Publicly Accessible Penn Dissertations

2013

Millimeter-Scale Epileptiform Spike Propagation Patterns and Their Relationship to Seizures

Ann Vanleer

University of Pennsylvania, achamb@seas.upenn.edu

Follow this and additional works at: <https://repository.upenn.edu/edissertations>

Part of the [Biomedical Commons](#)

Recommended Citation

Vanleer, Ann, "Millimeter-Scale Epileptiform Spike Propagation Patterns and Their Relationship to Seizures" (2013). *Publicly Accessible Penn Dissertations*. 934.
<https://repository.upenn.edu/edissertations/934>

This paper is posted at ScholarlyCommons. <https://repository.upenn.edu/edissertations/934>
For more information, please contact repository@pobox.upenn.edu.

Millimeter-Scale Epileptiform Spike Propagation Patterns and Their Relationship to Seizures

Abstract

Currently, clinical electrode arrays with a sparse spatial density (1 cm) are used to map the seizure onset zone (SOZ) and epileptic network in patients prior to epilepsy surgery. However, recent research demonstrates that submillimeter, cortical-column-scale domains have a role in seizure generation that may be clinically significant. We used novel high-resolution (500 μm), active, flexible, 1-cm², surface, electrode grid-arrays to explore the behavior of these domains. We employed this new technology to analyze the two-dimensional (2-D), wave-propagation patterns of epileptiform local field potential spikes (LFP spikes). Subdural micro-electrocorticographic (μECoG) signals were recorded in vivo from anesthetized cats. A GABA antagonist, picrotoxin, was applied to induce acute neocortical epileptiform activity leading up to discrete seizures. Nine hours of data yielding 26,331 LFP spikes was analyzed. Features characteristic of spatio-temporal (ST) patterns were extracted from these events and k-medians clustering was employed to separate the data into 10 distinct classes. We tested the hypothesis that 2-D spike patterns during seizures (ictal spikes) are different from those between seizures (interictal spikes). A permutation test ($n=1,000,000$) confirmed this hypothesis. A frequent episode discovery algorithm (Temporal Data Mining) was then applied to investigate the relationship of sequences of these patterns to seizure generation, progression and termination. We found that sub-millimeter-scale ST spike wave-propagation patterns reveal network dynamics that may elucidate mechanisms underlying local circuit activity generating seizures. We conclude that sequences of patterns of similar type are less likely to precede seizure generation than sequences of patterns of differing types. Temporal analysis of these patterns also suggests that seizures in this model are not initiated by a single 2-D pathway, but rather by a number of different ST-initiating events. While these findings may be model-specific, reflecting diffusion of picrotoxin across the feline neocortex, the tools we have developed to interpret these events are directly portable to the human condition. We are confident that recording LFP spike ST wave-propagation patterns at high resolution provides a fruitful direction for continued analysis of epileptiform network dynamics and we propose that further study may provide a novel opportunity for therapeutic intervention at the micro-scale to treat epilepsy.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Bioengineering

First Advisor

Brian Litt

Subject Categories

Biomedical

MILLIMETER-SCALE EPILEPTIFORM
SPIKE PROPAGATION PATTERNS
AND THEIR RELATIONSHIP TO SEIZURES

Ann C. Vanleer

A DISSERTATION

in

Bioengineering

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2013

Supervisor of Dissertation

Brian Litt, MD
Professor of Bioengineering and
Neurology

Graduate Group Chairperson

Jason Burdick, PhD
Professor of Bioengineering

Dissertation Committee:

Gershon Buchsbaum, PhD
Professor of Bioengineering

Diego Contreras, MD PhD
Professor of Neuroscience

Marc Dichter, MD PhD
Emeritus Professor of Neurology

Abba M. Krieger, PhD
Professor of Statistics and Operations
Research

Jonathan Viventi, PhD
Assistant Professor of Electrical and
Computer Engineering and Neural
Science

MILLIMETER-SCALE EPILEPTIFORM
SPIKE PROPAGATION PATTERNS
AND THEIR RELATIONSHIP TO SEIZURES
COPYRIGHT © 2013 Ann C. Vanleer

This work is licensed under the
Creative Commons Attribution-
NonCommercial-ShareAlike 3.0
License
To view a copy of this license, visit

<http://creativecommons.org/licenses/by-nc-sa/3.0/>

To Dad

Who taught me to trust in the slow work of God

Acknowledgment

I would first like to thank Dr. Brian Litt for his encouragement and mentorship through these last five years. I would also like to thank the members of my committee: Diego Contreras, Gershon Buchsbaum, Marc Dichter, Abba Krieger and Jon Viventi. Next I would like to thank the lab mates and friends that I have had the pleasure to know and work with throughout this endeavor, to include: Justin Blanco, Leif Vigeland, Tom Murphy, Drausin Wulsin, Mark Lippmann, Bill Stacey, Ruthie Krieger, Carolyn Wilkinson, Jeff Keating, Hal Juul, Kate Davis, Duygu Kuzum, Joost Wagenaar, Hank Bink, Ankit Khambhati, Cat Rojvirat, Maria Murray, Stephanie Majkut, Kristy Simmons, Tina Mazzanti, Alyce Fernebok and Joe McGowan. Finally and most importantly I would like to thank my family for their unconditional support. Thank you especially to: Susan Chamberlain, Catherine Chamberlain, Robert Chamberlain, Mom and Dad. And last but most definitely not least, thank you to Brian Vanleer and Bryson Vanleer. I am so fortunate to have you both in my corner!

ABSTRACT

MILLIMETER-SCALE EPILEPTIFORM SPIKE PROPAGATION PATTERNS AND THEIR RELATIONSHIP TO SEIZURES

Ann C. Vanleer

Brian Litt, MD

Currently, clinical electrode arrays with a sparse spatial density (1 cm) are used to map the seizure onset zone (SOZ) and epileptic network in patients prior to epilepsy surgery. However, recent research demonstrates that submillimeter, cortical-column-scale domains have a role in seizure generation that may be clinically significant. We used novel high-resolution (500 μm), active, flexible, 1-cm², surface, electrode grid-arrays to explore the behavior of these domains. We employed this new technology to analyze the two-dimensional (2-D), wave-propagation patterns of epileptiform local field potential spikes (LFP spikes). Subdural micro-electrocorticographic (μECoG) signals were recorded *in vivo* from anesthetized cats. A GABA antagonist, picrotoxin, was applied to induce acute neocortical epileptiform activity leading up to discrete seizures. Nine hours of data yielding 26,331 LFP spikes was analyzed. Features characteristic of spatio-temporal (ST) patterns were extracted from these events and k-medians clustering was employed to separate the data into 10 distinct classes. We tested the hypothesis that 2-D spike patterns during seizures (ictal spikes) are different from those between seizures (interictal spikes). A permutation test ($n=1,000,000$) confirmed this hypothesis. A frequent episode discovery algorithm (Temporal Data Mining) was then applied to investigate the relationship of sequences of these patterns to seizure generation, progression and termination. We found that sub-millimeter-scale ST spike wave-

propagation patterns reveal network dynamics that may elucidate mechanisms underlying local circuit activity generating seizures. We conclude that sequences of patterns of similar type are less likely to precede seizure generation than sequences of patterns of differing types. Temporal analysis of these patterns also suggests that seizures in this model are not initiated by a single 2-D pathway, but rather by a number of different ST-initiating events. While these findings may be model-specific, reflecting diffusion of picrotoxin across the feline neocortex, the tools we have developed to interpret these events are directly portable to the human condition. We are confident that recording LFP spike ST wave-propagation patterns at high resolution provides a fruitful direction for continued analysis of epileptiform network dynamics and we propose that further study may provide a novel opportunity for therapeutic intervention at the micro-scale to treat epilepsy.

TABLE OF CONTENTS

Acknowledgment	iv
Abstract	v
List of Tables	xi
List of Figures	xii
1 Motivation.....	1
1.1 Problem and Clinical Significance	1
1.2 Dissertation Goals and Roadmap	6
2 Background.....	9
2.1 History of Research and Clinical EEG Arrays	9
2.2 Recording Neural Activity with an Active Electrode Array	14
2.3 Additional Sensor Modalities Utilized in the Study of Epilepsy	19
2.4 Methods of EEG Data Analysis in Epilepsy	22
2.5 Conclusion	26

3 Description of the Algorithm..... 27

3.1	Summary.....	27
3.2	Introduction	29
3.3	Methods	34
3.3.1	Animal Model.....	34
3.3.2	Data Acquisition.....	35
3.3.3	Overview of Data Analysis.....	37
3.3.4	Pre-processing.....	37
3.3.5	LFP Spike Detection	39
3.3.6	Feature Extraction	41
3.3.7	Clustering	44
3.4	Results.....	46
3.5	Discussion	49

4 Statistical Testing..... 53

4.1	Summary.....	53
4.2	Introduction	54
4.3	Methods	59
4.3.1	The Gap Statistic	59
4.3.2	The Permutation Test	60
4.4	Results.....	61
4.5	Discussion	66

5 Temporal Data Mining..... 68

5.1	Summary.....	68
5.2	Introduction	70
5.3	Methods	71
5.3.1	Data description.....	71
5.3.2	TDM-5 algorithm.....	72
5.3.3	TDM-2 (Analysis of Transition Matrices).....	75
5.3.4	Various correlations across SZ-aligned data	76
5.4	Results.....	80
5.4.1	Most frequent TDM sequences of length 5.....	80
5.4.2	TDM sequences of length 2	86
5.4.3	Correlations across SZ-aligned data	91
5.4.4	Spike rasters of seizure-aligned windows (Cat 3a, 3b data).....	95
5.5	Discussion	98

6 Conclusion 103

6.1	Summary of Findings.....	103
6.2	Insights and Future Directions	106
6.2.1	Continued application of developed algorithms	106
6.2.2	Adaptation of Schiff-Wilson-Cowan computational model.....	107
6.2.3	Extension of analysis to integrate simultaneous depth recordings	108
6.2.4	Inform stimulation protocol in implantable devices	110

Bibliography	112
--------------------	-----

List of Tables

Table 3.1 Collected data summary	36
Table 3.2 PCA dimensions retained capturing 99% of variance in the original data	44
Table 4.1 Cat 1, 10 cluster permutation test results	64
Table 4.2 Cat 2, 10 cluster permutation test results	64
Table 4.3 Cat 3a, 10 cluster permutation test results	65
Table 4.4 Cat 3b, 10 cluster permutation test results	65
Table 5.1 TDM-5 data summary	84

List of Figures

Figure 1.1 Photographs of 2 different recording arrays	3
Figure 1.2 Hybrid subdural grid array	4
Figure 2.1 Schematic circuit diagram of four unit tiles of Viventi array	15
Figure 2.2 Multiple layers of circuitry in tiles of Viventi array	17
Figure 2.3 Photograph of high spatial density array used in recordings	18
Figure 3.1 Representative average spike waveforms following spike detection.....	32
Figure 3.2 Delay maps for detected spikes closest to the centers of 16 clusters.....	32
Figure 3.3 Schematic of the experimental set-up.....	36
Figure 3.4 Block diagram of an overview of the data analysis algorithm.....	37
Figure 3.5 Electrode yields for the arrays used during the experiments for Cats 1-3	38
Figure 3.6 Average micro-ECoG trace for Cat 1	38
Figure 3.7 Spike detection resulting from a voltage threshold crossing	41
Figure 3.8 Delay and power maps resulting from the spike detected in figure 3.7.....	42
Figure 3.9 Clustering results from the 2894 spikes of Cat 1	46
Figure 3.10 Combined delay/power maps of spikes closest to cluster centers, Cat 1	47
Figure 3.11 Combined delay/power maps of spikes closest to cluster centers, Cat 2	47

Figure 3.12 Combined delay/power maps of spikes closest to cluster centers, Cat 3a	47
Figure 3.13 Combined delay/power maps of spikes closest to cluster centers, Cat 3b	48
Figure 3.14 Cluster homogeneity of delay maps from Cat 1	48
Figure 4.1 Tibshirani's Gap Statistic applied to toy dataset	55
Figure 4.2 Tibshirani's Gap Statistic applied to uniform toy dataset	57
Figure 4.3 Tibshirani Gap Statistic applied to Cat 1 dataset.....	62
Figure 4.4 Ratios of inter-ictal or ictal spikes to total spikes, by cluster	63
Figure 5.1 Spike detection rasters for each of the four datasets (Cat 1-3b)	74
Figure 5.2 Seizure-start aligned spike rasters for Cats 1 and 2	77
Figure 5.3 Most frequent TDM-5 pattern rasters (Cats 1-3b).....	83
Figure 5.4 Most frequent TDM-5 'state' rasters (Cat 1)	85
Figure 5.5 Transition matrix for spikes in the Cat 1 dataset	86
Figure 5.6 Relationship of repeating and non-repeating spike transitions to seizure onset	87
Figure 5.7 Relationship of spike transitions to individual SZ onsets (Cat 1).....	90
Figure 5.8 First and second correlation metric results for Cat 1	91
Figure 5.9 Third correlation metric result for Cat 1	93
Figure 5.10 Fourth correlation metric result for Cat 1	94
Figure 5.11 Raw traces average for most correlated seizure windows from Cat 1 using correlation metric 4	95
Figure 5.12 Spike rasters from Cat 3a and 3b depicting change in cluster composition of seizures as status epilepticus develops	97

1 Motivation

1.1 Problem and Clinical Significance

Epilepsy is a disorder characterized by seizures, globally synchronous electrical activity in the brain, which disrupt its normal functioning [1, 2]. Millions of people suffer from epilepsy worldwide; clinically, 68% of patients with epilepsy respond well to anti-epileptic drugs, leaving 32% who are medically refractory and must seek surgery or experimental therapies to treat their symptoms [3, 4]. Surgery is successful in reducing the occurrence of seizures in 44% of medically refractory patients [5]. However, surgery is not an option for many patients either because their epileptic network cannot be localized, their epileptic network consists of multiple wide-spread regions, or because the patient is at a high risk for incurring unacceptable neurological damage as a consequence of resective surgery [6]. In addition, potential deficits from epilepsy surgery, most

commonly cognitive or psychiatric deficits, understandably deter some patients from assuming the risks associated with these irreversible procedures. The above facts aggressively drive research efforts to improve surgical techniques and to better inform new non-surgical therapies. Improvement in the clinical treatment of patients with epilepsy begins with the need for a better understanding of seizure generation and progression [1, 7].

Despite best efforts, attempts to understand the neuro-electrical mechanisms of action resulting in seizure generation, progression and termination using standard clinical data, have reached a plateau. Recent advances in technology provide us with improved tools, specifically novel electrode arrays capable of resolving brain function at high temporal and spatial resolution, to map and modulate brain function. These technologies generate huge amounts of data, impenetrable using standard manual methods applied in clinical care. Rather, these new technologies naturally lend themselves to new avenues of data mining to shed light on the nature of changes in brain network electrophysiology leading to the epileptic state and full-blown clinical seizures.

The Electroencephalogram (EEG) is used in clinical evaluation to guide epilepsy surgery; however, the optimal electrophysiological scale to sense epileptic networks is the subject of debate [8]. EEG voltages result from summed synchronous synaptic activity in the cortical regions under surface or intracranial recording electrodes [7, 9]. Prior to epilepsy surgery, clinicians map the “seizure onset zone” (SOZ), “epileptogenic zone” and “seizure focus” by implanting subdural or intra-cortical electrodes (iEEG/ECOG).

Currently, electrode arrays with a sparse spatial density (1 cm spacing) are used for these implants [9]. Figure 1.1.A depicts a standard clinical “grid”

Figure 1.1 Photographs of 2 different recording arrays

A. Photograph of a standard clinical subdural grid-array commonly used prior to epilepsy surgery to map the “seizure onset zone,” “epileptogenic zone” and “seizure focus.” These electrode arrays have a sparse spatial density (1 cm); the size of the electrodes in these arrays is 0.5 cm in diameter. Photo Credit: University of Utah, Department of Neurosurgery. **B.** A photograph of a Utah array with its 100 penetrating electrodes (top) and a top-down view of this same array next to a US penny to illustrate relative size [10].

electrode array. Modest successes in achieving seizure freedom in all but lesional epilepsies (where larger strides have been made), have led clinical investigators to question standard centimeter-scale ECoG as the best spatial resolution for mapping epileptic networks and their dynamics.

Advances in neural electrode technology enable brain recordings with increasingly fine spatial and temporal resolution. Recent landmark studies utilize the Utah array (covering an area of 4mm x 4mm) [11, 12], which records multi-unit activity (MUA) from 100 penetrating microelectrodes spaced 400μm apart. Figure 1.1.B depicts a Utah array. Still other studies utilize research subdural grids (covering an area of 3.4cm x 5.4cm) that contain up to 128, 40-micron diameter microwires in groups of 4-16, with a spatial

density of 1.0mm, interspersed between 16-24 macro-electrodes (1cm spacing, 0.4cm in diameter) [13]. Figure 1.2 depicts this type of hybrid array. Both of these types of arrays (figure 1.1.B and figure 1.2) are currently confined to research applications. Findings

Figure 1.2 Hybrid subdural grid array

A. Schematic of a hybrid subdural grid array containing 16 macro-electrodes (1 cm spacing, 0.4 cm in diameter) interspersed with 128 40-micron diameter microwires in groups of 12 and 16, with a spatial density of 1.0 mm. **B.** A photograph of the hybrid subdural grid array schematisized on the left [13].

from these investigations indicate that seizures appear to be generated from dispersed microdomains on the scale of distinct cortical columns [8] or neuronal ensembles [11, 14, 15] arguing for electrode arrays with high, sub-millimeter scale, spatial density. Despite this work, a clear picture of the process of seizure generation has not emerged, perhaps because sampling is too spatially constrained in multi-unit recordings (see relative size of Utah array in figure 1.1.B) and too sparsely sampled in local field potential (LFP) studies to understand a more global network process (figure 1.1.A and figure 1.2). Additionally, subdural grids are limited in the number of 40-micron LFP-recording electrodes they can contain, as they require 1 wire per electrode contact.

This work leverages a new, high-spatial density subdural active electrode array of 360 channels covering an area of 10mm x 9mm to measure LFP-scale electrical signals *in vivo* [16]. The electrode size and spacing is 300 μ m x 300 μ m and 500 μ m respectively; additionally, the multiplexed design of the array eliminates the requirement that each contact be individually wired to data recording circuitry. This electrode array is described in more detail in section 2.2. Data recorded and analyzed in this dissertation are from an acute *in vivo* model of epilepsy in the cat.

Increases in electrode density and sampling frequency introduce new challenges in data analysis due to the increased amount of data to be processed. However, increases in spatial and temporal resolution also introduce new opportunities to understand neural dynamics between closely spaced adjacent areas on the cortical surface of the brain. Traditional analysis techniques, such as tracking waveform morphology and spectral features of the recorded signal, may be less enlightening in this setting, and their computational overhead may be prohibitive when applied to datasets that span multiple terabytes. Machine learning algorithms are ideally suited to extract information from extremely large datasets and present the results in an easily interpretable way [17].

To address the challenges presented by larger datasets, this dissertation applies machine learning techniques to extract features from and further reduce the dimensionality of the data. The algorithm presented below also implements unsupervised clustering to reveal two-dimensional (2-D) patterns that might otherwise escape detection. Representing

pathological neuro-electrical discharges as 2-D spatio-temporal wave-propagation patterns provides a new opportunity to investigate mechanisms of action for seizure generation, progression and termination.

