The Dissertation Committee for Yin-Jui Chang Certifies that this is the approved version of the following dissertation:

Development of the multi-scale dynamic modeling techniques for multi-modal brain recordings

Committee:

Samantha R. Santacruz, Supervisor

José del R. Millán

Edward Castillo

Liberty Hamilton

Development of the multi-scale dynamic modeling techniques for multi-modal brain recordings

by

Yin-Jui Chang

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2023

Acknowledgements

I would like to thank my advisor, Dr. Samantha Santacruz, for her invaluable advice and support during my Ph.D. study. She made an excellent demonstration of how to survive as an assistant professor in a top 15 engineering school. She always guides me, and all other students as well, with her wisdom and her belief. I would like to thank Dr. José del R. Millán for his advice and support throughout my graduate study. My first meeting with him was impressive and his words have been imprinted in my mind and made a positive influence on my research. I would like to express my gratitude to Dr. Edward Castillo for his guidance and feedback on my main research project at UT Austin. I would also like to thank Dr. Liberty Hamilton for serving on my committee. With her suggestions and support, I can explore the potential of utilizing my biomedical research to clinical applications. I would like to thank Dr. Tim Yeh for his invaluable advice and support on my first co-authored publication at UT Austin. I would like to thank Dr. Thomas Yankeelov, Dr. Atlas Wang, Dr. James Sulzer, and Dr. Benito Fernandez for their advice. Without their support on my research, I would not have been able to carry through with the major accomplishments and publications.

I am so grateful having several excellent collaborators. Dr. Yuan-I Chen from Dr. Tim Yeh's lab led me in the field of fluorescence microscopy, allowing me to explore more fundamental biological research. Yuan-I and I have several brainstorming discussion throughout my graduate study and she can always provide me with the advice on the biomedical background to achieve my project goals. I also learn a lot of research and management skills from her. Dr. Jeff Liao from ISS, Inc. always replied to my request and questions immediately and debugged the fluorescence lifetime imaging system no matter

what time it was. Teamwork with these collaborators leads to numerous achievements and publications.

I feel very fortunate to have Dr. Enrique Contreras, Dr. Rodrigo Osuna-Orozco, Hannah Stealey, Yi Zhao, Hung-Yun Lu, Grace Jeanpierre, Megan Baker, Cole Barnett as lab mates. Enrique and Hannah taught me the key points when working with animals, and helped me to understand underlying behavioral tasks our animals are performing. Yi, Hung-Yun and I had several brainstorming discussions on our research projects. Grace's and Megan's solid background in fabrication helped me identify numerous challenges and current trends about the brain recording devices. Rodrigo and Cole helped me to identify the missing links in my research projects, allowing me to keep solidifying my projects. I would also like to thank my undergraduate over the years. Thanks to Kylie Hoyst for her work on computational modeling of spike trains and brain network graph. Without her effort, I would not have achieved so much. They genuinely care about each other and are willing to help whenever I need it. Working with such wonderful labmates made my Ph.D. life a fruitful journey.

Abstract

Development of the multi-scale dynamic modeling techniques for multi-modal brain recordings

Yin-Jui Chang, Ph.D.

The University of Texas at Austin, 2023

Supervisor: Samantha R. Santacruz

Fundamental principles underlying computation in multi-scale brain networks illustrate how multiple brain areas and their coordinated activity give rise to complex cognitive functions. Whereas brain activity has been studied in the micro- to meso-scale in studying the connections between the dynamical patterns and the behaviors, the investigation of the neural population dynamics is mainly limited by the single-scale analysis. My goal is to develop a multi-scale dynamical model for the collective activity of neuronal populations. First, I introduced a bio-inspired deep learning approach, termed NeuroBondGraph Network (NBGNet), to capture cross-scale dynamics that can infer and map the neural data from multiple scales (**Chapter 2**). The NBGNet not only exhibits more than an 11-fold improvement in reconstruction accuracy, but also predicts synchronous neural activity and preserves correlated low-dimensional latent dynamics. I also show that the NBGNet robustly predicts held-out data across a long time scale (two weeks) without

retraining. The effective connectivity defined from the presented model agrees with the established neuroanatomical hierarchy of motor control in the literature.

I then introduced the multi-scale neural dynamics neural ordinary differential equation (msDyNODE) to uncover multiscale brain communications governing cognitive behaviors (**Chapter 3**). I demonstrated that msDyNODE successfully captured multiscale activity using both simulations and electrophysiological experiments. The msDyNODE-derived casual interactions between recording channels and scales not only aligned well with the abstraction of the hierarchical anatomy of the mammalian nervous system but also exhibited behavioral dependences. This work offers a new approach for in depth mechanistic studies on the brain where simultaneous acquisition of multi-scale neural activity is available.

While the traditional neuronal tracing method serves as gold standard to assess the neural connectivity, toxicity and the limited efficiency prevent it from serving as a reliable tool to validate the multi-scale connectivity. It has been demonstrated that dendritic spines increase or become bigger after long-term potentiation while decrease or become smaller after long-term depression. Thus, we hypothesized the existence of excitatory and inhibitory connectivity can be revealed by the dynamics of dendritic spines. Since the dendritic spines can be smaller than the Abbe diffraction limit (~200 nm), we need an *in-vivo* super-resolution microscopy to achieve our goal. As a super-resolution imaging method, stimulated emission depletion (STED) microscopy has unraveled fine intracellular structures and provided insights into nanoscale organizations in cells. Although image resolution can be further enhanced by continuously increasing the STED-beam power, the resulting photodamage and phototoxicity are major issues for real-world applications of STED microscopy. In **Chapter 4**, I demonstrated that, with 50% less STED-beam power,

the STED image resolution can be improved up to 1.45-fold using the separation of photons by a lifetime tuning (SPLIT) scheme combined with a deep learning-based phasor analysis algorithm termed *flimGANE* (fluorescence lifetime imaging based on a generative adversarial network). This work offers a new approach for STED imaging in situations where only a limited photon budget is available.

In summary, I emphasized that this is the first time we are able to infer the nonlinear multiscale interactions in the brain networks by our multi-scale dynamic modeling approach. This opens a window of opportunity that multi-scale connectivity can provide a mechanistic understanding of brain computations underlying behaviors or mental states.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Overview of brain measurements

1.1.1 Single-modal measurement

The interface of the connection between human and system requires the technology of biosensing. Sensors detect the activity or events of targeting components and send information to a system for advanced signal conditioning, processing, and analysis. Brain activity can be detected from extracellular fields shaped by transmembrane current from several sources, including synaptic activity, fast action potentials, calcium spikes, intrinsic resonance, gap junctions[1]. Those electrical potentials can be recorded from many possible locations, including the scalp, on the surface of the cerebral cortex, deep in the brain, or on the surface of muscles.

Electroencephalography (EEG) is one of the commonly used approaches for detecting electrical activity in the human brain[2]. It is a safe and convenient method to record signals because of its noninvasive approach. The recording electrodes are placed on the scalp and thus surgery is not required. However, due to the distortion effects of the tissues between the signal source and the recording electrodes, it is impossible to detect the firing patterns of individual neurons. Newer technologies for EEG recording such as dry electrodes, ball cap-like head arrays, and wireless transmission of signals have made this technology more field deployable.

Magnetoencephalography (MEG) is another noninvasive approach for recording neural activity in the brain[3]. The tiny magnetic fields outside the skull generated by electrical currents in neurons can be measured by a superconducting quantum interference

device (SQUID). Since MEG provides high spatial and temporal resolution, it is considered useful for investigating human neurophysiology and information processing. However, MEG is not yet field-deployable and requires expensive technologies and a highly shielded chamber.

Another way to improve the spatial resolution of recording is through electrocorticography (ECoG), or intracranial electroencephalography (iEEG)[1]. This method uses stainless steel or platinum electrodes to record brain activity directly from the surface of depth of the cerebral cortex, which eliminates the distortion effect from the skull and intermediate tissue. Thus, ECoG provides better capability to map important functional areas of the brain.

To obtain even higher signal fidelity, multielectrode arrays (MEAs), also called microelectrode arrays, have been developed. Generally, they can be divided into three categories: microwire, micromachined, and flexible arrays[4]. Microwires are used to record individual neurons by applying them deep into the brain structure[5]. Micromachined arrays reduce the size of the array and provide higher spatial resolution. The "Michigan probe" is an example of micromachined arrays with electrodes placed on each shank for recording or stimulating in the central nervous system[6]. Another type of micromachined arrays is the "Utah arrays" that consist of conductive silicon needles electrically isolated from each other[7]. Due to its architecture, the Utah arrays are implanted easily into the cerebral cortex or peripheral nerves. However, most MEAs have a rigid structure, resulting in an undesired immune response and tissue encapsulation, which decreases signal quality over time[8]. Flexible arrays hold the promise of being a solution to these issues. "Soft" materials provide high flexibility and good biocompatibility, which allow electrodes to fit on the surface of the brain for chronic recording.

Beyond electrophysiology, neuroradiology is another category that monitors the brain activity. Magnetic resonance imaging (MRI) is a noninvasive medical imaging technique that produces detailed images of internal structure in the brain using the strong magnetic fields, magnetic field gradients and radio waves[9]. The extension of MRI captures different perspectives of the nervous system. Diffusion-weighted MRI can measure the restricted diffusion of water in tissue to produce neural tract images[10]. On the other hand, as the cerebral blood flow and neuronal activation are associated, functional MRI monitors brain activity *via* blood-oxygen-level dependent (BOLD) signals by detecting the changes in intravascular oxyhemoglobin concentration[11]. Unlike MRI, computed tomography (CT) employs the combination of X-rays and computer technology to generate the detailed internal images[12]. With the radioactive tracers, positron emission tomography (PET) scan provides functional information such as glucose or amino acid metabolism, dopamine receptors, amyloid, and tau deposits in the brain[13].

In order to enable longitudinal study of animal models, calcium imaging serves as a powerful tool to track activity of neurons over time[14,15]. It allows optical measurement of calcium concentration shift with neurons and neuronal tissue by using the fluorescent dyes that respond to the binding of Ca^{2+} ions by fluorescence properties.

1.1.2 Multi-modal measurement

As each modality has a unique set of strengths and weaknesses, multi-modal measurement, either recorded simultaneously or acquired separately but fused analytically, has become an emerging trend. For example, while the spiking activity and field potential refers to the neural activity at distinct spatial scales (individual neurons versus population of neurons),

they can be obtained by applying different post-processing pipelines from the measurement of MEAs.

Simultaneous measurement of EEG and fMRI (EEG-fMRI) has enabled the linking of electrical activity and hemodynamic responses since the mid-1990s[16,17]. EEG-fMRI provided complementary strengths in temporal and spatial resolutions. Furthermore, it can identify distinct neural signatures underlying the same behavior or in response to the same external stimuli. By characterizing the relationship between temporally precise electrical activity and spatially specific hemodynamic activity, EEG-fMRI can yield insights that would not be accessible within the single modality.

While fMRI and PET provide different functional measures of neural activity, where the former reveals regional time-varying changes in neural activity and the latter visualizes metabolic processes, the combination of both modalities allow us to study the relation between BOLD activation and the release of neurotransmitter[18–20].

Furthermore, trimodal measurements such as Spike-LFP-ECoG[21], EEG-fNIRS-fMRI[22], EEG-PET-fMRI[23], spike-LFP-fMRI[24], EEG-PET-MRI[25] have also been achieved by several research groups. With current advancement of sensing technology that allows multi-modal brain measurements, the integrative framework to analyze multi-modal dataset is required to fully leverage the benefits of simultaneous multi-modal measurements.

1.2 OVERVIEW OF DYNAMIC MODELING TECHNIQUES

1.2.1 Single-scale model analysis

Recent work has largely focused on developing robust statistical methods for obtaining the latent structure underlying high-dimensional data. This provides insight into the dynamics governing large-scale activity patterns. Gaussian-observation linear dynamical system model (GLDS) assumes the latent variable is governed by the linear dynamical system and the observed variables depend linearly on the state with the Gaussian noise[26]. Poisson linear dynamical system model (PLDS) instead assumes the observations conditioned on the latent variables are Poisson distributed[27]. By selecting the observation model as separate nonlinear dependence upon the latent variable, PfLDS was demonstrated outperforming PLDS in terms of predictive performance[28]. Furthermore, latent factor analysis via dynamical system (LFADS) employs recurrent neural networks (RNN) to produce the underlying dynamics[29]. Several recent studies have also posited the low-dimensional latent variable evolving in time according to a Gaussian process prior[30–33]. While most of the model were developed for spike trains, the observation model can be modified in order to apply to different brain activity.

1.2.2 Multi-scale model analysis

Source localization serves as standard cross-scale analysis, which goal is to localize the generator of the scalp measurements[34,35]. To enable precise localization, a large number of inverse solutions with multiple steps have been proposed. The first step is to use MRI to obtain anatomical information such as skull thickness. Then two solutions were widely used to solve the inverse problem (to localize the current dipoles): (**1**) dipole source localization which requires a priori assumption that only a few active brain areas generated

the measured EEG signals, and (**2**) distributed source localization methods such as low resolution electromagnetic tomography (LORETA)[36] and Local AutoRegressive Average (LAURA)[37], where the estimated locations of dipoles are restricted to the gray matter.

Cross-level coupling (CLC) is another cross-scale analysis that studies the statistical dependence between the micro-scale of single neurons and the meso- and macro-scale of oscillatory network activity[38]. Empirical findings of CLC reveal that the multi-scale coupling can be dynamically rebuilt to support new functional roles.

Recently, Shanechi and her group developed a multi-scale model-based Granger-like causality method to recover multi-scale neural causality during behavior[39]. By learning the point-process generalized linear models for the spike events and learning the linear Gaussian models for the field potential signals based on history of spike trains and field potential signals, they demonstrated their method to reveal the true multi-scale causality network structure in simulations.

1.3 OVERVIEW OF DISSERTATION STUDIES

The goal of this study is to develop a multi-scale modeling technique to reveal a comprehensive understanding of brain computation, where network mechanisms of multi-scale activity are critical.

In **Chapter 2**, the development and implementation of cross-scale dynamical model, NBGNet, are discussed. I also analytically validate our NBGNets by yielding small root mean squared errors; reproducing features commonly seen in neuroscientific analyses (cross-correlation, phase synchrony); capturing cross-scale interactions aligning well with the abstraction of the hierarchical anatomy of the mammalian nervous system;

reconstructing low-dimensional latent dynamics; inferring details of behavior; and predicting out-of-sample conditions. For all results in this chapter, I train NBGNets without any information about task conditions or behavioral parameters (e.g., real kinematics or eye-tracker data) and present the results from testing data.

In **Chapter 3**, I extend cross-scale dynamic modeling to multi-scale dynamic modeling, which captures within- and cross-scale neural dynamics all at once. The chapter covers the development of msDyNODE, including the selection of dynamic systems and the implementation of neural ordinary differential equations (NODE), a new type of deep learning method to model the continuous dynamic of hidden units. Validation of fitting performance is performed with both synthetic and experimental data. In addition, in this chapter, the multi-scale effective connectivity is also demonstrated exhibiting both common and unique patterns underlying behavioral tasks.

In **Chapter 4**, I develop a new method, termed STED-*flimGANE*, for the stimulated emission depletion (STED) microscopy. Current issue for *in-vivo* STED imaging and its solution has been discussed. I demonstrate the spatial resolution enhancement using two samples: fluorescent beads and nuclear pores.

This work details the development of msDyNODE, starting from cross-scale dynamics inference toward multi-scale dynamics inference. The findings inspire us to further invent a new method to validate the characterized multi-scale connectivity in the brain, in which we need to tackle the challenges of STED imaging in live cells. Furthermore, *in-vivo* STED imaging also holds great potential to studying the dynamics of the synaptic components in the nervous system at micro-scale, providing a new approach to access neural dynamics that has not been achievable yet.

Chapter 2: Cross-scale dynamic modeling using neurobiologically realistic recurrent neural networks

2.1 INTRODUCTION

Billions of individual neurons coordinate activity at multiple scales, either directly or indirectly, to drive behaviors such as motor preparation[40,41], motor adaptation[42], motor timing[43,44], decision-making[45], and working memory[46,47]. However, current techniques for capturing neural population dynamics are mainly limited by the single-scale analysis, typically with the simplified assumptions of linear[27] or log-linear[30] dynamics. While recurrent neural networks (RNNs) have been introduced to infer nonlinear latent dynamics that encode rich information giving rise to motor behavior[29], we lack a broadly accepted approach to explore cross-level activity for a deeper understanding of system-level nonlinear neural mechanisms[48,49]. Since the brain exhibits computational structure across a variety of scales: from single neurons (micro-scale) to functional areas (meso-scale) to cortical networks (macro-scale), a tool that can uncover multi-scale dynamics is critically important for illuminating the mechanistic understanding of brain activity[50].

Until recently, only a limited number of studies focused on cross- or multi-scale interactions in brain networks. For example, source localization (e.g., sphere head model[51]) aims to identify the brain areas or individual neurons generating the recorded electrical potentials such as electroencephalography[35]. However, the requirements of high-density recordings, unrealistic assumptions, and uncertainty on conductivity value[52] limit the fidelity of experimental data. In addition, cross-level coupling (CLC)[38] has shown evidence of cross-scale interactions between single neurons and oscillatory network activity. In contrast, no information about how the activity communicates across levels is provided. Recent work developed a generalized linear model-based method to reveal the directed

interactions across spatiotemporal scales of brain activity[39]. Nevertheless, brain dynamics are characterized by nonlinear coupling among neuronal populations[53,54]. Linear model-based approach may fail to capture the associated nonlinearity in the multi-scale brain networks.

The bond Graph (BG) is a graphical approach widely used to model multi-domain dynamical systems (e.g., electrical, fluid, mechanical, magnetic, thermal, and hydraulic) via energy exchange[55]. BG allows a compact and explicit representation of the complex system and provides analogous applicability to different domains using the common constitutive relations. With the analogy between multi-domain modeling and multi-scale modeling, we extend the BG approach to model multi-scale dynamical systems in brain networks, yielding a neurobiological-inspired state-space model with a priori knowledge of signal translations between multi-scale signals. Ultimately, we leverage the NeuroBondGraph Network (NBGNet)[56], a deep learning framework consisting of recurrent neural networks (RNNs) and multi-layer perceptrons (MLPs), to capture the temporal evolution and the nonlinearity of the system dynamics. Unlike source localization, incorporating neurobiological knowledge (specifically tissue electrical impedance) eliminates bias due to unrealistic assumptions (e.g., homogeneous tissue conductivity and ignorance of tissue capacitance). Compared to CLC, the NBGNet models the causal contributions which describe how individual and populations of neurons communicate in a cross-scale network. While purely data-driven methods, such as generalized linear models or black-box RNNs, may achieve similar performance, the NBGNet approach provides rigorous interpretability to evaluate both within- and cross-scale causal interactions.

The NBGNet model is universal in that it can be used for any combination of neural activity at different scales (or even the same scale) with the appropriate modification to the BG structure and its derived dynamic equations. To demonstrate the power of our approach, we employ two specific types of simultaneously recorded real neural data in this work. Namely, we use local field potentials (LFPs; in the spatial scale of 10^{-4} to 10^{-5} m) and signals recorded from intracranial "screw-type" macroelectrodes implanted in the skull (screw electrocorticography or screw ECoG[57,58]; in the spatial scale of 10^{-2} to 10^{-4} m) acquired from a rhesus macaque performing a simple motor task. Screw ECoG, rather than electroencephalography, is chosen due to its improved signal-to-noise ratio and stability. The structure of the NBGNet for these two particular data types is easily extended to other field potential signals, as well as spiking data with minor modifications.

We demonstrate that the NBGNet provides superior reconstruction accuracy with up to 11.1-fold decrease in root mean square error and 1.8- and 1.4-fold stronger similarity in time- and phase-domain compared to alternative methods. We show the NBGNet-derived causal interactions align well with the neuroanatomical hierarchy of motor control[59], demonstrating the interpretability of the model structure. We further validate the capability of the NBGNet to capture and reconstruct single-trial low-dimensional neural dynamics. Behavioral variables can also be detected by NBGNet-predicted activity as accurately as using empirical measurements. Finally, we examine the stability of the NBGNet and reveal that the learned dynamical system maintains predictive power over more than two weeks without model retraining.

2.2 DATA ACQUISITION

2.2.1 Experimental subjects and behavioral tasks

A male rhesus macaque is used in these experiments. The macaque is trained to perform a center-out task (**Figure 2.1 e-f**). Briefly, the subject is trained to use a joystick to move a cursor on a computer screen from a center target to a peripheral target. The joystick is attached to the front of the primate chair and the subject is free to use either hand to control the joystick during the experiment. In the task, the subject is trained to hold the cursor at the center target shown on the screen for 320 ms. Then the subject is presented with one of the eight outer targets, equally spaced in a circle, and selected randomly with uniform probability. The subject moves the cursor to the peripheral target and holds the cursor inside the target for 320 ms. A trial is successful if the subject completes the 320 ms center-hold followed by holding at the peripheral target for 320 ms. The reward is scheduled after a successful trial, where a custom-programmed Arduino triggered the reward system to deliver a small amount of juice to the subject. All experiments were approved by the Animal Care and Use Committee at the University of California, Berkeley. The subject was approximately 6 years of age at the time of data collection.

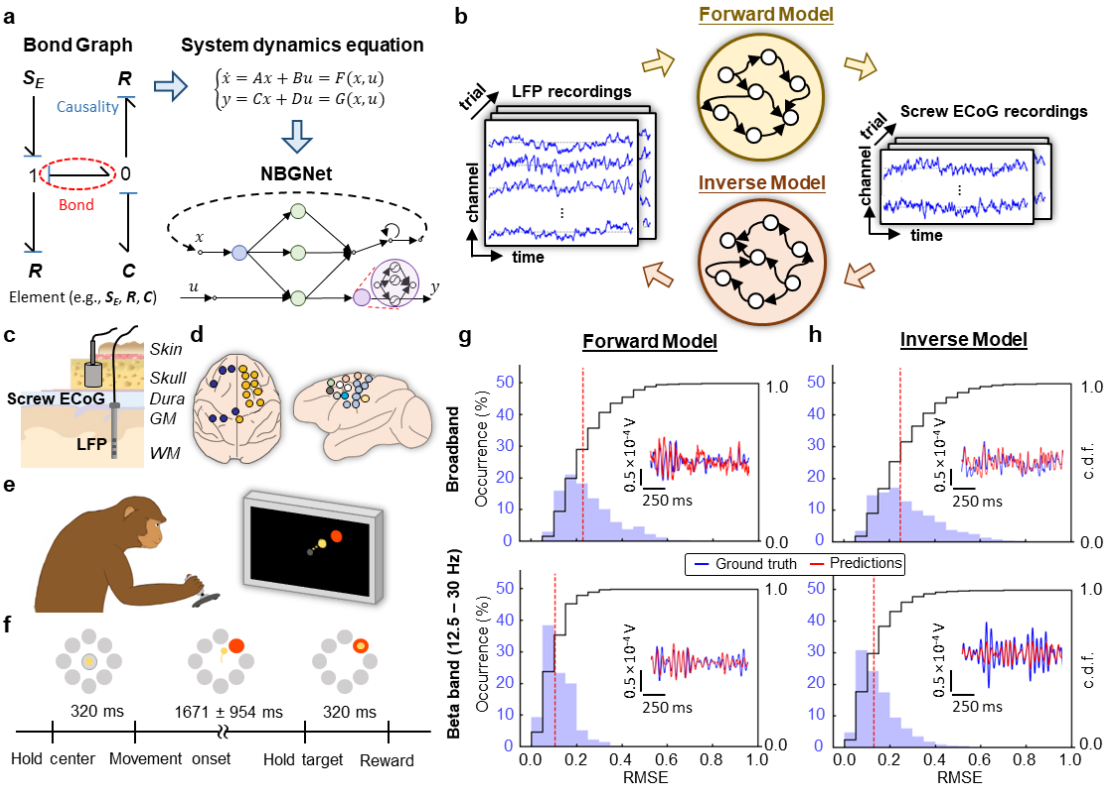

Figure 2.1 | The NBGNet is a neurobiologically realistic recurrent neural network that utilizes nonlinear dynamics to model the translation between multi-scale brain activities. (**a,b**) Schematic overview of the NBGNet architecture for forward and inverse modeling between LFP and screw ECoG. Details are provided in the main text. (**c**) Spatial relationships between LFP and screw ECoG. (**d**) Screw ECoG signals were recorded across both hemispheres and LFP data was acquired from one hemisphere. Only 16 LFP channels were shown as a schematic. (**e**) Monkey performed a center-out reaching task using a joystick. (**f**) Schematic of protocol for the experiments. (**g,h**) Histogram and cumulative distribution function (c.d.f.) of RMSE in broadband and beta band (12.5 – 30 Hz) for forward model (**g**) and inverse model (**h**; red dashed line: median). The insets show the representative examples with the RMSE close to the red dashed line.

2.3 NEUROBONDGRAPH NETWORK (NBGNET)

2.3.1 Modeling of tissues between sources

The interactions between measurements are modeled based on the physiology of brain tissue and its effect on the electrical signal flow. In this work, screw ECoG signals are recorded within the skull while LFP signals are measured within the cortical and subcortical structures (**Figure 2.1 c-d**). Therefore, the biological medium between the recording locations consists of skull, dura mater, and/or cortex. We then model the signal translations as an electrical circuit with the LFP as the source, the brain tissues as effective impedance, and the screw ECoG as the voltage measurement. Since the skull contains sinus cavities and numerous foramina, a three-layer structure is utilized: a spongy bone layer in the middle of two compact bone layers. The cavities in the spongy bone are modeled as a capacitance that provided potentials inside them. In contrast, the compact bone and the trabeculae of the spongy bone are modeled as resistances. All potential paths for electrical signals to travel are considered to model the signal propagation. As a thick membrane surrounding the brain, dura mater is represented with the effective resistance and capacitance in parallel. Although the cortex is composed of folded grey matter, we model it as an effective resistance to simplify the complexity.

2.3.2 Bond graph modeling

BG is a graphical representation of a physical system that allows easy access to the state-space representation. BG consists of the bonds and the elements (**Figure 2.1 a**). The bonds represent the power, and each of them has two features: half-arrow and causality. The power is broken down into two pairs: flow (e.g., current in electrical domain) and effort (e.g., voltage). Half-arrow indicates the sign convention for the work being done.

Accordingly, sources will always have the arrow pointing away from the element, while others will have the arrow pointing into the elements. Causality in BG denotes which side of the bond governs the instantaneous power. There are multiple categories for elements, including (1) sources, denoted as **S**, serving as the input to the system, (2) sinks, denoted as **S** as well while serving as the output of the system, (3) inertia elements (e.g., inductance), denoted as **I**, which store energy, (4) resistance elements (e.g., resistance), denoted as **R**, which dissipate energy, (5) compliance elements (e.g., capacitance), denoted as **C**, which store potential energy, and (6) 0- or 1-junctions which split the power. Specifically, 0-junctions are that all efforts are equal across the bonds and the sum of flow in equals to the sum of flow out. In contrast, 1-junctions represent that all flows are equal across the bonds and the sum of effort in equals to the sum of effort out. Furthermore, as two passive components, **I** and **C**, exhibit time-dependence behavior, there exists preferred causal orientations with **C** defining the effort and **I** defining the flow. Since energy in different domains can be transferred into each other with a constant, BGs enable modeling of physical systems in distinct domains.

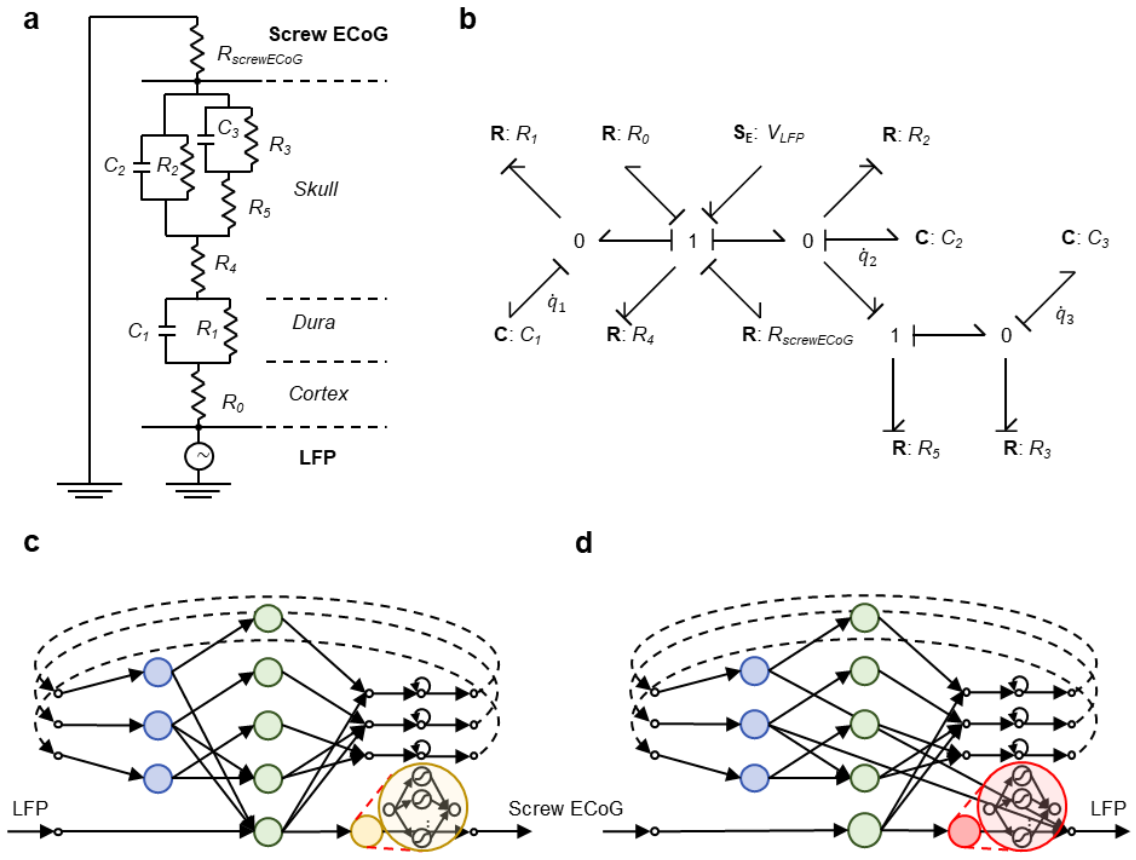

Figure 2.2 | Development of the NBGNet. (**a**) LFP-screw ECoG transmission electrical circuit was established based on the effective electrical signal pathway. (**b**) Bond Graph of the physical system illustrated in **a**. (**c-d**) Both forward- and inverse-NBGNet are derived from the system dynamics equations for the LFP-screw ECoG transmission model. (**c**) Schematic of forward-NBGNet architecture, where the colored circles represent a multi-layer perceptron unit. (**d**) Same as **c** for inverse-NBGNet architecture.

2.3.3 Forward and inverse modeling

Here, we develop the state-of-the-art deep learning technique termed NBGNet to approximate the unknown nonlinear relationship. In NBGNet, MLP units are adopted to approximate the nonlinearity of the brain dynamics. The RNN framework is then utilized

to capture the cross-scale interactions by maximizing the likelihood of the observed brain signals with its internal states. The modeling of brain tissue impedances makes NBGNets neurobiologically realistic to analyze neural signals. Here we demonstrate the capability of extracting bidirectional cross-scale dynamics using NBGNets for forward and inverse models, respectively.

Combing together, LFP-screw ECoG transmission electrical circuit can be established (**Figure 2.2a**). With the effective electrical circuit for electrical signal pathways, Bond Graph is then generated (**Figure 2.2b**). The compliance components ***C*** indicate the hidden state variables in the dynamic equations. Ultimately, we obtain a 3rd order ordinary differential equation describing the dynamics underlying multi-scale system based on the constitutive equation for each element and connection. The ordinary differential equations for the system are derived as follows:

$$\dot{q}_1 = \dot{q}_2 + \frac{q_2}{R_2C_2} + \frac{1}{R_5}\left(\frac{q_2}{C_2} - \frac{q_3}{C_3}\right) - \frac{q_1}{R_1C_1}$$

$$\dot{q}_2 = \frac{1}{1+\frac{R_0+R_{screwECoG}}{R_4}}\left(\frac{1}{R_4}\left(V_{LFP} - \frac{q_1}{C_1} - \frac{q_2}{C_2} - (R_0 + R_{screwECoG})\left(\frac{q_2}{R_2C_2} + \frac{1}{R_5}\left(\frac{q_2}{C_2} - \frac{q_3}{C_3}\right)\right)\right) - \frac{q_2}{R_2C_2} - \frac{1}{R_5}\left(\frac{q_2}{C_2} - \frac{q_3}{C_3}\right)\right) \quad (2.1)$$

$$\dot{q}_3 = \frac{1}{R_5}\left(\frac{q_2}{C_2} - \frac{q_3}{C_3}\right) - \frac{q_3}{R_3C_3}$$

$$V_{screwECoG} = R_{screwECoG}\left(\frac{1}{R_0+R_4+R_{screwECoG}}\left(V_{LFP} - \frac{q_1}{C_1} - \frac{q_2}{C_2}\right)\right) \quad (2.2)$$

While the equations above represent the ideal condition where the resistance *R* and the capacitance *C* are linear. Considering the uncertainty and continuous changing of human's brain tissue, nonlinearity is introduced in the equation:

$$\begin{aligned}\dot{q}_1 &= F_{R_T}^{-1}\left(u - F_{C_1}^{-1}(q_1) - F_{C_2}^{-1}(q_2)\right) - F_{R_1}^{-1}\left(F_{C_1}^{-1}(q_1)\right)\\ \dot{q}_2 &= F_{R_T}^{-1}\left(u - F_{C_1}^{-1}(q_1) - F_{C_2}^{-1}(q_2)\right) - F_{R_2}^{-1}\left(F_{C_2}^{-1}(q_2)\right) - F_{R_5}^{-1}\left(F_{C_2}^{-1}(q_2) - F_{C_3}^{-1}(q_3)\right)\\ \dot{q}_3 &= F_{R_5}^{-1}\left(F_{C_2}^{-1}(q_2) - F_{C_3}^{-1}(q_3)\right) - F_{R_3}^{-1}\left(F_{C_3}^{-1}(q_3)\right)\end{aligned} \quad (2.3)$$

$$y = F_{R_{screwECoG}}\left(F_{R_T}^{-1}\left(u - F_{C_1}^{-1}(q_1) - F_{C_2}^{-1}(q_2)\right)\right) \quad (2.4)$$

where, R_T represents $R_0 + R_4 + R_{screwECoG}$, u represents V_{LFP}, y represents $V_{screwECoG}$, and $F(\cdot)$ is a nonlinear function to be determined.

