Real-Time Neuroimaging and Cognitive Monitoring Using Wearable Dry EEG

Tim R. Mullen*, *Member, IEEE*, Christian A. E. Kothe, *Member, IEEE*, Yu Mike Chi, *Member, IEEE*,
Alejandro Ojeda, Trevor Kerth, Scott Makeig, Tzyy-Ping Jung, *Fellow, IEEE*, and Gert Cauwenberghs, *Fellow, IEEE*

Abstract—Goal: **We present and evaluate a wearable high-density dry-electrode EEG system and an open-source software framework for online neuroimaging and state classification.** *Methods:* **The system integrates a 64-channel dry EEG form factor with wireless data streaming for online analysis. A real-time software framework is applied, including adaptive artifact rejection, cortical source localization, multivariate effective connectivity inference, data visualization, and cognitive state classification from connectivity features using a constrained logistic regression approach (ProxConn). We evaluate the system identification methods on simulated 64-channel EEG data. Then, we evaluate system performance, using ProxConn and a benchmark ERP method, in classifying response errors in nine subjects using the dry EEG system.** *Results:* **Simulations yielded high accuracy (AUC $= 0.97 \pm 0.021$) for real-time cortical connectivity estimation. Response error classification using cortical effective connectivity [short-time direct-directed transfer function (sdDTF)] was significantly above chance with similar performance (AUC) for cLORETA (0.74 ± 0.09) and LCMV (0.72 ± 0.08) source localization. Cortical ERP-based classification was equivalent to ProxConn for cLORETA (0.74 ± 0.16) but significantly better for LCMV (0.82 ± 0.12).** *Conclusion:* **We demonstrated the feasibility for real-time cortical connectivity analysis and cognitive state classification from high-density wearable dry EEG.** *Significance:* **This paper is the first validated application of these methods to 64-channel dry EEG. This study addresses a need for robust real-time measurement and interpretation of complex brain activity in the dynamic environment of the wearable setting. Such advances can have broad impact in research, medicine, and brain–computer interfaces. The pipelines are made freely available in the open-source SIFT and BCILAB toolboxes.**

*T. R. Mullen was with the Swartz Center for Computational Neuroscience, Institute for Neural Computation, and Department of Cognitive Science, University of California San Diego, La Jolla, CA 92093 USA. He is now with Syntrogi Labs, San Diego, CA 92121 USA (e-mail: tim.mullen@syntrogi.com).

C. A. E. Kothe and A. Ojeda were with the Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California San Diego. They are now with Syntrogi Labs.

Y. M. Chi is with Cognionics, Inc.

T. Kerth was with Cognionics, Inc. He is now at Kingston University, London, UK.

S. Makeig and T.-P. Jung are with the Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California San Diego.

G. Cauwenberghs is with the Department of Bioengineering, and Institute for Neural Computation, University of California San Diego.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2015.2481482

Index Terms—**Adaptive systems, brain–computer interfaces (BCI), connectivity analysis, dry-contact electrode, electroencephalography (EEG), neuroimaging, wearable sensors.**

I. INTRODUCTION

IN recent years, advances in dry-electrode electroencephalography (EEG) and wireless integrated acquisition systems [1], [2] have spurred increasing development of a new generation of wearable, mobile applications of EEG for real-world cognitive state monitoring, clinical diagnostics and therapeutics, and brain–computer interfaces (BCI), among others [3]–[7]. Concomitant with this is an increasing scientific appreciation for the importance of measuring complex dynamic interactions (e.g., functional or effective connectivity) between brain processes. These advances may provide key predictive information regarding brain function and dysfunction [8]–[11]. In particular, measuring interactions at the level of cortical sources, rather than sensors can offer increased interpretability, while reducing confounding factors of volume conduction [12]–[14].

However, many practical applications of EEG call for further developments in signal processing and machine learning to improve real-time (and online) measurement and classification of brain and behavioral states from small samples of noisy EEG data. Such developments present significant challenges, which we comprehensively review in [15]. Methods for motion artifact rejection and neuronal system identification in the highly dynamic environments of mobile wearable EEG settings must be fully automatable and capable of adapting to changes in measured data distributions. Robust statistical machine learning approaches are required for modeling relationships between high-dimensional neuronal features and cognitive or behavioral states. For real-time applications, such methods must be capable of operating efficiently with minimal computational delay. Finally, the integration of data acquisition, processing, classification, and visualization pipelines within a unified interoperable software framework is a key to reduce barriers to real-world application and reproducibility.

Of similar importance is the development of wearable (wireless, lightweight, dry) EEG hardware capable of comparable signal quality to research-grade wet systems. High channel density and spatial coverage are particularly important for effective artifact rejection and for high-resolution EEG source localization [12].

Over the last decade, an increasing number of studies have explored the application of multivariate functional and effective connectivity estimation in the EEG source domain (reviewed in He *et al.* [16]). For example, Babiloni *et al.* [17] studied

spectral directed information transfer between cortical regions of interest (ROIs) in a finger-tapping task. Astolfi *et al.* [18] performed a detailed performance analysis of three commonly used multivariate spectral effective connectivity estimators [including dDTF and partial directed coherence (PDC) estimators used in this study] applied to cortical ROI activity. They demonstrated reliable recovery of cortical connectivity patterns in simulations and Stroop experimental task data. Haufe *et al.* [14] provided a critical simulation-based assessment of phase slope index and granger-causality connectivity measures in both sensor and source space. Hassen *et al.* [12] performed a comparative study of several approaches for source localization and connectivity analysis, applied to a well-characterized (picture recognition and naming) experimental task dataset.

However, these studies applied source connectivity models to ensembles of multitrial data, confining applications to offline analysis. Less common is the online application of source-level multivariate connectivity inference at the level of single trials and in real time. Furthermore, the use of single trial multivariate source connectivity as predictive features for BCIs still remains relatively unexplored. One exception is a 2013 paper by Billinger *et al.* [19] that described and evaluated a system for single-trial source connectivity analysis applied to motor imagery classification. While this system shares some features with our own, there are also ample differences, which we note in the discussion section. We also point to an innovative paper by Stopczynski *et al.* [7] demonstrating online low-resolution cortical source localization on a mobile phone using 14-channel (Emotiv Epoch) wet EEG.

The objective of this paper is to describe and demonstrate 1) a novel high-density (64-channel) dry EEG hardware system, and 2) a software framework for real-time artifact rejection, source localization and connectivity analysis, cognitive/behavioral state classification, and data visualization. Outside a preliminary case study by our group [20], this is the first demonstration of such a framework applied to high-density dry wearable EEG data.

