



(51) International Patent Classification:

A61N 1/36 (2006.01) A61F 9/08 (2006.01)  
A61N 1/05 (2006.01) G06N 3/00 (2006.01)

(21) International Application Number:

PCT/AU2020/050287

(22) International Filing Date:

26 March 2020 (26.03.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

2019901065 29 March 2019 (29.03.2019) AU

(71) Applicant: **THE UNIVERSITY OF MELBOURNE**  
[AU/AU]; Grattan Street, The University of Melbourne,  
Victoria 3010 (AU).

(72) Inventor: **ESLER, Timothy Bede**; c/- The University of  
Melbourne, Intellectual Property & Licences, Research, In-  
novation & Commercialisation, The University of Mel-  
bourne, Victoria 3010 (AU).

(74) Agent: **FB RICE PTY LTD**; Level 23, 44 Market St, Syd-  
ney, New South Wales 2000 (AU).

(81) Designated States (*unless otherwise indicated, for every  
kind of national protection available*): AE, AG, AL, AM,  
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ,  
CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,  
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,

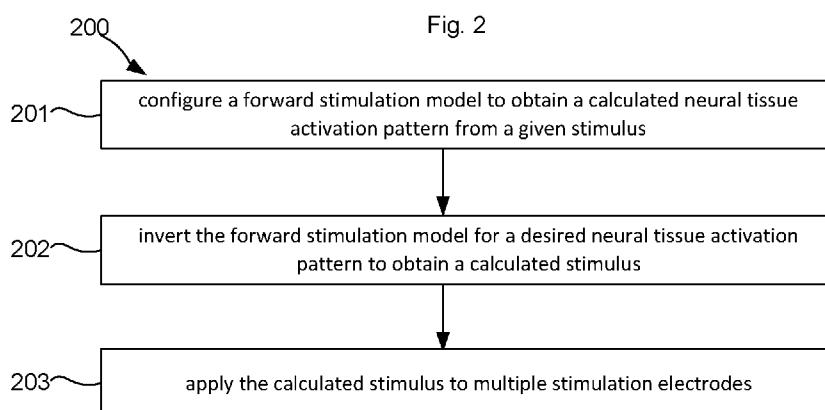
HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP,  
KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME,  
MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,  
OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,  
SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR,  
TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every  
kind of regional protection available*): ARIPO (BW, GH,  
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ,  
UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,  
TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,  
EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,  
MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,  
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,  
KM, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: MULTI-ELECTRODE NEURAL STIMULATION



(57) Abstract: This disclosure relates to a neural stimulation device. Multiple electrodes provide stimulation to neural tissue. A processor configures a forward stimulation model to obtain a calculated neural tissue activation pattern from a given stimulus. The processor then inverts the forward stimulation model for a desired neural tissue activation pattern to obtain a calculated stimulus and applies the calculated stimulus to multiple stimulation electrodes. The disclosed neural stimulation method uses a neural network that models the evoked neural response based on known stimuli. The neural network has the stimuli as weight parameters, which means that the weights can be determined by numerical training for any arbitrary desired output image.



## **“Multi-electrode neural stimulation”**

### **Cross-Reference to Related Applications**

[0001] The present application claims priority from Australian Provisional Patent Application No 2019901065 filed on 29 March 2019, the contents of which are incorporated herein by reference in their entirety.

### **Technical Field**

[0002] This disclosure relates to methods, systems and implantable devices for multi-electrode neural stimulation, including, but not limited to, retinal stimulation.

### **Background**

[0003] Neural stimulation can be used in many applications including therapeutic applications, such as pain relief, and perception applications, such as visual support via retinal stimulation. Typically, electrodes are in contact with or near neural tissue, such as the retina, and deliver a stimulation current to evoke a desired neural activation pattern.

[0004] However, most neural tissue has a complex structure and, as a result, the stimulation current delivered by one electrode at one electrode location does not necessarily result a neural activation only at that electrode location. Instead, the axons of other neurons, which have distantly located cell bodies, also cross the electrode locations and get activated. In the case of retinal stimulation, this results in blurry perception for the user even if the intensity of stimulation currents at the electrodes represent the desired visual scene perfectly.

[0005] Fig. 1 illustrates an electrode array 101 and a nervous tissue 102 including a nerve fiber layer 103 and a ganglion cell layer 104. One example electrode 105 is activated which results in activation of three nerve fibers 105 and consequently activation of three ganglion cells 106, 107, 108. Since these three ganglion cells 106,

107, 108 are not under the activated electrode 105, they are considered unwanted in this example and lead to blurring of the perceived image.

[0006] Further, stimulation devices are often implanted, which means they have a limited amount of resources available. In particular, energy from the battery is limited and processing power is low due to the use a low-power, implantable processor. This means computationally expensive calculations are impractical, while at the same time, a satisfactory framerate, such as at least 10fps, is desired.

[0007] Any discussion of documents, acts, materials, devices, articles or the like which has been included in the present specification is not to be taken as an admission that any or all of these matters form part of the prior art base or were common general knowledge in the field relevant to the present disclosure as it existed before the priority date of each claim of this application.

[0008] Throughout this specification the word “comprise”, or variations such as “comprises” or “comprising”, will be understood to imply the inclusion of a stated element, integer or step, or group of elements, integers or steps, but not the exclusion of any other element, integer or step, or group of elements, integers or steps.

## **Summary**

[0009] This disclosure provides a method for neural stimulation that uses a neural network that models the evoked neural response based on known stimuli. The neural network has the stimuli as weight parameters, which means that the weights can be determined by numerical training for any arbitrary desired output image.

[0010] A method for neural stimulation comprises:  
configuring a forward stimulation model to obtain a calculated neural tissue activation pattern from a given stimulus;  
inverting the forward stimulation model for a desired neural tissue activation pattern to obtain a calculated stimulus; and

applying the calculated stimulus to multiple stimulation electrodes.

[0011] It is an advantage that the forward stimulation model is inverted to obtain the calculated stimulus, as opposed to using a fixed inverse model, as this allows the use of accurate forward stimulation models.

[0012] Inverting the forward stimulation model may comprise obtaining multiple samples of the desired nervous stimulation pattern and optimising the forward stimulation model for each sample.

[0013] It is an advantage that the sampling of patterns allows for efficient optimisation of the model. This approach allows for an approximately optimal inversion to be obtained rapidly, which is then improved with further sampling if time permits before the next target pattern arrives or the target pattern does not change greatly.

[0014] Optimising the forward stimulation model may be performed iteratively for the multiple samples.

[0015] The forward stimulation model may comprise an input, an output and model parameters and the calculated stimulus is represented by the model parameters of the forward stimulation model such that optimising the parameters of the forward stimulation model results in optimising the calculated stimulus.

[0016] It is an advantage that the model parameters represent the stimulus because the optimisation then provides the calculated stimulus as the optimised model parameters. This allows the use of efficient numerical optimisation techniques that have been designed for, for example, neural networks.

[0017] The neural stimulation may comprise neural stimulation of retinal nerves of the visual system. The desired nervous stimulation pattern is a desired image represented by the stimulation pattern on a retina. Inverting the forward stimulation

model may comprise obtaining multiple sample images of the desired nervous stimulation pattern and optimising the forward stimulation model for each sample image.

[0018] The forward stimulation model may be an artificial neural network. It is an advantage that the forward stimulation model is represented as an artificial neural network as this allows for model inversion to be performed using efficient iterative optimization algorithms designed for online training of artificial neural networks, such as stochastic gradient descent (SGD) and Adam.

[0019] The artificial neural network may comprise a linear input, a network layer representing a differentiable nonlinearity and an output representing the neural tissue activation pattern. The linear input may comprise a linear representation of the neural tissue and a differentiable nonlinearity may comprise a representation of nonlinear neural tissue activation.

[0020] The artificial neural network may comprise network parameters that represent a stimulus and inverting the forward stimulation model comprises optimising the network parameters to thereby obtain the calculated stimulus.

[0021] Optimising the network parameters may comprise applying a constraint to the network parameters to ensure the calculated stimulus remains within clinical safety bounds or device power limits or both.

[0022] The artificial neural network may comprise a network layer, the network layer may comprise a non-overlapping convolutional layer, and optimising the network parameters may comprise optimising only parameters of the non-overlapping convolutional layer.

[0023] The method may be continuously repeated for a sequence of images to be perceived, each image representing a desired neural activation pattern, and the forward stimulation model being inverted for each desired neural activation pattern.

[0024] The method may comprise inverting the forward stimulation model for a first region of the neural tissue and minimising activation in a second region of the neural tissue.

[0025] Inverting the forward stimulation model may comprise minimizing logarithmic loss.

[0026] The forward stimulation model may comprise a three-dimensional representation of neural tissue.

[0027] Optional features described of any aspect of method, computer readable medium or computer system, where appropriate, similarly apply to the other aspects also described here.

### **Brief Description of Drawings**

[0028] An example will be described with reference to the following figures.

[0029] Fig. 1 illustrates unwanted stimulation of retinal ganglion cell axons of passage. Retinal ganglion cell somas and axon initial segments represent the target regions for epiretinal stimulation. Activation of passing axons in the nerve fiber layer results in long, arc-shaped visual percepts and degradation of the quality of artificial vision. Retinal ganglion cell axon bundles in the nerve fiber layer that pass close to stimulating electrodes may be stimulated preferentially to target locations in the ganglion cell layer. Note that the orientation of initial axonal segments is more varied in reality than shown in this schematic.

[0030] Fig. 2 illustrates a method for neural stimulation.

[0031] Fig. 3 illustrates an example forward stimulation model.

[0032] Fig. 4 illustrates a linear-nonlinear neural network architecture. The input dataset has the dimensions  $R \times M \times N$ , where  $R$  is the number of two-dimensional

representations of the retina,  $M$  is the number of locations sampled in each two-dimensional representation, and  $N$  is the number of electrodes. A single square in the input,  $V_r$ , is the estimated membrane depolarization induced at a single location in the retina in response to a standard stimuli (e.g.  $1 \mu A$ ) delivered by a single electrode. The vector of these values for each electrode at a single location is the electrical receptive field (ERF) for that location. Each layer of the network adds a component to the final mathematical description of the output as shown in the right column. The layers of the model are (1) the input layer, (2) a *non-overlapping* convolutional layer, (3) a replication layer which creates a positive and negative copy of its input, (4) an offset layer which applies the positive and negative spiking thresholds for each representation, (5) a double-sided sigmoidal nonlinearity, and (6) the averaged output. The convolutional layer does not strictly perform a convolution over the input (as the filter is applied in a non-overlapping manner), however, the layer weights,  $\vec{S}_r$ , are fixed across all  $R$  representations in the manner of a typical convolutional layer. Note that for monophasic stimulation, layer 3 is not needed and layer 5 is simplified to a single-sided sigmoid.

[0033] Fig. 5 is a schematic of transformation of two-dimensional electrical receptive field (ERF) data. For each orientation in the GCL and for the NFL (i.e. for all representations), simulations were run of stimulation by each electrode, one by one, with a stimulus amplitude of  $1 \mu A$ . Each two-dimensional simulation generates a single element in the ERF for each of the  $M$  locations. Each two-dimensional grid of passive membrane depolarizations provided by simulations is reshaped into a one-dimensional vector to prepare data for the neural network. Shades indicate the mapping of calculated values. The mapping of individual values is indicated by lettered elements, a-i, for the blue plane. The mapping of data from stimulation with each electrode is indicated by numbers 1-5.

[0034] Fig. 6 illustrates a Linear-nonlinear network with GCL and NFL outputs. The GCL is represented by a range of uniformly sampled orientations. Each of these orientations is a two-dimensional representation of the GCL. The output of the linear-

nonlinear network for the GCL is the average activation across all modeled orientations. Since the NFL is better modeled as having a single orientation of fibers, only one representation is used. The activation achieved for this one representation is the output of the model for the NFL. The passage of the single NFL representation through the network is indicated by dashed lines.

[0035] Fig. 7 illustrates the distribution of the ERF from a single electrode in space. The matrix  $V$  contains the ERF vectors for locations in the retina, where each ERF is represented in a row. Hence, each column represents the element of each ERF contributed by a single electrode. Shown here is the value of one column in  $V$  across a plane in the GCL. This is equivalent to the passive membrane potential induced by stimulation with one electrode. The center-to-center electrode pitch is 200  $\mu\text{m}$  and the electrode array is located epiretinally, 100  $\mu\text{m}$  from the surface of the retina.

[0036] Fig. 8 is a map of the objective surface. The objective function is shown with respect to two electrode current amplitudes. As shown in the inset, this mapping was performed with only two electrodes for the purposes of visualization and the target activation pattern consisted of a single phosphene in the GCL. The properties of the shown optimization surfaces are consistent for more complex targets. (a) For monophasic stimulation, a single global stationary point was found, in agreement with the mathematical result. The value of the objective function along the dashed and dotted lines is shown in (b). (c) For biphasic stimulation, a potential  $2^M$  locally optimal solutions exist, of which four can be clearly seen here. The value of the objective function along the dashed and dotted lines is shown in (d).

[0037] Fig. 9 illustrates the optimality of fitting procedure. The optimality of solutions obtained using least-squares initialization followed by the Adam gradient descent algorithm (thick black lines) is compared to solutions obtained from genetic algorithms (dashed line) and from gradient descent within each orthant in the weight space (gray lines). The grey lines represent the loss achieved by initializing the weights in each orthant, followed by Adam gradient descent constrained to that orthant. For three sample images, (a), (c), and (e) show histograms of the log loss achieved from the

orthant search solutions, with the loss obtained by genetic algorithm optimization and unconstrained Adam gradient descent shown. (b), (d), and (f) show the trajectory of the loss during training of the network for each gradient descent.

[0038] Fig. 10 illustrates the optimization of currents for focal stimulation. Using biphasic stimulation, the linear-nonlinear network was trained to stimulate progressively smaller regions of the GCL. The target activation was based on the simulated activation resulting from stimulation with a single electrode. This target was then shrunk by a proportional spread factor (*s.f.*) to 75% and 50% of its size. (a-c) show the target activation pattern, achieved activation, and optimized stimulus currents (trained network weights), respectively, for a rectangular electrode arrangement. (d-f) show the target activation pattern, achieved activation, and optimal stimulus currents for a hexagonal electrode arrangement. Solid black lines represent the 0.25 activation (i.e. 25%) contour lines (1 represents maximum activation).

[0039] Fig. 11 illustrates the optimization of currents for creating virtual electrodes. For several target locations, virtual electrodes with a focal spread factor of 0.75 are achieved using biphasic stimulation via inversion of the linear-nonlinear network model. Virtual electrodes centered in-line between two electrodes and between four electrodes are shown in panels (a) and (b), respectively. Solid black lines represent the 0.25 activation (i.e. 25%) contour lines.

[0040] Fig. 12 illustrates the minimizing activation of passing axons. Using separate target outputs from the GCL and NFL, biphasic electrode currents were optimized for a range of weightings between the two targets. In each case presented, the NFL target is a vector of zeros and the GCL target is a single phosphene (identical to that shown in (a)). (a-d) show the activation achieved in the NFL and GCL for  $\lambda \in 1.0, 0.75, 0.5, 0.25$ , where  $\lambda$  represents the weight assigned to the optimization of activation in the GCL. (e-h) represent the same analysis with the current to the center electrode constrained to 200  $\mu$  A. Solid black lines represent the 0.25 activation (i.e. 25%) contour lines.

[0041] Fig. 13 illustrates the optimization of complex stimuli for epiretinal stimulation. An image of a letter (a) and a real-world scene (b) were used to test the performance of the biphasic linear-nonlinear network inversion scheme. (c) and (d) show the pattern of activation achieved using a naive stimulation strategy, in which the stimulus current delivered to each electrode is proportional to the intensity of the image in the surrounding region. (e-f), (g-h), (i-j), and (k-l) show optimized GCL activation, NFL activation, and stimulus currents for  $\lambda$ -values of 1.0, 0.75, 0.5, and 0.25, respectively, where  $\lambda$  represents the fractional weight applied to the GCL target, and  $1 - \lambda$  is the NFL weight.

[0042] Fig. 14 illustrates the interaction of NFL constraints and current constraints. For the real-world visual scene shown in (b), the patterns of GCL and NFL activation were optimized for various values of the GCL/NFL weight factor,  $\lambda$ , and an inequality constraint applied to the norm of the stimulus current vector,  $\vec{S}_t$ . For each set of two rows, the top row shows activation in the GCL and the bottom row shows activation in the NFL.

[0043] Fig. 15 illustrates the optimization of subretinal stimulation. For the text (a) and real-world (b) scene target images shown in Fig. 13, the biphasic linear-nonlinear network was inverted to optimize subretinal stimuli. In these optimizations, only the target pattern for the GCL (i.e. the target image) was considered, with activation in the NFL being ignored.

[0044] Fig. 16 illustrates the comparison of stimulation of rat and human retinas. The biphasic linear-nonlinear network was inverted to optimize epiretinal stimulus currents for human and rat NFL thicknesses of 100  $\mu$ m and 45  $\mu$ m, respectively. For each retinal geometry, optimizations were performed by each of the target images shown in Fig. 13. (a-d) Achieved GCL and NFL activation pattern for rat-like retinal geometry for  $\lambda$ -values of 1.0 and 0.75. (a-d) Achieved GCL and NFL activation pattern for human-like retinal geometry for  $\lambda$ -values of 1.0 and 0.75.

[0045] Fig. 17 illustrates a computer system for neural stimulation.

### **Description of Embodiments**

[0046] Fig. 2 illustrates a method for neural stimulation. Neural stimulation in this context means the application of an electrical stimulus to neural tissue, such as by a controlled current or a controlled voltage. This can be achieved by a controlled current source that in effect delivers a controlled amount of energy from a battery to a conductive (e.g. metal) electrode. In most cases there are multiple electrodes in an array structure, which may be a linear, two-dimensional or three-dimensional array. The method may be performed by a processor of an implanted stimulation device.

