

ABSTRACT

Title of Dissertation: **BAYESIAN INFERENCE OF LATENT SPECTRAL
AND TEMPORAL NETWORK ORGANIZATIONS
FROM HIGH DIMENSIONAL NEURAL DATA**

Rupasinghe Arachchige Anuththara Rupasinghe
Doctor of Philosophy, 2022

Dissertation Directed by: **Professor Behtash Babadi**
Department of Electrical and Computer Engineering

The field of neuroscience has striven for more than a century to understand how the brain functionally coordinates billions of neurons to perform its many tasks. Recent advancements in neural data acquisition techniques such as multi-electrode arrays, two-photon calcium imaging, and high-speed light-sheet microscopy have significantly contributed to this endeavor's progression by facilitating concurrent observation of spiking activity in large neuronal populations. However, existing methods for network-level inference from these data have several shortcomings: including undermining the non-linear dynamics, ignoring non-stationary brain activity, and causing error propagation by performing inference in a multi-stage fashion. The goal of this dissertation is to close this gap by developing models and methods to directly infer the dynamic spectral and temporal network organizations in the brain, from these ensemble neural data.

In the first part of this dissertation, we introduce Bayesian methods to infer dynamic frequency-domain network organizations in neuronal ensembles from spiking observations, by

integrating techniques such as point process modeling, state-space estimation, and multitaper spectral estimation. Firstly, we introduce a semi-stationary multitaper multivariate spectral analysis method tailored for neuronal spiking data and establish theoretical bounds on its performance. Building upon this estimator, we then introduce a framework to derive spectrotemporal Granger causal interactions in a population of neurons from spiking data. We demonstrate the validity of these methods through simulations, and applications on real data recorded from cortical neurons of rats during sleep, and human subjects undergoing anesthesia. Finally, we extend these methods to develop a precise frequency-domain inference method to characterize human heart rate variability from electrocardiogram data.

The second part introduces a methodology to directly estimate signal and noise correlation networks from two-photon calcium imaging observations. We explicitly model the observation noise, temporal blurring of spiking activities, and other underlying non-linearities in a Bayesian framework, and derive an efficient variational inference method. We demonstrate the validity of the resulting estimators through theoretical analysis and extensive simulations, all of which establish significant gains over existing methods. Applications of our method on real data recorded from the mouse primary auditory cortex reveal novel and distinct spatial patterns in the correlation networks. Finally, we use our methods to investigate how the correlation networks in the auditory cortex change under different stimulus conditions, and during perceptual learning.

In the third part, we investigate the respiratory network and the swimming-respiration coordination in larval zebrafish by applying several spectro-temporal analysis techniques, on whole-brain light-sheet microscopy imaging data. Firstly, using multitaper spectrotemporal analysis techniques, we categorize brain regions that are synchronized with the respiratory rhythm based on their distinct phases. Then, we demonstrate that zebrafish swimming is phase-locked to breath-

ing. Next, through the analysis of neural activity and behavior under optogenetic stimulations and two-photon ablations, we identify the brain regions that are key for this swimming-respiration coordination. Finally, using the Izhikevich model for spiking neurons, we develop and simulate a circuit model that replicates this swimming-respiration coupling phenomenon, providing new insights into the possible underlying neural circuitry.

BAYESIAN INFERENCE OF
LATENT SPECTRAL AND TEMPORAL NETWORK ORGANIZATIONS
FROM HIGH DIMENSIONAL NEURAL DATA

by

Rupasinghe Arachchige Anuththara Rupasinghe

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2022

Advisory Committee:

Professor Behtash Babadi, Chair/Advisor
Professor Jonathan Z. Simon
Professor Shihab Shamma
Professor Patrick O. Kanold
Dr. Misha Ahrens
Professor Daniel A. Butts, Dean's Representative

© Copyright by
Rupasinghe Arachchige Anuththara Rupasinghe
2022

Dedication

To *my husband* and *my parents*, for their endless love and support.

Acknowledgments

First and foremost, I would like to take this opportunity to thank my Ph.D. advisor, Professor Behtash Babadi for his continuous guidance, kind support, and untiring supervision throughout my Ph.D. This dissertation would not have been possible if not for his brilliant ideas, exceptional signal processing expertise, constant accessibility for discussions, and most importantly never giving-up attitude. His passion, dedication, enthusiasm, and diligence toward work greatly inspired me and brought out the best in me. He has played many roles in my life during the past five years in addition to being my research advisor: a mentor, a well-wisher, a confidant, and a role model. Words fail to express my profound gratitude toward him, and I am truly honored to have had the unique privilege of being one of his early Ph.D. students.

I am also very grateful to my committee members Prof. Jonathan Simon, Prof. Shihab Shamma, Prof. Patrick Kanold, Dr. Misha Ahrens, and Prof. Daniel Butts who kindly agreed to serve on my dissertation committee and provide feedback to enrich this thesis, amidst their busy schedules. I would also like to thank Prof. Patrick Kanold, Dr. Misha Ahrens, Prof. Melissa Caras, and Prof. Nicholas Schiff for all the productive collaborations, and their helpful inputs to this dissertation. I would also like to express my gratitude to all my other collaborators: Nikolas Francis, Ji Liu, Zac Bowen, Peter Jendrichovsky, Rongwei Zhang, Jackie Gottshall, Natasha Recorder, and Shoutik Mukherjee, for sharing their expertise and valuable data recordings.

During my Ph.D. graduate studies, I had the pleasure of learning diverse courses from very

experienced professors at UMD. While thanking them all, I specifically express my heartfelt gratitude to Prof. Armand Makowski, for encouraging me to explore beyond my comfort zones and identify my strengths during the first year of my Ph.D. I also would like to thank the staff of the ECE department at UMD: Melanie Prange, Emily Irwin, Vivian Lu, Guerda Mervilus, Jane Wang, and Maria Hoo; for being extremely friendly and helpful to me over the past five years.

I am also very grateful to the senior students of my lab group: Sina Miran and Proloy Das, for all their guidance and friendly support with my research and graduate work. I was also very lucky to have had the companionship of many colleagues including Shoutik, Behrad, Sahar, Ruwanthi, Pardis, and many friends in the DC area including Nadee, Umesha, Pranjeevan, Sajani, Dilhara, Dushyanthi, Yashish, Shavindra, Kasuni, Suresh, and many others back at home in Sri Lanka; they made this journey very memorable, pleasant, and colorful. I am also truly indebted to all my teachers from kindergarten to graduate school, for leading the way to where I am today.

Finally, and most importantly, I owe my deepest gratitude to my family, my parents – Sudharma and Nihal Rupasinghe, parents-in-law – Prabha and S. B. Weerakoon, sisters – Madhuwanthi and Sawandika, and brothers – Nishantha, Rajitha, Heshan, and Buddhika, for providing me endless love and support all these years, from far away. This thesis would not have been possible if not for all the sacrifices they had to make, by constantly supporting and encouraging me to follow my dreams. Lastly, I would like to dedicate this thesis to my husband – Lasitha Weerakoon, for being my emotional, intellectual, and physical support pillar all these years. I would not have been able to endure this journey if not for his endless love, support, friendship, care, and understanding. I am forever grateful to be blessed with such a selfless family, I love you all!

- Anuththara Rupasinghe (June 2022)

Table of Contents

Dedication	ii
Acknowledgements	iii
Table of Contents	v
List of Tables	ix
List of Figures	x
List of Abbreviations	xiii
Chapter 1: Introduction	1
1.1 Limitations of existing methods	2
1.2 Contributions	4
1.2.1 Part I: Frequency-domain inference	5
1.2.2 Part II: Signal and Noise Correlation inference	7
1.2.3 Part III: Zebrafish respiration-locomotion network coordination	10
1.3 Broader Impact	10
I Frequency-domain inference	12
Chapter 2: Multitaper Analysis of Semi-Stationary Spectra from Multivariate Neuronal Spiking Observations	13
2.1 Problem Formulation	16
2.2 Multitaper Estimation of the SSD Matrix	20
2.2.1 The Proposed Multitaper Framework	20
2.2.2 MAP Estimation of the Parameters via the EM Algorithm	23
2.3 Theoretical Analysis	26
2.4 Simulation Studies	32
2.4.1 Existing Methods and Benchmarks for Comparison	32
2.4.2 Simulation study 1: Estimating the SSD of a Latent Trivariate Process	34
2.4.3 Robustness to the Number of Realizations and Spiking Rate	38
2.4.4 Choice of Design Parameters and Hyper-parameters	40
2.4.5 Simulation study 2: Estimating the SSD of a Bivariate Process, with Only One Directly Observable Component	41

2.5	Real data studies	43
2.5.1	Real data study 1: Application to Experimentally-recorded Data from the Rat Cortical Neurons During Sleep	43
2.5.2	Real data study 2: Application to experimentally-recorded data from general anesthesia	47
2.6	Concluding Remarks	50
Chapter 3: Adaptive Frequency-domain Granger Causal Inference from Neuronal Ensemble Data		52
3.1	Methods	54
3.1.1	Geweke's Conditional Granger Causality framework	54
3.1.2	Proposed forward model	55
3.1.3	MAP estimation via Expectation-Maximization	57
3.1.4	Testing statistical significance	59
3.2	Results	60
3.2.1	Simulation study	60
3.2.2	Application to experimentally recorded data	62
3.3	Concluding Remarks	64
Chapter 4: Denoised Multitaper Frequency-domain Heart Rate Variability inference		65
4.1	Methods	67
4.1.1	Data Pre-processing and Preliminary Assumptions	67
4.1.2	Denoised Multitaper Spectral Estimation	67
4.1.3	Confidence Intervals	69
4.2	Results	70
4.3	Concluding remarks	72

II Signal and Noise Correlation inference 74

Chapter 5: Direct Inference of Signal and Noise Correlations from Two-Photon Calcium Imaging of Ensemble Neuronal Activity		75
5.1	Results	78
5.1.1	Signal and Noise correlations	78
5.1.2	Existing methods used for performance comparison	81
5.1.3	Simulation study 1: Neuronal ensemble driven by external stimulus	82
5.1.4	Simulation study 2: Spontaneous activity	87
5.1.5	Real data study 1: Mouse auditory cortex under random tone presentation	88
5.1.6	Real data study 2: Spontaneous vs. stimulus-driven activity in the mouse A1	95
5.1.7	Real data study 3: Spatial analysis of signal and noise correlations in the mouse A1	97
5.1.8	Theoretical analysis of the bias and variance of the proposed estimators	102
5.2	Discussion	105
5.3	Methods and Materials	110

5.3.1	Proposed forward model	110
5.3.2	Overview of the proposed estimation method	112
5.3.3	Preliminary assumptions	114
5.3.4	Decoupling via Pólya-Gamma augmentation	114
5.3.5	Deriving the optimal variational densities	115
5.3.6	Low-complexity parameter updates	117
5.3.7	Guidelines for model parameter settings	120
5.3.8	Performance evaluation	126
5.3.9	Hyper-parameter tuning	129
5.3.10	Experimental procedures	131
Chapter 6:	Applications of Signal and Noise Correlation analysis	134
6.1	Pure-tone vs two-tone network organizations	134
6.1.1	Methods	135
6.1.2	Results	138
6.1.3	Concluding remarks	155
6.2	Network-level changes during perceptual learning	157
6.2.1	Results	158
6.2.2	Concluding remarks	164
III	Zebrafish respiration-locomotion network coordination	166
Chapter 7:	Investigating and modeling larval zebrafish respiration-locomotion coordi- nation neural mechanisms	167
7.1	Motivation: Preliminary observations	168
7.2	Cross Spectral and Spectral Granger Causality analysis	172
7.2.1	Replicating the spectro-temporal coupling in real data using AR modeling	175
7.2.2	Relative time-shifts between the three oscillatory processes	178
7.3	Swimming occurs at a distinct phase of the breathing rhythm	181
7.4	Timing of grating induced swimming depends on pre-stimulus EEG	185
7.5	Exploring breathing activity at the individual-cell level	190
7.5.1	Reconstructing EEG through regression	193
7.6	Optogenetic stimulation of the middle C-MO changes the breathing rhythm	196
7.7	Ablation of Lateral-MO does not affect phase coupling between swimming and breathing	199
7.8	Proposed circuit model for respiratory-swimming coupling	201
7.9	Concluding Remarks	208
Chapter 8:	Concluding Remarks and Future Directions	211
8.1	Summary and Extensions of our Contributions	211
8.2	Limitations of our approaches	213
8.3	Future Directions	215
Appendices		218

Appendix A: Supplementary Material on Chapter 2	218
A.1 Proof of Theorem 2.1	218
A.2 Proof of Corollaries 2.1 and 2.2	224
Appendix B: Supplementary Material on Chapter 5	226
B.1 Relationship to existing definitions of Signal and Noise correlations	226
B.2 Proof of Theorem 5.1	230
B.3 Adapting the proposed signal and noise correlation estimates to spiking observations	254
B.4 Supplementary material of the Simulation study 1 in Section 5.1.3 (Supplemen- tary figures to Figure 5.2)	256
B.4.1 Sensitivity of two-stage estimates to the choice of the underlying spike deconvolution technique	256
B.4.2 Performance of two-stage estimates based on ground truth spikes	257
B.4.3 Performance comparison under stimulus integration model mismatch	259
B.4.4 Proposed estimates based on simulated data with model mismatch and at lower SNR	260
B.4.5 Performance comparison under varying SNR levels and firing rates	261
B.4.6 Performance comparison under observation noise model mismatch	262
B.5 Supplementary material of the real data study 1 in Section 5.1.5 (Supplementary figures to Figure 5.4)	264
B.5.1 Probing the effect of stimulus integration window length on the perfor- mance of the proposed estimates	264
B.5.2 Inspecting the inferred latent processes under high fluorescence activity due to rapid increase in firing rate	266
B.6 Supplementary material of the real data study 2 in Section 5.1.6 (Supplementary figures to Figure 5.6)	267
B.6.1 Histograms of the similarity/dissimilarity metrics under the shuffling pro- cedure	267
B.7 Supplementary material of the real data study 3 in Section 5.1.7 (Supplementary figures to Figure 5.7)	268
B.7.1 Comparing the marginal distributions of signal and noise correlations along the dorsoventral and rostrocaudal axes	268
B.7.2 Marginal angular distributions of signal and noise correlations	269
Bibliography	271

List of Tables

2.1	Comparison of MSE and spectral leakage in study 1	37
5.1	Dissimilarity metric statistics for the estimates in Figure 5.4-A and linear regression statistics of the comparison between signal and noise correlations in Figure 5.4-E.	91
5.2	The average NMSE across 50 trials used in the shuffling procedure illustrated in Figure 5.5-A.	95
5.3	Similarity/dissimilarity metric statistics for the estimates in Figure 5.6.	97
5.4	Linear regression statistics for the analysis of correlations vs. cell-pair distance .	100
7.1	Relative phase and time shifts at peak EEG oscillation frequency	180
7.2	Linear regression statistics	187

List of Figures

2.1	Samples from data used for spectral estimation in simulation study 1	35
2.2	Estimated spectrograms in simulation study 1	35
2.3	A snapshot of the spectrograms estimated in simulation study 1	36
2.4	Analysis of the robustness of the estimates in simulation study 1	39
2.5	Estimated spectrograms in two cases of the robustness analysis	39
2.6	Estimated spectrograms in simulation study 2	42
2.7	Sample data from the rat cortical neurons during sleep, data used for the first real data study	44
2.8	Estimated spectrograms in the analysis of rat cortical activity during sleep	45
2.9	Sample data from a human subject under Propofol-induced general anesthesia, data used for the second real data study	48
2.10	Estimated spectrograms in the analysis of data from a human subject under propofol-induced general anesthesia	49
3.1	Samples from data used for the spectral GC simulation study	61
3.2	Ground truth spectrograms and conditional spectral GC links of the simulation study	61
3.3	Comparison of estimated spectrograms and conditional GC estimates in the simulation study	62
3.4	Spectral GC analysis of rat cortical spiking during sleep	63
4.1	HRV time-domain pre-processing steps	71
4.2	Multitaper frequency-domain HRV analysis results	72
5.1	The proposed generative model for signal and noise correlations and the inverse problem	79
5.2	First simulation study: neuronal ensemble driven by external stimulus	83
5.3	Second simulation study: spontaneous activity	88
5.4	First real data study: application to experimentally-recorded data from the mouse A1	89
5.5	Assessing the specificity of different estimation results in the first real data study .	94
5.6	Second real data study: comparison of spontaneous and stimulus-driven activity in the mouse A1	96
5.7	Third real data study: comparison of signal and noise correlations across layers 2/3 and 4	99
5.8	Probabilistic graphical model of the proposed forward model	111

6.1	The design of the experiment and preliminary observations	139
6.2	Variation of correlations vs cell-pair distances	141
6.3	Investigating spatial patterns in 2D spatial spreads of correlations	143
6.4	Radial and angular marginal distributions	145
6.5	Best frequency same vs best frequency different interactions	147
6.6	Contributions of suppressive vs facilitative interactions	149
6.7	Results under non-harmonic tones	151
6.8	Control experiment: hearing-impaired mice	154
6.9	Sample signal and noise correlation matrices from gerbil auditory cortex	159
6.10	Comparing disengaged sessions with engaged sessions	160
6.11	Similarity of correlation networks across modulation levels vs average hit rate	161
6.12	Similarity of signal correlation networks across modulation levels on different days	162
6.13	Identifying network-level changes in the gerbil auditory cortex during perceptual learning	163
7.1	Rhythmic activity in the EEG recorded from the zebrafish tectum	168
7.2	Identifying the main brain regions with oscillations in sync with the EEG	169
7.3	Sample time-domain signals from the EEG, Caudal-MO, and Lateral-MO	170
7.4	Comparison of the breathing oscillatory signals with the EMG recording	171
7.5	Cross power spectral density analysis	173
7.6	Spectral Granger Causality analysis	174
7.7	Tri-variate oscillatory process simulated using AR modeling	176
7.8	Cross power spectral density analysis of the simulation	177
7.9	Spectral Granger Causality analysis of the simulation	178
7.10	Sign correction of the Caudal-MO activity	179
7.11	Changes in the phase spectrograms with sign correction	180
7.12	Swim events occur at a particular breathing oscillation phase	181
7.13	Swim triggered averages of activity and phase of oscillations	182
7.14	Swim triggered averages of instantaneous oscillation phase	183
7.15	Instantaneous-phase triggered averages of EMG activity	184
7.16	Grating stimulus onset triggered averages	185
7.17	Grating induced swimming timing depends on pre-stimulus EEG	186
7.18	Linear regression between t_1 and t_2	187
7.19	Example trials for swimming under grating during hypoxia	188
7.20	Pre-stimulus EEG peak triggered averages	189
7.21	Extracting cells having significant coherence with EEG	191
7.22	Magnitude and phase spectrograms of cells with significant coherence	192
7.23	Phase sequence of cells with significant coherence	193
7.24	Cells selected for the regression analysis	194
7.25	Comparison between the observed EEG and the regressed EEG	195
7.26	Cross power spectral density analysis between true EEG and reconstructed EEG	196
7.27	Optogenetic stimulation experiment	197
7.28	Results of the optogenetic stimulation experiment	198
7.29	Ablation of Lateral-MO experiment	200
7.30	Effect of L-MO ablations on phase triggered averages of swimming	201

7.31	Proposed circuit model for respiratory-swimming coupling	202
7.32	Sample activity simulated from Izhikevich neuronal model	205
7.33	Simulation under different oxygen levels	206
7.34	Simulation under L-MO ablation	206
7.35	Simulation results across multiple trials	207
B.1	Sensitivity of two-stage estimates to the choice of the underlying spike deconvolution technique, in the first simulation study	257
B.2	Performance of two-stage estimates based on ground truth spikes, in the first simulation study	258
B.3	Performance comparison under stimulus integration model mismatch, in the first simulation study	259
B.4	Proposed estimates based on simulated data with model mismatch and at lower SNR, in the first simulation study	260
B.5	Performance comparison at different SNR levels and average firing rate settings, in the first simulation study	262
B.6	Performance comparison under observation noise model mismatch, in the first simulation study	263
B.7	Probing the effect of stimulus integration window length on the performance of the proposed estimates, in the first real data study.	265
B.8	Inspecting the inferred latent processes under high fluorescence activity due to rapid increase in firing rate, in the first real data study.	267
B.9	Histograms of the similarity/dissimilarity metrics under the shuffling procedure, in the second real data study	268
B.10	Comparing the marginal distributions of signal and noise correlations along the dorsoventral and rostrocaudal axes, in the third real data study	269
B.11	Marginal angular distributions of signal and noise correlations, in the third real data study	270

List of Abbreviations

AIC	Akaike Information Criterion
AM	Amplitude Modulation
ANS	Autonomous Nervous System
AR	Auto-Regressive
BIC	Bayesian Information Criterion
BF	Best Frequency
BP	Blood Pressure
CAVI	Coordinate Ascent Variational Inference
CIF	Conditional Intensity Function
C-MO	Caudal Medulla Oblongata
CPSD	Cross Power Spectral Density
DC	Direct Current
dpss	discrete prolate spheroidal sequences
EEG	Electroencephalography
EKG	Electrocardiogram
ELBO	Evidence Lower Bound
EM	Expectation Maximization
EMG	Electromyography
FCSS	Fast Compressible State Space
FDB	Frequency Domain Bootstrapping
FFT	Fast Fourier Transform
FIS	Fixed Interval Smoothing
GC	Granger Casual
GPFA	Gaussian Process Factor Analysis
HF	High Frequency
HRV	Heart Rate Variability

iid	independent and identically distributed
IRLS	Iteratively Re-weighted Least Squares
KL	Kullback-Leibler
KS	Kolmogorov–Smirnov
LF	Low Frequency
LFP	Local Field Potential
LRC	Locomotion Respiratory Coupling
L-MO	Lateral Medulla Oblongata
MAP	Maximum A Posteriori
ML	Maximum Likelihood
MO	Medulla Oblongata
MSE	Mean Squared Error
NMSE	Normalized Mean Squared Error
pE	putative pyramidal cells
pI	putative interneurons
PG	Pólya-Gamma
PPMT-SSD	Point Process Multi-Taper Semi-stationary Spectral Density
PSD	Power Spectral Density
PSTH	Peristimulus Time Histogram
PSTH-SSD	Peristimulus Time Histogram Semi-stationary Spectral Density
REM	Rapid Eye Movement
ROI	Region of Interest
SAM	Sinusoidal Amplitude Modulated
SNR	Signal-to-Noise Ratio
SSD	Semi-stationary Spectral Density
SS-SSD	State-Space Semi-stationary Spectral Density
std	standard deviation
VI	Variational Inference

Chapter 1: Introduction

The brain is undeniably the most complicated, fascinating, and mysterious machinery existing in nature. Understanding how the brain coordinates the activity of billions of neurons to perform cognitive and motor functions has interested humans for many years due to numerous reasons. Studying how the brain's functional connectivity changes under different behavioral tasks and conditions provide us insights into how the neuronal networks in the brain effortlessly adapt to different circumstances. These insights in turn facilitate advancements in artificial intelligence and machine learning, where the ultimate goal is indeed to develop artificial systems replicating the brain's functionality. Understanding the functional architecture of the brain is also crucial for the progression of emerging neural engineering developments, such as brain-machine interfacing. Moreover, investigating the functional connectivity of the brain also facilitates us to better understand, detect and treat neurological diseases. For instance, functional connectivity is significantly related to a variety of diseases such as depression, schizophrenia, and Alzheimer's disease [1, 2]. Thus, methodological advancements for inference of the functional network organizations in the brain have become very consequential.

The ability to investigate the functional architecture of the brain has recently been further enhanced by the rapid progressions in different forms of neural data recording methods. With the advancements of modern neural data acquisition techniques such as two-photon calcium imaging

[3–6], high-speed light-sheet microscopy [6, 7] and high-density electrophysiology recordings [8–10], today we have the ability to concurrently observe spiking activity in large populations of neurons. These different data modalities have indeed facilitated the unique opportunity to investigate network organizations in the brain in diverse settings. Extracting the underlying spectral and temporal network organizations from these data is key to deciphering computations carried out by neuronal networks and understanding how the brain optimizes and coordinates the activity of large populations of neurons to simultaneously perform many diverse tasks.

1.1 Limitations of existing methods

While these modern data recording techniques have opened up many avenues, existing methods for network-level inference from these data have several shortcomings. Firstly, two-photon calcium imaging is currently becoming very popular due to its capacity of recording thousands of neurons simultaneously [3–5]. However, fluorescence observations recorded with two-photon imaging are temporally blurred and noisy surrogates of neuronal spiking activities. Some existing methods assume the fluorescence to be equivalent to spiking events and directly perform network-level analysis on the fluorescence [11–14]. Due to the blurring and typically low Signal-to-Noise ratio of fluorescence activity, this assumption could induce bias in the estimates. On the other hand, even though it is possible to recover spikes first from spike deconvolution and then infer network organizations, since deconvolution methods [15–26] are typically tailored to be optimal to the recovery of accurate first-order statistics (i.e., spike timings) via spatiotemporal priors, they could have detrimental effects on recovering higher-order statistics such as correlations. Thus, developing neural encoding models and inference methods to infer latent neuronal

connectivity directly from calcium imaging data without requiring intermediate spike deconvolution is consequential [27].

Secondly, it is well known that the mapping between the stimuli integration, past activity, and other latent neural covariates and the resulting neuronal spiking activities is non-linear [28]. Most existing methods assume spiking activities to be linearly related to these underlying latent covariates, inducing significant bias on the connectivity estimates [29–32]. Thus, even if the spikes were perfectly deconvolved or directly recorded using electrophysiological methods, most conventional methods for network-level inference from neuronal spiking observations inherently undermine the underlying non-linear dynamics.

Thirdly, most existing approaches for network-level inference perform the estimation in a stage-by-stage fashion. For instance, estimating cross-spectral coupling from spiking observations is typically done by first estimating the latent underlying continuous covariates, and then applying spectral analysis on estimated time-domain signals [29, 33, 34]. It is well-known that time-domain smoothing can induce bias in the subsequent frequency-domain analysis. Existing methods for network-level inference from two-photon calcium imaging observations that first require intermediate spike estimation are another such example. We note that the downstream networks inferred from those methods are highly sensitive to slight variations in estimated spikes; i.e., the connectivity estimates when using different spike deconvolution techniques [15–26] tend to be significantly different from each other. Multi-stage treatments of this nature frequently cause an accumulation of uncertainty at each level of estimation, resulting in the downstream estimates of network organizations being highly unreliable.

Lastly, while it is well known that the brain activity is non-stationary and may exhibit rapid changes corresponding to the brain state or behavioral dynamics [34, 35], existing meth-

ods generally assume the data to be stationary, or piece-wise stationary. Since this assumption renders independent analysis of segmented observation windows, these methods will undermine the stochastic continuity across the entire observation duration. Thus they provide sub-optimal temporal resolution, and may even fail to recover the true underlying brain dynamics.

1.2 Contributions

In this dissertation, we address the aforementioned challenges by introducing Bayesian models and methods to *directly* infer latent spectral and temporal network organizations, from these high dimensional neural data. Bayesian techniques intrinsically have numerous well-known [36] advantages including 1) Performing well under sparse and limited data, which is indeed the typical setting of these neuronal data recordings, 2) Being highly interpretable and understandable due to the so-called “white-box” approach, 3) Providing a principled framework for combining observations with prior knowledge, i.e., facilitating a systematic framework for incorporating all problem-specific information, and 4) Facilitating further statistical analyses such as the construction of confidence intervals.

For the clarity of exposition, we have organized our contributions into three parts. In Part I (Chapters 2,3 and 4), we first introduce Bayesian models and methods for frequency-domain network analysis from neuronal spiking data, and then extend these ideas for frequency-domain inference of human heart rate variability. Next, in Part II (Chapters 5 and 6), we introduce Bayesian models and methods to directly infer signal and noise correlation networks from two-photon calcium imaging data and neuronal spiking data and then present several applications of our methods. Finally, in Part III (Chapter 7), we demonstrate how we applied several spectro-

temporal analysis techniques on whole-brain light-sheet microscopy imaging data, to explore and model swimming-respiration coupling in larval zebrafish.

1.2.1 Part I: Frequency-domain inference

The first part of this dissertation introduces new models and methods for inference of dynamic frequency-domain network organizations from high dimensional neuronal spiking data. Neural oscillations are well known to play a significant role in mediating the cognitive and motor functions of the brain [37–39]. While the advent of high-density electrophysiology recordings [8–10] from multiple locations in the brain has opened a unique window of opportunity to probe these oscillations at the neuronal scale with a high temporal resolution, spectral analysis techniques tailored for these data are in need to uncover the underlying spectro-temporal network organizations. In this part of this dissertation, we introduce methods to infer underlying spectro-temporal connectivity in neuronal ensembles from spiking observations, by integrating techniques such as point process modeling [28], state-space estimation [33] and multitaper spectral estimation [40–44].

Specifically, in Chapter 2 we introduce a semi-stationary multi-taper spectral estimation method tailored for non-stationary multivariate spiking data, called the Point Process Multi-Taper Semi-Stationary Spectral Density (PPMT-SSD) estimator. While the spectral estimation of continuous-time series is well studied, inferring the spectral representation of latent non-stationary processes based on spiking observations is challenging due to the underlying nonlinearities that limit the spectrotemporal resolution of existing methods. In this chapter, we address this issue by developing the PPMT-SSD estimator: a multitaper spectral estimation [40–44]

methodology that can be directly applied to multivariate spiking observations to extract the semi-stationary spectral density [45] of the latent non-stationary processes that govern spiking activity. We explicitly model the underlying non-linearities and non-stationarity of data using point process modeling methods [28], state-space modeling [33] and Priestly’s formulation of semi-stationary spectra [45], and then introduce an inference framework based on expectation-maximization [46]. Further, we establish theoretical bounds on the bias-variance trade-off of the PPMT-SSD estimator and demonstrate that it outperforms existing methods through extensive simulations. Finally, we apply our estimator to infer dynamic spectro-temporal characterizations in two real data experiments. While applications on data recorded from the rat brain [35] reveal new insights on the coupling between two cortical neuronal populations across frequencies during different types of sleep episodes, application to multi-unit spike and local field potential recordings from a human subject undergoing general anesthesia establish previously reported phase-locking effects [34]. The results of this chapter were presented in part at the *2019 IEEE Data Science Workshop* [47], have been published in the *IEEE Transactions on Signal Processing* [48], and an implementation is publicly available at [49].

Building upon PPMT-SSD, in Chapter 3 we then introduce a model and method to infer spectrotemporal directional interactions in a neuronal network from ensemble spiking observations. Granger causality is an increasingly prevalent tool in extracting the functional networks that underlie neural processes. While its time-domain formulation yields useful insights into these functional networks, the inferred Granger causal influences leave the spectral properties of said functional networks ambiguous: this is a question of particular interest when neural processes exhibit oscillatory behavior [37–39]. The frequency-domain formulation of Granger causality proposed by Geweke has addressed the spectral properties of functional links between stationary

processes [50, 51]. Based on Geweke’s method of conditional Granger causality [51], we introduce a framework to derive direct dynamic spectro-temporal causal interactions in a population of neurons from multivariate ensemble spiking observations, using point process modeling [28], state-space estimation [33] and multitaper spectral analysis [40–44]. We demonstrate the utility of our methods through a simulation study, and an application to rat cortical neuronal data [35] acquired during sleep provides insights into the dynamics in directional functional connectivity under different brain states. The results in this chapter were presented at the *2020 Asilomar Conference on Signals, Systems, and Computers* [52].

Next, in Chapter 4 we extend our multitaper spectral analysis methods to develop a precise inference method to characterize spectral components of human heart rate variability (HRV) from electrocardiogram data. The well-known oscillatory nature of HRV has led to the wide usage of its frequency-domain analysis for the identification of features such as the interplay between sympathetic and parasympathetic nervous system activation [53–55]. By posing the problem in a Bayesian framework similar to Chapter 2, we introduce a denoised multitaper spectral estimation method for frequency-domain HRV analysis. In addition, we provide two different methods for deriving the confidence bounds of our estimates. Finally, we demonstrate the applicability of our methods by deriving the denoised multitaper spectral components of experimentally recorded electrocardiogram data (data from Schiff lab, Weill Cornell Medicine, New York City, NY).

1.2.2 Part II: Signal and Noise Correlation inference

The second part of this dissertation introduces new models and methods for signal and noise correlation network inference from two-photon calcium imaging data. Neuronal activity

correlations are key to understanding how groups of neurons collectively encode information [56–66], and are often categorized in two groups: *signal* correlations and *noise* correlations [58, 62, 67–70]. Given two neurons, signal correlation quantifies the similarity of neural responses that are time-locked to a repeated stimulus across trials, whereas noise correlation quantifies the stimulus-independent trial-to-trial variability shared by neural responses that are believed to arise from common latent inputs. While two-photon calcium imaging has created a unique opportunity to record the activity of large populations of neurons [3–6], existing methods for inferring correlations from these data face several challenges. First, the observations of spiking activity produced by two-photon imaging are temporally blurred and noisy. Secondly, even if the spiking data were perfectly recovered via deconvolution, inferring underlying signal and noise correlation networks from binary spiking data is a challenging task due to the non-linear relation of neuronal spiking to endogenous and exogenous inputs [28].

Taking into account these challenges, in Chapter 5 we propose a methodology to explicitly model and directly estimate signal and noise correlations from two-photon calcium imaging observations, without requiring intermediate spike deconvolution. We avoid the need to perform spike deconvolution by integrating techniques from point processes [28] and state-space modeling that explicitly relates the signal and noise correlations to the observed fluorescence traces in a multi-tier model. Thus, we cast signal and noise correlations within a parameter estimation setting. To solve the resulting estimation problem in an efficient fashion, we develop a solution method based on variational inference [71, 72], by combining techniques from Pólya-Gamma augmentation [73] and compressible state-space estimation [20, 74, 75]. Further, we provide theoretical guarantees on the performance of the proposed estimator and demonstrate through extensive simulations that our proposed method has significant performance gains over existing

methods under a variety of settings. Applications of our method on real data recorded from the mouse primary auditory cortex reveal that signal and noise correlation networks are more distinct from each other than reported in previous work, while also exposing novel and distinct spatial patterns in the correlation structures of two different layers in the mouse auditory cortex. We presented these results in part at the *2020 Annual Conference on Information Sciences and Systems (CISS)* [76], *2021 Annual Neuroscience Meeting (SfN)*, and also published in *eLife* [77]. Further, an implementation of our proposed method as well as the data used in this chapter are made publicly available [78, 79].

Next, in Chapter 6 we apply our signal and noise correlation inference framework to investigate network-level changes in the auditory cortex under two different experiments. Firstly, we present a comparison between correlation networks inferred by applying our methods to two-photon calcium imaging data recorded from the mouse primary auditory cortex, under two different stimulus conditions: pure-tone versus two-tone presentation (data from Kanold Lab, Johns Hopkins University, Baltimore, MD). We uncover distinct spatial patterns in the signal and noise correlation networks under each stimulus setting, providing new insights into how the network organizations change under different stimulus conditions. Next, we illustrate how our correlation analysis methods provide insights into network-level changes in the course of perceptual learning, by applying our method to electrophysiological data recorded from the auditory cortex of gerbils under auditory task performance [80]. Through this analysis, we demonstrate that while our method was originally developed to be applied to two-photon calcium imaging data, it can easily be adapted to infer signal and noise correlations even from neuronal spiking data.

1.2.3 Part III: Zebrafish respiration-locomotion network coordination

In the last part of this dissertation (Chapter 7), we investigate the respiratory circuitry and the swimming-respiration coupling phenomena in larval zebrafish by applying several spectrotemporal analyses techniques, on whole-brain light-sheet microscopy imaging data (data from Ahrens Lab, Janelia Research Campus, Ashburn, VA). Firstly, using multitaper spectrotemporal analysis of whole-brain imaging data and simultaneously recorded electrophysiological data, we discover brain regions that are synchronized with the respiratory rhythm. Then, using cross-spectral density phase analysis, we introduce a method to categorize the respiratory neurons based on their distinct phases relative to the breathing rhythm. Using several spectrotemporal analysis methods, we also demonstrate that zebrafish swimming is phase-locked to breathing and is gated by the respiratory rhythm. Then, through the analysis of whole-brain imaging data recorded under optogenetic stimulation and two-photon ablations, we distinguish the brain regions that are key for this swimming-respiration coupling. Finally, using the Izhikevich model [81] for spiking neurons, we develop a circuit model that replicates the swimming-respiration coupling behavior observed in real data, providing new insights into the possible underlying neural circuitry.

1.3 Broader Impact

Finally, we close this thesis in Chapter 8 by summarizing our contributions, discussing the limitations of our approaches, and outlining potential future directions. Firstly, we discuss the challenges that could potentially arise when using our methods due to factors such as modeling assumptions, inexact inverse solutions, computational complexity, and the biological interpretability of the inferred functional networks. Then, we discuss several possible extensions of

our work and outline future directions of research that are brought to light as a byproduct of this dissertation. It is also noteworthy that the challenges we addressed through this dissertation are shared among many other disciplines such as econometrics, climatology, seismology, social network analysis, gene regulatory networks, astronomical imaging, and communication systems. Thus, the developed ideas and algorithms in this thesis have potential applications in other domains outside neuroscience and can be readily adopted thanks to the plug-and-play nature of our algorithms. Furthermore, in the spirit of reproducibility, we have archived implementations of our algorithms as open source repositories on <https://github.com/Anuththara-Rupasinghe>.

Part I

Frequency-domain inference

Chapter 2: Multitaper Analysis of Semi-Stationary Spectra from Multivariate Neuronal Spiking Observations

Neural oscillations are known to play a significant role in mediating the cognitive and motor functions of the brain [37–39]. The advent of high-density electrophysiology recordings [8–10] from multiple locations in the brain has opened a unique window of opportunity to probe these oscillations at the neuronal scale. To exploit such experimental data for inferring the mechanisms of brain function, spectral analysis techniques tailored for such neuronal spiking data are required [82].

While there exists a wide range of mathematical models for capturing neuronal spiking dynamics [83], statistical approaches based on the theory of point processes or generalized linear models have gained popularity in recent years [28, 82, 84–86]. These models relate the statistics of the neural stimulus (input) and response (output) in a model-based fashion to facilitate further analyses such as estimation, prediction, and decoding. The input may consist of explicit (e.g., sensory stimuli) and latent (e.g., internal neural processes) portions. Existing methods for spectral analysis under the point process framework typically consider univariate spiking observations and take the input to be a latent second-order stationary process during the observation period [29, 33, 86, 87]. Despite their relative success in capturing the spectral properties of spiking activity, these methods have several shortcomings to be addressed.

First of all, some of the existing methods first estimate the latent input processes in the time domain, followed by spectral analysis of these time-domain estimates [29,33,34]. The time-domain estimation techniques vary from averaging across different trials [29] to estimating the latent process using Kalman filtering and smoothing [33,34]. Due to the time-domain smoothing procedures used by these methods, the resulting power spectral density (PSD) estimates undergo distortion in the spectral domain.

Secondly, some of the existing methods that overcome the aforementioned challenge and aim at directly estimating the PSD without recourse to time-domain smoothing can only operate on univariate time-series [86,87]. As such, they are not able to capture the cross-spectral couplings of multiple spike trains, which is crucial in understanding the underlying circuit mechanisms of neural activity.

Finally, it is known that the brain oscillations that underlie neuronal spiking are non-stationary and may exhibit rapid changes corresponding to the brain state or behavioral dynamics [34,88]. Existing methods apply sliding window adaptations of stationary spectral estimation methods to capture these changes in the spectral domain, and thus provide sub-optimal spectrotemporal resolutions due to the well-known limitations of sliding window analysis. Indeed, non-stationary time-series analysis has been well studied for multivariate continuous signals and various methods have been proposed to quantify the energy-frequency-time distributions [45,89–97]. One notable example is the evolutionary power spectral characterization [45,94], which defines a time-varying spectral density matrix in order to quantify the local spectral energy distributions at each instant of time for a multivariate oscillatory process.

In light of the foregoing challenges, a unified framework for capturing the non-stationary spectral properties of multivariate neuronal spiking data with a high spectrotemporal resolution

is lacking, but highly desired due to the emerging demands of modern neuronal data analysis. In this chapter, we close this gap by developing a framework under the point process framework to estimate the semi-stationary spectral density (SSD) matrix of a multivariate non-stationary latent process, given spiking observations.

To address the first two challenges, we model the spiking observations as multiple realizations of Bernoulli processes with logistic links to the latent multivariate process. This model is motivated by its univariate counterpart, which has been previously utilized in the statistical analysis of spiking data [28, 98–100]. We then pose the problem of spectral estimation within a multitaper framework [40–44] to directly estimate the SSD from multivariate spiking observations, and thus obviating the need for intermediate time-domain smoothing.

To address the third challenge, we adopt the working hypothesis that the latent processes under study conform to the commonly-used regularity assumption in evolutionary spectral estimation known as semi-stationarity. We employ a state-space model to characterize the dynamics of the semi-stationary spectra [96] and derive an Expectation-Maximization (EM) algorithm for efficiently computing the maximum a posteriori (MAP) estimate of the SSD matrix. We also establish theoretical bounds on the bias-variance performance of our proposed estimator and recover the favorable asymptotic properties of the classical multitaper framework.

We compare the performance of our proposed method to existing techniques through simulation studies and application to experimentally-recorded neuronal data. We present simulated case studies using non-stationary multivariate auto-regressive processes, whose dynamics are inspired by neural oscillations. These studies reveal that the proposed method outperforms two of the widely used methods for deriving spectral representations from spiking data. Next, we apply our proposed estimator to multi-unit recordings from the rat cortical neurons during sleep

[35, 101]. Our proposed method corroborates existing results on spectrotemporal dynamics of brain state transitions obtained by local field potential analysis, by directly inferring said dynamics from spike recordings at the neuronal scale and high spectrotemporal resolutions. Finally, we apply our proposed estimator to multi-unit spike and local field potential (LFP) recordings from a human subject undergoing general anesthesia [34, 86]. Our proposed estimates establish existing hypotheses on the relation between the LFP signals and spiking dynamics under general anesthesia [34, 86], validating its utility in obtaining reliable spectral representations of neural spiking data.

In closing, our contributions in this chapter can be summarized as follows: 1) Developing an algorithm for estimating the SSD matrix directly from spiking observations with no recourse to time-domain smoothing procedures, 2) Enhancing the spectrotemporal resolution by integrating the multi-tapering and state-space modeling frameworks, and 3) Establishing theoretical guarantees on the bias-variance performance of our proposed estimator.

2.1 Problem Formulation

Let $N(t)$ and $H(t)$ denote the point process representing the number of spikes and spiking history of a neuron in $[0, t)$, respectively, where $t \in [0, T]$ and T denotes the observation duration. The Conditional Intensity Function (CIF) [102] of a point process $N(t)$ is defined as:

$$\lambda(t|H_t) := \lim_{\Delta \rightarrow 0} \frac{P[N(t + \Delta) - N(t) = 1|H_t]}{\Delta}. \quad (2.1)$$

To discretize the continuous process, we consider time bins of length Δ , small enough that the probability of having two or more spikes in an interval of length Δ is negligible. A choice

of $\Delta \sim 1$ ms renders this assumption valid, due to the absolute refractory period property of biological neurons [28]. Considering the binary nature of spiking data, this discretized point process can be modeled by a Bernoulli process with success probability $\lambda_k := \lambda(k\Delta|H_k)\Delta$, for $1 \leq k \leq K$, where $K := T/\Delta$ is an integer (with no loss of generality). We refer to λ_k as CIF hereafter for brevity.

In a similar fashion, we consider spiking observations from an ensemble of J neurons, with CIFs $\{\lambda_{k,j}\}_{k=1}^K$, for $j = 1, 2, \dots, J$. Additionally, for each neuron, we assume that L independent realizations of spiking activity are observed. Thus, we represent the l^{th} realization corresponding to the CIF $\lambda_{k,j}$ by $n_{k,j}^{(l)}$. Further, we model $\lambda_{k,j}$ by a logistic link to the underlying latent random process, $x_{k,j}$. The logistic function is often referred to as the *canonical link* function for a Bernoulli process in the generalized linear model and point process frameworks, and significantly facilitates procedures such as regression and maximum likelihood estimation [28]. Accordingly, we have

$$n_{k,j}^{(l)} \sim \text{Bernoulli}(\lambda_{k,j}), \quad (2.2)$$

where $\lambda_{k,j} = \ell(x_{k,j}) := 1/(1 + \exp(-x_{k,j}))$, with $\ell(\cdot)$ representing the logistic function. Our goal is to estimate the time-varying power spectral density of each process $x_{k,j}$, for $1 \leq j \leq J$, as well as the time-varying cross spectra between each pair of processes. To that end, we assume that the process $x_{k,j}$ is a *semi-stationary process* in the sense of Priestly [45], which implies a representation of the form,

$$x_{k,j} - \mu_{k,j} = \int_{-\pi}^{\pi} e^{ik\omega} A_{k,j}(\omega) dZ_j(\omega), \quad (2.3)$$

where $\mu_{k,j}$ is the mean of $x_{k,j}$, $A_{k,j}(\omega)$ is the time-varying amplitude function and $dZ_j(\omega)$ is an orthogonal increment process. The semi-stationarity assumption restricts the rate of change of $A_{k,j}(\omega)$ in the sense that for each fixed ω , the Fourier transform of $A_{k,j}(\omega)$ with respect to k , is highly concentrated around zero [91]. Note that this assumption is the basis for the widely-used short-time periodogram analysis as well as constructing estimators of the evolutionary spectra [45]. Given that the dynamics of the neuronal processes underlying spiking activity are typically much slower than the sampling rate, the assumption of semi-stationarity is commonly adopted in neural data analysis.

To obtain a discrete-parameter harmonic process, we approximate $Z_j(\omega)$ by a jump process over N frequency bins [86], and thereby replace $dZ_j(\omega_n)$ with $\frac{\pi}{N}(a_{j,n} + ib_{j,n})$, at $\omega_n = n\pi/N$, $1 \leq n \leq N-1$, where $a_{j,n}$ and $b_{j,n}$ are real-valued random variables. Since the random process is real, using the symmetry $Z_j(\omega) = Z_j(-\omega)$, we express the discretization of Eq. (2.3) as

$$x_{k,j} = \mu_{k,j} + \frac{2\pi}{N} \sum_{n=1}^{N-1} A_{k,j}(\omega_n) (a_{j,n} \cos(\omega_n k) - b_{j,n} \sin(\omega_n k)).$$

To explicitly utilize the semi-stationarity assumption, we consider a piece-wise constant approximation to $A_{k,j}(\omega)$, implying that the J -variate random process $\{x_{k,j}\}_{k,j=1}^{K,J}$ is jointly stationary in windows of small enough length W [96], and divide the total data duration K into M non-overlapping segments of length W , with $K = MW$. We thus have:

$$x_{k,j} = \mu_{m,j} + \frac{2\pi}{N} \sum_{n=1}^{N-1} (p_{m,j,n} \cos(\omega_n k) - q_{m,j,n} \sin(\omega_n k)),$$

for $(m-1)W+1 \leq k \leq mW$, $1 \leq m \leq M$, where $p_{m,j,n}$ and $q_{m,j,n}$ are real-valued random variables. Defining $\mathbf{X}_{m,j} := [x_{(m-1)W+1,j}, x_{(m-1)W+2,j}, \dots, x_{mW,j}]^\top$, $\mathbf{v}_{m,j} := [\frac{N}{2\pi}\mu_{m,j}, p_{m,j,1}, q_{m,j,1},$

$\cdots, p_{m,j,N-1}, q_{m,j,N-1}]^\top$, we have the compact representation $\mathbf{X}_{m,j} = \mathbf{A}_m \mathbf{v}_{m,j}$, where $\mathbf{A}_m$ is a $W \times (2N-1)$ matrix with the first column equal to $\frac{2\pi}{N} \mathbf{1}$, $(\mathbf{A}_m)_{w,2n} = \frac{2\pi}{N} \cos\left(\frac{n(m-1)W+w}{N}\pi\right)$ and $(\mathbf{A}_m)_{w,2n+1} = -\frac{2\pi}{N} \sin\left(\frac{n(m-1)W+w}{N}\pi\right)$, for $w = 1, 2, \dots, W$ and $n = 1, 2, \dots, N-1$.

The evolutionary spectrum of an oscillatory process $x_{k,j}$, at frequency ω_n , is defined as $\psi_{k,j}(\omega_n) d\omega_n = |A_{k,j}(\omega_n)|^2 \mathbb{E}[|dZ_j(\omega_n)|^2]$ [45]. Moreover, for a J -variate stationary vector-valued orthogonal increment process $\mathbf{Z}(\omega_n) := [Z_1(\omega_n), \dots, Z_J(\omega_n)]^\top$, the spectral density matrix has been expressed as $\Psi(\omega_n) d\omega_n := \mathbb{E}[d\mathbf{Z}(\omega_n) d\mathbf{Z}(\omega_n)^H]$ [103]. Extending this notion to the semi-stationary spectra, we formulate the SSD matrix as:

$$\Psi_m(\omega_n) = \frac{\pi}{N} \mathbb{E}[(\mathbf{p}_{m,n} + i\mathbf{q}_{m,n})(\mathbf{p}_{m,n} - i\mathbf{q}_{m,n})^\top], \quad (2.4)$$

where $\mathbf{p}_{m,n} = [p_{m,1,n}, p_{m,2,n}, \dots, p_{m,J,n}]^\top$ and $\mathbf{q}_{m,n} = [q_{m,1,n}, q_{m,2,n}, \dots, q_{m,J,n}]^\top$.

For brevity, we concatenate the vectors $\mathbf{p}_{m,n}$ and $\mathbf{q}_{m,n}$ as $\mathbf{w}_{m,n} := [\mathbf{p}_{m,n}^\top, \mathbf{q}_{m,n}^\top]^\top$, $1 \leq n \leq N-1$, and define the vector $\mathbf{w}_m := [\mathbf{w}_{m,0}^\top, \mathbf{w}_{m,1}^\top, \dots, \mathbf{w}_{m,N-1}^\top]^\top$, with $\mathbf{w}_{m,0} := [\mu_{m,1}, \mu_{m,2}, \dots, \mu_{m,J}]^\top$. Thus, we see that the SSD matrix in Eq. (2.4), can be determined if $\mathbb{E}[\mathbf{w}_{m,n} \mathbf{w}_{m,n}^\top]$ is estimated for $1 \leq n \leq N-1$. As such, the task of estimating the semi-stationary spectra of the J -variate random process is reduced to computing $\mathbb{E}[\mathbf{w}_m \mathbf{w}_m^\top]$, for $m = 1, 2, \dots, M$, given the ensemble spiking data $\mathcal{D} = \{n_{k,j}^{(l)}\}_{l,k,j=1}^{L,K,J}$. To this end, we first develop a multitaper version of the foregoing SSD estimator to enhance spectral resolution, and then complement it with a suitable estimation algorithm within a Bayesian setting.

2.2 Multitaper Estimation of the SSD Matrix

2.2.1 The Proposed Multitaper Framework

The estimate $\Psi_m(\omega_n)$ in Eq. (2.4) is based on conventional Fourier analysis, which has its inherent bias and variance limitations. Multitapering is a technique widely used in power spectral estimation of stationary random processes, to overcome these limitations [40]. Given that the latent process $x_{k,j}$ is not directly observable, forming its multitaper estimate is not straightforward. We will indeed develop multitaper spectral estimates of $x_{k,j}$ based on its indirect observations $n_{k,j}^{(l)}$.

For a stationary continuous process, $x_1, x_2, \dots, x_K$ the classical multitaper estimator is given by:

$$S^{\text{mt}}(\omega) = \frac{1}{P} \sum_{p=1}^P |y^{(p)}(\omega)|^2 \quad (2.5)$$

with $y^{(p)}(\omega) = \sum_{k=1}^K \nu_k^{(p)} x_k e^{-i\omega k}$, where $\nu_k^{(p)}$ is the k^{th} time sample of the p^{th} discrete prolate spheroidal sequence (dpss) [41], for $1 \leq p \leq P$. The dpss tapers are a set of orthogonal tapers that maximally concentrate their spectral power within a resolution bandwidth of $[-\frac{\xi f_s}{K}, \frac{\xi f_s}{K}]$, for some positive constant ξ . Tapering with the dpss sequences results in approximately unbiased spectral estimates with minimal spectral leakage [42–44], and averaging the eigen-spectra $|y^{(p)}(\omega)|^2$, significantly reduces the variance of the estimated PSD [40, 42, 43].

While originally designed for univariate time series, multitapering has been likewise extended to multivariate time series [104]. Given a second order jointly stationary J -dimensional random process $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K$, where, $\mathbf{x}_k = [x_{k,1}, x_{k,2}, \dots, x_{k,J}]$, the multitaper cross-spectral

estimate between the r^{th} process and the t^{th} process ($r, t \in \{1, 2, \dots, J\}$) has been defined as,

$$S_{r,t}^{\text{mt}}(\omega) = \frac{1}{P} \sum_{p=1}^P y_r^{(p)}(\omega) (y_t^{(p)}(\omega))^* \quad (2.6)$$

where $y_r^{(p)}(\omega) := \sum_{k=1}^K \nu_k^{(p)} x_{k,r} e^{-i\omega k}$.

These classical estimators are not directly applicable in our framework, since the latent processes $x_{k,j}$ are not directly observable. Our objective is to estimate the semi-stationary spectra $\Psi_m(\omega_n)$ directly from the spiking observations, avoiding intermediate time-domain estimation of the latent processes. Thus, we need to develop an alternative approach that incorporates the effect of indirect observation of $x_{k,j}$ through spiking activity. The data log-likelihood in our model,

$$\log f(\mathcal{D} | \{x_{k,j}\}_{k,j=1}^{K,J}) = \sum_{k,j=1}^{K,J} L \left\{ \bar{n}_{k,j} x_{k,j} - \log(1 + \exp(x_{k,j})) \right\}, \quad (2.7)$$

depends on the observations only through the ensemble average, $\bar{n}_{k,j} := \frac{1}{L} \sum_{l=1}^L n_{k,j}^{(l)}$. Hence, $\bar{n}_{k,j}$ is a sufficient statistic. Thus, we need to characterize the effect of tapering the latent time series $x_{k,j}$ on the observed ensemble average spiking $\bar{n}_{k,j}$. Note that tapering the observed process $\bar{n}_{k,j}$, i.e., $\nu_{k,j}^{(p)} \bar{n}_{k,j}$, is not admissible as the dpss tapers take negative values for $p > 1$, whereas spiking observations are non-negative. In addition, $\nu_{k,j}^{(p)} \bar{n}_{k,j}$ does not relate to the desired tapered latent process $\nu_{k,j}^{(p)} x_{k,j}$ in a straightforward fashion.

Consider the log-likelihood $\log f(\{n_{k,j}^{(l)}\}_{l=1}^L | x_{k,j}) = L \left\{ \bar{n}_{k,j} x_{k,j} - \log(1 + \exp(x_{k,j})) \right\}$, in

Eq. (2.7). Noting that

$$\left. \frac{\partial \log f(\{n_{k,j}^{(l)}\}_{l=1}^L | x_{k,j})}{\partial x_{k,j}} \right|_{x_{k,j} = \ell^{-1}(\bar{n}_{k,j})} = 0,$$

the logit of $\bar{n}_{k,j}$ defined by $\ell^{-1}(\bar{n}_{k,j}) := \log(\bar{n}_{k,j}/(1 - \bar{n}_{k,j}))$ is the maximum likelihood estimator of $x_{k,j}$. Accordingly, we use $\ell^{-1}(\bar{n}_{k,j})$ as the estimator of $x_{k,j}$. Furthermore, $\bar{n}_{k,j}$ converges almost surely to $\ell(x_{k,j})$, by the strong law of large numbers. To address the foregoing challenge, we propose

$$(\bar{n}_{k,j})^{(p)} := \ell(\nu_{k \bmod W+1}^{(p)} \ell^{-1}(\bar{n}_{k,j})), \quad (2.8)$$

to be the estimator of the ensemble average, had the latent process been tapered by the p^{th} dpss taper, i.e., had the spiking activity been governed by $\nu_{k \bmod W+1}^{(p)} x_{k,j}$. Note that when $\bar{n}_{k,j} = 0$ or $\bar{n}_{k,j} = 1$, the function $\ell^{-1}(\bar{n}_{k,j})$ is not defined. In such cases, we directly estimate $(\bar{n}_{k,j})^{(p)}$ by $\bar{n}_{k,j}$. The dpss tapers are sandwiched between $\ell(\cdot)$ and $\ell^{-1}(\cdot)$ in Eq. (2.8), which makes it an admissible non-negative ensemble average and is key to developing the multitaper estimator of the spectra of $x_{k,j}$. If one would switch the order of evaluating $\ell(\cdot)$ and tapering in Eq. (2.8), the process $(\bar{n}_{k,j})^{(p)}$ would reduce to $\nu_{k \bmod W+1}^{(p)} \bar{n}_{k,j}$, which is inadmissible for $p > 1$.

The problem is now reduced to computing the semi-stationary spectra of $\{x_{k,j}\}$ corresponding to each dpss taper, $\Psi_m^{(p)}(\omega_n)$, from $(\bar{n}_{k,j})^{(p)}$, for $p = 1, 2, \dots, P$. Then, the multitaper SSD estimate can be formed by averaging the P eigen-spectral estimates, $\Psi_m^{\text{mt}}(\omega_n) = \frac{1}{P} \sum_{p=1}^P \widehat{\Psi}_m^{(p)}(\omega_n)$. In the next subsection, we first develop a method for estimating the SSD matrix of the untapered latent process $\{x_{k,j}\}$ from $\{\bar{n}_{k,j}\}$, and then extend it to the case of P tapers by replacing $\{\bar{n}_{k,j}\}$ with $\{(\bar{n}_{k,j})^{(p)}\}$, which gives the eigen-spectral SSD estimate $\widehat{\Psi}_m^{(p)}(\omega_n)$, for $p = 1, 2, \dots, P$.

2.2.2 MAP Estimation of the Parameters via the EM Algorithm

In order to capture the evolution of the SSD $\Psi_m(\omega_n)$, we impose a stochastic continuity constraint on the random variables $\mathbf{w}_m$ using the following state-space model:

$$\mathbf{w}_m = \Phi \mathbf{w}_{m-1} + \boldsymbol{\eta}_m, \quad (2.9)$$

where the state transition matrix Φ is a constant matrix and $\boldsymbol{\eta}_m \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_m)$. We consider the special case where Φ takes the form $\alpha \mathbf{I}$, for simplicity. Nevertheless, Φ can also be estimated from data within the same Expectation-Maximization [46] framework that follows next.

The parameters $\boldsymbol{\theta} := \{\mathbf{Q}_m, 1 \leq m \leq M\}$ need to be estimated from the observations $\mathcal{D} = \{n_{k,j}^{(l)}\}_{k,j,l=1}^{K,J,L}$. Considering $\mathbf{w} = \{\mathbf{w}_m, 1 \leq m \leq M\}$ to be the hidden variable, we aim at recovering $\boldsymbol{\theta}$ via MAP estimation. We consider a diagonal covariance matrix $\mathbf{Q}_m$, whose i^{th} diagonal entry is denoted by $Q_{m,i}$. This assumption is due to the orthogonality of the increment process $dZ_j(\omega)$ in Eq. (2.3) and eliminates undesired spectral coupling. In addition, it reduces the number of parameters to be estimated, thereby improving the convergence of the estimator. Given the typical sparsity of neuronal spiking in time and the low number of observed realizations (L), an appropriate prior distribution on $\mathbf{Q}_m$ helps in reducing the estimation variance. Thus, we assume $\mathbf{Q}_m$ to be independent and identically distributed for $1 \leq m \leq M$, with a density:

$$f(\mathbf{Q}_m) \propto \exp \left(-\rho \sum_{j=1}^{2J} \sum_{n=1}^{N-2} \left(\log(Q_{m,J(2n-1)+j}) - \log(Q_{m,J(2n+1)+j}) \right)^2 \right).$$

This prior distribution encourages continuity in log scale of the spectral estimates corre-

sponding to adjacent frequency bins, and can be controlled by appropriately selecting the hyper-parameter ρ . MAP estimation corresponds to maximizing the log-likelihood of the complete data $(\mathcal{D}, \mathbf{w})$, given by:

$$\begin{aligned} \log f(\mathcal{D}, \mathbf{w}, \boldsymbol{\theta}) &= \log f_{\mathbf{w}|\boldsymbol{\theta}}(\mathbf{w}|\boldsymbol{\theta}) + \log f(\boldsymbol{\theta}) + C_1 \\ &= -\frac{1}{2} \sum_{m=1}^M \left\{ \log |\mathbf{Q}_m| + (\mathbf{w}_m - \Phi \mathbf{w}_{m-1})^\top \mathbf{Q}_m^{-1} (\mathbf{w}_m - \Phi \mathbf{w}_{m-1}) \right\} \\ &\quad - \rho \sum_{m=1}^M \sum_{j=1}^{2J} \sum_{n=1}^{N-2} \left(\log(Q_{m,J(2n-1)+j}) - \log(Q_{m,J(2n+1)+j}) \right)^2 + C_2, \end{aligned}$$

where C_1 and C_2 represent terms that do not depend on $\boldsymbol{\theta}$. We next construct the EM algorithm for solving this MAP estimation problem:

2.2.2.1 E-Step

Suppose that the current estimate of $\boldsymbol{\theta}$ at the r^{th} iteration is given by $\hat{\boldsymbol{\theta}}^{(r)}$. Then, the Q-function

$$Q^{(r)} := \mathbb{E}[\log f(\mathcal{D}, \mathbf{w}, \boldsymbol{\theta}) | \mathcal{D}, \hat{\boldsymbol{\theta}}^{(r)}] \quad (2.10)$$

can be evaluated if the conditional expectations

$$\begin{aligned} \mathbf{w}_{m|M} &:= \mathbb{E}[\mathbf{w}_m | \mathcal{D}, \hat{\boldsymbol{\theta}}^{(r)}], \\ \boldsymbol{\Sigma}_{m|M} &:= \mathbb{E}[(\mathbf{w}_m - \mathbf{w}_{m|M})(\mathbf{w}_m - \mathbf{w}_{m|M})^\top | \mathcal{D}, \hat{\boldsymbol{\theta}}^{(r)}], \\ \boldsymbol{\Sigma}_{m,m-1|M} &:= \mathbb{E}[(\mathbf{w}_m - \mathbf{w}_{m|M})(\mathbf{w}_{m-1} - \mathbf{w}_{m-1|M})^\top | \mathcal{D}, \hat{\boldsymbol{\theta}}^{(r)}], \end{aligned}$$

are known. To compute these conditional expectations we utilize the Fixed Interval Smoothing [74] and the Covariance Smoothing [105] algorithms. However, considering that the forward model is not Gaussian in our formulation, we cannot directly use these algorithms to estimate $\mathbf{w}_{m|m}$ and $\Sigma_{m|m}$.

The density $f(\{\mathbf{w}_s\}_{s=1}^m | \mathcal{D}_1^m, \hat{\boldsymbol{\theta}}^{(r)})$, with $\mathcal{D}_1^m = \{n_{k,j}^{(l)}\}_{k,j,l=1}^{mW,J,L}$, is proportional to the product of the two densities $f(\mathcal{D}_1^m | \{\mathbf{w}_s\}_{s=1}^m, \hat{\boldsymbol{\theta}}^{(r)})$ and $f(\{\mathbf{w}_s\}_{s=1}^m | \hat{\boldsymbol{\theta}}^{(r)})$, which are Binomial and Gaussian distributed, respectively. Utilizing the unimodality of the density of $\{\mathbf{w}_s\}_{s=1}^m | \mathcal{D}_1^m, \hat{\boldsymbol{\theta}}^{(r)}$, we approximate it by a multivariate Gaussian, and derive the mean of the distribution, $\mathbf{w}_{m|m}^{(r)}$ by the mode of $\log f(\{\mathbf{w}_s\}_{s=1}^m | \mathcal{D}_1^m, \hat{\boldsymbol{\theta}}^{(r)})$:

$$\begin{aligned} \underset{\mathbf{w}_m}{\operatorname{argmax}} \quad & \sum_{j,s,w=1}^{J,m,W} L \left\{ \bar{n}_{(s-1)W+w,j} (\mathbf{A}_s \mathbf{v}_{s,j})_w - \log(1 + \exp(\mathbf{A}_s \mathbf{v}_{s,j})_w) \right\} \\ & - \frac{1}{2} \sum_{s=1}^m \left\{ \log |\mathbf{Q}_s^{(r)}| + (\mathbf{w}_s - \Phi \mathbf{w}_{s-1})^\top (\mathbf{Q}_s^{(r)})^{-1} (\mathbf{w}_s - \Phi \mathbf{w}_{s-1}) \right\}, \end{aligned} \quad (2.11)$$

and its covariance, $\Sigma_{m|m}^{(r)}$ by the negative of the inverse Hessian of $\log f(\{\mathbf{V}_s\}_{s=1}^m | \mathcal{D}_1^m, \hat{\boldsymbol{\theta}}^{(r)})$. Observing that the objective function is a combination of convex functions and is differentiable, we perform the above optimization for $\mathbf{w}_{m|m}^{(r)}$ using the Newton-Raphson method. Concurrently, we estimate $\Sigma_{m|m}^{(r)}$ using the Hessian matrix of the objective.

2.2.2.2 M-Step

Due to the separability of the Q-function $\mathcal{Q}^{(r)}$ in terms of $\mathbf{Q}_m$'s, we can update $\mathbf{Q}_m$'s independently as:

$$\mathbf{Q}_m^{(r+1)} = \underset{\mathbf{Q}_m}{\operatorname{argmax}} \quad \mathcal{Q}^{(r)}, \quad 1 \leq m \leq M. \quad (2.12)$$

Observing that $Q^{(r)}$ is differentiable in $\mathbf{Q}_m$, we employ the multivariate Newton-Raphson algorithm to perform this maximization and derive the updates for $\mathbf{Q}_m$, $1 \leq m \leq M$.

An implementation of this EM procedure is outlined in Algorithm 2.1. Following convergence, we use the final estimates $\mathbf{w}_{m|M}$ and $\Sigma_{m|M}$ to evaluate the SSD matrix $\hat{\Psi}_m(\omega_n)$. As mentioned in Section 2.2.1, the same EM procedure can be carried out for $\{\bar{n}_{k,j}^{(p)}\}$, for $p = 1, 2, \dots, P$ in order to estimate the multivariate eigen-spectra $\hat{\Psi}_m^{(p)}(\omega_n)$. Finally, the multitaper spectral density matrix is formed by averaging the eigen-spectral estimates as outlined in Algorithm 2.2. We refer to our proposed algorithm as the Point Process Multi-Taper Semi-stationary Spectral Density (PPMT-SSD) estimator.

2.3 Theoretical Analysis

In this section, we derive bounds on the bias and variance of the proposed estimator. We first briefly review the corresponding bounds for the classical multitaper estimator (Eq. (2.5)). As proven in [106], the bias and variance of the multitaper estimate of a stationary process $x_1, x_2, \dots, x_K$ with a uniformly continuous PSD $S(\omega)$ are bounded as follows:

$$|\text{bias}(S^{\text{mt}}(\omega))| \leq (\sup_{\omega} S(\omega)) \left\{ 1 - \frac{1}{P} \sum_{p=1}^P c_p \right\} + o(1), \quad (2.13)$$

$$\text{Var}(S^{\text{mt}}(\omega)) = (1 + \beta(\omega)) \frac{1}{P^2} \sum_{p=1}^P c_p^2 S(\omega)^2 + \mathcal{O} \left(\frac{1}{P} \sum_{p=1}^P (1 - c_p) \right) + o(1), \quad (2.14)$$

as $K \rightarrow \infty$. Here c_p is the eigenvalue corresponding to the taper $\nu^{(p)}$ and $\beta(\omega) = 0$ if $\frac{\omega}{2\pi} \neq 0, 1/2 \bmod 1$ and is equal to 1 otherwise. It is evident that the multitaper estimator $\hat{S}^{\text{mt}}(\omega)$ is asymptotically unbiased.

Algorithm 2.1 Estimation of the Semi-stationary Spectral Density Matrix via the EM Algorithm

Inputs: Ensemble averages of the spiking observations $\{\bar{n}_{k,j}\}_{k,j=1}^{K,J}$, hyper-parameters ρ, ζ and α , maximum number of EM iterations $R_{\max}$.

Outputs: Estimates of the SSD matrices $\hat{\Psi}_m(\omega_n)$ for $1 \leq m \leq M, 1 \leq j \leq N - 1$

Initialization: Initial choice of $\mathbf{Q}_m^{(0)} = \zeta \mathbf{I}$, $\mathbf{w}_{0|0} = \mathbf{0}$, $\Sigma_{0|0} = \mathbf{0}$, $r = 1$.

1: **for** $r \leq R_{\max}$ **do**

2: Forward filter, for $m = 1, 2, \dots, M$:

$$\mathbf{w}_{m|m-1} = \Phi \mathbf{w}_{m-1|m-1}$$

$$\Sigma_{m|m-1} = \Phi \Sigma_{m-1|m-1} \Phi^\top + \mathbf{Q}_m^{(r)}$$

 Compute $\mathbf{w}_{m|m}$ and $\Sigma_{m|m}$ using Newton's method as described in Eq. (2.11).

3: Backward smoother, for $m = M - 1, M - 2, \dots, 1$:

$$\mathbf{B}_m = \Sigma_{m|m} \Phi^\top \Sigma_{m+1|m}^{-1}$$

$$\mathbf{w}_{m|M} = \mathbf{w}_{m|m} + \mathbf{B}_m (\mathbf{w}_{m+1|M} - \mathbf{w}_{m+1|m})$$

$$\Sigma_{m|M} = \Sigma_{m|m} + \mathbf{B}_m (\Sigma_{m+1|M} - \Sigma_{m+1|m}) \mathbf{B}_m^\top$$

4: Covariance smoother, for $m = M - 1, M - 2, \dots, 1$:

$$\Sigma_{m,m-1|M} = \Sigma_{m|M}^\top \mathbf{B}_{m-1}^\top$$

5: Update the $\mathbf{Q}_m$'s independently, for $m = 1, 2, \dots, M$ using the multivariate Newton-Raphson method to solve:

$$\mathbf{Q}_m^{(r+1)} = \underset{\mathbf{Q}_m}{\operatorname{argmax}} \quad Q^{(r)}$$

6: Set $r \leftarrow r + 1$

7: **end for**

8: **for** $1 \leq m \leq M$ **do**

9: $\mathbf{R}_m = \Sigma_{m|M} + \mathbf{w}_{m|M} \mathbf{w}_{m|M}^\top$

10: **for** $1 \leq n \leq N - 1$ **do**

$$\mathbf{R}_m^n = (\mathbf{R}_m)_{(J(2n-1)+1:J(2n+1), J(2n-1)+1:J(2n+1))}$$

$$\begin{aligned} \hat{\Psi}_m(\omega_n) = & \frac{\pi}{N} \left\{ \mathbf{R}_{m(1:J,1:J)}^n + \mathbf{R}_{m(J+1:2J,J+1:2J)}^n \right. \\ & \left. + i \left(\mathbf{R}_{m(J+1:2J,1:J)}^n - \mathbf{R}_{m(1:J,J+1:2J)}^n \right) \right\} \end{aligned}$$

11: **end for**

12: **end for**

13: Return $\hat{\Psi}_m(\omega_n)$ for $1 \leq m \leq M, 1 \leq n \leq N - 1$.

Algorithm 2.2 Estimation of the Multi-Taper Semi-stationary Spectral Density Matrix (PPMT-SSD)

Inputs: Collection of ensemble averages of the observations $\{\bar{n}_{k,j}\}_{k,j=1}^{K,J}$, the set of P dpss tapers of length W $\{\nu_w^{(p)}\}_{w,p=1}^{W,P}$

Outputs: The multitaper estimates of the SSD matrices $\hat{\Psi}_m^{\text{mt}}(\omega_n)$ for $1 \leq m \leq M$, $1 \leq n \leq N - 1$

1: **for** $p = 1, 2, \dots, P$ **do**

2: **for** $1 \leq w \leq W$, $1 \leq m \leq M$, $1 \leq j \leq J$ **do**

3: $k = ((m - 1)W + w)$

4: **if** $\bar{n}_{k,j} \neq 0$ and $\bar{n}_{k,j} \neq 1$ **then**

$$(\bar{n}_{k,j})^{(p)} = \ell(\ell^{-1}(\bar{n}_{k,j})\nu_w^{(p)})$$

5: **else**

$$(\bar{n}_{k,j})^{(p)} = \bar{n}_{k,j}$$

6: **end if**

7: **end for**

8: Compute the p^{th} tapered spectral density matrix estimate, $\hat{\Psi}_m^{(p)}(\omega_n)$ for $1 \leq m \leq M$, $1 \leq n \leq N - 1$, using Algorithm 2.1, with $\{\bar{n}_{k,j}^{(p)}\}_{k,j=1}^{K,J}$ as the input collection of ensemble averages.

9: **end for**

10: **for** $1 \leq m \leq M$, $1 \leq n \leq N - 1$ **do**

$$\hat{\Psi}_m^{\text{mt}}(\omega_n) = \frac{1}{P} \sum_{p=1}^P \hat{\Psi}_m^{(p)}(\omega_n)$$

11: **end for**

12: Return $\hat{\Psi}_m^{\text{mt}}(\omega_n)$ for $1 \leq m \leq M$, $1 \leq n \leq N - 1$

We state our main theorem for a univariate second-order stationary process $x_1, x_2, \dots, x_K$, corresponding to the special case of $J = 1$, for the clarity of exposition. We later on provide extensions to the multivariate and semi-stationary cases. To proceed with our theoretical analysis, we need to make two extra technical assumptions.

Assumption (I). From Eq. (2.7), the data likelihood for the univariate case can be expressed

as

$$f(\mathcal{D}) := \int \exp \left(\sum_{k=1}^K L \left\{ \bar{n}_k x_k - \log(1 + \exp(x_k)) \right\} \right) \prod_{k=1}^K dx_k,$$

where $\bar{n}_k := \frac{1}{L} \sum_{l=1}^L n_k^{(l)}$. Given the nonlinear functional form of the integrand, we consider the saddle point approximation [107] of the integral, and take $\ell^{-1}(\bar{n}_k)$ as an estimator of x_k . Under this approximation, the proposed multitaper spectral estimator takes the simpler form:

$$\widehat{S}^{\text{mt}}(\omega) = \frac{1}{P} \sum_{p=1}^P |\widehat{y}^{(p)}(\omega)|^2, \quad (2.15)$$

where $\widehat{y}^{(p)}(\omega) := \sum_{k=1}^K \nu_k^{(p)} \ell^{-1}(\bar{n}_k) e^{-i\omega k}$.

Assumption (2). We assume that $|x_k| \leq B$ for all $k = 1, 2, \dots, K$, for some fixed upper bound B . In defining the bias and variance, we condition the expectations on the event $A := \{\bar{n}_k \mid \bar{n}_k \neq 0, \bar{n}_k \neq 1, 1 \leq k \leq K\}$, which is highly probable due to the absolute boundedness of x_k , for $L > 2(1 + \exp(B))$ (See Appendix A.1 for details). We denote the conditional bias and variance by $\text{bias}_A(\cdot)$ and $\text{Var}_A(\cdot)$, respectively. Note that for the multivariate case, we naturally extend this assumption to $|x_{k,j}| \leq B$ for all k, j , and define the set A as $A := \{\bar{n}_{k,j} \mid \bar{n}_{k,j} \neq 0, \bar{n}_{k,j} \neq 1, 1 \leq k \leq K, 1 \leq j \leq J\}$.

Theorem 2.1 (Univariate Case). *Under the Assumptions (1) and (2) and for $L > \max\{2(1 + \exp(B)), 2B/\log(1 + \exp(B))\}$, the conditional bias and variance of $\widehat{S}^{\text{mt}}(\omega)$ in Eq. (2.15) are bounded with respect to those of the direct multitaper estimate $S^{\text{mt}}(\omega)$ given in Eqs. (2.13) and (2.14) as:*

$$|\text{bias}_A(\widehat{S}^{\text{mt}}(\omega))| \leq g_1 K \frac{\log L}{\sqrt{L}} + |\text{bias}(S^{\text{mt}}(\omega))|, \quad (2.16)$$

$$\text{Var}_A(\widehat{S}^{\text{mt}}(\omega)) \leq \left\{ g_2 K \frac{\log L}{\sqrt{L}} + \sqrt{\text{Var}(S^{\text{mt}}(\omega))} \right\}^2 \quad (2.17)$$

where g_1 and g_2 are bounded constants depending on B , K and L , explicitly given in Appendix A.1.

Proof of Theorem 2.1. The proof of Theorem 2.1 is given in Appendix A.1. $\square$

Remark. Theorem 2.1 states that the cost of the indirect access to the process $\{x_k\}_{k=1}^K$ through spiking observations with L trials appears as excess terms in both the bias and variance, which would go to zero as $\frac{L}{K^2 \log^2 L} \rightarrow \infty$. Hence, for a large enough number of realizations L , one can expect a performance close to the direct multitaper estimate of $\{x_k\}_{k=1}^K$. While Assumption (2) on the boundedness of the time-series is natural in practical scenarios, as long as $|x_k| \leq \epsilon \log L$, for some $\epsilon < 1/2$, the excess bias and variance terms will be bounded by $\mathcal{O}(\log^2 L / L^{1/2-\epsilon})$, which implies that the bias and variance of the estimator $\widehat{S}^{\text{mt}}(\omega)$ will converge to those of $S^{\text{mt}}(\omega)$ even under milder conditions.

The following corollary extends Theorem 2.1 to the multivariate case:

Corollary 2.1 (Stationary Multivariate Case). *Consider a second order jointly stationary J -variate process $\{\mathbf{x}_k\}_{k=1}^K$, where, $\mathbf{x}_k = [x_{k,1}, x_{k,2}, \dots, x_{k,J}]^\top$. Suppose that the observations are binary spiking data, $\{n_{k,j}^{(l)}\}_{k,j,l=1}^{K,J,L}$ with $n_{k,j}^{(l)} \sim \text{Bernoulli}(\ell(x_{k,j}))$ and $\bar{n}_{k,j} := \frac{1}{L} \sum_{l=1}^L n_{k,j}^{(l)}$. Then, under Assumptions (1) and (2), the bias and variance of the multitaper cross-spectral estimate between the r^{th} and t^{th} processes, given by*

$$\widehat{S}_{r,t}^{\text{mt}}(\omega) = \frac{1}{P} \sum_{p=1}^P \widehat{y}_r^{(p)}(\omega) (\widehat{y}_t^{(p)}(\omega))^*$$

with $\hat{y}_r^{(p)}(\omega) = \sum_{k=1}^K \nu_k^{(p)} \ell^{-1}(\bar{n}_{k,r}) e^{-i\omega k}$, are bounded as follows:

$$\begin{aligned} |\text{bias}_A(\hat{S}_{r,t}^{\text{mt}}(\omega))| &\leq g'_1 K \frac{\log L}{\sqrt{L}} + |\text{bias}(S_{r,t}^{\text{mt}}(\omega))|, \\ \text{Var}_A(\hat{S}_{r,t}^{\text{mt}}(\omega)) &\leq \left\{ g'_2 K \frac{\log L}{\sqrt{L}} + \sqrt{\text{Var}(S_{r,t}^{\text{mt}}(\omega))} \right\}^2, \end{aligned}$$

where g'_1 and g'_2 are bounded constants that depend only on B , K , and L , explicitly given in Appendix A.2.

Proof. The proof is outlined in Appendix A.2. □

It is not difficult to verify that $\text{bias}(S_{r,t}^{\text{mt}}(\omega))$ and $\text{Var}(S_{r,t}^{\text{mt}}(\omega))$ can be upper bounded in a similar fashion to Eqs. (2.13) and (2.14), with the true cross-spectra $S_{r,t}(\omega)$ replacing $S(\omega)$. Before extending the result of Corollary 2.1 to the semi-stationary case, we need an additional assumption:

Assumption (3). Given that Corollary 2.1 holds for large L , in this regime we relax the prior distribution on $\mathbf{Q}_m$ to be flat, i.e., $f(\mathbf{Q}_m) \propto 1$. Recall that the rationale for using a prior on $\mathbf{Q}_m$ in Section 2.2.2 was to reduce the variance of the estimates in the low spiking regime, i.e., small L .

Finally, combining Corollary 2.1 and the treatment of [96], we have the following corollary on the bias and variance of the PPMT-SSD estimator:

Corollary 2.2 (Semi-stationary Multivariate Case). *Suppose that the J -variate process in Corollary 2.1 is semi-stationary (jointly stationary within consecutive windows of length W). Let $\Psi_{m,r,t}(\omega_n)$ be the cross-spectra between the r^{th} and t^{th} processes over window m , $1 \leq m \leq K/W$, and $\hat{\Psi}_{m,r,t}^{\text{mt}}(\omega_n)$ be the corresponding multitaper estimate obtained from spiking obser-*

uations. Then, under Assumptions (1)–(3), the bias and variance of the proposed PPMT-SSD estimator at window m can be bounded as,

$$\begin{aligned} |\text{bias}_A(\widehat{\Psi}_{m,r,t}^{\text{mt}}(\omega_n))| &\leq g_1''(\omega_n)W \frac{\log L}{\sqrt{L}} + |\Psi_{m,r,t}(\omega_n)| |1 - \kappa_m(\omega_n)| \\ &\quad + \kappa_m(\omega_n) \left\{ \sup_{\omega} \{|\Psi_{m,r,t}(\omega)|\} \left(1 - \frac{1}{P} \sum_{p=1}^P c_p\right) + o(1) \right\}, \\ \text{Var}_A(\widehat{\Psi}_{m,r,t}^{\text{mt}}(\omega_n)) &\leq \left\{ g_2''(\omega_n)W \frac{\log L}{\sqrt{L}} + \sqrt{\frac{2}{P}} \sup_{\omega} \{\kappa_m(\omega) |\Psi_{m,r,t}(\omega)|\} \right\}^2, \end{aligned}$$

where $g_1''(\omega)$, $g_2''(\omega)$ are bounded functions of B , L , W and $\kappa_m(\cdot)$ is a function of ω , explicitly given in Appendix A.2.

Proof. The proof is mainly based on Theorem 1 of [96]. The proof sketch is given in Appendix A.2. □

2.4 Simulation Studies

In this section, we present two simulation studies based on non-stationary AR processes and compare the performance of the proposed PPMT-SSD estimator to those of two existing methods and benchmarks.

2.4.1 Existing Methods and Benchmarks for Comparison

State-Space SSD (SS-SSD) Estimator

This estimator is based on [33], where the latent process x_k is modeled as a first-order autoregressive process, $x_k = x_{k-1} + \epsilon_k$, with $\epsilon_k \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$ for $1 \leq k \leq K$. Following an

EM algorithm developed in [33], the MAP estimate of x_k given the observed data is obtained and its multitaper PSD is computed. For the multivariate non-stationary case, for each process $x_{k,j}$, $1 \leq j \leq J$, we assume joint stationarity within non-overlapping consecutive windows of length W . The MAP estimates of all processes are obtained and an estimate of the SSD matrix is derived using their PSDs within each window.

Peristimulus Time Histogram SSD (PSTH-SSD) Estimator

This estimator is derived by directly considering the ensemble mean of the spiking observations $\bar{n}_{k,j}$, referred to as the peristimulus time histogram (PSTH), as an estimate of the random signal $x_{k,j}$, for $1 \leq k \leq K$ and $1 \leq j \leq J$ [29]. With a similar joint stationarity assumption in windows of length W , the non-overlapping sliding window multitaper spectral estimate of the PSTH forms the PSTH-SSD estimator.

Benchmarks (True SSD & Oracle SSD)

To benchmark our comparison, we consider the theoretical spectra of the AR processes derived using closed-form expressions (True SSD) as well as the non-overlapping sliding window direct multitaper estimates of the processes $x_{k,j}$ that have been used to generate the spikes. We refer to the latter benchmark as the Oracle SSD as if an oracle could directly observe the latent processes and estimate their SSD.

2.4.2 Simulation study 1: Estimating the SSD of a Latent Trivariate Process

In this study, we model the latent stimuli by AR processes, so that the ground truth is known for performance comparison. We generate three processes ($J = 3$) with spectral couplings, by considering different linear combinations of a set of AR(6) processes, $\{\{y_k^{(i)}\}_{k=1}^K, 1 \leq i \leq 6\}$, where $y_k^{(i)}$ is tuned around the frequency f_i , with $f_0 = 0.0008$ Hz, $f_1 = 1.15$ Hz, $f_2 = 0.95$ Hz, $f_3 = 1.3$ Hz, $f_4 = 1.5$ Hz, $f_5 = 0.65$ Hz and $f_6 = 1.85$ Hz. These linear combinations are given as:

$$\begin{aligned} x_{k,1} &= y_k^{(1)} \cos\left(2\pi \frac{f_0}{f_s} k\right) + 1.2y_k^{(4)} + 1.2y_k^{(5)} u_{k-0.4K} + \sigma_{x1} \nu_{1,k} + x_{1,dc} \\ x_{k,2} &= 0.83 y_k^{(2)} + 0.83 y_{k-6}^{(4)} + 0.83 y_k^{(5)} + 0.83 y_k^{(6)} + \sigma_{x2} \nu_{2,k} + x_{2,dc} \\ x_{k,3} &= y_k^{(3)} + y_k^{(5)} + y_{k-10}^{(6)} u_{0.5K-1-k} + y_k^{(6)} u_{k-0.5K} + \sigma_{x3} \nu_{3,k} + x_{3,dc} \end{aligned}$$

where $x_{i,dc}$ are the DC components, u_k is the unit step function, $\nu_{i,k}$ is a zero mean white Gaussian noise with unit variance, and σ_i is a scaling standard deviation to set the SNR of all signals at 20 dB, for $i = 1, 2$ and 3 . To induce non-stationarity, we have included an amplitude modulated component based on $y_k^{(1)}$ and an abruptly appearing version of $y_k^{(5)}$ in $x_{k,1}$. Likewise, we incorporated non-stationarity and coupling among the processes $\{x_{k,i}\}_{i=1}^3$, by adding time-varying weights to the different $y_k^{(i)}$ s using step functions.

