Title

Spectral Characterization and Delay Differential Analysis of Human Brain Dynamics

Author

Gonzalez, Christopher Ervey

Publication Date

2021

UNIVERSITY OF CALIFORNIA SAN DIEGO

Spectral Characterization and Delay Differential Analysis of Human Brain Dynamics

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Neurosciences with a Specialization in Computational Neurosciences

by

Christopher Ervey Gonzalez

Committee in Charge:

Eric Halgren, Chair
Terrence Sejnowski, Co-Chair
Maxim Bazhenov
Claudia Lainscsek
John Serences
Tatyana Sharpee

2021

The dissertation of Christopher Gonzalez is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

This dissertation is dedicated to:

My parents, James, and Will for providing endless support.

EPIGRAPH

Suddenly he closed his eyes, raised his hand with an effort, and threw away the heavy end of his cigar. A great brush swept smooth across his mind, sweeping across it moving branches, children’s voices, the shuffle of feet, and people passing, and humming traffic, rising and falling traffic. Down, down he sank into the plumes and feathers of sleep, sank, and was muffled over.

- Virginia Woolf

Changing, it rests.

- Heraclitus

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

NREM: non-REM sleep

DS: Downstate

US: Upstate

TB: Theta burst

SS: Sleep spindle

EEG: Electroencephalography

SEEG: stereoelectroencephalography

MEG: Magnetoencephalography

MRI: Magnetic Resonance Imaging

CT: Computerized Tomography

N2/N3: Stage 2/Stage 3 sleep

PAC: Phase-amplitude coupling

PLV: Phase-locking value

HG: High gamma

LFP: Local field potential

PSD: Power spectral density

KC: K-complex

HC: Hippocampus

NC: Neocortex

DDA: Delay differential analysis

CD-DDA: Cross-dynamical delay differential analysis

GC: Granger Causality

CCM: Convergent cross-mapping

HR: Hindmarsh-Rose

SOC: Symbolic observability coefficient

RC: Reservoir computing

SVDO: Singular value decomposition observability

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Eric Halgren for his support as the chair of my committee, as well his close mentorship, clear questions, and kind leadership. I would also like to thank Terrence Sejnowski for his leadership and providing a scientifically diverse environment with creative and deep thinkers. I would like to thank Claudia Lainscsek for her invaluable mentorship in dynamical and chaotic systems, which expanded my outlook on science and the world. I would also like to thank the rest of my committee, Maxim Bazhenov, Tatyana Sharpee, and John Serences, for their astute and helpful feedback over the years.

I'd like to thank members of both of the labs I've had the honor of working in as well as collaborators, for the stimulating scientific discussions, help tackling problems, and for providing a warm community to work in: Burke Rosen, Charlie Dickey, Xi Jiang, Jacob Garrett, Sophie Kajfez, Ilya Verzhbinsky, Rachel-Mak McCully, Eric Kastner, Anna Sargsyan, Aaron Sampson, Nuttida Rungratsameetaweemana, Robert Kim, Sarah Schoch, Yusi Chen, Mariam Ordyan, Ben Tsuda, Lyle Muller, Margot Wagner, Oscar Gonzalez, and Ryan Golden.

I also want to thank Steph Nelli for her friendship and support, and for our scientific/mathematical/philosophical conversations. In addition to Steph, I want to recognize Katie Fife and Jessica Bui for giving myself and Will a home in San Diego. I am also thankful for the friends I've made in the Neuroscience Graduate Program at UCSD, especially the support of Javier How, Burke Rosen, and Rachel Cassidy. Last, I am so grateful to my partner Will for moving with me from Austin to Baltimore to San Diego, for loving and supporting me every step of the way, and for reminding me not to take work *too* seriously.

I would also like to thank the National Science Foundation Graduate Research Fellowship, the National Institute of Mental Health Predoctoral Training Grant in Cognitive Neuroscience, and the San Diego Matching Fellowship for support.

Chapter 1, in full, is a reprint of the material as it appears in The Journal of Neuroscience, titled “Theta Bursts Precede, and Spindles Follow, Cortical and Thalamic Downstates in Human NREM Sleep”, November 2018.; 38(46). Gonzalez, Christopher; Mak-McCullly, Rachel; Rosen, Burke; Cash, Sydney; Chauvel, Patrick; Bastuji, Hélène; Rey, Marc; Halgren, Eric. The dissertation author was one of the primary investigators and authors of this paper.

Chapter 2, in full, is currently being prepared for submission for publication. Co-authors include: Gonzalez, Christopher; Jiang, Xi; Rosen, Burke; Gonzalez-Martinez, Jorge; Halgren, Eric. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in Chaos, titled “Causality detection in cortical seizure dynamics using cross-dynamical delay differential analysis”,2019; 29. Lainscsek, Claudia; Gonzalez, Christopher; Sampson, A; Cash, Sydney; Sejnowski, Terrence. The dissertation author was one of the primary investigators and authors of this paper.

Chapter 4, in full, is a reprint of the material as it appears in Chaos, titled “Assessing observability of chaotic systems using Delay Differential Analysis”,2020; 30. Gonzalez, Christopher; Lainscsek, Claudia; Sejnowski, Terrence; Letellier, Christopher. The dissertation author was one of the primary investigators and authors of this paper.

VITA

2013 Bachelor of Arts, Philosophy, The University of Texas at Austin

2013 Bachelor of Science, Biology (Neurobiology Spec.), The University of Texas at Austin

2021 Doctor of Philosophy, Neurosciences, University of California San Diego

PUBLICATIONS

Gonzalez, C.E.*, Lainscsek, C*, Sejnowski, T.J., Letellier, C. (2020). Assessing observability of chaotic systems using Delay Differential Analysis. *Chaos*. *indicates co-first author

Lainscsek, C.*, **Gonzalez, C.E.***, Sampson, A., Cash, S., Sejnowski, T.J.(2019). Causality detection in cortical seizure dynamics using cross-dynamical delay differential analysis. *Chaos.* *indicates co-first author

Sampson, A.L., Lainscsek C., **Gonzalez, C.E**., Ulbert, I., Devinsky, O., Fabo, D., Halgren E., Cash, S.S.,Sejnowski, T.J (2019). Delay differential analysis for dynamical sleep spindle detection. *Journal of Neuroscience Methods.*

Gonzalez, C.E.*, Mak-McCully, R.A.*, Rosen, B.Q., Cash, S.S., Chauvel, P., Bastuji, H., Rey, M., Halgren, E (2018). Theta bursts precede, and spindles follow, cortical and thalamic downstates in human NREM sleep. *Journal of Neuroscience.* *indicates co-first author.

Mak-McCully, R.A., Rolland, M., Sargsyan, A., **Gonzalez, C.E**., Magnin, M., Chauvel, P., Rey, M.,Bastuji,H., Halgren, E (2017). Coordination of cortical and thalamic activity during non-REM sleep in humans. *Nature Communications.*

Schnyer, D., Clasen, P., **Gonzalez, C.E.**, Beevers, C (2017). Evaluating the diagnostic utility of applying a machine learning algorithm to diffusion tensor MRI measures in individuals with major depressive disorder. *Psychiatry Res Neuroimaging.*

Wolf, C., An, Y., Tanaka, T., Bilgel, M., **Gonzalez, C.E.**, Kitner Triolo, M., Resnick, S.M. (2017). Cross-sectional and longitudinal effects of CREB1 genotypes on individual differences in memory and executive function: findings from the BLSA. *Frontiers in Aging Neuroscience.*

Chuang, Y., Elango, P., **Gonzalez, C.E.**, Thambisetty, M (2017). Midlife anticholinergic drug use, risk of Alzheimer's disease, and brain atrophy in comunity-dwelling older adults. *Alzheimer's and Dementia:Translational Research and Clinical Interventions.*

Gonzalez, C.E., Venkatraman, V.K., An Y., Landman, B.A., Ratnam Bandarua, V.V., Haughey N.J., Ferrucci, L., Mielke, M.M., Resnick, S.M (2016). Peripheral sphingolipids predict age-related differences in white matter microstructure. *Neurobiology of Aging.*

Spira, A.P., **Gonzalez, C.E.,** Venkatraman, V.K., Wu, M.N., Pacheco, J., Simonsick, E.M., Ferrucci, L., Resnick, S.M. (2016). Sleep duration and subsequent cortical thinning in community-dwelling older adults. *SLEEP*.

Tian, Q., Resnick, S.M., Landman, B.A., Huo, Y., Venkatraman, V.K., **Gonzalez, C.E.**, Simonsick, E.M., Shardell, M.D., Ferrucci, L., Studenski, S.A (2016). Lower gray matter integrity is associated with greater lap time variation in high-functioning older adults. *Experimental Gerontology.*

Gonzalez, C.E., Pacheco, J., Beason-Held, L.L., Resnick, S.M (2015). Longitudinal changes in cortical thinning associated with hypertension. *Journal of Hypertension*.

Venkatraman, V.K., **Gonzalez, C.E.**, Landman, B.A., An., Y., Reiter, D., Goh, J., Resnick, S.M (2015). Region of interest correction factors improve reliability of diffusion imaging measures within and across scanners and field strengths. *Neuroimage*.

Tian, Q., Ferrucci, L., Resnick, S., Simonsick, E.M., Shardell, M., Landman, B., Venkatraman, V.K., **Gonzalez, C.E.**, Studenski, S (2015). The effect of age and microstructural white matter integrity on lap time variation and fast-paced walking speed. *Brain Imaging and Behavior.*

Plassard, A.J., Hinton, K.E., Venkatraman V., **Gonzalez C.E.**, Resnick, S.M., Landman, B.A (2015). Evaluation of atlas-based white matter segmentation with Eve. *Proc SPIE Int Soc Opt Eng.*

FIELDS OF STUDY

Major Field: Neuroscience

Computational Neuroscience Specialization
Professors Eric Halgren and Terrence Sejnowski

ABSTRACT OF THE DISSERTATION

Spectral Characterization and Delay Differential Analysis of Human Brain Dynamics

by

Christopher Ervey Gonzalez

Doctor of Philosophy in Neurosciences with a Specialization in Computational Neurosciences

University of California San Diego, 2021

Professor Eric Halgren, Chair

The brain endogenously generates electrical activity that arises from the complex, nonlinear interactions of its components. During sleep, large amplitude, slow oscillations as well as 10-16 Hz rhythms known as sleep spindles are generated in the cortex and thalamus respectively, and their coupling has been shown to bolster our memory capacities by facilitating

cortical plasticity. Identifying where particular sleep rhythms are generated, how they co-occur with other regions, and whether rhythms differ in frequency or other characteristics can inform mechanisms for how they coordinate information exchange during sleep. In this dissertation, I characterized the cortical and thalamic activity of sleep spindles, theta bursts (~6 Hz), a novel sleep rhythm identified here, and the coupling of spindles and theta bursts with slow waves using intracranial recordings from epileptic patients. I also report regional differences in spindle properties, largely inaccessible to non-invasive recordings, that propose a modified view of spindle dynamics across the cortex.

The most common characterizations of brain dynamics, including the sleep rhythms reported here, are largely based on linear time-frequency analyses. However, because the brain is a high-dimensional, nonlinear system, applying linear techniques alone may not sufficiently capture the relevant dynamical features. To address this, I helped develop nonlinear tools, based on Delay Differential Analysis (DDA), for analyzing neural time series. I evaluated these tools in simulated, chaotic systems, which are suitable models for recurrent, continuous, and nonlinear dynamics. Specifically, I investigated whether given a set of recorded time series, can we (1) assess whether two signals are causally interacting and (2) rank signals by their amount of dynamical information about the original system. These applications of DDA, in tandem with traditional linear techniques, can improve our understanding of underlying brain activity during seizures and sleep.

In Chapter 1, I characterize a novel sleep rhythm, the theta burst, that is distinct from sleep spindles, recorded in both the cortex and thalamus, and which in both structures, precedes downstates. In Chapter 2, I report distinct sources of spindle variability, including variability across channels within a region, across spindles within a single recording site, and across cycles within a spindle, and how these sources are of a size comparable to the frontal-parietal difference typically used to summarize spindle dynamics. Chapter 3 introduces a novel nonlinear signal processing technique, Cross-Dynamical Delay Differential Analysis (CD-DDA), for inferring causal interactions between time series and applies this approach to track seizure spread in a patient with epilepsy. Chapter 4 applies DDA in simulated chaotic dynamical systems to assess the observability of a time series, i.e. how much dynamical information a variable has about the original system it is a part of.

CHAPTER 1: Theta bursts are distinct sleep rhythms in NREM sleep

Systems/Circuits

Theta Bursts Precede, and Spindles Follow, Cortical and Thalamic Downstates in Human NREM Sleep

Christopher E. Gonzalez,[1]* Rachel A. Mak-McCully,[2]* Burke Q. Rosen,[1] Sydney S. Cash,[3] Patrick Y. Chauvel,[4] Hélène Bastuji,[5] Marc Rey,[4] and Eric Halgren[6]

[1]Department of Neurosciences, University of California San Diego, La Jolla, California 92093, [2]University California Berkeley, Berkeley, California 94720, [3]Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Harvard University, Boston, Massachusetts 02114, [4]Aix-Marseille Université, Marseille 13385, France, [5]Central Integration of Pain, Lyon Neuroscience Research Center, INSERM, U1028, CNRS, UMR5292, Université Claude Bernard, Lyon, Bron, France, and [6]Departments of Radiology and Neurosciences, University of California, San Diego, California 92093

Since their discovery, slow oscillations have been observed to group spindles during non-REM sleep. Previous studies assert that the slow-oscillation downstate (DS) is preceded by slow spindles (10–12 Hz) and followed by fast spindles (12–16 Hz). Here, using both direct transcortical recordings in patients with intractable epilepsy ($n = 10$, 8 female), as well as scalp EEG recordings from a healthy cohort ($n = 3$, 1 female), we find in multiple cortical areas that both slow and fast spindles follow the DS. Although discrete oscillations do precede DSs, they are theta bursts (TBs) centered at 5–8 Hz. TBs were more pronounced for DSs in NREM stage 2 (N2) sleep compared with N3. TB with similar properties occur in the thalamus, but unlike spindles they have no clear temporal relationship with cortical TB. These differences in corticothalamic dynamics, as well as differences between spindles and theta in coupling high-frequency content, are consistent with NREM theta having separate generative mechanisms from spindles. The final inhibitory cycle of the TB coincides with the DS peak, suggesting that in N2, TB may help trigger the DS. Since the transition to N1 is marked by the appearance of theta, and the transition to N2 by the appearance of DS and thus spindles, a role of TB in triggering DS could help explain the sequence of electrophysiological events characterizing sleep. Finally, the coordinated appearance of spindles and DSs are implicated in memory consolidation processes, and the current findings redefine their temporal coupling with theta during NREM sleep.

Key words: corticothalamic; downstates; iEEG; sleep; spindles; theta

Significance Statement

Sleep is characterized by large slow waves which modulate brain activity. Prominent among these are downstates (DSs), periods of a few tenths of a second when most cells stop firing, and spindles, oscillations at ∼12 times a second lasting for ∼a second. In this study, we provide the first detailed description of another kind of sleep wave: theta bursts (TBs), a brief oscillation at ∼six cycles per second. We show, recording during natural sleep directly from the human cortex and thalamus, as well as on the scalp, that TBs precede, and spindles follow DSs. TBs may help trigger DSs in some circumstances, and could organize cortical and thalamic activity so that memories can be consolidated during sleep.

Introduction

During NREM sleep, the brain endogenously produces electrical activity dominated by larger amplitude, lower frequency (0.1- 16 Hz) rhythms compared with wake or REM states. Downstates (DSs) (periods of neuronal quiescence lasting a few hundred milliseconds) and sleep spindles (∼0.5–2 s, 10–16 Hz oscillations) are two canonical NREM events with initiating mechanisms largely attributed to cortical and thalamic activities, respectively. However, the extensive bidirectional connections between cortex and thalamus precludes a simple entraining mechanism for ei-

Received Feb. 20, 2018; revised Aug. 10, 2018; accepted Aug. 28, 2018.

Author contributions: C.E.G. and R.A.M.-M. wrote the first draft of the paper; C.E.G., R.A.M.-M., S.S.C., P.Y.C., H.B., M.R., and E.H. edited the paper; C.E.G., R.A.M.-M., and E.H. designed research; C.E.G., R.A.M.-M., B.Q.R., P.Y.C., H.B., and M.R. performed research; S.S.C., P.Y.C., H.B., and M.R. contributed unpublished reagents/analytic tools; C.E.G., R.A.M.-M., and B.Q.R. analyzed data; C.E.G. and R.A.M.-M. wrote the paper.

This work was supported by the National Institutes of Health (Grants R01-MH-099645 and R01-EB-009282), the U.S. Office of Naval Research (Grant N00014-13-1-0672), the National Science Foundation Graduate Research Fellowships Program, and the National Institute of Mental Health–NIH (T32 Cognitive Neuroscience Training Grant). We thank Nima Dehghani for EEG data, Donald Hagler for spindle detection scripts, Fabrice Bartolomei for access to data and analysis input, Catherine Liegeois-Chauvel for research access, and Jean Regis for electrode localization for the Marseille patient.

The authors declare no competing financial interests.

*C.E.G. and R.A.M.-M. contributed equally to this work.

Correspondence should be addressed to Christopher E. Gonzalez, Department of Neurosciences Graduate Program, University of California, 9500 Gilman Drive 0634, La Jolla, CA 92093. E-mail: cegonzalez@ucsd.edu.

https://doi.org/10.1523/JNEUROSCI.0476-18.2018

Table 1. Demographic and clinical information

Patient	Sex	Age	Handedness	Clinical diagnosis	Pathological diagnosis	Imaging	Focus
1	M	35	—	Right temporal lobe epilepsy	No pathology obtained	Normal	Temporo-parieto-occipital junction
2	F	37	R	Temporal lobe epilepsy	No pathology obtained	Normal	Hippocampus
3	M	45	L	CPS; bitemporal	No pathology obtained	—	Left and right mesial temporal lobes
4	F	45	R	CPS; multifocal	Multifocal: temporal, parietal, occipital	—	Temporal
5	F	65	R	CPS; temporal lobe epilepsy with two foci: left mesial temporal structures and right subfrontal region	No pathology obtained	—	Right subfrontal and anterior temporal
6	F	32	R	Left temporal lobe epilepsy	Nonspecific gliosis	Normal	Hippocampus, entorhinal cortex, amygdala
7	F	23	R	Right temporal lobe epilepsy	Type I focal dysplasia	Normal	Hippocampus, entorhinal cortex, anterior insula
8	F	28	R	Right temporal occipital epilepsy	Type II focal dysplasia	Focal cortical dysplasia in right fusiform	Fusiform gyrus, entorhinal cortex
9	F	50	R	Right temporal occipital epilepsy	No pathology obtained	Normal	Right fusiform gyrus
10	F	27	R	Right temporal lobe epilepsy	No pathology obtained	Right hippocampal sclerosis	Hippocampus, entorhinal cortex, amygdala

CPS, Complex partial seizures.

ther event and properties such as duration (Bazhenov et al., 2002; Bonjean et al., 2011; Barthó et al., 2014), frequency of occurrence (Timofeev et al., 2000), and synchrony (Contreras et al., 1996) are all shaped by their cooperative dynamics (Steriade, 1997; Crunelli and Hughes, 2010). These rhythms are believed to serve functional roles in sleep-dependent memory consolidation (Sejnowski and Destexhe, 2000; Diekelmann and Born, 2010; Hanert et al., 2017). In particular, the specific grouping of spindles by slow waves has been associated with improved declarative memory in humans (Mölle et al., 2011; Niknazar et al., 2015) and fear conditioning in mice (Latchoumane et al., 2017).

Although several studies report faster spindle frequency activity (>12 Hz) occurs on the transition from the down to upstate (US) (Andrillon et al., 2011; Mölle et al., 2011; Cox et al., 2014a; Klinzing et al., 2016), the temporal relation of lower frequency (4–12 Hz) content to DSs has not been definitively established. This relationship is more clear during NREM stage 2 (N2) sleep, when DSs usually occur without preceding USs as the main component of the K-complex (KC) (Cash et al., 2009; Mak-McCully et al., 2015). Some investigators have observed "polyphasic waves [. . .]just before the onset of the negative K-complex sharp wave" (Rodenbeck et al., 2006) or an ~7 Hz short-lasting "intra-KC" oscillation (Kokkinos and Kostopoulos, 2011; Kokkinos et al., 2013). More recently, however, the US to DS transition has been associated with slow spindle activity (9–12 Hz) (Mölle et al., 2011; Klinzing et al., 2016; Yordanova et al., 2017). Findings from the latter studies bolster the hypothesis that there are distinct types of spindles, slow and fast, which could have their own rhythmogenesis mechanisms (Timofeev and Chauvette, 2013; Fogerson and Huguenard, 2016). Here, we present findings that challenge this hypothesis using bipolar transcortical and thalamic recordings from epileptic patients to obtain focal measures of DSs and spindles. We confirm these findings at the scalp using EEG recordings from a nonclinical population. We propose that, regardless of frequency, spindles recorded either intracranially or at the scalp are more likely to start on the down to up transition and we describe a theta burst (TB) distinct from spindles that can accompany the up to down transition, especially in N2 sleep.

Materials and Methods

Intracranial recordings. Stereoencephalography (SEEG) was obtained in 10 patients (8 female; mean ± SD age: 38.7 ± 12.7; Table 1) undergoing evaluation for pharmacoresistant epilepsy at Massachusetts General Hospital; La Timone Hospital, Marseille, France; or Neurological Hospital, Lyon, France. At Massachusetts General Hospital, electrode contacts were localized using CT of the implanted electrodes superimposed on preoperative MRI (Dykstra et al., 2012). Each SEEG electrode had either 8 (5 mm center-to-center spacing) or 6 (8 mm spacing) contacts. Each contact was 1.28 mm in diameter and 2.4 mm long. Signals were sampled at 500 Hz and band-pass filtered from 0.33 to 128 Hz. At La Timone Hospital, localization of electrode contacts was performed using MRI and CT of implanted electrodes. For one patient, localization was determined using preoperative MRI and surgical planning. Each electrode had either 10 or 15 contacts (3.5 mm center-to-center spacing). Each contact was 0.8 mm in diameter and 2 mm long. The recordings were sampled at 256, 512, or 1024 Hz. At the Neurological Hospital, electrode localization was determined directly from stereotactic teleradiographs without parallax performed within the stereotactic frame (Talairach and Tournoux, 1998). These locations were superimposed onto the preimplantation 3T structural MRI (3D MPRAGE T1 sequence) after alignment with the skull. The locations of cortical and thalamic contacts were determined by reference to the atlases of Duvernoy (1999) and Morel et al. (1997). Each electrode had either 10 or 15 contacts (3.5 mm center-to-center spacing). Each contact was 0.8 mm in diameter and 2 mm long. The recordings were sampled at 256 Hz and bandpass filtered from 0.33 to 128 Hz. Informed consent was obtained from all patients.

SEEG recordings were bipolar between adjacent contacts spanning the cortical ribbon. These "transcortical bipolar contacts" provide relatively focal measurements of the local field potentials generated in the transected cortex (Mak-McCully et al., 2015). The polarity of bipolar derivations was inverted if necessary to ensure that DS [as confirmed with decreased high gamma (HG); see below] were negative. This produced recordings with a relatively consistent relationship to the underlying cortical generators, as evidenced by the consistent phase relation between HG and TB (see below) and chosen as described previously (Mak-McCully et al., 2015). Only channels with both slow oscillations and spindles apparent to visual inspection were included for analysis. Channels were excluded if a clinical electroencephalographer judged significant interictal activity, pathological background changes, or early involvement in the ictal discharge. At each site, sleep scoring was performed by clinical experts (S.S.C., H.B., or M.R.) using examination of

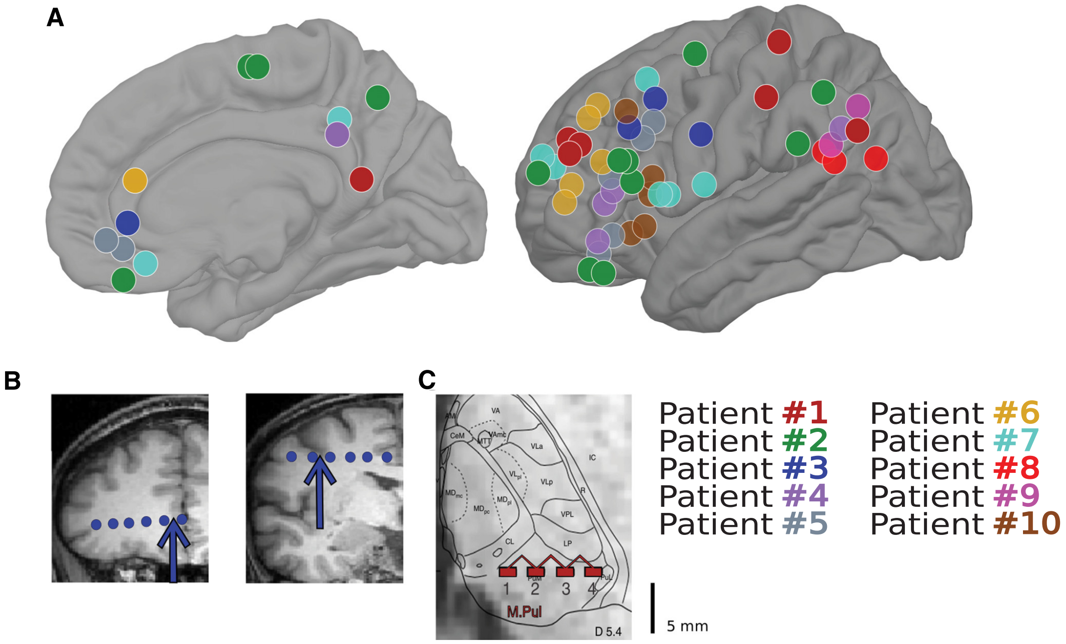


Figure 1. Bipolar SEEG recordings obtained from epileptic patients. ***A***, Locations of 60 cortical bipolar recordings from 10 patients with intracranial electrodes. ***B***, Illustration of bipolar transcortical derivation used throughout the study, with arrows indicating two cortical bipolar channels from Patient 3. ***C***, Illustration of bipolar recordings from Patient 1 from the pulvinar of the thalamus.

bipolar electrodes ($n = 7$) or scalp EEG with EOG/EMG when available ($n = 3$). Analyses were performed on the N2 and N3 sleep periods only (Silber et al., 2007).

In total, 60 cortical bipolar SEEG channels obtained from 10 patients and eight thalamic (mainly pulvinar) bipolar channels from three of the 10 patients during NREM sleep were included for analysis. The locations of all cortical bipolar channels are shown in Figure 1*A*, as well as examples of bipolar derivations for cortical (Fig. 1*B*) and thalamic (Fig. 1*C*) sites. On average, each patient had six cortical channels (ranging from two to 13) and 120.6 min of NREM sleep from one night (ranging from 26.5 to 240 min). DSs and spindles were detected on each channel separately using previously described methods (and see below) (Mak-McCully et al., 2017).

Scalp EEG recordings. Scalp EEG data were recorded during sleep from three healthy subjects (one female). Written informed consent approved by the Partners Healthcare Network was obtained for all subjects before their participation. Subjects wore a 70-channel EEG cap with a modified 10–20 montage (Elekta Neuromag). Data were referenced to left mastoid. Magnetoencephalographic data were collected simultaneously but are not reported here. Periods of N2 sleep were identified according to standard criteria (Iber et al., 2007). Gross artifacts were removed by visual inspection.

DS detection. DSs were detected on each channel as follows: (1) apply a zero-phase eighth-order Butterworth filter from 0.1 to 4 Hz; (2) select consecutive zero crossings within 0.25- 3 s; and (3) calculate amplitude peak between zero crossings and retain only the bottom 20% of peaks for intracranial recordings or the bottom 10% of peaks for scalp EEG recordings.

For DS detection in intracranial recordings, only periods of N2 and N3 sleep free of visually identified epileptiform discharges were used. Bipolar SEEG channels exhibiting DSs were also required to show decreases in power within HG, defined as 60–100 Hz, exceeding 1 dB within ±250 ms of the negative DS peak. Sixty such channels were identified, showing mean decreases in HG power during DS troughs of −3.18 dB (range, −1 to −8).

Spindle detection. The current clinical standard for sleep scoring adopted by the American Academy of Sleep Medicine is 11–16 Hz (Silber et al., 2007), but the major previous studies describing sleep spindles using intracranial recordings in humans adopted 9–16 Hz (Andrillon et al., 2011; Piantoni et al., 2017) or 10–16 Hz (Mak-McCully et al., 2017; Hagler et al., 2018). Recent analyses with scalp recordings of fast versus slow spindles have also used either 9–15 Hz (Mölle et al., 2011; Klinzing et al., 2016) or 10–16 Hz (Cox et al., 2014a) or even 9–16 Hz(Yordanova et al., 2017), but in the majority of cases, with the division between fast and slow spindles at 12 Hz. Here, we consider spindles as 10–16 Hz events and define spindles ≤12 Hz as slow. Recordings were notch filtered (either 49–51 Hz or 59–61 Hz, depending on country of origin) and then band-pass filtered at 10–16 Hz using a zero-phase frequency domain filter (transition bands 30% of cutoff frequency). Taking the absolute value of this filtered signal produced a spindle-band amplitude envelope. This envelope was convolved with a 400 ms Tukey window and the median amplitude was subtracted and normalized by the median absolute deviation. This signal was used to detect the onset and offset of putative spindle epochs. To detect the middle of spindle epochs, we convolved the amplitude signal with a 600 ms Tukey window, normalized as before, and identified peaks with magnitude larger than 2 for intracranial recordings and larger than 1 for scalp EEG recordings. Then, we defined the onset and offset as 40% of the peak amplitude of the original spindle amplitude envelope. Any overlapping or duplicate epochs were resolved and epochs <300 ms were excluded. We then applied a series of strict exclusion criteria for putative spindle epochs. These included removing any epochs that also exceeded 5 for a low (4–8 Hz)- or high (18–25 Hz)-amplitude envelope. We also required 5 peaks in a broad-band filtered signal (4–25 Hz) with an amplitude greater than the median absolute deviation per channel and at least 25% amplitude of the largest peak. Spindles in the thalamus were detected using a modified version of this detector as described previously (Mak-McCully et al., 2017).

TB detection. TBs were detected by modifying a previously reported spindle detector (Andrillon et al., 2011). Our procedure was as follows:

(1) apply a zero-phase eighth-order Butterworth filter from 5 to 8 Hz (range selected to minimize overlap with delta and spindle content); (2) calculate the mean of the Hilbert envelope of this signal smoothed with a Gaussian kernel (300 ms window; 40 ms σ); (3) detect events with a ±3 SD threshold for the peak and identify the start and stop times with a ±1 SD threshold; (4) only include events with a duration between 400 ms and 1 s; and (5) for each band-pass-filtered peak in putative burst, calculate the preceding trough-to-peak deflection and only take events that have at least 3 peaks exceeding 25% of the maximum deflection. We also required bipolar SEEG recordings to have at least 50 TBs before calculating average theta frequency or the proportion of theta events associated with DSs, which excluded four of 60 cortical channels.

Phase–amplitude coupling (PAC). For the intracranial recordings, we correlated the phase of either spindle (10–16 Hz) or theta (5–8 Hz) with the analytic amplitude in 60–100 Hz for all cortical channels across patients. We chose 60–100 Hz to keep the range consistent across patients and because our lowest sampling frequency was 256 Hz. Only channels that had at least 30 theta events and 30 spindles were included (58 of 60 cortical channels). We band-pass filtered our data in the theta-, spindle-, and gamma-frequency ranges using finite impulse response filters with an order equal to the duration of three cycles of the lowest frequency. We used the Hilbert transform to extract the analytic signal from our band-passed data, took the phase angle from our theta and spindle band-passed data, and the amplitude from our HG band-pass signal. Correlations between the phase of the lower frequency signal and HG amplitude were evaluated across a duration equal to the first two cycles of the lower frequency range per signal across all detected events for each channel. The observed PAC measure was calculated by taking the length of the average complex vector of the low-frequency phase, weighted by the corresponding high-frequency power time series (Canolty et al., 2006). Significance was assessed for each channel and event type using nonparametric permutation statistics. Specifically, for each channel and event type, the phase time series was randomly offset relative to the power time series and PAC was recalculated 1000 times, generating a null distribution against which to compare our observed PAC measure. The preferred phase was determined only for channels with a PAC-Z value >3 ($p < 0.002$).

Time domain and spectral analyses. We created event-related histograms to quantify the timing of intracranially recorded spindle or theta events relative to all intracranially recorded DS troughs. For each bipolar SEEG recording, we required at least 30 events associated with a DS within ±1 s of the DS trough to be included in the grand average histograms. Furthermore, we required at least 20 events within ± 500 ms to assess whether spindle (theta) events were more likely to start 500 ms after (before) the DS trough using a two-way binomial test. Binomial tests were corrected for multiple comparisons using Bonferroni correction ($\alpha = 0.05$) for each grapho-element type. Data processing and analysis were performed in MATLAB and time–frequency plots were created using EEGLAB (Delorme and Makeig, 2004).

We also examined the spectral profile of periods before DSs. Power-spectral densities (PSDs) were computed for the epoch 500 to 0 ms before the DS trough. For each channel, these were ranked and quartiled by the power in the 4–12 Hz band. PSDs in the first and fourth quartile were normalized as *z*-scores and averaged over DSs. The difference between the quartiles is plotted in Figures 2*G* (SEEG) and 5*E* (scalp EEG).

Experimental design and statistical analysis. Linear mixed-effects models with patient specified as random effect were implemented in R to estimate descriptive statistics such as overall frequency, duration, and rate of occurrence, as well as for testing differences between frontal and parietooccipital electrodes.

Results

Identifying short TBs before DSs in the cortex

Visual inspection of average spectrograms ± 2 s relative to all DS troughs in intracranial recordings showed increases within the 5–10 Hz range ~250 ms before the negative peak (Fig. 2*A*,*D*, cyan arrows). However, inferring oscillatory activity from such representations can be misleading (Jones, 2016) because the sharp decline in the pretrough part of the DS could contain nonoscillatory power in the theta band (Cox et al., 2014b). To confirm that this increase was associated with an oscillatory component and was not just an artifact of the ensuing DS LFP waveform, we looked for the presence of oscillations in the average of the original LFP time locked to a filtered theta peak (5–8 Hz). We implemented this by sorting DSs according to previous theta power and averaging across events within a quartile time locked to the first band-passed theta peak preceding the DS trough. This unfiltered average revealed clear oscillatory activity within the theta range, showing two to three peaks across the majority of channels for the top quartile and often absent in the lowest (Fig. 2*B*,*E*). Theta oscillations were also apparent at the level of single DSs selected from the top quartile (Fig. 2*B*,*E*, insets, cyan arrows).

To further characterize this observed intracranially recorded theta oscillation during NREM sleep, we applied a theta detector to each channel (see Materials and Methods). Descriptive statistics for detected theta events are shown per patient in Table 2 and the overall estimates for frequency, duration, and rate of occurrence are as follows (mean ± SD): 6.33 ± 0.45 Hz, 672 ± 28 ms, 1.25 ± 0.39 /min. Interestingly, frontal channels exhibited lower overall frequency compared with parietooccipital channels, with estimated frequencies of 6.26 and 6.5, respectively ($t = -5.6$, $p =$ 2.14e-08, mixed-effects model with patient as random effect). As expected, frontal channels also showed lower overall spindle frequency compared with parietooccipital ($t = -3.14$, $p = 0.002$), with estimated overall frequencies of 12.26 and 12.65 Hz. To further verify that the pre-DS oscillations were the result of theta and not slower spindles, for each channel, we sorted DSs by power in 4–12 Hz 500 ms before the DS trough. This range includes both theta and slow spindle frequencies and thus would detect either. We then calculated the average PSD across all DSs, per quartile. The difference of the top and bottom quartiles reveals a center of frequency at 6 Hz across channels (2G), indicating that the pre-DS oscillations are theta rather than slow spindles.