The software is made freely available within open-source toolboxes by the authors, including BCILAB [21] and SIFT [5], [22].

In [1], a first version of a 64-channel dry EEG system was introduced, focusing on the physical properties of the dry electrode, and briefly highlighting the wearable headset and compact electronics. Here, we present an extended version with a detailed description of the complete headset system, including operational mechanics to minimize motion artifacts; system specifications and electronics, including analog front-end and shielding for obtaining high-quality signals from dry electrodes; and a wireless communications system, necessary for transmitting accurate time-marked data in a wireless environment.

We further demonstrate the use of the dry EEG system with the aforementioned real-time framework for artifact rejection and neuronal system identification, expanding on our earlier 2013 report [20], in which we provided a brief introduction and preliminary (single subject) evaluation of the system. In this paper, we present mathematical details of key methods, including the artifact subspace reconstruction (ASR) method for arti-

fact rejection; an efficient implementation of anatomically constrained LORETA for source localization; and the application of the alternating direction method of multipliers (ADMM) for efficient sparse neuronal system identification and connectivity-based cognitive state classification. Additionally, we evaluate system performance in a nine-subject BCI study.

We note that alternative open-source software solutions are available for inferring single-trial effective connectivity in the source domain. These include the MATLAB-based eConnectome toolbox [23] and the Python-based SCoT toolbox [24]. The purpose of this paper is not to compare those useful toolboxes with BCILAB or SIFT, or advocate for any specific toolbox. However, we note that, to our knowledge, alternative toolboxes are designed primarily for offline data analysis, and have not yet been optimized for online (streaming) or real-time application. Further, eConnectome does not offer methods for cognitive state classification. While other toolboxes offer methods unavailable in BCILAB or SIFT, the integration of BCILAB and SIFT offers a uniquely comprehensive selection of methods for EEG signal processing, neuronal system identification, and machine learning which may be easily combined into standard BCILAB pipelines for online or offline application. The pipelines in this paper demonstrate just a few possible combinations of such methods.

The outline of this paper is as follows. First, we detail the design and implementation of the wireless 64-channel dry-electrode EEG system. Then, we provide details on the signal processing and machine-learning framework supporting real-time analysis of the streaming data. Next, we describe two validation studies: a 64-channel simulated EEG experiment and an EEG BCI experiment (detecting behavioral response error commission in a modified Eriksen Flanker task) using the wearable system. Finally, we present and discuss the results of experiments including exposition and interpretation of neuronal features that discriminate between correct and erroneous responses.

II. MATERIAL AND METHODS

The proposed real-time analysis framework is outlined in Fig. 1. EEG data are acquired from the wearable dry EEG system via the open-source Lab Streaming Layer (LSL) software[1]. The data stream feeds to a data analysis and classification pipeline consisting of preprocessing, source localization, dynamical model fitting and connectivity estimation, and cognitive state classification. Supporting tools for 2-D and 3-D data visualization augment this, allowing examination of task-relevant brain network dynamics and activity across time, frequency, and anatomical location.

The framework is implemented in the MATLAB-based (The Mathworks, Natick, MA, USA) BCILAB[2] [21] and SIFT[3] [5] toolboxes for EEGLAB. We have made the workflow available as the BCILAB "Source Information Flow Toolbox Adapter" paradigm (ParadigmSIFT class). For an example of online data

---

[1]"LSL" https://github.com/sccn/labstreaminglayer.

[2]"BCILAB" http://sccn.ucsd.edu/wiki/BCILAB
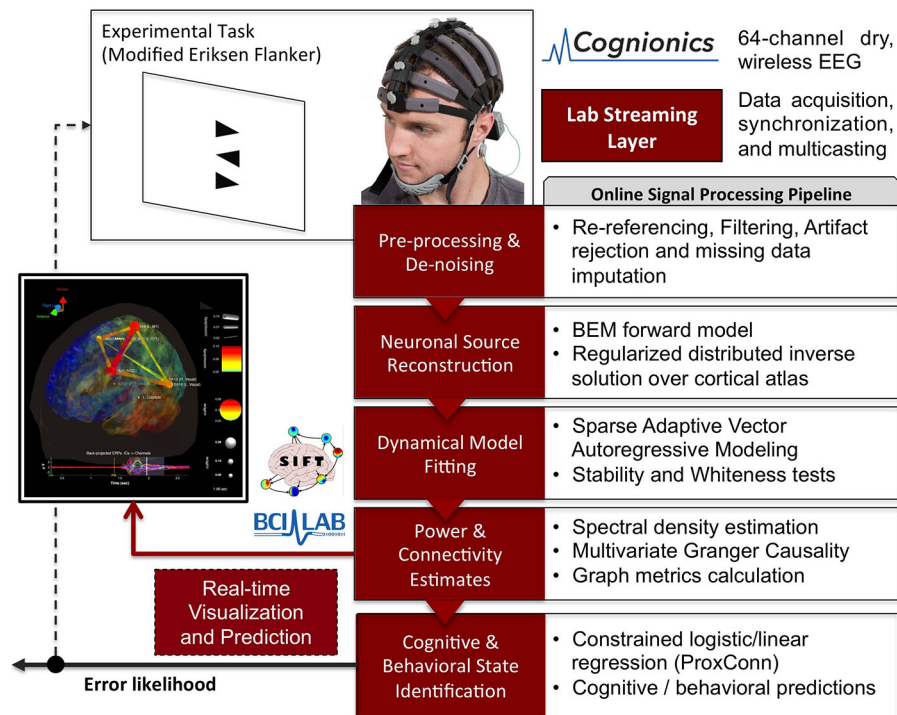
[3]"SIFT" http://sccn.ucsd.edu/wiki/SIFT

Fig. 1. Schematic of the real-time data processing pipeline used in this study.

acquisition and source connectivity analysis, we point to SIFT's /scripts/BCILAB_MINIMAL_DEMO example.

A. Wearable EEG Hardware

While dry-electrode EEG systems have been explored for BCIs and are commercially available for "consumer" gaming applications, few such systems feature more than a handful of EEG electrodes. Cognionics has developed the HD-72 dry wireless high-density EEG headset, shown in Fig. 2(a). The system features 64 EEG electrodes [see Fig. 2(d)] plus reference and ground. An additional eight recording channels are available providing ECG, EMG, respiration, and other physiological variables for mobile brain-body activity monitoring.