All signals have been sampled at $f_s = 32$ Hz, for a total duration of 2000 seconds ($K = 64000$). We take the window length of stationarity to be 100 seconds ($W = 3200$), resulting in a total number of $M = 20$ windows. We assume that $L = 20$ spike train realizations are observed from each latent process. All DC components have been set to -5.5 , so that the average spiking

rate of the ensemble, denoted by f_r , is ≈ 0.28 spikes/s, consistent with the low spiking rate of experimentally recorded data. A 30 s sample window of the process $x_{k,1}$ and the corresponding spiking raster plot are shown in Figure 2.1.

Figure 2.2 shows the main results of this study, and is formatted as a grid with columns representing (from left to right) the True SSD, Oracle SSD, PPMT-SSD, SS-SSD, and PSTH-SSD estimates, and rows representing $(\Psi_m)_{i,j}(\omega)$, i.e., the magnitude of the $(i, j)^{\text{th}}$ block of the

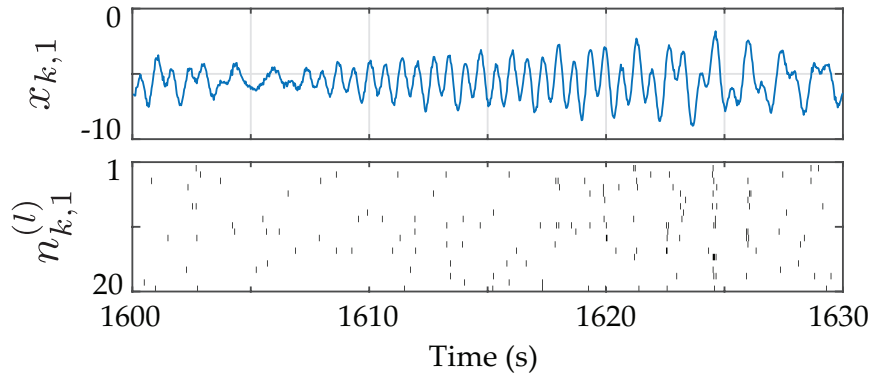


Figure 2.1: Samples of the signal $x_{k,1}$ (top) and the raster plot of the corresponding spikes (bottom) from $t = 1600$ s to $t = 1630$ s.

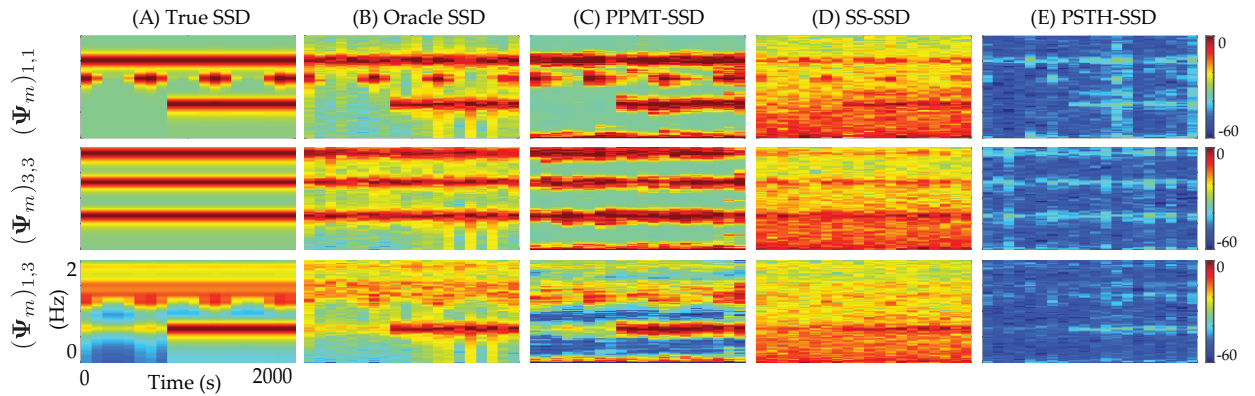


Figure 2.2: Estimation results for Study 1. Each panel shows the magnitude spectra in dB scale. Columns from left to right: (A) True SSD, (B) Oracle SSD, (C) PPMT-SSD, (D) SS-SSD, and (E) PSTH-SSD. Rows from top to bottom: $(\Psi_m)_{1,1}(\omega)$, $(\Psi_m)_{3,3}(\omega)$ and $(\Psi_m)_{1,3}(\omega)$.

SSD matrix for $i, j = 1, 3$. Furthermore, for a closer inspection, the magnitude of the spectra corresponding to a window of $t = 700$ s to $t = 800$ s for $(\Psi_m)_{i,j}(\omega)$, for $i, j = 1, 2$ are shown in Figure 2.3. It can be observed that the proposed PPMT-SSD estimator (Figure 2.2-C) results in much less background noise compared to all the others, while precisely capturing the evolution of the spectra and properly resolving the various spectral components. The latter is more evident from Figure 2.3, where the PPMT-SSD (black trace) closely matches the true SSD (blue trace) on par with the Oracle SSD (red trace), while the SS-SSD (green trace) and PSTH-SSD (orange trace) show significant bias and variability. It is worth mentioning that the erroneous spectral peak near the DC component in Figure 2.3 (black trace) is due to the estimation error of the DC component in the low spiking regime of our setting, and is mitigated as the spiking rate increases.

Due to time-domain smoothing carried out by the SS-SSD estimator, the SSD rapidly decays with frequency (Figure 2.3, green trace). As such, the SS-SSD estimate (Figure 2.2-D) heavily amplifies non-existing low-frequency components that arise from the intrinsic noise in

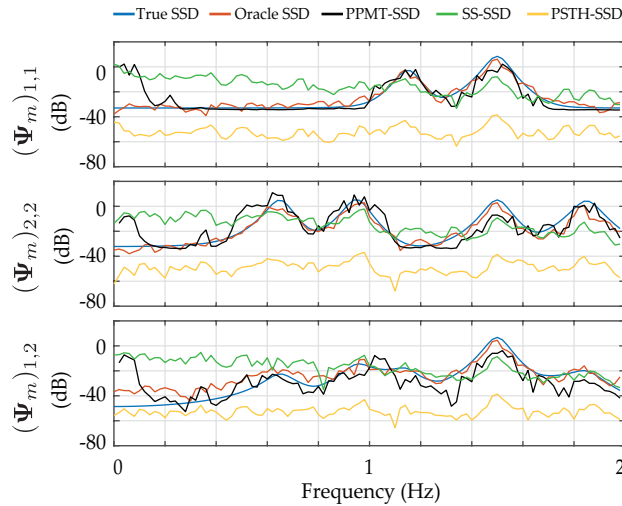


Figure 2.3: A snapshot of the spectrograms of Figure 2.2 at the 8th window ($t = 700$ s–800 s). Rows from top to bottom: $(\Psi_m)_{1,1}(\omega)$, $(\Psi_m)_{2,2}(\omega)$, and $(\Psi_m)_{1,2}(\omega)$.

spiking observations, while suppressing the higher frequency components that exist in the true SSD (Figure 2.2-A). Similarly, the PSTH-SSD estimate shown in Figure 2.2-E fails to capture most of the spectrotemporal features of the SSD, since it does not account for the binary nature of the observations.

To quantify the performance of these estimators, we repeated this numerical experiment for a total of 50 trials, generating independent realizations of the AR processes and spiking observations in each trial. The average Mean Squared Error (MSE) relative to the True SSD (in dB scale) and the average spectral leakage across trials are presented in Table 2.1. To compute the spectral leakage, we first define the *in-band* spectra by the components of the True SSD above -10 dB and then take the relative power of each estimate outside the in-band spectra as the spectral leakage. The MSE values are normalized relative to the total power of the True SSD (in dB scale). The proposed PPMT-SSD estimator achieves the lowest MSE, followed by the SS-SSD estimator with a significant gap. The PSTH-SSD estimator exhibits the poorest performance, in terms of both average MSE and variance, which is also visually evident in Figure 2.3. The spectral leakage of the SS-SSD and PSTH-SSD estimates are an order of magnitude higher than that of the PPMT-SSD, which is comparable to that of the Oracle SSD.

It is noteworthy that the superior performance of PPMT-SSD comes at the cost of higher computational complexity: Algorithm 2.2 has a computational complexity of $\mathcal{O}(JPMWN^2)$,

Table 2.1: Comparison of MSE and spectral leakage in study 1

Estimation method	MSE	Spectral Leakage
Oracle SSD	0.0490 ± 0.0028	$6.79\% \pm 0.36\%$
PPMT-SSD	0.1868 ± 0.0075	$7.69\% \pm 0.89\%$
SS-SSD	0.3906 ± 0.0036	$67.06\% \pm 1.78\%$
PSTH-SSD	1.4777 ± 0.0315	$46.59\% \pm 0.89\%$

whereas the complexity of SS-SSD and PSTH-SSD are considerably lower at $\mathcal{O}(JPMW \log W)$. As such, the performance gain achieved by the PPMT-SSD estimator comes at the cost of a factor of $\mathcal{O}(\frac{N^2}{\log W})$ increase in computational complexity. In the spirit of easing reproducibility, a MATLAB implementation that regenerates the data, results, and figures outlined in this section has been made publicly available on the open-source repository GitHub [49].

2.4.3 Robustness to the Number of Realizations and Spiking Rate

Here, we evaluate the robustness of the various algorithms with respect to two main data attributes: average spiking rate (f_r) and number of realizations (L). For the same latent processes considered in Study 1, we first vary L , while fixing the average spiking rate at $f_r = 0.28$ spikes/s. Next, we fix $L = 10$, and vary f_r by changing the DC components of the latent processes. Finally, we consider $L = 1$ and vary f_r . The corresponding MSE and spectral leakage performance curves are shown in Figure 2.4-A, B, and C, respectively.

The PPMT-SSD estimates uniformly outperform the PSTH-SSD estimates in terms of both MSE and spectral leakage. They similarly outperform the SS-SSD estimates, except in terms of the MSE for the case of $L = 1$ and $f_r \leq 1$ spikes/s (Figure 2.4-A and C). The spectrograms shown in Figure 2.2 correspond to the case (a) marked in Figure 2.4-A ($L = 20$, $f_r = 0.28$ spikes/s). For visual comparison, in Figure 2.5 we also depict the spectrograms of two other cases from Figure 2.4: ($L = 10$, $f_r = 0.28$ spikes/s), marked as case (b) in Figure 2.4-B, and ($L = 1$, $f_r = 5.5$ spikes/s) marked as case (c) in Figure 2.4-C. Overall, all estimates improve by increasing L and/or the spiking rate, as expected. It is noteworthy that all estimators, including

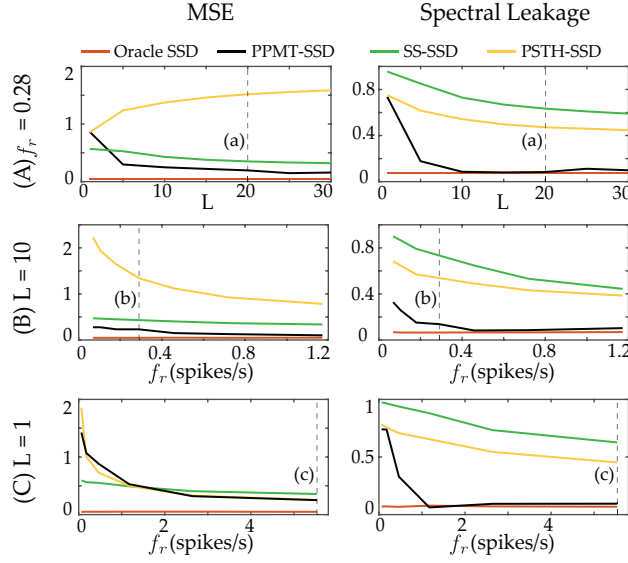


Figure 2.4: Analysis of the robustness of the estimates in Study 1 to L (number of realizations) and f_r (average spiking rate). MSE (left) and Spectral Leakage (right) performance curves are computed by varying: (A) L , for $f_r = 0.28$ spikes/s, (B) f_r , for $L = 10$, and (C) f_r , for $L = 1$.

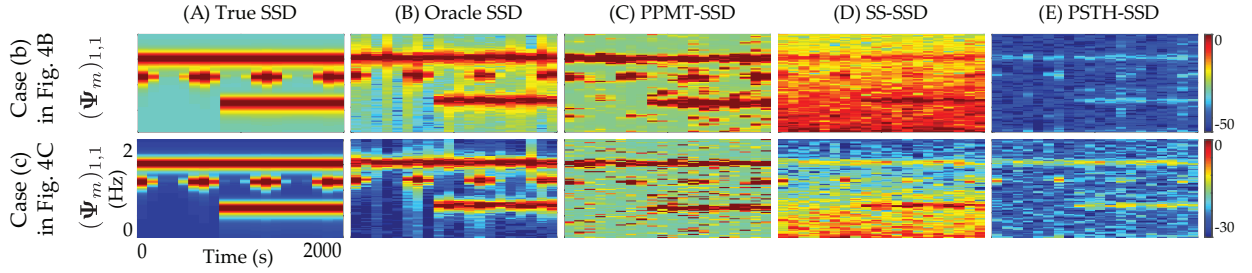


Figure 2.5: Estimated $(\Psi_m)_{1,1}(\omega)$ for two selected cases from Figure 2.4. Each panel shows the magnitude spectra in dB scale. Top: Case (b) in Figure 2.4-B, $L = 10$ and $f_r = 0.28$ spikes/s. Bottom: Case (c) in Figure 2.4-C, $L = 1$ and $f_r = 5.5$ spikes/s. Columns, left to right: (A) True SSD, (B) Oracle SSD, (C) PPMT-SSD, (D) SS-SSD, and (E) PSTH-SSD.

PPMT-SSD, perform poorly when both f_r and L are small (Figure 2.4-C). However, even in this regime, the PPMT-SSD estimator exhibits significantly lower spectral leakage, even for $L = 1$. In conclusion, maintaining a trade-off between L and the spiking rate results in substantially improved performance.

2.4.4 Choice of Design Parameters and Hyper-parameters

We have chosen $N = 800$ bins to have a densely sampled spectral representation. Note that the choice of N directly affects the computational complexity of the estimator. In cases where the spectral range is known *a priori* (e.g., $[0, 2]$ Hz in our simulation setting), the computational complexity can be significantly reduced by using only the frequency bins in the relevant range when constructing the matrix A [86].

The time-bandwidth product of the multitaper framework has been chosen as 2 ($\xi = 2$), and we have used the first three tapers. Note that the design bandwidth of the multitaper framework $\frac{\xi f_s}{W}$ must be chosen to match the spectral spacing $\frac{f_s}{2N}$. Thus, ξ could be determined based on W and N . For any given ξ , the number of tapers (P) with high spectral concentration within the main lobe is bounded by $2\xi - 1$ [42].

The hyper-parameters in Algorithm 2.1 are ρ , ζ and α . By appropriately adjusting the magnitude of the state transition parameter $\alpha \in [0, 1]$, using cross-validation or data-driven tuning within the EM framework, the degree of temporal dependency between adjacent time windows can be controlled. Thus, α was chosen as 0.4. The optimal choices of ρ and ζ depend on the average spiking rate and the number of realizations (L). Accordingly, we have set $\rho = 0.2$ and $\zeta = 0.02$ in Study 1. In general, the parameter ζ affects the average noise floor of the resulting spectral estimates and ρ controls the degree of fluctuations. When the spiking data are highly sparse (i.e., average spiking rate and L are small), choosing higher values of ζ and lower values of ρ results in more robust estimates. These parameters can also be systematically tuned through

cross-validation via Monte Carlo methods as shown in [86].

2.4.5 Simulation study 2: Estimating the SSD of a Bivariate Process, with Only One Directly Observable Component

While the previous study was a natural choice for performance comparison, this study is of particular interest in the joint analysis of neural spiking and continuous signals, such as the local field potential (LFP). The LFP signal corresponds to the electrical field potential measured at the cortical surface, and mesoscale dynamics of cortical activity. We consider a bivariate random process, whose first component $x_{k,1}$ is observed through spiking activity $\{n_{k,1}^{(l)}\}_{k,l=1}^{K,L}$, while its second component $x_{k,2}$ is directly observable in i.i.d. zero-mean Gaussian noise, i.e., $\tilde{x}_{k,2} := x_{k,2} + \nu_k$, with $\nu_k \sim \mathcal{N}(0, \sigma_\nu^2)$. Explicitly, the two processes are given by,

$$\begin{aligned} x_{k,1} &= y_k^{(1)} \cos\left(2\pi \frac{f_0}{f_s} k\right) + y_k^{(4)} + y_k^{(7)} + \sigma_{x1} \nu_{1,k} + x_{1,dc} \\ x_{k,2} &= 0.83 y_k^{(2)} + 0.83 y_{k-6}^{(4)} + 0.83 y_k^{(5)} + 0.83 y_k^{(6)} + \sigma_{x2} \nu_{2,k} + x_{2,dc}. \end{aligned}$$

The process $x_{k,2}$ here is exactly the same as that in Section 2.4.2. To induce additional non-stationarity, we slightly modified the process $x_{k,1}$ by including an additional frequency modulated component. The AR component $y_k^{(7)}$, has been tuned around the frequency f_7 , which changes by decrements of 0.06 Hz every 200 seconds, starting at 0.9 Hz at $t = 0$ s.

The SSD matrix can be estimated by Algorithm 2.2 with a minor modification in the forward filtering step (step 2) of Algorithm 2.1: Given that the second process is directly observable, the distribution $f(\mathcal{D}_1^m | \{\mathbf{w}\}_1^m, \hat{\boldsymbol{\theta}}^{(r)})$ needs to be modified, and accordingly, the log-posterior in Eq.

(2.11) changes to:

$$\sum_{s,w=1}^{m,W} L \left\{ \bar{n}_{(s-1)W+w,1} (\mathbf{A}_s \mathbf{v}_{s,1})_w - \log(1 + \exp(\mathbf{A}_s \mathbf{v}_{s,1})_w) \right\} \\ - \sum_{s,w=1}^{m,W} \frac{1}{2\sigma_\nu^2} \left(\tilde{x}_{(s-1)W+w,2} - (\mathbf{A}_s \mathbf{v}_{s,2})_w \right)^2 - \frac{1}{2} \sum_{s=1}^m (\mathbf{w}_s - \Phi \mathbf{w}_{s-1})^T (\mathbf{Q}_s^{(r)})^{-1} (\mathbf{w}_s - \Phi \mathbf{w}_{s-1}).$$

Figure 2.6 shows the corresponding estimation results, similarly-formatted as in Figure 2.3. Note that in this case, we take the SS-SSD and PSTH-SSD estimates of $(\Psi_m)_{2,2}(\omega)$ to be the same as its Oracle SSD estimate, given that these methods are based on estimating the process $x_{k,2}$ in the time domain (which is directly observable here). Similar to previous study, the proposed PPMT-SSD estimator (Figure 2.6-C) captures the dynamics of the spectra $(\Psi_m)_{1,1}(\omega)$ and $(\Psi_m)_{1,2}(\omega)$ accurately, closely matching the True SSD (Figure 2.6-A). As before, the SS-SSD estimator (Figure 2.6-D) is not able to capture the SSD dynamics, especially at high frequencies. Though some frequency components at certain time windows are detected by the PSTH-SSD estimates (Figure 2.6-E), most of the frequency content is concealed by background noise.

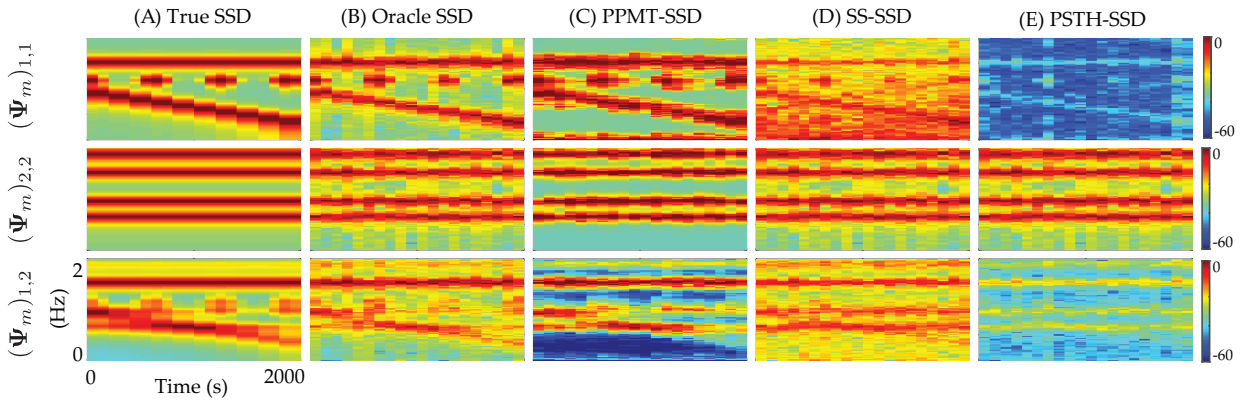


Figure 2.6: Estimation results for Study 2. Each panel shows the magnitude spectra in dB scale. Columns, left to right: (A) True SSD, (B) Oracle SSD, (C) PPMT-SSD, (D) SS-SSD, and (E) PSTH-SSD. Rows, top to bottom: $(\Psi_m)_{1,1}(\omega)$, $(\Psi_m)_{2,2}(\omega)$, and $(\Psi_m)_{1,2}(\omega)$.

2.5 Real data studies

2.5.1 Real data study 1: Application to Experimentally-recorded Data from the Rat Cortical Neurons During Sleep

Next, we apply our proposed PPMT-SSD estimator to multi-unit recordings from the rat cortical neurons during sleep (data from [35], publicly available in the Collaborative Research in Computational Neuroscience data sharing website [101]). The data set includes spiking activities of putative pyramidal cells (pE) and putative interneurons (pI), as well as LFP signals recorded from deep layers of frontal cortical areas [101]. To examine the homeostatic effects of sleep on cortical activity, three main brain states of waking (WAKE), rapid eye movement (REM) sleep, and non-rapid eye movement (nonREM) sleep have been identified and labeled throughout the recordings (See [35] for details). Analyzing the dynamics of neuronal activity during REM and nonREM sleep provides insights into how these brain states differentially contribute to homeostasis. For instance, it has been shown in [35] that the LFP signals exhibit increased delta activity (0.5–4 Hz) during nonREM sleep, as compared to WAKE and REM sleep episodes.

While the results of [35] based on spectrotemporal analyses of the LFPs give significant insights into the cortical dynamics during sleep at the mesoscale, our proposed methodology could be utilized to capture the spectrotemporal dynamics of neuronal activity at the neuronal scale and across the pE and pI cell populations. To this end, we consider a bivariate setting ($J = 2$), where spiking observations of pE and pI cells are represented by $n_{k,1}^{(l)}$ and $n_{k,2}^{(l)}$, respectively. We choose 10 spike trains ($L = 10$) from each cell type for the analysis, for a total observation duration of 35 minutes (data from animal BWRat19, premotor cortex/M2 area [101]). Note that

we have assumed the activity of the L neurons to constitute independent realizations of a process governed by the same underlying intensity. This assumption is motivated by the fact that cortical neurons are known to phase-lock to global oscillatory signals (governed by subcortical processes) under brain states such as anesthesia and sleep [34,35,108], and exhibit the so-called UP-DOWN state dynamics [109,110].

The average spiking rates of the pE and pI populations amount to $f_r = 2.92$ spikes/s and $f_r = 7.86$ spikes/s, respectively. The spike trains are sampled at $f_s = 64$ Hz. The first 7.5 minutes of the observation duration pertain to the WAKE state, followed by a sleep episode consisting of REM and nonREM epochs. Figure 2.7 shows a zoomed-in view of the raster plots of the two neuronal ensembles corresponding to a 30 s window starting at $t = 11$ min.

We assume stationarity within windows of length 30 s, resulting in total $M = 70$ non-overlapping time windows. Focusing our analysis to the delta band (0.5–4 Hz) [35], we estimate the SSDs up to 4 Hz while $N = 512$. Further, we choose $\xi = 2$, and use the first three tapers in the analysis. We set $\alpha = 0.1$, $\zeta = 0.02$ and $\rho = 0.2$. Figure 2.8 shows the results of our analysis. The first three columns (from left to right) represent the PSTH-SSD, SS-SSD and

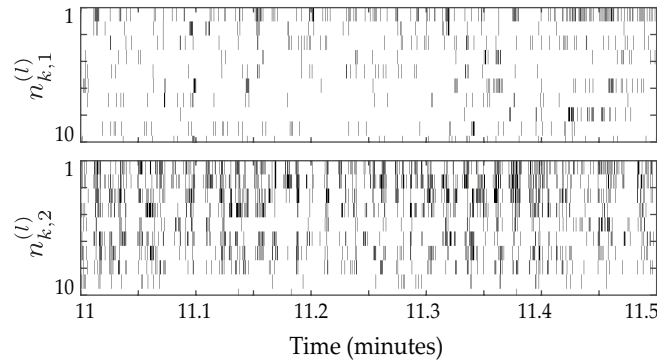


Figure 2.7: Spike raster plots of pE cells (top: $n_{k,1}^{(l)}$) and pI cells (bottom: $n_{k,2}^{(l)}$) within the time window of 11–11.5 min, from the data used in the spectrotemporal analysis of Section 2.5.1.

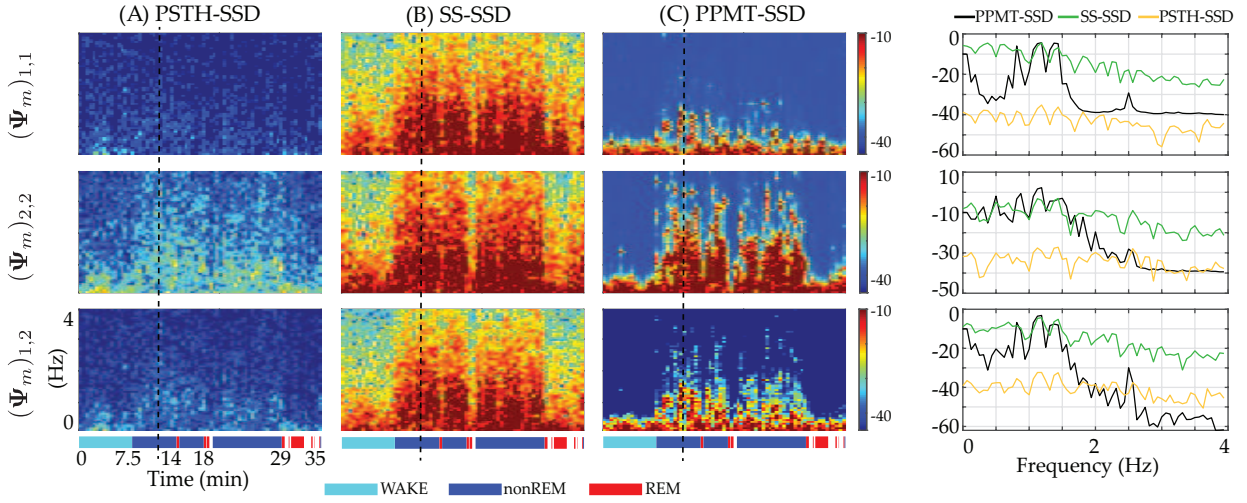


Figure 2.8: SSD analysis of rat cortical activity during sleep. Each panel shows the magnitude spectra in dB scale. The first three columns from left to right: (A) PSTH-SSD, (B) SS-SSD, and (C) PPMT-SSD. State labels of WAKE (cyan), nonREM (blue) and REM (red) are indicated at the bottom of each column. The fourth column shows a snapshot of the spectrograms at the 23rd window (11–11.5 min), whose endpoint is marked by dashed vertical lines in the first three columns. Rows from top to bottom: $(\Psi_m)_{1,1}(\omega)$, $(\Psi_m)_{2,2}(\omega)$, and $(\Psi_m)_{1,2}(\omega)$.

PPMT-SSD estimates. The fourth column illustrates the magnitude spectra corresponding to the time window of 11–11.5 min shown in Figure 2.7. Rows from top to bottom correspond to the estimates of $(\Psi_m)_{1,1}(\omega)$, $(\Psi_m)_{2,2}(\omega)$, and $(\Psi_m)_{1,2}(\omega)$, respectively.

The PSTH-SSD estimate (Figure 2.8-A) is heavily attenuated and concealed by background noise (first column, and orange traces in the fourth column) and does not capture the spectral variations across REM-nonREM transitions. On the other hand, the SS-SSD estimate (Figure 2.8-B) captures these variations at the expense of significantly amplifying a wide band of low frequencies and introducing spurious background noise (green traces in the fourth column). The PPMT-SSD estimates (Figure 2.8-C), however, provide a denoised and well-delineated spectrotemporal representation within and across the brain states. Both pE and pI cells (Figure 2.8-C, first and second row, respectively), exhibit increased power in the delta band over the nonREM epochs.

This is consistently observed in the three main nonREM epochs of 7.5–14 min, 14.5–18 min, and 19–29 min, manifestly within the 0.5–2 Hz frequency band. The cross-spectral coupling between the two cell groups (Figure 2.8-C, third row) similarly increases within the 0.5–2 Hz band during nonREM epochs, as compared to the REM and WAKE states. These observations are consistent with the changes in the LFP spectra across REM/nonREM transitions reported in [35], which indicate a steady increase of delta power during nonREM episodes. Note that the spectral power in the 2–4 Hz band exhibits fluctuation during the nonREM epochs, with increased power in only some of the time windows. This variability is likely due to the non-rhythmic and burst-like spiking activity of the ensemble, perhaps modulated by finer state dynamics endogenous to the nonREM episodes.

The spectral snapshots shown in the fourth column of Figure 2.8 show a consistent performance pattern as in Figure 2.3 from our simulation studies: the spectral peaks are distinguishable in the PPMT-SSD estimates, whereas the SS-SSD and PSTH-SSD estimates flatten the peaks across a wide band of frequencies, with the former undergoing significant attenuation. It is noteworthy that the attenuation of the PPMT-SSD estimates $(\Psi_m)_{1,1}(\omega)$ and $(\Psi_m)_{1,2}(\omega)$ (Figure 2.8, fourth column, first and third rows) around 0.5 Hz is likely due to the increase of burst-like activity of the pE neurons in the particular window chosen for generating the snapshots. This attenuation is not consistent across windows, as evident in the spectrogram plots (Figure 2.8-C, first and third rows), whereas the increased activity of the pI neurons in the 0.5–2 Hz band is persistent during the nonREM episodes (Figure 2.8-C, second row). Our analysis validates the utility of PPMT-SSD estimation as an alternative to existing methods, which excels in providing spectrotemporal characterizations of multivariate spiking data with high resolution and favorable denoising performance.

2.5.2 Real data study 2: Application to experimentally-recorded data from general anesthesia

Finally, we apply our proposed PPMT-SSD estimator to multi-unit recordings from a human subject under Propofol-induced general anesthesia (data from [34, 86]). The dataset includes the spiking activity of 41 neurons, as well as the LFP, recorded from a patient undergoing intracranial monitoring for surgical treatment of epilepsy using a multi-channel microelectrode array implanted in the temporal cortex [34]. Recordings were conducted during the administration of Propofol for induction of anesthesia (see [34] for details). It is known that the neuronal spiking under general anesthesia is highly phase-locked to the LFP signal, with a significant level of coupling between the two processes. Thus, we utilize this virtue to compare the estimated spectra of these two processes as a means of validation.

We first downsample the LFP and spike recordings to the same sampling frequency of 25 Hz [86]. From the spike recordings, we select 25 neurons ($L = 25$) with the highest spiking rates for analysis. Note that we have assumed the activity of the L neurons to constitute independent realizations of a process governed by the same underlying intensity. This assumption is justified by the fact that cortical neurons highly decouple under general anesthesia and phase-lock to global oscillatory signals (governed by subcortical processes) [34]. The average spiking rate of this sub-population of neurons is 0.1732 spikes/second. Figure 2.9 shows the LFP signal for a total duration of 800 s considered in the analysis, as well as a zoomed-in view of the LFP and the raster plot of the neuronal ensemble corresponding to a 40 s window from $t = 480$ s to $t = 520$ s.

Similar to Section 2.4.5, we model the activity of the neuronal ensemble and the LFP using a bivariate process ($J = 2$), where the LFP is directly observable (via a single realization), but the

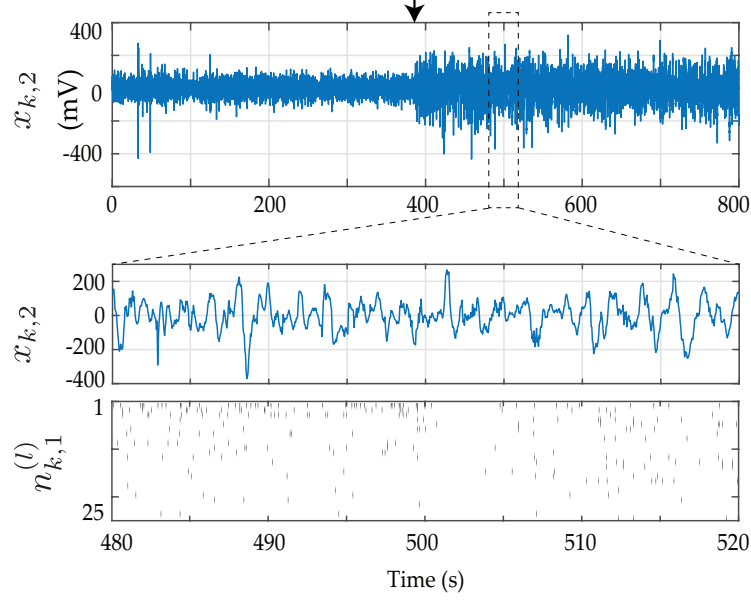


Figure 2.9: Top: LFP signal. The downward arrow marks the induction of the anesthetic. Bottom: zoomed-in view of the experimentally recorded LFP and the spike trains for $t = 480$ s to $t = 520$ s, , from the data used in the spectrotemporal analysis of Section 2.5.2.

latent process governing the neuronal activity is observed through spiking ($L = 25$ realizations). We assume the process to be stationary within windows of length 40 s ($W = 1000$), resulting in $M = 20$ non-overlapping windows. Given that the main spectral content is known to be in the range of 0–2 Hz [34, 86], we estimate the SSDs up to 2.5 Hz while $N = 500$. Further, we choose $\xi = 2$ and use the first three tapers in the analysis. We set $\alpha = 0.85$ and $\rho = 0.02$.

Figure 2.10 shows the results of our analysis. The first column illustrates the SSD estimates, where the dashed vertical line marks the induction of anesthetic. The second column shows snapshots of the normalized magnitude spectrum corresponding to the time window shown in Figure 2.9. The first three rows (from top to bottom) show the estimates of $(\Psi_m)_{1,1}(\omega)$, i.e., the SSD of the latent process governing spiking activity, for PSTH-SSD, SS-PSD, and the proposed PPMT-SSD, respectively. The fourth row shows the multitaper estimate of the LFP signal $((\Psi_m)_{2,2}(\omega))$, and the fifth row depicts the cross-spectra between the LFP and the latent process $((\Psi_m)_{1,2}(\omega))$,

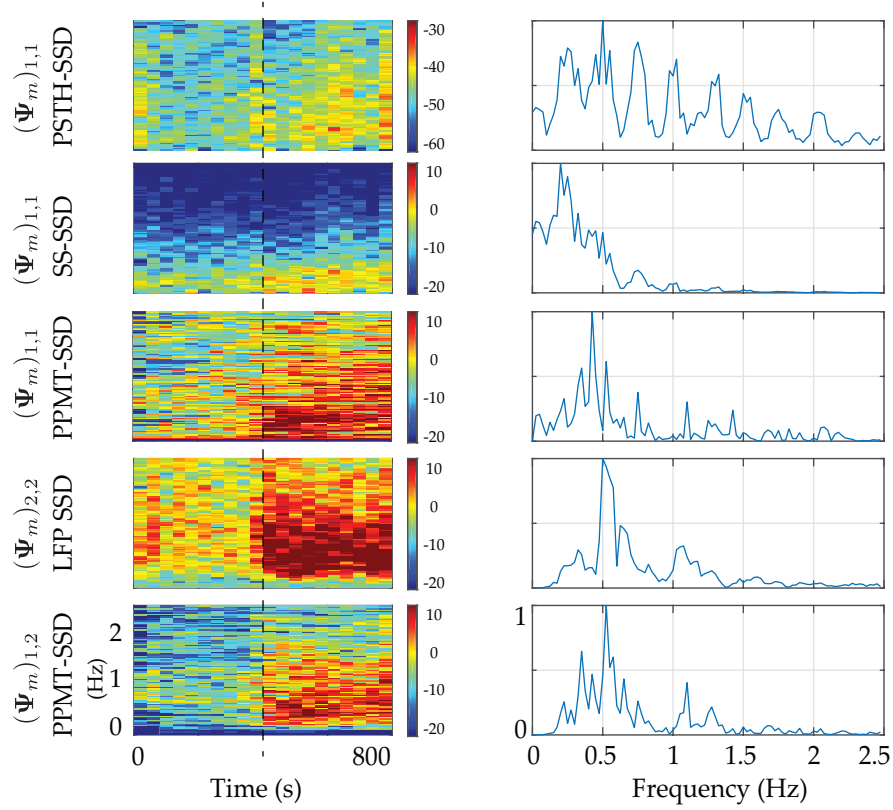


Figure 2.10: SSD analysis of real data from anesthesia. Left: magnitude of the spectrograms in dB scale. Right: a snapshot of the normalized SSD estimates for a 40 s window starting at $t = 480$ s. Rows from top to bottom: PSTH-SSD, SS-SSD, PPMT-SSD of spiking observations, the multitaper SSD of the LFP signal, and the PPMT-SSD estimate of the cross-spectrum between the spiking observations and the LFP.

using the proposed PPMT-SSD method.

The PSTH-SSD estimate (Figure 2.10, first row) captures a significant number of spurious spectral harmonics, which are known to be absent during Propofol general anesthesia [34], masking the relevant spectral content. Also, the SS-SSD estimate amplifies the low-frequency components (Figure 2.10, second row). Even though the dominant spectral content of the LFP is around 0.5 Hz and 1.1 Hz (Figure 2.10, fourth row, right panel), neither of these peaks are present in the SS-SSD estimate (Figure 2.10, second row, right panel). In addition, the rapid change in the LFP spectrogram (fourth row, left panel) marked by the dashed line, is not captured by the

SS-SSD estimate (second row, left panel).

The proposed PPMT-SSD estimate (third row) closely resembles the spectrum of the LFP (fourth row), evident in both the spectrogram and the spectral snapshot. The dominant frequency components of the PPMT-SSD around 0.5 Hz and 1.1 Hz match those in the LFP spectrum. The temporal variations of the PPMT-SSD are also consistent with those of the LFP. The estimated cross-spectra (fifth row) emphasizes this observation, by exhibiting increased spectral coupling between the LFP and the latent process governing the spiking activity following unconsciousness. This result serves as validation that the PPMT-SSD method reliably characterizes the spectra of the latent process governing spiking activity, and affirms the previously reported phase-locking effect, albeit at higher spectrotemporal resolutions [34].

2.6 Concluding Remarks

In this chapter, we proposed a methodology for estimating the SSD matrix of a multivariate non-stationary latent process *directly* from binary spiking observations, thus obviating the need for intermediate time-domain smoothing procedures used by existing techniques, which in turn result in biased estimates of the spectra. To this end, we integrated techniques from state-space modeling, multitaper analysis, and point processes. We established theoretical bounds on the bias and variance performance of the proposed estimator and compared its performance with the existing techniques through application to simulated and experimentally-recorded neural data. Our simulation studies confirmed our theoretical analysis and revealed the favorable performance of our proposed method over existing approaches. Our application to real data provided a highly-resolved characterization of the spectrotemporal dynamics of cortical activity during sleep and

general anesthesia at the neuronal scale. It is worth noting that in the emerging neural recording technologies, such as two-photon calcium and voltage imaging, the LFP signals are not recorded. Spiking activity, however, can be recovered from these recordings using signal deconvolution techniques, which renders our proposed methodology highly desirable in analyzing these data. A key limitation of our proposed method is the assumption of semi-stationarity, which limits its applicability to more general classes of non-stationary processes. Nevertheless, this assumption seems to be plausible for neuronal spiking data under anesthesia or sleep, where the underlying cortical activity pertains to slowly-varying states, as well as other types of slowly-varying binary oscillatory processes such as heartbeat data [111].

Chapter 3: Adaptive Frequency-domain Granger Causal Inference from Neuronal Ensemble Data

The advent of neural data acquisition techniques such as two-photon calcium imaging and multi-electrode array recording has enabled the observation of brain oscillations at the neuronal scale, manifest in ensemble spiking activity. As we also demonstrated in Chapter 2, cross-spectral and coherence analyses of such neural data provide insights into the spectro-temporal coupling between neurons, and have been the focus of active research [34, 48, 87]. However, such analyses do not capture the directionality of neuronal interactions, an essential component of understanding the roles of different neurons within a population. Granger-Geweke causality [30–32, 50, 51, 112–116] is a widely used statistical method of characterizing such frequency domain directional interactions in continuous processes.

In time-domain analysis, if the prediction of one time series is improved by incorporating the past samples of a second one, the second process is said to have a Granger Causal (GC) influence on the first one. Geweke [50, 51] extended this notion to frequency-domain analysis, further introducing a statistical characterization of conditional GC links [51]. The conditional measure quantifies the fraction of total power of the first process at a given frequency that is *directly* contributed by the second process [51]. Even though various applications and extensions of this method have been introduced in the literature [112–116], they are tailored for the analysis

of continuous observations and hence are not directly applicable to binary spiking observations.

Point processes and generalized linear models are widely-applied methods for modeling the underlying non-linear mappings between binary spiking observations and latent intrinsic/extrinsic oscillatory neural covariates [28, 33, 48, 87]. Existing work [30–32] on conditional GC analysis of spiking observations applies methods developed for continuous processes to spike trains, undermining the aforementioned non-linearities and thus leading to biased estimates that cannot be guaranteed to identify the true directional interactions underlying neuronal oscillations. Further, existing methods analyze spectro-temporal coupling using sliding windows, independently estimating GC links at each time window. Brain oscillations underlying neuronal spiking are in general non-stationary and may exhibit dynamics corresponding to brain state or behavior [34, 35]. These dynamics may not be accounted for if the selected window length is too large; at the same time, smaller window lengths can adversely affect the accuracy of spectral estimation.

To address the aforementioned challenges, in this chapter we introduce a unified method to estimate spectro-temporal conditional GC links from binary spiking observations in a generalized linear model framework. We employ state-space modeling techniques to characterize the latent process driving ensemble spiking activity, and use the multitaper method [40, 43, 104] for spectral estimation to subsequently infer conditional Geweke-Granger causal interactions non-parametrically [114]. We quantify the statistical significance of the inferred links using an empirical method. Finally, we demonstrate the utility of the proposed methods in a simulation study and through application to multi-unit recordings from rat cortical neurons during sleep [35, 101].

3.1 Methods

In this section, we summarize the proposed methods for frequency domain conditional GC analysis. We denote the i^{th} element of vector $\mathbf{v}$ by $v^{(i)}$ and the $(i, j)^{\text{th}}$ element of matrix $\mathbf{M}$ by $M^{(i,j)}$ in subsequent expositions. Suppose that the spiking activity of N neurons is observed over L independent trials. Let $n_{t,l}^{(i)}$ denote the spiking activity of the i^{th} neuron at discrete time frame t in the l^{th} trial. Further, suppose that spiking events $n_{t,l}^{(i)}$ for $l = 1, \dots, L$, are driven by an underlying latent process $x_t^{(i)}$ corresponding to latent intrinsic/extrinsic oscillatory covariates. Given the ensemble spiking observations $\{n_{t,l}^{(i)}\}_{i,t,l=1}^{N,T,L}$, we seek to identify the direct causal interactions in the latent oscillations at different frequencies and time windows. To this end, we employ conditional Granger-Geweke Causality (GC) [51].

3.1.1 Geweke's Conditional Granger Causality framework

First, we briefly review the non-parametric formulation of conditional Granger-Geweke causality [114]. Let $\text{GC}_{j \rightarrow i|\mathbf{w}}(f, m)$ denote the conditional GC from the j^{th} neuron to the i^{th} neuron, at frequency f and time window m . Here $\mathbf{w}$ represents the $N - 2$ neurons in the population excluding i and j . To compute the conditional link $\text{GC}_{j \rightarrow i|\mathbf{w}}(f, m)$, we first derive the Power Spectral Density (PSD) matrix [103] within the m^{th} time window of the process $\mathbf{x}_t := [x_t^{(1)}, \dots, x_t^{(N)}]^\top$, denoted by $\mathbf{S}_m(f)$. Re-arranging the entries of $\mathbf{S}_m(f)$, the full ($\mathbf{S}_m^{\text{full}}(f)$)

and reduced ($\mathbf{S}_m^{\text{red}}(f)$) PSD matrices can be formed as:

$$\mathbf{S}_m^{\text{full}}(f) = \begin{bmatrix} S_m^{(i,i)}(f) & S_m^{(i,j)}(f) & \mathbf{S}_m^{(i,\mathbf{w})}(f) \\ S_m^{(j,i)}(f) & S_m^{(j,j)}(f) & \mathbf{S}_m^{(j,\mathbf{w})}(f) \\ \mathbf{S}_m^{(\mathbf{w},i)}(f) & \mathbf{S}_m^{(\mathbf{w},j)}(f) & \mathbf{S}_m^{(\mathbf{w},\mathbf{w})}(f) \end{bmatrix}, \quad \mathbf{S}_m^{\text{red}}(f) = \begin{bmatrix} S_m^{(i,i)}(f) & \mathbf{S}_m^{(i,\mathbf{w})}(f) \\ \mathbf{S}_m^{(\mathbf{w},i)}(f) & \mathbf{S}_m^{(\mathbf{w},\mathbf{w})}(f) \end{bmatrix}.$$

Next, the symmetric minimum phase factorizations of these matrices, defined as:

$$\mathbf{S}_m^{\text{full}}(f) = \mathbf{H}_m(f) \mathbf{\Sigma}_m \mathbf{H}_m^*(f), \quad \mathbf{S}_m^{\text{red}}(f) = \mathbf{G}_m(f) \mathbf{\Omega}_m \mathbf{G}_m^*(f),$$

are obtained using Wilson's algorithm for factorization of matrix spectral densities [117]. This method has been used extensively in previous work [114–116], but, importantly, requires a high spectral sampling frequency and smooth PSD estimates to ensure stability and convergence. Finally, the conditional Granger causal influence of the j^{th} neuron on the i^{th} neuron is defined as:

$$\text{GC}_{j \rightarrow i|\mathbf{w}}(f, m) = \log \left(\frac{\Omega_m^{(i,i)}}{\left(Q_m^{(i,i)}(f) \right) \Sigma_m^{(i,i)} \left(Q_m^{(i,i)}(f) \right)^*} \right), \quad (3.1)$$

where the matrix $\mathbf{Q}_m(f)$ is a function of the spectral factors $\mathbf{H}_m(f)$ and $\mathbf{G}_m(f)$ whose explicit form is given in [114].

3.1.2 Proposed forward model

Note that the oscillatory process $\mathbf{x}_t$ is latent, indirectly observed through spiking processes $n_{t,l}^{(i)}$; hence, recovering the PSD matrix is the key challenge in computing conditional GC links as described previously. Though a method for direct spectral estimation from spiking observations

without requiring estimation of $\mathbf{x}_t$ was introduced in Chapter 2, the computational complexity of this method increases considerably with the increase of frequency bins, and consequently is not well-suited to the conditional GC estimation framework. Alternatively, PSD estimates with minimum bias and variance can be obtained with considerably lower computational complexity and higher sampling frequency using direct multitaper spectral estimation [40, 43, 104] if accurate estimates of $\mathbf{x}_t$ are available. Hence, we propose the use of generalized linear models for point processes [28, 33, 34, 48] to estimate the latent oscillations $\mathbf{x}_t$ from the spiking observations $n_{t,l}^{(i)}$, and subsequently estimate the PSD matrix using multitaper spectral estimation as in [34].

We model the binary spiking process $n_{t,l}^{(i)}$ as a Bernoulli process with a logistic link to the latent oscillatory continuous process $\mathbf{x}_t$:

$$n_{t,l}^{(i)} \sim \text{Bernoulli} \left(\frac{\exp \left(x_t^{(i)} + \mu^{(i)} \right)}{1 + \exp \left(x_t^{(i)} + \mu^{(i)} \right)} \right),$$

where $\mu^{(i)}$, for $i = 1, \dots, N$ are the baseline firing rate parameters; these are taken to be hyperparameters in the proposed model, but in practice may be separately estimated based on the average firing rates of neurons. The latent process $\mathbf{x}_t$ is assumed to be jointly stationary within windows of length W . The oscillatory process is explicitly modeled at the m^{th} time window as a multivariate autoregressive (AR) process:

$$\mathbf{x}_t = \sum_{p=1}^P \mathbf{A}_{m,p} \mathbf{x}_{t-p} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_m), \quad (3.2)$$

for $(m-1)W + 1 \leq t < mW$. Stochastic continuity across time windows is imposed by a first

order state space model on the AR coefficients across time windows:

$$A_{m,p}^{(i,j)} = \beta A_{m-1,p}^{(i,j)} + v_{m,p}^{(i,j)}, \quad v_{m,p}^{(i,j)} \sim \mathcal{N}(0, \sigma^2),$$

where $\beta \in [0, 1]$ is the state transition parameter. Finally, to compensate for potential over-fitting arising from the sparsity of typical spiking observations, we impose a conjugate prior over the noise covariance:

$$\Sigma_m \sim \text{InverseWishart}_N(\boldsymbol{\psi}, \gamma),$$

where $\boldsymbol{\psi}$ and γ are hyper-parameters.

3.1.3 MAP estimation via Expectation-Maximization

Integrating the various facets of the forward model, the joint log-likelihood of the spiking observations $\mathbf{n} := \{n_{t,l}^{(i)}\}_{i,t,l=1}^{N,T,L}$, the latent variables $\mathbf{x} := \{x_t^{(i)}\}_{i,t=1}^{N,T}$ and the parameters $\mathbf{A} := \{\mathbf{A}_{m,p}\}_{m,p=1}^{M,P}$ and $\Sigma := \{\Sigma_m\}_{m=1}^M$ takes the form:

$$\begin{aligned} \log p(\mathbf{n}, \mathbf{x}, \mathbf{A}, \Sigma | \beta, \sigma^2) = & -\frac{1}{2} \sum_{m=1}^M \left((W + \gamma + N + 1) \log |\Sigma_m| + \text{Tr}(\boldsymbol{\psi} \Sigma_m^{-1}) \right) \\ & + \sum_{i,t=1}^{N,T} \left(\sum_{l=1}^L n_{t,l}^{(i)} \left(x_t^{(i)} + \mu^{(i)} \right) - L \log \left(1 + \exp \left(x_t^{(i)} + \mu^{(i)} \right) \right) \right) \\ & - \frac{1}{2} \sum_{m=1}^M \sum_{t=(m-1)W+1}^{mW} \left(\mathbf{x}_t - \sum_{p=1}^P \mathbf{A}_{m,p} \mathbf{x}_{t-p} \right)^\top \Sigma_m^{-1} \left(\mathbf{x}_t - \sum_{p=1}^P \mathbf{A}_{m,p} \mathbf{x}_{t-p} \right) \\ & - \frac{1}{2} \left(N^2 P M \log \sigma^2 + \frac{1}{\sigma^2} \sum_{i,j,m,p=1}^{N,N,M,P} \left(A_{m,p}^{(i,j)} - \beta A_{m-1,p}^{(i,j)} \right)^2 \right) + \Delta, \end{aligned} \quad (3.3)$$

where Δ cumulatively represents terms independent of $\mathbf{n}, \mathbf{x}, \mathbf{A}, \Sigma, \beta$ or σ^2 . We fix the hyperparameters μ, ψ, γ and P , and derive the maximum *a posteriori* (MAP) estimates of the latent variables $\mathbf{x}$ and the parameters $\mathbf{A}, \Sigma, \beta$ and σ^2 using Expectation-Maximization (EM) [46].

3.1.3.1 E-step

The autocovariance function of $\mathbf{x}_t$ up to lag P will be required for subsequent computations, so we first perform a state augmentation $\tilde{\mathbf{x}}_t := [\mathbf{x}_t^\top, \mathbf{x}_{t-1}^\top, \dots, \mathbf{x}_{t-P+1}^\top]^\top$, to derive the first order formulation equivalent to Eq. (3.2): $\tilde{\mathbf{x}}_t = \tilde{\mathbf{A}}_m \tilde{\mathbf{x}}_{t-1} + \tilde{\boldsymbol{\varepsilon}}_t$. The first two conditional moments of the augmented state variable $\tilde{\mathbf{x}}_t$, i.e. $\mathbb{E}[\tilde{\mathbf{x}}_t | \mathbf{n}, \hat{\mathbf{A}}, \hat{\Sigma}, \hat{\beta}, \hat{\sigma}^2]$ and $\mathbb{E}[\tilde{\mathbf{x}}_t \tilde{\mathbf{x}}_t^\top | \mathbf{n}, \hat{\mathbf{A}}, \hat{\Sigma}, \hat{\beta}, \hat{\sigma}^2]$, are obtained by point process smoothing [33] and covariance smoothing [105]. The conditional mean and autocovariance of $\mathbf{x}_t$ can subsequently be recovered using the first two conditional moments of $\tilde{\mathbf{x}}_t$.

3.1.3.2 M-step

The parameter updates for $\mathbf{A}, \Sigma, \beta$ and σ^2 are derived by maximizing the expected log-likelihood $\mathbb{E}_{\mathbf{x}}[\log p(\mathbf{n}, \mathbf{x}, \mathbf{A}, \Sigma | \beta, \sigma^2)]$. First, we address the updates for $\mathbf{A}_{m,p}$, for $m = 1, \dots, M$ and $p = 1, \dots, P$. Note that the joint log-likelihood in Eq. (3.3) is quadratic in $\mathbf{A}_{m,p}$, and hence the posterior distribution $p(\mathbf{A} | \mathbf{n}, \mathbf{x}_{1:T}, \hat{\Sigma}, \hat{\beta}, \hat{\sigma}^2)$ is in fact Gaussian. Thus, the maximization of $\mathbf{A}_{m,p}$ can be obtained by computing the conditional mean, for which we use fixed interval smoothing [74]. The updated $\hat{\mathbf{A}}$ is then used to derive closed form expressions for the updates of Σ, σ^2 , and β by maximizing $\mathbb{E}_{\mathbf{x}}[\log p(\mathbf{n}, \mathbf{x}, \hat{\mathbf{A}}, \Sigma | \beta, \sigma^2)]$.

We iterate between the aforementioned expectation and maximization steps until conver-

gence. The resulting estimates are used to obtain the MAP estimate of the latent oscillatory process $(\hat{\mathbf{x}}_t)_{\text{MAP}} := \mathbb{E}[\mathbf{x}_t | \mathbf{x}_{1:T}, \hat{\mathbf{A}}, \hat{\mathbf{\Sigma}}, \hat{\beta}, \hat{\sigma}^2]$. Then, the PSD matrix $(\mathbf{S}_m(f))_{\text{proposed}}$ is estimated non-parametrically by performing sliding window multitaper spectral estimation [43] on $(\hat{\mathbf{x}}_t)_{\text{MAP}}$. Notably, the proposed model enables the full PSD matrix to alternatively be parametrically obtained using the estimated AR coefficients $\hat{\mathbf{A}}_{m,p}$ and the estimated noise covariance $\hat{\mathbf{\Sigma}}_m$ [112]. However, $(\mathbf{S}_m(f))_{\text{proposed}}$ is in fact the better PSD estimate, since $(\hat{\mathbf{x}}_t)_{\text{MAP}}$ for $t = 1, \dots, T$ capture the underlying dynamics with a much larger number of parameters than $\hat{\mathbf{A}}$ and $\hat{\mathbf{\Sigma}}$ combined. Thus, the presented results focus on non-parametric spectral estimation using $(\hat{\mathbf{x}}_t)_{\text{MAP}}$, which we denote by $(\mathbf{S}_m(f))_{\text{proposed}}$. Finally, employing Wilson’s factorization method and conditional Granger-Geweke analysis as described in Eq. (3.1), we derive the proposed estimates of the GC links $\text{GC}_{j \rightarrow i | \mathbf{w}}^{\text{proposed}}(f, m)$.

3.1.4 Testing statistical significance

Estimated GC links with non-zero values do not necessarily indicate a meaningful effect; this necessitates statistical inference to identify the salience of estimated influences. The significance of estimated links, $\text{GC}_{j \rightarrow i | \mathbf{w}}^{\text{proposed}}(f, m)$, is tested using the following data-driven approach. We empirically obtain the distribution of GC links under the null hypothesis that there is no contribution from process j at frequency f to the power spectrum of i over time window m . Time indices were shuffled for each neuron individually to decorrelate neuronal activity and spectral GC analysis was applied to shuffled data. The distribution of the resulting links are used as the empirical null distribution against which estimated links $\text{GC}_{j \rightarrow i | \mathbf{w}}^{\text{proposed}}(f, m)$ are tested.

3.2 Results

In this section, we present a simulation and a real data study demonstrating the utility of the proposed method. In both studies, we compare the performance to an existing, widely used method [30–32] whose results are based on the peristimulus time histogram (PSTH) of spiking activity, which we refer to as PSTH estimates, $\text{GC}_{j \rightarrow i | \mathbf{w}}^{\text{PSTH}}(f, m)$. The PSTH is given by the ensemble average of spiking activity $(\hat{x}_t^{(j)})_{\text{PSTH}} = \frac{1}{L} \sum_l n_{t,l}^{(j)}$, which is used to estimate the latent oscillatory process, followed by sliding window multitaper spectral estimation. We additionally benchmark the performance of the estimators with respect to the oracle estimator, $\text{GC}_{j \rightarrow i | \mathbf{w}}^{\text{oracle}}(f, m)$, in the simulation study. The oracle estimator, which uses the true latent process as an observable, is derived using the multitaper estimates of $\mathbf{x}_t$ in simulations.

3.2.1 Simulation study

We simulated the activity of three neurons as a tri-variate ($N = 3$) AR process, with $L = 30$ independent trials per latent process. Spikes were sampled at $f_s = 30$ Hz for a total duration of 70 minutes. Figure 3.1 shows a one-minute sample duration of the three latent oscillatory processes and the raster plots. The ground truth PSD matrix and conditional GC links in Figure 3.2 respectively show the true dynamics in spectral coupling and GC connectivity across frequencies. Assuming stationarity within windows of length $W = 2100$, direct causal links were estimated between each pair of processes using the oracle, proposed, and PSTH methods.

The magnitude spectrograms and the statistically significant conditional GC links inferred by each of the three methods are illustrated in Figure 3.3. The proposed estimates closely follow the oracle estimates and the ground truth, recovering most of the spectro-temporal dynamics of

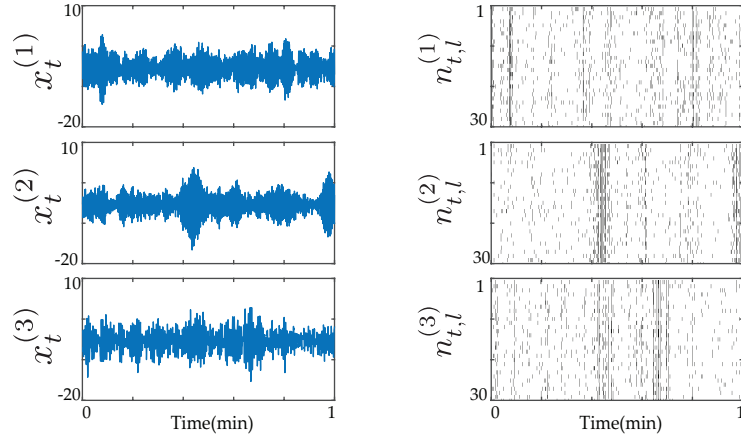


Figure 3.1: Samples of the latent signal $x_t^{(i)}$ (left) and the raster plot of the corresponding spikes $n_{t,l}^{(i)}$ (bottom) from a one minute duration.

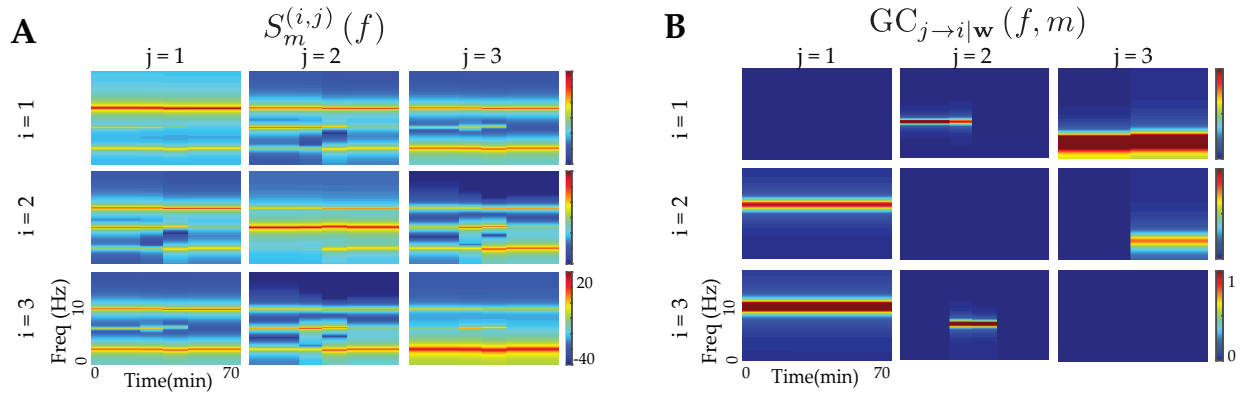


Figure 3.2: (A) Magnitude spectrograms of the ground truth PSD matrix $S_m(f)$ in dB scale and (B) the ground truth conditional GC links $GC_{j \rightarrow i|w}(f, m)$ between each pair of neurons, in the simulation study.

ground truth spectrograms and the conditional GC links. However, the PSTH-based approach fails to identify links that are significant according to either statistical test. In fact, PSTH spectral estimates generally capture spurious harmonics masking the true spectral content as was also seen in Chapter 2, which in turn adversely affects the corresponding conditional GC estimates, as seen in Figure 3.3.

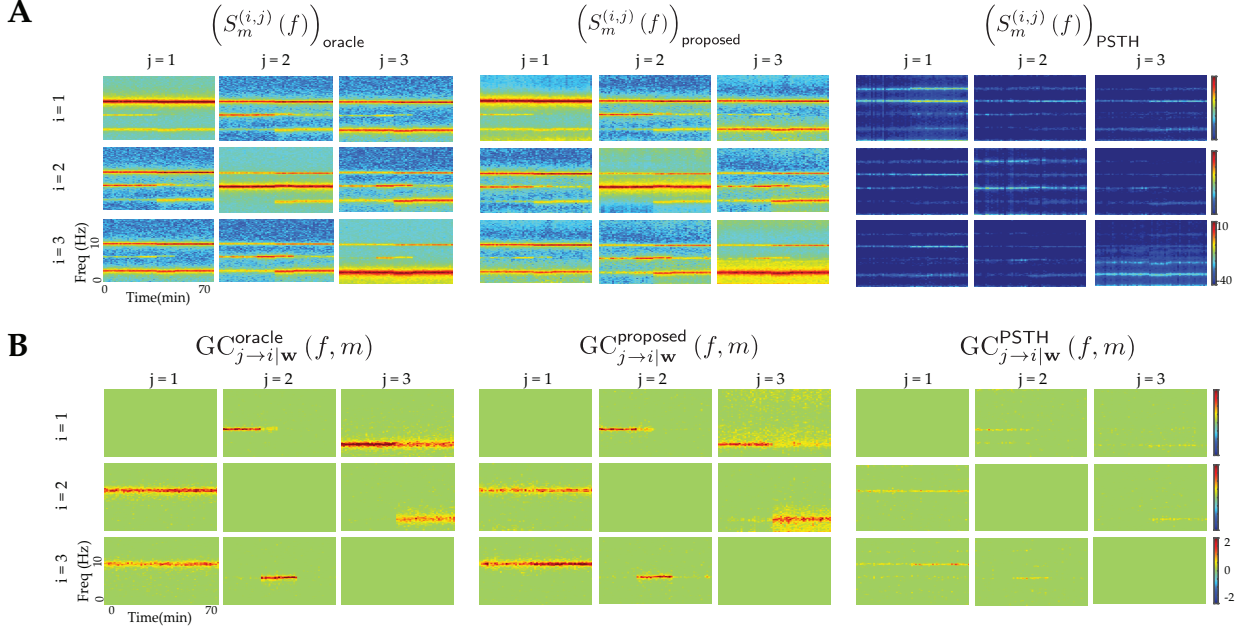


Figure 3.3: (A) Magnitude spectrograms of the estimates in dB scale (B) Statistically significant conditional GC links according to the empirical test with significance level $\alpha = 0.01$, in the simulation study. Columns from left to right: the oracle estimates, proposed estimates and PSTH estimates.

3.2.2 Application to experimentally recorded data

Finally, we applied our method to multi-unit recordings of rat cortical neurons during sleep (data from [35], publicly available in the Collaborative Research in Computational Neuroscience data sharing website [101], also analyzed in Section 2.5.1 of Chapter 2). The dataset included the spiking activity of putative pyramidal cells (pE) and putative interneurons (pI) recorded during three main brain states: waking (WAKE), rapid eye movement (REM) sleep, and non-rapid eye movement (nonREM) (See [35] for details). We considered a bivariate setting ($N = 2$), denoting spiking observations of pE and pI cells by $n_{k,l}^{(1)}$ and $n_{k,l}^{(2)}$, respectively. We chose 10 spike trains ($L = 10$) from each cell type for the analysis, and considered an observation period of 35 minutes.

Figure 3.4 shows the cross power spectral density estimates and the conditional GC esti-

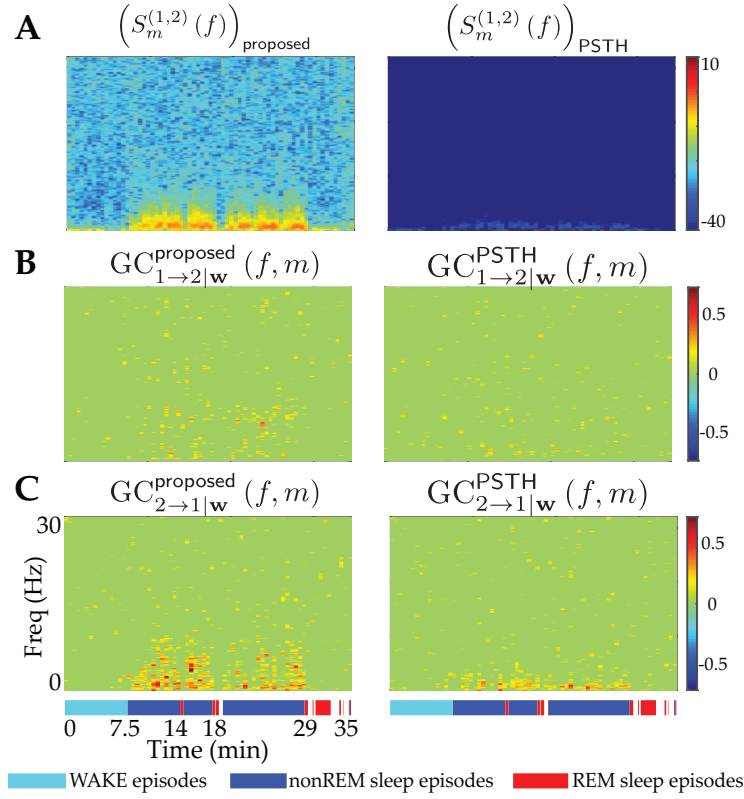


Figure 3.4: Spectral GC analysis of rat cortical spiking during sleep. (A) The magnitude spectrogram of estimated cross PSD $(S_m^{(1,2)}(f))$ in dB scale, (B) the estimated GC links from pE cells to pI cells $(GC_{1 \rightarrow 2|w}(f, m))$ and (C) the estimated GC links from pI cells to pE cells $(GC_{2 \rightarrow 1|w}(f, m))$. The Proposed (resp., PSTH) estimates are shown in the left (right) column. State labels of WAKE (cyan), nonREM (blue) and REM (red) are indicated at the bottom of each column. All GC links shown were statistically significant with respect to the empirical test for $\alpha = 0.01$.

mates from the proposed and PSTH methods during different brain states. The GC link estimates in Figure 3.4 are statistically significant with respect to the empirical test at level $\alpha = 0.01$. Similar to the results in Section 2.5.1 of Chapter 2, the cross PSD estimated from the proposed method (Figure 3.4-A left) has greater cross power in low frequencies during nonREM sleep episodes. The GC links estimated with the proposed method, shown in Figure 3.4-B (left) and Figure 3.4-C (left), indicate steady low-frequency links from interneurons to pyramidal cells during nonREM sleep episodes. In contrast, PSTH estimates (Figure 3.4 right) of cross spectra and GC links are

unable to characterize these spectro-temporal dynamics.

3.3 Concluding Remarks

In this chapter, we introduced a framework that combines point processes, state-space modeling, multitaper spectral estimation, and conditional Granger-Geweke Causality analysis to infer spectro-temporal causal interactions from ensemble neuronal spiking activity, addressing shortcomings in existing methods. We also introduced a method to test the statistical significance of inferred functional links. The utility of the proposed methods was demonstrated through a simulation study. In an application to cortical neuronal data acquired during sleep, the proposed method identified dynamics in functional connectivity under different brain states. These studies highlight the ability of the proposed method to accurately infer the underlying direct causal spectro-temporal interactions in a network of neurons, outperforming existing methods for spiking observations.

Chapter 4: Denoised Multitaper Frequency-domain Heart Rate Variability inference

Heart rate variability (HRV) describes the fluctuations in the time intervals between adjacent heartbeats [53–55]. The beat-to-beat variability in heart rate indexes neurocardiac functionality and is known to be generated by heart-brain interactions and dynamic non-linear autonomic nervous system (ANS) processes [53–55]. HRV in general reflects the regulation of autonomic balance, blood pressure (BP), gas exchange, gut, heart, and vascular tone, and possibly facial muscles [53–55]. Further, the well-known oscillatory nature of HRV has led to the wide usage of its frequency-domain analysis for the identification of key features [53–55]. For instance, two critical parameters obtained from frequency-domain HRV analysis being widely used are the low frequency (LF) power ($0.04 - 0.15$ Hz) which represents both sympathetic and vagal influences; and the high frequency (HF) power ($0.15 - 0.40$ Hz) which reflects the modulation of vagal tone [55]. In addition, the ratio between LF and HF components has been proposed to represent the interplay between sympathetic and parasympathetic nervous system activation, with sympathetic tone tending toward the low-frequency component and parasympathetic tone toward the high-frequency component [53–55]. Thus, spectral analysis of HRV is a non-invasive and easy-to-perform tool that is widely being used for evaluating cardiac autonomic activity [53–55].

While there exist various parametric and non-parametric methods for spectral estimation of

stationary processes, non-parametric approaches are often preferred because they explicitly estimate the frequency spectrum of the process without assuming any particular a priori parametric structure [91]. Out of the many existing non-parametric spectral estimation methods, multitaper spectral analysis has gained significant popularity in recent years due to its various favorable properties [40–43]. Multitapering simultaneously reduces the estimation bias and variance incurred by using data of finite duration to estimate the spectra [42, 43]. Moreover, multitapering also admits a principled framework for constructing confidence intervals about any specific frequency of interest in the spectra [42, 43]. However, a systematic approach to performing HRV analysis in the frequency-domain in a multitaper framework is not yet available, to the best of our knowledge.

We close this gap in this chapter by introducing multitaper methods for frequency-domain HRV analysis. Further, although the classical multitaper framework has the aforementioned favorable characteristics, it does not intrinsically have a mechanism to suppress the additive measurement noise that commonly contaminates empirical observations. As we also illustrated in Chapter 2, the classical multitaper estimates can be further improved by posing the multitaper spectral estimation problem in a Bayesian framework, where observation noise and other additive artifacts are explicitly taken into account and suppressed. Accordingly, we formulate the problem of denoised multitaper spectral estimation from the HRV time series in a Bayesian framework and then solve the inverse problem utilizing Expectation-Maximization (EM) following similar methods to Chapter 2. In addition, we provide two different methods for deriving confidence bounds of the denoised multitaper spectral estimates. Finally, we demonstrate the applicability of these methods by identifying the denoised multitaper frequency-domain components with confidence bounds, from short-term (5 minute) single-lead electrocardiogram (EKG) recordings (data from

Nicholas Schiff lab, Weill Cornell Medicine, New York City, NY).

4.1 Methods

4.1.1 Data Pre-processing and Preliminary Assumptions

We performed the frequency-domain HRV analysis on the tachogram of inter-beat-intervals, which is simply a recording of the beat-to-beat heart rate fluctuations. More specifically, the tachogram of inter-beat-intervals is a time series that records the time difference between consecutive R-peaks in the EKG as a function of the R-peak arrival times [53]. It is noteworthy that both independent and dependent variables in this time series recording are time intervals, with one being related to the other. Furthermore, since the variability in heart rate occurs on a beat-to-beat basis, this time series is inherently unevenly spaced along the horizontal axis [53]. Thus, a typical preliminary preprocessing step before frequency-domain analysis is to evenly resample the tachogram of inter-beat-intervals to have a uniform sampling rate across the horizontal axis. To that end, we utilized linear interpolation. Further, since all frequency components of interest for HRV analysis are less than 2 Hz, we downsampled the time series to a sampling rate of 4 Hz. Following standard procedure for frequency-domain HRV analysis, we then assumed these short segments (5 minute) of heart rate variability data to be stationary processes.

4.1.2 Denoised Multitaper Spectral Estimation

We used similar methods to Chapter 2 to derive the denoised multitaper spectral estimates. Specifically, let x_k denote the k^{th} time sample of the linearly interpolated and downsampled (sampling rate: f_s Hz) HRV time series. Suppose that we use L data tapers for multitapering,

and let $\nu_k^{(l)}$ denotes the k^{th} time sample of the l^{th} discrete prolate spheroidal sequence (dpss) / data taper [41]. For each of the L tapers, first we derived the tapered time series $x_k^{(l)} := \nu_k^{(l)} x_k$, by multiplying the HRV time series x_k with the corresponding data taper $\nu_k^{(l)}$. Then, let $\mathbf{x}^{(l)} := [x_1^{(l)}, x_2^{(l)}, \dots, x_K^{(l)}]^\top$ denote the vector containing all K time points of the l^{th} tapered time series. Following a treatment similar to Chapter 2, we related this time series to the denoised spectral eigen-coefficients $\mathbf{w}^{(l)} := [w_1^{(l)}, w_2^{(l)}, \dots, w_{2N-1}^{(l)}]^\top$ using the formulation:

$$\mathbf{x}^{(l)} = \mathbf{A} \mathbf{w}^{(l)} + \boldsymbol{\epsilon}^{(l)},$$

where $\mathbf{A}$ is a $K \times (2N - 1)$ inverse Fourier transform matrix with the first column equal to $\frac{1}{N}$ (N is the total number of frequency bins), $A_{k,2n} := \frac{1}{N} \cos\left(\frac{\pi kn}{N}\right)$ and $A_{k,2n+1} := \frac{1}{N} \sin\left(\frac{\pi kn}{N}\right)$ for $1 \leq k \leq K$ and $1 \leq n \leq N - 1$, and $\boldsymbol{\epsilon}^{(l)} := [\epsilon_1^{(l)}, \epsilon_2^{(l)}, \dots, \epsilon_K^{(l)}]^\top$ with $\epsilon_k^{(l)} \sim \mathcal{N}\left(0, \sigma^{2(l)}\right)$ modeling additive observation noise. Finally, we assumed that the denoised eigen-coefficients are independent and follow a Gaussian distribution:

$$\mathbf{w}^{(l)} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}^{(l)}\right),$$

where $\mathbf{Q}^{(l)}$ is a diagonal matrix. Under this formulation, the denoised multitaper spectral estimator for the n^{th} frequency bin ($1 \leq n \leq N - 1$) can be derived as:

$$S^{\text{mt}}\left(\frac{f_s n}{2N}\right) = \frac{1}{L} \sum_{l=1}^L \left(\left(w_{2n}^{(l)}\right)^2 + \left(w_{2n+1}^{(l)}\right)^2 \right). \quad (4.1)$$

Next, we solved the inverse problem using Expectation-Maximization (EM) [46], as in Chapter 2. In brief, for each of the L tapers, we first estimated the expected values of $\mathbf{w}^{(l)}$ in the

E-step, then updated the parameter estimates $\sigma^{2(l)}$ and $\mathbf{Q}^{(l)}$ in the M-step, and iterated between the E-step and M-step until convergence. Once the estimates of $\mathbf{w}^{(l)}$ are obtained for $1 \leq l \leq L$, we derived our final denoised multitaper spectral estimates using the formula in Eq. (4.1).

4.1.3 Confidence Intervals

Next, we derived the confidence bounds of the denoised multitaper spectral estimates using two different methods. The first approach is to use the frequency-domain bootstrap (FDB) procedure introduced in [118]. We followed the same steps outlined in [118], performing the bootstrapping procedure for 5000 repeats.

The second approach is to utilize the chi-square approximation to the distribution of multitaper spectral estimates [42]. In brief, it is well-known [42] that the multitaper spectral estimator $S^{\text{mt}}(f)$ approximately follows a chi-square distribution with $2L$ degrees of freedom and a scaling factor of $\frac{S(f)}{2L}$, where $S(f)$ is the true spectrum and L is the number of tapers. Hence:

$$\begin{aligned} P\left(\phi_{2L}(\alpha) \leq \frac{2LS^{\text{mt}}(f)}{S(f)} \leq \phi_{2L}(1-\alpha)\right) &= 1 - 2\alpha \\ P\left(\frac{2LS^{\text{mt}}(f)}{\phi_{2L}(1-\alpha)} \leq S(f) \leq \frac{2LS^{\text{mt}}(f)}{\phi_{2L}(\alpha)}\right) &= 1 - 2\alpha, \end{aligned} \quad (4.2)$$

where $\phi_{2L}(\alpha)$ is the $\alpha \times 100\%$ percentage point of the inverse χ_{2L}^2 (chi-square distribution with $2L$ degrees of freedom) cumulative distribution function. Accordingly, the $(1 - 2\alpha) \times 100\%$ confidence bounds of the multitaper spectral estimates are given by $\left[\frac{2LS^{\text{mt}}(f)}{\phi_{2L}(1-\alpha)}, \frac{2LS^{\text{mt}}(f)}{\phi_{2L}(\alpha)}\right]$. For instance, setting $\alpha = 0.025$ in Eq. (4.2) gives the 95% confidence bounds of the multitaper spectral estimates based on the chi-square approximation.

4.2 Results

In this section, we demonstrate the applicability of our method by deriving the denoised multitaper spectral estimates with confidence bounds of a 5 minute-long EKG segment from a healthy individual (data from Nicholas Schiff lab, Weill Cornell Medicine, New York City, NY). Figure 4.1 displays the time domain pre-processing steps followed for deriving the interpolated inter-beat-interval tachogram (fourth row) from the raw EKG data (first row), on a sample 15 second long data segment.

The first step performed was to extract the R-peak arrival times (Figure 4.1 second row) from the experimentally recorded raw EKG data (Figure 4.1 first row). Next, we derived the inter-beat interval time series (Figure 4.1 third row), by recording the time differences (in milliseconds) between consecutive R-peaks as a function of the R-peak arrival times. Then, we performed linear interpolation on this inter-beat interval tachogram (Figure 4.1 fourth row) to have even sampling across the time axis. Finally, we performed denoised multitaper spectral estimation on this linearly interpolated inter heartbeat interval tachogram.

For an inter-beat-interval tachogram that is 5 minutes long, if the sampling frequency is $f_s = 4$ Hz, the total number of data points amounts to $K = 1200$ samples. Thus, we set the number of frequency bins to be estimated to $N = 512$, the multitaper time-half-bandwidth product to be 2, and the number of data tapers to be $L = 3$, to achieve an optimal spectral resolution with minimal computational complexity (see Chapter 2 for more details). Under these parameter settings, the multitaper spectral resolution amounts to 0.0067 Hz, while the spectral spacing is 0.0039 Hz. Figure 4.2-A shows the denoised multitaper spectral estimates thus obtained. The top row of Figure 4.2-A shows the denoised multitaper spectral estimates with 95% confidence

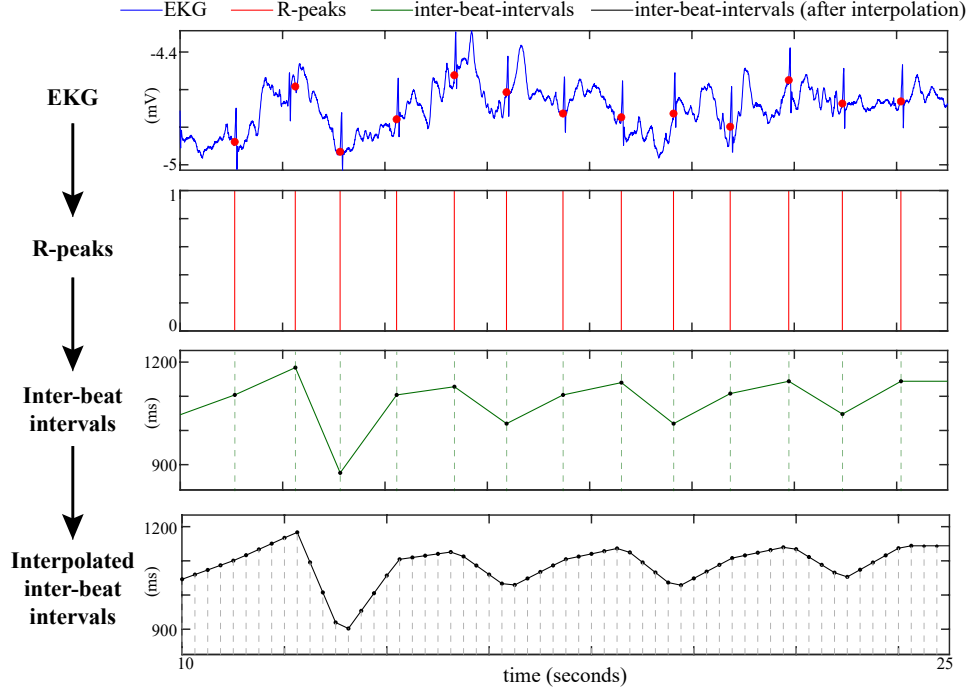


Figure 4.1: Time domain pre-processing steps performed before frequency-domain HRV analysis. Rows from top to bottom: sample 15 second long EKG recording (blue) from a healthy individual and the detected R-peaks (red), extracted R-peak arrival time points (red), the corresponding inter-beat interval tachogram which records the time differences (in milliseconds) between consecutive R-peaks as a function of the R-peak arrival times (green), and the linearly interpolated inter-beat interval time series with even sampling across the time axis (black).

bounds derived through bootstrapping and the bottom row shows the denoised multitaper spectral estimates with 95% confidence bounds obtained using the chi-squared approximation. The dashed vertical lines indicate the frequency boundaries of the low frequency (LF) and high frequency (HF) spectral bands.

For comparison, we also derived the classical multitaper spectral estimates [43], as shown in Figure 4.2-B. While the denoised multitaper spectral estimates (Figure 4.2-A) closely follow the classical multitaper spectral estimates (Figure 4.2-B), we also see some notable differences between the two. For instance, the classical estimates clearly outline two distinct spectral peaks in the frequency range $[0, 0.04]$, while the denoised estimates outline a uni-modal spectral distri-

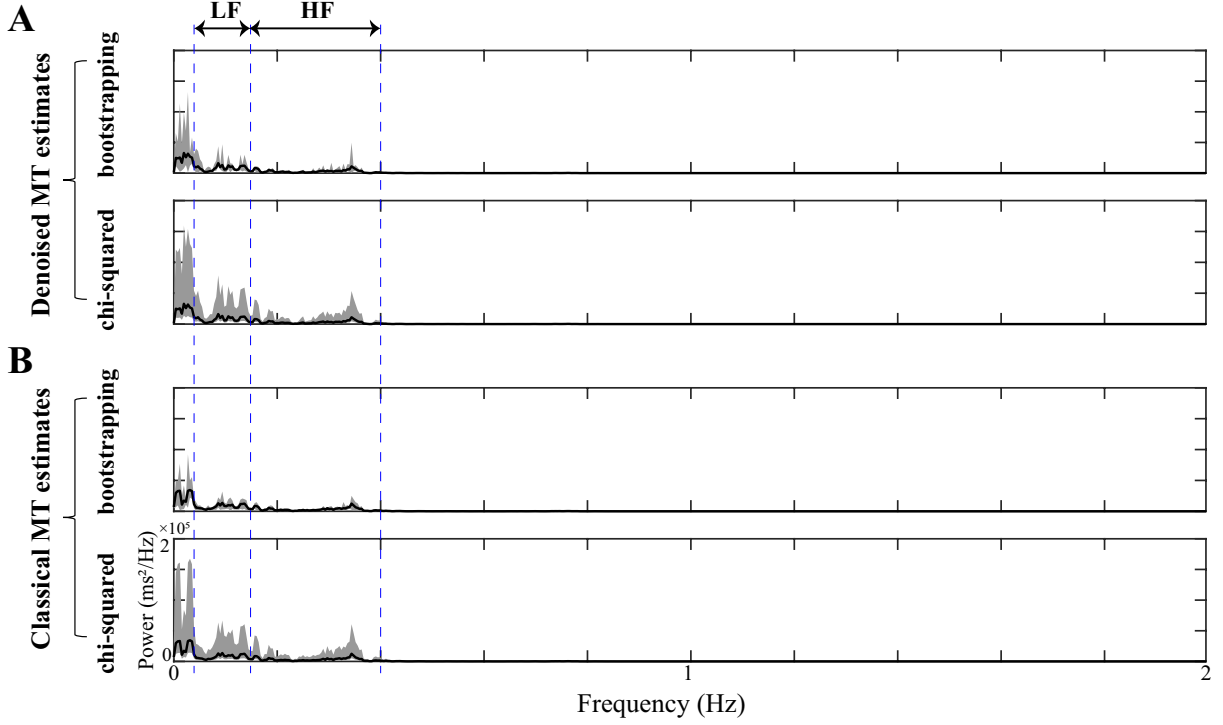


Figure 4.2: Multitaper frequency-domain HRV analysis results of EKG data from a healthy individual. (A) Denoised multitaper spectral estimates derived from the EKG recording with 95% confidence bounds based on bootstrapping (top) and the chi-squared approximation (bottom) (B) The classical multitaper spectral estimates, in the same organization as panel A. The dashed vertical lines indicate the frequency boundaries of the low frequency (LF) and high frequency (HF) spectral bands.

bution within the same frequency range. We believe that these spurious effects in the classical multitaper estimates could largely be artifacts induced by noise and jitter at the observation level. The denoised multitaper spectral estimation method explicitly accounts for and suppresses observation noise, thus providing more accurate and robust spectral estimates.

