

An Electric Field Model for Prediction of Somatosensory (S1) Cortical Field Potentials Induced by Ventral Posterior Lateral (VPL) Thalamic Microstimulation

John Stephen Choi* Marcello Michael DiStasio* Austin J. Brockmeier and Joseph Thachil Francis

Abstract—Microstimulation (MiSt) is used experimentally and clinically to activate localized populations of neural elements. However, it is difficult to predict—and subsequently control—neural responses to simultaneous current injection through multiple electrodes in an array. This is due to the unknown locations of neuronal elements in the extracellular medium that are excited by the superposition of multiple parallel current sources. We therefore propose a model that maps the computed electric field in the three-dimensional space surrounding the stimulating electrodes in one brain region to the local field potential (LFP) fluctuations evoked in a downstream region. Our model is trained with the recorded LFP waveforms in the primary somatosensory cortex (S1) resulting from MiSt applied in multiple electrode configurations in the ventral posterolateral nucleus (VPL) of the quiet awake rat. We then predict the cortical responses to MiSt in “novel” electrode configurations, a result that suggests that this technique could aid in the design of spatially optimized MiSt patterns through a multi-electrode array.

I. INTRODUCTION

Microstimulation (MiSt) is a technique used in the functional analysis of neural activity, and shares its biophysical basis with clinical methods of deep brain stimulation (DBS). Experimentally, the application of MiSt has been employed to demonstrate the causal role of locally defined neuronal populations in the production of behaviors and conscious perceptions [13] [17]. In the context of neuroprosthetics research MiSt provides a means by which information can be delivered into

This work was supported in part by the Joint Graduate Program in Biomedical Engineering at SUNY Downstate/NYU Polytechnic and DARPA REPAIR project N66001-10-C-2008.

J. S. Choi, M. M. DiStasio and J. T. Francis are with the Department of Physiology and Pharmacology, SUNY Downstate Medical Center, Brooklyn, NY 11203 USA. {}

A. J. Brockmeier is with the Department of Electrical and Computer Engineering, University of Florida, P.O. Box 116130 NEB 486, Bldg 33, University of Florida, Gainesville, FL 32611 USA.

*These authors contributed equally to this work.

the central nervous system (CNS) [16] [15]. Designing MiSt patterns for rich sensory feedback, however, is a difficult problem. Often spatial resolution is low, and each stimulating electrode affects many cells. There are also limits to current amplitudes, beyond which electrode corrosion and tissue damage occur. To maximize MiSt’s utility under these constraints, an accurate multi-electrode model of spatiotemporal patterns of MiSt current input and evoked neural activity is needed.

Computational modeling studies of deep-brain stimulation [11] indicate that electrical stimulation activates projecting axons while inhibiting the activity of somata. This suggests that the effects of MiSt should be characterized by induced activity in downstream areas. For the purposes of MiSt, these are areas contacted by efferent (or afferent) axons from the stimulation location. In this report we focus on the somatosensory system, where feedback from sensors on a brain-controlled device could be delivered directly to the user via MiSt as haptic and tactile feedback. We aim to influence primary somatosensory cortex using the VPL thalamus as our locus of control. As natural drivers of cortical activity, thalamic relays appear to be promising candidates for prosthetic input to cortex. Primary sensory cortex circuits are influenced by thalamic relay cells via a recurrent circuit that allows the cortex to modulate or gate thalamic activity [7]. It may therefore be easier to achieve exogenous control of cortical activity by driving thalamocortical inputs with MiSt, rather than by directly stimulating the cortex itself. In this paper we present a model that locates the most sensitive regions (to current density amplitude) around a MiSt electrode array in thalamus, as measured by ability to modulate cortical LFP.

A. Background for model

The amount of activation of neural tissue by microstimulation is dictated by the physical extent of current spread and by the electrical excitability of the elements in

that volume. The excitability properties of many neural elements have been described, often by determination of chronaxie for neurons or their compartments (for reviews see [19], [14]). The relationship between many microstimulation parameters and patterns of neural activation have been investigated, including pulse duration [2], current polarity relative to stimulated elements [14], and inter-pulse intervals (both fixed [12] and variable [8]). In some studies, behaviors elicited (e.g. saccades [12]) have been used in conjunction with knowledge of cortical functional anatomy to estimate the spread of neural activation caused by a microstimulation input. Furthermore, extensive computational modeling of neural responses to electric stimulation has been performed, using techniques like finite element modeling (FEM) to account for electrode geometry and tissue anisotropy [5]. Compartmental neuron models have also been employed to describe complex excitable membrane responses to electrical activation [10]. Such studies demonstrate the improvement in quantitative conclusions that can be drawn about MiSt-induced neuromodulation by taking electric field effects into account. This study aims to extend some of these conclusions in an intact animal (rat) model. In light of evidence that thalamic relay cells function in different modes during sleep and waking behavior [18] we describe the effects of thalamic MiSt on cortical LFP in the quiet but awake state.

B. Simultaneously stimulating electrodes and motivation for study

Passing current through multiple electrodes in parallel produces complex field patterns resulting from the vector-additive interactions among current sources (see Fig. 1). We refer to this technique as “simultaneously stimulating electrodes” (SSE). For example, in bipolar stimulation configurations two electrodes are simultaneously sourcing and sinking equal amounts of current. In this case, the neural response is not simply a linear combination of responses to the same two currents delivered individually as monopoles referenced to ground [1]. A tripolar arrangement involves three electrodes where one electrode sinks the sum of the currents sourced by the remaining two. The most general case is when parallel electrodes are independently sourcing or sinking current, with a common distant reference/ground as a return path.

It is often difficult and prohibitively time-consuming to test all possible configurations of SSEs to find one that produces a desirable physiological activation, so a more guided approach is required. The relevant design choices are the subset of electrodes to use and the amplitude of current to be sinked/sourced from each. Clearly there is an infinite number of (configuration, amplitude) combinations, which is probably why much attention has been devoted to designing array geometries and constrained SSE configurations (e.g. tripolar, hexagonal return) that produce fields with desirable characteristics, such as spatial specificity. A model that is based on features of the imparted electric field, regardless of the SSE configuration that produced it, is necessary to capture more complex geometries. Fortunately, the electrode contact locations in a multi-electrode array are known to a certain degree, and hence the electric field information, despite being subject to theoretical simplifications, is available.

