Imitation learning of dexterous hand control uncovers muscle-level representations in primate sensorimotor cortex

Alessandro Marin Vargas[*], **Adriana Perez Rotondo, Alberto Silvio Chiappa, Mackenzie Weygandt Mathis**, **Alexander Mathis**[†]
Brain Mind Institute and Neuro-X Institute, EPFL, Lausanne, Switzerland

Abstract

Dexterous grasping requires the integration of proprioceptive feedback with predictive motor commands. Yet, how cortical circuits combine afferent feedback with efference signals during grasp remains poorly understood. Here we combine deep reinforcement learning, biomechanics, and neural recordings to build a closed-loop, muscle-level controller of primate grasping. A neural network policy, trained via imitation learning on a 39-muscle musculoskeletal hand, reproduces naturalistic pre-contact shaping and develops internal dynamics that explain single-neuron activity in primary motor (M1) and somatosensory (S1) cortex. Three principles emerged. First, muscle-based control aligns more closely with cortical dynamics than joint-based control, even when the latter achieves higher tracking accuracy. Second, imitation learning drives the emergence of internal representations that capture both trial-to-trial variability and object-specific tuning. Third, trajectory embeddings learned within the modular policy can be decoded from M1 to directly drive the controller from only tens of neurons. Together, these findings establish a stimulus-computable framework that reveals integrated, muscle-centric and goal-latent codes in S1/M1 while opening a novel route for creating brain-body models.

1 Introduction

Dexterous hand use requires the integration of proprioceptive feedback with predictive motor commands [1–3], engaging both motor (M1) and somatosensory (S1) cortices during grasping [4, 5, 2]. Proprioception reflects the convergence of afferent feedback from muscle spindles and Golgi tendon organs with reafferent signals driven by motor commands [1, 6]. During grasp pre-shaping, this integration is especially critical: the hand must be configured in anticipation of object contact, requiring dynamic feedback loops between S1 and M1 as well as parietal and premotor areas [7, 8, 4]. Yet, how cortical circuits combine afferent and efferent signals to support closed-loop control remains unclear.

Most prior computational models capture only fragments of this loop. Joint-level controllers provide insight into kinematic planning [9, 10], while task- or data-driven neural networks have modeled sensory or motor encoding separately [11, 12, 6]. However, such approaches often omit the bidirectional feedback central to sensorimotor integration (but see DeWolf et al. [13]). In parallel, biomechanical simulators and reinforcement learning are increasingly enabling the study of dexterous hand control [14, 15], yet learned policies often do not move in a biologically plausible way [16, 15, 17]. In

[*]Current address: Neural Prosthetics Translational Lab, Stanford University; alessmv@stanford.edu
[†]Corresponding author: alexander.mathis@epfl.ch

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Foundation Models for the Brain and Body.