In Figure 2, *C* and *F*, we superimpose all detected theta traces, unfiltered and locked to the deepest trough in the theta event for two example bipolar SEEG channels. Some channels exhibited a downward trend in the average (Fig. 2*F*), suggesting that DS tends to follow the deepest trough, whereas others do not (Fig. 2*C*). This difference illustrates that, whereas some detected theta events were associated with DSs, others were not, and this varied within and between channels. We also observed from *post hoc* analyses that some channels exhibited larger amplitude and more prolonged theta oscillations in the raw LFP for N2 versus N3 DSs (Fig. 3*A*). This was corroborated with a greater number of peaks in detected TBs on average per channel for N2 compared with N3 ($t = -4.33$, $p =$ 1.49e-05, mixed-effects model with patient as random effect and channel as nested random effect). Additionally, we found that there was significantly greater theta power prior (−500 ms to 0) to the DS trough for N2 compared with N3 DSs (Fig. 3*B*; $t = -3.13$, $p = 0.002$, mixed-effects model). However, the rate of detected theta event occurrence was not different between N2 and N3 ($t = -1.2$, $p = 0.23$). That is, although the number of events did not differ between N2 and N3, TBs were longer and larger in N2.

Theta and spindles show different temporal relationships with DSs

Because slow oscillations during sleep are known to group higher-frequency rhythms, we investigated how often our de-

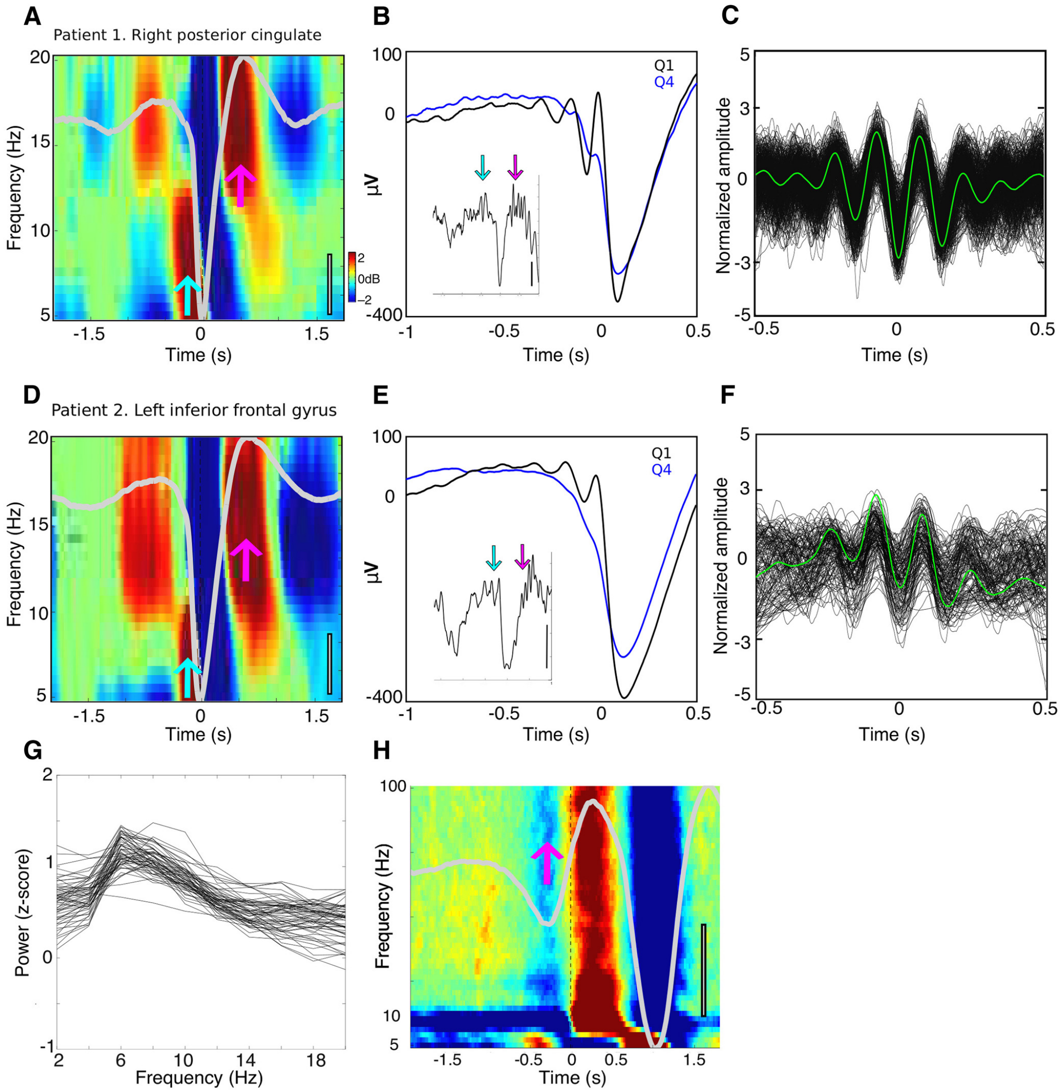


Figure 2. Cortical theta during NREM sleep. ***A***–***C***, Patient 1, right posterior cingulate. ***D***–***F***, Patient 2, left inferior frontal gyrus. For both ***A*** ($n = 2084$) and ***D*** ($n = 1265$), time–frequency plots locked to DS troughs reveal increases in the theta range (cyan arrows) immediately before the trough and centered within spindle range (pink arrows) from 0.5 to 1 s after the trough (and more weakly from −1 to −0.5 s before the trough due to a preceding DS (see ***H***)). The entire epoch was used as baseline. The average of all DSs is overlaid in light gray, gray scale bars indicate 100 μV. ***B***, ***E***, Average, unfiltered LFP time-locked to the first filtered positive theta peak before the DS trough. Black indicates the top quartile ($n = 521$ for ***B***, $n = 316$ for ***E***), blue the bottom quartile of DSs with prior theta power. Insets reveal a single trace from the top quartile. Scale bars, 200 μV for inset, time ranges from −2 to +0.5 s of the DS trough. Cyan and pink arrows indicate theta events and spindles, respectively. Positive potentials indicate cortical surface positivity. ***C***, ***F***, Raw LFP of detected theta events locked to the deepest trough, *z*-normalized. Average shown in green for ***C*** ($n = 509$) and ***F*** ($n = 148$). ***G***, Pre-DS spectra. Plotted are the differences of the first and fourth quartile average power spectral densities of 500 ms pre-DS, when ranked by the amount of 4–12 Hz power they contain. Each trace represents the difference of average PSDs for one channel in one patient; all 60 cortical channels are shown. In nearly all channels, power peaks at 6 Hz. ***H***, Average time–frequency plot across patients time locked to spindle starts from −1.25 to −0.75 s before DS troughs. DSs can be seen at −0.25 s before these spindle starts, marked by the pink arrow. Note that these spindle-preceding DSs also precede the TBs and the DSs shown in ***A*** and ***D*** by ~1 s, as shown by the two negative deflections in the superimposed gray waveform (color scale indicates ± 1 dB). Scale bar, 100 μV.

tected TBs occurred in relation to DSs recorded intracranially. To quantify this, we created event-related histograms for each bipolar channel by relating the start of TBs to DS troughs. For example, a recording from Patient 1 in the right middle frontal gyrus shows that, when TBs are detected around DSs, they are more likely to start ~300 ms before the DS trough (Fig. 4*A*). At the same cortical location, the likelihood of slow spindles (Fig. 4*B*), fast spindles (Fig. 4*C*), or all spindles (Fig. 4*D*) starting between

Table 2. Theta burst characteristics

Patient	No. of cortical channels	No. of thalamic channels	NREM time (min)	Cortical theta rate (1/min)	Cortical theta frequency (Hz)	Cortical theta duration (ms)
1	7	3	240	1.83 (0.43)	6.44 (0.2)	669 (22)
2	13	3	146	0.78 (0.3)	6.13 (0.1)	679 (29)
3	4	0	40	1.24 (0.36)	6.05 (0.1)	688 (21)
4	5	0	26	1.09 (0.14)	6.39 (0.14)	646 (40)
5	7	0	58	1.07 (0.27)	6.37 (0.12)	622 (19)
6	7	0	144	0.61 (0.34)	6.32 (0.14)	701 (29)
7	8	0	161	1.28 (0.35)	6.21 (0.19)	696 (22)
8	3	0	90	1.44 (0.2)	6.62 (0.14)	640 (25)
9	2	2	202	1.64 (0.79)	6.65 (0.13)	672 (1)
10	5	0	98	0.95 (0.22)	6.25 (0.1)	667 (27)
Avg:	6.1 (3.1)	—	120.5 (70.27)	1.2 (0.38)	6.34 (0.2)	668 (25.3)

Parentheses indicate SD.

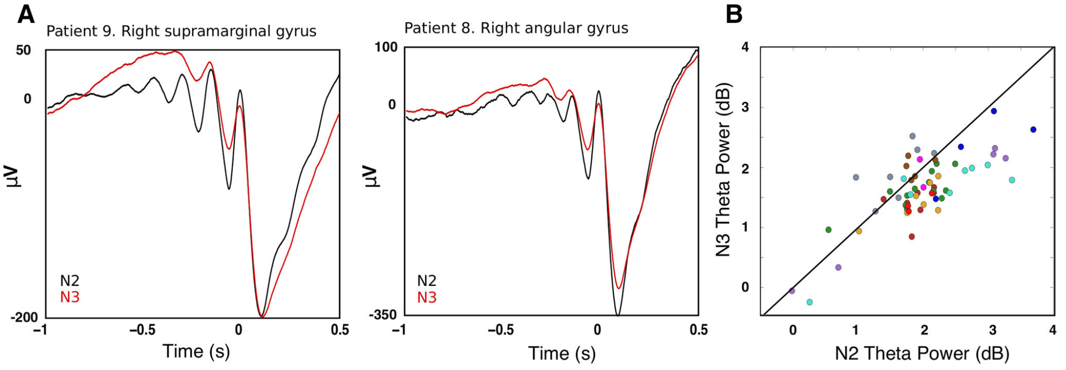


Figure 3. Comparing N2 and N3 TBs. ***A***, Theta oscillations before DS troughs were more pronounced for N2 than for N3 DSs. Average waveforms of the DS with the top quartile of prior theta power are shown from two patients within N2 in black ($n = 677$ in Patient 9 and 174 in Patient 8) and N3 in red ($n = 913$ in Patient 9 and 635 in Patient 8). Unfiltered averages are locked to the filtered theta peak just preceding the DS trough. ***B***, Power within theta (5–8 Hz) range −500 ms to 0 relative to DS trough for each channel (channels from a given patient have the same color). Most dots are below the diagonal, indicating that theta power was greater for N2 than for N3 ($t = -3.13$, $p = 0.002$, mixed-effects model with patient as random effect and channel as nested random effect).

−500 ms and 0 is greatly reduced compared with after the DS trough.

Of 60 cortical recordings, 35 channels from seven patients had at least 20 TBs within ± 500 ms of the DS trough, none of which was significantly more likely to start after the DS trough. However, 14 of the 35 channels from five patients were more likely to have TBs start before the DS trough (binomial test, Bonferroni adjusted $p < 0.05$). The proportion of TBs occurring before the DS trough for these channels was not significantly different between the frontal and parietooccipital regions ($t = 1.1$, $p = 0.27$).

Previous work (Mölle et al., 2011; Klinzing et al., 2016) asserts that spindles have different temporal relationships with DSs depending on spindle frequency; with slower spindles ($\leq$ 12 Hz) occurring on the up to down transition and faster spindles (>12 Hz) occurring on the down to up transition. To test this hypothesis, we selected from the 60 cortical recordings, the 35 channels from seven patients who had at least 20 slow spindles within ± 500 ms of the DS trough. None of these channels showed slower spindles that were significantly more likely to start before the DS trough. However, slower spindles recorded by 20 of the 35 channels from 6 patients were more likely to start after the DS trough (Bonferroni-adjusted $p < 0.05$). If all spindles were grouped together, then those recorded by 39/49 channels from 10 patients were significantly more likely to start after the DS trough (Bonferroni-adjusted $p < 0.05$). Neither slower spindles ($t = -0.44$, $p = 0.66$) nor all spindles ($t = -1.25$, $p = 0.21$) showed significant differences between frontal and parietooccipital regions in the proportion of spindle events occurring after DS troughs.

Normalizing each event-related histogram per channel by the total number of counts in the ±1 s time window and pooling histograms across all cortical bipolar channels from 10 patients revealed similar results (Fig. 4*E*–*H*): TBs initiate before DS troughs and spindles, both slow and fast, initiate after. Significance was assessed by Bonferroni correction ($p < 0.05$) across time bins within each event type. Time bins with significant likelihood of TBs starting were centered around −450 and −350 ms before the DS trough (Fig. 4*E*). In contrast, slower spindles were significantly more likely to start 350–450 ms after the DS trough (Fig. 4*F*) and faster spindles were significantly likely to start between bins centered 150–550 ms after the DS trough (Fig. 4*G*). This confirms that spindles are more likely to start during the transition from down to USs, whereas TBs start just before the transition to a DS.

Some time bins (Fig. 4*B*–*D*, *G*,*H*) showed an increased likelihood of spindles starting from ∼−1.25 s to −0.75 s before DS troughs. To determine whether this increase was due to a previous DS, we identified the "early" spindles in question as those that began at ∼−1 s before down state troughs (between −1.25 s and −0.75 s relative to the DS trough). Then, we generated a time–frequency plot for the activity surrounding these spindles for each

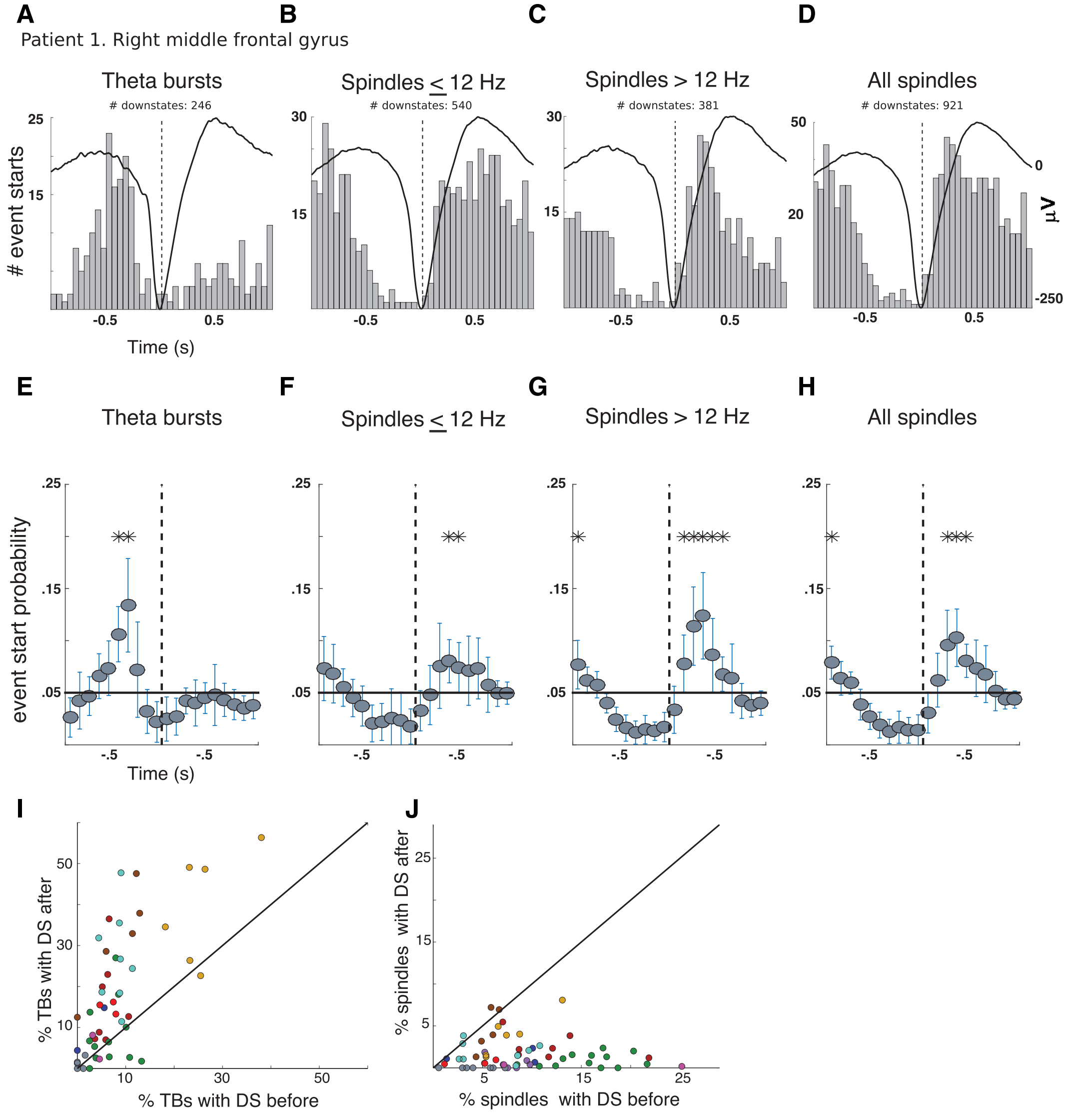


Figure 4. Cortical theta precedes DSs and both slow and fast spindles follow. ***A***–***D***, Event-related histograms showing the timing of the start of TBs (***A***), spindles $\leq$ 12 Hz (***B***), spindles $>$12 Hz (***C***), and all spindles (***D***) relative to the DS trough at a single bipolar channel in the right middle frontal gyrus of Patient 1 during both N2 and N3 sleep. ***E***–***H***, Histograms from all channels normalized by the total number of counts ± 1 s and pooled in 100 ms bins. The black line indicates chance level and probability estimates per bin across channels were calculated using linear mixed-effects models. Stars indicate bins where events occur significantly more than chance (Bonferroni adjusted, $p < 0.05$). Error bars indicate 95% confidence intervals (CIs). ***I***, Scatterplot for each channel of the proportion of TBs with the DSs within 500 ms after (*y*-axis) versus the proportion with the DS within 500 ms before (*x*-axis). ***J***, Same as ***I*** but for spindles. All cortical channels with at least 50 events were included. Most channels (dots) in ***I*** are above the diagonal, indicating that the DS trough usually occurs after TB onset; in contrast, most channels in ***J*** are below the diagonal, indicating that the DS trough usually occurs before spindle onsets.

channel and averaged them across the 33 channels in six patients who had at least 50 such spindles (Fig. 2*H*). A DS is clearly present in the average time–frequency plot, peaking ~.25 s before the onset of these early spindles. The same plot shows another DS at ~1 s following early spindle onset; this is the DS that was used to identify the early spindles in the first place, corresponding to the DS at 0.0 s in plots 2AD. The early spindle power increase precedes the theta band increase in this plot, as well as those triggered on the main DS (Fig. 2*A*,*D*). In summary, we demonstrate that there is a sequence of sleep grapho-elements, typically theta–DS–spindle, but sometimes DS1–spindle1–theta–DS2–spindle2. The second, longer sequence is expected given that, especially in stage N3, DSs are well known to occur rhythmically at ~1 Hz, comprising the slow oscillation.

Next we investigated how often TBs or spindles occur around DSs. Only channels that had at least 50 occurrences of each event type (theta or spindle) were included ($n = 9$ patients, 49 cortical channels). DS masks were marked by DS troughs and spindle and theta event masks marked by estimated start times. We found that, on average across patients, 24% of detected TBs began within ± 500 ms of a DS trough, with 7.7% of TBs occurring after a DS and 17.8% occurring before (Fig. 4*I*). In contrast, 10% of detected spindles began within ± 500 ms of a DS trough, with 8.7% of spindles occurring after a DS and 1.7% occurring before (Fig. 4*J*). There was a significantly greater proportion of TBs that fell within ± 500 ms of a DS trough than spindles (paired *t* test, $p = 2.4\text{e-}05$). There was substantial interpatient variability for both events, especially for TBs. Despite this variability, most patients exhibited similar temporal relationships with DSs.

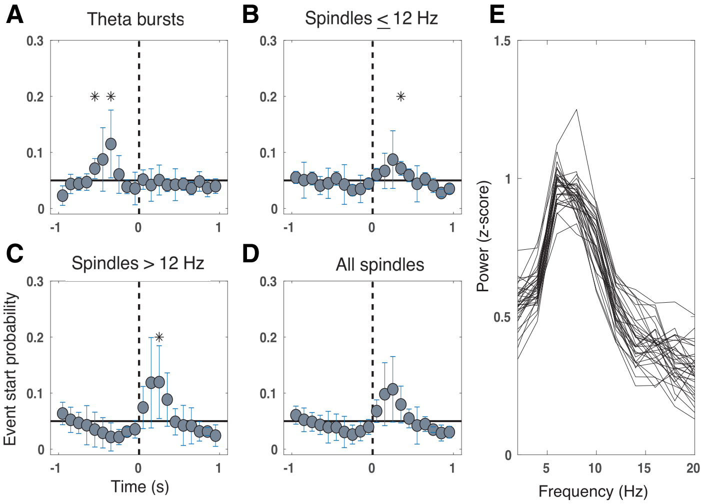


Figure 5. Scalp EEG theta precedes DSs and spindles follow. As in Figure 4, ***E–H***, DS-locked event-related histograms from all scalp EEG channels normalized by the total number of counts ±1 s and pooled in 100 ms bins are plotted for the start times of: TBs (***A***), spindles ≤ 12 Hz (***B***), spindles > 12 Hz (***C***), and all spindles (***D***). The black line indicates chance level and probability estimates per bin across channels were calculated using linear mixed-effects models. Stars indicate bins where events occur significantly more than chance (Bonferroni-adjusted $p < 0.05$). Error bars indicate 95% confidence intervals (CIs). ***E***, Pre-DS spectra. Plotted are the differences of the first and fourth quartile average power spectral densities of 500 ms pre-DS, when ranked by the amount of 4–12 Hz power they contain. Each trace represents the difference of PSDs for one channel averaged across all subjects. Only channels common to all subjects are included.

Relation of TBs and spindles recorded in scalp EEG to DSs

To ensure that that our findings generalize to nonpatient populations and to contextualize our results within the framework of more commonly recorded noninvasive measures, we detected events recorded using scalp EEG in three healthy human participants and recalculated event-related histograms, as shown in Figure 4, *E–H*. Replicating the SEEG results reported above, normalized peri-DS histograms revealed divergent patterns for TBs and spindles recorded at the scalp (Fig. 5): TBs initiate before DS troughs, whereas spindles begin afterward. Significance was assessed by Bonferroni correction ($p < 0.05$) across time bins within each event type. TBs had a significant likelihood of starting 500–300 ms before the DS trough (Fig. 5*A*). In contrast, slower and faster spindles were significantly more likely to start during time bins centered on 450 and 350 ms, respectively, after the DS trough (Fig. 5*B*,*C*). These results are consistent with our conclusions from SEEG: spindles are more likely to start during the transition from down to USs, whereas TBs start just before the transition to a DS. Furthermore, as with Figure 2*G*, we calculated PSDs over 500 ms before DS troughs, sorted DSs by power in 4–12 Hz, took the difference of the top and bottom quartile average PSDs, and found that power is centered within theta range (Fig. 5*E*). Each trace is a scalp channel averaged across subjects.

Relating thalamic theta with cortical theta

Three patients also had SEEG electrodes that recorded from the thalamus and we recently characterized the coordination of NREM DS and spindles between cortex and thalamus in these patients (Mak-McCully et al., 2017). Here, we investigated whether the thalamus also exhibits similar TBs to the cortex and, if so, how they relate to cortical TBs.

Both time–frequency representations locked to DSs (Fig. 6*A*, example channel) and averaging the raw LFP for the top and bottom quartiles of prior DS theta power (Fig. 6*B*, example channel) suggested that there are similar TBs before DSs in the thalamus. These oscillations could also be seen at the level of single DSs (Fig. 6*B*, inset) and often revealed a downward slope in the average LFP of detected TBs (Fig. 6*C*), similar to some cortical channels. Compared with cortical channels from the same patients (shown in Table 2), thalamic channels showed no difference in the rate of theta occurrence (mean ± SD for thalamus = 1.07 ± 0.66; for cortex = 1.39 ± 0.62; $t = 1.4, p = 0.16$), average frequency (mean ± SD for thalamus = 6.25 ± 0.13; for cortex = 6.33 ± 0.22; $t = 1.12, p = 0.26$), or duration (mean ± SD for thalamus = 661 ± 17 ms; for cortex = 675 ± 25 ms; $t = 1.43, p = 0.15$) in detected events. This is in contrast with DSs, which showed greater rates of occurrence in the cortex, and spindles, which occurred more frequently in the thalamus, as shown in our previous study (Mak-McCully et al., 2017).

We found that cortical TBs only slightly overlapped with thalamic TBs after correcting for the overlap expected by chance (average overlap per patient = 1.9%, 3.7%, and 5.1% above chance), but this overlap was nonetheless significant ($t = 4.04$, $p = 5.3\text{e-}05$; patient as random effect, corticothalamic pair as nested random effect). In contrast, cortical spindles in most sites overlapped strongly with thalamic spindles over the proportion expected by chance (average overlap per patient = 24.2%, 15.6%, and 26.6% above chance) and, again, this was highly significant ($t = 6.22$, $p = 5\text{e-}10$; Fig. 6*D*). This greater corticothalamic overlap for spindles compared with TBs is significant ($t = 14.2$, $p = 1.2\text{e-}45$).

We also examined the cooccurrence of thalamic and cortical TBs using their joint occurrence histograms. Only 5/64 corticothalamic pairs had at least 20 thalamic TBs starting within ± 500 ms of the start of cortical TBs, and none of these 5 pairs had histograms with a significant difference between leading versus lagging peaks as assessed with a binomial test ($p < 0.05$, Bonferonni corrected). This is in contrast to spindles, which showed

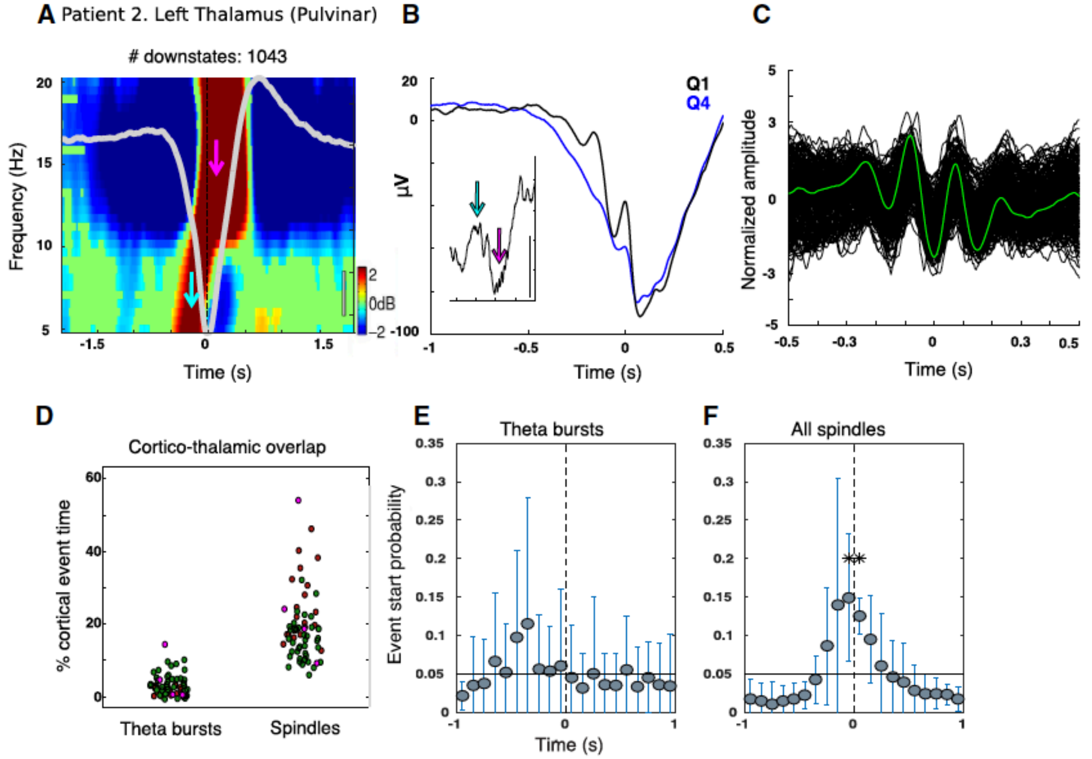


Figure 6. Thalamic theta during NREM sleep. ***A***, Example time–frequency representations locked to DS troughs for a bipolar thalamic channel from Patient 2 showing increased power in the theta range before the trough (cyan arrow) and in the spindle range just after the trough (pink arrow). The average DS is overlaid in light gray. Black scale bar indicates 20 μV. (See Fig. 2*B* in Mak-McCully et al., 2017, for time–frequency plots extending to 100 Hz). ***B***, Prior theta power is not solely driven by DS waveform shape, but displays oscillatory activity in the theta range after averaging the raw LFP locked to the first theta peak preceding the DS trough (as in Fig. 2*B*, *E*). The inset displays a single trace from the top quartile for ± 1 s around DS trough. Scale bar indicates 100 μV. Cyan and pink arrows show individual theta and spindle events, respectively. ***C***, All detected theta events locked to the deepest trough for the same channel as ***A*** and ***B***. ***D***, Proportion of cortical and thalamic overlap for TBs and spindles. Each dot is a different channel; dots from the same patient share the same color. *y*-axis indicates the proportion of time a given cortical channel has a TB (spindle) overlapping with a thalamic TB (spindle) after subtracting the amount of overlap expected by chance. Spindles show a greater degree of corticothalamic overlap compared with TBs ($t = 14.2$, $p =$ 1.2e-45; patients as random effects, corticothalamic pairs as nested random effects). As in Figure 4, *E–H*, DS-locked event-related histograms from 8 thalamic channels, normalized by the total number of counts ±1 s and pooled in 100 ms bins are plotted for the start times of: TBs (***E***) and all spindles (***F***).

significant ordering effects with the thalamus preceding the cortex, and DSs, which usually started in the cortex before thalamus (Mak-McCully et al., 2017). These findings provide support that our detected theta events are separate entities from DSs and spindles because they appear to have distinct corticothalamic dynamics.

When relating thalamic TBs to thalamic DSs, six of eight thalamic channels from three patients had at least 20 TBs within ± 500 ms of the DS trough. Of these six, two channels from two patients were more likely to have TBs start before the DS trough (binomial test, Bonferroni adjusted $p < 0.05$); no channels were more likely to have TBs start after the trough. When pooled across patients, TBs were not more likely to start in any time bin over chance relative to DS troughs (Fig. 6*E*). In contrast, peak occurrence of thalamic spindle onset was at the thalamic DS peak (Fig. 6*F*), as has been reported previously (Mak-McCully et al., 2017).

Theta and spindles differ in high-frequency coupling

To determine whether the TBs before DSs modulated HG power, we took the top quartile of DSs with prior theta power (5–8 Hz, −500 ms to DS trough) and averaged both the HG (60–100 Hz) power and LFP locked to the band-passed theta peak first preceding the trough in the bipolar cortical channels. This revealed clear oscillations within the theta range in the HG power (Fig. 7*A*, example channel). We performed a similar analysis for spindles in the same channels, only with DSs sorted by spindle power 0–500 ms after DSs (Fig. 7*B*), and averaged time-locked to the first post-DS filtered peak in 10–16 Hz. To quantify the extent of this modulation, we used traditional cross-frequency coupling metrics at the beginning (first two cycles) of detected spindle and TBs (see Materials and Methods).

We found that, for TBs, 42 cortical channels from 10 patients showed significant PAC-Z ($p < 0.002$, uncorrected for multiple comparisons), and 45 channels from 10 patients were significant for spindle events. There was a strong correlation between PAC-Z measures from the two event types across channels (Fig. 7*C*; Spearman's $\rho = 0.59$, 10 patients, 57 channels) and no difference in amplitude of PAC between theta and spindle events ($p = 0.15$; $n = 10$, 57 channels; nested random effects of channel in patient). However, there was difference in the preferred phase for HG power for TB (at ~200°) compared with spindles (at ~140°; Fig.

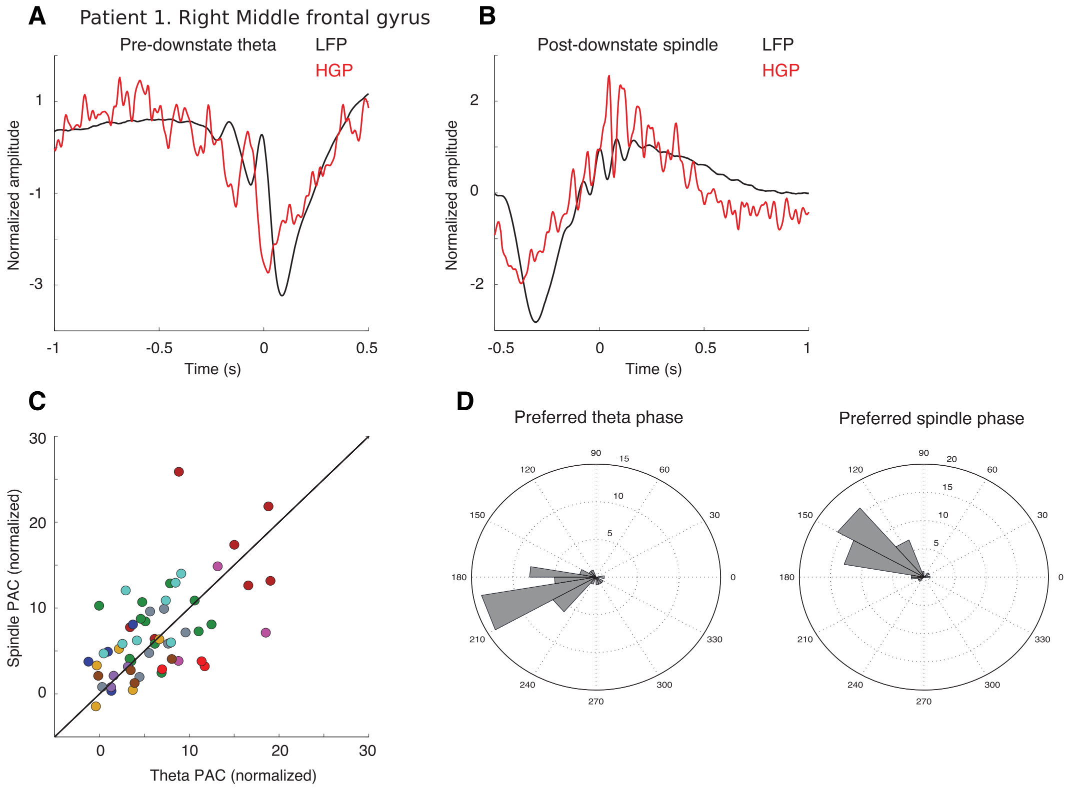


Figure 7. Theta and spindles couple to HG power differently. ***A***, ***B***, Patient 1, right middle frontal gyrus. ***A***, Average of the top quartile of DSs (ranked by pre-DS theta power) locked to the first prior theta peak (black), as well as the average of HG power (60 –100 Hz) for the same events (red). ***B***, Average of the top quartile of DSs (ranked by post-DS spindle power) locked to the first filtered spindle peak following the DS trough (black), as well as the average of the HG power. ***C***, Magnitude of PAC between theta phase and HG (*x*-axis) versus spindle phase and HG (*y*-axis) for all bipolar channels with unity line overlaid for reference. Overall, a similar level of coupling is observed for TBs versus spindles. ***D***, Preferred phase of either theta or spindle events for each channel that had significant PAC. Radial scale is number of channels. The preferred phase is consistent for each type of wave and differs significantly between TBs and spindles (parametric Watson–Williams test, $F = 26.5$, $p =$ 1.6e-6).

7*D*). This difference was highly significant (parametric Watson–Williams test, $F = 26.5$, $p = 1.6e\text{-}6$).

Discussion

Loomis' original study recording sleep EEGs in 1939 commented on sleep spindles following KCs (i.e., DSs; Cash et al., 2009). This coupling was later quantified (Mölle et al., 2002) and then extended to assert that, whereas "fast spindles" occur after DSs, "slow spindles" occur before (Mölle et al., 2011; Klinzing et al., 2016; Yordanova et al., 2017). Like these studies, we found contrasting patterns of oscillatory activity before versus after DS. However, unlike these studies, we found that both slow and fast spindles occurred post-DS in direct cortical recordings from both frontal and occipitoparietal sites (Fig. 4), as well as scalp EEG (Fig. 5). Rather than slow spindles, we found that short TBs precede DSs whether in cortical or scalp recordings. Increased spectral power preceding DS is mainly within the theta band centered at ~6 Hz (Figs. 2*G*, 5*E*), but extends into the spindle range. Therefore, if this activity is band-pass filtered in the spindle range, then low-frequency "spindles" could be detected despite the center frequency of unfiltered recordings being in the theta band.

