# Closed-Loop Reinforcement Learning Based Deep Brain Stimulation Using SpikerNet: A Computational Model

Brandon S. Coventry and Edward L. Bartlett

***Abstract*— Clinical deep brain stimulation (DBS) has proven effective in treating neurological and neuropsychiatric conditions such as Parkinson's disease and obsessive compulsive disorder. However, current DBS systems operate in open-loop conditions, using stimulation that is agnostic to patient symptomology. Furthermore, current DBS systems require precise parameter tuning for effective stimulation and may require retuning due to device age and long term neural adaptation. The relatively simple control algorithms of current gold-standard closed-loop and adaptive DBS systems sometimes have better clinical outcomes but do not track the patient's dynamic brain states and may be associated with interruptions of voluntary movement. We recently introduced SpikerNet, a closed-loop neuromodulation platform which uses deep reinforcement learning to track brain state changes and actively tune stimulus parameters in real time to achieve targeted neural firing patterns, with the ability to adapt to changing neural dynamics. In this paper, we evaluate SpikerNet in a computational basal ganglia-thalamocortical model of DBS to assess run time and stimulation dynamics. We find that SpikerNet rapidly finds stimulus parameters to restore dopamine depleted activity back to naturalistic states, is stable against rapid and drastic changes in the neural environment, and generates novel stimulation hypotheses for evaluation in preclinical models to derive novel stimulation regimes.**

***Index Terms*— Closed-Loop DBS, Deep Brain Stimulation, Deep Learning, Reinforcement Learning**

## I. Introduction

Deep brain stimulation (DBS) is a proven and effective treatment for movement disorders such as Parkinson's disease and essential tremor and has recently received FDA approval for treatment of drug refractory obsessive-compulsive disorder. However, clinical DBS paradigms operate in an open-loop fashion, not accounting for patient symptomology and requiring in-clinic parameter tuning both when the stimulator is initially turned on and as stimulation efficacy changes due to glial encapsulation of the electrode and neural adaptation to applied stimuli[1,2]. Furthermore, open-loop DBS has been found to increase brain-wide metabolic demands, potentially leading to cell death and fundamental alterations of neural dynamics in sensorimotor circuits[3].

Closed-loop adaptive DBS strategies have been introduced to mitigate deficits found in open-loop DBS. In particular, β band activity across every region in the basal ganglia-thalamocortical (BTC) circuit has been found to be correlated to Parkinsonian symptomology [4–6] and has been used as a biomarker for closed-loop feedback control [7]. Current closed-loop DBS systems use relatively simple control algorithms, largely regulated to turning on or off stimulation based on single or dual threshold measurements of β band activity [8–9]. Initial clinical trials show improvement in stimulation efficacy[10], reductions in side effects[11], and improved battery life in closed-loop versus open-loop stimulation. However, there is evidence that closed-loop DBS can interfere with volitional movement[9] likely due to interactions with β-burst activity arising from planned movement in subthalamic nucleus (STN) and the internal globus pallidus (GPi)[12] suggesting a need for systems with more robust neural feature selection methods. Moreover, current systems do not track patient response or titrate stimulation based on β band correlated symptomology with limited ability to search for optimal parameters that could provide effective stimulation with reduced patient side-effects.

To account for these deficits, we recently introduced SpikerNet [13], a closed-loop, reinforcement learning (RL) based neuromodulation system. RL is a method by which an artificial agent learns optimal policies to maximize rewards through repeated iterations of a test environment, developing policies of the best action to take to maximize short and long term rewards. In the case of SpikerNet, agents learn real-time statistical models of the measured neural environment and its dynamical response to applied stimulation with the policy mapping out best sequential stimuli to drive neural patterns to desired states. As policy functions are often nonlinear and difficult to estimate, SpikerNet uses deep RL in which policies are learned using deep feedforward neural networks as expressive nonlinear functional approximators. Importantly, RL has the ability to adapt to changing environments [14] allowing for tracking of short-term changes in neural firing, such as sleep-wake cycles or long-term changes due to neural adaptation or electrode encapsulation.

The goal of this work is to assess runtime costs, search efficiency, and the generation of DBS stimulation parameters

B.S. Coventry was with the Weldon School of Biomedical Engineering and the Institute for Integrative Neuroscience, Purdue University West Lafayette. He is now with the Department of Biomedical Engineering and the Wisconsin Institute for Translational Neuroengineering, University of Wisconsin-Madison, Madison, WI 53907 USA (e-mail: coventry@wisc.edu).