Therefore, the multi-variable time varying Bond Graph forward model, Equations 2.3-2.4, can be expressed as the state-space representation,

$$\dot{x} = \mathbf{A}x + \mathbf{B}u \quad (2.5)$$

$$y = \mathbf{C}x + \mathbf{D}u \quad (2.6)$$

where, $x = [q_1,\ q_2,\ q_3]^T$, $u = V_{LFP}$, $y = V_{screwECoG}$,

$$\mathbf{A} = \begin{bmatrix} \left(\frac{-1}{R_{screwECoG}+R_0+R_4} + \frac{-1}{R_1}\right)\frac{1}{C_1} & \left(\frac{-1}{R_{screwECoG}+R_0+R_4}\right)\frac{1}{C_2} & 0 \\ \left(\frac{-1}{R_{screwECoG}+R_0+R_4}\right)\frac{1}{C_1} & \left(\frac{-1}{R_{screwECoG}+R_0+R_4} + \frac{-1}{R_2} + \frac{-1}{R_5}\right)\frac{1}{C_2} & \frac{1}{R_5 C_3} \\ 0 & \frac{1}{R_5 C_2} & \left(\frac{-1}{R_5} + \frac{-1}{R_3}\right)\frac{1}{C_3} \end{bmatrix} \quad (2.7)$$

$$\mathbf{B} = \begin{bmatrix} \frac{1}{R_{screwECoG}+R_0+R_4} & \frac{1}{R_{screwECoG}+R_0+R_4} & 0 \end{bmatrix}^T \quad (2.8)$$

$$\mathbf{C} = \begin{bmatrix} \frac{-R_{screwECoG}}{(R_{screwECoG}+R_0+R_4)C_1} & \frac{-R_{screwECoG}}{(R_{screwECoG}+R_0+R_4)C_2} & 0 \end{bmatrix} \quad (2.9)$$

$$\mathbf{D} = \begin{bmatrix} \frac{R_{screwECoG}}{R_{screwECoG}+R_0+R_4} \end{bmatrix} \quad (2.10)$$

The inversion algorithm for multi-variable system were obtained by the following,

$$\begin{aligned}\dot{q}_1 &= -\frac{q_1}{R_1C_1} + \frac{V_{screwECoG}}{R_{screwECoG}} \\ \dot{q}_2 &= -\frac{q_2}{R_2C_2} - \frac{1}{R_5}\left(\frac{q_2}{C_2} - \frac{q_3}{C_3}\right) + \frac{V_{screwECoG}}{R_{screwECoG}} \\ \dot{q}_3 &= \frac{1}{R_5}\left(\frac{q_2}{C_2} - \frac{q_3}{C_3}\right) - \frac{q_3}{R_3C_3}\end{aligned} \tag{2.11}$$

$$V_{LFP} = \frac{q_1}{C_1} + \frac{q_2}{C_2} + \frac{R_{screwECoG}+R_0+R_4}{R_{screwECoG}}\, V_{screwECoG} \tag{2.12}$$

As forward model, nonlinearity is introduced in the equation as well,

$$\begin{aligned}\dot{q}_1 &= F^{-1}_{R_{screwECoG}}(y) - F^{-1}_{R_1}\left(F^{-1}_{C_1}(q_1)\right) \\ \dot{q}_2 &= F^{-1}_{R_{screwECoG}}(y) - F^{-1}_{R_2}\left(F^{-1}_{C_2}(q_2)\right) - F^{-1}_{R_5}\left(F^{-1}_{C_2}(q_2) - F^{-1}_{C_3}(q_3)\right) \\ \dot{q}_3 &= F^{-1}_{R_5}\left(F^{-1}_{C_2}(q_2) - F^{-1}_{C_3}(q_3)\right) - F^{-1}_{R_3}\left(F^{-1}_{C_3}(q_3)\right)\end{aligned} \tag{2.13}$$

$$u = F^{-1}_{C_1}(q_1) + F^{-1}_{C_2}(q_2) + F_{R_T}\left(F^{-1}_{R_{screwECoG}}(y)\right) \tag{2.14}$$

where, R_T represents $R_0 + R_4 + R_{screwECoG}$, u represents V_{LFP}, y represents $V_{screwECoG}$, and $F(\cdot)$ is a nonlinear function to be determined.

Here we develop the deep learning technique termed NBGNet to approximate the unknown nonlinear relationship. In NBGNet, the network implements the causal form of the dynamic equations where the unknown nonlinear mappings are realized by the MLP units. The RNN framework is then utilized to capture the cross-scale interactions by maximizing the likelihood of the observed brain signals with its internal states. The modeling of brain tissue impedances makes NBGNets neurobiologically realistic to

analyze neural signals. Here we demonstrate the capability of extracting bidirectional cross-scale dynamics using NBGNets for forward and inverse models, respectively. The evolution of latent variables and the output is described by the nonlinear functions approximated by NBGNets.

$$\boldsymbol{V}_{screwECoG,t}, \boldsymbol{q}_{t+1}^{forward} = NBG^{forward}(\boldsymbol{q}_t^{forward}, \boldsymbol{V}_{LFP,t}) \tag{2.15}$$

$$\boldsymbol{V}_{LFP,t}, \boldsymbol{q}_{t+1}^{inverse} = NBG^{inverse}(\boldsymbol{q}_t^{inverse}, \boldsymbol{V}_{screwECoG,t}) \tag{2.16}$$

where $\boldsymbol{q}$ represents the latent states of the system, and $\boldsymbol{V}$ represents the electrical recordings. The forward-NBGNet serves as a forward solution that models the single-trial screw ECoG as a nonlinear recursive mapping from the multivariate LFP (**Figure 2.1 a-b**). The network's units to approximate such a mapping depend on three elements: a trial-specific initial state, input signals, and the parameters defining the system dynamics. To mimic the real-time modeling and abide by causality constraints, the network only runs through the trial forward for estimation. By inverting the forward solution, the inverse-NBGNet is then developed to predict LFP from screw ECoG (**Figure 2.1b**). As inverse computation is an ill-posed problem which can lead to a non-unique and unstable solution[34], we expect a relatively poorer performance as compared with the forward solution.

To optimize the NBGNet, we train our model to minimize the mean-squared-error between predicted activity and the ground-truth using simultaneously recorded LFP data from the left hemisphere and screw ECoG data from both hemispheres. The major hyperparameters for forward and inverse model are the number of hidden nodes in the MLP unit for nonlinear mapping estimation and the time step. For both forward and inverse model, 7 nodes are utilized in MLP units, and the time step of NBGNet is equal to the data

sampling rate. Over-fitting occurs when we train the model with the same batch of data for excessive iterations. To avoid over-fitting, we select different trials of data for training when NBGNet has been updated for twenty times. The NBGNets are randomly initialized by Glorot uniform initializer and optimized using adaptive moment estimation (Adam) optimizer with a starting learning rate of 1 x 10^{-3}. A portion of the data serve as the validation set and to determine if the model was overfit. Here we used a ratio of 9:1 between training and validation (held-out) data. After model training, the parameters of the NBGNet remain fixed for further analysis.

2.4 MODEL VALIDATION

To demonstrate how well the proposed method reconstructs the ground-truth signals and captures key characteristics in neural signals, we utilize several metrics, including root mean square error, similarity, phase synchrony, and decoding capabilities. Additionally, we compare our approach with two existing algorithms, sphere head model and RNN.

2.4.1 Evaluation metrics – Scale dependent analysis

To evaluate how close the model predictions are to the ground-truth signals, root mean square error (RMSE) is commonly used to indicate the absolute fit of the model. RMSE is defined as the square root of the mean of the square of the error,

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left[\boldsymbol{Y}_{gt}(t) - \boldsymbol{Y}_{pre}(t)\right]^2} \tag{2.17}$$

where $\boldsymbol{Y}_{gt}$ represents ground-truth measurement, $\boldsymbol{Y}_{pre}$ represents the model prediction, and T is the number of time points in the given trial.

We evaluate the prediction accuracy of the presented NBGNet by calculating the RMSE with the acquired broadband signals over 150 individual reach trials (**Figure 2.1**). Since the beta frequency band (12.5–30 Hz) is strongly implicated in motor behaviors[60,61], we also examine the performance specifically within the beta band activity (**Figure 2.1g**). Gated recurrent unit based RNN (GRU-RNN) is utilized as baseline for quantitative comparison. For the forward solution, the NBGNet yielded 17% and 42% higher accuracy than GRU-RNN and sphere head model in trial-wise (RMSE = 0.12 ± 0.06 for NBGNet, 0.14 ± 0.06 for GRU-RNN, and 0.17 ± 0.07 for GRU-RNN; mean ± s.d. in the unit of 10^{-4} V), 7% and 53% in session-averaged comparison (RMSE = 0.15 for NBGNet, 0.16 for GRU-RNN, and 0.23 for sphere head model), respectively.

We also assess the capability of reconstructing LFP using inverse-NBGNet and screw ECoG recordings (**Figure 2.1h**). Similarly, inverse-NBGNet outperform GRU-RNN and sphere head model in both trial-wise (RMSE = 0.15 ± 0.10 for NBGNet, 0.17 ± 0.09 for GRU-RNN, and 1.68 ± 8.03 for sphere head model) and session-averaged comparisons (RMSE = 0.19 for NBGNet, 0.20 for GRU-RNN, and 2.42 for sphere head model). Interestingly, similarly small reconstruction error reveals that the inverse-NBGNet is able to transform the lower-dimensional screw ECoG into the higher-dimensional LFP.

2.4.2 Evaluation metrics – Similarity analysis

Similarity of two time series signals also measures whether two time series signals exhibit similar shape of oscillation. Here we use the Pearson correlation coefficient to measure how highly linearly correlated two time series signals are:

$$\rho(\boldsymbol{Y}_{gt}, \boldsymbol{Y}_{pre}) = \frac{\sum \boldsymbol{Y}_{gt}\boldsymbol{Y}_{pre} - \frac{\sum \boldsymbol{Y}_{gt}\sum \boldsymbol{Y}_{pre}}{T}}{\sqrt{\left(\sum \boldsymbol{Y}_{gt}^2 - \frac{\left(\sum \boldsymbol{Y}_{gt}\right)^2}{T}\right)\left(\sum \boldsymbol{Y}_{pre}^2 - \frac{\left(\sum \boldsymbol{Y}_{pre}\right)^2}{T}\right)}}. \quad (2.18)$$

Similarity of oscillation dynamics is an alternative approach to evaluate the integrity of predicted signals using cross-correlations computed on a single-trial single-channel basis. From the representative session (**Figure 2.3a-b**), NBGNet-predicted signals from most of the channels are moderately to strongly correlated with the ground-truth signals (average correlation greater than 0.4[62,63]). Strong correlation (correlation coefficient > 0.6) is found on 63% of channels. Interestingly, due to movement-induced activation, channels at anterior brain regions exhibit greater correlation than those at posterior brain regions. The predicted screw ECoG match well with the raw screw ECoG in both trial-wise (**Figure 2.3c**) and session-averaged comparison (**Figure 2.3d**). We note that the performance is relatively poor during 0.3-0.4 s when the subject is searching for the correct direction of cursor's movement. However, the performance is better in the remainder of the time interval of interest when the direction of movement aligns with the target direction.

We also examine the correlation between the inverse-NBGNet-inferred and the ground-truth LFPs. Channel 29 (white matter) provides the highest correlation as 0.90 ± 0.07 (mean ± s.d.; **Figure 2.3e-f**); whereas channel 82 (M1) exhibits no correlation (0.00 ± 0.30) due to unexpectedly larger amplitude. However, the predicted signals on most of the other channels are moderately correlated with ground-truth activity. Strong correlations are found on 54% of channels. As the more lateral brain recording sites are also further away from the surface (compared with more medial regions) where screw ECoG was recorded, the channels in these regions show smaller correlations (**Figure 2.3b**). In

summary, the correlation analysis confirms the NBGNet's ability to capture the beta-frequency dynamic features.

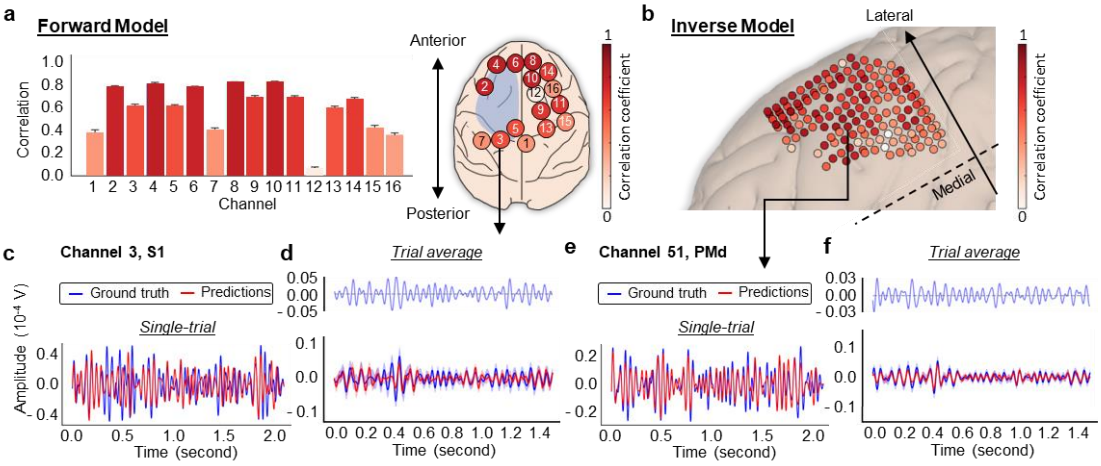

Figure 2.3 | Cross-correlation analysis indicated the similarity between NBGNet inference and ground-truth recordings. (**a**) Average correlation coefficient across all the trials (error bars, s.d.; n = 150). Screw ECoG electrodes layout labeled with the channel number. Blue shaded area represents the coverage of LFP channels. (**b**) Same as (**a**) for the inverse model. A screw ECoG channels (3: S1) was selected for single trial-based comparison (**c**): ground truth (blue trace) versus model prediction (red trace) in the 3rd trial, and grand average-based comparison (**d**): ground truth (blue trace; mean ± s.e.m.) versus network output (red trace; mean ± s.e.m.) and the corresponding error trace (top). (**e,f**) Same as (**c**,**d**) for representative comparison for the inverse model (51: PMd)

2.4.3 Evaluation metrics – Phase analysis

Phase-domain reveals other characteristics that are not visible in time-domain. Phase synchronization between neurons is a fundamental neural mechanism that supports neural communication and plasticity[64]. Given a pair of signals, $s_1(t)$ and $s_2(t)$, which have been band-pass filtered to a frequency range of interest, the Hilbert transform, HT[•], is applied to obtain the corresponding analytical signals, $z_1(t)$ and $z_2(t)$:

$$\boldsymbol{z}_i(t) = \boldsymbol{s}_i(t) + j\,\boldsymbol{HT}[\boldsymbol{s}_i(t)] = \boldsymbol{A}_i(t)e^{j\boldsymbol{\phi}_i(t)} \tag{2.19}$$

$$\boldsymbol{HT}[\boldsymbol{s}_i(t_k)] = \boldsymbol{s}_i(t_k) * \frac{1}{2\pi}\left[\int_{-\pi}^{0} j \cdot e^{jwk}dw - \int_{0}^{\pi} j \cdot e^{jwk}dw\right] \tag{2.20}$$

where k = 1 to T, A_i(t) represents the instantaneous amplitude, and ϕ_i(t) represents the instantaneous phase. In order to obtain a comprehensive view, we utilize two metrics: phase-locking value and phase synchrony index. Phase locking value (PLV)[65] or so-called mean phase coherence[66], is defined as,

$$PLV = \left|\frac{1}{T}\sum_{t=0}^{T-1} e^{j(\Delta\boldsymbol{\phi}(t))}\right| \tag{2.21}$$

where $\Delta\phi$(t) represents the phase difference between pair of signals.

This metric characterizes the intra-trial variability of the phase difference between two signals, where a larger PLV indicates a stronger synchrony between them. In addition, the phase of phase-locking can be extracted to evaluate the mean phase difference across time.

In addition to the PLV, we are also interested in the instantaneous performance, and thus we consider phase synchrony index. First, provided with the instantaneous phase of two time series signals, ϕ_1(t) and ϕ_2(t), the instantaneous phase synchrony (IPS)[67], which measured the phase similarity at each timepoint, is calculated by

$$IPS(t) = 1 - sin\left(\frac{|\boldsymbol{\phi}_1(t) - \boldsymbol{\phi}_2(t)|}{2}\right) \tag{2.22}$$

where the phase is in the unit of degree. IPS spans the range of 0 – 1, where a larger value indicates a stronger synchrony. We define a quarter of the whole range of phase difference (180^o), 45^o, as the threshold. When the phase difference is less than 45^o, IPS was greater

than 0.62, thus revealing a better performance. We then calculated the ratio of the time with the IPS greater than 0.62, termed phase synchrony index (PSI; **Figure 2.4**),

$$PSI = \frac{t_{IPS>0.62}}{T} \tag{2.23}$$

To determine the level of the phase synchrony, we categorized the two-dimensional scatter plot of PSI and PLV into four sections with both thresholds as 0.5: Zone 1 (low PSI and low PLV) indicates poor synchronization, Zone 2 (low PSI and high PLV) and Zone 3 (high PSI and low PLV) indicates medium synchronization, and Zone 4 (high PSI and high PLV) indicates very good synchronization (**Figure 2.4**).

As phase-domain coherence is an important tool to determine the functional connectivity in brain networks, we examine whether the predicted and the recorded signals were phase-synchronized. PLV has been widely used to measure the inter-trial variability of phase difference, where 1 represents no change in phase difference and 0 reflects the opposite case[65,66]. To assess the intra-trial variability, we adapt PLV by averaging the phase difference across the time rather than the trials (**Figure 2.4a,e**). We also evaluate the phase of phase-locking to compare the average phase difference. Furthermore, IPS is applied to obtain the phase similarity at each time point. To quantitively assess the phase similarity, we examine PSI. If the phase difference seldom exceeds 45^{o}, PSI is close to 1; it is close to zero otherwise. Forward-NBGNet-predicted signals are in sync with the ground truths (73% of average phase difference < 22.5°; average PLV = 0.59; average PSI = 0.59; **Figure 2.4b**). Notably, the phase-synchronized predictions are generated in channel 3 (PSI = 0.51; **Figure 2.4c**). We further assess the phase locking and phase synchrony simultaneously for each channel and each trial in the form of a scatter plot (**Figure 2.4d**). Segmentation of the

scatter plot enables us to study further details. A larger fraction (74%) of predictions exhibits moderate or strong phase synchronization. We next evaluate the inverse-NBGNet's inference of the synchronous LFPs. Similarly, the predictions are in sync with the ground-truth LFPs (75% of average phase difference < 22.5°; average PLV = 0.60; average PSI = 0.60; **Figure 2.4f**). Notably, highly synchronized predictions at a representative channel are also observed (PSI = 0.83 for channel 51; **Figure 2.4g**). Furthermore, approximately half of the predictions have strong synchronization (**Figure 2.4h**). Our phase analysis comprehensively validates that the model predictions are phase-synchronized with the ground truth.

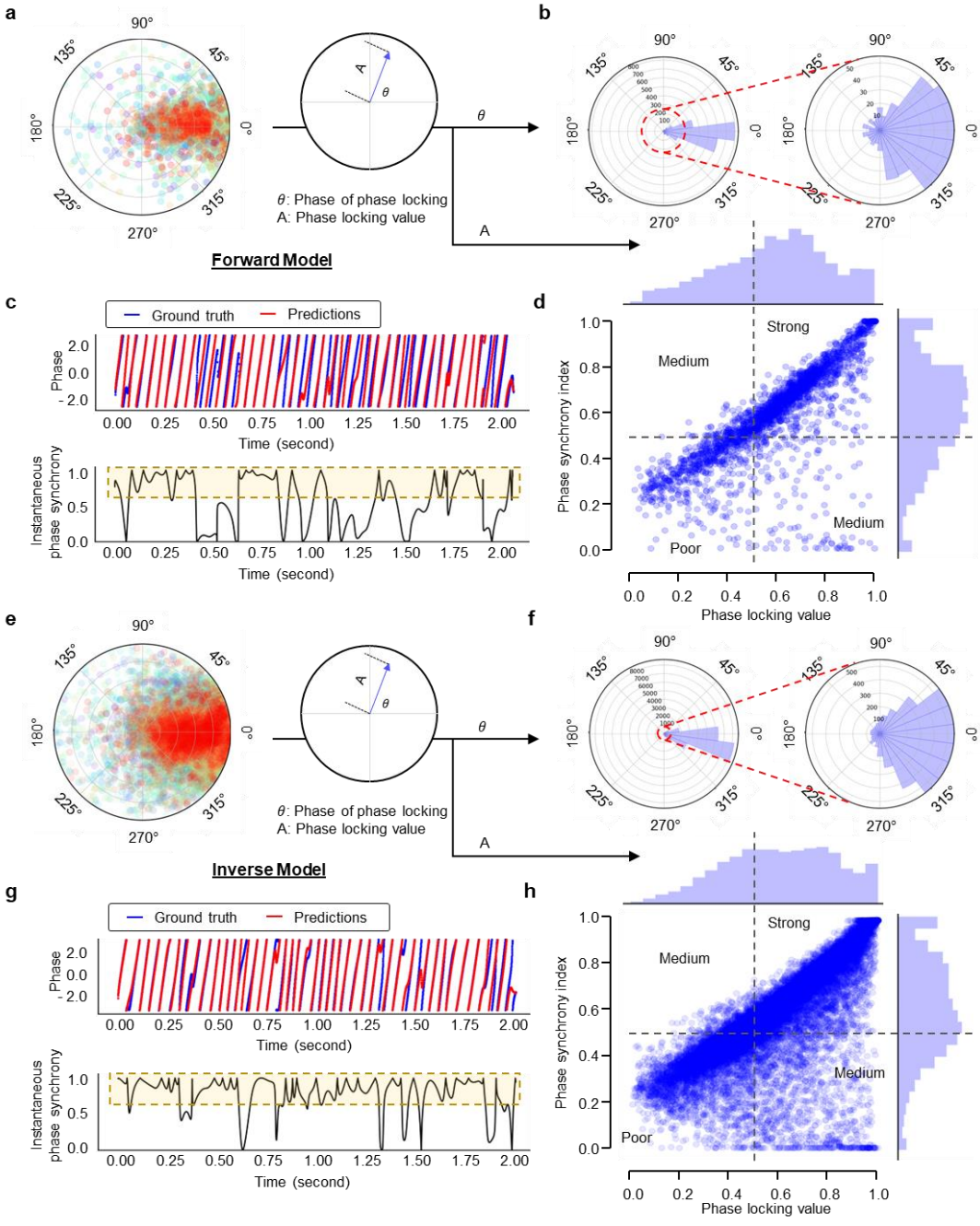
a
b
c
d
e
f
g
h
θ: Phase of phase locking
A: Phase locking value
Forward Model
Inverse Model
Ground truth
Predictions
Phase
Instantaneous phase synchrony
Time (second)
Phase synchrony index
Phase locking value
Strong
Medium
Poor

Figure 2.4 | Strong phase synchrony between NBGNet estimations and the experimental recordings. (**a**) Polar plots of the mean phase difference averaging across time in each trial for the forward model (n = 2,400). (**b**) Angular and corresponding zoom-in histogram of the phase of phase-locking derived from (**a**). (**c**) A screw ECoG channel (3: S1) was selected for demonstrating that NBGNet made predictions in sync with the ground truth in the 4th trial. Instantaneous phase of the ground truth (blue trace) and the model inference (red trace) at each timepoint (upper) was employed to obtain the instantaneous phase synchrony (lower; black trace) across the time. Yellow area showed a strong synchronization utilized to compute the phase synchrony index (PSI). (**d**) A scatter plot of phase analysis on each channel and each trial, respectively (n = 2,400), revealing the expected and hidden relations between PSI and phase-locking value (PLV). Histograms of both PLV and PSI are represented on the *x* and *y* axes, respectively. 0.5 was set as thresholds for both PSI and PLV (black dashed line) to identify strong, medium, and poor synchrony regions. (**e**) Same as (**a**) for the inverse model (n = 23,550). (**f**) Same as (**b**) for the inverse model, where the histogram was derived from (**e**. **g**) Same as (**c**), where the chosen LFP channel for demonstration was the same as **Figure 2.3**. (**h**) Same as (**d**) for the inverse model (n = 23,550).

2.4.4 Evaluation metrics – Neural latent dynamics analysis

To characterize the latent dynamics associated with the recorded or reconstructed neural activity in each trial, we analyze the filtered signals, which are obtained by applying a bandpass filter with cutoffs at 12.5 Hz and 30 Hz, in the window starting at movement onset and ending 600 ms after movement onset. Such a window is selected due to the interest in movement execution during the trial. For each trial, we obtain the data matrix D of dimension n by T, where n was the number of recorded channels, T was the number of time points in the given trial. Then we compute the low-dimensional manifold by applying principal component analysis (PCA)[68] to D. The resulting PCs are the linear combination of measurements of all the channels. We then rank these PCs based on the amount of neural variance explained by each PC. We keep only the three leading PCs to represent the low-dimensional manifold, where these three leading PCs, referred to as neural modes, explain most of the variance in the data matrix.

Differences between the neural recordings and the NBGNet's predictions necessarily cause a change in the estimated manifold and latent dynamics; however, a simple linear transformation can be applied to compensate for these differences[69]. Here we expect to identify the embedding space where true latent dynamics are located by using canonical correlation analysis (CCA). In CCA, given a pair of two latent trajectories, P_A and P_B, linear transformations for each trajectory are identified to make the linearly transformed latent trajectories, $\widetilde{\mathbf{P}}_A$ and $\widetilde{\mathbf{P}}_B$, maximally correlated. First, QR decomposition[70] is applied to both latent trajectories,

$$\boldsymbol{P}_A^T = \boldsymbol{Q}_A \boldsymbol{R}_A \tag{2.24}$$

$$\boldsymbol{P}_B^T = \boldsymbol{Q}_B \boldsymbol{R}_B \tag{2.25}$$

Then the singular value decomposition is performed on the inner product of Q_A and Q_B:

$$\boldsymbol{Q}_A^T \boldsymbol{Q}_B = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T \tag{2.26}$$

The transformation matrix, T_A and T_B, is then obtained by:

$$\boldsymbol{T}_A = \boldsymbol{R}_A^{-1}\boldsymbol{U} \tag{2.27}$$

$$\boldsymbol{T}_B = \boldsymbol{R}_B^{-1}\boldsymbol{V} \tag{2.28}$$

Accordingly, the transformed latent trajectories are given by:

$$\widetilde{\boldsymbol{P}}_A^T = \boldsymbol{P}_A^T \boldsymbol{T}_A \tag{2.29}$$

$$\widetilde{\boldsymbol{P}}_B^T = \boldsymbol{P}_B^T \boldsymbol{T}_B \tag{2.30}$$

The correlation between the transformed latent trajectories, termed canonical correlation (CC), is obtained by the Pearson correlation coefficient. As CC was sorted from the largest to the smallest in CCA, we expect to observe a descending order from neural mode 1 to mode 3.

Since low-dimensional latent dynamics have been widely used to illuminate the relationship between neural population activity and behavior[71–75], we also test whether NBGNet captures latent dynamics. The window of interest starting from movement onset and ending 600 ms after movement onset is selected, and there is no issue of imbalance target directions (**Figure 2.5a**). We compute the neural manifold and the latent dynamics within it using PCA[68,76]. The resulting PCs are termed the neural modes. The first three neural modes capture the majority of the variance, and are used to define the axes of the neural manifold. We then perform CCA[77–79] to align the latent dynamics (**Figure 2.5b**). Correlation analysis (Pearson's ρ) is utilized to quantify the similarity between these latent dynamics. Since canonical correlations are sorted from the largest to the smallest, we expect the same trend in the evaluation. First, we show that the single-trial and session-average latent trajectories of ground-truth and reconstructed screw ECoG are similar for all the target directions (**Figure 2.5c,d**). A strong and a moderate correlation are obtained for neural mode 1 (0.80) and mode 2 (0.60), respectively (**Figure 2.5e**). To assess the effects of behavioral states on the model performance, we calculate the instantaneous correlation across time for each trial. We demonstrate both mode 1 and 2 exhibit a consistently strong correlation (**Figure 2.5f**). These observations hold for the inverse-NBGNet, where latent trajectories derived from the inferred and the ground-truth LFPs are

highly correlated (**Figure 2.5g**). Similarly, session-averaged latent traces for the first neural mode are almost the same for all the targets (**Figure 2.5h**). A strong correlation, as well as correlated instantaneous correlation, are also observed in neural mode 1 (0.69; **Figure 2.5i-j**). Neural mode 2 exhibits a relatively poorer performance due to less precise inference from inverse-NBGNet. The results indicate that the NBGNet captures the latent dynamics. As expected, a stronger correlation is associated with the higher ratio of variance that the neural mode explained.

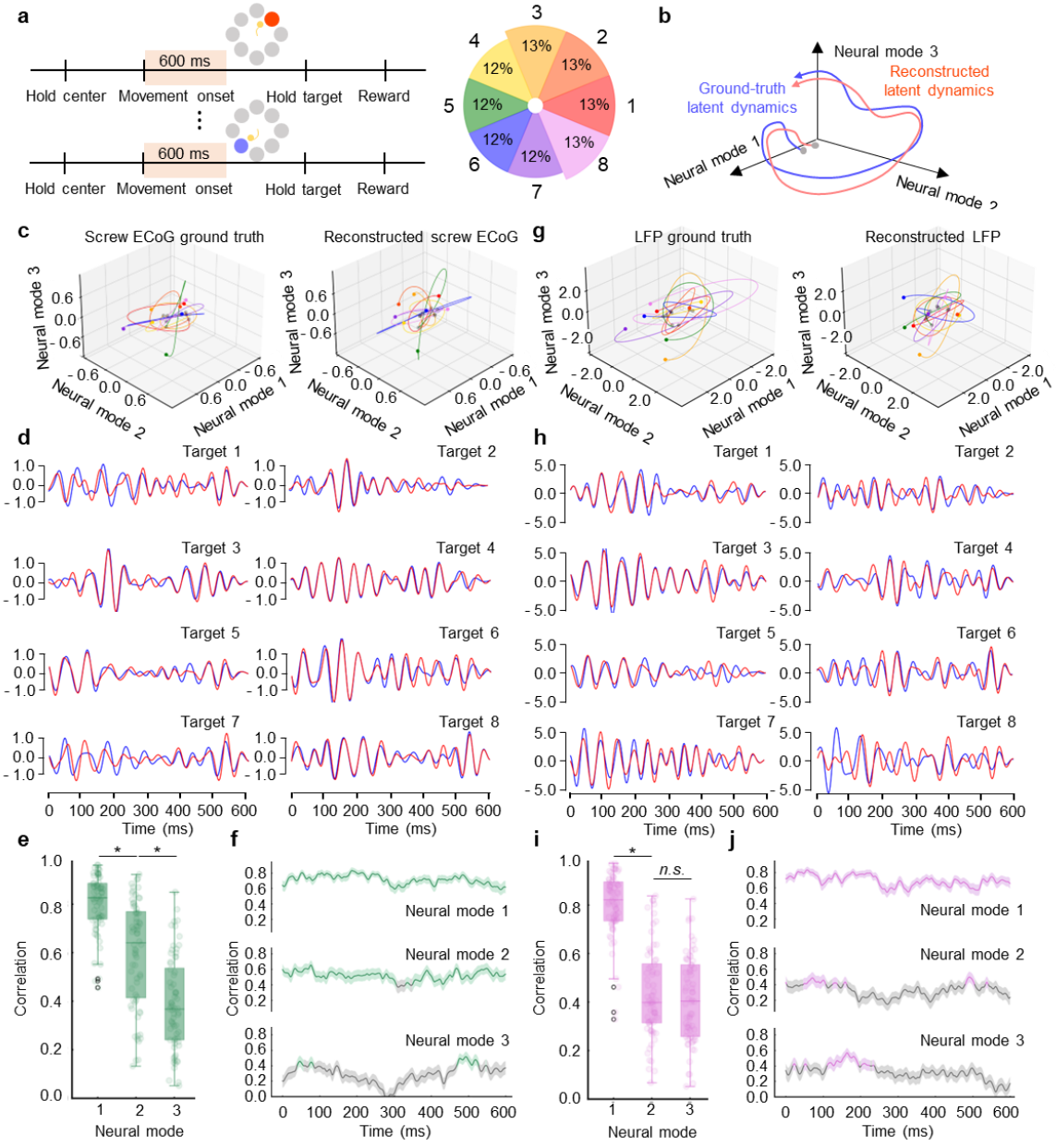
a
600 ms
Hold center
Movement onset
Hold target
Reward
600 ms
Hold center
Movement onset
Hold target
Reward
13%
13%
13%
13%
13%
12%
12%
12%
12%
1
2
3
4
5
6
7
8
b
Neural mode 3
Reconstructed latent dynamics
Ground-truth latent dynamics
Neural mode 1
Neural mode 2
c
Screw ECoG ground truth
Reconstructed screw ECoG
Neural mode 3
Neural mode 2
Neural mode 1
g
LFP ground truth
Reconstructed LFP
d
Target 1
Target 2
Target 3
Target 4
Target 5
Target 6
Target 7
Target 8
Time (ms)
h
e
Correlation
Neural mode
f
Neural mode 1
Neural mode 2
Neural mode 3
Time (ms)
i
n.s.
j

Figure 2.5 | NBGNet captures and reconstructs the latent dynamics in the reaching-out task. (**a**) Schematic of protocol indicates the time window used for analysis. The probability of each target direction is uniform. (**b**) We predicted that the latent dynamics can be recovered. (**c**) Representative latent trajectories derived from the ground-truth screw ECoG (left) and reconstructed screw ECoG (right). Each color represents each target direction in (**a**). (**d**) Projection of average ground-truth (blue trace) and reconstructed (red trace) latent trajectories for each target on the first mode. (**e**) Bar plot showing the strong magnitude of the correlations between the ground-truth and reconstructed latent trajectories (error bars, s.e.m.; n = 68). (**f**) Temporal correlation trajectories for each neural mode (green trace when above the threshold as 0.4; grey trace as below the threshold; mean ± s.e.m.). (**g**) Same as (**c**) for the inverse model to reconstruct the latent trajectories derived from LFPs. (**h**) Same as (**d**) for the projection of average ground-truth LFPs-derived (blue trace) and reconstructed LFPs-derived (red trace) latent trajectories. (**i**) Same as (**e**) for the correlation between the latent trajectories obtained from recorded LFPs and estimated LFPs. (**j**) Same as (**f**) for the inverse model (purple trace when above the threshold as 0.4; grey trace as below the threshold). *$p < 0.05$ using two-sided Wilcoxon's rank-sum test. *n.s.* indicates no significant difference.