1.2 Dissertation Goals and Roadmap

Both the clinical utility of high-density electrodes and the novel signal analysis algorithms they make possible will ultimately need to be validated. They must increase our understanding of the neural mechanisms underlying epileptic network dynamics and advance the diagnostic and therapeutic reach of clinicians. It is hoped that results of research studies utilizing large high-density electrode grid-arrays will be ultimately incorporated in closed-loop feedback devices capable of anticipating and arresting abnormal electrical activity in the cortex. However, there are many challenges to be overcome before such goals can be achieved. These challenges include: developing signal analysis algorithms to efficiently leverage the dramatic increase in the amount of data collected and to extract relevant information about underlying electrical activity; developing hypotheses of the mechanisms of action driving observed abnormal electrical discharges in analyzed data; and developing computational models based on these hypotheses that can expand our understanding and guide the design of new therapies. The ultimate proof, of course, is to employ these results and models based upon them, to demonstrate effective stimulation protocols to prevent or arrest seizures in animal models and humans.

This dissertation addresses a few of these opening challenges. It is organized in the following manner. Chapter 2 is a concise review of the scientific literature on the recording of pathological signals of epilepsy from various modalities; the application of various signal analysis algorithms for relevant feature extraction from EEG recordings; and the variety of available ECoG electrode arrays currently in use clinically as well as in neuro-engineering and neuroscience research labs. The review is not comprehensive since each succeeding chapter contains a brief introduction section discussing in further detail the aspects of current electrode array technology and/or signal analysis algorithm development that are pertinent to the specific chapter. Chapter 2 introduces the reader to the history of the development of EEG electrode arrays culminating in our choice of array that is used in this thesis. In addition it touches briefly on the aspects of our increasing understanding of EEG signals via improved analysis techniques that have broadened the use of these arrays, both clinically and in research. There is also a brief discussion describing why these two arenas have diverged slightly in their choice of preferred electrode array. I believe the array chosen for this thesis has distinct advantages over standard electrodes used both in the research and clinical settings.

Chapter 3 introduces a new signal analysis algorithm for representing and classifying detected epileptiform LFP spikes from data collected using novel electrode grid-arrays. The core of the techniques employed are not novel; however, the manner in which these techniques are collectively applied to extract relevant information is unique and original. This application selects spatio-temporal features from the data that leverages the high-spatial density of the recordings.

Chapter 4 presents the statistical methods employed to test hypotheses formulated from the results of the data analysis in Chapter 3. Chapter 4 presents new findings about the spatio-temporal patterns of epileptiform spikes detected and their classification into a reduced and separable set of stereotypical patterns. We also present findings that subsets of these patterns indicate the ictal vs. non-ictal states (a difference not distinguishable by the temporal morphology of the epileptic waveforms detected by current clinical electrode arrays.)

Chapter 5 employs temporal data mining (frequent episode discovery) techniques to our experimental data and discusses insight gleaned from observing *sequences* of spatio-temporal patterns of detected spikes. We test the hypothesis of the existence of an underlying sequence of neuro-electrical events characteristic of a transition of a cortical epileptic network from a seizure resistant state to a seizure prone state, increasing the susceptibility of the underlying cortex to seizures. Chapter 6 concludes with a discussion of future extensions of this work.

2 Background

2.1 History of Research and Clinical EEG Arrays

Electroencephalography (EEG) records electrical signals from the brain. EEG signals are generated when neurons transport ions across their membranes during normal brain activity. Scalp EEG senses post-synaptic potentials resulting from the summed synchronous activity of vertically oriented pyramidal cell cell bodies and dendrites in cortical layers 3 to 5 [18]. Extracellular ionic currents caused by dendritic electrical activity results in differential electrical potentials between EEG electrodes and a reference electrode. These local field potentials (LFPs) are recorded as EEG voltages [19].

Because EEG directly measures brain activity (as opposed to a correlate of brain activity such as cerebral blood flow or glucose metabolism, etc.), it has the advantage of the highest temporal resolution of all imaging modalities (on the order of tens of milliseconds in the case of intracranial EEG). The temporal resolution is limited only by the analog-to-digital sampling rate of the data acquisition system used to record the brain signals. And although magnetoencephalography (MEG) also has excellent temporal resolution, it has the disadvantage of a lower signal-to-noise ratio (SNR) compared to EEG, because magnetic fields sensed by the magnetometer experience more decay than electric fields, with increasing distance from the source of brain activity [20]. EEG has the added advantage that it can be recorded from ambulatory subjects in a natural environment as opposed to the bulky machinery required by other imaging modalities such as MEG and functional magnetic resonance imaging (fMRI). This allows for the future adaptation of any advances in signal analysis to be integrated into a portable and/or implantable diagnostic, monitoring and/or therapeutic device.

Historically, the disadvantage of EEG has been its limited spatial resolution. Hans Berger first introduced scalp EEG for human recordings in 1924 [21]. However, scalp EEG is limited in its spatial resolution (~10mm, inferior to other imaging modalities) because of distortions of electrical potentials by the skull, scalp, cerebrospinal fluid (CSF) and soft tissues [9, 19]. This prompted the pioneering work of Wilder Penfield and Herbert Jasper in the early 1950's at the Montreal Neurological Institute, developing intracranial EEG (iEEG) or electrocorticography (ECoG) where EEG electrodes now assembled in strips, grids or penetrating depth electrodes (also called “stereo EEG”) are surgically implanted

[22]. Today, scalp EEG continues to be the gold standard in the clinical diagnosis of epilepsy in the ambulatory clinical setting. However, because of its limited spatial resolution, more invasive techniques are used prior to any resective surgery for epilepsy and in functional surgery for tumors and other lesions. These techniques are now highly protocolized. For patients refractory to pharmaceutical therapies a high-resolution MRI of the brain is first obtained to determine if there are any clearly identifiable structural abnormalities. In cases where abnormalities cannot be easily identified or eloquent brain regions may be involved, iEEG grids and/or depth electrode strip arrays are usually implanted to record neural activity associated with seizures in order to map the putative brain networks involved prior to resective surgery [9, 19]. Scalp EEG remains the least invasive, but also least informative of clinical electrophysiology for epilepsy. Subdural and intra cortical EEG yield much higher signal quality and spatial resolution, but require surgery. Depth electrodes are the most invasive, creating tracks as electrodes are driven into the brain; however, they are usually required when the sources of epileptic activity are suspected to exist in deeper structures of the brain, though some centers preferentially use this technique (stereo EEG) to delineate neocortical networks preferentially by tradition.

One reason for an increased emphasis in research for using electrode arrays with an increased spatial resolution is to better inform epilepsy surgery. During resective surgery, due to a continued lack of understanding of the mechanisms of seizure generation in epilepsy, surgeons resect as large an area as is safe and prudent, taking care to minimize the chances of neurological deficits. The size of the resection continues to be the most

predictive factor of positive surgical outcome in terms of a reduction in seizure frequency or (in some cases) of complete seizure freedom. The need persists for a better understanding of epileptic networks to enable more informed resections that minimize morbidity, with equally positive outcomes. However, because of the morbidity of resective surgery, primarily cognitive and psychiatric dysfunction in addition to routine risks of surgery, it would be desirable to eliminate resective therapy when possible, and to replace it with safer alternatives. Intelligent, implantable neurostimulation devices hold great promise in this role. We hypothesize that the efficacy of such devices, the first generation of which are only now coming before the FDA, may be substantially enhanced by use in conjunction with high resolution implantable electrode grids.

In the research setting, electrophysiological studies record neural data from electrode grids that range in size and scale from scalp EEG to the Utah array, which records multi-unit activity from penetrating electrodes 400 μm apart [12]. A limitation of the Utah array is its total array size of 4mm x 4mm which limits studies using this array to recordings from very small regions of cortex. Worrell et al. [23] have demonstrated another approach, a standard clinical iEEG grid with macro-electrodes spaced 10 millimeters apart augmented with imbedded arrays of 40 micron microwires placed between the macro-electrodes [24]. However, these particular research micro-electrode or “hybrid” arrays are not able to have more than 16 (4 x 4) electrodes in a local area (3mm x 3mm) because of limitations in wiring density. While the above devices give improved spatial resolution over a few patches of cortex, they do not address the clinical need for devices to map epileptic networks over greater portions of the cortical surface (up to 64 cm^2 for

some clinical electrode grids). In cases of non-lesional extratemporal epilepsy the location of the seizure onset zone must be discerned without the help of structural abnormalities on MRI.

These factors are important drivers of newer electrode technologies capable of recording brain activity at high resolution over large, spatially distributed regions. Another reason for an increased emphasis on research using electrode arrays with an increased spatial resolution is recent findings suggesting that epileptic networks are composed of dispersed microdomains on the scale of cortical columns [8], driving development of devices with high spatial resolutions (on the order of 500 μ m).

As research surges forward to increase the spatial resolution of brain recording arrays there is also a concern that this must be done without sacrificing the temporal resolution of research arrays, which are capable of per channel sampling rates of 32kHz (e.g. the Cheetah system by Neuralynx, Bozeman Montana). Recent studies point towards the importance of detecting features characteristic of high-frequency oscillations (HFOs) to interpret the role these phenomena play in continuing to develop our understanding of epilepsy [25-28].

The array we chose to use in the experiments described below has the advantage of a high spatial density (500 μ m); it is subdural and therefore does not suffer from the signal dampening distortion affects present in scalp EEG; and it is less invasive in comparison to the Utah array. Another advantage over the Utah array is that the array we chose

contains 360 channels over $\sim 1 \text{ cm}^2$ as opposed to only 100 channels over 0.16 cm^2 . We chose an acute animal model of induced cortical seizures for the experiments described below. The theoretical limitation on the spatial resolution of research arrays imposed by the requirement that each electrode have a dedicated output wire has been addressed by the active electronic/ multiplexed design of the electrode array used in this work.

2.2 Recording Neural Activity with an Active Electrode Array

In this study, we use a new, high-spatial density subdural surface active electrode array of 360 channels covering an area of $10 \text{ mm} \times 9 \text{ mm}$ to measure local field potential (LFP)-scale electrical signals *in vivo* from an acute feline model of epilepsy. We apply high-density, flexible active electrode arrays, developed by Viventi et al. [16] to the task of collecting micro-electrocorticographic (μECoG) data continuously (hours) at a mesoscopic scale ($10 \text{ mm} \times 9 \text{ mm}$) at a high spatial resolution ($500 \mu\text{m}$). The ability to record simultaneously from hundreds of high-density electrodes over a large local area this size is unprecedented. Leveraging advances in the technology of high-performance, ultrathin and flexible silicon nanomembrane transistors, these arrays can record from 360 independent channels (arranged in 20 columns and 18 rows) spaced $500 \mu\text{m}$ apart without the constraint of needing to wire each sensor to data recording circuitry due to their advanced multiplexed design.

A schematic circuit diagram of four unit tiles is illustrated in figure 2.1.

Figure 2.1 Schematic circuit diagram of four unit tiles of Viventi array

Each of the above four tiles contains two identically fabricated silicon nanomembrane transistors. The ‘inputs’ in each tile are the platinum electrodes in contact with the cortex. The vertical white connections comprise the column lines of the array. These lines are always ‘on.’ The red and green lines connect row 0 and row 1, respectively. They are ‘on’ only when enabled. ‘col 0’ and ‘col 1’ are the outputs of the array; they carry the electrode voltages sensed by an enabled row. The upper right of each tile illustrates the shared ground connection. At the bottom of the figure, each column is connected to a shared constant current sink which is then connected to a system negative power supply represented by $-V$.

Each unit tile in the array contains two identical N-MOSFET transistors that serve different purposes: one for the purpose of buffering the detected electrical signal and the other for the purpose of enabling multiplexing resulting in the realization of a high-density array with fewer output wires. The transistor whose gate is connected to the ‘input’ is connected to the platinum electrode in contact with the cortex and serves the role of a voltage buffer, also referred to as a common drain amplifier or source follower. This transistor operates in the linear range whereby the output voltage of the transistor is

directly related to input voltage sensed by the electrode. A voltage buffer holds the sensed electrical potential so the value can be ‘read’ on the output line when the multiplexing transistor (the transistor whose gate is connected to a row line) is enabled. Without a buffering transistor, it would not be possible to ‘read’ a given electrode voltage until after the settling time of the electrode passes. The result would be an upper limit on the sampling frequency of multiplexed arrays. However, with the inclusion of buffering transistors at each tile, the electrode voltages are continuously available and the sampling frequency of the array is no longer tied to the settling time of the electrodes.

As mentioned above, the second transistor in each tile, whose gate is connected to a row line, serves the role of a multiplexing transistor. This transistor operates either in the cutoff (‘off’) or saturation range (‘on’). When an ‘on’ signal is present on a given row line, an ‘off’ signal will be present on all of the remaining row lines and the output of the array will be of the voltages present on the column lines of the ‘on’ row only. During the next timing signal, the subsequent row will be turned ‘on’ and the previous row (along with the remain rows) is turned ‘off.’

The sampling frequency of the electrodes on this array is determined by the sampling frequency of the data acquisition system, the total number of electrodes on the array and our choice of over sampling rate for each row (number of clock cycles used to record data from electrodes in a given row before switching rows). The sampling frequency of our data acquisition system, constructed by our laboratory specifically for these experiments, is 100 kHz. During each of our experiments, we set an over sampling rate of either 20, 10

or 6 cycles per row resulting in a row-switching frequency of 5 kHz, 10 kHz and ~ 16.7 kHz respectively. Lower over sampling rates, without additional loss in signal-to-noise ratio of recorded data, was made possible due to increasing quality in fabrication over time. Finally, our array has 18 rows and therefore our effective sampling frequency per row and by extension effective sampling frequency per electrode/channel was ~ 278 Hz, ~ 556 Hz and ~ 926 Hz respectively.

Figure 2.2 shows the schematic of an exploded view (left) and corresponding microscope image of each fabrication layer of a few tiles of the electrode grid-array. Each electrode contact is composed of a $300\mu\text{m}$ square of platinum that is exposed on the surface of easily duplicated individual tiles comprising the 360 contacts of the grid-array. The circuitry of each tile (illustrated in figure 2.1) is constructed with multiple layers.

Figure 2.2 Multiple layers of circuitry in tiles of Viventi array

Schematic (left) and corresponding microscope images of each layer (right). Top images show the layer closest to the brain and bottom images show the layer furthest away from the brain. The tiles were

fabricated from the bottom to the top, starting with the layer of doped silicon (Si) on polyimide and ending with the deposition of platinum (Pt) onto the surface electrodes. Horizontal and vertical metal interconnects on different layers, 1st and 2nd metal-layers (1st and 2nd ML), comprised the ‘wiring’ of hundreds of silicon nanomembrane transistors, two transistors per tile/electrode. Following the fabrication of the metal layers but before the platinum deposition, a multilayer offset vertical interconnect access (VIA) structure was fabricated in order to connect the buffer transistor from each tile to its corresponding electrode while preventing electrical leakage currents when the array is placed on the surface of the cortex which is immersed in highly conductive bio-fluids [16].

The total array size of 10 mm × 9 mm approximates the area recorded from underneath one currently utilized clinical intracranial “grid” electrode. Therefore, the signal recorded from a single clinical electrode, which we consider analogous to the average signal from all 360 channels of the Viventi array, is incapable of recording any spatio-temporal (ST) dynamics of local field potential (LFP) spikes that may be present within this area. Figure 2.3 is a photograph of the 360 channel, high-density neural electrode array used in the experiments.

Figure 2.3 Photograph of high spatial density array used in recordings

The array is a 360 channel, high density neural electrode grid used in a feline model of epilepsy. It was placed on the surface of visual cortex. The electrode size and spacing is 300 μm x 300 μm and 500 μm , respectively.

2.3 Additional Sensor Modalities Utilized in the Study of Epilepsy

Electrical signals are not the only signals available for studying the abnormalities associated with epilepsy. In recent years (1990's), the study of the structural abnormalities (such as hippocampal sclerosis) underlying many seizure disorders made significant advances as improvements were made in the technology behind Magnetic Resonance Imaging (MRI). MRI is safe, non-invasive and accurate. Today, it is widely used clinically for diagnosis and pre-surgical evaluation for patients with refractory epilepsy and in the research setting as developments continue to be made that improve upon the sensitivity and spatial resolution of the technique [29]. However, many cryptogenic cases (~15% of all clinical cases) persist, cases where no structural abnormalities are visible and it is suspected that subtle malformations of cortical development or gliosis remain undetectable by MRI [29, 30]. For these cases, modalities that sense the dynamics of pathological brain signals, temporally correlated with the occurrence of epileptic seizures, or that indicate a functional abnormality must guide diagnosis and therapy [31]. The pathological brain signals that we have come to associate with this disease include: abnormal cerebral blood flow (CBF), abnormal glucose metabolism, abnormal levels of neurochemical signals, abnormal brain electrical activity and its related abnormal magnetic field activity, to name a few [31, 32]. The modalities we employ to sense these signals in the clinical setting include: positron emission tomography (PET), single-photon emission computed tomography (SPECT), functional magnetic resonance imaging (fMRI), magnetic resonance spectroscopy (MRS), electroencephalography (EEG) and magnetoencephalography (MEG). Additional

research techniques employed to study epileptic signals include Calcium imaging and voltage sensitive dye (VSD) imaging [31-33].

PET, SPECT, fMRI, and MRS are all tomographic-based data acquisition modalities whereby a signal of interest within the brain is imaged along various projections and spatially reconstructed in post-processing. PET senses electron-positron annihilation, SPECT senses gamma rays, fMRI and MRS sense radio-frequency waves [33]. EEG, MEG, Calcium imaging and VSD imaging are surface-based data acquisition modalities in that they cannot provide accurate 3-D localization of the source of the signals detected.

PET is a molecular imaging tool that uses radioactively labeled organic molecules to study brain metabolism and neurochemistry, non-invasively. It has the added advantage of whole brain imaging which allows a simultaneous analysis of local and network behavior. However, PET has the disadvantage of a low spatial resolution (2 mm) and a temporal resolution limited by the dynamics of the metabolic or neurochemical process being studied, most often on the order of several minutes [33].

SPECT and fMRI are primarily employed to sense abnormal cerebral blood flow (CBF) which, because of the brain's autoregulation of blood flow with metabolic demand, is used as a surrogate marker for neuronal activity [31]. The spatial resolution of SPECT is on the order of 7-8mm, whereas 1.5-4.5T fMRI images (higher magnetic fields improve signal-to-noise ratio (SNR)) yield a spatial resolution on the order of 1-3mm resulting in their more widespread clinical use. Studies have shown regions of reduced CBF during

interictal epochs and increased CBF during ictal epochs. However, these findings help only to localize areas of abnormality not to study brain dynamics; as the temporal resolution is limited by the dynamics of blood flow and is on the order of 1-3 seconds [34].

MEG measures the magnetic induction outside the head produced by electrical activity in neural cell assemblies. MEG signals are associated with intracellular ionic currents [35]. While EEG measures the electrical local field potentials (LFP) in the extracellular ionic currents produced by the electrical activity in these same neural cell assemblies. [36, 37]. Both MEG and EEG directly measure brain activity and therefore have the highest temporal resolution of all imaging modalities (on the order of tens of milliseconds). The temporal resolution is limited only by the analog-to-digital sampling rate of the data acquisition system used to record the brain signals. Clinically, the spatial resolution of MEG and EEG (~10mm) is inferior to fMRI and PET, limited by sparse placement of sensors. Some investigators use complex dipole modeling strategies to infer fields not directly imaged by these recording devices, however their accuracy can be somewhat limited by the assumptions underlying these field models. However, as mentioned in section 2.1, in research settings, the development of newer devices for intracranial EEG is resulting in spatial resolutions of 500 μ m, surpassing other modalities. An additional disadvantage of MEG is a lower signal-to-noise ratio (SNR) of the signals detected compared to EEG, particularly for deep brain sources [20].