[0047] The processor first configures 201 a forward stimulation model to obtain a calculated neural tissue activation pattern from a given stimulus. “Configure” in this sense means that the processor creates, maintains or stores a data structure that represents a forward stimulation model in a form that allows calculation of neural tissue activation patterns from given stimuli. Fig. 3 illustrates an example forward stimulation model. The forward stimulation model may take a variety of different forms and one example is described further below. Generally, the forward stimulation model can be used to calculate, estimate, or predict an activation pattern given a known stimulus. For example, the model can calculate a perceived image for a given stimulation pattern. That is, the model output can be inspected to assess the amount of blurring or other quality parameters of the perceived image.

[0048] The processor then inverts 202 the forward stimulation model for a desired neural tissue activation pattern to obtain a calculated stimulus. It is important to note that the processor does not invert the model in a general sense so that the inverted model can be used to calculate the stimulus for any desired activation pattern. Instead, the processor inverts the model for a desired activation pattern. This means, when the desired activation pattern changes because the image changes, for example, the processor inverts the forward model again.

[0049] Finally, the processor applies the calculated stimulus to multiple stimulation electrodes. This may involve generating a digital signal that controls a stimulation circuitry, such as a controlled current source, to deliver the stimulation current according to the calculated stimulus. This may be an on/off signal for each electrode for a “black and white” image perception or a multi-valued numerical signal for a “greyscale” image perception.

### Overview

[0050] This disclosure provides a form of the linear-nonlinear model for modelling the activation of a population of retinal ganglion cells, using either experimentally- or computationally-estimated electrical receptive fields (ERF) for each cell in the population. A complication in applying models of spiking activity to the problem of spatial shaping may be that desired spatial patterns of activity are likely to be specified in two-dimensions (such as images recorded from a camera), but a higher-dimensional representation of the retina may be desired (such as multiple two-dimensional layers, or multiple distinct cell types within the same layer). For example, neurites in the ganglion cell layer are modelled using a series of overlaid two-dimensional representations with different characteristic orientations, while mapping activity to a single two-dimensional target image.

[0051] To overcome this complication, the linear-nonlinear model can be generalized to enable for the specification of one or more two-dimensional target activity patterns while modelling a higher-dimensional representation of the retina and its ERFs. This linear-nonlinear model may be identical to a particular type of convolutional neural network and can be used as the forward model in step 201 of method 200. It is noted here that the term neural network in this disclosure is meant to refer to an ‘artificial’ neural network, such as a computer implemented data/software network or electronics hardware network or neuromorphic electronics network, but not a natural neural network, such as a human brain.

[0052] Methods for efficient spatial shaping of retinal ganglion cell (RGC) responses can be demonstrated via the constrained iterative inversion of the neural network model to yield optimized stimulus currents for arbitrary target activation patterns. The inversion problem may be equivalent to the training of the neural network, where the network weights are equal to the optimal set of stimulus currents. The conditions under which the inversion is globally optimal are explored.

[0053] The quality of the achieved activation using the inversion strategy can be tested with a wide variety of target activity patterns ranging from single phosphenes to complex real-world images. A range of stimulation conditions will also be varied to test their importance. These include electrode array location (epiretinal or subretinal), electrode arrangement (hexagonal or grid), stimulation form (monophasic or biphasic), and animal model (human or rodent).

#### The linear-nonlinear model

[0054] As shown in Maturana, M. I., Apollo, N. V., Hadjinicolaou, A. E., Garrett, D. J., Cloherty, S. L., Kameneva, T., Grayden, D. B., Ibbotson, M. R., and Me n, H. (2016). “A simple and accurate model to predict responses to multi-electrode stimulation in the retina”. PLOS Comput. Biol., 12(4):e1004849, which is incorporated herein by reference, the linear-nonlinear model of direct RGC activation under biphasic stimulation for a single neural cell is given by,

$$\mathcal{N}(\vec{v}_1 \cdot \vec{S}_t) = \frac{1}{1 + \exp(-(\vec{v}_1 \cdot \vec{S}_t - c^+))} + \frac{1}{1 + \exp(-(-\vec{v}_1 \cdot \vec{S}_t - c^-))}, \quad (1)$$

where  $\vec{v}_1$  is the electrical receptive field (ERF) and  $\vec{S}_t$  is the vector of first-phase electrode amplitudes in the stimulus delivered at time  $t$ .  $c^+$  and  $c^-$  are the positive and negative thresholds for the spiking nonlinearity, respectively. This captures spiking behavior resulting from both net cathodic-first and net anodic-first stimulation. To simulate the response to monophasic stimulation, the second term on the right-hand-side of Eq. (1) is removed. The ERF,  $\vec{v}_1$ , is an approximation of the cell's response to

stimulation with each electrode. For  $N$  electrodes, the ERF for a single cell is a  $1 \times N$  vector.

[0055] Using the experimental approach described by Maturana (see above) to estimate ERFs at multiple locations in the retina (using multiple recording electrodes) yields a series of ERF vectors. These vectors can be combined into a matrix representation, where each row is a recovered ERF,

$$\mathbf{V} = \begin{bmatrix} \vec{v}_{1,1} \\ \vec{v}_{1,2} \\ \vdots \\ \vec{v}_{1,M} \end{bmatrix}. \quad (2)$$

[0056] Similar to the measuring of ERFs using the experimental approach, a biophysical model can be used to simulate the ERF at arbitrarily many locations in the retina yielding a matrix structure,

$$\mathbf{V} = \begin{bmatrix} \vec{u}_{m,1} \\ \vec{u}_{m,2} \\ \vdots \\ \vec{u}_{m,M} \end{bmatrix}, \quad (3)$$

where each simulated ERF,  $\vec{u}_{m,i}$ , is the vector of membrane potential depolarizations at location  $i$  in response to stimulation from each electrode with a nominal stimulation amplitude. In this case, the number of locations at which the ERF is determined,  $M$ , may be very large. For example, the model may calculate ERFs over a  $10 \mu\text{m} \times 10 \mu\text{m}$  sampling of a plane in the retina. Furthermore, there may be a distinction between different cell types or axonal orientations within that plane, yielding multiple overlaid two-dimensional sets of ERFs.

[0057] Using either of the matrix representations of the ERF described above, the multi-cell/population form of the linear-nonlinear model is:

$$\text{Monophasic: } Y' = \mathcal{N}(\mathbf{V}\vec{S}_t) = \frac{1}{1 + \exp(-(\mathbf{V}\vec{S}_t - c))} \quad (4a)$$

$$\begin{aligned} \text{Biphasic: } Y' = \mathcal{N}(\mathbf{V}\vec{S}_t) = & \frac{1}{1 + \exp(-(\mathbf{V}\vec{S}_t - c^+))} \\ & + \frac{1}{1 + \exp(-(-\mathbf{V}\vec{S}_t - c^-))}, \end{aligned} \quad (4b)$$

[0058] The task of spatial shaping is to find the set of electrode stimuli,  $\vec{S}_t$ , that minimize the error between a desired pattern of activation,  $Y$ , and the activation achieved,  $Y' = \mathcal{N}(\mathbf{V}\vec{S}_t)$ . Importantly, the matrices defining the target activity, the achieved activity, and the ERFs all have the same number of observations or sampled location,  $M$ .

#### The linear-nonlinear model for spatial shaping

[0059] The problem of spatial shaping may be complicated further when attempting to map multiple distinct representations (e.g. cell types, axonal orientations, or different retinal depths) to a single target activity (e.g. a grayscale image). In some examples it is desired that the achieved neural activity resembles to a two-dimensional target image, such as that recorded from a device camera, however, the model of the ERF and spiking activity may include a three-dimensional volume in the retina or distinct descriptions for multiple cell types.

[0060] When attempting to map multiple distinct two-dimensional sets of ERFs, such as multiple axonal orientations or multiple retinal depths, to a single two-dimensional achieved activity equation 3 may no longer hold ( $Y'$  and  $\mathbf{V}$  have incompatible dimensionality). In this case, the desired activity at each two-dimensional location can be approximated by the average of the multiple representations of the ERF at that location: the average two-dimensional activation across cell types and retinal depths. The model to be inverted then becomes

$$\text{Monophasic: } Y' = \frac{1}{R} \sum_r \frac{1}{1 + \exp(-(\mathbf{V}_r \vec{S}_t - c_r))}, \quad (5a)$$

$$\text{Biphasic: } Y' = \frac{1}{R} \sum_r \left( \frac{1}{1 + \exp(-(\mathbf{V}_r \vec{S}_t - c_r^+))} + \frac{1}{1 + \exp(-(-\mathbf{V}_r \vec{S}_t - c_r^-))} \right), \quad (5b)$$

where  $Y'$  is the two-dimensional achieved activity,  $R$  is the number of distinct representations, and  $\mathcal{N}_r(x)$  is the nonlinearity applied to the  $r$ th two-dimensional representation, with spiking threshold,  $c_r$ . This enables, for example, different cells in different retinal layers to have different membrane thresholds.  $\mathbf{V}_r$  is the set of  $M$  ERFs in the  $r$ th two-dimensional representation. Hence, each of the summands in the right-hand-side of Eq. 5 has the same dimension as the output,  $Y'$ .

[0061] Furthermore, it may be advantageous to map multiple distinct representations of the retina to several, but fewer, target activity patterns. An example of this is the specification of separate target activity patterns in the nerve fiber layer (NFL) and the ganglion cell layer (GCL). Epiretinal stimulation may be challenging due to the overlying layer of passing axons in the NFL that may be preferentially stimulated. However, activation of these passing fibers could be avoided using carefully chosen simultaneous multi-electrode stimulation. To address challenges such as this, when different patterns of activation are desired in different retinal layers or cell types, a spatial shaping solution should enable for the specification of multiple two-dimensional target activities.

[0062] For multiple two-dimensional target activities, such as activation in the the GCL and NFL, Eq. (5) may be further extended to be:

$$\text{Monophasic: } Y'_1 = \frac{1}{R_1} \sum_{r \in S_1} \frac{1}{1 + \exp(-(\mathbf{V}_r \vec{S}_t - c_r))}, \quad (6a)$$

16

$$Y'_2 = \frac{1}{R_2} \sum_{r \in \mathcal{S}_2} \frac{1}{1 + \exp(-(\mathbf{V}_r \vec{S}_t - c_r))}, \quad (6b)$$

$$\vdots$$

$$\text{Biphasic: } Y'_1 = \frac{1}{R_1} \sum_{r \in \mathcal{S}_1} \left( \frac{1}{1 + \exp(-(\mathbf{V}_r \vec{S}_t - c_r^+))} + \frac{1}{1 + \exp(-(-\mathbf{V}_r \vec{S}_t - c_r^-))} \right), \quad (6c)$$

$$Y'_2 = \frac{1}{R_2} \sum_{r \in \mathcal{S}_2} \left( \frac{1}{1 + \exp(-(\mathbf{V}_r \vec{S}_t - c_r^+))} + \frac{1}{1 + \exp(-(-\mathbf{V}_r \vec{S}_t - c_r^-))} \right), \quad (6d)$$

$$\vdots$$

where  $Y'_i$  is the  $i$ th achieved activation pattern,  $R_i$  is the number of ERF representations that are mapped to the  $i$ th activation pattern, and  $\mathcal{S}_i$  is the set of representations that are mapped to the  $i$ th activation pattern. For example, if each of the multiple retinal representations,  $\mathbf{V}_r$ , are depths in the retina, then  $\mathcal{S}_1$  may correspond to the simulated depths occupying the GCL and  $\mathcal{S}_2$  may correspond to the simulated depths occupying the NFL. The achieved two-dimensional GCL and NFL activity,  $Y'_1$  and  $Y'_2$ , respectively, are the average of the activity achieved in those three-dimensional retinal layers. Similar approaches may be used to target activation to certain cell subtypes while limiting activation of others. During model inversion, each

specified target pattern,  $Y_i$ , may be weighted according to its importance. For simplicity, the following discussion of model inversion will focus on a single output activity pattern,  $Y'$ , and target activity,  $Y$ , as in Eq. (5), but will later be generalized to the case of multiple output activity patterns.

#### The linear-nonlinear neural network

[0063] In the case of monophasic stimulation, the problem of inverting the model can be solved using existing model training procedures in machine learning. Inspection of Eq. (4a) reveals that it is equivalent to a logistic regression model in the form of:

$$y' = P(y | \theta) = \frac{1}{1 + e^{-\theta^T x}}, \quad yK \sim \mathcal{B}(K, p) \quad (7)$$

where  $x$  is a vector of observed regressors,  $\theta$  is a vector of regression coefficients,  $y'$  is the predicted probability/proportion,  $y$  is the true value, and  $\mathcal{B}$  is the binomial distribution. In this form of logistic regression, the distribution if  $y$  is free to follow either a Bernoulli distribution (i.e.  $K=1$  and  $y \in \{0,1\}$ ) or a Binomial proportion. To translate this into terms consistent with the linear-nonlinear model,  $y$  (and  $y'$ ) are activation probabilities or proportions,  $\theta$  is the electrode current vector,  $\vec{S}_i$ , and  $x$  is an ERF.

[0064] Hence, given a set of target activations,  $Y$ , solving this problem for monophasic stimulation is equivalent to fitting a logistic regression model, with  $Y$  and  $\mathbf{V}$  the dependent and independent variables, respectively. The fitting procedure minimizes some measure of distance between the desired activation,  $Y$ , and the achieved activation,  $Y'$ . The fitted model coefficients are the optimized electrode currents,  $\vec{S}_i$ .

[0065] It is important to note the difference between this application of logistic regression and its use in predictive modelling and inference. In the case of prediction or

inference, models are typically trained on a single large set of data against pre-coded binary labels or probabilities. Trained models are then fixed and used to either infer relationships in the training data (as for inference) or used to predict labels on future data (as in predictive modelling). In contrast, for its application in optimization of stimulus currents (i.e. model coefficients), the labelled probabilities in the training data are rapidly changing as target images change. This means that the trained model is not fixed, but continuously retrained against new images/labels. As such, for stimulus optimization, it is the algorithm used for model training (its accuracy and efficiency) that is of interest, as opposed to the predictive power of the model itself.

[0066] Similarly, equations 5 or 6, which represent a mapping from a higher-dimensional retinal representation to one or several output images, are equivalent to a particular type of feed-forward convolutional neural network. This equivalent network is shown in Fig. 4 and consists of an input layer 401, a *non-overlapping* convolution layer 402, a replication layer 403, an offset layer 404, a sigmoidal activation layer 405 and an average-pooling layer 406. Henceforth, this neural network architecture will be referred to as the *linear-nonlinear network*. The convolutional layer 402 corresponds to the scaling of the multiple different ERF representations by a common set of electrode currents (i.e. the same stimuli is experienced by the whole retina) to yield the approximate passive membrane potential. The offset layer 404 and sigmoidal activation layer 405 correspond to the application of the spiking threshold to the membrane potential and the spiking nonlinearity, respectively. Finally, the average pooling layer 406 combines the multiple representations using their average to determine the average two-dimensional spiking probability. Each of the layers in the neural network 400 can be equated to mathematical operations in Eq. (5) as shown in the right hand panel of Fig. 4.

[0067] In one example, the only network weights that are trained are those in the convolutional layer 403,  $\vec{S}_t$ , and these correspond, once trained, to the optimized electrode currents. Note that the convolutional layer 402 does not strictly perform a convolution over the input (as the filter is applied in a non-overlapping manner),

however, the layer weights,  $\vec{S}_l$ , are fixed across all  $R$  representations in the manner of a typical convolutional layer. Furthermore, since Eq. 4 is a special case of Eq. 5 in which there is only a single representation,  $r$ , the neural network approach can be applied in all cases considered here.

[0068] As for logistic regression, this application of neural networks is unconventional. Training/optimization/inversion is not performed once using stationary data for prediction purposes, but is retrained for each target image. Hence, the efficiency of the chosen training or optimization algorithm is of key importance. The translation of Eq. 5a into an equivalent neural network enables ready application of the wealth of neural network optimization algorithms available.

[0069] In the case of biphasic stimulation, in which stimuli of opposite polarity may result in approximately equal activation, the spiking nonlinearity of the linear-nonlinear model is double sided, as in Eq. 4b and Eq. 5b. As a result, no unique inverse may exist and the model may not be able to be represented as an equivalent logistic regression model. However, the linear-nonlinear network described above may be modified, by applying a double-sided nonlinearity, to model the response to biphasic stimuli, as shown in Fig. 4. An additional neural network layer 405 is added to sum the output of each side of the nonlinearity as in Eq. 4b.

[0070] Although no globally optimal set of weights/currents may exist due to the double-sided nonlinearity of Eq. 5b, backpropagation combined with an iterative optimization algorithm may be used to find locally optimal weights. To determine the effectiveness of this approach, locally optimal solutions found using iterative approaches were compared to globally, or near-globally, optimal solutions obtained by search methods, or approximately optimal solutions obtained using evolutionary algorithms, as described below.

#### Neural network training algorithm

[0071] For some examples, the model output,  $Y'$ , and the target activation,  $Y$ , represent a two-dimensional map of spiking probability between 0 and 1. Alternatively, these values may represent a proportion of cells that are spiking. As a result, error in the fit of a model to  $Y$  is determined using binomial logarithmic loss :

$$\mathcal{L}(\vec{S}_t) = -\frac{1}{M} \sum_m^M (1 - y_m) \log(1 - y_{m'}) + y_m \log y_{m'}, \quad (8)$$

where  $M$  is the number of locations at which the ERF is estimated,  $y_m$  is the target activation at location  $m$ ,  $y_{m'}$  is the activation achieved with stimulus currents,  $\vec{S}_t$ , at location  $m$ . Depending on context, this measure is sometimes referred to as negative binomial log-likelihood or binary cross-entropy. Eq. 8 is the objective function used for training models and is minimized in order to determine an optimal set of electrode currents. The optimization/inversion problem may be expressed as,

$$\begin{aligned} & \underset{\vec{S}_t}{\text{minimize}} \quad \mathcal{L}(\vec{S}_t; Y) \\ & \text{subject to} \quad \vec{S}_t \in C, \end{aligned} \quad (9)$$

where  $C$  is the set of allowed values for the stimulus currents  $\vec{S}_t$ . If unconstrained,  $C = \mathbb{R}$ . The possibility of constraining the optimization using  $C$  is addressed below. Hereafter, the terms training, fitting, and inverting will be used interchangeably.