2.4.5 Evaluation metrics – Feature selection and motion classification

We consider several features per channel as candidates for the decoder and select the leading number of features for further analysis. For each channel, we obtain a total of 34 features, including root mean square (RMS), mean frequency (MF), waveform length (WL), and the power at certain frequency ranged from 10 – 40 Hz (step size as 1 Hz):

$$RMS = \sqrt{\frac{1}{T}\sum_{t=0}^{T-1} \boldsymbol{Y}(t)^2} \tag{2.31}$$

$$MF = \frac{\sum_{t=0}^{T-1} f_t p_t}{\sum_{t=0}^{T-1} p_t} \tag{2.32}$$

$$WL = \sum_{t=1}^{T-1} |\boldsymbol{Y}(t) - \boldsymbol{Y}(t-1)| \tag{2.33}$$

where $Y(t)$ represents the neural signals, T is the number of time points in the given trial, and f_t and p_t are the frequencies of the power spectrum and the corresponding amplitude.

To determine the subset of features selected for decoders, we calculate the Fisher score[80] for each candidate feature. The Fisher score, $F(x_i)$, for the ith feature, x_i, is computed by

$$F(x^i) = \frac{\sum_{j=1}^{c} n_j \left(\mu_j^i - \mu^i\right)^2}{\sum_{j=1}^{c} n_j \left(\sigma_j^i\right)^2} \tag{2.34}$$

where μ_{ij} and σ_{ij} are the mean and standard deviation of the j^{th} class corresponding to the i^{th} feature, μ_i denotes the mean of the whole data set corresponding to the i^{th} feature, n_j represents the size of the j^{th} class, and c is the total number of classes. After computing the Fisher score for each feature, we select the top fourteen ranked features to predict the subject's behavior. Number of features is determined by maximizing the classification accuracy via grid search.

To test whether the reconstructed activity from the NBGNet maintain movement-related information, we build linear decoders to predict the direction of the movement based on the neural activity. Our hypothesis is that our NBGNet inference and the neural recordings will yield a comparable classification accuracy. To test this hypothesis, we compare the predictive accuracy of seven types of decoders: (1) a decoder trained and tested based on screw ECoG; (2) a decoder trained and tested based on reconstructed screw ECoG inferred by forward-NBGNet; (3) a decoder trained and tested based on LFP; (4) a decoder trained and tested based on reconstructed LFP inferred by inverse-NBGNet; (5) a decoder trained and tested based on screw ECoG and LFP; (6) a decoder trained and tested based on reconstructed screw ECoG and LFP; and (7) a decoder trained and tested based on reconstructed LFP and screw ECoG. All decoders are defined using linear discriminant

analysis with the selected features as inputs to predict the direction of cursor's movement. They are trained and tested on the same day, using a 4-fold cross-validation procedure to protect against overfitting. Chance-level performance is obtained by shuffling the dataset. As expected, all predictive accuracy is higher than chance-level (~12.5%).

To understand the information encoded within neural populations, decoding cortical activity is of particular interest[78]. We wondered how accurately linear decoders trained with the model-inferred neural activities would perform. We first extract candidate features from the dataset and picked fourteen of them using Fisher score[81], where fourteen features yields the highest classification accuracy *via* grid search. Linear discriminant analysis (LDA)[81] classifiers are then trained with the selected features to predict the direction of cursor's movement. The classification accuracy is evaluated using 4-fold cross-validation. Candidate features are arranged in descending order based on Fisher score averaging across all the channels. LDAs are trained with seven conditions: (1) screw ECoG only, (2) reconstructed screw ECoG only, (3) LFP only, (4) reconstructed LFP only, (5) screw ECoG + LFP, (6) reconstructed screw ECoG + LFP, and (7) screw ECoG + reconstructed LFP. Fourteen features are selected for classifiers 1 – 4; while twenty-eight features (two-fold increase due to more candidate features available from two dataset) are selected for classifiers 5 – 7. We demonstrate that no significant difference in classification accuracy between the model inference and the ground truth is observed ($p > 0.05$; **Figure 2.6**), indicating that NBGNet's inference maintains the discriminant power. As expected, the classifier trained with LFP and screw ECoG outperforms the other conditions. Surprisingly, the classifiers trained with both real signals and with the inclusion of NBGNet's predictions (reconstructed screw ECoG + LFP and screw ECoG + reconstructed

LFP) yield a comparable decoding capability. Together, we show that the presented model maintained the integrity of information represented by the neural activity.

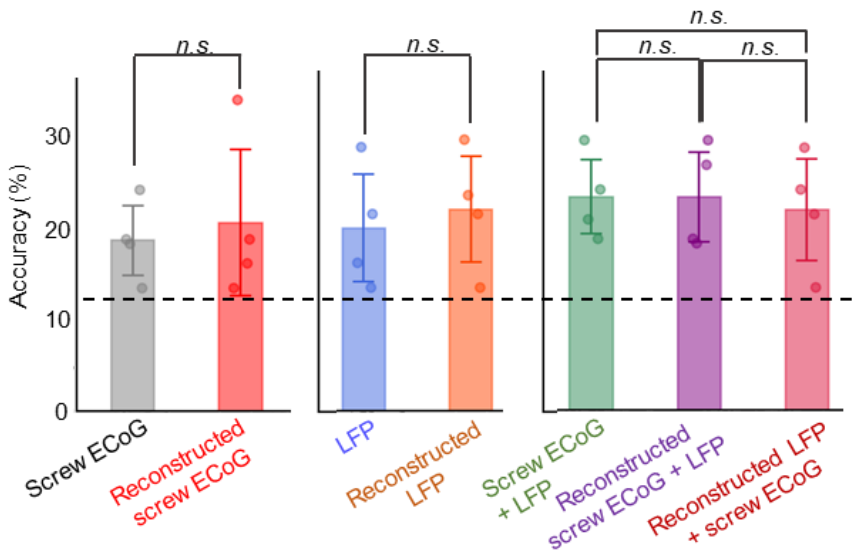

Figure 2.6 | NBGNet inference can be used to predict the movement behavior. Bar plot showing the classification accuracy for each dataset (dashed line, chance performance; error bars, s.d.; n = 9). *$p < 0.05$ using two-sided paired T-test. *n.s.* indicates no significant difference.

2.4.6 Evaluation metrics – Comparison methods

The multi-scale modeling is relatively new in the neuroscience field. To benchmark performance of NBGNet relative to other existing algorithms, we implement two approaches: the electrophysiology-based sphere head model[51] and data-driven recurrent neural network[82]. The sphere head model is widely used to either compute the contribution from the current dipoles to the electrical potentials recorded at scalp electroencephalography (EEG) or estimate the current dipole sources based on the scalp potentials[83]. Typically, the sphere head model assumes the head to be modeled as a four-

layered sphere (brain, cerebrospinal fluid, skull, and scalp). Using the quasi-static approximation of Maxwell's equations and the volume-conductor theory, the electrical potential, $\Phi(r,t)$, is obtained by the following Poisson equation[84]:

$$\nabla \cdot \sigma(\boldsymbol{r}) \nabla \boldsymbol{\Phi}(\boldsymbol{r}, t) = -C(\boldsymbol{r}, t) \tag{2.40}$$

where $\sigma(r)$ represents the position-dependent conductivity of the medium, and $C(r,t)$ is the density of the current sources. Assuming the conductivity to be isotropic, the boundary conditions to the sphere head model was

$$\boldsymbol{\Phi}^{s+1}(\boldsymbol{r}_s, t) = \boldsymbol{\Phi}^{s}(\boldsymbol{r}_s, t) \tag{2.41}$$

$$\sigma_{s+1} \frac{\partial \boldsymbol{\Phi}^{s+1}(\boldsymbol{r}_s, t)}{\partial r} = \sigma_s \frac{\partial \boldsymbol{\Phi}^{s}(\boldsymbol{r}_s, t)}{\partial r} \tag{2.42}$$

$$\frac{\partial \boldsymbol{\Phi}^{4}(\boldsymbol{r}_4, t)}{\partial r} = 0 \tag{2.43}$$

where each layer is labelled by s = 1 to 4. Here we assume the dipole is radial with magnitude $p(t)$ at location r_z. The analytical solution is then given by:

$$\boldsymbol{\Phi}^{1}(r, \theta, t) = \frac{p(t)}{4\pi\sigma_1 r_z} \sum_{n=1}^{\infty} \left[A_n^1 \left(\frac{r}{r_1} \right)^n + \left(\frac{r_z}{r} \right)^{n+1} \right] n P_n(cos(\theta)) \qquad r_z < r \leq r_1 \tag{2.44}$$

$$\boldsymbol{\Phi}^{s}(r, \theta, t) = \frac{p(t)}{4\pi\sigma_1 r_z} \sum_{n=1}^{\infty} \left[A_n^s \left(\frac{r}{r_s} \right)^n + B_n^s \left(\frac{r_s}{r} \right)^{n+1} \right] n P_n(cos(\theta)) \qquad r_{s-1} \leq r \leq r_s \tag{2.45}$$

where $\Phi^s(r,\theta,t)$ is the extracellular potential measured at radius r and the angle θ between the measurement and dipole location vectors in the shell s, r_s represent the radius of sphere

s, A^s_n and B^s_n are the coefficients depending on the radius and conductivities of each medium (defined in[51]), and $P_n(cos(\theta))$ represents the n^{th} Legendre Polynomial. As the solution is implemented into the case where we had n_d current dipoles and nr recording units, a linear transformation matrix F of dimension n_r by n_d is obtained and utilized to convert the dipole moment vectors X into the electrical potential Y, given by Y = FX. This is a so-called forward mapping. When we perform inverse mapping to estimate X from Y, we need to solve an underdetermined system with pseudo-inverse by minimizing the following equation,

$$\|\boldsymbol{X}\|^2 + \lambda(\boldsymbol{Y} - F\boldsymbol{X}) \tag{2.46}$$

The solution to minimizing the above equation is given by,

$$\boldsymbol{X} = F^T(FF^T)^{-1}\boldsymbol{Y} \tag{2.47}$$

Here we segment the brain (n_d = 3600), where each segment includes a potential current dipole source. Since our data for comparison does not include dipole sources, we adapt the algorithms into two-step computation for both the forward and inverse models. In the forward model, we perform inverse mapping from LFP toward estimated dipole sources, followed by a forward mapping from the estimated dipole sources toward screw ECoG recordings. Similarly, in the inverse model, we perform inverse mapping from screw ECoG toward estimated dipole sources followed by a forward mapping from the estimated dipole sources toward LFP recordings. The parameters are summarized in **Table 2.1**.

Label	Tissue	Radius (mm)	σ (S/m)
1	Brain	27.88	0.33
2	Cerebrospinal fluid	28.24	1.65
3	Skull	30.00	0.00825
4	Scalp	31.76	0.33

Table 2.1: Parameters for sphere head model.

RNN is a deep learning method widely used to model a nonlinear dynamical system that included nonlinearity, recurrent connection, and hidden dynamic states[85,86]. In order to handle the long-term dependency embedded in the neural activity, Gated recurrent unit (GRU)[82] is often implemented, where in each time point, model can determine the information to be stored and filtered. GRU is chosen over long short-term memory (LSTM) by its speed and the simpler structure. GRU-based RNN utilized in this work for comparison consists of two GRU layers with 64 and 32 units, two hidden layers with 32 and 16 nodes for forward model or 32 and 128 nodes for inverse model, and the output layer. To avoid overfitting, we train the GRU-RNN with L2 regularization[87,88] and dropout[89]. The relevant hyperparameters were optimized via Bayesian optimization. The training details, including training iteration, the split ratio of training and validation data, and the choice of optimizer, are set to be the same as NBGNet to ensure a fair comparison.

Here we compare the NBGNet with two conventional alternatives, specifically a sphere head model[34,90] and GRU-RNN. The sphere head model provides analytical formulas describing the contribution from current sources to EEG potentials with the

assumption of a multi-layered spherical head where each layer represents each brain tissue. The inverse computation can then be achieved. We also apply the state-of-the-art deep learning technique, GRU-RNN, as a purely date-driven alternative. The NBGNet outperforms the purely data-driven GRU-RNN and electrophysiology-based sphere head model for multiple days (**Figure 2.7j**). As expected, the performance of the analytical solution is the poorest due to the non-high-density recordings and the unrealistic assumptions (e.g., isotropic conductivity of the medium). While GRU-RNN clearly performs better than the sphere head model, the NBGNet gives a more accurate inference consistently over multiple days.

2.4.7 Evaluation metrics – Stability across days

To benchmark performance of NBGNet relative to other existing algorithms, we implement two approaches: the electrophysiology-based sphere head model[51] and data-driven recurrent neural network[82]. The sphere head model is widely used to either compute the contribution from the current dipoles to the electrical potentials recorded at scalp electroencephalography (EEG) or estimate the current dipole. As experiments are often conducted across multiple sessions or days, whether the trained model could generate a reliable and robust result is crucial. Here we examine the stability of NBGNet using the same metrics presented in the previous sections. We would like to emphasize that the NBGNet was trained on day 1 and remains fixed for testing in subsequent days. First, the average RMSE for both forward and inverse model are consistent over weeks to a degree almost indistinguishable from that in Day 1 (**Figure 2.7a-b**). As is the case for RMSE, the beta correlation is stable as well even with a few individual trial exceptions (**Figure 2.7c-d**). Specifically, we find an unexpected decrease in correlation of a specific channel (Day

2 for forward model; Day 16 for inverse model) due to the change of the order of magnitude in the measurements. Overall, the predicted neural activities are still highly correlated with the empirical recordings (ρ= 0.47 and 0.52 for forward and inverse model).

We then test the stability in phase analysis. While the performance slightly dropped with the time, Forward-NBGNet-inferred screw ECoG are still highly synchronized with the real recordings (**Figure 2.7e**). More predictions are mostly categorized in the moderate to strong synchrony zone than in the poor one (+48%, +13%, +17%, +9%, -4% for Day 1, 2, 4, 12, 16, respectively). Similarly, reconstructed LFPs are in sync across sessions (+10%/session; **Figure 2.7f**).

The NBGNet maintains the capability of reconstructing latent dynamics during the repeated movement generation for the full length of recordings from the monkey (**Figure 2.7g,h**). The stability holds for a range of manifold dimensionalities from 1 to 3. As we find in Day 1 (**Figure 2.7e,i**), the descending trend in the correlations of neural modes is observed for multiple days. The average temporal correlations also show similar results for both forward and inverse model. We then test whether NBGNet inferences predict behavior in different sessions. It is noted that the classifiers performed as well as that trained in different sessions (**Figure 2.7i**). These results provide evidence that NBGNet-derived signals predict behavioral variables with similar accuracy as compared with the ground-truth signals for multiple sessions.

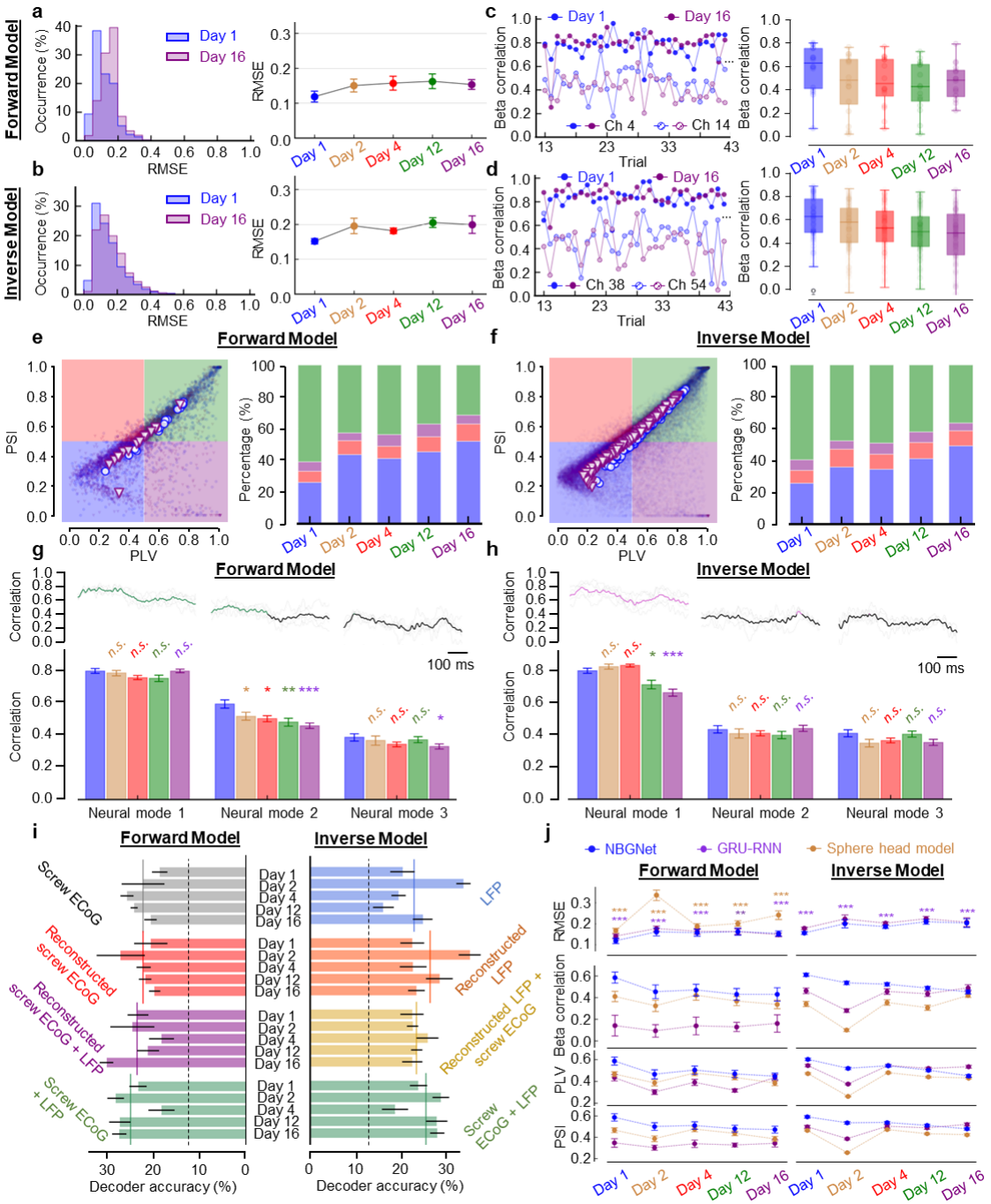
a
Forward Model
Occurrence (%)
RMSE
Day 1
Day 16
b
Inverse Model
c
Beta correlation
Trial
Ch 4
Ch 14
d
Ch 38
Ch 54
e
Forward Model
PSI
PLV
Percentage (%)
f
Inverse Model
g
Correlation
100 ms
n.s.
Neural mode 1
Neural mode 2
Neural mode 3
h
i
Screw ECoG
Reconstructed screw ECoG
Reconstructed screw ECoG + LFP
Screw ECoG + LFP
LFP
Reconstructed LFP
Reconstructed LFP + screw ECoG
Screw ECoG + LFP
Decoder accuracy (%)
j
NBGNet
GRU-RNN
Sphere head model
PLV
PSI
Day 1
Day 2
Day 4
Day 12
Day 16

Figure 2.7 | Stability of NBGNet's predictions for multiple days. (**a**,**b**) Histogram of RMSE (left) at Day 1 (blue) and 16 (purple) for the forward (**a**) and the inverse (**b**) model. Scatter plot of average RMSE (right) showing no significant difference (error bars, s.e.m.; n = 16 and 157 for **a** and **b;** p = 0.29 and 0.07 for **a** and **b** using one-way ANOVA test). (**c**,**d**) Beta correlation (left) at Day 1 (blue) and 16 (purple) for the forward (**c**) and the inverse (**d**) model. Scatter plot of average beta correlation, where black solid line is obtained by averaging over the channels and black dashed line left the poorest channel out (error bars, s.e.m.; n = 7 and 16 for (**c**) and (**d**)). p = 0.32 and < 0.05 for **c** and **d** using one-way ANOVA test. While there are some significant decreases for the inverse model, the effect size is small. (**e**,**f**) Scatter plot of PSI versus PLI (left) at Day 1 (blue circle) and 16 (purple triangle) for the forward (**e**) and the inverse (**f**) model. Stacked bars (right) demonstrate the percentage of predictions locating in each section. (**g**,**h**) Temporal correlation averaging across days (upper), where colored segments represents stronger correlation as compared with the grey counterparts. Bar plot of average correlation (lower) exhibiting stable performance (error bars; s.e.m.; n = 68, 49, 128, 78, 135 at Day 1, 2, 4, 12, 16). *p < 0.05, **p < 0.01, ***p < 0.001 using two-sided Wilcoxon's rank-sum test. *n.s.* indicates no significant difference. (**i**) Bar plots showing the classification accuracy of linear classifier to predict the target direction (error bars, s.d.; n = 5). Solid line represents the average performance across days. (**j**) Performance comparisons between the NBGNet, GRU-RNN, and the sphere head model. **p < 0.01, ***p < 0.001 using two-sided Wilcoxon's rank-sum test.

2.5 CROSS-SCALE CONNECTIVITY

The complex coordination of brain functions, such as vision, motor preparation, and attention requires the control of causal interactions between areas[91]. Effective connectivity, which represents the influence that a neural system exerts over another[92], is thus a powerful measure to evaluate the brain computations. In the NBGNet, we are able to derive the cross-scale effective connectivity that depicts how the latent states of sources change those of targets. Positive and negative connection strengths correspond to excitatory and inhibitory effects, respectively. The cross-scale effective connectivity exhibits patterns of visual feedback (unique to target position) and voluntary movement (shared across target position; **Figure 2.8**) for different movement directions in the center-out joystick task. During rightward movement, a unique inverse connectivity from lateral prefrontal cortex

to frontal eye field is observed, exhibiting a strong preference for contralateral visual space[93]. Furthermore, the identification of multiple shared causal interactions (e.g., prefrontal cortex-supplementary motor area, prefrontal cortex-motor cortex, and somatosensory cortex-motor cortex) over all the target directions aligns well with the abstraction of the hierarchical anatomy of the mammalian nervous system[59,94]. Thus, NBGNet-derived effective connectivity holds great potential to illuminating the cross-scale computations underlying brain functions.

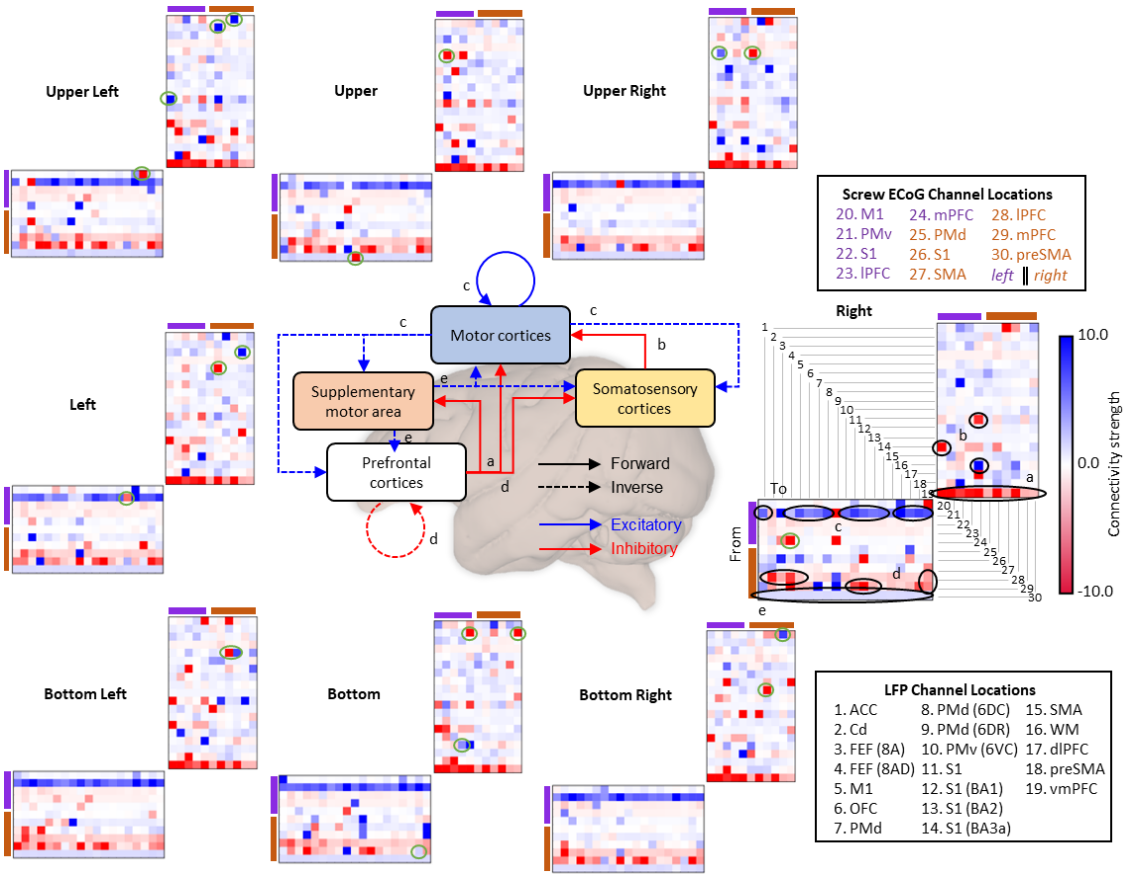

Figure 2.8 | Bi-directional effectivity connectivity extracted from NBGNets exhibited unique and shared patterns in the center-out joystick task. The bi-directional effective connectivity for each target direction was obtained from the NBGNet's parameters and was averaged over the trials reaching the same target. Each subfigure corresponds to a target position. The vertical axes represent the channels where the connection originates; the horizontal axes represent the channels where the connection contributes to. The shared patterns were indicated with black circles; the unique patterns were indicated with the green circles. The circuitry diagram (middle) depicts the hierarchical interactions between brain regions from the shared patterns of effective connectivity.

2.6 DISCUSSION

The brain consists of a hierarchical system with multiple levels of organization[95]. Growing interest in multi-scale interactions among the genetic, cellular, and macroscale levels has recently inspired a shift from emphasizing neural communication in individual scales to exploring the potential associations between scales. However, how these scales are interlinked is still an open question. In this work, the NBGNet addresses an unmet need to capture the implicit relations of multi-scale brain activity. We demonstrate that the neural activity at one scale can be inferred from one another with consistent performance across multiple days without model retraining.

As neuronal coupling among distinct populations can be linear (synchronous) or nonlinear (asynchronous)[96], a powerful tool capable of capturing nonlinear interactions is imperative. Here we utilize a BG approach to derive the nonlinear system dynamics in multi-scale brain network and employ a deep learning technique to approximate the nonlinear mapping. The BG method enables the modeling of multi-domain physical systems by specifying the transfer of energy between system components. To the best of our knowledge, this is the first time that the BG is applied to the brain. We model the transfer of electrical energies among brain tissues, but these energies can be measured at different scales or by different approaches. Inherited from BG modeling, another important feature of the NBGNet is system identification[56,97,98]. After network training, the system parameters are extracted and further utilized to interpret the temporal evolution of the underlying dynamical system. The embedded dynamics in the NBGNet are thus able to illustrate how the activity at one scale communicates with other scales, serving a key factor in uncovering the mechanistic understanding of brain computations and the mediation of the behaviors.

The guiding factor in model evaluation is utilizing comprehensive metrics. This is especially important for neuroscientific research. A perfect performance in one metric may not guarantee the same observation in another. RMSE (**Figure 2.1, 2.7**) is used to indicate the absolute measure of fit. For similarity analysis in time- and phase-domain, we assess cross-correlation (**Figure 2.3**) and phase synchrony (**Figure 2.4**) between model predictions and ground truths. As a key to understanding neural mechanisms, the capability of reconstructing the low-dimensional latent dynamics is also examined (**Figure 2.5**). Additionally, we consider the decoding accuracy as an indicator of the applicability to brain-machine interface paradigms (**Figure 2.6**). Finally, the predictive power without retraining the model over a long period has recently drawn growing attention in the field of neural engineering. We validate the NBGNet as a reliable approach with aforementioned metrics and show its broad applicability (**Figure 2.7**).

As the model performance with limited channels is of particular interest, we randomly chose 16 LFP channels and 7 screw ECoG channels from distinct areas to obtain a subset of anatomically spatially distributed signals. It is worth noting that the NBGNet still yields a similar performance if only a subset of measurements is accessible. Interestingly, NBGNet's performance is not dependent upon the depth, but on the regions (**Figure 2.3**). Forward-NBGNet captures the internal dynamics for performing a center-out task and thus accurately reconstructs the task-related neural activity in premotor, prefrontal cortex, and primary motor cortex. Furthermore, inverse-NBGNet-inferred activity matches the ground truth not only at the cortical region but also at a deeper subcortical area. As inverse model is developed by nonlinearizing the inversion of linear forward mapping rather than the direct inversion of nonlinear forward mapping, a slightly poorer performance is expected at more ventral brain region. Additionally, evidenced with the

failure of capturing the noise from unstable recordings, dynamics embedded in the NBGNet are useful for disambiguating brain computations.

The bias-variance trade-off is a critical problem in statistics and machine learning[99], where the simple models have a lower variance yet a higher bias, and the complexity of the model can reduce the bias but increase the variance. It is thus expected that the NBGNet outperforms the analytical sphere head model and the GRU-RNN. With the assumptions of dipole as the signal sources and the conductivity of the brain tissues, the sphere head model provides a simple solution but leads to a large bias error. The data-driven GRU-RNN enables the approximation of nonlinear dynamics; however, a large variance, or the so-called "overfitting," can be observed. Therefore, to make a fair comparison, we train the GRU-RNN with appropriate regularization. However, the regularized GRU-RNN is still a black box without any physiological interpretation of the model. Combining both neurobiological modeling and deep learning technique, the NBGNet succeeds in capturing the patterns in the training data and adapting itself to unseen data. With the complexity lying between sphere head model and GRU-RNN, the NBGNet holds great potential to resolving the bias-variance dilemma.

In sum, the NBGNet is powerful for investigating the underlying dynamics in multi-scale brain networks. Modeling the neural activity at disparate scales yields the causal interactions among multiple levels, which is crucial in illuminating the mechanistic understanding of brain computation. Effective connectivity extracted from the NBGNet exhibits both unique and shared patterns of both visual feedback and voluntary movement, suggesting that the NBGNet serves as a useful tool to study the brain computation. Whereas current work focuses on cross-scale interaction, within-scale communication can be incorporated for comprehensive modeling. Additionally, NBGNet can potentially improve

the applicability of brain-machine interfaces by inferring the brain activity with increased signal-to-noise ratio and even combining multi-scale activity[100]. Moreover, the inverse computation to reconstruct the activity at the uncovered brain regions makes LFP-derived whole-brain dynamics available. Taken together, our work represents an important step forward towards the mechanistic modeling of multi-scale neural activity, which may facilitate our understanding of neuropathological activity and the development of clinical devices and rehabilitative therapies to treat abnormal neural activity underlying dysfunctional behaviors.

Chapter 3: Multiscale dynamic modeling using neurobiologically realistic neural ordinary differential equations

3.1 INTRODUCTION

The brain is a formidable complex system exhibiting computational structure involving multiple spatial scales (from molecules to whole brain) and temporal scales (submilliseconds to the entire lifespan)[101]. Effective connectivity (EC) is a type of brain connectivity that characterizes the relations between brain regions. Unlike structural connectivity for anatomical links and functional connectivity for statistical dependencies, EC refers to a patten of causal interactions between distinct units within the nervous system. Multiscale effective connectivity (msEC) among brain regions provides essential information about human cognition[102] and behaviors such as motor preparation[41], motor adaptation[42], motor timing[43], decision making[45], and working memory[46]. To date, work has primarily focused on extracting EC from a single modality (e.g., electrophysiology, functional magnetic resonance imaging, and ^{18}F-fludeoxyglucose positron emission tomography[102]) and typically makes simplified assumptions in which neural dynamics are linear[27] or log-linear[30]. However, the lack of the communications between multiple modalities and the nonlinearity prevents us from uncovering a deeper and more comprehensive understanding of system-level neural mechanisms of motion behavior[48,49].

msEC can be divided into within-scale and cross-scale ECs, where the former indicates the causal interactions between neural elements at the same spatial and temporal scales and the latter specifies the causal interactions between neural elements at different spatial or temporal scales. Previous work has largely focused on inferring within-scale EC *via* multivariate autoregressive models[103], vector autoregressive models[104], psycho-physiological interactions[105], structural equation modeling[106–109], or dynamic causal

modeling[110]. Despite emergence of the cross-scale analyses such as source localization[111] and cross-level coupling (CLC)[38], the fidelity of experimental implementation of source localization is limited and only the statistical dependencies are quantified by CLC. To reveal the directed interactions across spatiotemporal scales of brain activity, recent work has developed the generalized linear model-based multi-scale method[39]. However, recent study indicates that local brain dynamics relies on nonlinear phenomena[112]. Nonlinear model may be required to generate rich temporal behavior matching that of the measured data[113]. Taking the nature of nonlinearity in brain computations, we have previously proposed the NBGNet, a sparsely-connected recurrent neural network (RNN) where the sparsity is based on the electrophysiological relationships between modalities, to capture cross-scale EC[114]. Despite the success of capturing complex dynamics using a nonlinear model, to the best of our knowledge, we still lack an integrative method that can infer nonlinear msEC.

To analyze multiscale neural activity in an integrative manner, we introduce a multiscale modeling framework termed msDyNODE (multiscale neural dynamics neural ordinary differential equation). Neural ordinary differential equation (NODE) is a new family of deep neural networks that naturally models the continuously-defined dynamics. In our method, within-scale dynamics was determined based on neurobiological models at each scale, and cross-scale dynamics was added as the connections between latent states at disparate scales. Using both simulation and experimental dataset, we demonstrated that msDyNODE not only well reconstructed the multi-scale data but also uncovered multi-scale causal interactions driving the cognitive behavior.

