Reinforcement learning algorithms for personalized virtual brain stimulation controller

Fu-Te Wong

Introduction

Overview

Brain stimulation has been successfully applied to several neuropathologies, such as Parkinson's disease (Timmermann et al., 2015; Vitek et al., 2020), medication-refractory epilepsy (Salanova et al., 2015), treatment-resistant major depressive disorder (Scangos et al., 2021), and obsessive-compulsive disorder (Anderson & Ahmed, 2003; Franzini et al., 2010). However, in live humans, finding the most responsive site following stimulation for an individual still needs several iterations of trials, which may increase the risk to incur damage to brain tissue, especially during the procedure of deep brain stimulation. The process may be slow and have a limited probability of converging to the solution of the best stimulation algorithm. Running simulations based on computational biophysical models can provide useful information to predict how dynamic systems respond to stimulation, to gain insight into potential mechanisms for the stimulation to be effective, and to test stimulation algorithms.

Neural mass models for simulating neural dynamics and neurostimulation

The Virtual Brain (TVB) is a neuroinformatic platform with a brain simulator that integrates large-scale structural brain connectivity and neural mass models (Ritter et al., 2013). The individualized information of the brain topology and coupling can be extracted from the connectome reconstructed from diffusion tensor imaging. Brain dynamics are generated from the neural mass models, which simulate a collection of neural activity in a mean-field/region. Interaction of different regions is simulated as the neural dynamics propagate through the coupling factor. Recently developed TVB-multiscale co-simulation toolbox (Schirner et al., 2022) further provides an interface connecting neural mass model and spiking network simulators (currently Neural Simulation Technology (NEST) (Eppler et al., 2008) and Artificial Neural Network architect (ANNarchy) (Vitay et al., 2015)). Depending on the types of stimulation to be simulated, TVB supplies stimulator surrogates using temporal equations, such as linear (constant as tDCS or DBS), sinusoid/cosine (DBS), and pulse-train (rTMS), to modulate and control neural dynamics. TVB serves as a nice environ-

ment to test the implementation of focal and distributed pathological changes, identify, and explore treatment strategies (e.g., policies of a controller) to counteract those unfavorable pathological processes.

Reinforcement learning for brain stimulation control

Hypothetically, neuromodulation techniques can move the brain network dynamics between the diseased and healthy state (Meier et al., 2021; Stefanovski et al., 2019). However, a system involving neuronal interactions is complex and non-linear. While traditional optimal control techniques utilizing the linear–quadratic regulator (LQR) perform well in a linear system, modeling and controlling large-scale brain dynamics would gain more benefits from data-driven optimization approaches. Reinforcement Learning (RL) provides powerful algorithms to search for optimal control strategies that can handle systems with nonlinear dynamics and nonquadratic cost functions. The basic elements of an RL system include the *agent* (analog to stimulation controller), the *environment* (here the virtual brain simulation), and four main subelements: a *policy*, a *reward signal*, and a *value function* (Sutton & Barto, 2018). The policy defines the learning agent's response/action to the environment, given its perceived states. The reward signal defines what are the good and bad events for the agent in an immediate sense, and the value function specifies what is good in the long run.

The core of an RL agent is the policy (π), which is often calculated with a popular RL algorithm called Q-learning (Watkins & Dayan, 1992). The dynamic programming approach to compute the optimal Q value via the Value iteration algorithm, using Bellman optimality equation (Buşoniu et al., 2018), is of the form:

$$Q^*(s,a) \leftarrow R(s,a) + \gamma \sum_{s' \in \mathcal{S}} P(s'|s,a) \max_{a'} Q^*(s',a'),$$

where s is the state dynamics, a is the action, P is the transition function, R is the reward function, and γ is the discount factor. Once estimation of $Q^*(s,a)$ has converged, we would have the optimal policy:

$$\pi^*(s) = \max_a Q^*(s,a)$$

We propose that the optimal policy for doing brain stimulation can be learned by an

RL algorithm that is coupled to a biophysical brain simulator, such as TVB. One important goal of the policy is to steer the neural dynamics from the disease state to the healthy state (see Figure 1).