[0072] Due to the greater generality of the linear-nonlinear network when compared to the logistic regression form of the linear-nonlinear model, this disclosure is focused on the inversion of the neural network. The linear-nonlinear network was trained using a modification of the mini-batch gradient descent algorithm known as Adam: Adaptive Moment Estimation as described in Kingma, D. P. and Ba, J. L. (2015). Adam: A method for stochastic optimization. In Int. Conf. Learn. Represent. 2015, pages 1-15, which is incorporated herein by reference. Mini-batch gradient descent is an algorithm which uses a randomly sampled subset of the training data for each update of the

network's weights,  $\vec{S}_t$ . At each step,  $s$ , mini-batch gradient descent updates the weights using the following general update rule:

$$\vec{S}_{t,s+1} = \vec{S}_{t,s} - \eta \cdot \gamma(\vec{S}_{t,s}; \mathbf{V}_{\text{batch}}; Y'_{\text{batch}}), \quad (10)$$

where  $\eta$  is the learning rate and the function  $\gamma$  is related to the gradient of the objective function,  $\mathcal{L}$ , with respect to  $\vec{S}_t$ .  $\mathbf{V}_{\text{batch}}$  and  $Y_{\text{batch}}$  are random subsamples of fixed size (batches) of the independent and dependent variables, respectively.

[0073] The Adam algorithm supplies definitions for  $\eta$  and  $\gamma$  in Eq. 10, and utilizes the gradient descent concepts momentum and adaptive learning. Adam enables for the adaption of the learning rate for each network weight individually. In addition, Adam incorporates a momentum-like mechanism in the form of an exponentially-decaying mean of past gradients of the objective function and an exponentially-decaying mean of the square of those gradients. These two means are estimates of the first moment (mean) and second moment (uncentered variance) of the objective function gradient. Momentum mechanisms such as this enable for those weights in  $\vec{S}_t$  for which the gradient is consistently positive (negative) to gain momentum in the positive (negative) direction. This enables for gradient descent to proceed more rapidly in directions that offer the most benefit.

[0074] Adam uses the following user-defined parameters:

- $\alpha$  : the step size.
- $\beta_1, \beta_2 \in [0,1)$  : the exponential decay rates for the moment estimates.
- $\mathcal{L}(\vec{S}_t)$  : the objective function.
- $\vec{S}_t$  : the initial weight vector.

[0075] Given the above parameters, the Adam algorithm is implemented as follows:

---

Algorithm 1: *Adam* gradient descent optimization algorithm (Kingma and Ba, 2015).

---

```

 $\mu_0 \leftarrow 0$  (Initialize first moment);
 $\nu_0 \leftarrow 0$  (Initialize second moment);
while  $\vec{S}_t$  has not converged do
     $s \leftarrow s + 1$ ;
     $g_s \leftarrow \nabla \mathcal{L}(\vec{S}_t)$  (Get gradients of objective function w.r.t. weights);
     $\mu_s \leftarrow \beta_1 \mu_{s-1} + (1 - \beta_1) g_s$ ;
     $\nu_s \leftarrow \beta_2 \nu_{s-1} + (1 - \beta_2) g_s^2$ ;
     $\hat{\mu}_s \leftarrow \mu_s / (1 - \beta_1^s)$ ;
     $\hat{\nu}_s \leftarrow \nu_s / (1 - \beta_2^s)$ ;
     $\vec{S}_t \leftarrow \vec{S}_t - \alpha \hat{\mu}_s / (\sqrt{\hat{\nu}_s} + \epsilon)$ ;
end

```

---

[0076] The solution is deemed to have converged when the step-on-step improvement in the objective function,  $\mathcal{L}$ , drops below a specified value, such as 0.0001. In the above algorithm, the gradient of the optimization function with respect to the network weights,  $g$ , is determined using backpropagation.

[0077] The linear-nonlinear neural network was specified and trained in statistical computing software R 3.4.2 using third-party R packages keras and tensorflow, which offer R interfaces to neural network software packages Keras and Tensorflow, respectively. Efficient manipulation of data within R was achieved using the third-party R package data.table. Genetic algorithms were generated using the third-party R package GA.

#### Existence of globally optimal solutions

[0078] To demonstrate that a fitting procedure or optimization algorithm can find the global optima, it is useful to examine the topology of the objective function, in our case  $\mathcal{L}$ , with respect to the model coefficients being fit, in our case  $\vec{S}_t$ . Since the Adam training procedure is a gradient descent method, it is useful to consider the possibility that gradient descent will get caught in any local optima and terminate with a sub-optimal solution.

[0079] As described previously, when the target activation,  $Y$ , is specified using the same number of two-dimensional representations as the input,  $V$ , spatial shaping can be achieved via the inversion/training of Eq. 4. In the case of monophasic stimulation, this inversion problem is identical to training a logistic regression model (also known as a binomial generalized linear model). The fitting of generalized linear models such as this by maximizing log-likelihood (or minimising logarithmic loss) is a convex problem. This means that every local optima is also a global optima and the gradient of the objective function is monotonically non-decreasing with respect to each model coefficient. Hence, for the case of monophasic stimulation and compatible input and output representations, any gradient descent algorithm (such as Adam) will obtain an estimate arbitrarily close to the global optima given sufficient descent iterations. This result holds for both logistic regression and neural network implementation of Eq. 4a as they are equivalent mathematically and minimize the same logarithmic loss function.

[0080] In the case of biphasic stimulation, gradient descent is not guaranteed to find the globally optimal set of stimulus currents. This can be seen from inspection of Eq. 4b. If we assume that  $c_r^+ = c_r^-$ , the same  $Y'$  is achieved with either of  $\vec{S}_r$  or  $-\vec{S}_r$ . Hence, if one local optima exists at  $\vec{S}_r$ , another will exist at  $-\vec{S}_r$ . Furthermore, since  $\mathbf{V}_r \vec{S}_r$  is a vector of length  $M$ , each element of which is passed through a double-sided nonlinearity, there are potentially  $2^M$  locally optimal solutions to the inversion problem for a given  $Y$ .

[0081] In the case of a single two-dimensional representation for  $Y$  and multiple two-dimensional representations included in  $V$ , see Eq. 5. As for above, in the case of biphasic stimulation, many locally optimal solutions to Eq. 9 exist. Since the sum of convex functions is convex, it suffices to show that the summand of Eq. 5 is convex. Furthermore, since the target,  $Y$ , is fixed and has no dependence on  $\vec{S}_r$ , I will consider only the logarithmic terms in Eq. 9. I will denote this simplified objective function as  $\mathcal{K}$ . Combining equation Eq. 5a and the summand of Eq. 9 gives,

$$\begin{aligned}
\mathcal{K}(\vec{S}_t) &= \log\left(1 - \sum_r \frac{\exp(\gamma(r, m))}{1 + \exp(\gamma(r, m))}\right) \\
&+ \log\left(\sum_r \frac{\exp(\gamma(r, m))}{1 + \exp(\gamma(r, m))}\right) \\
&= \log\left(1 - \frac{\sum_r \exp(\gamma(r, m)) \prod_{q \neq r} (1 + \exp(\gamma(q, m)))}{\prod_r (1 + \exp(\gamma(r, m)))}\right) \\
&+ \log\left(\frac{\sum_r \exp(\gamma(r, m)) \prod_{q \neq r} (1 + \exp(\gamma(q, m)))}{\prod_r (1 + \exp(\gamma(r, m)))}\right) \\
&= \log\left(\prod_r (1 + \exp(\gamma(r, m)))\right) \\
&- \sum_r \exp(\gamma(r, m)) \prod_{q \neq r} (1 + \exp(\gamma(q, m))) \\
&- 2 \log\left(\prod_r (1 + \exp(\gamma(r, m)))\right) \\
&+ \log\left(\sum_r \exp(\gamma(r, m)) \prod_{q \neq r} (1 + \exp(\gamma(q, m)))\right),
\end{aligned}$$

where

$$\gamma(r, m) = \vec{v}_{r, m} \cdot \vec{S}_t - c_r,$$

and where  $\vec{v}_{r,m}$  is the ERF at location  $m$  for representation  $r$ . Notice that each of the logarithm terms in the equation above can be represented in the form,

$$\log(\sum_i \exp(a_i \cdot \vec{S}_t - b_i)).$$

[0082] Since the softmax function,  $\log(\sum_i \exp(x_i))$ , is convex and since weighted summation and composition with an affine function preserves convexity,  $\mathcal{K}(\vec{S}_t)$ , and hence the objective function,  $\mathcal{L}(\vec{S}_t)$ , are convex functions. Therefore, gradient descent will obtain the global optima for monophasic stimulation using the linear-nonlinear neural network.

[0083] To demonstrate the existence of a single (multiple) optima for monophasic (biphasic) stimulation, network weights were varied over a grid to construct a map of the objective function with respect to the weights. This analysis is presented below.

#### Network weight initialization

[0084] To ensure reproducibility of results and decrease training time, the weights of the linear-nonlinear network were initialized using multiple linear regression. In the case of monophasic stimulation with an equal number of input and output representations, Eq. 4a was rearranged to give,

$$\mathbf{V}\vec{S}_t = -\log(1/Y' - 1) + c. \quad (11)$$

[0085] Based on this expression, an initialized set of weights,  $\vec{S}_t$ , is found by solving the multiple linear regression problem,

$$\underset{\vec{S}_t}{\text{minimize}} \quad \|\mathbf{V}\vec{S}_t - (-\log(1/Y - 1) + c)\|. \quad (12)$$

[0086] In practice, to ensure that  $\log(1/Y-1)$  is defined,  $Y$  is replaced by  $Y(1-2\epsilon)+\epsilon$ , where  $\epsilon$  is a small positive value ( $10^{-10}$  in this example).

[0087] In the case of a single output representation and multiple input representations, a direct inversion as in Eq. 11 may not exist. Instead, an approximate inversion is obtained by averaging  $\mathbf{V}$  across representations, before using the same approach as in Eqs. 11 and 12. This is equivalent to moving the average-pooling layer shown in Fig. 4 to be before the convolutional layer. Although this process does not find the optimal solution, it seeds the network weights at a more sensible location than random initialization and significantly reduces the required number of gradient descent training iterations.

[0088] As discussed above, for biphasic stimulation there are potentially many locally optimal solutions to the inversion problem. In the absence of a method for choosing one solution over another, the aforementioned initialization approach for monophasic stimulation was also used for biphasic stimulation. The optimality of the solution achieved from this this initialization is explored below.

#### Stimulus current constraints

##### *Inequality constraints*

[0089] As introduced in Eq. 9, the optimization of electrode currents may be constrained. This enables the training algorithm to conform to specific power and safety constraints on the amount of current delivered by an electrode array to tissue. Inequality constraints may be applied in some examples to limit the L2-norm of the stimulus current vector,  $\|\vec{S}_t\|_2$ , to remain under a fixed value. Other constraints, such as limiting the infinity-norm (i.e. limiting the maximum electrode current magnitude) and the L1-norm (i.e. limiting the sum of electrode current magnitudes) may be more directly applicable to the problem of multi-electrode array stimulation, may be used as constraints in the neural network.

### *Equality constraints*

[0090] It is also advantageous to be able to apply equality constraints on a subset of network weights (i.e. electrode currents) during model training. An example use case is the development of methods for focal stimulation. Current delivered to a single electrode can be fixed at a pre-specified value, and the current delivered to surrounding electrodes may be optimized in order to narrow the achieved region of activation.

[0091] To impose equality constraints on a subset of electrodes, it is possible to split the electrode amplitudes  $\vec{S}_t$ , and the components of the ERF sets,  $\mathbf{V}_r$ , into fixed and trainable components. Eq. 5b can be modified to,

$$Y' = \frac{1}{R} \sum_r \left( \frac{1}{1 + \exp(-(\mathbf{V}_r^f \vec{S}_t^f + \mathbf{V}_r^f \vec{S}_t^f - c_r^+))} + \frac{1}{1 + \exp(-(-\mathbf{V}_r^f \vec{S}_t^f - \mathbf{V}_r^f \vec{S}_t^f - c_r^-))} \right), \quad (13)$$

where  $\vec{S}_t^f$  is the set of fixed electrode currents and  $\mathbf{V}_r^f$  is the ERF elements corresponding to those electrodes. In order to include this type of constraint in the neural network shown in Fig. 2.3, an additional offset layer is added between the convolutional and replication layers which adds in the fixed terms,  $\mathbf{V}_r^f \vec{S}_t^f$ .

### *Inversion using genetic algorithms*

[0092] As mentioned above, there may be many locally optimal solutions to the inversion of the linear-nonlinear network for biphasic stimulation. Hence, iterative gradient descent optimization algorithms may not be guaranteed to find the global optima. To assess the optimality of the stimulus currents obtained using gradient descent, they were compared to the stimulus currents obtained by inverting Eq. 5b using a genetic algorithm.

[0093] Genetic algorithms are a universal search algorithm, meaning that they search across the entire weight space for the best solution. They are commonly used to generate high-quality solutions by relying on evolution-inspired processes including mutation, crossover and selection. Although a genetic algorithm is not guaranteed to find the global optima, it searches through a much wider class of potential solutions, generally at the expense of efficiency.

[0094] Starting with a randomly generated population of weight sets,  $\{S_{t,1}, S_{t,2}, \dots, S_{t,n}\}$ , a genetic algorithm was employed to *evolve* the the population toward an optimal solution. This is achieved by:

1. Calculating the value of the loss function,  $\mathcal{L}$ , for each set of weights in the population.
2. Selecting a proportion,  $p_s$ , of the population based on the value of the loss function. This subset is *bred* with each using crossover: randomly switching weights between weight sets (with probability  $p_c$ ) to maintain the same population size. A small proportion,  $p_e$ , of the very best weight sets from the previous generation are carried forward unchanged.
3. Mutating the new generation by randomly altering a proportion,  $p_m$ , of weights in each weight set.
4. Repeating steps 1-3 until a satisfactory solution is found.

[0095] The results presented here used the following parameters to control the algorithm:  $p_s = 0.5$ ,  $p_c = 0.8$ ,  $p_e = 0.05$ ,  $p_m = 0.1$ .

#### *Biophysical model of the ERF*

[0096] To this point, methods relating to the linear-nonlinear model, the linear-nonlinear network and their inversion are not specific to the method of estimation of the ERFs across the retina. The above descriptions may be used to develop an approach to spatial shaping for ERFs determined either experimentally or derived from biophysical simulation. A four-layer biophysical model of the retina can be used to

derive a biophysical approximation of the set of required ERFs,  $V_r$ , for all representations,  $r$ .

[0097] The linear component of the model corresponds to a linear approximation of the integration of stimulus-induced transmembrane currents in a cell. That is, the linear component is the linear subthreshold membrane dynamics and the nonlinear component is a fixed spiking nonlinearity. The model may be based on several underlying assumptions, that supplies a linear, closed-form description of subthreshold membrane depolarization resulting from direct electrical stimulation. This model may be used to approximate direct RGC activation from an arbitrary set of stimuli across the entire retina at a desired resolution.

[0098] The passive membrane potential in response to direct electrical stimulation is given by:

$$\hat{V}_m(k_x, k_y, z, \omega) = -\frac{k_y^2 \lambda^2(\omega)}{1 + k_y^2 \lambda^2(\omega)} \hat{\phi}_e(k_x, k_y, z, \omega), \quad (14)$$

where  $\hat{V}_m$  is the membrane potential in the  $x, y, t$ -Fourier domain and  $\hat{\phi}_e$  is the extracellular voltage in the  $x, y, t$ -Fourier domain resulting from stimulation with  $N$  electrodes, where the  $x$  and  $y$ -directions are parallel to the surface of the retina and  $z$  represents depth.  $k_x$ ,  $k_y$ , and  $\omega$  are the Fourier domain pairs of  $x$ ,  $y$ , and  $t$ , respectively. The frequency-dependent length constant,  $\lambda(\omega)$ , is given by:

$$\lambda(\omega) = \frac{\sqrt{r_m}}{\sqrt{r_i(1 + 2\pi a j \omega r_m C_m)}},$$

where  $r_m$  and  $r_i$  are the membrane and intracellular resistance per unit length, respectively,  $C_m$  is the membrane capacitance per unit area, and  $a$  is the radius of the simulated neurites. The calculation of  $\phi_e(k_x, k_y, z, \omega)$  in Eq. 14 depends on the stimuli

delivered to each electrode and the type of retinal stimulation (epiretinal or subretinal) being simulated.

[0099] Eq. 14 can be used to determine the Fourier-domain representation of the passive membrane depolarization at a particular depth,  $z$ , in the retina in response to stimulation with electrodes. The inverse Fourier transform with respect to  $k_x$ ,  $k_y$ , and  $\omega$  is then calculated. This yields the time course of the passive membrane depolarization in an  $x - y$  plane. The time course is sampled at  $t_{\text{meas}} = 0.1$  ms after stimulus onset to obtain a single depolarization value per location.

[0100] Fig. 5 demonstrates how a set of ERFs,  $\mathbf{V}_r$ , were generated using the four-layer biophysical model. Since each element of the ERF is the approximate passive membrane depolarization in response to stimulation with a single electrode, simulations were conducted in which each electrode was used to deliver  $1 \mu\text{A}$  of current and the resulting membrane depolarization at all desired locations was calculated using Eq. 14. In this way, the set of ERFs for a particular representation,  $r$ , is:

$$\mathbf{V}_r = \begin{bmatrix} \vec{u}_{m,1} \\ \vec{u}_{m,2} \\ \vdots \\ \vec{u}_{m,M} \end{bmatrix} = \begin{bmatrix} V_{m,1}(x_1, y_1, z_r, t_{\text{meas}}) & V_{m,2}(x_1, y_1, z_r, t_{\text{meas}}) & \dots \\ V_{m,1}(x_2, y_2, z_r, t_{\text{meas}}) & V_{m,2}(x_2, y_2, z_r, t_{\text{meas}}) & \dots \\ \vdots & \vdots & \vdots \\ V_{m,1}(x_M, y_M, z_r, t_{\text{meas}}) & V_{m,2}(x_M, y_M, z_r, t_{\text{meas}}) & \dots \end{bmatrix}, \quad (15)$$

where  $V_{m,e}(x_i, y_i, z_r, t_{\text{meas}})$  is the membrane depolarization at  $x - y$  location  $(x_i, y_i)$  induced by a  $1 \mu\text{A}$  stimulation current delivered by electrode  $e$ . Note that a representation here corresponds to a particular fiber orientation and retinal depth (i.e. the GCL or the NFL).