The superior temporal resolution of EEG, and now in the research setting a comparable spatial resolution of modalities such as MRI, are enabling us to observe changes in patterns of brain activation over time and in greater spatial detail, thus allowing investigators to hypothesize about the mechanisms of action leading to seizure genesis. We chose EEG as the sensory modality in the work described below because of its superior temporal resolution, its wide-spread use in clinical settings, acceptance within the research community and its portability [19, 33]. Alternative future research directions in many labs include the integration of multiple imaging techniques. However, because it requires the least amount of bulky machinery, only EEG used alone holds the potential to one day be adopted by patients for their own personal home care and even one day to be the sensing modality employed in implantable monitoring and therapeutic devices for patient care.

2.4 Methods of EEG Data Analysis in Epilepsy

Addressing the history of algorithm development in the post-processing of EEG signals, much of the epilepsy literature is focused on the detection and description of local field potential (LFP) spikes in neural data recorded from patients suffering from the occurrence of seizures [38-42]. LFPs are the result of transient local changes in the extracellular voltage of populations of neurons, which result when they behave synchronously [43]. The shape, timing, frequency and/or location of LFP spikes (occurring both interictally and ictally) have been actively investigated as potential biomarkers useful to predict and/or understand seizures [42, 44]. However, despite

sustained efforts, no single quantitative feature (or combination of features) extracted from clinical recordings has elucidated the relationship between spikes and seizures to date. In 2001, Janszky et al. reviewed the traditional approaches for extracting salient spike features from EEG data [41]. They summarize that since 1958, studies have sought to understand the relationship between interictal spikes and seizures. Early hypotheses postulated that the temporal frequency of LFP spikes in conjunction with their ‘spatial spread’ is what eventually leads to seizure. Yet, more recent studies refute any causality and pose that although LFP spikes and seizures are correlated, LFP spike feature analysis has not yielded a seizure precursor [45]. Gotman’s 1980 paper exploring epileptic spike morphology in human EEG concluded that a statistically significant asymmetry exists in amplitude and width (duration) but not in slope, between the first and second halves of spike slow waves (SSW) detected interictally [42]. However, this asymmetry provided no insight into seizure genesis, progression or termination. An interesting alternative hypothesis concerning the connection between interictal spikes and seizures is that there is no connection; the mechanisms underlying spike generation and seizures are different [46]. Still another hypothesis is that seizures cause interictal spikes and therefore, spikes are not capable of providing insights in any subsequent seizures, only preceding ones [41].

More recent signal analysis techniques have adopted the use of wavelet transforms to attempt to understand a generative perspective of neural electrical components that comprise the appearance of interictal and ictal LFP spikes. Neural network algorithms have also been employed in order to try to understand the dynamics of transient electrical

similarities between disparate and distant channels and the possible patterns of temporary interictal connectivities giving rise to the spread of epileptic discharges leading up to seizure onset.

Clinically, manual scoring by eye, the most crude of all analysis techniques was and continues to be employed. Neurologists are trained to inspect visually temporal EEG traces from tens to hundreds of channels recorded over a few hours to many days from patients in epilepsy monitoring units during pre-surgical analysis. A determination from these traces, which doctors spend many hours analyzing by eye, is made as to where in the brain seizures are thought to originate [19]. It is difficult to ascertain any specific or definitive conclusions about the nature of seizures from this time consuming, subjective and ubiquitously employed analysis technique.

One challenge all of the analysis techniques mentioned above have in common is the inability to analyze the spatial relationships between data on different channels, to leverage the spatial structure of data. Efforts to view clinical EEG in two dimensions have been of variable utility, as limited spatial resolution of early recording arrays leaves much of this imaging to interpolation, with variable relation to the actual temporal and spatial dynamics of seizures. A key concern is the capability 1 cm spaced electrodes have to resolve fine cortical activity in such a way as to give meaningful mechanistic information on seizure generation, as the distance between adjacent electrodes is so great, with millions of neurons between them, that interpolation between these likely unrelated recording sites becomes of questionable utility. The algorithm employed in this

dissertation attempts to exploit the spatial relationships between channels, primed by much higher resolution data.

There are two important consequences of advances in sensors that increase the spatial and temporal resolution of neural recordings: (1) they dramatically increase the amount of data to be analyzed and (2) there is increased pressure to explore automated methods of data analysis. Large amounts of data spanning Terabytes per patient preclude standard clinical analysis “by eye.” Automated methods of analysis are required both to increase the efficiency/ effectiveness of neurologists time/ expertise and to standardize observational/ evaluation criteria. This in turn drives data interpretation, reducing the risk of inconsistent or poorly reproducible interpretation from one expert to another, which is common in human scoring [24]. In addition, automated methods for analysis enable interpretation of complex data in real time, which will eventually facilitate implementation in implantable devices. Addressing the first challenge of an increased amount of data has helped spawn tremendous activity in the field of computational neuroscience. It has also encouraged the application of digital signal processing techniques traditionally reserved for communications to biomedical signals and the adaptation of machine learning algorithms to try and discover hidden patterns where traditional techniques have failed. Such methods include unsupervised learning for data classification [24] and mining intrinsic structure using Bayesian estimates [47-49]. The hope is that these techniques will lead to an increased understanding of the mechanisms of action underlying seizure disorders.

2.5 Conclusion

Due to our continued lack of understanding of the epileptic network, which leads to the unpredictable occurrence of debilitating seizures, we look for new ways to gather more data from the brain to overcome the current sparseness of information. We hypothesize that this sparse sampling, in part, limits our ability to improve diagnoses and therapy for this disorder. Recent improvements in technology have led to the development of multiple electrode arrays capable of recording from more channels at a higher spatial density while covering a larger cortical surface area without degradation in temporal resolution or signal-to-noise ratio. The development of new high-density electrode arrays makes it increasingly important to develop algorithms for data analysis that take advantage of the increase in data availability. In the following chapters, we describe our contributions in developing new algorithms to leverage the increased amount of available data recorded from these newly developed arrays.

3 Description of the Algorithm

3.1 Summary

Local field potential (LFP) epileptiform spikes have long been scrutinized for their potential to lead to an increased understanding of time to next seizure, mechanisms sustaining seizures, seizure progression, and termination. In this thesis, because recordings are made at a much higher spatial density, a new opportunity exists when extracting salient features from LFP spikes. It is possible that these new features will illuminate a previously unobservable relationship between LFP spikes and seizures.

In this chapter, we introduce an algorithm for extracting spatio-temporal features of LFP spikes in order to describe patterns of spike propagation and the spatial distributions of spike power as these events cross beneath a high spatial density array. LFP spikes were

detected from data recorded during experiments utilizing an acute *in vivo* model of epilepsy. In brief, our algorithm utilizes principal components analysis (PCA) to reduce the dimensionality that results when extracting large numbers of features per LFP epileptiform spike. Our algorithm then applies k-medians clustering, an unsupervised learning technique from machine learning, to separate differing classes of LFP spikes. The results of applying these techniques show distinct clusters of spatio-temporal patterns of epileptiform LFP spikes present within the micro-electrocorticographic (μ ECoG) data. Our new approach exceeds the basic examination of shape and frequency analysis of individual spikes on spatially isolated, independent channels and leverages the increased amount of data and their spatial relationships that are available from high-density arrays.

The primary purpose of this chapter is to present our method for automatically detecting LFP epileptiform spikes and classifying them into distinct clusters based on the spatio-temporal features we extract. We present the animal model, the data acquisition system, and the software algorithms developed to analyze the recorded data, along with our results. Our results demonstrate that LFP epileptiform spike data is separable into distinct clusters based on the extraction of spatio-temporal features that describe the process of seizure generation in the experimental preparation. Additionally, our results demonstrate that our data analysis algorithm is a useful tool for future datasets, recorded from high-spatial density arrays of any scale.

3.2 Introduction

The mainstay of research into mechanisms of seizure generation uses animal models [1]. These models employ chemical convulsants, electrical stimulation, genetic manipulation, aberrations of development (febrile seizures, hypoxic-ischemia induced in rat pups, etc.), and trauma to induce seizures and/or epilepsy. Each model targets a specific type of human epilepsy or addresses underlying disease mechanisms, such as seizure generation, spread, progression, termination, or comorbidities (e.g cognitive dysfunction) [1]. Models separate into acute, where single or multiple seizures are studied during a time-limited session, or chronic, where manipulations generate repeated seizures that arise spontaneously over time, often following a latent period, and are studied over multiple sessions spanning days, weeks or months. Many of these models, both acute and chronic, generate repetitive seizures over hours, enabling investigators to carefully study the process by which seizures arise and to attempt various interventions to disrupt them. Topical application of chemoconvulsants, such as picrotoxin on cat neocortex, simulate acute simple partial seizures in intact brains of mammals [7], enabling investigators to study the propagation patterns of local field potential (LFP) spikes leading up to ictal events. All recorded data in this thesis were obtained from an acute *in vivo* feline model of epilepsy. Our surgical procedures and specific method of inducing seizures are described in section 3.3.1.

Our data acquisition system is described in section 3.3.2.

Upon our initial exposure to data recorded from preliminary experiments, we opted to implement an initial version of our algorithm on a subset of the data to gain an understanding of the difference between traditionally extracted features from recorded LFP spikes detected on clinical sized electrodes and our proposed novel features to be extracted from the spikes recorded from grid electrodes of much higher spatial density. The subset of the data we selected was the final 13 minutes and 40 seconds from Cat 1 (final 25% of the Cat 1 data).

To approximate the data recorded from a clinical electrode, we averaged the signal recorded from all 360 channels of the grid array. In doing this we propose that a clinical electrode of 5 mm diameter is equivalent to the average of all 360 channels of the 1 cm² grid array. To detect and extract LFP spikes, we applied a voltage threshold of 500 μ V on this average signal. This threshold is lower relative to the 1 mV threshold we later implement in our fully developed algorithm; it was selected because spikes are not as prominent on the average signal when compared to spikes detected on individual channels. Detected spikes have lower amplitudes on the average signal because most LFP spikes in our data were detected only under a small area of the array. Therefore, their amplitude is diminished when averaged with channels outside of the spike detection area. This points to one limitation of large clinical arrays; LFP spikes detected by areas smaller than 1 cm² run the risk of not being detected at all due to signal averaging inherent when recording iEEG using larger electrodes. It is possible we could detect the start of focal seizures earlier in the clinical setting if we used high-density arrays with small electrodes

and with large areas of coverage in our pre-surgical evaluations. In our implementation of the initial version of our algorithm on a subset of data recorded from a preliminary experiment (Cat 1), we detected 724 spikes.

For each of the 724 spikes, we extracted a 160 ms window (60 ms prior to the threshold crossing and 100 ms post-crossing). 160 ms was selected as the window size because it was wide enough to capture the whole spike while narrow enough to successfully temporally separate detected spikes (equates to a spike frequency of approximately 6 Hz). In the later implementation of the fully developed algorithm, this window size was reduced to approximately 50 ms, equating to a spike frequency of 20 Hz, due to the higher frequency of LFP spikes encountered when including the entire dataset. Additionally, for this initial implementation of our data analysis algorithm we included the constraint that all detected spikes cover a majority of the array (requiring individual spikes of an LFP to be detected on at least 50% of the 360 electrodes; most detected spikes satisfied this constraint since spikes covering less than 50% of the array typically resulted in an average amplitude too low to cross the voltage threshold for detection). Figure 3.1 depicts 16 representative detected spike waveforms (average of all channels in the array).

Delay maps and standard deviation maps were then generated using the feature extraction techniques described in detail in section 3.3.6. The detected spikes were subsequently clustered, via k-medians clustering described in section 3.3.7. Tibshirani's gap statistic, described in section 4.3.1, was applied to determine the number of clusters present within

this subset of the data, yielding 16 distinct clusters based on their ST features. Figure 3.2 displays the spike ST delay patterns corresponding to the 16 clusters that resulted.

Figure 3.1 Representative average spike waveforms following spike detection

Each of the spike waveforms is 160ms clipped from the average voltage recording of all 360 channels. Within each window there is only one negative-going crossing followed by one positive-going crossing. Negative is plotted up by convention. These 16 waveforms are of the spikes closest (in the L1 sense) to each of the cluster centers and correspond to the spatio-temporal delay maps in Figure 3.2.

Figure 3.2 Delay maps for detected spikes closest to the centers of 16 clusters

Color shading represents relative timing of peak voltage in each spike, but not speed of propagation. Blue represents the earliest onset time of each waveform, progressing to red as the latest portion of the

waveform. As an example, the delay map for cluster 1 displays a spike ST pattern of propagation across the array of a spike that enters on the bottom left and proceeds in a sweeping arc until it exits the array in the top left. The corresponding standard deviation maps have not been included.

The 16 average spike waveforms depicted in figure 3.1 correspond to the 16 delay maps depicted in figure 3.2 which illustrate the delay features from the spikes closest to the cluster centers of the 16 clusters that resulted, following k-medians clustering.

These preliminary results were the motivation to pursue and further refine clustering based on ST features for the entire dataset. An important observation gleaned from these results was that spikes which were clearly separable based on ST features were very difficult to distinguish based on traditional/historical analysis techniques of extracting temporal, frequency and shape characteristics of spike waveforms. For example, the shapes of the spike waveforms from clusters 1 and 11 (figure 3.1) appear to be very similar, however, their delay maps (figure 3.2) clearly depict the dissimilarity of the wave propagation path taken across the array. A second example holds for the spike waveforms from clusters 9 and 13. However, we acknowledge that some waveforms which are similar do have correspondingly similar delay maps (for example clusters 4 and 8); ultimately, what motivated us to continue to develop this novel analysis approach were the cases where spikes would not have been differentiated based on their shape alone. The following section outlines our experimental protocol, data acquisition system, and software algorithms developed for data analysis.

3.3 Methods

3.3.1 Animal Model

Surgical procedures: Three adult cats were anesthetized with an intraperitoneal injection of nembutal (25-30 mg/kg) and supplemental isoflurane (2-5% in 70% N₂O and 30% O₂), for 15 minutes while placing an intravenous cannula. Subsequently, the animal was paralyzed with gallamine triethiodide (Flaxedil) and artificially ventilated (end tidal CO₂ held at 3.8-4.0%). A stereotaxic frame was used to stabilize the head of the animal and a craniotomy and durotomy were performed to expose at least a 2 x 3 cm region of cortex. During the surgery, the cat received additional Nembutal as needed in approximately 5 mg boluses to maintain a stable depth of anesthesia. Following the surgery and for the remainder of the experiment (8-12 h), anesthesia was maintained by continuous infusion of intravenous thiopental or nembutal (3-15 mg/kg-hr). Heart rate, blood pressure and EEG were continuously monitored. Rectal temperature was maintained at 37-38°C. All experiments were conducted in accordance with the ethical guidelines of the National Institutes of Health and with the approval of the Institutional Animal Care and Use Committee of the University of Pennsylvania.

Induction of seizures: Seizures were induced using picrotoxin, a GABA-A receptor antagonist that blocks inhibition. Picrotoxin was applied topically to induce focal neocortical seizures, as described by Raol and Fisher, to study mechanisms of seizure generation and spread [1, 7]. Cortical activity was monitored continuously after administration and the picrotoxin dosage was increased gradually until epileptiform

activity was observed. Picrotoxin was applied adjacent to the electrode array in an amount sufficient to induce abnormal electrical spikes and seizures from the covered region [1, 7]. During one of the three experiments (Cat 3), data were collected from two different locations on the neocortex, yielding two separate datasets (Cat 3a and Cat 3b). Table 1 describes datasets collected from the above experiments. Each experiment induced repeated seizures, and spanned 8-12 hours.

3.3.2 Data Acquisition

Data were recorded with active, flexible, high-resolution electrode arrays covering approximately 1cm^2 of neocortex. Each array was placed on the exposed cortex and held in place using a micro-manipulator. The arrays collected micro-electrocorticographic (μECoG) data continuously from 360 independent channels arranged in 20 columns and 18 rows, spaced $500\text{ }\mu\text{m}$ apart. Each electrode contact was composed of a $300\text{ }\mu\text{m} \times 300\text{ }\mu\text{m}$ square of platinum. Two high-performance, flexible silicon transistors and a single recording electrode were fabricated in repeating units on the array, to transduce, buffer and multiplex the recorded signals on each channel [16]. The total array size was $10\text{ mm} \times 9\text{ mm}$, which approximates the area recorded from underneath one currently utilized clinical ECoG electrode [13]. Signals were recorded with an effective sampling rate between 277.7 to 925.9 Hz per channel (see Table 3.1).

Custom hardware was designed and developed to drive the active electrodes and interface with a National Instruments PXI system (National Instruments Corporation, Austin, TX).

Table 3.1 Collected data summary

	Post-Trim File Length (h:m:s)	Number of Seizures	Number of Spikes	Seizure Length (seconds)	Sampling Frequency (Hz)	Array Channel Yield
Cat 1	0:53:37	7	2894	5-42	277.78	295/360 (81.9%)
Cat 2	3:15:30	17	255	16-53	555.56	318/360 (88.3%)
Cat 3a	4:24:15	76	14,528	6-65	925.93	294/360 (81.6%)
Cat 3b	0:37:49	23	8,654	7-70	925.93	294/360 (81.6%)
Total	9:11:11	123	26,331	n/a	n/a	n/a

Post-Trim File Length is the length of the file after segments of noise were clipped. Seizures were marked by a neurologist/ epileptologist boarded in both neurology and EEG/ Electrophysiology (ABPN Added qualification and ABCN) and were defined to be at least 5 seconds in length. Spikes were detected via a threshold detector described in section 3.3.5. Sampling Frequency was per channel. Array Channel Yield gives the number of working electrodes on an array. Data for the non-functioning channels were interpolated as described in section 3.3.4. The array was placed in two different locations during the experiment on Cat 3, yielding two datasets.

Software was developed using National Instruments' LabVIEW to de-multiplex, visualize and store recorded signals. A schematic of the experimental set-up is depicted in figure 3.3.

Figure 3.3 Schematic of the experimental set-up

The electrode interface board, headstage boards and digital boards were all designed and developed in the lab in order to drive the active electrodes and interface with the National Instruments (NI) digital acquisition (DAQ) PXI system. Additional software was developed using NI's LabVIEW to de-multiplex, visualize and store recorded neural signals.

3.3.3 Overview of Data Analysis

Figure 3.4 illustrates the four major stages of the data analysis algorithm. We implemented a novel machine-learning approach identifying and tracking 2-dimensional spatio-temporal (ST) patterns associated with epileptiform local field potential (LFP) spikes recorded by the high-density neural electrode array. Following a few initial pre-processing steps, this algorithm detects LFP spikes and extracts delay and power features from their temporal signals within a window surrounding each spike. Finally, using machine learning techniques, during the clustering stage, the data is reduced (retaining 99% of its variability) and clustered via k-medians. The resulting clusters separate the spikes by their pattern of propagation and their distribution of power across the array. Data analysis was performed using custom developed MATLAB (MathWorks, Natick, MA) scripts.

