy P, Mathieson K, Gunning D E, Dąbrowski W, Litke A M and Chichilnisky E J 2014 Spatially patterned electrical stimulation to enhance resolution of retinal prostheses *J. Neurosci.* **34** 4871–81

- [27] Jepson L H, Hottowy P, Weiner G A, Dabrowski W, Litke A M and Chichilnisky E J 2014 High-Fidelity Reproduction of Spatiotemporal Visual Signals for Retinal Prosthesis *Neuron* **83** 87–92
- [28] Halupka K 2017 Prediction and shaping of visual cortex activity for retinal prostheses, PhD Thesis, University of Melbourne.
- [29] Halupka K J, Shivdasani M N, Cloherty S L, Grayden D B, Wong Y T, Burkitt A N and Meffin H 2017 Prediction of cortical responses to simultaneous electrical stimulation of the retina *J. Neural Eng.* **14** 016006
- [30] Halupka K J, Abbott C J, Wong Y T, Cloherty S L, Grayden D B, Burkitt A N, Sergeev E N, Luu C D, Brandli A, Allen P J, Meffin H and Shivdasani M N 2017 Neural Responses to Multielectrode Stimulation of Healthy and Degenerate Retina *Invest. Ophthalmol. Vis. Sci.* **58** 3770–84
- [31] Maturana M I, Apollo N V, Hadjinicolaou A E, Garrett D J, Cloherty S L, Kameneva T, Grayden D B, Ibbotson M R and Meffin H 2016 A Simple and Accurate Model to Predict Responses to Multi-electrode Stimulation in the Retina *PLOS Comput. Biol.* **12** e1004849
- [32] Maturana M I, Apollo N V, Garrett D J, Kameneva T, Cloherty S L, Grayden D B, Burkitt A N, Ibbotson M R and Meffin H 2018 Electrical receptive fields of retinal ganglion cells: Influence of presynaptic neurons *PLOS Comput. Biol.* **14** e1005997
- [33] Sekhar S, Jalligampala A, Zrenner E and Rathbun D L 2016 Tickling the retina: integration of subthreshold electrical pulses can activate retinal neurons *J. Neural Eng.* **13** 046004
- [34] Vanderbei R J and Carpenter T J 1993 Symmetric indefinite systems for interior point methods *Math. Program.* **58** 1–32
- [35] Altman A and Gondzio J 1998 Regularized Symmetric Indefinite Systems in Interior Point Methods for Linear and Quadratic Optimization *Optim. Methods Softw.* **11–12** 275–302
- [36] Enroth-Cugell C and Robson J G 1966 The contrast sensitivity of retinal ganglion cells of the cat *J. Physiol.* **187** 517–52
- [37] Fried S I, Hsueh H A and Werblin F S 2006 A Method for Generating Precise Temporal Patterns of Retinal Spiking Using Prosthetic Stimulation *J. Neurophysiol.* **95** 970–8
- [38] Masland R H 2001 Neuronal diversity in the retina *Curr. Opin. Neurobiol.* **11** 431–6
- [39] Esler T B, Kerr R R, Tahayori B, Grayden D B, Meffin H and Burkitt A N 2018 Minimizing activation of overlying axons with epiretinal stimulation: The role of fiber orientation and electrode configuration *PLOS ONE* **13** e0193598

- [40] Meng K, Fellner A, Rattay F, Ghezzi D, Meffin H, Ibbotson M R and Tatiana Kameneva
2018 Upper stimulation threshold for retinal ganglion cell activation *J. Neural Eng.* **15**
046012