E.L. Bartlett is with the Department of Biological Sciences, the Weldon School of Biomedical Engineering, and the Institute for Integrative Neuroscience, Purdue University, West Lafayette IN 49707 USA (e-mail: ebartle@purdue.edu).

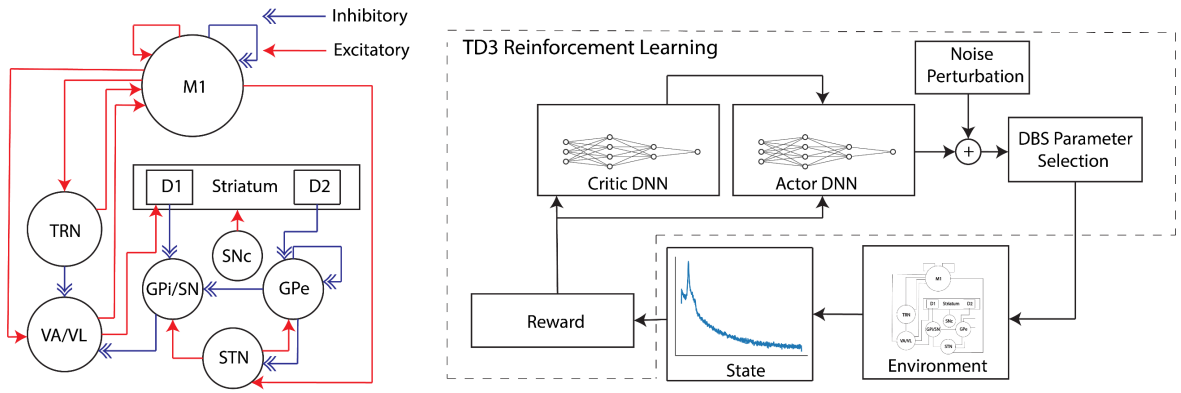


**Fig. 1**. Schematics of basal-ganglia thalamocortical DBS model and SpikerNet. Left: Basal-ganglia thalamocortical model circuit with excitatory (red) and inhibitory (blue) connections. STN: Subthalamic nucleus, GPi/SN: Internal globus pallidus/substantia nigra reticulata, GPe: External globus pallidus, SNc: Substantia nigra pars compacta, D1/D2: Striatum neurons expressing D1/D2 receptors respectively, M1: Motor Cortex, VA/VL: Ventral anterior/ventral lateral nucleus of thalamus, TRN: Thalamic reticular network. DBS stimulation is performed in STN with LFP recordings modeled from GPi Right: Schematic description of SpikerNet using the basal-ganlia thalamocortical model as the environment. Evoked LFP power spectral densities from GPi were used as measure of the state with action selection corresponding to choice of DBS parameters. twin-delayed deep deterministic policy gradients (TD3) was used as the reinforcement learning method.

in SpikerNet. To achieve these goals, a computational model of the SpikerNet system was developed using a BTC DBS model as the stimulation environment.

## II. Methods

### A. Basal Ganglia Thalamocortical Model

In order to establish computational baselines of SpikerNet, a biophysically relevant BTC circuit model was adapted from Grado et al. [15]. This model incorporates all major nuclei in the basal ganglia circuit and includes conventional DBS in STN, along with options for normal and dopamine depleted circuit conditions. BTC models utilized mean-field estimation[16] of regional responses, with response of each neuron modeled as

$$Q_n(V_n) = Q_n^{\max} \int H(V_n - v)p(v)dv$$

where $Q_n, Q_n^{\max}$ are the mean firing rate and max firing rate of neuron n, respectively, $H(*)$ is the Heavyside step function, $V_n$is the membrane potential, $v$ is the threshold potential, and $p(*)$ is the distribution of threshold potentials. Changes in cellular membrane potential from synaptic input is modeled as:

$$\Delta V = \sum \eta\varphi(t-\tau)$$

where $\eta$ is the aggregate change in membrane potential to a single synaptic event across all synaptic inputs and $\varphi$ is the arrival rate of synaptic events accounting for axon latency $\tau$. Details of parameter changes leading to dopamine depleted states can be found in [16]. Local field potentials (LFP) were simulated and recorded from GPi with LFP power spectral densities estimated using Welch's method. A Schematic of the model is given in Fig 1.