[0101] To assess the degree to which desired stimulation was occurring in the GCL and unwanted stimulation was occurring in the NFL, ERFs were calculated for a two-dimensional plane in each retinal layer. The GCL is represented by considering fibers with a uniform distribution of orientations, whereas the NFL may be approximated as having a single parallel fiber orientation. To account for this, multiple ERF representations were used to describe the GCL, with each representation describing the set of ERFs for fibers with a specific orientation. One representation was used for the NFL, resulting in a total of  $R = R_{\text{GCL}} + 1$  representations.

[0102] One example neural network architecture is shown in Fig. 6. As can be seen, the model consisted to two outputs, one for the GCL activation and one for the NFL activation. To fit this model, two target activations,  $Y_{\text{GCL}}$  and  $Y_{\text{NFL}}$  are specified. Since it is ideal to minimize activation of the NFL, for all optimizations presented here in which activation in the NFL is considered, the target for the NFL was the zero matrix (i.e. no activation). A range of different targets were used for the GCL target activity.

[0103] Finally, since the linear-nonlinear network was being fit to multiple target patterns of activation, a weighting factor,  $\lambda$ , was applied which varied the relative importance of each target in the optimization problem. The objective function in Eq. 72 then becomes:

$$\begin{aligned} \mathcal{L}(\vec{S}_t) = & -\frac{\lambda}{M} \sum_m (1 - y_{\text{GCL},m}) \log(1 - y_{\text{GCL},m}) \\ & + y_{\text{GCL},m} \log y_{\text{GCL},m} \\ & - \frac{(1-\lambda)}{M} \sum_m (1 - y_{\text{NFL},m}) \log(1 - y_{\text{NFL},m}) \\ & + y_{\text{NFL},m} \log y_{\text{NFL},m}, \end{aligned} \quad (16)$$

[0104] The importance weighting,  $\lambda$ , was varied from 0 to 1, where 1 gives importance only to the GCL target and ignores activation of the NFL, whereas 0 gives importance only to the NFL target. For the network weight initialization described

above, the same  $\lambda$  weighting was used to solve the norm minimization problem in Eq. 12.

[0105] Along with assessing the quality of inversion, analyses were also conducted to assess the differences in achievable stimulation between epiretinal and subretinal stimulation. This was achieved by reversing the order of the retinal layers in the four-layer biophysical model used to calculate the ERFs. All biophysical simulations of the ERF were conducted using MATLAB (The Mathworks, Release 2016a).

#### *Analytic parameters*

[0106] Unless otherwise stated, all simulations and optimizations employed the following parameters:

[0107] 1. **Electrode location.** Except for comparisons between epiretinal and subretinal stimulation, epiretinal electrode array placement was used. For epiretinal stimulation, the electrode array was located 100  $\mu$  m away from the inner surface of the retina in the vitreous body. For subretinal stimulation, the electrode array was located on the outer surface of the modeled retinal layers. In both cases, the array is assumed to be flat and parallel to the retinal surface.

[0108] 2. **Electrode array geometry.** Arrays had electrodes with diameters of 100  $\mu$  m and a center-to-center electrode pitch of 200  $\mu$  m. Both rectangularly- and hexagonally-arrange electrode arrays were tested. In both cases, the electrode pitch was kept constant at 200  $\mu$  m. For all simulations/optimization of rectangularly-arranged electrodes, a 9×9 grid of regularly spaced electrodes was used.

[0109] 3. **Retinal anatomy.** This research focused on human or primate-like retinal geometry. As such, the thickness of the NFL was 100  $\mu$  m.

[0110] 4. **RGC spiking thresholds.** In all results presented here, the sensitivity of the cells was assumed to be equal for net-cathodic and net-anodic first stimuli:  $c_r^+ = c_r^-$ .

[0111] 5. **Target locations.** Activation in the GCL was targeted to a plane a short distance into the GCL on the side of the layer closest to the electrode. For epiretinal stimulation, the target plane was 10  $\mu$  m into the GCL from the NFL. Similarly, the resulting activation in the NFL was calculated at 10  $\mu$  m from the the retinal surface.

[0112] 6. **Stimulus waveform.** Biphasic stimulation was used for all presented results, except where comparisons between monophasic and biphasic stimulation are presented.

## Results

### *Spatial distribution of ERFs*

[0113] A primary reason that spatial shaping methods are used for effective multi-electrode stimulation is that there is significant interaction of the electric fields and stimulating currents generated by adjacent electrode by the time they reach the target neural tissue. If the distance were such that this interaction could be neglected, spatial shaping may not be necessary. However, due to the push in recent years for higher-density electrode arrays, this is unlikely to be a feasible approach. Furthermore, large separations between electrodes will result in discrete spotted percepts, precluding any recovery of complex, continuously spatially varying visual scenes.

[0114] Fig. 7 demonstrates the resulting passive membrane potential from stimulation with a single electrode. This is equivalent to the elements of the ERF for this electrode (a single column) across all positions (rows) in the matrix  $V$ . In this simulation, the center-to-center electrode pitch was 200  $\mu$  m, and the electrode array was located epiretinally, 100  $\mu$  m from the surface of the retina. This geometry is used for most of the results presented in the remainder of this Chapter. The spread shown in Figure 7 is that observed in the GCL, the region being targeted by stimulation. As can be seen, due to the spread of the electric field from one electrode and the corresponding distribution of the ERF, interactions between adjacent electrodes are significant for the electrode

geometry and placement modeled here, necessitating methods for spatial shaping in selecting simultaneous multielectrode stimulation parameters.

#### *Visualization of the objective function*

[0115] To validate the analysis presented in the Methods regarding the existence of local and global optima for biphasic and monophasic stimulation, respectively, the surface of the objective function,  $\mathcal{L}$ , was mapped, as shown in Fig. 8. The neural network architecture shown in Fig. 4 was defined, along with a simple target stimuli which is shown in the inset. For simplicity, this experiment was conducted using only two electrodes (so that it might be visualized) and with a single centered phosphene as the activation target. The electrode currents/network weights assigned to each electrode were varied in a regular, two-dimensional grid with a resolution of  $40 \mu\text{A}$  and the value of the objective function was calculated at each grid point.

[0116] As can be seen in Fig. 8, for monophasic stimulation, there is a single globally optimal solution to the inversion problem (at  $[80 \mu\text{A}, 80 \mu\text{A}]$ ). Hence, a gradient descent algorithm is guaranteed to find the optimal set of stimulus currents given enough optimization iterations. In contrast, there are four distinct local optima for biphasic stimulation; one in each quadrant of the stimulus space. In addition, since the positive and negative thresholds are equal, the loss function is reflected about the origin (i.e. each local optima has a pair in the opposite quadrant). For biphasic stimulation, different optima will be obtained by gradient descent depending on the starting position.

#### *Optimality of gradient descent solutions*

[0117] The large number of potential local optima for the biphasic optimization can be addressed by an assessment of the level of optimality of solutions obtained using the Adam gradient descent algorithm. This was achieved by comparing the performance of solutions obtained using Adam with those obtained using (1) a genetic algorithm and (2) a wide range of weight initializations followed by gradient descent. For (2), to

ensure that a suitable amount of the potential solution space is explored, a gradient descent optimization was performed in each orthant. This was achieved by initializing the electrode currents in each orthant and performing gradient descent from this point while constraining the optimization to remain in the same orthant. To make this approach feasible, optimizations were performed with only 9 electrodes, yielding  $2^9 = 512$  orthants (and 512 gradient descent solutions for comparison).

[0118] Fig. 9 shows the results of this analysis for three sample target images. Fig 9(a), (c), and (e) are histograms which demonstrate that in each case, the loss achieved via Adam gradient descent (thick black lines) is as low or lower than that achieved using either the genetic algorithm (dashed line) or orthant search (grey lines) approaches. Fig 34(b), (d), and (f) show the trajectory of the loss during the gradient descent procedure.

#### *Current steering and virtual electrodes*

[0119] Two enhancements to retinal stimulation are current steering and virtual electrodes. Both of these techniques aim to control the local or spread of retinal neural activation under electrical stimulation through simultaneous stimulation with multiple electrodes. As such, the optimization framework described here is suitable for achieving the outcomes intended by both techniques. Current steering (or focused multipolar stimulation) aims to recruit multiple neighbouring electrodes to narrow the region of activation, thereby increasing the achievable resolution from a retinal prosthesis. Virtual electrodes refer to stimuli aimed at eliciting activation at locations between electrodes by recruiting multiple neighbouring electrodes.

#### *Focal activation*

[0120] Fig. 10 demonstrates the optimization of electrode currents for focal stimulation (current steering) for a range of decreasing spread factors. Here, a spread factor of 1 corresponds to the radius of activation achieved using stimulation with a single electrode (with a stimulus of 200  $\mu$  A). A lower spread factor corresponds to a

smaller radius of activation. In agreement with experimental research on current steering and focal activation, optimization of stimulus currents achieves focal activation by stimulating with surrounding electrodes at an opposite polarity to the center electrode. No significant difference could be seen between rectangular or hexagonally-arranged electrodes. Due to the tissue anisotropy introduced by the NFL, the optimal current delivered to the surrounding electrodes is not quite radially symmetric about the center electrode.

#### *Virtual electrodes*

[0121] Fig. 11 demonstrates the optimization of electrode currents for creating virtual electrodes. This is achieved while simultaneously focusing the field of activation. Virtual electrodes are targeted at multiple locations: mid-way between two electrodes and mid-way between four electrodes. In each case, inversion of the linear-nonlinear network recovers electrode currents which accurately reproduce the target activity.

#### *Minimizing activation of the NFL*

[0122] The above analyses have aimed at optimizing patterns of activation in the GCL, but have neglected any collateral activation that may have occurred in the NFL. In order to simultaneously optimize GCL target activation and minimize activation of the NFL, target patterns are specified for both the GCL and NFL, corresponding to multiple outputs from the neural network, as shown in Fig. 4. In this case, the NFL target is a vector of zeros, since no activation of the NFL is desired.

[0123] Fig. 12 shows the result of stimulus current optimization for a range of values for the weighting factor,  $\lambda$ , between the GCL and NFL targets. As described in the Methods, a value of  $\lambda$  of 1 corresponds to ignoring activation of the NFL and a value of 0 corresponds to ignoring the GCL. Fig. 12(a-d) shows the results for  $\lambda$  values of 1, 0.75, 0.5, and 0.25, respectively. In agreement with the results presented in Chapter 1 of this thesis, activation of the NFL is minimized by stimulating with multiple electrode in-line with the orientation of passing fibers in the NFL. A consequence of this

stimulation pattern is the elongation of activation in the GCL. Fig. 12(e-h) presents similar results, but where the current delivered to the central electrode is fixed at  $200 \mu A$  and the surrounding electrodes are optimized. This approach appears to perform less favourably, resulting in greater elongation of GCL activation and greater NFL activation.

#### *Optimization of complex stimuli*

[0124] A more relevant test of the optimization regime presented here is its performance with more complicated target patterns, such as text and real-world visual scenes. Fig. 13 shows the performance of the algorithm for each of these cases, where activity in both the GCL and NFL is being optimized. For comparison, the pattern of activation achieved using a naive stimulation strategy is shown in Fig. 13(c) and (d). Fig. 13(e)-(l) demonstrate the achievable patterns of stimulation in the GCL and NFL as the importance of each target is changed by varying  $\lambda$ .

[0125] From inspection of Fig. 13, it is clear that a side-effect of increasing the importance of unwanted activation in the NFL is the reducing of overall stimulus current amplitudes. In this sense, dual optimization of activation in the GCL and the NFL acts in a similar way to constrained optimization, with constraints applied to stimulus magnitudes. To investigate this, Fig. 13 shows the interaction of the GCL/NFL weighting term,  $\lambda$ , and the an inequality constraint applied to the the norm of the stimulus current vector,  $|\vec{S}_t|$ . In Fig 13, the norm constraint varies along the x-axis and  $\lambda$  varies along the y-axis. Each set of two rows shows the optimized activation in the GCL (top) and the NFL (bottom).

#### *Epiretinal versus subretinal stimulation*

[0126] The challenge of avoiding stimulation of the NFL when conducting epiretinal stimulation, as illustrated in Figs. 12, 13, and 14, allows a comparison of epiretinal and subretinal stimulation using the methodology developed here. Due to their greater distance from the NFL, subretinal electrode arrays are less likely to activate passing

fibers in the NFL. This should have the effect of reducing the patterns of unwanted stimulation seen above.

[0127] Fig. 15 shows the achieved activation patterns in the GCL and NFL under subretinal stimulation, using the two target images shown in Fig. 13(a) and (b). For each image, very little activation is observed in the NFL, even when NFL activity is ignored by the optimization (i.e.  $\lambda = 1.0$ ). Further, from inspection of the optimized patterns of activation in the GCL, a similar quality of image reproduction is achieved for subretinal stimulation as was observed for epiretinal stimulation.

#### *Comparison of animal model anatomies*

[0128] Due to the variation in the thickness of the NFL between different species (human: 100  $\mu\text{m}$ , rat: 45  $\mu\text{m}$ ), a similar analysis can be performed to that presented above using a rat-like geometry for the NFL by setting the thickness to 45  $\mu\text{m}$ . Fig. 3.7 shows the results of this analysis for the two target images from Fig. 3.5. Despite the lower NFL thickness of the rat and lower distance between electrodes and GCL, for both target images, similar results were observed for both animal models. However, for the rat geometry, the resolution of the achieved activation patterns seems marginally better.

#### Discussion

[0129] In this disclosure, there is provided a method for the spatial shaping of neural activity in retinal ganglion cells via inversion of the linear-nonlinear network. The presented approach is capable of rapidly optimizing stimulus currents to desired complex images or patterns using a complex biophysical model of the retina. Furthermore, the strategy can be used to fit to multiple target activation patterns for multiple retinal layers, and can be constrained to ensure clinically-acceptable stimulus parameters. The artificial neural network form of the linear-nonlinear model used here is flexible, and is straightforward to generalize to account for complexities such as asymmetric stimulation thresholds and temporally-dynamic ERFs. Furthermore, the

approach developed in this study is not applicable only to linear-nonlinear models of retinal activation, but any application in which linear-nonlinear models are relevant, including cortical stimulation.

#### *Optimality of solutions*

[0130] Inversion of the linear-nonlinear model with biphasic stimulation can be challenging due to the potentially large number of locally-optimal solutions. As a result, there is no guarantee that a gradient-descent algorithm, such as Adam, will find the globally optimal inversion. The inversion strategy achieves solutions with a range of loss values when constrained to localized regions of the solution space. In contrast, when allowed to traverse the entire solution space, the inversion strategy finds an optimal or near-optimal solution for each of the target images tested.

[0131] This result can be explained in two possible ways: 1) the Adam gradient-descent algorithm is effective at searching across the solution space to find the best solution, avoiding getting caught by local optima or 2) the least-squares initialization used in this work is effective at initializing the algorithm at a location in the vicinity of the global optima. Similar results were achieved when the analysis was repeated with random initialization of network weights (despite requiring longer to converge for some targets), indicating that it is the robustness of the Adam algorithm which is responsible for the optimality of the results. The Adam algorithm contains a mechanism for *momentum* which acts to prevent the algorithm from converging prematurely into local minima. Due to the large number of local optima for biphasic stimulation, the objective/optimization surface may have a *bumpiness*. The second order momentum-like mechanism in the Adam algorithm is able to self-tune in a way that enables it to traverse that surface without getting caught in localized troughs (local optima).

[0132] If, in the application of a strategy such as this, low-quality inversions are observed, alternative, non-biphasic stimulation strategies may be considered. For example, the use of charge-balanced tri-phasic stimuli may limit activation to a single stimulus phase, reducing or eliminating the need for a double-sided nonlinearity.

*Constrained optimization*

[0133] Inversion of the linear-nonlinear neural network can be constrained to ensure the clinical safety of optimized stimuli. However, it is also evident from the interaction of  $\lambda$  and norm constraints that constraining solutions by penalizing unwanted NFL activation has a similar, but more marked, effect. The dominance of the effect of  $\lambda$  demonstrates that co-targeting both the GCL and the NFL reduces the need to constrain the currents. Furthermore, unlike a hard constraint on stimulus amplitudes, constraining based on NFL activation is informed by desired patterns of activation, and so is a preferable method of constraining currents. However, since including a term for the NFL does not *guarantee* clinical constraints on current amplitude and power are met, a best solution may be include both approaches, with a fixed current constraint serving predominantly as a backup measure.

[0134] Activation of the NFL could be reduced by simultaneously stimulating with multiple electrodes aligned with the axons of passage in the NFL. In agreement with this result, as activation in the NFL is given more weight in the optimization scheme, multiple electrodes aligned with fibers in the NFL are recruited. Stimulus currents may be optimized such that horizontally adjacent electrodes (aligned with the NFL) receive stimuli of the same polarity, whereas vertically adjacent electrode receive stimuli with alternating polarity.

[0135] The use of constraining terms may also be used to invert an undetermined form of the linear-nonlinear system. When using experimentally-determined ERFs as opposed to ERFs estimated with a computational model, it may be that fewer recording electrodes are used than stimulating electrodes, resulting in fewer sampled locations,  $M$ , than stimulating electrodes,  $N$ . In this situation, the problem of selecting an optimal set of  $N$  electrode currents given a desired pattern of activation across  $M$  locations in the retina is an undetermined problem, yielding infinitely many solutions. A method for solving this problem is to use ridge regression (also known as L2-regularization, Tikhonov regularization, or weight decay). Ridge regression uses an altered objective function in which a term is included that penalizes based on the size of

the weights (current amplitudes in our case). This gives an inversion algorithm a preference for a smaller subset of potential solutions: those that require less current. Following appropriate tuning of the regularization parameter, this allows for the approximate inversion of the undetermined system. For example, the regularization parameter,  $\lambda$ , may be increased gradually from 0 (i.e., no regularization) to some (typically small) positive value at which stimulus currents are maintained below a suitable value.