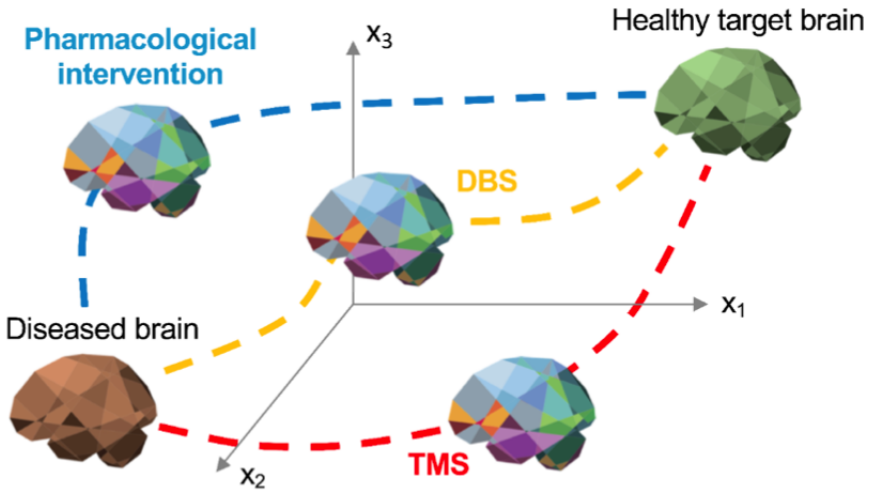

Figure 1: Schematic overview of steering brain dynamics from one state to another using The Virtual Brain. Invasive stimulation (e.g., DBS), non-invasive stimulation (e.g., TMS), and pharmacological interventions hold the potential to drive the brain dynamics from disease state to healthy state via different trajectories (Meier et al., 2021). Biophysical brain simulations allow for a comprehensive mapping of potential trajectories through this space, as well as testing and development of stimulator control algorithms, such as the RL approach developed here.

Project 1 Connecting the RL agent with TVB in biophysical simulations of tDCS and tACS effects, demonstrating the neural modulation effect controlled by the RL agent for the simple case of a single, disconnected, oscillatory brain region.

Project 2 Extension of the single-region, single-agent stimulation case in Project 1 to multi-region, multi-agent stimulation and control within a whole-brain network, corresponding to multi-focal tDCS/tACS stimulation.

Project 3 Consolidating the developments of Projects 1 and 2, develop a new Python-based open-source software library that integrates RL algorithms (such as Deep Q-Learning for discrete action space, and Actor-Critic for continued action space) and virtual stimulation methods (such as tDCS, TMS, and DBS).

Methods

Deep Q-learning

In the RL Q-Learning process, learning the Q-function is like constructing a table containing values for each combination of state and action. Representation of the state and action pair with the Q-value is, however, a difficult task and the learning process is impractical in the real world. A more efficient RL algorithm is called Deep Q-Learning (Mnih et al., 2015), which deep neural network with parameters θ as a function approximator to estimate the Q-value (i.e., $Q(s, a; \theta) \approx Q^*(s, a)$). The learning process is iterated by minimizing the following loss at each step i:

$$L_i(\theta_i) = \mathbb{E}_{s,a,r,s' \sim \rho(.)} \left[(y_i - Q(s, a; \theta_i))^2 \right],$$

$$y_i = r + \gamma \max_{a'} Q(s', a'; \theta_{i-1})$$

In the above equation, y_i is the TD (temporal difference) target, $y_i - Q$ is the TD error term, and ρ represents the behavior distribution. The distribution over transitions $\{s, a, r, s'\}$ is collected from the environment.

Fitzhugh-Nagumo model of neural dynamics

As a starting point for this new simulation control approach, we began by modulating the oscillation frequency of an individual simulated neuron. The RL agent performs one of two actions: applying a fixed-amount increase or decrease in the level of static current injection to the system. This single node was represented by a FitzHugh-Nagumo (FHN) oscillator model, which is a simplified 2D version of the Hodgkin–Huxley model that models the activation and deactivation of the dynamics of a spiking neuron, with the equation:

$$\dot{x} = \alpha \left[y + x - \frac{x^3}{3} + z \right]$$

$$\dot{y} = -\frac{1}{\alpha} \left[w^2 x - a + by \right]$$

Here, x represents membrane potential, y is the recovery variable, and z is the injected current.

Results

Non-RL Biophysical tDCS Simulations

To begin, we examined a simulation of brain network dynamics within a whole-brain model and non-RL-based tDCS stimulation, following the approach of Kunze et al. (2016). Results of this are shown in Figure 2. As indicated by the power spectrum of zoomed-in EEG channel, compared with simulation of resting state, the amplitude of power was shrunk, and the oscillation frequency slowed down during tDCS simulation.

