Improving Scalability in Systems Neuroscience

Zhe Sage Chen[1,2,5,*] and Bijan Pesaran[2,3,4*]

[1]Department of Psychiatry, Department of Neuroscience and Physiology, New York University School of Medicine, New York, NY 10016, USA
[2]Neuroscience Institute, NYU School of Medicine, New York, NY 10016, USA
[3]Center for Neural Science, New York University, New York, NY 10003, USA
[4]Department of Neurology, New York University School of Medicine, New York, NY 10016, USA
[5]Lead Contact

* Correspondence: zhe.chen@nyulangone.org (Z.S.C.), bijan@nyu.edu (B.P.)

In Brief

Advances in neurotechnology for exponential growth of neural data present both opportunities and challenges in systems neuroscience. Chen and Pesaran argue that active, adaptive closed-loop experiments offer a solution to improve scalability for knowledge discovery and overcome the dimensionality bottleneck.

Summary

Emerging technologies to acquire data at increasingly greater scales promise to transform discovery in systems neuroscience. However, current exponential growth in the scale of data acquisition is a double-edged sword. Scaling up data acquisition can speed up the cycle of discovery, but can also misinterpret the results or possibly slow down the cycle due to challenges presented by the curse of high-dimensional data. Active, adaptive, closed-loop experimental paradigms employ hardware and algorithms optimized to enable time-critical computation to provide feedback that interprets the observations and tests hypotheses, to actively update the stimulus or stimulation parameters. In this perspective, we review important concepts of active and adaptive experiments, and discuss how selectively constraining the dimensionality and optimizing strategies at different stages of discovery loop can help mitigate the curse of high-dimensional data. Active and adaptive closed-loop experimental paradigms can speed up discovery in spite of an exponentially increasing data scale, offering a roadmap to timely and iterative hypothesis revision and discovery in an era of exponential growth in neuroscience.

Introduction

Systems neuroscience faces the daunting challenge of understanding brain networks of complex and poorly understood topologies. Over the last decade, however, a technological revolution in neuroscience has enabled tremendous growth in volume and quality of scientific data. Experimental tools allowing measurements of large-scale *in vivo* neuronal population activity at high-resolution using multiple (e.g. electrical, optical, magnetic) modalities and across multiple brain regions are becoming widespread. Alongside advances in instrumentation, methods to efficiently preprocess, characterize and fit models to large-scale neuroscientific data are also being developed (Stevenson and Kording, 2011; Paninski and Cunningham, 2018). How should we use large-scale neurotechnologies to understand brain network mechanisms? This Perspective reviews the state-of-the-art in neurotechnology through the lens of the curse of high-dimensional neural data analysis. The curse of high-dimensional data arises from the consequences of scaling data dimensionality (Vershynin, 2018; Wainwright, 2019) and leads to exponentially increasing computation time. We propose that a new generalization of closed-loop experiments, which we term active, adaptive closed-loop (AACL) experiments, will be important to successfully mitigating the scalability in neuroscience, especially for discovering brain network mechanisms.

Discovery is a process of obtaining new knowledge based on multiple steps of verification. In systems neuroscience, knowledge can be expressed in multiple forms, ranging from the understanding of the animal's behavior, the effectiveness of experimental stimuli, the regularity of the neural response, to the causal link between neural codes to behavior. Discovery can revise the existing theories or hypotheses, or even create a paradigm shift in the research practice. The standard discovery cycle features data acquisition, analysis and interpretation to test hypotheses and update concepts, which is fundamental to scientific progress (*Conceive-Acquire-Analyze-Test-Revise*; Figure 1A**)**. However, the concept of "loop" is underemphasized in discovery cycle for two important reasons: first, there is no nested internal loop containing feedback; second, there is no strict time constraint between steps. Large-scale neuroscience presents a challenge for the cycle of discovery. Counterintuitively, the growth of neuroscience data (in dimensionality, volume and size) can slow and even impede the cycle of discovery. High-dimensionality of data can overwhelm the analysis because of the bandwidth bottleneck and efforts to address the bottleneck can effectively decrease statistical power. The loss of statistical power is because unless simplification (such as averaging or subsampling) is assumed, the statistical estimate of model variables may become increasingly biased, which may further misinterpret the results. The alternative involves increasing the recording duration to account for the increased dimensionality (thereby increasing the acquisition time), or processing all collected data sequentially without time constraints (thereby increasing the analysis time), either of which can slow down the hypothesis testing or revision and progress toward the scientific goals. This is particularly disruptive in neuroscience because the nervous system is dynamic and plastic. We cannot revise and test important classes of hypotheses, such as specificity and causality, until we verify the steps to complete an iteration. For instance, if the tested hypothesis is circuit A is responsible for behavior B, even we observe that a neurostimulation (as perturbation of "A") disrupts a task ("B"), we still need to verify several important questions to fully test the hypothesis: first, whether neurons collected from A show significantly changed patterns that correlate with the behavior; second, what types of neurons and how many of them contribute to such changes; third, how does the specific stimulation input (e.g., timing and intensity) causally alter the task behavior. Much like how we cannot step into the same river twice, by the time the original hypothesis has been tested the context within which the results were obtained cannot be revisited. This substantially limits, and can potentially even make impossible, our ability to test alternative hypotheses.

AACL experiments are different from open-loop or passive closed-loop experiments in that not only strict time constraints are imposed on every step (e.g., acquisition, analysis, stimulation), but also an optimization procedure is employed in some or even all steps based on active feedback (Figure 2A). AACL feedback signals can be iteratively used for many purposes: to optimize experimental stimuli and other experimental parameters (Walker et al.., 2019; Ponce et al., 2019), to select neural channels for recording and/or telemetry (Choi et al., 2020), to perform decoder adaptation (Dangi et al., 2013), to optimize stimulation parameters (Tafazoli et al., 2020), and to optimize objective functions and other aspects of control policies (Bolus et al., 2018). In principle, each step of an AACL experiment may contain nested inner loops. In contrast, passive closed-loop experiments employ a fixed policy. The experimental stimulus is predetermined and cannot adapt. The decoder is fixed. Control and stimulation parameters are predetermined.

In this Perspective, we discuss how *jointly* scaling up data acquisition and data analysis in an active and adaptive manner can speed up the cycle and enable AACL experiments. We first review the scalability in neurotechnology and instrumentation, highlighting how multiple trends increase the size, volume and dimensionality of experimental observations. We then point to our main thesis - that scaling is a double-edged sword: it can speed up the cycle of discovery in systems neuroscience, but involves defining and following a sequence of predetermined experimental steps. In considering the cycle of discovery, each step in an AACL experiment has a respective "scaling-speed limit". The overall rate of discovery is limited by the slowest factor in each step (Figure 1D), and can be exacerbated by large data volumes and high dimensionality, which can overwhelm our capacity for analysis and interpretation. Consequently, the lack of scalability of data analytic tools introduces barriers to scientific discovery. Finally, we discuss the features and limitations of AACL experiments, and review strategies to speed up data analysis.

Size, Depth, and Multi-site in Neurophysiological and Imaging Recordings

Neurotechnologies employ a range of physical modalities spanning sound, light, electricity, magnetism as well as multimodal mechanisms such as optoacoustics/photoacoustics and magnetoacoustics (Marblestone et al., 2013; Gottschalk et al., 2019). Modern neural interfaces that can record and/or stimulate the nervous system are dramatically expanding the number of neural signal channels that can be monitored and manipulated. When the word "scale" is used, we refer to the dimensionality/size/volume of neural signals, which should not be confused with the spatial or temporal granularity at which data are acquired.

Accessing brain tissue at single cell resolution has traditionally involved implanting electrodes directly into the brain. Multielectrode array recording devices remain the gold-standard approach to recording *in vivo* electrophysiological cellular activity (Hong and Lieber, 2019). Growth in the number of simultaneously-recordable signal channels has been driven by electrode fabrication, packaging, materials, and implementation. Neuron density, brain area size and tissue displacement due to wiring and other physical device properties impose fundamental limits on the number of recordable neurons. New neural recording technologies that exploit nanoscale features and integrated electronics are significantly increasing the number of single cells that can be recorded concurrently in single or multi-sites of the brain. Two fundamental factors are paving the way towards large-scale neurophysiology. One factor is to increase the number of electrode/channels through advanced packaging and new materials (Sholvin et al., 2016). To date, hundreds to thousands of electrodes have been implanted to record neural activity *in vivo* (Figure 1B; Berenyi et al., 2014; Shobe et al., 2015; Jun et al., 2017; Chung et al., 2019; Steinmetz et al., 2018; Chiang et al., 2020). The other factor develops 3D electrode array technologies, by combining laminar and movable penetrating electrodes and 2D electrode arrays, to record depth- and layer-specific areas in brain circuits (Hoogerwerf and Wise, 1994; Rios et al., 2016).

While electrophysiology traditionally has been used to collect neural activity of local brain area with high temporal resolution, various optical imaging techniques make possible whole brain recordings, focusing on the network and circuit levels (Yan and Yuste, 2017). Multiphoton imaging also enables cellular-resolution chronic recordings of large-scale neuronal ensembles *in vivo* across days and weeks (Jercog et al., 2016; Kim et al., 2016; Huang et al. 2018; Pachitariu et al., 2017; Weisenburger and Vaziri, 2018). Recently, modern technologies have also rapidly improved the spatiotemporal resolution and sampling speed of optical imaging and microscopy (Rumyantsev et al., 2020; Wu et al., 2020). Ultimately, physical constraints will impose a limit on the effectiveness of optical imaging, as any imaging techniques encounter the tradeoff between the imaging speed, field-of-view and depth.

For all neural interfaces that rely on electrophysiology or optical imaging, technological factors constrain the number of signal channels that can be recorded or controlled simultaneously (Marblestone et al., 2013; Kleinfeld et al., 2019). These constraints involve, for example, power and thermal dissipation for implanted wireless arrays (Zhou et al., 2019), sampling frequency or optical paths for microscopes (Figure 1C; Tsai et al., 2015; Sofroniew et al., 2016; Stirman et al., 2016; Lecoq et al., 2019) and wiring constraints for electrode arrays (Marblestone et al., 2013; Hong and Lieber, 2019; Raducanu et al., 2017). Constraints on simultaneous access lead to a selection problem involving how to use the available signal channels to optimally monitor and manipulate the neural population of interest (Saxena and Cunningham, 2019; Moreaux et al., 2020). If there were no constraints, one could simply measure from all signal channels and there would be no selection problem. If there were too many constraints, there would be very few simultaneously accessible signal channels obviating the problem of selection. For most modern neurotechnologies, however, the space of possible selections is combinatorial. For example, Neuropixel electrode arrays contain 960 electrodes (Steinmetz et al., 2018); however, only 384 recording channels can be acquired simultaneously. Subject to other constraints, there are 2.5^{149} different possible selections for this array (Choi et al., 2020). Similarly, the two-photon random access mesoscope (2p-RAM) has a 5-mm field of view cellular-resolution microscope that makes available up to a million neurons in the transgenic mouse expressing GCaMP in neurons (Sofroniew et al., 2016). However, adaptive optics strategies are necessary to flexibly and rapidly deliver light and make available neurons for simultaneous investigation. For instance, a system using custom optics and independently repositionable temporally multiplexed imaging beams offers an expanded field-of-view (>9.5 mm^2), with multi-site imaging of tens of thousands of neurons across multiple mouse cortical areas (Stirman et al., 2016). In each of the above-mentioned cases, modern instrumentation leads to a combinatorial explosion of possible selections.