Figure 3.4 Block diagram of an overview of the data analysis algorithm

3.3.4 Pre-processing

During pre-processing, all recordings were band-pass filtered between 1 and 50 Hz using a 6th-order butterworth filter in the forward and reverse directions to achieve zero-phase filtering (and effectively doubling the order of the filter to a 12th order filter), by

employing MATLAB's `filtfilt` function. Next, any missing data, as a result of non-functional channels (accounting for $< 15\%$ of the array in each experiment, see Table 1 and figure 3.5), were interpolated from surrounding electrodes using a 2-D averaging spatial filter of window size 3×3 pixels.

Figure 3.5 Electrode yields for the arrays used during the experiments for Cats 1-3. The same array was used during the Cat 3 experiment to collect the Cat 3a and Cat 3b datasets. Clear cells in the array demonstrate non-functioning electrode contacts that were interpolated for the analyses that follow below.

Finally, noisy data were trimmed by visual inspection. When noise was observed across all channels, those segments were clipped and removed prior to continuing with the analysis. An example of the amount of data removed as a result of noise is illustrated in figure 3.6, for the experiment from Cat 1, where $\sim 5.43\%$ of the data were removed. 1.99%, 1.03% and 0.44% of the data were removed for Cat 2, 3a and 3b respectively.

Figure 3.6 Average micro-ECoG trace for Cat 1. Data from the Cat 1 experiment was collected in 2 files. The segments in green are observer-identified epochs of noise across all channels of the array. Those segments of data were removed and were not

included in the subsequent stages of data analysis. A similar process was implemented for the data from Cats 2, 3a and 3b. Figures not shown.

There were times when excessive amounts of noise consistently affected only 1 or 2 channels out of all 360 channels. When this was the case, that entire channel was deemed to be non-functioning and was included amongst the non-working channels mentioned previously in this section. Subsequently, the missing data from those channels were interpolated.

The last type of noise encountered was noise that would only sporadically appear on a few channels at a time for a limited amount of time. The origins of this noise are unclear and are thought to be a result of imperfect manufacturing of the array at the time of the experiments. This noise was not removed but rather data containing this type of noise proceeded on to the subsequent stages of data analysis. Since data analysis following this initial stage of pre-processing is automatic, this noise would often trigger the threshold detector used for LFP spike detection and windows surrounding these noise snippets would then be forwarded on to clustering along with ‘true’ LFP spikes. We then relied on the processing of clustering the data to group all of the noise snippets together in one easily identifiable ‘junk’ cluster. We revisit the topic of these noisy snippets in the Results section (section 3.4).

3.3.5 LFP Spike Detection

We used a voltage threshold to detect spikes from recorded signals. The voltage threshold was set by visual inspection at -1.0 mV for all datasets (with the exception of the dataset

from Cat 2, for which we set the voltage threshold at -0.5 mV). In figure 3.7, the threshold appears as a positive number since all recorded voltage amplitudes were negated by convention. This threshold was set for spikes detected on individual channels, not on the average of all 360 channels. After detecting a spike, a 50 msec window was copied and stored from all channels (2 msec prior to the crossing and 48 msec post-crossing). We later extracted features from these stored windows of data. A 50 msec window size was chosen to minimize the detection of multiple epileptiform spikes within the detection window while capturing as much of a given spike temporal waveform as possible. Additionally, spikes detected on fewer than 10% of the electrodes of the array were not included since upon visual inspection, the majority of these spikes were found to be the result of transient noise on a few channels. Table 1 shows the number of spikes detected in each dataset. Figure 3.7 depicts a typical detected spike. The multicolored traces are from each of the 360 electrodes on the array. The heavy black trace is of the average of all 360 traces. And the horizontal blue line shows the voltage threshold crossed, resulting in this spike detection. In this example, approximately 283 channels (78.6% of the electrodes on the array) detected the LFP spike with a voltage amplitude high enough to cross the threshold. The figure window is approximately 150 msec; the center third contains the 50 msec window retained following the spike detection. The trailing spike, on the right, was handled as a separate detection.

Figure 3.7 Spike detection resulting from a voltage threshold crossing

The multicolored traces are from each of the 360 electrodes on the array. The heavy black trace is from an average of all 360 traces. The horizontal blue line shows the voltage threshold crossed (1 mV). The y-axis depicts Voltage in mV. The x-axis depicts samples; the time between samples is equal to 3.6 msec (277.78 Hz sampling frequency).

3.3.6 Feature Extraction

To characterize detected spikes by a set of features we created “delay maps” to capture the direction of spike wave propagation and “standard deviation maps” to represent the spatial distribution of power of the spike recorded by the array’s 360 channels. Delay maps were generated for each extracted 50 msec window by determining the relative time of the peaks of spikes on individual channels. This feature captures both the direction of movement of spikes across the array and the speed of propagation. To generate standard deviation maps, we calculated the zero-meaned RMS power for each channel within the 50 msec analysis window. This resulted in a single value per channel representing the

power of each spike within the time window. The zero-meaned RMS power can be written as:

$$x_{rms} = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}} \quad (3.1)$$

where $x_1 - x_n$ are the voltage amplitudes of a given spike within the detection window for each time point (1- n) and n is the total number of time points. Prior to calculating the zero-meaned RMS power we subtract the mean of the 360 voltage traces within each detection window. Therefore, the zero-meaned RMS power is equal to the standard deviation of the signal as can be seen below:

$$\sigma = \sqrt{\frac{1}{N}[(x_1 - \mu)^2 + (x_2 - \mu)^2 + \dots + (x_N - \mu)^2]}, \text{ where } \mu = \frac{1}{N}(x_1 + \dots + x_N) \quad (3.2)$$

where μ is the mean of the signal (now equal to zero) and N is the total number of time points. Figure 3.8 depicts the delay and standard deviation map corresponding to the spike detected in figure 3.7.

Figure 3.8 Delay and power maps resulting from the spike detected in figure 3.7

A. Delay map. The colors represent the relative delay of the spike peak for each detected spike from the 283 electrodes which had a voltage amplitude large enough to cross the voltage threshold. Blue indicates an electrode with an early detection and red an electrode with a late detection. The deep blue zone at the bottom of the figure represents the 77 electrodes whose voltage traces were not large enough to cross the threshold. **B.** Power map. The colors represent the relative amount of power recorded from each electrode within the 50 msec spike detection window. Blue is relatively low power; red is relatively high power.

Each set of maps (delay and standard deviation) was then normalized across all spikes in the dataset. Following normalization, we concatenated the two sets, yielding 720 features per spike (360 delay, 360 power). To reduce the dimensionality of this comprehensive feature set we implemented Principle Components Analysis (PCA). PCA determines an orthogonal transformation to convert a dataset with a large number of observations, each defined by a set of possibly correlated feature variables, to an alternative representation of the dataset, with the same number of observations, now defined by a set of linearly uncorrelated variables known as principal components (the resulting features are linear combinations of the original features). Principal components are calculated by determining the combination of feature-dimension directions along which the lengths of the orthogonal projections of the data observations have the maximum variance. Each successive principal component is constrained to be orthogonal to the previous component. Principal components are the eigenvectors of the covariance matrix of the data, C :

$$C = \frac{1}{N} \sum_{n=1}^N (\mathbf{b}_n - \bar{\mathbf{b}}) (\mathbf{b}_n - \bar{\mathbf{b}})^T \quad (3.3)$$

where $\mathbf{b}_n$ is the 720 x 1 concatenated delay and power feature vector of each detected spike and $\bar{\mathbf{b}}$ is the mean feature vector of all N detected spikes. The new coordinates for each detected spike are computed as:

$$\mathbf{X} = \mathbf{B}\mathbf{U} \quad (3.4)$$

where $\mathbf{X}$ is the new $N \times P$ data matrix of spike-detection representations, $\mathbf{B}$ is the $N \times D$ original data matrix of spike detections, each detection defined by 720 concatenated

delay and power features, and U is the $D \times P$ matrix whose columns are the unit-normalized eigenvectors of C corresponding to the P largest eigenvalues ($P \leq D$). We retained only the number of eigenvectors necessary to account for 99% of the variance within the data [17]. Table 3.2 displays the number of PCA dimensions retained.

Table 3.2 PCA dimensions retained capturing 99% of variance in the original data

	Number of Spikes Detected (N)	Original Number of Feature Dimensions (D)	PCA Dimensions Retained (P)
Cat 1	2894	720	251
Cat 2	255	720	155
Cat 3a	14528	720	281
Cat 3b	8654	720	316

Number of Spikes Detected corresponds to N in the description of equation 3.4 above. Original Number of Feature Dimensions corresponds to D . And PCA Dimensions Retained corresponds to P .

3.3.7 Clustering

Clustering was performed using the k-medians algorithm, an adaptation of the k-means algorithm [50]. The goal of the k-means algorithm is to minimize the quantity:

$$W_K = \sum_{k=1}^K \frac{1}{2N_k} \sum_{i,i' \in C_k} d_{ii'} \quad (3.5)$$

where K is the total number of clusters, C_k is the set of indices of the data points in cluster k , and N_k is the total number of data points within cluster k . $d_{ii'}$ in the k-means algorithm is the squared Euclidean distance between P -dimensional observations x_i and $x_{i'}$:

$$d_{ii'} = \sum_{p=1}^P (x_{ip} - x_{i'p})^2 = \|(\mathbf{x}_i - \mathbf{x}_{i'})\|^2 \quad (3.6)$$

Minimizing W_K in (3.5) is the equivalent to minimizing the pooled within-cluster sum of

squared distances from the cluster means and can alternatively be written as:

$$W_K = \sum_{k=1}^K \sum_{i \in C_k} \|\mathbf{x}_i - \bar{\mathbf{x}}_k\|^2 \quad (3.7)$$

where $\bar{\mathbf{x}}_k$ is the mean of the cluster elements in cluster k .

The k-medians algorithm is similar to the widely used k-means algorithm but employs the L1 (city block) rather than L2 (Euclidean) norm as its distance metric, and computes medians rather than means. K-medians is more robust to the presence of outliers in the data [50].

To describe the k-medians algorithm, we will use an example where we know the number of clusters, k , present in the data is equal to 10. K-medians initializes a search for cluster centers by randomly selecting 10 points within the dataset as cluster centers. Next, an iterative descent algorithm for minimizing W_K is used in order to converge towards the true cluster centers (convergence at least to a local minimum is assured). The iterative descent algorithm uses a two-phase scheme. The first phase uses batch updates to assign points to their nearest cluster center, followed by recalculating all cluster centers. The two steps of this first phase are repeated until cluster centers no longer move from one iteration to the next. The second phase uses online updates, where points are individually reassigned in an attempt to further reduce the sum of distances to cluster centers, which are recomputed after each reassignment. The iterations stop when the algorithm converges on a local minimum or when the algorithm reaches a maximum number of iterations. We set our maximum number of iterations to 750. And, in order to increase our

chances of finding the global minimum, we repeated the number of random initializations of cluster centers 30 times. Additionally, we chose to seed the k-medians clustering algorithm with 10 clusters in order to aid in comparison across datasets.

Figure 3.9 shows the cluster assignments for detected spikes in the Cat 1 dataset; only the first 3 principal component projections (of the 251 PCA-dimensions retained and employed in the k-medians algorithm) are depicted.

Figure 3.9 Clustering results from the 2894 spikes of Cat 1

The 10 colors in each plot represent the cluster assignment for each spike. **A.** Projection of the 2894 spikes onto PCA dimensions 1 and 2. **B.** Projection of the 2894 spikes onto PCA dimensions 1 and 3. **C.** Projection of the 2894 spikes onto PCA dimensions 2 and 3. A total of 251 PCA dimensions were used in the k-medians clustering resulting in these assignments.

3.4 Results

We were able to extract novel features characterizing the spatio-temporal wave propagation dynamics of epileptiform spikes by leveraging a high-spatial density 360-electrode grid array. We were also able to cluster each dataset, from 3 separate cat experiments, into 10 distinct clusters of LFP spike detections characterized by these novel features. Figure 3.10 depicts our results from the Cat 1 dataset. Figure 3.11 depicts our results from the Cat 2 dataset. Figure 3.12 depicts our results from the Cat 3a dataset. Finally, figure 3.13 depicts our results from the Cat 3b dataset.

Cat 1 – 10 Clusters of 2,894 Spikes

Figure 3.10 Combined delay/power maps of spikes closest to cluster centers, Cat 1. Similar to figure 3.8.B, the colors represent the relative amount of power recorded from each electrode within the 50 msec spike detection window. Blue is relatively low power; red is relatively high power. The black quivers on each plot depict the direction of spike wave propagation (the quivers are not scaled by velocity).

Cat 2 – 10 Clusters of 255 Spikes

Figure 3.11 Combined delay/power maps of spikes closest to cluster centers, Cat 2. See figure description for figure 3.10.

Cat 3a – 10 Clusters of 14,528 Spikes

Figure 3.12 Combined delay/power maps of spikes closest to cluster centers, Cat 3a. See figure description for figure 3.10.

Cat 3b – 10 Clusters of 8,654 Spikes

Figure 3.13 Combined delay/power maps of spikes closest to cluster centers, Cat 3b
See figure description for figure 3.10.

Figure 3.14 depicts the delay maps for the 25 detected spikes closest (using L1 distance) to their respective cluster centers for four different clusters from the Cat 1 dataset. This

Figure 3.14 Cluster homogeneity of delay maps from Cat 1
Delay maps from 25 detected spikes closest to their respective cluster centers (L1 distance) are depicted for four different clusters. Similar to figure 3.8.A, the colors represent the relative delay of the spike peak for

each detected spike with voltage amplitudes large enough to cross the voltage threshold. Blue indicates an electrode with an early detection and red an electrode with a late detection. The deep blue zone represents the electrodes whose voltage traces were not large enough to cross the threshold.

figure is included to show the homogeneity of features from detected spikes within distinct clusters.

3.5 Discussion

When implementing the clustering step of our data analysis algorithm, we initially chose to seed k-medians with the possible number of clusters within a dataset equal to anywhere from 1 to 30 clusters and to implement a gap statistic to arrive at an intrinsic number of clusters within each dataset. However, as can be observed in figure 3.9, we found that clusters were neither very distinct nor compact (with the caveat that we are only viewing the first three PCA dimensions in this figure). This is due to the nature of the noisy dataset. Our observation of tightly packed cluster centers with loosely dispersed cluster borders, ‘cluster dust,’ prompted us to realize we had a scenario for which the implementation of a gap statistic was ill suited. It was partially for this reason that we chose to fix the number of clusters at 10. (The other reason, as mentioned earlier in section 3.3.7 was in order to aid in comparison across datasets). In figure 3.14 we observed that spikes nearer the center of each cluster are comprised of adequately homogenous spike ST-patterns. In the following chapter, we later conclude that the cluster homogeneity was sufficient to statistically reject our null hypothesis despite the noisiness of the data and the fact that we fixed the number of clusters, a priori.

From figure 3.10, we observed the likelihood that spikes comprising clusters 8, 9 and 10 should belong to a single cluster rather than belonging to 3 distinct clusters. We conclude that it is possible, in heuristically seeding the k-medians algorithm with the information that 10 clusters exist within the data, our algorithm resulted in over-splitting a cluster. Clusters 8, 9 and 10 could be more accurately combined resulting in a total of 8 versus 10 clusters present within the Cat 1 dataset. However, again, we shall see in Chapter 4, that despite this possible over-splitting, we were still able to statistically reject the null hypothesis posed in the beginning of that chapter. Therefore, we have not lost any statistical power due to possible over-splitting. In summary, given the noisiness of the data, we were satisfied with our decision to heuristically select a priori the number of clusters present in the data.

In figure 3.11, we again made the observation of a possible over-splitting of clusters 3, 4, 6 and 7. A second and more general observation is that of lower zero-meaned RMS power present in each of the detected spikes; and that virtually none of the detections cover the entire array. We speculate that this difference to the other three datasets is due to the reduced amount of picrotoxin application during this experiment. Additionally, this experiment was unique due to injecting picrotoxin subcortically as opposed to topical application in the other experiments. Therefore, it is possible that inhibition was blocked in a deeper cortical layer resulting in epileptiform spikes that were more difficult to detect from the surface of the cortex. One final observation from this dataset is that cluster 9 appears to have only captured spikes generated by spurious noise.

In figure 3.12, we made the observation that possibly clusters 5 and 6 have been over-split. We also made the observation that most of the wave propagation patterns appear to lie along a mid-horizontal band across the array with the majority of the power in the spike sometimes evenly distributed across this band (cluster 10), sometimes more prevalent on the left side of the array (clusters 7 and 9) and sometimes more prevalent on the right side of the array (clusters 1, 2, 4, 5, 6 and 8).

In figure 3.13, we made the observation that the majority of cluster center patterns are situated closer to the top of the array when compared to patterns for Cat 3a. This is because the array was repositioned with a vertical downward shift for this dataset, Cat 3b, during a continuation of the Cat 3 experiment.

The results observed in this chapter add to the body of evidence that reports the presence of LFP spike propagation waves in cortex. Prechtl et al. described finding waves of electrical activity in turtle visual cortex induced by visual stimuli [51]. They analyzed data in the frequency domain and isolated specific bands, to more fully understand various phase relationships of electrical activity within a local neighborhood. Using Voltage Sensitive Dyes (VSDs), Prechtl et al. found different types of waves of activity including plane waves, spiral waves and more complex patterns. More recently, Huang et al. demonstrated the use of VSD imaging in the analysis of unique patterns of brain activity in induced sleep states [52]. They postulated that the patterns point toward underlying cortical mechanisms. Spiral waves detected in this study were thought to

explain mechanisms occurring on the mesoscopic scale. They proposed that patterns which modify cortical activity via disruptions in frequency, spatial coherence and modulation of LFP amplitude result in local organization, impacting small networks. Spiral waves occurred with other wave patterns, as we have found in our research. A highly excitable medium is required in order to propagate spiral waves and in the case of Huang et al., this was achieved using a bicuculline model of epilepsy. Bicuculline is a GABA-A receptor antagonist that blocks inhibition similar to picrotoxin. There is still much more to be understood concerning the dynamics of neuronal network processing, both in response to external input and when associated with pathological states of the brain, such as seizures. High-resolution recordings of this nature have yet to be performed in humans, and we are hoping to push this forward in our “next steps” research plan. At this time it is not known how much of the spike patterns we have observed apply to spontaneous human seizures, though human ictal events look quite similar, at the macro level, to the acute seizures recorded in our work.

4 Statistical Testing

4.1 Summary

In order to address the uncertainty of the true number of clusters present in the data, we initially employed a gap statistic developed by Tibshirani [53]. The conclusion of our analysis using this method was that the application of the Tibshirani gap statistic was limited in its ability to reveal the intrinsic number of clusters present within the data due to the noisiness of the data. As a result, we decided to fix the number of clusters for each dataset heuristically at ten and then to perform a statistical validation of our cluster assignments via the rejection of our null hypothesis. Ten is unlikely to be the true number of clusters present within any one dataset. We most likely over-split clusters in some of the datasets and/ or combined two clusters that should be split in other datasets. However, following the application of a permutation test on the results of the cluster assignments,

we concluded that we retained enough statistical power to reject our null hypothesis that ‘the proportion of spikes occurring during seizures is equal across clusters.’ As a result of this conclusion, we choose to overlook the noisiness of these particular datasets and our subsequent inability to discover the intrinsic number of ST-pattern clusters present within them, and encouraged by our findings to date, we will proceed with confidence in collecting newer lower noise data recordings from future experiments.