3.2 DATA ACQUISITION

3.2.1 Experimental protocols

A male rhesus macaque was implanted with a 64- or 128-channel chronic array (Innovative Neurophysiology, Inc., Durham, NC) in the left dorsal premotor (PMd) and primary motor cortex (M1) regions (**Figure 3.1**). Before the experimental session, we ran the calibration session. During the calibration, the rhesus macaque passively observed a cursor moving from the center target toward a randomly generated peripheral target, followed by the cursor movement back to the center. In addition to providing continuous visual feedback, we also reinforced the behavior and neural activity by delivering a small juice reward directly into the subject's mouth. The neural data was recorded for approximately three and a half minutes (or reaching 6 trials per target direction). A Kalman filter (KF) was employed as the decoder to map the spike count from each unit to a two-dimensional cursor output. While the KF decodes both intended position and velocity, only the velocity was used to estimate the position at the next time point based on kinematic equations of motion. To increase the initial performance and reduce directional bias, we conducted daily, 10-minute closed-loop decoder adaptation (CLDA)[115–119] sessions. Both the decoder and neural activity adapted to complete center-out tasks with consistent trial times and straight path lengths to each target. After the calibration session, the main task was manually initiated. The rhesus macaque was trained in a BMI task called "center-out"[120–122]. During the task, spiking activity was recorded online to produce cursor commands in real-time. Spikes for each unit were summed over a window of 100 millisecond and served as the input to the decoder. The neural activity was then transformed into a "neural command" by applying the dot product of the spike count vector to the Kalman gain matrix. Cursor position was updated iteratively by adding the cursor position to the product of velocity,

which was determined by the neural command, times update time (100ms). In each trial, they controlled the velocity of a computer cursor to move from the center target toward one of eight outer targets. Only one peripheral target was presented on a given trial. The order of the appearance of the target was randomly selected; however, for every eight consecutive trials, each target was shown once. The 8 targets were radially distributed from 0° to 360° (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°) at equal distances from the center (10 centimeters). Upon successful completion of reaching and holding the cursor at the peripheral target for 0.2 seconds, the target turned green, and a small juice reward was dispensed directly into the subject's mouth. The cursor then automatically appeared at the center of the screen to initiate another new trial. Subjects can fail the task in two ways: (1) failure in holding the cursor at the center target or the peripheral target for 0.2 seconds or (2) failure in reaching the peripheral target within specified time (10 seconds). The subject has 10 chances to complete a successful trial before the task automatically moves onto the next target.

All the experiments were performed in compliance with the regulation of the Animal Care and Use Committee at the University of Texas at Austin.

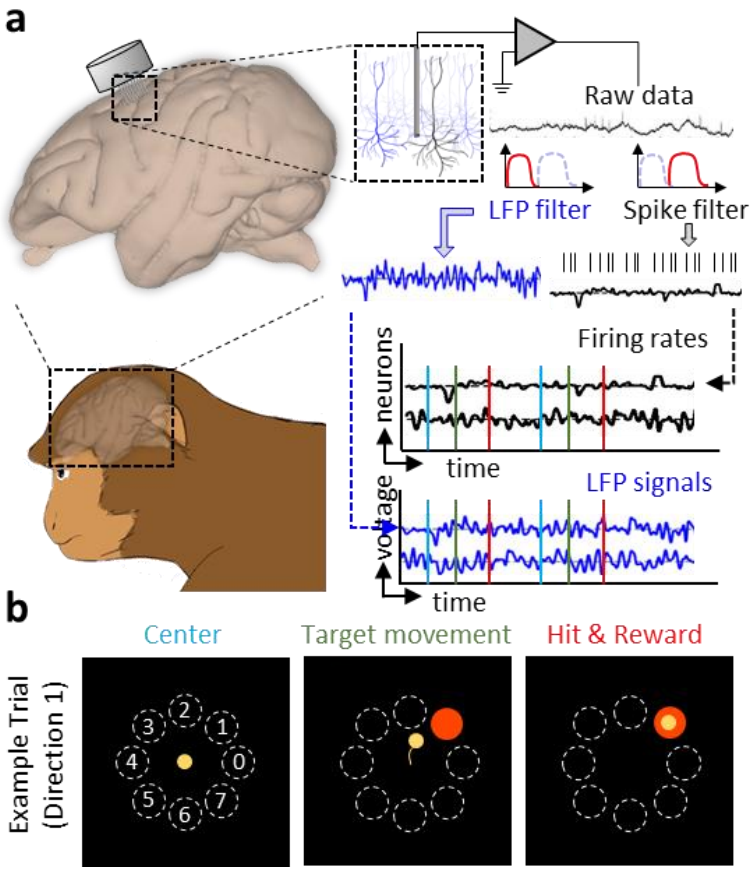

Figure 3.1 | Data acquisition and experimental task design for multiscale neural signals. (**a**) Simultaneous recording of firing rates and LFP signals. (**b**) The visual feedback task contains eight different cursor movements, each corresponding to one of the eight outer targets. The color-coded tasks are also indicated in **a**.

3.2.2 Spike trains and LFP data

The extracellular single and multi-unit activity in the left primary motor cortex (M1) and dorsal premotor cortex (PMd) were recorded using a 64- or 128-channel chronic array (Innovative Neurophysiology, Inc., Durham, NC; **Figure 3.1a**). The spike trains are acquired at 30 kHz sampling frequency, and the LFPs are acquired at 1 kHz sampling frequency. After excluding the recording channels that fail to capture activity (average

firing rate < 1Hz), 10 (Monkey A) and 38 (Monkey B) channels are considered for analysis. Cursor movements were tracked using the custom-built Python-based BMI3D interface. Neuronal signals were collected using Trellis (Ripple Neuro, UT, USA) interfacing with Python (v3.6.5) via Xipppy (v1.2.1), amplified, digitized, and filtered with the Ripple Grapevine System (Ripple Neuro, UT, USA).

3.3 MULTISCALE NEURAL DYNAMICS NEURAL ORDINARY DIFFERENTIAL EQUATION

3.3.1 Multiscale dynamics modeling with neurobiological constraints

We defined a multi-scale dynamics network as a collection of neural recordings from different modalities (e.g., spike trains, LFPs, EEGs, fast-scan cyclic voltammetry, calcium imaging, functional magnetic resonance imaging, and functional near-infrared spectroscopy). A generic multi-scale dynamics system, where the evolution of latent variables and the output was described by the nonlinear functions of latent states and corresponding inputs, for *M* modalities is as follows,

$$\dot{\mathbf{x}}_i = \sum_{j=1}^{M} f_{ij}(\mathbf{x}_j, \theta_j, t)\,,\ i \in [1, 2, \dots, M] \tag{3.1}$$

$$\mathbf{y}_i = g_{ii}(\mathbf{x}_i, \theta_i, t),\ i \in [1, 2, \dots, M] \tag{3.2}$$

where $\mathbf{x}_i$, $\mathbf{y}_i$ represent the latent state variables and the observations for i^{th} modality, respectively, f_{ij} denotes within-scale ($i = j$) and cross-scale ($i \neq j$) dynamics parameterized by θ_j, and g_{ii} is the observation model in each modality. In this work, we focused on the firing rates and LFPs (**Figures. 3.1-3.2**), referred to as multi-scale signals.

In addition, to enable the interpretability of the deep learning model, we introduced neurobiological constraints in our proposed network. Constraints including integration of modeling across different scales, the nature of the neuron model, regulation and control through interplay between excitatory and inhibitory neurons, and both local within- and global between-area connectivity have been reported to make neural network models more biologically plausible[123]. How are these neurobiological constraints implemented in the proposed approach will be shown in the following sections.

The multi-scale dynamics modeling for firing rate activity and LFP based on well-established neurobiological models can be divided into three parts: (1) firing rate-firing rate within-scale model, (2) LFP-LFP within-scale model, and (3) firing rate-LFP cross-scale model. The rate model was employed as the firing rate-firing rate inference model with N_{tol} coupled neurons[124–126].

$$\frac{dx_{firing\ rate,i}}{dt} = \frac{x_{firing\ rate,i} + \mathrm{sigm}\left(\sum_{j=1}^{N_{tol}}\left(C_{hidden\ firing\ rate,ij}\frac{dx_{firing\ rate,j}}{dt}(t) + C_{firing\ rate,ij}x_{firing\ rate,j}(t)\right)\right)}{-\tau_m} \quad (3.3)$$

where $x_{firing\ rate,i}$ represents the membrane voltage of neuron i, τ_m denotes the membrane time constant, $C_{firing\ rate,ij}$ and $C_{hidden\ firing\ rate,ij}$ represents two types of coupling strengths between presynaptic neuron j and postsynaptic neuron i. For the LFP-LFP within-scale model, we implemented the Jasen-Rit model to describe the local cortical circuit by second-order ODEs[127].

$$\ddot{x}_{LFP,i0} = Aa\,\mathrm{sigm}_0\left(x_{LFP,i1}(t) - x_{LFP,i2}(t)\right) - 2a\dot{x}_{LFP,i0} - a^2 x_{LFP,i0}(t) \quad (3.4)$$

$$\ddot{x}_{LFP,i1} = Aa[p_i(t) + C_2\,\mathrm{sigm}_1\left(C_1 x_{LFP,i0}(t)\right) + p_{mu} + C_{FR-LFP,ij}x_{firing\,rate,j}] - 2a\dot{x}_{LFP,i1} - a^2 x_{LFP,i1}(t) \quad (3.5)$$

$$\ddot{x}_{LFP,i2} = BbC_4\,\mathrm{sigm}_2\left(C_3 x_{LFP,i0}(t)\right) - 2b\dot{x}_{LFP,i2} - b^2 x_{LFP,i2}(t) \quad (3.6)$$

$$p_i(t) = \sum_j^{N_{tol}} C_{LFP,ij}\,\mathrm{sigm}_0\left(x_{LFP,j1}(t) - x_{LFP,j2}(t)\right) \quad (3.7)$$

where *sigm()* is a sigmoid function, *A* and *B* represent the maximum amplitude of the excitatory and inhibitory postsynaptic potentials (PSPs), *a* and *b* denote the reciprocal of the time constants of excitatory and inhibitory PSPs, $p_{mu}(t)$ represents the excitatory input noise of the neuron *i*, and *p(t)* represents the excitatory input of the neuron *i* from other neurons.

For the cross-scale model that identified and quantified cross-scale communications, we considered the coupling strengths between the hidden states (membrane voltage of single neuron for spike; membrane potential of pyramidal, inhibitory, and excitatory neurons) as the effective connectivity.

$$h_{LFP} = Ch_{spike} + \varepsilon \quad (3.8)$$

where *C* represents the cross-scale coupling strengths, and ε denotes the inputs from other units which are not taken into account plus error. Taken together, combining above equations, our multiscale dynamics model for spike and field potential could be written as

follows, where $F_{spike-spike}$ and $F_{LFP-LFP}$ represent the within-scale dynamics equations while $F_{spike-LFP}$ denotes the cross-scale dynamics equations,

$$\frac{d\mathbf{x}}{dt} = \begin{bmatrix} \frac{d\mathbf{x}_{spike}}{dt} \\ \frac{d\mathbf{x}_{LFP}}{dt} \end{bmatrix} = \begin{bmatrix} F_{spike-spike} & \mathbf{0} \\ \mathbf{0} & F_{LFP-LFP} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{spike} \\ \mathbf{x}_{LFP} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & F_{spike-LFP} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \frac{d\mathbf{x}_{spike}}{dt} \\ \frac{d\mathbf{x}_{LFP}}{dt} \end{bmatrix} + \begin{bmatrix} \mathbf{b}_{spike} \\ \mathbf{0} \end{bmatrix} \tag{3.9}$$

3.3.2 Multiscale neural dynamics neural ordinary differential equation (msDyNODE)

Popular models such as recurrent neural networks and residual networks learn a complicated transformation by applying a sequence of transformations to the hidden states: $\mathbf{h}_{t+1} = \mathbf{h}_t + f(\mathbf{h}_t, \theta_t)$. Such iterative updates can be regarded as the discretization of a continuous transformation. In the case of infinitesimal update steps, the continuous dynamics of the hidden states can be parameterized with an ordinary differential equation (ODE):

$$\frac{d\mathbf{h}(t)}{dt} = f(\mathbf{h}(t), \theta, t) \tag{3.10}$$

A new family of deep neural networks, termed the neural ODE (NODE), was thus introduced to parameterize the f using a neural network. The output of the NODE was then computed using any differential equation solver (e.g., Euler, Runge-Kutta methods). In this work, we utilize Runge-Kutta method with a fixed time step of 1 ms. The resulting

msDyNODE model consists of 7 layers with 1,480 and 18,392 trainable parameters for Monkey A and B, respectively. NODE exhibits a number of benefits, including memory efficiency, adaptive computation, and the capability of incorporating data arriving at arbitrary times. Moreau *et al.* proposed a NODE-based approach with a Bayesian update network to model the *sporadically* observed (i.e., irregular sampling) multi-dimensional time series dataset[128]. Therefore, NODE serves as a powerful tool for multi-scale data analysis.

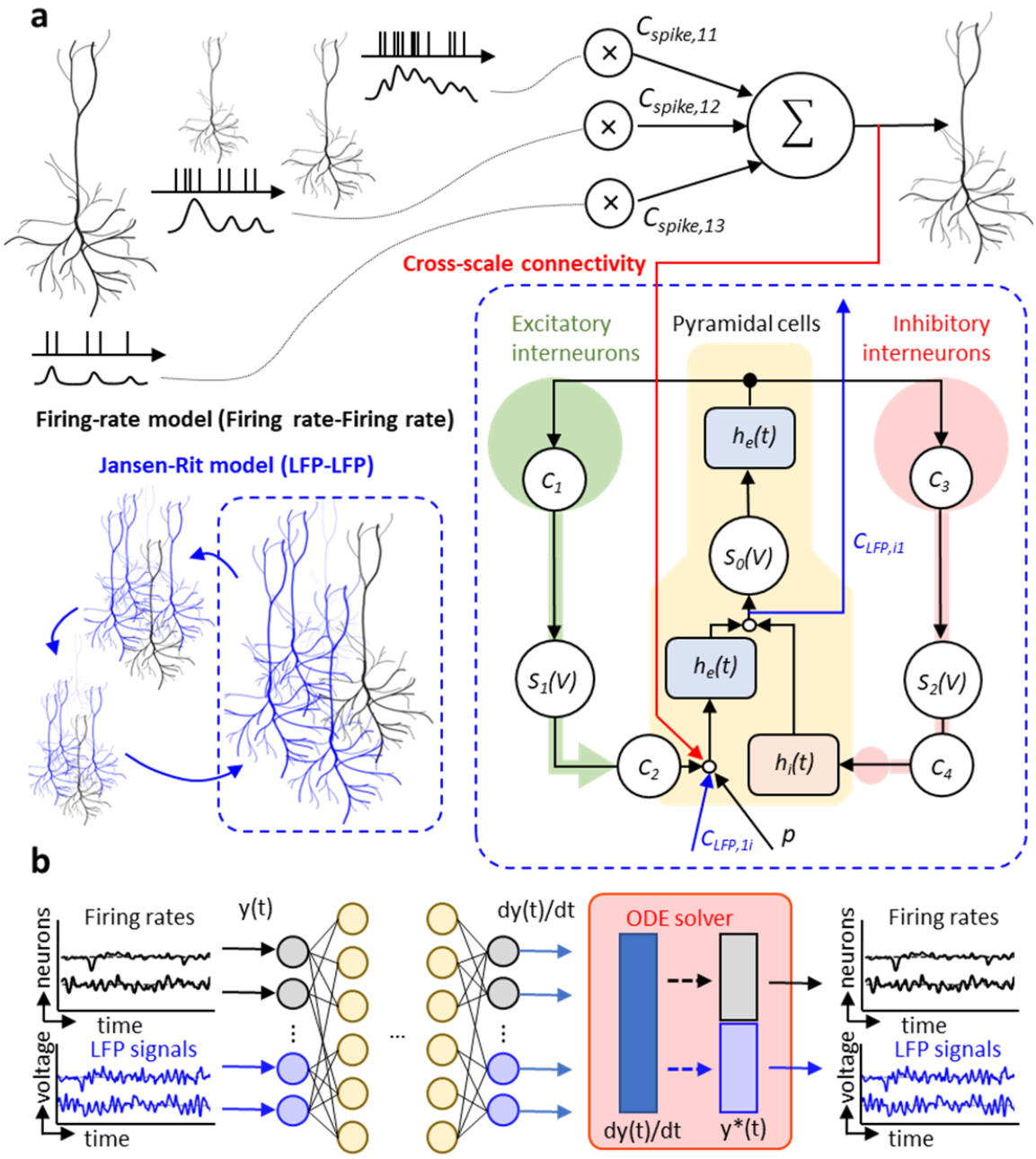

Figure 3.2 | The architecture of msDyNODE applied to multiscale LFP and firing rate. (**a**) Firing rate-Firing rate model follows the firing-rate model. LFP-LFP model follows the Jansen-Rit model. Cross-scale connectivity between firing rates and LFPs is added between latent variables of two systems. (**b**) The schematics of msDyNODE for multiscale firing rate-LFP model.

3.4 RESULTS

3.4.1 Synthetic Lorenz attractor

The Lorenz attractor model is becoming standard nonlinear dynamical system in the field with its simplicity and the easiness of state space visualization[29,31]. The Lorenz attractor is a simple but standard model of a nonlinear, chaotic dynamical system in the field. It consists of nonlinear equations for three dynamic variables. The state evolutions are derived as follows,

$$\dot{x}_1 = \sigma(x_2 - x_1) \tag{3.11}$$

$$\dot{x}_2 = x_1(\rho - x_3) - x_2 \tag{3.12}$$

$$\dot{x}_3 = x_1 x_2 - \beta x_3 \tag{3.13}$$

The standard parameters are σ = 10, ρ = 28, and β = 8/3. The Euler integration is used with Δt = 0.001. We first simulated two sets of Lorenz attractor systems with different parameter sets (σ_1 = 10, ρ_1 = 28, β_1 = 8/3, σ_2 = 8, ρ_2 = 20, and β_2 = 10/3) but without cross-scale interactions,

$$\dot{x}_1 = \sigma_1(x_2 - x_1) \tag{3.14}$$

$$\dot{x}_2 = x_1(\rho_1 - x_3) - x_2 \tag{3.15}$$

$$\dot{x}_3 = x_1 x_2 - \beta_1 x_3 \tag{3.16}$$

$$\dot{x}_4 = \sigma_2(x_5 - x_4) \tag{3.17}$$

$$\dot{x}_5 = x_4(\rho_2 - x_6) - x_5 \tag{3.18}$$

$$\dot{x}_6 = x_4 x_5 - \beta_2 x_6 \tag{3.19}$$

where one for a population of neurons with firing rates given by the Lorenz variables and another one for LFPs given by the Lorenz variables (**Figure 3.3**). We started the Lorenz system with a random initial state vector and ran it for 6 seconds. We hypothesize that the neural activity consists of multiple marginally stable modes[129,130]. The last five seconds were selected to ensure marginal stability in the simulation. Three different firing rates and LFPs were then generated with different sampling rates (1,000 Hz for spikes and 100 Hz for LFPs). Models were trained with ten batches of 1-second data with randomly selected starting points for 1,000 iterations.

To evaluate the fitting performance of the msDyNODE with the Lorenz systems with cross-scale interactions, we then simulated two sets of Lorenz attractor systems with different parameter sets (σ_1 = 10, ρ_1 = 28, β_1 = 8/3, σ_2 = 8, ρ_2 = 30, and β_2 = 10/3) and cross-scale interactions,

$$\dot{x}_1 = \sigma_1(x_2 - x_1) + 0.2x_4 + 0.6x_5 \tag{3.20}$$

$$\dot{x}_2 = x_1(\rho_1 - x_3) - x_2 + 0.8x_4 - 0.4x_6 \tag{3.21}$$

$$\dot{x}_3 = x_1 x_2 - \beta_1 x_3 - x_4 + 0.5x_5 + x_6 \tag{3.22}$$

$$\dot{x}_4 = \sigma_2(x_5 - x_4) + 0.5x_3 \tag{3.23}$$

$$\dot{x}_5 = x_4(\rho_2 - x_6) - x_5 - 0.4x_2 + 0.1x_3 \tag{3.24}$$

$$\dot{x}_6 = x_4 x_5 - \beta_2 x_6 + x_1 - 0.6x_2 \tag{3.25}$$

All the other simulation settings remain the same as above.

The msDyNODE framework was first demonstrated using the simulated Lorenz attractor dataset. A Python program was employed to generate synthetic stochastic neuronal firing rates and local field potentials from deterministic nonlinear system. Two sets of Lorenz attractor systems were simulated, representing firing rates at the single-neuron scale and LFPs at the neuronal population scale, respectively. Without causal interactions between scales, the msDyNODE well reconstructed the Lorenz attractor parameters, simulated firing rates and LFPs (mean absolute error = 0.64 Hz for firing rate; = 0.18 μV for LFPs; **Figure 3.3a**). To evaluate the performance of the msDyNODE in the multiscale system, we mimicked cross-scale interactions by adding causal connections between latent states of the two systems (**Figure 3.3b**). Although the fitting accuracy was relatively poorer than the systems without causal interactions (mean absolute error = 7.75 Hz for firing rate; = 6.08 μV for LFPs), the msDyNODE still captured the signals and the Lorenz attractor parameters (**Table 3.1**). Furthermore, we assessed if the msDyNODE can identify the types (excitatory or inhibitory) and the strength of the causal interactions (**Figure 3.3c**). Positive and negative causal strengths correspond to excitatory and inhibitory effect, respectively. The positive causality identified by the msDyNODE was true positive when the ground truth was also positive. It became a false positive if the ground truth was negative. The identification accuracy was 77±6% (**Figure 3.3c** left). However, if the causality with the coupling strengths less than 0.05 was counted as no connections, the identification accuracy became 100%. The msDyNODE was also demonstrated successfully to capture the cross-scale coupling strengths (mean absolute difference between the ground-truth and estimated causality = 0.07; **Figure 3.3c** right). These simulations verified that msDyNODE is a reliable framework for modeling multiscale systems.

Model parameters	Ground truth	Predictions (n = 10)
$\boldsymbol{\sigma_1}$	10	10.09 ± 0.04
$\boldsymbol{\rho_1}$	28	28.02 ± 0.06
$\boldsymbol{\beta_1}$	2.67	2.69 ± 0.03
$\boldsymbol{\sigma_2}$	8	7.87 ± 0.08
$\boldsymbol{\rho_2}$	20	19.82 ± 0.06
$\boldsymbol{\beta_2}$	3.33	3.45 ± 0.03

Table 3.1: msDyNODE captures the Lorenz attractor parameters.

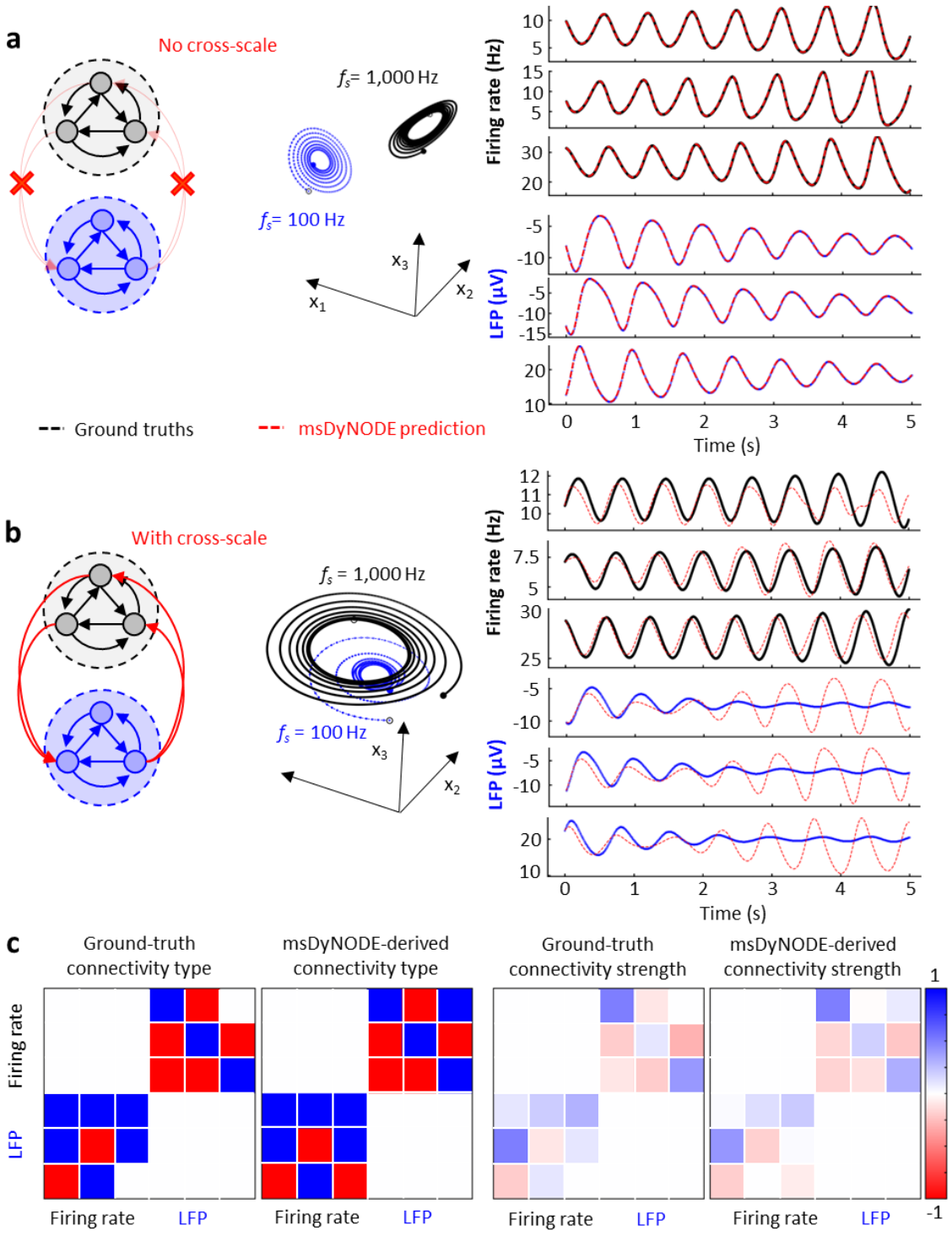
a
No cross-scale
f_s= 1,000 Hz
f_s= 100 Hz
x_1
x_2
x_3
Firing rate (Hz)
LFP (μV)
Time (s)
Ground truths
msDyNODE prediction
b
With cross-scale
f_s = 1,000 Hz
f_s = 100 Hz
x_1
x_2
x_3
Firing rate (Hz)
LFP (μV)
Time (s)
c
Ground-truth connectivity type
msDyNODE-derived connectivity type
Ground-truth connectivity strength
msDyNODE-derived connectivity strength
Firing rate
LFP
1
-1

Figure 3.3 | msDyNODE applied to Lorenz attractor. (**a**) The evolution of the Lorenz system in its 3-dimensional state space for firing rates (black) and LFPs (blue; top). The synthetic firing rates and LFPs, as well as the msDyNODE predictions, were plotted as a function of time (bottom). (**b**) The same as **a** but with cross-scale causal interactions. (**c**) Ground-truth and identified cross-scale communication types (left) and coupling strengths (right) between synthetic firing rates and LFPs.

3.4.2 msDyNODE outputs well reconstruct experimental firing rate and field potential signals

Firing rate and LFP activity are simultaneously recorded in the left dorsal premotor (PMd) and primary motor cortex (M1) regions of the rhesus macaque (N=2) while performing a center-out BMI task[131–135] (**Figure 3.1**). Multi-scale firing rate and LFP are acquired with the same set of electrodes but undergoing different pre-processing procedures. During the center-out BMI task, the subjects volitionally modulate brain activity to move the cursor from the center to one of the eight peripheral targets. The msDyNODE for the firing rate-LFP modeling was developed based on firing-rate model[124–126] and Jansen-Rit model[127] (**Figure 3.2**). By fitting the msDyNODE to the experimental datasets, we showed the success of the proposed multiscale framework on modeling the multiscale brain activity (**Figure 3.4**). We find the msDyNODE primarily captured the LFP activity with frequency less than 30 Hz by evaluating the correlation between the msDyNODE predictions before and after a low-pass filter (**Figure 3.4a**). This observation may be attributed to the neural dynamics that is dominated by lower temporal frequencies. Therefore, for the rest of the evaluations, we focused on the performance in the frequency range of 0 and 30 Hz. Overall, the msDyNODE well reconstructed the firing rates (median of mean absolute error (MAE) = 0.74 Hz) and LFPs (median MAE = 24.23 μV; **Figure 3.4b**). In addition, we find that the performance of the msDyNODE is target direction-

independent, with a similar MAE over eight target directions for both firing rates and LFPs (**Figure 3.4c**). Interestingly, the reconstruction performances for firing rates and LFPs were independent (**Figure 3.4d**). Good performance on certain channels did not indicate similarly good performance for different signal types, and vice versa. Surprisingly, the modeling performance for firing rates remained high over hundreds of trials even perturbation is introduced to increase the task difficulty (**Figure 3.4e**). However, the modeling performance for LFP gradually improves over trials, which may indicate that LFP dynamics became more predictable. Furthermore, the performance holds when applying the msDyNODE to a different monkey dataset (i.e., that it is not trained on), indicating that the msDyNODE is generalizable across different sessions and subjects (**Fig. 3.5**). With a larger number of spiking units and LFPs recorded in this subject, it is expected that the msDyNODE can reconstruct LFP more accurately. The only difference in the reconstruction performance is that the firing rate predictions are worse during the first half of the experimental sessions, followed by increasing accuracy for the second half of the recording sessions (**Fig. 3.5e**). This observation may indicate the neural dynamics is more chaotic and thus more challenging to be captured during the first half of the sessions. These experimental results validated that msDyNODE could capture the dynamics hidden in the multiscale brain systems.

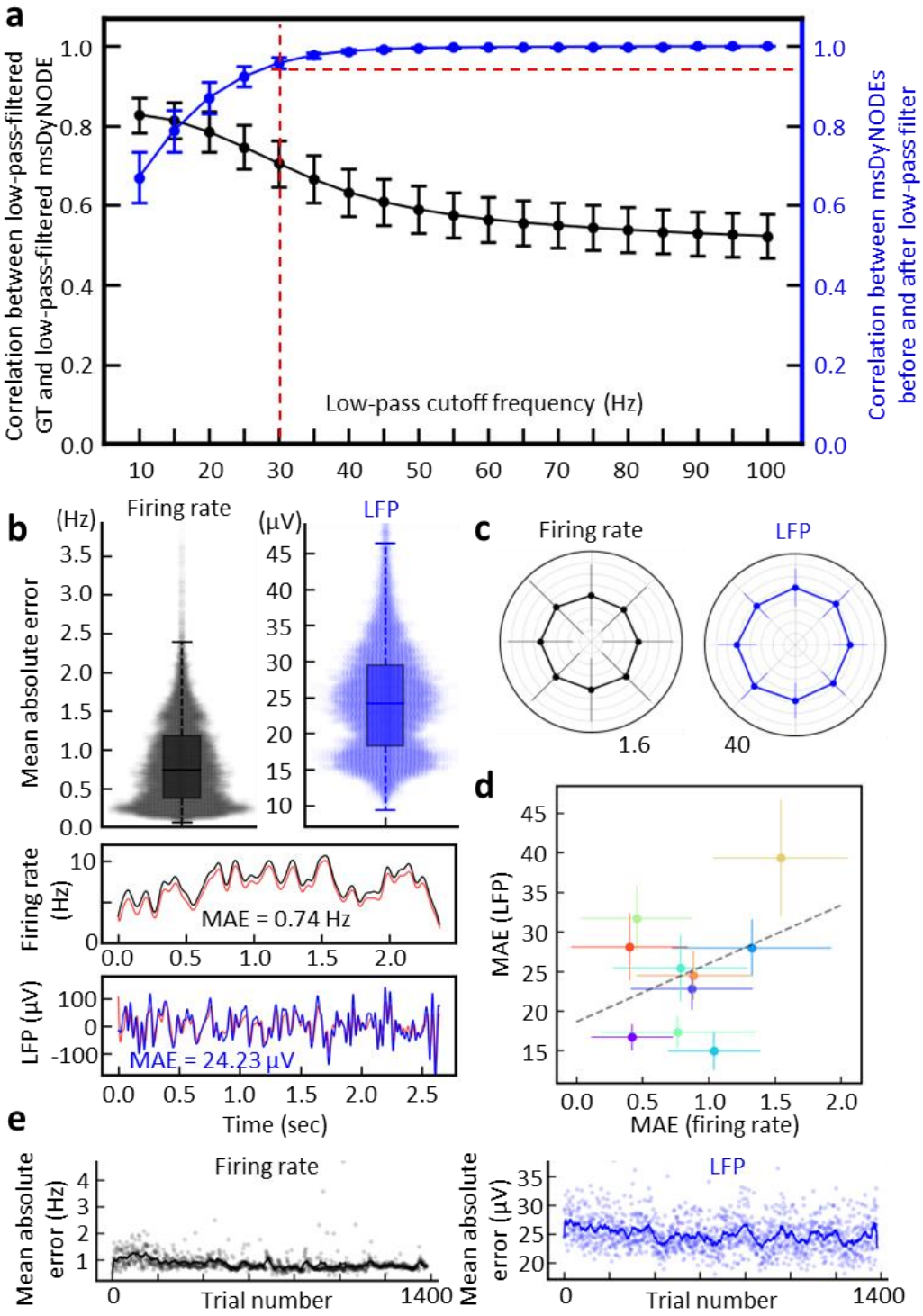
a
Correlation between low-pass-filtered GT and low-pass-filtered msDyNODE
Correlation between msDyNODEs before and after low-pass filter
Low-pass cutoff frequency (Hz)
b
Firing rate
LFP
Mean absolute error
MAE = 0.74 Hz
MAE = 24.23 µV
Time (sec)
c
Firing rate
LFP
1.6
40
d
MAE (LFP)
MAE (firing rate)
e
Mean absolute error (Hz)
Firing rate
Trial number
Mean absolute error (µV)
LFP
Trial number

Figure 3.4 | msDyNODE captures and reconstructs the latent dynamics in the center-out BMI task for Monkey A. (**a**) Correlation coefficient between low-pass-filtered ground truths (GT) and low-pass-filtered msDyNODE (black) and between msDyNODE before and after low-pass filter (blue) as a function of low-pass cutoff frequency (error bars, s.t.d.). (**b**) Boxplots and swarmplots of the mean absolute errors in firing rates and LFPs (top). The representative GT and msDyNODE with the MAE equaling to the median values of all the MAEs (bottom). (**c**) Error bars of the MAE over eight different target directions presented in polar coordinates (error bars, s.t.d.). (**d**) Bar plots of the MAE over recording channels (error bars, s.t.d.). (**e**) MAE values of firing rates and LFPs over trials. Dim points represent average MAE (n = 10) at each trial.