Chronic Experiments, Task Complexity, Naturalistic Behavior

Increasingly, modern neurotechnologies are being deployed chronically in implanted systems (Schwartz et al., 2014; Tybrandt et al., 2018; Chiang et al., 2020). The main concerns of chronic electrophysiological recordings are the unit yield, longevity, stability and quality of neural signals (Juavinett et al., 2019; Luo et al., 2020). In all recording devices, the interfaces between the nervous system and a synthetic sensor involve innovations in advanced materials (Chen et al., 2018). Advanced microelectrode technologies have been invented for recording interfaces to improve biocompatibility and stability (Fattachi et al., 2014), which enable us to repeatedly sample the activity of the same population of neurons. In addition, wireless recording devices have become increasingly available for chronic data acquisition.

The challenge of neuroscience data analysis is further magnified by the complexity of behavior. New technologies allow complex, naturalistic and unconstrained behaviors to be measured with increasing detail at the individual and group levels (Tseng et al., 2018). Some behaviors, such as navigation, can involve multiple animals in social interactions (Danjo et al. 2018) or in three-dimensional spaces (Omer et al., 2018). Skeletal movements involve joint rotations with as many as 27 different joint angles for the primate arm and hand. Other task behaviors, such as motor learning, can last hours, days and even weeks (Sandler, 2008). As the temporal duration increases, the task complexity also scales up.

Naturalistic behavior introduces additional issues. To be considered naturalistic, a behavior should not depend on training to follow experimenter-defined instructions. In the absence of instructions, however, preferred behaviors will be acquired and behavioral stereotypy can emerge; namely, subjects can choose to repeatedly make the same, potentially optimal, action sequences, such as “look-then-reach” when picking up a cup. To more completely study the underlying neural mechanisms, investigations of naturalistic behaviors may need to adaptively deliver instructions in an AACL experiment. In active sampling behaviors (such as sniffing or shifting gaze), subjects actively use attention and active sensing strategies to sample relevant cues for information seeking or decision making. While animals can learn a sampling policy through attentional learning and reward maximization, it poses a challenge for experimenters to study the neural correlates underlying such behaviors.

Subsampling and Resampling of Neural Space

As the number of simultaneously recorded neurons from electrophysiology or calcium imaging becomes very large (e.g., 10,000-1,000,000), redundancy will arise. Given a specific recorded brain target, identification of a high- or low-dimensional neural code will vary according to the question of interest. For instance, the visual cortex may have a high-dimensional representation for visual signals, yet a low-dimensional representation for other nonvisual behavioral variables (Stringer et al., 2019a, 2019b). Random sampling is a widely used statistical strategy for estimating the properties of a large network or system. Supported by the law of large numbers and distribution invariance, subsampling assumes exchangeability and ergodicity of a stochastic dynamical system. In data acquisition, large-scale sampling of neural signals enables us to examine the resampling axis in order to assess neural dimensionality and coding sufficiency. For instance, a theoretical question regarding the neural code is "what is the dimensionality of odor space?" (Meister, 2015), or "what is the intrinsic multi-neuronal dimensionality or the complexity of dynamics that relates to the task behavior?" (Gao and Ganguli, 2015; Gao et al., 2017). Unlike traditional data-replacement resampling techniques, sequential neural resampling opens the door to measuring neuronal populations in an integrated manner to generate datasets that are sufficient to rigorously test hypotheses about brain functions. Additionally, researchers may test if subsampling of neuronal populations can preserve the invariant structure of network structure or neural dynamics (Chen et al. 2014; Williamson et al., 2016; Gao et al. 2017; Liu et al., 2019).

Curse of High-Dimensional Data Analysis

The combination of task complexity, multimodality, and large-scale chronic experimental paradigms can quickly generate high-dimensional, structured neural and behavioral data whose analysis and interpretation can outpace computational capabilities. A statistical curse of dimensionality (CoD) arises to impede the discovery cycle within the *Analyze* step.

The common theme of CoD problems is that when the dimensionality increases, the volume of the space increases so rapidly that the available data become very sparse. For instance, to study d-dimensional behavioral variables, we design N experimental trials and record m neurons. If we increase d and m separately or jointly while keeping N unchanged, the insufficient sample size will make it difficult to relate a neural space R^m to a behavioral space R^d. In this case, in order to establish statistical significance, the number of samples (trial-by-duration) needed to support the result often grows exponentially with the dimensionalities d and m.

High-dimensional neural data imposes a CoD across many statistical analyses. First, neural data analysis depends on second and higher-order computations critical to understanding networks, such as functional connectivity. However, the number of trials and duration of trials needed for a reliable statistical estimate does not scale with data dimensionality. Statistical estimation of the covariance matrix in a principal component analysis (PCA) can suffer strong bias and/or high variance when the sample size is insufficient given the data dimensionality (BOX 1). Second, statistical estimation, either by model-free or model-based approach, can be ill-posed when analyzing high-dimensional data. While model-free approximations can have a small number of parameters, they may lack neuroscientific validity. In contrast, model-based approaches can involve many parameters, but pose challenges for model fitting when the data are high-dimensional. Therefore, incorporation of hypothesis-driven theories, priors, and constraints into the model may help solve ill-posed estimation problems. Dimensionality reduction techniques are important tools to tackle large-scale neural recordings on a single-trial basis (BOX 2; Cunningham and Yu, 2014). Third, the complexity and long timescales of task behaviors will introduce plasticity or non-stationarity in neural recordings, posing additional estimation challenges.

Scaling data acquisition and analysis should accelerate the rate of discovery (Figure 1E). However, the curse of high-dimensional data exponentially increases the time necessary to obtain each discovery. As a result, the discovery rate may saturate as data acquisition and analysis increase in scale. The challenge is to maintain an increasing rate of discovery while increasing the scale of data acquisition and analysis. As we discuss below, AACL experiments may offer a solution.

AACL Experimental Paradigms

Closed-loop experiments represent a paradigm shift from open-loop experiments. In closed-loop experiments, neural signals are processed to algorithmically generate feedback signals that are delivered to the subject according to a policy (Zrenner et al., 2016; Yang and Shanechi, 2016; Ciliberti et al., 2018; Srinivasan et al., 2018; Kane et al. 2020; Bolus et al., 2018; Walker et al., 2019; Ponce et al., 2019; Tafazoli et al., 2020; see also reviews in Potter et al., 2014; El Hady, 2016). Traditionally, feedback in a closed-loop experiment can take a variety of forms. If the purpose of a brain-machine interface (BMI) is to control an external actuator, the feedback can be the delivery of stimulation to the nervous system; if the goal of BMI is to control sensory feedback, the feedback can be the timing of sensorimotor information. However, in all closed-loop BMIs, data acquisition is subject to a signal bandwidth constraint, and analysis and feedback are subject to a time constraint. The timescale of feedback is often on the order of milliseconds or seconds that map from circuit functions to behavior. Here we argue that passive closed-loop experiments are still insufficient and inefficient. Specifically, we introduce AACL experiments which generalize concepts familiar to traditional closed-loop experimental designs and include active feedback that is based on multiple stages of knowledge discovery. The terms “active” and “adaptive” are subtly different yet often exchangeable in the literature. By “active”, the experimenter can manipulate the instrumentation or experimental stimuli according to a predefined or optimized policy. Unlike passive feedback that arises automatically regardless of the user’s intention, active feedback emphasizes the effort of seeking valuable information from the feedback signal, and then iteratively optimizes the discovery process at various stages (e.g., sampling, resampling analysis, stimulation). By “adaptive”, the experimenter can modify the decoder or stimulation parameters over time based on feedback or error-correction learning.

AACL experiments enable the testing of hypotheses that cannot be tested by non-AACL experiments in two distinct ways. Some hypotheses can, in principle, be tested using both AACL and non-AACL experiments but non-AACL experiments are sufficiently inefficient that, in practice, the hypothesis cannot be tested due to lack of time. For example, hypotheses that depend on neurostimulation efficacy which requires estimating a map of responses to stimulation at different stimulation sites. Other examples of new knowledge acquired in practice by AACL experiments include neuron-stimulus sensitivity, maximal electrode channel unit yields, and system controllability. Other hypotheses cannot be tested by non-AACL experiments even in principle and require the use of AACL experiments. For example, hypotheses that depend on learning especially when learning occurs rapidly and when learning is irreversible. When both AACL and non-AACL experiments can be performed in principle, then the nature of knowledge gained is similar except that AACL experiments obtain knowledge at a faster rate due to their improved efficiency. In cases where AACL experiments cannot be performed by non-AACL experiments, the nature of knowledge gained is distinct.

To use neurostimulation again as the example, traditional closed-loop stimulation is designed in an on/off stimulation fashion triggered by predetermined stimulation parameters. In contrast, AACL experiments can actively seek the feedback from neuronal firing and adjust the stimulation parameters or control policy to optimize the "natural" cost function (Choi et al., 2016; Bolus et al., 2018; Tafazoli et al., 2020). The cost function is defined by the difference between the observed neural responses and predicted neural responses, where the predictor can be, for example, a simple linear-nonlinear Poisson (LNP) model, or an artificial neural network.

We propose that AACL experiments offer a natural solution to the scaling bottleneck and improve the scalability. In contrast to the standard "*Conceive-Acquire-Analyze-Test-Revise*" paradigm that does not impose strict time constraints on each step, AACL experiments collect and analyze neural data in a sequential manner with time constraints, and test adaptive hypotheses with timely neurofeedback that accounts for neural plasticity during the course of learning and adaptation (Figure 2A). The active and adaptive strategies can be implemented, independently or jointly, throughout the acquisition, analysis, and feedback steps. The form of feedback may be diverse, in terms of stimulus optimization, experimental design, decoder adaptation, neurostimulation, and other user-defined feedback control. The discovered knowledge accumulates with completion of each step. The discovery cycle continues until the experimental subject reaches the predefined experimental goal algorithmically according to the policy. Notably, certain stages of the AACL experiments accommodate many other names proposed in the literature as special cases, such as active experimental design, active stimulus selection, closed-loop feedback control, and closed-loop decoder adaptation. AACL experiments therefore generalize the concept of closed-loop experiments across time-scales for closing the loop and iterating the discovery cycle, as quickly as a fraction of a second, to chronic experimental preparations, as long as months and years.