Previous scalp studies have reported increased theta power before DS troughs (Cox et al., 2014b; Klinzing et al., 2016). We confirmed that these were true oscillations by averaging the raw LFP locked to a theta peak before the DS (Fig. 2*B*,*E*) and by requiring each potential TB contain at least three peaks. TBs thus consist of multiple 4–8 Hz waves with a mean duration of ~670 ms, shorter than most spindles. They have fewer cycles and lower amplitude in N3 compared with N2. The density of TB is ~7-fold less than spindles and, unlike spindles (Mak-McCully et al., 2017), TB density does not differ between cortex and thalamus. In addition, whereas spindles are tightly coupled between thalamus and cortex, no significant relationship can be observed for TB. Finally, cortical TBs have a significantly different phase relation to HG than spindles, indicating distinct generators (Fig. 7*D*). Therefore, TBs are distinguished from spindles in their internal frequency, duration, density, position relative to the DS, lack of overlap or consistent sequencing between thalamus and cortex, and distinct HG phase preferences.

If slow versus fast spindles do not differ in their relation to the DS, then is it still tenable to claim that they represent distinct neurophysiological phenomena rather than variations within a continuum? Our study confirms the slightly but significantly higher average frequency of spindles in parietal versus frontal cortex, consistent with previous EEG and MEG recordings (Dehghani et al., 2011) and intracranial recordings (Andrillon et al., 2011; Peter-Derex et al., 2012; Piantoni et al., 2017). The individual waves in EEG and MEG spindle bursts also vary in frequency, with later waves also ~1 Hz slower on average (Dehghani et al., 2011). In SEEG, within a given cortical location and often within the same spindle, spindle waves with frequencies both above and below the fast/slow division are typically observed. Two kinds of spindles can be identified in laminar recordings, involving mainly upper or middle layers (Hagler et al., 2018). The average frequency of upper versus middle channel spindles does not differ significantly and both include both slow and fast spindle waves. In all of these circumstances, fast and slow spindles occur in a continuum rather than a dichotomy.

Within this framework, it is not clear how to explain how, in some subjects, two peaks in the spindle spectrogram can be discerned at the scalp (Cox et al., 2017). The cortical origin of scalp EEG spindles is not yet well understood due to the lack of detailed information regarding the amplitude, density, synchrony, phase, and orientation of the generating cortical patches. It is thus theoretically possible that the slower spindles reported at the scalp are from a location where we did not record. However, the parietal and frontal cortices where we recorded have been proposed to be the generators of fast and slow scalp spindles (Mölle et al., 2011; Klinzing et al., 2016) and generate spindle band activity most related to scalp EEG spindles (Frauscher et al., 2015). Because the inverse problem is ill posed, it is possible to model the scalp EEG spindle distribution as being due to either anatomically distinct generators, each with a single frequency, or distributed generators, each with a range of overlapping frequencies, changing slightly across areas. Our results clearly support the second model.

This supposed dichotomy between slow pre-DS spindles and fast post-DS spindles in humans has been homologized to the clear dichotomy in rodents between high-voltage slow spindles (~8 Hz) versus low-voltage fast spindles (~14 Hz) (Timofeev and Chauvette, 2013). Only the fast spindles are associated with memory replay and consolidation (Eschenko et al., 2006; Johnson et al., 2010) or DS (Johnson et al., 2010). The sharp waveforms and other epileptiform characteristics of slow rodent spindles (Polack and Charpier, 2006) suggest that they do not have an homology in healthy human recordings. Therefore, it appears that fast rodent spindles correspond to both faster and slower spindles in humans.

The TBs preceding DS in N2 comprise an augmenting oscillation between cortical excitation and inhibition as indexed by phase-locked HG, which is correlated with neuronal firing (Lachaux et al., 2012). The greatest HG decrease occurs at the final surface-negative TB trough, which coincides with the DS trough. Therefore, in TB–DS sequences, the DS does not arise as a sudden decline from baseline, but as the culmination of an escalating TB oscillation, suggesting that the TB may play a role in helping to trigger the DS as its final cycle. This possibility receives some indirect support from the fact that they are generated by the same cortical layers in laminar recordings (Csercsa et al., 2010; Halgren et al., 2018) and thus may be engaging the same circuits. The hypothesis that theta waves may trigger DS in N2 provides a solution to a difficult question: how do KCs arise? Current theories model DS onset as a response to the preceding US (Neske, 2015). This view describes the usual *in vitro* or anesthetized recordings where USs arise from a flat depressed baseline, which is considered the DS (Lemieux et al., 2014). In contrast, in unanesthetized humans (Mak-McCully et al., 2015) and animals (Chauvette et al., 2011), DS appear as stereotyped events in a chronically active cortex during NREM sleep. Laminar recordings in humans demonstrate that KCs are DS without a preceding US (Cash et al., 2009). We suggest here that the last positive peak of the TB may replace the US as the DS trigger. For example, the calcium influx associated with this cortical excitation during the final positive peak would trigger hyperpolarizing K^+ currents (Cunningham et al., 2006), which may tip the circuit into the DS. However, although TBs could serve as a trigger for DSs in N2, the majority of DSs are not preceded by TBs, so this triggering phenomenon may only occur in some circumstances.

The consistent relation of TB to DS and DS to spindles may be related to the appearance of these waves in successive stages of sleep. The transition from quiet waking to N1 is marked by the replacement of alpha by theta waves. N2 then appears when KCs and sleep spindles appear. Simultaneous corticothalamic recordings in natural human sleep show that converging cortical DS precede thalamic DS, that thalamic spindles are tightly coupled to begin at the thalamic DS trough, and that thalamic spindles drive cortical (Mak-McCully et al., 2017). In this view, the N1 to N2 transition would occur when theta begins to trigger cortical DSs (i.e., KCs), which in turn trigger successively thalamic DS, thalamic spindles, and cortical spindles. N2 transitions to N3 when DS and US recur rhythmically as the slow oscillation. Both TBs (shown here) and spindles (Andrillon et al., 2011; Mak-McCully et al., 2017; Piantoni et al., 2017) continue during N3, but become abbreviated. We hypothesize that, as the neuromodulatory state deepens, the US triggered by the DS becomes capable of triggering the following DS, resulting in less time for elaboration of the TB or spindles.

During NREM sleep, hippocampal cells replay events from the preceding waking period during sharp-wave ripples, which arrive at the cortex during the down to US transition, as the spindle is beginning (Maingret et al., 2016; Jiang et al., 2017). This conjunction of hippocampal input with cortical modulation is thought to underlie consolidation of cortical memory circuits (Diekelmann and Born, 2010; Mölle et al., 2011; Niknazar et al., 2015; Hanert et al., 2017; Latchoumane et al., 2017). Our results suggest that TBs may precede and could help to initiate the DS–ripple–spindle–US sequence. During waking, theta occurs during tasks requiring sustained processing (Kahana et al., 1999; Raghavachari et al., 2006) and may underlie prominent cognitive event-related potentials (Cavanagh et al., 2012; Halgren et al., 2015). If theta serves a similar function during NREM, then TB may organize the gathering of cortical information before the DS and consequently help to select related hippocampal traces to be activated and sent back to the cortex for integration with currently active neurons during the US and spindle.

References

Andrillon T, Nir Y, Staba RJ, Ferrarelli F, Cirelli C, Tononi G, Fried I (2011) Sleep spindles in humans: insights from intracranial EEG and unit recordings. J Neurosci 31:17821–17834. CrossRef Medline

Barthó P, Slézia A, Mátyás F, Faradzs-Zade L, Ulbert I, Harris KD, Acsády L (2014) Ongoing network state controls the length of sleep spindles via inhibitory activity. Neuron 82:1367–1379. CrossRef Medline

Bazhenov M, Timofeev I, Steriade M, Sejnowski TJ (2002) Model of thalamocortical slow-wave sleep oscillations and transitions to activated states. J Neurosci 22:8691–8704. CrossRef Medline

Bonjean M, Baker T, Lemieux M, Timofeev I, Sejnowski T, Bazhenov M (2011) Corticothalamic feedback controls sleep spindle duration in vivo. J Neurosci 31:9124–9134. CrossRef Medline

Canolty RT, Edwards E, Dalal SS, Soltani M, Nagarajan SS, Kirsch HE, Berger

MS, Barbaro NM, Knight RT (2006) High gamma power is phase-locked to theta oscillations in human neocortex. Science 313:1626–1628. CrossRef Medline

Cash SS, Halgren E, Dehghani N, Rossetti AO, Thesen T, Wang C, Devinsky O, Kuzniecky R, Doyle W, Madsen JR, Bromfield E, Eross L, Halász P, Karmos G, Csercsa R, Wittner L, Ulbert I (2009) The human K-complex represents an isolated cortical down-state. Science 324:1084–1087. CrossRef Medline

Cavanagh JF, Zambrano-Vazquez L, Allen JJ (2012) Theta lingua franca: A common mid-frontal substrate for action monitoring processes. Psychophysiology 49:220–238. CrossRef Medline

Chauvette S, Crochet S, Volgushev M, Timofeev I (2011) Properties of slow oscillation during slow-wave sleep and anesthesia in cats. J Neurosci 31:14998–15008. CrossRef Medline

Contreras D, Destexhe A, Sejnowski TJ, Steriade M (1996) Control of spatiotemporal coherence of a thalamic oscillation by corticothalamic feedback. Science 274:771–774. CrossRef Medline

Cox R, Hofman WF, de Boer M, Talamini LM (2014a) Local sleep spindle modulations in relation to specific memory cues. Neuroimage 99:103–110. CrossRef Medline

Cox R, van Driel J, de Boer M, Talamini LM (2014b) Slow oscillations during sleep coordinate interregional communication in cortical networks. J Neurosci 34:16890–16901. CrossRef Medline

Cox R, Schapiro AC, Manoach DS, Stickgold R (2017) Individual differences in frequency and topography of slow and fast sleep spindles. Front Hum Neurosci 11:433. CrossRef Medline

Crunelli V, Hughes SW (2010) The slow (below 1 Hz) rhythm of non-REM sleep: a dialogue between three cardinal oscillators. Nat Neurosci 13:9–17. CrossRef Medline

Csercsa R, Dombovári B, Fabó D, Wittner L, Eross L, Entz L, Sólyom A, Rásonyi G, Szucs A, Kelemen A, Jakus R, Juhos V, Grand L, Magony A, Halász P, Freund TF, Maglóczky Z, Cash SS, Papp L, Karmos G et al. (2010) Laminar analysis of slow wave activity in humans. Brain 133:2814–2829. CrossRef Medline

Cunningham MO, Pervouchine DD, Racca C, Kopell NJ, Davies CH, Jones RS, Traub RD, Whittington MA (2006) Neuronal metabolism governs cortical network response state. Proc Natl Acad Sci U S A 103:5597–5601. CrossRef Medline

Dehghani N, Cash SS, Halgren E (2011) Topographical frequency dynamics within EEG and MEG sleep spindles. Clin Neurophysiol 122:229–235. CrossRef Medline

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134:9–21. CrossRef Medline

Diekelmann S, Born J (2010) The memory function of sleep. Nat Rev Neurosci 11:114–126. CrossRef Medline

Duvernoy H (1999) The human brain: surface, blood supply, and three-dimensional sectional anatomy, Ed 22. New York: Springer.

Dykstra AR, Chan AM, Quinn BT, Zepeda R, Keller CJ, Cormier J, Madsen JR, Eskandar EN, Cash SS (2012) Individualized localization and cortical surface-based registration of intracranial electrodes. Neuroimage 59:3563–3570. CrossRef Medline

Eschenko O, Mölle M, Born J, Sara SJ (2006) Elevated sleep spindle density after learning or after retrieval in rats. J Neurosci 26:12914–12920. CrossRef Medline

Fogerson PM, Huguenard JR (2016) Tapping the brakes: cellular and synaptic mechanisms that regulate thalamic oscillations. Neuron 92:687–704. CrossRef Medline

Frauscher B, von Ellenrieder N, Dubeau F, Gotman J (2015) Scalp spindles are associated with widespread intracranial activity with unexpectedly low synchrony. Neuroimage 105:1–12. CrossRef Medline

Hagler DJ Jr, Ulbert I, Wittner L, Erőss L, Madsen JR, Devinsky O, Doyle W, Fabó D, Cash SS, Halgren E (2018) Heterogeneous origins of human sleep spindles in different cortical layers. J Neurosci 38:3013–3025. CrossRef Medline

Halgren E, Kaestner E, Marinkovic K, Cash SS, Wang C, Schomer DL, Madsen JR, Ulbert I (2015) Laminar profile of spontaneous and evoked theta: rhythmic modulation of cortical processing during word integration. Neuropsychologia 76:108–124. CrossRef Medline

Halgren M, Fabó D, Ulbert I, Madsen JR, Erőss L, Doyle WK, Devinsky O, Schomer D, Cash SS, Halgren E (2018) Superficial slow rhythms integrate cortical processing in humans. Sci Rep 8:2055. CrossRef Medline

Hanert A, Weber FD, Pedersen A, Born J, Bartsch T (2017) Sleep in humans stabilizes pattern separation performance. J Neurosci 37:12238–12246. CrossRef Medline

Iber C, Ancoli-Israel S, Chesson AL, Quan SF (2007) The AASM manual for the scoring of sleep and associated events: rules, terminology, and technical specifications, Ed 1. Westchester, IL: American Academy of Sleep Medicine.

Jiang X, Shamie I, K Doyle W, Friedman D, Dugan P, Devinsky O, Eskandar E, Cash SS, Thesen T, Halgren E (2017) Replay of large-scale spatiotemporal patterns from waking during subsequent NREM sleep in human cortex. Sci Rep 7:17380. CrossRef Medline

Johnson LA, Euston DR, Tatsuno M, McNaughton BL (2010) Stored-trace reactivation in rat prefrontal cortex is correlated with down-to-up state fluctuation density. J Neurosci 30:2650–2661. CrossRef Medline

Jones SR (2016) When brain rhythms aren't "rhythmic": implication for their mechanisms and meaning. Curr Opin Neurobiol 40:72–80. CrossRef Medline

Kahana MJ, Sekuler R, Caplan JB, Kirschen M, Madsen JR (1999) Human theta oscillations exhibit task dependence during virtual maze navigation. Nature 399:781–784. CrossRef Medline

Klinzing JG, Mölle M, Weber F, Supp G, Hipp JF, Engel AK, Born J (2016) Spindle activity phase-locked to sleep slow oscillations. Neuroimage 134:607–616. CrossRef Medline

Kokkinos V, Kostopoulos GK (2011) Human non-rapid eye movement stage II sleep spindles are blocked upon spontaneous K-complex coincidence and resume as higher frequency spindles afterwards. J Sleep Res 20:57–72. CrossRef Medline

Kokkinos V, Koupparis AM, Kostopoulos GK (2013) An intra-K-complex oscillation with independent and labile frequency and topography in NREM sleep. Front Hum Neurosci 7:163. CrossRef Medline

Lachaux JP, Axmacher N, Mormann F, Halgren E, Crone NE (2012) High-frequency neural activity and human cognition: past, present and possible future of intracranial EEG research. Prog Neurobiol 98:279–301. CrossRef Medline

Latchoumane CV, Ngo HV, Born J, Shin HS (2017) Thalamic spindles promote memory formation during sleep through triple phase-locking of cortical, thalamic, and hippocampal rhythms. Neuron 95:424–435.e6. CrossRef Medline

Lemieux M, Chen JY, Lonjers P, Bazhenov M, Timofeev I (2014) The impact of cortical deafferentation on the neocortical slow oscillation. J Neurosci 34:5689–5703. CrossRef Medline

Maingret N, Girardeau G, Todorova R, Goutierre M, Zugaro M (2016) Hippocampo-cortical coupling mediates memory consolidation during sleep. Nat Neurosci 19:959–964. CrossRef Medline

Mak-McCully RA, Rolland M, Sargsyan A, Gonzalez C, Magnin M, Chauvel P, Rey M, Bastuji H, Halgren E (2017) Coordination of cortical and thalamic activity during non-REM sleep in humans. Nat Commun 8:15499. CrossRef Medline

Mak-McCully RA, Rosen BQ, Rolland M, Régis J, Bartolomei F, Rey M, Chauvel P, Cash SS, Halgren E (2015) Distribution, amplitude, incidence, co-occurrence, and propagation of human k-complexes in focal transcortical recordings. eNeuro 2:ENEURO.0028–15.2015. CrossRef Medline

Mölle M, Bergmann TO, Marshall L, Born J (2011) Fast and slow spindles during the sleep slow oscillation: disparate coalescence and engagement in memory processing. Sleep 34:1411–1421. CrossRef Medline

Mölle M, Marshall L, Gais S, Born J (2002) Grouping of spindle activity during slow oscillations in human non-rapid eye movement sleep. J Neurosci 22:10941–10947. CrossRef Medline

Morel A, Magnin M, Jeanmonod D (1997) Multiarchitectonic and stereotactic atlas of the human thalamus. J Comp Neurol 387:588–630. CrossRef Medline

Neske GT (2015) The slow oscillation in cortical and thalamic networks: mechanisms and functions. Front Neural Circuits 9:88. CrossRef Medline

Niknazar M, Krishnan GP, Bazhenov M, Mednick SC (2015) Coupling of thalamocortical sleep oscillations are important for memory consolidation in humans. PLoS One 10:e0144720. CrossRef Medline

Peter-Derex L, Comte JC, Mauguière F, Salin PA (2012) Density and frequency caudo-rostral gradients of sleep spindles recorded in the human cortex. Sleep 35:69–79. CrossRef Medline

Piantoni G, Halgren E, Cash SS (2017) Spatiotemporal characteristics of sleep spindles depend on cortical location. Neuroimage 146:236–245. CrossRef Medline

Polack PO, Charpier S (2006) Intracellular activity of cortical and thalamic neurones during high-voltage rhythmic spike discharge in Long-Evans rats in vivo. J Physiol 571:461–476. CrossRef Medline

Raghavachari S, Lisman JE, Tully M, Madsen JR, Bromfield EB, Kahana MJ (2006) Theta oscillations in human cortex during a working-memory task: evidence for local generators. J Neurophysiol 95:1630–1638. CrossRef Medline

Rodenbeck, A. et al., (2006) A review of sleep EEG patterns. part I: aAcompilation of amended rules for their visual recognition according to Rechtschaffen and Kales. Somnologie 10:159–175. CrossRef

Sejnowski TJ, Destexhe A (2000) Why do we sleep? Brain Res 886:208–223. CrossRef Medline

Silber MH, Ancoli-Israel S, Bonnet MH, Chokroverty S, Grigg-Damberger MM, Hirshkowitz M, Kapen S, Keenan SA, Kryger MH, Penzel T, Pressman MR, Iber C (2007) The visual scoring of sleep in adults. J Clin Sleep Med 3:121–131. Medline

Steriade M (1997) Synchronized activities of coupled oscillators in the cerebral cortex and thalamus at different levels of vigilance. Cereb Cortex 7:583–604. CrossRef Medline

Talairach J, Tournoux P (1998) Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: an approach to cerebral imaging. New York, NY: Thieme.

Timofeev I, Chauvette S (2013) The spindles: are they still thalamic? Sleep 36:825–826. CrossRef Medline

Timofeev I, Grenier F, Bazhenov M, Sejnowski TJ, Steriade M (2000) Origin of slow cortical oscillations in deafferented cortical slabs. Cereb Cortex 10:1185–1199. CrossRef Medline

Yordanova J, Kirov R, Verleger R, Kolev V (2017) Dynamic coupling between slow waves and sleep spindles during slow wave sleep in humans is modulated by functional pre-sleep activation. Scientific Reports, 7, pp. 1–14.

Chapter 1, in full, is a reprint of the material as it appears in The Journal of Neuroscience, titled “Theta Bursts Precede, and Spindles Follow, Cortical and Thalamic Downstates in Human NREM Sleep”, November 2018.; 38(46). Gonzalez, Christopher; Mak-McCullly, Rachel; Rosen, Burke; Cash, Sydney; Chauvel, Patrick; Bastuji, Hélène; Rey, Marc; Halgren, Eric. The dissertation author was one of the primary investigators and authors of this paper.

CHAPTER 2: Human spindle variability

Abstract

In humans, sleep spindles are 10-16 Hz oscillations lasting approximately 0.5-2s. Spindles, along with cortical slow oscillations, facilitate memory consolidation by enabling synaptic plasticity. Early recordings of spindles at the scalp found anterior channels had overall slower frequency than central-posterior channels. This robust, topographical finding led to dichotomizing spindles as ‘slow’ versus ‘fast’, modelled as two distinct spindle generators in frontal versus posterior cortex. Using a large dataset of intracranial sEEG recordings (n=20, 365 bipolar recordings), we show that the difference in spindle frequency between frontal and parietal channels is comparable to the variability in spindle frequency within the course of individual spindles, across different spindles recorded by a given site, and across sites within a given region. Thus, fast and slow spindles only capture average differences that obscure a much larger underlying overlap in frequency. Furthermore, differences in mean frequency are only one of several ways that spindles differ. For example, compared to parietal, frontal spindles are smaller, tend to occur after parietal when both are engaged, and show a larger decrease in frequency within-spindles. Conversely, frontal and parietal spindles are similar in being longer, less variable, and more widespread than occipital, temporal, and Rolandic spindles. Lastly, spindles which are highly phase-locked to posterior hippocampal spindles are faster, longer, and less variable. We propose that rather than a strict parietal-fast/frontal-slow dichotomy, spindles differ continuously and quasi-independently in multiple dimensions, with variability due about equally to within-spindle, within-region and between-region factors.

Introduction

Spontaneous, electrical brain rhythms generated during sleep have been shown to play an active role in organizing and strengthening our memory capacities (Marshall et al., 2020; Rasch and Born, 2013). Two such rhythms are cortico-thalamic slow waves and sleep spindles (SS). Slow waves are large, ~0.5-4Hz rhythms composed of alternating downstates (DSs), or periods of neuronal quiescence, followed by upstates, where neuronal activity is similar to waking activity(Steriade et al., 1993). SSs are 10-16 Hz oscillations lasting 0.5-2s, and are generated by the inhibitory thalamic reticular nucleus interacting with excitatory thalamocortical cells (Steriade, 2003; Steriade et al., 1993). Some properties of SSs such as duration(Bonjean et al., 2011) and synchronization across the thalamus (Contreras et al., 1996), are shaped by corticothalamic feedback. SSs are grouped by slow waves, such that SSs begin on the down-to-upstate transition(Andrillon et al., 2011; Contreras and Steriade, 1995; Mölle et al., 2002). This grouping has been shown to facilitate memory consolidation(Mölle et al., 2009; Niknazar et al., 2015), likely mediated through massive calcium influx into apical dendrites and enabling synaptic plasticity (Seibt et al., 2017; Sejnowski and Destexhe, 2000).

In early scalp recordings, Gibbs and Gibbs (1950s) distinguished slower SSs occurring at more anterior sensors from faster SSs at central and posterior sensors. This observation has promoted a model of SS dynamics as two distinct SS generators, a slow/frontal and fast/central-parietal generator(Anderer et al., 2001; Ayoub et al., 2012; Mölle et al., 2011; Timofeev and Chauvette, 2013). However, an alternative explanation is a more unified mechanism for SS generation(Steriade, 2003), comprised of many local generators with overlapping frequency distributions spread across the cortex(Dehghani et al., 2011a, 2010; Frauscher et al., 2015; Gennaro and Ferrara, 2003; Peter-Derex et al., 2012; Piantoni et al., 2017). In this case, average frequency differences between frontal and parietal sites would not be attributable to two generators with uniform frequency, but merely obscure a much larger underlying overlap in frequency and location. In this study, we explicitly estimate the degree of frequency variation between regions, within a region, at individual cortical sites, and within individual SSs, to better adjudicate between these two generating mechanisms.

Frequency variability within SS includes a tendency to slow over the course of the SS. In both MEG and EEG, previous work found power is maximal at higher frequencies (13-15 Hz) earlier on in the SS, especially at central sensors, and maximal at lower frequencies (10-12 Hz) later in the SS, especially at frontal sites(Dehghani et al., 2011a; Zygierewicz et al., 1999). Here we report differences in intra-SS frequency variability across the cortical surface, and show it is quite large compared to differences in frequency due to region. Some studies at the scalp(O' Reilly and Nielsen, 2014; Schönwald et al., 2011; Souza et al., 2016) as well as intracranial (Andrillon et al., 2011)have also reported a systematic decrease or slowing in frequency during a SS. However, likely due to smaller sampling, these results appear inconsistent with one indicating slower SSs slow more(Souza et al., 2016), and another showing no difference in slowing between slower and faster SSs(Andrillon et al., 2011). We provide a more comprehensive study in SS slowing across the cortex, as well as report how more widespread SSs slow more than local SSs.

The active process of sleep contributing to memory consolidation and structuring depends on communication transfer between the hippocampus and cortex(Diekelmann and Born, 2010; Rasch and Born, 2013). Previous work in our lab identified a subset of parietal channels that have large phase-locking value (PLV > 0.4) in SSs, especially during N2, with posterior hippocampal SSs(Jiang et al., 2019a). We observed a cluster of parietal channels with faster frequency and lower SD that appeared to overlap with this group. We characterize the unique SS properties of cortical sites that have strong communication with the hippocampus during SSs.

These findings support updating a model of cortical SS from two SS generators to a model with many generators with varied and overlapping frequency characteristics. We identify different sources of SS variability and their magnitude, as well as how types of SSs defined in various ways have distinct SS properties. These findings could improve our understanding of the mechanisms by which SSs organize memory consolidation during sleep.

Methods

Patient selection

Patients with intractable, pharmaco-resistant epilepsy were implanted with stereoelectroencephalographic (sEEG) electrodes to determine seizure onset for subsequent resection for treatment. Patients were selected from an original group of 54, excluding patients that had pronounced diffuse slowing, widespread interictal discharges, or highly frequent seizures. The selected 20 patients each had at least one hippocampal contact in a hippocampus not involved in seizure initiation. The 20 patients includes 7 males, aged 29.8+/-11.9 years old (range 16-38 years). For demographic and clinical information, see Table 1 (originally published in Jiang 2019a). All electrode implants and duration of recordings were selected for clinical purposes (Gonzalez-Martinez 2013). All patients gave fully informed consent for data usage as monitored by the local Institutional Review Board, in accordance with clinical guidelines and regulations at Cleveland Clinic.

Electrode localization

Electrodes were localized by registering a post-operative CT scan with a pre-operative 3D T1-weighted MRI with ~1mm^3 voxel size (Dykstra et al 2012) using Slicer

(RRID:SCR_005619). This allowed visualization of individual contacts with respect to the HC cross-sectional anatomy, which was interpreted in reference to the atlas of Duvernoy (1988). The posterior limit of the uncal head served as the border for labeling depth contacts as anterior or posterior. Recordings were obtained for 32 HC contacts, 20 anterior (11 left) and 12 posterior (7 left). The CT-visible cortical contacts were identified as previously described (Jiang et al 2019a), to ensure activity recorded by bipolar transcortical pairs is locally generated (Mak-McCully 2015). Electrode contacts were excluded if they were involved in early stages of seizure discharge or had frequent interictal activity. Of the 2844 contacts implanted in the selected 20 patients, 366 transcortical pairs (18.3 +/- 4.7 per patient) were accepted for further analysis. Polarity was adjusted, if necessary, to "pial surface minus white matter" according to MRI localization, confirmed with decreased high gamma power during for surface-negative DSs.

Freesurfer was used to reconstruct pial and white matter surfaces from individual MRI scans and to parcellate the cortical surface into anatomical areas (Desikan 2006). An average surface from all 20 patients was generated to serve as the basis for all 3D maps. While each cortical SEEG electrode contact's location was obtained through direct correlation of CT and MRI as described earlier, we obtained the cortical parcellation labels corresponding to each contact by morphing the right-anterior-superior-oriented anatomical coordinates from individual surfaces to the average surface space (Fischl et al 1999b). For displaying transcortical recordings on a 3D surface, the vertex closest to the midpoint to each contact pair on the native surface was registered to the average surface. As some contact pair markers are located within sulci, all markers were moved to the same plane, for the medial and lateral surfaces separately. This allows visualizing all contact markers while maintaining anatomical fidelity. For a priori statistical analysis of SS characteristics by cortical region, insular transcortical pairs were assigned to temporal cortex, and paracentral, postcentral and precentral labels constituted the Rolandic cortex.

Data processing

Continuous recordings from SEEG depth electrodes were made with a cable telemetry system (JE-120 amplifier with 128 or 256 channels, 0.016-3000 Hz bandpass, Neurofax EEG-1200,Nihon Kohden) across multiple nights (Table 1). Patients were recorded over the course of clinical monitoring for spontaneous seizures, with 1000Hz sampling rate. The total NREM sleep durations vary across patients due to intrinsic variability and sleep deprivation due to clinical environment. We confirmed the percentages of NREM in total sleep from 28 sleeps across 16 of our patients were comparable to (i.e. within 2 SD of) normative data (Moraes et al 2014) in terms of N2 and N3 durations. Furthermore, we did not observe any significant differences in sleep graphoelements (GE) compared with normative data (Jiang 2019a). Recordings were anonymized and converted into the European Data Format. Subsequent data processing was performed in MATLAB (RRID:SCR_001622); the Fieldtrip toolbox (Oostenveld et al 2011) was used for line noise removal and visual inspection. Separation of patient NREM sleep/wake states from intracranial LFP alone was achieved by previously described methods using clustering of first principal components of delta-to-spindle and delta-to-gamma power ratios across multiple LFP-derived signal vectors (Gervasoni et al 2004; Jiang et al 2017), with the addition that separation of N2 and N3 was empirically determined by the proportion of DSs that are also part of slow oscillations (at least 50% for N3) (Silber et al 2007), since isolated DSs in the form of K-complexes are predominantly in N2 (Cash et al 2009).

Cortical graphoelement detection

Spindles were detected as reported in Gonzalez 2018 and Jiang 2019a. For each sleep period, each channel's signal was filtered in 4-8 Hz, 10-16 Hz, and 18-30 Hz bands using a zero-phase shift frequency domain filter. The width of the filter transition bands relative to the cut-off frequencies was 0.3 and a Hanning window was used for the transition. To calculate the power envelope for each narrow band signal, the absolute value of the filtered data was calculated and smoothed by convolution with a 400ms Tukey window. To detect peaks in the power time series, this signal was subsequently smoothed using a 600ms Tukey window and a robust estimate of deviation for each channel was calculated by subtracting the median and dividing by the median absolute deviation. Putative SS peaks were identified as exceeding 3 in the normalized median power time series and a relative edge threshold of 40% of the peak amplitude defined SS onsets and offsets. SS epochs that co-occurred with greater than 3 in theta (4-8 Hz) and beta (18-30 Hz) ranges were excluded. SS detections were performed on each sleep period separately. Only SSs longer than 0.5s were analyzed.

Downstates and theta bursts were detected as previously described(Gonzalez et al., 2018; Jiang et al., 2019b). DSs were detected on each channel as follows: (1)apply a zero-phase shift, eighth order (after forward and reverse filtering) Butterworth IIR bandpass filter from 0.1 to 4 Hz; (2) select consecutive zero crossings within 0.25-3s; and (3) calculate amplitude trough between zero-crossings and retain only the bottom 20% of troughs. Theta bursts were detected as follows: (1) apply a zero-phase shift, eighth order (after forward and reverse filtering) Butterworth IIR bandpass filter from 5 to 8 Hz (range selected to minimize overlap with delta and SS content); (2) calculate the mean of the Hilbert envelope of this signal smoothed with a Gaussian kernel (300ms window; 40ms sigma); (3) detect events with a +/- 3 SD threshold for the peak and identify the start and stop times with a +/-1 SD threshold; (4) only include events with a duration between 400ms and 1s; and (5) for each bandpass filtered peak in a putative burst, calculate the preceding trough-to-peak deflection and only take events that have at least 3 peaks exceeding 25% of the maximum deflection.

Spindle characteristics

Several SS characteristics were estimated for each SS, including: duration, overall frequency, frequency change, SS frequency variability, waveform shape measures (described in *Waveform shape*), power, and amplitude. The duration of the SS is defined as the onset and offset of the SS as reported in *SS detection*. We applied a zero-phase shift, eighth order (after forward and reverse filtering) Butterworth IIR bandpass filter at 10-16Hz and segmented out the detected SSs. As also described in *Waveform shape*, the amplitude and frequency of each cycle was assessed by uniformly cropping SSs such that they start and end with troughs. The narrow-band trough to peak amplitude was then defined as sum of the narrow-band peak amplitude and the preceding trough amplitude. For a cycle to be included, it needed to exceed 30% of the maximum trough-to-peak amplitude. The frequency of a cycle was calculated by dividing the sampling rate by the trough-to-trough period and limiting the frequency precision to two decimal places. The frequency of each SS was calculated as the number of cycles surviving the amplitude threshold divided by the trough-to-trough duration of the SS. Additional frequency metrics per SS calculated included the maximum trough-to-peak narrow band amplitude, to assess frequency variability, the difference of the fastest and slowest cycle in Hz and the standard deviation of cycle frequencies. These measures were only calculated for SSs with at least 4 cycles that exceeded 30% of the maximum amplitude. Frequency change per SS was estimated using least squares. For each SS a design matrix, A, was constructed as an intercept column and a column indicating the time of each peak since the first peak (in seconds) in the narrow-band signal. The frequency at each cycle was coded as the dependent variable, b, and the coefficients for frequency intercept and change was estimated using the MATLAB 'A\b' operation. For each channel this yielded a distribution of frequency slopes and a one-sample t-test was used to assess significant slowing or speeding (in Hz/s). The average of all SS characteristics was calculated for each channel. The NREM, N2, and N3 SS density (SSs per minute) was also calculated.

Waveform shape

Each channel was band-passed filtered in a broadband signal of 1-30Hz via finite-impulse response filters (duration minimum of 3 cycles of 10 Hz). This signal was band-pass filtered in 10-16Hz in order to identify times of zero-crossings. These times were mapped back to the wideband signal, and peaks (troughs) were identified as the maxima (minima) within these zero-crossings. SS epochs started and ended with troughs to set the number of cycles equal to the number of peaks. Because peaks are marked as maxima on the broad-band signal, they could be marked during steep broadband rises or falls and not reflect true SS cycles. To mitigate these effects, two rejection criteria were applied to each cycle within a SS before estimating SS metrics: 1) the cycle deflection (μV), defined as the average of the trough-to-peak and peak-to-trough amplitude, must exceed 30% of the largest cycle deflection within a SS, and 2) the amplitude of the rising (falling) phase must exceed 15% of the amplitude of the falling (rising) phase. Applying an average amplitude threshold (1) removes smaller cycles, and (2) removes cycles that may have large amplitude but fall on steep rises or falls in the broadband signal. This step avoids analyzing cycles that would otherwise have rise-decay-symmetry measures close to 0 or 1. These steps mitigate the influence from overlapping large, slower rhythms on the SS waveform shape features. For a SS to be included in waveform shape analysis, we required a density of 8 good cycles /s, and to estimate mean and standard deviation within a SS, at least 5 good cycles.

Statistical analysis

These SS characteristics were estimated for each SS and subsequently averaged across SSs for each channel and for major cortical lobes. We applied linear-mixed effects models to account for measuring multiple cortical channels within patients. This was calculated in *RStudio* Version 1.4.1103 using the *lme4* package at the channel level as ' lmer(Dependent Var ~ 1 + Independent Fixed Effect + (1|Patient),data)'.For example, when comparing differences across cortical regions, cortical region was the fixed effect of interest, patients coded as random intercepts, and each observation was the average measure for a channel. When evaluating associations between SS characteristics, each observation was a SS and a nested random effect structure was applied as follows:' lmer(Dependent Var ~ 1 + Independent Fixed Effect + (1|Patient/Channel),data)'. When comparing the start versus end of SS features (waveform shape measures, frequency), for each channel, paired-t-test (or Wilcoxon signed rank when data not normally distributed) assessed significant differences across SSs. Descriptive tables and summary results of mixed effects models were created using qwraps2 and sJPlot, respectively.

Widespread spindles

The number of cortical channels spindling was calculated for each sample, and sleep SS epochs were defined as the non-zero onset and offset periods. For each SS, we determined the maximum number of channels spindling within the same epoch as well as the number of channels with minimum 100 ms overlap. We recorded whether a SS was the leading SS within the epoch, as well as its latency to start from the beginning of the epoch. Lastly, we coded whether a SS was in a frontal (parietal) channel and had a SS preceding or following a SS from a frontal or parietal channel within the same SS epoch. We then found the maximum number of channels spindling within this epoch, as well as the greatest number of channels spindling with an overlap of at least 100ms. Both measures were expressed as the number of cortical channels co-spindling and the proportion of cortical channels.