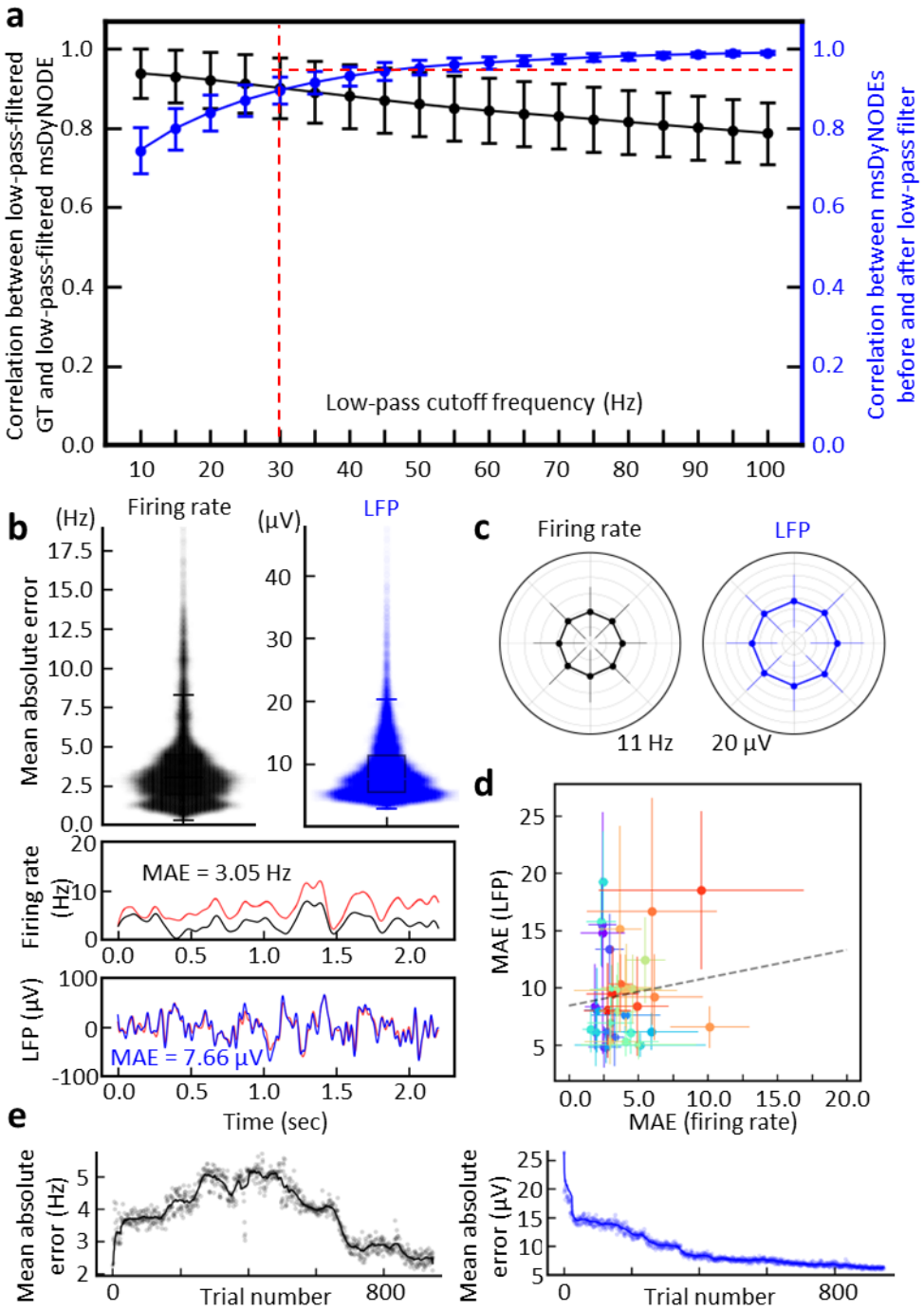
a
Correlation between low-pass-filtered GT and low-pass-filtered msDyNODE
Correlation between msDyNODEs before and after low-pass filter
Low-pass cutoff frequency (Hz)
b
Firing rate
LFP
Mean absolute error
MAE = 3.05 Hz
MAE = 7.66 µV
Time (sec)
c
Firing rate
LFP
11 Hz
20 µV
d
MAE (LFP)
MAE (firing rate)
e
Mean absolute error (Hz)
Mean absolute error (µV)
Trial number

Figure 3.5 | msDyNODE captures and reconstructs the latent dynamics in the center-out BMI task for Monkey B. (**a**) Correlation coefficient between low-pass-filtered ground truths (GT) and low-pass-filtered msDyNODE (black) and between msDyNODE before and after low-pass filter (blue) as a function of low-pass cutoff frequency (error bars, s.t.d.). (**b**) Boxplots and swarmplots of the mean absolute errors in firing rates and LFPs (top). The representative GT and msDyNODE with the MAE equaling to the median values of all the MAEs (bottom). (**c**) Error bars of the MAE over eight different target directions presented in polar coordinates (error bars, s.t.d.). (**d**) Bar plots of the MAE over recording channels (error bars, s.t.d.). (**e**) MAE values of firing rates and LFPs over trials. Dim points represent average MAE (n = 38) at each trial.

3.4.3 msDyNODE decodes underlying behavior via multiscale effective connectivity

In msDyNODE, the msEC can be derived from the parameters that indicate the causal influence that the latent states of a neural system exert over those of another system. The average connectivity for each target direction is calculated by subtracting the grand-averaged connectivity from the average connectivity within each target (**Fig. 3.6a**). For each direction, the bi-directional msEC is divided into two parts (upper and lower triangular connectivity matrix) and visualized respectively (**Fig. 3.6b**). Most of the msEC remained similar across target directions, indicating the common patterns of voluntary movement. To investigate if there exist unique patterns of excitatory and inhibitory subnetworks across directions, we quantify the individual subnetworks using common graph properties such as number of edges, average clustering, and total triangles (**Fig. 3.6c**). Interestingly, these graph properties are different across the eight target directions, revealing the excitatory and inhibitory neural dynamics exhibit unique connectivity patterns underlying target directions. Thus, msDyNODE is demonstrated to be capable of capturing the multiscale effective connectivity patterns underlying behaviors.

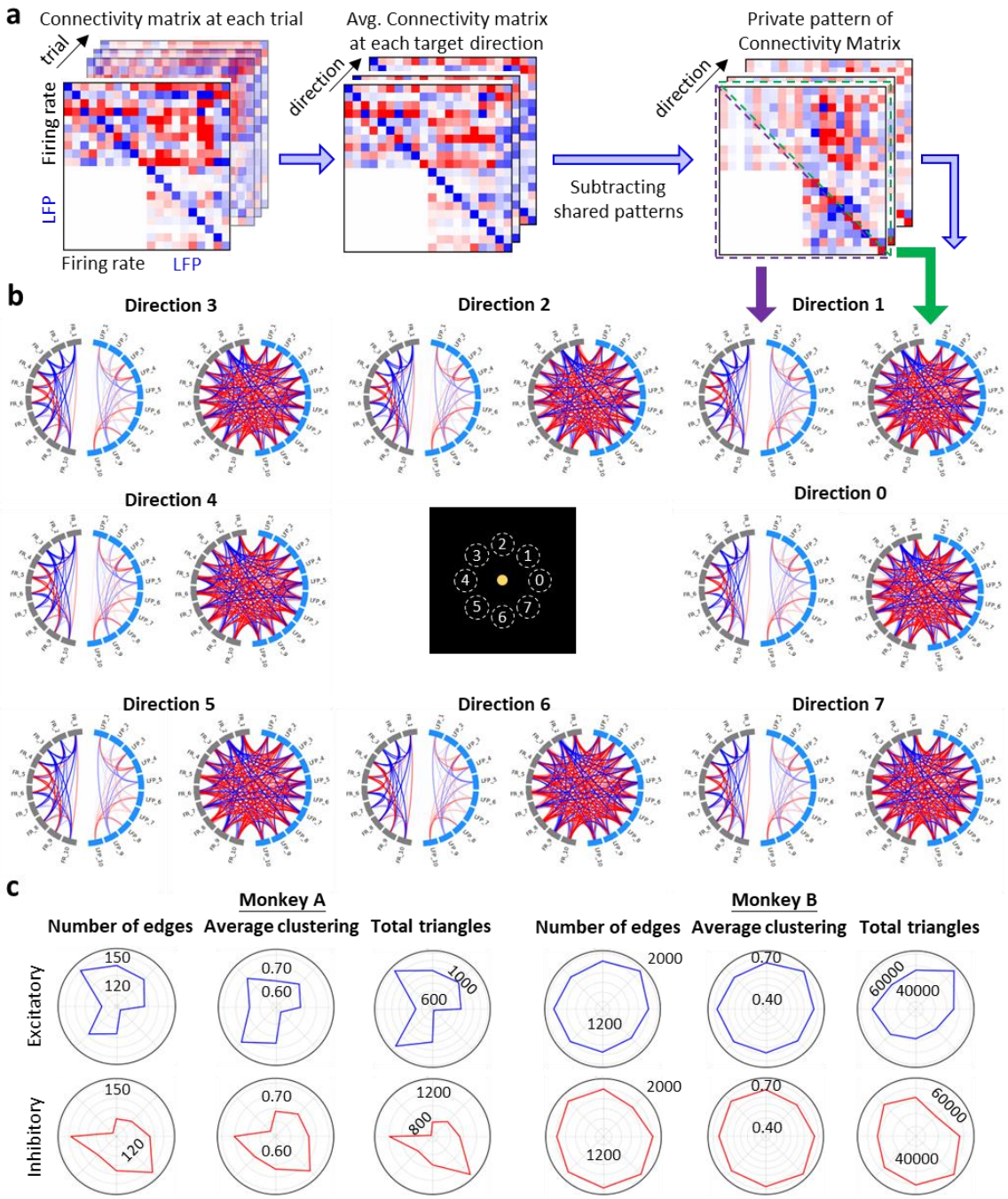
a
Connectivity matrix at each trial
trial
Firing rate
LFP
Firing rate
LFP
Avg. Connectivity matrix
at each target direction
direction
Subtracting
shared patterns
Private pattern of
Connectivity Matrix
direction
b
Direction 3
Direction 2
Direction 1
Direction 4
Direction 0
Direction 5
Direction 6
Direction 7
c
Monkey A
Monkey B
Number of edges
Average clustering
Total triangles
Excitatory
Inhibitory
150
120
0.70
0.60
1000
600
2000
1200
0.70
0.40
60000
40000
150
120
0.70
0.60
1200
800
2000
1200
0.70
0.40
60000
40000

Figure 3.6 | msDyNODE captures msEC patterns underlying behaviors. (**a**) Workflow to obtain the private pattern of connectivity matrix for each target direction from msDyNODE-inferred msEC. (**b**) Circular connectivity graphs of lower (left) 丂and upper (right) triangular msEC matrix for each target direction. (**c**) Graph properties (number of edges, average clustering, number of total triangles) over eight different target directions presented in polar coordinates for Monkey A and B, and excitatory and inhibitory subnetworks, respectively.

3.5 DISCUSSION

Billions of individual neurons coordinate their activity together to achieve a specific cognitive task. But such a task cannot grow so large as to recruit all of the brain. Only the associated neurons will be recruited while the irrelevant ones will be left, highlighting the importance of studying the coordination of neural activity. Over the past decades, we have learned much about the human cognitive behaviors and viewed an explosive growth in the understanding of single neurons and synapses[136,137]. However, we still lack a fundamental understanding of multiscale interactions. A critical barrier to multiscale study was recording technologies for decades, forcing the scientists to choose either the microscale or macroscale, with few researchers addressing on the interactions between scales. Neurophysiologists, for example, often focused on single-neuronal activity to investigate the sensory consequences of motor commands with the bottom-up approach[138], without the consideration of brain rhythm. Instead, cognitive neuroscientists pay attention to the neural oscillations at a larger scale (e.g., electroencephalography) with the top-down approach to establish the links between brain rhythm and cognitive behaviors[139], disregarding the spiking activity of single neurons. With the advancement of multi-modal measurements, there is an unmet need for integrative framework to analyze multiscale systems. In the present study, we proposed the msDyNODE to model the multiscale signals of firing rates and field potentials, and then inferred multiscale causal interactions that exhibited distinct patterns for different motor behaviors.

To the best of our knowledge, this is the first demonstration of a NODE applied to model multiscale neural activity. Assuming the brain computation as the nonlinear operator[140–146], we employ the deep learning technique to approximate the nonlinear mapping of the state variables in dynamic systems. While recurrent neural network (RNN)

is a powerful approach to solve the dynamic equations[147,148], it may loss faster dynamics information or introduce artifacts by matching the sampling rates between signals. In contrast to the RNN which describes the complicated state transformation at discretized steps for time-series inference, the proposed msDyNODE rather model continuous dynamics by learning the derivative of the state variables[149]. It is worth noting that the nonconstant sampling can be addressed by preprocessing the NODE output with the observation mask[128]. Therefore, unmatched sampling rate between modalities can be resolved by feeding individual observation masks, respectively. Furthermore, in the real world, not all the signals can be measured at fixed time intervals. The missing data issue can thus introduce artefacts using conventional approach which assumes the signals are sampled regularly. While there exists several methods such as dropping variables, last observation carried forward and next observation carried backward, linear interpolation, linear regression, or imputation dealing with missing data[150], none of them serves as good way to deal with this issue because they adds no new information but only increases the sample size and leads to an underestimate of the error. The proposed framework also holds great potential to be an alternative approach dealing with missing data commonly seen in the real world.

While msDyNODE provides accurate analysis for multiscale systems, its cost lies in the optimal selections of neural models. At the scale of firing rate, integrate-and-fire model and its variants (leaky integrate-and-fire[151,152] and quadratic integrate-and-fire[153,154]) are all plausible options. At the scale of field potential activity, the candidate model includes Jansen-Rit model[127] that characterizes three populations (pyramidal cells, excitatory interneurons, and inhibitory interneurons) and Wilson-Cowan model[155] that refers to two coupled populations (excitatory and inhibitory). Suboptimal selections of

neural models may result in misleading conclusions. To avoid suboptimal model selection, probabilistic statistical measures such as Akaike Information Criterion[156,157], Bayesian Information Criterion[158], and minimum description length[159,160] can be implemented to ensure the correct selection of the neural models.

Recent evidence has suggested that the signal changes on multiple timescales at multiple levels in the motor system allow arbitration between exploration and exploitation to achieve the goal[161–164]. Still, the role of cross-scale, as well as within-scale, causal interactions in motor learning remains incompletely understood[162,163,165]. In this work, we utilized msDyNODE to study the essential brain function that modulates the motor commands to achieve desired actions, showing distinct dynamic patterns underlying different behaviors. Taken together, our work represents an important step forward towards multiscale modeling of brain networks for mechanistic understanding of brain computations. The underlying multiscale dynamics embedded in msDyNODE illustrate how the individual neurons and populations of neurons communicate across scales, which is a key factor in uncovering the mechanisms of brain computations and the mediation of the behaviors.

Chapter 4: Spatial resolution enhancement in photon-starved STED imaging

4.1 INTRODUCTION

Super-resolution imaging techniques allow fine structures of subcellular organelles, protein complexes, and membrane domains to be studied in their native environments, revealing new functions of these molecular assembles that were not previously known[166–169]. Among these techniques, stimulated emission depletion (STED) microscopy is particularly suited to image dynamic processes as its light-driven off-switching mechanism well supports a high recording speed[169–172]. STED microscopy keeps the fluorescent molecules within a sub-diffraction volume at an "on" state whilst switching "off" the neighboring molecules (or, more precisely, returning the excited fluorophores back to the ground state without emitting fluorescence) using a donut-shaped depletion laser (i.e., the STED beam)[166,173]. Although the spatial resolution of STED microscopy can be further enhanced by continuously increasing the STED-beam power, it comes at the price of severe photobleaching and phototoxcity[174,175]. For example, to achieve 50% of probability of returning an excited fluorophore to the ground state within its excited-state lifetime (τ, about 4 ns) using stimulated emission (with energy normalized cross section of 10 cm^2/J), a STED-beam intensity of 25 MW/cm^2 is required[175]. Such an intense excitation either directly damages the live samples or prohibits any long-term observation of the samples[176]. Several techniques have been developed to maintain or improve the spatial resolution while lowering the STED-beam power[177,178]. The key lies in reducing the noise in the imaging acquisition process. One source of noise is the anti-Stokes excitation, which can be easily removed by subtracting the STED-beam-only background from the STED images[179–183]. Another source of noise is the "early fluorescence" from the donut-shaped peripheral

region[179]. Using a pulsed laser and the time-correlated single-photon counting (TCSPC) scheme, this early fluorescence from the peripheral region can be eliminated by time-gated analysis, creating a method called gated-STED (g-STED) that shows improved STED imaging resolution (**Figure 4.1a**)[173,184,185]. However, not limited to the unwanted photons from the peripheral region, g-STED also discards the wanted photons from the central region that arrive at the detector early. As both noise and signal are reduced, the signal-to-noise ratio (SNR) is not necessarily improved[186].

To differentiate the central fluorescence (signal) from the peripheral fluorescence (noise, due to imperfect depletion), Lanzanò et al. developed a method termed SPLIT (separation of photons by lifetime tuning) that integrates a low-excitation-power continuous-wave STED beam (CW-STED) with a phasor strategy to separate photons[187]. Since CW-STED exhibited less depletion efficiency and was more susceptible to the background noise, the continuous-wave depletion beam was later replaced by the pulsed beam (pSTED, **Figure 4.1b** and **Figure 4.2**)[188,189]. The pSTED-SPLIT method relied on a phasor plot for photon separation, where the decay histogram at each pixel is converted into a phasor point by Fourier transform[190]. Considering the instrument response function (IRF), the noise, $n(t)$, and the measured decay signals, $m(t)$, the response was described as follows,

$$m(t) = IRF \otimes I(t) + n(t) \tag{4.1}$$

where $I(t)$ represented fluorescence decay before the convolution with the IRF. After applying the fast Fourier transform to the measured data,

$$\vec{M}(\omega) = \mathcal{F}\big(m(t)\big) = \mathcal{F}(IRF \otimes I(t) + n(t)) = \mathcal{F}(IRF) \times \mathcal{F}\big(I(t)\big) + \mathcal{F}\big(n(t)\big) \tag{4.2}$$

We noticed that in the frequency domain, the convolution operation with the IRF was just a simple multiplication of the IRF vector. Therefore, during the calibration procedure, we would obtain the IRF. Then for the rest of the experiment, we simply subtracted the noise vector and divided the denoised vector by the IRF vector to get the emission fluorescent vector. Alternatively, the Fourier Transform of the fluorescence decay histogram could be separated into a real number, $g(\omega)$, and an imaginary number, $s(\omega)$.

$$g(\omega) = \frac{\int_0^\infty m(t)\cos\omega t dt}{\int_0^\infty m(t)dt} \tag{4.3}$$

$$s(\omega) = \frac{\int_0^\infty m(t)\sin\omega t dt}{\int_0^\infty m(t)dt} \tag{4.4}$$

After calibration, the fluorescent decay histogram (without IRF and noise) at each pixel could be plotted as a single point, termed phasor, on the phasor plot by applying the sine and cosine transforms to the measured decay data, where the modulation frequency ω was the laser repetition angular frequency and was calculated by multiplying the laser repetition rate with 2π.

A linear decomposition algorithm was then employed to separate the long-lifetime photons embedded in each phasor point (wanted photons presumably from the center, which is denoted as the $\boldsymbol{P_1}$ phasor component) from the short-lifetime photons (unwanted photons from the periphery, denoted as $\boldsymbol{P_2}$ phasor component). When the $\boldsymbol{P_1}$ component was restored and put back to the original scanned image, a STED image with improved resolution was obtained[191]. While SPLIT clearly outperformed g-STED in its imaging quality, the unmixing accuracy of SPLIT was still limited by the shot noise (square root of the number of photons collected at each pixel)[177,192]. Due to the limited photon budgets in

imaging biological samples, widely scattered and dislocated phasor points were often seen in a phasor plot, making high-fidelity SPLIT analysis challenging.

To overcome the low-photon-budget issue, we previously developed a deep-learning framework termed *flimGANE* (fluorescence lifetime imaging based on Generative Adversarial Network Estimation) to denoise the time-resolved measurements, generating high-quality fluorescence lifetime images[193]. Here we are combining *flimGANE* with SPLIT, creating a method termed STED-*flimGANE* (**Figure 4.1c**) that can achieve an enhanced pSTED imaging resolution under a low STED-beam power (< 30 mW) and photon-starved conditions (< 200 photons per pixel). In addition, our STED-flimGANE approach shows minimum dependence of spatial resolution on the STED power. In our method, a GAN model is trained to transform low-photon-count fluorescence decays into highly realistic artificial high-photon-count decays, thus resulting in denoised phasors that can be used to restore fluorophores' true distributions within nanoscale domains. Using our STED-*flimGANE* approach, an average resolution of 64 ± 5 nm is obtained when imaging the 60 nm fluorescent beads (n > 100). We also demonstrate STED-*flimGANE* imaging of nuclear pore complexes (NPCs; having an average size of 60 nm)[194] on COS-7 cells and achieved a spatial resolution of 77 nm. It is verified that STED-*flimGANE* can provide up to 1.45-fold resolution enhancement with less dependence on STED-beam power as compared to the traditional pSTED-SPLIT method.

4.2 DATA ACQUISITION FOR STED IMAGING

4.2.1 The time-resolved pSTED microscope

The time-resolved pulsed STED (pSTED) was implemented on an ISS Alba v5 laser scanning system, which was built either upon a Nikon TE2000 microscope equipped with a 60X NA=1.4 oil objective lens (CFI Plan APO λ, Nikon) for imaging 60 nm fluorescence beads, or a Zeiss Axiovert 200m microscope equipped with a Zeiss Plan-Apochromat 100X NA=1.4 oil objective lens for fixed COS-7 cells. A 642 nm diode laser with pulse duration around 120 ps (QuixX 642-140PS, Omicron-Laserage) and a 775 nm fiber laser with duration around 600 ps (Katana-775HP, NKT Phontonics) were used for fluorescence excitation and depletion, respectively (**Figure 4.1b; Figure 4.2**). A set of Galvo mirrors and a Z-piezo stage (Nano-Z200, Mad City Labs) were employed for 3D scanning of the samples. Both the excitation and the STED lasers were in sync with the lifetime acquisition/analysis module (FastFLIM, ISS), which was triggered at the clock rate of either 50 MHz (for beads samples) or 80 MHz (for biological samples) by the STED laser. The 642 nm laser had a built-in picosecond delayer (10 ps resolution, 0-28 ns tuning range) for precise tuning of the temporal separation between the excitation and depletion pulses. The optimal delay was obtained by comparing the pSTED results of the beads while adjusting the delay in 50-ps steps. The two beams were combined by a 670 nm long-pass dichroic mirror (DM1 in **Figure 4.1b**, FF685-Di02, Semrock). A custom-made multi-band dichroic mirror (DM2 in **Figure 4.1b**, zt473-491/561/640/2p-trans-pc-uf2, custom-made by Chroma) was used to separate fluorescence emission and stimulated emission and reflect the de-scanned emission light. The fluorescence was collected by an avalanche photodiode (SPCM-ARQH-15, Excelitas), after being filtered by a 720 nm short-pass filter (OD-8 720 nm, Chroma) and a band-pass emission filter (679/41 nm, Semrock). A tuneable

and motorized confocal pinhole (with size tuning range from 20 μm to 1 mm) was placed in front of the detector to suppress the background from out-of-focus planes. For the pSTED imaging, the pinhole size was fixed at 50 μm. Photon histograms acquired by the FastFLIM module were converted into a phase plot. A diluted Cy5 solution, with ~1 ns lifetime, was used to calibrate the FastFLIM system before imaging the samples. Each bead or cellular sample was scanned only once with a dwelling time of 0.4 ms per pixel, generating a 512 x 512 image (with an image size of approximately 10.63 x 10.63 μm or 14.64 x 14.64 μm, depending on the location of field-of-view). Data acquisition and a part of the analyses were performed using the ISS VistaVision 64-bit software. The rest of the analyses (*flimGANE*) were performed in Python.

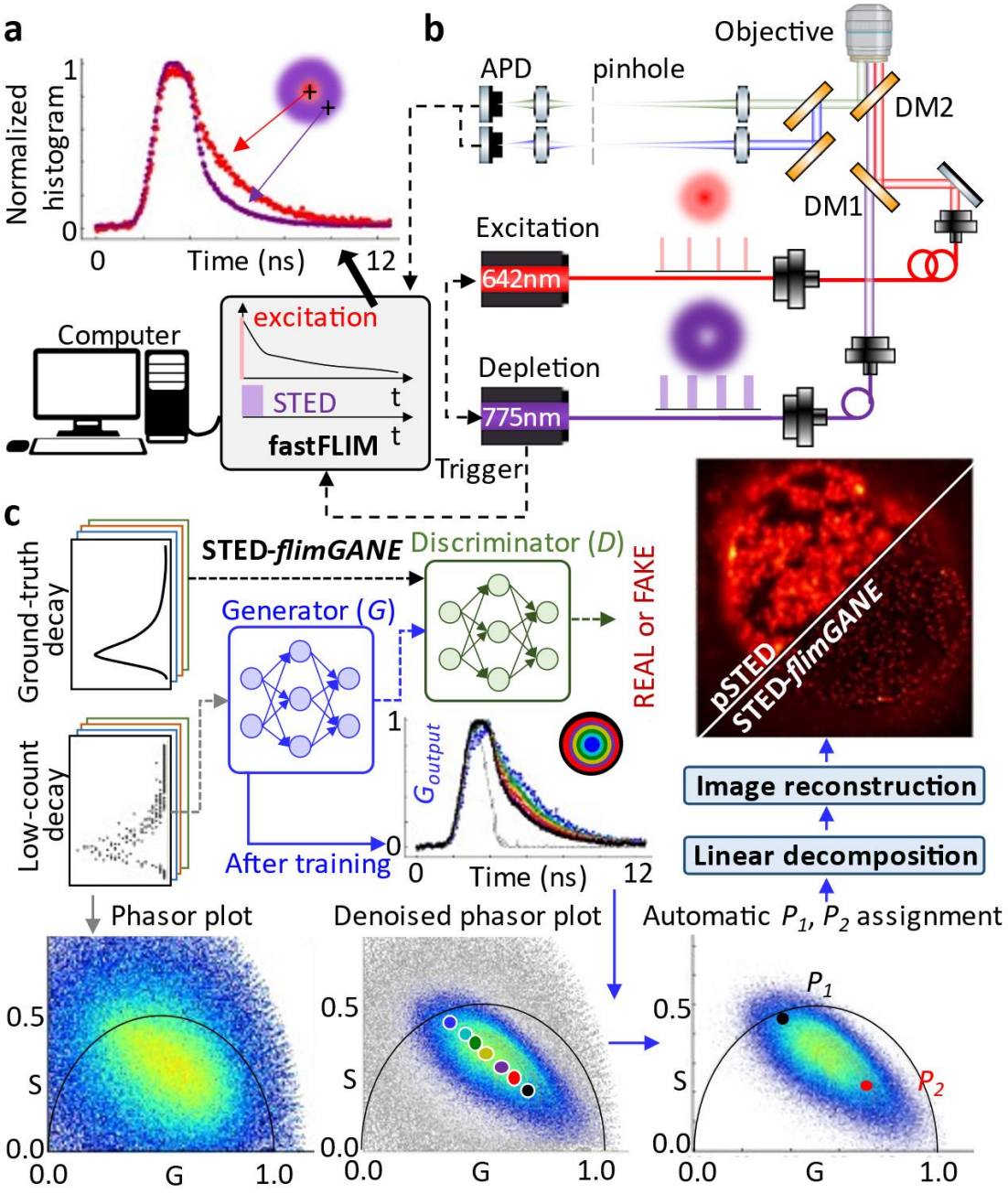
a
Normalized histogram
1
0
0
Time (ns)
12
b
Objective
APD
pinhole
DM2
DM1
Excitation
642nm
Depletion
775nm
Trigger
Computer
excitation
STED
t
t
fastFLIM
c
STED-flimGANE
Discriminator (D)
Generator (G)
Ground-truth decay
Low-count decay
After training
REAL or FAKE
1
0
G_{output}
0
Time (ns)
12
pSTED
STED-flimGANE
Image reconstruction
Linear decomposition
Phasor plot
Denoised phasor plot
Automatic P_1, P_2 assignment
0.5
S
0.0
0.0
G
1.0
P_1
P_2

Figure 4.1 | **Principle of the STED-*flimGANE*.** (a) The temporal dynamics of fluorescence can be used to improve the STED imaging resolution. When the distance of a fluorophore to the center of excitation increases, the depletion effect kicks in and increases the decay rate. (b) pSTED microscopy setup uses a pulsed excitation and pulsed depletion beam (i.e., the STED beam). The two laser beams are combined by a dichroic mirror (DM1), forming diffraction-limited Gaussian (red) and doughnut-shaped focal intensity distribution (purple), respectively. The fluorescence (green and blue) is detected by the avalanche photodiodes (APDs) and registered by a digital-frequency-domain lifetime acquisition module. (c) Our STED-*flimGANE* method integrates the SPLIT (separation of photons by lifetime tuning) scheme with a deep learning-based phasor analysis algorithm termed *flimGANE* (fluorescence lifetime imaging based on a generative adversarial network) to improve the STED imaging resolution while using a lower STED-beam power. The well-trained STED-*flimGANE* model can generate a robust phasor representation (*G*-generated phasor plot) with less scatter for reconstructing STED images with higher spatial resolution.

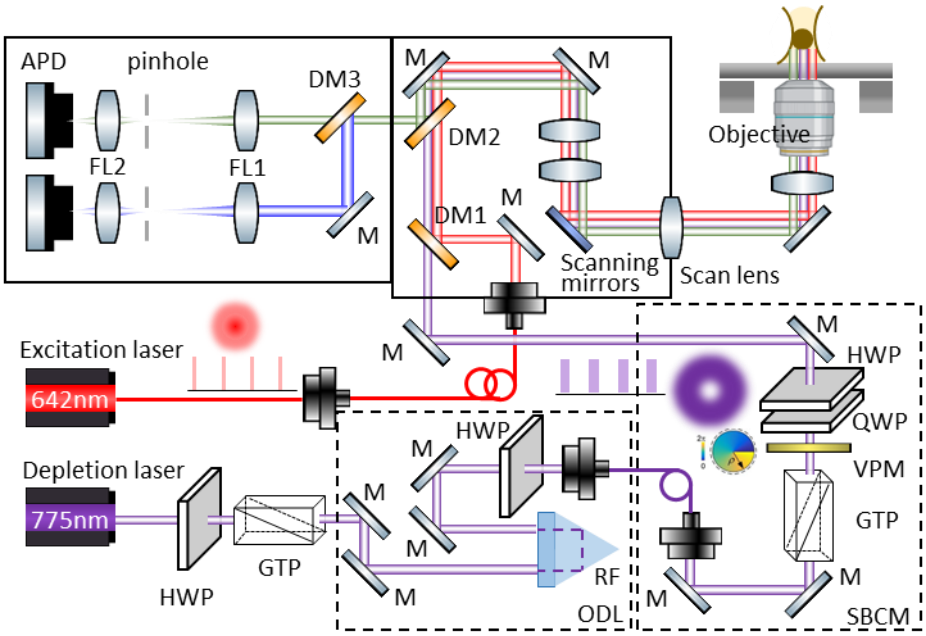

Figure 4.2 | **Details of the pulsed STED (pSTED) microscope.** HWP: halfwave plate. GTP: Glan-Thompson polarizer. ODL: optical delay line. M: mirror. DM: dichroic mirror. FL: focal lens. RF: retroreflector. SBCM: STED beam conditioning module. QWP: quarter wave plate. VPM: vortex phase mask. APD: avalanche photodiode.

4.2.2 Sample preparation and image acquisition

We performed Two types of samples, 60 nm fluorescent beads and STAR-635P-stained nuclear pore complexes (NPCs) on fixed COS-7 cells, were tested using the STED-*flimGANE* method. The beads were first diluted in water and sparsely coated on a poly-L-lysine-treated glass coverslip. The fixed COS-7 cells were stained with a primary antibody against the nuclear pore complex protein Nup 153 and then a secondary antibody conjugated with STAR-635P. The power of the 642 nm laser for fluorescence excitation was fixed at 2.2 μW (measured at the back aperture of the objective), while the power of the 775 nm depletion beam varied for spatial resolution investigation.

4.3 DATA ANALYSIS

4.3.1 STED-*flimGANE*

To further enhance the resolution of the pSTED-SPLIT method, the noise of the phasor plot needs to be suppressed and the assignment of the $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$ components must be accurate and highly reproducible. To improve the quality of the phasor plot, here we applied the generative adversarial network (GAN) model to denoise the phasors under photon-starved conditions (**Figure 4.1c**). The generative model learned to generate realistic objects from noisy inputs via an adversarial process[195].31 Based on the Wasserstein GAN framework[196], the generator (*G*) in our *flimGANE* algorithm was trained to produce an "artificial" high-photon-count fluorescence decay histogram based on a low-photon-count input (photon counts < 200 per pixel), while the discriminator (*D*) distinguished the artificial decay histogram from the ground-truth (**Figure 4.1c**)[193]. Described in more details in Section 4.3.2 and **Figure 4.3**, our generator *G* was built on convolutional neural

networks (CNN), fully connected neural networks (FCNN), and residual neural networks (ResNet), while the discriminator *D* was comprised of four fully connected layers. To achieve the training goal that maximized the probability of the discriminator making a mistake, the GAN model was trained by minimizing both the generator loss and discriminator loss, which were defined as:

$$G_{loss} = \frac{1}{n}\sum_{i=1}^{n}\left[-f\left(G(z_i)\right)\right], \tag{4.5}$$

$$D_{loss} = \frac{1}{n}\sum_{i=1}^{n}\left(-f(x_i) + f\left(G(z_i)\right)\right), \tag{4.6}$$

where z_i represented the i^{th}normalized low-photon-count fluorescence decay histogram, x_i was the i^{th} normalized ground-truth fluorescence decay histogram (obtained by simulations), and $f(x)$ was the 1-Lipschitz function approximated by the discriminative model. In order to avoid mode collapse, a common issue in GAN where the *G* got stuck in a small space with low variety[196], we included additional loss functions between the generator output and the ground truths. Eventually, a well-trained *G* enabled the production of very realistic, artificial high-photon-count decays that could be used to generate a robust phasor representation with less scatter in the frequency domain.