The concept of adaptive experiments is not new in neuroscience. For instance, design of adaptive experiments is a long-established standard for psychometric testing, such as the use of QUEST procedure (Watson and Pelli, 1983). At slower feedback time-scales, iterative closed-loop paradigms are already well-established in various domains within systems neuroscience. Neuronal stimulus selectivity in the ventral visual pathway exists in a high-dimensional space of sensory stimuli. To assess ventral stream selectivity, Qian-Quiroga and colleagues systematically searched for responses of single neurons to stimuli (Qian-Quiroga et al., 2005). The limited time available for experiments required closing the loop in two-stages of correlational studies. First, responses in a screening session were analyzed, and then used to select target stimuli for the testing session. While each stage was an open-loop experiment, closing the feedback loop across stages tamed the explosion of potential experiments and made possible a more focused investigation. Dramatic increases in the number of neural signal channels that can be monitored and manipulated means neuroscience investigations increasingly lie within a high-dimensional space of experimental designs. These capabilities are opening the door to new applications of closed-loop experimental paradigms to map networks as part of large-scale investigations of multiregional communication (BOX 3). To follow the similar philosophy but with improved efficiency, an analogue of AACL experiment is to identify sensory stimuli that optimize visual neuronal responses at a fast, sub-second timescale. Specifically, Walker and colleagues developed "inception loops", a closed-loop paradigm combining in vivo recording from thousands of neurons with in silico nonlinear response modeling (Walker et al., 2019). The closed-loop model-based response prediction enabled them to generate synthetic yet optimal stimuli (Figure 2B). Therefore, designing adaptive closed-loop image synthesis systems to explore the single or population neuronal response properties represents a new paradigm in visual neuroscience (Ponce et al., 2019; Bashivan et al., 2019).

Closed-loop BMIs can not only learn optimizing sensory stimuli, but also learn active sensing strategies (Richardson et al., 2019). Specifically, experimental manipulation of task-relevant sensory feedback, provided by intracortical microstimulation (ICMS) that encoded egocentric bearing to the hidden goal direction, can reveal distinct motor strategy adaptation to match novel sensor properties for goal-directed behavior. Additionally, BMIs seek to deliver either neural feedback by stimulating neural activity (SENSE-STIMULATE) or user feedback through an external interface that the user controls (SENSE-CONTROL). In neural feedback BMIs, subjects do not need to be aware of the operation of the interface. The BMI seeks to disrupt on-going network excitation or inhibition, e.g. seizure control or optogenetic control (Berenyi et al., 2012; Paz et al., 2013; Grosenick et al., 2015), and/or shape neural plasticity, e.g. mood regulation (Zhang et al., 2021; Shanechi, 2019). In contrast, user feedback BMIs, e.g. visual and motor prostheses, depend on how the user learns to use the interface (Carmena et al., 2003; Koralek et al., 2012; Shenoy and Carmena, 2014). Another example of AACL experiment is closed-loop decoder adaptation (CLDA) used in BMI systems, which can accelerate learning and improve performance by iteratively updating a BMI decoder's parameter (Dangi et al., 2013; Figure 2F). In these cases, volition, awareness and agency play an important role as the subject controls the relevant patterns of neural activity decoded by the BMI. In principle, neural-feedback and user-feedback-based BMIs can be combined. For example, BMIs based on feedback that the user controls could also feature neural feedback protocols designed to recruit brain plasticity and enhance learning (Shenoy and Carmena, 2014).

BMIs offer clinical opportunities as neuroprosthetic devices (Collinger et al., 2013; Ajiboye et al., 2017). Additionally, BMIs provide a novel experimental platform for performing adaptive perturbations and causal circuit manipulations. One successful AACL application is to employ an adaptive closed-loop patterned stimulation strategy (Tafazoli et al., 2020), which learns to use multi-site electrical stimulations to control the pattern of a population of neurons. Additionally, BMIs can help reveal important circuit mechanisms and are particularly useful when studying learned behaviors and sensorimotor control (Jarosiewicz et al., 2008; Koralek et al., 2012; Sadtler et al., 2014; Golub et al., 2016). By making explicit the system inputs and outputs, BMIs allow us to resolve the neural computations that drive learning and test how network structure influences learning (Orsborn and Pesaran, 2017). Since BMIs require low-latency feedback, they can also be used with causal circuit manipulations to stimulate or inactivate in a state-dependent manner. State-dependence allows manipulations to be sensitive to the dynamic properties of brain processes (Qiao et al., 2020) and is an essential component of closed-loop feedback control algorithms (Shanechi et al., 2016; Srinvivasan et al., 2018; Yang et al., 2021). Therefore, BMIs can enable us to conditionally test specific causal functional roles for neural circuits or their plausible links to behaviors.

Features and Limitations of AACL Experiments

Discovery does not rely on closed-loop experiments per se; neither do closed-loop experiments automatically lead to discovery in neuroscience. However, AACL experiments can provide a timely feedback, and update new hypotheses *iteratively* during the course of discovery process.

High-dimensional capabilities enabled by modern neurotechnologies present not only opportunities in establishing the links between neuronal activity and behavior, but also challenges and paradigm shifts in neural data analysis and interpretation. Traditional neuroscience paradigms based on spike sorting and tuning curve estimation will inevitably fail to capture the complexity and dynamics of naturalistic behaviors because the behaviors occupy high-dimensional spaces. AACL experiments offer opportunities to perform 'active' experimental designs that algorithmically select experimental parameters from a high-dimensional space of configurations. In traditional 'passive' experimental designs, each step of the cycle has a predetermined policy. For instance, the *Acquire* step uses the fixed stimulus configurations; whereas in the *Analyze* step, the stimulation configuration or control strategy is fixed. Active experimental designs feature adaptive selection strategies that optimize each step in a closed-loop using real-time neurofeedback. For instance, animal training can be optimized (Bak et al., 2016), experimental stimulus design can be optimized in a sequential manner (Lewi et al., 2009; 2011); and feedback control or neurostimulation can be optimized on the fly (Cunningham et al., 2011; Swann et al., 2018). As a result, we can efficiently test hypotheses sequentially, and potentially even in parallel.

The challenge presented by high-dimensional experimental configurations is particularly acute in the case of neurostimulation experiments. Unlike neural recordings which can be performed at multiple sites simultaneously, neurostimulation experiments can only performed one at a time by choosing "when", "how", and "where" to stimulate. The resulting spatiotemporal patterns of stimulation occupy a particularly high-dimensional configuration space which cannot necessarily be probed simultaneously. In a sense, stimulating all the electrodes at once is not analogous to recording from them all at once. As a result, in the general case only a relatively small number of stimulation configurations can be tested in a single experimental session. Since the nervous system is adaptive and plastic, with constantly changing neural responses, we cannot necessarily rely on comparing stimulation responses to different configurations in different sessions. Novel AACL experimental designs will be critical for progress toward identifying causal roles for neural codes.

Another important approach features AACL experiments with active designs to guide neuronal subsampling and resampling (Figure 2C). A central issue in these experimental designs is whether the properties of the repeatedly sampled populations reflect properties of the underlying distribution. Closed-loop acquisition is like an active search in the space of neural activity to maximize the signal-to-noise ratio. Unlike active sensing in behavior that reflects the animal's behavioral policy, neuronal subsampling is guided by the experimenter's policy, subject to physical, time and bandwidth constraints. For instance, we can design an algorithm that optimizes the joint electrode selections for all recording channels according to the experimenter's policy in order to maximize the isolation quality of detected neurons (Choi et al., 2020). Analyzing resampled populations is very effective when performing dimensionality reduction. Subsampling *m* neurons from a population of *n* neurons can be viewed as a random projection from an *n*-dimensional manifold (Ganguly and Sompolinsky, 2012; Gao and Ganguly, 2015); in this sense, resampling can be viewed as multiple random projections of *n* neurons. The Johnson-Lindenstrauss lemma states that random projections preserve the pairwise distances of high-dimensional data (Bingham and Mannila, 2001). As a result, properties of the underlying distribution that depend on pairwise distances, such as in dimensionality reduction techniques, are preserved by resampled populations. Modes estimated from resampled populations may share other distributional properties with the underlying population. Notably, neural activity often follows a log-dynamic law (Buzsaki and Mizuseki, 2014), and linear combinations of subsampled lognormal distributed neural responses can also be approximated by a lognormal distribution (Asmussen and Rojas-Nandayapa, 2008).

Establishing causality is the holy grail for many questions in systems neuroscience. When doing so, it is important to distinguish between how correlation and causation arise in closed-loop experiments. Correlational dependencies describe associations of measurements that experiments do not control, whereas causal dependencies link a dependent variable to an experimentally controlled variable (Jazayeri and Fraz, 2017). The key concept in causal inference is randomization, such as a random external stimulus or random perturbation (e.g., microstimulation or optogenetic stimulation). The relationship between every dependable variable and the randomized variable is causal, whereas the relationship between non-randomized variables and behavior, remains correlational. As the brain activity is high-dimensional, correlations within massively under-sampled neuronal recordings cannot fully reveal circuit mechanisms. Although closed-loop experiments can contain both correlation and causation components, they can be distinguished from open-loop perturbation experiments in timing and specificity, thereby narrowing the search space of neural code-behavior relationship--- i.e., mitigating the dimensionality bottleneck.

It is also important to point out the limitations of closed-loop perturbation-based approaches for the causal dissection of circuit and behavior. First, the brain is complex, and many brain areas can engage in even a simple task or spontaneous behavior (Stringer et al., 2019b). Therefore, even large-scale neural recordings can only provide a small window of the brain activity, and our target system is partially observable. The presence of latent variables can bring an additional degree of complexity to precisely controlling variables of interest (either neural activity or behavior); induced unexpected network-level side effects complicate data interpretation. Second, the brain is nonlinear and plastic. Consequently a control strategy that works in a certain condition may not necessarily generalize well in other tasks or behavioral states. Third, behavior can also be complex (although the dimensionality of behavior is much smaller than the neural dimensionality), and each axis of behavioral space may be jointly or independently controlled by neural correlates. Fourth, electrical or optical stimulation may create undesirable lasting side effects (e.g., heat, toxicity, cell death, change in excitation-inhibition balance) that bring additional uncertainties to the specific brain functions under study. Finally, even the most sophisticated neural stimulation technologies available today suffer the limitation that they artificially activate or suppress neural activity. Such manipulations are artificial in that they highjack the system and drive the neuronal network into "unnatural" regimes. Many BMI technologies based on non-specific and unnatural perturbations may face challenges in how to interpret the results, as traditional non-adaptive neurostimulations may cause varying degrees of side effects on behavior, or even generate "false discoveries". To mitigate such concerns, recent work performing causal network analyses emphasizes the use of minimal perturbations and delivers single stimulation pulses (Qiao et al, 2020). Related work on multiregional network system identification shows that pairs of sites that share correlated neural activity patterns also tend to share stimulation responses (Yang et al, 2021), indicating that causal manipulations and recordings can be mutually informative and constrain network inferences. While acknowledging that fundamental constraints limit testing causality dependencies between brain and behavior, trial-and-error active manipulations remain the most important source of evidence that a brain area supports one aspect of behavior. The development of next-generation active and adaptive BMIs that deliver "naturalistic" patterned neurostimulations and incorporate appropriate control experiments would help alleviate the illusion of false discovery. A type of new closed-loop feedback for neurostimulation, for instance, can be the output of neuronal firing patterns or the local network connectivity (Vlachos et al., 2016; Choi et al., 2016; Bolus et al., 2018; Tafazoli et al., 2020).