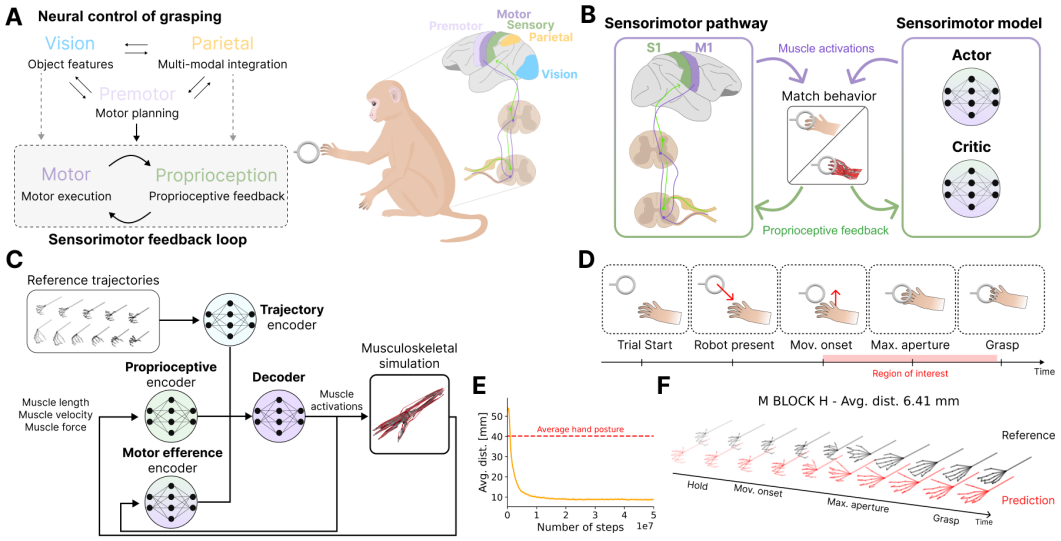

Figure 1: **Closed-loop imitation-learning controller for investigating sensorimotor processing.** **A.** Neural control of grasping schematic highlighting the sensorimotor feedback loop linking vision, parietal, premotor, motor, and proprioceptive pathways. **B.** Conceptual link between biological sensorimotor pathways (S1/M1 integration) and the model's closed-loop controller. **C.** Architecture: reference trajectories feed a trajectory encoder; proprioceptive and motor-efference streams are encoded separately; a decoder integrates all streams to output muscle activations to the musculoskeletal simulation. **D.** Trial timeline from start to grasp. Red area marks the pre-contact region of interest. **E.** Average egocentric distance versus training steps; red dashed line indicates a control of maintaining the average hand posture. **F.** Example trial for object *M BLOCK H*: reference vs prediction hand postures across phases.

contrast, imitation learning enables learning control policies that reproduce the kinematic structure of natural movements [18, 19].

Here, we develop imitation-learning controllers for a 39-muscle musculoskeletal hand [20] trained on complex pre-shaping trajectories [4]. These controllers need to integrate continuous proprioceptive feedback and output muscle activations. We demonstrate that their internal states align with single-neuron activity in M1 and S1, that optimization induces representations capturing both variability and object-specific tuning, and that trajectory embeddings decoded from cortex can directly drive the policy. This closed-loop, stimulus-computable framework bridges biomechanics, reinforcement learning, and neural recordings, providing a platform to uncover principles of cortical coding and to inform the design of brain-machine interfaces.

2 A closed-loop imitation-learning framework for sensorimotor control

Primate grasping relies on a feedback loop that integrates vision, parietal planning, premotor coordination, and sensorimotor integration in M1 and S1 [7, 8, 4] (Fig. 1A). To model this process in silico, we developed a framework combining a musculoskeletal hand model with imitation learning (Fig. 1B). Imitation learning allows controllers to mimic expert demonstrations [21, 18], producing biologically realistic kinematics that are difficult to obtain with reinforcement learning alone. Our design incorporates three input streams: trajectory goals (parietal/premotor-like from visual information), proprioceptive feedback (afferent signals to S1/M1), and motor efference. These streams converge in a decoder (policy) that outputs muscle activations, mirroring the integration of predictive goals with afferent and efferent feedback thought to occur in sensorimotor cortex (Fig. 1C).

To provide the demonstrations, we used a pose-tracking dataset of macaques grasping 35 objects with simultaneous recordings in M1 and S1 [4]. We focused on the pre-contact period, from movement onset until object touch, when proprioceptive and motor signals shape the hand in anticipation of grasp (Fig. 1D). This choice mitigates the influence of touch.

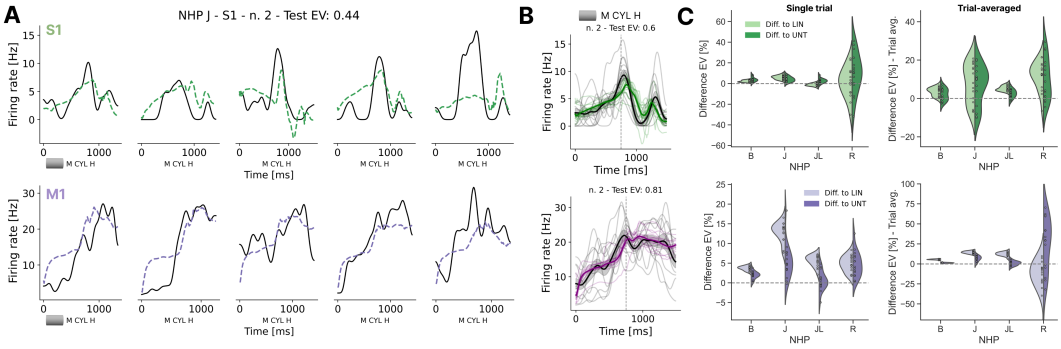

Figure 2: **Trained network representations explain S1/M1 activity at single-trial and trial-averaged levels. A.** Single-trial predictions on held-out OOD trials for example units in S1 (top) and M1 (bottom). Black: recorded firing rates; dashed: predictions from the decoder (integration) layer of the value network. **B.** Trial-averaged predictions for the same example unit and object of **A.**. Thick black: trial-averaged firing rate; colored line: trial-averaged prediction; thin lines: individual trials; shaded area: s.e.m. **C.** Distribution of explained variance gain (trained minus linear - trained minus untrained) across 15 networks; violins show distributions across models.