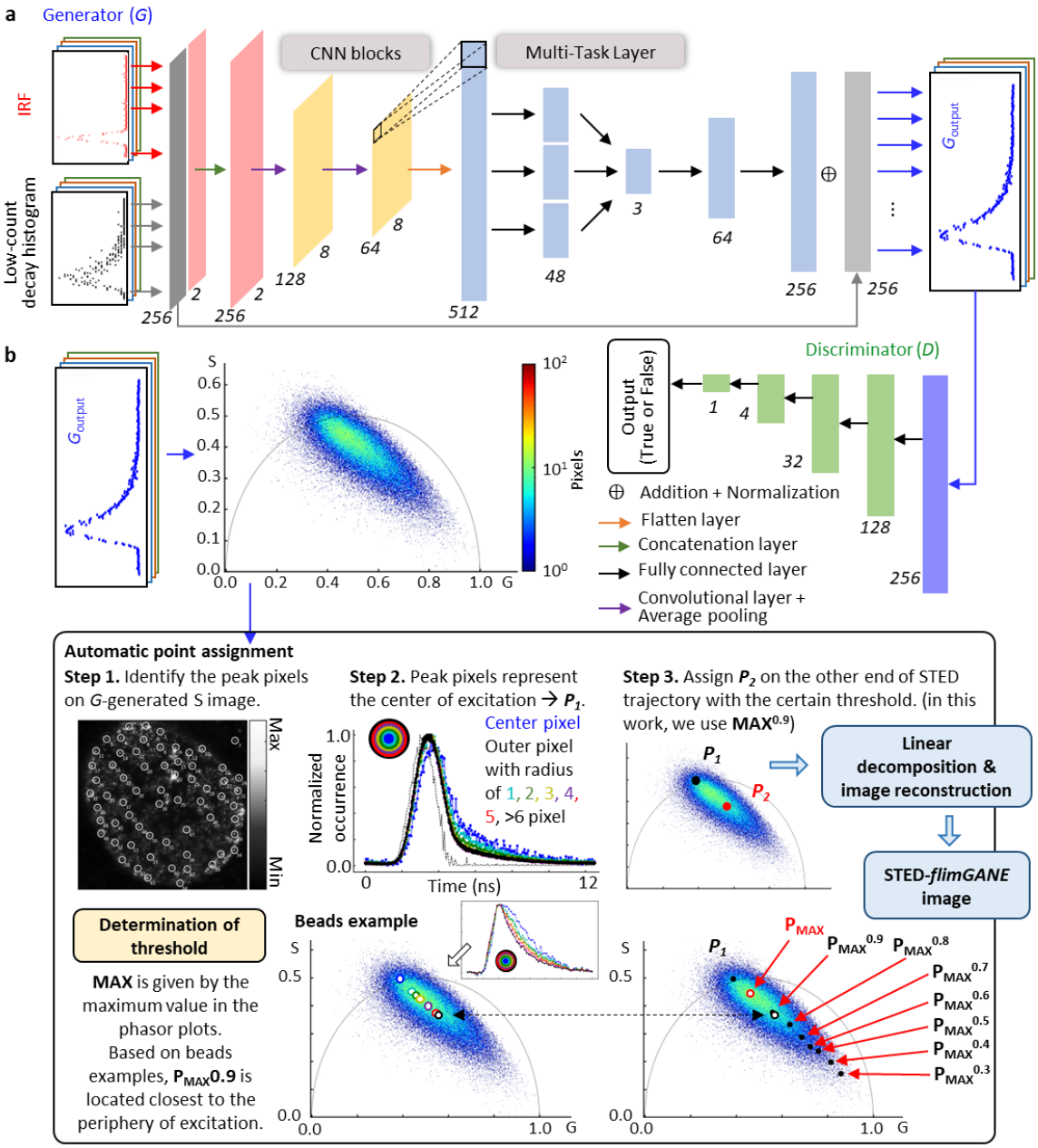
a
Generator (G)
IRF
Low-count decay histogram
CNN blocks
Multi-Task Layer
256
2
256
2
128
8
64
8
512
48
3
64
256
256
G output
b
G output
S
Pixels
Output (True or False)
Discriminator (D)
1
4
32
128
256
Addition + Normalization
Flatten layer
Concatenation layer
Fully connected layer
Convolutional layer + Average pooling
Automatic point assignment
Step 1. Identify the peak pixels on G-generated S image.
Max
Min
Step 2. Peak pixels represent the center of excitation → P1.
Center pixel
Outer pixel with radius of 1, 2, 3, 4, 5, >6 pixel
Normalized occurrence
Time (ns)
Step 3. Assign P2 on the other end of STED trajectory with the certain threshold. (in this work, we use MAX0.9)
Linear decomposition & image reconstruction
STED-flimGANE image
Determination of threshold
MAX is given by the maximum value in the phasor plots. Based on beads examples, PMAX0.9 is located closest to the periphery of excitation.
Beads example
PMAX
PMAX0.9
PMAX0.8
PMAX0.7
PMAX0.6
PMAX0.5
PMAX0.4
PMAX0.3

Figure 4.3 | Schematic of the STED-*flimGANE* method. (a) A generator (*G*) transformed the acquired decay curve into a high-photon-count decay. It comprised a concatenation layer, two CNN blocks, each of which comprised one convolutional layer followed by an average pooling layer. The CNN section was followed by a flatten layer. Then a multi-task layer converted data into virtual lifetime parameters, followed by two fully connected layers. Skip connection was used to pass data between layers of the same level. Discriminator (*D*) consisted of four fully connected layers. (b) G_{output} from the well-trained generator was converted into phasor domain for automatic assignment of $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$. First, peak pixels on *G*-generated S image were identified. Second, their corresponding decay histograms and the neighboring pixels were extracted to ensure the selected peaks represented the center of excitation. The phasor coordinates of these center pixels were determined to be $\boldsymbol{P_1}$. Third, $\boldsymbol{P_2}$ was assigned on the other end of the STED trajectory with a certain threshold (e.g., $MAX^{0.9}$, where MAX represented the maximum occurrence on the phasor plot). The representative beads example showed that the location with the threshold of $MAX^{0.9}$ matched well with the fluorescence decays at the periphery of the excitation. Finally, the STED-*flimGANE* image was obtained by photon separation with our automatic assignment of $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$.

4.3.2 Generative adversarial network structure and training

In this work, we trained a deep neural network using a Wasserstein generative adversarial network (WGAN)[196]. While GAN has shown great success with Jensen-Shannon (JS) divergence as the loss function, the difficulty of achieving Nash equilibrium and vanishing gradient make the GAN training challenging. Instead of JS divergence, Wasserstein distance as the loss function enables the training process to offer strong enough gradients to train the generator as compared with the original GAN. The loss functions of WGAN for the gradient update to the generator and discriminator were shown in Equation 4.5- 4.6, where z_i was the normalized low-photon-count fluorescence decay histogram, and x_i represented the normalized ground truth. $G(z)$ was the normalized ground-truth mimicking histogram and $D(x)$ was the probability that x came from the ground-truth decay histogram. Additionally, $f(x)$ was a 1-Lipschitz function that met the following formula: $|f(x_1) - f(x_2)| \leq |x_1 - x_2|$. In the WGAN framework, the function f to calculate the Wasserstein distance was approximated by the discriminator with the weights determined

by the clipping parameters. Typically, a high value of the discriminator output was regarded as the ground-truth data, while a lower value meant the low-photon-count histogram.

Another issue in the GAN framework was the mode collapse, where the generator could only produce a single output type. We added two additional loss functions to the G_{loss} to avoid the mode collapse issue. The first one was the mean squared error between the $G(z_i)$ and the x_i in the time domain. The second additional loss was the mean squared error between the $G(z_i)$ and the x_i in the frequency domain. This additional information in the training process forced the generator model to learn the variability in the training dataset, thus stabilizing the training schedule and leading the GAN model to converge.

The model architecture for the generative model consisted of convolutional neural network layers, a multi-task layer with three nodes, a fully-connected decoding layer, and a residual connection (**Figure 4.3**). There were two convolutional layers (each had a rectified linear unit activation function, $ReLU(x)$) and pooling layers (with average operation with the pool size of 2) followed by a flatten layer.

$$ReLU(x) = x^+ = \max(0, x) \tag{4.7}$$

The output of the flatten layer was then fed into a multi-task neural network with hard parameter sharing, transforming the high-dimensional output into three tasks. The last layer was the multilayer perceptron with the activation functions as tanh () to force the output range to lie within -1 and 1. Since this last layer mapped the 3 tasks into 256 channels of the fluorescence decay histogram, we termed the layer a decoding layer. Finally, instead of learning a direct mapping from the normalized low-photon-count fluorescence decay histogram to the ground-truth fluorescence decay histogram, we reframed the process with

the residual learning framework by introducing a residual connection between the normalized low-photon-count decay histogram and the model's output.

The model architecture for the discriminative model consisted of four densely connected neural network layers with 128, 32, 4, and 1 nodes (**Figure 4.3**). All the layers except the last one had a sigmoid activation, $S(x)$, forcing the output to lie within the range from 0 to 1, defined as follows,

$$S(x) = \frac{1}{1+e^{-x}} \tag{4.8}$$

In contrast, the last layer had the linear activation function to output the scores corresponding to the fluorescence decay histogram fed into *D*.

In this work, we focused on the training of the generative model. In the generative model training stage, the total iteration was 2,000. We randomly selected ~10% of training samples from the dataset pool within each iteration. The discriminative model was updated five times while the generative model was kept untrainable, then the generative model was updated once while keeping the other one untrainable. The Glorot uniform initializer randomly initialized the generative and discriminative models. The generative model was optimized using Adam optimizer with a learning rate of 1×10^{-4}. The discriminative model was optimized using RMSprop optimizer with a starting learning rate of 5×10^{-5}. The final generative and discriminative models for each application in this work were selected at around the 700^{th} iteration, which took ~2h to train the model. Training without the discriminative loss and predictive cost could result in over-smoothed images, as the generative model optimized only a specific group of statistical metrics. Therefore, it was imperative to incorporate a discriminator to train the generative model.

4.3.3 The separation of photons by a lifetime tuning (SPLIT)

As shown in **Figure 4.1c**, when the distance of the fluorophore to the center of excitation increased, the depletion effect kicked in and increased the decay rate. The resulting phasor points thus lay along a chord of the G-S semicircle, moving from the top of the semicircle to the lower part of the semicircle, termed a STED trajectory (**Figure 4.3**). The two ends of the STED trajectory were the $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$ components, which represent central and peripheral emissions, respectively. Typically, the positions of $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$ were determined manually. To improve accuracy and reproducibility, our STED-*flimGANE* adopted an algorithm to assign $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$ positions in the denoised phasor plot. Previously, the SPLIT method relied on the manual assignment of the locations of two phasors, $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$, which corresponded to the center and the periphery of the excitation spot, respectively. With our high-quality phasor plot, we proposed an automatic procedure to assign the locations with our high-quality phasor plot and then retrieve the desired signals *via* photon separation. First, as the center of the excitation had the largest S coordinate according to the STED trajectory (**Figures 4.3-4.4**), we identified multiple peak pixels on the *G*-generated S image *via* a local maximum filter. Then the exclusion criteria were applied to exclude the identified peaks if any of the following situations has been met: (1) the distance between peaks was less than 40 pixels, (2) the distance to the border of the image was less than 15 pixels. Second, the decay histograms at these selected peaks and the neighboring pixels were extracted to ensure the identified peaks represented the center of excitation. When the distance of the fluorophore to the center of excitation increased, the depletion effect kicked in and increased the decay rate. After summing all the decay histograms at peak pixels, followed by the normalization, we converted the normalized decay into a phasor. The location of this phasor was determined to be $\boldsymbol{P}_1$. Third, according to the STED trajectory,

the phasor of the periphery of the excitation spot was located at the other end of the trajectory. Therefore, $\boldsymbol{P}_2$ was assigned to the other end of the STED trajectory with a certain threshold (e.g., $MAX^{0.9}$, where MAX represents the maximum occurrence on the phasor plot; **Figure 4.3**). When evaluating the fluorescence decays at the pixels away from the identified peaks, the phasor location with the threshold of $MAX^{0.9}$ matched well with the peripheral region of the excitation spot. Finally, the STED-*flimGANE* image was obtained by performing the linear decomposition algorithm on the *G*-generated phasors smoothed with a Gaussian filter ($\sigma = 1$).

Assume that a pixel (with the total number of collected photons N) had a phasor located on $\boldsymbol{P}$, equal to $\boldsymbol{P} = f_1\boldsymbol{P}_1 + f_2\boldsymbol{P}_2$, where f_1 and f_2 represented the fractional components of the detected photons. Given $\boldsymbol{f} = [f_1; f_2]$ and $M = [\boldsymbol{P}_1^T, \boldsymbol{P}_2^T]$, we could write the linear system in the matrix form $\boldsymbol{P} = M\boldsymbol{f}$. The solution $\boldsymbol{f} = M^{-1}\boldsymbol{P}$ enabled us to separate the photons emitted by molecules in the center of the excitation ($= f_1N$). Eventually, the resulting image with improved resolution could be obtained by iterating this process for each pixel in the image. Here we first validated our model using a synthetic dataset (**Figures 4.4-4.6**), followed by applications to real samples (fluorescent beads in **Figure 4.7** and biological samples in **Figures 4.8-4.13**) under various depletion conditions.

4.4 RESULTS

4.4.1 Synthetic data and the STED-*flimGANE* model training

The STED-*flimGANE* model was first trained using a Monte Carlo simulation dataset (**Figure 4.4**). A Python program was employed to simulate the photon collection process in the counting device with either 64- (for fluorescent beads) or 256-time bins (for stained

NPCs), followed by the probability mass function calculation based on the convolution of an experimentally obtained instrument response function (IRF) and a theoretical fluorescence decay in the central and peripheral regions, generating the following piecewise function:

$$I_i(t) \propto \begin{cases} e^{-t/\tau_{i,STED}} & \text{if } 0 \le t < T_{STED} \\ e^{-T_{STED}/\tau_{i,STED}} e^{-(t-T_{STED})/\tau_{i,fl}} & \text{if } t \ge T_{STED} \end{cases} \quad (4.9)$$

Where I_i represented the intensity at time t and location i, $\tau_{i,fl}$ represented the natural fluorescence lifetime of a fluorophore at location i, $\tau_{i,STED}$ was the shortened fluorescence lifetime under the STED-beam depletion at location i, and T_{STED} was the STED-beam pulse duration (600 ps in this work). Location i = 1 represented the center while $\boldsymbol{i} > 1$ indicated the ($\boldsymbol{i}$ *-1*) pixels toward the periphery. Depending on the fluorophores that users wanted to image (0.5-3.5 ns for natural fluorescence lifetime of most fluorophores) and the available photon budgets (50-5,000 photon counts per pixel), 300 normalized ground truths and 150,000 degraded decays were generated for training *G* and *D*. The adversarial network training was completed in 2.3 hours.

The normalized degraded decay was transformed into the normalized "ground-truth mimicking" histogram, termed G_{output}, within 0.08 ms per pixel. After the training process, G_{output} became indistinguishable by *D* from the ground truth dataset (**Figure 4.1c** and **Figure 4.5**). Two validation steps were employed to evaluate the accuracy and reliability of the STED-*flimGANE* imaging. First, success in training was reflected by the quick drop of training loss (mean-squared error, MSE; **Figure 4.5a**), followed by the convergence to 0.01 count^2 after training iterations. Second, Sarmento *et al.* evaluated the SNR of a phasor

plot based on the spread along the direction of the phasor elongation. The phasor elongation along the direction of the STED trajectory, Δ_d, was mostly affected by STED-induced temporal dynamics modulation (**Figure 4.6**). The higher depletion power, the larger Δ_d. The spread along the direction perpendicular to the STED trajectory, Δ_N, represented the SNR, where higher SNR led to smaller Δ_N. We first identified the STED trajectory based on the given phasor plot. We then calculated Δ_N for each pixel. Ultimately, the average Δ_N served as a metric to compare the SNR of the phasor plot among different methods. Adopting the same strategy, we assessed the denoising performance of our STED-*flimGANE* algorithm[197]. When tested on an unseen dataset, STED-*flimGANE* showed 2.9-, 2.2-, and 2-fold improvement in SNR of the denoised phasor plots as compared to the plots before denoising, for the ultra-low- (< 100 photons per pixel), low- (100-200 photons per pixel), and medium-photon-count (200-300 photons per pixel) conditions, respectively, and a positive correlation (0.95; Pearson correlation coefficient) between the model performance and the SNR (**Figure 4.6**). These metrics verified that the STED-*flimGANE* is a reliable model for STED imaging analysis and reconstruction.

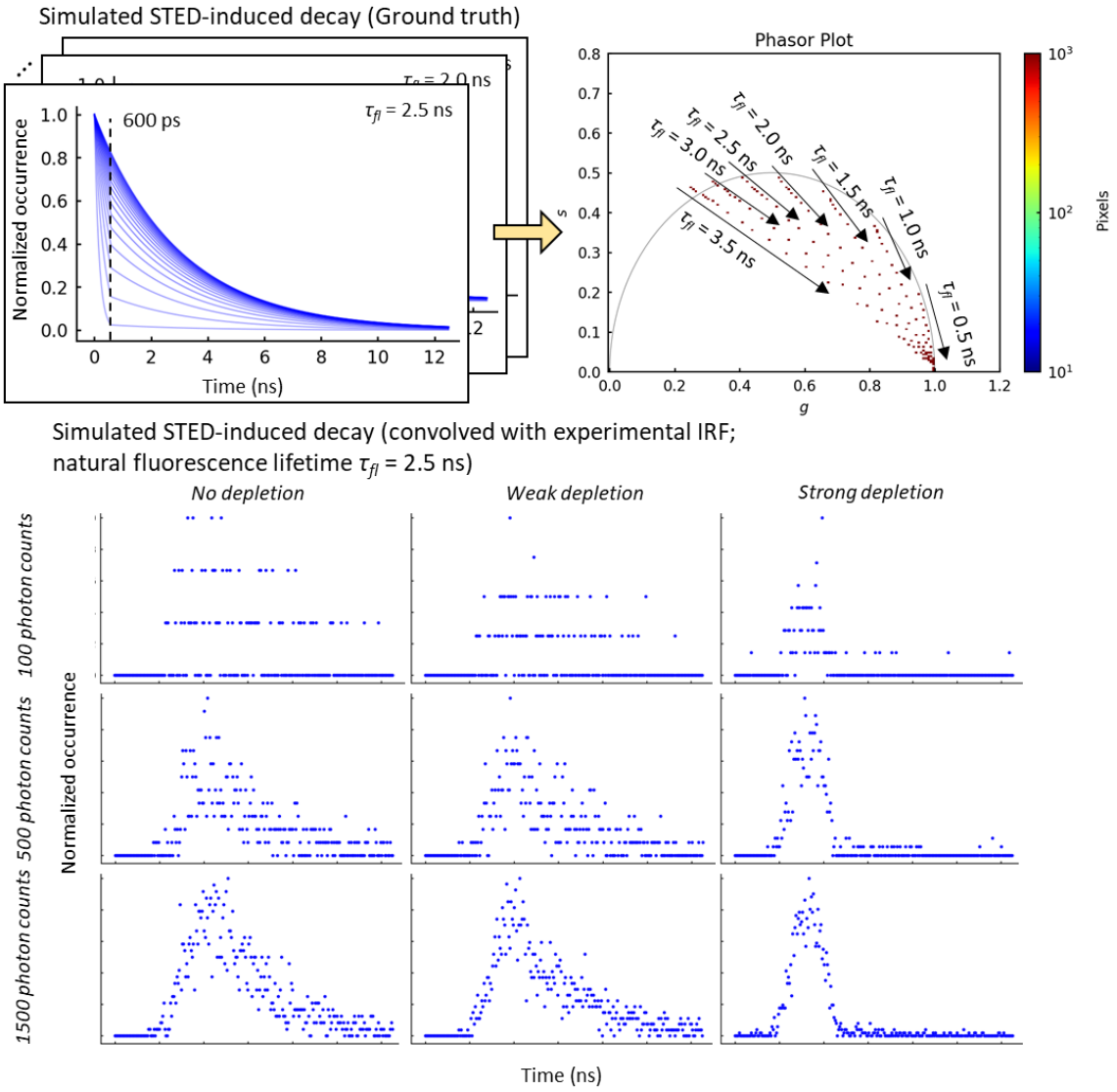

Figure 4.4 | **Simulated dataset for STED-*flimGANE* network training.** Due to the pulsed STED laser (600 ps), the originally exponential decay was modulated by the STED-induced decay rate during the depletion process. With different simulated natural fluorescence lifetimes and STED-induced decay rates, we observed the corresponding STED trajectory on the phasor plot. Then we performed a Monte Carlo simulation on the instrumental response function-convolved STED-induced decay to generate a large dataset for our model training. A clear and distinct pattern between disparate depletion levels was observed from nine representative simulated STED-induced decays.

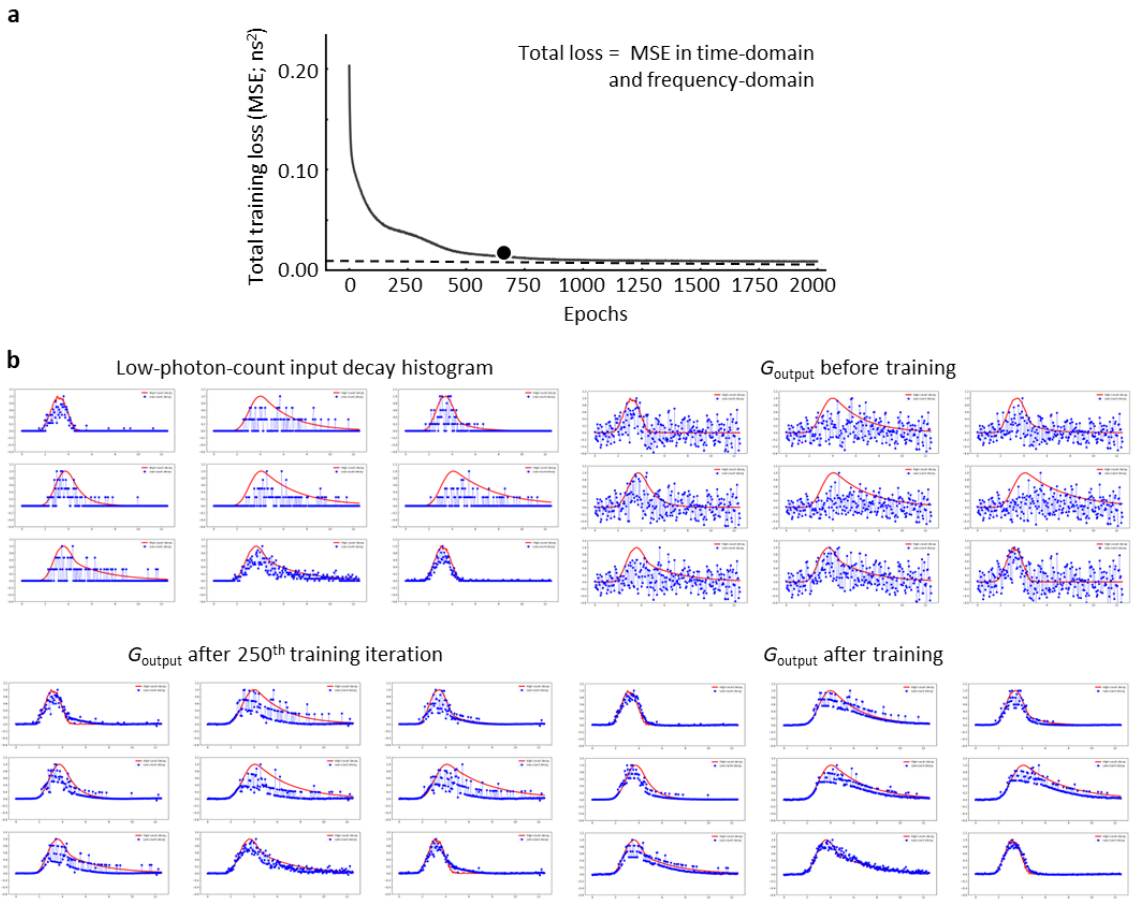

Figure 4.5 | Generator (*G*) can transform a low-photon-count decay histogram into a high-photon-count one. (a) Training loss of STED-*flimGANE* over iterations. The mean squared error between G_{output} and the ground-truth fluorescence decay histograms was visualized over training iterations. We could observe a rapid drop at the beginning of the training, and the loss converged to a certain value (~0.01 ns^2). The model was selected after 700 training iterations (indicated by the back dot). (b) Given the GAN framework, the normalized low-photon-count decay histogram was transformed into the normalized ground-truth mimicking histogram. At the beginning of the training stage, the output from the *G* was chaotic. The generator-inferred fluorescence decay histogram gradually matched the ground truth during the training process.

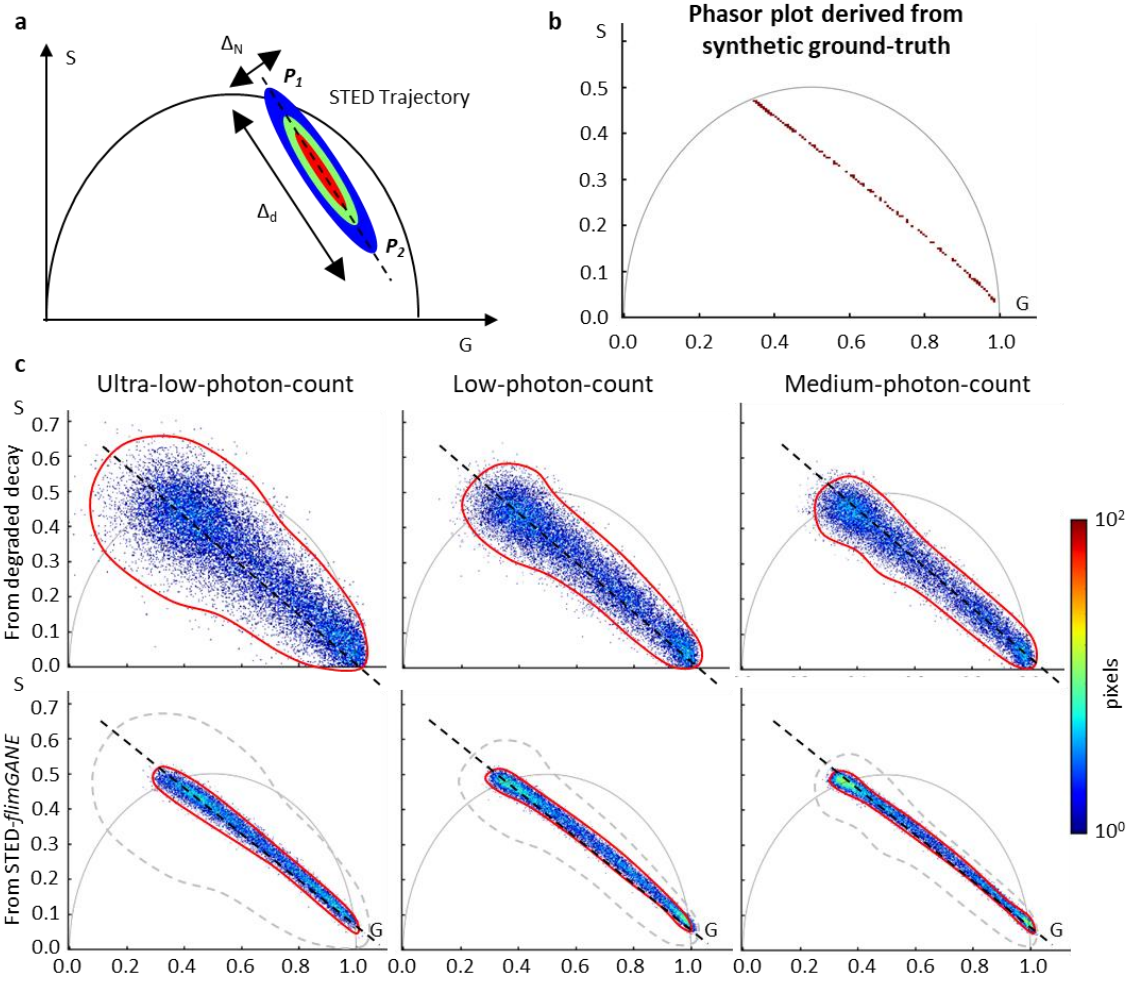

Figure 4.6 | Characterization of SNR improvement on phasor plots. (a) Schematic of SNR characterization in phasor plot. (b) The phasor plot was obtained from the synthetic ground truths for evaluation. (c) The phasor analysis of the synthetic degraded fluorescence decay histogram and the STED-*flimGANE*'s G_{output} under ultra-low- (30-100 photons per pixel), low- (100-200 photons per pixel), and medium-photon-count (200-300 photons per pixel) conditions. Scatter error was significantly reduced after STED-*flimGANE*.

4.4.2 Validation of STED-*flimGANE* using 60 nm fluorescent beads

When imaging fluorescent beads, the power of the 775 nm depletion beam was fixed at 120 mW. Under such a strong STED-beam power, the acquired average photon count per pixel reduced to 41 photons (**Figure 4.7**). In the confocal imaging (equivalent to STED imaging without the depletion beam), the phasors mapped from all pixels scattered around the long lifetime ($\boldsymbol{P_1}$) components in the semicircle. In contrast, when the STED beam was employed, the phasor distribution elongated toward the short lifetime ($\boldsymbol{P_2}$) component (**Figure 4.7a**). The formation of a clear STED trajectory clearly facilitated photon separation in the pSTED-SPLIT and STED-*flimGANE* schemes, thus improving the spatial resolution (**Figure 4.7b-c**). By fitting the line profiles with a multi-peak Gaussian function (**Figure 4.7d**), we estimated the full-width-at-half-maximum (FWHM) of these fluorescent beads ($n > 100$) to be 64 ± 5 nm when using STED-*flimGANE*, which was 1.17-fold and 1.09-fold improvement as compared to the conventional pSTED imaging (75 ± 5 nm) and the pSTED-SPLIT imaging (70 ± 6 nm), respectively.

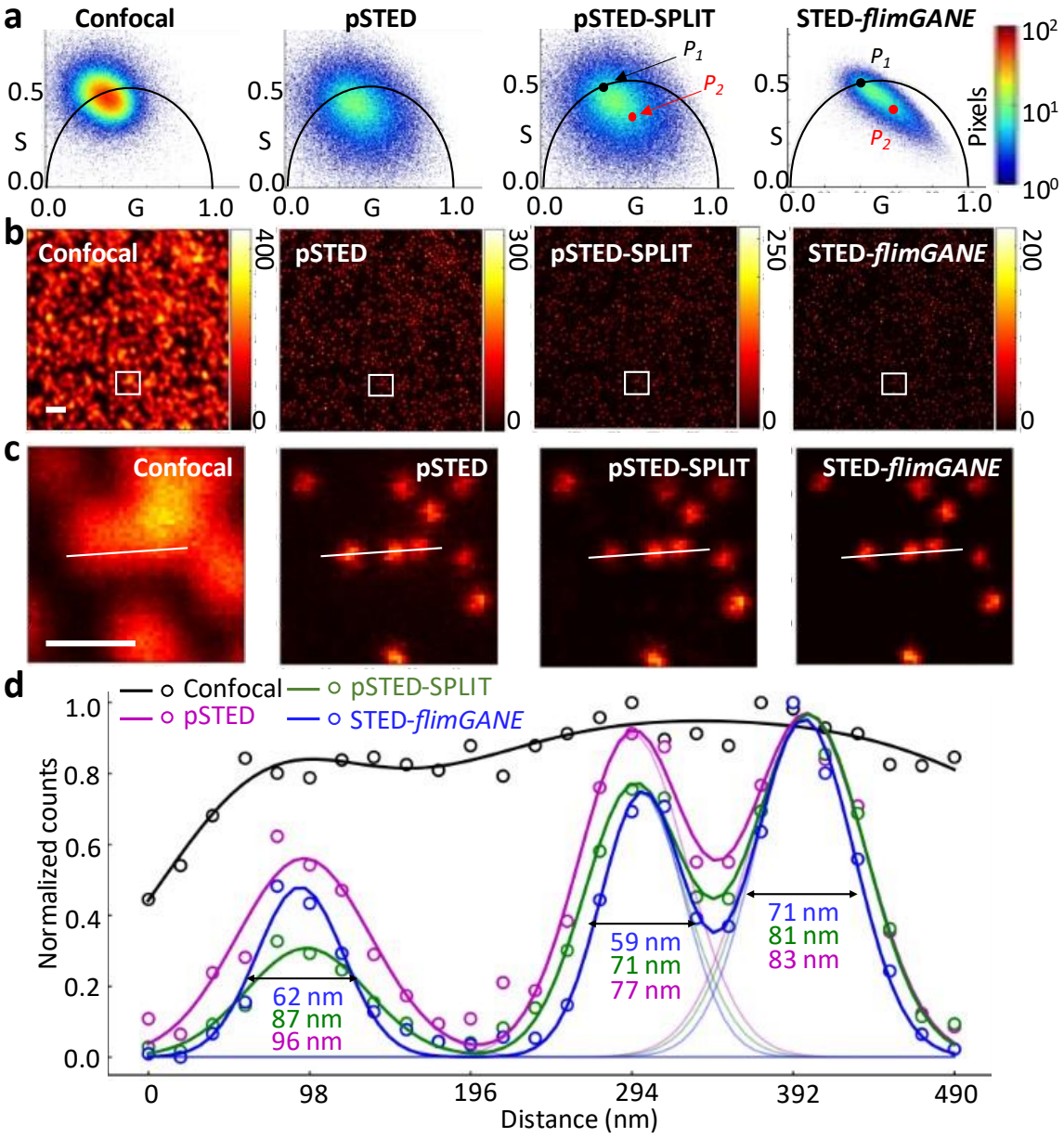
a
Confocal
pSTED
pSTED-SPLIT
STED-flimGANE
P1
P2
S
G
0.0
0.5
1.0
Pixels
10^2
10^1
10^0
b
400
300
250
200
0
c
d
Confocal
pSTED
pSTED-SPLIT
STED-flimGANE
Normalized counts
Distance (nm)
62 nm
87 nm
96 nm
59 nm
71 nm
77 nm
71 nm
81 nm
83 nm

Figure 4.7 | Fluorescent bead imaging results using confocal, pSTED, sSTED-SPLIT and STED-*flimGANE* microscopy. (a-b) The phasor representations (a) and the intensity images (b) of the confocal, pSTED, pSTED-SPLIT, and STED-*flimGANE* images for 60 nm fluorescent beads. The STED-beam power (P_{STED}) was fixed at 120 mW. Scale bar is 1 μm. (c) Zoom-in views of the box regions in (b) show that the STED-*flimGANE* achieves the highest spatial resolution. Scale bar is 400 nm. (d) Line profiles of the confocal, pSTED, pSTED-SPLIT, and STED-*flimGANE* images in (c). A two-component Gaussian fitting was utilized to obtain the FWHM of the fluorescent beads. The average resolutions are 75 ± 5, 70 ± 6 and 64 ± 5 nm (n = 125) for pSTED, pSTED-SPLIT and STED-*flimGANE* microscopy, respectively.