Speeding Up Neural Data Analysis

Another important issue of scalability in knowledge discovery involves speed. Even if the dimensionality of data remains constant, the increasing amount of data may still create an analysis bottleneck for knowledge discovery. In data analysis and interpretation, we aim to avoid an exponential complexity or computation latency with respect to the number of neurons, seeking a linear or sub-linear order of complexity. In developing efficient analytic tools, computation speed and scalability are key considerations.

First, closed-loop BMIs impose low-latency constraints in all experimental steps. Computational overhead jointly depends on data size, CPU architecture, memory, and bandwidth. Overall, computation latency is composed of two parts: *Total cost = fixed cost + scaling cost*, where the first term is independent of the scale of data, and the second term increases with the scale of data. Therefore, scaling up data acquisition can impose a great challenge in computation speed due to limited resources in memory, bandwidth and computing power. To accommodate scalable ultrafast neural data analysis, modern computing resources and dedicated hardware can help meet these resource requirements. According to a current estimate of the doom of Moore's law, the computing power of single CPU will similarly reach the physical limit around 2022. In contrast, high-performance computing devices based on graphic processing unit (GPU) and field programmable gate arrays (FPGA) have become widely adopted for data analyses (Hu et al., 2018; Giovannucci e al., 2018).

In the *Acquire* step, open-source low-latency hardware (e.g., Open Ephys 2.0) has managed to minimize high-speed sampling delay with microsecond latency. Automated and scalable hardware-empowered spike sorting can accommodate real-time processing for large-scale data acquisition (Pachitariu et al., 2016; Chung et al., 2017; Jun et al., 2017; Yger et al., 2018). In the

Analyze step, computational tasks can be operated in real time for ultrafast decoding, detection and control. Using the rodent hippocampus as an example, hippocampal replays during sharp-wave ripples are known to contribute to memory consolidation, planning and future decision making (Buzsaki, 2015). Closed-loop perturbation experiments that aim to investigate the contribution of these replay events may narrow the search space of relationship between neural code and behavior, or eliminate alternative competing hypotheses (Girardeau et al., 2009; Fernandez-Ruiz et al., 2019). Therefore, it is important to develop scalable methods that enable real-time decoding and assessment of these hippocampal replay contents to match the complexity of neural data, in the form of large-scale unsorted ensemble spikes (Ciliberti et al., 2018; Hu et al., 2018; Figure 2D,E) or high-density field potential recordings (Cao et al. 2020; Frey et al., 2019), or large-scale calcium imaging (Tu et al., 2020). Take the primate motor cortex as another example, closed-loop BMIs have provided mechanistic insight into learning, plasticity and functional reorganization (Jarosiewicz et al., 2008; Sadtler et al., 2014; Shenoy and Carmena, 2014). The development of scalable methods for decoding arm or hand movement or assessing neural population dynamics can greatly advance research in motor control (Trautmann et al., 2019; Sussillo et al., 2016). The key component of BMIs is feedback, in the form of neurostimulation (Berenyi et al., 2012; Paz et al., 2013; Grosenick et al., 2015; Zhang et al., 2021), user-defined feedback control (Figure 2G; Carmena et al., 2003; Dangi et al., 2013; Shanechi et al., 2016), or the prediction error of neural responses (Figure 2H; Tafazoli et al., 2020), which can be further used to perturb the circuit or causally change the behavior. Finally, the time window of closed-loop feedback is critical as it allows interaction with neurons and circuits differently. Sub-millisecond feedback stimulation may prevent recurrent inhibition, but the same setup with an order of second delay may affect the system in a completely different manner. These uncertainties of mechanistic inquiry grow in time especially when the casual chain between the cause (stimulation) and effect is long. Therefore, timing imposes a strict low-latency constraint on closed-loop BMIs (Muller et al., 2012; Kane et al., 2020).

Second, scalable data-intensive computation demands fast and efficient computing strategies. Even though real-time operation may not be always required, off-line processing of high-throughput high-dimensional neural data can still be prohibitive, this may include neural-behavior mapping (Vogelstein et al., 2014), large-scale model fitting, data visualization and computer simulations. For instance, structural data are fundamentally high-dimensional, including 2-D images, 3-D volumes, and 4-D and 5-D hypervolumes for multispectral data. Large-scale neural circuit mapping may require both structural and functional data (Shi et al., 2015). High-performance computing is required to analyze high-resolution high-throughput neuroanatomy and neuroimaging data. Behavioral data can be also high dimensional, especially when they are captured via high-fidelity video recordings. Data-intensive, automated image segmentation and 3-D morphological reconstruction have been empowered by powerful deep learning methods for behavioral video or imaging analyses (Mathis et al., 2018; Pereira et al., 2019; Zhou et al. 2018; Arac et al., 2019). Another source of high-dimensionality arises from multimodal measurements, such as concurrent EEG/MEG source localization (Antelis and Minguez, 2013). Furthermore, large-scale biologically-inspired neuronal network modeling and computer simulations may leverage high-performance GPU or FPGA computing (Hoang et al., 2013; Sripad et al., 2018). Finally, distributed data analytics platform and computing infrastructure can help achieve fast and scalable data analysis of massive size (Freeman et al., 2014; Freeman 2015).

Third, artificial intelligence (AI) and machine learning can help accelerate the pace of neuroscience discoveries (Marblestone et al., 2016; Richards et al., 2019; Cichy and Kaiser, 2019) and scale up innovation (Kittur et al., 2019). On the one hand, AI and deep learning can help or automate complex and large-scale neural data analyses to uncover the patterns in brain activity. For instance, in neural encoding, deep learning can help link complex patterns of neural activity and/or cortical anatomy to complex behavior (Minderer et al. 2019; Pandarinath et al., 2018), as well as control neuronal spiking or internal brain states (Bashivan et al., 2019). On the other hand, neuroscience can drive AI forward for knowledge discovery; neuroscience-inspired AI has achieved a professional human-level intelligence for playing chess and computer games (Silver et al., 2016; Hassabis et al. 2017). The *AlphaGo*, motivated from deep reinforcement learning, discovered a remarkable level of Go knowledge through a self-taught training process. Therefore, brain-inspired deep learning architectures can not only provide a new computational framework for brain information processing (Kriegeskorte, 2015; Banino et al., 2018), but also generate new insight in systems neuroscience and provide rapid theoretical and experimental progress (Richard et al., 2019).

Concluding Thoughts: Scaling to the Human Brain

Currently, our understanding of brain mechanisms in animal models and in the human brain are separated by a divide. This is partly due to the additional ethical, safety and efficacy, and financial constraints that govern the development of neurotechnologies for use in humans. Nevertheless, progress in neurotechnology is increasingly making possible studies of the human brain. The vast scale and complexity of the human brain inevitably means that understanding how to jointly scale data acquisition and data analysis will play an essential role in progress. To date, high-density biocompatible and stretchable electrode grids can record spikes and LFPs at the surface of human brain (Khodaholdy et al., 2015; Tybrandt et al., 2018). Scaling up data acquisition via high-density interfaces may further improve the spatiotemporal resolution of human brain mechanisms (Robinson et al., 2017; Matsuhita et al., 2018; Escabi et al., 2014; Even-Chen et al., 2020; Sohrabour et al., 2020). Concurrent multimodal and multisite recordings, neuroimaging and neurostimulation will also drive progress (Chang, 2015; Krook-Magnuson et al., 2015; Swann et al., 2018). Basic brain mechanisms will play a role in our understanding of the diseased brain. In translational or therapeutic applications, closed-loop human BMI systems have been widely adopted for restoring or enhancing sensory, motor or cognitive brain functions, as well as delivering anesthesia drug (Shanechi, 2019; Moses et al., 2019; Gilja et al., 2015; Liberman et al., 2013; Yang and Shanechi, 2016). As a result, AACL experimental or adaptive BMI paradigms may have significant impacts on human brain science. The development of stable, secure, real-time brain-cloud interfaces similar to current mobile voice and image-based interfaces will be critical to updating model-based inferences based on new observations (Martins et al., 2019).

In summary, knowledge discovery in systems neuroscience is being transformed by advances in neurotechnology. Fundamentally, the scale of data acquisition and speed of data analysis jointly determine the rate of hypothesis testing or revision, and ultimately the rate of discovery. The peril arises from how scaling up data acquisition can slow down data analysis. AACL experiments offer a solution to improve scalability for knowledge discovery. Achieving this vision requires the coordination of scalable computation, active and adaptive experimental designs in real-time systems and interfaces. Ultimately, the successful scaling of knowledge discovery is essential to understand the complex brain mechanisms supporting cognition and behavior in health and disease.

BOX 1. Correlation Matrix Estimation (303 words)

The correlation matrix is a central statistical measure central to principal component analysis (PCA). As the dimensionality of observed variables, *n,* increases, the number of estimated parameters scales quadratically in *n*. A curse of dimensionality (CoD) arises when the sample size *N* is small in relation to the dimensionality, *n*; and the correlation structure of neuronal populations cannot be reliably estimated because of insufficient number of experimental trials or duration in neuroscience experiments.

A simple model-free solution to the CoD imposes local proximity structure onto the correlations; namely, only neurons recorded from nearby electrodes are connected. Consequently, the number of parameters scales linearly with *n*. However, the spatial proximity assumption may not hold in practice. For instance, two hippocampal place cells recorded from distant electrodes may share overlapping place fields and hence correlated activity.

A model-based approach to the correlation matrix estimation is via partial correlation. Partial correlation is equivalent to conditional correlation when the random variables are multivariate normal distributed. If the observations are discrete (e.g. multinomial), the equivalence also holds when the conditional expectation of the random variables is linear (Baba et al., 2004). By using partial correlation, one can solve *n* linear regression problems, each of which involving *n*-1 regressors and 1 predictor. Since the *n* regression problems can be solved independently, the computation can be scaled up using parallel computing. Therefore, by mitigating the CoD, partial correlation estimates may be not only more reliable, but also computationally efficient.