### B. SpikerNet

SpikerNet was developed as a system to implement closed-loop reinforcement learning based control that was agnostic to stimulation modality and target. SpikerNet is schematized in Fig 1. All models were implemented in Python with PyTorch deep learning backend[17]. To validate SpikerNet's performance, a custom Open AI gym basal ganglia-thalamocortical environment was implemented. The action space was defined as a continuous environment with DBS stimulation amplitude, pulse width, and stimulation frequency as adjustable parameters. The observation state was defined as a continuous environment of the evoked LFP power spectral density of GPi in response to STN stimulation. For studies assessing whether SpikerNet could find stimulation patterns to return β band oscillations from dopamine depleted to normal states, a reward function was defined as:

$$Reward = \left[\sum_{i}^{\beta}\left(\hat{P}_i - P_i\right)^2\right]^{-1}$$

where $\hat{P}_i$ and $P_i$ are the target and measured power at frequency $i$ within the β band respectively. β band reduction experiments were given a reward target of

$$Reward = \left|\frac{P_{DD} - P_{Obs}}{P_{DD}} * 100\right|$$

where $P_{DD}, P_{Obs}$ are the mean β band power of dopamine depleted and stimulation induced LFPs, respectively. Reinforcement learning was performed via the twin delayed deep deterministic policy gradients (TD3) method [18] implemented using GenRL[19]. Experiments were run for 50 episodes, where an episode was completed if the reward exceeded a reward threshold. Reward threshold for template matching experiments was set to 1.5, equating to a mean square error of 0.067 between target and observed β activity. β band reduction experiments were given a target reward of 25% reduction in β activity as referenced to dopamine depleted, stimulus free models. It is important to note that SpikerNet rewards are encouraged via reward maximization to go above reward threshold, with thresholds demarcating the observation of a good fit, ending an episode and allowing the agent to reinitialize and attempt to find better stimulus trajectories and fits.

## III. Results

### A. SpikerNet fits target responses

We first evaluated SpikerNet's ability to find DBS parameters that drive BTC model responses from a dopamine

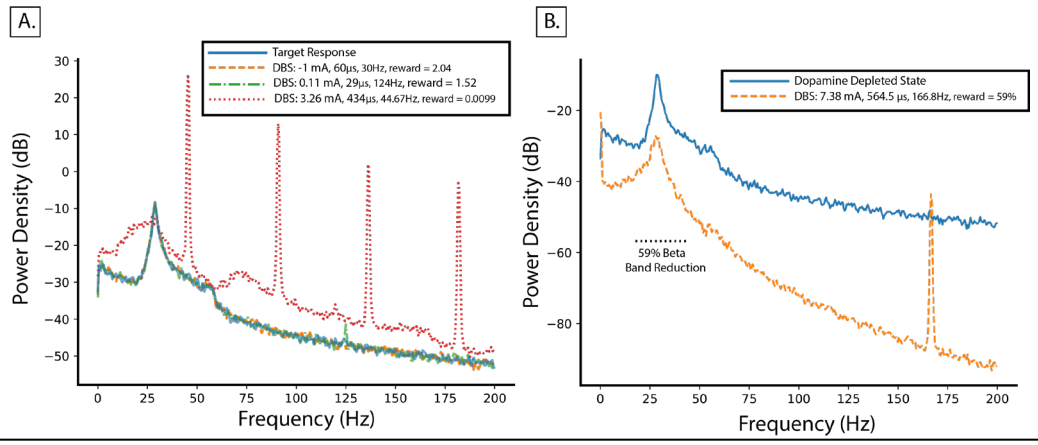


**Fig 2**. A. SpikerNet finds parameters which move BTC model responses from dopamine-depleted to naïve states. Example fit responses and DBS parameters with one example of non-effective stimulation (red dotted line). B. Reward function only observing reductions in β band power also results in reductions of off-target LFP frequency bands.

depleted to normal state. The target state was obtained by evaluating the model under normal dopamine conditions with no DBS applied (Fig 2A). Training was performed under dopamine depleted conditions. SpikerNet was able to find DBS parameters which moved dopamine depleted model responses back to normal states (Fig 2A). RL training mixes exploration and targeting phases in parameter searching, such that SpikerNet was able to find a diversity of parameters which correct dopamine depleted states, and thus generates falsifiable hypotheses which can be used to evaluate efficacy of novel stimulation parameters.