4.4.3 Performance of STED-*flimGANE* in imaging biological samples

When imaging nuclear pore complexes (NPCs, with an average size of 60 nm) on fixed COS-7 cells, STAR-635p was used as the STED dye that stained nucleoporin Nup 153 through a primary and a secondary antibody (**Figure 4.8a-b**). As expected, with a more effective algorithm to separate unwanted photons, STED-*flimGANE* outperformed pSTED, and pSTED-SPLIT in resolving NPCs under low-photon-count (133 average photons per pixel) to ultra-low-count (45 average photons per pixel) conditions (**Figure 4.8c**). Both pSTED and pSTED-SPLIT schemes failed to reconstruct a clear NPC image at a low STED-beam power (20 mW) and high average photon count (133 per pixel), indicating a reduced resolution due to the early fluorescence background. However, under this low depletion condition, STED-*flimGANE* still well resolved the adjacent NPCs (**Figure 4.8b**). The line profile from the STED-*flimGANE* image exhibited a clear trough between the adjacent NPCs (**Figure 4.8c**). When the distance between adjacent NPCs is down to 140 nm, only STED-*flimGANE* can better differentiate these NPCs (**Figure 4.9**). In contrast, pSTED and pSTED-SPLIT could resolve the same pair of NPC only when the STED-beam power was higher (≥60 mW). Although the resolution of the STED-*flimGANE* slightly improved under high STED-beam power (**Figure 4.8c**), its spatial resolution dependence on the STED-beam power was much less than that of pSTED and pSTED-SPLIT.

Specifically, the STED-*flimGANE* showed 1.45-fold improvement in spatial resolution (77 nm) as compared to that of pSTED-SPLIT (112 nm) under the STED-beam power of 60 mW, indicating that STED-*flimGANE* could further enhance the spatial resolution when STED-beam power was kept low.

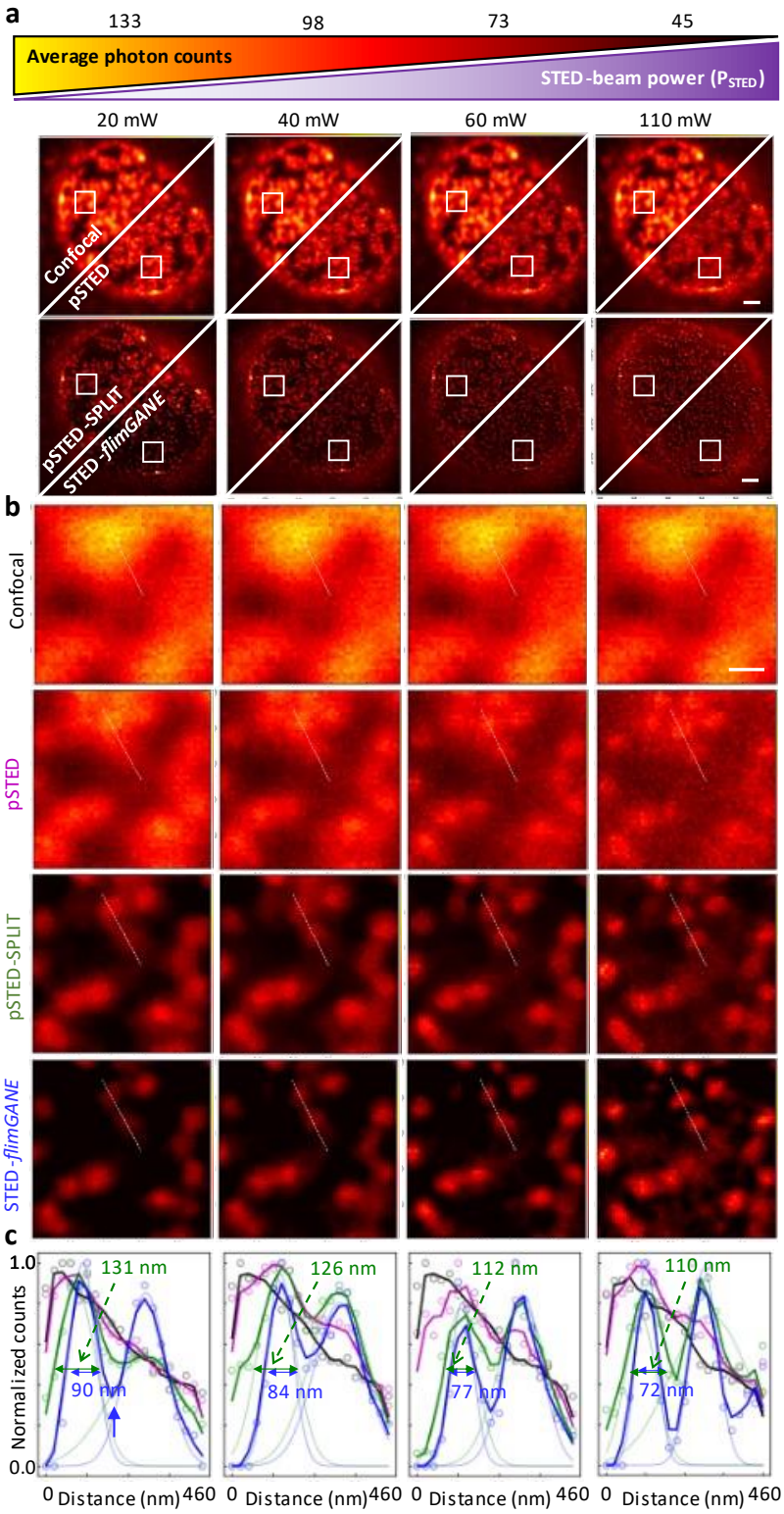
a
133
98
73
45
Average photon counts
STED-beam power (P_STED)
20 mW
40 mW
60 mW
110 mW
Confocal
pSTED
pSTED-SPLIT
STED-flimGANE
b
Confocal
pSTED
pSTED-SPLIT
STED-flimGANE
c
1.0
0.0
Normalized counts
131 nm
90 nm
126 nm
84 nm
112 nm
77 nm
110 nm
72 nm
0
Distance (nm)
460

Figure 4.8 | Nuclear pore complex (NPC) imaging results on COS-7 cells under low-to-medium depletion power conditions. (a) The intensity images of NPCs using the confocal, pSTED, pSTED-SPLIT, and STED-*flimGANE* microscopy under various depletion powers. COS-7 cells were stained with a primary antibody against Nup 153 and a STAR-635P-labeled secondary antibody. Scale bars are 1 μm. (b) Zoom-in views of the box regions in (a) show that only the STED-*flimGANE* could well resolve the two adjacent nuclear pore complexes (indicated by white dashed lines) at P_{STED} = 20 mW. Scale bar is 200 nm. (c) Line profiles of the confocal, pSTED, pSTED-SPLIT, and STED-*flimGANE* images in (b). The average resolutions are 116 ± 1 and 101 ± 1 nm (n = 100) for pSTED-SPLIT and STED-*flimGANE* microscopy, respectively, at PSTED = 20 mW.

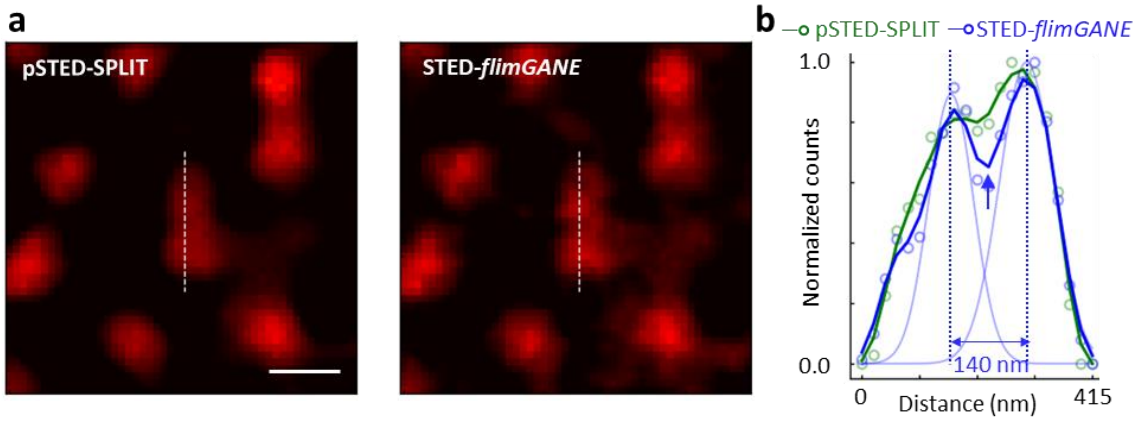

Figure 4.9 | The STED-*flimGANE* enables differentiation of the nuclear pore complex. (**a**) The zoom-in views of the intensity images of NPCs using the pSTED-SPLIT and STED-*flimGANE* microscopy at P_{STED} = 20 mW. Scale bar is 200 nm. (**b**) Line profiles of the confocal, pSTED, pSTED-SPLIT, and STED-*flimGANE* images in (a). Blue arrow indicates a clear trough between the adjacent nuclear pore complexes.

4.4.4 Resolution comparison under extreme depletion power conditions

Here we compared the spatial resolution of the STED-*flimGANE* imaging with that of the confocal, pSTED, and pSTED-SPLIT imaging under extremely low (P_{STED} = 14 mW, 120 average photons per pixel) and extremely high (P_{STED} = 140 mW, 55 average photons per pixel) depletion power conditions (**Figure 4.10a-b**). As expected, the $\boldsymbol{P_2}$ component moved closer to (1, 0) under the high depletion power (**Figures 4.11-4.12**). Although both pSTED-

SPLIT and STED-*flimGANE* could differentiate closely packed NPCs under the high depletion power, some closely packed NPCs were not resolved in pSTED-SPLIT imaging under the low depletion power (**Figure 4.10c-d and Figure 4.13**). In contrast, STED-*flimGANE* well resolved these closely packed NPCs even under a low depletion power. In addition, STED-*flimGANE* offered more background suppression in the surroundings of the NPCs and the outer regions of the cells (**Figure 4.10c**).

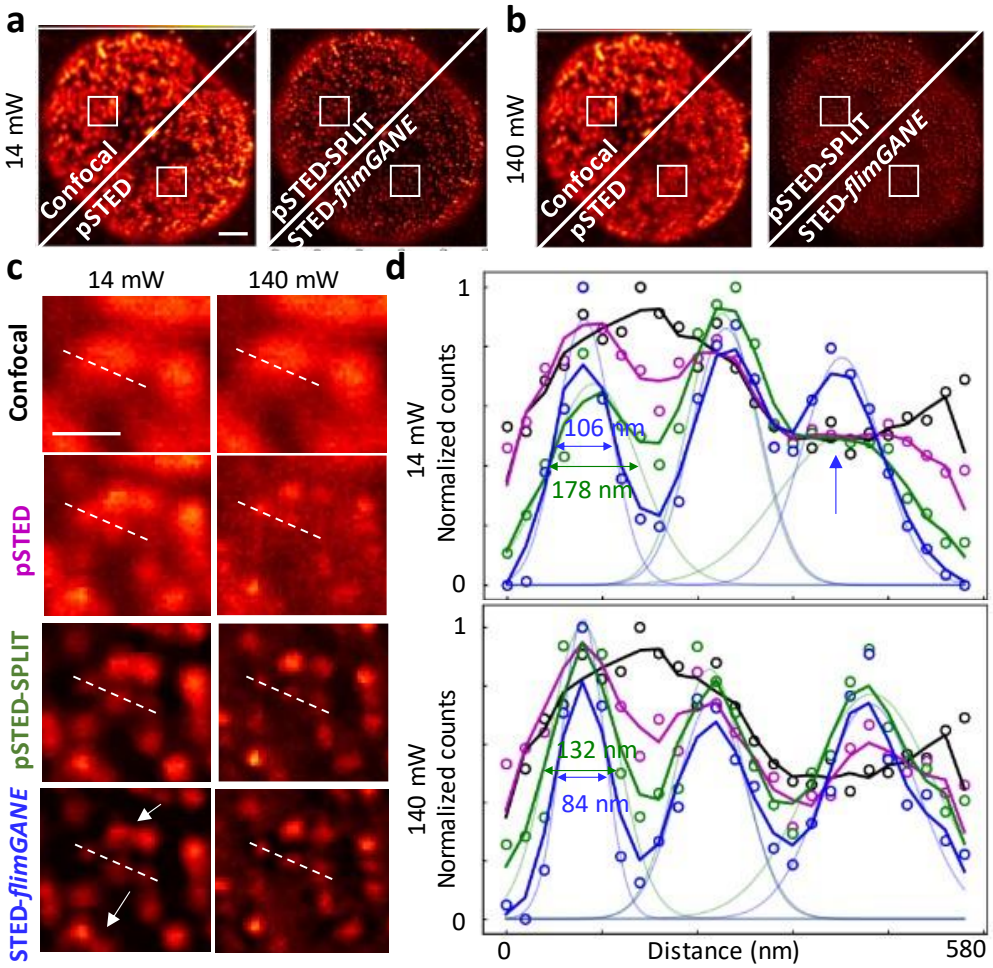

Figure 4.10 | Nuclear pore complex (NPC) imaging results on COS-7 cells at extreme depletion power conditions. (a) The intensity images of the confocal, pSTED, pSTED-SPLIT, and STED-*flimGANE* for the STED probe under extremely low depletion conditions (P_{STED} = 14 mW). Scale bars, 1 μm. (b) The same as (a) but under extremely high depletion conditions (P_{STED} = 140 mW). (c) Zoom-in views from the solid box region in (a, b) demonstrated that the STED-*flimGANE* achieved a similarly high spatial resolution using extreme depletion power. Obvious improvements were observed from extremely low to high depletion for both pSTED and pSTED-SPLIT. Scale bars, 400 nm. The arrows indicate that only STED-*flimGANE* was able to resolve the closely packed NPCs under extremely low depletion conditions. (d) Line profile of the confocal, pSTED, pSTED-SPLIT, and STED-*flimGANE* images in (c).

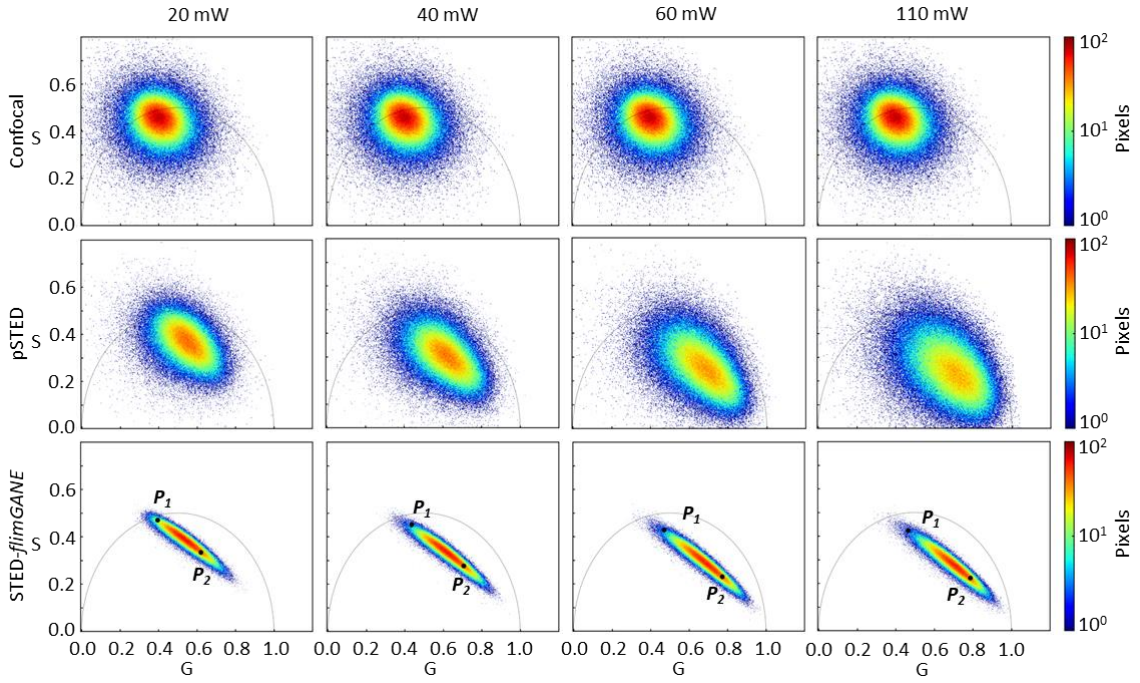

Figure 4.11 | Phasor plots for confocal, pSTED, and STED-*flimGANE* in Figure 4.8. The *G*-generated phasor plots, ***P₁*** and ***P₂*** assigned by our automatic approach allowed us to determine the photons emitted from the natural fluorophore in the STED experiments with disparate levels of STED power (20, 40, 60, and 110 mW) and the excitation power of 2.2 μW.

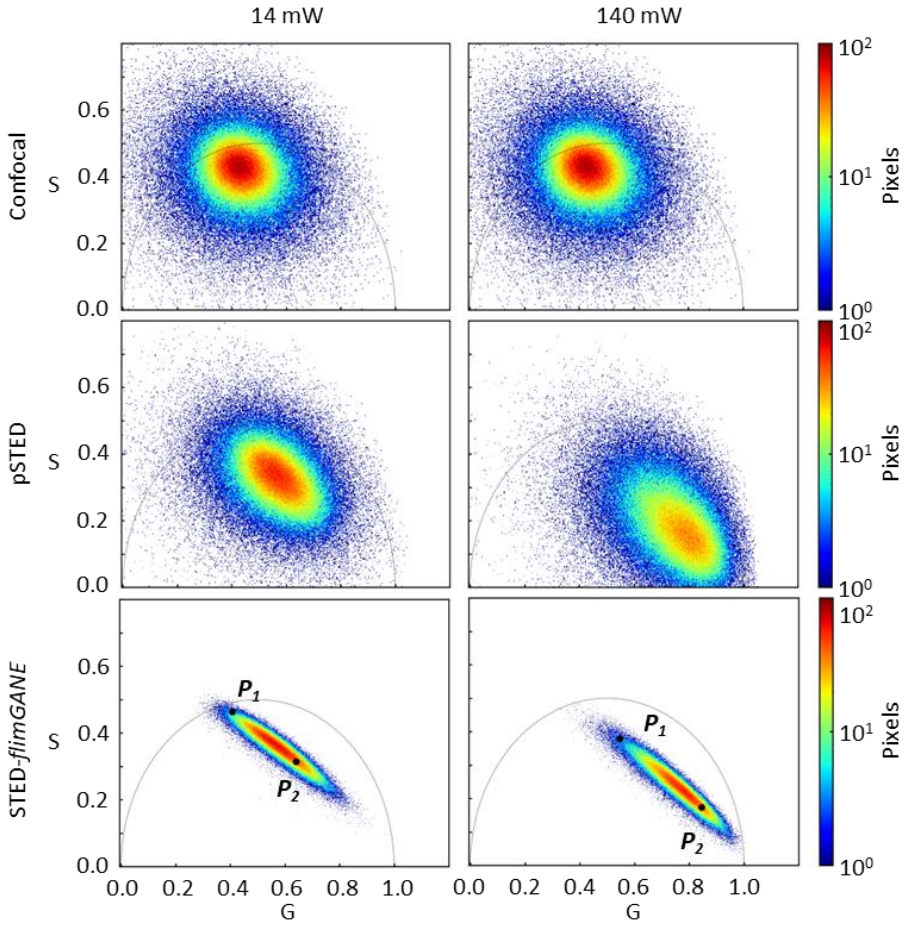

Figure 4.12 | Phasor plots for confocal, pSTED, and STED-*flimGANE* in Figure 4.10. The *G*-generated phasor plots, ***P*₁** and ***P*₂** assigned by our automatic approach allowed us to determine the photons emitted from the natural fluorophore in the STED experiments with disparate levels of STED power (14 and 140 mW) and the excitation power of 2.2 μW.

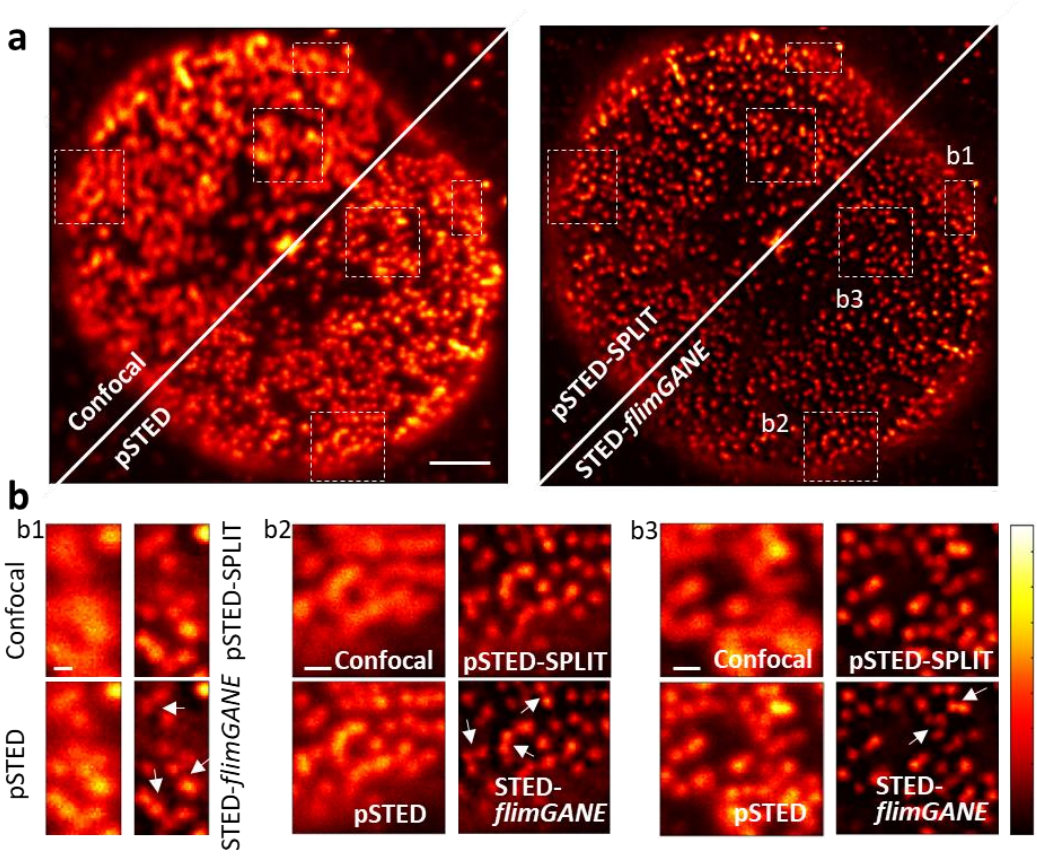

Figure 4.13 | Nuclear pore complex (NPC) imaging results on COS-7 cells at extremely-low depletion power. (a) The intensity images of the confocal, pSTED, pSTED-SPLIT, and STED-*flimGANE* for the STED probe under extremely-low depletion conditions (14 mW). Scale bars, 1 μm. (b) Zoom-in views from the dashed box regions in (a) demonstrated that only the STED-*flimGANE* could discriminate adjacent nuclear pores. Scale bars, 200 nm for b1, 400 nm for b2 and b3.

4.4.5 Evaluation of the spatial resolution dependence on STED power

Using as the resolution estimate, FWHM of the point spread function was determined by fitting the intensity profiles of 100 randomly selected NPCs with a Gaussian function (**Figure 4.14**), where the FWHM was equal to 2.35σ. Under all STED-beam powers used in this investigation, STED-*flimGANE* not only achieved super resolution imaging (i.e., FWHM < 200 nm, **Figure 4.14a**), but showed 1.13-fold improvement in spatial resolution

(on average) over the traditional pSTED-SPLIT imaging. When the same resolution was maintained, STED-*flimGANE* only required one-fourth the depletion power as compared to pSTED-SPLIT (30 mW vs. 120 mW). While the resolution of both methods could be enhanced by increasing the STED-beam power, STED-*flimGANE* clearly showed less dependence on the depletion power (the slope of the linear regression fit was only half of that of the pSTED-SPLIT fit). In addition, the FWHM distributions also indicated STED-*flimGANE* had less dependence on the STED-beam power (**Figure 4.14b**). All these results suggested that STED-*flimGANE* is a better form of the SPLIT method and is more suitable for imaging live samples (although here we only tested STED-*flimGANE* on fixed cells).

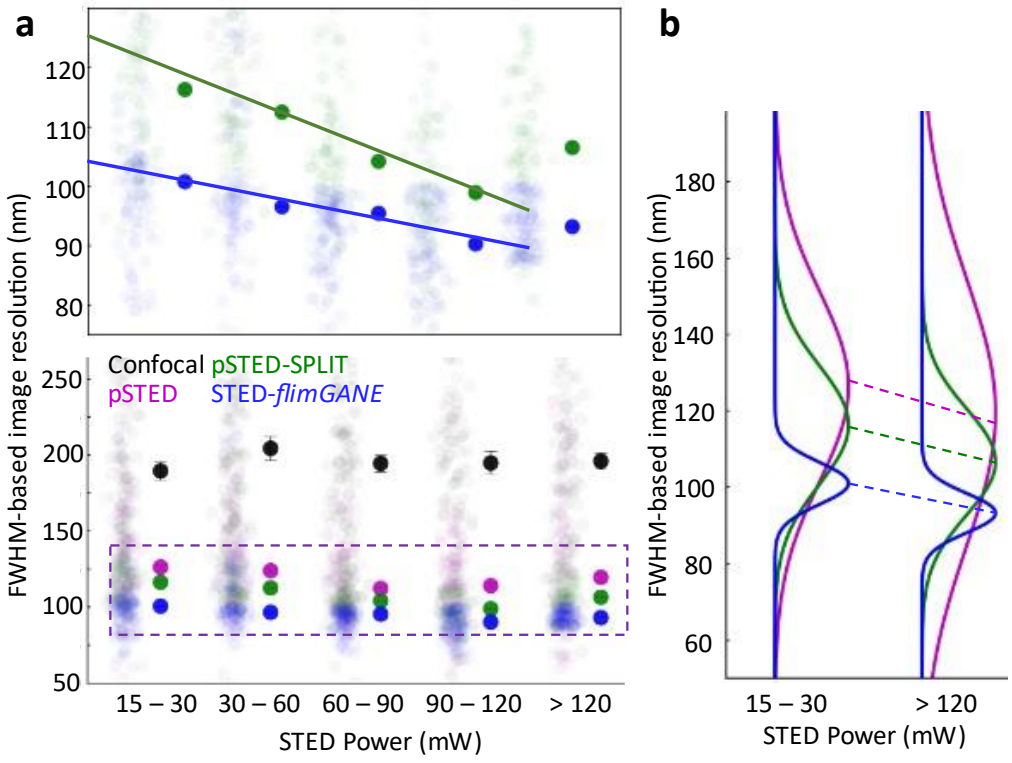

Figure 4.14 | The STED-*flimGANE* provided reliable super-resolution images across various STED depletion conditions. (a) Qualification of the resolution in the confocal, pSTED, pSTED-SPLIT, and STED-*flimGANE* images. The optical resolution is determined by the FWHM of the intensity profiles of the nuclear pores. STED-*flimGANE* exhibited 50% less dependency of the image resolution on the STED power. The transparent points in the raincloud plots represent individual FWHM values for 100 randomly selected NPCs while the darker points represent the average FWHM values from Gaussian fitting. (b) The fitted Gaussian distributions of pSTED, pSTED-SPLIT, and STED-*flimGANE* on the FWHMs under two P_{STED} conditions: 15-30 mW and >120 mW depletion power.

4.5 DISCUSSION

In this work, we have demonstrated that the STED-*flimGANE* can achieve 77 nm lateral resolution in imaging nuclear pore complexes on fixed COS-7 cells, under the conditions of 73 photons per pixel and 60 mW STED-beam power. These results indicated that the STED-*flimGANE* is a robust, fit-free and user-friendly method for enhancing the spatial resolution with minimum dependence on the STED-beam power, even in the presence of uncorrelated background and shot noise. The denoising capability of the STED-*flimGANE* on the noisy fluorescence decays *via* GAN is the key to providing robust phasor representations for downstream photon separation (**Figure 4.5**). While artificial neural network or convolutional neural network have been employed to achieve rapid fluorescence lifetime analysis, the performance may become unreliable when analyzing the low-photon-count data. Although GAN-based approaches have been used to transform the confocal images to match the resolution of STED images[198,199], to the best of our knowledge, this is the first report that the GAN is applied to improve the spatial resolution of STED imaging based on the fluorescence temporal dynamics.

It is worth noting that STED-*flimGANE* maintains good spatial resolution even when photon counts are low. Insufficient photons result in scattered phasor points in the phasor plot, leading to so-called scatter error[200]. Since the scatter error scales inversely with the square root of the total photon counts per pixel, filtering is a typical strategy to denoise phasor plots under photon-starved conditions. For instance, a median filter is commonly used for phasor denoising[201]; however, high-spatial-frequency components such as the edge of features are often diminished. To preserve these high-frequency components, STED-*flimGANE* employs Wasserstein distance as the loss function, which provides a smoother gradient for GAN network training under all conditions. As a result, the well-trained generator in STED-*flimGANE* seamlessly converts low-photon-count decays into highly realistic high-photon-count ones, leading to a denoised phasor plot with much less scatter error[193]. Although other strategies such as the complex wavelet transform methods can also preserve fine structures and concurrently clean up noisy images[202,203], they require additional steps, such as selection and optimization of the basis functions, before the transformation.

A few reports demonstrated low-power STED imaging through modulating the STED beam. One example is the adaptive STED-illumination strategy DyMIN (Dynamic intensity MINimum) that provides the imaging of NPCs at a resolution of 73 nm with 30 and 344 mW STED-beam powers for two cycles. Similarly, modulation-enhanced STED (M-STED) also achieves 87 nm resolution of NPCs by analyzing the STED trajectory under various STED-beam powers[197,204]. However, changing the STED-beam power not only increases the data acquisition time but also complicates the analysis process. In contrast, our STED-*flimGANE* only relies on post-processing, which provides a comparable resolution without the need of spending additional time modulating the STED-

beam power. We emphasize that this method is simple to implement in the existing optical system such as modulated STED (mSTED) to facilitate the observation of the rapid process of the biological samples[205]. Although the network training time may be the bottleneck for STED-*flimGANE*, this issue can be overcome by using more advanced hardware (e.g., graphic processing unit or tensor processing unit) or transfer learning techniques that take advantage of the previously trained networks[206].

The STED-*flimGANE* is a versatile and efficient method for improving the resolution of STED microscopy, which can be easily adapted on a variety of standard and custom setups. Bayesian optimization provides an efficient means to select the optimal hyperparameters. The $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$ phasor components in the traditional SPLIT are either assigned manually or determined by the characteristic limits of the operative range. As a result, these phasor components can be skewed due to bias, or require recalibration once the sample or the excitation/depletion power is changed. In contrast, STED-*flimGANE* automatically assigns the $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$ phasor components without the need of recalibration, thus eliminating any user or experimental bias. Similar to the existing photon-separation approaches, we expect STED-*flimGANE* to be capable of improving the resolution in both CW and pulsed STED. While the implementation of CW-STED is more straightforward, less depletion efficiency and more susceptibility to the background noise limit its widespread use in various applications[177,181,207]. Here we demonstrated STED-*flimGANE* using pSTED, achieving up to 1.45-fold improvement in spatial resolution under the conditions of both 20 mW (90 nm vs 131 nm) and 60 mW (77 nm vs 112 nm) STED-beam power (**Figure 4.8c**). The higher STED-beam power depletes more photons from the periphery at the expense of SNR of the phasor plot. In contrast, the lower STED-beam power yields a better quality of phasor plot, but at the cost of the spatial resolution.

As live-cell imaging usually involves a compromise between image quality and cellular health, the labeling protocols and imaging conditions are required to be optimized according to the users' purpose[169].

Here we introduce a new method termed STED-*flimGANE* (combination of a deep learning-based phasor analysis, fluorescence lifetime imaging based on Generative Adversarial Network Estimation, and the separation of photons by lifetime tuning scheme) that can achieve enhanced STED imaging resolution under a low STED-beam power (< 30 mW) and photon-starved conditions (< 200 photons per pixel). Our STED-*flimGANE* method can rapidly generate robust phasor representations with less scatter error. Images reconstructed by these denoised phasor plots clearly show improved resolution and quality. Our work represents an important step towards low-power STED imaging for live-cell or live-tissue imaging applications, allowing users to achieve desired spatial resolution without significantly causing phototoxicity and photodamage to the samples. Our next task is to demonstrate STED-*flimGANE* in imaging organelles in live cells. As one of the key advantages of phasor analysis is the differentiation of multiple fluorophores with different lifetimes excited by a single excitation source[208,209], with recent development in fluorescence lifetime tuning strategies in fluorophores and protein tags[210–213], we envision that our STED-*flimGANE* can also be adapted for multiplexed STED imaging.

Bibliography

1. Buzsáki, G., Anastassiou, C. A. & Koch, C. The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience* **13**, 407–420 (2012).
2. Niedermeyer, E. & da Silva, F. L. *Electroencephalography: basic principles, clinical applications, and related fields*. (Lippincott Williams & Wilkins, 2005).
3. Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J. & Lounasmaa, O. V. Magnetoencephalography---theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Mod. Phys.* **65**, 413–497 (1993).
4. Ghane-Motlagh, B. & Sawan, M. Design and implementation challenges of microelectrode arrays: a review. *Materials Sciences and Applications* **4**, 483 (2013).
5. Strumwasser, F. Long-term recording from single neurons in brain of unrestrained mammals. *Science* **127**, 469–470 (1958).
6. Wise, K. D. *et al.* Microelectrodes, Microelectronics, and Implantable Neural Microsystems. *Proceedings of the IEEE* **96**, 1184–1202 (2008).
7. Campbell, P. K., Jones, K. E., Huber, R. J., Horch, K. W. & Normann, R. A. A silicon-based, three-dimensional neural interface: manufacturing processes for an intracortical electrode array. *IEEE Transactions on Biomedical Engineering* **38**, 758–768 (1991).
8. Sanchez, J. C., Alba, N., Nishida, T., Batich, C. & Carney, P. R. Structural modifications in chronic microwire electrodes for cortical neuroprosthetics: a case

study. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **14**, 217–221 (2006).