Assessment of hippocampal-cortical phase-locking value

Significant phase-locking between HC and NC SSs was determined as reported in Jiang et al 2019b. This analysis focuses on coupling of NC channels to posterior HC during N2. Briefly, for each NC-HC pair, SSs detected in both structures that overlapped for at least 1 SS cycle (here 160 ms) had phase-locking values (PLVs) (Lachaux et al., 1999) calculated over 3s trials centered on all hippocampal SS event starts in NREM. We also computed PLV for the same NC-HC channel pairs over the same number of trials centered on random times in NREM to create a baseline estimate; and for each non-overlapping 50 ms time bin, a two-sample t test was performed between the actual PLV and the baseline estimate, with the resulting p values undergoing FDR correction. A given channel pair would be considered significantly phase-locking if: (1) > 40 trials were used in the PLV computation; or (2) at least 3 consecutive time bins yield post-FDR p values <0.05.

Results

Data characterization

We analyzed intracranial SEEG recordings from twenty patients with pharmaco-resistant epilepsy, and their full demographic information can be found in our previous work (Jiang 2019a,b). Only cortical channels and sleep periods free of epileptic activity were selected for SS analysis. Information pertinent to this study, including number of bipolar cortical recordings, number of sleep periods, and hours in stage 2 and 3 are shown in Table 2.1. In total we recorded from 366 cortical sites, and as shown in Figure 2.1, we have broad coverage across the cortical surface, including both hemispheres and medial and lateral surfaces. Of these 366 sites, 5 were excluded due to anatomical labels that were assigned “Medial wall” or “unknown” and 4 excluded because they did not have at least 50 SSs with durations greater than 500 ms, resulting in 357 cortical sites analyzed.

Primary SS characteristics

SSs were detected across the cortical surface in both N2 and N3 sleep. The average number of SSs and standard deviation across channels is reported in Table 2.2. The median number of SSs per channel and interquartile ranges for NREM, N2, and N3 are, respectively: 907 (357,2100), 649 (218,1315), and 209 (54,569). SS density, or number of SSs per minute, is shown for NREM sleep stages separately (Table 2.2). Regional differences in SS density were observed, with frontal and parietal sites showing greater densities than temporal, Rolandic, or occipital areas (Figure 2.1 C, Table 2.2). Overall, SS density was greater in N2 than N3 (paired t-test, 0.4 SSs/min, N=357, t=9.2, $p<0.0001$).

The average amplitude and duration of SSs are shown in Table 2.2 as well as Figure 2.1. The average SS amplitude (defined as maximum trough-to-peak deflection) increases anterior-to-posterior and is largest at occipital sites. SSs have the longest duration at frontal, Rolandic, and parietal sites, and shortest at temporal sites. SSs were slightly longer (linear mixed effects model ; $\beta = 0.02$ s, Std. Error= 0.002, No. obs=711, No. ch=357,N=20) in N2 than N3, however there were no significant differences in amplitude (linear mixed effects model; $\beta = 0.29$ µV, Std. Error= 0.21, No. obs=711, No. ch=357,N=20).

The overall frequency of SSs showed a clear increase along the anterior-to-posterior axis, from frontal (12.1 Hz) to Rolandic (12.47 Hz) to parietal (12.74 Hz) regions, though showing intermediate frequencies at occipital sites (12.28 Hz). Temporal cortex showed the lowest overall frequency (11.89 Hz). Additionally, there appears to be a cluster of parietal channels with much greater frequency than surrounding cortex. These sites overlap with a subset of channels our group previously identified (Jiang et al 2019b) as showing strong phase-locking to hippocampal SSs and will be further discussed in section *Cortical-hippocampal SS phase locking*. These regional differences in frequency were present in both N2 and N3. After controlling for regional differences, N2 estimates of overall frequency were higher than N3 (linear mixed effects model; β = 0.08 Hz, Std. Error= 0.01, No. obs=711, No. ch=357,N=20).

Variability in spindle frequency

In addition to SS frequency varying between regions, we found there was substantial variability across sites within a region, and across SSs measured at a single cortical site, as well as within individual SSs. Here we compare the magnitude of these sources of variability with our reported difference between frontal and parietal of 0.64 Hz, which is indicated across the color bars in Figure 2.2.

Frequency variability across sites within a single cortical region

The variation within a region across channels was lower in frontal and temporal cortices, ranging from 0.28 to 0.34 Hz standard deviations respectively, compared to parietal and occipital sites ranging from 0.68 to 0.8 Hz respectively. Notably, the standard deviation across channels within a site is larger at parietal and occipital channels than the average difference between frontal and parietal (0.64 Hz).

Frequency variability at a single cortical site

To assess how much individual sites vary in SS frequency, we calculated the standard deviation in overall frequency across SSs at each cortical channel (Fig 2.2A, Table 2.3). Notably, the average inter-SS frequency standard deviation is 0.87 Hz in NREM across the cortex. Rolandic sites had the largest inter-SS standard deviation (0.98 Hz), whereas frontal sites had the lowest with 0.79 Hz. Notably the amount that SSs vary in frequency at a single site typically exceeds the reported difference between frontal and parietal recordings. After controlling for differences due to brain region, cortical sites showed a greater inter-SS standard deviation in frequency for N2 than N3 (linear mixed effects model; β = -0.05Hz, Std. Error= 0.01, No. obs=711, No. ch=357,N=20).

Frequency variability within a spindle

Next, we evaluated measures of intra-SS variability. This includes the intra-SS standard deviation, the range of SS cycle frequency, and the linear change in frequency. The cortical average intra-SS standard deviation, that is, across SS cycles within a SS, was 0.94Hz during NREM (Figure 2.2 B; Table 2.3). The average frequency range within a SS, calculated as the difference of the fastest and slowest cycle, was 2.7 Hz(Figure 2.2 C; Table 2.3). For both of these measures, the temporal and occipital cortical sites had the largest intra-SS variation, whereas the frontal and parietal had the lowest (Figure 2.2B,C; Table 2.3). While there were no differences between N2 and N3 for intra-SS SD (linear mixed effects model; β =-0.01,Std Error=0.01, No. obs=711, No. ch=357,N=20), the frequency range was smaller in N3 (linear mixed effects model; β =-0.05, Std Error=0.02, No. obs=711, No. ch=357,N=20).

Intra-SS variability is not completely random, but partially reflects a linear change in frequency. Previous work has reported that SSs decrease in frequency during their evolution. However, estimates across the cortex intracranially have not been systematically reported. 45.3%

(163/360) of cortical channels showed SS frequencies that were significantly different at the start versus at the end of the SS (paired-t test, $p<0.05$, FDR adjusted). Of these, 98.2% (160/163) showed slower frequencies at the end versus the start. The three channels that showed faster frequencies at the end of the SS were located in the lateral inferior temporal sulcus and pericalcarine cortex. For more precise estimates of linear change in frequency, we regressed cycle frequency against time since the first cycle for each SS. The average change estimates across SSs for a given cortical site are shown in Figure 2.2 D, and summarized in Table 2.3. Using this approach, we found that 46.4% of cortical sites had a significant linear change in SS frequency (one-sample t-test, $p<0.05$ FDR adjusted), with 95.2% (159/167) showing SS slowing and 4.8% (8) channels showing SS speeding. These 8 speeding channels were in the ventral and lateral temporal (2), pericalcarine (2), superior and inferior parietal (2), insula (1), and superior frontal cortex (1) across 7 different patients. As evident in Figure 2.2 D, the majority of recordings across cortical regions tended to decrease in frequency, and significant slowing in SS frequency occurred across frontal cortex (-0.74 Hz/s). There were no significant differences in rates of change in frequency between N2 and N3 (linear mixed effects model; β =-0.01, Std Error=0.03, No. obs=711, No. ch=357, N=20).

Waveform shape

Brain rhythms are not pure sinusoidal oscillations, and as such there may be non-linear features of the signal which Fourier analysis fails to capture. To test for significant waveform shape features, we assessed the rise-decay and peak-trough symmetries of sleep SSs (Cole). Rise-decay asymmetry indicates a sawtooth shape. Peak-trough asymmetry indicates a rounded peak with sharp trough, or *vice versa*. Progressive changes in these parameters could conceivably result in the frequency change we observed, thus we compared these measures at the start and end of SSs. We found that 16.1% (58/360) of cortical channels showed a significant (paired t-test, p<0.05, FDR adjusted) difference in peak-trough symmetry at the start versus the end of a SS. Of these 58 significant channels, in 41 peak-trough asymmetry increased (p=.0022, binomial test). However, the changes were small, with the average change of peak-trough symmetry in the 17 with decreases being from 0.49 to 0.48, and in the 41 with increases, from 0.52 to 0.53.. For rise-decay symmetry, 24.4% (88/360) of channels showed significant differences at the start versus the end. Of these, 38.6% (54/88) on averaged changed from 0.52 to 0.5 and the other 34 cortical sites on average from 0.49 to 0.5. This indicates as the SS progresses, it becomes slightly more symmetrical in its rise and decay. Overall any significant changes during a SS in waveform shape measures are small, and the differences within and between regions in these measures also appear minor. Thus it is unlikely these measures significantly affect our previous estimates of intra-SS variability.

Spindle co-occurrence

While originally described as a global phenomenon, MEG and intracranial work have identified SSs as primarily local events. We found that indeed the majority of SSs occur at small proportion of cortical sites recorded (Figure 2.3 A,B). Typically, individual SSs occurred in only a single or a few channels, with 50% of SSs occurring in under 16% of channels and 75% in under 25% of channels (on average 18 cortical channels per patient analyzed for SSs). Frontal and parietal sites showed the greatest proportion of multiple channels participating in SS events (Figure 2.3 A). This effect is also clearly shown in Figure 2.3 C where frontal and parietal sites have the smallest proportion of SSs occurring in only a single channel, and the largest proportion of SSs occurring in four or more channels.

We also found that the proportion of SSs that lead or initiated co-spindling events (e.g. SSs detected in multiple channels) was greatest in parietal and especially medial parietal regions (Figure 2.3 D). In contrast, frontal sites showed the lowest proportion of leading SSs in co-SS events. When SSs co-occur in multiple channels, parietal sites tend to lead and frontal sites to follow.

SSs that co-occurred in multiple channels also had unique SS characteristics (Table 2.3). SSs that occurred in successively more channels were faster in frequency (linear mixed effects model with nested random effects for channels within patients; No. observations=554,794, No. Ch= 365, N= 20; for single channel vs 6+ channels β=0.36 Hz,Std. Error=4e-03,t=86.86), longer in duration (for 6+ channels β=0.13 s, Std. Error=1e-03,t=126.86), had lower intra-SS frequency SD (for 6+ channels β=-0.14 Hz, Std. Error=2e-03,t=-72.98), and showed significantly greater SS slowing (for 6+ channels β=-0.45 Hz/s, Std. Error=0.01, t=-31.68).

Relationships between spindle characteristics

We also investigated the relationships between SS frequency and other SS characteristics. We modeled SS frequency as a function of SS duration, intra-SS variation, frequency change, amplitude, and high gamma power using nested linear mixed effects models, with observations at the SS level (Table 2.3). All variables were entered as fixed effects into a full model with nested random effects for channels within patients. Overall, faster SSs were shorter duration (β=-0.17,Std. Error= 5.4e-03,t=-30.99), showed lower intra-SS standard deviation (β=-0.43,Std. Error=2.9e-03,t=-147.9), lower amplitude(β=-6.7e-03,Std. Error=9.3e-05,t=-71.97), greater high gamma power (β=.23,Std. Error=3.7e-03,t=63.3), and showed a small effect for slowing more (β=-3.9e-03,Std. Error=3.9e-04,t=-10.2). We found these results did not change after controlling for differences due to brain region or sleep stage. These findings demonstrate SS frequency covaries with a number of SS metrics across the brain.

Cortical-hippocampal spindle phase locking

In previous work, we found 37% of NC channels showed significant PLV with NC-HC SSs; 67% of these were with posterior hippocampus in N2. The latter channels are shown in Figure 2.4 A, with color indicating the peak PLV during N2 with posterior hippocampal SSs. This PLV of channels with significant HC-NC SS PLV significantly, positively covaried with overall SS frequency (linear mixed effects, No. channels=76, N=12; β =1.44, t=3.1), however after controlling for differences in frequency due to cortical region, this effect size was reduced (β = 0.78, t=1.92; Figure 2.4 B). Intra-SS frequency SD negatively covaried with PLV (β =-0.29, t=-2.69; Figure 2.4 C), and similarly, this effect was reduced after controlling for differences due to cortical region (β =-0.2, -1.96). Neither SS duration, frequency change, or N2 density significantly covaried with PLV with or without controlling for cortical region ($|t|<1.95$). This indicates there is some relationship among cortical channels with significant PLV for frequency and intra-SS frequency variation. By visualizing PLV magnitude along these two dimensions (Figure 2.4 D), channels with PLV under around 0.25 are predominantly fronto-temporal and clustered in one quadrant, while those around 0.5 and greater are predominantly parietal-occipital and are clustered in another. Furthermore, we previously found 5% of all NC channels showed high NC-HC PLV (peak >0.4), 70% of which were in parietal channels. Post-hoc analyses found after controlling for regional differences, this high PLV subset compared to all other cortical channels were significantly faster (linear mixed effects, No. channels= 341, N=20, β =0.57,t=5.01; Figure 2.4 E), had lower intra-SS frequency variation (β =-0.15,t=-4.7; Figure 2.4 F), had slightly longer duration (β =0.02, t=2.82; Figure 2.4 G), and greater SS density in N2 (β =0.76,t=2.48; Figure 2.4 H) as well as in N3 (β =0.72,t=2.25). They did not differ in linear change in frequency (β =-0.1,t=-1.06) , inter-SS frequency standard deviation (β =0.03,t=0.77), or amplitude (β =-2.9,t=-0.75). These results did not change when restricting to just parietal channels (linear mixed effects, No. channels=113, N=18). These sites with large HC-NC SS PLV, predominately parietal sites, have unique SS characteristics compared to other cortical sites after controlling for regional differences.

Slower and faster spindles show similar coupling to DSs

Putative evidence for dichotomizing SSs as slow or fast includes that they have different coupling to SOs, wherein frontal slower SSs precede and faster centro-parietal SSs follow downstates (Mölle et al 2011, however see Gonzalez et al 2018). Here we evaluate whether SS frequency determines whether SSs differentially couple DSs, and if this coupling varies by frontal or parietal sites.

We found that of the 365 channels, 320 had at least 40 SSs starting within +/- 0.5s of cortical DSs. There were significantly more SSs starting 0.5s after DSs compared with starting before (linear mixed effects model; β =180.5 SSs, Std. Error= 28.2, t=6.4, N=20, 320 channels). Of these 320, 208 showed a significant tendency for SSs to start after (77%, 160 channels) or before (23%, 48 channels) DSs (two-sided binomial test; $p<0.05$, FDR-adjusted). The majority of significant sites were from frontal (29%,60 channels) and parietal (31%, 64 channels) cortex. All significant channels from frontal and parietal sites, regardless of preferred latency, are shown in Figure 2.5. To compare differences in coupling of SSs to SOs by frequency, only channels with at least 100 SSs <12 Hz and 100 SSs > 12 Hz within +/- 1s DSs are shown. Gray lines separate patients. Because the sample sizes were unbalanced for some channels, we bootstrapped the SS latency times for both frequency groups over 10,000 iterations, using the size of the smaller group. With each iteration, a histogram of SS start latencies was generated, and the average of these histograms was normalized across bins for each channel. To assess whether there were significant differences between either frequency group and chance, or between the two groups, we fit a linear mixed effects model on the probability of SS starting at each time bin and display error bars of standard error. Both SSs greater than 12 Hz (pink, right subpanels) and less than 12 Hz (black, left subpanels) significantly start following DS troughs, in both frontal and parietal sites. Whether SSs are overall slower or faster, they show similar temporal relationships to DSs.

Discussion

This study uses sEEG recordings to investigate SS characteristics including multiple sources of variability in SS frequency, as well as widespread cortical and cortical-hippocampal features that moderate SS frequency. SSs were detected across the entire cortex sampled, however temporo-occipital sites showed lower density, more variable frequency, and shorter duration SSs. We found fronto-parietal regions had the greatest SS duration, density, and proportion of SSs co-occurring in multiple channels; all indicating these regions would make the largest contribution to EEG sensors. SS frequency variability was assessed as overall frequency across SSs within a channel, the average intra-SS variability across cycles, and linear change in frequency. While we observed faster overall frequency in Rolandic (0.37Hz) and parietal (0.64 Hz) cortex compared to frontal channels, the SD across SSs within a site and across cycles within a SS was larger (0.87 and 0.94 Hz, respectively). We also found SSs that occurred in multiple channels were faster, showed greater SS frequency slowing, and had lower intra-SS variability, thus indicating local SSs differ from more global events. Previous work identified a subset of parietal channels with high phase locking to posterior hippocampal SSs (Jiang et al., 2019a), and here we describe how these cortical sites have SSs with faster frequency, longer duration, lower intra-SS frequency variability, and greater density. These unique characteristics could enable differential coordination of NC-HC or NC-NC SSs. Although our findings are from epileptic patients and not a healthy population, we analyzed a large number of patients (n=20) with unique epileptiform etiologies, had broad cortical coverage, excluded sleep periods or channels with pervasive epileptic activity, and the appearance as well as other SS characteristics such as density were within healthy ranges.

The dichotomy of fast and slow SS types was introduced by Gibbs and Gibbs in 1950 to summarize the clear observation of faster SS centro-parietally and slower SS anteriorly. However, most intracranial work has shown there is a gradient of SS frequencies (Peter-Derex 2012; Frauscher 2015; Piantoni 2017) with an exception being Andrillon et al 2011 who reported a sharp frequency boundary in the medial supplementary motor area. Our findings suggest a modification to the model of SSs as dichotomous slow and fast systems. We found individual locations exhibit both faster and slower SS, and that there is a large overlap in the distribution of frequencies between frontal and parietal sites (Figure 2.1A). We did observe differences in the width of frequency distributions at individual sites, with Rolandic-parietal-occipital being broader than frontal-temporal sites (Figure 2.2 A), which has been previously reported along medial structures (Andrillon et al., 2011). In addition to the variability in SS frequency at individual sites, there is also substantial variability within a SS. We found the fastest and slowest cycles within a SS on average differed by 2.7 Hz (Table 2.3, Figure 2.2 C), with the greatest intra-SS variability in temporal and occipital sites, and the lowest at frontal and parietal (Figure 2.2 B,C). These differences could reflect higher quality (i.e. signal-to-noise) SSs in these structures. Previous work in EEG and MEG reported differences of ~ 1Hz slower SS cycles at the end versus the start of the SS(Dehghani et al., 2011a). Work at the scalp has reported SS decrease in frequency during the SS(O' Reilly and Nielsen, 2014; Souza et al., 2016), one study estimating slower SSs slow more than faster (-0.61 Hz/s vs -0.17 Hz/s), whereas an intracranial study found no difference in SS slowing rates between slower and faster SSs (-0.8Hz/s)(Andrillon et al., 2011). Our intracranial study provided a greater number of channels over both medial and lateral surfaces, unlike Andrillon et al 2011 which only records from medial, allowing us to investigate the spatial extent and degree of change in frequency within a SS. We observed 46% of cortical channels had a significant change in linear frequency, with 95% of those showing SS slowing on average of -0.34 Hz/s, and greatest in frontal cortex at -0.74 Hz/s. We also found a statistically significant but small effect for faster SSs to slow more (Table 2.4). We've shown how different sources of variability, inter-SS, intra-SS expressed as frequency range, and intra-SS slowing vary by cortical region, and in the majority of channels, exceed our 0.64 Hz frontal-parietal difference (Figure 2.2, see black triangles on color bars).

Research with MEG and intracranial recordings has shown SSs are largely local phenomena. In the current study, we found the majority of SSs occurred in only a single or a few channels (Figure 2.3 B). Dehghani et al 2011b and Frauscher et al 2015 described how multiple asynchronous SS generators could underly the large synchronous waves observed with EEG. Specifically, intracranial work found that scalp SSs were associated with asynchronous sigma activity, predominantly at frontoparietal sites(Frauscher et al., 2015), and simultaneous MEG/EEG found SSs detected in both modalities had a 66% increase in the number of MEG sensors involved in the SS, especially over frontal sensors(Dehghani et al., 2011b). Similarly, we found that frontal and parietal sites showed the greatest proportion of SSs occurring in multiple channels (Figure 2.3 A,C). We also replicated the phenomenon, previously observed at the scalp

(Dehghani et al., 2011a; Mölle et al., 2011) and intracranially(Andrillon et al., 2011), that SSs at central-posterior sites precede anterior SSs, by showing that frontal sites are least likely to lead co-spindling events(Figure 2.3 D). Furthermore, we found that SSs occurring across multiple channels exhibit unique SS characteristics, including faster overall frequency, lower within-SS frequency variability, and much greater SS slowing. These differences could reflect stereotyped, global propagation patterns, such as rotating front-temporo-parietal waves(Muller et al., 2016), however these patterns were identified with electrocorticography and our sEEG recordings have more irregular spacing across the cortex.

Communication between hippocampal and cortical brain rhythms such as sharp-wave ripple and SSs could serve as a substrate for restructuring recent experiences into long-term memory. Previous work has shown that ripples in the hippocampus are phase locked to hippocampal SSs(Jiang et al., 2019a; Staresina et al., 2015), and that these hippocampal SSs are in turn coordinated with cortical SSs, especially at parietal sites such as precuneus, angular gyrus, and retrosplenial cortex(Jiang et al., 2019a). We found that these cortical sites also had unique SS characteristics, including faster frequency, lower intra-SS frequency variation, and greater density (Figure 2.4). We speculate sites with these SS characteristics may lend themselves to more optimal NC-pHC coordination, or that establishing strong, synchronized NC-HC coordination alters the local cortical SS characteristics. Regardless of how these correlations arise, given the strong NC-HC communication and the involvement of these structures in recollective experiences(Gilmore et al., 2015; Hoppstädter et al., 2015), it could be that SSs at these cortical sites play a unique role in processing such detailed, episodic information, perhaps by coordinating or driving SS dynamics at other cortical structures.

The assertion that slow SSs precede DSs and fast SSs follow is often cited as evidence for their distinct generating mechanisms(Klinzing et al., 2016; Mölle et al., 2011; Timofeev and Chauvette, 2013). Studies vary in the cutoff frequency that separates slow from fast SSs, as SS peak frequencies varies across subjects(Ujma et al., 2015) and some subjects do not show distinct slow and fast SS peaks in averaged power spectra(Cox et al., 2017; Gennaro and Ferrara, 2003; Mölle et al., 2011; Werth et al., 1997). We chose 12 Hz to demarcate slower and faster SSs, as 12 or 13 Hz is typical(Ayoub et al., 2012; Barakat et al., 2011; Mölle et al., 2011; Schabus et al., 2007). Replicating our previous work(Gonzalez et al., 2018; Mak-McCully et al., 2017), we found that SSs were significantly more likely to start after DSs across all cortical sites. We showed that within the same channels, and in both frontal and parietal sites, that SSs above and below 12 Hz both follow cortical DSs (Figure 2.5). There is also a significant effect for faster SSs to show a greater likelihood of initiating after DSs compared to slower SSs, most apparent in the parietal cortex. Overall, these findings affirm SSs, regardless of frequency, show similar temporal relationships to DSs and are consistent with slower and faster SSs existing along a continuum instead of arising from distinct neurophysiological generators.

A related question as to whether slower or faster SSs have distinct generating mechanisms is whether SS frequency moderates cortical plasticity and memory, and if so, whether it acts as a dichotomous or continuous variable. Some work has asserted that fast but not slow SSs are involved in memory consolidation(Barakat et al., 2011; Mölle et al., 2011). However, these studies conflated SS frequency with sensors (i.e. did not look at faster frontal SSs or slower central-parietal SSs)(Mölle et al., 2011). Additionally, analyses often treated frequency as either slow or fast, whether applying fixed ranges for all subjects(Barakat et al., 2011), or identifying two subject specific frequency bands based on power spectra(Mölle et al.,

2011) instead of applying a more broad (10-16 Hz) SS bandpass and treating frequency as a continuous variable. Such analyses cannot exclude frequency continuously moderates learning and memory, and identifying any differences reinforces the mechanism for two SS types with distinct generators. In our work, we found faster SSs are associated with greater high gamma power modulation (Table 2.3) and are more tightly associated with initiating on the down-to-upstate transition (Figure 2.5). Since SSs occurring on down-to-upstate transitions were shown to permit greater calcium influx to layer 2/3 pyramidal neurons in transgenic mice compared to SSs alone(Niethard et al., 2018), faster SSs could be better suited to facilitate cortical plasticity. Updating the model for SS dynamics as not dichotomous slow and fast but a continuum with regional differences along several SS characteristics can inform mechanisms for how SS contribute to memory restructuring.

References

Anderer, P., Klosch, G., Gruber, G., Trenker, E., Pascual-Marqui, R.D., Zeitlhofer, J., Barbanoj, M.J., Rappelsberger, P., Saletu, B., 2001. Low-resolution brain electromagnetic tomography revealed simultaneously active frontal and parietal sleep spindle sources in the human cortex. Neuroscience 103, 581–592. https://doi.org/S0306452201000288

Andrillon, T., Nir, Y., Staba, R.J., Ferrarelli, F., Cirelli, C., Tononi, G., Fried, I., 2011. Sleep spindles in humans: insights from intracranial EEG and unit recordings. J. Neurosci. 31, 17821–17834.

Ayoub, A., Aumann, D., Hörschelmann, A., Kouchekmanesch, A., Paul, P., Born, J., Marshall, L., 2012. Differential effects on fast and slow spindle activity, and the sleep slow oscillation in humans with carbamazepine and flunarizine to antagonize voltage-dependent Na+ and Ca2+ channel activity. Sleep 36, 905–11. https://doi.org/10.5665/sleep.2722

Barakat, M., Doyon, J., Debas, K., Vandewalle, G., Morin, A., Poirier, G., Martin, N., Lafortune, M., Karni, A., Ungerleider, L.G., Benali, H., Carrier, J., 2011. Fast and slow spindle involvement in the consolidation of a new motor sequence. Behav. Brain Res. 217, 117–121. https://doi.org/10.1016/j.bbr.2010.10.019

Bonjean, M., Baker, T., Lemieux, M., Timofeev, I., Sejnowski, T., Bazhenov, M., 2011. Corticothalamic feedback controls sleep spindle duration in vivo. J. Neurosci. 31, 9124–9134.

Contreras, D., Destexhe, A., Sejnowski, T.J., Steriade, M., 1996. Control of spatiotemporal coherence of a thalamic oscillation by corticothalamic feedback. Science 274, 771–774. https://doi.org/10.1126/science.274.5288.771

Contreras, D., Steriade, M., 1995. Cellular Basis of EEG Slow Rhythms : Corticothalamic Relationships. J. Neurosci. 15, 604–622.

Cox, R., Schapiro, A.C., Manoach, D.S., Stickgold, R., 2017. Individual differences in frequency and topography of slow and fast sleep spindles. Front. Hum. Neurosci. 11, 1–22.

Dehghani, N., Cash, S.S., Halgren, E., 2011a. Topographical frequency dynamics within EEG and MEG sleep spindles. Clin. Neurophysiol. 122, 229–235.

Dehghani, N., Cash, S.S., Halgren, E., 2011b. Emergence of synchronous EEG spindles from asynchronous MEG spindles. Hum. Brain Mapp. 32, 2217–27.

Dehghani, N., Cash, S.S., Rossetti, A.O., Chen, C.C., Halgren, E., 2010. Magnetoencephalography demonstrates multiple asynchronous generators during human sleep spindles. J Neurophysiol 104, 179–188. https://doi.org/10.1152/jn.00198.2010

Diekelmann, S., Born, J., 2010. The memory function of sleep. Nat. Rev. Neurosci. 11, 114–126.

Frauscher, B., von Ellenrieder, N., Dubeau, F., Gotman, J., 2015. Scalp spindles are associated with widespread intracranial activity with unexpectedly low synchrony. Neuroimage 105, 1–12. https://doi.org/10.1016/j.neuroimage.2014.10.048

Gennaro, L. De, Ferrara, M., 2003. Sleep spindles : an overview. Sleep Med. 7, 422–440. https://doi.org/10.1016/S1087-0792(02)00116-8

Gilmore, A.W., Nelson, S.M., McDermott, K.B., 2015. A parietal memory network revealed by multiple MRI methods. Trends Cogn. Sci. 19, 534–543.

https://doi.org/10.1016/j.tics.2015.07.004

Gonzalez, C.E., Mak-McCully, R.A., Rosen, X.B.Q., Cash, X.S.S., Chauvel, P.Y., Rey, M., Halgren, E., 2018. Theta Bursts Precede , and Spindles Follow , Cortical and Thalamic Downstates in Human NREM Sleep. J. Neurosci. 38, 9989–10001.

Hoppstädter, M., Baeuchl, C., Diener, C., Flor, H., Meyer, P., 2015. Simultaneous EEG-fMRI reveals brain networks underlying recognition memory ERP old/new effects. Neuroimage 116, 112–122. https://doi.org/10.1016/j.neuroimage.2015.05.026

Jiang, X.X., Gonzalez-Martinez, J., Halgren, E., 2019a. Posterior Hippocampal Spindle Ripples Co-occur with Neocortical Theta Bursts and Downstates-Upstates , and Phase-Lock with Parietal Spindles during NREM Sleep in Humans 39, 8949–8968.

Jiang, X.X., Gonzalez-Martinez, J., Halgren, E., 2019b. Coordination of Human Hippocampal Sharpwave Ripples during NREM Sleep with Cortical Theta Bursts , Spindles , Downstates , and Upstates 39, 8744–8761.

Klinzing, J.G., Mölle, M., Weber, F., Supp, G., Hipp, J.F., Engel, A.K., Born, J., 2016. Spindle activity phase-locked to sleep slow oscillations. Neuroimage 134, 607–616. https://doi.org/10.1016/j.neuroimage.2016.04.031

Lachaux, J.P., Rodriguez, E., Martinerie, J., Varela, F.J., 1999. Measuring phase synchrony in brain signals. Hum. Brain Mapp. 8, 194–208. https://doi.org/10.1017/S0007680500048066

Mak-McCully, R.A., Rolland, M., Sargsyan, A., Gonzalez, C., Magnin, M., Halgren, E., 2017. Coordination of cortical and thalamic activity during non-REM sleep in humans. Nat. Commun. 8. https://doi.org/10.1038/ncomms15499

Marshall, L., Cross, N., Binder, S., Dang-vu, T.T., 2020. Brain Rhythms During Sleep and Memory Consolidation : Neurobiological Insights. Physiology 35, 4–15. https://doi.org/10.1152/physiol.00004.2019

Mölle, M., Bergmann, T.O., Marshall, L., Born, J., 2011. Fast and slow spindles during the sleep slow oscillation: disparate coalescence and engagement in memory processing. Sleep 34, 1411–21.

Mölle, M., Eschenko, O., Gais, S., Sara, S.J., Born, J., 2009. The influence of learning on sleep slow oscillations and associated spindles and ripples in humans and rats. Eur. J. Neurosci. 29, 1071–1081. https://doi.org/10.1111/j.1460-9568.2009.06654.x

Mölle, M., Marshall, L., Gais, S., Born, J., 2002. Grouping of spindle activity during slow oscillations in human non-rapid eye movement sleep. J. Neurosci. 22, 10941–10947.

Muller, L., Piantoni, G., Koller, D., Cash, S.S., Halgren, E., Sejnowski, T.J., 2016. Rotating waves during human sleep spindles organize global patterns of activity that repeat precisely through the night. Elife 5, 1–16. https://doi.org/10.7554/eLife.17267

Niethard, N., Ngo, H.V. V., Ehrlich, I., Born, J., 2018. Cortical circuit activity underlying sleep slow oscillations and spindles. Proc. Natl. Acad. Sci. U. S. A. 115, E9220–E9229. https://doi.org/10.1073/pnas.1805517115

Niknazar, M., Krishnan, G.P., Bazhenov, M., Mednick, S.C., 2015. Coupling of thalamocortical sleep oscillations are important for memory consolidation in humans. PLoS One 10, 1–14. https://doi.org/10.1371/journal.pone.0144720

O' Reilly, C., Nielsen, T., 2014. Assessing EEG sleep spindle propagation . Part 2 : Experimental characterization. J. Neurosci. Methods 221.

Peter-Derex, L., Comte, J.-C., Mauguière, F., Salin, P.A., 2012. Density and frequency caudo-rostral gradients of sleep spindles recorded in the human cortex. Sleep 35, 69–79.

Piantoni, G., Halgren, E., Cash, S.S., 2017. Spatiotemporal characteristics of sleep spindles depend on cortical location. Neuroimage 146, 236–245.

Rasch, B., Born, J., 2013. About Sleep's Role in Memory. Physiol Rev 93, 681–766. https://doi.org/10.1152/physrev.00032.2012

Schabus, M., Dang-Vu, T.T., Albouy, G., Balteau, E., Boly, M., Carrier, J., Darsaud, A., Degueldre, C., Desseilles, M., Gais, S., Phillips, C., Rauchs, G., Schnakers, C., Sterpenich, V., Vandewalle, G., Luxen, A., Maquet, P., 2007. Hemodynamic cerebral correlates of sleep spindles during human non-rapid eye movement sleep. Proc. Natl. Acad. Sci. U. S. A. 104, 13164–13169. https://doi.org/10.1073/pnas.0703084104

Schönwald, S. V, Carvalho, D.Z., Dellagustin, G., de Santa-Helena, E.L., Günther, G.J., 2011. Quantifying chirp in sleep spindles. J. Neurosci. Methods 197, 158–164. https://doi.org/10.1016/j.jneumeth.2011.01.025

Seibt, J., Richard, C.J., Sigl-glöckner, J., Takahashi, N., Kaplan, D., Doron, G., Limoges, D. De, Bocklisch, C., Larkum, M.E., 2017. Cortical dendritic activity correlates with spindle-rich oscillations during sleep in rodents. Nat. Commun. 8. https://doi.org/10.1038/s41467-017-00735-w

Sejnowski, T.J., Destexhe, a, 2000. Why do we sleep? Brain Res 886, 208–223. https://doi.org/S0006-8993(00)03007-9

Souza, R.T.F. de, Gerhardt, G.J.L., Schönwald, S.V., Rybarczyk-Filho, J.L., Lemke, N., 2016. Synchronization and Propagation of Global Sleep Spindles. PLoS One 11, e0151369. https://doi.org/10.1371/journal.pone.0151369

Staresina, B.P., Ole Bergmann, T., Bonnefond, M., van der Meij, R., Jensen, O., Deuker, L., Elger, C.E., Axmacher, N., Fell, J., 2015. Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. Nat. Neurosci. 18, 1679–1686. https://doi.org/10.1038/nn.4119

Steriade, M., 2003. The CorticoThalamic System in Sleep. Front. Biosci. 878–899.

Steriade, M., Mccormick, D.A., Sejnowski, T.J., 1993. Thalamocortical Oscillations in the Sleeping and Aroused Brain. Science (80-.). 262, 679–685.