Many closed-loop DBS systems under clinical investigation utilize β band power as the biomarker for patient symptomology with reduction of β band activity serving as the goal of the control system, a looser constraint than template matching. To recapitulate this biomarker, SpikerNet was trained on a reward of percent reduction in β band power within the context of the dopamine depleted model. We found that this looser constraint can lead to significant reductions in β band power, but at the cost of inhibition of power in unrelated frequency bands (Fig 2B). This suggests that more informative reward functions using established models of naturalistic neural firing patterns provide more optimal control of pathologic firing patterns.

### *B. SpikerNet Evaluation Time*

As reinforcement learning trains with repeated iterations through an environment, the average number of iterations until first success and the average total number of iterations needed to complete a 50 episode experiment in the template matching condition were calculated. The median number of trials to first success above a reward threshold of 1.5 was 6.0 (s.d. = 7.18, n = 23) and median number of trials to experiment completion was 327.0 (s.d. = 674.60, n = 23). Distributions for first success and trials to experiment completion are given in Fig 3A. However, choice of total number of episodes in RL training is a heuristic based on domain knowledge of the environment under test, and as such should be subject to sensitivity analyses to ensure the neural environment is sampled adequately. In order to evaluate time to stimulus stability and number of trials needed for training, stimulus parameters were plotted as a function of trial number through the duration of the experiment (Fig 3B). It was found that parameter selection was largely performed between onset of training and the first crossing of the reward threshold, with smaller refinements made through the rest of training. This suggests optimal parameters can be found quickly, and training made reasonably short, supporting initial goals of SpikerNet to train to find effective stimuli, only retraining when a significant state change has occurred or previous stimulation is no longer effective. More episodes than strictly necessary can also be added to develop a more robust statistical model of the neural environment as a trade off with training time.

### *C. SpikerNet is Stable Across Changing Brain States*

Neural firing patterns are subject to change from normal circadian rhythms as well as patient arousal and valence. To test the stability of SpikerNet to changing neural firing patterns, runs were conducted in which the BTC model was

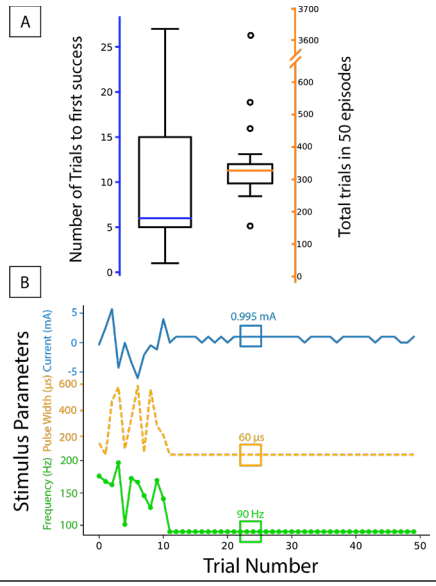


**Fig 3**. A: Distribution of number of trials until first successful crossing of reward threshold (left) and total number of trials needed to complete a 50 episode experiment (Right). Total number of experiments = 23. B: Stimulus parameter rapidly converge after first threshold crossing.

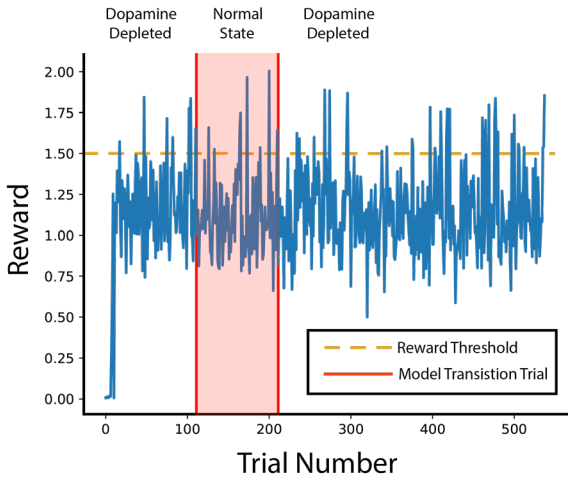


**Fig 4**. SpikerNet is stable to drastically changed neural firing patterns as tested by rapid changes between dopamine depleted and normal model states.