#### *Application*

[0136] It is noted that in a real-time implementation the target images are likely to be constantly changing. In this respect, a neural network and an iterative optimization algorithm such as Adam are suitable. Neural networks are well suited for online applications such as this, for both training (relevant to spatial shaping) and prediction (not relevant to spatial shaping). A sensible approach to the online application of such as model may be to initially fit the stimulus currents to a static image (e.g. by asking the user of the retinal prostheses to look at a still object), and then allow the stimulus currents to be updated online as the user moves their focus. Following the initial fitting of optimized stimuli, it is expected that, as the target changes, the stimulus currents will be able to track the optima as it moves through the stimulus space. Provided a single iteration of the optimization algorithm is fast enough, this will maintain an effective set of stimuli in response to changing target images. Based on the optimizations run during this analysis, iterative optimization updates can be provided at a rate of greater than 50 Hz. Furthermore, online training of neural networks can potentially be implemented far more efficiently on a field-programmable gate array, allowing for extremely efficient, hardware-based, stimulus current optimization.

[0137] In contrast to this, non-iterative approaches, such as those that involve matrix inversion (e.g. least squares approaches) may not be well-suited to online optimization. Unlike an online approach, which allows for the gradual/continuous changing of stimulus currents, non-online approaches are likely to result in sharp and large changes

in stimulus patterns. This is due to the fact that there is no connection between one solution and the next which may result in confusing percept transitions.

[0138] Some wearers may experience a greater sensitivity to certain electrodes due to their particular retinal anatomy or irregular device placement. In this situation, an approximate sensitivity matrix can be determined from the individual by stimulating with one electrode at a time. This matrix can then be implemented in the linear-nonlinear network as an additional network layer so that model inversion automatically accounts for varying sensitivities across the electrode array.

### Conclusion

[0139] In this disclosure, a method was demonstrated for the spatial shaping of neural activity in retinal ganglion cells via inversion of a generalized form of the linear-nonlinear model. The linear-nonlinear model was first extended to allow for a high-dimensional representation of the retinal tissue (e.g. multiple volumes of multiple cell types/orientations). This model was then translated into an equivalent artificial neural network: the linear-nonlinear network, which is applicable to both experimentally- and computationally-estimated electrical receptive fields. Based on biophysical validation, electrical receptive fields were computed for a population of retinal ganglion cells using a biophysical model of passive activation. This allowed for a high-resolution representation of the retina and target patterns of activation. Using the Adam optimization algorithm, developed for training neural networks, an efficient and optimal method for the inversion of the linear-nonlinear network was demonstrated. The presented approach is capable of rapidly optimizing stimulus currents to desired complex images or patterns using a complex biophysical model of the retina. Furthermore, the strategy can be used to fit to multiple target activation patterns for multiple retinal layers, and can be constrained to ensure clinically-acceptable stimulus parameters. The artificial neural network form of the linear-nonlinear model used here is flexible, and is straightforward to generalize to account for complexities such as asymmetric cathodic/anodic stimulation thresholds and temporally-dynamic ERFs.

[0140] Fig. 17 illustrates a computer system 1700 for neural stimulation, such as an implantable neural stimulation device. System 1700 comprises a processor 1701 connected to a program memory 1702, a data memory 1703, a communication port 1704 connected to a camera 1705 and a stimulation array 1706. The camera may be mounted on glasses worn by the user so that the captured image is similar to what the user would see. The stimulation array 1706 may be a retinal stimulation array as described herein. The program memory 1702 is a non-transitory computer readable medium, such as solid state disk or FLASH-ROM. Software, that is, an executable program stored on program memory 1702 causes the processor 1701 to perform the method in Fig. 2, that is, processor 1701 receives an image from camera 1705, trains a forward stimulation model to invert it and thereby determine stimuli and apply the determined stimuli to the stimulation array 1706. The term “determining a stimulus” refers to calculating a value that is indicative of the stimulus. This also applies to related terms.

[0141] The processor 1071 may then store the stimulus and/or the forward model on data store 1702, such as on RAM or a processor register. Processor 1701 may also send the determined stimulus via communication port 1704 to a stimulation circuit, such as digitally controlled current source.

[0142] The processor 1071 may receive data, such as image data, from data memory 1703 as well as from the communications port 1704. In one example, the processor 1701 receives and processes the image data in real time. This means that the processor 1701 determines the stimuli every time camera 1705 sends a fresh image and completes this calculation before the camera 1705 sends the next image data update.

[0143] Although communications port 1704 is shown as a distinct entity, it is to be understood that any kind of data port may be used to receive data, such as a network connection, a memory interface, a pin of the chip package of processor 1701, or logical ports, such as IP sockets or parameters of functions stored on program memory 1702 and executed by processor 1701. These parameters may be stored on data memory

1703 and may be handled by-value or by-reference, that is, as a pointer, in the source code.

[0144] The processor 1701 may receive data through all these interfaces, which includes memory access of volatile memory, such as cache or RAM, or non-volatile memory, such as an optical disk drive, hard disk drive, storage server or cloud storage. The computer system 1700 may further be implemented within a cloud computing environment, such as a managed group of interconnected servers hosting a dynamic number of virtual machines. In this sense, the implanted device or a device worn by the user, may send the image data to a cloud computing platform and receive the calculated stimuli.

[0145] It is to be understood that any receiving step may be preceded by the processor 1701 determining or computing the data that is later received. For example, the processor 1701 determines the image data, such as by pre-processing or filtering, and stores the image in data memory 1703, such as RAM or a processor register. The processor 1701 then requests the data from the data memory 1703, such as by providing a read signal together with a memory address. The data memory 1703 provides the data as a voltage signal on a physical bit line and the processor 1701 receives the image data via a memory interface.

[0146] It is to be understood that throughout this disclosure unless stated otherwise, nodes, edges, graphs, solutions, variables, networks, parameters, weights and the like refer to data structures, which are physically stored on data memory 1703 or processed by processor 1701. Further, for the sake of brevity when reference is made to particular variable names, such as “stimulus” or “parameter” this is to be understood to refer to values of variables stored as physical data in computer system 1700.

[0147] Further, Fig. 2 is to be understood as a blueprint for the software program and may be implemented step-by-step, such that each step in Fig. 2 is represented by a function in a programming language, such as C++ or Java. The resulting source code is

then compiled and stored as computer executable instructions on program memory 1702.

[0148] It is noted that for most humans performing the method 200 manually, that is, without the help of a computer, would be practically impossible. Therefore, the use of a computer is part of the substance of the invention and allows performing the necessary calculations that would otherwise not be possible due to the large amount of data and the large number of calculations that are involved.

[0149] It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the above-described embodiments, without departing from the broad general scope of the present disclosure. The present embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.

**CLAIMS:**

1. A method for neural stimulation, the method comprising:  
configuring a forward stimulation model to obtain a calculated neural tissue activation pattern from a given stimulus;  
inverting the forward stimulation model for a desired neural tissue activation pattern to obtain a calculated stimulus; and  
applying the calculated stimulus to multiple stimulation electrodes.
2. The method of claim 1, wherein inverting the forward stimulation model comprises obtaining multiple samples of the desired nervous stimulation pattern and optimising the forward stimulation model for each sample.
3. The method of claim 2, wherein optimising the forward stimulation model is performed iteratively for the multiple samples.
4. The method of claim 2 or 3, wherein the forward stimulation model comprises an input, an output and model parameters and the calculated stimulus is represented by the model parameters of the forward stimulation model such that optimising the parameters of the forward stimulation model results in optimising the calculated stimulus.
5. The method of any one of the preceding claims, wherein the neural stimulation comprises neural stimulation of retinal nerves of the visual system.
6. The method of claim 5, wherein the desired nervous stimulation pattern is a desired image represented by the stimulation pattern on a retina.
7. The method of claim 6, wherein inverting the forward stimulation model comprises obtaining multiple sample images of the desired nervous stimulation pattern and optimising the forward stimulation model for each sample image.

8. The method of any one of the preceding claims, wherein the forward stimulation model is an artificial neural network.
9. The method of claim 8, wherein the artificial neural network comprises a linear input, a network layer representing a differentiable nonlinearity and an output representing the neural tissue activation pattern.
10. The method of claim 9, wherein the linear input comprises a linear representation of the neural tissue and a differentiable nonlinearity comprises a representation of nonlinear neural tissue activation.
11. The method of claim 8, 9 or 10 wherein the artificial neural network comprises network parameters that represent a stimulus and inverting the forward stimulation model comprises optimising the network parameters to thereby obtain the calculated stimulus.
12. The method of claim 11, wherein optimising the network parameters comprises applying a constraint to the network parameters to ensure the calculated stimulus remains within clinical safety bounds or device power limits or both.
13. The method of claim 11 or 12, wherein  
the artificial neural network comprises a network layer,  
the network layer comprises a non-overlapping convolutional layer, and  
optimising the network parameters comprises optimising only parameters of  
the non-overlapping convolutional layer.
14. The method of any one of claims 6 to 13, wherein the method is continuously repeated for a sequence of images to be perceived, each image representing a desired neural activation pattern, and the forward stimulation model being inverted for each desired neural activation pattern.

15. The method of any one of the preceding claims, wherein the method comprises inverting the forward stimulation model for a first region of the neural tissue and minimising activation in a second region of the neural tissue.
16. The method of any one of the preceding claims, wherein inverting the forward stimulation model comprises minimizing logarithmic loss.
17. The method of any one of the preceding claims, wherein the forward stimulation model comprises a three-dimensional representation of neural tissue.
18. A neural stimulation device comprising:
  - multiple electrodes to provide stimulation to neural tissue;
  - a processor configured to:
    - configure a forward stimulation model to obtain a calculated neural tissue activation pattern from a given stimulus;
    - invert the forward stimulation model for a desired neural tissue activation pattern to obtain a calculated stimulus; and
    - apply the calculated stimulus to multiple stimulation electrodes.