Obtaining high-quality EEG signals in real-world environments is challenging due to the various sources of electrical, mechanical, and physiological artifacts, especially in real-world environments. The EEG headset is designed to mitigate these challenges by optimizing electromechanical design in a single, integrated, and wearable form factor.

In terms of electronics, a practical wearable EEG system must not only be lightweight but also able to reject electrical interference and cope with variable and changing electrode contact qualities. External electrical noise is often the first sign of poor signal quality, commonly observed as 50/60 Hz line noise. While notch filtering has some utility in removing known line noise, many other sources of external interference (e.g., static charging as the subject moves) are unpredictable and cannot be removed via simple filtering. To minimize the influence of external electrical fields, the headset utilizes an actively driven ground system to sense and cancel out common-mode poten-

tials on the subject's body. In addition, the internal wiring of the headset itself sits within a local Faraday cage-like enclosure formed by a conductive layer, spanning the headset, driven by the output of a reference amplifier. This further eliminates differential interference, which is particularly problematic with high-impedance dry electrodes.

In addition to rejecting external noise, the headset electronics provide a high dynamic range input (± 400 mV) to cope with the potentially large dc offsets encountered with dry electrodes. The use of a 24-bit ADC enables the use of low analog amplification and the elimination of ac-coupling within the signal path. Large transient artifacts (e.g., sweating or movement) recover quickly as there are no filter settling or amplifier saturation issues.

The headset provides an optional real-time impedance measurement for monitoring of the electrode contact quality prior to and during recording. This can significantly reduce the time required for setup, while allowing for improved automated channel rejection during recording. Each channel contains a precision ac current source (± 24 nA) operating at one-fourth the sample rate. Measuring the induced voltage drop, with respect to the reference electrode, isolates the local electrode impedance. The impedance check signal is superimposed as a carrier wave on top of the EEG and notch filtered out by the acquisition software.

Signals are digitized at 300 samples/s with a bandwidth from dc to 50 Hz (80 Hz with impedance check off) and transmitted via an onboard Bluetooth transmitter. A secondary radio is also onboard to receive event markers/triggers that require precise timing. The trigger receiver also operates within the 2.4 GHz band but uses a custom protocol optimized for the reliable and deterministic transmission of short data packets. Markers are sent by a dedicated transmitter box [see Fig. 2(a), right] with
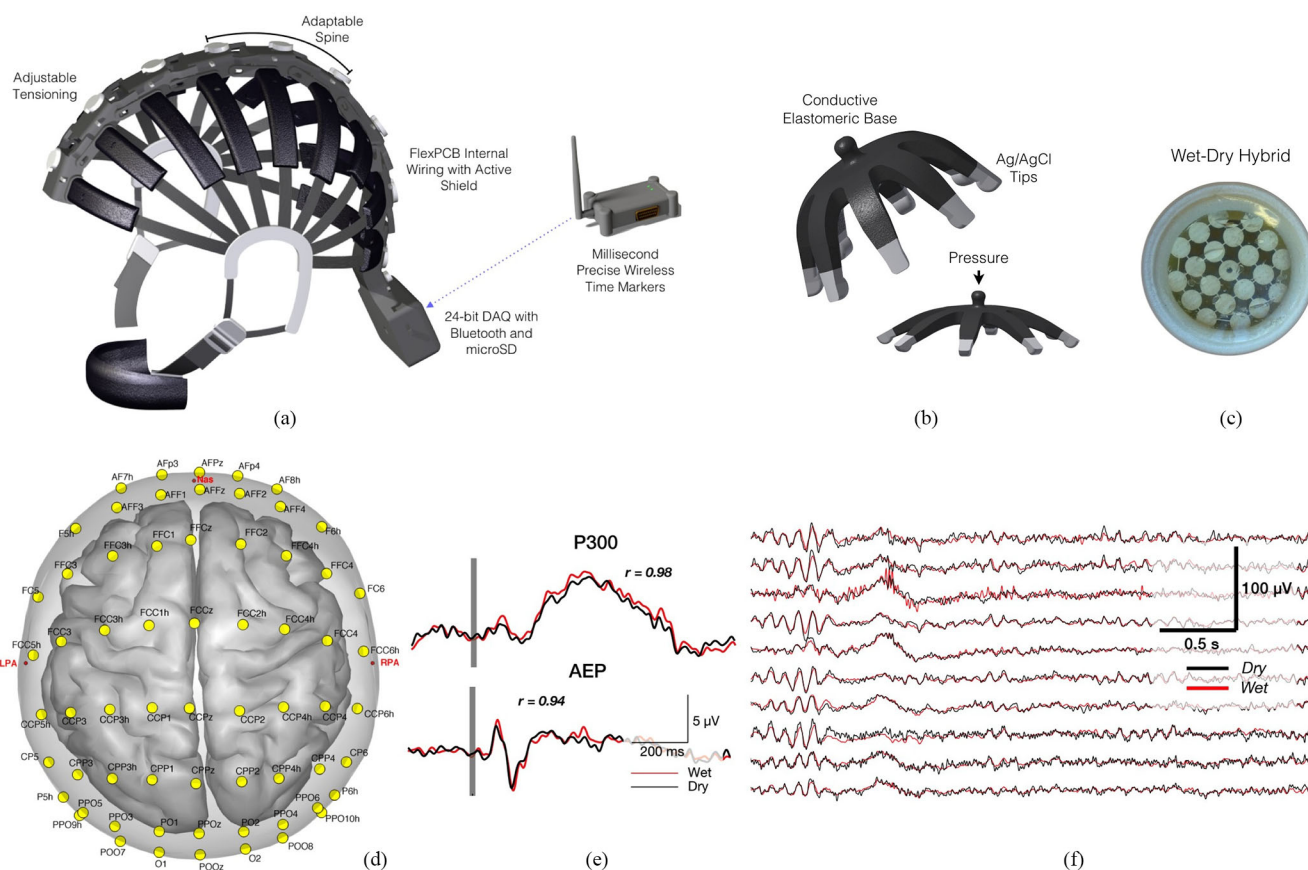
Fig. 2. Cognionics HD-72 64-channel mobile EEG system. (a) EEG headset harness with adjustable tensioning of dry electrodes contacting the scalp, and with Bluetooth wireless transmission and data synchronization. (b) Flexible active dry-contact Ag/AgCl EEG electrodes, and pressure-induced flexing mechanism to reach scalp contact through hair [1]. (c) Hybrid wet-dry electrode with ion-permeable membrane separating conductive gel inside from skin outside. (d) 64-channel sensor montage, coregistered with MNI "Colin27" brain. Average sensor locations were obtained by averaging 3-D digitized (ELPOS, Zebris Medical GmbH) electrode locations from ten individuals. Electrodes labels are assigned based on the nearest neighbor mapping to the standard 10/5 montage. *Nas*, *LPA*, and *RPA* denote nasion and left/right preauricular fiducials. (e) Standard wet (3Mdot Ag/AgCl) and the flexible active dry electrodes produce comparable averaged evoked response potentials and (f) good agreement between simultaneously recorded continuous wet and dry data.

standard RS232 serial (DB-9) and TTL parallel-type (DB-25) inputs. Timing accuracy is less than 2 ms, independent of the large latencies and jitter encountered with Bluetooth and without the use of a wireline. The transmitter unit also provides a virtual serial interface over USB.