4.2 Introduction

Tibshirani’s Gap Statistic is a method used to estimate the number of clusters present within a dataset when the number of clusters is not known beforehand. The method is applied by running k-medians multiple times, seeding it with varying values of k (number of clusters) and comparing the observed within cluster dispersion, W_k (equation 3.5), at a given k to the expected dispersion under an appropriate reference null distribution. [53] To generate an appropriate reference null distribution a number of randomly distributed datasets are generated (we later use the average of these multiple distributions for subsequent calculations) and clustered, with the same varying values of k used when clustering the observed data. By comparing the observed within cluster dispersion values (O_k in figure 4.1) to the average reference null dispersion within cluster dispersion values (E_k in figure 4.1) for varying number of clusters, it is possible to identify the point at which a difference is statistically maximal. The k at which this maximal difference occurs indicates the proposed optimal number of clusters that identify various distinct wave propagation patterns of LFP spikes, intrinsic within the structure of each dataset.

Figure 4.1 illustrates, with a toy dataset, how Tibshirani's Gap Statistic can be employed to correctly determine that the intrinsic number of clusters present in this example is two.

Figure 4.1 Tibshirani's Gap Statistic applied to toy dataset

A. Toy dataset containing two clusters in 2-dimensional feature space. **B.** Within sum of squares function, W_k , for the observed dataset. **C.** Log(W_k) functions for the observed (O_k) and expected (E_k) datasets. (E_k is calculated as the average of multiple reference null distributions.) **D.** Gap curve [53].

Figure 4.1.A shows the scatter plot of the toy data in 2-dimensional feature space. Figure 4.1.B displays the within sum of squares function (W_k vs. k), for the observed dataset. W_k is a monotonically decreasing function of k . This is apparent when considering the extreme case where k equals the number of data points within the dataset, in which case

W_k is equal to zero. This function often shows a characteristic ‘elbow’ where previous relatively larger decreases in the function suddenly flatten and become more gradual. The ‘elbow’ typically occurs at the value of k when W_k is no longer decreasing due to the discovery of new intrinsic clusters, but is decreasing due to the gradual reduction that results when over-splitting existing clusters. In figure 4.1.B, the ‘elbow’ in the within sum of squares function is apparent when $k=2$; however the observation of an ‘elbow’ in this curve is not statistically robust enough to conclude that two clusters are intrinsic in the data. Tibshirani’s Gap Statistic is a measure to statistically arrive at this ‘elbow.’ Figure 4.1.C, illustrating the $\log(\text{within sum of squares})$ functions for the observed and expected datasets, confirms the presence of two clusters statistically, since it can be observed that when $k=2$, the gap between these two curves is largest. Mathematically, it is the first time that the difference between the two curves exceeds the difference of the two curves at the next value of k minus the standard deviation at the next k (the standard deviation is calculated from reference null distribution). It is easier to observe that this first occurs when $k=2$ in figure 4.1.D which illustrates the gap curve resulting when taking the difference between the two curves in figure 4.1.C. The method of Tibshirani’s Gap Statistic is explained in more detail in section 4.3.1.

Figure 4.2 illustrates same method applied to a noisier dataset and the inconclusive results, which the Tibshirani gap statistic yields in these instances. In this second

Figure 4.2 Tibshirani's Gap Statistic applied to uniform toy dataset

Similar results might be achieved from a very noisy two-cluster dataset. **A.** Toy data in 2-dimensional feature space. **B.** Within sum of squares function, W_k , for the observed dataset. **C.** $\log(W_k)$ functions for the observed (O) and expected (E, average of all reference null distributions) datasets. **D.** Gap curve [53].

example, a rigid application of the gap statistic yields $k=1$, describing the uniformity of the dataset. This could be the result of a dataset where only one cluster is truly present, according to the features selected to define the data, or it could be the result of a dataset where the intrinsic clusters are not clearly defined, having many data points existing in the boundaries between clusters (figure 4.2.A). Figure 4.2.B illustrates how a dataset fitting this description yields a within sum of squares function for the observed data that does not contain an easily recognizable ‘elbow.’ A noisier dataset inflates the within

cluster sum of squares function. In this case, although intrinsic clusters may exist, it is difficult to distinguish the distribution of the observed points from the uniform distribution of the null reference sets (figure 4.2.C). Figure 4.2 is relevant when we later present our application of Tibshirani's Gap Statistic to our data. As a result of the inconclusiveness of our results using this method, we opted for an alternative method to reject our null hypothesis, which we proposed after our observation of differing ST wave propagation patterns present in our data.

To employ a permutation test, we needed to propose a likely number of clusters within a dataset and to have assigned our data to their respective clusters. We then needed to postulate a null hypothesis. To test the null hypothesis, we held the cluster membership of each data point (LFP spike ST pattern) fixed while randomly permuting our labels of interest (the occurrence of spikes during seizure vs. between seizures, our labels are derived from our null hypothesis), for one million permutations. For every permutation we recorded the proportion of data with each label within each cluster to obtain the null distribution. We then compared our observed data (label proportions within each cluster) to this null distribution. In this manner we determined not only that our null hypothesis was rejected but also which clusters in particular were the cause of this rejection.

4.3 Methods

4.3.1 The Gap Statistic

To reiterate from the introduction, Tibshirani's Gap Statistic is a measure to statistically arrive at the 'elbow' traditionally observed in the within cluster sum of squares function (W_k vs. k). In our application of Tibshirani's Gap Statistic, we chose values of k ranging from 1 to 30.

Tibshirani defines the gap statistic for sample size I , $G_I(k)$, as:

$$G_I(k) = E_I^*[\log(W_k)] - \log(W_k) \quad (4.1)$$

where $E_I^*[\log(W_k)]$ is the expectation of $\log(W_k)$ under a suitable null reference distribution of the data. Tibshirani proposes that an optimal k is determined when the gap statistic is the largest (taking its sampling distribution into account). To generate an appropriate reference null distribution, Tibshirani et al. recommends a uniform distribution over a box aligned with the principal components of the data. We generated 70 reference sets of randomly distributed data (uniformly covering the PCA dimension range for each feature representing each spike in the recorded data). We ran the k -medians algorithm using 30 random initializations for the number of cluster values, $k = 1$ through 30. We then chose the run that yielded the lowest $\log(W_k)$ for each k (for each of the 70 reference sets). The expected value for each $\log(W_k)$ was calculated as the average over the 70 reference sets. And, in accordance with Tibshirani et al., a corrected

standard deviation, s_k , accounting for simulation error, was computed as:

$$s_k = sd_k \sqrt{1 + \frac{1}{B}} \quad (4.2)$$

where sd_k is the standard deviation of the $\log(W_k)$ across the 70 reference sets and B is the number of reference sets, in our case 70. Finally, we determined the optimal number of clusters within the data by ascertaining the value of k that met the following criterion:

$$\hat{k} = \min\{k \mid Gap(k) \geq Gap(k + 1) - s_{k+1}\} \quad (4.3)$$

The chosen number of clusters is $\hat{k}$ (which we considered to be the optimal number of clusters within the data). If the above set were empty, the optimal number of clusters was set equal to the largest value of k tested, which in our case was 30.

4.3.2 The Permutation Test

The permutation test described below requires specifying the number of clusters. We did not know how many clusters were present in the dataset a priori. As a result, we elected to propose 10 clusters were present after our initial observation of numerous feature sets of detected spikes. We deemed that determining the exact number of clusters intrinsic in the data would not be crucial to our hypothesis. We opted to compare the cluster assignments of the ST patterns of LFP spikes within our dataset with the occurrence of seizures. Specifically, we elected to test the hypothesis that 2-D spike patterns during seizures (ictal spike patterns) could be differentiated from those occurring between seizures (interictal spike patterns) within the constraint of our assumption of 10 intrinsic clusters present within the data. Stated another way, we hypothesized that ST pattern of

only certain clusters occurred preferentially during seizure epochs. (As noted in the caption to Table 3.1, seizures were marked by a board-certified expert clinical neurophysiologist and were defined to be at least 5 seconds in length.) We employed a permutation test to test the null hypothesis that the proportion of spikes occurring during seizures is equal across clusters (i.e. ST patterns). As previously mentioned, during the permutation test, we held the cluster membership of each data point (spike ST pattern) fixed while randomly permuting the seizure/non-seizure-labels for one million permutations. For every permutation we recorded the proportion of ST patterns within each cluster to obtain the null distribution. We then compared the observed proportion of seizure and non-seizure spike patterns within each cluster to this null distribution.

4.4 Results

Figure 4.3 depicts a sequence of plots illustrating the results of applying Tibshirani's Gap Statistic to our data from the Cat 1 dataset. The layout of the four plots is identical to the layout in figures 4.1 and 4.2 for ease of comparison. Figure 4.3.A is a scatter plot of the data points in the Cat 1 data set in 2-dimensional feature space (first 2 principal components). This is the same plot as in figure 3.9.A, however with different colors representing different cluster assignments. Figure 4.3.B displays the within sum of squares function (W_k vs. k), for our observed dataset. Figure 4.3.C, illustrates the $\log(\text{within sum of squares})$ functions for the observed and expected datasets (averaged over all 70 reference sets). Finally, the plot in figure 4.3.D results when we took the difference of the two curves in figure 4.3.C according to equation 4.1. Next, using

equation 4.3, we identified the point at which this difference was statistically maximal, the cluster number, k , which separates our observed data most from the null distributions. This method yielded an estimate of the number of clusters to be thirteen indicating thirteen distinct patterns of wave propagation of LFP spikes across the array.

Figure 4.3 Tibshirani Gap Statistic applied to Cat 1 dataset

Results achieved following application to Cat 1 dataset. There were deemed to be 13 intrinsic clusters present. **A.** Cat 1 data in 2-dimensional feature space (first 2 principal components). Colors assigned to the data correspond to the 13 clusters revealed. (Same plot as in figure 3.9.A, however in figure 3.9.A the data were colored to correspond to the 10 fixed clusters later chosen to exist within the data). **B.** Within sum of squares function, W_k , for the observed dataset. **C.** $\log(W_k)$ functions for the observed (O) and expected (E, average of all reference null distributions) datasets. **D.** Gap curve.

In the figures capturing the results of Tibshirani's Gap Statistic on our dataset (figure 4.3), we observed that the results are much less conclusive when comparing them to the

results from the toy dataset depicted in figure 4.1. We explore our thoughts about these results in the discussion, section 4.5. The immediate consequence of the lack of convincing evidence following the application of Tibshirani's Gap Statistic on our Cat 1 dataset was that we chose to discontinue the application of Tibshirani's Gap Statistic and instead to fix the number of clusters at 10 across all datasets. We then chose to perform a permutation test to reject our null hypothesis. Our results are captured in figure 4.4 and tables 4.1-4.4.

Figure 4.4 Ratios of inter-ictal or ictal spikes to total spikes, by cluster

In the Cat 1 and 2 datasets, the majority of spikes detected occurred interictally (86.3% and 94.1% respectively, represented by the green horizontal lines in **A.** and **B.**) In the Cat 3a and 3b datasets, the majority of spikes detected occurred ictally (81.4% and 78.1% respectively, represented by the purple horizontal lines in **C.** and **D.**). The blue bars in **A.** and **B.** depict the observed ratio of interictal spikes to total spikes within a given cluster. The red bars in **C.** and **D.** depict the observed ratio of ictal spikes to total spikes within a given cluster. In all bar graphs, a red asterisk identifies the clusters that showed a statistically significant ($p < .007$) preference for containing LFP spikes that occurred during seizure (ictal epochs); a black asterisk identifies the clusters that showed a statistically significant preference for containing LFP spikes that occurred between seizures (interictal epochs). These graphs were generated from the data contained in Tables 4.1-4.4.

Table 4.1 Cat 1, 10 cluster permutation test results
(n=1,000,000)

Cluster	Spike Ratio (Interictal/Total)	Ictal Percentage (13.7%)	Number of Occurrences in Test	P Value	Interictal Percentage (86.3%)	Number of Occurrences in Test	P Value
1	80/215	62.79%	0	<<0.001	37.21%	n/a	n/a
2	220/273	19.41%	3426	.003426	80.59%	n/a	n/a
3	372/439	15.26%	166399	.166399	84.74%	n/a	n/a
4	313/357	12.32%	n/a	n/a	87.68%	190023	0.19
5	129/147	12.24%	n/a	n/a	87.76%	265737	0.27
6	440/484	9.09%	n/a	n/a	90.91%	549	.000549
7	340/357	4.76%	n/a	n/a	95.24%	0	<<0.001
8	253/265	4.53%	n/a	n/a	95.47%	1	0.000001
9	202/206	1.94%	n/a	n/a	98.06%	0	<<0.001
10	149/151	1.32%	n/a	n/a	98.68%	0	<<0.001

The total number of spikes in this dataset is 2894 (2498 spikes occurred interictally and 396 spikes occurred ictally; ictal epochs (seizures) were marked by a board-certified epileptologist). As stated in the figure caption for figure 4.4, the majority of spikes detected in the Cat 1 dataset occurred interictally (86.3%). The p-values in red identify the clusters which showed a statistically significant composition of spikes occurring ictally, corresponding to the blue bars with a red asterisk in figure 4.4; and the p-values in blue identify the clusters which showed a statistically significant composition of spikes occurring interictally, corresponding to the blue bars with a black asterisk in figure 4.4.

Table 4.2 Cat 2, 10 cluster permutation test results
(n=1,000,000)

Cluster	Spike Ratio (Interictal/Total)	Ictal Percentage (5.9%)	Number of Occurrences in Test	P Value	Interictal Percentage (94.1%)	Number of Occurrences in Test	P Value
1	18/23	21.74%	6211	.006211	78.26%	n/a	n/a
2	26/29	10.34%	236357	.236357	89.66%	n/a	n/a
3	33/35	5.71%	n/a	n/a	94.29%	661365	.661365
4	35/37	5.41%	n/a	n/a	94.59%	625392	.625392
5	26/27	3.70%	n/a	n/a	96.30%	513143	.513143
6	30/31	3.23%	n/a	n/a	96.77%	432780	.432780
7	55/56	1.79%	n/a	n/a	98.21%	119469	.119469
8	1/1	0%	n/a	n/a	100%	941077	.941077
9	7/7	0%	n/a	n/a	100%	650875	.650875
10	9/9	0%	n/a	n/a	100%	573831	.573831

The total number of spikes in this dataset is 255 (240 spikes occurred interictally and 15 spikes occurred ictally; ictal epochs (seizures) were marked by a board-certified epileptologist). As stated in the figure caption for figure 4.4, the majority of spikes detected in the Cat 2 dataset occurred interictally (94.1%). The p-value in red identifies the cluster which showed a statistically significant composition of spikes occurring ictally and corresponds to the blue bar with a red asterisk in figure 4.4.

Table 4.3 Cat 3a, 10 cluster permutation test results
(n=1,000,000)

Cluster	Spike Ratio (Interictal/Total)	Ictal Percentage (81.4%)	Number of Occurrences in Test	P Value	Interictal Percentage (18.6%)	Number of Occurrences in Test	P Value
1	88/2145	95.90%	0	<<0.001	4.10%	n/a	n/a
2	134/1355	90.11%	0	<<0.001	9.89%	n/a	n/a
3	155/1491	89.60%	0	<<0.001	10.4%	n/a	n/a
4	227/1903	88.07%	0	<<0.001	11.93%	n/a	n/a
5	149/1175	87.32%	0	<<0.001	12.68%	n/a	n/a
6	194/1409	86.23%	0	<<0.001	13.77%	n/a	n/a
7	335/2235	85.01%	0	<<0.001	14.99%	n/a	n/a
8	311/1431	78.27%	n/a	n/a	21.73%	980	.000980
9	515/733	29.74%	n/a	n/a	70.26%	0	<<0.001
10	599/651	7.99%	n/a	n/a	92.01%	0	<<0.001

The total number of spikes in this dataset is 14528 (2707 spikes occurred interictally and 11821 spikes occurred ictally; ictal epochs (seizures) were marked by a board-certified epileptologist). As stated in the figure caption for figure 4.4, the majority of spikes detected in the Cat 3a dataset occurred ictally (81.4%). The p-values in red identify the clusters which showed a statistically significant composition of spikes occurring ictally, corresponding to the red bars with a red asterisk in figure 4.4; and the p-values in blue identify the clusters which showed a statistically significant composition of spikes occurring interictally, corresponding to the red bars with a black asterisk in figure 4.4.

Table 4.4 Cat 3b, 10 cluster permutation test results
(n=1,000,000)

Cluster	Spike Ratio (Interictal/Total)	Ictal Percentage (78.1%)	Number of Occurrences in Test	P Value	Interictal Percentage (21.9%)	Number of Occurrences in Test	P Value
1	56/903	93.80%	0	<<0.001	6.20%	n/a	n/a
2	157/1487	89.44%	0	<<0.001	10.56%	n/a	n/a
3	116/767	84.88%	1	0.000001	15.12%	n/a	n/a
4	196/1175	83.32%	1	0.000001	16.68%	n/a	n/a
5	224/1303	82.81%	3	0.000003	17.19%	n/a	n/a
6	134/687	80.49%	60354	0.060354	19.51%	n/a	n/a
7	223/1033	78.41%	412965	0.412965	21.59%	n/a	n/a
8	114/215	46.98%	n/a	n/a	53.02%	0	<<0.001
9	397/671	40.83%	n/a	n/a	59.17%	0	<<0.001
10	279/413	32.45%	n/a	n/a	67.55%	0	<<0.001

The total number of spikes in this dataset is 8654 (1896 spikes occurred interictally and 6758 spikes occurred ictally; ictal epochs (seizures) were marked by a board-certified epileptologist). As stated in the figure caption for figure 4.4, the majority of spikes detected in the Cat 3b dataset occurred ictally (78.1%). The p-values in red identify the clusters which showed a statistically significant composition of spikes

occurring ictally, corresponding to the red bars with a red asterisk in figure 4.4; and the p-values in blue identify the clusters which showed a statistically significant composition of spikes occurring interictally, corresponding to the red bars with a black asterisk in figure 4.4.

4.5 Discussion

We concluded that the most likely reason behind our inconclusive results following the application of Tibshirani’s Gap Statistic on our Cat 1 dataset was due to the noisiness of the data. In our Cat 1 dataset, the distribution of LFP spike wave propagation patterns was fairly uniform and therefore the distribution approximated a null distribution. This could be a result of our choice of features not being specific enough to distinguish one pattern from another. This could also be the result of too few LFP spike detections (2894) relative to the number of PCA dimensions (251, see table 3.2) defining the space within which the clustering was performed. Both possibilities point toward opportunities for future improvements in our data analysis and a short-term necessity to collect larger and less noisy datasets.

As mentioned in section 4.4, the immediate consequence of the lack of convincing evidence following the application of Tibshirani’s Gap Statistic was our decision to fix the number of clusters at 10 across all datasets and to perform a permutation test to reject our null hypothesis. Our null hypothesis was that ‘the proportion of spikes occurring during seizures is equal across clusters.’

One advantage of fixing the number of clusters followed by the application of a statistical test (as opposed to applying a statistical test to arrive at a variable number of clusters

intrinsic to different datasets) was the ease of comparing statistical results across datasets. One disadvantage of fixing the number of clusters was the increased probability that certain clusters were combinations of multiple smaller clusters that should be separate; and the increased probability that other, separate clusters, were actually members of a single cluster which was over-split. However, despite this disadvantage, 10 clusters were deemed statistically sufficient and our permutation test was successful in rejecting our null hypothesis. As a result, we can conclude that these epileptic LFP spike wave propagation patterns are informative and we can continue to explore their relationship to other observed phenomena of epilepsy.