Policies were trained with Proximal Policy Optimization (PPO) [22] in an actor-critic setup, where the actor generated muscle activations and the critic estimated expected returns. The reward combined penalties on joint-angle and keypoint error with a cost on muscle activation, encouraging accurate yet smooth, energy-efficient movements. This reward structure is critical for producing biologically realistic dynamics, as animals minimize both error and effort. Training converged after 20 million steps, at which point the controller achieved an average egocentric keypoint error of 9 mm; for comparison that's about 40% of the pinky tip segment (distal phalanx) length (Fig. 1E).

Simulated postures closely matched experimental pre-shaping across objects, and trial examples demonstrated accurate reproduction of movement phases, including hold, opening, and shaping (Fig. 1F). Together, these results demonstrate that imitation learning yields high-fidelity closed-loop controllers that model the sensorimotor feedback loop, which can serve as computational proxies for testing hypotheses.

3 Behavior-aligned network dynamics predict cortical activity

We next asked whether the internal representations learned by the controller account for neural dynamics in sensorimotor cortex. Predictions were compared against two baselines: (*i*) linear regressors trained directly on muscle kinematics, and (*ii*) untrained networks with the same architecture but initialized-random weights. For each layer of the trained networks, we extracted activations during the pre-shaping window, reduced them to 78 principal components (to match the dimensionality of linear baselines), and used these features in ridge regression to predict single-neuron firing rates in M1 and S1.

Example neurons dynamics illustrate two complementary results. First, internal network representations can capture trial-to-trial variability in firing rates, aligning with fluctuations observed on individual grasps. Second, the same states reproduce condition-averaged dynamics, showing robust object-specific modulation across repetitions (Fig. 2A-B). This dual alignment is important: whereas prior models typically captured condition-averaged neural trajectories without modeling single-trial variability, imitation-learning trained networks account for both levels of neural dynamics.

To quantify performance, we averaged EV across all neurons for each trained network (picking for each neuron the best layer on the train set) and compared this to the baseline models. Trained controllers consistently outperformed both linear and untrained baselines for both M1 and S1 (Fig. 2C). Gains were observed whether predictions were evaluated on single-trial firing rates or on condition-averaged responses, showing that learned representations capture both fast fluctuations and stable task-related modulations.

In summary, these results demonstrate that optimization for naturalistic grasping yields internal states that better align with cortical dynamics than muscle kinematic regressors or untrained architectures.