While the electronics provide a high degree of electrical shielding and low-noise signal amplification, a dry-electrode system is also highly dependent on the mechanics to provide adequate contact between sensor and subject, particularly during movement. Unlike wet electrodes, dry electrodes lack the benefit of a fluid coupling medium to fill gaps between the electrode metal and the surface of the scalp. A dry-electrode system is critically dependent on a harness to hold the electrodes in place and maintain direct skin contact. Building a high-density dry-electrode array is especially difficult given the many variations in head size and shape.

To adapt to a wide range of subjects, the EEG headset starts with a mechanically flexible "spine" running from the forehead to the base of the neck [shown in Fig. 2(a)]. The spine is made from a series of plastic pods that are hinged together to form a single easy to handle unit. Each pod contains a pair of bands that

run laterally out to the sides of the subject's head and contain a row of sensors. A knob at each pod adjusts the tension and sensor contact pressure. Providing independent tension adjustment at each pod enables the headset to conform to different individuals and use cases (e.g., more tension for ambulatory use and less for more comfortable stationary recordings). To minimize weight, the internal wiring is provided by a flexible printed circuit board which is enclosed inside the headset. The base of the headset at the neck [shown in Fig. 2(a)], houses the electronics module and provides two-wire connections that terminate in standard ECG-sized snap connectors, for reference and active ground. The entire system weighs only 350 g, including batteries, enabling it to be easily worn by a mobile subject.

Two types of sensors are used with the headset: one with flexible prongs designed to go through hair shown in Fig. 2(b), and one with a flat surface designed for use on bare skin shown in Fig. 2(c). Both sensors connect to the headset via a miniature snap receptacle, enabling it to be easily interchanged as needed. The flexible sensor contains an array of angled legs. Mild to moderate pressure, from the headset onto the scalp, causes the legs to spread to better push aside hair and make contact to the

scalp. Hard pressure causes the sensors to completely flatten, making it potentially safer than conventional straight metal pin electrodes. The body of the electrode is made from a flexible polymer and coated with a conductive outer layer. The tips, which make skin contact, are further coated with Ag/AgCl for the best possible signal quality. Typical contact impedances range from a few megohms to hundreds of kilohms depending on the condition of the subject's skin and contact pressure. The high input impedance of the amplifier and the use of an integrated active shield enables the system to obtain acceptable signal quality despite higher contact impedances than traditional wet electrodes.

For areas of the head with bare skin, a pad-like sensor is used instead for maximum comfort and signal quality. The pad sensor contains a layer of hydrogel sandwiched between a semipermeable membrane and a plate of Ag/AgCl. The membrane enables ionic conduction with the skin for high signal quality, while retaining a reusable and dry exterior. Due to larger surface area of the pad sensor, the impedances are typically lower, in the range of tens to hundreds of kilohms.

System evaluation demonstrated ability to acquire high-fidelity EEG signals even with the use of high-impedance dry electrodes. Average evoked potentials (AEP, SSVEP, P300) showed a correlation of $r > 0.9$ with signals measured simultaneously with nearby wet electrodes [see Fig. 2(e)]. An additional plot of simultaneously acquired wet and dry raw EEG data is also shown in Fig. 2(f) demonstrating comparable single-trial signal quality for wet and dry electrodes. Preprocessing and Artifact Rejection

Despite the use of an artifact-mitigating form factor and electronics design, motion artifacts and poor-contacting EEG sensors can remain a challenge for both wet- and dry-electrode EEG data in mobile wearable settings. Furthermore, physiological artifacts, such as EMG and skin potentials, are inherently a part of the recording. We employ online preprocessing in our BCILAB pipeline to further remove such artifacts.

The preprocessing framework supports several methods for artifact removal. This includes rejecting a subspace of ICA components precomputed using an (possibly overcomplete) decomposition [25] on calibration data or adaptively estimated using online recursive ICA [26], [27]. In this paper, we describe an adaptive spatial filtering approach called ASR, which we briefly introduced in [20]. The ASR filter operates online and is designed to detect and remove high-amplitude data components (for instance, stemming from eye blinks, muscle, and sensor motion) of high amplitude relative to some artifact-free reference data, while recovering EEG background activity that lies in the subspace spanned by the artifact components (see Fig. 8 for an example). Fig. 3 graphically demonstrates the ASR procedure, which we outline next.

Let $X_c = \mathbb{R}^{Q \times M}$ be a reference (e.g., artifact-free) signal with low artifact content. Typically, this may be a short (e.g., 1 min) segment of data collected at during a "calibration" period at the start of an online session, or it may be heuristically extracted from a longer data segment containing artifacts. Let $x(t) \in \mathbb{R}^Q$ be a Q-channel EEG sample measured at time point
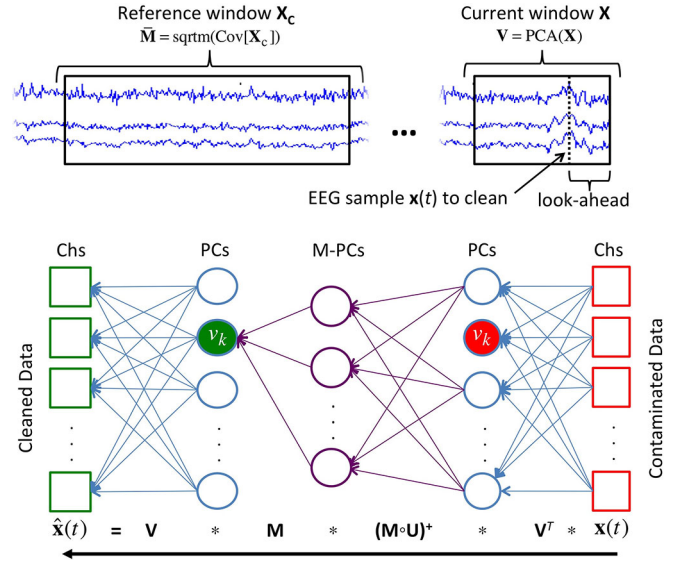

Fig. 3. The ASR method. High-variance artifacts (relative to a reference dataset or window) are identified and adaptively removed from the data using a series of linear subspace projections.

t, and let $X \in \mathbb{R}^{Q \times N}$ be a short sliding window of data containing $x(t)$. We apply principal component analysis decomposition to X and obtain components $V = [v_1 \ldots v_Q] \in \mathbb{R}^{Q \times Q}$. We then remove the subspace of "artifact" components whose short-window variance σ_k exceeds a (spatially varying) threshold $t(v_k)$, itself derived from X_c, and impute each removed component with a linear combination of activity of the remaining nonartifact components. Finally, we back-project components into channel space.