Exploration of RL control for a network node with FHN dynamics

The whole-brain simulation results in Figure 2 provide a starting point for simulation of neural mass model-based simulation of tDCS stimulation.

Next, we began our incorporation of the RL control framework into this system. We began with a virtual single node, following the FHN dynamics described earlier. As shown in Figure 3, the RL agent can successfully learn to control the stimulation so as to steer the oscillation from one point to another in the FHN frequency space. The task was broken down into small steps, incrementally adding task difficulty. First, the goal of the agent is to keep the oscillation frequency in a fixed state with 2 control actions (i.e., increasing or decreasing injection currents) (Figure 3a). Then, we added environmental and internal perturbations (Figure 3b and Figure 3c, respectively). Environmental perturbations add random noise to the information seen by the RL agent, and can be understood as testing its sensitivity and accuracy to make correct control decisions, without any actual change to the neural system activity itself. At the next level, the RL agent is equipped with 3 actions, which are increasing, decreasing, and turning on/off the injected currents. The oscillation frequency still can be maintained under 3 conditions, i.e., no perturbations, environmental perturbations, and internal perturbations, as shown in Figure 3d, Figure 3e, and Figure 3f, respectively. Finally, we showed that the FHN controlled by the RL agent can generate oscillation frequency at any point in the legitimate frequency space (Figure 3g).