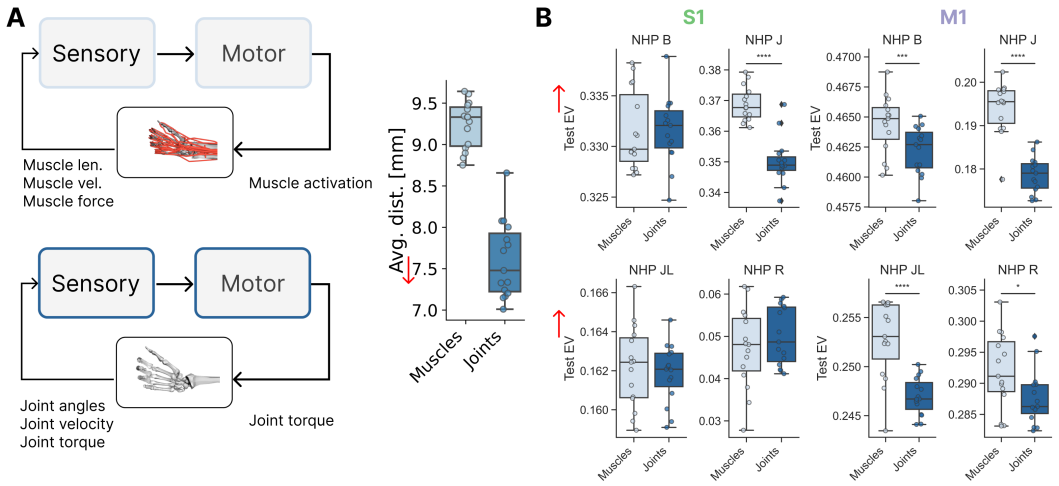

Figure 3: **Muscle-based representation outperform joint-based ones. A.** Control diagram of muscle-based (top) and joint-based (bottom) control policies. **B.** Task performance distributions show lower imitation errors for joint control policies. **C.** Neural explainability distributions for muscle- and joint-based control policies, with muscle control significantly outperforming joint control. Each dot represents one network. Significance assessed with Mann-Whitney tests.

They suggest that M1 and S1 encode feedback-driven sensorimotor representations, rather than simply reflecting external kinematics, that can be inferred from closed-loop imitation learning policies.

4 Muscle-based control develops more brain-like representations

We next asked how control modality, muscle- versus joint-based, affects neural alignment. Joint-based controllers, which output torques, achieved lower imitation errors and thus tracked reference trajectories with higher fidelity (Fig. 3A-B). This was expected, as joint control directly manipulates fewer degrees of freedom and bypasses the nonlinear dynamics of muscles. However, when comparing neural predictions, the opposite pattern emerged: muscle-based controllers consistently yielded higher explained variance in both M1 and S1 across NHPs (Fig. 3C). Thus, while joint-based control was behaviorally more accurate, muscle-based control produced internal states that aligned more closely with cortical activity.

This dissociation parallels neurophysiological findings that cortical activity correlates more strongly with muscle-level variables than with joint kinematics [23, 24, 13]. Our results extend this evidence by showing that when controllers are optimized in closed loop, muscle-level policies give rise to neural representations that better resemble those observed in sensorimotor cortex (vs. torque-based policies).

5 Brain-controlled policy

So far, we have shown that internal representations of the learned controller align with neural activity, establishing the model as a powerful *encoding model* of S1/M1. A natural next question is whether this link can be turned around: can cortical activity be used to *drive* the controller to control the musculoskeletal model? To test this, we trained linear decoders that mapped S1/M1 population activity either to joint angles (fed into the trajectory encoder) or directly to the network's latent trajectory embedding (Fig. 4A-B).

Both joint angles and trajectory embeddings could be decoded from neural populations, with M1 consistently outperforming S1 (Fig. 4C). Decoded signals, when fed into the controller, gave rise to brain-controlled trajectories that closely tracked reference grasps, with errors of 20 mm for M1 and 25 mm for S1. Notably, feeding the latent goal embedding (decoded from neural data) into the trained policy yielded performance comparable to directly decoding joint angles, with M1 having more information than S1 (Fig. 4D). Good tracking performance could be achieved with relatively

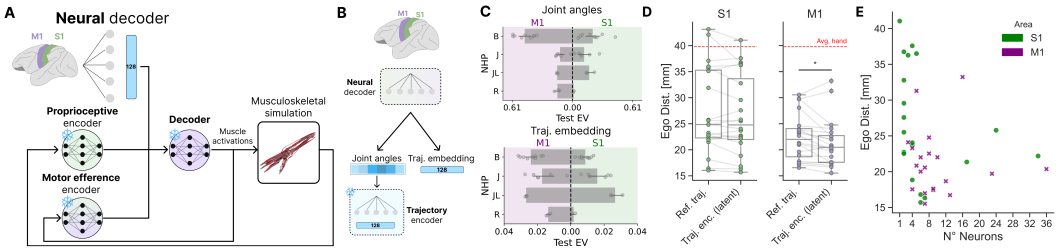

Figure 4: **Brain-Controlled Policy. A.** Diagram showing how the neural network policy can be controlled by decoding the "trajectory" signal directly from neural activity in sensory or motor brain areas. **B.** Diagram illustrating the two neural decoder variants. Neural activity can be used to decode joint angles for the trajectory encoder or the trajectory embedding of the policy. **C.** Distribution of test EV for decoding joint angles and trajectory embeddings from sensory and motor brain areas for each NHP. **D.** Imitation performance between the reference trajectory and the brain-controlled trajectory for all NHPs (each dot represents a single recording session). The ref. traj. decoding represent the performance achieved by directly decoding joint angles from S1 (green) and M1 (purple) and setting the musculoskeletal model to those positions. The red dashed line shows the error when the hand posture is fixed as the average pose across all time points and trajectories. **E.** Decoding performance as a function of neuron count, separated by brain area (S1: green, M1: purple).

few neurons (tens per session; Fig. 4E), showing that compact neural subpopulations already contain sufficient information to drive high-dimensional musculoskeletal control. This suggests that M1 is a high-level controller.