5 Temporal Data Mining

5.1 Summary

Temporal Data Mining (TDM) involves finding significant sequential patterns from large amounts of symbolic time series data [54]. TDM has previously been employed to infer network connectivity from neuronal spike data [55]. We were interested in TDM to determine if there are frequently occurring LFP spatio-temporal (ST)-pattern sequences that identify current or imminent seizure states. After clustering to identify ST patterns we analyzed the progression of these patterns (characterized by pattern label and time) under the hypothesis that we would detect statistically significant sequences of patterns repeating throughout the data. We aimed to identify and correlate these significant sequences to current or imminent seizure state and thereby to postulate mechanisms of seizure generation, progression and termination on a mesoscopic scale. As a concrete example of the task, cardiac electrophysiologists query the heart to look for sequences of

waveforms that create re-entrant rhythms and consequently dangerous arrhythmias. In these patients there are specific waveforms that trigger these events, corresponding to specific spatio-temporal patterns. We hypothesized that seizures might be triggered in the same way, by specific ST patterns or sequences of patterns that trigger clinically significant events.

In this chapter, we present our four main results: (1) When analyzing the data for statistically significant sequences of spike ST-patterns of length 5, the number of observations was dramatically reduced from the number of spikes detected and the resulting temporal occurrence of these sequences showed no statistically significant relationship to seizure onset or termination; (2) When analyzing the data for spike ST-patterns of length 2, we observed a statistically significant cluster assignment repeatability/stability (i.e. if a given spike was assigned to cluster 3, the following spike had a much higher than chance probability of also being assigned to cluster 3); (3) Following the application of three different correlation metrics that take into account the cluster assignments of spike wave propagation patterns we observed seizure windows in the data with higher relative similarity; and (4) We observed that the cluster composition of the ST-patterns in the seizures occurring during status epilepticus changed over time (found in the Cat 3a dataset).

5.2 Introduction

Frequent sequence discovery is a challenging statistical and computational problem and we drew on the work of Patnaik et al. to address it [55]. Patnaik et al. leveraged TDM in analyzing simultaneous recordings of spike train data from hundreds of neurons in order to infer underlying neuronal connectivity. We applied their algorithms for discovering serial sequences. Similar to Patnaik, we were motivated to look for causative chains of events (ST patterns) that happen at different times, repeatedly. Therefore, the TDM algorithm we implemented was designed with the constraint to find non-overlapping occurrences of ordered series of events, which we called episodes. Episodes are considered to be non-overlapping when no event in one occurrence appears simultaneously as a member of the previous or following occurrence. In addition to revealing causative chains, another important benefit of this constraint is the capability to apply computationally efficient TDM techniques to circumvent the infeasible task (due to combinatorial explosion) of exhaustively counting all of the possible episodes present in the data. Non-overlapping episodes must necessarily be comprised of frequent sub-episodes. This gives rise to a ‘level-wise’ algorithm.

To implement a ‘level-wise’ algorithm, we started by generating a histogram of all single-pattern events. Then, we built a set of candidate 2-pattern episodes. The single-pattern events used to build these episodes were only those whose frequency exceeded a user-defined threshold. We then generated a frequency histogram of 2-pattern episodes. Again, we applied a frequency threshold before using the 2-pattern episodes to build a set

of candidate 3-pattern episodes. We continued the application of this iterative algorithm until no n -length candidate episodes remained.

In our attempt to quantify the capability of these episodes to indicate current or imminent seizure state, we followed by determining the times of the occurrences of these episodes relative to the times of seizure onset and termination within the data. We also opted to calculate several correlation metrics between seizure-start aligned windows to quantify the similarity between them.

5.3 Methods

5.3.1 Data description

The dataset we analyzed was comprised of the results obtained from Chapter 3. We analyzed the 2,894 spikes detected and clustered into 10 clusters of distinct ST-patterns from Cat 1, the 255 spikes from Cat 2, the 14,528 spikes from Cat 3a and the 8,654 spikes from Cat 3b (spikes from a separate portion of the third animal experiment collected after moving the position of the recording array on the cortex of the cat). The amount of recording time for each dataset (summarized in Table 3.1) was: 53 minutes and 37 seconds from Cat 1; 3 hours, 15 minutes and 30 seconds for Cat 2; and a total of 5 hours, 2 minutes and 4 seconds from Cat 3 (4 hours, 24 minutes and 15 seconds prior to repositioning the array and 37 minutes and 49 seconds post repositioning the array). Because of minor variations in experimental protocol employed while obtaining these datasets, the TDM analysis was performed on each dataset independently.

Spike detection rasters and their respective cluster assignments are depicted in figure 5.1.

5.3.2 TDM-5 algorithm

As outlined in the introduction, we implemented a ‘level-wise’ algorithm. For each dataset, we started by generating a histogram of all single-pattern events. This histogram contained ten bars corresponding to the ten clusters. We set our user-defined threshold to retain only the top five most frequent clusters. Combining five 1st-clusters with ten possible 2nd-clusters yielded 50 candidate two-pattern episodes. We then generated a frequency histogram of these 50 two-pattern episodes. Again, we applied our threshold to retain only the top five most frequent two-pattern episodes before building a set (50) of candidate three-pattern episodes. We continued this iterative algorithm for the datasets from Cat 1, 3a and 3b until we arrived at the five most frequent five-pattern episodes. Because of the smaller initial dataset for Cat 2 (255 detected spikes), we stopped the iterative ‘level-wise’ algorithm when we arrived at the three most frequent four-pattern episodes.

We then applied a slight modification to the above algorithm for the Cat 1 dataset only. We repeated the above algorithm, however, we defined a repetitive sequence of cluster labeled spikes (spikes belonging to the same cluster) as a single ‘state.’ For example, a solitary spike belonging to cluster 5 (i.e. within the spike sequence 11132544) was assigned the ‘state’ 5 just as 2 sequential spikes assigned to cluster 5 (i.e. within the spike

Figure 5.1 Spike detection rasters for each of the four datasets (Cat 1-3b)

The segments in green are observer-identified epochs of noise across all channels of the array. Noise segments were removed and were not included in any stages of data analysis. The segments in yellow represent gaps in between files recorded during the experiments; the gaps are of arbitrary length for visualization purposes only. Seizure starts and stops are depicted with red and black vertical lines, respectively. Cat 1 experienced 7 seizures, Cat 2, 17 seizures and Cat 3, 76 seizures prior to repositioning the array and 23 seizures post repositioning the array (datasets Cat 3a and Cat 3b, respectively). Spikes were detected according to the algorithm in Chapter 3. Each spike was then assigned to one of 10 clusters of ST wave propagation patterns, again according to the algorithm in Chapter 3. Cluster assignment is on the y-axis.

CAT 3a: ~21.5 + ~19.6 + ~18.8 + ~56.7 + ~22.1 + ~35.8 + ~26.0 + ~11.4 mins of recording (No seizures or spikes detected for first 52.5 mins of recording). Number of detected spikes = 4 + 17 + 2 + 851 + 8090 + 4307 + 1257

CAT 3b: ~37.8 mins of recording, Number of detected spikes = 8654

Figure 5.1 Spike detection rasters for each of the four datasets (Cat 1-3b)

The segments in green are observer-identified epochs of noise across all channels of the array. Noise segments were removed and were not included in any stages of data analysis. The segments in yellow represent gaps in between files recorded during the experiments; the gaps are of arbitrary length for visualization purposes only. Seizure starts and stops are depicted with red and black vertical lines, respectively. Cat 1 experienced 7 seizures, Cat 2, 17 seizures and Cat 3, 76 seizures prior to repositioning the array and 23 seizures post repositioning the array (datasets Cat 3a and Cat 3b, respectively). Spikes were detected according to the algorithm in Chapter 3. Each spike was then assigned to one of 10 clusters of ST wave propagation patterns, again according to the algorithm in Chapter 3. Cluster assignment is on the y-axis.

sequence 11132554) were also assigned the ‘state’ 5; however, the 2 examples would have variable lengths of time within the singular ‘state,’ 5. The onset time of the cluster state was defined as the time of the first spike in a series of identically assigned spikes and the offset time of the cluster state was defined as the time of the first spike immediately following the end of a series of identically assigned spikes.

5.3.3 TDM-2 (Analysis of Transition Matrices)

We also chose to look at TDM pattern sequences of length two (TDM-2). In this analysis, we did not focus on particular sequences; instead we differentiated between sequences where two spikes were assigned to the same cluster and sequences where two spikes were assigned to different clusters (i.e. the sequences 1-1 and 3-3 would be members of the first category, while sequences 9-1 and 3-4 would be members of the second). Our objective was to address the question, ‘is the probability of a repeat cluster assignment for any two given spikes greater than chance?’ A variation of this question is, ‘what is the stability of cluster membership in sequential spikes?’ To answer this question we first constructed a transition matrix (a data matrix with cluster membership of the first spike on the y-axis, cluster membership of the second spike on the x-axis and populated with frequency counts for all combinations of two-pattern sequences). The diagonal of this transition matrix contained the counts of TDM-2 sequences comprised of spikes assigned to identical clusters and all off-diagonal count entries contained the counts of TDM-2 sequences comprised of spikes assigned to different clusters. We conducted this analysis only on the Cat 1 dataset.

To test the null hypothesis that ‘a spike belonging to a given cluster has a probability equal to chance of being followed by a spike belonging to any cluster,’ we conducted a permutation test ($n=1,000,000$), fixing the number of spikes per cluster to equal the number found within the dataset but shuffling the temporal order of cluster labels.

5.3.4 Various correlations across SZ-aligned data

Finally, we decided to calculate a few correlations between seizure-aligned data segments. We began by generating spike detection rasters, aligned by seizure onset for each dataset (the fixed windows containing onset in the center were set wide enough to capture the length of the longest marked seizure within a given dataset). The rasters for Cat 1 and 2 are depicted in figure 5.2. The rasters for Cat 3a and 3b are depicted in figure 5.12.

Each spike detection was colored according to its cluster assignment. The areas shaded in grey illustrate the time during marked seizures; the areas shaded in pink illustrate epochs of noise and the areas shaded in dusty blue mark gaps between consecutive files of recorded data.

Figure 5.2 only includes the rasters for Cat 1 and 2 because, ultimately, we implemented the correlation metrics only on the Cat 1 dataset. We decided, regardless of the results, that we would not implement the correlation analysis methods on the Cat 2 dataset due to the sparseness of spikes detected during these seizure-aligned windows, as can be seen in figure 5.2.B. Additionally, we found the results of this analysis method were vague for

Figure 5.2 Seizure-start aligned spike rasters for Cats 1 and 2

A. Cat 1 spike rasters for 7 identified seizures. Time along the x-axis is from seizure onset $\pm$ 45 secs. **B.** Cat 2 spike rasters for 17 identified seizures. Time along the x-axis is from seizure onset $\pm$ 65 secs. Seizure-start aligned spike rasters for Cat 3a and 3b are depicted in figure 5.12, section 5.4.4. The legend for both datasets is positioned on the bottom right of the above figure. The colors that identify each of the various clusters are consistent across cats; however Cluster 1 in Cat 1 is not equivalent to Cluster 1 in Cat 2, etc. since each dataset was clustered independently.

Cat 1. This deterred us from proceeding to implement these correlation metrics on the remaining datasets (Cat 3a and 3b). The seizure-start aligned spike rasters for Cat 3a and 3b are included later in the chapter (figure 5.12) when we discuss our observations regarding spike cluster membership composition of seizures in an episode of status epilepticus.

We implemented four different correlation metrics to the Cat 1 dataset to quantify the similarity between spike sequences across marked seizures.

The first metric we implemented was a cross-correlation of spike cluster membership between the 7 segments containing the marked seizures from Cat 1. Each data segment was reduced to a variable-length string of cluster labels. Time of spike occurrence was not retained, only the sequence order of cluster membership. When two spike sequences were of differing lengths, the shorter sequence was zero-padded. Fixed length 90-second data segments yielded variable length spike sequences because the number of spike detections varied according to spike frequency throughout the dataset. The cross-correlation was then conducted by sliding one string across the other. When spike cluster labels were in agreement, a value of 1 was assigned and when spike cluster labels were not in agreement, a value of 0 was assigned. These values were then summed for all overlapping spikes within a given shift in relative position of the two sequences. This is summarized in the following cross-correlation equation for discrete functions:

$$(f \star g)[n] \stackrel{\text{def}}{=} \sum_{m=-\infty}^{\infty} f[m] g[n + m]. \quad (5.1)$$

where f and g represent the two sequences being cross-correlated and n is the relative shift between them. Finally, to arrive at a single value for our metric (as opposed to a function of shifts) we summed across all shifts (all values of n). Then, for visualization purposes, we normalized the results (dividing all 49 cross-correlation values between the 7 seizure segments by the maximum).

The second metric we implemented was a slight variation on the first. As opposed to ignoring the time of each spike occurrence, we wanted to preserve the relative timing between detected spikes. Therefore, each spike sequence was padded with ‘not a number’ (NaN, serving as temporal placeholders) for all the instances that a spike was not detected within a given train (every 10 msecs). As a result of this padding, all sequences were of equal length prior to calculating the cross-correlation between seizure segments. The calculations used in applying the first metric were then repeated.

For the third metric, we calculated the correlation coefficient to quantify the correlation between transition matrices of the spike sequences of the 7 seizure segments. We generated the transition matrices in the same manner as was described in section 5.3.3. We then calculated the correlation coefficient matrix by determining the covariance between the 49 pair combinations of the 7 transition matrices and dividing by the product

of a given pair's standard deviations. This is summarized in the following correlation coefficient equation:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} \quad (5.2)$$

where $\rho_{X,Y}$ is the correlation coefficient calculated from seizure segments X and Y; and σ_X and σ_Y are their respective standard deviations.

Our fourth and final correlation metric calculated the correlation coefficient (equation 5.2) of the 49 pair combinations of the average of the raw seizure segment traces (mean signal across the 360 electrodes of the array), to approximate a comparison of our first three metrics to a metric that would be possible from a clinical recording lacking the high spatial-density of our recording array.

5.4 Results

5.4.1 Most frequent TDM sequences of length 5

Figure 5.3 depicts the rasters of sequences, which resulted from the application of the TDM-5 algorithm, locating the most frequent pattern sequences in the four cat spike datasets. Table 5.1 summarizes our observations from figure 5.3. The three columns for temporal distance of pattern sequences to points of clinical significance within the dataset are for comment only since we concluded that the number of occurrences of the most frequent TDM-5 sequences were too few to detect any trends with statistical significance. Following the observation of a lack of quantitative conclusions from our initial results, we applied a slight modification to the TDM-5 algorithm for the Cat 1 dataset only. In the

modified algorithm we defined a series of cluster labeled spikes belonging to the same cluster as a single 'state' prior to locating the most frequent 'state' sequences (section 5.3.2). This modification yielded results that were no more informative than the raster of individual spike pattern sequences. The raster from this modified TDM-5 algorithm is depicted in figure 5.4.

Figure 5.3 Most frequent TDM-5 pattern rasters (Cats 1-3b)

The most frequent pattern sequences of length 5 were detected according to the algorithm described in section 5.3.2. Pattern sequences are identified on the y-axes of the above rasters. Note: the dataset for Cat 2, only depicts the rasters of the 3 most frequent pattern sequences of length 4 due the small size of the dataset. Cat 1 experienced 7 seizures, Cat 2 17 seizures and Cat 3, 76 seizures prior to repositioning the array and 23 seizures post repositioning the array (datasets Cat 3a and Cat 3b, respectively).

The segments in green are observer-identified epochs of noise across all channels of the array, which were removed and were not included in any stages of data analysis. The segments in yellow represent gaps in between files recorded during the experiments; the gaps are of arbitrary length for visualization purposes only. Seizure start and stop times identified in the data are depicted with red and black vertical lines, respectively.

Figure 5.3 Most frequent TDM-5 pattern rasters (Cats 1-3b)

The most frequent pattern sequences of length 5 were detected according to the algorithm described in section 5.3.2. Pattern sequences are identified on the y-axes of the above rasters. Note: the dataset for Cat 2, only depicts the rasters of the 3 most frequent pattern sequences of length 4 due the small size of the dataset. Cat 1 experienced 7 seizures, Cat 2 17 seizures and Cat 3, 76 seizures prior to repositioning the array and 23 seizures post repositioning the array (datasets Cat 3a and Cat 3b, respectively).

The segments in green are observer-identified epochs of noise across all channels of the array, which were removed and were not included in any stages of data analysis. The segments in yellow represent gaps in between files recorded during the experiments; the gaps are of arbitrary length for visualization purposes only. Seizure start and stop times identified in the data are depicted with red and black vertical lines, respectively.

Table 5.1 TDM-5 data summary

	Top 5 Most Frequent Sequences	Frequency (Number of Observations)	Temporal Distance from Termination of Previous Seizure (seconds)**	Temporal Distance to Onset of Next Seizure (seconds)**	Temporal Distance to Termination of Current Seizure (seconds)***	Number of Detected Spikes/ Seizures in Dataset
Cat 1	1-1-1-1-1	13	70+/-46 (n=11)	45+/-37 (n=6)	5 (n=1)	2894/7
	1-1-1-1-8	10	68+/-46 (n=9)	57+/-24 (n=5)	n/a (n=0)	
	8-8-8-8-8	15	331+/-129 (n=12)	274+/-137 (n=12)	24+/-5 (n=3)	
	8-8-10-8-8	11	386+/-121 (n=11)	257+/-117 (n=11)	n/a (n=0)	
	10-8-8-8-8	12	312+/-163 (n=12)	261+/-119 (n=12)	n/a (n=0)	
Cat 2*	1-5-5-1	2	n/a (n=0)	4635+/-3262 (n=2)	n/a (n=0)	255/17
	5-5-1-5	2	n/a (n=0)	5693+/-1810 (n=2)	n/a (n=0)	
	5-5-5-5	3	n/a (n=0)	2211+/-2166 (n=3)	n/a (n=0)	
Cat 3a	4-4-4-4-4	30	n/a (n=0)	n/a (n=0)	9+/-6 (n=28)	14,528/76
	4-4-4-4-10	23	n/a (n=0)	n/a (n=0)	7+/-7 (n=23)	
	4-4-10-4-4	27	n/a (n=0)	n/a (n=0)	9+/-7 (n=27)	
	9-9-9-9-6	23	n/a (n=0)	n/a (n=0)	33+/-15 (n=22)	
	9-9-9-9-9	47	n/a (n=0)	n/a (n=0)	40+/-14 (n=42)	
Cat 3b	9-9-8-8-8	8	104 (n=1)	2 (n=1)	17+/-10 (n=7)	8,654/23
	9-9-9-8-9	7	n/a (n=0)	n/a (n=0)	34+/-19 (n=7)	
	9-9-9-9-7	7	n/a (n=0)	n/a (n=0)	29+/-20 (n=7)	
	9-9-9-9-9	10	n/a (n=0)	n/a (n=0)	24+/-16 (n=10)	
	9-9-9-9-10	7	30+/-30 (n=2)	43+/-32 (n=2)	20+/-22 (n=5)	

Three columns list temporal distance of the most frequent pattern sequences to points of clinical significance, termination of previous seizure, onset of next seizure and termination of current seizure.

**Inter-ictal occurrences only; not calculated if a file gap existed between pattern occurrence and seizure.

***Ictal occurrences only.

Number of detected spikes and seizures are shown in the last column as a reference.