Use of SSE patterns confers improved control over neural activity versus monopolar stimulation. Bierer et al. have demonstrated that using tripolar configurations in cochlear implants produces more focal auditory cortex activation patterns than seen with bipolar or monopolar stimulations [3]. In computer simulated myelinated fibers of varying sizes around DBS electrodes, bipolar stimulation produced different, more complex (versus monopolar) activation zones [10]. It has also been shown in simulation that a hexagonal configuration (See Fig. 1) designating one current source electrode and six surrounding return electrodes decreases the amount of current leakage to surrounding areas considerably [9], indicating an increase in specificity. Thus, SSEs allow finer control over a region of neural tissue than serially

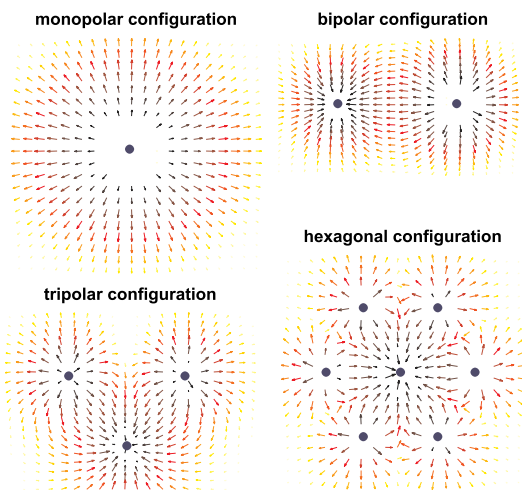


Fig. 1. Two or more simultaneously-stimulating electrodes (SSE) are capable of producing a wide range of field patterns that monopolar configurations acting individually cannot. The static current density field lines resulting from bipolar, tripolar, and hexagonal arrangements reveal the vector-additive interactions among individual stimulating monopoles.

inputting monopolar stimulations.

C. Application to MiSt of somatosensory thalamocortical afferents

We propose a parametric model that predicts downstream neural response strength as a function of electric field. This electric field could be produced by any number of stimulating electrodes in any configuration. For the purposes of this report only monopolar and bipolar configurations were used, but were applied on varying combinations of electrodes. The electric fields produced by various SSE configurations in VPL serves as the input, and the resulting S1 LFP fluctuations serve as the output. The model makes these response predictions based on a weighted summation of electric field values calculated around the stimulating array. Data from multiple stimulation configurations are used in learning this mapping. This procedure also identifies where (in 3D space) the induced electric field has the largest influence on cortical response power. Though we present here the results from MiSt in the VPL, ongoing work with MiSt in the S1 cortex and the dorsal column nuclei (DCN) will help us determine which region is the best target for somatosensory neuroprostheses.

II. METHODS

A. Static modeling problem

The general problem we address is how to construct a model that maps array microstimulation in one brain region to the resulting responses in a downstream one. A good model should accurately predict responses for arbitrary simultaneous stimulation electrode (SSE) configurations. It should do so using data from only a small number of such configurations, i.e., we would like to obviate the need for exhaustive SSE sampling. For simplicity, let us assume that the responses on multiple recording channels are conditionally independent given the stimulus. We can then train a separate multiple input single output (MISO) model for each recording channel.

For now, we focus our attention on *static* modeling, i.e., the “responses” that the models predict are time-independent measures of magnitude. By collapsing the full dynamic response into a single value, we can greatly simplify the learning problem.

Introducing our notation, we have N_s stimulating electrodes and N_r downstream recording electrodes. We record local field potential (LFP) waveforms from these N_r channels while stimulating in a variety of configurations and current amplitudes on the N_s electrodes. Let us denote the recorded LFP signal for one channel as $x(t), t = 1, 2, \dots$

For each stimulation pulse, we know the current waveforms through the stimulating electrodes, the (approximate) locations of the electrode tips, and the recorded responses in downstream electrodes. We stimulated with symmetric charge balanced biphasic pulses (width = $200\mu s$) delivered in monopolar and bipolar configurations. Since these are stereotyped waveforms, we assign a single current value I to each stimulating electrode that is positive when the current waveform through it is negative-first. For a single (configuration, amplitude) variation, let us define $\mathbf{I} = (I_1, \dots, I_{N_s})$ as the multichannel currents through all the electrodes. Each electrode is assigned one current value according to the sign convention above. For m unique stimulus variations, we denote our input data as $\{\mathbf{I}^{(1)}, \dots, \mathbf{I}^{(m)} \in \mathbb{R}^{N_s}\}$. Throughout the recording session, each variation is delivered N_T times so the response waveforms can be averaged.

We denote the post-stimulus response strengths as $\{y^{(1)}, \dots, y^{(m)} \in \mathbb{R}\}$, where $y^{(i)}$ is given by the RMS power of the averaged response waveform in a window between the T_a th and T_b th post-stimulus samples. Let $\mathcal{T} = \{t_s^{(1)}, \dots, t_s^{(N_T)}\}$ be a set of time indices during which pulses of a particular stimulus variation occurred. Equation (1) shows the formula for the RMS response:

$$RMS(x, \mathcal{T}) = \frac{1}{\sqrt{T_b - T_a + 1}} \left\| \frac{1}{N_T} \sum_{t_s \in \mathcal{T}} \mathbf{x}(t_s) \right\| \quad (1)$$

$$\mathbf{x}(t_s) = \begin{bmatrix} x(t_s + T_a) \\ \vdots \\ x(t_s + T_b) \end{bmatrix}$$

The static modeling data thus consists of m experimentally observed input/output (current/response) pairs, as shown in (2).

$$(\mathbf{I}^{(1)}, y^{(1)}), \dots, (\mathbf{I}^{(m)}, y^{(m)}) \quad (2)$$

$$y^{(i)} = RMS(x, \mathcal{T}^{(i)})$$

where $\mathcal{T}^{(i)}$ is the set of time indices in which the i th stimulus variation occurred.

B. Absolute current model (field-naive)

In one MISO model we consider, the relevant features are the absolute channel currents out of all stimulating channels. This model assigns no importance to the spatial locations of the stimulating electrodes and assumes independence among them. This is a reasonable assumption, considering that in places where the field is highest,

the observed field is dominated by just one channel. This model also assumes that both polarities of biphasic pulses elicit similar effects.