4.3 Concluding remarks

In this chapter, we introduced a denoised multitaper spectral estimation method to perform frequency-domain HRV analysis, by extending ideas developed in Chapter 2. Through application to real EKG data, we demonstrated that our proposed denoised estimates are more robust and

reliable than the classical multitaper spectral estimates. In addition, we derived the confidence bounds of the multitaper spectral estimates using two different methods: frequency-domain bootstrapping and the chi-squared approximation. We note that in general the bounds derived from the frequency-domain bootstrapping method are tighter than those given by the chi-square approximation. The confidence intervals thus derived have many potential applications, such as refining the predefined frequency bands (such as the low-frequency band and the high-frequency band) in a principled manner at the subject level, and comparing the frequency-domain components across subjects and different conditions. The methods introduced in this chapter thus facilitate a more accurate and reliable frequency-domain analysis of heart rate variability, which could potentially provide better insights into neurocardiac functionality.

Part II

Signal and Noise Correlation inference

Chapter 5: Direct Inference of Signal and Noise Correlations from Two-Photon Calcium Imaging of Ensemble Neuronal Activity

Neuronal activity correlations are essential in understanding how populations of neurons encode information. Correlations provide insights into the functional architecture and computations carried out by neuronal networks [56–66]. Neuronal activity correlations are often categorized in two groups: *signal* correlations and *noise* correlations [58, 62, 67–70]. Given two neurons, signal correlation quantifies the similarity of neural responses that are time-locked to a repeated stimulus across trials, whereas noise correlation quantifies the stimulus-independent trial-to-trial variability shared by neural responses that are believed to arise from common latent inputs.

Two-photon calcium imaging has become increasingly popular in recent years to record *in vivo* neural activity simultaneously from hundreds of neurons [3–6]. This technology takes advantage of intracellular calcium flux mostly arising from spiking activity and captures calcium signaling in neurons in living animals using fluorescence microscopy. The observed fluorescence traces of calcium concentrations, however, are indirectly related to neuronal spiking activity. Extracting spiking activity from fluorescence traces is a challenging signal deconvolution problem, and has been the focus of active research [15–26].

The most commonly used approach to infer signal and noise correlations from two-photon

data is to directly apply the classical definitions of correlations for firing rates [62], to fluorescence traces [11–14]. However, it is well known that fluorescence observations are noisy and blurred surrogates of spiking activity, because of their dependence on observation noise, calcium dynamics, and the temporal properties of calcium indicators. Due to temporal blurring, the resulting signal and noise correlation estimates are highly biased. An alternative approach is to carry out the inference in a two-stage fashion: first, infer spikes using a deconvolution technique, and then compute firing rates and evaluate the correlations [66, 119–122]. These two-stage estimates are highly sensitive to the accuracy of spike deconvolution, and require high temporal resolution and signal-to-noise ratios [21, 123]. Furthermore, these deconvolution techniques are biased towards obtaining accurate first-order statistics (i.e., spike timings) via spatiotemporal priors, which may be detrimental to recovering second-order statistics (i.e., correlations). Finally, both approaches also undermine the non-linear dynamics of spiking activity as governed by stimuli, past activity and other latent processes [28]. There are a few existing studies that aim at improving the estimation of neuronal correlations, but they either do not consider signal correlations [66, 76], or aim at estimating surrogates of correlations from spikes such as the connectivity/coupling matrix [122, 124–126].

In this chapter, we propose a methodology to *directly* estimate both signal and noise correlations from two-photon imaging observations, without requiring an intermediate step of spike deconvolution. We pose the problem under the commonly used experimental paradigm in which neuronal activity is recorded during trials of a repeated stimulus. We avoid the need to perform spike deconvolution by integrating techniques from point processes and state-space modeling that explicitly relate the signal and noise correlations to the observed fluorescence traces in a multi-tier model. Thus, we cast signal and noise correlations within a parameter estimation setting. To

solve the resulting estimation problem in an efficient fashion, we develop a solution method based on variational inference [71,72], by combining techniques from Pólya-Gamma augmentation [73] and compressible state-space estimation [20, 74, 75]. We also provide theoretical guarantees on the bias and variance performance of the resulting estimator.

We demonstrate the utility of our proposed estimation framework through application to simulated and real data from the mouse auditory cortex during presentations of tones and acoustic noise. In application to repeated trials under spontaneous and stimulus-driven conditions within the same experiment, our method reliably provides noise correlation structures that are invariant across the two conditions. In addition, our joint analysis of signal and noise correlations corroborates existing hypotheses regarding the distinction between their structures [126–128]. Moreover, while the application of our proposed method to spatial analysis of signal and noise correlations in the mouse auditory cortex is consistent with existing work [14], it reveals novel and distinct spatial trends in the correlation structure of layers 2/3 and 4. In summary, our method improves on existing work by: 1) explicitly modeling the fluorescence observation process and the nonlinearities involved in spiking activity, as governed by both the stimulus and latent processes, through a multi-tier Bayesian forward model, 2) joint estimation of signal and noise correlations directly from two-photon fluorescence observations through an efficient iterative procedure, without requiring intermediate spike deconvolution, 3) providing theoretical guarantees on the performance of the proposed estimator, and 4) gaining access to closed-form posterior approximations, with low-complexity and iterative update rules and minimal dependence on training data. Our proposed method can thus be used as a robust and scalable alternative to existing approaches for extracting signal and noise correlations from two-photon imaging data.

5.1 Results

In this section, we first demonstrate the utility of our proposed estimation framework through simulation studies as well as applications on experimentally - recorded data from the mouse auditory cortex (data from the Kanold Lab, Department of Biology, University of Maryland, College Park, MD). Then, we present theoretical performance bounds on the proposed estimator. Before presenting the results, we will give an overview of the proposed signal and noise correlation inference framework, and outline our contributions and their relationship to existing work. For the ease of reproducibility, we have archived a MATLAB implementation of our proposed method in [78], and have deposited the data used in this chapter in the Digital Repository at the University of Maryland [79].

5.1.1 Signal and Noise correlations

We consider a canonical experimental setting in which the same external stimulus, denoted by s_t , is repeatedly presented across L independent trials and the spiking activity of a population of N neurons is indirectly measured using two-photon calcium fluorescence imaging. Figure 5.1 (forward arrow) shows the generative model that is used to quantify this procedure. The fluorescence observation in the l^{th} trial from the j^{th} neuron at time frame t , denoted by $y_{t,l}^{(j)}$, is a noisy surrogate of the intracellular calcium concentrations. The calcium concentrations in turn are temporally blurred surrogates of the underlying spiking activity $n_{t,l}^{(j)}$, as shown in Figure 5.1.

In modeling the spiking activity, we consider two main contributions: 1) the common known stimulus s_t affects the activity of the j^{th} neuron via an unknown kernel $\mathbf{d}_j$, akin to the receptive field; 2) the trial-to-trial variability and other intrinsic/extrinsic neural covariates that

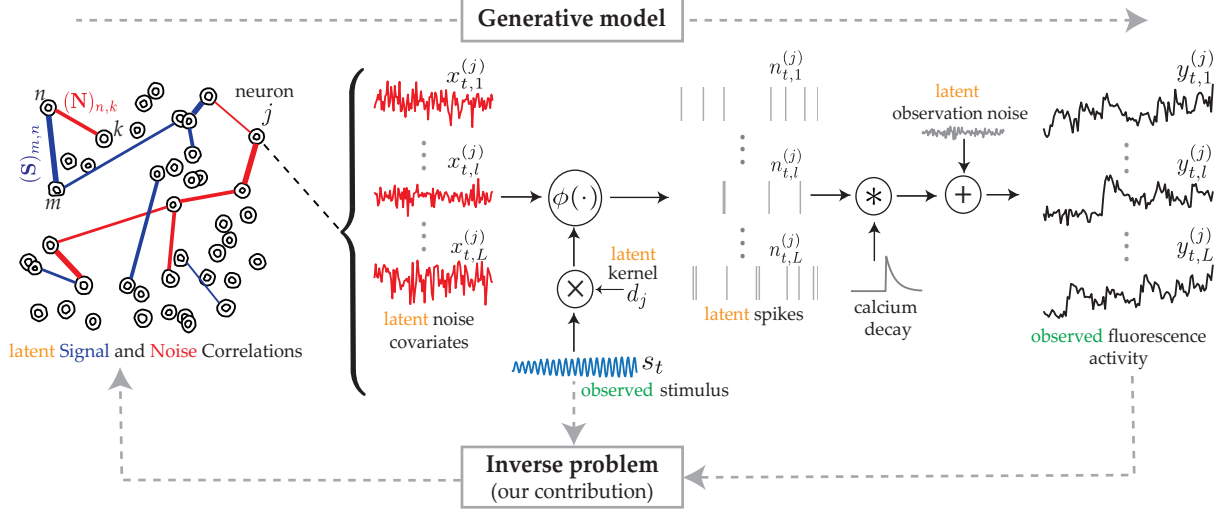


Figure 5.1: The proposed generative model and inverse problem. Observed (green) and latent (orange) variables pertinent to the j^{th} neuron are indicated, according to the proposed model for estimating the signal (blue) and noise (red) correlations from two-photon calcium fluorescence observations. Calcium fluorescence traces $\left(y_{t,l}^{(j)}\right)$ of L trials are observed, in which the repeated external stimulus (s_t) is known. The underlying spiking activity $\left(n_{t,l}^{(j)}\right)$, trial-to-trial variability and other intrinsic/extrinsic neural covariates that are not time-locked with the external stimulus $\left(x_{t,l}^{(j)}\right)$, and the stimulus kernel (d_j) are latent. Our main contribution is to solve the inverse problem: recovering the underlying latent signal (S) and noise (N) correlations directly from the fluorescence observations, without requiring intermediate spike deconvolution.

are not time-locked to the stimulus s_t are captured by a trial-dependent latent process $x_{t,l}^{(j)}$. Then, we use a Generalized Linear Model to link these underlying neural covariates to spiking activity [28]. More specifically, we model spiking activity as a Bernoulli process:

$$n_{t,l}^{(j)} \sim \text{Bernoulli} \left(\phi \left(x_{t,l}^{(j)}, \mathbf{d}_j^\top \mathbf{s}_t \right) \right),$$

where $\phi(\cdot)$ is a mapping function, which could in general be non-linear.

The *signal* correlations aim to measure the correlations in the temporal response that are time-locked to the repeated stimulus, s_t . On the other hand, *noise* correlations in our setting

quantify connectivity arising from covariates that are unrelated to the stimulus, including the trial-to-trial variability [126]. Based on the foregoing model, we propose to formulate the signal $((\Sigma_s)_{i,j})$ and noise $((\Sigma_x)_{i,j})$ covariance between the i^{th} neuron and j^{th} neuron as:

$$(\Sigma_s)_{i,j} := \mathbf{d}_i^\top \text{cov}(\mathbf{s}_t, \mathbf{s}_t) \mathbf{d}_j, \quad (\Sigma_x)_{i,j} := \text{cov}\left(x_{t,l}^{(i)}, x_{t,l}^{(j)}\right), \quad (5.1)$$

where $\text{cov}(\cdot)$ is the empirical covariance function defined as $\text{cov}(\mathbf{u}_t, \mathbf{v}_t) := \frac{1}{T} \sum_{t=1}^T \left(\mathbf{u}_t - \frac{1}{T} \sum_{t=1}^T \mathbf{u}_t \right) \left(\mathbf{v}_t - \frac{1}{T} \sum_{t=1}^T \mathbf{v}_t \right)^\top$, for a total observation duration of T time frames.

Our main contribution is to provide an efficient solution for the so-called inverse problem: direct estimation of Σ_s and Σ_x from the fluorescence observations, without requiring intermediate spike deconvolution (Figure 5.1, backward arrow). The signal and noise correlation matrices, denoted by $\mathbf{S}$ and $\mathbf{N}$, can then be obtained by standard normalization of Σ_s and Σ_x :

$$(\mathbf{S})_{i,j} := \frac{(\Sigma_s)_{i,j}}{\sqrt{(\Sigma_s)_{i,i}(\Sigma_s)_{j,j}}}, \quad (\mathbf{N})_{i,j} := \frac{(\Sigma_x)_{i,j}}{\sqrt{(\Sigma_x)_{i,i}(\Sigma_x)_{j,j}}}, \quad \forall i, j = 1, 2, \dots, N. \quad (5.2)$$

We note that when spiking activity is directly observed using electrophysiology recordings, the conventional signal $((\Sigma_s^{\text{con}})_{i,j})$ and noise $((\Sigma_x^{\text{con}})_{i,j})$ covariances of spiking activity between the i^{th} and j^{th} neuron are defined as [62]:

$$\begin{aligned} (\Sigma_s^{\text{con}})_{i,j} &:= \text{cov}\left(\frac{1}{L} \sum_{l=1}^L n_{t,l}^{(i)}, \frac{1}{L} \sum_{l=1}^L n_{t,l}^{(j)}\right), \\ (\Sigma_x^{\text{con}})_{i,j} &:= \frac{1}{L} \sum_{l=1}^L \text{cov}\left(n_{t,l}^{(i)} - \frac{1}{L} \sum_{l=1}^L n_{t,l}^{(i)}, n_{t,l}^{(j)} - \frac{1}{L} \sum_{l=1}^L n_{t,l}^{(j)}\right), \end{aligned} \quad (5.3)$$

which after standard normalization in Eq. (5.2) give the conventional signal $((\mathbf{S}^{\text{con}})_{i,j})$ and noise

$((\mathbf{N}^{\text{con}})_{i,j})$ correlations. While at first glance our definitions of signal and noise covariances in Eq. (5.1) seem to be a far departure from the conventional ones in Eq. (5.3), we show that the conventional notions of correlation indeed approximate the same quantities as in our definitions:

$$\mathbf{S}^{\text{con}} \approx \mathbf{S} \quad \text{and} \quad \mathbf{N}^{\text{con}} \approx \mathbf{N},$$

under asymptotic conditions (i.e., T and L sufficiently large). We prove this assertion of asymptotic equivalence in Appendix B.1, which highlights another facet of our contributions: our proposed estimators are designed to robustly operate in the regime of finite (and typically small) T and L , aiming for the very same quantities that the conventional estimators could only recover accurately under ideal asymptotic conditions.

5.1.2 Existing methods used for performance comparison

To compare the performance of our proposed method with existing work, we consider three widely available methods for extracting neuronal correlations. In simulation studies, we additionally benchmark these estimates with respect to the known ground truth. The existing methods considered are the following:

5.1.2.1 Pearson Correlations from the Two-Photon Data

In this method, fluorescence observations are assumed to be the direct measurements of spiking activity, and thus empirical Pearson correlations of the two-photon data are used to compute the signal and noise correlations [12–14, 129]. Explicitly, these estimates are obtained by simply replacing $n_{t,l}^{(j)}$ in Eq. (5.3) by $y_{t,l}^{(j)}$, without performing spike deconvolution.

5.1.2.2 Two-stage Pearson Estimation

Unlike the previous method, in this case, spikes are first inferred using a deconvolution technique. Then, following temporal smoothing via a narrow Gaussian kernel the Pearson correlations are computed using the conventional definitions of Eq. (5.3). For spike deconvolution, we primarily used the FCSS algorithm [20]. To also demonstrate the sensitivity of these estimates to the deconvolution technique that is used, we provide a comparison with the f-oopsi deconvolution algorithm [22] in Appendix B.4.1.

5.1.2.3 Two-stage GPFA Estimation

Similar to the previous method, spikes are first inferred using a deconvolution technique. Then, a latent variable model called Gaussian Process Factor Analysis (GPFA) [130] is applied to the inferred spikes to estimate the latent covariates and receptive fields. Based on those estimates, the signal and residual noise correlations are derived through a formulation similar to Eq. (5.1) and Eq. (5.2) [131].

5.1.3 Simulation study 1: Neuronal ensemble driven by external stimulus

We simulated fluorescence observations according to the proposed generative model given in Section 5.3, from an ensemble of $N = 8$ neurons for a duration of $T = 5000$ time frames. We considered $L = 20$ repeated trials driven by the same external stimulus, which we modeled by an autoregressive process (see Section 5.3 for details). Figure 5.2 shows the corresponding estimation results.

The first column of Figure 5.2-A shows the ground truth noise (top) and signal (bottom)

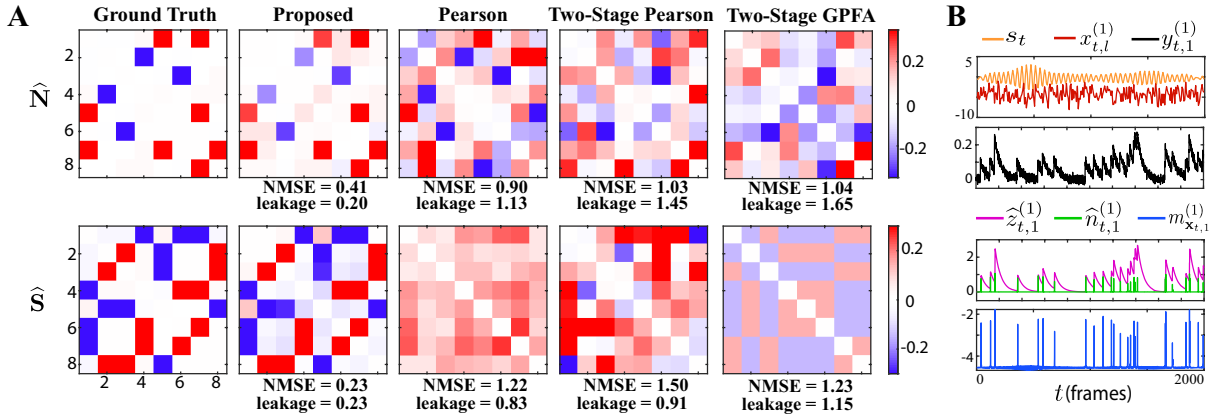


Figure 5.2: Results of simulation study 1. (A) Estimated noise and signal correlation matrices from different methods. Rows from left to right: ground truth, proposed method, Pearson correlations from two-photon recordings, two-stage Pearson estimates, and two-stage GPFA estimates. The normalized mean squared error (NMSE) of each estimate with respect to the ground truth and the leakage effect quantified by the ratio between out-of-network and in-network power (leakage) are indicated below each panel. (B) Simulated external stimulus (orange), latent trial-dependent process (red), fluorescence observations (black), estimated calcium concentrations (purple), putative spikes (green), and estimated mean of the latent state (blue) by the proposed method, for the first trial of neuron 1.

correlations (diagonal elements are all equal to 1 and omitted for visual convenience). The second column shows estimates of the noise and signal correlations using our proposed method, which closely match the ground truth. The third, fourth, and fifth columns, respectively, show the results of the Pearson correlations from the two-photon data, two-stage Pearson, and two-stage GPFA estimation methods. Through a qualitative visual inspection, it is evident that these methods incur high false alarms and mis-detections of the ground truth correlations.

To quantify these comparisons, the normalized mean square error (NMSE) of different estimates with respect to the ground truth are shown below each of the subplots (Figure 5.2-A). Our proposed method achieves the lowest NMSE compared to the others. Furthermore, we observed a significant mixing between signal and noise correlations in these other estimates. To quantify this leakage effect, we first classified each of the correlation entries as in-network or out-of-network, based on being non-zero or zero in the ground truth, respectively (see Section 5.3).

We then computed the ratio between the power of out-of-network components and the power of in-network components as a measure of leakage. The leakage ratios are also reported in Figure 5.2-A. The leakage of our proposed estimates is the lowest of all four techniques, in estimating both the signal and noise correlations. In order to further probe the performance of our proposed method, the external stimulus s_t , latent trial-dependent process $x_{t,1}^{(1)}$, simulated observations $y_{t,1}^{(1)}$, estimated calcium concentration $\hat{z}_{t,1}^{(1)}$, the putative spikes $\hat{n}_{t,1}^{(1)} := \hat{z}_{t,1}^{(1)} - \alpha \hat{z}_{t-1,1}^{(1)}$, and the estimated mean of the latent state $m_{\mathbf{x}_{t,1}}^{(1)}$, for the first trial of the first neuron are shown in Figure 5.2-B. These results demonstrate the ability of the proposed estimation framework to accurately identify the latent processes, which in turn leads to an accurate estimation of the signal and noise correlations as shown in Figure 5.2-B.

The main sources of the observed performance gap between our proposed method and the existing ones are the bias incurred by treating the fluorescence traces as spikes, low spiking rates, and non-linearity of spike generation with respect to intrinsic and external covariates, and sensitivity to spike deconvolution. For the latter, we demonstrated the sensitivity of the two-stage Pearson estimates to the choice of the deconvolution technique in Appendix B.4.1. Furthermore, in order to isolate the effect of said non-linearities on the estimation performance, we applied the two-stage methods to ground truth spikes in Appendix B.4.2. Our analysis showed that both two-stage estimates incur significant estimation errors even if the spikes were recovered perfectly, mainly due to the limited number of trials ($L = 20$ here). In accordance with our theoretical analysis of the asymptotic behavior of the conventional signal and noise correlation estimates given in Appendix B.1, we also showed in Appendix B.4.2 that the performance of the two-stage Pearson estimates based on ground truth spikes, but using $L = 1000$ trials, dramatically improves. Our proposed method, however, was capable of producing reliable estimates with the

number of trials as low as $L = 20$, which is typical in two-photon imaging experiments.

5.1.3.1 Analysis of robustness with respect to modeling assumptions

While the preceding results are quite favorable to our proposed method, the underlying generative models were the same as those used to estimate signal and noise correlations, which is in contrast to conventional real data validation with known ground truth. Access to ground truth correlations in two-photon imaging experimental settings, however, is quite challenging. To further probe the robustness of our proposed method in the absence of ground truth data, we utilized surrogate data that parallel the setting of Figure 5.2 but deviate from our modeling assumptions.

1) *Robustness to stimulus integration model mismatch.* First, we considered surrogate data generated with a non-linear stimulus integration model by replacing the linear receptive field component $\mathbf{d}_j^\top \mathbf{s}_t$ with $\mathbf{d}_j^\top \mathbf{s}_t + (\tilde{\mathbf{d}}_{j,1}^\top \mathbf{s}_t)^2 + (\tilde{\mathbf{d}}_{j,2}^\top \mathbf{s}_t)^2$, where $\tilde{\mathbf{d}}_{j,1}$ and $\tilde{\mathbf{d}}_{j,2}$ are akin to *quadratic* receptive field components. We assumed a linear stimulus integration model in our estimation framework (i.e., $\tilde{\mathbf{d}}_{j,1} = \tilde{\mathbf{d}}_{j,2} = \mathbf{0}$). Appendix B.4.3 shows the resulting correlation estimates. While the performance of our proposed signal correlation estimates degrade under this setting as compared to Figure 5.2, our proposed estimates still outperform existing methods. In addition, the model mismatch in the stimulus integration component does not affect the accuracy of noise correlation estimation in our method.

2) *Robustness to calcium decay model mismatch.* Next, we tested our proposed estimation framework on data simulated with a different calcium decay model. Specifically, we simulated data with second-order autoregressive calcium dynamics, and at a lower signal-to-noise ratio

(SNR) compared to the setting of Figure 5.2, and used our inference framework which assumes first-order calcium dynamics for estimation. Appendix B.4.4 shows the corresponding noise and signal correlations estimated by the proposed method under these conditions. Even though the performance slightly degrades (in terms of NMSE and leakage ratio), our method can recover the underlying correlations faithfully under this setting.

3) *Robustness to SNR level and firing rate.* Next, we compared the performance of Pearson and Two-Stage Pearson methods with our proposed method under varying SNR levels and average firing rates, as shown in Appendix B.4.5. While the performance of all methods degrades at low SNR levels or firing rates ($\text{SNR} < 10$ dB, firing rate < 0.5 Hz), our proposed method outperforms the existing methods for a wide range of SNR and firing rate values. To quantify this comparison, we have also indicated the mean and standard deviation of the relative performance gain of our proposed estimates across SNR levels and firing rates as insets in Figure B.5.

4) *Robustness to observation noise model mismatch.* Finally, we repeated the foregoing comparisons under varying SNR levels and firing rates, only now we included an additional observation noise model mismatch (Appendix B.4.6). Similar to the treatment in [15], we considered two temporally correlated observation noise models: white noise with a low-frequency drift (Figure B.6, top panels) and pink noise (Figure B.6, bottom panels). In accordance with the results in Appendix B.4.5, our proposed method outperforms the existing ones for a wide range of SNR and firing rate values and under both observation noise model mismatch conditions. From Figure B.6-C and F, it can be observed that the ground truth spikes are favorably recovered as a byproduct of our method, even though the estimated calcium concentrations are contaminated by the temporally correlated fluctuations in observation noise. This in turn results in accurate signal and noise correlation estimates.

5.1.4 Simulation study 2: Spontaneous activity

Next, we present the results of a simulation study in the absence of external stimuli (i.e. $s_t = 0$), pertaining to the spontaneous activity condition. It is noteworthy that the proposed method can readily be applied to estimate noise correlations during spontaneous activity, by simply setting the external stimulus s_t and the receptive field d_j to zero in the update rules (see Section 5.3 for details). We simulated the ensemble spiking activity based on a Poisson process [33] using a discrete time-rescaling procedure [33, 132] so that the data are generated using a different model than that used in our inference framework (i.e., Bernoulli process with a logistic link as outlined in Section 5.3). As such, we eliminated potential performance biases in favor of our proposed method by introducing the aforementioned model mismatch. We simulated $L = 20$ independent trials of spontaneous activity of $N = 30$ neurons, observed for a time duration of $T = 5000$ time frames. The number of neurons in this study is notably larger than that used in the previous one, to examine the scalability of our proposed approach with respect to the ensemble size.

Figure 5.3 shows the comparison of the noise correlation matrices estimated by our proposed method, Pearson correlations from two-photon recordings, two-stage Pearson, and two-stage GPFA estimates, with respect to the ground truth. The Pearson and the two-stage estimates are highly variable and result in excessive false detections. Our proposed estimate, however, closely follows the ground truth, which is also reflected by the comparatively lower NMSE and leakage ratios, despite the mismatch between the models used for data generation and inference. It is noteworthy that the proposed method exhibits favorable scaling with respect to the ensemble size, thanks to the underlying low-complexity variational updates (see Section 5.3).

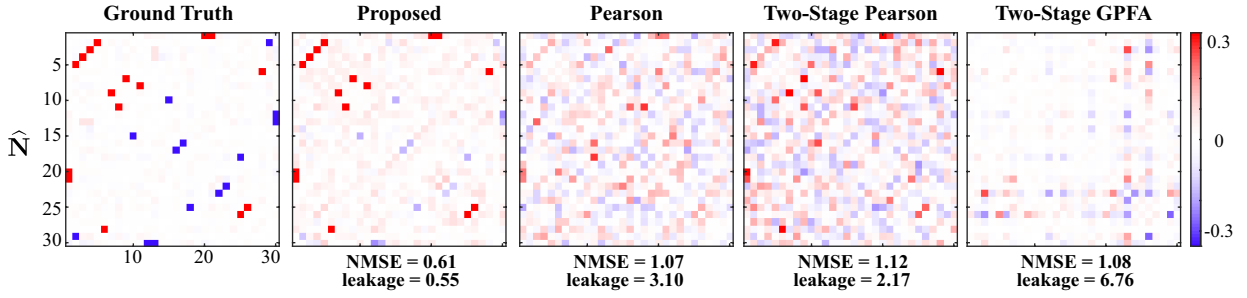


Figure 5.3: Results of simulation study 2. Estimated noise correlation matrices using different methods based on spontaneous activity data. Rows from left to right: ground truth, proposed method, Pearson correlations from two-photon recordings, two-stage Pearson, and two-stage GPFA estimates. The normalized mean squared error (NMSE) of each estimate with respect to the ground truth and the ratio between out-of-network power and in-network power (leakage) are shown below each panel.

5.1.5 Real data study 1: Mouse auditory cortex under random tone presentation

We next applied our proposed method to experimentally recorded two-photon observations from the mouse primary auditory cortex (A1). The dataset consisted of recordings from 371 excitatory neurons in layer 2/3 A1, from which we selected $N = 16$ responsive neurons (i.e., neurons that exhibited at least one spiking event in at least half of the trials considered; see Section 5.3.7). A random sequence of four tones was presented to the mouse, with the same sequence being repeated for $L = 10$ trials. Each trial consisted of $T = 3600$ time frames, and each tone was two seconds long followed by a four-second silent period (see Section 5.3.10 for details). We considered an integration window of $R = 25$ frames for stimulus encoding (see Section 5.3.7 for details). The comparison of the noise and signal correlation estimates obtained by our proposed method, Pearson correlations from two-photon recordings, two-stage Pearson and two-stage GPFA methods is shown in Figure 5.4-A. The spatial map of the 16 neurons considered in the analysis in the field of view is shown in Figure 5.4-B. Figure 5.4-C shows the stimulus tone

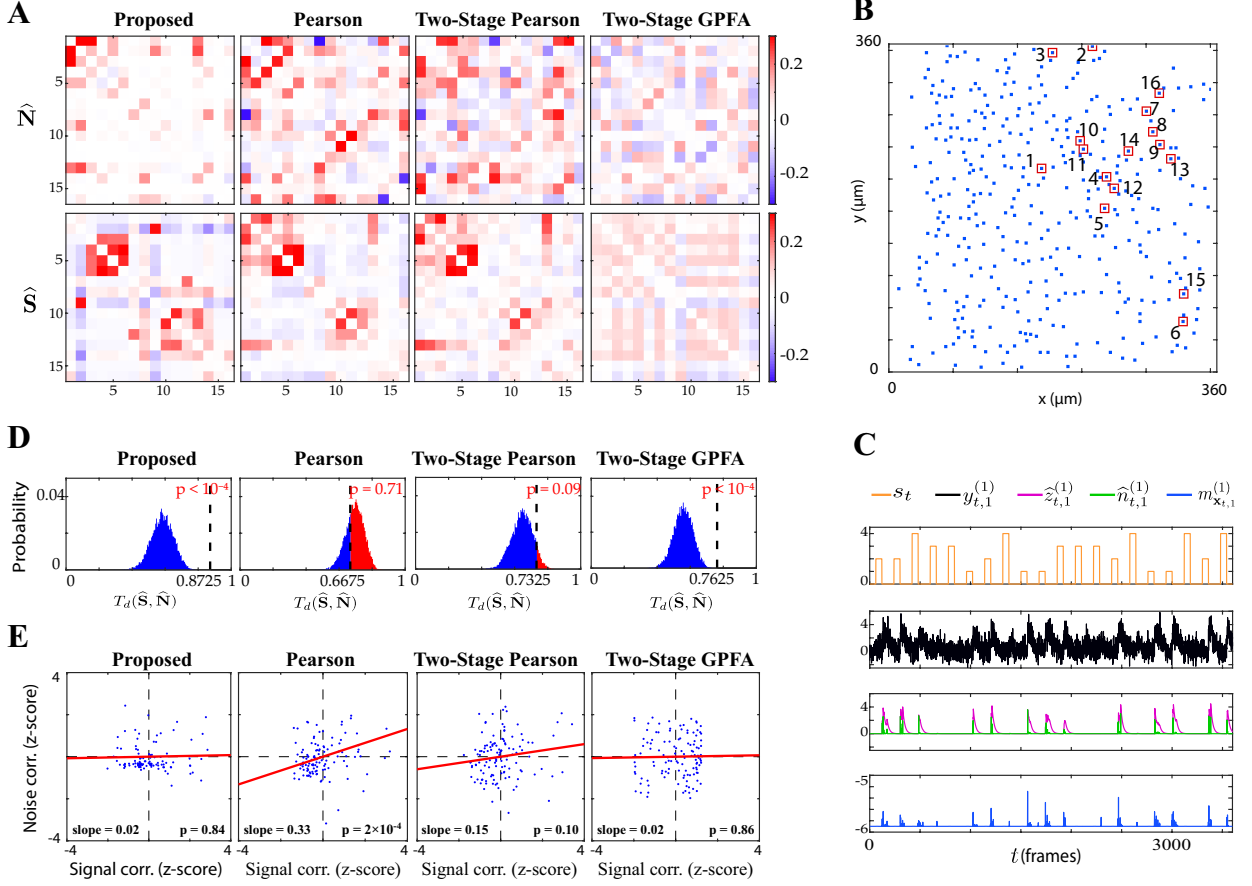


Figure 5.4: Application to experimentally-recorded data from the mouse A1. (A) Estimated noise (top) and signal (bottom) correlation matrices using different methods. Rows from left to right: proposed method, Pearson correlations from two-photon data, two-stage Pearson, and two-stage GPFA estimates. (B) Location of the selected neurons with the highest activity in the field of view. (C) Presented tone sequence (orange), observations (black), estimated calcium concentrations (purple), putative spikes (green), and estimated mean latent state (blue) in the first trial of the first neuron. (D) Null distributions of chance occurrence of dissimilarities between signal and noise correlation estimates using different methods. The observed test statistic in each case is indicated by a dashed vertical line. (E) Scatter plots of signal vs. noise correlations for individual cell pairs (blue dots) corresponding to each method. Data were normalized for comparison by computing z-scores. For each case, the linear regression model fit is shown in red, and the slope and p-value of the t-test are indicated as insets.

sequence s_t , two-photon observations $y_{t,1}^{(1)}$, estimated calcium concentration $\hat{z}_{t,1}^{(1)}$, putative spikes $\hat{n}_{t,1}^{(1)} := \hat{z}_{t,1}^{(1)} - \alpha \hat{z}_{t-1,1}^{(1)}$ and the estimated mean of the latent state $m_{\mathbf{x}_{t,1}}^{(1)}$, for the first trial of the first neuron.

We estimated the Best Frequency (BF) of each neuron as the tone that resulted in the highest level of fluorescence activity. The results in Figure 5.4-A are organized such that the neurons with the same BF are neighboring, with the BF increasing along the diagonal. Thus, expectedly [129] our proposed method, as well as the Pearson and two-stage Pearson estimates show high signal correlations along the diagonal. However, the two-stage GPFA estimates do not reveal such a structure. By visual inspection, as also observed in the simulation studies, the Pearson correlations from two-photon recordings, two-stage Pearson, and two-stage GPFA estimates have significant leakage between the signal and noise correlations, whereas our proposed signal and noise correlation estimates in Figure 5.4-A suggest distinct spatial structures.

To quantify this visual comparison, we used a statistic based on the Tanimoto similarity metric [133], denoted by $T_s(\mathbf{X}, \mathbf{Y})$ for two matrices $\mathbf{X}$ and $\mathbf{Y}$. As a measure of dissimilarity, we used $T_d(\mathbf{X}, \mathbf{Y}) := 1 - T_s(\mathbf{X}, \mathbf{Y})$ (see Section 5.3). The comparison of $T_d(\hat{\mathbf{S}}, \hat{\mathbf{N}})$ for the four estimates is presented in the second column of Table 5.1. To assess statistical significance, for each comparison we obtained null distributions corresponding to chance occurrence of dissimilarities using a shuffling procedure as shown in Figure 5.4-D, and then computed one-tailed p -values from those distributions (see Section 5.3 for details). Table 5.1 and Figure 5.4-D includes these p -values, which show that the proposed estimates (boldface numbers in Table 5.1, second column) indeed have the highest dissimilarity between signal and noise correlations. The higher leakage effect in the other three estimates is also reflected in their smaller $T_d(\hat{\mathbf{S}}, \hat{\mathbf{N}})$ values. It is also noteworthy that while both our proposed and the two-stage GPFA estimates indicate that this dissimilarity is statistically significant (p -values in Table 5.1 and Figure 5.4-D), the dissimilarity metric statistic of our estimates is considerably larger than that of the two-stage GPFA estimates.

To further investigate this effect, we have depicted the scatter plots of signal vs. noise

Table 5.1: Dissimilarity metric statistics for the estimates in Figure 5.4-A (also illustrated in Figure 5.4-D) and linear regression statistics of the comparison between signal and noise correlations in Figure 5.4-E.

Estimate	Dissimilarity $T_d(\hat{\mathbf{S}}, \hat{\mathbf{N}})$ (Figure 5.4-D)	Regression statistics (Figure 5.4-E)	
		slope (p-value)	R^2 value
Proposed	0.8725 ($p < 10^{-4}$)	0.02 ($p = 0.84$)	4×10^{-4}
Pearson	0.6675 ($p = 0.71$)	0.33 ($p = 2 \times 10^{-4}$)	0.11
Two-Stage Pearson	0.7325 ($p = 0.09$)	0.15 ($p = 0.10$)	0.02
Two-Stage GPFA	0.7625 ($p < 10^{-4}$)	0.02 ($p = 0.86$)	3×10^{-4}

correlations estimated by each method in Figure 5.4-E. To examine the possibility of the leakage effect on a pairwise basis, we performed linear regression in each case. The slope of the model fit, the p-value for the corresponding t-test and the R^2 values are reported in the third and fourth columns of Table 5.1 (the slope and p-values are also shown as insets in Figure 5.4-E). Consistent with the results of [14], the Pearson estimates suggest a significant correlation between the signal and noise correlation pairs (as indicated by the higher slope in Figure 5.4-E). However, it is noteworthy that none of the other estimates (including the proposed estimates) in Figure 5.4-E register a significant trend between signal and noise correlations. This further corroborates our assessment of the high leakage between signal and noise correlations in Pearson estimates since such a leakage effect could result in overestimation of the trend between the signal and noise correlation pairs. It is noteworthy that the signal and noise correlations estimated by the proposed, the two-stage Pearson and the two-stage GPFA methods show no such pairwise trend, suggesting distinct patterns of stimulus-dependent and stimulus-independent functional connectivity [13, 60, 63, 126].

A key advantage of our proposed method over the Pearson and two-stage approaches is the explicit modeling of stimulus integration. The relevant parameter in this regard is the length of the stimulus integration window R . While in our simulation studies the value of R was known,

it needs to be set by the user in real data applications. To this end, domain knowledge or data-driven methods such as cross-validation and model order selection can be utilized (see Section 5.3.7 for details). Noting that the number of parameters to be estimated linearly scales with R , it must be chosen large enough to capture the stimulus effects, yet small enough to result in favorable computational complexity. Here, given that the typical tone response duration of mouse A1 neurons is < 1 s [134–136], with a sampling frequency of $f_s = 30$ Hz, we surmised that a choice of $R \sim 30$ suffices to capture the stimulus effects. We further examined the effect of varying R on the proposed correlation estimates in Appendix B.5.1. As shown, small values of R (e.g., $R = 1$ or 10) may not be adequate to fully capture stimulus integration effects. By considering values of R in the range $25 - 50$, we observed that the correlation estimates remain stable. We thus chose $R = 25$ for our analysis.

Careful inspection of the second panel in Figure 5.4-C shows that the fluorescence activity often saturates to ~ 4 times its baseline value. This effect is due to successive closely-spaced spikes, which implies the occurrence of more than one spike per frame and thus violates our Bernoulli modeling assumption. To inspect the performance of our method more carefully under this scenario, we show in Appendix B.5.2 a zoomed-in view of the estimated latent processes $\hat{z}_{t,1}^{(1)}$ (calcium concentration) and $\hat{n}_{t,1}^{(1)}$ (putative spikes) for a sample data segment with high fluorescence activity. The estimated latent processes reveal two mechanisms leveraged by our inference method to mitigate the aforementioned model mismatch: first, our proposed method predicts spiking events in adjacent time frames to compensate for the rapid increase in firing rate and thus infers calcium concentration levels that match the observed fluorescence; secondly, even though our generative model assumes that there is only one spiking event in a given time frame, this restriction is mitigated in our inference framework by relaxing the constraint

$\hat{n}_{t,l}^{(j)} := \hat{z}_{t,l}^{(j)} - \alpha \hat{z}_{t-1,l}^{(j)} \leq 1$, as explained in Section 5.3.6. While this relaxation was performed for the sake of tractability of the inverse solution, it in fact leads to improved estimation results under episodes of rapid increase in firing rate, by allowing the putative spike magnitudes $\hat{n}_{t,l}^{(j)}$ to be greater than 1. The latter is evident in the magnitude of the inferred spikes in Figure B.8, following the rise of fluorescence activity.

Given that the ground truth correlations are not available for a direct comparison, we instead performed a test of specificity that reveals another key limitation of existing methods. Fluorescence observations exhibit structured dynamics due to the exponential intracellular calcium concentration decay (as shown in Figure 5.4-C, for example), which are in turn related to the underlying spikes that are driven non-linearly by intrinsic/extrinsic stimuli as well as the properties of the indicator used. As such, an accurate inference method is expected to be specific to this temporal structure. To test this, we randomly shuffled the T time frames consistently in the same order in all trials, to fully break the temporal structure governing calcium decay dynamics, and then estimated correlations from these shuffled data using the different methods. The resulting estimates of noise correlations are shown in Figure 5.5-A for one instance of such shuffled data. The average NMSE for a total of 50 shuffled samples with respect to the original un-shuffled estimates (in Figure 5.4-A) are tabulated in Table 5.2, and are also indicated below each panel in Figure 5.5-A.

A visual inspection of Figure 5.5-A shows that the Pearson correlations from two-photon recordings expectedly remain unchanged. Since this method treats each time frame to be independent, temporal shuffling does not impact the correlations in any way. On the other extreme, both of the two-stage estimates seem to detect highly variable and large correlation values, despite operating on data that lacks any relevant temporal structure. Our proposed method, however,

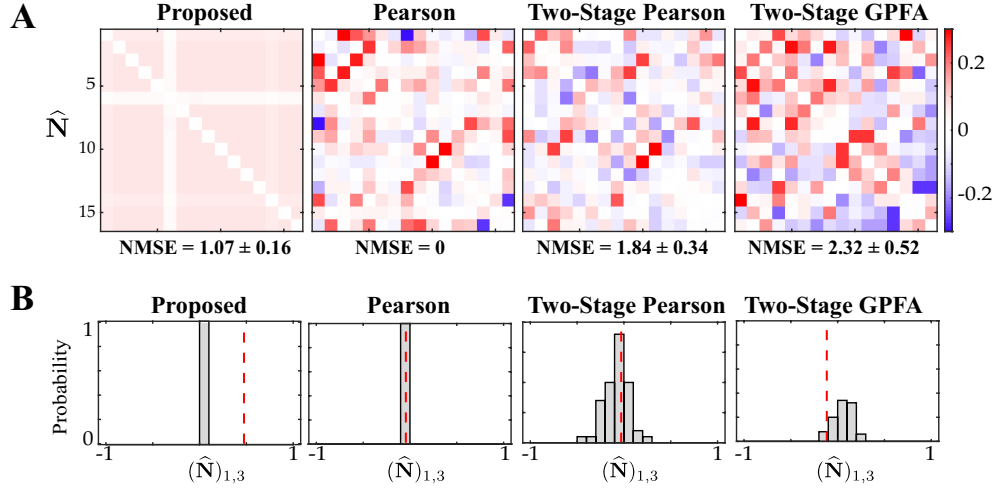


Figure 5.5: Assessing the specificity of different estimation results shown in Figure 5.4. Rows from left to right: proposed method, Pearson correlations from two-photon data, two-stage Pearson, and two-stage GPFA estimates. (A) The estimated noise correlations using different methods after random temporal shuffling of the observations. The mean and standard deviation of the NMSE across 50 trials are indicated below each panel. (B) Histograms of the noise correlation estimates between the first and third neurons over the 50 temporal shuffling trials. The estimate based on the original (un-shuffled) data in each case is indicated by a dashed vertical line.

remarkably produces negligible correlation estimates. Although both the two-stage and proposed estimates show variability with respect to the shuffled data (Table 5.2), the standard deviations of the NMSE values of our proposed method are considerably smaller than those of the two-stage methods (Table 5.2). For further inspection, the histograms of a single element $((\hat{\mathbf{N}})_{1,3})$ of the estimated correlation matrices across the 50 shuffling trials are shown in Figure 5.5-B. The original un-shuffled estimates are marked by the dashed vertical lines in each case. The proposed estimate in Figure 5.5-B is highly concentrated around zero, even though the un-shuffled estimate is non-zero. However, the two-stage estimates produce correlations that are widely variable across the shuffling trials. This analysis demonstrates that our proposed method is highly specific to the temporal structure of fluorescence observations, whereas the Pearson correlations from two-photon recordings, two-stage Pearson, and two-stage GPFA methods fail to be specific.

Table 5.2: The average NMSE across 50 trials used in the shuffling procedure illustrated in Figure 5.5-A.

Estimate	NMSE in $\hat{\mathbf{N}}$	NMSE in $\hat{\mathbf{S}}$
Proposed	1.07 ± 0.16	1.32 ± 0.19
Pearson	0	0
Two-Stage Pearson	1.84 ± 0.34	0.55 ± 0.12
Two-Stage GPFA	2.32 ± 0.52	2.26 ± 0.51

5.1.6 Real data study 2: Spontaneous vs. stimulus-driven activity in the mouse

A1

To further validate the utility of our proposed methodology, we applied it to another experimentally recorded dataset from the mouse A1 layer 2/3. This experiment pertained to trials presenting a sequence of short white-noise stimuli, randomly interleaved with silent trials of the same duration. Figure 5.6-A shows a sample trial sequence. The two-photon recordings thus contained episodes of stimulus-driven and spontaneous activity (see Section 5.3.10 for details). Under this experimental setup, it is expected that the noise correlations are invariant across the spontaneous and stimulus-driven conditions. By the foregoing results of real data study 1, we also expect the signal and noise correlation patterns to be distinct. Each trial considered in the analysis consisted of $T = 765$ frames (see Section 5.3.10 for details). We selected $N = 10$ responsive neurons (according to the criterion described in Section 5.3.7), each with $L = 10$ trials. Similar to real data study 1, we chose a stimulus integration window of length $R = 25$ frames.

Figure 5.6-B shows the resulting noise and signal correlation estimates under the spontaneous ($\hat{\mathbf{N}}_{\text{spn}}$, top) and stimulus-driven ($\hat{\mathbf{N}}_{\text{stim}}$ and $\hat{\mathbf{S}}_{\text{stim}}$, bottom) conditions. Figure 5.6-C shows the spatial map of the 10 neurons considered in the analysis in the field of view. A visual inspection of the first column of Figure 5.6-B indeed suggests that $\hat{\mathbf{N}}_{\text{spn}}$ and $\hat{\mathbf{N}}_{\text{stim}}$ are saliently

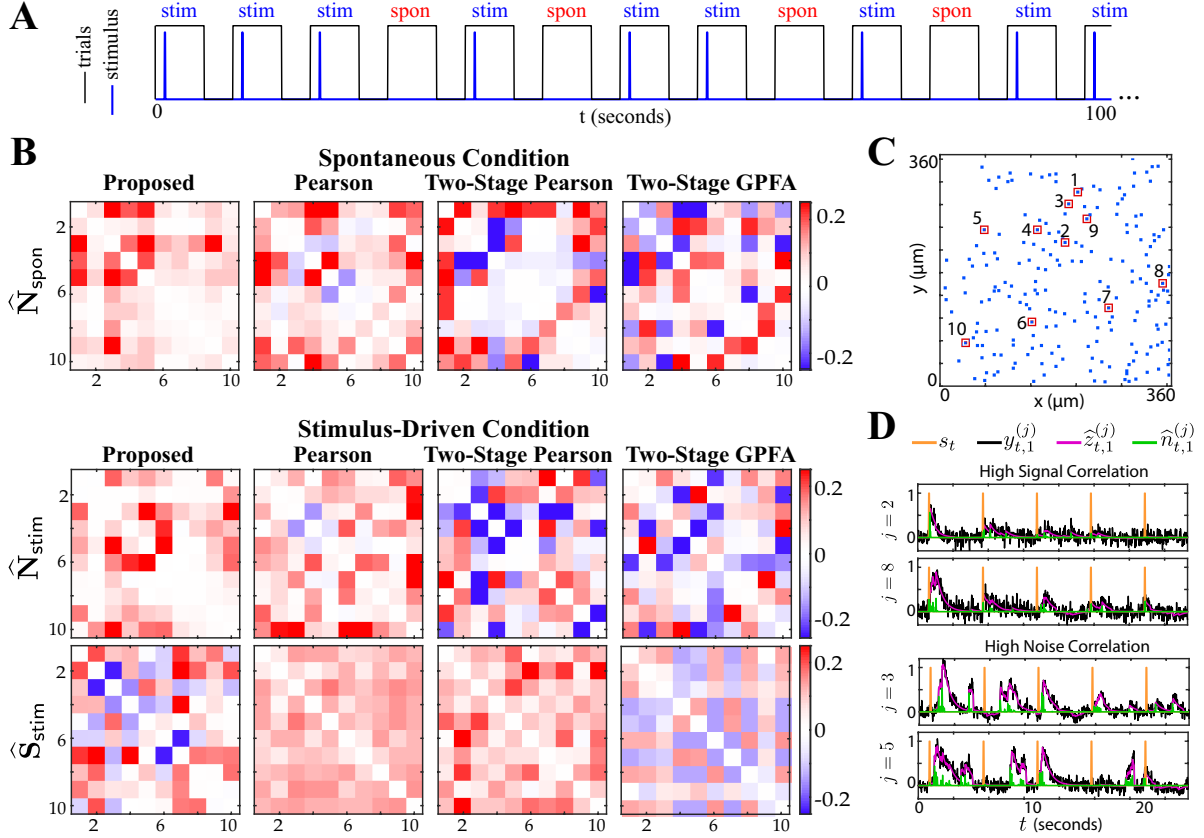


Figure 5.6: Comparison of spontaneous and stimulus-driven activity in the mouse A1. (A) A sample trial sequence in the experiment. Stimulus-driven (stim) trials were recorded with randomly interleaved spontaneous (spon) trials of the same duration. (B) Estimated noise and signal correlation matrices under spontaneous (top) and stimulus-driven (bottom) conditions. Rows from left to right: proposed method, Pearson correlations from two-photon data, two-stage Pearson, and two-stage GPFA estimates. (C) Location of the selected neurons with the highest activity in the field of view. (D) Stimulus onsets (orange), observations (black), estimated calcium concentrations (purple), and putative spikes (green) for the first trial from two pairs of neurons with high signal correlation (top) and high noise correlation (bottom), as identified by the proposed estimates.

similar, and distinct from $\hat{\mathbf{S}}_{\text{stim}}$. The Pearson correlations obtained from two-photon data (second column) and the two-stage Pearson and GPFA estimates (third and fourth columns, respectively), however, evidently lack this structure. As in the previous study, we quantified this visual comparison using the similarity metric $T_s(\mathbf{X}, \mathbf{Y})$ and the dissimilarity metric $T_d(\mathbf{X}, \mathbf{Y})$ (see Section 5.3 for details). These statistics are reported in Table 5.3 along with the p -values (null distributions are shown in Appendix B.6.1), which show that the only significant outcomes (boldface numbers)

Table 5.3: Similarity/dissimilarity metric statistics for the estimates in Figure 5.6.

Estimation Method	$T_s(\hat{\mathbf{N}}_{\text{spon}}, \hat{\mathbf{N}}_{\text{stim}})$	$T_d(\hat{\mathbf{S}}_{\text{stim}}, \hat{\mathbf{N}}_{\text{stim}})$
Proposed	0.5716 ($p = 0.003$)	0.7946 ($p = 0.004$)
Pearson	0.3031 ($p = 0.61$)	0.5032 ($p = 0.92$)
Two-Stage Pearson	0.2790 ($p = 0.05$)	0.7862 ($p = 0.39$)
Two-Stage GPFA	0.2008 ($p = 0.50$)	0.7792 ($p = 0.22$)

are those of our proposed method. While it is expected from the experiment design for the noise correlations under the two settings to be similar, the only method that detects this expected outcome with statistical significance is our proposed method. Moreover, the statistically significant dissimilarity between the signal and noise correlations of our proposed estimates corroborate the hypothesis that signal and noise are encoded by distinct functional networks [13, 60, 63, 126].

Furthermore, Figure 5.6-D shows the time course of the stimulus, observations, estimated calcium concentrations, and putative spikes for the first trial from two pairs of neurons with high signal correlation ($j = 2, 8$, top) and high noise correlation ($j = 3, 5$, bottom). As expected, the putative spiking activity of the neurons with high signal correlation (top) is closely time-locked to the stimulus onsets. The activity of the two neurons with high noise correlation (bottom), however, is not time-locked to the stimulus onsets, even though the two neurons exhibit highly correlated activity. The correlations estimated via the proposed method thus encode substantial information about the inter-dependencies of the spiking activity of the neuronal ensemble.

5.1.7 Real data study 3: Spatial analysis of signal and noise correlations in the mouse A1

Lastly, we applied our proposed method to examine the spatial distribution of signal and noise correlations in the mouse A1 layers 2/3 and 4 (data from [129]). The dataset included

fluorescence activity recorded during multiple experiments of presenting sinusoidal amplitude-modulated tones, with each stimulus being repeated across several trials (see Section 5.3.10 and [129] for experimental details). In each experiment, we selected on average around 20 responsive neurons for subsequent analysis (according to the criterion described in Section 5.3.7). For brevity, we compare the estimates of signal and noise correlations using our proposed method only with those obtained by Pearson correlations from the two-photon data. The latter method was also used in previous analyses of data from this experimental paradigm [14].

In parallel to the results reported in [14], Figure 5.7-A and Figure 5.7-B illustrate the correlation between the signal and noise correlations in layers 2/3 and 4, respectively. Consistent with the results of [14], the signal and noise correlations exhibit positive correlation in both layers, regardless of the method used. However, the correlation coefficients (i.e., slopes in the insets) identified by our proposed method are notably smaller than those obtained from Pearson correlations, in both layer 2/3 (Figure 5.7-A) and layer 4 (Figure 5.7-B). Comparing this result with our simulation studies suggests that the stronger linear trend between the signal and noise correlations observed using the Pearson correlation estimates is likely due to the mixing between the estimates of signal and noise correlations. As such, our method suggests that the signal and noise correlations may not be as highly correlated with one another as indicated in previous studies of layers 2/3 and 4 in mouse A1.

Next, to evaluate the spatial distribution of signal and noise correlations, we plotted the correlation values for pairs of neurons as a function of their distance for layer 2/3 (Figure 5.7-C) and layer 4 (Figure 5.7-D). The distances were discretized using bins of length $10\ \mu m$. The scatter of the correlations along with their median at each bin are shown in all panels. Then, to examine the spatial trend of the correlations, we performed linear regression in each case. The

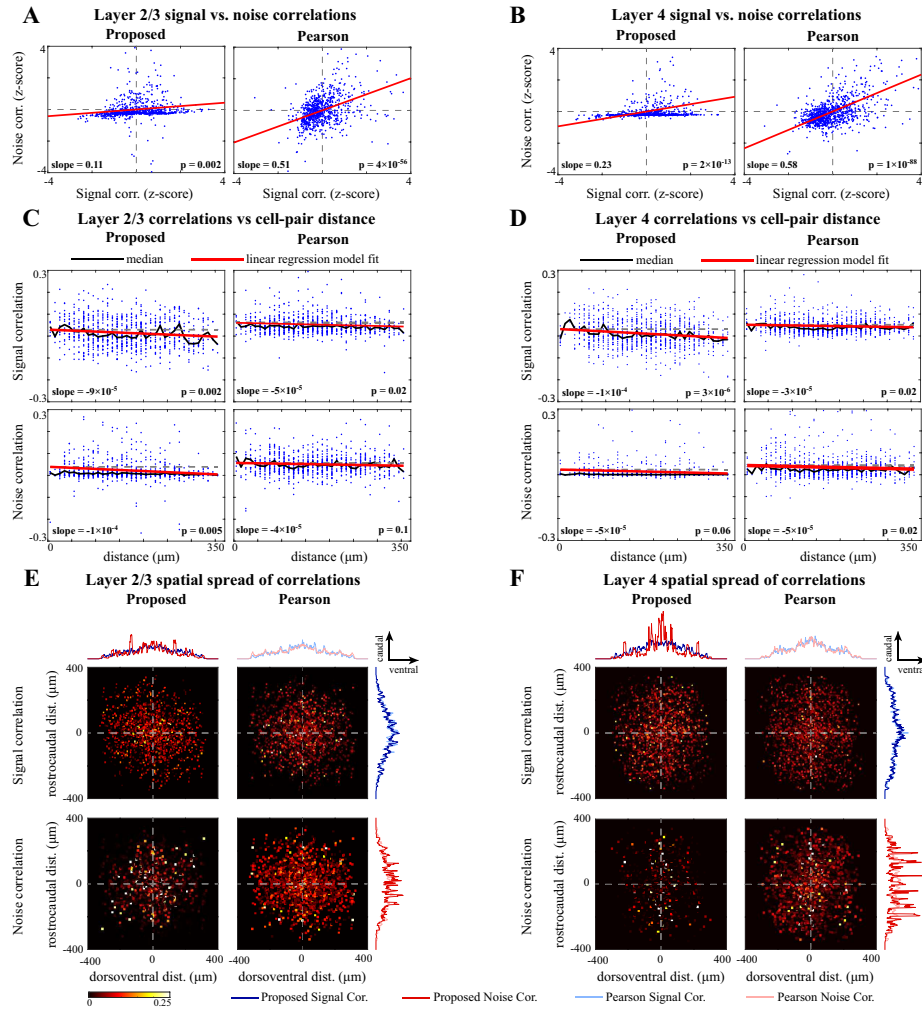


Figure 5.7: Comparison of signal and noise correlations across layers 2/3 and 4. (A) Scatter-plot of noise vs. signal correlations (blue) for individual cell pairs in layer 2/3, based on the proposed (left) and Pearson estimates (right). Data were normalized for comparison by computing z-scores. The linear model fits are shown in red, and the slope and p-value of the t-tests are indicated as insets. Panel B corresponds to layer 4 in the same organization as panel A. (C) Signal (top) and noise (bottom) correlations vs. cell-pair distance in layer 2/3, based on the proposed (left) and Pearson estimates (right). Distances were binned to $10 \mu m$ intervals. The median of the distributions (black) and the linear model fit (red) are shown in each panel. The slope of the linear model fit, and the p-value of the t-test are also indicated as insets. Dashed horizontal lines indicate the zero-slope line for ease of visual comparison. Panel D corresponds to layer 4 in the same organization as panel C. (E) Spatial spread of signal (top) and noise (bottom) correlations in layer 2/3, based on the proposed (left) and Pearson estimates (right). The horizontal and vertical axes in each panel respectively represent the relative dorsoventral and rostrocaudal distances between each cell pair, and the heat-map indicates the magnitude of correlations. Marginal distributions of the signal (blue) and noise (red) correlations along the dorsoventral and rostrocaudal axes for the proposed method (darker colors) and Pearson method (lighter colors) are shown at the top and right sides of the sub-panels. Panel F corresponds to layer 4 in the same organization as panel E.

Table 5.4: Linear regression statistics for the analysis of correlations vs. cell-pair distance

Correlations	Statistics of layer 2/3 correlations		Statistics of layer 4 correlations	
	slope (p-value)	R ² value	slope (p-value)	R ² value
Proposed Signal Corr.	-9×10^{-5} ($p = 0.002$)	0.012	-1×10^{-4} ($p = 3 \times 10^{-6}$)	0.023
Pearson Signal Corr.	-5×10^{-5} ($p = 0.02$)	0.007	-3×10^{-5} ($p = 0.02$)	0.005
Proposed Noise Corr.	-1×10^{-4} ($p = 0.005$)	0.010	-5×10^{-5} ($p = 0.06$)	0.004
Pearson Noise Corr.	-4×10^{-5} ($p = 0.1$)	0.003	-5×10^{-5} ($p = 0.02$)	0.005

slope of the model fit, the p-value for the corresponding t-test and the R² values are reported in Table 5.4 (the slope and p-values are also shown as insets in Figure 5.7-C & D).

From Table 5.4 and Figure 5.7-C & D (upper panels), it is evident that the signal correlations show a significant negative trend with respect to distance, using both methods and in both layers. However, the slope of these negative trends identified by our method (boldface numbers in Table 5.4) is notably steeper than those identified by Pearson correlations. On the other hand, the trends of the noise correlations with distance (bottom panels) are different between our proposed method and Pearson correlations: our proposed method shows a significant negative trend in layer 2/3, but not in layer 4, whereas the Pearson correlations of the two-photon data suggest a significant negative trend in layer 4, but not in layer 2/3. In addition, the slopes of these negative trends identified by our method (boldface numbers in Table 5.4) are steeper than or equal to those identified by Pearson correlations.

Our proposed estimates indicate that noise correlations are sparser and less widespread in layer 4 (Figure 5.7-D) than in layer 2/3 (Figure 5.7-C). To further investigate this observation, we depicted the two-dimensional spatial spread of signal and noise correlations in both layers and for both methods in Figure 5.7-E & F, by centering each neuron at the origin and overlaying the individual spatial spreads. The horizontal and vertical axes in each panel respectively represent the relative dorsoventral and rostrocaudal distances, and the heat maps represent the magnitude

of correlations. Comparing the proposed noise correlation spread in Figure 5.7-E with the corresponding spread in Figure 5.7-F, we observe that the noise correlations in layer 2/3 are indeed more widespread and abundant than in layer 4.

It is notable that the spatial spreads of signal and noise correlations based on the Pearson estimates are remarkably similar in both layers (Figure 5.7-E & F, right panels), whereas they are saliently different for our proposed estimates (Figure 5.7-E & F, left panels). This further corroborates our hypothesis on the possibility of high mixing between the signal and noise correlation estimates obtained by the Pearson correlation of two-photon data. To further examine the differences between the signal and noise correlations, the marginal distributions along the dorsoventral and rostrocaudal axes are shown in Figure 5.7-E & F, selectively overlaid for ease of visual comparison. To quantify the differences between the spatial distributions of signal and noise correlations estimated by each method, we performed Kolmogorov–Smirnov (KS) tests on each pair of marginal distributions, which are summarized in Appendix B.7.1. Although the marginal distributions of signal and noise correlations are significantly different in all cases from both methods, the effect sizes of their difference (KS statistics) are notably higher for our proposed estimates compared to those of the Pearson estimates.

Finally, it is noteworthy that the spatial spreads of correlations for either method and in each layer suggest non-uniform angular distributions with possibly directional bias. To test this effect, we computed the angular marginal distributions and performed KS tests for non-uniformity, which are reported in Appendix B.7.2. These tests indicate that all distributions are significantly non-uniform. In addition, the angular distributions of both signal and noise correlations in layer 4 exhibit salient modes in the rostrocaudal direction, whereas they are less directionally selective in layer 2/3 (Appendix B.7.2).

In summary, the spatial trends identified by our proposed method are consistent with empirical observations of spatially heterogeneous pure-tone frequency tuning by individual neurons in the auditory cortex [14]. The improved correspondence of our proposed method compared to results obtained using Pearson correlations could be the result of the demixing of signal and noise correlations in our method. As a result of the demixing, our proposed method also suggests that noise correlations have a negative trend with distance in layer 2/3, but are much sparser and spatially flat in layer 4. In addition, the spatial spread patterns of signal and noise correlations are more structured and remarkably more distinct for our proposed method than those obtained by the Pearson estimates.

5.1.8 Theoretical analysis of the bias and variance of the proposed estimators

Finally, we present a theoretical analysis of the bias and variance of the proposed estimator. Note that our proposed estimation method has been developed as a scalable alternative to the intractable maximum likelihood (ML) estimation of the signal and noise covariances (see Section 5.3). To benchmark our estimates, we thus need to evaluate the quality of said ML estimates. To this end, we derived bounds on the bias and variance of the ML estimators of the kernel $\mathbf{d}_j$ for $j = 1, \dots, N$ and the noise covariance Σ_x . To simplify the treatment, we posit the following mild assumptions:

Assumption (1). We assume a scalar time-varying external stimulus (i.e. $\mathbf{s}_t = s_t$, and hence $\mathbf{d}_j = d_j$, $\mathbf{d} = [d_1, d_2, \dots, d_N]^\top$). Furthermore, we set the observation noise covariance to be $\Sigma_w = \sigma_w^2 \mathbf{I}$, for notational convenience.

Assumption (2). We derive the performance bounds in the regime where T and L are large,

and thus do not impose any prior distribution on the correlations, which are otherwise needed to mitigate overfitting (see Section 5.3).

Assumption (3). We assume the latent noise process and stimulus to be slowly varying signals, and thus adopt a piece-wise constant model in which these processes are constant within consecutive windows of length W (i.e., $\mathbf{x}_{t,l} = \mathbf{x}_{W_k,l}$ and $s_t = s_{W_k}$, for $(k-1)W + 1 \leq t < kW$ and $k = 1, \dots, K$ with $W_k = (k-1)W + 1$ and $KW = T$) for our theoretical analysis, as is usually done in spike count calculations for conventional noise correlation estimates.

Our main theoretical result is as follows:

Theorem 5.1 (Performance Bounds). *Let $q > \frac{1}{64}$, $0 < \epsilon < 1/2$, and $0 < \eta \leq 1/2$ be fixed constants, $\sigma_m^2 := \max_i (\Sigma_x)_{i,i}$ and $\sigma_s^2 := \frac{1}{K} \sum_{k=1}^K s_{W_k}^2$. Then, under Assumptions (1) - (3), the bias and variance of the maximum likelihood estimators $\hat{\mathbf{d}}$ and $\hat{\Sigma}_x$, conditioned on an event $\mathcal{A}_W$ with $\mathbb{P}(\mathcal{A}_W) \geq 1 - \eta$ satisfy:*

$$\begin{aligned} |\text{bias}_{\mathcal{A}_W}(\hat{d}_j)| &\leq \frac{1}{\sqrt{W^{1-2\epsilon}}} C_1 (2\sigma_w \sqrt{1 + \alpha^2} + 1) + \tau_j, \\ \sqrt{\text{Var}_{\mathcal{A}_W}(\hat{d}_j)} &\leq \sqrt{\frac{(\Sigma_x)_{j,j}}{KL\sigma_s^2(1-\eta)}} + \frac{1}{\sqrt{W^{1-2\epsilon}}} C_2 (2\sigma_w \sqrt{1 + \alpha^2} + 1) + \tilde{\tau}_j, \\ |\text{bias}_{\mathcal{A}_W}((\hat{\Sigma}_x)_{i,j})| &\leq \frac{|(\Sigma_x)_{i,j}|}{KL(1-\eta)} + \sqrt{\frac{\log W}{W^{1-2\epsilon}}} C_3 (14\sigma_w \sqrt{1 + \alpha^2} + 3) + \xi_{i,j}, \\ \sqrt{\text{Var}_{\mathcal{A}_W}((\hat{\Sigma}_x)_{i,j})} &\leq \sqrt{\frac{(KL-1)((\Sigma_x)_{i,j}^2 + (\Sigma_x)_{i,i}(\Sigma_x)_{j,j})}{K^2L^2(1-\eta)}} + \sqrt{\frac{\log W}{W^{1-2\epsilon}}} C_4 (2\sigma_w \sqrt{1 + \alpha^2} + 1) + \tilde{\xi}_{i,j}, \end{aligned}$$

for all $i, j = 1, 2, \dots, N$, if $\log W \geq \max \left\{ \log 2, \frac{\max\{6.25, 4(\|\mu_x\|_\infty + \max_{k,j} \{|s_{W_k} d_j| \})^2\}}{8q\sigma_m^2}, \frac{32\sigma_m^2 q}{\epsilon^2}, \frac{2 \log(64q)}{1-2\epsilon}, \frac{\log(8KLN/\eta)}{q} \right\}$, where τ_j and $\tilde{\tau}_j$ denote bounded terms that are $\mathcal{O}(\sigma_w^2)$ or $\mathcal{O}(\frac{1}{W})$, $\xi_{i,j}$ and $\tilde{\xi}_{i,j}$ denote bounded terms that are $\mathcal{O}(\sigma_w^2)$ or $\mathcal{O}(\frac{1}{W^{1-2\epsilon}})$ and C_1, C_2, C_3 and C_4 are bounded constants given in Appendix B.2.

Proof. The proof of Theorem 5.1 is provided in Appendix B.2. □

To discuss the implications of this theoretical result, several remarks are in order:

5.1.8.1 Remark 1: Achieving near oracle performance

A common benchmark in estimation theory is the performance of the idealistic *oracle* estimator, in which an oracle directly observes the true latent process $\mathbf{x}_{t,l}$ and the true kernel d_j and forms the correlation estimates. In this case, the oracle would incur zero bias and variance of order $\mathcal{O}(1/KL)$ in estimating d_j , and outputs an estimate of Σ_x with bias and variance in the order of $\mathcal{O}(1/KL)$. Theorem 5.1 indeed states that for sufficiently large W and small σ_w , the bias and variance of the ML estimators are arbitrarily close to those of the oracle estimator. Recall that our variational inference framework is in fact a solution technique for the regularized ML problem. Hence, the bounds in Theorem 5.1 provide a benchmark for the expected performance of the proposed estimators, by quantifying the excess bias and variance over the performance of the oracle estimator.

5.1.8.2 Remark 2: Effect of the observation noise and observation duration

As the assumed window of stationarity $W \rightarrow \infty$ (and hence the observation duration $T \rightarrow \infty$), the loss of performance of the proposed estimators only depends on σ_w^2 , the variance of the observation noise. As a result, at a given observation noise variance σ_w^2 , these bounds provide a sufficient upper bound on the time duration of the observations required for attaining the desired level of estimation accuracy. It is noteworthy that σ_w^2 is typically small in practice, as it pertains to the effective observation noise and is significantly diminished by pixel averaging of

the fluorescence traces following cell segmentation.

5.1.8.3 Remark 3: Effect of the number of trials

Finally, note that the bounds in Theorem 5.1 have terms that also drop as the number of trials L grows. These terms pertain to the performance of the oracle estimator. As the number of trials grows ($L \rightarrow \infty$), the oracle estimates become arbitrarily close to the true parameters Σ_x and $\mathbf{d}_j$. Thus, our theoretical performance bounds also provide a sufficient upper bound on the number of trials L required for the oracle estimator to attain the desired level of estimation accuracy.

5.2 Discussion

We developed a novel approach for the joint estimation of signal and noise correlations of neuronal activities directly from two-photon calcium imaging observations and tested our method with experimental data. Existing widely used methods either take the fluorescence traces as surrogates of spiking activity or first recover the unobserved spikes using deconvolution techniques, both followed by computing Pearson correlations or connectivity matrices. As such, they typically result in estimates that are highly biased and are heavily dependent on the choice of the spike deconvolution technique. We addressed these issues by explicitly relating the signal and noise covariances to the observed two-photon data via a multi-tier Bayesian model that accounts for the observation process and non-linearities involved in spiking activity. We developed an efficient estimation framework by integrating techniques from variational inference and state-space estimation. We also established performance bounds on the bias and variance of the proposed es-

timators, which revealed favorable scaling with respect to the observation noise and trial length.

We demonstrated the utility of our proposed estimation framework on both simulated and experimentally-recorded data from the mouse auditory cortex. In our simulation studies, we evaluated the robustness of our proposed method with respect to several model mismatch conditions induced by the stimulus integration model, calcium decay, SNR level, firing rate, and temporally correlated observation noise. In all cases, we observed that our proposed estimates outperform the existing methods in recovering the signal and noise correlations.

There are two main sources for the observed performance gap between our proposed method and existing approaches. The first source is the favorable *soft* decisions on the timing of spikes achieved by our method as a byproduct of the iterative variational inference procedure. Accurate probabilistic decoding of spikes results in better estimates of the signal and noise correlations, and conversely having more accurate estimates of the signal and noise covariances improves the probabilistic characterization of spiking events. This is in contrast with both the Pearson correlations computed from two-photon data and two-stage methods: in computing the Pearson correlations from two-photon data, spike timing is heavily blurred by the calcium decay; in the two-stage methods, erroneous *hard* decisions on the timing of spikes result in biases that propagate to and contaminate the downstream signal and noise correlation estimation and thus result in significant errors.

The second source of performance improvement is the explicit modeling of the non-linear mapping from the stimulus and latent covariates to spiking through a canonical point process model, which is in turn tied to a two-photon observation model in a multi-tier Bayesian fashion. Our theoretical analysis in Theorem 5.1 corroborates that this virtue of our proposed methodology results in robust performance under a limited number of trials. As we have shown in Appendix

B.1, as the number of trials L and trial duration T tends to infinity, conventional notions of signal and noise correlation indeed recover the ground truth signal and noise correlations, as the biases induced by non-linearities average out across trial repetitions. However, as exemplified in Appendix B.4.2, in order to achieve comparable performance to our method using few trials (e.g., $L = 20$), the conventional correlation estimates require considerably more trials (e.g., $L = 1000$).

Application to two-photon data recorded from the mouse primary auditory cortex showed that, unlike the aforementioned existing methods, our estimates provide noise correlation structures that are expectedly invariant across spontaneous and stimulus-driven conditions within an experiment, while producing signal correlation structures that are largely distinct from those given by noise correlation. These results provide evidence for the involvement of distinct functional neuronal network structures in encoding the stimulus-dependent and stimulus-independent information.

Our analysis of the relationship between the signal and noise correlations in layers 2/3 and 4 in mouse A1 indicated a smaller correlation between signal and noise correlations than previously reported [14]. Thus, our proposed method suggests that the signal and noise correlations reflect distinct circuit mechanisms of sound processing in layers 2/3 vs 4. The spatial distribution of signal correlations obtained by our method was consistent with previous work showing significant negative trends with distance [14]. However, in addition, our proposed method revealed a significant negative trend of noise correlations with distance in layer 2/3, but not in layer 4, in contrast to the outcome of Pearson correlation analysis. The lack of a negative trend in layer 4 could be attributed to the sparse nature of the noise correlation spread in layer 4, as revealed by our analysis of two-dimensional spatial spreads. The latter analysis indeed revealed that the noise correlations in layer 2/3 are more widespread than those in layer 4, consistent with existing work

based on whole-cell patch recordings [137, 138].

The two-dimensional spatial spreads of signal and noise correlations obtained by our method are more distinct than those obtained by Pearson correlations. The spatial spreads also allude to the directionality of the functional connectivity patterns, with a notable rostrocaudal preference in layer 4. This result seems surprising in light of existing evidence for quasi-rostrocaudal organization of the tonotopic axis in mouse A1 [139]. However, given the heterogeneity of tuning in both layers, 2/3 and 4 with a best frequency interquartile range of ~ 1 -1.5 octaves over the imaging field [129] and using supra-threshold tones, we expect that the tones will drive not only neurons with the corresponding best frequency but also neurons tuned to neighboring frequencies. Moreover, there is high connectivity between layer 4 cells within a few 100 μm across the tonotopic axis [137, 140], potentially amplifying and broadening the effect of supra-threshold tones.

Our proposed method can scale up favorably to larger populations of neurons, thanks to the underlying low-complexity variational updates in the inference procedure. Due to its minimal dependence on training data, our estimation framework is also applicable to the single-session analysis of two-photon data with a limited number of trials and duration. Another useful byproduct of the proposed framework is gaining access to approximate posterior densities in closed-form, which allows further statistical analyses such as the construction of confidence intervals. Our proposed methodology can thus be used as a robust and scalable alternative to existing approaches for extracting neuronal correlations from two-photon calcium imaging data.

A potential limitation of our proposed generative model is the assumption that there is at most one spiking event per time frame for each neuron, in light of the fact that typical two-photon imaging frame durations are in the range of 30–100 ms. Average spike rates of excitatory neurons in mouse A1 layers 2/3 and 4 are of the order of < 10 Hz [136, 141] and thus our model

is reasonable for the current study, although it might not be optimal during bursting activity. It is noteworthy that we relax this assumption in the inference framework by allowing the magnitude of putative spikes to be greater than one, thus alleviating the model mismatch during episodes of rapid increase in firing rate. This assumption can also be made more precise by adopting a Poisson model, but that would render closed-form variational density updates intractable.

Furthermore, in the regime of extremely low spiking rate and high observation noise, the proposed method may fail to capture the underlying correlations faithfully and its performance degrades to those of existing methods based on Pearson correlations, as we have shown through our simulation studies. Nevertheless, our method addresses key limitations of conventional signal and noise correlation estimators that persist even in high spiking rates and high SNR conditions.

Our proposed estimation framework can be used as groundwork for incorporating other notions of correlation such as the connected correlation function [142], and to account for non-Gaussian and higher-order structures arising from spatiotemporal interactions [143, 144]. Other possible extensions of this work include leveraging variational inference beyond the mean-field regime [145], extension to time-varying correlations that underlie rapid task-dependent dynamics [146], and extension to non-linear models such as those parameterized by neural networks [124]. In the spirit of easing reproducibility, a MATLAB implementation of our proposed method as well as the data used in this chapter are made publicly available [78, 79].

5.3 Methods and Materials

5.3.1 Proposed forward model

Suppose we observe fluorescence traces of N neurons, for a total duration of T discrete-time frames, corresponding to L independent trials of repeated stimulus. Let $\mathbf{y}_{t,l} := [y_{t,l}^{(1)}, y_{t,l}^{(2)}, \dots, y_{t,l}^{(N)}]^\top$, $\mathbf{z}_{t,l} := [z_{t,l}^{(1)}, z_{t,l}^{(2)}, \dots, z_{t,l}^{(N)}]^\top$, and $\mathbf{n}_{t,l} := [n_{t,l}^{(1)}, n_{t,l}^{(2)}, \dots, n_{t,l}^{(N)}]^\top$ be the vectors of noisy observations, intracellular calcium concentrations, and ensemble spiking activities, respectively, at trial l and frame t . We capture the dynamics of $\mathbf{y}_{t,l}$ by the following state-space model:

$$\mathbf{y}_{t,l} = \mathbf{A} \mathbf{z}_{t,l} + \mathbf{w}_{t,l}, \quad \mathbf{z}_{t,l} = \alpha \mathbf{z}_{t-1,l} + \mathbf{n}_{t,l},$$

where $\mathbf{A} \in \mathbb{R}^{N \times N}$ represents the scaling of the observations, $\mathbf{w}_{t,l}$ is zero-mean i.i.d. Gaussian noise with covariance Σ_w , and $0 \leq \alpha < 1$ is the state transition parameter capturing the calcium dynamics through a first order model. Note that this state-space is non-Gaussian due to the binary nature of the spiking activity, i.e., $n_{t,l}^{(j)} \in \{0, 1\}$. We model the spiking data as a point process or Generalized Linear Model with Bernoulli statistics [28, 33, 84, 99]:

$$n_{t,l}^{(j)} \sim \text{Bernoulli} \left(\lambda_{t,l}^{(j)} \right), \quad \lambda_{t,l}^{(j)} = \phi \left(x_{t,l}^{(j)}, \mathbf{d}_j^\top \mathbf{s}_t \right)$$

where $\lambda_{t,l}^{(j)}$ is the conditional intensity function [28], which we model as a non-linear function of the known external stimulus $\mathbf{s}_t$ and the other latent intrinsic and extrinsic trial-dependent covariates, $\mathbf{x}_{t,l} := [x_{t,l}^{(1)}, x_{t,l}^{(2)}, \dots, x_{t,l}^{(N)}]^\top$. While we assume the stimulus $\mathbf{s}_t \in \mathbb{R}^M$ to be common to all neurons, we model the distinct effect of this stimulus on the j^{th} neuron via an unknown kernel

$\mathbf{d}_j \in \mathbb{R}^M$, akin to the receptive field.

The non-linear mapping of our choice is the logistic link, which is also the canonical link for a Bernoulli process in the point process and Generalized Linear Model frameworks [28].

Thus, we assume:

$$\phi\left(x_{t,l}^{(j)}, \mathbf{d}_j^\top \mathbf{s}_t\right) = \frac{\exp\left(x_{t,l}^{(j)} + \mathbf{d}_j^\top \mathbf{s}_t\right)}{1 + \exp\left(x_{t,l}^{(j)} + \mathbf{d}_j^\top \mathbf{s}_t\right)}.$$

Finally, we assume the latent trial dependent covariates to be a Gaussian process $\mathbf{x}_{t,l} \sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$, with mean $\boldsymbol{\mu}_x := [\mu_x^{(1)}, \mu_x^{(2)}, \dots, \mu_x^{(N)}]^\top$ and covariance $\boldsymbol{\Sigma}_x$.

The probabilistic graphical model in Figure 5.8 summarizes the main components of the aforementioned forward model. According to this forward model, the underlying noise covari-

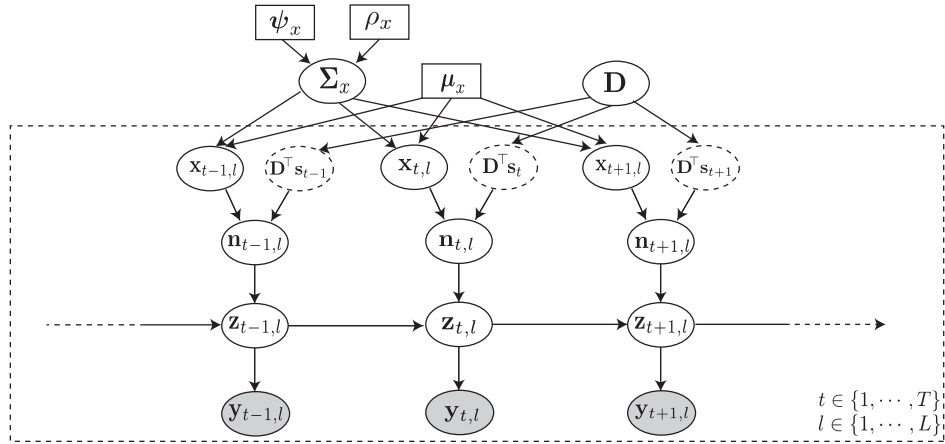


Figure 5.8: Probabilistic graphical model of the proposed forward model. The fluorescence observations at the t^{th} time frame and l^{th} trial: $\mathbf{y}_{t,l}$, are noisy surrogates of the intracellular calcium concentrations: $\mathbf{z}_{t,l}$. The calcium concentration at time t is a function of the spiking activity $\mathbf{n}_{t,l}$, and the calcium activity at the previous time point $\mathbf{z}_{t-1,l}$. The spiking activity is driven by two independent mechanisms: latent trial-dependent covariates $\mathbf{x}_{t,l}$, and contributions from the known external stimulus $\mathbf{s}_t$, which we model by $\mathbf{D}^\top \mathbf{s}_t$ (in which the receptive field $\mathbf{D}$ is unknown). Then, we model $\mathbf{x}_{t,l}$ as a Gaussian process with constant mean $\boldsymbol{\mu}_x$, and unknown covariance $\boldsymbol{\Sigma}_x$. Finally, we assume the covariance $\boldsymbol{\Sigma}_x$ to have an inverse Wishart prior distribution with hyper-parameters ψ_x and ρ_x . Based on this forward model, the inverse problem amounts to recovering the signal and noise correlations by directly estimating $\boldsymbol{\Sigma}_x$ and $\mathbf{D}$ (top layer) from the fluorescence observations $\{\mathbf{y}_{t,l}\}_{t=1,L}^{T,L}$ (bottom layer).

ance matrix that captures trial-to-trial variability can be identified as Σ_x . The signal covariance matrix, representing the covariance of the neural activity arising from the repeated application of the stimulus $\mathbf{s}_t$, is given by $\Sigma_s := \mathbf{D}^\top \text{cov}(\mathbf{s}_t, \mathbf{s}_t) \mathbf{D}$, where $\mathbf{D} := [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_N] \in \mathbb{R}^{M \times N}$. The signal and noise correlation matrices, denoted by $\mathbf{S}$ and $\mathbf{N}$, can then be obtained by standard normalization of Σ_s and Σ_x :

$$(\mathbf{S})_{i,j} := \frac{(\Sigma_s)_{i,j}}{\sqrt{(\Sigma_s)_{i,i}(\Sigma_s)_{j,j}}}, \quad (\mathbf{N})_{i,j} := \frac{(\Sigma_x)_{i,j}}{\sqrt{(\Sigma_x)_{i,i}(\Sigma_x)_{j,j}}}, \quad \forall i, j = 1, 2, \dots, N.$$

The main problem is thus to estimate $\{\Sigma_x, \mathbf{D}\}$ from the noisy and temporally blurred data $\{\mathbf{y}_{t,l}\}_{t=1, l=1}^{T,L}$.

5.3.2 Overview of the proposed estimation method

First, given a limited number of trials L from an ensemble with typically low spiking rates, we need to incorporate suitable prior assumptions to avoid overfitting. Thus, we impose a prior $p_{\text{pr}}(\Sigma_x)$ on the noise covariance, to compensate sparsity of data. A natural estimation method to estimate $\{\Sigma_x, \mathbf{D}\}$ in a Bayesian framework is to maximize the observed data likelihood $p(\{\mathbf{y}_{t,l}\}_{t,l=1}^{T,L} | \Sigma_x, \mathbf{D})$, i.e., maximum likelihood (ML). Thus, we consider the joint likelihood

of the observed data and latent processes to perform Maximum a Posteriori (MAP) estimation:

$$\begin{aligned}
& p(\mathbf{y}, \mathbf{z}, \mathbf{x}, \Sigma_x | \mathbf{D}) \\
&= p_{\text{pr}}(\Sigma_x) \prod_{t,l=1}^{T,L} \frac{1}{\sqrt{(2\pi)^N |\Sigma_w|}} \exp \left(-\frac{1}{2} (\mathbf{y}_{t,l} - \mathbf{A} \mathbf{z}_{t,l})^\top \Sigma_w^{-1} (\mathbf{y}_{t,l} - \mathbf{A} \mathbf{z}_{t,l}) \right) \\
&\times \prod_{t,l,j=1}^{T,L,N} \frac{\left(\exp \left(x_{t,l}^{(j)} + \mathbf{d}_j^\top \mathbf{s}_t \right) \right)^{z_{t,l}^{(j)} - \alpha z_{t-1,l}^{(j)}}}{1 + \exp \left(x_{t,l}^{(j)} + \mathbf{d}_j^\top \mathbf{s}_t \right)} \prod_{t,l=1}^{T,L} \frac{1}{\sqrt{(2\pi)^N |\Sigma_x|}} \exp \left(-\frac{1}{2} (\mathbf{x}_{t,l} - \boldsymbol{\mu}_x)^\top \Sigma_x^{-1} (\mathbf{x}_{t,l} - \boldsymbol{\mu}_x) \right)
\end{aligned} \tag{5.4}$$

Inspecting this MAP problem soon reveals that estimating Σ_x and $\mathbf{D}$ is a challenging task:

1) standard approaches such as Expectation-Maximization (EM) [46] are intractable due to the complexity of the model, arising from the hierarchy of latent processes and the non-linearities involved in their mappings, and 2) the temporal coupling of the likelihood in the calcium concentrations makes any potential direct solver scale poorly with T .