This sequence of operations is collapsed into a linear operator $R = VM(M \circ U)^+ V^T$ which is applied to $x(t)$ as $\hat{x}(t) = Rx(t)$. $M = V^T \bar{M}$ is the projected matrix square root of the covariance matrix C of X_c, such that $\bar{M}\bar{M}^T = C$. For improved robustness to artifacts in X_c, we estimate C using the ℓ_1 median [28] of samplewise covariance matrices $X_c X_c^T$ rather than the mean covariance.

The threshold operator $U \in \mathbb{R}^{Q \times Q}$ is chosen such that $U_{kl} = 0$, if $\sigma_k > t(v_k)$, otherwise $U_{kl} = 1$. Thresholds $t(v_k)$ are computed from reference data as follows: we first obtain principal components $W = [w_1 \ldots w_Q] \in \mathbb{R}^{Q \times Q}$, for X_c from C. Next, we obtain component activations $Y = X_c W^T$. For each component y_k we estimate root-mean-square amplitudes of successive overlapping short windows (e.g., half-second), as well as the robust mean m_k and standard deviation s_k of these values. Use of the median and median absolute deviation, respectively, typically yields good results. However, to support data with more than 50% of time windows affected by artifacts, an alternative maximum *a posteriori* (MAP) estimator is used. This fits a truncated exponential power distribution with a data-dependent prior that is based on EEG-specific heuristics. Given the robust per-component amplitude mean $m = [m_1 \ldots m_Q]$ and standard deviation $s = [s_1 \ldots s_Q]$, we estimate a vector of per-component thresholds $z = m + cs$ and threshold matrix $Z = \text{diag}(z)W^T$, where c is a tunable "cut-off" parameter, typ-

ically set between 5 and 7. The direction-dependent threshold is now simply $t(v) := \|Zv\|_2^2$. It is also possible to derive a usable threshold from C alone as a simpler approximation.

To attain real-time performance, a new filter R is estimated every ca. 100 ms (typically after a new signal block has been received from the hardware). EEG samples between any two updates of R are filtered by applying a raised-cosine blend $\tilde{R}$ of the two neighboring R operators as $\hat{x}(t) = \tilde{R}x(t)$. To further increase sensitivity to artifacts while decreasing sensitivity to natural high-amplitude brain signal components, the signals X and X_c can be spectrally reweighted using an eighth-order IIR filter designed to boost known artifact frequencies, mitigate frequencies of known high-amplitude brain idle rhythms, and suppress frequencies below what is captured by the sliding window. Note, however, that this filter is not applied to $x(t)$, and, therefore, does not affect the spectrum of the output signal. Generally, ASR signal processing is applied after high-pass filtering.

We also provide functionality for detecting and removing broken or otherwise corrupted data channels, based on channel correlation within a reference data segment (e.g., X_c above) using the RANSAC method presented in [29]. An FIR notch prefilter can be optionally enabled to suppress influence of the line noise on covariance. Missing or otherwise removed channels may then be spatially reconstructed from the activity of neighboring channels using a Gaussian spline function.

B. Distributed Source Reconstruction

Following ASR preprocessing, we apply inverse methods to a forward head model to infer source neuronal activity from the EEG data. We estimate the primary current source density (CSD) using a medium- to high-resolution (3,000—12,000 dipole) source space homogenously distributed over the cortical surface. Our forward model consisted of a four-layer (skull, scalp, cerebral spinal fluid, and cortex) boundary element method (BEM) model obtained through a nonlinear coregistration of the MNI "Colin 27" brain with the Cognionics HD-72 sensor montage [see Fig. 2(d)]. The BEM forward solution was computed using OpenMEEG [30]. We additionally segmented the cortical source space into (here, 90) ROIs using automated anatomical labeling (AAL) [31]. Arbitrary user-defined atlases are also supported. The pipeline makes use of modified routines and objects from the MoBILAB toolbox freely available online [32].

For inverse modeling, our framework supports several methods, including anatomically constrained low-resolution electromagnetic tomography (cLORETA) and regularized linearly constrained minimum-variance (LCMV) beamforming, which we utilize in this study.

1) Anatomically Constrained LORETA: cLORETA is well suited for real-time estimation, and automatically controls the level of regularization for each inverse solution. We briefly outline the procedure and refer to [33] for further details.

Let $X \in \mathbb{R}^{Q \times N}$ be a length-N sequence of EEG observations from Q electrodes. Let $S \in \mathbb{R}^{J \times N}$ be an unobserved matrix of

current density estimates for J sources. We adopt the conventional linear generative forward model

$$X = LS + \Upsilon \tag{1}$$

where $L \in \mathbb{R}^{Q \times J}$ is a forward (lead-field) matrix, and $\Upsilon \in \mathbb{R}^{Q \times N}$ is a zero mean i.i.d Gaussian sensor noise. Our objective is to obtain the MAP estimate of S given the Bayesian parameterization

$$p(S|X, \alpha, \beta) \approx p(X|S, \beta)p(S|\alpha). \tag{2}$$

Gaussian assumptions on the noise and prior yield the likelihood and prior densities

$$
\begin{aligned}
p(X|S, \beta) &= N(X|LS, \beta^{-1}I) \\
p(S|\alpha) &= N(0, \alpha^{-1}(H^T H)^{-1})
\end{aligned} \tag{3}
$$

where the hyperparameters β and α, respectively, express the precision (inverse variance) of the sensor observations and source estimates, and $H^T H$ is a sparse $J \times J$ precision matrix encoding prior variance-covariance assumptions on the sources. The entries of H also express anatomical constraints. For instance, anatomical regions that are extremely unlikely to contain an EEG generator may have their corresponding prior source (co)variances set to zero (by setting entries of H to infinity). Prior assumptions on nonzero source correlation structure (for instance, due to known interareal structural or functional connectivity) may also be encoded in H.