randomly switched from dopamine depleted to normal states during training. Such a state change represents an extreme scenario, with firing pattern changes more drastic than in normal circadian or arousal changes. The model was allowed to train normally beginning in the dopamine depleted state with every reward threshold crossing having a 10% probability of switching from dopamine depleted to normal model environment states. After the first switch, every subsequent reward threshold crossing had a 10% chance to change back from normal to dopamine depleted model state. It was found that SpikerNet is resistant to extreme state changes, with stimulation rapidly adjusting to the changing environment as suggested by stable, high reward fits after state changes (Fig 4) with a mean reward change from dopamine depleted to normal state of -0.0077 (std = 0.34, n = 12) and normal to dopamine depleted of -0.07 (std = 0.19, n = 9).

## IV. Discussion and Conclusions

The present study uses computational BTC models to assess SpikerNet in Parkinsonian DBS applications, finding that SpikerNet obtains viable stimulation parameters in few iterations and is stable against drastic state changes. SpikerNet was developed to rapidly find a target brain state and learn stimulus trajectories to continuously titrate stimulation to maintain a target neural state. Training is reinitialized online when stimulation is no longer achieving its target state, adapting to patient needs and reducing return visits to the clinic. This study shows that training is rapid, stable and is able to generate a diversity of effective stimulus parameters. Our finding that simpler β band reduction reward functions may produce compromised neural responses generalize to any closed-loop DBS systems and potentially describes maladaptive closed-loop interactions with volitional movement[6]. Our findings then suggest that even slightly more informative control functions provide better neural control. While normal patient brain state cannot be obtained after disease onset, general reward functions can be obtained through models of BTC function through clinical and basic science collaborative studies. β band reduction is a prerequisite for volitional movement and thus should be accounted for in neural control algorithms. SpikerNet's ability to learn and adapt from a multiplicity of neural firing states suggests an ability to adapt to resting and volitional movement states, which will be tested in subsequent *in vivo* models.

A caveat to this study is that DBS parameter search was constrained to ensure that stimulation would be achievable in implantable pulse generator hardware, but no checks were made to ensure predicted stimuli would fall within clinically safe boundaries[20–21]. Past *in vivo* realizations of SpikerNet do include stimulation parameter constraints which can be adjusted to fit that target model and stimulation paradigm[13]. Future studies will test derived SpikerNet parameters in this modeling study with *in vivo* preclinical Parkinsonian DBS models.

## V. Disclosures and Code Availability

B.S.C. and E.L.B. hold a provisional patent on the reinforcement learning based closed-loop neuromodulation system used in SpikerNet. Code used in this study is available upon request.

## VI. References


[1] T. J. van Hartevelt *et al.*, "Neural Plasticity in Human Brain Connectivity: The Effects of Long Term Deep Brain Stimulation of the Subthalamic Nucleus in Parkinson's Disease," *PLoS One*, vol. 9, no. 1, p. e86496, Jan. 2014, doi: 10.1371/journal.pone.0086496.

[2] K. van Kuyck, M. Welkenhuysen, L. Arckens, R. Sciot, and B. Nuttin, "Histological Alterations Induced by Electrode Implantation and Electrical Stimulation in the Human Brain: A Review," *Neuromodulation: Technology at the Neural Interface*, vol. 10, no. 3, pp. 244–261, Jul. 2007, doi: 10.1111/j.1525-1403.2007.00114.x.

[3] J. Liu, L. Li, Y. Li, Q. Wang, R. Liu, and H. Ding, "Metabolic Imaging of Deep Brain Stimulation in Meige Syndrome," *Front Aging Neurosci*, vol. 14, Mar. 2022, doi: 10.3389/fnagi.2022.848100.

[4] Y. Yu *et al.*, "Parkinsonism Alters Beta Burst Dynamics across the Basal Ganglia–Motor Cortical Network," *The Journal of Neuroscience*, vol. 41, no. 10, pp. 2274–2286, Mar. 2021, doi: 10.1523/JNEUROSCI.1591-20.2021.

[5] R. S. Eisinger *et al.*, "Parkinsonian Beta Dynamics during Rest and Movement in the Dorsal Pallidum and Subthalamic Nucleus," *The Journal of Neuroscience*, vol. 40, no. 14, pp. 2859–2867, Apr. 2020, doi: 10.1523/JNEUROSCI.2113-19.2020.