An alternative approach involves random projection or subsampling, in which one selects *m* variables ($m \ll n$), and repeats the linear regression procedure with different subsets. Statistically, random projection-based correlation estimates assume sufficient sparsity in order to stably embed the subsets into a low-dimensional subspace. Therefore, the correlation estimates are robust with respect to varying number of neurons due to recording instability.

BOX 2. Dimensionality Reduction (315 words)

In systems neuroscience, dimensionality reduction methods are important to answering the neural dimensionality question: i.e., how many neurons are required to resolve the dynamics underlying a behavioral task? The answer may depend on the coding specificity of the stimulus or behavioral variables (Stringer et al, 2019a, 2019b). Knowing the answer can improve our understanding of the scaling property of neuronal population in both encoding and decoding (Williamson et al., 2016).

The choice of neural decoding methods also leads to dimensionality concerns. Linear decoding methods (such as factor analysis and Kalman filter) are commonly used because of their simplicity. In contrast, despite potentially better performance, many nonlinear decoding methods are less commonly used. One important reason is that nonlinear methods suffer a curse of dimensionality. For instance, nonlinear function estimation scales polynomially or even exponentially in terms of dimensionality. Moreover, fitting nonlinear functions requires parameter search in the presence of local minima which also scales with *n*. As a result, nonlinear methods often lack scalability.

Dimensionality reduction methods can help alleviate the computational curse (Cunningham and Yu, 2014). This has motivated the development of a variety of advanced nonlinear dimensionality reduction methods to examine neuronal population activities (Yu et al., 2009; Gao et al, 2016; Wu et al., 2017). However, nonlinear dimensionality reduction approaches are computationally expensive and depend on strong assumptions, such as the ability to conceptualize experimental measurement as a random projection of neural activity.

Adaptive subsampling provides a complementary approach to measurement by random projection and can address the CoD present for dimensionality reduction, which is critical in the context of AACL experiments. Since neurons exhibit log-normal firing rate distributions, applying dimensionality reduction methods to large numbers of neurons may not sufficiently capture the long-tail behavior. As a result, neuronal representations may be incompletely characterized. By allowing for better sampling in the tails, adaptive subsampling of neurons can provide a more complete picture.

BOX 3. Mapping Networks in Neuroscience (308 words)

Brain function and dysfunction are increasingly understood as being due to the expression of multiple overlapping network mechanisms. Network mechanisms of multiregional communication are most often inferred from the structure of correlations in neural activity. The availability of recordings from many signal channels has fueled progress. However, functionally connectivity analyses have been typically applied to signals that measure neuronal function indirectly and do not necessarily scale due to fundamental limits on signal resolution – such as blood oxygen level-dependent (BOLD) functional magnetic resonance imaging (fMRI) and widefield calcium imaging signals. Inferring network mechanisms from high-dimensional neuronal recordings is hampered by the CoD. Moreover, using correlations to interpret activity patterns as being due to interactions is subject to significant confounds. Correlations are sensitive to the confounding influence of common inputs from other brain regions, yielding network edges even when the receiver does not receive any input from the sender.

Recent work maps large-scale brain networks and studies the mechanisms of multiregional communication by recording neural responses while delivering low-amplitude stimulation pulses in a causal network analysis (Qian et al, 2020). Taking a causal sampling approach offers important advantages. Causal responses cannot be due to common input. Delivering isolated low-amplitude microstimulation pulses also offers the opportunity to more directly probe network excitability while avoiding the confounding effects and network responses. Inferences from large-amplitude stimulation pulses or pulse trains may recruit network responses that do not reflect direct functional interactions between the stimulation and recording sites (Lozano et al., 2019). Large amplitude pulses and pulse trains can effectively change the interaction instead of measuring the interaction.

Because mapping networks using a causal network analysis allows a selective targeting of neurons and neural circuits for investigation based on their role in the network, we may be able to mitigate the curse of dimensionality associated with scaling up data acquisition and analysis without constraints.

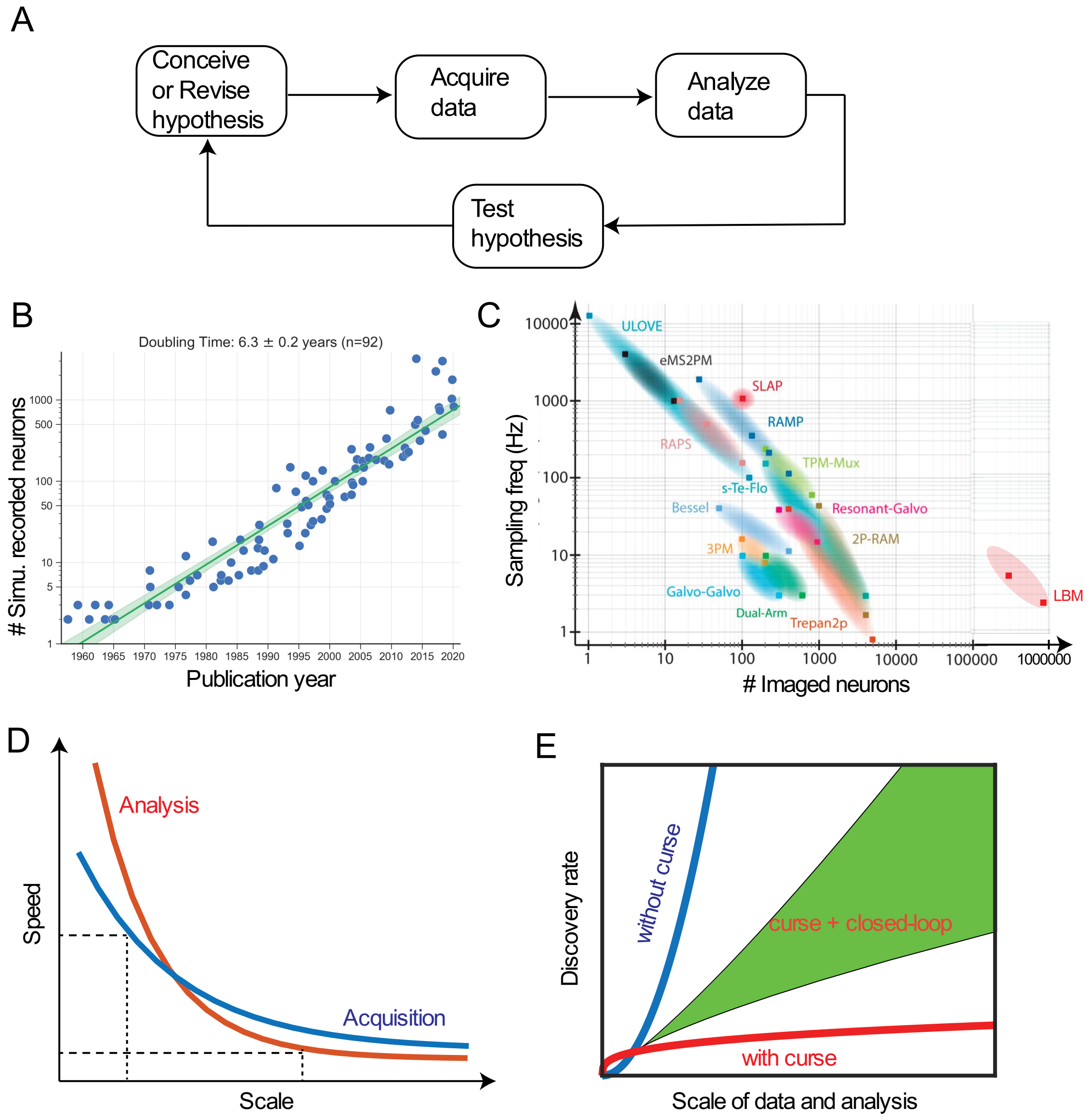

Figure 1. Scaling in Neural Data Acquisition

(A) Cycle of knowledge discovery (*Conceive-Acquire-Analyze-Test-Revise)*. The *Acquire* step consists of recording large-scale neuronal activity during behavior. The *Analyze* step consists of data analysis and interpretation. The *Test and Revise* step involves testing the hypothesis and revising it as necessary to close the cycle.

(B) Near or faster than exponential growth in the number of recorded neurons based on *in vitro* or *in vivo* electrophysiology (up-to-year update from http://stevenson.lab.uconn.edu/scaling/). Recent neurotechnology development for simultaneous neuronal recordings suggested that a further jump from the exponential growth.

(C) Trade-off between sampling frequency and the number of recorded neurons based on microscopy imaging (From Lecoq et al., 2019, Society for Neuroscience). Based on the new Light Beads Microscopy (LBM) technique, ~1,000,000 neurons were recorded within ~5.4 × 6 × 0.5 mm^3 volumes at ~2 Hz (Demas et al., 2021).

(D) Schematic of “scale-speed limit” for data acquisition and analysis steps. For a fixed scale, the pace of discovery is determined by slowest scale-speed factor.

(E) Discovery rate (DR) scaling with data acquisition and analysis: $\mathrm{DR} = \frac{\text{number of discoveries}}{\text{time per discovery}}$. The number of discoveries scales proportionally to the scale of data acquisition. The time per discovery scales inversely proportionally to the rate at which data can be analyzed. Scaling the rate of data analysis with the rate of data acquisition, should lead to an accelerating DR (blue curve). However, the CoD effectively slows down discovery rate and DR scaling requires a correction: This correction to the DR can dominate. If the CoD correction scales faster than the rate of data analysis, the otherwise-accelerating DR flattens (red curve). Closed-loop experimental designs can mitigate the CoD and restore an accelerating DR (green shaded area).