Thus, we propose an alternative solution based on Variational Inference (VI) [71, 72, 147].

VI is a method widely used in Bayesian statistics to approximate unwieldy posterior densities using optimization techniques, as a low-complexity alternative strategy to Markov Chain Monte Carlo sampling [148] or empirical Bayes techniques such as EM. To this end, we treat $\{\mathbf{x}_{t,l}\}_{t,l=1}^{T,L}$ and Σ_x as latent variables and $\{\mathbf{z}_{t,l}\}_{t,l=1}^{T,L}$ and $\mathbf{D}$ as unknown parameters to be estimated. We introduce a framework to update the latent variables and parameters sequentially, with straightforward update rules. We will describe the main ingredients of the proposed framework in the following subsections. Hereafter, we use the shorthand notations $\mathbf{y} := \{\mathbf{y}_{t,l}\}_{t,l=1}^{T,L}$, $\mathbf{z} := \{\mathbf{z}_{t,l}\}_{t,l=1}^{T,L}$, and $\mathbf{x} := \{\mathbf{x}_{t,l}\}_{t,l=1}^{T,L}$.

5.3.3 Preliminary assumptions

For the sake of simplicity, we assume that the constants α , $\mathbf{A}$, Σ_w and μ_x are either known or can be consistently estimated from pilot trials. For example, we assume $\mathbf{A}$ to be diagonal and estimate the diagonal elements based on the magnitudes of spiking events. Further, we estimate μ_x from the average firing rate and Σ_w using the background fluorescence in the absence of spiking events. Next, we take $p_{\text{pr}}(\Sigma_x)$ to be an Inverse Wishart density:

$$\Sigma_x \sim \text{InvWish}_N(\psi_x, \rho_x),$$

which turns out to be the conjugate prior in our model. Thus, ψ_x and ρ_x will be the hyper-parameters of our model. Procedures for hyper-parameter tuning and choosing the key model parameters are given in Section 5.3.9 and Section 5.3.7, respectively.

5.3.4 Decoupling via Pólya-Gamma augmentation

Direct application of VI to problems containing both discrete and continuous random variables results in intractable densities. Specifically, finding a variational distribution for $\mathbf{x}_{t,l}$ in our model with a standard distribution is not straightforward, due to the complicated posterior arising from co-dependent Bernoulli and Gaussian random variables. In order to overcome this difficulty, we employ Pólya-Gamma (PG) latent variables [73, 149, 150]. We observe from the joint likelihood that the posterior density, $p(\mathbf{x}|\mathbf{z}, \mathbf{D}, \Sigma_x)$ is conditionally independent in t, l with:

$$p(\mathbf{x}_{t,l}|\mathbf{z}, \mathbf{D}, \Sigma_x) \propto p(\mathbf{x}_{t,l}|\Sigma_x) \prod_{j=1}^N \frac{\left(\exp \left(x_{t,l}^{(j)} + \mathbf{d}_j^\top \mathbf{s}_t \right) \right)^{z_{t,l}^{(j)} - \alpha z_{t-1,l}^{(j)}}}{1 + \exp \left(x_{t,l}^{(j)} + \mathbf{d}_j^\top \mathbf{s}_t \right)}.$$

Thus, upon careful inspection we see that this density has the desired form for the PG augmentation scheme [73]. Accordingly, we introduce a set of auxiliary PG-distributed i.i.d. latent random variables $\boldsymbol{\omega}_{t,l} := [\omega_{t,l}^{(1)}, \omega_{t,l}^{(2)}, \dots, \omega_{t,l}^{(N)}]^\top$, $\omega_{t,l}^{(j)} \sim \text{PG}(1, 0)$ for $1 \leq j \leq N$, $1 \leq t \leq T$ and $1 \leq l \leq L$, to derive the complete data log-likelihood:

$$\begin{aligned}
& \log p(\mathbf{y}, \mathbf{z}, \mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\Sigma}_x | \mathbf{D}) \\
&= -\frac{TL}{2} \log |\boldsymbol{\Sigma}_x| + \log p_{\text{pr}}(\boldsymbol{\Sigma}_x) + \sum_{t,l=1}^{T,L} \left\{ -\frac{1}{2} (\mathbf{y}_{t,l} - \mathbf{A}\mathbf{z}_{t,l})^\top \boldsymbol{\Sigma}_w^{-1} (\mathbf{y}_{t,l} - \mathbf{A}\mathbf{z}_{t,l}) \right. \\
&\quad -\frac{1}{2} (\mathbf{x}_{t,l} - \boldsymbol{\mu}_x)^\top \boldsymbol{\Sigma}_x^{-1} (\mathbf{x}_{t,l} - \boldsymbol{\mu}_x) + \sum_{j=1}^N \left\{ \left(z_{t,l}^{(j)} - \alpha z_{t-1,l}^{(j)} - \frac{1}{2} \right) \left(x_{t,l}^{(j)} + \mathbf{d}_j^\top \mathbf{s}_t \right) \right. \\
&\quad \left. \left. - \frac{1}{2} \omega_{t,l}^{(j)} \left(x_{t,l}^{(j)} + \mathbf{d}_j^\top \mathbf{s}_t \right)^2 + \log p_{\text{PG}(1,0)} \left(\omega_{t,l}^{(j)} \right) \right\} \right\} + C, \tag{5.5}
\end{aligned}$$

where $\boldsymbol{\omega} := \{\boldsymbol{\omega}_{t,l}\}_{t,l=1}^{T,L}$ and C accounts for terms not depending on $\mathbf{y}, \mathbf{z}, \mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\Sigma}_x$ and $\mathbf{D}$. The complete data log-likelihood is notably *quadratic* in $\mathbf{z}_{t,l}$, which as we show later admits efficient estimation procedures with favorable scaling in T .

5.3.5 Deriving the optimal variational densities

In this section, we will outline the procedure of applying VI to the latent variables $\mathbf{x} = \{\mathbf{x}_{t,l}\}_{t,l=1}^{T,L}$, $\boldsymbol{\omega} = \{\boldsymbol{\omega}_{t,l}\}_{t,l=1}^{T,L}$ and $\boldsymbol{\Sigma}_x$, assuming that the parameter estimates $\hat{\mathbf{z}}$ and $\hat{\mathbf{D}}$ of the previous iteration are available. The methods that we propose to update the parameters $\hat{\mathbf{z}}$ and $\hat{\mathbf{D}}$ subsequently, will be discussed in the next section.

The objective of variational inference is to posit a family of approximate densities $\mathcal{Q}$ over the latent variables, and to find the member of that family that minimizes the Kullback-Leibler

(KL) divergence to the exact posterior:

$$q^*(\mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\Sigma}_x | \hat{\mathbf{z}}, \hat{\mathbf{D}}) = \underset{q \in \mathcal{Q}}{\operatorname{argmin}} \operatorname{KL} \left(q(\mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\Sigma}_x | \hat{\mathbf{z}}, \hat{\mathbf{D}}) \parallel p(\mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\Sigma}_x | \mathbf{y}, \hat{\mathbf{z}}, \hat{\mathbf{D}}) \right).$$

However, evaluating the KL divergence is intractable, and it has been shown [72] that an equivalent result to this minimization can be obtained by maximizing the alternative objective function, called the evidence lower bound (ELBO):

$$\operatorname{ELBO}(q) = \mathbb{E}[\log p(\mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\Sigma}_x, \mathbf{y} | \hat{\mathbf{z}}, \hat{\mathbf{D}})] - \mathbb{E}[\log q(\mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\Sigma}_x | \hat{\mathbf{z}}, \hat{\mathbf{D}})].$$

Further, we assume $\mathcal{Q}$ to be a mean-field variational family [72], resulting in the overall variational density of the form:

$$q(\mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\Sigma}_x) = q(\boldsymbol{\Sigma}_x) \prod_{t,l=1}^{T,L} \left(q(\mathbf{x}_{t,l}) \prod_{j=1}^N q(\omega_{t,l}^{(j)}) \right). \quad (5.6)$$

Under the mean field assumptions, the maximization of the ELBO can be derived using the optimization algorithm ‘Coordinate Ascent Variational Inference’ (CAVI) [72, 151]. Accordingly, we see that the optimal variational densities in Eq. (5.6) take the forms:

$$\begin{aligned} \log q^*(\mathbf{x}_{t,l}) &\propto \mathbb{E}_{q^*(\boldsymbol{\Sigma}_x)q^*(\boldsymbol{\omega}_{t,l})} \left[\log p(\mathbf{x}_{t,l} | \boldsymbol{\omega}_{t,l}, \boldsymbol{\Sigma}_x, \mathbf{y}, \hat{\mathbf{z}}, \hat{\mathbf{D}}) \right] \\ \log q^*(\omega_{t,l}^{(j)}) &\propto \mathbb{E}_{q^*(\mathbf{x}_{t,l})} \left[\log p(\omega_{t,l}^{(j)} | \mathbf{x}_{t,l}, \boldsymbol{\Sigma}_x, \mathbf{y}, \hat{\mathbf{z}}, \hat{\mathbf{D}}) \right] \\ \log q^*(\boldsymbol{\Sigma}_x) &\propto \mathbb{E}_{q^*(\mathbf{x})} \left[\log p(\boldsymbol{\Sigma}_x | \mathbf{x}, \mathbf{y}, \hat{\mathbf{z}}, \hat{\mathbf{D}}) \right] \end{aligned}$$

Upon evaluation of these expectations, we derive the optimal variational distributions as:

$$q^*(\mathbf{x}_{t,l}) \sim \mathcal{N}(\mathbf{m}_{\mathbf{x}_{t,l}}, \mathbf{Q}_{\mathbf{x}_{t,l}}), \quad q^*(\omega_{t,l}^{(j)}) \sim \text{PG}(1, c_{t,l}^{(j)}), \quad q^*(\Sigma_x) \sim \text{InvWish}_N(\mathbf{P}_x, \gamma_x).$$

whose parameters $\mathbf{m}_{\mathbf{x}_{t,l}} := [m_{\mathbf{x}_{t,l}}^{(1)}, m_{\mathbf{x}_{t,l}}^{(2)}, \dots, m_{\mathbf{x}_{t,l}}^{(N)}]^\top$, $\mathbf{Q}_{\mathbf{x}_{t,l}}$, $c_{t,l}^{(j)}$, $\mathbf{P}_x$, and γ_x can be updated given parameter estimates $\hat{\mathbf{D}}$ and $\hat{\mathbf{z}}$:

$$\begin{aligned} \mathbf{Q}_{\mathbf{x}_{t,l}} &= (\tilde{\Omega}_{t,l} + \gamma_x \mathbf{P}_x^{-1})^{-1}, \\ \mathbf{m}_{\mathbf{x}_{t,l}} &= \mathbf{Q}_{\mathbf{x}_{t,l}} \left(\hat{\mathbf{z}}_{t,l} - \alpha \hat{\mathbf{z}}_{t-1,l} - \frac{1}{2} \mathbf{1} - \tilde{\Omega}_{t,l} \hat{\mathbf{D}}^\top \mathbf{s}_t + \gamma_x \mathbf{P}_x^{-1} \boldsymbol{\mu}_x \right), \\ \mathbf{P}_x &= \psi_x + \sum_{t,l=1}^{T,L} \left\{ \mathbf{Q}_{\mathbf{x}_{t,l}} + \mathbf{m}_{\mathbf{x}_{t,l}} \mathbf{m}_{\mathbf{x}_{t,l}}^\top - \boldsymbol{\mu}_x \mathbf{m}_{\mathbf{x}_{t,l}}^\top - \mathbf{m}_{\mathbf{x}_{t,l}} \boldsymbol{\mu}_x^\top + \boldsymbol{\mu}_x \boldsymbol{\mu}_x^\top \right\}, \\ c_{t,l}^{(j)} &= \sqrt{(\mathbf{Q}_{\mathbf{x}_{t,l}})_{j,j} + \left(m_{\mathbf{x}_{t,l}}^{(j)} + \hat{\mathbf{d}}_j^\top \mathbf{s}_t \right)^2}, \end{aligned}$$

and $\gamma_x := \rho_x + TL$, with $\tilde{\Omega}_{t,l} \in \mathbb{R}^{N \times N}$ denoting a diagonal matrix with entries $(\tilde{\Omega}_{t,l})_{j,j} := \frac{1}{2c_{t,l}^{(j)}} \tanh\left(\frac{c_{t,l}^{(j)}}{2}\right)$ and $\mathbf{1} \in \mathbb{R}^N$ denoting the vector of all ones.

5.3.6 Low-complexity parameter updates

Note that even though $\mathbf{z}$ is composed of the latent processes $\mathbf{z}_{t,l}$, we do not use VI for its inference, and instead consider it as an unknown parameter. This choice is due to the temporal dependencies arising from the underlying state-space model in Eq. (5.4), which hinders a proper assignment of variational densities under the mean-field assumption. We thus seek to estimate both $\mathbf{z}$ and $\mathbf{D}$ using the updated variational density $q^*(\mathbf{x}, \omega, \Sigma_x)$.

First, note that the log-likelihood in Eq. (5.5) is decoupled in l , which admits independent updates to $\{\mathbf{z}_{t,l}\}_{t=1}^T$, for $l = 1, \dots, L$. As such, given an estimate $\hat{\mathbf{D}}$, we propose to estimate

$\{\mathbf{z}_{t,l}\}_{t=1}^T$ as:

$$\begin{aligned} \{\widehat{\mathbf{z}}_{t,l}\}_{t=1}^T &= \underset{\{z_{t,l}\}_{t=1}^T}{\operatorname{argmax}} \quad \mathbb{E}_{q^*(\mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\Sigma}_x)} \left[\log p(\mathbf{y}, \mathbf{z}, \mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\Sigma}_x | \widehat{\mathbf{D}}) \right] \\ &= \underset{\{z_{t,l}\}_{t=1}^T}{\operatorname{argmin}} \quad \sum_{t=1}^T \left\{ \frac{1}{2} (\mathbf{y}_{t,l} - \mathbf{A} \mathbf{z}_{t,l})^\top \boldsymbol{\Sigma}_w^{-1} (\mathbf{y}_{t,l} - \mathbf{A} \mathbf{z}_{t,l}) \right. \\ &\quad \left. - \sum_{j=1}^N \left(m_{\mathbf{x}_{t,l}}^{(j)} + \widehat{\mathbf{d}}_j^\top \mathbf{s}_t \right) \left(z_{t,l}^{(j)} - \alpha z_{t-1,l}^{(j)} \right) \right\}, \end{aligned} \quad (5.7)$$

under the constraints $0 \leq z_{t,l}^{(j)} - \alpha z_{t-1,l}^{(j)} \leq 1$, for $t = 1, \dots, T$ and $j = 1, \dots, N$. These constraints are a direct consequence of $n_{t,l}^{(j)} = z_{t,l}^{(j)} - \alpha z_{t-1,l}^{(j)}$ being a Bernoulli random variable with $\mathbb{E} \left[n_{t,l}^{(j)} \right] \in [0, 1]$. While this problem is a quadratic program and can be solved using standard techniques, it is not readily decoupled in t , and thus standard solvers would not scale favorably in T .

Instead, we consider an alternative solution that admits a low-complexity recursive solution by relaxing the constraints. To this end, we relax the constraint $\mathbf{z}_{t,l} - \alpha \mathbf{z}_{t-1,l} \preceq \mathbf{1}$ and replace the constraint $\mathbf{z}_{t,l} - \alpha \mathbf{z}_{t-1,l} \succeq \mathbf{0}$ by penalty terms proportional to $|z_{t,l}^{(j)} - \alpha z_{t-1,l}^{(j)}|$. The resulting relaxed problem is thus given by:

$$\min_{\{z_{t,l}\}_{t=1}^T} \quad \sum_{t=1}^T \left\{ \frac{1}{2} (\mathbf{y}_{t,l} - \mathbf{A} \mathbf{z}_{t,l})^\top \boldsymbol{\Sigma}_w^{-1} (\mathbf{y}_{t,l} - \mathbf{A} \mathbf{z}_{t,l}) + \sum_{j=1}^N \nu_{t,l}^{(j)} \left| z_{t,l}^{(j)} - \alpha z_{t-1,l}^{(j)} \right| \right\}, \quad (5.8)$$

where $\nu_{t,l}^{(j)} := \beta |m_{\mathbf{x}_{t,l}}^{(j)} + \widehat{\mathbf{d}}_j^\top \mathbf{s}_t|$ with $\beta \geq 1$ being a hyper-parameter. Given that the typical spiking rates are quite low in practice, $m_{\mathbf{x}_{t,l}}^{(j)} + \widehat{\mathbf{d}}_j^\top \mathbf{s}_t$ is expected to be a negative number. Thus, we have assumed that $-m_{\mathbf{x}_{t,l}}^{(j)} - \widehat{\mathbf{d}}_j^\top \mathbf{s}_t = |m_{\mathbf{x}_{t,l}}^{(j)} + \widehat{\mathbf{d}}_j^\top \mathbf{s}_t|$.

The problem of Eq. (5.8) pertains to *compressible* state-space estimation, for which fast recursive solvers are available [20]. The solver utilizes the Iteratively Re-weighted Least Squares

(IRLS) [75] framework to transform the absolute value in the second term of the cost function into a quadratic form in $\mathbf{z}_{t,l}$, followed by Fixed Interval Smoothing (FIS) [74] to find the minimizer. At iteration k , given a current estimate $\mathbf{z}^{[k-1]}$, the problem reduces to a Gaussian state-space estimation of the form:

$$\mathbf{y}_{t,l} = \mathbf{A}\mathbf{z}_{t,l} + \mathbf{w}_{t,l}, \quad \mathbf{z}_{t,l} = \alpha\mathbf{z}_{t-1,l} + \mathbf{v}_{t,l}, \quad (5.9)$$

with $\mathbf{w}_{t,l} \sim \mathcal{N}(0, \Sigma_w)$ and $\mathbf{v}_{t,l} \sim \mathcal{N}(0, \Sigma_{\mathbf{v}_{t,l}}^{[k]})$, where $\Sigma_{\mathbf{v}_{t,l}}^{[k]} \in \mathbb{R}^{N \times N}$ is a diagonal matrix with $(\Sigma_{\mathbf{v}_{t,l}}^{[k]})_{j,j} := \sqrt{(\hat{z}_{t,l}^{(j)[k-1]} - \alpha\hat{z}_{t-1,l}^{(j)[k-1]})^2 + \varepsilon^2} / \nu_{t,l}^{(j)}$, for some small constant $\varepsilon > 0$. This problem can be efficiently solved using FIS, and the iterations proceed for a total of K times or until a standard convergence criterion is met [20]. It is noteworthy that our proposed estimator of the calcium concentration $\mathbf{z}_{t,l}$ can be thought of as *soft* spike deconvolution, which naturally arises from our variational framework, as opposed to the *hard* spike deconvolution step used in two-stage estimators.

Finally, given $q^*(\mathbf{x}, \boldsymbol{\omega}, \Sigma_x)$ and the updated $\hat{\mathbf{z}}$, the estimate of the kernels $\mathbf{d}_j$ for $j = 1, 2, \dots, N$ can be updated in closed-form by maximizing the expected complete log-likelihood $\mathbb{E}_{q^*(\mathbf{x}, \boldsymbol{\omega}, \Sigma_x)} [\log p(\mathbf{y}, \hat{\mathbf{z}}, \mathbf{x}, \boldsymbol{\omega}, \Sigma_x | \mathbf{D})]$:

$$\hat{\mathbf{d}}_j = \left(\sum_{t,l=1}^{T,L} \left((\tilde{\boldsymbol{\Omega}}_{t,l})_{j,j} \mathbf{s}_t \mathbf{s}_t^\top \right) \right)^{-1} \left(\sum_{t,l=1}^{T,L} \left\{ \left(\hat{z}_{t,l}^{(j)} - \alpha\hat{z}_{t-1,l}^{(j)} - \frac{1}{2} \right) \mathbf{s}_t - (\tilde{\boldsymbol{\Omega}}_{t,l})_{j,j} m_{\mathbf{x}_{t,l}}^{(j)} \mathbf{s}_t \right\} \right).$$

The VI procedure iterates between updating the variational densities and parameters until con-

vergence, upon which we estimate the noise and signal covariances as:

$$\hat{\Sigma}_x := \text{mode}\{q^*(\Sigma_x)\} = \frac{\mathbf{P}_x}{\gamma_x + N + 1}, \quad \hat{\Sigma}_s := \hat{\mathbf{D}}^\top \mathbb{E}[\mathbf{s}_t \mathbf{s}_t^\top] \hat{\mathbf{D}}.$$

The overall combined iterative procedure is outlined in Algorithm 5.1. Furthermore, a MATLAB implementation of this algorithm is publicly available in [78]. It is worth noting that a special case of our proposed variational inference procedure can be used to estimate signal and noise correlations from electrophysiology recordings. Given that spiking activity, i.e. $\{\mathbf{n}_{t,l}\}_{t,l=1}^{T,L}$, is directly observed in this case, the solution to the optimization problem in Eq. (5.8) is no longer required. Thus, the parameters Σ_x and $\mathbf{D}$ can be estimated using a simplified variational procedure, which is outlined in Appendix B.3.

5.3.7 Guidelines for model parameter settings

There are several key model parameters that need to be set by the user prior to the application of our proposed method. Here, we provide our rationale and criteria for choosing these parameters, which could also serve as guidelines in facilitating the applicability and adoption of our method by future users. We will also provide the specific choices of these parameters used in our simulation studies and real data analyses.

Number of neurons selected for the analysis (N)

While our proposed method scales up well with the population size due to the low complexity update rules involved, including neurons with negligible spiking activity in the analysis would only increase the complexity and potentially contaminate the correlation estimates. Thus,

Algorithm 5.1 Estimation of Σ_x and $\mathbf{D}$ through the proposed iterative procedure

Inputs: Ensemble fluorescence measurements $\{\mathbf{y}_{t,l}\}_{t,l=1}^{T,L}$, constants $\alpha, \mathbf{A}, \Sigma_w$ and $\boldsymbol{\mu}_x$, hyper-parameters ψ_x, ρ_x, β and ϵ , tolerance at convergence δ and external stimulus $\mathbf{s}_t$

Outputs: $\hat{\Sigma}_x$ and $\hat{\mathbf{D}}$

Initialization: Initial choice of $\Sigma_{\mathbf{v}_{t,l}}, \tilde{\Omega}_t, \hat{\Sigma}_x$ and $\hat{\mathbf{D}}$, residual = 10δ , $\gamma_x = \rho_x + LT$

1: **while** residual $\geq \delta$ **do**

2: **Estimate calcium concentrations using Fixed Interval Smoothing**

Forward filter, for $t = 1, \dots, T$:

$$\begin{aligned} \mathbf{z}_{(t|t-1),l} &= \alpha \mathbf{z}_{(t-1|t-1),l} \\ \mathbf{P}_{(t|t-1),l} &= \alpha^2 \mathbf{P}_{(t-1|t-1),l} + \Sigma_{\mathbf{v}_{t,l}} \\ \mathbf{B}_{t,l} &= \mathbf{P}_{(t|t-1),l} \mathbf{A}^\top (\mathbf{A} \mathbf{P}_{(t|t-1),l} \mathbf{A}^\top + \Sigma_w)^{-1} \\ \mathbf{z}_{(t|t),l} &= \mathbf{z}_{(t|t-1),l} + \mathbf{B}_{t,l} (\mathbf{y}_{t,l} - \mathbf{A} \mathbf{z}_{(t|t-1),l}) \\ \mathbf{P}_{(t|t),l} &= (\mathbf{I} - \mathbf{B}_{t,l} \mathbf{A}) \mathbf{P}_{(t|t-1),l} \end{aligned}$$

Backward smoother, for $t = T-1, \dots, 1$:

$$\hat{\mathbf{z}}_{t,l} = \mathbf{z}_{(t|t),l} + \alpha \mathbf{P}_{(t|t),l} \mathbf{P}_{(t+1|t),l}^{-1} (\hat{\mathbf{z}}_{t+1,l} - \mathbf{z}_{(t+1|t),l})$$

3: **Update variational parameters**, for $t = 1, \dots, T$ and $l = 1, \dots, L$:

$$\begin{aligned} \mathbf{Q}_{\mathbf{x}_{t,l}} &= (\tilde{\Omega}_{t,l} + \gamma_x \mathbf{P}_x^{-1})^{-1} \\ \mathbf{m}_{\mathbf{x}_{t,l}} &= \mathbf{Q}_{\mathbf{x}_{t,l}} \left(\hat{\mathbf{z}}_{t,l} - \alpha \hat{\mathbf{z}}_{t-1,l} - \frac{1}{2} \mathbf{1} - \tilde{\Omega}_{t,l} \hat{\mathbf{D}}^\top \mathbf{s}_t + \gamma_x \mathbf{P}_x^{-1} \boldsymbol{\mu}_x \right) \\ \nu_{t,l}^{(j)} &:= \beta |m_{\mathbf{x}_{t,l}}^{(j)} + \hat{\mathbf{d}}_j^\top \mathbf{s}_t| \quad (\text{for } j = 1, \dots, N) \\ c_{t,l}^{(j)} &= \sqrt{(\mathbf{Q}_{\mathbf{x}_{t,l}})_{j,j} + \left(m_{\mathbf{x}_{t,l}}^{(j)} + \hat{\mathbf{d}}_j^\top \mathbf{s}_t \right)^2} \quad (\text{for } j = 1, \dots, N) \\ (\tilde{\Omega}_{t,l})_{j,j} &:= \frac{1}{2c_{t,l}^{(j)}} \tanh \left(\frac{c_{t,l}^{(j)}}{2} \right) \quad (\text{for } j = 1, \dots, N) \end{aligned}$$

$$\mathbf{P}_x := \psi_x + \sum_{t,l=1}^{T,L} \{ \mathbf{Q}_{\mathbf{x}_{t,l}} + \mathbf{m}_{\mathbf{x}_{t,l}} \mathbf{m}_{\mathbf{x}_{t,l}}^\top - \boldsymbol{\mu}_x \mathbf{m}_{\mathbf{x}_{t,l}}^\top - \mathbf{m}_{\mathbf{x}_{t,l}} \boldsymbol{\mu}_x^\top + \boldsymbol{\mu}_x \boldsymbol{\mu}_x^\top \}$$

4: **Update IRLS covariance approximation**, for $l = 1, \dots, L$, $t = 1, \dots, T$ and $j = 1, \dots, N$:

$$(\Sigma_{\mathbf{v}_{t,l}})_{j,j} := \sqrt{\left(\hat{z}_{t,l}^{(j)} - \alpha \hat{z}_{t-1,l}^{(j)} \right)^2 + \epsilon^2} / \nu_{t,l}^{(j)}$$

5: **Update outputs and the convergence criterion:**

$$\hat{\mathbf{d}}_j = \left(\sum_{t,l=1}^{T,L} \left((\tilde{\Omega}_{t,l})_{j,j} \mathbf{s}_t \mathbf{s}_t^\top \right) \right)^{-1} \left(\sum_{t,l=1}^{T,L} \left\{ \left(\hat{z}_{t,l}^{(j)} - \alpha \hat{z}_{t-1,l}^{(j)} - \frac{1}{2} \right) \mathbf{s}_t - (\tilde{\Omega}_{t,l})_{j,j} m_{\mathbf{x}_{t,l}}^{(j)} \mathbf{s}_t \right\} \right)$$

$$(\hat{\mathbf{D}})_{\text{prev}} = \hat{\mathbf{D}}, \quad \hat{\mathbf{D}} = [\hat{\mathbf{d}}_1, \hat{\mathbf{d}}_2, \dots, \hat{\mathbf{d}}_N]$$

$$(\hat{\Sigma}_x)_{\text{prev}} = \hat{\Sigma}_x, \quad \hat{\Sigma}_x = \frac{\mathbf{P}_x}{\gamma_x + N + 1}$$

$$\text{residual} = \|(\hat{\Sigma}_x)_{\text{prev}} - \hat{\Sigma}_x\|_2 / \|(\hat{\Sigma}_x)_{\text{prev}}\|_2 + \|(\hat{\mathbf{D}})_{\text{prev}} - \hat{\mathbf{D}}\|_2 / \|(\hat{\mathbf{D}})_{\text{prev}}\|_2$$

6: **end while**

7: Return $\hat{\Sigma}_x$ and $\hat{\mathbf{D}}$

we performed an initial pre-processing step to extract N neurons that exhibited at least one spiking event in at least half of the trials considered.

Stimulus integration window length (R)

The number of lags R considered in stimulus integration is a key parameter that can be set through data-driven approaches or using prior domain knowledge. Examples of common data-driven criteria include cross-validation, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC), which balance the estimation accuracy and model complexity [152, 153].

To quantify the effect of R on model complexity, we first describe the stimulus encoding model in our framework. Suppose that the onset of the p^{th} tone in the stimulus set ($p = 1, \dots, P$, where P is the number of distinct tones) is given by a binary sequence $f_t^{(p)} \in \{0, 1\}$. The choice of R implies that the response at time t post-stimulus depends only on the R most recent time lags. As such, the effective stimulus at time t corresponding to tone p is given by $\mathbf{s}_t^{(p)} := [f_t^{(p)}, f_{t-1}^{(p)}, \dots, f_{t-R+1}^{(p)}]^\top \in \mathbb{R}^R$. By including all the P tones, the overall effective stimulus at the t^{th} time frame is given by $\mathbf{s}_t := [\mathbf{s}_t^{(1)\top}, \dots, \mathbf{s}_t^{(P)\top}]^\top \in \mathbb{R}^{RP}$. The stimulus modulation vector $\mathbf{d}_j$ would thus be RP -dimensional. As a result, the number of parameters ($M = RP$) to be estimated linearly increases with R . By using additional domain knowledge, we chose R to be large enough to capture the stimulus effects, and at the same time to be small enough to control the complexity of the algorithm.

As an example, given that the typical tone response duration of mouse primary auditory neurons is < 1 s [134–136], with a sampling frequency of $f_s = 30$ Hz, a choice of $R \sim 30$

would suffice to capture the stimulus effects. By further examining the effect of varying R on the proposed correlation estimates in Appendix B.5.1, we chose $R = 25$ for our real data analyses.

Observation noise covariance (Σ_w) and scaling matrix ($\mathbf{A}$)

We assumed that the observation noise covariance Σ_w is diagonal, and estimated the diagonal elements using the background fluorescence in the absence of spiking events, for each neuron. We set $\mathbf{A} = a\mathbf{I}$, where $\mathbf{I} \in \mathbb{R}^{N \times N}$ represents the identity matrix, and estimated a by considering the average increase in fluorescence after the occurrence of isolated spiking events. Specifically, we derived the average fluorescence activity of multiple trials triggered to the fluorescence rise onset, and set a as the increment in the magnitude of this average fluorescence immediately following the rise onset.

State transition parameter (α)

We chose α in the range $[0.95, 0.98]$, which match the slow dynamics of the calcium indicator in our data. We tested the robustness of our estimates under different choices of α in this range through the method outlined in Section 5.3.9 and accordingly chose the optimal value of α .

Mean of the latent trial-dependent process (μ_x)

We estimated μ_x as a constant that is proportional to the average firing rate. To this end, we parametrized each component of μ_x as $\mu_x^{(j)} = -a_\mu + b_\mu \frac{1}{LT} \sum_{t,l=1}^{T,L} y_{t,l}^{(j)}$, for $j = 1, \dots, N$. The constants a_μ and b_μ were chosen such that $-2 \leq \mu_x^{(j)} \leq -10$, which gives the range of baseline

parameters compatible with observed firing rates in our experimental data.

5.3.7.1 Parameter choices for simulation study 1

In the first simulation study, we set $\alpha = 0.98$, $\beta = 8$, $\mathbf{A} = 0.1\mathbf{I}$, $\boldsymbol{\mu}_x = -4.5\mathbf{1}$ and $\boldsymbol{\Sigma}_w = 2 \times 10^{-4}\mathbf{I}$ ($\mathbf{I} \in \mathbb{R}^{8 \times 8}$ represents the identity matrix and $\mathbf{1} \in \mathbb{R}^8$ represents the vector of all ones), so that the SNR of simulated data was in the same range as that of experimentally-recorded data. We used a 6th order autoregressive process with a mean of -1 as the stimulus (s_t), and considered $R = 2$ ($M = 2$) lags of the stimulus (i.e., $\mathbf{s}_t = [s_t, s_{t-1}]^\top$) in both the generative model and inference procedure. The components of the linear and quadratic stimulus modulation vectors, i.e. $\mathbf{d}_j$, $\tilde{\mathbf{d}}_{j,1}$ and $\tilde{\mathbf{d}}_{j,2}$, were chosen at random uniformly in the range $[-0.5, 0.5]$. The variance of s_t was set in each case such that the average power of the overall signal component ($\mathbf{d}_j^\top \mathbf{s}_t$ for the linear model, and $\mathbf{d}_j^\top \mathbf{s}_t + (\tilde{\mathbf{d}}_{j,1}^\top \mathbf{s}_t)^2 + (\tilde{\mathbf{d}}_{j,2}^\top \mathbf{s}_t)^2$ for the non-linear model) was comparable to the average power of the noise component $\left(x_{t,l}^{(j)}\right)$.

5.3.7.2 Parameter choices for simulation study 2

In the second simulation study, we set $\alpha = 0.98$, $\mathbf{A} = 0.1\mathbf{I}$, $\boldsymbol{\mu}_x = -4.5\mathbf{1}$ and $\boldsymbol{\Sigma}_w = 10^{-4}\mathbf{I}$ ($\mathbf{I} \in \mathbb{R}^{30 \times 30}$ represents the identity matrix and $\mathbf{1} \in \mathbb{R}^{30}$ represents the vector of all ones) when generating the fluorescence traces $\{\mathbf{y}_{t,l}\}_{t,l=1}^{T,L}$, so that the SNR of the simulated data was in the same range as of real calcium imaging observations. Furthermore, we simulated the spike trains based on a Poisson process [33] using the discrete time re-scaling procedure [33, 132]. Following the assumptions in [132], we used an exponential link to simulate the observations:

$$n_{t,l}^{(j)} \sim \text{Poisson} \left(\lambda_{t,l}^{(j)} \right), \quad \lambda_{t,l}^{(j)} = \exp \left(x_{t,l}^{(j)} \right).$$

as opposed to the Bernoulli-logistic assumption in our recognition model. Then, we estimated the noise covariance $\hat{\Sigma}_x$ using the Algorithm 5.1, with a slight modification. Since there are no external stimuli, we set $\mathbf{s}_t = \mathbf{0}$ and $\mathbf{D} = \mathbf{0}$. Accordingly, in Algorithm 5.1, we initialized $\hat{\mathbf{D}} = \mathbf{0}$ and did not perform the update on $\hat{\mathbf{D}}$ in the subsequent iterations.

5.3.7.3 Parameter choices for real data study 1

The dataset consisted of recordings from 371 excitatory neurons, from which we selected $N = 16$ responsive neurons for the analysis. Each trial consisted of $T = 3600$ time frames (the sampling frequency was 30 Hz, and each trial had a duration of 120 seconds), with the presentation of a random sequence of four tones ($P = 4$). The spiking events were very sparse and infrequent, and hence this dataset fits our model with at most one spiking event in a time frame.

We considered $R = 25$ ($M = 100$) time lags in this analysis and further examined the effect of varying R in Appendix B.5.1. We set $\alpha = 0.95$ and $\mathbf{A} = \mathbf{I}$ ($\mathbf{I} \in \mathbb{R}^{16 \times 16}$ represents the identity matrix).

5.3.7.4 Parameter choices for real data study 2

Each trial consisted of $T = 765$ frames (25.5 seconds) at a sampling frequency of 30 Hz. The A1 neurons studied here had low response rates (in both time and space), with only ~ 10 neurons exhibiting spiking activity in at least half of the trials. Thus, we selected $N = 10$ neurons and $L = 10$ trials for the analysis, and chose $R = 25$ lags of the stimulus ($M = 25$) in the model for the stimulus-driven condition. We set $\alpha = 0.95$ and $\mathbf{A} = 0.75\mathbf{I}$ ($\mathbf{I} \in \mathbb{R}^{10 \times 10}$ represents the

identity matrix).

5.3.7.5 Parameter choices for real data study 3

Each experiment consisted of $L = 5$ trials of $P = 9$ different tone frequencies repeated at 4 different amplitude levels, resulting in each concatenated trial being ~ 180 second long (see [129] for more details). We set the number of stimulus time lags considered to be $R = 25$ ($M = 225$). For each layer, we analyzed fluorescence observations from six experiments. In each experiment, we selected the most responsive $N \sim 20$ neurons for the subsequent analysis. We set $\alpha = 0.95$ and $\mathbf{A} = \mathbf{I}$.

5.3.8 Performance evaluation

5.3.8.1 Simulation studies

Since the ground truth is known in simulations, we directly compared the performance of each signal and noise correlation estimate with the ground truth signal and noise correlations, respectively. Suppose the ground truth correlations are given by the matrix $\mathbf{X}$ and the estimated correlations are given by the matrix $\hat{\mathbf{X}}$. To quantify the similarity between $\mathbf{X}$ and $\hat{\mathbf{X}}$, we defined the following two metrics:

Normalized Mean Squared Error (NMSE): The NMSE computes the mean squared error of $\hat{\mathbf{X}}$ with respect to $\mathbf{X}$ using the Frobenius Norm:

$$\text{NMSE} := \frac{\|\mathbf{X} - \hat{\mathbf{X}}\|_F^2}{\|\mathbf{X}\|_F^2}.$$

Ratio between out-of-network power and in-network power (leakage): First, we identified the in-network and out-of-network components from the ground truth correlation matrix $\mathbf{X}$. Suppose that if the true correlation between the i^{th} neuron and the j^{th} neuron is non-zero, then $|\mathbf{X}_{i,j}| > \delta_x$, for some $\delta_x > 0$. Thus, we formed a matrix $\mathbf{X}^{\text{in}}$ that masks the in-network components, by setting $(\mathbf{X}^{\text{in}})_{i,j} = 1$ if $|\mathbf{X}_{i,j}| > \delta_x$ and $(\mathbf{X}^{\text{in}})_{i,j} = 0$ if $|\mathbf{X}_{i,j}| \leq \delta_x$. Likewise, we also formed a matrix $\mathbf{X}^{\text{out}}$ that masks the out-of-network components, by setting $(\mathbf{X}^{\text{out}})_{i,j} = 1$ if $|\mathbf{X}_{i,j}| \leq \delta_x$ and $(\mathbf{X}^{\text{out}})_{i,j} = 0$ if $|\mathbf{X}_{i,j}| > \delta_x$. Then, using these two matrices we quantified the leakage effect of $\hat{\mathbf{X}}$ comparative to $\mathbf{X}$ by:

$$\text{leakage} := \frac{\|\hat{\mathbf{X}} \cdot \mathbf{X}^{\text{out}}\|_F^2}{\|\hat{\mathbf{X}} \cdot \mathbf{X}^{\text{in}}\|_F^2},$$

where $(\cdot)$ denotes element-wise multiplication.

5.3.8.2 Real data studies

To quantify the similarity and dissimilarity between signal and noise correlation estimates, we used a statistic based on the Tanimoto similarity metric [133], denoted by $T_s(\mathbf{X}, \mathbf{Y})$ for two matrices $\mathbf{X}$ and $\mathbf{Y}$. For two vectors $\mathbf{a}$ and $\mathbf{b}$ with *non-negative* entries, the Tanimoto coefficient [133] is defined as:

$$T(\mathbf{a}, \mathbf{b}) := \frac{\mathbf{a}^\top \mathbf{b}}{\mathbf{a}^\top \mathbf{a} + \mathbf{b}^\top \mathbf{b} - \mathbf{a}^\top \mathbf{b}}.$$

The Tanimoto similarity metric between two matrices can be defined similarly, by vectorizing the matrices. Thus, we formulated a similarity metric between two correlation matrices $\mathbf{X}$ and $\mathbf{Y}$ as follows. Let $\mathbf{X}^+ := \max\{\mathbf{X}, 0\mathbf{I}\}$ and $\mathbf{X}^- := \max\{-\mathbf{X}, 0\mathbf{I}\}$, with the $\max\{\cdot, \cdot\}$ operator interpreted element-wise. Note that $\mathbf{X} = \mathbf{X}^+ - \mathbf{X}^-$, and $\mathbf{X}^+, \mathbf{X}^-$ have non-negative entries.

We then defined the similarity matrix by combining those of the positive and negative parts as follows:

$$T_s(\mathbf{X}, \mathbf{Y}) := \varepsilon T(\mathbf{X}^+, \mathbf{Y}^+) + (1 - \varepsilon) T(\mathbf{X}^-, \mathbf{Y}^-)$$

where $\varepsilon \in [0, 1]$ denotes the percentage of positive entries in $\mathbf{X}$ and $\mathbf{Y}$. As a measure of dissimilarity, we used $T_d(\mathbf{X}, \mathbf{Y}) := 1 - T_s(\mathbf{X}, \mathbf{Y})$. The values of $T_d(\hat{\mathbf{S}}, \hat{\mathbf{N}})$ in Table 5.1 and $T_s(\hat{\mathbf{N}}_{\text{spon}}, \hat{\mathbf{N}}_{\text{stim}})$ and $T_d(\hat{\mathbf{S}}_{\text{stim}}, \hat{\mathbf{N}}_{\text{stim}})$ reported in Table 5.3 were obtained based on the foregoing definitions.

To further assess the statistical significance of these results, we performed the following randomized tests. To test the significance of $T_s(\hat{\mathbf{N}}_{\text{spon}}, \hat{\mathbf{N}}_{\text{stim}})$, for each comparison and each algorithm, we fixed the first matrix (i.e. $\hat{\mathbf{N}}_{\text{spon}}$) and randomly shuffled the entries of the second one ($\hat{\mathbf{N}}_{\text{stim}}$ in both cases) while respecting symmetry. We repeated this procedure for 10000 trials, to derive the null distributions that represented the probabilities of chance occurrence of similarities between two random groups of neurons.

To test the significance of $T_d(\hat{\mathbf{S}}, \hat{\mathbf{N}})$ and $T_d(\hat{\mathbf{S}}_{\text{stim}}, \hat{\mathbf{N}}_{\text{stim}})$, for each comparison and each algorithm, again we fixed the first matrix (i.e. signal correlations). Then, we formed the elements of the second matrix (akin to noise correlations) as follows. For each element of the second matrix, we assigned either the same element as the signal correlations (to model the leakage effect) or a random noise (with the same variance as the elements in the noise correlation matrix) with equal probability. As before, we repeated this procedure for 10000 trials, to derive the null distributions that represent the probabilities of chance occurrence of dissimilarities between two matrices that have some leakage between them.

5.3.9 Hyper-parameter tuning

The hyper-parameters that directly affect the proposed estimation are the inverse Wishart prior hyper-parameters: ψ_x and ρ_x . Given that ρ_x appears in the form of $\gamma_x := TL + \rho_x$, we will consider ψ_x and γ_x as the main hyper-parameters for simplicity. Here, we propose a criterion for choosing these two hyper-parameters in a data-driven fashion, which will then be used to construct the estimates of the noise covariance matrix $\hat{\Sigma}_x$ and weight matrix $\hat{\mathbf{D}}$. Due to the hierarchy of hidden layers in our model, an empirical Bayes approach for hyper-parameter selection using a likelihood-based performance metric is not straightforward. Hence, we propose an alternative empirical method for hyper-parameter selection as follows.

For a given choice of ψ_x and γ_x , we estimate $\hat{\Sigma}_x$ and $\hat{\mathbf{D}}$ following the proposed method. Then, based on the generative model in Section 5.3.1, and using the estimated values of $\hat{\Sigma}_x$ and $\hat{\mathbf{D}}$, we sample an ensemble of simulated fluorescence traces $\hat{\mathbf{y}} = \{\hat{\mathbf{y}}_t^{(l)}\}_{t,l=1}^{T,L}$, and compute the metric $d(\psi_x, \gamma_x)$:

$$d(\psi_x, \gamma_x) := D_{\text{frob}}(\text{cov}(\hat{\mathbf{y}}, \hat{\mathbf{y}}), \text{cov}(\mathbf{y}, \mathbf{y})),$$

where $\text{cov}(\cdot)$ denotes the empirical covariance and $D_{\text{frob}}(\mathbf{X}, \mathbf{Y}) := \|\mathbf{X} - \mathbf{Y}\|_F^2$. Note that $D_{\text{frob}}(\mathbf{X}, \mathbf{Y})$ is strictly convex in $\mathbf{X}$. Thus, minimizing $D_{\text{frob}}(\mathbf{X}, \mathbf{Y})$ over $\mathbf{X}$ for a given $\mathbf{Y}$ has a unique solution. Accordingly, we observe that $d(\psi_x, \gamma_x)$ is minimized when $\text{cov}(\hat{\mathbf{y}}, \hat{\mathbf{y}})$ is nearest to $\text{cov}(\mathbf{y}, \mathbf{y})$. Therefore, the corresponding estimates $\hat{\Sigma}_x$ and $\hat{\mathbf{D}}$ that generated $\hat{\mathbf{y}}$, best match the second-order statistics of $\mathbf{y}$ that was generated by the true parameters Σ_x and $\mathbf{D}$.

The typically low spiking rate of sensory neurons observed in practice may render the estimation problem ill-posed. It is thus important to have an accurate choice of the scale matrix

ψ_x in the prior distribution. However, an exhaustive search for optimal tuning of ψ_x is not computationally feasible, given that it has $N(N+1)/2$ free variables. Thus, the main challenge here is finding the optimal choice of the scale matrix $\psi_{x,\text{opt}}$.

To address this challenge, we propose the following method. First, we fix $\psi_{x,\text{init}} = \tau \mathbf{I}$, where τ is a scalar and $\mathbf{I} \in \mathbb{R}^{N \times N}$ is the identity matrix. Next, given $\psi_{x,\text{init}}$ we find the optimal choice of γ_x as:

$$\gamma_{x,\text{init}} = \underset{\gamma_x \in \mathcal{S}_\gamma}{\operatorname{argmin}} d(\psi_{x,\text{init}}, \gamma_x),$$

where $\mathcal{S}_\gamma$ is a finite set of candidate solutions for $\gamma_x > N - 1$. Let $\hat{\Sigma}_{x,\text{init}}$ denote the noise covariance estimate corresponding to hyper-parameters $(\psi_{x,\text{init}}, \gamma_{x,\text{init}})$. We will next use $\hat{\Sigma}_{x,\text{init}}$ to find a suitable choice of ψ_x . To this end, we first fix $\gamma_{x,\text{opt}} := TL + \tilde{\rho}_x$, for some $N - 1 < \tilde{\rho}_x \ll TL$. Note that by choosing $\tilde{\rho}_x$ to be much smaller than TL , the final estimates become less sensitive to the choice of γ_x . Then, we construct a candidate set $\mathcal{S}_\psi$ for $\psi_{x,\text{opt}}$ by scaling $\hat{\Sigma}_{x,\text{init}}$ with a finite set of scalars $\eta \in \mathbb{R}^+$: $\mathcal{S}_\psi = \left\{ \eta \hat{\Sigma}_{x,\text{init}}, \eta \in \mathbb{R}^+ \right\}$. To select $\psi_{x,\text{opt}}$, we match it with the choice of $\gamma_{x,\text{opt}}$ by solving:

$$\psi_{x,\text{opt}} = \underset{\psi_x \in \mathcal{S}_\psi}{\operatorname{argmin}} d(\psi_x, \gamma_{x,\text{opt}}).$$

Finally, we use these hyper-parameters $(\psi_{x,\text{opt}}, \gamma_{x,\text{opt}})$ to obtain the estimators $\hat{\Sigma}_x$ and $\hat{\mathbf{D}}$ as the output of the algorithm.

5.3.10 Experimental procedures

All procedures were approved by the University of Maryland Institutional Animal Care and Use Committee (data from the Kanold Lab, Department of Biology, University of Maryland, College Park, MD). Imaging experiments were performed on a P60 (for real data study 1) and P83 (for real data study 2) female F1 offspring of the CBA/CaJ strain (The Jackson Laboratory; stock #000654) crossed with transgenic C57BL/6J-Tg (Thy1-GCaMP6s)GP4.3Dkim/J mice (The Jackson Laboratory; stock #024275) (CBAxThy1), and F1 (CBAxC57). The third real data study was performed on data from P66-P93 and P166-P178 mice (see [129] for more details). We used the F1 generation of the crossed mice because they have good hearing into adulthood [154].

We performed cranial window implantation and 2-photon imaging as previously described in [12, 155, 156]. Briefly, we implanted a cranial window of 3 mm in diameter over the left auditory cortex. We used a scanning microscope (Bergamo II series, B248, Thorlabs) coupled to Insight X3 laser (Spectra-physics) (study 1) or pulsed femtosecond Ti:Sapphire 2-photon laser with dispersion compensation (Vision S, Coherent) (studies 2 and 3) to image GCaMP6s fluorescence from individual neurons in awake head-fixed mice with an excitation wavelength of $\lambda = 920$ nm and $\lambda = 940$ nm, respectively. The microscope was controlled by ThorImageLS software. The size of the field of view was $370 \times 370 \mu\text{m}$. Imaging frames of 512×512 pixels (pixel size $0.72 \mu\text{m}$) were acquired at 30 Hz by bidirectional scanning of an 8 kHz resonant scanner. The imaging depth was around $200 \mu\text{m}$ below pia.

5.3.10.1 Data pre-processing

A circular ROI was manually drawn over each cell body to extract raw fluorescence traces from individual cells. Neuropil contamination subtraction and baseline correction were performed on the raw fluorescence traces of each cell [12,129,155] according to $\frac{F_{\text{cell}} - \alpha_n F_{\text{neuropil}} - \text{baseline}}{\text{baseline}}$, where α_n was set to 0.7 in real data study 1 [12], 0.8 in real data study 2 [155] and 0.9 in real data study 3 [129]. The two-photon observations $\{\mathbf{y}_{t,l}\}_{t,l=1}^{T,L}$ used in our analyses are the output of this pre-processing step.

5.3.10.2 Stimuli for real data study 1

During imaging experiments, we presented 4 tones (4, 8, 16 and 32 kHz) at 70 dB SPL. The tones were 2 s in duration with an inter-trial silence of 4 s. For the sequence of tones, we first generated a randomized sequence that consisted of 5 repeats for each tone (20 tones in total), and then the same sequence was repeated for 10 trials.

5.3.10.3 Stimuli for real data study 2

During imaging experiments, we presented a 75 dB SPL 100 ms broadband noise (4–48 kHz) as the auditory stimulus. Each trial was 5.1 s long (1 s pre-stimulus silence + 0.1 s stimulus + 3 s post-stimulus silence), and the inter-trial duration was 3 s. Spontaneous neuronal activity was collected from activity during randomly interleaved no-stimuli trials of the same duration, and these trials had complete silence throughout the trial duration (5.1 s long).

Then, we extracted 50 such trials from each type, and formed 10 ($L = 10$) trials each of 25.5 s duration ($T = 765$ frames) for the subsequent analysis, by concatenating five 5.1 s trials.

This final step was performed to increase the effective trial duration.

5.3.10.4 Stimuli for real data study 3

During imaging experiments, sounds were played at four sound levels (20, 40, 60, and 80 dB SPL). Auditory stimuli consisted of sinusoidal amplitude-modulated (SAM) tones (20 Hz modulation, cosine phase), ranging from 3–48 kHz. The frequency resolution was 2 tones/octave (0.5-octave spacing) and each of these tonal stimuli was 1 s long, repeated five times with a 4–6 s inter-stimulus interval (see [[129](#)] for details).

Chapter 6: Applications of Signal and Noise Correlation analysis

In Chapter 5, we introduced a new method to directly extract signal and noise correlations from two-photon calcium imaging data. Moreover, we also explained how our method can easily be adapted to extract the same correlations from electrophysiology data. In this chapter, we use those methods to understand how the network-level changes in the auditory cortex occur in diverse settings. Specifically, in Section 6.1, we present a comparison between correlation networks inferred by applying our methods on two-photon calcium imaging data recorded from the mouse primary auditory cortex, under two different auditory stimulus conditions (data from the Kanold Lab, Johns Hopkins University, Baltimore, MD). Then, in Section 6.2, we investigate network-level changes in the auditory cortex in the course of perceptual learning, by applying our methods to multi-unit spiking data recordings from the gerbil auditory cortex under auditory task performance (data from [80]).

6.1 Pure-tone vs two-tone network organizations

In this section, we use the signal/noise correlation estimation methods introduced in Chapter 5 to identify how the network-level properties in the mouse primary auditory cortex layer 2/3 change under two different auditory stimulus conditions: pure-tone vs two-tone presentation (i.e., presenting two pure-tones simultaneously). First, by combining several spatial filtering

methods, we uncover distinct spatial patterns in the signal and noise correlation networks under each stimulus condition, providing new insights into how the network organizations change under the two different stimulus conditions. Then, by separately analyzing subpopulations of interest, we closely inspect the nature and sources of these spatial patterns. Next, by performing the same analysis on a different set of frequencies (non-harmonics), we demonstrate that those spatial patterns are largely invariant to the choice of frequencies. Finally, by comparing the network patterns inferred from the healthy mice with an equivalent analysis performed on hearing-impaired mice (i.e., animals diagnosed with OTOF-related hearing loss), we further illustrate that the spatial patterns inferred using our methods provide meaningful interpretations and insights into the underlying network organizations.

6.1.1 Methods

We analyzed two-photon calcium imaging data recorded from the mouse primary auditory cortex layer 2/3, in eight healthy (five female and three male) and five hearing-impaired (three female and two male) animals. The experiment consisted of presenting a pseudo-random sequence of pure-tones (the four frequencies 4, 8, 16 and 32 kHz, each presented five times), and a similar sequence of two-tones (the six pairwise combinations of the four pure-tones, each presented five times), and both sequences were repeated across ten trials. Each tone presentation lasted for one second, and the tones were interleaved by a four-second duration of silence. The pure-tone and two-tone trials were presented randomly interleaved rather than in separate blocks, but the overall stimulus sequence was kept unchanged across all experiments. While the same stimulus ordering was used in the experiments of non-harmonic frequencies as well, the four presented frequencies

were changed to 4, 9.5, 22.6, and 53.8 kHz in those experiments.

6.1.1.1 Data pre-processing

In each dataset, we performed several initial data pre-processing steps. Firstly, neuropil contamination subtraction and baseline correction were performed on the raw fluorescence traces of each cell [129] according to $\frac{F_{\text{cell}} - \alpha_n F_{\text{neuropil}} - \text{baseline}}{\text{baseline}}$, where α_n was set to 0.7. Then, we used the following method to identify the neurons that had statistically significant responses under pure-tones and two-tones. Recall that each tone (4 pure-tones and 6 two-tones) was repeated 50 times (5 times within a trial $\times$ 10 trials), in each experiment. For each neuron and each tone, we tested whether the responses were statistically significant by performing a t-test (with $\alpha = 0.01$) comparing the average fluorescence activity one second after the tone onset with the average fluorescence activity one second before the tone onset, over the 50 repeats. For the subsequent analysis, we considered only the neurons that responded with statistical significance to at least one of the four pure-tones *and* one of the six two-tones. When finding the responsive neurons in the hearing-impaired datasets, we relaxed the threshold of the t-test to $\alpha = 0.05$, since relatively fewer neurons were responsive.

6.1.1.2 Signal and Noise correlation inference

We used the iterative procedure introduced in Chapter 5 to derive signal and noise correlations directly from fluorescence observations. We first separated the 10 trials corresponding to pure-tone vs two-tone presentation and then derived signal and noise correlation matrices under each condition following the methods in Chapter 5.

6.1.1.3 Identifying significant spatial patterns

To identify spatial patterns in correlations, we first derived the two-dimensional (2D) relative spatial spreads of correlations, by centering each neuron at the origin and overlaying the individual spatial spreads. Then, to investigate the existence of significant spatial patterns in these 2D spatial spreads, we adopted a procedure similar to [157]. In brief, we derived the 2D FFT of the 2D spatial maps and compared it with a null distribution ($\alpha = 0.05$) to extract the statistically significant FFT components. To form the null distribution, we randomly shuffled the correlations in each dataset (while respecting the symmetry), derived the 2D spatial spreads of the shuffled correlations, obtained the 2D FFT from those shuffled spatial spreads, and considered the magnitude of the 2D FFT. We formed the null distribution corresponding to the chance occurrence of spatial patterns by performing 500 repeats of this procedure. Then, we obtained the 95% percentile points from the null distribution and extracted the 2D FFT components of the original un-shuffled spatial maps with a magnitude exceeding those thresholds. Next, we performed a smoothing in the FFT-domain using a 2D low-pass Gaussian filter and then performed an inverse 2D FFT to recover the filtered significant spatial patterns. Finally, to emphasize the significant spatial patterns, we visualized these filtered spatial patterns in the form of 2D contour plots.

6.1.1.4 Identifying best frequency and suppressive/facilitative interactions

To identify the best frequency of each neuron, we used the tuning curves inferred as an additional outcome of the correlation estimation method in Chapter 5. For each neuron, we compared the tuning curve levels inferred for each of the four tones under pure-tone presentation

and determined the mode of the tuning curve to be the best frequency (BF) of that neuron.

Then, to identify suppressive and facilitative interactions, we performed the following analysis. For each neuron considered, we compared the tuning curve level inferred for its best frequency tone under pure-tone presentation, with those inferred for its best frequency tone along with a second tone under two-tone presentation. If the two-tone tuning curve level is significantly larger/smaller than the corresponding pure-tone tuning curve level (according to a Wilcoxon signed-rank test with $\alpha = 0.05$), we considered the frequency of the second tone under the two-tone presentation to be facilitative/suppressive to the best frequency. Following this convention, we termed all pairwise interactions between the target neuron and all other neurons whose best frequency is equal to that of the second tone, as facilitative/suppressive interactions.

6.1.2 Results

We analyzed two-photon calcium imaging data (11 datasets in total) recorded from layer 2/3 of the primary auditory cortex of healthy mice under two different stimulus conditions, specifically pure-tones vs two-tones. Figure 6.1-A shows the pure-tone and two-tone sequences that were repeated in each dataset across 10 trials. The pure-tone sequence was a pseudo-random sequence of four different frequencies (4, 8, 16 and 32 kHz, each presented five times) and the two-tone sequence consisted of the six pairwise combinations of the same four frequencies (each presented five times also in a pseudo-random order). Each tone presentation lasted for one second, and the tones were interleaved by a four-second duration of silence (Figure 6.1-A).

Firstly, from each dataset, we extracted the neurons that significantly responded to at least one of the four pure-tones *and* one of the six two-tones, by considering the stimulus-locked

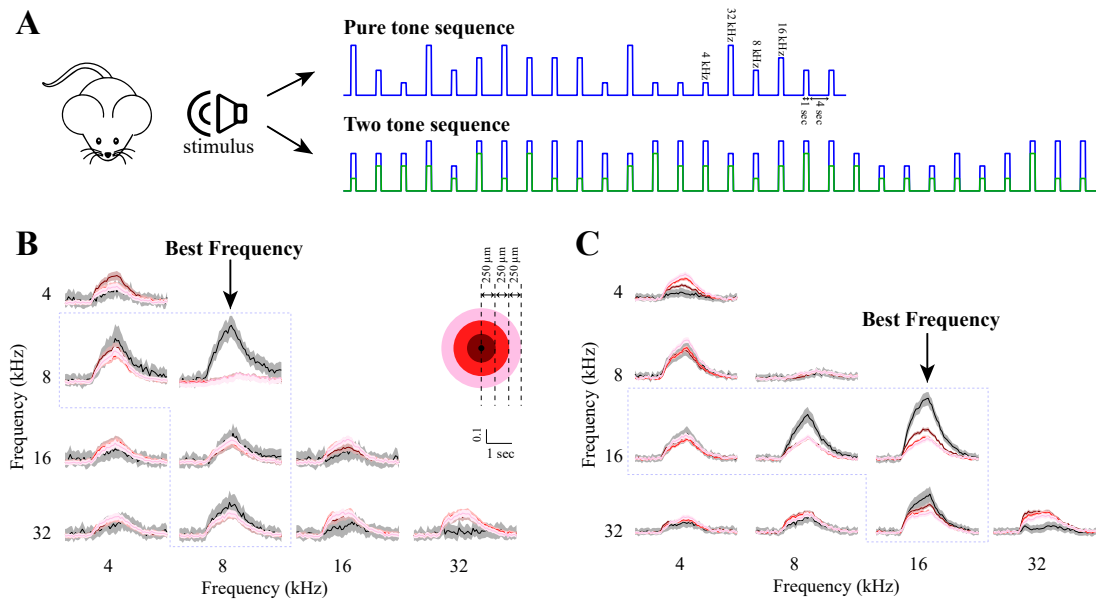


Figure 6.1: The design of the experiment and preliminary observations. (A) Mice were presented with pseudo-random sequences of pure-tones and two-tones while recording activity in the primary auditory cortex layer 2/3 via two-photon calcium imaging. (B) The average fluorescence activity of neurons with the best frequency 8 kHz (black) for each of the four pure-tone (diagonal) and six two-tone (off-diagonal) presentations. The maroon, red and pink traces respectively show the average fluorescence activity of neurons that are within a distance of 0 – 250 μm , 250 – 500 μm , and 500 – 750 μm away from neurons with the best frequency 8 kHz. The error bars indicate the 95% confidence intervals of the average activity, derived through bootstrapping. (C) Same as panel B for neurons with the best frequency of 16 kHz.

fluorescence activity. We restricted our subsequent analysis to this set of responsive neurons. To understand how the fluorescence responses change under the pure-tone and two-tone stimuli, we performed the following preliminary analysis. We classified the neurons based on their best frequency, and derived the average fluorescence activity of neurons with the same best frequency, under the ten different tones (four pure-tones and six two-tones). Figure 6.1-B and C (black curves) show the comparisons thus obtained, for neurons with the best frequency 8 kHz and 16 kHz respectively. Then, in each case, we also extracted the average fluorescence activity of neurons that are within a distance of 0 – 250 μm , 250 – 500 μm and 500 – 750 μm away from the target neuron, as illustrated respectively by the maroon, red and pink traces in Figure 6.1-B

and C.

Careful inspection of the activity traces in Figure 6.1-B and C reveals several interesting phenomena. When closely observing the pure-tone and two-tone responses to frequency 8 kHz in Figure 6.1-B, we see that the response to 8 kHz of the neighboring neurons (shown by the maroon, red and pink traces) is significantly higher under two-tone conditions ((8, 4), (8, 16), (8, 32) kHz) than the pure tone condition (i.e., the gap between the black and colored traces is considerably smaller under the two-tone cases than the corresponding pure-tone case). This suggests that neurons that did not originally significantly respond to 8 kHz, have a clear increase in response when the same frequency (8 kHz) is presented with another tone. We noted the same response pattern from a similar analysis on the other frequencies as well (Figure 6.1-C shows the activity under 16 kHz). In addition, Figure 6.1-C reveals another interesting aspect. Note that the average response of neurons tunes to 16 kHz to both pure-tone frequencies 4 kHz and 8 kHz are negligible, as indicated by the corresponding black traces in Figure 6.1-C. On the other hand, the average response of the same neurons to the pairwise combination of 4 kHz and 8 kHz is remarkably higher (the black trace of the two-tone combination (4, 8) kHz is as large as the corresponding colored traces). This validates that neurons that did not originally respond to 4 kHz or 8 kHz, do respond when the same two tones are presented together. These preliminary results clearly illustrate that the responses in the mouse primary auditory cortex under the pure-tone stimuli and two-tone stimuli have notable differences, in particular, neurons further apart from each other seemingly have higher stimulus-locked connectivity under two-tone stimuli than pure-tone ones. In the subsequent sections, we explore how these response changes manifest at the network level.

6.1.2.1 Signal/Noise Correlations vs cell pair distance

To investigate network-level changes under pure-tone and two-tone stimuli, we first derived the signal and noise correlation networks following the methods in Chapter 5. Then, to evaluate the spatial distribution of correlations under the two different stimulus conditions, we plotted the correlation values for pairs of neurons as a function of their cell-pair distance, separately for signal (Figure 6.2-A) and noise (Figure 6.2-B) correlations. The distances were discretized using bins of length $10 \mu m$. The scatter of the correlations along with their median (shown in black) at each bin are shown in all panels. Next, to examine the spatial trend of the correlations, we performed linear regression (as shown in red) in each case. The slope of the model fit, and the p-value for the corresponding t-test, are also reported as insets in all panels.

Consistent with the results reported in Chapter 5, we see that signal correlations have a statistically significant decreasing trend with distance under pure-tone presentation (Figure 6.2-A left panel). Interestingly, although the same trend exists in the two-tone signal correlations

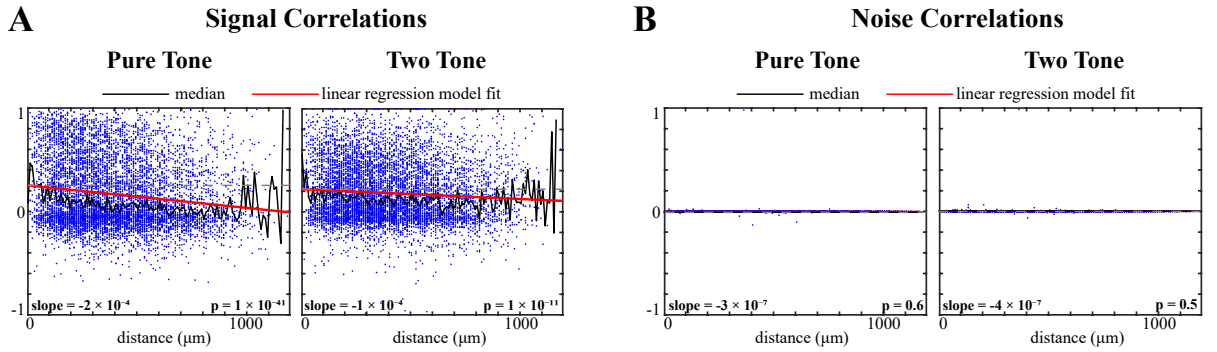


Figure 6.2: Variation of correlations vs cell-pair distance. (A) Variation of signal correlations vs. cell pair distance, under pure-tone stimuli (left) and two-tone stimuli (right). Distances were binned to $10 \mu m$ intervals. The median of the distributions (black) and the linear model fit (red) are shown in each panel. The slope of the linear model fit, and the p-value of the t-test to identify the statistical significance of the slope are also indicated as insets. Dashed horizontal lines indicate the corresponding zero-slope lines for the ease of visual comparison. (B) The same configuration as panel A, for noise correlations.

(Figure 6.2-A right panel) as well, it is comparatively much smaller (the slope under two-tone stimuli is half of that under pure-tone stimuli in Figure 6.2-A). This indeed corroborates our preliminary observations in Figure 6.1; i.e., stimulus-induced correlations under two-tone stimuli sustain over longer distances than under pure-tone stimuli, suggesting that two-tone signal correlation networks are more anatomically widespread than the pure-tone ones.

Further, from Figure 6.2-B we see that noise correlations under both pure-tone and two-tone stimuli are very sparse compared to the signal correlations. This is also expected since our analysis is focused specifically on neurons that significantly responded to stimuli. Given that these neurons are well-tuned to encode the stimuli, it is highly likely that they have comparatively weaker spontaneous activity, thus resulting in the noise correlations being insignificant. As a result of this sparseness, the distributions of noise correlations vs cell-pair distance (Figure 6.2-B) do not seem to exhibit statistically significant trends under either stimuli.

6.1.2.2 Spatial patterns in 2D spatial spreads of correlations

Next, to get an understanding of the relative configuration of correlations in the rostrocaudal-dorsoventral axes system, we derived the 2D relative spatial spreads of correlations by centering each neuron at the origin and overlaying the individual spatial spreads. Figure 6.3-A and -B respectively shows the spatial spreads of signal and noise correlations thus obtained. The horizontal and vertical axes in each panel represent the relative rostrocaudal and dorsoventral distances respectively, and the heat maps represent the average magnitude of correlations. In accordance with the results illustrated in Figure 6.2, we see from these spatial spreads as well that the noise correlations (Figure 6.3-B) under both pure-tone and two-tone conditions are notably weaker than

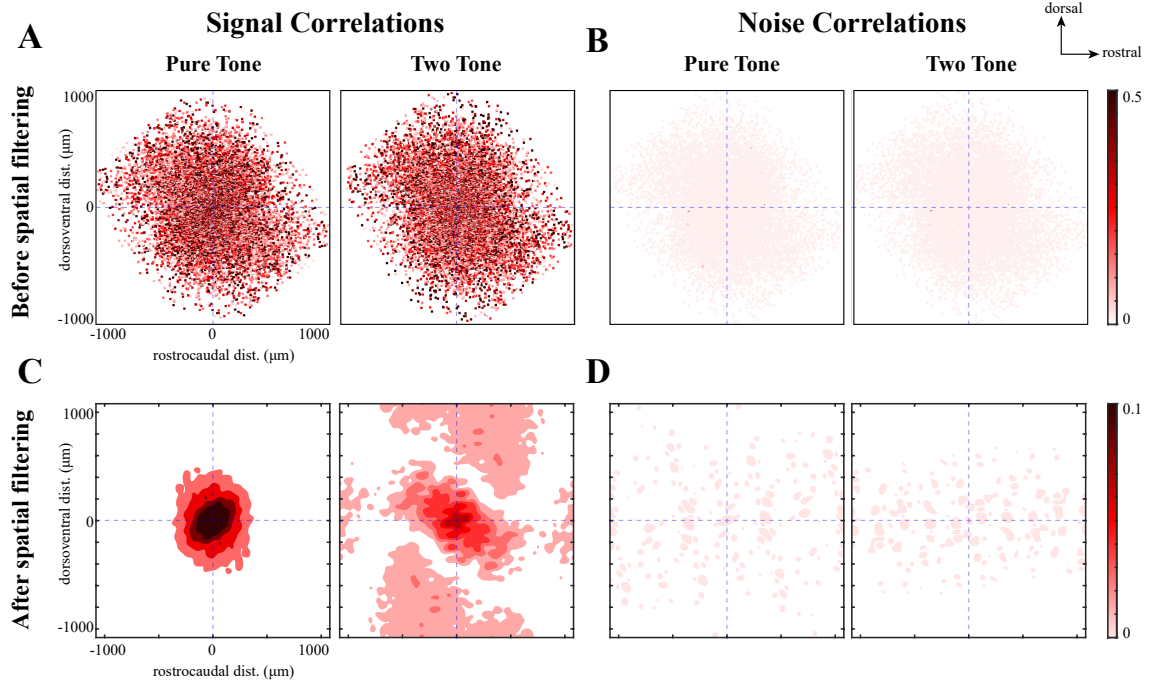


Figure 6.3: Investigating spatial patterns in 2D spatial spreads of correlations. (A) 2D spatial spread of signal correlations inferred under pure-tone stimuli (left) and two-tone stimuli (right), before performing filtering in the 2D FFT domain. The horizontal and vertical axes in each panel respectively represent the relative rostrocaudal and dorsoventral distances between each cell pair, and the heat maps indicate the average magnitude of correlations. (B) Same organization as panel A, for noise correlations. (C) Same organization as panel A, for contour plots of statistically significant 2D spatial patterns of signal correlations after performing filtering in the 2D FFT domain. (D) Same organization as panel C, for noise correlations.

the corresponding signal correlations (Figure 6.3-A). Furthermore, it is not possible to immediately identify the differences/similarities between the spatial patterns of signal correlations under pure-tone and two-tone conditions from an initial inspection of Figure 6.3-A, since both of these 2D spatial distributions are extremely dense and noisy.

Thus, we performed the following analysis on each 2D spatial spread, to denoise and extract the underlying spatial patterns. For each case, we derived the null distribution of spatial spreads corresponding to the chance occurrence of spatial patterns by shuffling the correlation estimates, for 500 repeats. Then, we derived the 2D FFT of the un-shuffled and shuffled spa-

tial spreads, and extracted the 2D FFT components of the original un-shuffled spatial spreads with magnitude greater than the 95% percentile point indicated by the null distribution (in other words, we extracted the components in the original un-shuffled spatial spread that rejected the null hypothesis with $\alpha = 0.05$). Next, we performed a 2D low-pass Gaussian filtering in the 2D FFT domain on the statistically significant 2D FFT components, to smoothen the spatial patterns. Finally, we performed an inverse 2D FFT to recover the spatial patterns in the spatial domain, and visualized the filtered maps in the form of contour plots as shown in Figure 6.3-C and -D.

The filtered spatial patterns of signal correlations in Figure 6.3-C reveal several key characteristics. The significant spatial patterns of pure-tone signal correlations (Figure 6.3-C left panel) have a uni-modal centered structure, whereas the ones derived from two-tone signal correlations (Figure 6.3-C right panel) are multi-modal (specifically, there is a strong centered component and several detached “islands” around it), and more widespread. To further investigate the differences between the pure-tone and two-tone filtered signal correlation spatial maps in Figure 6.3-C, we also derived the marginal radial and angular distributions of the spatial maps as shown in Figure 6.4-A and -B respectively. The marginal radial distributions in Figure 6.4-A clearly show that the pure tone activity (left panel) decays very fast, while the two-tone activity (right panel) sustains considerably over longer distances. The marginal angular distributions in Figure 6.4-B also reveal several interesting features. While both pure-tone and two-tone angular distributions are significantly different from a uniform distribution (as indicated by the p-values in Figure 6.4-B), note that comparing p-values is not useful, as the alternative distributions are unknown. The KS-statistic (effect size) of the two-tone case being larger compared to the pure-tone case validates that the two-tone distribution deviates from a uniform distribution far more than the pure-tone distribution, which was also visually evident in Figure 6.3-C. In addition, from both Figure 6.3-C

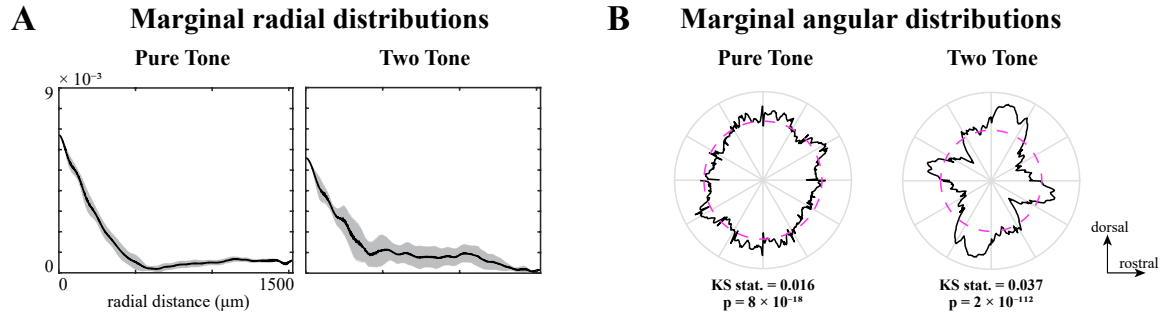


Figure 6.4: Marginal radial and angular distributions of filtered spatial patterns of signal correlations in Figure 6.3-C. (A) Marginal radial distributions of filtered spatial patterns of signal correlations under pure-tone stimuli (left) and two-tone stimuli (right). The error bars indicate the standard deviation. (B) Marginal angular distributions of filtered spatial patterns of signal correlations under pure-tone stimuli (left) and two-tone stimuli (right). The KS test statistic comparing each polar distribution with a uniform distribution (shown in magenta), along with the corresponding p-values are indicated below each polar plot.

and Figure 6.4-B, we see that the two-tone spatial patterns tend to have a dorso-ventral direction preference, while the pure-tone spatial patterns are less directionally selective.

On the other hand, similar to the un-filtered noise correlation spatial spreads in Figure 6.3-B, the spatial patterns of noise correlation maps after filtering shown in Figure 6.3-D are also very weak and insignificant in comparison to their signal counterparts. It is also noteworthy that while they are very weak, the noise correlation spatial patterns under pure-tone (Figure 6.3-D left panel) and two-tone (Figure 6.3-D right panel) presentation have the same salient features. This is indeed what one would expect; since the pure-tone and two-tone trials were recorded randomly interleaved rather than in separate blocks, the spontaneous activity (and thus the noise correlations) under the two settings tend to be quite similar to each other. Considering the sparseness, very low strength, and lack of structure in noise correlation spatial patterns, we restrict the subsequent analysis of further investigating the source of spatial patterns to signal correlations.

6.1.2.3 Network spatial patterns of neurons with same/different best frequency

In this section, we investigate the contributions of pairwise interactions between neuronal pairs having the same best frequency vs neuronal pairs with different best frequencies, to the spatial patterns of signal correlations observed in Figure 6.3-C. We categorized all pairwise interactions into two groups based on their best frequencies: pairwise interactions between neuronal pairs with the same best frequency, and pairwise interactions between neuronal pairs having different best frequencies. Figure 6.5 shows the contour maps corresponding to each of these two cases, for pure-tone and two-tone conditions. The pie charts in Figure 6.5 indicating the proportion of pairwise interactions that fall under each case show that only 34% of the pairwise interactions had the same best frequency, whereas the other 66% had been contributed by neuronal pairs with different best frequencies.

The spatial patterns in Figure 6.5 reveal several interesting phenomena. Firstly, when looking only at the pairwise interactions with the same best frequency (Figure 6.5-A), we see that both pure-tone and two-tone signal correlation spatial patterns are very similar to each other, with a uni-modal centered structure (similar to the initial pure-tone signal correlation spatial patterns in Figure 6.3-C). The fact that the off-centered islands that were observable in the two-tone spatial patterns in Figure 6.3-C no longer exist in the two-tone spatial patterns in Figure 6.5-A, clearly indicates that that widespread two-tone connectivity is not an outcome of the interaction between neurons with the same best frequency. In other words, this affirms that the widespread connectivity in the two-tone spatial spreads in Figure 6.3-C had been primarily contributed by pairwise interactions between neuronal pairs with different best frequencies. On the other hand, the main lobe at the center of the two-tone spatial patterns in Figure 6.3-C is likely to be majorly

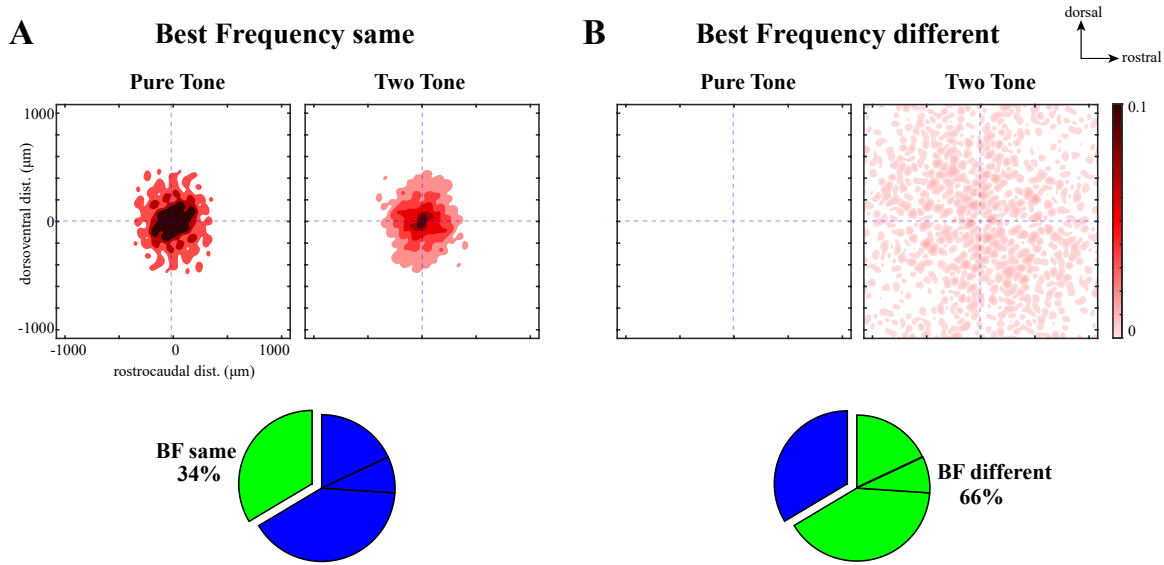


Figure 6.5: Contributions of neuronal pairs with the same best frequency (BF) vs different best frequencies to the spatial patterns of signal correlations in Figure 6.3-C. (A) Contour maps of statistically significant 2D spatial patterns of signal correlations after performing filtering in the 2D FFT domain, when considering only the pairwise interactions between neuronal pairs with the same best frequency, under pure-tone (left) and two-tone conditions (right). The pie chart at the bottom indicates (in green) the proportion of interactions with the same best frequency. (B) Same organization as panel A, for neuronal pairs having different best frequencies.

contributed by the pairwise interactions between neuronal pairs with the same best frequency, as evident from the uni-modal centered spatial distribution in Figure 6.5-A right panel.

Moreover, from the left panel of Figure 6.5-B we see that there is no significant contribution to the spatial patterns of the pure-tone activity from interactions between neuronal pairs having different best frequencies. In comparison, we do see widespread connectivity in the corresponding two-tone case (Figure 6.5-B right panel), further corroborating our hypothesis regarding the source of the broader connectivity noted in the two-tone maps. It is also noteworthy that while the augmentation of the two populations considered in Figure 6.5-A and B will indeed bring together the entire set of pairwise connections between all neurons, the combination of the two filtered maps in Figure 6.5-A and B will not reproduce the exact spatial patterns in Figure 6.3-C, essentially because the spatial filtering method we perform is not a linear operation.

6.1.2.4 Contributions of suppressive/facilitative interactions to spatial patterns

From the analysis in the previous section, we concluded that the widespread connectivity in the two-tone signal correlation spatial patterns is mostly contributed by pairwise interactions between neuronal pairs with different best frequencies. Here, we further probe into the neuronal pairs with different best frequencies, specifically to see whether those pairwise interactions are primary suppressive or facilitative. To classify interactions as suppressive or facilitative, for each neuron, we compared the tuning curve level inferred for its best frequency under pure-tone conditions, with those inferred for its best frequency tone along with a second tone under two-tone conditions. If the two-tone tuning curve level is significantly larger/smaller than the corresponding pure-tone tuning curve level, we considered the frequency of the second tone under the two-tone presentation to be facilitative/suppressive. Following this convention, we termed all pairwise interactions between the target neuron and other neurons having the best frequency equal to that second tone as facilitative/suppressive interactions. Figure 6.6-A and -B respectively illustrated the two-tone signal correlation spatial patterns inferred based on suppressive and facilitative interactions. Specifically, in each case, we have shown the spatial patterns under two settings: when including only the considered type of pairwise interactions (left panels), and when including all other interactions except the type of interest (right panels).

From the pie charts at the bottom of Figure 6.6, we see that while 40% of total pairwise interactions are contributed by suppressive interactions, facilitative interactions in total amount to only 8%. Thus, the majority of the best frequency different interactions are suppressive. The spatial patterns of suppressive interactions in Figure 6.6-A reveal how important these suppressive interactions are to the spatial patterns of two-tone signal correlations in Figure 6.3-C. Note that

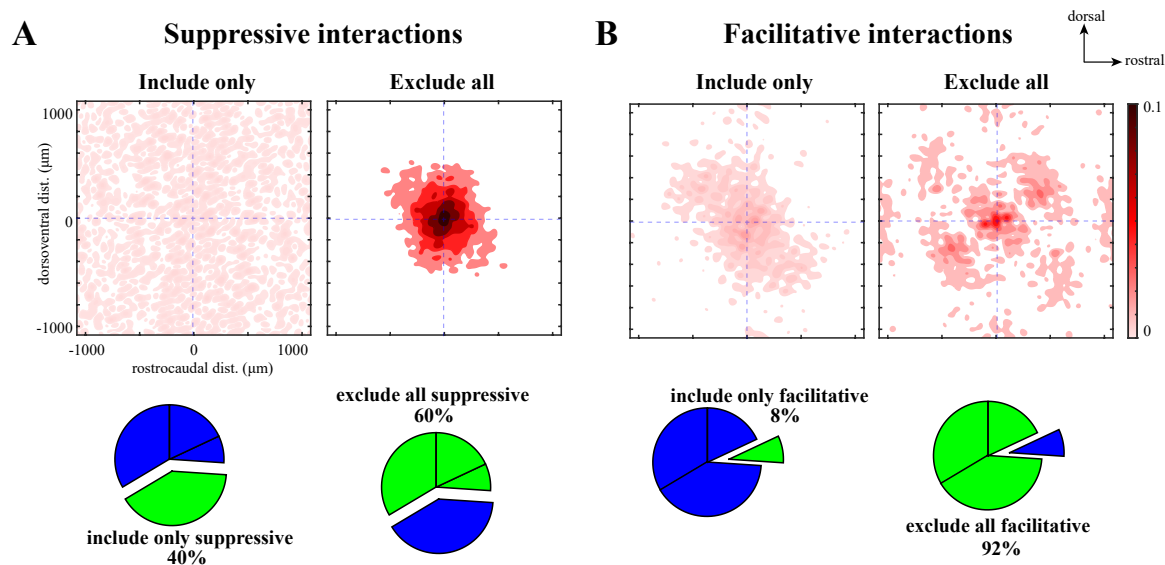


Figure 6.6: Contributions of suppressive vs facilitative interactions to the spatial patterns of two-tone signal correlations. (A) Contour maps of statistically significant 2D spatial patterns of two-tone signal correlations after performing filtering in the 2D FFT domain, when including only the suppressive pairwise interactions (left) and when excluding all suppressive pairwise interactions (right). The pie chart at the bottom indicates (in green) the proportion of interactions considered in each case. (B) Same organization as panel A, for facilitative pairwise interactions.

excluding all suppressive interactions (Figure 6.6-A right panel) eliminates the off-centered islands in the Figure 6.3-C (right panel), while the spatial patterns of considering only suppressive interactions (Figure 6.6-A left panel) has a widespread structure that is very similar to the spatial patterns of all best frequency different interactions (Figure 6.5-B right panel). This indeed implies that the widespread two-tone signal correlation connectivity is predominantly contributed by suppressive interactions. In contrast, the spatial distribution of only facilitative interactions (Figure 6.6-B left panel) is more unimodal and centered, suggesting that facilitative interactions primarily occur over shorter distances. Consequently, the off-centered isolated islands in the two-tone signal correlation patterns still notably persist even when observing the spatial patterns of all other interactions except the facilitative ones (Figure 6.6-B right panel).