The efficiency highlights the advantage of embedding-based decoding: by leveraging the policy's internal goal representation, it provides a robust, low-bandwidth interface. Importantly, the differential performance of S1 and M1 decoders is consistent with their functional roles, M1 being closer to motor control, and S1 representing sensory feedback and integrating it with efference [6]. Together, these results demonstrate a practical route from neural recordings to embodied control, strengthening the link between cortical representations and closed-loop biomechanical policies.

6 Conclusion

This work demonstrates how closed-loop imitation learning can serve as a powerful framework for understanding the neural dynamics during primate grasping. By integrating biomechanics with neural recordings, we show that muscle-level controllers trained by imitation learning infer internal dynamics that align with single-neuron activity in M1 and S1. Importantly, these states capture both trial-level variability and condition-averaged tuning, extend evidence for muscle-centric encoding in cortex, and can be harnessed to drive the controller directly from cortical activity.

Our results suggest that M1 and S1 encode signals necessary to drive muscle activations rather than forces, and that both areas integrate afferent and efferent signals to estimate body state [5, 6, 13]. By reproducing these principles in a behavior-constrained model, we provide computational support for the view that sensorimotor cortex acts as a feedback controller. This connection highlights how imitation-learned controllers can serve as mechanistic hypotheses about cortical computation rather than only as behavioral simulators.

Beyond providing a stimulus-computable model of sensorimotor integration, our approach unifies encoding and decoding within the same framework [25], offering both mechanistic insights and a functional platform for brain-machine interfaces. Future extensions could incorporate hierarchical control or vision-conditioned inputs. More broadly, imitation-learned controllers may provide a foundation for models of the brain and body that capture not only the statistics of natural movement, but also the computations that underlie adaptive sensorimotor control.

Acknowledgments: We are indebted to Sliman Bensmaia for generously sharing the primate data as well as James Goodman for discussions. We thank members of the Mathis Group for Computational Neuroscience & AI (EPFL) for their feedback throughout the project. We thank Mark Churchland for valuable discussions. This work was funded by Swiss SNF grant (310030_212516) and Swiss Government Excellence Scholarship (A.M.V.).

References

[1] Uwe Proske and Simon C Gandevia. The proprioceptive senses: their roles in signaling body shape, body position and movement, and muscle force. *Physiological reviews*, 2012.

[2] Anton R Sobinov and Sliman J Bensmaia. The neural mechanisms of manual dexterity. *Nature Reviews Neuroscience*, 22(12):741–757, 2021.

[3] Michael Dimitriou. Human muscle spindles are wired to function as controllable signal-processing devices. *Elife*, 11:e78091, 2022.

[4] James M Goodman, Gregg A Tabot, Alex S Lee, Aneesha K Suresh, Alexander T Rajan, Nicholas G Hatsopoulos, and Sliman Bensmaia. Postural representations of the hand in the primate sensorimotor cortex. *Neuron*, 104(5):1000–1009, 2019.

[5] Tatsuya Umeda, Tadashi Isa, and Yukio Nishimura. The somatosensory cortex receives information about motor output. *Science advances*, 5(7):eaaw5388, 2019.

[6] Alessandro Marin Vargas, Axel Bisi, Alberto S Chiappa, Chris Versteeg, Lee E Miller, and Alexander Mathis. Task-driven neural network models predict neural dynamics of proprioception. *Cell*, 187(7): 1745–1761, 2024.

[7] Thomas Brochier and Maria Alessandra Umiltà. Cortical control of grasp in non-human primates. *Current opinion in neurobiology*, 17(6):637–643, 2007.

[8] Stefan Schaffelhofer and Hansjörg Scherberger. Object vision to hand action in macaque parietal, premotor, and motor cortices. *elife*, 5:e15278, 2016.