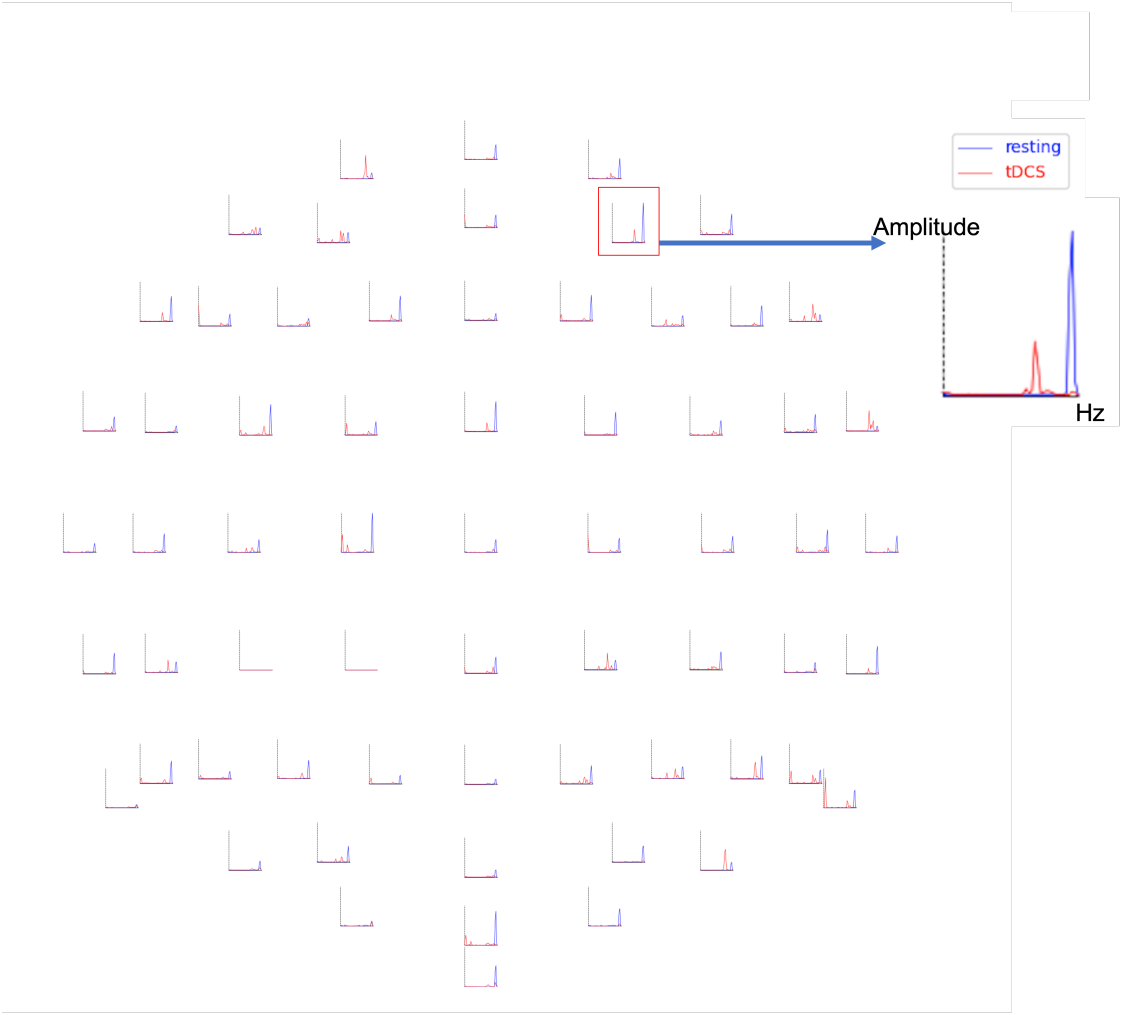

Figure 2: Simulated brain dynamics (represented with a power spectrum of 64 channels) in resting state and stimulation (tDCS) with TVB. The indicated zoom-in channel shows decreasing power amplitude and frequency.

Summary and Future Directions

In summary, we have developed a novel framework for RL stimulator control in the context of biophysical brain simulations and experiments.

The extended applications of the projects include transforming invasive stimulation into non-invasive stimulation based on simulation equivalence. A key long-term vision for this work is the idea that, based on simulation equivalence, it should be possible to find an optimal policy for non-invasive brain stimulation strategy that has a similar therapeutic effect as a given invasive brain stimulation strategy (Figure 1). For example, a patient with the treatment-resistant major depressive disorder may be subjected to the treatment of deep brain stimulation (Scangos et al., 2021). Considering the results of Scangos et al., imag-

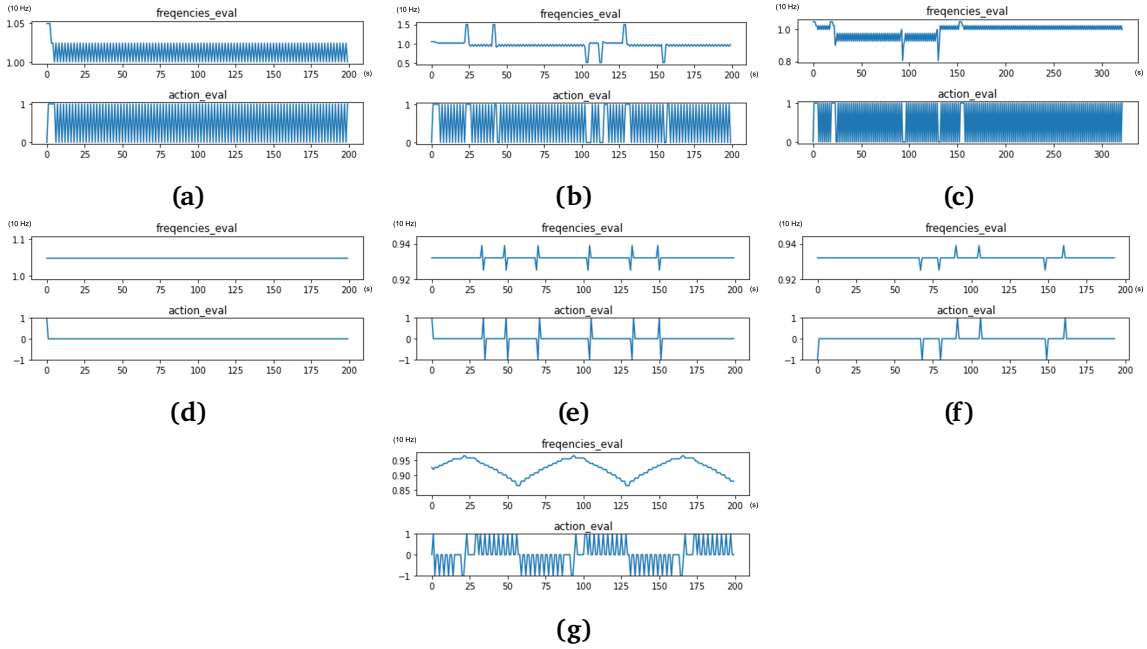

Figure 3: Oscillation frequency controlled by RL agent. The first row showed a fixed state of oscillation frequency has been maintained by the RL agent with 2 control actions in the 3 conditions, where no perturbations applied in (a), environmental perturbations in (b), and internal perturbations in (c). In the second row, the RL agent was equipped with 3 control actions and successfully maintained a fixed state of oscillation frequency in the 3 conditions same as in the first row (environmental perturbations in (e) and internal perturbations in (f)). In (g), the oscillation frequency under control of the RL agent 'walks' through the legitimate frequency range.