Given observation X, finding the MAP estimator of S reduces to solve a regularized least squares problem

$$
\begin{aligned}
S_{\text{MAP}} &= \underset{S}{\operatorname{argmin}} \ -\log p(S|X, \alpha, \beta) \\
&= \underset{S}{\operatorname{argmin}} \ -\log p(X|S, \beta)p(S|\alpha) \\
&= \underset{S}{\operatorname{argmin}} \ \|X - LS\|_F^2 + \lambda \|HS\|_F^2
\end{aligned} \tag{4}
$$

where $\|A\|_F^2 = tr(A^T A)$ denotes the square of the Frobenious norm of A and $\lambda = \alpha/\beta$ is a regularization parameter. The analytic solution of (4) is given by $S_{\text{MAP}} = WX$, where $W = (\lambda H^T H + L^T L)^{-1}L^T$. However, since the noise characteristics and, thus, the optimal value of λ, may change for different data segments X, this requires inversion of a $J \times J$ matrix for every X. Even for moderate J, this can be too costly for real-time application. To address this, we can express (4) in terms of the singular value decomposition (SVD) of the standardized lead field matrix, $LH^{-1} = U \operatorname{diag}(s_i)V^T$ for $i \in \{1 \dots Q\}$. This yields a more efficient estimator

$$S_{\text{MAP}} = H^{-1}V \operatorname{diag}\left(\frac{s_i}{s_i^2 + \lambda}\right) U^T X. \tag{5}$$

The SVD of LH^{-1} as well as the matrix $H^{-1}V$ need only be precomputed once, prior to online processing. An optimal value of λ is computed for a data block X by minimizing the generalized cross-validation function [34].

2) Regularized LCMV Beamforming: Alternatively, LCMV beamforming [35] attempts to learn an inverse solution $S_j = W_j^T X$ for each source $j \in \{1 \dots J\}$ by minimizing the

beamformer output power

$$\min_{W_s} tr\left(W_s^T C W_s\right) \qquad (6)$$

subject to a unity gain constraint $W_s L_s = I$. C is a channel covariance matrix, which in our implementation is regularized as $C = (1 - \lambda) XX^T + \lambda \left(tr\left(XX^T\right)/Q\right) I$, where λ is a small constant (here, we fix $\lambda = 0.001$). The solution to (6) is given by $W_s^T = \left(L_s^T C^{-1} L_s\right)^{-1} L_s^T C^{-1}$.

For the above inverse methods, the choice of block size N reflects a tradeoff between temporal stationarity assumptions on the source distributions and numerical stability of the inverse solution. Typical values may range from 100 ms to the length of a trial (1–2 s). Following current density estimation, we can compute spatially averaged, median, or maximal CSD for any subset of the AAL ROIs, which are then subjected to further analysis.

C. Sparse Dynamical System Identification

Having inferred source activity within a desired set of ROIs, we next model their multivariate spatiotemporal linear dynamics, including spectral power and functional or effective connectivity, using routines implemented in SIFT [36], operating as a BCILAB filter plugin. In brief, let $S = [s_1 \ldots s_N] \in \mathbb{R}^{Q \times N}$ be a Q-dimensional, zero mean, weakly stationary stochastic process of length N (e.g., data from Q ROIs or channels). Then, we model the linear dynamics of the state vector $s_t = [s_t^{(1)} \ldots s_t^{(Q)}]^T$ as a VAR process of order p

$$s_t = \sum_{l=1}^{p} B_l s_{t-l} + u_t. \qquad (7)$$

From the estimated VAR[p] model coefficients $\{B_1 \ldots B_p\}$ and the noise covariance matrix $\Sigma = \langle u_l u_l^T \rangle$, we may derive a number of dynamical measures, including spectral density, coherency, and multivariate Granger causality (see Chapter 1 and Supplementary Material of [5] for a detailed review).

In the pipeline described in this paper, S is a short segment of recent data, yielding a sliding-window VAR model. We note that a number of alternatives for time-varying VAR estimation exist, including Kalman or RLS filtering [37] and minimum-phase factorization of spectral matrices [38]. Implementations of linear and nonlinear Kalman filtering approaches are available in SIFT for use in an online pipeline.

1) Regularized Optimization Using Group Lasso: Equation (7) may be solved using a variety of unconstrained or constrained optimization methods [5]. However, for online applications, it is common for the number of model parameters to significantly exceed the number of data samples, i.e., $Q^2 p > QN$. Then, the solution to (7) is underdetermined, and additional model constraints (i.e., regularization) must be imposed in order to obtain a unique solution. A common approach is to impose various nonuniform prior distributions over the VAR parameters [39]. Typical choices include the Gaussian, as in Tikhonov regularization; Laplacian, as in the Lasso; or a combination of both, as in Group Lasso or Elastic Net. Alternatively, generalized Gaussian priors can be employed, as in (block) sparse Bayesian learning

[40]. We refer to [41] for an excellent assessment of regularization methods for accurate parameter estimation of highly underdetermined VAR models, such as in this paper.

The framework supports several of these regularization approaches. In this paper, we follow previous work [41], [42] and employ the Group Lasso (*sum-of-norms*) penalty [43] to solve (7). This assumes the source-level dynamical system has a globally sparse topology (few nonzero interactions between brain regions), with smooth (jointly Gaussian) transfer functions, ensuring preservation of important spectral properties, including positive definiteness of spectral densities.

To apply the regularization, we first transform the VAR[p] problem of (7) into a VAR[1] problem

$$\mathbf{Y} = \mathbf{SB} \qquad (8)$$

where $\mathbf{B} = [B_1 \ldots B_p]^T$ denotes a matrix of all VAR[p] coefficients, and with multivariate data matrices $\mathbf{S} = [S_1 \ldots S_p]$ and $\mathbf{Y} = S_0$, where $S_l = [s_{p+1-l} \ldots s_{N-l}]^T$ are delay-embedded time series. We obtain a unique solution to (8) with respect to $\mathbf{B}$ by minimizing a global cost function

$$\hat{\mathbf{B}} = \underset{\mathbf{B}}{\mathrm{argmin}} \left(\|\mathrm{vec}\left(\mathbf{Y} - \mathbf{SB}\right)\|_2^2 \right.$$
$$\left. + \lambda \sum_{i \neq j} \left\|\mathbf{B}_{1,(ij)} \ldots \mathbf{B}_{p,(ij)}\right\|_2 \right). \qquad (9)$$

Here, $\{\mathbf{B}_{1,(ij)} \ldots \mathbf{B}_{p,(ij)}\}$ are the VAR filter coefficients expressing dynamical interactions from process j to i. The regularization parameter λ determines the relative tradeoff between the model prediction error and the Group Lasso penalty and reflects a prior assumption on the degree of sparsity of the system (or similarly, the noise variance).