6.1.2.5 Network spatial patterns in data recorded under non-harmonic tones

Next, to investigate whether the spatial patterns identified in the previous sections are unique to the choice of stimulus (in particular, note that since the four frequencies we considered 4, 8, 16, 32 kHz are one octave apart from each other, they are all harmonics), we performed the same set of analyses on five other datasets recorded under a slightly different stimulus. Specifically, we changed the four frequencies used in the stimuli to 4, 9.5, 22.6, 53.8 kHz for both pure-tone and two-tone experiments, while using the same pseudo-random sequences as before (Figure 6.1-A). Note that while the four frequencies used in our initial analyses were harmonics of each other, none of the frequencies in this new setting have integer octave relative frequency spacing. Thus, this eliminates any biases induced in our previous results due to tones presented being harmonics of each other. Figure 6.7 shows the spatial patterns inferred under this new stimulus setting.

The correlation vs cell pair analysis of these new datasets (Figure 6.7-A and -B) interestingly reveals that all trends in correlations (signal/noise under pure-tone/two-tone) with cell-pair distance under non-harmonic tones are identical to the corresponding results under harmonic tones (Figure 6.2). Specifically, we note that even under the non-harmonic tones, both pure-tone and two-tone signal correlations have a statistically significant trend to decrease with distance, but the slope under the pure-tone case (Figure 6.7-A left panel) is around twice as large as the slope under the two-tone case (Figure 6.7-A right panel). These results indicate that even under non-harmonic frequencies, two-tone signal correlations are more widespread than the corresponding pure-tone ones. Moreover, similar to the noise correlations under the harmonic setting (Figure 6.2-B), noise correlations under both pure-tone (Figure 6.7-B left panel) and two-tone (Figure

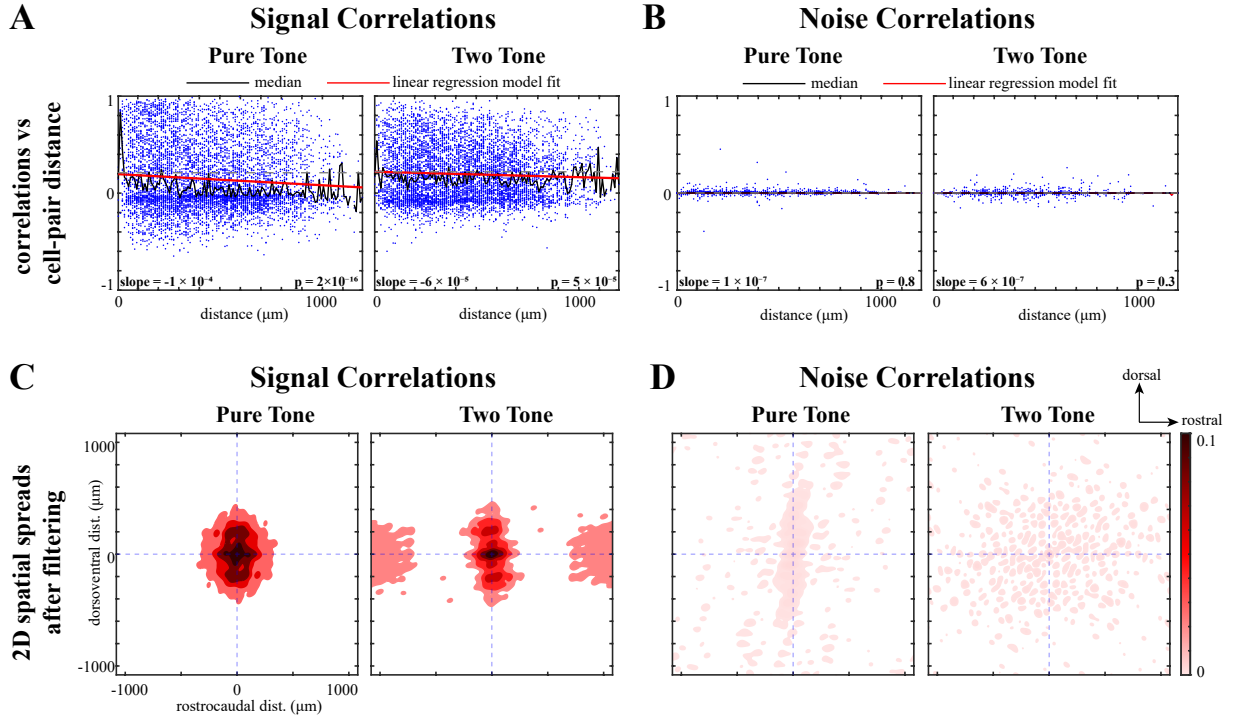


Figure 6.7: Spatial patterns of correlation networks inferred from datasets recorded under non-harmonic tones (specifically, the four frequencies of the tones are 4, 9.5, 22.6, 53.8 kHz). (A) Variation of signal correlations vs. cell pair distance, under pure-tone stimuli (left) and two-tone stimuli (right). Distances were binned to 10 μm intervals. The median of the distributions (black) and the linear model fit (red) are shown in each panel. The slope of the linear model fit, and the p-value of the t-test to identify the statistical significance of the slope are also indicated as insets. Dashed horizontal lines indicate the corresponding zero-slope lines for ease of visual comparison. (B) The same configuration as panel A, for noise correlations. (C) Contour plots of statistically significant 2D spatial patterns of signal correlations inferred under pure-tone stimuli (left) and two-tone stimuli (right), after performing filtering in the 2D FFT domain. The horizontal and vertical axes in each panel respectively represent the relative rostrocaudal and dorsoventral distances between each cell pair. (D) Same organization as panel C, for noise correlations.

6.7-B right panel) non-harmonic conditions are also very sparse, with no significant spatial patterns.

To further investigate the nature of the spatial patterns under non-harmonic frequencies, we also derived the statistically significant spatial patterns as illustrated in Figure 6.7-C and -D. Note that the spatial patterns in Figure 6.7-C and -D are also very similar to their harmonic counterparts (Figure 6.3-C and -D). In brief, pure-tone signal correlations are uni-modal and centered,

two-tone signal correlations are multi-modal (have a strong central component along with several detached islands) and widespread, and noise correlations under both pure-tone and two-tone cases are very weak with no identifiable structure. On the other hand, the orientation of the widespread connectivity under two-tone non-harmonic signal correlations (Figure 6.7-C right panel) is notably different from its harmonic equivalent (Figure 6.3-C right panel). While the harmonic two-tone signal correlation widespread patterns were primarily oriented in the dorsoventral direction (Figure 6.3-C right panel), the corresponding non-harmonic spatial patterns (Figure 6.7-C right panel) are displaying a significant rostrocaudal directional preference.

In summary, the analysis presented in this section signifies two key factors. Firstly, the overall spatial patterns of pure-tone/two-tone signal/noise correlations that we illustrated for the dataset with harmonic frequencies in previous sections preserve uniformly even with the dataset under non-harmonic frequencies, validating that the spatial trends we outlined in previous sections are not singular to our initial choice of frequencies. Secondly, we see that while indeed two-tone signal correlation spatial patterns are multi-modal and widespread under both harmonic and non-harmonic cases, the preferred direction of the widespread connectivity seemingly changes under the two settings.

6.1.2.6 Control experiment: network spatial pattern analysis in data recorded from animals diagnosed with OTOF-related hearing loss

Finally, we performed the same signal/noise correlation spatial pattern analysis on a control experiment, data collected from hearing-impaired mice (diagnosed with OTOF-related hearing loss) under identical experimental settings (with tone frequencies 4, 8, 16 and 32 kHz similar to

the initial setting). OTOF gene is key for the healthy maturation of synapses of the peripheral auditory system, and its mutations encoding for the protein otoferlin are known to cause significant hearing impairment [158]. However, the entire auditory system is not completely dysfunctional in the absence of otoferlin, and a weaker form of hearing capability is preserved in a majority of the individuals with a lack of this gene [158]. Here, we investigate how the underlying network organizations in the mouse primary auditory cortex change under this hearing impairment condition.

We analyzed data from five datasets and used a less stringent threshold when selecting the responsive neurons compared to the healthy datasets, to reduce any potential biases incurred due to a smaller population being inherently strongly responsive to auditory stimuli in the hearing-impaired datasets. Figure 6.8-A and -B shows the analysis of correlations vs cell pair distance and Figure 6.8-C and -D show the significant spatial patterns inferred from the 2D relative spatial spreads of correlations.

The correlations vs cell-pair distance distributions in Figure 6.8-A and -B expose several striking features. Note that the statistically significant trend of signal correlations decreasing with the distance that was observed in healthy animals (Figure 6.2-A) no longer exists in the equivalent analysis of the hearing-impaired mice (Figure 6.8-A), under both pure-tone and two-tone conditions. This indicates that the spatial structure of nearby neurons tending to have higher signal correlations in healthy animals does not prevail in the signal correlation networks inferred from hearing-impaired animals. On the other hand, while there was no significant trend in noise correlations in healthy mice (Figure 6.2-B), there does exist a significant trend for noise correlations to decrease with distance in hearing-impaired mice (Figure 6.8-B). In addition, we see that noise correlations are less sparse in hearing-impaired animals (Figure 6.8-B) than in the healthy ones

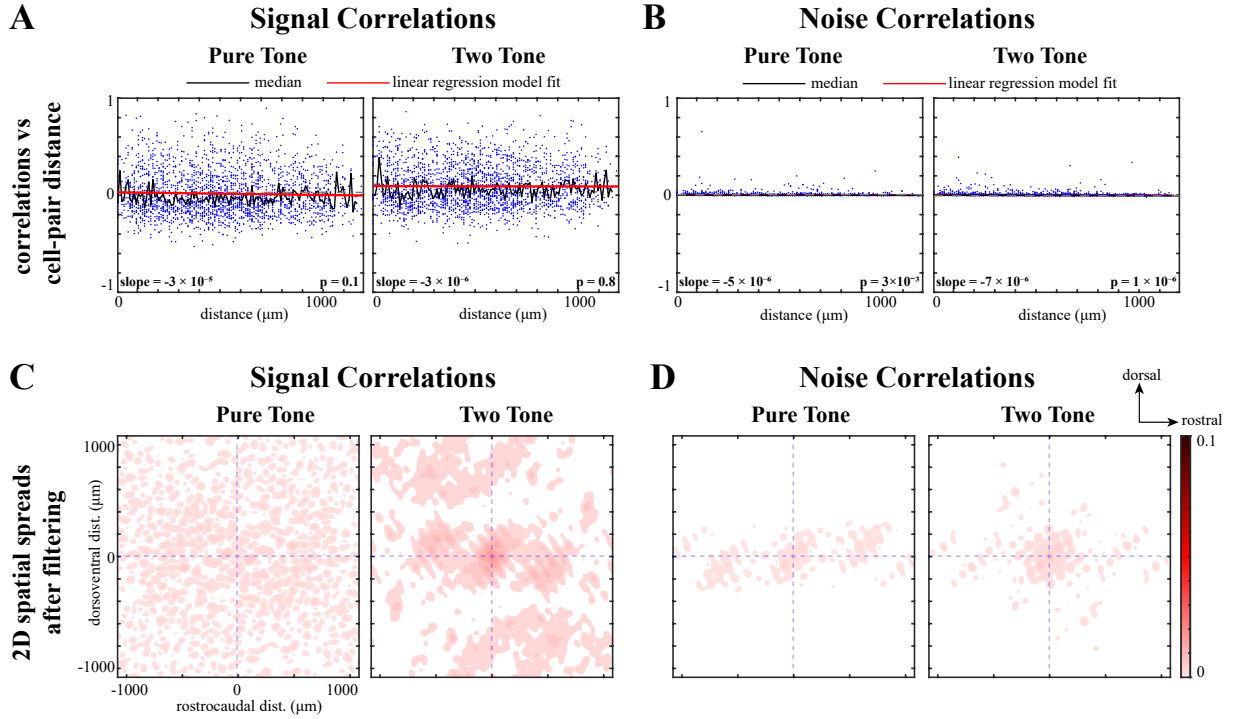


Figure 6.8: Spatial patterns of correlation networks inferred from mice diagnosed with OTOF-related hearing loss. (A) Variation of signal correlations vs. cell pair distance, under pure-tone stimuli (left) and two-tone stimuli (right). Distances were binned to $10 \mu\text{m}$ intervals. The median of the distributions (black) and the linear model fit (red) are shown in each panel. The slope of the linear model fit, and the p-value of the t-test to identify the statistical significance of the slope are also indicated as insets. Dashed horizontal lines indicate the corresponding zero-slope lines for ease of visual comparison. (B) The same configuration as panel A, for noise correlations. (C) Contour plots of statistically significant 2D spatial patterns of signal correlations inferred under pure-tone stimuli (left) and two-tone stimuli (right), after performing filtering in the 2D FFT domain. The horizontal and vertical axes in each panel respectively represent the relative rostrocaudal and dorsoventral distances between each cell pair. (D) Same organization as panel C, for noise correlations.

(Figure 6.2-B). This suggests that in hearing-impaired animals there are more spontaneous connections than in healthy ones, while spontaneous connectivity also being stronger over a shorter distance (similar to the stimulus-driven connectivity in healthy animals, thus it is as if the structure of the signal and noise correlations are reversed in hearing-impaired animals comparative to the healthy ones).

Consistent with the changes in the distribution of correlations vs cell-pair distances noted

in Figure 6.8-A and -B, we see equivalent transformations in the analysis of significant spatial patterns in correlation networks (Figure 6.8-C and -D) as well. The central lobe that was strongly noted in both pure-tone and two-tone signal correlation spatial patterns in healthy mice (Figure 6.3-C), is completely absent in the corresponding spatial patterns of hearing-impaired mice (Figure 6.8-C). This indeed corroborates that the tendency of signal correlations to be stronger between nearby neurons no longer persists in this control dataset. On the other hand, the spatial patterns of noise correlations in Figure 6.8-D have a more distinct structure than those in Figure 6.3-D. While the noise correlation spatial patterns in Figure 6.3-D display a sporadic nature, the ones in Figure 6.8-D have a multi-modal structure with a distinguishable central component (under both pure-tone and two-tone conditions).

In conclusion, we see that in the case of hearing-impaired mice, spontaneous connectivity has a more evident network organization than stimulus-driven connectivity, whereas we noted the inverse in healthy mice. This is indeed what one would also intuitively expect; since stimulus encoding and decoding are much weaker in the auditory cortex of hearing-impaired mice, it is natural for the stimulus-driven responses to be much weaker than spontaneous activity. Moreover, this also shows that spontaneous networks are stronger and better organized in the primary auditory cortex of hearing-impaired mice than in healthy mice.

6.1.3 Concluding remarks

In summary, we demonstrated in this section how network-level organizations in the mouse primary auditory cortex change under two different auditory stimuli: pure-tones vs two-tones. By closely inspecting the spatial patterns of the underlying signal and noise correlation network pat-

terns, we concluded that neuronal pairs at a larger cell-pair distance are more likely to have high signal correlations under two-tone conditions than pure-tone conditions. We also showed that the relative 2D spatial patterns of signal correlations under pure-tone conditions have a uni-modal centered structure, whereas the spatial patterns of signal correlations under two-tone stimuli were multi-modal and widespread. Further investigation into the source of these differences revealed that while the central lobe of the two-tone signal correlation spatial patterns was majorly contributed by interactions between neuronal pairs with the same best frequency, the off-centered widespread connections were predominantly a consequence of the pairwise interactions between neuronal pairs with different best frequencies. By diving into this further, we demonstrated that the dominant source of the widespread connectivity in the two-tone signal correlation spatial patterns was the suppressive frequency interactions, whereas the facilitative frequency interactions seemed to primarily occur over shorter distances.

Next, by performing the same spatial analysis on an equivalent experiment with a different set of frequencies, we demonstrated that the overall spatial trends remain largely invariant. In particular, while the original set of frequencies we considered were harmonics, in this case, we used a set of frequencies that had non-integer octave frequency spacing, to eliminate any biases incurred by the frequencies being harmonics. From this analysis, we observed that the uni-modal structure of pure-tone signal correlation spatial patterns and the multi-modal, widespread patterns in the two-tone signal correlation networks were consistently preserved even under the non-harmonic frequencies. However, while the widespread connectivity under the harmonic frequencies was primarily orientated in the dorsoventral direction, the directional preference of widespread connectivity under the non-harmonic frequencies was in contrast rostrocaudal. Thus, this reveals that the orientations of the widespread connectivity in the two-tone signal correlation patterns are

sensitive to the choice of frequencies, while the overall patterns of activity are not.

Finally, by performing an equivalent analysis on a similar dataset collected from hearing-impaired animals, we illustrated that none of the aforementioned significant spatial patterns of signal correlations inferred from healthy mice persisted in hearing-impaired mice. On the other hand, noise correlations were much stronger in hearing-impaired mice than in healthy ones, and noise correlations displayed notable spatial patterns, unlike in healthy mice. This suggests that stronger spontaneous activity coordination exists in hearing-impaired animals than in healthy ones, possibly due to stimulus-driven activity encoding being much weaker. All these results facilitate us to gain a better understanding of how stimulus-encoding and spontaneous networks in the neurons of the auditory cortex depend on the type of stimulus, and also on the hearing capabilities of the subjects. In conclusion, the work presented in this section provides us with novel meaningful intuitions about how the network-level organizations in the primary auditory cortex change under different stimulus conditions.

6.2 Network-level changes during perceptual learning

In the second part of this chapter, we apply the signal and noise correlation methods introduced in Chapter 5 on electrophysiological data recorded from the auditory cortex of gerbils under auditory task performance (data from [80]). The animals in the dataset had been trained to drink from a water spout while in the presence of the *safe* stimulus (unmodulated noise), and to withdraw from the spout when the sound changed to the *warn* stimulus (AM - Amplitude Modulated noise), to avoid an aversive shock (see [80] for details). During each day, the animals had been presented with five different warn stimuli of varying AM depths (the level of task difficulty

depends on the AM depth), and the depths had been changed from day to day based on the animal's performance [80]. The dataset also included multi-unit recordings from the left auditory cortex, as animals trained and improved on the AM detection task [80]. In addition, auditory cortex activity during disengaged listening sessions that occurred just before (*pre-passive*) and just after (*post-passive*) training sessions had also been recorded. The spout and metal floor plate had been removed from the test cage during the disengaged sessions; otherwise, the sound stimuli presented and the position of the recording electrodes had been identical to behaviorally engaged sessions [80].

6.2.1 Results

For each day in the course of perceptual training, we first extracted the signal and noise correlation matrices from the multi-unit recordings under different AM modulation levels, separately for pre-passive, actively engaged, and post-passive conditions, using the method outlined in Algorithm B.1 of Chapter 5. Figure 6.9 shows sample correlation matrices thus obtained, from data recorded during the first day of perceptual training. While the signal correlation matrices (Figure 6.9-A) are not very similar across the two modulation levels and engaged-disengaged sessions, the noise correlation structure (Figure 6.9-B) seems to be quite consistent across the different conditions. This is indeed expected from the design of the experiment because the trials corresponding to different modulation levels had been presented to the animal randomly interleaved rather than in separate blocks [80], and noise correlations account for ongoing spontaneous activity.

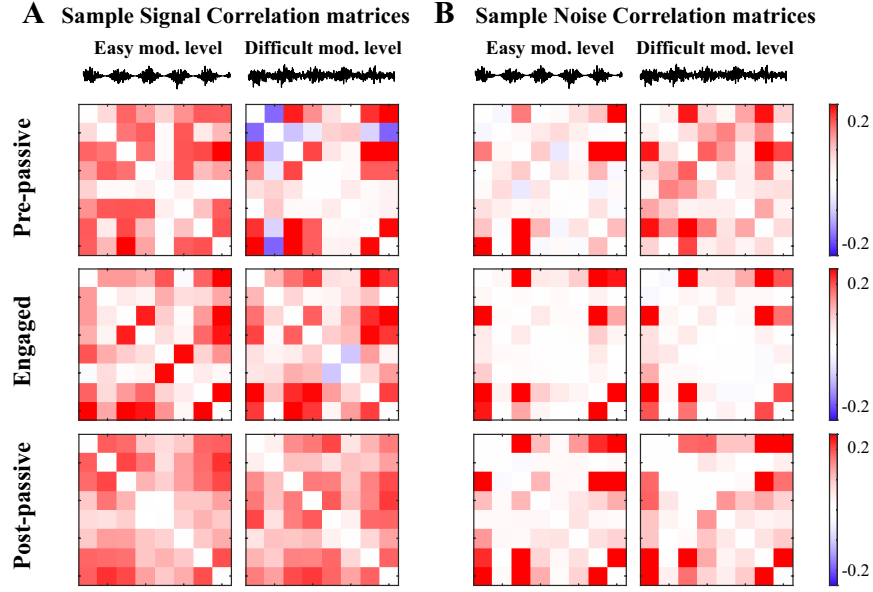


Figure 6.9: Sample signal and noise correlation matrices estimated from multi-unit recordings from the gerbil auditory cortex under auditory task performance. (A) Sample signal correlation matrices extracted from data recorded while the animal was in pre-passive (top), actively engaged (middle), and post-passive (bottom) sessions on the first day of perceptual training, for an easy (left) and a difficult (right) modulation level. (B) Sample noise correlation matrices organized in the same layout as panel A.

6.2.1.1 Pre/Post-passive vs actively-engaged network similarity

Next, to compare the similarity of correlation networks inferred under the two disengaged sessions with the actively-engaged session, we performed the following analysis. On each day, for each AM modulation level, we computed the similarity (via pairwise correlation) between the correlation matrices inferred under the pre-passive session and those inferred under the actively-engaged session, and also likewise between the post-passive and actively-engaged sessions. Figure 6.10 shows the corresponding results. While a notable trend does not seem to exist in the variations of network similarity between disengaged sessions and engaged sessions across days, from the rightmost panels in Figure 6.10 we see that the average similarity between post-passive and active networks are consistently higher than the average similarity between pre-passive and

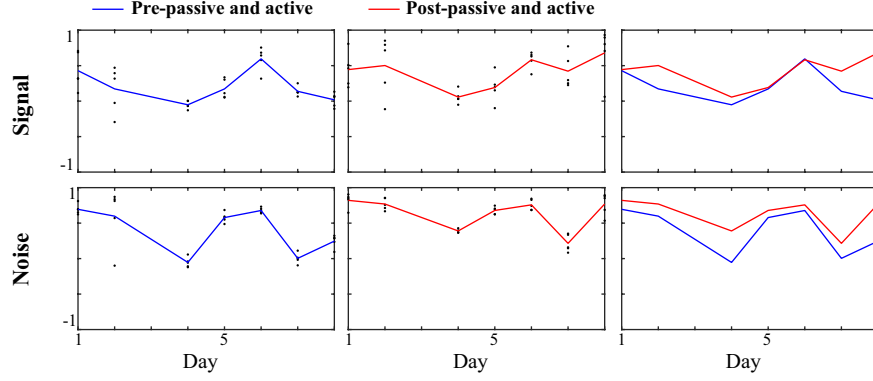


Figure 6.10: Pairwise correlations between the signal (top) / noise (bottom) correlation networks inferred for each AM depth on each day, under the pre-passive and actively engaged sessions (left), and under the post-passive and actively engaged sessions (middle). The black dots show the pairwise correlations corresponding to the five different modulation depths on each day, and the solid lines (blue: pre-passive vs active, and red: post-passive vs active) indicate the variation of the mean across the five modulation depths, with days of perceptual training. The rightmost panels show the mean pre-passive vs active and post-passive vs active variations on a single plot, for the ease of visual comparison.

active networks, for all days and both signal and noise correlation networks. This suggests that both signal and noise correlation networks under actively engaged sessions on average are more similar to the corresponding networks under post-passive sessions than those inferred from pre-passive sessions. This result is indeed consistent with the task-specific plasticity effects illustrated in [80], corroborating that disengaged neural sensitivity is better if measured immediately after (post) behavioral testing, rather than before (pre).

6.2.1.2 Similarity of correlation networks across modulation levels vs behavior

Next, to investigate the relationship between network-level changes with behavior, we performed the following analyses on the correlation networks inferred under actively-engaged sessions. Firstly, for each day, we computed the pairwise correlation between signal/noise correlation matrices extracted for each pair of AM depths presented during that day. Then, we observed

the variation of those pairwise correlations relative to the corresponding performance of the animal, by plotting them as a function of the average hit rate in those trials, as shown in Figure 6.11. Note that the results in Figure 6.11 are pooled across days of training. Consistent with what was also noted in Figure 6.9, we see in Figure 6.11 that the pairwise correlations between the noise correlation matrices across AM depths are very high (~ 1) independent of the performance. On the other hand, the similarity between signal correlation matrices across AM depths (Figure 6.11 left) seems to have a clear dependency on the performance. While the signal correlation networks across modulation depths are also very similar to each other (as suggested by pairwise correlations ~ 1) consistently at modulation depths with small hit rates (< 0.5), there is no such notable trend between the signal correlation networks inferred on modulation levels with better performance (hit rate > 0.5). This indeed suggests that the similarity of signal correlation networks across AM depths has a dependency on the performance and that signal correlation networks under AM depths with poor performance seem to be always similar to each other.

For a closer inspection of the similarity between signal correlation matrices across modulation levels, we illustrate the pairwise correlations between the signal correlation networks extracted at the five different AM depths (ranked from 1 to 5 based on the level of task diffi-

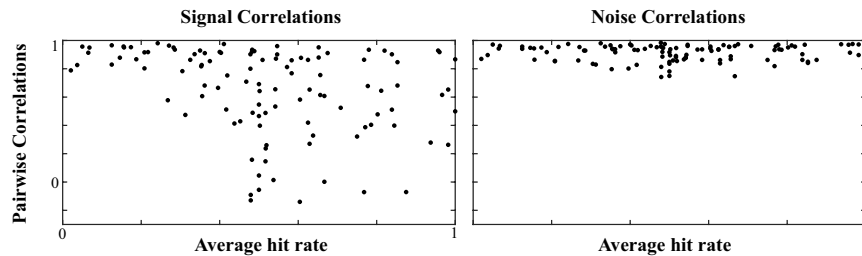


Figure 6.11: Pairwise correlations between signal (left) / noise (right) correlation matrices across different modulation depths vs the average hit rate of the corresponding modulation depths. The comparisons were performed independently for each day, and the plots show the results pooled across days.

culty) from day 1 to day 4 of perceptual training, in Figure 6.12 (diagonal elements are all equal to 1 and omitted for visual convenience). From Figure 6.12, we see that the two most difficult modulation levels (ranked 4 and 5) have high similarity steadily from day 1. This is indeed very consistent with the phenomena we also noted through Figure 6.11 left panel; the pairwise correlations between signal correlation matrices are consistently high in cases with low average hit rates (corresponding to the difficult modulation levels). On the other hand, the similarity between the signal correlation structures extracted from the relatively easier modulation depths (ranked 1-3 in Figure 6.12) seemingly increases with days of perceptual learning.

To further investigate this effect, Figure 6.13-A shows the signal correlation matrices of an easy modulation level (left - level 1) and a difficult modulation level (right - level 3) in an early day (top: day 2) and a later day (bottom: day 7) in the course of perceptual learning. Even visually, one can see that the two networks inferred on day 7 are more similar to each other than the two networks inferred on day 2. Next, to quantify this visual comparison, for each day we computed the mean and standard deviation of the pairwise correlations between signal correlation networks across the three modulation levels ranked 1 – 3 and observed how that varies across days (Figure 6.13-B top panel). Note that we excluded the two most difficult modulation levels

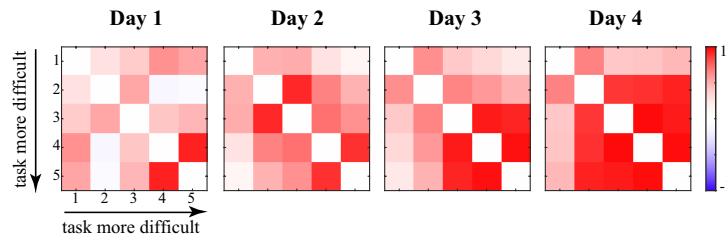


Figure 6.12: Similarity of signal correlation networks across modulation levels on different days (day 1 to day 4 from left to right). Each panel shows the pairwise correlations between the signal correlation networks extracted at the five different AM depths (ranked from 1 to 5 based on the level of task difficulty) presented on that day.

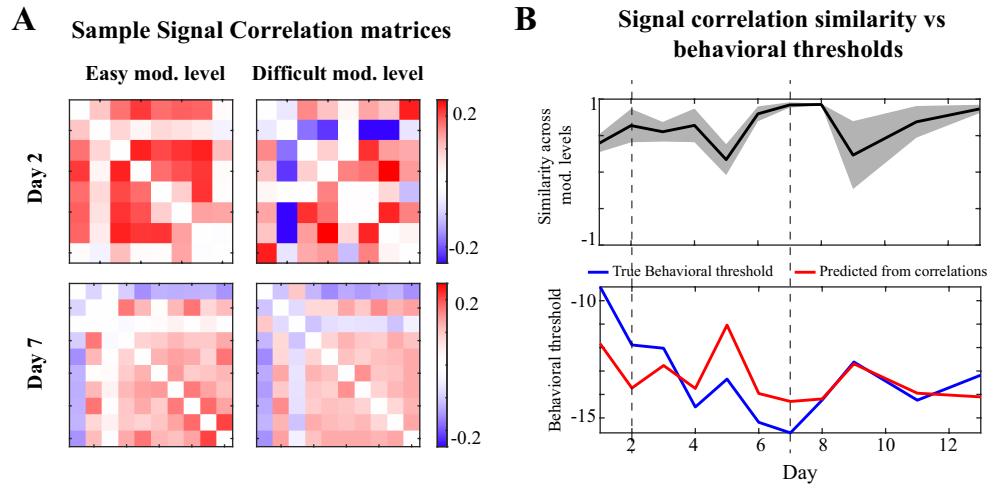


Figure 6.13: Identifying network-level changes during perceptual learning (A) Sample signal correlation matrices estimated for an easy modulation level (left - level ranked 1) and a difficult modulation level (right - level ranked 3) in an early day (top: day 2) and a later day (bottom: day 7) during perceptual training. (B) Top: The variation of the similarity of signal correlation networks across the modulation levels ranked 1-3, with days. The black line indicates the mean and the shaded grey area indicates the standard deviation. Bottom: The comparison of the true behavioral thresholds (blue) with the behavioral thresholds predicted (using linear regression) by the mean and standard deviation of the similarity of signal correlation networks (red) across different modulation levels. The dashed vertical lines indicate the two days shown in Panel A.

(ranked 4 and 5) in this analysis since they displayed a different trend as we illustrated in Figure 6.12. Consistent with what we also visually noted in Figure 6.12, we see from the top panel of Figure 6.13-B that the similarity across modulation levels indeed increases from day 1 to day 4.

However, this increasing trend with days is not very consistent across the later days of training, as evident from the fluctuations (for instance, on days 5 and 9 in the top panel of Figure 6.13-B). To investigate the source of these fluctuations, next, we also observed the variation of behavioral thresholds (see [80] for details on the procedure followed in deriving the thresholds) across the same days, as shown in blue in the bottom panel of Figure 6.13-B. Interestingly, we note that the behavioral thresholds also have analogous fluctuations to those in the similarity across modulation levels. To further investigate this, we performed a linear regression on the behavioral thresholds using the mean and standard deviation of the similarity of signal correlation

networks as covariates. The behavioral thresholds thus predicted from network-level statistics are shown in red in the bottom panel of Figure 6.13-B. The resemblance between the fluctuations in the blue and red curves in the bottom panel of Figure 6.13-B indeed suggests that the similarity of signal correlation networks across different modulation levels follows a trend that is very close to the fluctuations in behavioral thresholds. In other words, this analysis suggests that the signal correlation networks across modulation levels converge better on days that the gerbils perform the auditory discrimination task more competently.

6.2.2 Concluding remarks

In this section, we presented our analysis of signal/noise correlation networks inferred (using methods introduced in Chapter 5) from spiking data recorded from the gerbil auditory cortex while the animals were learning an amplitude modulation detection task. An initial comparison between the correlation networks inferred under pre-passive, actively engaged, and post-passive conditions revealed that the actively engaged and post-passive networks are more similar to each other than the actively engaged and pre-passive ones. This indeed provides further validation to the previously reported task-specific plasticity effects [80]; i.e, post disengaged neural sensitivity is better than the pre disengaged sensitivity.

Next, we closely inspected how the correlation networks inferred during actively engaged sessions changed with the animal performance. We noted that the noise correlation networks across different modulation levels were always highly similar to each other independent of the performance, which was expected since noise correlations encoded spontaneous neural activity. On the other hand, signal correlation networks across different modulation levels indicated a clear

dependency on animal performance. One notable pattern was that the signal correlation networks across different modulation levels became more similar to each other in the course of learning. Upon close inspection, we observed that the variation of the similarity of signal correlation networks across days had identical fluctuations to the variation of the behavioral threshold (derived based on the animal's daily performance) across days. This indicates that on the days that the animal learns the task better, the signal correlation networks in the gerbil auditory cortex are more similar across modulation levels. Thus, we demonstrated through this section how the methods introduced in Chapter 5 can be used to inspect network-level changes that occur in the auditory cortex with perceptual learning, using electrophysiological data. Moreover, through these results, we also exemplify the fact that while the methods in Chapter 5 were originally developed to be applied to two-photon calcium imaging data, they can easily be adapted to infer the signal and noise correlation networks from neuronal spiking data as well.

Part III

Zebrafish respiration-locomotion network coordination

Chapter 7: Investigating and modeling larval zebrafish respiration-locomotion coordination neural mechanisms

The recent advancements of high-speed light-sheet microscopy techniques to image almost all neurons in the intact larval zebrafish brain at single-cell resolution have facilitated the exciting prospect of examining the functional and anatomical structure of activity patterns across the entire brain [6, 7]. The whole-brain imaging data thus obtained have been used in previous work to investigate various functional circuits of the zebrafish brain; such as locomotion, sensory motor-transformation, and motor adaptation [6, 159–162]. Moreover, the ability to simultaneously monitor the motor response of the zebrafish by the means of electrophysiological techniques (see [160, 163] for details), provides us the unique opportunity of observing behavior along with the concurrent whole-brain neural activity.

While neural oscillations are known to play a significant role in mediating the cognitive and motor functions of the brain [37–39], there is very limited evidence of the presence of neural oscillations in the zebrafish brain. Upon closer inspection of whole-brain light-sheet imaging data and electrophysiological data recorded from the zebrafish brain (data from the Ahrens Lab, Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA) we observed consistent widespread synchronous brain oscillations, that have not been discovered and analyzed in previous studies to the best of our knowledge. Subsequently, we noted that these oscillations are

also synchronous with the jaw movement during breathing, from which we deduced that these oscillations corresponded to the zebrafish respiratory rhythm. In the rest of this chapter, we present the analyses we performed on whole-brain imaging data and electrophysiological recordings, to investigate and understand the functionality and circuitry of these respiratory brain oscillations.

7.1 Motivation: Preliminary observations

Firstly, we noted that an EEG (Electroencephalography) signal recorded from the tectum of the zebrafish brain displayed very consistent rhythmic oscillations. By performing a sliding window multitaper spectral analysis [40,43], we discovered these oscillations to be in the 2–4 Hz range and persist at all instances of time. Figure 7.1 illustrates the Power Spectral Density (PSD) derived through multitaper spectral estimation, in an experiment that was 15 min long. Both the time domain signal (Figure 7.1 bottom) and the PSD indeed validates the consistency of these EEG oscillations. Subsequently, we noted that these oscillations are also in sync with the jaw movement, from which we deduced that they correspond to the respiratory rhythm of the

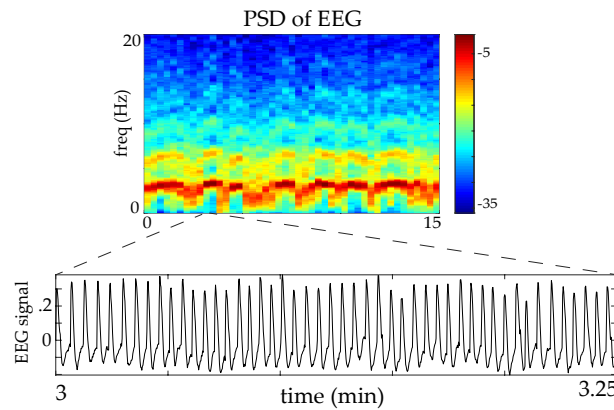


Figure 7.1: Rhythmic activity in the EEG recorded from the zebrafish tectum. The power spectral density (PSD) of the EEG recording in a 15 min experiment is shown on the top. The bottom figure shows the time domain EEG oscillations in a sub-interval (3 – 3.25 min).

zebrafish. Since the respiratory rhythm could be recorded very robustly and regularly at the macro-scale, we were interested in identifying the neural sources of this EEG. To that end, we utilized whole-brain light-sheet imaging recordings of larval zebrafish.

Accordingly, calcium fluorescence observations of each neuron in the zebrafish brain were recorded along with the EEG, under spontaneous conditions. For example, in the above experiment, we observed the fluorescence activity of 9951 cells in the brain (from plane 10), spanning from the hind-brain to the fore-brain. To identify the neurons that contain the respiratory rhythm, we performed the following analysis. We divided the 15 min experiment into windows of length 16 seconds, and then for each cell, computed the coherence [103] between the fluorescence observations and the recorded EEG signal, at the peak EEG oscillation frequency within that time window. Then, we computed the average coherence of each cell with EEG across all time windows, which is indicated in Figure 7.2-A. Based on this coherence analysis, we identified two primary neuronal clusters in the hind-brain with very high coherence: the Caudal-Medulla Oblongata (C-MO in Figure 7.2-B) and the Lateral-Medulla Oblongata (L-MO in Figure 7.2-C).

Then, to verify whether the fluorescence recordings of the cells within each of these clusters

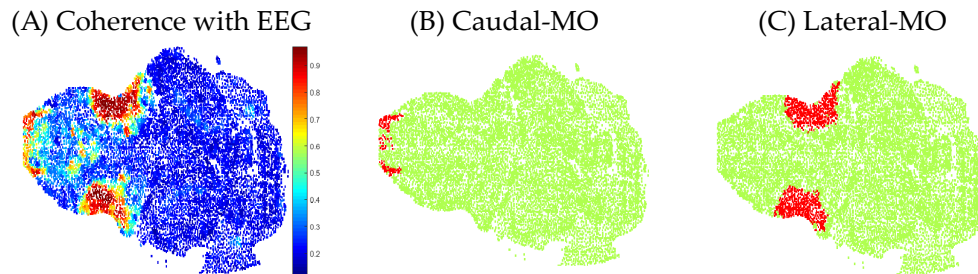


Figure 7.2: Identifying the main brain regions that have oscillations in sync with the EEG. (A) The intensity of the coherence between the fluorescence observations of individual cells and the EEG. (B) and (C): two neuronal clusters with high coherence: (B) Caudal-Medulla Oblongata (C-MO) and (C) Lateral-Medulla Oblongata (L-MO).

are similar to each other, we performed K-means clustering on the fluorescence observations of all 9951 cells, setting the number of clusters to an arbitrary large value ($K = 80$). Interestingly, the Caudal-MO and Lateral-MO neuronal clusters grouped into two distinct clusters, while not grouping with the other brain areas. This further suggested that the Lateral-MO and Caudal-MO are indeed two distinct neuronal clusters, with similar functionality within each group.

Then, we derived the average fluorescence activity within each of these two neuronal clusters. Figure 7.3 shows the average fluorescence activity of the Caudal-MO and Lateral-MO in comparison to the EEG in a sample time window. Even though neuronal recordings have more observation noise than the EEG, from Figure 7.3 we see that these two brain regions do have very similar oscillations to the EEG. This is further evident after low-pass filtering of the signals (Figure 7.3 right), which demonstrates that there indeed exist oscillations that are in sync with EEG in these two brain regions. Moreover, from Figure 7.3 we also note that the oscillations in Caudal-MO and Lateral-MO seem to be anti-phasic relative to each other.

Along with the EEG, EMG (electromyography) from the tail was also recorded to identify fictive swim events [163]. When a fictive swim event occurs, the EMG has a significant increase

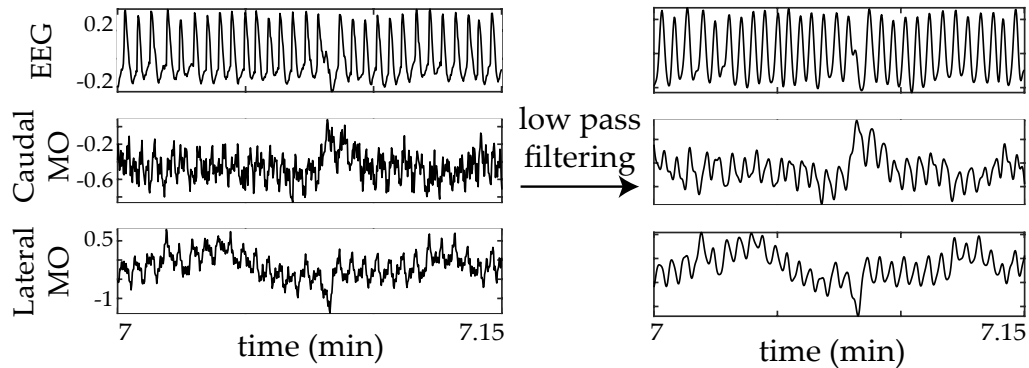


Figure 7.3: The EEG and the average fluorescence observations from Caudal-MO and Lateral-MO within the time window 7 – 7.15 min are shown on the left. The right panel shows the same, after low-pass filtering to reduce the observation noise in fluorescence activity.

in magnitude. Figure 7.4 shows a comparison of the EEG, Caudal-MO, and Lateral-MO with the EMG activity, in a one-minute time window. We noted several interesting phenomena through this comparison. In the occurrence of a swim bout (observed as spiking events in EMG), there are significant changes in the EEG, Caudal-MO and Lateral-MO oscillations. Specifically, we observed that the frequency of EEG oscillation changes when swim bouts occur. Further, swim events seem to cause an increase in the magnitude of fluorescence activity in both brain regions, more prominently observed in the Caudal-MO.

It was evident from all the aforementioned preliminary observations that these respiratory brain oscillations play a very important role and could be key to understanding both respiration and motor circuitry in zebrafish. In what follows, we interchangeably refer to this EEG-sync brain rhythm as *oscillations* and *respiratory-rhythm*. Understanding the relationships between the Lateral-MO and Caudal-MO oscillations, the relationships between the EEG and the oscillations from different brain regions, and the relationships between swim bouts and the respiratory rhythm

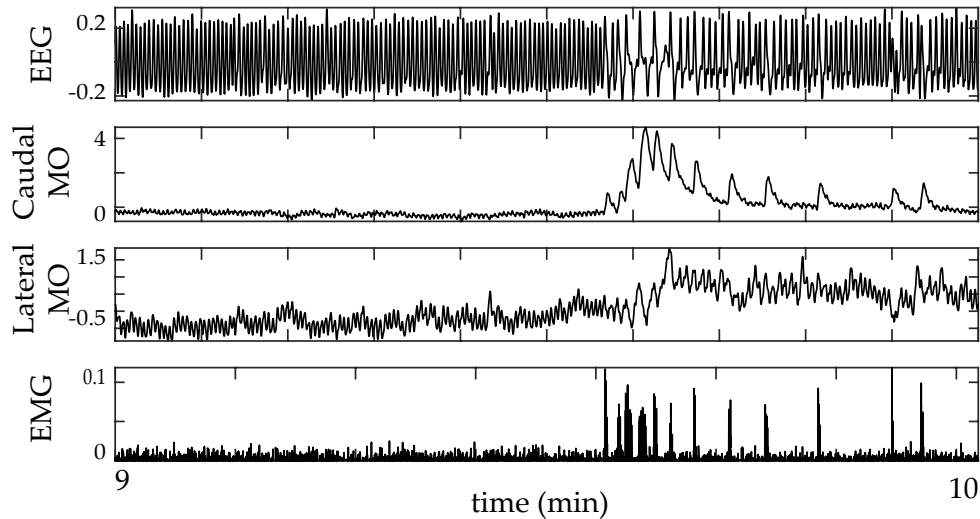


Figure 7.4: Comparison of the breathing-related oscillatory signals with the EMG recording (swim bouts). The EEG, the average fluorescence observations from Caudal-MO and Lateral-MO, and the EMG recording within the time window 9 – 10 min are shown from top to bottom.

were some of the interesting features to explore. In the subsequent sections, we first present the deductions we formed based on these analyses and finally introduce the circuit model that we propose combining all empirical evidence.

7.2 Cross Spectral and Spectral Granger Causality analysis

To understand the nature of the spectral coupling between the EEG, Caudal-MO, and Lateral-MO, we first performed a cross power spectral density analysis. We divided the experimental duration into non-overlapping windows of length 20 seconds and performed sliding window multitaper spectral analysis to derive the PSD matrix. Figure 7.5 shows the magnitude and phase spectra between each pair of processes over different time windows. The time windows in Figure 7.5 are ranked in the decreasing order of EMG activity, to understand the nature of spectral coupling in the presence and the absence of swimming.

From the magnitude spectrograms in Figure 7.5, we see that all three processes have highly coupled activity. In time windows with more swim activity, the spectral coupling between the processes occurs at a lower frequency, around 1 Hz. Further, the power of this swimming-related activity seems to be greater in the Caudal-MO than in the Lateral-MO. On the other hand, in the time windows with less swim activity, all three processes display consistently coupled oscillations around 3 Hz. While the EEG signal displays the strongest oscillations, the spectra of the Caudal-MO and Lateral-MO also follow it very closely. All these observations indeed corroborate that the breathing rhythm changes with swimming, consistent with the changes noted in the temporal sequences in Figure 7.4. The phase spectrograms in Figure 7.5 also signifies a key phenomenon. Note that the phase shift between each pair of processes is very consistent across

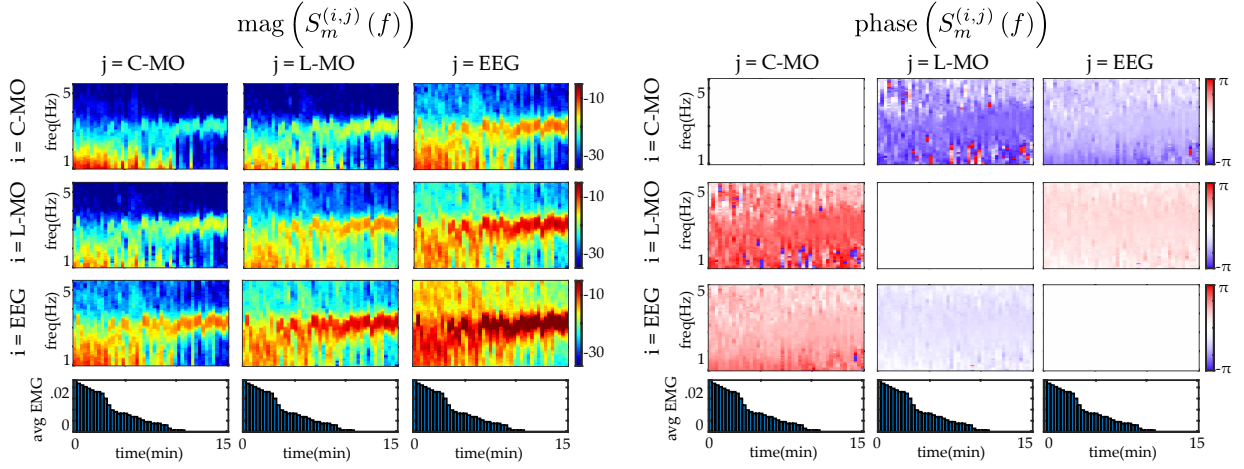


Figure 7.5: Cross power spectral density analysis. The magnitude spectrograms $\left(\text{mag}\left(S_m^{(i,j)}(f)\right)\right)$ in decibels are shown on the left and the phase spectrograms $\left(\text{phase}\left(S_m^{(i,j)}(f)\right)\right)$ in radians are shown on the right. In both cases $S_m^{(i,j)}(f)$ refers to the cross power spectrum between the i^{th} process and the j^{th} process in time window m and at frequency f . For ease of comparison, time windows are ranked in decreasing order of swimming activity, as indicated by the average EMG activity below each column.

all time windows. This implies that these three signals are highly correlated with each other, and could potentially be time-shifted versions of the same stochastic process.

Despite cross-spectral analysis being instrumental for identifying spectral content and correlations at different frequencies, it does not imply the directions of these interactions. To identify the direction of these functional links, we also performed a spectral Granger Causality (GC) analysis [50, 51, 114]. We derived both the pairwise GC links and the conditional GC links from each process to the other two, as shown in Figure 7.6. We tested the statistical significance of the derived links through an empirical test, as in Chapter 3. The Spectral GC links thus derived revealed several interesting phenomena.

We saw pairwise GC links from the Caudal-MO to Lateral-MO, specifically around 3 Hz during time windows with less swimming, and near 1 Hz in time windows with high EMG activ-

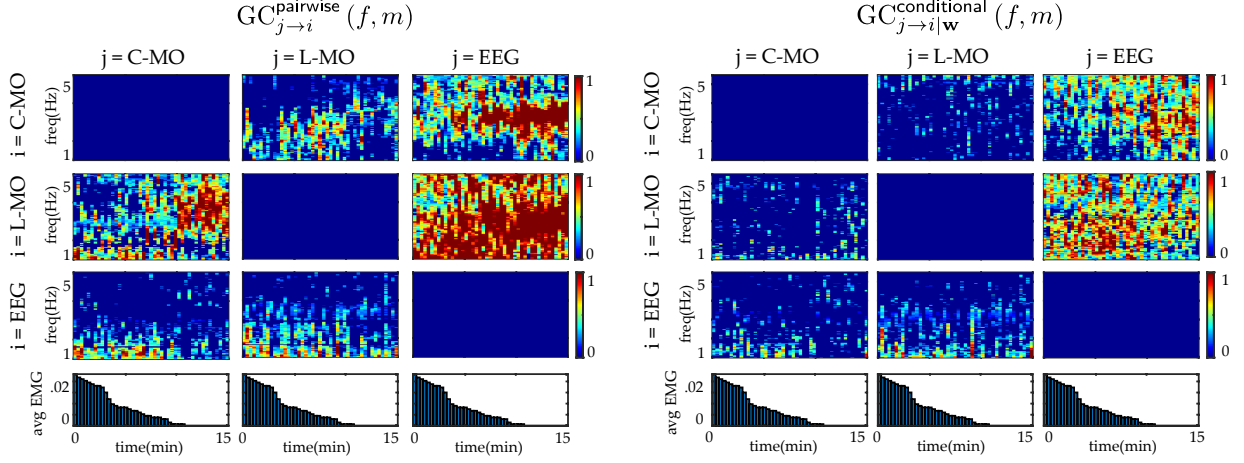


Figure 7.6: Spectral Granger Causality analysis. The pairwise Granger Causal links ($GC_{j \rightarrow i}^{\text{pairwise}}(f, m)$) are shown on the left and the conditional Granger Causal links ($GC_{j \rightarrow i|\mathbf{w}}^{\text{conditional}}(f, m)$) are shown on the right. In the case of conditional GC links, $\mathbf{w}$ refers to the other processes excluding i and j . All GC links are statistically significant with $\alpha = 0.01$ according to the empirical test. For ease of comparison, time windows are ranked in descending order of swimming activity, as indicated by the average EMG activity below each column.

ity. However, when conditioned on the EEG, these links no longer seem to exist, as apparent from the conditional GC results. Further, there exists strong and consistent pairwise GC links from the EEG to both Cadual-MO and Lateral-MO, specifically around 3 Hz. Even though these links are much weaker in the conditional GC analysis, we can still see some traces of them in the conditional GC results. These observations further strengthen our initial hypothesis that these three processes are indeed strongly coupled, and they in fact could correspond to time-shifted versions of a single underlying oscillatory stochastic process. To further validate this, we performed the following simulation as an attempt to replicate these empirical observations.

7.2.1 Replicating the spectro-temporal coupling in real data using AR modeling

The objective of this simulation is to replicate the spectro-temporal coupling (near 3 Hz) observed between Caudal-MO, Lateral-MO, and EEG in time windows with fewer swimming events, using Auto-Regressive (AR) processes. Note that the lower frequency (near 1 Hz) couplings during time windows with more swimming activities also displayed a very similar phase and GC coupling. Thus, these simulations indeed provide insights into the overall coupling between the different processes, during both swimming and non-swimming periods. Upon careful inspection of the temporal patterns (Figure 7.3), cross power spectral density analysis (Figure 7.5) and the spectral Granger Causality analysis (Figure 7.6), we concluded that the following model can generate oscillations with similar characteristics to the observed respiratory rhythms.

Let z_t be a latent oscillatory process, which we model by an AR(2) process that is tuned around 3.5 Hz. Then, suppose there exist three processes $x_t^{(1)}$, $x_t^{(2)}$ and $x_t^{(3)}$ which are differently time shifted and scaled versions of z_t :

$$\begin{aligned}x_t^{(1)} &= -0.1z_{t-\delta} + w_t^{(1)} + \epsilon_t, \\x_t^{(2)} &= 0.05z_{t-\delta-2} + w_t^{(2)} + \epsilon_t, \\x_t^{(3)} &= z_{t-\delta-4} + w_t^{(3)}.\end{aligned}\tag{7.1}$$

The three processes $x_t^{(1)}$, $x_t^{(2)}$ and $x_t^{(3)}$ hypothetically represent the breathing rhythms in Caudal-MO, Lateral-MO and EEG respectively. Note that all three processes have shifted versions of z_t with different time delays. Further, to replicate the anti-phasic nature of the oscillations in Caudal-MO compared to those in Lateral-MO and EEG (see Figure 7.3), we have included a

negative one factor in $x_t^{(1)}$. The noise components $w_t^{(i)}$ for $i = 1, 2, 3$ have the same variance and model their independent variations. The common noise ϵ_t in process 1 and process 2 models artifacts induced by the calcium imaging set-up. Different time shifts were included for each process considering the time shifts observed in the time-domain signal in true processes (Figure 7.3) and the phase shifts observed in the true phase spectrograms (Figure 7.5). Figure 7.7 shows a sample time-domain comparison of the three simulated processes.

Note that $x_t^{(3)}$ in Figure 7.7 has stronger oscillations comparative to $x_t^{(1)}$ and $x_t^{(2)}$, as a result of the higher SNR (see Equation 7.1). This is very similar to the nature of EEG, which displays very strong oscillations with significantly less noise than the Lateral-MO and Caudal-MO. Figure 7.8 and Figure 7.9 respectively show the cross power spectral density and spectral Granger Causality analysis of these simulated processes. Since the simulations only account for the ~ 3 Hz breathing rhythm and the true neural recordings could have many other underlying spontaneous activities in addition to these oscillations, as one would expect the magnitude spectra of the simulated processes (Figure 7.8 left) have fewer perturbations than the magnitude spectra

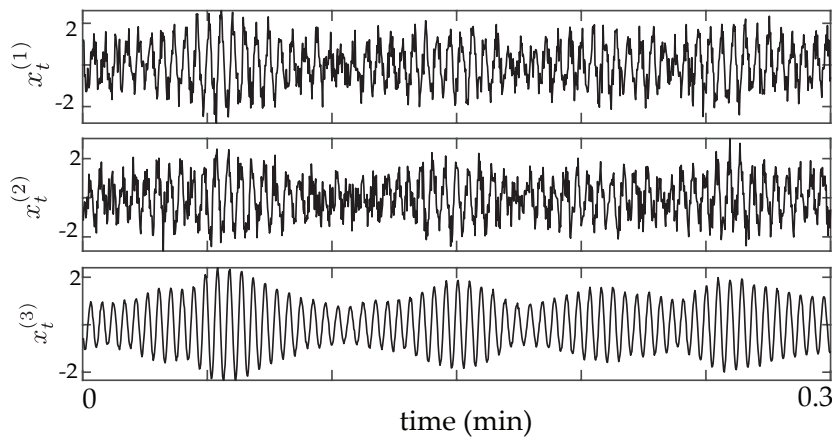


Figure 7.7: Tri-variate oscillatory process simulated using AR modeling. The processes $x_t^{(1)}$, $x_t^{(2)}$ and $x_t^{(3)}$ respectively replicate oscillations in Caudal-MO, Lateral-MO and EEG.

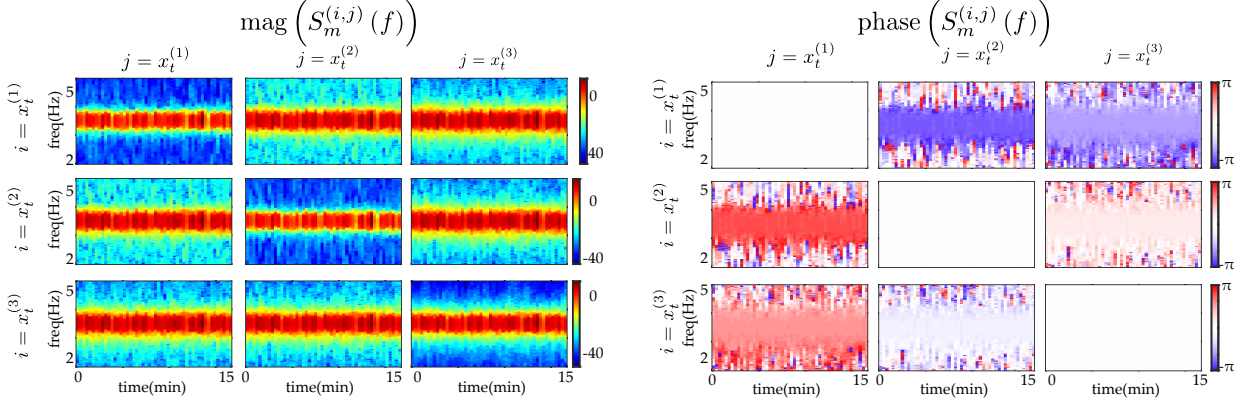


Figure 7.8: Cross power spectral density analysis of the simulated AR processes. The magnitude spectrograms $\left(\text{mag}\left(S_m^{(i,j)}(f)\right)\right)$ in decibels are shown on the left and the phase spectrograms $\left(\text{phase}\left(S_m^{(i,j)}(f)\right)\right)$ in radians are shown on the right.

of the real data (Figure 7.5 left). On the other hand, the phase spectrograms of the simulated processes (Figure 7.8 right) replicates the phase spectrograms of the real signals (Figure 7.5 left) with striking similarity.

The pairwise and conditional Granger Causal links reported in Figure 7.9 also very closely follow the corresponding results from real data (Figure 7.6). Note that spectral Granger Causality essentially quantifies the fraction of total power in one process at a given frequency that is (directly, in the case of conditional links) contributed by the other process. Thus, the presence of a GC link from j to i does not necessarily mean that the process j precedes process i in time. It is possible for the said link to exist, if process j is with a much higher SNR compared to i , even if j might be lagging i (in our simulation, $x_t^{(3)}$ displays GC links to the other two processes due to the higher SNR, despite being the most time-lagged process). This is indeed the source of the GC links from the EEG to Caudal-MO and Lateral-MO (Figure 7.6). Further, this simulation also explains why the pairwise GC links from Caudal-MO to Lateral-MO (Figure 7.6 left) do not appear in the conditional analysis (Figure 7.6 right). We see the same pattern in the GC analysis

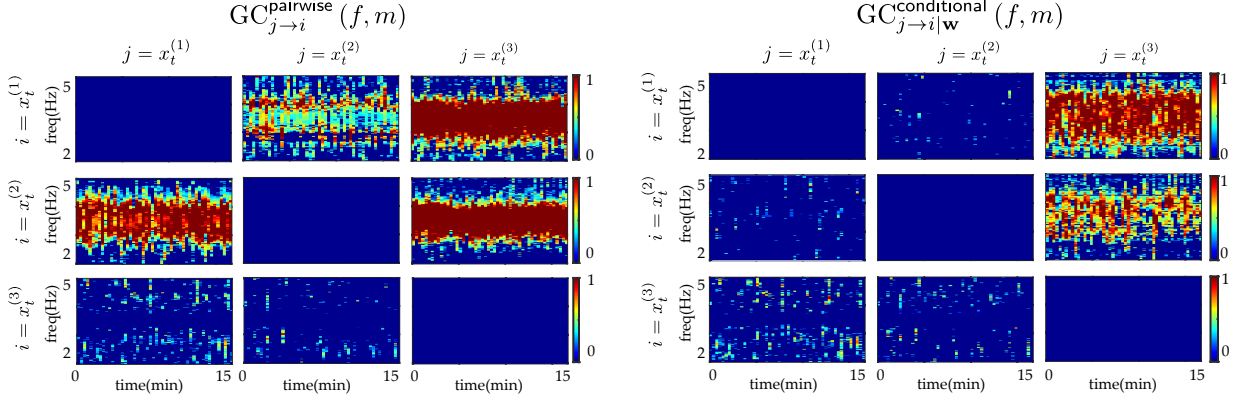


Figure 7.9: Spectral Granger Causality analysis of the simulated AR processes. The pairwise Granger Causal links $\left(GC_{j \rightarrow i}^{\text{pairwise}}(f, m)\right)$ are shown on the left and the conditional Granger Causal links $\left(GC_{j \rightarrow i | \mathbf{w}}^{\text{conditional}}(f, m)\right)$ are shown on the right. In the case of conditional GC links, $\mathbf{w}$ refers to the other processes excluding i and j . All GC links are statistically significant with $\alpha = 0.01$ according to the empirical test.

of the simulation (Figure 7.9) as well. Since all of these three processes are time-shifted versions of one another, conditioned on $x_t^{(3)}$ (which has very high SNR), the contributions from the other process become negligible.

Finally, note that even though we modeled the underlying latent oscillations by an abstract process z_t , z_t can be thought of as the breathing rhythm that pertains to the Caudal-MO. This would have been the case if the Caudal-MO were the source of the breathing rhythm (i.e., the breathing pacemaker is located in the Caudal-MO). In that case, z_t would represent the true de-noised oscillations in the Caudal-MO, with $\delta = 0$.

7.2.2 Relative time-shifts between the three oscillatory processes

From the previous simulation, we concluded that the oscillations in Caudal-MO, Lateral-MO, and EEG are essentially phase-shifted versions of a single rhythm, and Caudal-MO has an opposite polarity to Lateral-MO and EEG. Thus, to find the relative time shifts between the

three activities using the phase spectrograms, first, it is necessary to perform a sign correction. To account for this anti-polarity, it is required to multiply the average fluorescence activity of Caudal-MO by a factor of minus one, as shown in Figure 7.10. This brings all the oscillations to the same polarity, as we see in the right panel of Figure 7.10. After this modification, it is evident that the Caudal-MO precedes the oscillations in time. Then, the Lateral-MO follows and the EEG is the most lagged version. For the ease of visual comparison, we have indicated some peaks of the Caudal-MO in Figure 7.10 (right) with dashed red vertical lines.

To quantify these time-shifts, we derived the phase spectrograms of the three signals after accounting for the sign correction, as shown in Figure 7.11. It is also important to note that the magnitude spectrograms and the Granger Causal analysis remain unchanged under this sign transformation. From Figure 7.11, we see that the true temporal sequence is correctly reflected in the phase spectrograms, after sign correction. We see that Caudal-MO oscillations precede both other oscillations in time, followed by the Lateral-MO, while the EEG is the most time-lagged oscillation.

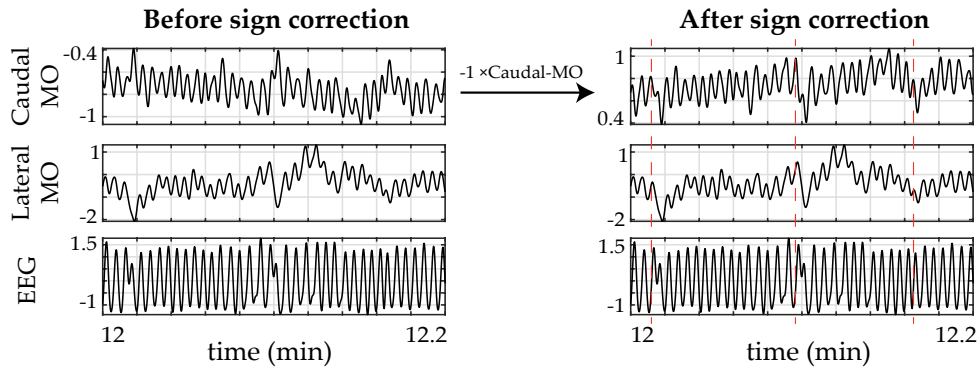


Figure 7.10: Sign correction of the Caudal-MO activity. The average fluorescence observations from Caudal-MO and Lateral-MO, and the EEG within the time window of 12 – 12.2 min are shown on left. The right panel shows the same, after performing the sign correction on the average Caudal-MO activity. For ease of visual comparison, some peaks of the Caudal-MO are indicated by dashed red vertical lines.

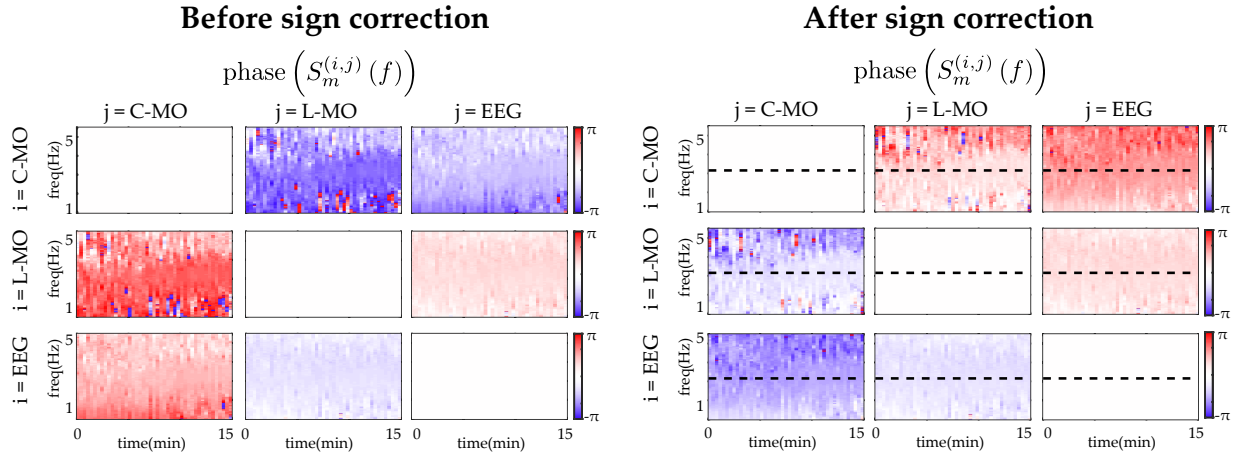


Figure 7.11: Changes in the phase spectrograms with sign correction. The phase spectrograms $\left(\text{phase}\left(S_m^{(i,j)}(f)\right)\right)$ in radians before the sign correction are shown on the left and after the sign correction of the Caudal-MO are shown on the right. The dashed horizontal lines on the right figure show the average peak EEG oscillation frequency, derived from the magnitude spectrogram of the EEG.

Table 7.1: Relative phase and time shifts at peak EEG oscillation frequency

		Phase (time) lag		
		C-MO	L-MO	EEG
Phase (time) lead	C-MO	0	0.86 rad (43 ms)	1.76 rad (89 ms)
	L-MO	-0.86 rad (-43 ms)	0	0.92 rad (46 ms)
	EEG	-1.76 rad (-89 ms)	-0.92 rad (-46 ms)	0

Next, to derive the relative time shifts, we estimated the average peak EEG oscillation frequency (f_{EEG}) using the magnitude spectrum of the EEG. Then, for each phase spectrogram, we calculated the average phase-shift ($\Delta\phi$) across all time windows at f_{EEG} , as indicated by the dashed horizontal lines in Figure 7.11 (right). Finally, phase-shifts were converted to time-shifts by $\Delta t = \frac{\Delta\phi}{2\pi f_{\text{EEG}}}$. Table 7.1 shows both phase-shifts ($\Delta\phi$) and time-shifts (Δt) thus calculated. From this, we see that the Lateral-MO oscillations lag the Caudal-MO oscillations by around 43 ms while the EEG has a further lag of 46 ms. These results indeed corroborate with the time

shifts that we originally observed between the three time series (Figure 7.10).

7.3 Swimming occurs at a distinct phase of the breathing rhythm

From our preliminary analysis (Figure 7.4), we noted that the breathing frequency changes with swim bouts. This was further confirmed by the cross-spectral (Figure 7.5) and spectral Granger Causality (Figure 7.6) analysis, by the display of lower frequency (near 1 Hz) coupling in time windows with higher EMG activity. Upon closer inspection of the breathing oscillations along with the EMG, we noted another interesting phenomenon.

Figure 7.12 shows the observed EMG, Caudal-MO, Lateral-MO, and the EEG activity, in a 0.3 min time window that had several swim bouts (spikes in the EMG). While the breathing frequency reduces when swim bouts occur, we also noted that the swim bouts seem to occur immediately after the negative peak of the Caudal-MO oscillations (or the positive peak of the sign corrected version, shown in Figure 7.12 right). For ease of visual comparison, we have

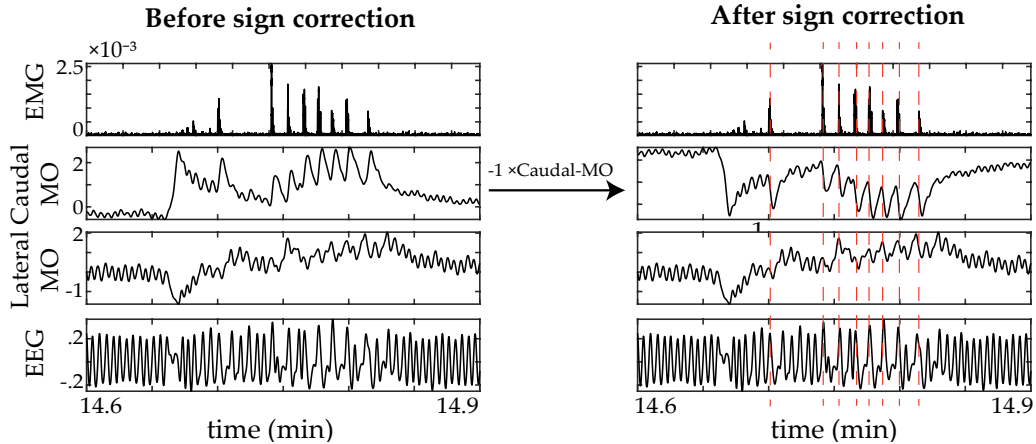


Figure 7.12: Swim events occur at a particular oscillation phase. The EMG, average fluorescence observations from Caudal-MO and Lateral-MO, and the EEG within the time window of 14.6 – 14.9 min are shown on the left. The right panel shows the same, after performing the sign correction on the average Caudal-MO activity. For ease of visual comparison, dashed vertical lines are indicated on the figure on the right, corresponding to the occurrences of swim bouts.

included some dashed vertical lines on the figure on the right, showing the occurrences of swim bouts. This pattern was consistent across all observations and different fish. To further verify this, we derived the swim triggered averages of each of these oscillatory signals, by averaging the activity three seconds before and after the peak of swim bouts. The mean and standard deviation of EMG, Caudal-MO, Lateral-MO, and EEG activity around the swim bouts are shown in Figure 7.13 (left). While the activity before and after swim bouts on average is random, that is not the case during the occurrence of swim bouts. Since swim bouts tend to occur at the negative peak of Caudal-MO oscillations (Figure 7.12 left), we observed the same pattern in the swim triggered averages (Figure 7.13 left) as well.

Next, we derived the instantaneous phases of Caudal-MO, Lateral-MO, and EEG by applying a Hilbert transform on the low-pass filtered average activity. Then, similar to swim triggered activity averages, we derived the swim triggered instantaneous phase averages as shown in Figure 7.13 right panel. This further validates the dependency of swim bouts on the phase of breathing.

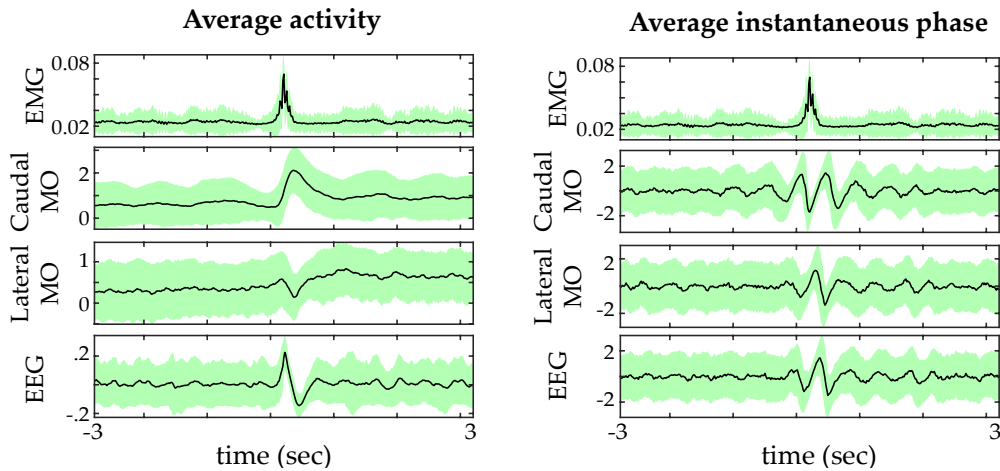


Figure 7.13: Swim triggered averages of activity and phase of oscillations. The swim triggered averages of activity three seconds before and after the EMG peak are shown on the left. Rows from top to bottom: the EMG, average fluorescence observations of Caudal-MO, average fluorescence observations of Lateral-MO, and the EEG activity. The right panel shows the swim-triggered averages of instantaneous phase in the same organization as the left panel.

While the average phase is near zero (since it is random) before and after the swim bouts, in the duration of the swim bout, the phase seems to be similar across trials (shown by non-zero mean and lower standard deviation). Further, this average instantaneous phase plot also reflects the opposite polarity of the Caudal-MO oscillations. Note that the EEG and Lateral-MO phase triggered averages are of the same shape, while the Caudal-MO phase triggered average seems flipped across the time axis. For a better exposition of this effect, we show in Figure 7.14 the comparison of the swim triggered averages of instantaneous phase, before and after correction of the sign of Caudal-MO oscillations. Note that after the sign correction, the triggered averages of Caudal-MO, Lateral-MO, and EEG all look very similar to each other in the overall pattern.

Finally, we derived the average EMG activity, triggered to the instantaneous phase of the oscillations. The phase triggered averages of the EMG relative to the Caudal-MO phase, Lateral-MO phase, and EEG phase are shown in Figure 7.15. In the case of Caudal-MO phase triggered averages, we have included the plots both before and after sign correction for comparison. Further, below each figure, we have plotted a polar histogram that shows the average EMG activity

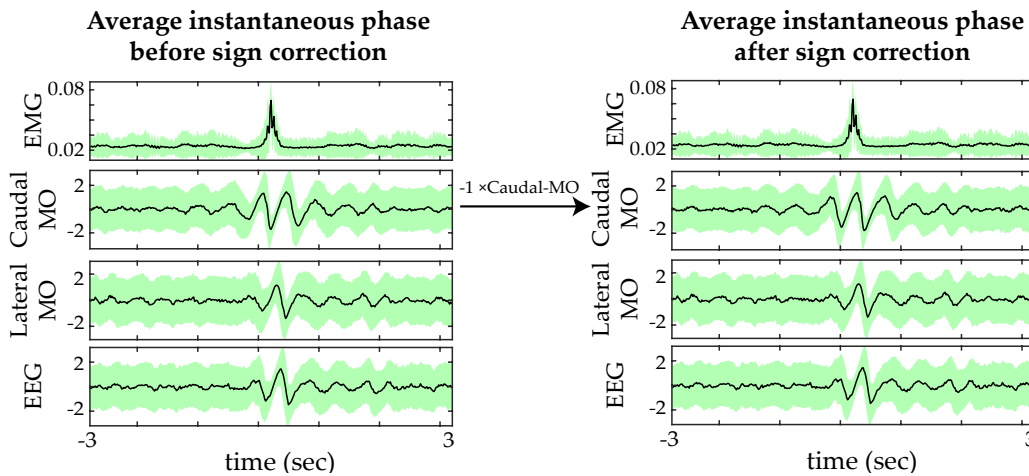


Figure 7.14: Swim triggered averages of instantaneous oscillation phases before performing sign correction on the Caudal-MO oscillations (left), and after performing sign correction on the Caudal-MO oscillations (right).

during each phase. These results further validate our hypothesis that swim bouts indeed occur at a particular phase of breathing oscillations. We consistently see this across all four distributions, with the EMG activity being significantly concentrated at a distinct phase. Further, comparing Caudal-MO after sign correction, Lateral-MO, and EEG, we see that the peak EMG activity occurs first in Caudal-MO, then in Lateral-MO, and finally in EEG; corroborating our previous results.

In summary, we see that the results derived in this section are consistent with those presented in previous sections, further validating our hypotheses regarding these oscillations. Further, from the analysis presented in this section, we see that the swim bouts are phase-locked and gated by the breathing rhythm. This suggests that the respiratory rhythm plays a major role in the locomotion gating of zebrafish.

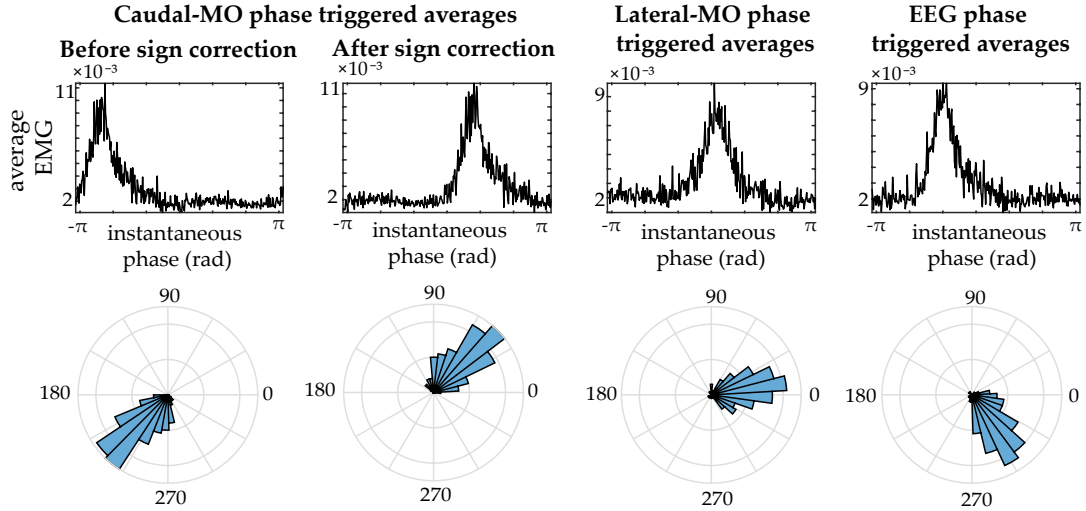


Figure 7.15: Instantaneous-phase triggered averages of EMG (swimming) activity. The average EMG activity for different instantaneous phases of the Caudal-MO oscillations (before and after sign correction), Lateral-MO oscillations, and EEG oscillations are shown from left to right. The bottom row shows the same, in the form of polar plots.

7.4 Timing of grating induced swimming depends on pre-stimulus EEG

It is well known that sensory inputs can induce swimming, and the auditory/visual-motor transformation and the brain circuitry that integrates the sensory inputs have been explored in recent work [161, 164]. However, given that the swimming circuit is phase-locked and coupled with respiration, it is also interesting to investigate the role of breathing in sensory-induced swimming. To that end, we presented a one-second visual grating stimulus to zebrafish across repeated trials, and recorded the breathing (through EEG) and swimming (through EMG) under the stimulus application. Figure 7.16 shows the averages of the EMG and EEG activity across all trials, aligned to the onset of the grating stimulus onset. The green shaded areas indicate the 95% empirical confident levels. The increased EMG activity in Figure 7.16 after stimulus presentation indicates that a grating stimulus can reliably induce swimming. Further, the corresponding changes in the EEG indicate that the breathing pattern also changes as a consequence of swimming.

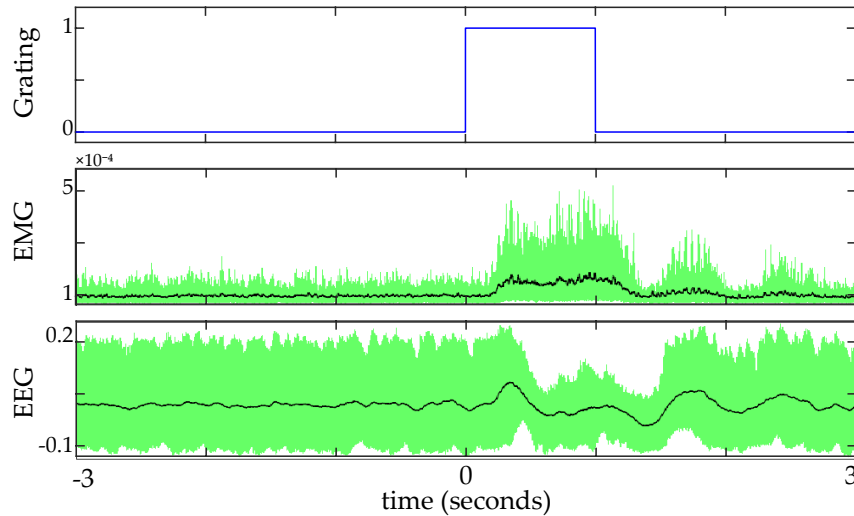


Figure 7.16: Grating stimulus onset triggered averages of EMG and EEG, three seconds before and after the stimulus presentation. The black lines show the average across multiple trials and the green shaded areas indicate the 95% empirical confident levels.

Next, to investigate the relationship between grating-induced swimming and breathing, we performed the following analysis. First, we extracted the trials that consisted grating triggered swimming, and extracted the swimming onsets of those trials as indicated by the green line in the example trial in Figure 7.17. Then, we also extracted the EEG peaks of the same trials, as shown by the yellow lines in Figure 7.17. To test whether the timing of the grating-induced swimming indeed depends on the pre-stimulus EEG, we extracted the following two time differences. As illustrated in Figure 7.17, t_1 is the time between the pre-stimulus EEG peak and the grating onset, and t_2 is the time between the grating onset and the onset of the grating induced swimming. If the swimming occurs independent of the ongoing breathing activity, one would expect t_2 to be constant independent of t_1 . To test whether t_2 is dependent or independent of t_1 , we performed a linear regression as shown in Figure 7.18. From the linear regression statistics reported in Table 7.2, we see that the linear fit is significantly different from a constant model, which shows that

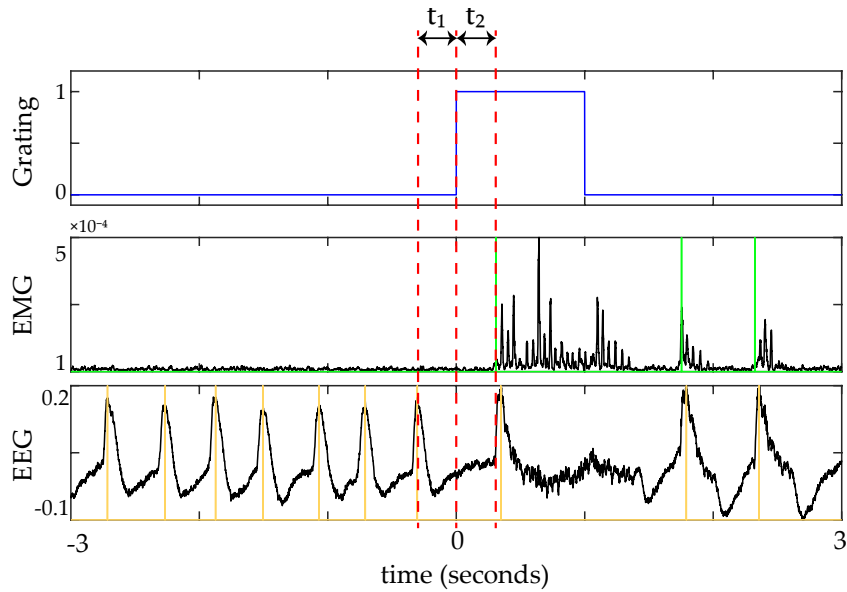


Figure 7.17: Comparing the timing of grating induced swimming with the timing of the pre-stimulus EEG. In each trial, we compare the time between the pre-stimulus EEG peak and grating onset (t_1) with the time between the grating onset and the grating induced swimming onset (t_2). The green lines indicate the swimming onset and the yellow lines indicate the EEG peaks.

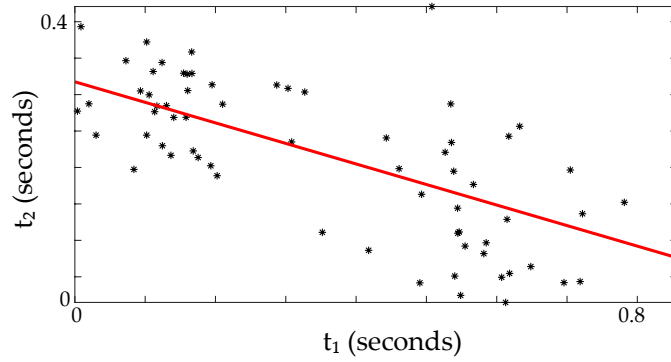


Figure 7.18: Linear regression between t_1 and t_2 to test the dependence of t_2 on t_1 .