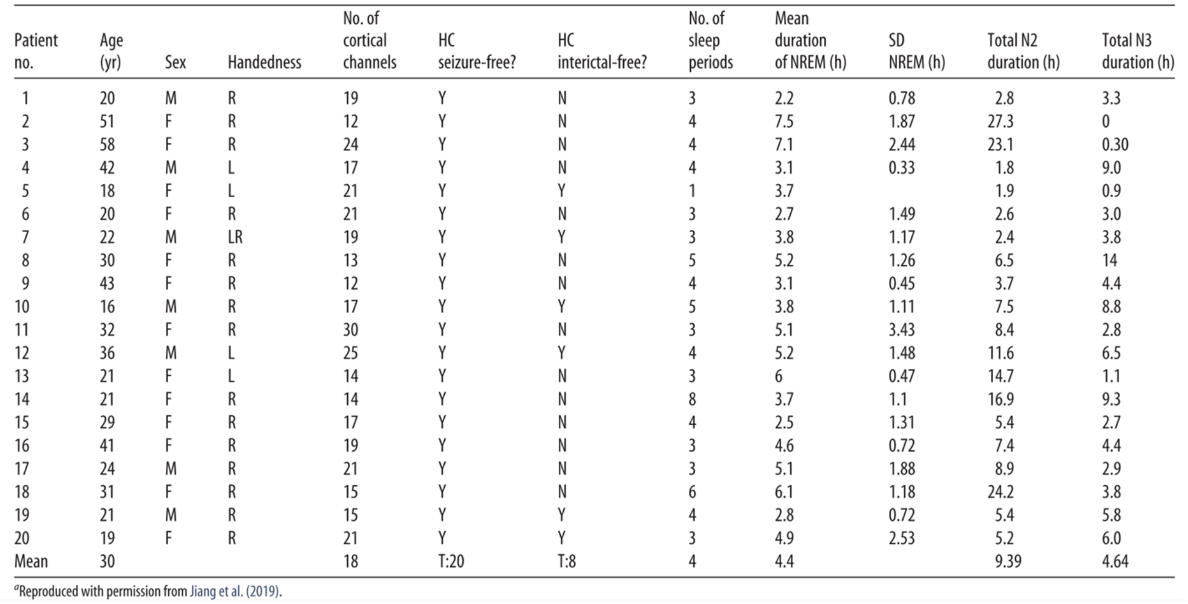
Timofeev, I., Chauvette, S., 2013. The spindles: are they still thalamic? Sleep 36, 825–6.

Ujma, P.P., Gombos, F., Genzel, L., Konrad, B.N., Simor, P., Steiger, A., Dresler, M., Bódizs, R., 2015. A comparison of two sleep spindle detection methods based on all night averages : individually adjusted vs . fixed frequencies. Front. Hum. Neurosci. 9, 1–11. https://doi.org/10.3389/fnhum.2015.00052

Werth, E., Achermann, P., Dijk, D.J., Borbély, A.A., 1997. Spindle frequency activity in the sleep EEG: Individual differences and topographic distribution. Electroencephalogr. Clin. Neurophysiol. 103, 535–542. https://doi.org/10.1016/S0013-4694(97)00070-9

Zygierewicz, J.S., Blinowska, K.J., Durka, P.J., Szelenberger, W., Niemcewicz, S., Androsiuk, W., 1999. High resolution study of sleep spindles. Clin. Neurophysiol. 110, 2136–2147. https://doi.org/10.1016/S1388-2457(99)00175-3

Table 2.1. List of patients, age, sex, handedness, language dominance, HC channel counts, and the lengths of sleep period recordings used.



Patient no.	Age (yr)	Sex	Handedness	No. of cortical channels	HC seizure-free?	HC interictal-free?	No. of sleep periods	Mean duration of NREM (h)	SD NREM (h)	Total N2 duration (h)	Total N3 duration (h)
1	20	M	R	19	Y	N	3	2.2	0.78	2.8	3.3
2	51	F	R	12	Y	N	4	7.5	1.87	27.3	0
3	58	F	R	24	Y	N	4	7.1	2.44	23.1	0.30
4	42	M	L	17	Y	N	4	3.1	0.33	1.8	9.0
5	18	F	L	21	Y	Y	1	3.7		1.9	0.9
6	20	F	R	21	Y	N	3	2.7	1.49	2.6	3.0
7	22	M	LR	19	Y	Y	3	3.8	1.17	2.4	3.8
8	30	F	R	13	Y	N	5	5.2	1.26	6.5	14
9	43	F	R	12	Y	N	4	3.1	0.45	3.7	4.4
10	16	M	R	17	Y	Y	5	3.8	1.11	7.5	8.8
11	32	F	R	30	Y	N	3	5.1	3.43	8.4	2.8
12	36	M	L	25	Y	Y	4	5.2	1.48	11.6	6.5
13	21	F	L	14	Y	N	3	6	0.47	14.7	1.1
14	21	F	R	14	Y	N	8	3.7	1.1	16.9	9.3
15	29	F	R	17	Y	N	4	2.5	1.31	5.4	2.7
16	41	F	R	19	Y	N	3	4.6	0.72	7.4	4.4
17	24	M	R	21	Y	N	3	5.1	1.88	8.9	2.9
18	31	F	R	15	Y	N	6	6.1	1.18	24.2	3.8
19	21	M	R	15	Y	Y	4	2.8	0.72	5.4	5.8
20	19	F	R	21	Y	Y	3	4.9	2.53	5.2	6.0
Mean	30			18	T:20	T:8	4	4.4		9.39	4.64

[a]Reproduced with permission from Jiang et al. (2019).

Table 2.2 Spindle characteristics for regions and non-REM sleep stage. Counts and means (SD) across channels shown.

	Cortex	Frontal	Temporal	Rolandic	Parietal	Occipital
Data						
No. Patients	20	12	19	15	17	14
No. Channels	357	80	101	33	102	41
No. SSs						
NREM	1,531.90 (1,725.98)	1,983.60 (1,342.65)	717.16 (738.15)	1,535.39 (1,018.27)	2,015.47 (2,151.13)	1,451.71 (2,527.84)
N2	1,035.79 (1,340.84)	1,239.12 (949.93)	582.27 (664.69)	1,091.61 (823.01)	1,280.22 (1,520.19)	1,103.22 (2,444.47)
N3	496.11 (707.19)	744.48 (658.88)	134.89 (184.73)	443.79 (494.25)	735.25 (962.72)	348.49 (591.08)
Density (min^{-1})						
NREM	1.64 (1.58)	2.12 (1.45)	0.71 (0.40)	1.73 (1.10)	2.30 (2.03)	1.28 (1.60)
N2	1.82 (1.67)	2.29 (1.50)	0.85 (0.49)	2.03 (1.20)	2.55 (2.14)	1.31 (1.56)
N3	1.43 (1.62)	2.03 (1.53)	0.49 (0.44)	1.30 (1.11)	2.08 (2.16)	1.01 (1.13)
Frequency (Hz)						
NREM	12.28 (0.63)	12.10 (0.34)	11.89 (0.28)	12.47 (0.59)	12.74 (0.68)	12.28 (0.80)
N2	12.30 (0.63)	12.15 (0.36)	11.90 (0.27)	12.54 (0.59)	12.75 (0.69)	12.29 (0.79)
N3	12.22 (0.66)	12.01 (0.34)	11.84 (0.35)	12.28 (0.63)	12.73 (0.74)	12.23 (0.77)
Duration (s)						
NREM	0.70 (0.05)	0.72 (0.05)	0.68 (0.02)	0.72 (0.04)	0.71 (0.06)	0.69 (0.03)
N2	0.71 (0.05)	0.73 (0.05)	0.68 (0.03)	0.72 (0.04)	0.73 (0.07)	0.70 (0.03)
N3	0.69 (0.06)	0.72 (0.04)	0.66 (0.03)	0.72 (0.09)	0.68 (0.06)	0.69 (0.06)
Amplitude (µV)						
NREM	35.31 (18.19)	30.14 (16.23)	30.12 (14.69)	36.90 (17.92)	39.22 (20.24)	47.19 (16.94)
N2	35.35 (18.16)	30.31 (16.05)	30.08 (14.58)	36.73 (17.95)	39.27 (20.27)	47.16 (17.18)
N3	35.61 (18.41)	30.47 (16.49)	30.32 (14.71)	38.69 (18.87)	39.14 (20.52)	47.18 (16.90)

Table 2.3. Spindle frequency variability. Means (SD) across channels shown.

	Cortex	Frontal	Temporal	Rolandic	Parietal	Occipital
Inter-SS SD (Hz)						
NREM	0.87 (0.16)	0.79 (0.13)	0.85 (0.10)	0.98 (0.19)	0.89 (0.19)	0.94 (0.15)
N2	0.88 (0.17)	0.79 (0.14)	0.86 (0.11)	0.99 (0.19)	0.91 (0.19)	0.96 (0.16)
N3	0.83 (0.19)	0.77 (0.13)	0.81 (0.18)	0.90 (0.18)	0.83 (0.22)	0.93 (0.22)
Intra-SS SD (Hz)						
NREM	0.94 (0.17)	0.88 (0.11)	1.05 (0.09)	0.94 (0.13)	0.85 (0.19)	0.99 (0.19)
N2	0.94 (0.17)	0.88 (0.11)	1.05 (0.09)	0.94 (0.13)	0.86 (0.20)	0.99 (0.18)
N3	0.93 (0.18)	0.89 (0.12)	1.03 (0.13)	0.95 (0.14)	0.83 (0.20)	1.00 (0.18)
Intra-SS range (Hz)						
NREM	2.70 (0.44)	2.55 (0.30)	2.96 (0.24)	2.77 (0.37)	2.48 (0.51)	2.85 (0.52)
N2	2.71 (0.43)	2.54 (0.30)	2.98 (0.24)	2.75 (0.37)	2.51 (0.50)	2.84 (0.50)
N3	2.66 (0.49)	2.56 (0.34)	2.89 (0.36)	2.75 (0.42)	2.39 (0.55)	2.87 (0.51)
Frequency change (Hz/s)						
NREM	-0.34 (0.49)	-0.74 (0.66)	-0.14 (0.29)	-0.40 (0.44)	-0.28 (0.40)	-0.17 (0.27)
N2	-0.35 (0.51)	-0.76 (0.68)	-0.13 (0.30)	-0.43 (0.46)	-0.30 (0.39)	-0.18 (0.30)
N3	-0.36 (0.66)	-0.73 (0.72)	-0.19 (0.63)	-0.26 (0.60)	-0.29 (0.59)	-0.29 (0.54)

Table 2.4. Spindle characteristics that covary with overall SS frequency. Results were unchanged after controlling for sleep stage and regional differences. We applied a linear mixed effects model with nested random effects, channels within patients on 550,475 SS observations 360 channels, and 20 patients. Estimates, β ,represent linear slopes of the predictors (rows) on overall SS frequency with 95% confidence intervals (CI). Results were unchanged after controlling for sleep stage and regional differences.

	Frequency (Hz)	
	β	*CI*
Intercept	12.35 ***	12.22 – 12.47
Duration (s)	-0.17 ***	-0.18 – -0.16
Intra-SS SD (Hz)	-0.43 ***	-0.43 – -0.42
Amplitude (μV)	-0.01 ***	-0.01 – -0.01
High Gamma Power (μV^2)	0.23 ***	0.23 – 0.24
Frequency Change (Hz/s)	-3.9e-03 ***	-4.7e-03 – -3.1e-03

* $p<0.05$ ** $p<0.01$ *** $p<0.001$

Table 2.5. Characteristics of widespread spindles. We applied a linear mixed effects model with nested random effects, channels within patients on 554,794 SS observations, 365 cortical channels, and 20 patients. Estimates, β ,represent contrasts in the dependent variable (columns) between 2 or more channels against a single channel. Results were unchanged after controlling for sleep stage and regional differences.

	Frequency (Hz)		**Duration (s)**		**Intra-spindle SD (Hz)**		**Frequency Change (Hz/s)**	
No. Ch	β	*CI*	β	*CI*	β	*CI*	β	*CI*
1	12.17 ***	12.04 – 12.29	0.66 ***	0.65 – 0.67	0.98 ***	0.95 – 1.01	-0.18 **	-0.30 – -0.06
2	0.09 ***	0.08 – 0.09	0.04 ***	0.03 – 0.04	-0.02 ***	-0.03 – -0.02	-0.11 ***	-0.13 – -0.09
3	0.16 ***	0.16 – 0.17	0.06 ***	0.06 – 0.06	-0.05 ***	-0.05 – -0.05	-0.22 ***	-0.24 – -0.19
4,5	0.25 ***	0.24 – 0.26	0.08 ***	0.08 – 0.09	-0.09 ***	-0.09 – -0.09	-0.36 ***	-0.38 – -0.33
6+	0.36 ***	0.35 – 0.37	0.13 ***	0.13 – 0.13	-0.14 ***	-0.14 – -0.14	-0.45 ***	-0.48 – -0.42

* $p<0.05$ ** $p<0.01$ *** $p<0.001$

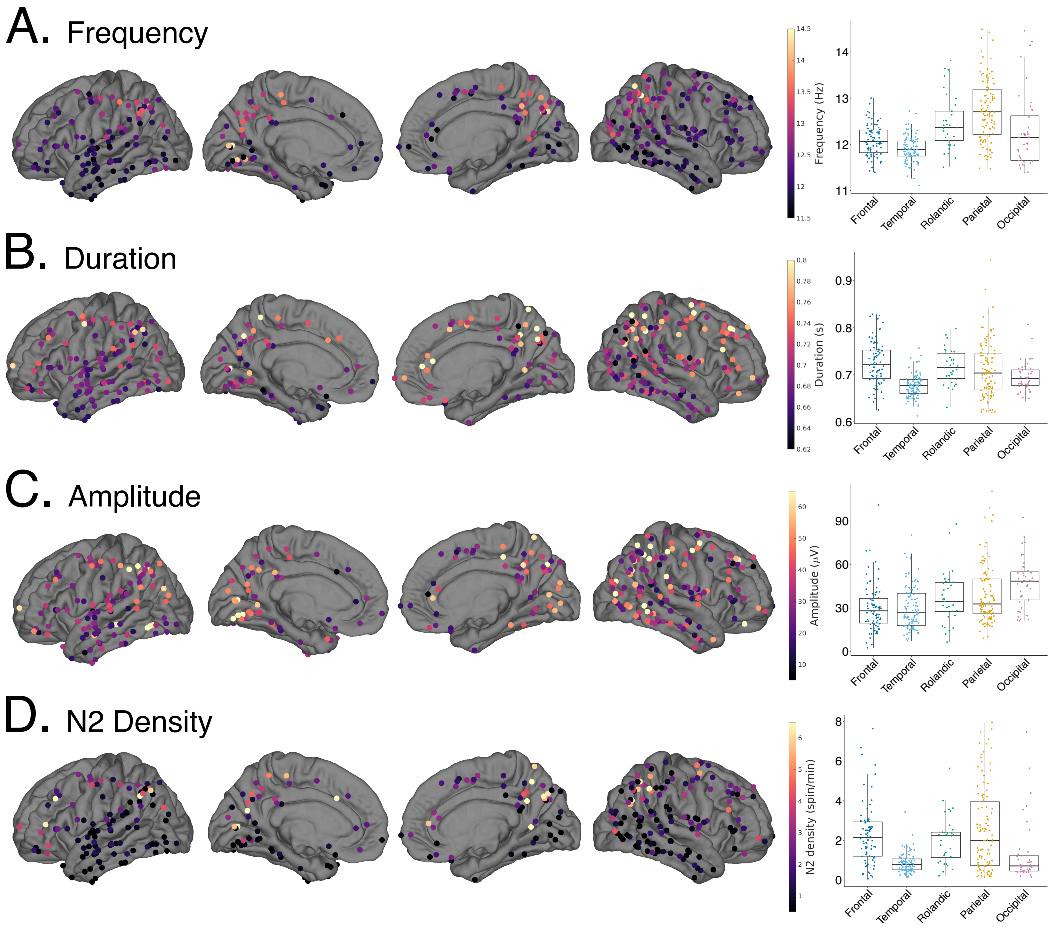


Figure 2.1. Primary spindle characteristics. Cortical bipolar SEEG recordings denoted as circles overlaid on an average surface, with warmer colors indicating greater values. Also indicated to the right of the brain surfaces for each measure are boxplots grouped by brain region. Different colors and columns denote different regions, box margins indicate inter-quartile ranges, and dots indicate individual cortical channels. **A.** Average overall frequency at individual sites is shown. Frequency gradually increases from fronto-temporal to Rolandic to parietal. **B.** Average SS duration across SSs at individual sites is shown. Compared to temporal and occipital, frontal-temporal-Rolandic structures have longer duration. **C.** Average maximal trough-to-peak voltage in a broadband (1-30Hz) increases along the anterior to posterior axis. **D.** Number of SSs per minute during stage 2 (N2), greatest for frontal and parietal sites.

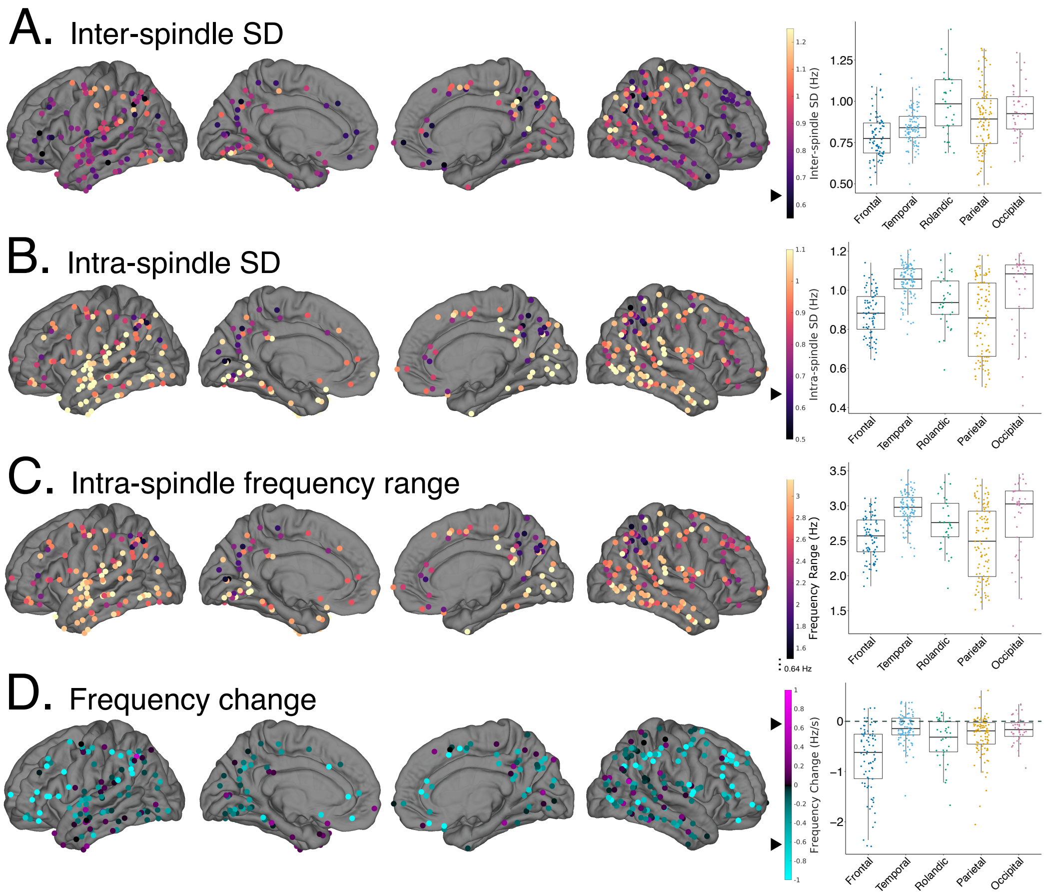


Figure 2.2. Sources of spindle frequency variability. Cortical bipolar SEEG recordings denoted as circles overlaid on an average surface, with warmer colors indicating greater values. Black triangles mark 0.64 Hz (or Hz/s) to indicate the average difference between frontal and parietal sites. Also indicated to the right of the brain surfaces for each measure are boxplots grouped by brain region. **A.** At individual cortical sites, the standard deviation of overall frequency across SSs. Rolandic sites show the greatest variability in frequency across SSs. **B.** The average standard deviation across cycle frequency within a SS. Intra-SS variability is largest for temporal and occipital sites. **C.** Another measure of intra-SS frequency variation, the average difference between the fastest and slowest cycle within a SS. **D.** Average estimated linear change in frequency within a SS. The majority of cortical sites show a decrease in frequency or slowing during a SS (cyan), with greatest slowing occurring at frontal sites. The amount of inter and intra-SS variability often exceeds the observed frontal-parietal overall difference in frequency.

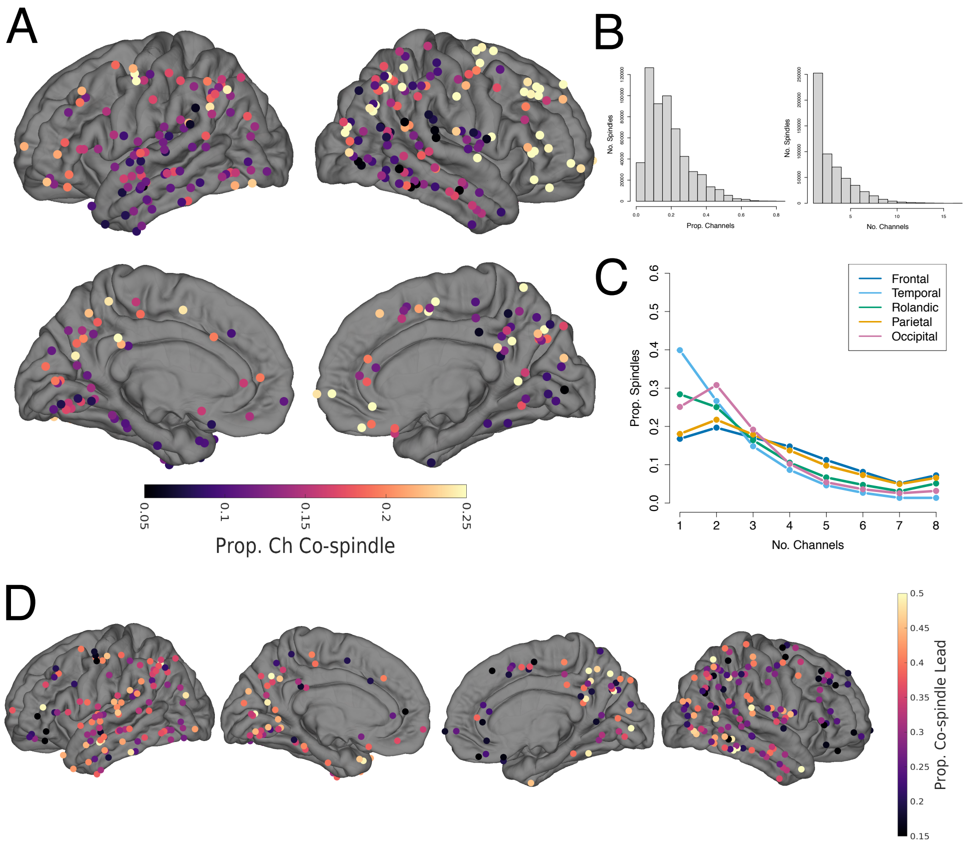
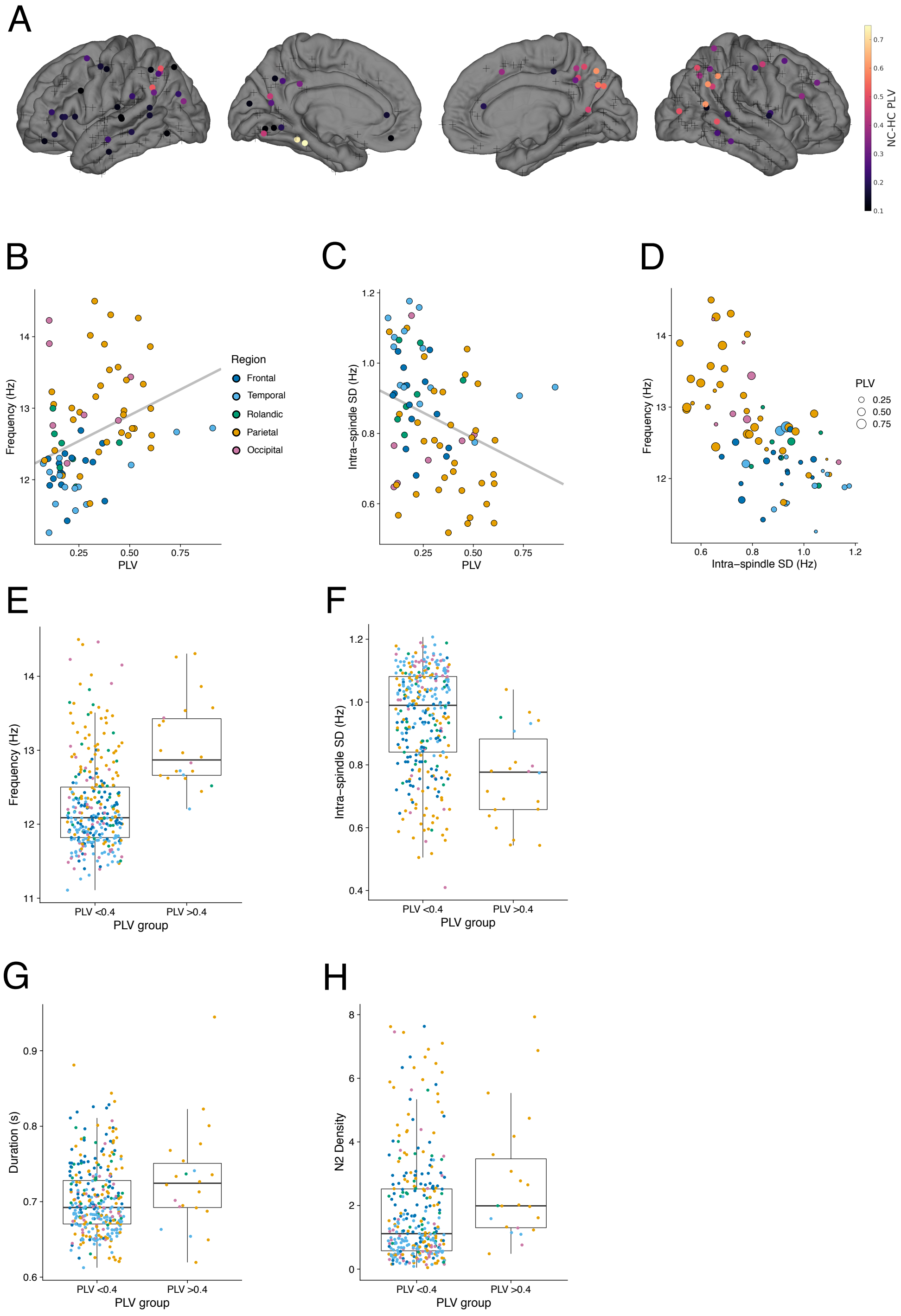


Figure 2.3. Spindle co-occurrences across the cortex. **A,D.** Brain surfaces overlaid with cortical SEEG recording sites. Warmer colors indicate greater values. **A.** The average proportion of channels participating in a SS at each cortical site; frontal and parietal show the greatest proportion. **B.** Distribution of the proportion of channels participating in a SS (left) and absolute number of channels participating in a SS (right) for all SSs from all channels. **C.** For each region, the proportion of SSs that occurred in one to eight channels. **D.** For each channel, the proportion of times a SS initiated a co-spindling event, i.e. when a SS co-occurred in at least two channels.

Figure 2.4. Cortical-hippocampal spindle phase locking. **A.** Channels with significant PLV during SSs with a posterior hippocampal channel during N2, as identified in Jiang et al 2019b, are indicated as circles, with the peak PLV shown in color. Non-significant channels are displayed as crosses. Areas with the highest PLV are apparent in parietal cortex as well as posterior ventral temporal cortex. **B.** Peak NC-HC PLV versus overall NC SS frequency for all significant channels, with colors indicating region. Overlaid in gray is the mixed effects model fit, with patient as random effect, without adjusting for the effect of brain region. After adjusting for brain region, there is only a slight positive association (β = 0.78, t=1.92). **C.** Peak NC-HC PLV versus average NC intra-SS frequency standard deviation. Overlaid in gray is the mixed effects model fit, with patient as random effect, without adjusting for the effect of brain region. After adjusting for brain region, there is only a slight negative association (β =-0.2, -1.96). **D.** For all channels with significant NC-HC PLV, the average SS frequency is plotted against the intra-SS frequency standard deviation, with colors indicating brain region and marker size indicating peak PLV magnitude. There is a clear cluster of parietal channels with high PLV that show faster frequency and lower intra-SS frequency variability. We performed post-hoc analyses on these high PLV channels and SS characteristics. **E-H** compare all cortical channels, including those with non-significant PLV, to the channels with peak PLV exceeding 0.4, with color denoting brain region. High PLV channels were (**E)** significantly faster (linear mixed effects, No. channels= 341, N=20, β =0.57,t=5.01;), (**F)** had lower intra-SS frequency variation (β =-0.15,t=-4.7), **(G)** had slightly longer duration (β =0.02, t=2.82), and **(H)** had greater SS density in N2 (β =0.76,t=2.48). These results did not change when restricting to just parietal channels.


A
NC-HC PLV
0.7
0.6
0.5
0.4
0.3
0.2
0.1
B
Frequency (Hz)
14
13
12
0.25
0.50
0.75
PLV
Region
Frontal
Temporal
Rolandic
Parietal
Occipital
C
Intra−spindle SD (Hz)
1.2
1.0
0.8
0.6
0.25
0.50
0.75
PLV
D
Frequency (Hz)
14
13
12
0.6
0.8
1.0
1.2
Intra−spindle SD (Hz)
PLV
0.25
0.50
0.75
E
Frequency (Hz)
14
13
12
11
PLV <0.4
PLV >0.4
PLV group
F
Intra−spindle SD (Hz)
1.2
1.0
0.8
0.6
0.4
PLV <0.4
PLV >0.4
PLV group
G
Duration (s)
0.9
0.8
0.7
0.6
PLV <0.4
PLV >0.4
PLV group
H
N2 Density
8
6
4
2
0
PLV <0.4
PLV >0.4
PLV group

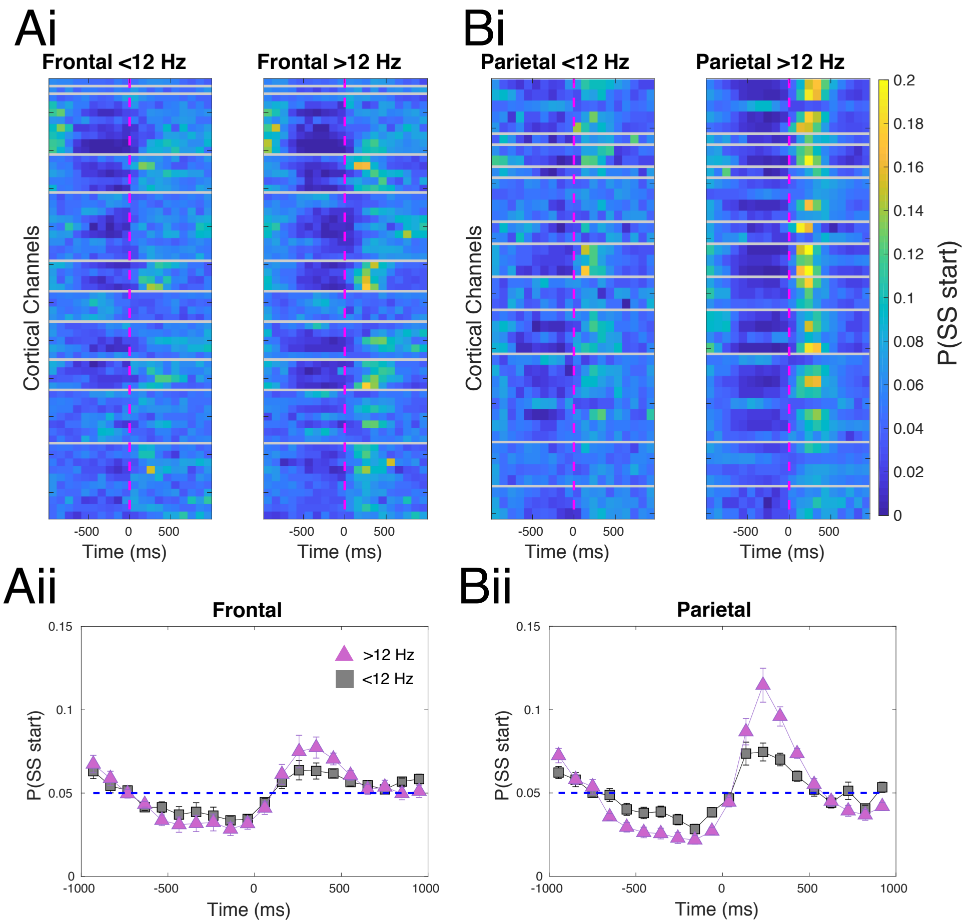


Figure 2.5. Slower and faster spindles show similar coupling to DSs. **A,B** Cortical channels with at least 100 SSs < 12 Hz and 100 SSs > 12 Hz starting within +/- 1s of DS trough are shown. **Ai,Bi** Rows indicate cortical channels, gray lines separate patients, and the dashed magenta line indicates time of the DS trough. Color indicates the probability of SSs starting within that channel. For both regions and frequency groups, there is a greater probability of SSs starting after the DS trough compared to before. This is statistically assessed in **Aii,Bii**, where linear mixed effects models, with patient as random effect, modeled the probability of SSs starting at each time bin for each frequency group separately, with pink triangles showing SSs >12 Hz and gray squares for SSs < 12 Hz. Error bars reflect standard error, and the blue dashed line indicates the probability of SSs starting by chance. Time bins where error bars for a group do not include the chance line indicate significance at $p<0.05$. Both frontal and parietal and frequency groups show a decreased probability of starting before, and an increase after DS trough. Notably, for both **Aii** and **Bii**, there are time bins where SSs > 12 Hz show a greater probability of starting after DS troughs compared to SSs < 12 Hz.

Chapter 2, in full, is currently being prepared for submission for publication. Co-authors include: Gonzalez, Christopher; Jiang, Xi; Rosen, Burke; Gonzalez-Martinez, Jorge; Halgren, Eric. The dissertation author was the primary investigator and author of this paper.

Chaos ARTICLE scitation.org/journal/cha

Causality detection in cortical seizure dynamics using cross-dynamical delay differential analysis

Cite as: Chaos **29**, 101103 (2019); doi: 10.1063/1.5126125
Submitted: 29 August 2019 · Accepted: 17 September 2019 ·
Published Online: 8 October 2019

View Online Export Citation CrossMark

Claudia Lainscsek,[1,2,a)] Christopher E. Gonzalez,[1,3,a)] Aaron L. Sampson,[1,3] Sydney S. Cash,[4]
and Terrence J. Sejnowski[1,2,5]

AFFILIATIONS

[1] Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA
[2] Institute for Neural Computation, University of California San Diego, La Jolla, California 92093, USA
[3] Department of Neurosciences, University of California San Diego, La Jolla, California 92093, USA
[4] Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA
[5] Division of Biological Sciences, University of California San Diego, La Jolla, California 92093, USA

[a)] **Contributions:** C. Lainscsek and C. E. Gonzalez contributed equally to this work.

ABSTRACT

Most natural systems, including the brain, are highly nonlinear and complex, and determining information flow among the components that make up these dynamic systems is challenging. One such example is identifying abnormal causal interactions among different brain areas that give rise to epileptic activities. Here, we introduce cross-dynamical delay differential analysis, an extension of delay differential analysis, as a tool to establish causal relationships from time series signals. Our method can infer causality from short time series signals as well as in the presence of noise. Furthermore, we can determine the onset of generalized synchronization directly from time series data, without having to consult the underlying equations. We first validate our method on simulated datasets from coupled dynamical systems and apply the method to intracranial electroencephalography data obtained from epilepsy patients to better characterize large-scale information flow during epilepsy.

Establishing and detecting information flow in a complex system with many interacting components is challenging yet crucial for many scientific fields. Currently available methods for inferring directionality and causality from observed time series signals are often limited by the number of data required and computational processing time. In this study, we present an efficient and accurate method for measuring causal interaction. We demonstrate that our method, cross-dynamical delay differential analysis (CD-DDA), can estimate the magnitude and direction of information flow in simulated datasets accurately. Furthermore, we applied our method to brain signals obtained from epilepsy patients to characterize previously unidentified seizure-related information flow.

I. INTRODUCTION

Determination of causality and direction of information flow is fundamental to various fields of science, from neuroscience to climate research. In neuroscience, for example, information flow and the nonlinear dynamical causal architecture in brain electroencephalography (EEG) data are important for understanding and predicting events (e.g., seizures). Various regions in the brain might be causally connected even if the data from those regions do not show strong correlations. This might be due to the spatiotemporal nonstationarity of the system.