ine that the simulated activity provides that the gamma-frequency power in Amygdala is a good biomarker signal for depressive symptoms, and the ventral capsule or ventral striatum is good as the stimulating site for the closed-loop system to modulate the brain network dynamics to move from a disease state to approximately close to a healthy state (S^*). S^* then can serve as a target for the RL agent to find an optimal non-invasive stimulation policy. To get a better result of the networked system control, an enhanced RL framework with multi-agents (Chu et al., 2020) can be implemented for the multi-focal brain stimulation scenario, as for example is now possible with new tACS, tDCS, TMS, and DBS systems (e.g., Jiang et al., 2013; Ruffini et al., 2018; Vitek et al., 2020).

In addition, the optimal RL algorithm allow us to explore the best sensing bio-marker, stimulation site, and stimulation parameters. The stimulation solution found by the RL agent can provide additional insights compared with the empirical closed-loop solution, for instance, the research from Scangos et al. (2021) aforementioned. Finally, an advantage of applying neural networks is that the features learned in one training environment could have predictive power in another environment. By exploiting the transfer learning advantage, we

will explore the possibility that a trained RL agent can adapt to other subjects more quickly than retraining the network on those subjects from scratch.

References

Anderson, D., & Ahmed, A. (2003). Treatment of patients with intractable obsessive-compulsive disorder with anterior capsular stimulation. Case report. *Journal of Neurosurgery*, *98*(5), 1104–1108. https://doi.org/10.3171/jns.2003.98.5.1104

Buşoniu, L., de Bruin, T., Tolić, D., Kober, J., & Palunko, I. (2018). Reinforcement learning for control: Performance, stability, and deep approximators. *Annual Reviews in Control*, *46*, 8–28. https://doi.org/10.1016/j.arcontrol.2018.09.005

Chu, T., Chinchali, S., & Katti, S. (2020, April 23). *Multi-agent Reinforcement Learning for Networked System Control*. Retrieved April 7, 2022, from http://arxiv.org/abs/2004.01339

Eppler, J. M., Helias, M., Muller, E., Diesmann, M., & Gewaltig, M.-O. (2008). PyNEST: A Convenient Interface to the NEST Simulator. *Frontiers in Neuroinformatics*, *2*, 12. https://doi.org/10.3389/neuro.11.012.2008

Franzini, A., Messina, G., Gambini, O., Muffatti, R., Scarone, S., Cordella, R., & Broggi, G. (2010). Deep-brain stimulation of the nucleus accumbens in obsessive compulsive disorder: Clinical, surgical and electrophysiological considerations in two consecutive patients. *Neurological Sciences: Official Journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, *31*(3), 353–359. https://doi.org/10.1007/s10072-009-0214-8

Jiang, R., Jansen, B. H., Sheth, B. R., & Chen, J. (2013). Dynamic multi-channel TMS with reconfigurable coil. *IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society*, *21*(3), 370–375. https://doi.org/10.1109/TNSRE.2012.2226914

Kunze, T., Hunold, A., Haueisen, J., Jirsa, V., & Spiegler, A. (2016). Transcranial direct current stimulation changes resting state functional connectivity: A large-scale brain network modeling study. *NeuroImage*, *140*, 174–187. https://doi.org/10.1016/j.neuroimage.2016.02.015

Meier, J. M., Perdikis, D., Blickensdörfer, A., Stefanovski, L., Liu, Q., Maith, O., Dinkelbach, H. Ü., Baladron, J., Hamker, F. H., & Ritter, P. (2021). Virtual deep brain stimulation: Multiscale co-simulation of a spiking basal ganglia model and a whole-brain mean-

field model with The Virtual Brain, 2021.05.05.442704. https://doi.org/10.1101/2021.05.05.442704

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M. A., Fidjeland, A., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*. https://doi.org/10.1038/nature14236

Ritter, P., Schirner, M., McIntosh, A. R., & Jirsa, V. K. (2013). The virtual brain integrates computational modeling and multimodal neuroimaging. *Brain Connectivity*, *3*(2), 121–145. https://doi.org/10.1089/brain.2012.0120