The model consists of a linear combination of currents followed by a nonlinearity. The feature vector $\phi_{current}$ of absolute currents in (5) is weighted by \mathbf{w} and biased by w_0 . A nonlinearity g is then applied, which is scaled by a gain factor α . In this case we chose a logistic nonlinearity for g . Equation (3) shows the predicted output under this model.

$$\hat{y}(\mathbf{I}, \theta) = \alpha g(\mathbf{w} \cdot \phi_{current}(\mathbf{I}) + w_0) \quad (3)$$

$$g(z) = (1 + \exp(-z))^{-1} \quad (4)$$

$$\phi_{current}(\mathbf{I}) = \begin{bmatrix} |I_1| \\ \vdots \\ |I_{N_s}| \end{bmatrix} \quad (5)$$

$$\theta = (\alpha, w_0, \mathbf{w})$$

The optimal values of the model parameters $\theta = (\alpha, w_0, \mathbf{w})$ can be found by minimizing the cost function in (6). A regularization parameter λ serves as a penalty on large values of \mathbf{w} , to prevent overfitting to individual inputs. We used standard gradient-based methods [4] to minimize this combined cost with respect to θ .

$$Cost(\theta) = \frac{1}{2} \sum_{i=1}^m (\hat{y}(\mathbf{I}^{(i)}, \theta) - y^{(i)})^2 + \frac{\lambda}{2} \|\mathbf{w}\|^2 \quad (6)$$

C. Electric field model (field-aware)

We also make a model that considers the electric field at points in the three dimensional space surrounding the array, thus explicitly using the electrode tip locations. A field of particular interest is current density \mathbf{J} , which is proportional to electric field [1]. We can use an analytical solution (7) if we make two assumptions: 1) The extracellular space is a uniform and purely resistive medium. 2) The electrode tips can be approximated as point sources. Let us denote the stimulating locations as $\mathbf{q}_1, \dots, \mathbf{q}_{N_s}$. The current density at a field point \mathbf{p} can hence be calculated using knowledge of the current \mathbf{I} through the electrodes:

$$\mathbf{J}(\mathbf{p}, \mathbf{I}) = \sum_{j=1}^{N_s} \frac{I_j(\mathbf{p} - \mathbf{q}_j)}{4\pi \|\mathbf{p} - \mathbf{q}_j\|^3} \quad (7)$$

Many stimulation studies have shown dependence on electric field [2][12][19]. Tehovnik et al. dubbed $K = \frac{I}{(\text{distance})^2}$ the ‘‘Excitability Constant’’ for a cell. It is a firing threshold that relates the stimulus current to the

distance to the stimulating tip. A simple rearrangement of (7) shows that K is proportional to the amplitude of \mathbf{J} if there is only one stimulus electrode. In this monopolar case, if there were a cell located at \mathbf{p} , then at the cell’s firing threshold, $\|\mathbf{J}\| = K$. We generalize this idea to the multi-electrode case in Eqn. (8) by taking the norm of \mathbf{J} , which we denote as f .

$$f(\mathbf{p}, \mathbf{I}) = \|\mathbf{J}(\mathbf{p}, \mathbf{I})\| \quad (8)$$

f is the scalar strength of the current density at any point \mathbf{p} . We form a field-based prediction (9) that is functionally similar to (3), but with field values at a grid of points $\mathbf{p}_1, \dots, \mathbf{p}_{N_f}$ as features (10).

$$\hat{y}(\mathbf{I}, \theta) = \alpha g(\mathbf{w} \cdot \phi_{field}(\mathbf{p}, \mathbf{I}) + w_0) \quad (9)$$

$$\phi_{field}(\mathbf{p}, \mathbf{I}) = \begin{bmatrix} f(\mathbf{p}_1, \mathbf{I}) \\ \vdots \\ f(\mathbf{p}_{N_f}, \mathbf{I}) \end{bmatrix} \quad (10)$$

Although this model now has spatial organization, it only differs from (3) in its input features. We learn \mathbf{w} using the same minimization methods as in (6). Hereafter, we refer to \mathbf{w} in the field-aware model as the ‘‘sensitivity map’’ for a recording channel.

To encourage smoother solutions of \mathbf{w} , i.e. ones with fewer sharp dips/peaks, we add a roughness penalty term to the cost (11). The penalty takes the squared norm of the 1st order differences of \mathbf{w} . Let $\mathbf{D}_x, \mathbf{D}_y, \mathbf{D}_z \in \mathbb{R}^{N_f \times N_f}$ be the 1st order spatial difference matrices in the x , y , and z directions, respectively. Taking the norm of a vector formed by $\mathbf{D}_x \mathbf{w}$, $\mathbf{D}_y \mathbf{w}$, and $\mathbf{D}_z \mathbf{w}$ has the effect of penalizing solutions that are rough in any direction. The contribution of this penalty is controlled by a hyperparameter μ .

$$Cost(\theta) = \frac{1}{2} \sum_{i=1}^m (\hat{y}(\mathbf{I}^{(i)}, \theta) - y^{(i)})^2 + \frac{\lambda}{2} \|\mathbf{w}\|^2 + \frac{\mu}{2} \left\| \begin{bmatrix} \mathbf{D}_x \mathbf{w} \\ \mathbf{D}_y \mathbf{w} \\ \mathbf{D}_z \mathbf{w} \end{bmatrix} \right\|^2 \quad (11)$$

D. Experimental validation

Two female Long-Evans rats (Hilltop, Scottsdale, PA), animals A and B, were implanted with two different 16-channel arrays (Fig. 2A). One was placed in S1 spanning an area activated by forepaw single digit stimulation. One was placed in the VPL nucleus, targeted to the same receptive field using a map in [6]. The thalamic array (MicroProbes Inc.) was a 2×8 grid of 70% platinum/30%

iridium $75\mu\text{m}$ diameter microelectrodes, with $500\mu\text{m}$ between the rows, and $250\mu\text{m}$ inter-electrode spacing within the rows (Fig.2B). The microelectrodes had a 25:1 taper on the distal 5mm , with a tip diameter of $3\mu\text{m}$. The approximate geometric surface area of the conducting tips was $1250\mu\text{m}^2$. The shank lengths were custom designed to fit the contour of the rat VPL as follows: Both rows were identical. The shaft lengths for each row, from posterior to anterior were (8, 8, 8, 8, 8, 7.8, 7.6, 7.4) mm .