Table 7.2: Linear regression statistics

Slope of the line	p-value	R-squared
-0.28	4×10^{-9}	0.42

the data fits the line with statistical significance.

Thus, these results clearly illustrate that the timing of the grating triggered swimming is dependent on the pre-stimulus EEG, in other words, the ongoing breathing activity. It is also noteworthy that the slope is less than one in magnitude, which indicates that the timing of the swimming is not completely independent of the stimulus presentation. This indicates that in general the next breathing cycle occurs faster than expected as a result of the stimulus presentation and the stimulus-induced swimming. However, the grating-triggered swimming cannot happen immediately after the stimulus presentation, and the timing of the swimming depends on the phase of the breathing at the stimulus onset.

We also noted that when the fish is in hypoxic conditions, sometimes grating-induced swimming does not happen immediately after the stimulus onset, but a little later. Figure 7.19 shows four example trials of activity recorded under such conditions. While in trial 1 of Figure 7.19 the grating-induced swimming occurs at the first EEG peak after the grating onset (as indi-

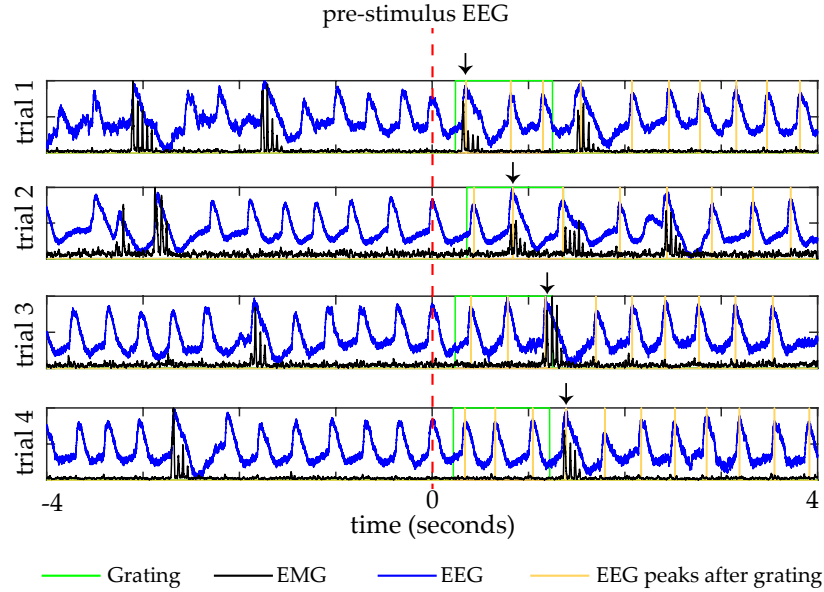


Figure 7.19: Example trials for swimming under grating during hypoxia. The one-second grating stimulus (green), EMG activity (black), EEG activity (blue), and the EEG peaks after grating (yellow) are indicated in each trial. The trials are centered on the pre-stimulus EEG for ease of comparison. The arrow indicates the time of the first swim induced by the grating visual stimulus.

cated by the arrow), it is not the case in the other three trials. It is evident from Figure 7.19 that the first swim induced by grating occurs respectively at the second, third, and fourth EEG peak after the grating onset, in trials 2, 3, and 4.

An initial inspection of the timing of swimming induced by grating revealed that it is indeed dependent on the pre-stimulus EEG, consistent with the preceding results. In addition, even the timing of swimming occurring with a delay (such as trials 2, 3, and 4 in Figure 7.19) seems to occur with minimal influence on the ongoing rhythmicity of breathing. To test this, we compared the averages across multiple trials aligned to the pre-stimulus EEG peak, with the averages aligned to the stimulus onset. Figure 7.20 shows the comparison of the pre-stimulus EEG peak triggered averages with the grating onset triggered averages, of the EMG and EEG activity until the occurrence of the first grating induced swimming event.

The multi-modal nature of the pre-stimulus EEG peak triggered average EMG activity in

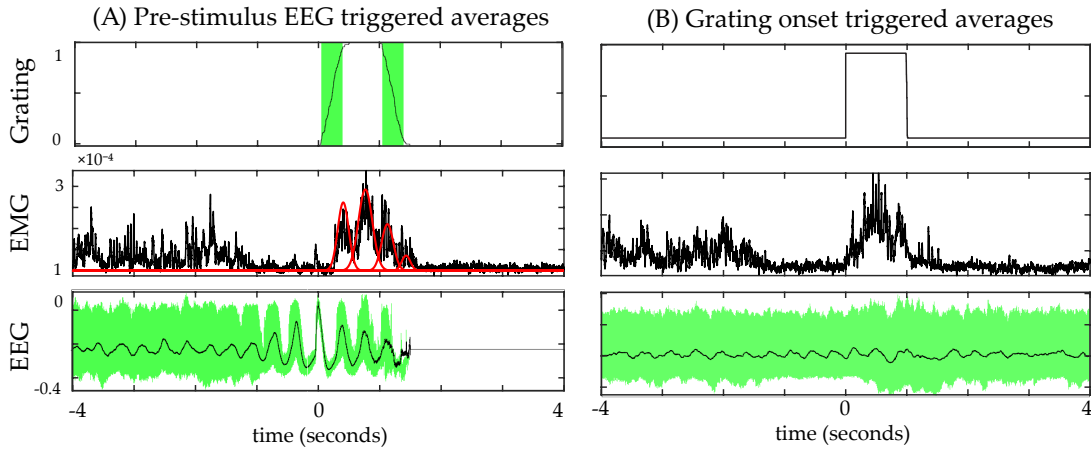


Figure 7.20: Comparison of (A) Pre-stimulus EEG peak triggered averages with (B) Grating onset triggered averages. The rows from top to bottom show the grating stimulus activity, EMG, and EEG activity respectively. The shaded green areas indicate the 95% empirical confidence levels. The results of fitting a Gaussian mixture model on the swim onsets are shown in red in panel A.

Figure 7.20-A illustrates that the timing of grating induced swimming is indeed governed by the ongoing EEG rhythmicity. To further validate this, we fit a Gaussian mixture model on the onsets of the grating induced swimming, as indicated in red in Figure 7.20-A. The peaks of the Gaussian fit indeed coincide with the peaks of the EEG, as evident when comparing the second and third rows of Figure 7.20-A. On the other hand, the EMG activity average triggered to the grating onset (Figure 7.20-B) is uni-modal with a larger variance, indicating that the timing of the swimming is not completely determined by the grating onset. While the swimming is indeed triggered by the grating, the timing of the swimming seems to be mainly determined by the ongoing breathing rhythm. Thus, these results suggest that the ongoing breathing rhythm acts as a predictor of the timing of stimulus-induced swimming.

7.5 Exploring breathing activity at the individual-cell level

In the previous sections, we looked in detail at the EEG-sync breathing oscillations that were observed in the Caudal-MO neuronal cluster and Lateral-MO neuronal cluster. However, it is also interesting to investigate the breathing oscillations at the individual-cell level, since it could be the case that there are minor variations among the individual cells in those neuronal clusters. Moreover, it is also possible that there could exist weaker breathing-related oscillations in other brain regions as well. In this section, we explore these possibilities by analyzing the fluorescence observations at the individual-cell level, as opposed to considering neuronal clusters. Furthermore, we compare the breathing networks under two different conditions: non-swim vs with-swimming, to further investigate the changes in the respiratory circuitry induced by swimming.

To that end, we first performed a coherence analysis to identify the cells with oscillations. We divided the observation duration into windows of length 4 s and then performed a sliding window multitaper analysis [43] to quantify the cross-spectral power between the fluorescence activity of each cell and the EEG. Then, we derived the average coherence [103] between the activity of each cell and the EEG, at peak EEG frequency, based on the cross power computations. Figure 7.21 (left) shows the average coherence thus obtained for each cell.

Next, we performed a statistical test to identify the cells with statistically significant coherence with the EEG. We obtained a null distribution corresponding to chance occurrences of coherence by temporally shuffling fluorescence activity of randomly selected cells. Then, based on that null distribution, we derived the coherence threshold that would reject the null hypothesis with $\alpha = 0.01$. Figure 7.21 (right) shows the cells that had statistically significant coherence

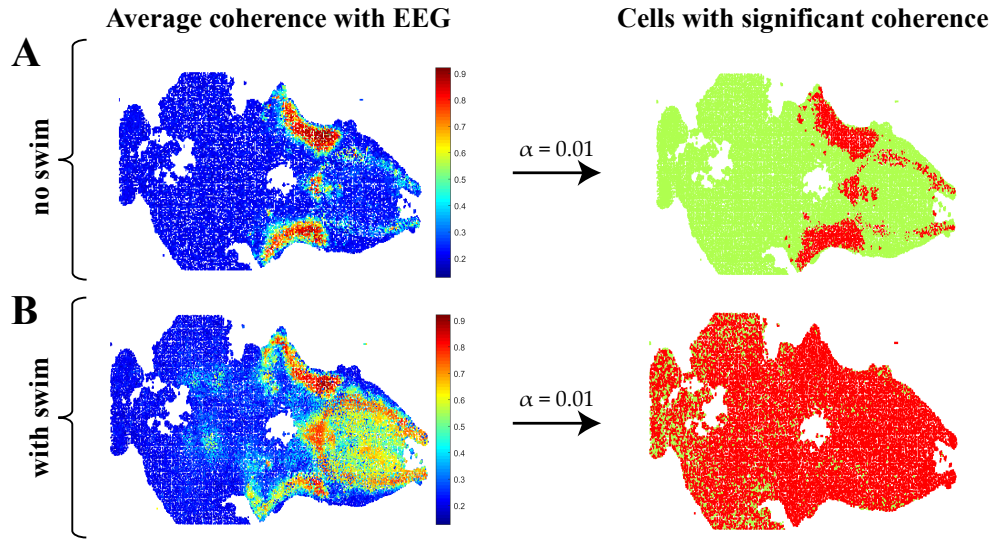


Figure 7.21: Extracting cells having significant coherence with EEG. (A) The left panel shows the average coherence of each cell with the EEG, for time windows without swimming. The right panel shows the cells that have high coherence with statistical significance ($\alpha = 0.01$) according to an empirical test, for time windows without swimming. (B) Same organization as panel A, for time windows with swimming.

according to this test ($\alpha = 0.01$). The top panels (Figure 7.21-A) show the coherence maps obtained using time windows without any fictive swim events, while the bottom panels (Figure 7.21-B) indicate the coherence maps derived from time windows with swimming.

From Figure 7.21 we see that the cells that have significant coherence under non-swim conditions are a subset of those with significant oscillations under swimming conditions. Next, we derived the magnitude spectrogram (of the CPSD wrt to EEG) in decibels and phase spectrogram (of the CPSD wrt to EEG) in radians of those significant cells at the peak EEG oscillation frequency, as shown in Figure 7.22. The magnitude spectrograms indeed look similar to the coherence maps, with higher cross-spectral power in the Lateral-MO and Caudal-MO areas. The phase spectrograms in Figure 7.22 reveal very interesting phenomena. From the phase spectrogram, we see that the regions around the Lateral-MO are in a similar phase and all the regions around the Caudal-MO are in a different phase. Moreover, from Figure 7.22-B we see that there

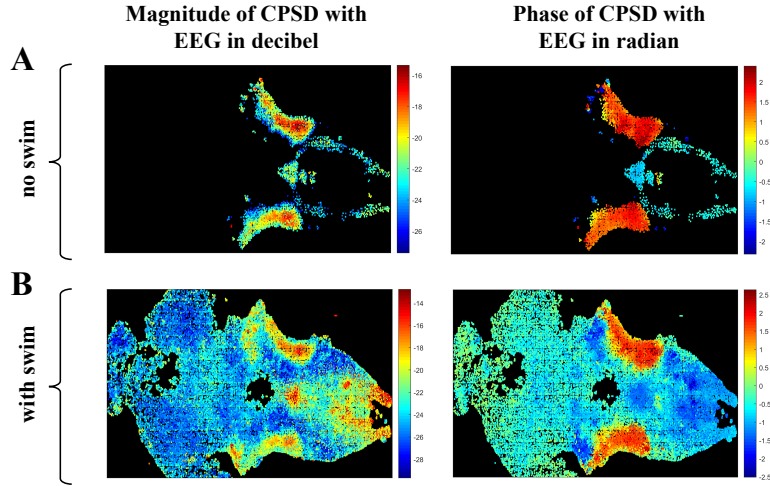


Figure 7.22: Magnitude and phase spectrograms of cells with significant coherence. (A) Average magnitude spectrogram in decibels (left) and average phase spectrogram in radians (right) of the Cross Power Spectral Density (CPSD) between the EEG and cells with significant coherence ($\alpha = 0.01$) at the peak EEG frequency, in time windows without swimming. (B) Same organization as panel A, for time windows with swimming.

exist cells in the mid-brain having a phase similar to the Caudal-MO region, under swimming conditions. These results strongly suggest that the respiratory rhythm is indeed distributed across a larger population of cells in the hind-brain, and is not only restricted to the two clusters Lateral-MO and Caudal-MO.

For a closer inspection of the phase spectrogram in Figure 7.22, we also looked into the contours of the phase spectrogram. Specifically, Figure 7.23 shows the different cells, within intervals of 0.4 rad in phase-shift relative to EEG. From Figure 7.23-A, we see that there are oscillations around the Lateral-MO and Caudal-MO with different phase shifts, even under the non-swim condition. Note that the negative phase shifts in the Caudal-MO arise due to the anti-polarity with the EEG, and do not imply that the Caudal-MO is lagging in time relative to EEG. In fact, in the previous sections, we provided several results that showed that the Caudal-MO precedes in time compared to all oscillations. Further, from Figure 7.23-B we see that the phase

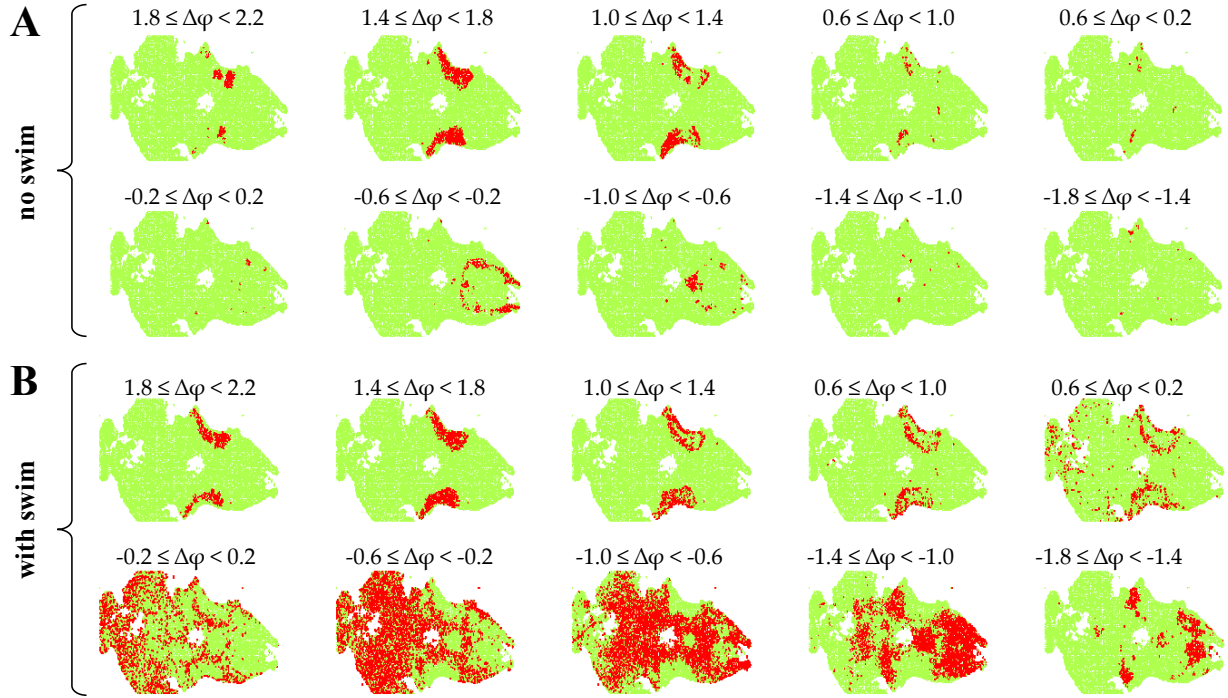


Figure 7.23: Phase sequence of cells with significant coherence. (A) Cells with different phase shifts ($\Delta\phi$ in radians) relative to the EEG, under non-swim conditions. The phase shift range in each panel is indicated above the panel. These plots correspond to different contours of the heat-map in Figure 7.22-A (right). (B) Same as panel A, for swimming conditions.

sequence of the breathing-related activity under swimming conditions is very similar to that under non-swim conditions, with a comparatively larger population of neurons being activated.

7.5.1 Reconstructing EEG through regression

To further validate that the EEG is indeed a surrogate of the oscillations in the above cells, we performed a regression analysis as follows. Firstly, we extracted the cells that have significant coherence with EEG under both non-swim and swimming conditions (Figure 7.21-A & -B), as shown in Figure 7.24. Since these cells are the ones that have the same frequency activity as EEG, the EEG is likely a surrogate of the oscillations in these cells. To test this hypothesis, we applied a regression on the fluorescence observations from those cells, to reconstruct the EEG

time series.

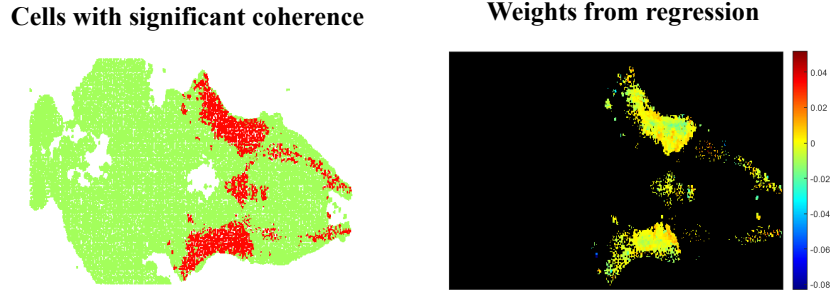


Figure 7.24: Cells selected for the regression analysis. The left panel shows the cells that have significant coherence with EEG under both non-swim and swimming conditions. The right panel shows the weight of each cell from the regression analysis.

Specifically, let $\mathbf{Y} \in \mathbb{R}^{T \times N}$ be a matrix of the fluorescence recordings from the cells in Figure 7.24 (left). Here N represents the total number of cells and T corresponds to the number of time samples observed. Then, suppose that the contribution of the n^{th} cell to the EEG is weighted by the weight w_n . Thus, $\mathbf{w} = [w_1, \dots, w_N] \in \mathbb{R}^N$ is a vector containing all the weights. Under this setting we performed a regression to find the optimal weights in the least-squares sense:

$$\mathbf{w}_{\text{opt}} = \underset{\mathbf{w}}{\operatorname{argmin}} \quad ||\text{EEG}_{\text{true}} - \mathbf{Y}\mathbf{w}||^2 + \lambda ||\mathbf{w}||^2, \quad (7.2)$$

where λ is a hyper-parameter and EEG_{true} is the observed EEG time series. We performed two-fold cross validation to find the optimal setting of λ . The optimal weights thus derived are shown in Figure 7.24 (right). Interestingly, we noted that the weights are mostly uniform, which implies that the oscillatory fluorescence activity in all these cells almost equally contribute to the EEG activity. Then, we derived the reconstructed EEG activity from the optimal weights and fluorescence observations:

$$\text{EEG}_{\text{recon}} = \mathbf{Y}\mathbf{w}_{\text{opt}}. \quad (7.3)$$

Figure 7.25 shows a comparison between the true EEG (EEG_{true}) and the reconstructed EEG from regression (EEG_{recon}). While the reconstructed EEG is somewhat noisier than the true EEG due to observation noise in fluorescence recordings, the reconstructed EEG follows the true observed EEG with remarkable accuracy.

To further check for the consistency of this observation, we performed a cross power spectral density analysis (CPSD) between the true EEG and reconstructed EEG, for the entire duration of the experiment. Figure 7.26 shows the magnitude spectrograms in decibel and phase spectrograms in radians corresponding to the CPSD analysis. The windows are ranked according to the increasing order of swim events. From the magnitude spectrograms, we see that the reconstructed EEG displays the same spectral content around the breathing frequency as the true EEG, consistently under both non-swim and swimming time windows. On the other hand, from the phase spectrograms we see that within the oscillation frequency band, the true EEG and the reconstructed EEG are indeed in-phase (with near-zero phase shift). Due to the presence of observation noise, the phases are different (in fact, random) at other frequencies.

In conclusion, the results of this analysis are key to understanding the potential source of EEG oscillations. This analysis shows that the EEG can be reconstructed with remarkable

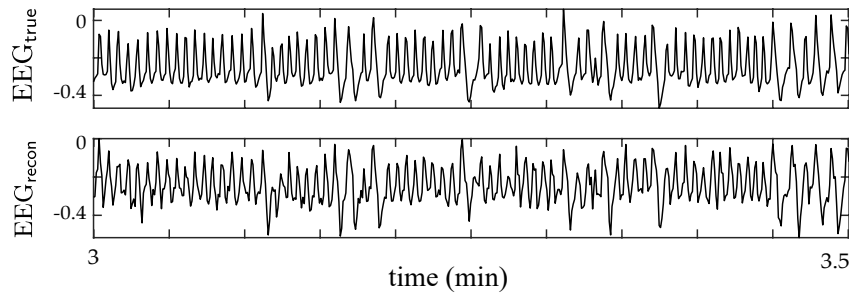


Figure 7.25: Comparison between the observed EEG and the EEG reconstructed from cells in Figure 7.24 using regression.

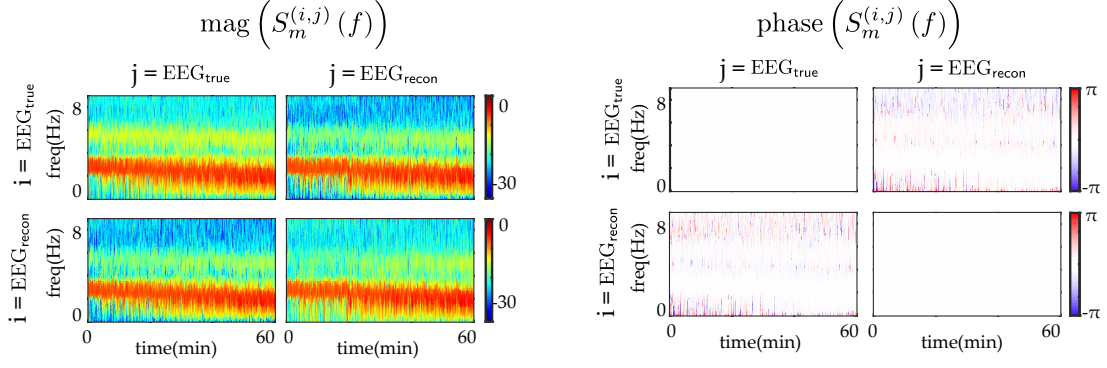


Figure 7.26: Cross power spectral density analysis between true EEG and reconstructed EEG. The magnitude spectrograms $\left(\text{mag}\left(S_m^{(i,j)}(f)\right)\right)$ in decibels are shown on the left and the phase spectrograms $\left(\text{phase}\left(S_m^{(i,j)}(f)\right)\right)$ in radians are shown on the right. In both cases $S_m^{(i,j)}(f)$ refers to the cross power spectrum between the i^{th} process and the j^{th} process in time window m and at frequency f . The windows are ranked from left to right in the order of increasing swimming events.

accuracy by combining the fluorescence activity recorded from several cells in the hindbrain. Thus, this indeed validates that the respiratory rhythm we capture through the EEG has a potential neural source, specifically the respiratory cells located in the hindbrain.

7.6 Optogenetic stimulation of the middle C-MO changes the breathing rhythm

In this section, we investigate the changes in breathing rhythmicity induced through optogenetic stimulation of different brain regions. Figure 7.27-A shows the four brain regions that were considered: middle C-MO, lateral C-MO, L-MO, and the cerebellum. The stimulation was one second long and each brain area was stimulated across multiple trials. Figure 7.27-B shows an example trial highlighting the key parameters of interest. Under the stimulation, we recorded the breathing pattern via EEG as shown in blue in Figure 7.27-B. Then, to analyze the effect of stimulation on the breathing rhythm, we extracted the following quantities. For each trial, we extracted the EEG peak-to-peak distance before stimulation (t^{pre}), during stimulation (t^{post1}) and

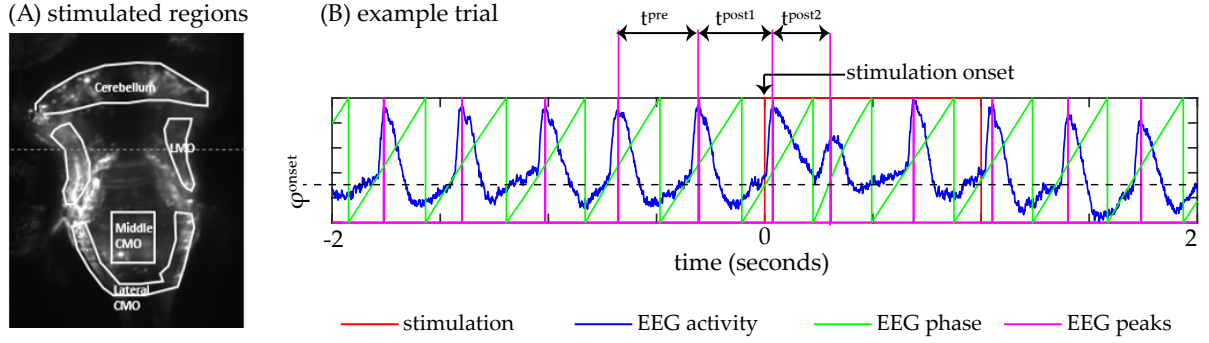


Figure 7.27: (A) The optogenetically stimulated brain regions. (B) An example trial outlining the optogenetic stimulation (red), EEG activity (blue), EEG phase (green) and the EEG peaks (magenta). The peak-to-peak distance before stimulation (t^{pre}), during stimulation (t^{post1}) and after stimulation (t^{post2}), and the phase of EEG at the stimulation onset (ϕ^{onset}) are also indicated.

after stimulation (t^{post2}) as shown for the example in Figure 7.27-B. Further, we derived the instantaneous phase of EEG using the Hilbert transform (as shown in green in Figure 7.27-B), and extracted the phase of EEG at the stimulation onset (ϕ^{onset}). Then, to identify effects on breathing rhythmicity induced by the stimulation, we derived the two quantities: $\Delta t^1 = t^{\text{post1}} - t^{\text{pre}}$ and $\Delta t^2 = t^{\text{post2}} - t^{\text{pre}}$. If the stimulation does not have a uniform effect on the breathing rhythmicity, one would expect Δt^1 and Δt^2 to be randomly distributed near zero. Moreover, to investigate whether the response could depend on the phase of EEG at stimulation onset, we plotted Δt^1 and Δt^2 in each trial as a function of ϕ^{onset} , as illustrated in Figure 7.28.

From Figure 7.28, we see that the stimulation of lateral C-MO, L-MO and cerebellum do not have notable effects on the breathing rhythm, since both Δt^1 and Δt^2 seem to be randomly distributed across multiple trials. However, there does exist a notable trend in the change in rhythmicity following middle C-MO stimulation (first column of Figure 7.28). During the middle C-MO stimulation, while Δt^1 is near zero and random when the phase of EEG at the stimulation onset is negative ($\phi^{\text{onset}} < 0$), we clearly see from Figure 7.28 that $\Delta t^1 < 0$ when $\phi^{\text{onset}} > 0$. On

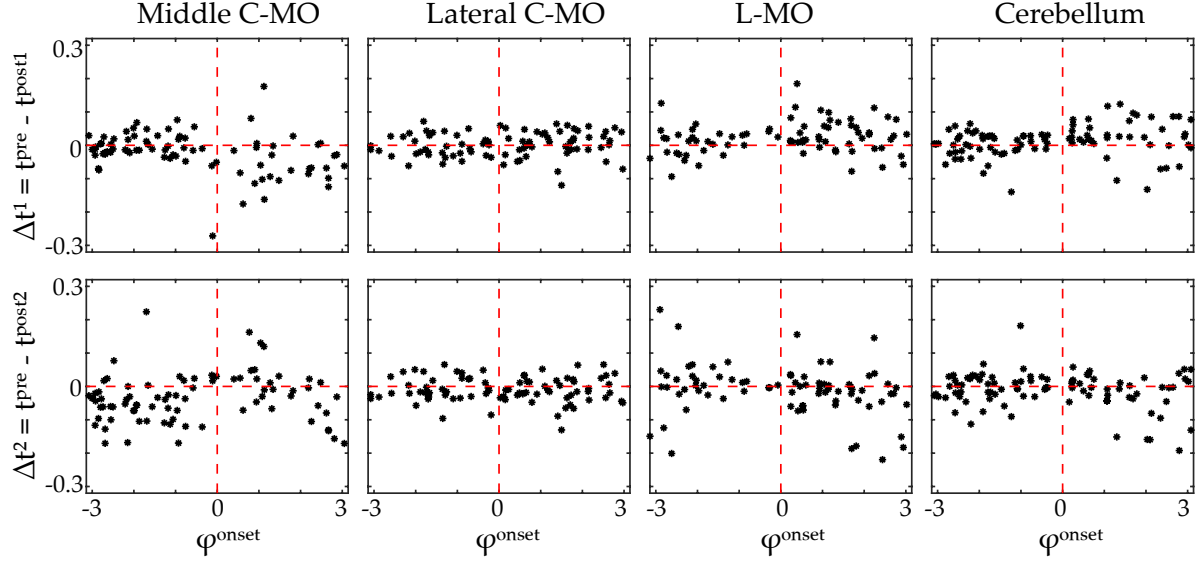


Figure 7.28: Results of optogenetically stimulating different brain areas. Each panel indicates time shifts $\Delta t^1 = t^{\text{post1}} - t^{\text{pre}}$ (top) and $\Delta t^2 = t^{\text{post2}} - t^{\text{pre}}$ (bottom) as a function of the phase of EEG at the stimulation onset (ϕ^{onset}). Columns from left to right show the results of stimulating the brain regions: middle C-MO, lateral C-MO, L-MO and the cerebellum respectively.

the other hand, while $\Delta t^2 < 0$ when $\phi^{\text{onset}} < 0$, Δt^2 is random when $\phi^{\text{onset}} > 0$.

This result can be understood as follows. When the phase of EEG at the onset of the stimulation of middle C-MO is positive, the change in rhythmicity induced by the stimulation is immediate; the EEG peak that immediately follows the stimulation occurs sooner than expected. On the other hand, when the phase of EEG at the onset of the stimulation is negative, the response to the stimulation is not immediate, it is not the first EEG peak that occurs sooner than expected, but the second. In conclusion, the changes in breathing rhythmicity induced by middle C-MO optogenetic stimulation depend on the phase of breathing at the onset of the stimulation. Moreover, the fact that the stimulation of middle C-MO causes the timing of the following breaths (and thus the breathing rhythmicity) to be changed indeed further suggests that the breathing pacemaker could potentially be located in that region.

7.7 Ablation of Lateral-MO does not affect phase coupling between swimming and breathing

Next, to investigate the role of the Lateral-MO in breathing and swimming, we performed ablation of L-MO cells and compared the breathing and swimming activities before and after the L-MO ablations. Consistently in all experiments, it was noted that the EEG is largely unaffected by the ablations, which implies that the other breathing cells can continue functioning independently of the L-MO. Further, it was noted that both spontaneous and stimulus-driven swimming increases significantly after L-MO ablations, which suggests that L-MO could have an inhibitory effect on swimming.

To investigate the effect of L-MO ablations on the oscillations in other brain regions, we performed whole-brain imaging of activity before and after L-MO ablations. Then, we extracted three main groups (Lateral-MO, Caudal-MO, Group-7) related to breathing using data recorded before ablations, as shown in Figure 7.29-A. We compared the average fluorescence activity of these three groups along with the EEG and EMG, before and after L-MO ablations. Figure 7.29-B & C show two such example trials, recorded under a grating visual stimulus. From Figure 7.29-B, we see that L-MO is anti-phasic relative to C-MO and G-7, and all three brain regions have breathing and swimming-related oscillations before L-MO ablation. On the other hand, after L-MO ablation (Figure 7.29-C), while the L-MO activity is random, the C-MO and G-7 neurons indeed have phase-locked oscillations in sync with EEG and EMG same as before.

To further investigate this, we derived the instantaneous phase of the L-MO, C-MO, G-7, and EEG, and computed the average swimming activity triggered to the phase of these oscilla-

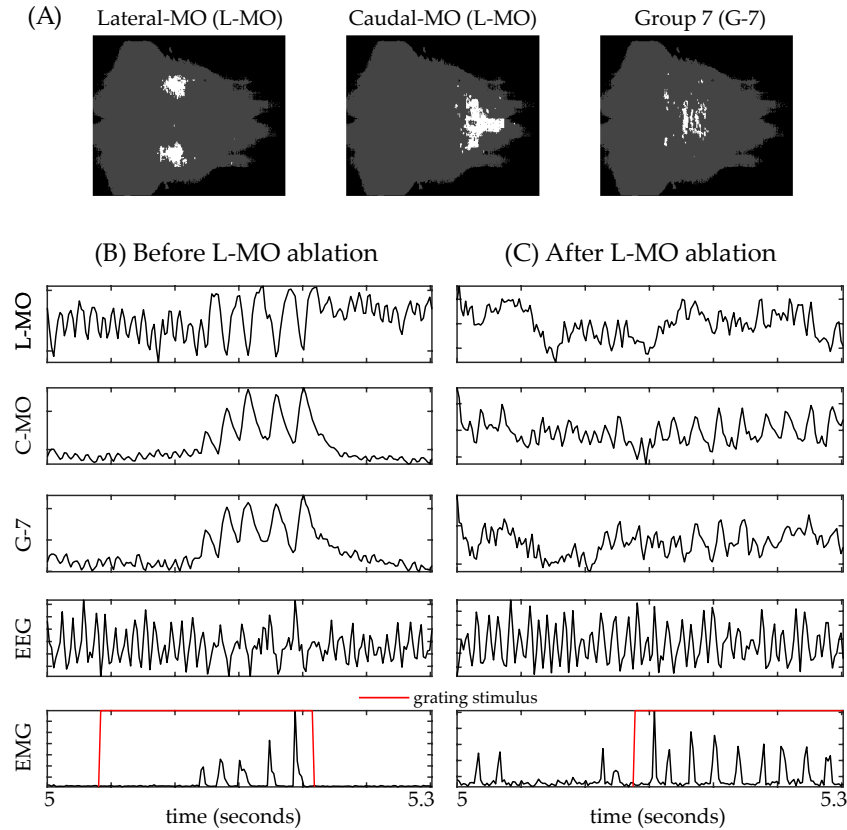


Figure 7.29: Ablation of Lateral-MO experiment. (A) Groups of cells with breathing activity under normal conditions: Lateral-MO (L-MO), Caudal-MO (C-MO), and Group-7 (G7) neurons. (B) Sample average fluorescence activity recorded from L-MO, C-MO, G-7, and the EEG and EMG under presentation of a grating visual stimulus, before L-MO ablations. (C) Same organization as panel B, for activity recorded after L-MO ablations.

tions, from data before and after L-MO ablations. Figure 7.30 shows the phase triggered averages of swimming thus obtained. As indeed one would expect, the swimming activity triggered in the L-MO phase is random after the L-MO ablations, while it is not the case before the ablation. More interestingly, the average swimming activity triggered by C-MO, G-7, and EEG phase seems to be largely unaffected by the L-MO ablations. These results suggest that swimming is phase-locked to the oscillations in C-MO and G-7 cells independent of the L-MO. Further, the breathing rhythm in C-MO and G-7 persists even after L-MO ablations, which further reinforces our hypothesis that the breathing pacemaker is located in this area. However, increased

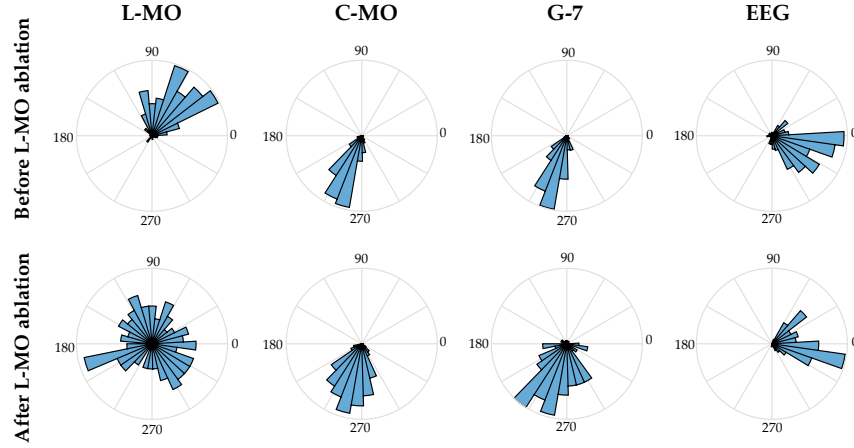


Figure 7.30: Average swimming activity triggered to the instantaneous phase of L-MO, C-MO, G-7 and EEG oscillations, before (top) and after (bottom) L-MO ablations.

swimming after L-MO ablations suggests that the L-MO plays a role in regulating swimming, by inhibiting swimming.

7.8 Proposed circuit model for respiratory-swimming coupling

Finally, we develop a circuit model for zebrafish respiratory-swimming coupling based on the observations and deductions of all the preceding sections. We focus on the two main respiratory brain regions (C-MO and L-MO) and create a model that can replicate their interactions with swimming. To that end, first, let us summarize the major conclusions that we deduced through all the analyses in the preceding sections.

- The C-MO is in-phase with swimming, and the fluorescence activity of C-MO is large during swimming.
- The breathing rhythm slows during swimming, and swimming is phase-locked to breathing.
- The L-MO activity is anti-phasic relative to C-MO activity during both swimming and

non-swimming conditions and is slightly delayed relative to C-MO.

- Swimming is facilitated by the ablation of L-MO, which suggests that an inhibitory connection exists from L-MO to swimming.
- The breathing frequency slows down during high oxygen conditions, and the fish can swim more easily in highly oxygenated water.

Based on the foregoing observations, we deduced the circuit model in Figure 7.31. The C-MO models the combination of the breathing pacemaker and other brain regions having activity in-phase with swimming. Since L-MO is anti-phasic relative to the C-MO, we hypothesize that L-MO activity is driven by a post-inhibitory rebound from C-MO. Then, since swimming is facilitated by L-MO ablation, swimming is hypothesized to be inhibited by the L-MO. On the other hand, since swimming is in-phase with C-MO and C-MO activity is large during swimming, we hypothesize an excitatory link between C-MO and swimming, with the swim drive being primarily modulated by the C-MO. Finally, since the breathing rhythm changes with oxygen conditions, we assume that an oxygen level sensor modulates the rhythm of the breathing pacemaker.

Next, we simulated the circuit model in Figure 7.31 using the Izhikevich model for spiking neurons [81]. The Izhikevich model is a two dimensional system of ordinary differential

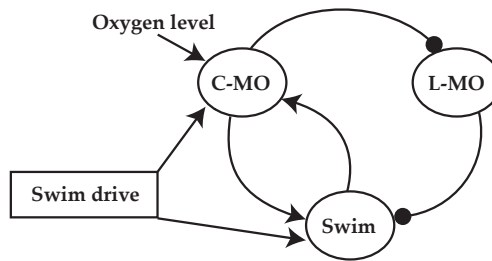


Figure 7.31: Proposed circuit model for respiratory-swimming coupling.

equations of the form,

$$\begin{aligned}\dot{v} &= 0.04v^2 + 5v + 140 - u + I \\ \dot{u} &= a(bv - u)\end{aligned}\tag{7.4}$$

with the after-spike resetting

$$\text{if } v > 30 \text{ mV, then } v \leftarrow c \text{ and } u \leftarrow u + d.$$

By changing the settings of the four parameters a, b, c and d , it is possible to simulate various intrinsic firing patterns observed in real neurons [81]. Considering that respiratory oscillations are generally known to be bursting events, we set the parameters such that the three populations: C-MO, L-MO and swim generate bursting events. Since the C-MO consists of the pacemaker and can generate the rhythm intrinsically, we modeled the C-MO as a tonic bursting unit, with the baseline tonic input level being determined by the oxygen level. Thus, the C-MO generates intrinsic bursting oscillations (respiratory rhythm) while modulating the frequency of oscillations with the oxygen level. A swim drive causes an instantaneous increase in C-MO input, which causes the C-MO bursts to be stronger (and hence larger calcium activity), which in turn triggers swimming.

To model the post inhibitory rebound effect in L-MO induced by C-MO, we used an additional linear differential equation:

$$\dot{r} = -pr + q\mathbb{I}[\text{C-MO} > 0]$$

and then modeled the input to the L-MO as

$$I_{L-MO} = r \times \mathbb{I}[C-MO < 0]$$

This causes the L-MO to be silent when the C-MO activity is high, and then to generate a bursting event after being released from the C-MO inhibition. Finally, we modeled the swimming activity as being generated by another bursting unit that is primarily driven by C-MO, along with an inhibitory input from L-MO.

Figure 7.32-A shows a sample case with all the variables, where we have simulated spiking activity with 1 ms resolution. Figure 7.32-B shows the corresponding low temporal resolution (sampling frequency 20 Hz) and temporally blurred calcium activity, similar to the calcium fluorescence activity observed in real data. In the example in Figure 7.32-A, the first, third, and fourth swim drives induce swimming while the second one fails to do so. This is because the second drive is not as strong as the other three, and the corresponding C-MO activity is not sufficiently strong enough to induce swimming activity by overcoming L-MO inhibition. It is evident from this simulation that the model replicates all the key properties that we observed in real data.

Next, to investigate the changes in the model with oxygen level and L-MO ablations, we extended our simulation to those conditions. Figure 7.33 shows simulated calcium activity under two different oxygen levels. To further make the setting realistic, we have modeled both spontaneous and stimulus-induced swim drives in this simulation. The spontaneous swim drives were modeled by a smoothed Bernoulli noise process (shown in blue in Figure 7.33 first row). The oscillation frequency of C-MO and L-MO is lower during high oxygen conditions, similar to real data. Due to the inhibition from L-MO to swim, weak swim drives cannot induce swimming

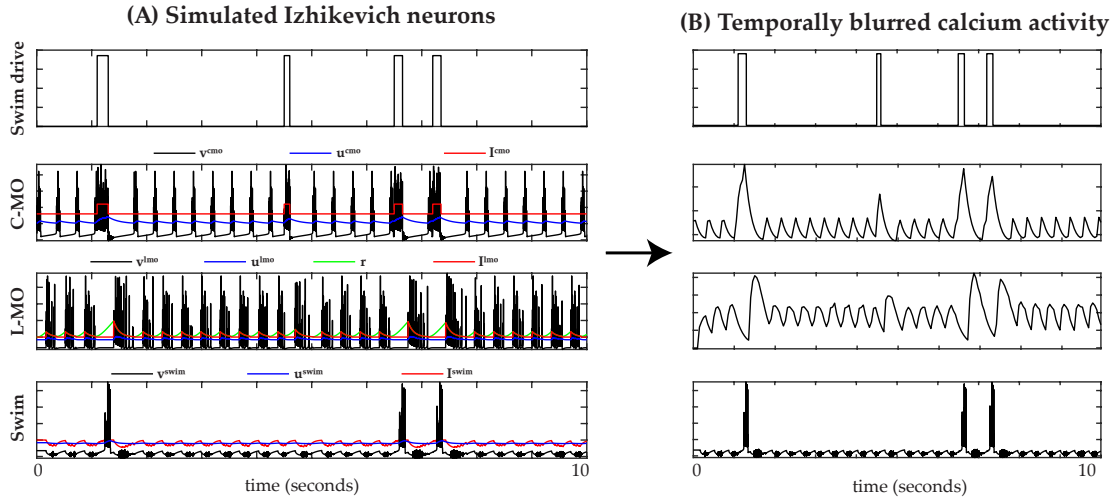


Figure 7.32: Sample simulated Izhikevich neuronal model activity. (A) Simulated Izhikevich neuronal spiking activity. Rows from top to bottom: swim drive, C-MO, L-MO, and swimming. Each panel also shows the variables and the inputs of the corresponding Izhikevich model. (B) The calcium activity with low temporal resolution and temporal blurring, in the same organization as panel A.

under the low oxygen condition in Figure 7.33. On the other hand, due to lower frequency oscillations under high oxygen conditions, swimming is mostly released from the L-MO inhibition by the time the excitation from C-MO is received, which facilitates swimming. Thus, even weak swim drives can induce swimming under high oxygen conditions, as shown in Figure 7.33. This explains the potential neural mechanism of how zebrafish adjust their swimming rates based on the oxygen availability in the environment.

Then, to test the role of L-MO in regulating swimming, we simulated the same setting as in Figure 7.33 by removing the inhibition from L-MO to swimming (which is equivalent to ablating the L-MO cells). Figure 7.34 shows a comparison of the swimming generated by our model, before and after L-MO ablations. After L-MO ablations, we see more spontaneous and stimulus-induced swimming even under low oxygen conditions, as a result of lack of inhibition from L-MO to swimming. From this, we see that the role of L-MO is to regulate swimming,

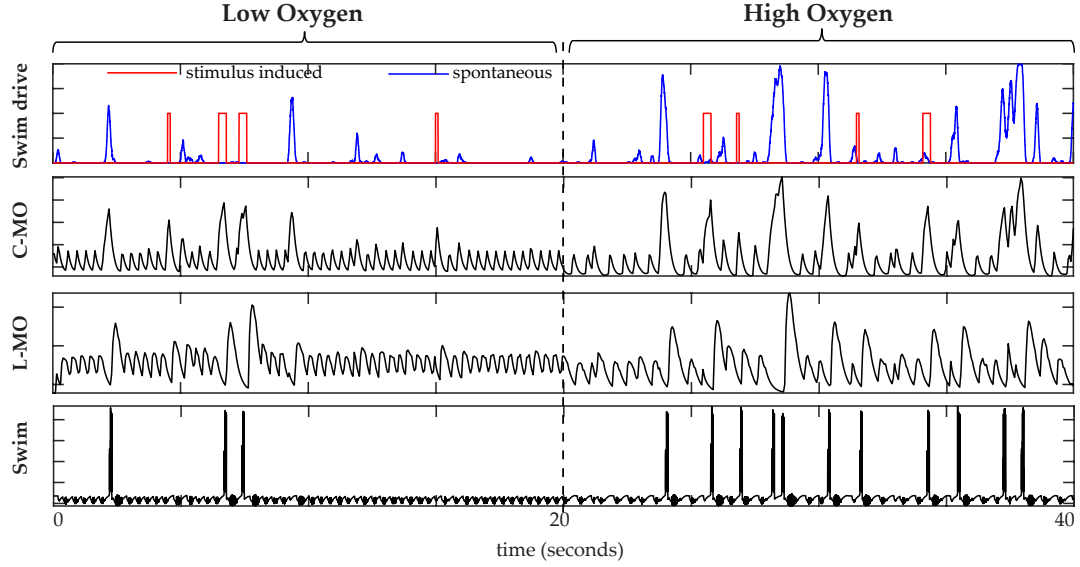


Figure 7.33: Simulated activity under different oxygen conditions. Rows from top to bottom: swim drives (spontaneous activity in blue and stimulus-induced activity in red), C-MO calcium activity, L-MO calcium activity, and swimming. The activity from 0 – 20 seconds replicates a low oxygen setting, while the activity from 20 – 40 seconds simulates a higher oxygen setting.

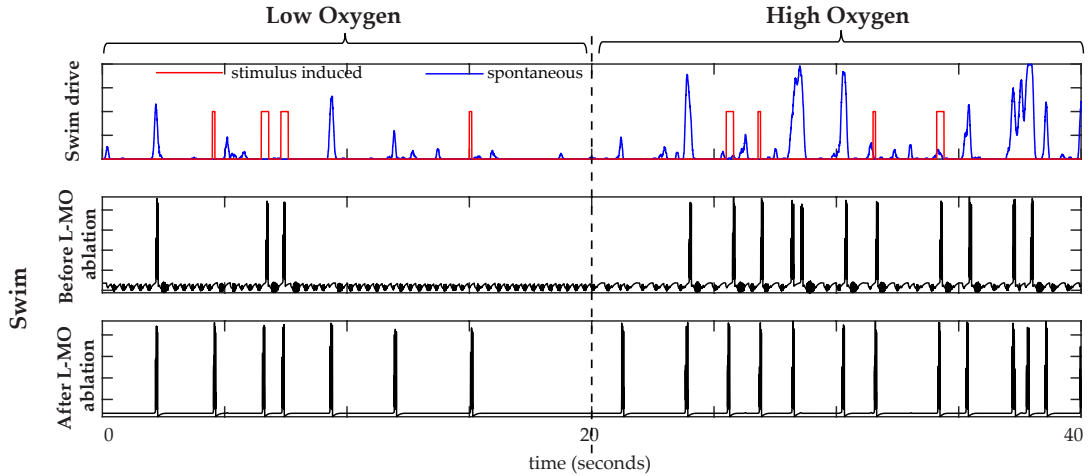


Figure 7.34: Comparing swimming generated by our model at the same settings as Figure 7.33, before and after L-MO ablations.

by suppressing swimming through inhibition during low-oxygen conditions. On the other hand, since swimming is primarily driven by the C-MO, the lack of inhibition from L-MO to swimming does not have a significant impact on the C-MO-swimming phase coupling, consistent with what we also observe in real data experiments.

Finally, we repeated a stimulus under 100 trials (with different spontaneous activity in each trial), under low and high oxygen conditions, both before and after L-MO ablations as shown in Figure 7.35. The second row of Figure 7.35 shows the swimming activity in each trial, and the third row shows the average swimming activity across all trials. These results indeed corroborate our initial observations: fish tend to swim more easily under high oxygen conditions than under hypoxia, and swimming is facilitated by ablation of L-MO irrespective of the oxygen level. In conclusion, the circuit model that we developed accurately reproduces the coupling between C-MO, L-MO, and swimming observed in real data, along with the changes induced in the network under different oxygen conditions.

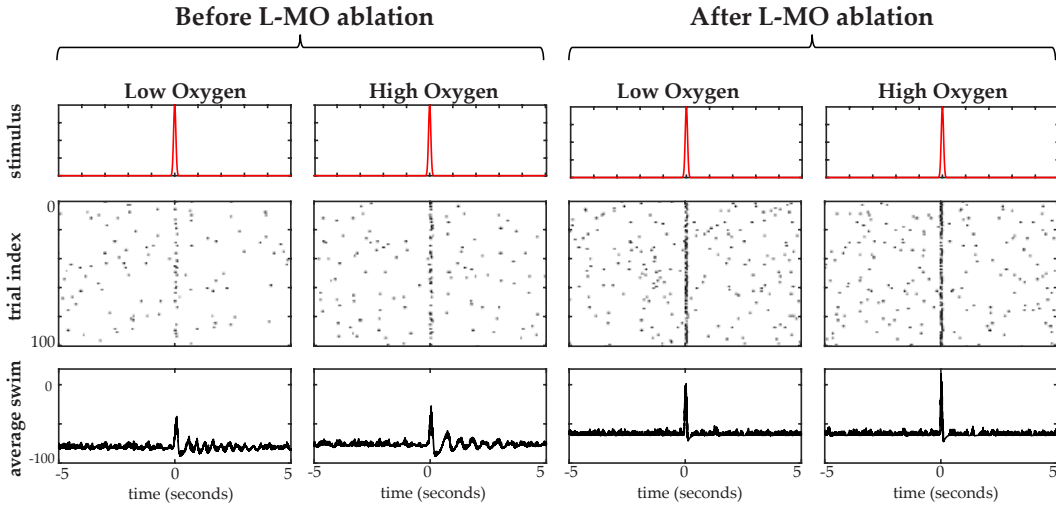


Figure 7.35: Comparing swimming generated by our model across multiple trials, before and after L-MO ablations, and under low and high oxygen conditions. The first row indicates the stimulus repeated across trials, and the second row shows the trial-to-trial swimming activity. The third row indicates the average swimming activity across all 100 trials.

7.9 Concluding Remarks

In this chapter, we investigated the respiratory circuitry and the swimming-respiration coupling phenomena in larval zebrafish by analyzing whole-brain light-sheet microscopy imaging data. Firstly, using multitaper spectrotemporal analysis of whole-brain imaging data and simultaneously recorded EEG data, we discovered brain regions that are synchronized with the EEG oscillations. Then, using cross-spectral and spectral Granger causality analysis, we established that the oscillations extracted from those brain regions and the oscillations in the EEG are indeed a single rhythm, which corresponded to the zebrafish respiratory rhythm. Using cross-spectral density phase analysis, we then introduced a new method to categorize the whole-brain respiratory neurons based on their distinct phases relative to the breathing rhythm. From this analysis, we deduced that the neural source of EEG is primarily located in the hind-brain and that a larger population of neurons spanning from the fore-brain to the hind-brain is in sync with breathing under swimming conditions in comparison to non-swim conditions. Using several spectrotemporal analysis methods, we also demonstrated that zebrafish swimming is phase-locked to breathing and is gated by the respiratory rhythm. Next, through the analysis of whole-brain imaging data under optogenetic stimulation and two-photon ablations, we identified the brain regions that are key for this swimming-respiration coupling. Finally, using the Izhikevich model for spiking neurons, we developed a circuit model that replicates the swimming-respiration coupling behavior observed in real data, providing new insights into the possible underlying neural circuitry.

While the results presented in this chapter are indeed very promising, several deductions may require further investigation to make our conclusions more concrete. Firstly, from our results, we hypothesized that the breathing pacemaker lies in the Caudal-MO area, based on mul-

multiple observations. While the observations indeed strongly suggest that, we do not yet have sufficient experimental evidence to distinguish the exact pacemaker neurons. Then, we introduced a minimal circuit model in Section 7.8 that could potentially explain the swimming-respiratory coupling behavior we noted in real data. It is noteworthy that we have not yet established the connections that we hypothesized at a synaptic level, and the true underlying circuitry is quite likely to be significantly more complicated than our proposed model. In addition, while we developed our model at a neuronal cluster level, our phase analysis at an individual-cell level in Section 7.5 revealed that there does exist notable variability in the respiratory rhythm at the scale of individual neurons. Nevertheless, our model is capable of reproducing all the main observations we have noted consistently across different datasets and experiments, in particular the effects related to swimming-respiration coupling. Future optogenetic stimulation and ablation experiments performed targeting individual neurons (as opposed to regions/clusters) could potentially aid in identifying the exact key respiratory-swimming neurons and their underlying connections more specifically and accurately.

Another interesting future direction would be investigating similarities and differences between the zebrafish respiratory system and the well-known mammalian three-phase respiratory rhythm/circuit [165–167]. The anti-phasic nature of the Caudal-MO and Lateral-MO suggests that one of them could correspond to inspiration activity while the other corresponds to expiration activity. It is possible that the different phases in oscillations we observed in Section 7.5 could correspond to different phases of the respiratory cycle. Since we believe that the Caudal-MO is the potential pacemaker area, it is very likely to correspond with the pre-Botzinger pacemaker neurons, that are known to initiate the inspiration phase. Likewise, it seems as if the Lateral-MO has functionality similar to the Post-Inspiratory Complex and the Botzinger neurons, regulating

expiration activity. Investigating if there indeed exists a mapping between the zebrafish respiratory circuit and the mammalian respiratory system, and identifying their common components is a potential extension of this work.

Finally, it is well known that in humans and other mammals, respiration is often coupled with locomotion; particularly during sustained, rhythmical activities such as cycling, rowing, running, and wheelchair propulsion [168–170]. This phenomenon is known as Locomotion-Respiratory coupling (LRC), and has been hypothesized to occur as a byproduct of optimizing the effective oxygen volume due to the usage of common respiratory and locomotion muscles [168, 170]. The swimming-breathing coupling in zebrafish that we observed in Section 7.3 and 7.4 too seems to be a very similar phenomenon, where we came to several deductions such as zebrafish swimming is phase-locked to breathing, the breathing frequency slows-down during swimming, and the timing of stimulus-induced swimming depends on the pre-stimulus breathing activity. Investigating the similarities and differences between the mammalian locomotion-respiratory coupling and zebrafish swimming-breathing coupling side-by-side could potentially help us in gaining a better understanding of both of these phenomena.

Chapter 8: Concluding Remarks and Future Directions

8.1 Summary and Extensions of our Contributions

In this dissertation, we introduced new models and methods to infer spectral and temporal network organizations from high dimensional neural data, specifically neuronal spiking data, two-photon calcium imaging data, and light-sheet imaging data. We identified that the existing methods for network-level inference from these data modalities have several shortcomings; including undermining the non-linear dynamics that govern neuronal spiking activity, requiring intermediate time-domain estimation, ignoring non-stationary brain activity, and causing downstream network-level analysis to be erroneous by solving the inverse problem in a multi-stage fashion. We addressed these issues by introducing Bayesian models and methods to *directly* infer the underlying network connectivity from the observed data, without requiring intermediate steps of latent activity estimation.

In the first part of this thesis, we introduced Bayesian methods to infer dynamic frequency-domain network organizations in neuronal ensembles from spiking observations, by integrating techniques such as point process modeling, state-space estimation, and multitaper spectral estimation. Firstly, we introduced a semi-stationary multi-taper method tailored for spiking data, called the Point Process Multi-Taper Semi-Stationary Spectral Density (PPMT-SSD) estimator, and established theoretical bounds on its performance. Building upon PPMT-SSD, we then in-

troduced a framework to derive dynamic spectrotemporal Granger causal interactions in a population of neurons from spiking data. We demonstrated the validity of these methods through simulations, and application to real data recorded from the rat brain revealed new insights on the coupling of different neuronal populations across frequency during different sleep episodes. Finally, we also extended our frequency-domain estimation methods for a completely different application: developing a precise inference method to characterize frequency-domain components of human heart rate variability from electrocardiogram data.

In the second part, we introduced a methodology to explicitly model and directly estimate signal and noise correlations from two-photon calcium imaging observations. Despite its remarkable spatial throughput, two-photon imaging only provides temporally blurred and noisy observations of spiking activity. Even if spiking data were perfectly recovered via deconvolution techniques, inferring the underlying neuronal correlations is challenging due to the non-linear dependence of neuronal spiking on endogenous and exogenous inputs. We addressed these challenges by posing the problem in a Bayesian framework and derived an efficient variational inference method. We demonstrated the validity of the resulting estimators through theoretical analysis and extensive simulations, all of which established significant gains over existing methods. Applications of our method on real data revealed novel and distinct spatial patterns in the correlation structures of two different layers in the mouse primary auditory cortex. Then, we used our methods to investigate how the spatial patterns of the correlation networks in the primary auditory cortex change under pure-tone vs two-tone stimulus conditions. Next, we extended our estimation framework to infer signal and noise correlation from electrophysiological data. By applying our method to electrophysiological data recorded from the auditory cortex of gerbils under auditory task performance, we illustrated how our method provides new insights

into network-level changes during perceptual learning.

Lastly, in the third part of this dissertation, we investigated the respiratory circuitry and the swimming-respiration coupling phenomena in larval zebrafish by applying several spectrotemporal analysis techniques, to whole-brain light-sheet microscopy imaging data. Firstly, using multitaper spectrotemporal analysis of whole-brain imaging data and simultaneously recorded electrophysiological data, we discovered brain regions that are synchronized with the respiratory rhythm and categorized them based on their distinct phases relative to the breathing rhythm. Using several spectrotemporal analysis methods, we then demonstrated that zebrafish swimming is phase-locked to breathing and is gated by the respiratory rhythm. Next, through the analysis of whole-brain imaging data under optogenetic stimulations and two-photon ablations, we identified the brain regions that are key for this swimming-respiration coupling. Finally, using the Izhikevich model for spiking neurons, we developed a circuit model that replicates the swimming-respiration coupling behavior observed in real data, providing new insights into the possible underlying neural circuitry.

8.2 Limitations of our approaches

It is noteworthy that while we indeed addressed most of the shortcomings in existing methods, there are some limitations inherent to our approaches as well. In what follows, we outline some important aspects to keep in mind when using our methods and interpreting our results.

1. *Computational Complexity*: While the approaches we introduced in Part I and Part II have significant performance gains in comparison to existing methods, it is important to note that we achieved this at the cost of higher computational complexity. Nonetheless, we

devised our inverse solutions so that computational complexity was tractable, and also used prior knowledge and data pre-processing to further reduce computational complexity; for example, in the Chapter 2 we restricted the estimated frequency range based on prior knowledge of the typical frequencies of brain oscillations. It is also noteworthy that all analyses presented in this dissertation were performed using a simple Intel Core i7 CPU, thus our implementations also do not require expensive computational resources despite having greater computational complexity in comparison to existing naive approaches.

2. *Modeling Assumptions*: Since our approaches are based on assuming an explicit generative model, it is important that the experimental data meaningfully follows our modeling assumptions. For example, in Chapter 2 and 3, we assume that the spiking activities are binary, with at most one spiking event per time bin. Thus, it is not recommended to use these methods under bursting and other rapidly firing conditions. Next, in Chapter 5 we noted that in the regime of extremely low spiking rate and high observation noise, our proposed method also fails to capture the underlying correlations faithfully and its performance degrades to those of existing methods. Thus, it is worth performing an initial inspection of whether the data to be analyzed indeed saliently falls under our modeling assumption, before applying our methods.
3. *Approximate inverse solutions*: Due to the multi-tier nature of our statistical models with the involvement of multiple latent variables and unknown parameters, the exact maximum likelihood solutions are intractable. Thus, we used alternative methods such as expectation-maximization (Part I) and variational inference (Part II) to solve the inverse problems. It is noteworthy that both these methods only guarantee convergence to a local optimum,

and thus are sensitive to initialization. We used extensive simulations to validate that our solutions indeed converged to the global optima, and also initialized the estimates using outputs of simpler models to ensure convergence to the desired solution. However, it is also important to keep in mind that a comprehensive theoretical understanding of initialization methods is lacking, and there is always a possibility of converging to local optima.

4. *Biological Interpretation:* Finally, it is important to keep in mind that the functional networks identified by our methods must not be readily interpreted as direct or synaptic connections. We derived all network organizations in this dissertation purely based on statistical relationships between neural signals over time. For example, the circuit model we propose for zebrafish respiration-swimming coordination in Part III was entirely based on empirical observation, with no reference to anatomical connectivity. Whether these network patterns also exist at the anatomical synaptic level is yet to be tested. An alternative and inclusive view is that these network organizations reflect a measure of the information transferred between neurons in the brain, rather than their underlying synaptic connections.

8.3 Future Directions

In closing, we outline some future directions of research that are brought to light as an additional outcome of this dissertation work:

1. Improving the temporal resolution of our frequency-domain network analysis methods by relaxing the semi-stationarity assumption, and integrating more general notions of non-stationarity such as Priestley’s evolutionary spectra [45] or Weyl spectra [94]
2. Introducing frameworks to infer dynamic frequency-domain network organizations directly

from two-photon calcium imaging observations, by combining the methods introduced in Part I and Part II of this dissertation.

3. Incorporating our direct inference framework of latent network organizations from two-photon calcium imaging data to network-level analysis beyond signal and noise correlations: such as connected correlation analysis [142], Granger causal network analysis [171], and synchrony analysis [172]
4. Extending our baseline signal and noise correlation inference framework to time-varying correlations that underlie rapid task-dependent dynamics [146], and to non-linear models such as those parameterized by neural networks [124]
5. Improving our signal and noise correlation inference framework by leveraging variational inference beyond the mean-field regime [145]
6. Investigating our proposed zebrafish swimming-respiration coupling circuit model at the individual-cell level and the anatomical synaptic level [173]
7. Comparing the zebrafish respiratory neural circuitry with the well-known mammalian three-phase respiratory rhythm/circuit [165–167]
8. Comparing the zebrafish swimming-respiration coupling with the mammalian locomotion-respiratory coupling phenomena [168–170]

Finally, note that the challenges we addressed through this dissertation are shared among many other disciplines such as econometrics, climatology, seismology, social network analysis, gene regulatory networks, astronomical imaging, and communication systems. Thus, the developed ideas and algorithms in this thesis have potential applications in other domains outside

neuroscience and can be readily adopted thanks to the plug-and-play nature of our algorithms. In the spirit of reproducibility, we have archived implementations of our algorithms as open source repositories on <https://github.com/Anuththara-Rupasinghe>.

Chapter A: Supplementary Material on Chapter 2

This appendix contains the proofs of the theorems characterizing the bias and variance of the PPMT-SSD estimator.

A.1 Proof of Theorem 2.1

Proof. Let $S(\omega)$ be the PSD of the process $\{x_k\}_{k=1}^K$. Then,

$$\begin{aligned} |\text{bias}(\hat{S}^{\text{mt}}(\omega))| &:= |\mathbb{E}[\hat{S}^{\text{mt}}(\omega)] - S(\omega)| \\ &\stackrel{(a)}{\leq} |\mathbb{E}[\hat{S}^{\text{mt}}(\omega) - S^{\text{mt}}(\omega)]| + |\text{bias}(S^{\text{mt}}(\omega))|, \end{aligned} \tag{A.1}$$

where (a) follows from the triangle inequality. Further,

$$\begin{aligned} \text{Var}(\hat{S}^{\text{mt}}(\omega)) &:= \mathbb{E}[|\hat{S}^{\text{mt}}(\omega) - \mathbb{E}[\hat{S}^{\text{mt}}(\omega)]|^2] \\ &\leq \mathbb{E}[|\hat{S}^{\text{mt}}(\omega) - S^{\text{mt}}(\omega)|^2] + \text{Var}(S^{\text{mt}}(\omega)) \\ &\quad + 2\mathbb{E}[(\hat{S}^{\text{mt}}(\omega) - S^{\text{mt}}(\omega))(S^{\text{mt}}(\omega) - \mathbb{E}[S^{\text{mt}}(\omega)])] \\ &\stackrel{(b)}{\leq} \left\{ \sqrt{\mathbb{E}[|\hat{S}^{\text{mt}}(\omega) - S^{\text{mt}}(\omega)|^2]} + \sqrt{\text{Var}(S^{\text{mt}}(\omega))} \right\}^2, \end{aligned} \tag{A.2}$$

where (b) follows from the Cauchy-Schwarz inequality. Thus, the desired bounds on the bias and variance can be established through bounding the first and second moments of $(\hat{S}^{\text{mt}}(\omega) -$

$S^{\text{mt}}(\omega)$). The first moment can be bounded by

$$\begin{aligned}
\left| \mathbb{E}[\widehat{S}^{\text{mt}}(\omega) - S^{\text{mt}}(\omega)] \right| &\stackrel{(c)}{\leq} \frac{1}{P} \sum_{p=1}^P \left| \mathbb{E} \left[|\widehat{y}^{(p)}(\omega)|^2 - |y^{(p)}(\omega)|^2 \right] \right| \\
&\stackrel{(d)}{\leq} \frac{1}{P} \sum_{p=1}^P \sum_{k=1}^K \sum_{m=1}^K \left| \mathbb{E}[\ell^{-1}(\bar{n}_k) \ell^{-1}(\bar{n}_m) - x_k x_m] \nu_k^{(p)} \nu_m^{(p)} e^{-i\omega(m-k)} \right| \\
&\stackrel{(e)}{\leq} \frac{1}{P} \sum_{p=1}^P \sum_{k=1}^K \left| \mathbb{E}[(\ell^{-1}(\bar{n}_k))^2 - x_k^2] \right| (\nu_k^{(p)})^2 \\
&\quad + \frac{1}{P} \sum_{p=1}^P \sum_{k \neq m} \left| \mathbb{E}[\ell^{-1}(\bar{n}_k) \ell^{-1}(\bar{n}_m) - x_k x_m] \right| \left| \nu_k^{(p)} \nu_m^{(p)} \right|, \tag{A.3}
\end{aligned}$$

where (c) and (d) follow from the triangle inequality and (e) follows by bounding the complex sinusoid. The main technical difficulty in further development of the bounds is due to the fact that $\ell^{-1}(z)$ does not have a Taylor series expansion for $z \in (0, 1)$. We thus have to find other algebraically useful bounds. To this end, we need to establish the following technical lemma.

Lemma A.1. *Consider the event $A = \{\bar{n}_k \mid \bar{n}_k \neq 0, \bar{n}_k \neq 1, 1 \leq k \leq K\}$. The following inequality holds true for all $\bar{n}_k \in A$:*

$$\varepsilon(\bar{n}_k) := \left| \ell^{-1}(\bar{n}_k) - x_k \right| \leq g(x_k, L) \left| \bar{n}_k - \lambda_k \right|,$$

where

$$g(x_k, L) = \max \left\{ \frac{1}{\lambda_k(1 - \lambda_k)}, \frac{|\log(L - 1) + x_k|}{|\lambda_k - 1/L|}, \frac{|\log(L - 1) - x_k|}{|1 - 1/L - \lambda_k|} \right\}.$$

Proof of Lemma A.1. First, consider the case $\lambda_k \leq 0.5$. We bound the function $\varepsilon(\bar{n}_k)$ in a piecewise fashion as follows. Note that $\ell^{-1}(\bar{n}_k)$ is convex for $\bar{n}_k \geq 0.5$ and concave for $\bar{n}_k \leq 0.5$.

Thus, it immediately follows that for $\bar{n}_k \leq \lambda_k$, $\varepsilon(\bar{n}_k)$ is convex and hence,

$$\varepsilon(\bar{n}_k) \leq \frac{|\log(L-1) + x_k|}{|\lambda_k - 1/L|} (\lambda_k - \bar{n}_k). \quad (\text{A.4})$$

Furthermore, for $\lambda_k \leq \bar{n}_k \leq 0.5$, $\varepsilon(\bar{n}_k)$ is concave, and hence is bounded by the tangent at λ_k as follows.

$$\varepsilon(\bar{n}_k) \leq \frac{1}{\lambda_k(1 - \lambda_k)} (\bar{n}_k - \lambda_k) \quad (\text{A.5})$$

Finally, for the case of $\bar{n}_k \geq 0.5$, consider the line

$$h(\bar{n}_k) := \frac{|\log(L-1) - x_k|}{|1 - 1/L - \lambda_k|} (\bar{n}_k - \lambda_k). \quad (\text{A.6})$$

From the convexity of $\varepsilon(\bar{n}_k)$, $h(\bar{n}_k)$ upper bounds $\varepsilon(\bar{n}_k)$ for $\bar{n}_k \geq 0.5$, since $h(0.5) \geq \varepsilon(0.5)$ for $\lambda_k \leq 0.5$. Combining the piece-wise bounds in Eqs. (A.4), (A.5) and (A.6), we conclude the claim in Lemma A.1 for $\lambda_k \leq 0.5$. Due to the symmetry of $\varepsilon(\bar{n}_k)$, through a similar argument, the bound can be established for $\lambda_k > 0.5$, which concludes the proof. $\square$

Given that $|x_k| \leq B$ and assuming that L is large enough so that $L \geq 2(1 + \exp(B))$, we can further simplify the bound of Lemma A.1. We have:

$$\begin{aligned} g(x_k, L) &\leq \max \left\{ \exp(B) (1 + \exp(-B))^2, \frac{|\log(L-1) + B|}{(1/(1 + \exp(B)) - 1/L)} \right\} \\ &\leq \max \left\{ \exp(B) (1 + \exp(-B))^2, 4(1 + \exp(B)) \log L \right\}. \end{aligned}$$

Thus, for sufficiently large L , we conclude that,

$$\varepsilon(\bar{n}_k) \leq 4(1 + \exp(B)) \log L |\bar{n}_k - \lambda_k|. \quad (\text{A.7})$$

Now, consider the expectations in the bounds of Eq. (A.3). Using iterated conditioning,

$$\begin{aligned}
|\mathbb{E}[(\ell^{-1}(\bar{n}_k))^2 - x_k^2]| &= |\mathbb{E}[\mathbb{E}[(\ell^{-1}(\bar{n}_k))^2 | x_k] - x_k^2]| \\
&= |\mathbb{E}[2x_k \mathbb{E}[(\ell^{-1}(\bar{n}_k) - x_k) | x_k] + \mathbb{E}[(\ell^{-1}(\bar{n}_k) - x_k)^2 | x_k]]| \\
&\stackrel{(f)}{\leq} \mathbb{E}[2|x_k| \mathbb{E}[|\ell^{-1}(\bar{n}_k) - x_k| | x_k]] + \mathbb{E}[\mathbb{E}[(\ell^{-1}(\bar{n}_k) - x_k)^2 | x_k]], \quad (\text{A.8})
\end{aligned}$$

where (f) follows from triangle and Jensen's inequalities. In order further simplify these bounds, we invoke the result of Lemma A.1. First, note that $\ell^{-1}(\bar{n}_k)$ is unbounded in the complement of event A . Provided $|x_k| \leq B$, $\mathbb{P}(\bar{n}_k \neq 0)$ and $\mathbb{P}(\bar{n}_k \neq 1)$ can be lower bounded by $1 - \exp(-L \log(1 + \exp(-B)))$, which implies that

$$\mathbb{P}(A) \geq 1 - 2 \exp(-L \log(1 + \exp(-B))).$$

Therefore, for sufficiently large L , we see that $\mathbb{P}(A)$ is exponentially close to 1. Thus, hereafter we condition the expectations on the highly probable event A . From Eq. (A.7), we get

$$\mathbb{E}[|\ell^{-1}(\bar{n}_k) - x_k| | x_k, A] \leq 4(1 + \exp(B)) \log L \mathbb{E}[|\bar{n}_k - \lambda_k| | x_k, A].$$

Note that the random variable $n_k = L\bar{n}_k$ is the sum of L independent Bernoulli random variables

given x_k . Thus, given x_k , $n_k \sim \text{Binomial}(L, \lambda_k)$. Accordingly,

$$\begin{aligned}
\mathbb{E}[|\bar{n}_k - \lambda_k| \mid x_k, A] &= \mathbb{E}[|\bar{n}_k - \lambda_k| \mathbb{1}_A \mid x_k] / \mathbb{P}(A) \\
&\stackrel{(g)}{\leq} \sqrt{\mathbb{E}[(\bar{n}_k - \lambda_k)^2 \mathbb{1}_A \mid x_k]} / \mathbb{P}(A) \\
&\stackrel{(h)}{\leq} \sqrt{\lambda_k(1 - \lambda_k)} / (\sqrt{L} \mathbb{P}(A)), \tag{A.9}
\end{aligned}$$

where (g) follows from the Jensen's inequality and (h) follows from substituting expression for the variance of a binomial random variable. Further, note that $\lambda_k(1 - \lambda_k) \leq 1/4$, for $\lambda_k \in [0, 1]$ and $\mathbb{P}(A) \geq 1/2$, if $L \geq 2B/\log(1 + \exp(B))$, which is satisfied for large enough L . Thus, combining the bounds in Eqs. (A.9) and (A.7), we get,

$$\mathbb{E}[|\ell^{-1}(\bar{n}_k) - x_k| \mid x_k, A] \leq 4(1 + \exp(B)) \frac{\log L}{\sqrt{L}}. \tag{A.10}$$

By a similar argument we can show that,

$$\mathbb{E}[(\ell^{-1}(\bar{n}_k) - x_k)^2 \mid x_k, A] \leq 8(1 + \exp(B))^2 \left(\frac{\log L}{\sqrt{L}} \right)^2.$$

Thus, the expectation in Eq. (A.8) is bounded as:

$$|\mathbb{E}[(\ell^{-1}(\bar{n}_k))^2 - x_k^2 \mid A]| \leq 8(1 + \exp(B)) \frac{\log L}{\sqrt{L}} \left(B + (1 + \exp(B)) \frac{\log L}{\sqrt{L}} \right). \tag{A.11}$$

Following a similar argument, one can show for $n \neq m$,

$$|\mathbb{E}[\ell^{-1}(\bar{n}_k)\ell^{-1}(\bar{n}_m) - x_k x_m | A]| \leq 8(1 + \exp(B)) \frac{\log L}{\sqrt{L}} \left(B + 2(1 + \exp(B)) \frac{\log L}{\sqrt{L}} \right). \quad (\text{A.12})$$

Finally, using the bounds of Eqs. (A.11) and (A.12) and noting that $\sum_{k=0}^{K-1} (\nu_k^{(p)})^2 = 1$, for all $0 \leq p \leq P$, we can upper bound the expectation in Eq. (A.3) as,

$$|\mathbb{E}[\widehat{S}^{\text{mt}}(\omega) - S^{\text{mt}}(\omega) | A]| \leq g_1 K \frac{\log L}{\sqrt{L}}, \quad (\text{A.13})$$

where

$$g_1 := 8(1 + \exp(B)) \left\{ \left(\frac{1}{K} + 1 \right) B + \left(\frac{1}{K} + 2 \right) (1 + \exp(B)) \frac{\log L}{\sqrt{L}} \right\}.$$

This concludes the proof of the bound on bias. Following similar bounding techniques, the second moment in Eq. (A.2) can be bounded by:

$$\sqrt{\mathbb{E}[|\widehat{S}^{\text{mt}}(\omega) - S^{\text{mt}}(\omega)|^2 | A]} \leq g_2 K \frac{\log L}{\sqrt{L}}, \quad (\text{A.14})$$

where

$$g_2 := 4(1 + \exp(B)) \left\{ \frac{\sqrt{2}}{K} \left[\sqrt{\frac{13}{3}} \frac{\log L}{\sqrt{L}} (1 + \exp(B)) + B \right] + \left[4 \left(\frac{\log L}{\sqrt{L}} (1 + \exp(B)) + B \right)^2 - B^2 \right]^{1/2} \right\}.$$

This concludes the proof of Theorem 2.1. $\square$

A.2 Proof of Corollaries 2.1 and 2.2

Proof of Corollary 2.1. Proof of Corollary 2.1 follows the proof of Theorem 2.1 closely, with the natural extension to the multivariate case. Following the proof of Theorem 2.1, the constants g'_1 and g'_2 in this case are given by:

$$g'_1 := 8(1 + \exp(B)) \times \left(B + 2(1 + \exp(B)) \frac{\log L}{\sqrt{L}} \right)$$

$$g'_2 := 4(1 + \exp(B)) \sqrt{4 \left(\frac{\log L}{\sqrt{L}} (1 + \exp(B)) + B \right)^2 - B^2}$$

$\square$

Proof of Corollary 2.2. As for Corollary 2.2, we work under the technical assumptions of Theorems 1 and 2 in [96]. Following [96], we assume that in Eq. (2.9), $\mathbf{Q}_m = \mathbf{Q}$ for all m in this proof, and that the EM algorithm finds estimates of $\mathbf{Q}$ and α close to their true value (for large enough K). Then, under Assumptions (1) and (2), we identify the effective observation $\tilde{\mathbf{y}}_m^{(p)}$ corresponding to the p^{th} taper at window m in [96] by the concatenation of $\nu_w^{(p)} \text{logit}(\bar{n}_{(m-1)W+w,j})$ for $w = 1, 2, \dots, W$ and $j = 1, 2, \dots, J$ in a vector of length WJ . We also assume, without loss of generality that $W = uN$, for some integer u . Then, we denote by Σ_∞ the steady state covariance of the backward smoother, and $\mathbf{\Lambda} := \alpha \Sigma_\infty (\alpha^2 \Sigma_\infty + \mathbf{Q})^{-1}$ and $\mathbf{\Gamma} = (\alpha^2 \Sigma_\infty + \mathbf{Q}) [\mathbf{I} - uW((\alpha^2 \Sigma_\infty + \mathbf{Q})^{-1} + uW\mathbf{I})^{-1}]$, as in [96]. Note that these matrices are $NJ \times NJ$ in our case. Under the same assumptions [96], we consider them to be diagonal with the i^{th} diagonal element

being γ_i and η_i respectively. Then, following the proof of Theorem 2.1, Corollary 2.1 and those of Theorems 1 and 2 in [96], it can be shown that the statement of the corollary holds with the constants:

$$g_1''(\omega_n) := 8(1 + \exp(B)) \left(B + 2(1 + \exp(B)) \frac{\log L}{\sqrt{L}} \right) \\ \times \eta_{(n-1)J+r} \eta_{(n-1)J+t} \sum_{s,s'=1}^M \gamma_{(n-1)J+r}^{|s-m|} \gamma_{(n-1)J+t}^{|s'-m|},$$

$$g_2''(\omega_n) := 4(1 + \exp(B)) \sqrt{4 \left(\frac{\log L}{\sqrt{L}} (1 + \exp(B)) + B \right)^2 - B^2} \\ \times \eta_{(n-1)J+r} \eta_{(n-1)J+t} \sum_{s,s'=1}^M \gamma_{(n-1)J+r}^{|s-m|} \gamma_{(n-1)J+t}^{|s'-m|},$$

and

$$\kappa_m(\omega_n) := \eta_{(n-1)J+r} \eta_{(n-1)J+t} \sum_{s,s'=1}^M \gamma_{(n-1)J+r}^{|s-m|} \gamma_{(n-1)J+t}^{|s'-m|} \alpha^{|s-s'|}.$$

We refer the interested readers to [96] for detailed derivations.

□

Chapter B: Supplementary Material on Chapter 5

This appendix contains the proofs of the connection between proposed estimators and conventional definitions, proofs of the theorems characterizing the bias and variance for the proposed estimators, the proposed signal and noise correlation estimation framework for electrophysiology recordings, and supplementary figures further elaborating the simulation and real data studies in Chapter 5.

B.1 Relationship to existing definitions of Signal and Noise correlations

Recall that the conventional definitions of signal and noise covariance of spiking activity between the i^{th} and j^{th} neuron are [62]:

$$\begin{aligned} (\Sigma_s^{\text{con}})_{i,j} &= \text{cov} \left(\frac{1}{L} \sum_l n_{t,l}^{(i)}, \frac{1}{L} \sum_l n_{t,l}^{(j)} \right), \\ (\Sigma_x^{\text{con}})_{i,j} &= \frac{1}{L} \sum_l \text{cov} \left(n_{t,l}^{(i)} - \frac{1}{L} \sum_l n_{t,l}^{(i)}, n_{t,l}^{(j)} - \frac{1}{L} \sum_l n_{t,l}^{(j)} \right), \end{aligned}$$

where $\text{cov}(\mathbf{u}_t, \mathbf{v}_t) := \frac{1}{T} \sum_{t=1}^T \left(\mathbf{u}_t - \frac{1}{T} \sum_{t=1}^T \mathbf{u}_t \right) \left(\mathbf{v}_t - \frac{1}{T} \sum_{t=1}^T \mathbf{v}_t \right)^\top$, is the empirical covariance. The correlations, are then derived by the standard normalization:

$$(\mathbf{S}^{\text{con}})_{i,j} := \frac{(\Sigma_s^{\text{con}})_{i,j}}{\sqrt{(\Sigma_s^{\text{con}})_{i,i} \cdot (\Sigma_s^{\text{con}})_{j,j}}}, (\mathbf{N}^{\text{con}})_{i,j} := \frac{(\Sigma_x^{\text{con}})_{i,j}}{\sqrt{(\Sigma_x^{\text{con}})_{i,i} \cdot (\Sigma_x^{\text{con}})_{j,j}}}, \forall i, j = 1, 2, \dots, N. \quad (\text{B.1})$$

Suppose that the spiking events follow the forward model:

$$n_{t,l}^{(j)} \sim \text{Bernoulli} \left(\lambda_{t,l}^{(j)} \right),$$

$$\lambda_{t,l}^{(j)} = \phi \left(x_{t,l}^{(j)}, \mathbf{d}_j^\top \mathbf{s}_t \right),$$

where $\phi : \mathbb{R}^2 \rightarrow [0, 1]$ is a differentiable non-linear mapping. We assume $\mathbf{x}_{t,l}$ and $\mathbf{s}_t$ to be independent. Without loss of generality, let $\mathbb{E}[\mathbf{s}_t] = \mathbf{0}$ and $\mathbb{E}[\mathbf{x}_{t,l}] = \boldsymbol{\mu}_x$. Further, we define the notation $X_t \approx Y_t$ to denote almost sure equivalence, i.e., $X_t \xrightarrow{\text{a.s.}} Z$ and $Y_t \xrightarrow{\text{a.s.}} Z$ for some random variable Z .

First, let us consider $(\mathbf{S}^{\text{con}})_{i,j}$. Noting that $\mathbb{E}[n_{t,l}^{(j)}] = \mathbb{E}[\lambda_{t,l}^{(j)}]$ and $\mathbb{E}[n_{t,l}^{(i)} n_{t,l}^{(j)}] = \mathbb{E}[\lambda_{t,l}^{(i)} \lambda_{t,l}^{(j)}]$, we conclude as $T \rightarrow \infty$:

$$(\boldsymbol{\Sigma}_s^{\text{con}})_{i,j} \approx \text{cov} \left(\frac{1}{L} \sum_l \lambda_{t,l}^{(i)}, \frac{1}{L} \sum_l \lambda_{t,l}^{(j)} \right),$$

from the law of large numbers. Then, if we consider the Taylor series expansion of $\phi \left(x_{t,l}^{(j)}, \mathbf{d}_j^\top \mathbf{s}_t \right)$ around the mean $(\mu_x^{(j)}, 0)$, we get:

$$(\boldsymbol{\Sigma}_s^{\text{con}})_{i,j} \approx \text{cov} \left(\phi \left(\mu_x^{(i)}, 0 \right) + \frac{1}{L} \sum_l \left(x_{t,l}^{(i)} - \mu_x^{(i)} \right) \phi_{(x_{t,l}^{(i)})} \left(\mu_x^{(i)}, 0 \right) + (\mathbf{d}_i^\top \mathbf{s}_t) \phi_{(\mathbf{d}_i^\top \mathbf{s}_t)} \left(\mu_x^{(i)}, 0 \right) + \epsilon_{t,l}^{(i)}, \right.$$

$$\left. \phi \left(\mu_x^{(j)}, 0 \right) + \frac{1}{L} \sum_l \left(x_{t,l}^{(j)} - \mu_x^{(j)} \right) \phi_{(x_{t,l}^{(j)})} \left(\mu_x^{(j)}, 0 \right) + (\mathbf{d}_j^\top \mathbf{s}_t) \phi_{(\mathbf{d}_j^\top \mathbf{s}_t)} \left(\mu_x^{(j)}, 0 \right) + \epsilon_{t,l}^{(j)} \right),$$

where $\epsilon_{t,l}^{(i)}$ and $\epsilon_{t,l}^{(j)}$ represent the higher order terms. Then, as $L \rightarrow \infty$, we get:

$$(\boldsymbol{\Sigma}_s^{\text{con}})_{i,j} \approx \text{cov} \left((\mathbf{d}_i^\top \mathbf{s}_t) \phi_{(\mathbf{d}_i^\top \mathbf{s}_t)} \left(\mu_x^{(i)}, 0 \right), (\mathbf{d}_j^\top \mathbf{s}_t) \phi_{(\mathbf{d}_j^\top \mathbf{s}_t)} \left(\mu_x^{(j)}, 0 \right) \right) + \epsilon_{t,l},$$

since $\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{l=1}^L \left(x_{t,l}^{(j)} \right) = \mu_x^{(j)}$ by the Law of Large numbers. Thus, we see that:

$$\begin{aligned} (\Sigma_s^{\text{con}})_{i,j} &\approx C_S \mathbf{d}_i^\top \text{cov}(\mathbf{s}_t, \mathbf{s}_t) \mathbf{d}_j + \epsilon_{t,l} \\ &= C_S (\Sigma_s)_{i,j} + \epsilon_{t,l}, \end{aligned}$$

where C_S is a constant and $\epsilon_{t,l}$ is typically small if the latent process $\mathbf{x}_{t,l}$ and the stimulus $\mathbf{s}_t$ are concentrated around their means. Then, the signal correlations are obtained by normalization of the signal covariance as in Eq. (B.1), through which the scaling factor C_S cancels and we get:

$$(\mathbf{S}^{\text{con}})_{i,j} \approx (\mathbf{S})_{i,j}.$$

Thus, as $T, L \rightarrow \infty$, we see that $\mathbf{S}$ is indeed the signal correlation matrix that is aimed to be approximated by the conventional definitions.