1/15

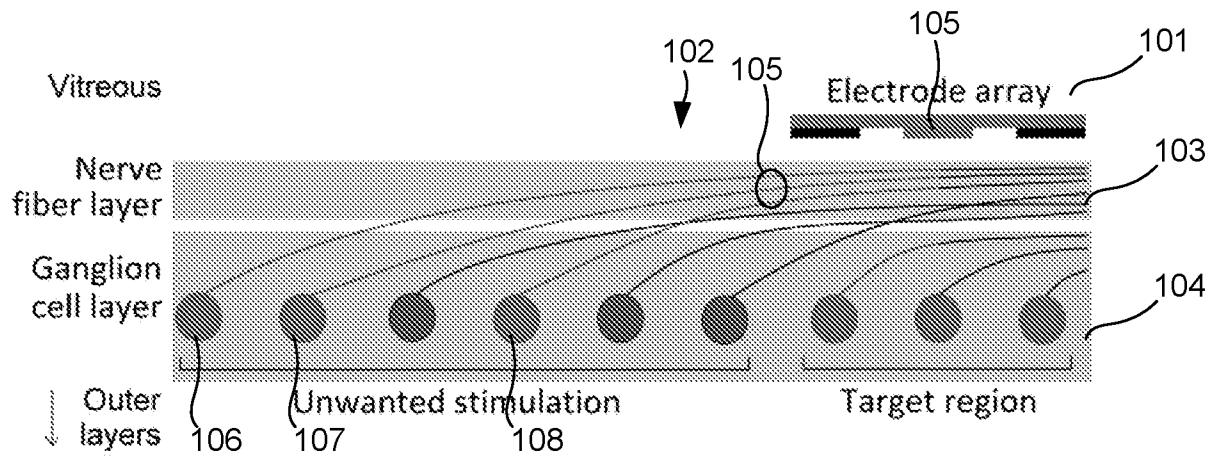


Fig. 1

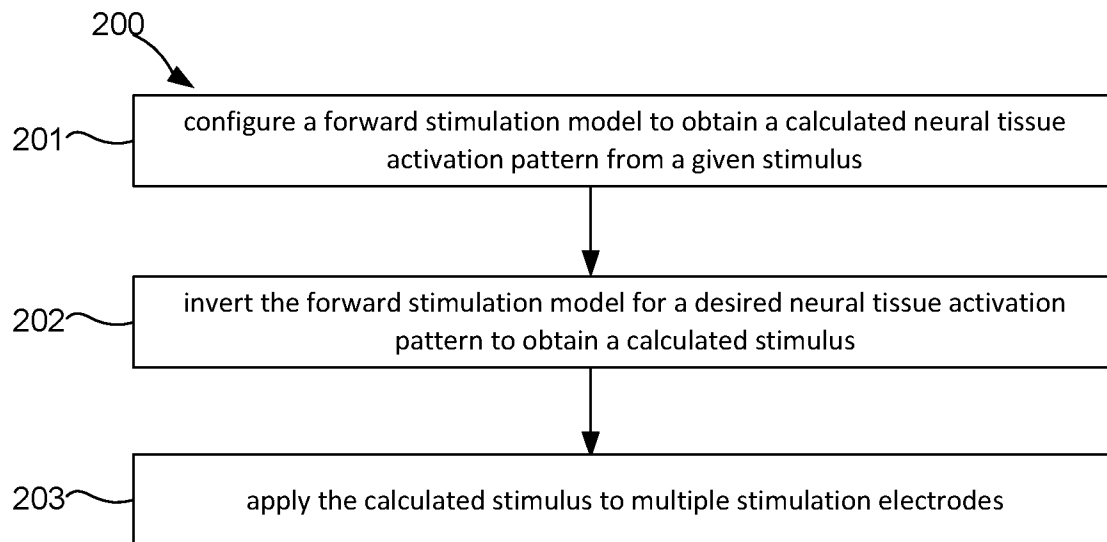


Fig. 2

2/15

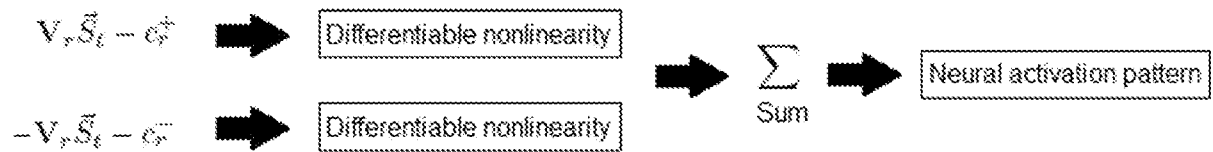


Fig. 3

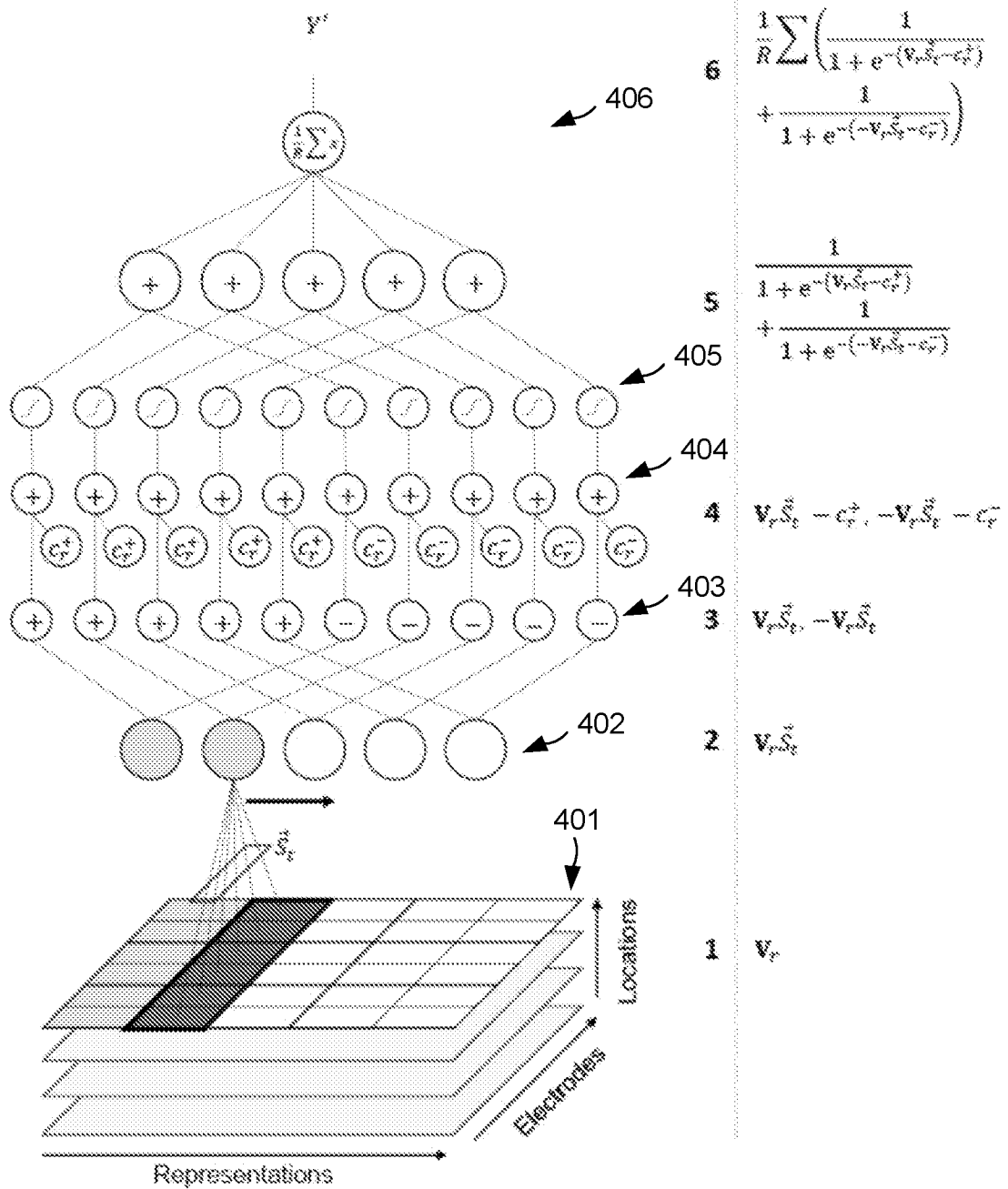


Fig. 4

3/15

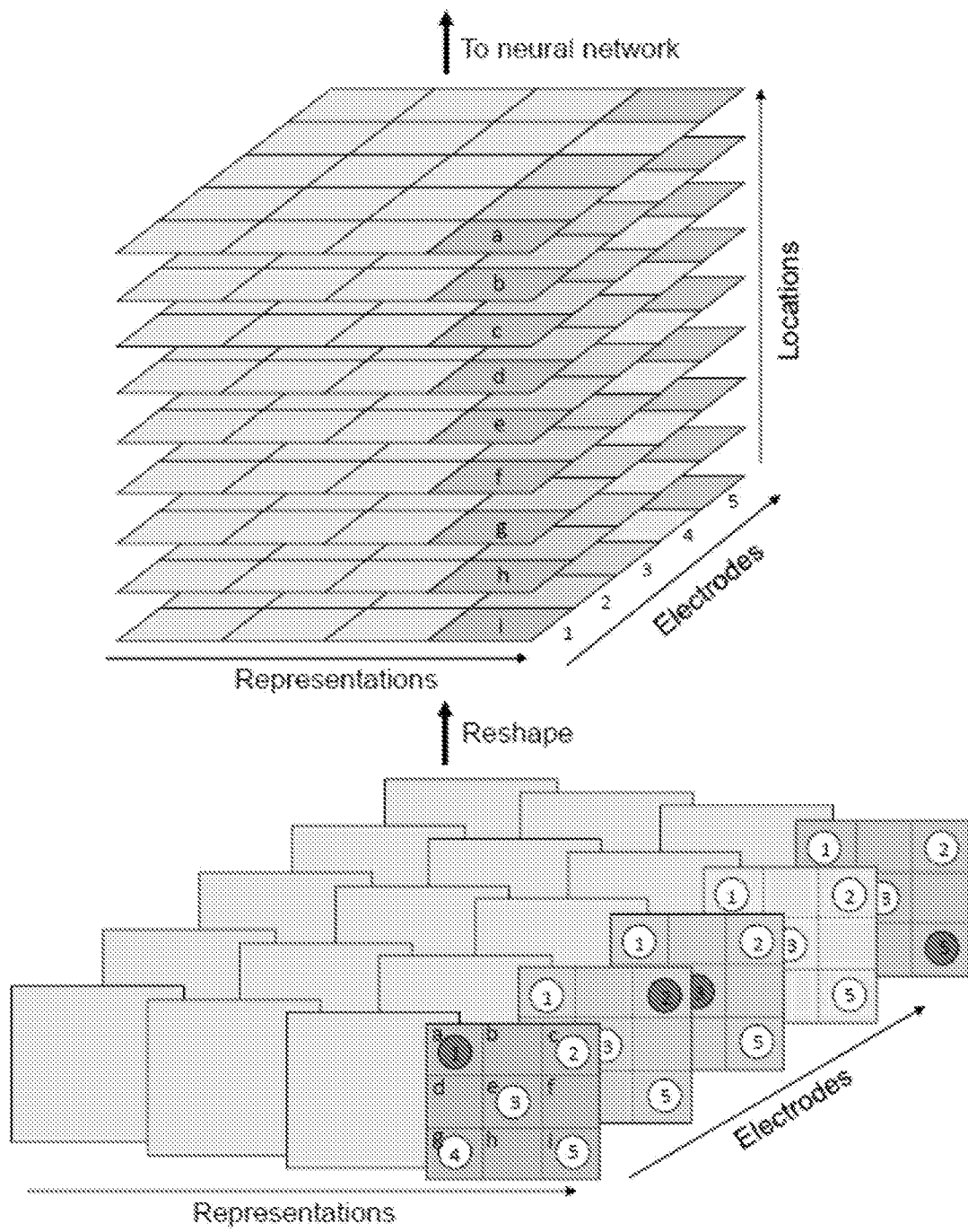


Fig. 5

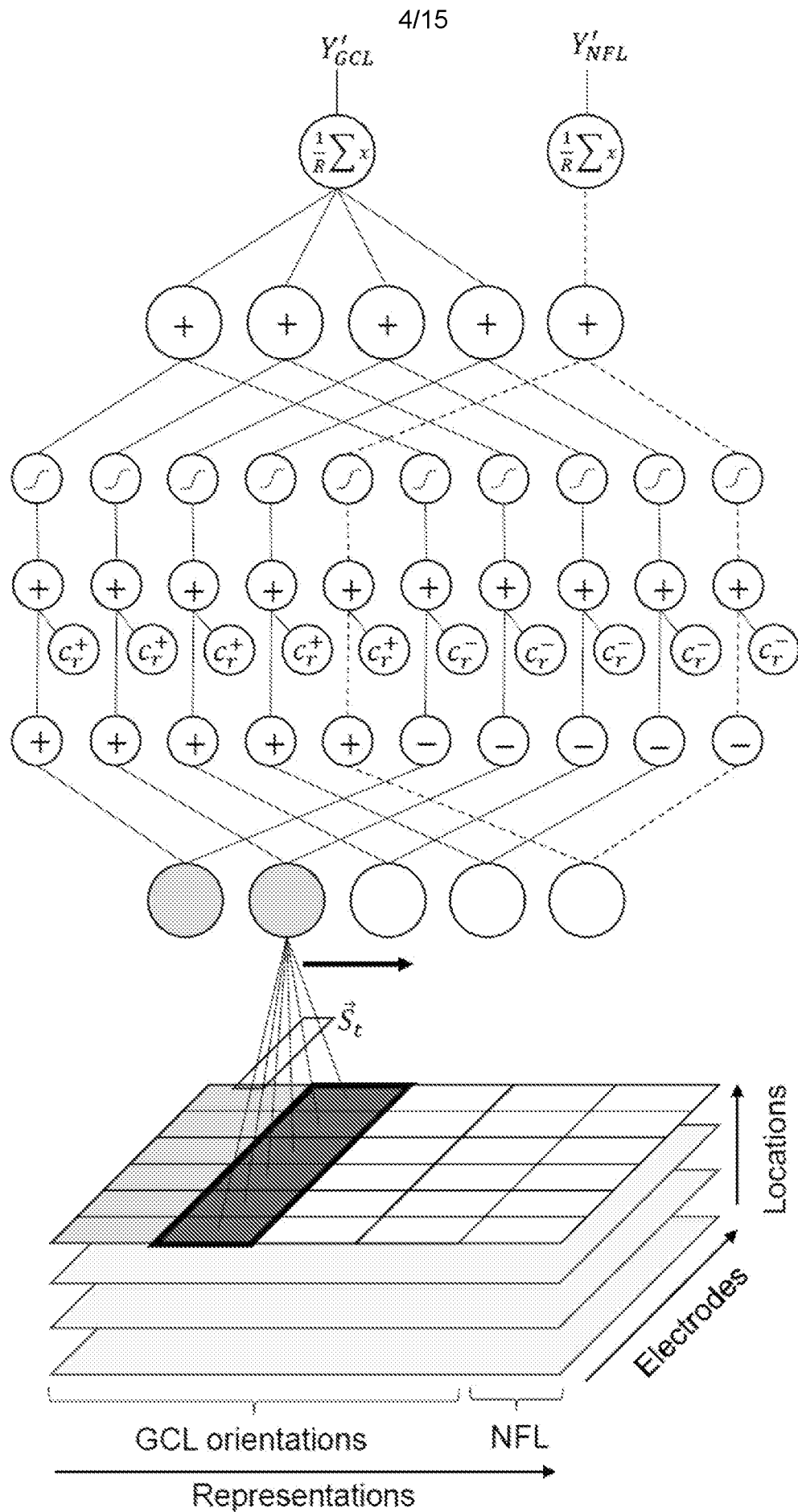


Fig. 6

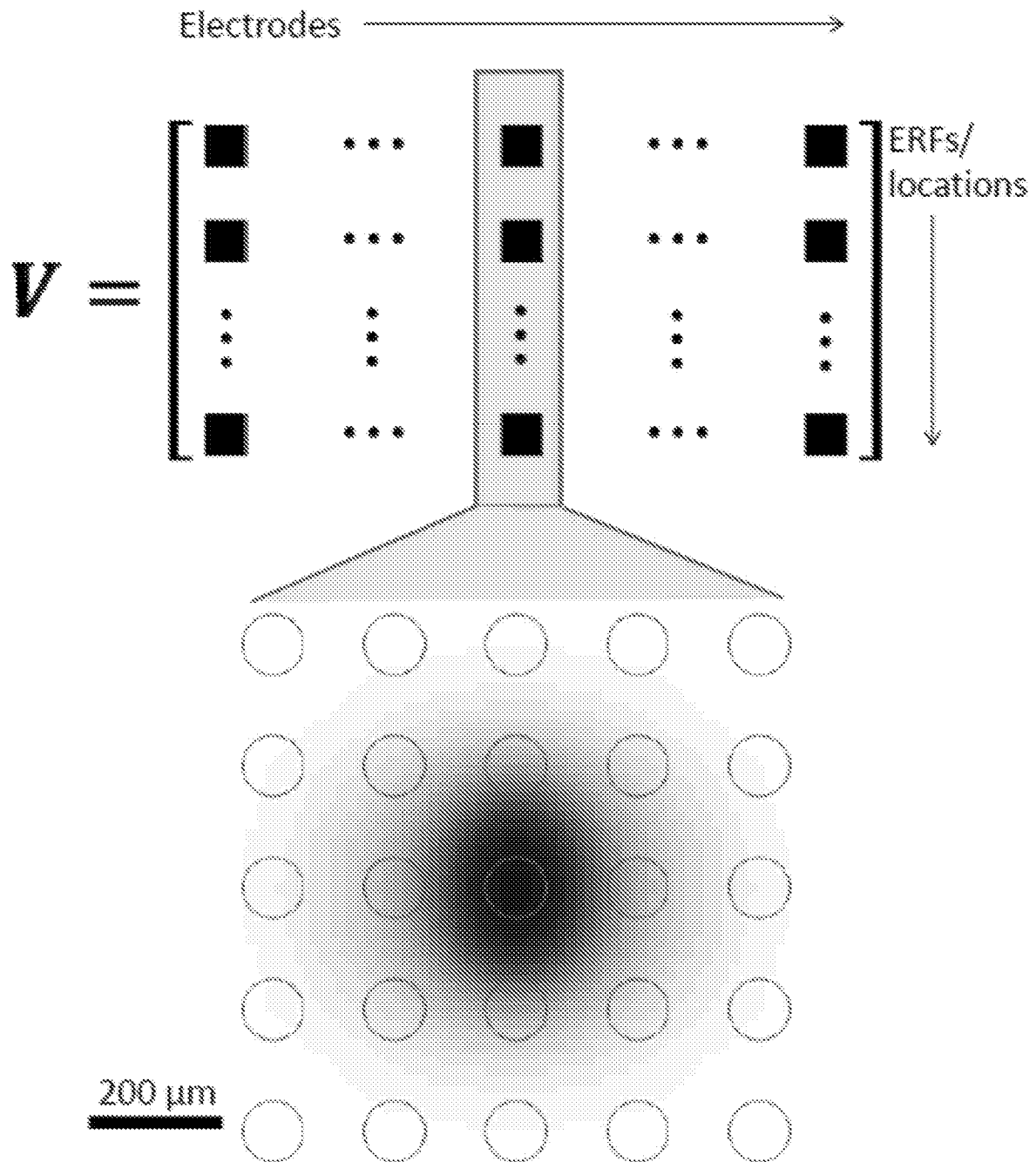


Fig. 7

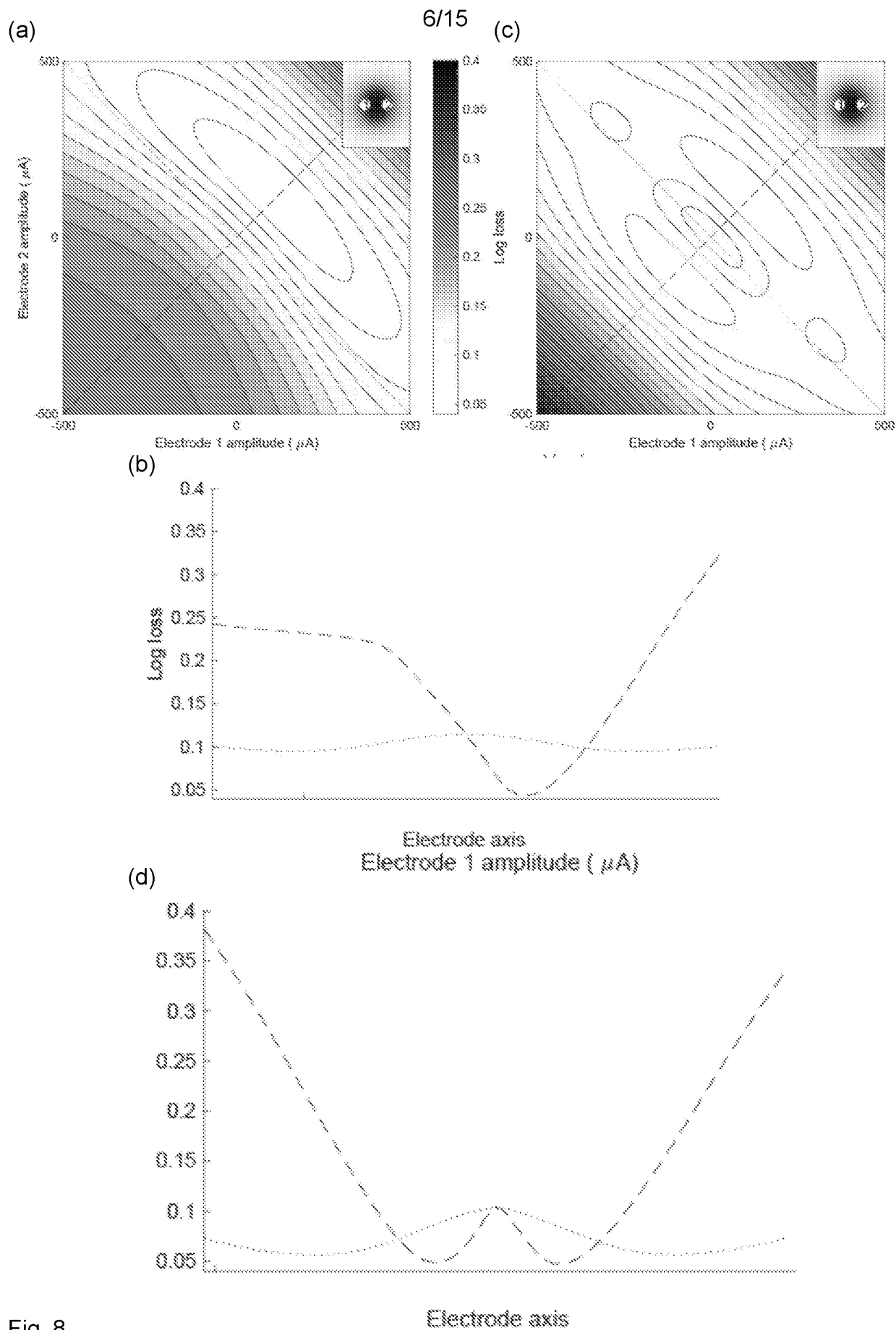


Fig. 8

7/15

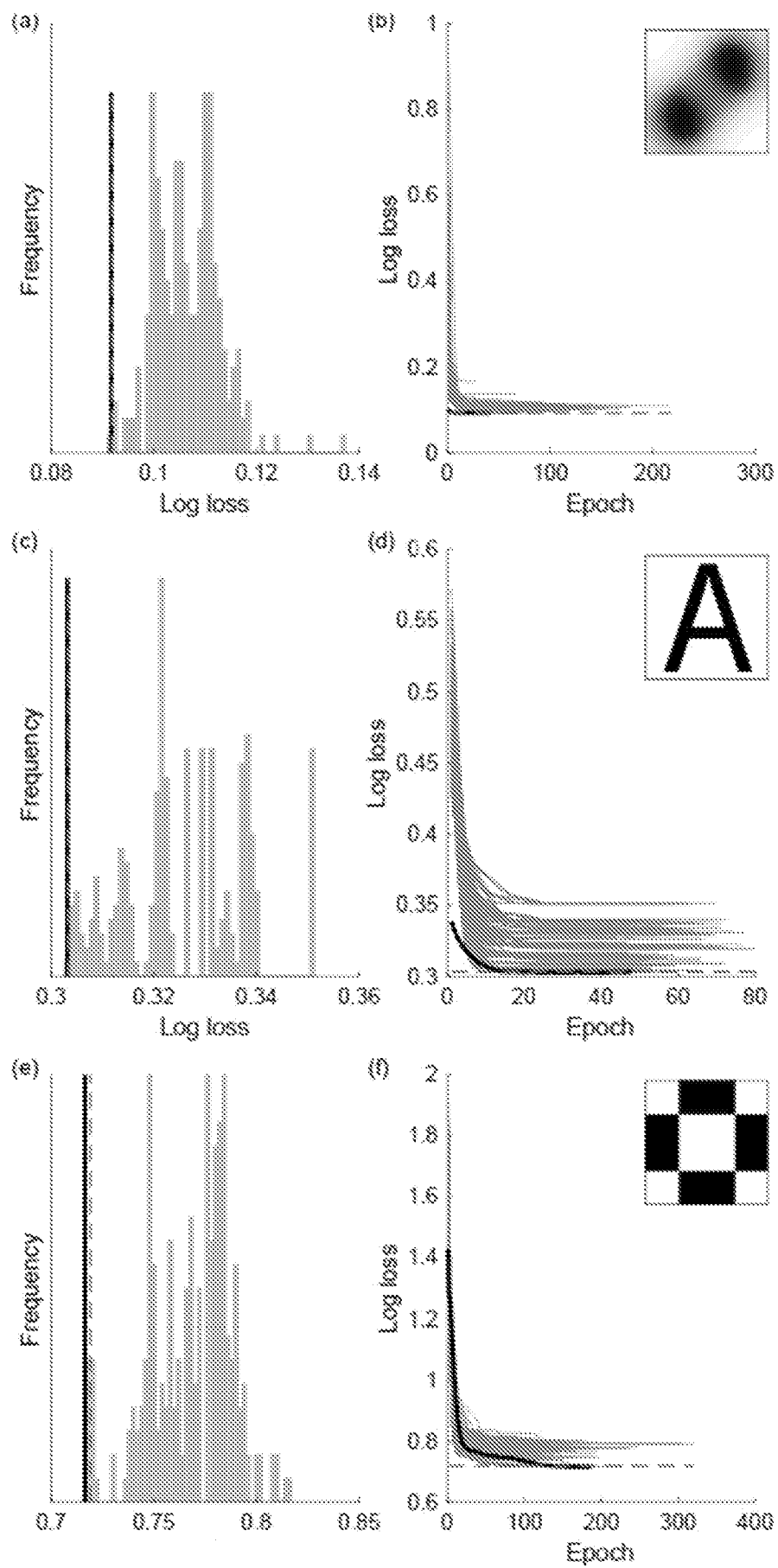


Fig. 9

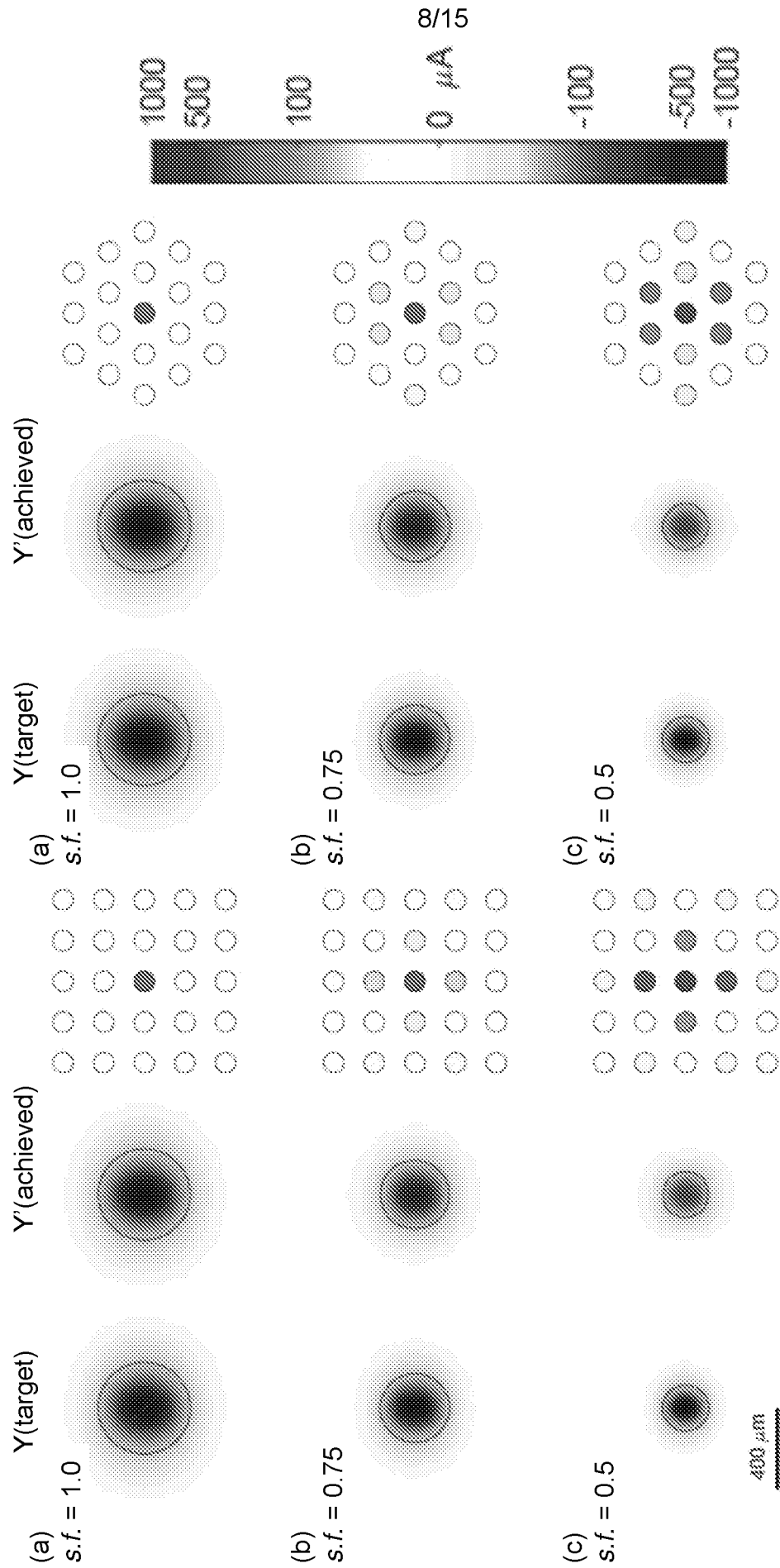


Fig. 10

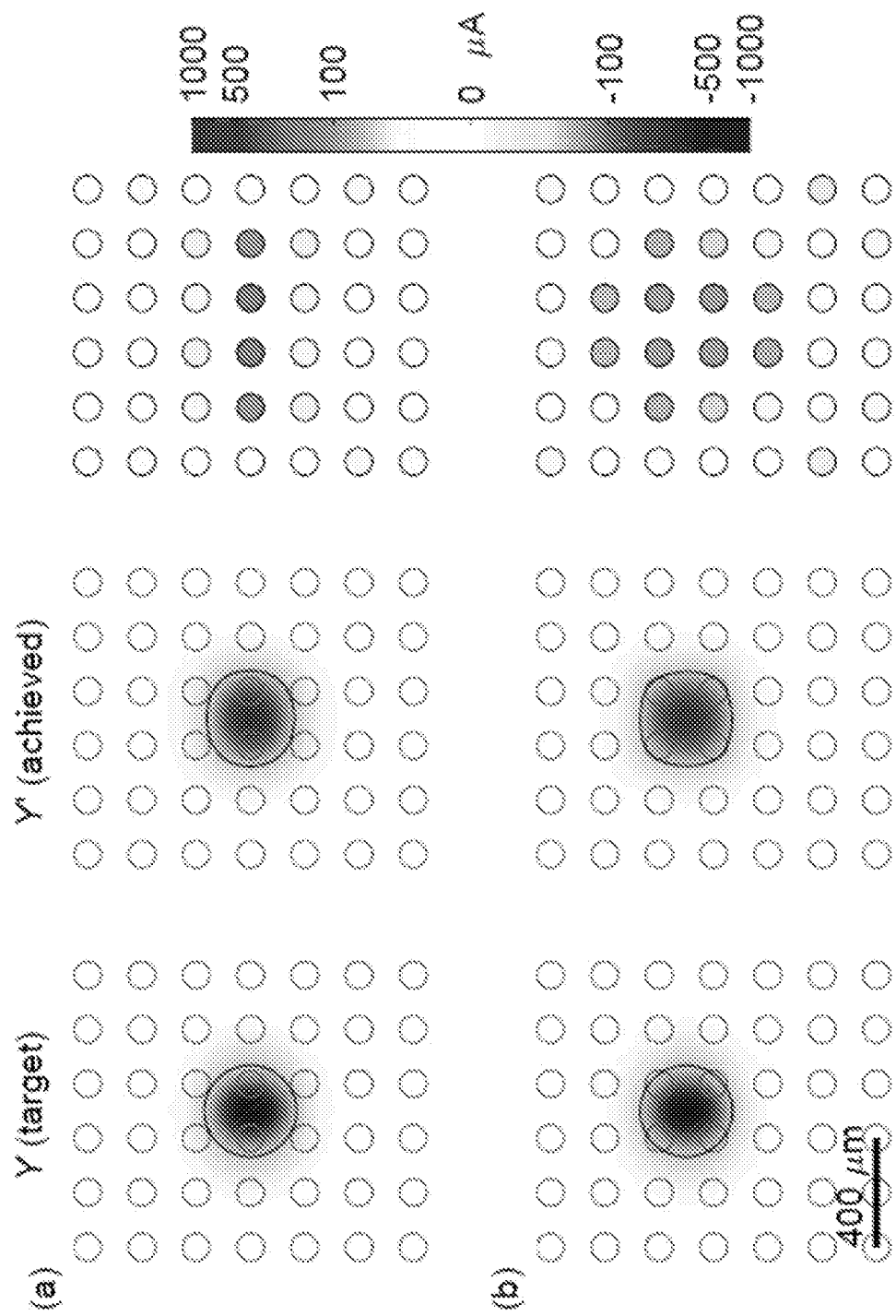


Fig. 11

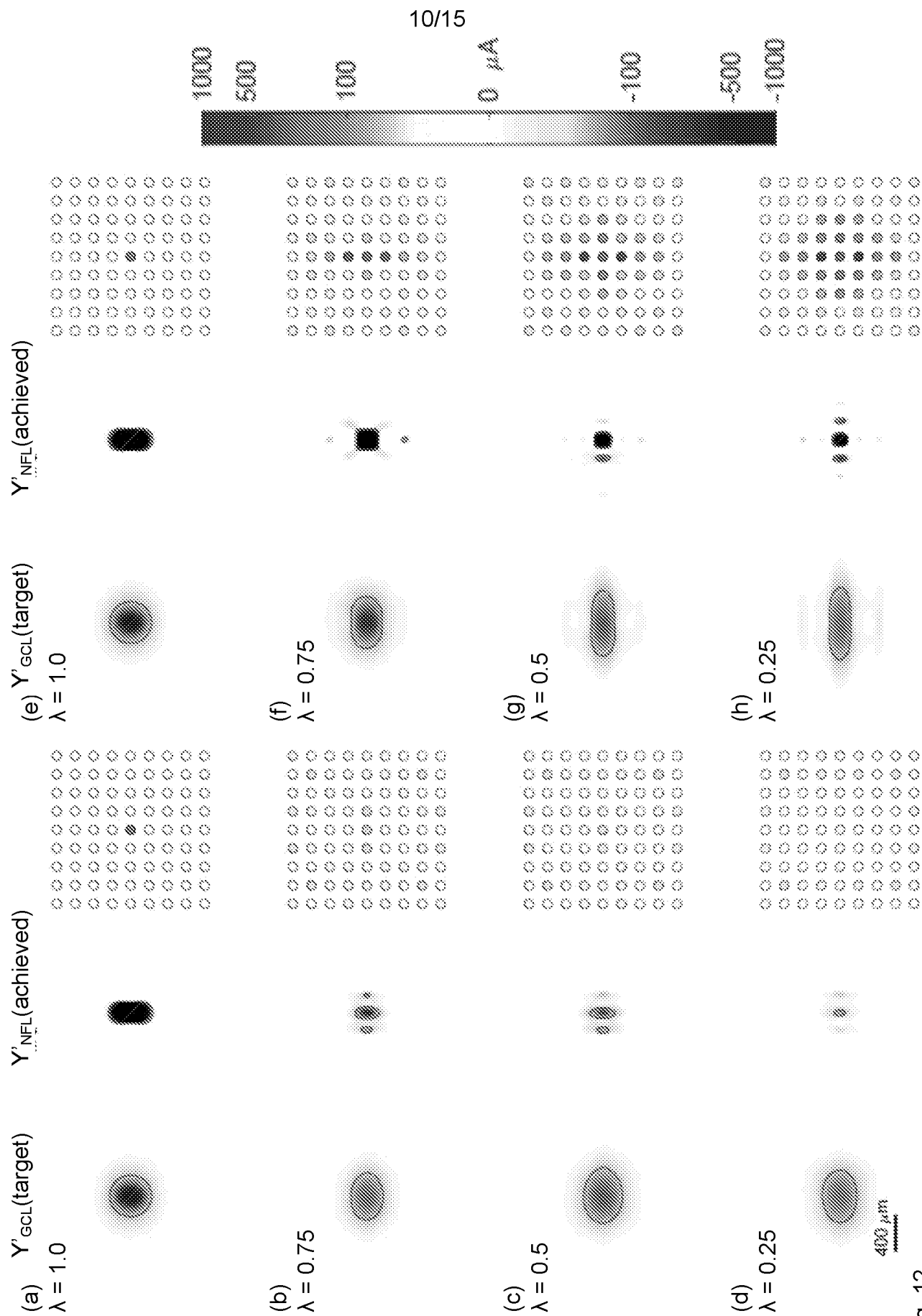


Fig. 12

11/15

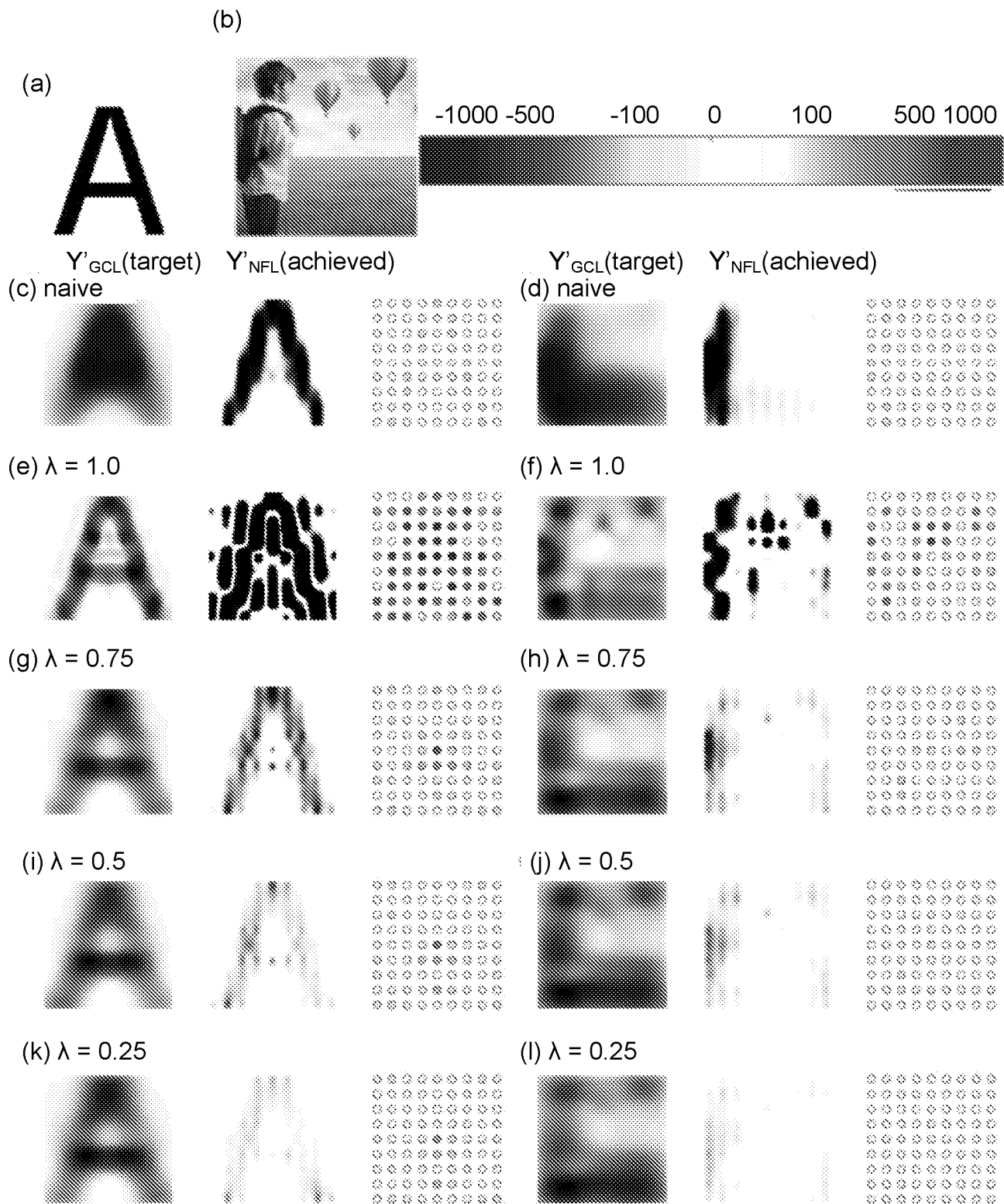


Fig. 13

12/15

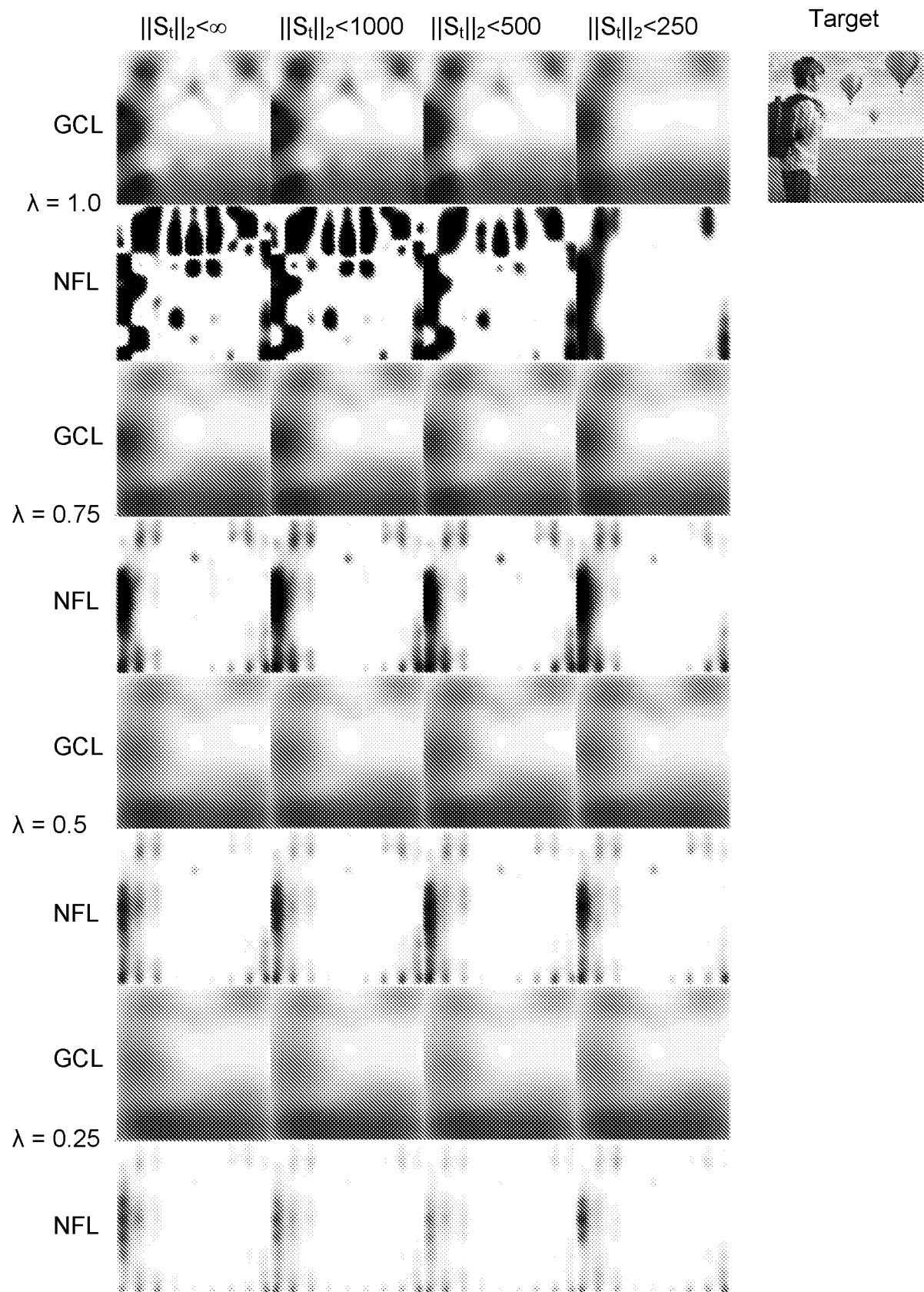


Fig. 14

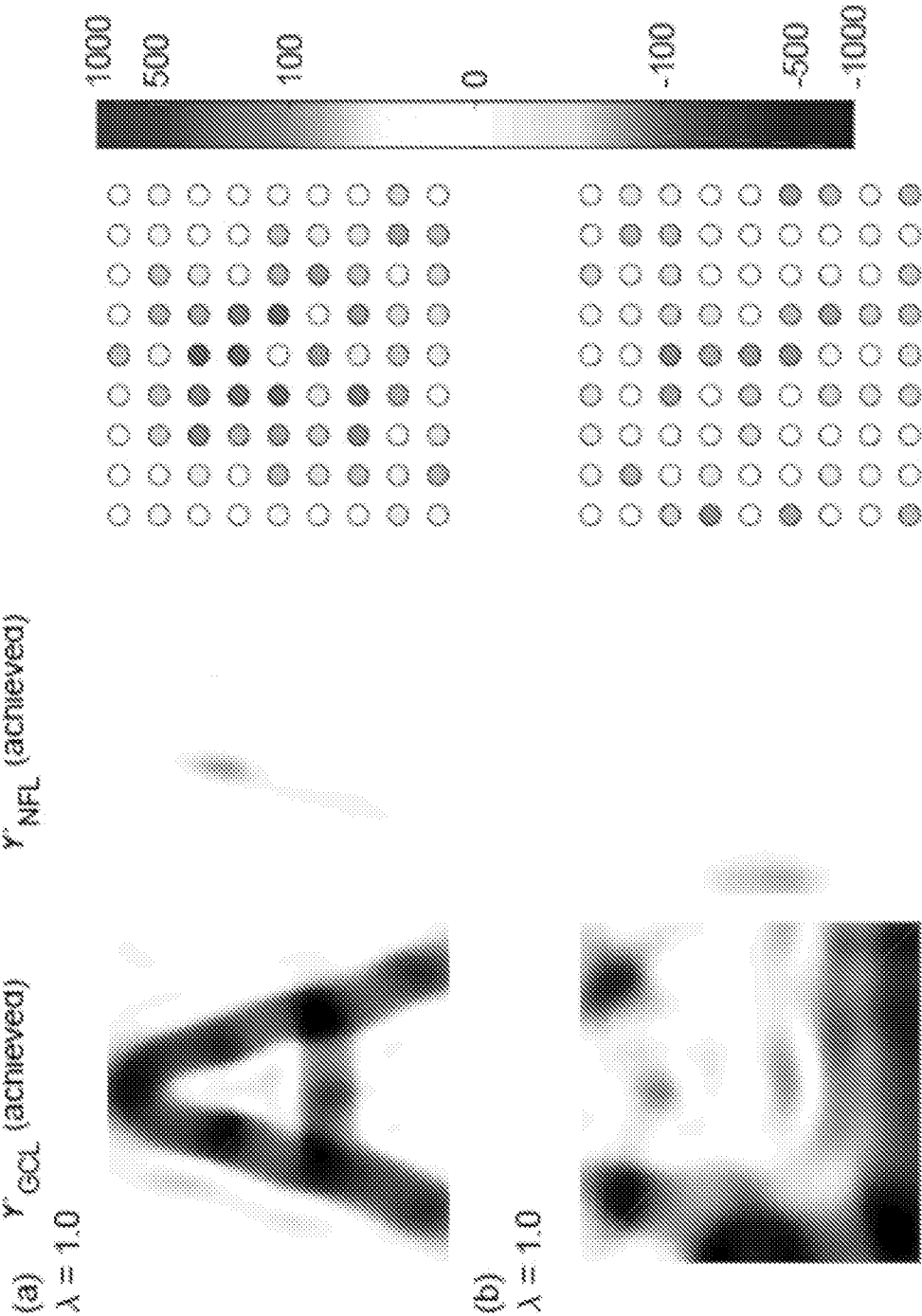


Fig. 15

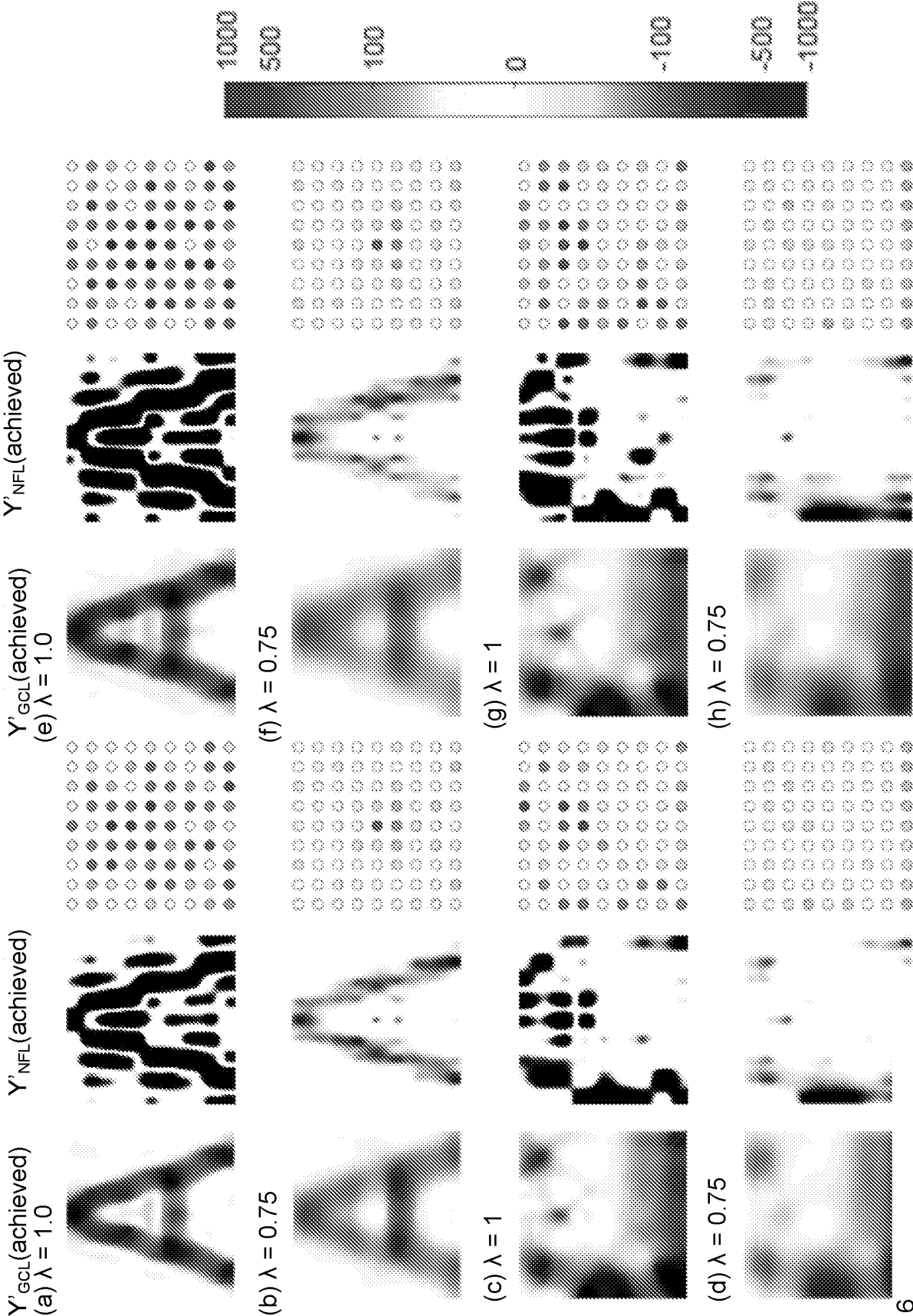


Fig. 16

15/15

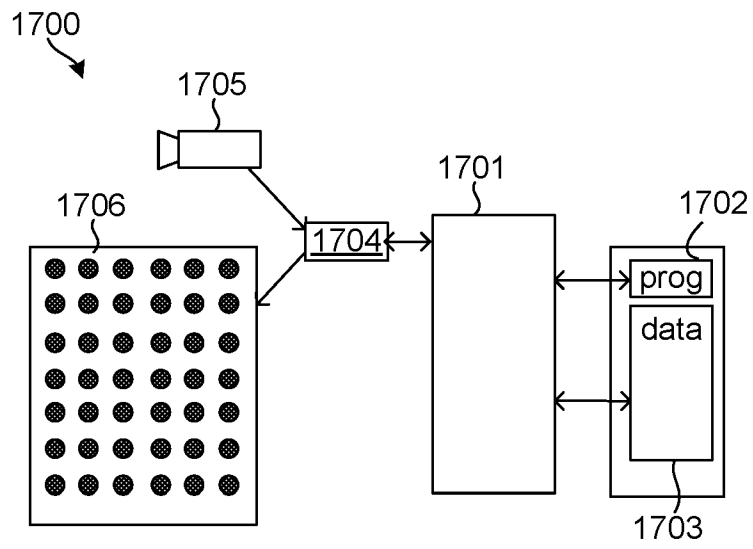


Fig. 17

## A. CLASSIFICATION OF SUBJECT MATTER

**A61N 1/36 (2006.01) A61N 1/05 (2006.01) A61F 9/08 (2006.01) G06N 3/00 (2006.01)**

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Database PATENW (EPOQUE search) and IPC /CPCS marks such as: A61N1/36, A61F9/08, A61N1/36046, A61N1/0543, G06N3/049, G06N3/008, G06N3/08, G06N3/084, G06K9/605, G06K9/0061, G06T, G06F17, G06F19/10, G06F19/20 and lower classification level IPC or CPC marks and the like. Keywords(kwds): retina, nerves, neural model, neural networks, multiple, sample image, build model, forward, model, CNN, Artificial networks, deep neural, pattern, convolution layers, linear input, non-linear model, inverted, inversion, inverse problem, optimising, optimizing, constrains, parameters, multiple electrode, stimuli, tissue activation, electrode currents and the like (in various Boolean & proximity operators). Inventors' & Applicant's names searched in PATENW. Databases NPL, MEDLINE, INSPEC (kwds): ganglion cell layer, retinal cell layer, retinal prosthesis, virtual retinal, inverse model, inverse decomposition, deep network, artificial networks and the like (in various Boolean & proximity operators). Google Patents (kwds): stimulation model, retinal stimulation optimised, iterative, inversion, calculated stimulus & the like combined with Inventor's name; Google Scholar (kwds): stochastic gradient descent, retinal stimulation model, ADAM, forward retinal stimulation model, inverse electrode parameters and the like. ESPACENET: Inventor's & Applicant's names searched in combination with keyword: electrode/neural and the like. Applicant(s)/Inventor(s) name searched in internal databases provided by IP Australia.

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|-----------|------------------------------------------------------------------------------------|-----------------------|
|           | Documents are listed in the continuation of Box C                                  |                       |



Further documents are listed in the continuation of Box C



See patent family annex

|                                                                                                                                                                         |                                                                                                                                                                                                                                                  |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| * Special categories of cited documents:                                                                                                                                |                                                                                                                                                                                                                                                  |
| "A" document defining the general state of the art which is not considered to be of particular relevance                                                                | "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention                                              |
| "D" document cited by the applicant in the international application                                                                                                    | "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone                                                                     |
| "E" earlier application or patent but published on or after the international filing date                                                                               | "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | "&" document member of the same patent family                                                                                                                                                                                                    |