As pointed out by Yule in 1926,[36] correlation does not imply causation. Yule also made a connection between the introduction of delays and causal relations between time series. In 1969, Granger[10] introduced a statistical measure of causality that is widely used in signal processing. This work is closely related to the work of Wiener, which was published in 1956.[35] Since Granger causality (GC) relies on linear autoregressive models, it may not yield good results for some nonlinear systems. To circumvent the limitations of the linear Granger causality test, Brock *et al.*[8,9] proposed a test based on correlation integrals[11] and Baek and Brock,[3] Hiemstra and Jones,[12] and Bai *et al.*[4,5] and then introduced nonlinear Granger causality. In 2000, Schreiber[32] introduced transfer entropy for information transfer between nonlinear dynamical systems. If the systems are linear Gaussian processes, GC and transfer entropy are equivalent.[6] Determining causality from dynamical attractors of nonlinear dynamical systems and the concept of generalized synchronization (GS) were introduced by Schiff *et al.*,[31] Arnhold *et al.*,[2] Hirata *et al.*,[13,14] and Sugihara *et al.*,[33] among others. One implementation is Sugihara's convergent cross-mapping (CCM).[33] CCM is based on standard uniform delay embeddings and is, therefore, limited to a subset of the dynamical systems found in nature. In 2018, a focus issue in Chaos was published that summarizes recent developments for causality detection.[7] Recent articles used complex network theory to characterize causality of multivariate data.[28,37]

Here, we determine causality through nonlinear, nonuniform, functional embeddings as an extension of delay differential analysis (DDA) and call this method *cross-dynamical DDA* (CD-DDA). We first test CD-DDA on simulated data from coupled dynamical systems and then apply this analysis to investigate spatiotemporal information flow in the brain before and during seizures.

This paper is organized as follows: Sec. II introduces classical DDA and the extension to CD-DDA to study causality. In Sec. III, causal interactions of coupled Lorenz and Rössler systems are investigated. In Sec. IV, CD-DDA is applied to epileptic seizures. Section V is the conclusion.

II. DETECTING CAUSALITY USING DELAY DIFFERENTIAL ANALYSIS (DDA)

DDA combines differential embeddings with linear and nonlinear nonuniform functional delay embeddings[22,30,34] to relate the current derivatives of a system to the current and past values of the system variables.[15,18] Inspired by Max Planck's "natural units,"[25] the DDA model maps experimental data onto a set of natural embedding coordinates.

The general nonlinear DDA model is

$$\dot{u} = \sum_{i=1}^{I} a_i \prod_{n=1}^{N} u_{\tau_n}^{m_{n,i}} + \rho_u = \mathcal{F}_u + \rho_u \tag{1}$$

for $\tau_n, m_{n,i} \in \mathbb{N}_0$, where N is the number of delays (usually 2), I is the number of terms (typically around 3), and $u_{\tau_n} = u(t - \tau_n)$, relating the signal derivative $\dot{u}$ to the signal nonuniformly shifted in time. We then use the coefficients a_i and the least square error ρ_u as features. Note that we explicitly added ρ_u to highlight its use in the causal DDA measure introduced below. To restrict complexity of the DDA model, most of the terms in Eq. (1) are set to zero. We, therefore, consider here DDA models with two delays τ_n, three terms, and a degree $\sum_i m_{n,i} \le 4$ of nonlinearity.

To put DDA in context, a general nonlinear, real-valued function can be expressed in a Taylor series expansion of functionals of increasing complexity around some fixed point. When the function $F(\star)$ represents the behavior of a dynamical system, that is, a time series model where the input is formed from past inputs [$u(t)$, $u(t-\tau_1), \ldots$], the expansion becomes a Volterra series. We have

$$\begin{aligned}\dot{u} = \mathbf{u}_0 &+ \sum_{i=0}^{\infty} g_i \mathbf{u}_{\tau_i} + \sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty} g_{i_1,i_2}\mathbf{u}_{\tau_{i_1}}\mathbf{u}_{\tau_{i_2}} \\ &+ \cdots \\ &+ \sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\cdots\sum_{i_q=0}^{\infty} g_{i_1,i_2,\ldots,i_q}\mathbf{u}_{\tau_{i_1}}\mathbf{u}_{\tau_{i_2}}\cdots\mathbf{u}_{\tau_q},\end{aligned} \tag{2}$$

with the linear and nonlinear data components modeled as separate model terms. To find a model that is a projection onto a stable manifold, we consider low-order models composed of a finite number of leading terms in Eq. (2), such as Eq. (1). This makes CD-DDA different to GC since we use (1) a derivative on the left side of the equation instead of the time series itself in GC, (2) nonuniform functional embeddings on the right side of the equation that were selected from the data instead of uniform linear embeddings of lags 1,2,... in GC, (3) two delays instead of a chain of around 10–15 delays in GC, and (4) three term models that are selected to fit each time series instead of around 15 general terms.

To be more explicit, let us give an example. A typical DDA model for the analysis of EEG and intracranial EEG (iEEG) data is the three term model [all a_i in Eq. (1) except three are set to zero]:

$$\dot{u} = a_1 u_1 + a_2 u_2 + a_3 u_1^4 + \rho_u\,, \tag{3}$$

where $u(t)$ is the time series from one EEG channel and $u_i = u(t - \tau_i)$. The derivative on the left side is computed using a 5-point center derivative algorithm.[21] The coefficients a_i are estimated with numerical singular value decomposition (SVD) to minimize the least square error.[26] The DDA model that best fits the overall dynamical properties of the system [e.g., (3) for EEG data] can be found by supervised (maximizing the classification performance) or unsupervised (minimizing the least square error ρ) structure selection from a list of candidate models (see, e.g., Refs. 18 and 17).

DDA is a nonlinear data analysis framework that (1) uses unprocessed data so as not to disturb the nonlinear properties of the data, (2) uses sparse models that match the macroscopic architecture of the underlying dynamical system, (3) disregards amplitude information to concentrate on the dynamical aspects of the data, and (4) can be extended for detection of dynamical causality in the data to understand information flow in the system. Previously, it has been demonstrated that DDA captures essential features of data to produce exceptional classification performance.

Let us consider two dynamical systems from which the time series $u(t)$ and $v(t)$ are measured, respectively. In order to consider whether there is a causal interaction between these two time series, we start by looking for a model

$$\dot{u} = \mathcal{F}_u + \rho_u, \tag{4}$$

minimizing the error for the time series u. We also minimize the error for $v(t)$ and get

$$\dot{v} = \mathcal{F}_v + \rho_v. \tag{5}$$

If we assume, for example, a unidirectional causal influence of $u(t)$ on $v(t)$,

$$\dot{u} = \mathcal{F}_u + \mathcal{F}_v + \rho_{uv}, \tag{6}$$

the errors ρ_u and ρ_{uv} should be similar since the coefficients of $\mathcal{F}_v$ are irrelevant. In contrast, for

$$\dot{v} = \mathcal{F}_v + \mathcal{F}_u + \rho_{vu}, \tag{7}$$

the errors ρ_v and ρ_{vu} are different.

We then define $\mathcal{C}_{uv}$ as our CD-DDA measure of the causal influence of $v(t)$ on $u(t)$ and $\mathcal{C}_{vu}$ for a causal influence of $u(t)$ on $v(t)$,

$$\mathcal{C}_{uv} = |\rho_u - \rho_{uv}|; \quad \mathcal{C}_{vu} = |\rho_v - \rho_{vu}|. \tag{8}$$

We additionally use the significance $\frac{\mu}{\sigma}$, where μ is the mean and σ is the standard deviation, of the estimated coefficients of the terms $\mathcal{F}_v$ in Eq. (6) and $\mathcal{F}_u$ in Eq. (7) to estimate the synchronization level, $\mathcal{S}$.

III. APPLICATIONS TO SIMULATED DATA

To test CD-DDA, we generate simulated data from coupled Rössler[27] and Lorenz[20] systems. The first example was previously studied in Ref. 24 and is an autonomous, chaotic Rössler system driving a periodic Rössler system. Our second simulated example is a Rössler system driving a Lorenz system.[19,23,24] For the coupled Rössler system, the functions $\mathcal{F}_u$ and $\mathcal{F}_v$ are the same. For the coupled Rössler and Lorenz system, we need to select individual functions $\mathcal{F}_u$ and $\mathcal{F}_v$ since these two systems are dynamically different.

We compared our measure with linear Granger causality,[10] transfer entropy or conditional mutual information,[32] and convergent cross-mapping[33] for a chaotic Rössler system driving a periodic Rössler system and for a Rössler system driving a Lorenz system. Our measure performed as well or better than the other three measures for both of these systems, and we report these findings in Figs. 2 and 4 in the supplementary material.

A. Chaotic Rössler system driving a periodic Rössler system

In this example, we unidirectionally couple two Rössler systems, where the driving system $\boldsymbol{R_1}$ is chaotic and the driven system $\boldsymbol{R_2}$ is periodic[24]

$$\begin{aligned}
&\boldsymbol{R_1}\begin{cases}\dot{x}_1 = -\omega_1 y_1 - z_1,\\ \dot{y}_1 = x_1 + a\, y_1,\\ \dot{z}_1 = b + z_1(x_1 - c),\end{cases}\\
&\boldsymbol{R_2}\begin{cases}\dot{x}_2 = -\omega_2 y_2 - z_2 + \epsilon(x_1 - x_2),\\ \dot{y}_2 = x_2 + a\, y_2,\\ \dot{z}_2 = b + z_2(x_2 - c),\end{cases}
\end{aligned} \tag{9}$$

with $a = 0.15$, $\omega_1^2 = 1.015$, $\omega_2^2 = 0.985$, $b = 0.2$, and $c = 10$. We modified the system in Ref. 24 to the system above as explained in the Appendix. The coupling strength ϵ was varied between 0 and 0.25 using 30 000 linearly spaced steps. The integration step size was set to 0.05 with a transient of 10^5 time points discarded. The remaining data were downsampled by a factor of 2 resulting in 10^5 data points for each coupling strength.

We selected the DDA models by minimizing the model error ρ_u from $u = x_1$ or $u = x_2$ and get

$$\mathcal{F}_u = a_1\, u_1 + a_2\, u_2 + a_3\, u_1^3, \tag{10}$$

with $u_j = u(t - \tau_j)$, $\tau_1 = 32\,\delta t$, $\tau_2 = 9\,\delta t$, and $\delta t = 0.025$.

We calculated $\mathcal{C}_{uv}$ and $\mathcal{C}_{vu}$ for $u = x_1$ and $v = x_2$ in sliding windows (window duration $3000\,\delta t$ or 50 pseudoperiods, window shift $1000\,\delta t$) and show these results in Fig. 1 (middle panel).

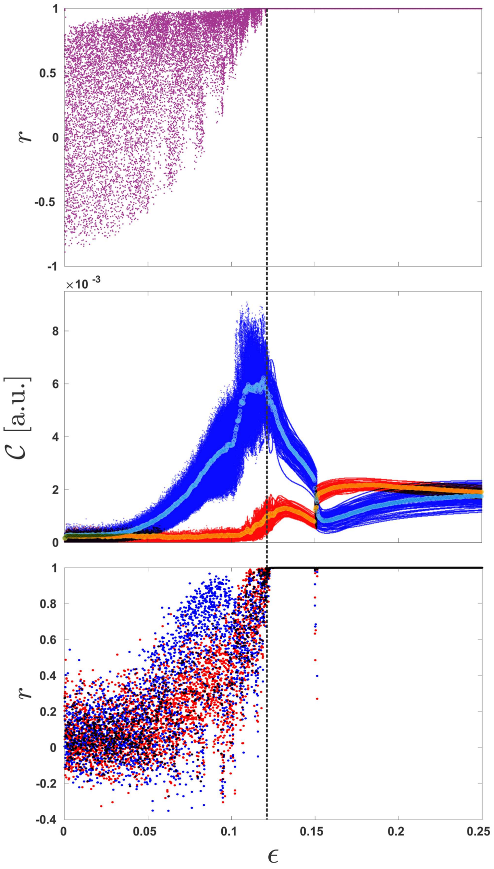


FIG. 1. Pearson correlation r between response system components x_2 (system $\boldsymbol{R_2}$) and x_3 (system $\boldsymbol{R_3}$) for determining the onset of synchronization (dashed line) between $\boldsymbol{R_1}$ and $\boldsymbol{R_2}$ (upper panel). Cross-dynamical measures for a driven periodic Rössler system by a chaotic Rössler system across coupling strength, ϵ. For each coupling strength, 97 time windows are plotted as dots. Blue dots indicate chaotic driving periodic ($\mathcal{C}_{vu}$), and red dots periodic driving chaotic ($\mathcal{C}_{uv}$) (middle panel). The average across time windows for each coupling strength is overlaid in cyan and orange. Pearson correlation between $\mathcal{C}_{uv}$ across time windows for ϵ_k and ϵ_{k+1} in red and for $\mathcal{C}_{vu}$ in blue. Both values converge to one after the onset of generalized synchronization.

Each coupling strength displays $\mathcal{C}_{vu}$ for all time windows in the direction of $u = x_1$ to $v = x_2$ (blue dots) and $\mathcal{C}_{uv}$ in the direction of v to u to (red dots). $\mathcal{C}_{vu}$ increases with increasing coupling strength while $\mathcal{C}_{uv}$ remains close to zero and begins to increase around the

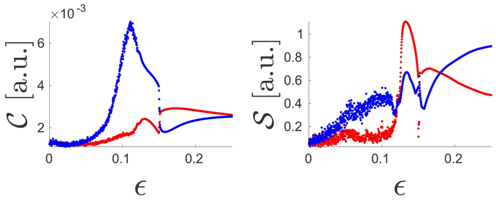


FIG. 2. Effect of short time windows: the left plot shows the CD-DDA causality measures $\mathcal{C}_{uv}$ and $\mathcal{C}_{vu}$ ($u = x_1$, $v = x_2$) for system (9) for a window length of 300 δt, or 5 pseudoperiods. This plot can be compared to the middle panel in Fig. 1. The right plots show our synchronization measure or significances $\mathcal{S}_{uv}$ and $\mathcal{S}_{vu}$ for the maximum of the additional three DDA parameters.

onset of generalized synchronization (GS), estimated as ϵ=0.12 in previous work based on conditional Lyapunov exponents[24] and based on the auxiliary systems approach[1] as shown in Fig. 1, upper panel. GS occurs in nonidentical, unidirectionally coupled oscillators when there is a map between trajectories on the driving attractor and trajectories on the response attractor.[29] The auxiliary systems approach requires introducing a second response system $\boldsymbol{R_3}$ [with the variables (x_3, y_3, z_3)] that is identical to $\boldsymbol{R_2}$, except for initial conditions. To determine the onset of synchronization between $\boldsymbol{R_1}$ and $\boldsymbol{R_2}$, we see when the two driven response systems undergo complete synchronization or when the Pearson correlation between x_2 and x_3 is 1 (Fig. 1, upper panel). We can also estimate the onset of generalized synchronization by correlating $\mathcal{C}_{uv}$ for ϵ_k with ϵ_{k+1} across all time windows, and doing the same for $\mathcal{C}_{vu}$ (Fig. 1, lower panel). After the onset of synchronization, neighboring ϵ values share the same synchronized manifold, whereas before, neighboring ϵ values follow distinct driven attractors.

In Fig. 2 (left panel), we investigate the effect of short time windows (see also Fig. S1 in the supplementary material). While the main effect of $\mathcal{C}_{vu}$ remains for even just five pseudoperiods, we also see $\mathcal{C}_{uv}$ unexpectedly increase before the onset of generalized synchronization.

In Fig. 2 (right panel), we show the onset of generalized synchronization in a purely data driven manner: first, we estimate the significances for the three additional, cross DDA parameters of $\mathcal{F}_v$ and $\mathcal{F}_u$ in Eqs. (6) and (7). Then, we recalculate the significance measures for two slightly different window lengths ($300 \pm 20\,\delta t$). When the two time series u and v are not synchronized, trajectories on the driving and driven attractors are unrelated, and different window lengths will yield similar significance measures. If the time series are synchronized, then there is a map between trajectories on the driving and driven attractors, and measures of significance will be sensitive to window length. To assess synchronization, we take the maximum of the three DDA parameters that shows the greatest difference across window lengths. We call this our measure of synchronization, $\mathcal{S}$.

We are able to recover the correct direction of causal inference for coupling between 0.05 and 0.15 for the longer (Fig. 1, middle panel) and shorter (Fig. 2, left panel) data windows, as well as after adding 20 dB noise (see Fig. S1 in the supplementary material).

B. Rössler system driving Lorenz system

In this example, we unidirectionally drove a Lorenz system $\boldsymbol{L}$ with a Rössler system $\boldsymbol{R}$[23]

$$
\begin{aligned}
&\boldsymbol{R}\begin{cases}\dot{x}_1 = -\alpha(y_1 + z_1),\\ \dot{y}_1 = \alpha(x_1 + ay_1),\\ \dot{z}_1 = \alpha(b + z_1(x_1 + c)),\end{cases}\\
&\boldsymbol{L}\begin{cases}\dot{x}_2 = \sigma(-x_2 + y_2),\\ \dot{y}_2 = Rx_2 - y_2 - x_2 z_2 + \epsilon y_1^2,\\ \dot{z}_2 = x_2 y_2 + \beta z_2,\end{cases}
\end{aligned}
\tag{11}
$$

where $a = b = 0.2$, $c = -5.7$, $\sigma = 10$, $R = 28$, and $\beta = -\frac{8}{3}$.

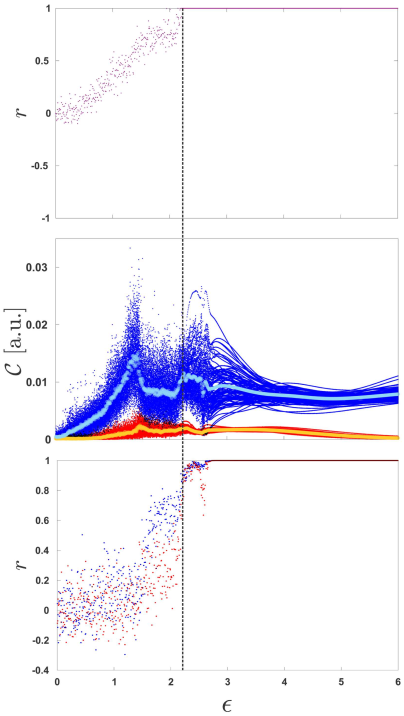


FIG. 3. Pearson correlation r between response system components y_2 (system $\boldsymbol{L}$) and y_3 (system $\boldsymbol{L_2}$) for determining onset of synchronization between $\boldsymbol{R}$ and $\boldsymbol{L}$. Cross-dynamical measures for a driven chaotic Lorenz system by a chaotic Rössler system across coupling strength, ϵ. For each coupling strength, $\mathcal{C}$ for 97 time windows are plotted as dots. Blue dots indicate $u = y_1$ driving $v = y_2$ ($\mathcal{C}_{vu}$), and red dots denote v driving u ($\mathcal{C}_{uv}$).The average across time windows for each coupling strength is overlaid in cyan and orange.

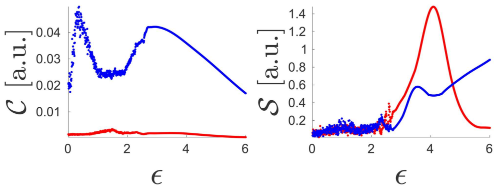


FIG. 4. Effect of short time windows: the left plot shows the CD-DDA causality measures $\mathcal{C}_{uv}$ and $\mathcal{C}_{vu}$ ($u = y_1$, $v = y_2$) for system (11). This plot can be compared to the middle panel in Fig. 3. The right plots show the significances $\mathcal{S}_{uv}$ and $\mathcal{S}_{vu}$ for the maximum of the additional three DDA parameters.

α adjusts the timescale of the Rössler system to match that of the Lorenz and is set to 6. We varied the coupling strength, ϵ, between 0 and 6 over 1000 linearly spaced values. We implemented an integration step size of 0.01, removed a transient of length 10^5, and evaluated 10^5 data points for each coupling strength. As with the coupled Rössler system, we introduce an identical copy of the response system $\boldsymbol{L}$, $\boldsymbol{L_2}$ (with components x_3,y_3, and z_3), which is also driven by $\boldsymbol{R}$ and only differs from $\boldsymbol{L}$ in initial conditions. We determined $\mathcal{F}_u$ and $\mathcal{F}_v$ for $u = y_1$ and $v = y_2$ from Eq. (12) by fitting all three-term, two-delay DDA models up to order 3 nonlinearity, scanning delays between 6 and 60 δt, and selecting the model and delays with the lowest error. This exhaustive model search was performed separately for u and v with no coupling, and is then applied to u and v for all subsequent coupling strengths

$$\begin{aligned} \mathcal{F}_u &= a_1\, u_1 + a_2\, u_2 + a_3\, u_1^2, \\ \mathcal{F}_v &= b_1\, v_1 + b_2\, v_1 v_2 + b_3\, v_1 v_2^2, \end{aligned} \tag{12}$$

where $\tau_{u,1} = 7\,\delta t$, $\tau_{u,2} = 6\,\delta t$, $\tau_{v,1} = 6\,\delta t$, and $\tau_{v,1} = 22\,\delta t$.

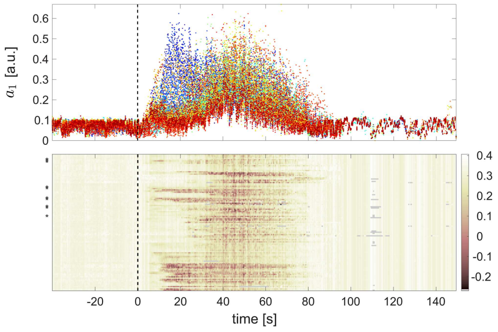


FIG. 5. Classical DDA: coefficient a_1 for 113 channels for Patient 1 in Ref. 17. In the upper panel, a_1 for each channel is shown in a different color. The lower panel shows the values of a_1 as color across all 113 channels (rows) to identify the onset channels. The onset channels marked by the neurologist are indicated by stars on the y-axis. The dotted line marks the seizure onset from the neurologist.

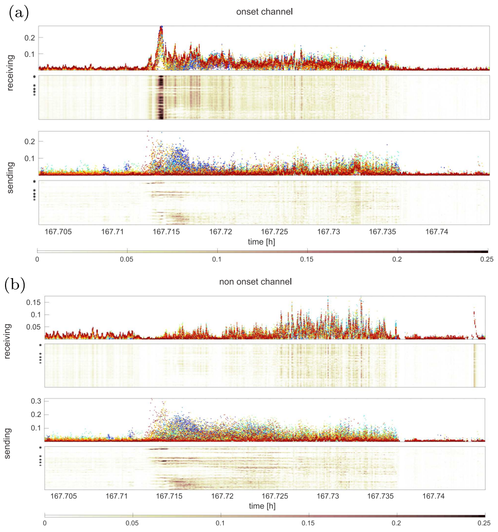


FIG. 6. CD-DDA: in the upper panels of (a) and (b), $\mathcal{C}_{vu}$ for all channels u to one channel v [onset channel in (a) and nononset channel in (b)] is shown and in the lower panels, $\mathcal{C}_{uv}$ for one channel v [onset channel in (a) and nononset channel in (b)] to all other channels u is shown. The onset channels marked by the neurologist are indicated by stars on the y-axis.

Once the models are selected, we evaluate Eq. (8) in a sliding window (window duration 3000 δt or 30 pseudoperiods, window shift 1000 δt) for each coupling strength, ϵ. In Fig. 3, each coupling strength displays $\mathcal{C}$ for all time windows in the direction of $u = y_1$ to $v = y_2$ ($\mathcal{C}_{vu}$, blue dots) and v to u ($\mathcal{C}_{uv}$, red dots). Our approach shows $\mathcal{C}_{vu}$ increases as the coupling strength increases, whereas $\mathcal{C}_{uv}$ slightly increases before the onset of GS for this system, which previous work estimates around $\epsilon = 2$,[23] and we have estimated to be $\epsilon = 2.2$ using the auxiliary systems approach. We also show in Fig. 3 lower panel the onset of synchronization by correlating $\mathcal{C}$ across time windows between ϵ_k and ϵ_{k+1}. We can also recover the correct direction of causal inference using significantly less data, that is, using 300 δt instead of 3000 δt, shown in the left panel of Fig. 4. Furthermore, we can also estimate the onset of synchronization with $\mathcal{S}_{uv}$ and $\mathcal{S}_{vu}$, as shown in the right panel of Fig. 4. Notably, $\mathcal{S}_{uv}$ in red goes up near the estimated onset of synchronization determined in Fig. 3. Overall, $\mathcal{C}$ determines the correct direction of causal inference for no noise and 20dB noise (see Fig. S3 in the supplementary material).

IV. DETECTING CAUSALITY IN EPILEPTIC SEIZURES

In Ref. 17, a genetic algorithm (GA) was used to select the model with minimum error from one second data segments for one hour periods centered on the seizure onset times. Around one million such

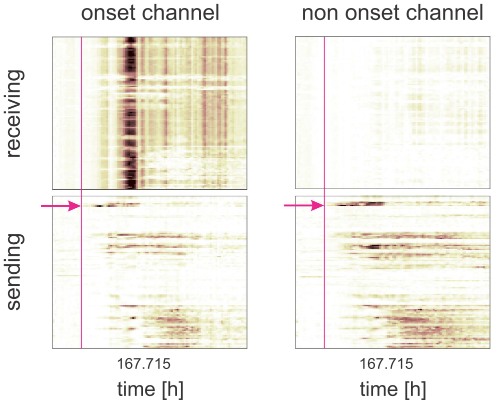


FIG. 7. CD-DDA: enlarged region around the seizure in Fig. 6.

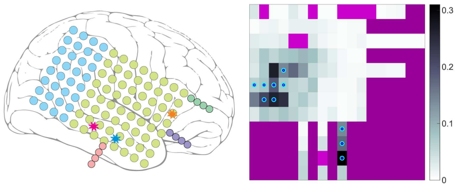


FIG. 8. The left panel shows the channel locations (for more information, see Patient 1 in Ref. 17) and the right panel shows a grid of these locations. The boxes in grayscale are the $\mathcal{C}_{uv}$ values or information sent by the onset channel to all other channels at the time of the highest peak in the upper panels in Fig. 6(a), at around 10 s after the seizure onset marked by the neurologist. The circles indicate the onset channels determined by the neurologist. The darker magenta boxes are background and the lighter magenta boxes indicate bad channels.

data segments (155 seizures, 730 iEEG channels from 13 patients) were analyzed in this way. The patient demographics and characteristics are described in Ref. 17 The DDA model selected in Ref. 17 for the characterization of epileptic seizures is

$$\dot{u} = a_1 u_1 + a_2 u_2 + a_3 u_1^4 + \rho_u = \mathcal{F}_u + \rho_u, \qquad (13)$$

with $u_i = u(t - \tau_i)$.

Model (13) bifurcates at the seizure onset, as shown in Ref. 17: it has a low error after seizure onset, but not before. This DDA model will be used in the remainder of this section.

In Ref. 17, eight delays were chosen to characterize and segment the data using truncated higher order SVD. In this paper, we only use the delay pair $\boldsymbol{\tau} = (7, 10)\,\delta t$, where $\delta t = \frac{1}{f_s}$ with the sampling rate $f_s = 500$ Hz since this delay pair proved to be sufficient for the characterization of epileptic seizures.

In Fig. 5, the coefficient a_1 from Eq. (13) is shown for 113 channels. The onset channels are marked with stars on the y-axis. As discussed in Ref. 17, these plots can be used to localize the seizure onset region in the brain.

We can extend this analysis to CD-DDA to determine the information flow patterns during seizures. $\mathcal{F}_u$ and $\mathcal{F}_v$ are the same model as Eq. (13) since all data are iEEG data. $u(t)$ and $v(t)$ are recordings from two different channels. Each channel is paired with all other channels and the information received and sent is computed for each pair.

The same seizure in Fig. 5 is shown in Fig. 6, where we show information received from all other channels (upper panels) and information sent to all other channels (lower panels) for an example onset channel (a) and for an example nononset channel (b). Figure 7 zooms into the time around seizure onset. While both channels send information to another onset channel marked with a magenta arrow just before the seizure starts, only the onset channel subsequently receives information from the majority of channels at the time of the seizure. This patient has one evoked seizure, five fully developed spontaneous seizures (first of these seizures is shown in Figs. 5–7),

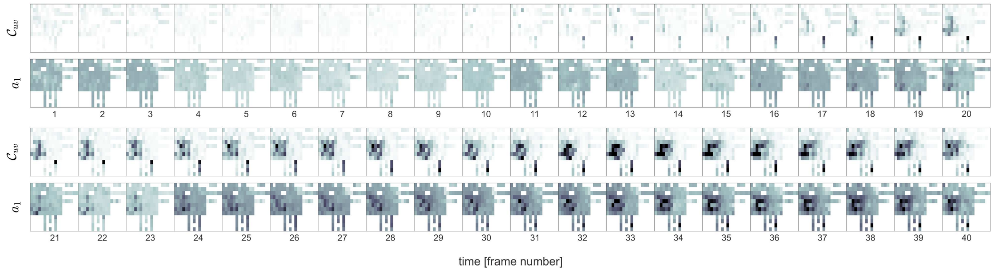


FIG. 9. Time line of the CD-DDA feature $\mathcal{C}_{uv}$ from Fig. 6(a) lower panels and the classical DDA feature a_1 from Fig. 5. The window length for each box is a quarter of a second with a window shift of half that length. The features $\mathcal{C}_{uv}$ and a_1 are mapped onto the same grid and color axis as in Fig. 8 with a white background.

and one nonfully developed seizure. In the supplementary material, we show $\mathcal{C}$ for all seizures as well as for stimulations. All spontaneous seizures except the nonfully developed one show the same signature, where the onset channel receives a large response from a majority of channels once the seizure starts. The last seizure, seizure 6, is different because the response from the nononset channels is too weak to keep the seizure developing.

In Fig. 8, the CD-DDA feature $\mathcal{C}_{uv}$ from Fig. 6 at the peak in the upper panels (around 10 s after marked onset) is mapped onto a grid (right panel) that was derived from the channel locations in the clinical report (left panel). Details and exact channel locations can be found in Ref. 17 (Patient 1). The circles denote the clinically marked onset channels. It is obvious that the region around the onset channels is sending most of the information (see also Fig. 7, magenta arrows). To investigate that further, we plotted such a grid for each sliding data window, where the window length is a quarter of a second and the window shift is half that length. Figure 9 shows such a time line for the CD-DDA feature $\mathcal{C}_{uv}$ from Fig. 6 and the classical DDA feature a_1 from Fig. 5. CD-DDA not only shows the start of the seizure earlier but also shows a constant information flow from the onset regions. Classical DDA, on the other hand, shows the generalization of the seizure and the involvement of the whole brain after seizure generalization.

V. CONCLUSION

We have developed a new tool for studying causality called CD-DDA and tested it on simulated data from dynamical systems and then further applied this technique to epileptic seizure data. Using CD-DDA, we can recover the direction of causal interaction between unidirectionally coupled systems with small parameter mismatch as well as nonequivalent systems, even with short time series and in the presence of noise (see supplementary material). Furthermore, CD-DDA can detect the onset of generalized synchronization (GS).

We have also shown that CD-DDA provides a useful measure of information flow in the brains of human patients experiencing seizures. By computing $\mathcal{C}$ for both directions for each channel, we can identify the channels that seem to be sending out information around the time of seizure onset. These channels match up well to the clinically determined onset channels. Future work will explore further applications of CD-DDA to brain data.

SUPPLEMENTARY MATERIAL

In the supplementary material, we present CD-DDA results for the simulated data from the Rössler and Lorenz systems in Sec. III for different window lengths and added white noise with a signal-to-noise ratio of SNR = 20 dB. We further show for the epilepsy patient of Sec. IV the CD-DDA outputs for 160 hours of data and plots around each of the seven seizures (from 30 s before to 2 min after seizure onset) as well as a plot for a stimulation.

ACKNOWLEDGMENTS

This work was supported by the NIH/NIBIB (Grant No. R01EB026899-01), NIH/NINDS (Grant No. R01NS104368), ONR (Grant No. N0014-15-1-2328), NIH/NINDS (Grant No. NS062092), NIH/NINDS (Grant No. NS088568), and the National Science Foundation (NSF) GRFP. The authors would like to thank Robert Kim for valuable discussions. They would also like to thank Christophe Letellier for helpful discussions and insight into working with coupled systems.

APPENDIX: EQUIVALENT RÖSSLER SYSTEMS

To remove the confusion of interpreting ω as a time scaling factor, we changed the Rössler system in Ref. 24 from

$$\begin{aligned} \dot{x} &= -\Omega\, y - z, \\ \dot{y} &= \Omega\, x + a\, y, \\ \dot{z} &= b + z_(x - c) \end{aligned} \tag{A1}$$

to

$$\begin{aligned} \dot{x} &= -\omega\, y - z, \\ \dot{y} &= x + a\, y, \\ \dot{z} &= b + z(x - c), \end{aligned} \tag{A2}$$

with $\omega = \Omega^2$. Both systems are exactly the same up to a scaling $y \to \Omega y$ of the y-component in the second system. As shown in Ref. 16, ω is one of the four possible bifurcation parameters of the Rössler system[27] since it can be written as[16]

$$\begin{aligned} \dot{x} &= -\frac{1}{B}\, y - \frac{1}{K}\, \frac{1}{C}\, z, \\ \dot{y} &= B\, x + a\, y, \\ \dot{z} &= C\, b - c\, z + K\, x\, z. \end{aligned} \tag{A3}$$

REFERENCES

[1] H. Abarbanel, N. Rulkov, and M. Sushchik, "Generalized synchronization of chaos: The auxiliary system approach," Phys. Rev. E **53**, 4528–4535 (1996).

[2] J. Arnhold, P. Grassberger, K. Lehnertz, and C. Elger, "A robust method for detecting interdependences: Application to intracranially recorded EEG," Physica D **134**(4), 419–430 (1999).

[3] E. Baek and W. Brock, "A general test for nonlinear Granger causality: Bivariate model," *Working Paper* (Korea Development Institute, University of Wisconsin-Madison, 1992), Vol. 15, pp. 197–235.

[4] Z. Bai, Y. Hui, D. Jiang, Z. Lv, W.-K. Wong, and S. Zheng, "A new test of multivariate nonlinear causality," PLoS One **13**(1), e0185155 (2018).

[5] Z. Bai, W.-K. Wong, and B. Zhang, "Multivariate linear and nonlinear causality tests," Math. Comput. Simul. **81**(1), 5–17 (2010).

[6] L. Barnett, A. B. Barrett, and A. K. Seth, "Granger causality and transfer entropy are equivalent for Gaussian variables," Phys. Rev. Lett. **103**, 238701 (2009).

[7] E. M. Bollt, J. Sun, and J. Runge, "Introduction to focus issue: Causation inference and information flow in dynamical systems: Theory and applications," Chaos **28**(7), 075201 (2018).

[8] W. Brock, W. Dechect, and J. Scheinkman, "A test for independence based on the correlation dimension," *Working Paper* (Department of Economics, 1989).

[9] W. Brock, W. Dechect, and J. Scheinkman, "A test for independence based on the correlation dimension," Econom. Rev. **15**, 197–235 (1996).

[10] C. W. Granger, "Investigating causal relations by econometric models and cross-spectral methods," Econometrica J. Econometric Soc. **37**, 424–438 (1969).

[11] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," Physica D **9**, 189 (1983).

[12] C. Hiemstra and J. D. Jones, "Testing for linear and nonlinear Granger causality in the stock price-volume relation," J. Finance **49**(5), 1639–1664 (1994).

[13] Y. Hirata and K. Aihara, Identifying hidden common causes from bivariate time series: A method using recurrence plots," Phys. Rev. E **81**, 016203 (2010).

[14] Y. Hirata, J. Amigó, Y. Matsuzaka, R. Yokota, H. Mushiake, and K. Aihara, "Detecting causality by combined use of multiple methods: Climate and brain examples," PLoS One **11**, e0158572 (2016).