The cortical probe (NeuroNexus Inc.) was an array of iridium electrodes on 4 $15\mu\text{m}$ thick silicon shanks, with 4 electrodes on each shank (See Fig.2C). The shanks were $68\mu\text{m}$ wide with $200\mu\text{m}$ between shanks (AP spacing, in our orientation) and $200\mu\text{m}$ between electrodes on a shank, giving a dorsal-ventral span of $600\mu\text{m}$. The circular electrode surfaces had a diameter of $40\mu\text{m}$.

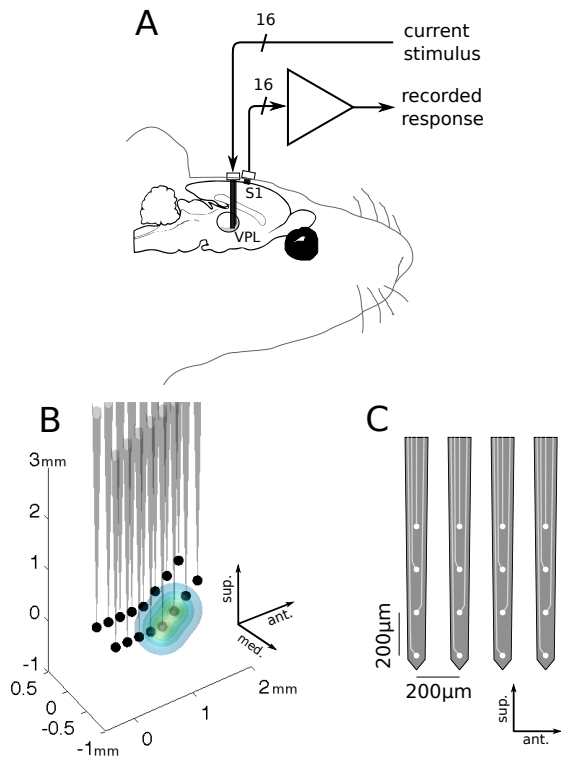


Fig. 2. Experimental Setup: **A**: Diagram of the neural recording and stimulation setup. Field potential recordings are collected at wide-band frequencies from 16 microelectrodes implanted in S1. 2 stimulation leads are routed to specific electrode channels in VPL, delivering pseudo-randomly drawn charge-balanced variable-amplitude biphasic current pulses. **B**: Scale drawing of microwire electrode array used for VPL stimulation, with isocontours of electric field strength for an example bipolar current. Electrode tip locations are highlighted with dots. **C**: Scale drawing of silicon microelectrode array (NeuroNexus Inc.) used for S1 recordings.

All animal procedures were approved by SUNY Downstate Medical Center IACUC and conformed to

National Institutes of Health guidelines. Neural recordings were made using a Multichannel Acquisition Processor system (Plexon, Inc.). Rats were placed into a small chamber with a mesh floor which was suspended above a table. This apparatus helped keep them calm and stationary even though they remained awake. Field potential data from each of the 16 cortical channels was filtered and amplified (gain=1000) through a bandpass filter with cutoffs at 0.7Hz and 8.8kHz. The output signal was then sampled at 20kHz (National Instruments PCI-6071E).

MiSt pulses were delivered with a stimulus isolation unit (AM Systems Model 2200) routed through a switchboard to the VPL array electrodes to create various monopolar and bipolar configurations. At each configuration, we stimulated with 250 biphasic pulses with pseudorandom current amplitudes drawn from $\{10\mu\text{A}, 20\mu\text{A}, 30\mu\text{A}, 40\mu\text{A}\}$. The inter-pulse times were drawn from an exponential distribution with a mean period of 0.5s. We stimulated on 9 monopolar and 9 bipolar configurations in animal A and 16 monopolar and 23 bipolar configurations in animal B.

VPL stimulation caused a small non-saturating artifact in recording channels. It was removed using adaptive noise cancellation [20]. The initially recorded neural signal is assumed to be corrupted additively by artifact, which in turn was caused by a known reference waveform (a stereotyped biphasic square pulse). A causal FIR filter was adapted through recursive least squares (RLS) to reproduce the artifact waveform. This estimate was then subtracted from the incoming signal. The artifact-free signal is then put through a 3rd order Butterworth band-pass filter with cutoffs at 5 and 200Hz and resampled at 800Hz.

Our field-point grid (as used in Eqn. (9)) spanned $200\mu\text{m}$ beyond the 3D extent of the electrode tips. We used an inter-point spacing of $50\mu\text{m}$ which resulted in a total of 18,800 field points. For all stimulus configurations (monopolar/bipolar), the fields were calculated according to (8). The RMS responses (Eqn. (1)) were taken on a window from 12.5ms to 100ms post-stimulus ($T_a = 10, T_b = 80$). We set the regularization parameter in (6) and (11) to a small value, 1×10^{-6} . For the roughness penalty μ , we tried five different values: $\{0.01, 0.05, 0.1, 0.5, 1\}$ to compare their relative performance. Before training the models, we divided the feature vectors ϕ_{field} and ϕ_{current} by their respective maximum (scalar) attempted values. This made all of the element values of both vectors exist between 0 and 1.

III. RESULTS

A. Microstimulation and natural responses were spatially similar

Microstimulation at various VPL configurations produced a wide range of LFP response strengths in S1. The RMS strength, as defined in Eqn. (1), ranged from 0.02 to $202.48\mu V$ in animal A and 5.01 to $349.92\mu V$ in animal B. Fig. 3(a) shows the variety of responses in one recording channel. Since only 4 amplitudes were used, the continuum of responses is largely due to stimulus configuration. Temporal responses on typical channels were highly stereotyped. Most consisted of short negative dips followed by positive segments lasting roughly 60-80ms. The RMS values mentioned hereafter are max-scaled to the largest observed response and hence only have values between 0 and 1. As mentioned before, the input features in both models are also max-scaled. The model parameters and validation results shown hereafter reflect these scaled versions of the inputs/outputs.