We note that the assumption of sparse functional connectivity in brain space has biological plausibility [44]–[46]. Numerical simulations additionally suggest that taking into account the group structure of VAR parameters (i.e., Group Lasso) can improve system identification over assuming unstructured sparsity (i.e., Lasso) [42]. Furthermore, Group Lasso aims to shrink nonsignificant parameter estimates exactly to zero, performing implicit feature selection. Since resulting connectivity tensors are sparse, this facilitates the use of sparsity assumptions in later classification and prediction stages. Conversely, assuming a (smooth) Gaussian prior guarantees strictly nonzero (if small) parameter estimates, and connectivity graphs may require post hoc statistical thresholding for interpretation.

2) ADMM Algorithm: Minimization of (9) may be achieved using a range of methods, including second-order cone programming with an active set solver [42] or the dual-augmented Lagrangian method [47]. We propose to use the ADMM, a flexible and efficient iterative framework for distributed convex optimization and parameter estimation [48]. In general, ADMM solves problems of the form

$$\text{minimize } f(x) + g(z)$$
$$\text{subject to } Ax + Bz = c \qquad (10)$$

where $x \in \mathbb{R}^n$, $z \in \mathbb{R}^m$, $A \in \mathbb{R}^{P \times m}$, $B \in \mathbb{R}^{p \times m}$.

In "scaled form," optimization consists of the following iterations:

$$x^{k+1} = \underset{x}{\operatorname{argmin}} \left(f(x) + \frac{\rho}{2} \left\| Ax + Bz^k - c + u^k \right\|_2^2 \right) \quad (11)$$

$$z^{k+1} = \underset{z}{\operatorname{argmin}} \left(g(z) + \frac{\rho}{2} \left\| Ax^{k+1} + Bz - c + u^k \right\|_2^2 \right) \quad (12)$$

$$u^{k+1} = u^k + Ax^{k+1} + Bz^{k+1} - c \quad (13)$$

where $\rho > 0$ is a *penalty parameter* and (11)–(13) are, respectively, x-minimization, z-minimization, and scaled dual variable updates.

In the case of Group Lasso, defining $b = \operatorname{vec}(\mathbf{Y})$, $x = \operatorname{vec}(\mathbf{B})$, and $\mathbf{A} = \mathbf{S} \otimes I_{Q \times Q}$, the minimization problem of (10) can be stated as follows:

$$\text{minimize } f(x) + g(z)$$

$$\text{subject to } x - z = 0 \quad (14)$$

where $f(x) = (1/2) \| b - \mathbf{A}x \|_2^2$ and $g(z) = \lambda^* \sum_{i=1}^{Q^2} \| z_i \|_2$ with scaled regularization parameter $\lambda^* = \lambda/2$ and where $z_q = [\mathbf{B}_{1,(ij)} \ldots \mathbf{B}_{p,(ij)}]$ is the vector of VAR coefficients for the qth pair of processes $i, j \in \{1 \ldots Q\}$. Note that $f(x)$ is the prediction error, while $g(z)$ is the Group Lasso regularization penalty.

The corresponding ADMM iterations are then as follows:

$$x^{k+1} = \left(\mathbf{A}^T \mathbf{A} + \rho I \right)^{-1} \left(\mathbf{A}^T b + \rho(z^k - u^k) \right)$$

$$z_i^{k+1} = \mathbf{S}_{\lambda^*/\rho} \left(x_i^{k+1} + u^k \right)$$

$$u^{k+1} = u^k + x^{k+1} - z^{k+1} \quad (15)$$

with vector *soft-thresholding* operator $\mathbf{S}_\kappa(a) = \max \left(0, 1 - \kappa / \| a \|_2 \right) a$. Convergence is achieved when the following criteria are met:

$$r^{\mathrm{pri}} = \left\| x^k - z^k \right\|_2 < \varepsilon^{\mathrm{pri}} \text{ and}$$

$$r^{\mathrm{dual}} = \left\| \rho \left(z^k - z^{k-1} \right) \right\|_2 < \varepsilon^{\mathrm{dual}} \quad (16)$$

where $\varepsilon^{\mathrm{pri}}$ and $\varepsilon^{\mathrm{dual}}$ are stopping criterion limits which may be defined absolutely, or relative to the norms of z, x, and u.

3) Warm Starting: The iterations of (15) can be "warm started" by initializing z and u with suitable values, for instance, a previously obtained solution to a similar problem. This can substantially reduce the number of iterations needed for convergence. In this study, we warm start ADMM for a given time window using the solution obtained for the previous time window.

4) Selection of Regularization Parameter λ: A suitable value for λ is often obtained through minimizing an objective value such as cross-validated prediction errors. However, since cross validation is not readily applicable for online inference with nonstationary data, we utilize heuristic approaches for adapting λ online.

Following [48], we may heuristically define lambda as a fraction of the critical value of λ for which $x = 0$ (i.e., the sparsest possible solution): $\lambda^{\mathrm{opt}} = \kappa \lambda^{\mathrm{max}}$, where $\kappa \in [0, 1]$ and $\lambda^{\mathrm{max}} = \max_i \| \mathbf{A}_{(i)}^T b_{(i)} \|_2$, where $\mathbf{A}_{(i)}$ and $b_{(i)}$ are regressors and regressands for the ith VAR coefficient group.

Alternatively, we propose a simple adaptive approach to select λ based on convergence properties of the ADMM algorithm. We initialize the iterations in (15) with a relatively large heuristic value for λ, corresponding to a strong sparsity assumption. If the absolute change in residual norms r^{pri} and r^{dual} in (16) remain below a predetermined threshold for a predetermined number of iterations, then we divide λ by a constant factor (e.g., 10). This process is repeated, thereby gradually relaxing the sparsity constraint, until convergence is accelerated (e.g., the gradient of residual norms is sufficiently large). While this by no means guarantees the "true" or optimal value for λ will be found in a statistically principled sense (only one that ensures rapid ADMM convergence), we find that, in practice, this yields reasonable VAR solutions, while accelerating convergence. In this study, we use this approach.