[15] M. Kremliovsky and J. Kadtke, "Using delay differential equations as dynamical classifiers," AIP Conf. Proc. **411**, 57 (1997).
[16] C. Lainscsek, "Nonuniqueness of global modeling and time scaling," Phys. Rev. E **84**, 046205 (2011).
[17] C. Lainscsek, J. Weyhenmeyer, S. S. Cash, and T. J. Sejnowski, "Delay differential analysis of seizures in multichannel electrocorticography data," Neural Comput. **29**(12), 3181–3218 (2017).
[18] C. Lainscsek, J. Weyhenmeyer, M. Hernandez, H. Poizner, and T. Sejnowski, "Non-linear dynamical classification of short time series of the Rössler system in high noise regimes," Front. Neurol. **4**, 182 (2013).
[19] M. Le Van Quyen, J. Martinerie, C. Adam, and F. J. Varela, "Nonlinear analyses of interictal EEG map the brain interdependences in human focal epilepsy," Physica D **127**, 250–266 (1999).
[20] E. N. Lorenz, "Deterministic nonperiodic flow," J. Atmos. Sci. **20**, 130–141 (1963).
[21] E. Miletics and G. Molnárka, "Taylor series method with numerical derivatives for initial value problems," J. Comp. Methods Sci. Eng. **4**(1–2), 105–114 (2004).
[22] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a time series," Phys. Rev. Lett. **45**, 712 (1980).
[23] M. Palus, V. Komarek, Z. Hrncier, and K. Sterbova, "Synchronization as adjustment of information rates: Detection from bivariate time series," Phys. Rev. E **63**, 046211 (2001).
[24] M. Palus and M. Vejmelka, "Directionality of coupling from bivariate time series: How to avoid false causalities and missed connections," Phys. Rev. E **75**, 056211 (2007).
[25] M. Planck, "Über irreversible Strahlungsvorgänge," Ann. Phys. **306**(1), 69–122 (1900).
[26] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes in C* (Cambridge University Press, New York, NY, 1990).
[27] O. E. Rössler, "An equation for continuous chaos," Phys. Lett. A **57**, 397 (1976).
[28] Y. Ruan, R. V. Donner, S. Guan, and Y. Zou, "Ordinal partition transition network based complexity measures for inferring coupling direction and delay from time series," Chaos **29**(4), 043111 (2019).
[29] N. Rulkov, M. Sushchik, L. Tsimring, and H. Abarbanel, "Generalized synchronization of chaos in directionally coupled chaotic systems," Phys. Rev. E **51**, 980–994 (1995).
[30] T. Sauer, J. A. Yorke, and M. Casdagli, "Embedology," J. Stat. Phys. **65**, 579 (1991).
[31] S. J. Schiff, P. So, T. Chang, R. E. Burke, and T. Sauer, "Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble," Phys. Rev. E **54**, 6708–6724 (1996).
[32] T. Schreiber, "Measuring information transfer," Phys. Rev. Lett. **85**, 461–464 (2000).
[33] G. Sugihara, R. May, H. Ye, C.-H. Hsieh, E. Deyle, M. Fogarty, and S. Munch, "Detecting causality in complex ecosystems," Science **338**(6106), 496–500 (2012).
[34] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, Lecture Notes in Mathematics, edited by D. A. Rand and L.-S. Young (Springer, Berlin, 1981), Vol. 898, pp. 366–381.
[35] N. Wiener, "The theory of prediction," in *Modern Mathematics for Engineers* (McGraw Hill, New York, 1956), pp. 165–190.
[36] G. U. Yule, "Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time-series," J. R. Stat. Soc. **89**(1), 1–63 (1926).
[37] Y. Zou, R. V. Donner, N. Marwan, J. F. Donges, and J. Kurths, "Complex network approaches to nonlinear time series analysis," Phys. Rep. **787**, 1–97 (2019).

Chapter 3, in full, is a reprint of the material as it appears in Chaos, titled “Causality detection in cortical seizure dynamics using cross-dynamical delay differential analysis”,2019; 29. Lainscsek, Claudia; Gonzalez, Christopher; Sampson, A; Cash, Sydney; Sejnowski, Terrence. The dissertation author was one of the primary investigators and authors of this paper.

CHAPTER 4: Assessing observability of chaotic systems

Chaos ARTICLE scitation.org/journal/cha

Assessing observability of chaotic systems using Delay Differential Analysis

Cite as: Chaos **30**, 103113 (2020); doi: 10.1063/5.0015533
Submitted: 28 May 2020 · Accepted: 24 September 2020 ·
Published Online: 13 October 2020

View Online Export Citation CrossMark

Christopher E. Gonzalez,[1,2,a)] Claudia Lainscsek,[1,3,a)] Terrence J. Sejnowski,[1,3,4] and Christophe Letellier[5]

AFFILIATIONS

[1] Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA
[2] Neurosciences Graduate Program, University of California San Diego, La Jolla, California 92093, USA
[3] Institute for Neural Computation, University of California San Diego, La Jolla, California 92093, USA
[4] Division of Biological Sciences, University of California San Diego, La Jolla, California 92093, USA
[5] CORIA, Rouen Normandie Université, Campus Universitaire du Madrillet, F-76800 Saint-Etienne du Rouvray, France

Note: This paper is part of the Focus Issue, Chaos: From Theory to Applications.

[a)] **Authors to whom correspondence should be addressed:** cegonzalez@ucsd.edu and claudia@salk.edu

ABSTRACT

Observability can determine which recorded variables of a given system are optimal for discriminating its different states. Quantifying observability requires knowledge of the equations governing the dynamics. These equations are often unknown when experimental data are considered. Consequently, we propose an approach for numerically assessing observability using Delay Differential Analysis (DDA). Given a time series, DDA uses a delay differential equation for approximating the measured data. The lower the least squares error between the predicted and recorded data, the higher the observability. We thus rank the variables of several chaotic systems according to their corresponding least square error to assess observability. The performance of our approach is evaluated by comparison with the ranking provided by the symbolic observability coefficients as well as with two other data-based approaches using reservoir computing and singular value decomposition of the reconstructed space. We investigate the robustness of our approach against noise contamination.

A popular approach for studying nonlinear dynamical systems from a recorded time series is to reconstruct the original system using delay or derivative coordinates. It is known that the choice of the measured variable can affect the quality of attractor reconstruction. Unlike in linear systems for which the state space is observable or not from the measurements, nonlinear systems are more or less observable from measurements depending on the state space location. Moreover, the observability strongly depends on the measured variables. It is, therefore, useful to assess the observability provided by a variable using a real number within the unit interval between two extreme values: 0 for nonobservable and 1 for full observability. Analytical techniques for determining observability require knowledge of the underlying equations, which are typically unknown when an experimental system is investigated. This is often the case for social and biological networks. It is thus of primary importance to assess observability directly from recorded time series. In this paper, we show how Delay Differential Analysis (DDA) can assess observability from time series. The performance of this approach is evaluated by comparing our results obtained for simulated chaotic systems with the symbolic observability coefficients obtained from the governing equations.

I. INTRODUCTION

Studying dynamical systems from real world data can be difficult as they are often high-dimensional and nonlinear; moreover, it is typically not possible to measure all the variables spanning the associated state space.[1–7] In theory, it is possible to reconstruct the non-measured variables by using delay or differential embeddings from a single measurement.[8] However, when performing state-space reconstruction, the dimension required to obtain a diffeomorphical equivalence—required for correctly distinguishing the different states of the system—with the original state space may depend on the measured variable(s).[9] Indeed, a d-dimensional system can be optimally reconstructed from a given variable with a d-dimensional embedding, but a higher-dimensional space may be required when another variable is measured. For instance, the Rössler attractor is easily reproduced with a three-dimensional global model from variable y, but a four-dimensional model[9] or a quite sophisticated procedure[10] is needed when variable z is measured. It was shown that data analysis often (if not always) depends on the observability provided by the measured variable.[11–13]

In the 1960s, the concept of observability and its mathematical dual, controllability, was introduced by Rudolf Kálmán in control theory for linear systems.[14] These concepts were later extended for nonlinear systems in the 1970s from the perspective of differential geometry.[15] Observability assesses whether different states of the original system can be distinguished from the measured variable. A system is said to be fully observable from some measurement if the rank of the observability matrix is equal to the dimension of the system.[16,17] With such an approach, the answer is either fully observable or nonobservable. This approach is sufficient for linear systems because the observability matrix does not depend on the location in the state space.

This is not true for nonlinear systems, and observability coefficients were introduced to overcome this discrepancy answer.[9,18] Observability coefficients are real numbers within the unit interval between two extreme values: 0 for nonobservable and 1 for fully observable. These coefficients are estimated at every point of the trajectory produced by the governing equations in the state space and then averaged along that trajectory.[9,18] It is also possible to construct symbolic observability coefficients from the Jacobian matrix of the system studied.[19,20] In this way, observability takes a graded value according to the probability with which the attractor intersects the singular observability manifold,[21] that is, the subset of the original space for which the determinant of the observability matrix is zero. The great advantage of these coefficients is that they allow comparing the observability provided by variables from different systems and they can be computed for high-dimensional systems.[7] It is then possible to rank the variables according to the observability of the original state space they provide. The dependency of the observability on the measured variable is due to the way variables are coupled in the original system.[22] Symmetries are often sources of difficulty for assessing observability, particularly because reconstructing the original symmetry is not possible from a single variable if the symmetry differs from an inversion.[23,24]

The weakness of these analytical approaches is that the governing equations must be known and it is not possible to assess observability from experimental data. A first attempt to overcome this was based on a singular value decomposition of some matrices built from local data.[25] Results were encouraging, but some slight discrepancies with analytical results were noticed. Another approach, based on a model built directly from the data using reservoir computing, was also proposed.[26] In both cases, some discrepancies with the symbolic observability coefficients were observed. It therefore, remains challenging to develop a reliable technique that always matches with theoretical results. In this work, we propose a measure for assessing observability from recorded data by using DDA and compare our results and those obtained—when available in the literature—with the two techniques discussed above with the symbolic observability coefficients computed for several well-studied chaotic systems. Here, DDA is based on a delay differential equation that approximates the dynamics underlying the measured time series. Contrary to what is done with global modeling[27] or reservoir computing,[28] there is no need for an accurate model. Previous work showed a rough model with a very limited number of terms (typically three) is sufficient to detect dynamical changes or classify various dynamical regimes.[29–31]

The subsequent part of this paper is organized as follows. Section II A is a brief introduction to the computation of symbolic observability coefficients. Section II B provides an introduction to DDA and explains how it can be used for ranking variables according to the observability of the state space they provide. Section III introduces the investigated chaotic systems and provides the corresponding symbolic observability coefficients. Section IV is the main section of this paper: it discusses the performance of DDA for assessing observability of the chaotic systems and compares it with those of the two other data-based techniques. Section V provides some conclusions.

II. THEORETICAL BACKGROUND

A. Symbolic observability coefficients

Let us consider a d-dimensional dynamical system represented by the state vector $\boldsymbol{x} \in \mathbb{R}^{\mathbf{d}}$ whose components are given by

$$\dot{x}_i = f_i(x_1, x_2, x_3, \ldots, x_d), \quad i = 1, 2, 3, \ldots, d, \tag{1}$$

where f_i is the ith component of the vector field $\boldsymbol{f}$. Let us introduce the measurement function $\boldsymbol{h}(\boldsymbol{x}) : \mathbb{R}^{\mathbf{d}} \mapsto \mathbb{R}^{\mathbf{m}}$ of m variables chosen among the d ones spanning the original state space. It is then required to reconstruct a space $\mathbb{R}^{d_r}$ ($d_r \geq d$) from the m measured variables. One has to choose $d_r - m$ derivatives of these m measured variables to get a d_r-dimensional vector $\boldsymbol{X}$ spanning the reconstructed space. Commonly, observability is assessed by using $d_r = d$.[16,17] In the present work, we are only working with scalar time series ($m = 1$). The change of the coordinate between the original state space and the reconstructed one is thus the map,

$$\Phi : \mathbb{R}^d(\boldsymbol{x}) \mapsto \mathbb{R}^{\mathbf{d}}(\boldsymbol{X}). \tag{2}$$

When the derivative coordinates are used for spanning the reconstructed space, the map can be analytically computed.[32] The observability of a system from a variable is defined as follows.[16,33] For the sake of simplicity, let us limit ourselves to the case $m = 1$ (a generalization to the other cases is straightforward).

Definition 1. The dynamical system (1) is said to be *state observable* at time t_f if every initial state $\boldsymbol{x}(0)$ can be uniquely determined from the knowledge of the vector $s(\tau)$, $0 \leq \tau \leq t_f$.

To test whether a system is observable or not is to construct the observability matrix,[15] which is defined as the Jacobian matrix of the Lie derivatives of $h(\boldsymbol{x})$. Differentiating $h(\boldsymbol{x})$ yields

$$\frac{\mathrm{d}}{\mathrm{d}t} h(\boldsymbol{x}) = \frac{\partial h(\boldsymbol{x})}{\partial \boldsymbol{x}} \dot{x} = \frac{\partial h(\boldsymbol{x})}{\partial \boldsymbol{x}} \boldsymbol{f}(\boldsymbol{x}) = \mathcal{L}_f h(\boldsymbol{x}),$$

where $\mathcal{L}_f h(\boldsymbol{x})$ is the Lie derivative of $h(\boldsymbol{x})$ along the vector field $\boldsymbol{f}$. The kth order Lie derivative is given by

$$\mathcal{L}_f^k h(\boldsymbol{x}) = \frac{\partial \mathcal{L}_f^{\mathbf{k-1}} h(\boldsymbol{x})}{\partial x} \boldsymbol{f}(\boldsymbol{x}),$$

being the zero order Lie derivative the measured variable itself, $\mathcal{L}_f^0 h(\boldsymbol{x}) = h(\boldsymbol{x})$. Therefore, the observability matrix $\mathcal{O} \in \mathbb{R}^{d\times d}$ is written as

$$\mathcal{O}(\boldsymbol{x}) = \begin{bmatrix} \mathbf{d}h(\boldsymbol{x}) \\ \mathbf{d}\mathcal{L}_f h(\boldsymbol{x}) \\ \vdots \\ \mathbf{d}\mathcal{L}_f^{d-1} h(\boldsymbol{x}) \end{bmatrix}, \tag{3}$$

where $\mathbf{d} \equiv \frac{\partial}{\partial x}$.

Theorem 1. *The dynamical system (1) is said to be state observable if and only if the observability matrix has full rank, that is, rank* $(\mathcal{O}) = d$.

The observability matrix $\mathcal{O}$ is equal to the Jacobian matrix of the change of coordinates $\Phi : \boldsymbol{x} \to \boldsymbol{X}$ when derivative coordinates are used.[32] In this approach, the observability is either full or zero. The term *structural* was introduced when the results do not depend on parameter values.[34] Computing the rank of the observability matrix is independent of parameter values and, consequently, is an example of structural observability.[35] Computing observability with graphs[1,34,36] is also a structural approach. We term observability assessed from recorded data—necessarily dependent on the parameter values used for simulating the trajectory of the system—as *dynamical* observability.[35] This type of approach returns a real number within the unit interval: variables can be ranked between the two extreme cases, 1.0 (0.0) for a full (null) observability. There is a third type of observability, *symbolic* observability, which does not depend on parameter values but allows ranking the variables.[20] All types of observability are not sensitive to symmetry-related problems. This is due to the fact that observability is a local property, while symmetry is a global one. Consequently, symmetry may degrade the assessment of observability.[24]

The procedure to compute symbolic observability coefficients is implemented in three steps as follows.[7,20] First, the Jacobian matrix $\mathcal{J}$ of the system (1), composed of elements J_{ij}, is transformed into the symbolic Jacobian matrix $\tilde{\mathcal{J}}$ by replacing each constant element J_{ij} by 1, each polynomial element J_{ij} by $\bar{1}$, and each rational element J_{ij} by $\bar{\bar{1}}$ when the jth variable is present in the denominator or by $\bar{1}$ otherwise. Rational terms in the governing equations (1) are distinguished from polynomial terms since the formers reduce more strongly the observability than the latter.[20]

Then, the symbolic observability matrix $\tilde{\mathcal{O}}$ is constructed. The first row of $\tilde{\mathcal{O}}$ is defined by the derivative of the measurement function $\mathbf{d}h(\boldsymbol{x})$; that is, $\tilde{\mathrm{O}}_{1j} = 1$ if $j = i$ and 0 otherwise when the ith variable is measured. The second row is the ith row of $\tilde{\mathcal{J}}$. The kth row is obtained as follows. First, each element of the ith row of $\tilde{\mathcal{J}}$ is multiplied by the corresponding ith component of the vector $\boldsymbol{v} = (\tilde{\mathrm{O}}_{\ell 1}, \ldots, \tilde{\mathrm{O}}_{\ell d})^{\mathrm{T}}$, where $\ell = k - 1$ refers to the $(k-1)$th row of the symbolic observability matrix $\tilde{\mathcal{O}}$. The rules to perform the symbolic product $\tilde{J}_{ij} \otimes v_i$ are such that[20]

$$\left|\begin{array}{lll} 0 \otimes a & = 0, & \\ 1 \otimes a & = a, & \\ \bar{1} \otimes a & = a & \text{for } a = \bar{1}, \bar{\bar{1}}, \\ \bar{\bar{1}} \otimes a & = \bar{\bar{1}} & \text{for } a \neq 0. \end{array}\right. \tag{4}$$

Second, the matrix $\tilde{\mathcal{J}}'$ is reduced into a row where each element $\tilde{\mathrm{O}}_{kj} = \sum_i \tilde{J}'_{ij}$ according to the addition law,[20]

$$\left|\begin{array}{lll} 0 \oplus a & = a, & \\ 1 \oplus a & = a & \text{for } a \neq 0, \\ \bar{1} \oplus a & = a & \text{for } a \neq 0, 1, \\ \bar{\bar{1}} \oplus a & = \bar{\bar{1}}. & \end{array}\right. \tag{5}$$

The last step is associated with the computation of the symbolic observability coefficients. The determinant of $\tilde{\mathcal{O}}$ is computed according to the symbolic product rule defined in (4) and expressed as products and addends of the symbolic terms 1, $\bar{1}$, and $\bar{\bar{1}}$, whose number of occurrences are N_1, $N_{\bar{1}}$, and $N_{\bar{\bar{1}}}$, respectively. It is convenient to impose that if $N_{\bar{1}} = 0$ and $N_{\bar{\bar{1}}} \neq 0$, then $N_{\bar{1}} = N_{\bar{\bar{1}}}$. The symbolic observability coefficient is thus defined as

$$\eta = \frac{1}{D} N_1 + \frac{1}{D^2} N_{\bar{1}} + \frac{1}{D^3} N_{\bar{\bar{1}}}, \tag{6}$$

with $D = N_1 + N_{\bar{1}} + N_{\bar{\bar{1}}}$. This coefficient is in the unit interval, $\eta = 1$ for a variable providing full observability of the original state space. An observability is said to be good when $\eta \geq 0.75$.[37]

B. Delay Differential Analysis

Let us assume that a time series $\{X_1\}$ is recorded in a d-dimensional system. From this time series, it is possible to obtain a global model reproducing the underlying dynamics. There are typically two main approaches working with either derivative or delay coordinates.[27,38] When derivatives are used, it is possible to construct a d-dimensional differential model,

$$\begin{cases} \dot{X}_1 = X_2, \\ \dot{X}_2 = X_3, \\ \vdots \\ \dot{X}_d = F\left(X_1, X_2, \ldots, X_d\right), \end{cases} \tag{7}$$

where X_i is the $(i-1)$th derivative of the measured variable X_1.[39] The function F can be numerically estimated by using a least-squares technique with a structure selection.[40,41] F can be polynomial[39,41] or rational.[42,43] This model requires d-ordinary differential equations whose variables are the d successive derivatives of X_1: this model works in a differentiable embedding.

Second, it is possible to construct a model whose equations have the form of a difference equation,

$$X(k+1) = F\left(X_{\tau_j}(k)\right) = \sum_{i=0}^{N} a_i \, \varphi_i, \tag{8}$$

where φ_i is a monomial of delay coordinates $X_{\tau_j}(k) = X(k - \tau_j)$ with $\tau_j = n\delta_t$ $(n \in \mathbb{N}^+)$ being a time delay expressed in terms of the sampling time δ_t with which the scalar time series $\{X_1(k)\}$ is recorded: k is the discrete time. Such a model has an auto-regressive form, and typically, the number N of terms is between 10 and 20. The space in which this model is working is thus spanned by delay coordinates: its dimension is very often significantly larger than the dimension d, the embedding dimension,[44] or even than the Takens criterion.[8] An optimal form of the difference equation (8) is developed under the form of a nonlinear autoregressive-moving average (NARMA) model.[45]

Recently, a third type of model was investigated under the name of *reservoir computing*.[46] This approach considers an oversized model with a functional structure based on a network whose nodes are characterized by some simple function. For instance, the Lorenz attractor was accurately reproduced with an Erdös–Rényi network of 300 nodes with a mean degree $\overline{\delta} = 6$, each node being made of a difference equation.[28] The model so-obtained corresponds to an accurate global model of the dynamics. Notably, this model was constructed from the measurements of all the variables of the Lorenz system. The main advantage of such a large model is its flexibility, that is, its ability to capture various dynamical regimes, but it has the disadvantage that the space in which it is working is not clearly defined and has a very large dimension ($d_r > 300$ in the work discussed above).

The DDA approach uses a kind of a mixed model between the differential model (7) and the difference equation (8), the left member of the latter being replaced with the left member of the former. It is, therefore, based on the delay differential equation,

$$\dot{X} = \mathcal{F}_X = \sum_{i=1}^{N} a_i \, \varphi_i(X_{\tau_j}), \tag{9}$$

where $X = X_1$ designates the measured variable and X_{τ_j} some delay coordinates. The purpose is not to construct a global model reproducing accurately the dynamics but only an approximated model for detecting dynamical changes (nonstationarity) or classifying different dynamical regimes.[29,30,47] We, therefore, use a rough model with very few terms ($N \leq 3$). Such sparsity in the model prevents overparametrization, that is, spurious dynamics induced by overly complex models.[48] Indeed, delay differential equations are known to already produce complex dynamics with only two terms.[49,50] Many characteristics of the measured dynamics can be captured with two or three terms and appropriate time delays.[31] Based on previous works,[29–31,47] it is assumed that these characteristics are sufficient to distinguish different dynamical regimes. This DDA model (9) is a differential equation whose state space is spanned by delay coordinates X_{τ_j}.

Model (9) has two sets of parameters, the fixed parameters τ_j (set during the structure selection) and the free parameters a_i (estimated independently from each data window). The structure of model (9) as well as the delays are determined for each time series. Then, the free coefficients a_i are determined for each window of the recorded time series. The data in each window $\{X_1\}$ are normalized to have a zero mean and unit variance to remove amplitude information before estimating the free parameters a_i by using a singular value decomposition (SVD). The least-squares error

$$\rho_X = \sqrt{\frac{1}{K} \sum_{k=1}^{K} \left(\dot{X}(k) - \mathcal{F}_X(k) \right)^2} \tag{10}$$

between the derivatives returned by the DDA model and the derivatives computed from the measured time series quantifies the ability of the model to capture the underlying dynamics. It is known that there is a relationship between the model quality and observability.[9,11,24] The signal derivative $\dot{X}_1$ is computed using a five-point center derivative.[51] In this work, structure selection [i.e., choosing the model form of Eq. (9] and the fixed parameters τ_j) was performed via an exhaustive search over all possible three-term models (three monomials: $N = 3$) with two delays such that $\tau_j \in [m+1; 60]\delta_t$, where $m = 5$ is equal to the number of points for estimating the derivative and δ_t is the sampling time. Function $\mathcal{F}$ is made of three monomials selected from the possible candidates,

$$\varphi_i \in \left\{ X_{\tau_1}, X_{\tau_2}, X^2_{\tau_1}, X_{\tau_1} X_{\tau_2}, X^2_{\tau_2}, X^3_{\tau_1}, X^2_{\tau_1} X_{\tau_2}, X_{\tau_1} X^2_{\tau_2}, X^3_{\tau_2} \right\}. \tag{11}$$

Monomials and delays are selected in an exhaustive search over all possible model forms, i.e., 44, and delay combinations under the restrictions specified above. Each model is thus characterized by the set of "fixed" parameters (τ_1, τ_2), the corresponding monomials φ_i, and the free parameters a_i, which are estimated for each time window of the measured data. In this work, the time window is the entire time series. The structure providing the model with the lowest ρ_X is retained to assess observability according to the model error ρ_X.

As used with reservoir computing,[26] the error ρ_X between the model and the measured data provides a measure of how the system dynamics may be reconstructed from these data. Indeed, to obtain a reliable deterministic model, it is necessary to distinguish every different state of the system for retrieving the underlying causality. Since the error is used as a relative measure, it is only needed to have a sufficiently flexible functional form for the model as observed with reservoir computing or with a delay differential equation. Consequently, the smaller the error ρ_X, the higher the observability provided by the variable X. This results from previous works where it was shown that the complexity of the model to approximate was correlated to the observability: the better the observability provided by the measured variable, the simpler the model to approximate.[11,24] The error ρ_X from the best DDA model is computed with an increasing noise amplitude. For each three-dimensional system and each signal-to-noise ratio (no noise, 20, 10, and 0 dB: where 0 dB indicates that the variance of the noise matches the variance of the signal), the error ρ_X was computed over several hundred pseudoperiods for each time series.

III. DYNAMICAL SYSTEMS AND OBSERVABILITY COEFFICIENTS

A. Low-dimensional systems

The governing equations of the systems here investigated are reported in Table I. The symbolic observability coefficients (SOCs) and the model error ρ_X are reported for each variable of every system in Table I. Parameter values are reported in Table II.

The Rössler 76,[52] Lorenz 84,[55] Cord,[56] Hindmarsh–Rose[57] (HR), and Fisher[58] systems have no symmetry. The Hindmarsh–Rose system is known to be problematic when variable x or z is measured, for two different reasons.[64] When variable z is measured, the observability matrix

$$\mathcal{O}_z = \begin{bmatrix} 0 & 0 & 1 \\ rs & 0 & -r \\ rs\,(x\chi - r) & rs & r(r-s) \end{bmatrix}, \tag{12}$$

where $\chi = 2b - 3ax$ becomes singular when r is too small (Det $\mathcal{O}_z = r^2 s^2$): the observability can be null for $r = 0$ and full for $r \neq 0$

TABLE I. Governing equations of each system for which the symbolic observability coefficients (SOCs) η_{s^3} and ρ_X between the DDA model and the measured data with no noise contamination are reported. The SOC for variable x of the Hindmarsh–Rose (HR) system is corrected as discussed in the main text. For the Chua system, $f(x) = bx + \frac{1}{2}(a-b)(|x+1| - |x-1|)$.

System	Equations	SOC	Error
Rössler 76[52]	$\dot{x} = -y - z$	0.84	0.037
	$\dot{y} = x + ay$	1.0	0.022
	$\dot{z} = b + z(x - c)$	0.56	0.106
Rössler 77[53]	$\dot{x} = -ax - y(1 - x^2)$	0.56	0.0009
	$\dot{y} = \mu(bx + y - cz)$	0.84	0.0005
	$\dot{z} = \mu(x + cy - dz)$	0.68	0.0007
Lorenz 63[54]	$\dot{x} = \sigma(y - x)$	0.78	0.02
	$\dot{y} = Rx - y - xz$	0.36	0.039
	$\dot{z} = -bz + xy$	0.36	0.071
Lorenz 84[55]	$\dot{x} = -y^2 - z^2 - ax + aF$	0.36	0.061
	$\dot{y} = xy - bxz - y + G$	0.36	0.205
	$\dot{z} = bxy + xz - z$	0.36	0.204
Cord[56]	$\dot{x} = -y - z - ax + aF$	0.68	0.108
	$\dot{y} = xy - bxz - y + G$	0.36	0.198
	$\dot{z} = bxy + xz - z$	0.36	0.232
HR[57]	$\dot{x} = y - ax^3 + bx^2 + I - z$	0.68	0.025
	$\dot{y} = c - dx^2 - y$	0.56	0.023
	$\dot{z} = r[s(x - x_{\mathrm{R}}) - z]$	1.00	0.002
Fisher[58]	$\dot{x} = y$	1.00	0.003
	$\dot{y} = -ax - by - z$	0.84	0.004
	$\dot{z} = b + x - \|x\|$	0.56	0.027
Chua[59]	$\dot{x} = \alpha(-x + y - f(x))$	1.00	0.05
	$\dot{y} = x - y + z$	0.84	0.068
	$\dot{z} = -\beta y$	1.00	0.066
Duffing[60,61]	$\dot{x} = y$	1.00	0.022
	$\dot{y} = -\mu y + x - x^3 + u$	0.86	0.08
	$\dot{u} = v$	0.00	0.00
	$\dot{v} = -\omega^2 u$	0.00	0.00
Rössler 79[62]	$\dot{x} = -y - z$	0.75	0.005
	$\dot{y} = x + ay + w$	0.83	0.001
	$\dot{z} = b + xz$	0.44	0.079
	$\dot{w} = -cz + dw$	0.63	0.006
Hénon–Heiles[63]	$\dot{x} = u$	0.64	0.0005
	$\dot{y} = v$	0.64	0.0004
	$\dot{u} = -x - 2xy$	0.44	0.0009
	$\dot{v} = -y - y^2 - x^2$	0.44	0.0008

(this is also true for s, but s is commonly significantly different from 0). When variable x is measured, although the observability matrix $\mathcal{O}_x$ is never singular (Det $\mathcal{O}_x = r - 1; r \neq 1$), the plane projection of the differential embedding induced by variable x does not reveal the chaotic nature of the underlying dynamics, contrary to what is clearly provided by variable z (Fig. 1). As discussed by Aguirre *et al.*,[64] the observability matrix

$$\mathcal{O}_x = \begin{bmatrix} 1 & 0 & 0 \\ \chi x & 1 & 1 \\ O^x_{31} & \chi x - 1 & -\chi x + r \end{bmatrix}, \tag{13}$$

TABLE II. Parameter values of the investigated systems.

Rössler 76	$a = 0.52$	$b = 2$	$c = 4$	
Rössler 77	$a = 0.03$	$b = 0.3$	$c = 2$	$d = 0.5$
	$\mu = 0.1$			
Lorenz 63	$\sigma = 10$	$b = 8/3$	$R = 28$	
Lorenz 84	$a = 0.28$	$b = 4$	$F = 8$	$G = 1$
Cord	$a = 0.28$	$b = 4$	$F = 8$	$G = 1$
HR	$a = 1$	$b = 3$	$c = 1$	$d = 5$
	$I = 3.29$	$x_{\mathrm{R}} = \frac{8}{5}$	$r = 0.003$	$s = 4$
Fisher	$a = 0.3$	$b = 0.097$		
Chua	$\alpha = 9$	$\beta = \frac{100}{7}$	$a = -\frac{8}{7}$	
	$b = -\frac{5}{7}$			
Duffing	$\mu = 0.3$	$\omega = 1.2$		
	$x_0 = 1$	$y_0 = 0$		
	$u_0 = 0.5$	$v_0 = 0$		
Rössler 79	$a = 0.25$	$b = 3$	$c = 0.5$	$d = 0.05$

where

$$O^x_{31} = \chi^2 x^2 - rs - 2bx + 2(b - 3a) \times \left[\left(I + x^2(b - ax) + y - z\right)\right], \tag{14}$$

has a determinant Det $\mathcal{O}_x$ whose polynomial nature is canceled by the contributions of O_{32} and O_{33}, but this is not structurally stable. Any perturbation in one of these two elements would lead to a determinant vanishing for a subset of the state space. This is not detected by the symbolic observability coefficients. If we keep the polynomial nature of elements O_{32} and O_{33}, the symbolic observability matrix would be

$$\mathcal{O}_x = \begin{bmatrix} 1 & 0 & 0 \\ \bar{1} & 1 & 1 \\ \bar{1} & \bar{1} & \bar{1} \end{bmatrix}. \tag{15}$$

The corresponding corrected symbolic observability coefficient is thus $\eta'_{x^3} = 0.68$. The corrected ranking of variables is, therefore, $z \rhd x \rhd y$. This ranking will be used in the subsequent analysis.

The other systems have symmetry properties as follows. The Lorenz 63 system[54] is equivariant under a $\mathcal{R}_z$ rotation symmetry around the z-axis.[65,66] Variables x and y are mapped into their opposite ($-x$ and $-y$, respectively), while variable z is invariant under the rotation symmetry. At least two variables must be measured to correctly reconstruct the rotation symmetry.[23] The Rössler 77,[53] the Chua circuit,[59] and the driven Duffing systems[60,61] are equivariant under an inversion symmetry. Such a symmetry can be recovered from a single variable and, consequently, should not blur the observability analysis. The driven Duffing system is in fact a four-dimensional system, a conservative harmonic oscillator driving the dissipative Duffing oscillator: it is thus a semi-dissipative (or semi-conservative) system.[61] When variable u (or v) is recorded, a periodic orbit is obtained, while variable x (or y) provides a chaotic state portrait. Since a chaotic driving signal necessarily implies a chaotic response, it is obvious that u drives x and not the opposite. It can, therefore, be concluded, without further analysis, that the system is not observable from u (or v). Thus, we only have to determine

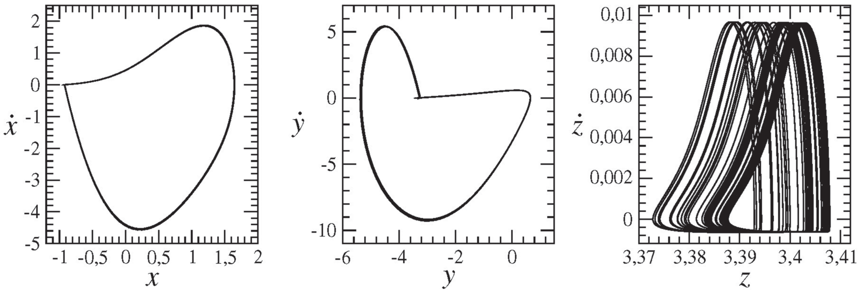


FIG. 1. Differential embedding induced by each of the three variables of the Hindmarsh–Rose system.

the observability from variables x and y, respectively. The Fisher system and the Chua circuit have a piecewise nonlinearity. They will be useful to test whether DDA is robust against discontinuous nonlinearity.

All these systems but three—the Lorenz 84, the Cord, and the Hénon–Heiles[63] systems—have at least one variable providing a good observability ($\eta > 0.75$) of the original state space. The Hénon–Heiles system is conservative, and one may guess that the observability problem will be more sensitive since the invariant domain of the state space has a dimension close to 3 and not 2 as for all the other systems that are strongly dissipative.

B. A higher-dimensional system

The Lorenz 63 system results from a Galerkin expansion of the Navier–Stokes equations for Rayleigh–Bénard convection.[67] It is also possible to have a higher-dimensional expansion in retaining more Fourier components. One of them lead to the 9D Lorenz system,[68]

$$\begin{cases} \dot{x}_1 = -\sigma(b_1x_1 + b_2x_7) + x_4(b_4x_4 - x_2) + b_3x_3x_5, \\ \dot{x}_2 = -\sigma x_2 + x_1x_4 - x_2x_5 + x_4x_5 - \dfrac{\sigma x_9}{2}, \\ \dot{x}_3 = \sigma(b_2x_8 - b_1x_3) + x_2x_4 - b_4x_2^2 - b_3x_1x_5, \\ \dot{x}_4 = -\sigma x_4 - x_2x_3 - x_2x_5 + x_4x_5 + \dfrac{\sigma x_9}{2}, \\ \dot{x}_5 = -\sigma b_5x_5 + \dfrac{x_2^2}{2} - \dfrac{x_4^2}{2}, \\ \dot{x}_6 = -b_6x_6 + x_2x_9 - x_4x_9, \\ \dot{x}_7 = -b_1x_7 - Rx_1 + 2x_5x_8 - x_4x_9, \\ \dot{x}_8 = -b_1x_8 + Rx_3 - 2x_5x_7 + x_2x_9, \\ \dot{x}_9 = -x_9 + (R + 2x_6)(x_4 - x_2) + x_4x_7 - x_2x_8, \end{cases} \tag{16}$$

where

$$\begin{cases} b_1 = 4\dfrac{1+a^2}{1+2a^2}, & b_2 = \dfrac{1+2a^2}{2(1+a^2)}, \\ b_3 = 2\dfrac{1-a^2}{1+a^2}, & b_4 = \dfrac{a^2}{1+a^2}, \\ b_5 = \dfrac{8a^2}{1+2a^2}, & b_6 = \dfrac{4}{1+2a^2}. \end{cases} \tag{17}$$

This 9D Lorenz system is equivariant.[69] Depending on the R-values, the attractor produced may be asymmetric [Fig. 2(a)] or symmetric [Fig. 2(b)]. The symbolic observability coefficients are

$$\begin{cases} \eta_{x_1^9} = \eta_{x_3^9} = \eta_{x_7^9} = \eta_{x_8^9} = 0.04, \\ \eta_{x_2^9} = \eta_{x_4^9} = 0.03, \\ \eta_{x_5^9} = \eta_{x_6^9} = \eta_{x_9^9} = 0, \end{cases} \tag{18}$$

leading to

$$x_1 = x_3 = x_7 = x_8 \triangleright x_2 = x_4 \triangleright x_5 = x_6 = x_9.$$

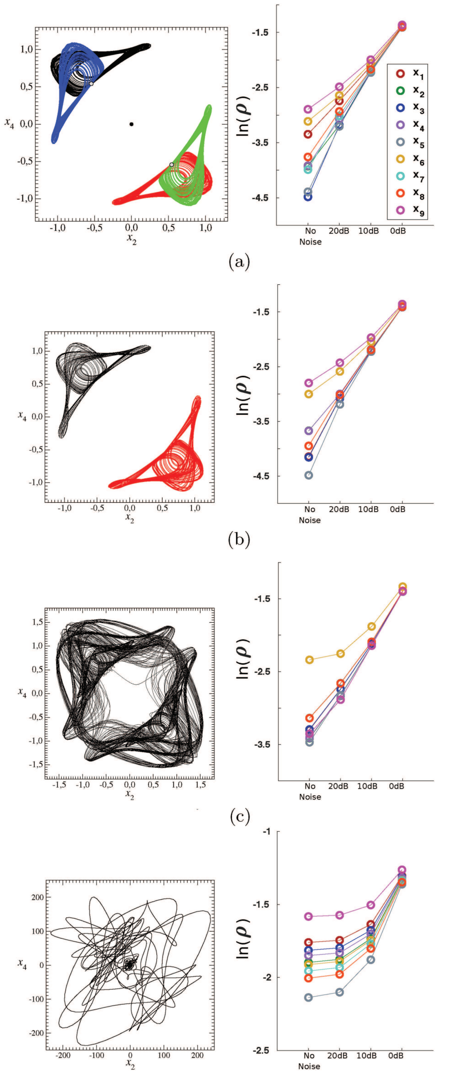
Notice that every variable offers an extremely poor observability of the original state space. It was shown that at least five variables need to be measured for having a good observability ($\eta > 0.75$) of the original state space.[7] Moreover, for a sufficiently large R-value ($R = 45$), the behavior is hyperchaotic. One of the characteristics of this highly developed behavior is that there are two different time scales. We will, therefore, investigate whether the observability assessed with DDA is dependent on parameter values, that is, on bifurcation affecting the symmetry properties (order-4 or order-2 asymmetric chaos, symmetric chaos, and hyperchaos).