Fig. 3(b) shows the average RMS responses across all stimulus variations. They are shown arranged by recording site. These recording channels also had similar responses to natural stimulation, as shown in Fig. 4. The corresponding VPL spike responses are shown in units of spikes/s over baseline. This was calculated by taking the mean increase (above baseline) in firing rate in a window of $80ms$ post-stimulus with $3ms$ bins. Roughly speaking, the strength and spatial pattern of natural S1 activation coincided with that of microstimulation.

B. Field-awareness increased generalization ability

We compared spatial generalization ability by leave-one-configuration-out (LOCO) validation. This measured a model's predictive ability for configurations not included in model fitting. In LOCO, all examples from one configuration at a time are excluded from the training set (2) and later tested upon. Fig. 5 shows a subset of real and predicted multichannel responses. The responses are sorted by real average response magnitude across channels. The field-aware model is more accurate in predicting general trends of strong and weak responses. Correlation (R^2) performance during LOCO also shows this trend (See Fig. 7). The mean and standard deviation of R^2 across channels for the aware and naive models is shown in Table I. In both animals, the field-aware models yielded significantly less (sign-rank $p < 0.01$) prediction error than the field-naive models.

The field-aware model's richer feature set is responsible for the improved prediction accuracy. As shown in Fig. 6, the weighted inputs ($w \cdot \phi$) to the nonlinearity yield less residual variance about the fitted curve,

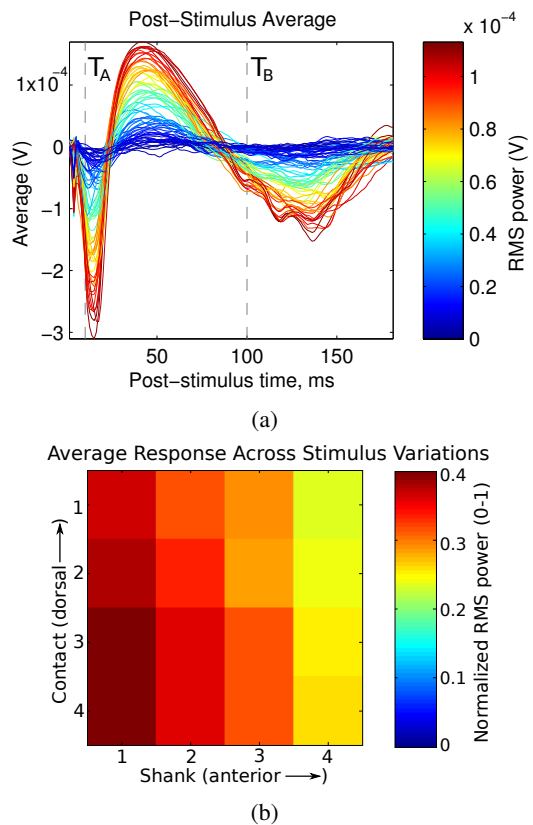


Fig. 3. (a) Averaged post-stimulus waveforms for 72 different stimulus variations, or (electrode configuration, current amplitude) pairs. The configurations used were monopolar and bipolar, and the current amplitudes used were 10, 20, 30, and $40\mu A$. Shown are averages across 62 presentations of each stimulus variation. The color represents the corresponding RMS amplitude of each response. The scalar RMS amplitude in the window delimited by the dotted lines is the output of the static model we discuss in this report. (b) Normalized average MiSt RMS response amplitudes arranged by cortical recording site. These amplitudes are max-scaled by the largest observed RMS amplitude, and hence are between 0 and 1. Our cortical recording array was a 4×4 NeuroNexus probe with 4 shanks ($200\mu m$ spacing) and 4 contacts ($200\mu m$ spacing) per shank (See Fig. 2C)

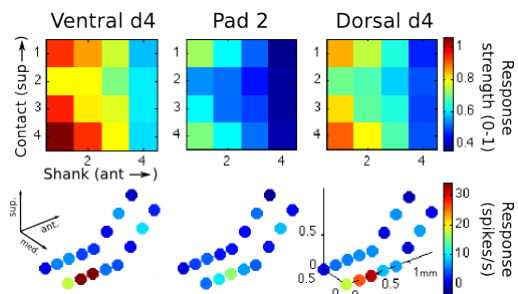


Fig. 4. Example of response strengths in animal A to tactile stimulation at three different sites on the hand, arranged by channel location. **Top row:** RMS LFP strength in cortex. **Bottom row:** Corresponding VPL spike responses measured in spikes/s above baseline.

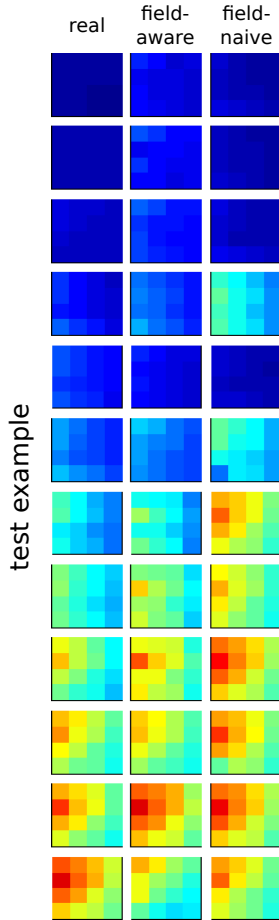


Fig. 5. LOCO output comparison for animal A. Each row shows responses to a single MiSt variation. A subset of cortical responses is shown in the left column. This subset is every fifth response out of a sorted list of real multichannel responses. The corresponding output predictions from the field-aware model and field-naive models are shown in the center and right columns, respectively.

compared to the field-naive model. The nonlinear curve for the field-aware model also showed consistently less saturation. In other words, the spatial feature mapping produced a more linear and more accurate relationship to responses.