Ruffini, G., Wendling, F., Sanchez-Todo, R., & Santarnecchi, E. (2018). Targeting brain networks with multichannel transcranial current stimulation (tCS). *Current Opinion in Biomedical Engineering*, *8*, 70–77. https://doi.org/10.1016/j.cobme.2018.11.001

Salanova, V., Witt, T., Worth, R., Henry, T. R., Gross, R. E., Nazzaro, J. M., Labar, D., Sperling, M. R., Sharan, A., Sandok, E., Handforth, A., Stern, J. M., Chung, S., Henderson, J. M., French, J., Baltuch, G., Rosenfeld, W. E., Garcia, P., Barbaro, N. M., … Group, F. t. S. S. (2015). Long-term efficacy and safety of thalamic stimulation for drug-resistant partial epilepsy. *Neurology*, *84*(10), 1017–1025. https://doi.org/10.1212/WNL.0000000000001334

Scangos, K. W., Khambhati, A. N., Daly, P. M., Makhoul, G. S., Sugrue, L. P., Zamanian, H., Liu, T. X., Rao, V. R., Sellers, K. K., Dawes, H. E., Starr, P. A., Krystal, A. D., & Chang, E. F. (2021). Closed-loop neuromodulation in an individual with treatment-resistant depression. *Nature Medicine*, *27*(10), 1696–1700. https://doi.org/10.1038/s41591-021-01480-w

Schirner, M., Domide, L., Perdikis, D., Triebkorn, P., Stefanovski, L., Pai, R., Prodan, P., Valean, B., Palmer, J., Langford, C., Blickensdörfer, A., van der Vlag, M., Diaz-Pier, S., Peyser, A., Klijn, W., Pleiter, D., Nahm, A., Schmid, O., Woodman, M., … Ritter, P. (2022). Brain simulation as a cloud service: The Virtual Brain on EBRAINS. *NeuroImage*, *251*, 118973. https://doi.org/10.1016/j.neuroimage.2022.118973

Stefanovski, L., Triebkorn, P., Spiegler, A., Diaz-Cortes, M.-A., Solodkin, A., Jirsa, V., McIntosh, A. R., Ritter, P., & Alzheimer's Disease Neuroimaging Initiative. (2019). Linking Molecular Pathways and Large-Scale Computational Modeling to Assess Candi-

date Disease Mechanisms and Pharmacodynamics in Alzheimer's Disease. *Frontiers in Computational Neuroscience*, *13*, 54. https://doi.org/10.3389/fncom.2019.00054

Sutton, R. S., & Barto, A. G. (2018, November 13). *Reinforcement Learning: An Introduction* (2nd ed.). A Bradford Book.

Timmermann, L., Jain, R., Chen, L., Maarouf, M., Barbe, M. T., Allert, N., Brücke, T., Kaiser, I., Beirer, S., Sejio, F., Suarez, E., Lozano, B., Haegelen, C., Vérin, M., Porta, M., Servello, D., Gill, S., Whone, A., Van Dyck, N., & Alesch, F. (2015). Multiple-source current steering in subthalamic nucleus deep brain stimulation for Parkinson's disease (the VANTAGE study): A non-randomised, prospective, multicentre, open-label study. *The Lancet. Neurology*, *14*(7), 693–701. https://doi.org/10.1016/S1474-4422(15)00087-3

Vitay, J., Dinkelbach, H. Ü., & Hamker, F. H. (2015). ANNarchy: A code generation approach to neural simulations on parallel hardware. *Frontiers in Neuroinformatics*, *9*, 19. https://doi.org/10.3389/fninf.2015.00019

Vitek, J. L., Jain, R., Chen, L., Tröster, A. I., Schrock, L. E., House, P. A., Giroux, M. L., Hebb, A. O., Farris, S. M., Whiting, D. M., Leichliter, T. A., Ostrem, J. L., San Luciano, M., Galifianakis, N., Verhagen Metman, L., Sani, S., Karl, J. A., Siddiqui, M. S., Tatter, S. B., … Starr, P. A. (2020). Subthalamic nucleus deep brain stimulation with a multiple independent constant current-controlled device in Parkinson's disease (INTREPID): A multicentre, double-blind, randomised, sham-controlled study. *The Lancet. Neurology*, *19*(6), 491–501. https://doi.org/10.1016/S1474-4422(20)30108-3

Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. *Machine Learning*, 279–292.