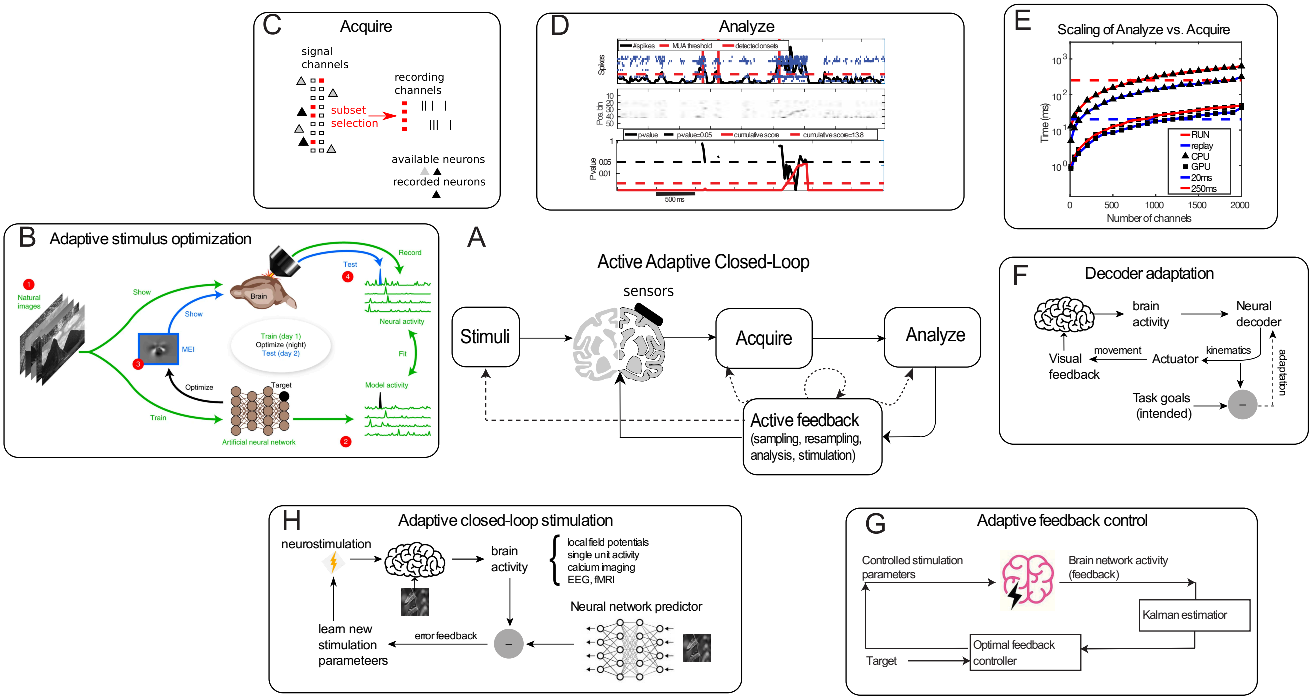

Figure 2. Active Adaptive Closed-Loop (AACL) Experimental Paradigms Provide Solutions to the Curse-of-Dimensionality Problems

(**A**) Schematic of AACL experimental paradigm. The active feedback iteratively updates or optimizes the sampling, resampling or stimulation parameters at each stage of the loop (marked by three dashed arrows). The feedback may have different timescales and forms: experimental design (Lewi et al., 2009; Lewi et al., 2011), adaptive stimulus optimization (Walker et al., 2019; Ponce et al., 2019), neurally-defined stimulation (Berenyi et al., 2012; Paz et al., 2013; Grosenick et al., 2015; Zhang et al., 2021), closed-loop decoder adaptation (Dangi et al., 2013), user defined control or prosthetics (Carmena et al., 2003; Shanechi et al., 2016), or adaptive closed-loop neurostimulation (Tafazoli et al., 2020).

(B) Schematic of adaptive stimulus optimization based on BMI with neural decoder adaptation. In the closed-lopp, an end-to-end trained neural network model predicted thousands of neuronal responses to arbitrary new natural input and synthesized optimal stimuli: most exciting inputs (MEIs) (Adapted with permission from Walker et al., 2019, Springer Nature).

(C) Illustration of subsampling/resampling in the closed-loop Acquire step (Choi et al., 2020). Recording channels can be selected from signal channels by optimizing the subset of signal channels selected to maximize the number of recorded neurons given the available neurons known when all signal channels are recorded. Optimization involves adaptively sampling the signal channels selected to maximize the yield of recorded neurons.

(**D**) A snapshot of the closed-loop Analyze step for large-scale rat hippocampal recordings. A GPU-powered population-decoding system was developed for ultrafast reconstruction of spatial positions from rodent's unsorted spatiotemporal spiking patterns, with real-time speed to decode rat's run position (latency < 250 ms) or memory replays (latency <20 ms). Furthermore, the approach enabled assessment of the statistical significance of online-decoded replay (Adapted with permission from Hu et al., 2018, Elsevier).

(E) Illustration of the scaling between the Acquire and Analyze steps to accommodate real-time operation (~fraction of millisecond per spike). The GPU decoding strategy can scale up to thousands of channels (Adapted with permission from Hu et al., 2018, Elsevier).

(F) Schematic of a BMI with neural decoder adaptation (Dangi et al., 2013). The adaptation design elements include the adaptation, timescale, selective parameter adaptation, smooth decoder updates, and intuitive decoder adaptation parameters.

(G) Schematic of a BMI with neural feedback (Adapted with permission from Yang et al., 2021, Springer Nature). Model-based closed-loop controllers can be designed and consist of a Kalman state estimator and a feedback controller. Brain network activity can be used as feedback and the model-based closed-loop controller identified the stimulation parameters to drive the internal brain state to a particular target.

(H) Schematic of adaptive closed-loop stimulation (Tafazoli et al., 2020). The system learns to use multi-site spatially patterned electrical stimulation to control the pattern of activity of a population of neurons.

ACKNOWLEDGMENTS
We thank G. Buzsaki and S. Shoham for valuable comments on the manuscript. We acknowledge the funding support from the US National Science Foundation (CBET-1835000 to Z.S.C.; BCS-150236 to B.P.), the National Institutes of Health (R01-NS100065 and R01-MH118928 to Z.S.C.; R01-EY024067, R01-NS104923 to B.P.).

REFERENCES

Ajiboye, A.B., Willett FR, Young DR*, et al.* (2017). Restoration of reaching and grasping movements through brain-controlled muscle stimulation in a person with tetraplegia: a proof-of-concept demonstration. *Lancet* 389, 1821-1830.

Antelis JM and Minguez J (2013). DYNAMO: concurrent dynamic multi-model source localization method for EEG and/or MEG. *J. Neurosci. Methods* 212, 28-42.

Arac A, Zhao P, Dobkin BH, *et al.* (2019). DeepBehavior: a deep learning toolbox for automated analysis of animal and human behavior imaging data. *Front. Syst. Neurosci.* 13, 20.

Asmussen, S. and Rojas-Nandayapa, L (2008). Asymptotics of sums of lognormal random variables with Gaussian copula. *Statistics and Probability Letters* 78, 2709-2714.

Baba K, Shibata R, Sibuya M (2004). Partial correlation and conditional correlation as measures of conditional independence. *Aust. N.Z. J. Stat*. 46, 657-664.

Bak JH, Choi JY, Akrami A, *et al.* (2016). Adaptive optimal training of animal behavior. *Advances in Neural Information Processing Systems* 29, pp. 1947-1955.

Banino, A., Barry, C., Uria, B. *et al.* (2018). Vector-based navigation using grid-like representations in artificial agents. *Nature* 557, 429-433.

Bashivan P, Kar K, DiCarlo JJ. (2019). Neural population control via deep image synthesis. *Science*, 364, eaav9436.

Berenyi, A., Somogyvari Z, Nagy AJ, *et al*. (2014). Large-scale, high-density (up to 512 channels) recording of local circuits in behaving animals. *J. Neurophysiol*. 111, 1132-1149.

Berenyi A, Belluscio M, Mao D, and Buzsaki G (2012). Closed-loop control of epilepsy by transcranial electrical stimulation. *Science* 337, 735-737.

Bingham E and Mannila, H (2001). Random projection in dimensionality reduction: applications to image and text data. *Proc. 7th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD'01*), pp. 245-250.

Bolus MF, Willats AA, Whitmire CJ, *et al.* (2018). Design strategies for dynamic closed-loop optogenetic neurocontrol in vivo. *J. Neural Eng*. 15, 026011.

Buzsaki G and Mizuseki K (2014). The log-dynamic brain: how skewed distributions affect network operations. *Nat. Rev. Neurosci.* 15, 264-278.

Buzsaki G (2015). Hippocampal sharp wave-ripple: a cognitive biomarker for episodic memory and planning. *Hippocampus,* 25, 1073-1188.

Cao L, Varga V and Chen Z (2020). Spatiotemporal patterns of rodent hippocampal field potentials uncover spatial representations. BioRxiv, https://doi.org/10.1101/82867

Carmena JM, Lebedev MA, Crist RE, *et al.* (2003). Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biol.* 1, e42.

Chang EF (2015). Towards large-scale, human-based, meoscopic neurotechnologies. *Neuron* 86, 68-78.

Chen R, Canales A, and Anikeeva P (2018). Neural recording and modulation technologies. *Nat. Rev. Materials* 2, 10693.

Chen Z, Gomperts SN, Yamamoto J, and Wilson MA (2014). Neural representation of spatial topology in the rodent hippocampus. *Neural Computat*. 26, 1-39.

Chiang C-H, Won, S-M, Orsborn, A, L. *et al.* (2020). Development of a neural interface for high-definition, long-term recording in rodents and nonhuman primates. *Sci. Trans. Med*. 12, eaay4682.

Choi JS, Brockmeier AJ, McNiel DB, Kraus LM, Príncipe JC, Francis JT (2016). Eliciting naturalistic cortical responses with a sensory prosthesis via optimized microstimulation. *J Neural Eng.* 13, 056007.

Choi, J, Kumar K, Khazali M, *et al.* (2020). Optimal adaptive electrode selection to maximize simultaneously recorded neuron yield. *Adv. Neural Info. Proc. Syst. (NeuralPS)*, https://www.biorxiv.org/content/10.1101/2020.10.06.328526v1

Chung JE, Joo HR, Fan JL, *et al*. (2019). High-density, long-lasting, and multi-region electrophysiological recordings using polymer electrode arrays. *Neuron* 101, 21-31.

Chung JE, Magland JF, Barnett AH, *et al*. (2017). A fully automated approach to spike sorting. *Neuron* 95, 1381-1394.

Cichy RM and Kaiser D (2019). Deep neural networks as scientific models. *Trends Cog. Sci*. 23, 305-317.

Ciliberti, D., Michon, F. and Kloosterman, F (2018). Real-time classification of experience-related ensemble spiking patterns for closed-loop applications. *eLife* 7, e36275.

Collinger, J.L., Wodlinger B, Downey JE*, et al.* (2013). High-performance neuroprosthetic control by an individual with tetraplegia. *Lancet* 381, 557-564.

Cunningham JP, Nuyujukian P, Gilja V, *et al.* (2011). A closed-loop human simulator for investigating the role of feedback control in brain-machine interfaces. *J. Neurophysiol.* 105, 1932-1949.

Cunningham, J. and Yu, B (2014). Dimensionality reduction for large-scale neural recordings. *Nat. Neurosci.* 17, 1500-1509.

Dangi D, Orsborn AL, Moorman HG, and Carmena JM (2013). Design and analysis of closed-loop decoder adaptation algorithms for brain-machine interfaces. *Neural Comput*. 25, 1693-1731.

Danjo T, Toyoizumi T and Fujisawa S (2018). Spatial representations of self and other in the hippocampus. *Science* 359, 213-218.

Demas J, Manley J, Tejera F, *et al*. (2021). High-speed, cortex-wide volumetric recording of neuroactivity at cellular resolution using light beams microscopy. BioRxiv, https://doi.org/10.1101/2021.02.21.432164

El Hady, A (2016). *Closed Loop Neuroscience*. Academic Press.