| "O" document referring to an oral disclosure, use, exhibition or other means                                                                                            |                                                                                                                                                                                                                                                  |
| "P" document published prior to the international filing date but later than the priority date claimed                                                                  |                                                                                                                                                                                                                                                  |

|                                                                                                                                                                |                                                                                                                                                   |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| Date of the actual completion of the international search<br>4 June 2020                                                                                       | Date of mailing of the international search report<br>04 June 2020                                                                                |
| <b>Name and mailing address of the ISA/AU</b><br><br>AUSTRALIAN PATENT OFFICE<br>PO BOX 200, WODEN ACT 2606, AUSTRALIA<br>Email address: pct@ipaustalia.gov.au | <b>Authorised officer</b><br><br>Viara Van Raad<br>AUSTRALIAN PATENT OFFICE<br>(ISO 9001 Quality Certified Service)<br>Telephone No. +61262832676 |

| INTERNATIONAL SEARCH REPORT                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | International application No. |
|-------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------|
| C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | PCT/AU2020/050287             |
| Category*                                             | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | Relevant to claim No.         |
| X                                                     | ALFARO, M. et al. "Wireless Trans-corneal Stimulus for the Optical Nerve Based on Adaptive Modeling using Continuous Neural Networks," 7th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE 2010), 2010, pages 236 -241<br>Abstract, p. 236-237, Part I, paragraphs 5-13; Figs. 1-3 and 5, esp. p. 237, Part III, 1st and 2nd paragraphs; Part III. sub-parts "Image acquisition and analysis," "Circuit 1 ON Response," "Data processing (CNN)," p. 238, col. 1, ln. 12-19, p. 238 ln. 26-34; p. 238, col. 1, ln. 29 - p. 240, col. 2, ln. 4; Part II, paragraphs 3-5, Part IV, paragraphs 1 to 5 p. 239-240, see: Part VI."Conclusion" | 1-15, 18                      |
| X                                                     | HUTH, J. et al. "Convis: A Toolbox to Fit and Simulate Filter-Based Models of Early Visual Processing." Frontiers in Neuroinformatics, 2018, Vol. 12, Article 9, pages 1-16 [online] [retrieved from internet on 25 May 2020]<URL <a href="https://www.frontiersin.org/articles/10.3389/fninf.2018.00009/full">https://www.frontiersin.org/articles/10.3389/fninf.2018.00009/full</a> ><br><a href="https://doi.org/10.3389/fninf.2018.00009">https://doi.org/10.3389/fninf.2018.00009</a><br>Abstract, part 1 sub-parts 1.1.1-1.1.3 and 1.2; part 2.3; sub-part 2.3.3. Fig. 4; part 3. sub-parts 3. 1-3.3.3; sub-part 3.5 & Figs. 6-8; part 4 and sub-part 3.1                          | 1-7, 14, 16-18                |
| X                                                     | US 2015/0352364 A1 (NATIONAL ICT AUSTRALIA LIMITED et al.) 10 December 2015<br>Abstract, Figs. 1, 2 and 4, para[0188-0203], para[0005-0023], esp. para[0022], para[0031-0038], para[0048-0132], para[0158-0187], para[0205-0208], Claim 20, elements (112), (120), (230), (440)                                                                                                                                                                                                                                                                                                                                                                                                          | 1-7, 14, 18                   |