Next, let us consider $(\mathbf{N}^{\text{con}})_{i,j}$. Similar to foregoing analysis of the signal covariance, as $T \rightarrow \infty$ we get:

$$(\Sigma_x^{\text{con}})_{i,j} \approx \frac{1}{L} \sum_l \text{cov} \left(\lambda_{t,l}^{(i)} - \frac{1}{L} \sum_l \lambda_{t,l}^{(i)}, \lambda_{t,l}^{(j)} - \frac{1}{L} \sum_l \lambda_{t,l}^{(j)} \right).$$

Then, from a Taylor series expansion, we get:

$$\begin{aligned} (\Sigma_x^{\text{con}})_{i,j} &\approx \frac{1}{L} \sum_l \text{cov} \left(x_{t,l}^{(i)} \phi_{(x_{t,l}^{(i)})}(\mu_x^{(i)}, 0) - \frac{1}{L} \sum_l x_{t,l}^{(i)} \phi_{(x_{t,l}^{(i)})}(\mu_x^{(i)}, 0) + \xi_{t,l}^{(i)}, \right. \\ &\quad \left. x_{t,l}^{(j)} \phi_{(x_{t,l}^{(j)})}(\mu_x^{(j)}, 0) - \frac{1}{L} \sum_l x_{t,l}^{(j)} \phi_{(x_{t,l}^{(j)})}(\mu_x^{(j)}, 0) + \xi_{t,l}^{(j)} \right), \end{aligned}$$

where $\xi_{t,l}^{(i)}$ and $\xi_{t,l}^{(j)}$ represent the higher order terms. Then, as $L \rightarrow \infty$:

$$(\Sigma_x^{\text{con}})_{i,j} \approx \frac{1}{L} \sum_l \text{cov} \left(\left(x_{t,l}^{(i)} - \mu_x^{(i)} \right) \phi_{(x_{t,l}^{(i)})} \left(\mu_x^{(i)}, 0 \right), \left(x_{t,l}^{(j)} - \mu_x^{(j)} \right) \phi_{(x_{t,l}^{(j)})} \left(\mu_x^{(j)}, 0 \right) \right) + \xi_{t,l},$$

from the law of large numbers. Accordingly, we see that:

$$\begin{aligned} (\Sigma_x^{\text{con}})_{i,j} &\approx C_N \frac{1}{L} \sum_l \text{cov} \left(x_{t,l}^{(i)} - \mu_x^{(i)}, x_{t,l}^{(j)} - \mu_x^{(j)} \right) + \xi_{t,l} \\ &= C_N (\Sigma_x)_{i,j} + \xi_{t,l}, \end{aligned}$$

where C_N is a constant and $\xi_{t,l}$ is typically small if the latent process $\mathbf{x}_{t,l}$ and the stimulus $\mathbf{s}_t$ are concentrated around their means. Then, the noise correlations are derived by normalization of the noise covariance given in Eq. (B.1). This cancels out the scaling factor C_N , and we get:

$$(\mathbf{N}^{\text{con}})_{i,j} \approx (\mathbf{N})_{i,j}.$$

Thus, we similarly conclude that as $T, L \rightarrow \infty$, the conventional definition of noise correlation $\mathbf{N}^{\text{con}}$ indeed aims to approximate $\mathbf{N}$.

As a numerical illustration, we demonstrated in Appendix B.4.2 that the conventional definitions of the correlations indeed approximate our proposed definitions, but require a much larger number of trials to be accurate. More specifically, to achieve comparable performance to our method using $L = 20$ trials, the conventional correlation estimates require $L = 1000$ trials.

B.2 Proof of Theorem 5.1

In what follows, we present a comprehensive proof of the Theorem 5.1. Recall the following key assumptions:

Assumption (1). We assume a scalar time-varying external stimulus (i.e. $s_t = s_t$, and hence $\mathbf{d}_j = d_j$, $\mathbf{d} = [d_1, d_2, \dots, d_N]^\top$). Furthermore, we set the observation noise covariance to be $\Sigma_w = \sigma_w^2 \mathbf{I}$, for notational convenience.

Assumption (2). We derive the performance bounds in the regime where T and L are large, and thus do not impose any prior distribution on the correlations (i.e., $p_{\text{pr}}(\Sigma_x) \propto 1$), which are otherwise needed to mitigate overfitting (see Section 5.3).

Assumption (3). We assume the latent noise process and stimulus to be slowly varying signals, and thus adopt a piece-wise constant model in which these processes are constant within consecutive windows of length W (i.e., $\mathbf{x}_{t,l} = \mathbf{x}_{W_k,l}$ and $s_t = s_{W_k}$, for $(k-1)W + 1 \leq t < kW$ and $k = 1, \dots, K$ with $W_k = (k-1)W + 1$ and $KW = T$) for our theoretical analysis, as is usually done in spike count calculations for conventional noise correlation estimates.

Proof of Theorem 5.1. First, recall the proposed forward model (see Section 5.3) under Assumption (1)–(3):

$$\mathbf{y}_{t,l} = \mathbf{A}\mathbf{z}_{t,l} + \mathbf{w}_{t,l},$$

$$\mathbf{z}_{t,l} = \alpha \mathbf{z}_{t-1,l} + \mathbf{n}_{t,l},$$

$$n_{t,l}^{(j)} \sim \text{Bernoulli} \left(\phi \left(x_{W_k,l}^{(j)} \right) \right),$$

$$\mathbf{x}_{W_k,l} \sim \mathcal{N}(\boldsymbol{\mu}_x + s_{W_k} \mathbf{d}, \Sigma_x),$$

where $\phi(\cdot) := \frac{\exp(\cdot)}{1+\exp(\cdot)}$, is the logistic function. Note that we have re-defined the latent process $\mathbf{x}_{t,l}$ by absorbing the stimulus activity $s_t \mathbf{d}$ to the mean of $\mathbf{x}_{t,l}$ for notational convenience, without loss of generality. Hereafter, we also assume that $\mathbf{A} = \mathbf{I}$ without loss of generality. For a truncation level B (to be specified later), consider the event $\mathcal{A}_W = \left\{ \left| x_{W_k,l}^{(j)} \right| \leq B \text{ and } \frac{1}{2(1+\exp(B))} \leq \bar{n}_{W_k,l}^{(j)} \leq 1 - \frac{1}{2(1+\exp(B))} \text{ for } j = 1, \dots, N, k = 1, \dots, K \text{ and } l = 1, \dots, L \right\}$, such that $\bar{\mathbf{n}}_{W_k,l} = \left[\bar{n}_{W_k,l}^{(1)}, \bar{n}_{W_k,l}^{(2)}, \dots, \bar{n}_{W_k,l}^{(N)} \right]^\top := \frac{1}{W} \sum_{w=1}^W \mathbf{n}_{(k-1)W+w,l}$. First, we derive convenient forms of the maximum likelihood estimators via the Laplace's approximations and asymptotic expansions [174] through the following lemma:

Lemma B.1. *Conditioned on event $\mathcal{A}_W$, the maximum likelihood estimators of the stimulus kernel of the j^{th} neuron and the noise covariance between the i^{th} and j^{th} neurons take the forms:*

$$\begin{aligned} \hat{d}_j &= \tilde{d}_j \left(1 + \mathcal{O}(\sigma_w^2) \right) \left(1 + \mathcal{O}\left(\frac{1}{W}\right) \right) \quad \text{and} \\ (\hat{\Sigma}_x)_{i,j} &= (\tilde{\Sigma}_x)_{i,j} \left(1 + \mathcal{O}(\sigma_w^2) \right) \left(1 + \mathcal{O}\left(\frac{1}{W}\right) \right), \end{aligned}$$

where

$$\begin{aligned} \tilde{d}_j &= \frac{1}{L \sum_{k=1}^K s_{W_k}^2} \sum_{k,l=1}^{K,L} s_{W_k} \left(\phi^{-1} \left(\tilde{n}_{W_k,l}^{(j)} \right) - \mu_x^{(j)} \right) \quad \text{and} \\ (\tilde{\Sigma}_x)_{i,j} &= \frac{1}{KL} \sum_{k,l=1}^{K,L} \left(\phi^{-1} \left(\tilde{n}_{W_k,l}^{(i)} \right) - \mu_x^{(i)} - s_{W_k} \tilde{d}_i \right) \left(\phi^{-1} \left(\tilde{n}_{W_k,l}^{(j)} \right) - \mu_x^{(j)} - s_{W_k} \tilde{d}_j \right), \end{aligned}$$

with $\tilde{\mathbf{n}}_{W_k,l} = \left[\tilde{n}_{W_k,l}^{(1)}, \tilde{n}_{W_k,l}^{(2)}, \dots, \tilde{n}_{W_k,l}^{(N)} \right]^\top := \frac{1}{W} \sum_{w=1}^W (\mathbf{y}_{(k-1)W+w,l} - \alpha \mathbf{y}_{(k-1)W+w-1,l})$ and $\phi^{-1}(z) := \ln(z/(1-z))$.

Proof of Lemma B.1. First, maximizing the data likelihood, we derive the estimators:

$$\begin{aligned}
\hat{d}_j &= \underset{d_j}{\operatorname{argmax}} p(\mathbf{y}|\Sigma_x, \mathbf{d}) \\
&= \frac{\int \left(\frac{1}{L \sum_{k=1}^K s_{W_k}^2} \sum_{k,l=1}^{K,L} s_{W_k} \left(x_{W_k,l}^{(j)} - \mu_x^{(j)} \right) \right) p(\mathbf{y}|\mathbf{n}) p(\mathbf{n}|\mathbf{x}) p(\mathbf{x}|\Sigma_x, \mathbf{d}) d\mathbf{n} d\mathbf{x}}{\int p(\mathbf{y}|\mathbf{n}) p(\mathbf{n}|\mathbf{x}) p(\mathbf{x}|\Sigma_x, \mathbf{d}) d\mathbf{n} d\mathbf{x}}, \quad (\text{B.2})
\end{aligned}$$

and

$$\begin{aligned}
(\hat{\Sigma}_x)_{i,j} &= \underset{(\Sigma_x)_{i,j}}{\operatorname{argmax}} p(\mathbf{y}|\Sigma_x, \mathbf{d}) \\
&= \frac{\int \left(\frac{1}{KL} \sum_{k,l=1}^{K,L} \left(x_{W_k,l}^{(i)} - \mu_x^{(i)} - s_{W_k} \hat{d}_i \right) \left(x_{W_k,l}^{(j)} - \mu_x^{(j)} - s_{W_k} \hat{d}_j \right) \right) p(\mathbf{y}|\mathbf{n}) p(\mathbf{n}|\mathbf{x}) p(\mathbf{x}|\Sigma_x, \mathbf{d}) d\mathbf{n} d\mathbf{x}}{\int p(\mathbf{y}|\mathbf{n}) p(\mathbf{n}|\mathbf{x}) p(\mathbf{x}|\Sigma_x, \mathbf{d}) d\mathbf{n} d\mathbf{x}}, \quad (\text{B.3})
\end{aligned}$$

where $W_k = (k-1)W + 1$. Then, we simplify these integrals based on the saddle point method of asymptotic expansions [174]. To that end, first consider the numerator of Eq. (B.2) denoted by $I_{\text{num}}^{(1)}$. First, we evaluate the integration in $I_{\text{num}}^{(1)}$ relative to the variable $\mathbf{n}$. To that end, note:

$$I_{\text{num}}^{(1)} = \int h_{\text{num}}^{(1)}(\mathbf{n}) \exp(A_1 f_1(\mathbf{n})) d\mathbf{n},$$

where $h_{\text{num}}^{(1)}(\mathbf{n}) = \frac{1}{\sqrt{(2\pi)^{TNL} \sigma_w^2 TNL}} \int \left(\frac{1}{L \sum_{k=1}^K s_{W_k}^2} \sum_{k,l=1}^{K,L} s_{W_k} \left(x_{W_k,l}^{(j)} - \mu_x^{(j)} \right) \right) p(\mathbf{n}|\mathbf{x}) p(\mathbf{x}|\Sigma_x, \mathbf{d}) d\mathbf{x}$, $A_1 = \frac{1}{\sigma_w^2}$, $f_1(\mathbf{n}) = -\frac{1}{2} \sum_{t,l,j} \left(y_{t,l}^{(j)} - \sum_{k=1}^t \alpha^{t-k} n_{t,l}^{(j)} \right)^2$ and $d\mathbf{n}$ is shorthand notation for the product measure of the discrete random vector $\mathbf{n}$. Observing that $\nabla f_1(\hat{\mathbf{n}}) = \mathbf{0}$ for $\hat{\mathbf{n}} := \{\hat{\mathbf{n}}_{t,l} = \mathbf{y}_{t,l} - \alpha \mathbf{y}_{t-1,l}\}_{t,l=1}^{T,L}$, using the method of asymptotic expansions, $I_{\text{num}}^{(1)}$ can be evaluated as:

$$I_{\text{num}}^{(1)} = h_{\text{num}}^{(1)}(\hat{\mathbf{n}}) \times \exp(A_1 f_1(\hat{\mathbf{n}})) \sqrt{\frac{(2\pi)^{TNL}}{-A_1 |H(f_1)|}} \left(1 + \mathcal{O}\left(\frac{1}{A_1}\right) \right), \quad (\text{B.4})$$

where the determinant of the Hessian matrix $|H(f_1)|$, is a negative function of α . Note that

the covariance of this Gaussian integral $(-(H(f_1))^{-1})$ is a function of $\alpha \in (0, 1)$, and hence is bounded. Thus, all higher order error terms in Eq. (B.4) are also bounded, as higher order moments of Gaussian distributions are functions of the covariance.

Next, we simplify the integral $h_{\text{num}}^{(1)}(\hat{\mathbf{n}})$ in Eq. (B.4) using a similar procedure. We have:

$$h_{\text{num}}^{(1)}(\hat{\mathbf{n}}) = \int r_{\text{num}}^{(1)}(\mathbf{x}) \exp(A_2 f_2(\mathbf{x})) d\mathbf{x}, \quad (\text{B.5})$$

where $f_2(\mathbf{x}) = \sum_{k,l,j} \left(\tilde{n}_{W_k,l}^{(j)} x_{W_k,l}^{(j)} - \log(1 + \exp(x_{W_k,l}^{(j)})) \right)$ with $\tilde{\mathbf{n}}_{W_k,l} = [\tilde{n}_{W_k,l}^{(1)}, \tilde{n}_{W_k,l}^{(2)}, \dots, \tilde{n}_{W_k,l}^{(N)}]^\top := \frac{1}{W} \sum_{w=1}^W \hat{\mathbf{n}}_{(k-1)W+w,l}$, $r_{\text{num}}^{(1)}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^{(W+1)KLN} \sigma_w^{2TNL} |\Sigma_x|^{KL}}} \times \exp\left(-\frac{1}{2} \sum_t (\mathbf{x}_{W_k,l} - \boldsymbol{\mu}_x - s_{W_k} \mathbf{d})^\top \Sigma_x^{-1} (\mathbf{x}_{W_k,l} - \boldsymbol{\mu}_x - s_{W_k} \mathbf{d})\right) \left(\frac{1}{L \sum_{k=1}^K s_{W_k}^2} \sum_{k=1}^K s_{W_k,l} \left(x_{W_k,l}^{(j)} - \mu_x^{(j)} \right) \right)$ and $A_2 = W$. Then, we note that the gradient of f_2 , $\nabla f_2(\hat{\mathbf{x}}) = \mathbf{0}$ for $\hat{\mathbf{x}} := \left\{ \hat{x}_{W_k,l}^{(j)} = \phi^{-1} \left(\tilde{n}_{W_k,l}^{(j)} \right) \right\}_{k,l,j=1}^{K,L,N}$, where $\phi^{-1}(z) := \text{logit}(z) = \ln(z/(1-z))$. Accordingly, by re-applying the saddle point method of asymptotic expansions, we evaluate the integral in Eq. (B.5) as:

$$h_{\text{num}}^{(1)}(\hat{\mathbf{n}}) = r_{\text{num}}^{(1)}(\hat{\mathbf{x}}) \times \exp(A_2 f_2(\hat{\mathbf{x}})) \sqrt{\frac{(2\pi)^{KLN}}{-A_2 |H(f_2(\hat{\mathbf{x}}))|}} \left(1 + \mathcal{O}\left(\frac{1}{A_2}\right) \right), \quad (\text{B.6})$$

where the determinant of the Hessian, $|H(f_2(\hat{\mathbf{x}}))| = -\prod_{k,l,j} \tilde{n}_{W_k,l}^{(j)} (1 - \tilde{n}_{W_k,l}^{(j)}) < 0$ when conditioned on event $\mathcal{A}_W$. The higher order terms in Eq. (B.6) will be bounded if the covariance of the saddle point approximation $(-(H(f_2(\hat{\mathbf{x}})))^{-1})$ is bounded, which we ensure by conditioning on event $\mathcal{A}_W$. This completes the evaluation of $I_{\text{num}}^{(1)}$.

Following the same sequence of arguments, we evaluate the denominator of Eq. (B.2)

denoted by $I_{\text{den}}^{(1)}$. Accordingly, we derive:

$$\begin{aligned} I_{\text{den}}^{(1)} &= h_{\text{den}}^{(1)}(\hat{\mathbf{n}}) \times \exp(A_1 f_1(\hat{\mathbf{n}})) \sqrt{\frac{(2\pi)^{TLN}}{-A_1 |H(f_1)|}} \left(1 + \mathcal{O}\left(\frac{1}{A_1}\right)\right), \\ h_{\text{den}}^{(1)}(\hat{\mathbf{n}}) &= r_{\text{den}}^{(1)}(\hat{\mathbf{x}}) \times \exp(A_2 f_2(\hat{\mathbf{x}})) \sqrt{\frac{(2\pi)^{KLN}}{-A_2 |H(f_2(\hat{\mathbf{x}}))|}} \left(1 + \mathcal{O}\left(\frac{1}{A_2}\right)\right), \end{aligned} \quad (\text{B.7})$$

where $r_{\text{den}}^{(1)}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^{(W+1)KLN} \sigma_w^2 T N L |\Sigma_x|^{KL}}} \exp\left(-\frac{1}{2} \sum_{k,l} (\mathbf{x}_{W_k,l} - \boldsymbol{\mu}_x - s_{W_k} \mathbf{d})^\top \Sigma_x^{-1} (\mathbf{x}_{W_k,l} - \boldsymbol{\mu}_x - s_{W_k} \mathbf{d})\right)$.

Finally, by combining Eq. (B.4), Eq. (B.6) and Eq. (B.7), the maximum likelihood estimator in Eq. (B.2) takes the form:

$$\hat{d}_j = \frac{I_{\text{num}}^{(1)}}{I_{\text{den}}^{(1)}} = \tilde{d}_j \frac{\left(1 + \mathcal{O}\left(\frac{1}{A_1}\right)\right) \left(1 + \mathcal{O}\left(\frac{1}{A_2}\right)\right)}{\left(1 + \mathcal{O}\left(\frac{1}{A_1}\right)\right) \left(1 + \mathcal{O}\left(\frac{1}{A_2}\right)\right)} = \tilde{d}_j (1 + \mathcal{O}(\sigma_w^2)) \left(1 + \mathcal{O}\left(\frac{1}{W}\right)\right).$$

Further, following the same sequence of reasoning, simplifying the numerator ($I_{\text{num}}^{(2)}$) and denominator ($I_{\text{den}}^{(2)}$) of Eq. (B.3) yields:

$$\begin{aligned} (\hat{\Sigma}_x)_{i,j} &= \frac{I_{\text{num}}^{(2)}}{I_{\text{den}}^{(2)}} = (\tilde{\Sigma}_x)_{i,j} \frac{\left(1 + \mathcal{O}\left(\frac{1}{A_1}\right)\right) \left(1 + \mathcal{O}\left(\frac{1}{A_2}\right)\right)}{\left(1 + \mathcal{O}\left(\frac{1}{A_1}\right)\right) \left(1 + \mathcal{O}\left(\frac{1}{A_2}\right)\right)} \\ &= (\tilde{\Sigma}_x)_{i,j} (1 + \mathcal{O}(\sigma_w^2)) \left(1 + \mathcal{O}\left(\frac{1}{W}\right)\right). \end{aligned}$$

This concludes the proof of Lemma B.1. □

Given that $\phi^{-1}(z)$ is unbounded for $z = 0$ or $z = 1$, we consider another truncation: $\phi_{B'}^{-1}(z) := \min\{\max\{\phi^{-1}(z), -B'\}, B'\}$, where $B' = 2 \log(2 \exp(B) + 1)$. This choice of B' guarantees that over $\mathcal{A}_W$, $\left|\phi_{B'}^{-1}\left(\bar{n}_{W_k,l}^{(j)}\right)\right| < B'$ for all $j = 1, \dots, N$, $k = 1, \dots, K$ and $l = 1, \dots, L$: and thus $\phi_{B'}^{-1}\left(\bar{n}_{W_k,l}^{(j)}\right) = \phi^{-1}\left(\bar{n}_{W_k,l}^{(j)}\right)$ on $\mathcal{A}_W$.

From Lemma B.1, the bias and variance of the maximum likelihood estimators, $\widehat{d}_j$ and $(\widehat{\Sigma}_x)_{i,j}$ are upper-bounded, if those of $\widetilde{d}_j$ and $(\widetilde{\Sigma}_x)_{i,j}$ are bounded:

$$\left| \text{bias} \left(\widehat{d}_j \right) \right| \leq \left| \text{bias} \left(\widetilde{d}_j \right) \right| + \zeta_j, \quad \text{Var} \left(\widehat{d}_j \right) \leq \text{Var} \left(\widetilde{d}_j \right) + \zeta_j, \quad (\text{B.8})$$

and

$$\left| \text{bias} \left((\widehat{\Sigma}_x)_{i,j} \right) \right| \leq \left| \text{bias} \left((\widetilde{\Sigma}_x)_{i,j} \right) \right| + v_{i,j}, \quad \text{Var} \left((\widehat{\Sigma}_x)_{i,j} \right) \leq \text{Var} \left((\widetilde{\Sigma}_x)_{i,j} \right) + \widetilde{v}_{i,j}, \quad (\text{B.9})$$

where ζ_j , $\widetilde{\zeta}_j$, $v_{i,j}$ and $\widetilde{v}_{i,j}$ represent terms that are $\mathcal{O}(\sigma_w^2)$ or $\mathcal{O} \left(\frac{1}{W} \right)$. Thus, we seek to derive the performance bounds of $\widetilde{d}_j$ and $(\widetilde{\Sigma}_x)_{i,j}$.

Bounding the bias of $\widehat{d}_j$

Let us first consider $\widetilde{d}_j$. Note that:

$$\begin{aligned} \left| \text{bias} \left(\widetilde{d}_j \right) \right| &:= \left| \mathbb{E} \left[\widetilde{d}_j \right] - d_j \right| \\ &\stackrel{(a)}{=} \left| \mathbb{E} \left[\widetilde{d}_j - (d_{\text{Oracle}})_j \right] \right| \\ &\stackrel{(b)}{\leq} \frac{1}{L \sum_{k=1}^K s_{W_k}^2} \sum_{k,l=1}^{K,L} |s_{W_k}| \mathbb{E} \left[\left| \phi_{B'}^{-1} \left(\widetilde{n}_{W_k,l}^{(j)} \right) - x_{W_k,l}^{(j)} \right| \right] \end{aligned} \quad (\text{B.10})$$

where (a) holds since the Oracle estimator, $(d_{\text{Oracle}})_j = \frac{1}{L \sum_{k=1}^K s_{W_k}^2} \sum_{k,l=1}^{K,L} s_{W_k} \left(x_{W_k,l}^{(j)} - \mu_x^{(j)} \right)$ (i.e., observing $\mathbf{x}_{t,l}$ directly) is unbiased and (b) follows through the application of Jensen's in-

equality and triangle inequality. To simplify this bound, the triangle inequality yields:

$$\mathbb{E} \left[\left| \phi_{B'}^{-1} \left(\tilde{n}_{W_k, l}^{(j)} \right) - x_{W_k, l}^{(j)} \right| \right] \leq \mathbb{E} \left[\left| \phi_{B'}^{-1} \left(\tilde{n}_{W_k, l}^{(j)} \right) - \phi_{B'}^{-1} \left(\bar{n}_{W_k, l}^{(j)} \right) \right| \right] + \mathbb{E} \left[\left| \phi_{B'}^{-1} \left(\bar{n}_{W_k, l}^{(j)} \right) - x_{W_k, l}^{(j)} \right| \right]. \quad (\text{B.11})$$

Then, to bound each of these terms, we establish a piece-wise linear Lipschitz-type bound on $\phi_{B'}^{-1}(z)$. First, consider the first term $\mathbb{E} \left[\left| \phi_{B'}^{-1} \left(\tilde{n}_{W_k, l}^{(j)} \right) - \phi_{B'}^{-1} \left(\bar{n}_{W_k, l}^{(j)} \right) \right| \right]$. We seek to upper-bound this expectation by bounding $\left| \phi_{B'}^{-1} \left(\tilde{n}_{W_k, l}^{(j)} \right) - \phi_{B'}^{-1} \left(\bar{n}_{W_k, l}^{(j)} \right) \right|$ via the following technical lemma:

Lemma B.2. *Conditioned on event $\mathcal{A}_W$, the following bound holds for all $j = 1, \dots, N$, $k = 1, \dots, K$ and $l = 1, \dots, L$:*

$$\varepsilon \left(\tilde{n}_{W_k, l}^{(j)}, \bar{n}_{W_k, l}^{(j)} \right) := \left| \phi_{B'}^{-1} \left(\tilde{n}_{W_k, l}^{(j)} \right) - \phi_{B'}^{-1} \left(\bar{n}_{W_k, l}^{(j)} \right) \right| \leq g(B) \left| \tilde{n}_{W_k, l}^{(j)} - \bar{n}_{W_k, l}^{(j)} \right|,$$

where

$$g(B) = \max \left\{ 4(1 + \exp(B))^2, 4 \exp(-B) \log(2 \exp(B) + 1) (1 + (2 \exp(B) + 1)^2) \right\}.$$

Proof of Lemma B.2. First, consider the case $\bar{n}_{W_k, l}^{(j)} \leq 0.5$.

We bound the function $\varepsilon \left(\tilde{n}_{W_k, l}^{(j)}, \bar{n}_{W_k, l}^{(j)} \right)$ in a piece-wise fashion as follows. Note that $\phi_{B'}^{-1} \left(\tilde{n}_{W_k, l}^{(j)} \right)$ is convex for $\tilde{n}_{W_k, l}^{(j)} \geq 0.5$ and concave for $\tilde{n}_{W_k, l}^{(j)} \leq 0.5$. Thus, it immediately

follows that for $\tilde{n}_{W_k,l}^{(j)} \leq \bar{n}_{W_k,l}^{(j)}$, $\varepsilon \left(\tilde{n}_{W_k,l}^{(j)}, \bar{n}_{W_k,l}^{(j)} \right)$ is convex and hence:

$$\varepsilon \left(\tilde{n}_{W_k,l}^{(j)}, \bar{n}_{W_k,l}^{(j)} \right) \leq \frac{\left| B' + \phi_{B'}^{-1} \left(\bar{n}_{W_k,l}^{(j)} \right) \right|}{\left| \bar{n}_{W_k,l}^{(j)} - \frac{1}{1+\exp(B')} \right|} \left(\bar{n}_{W_k,l}^{(j)} - \tilde{n}_{W_k,l}^{(j)} \right). \quad (\text{B.12})$$

Furthermore, for $\bar{n}_{W_k,l}^{(j)} \leq \tilde{n}_{W_k,l}^{(j)} \leq 0.5$, $\varepsilon \left(\tilde{n}_{W_k,l}^{(j)}, \bar{n}_{W_k,l}^{(j)} \right)$ is concave, and hence is bounded by the tangent at $\bar{n}_{W_k,l}^{(j)}$:

$$\varepsilon \left(\tilde{n}_{W_k,l}^{(j)}, \bar{n}_{W_k,l}^{(j)} \right) \leq \frac{1}{\bar{n}_{W_k,l}^{(j)}(1 - \bar{n}_{W_k,l}^{(j)})} \left(\tilde{n}_{W_k,l}^{(j)} - \bar{n}_{W_k,l}^{(j)} \right). \quad (\text{B.13})$$

Finally, for the case of $\tilde{n}_{W_k,l}^{(j)} \geq 0.5$, consider the line,

$$h \left(\tilde{n}_{W_k,l}^{(j)}, \bar{n}_{W_k,l}^{(j)} \right) := \frac{\left| B' - \phi_{B'}^{-1} \left(\bar{n}_{W_k,l}^{(j)} \right) \right|}{\left| \frac{1}{1+\exp(-B')} - \bar{n}_{W_k,l}^{(j)} \right|} \left(\tilde{n}_{W_k,l}^{(j)} - \bar{n}_{W_k,l}^{(j)} \right). \quad (\text{B.14})$$

From the convexity of $\varepsilon \left(\tilde{n}_{W_k,l}^{(j)}, \bar{n}_{W_k,l}^{(j)} \right)$, $h \left(\tilde{n}_{W_k,l}^{(j)}, \bar{n}_{W_k,l}^{(j)} \right)$ upper bounds $\varepsilon \left(\tilde{n}_{W_k,l}^{(j)}, \bar{n}_{W_k,l}^{(j)} \right)$ for $\tilde{n}_{W_k,l}^{(j)} \geq 0.5$, since $h \left(0.5, \bar{n}_{W_k,l}^{(j)} \right) \geq \varepsilon \left(0.5, \bar{n}_{W_k,l}^{(j)} \right)$ for $\bar{n}_{W_k,l}^{(j)} \leq 0.5$. Combining the piece-wise bounds in Eq. (B.12), Eq. (B.13) and Eq. (B.14), we conclude that for $\bar{n}_{W_k,l}^{(j)} \leq 0.5$:

$$\varepsilon \left(\tilde{n}_{W_k,l}^{(j)}, \bar{n}_{W_k,l}^{(j)} \right) \leq \tilde{g} \left(\bar{n}_{W_k,l}^{(j)}, B' \right) \left| \tilde{n}_{W_k,l}^{(j)} - \bar{n}_{W_k,l}^{(j)} \right|, \quad (\text{B.15})$$

where

$$\tilde{g} \left(\bar{n}_{W_k,l}^{(j)}, B' \right) = \max \left\{ \frac{1}{\bar{n}_{W_k,l}^{(j)}(1 - \bar{n}_{W_k,l}^{(j)})}, \frac{\left| B' + \phi_{B'}^{-1} \left(\bar{n}_{W_k,l}^{(j)} \right) \right|}{\left| \bar{n}_{W_k,l}^{(j)} - \frac{1}{1+\exp(B')} \right|}, \frac{\left| B' - \phi_{B'}^{-1} \left(\bar{n}_{W_k,l}^{(j)} \right) \right|}{\left| \frac{1}{1+\exp(-B')} - \bar{n}_{W_k,l}^{(j)} \right|} \right\}.$$

Due to the symmetry of $\varepsilon \left(\tilde{n}_{W_k,l}^{(j)}, \bar{n}_{W_k,l}^{(j)} \right)$, the same bound in Eq. (B.15) can be established for $\bar{n}_{W_k,l}^{(j)} > 0.5$ as well.

Then, using $\left| \phi_{B'}^{-1} \left(\bar{n}_{W_k,l}^{(j)} \right) \right| \leq B'$ and conditioning on event $\mathcal{A}_W$, we simplify this bound as:

$$\tilde{g} \left(\bar{n}_{W_k,l}^{(j)}, B' \right) \leq \max \left\{ 4 (1 + \exp(B))^2, \frac{4B' (1 + \exp(B')) (1 + \exp(B))}{\exp(B') - (2 \exp(B) + 1)} \right\}.$$

Finally, based on the fact that $B' = 2 \log(2 \exp(B) + 1)$, the latter is further upper bounded as:

$$\tilde{g} \left(\bar{n}_{W_k,l}^{(j)}, B' \right) \leq g(B),$$

where

$$g(B) = \max \left\{ 4 (1 + \exp(B))^2, 4 \exp(-B) \log(2 \exp(B) + 1) (1 + (2 \exp(B) + 1)^2) \right\}.$$

This concludes the proof of Lemma B.2. □

Following Lemma B.2, by conditioning on the event $\mathcal{A}_W$ we have:

$$\mathbb{E}_{\mathcal{A}_W} \left[\left| \phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(j)} \right) - \phi_{B'}^{-1} \left(\bar{n}_{W_k,l}^{(j)} \right) \right| \right] \leq g(B) \mathbb{E}_{\mathcal{A}_W} \left[\left| \tilde{n}_{W_k,l}^{(j)} - \bar{n}_{W_k,l}^{(j)} \right| \right]. \quad (\text{B.16})$$

Then, we note that:

$$\begin{aligned} \mathbb{E} \left[\left| \tilde{n}_{W_k, l}^{(j)} - \bar{n}_{W_k, l}^{(j)} \right| \right] &\stackrel{(c)}{\leq} \sqrt{\mathbb{E} \left[\left| \tilde{n}_{W_k, l}^{(j)} - \bar{n}_{W_k, l}^{(j)} \right|^2 \right]} \\ &\stackrel{(d)}{=} \frac{\sigma_w \sqrt{1 + \alpha^2}}{\sqrt{W}}, \end{aligned} \quad (\text{B.17})$$

wherein (c) we have used the Cauchy-Schwarz inequality, and in (d) we have used the fact that the observation noise across the W time instances is i.i.d. and white. From the bounds in Eq. (B.16) and Eq. (B.17), we conclude that the first expectation in Eq. (B.11), conditioned on event $\mathcal{A}_W$ is bounded as:

$$\begin{aligned} \mathbb{E}_{\mathcal{A}_W} \left[\left| \phi_{B'}^{-1} \left(\tilde{n}_{W_k, l}^{(j)} \right) - \phi_{B'}^{-1} \left(\bar{n}_{W_k, l}^{(j)} \right) \right| \right] &\leq g(B) \mathbb{E}_{\mathcal{A}_W} \left[\left| \tilde{n}_{W_k, l}^{(j)} - \bar{n}_{W_k, l}^{(j)} \right| \right] \\ &\leq g(B) \frac{\sigma_w \sqrt{1 + \alpha^2}}{\sqrt{W} \mathbb{P}(\mathcal{A}_W)}. \end{aligned} \quad (\text{B.18})$$

The foregoing sequence of reasoning similarly follows for $\mathbb{E} \left[\left| \phi_{B'}^{-1} \left(\bar{n}_{W_k, l}^{(j)} \right) - x_{W_k, l}^{(j)} \right| \right]$, since $\frac{1}{1+\exp(B)} \leq \phi \left(x_{W_k, l}^{(j)} \right) \leq 1 - \frac{1}{1+\exp(B)}$ for $k = 1, \dots, K$, $l = 1, \dots, L$ and $j = 1, \dots, N$ (as a consequence of $|x_{W_k, l}^{(j)}| < B$ for $k = 1, \dots, K$, $l = 1, \dots, L$ and $j = 1, \dots, N$, conditioned on $\mathcal{A}_W$). Accordingly, we derive the upper bound on the second term in Eq. (B.11), conditioned on

event $\mathcal{A}_W$:

$$\begin{aligned}
\mathbb{E}_{\mathcal{A}_W} \left[\left| \phi_{B'}^{-1} \left(\bar{n}_{W_k,l}^{(j)} \right) - x_{W_k,l}^{(j)} \right| \right] &\leq g(B) \mathbb{E}_{\mathcal{A}_W} \left[\left| \bar{n}_{W_k,l}^{(j)} - \phi \left(x_{W_k,l}^{(j)} \right) \right| \right] \\
&\stackrel{(e)}{\leq} \frac{g(B)}{W \mathbb{P}(\mathcal{A}_W)} \sqrt{\mathbb{E} \left[\left(\sum_{w=1}^W n_{(k-1)W+w,l}^{(j)} - W \phi \left(x_{W_k,l}^{(j)} \right) \right)^2 \right]} \\
&\stackrel{(f)}{=} \frac{g(B)}{W \mathbb{P}(\mathcal{A}_W)} \sqrt{\mathbb{E} \left[W \phi \left(x_{W_k,l}^{(j)} \right) \left(1 - \phi \left(x_{W_k,l}^{(j)} \right) \right) \right]} \\
&\stackrel{(g)}{\leq} \frac{g(B)}{2\sqrt{W} \mathbb{P}(\mathcal{A}_W)}, \tag{B.19}
\end{aligned}$$

where (e) follows from the application of Jensen's inequality, (f) follows from the formula for the variance of a Binomial random variable, and (g) follows from the inequality $\phi \left(x_{W_k,l}^{(j)} \right) \times \left(1 - \phi \left(x_{W_k,l}^{(j)} \right) \right) \leq 1/4$, for $\phi \left(x_{W_k,l}^{(j)} \right) \in [0, 1]$. Combining the results in Eq. (B.18) and Eq. (B.19), the overall expectation in Eq. (B.11), conditioned on the event $\mathcal{A}_W$ is upper-bounded by:

$$\mathbb{E}_{\mathcal{A}_W} \left[\left| \phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(j)} \right) - x_{W_k,l}^{(j)} \right| \right] \leq \frac{2g(B)}{\sqrt{W}} \left(\sigma_w \sqrt{1 + \alpha^2} + \frac{1}{2} \right), \tag{B.20}$$

where we have lower bounded the probability of the event $\mathcal{A}_W$ by $1/2$ (that is, $\mathbb{P}(\mathcal{A}_W) > 1/2$).

Thus, from Eq. (B.10) and Eq. (B.20) we derive:

$$\begin{aligned}
\left| \text{bias}_{\mathcal{A}_W} \left(\tilde{d}_j \right) \right| &\leq \frac{2g(B)}{\sqrt{W}} \left(\sigma_w \sqrt{1 + \alpha^2} + \frac{1}{2} \right) \frac{\sum_{k,l=1}^{K,L} |s_{W_k}|}{L \sum_{k=1}^K s_{W_k}^2} \\
&\stackrel{(h)}{\leq} \frac{2g(B)}{\sigma_s \sqrt{W}} \left(\sigma_w \sqrt{1 + \alpha^2} + \frac{1}{2} \right),
\end{aligned}$$

where in (h) we have used the Cauchy-Schwarz inequality $\sum_{k=1}^K |s_{W_k}| \leq \sqrt{K} \sqrt{\sum_{k=1}^K s_{W_k}^2}$ while defining $\sigma_s^2 := \frac{1}{K} \sum_{k=1}^K s_{W_k}^2$.

Then, for $B \geq 2.5$, we have $g(B) = 4(1 + \exp(B))^2$ and $B' = 2 \log(2 \exp(B) + 1) \leq 3B$.

Let $B := \sigma_m \sqrt{8q \log W}$ for some $q > \frac{1}{64}$. Further, for some $\epsilon < 1/2$, suppose that:

$$\log W \geq \max \left\{ \frac{\log(8KLN/\eta)}{q}, \frac{32\sigma_m^2 q}{\epsilon^2}, \frac{2\log(64q)}{1-2\epsilon}, \frac{\max \{6.25, 4(\|\boldsymbol{\mu}_x\|_\infty + \max_{k,j} \{ |s_{W_k} d_j| \})^2\}}{8q\sigma_m^2}, \log 2 \right\}. \quad (\text{B.21})$$

Under these conditions,

$$\begin{aligned} g(B) &\leq 4 \left(1 + \exp(\sigma_m \sqrt{8q \log W}) \right)^2 \\ &\stackrel{(i)}{\leq} 16 \exp \left(2\sigma_m \sqrt{8q \log W} \right) \\ &\leq 16W^\epsilon, \end{aligned} \quad (\text{B.22})$$

where in (i) we have used the fact that $e^x \geq 1$ for $x \geq 0$. Thus, under the conditions in Eq.

(B.21), we have:

$$\left| \text{bias}_{\mathcal{A}_W}(\tilde{d}_j) \right| \leq \frac{32}{\sigma_s \sqrt{W^{1-2\epsilon}}} \left(\sigma_w \sqrt{1 + \alpha^2} + \frac{1}{2} \right). \quad (\text{B.23})$$

Finally, from Eq. (B.8) and Eq. (B.23), we conclude that:

$$\left| \text{bias}_{\mathcal{A}_W}(\hat{d}_j) \right| \leq \frac{1}{\sqrt{W^{1-2\epsilon}}} C_1 \left(2\sigma_w \sqrt{1 + \alpha^2} + 1 \right) + \mathcal{O}(\sigma_w^2) + \mathcal{O}\left(\frac{1}{W}\right),$$

where $C_1 := \frac{16}{\sigma_s}$.

Bounding the variance of $\widehat{d}_j$

Next, we prove the upper bound on the variance of the maximum likelihood estimator, $\widehat{d}_j$.

To that end, we upper-bound the variance of $\widetilde{d}_j$. First, using the Cauchy-Schwarz inequality, we have:

$$\text{Var}(\widetilde{d}_j) := \mathbb{E} \left[\left| \widetilde{d}_j - \mathbb{E}[\widetilde{d}_j] \right|^2 \right] \leq \left\{ \sqrt{\mathbb{E} \left[\left| \widetilde{d}_j - (d_{\text{Oracle}})_j \right|^2 \right]} + \sqrt{\text{Var}((d_{\text{Oracle}})_j)} \right\}^2. \quad (\text{B.24})$$

Then, we upper-bound the conditional second moment of $\left| \widetilde{d}_j - (d_{\text{Oracle}})_j \right|$ using the same techniques as we used in bounding the first moment. Accordingly, we get:

$$\begin{aligned} \mathbb{E}_{\mathcal{A}_W} \left[\left| \widetilde{d}_j - (d_{\text{Oracle}})_j \right|^2 \right] &= \frac{1}{\left(L \sum_{k=1}^K s_{W_k}^2 \right)^2} \mathbb{E}_{\mathcal{A}_W} \left[\left| \sum_{k,l=1}^{K,L} s_{W_k} \left(\phi_{B'}^{-1} \left(\widetilde{n}_{W_k,l}^{(j)} \right) - x_{W_k,l}^{(j)} \right) \right|^2 \right] \\ &\stackrel{(k)}{\leq} \frac{1}{\left(L \sum_{k=1}^K s_{W_k}^2 \right)^2} \left\{ \sum_{k,l=1}^{K,L} |s_{W_k}| \sqrt{\mathbb{E}_{\mathcal{A}_W} \left[\left| \phi_{B'}^{-1} \left(\widetilde{n}_{W_k,l}^{(j)} \right) - x_{W_k,l}^{(j)} \right|^2 \right]} \right\}^2 \\ &\stackrel{(l)}{\leq} \left\{ \frac{\sqrt{2}g(B)}{\sigma_s \sqrt{W}} \left(\sigma_w \sqrt{1 + \alpha^2} + \frac{1}{2} \right) \right\}^2 \end{aligned} \quad (\text{B.25})$$

where in (k) we have used the Cauchy-Schwarz inequality and (l) follows from

$\mathbb{E}_{\mathcal{A}_W} \left[\left| \phi_{B'}^{-1} \left(\widetilde{n}_{W_k,l}^{(j)} \right) - x_{W_k,l}^{(j)} \right|^2 \right] \leq \frac{2(g(B))^2}{W} \left(\sigma_w \sqrt{1 + \alpha^2} + \frac{1}{2} \right)^2$, which can be proven by the same techniques as before.

Next, we note that the variance of the Oracle estimator $(d_{\text{Oracle}})_j$:

$$\begin{aligned}\text{Var}((d_{\text{Oracle}})_j) &= \frac{1}{\left(L \sum_{k=1}^K s_{W_k}^2\right)^2} \sum_{k,l=1}^{K,L} s_{W_k}^2 \text{Var}\left(\left(x_{W_k,l}^{(j)} - \mu_x^{(j)}\right)\right) \\ &= \frac{(\Sigma_x)_{j,j}}{L \sum_{k=1}^K s_{W_k}^2} = \frac{(\Sigma_x)_{j,j}}{LK\sigma_s^2}\end{aligned}\tag{B.26}$$

Combining Eq. (B.24), Eq. (B.25) and Eq. (B.26), we can upper-bound the conditional variance of $\tilde{d}_j$ as follows, based on Eq. (B.24):

$$\sqrt{\text{Var}_{\mathcal{A}_W}(\tilde{d}_j)} \leq \sqrt{\frac{(\Sigma_x)_{j,j}}{KL\sigma_s^2(1-\eta)}} + \frac{\sqrt{2}g(B)}{\sigma_s\sqrt{W}} \left(\sigma_w\sqrt{1+\alpha^2} + \frac{1}{2}\right)$$

Then, following Eq. (B.8), under the conditions for W in Eq. (B.21), we conclude the proof of the conditional variance of $\hat{d}_j$:

$$\begin{aligned}\sqrt{\text{Var}_{\mathcal{A}_W}(\hat{d}_j)} &\leq \sqrt{\frac{(\Sigma_x)_{j,j}}{KL\sigma_s^2(1-\eta)}} + \frac{1}{\sqrt{W^{1-2\epsilon}}} C_2 \left(2\sigma_w\sqrt{1+\alpha^2} + 1\right) \\ &\quad + \mathcal{O}(\sigma_w^2) + \mathcal{O}\left(\frac{1}{W}\right),\end{aligned}$$

where $C_2 := \frac{8\sqrt{2}}{\sigma_s}$.

Bounding the bias of $(\widehat{\Sigma}_x)_{i,j}$

Next, following the foregoing techniques, we upper-bound the bias and variance of the noise covariance estimator $(\widehat{\Sigma}_x)_{i,j}$. To that end, we first note:

$$\begin{aligned} \left| \text{bias} \left((\widetilde{\Sigma}_x)_{i,j} \right) \right| &:= \left| \mathbb{E} \left[(\widetilde{\Sigma}_x)_{i,j} \right] - (\Sigma_x)_{i,j} \right| \\ &\stackrel{(m)}{\leq} \left| \mathbb{E} \left[(\widetilde{\Sigma}_x)_{i,j} - (\Sigma_{\text{Oracle}})_{i,j} \right] \right| + \left| \text{bias} \left((\Sigma_{\text{Oracle}})_{i,j} \right) \right| \end{aligned} \quad (\text{B.27})$$

where (m) follows from the triangle inequality, with the Oracle noise covariance estimator (i.e., observing $\mathbf{x}_{t,l}$ directly), being defined as: $(\Sigma_{\text{Oracle}})_{i,j} = \frac{1}{KL} \sum_{k,l=1}^{K,L} \left(x_{W_k,l}^{(i)} - \mu_x^{(i)} - s_{W_k}(d_{\text{Oracle}})_i \right) \times \left(x_{W_k,l}^{(j)} - \mu_x^{(j)} - s_{W_k}(d_{\text{Oracle}})_j \right)$. Then, to simplify the first term in Eq. (B.27), we use similar techniques as before. Accordingly,

$$\begin{aligned} &\left| \mathbb{E} \left[(\widetilde{\Sigma}_x)_{i,j} - (\Sigma_{\text{Oracle}})_{i,j} \right] \right| \\ &= \left| \mathbb{E} \left[\frac{1}{KL} \sum_{k,l=1}^{K,L} \left(\phi_{B'}^{-1} \left(\widetilde{n}_{W_k,l}^{(i)} \right) - \mu_x^{(i)} - s_{W_k} \widetilde{d}_i \right) \left(\phi_{B'}^{-1} \left(\widetilde{n}_{W_k,l}^{(j)} \right) - \mu_x^{(j)} - s_{W_k} \widetilde{d}_j \right) \right. \right. \\ &\quad \left. \left. - \frac{1}{KL} \sum_{k,l=1}^{K,L} \left(x_{W_k,l}^{(i)} - \mu_x^{(i)} - s_{W_k}(d_{\text{Oracle}})_i \right) \left(x_{W_k,l}^{(j)} - \mu_x^{(j)} - s_{W_k}(d_{\text{Oracle}})_j \right) \right] \right| \\ &\stackrel{(n)}{\leq} \frac{1}{KL} \sum_{k,l=1}^{K,L} \mathbb{E} \left[\left| \left(\phi_{B'}^{-1} \left(\widetilde{n}_{W_k,l}^{(i)} \right) - \mu_x^{(i)} \right) \left(\phi_{B'}^{-1} \left(\widetilde{n}_{W_k,l}^{(j)} \right) - \mu_x^{(j)} \right) - \left(x_{W_k,l}^{(i)} - \mu_x^{(i)} \right) \left(x_{W_k,l}^{(j)} - \mu_x^{(j)} \right) \right| \right] \\ &\quad + \frac{1}{KL^2 \sum_{k=1}^K s_{W_k}^2} \mathbb{E} \left[\left| \sum_{k,l=1}^{K,L} s_{W_k} \left(\phi_{B'}^{-1} \left(\widetilde{n}_{W_k,l}^{(i)} \right) - \mu_x^{(i)} \right) \sum_{k',l'=1}^{K,L} s_{W_{k'}} \left(\phi_{B'}^{-1} \left(\widetilde{n}_{W_{k'},l'}^{(j)} \right) - \mu_x^{(j)} \right) \right. \right. \\ &\quad \left. \left. - \sum_{k,l=1}^{K,L} s_{W_k} \left(x_{W_k,l}^{(i)} - \mu_x^{(i)} \right) \sum_{k',l'=1}^{K,L} s_{W_{k'}} \left(x_{W_{k'},l'}^{(j)} - \mu_x^{(j)} \right) \right| \right], \end{aligned} \quad (\text{B.28})$$

where (n) follows through the application of Jensen's inequality and triangle inequality. Next, we have:

$$\begin{aligned}
& \mathbb{E}_{\mathcal{A}_W} \left[\left| \left(\phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(i)} \right) - \mu_x^{(i)} \right) \left(\phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(j)} \right) - \mu_x^{(j)} \right) - \left(x_{W_k,l}^{(i)} - \mu_x^{(i)} \right) \left(x_{W_k,l}^{(j)} - \mu_x^{(j)} \right) \right| \right] \\
& \leq \mathbb{E}_{\mathcal{A}_W} \left[\left| \phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(i)} \right) \phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(j)} \right) - x_{W_k,l}^{(i)} x_{W_k,l}^{(j)} \right| \right] + \mu_x^{(j)} \mathbb{E}_{\mathcal{A}_W} \left[\left| \phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(i)} \right) - x_{W_k,l}^{(i)} \right| \right] \\
& \quad + \mu_x^{(i)} \mathbb{E}_{\mathcal{A}_W} \left[\left| \phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(j)} \right) - x_{W_k,l}^{(j)} \right| \right] \\
& \leq \mathbb{E}_{\mathcal{A}_W} \left[\left| \phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(i)} \right) \phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(j)} \right) - \phi_{B'}^{-1} \left(\bar{n}_{W_k,l}^{(i)} \right) \phi_{B'}^{-1} \left(\bar{n}_{W_k,l}^{(j)} \right) \right| \right] \\
& \quad + \mathbb{E}_{\mathcal{A}_W} \left[\left| \phi_{B'}^{-1} \left(\bar{n}_{W_k,l}^{(i)} \right) \phi_{B'}^{-1} \left(\bar{n}_{W_k,l}^{(j)} \right) - x_{W_k,l}^{(i)} x_{W_k,l}^{(j)} \right| \right] \\
& \quad + 2\mu_m \mathbb{E}_{\mathcal{A}_W} \left[\left| \phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(j)} \right) - x_{W_k,l}^{(j)} \right| \right] \\
& \leq 2g(B) \frac{\sigma_w \sqrt{1+\alpha^2}}{\sqrt{W}} \left(g(B) \frac{\sigma_w \sqrt{1+\alpha^2}}{\sqrt{W}} + 4 \log(2 \exp(B) + 1) \right) \\
& \quad + \frac{2g(B)}{\sqrt{W}} \left\{ \frac{g(B)}{4\sqrt{W}} + B \right\} + \frac{4\mu_m g(B)}{\sqrt{W}} \left(\sigma_w \sqrt{1+\alpha^2} + \frac{1}{2} \right), \tag{B.29}
\end{aligned}$$

where $\mu_m = \|\boldsymbol{\mu}_x\|_\infty$ and we have used $B' = 2 \log(2 \exp(B) + 1)$. Similarly, the second term in Eq. (B.28) can be bounded as:

$$\begin{aligned}
& \mathbb{E}_{\mathcal{A}_W} \left[\left| \sum_{k,l=1}^{K,L} s_{W_k} \left(\phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(i)} \right) - \mu_x^{(i)} \right) \sum_{k',l'=1}^{K,L} s_{W_{k'}} \left(\phi_{B'}^{-1} \left(\tilde{n}_{W_{k'},l'}^{(j)} \right) - \mu_x^{(j)} \right) \right. \right. \\
& \quad \left. \left. - \sum_{k,l=1}^{K,L} s_{W_k} \left(x_{W_k,l}^{(i)} - \mu_x^{(i)} \right) \sum_{k',l'=1}^{K,L} s_{W_{k'}} \left(x_{W_{k'},l'}^{(j)} - \mu_x^{(j)} \right) \right| \right] \\
& \leq \sum_{k,k',l,l'=1}^{K,K,L,L} |s_{W_k} s_{W_{k'}}| \mathbb{E}_{\mathcal{A}_W} \left[\left| \phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(i)} \right) \phi_{B'}^{-1} \left(\tilde{n}_{W_{k'},l'}^{(j)} \right) - x_{W_k,l}^{(i)} x_{W_{k'},l'}^{(j)} \right| \right] \\
& \quad + \mu_x^{(i)} \sum_{k,k',l,l'=1}^{K,K,L,L} |s_{W_k} s_{W_{k'}}| \mathbb{E}_{\mathcal{A}_W} \left[\left| \phi_{B'}^{-1} \left(\tilde{n}_{W_{k'},l'}^{(j)} \right) - x_{W_{k'},l'}^{(j)} \right| \right] \\
& \quad + \mu_x^{(j)} \sum_{k,k',l,l'=1}^{K,K,L,L} |s_{W_k} s_{W_{k'}}| \mathbb{E}_{\mathcal{A}_W} \left[\left| \phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(i)} \right) - x_{W_k,l}^{(i)} \right| \right] \tag{B.30}
\end{aligned}$$

$$\begin{aligned}
& \leq \left(L \sum_{k=1}^K |s_{W_k}| \right)^2 \left\{ 2g(B) \frac{\sigma_w \sqrt{1+\alpha^2}}{\sqrt{W}} \left(g(B) \frac{\sigma_w \sqrt{1+\alpha^2}}{\sqrt{W}} + 4 \log(2 \exp(B) + 1) \right) \right. \\
& \quad \left. + \frac{2g(B)}{\sqrt{W}} \left\{ \frac{g(B)}{4\sqrt{W}} + B \right\} + \frac{4\mu_m g(B)}{\sqrt{W}} \left(\sigma_w \sqrt{1+\alpha^2} + \frac{1}{2} \right) \right\}. \tag{B.31}
\end{aligned}$$

Then, by combining the bounds in Eq. (B.29) and Eq. (B.31) and using an instance of Cauchy-Schwarz inequality $\left(\sum_{k=1}^K |s_{W_k}| \right)^2 \leq K \sum_{k=1}^K s_{W_k}^2$, we see that the bound in Eq. (B.28) can be expressed as:

$$\begin{aligned}
& \left| \mathbb{E}_{\mathcal{A}_W} \left[(\tilde{\boldsymbol{\Sigma}}_x)_{i,j} - (\boldsymbol{\Sigma}_{\text{Oracle}})_{i,j} \right] \right| \\
& \leq 4g(B) \frac{\sigma_w \sqrt{1+\alpha^2}}{\sqrt{W}} \left(g(B) \frac{\sigma_w \sqrt{1+\alpha^2}}{\sqrt{W}} + 4 \log(2 \exp(B) + 1) + 2\mu_m \right) \\
& \quad + \frac{4g(B)}{\sqrt{W}} \left\{ \frac{g(B)}{4\sqrt{W}} + B + \mu_m \right\}. \tag{B.32}
\end{aligned}$$

Next, we see that the oracle estimator follows an Inverse Wishart distribution, that is

$KL\Sigma_{\text{Oracle}} \sim \text{InvWish}_N(\Sigma_x, KL - 1)$. Therefore, we get:

$$\mathbb{E}[\Sigma_{\text{Oracle}}] = \frac{(KL - 1)}{KL} \Sigma_x.$$

Thus, the bias of the oracle estimator is given by:

$$\left| \text{bias} \left((\Sigma_{\text{Oracle}})_{i,j} \right) \right| = \frac{1}{KL} \left| (\Sigma_x)_{i,j} \right|. \quad (\text{B.33})$$

Combining the results in Eq. (B.32) and Eq. (B.33), the bias of $(\tilde{\Sigma}_x)_{i,j}$ can be bounded as:

$$\begin{aligned} \left| \text{bias}_{\mathcal{A}_W} \left((\tilde{\Sigma}_x)_{i,j} \right) \right| &\leq \frac{|(\Sigma_x)_{i,j}|}{KL(1-\eta)} + 4g(B) \frac{\sigma_w \sqrt{1+\alpha^2}}{\sqrt{W}} (4 \log(2 \exp(B) + 1) + 2\mu_m) \\ &\quad + \frac{4g(B)}{\sqrt{W}} (B + \mu_m) + \mathcal{O} \left(\frac{g(B)^2}{W} \right). \end{aligned} \quad (\text{B.34})$$

Finally, under the conditions for W in Eq. (B.21), the latter inequality simplifies to:

$$\begin{aligned} \left| \text{bias}_{\mathcal{A}_W} \left((\tilde{\Sigma}_x)_{i,j} \right) \right| &\stackrel{(p)}{\leq} \frac{|(\Sigma_x)_{i,j}|}{KL(1-\eta)} + \frac{Bg(B)}{\sqrt{W}} \left(28\sigma_w \sqrt{1+\alpha^2} + 6 \right) + \mathcal{O} \left(\frac{g(B)^2}{W} \right) \\ &\stackrel{(q)}{\leq} \frac{|(\Sigma_x)_{i,j}|}{KL(1-\eta)} + 64\sigma_m \sqrt{\frac{2q \log W}{W^{1-2\epsilon}}} \left(14\sigma_w \sqrt{1+\alpha^2} + 3 \right) + \mathcal{O} \left(\frac{1}{W^{1-2\epsilon}} \right) \end{aligned} \quad (\text{B.35})$$

where in (p) we have used $2 \log(2 \exp(B) + 1) \leq 3B$ and $B > 2\mu_m$ and in (q) we have used

$Bg(B) \leq 16L^\epsilon \sigma_m \sqrt{8q \log L}$, which follows from Eq. (B.22). Thus, following Eq. (B.9) we

derive the bound on the bias of the maximum likelihood estimator:

$$\begin{aligned} \left| \text{bias}_{\mathcal{A}_W} \left((\hat{\Sigma}_x)_{i,j} \right) \right| &\leq \frac{|(\Sigma_x)_{i,j}|}{KL(1-\eta)} + \sqrt{\frac{\log W}{W^{1-2\epsilon}}} C_3 \left(14\sigma_w \sqrt{1+\alpha^2} + 3 \right) \\ &\quad + \mathcal{O}(\sigma_w^2) + \mathcal{O}\left(\frac{1}{W^{1-2\epsilon}}\right), \end{aligned}$$

where $C_3 := 64\sigma_m\sqrt{2q}$.

Bounding the variance of $(\hat{\Sigma}_x)_{i,j}$

Next, we establish an upper bound on the variance of the maximum likelihood estimator of the noise covariance. To that end, we upper-bound the variance of $(\tilde{\Sigma}_x)_{i,j}$. First, using the Cauchy-Schwarz inequality, we get:

$$\text{Var} \left((\tilde{\Sigma}_x)_{i,j} \right) := \mathbb{E} \left[\left| (\tilde{\Sigma}_x)_{i,j} - \mathbb{E} \left[(\tilde{\Sigma}_x)_{i,j} \right] \right|^2 \right] \leq \left\{ \sqrt{\mathbb{E} \left[\left| (\tilde{\Sigma}_x)_{i,j} - (\Sigma_{\text{Oracle}})_{i,j} \right|^2 \right]} + \sqrt{\text{Var} \left((\Sigma_{\text{Oracle}})_{i,j} \right)} \right\}^2. \quad (\text{B.36})$$

Then, we upper-bound the conditional second moment of $\left|(\tilde{\Sigma}_x)_{i,j} - (\Sigma_{\text{Oracle}})_{i,j}\right|$ using the same techniques used in bounding its first moment. Accordingly, we derive:

$$\begin{aligned}
& \mathbb{E}_{\mathcal{A}_W} \left[\left| (\tilde{\Sigma}_x)_{i,j} - (\Sigma_{\text{Oracle}})_{i,j} \right|^2 \right] \\
&= \frac{1}{K^2 L^2} \mathbb{E}_{\mathcal{A}_W} \left[\left\{ \sum_{k,l=1}^{K,L} \left(\left(\phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(i)} \right) - \mu_x^{(i)} - s_{W_k} \tilde{d}_i \right) \left(\phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(j)} \right) - \mu_x^{(j)} - s_{W_k} \tilde{d}_j \right) \right. \right. \right. \\
&\quad \left. \left. \left. - \left(x_{W_k,l}^{(i)} - \mu_x^{(i)} - s_{W_k} (d_{\text{Oracle}})_i \right) \left(x_{W_k,l}^{(j)} - \mu_x^{(j)} - s_{W_k} (d_{\text{Oracle}})_j \right) \right) \right\}^2 \right] \\
&\leq \frac{1}{K^2 L^2} \left\{ \sum_{k,l=1}^{K,L} \left\{ \mathbb{E}_{\mathcal{A}_W} \left[\left(\left(\phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(i)} \right) - \mu_x^{(i)} - s_{W_k} \tilde{d}_i \right) \left(\phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(j)} \right) - \mu_x^{(j)} - s_{W_k} \tilde{d}_j \right) \right. \right. \right. \right. \\
&\quad \left. \left. \left. - \left(x_{W_k,l}^{(i)} - \mu_x^{(i)} - s_{W_k} (d_{\text{Oracle}})_i \right) \left(x_{W_k,l}^{(j)} - \mu_x^{(j)} - s_{W_k} (d_{\text{Oracle}})_j \right) \right) \right]^2 \right\}^{\frac{1}{2}} \right\}^2 \tag{B.37}
\end{aligned}$$

where the last bound follows from the Cauchy-Schwarz inequality. Then, we derive:

$$\begin{aligned}
& \mathbb{E}_{\mathcal{A}_W} \left[\left\{ \left(\phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(i)} \right) - \mu_x^{(i)} - s_{W_k} \tilde{d}_i \right) \left(\phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(j)} \right) - \mu_x^{(j)} - s_{W_k} \tilde{d}_j \right) \right. \right. \\
& \quad \left. \left. - \left(x_{W_k,l}^{(i)} - \mu_x^{(i)} - s_{W_k} (d_{\text{Oracle}})_i \right) \left(x_{W_k,l}^{(j)} - \mu_x^{(j)} - s_{W_k} (d_{\text{Oracle}})_j \right) \right\}^2 \right] \\
&= \mathbb{E}_{\mathcal{A}_W} \left[\left\{ \left(\phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(i)} \right) - \mu_x^{(i)} - s_{W_k} \frac{1}{L \sum_{k'=1}^K s_{W_{k'}}^2} \sum_{k',l'=1}^{K,L} s_{W_{k'}} \left(\phi_{B'}^{-1} \left(\tilde{n}_{W_{k'},l'}^{(i)} \right) - \mu_x^{(i)} \right) \right) \right. \right. \\
& \quad \times \left(\phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(j)} \right) - \mu_x^{(j)} - s_{W_k} \frac{1}{L \sum_{k''=1}^K s_{W_{k''}}^2} \sum_{k'',l''=1}^{K,L} s_{W_{k''}} \left(\phi_{B'}^{-1} \left(\tilde{n}_{W_{k'',l''}}^{(j)} \right) - \mu_x^{(j)} \right) \right) \\
& \quad - \left(x_{W_k,l}^{(i)} - \mu_x^{(i)} - s_{W_k} \frac{1}{L \sum_{k'=1}^K s_{W_{k'}}^2} \sum_{k',l'=1}^{K,L} s_{W_{k'}} \left(x_{W_{k'},l'}^{(i)} - \mu_x^{(i)} \right) \right) \\
& \quad \times \left. \left(x_{W_k,l}^{(j)} - \mu_x^{(j)} - s_{W_k} \frac{1}{L \sum_{k''=1}^K s_{W_{k''}}^2} \sum_{k'',l''=1}^{K,L} s_{W_{k''}} \left(x_{W_{k'',l''}}^{(j)} - \mu_x^{(j)} \right) \right) \right\}^2 \right] \\
&= \mathbb{E}_{\mathcal{A}_W} \left[\left\{ \left(\left(\phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(i)} \right) - \mu_x^{(i)} \right) \left(\phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(j)} \right) - \mu_x^{(j)} \right) - \left(x_{W_k,l}^{(i)} - \mu_x^{(i)} \right) \left(x_{W_k,l}^{(j)} - \mu_x^{(j)} \right) \right) \right. \right. \\
& \quad - \frac{s_{W_k}}{L \sum_{k''=1}^K s_{W_{k''}}^2} \sum_{k'',l''=1}^{K,L} s_{W_{k''}} \left(\left(\phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(i)} \right) - \mu_x^{(i)} \right) \left(\phi_{B'}^{-1} \left(\tilde{n}_{W_{k'',l''}}^{(j)} \right) - \mu_x^{(j)} \right) \right) \\
& \quad - \left(x_{W_k,l}^{(i)} - \mu_x^{(i)} \right) \left(x_{W_{k'',l''}}^{(j)} - \mu_x^{(j)} \right) \Big) \\
& \quad - \frac{s_{W_k}}{L \sum_{k'=1}^K s_{W_{k'}}^2} \sum_{k',l'=1}^{K,L} s_{W_{k'}} \left(\left(\phi_{B'}^{-1} \left(\tilde{n}_{W_{k'},l'}^{(i)} \right) - \mu_x^{(i)} \right) \left(\phi_{B'}^{-1} \left(\tilde{n}_{W_k,l}^{(j)} \right) - \mu_x^{(j)} \right) \right) \\
& \quad - \left(x_{W_{k'},l'}^{(i)} - \mu_x^{(i)} \right) \left(x_{W_k,l}^{(j)} - \mu_x^{(j)} \right) \Big) \\
& \quad + \frac{s_{W_k}^2}{L^2 \sum_{k'=1}^K s_{W_{k'}}^2 \sum_{k''=1}^K s_{W_{k''}}^2} \sum_{k',k'',l',l''=1}^{K,K,L,L} s_{W_{k'}} s_{W_{k''}} \left(\left(\phi_{B'}^{-1} \left(\tilde{n}_{W_{k'},l'}^{(i)} \right) - \mu_x^{(i)} \right) \right. \\
& \quad \times \left. \left(\phi_{B'}^{-1} \left(\tilde{n}_{W_{k'',l''}}^{(j)} \right) - \mu_x^{(j)} \right) - \left(x_{W_{k'},l'}^{(i)} - \mu_x^{(i)} \right) \left(x_{W_{k'',l''}}^{(j)} - \mu_x^{(j)} \right) \right) \Big\}^2 \right] \\
&\leq \frac{2(g(B))^2}{W} \left(\sigma_w \sqrt{1 + \alpha^2} + \frac{1}{2} \right)^2 \left(\frac{g(B)}{\sqrt{W}} \left(\sigma_w \sqrt{1 + \alpha^2} + \frac{1}{2} \right) + 2(B + \mu_m) \right)^2 \\
& \quad \times \left(1 + \frac{s_{W_k} L \sum_{k'=1}^K s_{W_{k'}}}{L \sum_{k'=1}^K s_{W_{k'}}^2} + \frac{s_{W_k} L \sum_{k''=1}^K s_{W_{k''}}}{L \sum_{k''=1}^K s_{W_{k''}}^2} + \frac{s_{W_k}^2 L \sum_{k'=1}^K s_{W_{k'}} L \sum_{k''=1}^K s_{W_{k''}}}{L^2 \sum_{k'=1}^K s_{W_{k'}}^2 \sum_{k''=1}^K s_{W_{k''}}^2} \right)^2.
\end{aligned}$$

(B.38)

Using the final bound of Eq. (B.38) in Eq. (B.37), we get:

$$\begin{aligned}
& \sqrt{\mathbb{E}_{\mathcal{A}_W} \left[\left| (\tilde{\Sigma}_x)_{i,j} - (\Sigma_{\text{Oracle}})_{i,j} \right|^2 \right]} \\
& \leq \frac{\sqrt{2}g(B)}{\sqrt{W}} \left(\sigma_w \sqrt{1 + \alpha^2} + \frac{1}{2} \right) \left(\frac{g(B)}{\sqrt{W}} \left(\sigma_w \sqrt{1 + \alpha^2} + \frac{1}{2} \right) + 2(B + \mu_m) \right) \\
& \quad \times \frac{1}{KL} \left(KL + \frac{L \sum_{k=1}^K |s_{W_k}| \sum_{k'=1}^K |s_{W_{k'}}|}{\sum_{k'=1}^K s_{W_{k'}}^2} + \frac{L \sum_{k=1}^K |s_{W_k}| \sum_{k''=1}^K |s_{W_{k''}}|}{\sum_{k''=1}^K s_{W_{k''}}^2} \right. \\
& \quad \left. + \frac{L \sum_{k=1}^K s_{W_k}^2 \sum_{k'=1}^K |s_{W_{k'}}| \sum_{k''=1}^K |s_{W_{k''}}|}{\sum_{k'=1}^K s_{W_{k'}}^2 \sum_{k''=1}^K s_{W_{k''}}^2} \right) \\
& \leq \frac{4\sqrt{2}g(B)}{\sqrt{W}} \left(\sigma_w \sqrt{1 + \alpha^2} + \frac{1}{2} \right) \left(\frac{g(B)}{\sqrt{W}} \left(\sigma_w \sqrt{1 + \alpha^2} + \frac{1}{2} \right) + 2(B + \mu_m) \right), \quad (\text{B.39})
\end{aligned}$$

where the last inequality follows from an instance of the Cauchy-Schwarz inequality, i.e.,

$$\left(\sum_{k=1}^K |s_{W_k}| \right)^2 \leq K \sum_{k=1}^K s_{W_k}^2.$$

Then, following the observation $KL \Sigma_{\text{Oracle}} \sim \text{InvWish}_N(\Sigma_x, KL - 1)$, we derive the variance of $(\Sigma_{\text{Oracle}})_{i,j}$:

$$\text{Var} \left((\Sigma_{\text{Oracle}})_{i,j} \right) = \delta_{i,j}^2 = \frac{(KL - 1) \left((\Sigma_x)_{i,j}^2 + (\Sigma_x)_{i,i}(\Sigma_x)_{j,j} \right)}{K^2 L^2}. \quad (\text{B.40})$$

Combining Eq. (B.36), Eq. (B.39) and Eq. (B.40), we express the upper bound on the conditional variance of $(\tilde{\Sigma}_x)_{i,j}$ as:

$$\sqrt{\text{Var}_{\mathcal{A}_W} \left((\tilde{\Sigma}_x)_{i,j} \right)} \leq \frac{1}{\sqrt{1 - \eta}} \delta_{i,j} + \frac{8\sqrt{2}g(B)}{\sqrt{W}} (B + \mu_m) \left(\sigma_w \sqrt{1 + \alpha^2} + \frac{1}{2} \right) + \mathcal{O} \left(\frac{g(B)^2}{W} \right).$$

Then, following Eq. (B.9) and the conditions in Eq. (B.21), we conclude the proof of the upper

bound on the conditional variance of $(\hat{\Sigma}_x)_{i,j}$:

$$\begin{aligned} \sqrt{\text{Var}_{\mathcal{A}_W} \left((\hat{\Sigma}_x)_{i,j} \right)} &\leq \frac{1}{\sqrt{1-\eta}} \delta_{i,j} + \sqrt{\frac{\log W}{W^{1-2\epsilon}}} C_4 \left(2\sigma_w \sqrt{1+\alpha^2} + 1 \right) \\ &\quad + \mathcal{O}(\sigma_w^2) + \mathcal{O}\left(\frac{1}{W^{1-2\epsilon}}\right), \end{aligned}$$

where $C_4 := 384\sigma_m\sqrt{q}$.

Finally, it only remains to prove that the event $\mathcal{A}_W$ occurs with a high probability for sufficiently large W :

Lemma B.3. *The probability of occurrence of the event $\mathcal{A}_W = \left\{ \left| x_{W_k,l}^{(j)} \right| \leq B \text{ and } \frac{1}{2(1+\exp(B))} \leq \bar{n}_{W_k,l}^{(j)} \leq 1 - \frac{1}{2(1+\exp(B))} \text{ for } j = 1, \dots, N, k = 1, \dots, K \text{ and } l = 1, \dots, L \right\}$ is upper-bounded as follows:*

$$\mathbb{P}(\mathcal{A}_W) \geq 1 - \eta,$$

for some constant $0 < \eta \leq 1/2$ satisfying the conditions of Eq. (B.21).

Proof of Lemma B.3. First, using the union bound, we have:

$$\begin{aligned} \mathbb{P}(\mathcal{A}_W) &\geq 1 - \sum_{k,l,j=1}^{K,L,N} \left\{ \mathbb{P}\left(\left|x_{W_k,l}^{(j)}\right| > B\right) + \mathbb{P}\left(\bar{n}_{W_k,l}^{(j)} < \frac{1}{2(1+\exp(B))}\right) \right. \\ &\quad \left. + \mathbb{P}\left(\bar{n}_{W_k,l}^{(j)} > 1 - \frac{1}{2(1+\exp(B))}\right) \right\}. \end{aligned} \tag{B.41}$$

Next, we bound the probabilities on the right hand side using Chernoff's inequality [175]. First,

note that:

$$\begin{aligned}
\mathbb{P}\left(x_{W_k,l}^{(j)} > B\right) &= \mathbb{P}\left(x_{W_k,l}^{(j)} - \mu_x^{(j)} - s_{W_k} d_j > B - \mu_x^{(j)} - s_{W_k} d_j\right) \\
&\stackrel{(r)}{\leq} \mathbb{P}\left(x_{W_k,l}^{(j)} - \mu_x^{(j)} - s_{W_k} d_j > \frac{B}{2}\right) \\
&\stackrel{(s)}{\leq} \exp\left(-\frac{B^2}{8\sigma_m^2}\right),
\end{aligned}$$

where (r) follows if $B > 2(\|\boldsymbol{\mu}_x\|_\infty + \max_{k,j} \{|s_{W_k} d_j|\})$ (which will hold under the conditions in Eq. (B.21)) and (s) has been derived by applying the Chernoff's bound on the Gaussian random variable $x_{W_k,l}^{(j)}$. From the same reasoning we see that $\mathbb{P}\left(x_{W_k,l}^{(j)} < -B\right) \leq \exp\left(-\frac{B^2}{8\sigma_m^2}\right)$. Combining these two results, we get the upper bound:

$$\mathbb{P}\left(\left|x_{W_k,l}^{(j)}\right| > B\right) \leq 2 \exp\left(-\frac{B^2}{8\sigma_m^2}\right). \quad (\text{B.42})$$

Next, note that:

$$\begin{aligned}
\mathbb{P}\left(\bar{n}_{W_k,l}^{(j)} < \frac{1}{2(1 + \exp(B))}\right) &\stackrel{(u)}{\leq} \mathbb{P}\left(\bar{n}_{W_k,l}^{(j)} - \phi\left(x_{W_k,l}^{(j)}\right) < \frac{-1}{2(1 + \exp(B))}\right) \\
&\stackrel{(v)}{\leq} \exp\left(-\frac{W}{16(1 + \exp(B))^2}\right),
\end{aligned} \quad (\text{B.43})$$

where (u) follows from the observation $\frac{1}{1 + \exp(B)} < \phi\left(x_{W_k,l}^{(j)}\right)$ (which is a consequence of $|x_{W_k,l}^{(j)}| < B$). Then, we note that the zero-mean random variable $\bar{n}_{W_k,l}^{(j)} - \phi\left(x_{W_k,l}^{(j)}\right)$ is sub-Gaussian with variance factor $\frac{2}{W}$. Thus, using the Chernoff's inequality on sub-Gaussian random variables [175], we derive the upper-bound (v) in Eq. (B.43). In a similar fashion, based on the

observation $\phi\left(x_{W_k,l}^{(j)}\right) < 1 - \frac{1}{1+\exp(B)}$, we conclude the bound:

$$\mathbb{P}\left(\bar{n}_{W_k,l}^{(j)} > 1 - \frac{1}{2(1+\exp(B))}\right) \leq \exp\left(-\frac{W}{16(1+\exp(B))^2}\right). \quad (\text{B.44})$$

By combining the bounds in Eq. (B.42), Eq. (B.43) and Eq. (B.44), the upper bound on $\mathbb{P}(\mathcal{A}_W)$ in Eq. (B.41) takes the form:

$$\mathbb{P}(\mathcal{A}_W) \geq 1 - 2KLN \exp\left(-\frac{W}{16(1+\exp(B))^2}\right) - 2KLN \exp\left(-\frac{B^2}{8\sigma_m^2}\right).$$

Finally, under the assumptions in Eq. (B.21), we further simplify this bound as:

$$\mathbb{P}(\mathcal{A}_W) \geq 1 - 2KLN \exp\left(-\frac{W^{1-\epsilon}}{64}\right) - \frac{2KLN}{W^q} \geq 1 - \frac{4KLN}{W^q},$$

where we have used $W \geq 2$ (which gives $\log W \geq 2 \log \log W$) and $\log W \geq \frac{2 \log(64q)}{1-2\epsilon}$ to show that $\frac{W^{1-\epsilon}}{64} \geq q \log W$. Thus, $\log W \geq \frac{\log(8KLN/\eta)}{q}$ ensures that $\mathbb{P}(\mathcal{A}_W) \geq 1 - \eta$, for $0 < \eta \leq \frac{1}{2}$. □

This concludes the proof of Theorem 5.1. □

B.3 Adapting the proposed signal and noise correlation estimates to spiking observations

While Algorithm 5.1 takes two-photon fluorescence observations as input and produces estimates of signal and noise correlation as output, it is possible to adapt it to spiking observations

obtained by electrophysiology recordings. The resulting algorithm is obtained by simplifying the variational inference procedure in Algorithm 5.1 and is given below (Algorithm B.1) for completeness.

Algorithm B.1 Estimation of Σ_x and $\mathbf{D}$ from spiking observations

Inputs: Ensemble of spiking observations $\{\mathbf{n}_{t,l}\}_{t,l=1}^{T,L}$, constant μ_x , hyper-parameters ψ_x and ρ_x , tolerance at convergence δ and the external stimulus $\mathbf{s}_t$

Outputs: $\hat{\Sigma}_x$ and $\hat{\mathbf{D}}$

Initialization: Initial choice of $\mathbf{P}_x$, $\tilde{\Omega}_t$, $\hat{\Sigma}_x$ and $\hat{\mathbf{D}}$, residual = 10δ , $\gamma_x = \rho_x + LT$

- 1: **while** residual $\geq \delta$ **do**
- 2: **Update variational parameters**, for $t = 1, \dots, T$ and $l = 1, \dots, L$:

$$\begin{aligned} \mathbf{Q}_{\mathbf{x}_{t,l}} &= (\tilde{\Omega}_{t,l} + \gamma_x \mathbf{P}_x^{-1})^{-1} \\ \mathbf{m}_{\mathbf{x}_{t,l}} &= \mathbf{Q}_{\mathbf{x}_{t,l}} \left(\mathbf{n}_{t,l} - \frac{1}{2} \mathbf{1} - \tilde{\Omega}_{t,l} \hat{\mathbf{D}}^\top \mathbf{s}_t + \gamma_x \mathbf{P}_x^{-1} \mu_x \right) \\ c_{t,l}^{(j)} &= \sqrt{(\mathbf{Q}_{\mathbf{x}_{t,l}})_{j,j} + \left(m_{\mathbf{x}_{t,l}}^{(j)} + \hat{\mathbf{d}}_j^\top \mathbf{s}_t \right)^2} \quad (\text{for } j = 1, \dots, N) \\ (\tilde{\Omega}_{t,l})_{j,j} &:= \frac{1}{2c_{t,l}^{(j)}} \tanh \left(\frac{c_{t,l}^{(j)}}{2} \right) \quad (\text{for } j = 1, \dots, N) \\ \mathbf{P}_x &:= \psi_x + \sum_{t,l=1}^{T,L} \{ \mathbf{Q}_{\mathbf{x}_{t,l}} + \mathbf{m}_{\mathbf{x}_{t,l}} \mathbf{m}_{\mathbf{x}_{t,l}}^\top - \mu_x \mathbf{m}_{\mathbf{x}_{t,l}}^\top - \mathbf{m}_{\mathbf{x}_{t,l}} \mu_x^\top + \mu_x \mu_x^\top \} \end{aligned}$$

- 3: **Update outputs and the convergence criterion:**

$$\begin{aligned} \hat{\mathbf{d}}_j &= \left(\sum_{t,l=1}^{T,L} \left((\tilde{\Omega}_{t,l})_{j,j} \mathbf{s}_t \mathbf{s}_t^\top \right) \right)^{-1} \left(\sum_{t,l=1}^{T,L} \left\{ \left(n_{t,l}^{(j)} - \frac{1}{2} \right) \mathbf{s}_t - (\tilde{\Omega}_{t,l})_{j,j} m_{\mathbf{x}_{t,l}}^{(j)} \mathbf{s}_t \right\} \right) \\ (\hat{\mathbf{D}})_{\text{prev}} &= \hat{\mathbf{D}}, \quad \hat{\mathbf{D}} = [\hat{\mathbf{d}}_1, \hat{\mathbf{d}}_2, \dots, \hat{\mathbf{d}}_N] \\ (\hat{\Sigma}_x)_{\text{prev}} &= \hat{\Sigma}_x, \quad \hat{\Sigma}_x = \frac{\mathbf{P}_x}{\gamma_x + N + 1} \\ \text{residual} &= \|(\hat{\Sigma}_x)_{\text{prev}} - \hat{\Sigma}_x\|_2 / \|(\hat{\Sigma}_x)_{\text{prev}}\|_2 + \|(\hat{\mathbf{D}})_{\text{prev}} - \hat{\mathbf{D}}\|_2 / \|(\hat{\mathbf{D}})_{\text{prev}}\|_2 \end{aligned}$$

- 4: **end while**
 - 5: **Return** $\hat{\Sigma}_x$ and $\hat{\mathbf{D}}$
-

B.4 Supplementary material of the Simulation study 1 in Section 5.1.3 (Supplementary figures to Figure 5.2)

B.4.1 Sensitivity of two-stage estimates to the choice of the underlying spike deconvolution technique

In this section, we present a comparison between the two-stage estimates of the signal and noise correlations in the simulation study of Section 5.1.3, using two different spike deconvolution techniques: FCSS [20] and f-oopsi [22].

Figure B.1-A show the noise and signal correlation corresponding to the ground truth, estimated by the two-stage Pearson method using the FCSS [20] (second column) and constrained f-oopsi [22] (third column) spike deconvolution techniques, for the simulation study in Figure 5.2. The NMSE and leakage ratios of the estimates are indicated below each panel. While the correlation estimates based on these two methods are comparable, there exist notable differences between them, as a result of the slight discrepancies in the deconvolved spikes. This demonstrates that the two-stage estimates are notably sensitive to minor differences in the estimated spikes obtained by different deconvolution techniques. In addition, both two-stage Pearson estimates fail to capture the ground truth correlations (as is also evident from the high NMSE and leakage values).

Figure B.1-B shows the simulated fluorescence trace (black) for the first trial of neuron 1, along with the ground truth spikes (blue). The second and third rows, show the deconvolved spikes using the FCSS and constrained f-oopsie methods, respectively. The estimated spikes closely match each other, and the ground truth.