Are vector field interactions due to the array geometry important for generalization? If they were not, then shuffling or perturbing the supposed electrode locations (before training) would not degrade prediction accuracy. Since the areas that see the strongest field would depend on just one electrode, shuffling these features would simply yield an equivalent model. The differences between the shuffled and non-shuffled situations would be in the field interactions resulting from the specific layout of the array. The LOCO performance of the field-aware model and a shuffled model are shown in Fig. 7. The shuffled model does indeed do worse ($p < 0.01$) than the non-shuffled version, which suggests that accurate

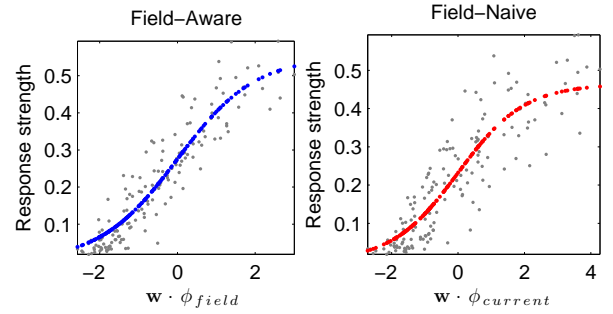


Fig. 6. 1-dimensional representation of both models showing only the nonlinear stage. (See Eqns. (3), (9)). The horizontal axis is the output from applying \mathbf{w} to the feature vectors ϕ_{field} or $\phi_{current}$. The vertical axis represents response strength. Gray dots are the training outputs used during fitting, and the model outputs are shown in blue or red. The nonlinearity is logistic with a scaling factor α . The field-aware models in all rats and experiments exhibited less saturation than the corresponding field-naive models.

geometric information produces more accurate models. This comparison was made across all LOCO test sets and recording channels. The shuffled model's mean and standard deviation of R^2 across channels is shown in Table I.

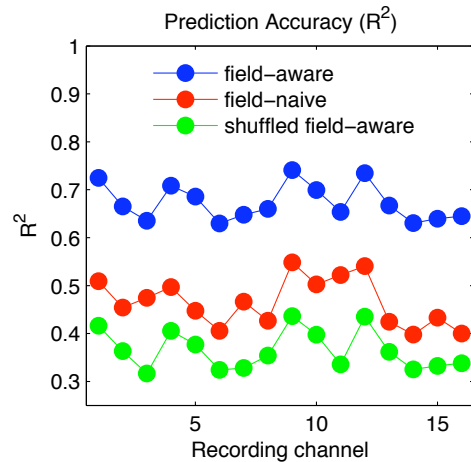


Fig. 7. Prediction accuracy measured by squared correlation coefficient (R^2) on LOCO validations for all recording channels. The field-aware model outperforms the field-naive and shuffled field-aware models.

How sensitive are the parameter estimates to removing stimulus configurations from the training set? We measured how much the predicted output of a model varied during LOCO compared to leaving no configurations out. The squared correlation coefficient R^2 was taken between values of \hat{y} during LOCO and values of \hat{y} when using all of the training data. This can be interpreted as a scalar measure of the stability of \hat{y} (and the model parameters) when single configurations are

removed from the training set. We refer to this quantity as the "LOCO stability" of a model, and it is shown for all models and animals in Table I. The field-aware model had higher stability than the naive or shuffled model. This also means that the field-aware model's parameters exhibit better stability when faced with limited training data.

Does the nonlinearity g help in producing better predictions or does it lead to overfitting? We tested this by training alternate linear models where g in Eqns. (3), (9) was replaced with identity. In this case, the predicted response was simply a linear combination of current/field inputs (with a constant bias term). The LOCO generalization results are shown in Table I. The nonlinear field-aware models outperformed their linear counterparts (and all other competing models) in both animals. They had significantly higher prediction accuracy (R^2) under a sign-rank test over recording channels ($p < 0.01$, $p = 0.03$ for animals A and B respectively). The nonlinear field-aware models also showed significantly ($p < 0.01$) better LOCO stability.

TABLE I

MODEL COMPARISON OF LOCO PREDICTION R^2 PERFORMANCE (MEAN \pm STD. DEVIATION ACROSS RECORDING CHANNELS) AND LOCO STABILITY.

| | | Prediction R^2 | Stability R^2 |
|----------|----------------------|------------------|-----------------|
| animal A | field-aware | 0.67 ± 0.04 | 0.89 ± 0.01 |
| | field-naive | 0.47 ± 0.05 | 0.76 ± 0.02 |
| | shuffled field-aware | 0.37 ± 0.04 | 0.74 ± 0.02 |
| | field-aware (linear) | 0.55 ± 0.04 | 0.77 ± 0.02 |
| | field-naive (linear) | 0.42 ± 0.04 | 0.84 ± 0.02 |
| animal B | field-aware | 0.68 ± 0.05 | 0.97 ± 0.01 |
| | field-naive | 0.41 ± 0.07 | 0.84 ± 0.05 |
| | shuffled field-aware | 0.55 ± 0.07 | 0.91 ± 0.02 |
| | field-aware (linear) | 0.66 ± 0.04 | 0.92 ± 0.01 |
| | field-naive (linear) | 0.51 ± 0.05 | 0.93 ± 0.08 |

C. Spatial characteristics of the field-aware model

Fig. 8 shows a color-coded optimal sensitivity map w (left) for a typical field-aware model. The adjacent plot (right) shows the corresponding optimal weights for the field-naive model. Although the two models resemble one another in gross spatial features, the field-aware model assigns some weight to extracellular space that is not in the immediate vicinity of an electrode. The optimal weights in the field-aware case are mostly positive with a positive skew. For example, in the sensitivity map shown in Fig. 8, the range of the 18,800 weights is $(-0.041, 0.163)$ with median 0.045. Similarly, the weight range in the field-naive model for the same data is $(0.000, 6.685)$ with median 0.794.

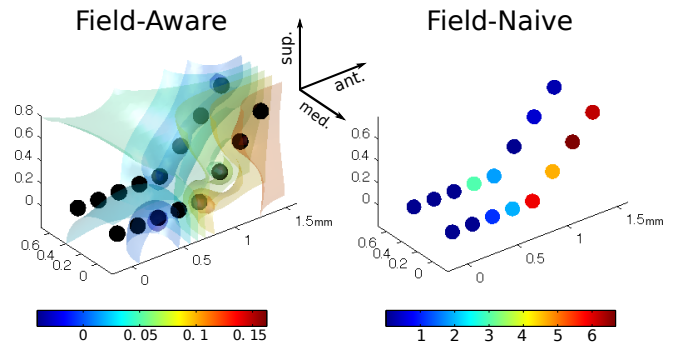


Fig. 8. 3D view of models for a single example cortical channel. Dimensions are in mm . **Left:** A typical 3D sensitivity map used for a field-aware model, trained with $\mu = 0.1$ (shown with isocontour surfaces). **Right:** As a point of comparison, the weights on absolute currents out of the 16 individual electrodes used in the field-naive model show a similar spatial mapping.