Escabi MA, Read HL, Viventi J, *et al*. (2014). A high-density, high-channel count, multiplexed uECoG array for auditory-cortex recordings. *J. Neurophysiol.* 112, 1566-1583.

Even-Chen N, Muratore DG, Stavisky SD, *et al.* (2020). Intracortical neural interface design opportunities for an order of magnitude power saving. *Nat. Biomed. Eng.* 4, 984-996.

Fattahi P, Yang G, Kim G, Abidian MR (2014). A review of organic and inorganic biomaterials for neural interfaces. *Advanced Materials* 26, 1846-1885.

Fernandez-Ruiz A, Oliva A, Fermino de Oliverira E, *et al.* (2019). Long-duration hippocampal sharp wave ripples improve memory. *Science*, 364, 1082-1086.

Freeman J, Vladimirov N, Kawashima T, *et al*. (2014). Mapping brain activity at scale with cluster computing. *Nat. Methods* 11, 941-950.

Freeman J. (2015). Open source tools for large-scale neuroscience*. Curr. Opin. Neurobiol.* 32, 156-163.

Frey M, Tanni S, Perrodin C, *et al.* (2019). DeepInsight: A general framework for interpreting wide-band neural activity. https://www.biorxiv.org/content/10.1101/871848v1.full.pdf.

Ganguli S and Sompolinsky H (2012). Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Ann. Rev. Neurosci.* 35, 485-508.

Gao P and Ganguli S (2015). On simplicity and complexity in the brave new world of large-scale neuroscience. *Curr. Opin. Neurobiol.* 32, 148-155.

Gao P, Trautmann E, Yu B, Santhanam G, Ryu S, Shenoy K and Ganguli S (2017). A theory of multineuronal dimensionality, dynamics and measurement. https://www.biorxiv.org/content/10.1101/214262v2 .

Gao Y, Archer EW, Paninski L, and Cunningham JP (2016). Linear dynamical neural population models through nonlinear embeddings. *Adv. Neural Info. Proc. Syst. (NeuralPS)*, pp. 163-171.

Gilja V, Pandarinath C, Blabe CH, *et al.* (2015). Clinical translation of a high performance neural prosthesis. *Nat. Med*. 21, 1142-1145.

Giovannucci, A., Friedrich, J., Gunn, P. *et al*. (2018). CalmAn: an open source tool for scalable calcium imaging data analysis. *eLife* 8, 339564.

Girardeau G, Benchenane K, Wiener SI, *et al.* (2009). Selective suppression of hippocampal ripples impairs spatial memory. *Nat. Neurosci*. 12, 1222-1223.

Golub, M.D., Chase, S.M., Batista, A.P. and Yu, B.M. (2016). Brain-computer interfaces for dissecting cognitive processes underlying sensorimotor control. *Curr. Opin. Neurobiol*. 37, 53-58.

Gottschalk, S., Degtyaruk O, Larney BM, *et al.* (2019). Rapid volumetric optoacoustic imaging of neural dynamics across the mouse brain. *Nat. Biomed. Eng*. 3, 392-401.

Grosenick L, Marshel JH and Deisseroth K (2015). Closed-loop and activity-guided optogenetic control. *Neuron* 86, 106-139.

Hassabis D, Kumaran D, Summerfield C, and Botvinick M (2017). Neuroscience-inspired artificial intelligence. *Neuron* 95, 245-258.

Hoang RV, *et al.* (2013). A novel CPU/GPU simulation environment for large-scale biologically realistic neural modeling. *Front. Neuroinform*. 7, 19.

Hong G and Lieber CM (2019). Novel electrode technologies for neural recordings. *Nat. Rev. Neurosci.* 20, 330-345.

Hoogerwerf AC and Wise KD (1994). A three-dimensional microelectrode array for chronic neural recording. *IEEE Trans. Biomed. Eng*. 41, 1136-1145.

Hu S, Zhang Q, Wang J, and Chen Z (2018a). Real-time particle filtering and smoothing algorithms for detecting abrupt changes in neural ensemble spike activity. *J. Neurophysiol.* 119, 1394-1410.

Hu S, Ciliberti D, Grosmark AD, *et al.* (2018b). Real-time read out for large-scale neural ensemble place codes. *Cell Rep.* 25, 2635-2642.

Huang, C., Maxey, J.R., Sinha, S. *et al.* (2018). Long-term optical brain imaging in live adult fruit flies. *Nat. Commun.* 9, 872.

Jarosiewicz, B., Chase, S.M., Fraser, G.W., *et al* (2008). Functional network reorganization during learning in a brain-computer interface paradigm. *Proc. Natl. Acad. Sci. USA* 105, 19486-19491.

Jazayeri M and Afraz A (2017). Navigating the neural space in search of the neural code. *Neuron*, 93, 1003-1014.

Jercog, P., Rogerson, T. & Schnitzer, M.J. (2016). Large-scale fluorescence calcium-imaging methods for studies of long-term memory in behaving mammals. *Cold Spring Harbor Perspectives in Biology* 8, a021824.

Juavinett AL, Beheet G, and Churchland AK (2019). Chronically implanted neuropixels probes enable high-yield recordings in freely moving mice. *eLife* 8, e47188.

Jun JJ, Steinmetz N, Siegle J, *et al*. (2017a). Fully integrated silicon probes for high-density recording of neural activity. *Nature* 551, 232-236.

Jun JJ, Mitelut C, Lai C, *et al.* (2017b). Real-time spike sorting platform for high-density extracellular probes with ground-truth validation and drift correction. https://www.biorxiv.org/content/10.1101/101030v1 .

Kane G, Lopes G, Saunders JL, *et al.* (2020). Real-time, low-latency closed-loop feedback using markerless posture tracking. *eLife* 9, e61909.

Khodaholy D, Gelinas JN, Thesen T, *et al.* (2015). Neurogrid: recording action potentials from the surface of the brain. *Nat. Neurosci*. 18, 310-315.

Kim TH, Zhang Y, Lecoq JA, *et al.* (2016). Long-term optical access to an estimated one million neurons in the live mouse cortex. *Cell Rep*. 17, 3385-3394.

Kittur A, Yu L, Hope T, *et al*. (2019). Scaling up analogical innovation with crowds and AI. *Proc. Natl. Acad. Sci. USA* 116, 1870-1877.

Kleinfeld D, Luan L, Mitra PP, *et al.* (2019). Can one concurrently record electrical spikes from every neuron in a mammalian brain? *Neuron* 103, 1005-1015.

Koralek AC, Jin X, Long JD, *et al.* (2012). Corticostriatal plasticity is necessary for learning intentional neuroprosthetic skills. *Nature* 483, 331-335.

Kriegeskorte N. (2015). Deep neural networks: a new framework for modeling biological vision and brain information processing. *Ann. Rev. Vision Sci.* 1, 417-446.

Krook-Magnuson E, Gelinas JN, Sltesz I, and Guzsaki G (2015). Neuroelectronics and biooptics: closed-loop technologies in neurological disorders. *JAMA Neurol*. 72, 823-829.

Lecoq J, Orlova N, and Grewe BF (2019). Wide. fast. Deep. Recent advances in multi-photon microscopy of in vivo neuronal activity. *J. Neurosci*. 39, 9042-9052.

Lewi J, Butera R, and Paninski L (2009). Sequential optimal design of neurophysiology experiments. *Neural Computat*. 21, 619-687.

Lewi J, Schneider DM, Woollley SMN, and Paninski L (2011). Automating the design of informative sequences of sensory stimuli. *J. Comput. Neurosci.* 30, 181-200.

Liberman MY, Ching S, Chemali J, and Brown EN (2013). A closed-loop anesthetic delivery system for real-time control of burst suppression. *J. Neural Eng.* 10, 046004.

Liu S, Iriate-Diaz J, Hatsopoulos NG, Ross CF, Takahashi K, and Chen Z (2019). Dynamics of motor cortical activity during naturalistic feeding behavior. *J. Neural Eng.* 16, 026038.

Lozano, A.M., Lipsman, N., Bergman, H. *et al.* (2019). Deep brain stimulation: current challenges and future directions. *Nat Rev. Neurol.* 15, 148-160.

Luo TZ, Bondy AG, Gupta D, *et al*. (2020). An approach for long-term, multi-probe neuropixels recordings in unrestrained rats. *eLife* 9, e59716.

Marblestone AH, Zamft BM, Maguire YG, *et al.* (2013). Physical principles for scalable neural recording. *Front. Comput. Neurosci*. 7, 137.

Marblestone AH, Wayne G and Kording KP (2016). Toward an integration of deep learning and neuroscience. *Front. Comput. Neurosci*. 10, 94.

Martins NRB, Angelica A, Chakravarthy K, *et al.* (2019). Human brain/cloud interface. *Front. Neurosci.* 13, 112.

Mathis A, Mamidanna P, Cury KM, *et al*. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* 21, 1281-1289.

Matsuhita K, Hirata M, Suzuki T, *et al.* (2018). A fully implantable wireless ECoG 128-channel recording device for human brain-machine interfaces: W-HERBS. *Front. Neurosci*. 12, 511.

Meister M (2015). On the dimensionality of odor space. *eLife* 4, e07865.

Minderer M, Brown KD and Harvey CD (2019). The spatial structure of neural encoding in mouse posterior cortex during navigation. *Neuron* 102, 232-248.

Moreaux LC, Yatsenko D, Sacher WD, *et al.* (2020). Integrated neurophotonics: toward dense volumetric interrogation of brain circuit activity—at depth and in real time. *Neuron* 108, 66-92.

Moses DA, Leonard MK, Makin JG, Chang EF (2019). Real-time decoding of question-and-answer speech dialogue using human cortical activity. *Nat. Commun*. 10, 3096.

Muller J, Bakkum DJ, and Hierlemann A (2012). Sub-millisecond closed-loop feedback stimulation between arbitrary sets of individual neurons. *Front. Neural Circuits*, 6, 121.

Omer DB, Maimon SR, Las L, and Ulanovsky N (2018). Social place-cells in the bat hippocampus. *Science* 359, 218-224.

Orsborn AL and Pesaran B (2017). Parsing learning in networks using brain-machine interfaces. *Curr. Opin. Neurobiol*. 46, 76-83.

Pachitariu M, Stringer C, Schroder S, Dipoppa M, Rossi LF, Carandini M, and Harris KD (2017). Suite2p: beyond 10,000 neurons with standard two-photon microscopy. https://www.biorxiv.org/content/10.1101/061507v2 .

Pachitariu M, Steinmetz N, Kadir S, Carandini M and Harris KD (2016). Kilsort: real-time spike-sorting for extracellular electrophysiology with hundreds of channels. https://www.biorxiv.org/content/10.1101/061481v1.

Pandarinath, C. *et al*. (2018). Inferring single-trial neural population dynamics using sequential auto-encoders. *Nat. Methods* 15, 805-815.