9. Westbrook, C. *Handbook of MRI technique*. (John Wiley & Sons, 2021).
10. Le Bihan, D. Looking into the functional architecture of the brain with diffusion MRI. *Nat Rev Neurosci* **4**, 469–480 (2003).
11. Bandettini, P. A. Twenty years of functional MRI: The science and the stories. *NeuroImage* **62**, 575–588 (2012).
12. Buzug, T. M. Computed Tomography. in *Springer Handbook of Medical Technology* (eds. Kramme, R., Hoffmann, K.-P. & Pozos, R. S.) 311–342 (Springer, 2011). doi:10.1007/978-3-540-74658-4_16.
13. Muehllehner, G. & Karp, J. S. Positron emission tomography. *Phys. Med. Biol.* **51**, R117 (2006).
14. Grienberger, C. & Konnerth, A. Imaging calcium in neurons. *Neuron* **73**, 862–885 (2012).
15. Sadakane, O. *et al.* Long-Term Two-Photon Calcium Imaging of Neuronal Populations with Subcellular Resolution in Adult Non-human Primates. *Cell Reports* **13**, 1989–1999 (2015).
16. Meir-Hasson, Y., Kinreich, S., Podlipsky, I., Hendler, T. & Intrator, N. An EEG Finger-Print of fMRI deep regional activation. *NeuroImage* **102**, 128–141 (2014).
17. Keynan, J. N. *et al.* Electrical fingerprint of the amygdala guides neurofeedback training for stress resilience. *Nat Hum Behav* **3**, 63–73 (2019).

18. Judenhofer, M. S. *et al.* Simultaneous PET-MRI: a new approach for functional and morphological imaging. *Nat Med* **14**, 459–465 (2008).
19. Sander, C. Y. & Hesse, S. News and views on in-vivo imaging of neurotransmission using PET and MRI. *Q J Nucl Med Mol Imaging* **61**, 414–428 (2017).
20. Wehrl, H. F., Sauter, A. W., Judenhofer, M. S. & Pichler, B. J. Combined PET/MR Imaging — Technology and Applications. *Technol Cancer Res Treat* **9**, 5–20 (2010).
21. Ibayashi, K. *et al.* Decoding Speech With Integrated Hybrid Signals Recorded From the Human Ventral Motor Cortex. *Frontiers in Neuroscience* **12**, (2018).
22. Moosmann, M. *et al.* Correlates of alpha rhythm in functional magnetic resonance imaging and near infrared spectroscopy. *NeuroImage* **20**, 145–158 (2003).
23. Golkowski, D. *et al.* Simultaneous EEG–PET–fMRI measurements in disorders of consciousness: an exploratory study on diagnosis and prognosis. *J Neurol* **264**, 1986–1995 (2017).
24. Wu, R., Yang, P.-F. & Chen, L. M. Correlated Disruption of Resting-State fMRI, LFP, and Spike Connectivity between Area 3b and S2 following Spinal Cord Injury in Monkeys. *J. Neurosci.* **37**, 11192 (2017).
25. Rajkumar, R. *et al.* Simultaneous trimodal PET-MR-EEG imaging: Do EEG caps generate artefacts in PET images? *PLOS ONE* **12**, e0184743 (2017).
26. Ghahramani, Z. & Hinton, G. E. Parameter estimation for linear dynamical systems. (1996).

27. Macke, J. H., Buesing, L. & Sahani, M. Estimating state and parameters in state space models of spike trains. in *Advanced State Space Methods for Neural and Clinical Data* (ed. Chen, Z.) 137–159 (Cambridge University Press, 2015). doi:10.1017/CBO9781139941433.007.

28. Gao, Y., Archer, E. W., Paninski, L. & Cunningham, J. P. Linear dynamical neural population models through nonlinear embeddings. in *Advances in Neural Information Processing Systems* vol. 29 (Curran Associates, Inc., 2016).

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

30. Byron, M. Y. *et al.* Gaussian-Process Factor Analysis for Low-Dimensional Single-Trial Analysis of Neural Population Activity. *Journal of Neurophysiology* **102**, 614–635 (2009).

31. Zhao, Y. & Park, I. M. Variational Latent Gaussian Process for Recovering Single-Trial Dynamics from Population Spike Trains. *Neural Computation* **29**, 1293–1316 (2017).

32. Lawrence, N. Gaussian Process Latent Variable Models for Visualisation of High Dimensional Data. in *Advances in Neural Information Processing Systems* vol. 16 (MIT Press, 2003).

33. Wu, A., Roy, N. A., Keeley, S. & Pillow, J. W. Gaussian process based nonlinear latent structure discovery in multivariate spike train data. in *Advances in Neural*

Information Processing Systems (eds. Guyon, I. et al.) 3496–3505 (Curran Associates, Inc., 2017).

34. Grech, R. *et al.* Review on solving the inverse problem in EEG source analysis. *Journal of NeuroEngineering and Rehabilitation* **5**, 25 (2008).
35. Michel, C. M. & Brunet, D. EEG Source Imaging: A Practical Review of the Analysis Steps. *Front. Neurol.* **10**, (2019).
36. Pascual-Marqui, R. D., Michel, C. M. & Lehmann, D. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *International Journal of Psychophysiology* **18**, 49–65 (1994).
37. de Peralta Menendez, R. G., Murray, M. M., Michel, C. M., Martuzzi, R. & Andino, S. L. G. Electrical neuroimaging based on biophysical constraints. *NeuroImage* **21**, 527–539 (2004).
38. Canolty, R. T., Ganguly, K. & Carmena, J. M. Task-Dependent Changes in Cross-Level Coupling between Single Neurons and Oscillatory Activity in Multiscale Networks. *PLOS Computational Biology* **8**, e1002809 (2012).
39. Wang, C., Pesaran, B. & Shanechi, M. M. Modeling multiscale causal interactions between spiking and field potential signals during behavior. *J. Neural Eng.* **19**, 026001 (2022).
40. Churchland, M. M., Santhanam, G. & Shenoy, K. V. Preparatory Activity in Premotor and Motor Cortex Reflects the Speed of the Upcoming Reach. *Journal of Neurophysiology* **96**, 3130–3146 (2006).

41. Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Ryu, S. I. & Shenoy, K. V. Cortical Preparatory Activity: Representation of Movement or First Cog in a Dynamical Machine? *Neuron* **68**, 387–400 (2010).

42. Vyas, S., O'Shea, D. J., Ryu, S. I. & Shenoy, K. V. Causal Role of Motor Preparation during Error-Driven Learning. *Neuron* **106**, 329-339.e4 (2020).

43. Mauk, M. D. & Buonomano, D. V. The Neural Basis of Temporal Processing. *Annual Review of Neuroscience* **27**, 307–340 (2004).

44. Remington, E. D., Egger, S. W., Narain, D., Wang, J. & Jazayeri, M. A Dynamical Systems Perspective on Flexible Motor Timing. *Trends in Cognitive Sciences* **22**, 938–952 (2018).

45. Chaisangmongkon, W., Swaminathan, S. K., Freedman, D. J. & Wang, X.-J. Computing by Robust Transience: How the Fronto-Parietal Network Performs Sequential, Category-Based Decisions. *Neuron* **93**, 1504-1517.e4 (2017).

46. Chaudhuri, R. & Fiete, I. Computational principles of memory. *Nature Neuroscience* **19**, 394–403 (2016).

47. Miller, E. K., Lundqvist, M. & Bastos, A. M. Working Memory 2.0. *Neuron* **100**, 463–475 (2018).

48. Buschman, T. J. & Kastner, S. From behavior to neural dynamics: An integrated theory of attention. *Neuron* **88**, 127–144 (2015).

49. Harbecke, J. The methodological role of mechanistic-computational models in cognitive science. *Synthese* 1–23 (2020) doi:10.1007/s11229-020-02568-5.

50. Lu, H.-Y. *et al.* Multi-scale neural decoding and analysis. *J. Neural Eng.* **18**, 045013 (2021).

51. Næss, S. *et al.* Corrected Four-Sphere Head Model for EEG Signals. *Front. Hum. Neurosci.* **11**, 490 (2017).

52. Vorwerk, J., Aydin, Ü., Wolters, C. H. & Butson, C. R. Influence of Head Tissue Conductivity Uncertainties on EEG Dipole Reconstruction. *Front. Neurosci.* **13**, 531 (2019).

53. Friston, K. J. Book Review: Brain Function, Nonlinear Coupling, and Neuronal Transients. *Neuroscientist* **7**, 406–418 (2001).

54. Yang, Q., Walker, E., Cotton, R. J., Tolias, A. S. & Pitkow, X. Revealing nonlinear neural decoding by analyzing choices. *Nat Commun* **12**, 6557 (2021).

55. Paynter, H. M. *Analysis and design of engineering systems*. (MIT Press, 1961).

56. Fernández, B., Prabhudesai, A. V., Murty, V. V., Gupta, R. & Chang, W. R. Neurobondgraphs: modeling environment of nonlinear dynamic systems using neural networks and bond graphs. in 75–90 (ASME, 1992).

57. Yeager, J. D., Phillips, D. J., Rector, D. M. & Bahr, D. F. Characterization of flexible ECoG electrode arrays for chronic recording in awake rats. *Journal of Neuroscience Methods* **173**, 279–285 (2008).

58. Choi, H. *et al.* Long-term evaluation and feasibility study of the insulated screw electrode for ECoG recording. *Journal of Neuroscience Methods* **308**, 261–268 (2018).

59. Merel, J., Botvinick, M. & Wayne, G. Hierarchical motor control in mammals and machines. *Nat Commun* **10**, 5489 (2019).

60. Sanes, J. N. & Donoghue, J. P. Oscillations in local field potentials of the primate motor cortex during voluntary movement. *PNAS* **90**, 4470–4474 (1993).

61. Khanna, P. & Carmena, J. M. Beta band oscillations in motor cortex reflect neural population signals that delay movement onset. *ELife* **6**, e24573 (2017).

62. Dancey, C. P. & Reidy, J. *Statistics without maths for psychology*. (Pearson education, 2007).

63. Akoglu, H. User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine* **18**, 91–93 (2018).

64. Fell, J. & Axmacher, N. The role of phase synchronization in memory processes. *Nat Rev Neurosci* **12**, 105–118 (2011).

65. Lachaux, J.-P., Rodriguez, E., Martinerie, J. & Varela, F. J. Measuring phase synchrony in brain signals. *Human Brain Mapping* **8**, 194–208 (1999).

66. Mormann, F., Lehnertz, K., David, P. & Elger, C. E. Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. *Physica D: Nonlinear Phenomena* **144**, 358–369 (2000).

67. Pedersen, M., Omidvarnia, A., Walz, J. M., Zalesky, A. & Jackson, G. D. Spontaneous brain network activity: Analysis of its temporal complexity. *Network Neuroscience* **1**, 100–115 (2017).

68. Wold, S., Esbensen, K. & Geladi, P. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems* **2**, 37–52 (1987).

69. Gallego, J. A., Perich, M. G., Chowdhury, R. H., Solla, S. A. & Miller, L. E. Long-term stability of cortical population dynamics underlying consistent behavior. *Nature Neuroscience* **23**, 260–270 (2020).

70. Goodall, C. R. 13 Computation using the QR decomposition. in *Handbook of Statistics* vol. 9 467–508 (Elsevier, 1993).

71. Afshar, A. *et al.* Single-Trial Neural Correlates of Arm Movement Preparation. *Neuron* **71**, 555–564 (2011).

72. Churchland, M. M. *et al.* Neural population dynamics during reaching. *Nature* **487**, 51–56 (2012).

73. Kaufman, M. T., Churchland, M. M., Ryu, S. I. & Shenoy, K. V. Cortical activity in the null space: permitting preparation without movement. *Nature Neuroscience* **17**, 440–448 (2014).

74. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78–84 (2013).

75. Sadtler, P. T. *et al.* Neural constraints on learning. *Nature* **512**, 423–426 (2014).

76. Gallego, J. A., Perich, M. G., Miller, L. E. & Solla, S. A. Neural Manifolds for the Control of Movement. *Neuron* **94**, 978–984 (2017).

77. Thompson, B. Canonical Correlation Analysis. in *Encyclopedia of Statistics in Behavioral Science* (eds. Everitt, B. S. & Howell, D.) (Wiley, 2005). doi:10.1002/0470013192.bsa068.

78. Gallego, J. A. *et al.* Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. *Nature Communications* **9**, 1–13 (2018).

79. Winkler, A. M., Renaud, O., Smith, S. M. & Nichols, T. E. Permutation Inference for Canonical Correlation Analysis. *arXiv reprint arXiv:2002.10046* (2020).

80. Gu, Q., Li, Z. & Han, J. Generalized Fisher Score for Feature Selection. Preprint at https://doi.org/10.48550/arXiv.1202.3725 (2012).

81. Duda, R. O., Hart, P. E. & Stork, D. G. *Pattern Classification. 2nd Edition.* (Hoboken: Wiley, 2000).

82. Cho, K., van Merrienboer, B., Bahdanau, D. & Bengio, Y. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. Preprint at https://doi.org/10.48550/arXiv.1409.1259 (2014).

83. Sanei, S. & Chambers, J. A. *EEG Signal Processing*. (John Wiley & Sons, 2013).

84. Nunez, P. L. & Srinivasan, R. *Electric Fields of the Brain: The neurophysics of EEG*. *Electric Fields of the Brain* (Oxford University Press, USA, 2006).

85. Pan, Y. & Wang, J. Model Predictive Control of Unknown Nonlinear Dynamical Systems Based on Recurrent Neural Networks. *IEEE Transactions on Industrial Electronics* **59**, 3089–3101 (2012).

86. Sussillo, D. Neural circuits as computational dynamical systems. *Current Opinion in Neurobiology* **25**, 156–163 (2014).

87. Hoerl, A. E. & Kennard, R. W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* **12**, 55–67 (1970).

88. Hastie, T. Ridge Regularization: An Essential Concept in Data Science. *Technometrics* **62**, 426–433 (2020).

89. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).

90. Hallez, H. *et al.* Review on solving the forward problem in EEG source analysis. *Journal of NeuroEngineering and Rehabilitation* **4**, 46 (2007).

91. Battaglia, D., Witt, A., Wolf, F. & Geisel, T. Dynamic Effective Connectivity of Inter-Areal Brain Circuits. *PLOS Computational Biology* **8**, e1002438 (2012).

92. Friston, K. J. Functional and Effective Connectivity: A Review. *Brain Connectivity* **1**, 13–36 (2011).

93. Funahashi, S., Bruce, C. J. & Goldman-Rakic, P. S. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology* **61**, 331–349 (1989).

94. Arce-McShane, F. I., Ross, C. F., Takahashi, K., Sessle, B. J. & Hatsopoulos, N. G. Primary motor and sensory cortical areas communicate via spatiotemporally coordinated networks at multiple frequencies. *PNAS* **113**, 5083–5088 (2016).

95. Heuvel, M. P. van den, Scholtens, L. H. & Kahn, R. S. Multiscale Neuroscience of Psychiatric Disorders. *Biological Psychiatry* **86**, 512–522 (2019).

96. Friston, K. J. The labile brain. I. Neuronal transients and nonlinear coupling. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **355**, 215–236 (2000).

97. Gawthrop, P. J. & Bevan, G. P. Bond-graph modeling. *IEEE Control Systems Magazine* **27**, 24–45 (2007).

98. *Bond Graphs for Modelling, Control and Fault Diagnosis of Engineering Systems*. (Berlin: Springer, 2017). doi:10.1007/978-3-319-47434-2.

99. Rashidi, H. H., Tran, N. K., Betts, E. V., Howell, L. P. & Green, R. Artificial Intelligence and Machine Learning in Pathology: The Present Landscape of Supervised Methods. *Academic Pathology* **6**, 2374289519873088 (2019).

100. Abbaspourazad, H., Choudhury, M., Wong, Y. T., Pesaran, B. & Shanechi, M. M. Multiscale low-dimensional motor cortical state dynamics predict naturalistic reach-and-grasp behavior. *Nat Commun* **12**, 607 (2021).

101. Presigny, C. & De Vico Fallani, F. Colloquium: Multiscale modeling of brain network organization. *Rev. Mod. Phys.* **94**, 031002 (2022).

102. Riedl, V. *et al.* Metabolic connectivity mapping reveals effective connectivity in the resting human brain. *Proceedings of the National Academy of Sciences* **113**, 428–433 (2016).

103. Harrison, L., Penny, W. D. & Friston, K. Multivariate autoregressive modeling of fMRI time series. *NeuroImage* **19**, 1477–1491 (2003).

104. Goebel, R., Roebroeck, A., Kim, D.-S. & Formisano, E. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magnetic Resonance Imaging* **21**, 1251–1261 (2003).

105. Friston, K. J. *et al.* Psychophysiological and Modulatory Interactions in Neuroimaging. *NeuroImage* **6**, 218–229 (1997).

106. McIntosh, A. R. & Gonzalez-Lima, F. Structural modeling of functional neural pathways mapped with 2-deoxyglucose: effects of acoustic startle habituation on the auditory system. *Brain Research* **547**, 295–302 (1991).

107. Büchel, C. & Friston, K. J. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. *Cerebral Cortex (New York, NY: 1991)* **7**, 768–778 (1997).

108. Bullmore, E. *et al.* How Good Is Good Enough in Path Analysis of fMRI Data? *NeuroImage* **11**, 289–301 (2000).

109. McLntosh, A. R. & Gonzalez-Lima, F. Structural equation modeling and its application to network analysis in functional brain imaging. *Human Brain Mapping* **2**, 2–22 (1994).

110. Penny, W. D., Stephan, K. E., Mechelli, A. & Friston, K. J. Modelling functional integration: a comparison of structural equation and dynamic causal models. *Neuroimage* **23 Suppl 1**, S264-74 (2004).

111. Michel, C. M. & Brunet, D. EEG Source Imaging: A Practical Review of the Analysis Steps. *Front. Neurol.* **10**, 325 (2019).

112. Characterization of regional differences in resting-state fMRI with a data-driven network model of brain dynamics | Science Advances. https://www.science.org/doi/10.1126/sciadv.abq7547.

113. Heitmann, S. & Breakspear, M. Putting the "dynamic" back into dynamic functional connectivity. *Network Neuroscience* **02**, 150–174 (2018).

114. Chang, Y.-J., Chen, Y.-I., Yeh, H.-C., Carmena, J. M. & Santacruz, S. R. Inferring system-level brain communication through multi-scale neural activity. 2020.11.30.404244 Preprint at https://doi.org/10.1101/2020.11.30.404244 (2021).

115. Orsborn, A. L. *et al.* Closed-Loop Decoder Adaptation Shapes Neural Plasticity for Skillful Neuroprosthetic Control. *Neuron* **82**, 1380–1393 (2014).

116. Dangi, S. *et al.* Continuous closed-loop decoder adaptation with a recursive maximum likelihood algorithm allows for rapid performance acquisition in brain-machine interfaces. *Neural Comput* **26**, 1811–1839 (2014).

117. Shanechi, M. M. Brain–machine interfaces from motor to mood. *Nature Neuroscience* **22**, 1554–1564 (2019).

118. Shenoy, K. V. & Carmena, J. M. Combining Decoder Design and Neural Adaptation in Brain-Machine Interfaces. *Neuron* **84**, 665–680 (2014).

119. Orsborn, A. L. & Pesaran, B. Parsing learning in networks using brain-machine interfaces. *Curr Opin Neurobiol* **46**, 76–83 (2017).

120. Schwarz, D. A. *et al.* Chronic, wireless recordings of large-scale brain activity in freely moving rhesus monkeys. *Nat Methods* **11**, 670–676 (2014).

121. Ganguly, K., Dimitrov, D. F., Wallis, J. D. & Carmena, J. M. Reversible large-scale modification of cortical networks during neuroprosthetic control. *Nat Neurosci* **14**, 662–667 (2011).

122. Inoue, Y., Mao, H., Suway, S. B., Orellana, J. & Schwartz, A. B. Decoding arm speed during reaching. *Nat Commun* **9**, 5243 (2018).

123. Pulvermüller, F., Tomasello, R., Henningsen-Schomers, M. R. & Wennekers, T. Biological constraints on neural network models of cognitive function. *Nat Rev Neurosci* **22**, 488–502 (2021).

124. Dayan, P. & Abbott, L. F. Theoretical neuroscience, vol. 806. (2001).

125. Nordbø, Ø., Wyller, J. & Einevoll, G. T. Neural network firing-rate models on integral form. *Biol Cybern* **97**, 195–209 (2007).

126. Nordlie, E., Tetzlaff, T. & Einevoll, G. Rate Dynamics of Leaky Integrate-and-Fire Neurons with Strong Synapses. *Frontiers in Computational Neuroscience* **4**, (2010).

127. Jansen, B. H. & Rit, V. G. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biol. Cybern.* **73**, 357–366 (1995).

128. De Brouwer, E., Simm, J., Arany, A. & Moreau, Y. GRU-ODE-Bayes: Continuous modeling of sporadically-observed time series. *arXiv:1905.12374 [cs, stat]* (2019).

129. Gray, R. & Robinson, P. Stability constraints on large-scale structural brain networks. *Frontiers in Computational Neuroscience* **7**, (2013).

130. Xu, T. & Barak, O. Dynamical Timescale Explains Marginal Stability in Excitability Dynamics. *J. Neurosci.* **37**, 4508–4524 (2017).

131. Ganguly, K. & Carmena, J. M. Emergence of a Stable Cortical Map for Neuroprosthetic Control. *PLOS Biology* **7**, e1000153 (2009).

132. Athalye, V. R., Ganguly, K., Costa, R. M. & Carmena, J. M. Emergence of Coordinated Neural Dynamics Underlies Neuroprosthetic Learning and Skillful Control. *Neuron* **93**, 955-970.e5 (2017).

133. Golub, M. D. *et al.* Learning by neural reassociation. *Nat Neurosci* **21**, 607–616 (2018).

134. Oby, E. R. *et al.* New neural activity patterns emerge with long-term learning. *Proceedings of the National Academy of Sciences* **116**, 15210–15215 (2019).

135. Zippi, E. L., You, A. K., Ganguly, K. & Carmena, J. M. Selective modulation of cortical population dynamics during neuroprosthetic skill learning. *Sci Rep* **12**, 15948 (2022).

136. Nichols, M. J. & Newsome, W. T. The neurobiology of cognition. *Nature* **402**, C35–C38 (1999).

137. Lisman, J. The challenge of understanding the brain: where we stand in 2015. *Neuron* **86**, 864–882 (2015).

138. Abbott, L. F. Theoretical Neuroscience Rising. *Neuron* **60**, 489–495 (2008).

139. Kriegeskorte, N. & Douglas, P. K. Cognitive computational neuroscience. *Nature Neuroscience* **21**, 1148–1160 (2018).

140. McKenna, T. M., McMullen, T. A. & Shlesinger, M. F. The brain as a dynamic physical system. *Neuroscience* **60**, 587–605 (1994).

141. Freeman, W. J. Mesoscopic neurodynamics: From neuron to brain. *Journal of Physiology-Paris* **94**, 303–322 (2000).

142. Stam, C. J. Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field. *Clinical Neurophysiology* **116**, 2266–2301 (2005).

143. Breakspear, M. Dynamic models of large-scale brain activity. *Nature Neuroscience* **20**, 340–352 (2017).

144. Roberts, J. A. *et al.* Metastable brain waves. *Nat Commun* **10**, 1056 (2019).

145. Bansal, K. *et al.* Cognitive chimera states in human brain networks. *Science Advances* **5**, eaau8535 (2019).

146. Lynn, C. W. & Bassett, D. S. The physics of brain network structure, function and control. *Nat Rev Phys* **1**, 318–332 (2019).

147. Voelker, A., Kajić, I. & Eliasmith, C. Legendre Memory Units: Continuous-Time Representation in Recurrent Neural Networks. in *Advances in Neural Information Processing Systems* vol. 32 (Curran Associates, Inc., 2019).

148. Chang, B., Chen, M., Haber, E. & Chi, E. H. AntisymmetricRNN: A Dynamical System View on Recurrent Neural Networks. Preprint at https://doi.org/10.48550/arXiv.1902.09689 (2019).

149. Chen, R. T. Q., Rubanova, Y., Bettencourt, J. & Duvenaud, D. Neural Ordinary Differential Equations. *arXiv:1806.07366 [cs, stat]* (2019).

150. Kang, H. The prevention and handling of the missing data. *Korean J Anesthesiol* **64**, 402–406 (2013).

151. Stein, R. B. & Hodgkin, A. L. The frequency of nerve action potentials generated by applied currents. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **167**, 64–86 (1997).

152. Potjans, T. C. & Diesmann, M. The Cell-Type Specific Cortical Microcircuit: Relating Structure and Activity in a Full-Scale Spiking Network Model. *Cerebral Cortex* **24**, 785–806 (2014).

153. Montbrió, E., Pazó, D. & Roxin, A. Macroscopic Description for Networks of Spiking Neurons. *Phys. Rev. X* **5**, 021028 (2015).

154. Schmidt, H., Avitabile, D., Montbrió, E. & Roxin, A. Network mechanisms underlying the role of oscillations in cognitive tasks. *PLOS Computational Biology* **14**, e1006430 (2018).

155. Wilson, H. R. & Cowan, J. D. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.* **12**, 1–24 (1972).

156. Arnold, T. W. Uninformative Parameters and Model Selection Using Akaike's Information Criterion. *The Journal of Wildlife Management* **74**, 1175–1178 (2010).

157. Vrieze, S. I. Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods* **17**, 228–243 (2012).

158. Chen, J. & Chen, Z. Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* **95**, 759–771 (2008).

159. Grünwald, P. Model Selection Based on Minimum Description Length. *Journal of Mathematical Psychology* **44**, 133–152 (2000).

160. Hansen, M. H. & Yu, B. Model selection and the principle of minimum description length. *Journal of the American Statistical Association* **96**, 746–774 (2001).

161. Sternad, D. It's not (only) the mean that matters: variability, noise and exploration in skill learning. *Current Opinion in Behavioral Sciences* **20**, 183–195 (2018).

162. Waschke, L., Kloosterman, N. A., Obleser, J. & Garrett, D. D. Behavior needs neural variability. *Neuron* **109**, 751–766 (2021).

163. Sun, X. *et al.* Cortical preparatory activity indexes learned motor memories. *Nature* 1–6 (2022) doi:10.1038/s41586-021-04329-x.

164. Churchland, M. M. *et al.* Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat Neurosci* **13**, 369–378 (2010).

165. Dhawale, A. K., Smith, M. A. & Ölveczky, B. P. The Role of Variability in Motor Learning. *Annual Review of Neuroscience* **40**, 479–498 (2017).

166. Hell, S. W. & Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt. Lett., OL* **19**, 780–782 (1994).

167. Betzig, E. *et al.* Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science* **313**, 1642–1645 (2006).

168. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods* **3**, 793–796 (2006).

169. Sahl, S. J., Hell, S. W. & Jakobs, S. Fluorescence nanoscopy in cell biology. *Nat Rev Mol Cell Biol* **18**, 685–701 (2017).

170. Westphal, V. *et al.* Video-Rate Far-Field Optical Nanoscopy Dissects Synaptic Vesicle Movement. *Science* **320**, 246–249 (2008).

171. Schneider, J. *et al.* Ultrafast, temporally stochastic STED nanoscopy of millisecond dynamics. *Nat Methods* **12**, 827–830 (2015).

172. Calovi, S., Soria, F. N. & Tønnesen, J. Super-resolution STED microscopy in live brain tissue. *Neurobiology of Disease* **156**, 105420 (2021).

173. Hell, S. W. Microscopy and its focal switch. *Nat Methods* **6**, 24–32 (2009).

174. Laissue, P. P., Alghamdi, R. A., Tomancak, P., Reynaud, E. G. & Shroff, H. Assessing phototoxicity in live fluorescence imaging. *Nat Methods* **14**, 657–661 (2017).

175. Vicidomini, G., Bianchini, P. & Diaspro, A. STED super-resolved microscopy. *Nat Methods* **15**, 173–182 (2018).

176. Kilian, N. *et al.* Assessing photodamage in live-cell STED microscopy. *Nat Methods* **15**, 755–756 (2018).

177. Ma, Y. & Ha, T. Fight against background noise in stimulated emission depletion nanoscopy. *Phys. Biol.* **16**, 051002 (2019).

178. Gao, P., Prunsche, B., Zhou, L., Nienhaus, K. & Nienhaus, G. U. Background suppression in fluorescence nanoscopy with stimulated emission double depletion. *Nature Photon* **11**, 163–169 (2017).

179. Vicidomini, G., Moneron, G., Eggeling, C., Rittweger, E. & Hell, S. W. STED with wavelengths closer to the emission maximum. *Opt. Express, OE* **20**, 5225–5236 (2012).

180. Ronzitti, E., Harke, B. & Diaspro, A. Frequency dependent detection in a STED microscope using modulated excitation light. *Opt. Express, OE* **21**, 210–219 (2013).

181. Coto Hernàndez, I. *et al.* A new filtering technique for removing anti-Stokes emission background in gated CW-STED microscopy. *Journal of Biophotonics* **7**, 376–380 (2014).

182. Castello, M. *et al.* Removal of anti-Stokes emission background in STED microscopy by FPGA-based synchronous detection. *Review of Scientific Instruments* **88**, 053701 (2017).

183. Heine, J. *et al.* Adaptive-illumination STED nanoscopy. *Proceedings of the National Academy of Sciences* **114**, 9797–9802 (2017).

184. Vicidomini, G. *et al.* Sharper low-power STED nanoscopy by time gating. *Nat Methods* **8**, 571–573 (2011).

185. Moffitt, J. R., Osseforth, C. & Michaelis, J. Time-gating improves the spatial resolution of STED microscopy. *Opt. Express, OE* **19**, 4242–4254 (2011).

186. Vicidomini, G. *et al.* STED Nanoscopy with Time-Gated Detection: Theoretical and Experimental Aspects. *PLOS ONE* **8**, e54421 (2013).

187. Lanzanò, L. *et al.* Encoding and decoding spatio-temporal information for super-resolution microscopy. *Nat Commun* **6**, 6701 (2015).

188. Sun, Y. *et al.* A novel pulsed STED microscopy method using FastFLIM and the phasor plots. in *Multiphoton Microscopy in the Biomedical Sciences XVII* vol. 10069 136–151 (SPIE, 2017).

189. Wang, L. *et al.* Resolution improvement in STED super-resolution microscopy at low power using a phasor plot approach. *Nanoscale* **10**, 16252–16260 (2018).

190. Digman, M. A., Caiolfa, V. R., Zamai, M. & Gratton, E. The Phasor Approach to Fluorescence Lifetime Imaging Analysis. *Biophysical Journal* **94**, L14–L16 (2008).

191. Tortarolo, G. *et al.* Photon-separation to enhance the spatial resolution of pulsed STED microscopy. *Nanoscale* **11**, 1754–1761 (2019).

192. Ma, Y., Lee, Y., Best-Popescu, C. & Gao, L. High-speed compressed-sensing fluorescence lifetime imaging microscopy of live cells. *Proceedings of the National Academy of Sciences* **118**, e2004176118 (2021).

193. Chen, Y.-I. *et al.* Generative adversarial network enables rapid and robust fluorescence lifetime image analysis in live cells. *Commun Biol* **5**, 1–11 (2022).

194. Moussavi-Baygi, R., Jamali, Y., Karimi, R. & Mofrad, M. R. K. Brownian Dynamics Simulation of Nucleocytoplasmic Transport: A Coarse-Grained Model for the Functional State of the Nuclear Pore Complex. *PLOS Computational Biology* **7**, e1002049 (2011).

195. Goodfellow, I. *et al.* Generative Adversarial Nets. in *Advances in Neural Information Processing Systems* vol. 27 (Curran Associates, Inc., 2014).

196. Arjovsky, M., Chintala, S. & Bottou, L. Wasserstein GAN. Preprint at https://doi.org/10.48550/arXiv.1701.07875 (2017).

197. Sarmento, M. J. *et al.* Exploiting the tunability of stimulated emission depletion microscopy for super-resolution imaging of nuclear structures. *Nat Commun* **9**, 3415 (2018).

198. Wang, H. *et al.* Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nature methods* **16**, 103–110 (2019).

199. Bouchard, C. *et al.* Resolution Enhancement with a Task-Assisted GAN to Guide Optical Nanoscopy Image Analysis and Acquisition. 2021.07.19.452964 Preprint at https://doi.org/10.1101/2021.07.19.452964 (2022).

200. Cutrale, F. *et al.* Hyperspectral phasor analysis enables multiplexed 5D in vivo imaging. *Nat Methods* **14**, 149–152 (2017).

201. Ranjit, S., Malacrida, L., Jameson, D. M. & Gratton, E. Fit-free analysis of fluorescence lifetime imaging data using the phasor approach. *Nat Protoc* **13**, 1979–2004 (2018).

202. Wang, P. *et al.* Complex wavelet filter improves FLIM phasors for photon starved imaging experiments. *Biomedical Optics Express* **12**, 3463–3473 (2021).

203. Selesnick, I. W., Baraniuk, R. G. & Kingsbury, N. C. The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine* **22**, 123–151 (2005).

204. Pelicci, S., Tortarolo, G., Vicidomini, G., Diaspro, A. & Lanzanò, L. Improving SPLIT-STED super-resolution imaging with tunable depletion and excitation power. *J. Phys. D: Appl. Phys.* **53**, 234003 (2020).

205. Wang, L. *et al.* Low-power STED nanoscopy based on temporal and spatial modulation. *Nano Res.* **15**, 3479–3486 (2022).

206. Pan, S. J. & Yang, Q. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering* **22**, 1345–1359 (2010).

207. Neupane, B. *et al.* Continuous-Wave Stimulated Emission Depletion Microscope for Imaging Actin Cytoskeleton in Fixed and Live Cells. *Sensors* **15**, 24178–24190 (2015).

208. Bückers, J., Wildanger, D., Vicidomini, G., Kastrup, L. & Hell, S. W. Simultaneous multi-lifetime multi-color STED imaging for colocalization analyses. *Optics express* **19**, 3130–3143 (2011).

209. Gonzalez Pisfil, M. *et al.* Stimulated emission depletion microscopy with a single depletion laser using five fluorochromes and fluorescence lifetime phasor separation. *Sci Rep* **12**, 14027 (2022).

210. Wang, C. *et al.* A photostable fluorescent marker for the superresolution live imaging of the dynamic structure of the mitochondrial cristae. *Proceedings of the National Academy of Sciences* **116**, 15817–15822 (2019).

211. Yang, X. *et al.* Mitochondrial dynamics quantitatively revealed by STED nanoscopy with an enhanced squaraine variant probe. *Nat Commun* **11**, 3699 (2020).

212. Frei, M. S. *et al.* Engineered HaloTag variants for fluorescence lifetime multiplexing. *Nat Methods* **19**, 65–70 (2022).

213. Glogger, M. *et al.* Synergizing Exchangeable Fluorophore Labels for Multitarget STED Microscopy. *ACS Nano* **16**, 17991–17997 (2022).