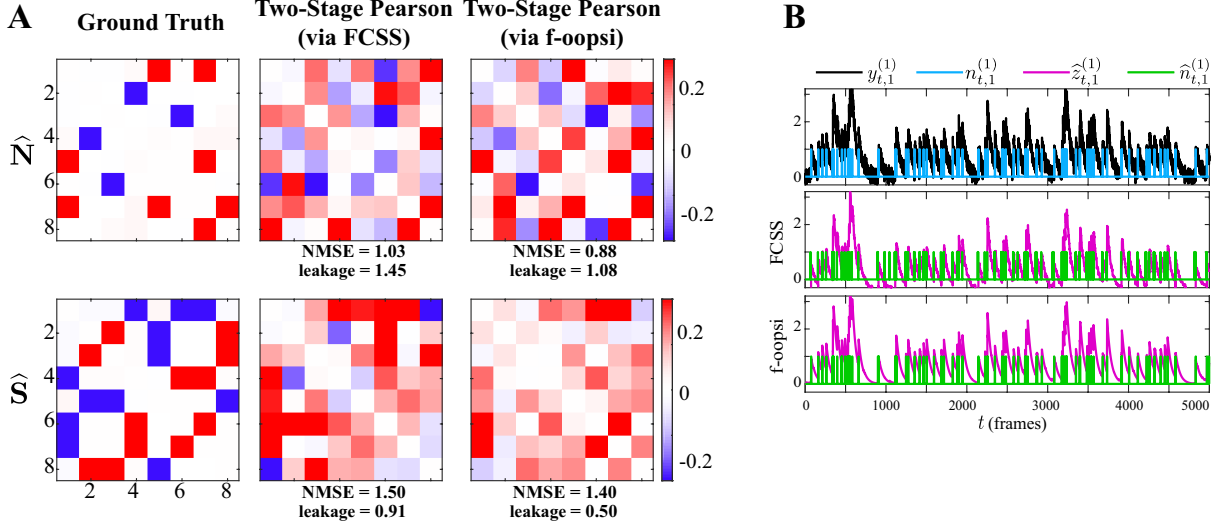


Figure B.1: Sensitivity of two-stage estimates to the choice of the underlying spike deconvolution technique: (A) Noise (first row) and signal (second row) correlations corresponding to the ground truth (first column), estimated by the two-stage Pearson method using the FCSS [20] (second column) and constrained f-oopsi [22] (third column) spike deconvolution techniques, for the simulation study in 5.2. The NMSE and leakage ratios of the estimates are indicated below each panel. (B) Simulated observations (black, re-scaled for ease of visual comparison) and ground truth spikes (blue), as well as the estimated calcium concentrations (purple) and putative spikes (green) for the 1st trial of neuron 1 in the simulation study of Figure 5.2, using the FCSS [20] (second row) and constrained f-oopsi [22] (third row) spike deconvolution methods.

B.4.2 Performance of two-stage estimates based on ground truth spikes

In this section, we apply the two-stage Pearson method on the ground truth spikes used in simulation study 1 (Section 5.1.3), as opposed to extracting the spikes via a deconvolution technique. Thus, these results isolate the effect of the non-linearities involved in spike generation on the estimation performance.

The first column of Figure B.2 shows the noise (first row) and signal (second row) correlations corresponding to the ground truth, which are repeated from Figure 5.2. The second and third columns show the results of two-stage GPFA and two-stage Pearson methods using $L = 20$ trials, respectively. The fourth column shows the results of the two-stage Pearson method using

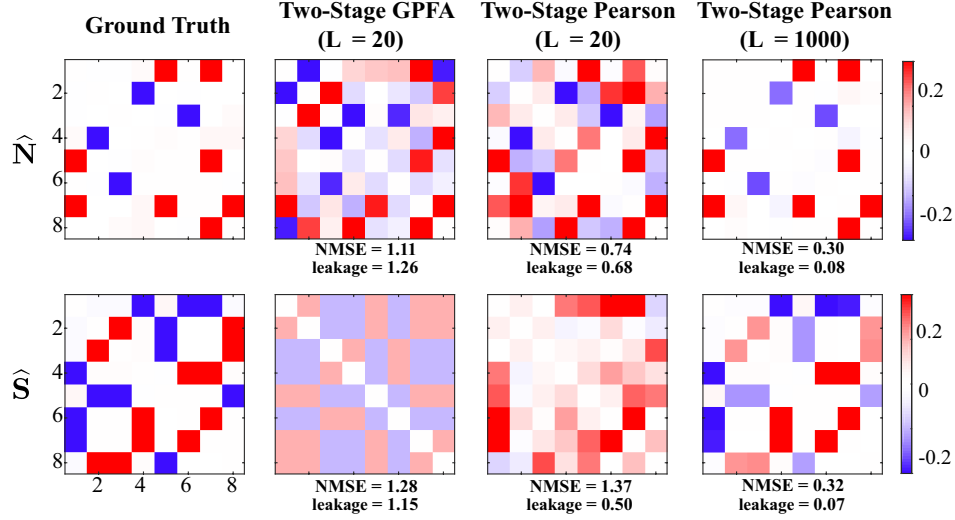


Figure B.2: Performance of two stage estimates based on ground truth spikes. Noise (first row) and signal (second row) correlations corresponding to the ground truth (first column) are repeated from Figure 5.2. The second and third columns show the results of two-stage GPFA and two-stage Pearson methods using $L = 20$ trials, respectively. The fourth column shows the results of the two-stage Pearson method using $L = 1000$ trials. All estimates were obtained using the ground truth spikes. The NMSE and leakage ratios of the estimates are indicated below each panel.

$L = 1000$ trials. All estimates were obtained using the ground truth spikes. The NMSE and leakage ratios of the estimates are indicated below each panel.

Even though the ground truth spikes are used, the NMSE and leakage ratios indicated in the second and third columns are remarkably high. This further shows that the usage of conventional definitions and GPFA estimates is not optimal for the recovery of signal and noise correlations. In accordance with our theoretical analysis in Appendix B.1, the performance of the two-stage Pearson method significantly improves as the number of trials is increased to $L = 1000$, a number that is unrealistic in the context of typical two-photon imaging experiments. However, our proposed method shown in Figure 5.2 achieves comparable performance with a number of trials as low as $L = 20$. In summary, these results suggest that the two-stage methods produce highly biased estimates under a limited number of trials, even if the ground truth spikes were ideally

deconvolved from the two-photon data.

B.4.3 Performance comparison under stimulus integration model mismatch

In this section we analyze the robustness of our estimates to stimulus integration model mismatch, in a setting similar to simulation study 1 (Section 5.1.3). Specifically, spikes were generated by replacing the linear receptive field model $\mathbf{d}_j^\top \mathbf{s}_t$ with a non-linear one given by $\mathbf{d}_j^\top \mathbf{s}_t + (\tilde{\mathbf{d}}_{j,1}^\top \mathbf{s}_t)^2 + (\tilde{\mathbf{d}}_{j,2}^\top \mathbf{s}_t)^2$, but a linear stimulus model was used for estimation (i.e., $\tilde{\mathbf{d}}_{j,1} = \tilde{\mathbf{d}}_{j,2} = 0$). Rows from left to right in Figure B.3 show the noise (first row) and signal (second row) correlations corresponding to the ground truth, proposed method, Pearson correlations from two-photon recordings, two-stage Pearson estimates and two-stage GPFA estimates. The NMSE and

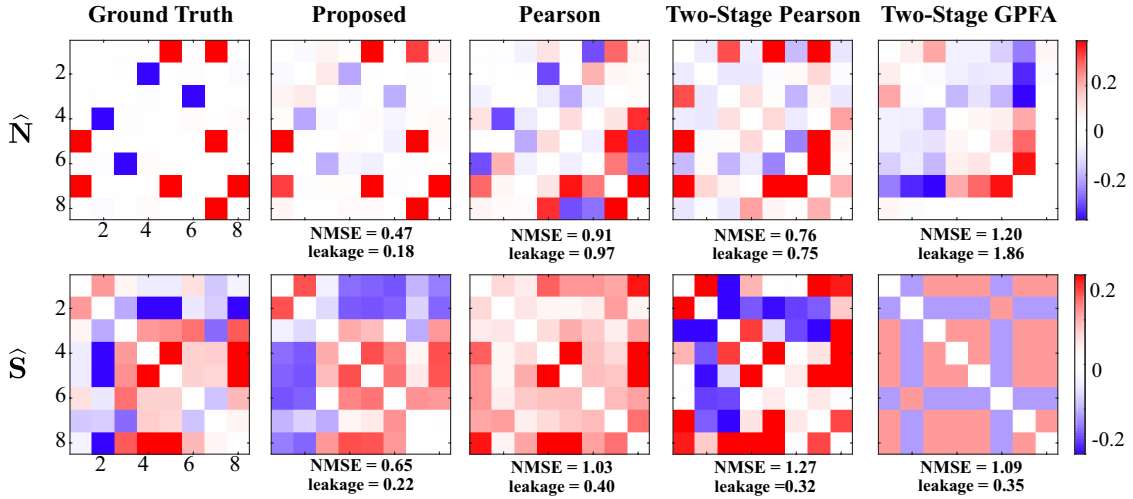


Figure B.3: Estimated noise and signal correlation matrices from different methods based on data generated with non-linear stimulus integration. Spikes were generated by replacing the linear receptive field model $\mathbf{d}_j^\top \mathbf{s}_t$ with a non-linear one given by $\mathbf{d}_j^\top \mathbf{s}_t + (\tilde{\mathbf{d}}_{j,1}^\top \mathbf{s}_t)^2 + (\tilde{\mathbf{d}}_{j,2}^\top \mathbf{s}_t)^2$, but a linear stimulus model was used for estimation (i.e., $\tilde{\mathbf{d}}_{j,1} = \tilde{\mathbf{d}}_{j,2} = 0$). Rows from left to right: ground truth, proposed method, Pearson correlations from two-photon recordings, two-stage Pearson estimates and two-stage GPFA estimates. The normalized mean squared error (NMSE) of each estimate relative to the ground truth and the leakage effect quantified by the ratio between out-of-network and in-network power (leakage) are indicated below each panel.

leakage ratios of the estimates are indicated below each panel.

While the NMSE in our proposed signal correlation estimates under this setting is greater than that in Figure 5.2 with no model mismatch, our proposed estimates still outperform existing methods. In addition, model mismatch in the stimulus integration component does not affect the accuracy of noise correlations estimated by our method.

B.4.4 Proposed estimates based on simulated data with model mismatch and at lower SNR

Here we illustrate the robustness of our proposed estimation procedure by applying our method to data generated by a mismatched model and at a lower SNR than the setting in Figure 5.2. Figure B.4-A shows the proposed noise and signal correlation estimates for data simulated at lower SNR than the setting of Figure 5.2 and model mismatch introduced by using a second-order auto-regressive model for the calcium decay. The ground truth correlations are the same as those in Figure 5.2. The NMSE and leakage ratio are given at the bottom.

Figure B.4-B shows the putative spikes (green) and estimated calcium concentrations (pur-

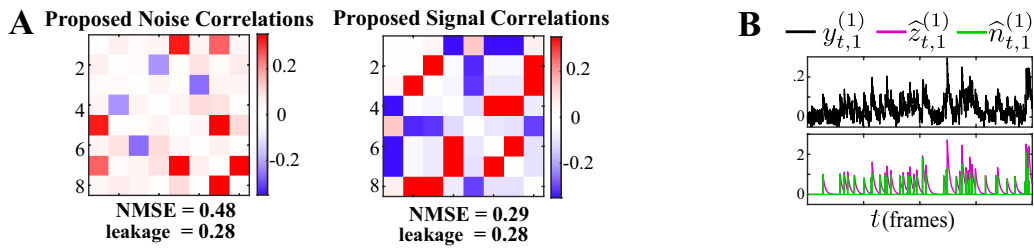


Figure B.4: (A) Proposed noise and signal correlation estimates for data simulated at lower SNR than the setting of Figure 5.2 and model mismatch introduced by using a second-order Auto-Regressive model for the calcium decay. The ground truth correlations are the same as those in Figure 5.2. The NMSE and leakage ratio are given at the bottom. (B) putative spikes (green) and estimated calcium concentrations (purple).

ple). Although the model mismatch and lower SNR result in slight performance degradation compared to Figure 5.2 (in terms of NMSE and leakage), and our method is capable of recovering the underlying correlations faithfully even under these settings.

B.4.5 Performance comparison under varying SNR levels and firing rates

Next, we evaluate the performance of our proposed estimates, Pearson correlations from two-photon data and the two-stage Pearson method under varying SNR levels and average firing rate settings, for the same signal and noise correlations shown in Figure 5.2. Figure B.5-A shows the NMSE (top) and the leakage ratios (bottom) for the noise (left) and signal (right) correlation estimates vs. SNR (in dB), for the proposed method, Pearson correlations from two-photon data and two-stage Pearson method. The SNR setting corresponding to Figure 5.2 is indicated by a dashed vertical line. The mean and standard deviation (std) of the normalized performance gain of the proposed method in comparison to the two existing methods are indicated as insets in each panel. Figure B.5-B illustrates the corresponding variations relative to different firing rate settings. While the performance of all methods degrades at low SNR levels or firing rates ($\text{SNR} < 10$ dB, firing rate < 0.5 Hz), our proposed method outperforms the existing methods for almost all SNR and firing rate settings considered. The sample time-domain signals shown in Figure B.5-C show that the latent underlying spiking activities are accurately recovered by our proposed method, which in turn facilitates robust and accurate recovering of the latent underlying signal and noise correlations.

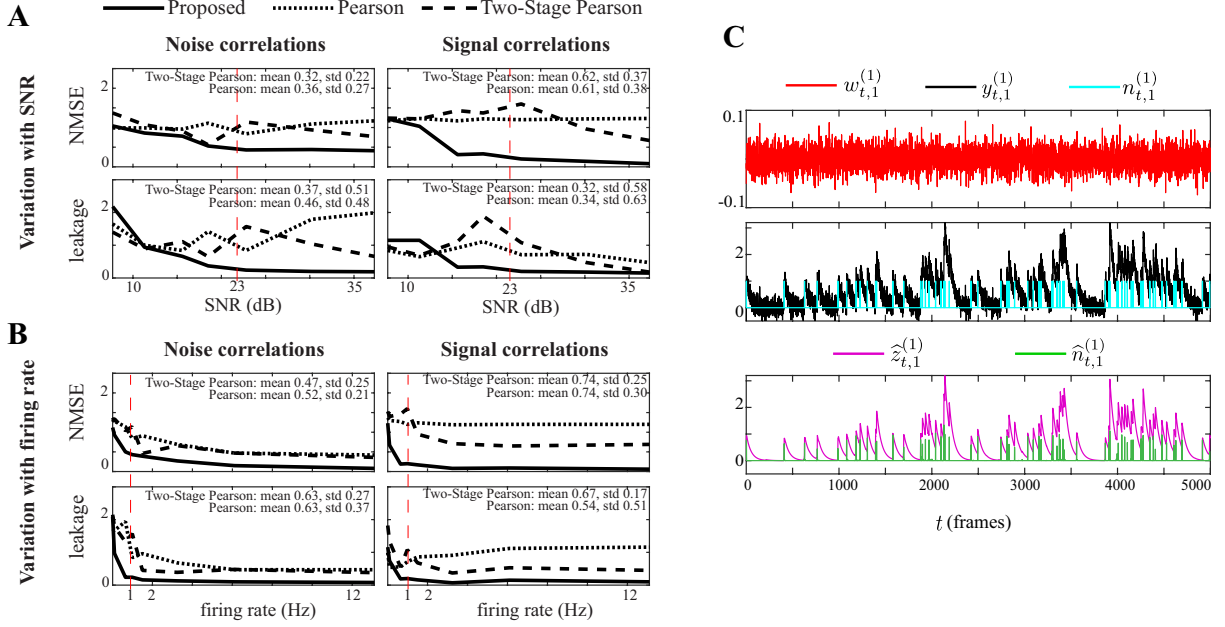


Figure B.5: Performance comparison at different SNR levels and average firing rate settings, in the first simulation study. (A) NMSE (top) and leakage ratios (bottom) for the noise (left) and signal (right) correlation estimates vs. SNR (in dB), for the proposed method, Pearson correlations from two-photon data and Two-stage Pearson method. The SNR setting corresponding to Figure 5.2 is indicated by a dashed vertical line. The mean and standard deviation (std) of the normalized performance gain of the proposed method in comparison to the two existing methods are indicated as insets in each panel. (B) Same organization as panel A, but with respect to varying firing rates (in Hz). (C) Sample simulated white observation noise (red), two-photon observations (black, re-scaled for ease of visual comparison), and ground truth spikes (cyan), as well as the estimated calcium concentrations (purple) and putative spikes (green) for the 1st trial of neuron 1.

B.4.6 Performance comparison under observation noise model mismatch

In this section, we compare the performance of different estimates at varying SNR levels and average firing rate settings, with additional observation noise model mismatch. Figure B.6-A shows the NMSE (top) and leakage ratios (bottom) for the noise (left) and signal (right) correlation estimates vs. SNR (in dB), for the proposed method, Pearson correlations from two-photon data and two-stage Pearson method, for cases where the observation noise was generated by a white noise signal with an additive drift component from a low-frequency auto-regressive

White + drift noise model

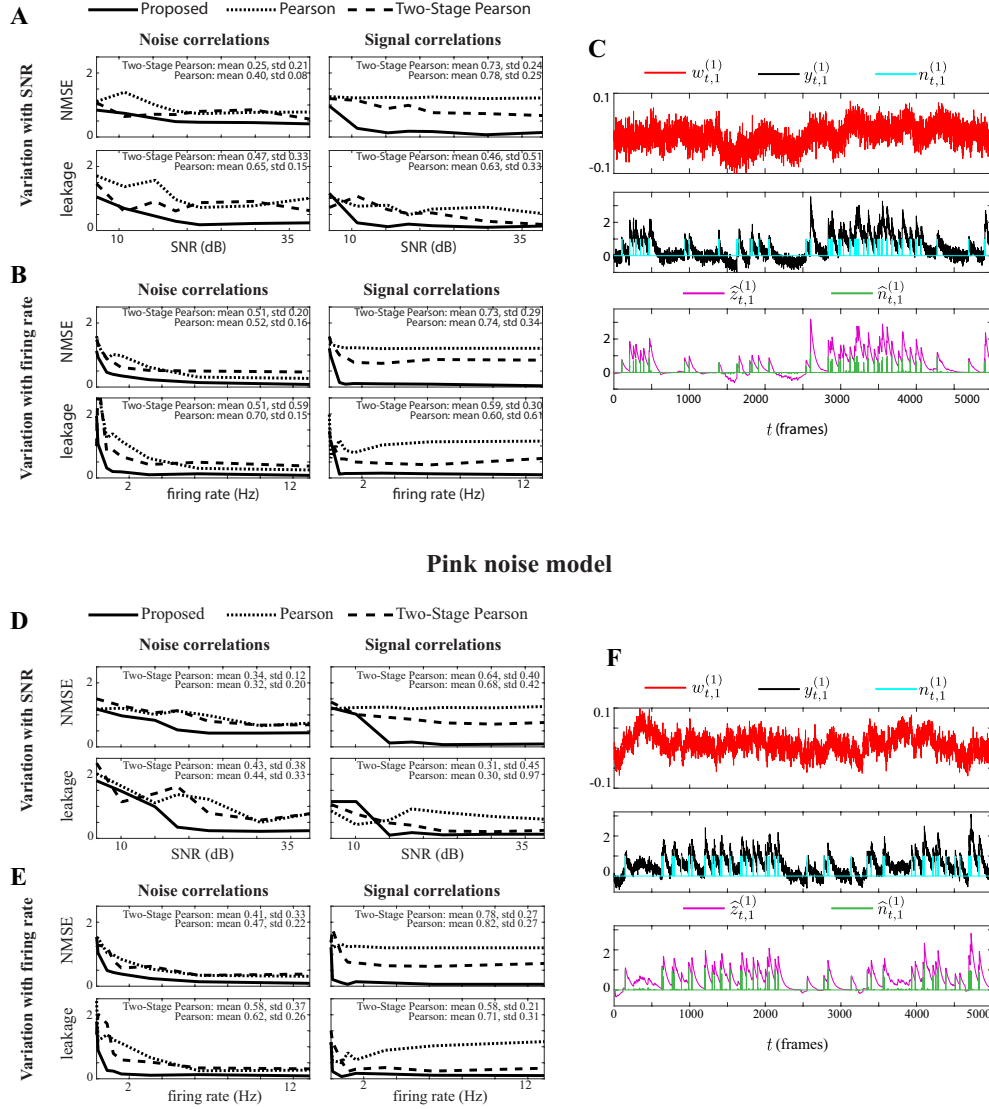


Figure B.6: Performance comparison under observation noise model mismatch. (A) NMSE (top) and leakage ratios (bottom) for the noise (left) and signal (right) correlation estimates vs. SNR (in dB), for the proposed method, Pearson correlations from two-photon data and Two-stage Pearson method. The observation noise is generated by a white noise signal with an additive drift component from a low-frequency auto-regressive process. The mean and standard deviation (std) of the normalized performance gain of the proposed method in comparison to the two existing methods are indicated as insets in each panel. (B) Same organization as panel A, but with respect to varying firing rates (in Hz). (C) Sample simulated observation noise (red), two-photon observations (black, re-scaled for ease of visual comparison), and ground truth spikes (cyan), as well as the estimated calcium concentrations (purple) and putative spikes (green) for the 1st trial of neuron 1. Panels D, E, and F are respectively in the same organization as panels A, B, and C, but the observation noise is generated by a pink noise process.

process. The mean and standard deviation (std) of the normalized performance gain of the proposed method in comparison to the two existing methods are indicated as insets in each panel. Figure B.6-B has the same organization as panel A, but with respect to varying firing rates (in Hz). Figure B.6-C shows sample simulated observation noise (red), two-photon observations (black, re-scaled for ease of visual comparison), and ground truth spikes (cyan), as well as the estimated calcium concentrations (purple) and putative spikes (green) for the 1st trial of neuron 1. Panels D-F are respectively in the same organization as panels A-C and show the comparisons corresponding to observation noise being generated by pink noise processes.

From the results in Figure B.6, we see that our proposed method outperforms the existing ones for a wide range of SNR and firing rate values and under both observation noise model mismatch conditions. Further, from Figure B.6-C and F, it can be observed that the ground truth spikes are favorably recovered as a byproduct of our method, even though the estimated calcium concentrations are contaminated by the temporally correlated fluctuations in observation noise. This in turn results in accurate signal and noise correlation estimates.

B.5 Supplementary material of the real data study 1 in Section 5.1.5 (Supplementary figures to Figure 5.4)

B.5.1 Probing the effect of stimulus integration window length on the performance of the proposed estimates

In this section, we observe the effect of stimulus integration window length (R) on our proposed noise and signal correlation estimates shown in Figure 5.4, Section 5.1.5. The rows

from left to right in Figure B.7-A shows the proposed noise (top) and signal (bottom) correlation estimates under settings $R = 1$, $R = 10$, $R = 25$ and $R = 50$. Figure B.7-B shows the corresponding null distributions of dissimilarities between proposed signal and noise correlation estimates. These results show that small values of $R = 1$ and $R = 10$ are not adequate to capture stimulus effect. However, both signal and noise correlation estimates exhibit consistency for $R = 25$ and $R = 50$.

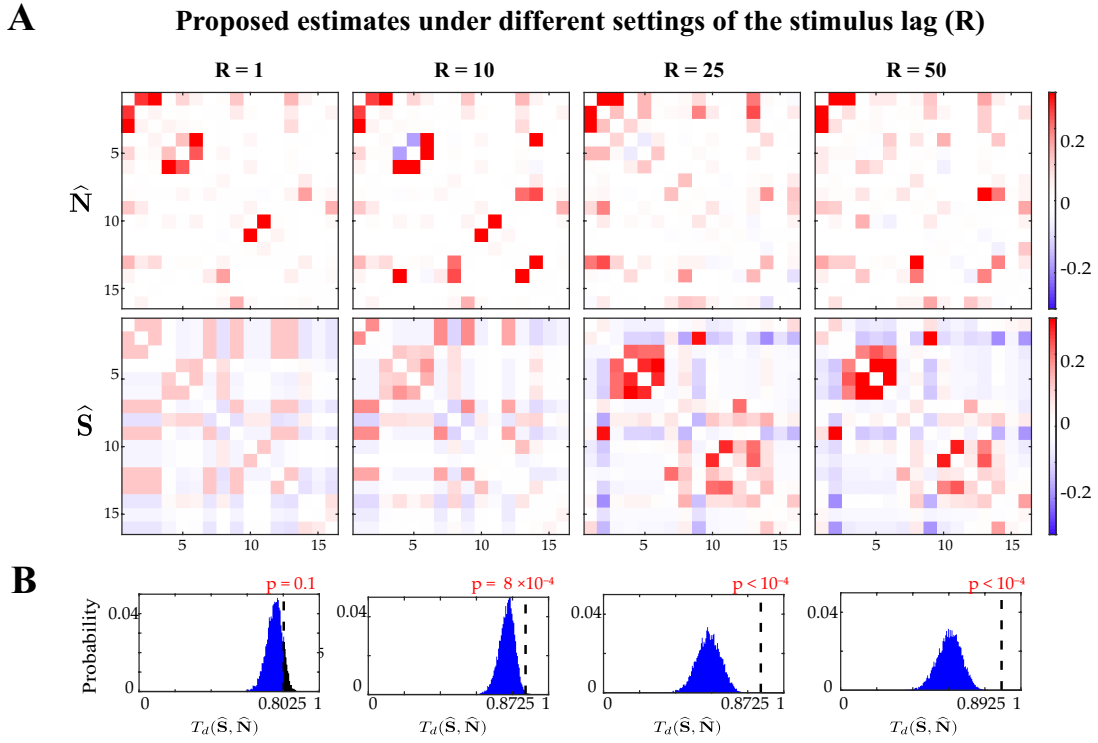


Figure B.7: Proposed noise and signal correlation estimates under different settings of the stimulus integration window length (R). (A) Proposed noise correlation (top) and signal correlation (bottom) estimates under different settings of R , from left to right: $R = 1$, $R = 10$, $R = 25$ and $R = 50$. (B) Null distributions of dissimilarities between proposed signal and noise correlation estimates corresponding to different choice of R . The observed test statistic in each case is indicated by a dashed vertical line, and the p-values are indicated above each panel.

B.5.2 Inspecting the inferred latent processes under high fluorescence activity due to rapid increase in firing rate

Here we investigate the inferred latent processes by the proposed method under settings of high fluorescence activity due to a rapid increase in firing rate. Figure B.8 shows a zoomed-in view of the estimated latent processes $\hat{z}_{t,1}^{(1)}$ (calcium concentration) and $\hat{n}_{t,1}^{(1)}$ (putative spikes) for a sample data segment with high fluorescence activity, from the data presented in Figure 5.4. The estimated latent processes reveal two mechanisms leveraged by our inference method to mitigate the aforementioned model mismatch: first, our proposed method predicts spiking events in adjacent time frames to compensate for the rapid increase in firing rate and thus infers calcium concentration levels that match the observed fluorescence; secondly, even though our generative model assumes that there is only one spiking event in a given time frame, this restriction is mitigated in our inference framework by relaxing the constraint $\hat{n}_{t,l}^{(j)} := \hat{z}_{t,l}^{(j)} - \alpha \hat{z}_{t-1,l}^{(j)} \leq 1$, as explained in Section 5.3.6. While this relaxation was performed for the sake of tractability of the inverse solution, it in fact leads to improved estimation results under episodes of rapid increase in firing rate, by allowing the putative spike magnitudes $\hat{n}_{t,l}^{(j)}$ to be greater than 1. The latter is evident in the magnitude of the inferred spikes in Figure B.8, following the rise of fluorescence activity.

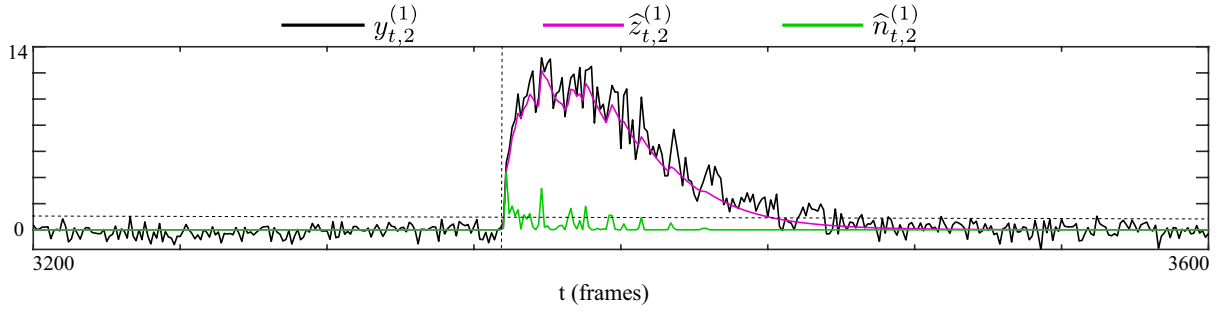


Figure B.8: The fluorescence observations (black), inferred calcium concentrations (purple), and putative spikes (green) by our proposed method, for a sample data segment with high fluorescence activity due to successive closely-spaced spikes. The rise onset of the fluorescence activity is marked by the vertical dashed line and the spiking magnitude level of 1 is indicated by the horizontal dashed line.

B.6 Supplementary material of the real data study 2 in Section 5.1.6 (Supplementary figures to Figure 5.6)

B.6.1 Histograms of the similarity/dissimilarity metrics under the shuffling procedure

In this section, we illustrate the histograms of the similarity/dissimilarity metrics under the shuffling procedure described in Section 5.3.8.2, applied on the real data study in Section 5.1.6. Figure B.9 shows the similarities between noise correlations N_{spn} and N_{stim} (first row: $T_s(\hat{N}_{\text{spn}}, \hat{N}_{\text{stim}})$) and the dissimilarities between signal and noise correlations $\hat{S}_{\text{stim}}$ and $\hat{N}_{\text{stim}}$ (second row: $T_d(\hat{S}_{\text{stim}}, \hat{N}_{\text{stim}})$), obtained by the shuffling procedure applied to the results of real data study 2 in Figure 5.6. The observed test statistic in each case is indicated by a dashed vertical line. These results show that the only statistically significant outcomes (with $p \leq 0.05$) are the similarities and dissimilarities obtained by our proposed method.

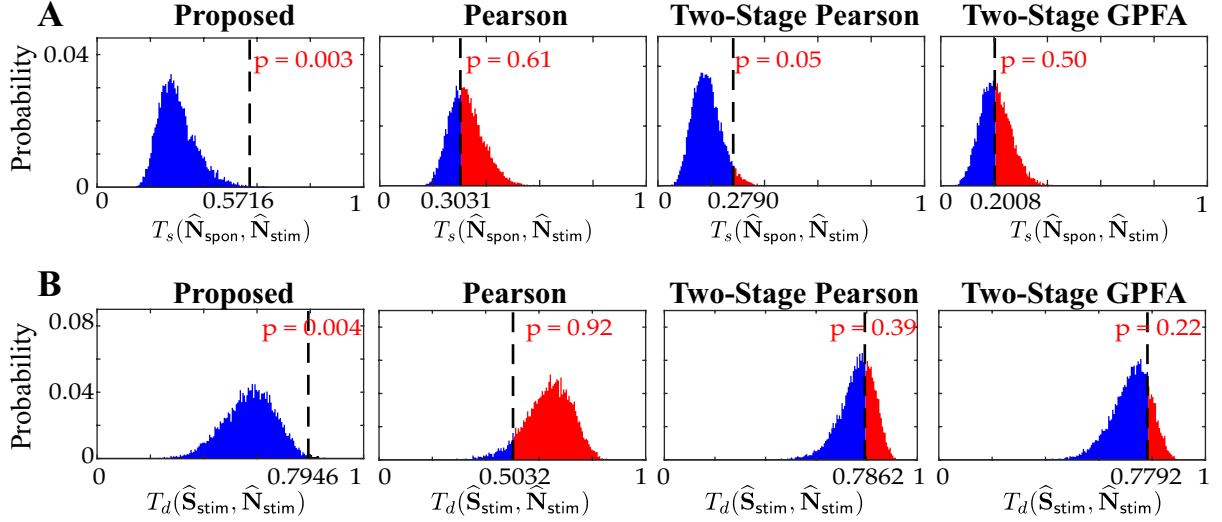


Figure B.9: Null distributions of (A) the similarities between N_{spon} and N_{stim} (top: $T_s(\hat{N}_{\text{spon}}, \hat{N}_{\text{stim}})$) and (B) the dissimilarities between $\hat{S}_{\text{stim}}$ and $\hat{N}_{\text{stim}}$ (bottom: $T_d(\hat{S}_{\text{stim}}, \hat{N}_{\text{stim}})$), obtained by the shuffling procedure applied to the results of real data study 2 in Figure 5.6. The observed test statistic in each case is indicated by a dashed vertical line. Rows from left to right: proposed method, Pearson correlations from two-photon data, two-stage Pearson correlations and two-stage GPFA estimates.

B.7 Supplementary material of the real data study 3 in Section 5.1.7 (Supplementary figures to Figure 5.7)

B.7.1 Comparing the marginal distributions of signal and noise correlations along the dorsoventral and rostrocaudal axes

In this section, we present the marginal distributions of signal and noise correlations along the dorsoventral and rostrocaudal axes, corresponding to the spatial distributions of correlations illustrated in Figure 5.7-E & -F. Figure B.10-A shows the cumulative marginal probability distributions of signal (blue) and noise (red) correlations along the rostrocaudal (top) and dorsoventral (bottom) directions, as estimated by the proposed method (left) and Pearson correlations from two-photon data (right), in layer 2/3 neurons. Figure B.10-B shows the same results for the layer

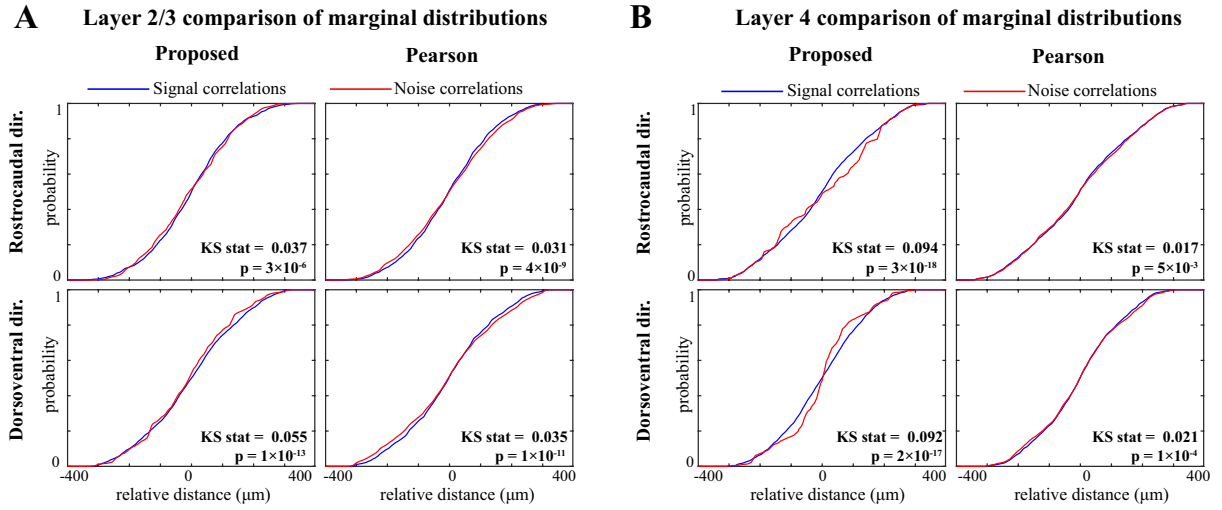


Figure B.10: Comparison of marginal distributions of signal and noise correlations. (A) Cumulative marginal probability distributions of signal (blue) and noise (red) correlations along the rostrocaudal (top) and dorsoventral (bottom) directions, as estimated by the proposed method (left) and Pearson correlations from two-photon data (right), in layer 2/3 neurons. The Kolmogorov–Smirnov (KS) test statistic along with the corresponding p-values are indicated as insets in each panel. Panel B shows the results for layer 4 in the same organization as panel A.

4 neurons. The Kolmogorov–Smirnov (KS) test statistic comparing each pair of distributions along with the corresponding p-values are indicated as insets in each panel.

These results show that along both directions and in both layers, the signal correlation distributions are significantly different from the corresponding noise correlation distributions, consistently for both methods. However, the KS statistics (i.e., effect sizes) for the proposed estimate are remarkably larger than those obtained from the Pearson estimates.

B.7.2 Marginal angular distributions of signal and noise correlations

Here, we present a comparison of the marginal angular distributions of signal and noise correlations, corresponding to the spatial distributions of correlations illustrated in Figure 5.7-E & -F. Figure B.11-A shows the polar histograms indicating the distribution of signal (top) and noise (bottom) correlations as a function of relative angle (in the dorsoventral-rostrocaudal

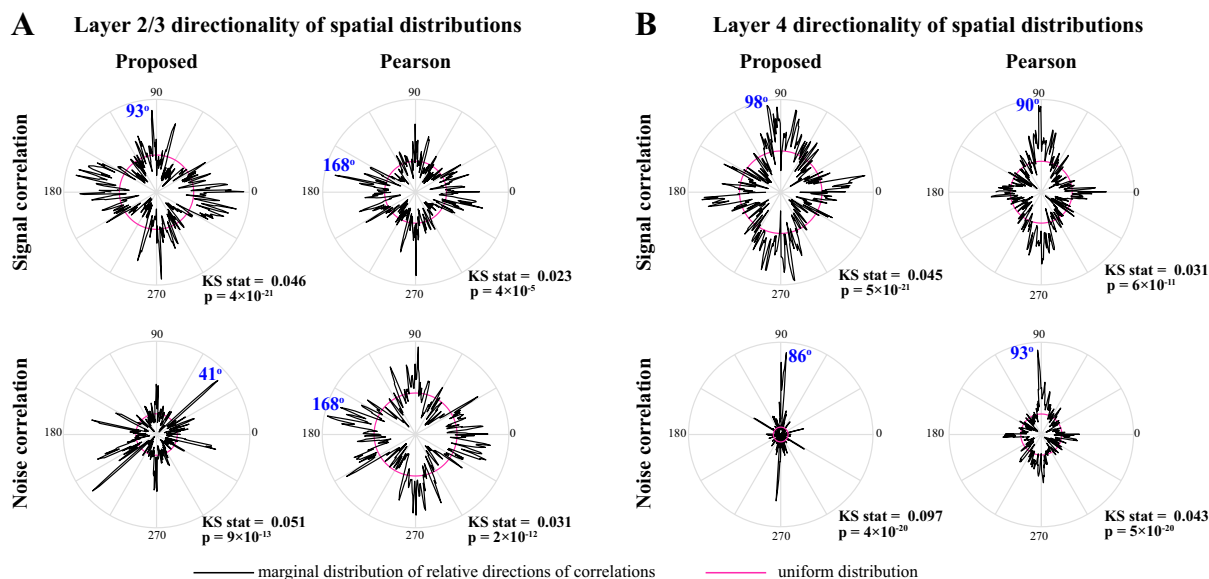


Figure B.11: Polar plots of the angular marginal distributions of correlations. (A) Polar histograms indicating the distribution of signal (top) and noise (bottom) correlations as a function of relative angle (in the dorsoventral-rostrocaudal coordinate system) between pairs of neurons in layer 2/3, as estimated by the proposed method (left) and Pearson correlations from two-photon data (right). The KS test statistic comparing each polar distribution with a uniform distribution (shown in magenta), along with the corresponding p-values are indicated below each polar plot. The mode of each probability distribution is also indicated in blue fonts. Panel B shows the results for layer 4 in the same organization as panel A.

coordinate system) between pairs of neurons in layer 2/3, as estimated by the proposed method (left) and Pearson correlations from two-photon data (right). Figure B.11-B shows the same results for the layer 4 neurons. The KS test statistic comparing each polar distribution with a uniform distribution (shown in magenta), along with the corresponding p-values are indicated below each polar plot. The mode of each probability distribution is also indicated in blue fonts.

All distributions are significantly non-uniform, and particularly indicate a rostrocaudal directionality in layer 4 (as indicated by the mode angles in panel B). The angular distributions of both signal and noise correlations in layer 4 exhibit salient modes in the rostrocaudal direction, whereas they are less directionally selective in layer 2/3.

Bibliography

- [1] R. M. Hutchison, T. Womelsdorf, E. A. Allen, P. A. Bandettini, V. D. Calhoun, M. Corbetta, S. Della Penna, J. H. Duyn, G. H. Glover, J. Gonzalez-Castillo, D. A. Handwerker, S. Keilholz, V. Kiviniemi, D. A. Leopold, F. de Pasquale, O. Sporns, M. Walter, and C. Chang, “Dynamic functional connectivity: Promise, issues, and interpretations,” *NeuroImage*, vol. 80, pp. 360–378, 2013, mapping the Connectome. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S105381191300579X>
- [2] M. P. van den Heuvel and H. E. Hulshoff Pol, “Exploring the brain network: A review on resting-state fmri functional connectivity,” *European Neuropsychopharmacology*, vol. 20, no. 8, pp. 519–534, 2010. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0924977X10000684>
- [3] S. A. Romano, V. Pérez-Schuster, A. Jouary, J. Boulanger-Weill, A. Candeo, T. Pietri, and G. Sumbre, “An integrated calcium imaging processing toolbox for the analysis of neuronal population dynamics,” *PLoS Computational Biology*, vol. 13, no. 6, p. e1005526, 2017.
- [4] C. Stosiek, O. Garaschuk, K. Holthoff, and A. Konnerth, “In vivo two-photon calcium imaging of neuronal networks,” *Proceedings of the National Academy of Sciences*, vol. 100, no. 12, pp. 7319–7324, 2003. [Online]. Available: <https://doi.org/10.1073/pnas.1232232100>
- [5] K. Svoboda and R. Yasuda, “Principles of two-photon excitation microscopy and its applications to neuroscience,” *Neuron*, vol. 50, no. 6, pp. 823 – 839, 2006. [Online]. Available: <https://doi.org/10.1016/j.neuron.2006.05.019>
- [6] M. B. Ahrens, M. B. Orger, D. N. Robson, J. M. Li, and P. J. Keller, “Whole-brain functional imaging at cellular resolution using light-sheet microscopy,” *Nature methods*, vol. 10, no. 5, p. 413, 2013. [Online]. Available: <https://doi.org/10.1038/nmeth.2434>
- [7] N. Vladimirov, Y. Mu, T. Kawashima, D. V. Bennett, C.-T. Yang, L. L. Looger, P. J. Keller, J. Freeman, and M. B. Ahrens, “Light-sheet functional imaging in fictively

- behaving zebrafish,” *Nature methods*, vol. 11, p. 883, 2014. [Online]. Available: <https://doi.org/10.1038/nmeth.3040>
- [8] R. P. Vertes and R. W. Stackman, *Electrophysiological Recording Techniques*. Humana Press, 2011, vol. 54.
 - [9] J. J. Jun, N. Steinmetz, J. H. Siegle, D. J. Denman, M. Bauza, B. Barbarits, A. K. Lee, C. Anastassiou, A. Andrei, C. Aydin, M. Barbic, T. Blanche, V. Bonin, J. Couto, B. Dutta, S. Gratiy, D. Gutnisky, M. Häusser, B. Karsh, and T. D. Harris, “Fully integrated silicon probes for high-density recording of neural activity,” *Nature*, vol. 551, pp. 232–236, 11 2017.
 - [10] J. Du, T. J. Blanche, R. R. Harrison, H. A. Lester, and S. C. Masmanidis, “Multiplexed, high density electrophysiology with nanofabricated neural probes,” *PLOS ONE*, vol. 6, no. 10, pp. 1–11, 10 2011.
 - [11] F. De Vico Fallani, M. Corazzol, J. R. Sternberg, C. Wyart, and M. Chavez, “Hierarchy of neural organization in the embryonic spinal cord: Granger-causality graph analysis of in vivo calcium imaging data,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 3, pp. 333–341, May 2015.
 - [12] N. A. Francis, D. E. Winkowski, A. Sheikhattar, K. Armengol, B. Babadi, and P. O. Kanold, “Small networks encode decision-making in primary auditory cortex,” *Neuron*, vol. 97, no. 4, pp. 885 – 897.e6, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0896627318300229>
 - [13] G. Rothschild, I. Nelken, and A. Mizrahi, “Functional organization and population dynamics in the mouse primary auditory cortex,” *Nature Neuroscience*, vol. 13, no. 3, p. 353, 2010.
 - [14] D. E. Winkowski and P. O. Kanold, “Laminar transformation of frequency organization in auditory cortex,” *Journal of Neuroscience*, vol. 33, no. 4, pp. 1498–1508, 2013. [Online]. Available: <https://www.jneurosci.org/content/33/4/1498>
 - [15] T. Deneux, A. Kaszas, G. Szalay, G. Katona, T. Lakner, A. Grinvald, B. Rózsa, and I. Vanzetta, “Accurate spike estimation from noisy calcium signals for ultrafast three-dimensional imaging of large neuronal populations in vivo,” *Nature Communications*, vol. 7, no. 1, pp. 1–17, 2016.
 - [16] J. Friedrich, P. Zhou, and L. Paninski, “Fast online deconvolution of calcium imaging data,” *PLOS Computational Biology*, vol. 13, no. 3, pp. 1–26, 03 2017. [Online]. Available: <https://doi.org/10.1371/journal.pcbi.1005423>
 - [17] B. F. Grewe, D. Langer, H. Kasper, B. M. Kampa, and F. Helmchen, “High-speed in vivo calcium imaging reveals neuronal network activity with near-millisecond precision,” *Nature Methods*, vol. 7, no. 5, p. 399, 2010.

- [18] S. W. Jewell, T. D. Hocking, P. Fearnhead, and D. M. Witten, “Fast nonconvex deconvolution of calcium imaging data,” *Biostatistics*, vol. 21, October 2020. [Online]. Available: <https://doi.org/10.1093/biostatistics/kxy083>
- [19] S. W. Jewell and D. M. Witten, “Exact spike train inference via ℓ_0 optimization,” *The Annals of Applied Statistics*, vol. 12, no. 4, pp. 2457–2482, 2018. [Online]. Available: <https://projecteuclid.org/euclid.aoas/1542078052>
- [20] A. Kazemipour, J. Liu, K. Solarana, D. A. Nagode, P. O. Kanold, M. Wu, and B. Babadi, “Fast and stable signal deconvolution via compressible state-space models,” *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 1, pp. 74–86, Jan 2018.
- [21] M. Pachitariu, C. Stringer, and K. D. Harris, “Robustness of spike deconvolution for neuronal calcium imaging,” *Journal of Neuroscience*, vol. 38, no. 37, pp. 7976–7985, 2018. [Online]. Available: <https://www.jneurosci.org/content/38/37/7976>
- [22] E. Pnevmatikakis, D. Soudry, Y. Gao, T. A. Machado, J. Merel, D. Pfau, T. Reardon, Y. Mu, C. Lacefield, W. Yang, M. Ahrens, R. Bruno, T. M. Jessell, D. Peterka, R. Yuste, and L. Paninski, “Simultaneous denoising, deconvolution, and demixing of calcium imaging data,” *Neuron*, vol. 89, no. 2, pp. 285–299, 2016.
- [23] C. Stringer and M. Pachitariu, “Computational processing of neural recordings from calcium imaging data,” *Current Opinion in Neurobiology*, vol. 55, pp. 22–31, 2019. [Online]. Available: <https://doi.org/10.1016/j.conb.2018.11.005>
- [24] L. Theis, P. Berens, E. Froudarakis, J. Reimer, M. Román Rosón, T. Baden, T. Euler, A. Tolias, and M. Bethge, “Benchmarking spike rate inference in population calcium imaging,” *Neuron*, vol. 90, no. 3, pp. 471–482, 2016. [Online]. Available: <https://doi.org/10.1016/j.neuron.2016.04.014>
- [25] J. T. Vogelstein, A. M. Packer, T. A. Machado, T. Sippy, B. Babadi, R. Yuste, and L. Paninski, “Fast nonnegative deconvolution for spike train inference from population calcium imaging,” *Journal of Neurophysiology*, vol. 104, no. 6, pp. 3691–3704, 2010. [Online]. Available: <https://doi.org/10.1152/jn.01073.2009>
- [26] J. T. Vogelstein, B. O. Watson, A. M. Packer, R. Yuste, B. Jedynek, and L. Paninski, “Spike inference from calcium imaging using sequential Monte Carlo methods,” *Biophysical Journal*, vol. 97, no. 2, pp. 636–655, 2009. [Online]. Available: <https://doi.org/10.1016/j.bpj.2008.08.005>
- [27] M. A. Triplett, Z. Pujic, B. Sun, L. Avitan, and G. J. Goodhill, “Model-based decoupling of evoked and spontaneous neural activity in calcium imaging data,” *PLOS Computational Biology*, vol. 16, no. 11, pp. 1–28, 11 2020. [Online]. Available: <https://doi.org/10.1371/journal.pcbi.1008330>
- [28] W. Truccolo, U. T. Eden, M. R. Fellows, J. P. Donoghue, and E. N. Brown, “A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects,” *Journal of Neurophysiology*, vol. 93, no. 2, pp. 1074–1089, 2005.

- [29] D. M. Halliday and J. R. Rosenberg, *Time and Frequency Domain Analysis of Spike Train and Time Series Data*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 503–543.
- [30] A. G. Nedungadi, G. Rangarajan, N. Jain, and M. Ding, “Analyzing multiple spike trains with nonparametric granger causality,” *Journal of Computational Neuroscience*, vol. 27, pp. 55–64, 2008.
- [31] Y. Cao, S. K. Maran, M. Dhamala, D. Jaeger, and D. H. Heck, “Behavior-related pauses in simple-spike activity of mouse purkinje cells are linked to spike rate modulation,” *Journal of Neuroscience*, vol. 32, no. 25, pp. 8678–8685, 2012.
- [32] M. Chen, Y. Yan, X. Gong, C. D. Gilbert, H. Liang, and W. Li, “Incremental integration of global contours through interplay between visual cortical areas,” *Neuron*, vol. 82, no. 3, pp. 682 – 694, 2014.
- [33] A. C. Smith and E. N. Brown, “Estimating a state-space model from point process observations,” *Neural Computation*, vol. 15, no. 5, pp. 965–991, 2003.
- [34] L. D. Lewis, V. S. Weiner, E. A. Mukamel, J. A. Donoghue, E. N. Eskandar, J. R. Madsen, W. S. Anderson, L. R. Hochberg, S. S. Cash, E. N. Brown, and P. L. Purdon, “Rapid fragmentation of neuronal networks at the onset of propofol-induced unconsciousness,” *Proc. Natl. Acad. Sci.*, vol. 109, no. 49, pp. 19 891–19 892, 2012.
- [35] B. O. Watson, D. Levenstein, J. P. Greene, J. N. Gelinas, and G. Buzsáki, “Network homeostasis and state dynamics of neocortical sleep,” *Neuron*, vol. 90, no. 4, pp. 839 – 852, 2016.
- [36] R. van de Schoot, S. Depaoli, R. King, B. Kramer, K. Märtens, M. G. Tadesse, M. Vannucci, A. Gelman, D. Veen, J. Willemsen, and C. Yau, “Bayesian statistics and modelling,” *Nature Reviews Methods Primers*, vol. 1, no. 1, p. 1, Jan 2021. [Online]. Available: <https://doi.org/10.1038/s43586-020-00001-2>
- [37] A. Schnitzler and J. Gross, “Normal and pathological oscillatory communication in the brain,” *Nat. Rev. Neurosci.*, vol. 6, pp. 285–296, 2005.
- [38] P. J. Uhlhaas and W. Singer, “Neural synchrony in brain disorders: Relevance for cognitive dysfunctions and pathophysiology,” *Neuron*, vol. 52, no. 1, pp. 155 – 168, 2006.
- [39] L. M. Ward, “Synchronous neural oscillations and cognitive processes,” *Trends in Cognitive Sciences*, vol. 7, no. 12, pp. 553 – 559, 2003.
- [40] D. J. Thomson, “Spectrum estimation and harmonic analysis,” *Proceedings of the IEEE*, vol. 70, no. 9, pp. 1055–1096, Sept 1982.
- [41] D. Slepian, “Prolate spheroidal wave functions, fourier analysis, and uncertainty — V: the discrete case,” *The Bell System Technical Journal*, vol. 57, no. 5, pp. 1371–1430, May 1978.

- [42] D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications*. Cambridge University Press, 1993.
- [43] B. Babadi and E. N. Brown, “A review of multitaper spectral analysis,” *IEEE Trans. on Biomedical Engineering*, vol. 61, pp. 1555–1564, 2014.
- [44] T. P. Bronez, “On the performance advantage of multitaper spectral analysis,” *IEEE Trans. on Signal Processing*, vol. 40, no. 12, pp. 2941–2946, Dec 1992.
- [45] M. B. Priestley, “Evolutionary spectra and non-stationary processes,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 27, no. 2, pp. 204–229, 1965. [Online]. Available: <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1965.tb01488.x>
- [46] R. H. Shumway and D. S. Stoffer, “An approach to time series smoothing and forecasting using the EM algorithm,” *Journal of time series analysis*, vol. 3, no. 4, pp. 253–264, 1982.
- [47] A. Rupasinghe and B. Babadi, “Multitaper analysis of evolutionary spectral density matrix from multivariate spiking observations,” in *2019 IEEE Data Science Workshop (DSW)*, Minneapolis, MN, June 2–5, 2019.
- [48] A. Rupasinghe and B. Babadi, “Multitaper analysis of semi-stationary spectra from multivariate neuronal spiking observations,” *IEEE Transactions on Signal Processing*, vol. 68, pp. 4382–4396, 2020.
- [49] “Multitaper Analysis of Evolutionary Spectra from Multivariate Spiking Observations MATLAB Codes,” 2019. [Online]. Available: <https://github.com/Anuththara-Rupasinghe/PPMT-ESD-Estimation>
- [50] J. Geweke, “Measurement of linear dependence and feedback between multiple time series,” *Journal of the American Statistical Association*, vol. 77, no. 378, pp. 304–313, 1982.
- [51] J. F. Geweke, “Measures of conditional linear dependence and feedback between time series,” *Journal of the American Statistical Association*, vol. 79, no. 388, pp. 907–915, 1984.
- [52] A. Rupasinghe, S. Mukherjee, and B. Babadi, “Adaptive frequency-domain granger causal inference from neuronal ensemble data,” in *2020 54th Asilomar Conference on Signals, Systems, and Computers*, 2020, pp. 101–105.
- [53] G. D. Clifford, “Signal processing methods for heart rate variability,” Ph.D. dissertation, Oxford University, UK, 2002.
- [54] F. Shaffer and J. P. Ginsberg, “An overview of heart rate variability metrics and norms,” *Frontiers in Public Health*, vol. 5, 2017. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fpubh.2017.00258>
- [55] K. Li, H. Rüdiger, and T. Ziemssen, “Spectral analysis of heart rate variability: Time window matters,” *Frontiers in Neurology*, vol. 10, 2019. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fneur.2019.00545>

- [56] L. F. Abbott and P. Dayan, “The effect of correlated variability on the accuracy of a population code,” *Neural Computation*, vol. 11, no. 1, pp. 91–101, 1999. [Online]. Available: <https://doi.org/10.1162/089976699300016827>
- [57] B. B. Averbeck, P. E. Latham, and A. Pouget, “Neural correlations, population coding and computation,” *Nature Reviews Neuroscience*, vol. 7, no. 5, pp. 358–366, 2006. [Online]. Available: <https://doi.org/10.1038/nrn1888>
- [58] M. Cohen and A. Kohn, “Measuring and interpreting neuronal correlations,” *Nature Neuroscience*, vol. 14, no. 7, pp. 811–819, 2011. [Online]. Available: <https://doi.org/10.1038/nn.2842>
- [59] B. J. Hansen, M. I. Chelaru, and V. Dragoi, “Correlated variability in laminar cortical circuits,” *Neuron*, vol. 76, no. 3, pp. 590–602, 2012.
- [60] A. Kohn, R. Coen-Cagli, I. Kanitscheider, and A. Pouget, “Correlations and neuronal population information,” *Annual Review of Neuroscience*, vol. 39, no. 1, pp. 237–256, 2016, pMID: 27145916.
- [61] A. Kohn and M. A. Smith, “Stimulus dependence of neuronal correlation in primary visual cortex of the macaque,” *Journal of Neuroscience*, vol. 25, no. 14, pp. 3661–3673, 2005. [Online]. Available: <https://www.jneurosci.org/content/25/14/3661>
- [62] D. R. Lyamzin, S. J. Barnes, R. Donato, J. A. Garcia-Lazaro, T. Keck, and N. A. Lesica, “Nonlinear transfer of signal and noise correlations in cortical networks,” *Journal of Neuroscience*, vol. 35, no. 21, pp. 8065–8080, 2015. [Online]. Available: <https://www.jneurosci.org/content/35/21/8065>
- [63] J. S. Montijn, M. Vinck, and C. M. A. Pennartz, “Population coding in mouse visual cortex: response reliability and dissociability of stimulus tuning and noise correlation,” *Frontiers in Computational Neuroscience*, vol. 8, p. 58, 2014. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fncom.2014.00058>
- [64] M. A. Smith and M. A. Sommer, “Spatial and temporal scales of neuronal correlation in visual area V4,” *Journal of Neuroscience*, vol. 33, no. 12, pp. 5422–5432, 2013. [Online]. Available: <https://doi.org/10.1523/JNEUROSCI.2929-08.2008>
- [65] H. Sompolinsky, H. Yoon, K. Kang, and M. Shamir, “Population coding in neuronal systems with correlated noise,” *Phys. Rev. E*, vol. 64, p. 051904, Oct 2001. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevE.64.051904>
- [66] D. Yatsenko, K. Josić, A. S. Ecker, E. Froudarakis, R. J. Cotton, and A. S. Tolias, “Improved estimation and interpretation of correlations in neural circuits,” *PLOS Computational Biology*, vol. 11, no. 3, pp. 1–28, 03 2015. [Online]. Available: <https://doi.org/10.1371/journal.pcbi.1004083>
- [67] M. R. Cohen and J. H. Maunsell, “Attention improves performance primarily by reducing interneuronal correlations,” *Nature Neuroscience*, vol. 12, no. 12, p. 1594, 2009.

- [68] T. Gawne and B. Richmond, “How independent are the messages carried by adjacent inferior temporal cortical neurons?” *Journal of Neuroscience*, vol. 13, no. 7, pp. 2758–2771, 1993. [Online]. Available: <https://www.jneurosci.org/content/13/7/2758>
- [69] K. Josić, E. Shea-Brown, B. Doiron, and J. de la Rocha, “Stimulus-dependent correlations and population codes,” *Neural Computation*, vol. 21, no. 10, pp. 2774–2804, 2009, pMID: 19635014. [Online]. Available: <https://doi.org/10.1162/neco.2009.10-08-879>
- [70] G. Vinci, V. Ventura, M. A. Smith, and R. E. Kass, “Separating spike count correlation from firing rate correlation,” *Neural computation*, vol. 28, no. 5, pp. 849–881, 2016. [Online]. Available: <https://www.mitpressjournals.org/doi/full/10.1162/NECO.a.00831>
- [71] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, “An introduction to variational methods for graphical models,” *Machine Learning*, vol. 37, no. 2, pp. 183–233, 1999.
- [72] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, “Variational inference: A review for statisticians,” *Journal of the American Statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.
- [73] N. G. Polson, J. G. Scott, and J. Windle, “Bayesian inference for logistic models using Pólya–Gamma latent variables,” *Journal of the American Statistical Association*, vol. 108, no. 504, pp. 1339–1349, 2013. [Online]. Available: <https://www.jstor.org/stable/24247065>
- [74] H. E. Rauch, C. T. Striebel, and F. Tung, “Maximum likelihood estimates of linear dynamic systems,” *AIAA Journal*, vol. 3, no. 8, pp. 1445–1450, aug 1965.
- [75] D. Ba, B. Babadi, P. L. Purdon, and E. N. Brown, “Convergence and stability of iteratively re-weighted least squares algorithms,” *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pp. 183–195, Jan 2014.
- [76] A. Rupasinghe and B. Babadi, “Robust inference of neuronal correlations from blurred and noisy spiking observations,” in *2020 54th Annual Conference on Information Sciences and Systems (CISS)*, 2020, pp. 1–5. [Online]. Available: <https://doi.org/10.1109/CISS48834.2020.1570617409>
- [77] A. Rupasinghe, N. Francis, J. Liu, Z. Bowen, P. O. Kanold, and B. Babadi, “Direct extraction of signal and noise correlations from two-photon calcium imaging of ensemble neuronal activity,” *eLife*, vol. 10, p. e68046, jun 2021. [Online]. Available: <https://doi.org/10.7554/eLife.68046>
- [78] A. Rupasinghe, “Direct extraction of signal and noise correlations from two-photon calcium imaging of ensemble neuronal activity MATLAB Codes,” 2020. [Online]. Available: <https://github.com/Anuththara-Rupasinghe/Signal-Noise-Correlation>
- [79] A. Rupasinghe, N. Francis, J. Liu, Z. Bowen, P. O. Kanold, and B. Babadi, “Experimental data from ‘direct extraction of signal and noise correlations from two-photon calcium imaging of ensemble neuronal activity’,” 2021. [Online]. Available: <http://hdl.handle.net/1903/26917>

- [80] M. L. Caras and D. H. Sanes, "Top-down modulation of sensory cortex gates perceptual learning," *Proceedings of the National Academy of Sciences*, vol. 114, no. 37, pp. 9972–9977, 2017. [Online]. Available: <https://www.pnas.org/doi/abs/10.1073/pnas.1712305114>
- [81] E. Izhikevich, "Simple model of spiking neurons," *IEEE Transactions on Neural Networks*, vol. 14, no. 6, pp. 1569–1572, 2003.
- [82] E. N Brown, R. Kass, and P. P Mitra, "Multiple neural spike train data analysis: state-of-the-art and future challenges," *Nature Neuroscience*, vol. 7, pp. 456–61, 06 2004.
- [83] A. V. Herz, T. Gollisch, C. K. Machens, and D. Jaeger, "Modeling single-neuron dynamics and computations: a balance of detail and abstraction," *science*, vol. 314, no. 5796, pp. 80–85, 2006.
- [84] L. Paninski, "Maximum likelihood estimation of cascade point-process neural encoding models," *Network: Computation in Neural Systems*, vol. 15, no. 4, pp. 243–262, 2004.
- [85] L. Paninski, J. Pillow, and J. Lewi, "Statistical models for neural encoding, decoding, and optimal stimulus design," *Progress in brain research*, vol. 165, pp. 493–507, 2007.
- [86] S. Miran, P. L. Purdon, E. N. Brown, and B. Babadi, "Robust estimation of sparse narrowband spectra from neuronal spiking data," *IEEE Trans. on Biomedical Engineering*, vol. 64, no. 10, pp. 2462–2474, Oct 2017.
- [87] P. Das and B. Babadi, "Multitaper spectral analysis of neuronal spiking activity driven by latent stationary processes," *Signal Processing*, vol. 170, p. 107429, 2020.
- [88] S. Grün, M. Diesmann, and A. Aertsen, "Unitary events in multiple single-neuron spiking activity: II. Nonstationary data," *Neural Computation*, vol. 14, no. 1, pp. 81–119, 2002.
- [89] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, 1998.
- [90] L. Cohen, *Time-frequency Analysis: Theory and Applications*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1995.
- [91] M. B. Priestley, *Spectral analysis and time series*. Academic Press London ; New York, 1981.
- [92] G. Matz and F. Hlawatsch, "Nonstationary spectral analysis based on time-frequency operator symbols and underspread approximations," *IEEE Trans. on Information Theory*, vol. 52, no. 3, pp. 1067–1086, March 2006.
- [93] W. Martin and P. Flandrin, "Wigner-Ville spectral analysis of nonstationary processes," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 33, no. 6, pp. 1461–1470, December 1985.

- [94] G. Matz, F. Hlawatsch, and W. Kozek, "Generalized evolutionary spectral analysis and the Weyl spectrum of nonstationary random processes," *IEEE Trans. on Signal Processing*, vol. 45, no. 6, pp. 1520–1534, June 1997.
- [95] J. Hammond and P. White, "The analysis of non-stationary signals using time-frequency methods," *Journal of Sound and Vibration*, vol. 190, no. 3, pp. 419 – 447, 1996.
- [96] P. Das and B. Babadi, "Dynamic Bayesian multitaper spectral analysis," *IEEE Trans. on Signal Processing*, vol. 66, no. 6, pp. 1394–1409, March 2018.
- [97] S.-E. Kim, M. K. Behr, D. Ba, and E. N. Brown, "State-space multitaper time-frequency analysis," *Proc. Natl. Acad. Sci.*, vol. 115, no. 1, pp. E5–E14, 2018.
- [98] E. N. Brown, R. Barbieri, U. T. Eden, and L. M. Frank, "Likelihood methods for neural spike train data analysis," *Computational neuroscience: A comprehensive approach*, pp. 253–286, 2003.
- [99] U. T. Eden, L. M. Frank, R. Barbieri, V. Solo, and E. N. Brown, "Dynamic analysis of neural encoding by point process adaptive filtering," *Neural computation*, vol. 16, no. 5, pp. 971–998, 2004.
- [100] A. Sheikhattar, J. B. Fritz, S. A. Shamma, and B. Babadi, "Recursive sparse point process regression with application to spectrotemporal receptive field plasticity analysis," *IEEE Trans. on Signal Processing*, vol. 64, no. 8, pp. 2026–2039, 2015.
- [101] B. O. Watson, D. Levenstein, J. P. Greene, J. N. Gelinas, and G. Buzsáki, "Multi-unit spiking activity recorded from rat frontal cortex (brain regions mPFC, OFC, ACC, and M2) during wake-sleep episode wherein at least 7 minutes of wake are followed by 20 minutes of sleep," *CRCNS.org*, 2016. [Online]. Available: <http://dx.doi.org/10.6080/K02N506Q>
- [102] D. J. Daley and D. Vere-Jones, *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Science & Business Media, 2007, vol. 2.
- [103] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*. Berlin, Heidelberg: Springer-Verlag, 1986.
- [104] A. T. Walden, "A unified view of multitaper multivariate spectral estimation," *Biometrika*, vol. 87, no. 4, pp. 767–788, 2000.
- [105] P. D. Jong and M. J. Mackinnon, "Covariances for smoothed estimates in state space models," *Biometrika*, vol. 75, no. 3, pp. 601–602, 1988.
- [106] K. Lii and M. Rosenblatt, "Prolate spheroidal spectral estimates," *Statistics & Probability Letters*, vol. 78, no. 11, pp. 1339 – 1348, 2008.
- [107] C. Goutis and G. Casella, "Explaining the saddlepoint approximation," *The American Statistician*, vol. 53, no. 3, pp. 216–224, 1999.

- [108] S. Chauvette, S. Crochet, M. Volgushev, and I. Timofeev, “Properties of slow oscillation during slow-wave sleep and anesthesia in cats,” *Journal of Neuroscience*, vol. 31, no. 42, pp. 14 998–15 008, 2011.
- [109] V. V. Vyazovskiy, B. A. Riedner, C. Cirelli, and G. Tononi, “Sleep homeostasis and cortical synchronization: II. A local field potential study of sleep slow waves in the rat,” *Sleep*, vol. 30, no. 12, pp. 1631–1642, 2007.
- [110] V. Crunelli and S. W. Hughes, “The slow (< 1 Hz) rhythm of non-REM sleep: a dialogue between three cardinal oscillators,” *Nature Neuroscience*, vol. 13, no. 1, p. 9, 2010.
- [111] R. Barbieri, E. C. Matten, A. A. Alabi, and E. N. Brown, “A point-process model of human heartbeat intervals: new definitions of heart rate and heart rate variability,” *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 288, no. 1, pp. H424–H435, 2005.
- [112] M. Ding, Y. Chen, and S. L. Bressler, *Granger Causality: Basic Theory and Application to Neuroscience*. John Wiley & Sons, Ltd, 2006, ch. 17, pp. 437–460.
- [113] M. Dhamala, G. Rangarajan, and M. Ding, “Analyzing information flow in brain networks with nonparametric granger causality,” *NeuroImage*, vol. 41, no. 2, pp. 354 – 362, 2008.
- [114] X. Wen, G. Rangarajan, and M. Ding, “Multivariate granger causality: an estimation framework based on factorization of the spectral density matrix,” *Phil. Trans. R. Soc. A*, vol. 371, no. 2, pp. 354 – 362, 2013.
- [115] L. Barnett and A. K. Seth, “The mvgc multivariate granger causality toolbox: A new approach to granger-causal inference,” *Journal of Neuroscience Methods*, vol. 223, pp. 50 – 68, 2014.
- [116] M. F. Pagnotta, M. Dhamala, and G. Plomp, “Benchmarking nonparametric granger causality: Robustness against downsampling and influence of spectral decomposition parameters,” *NeuroImage*, vol. 183, pp. 478 – 494, 2018.
- [117] G. T. Wilson, “The factorization of matricial spectral densities,” *SIAM Journal on Applied Mathematics*, vol. 23, no. 4, pp. 420–426, 1972.
- [118] S.-E. Kim, D. Ba, and E. N. Brown, “A multitaper frequency-domain bootstrap method,” *IEEE Signal Processing Letters*, vol. 25, no. 12, pp. 1805–1809, 2018.
- [119] A. Kerlin, B. Mohar, D. Flickinger, B. J. MacLennan, M. B. Dean, C. Davis, N. Spruston, and K. Svoboda, “Functional clustering of dendritic activity during decision-making,” *eLife*, vol. 8, p. e46966, oct 2019. [Online]. Available: <https://doi.org/10.7554/eLife.46966>
- [120] F. Najafi, G. F. Elsayed, R. Cao, E. Pnevmatikakis, P. E. Latham, J. P. Cunningham, and A. K. Churchland, “Excitatory and inhibitory subnetworks are equally selective during decision-making and emerge simultaneously during learning,” *Neuron*, vol. 105, no. 1, pp. 165 – 179.e8, 2020. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0896627319308487>

- [121] R. N. Ramesh, C. R. Burgess, A. U. Sugden, M. Gyetvan, and M. L. Andermann, “Intermingled ensembles in visual association cortex encode stimulus identity or predicted outcome,” *Neuron*, vol. 100, no. 4, pp. 900–915, 2018.
- [122] D. Soudry, S. Keshri, P. Stinson, M.-h. Oh, G. Iyengar, and L. Paninski, “Efficient ”shotgun” inference of neural connectivity from highly sub-sampled activity data,” *PLoS Computational Biology*, vol. 11, no. 10, 2015. [Online]. Available: <https://doi.org/10.1371/journal.pcbi.1004464>
- [123] H. Lütcke, F. Gerhard, F. Zenke, W. Gerstner, and F. Helmchen, “Inference of neuronal network spike dynamics and topology from calcium imaging data,” *Frontiers in Neural Circuits*, vol. 7, p. 201, 2013. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fncir.2013.00201>
- [124] L. Aitchison, L. Russell, A. M. Packer, J. Yan, P. Castonguay, M. Hausser, and S. C. Turaga, “Model-based Bayesian inference of neural activity and connectivity from all-optical interrogation of a neural circuit,” in *Advances in Neural Information Processing Systems*, 2017, pp. 3486–3495.
- [125] Y. Mishchenko, J. Vogelstein, and L. Paninski, “A Bayesian approach for inferring neuronal connectivity from calcium fluorescent imaging data,” *Annals of Applied Statistics*, vol. 5, no. 2B, pp. 1229–1261, 2011. [Online]. Available: <https://www.jstor.org/stable/23024850>
- [126] S. L. Keeley, M. C. Aoi, Y. Yu, S. L. Smith, and J. W. Pillow, “Identifying signal and noise structure in neural population activity with gaussian process factor models,” in *Advances in Neural Information Processing Systems 33*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds. Curran Associates, Inc., 2020. [Online]. Available: <https://papers.nips.cc/paper/2020/file/9eed867b73ab1eab60583c9d4a789b1b-Paper.pdf>
- [127] O. I. Rumyantsev, J. A. Lecoq, O. Hernandez, Y. Zhang, J. Savall, R. Chrapkiewicz, J. Li, H. Zeng, S. Ganguli, and M. J. Schnitzer, “Fundamental bounds on the fidelity of sensory cortical coding,” *Nature*, vol. 580, no. 7801, p. 100–105, April 2020. [Online]. Available: <https://doi.org/10.1038/s41586-020-2130-2>
- [128] R. Bartolo, R. C. Saunders, A. R. Mitz, and B. B. Averbeck, “Information-limiting correlations in large neural populations,” *Journal of Neuroscience*, vol. 40, no. 8, pp. 1668–1678, 2020. [Online]. Available: <https://www.jneurosci.org/content/40/8/1668>
- [129] Z. Bowen, D. E. Winkowski, and P. O. Kanold, “Functional organization of mouse primary auditory cortex in adult c57bl/6 and f1 (cbaxc57) mice.” *Scientific Reports*, vol. 10, no. 10905, 2020. [Online]. Available: <https://doi.org/10.1038/s41598-020-67819-4>
- [130] B. M. Yu, J. P. Cunningham, G. Santhanam, S. I. Ryu, K. V. Shenoy, and M. Sahani, “Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity,” *Journal of Neurophysiology*, vol. 102, no. 1, pp. 614–635, 2009. [Online]. Available: <https://doi.org/10.1152/jn.90941.2008>

- [131] A. S. Ecker, P. Berens, R. J. Cotton, M. Subramaniyan, G. H. Denfield, C. R. Cadwell, S. M. Smirnakis, M. Bethge, and A. S. Tolias, “State dependence of noise correlations in macaque primary visual cortex,” *Neuron*, vol. 82, no. 1, pp. 235 – 248, 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0896627314001044>
- [132] E. N. Brown, R. Barbieri, V. Ventura, R. E. Kass, and L. M. Frank, “The time-rescaling theorem and its application to neural spike train data analysis,” *Neural Comput.*, vol. 14, no. 2, pp. 325–346, Feb. 2002. [Online]. Available: <https://doi.org/10.1162/08997660252741149>
- [133] A. H. Lipkus, “A proof of the triangle inequality for the tanimoto distance,” *Journal of Mathematical Chemistry*, vol. 26, no. 1, pp. 263–265, Oct 1999. [Online]. Available: <https://doi.org/10.1023/A:1019154432472>
- [134] J. F. Linden, R. C. Liu, M. Sahani, C. E. Schreiner, and M. M. Merzenich, “Spectrotemporal structure of receptive fields in areas ai and aaf of mouse auditory cortex,” *Journal of neurophysiology*, vol. 90, no. 4, pp. 2660–2675, 2003, pMID: 12815016. [Online]. Available: <https://doi.org/10.1152/jn.00751.2002>
- [135] M. R. DeWeese, M. Wehr, and A. M. Zador, “Binary spiking in auditory cortex,” *Journal of Neuroscience*, vol. 23, no. 21, pp. 7940–7949, 2003. [Online]. Available: <https://www.jneurosci.org/content/23/21/7940>
- [136] E. Petrus, A. Isaiah, A. P. Jones, D. Li, H. Wang, H.-K. Lee, and P. O. Kanold, “Crossmodal induction of thalamocortical potentiation leads to enhanced information processing in the auditory cortex,” *Neuron*, vol. 81, no. 3, pp. 664 – 673, 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S089662731301091X>
- [137] X. Meng, J. P. Kao, H.-K. Lee, and P. O. Kanold, “Intracortical circuits in thalamorecipient layers of auditory cortex refine after visual deprivation,” *Eneuro*, vol. 4, no. 2, 2017. [Online]. Available: <https://www.eneuro.org/content/4/2/ENEURO.0092-17.2017>
- [138] X. Meng, D. E. Winkowski, J. P. Kao, and P. O. Kanold, “Sublaminar subdivision of mouse auditory cortex layer 2/3 based on functional translaminar connections,” *Journal of Neuroscience*, vol. 37, no. 42, pp. 10 200–10 214, 2017. [Online]. Available: <https://www.jneurosci.org/content/37/42/10200>
- [139] S. Romero, A. E. Hight, K. K. Clayton, J. Resnik, R. S. Williamson, K. E. Hancock, and D. B. Polley, “Cellular and widefield imaging of sound frequency organization in primary and higher order fields of the mouse auditory cortex,” *Cerebral Cortex*, vol. 30, no. 3, pp. 1603–1622, 2020. [Online]. Available: <https://europepmc.org/articles/PMC7132909>
- [140] M. B. Kratz and P. B. Manis, “Spatial organization of excitatory synaptic inputs to layer 4 neurons in mouse primary auditory cortex,” *Frontiers in Neural Circuits*, vol. 9, p. 17, 2015. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fncir.2015.00017>
- [141] A. Forli, D. Vecchia, N. Binini, F. Succol, S. Bovetti, C. Moretti, F. Nespoli, M. Mahn, C. A. Baker, M. M. Bolton *et al.*, “Two-photon bidirectional control and imaging of

- neuronal excitability with high spatial resolution in vivo,” *Cell reports*, vol. 22, no. 11, pp. 3087–3098, 2018. [Online]. Available: <https://doi.org/10.1016/j.celrep.2018.02.063>
- [142] D. A. Martin, T. L. Ribeiro, S. A. Cannas, T. S. Grigera, D. Plenz, and D. R. Chialvo, “Box-scaling as a proxy of finite-size correlations,” 2020. [Online]. Available: <https://arxiv.org/abs/2007.08236>
- [143] B. Kadirvelu, Y. Hayashi, and S. J. Nasuto, “Inferring structural connectivity using ising couplings in models of neuronal networks,” *Scientific Reports*, vol. 7, no. 1, aug 2017.
- [144] S. Yu, H. Yang, H. Nakahara, G. S. Santos, D. Nikolić, and D. Plenz, “Higher-order interactions characterized in cortical activity,” *Journal of Neuroscience*, vol. 31, no. 48, pp. 17 514–17 526, 2011. [Online]. Available: <https://www.jneurosci.org/content/31/48/17514>
- [145] C. Wang and D. M. Blei, “Variational inference in nonconjugate models,” *J. Mach. Learn. Res.*, vol. 14, no. 1, pp. 1005–1031, Apr. 2013.
- [146] J. Windle and C. M. Carvalho, “A Tractable State-Space Model for Symmetric Positive-Definite Matrices,” *Bayesian Analysis*, vol. 9, no. 4, pp. 759 – 792, 2014. [Online]. Available: <https://doi.org/10.1214/14-BA888>
- [147] M. J. Beal, “Variational algorithms for approximate Bayesian inference,” Ph.D. dissertation, University of London, University College London (United Kingdom), 2003. [Online]. Available: <https://discovery.ucl.ac.uk/id/eprint/10101435>
- [148] W. K. Hastings, “Monte Carlo sampling methods using Markov chains and their applications,” *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970. [Online]. Available: <https://www.jstor.org/stable/2334940>
- [149] J. W. Pillow and J. Scott, “Fully bayesian inference for neural models with negative-binomial spiking,” in *Advances in Neural Information Processing Systems* 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1898–1906. [Online]. Available: <http://papers.nips.cc/paper/4567-fully-bayesian-inference-for-neural-models-with-negative-binomial-spiking.pdf>
- [150] S. Linderman, R. P. Adams, and J. W. Pillow, “Bayesian latent structure discovery from multi-neuron recordings,” in *Advances in Neural Information Processing Systems*, 2016, pp. 2002–2010. [Online]. Available: <http://papers.nips.cc/paper/6185-bayesian-latent-structure-discovery-from-multi-neuron-recordings.pdf>
- [151] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.
- [152] S. Arlot, A. Celisse *et al.*, “A survey of cross-validation procedures for model selection,” *Statistics surveys*, vol. 4, pp. 40–79, 2010. [Online]. Available: <https://doi.org/10.1214/09-SS054>
- [153] J. Ding, V. Tarokh, and Y. Yang, “Model selection techniques: An overview,” *IEEE Signal Processing Magazine*, vol. 35, no. 6, pp. 16–34, 2018.

- [154] R. D. Frisina, A. Singh, M. Bak, S. Bozorg, R. Seth, and X. Zhu, “F1 (cba× c57) mice show superior hearing in old age relative to their parental strains: Hybrid vigor or a new animal model for “golden ears”?” *Neurobiology of aging*, vol. 32, no. 9, pp. 1716–1724, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0197458009003157>
- [155] J. Liu, M. R. Whiteway, A. Sheikhattar, D. A. Butts, B. Babadi, and P. O. Kanold, “Parallel processing of sound dynamics across mouse auditory cortex via spatially patterned thalamic inputs and distinct areal intracortical circuits,” *Cell Reports*, vol. 27, no. 3, pp. 872 – 885.e7, 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2211124719303997>
- [156] Z. Bowen, D. E. Winkowski, S. Seshadri, D. Plenz, and P. O. Kanold, “Neuronal avalanches in input and associative layers of auditory cortex,” *Frontiers in Systems Neuroscience*, vol. 13, p. 45, 2019. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fnsys.2019.00045>
- [157] P. Watkins, J. Kao, and P. Kanold, “Spatial pattern of intra-laminar connectivity in supragranular mouse auditory cortex,” *Frontiers in Neural Circuits*, vol. 8, 2014. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fncir.2014.00015>
- [158] U. Stalman, A. J. Franke, H. Al-Moyed, N. Strenzke, and E. Reisinger, “Otoferlin is required for proper synapse maturation and for maintenance of inner and outer hair cells in mouse models for dfnb9,” *Frontiers in Cellular Neuroscience*, vol. 15, 2021. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fncel.2021.677543>
- [159] P. J. Keller and M. B. Ahrens, “Visualizing whole-brain activity and development at the single-cell level using light-sheet microscopy,” *Neuron*, vol. 85, no. 3, pp. 462–483, 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0896627314011489>
- [160] T. W. Dunn, Y. Mu, S. Narayan, O. Randlett, E. A. Naumann, C.-T. Yang, A. F. Schier, J. Freeman, F. Engert, and M. B. Ahrens, “Brain-wide mapping of neural activity controlling zebrafish exploratory locomotion,” *eLife*, vol. 5, p. e12741, mar 2016. [Online]. Available: <https://doi.org/10.7554/eLife.12741>
- [161] X. Chen, Y. Mu, Y. Hu, A. T. Kuan, M. Nikitchenko, O. Randlett, A. B. Chen, J. P. Gavornik, H. Sompolsky, F. Engert, and M. B. Ahrens, “Brain-wide organization of neuronal activity and convergent sensorimotor transformations in larval zebrafish,” *Neuron*, vol. 100, no. 4, pp. 876–890.e5, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0896627318308444>
- [162] N. Vladimirov, C. Wang, B. Höckendorf, A. Pujala, M. Tanimoto, Y. Mu, C.-T. Yang, J. D. Wittenbach, J. Freeman, S. Preibisch, M. Koyama, P. J. Keller, and M. B. Ahrens, “Brain-wide circuit interrogation at the cellular level guided by online analysis of neuronal function,” *Nature methods*, vol. 15, p. 1117, 2018. [Online]. Available: <https://doi.org/10.1038/s41592-018-0221-x>

- [163] M. B. Ahrens, J. M. Li, M. B. Orger, D. N. Robson, A. F. Schier, F. Engert, and R. Portugues, “Brain-wide neuronal dynamics during motor adaptation in zebrafish,” *Nature*, vol. 483, p. 471, 2012. [Online]. Available: <https://doi.org/10.1038/nature11057>
- [164] M. Privat, S. A. Romano, T. Pietri, A. Jouary, J. Boulanger-Weill, N. Elbaz, A. Duchemin, D. Soares, and G. Sumbre, “Sensorimotor transformations in the zebrafish auditory system,” *Current Biology*, vol. 29, no. 23, pp. 4010–4023.e4, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0960982219313600>
- [165] J. Smith, H. Ellenberger, K. Ballanyi, D. Richter, and J. Feldman, “Pre-bötzinger complex: a brainstem region that may generate respiratory rhythm in mammals,” *Science*, vol. 254, no. 5032, pp. 726–729, 1991. [Online]. Available: <https://science.sciencemag.org/content/254/5032/726>
- [166] J. C. Smith, A. P. L. Abdala, H. Koizumi, I. A. Rybak, and J. F. R. Paton, “Spatial and functional architecture of the mammalian brain stem respiratory network: A hierarchy of three oscillatory mechanisms,” *Journal of Neurophysiology*, vol. 98, no. 6, pp. 3370–3387, 2007, pMID: 17913982. [Online]. Available: <https://doi.org/10.1152/jn.00985.2007>
- [167] J.-M. Ramirez and N. Baertsch, “Defining the rhythmogenic elements of mammalian breathing,” *Physiology*, vol. 33, no. 5, pp. 302–316, 2018, pMID: 30109823. [Online]. Available: <https://doi.org/10.1152/physiol.00025.2018>
- [168] A. Daffertshofer, R. Huys, and P. J. Beek, “Dynamical coupling between locomotion and respiration,” *Biological Cybernetics*, vol. 90, no. 3, pp. 157–164, Mar 2004. [Online]. Available: <https://doi.org/10.1007/s00422-004-0462-x>
- [169] C. P. Hoffmann, G. Torregrosa, and B. G. Bardy, “Sound stabilizes locomotor-respiratory coupling and reduces energy cost,” *PLOS ONE*, vol. 7, no. 9, pp. 1–12, 09 2012. [Online]. Available: <https://doi.org/10.1371/journal.pone.0045206>
- [170] C. P. Hoffmann and B. G. Bardy, “Dynamics of the locomotor–respiratory coupling at different frequencies,” *Experimental Brain Research*, vol. 233, no. 5, pp. 1551–1561, May 2015. [Online]. Available: <https://doi.org/10.1007/s00221-015-4229-5>
- [171] A. Sheikhattar, S. Miran, J. Liu, J. B. Fritz, S. A. Shamma, P. O. Kanold, and B. Babadi, “Extracting neuronal functional network dynamics via adaptive granger causality analysis,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 17, pp. E3869–E3878, 2018.
- [172] S. Mukherjee and B. Babadi, “Dynamic analysis of higher-order coordination in neuronal assemblies via de-sparsified orthogonal matching pursuit,” in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34. Curran Associates, Inc., 2021, pp. 4120–4133. [Online]. Available: <https://proceedings.neurips.cc/paper/2021/file/2172fde49301047270b2897085e4319d-Paper.pdf>

- [173] O. Randlett, C. L. Wee, E. A. Naumann, O. Nnaemeka, D. Schoppik, J. E. Fitzgerald, R. Portugues, A. M. B. Lacoste, C. Riegler, F. Engert, and A. F. Schier, “Whole-brain activity mapping onto a zebrafish brain atlas,” *Nature Methods*, vol. 12, no. 11, pp. 1039–1046, Nov 2015. [Online]. Available: <https://doi.org/10.1038/nmeth.3581>
- [174] R. Wong, *Asymptotic Approximations of Integrals*. Society for Industrial and Applied Mathematics, 2001. [Online]. Available: <https://epubs.siam.org/doi/abs/10.1137/1.9780898719260>
- [175] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*. OUP Oxford, 2013.