| A                                                     | ESLER, T. et al. "Minimizing activation of overlying axons with epiretinal stimulation: The role of fiber orientation and electrode configuration," PLOS ONE, 2018, pages 1-27. [online] [retrieved from internet on 25 May 2020]<br><URL <a href="https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0193598">https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0193598</a> ><br><a href="https://doi.org/10.1371/journal.pone.0193598">https://doi.org/10.1371/journal.pone.0193598</a><br>Entire document                                                                                                                                              | 1-18                          |
| A                                                     | KAWATO, M. et al. "A forward -inverse optics model of reciprocal connections between visual cortical areas," Network, 1993, Vol. 4, pages 415 -422<br>Entire Document                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 1-18                          |

|                                                                                                                                                                                                                                                                       |                         |                                                           |                         |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------|-----------------------------------------------------------|-------------------------|
| <b>INTERNATIONAL SEARCH REPORT</b><br>Information on patent family members                                                                                                                                                                                            |                         | International application No.<br><b>PCT/AU2020/050287</b> |                         |
| This Annex lists known patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information. |                         |                                                           |                         |
| <b>Patent Document/s Cited in Search Report</b>                                                                                                                                                                                                                       |                         | <b>Patent Family Member/s</b>                             |                         |
| <b>Publication Number</b>                                                                                                                                                                                                                                             | <b>Publication Date</b> | <b>Publication Number</b>                                 | <b>Publication Date</b> |
| US 2015/0352364 A1                                                                                                                                                                                                                                                    | 10 December 2015        | US 2015352364 A1                                          | 10 Dec 2015             |
|                                                                                                                                                                                                                                                                       |                         | US 9517345 B2                                             | 13 Dec 2016             |
|                                                                                                                                                                                                                                                                       |                         | EP 2950880 A1                                             | 09 Dec 2015             |
|                                                                                                                                                                                                                                                                       |                         | WO 2014117208 A1                                          | 07 Aug 2014             |
| <b>End of Annex</b>                                                                                                                                                                                                                                                   |                         |                                                           |                         |
| Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001.<br>Form PCT/ISA/210 (Family Annex)(July 2019)                                                                                               |                         |                                                           |                         |