Several settings of the roughness penalty factor μ were explored. Its smoothing effect on the sensitivity map is demonstrated in Fig. 9 for two settings of μ . Both plots show the same horizontal slice located $0.118mm$ superior to the bottom of the array. Low values of μ led to a detailed spatial map. However, this improved resolution tended to decrease generalization performance across all channels. The attempted values of μ and the corresponding prediction errors are shown for animal A in Fig. 10. The minimal settings (from this discrete list) of μ were 0.1 and 0.5 for animals A and B, respectively. Despite the simplicity of the spatial gradients involved in this roughness penalty (Eqn. (11)), the best settings of μ improved generalization performance by 2.4% and 9.8% in rats A and B, respectively (% reduction in test error from the $\mu = 0$ case across all channels and LOCO test sets). This suggests that cortical areas have an intrinsic limit on spatial resolution for microstimulation in thalamus. Letting the model capture more detailed field features only leads to overfitting.

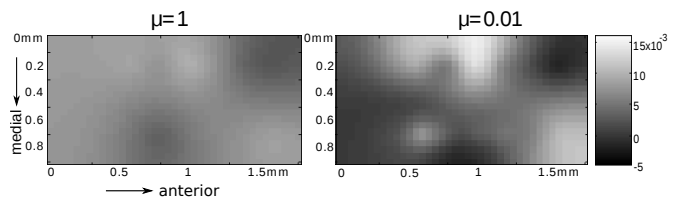


Fig. 9. Comparison of sensitivity maps generated using different roughness penalties for the same data set. Both panels show the same horizontal slice located at $0.118mm$ superior to the bottom of the array. The left panel shows the sensitivity map for a high roughness penalty ($\mu = 1$), and the right panel uses a low roughness penalty ($\mu = 0.01$). Each map is scaled to its norm, and thus represents the shape and amplitude for the unit norm electric field stimulus that optimally excites the cortical channel.

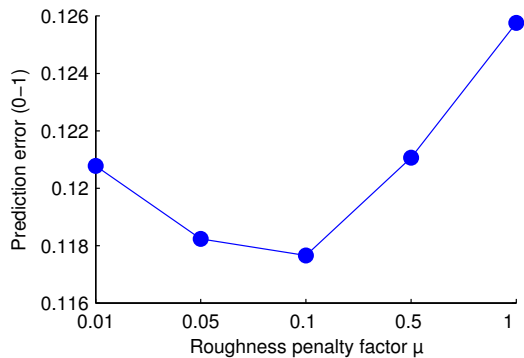


Fig. 10. Tuning the roughness penalty μ . Average prediction error (across channel models and LOCO test sets) versus selected μ settings. Since the outputs were max-scaled, these errors values signify the ratio of error versus the largest encountered RMS value.

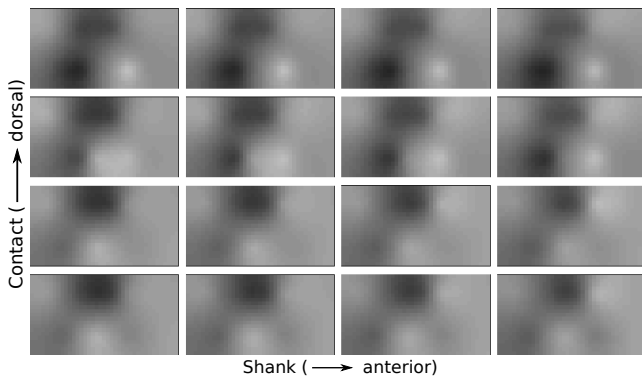


Fig. 11. Sensitivity maps for all 16 cortical channels in animal B, arranged by recording site. The model parameter $\mu = 0.1$ was used in all models. Each plot shows the same horizontal section with the same spatial and color scales as in Fig. 9.

We note that the sensitivity maps for all cortical channels are quite similar, but have qualitative features (such as the centers, widths, and amplitudes of bumps and valleys) that show trends according to recording location (see Fig. 11). For example, in the third row of Fig. 11, the light bump in the top right of the map appears to get brighter for increasingly anterior cortical channels. We expect the maps to be mostly similar, since the recording electrodes were so close to one another. The horizontal span of the whole cortical array was only $0.6mm$.

IV. DISCUSSION

A. Generalization ability

We have demonstrated that field-aware models of MiSt can generalize to un-encountered electric fields. These may result from arbitrary stimulation configurations as long as the spatial sampling of the training data is

sufficiently rich. For neuroprosthetic applications, this provides a systematic way to select stimulation configurations likely to generate desired responses without having to try each one first. In bipolar stimulations, the number of possible configurations already reaches $\frac{1}{2}N(N-1)$ where N is the number of electrodes (120 for a 16 channel array). Of course, some educated guesswork may narrow down the number of plausibly efficacious pairs. However, for a higher number of simultaneously-stimulating electrodes (SSE), the number of possible configurations becomes so large that it is arguably prohibitive to sample exhaustively. In order to capitalize on the full range of spatially varied SSE patterns, a generalizable model is required. Here we have reported a model that can be trained on a limited subset of current input configurations yet performs reliably in predicting responses to novel inputs. Electric field is what unites different SSE configurations into a common space. Thus, by fitting responses to electric fields, our method can accommodate new, arbitrarily complex MiSt configurations. It is notable that the field-aware and field-naive models had access to the same adaptable non-linearity (compare Eqns. (3) and (9)), so prediction improvement in the field-aware case is attributable to the lattice of electric field sample points acting as a better feature set.

B. Sensitivity maps as functional brain mapping

For a given cortical recording, the sensitivity maps (e.g. Figs. 11 and 8) can be viewed as highlighting regions of most efficacious input to the thalamocortical column. This can be roughly interpreted as the cortical “receptive fields” of MiSt. These maps could be interpreted anatomically if an image of the implanted array (e.g. CT scan) were taken and used to register them to the brain structures they occupy. Such a procedure offers a means for obtaining higher-resolution functional mapping of brain connections than mapping based on electrode tip locations alone. This is naturally subject to the caveat that we treat the tissue as a uniform conductive medium.