*Analysis for Cat 2 was halted after identifying the top 3 pattern sequences of length 4. This is because the dataset for Cat 2 contained very few spike detections. The next 19 most frequent pattern sequences of length 4 (after the top 3 listed in the table) in this dataset were all tied with a frequency of 1.

Figure 5.4 Most frequent TDM-5 'state' rasters (Cat 1)

'State' sequences are identified on the y-axes. A 'state' was defined as a string of spikes belonging to the same cluster (section 5.3.2). The green, yellow, red and black colors represent the same features of the data as described in figure 5.3. Cat 1 experienced 7 seizures.

5.4.2 TDM sequences of length 2

Figure 5.5 is of the transition matrix for cluster assignments for Cat 1 spikes.

Figure 5.5 Transition matrix for spikes in the Cat 1 dataset

The left axis depicts the cluster assignment of the first of two spikes in a TDM sequence of length 2; and the top axis depicts the cluster assignment of the second of the two spikes within the sequence. The colorbar on the right details the spike counts associated with each color.

We observed the highest spike counts along the diagonal of the transition matrix. We chose to sum the diagonal entries and to perform a permutation test ($n=1,000,000$) to test the hypothesis that ‘TDM sequences of length 2 show a higher than chance probability of containing 2 spikes from the same cluster.’ We found that 37.4% ($n=1082$) of the total number of spike sequence transitions was of spikes from a given cluster to spikes of the same cluster (diagonal entries). The results of our permutation test yielded a maximum diagonal count of 14.3% of the total number of transitions ($p \ll 0.001$), leading us to speculate that our observation of a much larger percentage of repeatable spike transitions (37.4%) is attributable to a biological cause, since its occurrence is much greater than chance.

We then plotted the relationship of this observed cluster assignment ‘stability’ of spike transitions (TDM-2) versus proximity to next seizure onset (grouping all 7 seizures from the Cat 1 dataset together). We wanted to ascertain if there is any change in the repeatability of cluster assignments of two sequential spikes as the brain approaches the seizure state. Our result is shown in figure 5.6.

Figure 5.6 Relationship of repeating and non-repeating spike transitions to seizure onset
The above figure depicts the percentage of non-repeating patterns and repeating patterns vs. time to next seizure onset after grouping together all 7 seizures detected in the Cat 1 dataset. There were variable lengths of time prior to each seizure onset. (957 seconds prior to seizure 1; and 128, 543, 694, 265, 87 and 178 seconds prior to seizures 2 through 7 respectively). Above, we opted to depict the relationship only from 265 seconds prior to seizure onset because extending the plot to the left merely showed a stable horizontal line for both functions, with the percentage of non-repeating patterns at 62.3% $\pm$ 0.7% and percentage of repeating patterns at 37.7% $\pm$ 0.7%, confirming the percentages calculated from the entire length of the dataset, initially observed in the transition matrix in figure 5.5.

From this plot we observed that the percentage of transitions with repeating cluster assignments decreased as the recording approached seizure onset (starting around 160 seconds prior until approximately 15 seconds prior to onset). We speculated that transitions are more chaotic just prior to seizure onset, possibly indicating a loss in cluster membership stability as a seizure approaches. Figure 5.6 prompted us to examine whether this observation held for individual seizures, which we address below. We also observed that the percentage of transitions with repeating cluster assignments abruptly increased to steady-state percentage (~37.4%) just prior to onset. However, we attributed this abrupt change to an artifact of the reduced number of transitions occurring within this small window; recorded percentages increased in volatility approaching onset as window size for the calculation decreased.

To examine whether our observation of a decrease in percentage of repeating cluster assignment transitions as the recording approached seizure onset holds for individual seizures, we plotted the percentage of non-repeating and repeating patterns within a sliding 30 second window vs. time for each of the 7 seizures in the Cat 1 dataset. Figure 5.7 depicts our results.

Figure 5.7 Relationship of spike transitions to individual SZ onsets (Cat 1)

A sliding window of 30 seconds was used to calculate the percentages of repeating and non-repeating patterns to total patterns for every two-pattern sequence leading up to the 7 seizures in the Cat 1 dataset. Percentages above were plotted at the end of each window (i.e. the percentages plotted at -50 seconds are from a window spanning 50 to 80 seconds prior to seizure onset). In **B.** and **F.**, the plot does not extend to 120 seconds prior to seizure as a result of the sliding window size and insufficient time between consecutive seizure onsets. (The time between the termination of SZ 1 and the onset of SZ 2 was 128 seconds; the time between the start of a new recording file and the onset of SZ 6 was 87 seconds).

In figure 5.7, we noted that the trend we observed in figure 5.6 did not hold across all individual seizures. Rather, although we observed a decrease in the percentage of transitions with repeating cluster assignments with increasing proximity to seizure onset for seizures 3 and 5 (figure 5.6.C. and E.), for seizures 1 and 2 (figure 5.6.A. and B.) we

observed the opposite and further, for seizures 6 and 7 (figure 5.6.F. and G.) we observed neither an increase nor a decrease. Seizure 4 (figure 5.6.D.) did not seem to fit any of the above categories, first showing a rapid decrease before gradually increasing as seizure onset neared.

5.4.3 Correlations across SZ-aligned data

Figure 5.8 illustrates the results of applying the first two correlation metrics to our seizure-aligned spike raster data from the Cat 1 dataset. (As a consequence of our algorithm, seizure sequences with more spikes had higher values of cross-correlation, i.e. the maximum occurred for the auto cross-correlation of seizure 7).

Ignoring auto cross-correlations along the diagonals, we observed the highest values of cross-correlation existed between seizures 4 and 6, 4 and 7, 5 and 7, and 6 and 7, in both of the matrices. The agreement between the results from the first and second correlation

metrics indicated that the order of cluster assignments of detected spikes was more influential than their temporal occurrences within these windows. Referring back to the spike rasters for Cat 1 in figure 5.2, visually the most similar seizure spike rasters appeared to be between seizure windows 4 and 6, and 5 and 7, confirming the results from the first two correlation metrics (the similarity between windows 4 and 7, and 6 and 7 was difficult to confirm visually).

Figure 5.9 illustrates the result of applying the third correlation metric to our seizure-aligned spike raster data from the Cat 1 dataset. Figure 5.9.A. is of the 7 spike-pattern transition matrices generated from the 7 seizure data segments. Figure 5.9.B. is a matrix of the correlation coefficients between the matrices in figure 5.9.A.

B
Figure 5.9 Third correlation metric result for Cat 1

A. Illustrates the 7 transition matrices generated from the 7 seizure windows. **B.** Matrix of correlation coefficients calculated from the 7 transition matrices in **A**.

In figure 5.9.A, we observed that not all entries were populated, calling our attention to the fact that only certain transitions in cluster membership occurred within a given seizure window. Additionally, we observed that the total number of unique spike transitions within a seizure window, in general tended to increase over the length of the dataset. The number of unique spike transitions for seizure windows 1 through 7 were 17, 10, 20, 33, 35, 44 and 28, respectively. We speculated that this increasing trend was a sign of a decline in organization, that could be due to the increasing effects over time of the decreased inhibition in the cortex resulting from the application and consequent diffusion of picrotoxin. One final observation we drew from figure 5.9.A was of a loose symmetry across the diagonal in each transition matrix (i.e. in transition matrix 5, the number of wave propagation pattern transitions from cluster 1 to cluster 9 appears to be roughly equivalent to the number of transitions from cluster 9 to cluster 1). This led us to remark that bi-directional cluster membership transitions were much more common than

uni-directional transitions; however, we have not been able to speculate on cortical mechanisms that would produce this observation.

In figure 5.9.B, we made the observation that the correlation coefficient (calculated between the 7 transition matrices of the 7 seizure windows) was highest between seizures 4 and 6, followed by seizures 5 and 7. This observation agreed with our previous results, following the calculations of the first two correlation metrics.

Figure 5.10 illustrates the outcome of applying the fourth correlation metric to our seizure-aligned data from the Cat 1 dataset. We included this metric because calculation of correlation coefficients between the averages of the 360 raw traces of each seizure window can be accomplished using recordings from standard clinical electrodes.

Figure 5.10 Fourth correlation metric result for Cat 1
Correlation coefficient values on the diagonal were zeroed out.

From figure 5.10, we observed the strongest correlation existed between seizure windows 1 and 4. The average spike traces for the most positively correlated pair of seizure windows are depicted in figure 5.11. We were not sure what to deduce from this outcome

other than to note that it contradicts with the results from the first three correlation metrics.

Most Correlated - SZs 1&4

Figure 5.11 Raw traces average for most correlated seizure windows from Cat 1 using correlation metric 4

Figure 5.10 revealed that seizure windows 1 and 4 resulted in the highest off-diagonal correlation coefficient. The colors depicted in the voltage traces represent the cluster assignments of the detected spikes, using the same legend as in figures 5.2 and 5.12 and is for visualization only, as cluster assignment was not utilized in the calculation of the correlation coefficients. The areas shaded in grey depict the marked seizure epochs within each window; the areas shaded in pink depict epochs of noise.

5.4.4 Spike rasters of seizure-aligned windows (Cat 3a, 3b data)

Figure 5.12 depicts spike detection rasters aligned by seizure onset for the Cat 3a and 3b datasets. (Rasters for Cat 1 and 2 were depicted earlier in the chapter, figure 5.2.) Each spike detection was colored according to its cluster assignment. The areas shaded in grey illustrate the time during marked seizures; the areas shaded in pink illustrate epochs of noise and the areas shaded in dusty blue mark gaps between consecutive files of recorded data. Figure 5.12 was included in order to discuss our observation regarding spike cluster

A

Figure 5.12 Spike rasters from Cat 3a and 3b depicting change in cluster composition of seizures as status epilepticus develops

A. Cat 3a spike rasters for 76 identified seizures. Time along the x-axis is from seizure onset $\pm$ 70 secs.

The legend for both datasets is positioned to the right of the Cat 3b raster. The colors that identify each of the various clusters are consistent across cats; however Cluster 1 in Cat 3a is not equivalent to Cluster 1 in Cat 3b, etc. since each dataset was clustered independently. The empty areas shaded in grey preceding seizure onsets in some of the above rasters represent the marked seizures immediately preceding the seizure of interest in a given window; their spike cluster compositions are depicted in earlier rasters (i.e. SZ 3 is greyed out in the SZ 4 raster of Cat 3b, however, it is not greyed out in the SZ 3 raster immediately above it). Many seizures in the Cat 3a and 3b datasets occurred with less than 70 seconds between them resulting in rasters which temporally overlapped.

B

Figure 5.12 Spike rasters from Cat 3a and 3b depicting change in cluster composition of seizures as status epilepticus develops

B. Cat 3b spike rasters for 23 identified seizures. Time along the x-axis is from seizure onset ± 75 secs.

The legend for both datasets is positioned to the right of the Cat 3b raster. The colors that identify each of the various clusters are consistent across cats; however Cluster 1 in Cat 3a is not equivalent to Cluster 1 in Cat 3b, etc. since each dataset was clustered independently. The empty areas shaded in grey preceding seizure onsets in some of the above rasters represent the marked seizures immediately preceeding the seizure of interest in a given window; their spike cluster compositions are depicted in earlier rasters (i.e. SZ 3 is greyed out in the SZ 4 raster of Cat 3b, however, it is not greyed out in the SZ 3 raster immediately above it). Many seizures in the Cat 3a and 3b datasets occurred with less than 70 seconds between them resulting in rasters which temporally overlapped.

membership composition of status epilepticus seizures. During the Cat 3a experiment, the animal developed status epilepticus. We noted from the rasters in figure 5.12.A that the cluster assignment of detected LFP spikes composing seizures in this dataset progressed from spikes primarily assigned to ‘pink/purple’ clusters (with some blue) early in the recording, to spikes primarily assigned to ‘blue/green/yellow’ clusters later in the recording. This shift is subtle and seemed to occur approximately around seizures 25-32. We did not observe a progression in cluster assignment of spikes composing the seizures of Cat 3b; we postulated that this is because during the time frame in the recording of Cat 3b, the animal had reached a status epilepticus steady-state.

5.5 Discussion

Following the collection of results from our TDM analysis, we concluded that although sequence lengths of 5 patterns did not yield a sufficient number of occurrences from which to observe trends, sequence lengths of 2 patterns resulted in a few interesting observations about the repeatability of cluster assignments from one spike to the next throughout the length of a dataset and as the brain approached the seizure state (identified by marked seizure onsets). Pursuing an alternative tack, following the application of four correlation metrics across seizure-aligned windows to quantify the similarities between them, we observed conflicting results. Finally, during the course of our analysis, we made an observation concerning the cluster assignments of the spikes composing the seizures from the datasets of the cat that developed status epilepticus (Cat 3). The remainder of this section provides more detail in our discussion regarding the specifics of our results.

Based upon our TDM-5 analysis we concluded that identifying frequent pattern sequences of length 5 yielded so few observations that it was not possible to determine trends with any statistical power (table 5.1, figure 5.3). We did observe that interictal occurrences of TDM-5 sequences in the Cat 1 dataset were detected closer to subsequent seizure onsets than to previous seizure terminations. However, this may be due to our algorithmic choice to assign the time of TDM-5 sequence occurrence to the detection of the last spike within the sequence. The relatively small number of observations in this experiment, 61 across 5 sequences in Cat 1, emphasized that we must record more data to arrive at any statistically significant conclusions. We did not observe any trends worth testing via a permutation test in any of our four datasets. From Table 5.1, we did note that the majority of TDM-5 sequences in the Cat 1 and Cat 2 datasets occurred interictally (57 of 61 and 7 of 7 occurrences, respectively) and the majority of TDM-5 sequences in the Cat 3a and 3b datasets occurred ictally (142 of 150 and 36 of 39 occurrences, respectively). This was likely due to selection bias, as 86.3% and 94.1% of the spikes occurred interictally in Cat 1 and 2; and 81.4% and 78.1% and of the spikes occurred during seizures (ictally) in the Cat 3a and 3b datasets (figure 4.4). Additionally, our subsequent application of a slight modification to the TDM-5 algorithm for the Cat 1 dataset, where we defined a series of cluster labeled spikes belonging to the same cluster as belonging to a single ‘state,’ yielded similarly inconclusive results, confirming the need to collect more data in future experiments.

The TDM-2 analysis demonstrated that the percentage of spike transitions with repeating cluster assignments in the Cat 1 dataset was much greater than chance (37.4% observed >

14.3% maximum from permutation test, $p < 0.001$). Reducing the sequence length from 5 to 2 in our TDM analysis and grouping the sequences into two categories (repeating vs. non-repeating), as opposed to 49 unique sequences, resulted in an increased number of total occurrences (2893 vs. 61, Cat 1 dataset). This enabled us to statistically validate our observation via a permutation test ($n=1,000,000$), and led to our speculation that the higher than expected repeatability of cluster assignments from one spike to the next is attributable to a biological cause.

To further investigate our observation from our TDM-2 analysis, we wanted to determine if the percentage of spike transitions with repeating cluster assignments varied as the brain approached the seizure state. From figure 5.6, we observed that the percentage of transitions with repeating cluster assignments decreased as the recording approached seizure onset, possibly indicating a loss in cluster membership stability as a seizure approaches. This led us to speculate that transitions are more chaotic just prior to seizure onset and prompted us to examine whether this observation held for individual seizures. In figure 5.7, we noted that the trend we observed in figure 5.6 did not hold across all individual seizures. However, we believe it is possible that in a larger dataset with a greater number of detected seizures, we may more clearly discern a decreasing trend in the percentage of spike transitions with repeating cluster assignments leading up to individual seizure onsets. From our sparse dataset, we hypothesize that following an initial seizure genesis settling time (accounting for the opposite trend observed leading up to seizures 1 and 2), the trend toward a decrease in percentage of transitions with repeating cluster assignments reflects an increased disorganization of wave propagation

patterns for detected epileptiform LFP spikes that may warn of seizure onset. However, we need to validate this hypothesis with more data from future experiments.

Applying four correlation metrics across seizure-aligned data from the Cat 1 dataset, we noted that when we compared the observed seizure window similarities based on the first three metrics, which take into account spike wave propagation pattern cluster assignments, to the observed seizure window similarities based off of the fourth metric, which was calculated from the average of the raw traces from the array, we were left with a contradictory result. The fourth correlation metric was applied after the additional high spatial-density information that the array provides was discarded, approximating a recording from a standard clinical electrode. This confirmed for us that the information provided by the high spatial-density array is not redundant of the information already available from standard clinical arrays; however, determining which result yields more useful information remains unclear. It was not apparent what more could be concluded from our observed seizure window similarities. However, in the course of applying our third correlation metric to the Cat 1 dataset, we noted a general increase in the number of unique spike transitions within each seizure window over the course of the recording. We speculated that the cause of this increase was a progressive decline in organization, possibly due to the effects over time of the decreased inhibition in the cortex resulting from the application and consequent diffusion of picrotoxin.

Finally, we present an observation we made when we generated the spike rasters of seizure-aligned windows in preparation for the application of our correlation metrics. We

believe the gradual change in cluster assigned spike composition of seizures over time in figure 5.12.A suggests that circuits causing status epilepticus may change in stereotyped ways, pointing us towards an additional research direction we would like to pursue in future work.

We close our discussion conceding that although we could continue by applying our TDM-2 analysis and our four correlation metrics to the Cat 2, 3a and 3b datasets, given the sparseness of data available in the Cat 2 dataset and the noisiness of the three remaining datasets, we believe it would be more fruitful to collect newer higher quality data from longer recordings in future experiments. Additionally, while it is possible that the noisiness of the data could be overcome by advanced signal processing methods, there is another fundamental issue that gives us pause. This is that in all of the experiments performed, picrotoxin diffused throughout the recording area over time, causing enlargement of the functional epileptic network. The progression of the network in this fashion, observed in all the animals of this model, is different from the human condition, which in most cases reflects a fixed network, that fires synchronously at unpredictable intervals. Therefore, we look forward to the application of our data analysis methods to assess the repeatability of our observations and conclusions following the collection of new data from more appropriate animal models and ultimately from human patients.

6 Conclusion

6.1 Summary of Findings

In this work, we introduced a new signal analysis algorithm for representing and classifying detected epileptiform local field potential (LFP) spikes. This was done in order to leverage data recorded using novel electrode high spatial-density arrays placed on the cortical surface of the brain. We ran the algorithm on over 3300 channel-hours of cat μ ECoG data recorded *in vivo*. We demonstrated that the proportion of spikes occurring during seizures is not equal across clusters, specifically that some clusters contain a higher than expected proportion of ictal spikes while others were disproportionally interictal. Finally, we implemented temporal data mining techniques to understand seizure generation in an acute animal model, as described by the two-dimensional wave propagation patterns of these spike detections.

We succeeded in extracting novel features from epileptiform LFP spikes, which characterized their spatio-temporal wave propagation dynamics. Spikes from three separate cat experiments were clustered within each dataset into 10 distinct patterns described by these novel features. Statistical analysis validated the preference of some clusters for the ictal state and others for the interictal state.