[9] Harsha Gurnani, Weixuan Liu, and Bingni W Brunton. Feedback control of recurrent dynamics constrains learning timescales during motor adaptation. *bioRxiv*, pages 2024–05, 2024.

[10] Olivier Codol, Jonathan A Michaels, Mehrdad Kashefi, J Andrew Pruszynski, and Paul L Gribble. Motornet, a python toolbox for controlling differentiable biomechanical effectors with artificial neural networks. *Elife*, 12:RP88591, 2024.

[11] David Sussillo, Mark M Churchland, Matthew T Kaufman, and Krishna V Shenoy. A neural network that finds a naturalistic solution for the production of muscle activity. *Nature neuroscience*, 18(7):1025–1033, 2015.

[12] Abigail A Russo, Sean R Bittner, Sean M Perkins, Jeffrey S Seely, Brian M London, Antonio H Lara, Andrew Miri, Najja J Marshall, Adam Kohn, Thomas M Jessell, et al. Motor cortex embeds muscle-like commands in an untangled population response. *Neuron*, 97(4):953–966, 2018.

[13] Travis DeWolf, Steffen Schneider, Paul Soubiran, Adrian Roggenbach, and Mackenzie Mathis. Neuromusculoskeletal modeling reveals muscle-level neural dynamics of adaptive learning in sensorimotor cortex. *bioRxiv*, pages 2024–09, 2024.

[14] Vittorio Caggiano, Guillaume Durandau, Huwawei Wang, Alberto Chiappa, Alexander Mathis, Pablo Tano, Nisheet Patel, Alexandre Pouget, Pierre Schumacher, Georg Martius, et al. Myochallenge 2022: Learning contact-rich manipulation using a musculoskeletal hand. In *NeurIPS 2022 Competition Track*, pages 233–250. Proceedings of Machine Learning Research, 2022.

[15] Alberto Silvio Chiappa, Pablo Tano, Nisheet Patel, Abigaïl Ingster, Alexandre Pouget, and Alexander Mathis. Acquiring musculoskeletal skills with curriculum-based reinforcement learning. *Neuron*, 112(23): 3969–3983.e5, 2024. ISSN 0896-6273.

[16] Xue Bin Peng, Erwin Coumans, Tingnan Zhang, Tsang-Wei Lee, Jie Tan, and Sergey Levine. Learning agile robotic locomotion skills by imitating animals. *arXiv preprint arXiv:2004.00784*, 2020.

[17] Pierre Schumacher, Thomas Geijtenbeek, Vittorio Caggiano, Vikash Kumar, Syn Schmitt, Georg Martius, and Daniel FB Haeufle. Natural and robust walking using reinforcement learning without demonstrations in high-dimensional musculoskeletal models. *arXiv preprint arXiv:2309.02976*, 2023.

[18] Xue Bin Peng, Pieter Abbeel, Sergey Levine, and Michiel Van de Panne. Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM (TOG)*, 37(4):1–14, 2018.

[19] Merkourios Simos, Alberto Silvio Chiappa, and Alexander Mathis. Reinforcement learning-based motion imitation for physiologically plausible musculoskeletal motor control. *arXiv preprint arXiv:2503.14637*, 2025.

[20] Vittorio Caggiano, Huawei Wang, Guillaume Durandau, Massimo Sartori, and Vikash Kumar. Myosuite: A contact-rich simulation suite for musculoskeletal motor control. In *Learning for Dynamics and Control Conference*, pages 492–507. PMLR, 2022.

[21] Stefan Schaal. Is imitation learning the route to humanoid robots? *Trends in cognitive sciences*, 3(6): 233–242, 1999.

[22] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[23] Edward V Evarts. Relation of pyramidal tract activity to force exerted during voluntary movement. *Journal of neurophysiology*, 31(1):14–27, 1968.

[24] Paul D Cheney and Eberhard E Fetz. Functional classes of primate corticomotoneuronal cells and their relation to active force. *Journal of Neurophysiology*, 44(4):773–791, 1980.

[25] Mackenzie Weygandt Mathis, Adriana Perez Rotondo, Edward F Chang, Andreas S Tolias, and Alexander Mathis. Decoding the brain: From neural representations to mechanistic models. *Cell*, 187(21):5814–5832, 2024.