IV. DDA RANKING

The structure of the best DDA models $\mathcal{F}_X$ under no noise is reported in Table V of the Appendix along with the corresponding time delays retained for identifying the free parameters. As examples, ρ_X for some systems with increasing noise is shown in Fig. 3. For no noise, ρ_X is reported in Table I.

The rankings for variables according to increasing symbolic observability coefficients (SOCs), decreasing ρ_X for DDA, and, when available in the literature, for decreasing reservoir computing (RC) and singular value decomposition observability (SVDO) are summarized in Table III for all low-dimensional systems ($d \leq 4$). The results for the Rössler 76, Rössler 77, Fisher, driven Duffing, and Rössler 79[62] systems are in a perfect agreement with the SOC. The discontinuity of the Fisher system does not perturb the analysis. The hyperchaotic nature of the Rössler 79 system was not problematic for correctly assessing observability.

The Lorenz 63, Lorenz 84, Cord, and Hindmarsh–Rose systems show close agreement between DDA and SOC. For the Lorenz



(a)

(b)

(c)

(d)

FIG. 2. Chaotic attractor produced by the 9D Lorenz system (18). (a) $R = 14.22$, (b) $R = 14.30$, (c) $R = 15.10$, and (d) $R = 45.00$. Other parameter values: $a = 0.5$ and $\sigma = 0.5$. When there are co-existing attractors, they are plotted in different colors in the plane projections of the state space.

63 system, variable x was correctly detected as providing the best observability, but variable z was found to offer worse observability than variable y, a feature that is not predicted by the SOC due to a problem inherent to the symmetry involved. For the Lorenz 84 system, all variables have equally low SOCs, however, for DDA variable x shows greater observability. For the Cord system, while no single variable provides good observability for the original state space, DDA correctly ranks x as providing the best observability. However, DDA ranks z as providing worse observability than variable y, while SOC ranks them with equivalent observability. For the Hindmarsh–Rose system, variable z provides full observability and is associated with the lowest ρ_X. However, there is some discrepancy between DDA and SOC since, as assessed with DDA, y provides a slightly higher observability than x. Results for the Hénon–Heiles system are quite equivalent to the SOC. Variables x and y are more observable than u and v; however, $y(v)$ is more observable than $x(u)$ instead of showing equivalent observability.

For the Chua circuit, the variable x contains a piecewise nonlinearity and has full observability, and DDA correctly ranks x as the most observable. DDA also ranks variable y with the worst observability, which is in agreement with SOC. However, variable z has only slightly better observability than y, whereas it should be equivalent to x.

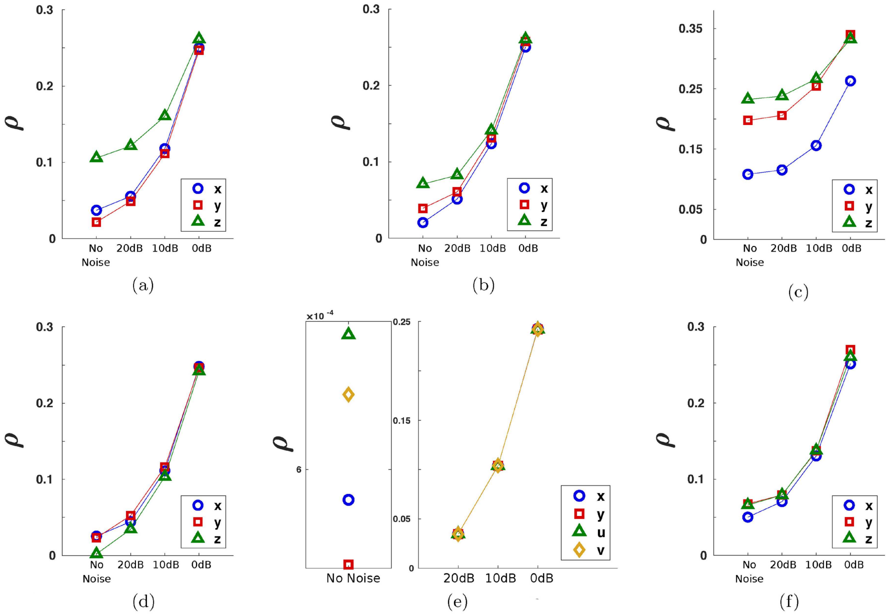
When compared to the two other data-based techniques, DDA performs better than RC for the Rössler 76, Rössler 79, and the Lorenz 63 systems but not for the Chua circuit. Compared to the SVDO, the DDA approach provides similar results for all systems investigated by these two techniques. DDA outperforms SVDO for the hyperchaotic Rossler 79 system in correctly identifying the variable y as providing the best observability, a feature missed by the SVDO, whereas the SVDO approach outperforms DDA for the Lorenz 84 and Hindmarsh-Rose systems.

For most of the systems, these results are robust against noise contamination, at least up to a signal-to-noise ratio greater than 10 dB: below this ratio, results can be blurred and observability can no longer be reliably assessed using DDA. A similar robustness was observed with SVDO. It was not investigated with RC.

Note that another interesting data-based technique for assessing observability was proposed by Parlitz *et al.*[70] It was only tested with the Rössler 76 system (and the Hénon map, not investigated here). It would be interesting to further investigate its performance, but this is out of the scope of this paper.

The results for the 9D Lorenz system are not so clear. The first reason is that this system is nearly unobservable from a single variable. The SOC is nearly saturated (close to 0) with nonlinear elements as revealed by the symbolic Jacobian matrix of the 9D Lorenz system (16), namely,

$$\mathcal{J}^{\rm sym} = \begin{bmatrix} 1 & \bar{1} & \bar{1} & \bar{1} & \bar{1} & 0 & 1 & 0 & 0 \\ \bar{1} & \bar{1} & 0 & \bar{1} & \bar{1} & 0 & 0 & 0 & 1 \\ \bar{1} & \bar{1} & 1 & \bar{1} & \bar{1} & 0 & 0 & 1 & 0 \\ 0 & \bar{1} & \bar{1} & \bar{1} & \bar{1} & 0 & 0 & 0 & 1 \\ 0 & \bar{1} & 0 & \bar{1} & \bar{1} & 0 & 0 & 0 & 0 \\ 0 & \bar{1} & 0 & \bar{1} & 0 & 1 & 0 & 0 & \bar{1} \\ 1 & 0 & 0 & \bar{1} & \bar{1} & 0 & 1 & \bar{1} & \bar{1} \\ 0 & \bar{1} & 1 & 0 & \bar{1} & 0 & \bar{1} & 1 & 0 \\ 0 & \bar{1} & \bar{1} & 0 & \bar{1} & \bar{1} & \bar{1} & \bar{1} & 1 \end{bmatrix}, \qquad (19)$$



(a) (b) (c)

(d) (e) (f)

FIG. 3. Error ρ_X vs a decreasing signal-to-noise ratio for some of the different systems investigated in this paper. (a) The Rössler 76 system, (b) the Lorenz 63 system, (c) the Cord system, (d) the Hindmarsh–Rose system, (e) the Hénon–Heiles system, and (f) the Chua system.

which illustrates that most of the couplings between variables are nonlinear. Considering only the observability provided by a single variable is here investigated and that the SOCs are all close to 0, one may conclude that the 9D Lorenz system is not observable from a single variable.

Results provided by DDA are shown in Fig. 2 where it is seen that variables cannot be easily ranked, particularly when R is increased. Results are summarized in Table IV as follows. For each R-value, the rankings of the variables are reported—from 1 for the variable offering the best observability to 9 for the one providing

TABLE III. Ranking variables according to the observability as assessed by the symbolic observability coefficients (SOCs), DDA analysis, reservoir computing (RC), and singular value decomposition observability (SVDO). A perfect agreement with the SOC is indicated by a •. When the variable providing the best observability is correctly detected or when = is replaced with ≈ or ▷, a ○ is reported.

System	SOC	DDA	RC	SVDO
Rössler 76	$y \triangleright x \triangleright z$	•	$x \triangleright y \triangleright z$	•
Rössler 77	$y \triangleright z \triangleright x$	•	…	…
Lorenz 63	$x \triangleright y = z$	○	$y \triangleright x \triangleright z$	○
Lorenz 84	$x = y = z$	○	…	•
Cord	$x \triangleright y = z$	○	…	○
Hindmarsh–Rose	$z \triangleright x \triangleright y$	○	…	•
Fisher	$x \triangleright y \triangleright z$	•	…	…
Chua	$x = z \triangleright y$	○	•	○
Duffing	$x \triangleright y \triangleright u = v$	•	…	…
Rössler 79	$y \triangleright x \triangleright w \triangleright z$	•	$x \triangleright y \triangleright z \triangleright w$	$x \triangleright y \triangleright w \triangleright z$
Hénon–Heiles	$x = y \triangleright u = v$	○	○	…

TABLE IV. Observability of the 9D Lorenz system as assessed with the symbolic observability coefficients (SOCs) and DDA.

	R	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
SOC	…	1	2	1	2	3	3	1	1	3
DDA	14.22	7	4	1	5	2	8	3	6	9
	14.30	3	7	2	6	1	8	5	4	9
	15.10	5	2	6	3	1	9	8	7	4
	45.00	8	5	7	6	1	4	3	2	9

the poorest observability—and compared to the ranking provided by the SOC. The results vary with the R-value but do not follow a clear trend. Variable x_5 with a null observability as assessed by the SOC (and analytically) is found to provide the best observability as assessed by DDA. Nevertheless, this is in agreement with the successful three-dimensional global model obtained from this variable for $R = 14.22$;[68] that is, at least for this R-value, the dynamics can be correctly reconstructed for recovering the underlying determinism.

It should be pointed out that looking for full observability (i.e., being able to "reconstruct" each of the non-measured variables) is not the same thing as looking for an embedding, especially for large d-dimensional systems producing an attractor that can be embedded within a space whose dimension d_R is lower than the dimension d of the original state space. Full observability ensures the existence of an embedding, and the opposite is not necessarily true. Here, DDA selects the variable that provides the best reconstructed space. If compared with the results provided by the SOC with multivariate measurements,[7] variables x_2, x_4, x_5, and x_6 are always among the six variables selected for providing a full observability. DDA returns three of them as providing the best observability, x_2, x_4, and x_5 (Table IV). Variable x_6, the single one that is invariant under the symmetry of this system, is identified as a variable providing a poor observability. Once again, symmetry induces difficulties for assessing observability.

V. CONCLUSION

The ability to infer the state of a system from a scalar output depends on the system variable that is measured. We have introduced a numerical approach using the error between a DDA model and measured data to assess the observability provided by the measured variables in several chaotic systems. The smaller the model error, the better the observability provided. We compared these measures with symbolic observability coefficients, which are determined directly from the system's equations. Our measure overall reliably ranks variables according to the observability they provide about the original state space. The largest discrepancy was obtained for a large-dimensional (9D Lorenz) system. The assessment of observability is quite robust against noise contamination in the majority of the systems here considered.

There are two situations in which our approaches may face some complications. The first one is a common one. Inconsistencies in assessing observability are known for systems with symmetry properties, particularly with variables left invariant. The second one is also a typical one: when the dimension of the system increases, the observability of the state space provided by a single variable becomes very poor and assessing observability is delicate. Our approach is thus very reliable for low-dimensional systems without symmetry properties, even with a signal-to-noise ratio as commonly encountered in experiments.

As in most of the other techniques, variables of different systems cannot be compared to each other. This is a common limitation in assessing observability that is only overcome by using an analytical approach, such as by computing explicitly the observability matrix or by using the symbolic observability coefficients. A kind of normalization should be considered to have, for instance, the error ρ_y of variable y of the Rössler 76 system (which has full observability) smaller than for variable y of the Rössler 77 system. This problem is more challenging than it may appear. It was, for instance, never solved for the observability coefficients computed along a trajectory using a relationship extracted from the system's equations or using SVD applied to a reconstructed space.

AUTHORS' CONTRIBUTIONS

C.E.G. and C. Lainscsek contributed equally to this work.

ACKNOWLEDGMENTS

C. Letellier wishes to thank Irene Sendiña-Nadal for her assistance in computing the symbolic observability coefficients for the 9D Lorenz system. This work was supported by the National Institutes of Health (NIH)/NIBIB (Grant No. R01EB026899-01) and by the National Science Foundation Graduate Research Fellowship (Grant No. DGE-1650112).

APPENDIX: FUNCTIONAL FORMS OF DDA MODELS

The functional forms of the DDA models for each variable of the systems investigated are shown in Table V.

TABLE V. Functional forms of the DDA models for each variable of the systems investigated. The time delays are expressed in terms of δ_t, the sampling time at which variable X is recorded.

		a_1	a_2	a_3	τ_1	τ_2
Rössler 76	$\mathcal{F}_x$	X_{τ_1}	X_{τ_2}	$X^3_{\tau_1}$	$6\,\delta_t$	$7\,\delta_t$
	$\mathcal{F}_y$	X_{τ_1}	X_{τ_2}	$X^3_{\tau_1}$	$6\,\delta_t$	$7\,\delta_t$
	$\mathcal{F}_z$	X_{τ_1}	X_{τ_2}	$X^3_{\tau_1}$	$6\,\delta_t$	$7\,\delta_t$
Rössler 77	$\mathcal{F}_x$	X_{τ_1}	X_{τ_2}	$X^3_{\tau_1}$	$6\,\delta_t$	$7\,\delta_t$
	$\mathcal{F}_y$	X_{τ_1}	X_{τ_2}	$X^3_{\tau_1}$	$7\,\delta_t$	$6\,\delta_t$
	$\mathcal{F}_z$	X_{τ_1}	X_{τ_2}	$X^3_{\tau_1}$	$6\,\delta_t$	$7\,\delta_t$
Lorenz 63	$\mathcal{F}_x$	X_{τ_1}	X_{τ_2}	$X^3_{\tau_1}$	$6\,\delta_t$	$7\,\delta_t$
	$\mathcal{F}_y$	X_{τ_1}	$X^3_{\tau_1}$	$X_{\tau_1}X^2_{\tau_2}$	$6\,\delta_t$	$19\,\delta_t$
	$\mathcal{F}_z$	X_{τ_1}	$X^2_{\tau_1}$	$X^2_{\tau_2}$	$18\,\delta_t$	$6\,\delta_t$
Lorenz 84	$\mathcal{F}_x$	X_{τ_1}	X_{τ_2}	$X^2_{\tau_1}$	$7\,\delta_t$	$6\,\delta_t$
	$\mathcal{F}_y$	X_{τ_1}	$X^3_{\tau_1}$	$X_{\tau_1}X^2_{\tau_2}$	$6\,\delta_t$	$28\,\delta_t$
	$\mathcal{F}_z$	X_{τ_1}	$X_{\tau_1}X_{\tau_2}$	$X_{\tau_1}X^2_{\tau_2}$	$6\,\delta_t$	$60\,\delta_t$

TABLE V. *(Continued.)*

		a_1	a_2	a_3	τ_1	τ_2
Cord	$\mathcal{F}_x$	X_{τ_1}	$X^3_{\tau_1}$	$X^2_{\tau_1}X_{\tau_2}$	$7\,\delta_t$	$51\,\delta_t$
	$\mathcal{F}_y$	X_{τ_1}	$X_{\tau_1}X_{\tau_2}$	$X_{\tau_1}X^2_{\tau_2}$	$6\,\delta_t$	$18\,\delta_t$
	$\mathcal{F}_z$	X_{τ_1}	X_{τ_2}	$X^2_{\tau_1}$	$6\,\delta_t$	$7\,\delta_t$
HR	$\mathcal{F}_x$	X_{τ_1}	$X_{\tau_1}X_{\tau_2}$	$X^3_{\tau_2}$	$6\,\delta_t$	$9\,\delta_t$
	$\mathcal{F}_y$	$X^2_{\tau_1}$	$X^2_{\tau_1}X_{\tau_2}$	$X^3_{\tau_2}$	$25\,\delta_t$	$6\,\delta_t$
	$\mathcal{F}_z$	X_{τ_1}	X_{τ_2}	$X^2_{\tau_1}$	$6\,\delta_t$	$7\,\delta_t$
Fisher	$\mathcal{F}_x$	X_{τ_1}	X_{τ_2}	$X^2_{\tau_1}$	$7\,\delta_t$	$6\,\delta_t$
	$\mathcal{F}_y$	X_{τ_1}	X_{τ_2}	$X^2_{\tau_1}$	$6\,\delta_t$	$7\,\delta_t$
	$\mathcal{F}_z$	X_{τ_1}	X_{τ_2}	$X^2_{\tau_1}$	$7\,\delta_t$	$6\,\delta_t$
Chua	$\mathcal{F}_x$	X_{τ_1}	X_{τ_2}	$X^3_{\tau_1}$	$6\,\delta_t$	$7\,\delta_t$
	$\mathcal{F}_y$	X_{τ_1}	X_{τ_2}	$X^3_{\tau_1}$	$7\,\delta_t$	$6\,\delta_t$
	$\mathcal{F}_z$	X_{τ_1}	X_{τ_2}	$X^3_{\tau_1}$	$13\,\delta_t$	$32\,\delta_t$
Duffing	$\mathcal{F}_x$	X_{τ_1}	X_{τ_2}	$X^3_{\tau_1}$	$6\,\delta_t$	$7\,\delta_t$
	$\mathcal{F}_y$	X_{τ_1}	X_{τ_2}	$X^3_{\tau_1}$	$6\,\delta_t$	$7\,\delta_t$
	$\mathcal{F}_u$	X_{τ_1}	X_{τ_2}	$X^2_{\tau_1}$	$38\,\delta_t$	$37\,\delta_t$
	$\mathcal{F}_v$	X_{τ_1}	X_{τ_2}	$X^2_{\tau_1}$	$38\,\delta_t$	$37\,\delta_t$
9D Lorenz	$\mathcal{F}_{1,3,5}$	X_{τ_1}	X_{τ_2}	$X^2_{\tau_1}$	$7\,\delta_t$	$6\,\delta_t$
$R = 14.22$	$\mathcal{F}_{4,7,8}$	X_{τ_1}	X_{τ_2}	$X^3_{\tau_1}$	$7\,\delta_t$	$6\,\delta_t$
	$\mathcal{F}_{6,9}$	X_{τ_1}	X_{τ_2}	$X^3_{\tau_1}$	$6\,\delta_t$	$7\,\delta_t$
	$\mathcal{F}_2$	X_{τ_1}	$X_{\tau_1}X_{\tau_2}$	$X^2_{\tau_2}$	$47\,\delta_t$	$14\,\delta_t$
9D Lorenz	$\mathcal{F}_{1-5,7,8}$	X_{τ_1}	X_{τ_2}	$X^3_{\tau_1}$	$6\,\delta_t$	$7\,\delta_t$
$R = 14.30$	$\mathcal{F}_{6,9}$	X_{τ_1}	X_{τ_2}	$X^3_{\tau_1}$	$7\,\delta_t$	$6\,\delta_t$
9D Lorenz	$\mathcal{F}_{1,3,7,8}$	X_{τ_1}	X_{τ_2}	$X^2_{\tau_1}$	$7\,\delta_t$	$6\,\delta_t$
$R = 15.10$	$\mathcal{F}_{2,4,5,9}$	X_{τ_1}	X_{τ_2}	$X^3_{\tau_1}$	$6\,\delta_t$	$7\,\delta_t$
	$\mathcal{F}_6$	X_{τ_1}	$X^2_{\tau_1}$	$X^2_{\tau_2}$	$25\,\delta_t$	$6\,\delta_t$
9D Lorenz	$\mathcal{F}_1$	X_{τ_1}	$X_{\tau_1}X_{\tau_2}$	$X^3_{\tau_1}$	$6\,\delta_t$	$11\,\delta_t$
$R = 45$	$\mathcal{F}_2$	X_{τ_1}	$X^3_{\tau_1}$	$X_{\tau_1}X^2_{\tau_2}$	$6\,\delta_t$	$57\,\delta_t$
	$\mathcal{F}_3$	X_{τ_1}	$X^3_{\tau_1}$	$X^3_{\tau_2}$	$6\,\delta_t$	$7\,\delta_t$
	$\mathcal{F}_4$	X_{τ_1}	$X^3_{\tau_1}$	$X_{\tau_1}X^2_{\tau_2}$	$6\,\delta_t$	$44\,\delta_t$
	$\mathcal{F}_5$	$X^3_{\tau_1}$	$X^2_{\tau_1}X_{\tau_2}$	$X^3_{\tau_2}$	$10\,\delta_t$	$6\,\delta_t$
	$\mathcal{F}_6$	$X^2_{\tau_1}$	$X_{\tau_1}X_{\tau_2}$	$X^3_{\tau_1}$	$10\,\delta_t$	$23\,\delta_t$
	$\mathcal{F}_7$	X_{τ_1}	$X_{\tau_1}X_{\tau_2}$	$X^3_{\tau_2}$	$6\,\delta_t$	$9\,\delta_t$
	$\mathcal{F}_8$	X_{τ_1}	$X_{\tau_1}X_{\tau_2}$	$X^3_{\tau_2}$	$6\,\delta_t$	$10\,\delta_t$
	$\mathcal{F}_9$	X_{τ_1}	$X^2_{\tau_1}X_{\tau_2}$	$X^3_{\tau_2}$	$6\,\delta_t$	$9\,\delta_t$
Rössler 79	$\mathcal{F}_x$	X_{τ_1}	X_{τ_2}	$X^2_{\tau_1}$	$6\,\delta_t$	$7\,\delta_t$
	$\mathcal{F}_y$	X_{τ_1}	X_{τ_2}	$X^3_{\tau_1}$	$7\,\delta_t$	$6\,\delta_t$
	$\mathcal{F}_z$	X_{τ_1}	X_{τ_2}	$X^2_{\tau_1}$	$6\,\delta_t$	$7\,\delta_t$
	$\mathcal{F}_w$	X_{τ_1}	X_{τ_2}	$X^3_{\tau_1}$	$7\,\delta_t$	$6\,\delta_t$
Hénon–Heiles	$\mathcal{F}_x$	X_{τ_1}	X_{τ_2}	$X^3_{\tau_1}$	$6\,\delta_t$	$7\,\delta_t$
	$\mathcal{F}_y$	X_{τ_1}	X_{τ_2}	$X^2_{\tau_1}$	$6\,\delta_t$	$7\,\delta_t$
	$\mathcal{F}_u$	X_{τ_1}	X_{τ_2}	$X^3_{\tau_1}$	$6\,\delta_t$	$7\,\delta_t$
	$\mathcal{F}_v$	X_{τ_1}	X_{τ_2}	$X^3_{\tau_1}$	$7\,\delta_t$	$6\,\delta_t$

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, "Observability of complex systems," Proc. Natl. Acad. Sci. U.S.A. **110**, 2460–2465 (2013).

[2] B. Wang, L. Gao, Y. Gao, Y. Deng, and Y. Wang, "Controllability and observability analysis for vertex domination centrality in directed networks," Sci. Rep. **4**, 5399 (2014).

[3] A. J. Whalen, S. N. Brennan, T. D. Sauer, and S. J. Schiff, "Observability and controllability of nonlinear networks: The role of symmetry," Phys. Rev. X **5**, 011005 (2015).

[4] D. Leitold, Á. Vathy-Fogarassy, and J. Abonyi, "Controllability and observability in complex networks—The effect of connection types," Sci. Rep. **7**, 151 (2017).

[5] A. Haber, F. Molnar, and A. E. Motter, "State observation and sensor selection for nonlinear networks," IEEE Trans. Control Netw. Syst. **5**, 694–708 (2018).

[6] D. Leitold, A. Vathy-Fogarassy, and J. Abonyi, "Design-oriented structural controllability and observability analysis of heat exchanger networks," Chem. Eng. Trans. **70**, 595–600 (2018).

[7] C. Letellier, I. Sendiña-Nadal, E. Bianco-Martinez, and M. S. Baptista, "A symbolic network-based nonlinear theory for dynamical systems observability," Sci. Rep. **8**, 3785 (2018).

[8] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, Lecture Notes in Mathematics Vol. 898 (Springer, 1981), pp. 366–381.

[9] C. Letellier, J. Maquet, L. L. Sceller, G. Gouesbet, and L. A. Aguirre, "On the non-equivalence of observables in phase-space reconstructions from recorded time series," J. Phys. A. Math. Gen. **31**, 7913–7927 (1998).

[10] C. Lainscsek, C. Letellier, and I. Gorodnitsky, "Global modeling of the Rössler system from the *z*-variable," Phys. Lett. A **314**, 409–427 (2003).

[11] C. Letellier, L. Aguirre, and J. Maquet, "How the choice of the observable may influence the analysis of nonlinear dynamical systems," Commun. Nonlinear Sci. Numer. Simul. **11**, 555–576 (2006).

[12] C. Letellier, "Estimating the Shannon entropy: Recurrence plots versus symbolic dynamics," Phys. Rev. Lett. **96**, 254102 (2006).

[13] L. L. Portes, A. N. Montanari, D. C. Correa, M. Small, and L. A. Aguirre, "The reliability of recurrence network analysis is influenced by the observability properties of the recorded time series," Chaos **29**, 083101 (2019).

[14] R. Kalman, "On the general theory of control systems," IRE Trans. Autom. Control **4**, 110–110 (1959).

[15] R. Hermann and A. Krener, "Nonlinear controllability and observability," IEEE Trans. Automat. Control **22**, 728–740 (1977).

[16] T. Kailath, *Linear Systems*, Information and System Sciences Series (Prentice-Hall, 1980).

[17] G. Chen and T. Ueta, "Yet another chaotic attractor," Int. J. Bifurcat. Chaos **9**, 1465–1466 (1999).

[18] L. A. Aguirre, "Controllability and observability of linear systems: Some noninvariant aspects," IEEE Trans. Educ. **38**, 33–39 (1995).

[19] C. Letellier and L. A. Aguirre, "Symbolic observability coefficients for univariate and multivariate analysis," Phys. Rev. E **79**, 066210 (2009).

[20] E. Bianco-Martinez, M. S. Baptista, and C. Letellier, "Symbolic computations of nonlinear observability," Phys. Rev. E **91**, 062912 (2015).

[21] M. Frunzete, J.-P. Barbot, and C. Letellier, "Influence of the singular manifold of nonobservable states in reconstructing chaotic attractors," Phys. Rev. E **86**, 026205 (2012).

[22] C. Letellier and L. A. Aguirre, "Graphical interpretation of observability in terms of feedback circuits," Phys. Rev. E **72**, 056202 (2005).

[23] G. King and I. Stewart, "Phase space reconstruction for symmetric dynamical systems," Physica D **58**, 216–228 (1992).

[24] C. Letellier and L. A. Aguirre, "Investigating nonlinear dynamics from time series: The influence of symmetries and the choice of observables," Chaos **12**, 549–558 (2002).

[25] L. A. Aguirre and C. Letellier, "Investigating observability properties from data in nonlinear dynamics," Phys. Rev. E **83**, 066209 (2011).

[26] T. L. Carroll, "Testing dynamical system variables for reconstruction," Chaos **28**, 103117 (2018).

[27] L. A. Aguirre and C. Letellier, "Modeling nonlinear dynamics and chaos: A review," Math. Prob. Eng. **2009**, 238960.

[28] J. Pathak, Z. Lu, B. R. Hunt, M. Girvan, and E. Ott, "Using machine learning to replicate chaotic attractors and calculate Lyapunov exponents from data," Chaos **27**, 121102 (2017).

[29] C. Lainscsek, C. Letellier, J. Kadtke, G. Gouesbet, and F. Schürrer, "Equivariance identification using delay differential equations," Phys. Lett. A **265**, 264–273 (2000).

[30] C. Lainscsek and T. J. Sejnowski, "Electrocardiogram classification using delay differential equations," Chaos **23**, 023132 (2013).

[31] C. Lainscsek, M. E. Hernandez, H. Poizner, and T. J. Sejnowski, "Delay differential analysis of electroencephalographic data," Neural Comput. **27**, 615–627 (2015).

[32] C. Letellier, L. A. Aguirre, and J. Maquet, "Relation between observability and differential embeddings for nonlinear dynamics," Phys. Rev. E **71**, 066213 (2005).

[33] C.-T. Chen, *Linear System Theory and Design* (Oxford University Press, Oxford, 1999).

[34] C.-T. Lin, "Structural controllability," IEEE Trans. Automat. Control **19**, 201–208 (1974).

[35] L. A. Aguirre, L. L. Portes, and C. Letellier, "Structural, dynamical and symbolic observability: From dynamical systems to networks," PLoS One **13**, e0206180 (2018).

[36] C. Letellier, I. Sendiña-Nadal, and L. A. Aguirre, "A nonlinear graph-based theory for dynamical network observability," Phys. Rev. E **98**, 020303 (2018).

[37] I. Sendiña-Nadal, S. Boccaletti, and C. Letellier, "Observability coefficients for predicting the class of synchronizability from the algebraic structure of the local oscillators," Phys. Rev. E **94**, 042205 (2016).

[38] J. P. Crutchfield and B. S. McNamara, "Equations of motion from a data series," Complex Syst. **1**, 417–452 (1987).

[39] G. Gouesbet and C. Letellier, "Global vector-field reconstruction by using a multivariate polynomial L_2 approximation on nets," Phys. Rev. E **49**, 4955–4972 (1994).

[40] S. Mangiarotti, R. Coudret, L. Drapeau, and L. Jarlan, "Polynomial search and global modeling: Two algorithms for modeling chaos," Phys. Rev. E **86**, 046205 (2012).

[41] S. Mangiarotti and M. Huc, "Can the original equations of a dynamical system be retrieved from observational time series?," Chaos **29**, 023133 (2019).

[42] C. S. M. Lainscsek, C. Letellier, and F. Schürrer, "Ansatz library for global modeling with a structure selection," Phys. Rev. E **64**, 016206 (2001).

[43] C. Lainscsek, "Nonuniqueness of global modeling and time scaling," Phys. Rev. E **84**, 046205 (2011).

[44] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," Physica D **110**, 43–50 (1997).

[45] L. A. Aguirre and S. A. Billings, "Retrieving dynamical invariants from chaotic data using NARMAX models," Int. J. Bifurcat. Chaos **5**, 449–474 (1995).

[46] M. Lukosevicius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," Comput. Sci. Rev. **3**, 127–149 (2009).

[47] C. Lainscsek, J. Weyhenmeyer, S. S. Cash, and T. J. Sejnowski, "Delay differential analysis of seizures in multichannel electrocorticography data," Neural Comput. **29**, 3181–3218 (2017).

[48] L. A. Aguirre and S. A. Billings, "Dynamical effects of overparametrization in nonlinear models," Physica D **80**, 26–40 (1995).

[49] M. C. Mackey and L. Glass, "Oscillation and chaos in physiological control systems," Science **197**, 287–289 (1977).

[50] J. D. Farmer, "Chaotic attractors of an infinite-dimensional dynamical system," Physica D **4**, 366–393 (1982).

[51] E. Miletics and G. Molnárka, "Taylor series method with numerical derivatives for initial value problems," J. Comput. Methods Sci. Eng. **4**, 105–114 (2004).

[52] O. E. Rössler, "An equation for continuous chaos," Phys. Lett. A **57**, 397–398 (1976).

[53] O. E. Rössler, *Continuous Chaos* (Springer-Verlag, Berlin, Germany, 1977), pp. 174–183.

[54] E. N. Lorenz, "Deterministic nonperiodic flow," J. Atmos. Sci. **20**, 130–141 (1963).

[55] E. N. Lorenz, "Irregularity: A fundamental property of the atmosphere," Tellus A **36**, 98–110 (1984).

[56] C. Letellier and L. A. Aguirre, "Required criteria for recognizing new types of chaos: Application to the "cord" attractor," Phys. Rev. E **85**, 036204 (2012).

[57] J. L. Hindmarsh and R. M. Rose, "A model of neuronal bursting using three coupled first order differential equations," Proc. R. Soc. Lond. B **221**, 87–102 (1984).

[58] S. Fischer, A. Weiler, D. Fröhlich, and O. E. Rössler, "Kleiner-attractor in a piecewise-linear C^1-system," Z. Naturforsch. A **54**, 268–269 (1999).

[59] L. O. Chua, Y. Yao, and Q. Yang, "Devil's staircase route to chaos in a non-linear circuit," Int. J. Circuit Theory Appl. **14**, 315–329 (1986).

[60] G. Duffing, *Erzwungene Schwingungen bei Veränderlicher Eigenfrequenz und ihre Technische Bedeutung* (Vieweg, Braunschweig, 1918).

[61] O. Ménard, C. Letellier, J. Maquet, L. L. Sceller, and G. Gouesbet, "Analysis of a non synchronized sinusoidally driven dynamical system," Int. J. Bifurcat. Chaos **10**, 1759–1772 (2000).

[62] O. E. Rössler, "An equation for hyperchaos," Phys. Lett. A **71**, 155–157 (1979).

[63] M. Hénon and C. Heiles, "The applicability of the third integral of motion: Some numerical experiments," Astron. J. **69**, 73–79 (1964).

[64] L. A. Aguirre, L. L. Portes, and C. Letellier, "Observability and synchronization of neuron models," Chaos **27**, 103103 (2017).

[65] C. Letellier, P. Dutertre, and G. Gouesbet, "Characterization of the Lorenz system, taking into account the equivariance of the vector field," Phys. Rev. E **49**, 3492–3495 (1994).

[66] C. Letellier and R. Gilmore, "Covering dynamical systems: Two-fold covers," Phys. Rev. E **63**, 016206 (2001).

[67] B. Saltzman, "Finite amplitude free convection as an initial value problem—I," J. Atmos. Sci. **19**, 329–341 (1962).

[68] P. Reiterer, C. Lainscsek, F. Schürrer, C. Letellier, and J. Maquet, "A nine-dimensional Lorenz system to study high-dimensional chaos," J. Phys. A Math. Gen. **31**, 7121 (1998).

[69] R. Gilmore and C. Letellier, *The Symmetry of Chaos* (Oxford University Press, 2007).

[70] U. Parlitz, J. Schumann-Bischoff, and S. Luther, "Local observability of state variables and parameters in nonlinear modeling quantified by delay reconstruction," Chaos **24**, 024411 (2014).

Chapter 4, in full, is a reprint of the material as it appears in Chaos, titled “Assessing observability of chaotic systems using Delay Differential Analysis”,2020; 30. Gonzalez, Christopher; Lainscsek, Claudia; Sejnowski, Terrence; Letellier, Christopher. The dissertation author was one of the primary investigators and authors of this paper.