The static spatial model presented in this paper could also provide a starting point for a dynamical model that captures both spatial and temporal effects of MiSt. Thus, armed with a field-aware model and knowledge of the somatotopic map of VPL [6], we aim to stimulate thalamic inputs to cortex in specific patterns to achieve more naturalistic modulation of S1.

ACKNOWLEDGMENT

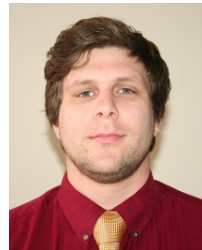
We thank L. von Kraus for useful comments and discussions.

REFERENCES

- [1] R C Barr and R Plonsey. *Bioelectricity, a Quantitative Approach*. Springer, New York, 3rd edition, 2007.
- [2] S L BeMent and J B Ranck. A quantitative study of electrical stimulation of central myelinated fibers with monopolar electrodes. *Exp Neurol*, 24:147–70, 1969.
- [3] J A Bierer, S M Bierer, and J C Middlebrooks. Partial tripolar cochlear implant stimulation: spread of excitation and forward masking in the inferior colliculus. *Hearing Research*, 270:134–42, Dec 2010.
- [4] C M Bishop. *Pattern Recognition and Machine Learning*. Springer, New York, 2006.
- [5] A Chaturvedi, CR Butson, SF Lempka, SE Cooper, and CC McIntyre. Patient-specific models of deep brain stimulation: influence of field model complexity on neural activation predictions. *Brain Stimulation*, 3:65–7, 2010.
- [6] J T Francis, S Xu, and J K Chapin. Proprioceptive and cutaneous representations in the rat ventral posterolateral thalamus. *Journal of Neurophysiology*, 99(5):2291–304, 2008.
- [7] R W Guillery and S M Sherman. Thalamic relay functions and their role in corticocortical communication: generalizations from the visual system. *Neuron*, 33(2):163–75, 2002.
- [8] D L Kimmel and T Moore. Temporal patterning of saccadic eye movement signals. *J Neuroscience*, 27:7619–30, 2007.
- [9] N H Lovell, E Cheng, G J Suaning, and S Dokos. Stimulation of parallel current injection for use in a vision prosthesis. *2nd Annual International Conference in Neural Engineering*, Mar 2005.
- [10] C C McIntyre, S Mori, D L Sherman, N Thakor, and J L Vitek. Electric field and stimulating influence generated by deep brain stimulation of the subthalamic nucleus. *Clinical Neurophysiology*, 115:589–95, 2004.
- [11] S Miciovic, M Parent M, C R Butson C R, P J Hahn, G S Russo, J L Vitek, and C C McIntyre. Computational analysis of subthalamic nucleus and lenticular fasciculus activation during therapeutic deep brain stimulation. *Journal of Neurophysiology*, 96(3):1569–80, 2006.
- [12] C M Murasugi, C D Salzman, and W T Newsome. Microstimulation in visual area mt: effects of varying pulse amplitude and frequency. *J Neuroscience*, 13(4):1719–29, 1993.
- [13] W T Newsome and M R Cohen. What electrical microstimulation has revealed about the neural basis of cognition. *Curr Opin Neurobiol*, 14(2):169–77, April 2004.
- [14] J B Ranck. Which elements are excited in electrical stimulation of mammalian central nervous system: a review. *Brain Research*, 98:417–40, 1975.
- [15] R Romo, A Hernandez, A Zainos, C D Brody, and L Lemus. Sensing without touching: psychophysical performance based on cortical microstimulation. *Neuron*, 26:273–8, 2000.
- [16] R Romo, A Hernandez, A Zainos, and E Salinas. Somatosensory discrimination based on cortical microstimulation. *Nature*, 392:387–90, 1998.
- [17] C D Salzman, K H Britten, and W T Newsome. Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, 346:174–77, 1990.
- [18] M Steriade. Corticothalamic resonance, states of vigilance and mentation. *Neuroscience*, 101(2):243–76, 2001.
- [19] E J Tehovnik, A S Tolia, M F Sultan, W M Slocum, and N K Logothetis. Direct and indirect activation of cortical neurons by microstimulation. *J Neurophysiology*, 96:512–21, 2006.
- [20] B Widrow, J R Glover, J M McCool, J Kaunitz, C S Williams, R H Hearn, J R Zeidler, E Dong, and R C Goodlin. Adaptive noise cancelling: principles and applications. *Proceedings of the IEEE*, 63(12):1692–716, 1975.



John S. Choi John S. Choi is a Ph.D. candidate in the joint Biomedical Engineering program between SUNY Downstate and NYU Polytechnic in Brooklyn, NY. He received his undergraduate degree in biomedical and electrical and computer engineering at Duke University in 2008. John’s focus is in computational neuroscience, neural prosthetic devices, and machine learning.



functional neural correlates of decision-making variables.

Marcello M. DiStasio Marcello DiStasio is an MD/PhD candidate in the biomedical engineering program at SUNY Downstate Medical Center and NYU-Polytechnic. He received his degree in neurobiology and behavior from Cornell University in 2005. His research interests span electrophysiology, motor control, information processing in neural circuits, computational modeling of spiking networks, and



computer science.

Austin’s research interests cover signal processing, machine learning, and computational neuroscience as well as mathematics, science, and engineering education. He is the co-author of an upcoming neural engineering textbook chapter on brain-machine interfaces.

Austin J. Brockmeier Austin J. Brockmeier is a Ph.D. student in the Electrical and Computer Engineering Department at the University of Florida. His undergraduate studies were at the University of Nebraska at Omaha and Peter Kiewit Institute; upon completion he received a B.S. degree in computer engineering from the University of NebraskaLincoln with a second major in mathematics and minor in



Joseph T. Francis Joseph T. Francis graduated from the honors program in biology at SUNY Buffalo in 1994. Subsequently he studied neural dynamics with an emphasis on non-linear dynamical systems theory applied to the nervous system, as well as ephaptic interactions at The George Washington University in Washington DC, PhD 2000.

He conducted two post doctoral fellowships, the first was in computational sensorimotor control and learning under the guidance of Reza Shadmehr at Johns Hopkins University. He then moved onto Brain Machine interfacing with John Chapin at SUNY Downstate, where he later took on a faculty position and continues to work today.