[6] R. Levy, P. Ashby, W. D. Hutchison, A. E. Lang, A. M. Lozano, and J. O. Dostrovsky, "Dependence of subthalamic nucleus oscillations on movement and dopamine in Parkinson's disease," *Brain*, vol. 125, no. 6, pp. 1196–1209, Jun. 2002, doi: 10.1093/brain/awf128.

[7] S. Little *et al.*, "Adaptive deep brain stimulation in advanced Parkinson disease.," *Ann Neurol*, vol. 74, no. 3, pp. 449–457, 2013, doi: 10.1002/ana.23951.

[8] M. Scherer *et al.*, "Single-neuron bursts encode pathological oscillations in subcortical nuclei of patients with Parkinson's disease and essential tremor," 2022, doi: 10.1073/pnas.

[9] S. Little and P. Brown, "Debugging Adaptive Deep Brain Stimulation for Parkinson's Disease," *Movement Disorders*, vol. 35, no. 4, pp. 555–561, Apr. 2020, doi: 10.1002/mds.27996.

[10] B. Rosin *et al.*, "Closed-loop deep brain stimulation is superior in ameliorating parkinsonism," *Neuron*, vol. 72, no. 2, pp. 370–384, Oct. 2011, doi: 10.1016/j.neuron.2011.08.023.

[11] S. Little *et al.*, "Adaptive deep brain stimulation for Parkinson's disease demonstrates reduced speech side effects compared to conventional stimulation in the acute setting," *Journal of Neurology, Neurosurgery and Psychiatry*, vol. 87, no. 12. BMJ Publishing Group, pp. 1388–1389, Dec. 01, 2016. doi: 10.1136/jnnp-2016-313518.

[12] R. S. Eisinger *et al.*, "Parkinsonian Beta Dynamics during Rest and Movement in the Dorsal Pallidum and Subthalamic Nucleus," *The Journal of Neuroscience*, vol. 40, no. 14, pp. 2859–2867, Apr. 2020, doi: 10.1523/JNEUROSCI.2113-19.2020.

[13] B. S. Coventry, "Closed-loop optical neuromodulation of the auditory thalamocortical pathway," Doctor of Philosophy, Purdue University, West Lafayette, 2021. doi: 10.25394/PGS.14831052.v1.

[14] S. Padakandla, "A Survey of Reinforcement Learning Algorithms for Dynamically Varying Environments," *ACM Comput Surv*, vol. 54, no. 6, pp. 1–25, Jul. 2021, doi: 10.1145/3459991.

[15] Grado Logan L., M. Johnson, and Netoff Theoden I, "Bayesian adaptive dual control of deep brain stimulation in a computational model of Parkinson's disease," *PLoS Comput Biol*, vol. 14, no. 12, pp. 1–23, Oct. 2018, doi: 10.1371/journal.pcbi.1006606.

[16] S. J. van Albada and P. A. Robinson, "Mean-field modeling of the basal ganglia-thalamocortical system. I Firing rates in healthy and parkinsonian states," *J Theor Biol*, vol. 257, no. 4, pp. 642–663, Apr. 2009, doi: 10.1016/j.jtbi.2008.12.018.

[17] A. Paszke *et al.*, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," Dec. 2019, [Online]. Available: http://arxiv.org/abs/1912.01703

[18] S. Fujimoto, H. van Hoof, and D. Meger, "Addressing Function Approximation Error in Actor-Critic Methods," *ArXiv*, Feb. 2018, [Online]. Available: http://arxiv.org/abs/1802.09477

[19] S. Chitlangia, A. Subramanian, S. Arthi, A. Sonwane, H. Shah, and R. Patra, "GenRL," *Github*, 2020. https://github.com/SforAiDl/genrl

[20] S. F. Cogan, S. Hara, and K. A. Ludwig, "The Safe Delivery of Electrical Currents and Neuromodulation," in *Neuromodulation*, Elsevier, 2018, pp. 83–94. doi: 10.1016/B978-0-12-805353-9.00007-3.

[21] S. F. Cogan, K. A. Ludwig, C. G. Welle, and P. Takmakov, "Tissue damage thresholds during therapeutic electrical stimulation," *J Neural Eng*, vol. 13, no. 2, p. 021001, Apr. 2016, doi: 10.1088/1741-2560/13/2/021001.