Paninski L and Cunningham JP (2018). Neural data science: accelerating the experiment-analysis-theory cycle in large-scale neuroscience. *Curr. Opin. Neurobiol.* 50, 232-241.

Paz JT, Davidson TJ, Frechette ES, *et al.* (2013). Closed-loop optogenetic control of thalamus as a tool for interrupting seizures after cortical injury. *Nat. Neurosci.* 16, 164-70.

Pereira T, Aldarondo DE, Willmore L, *et al.* (2019). Fast animal pose estimation using deep neural networks. *Nat. Methods* 16, 117-125.

Ponce CR, Xiao W, Schade PF, *et al.* (2019). Evolving images for visual neurons using a deep generative network reveals coding principles and neuronal preferences. *Cell,* 177, 999-1009.

Potter, S. M., El Hady, A. and Fetz, E. E. (2014). Closed-loop neuroscience and neuroengineering. *Front. Neural Circuits* 8, 115.

Qiao S, Sedillo JI, Brown KA, Ferrentino B, and Pesaran B (2020). A causal network analysis of neuromodulation in the mood processing network. *Neuron* 107, 972-985.

Quian Quiroga, R., Reddy, L., Kreiman, G., Koch, C. and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature* 435, 1102-1107.

Raducanu BC, Yazicioglu RF, Lopez CM, *et al.* (2017). Time multiplexed active neural probe with 1356 parallel recording sites. *Sensors* 17, 2388.

Richards, B.A., Lillicrap, T.P., Beaudoin, P. *et al.* (2019). A deep learning framework for neuroscience. *Nat. Neurosci.* 22, 1761-1770.

Richardson AG, Ghenbot Y, Liu X, *et al.* (2019). Learning active sensing strategies using a sensory brain-machine interface. *Proc. Natl. Acad. Sci. USA* 116, 17509-17514.

Rios G, Lubenov EV, Chi D, *et al.* (2016). Nanofabricated neural probes for dense 3D recordings of brain activity. *Nano Lett.* 16, 6857-6862.

Robinson, A.K., Venkatesh, P., Boring, M.J. *et al.* (2017). Very high density EEG elucidates spatiotemporal aspects of early visual processing. *Sci. Rep.* 7, 16248.

Rumyantsev OI, Lecoq JA, Hernandez O, *et al.* (2020). Fundamental bounds on fidelity of sensory cortical coding. *Nature* 580, 100-105.

Sadtler, P.T., Quick, K.M., Golub, M.D.*, et al.* (2014). Neural constraints on learning. *Nature* 512, 423-426.

Sandler AJ (2008). Chronic recording during learning. In Nicolelis MAL, editor. *Methods for Neural Ensemble Recordings* (2nd edition). CRC Press.

Saxena S and Cunningham JP (2019). Towards the neural population doctrine. *Curr. Opin. Neurobiol*. 55, 103-111.

Scholvin, J., Kinney, J. P., Bernstein, J. G., *et al.* (2016). Close-packed silicon microelectrodes for scalable spatially oversampled neural recording. *IEEE Trans. Biomed. Eng.* 63, 120-130.

Schwartz DA, Lebedev MA, Hanson TL, *et al.* (2014). Chronic, wireless recordings of large-scale brain activity in freely moving rhesus monkeys. *Nat. Methods* 11, 670-676.

Shanechi MM, Orsborn AL and Carmena JM. (2016). Robust brain-machine interface design using optimal feedback control modeling and adaptive point process filtering. *PLoS Comput. Biol.* 12, e1004730.

Shanechi, M.M. (2019). Brain-machine interfaces from motor to mood. *Nat. Neurosci.* 22, 1554-1564.

Shenoy KV and Carmena JM (2014). Combining decoder design and neural adaptation in brain-machine interfaces. *Neuron* 84, 665-680.

Shi Y, Veidenbaum AV, Nicolau A, and Xu X (2015). Large-scale neural circuit mapping data analysis accelerated with the graphical processing unit (GPU). *J. Neurosci. Methods* 239, 1-10.

Shobe, J.L., Claar, L.D., Parhami, S., Bakhurin, K.I. and Masmanidis, S. C. (2015). Brain activity mapping at multiple scales with silicon microprobes containing 1,024 electrodes. *J. Neurophysiol*. 114*,* 2043-2052.

Silver D, Huang A, Maddison CJ, *et al.* (2016). Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 484-489.

Sofroniew NJ, Flickinger D, King J, and Svoboda K (2016). A large field of view two-photon mesoscope with subcellular resolution for in vivo imaging. *eLife* 5, e14472.

Sohrabpour, A., Cai, Z., Ye, S. *et al.* (2020). Noninvasive electromagnetic source imaging of spatiotemporally distributed epileptogenic brain sources. *Nat. Commun.* 11, 1946.

Srinivasan, S.S., Maimon, B.E., Diaz, M. *et al.* (2018). Closed-loop functional optogenetic stimulation. *Nat Commun* 9, 5303.

Sripad A, *et al*. (2018). SNAVA---A real-time multi-FPGA multi-model spiking neural network stimulation architecture. *Neural Netw*. 97, 28-45.

Steinmetz NA, Koch C, Harris KD, and Carandini M (2018). Challenges and opportunities for large-scale electrophysiology with Neuropixels probes. *Curr. Opin. Neurobiol*. 50, 92-100.

Stevenson, IH and Kording, KP (2011). How advances in neural recording affect data analysis. *Nat. Neurosci*. 14, 139-142.

Stirman JN, Smith IT, Kudenov MW, and Smith SL (2016). Wide field-of-view, multi-region, two-photon imaging of neuronal activity in the mammalian brain. *Nat. Biotechnol*. 34, 857-862.

Stringer C, Pachitariu M, Steinmetz N, Carandini M, and Harris KD (2019a). High-dimensional geometry of population responses in visual cortex. *Nature* 571, 361-365.

Stringer C, Pachitariu M, Steinmetz N, Reddy CB, Carandini M, Harris KD (2019b). Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, 364, eaav7893.

Sussillo, D., Stavisky, S., Kao, J. *et al.* (2016). Making brain-machine interfaces robust to future neural variability. *Nat. Commun* 7, 13749.

Swann NC, de Hemptinne C, Miocinovic S, *et al.* (2018a). Chronic multisite brain recordings from a totally implantable bidirectional neural interface: experience in 5 patients with Parkinson's disease. *J. Neurosurg*. 128, 605-616.

Swann NC, de Hemptinne C, Thompson MC, *et al.* (2018b). Adaptive deep brain stimulation for Parkinson's disease using motor cortex sensing. *J. Neural Eng*. 15, 046006.

Tafazoli S, MacDowell CJ, Che Z, *et al.* (2020). Learning to control the brain through adaptive closed-loop patterned stimulation. *J. Neural Eng*. 17, 056007.

Tsai PS, Mateo C, Field JJ, *et al.* (2015). Ultra-large field-of-view two-photon microscopy. *Opt. Express* 23, 13833-13847.

Tseng, P., Rajangam, S., Lehew, G. *et al.* (2018). Interbrain cortical synchronization encodes multiple aspects of social interactions in monkey pairs. *Sci. Rep.* 8, 4699.

Trautmann EM, Stavisky SD, Lahiri S, *et al*. (2019). Accurate estimation of neural population dynamics without spike sorting. *Neuron* 103, 1-17.

Tu M, Zhao R, Adler A, Gan WB, and Chen Z (2020). Efficient position decoding methods based on fluorescence calcium imaging in the mouse hippocampus. *Neural Computat.* 30, 1144-1167.

Tybrandt K, Khodagholy D, Dielacher B, *et al.* (2018). High-density stretchable electrode grids for chronic neural recordings. *Advanced Materials* 30, 1706520.

Vershynin, R (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Univ. Press.

Vlachos I, Deniz T, Aertsen A, and Kumar A (2016). Recovery of dynamics of function in spiking neural networks with closed-loop control. *PLoS Comput. Biol*. 12, e1004720.

Volgelstein JT, Park Y, Ohyama T, *et al.* (2014). Discovery of brainwide neural-behavioral maps via multiscale unsupervised structure learning. *Science,* 344, 386-392.

Wainwright, M (2019). *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Univ. Press.

Walker, E.Y., Sinz, F.H., Cobos, E. *et al.* (2019). Inception loops discover what excites neurons most using deep predictive models. *Nat. Neurosci.* 22, 2060-2065.

Watson, A. B. and Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33, 113-120.

Weisenburger S and Vaziri A (2018). A guide to emerging technologies for large-scale and whole-brain optical imaging of neuronal activity. *Ann. Rev. Neurosci.* 41, 431-452.

Williamson RC, Cowley BR, Litwin-Kumar A, *et al.* (2016). Scaling properties of dimensionality reduction for neural populations and network models. *PLoS Comput. Biol*. 12, e1005141.

Wu A, Roy NA, Keeley S and Pillow JW (2017). Gaussian process based nonlinear latent structure discovery in multivariate spike trains. *Adv. Neural Info. Proc. Syst. (NeuralPS)*, pp. 3499-3508.

Wu J, Liang Y, Chen S, *et al.* (2020). Kilohertz two-photon fluorescence microscopy imaging of neural activity in vivo. *Nat. Methods* 17, 287-290.

Yang W and Yuste R (2017). In vivo imaging of neural activity. *Nat. Methods*, 14, 349-359.

Yang Y and Shanechi MM (2016). An adaptive and generalizable closed-loop system for control of medically induced coma and other states of anesthesia. *J. Neural Eng.* 13, 066019.

Yang Y, Qiao S, Sani OG, *et al.* (2021). Motor-based prediction of large-scale brain network dynamic response to direct electrical stimulation. *Nat. Biomed. Eng*., https://doi.org/10.1038/s41551-020-00666-w

Yger P, Spampinato GLB, Esposito E, *et al*. (2018). A spike sorting toolbox for up to thousands of electrodes validated with ground truth recordings in vitro and in vivo. *eLife* 7, e34518.

Yu MB, Cunningham JP, Santhanam G, Ryu SI, Shenoy K V, and Sahani M (2009). Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *J. Neurophysiol.* 102, 614-635.

Zhang Q, Hu S, Talay R, *et al.* (2021). A closed-loop brain-machine interface for the study and treatment of pain. *Nat. Biomed. Eng.*, in press.

Zhou A, Santacruz SR, Johnson BC, *et al*. (2019). A wireless and artefact-free 128-channel neuromodulation device for closed-loop stimulation and recording in non-human primates. *Nat. Biomed. Eng*. 3, 15-26.

Zhou Z, Kuo H-C, Peng H, and Long F (2018). DeepNeuron: an open deep learning toolbox for neuron tracing. *Brain Inform*. 5, 3.

Zrenner C, Belardinelli P, Muller-Dahlhaus F, and Ziemann U (2016). Closed-loop neuroscience and non-invasive brain stimulation: a tale of two loops. *Front. Cell. Neurosci.* 10, 92.