Results from temporal data mining analysis demonstrated a trend toward decreasing percentage of spike transitions with repeating cluster assignments as the brain approached the seizure state. This finding suggests a spread of excitatory activity to novel neuronal pathways as seizures approach. This finding might also explain the observation in human epilepsy that there is often a disappearance of normal cortical oscillations and background activity in the seizure onset region prior to the incidence of spontaneous seizures. Although our observation did not appear to hold for all individual seizures, we believe that future experiments will validate our hypothesis that loss in cluster membership stability of detected spikes may be a biomarker of increasingly chaotic brain state. And this brain state may indicate a lower resistance to organizing events that initiate seizure onset. It may be precisely the loss of normal functional brain connectivity, destroying oscillations that stabilize networks in a carefully arranged infrastructure of refractory states, that enables widespread pathological synchronization of epileptic brain.

We acknowledge our limited ability to address current clinical questions with our research, given the artificial nature of the picrotoxin-induced model of spontaneous seizures we employed. However, addressing the question, ‘do seizures start the same way

every time,’ we feel confident in responding that results from our analysis indicate that they do not. We found that seizures within cats were neither generated by stereotyped two-dimensional LFP spikes nor by stereotypical sequences of wave-propagation patterns (using our selection of features extracted from the data to represent these 2-D patterns). Additionally, there did not appear to be any identifiable, deterministic wave-propagation pattern motifs of seizure progression or termination across seizures. In this way, despite the tempting analogy, seizures in our model were different than observations in some models of cardiac arrhythmias that have stereotyped patterns of arrhythmia initiation and termination. Additionally, our analysis did not reveal any prominent similarities between seizures across cats. To the contrary, our experiments suggested that seizures arise from increasingly chaotic cortical disorganization, perhaps supporting the idea that for the synchronization of seizures to begin, the natural organization of activity in the brain must first be disabled.

Despite all we have gleaned from a new approach to data analysis afforded by these novel high spatial-density arrays, there is still much more to be understood concerning the dynamics of pathological neuronal processing in the epileptic network. High resolution spatio-temporal recording may be key to this understanding. Some additional knowledge will come from new experiments. More will come from similar observations in naturally occurring epilepsy. Going forward, the issue of sampling error inherent when relying solely on surface recordings might eventually require experiments that simultaneously record from surface and penetrating arrays at high resolution to understand phenomena driving seizure generation. Even this approach has drawbacks, as

penetrating electrodes injure brain tissue in such a way as to potentially preclude observing the natural state of these complex systems. We feel confident that much remains to be revealed as we continue to probe the two- and eventually three-dimensional characteristics of brain networks at high resolution.

6.2 Insights and Future Directions

6.2.1 Continued application of developed algorithms

The algorithms in this dissertation provide a foundational set of tools for data mining of multi-dimensional EEG data. This toolset can be easily adapted and applied to future experiments from high spatial-density arrays that cover even larger cortical areas. These larger arrays are currently in fabrication courtesy of the flexible design facilitating the scalability of the multiplexed unit tiles comprising the array.

Additionally, opportunity exists to apply this algorithmic toolset to focus analysis on other sections of interest within the data, for example the period between time of injury (i.e., picrotoxin application) and the time of first seizure onset, or in other more natural chronic models of epileptogenesis. Comparison of the results from the latent period after injury to controls or to the period after the development of spontaneous seizures could lead to new understanding of mechanisms underlying epileptogenesis. Important models to human translation that could be analyzed in this way include animal models of traumatic brain injury or chemically-induced status epilepticus.

We are also interested in applying our analysis techniques to more natural animal models of idiopathic epilepsy. Naturally epileptic dogs experience seizures with semiologies and electrographic phenotypes that are strikingly similar to those observed in human patients with intractable epilepsy [56].

Our ultimate goal is translating our work to treat epilepsy patients. Employing our high-resolution arrays to cover large areas of cortex will enable us to ascertain how the spike patterns we have observed in the recordings from our acute animal model apply to spontaneous human seizures. Although human ictal events look quite similar at the macro level to the acute seizures we recorded, to further our understanding of epilepsy, we will need to pursue an investigational device exemption (IDE) from the FDA to use these novel arrays in patients.

6.2.2 Adaptation of Schiff-Wilson-Cowan computational model

As mentioned in the introduction, we believe in the ability of computational models to expand our understanding of seizures and guide the design of new therapies. They are extremely useful to test neuroscience-driven hypotheses of the mechanisms of action driving epileptic discharges. Specifically, we are interested in adapting and implementing the Wilson-Cowan model within a control framework as outlined by Schiff [57]. The Wilson-Cowan model is a general two-dimensional model of cortex with a uniform and equal distribution of excitatory and inhibitory nodes (representing populations of neurons) that are comprehensively interconnected with the strength of connection falling off exponentially with distance. The model presents three key assumptions: (1) the

resting state of the interconnected network is stable to small perturbations, (2) there is no uniformly excited state in the absence of a persistent stimulus, i.e. the seizure state is not a high probability state of an undamaged network, and (3) the inhibitory connections are longer range than the excitatory connections. The model also includes a parameter to represent the slow recovery of excitatory nodes that mimics the experimentally observed recovery of excitatory neuronal populations. The implementation of this model within a control framework provides for the introduction of closed loop feedback. Schiff's modification of the Wilson-Cowan model introduces the use of a Kalman filter observer to minimize the destabilizing effects of noise, a common characteristic of EEG recordings. Inhibition is blocked via the removal of the inhibitory nodes of the model to approximate the behavior of the network in a hyperexcitable state susceptible to seizures.

We have to begun to estimate parameters for the Wilson-Cowan equations to fit the model to our recorded experimental data. We believe this type of modeling will be essential to understand the neuronal network activity that drives the two-dimension wave propagation patterns we observed. Leveraging a control framework, in future experiments, we eventually hope to employ this model to intelligently propose effective electrical stimulation protocols to alter abnormal electrical discharges and to prevent or arrest seizures.

6.2.3 Extension of analysis to integrate simultaneous depth recordings

As noted above, we look forward to the future extension of our fundamental set of computational tools for data analysis in two-dimensions to describe the three-dimensional

characteristics of electroencephalography. We are certain this logical extension, necessary for experiments that simultaneously record from surface and penetrating arrays at high resolution, will add to our understanding of mechanisms driving abnormal cortical electrical discharges. Recent studies analyzing data from simultaneous recordings from depth and cortical surface electrodes have called into question certain ‘truths’ concerning previously accepted hallmarks of seizure activity [11, 14]. Schevon et al. [14] find evidence to suggest that iEEG may be misleading when used to precisely localize epileptogenic zone. Electroencephalography senses extracellular synaptic currents that are thought to coincide with high-levels of synchronous postsynaptic neuronal firing. Although Schevon et al. delineate instances where this is indeed the case, they also present findings where high-amplitude voltages in iEEG coincide with low-level low-synchronization of local neuronal firing. They suggest that conventional interpretation of iEEG in locating the seizure onset zone has led to the demarcation of exaggerated areas, caused by the rapid spread of postsynaptic currents far from the ‘true’ location of the presynaptic discharging ictal neurons. Cash et al. [11] report their finding of an unexpected increase in the heterogeneity of ictal neuronal spiking rate in areas of seizure propagation (as identified from the iEEG recordings). Additionally, they comment that this heterogeneity occurs when the individual neurons were observed to have a stereotyped response to seizures. Their work raises questions concerning the relative influence of local connectivity and of more distributed networks in the propagation of pathological cortical seizure activity and in the observed increased heterogeneity in neuronal firing as new cortical territory is recruited in the course of this propagation.

While these studies have raised interesting and important questions concerning the neuronal and extracellular synaptic current characteristics of seizures, we postulate that as a result of employing standard clinical subdural iEEG grids and the Utah array, these studies were limited in their ability to resolve these questions due to the sparse spatial sampling at the cortical surface and the severely constrained recording area of the Utah array (16 mm²) in deeper layers of cortex. We are currently collecting data from simultaneous recordings from high-density subdural and penetrating arrays to better address these questions. We believe that this project will help us better understand the nature of seizures and their generation in cortical networks and enable us to contribute to a resolution of the questions raised from the observations of Cash et al. and Schevon et al., which challenge long held beliefs concerning the neuronal activity underlying seizures.

6.2.4 Inform stimulation protocol in implantable devices

Among the most promising alternative treatments in the therapy of human epilepsy is the use of implantable devices designed to detect, predict and warn of approaching seizure onset [58]. We believe that reliable algorithms for accomplishing these tasks may ultimately be used to prevent or abort seizures. Although, widespread clinical adoption of these implantable devices has the potential to dramatically improve the quality of life for sufferers of medically intractable epilepsy for whom surgery is not a viable option, clinical trials of several seizure control devices have been reported in the literature with only modest outcomes [59, 60]. Optimizing their clinical efficacy requires a more complete understanding of where, when, and how seizures start. It also requires

knowledge of when and how to deliver therapy within the brain. These challenges herald room for improvement in the continued development and eventual incorporation of our data analysis algorithms. We hope that the above work, and the tools we have produced, constitutes an important step in this direction.

BIBLIOGRAPHY

1. Raol, Y.H. and A.R. Brooks-Kayal, *Experimental models of seizures and epilepsies*. Prog Mol Biol Transl Sci. **105**: p. 57-82.
2. Fisher, R.S., et al., *Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE)*. Epilepsia, 2005. **46**(4): p. 470-2.
3. Begley, C.E., et al., *The cost of epilepsy in the United States: an estimate from population-based clinical and survey data*. Epilepsia, 2000. **41**(3): p. 342-51.
4. Brodie, M.J., et al., *Patterns of treatment response in newly diagnosed epilepsy*. Neurology. **78**(20): p. 1548-54.
5. Schmidt, D. and K. Stavem, *Long-term seizure outcome of surgery versus no surgery for drug-resistant partial epilepsy: a review of controlled studies*. Epilepsia, 2009. **50**(6): p. 1301-9.
6. Bien, C.G., et al., *Assessment of the long-term effects of epilepsy surgery with three different reference groups*. Epilepsia, 2006. **47**(11): p. 1865-9.

7. Fisher, R.S., *Animal models of the epilepsies*. Brain Res Brain Res Rev, 1989. **14**(3): p. 245-78.
8. Stead, M., et al., *Microseizures and the spatiotemporal scales of human partial epilepsy*. Brain. **133**(9): p. 2789-97.
9. Engel, J. and T.A. Pedley, *Epilepsy : a comprehensive textbook*. 2nd ed. 2008, Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.
10. Kelly, R.C., et al., *Comparison of recordings from microelectrode arrays and single electrodes in the visual cortex*. J Neurosci, 2007. **27**(2): p. 261-4.
11. Truccolo, W., et al., *Single-neuron dynamics in human focal epilepsy*. Nat Neurosci, 2011. **14**(5): p. 635-41.
12. Campbell, P.K., et al., *A silicon-based, three-dimensional neural interface: manufacturing processes for an intracortical electrode array*. IEEE Trans Biomed Eng, 1991. **38**(8): p. 758-68.
13. Van Gompel, J.J., et al., *Phase I trial: safety and feasibility of intracranial electroencephalography using hybrid subdural electrodes containing macro- and microelectrode arrays*. Neurosurg Focus, 2008. **25**(3): p. E23.
14. Schevon, C.A., et al., *Evidence of an inhibitory restraint of seizure activity in humans*. Nat Commun, 2012. **3**: p. 1060.
15. Bower, M.R. and P.S. Buckmaster, *Changes in granule cell firing rates precede locally recorded spontaneous seizures by minutes in an animal model of temporal lobe epilepsy*. J Neurophysiol, 2008. **99**(5): p. 2431-42.
16. Viventi, J., et al., *Flexible, foldable, actively multiplexed, high-density electrode array for mapping brain activity in vivo*. Nat Neurosci. **14**(12): p. 1599-605.

17. Bishop, C.M., *Pattern recognition and machine learning*. Information science and statistics. 2006, New York: Springer. xx, 738 p.
18. Lopes da Silva, F., *Neural mechanisms underlying brain waves: from neural membranes to networks*. *Electroencephalogr Clin Neurophysiol*, 1991. **79**(2): p. 81-93.
19. Ebersole, J.S. and T.A. Pedley, *Current practice of clinical electroencephalography*. 3rd ed. 2003, Philadelphia: Lippincott Williams & Wilkins. xvi, 974 p., [3] p. of plates.
20. Ossadtchi, A., et al., *Automated interictal spike detection and source localization in magnetoencephalography using independent components analysis and spatio-temporal clustering*. *Clin Neurophysiol*, 2004. **115**(3): p. 508-22.
21. Haas, L.F., *Hans Berger (1873-1941), Richard Caton (1842-1926), and electroencephalography*. *J Neurol Neurosurg Psychiatry*, 2003. **74**(1): p. 9.
22. Palmini, A., *The concept of the epileptogenic zone: a modern look at Penfield and Jasper's views on the role of interictal spikes*. *Epileptic Disord*, 2006. **8 Suppl 2**: p. S10-5.
23. Worrell, G.A., et al., *High-frequency oscillations in human temporal lobe: simultaneous microwire and clinical macroelectrode recordings*. *Brain*, 2008. **131**(Pt 4): p. 928-37.
24. Blanco, J.A., et al., *Unsupervised classification of high-frequency oscillations in human neocortical epilepsy and control patients*. *J Neurophysiol*. **104**(5): p. 2900-12.
25. Allen, P.J., D.R. Fish, and S.J. Smith, *Very high-frequency rhythmic activity during SEEG suppression in frontal lobe epilepsy*. *Electroencephalogr Clin Neurophysiol*, 1992. **82**(2): p. 155-9.
26. Blanco, J.A., et al., *Data mining neocortical high-frequency oscillations in epilepsy and controls*. *Brain*. **134**(Pt 10): p. 2948-59.

27. Bragin, A., et al., *Hippocampal and entorhinal cortex high-frequency oscillations (100--500 Hz) in human epileptic brain and in kainic acid--treated rats with chronic seizures*. *Epilepsia*, 1999. **40**(2): p. 127-37.
28. Staba, R.J., et al., *Quantitative analysis of high-frequency oscillations (80-500 Hz) recorded in human epileptic hippocampus and entorhinal cortex*. *J Neurophysiol*, 2002. **88**(4): p. 1743-52.
29. Duncan, J.S., *Imaging and epilepsy*. *Brain*, 1997. **120 (Pt 2)**: p. 339-77.
30. Van Paesschen, W., et al., *The spectrum of hippocampal sclerosis: a quantitative magnetic resonance imaging study*. *Ann Neurol*, 1997. **41**(1): p. 41-51.
31. Toga, A.W., J.C. Mazziotta, and R.S.J. Frackowiak, *Brain mapping : the disorders*. 2000, San Diego, Calif. ; Academic. ix, 669 p.
32. Sakamoto, S., et al., *Interictal patterns of cerebral glucose metabolism, perfusion, and magnetic field in mesial temporal lobe epilepsy*. *Epilepsia*, 2003. **44**(9): p. 1196-206.
33. Toga, A.W. and J.C. Mazziotta, *Brain mapping : the methods*. 2nd ed. 2002, Amsterdam ; Boston: Academic Press. xvii, 877 p.
34. Howseman, A.M. and R.W. Bowtell, *Functional magnetic resonance imaging: imaging techniques and contrast mechanisms*. *Philos Trans R Soc Lond B Biol Sci*, 1999. **354**(1387): p. 1179-94.
35. Buzsaki, G. and B.O. Watson, *Brain rhythms and neural syntax: implications for efficient coding of cognitive content and neuropsychiatric disease*. *Dialogues Clin Neurosci*. **14**(4): p. 345-67.
36. Nunez, P.L. and R. Srinivasan, *Electric fields of the brain : the neurophysics of EEG*. 2nd ed. 2006, Oxford ; New York: Oxford University Press. xvi, 611 p.

37. Baumgartner, C., et al., *Magnetoencephalography in focal epilepsy*. Epilepsia, 2000. **41 Suppl 3**: p. S39-47.
38. Wheless, J.W., et al., *A comparison of magnetoencephalography, MRI, and V-EEG in patients evaluated for epilepsy surgery*. Epilepsia, 1999. **40**(7): p. 931-41.
39. Tao, J.X., et al., *Intracranial EEG substrates of scalp EEG interictal spikes*. Epilepsia, 2005. **46**(5): p. 669-76.
40. McCormick, D.A. and D. Contreras, *On the cellular and network bases of epileptic seizures*. Annu Rev Physiol, 2001. **63**: p. 815-46.
41. Janszky, J., et al., *Spatiotemporal relationship between seizure activity and interictal spikes in temporal lobe epilepsy*. Epilepsy Res, 2001. **47**(3): p. 179-88.
42. Gotman, J., *Quantitative measurements of epileptic spike morphology in the human EEG*. Electroencephalogr Clin Neurophysiol, 1980. **48**(5): p. 551-7.
43. Kandel, E.R. and J.H. Schwartz, *Principles of neural science*. 4th ed. 2000, New York: McGraw-Hill. xli, 1414 p.
44. Faure, C., *Attributed strings for recognition of epileptic transients in EEG*. Int J Biomed Comput, 1985. **16**(3-4): p. 217-29.
45. Gotman, J., *Relationships between interictal spiking and seizures: human and experimental evidence*. Can J Neurol Sci, 1991. **18**(4 Suppl): p. 573-6.
46. Jensen, M.S. and Y. Yaari, *The relationship between interictal and ictal paroxysms in an in vitro model of focal hippocampal epilepsy*. Ann Neurol, 1988. **24**(5): p. 591-8.
47. Lippmann, M.D., *Seizure quantification and clinical judgment*, in *Bioengineering*. 2013, University of Pennsylvania: ProQuest Dissertation AAI3565125.

48. Wulsin, D.F., *Bayesian Nonparametric Modeling of Epileptic Events*, in *Bioengineering*. 2013, University of Pennsylvania: ProQuest Dissertation AAI3566348.
49. Wulsin, D.F., et al., *Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement*. J Neural Eng. **8**(3): p. 036015.
50. MacQueen, J., *Some methods for classification and analysis of multivariate observations*. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, 1967. **vol. 1**: p. 281-297.
51. Prechtl, J.C., et al., *Visual stimuli induce waves of electrical activity in turtle cortex*. Proc Natl Acad Sci U S A, 1997. **94**(14): p. 7621-6.
52. Huang, X., et al., *Spiral wave dynamics in neocortex*. Neuron. **68**(5): p. 978-90.
53. Tibshirani, R., G. Walther, and T. Hastie, *Estimating the number of clusters in a data set via the gap statistic*. Journal of the Royal Statistical Society Series B-Statistical Methodology, 2001. **63**: p. 411-423.
54. Morchen, F., *Unsupervised pattern mining from symbolic temporal data*. SIGKDD Explorations, 2007. **9**(1): p. 41-55.
55. Patnaik, D., P.S. Sastry, and K.P. Unnikrishnan (2008) *Inferring Neuronal Network Connectivity from Spike Data: A Temporal Datamining Approach*. Computing Research Repository (CoRR) **arXiv: 0803.0450**.
56. Davis, K.A., et al., *A novel implanted device to wirelessly record and analyze continuous intracranial canine EEG*. Epilepsy Res. **96**(1-2): p. 116-22.
57. Schiff, S.J., *Neural Control Engineering: The Emerging Intersection between Control Theory and Neuroscience*. 2011, Cambridge, Massachusetts: The MIT Press.

58. Cook, M.J., et al., *Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study*. Lancet Neurol, 2013. **12**(6): p. 563-71.
59. Fisher, R., et al., *Electrical stimulation of the anterior nucleus of thalamus for treatment of refractory epilepsy*. Epilepsia, 2010. **51**(5): p. 899-908.
60. Morrell, M.J., *Responsive cortical stimulation for the treatment of medically intractable partial epilepsy*. Neurology, 2011. **77**(13): p. 1295-304.