We also note that we exploit several additional optimizations, including an adaptive update scheme for the penalty parameter ρ (see [48, Sec. 3.4.1]) and caching factorizations of the coefficient matrix $F = \mathbf{A}^T \mathbf{A} + \rho I$ (see [48, Sec. 4.2.3]). Note that when $\mathbf{A}$ is "fat" (wide), rather than "skinny" (tall), a more efficient factorization may be carried out by applying the matrix inversion lemma to the x-update in (15) as in [48, Sec. 11.1.1]. Finally, we exploit the sparse block-Toeplitz structure of the data matrix $\mathbf{A}$ for much more efficient iterative operations on reduced submatrices. We refer the interested reader to [48] for further details on the ADMM method and its application to Group Lasso.

5) Model Order Selection, Validation, and Power and Connectivity Estimation: In our framework, the VAR model order can be automatically selected by minimizing information criteria (e.g., AIC or BIC), either online or on offline calibration data. Alternatively, one may just set the model order to a reasonably high value and allow the Group Lasso regularization to select a suitably parsimonious submodel by shrinking uninformative coefficients.

Following model fitting and (optional) tests of model stability and residual whiteness (autocorrelation function or Portmanteau), we may obtain the spectral density matrix and any of (to date) over 15 frequency-domain functional and effective connectivity measures implemented in SIFT. These include ordinary and partial coherence, Granger–Geweke causality, and several related multivariate causality measures including several variants of PDC, the directed transfer function, and the direct directed transfer function [5].

The connectivity estimates take the form of a tensor $C \in \mathbb{R}^{Q \times Q \times T \times F}$, where Q is the number of sources, ROIs, or channels, T is the number of overlapping time windows within a data chunk or trial, and F is the number of selected frequencies. We note that tensor diagonals $C_{ii,:}$ reflect autoconnectivity measures, which can be regarded as the fraction of a source's variance (power) that cannot be explained by causal inputs from other measured sources. This can also be interpreted as a measure of a processes' autonomy within a complex system [49]. We also note that the framework allows for graph-theoretic measures [50], such as degree, flow, and asymmetry ratio to be easily applied to connectivity matrices, although we do not study these here. The various measures may then be directly visualized,

transmitted (e.g., via LSL), or stored for research or monitoring purposes. They may also be subsequently used by BCILAB as features for classification or prediction an individual's state (e.g., behavioral, cognitive, or affective state) within a BCI.

D. Connectivity-Based Classification With ProxConn

To learn robust BCI-relevant predictive models on a high-dimensional multivariate (e.g., connectivity-based) feature space from only a few trials, strong prior assumptions are required. We developed a method, which we refer to as *ProxConn*, consisting of applying regularized logistic or linear regression to log-transformed time/frequency (T/F) connectivity measures (yielding a 4-D feature tensor across pairwise connectivity, time, and frequency). The regularization simultaneously employs a sparsifying $\ell_{1,2} + \ell_1$ norm with one group for each connectivity edge, containing its associated T/F weights, plus two trace norm terms to couple the T/F weights for all out-edges of a node and all in-edges of a node, plus an ℓ_2 smoothness term across time and frequency, respectively.

More formally, single-trial tensors C of log-transformed connectivity features are classified with binary label y by a generalized linear model with logistic link function

$$q_\theta(Y = y|C) = \frac{1}{1 + e^{-y(\langle C, \theta \rangle + b)}}, y \in \{-1, +1\}. \quad (17)$$

The weight tensor θ (of same dimensionality as C) and unregularized bias b are learned in a jointly convex optimization problem of the form

$$\{\theta^*, b^*\} = \underset{\theta, b}{\operatorname{argmin}} \left(\begin{array}{l} \frac{1}{N\lambda_D} \sum_t \log\left(1 + e^{-y_t(\langle C_t, \theta \rangle + b)}\right) \\[2mm] + \lambda_1 \sum_{i=1}^M \sum_{j=1}^M \|\theta_{i,j}\|_2^2 \\[2mm] + \lambda_2 \|\theta\|_1 \\[2mm] + \lambda_3 \left(\sum_{j=1}^M \|\theta_{:,j}\|_* + \sum_{i=1}^M \|\theta_{i,:}\|_* \right) \\[2mm] + \lambda_4 \|\mathcal{D}_{tf}\theta\|_2^2 \end{array} \right) \quad (18)$$

where

$\theta_{i,j}$ denotes the $T \times F$ matrix of time/frequency weights for connectivity $j{\to}i$;

$\theta_{i,:}$ denotes the $[Q{-}1] \times TF$ matrix of *inflow* weights for node i;

$\theta_{i,:}$ denotes the *outflow* weights for node j;

$\mathcal{D}_{tf}$ is a time/frequency finite difference operator enforcing T–F smoothness;

$\|x\|_*$ denotes the trace norm of x, and λ_D and $\{\lambda_k\}$ are respective regularization parameters for data loss and constraint terms.

We perform minimization of (18) via consensus ADMM with proximal splitting [48]. Regularization parameters are typically learned via nested cross validation, although in practice we may heuristically set $\lambda_k = 1$ for some k. We note that simpler or more complex variations of (18) may also be used, depending

$$x_1(t) = 0.9x_1(t-1) + 0.3x_2(t-2) + \varepsilon_1(t),$$
$$x_2(t) = 1.3x_2(t-1) - 0.8x_2(t-2) + \varepsilon_2(t),$$
$$x_3(t) = 0.3x_1(t-2) + 0.6x_2(t-1) + \varepsilon_3(t),$$
$$x_4(t) = -0.7x_4(t-3) - 0.7x_1(t-3) + 0.3x_5(t-3) + \varepsilon_4(t),$$
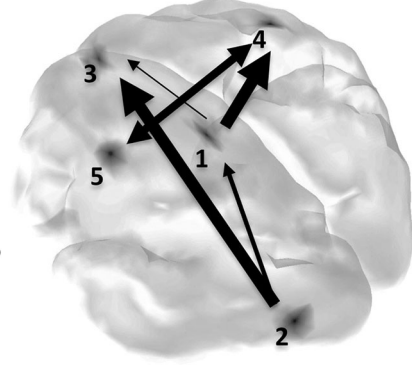$$x_5(t) = 1.0x_5(t-1) - 0.4x_5(t-2) + 0.3x_4(t-2) + \varepsilon_5(t),$$

Fig. 4. EEG simulation ground truth (upper): VAR [3] dynamical equations. Gaussian source patches and directed connectivity graph (lower). Line width reflects peak connectivity strength across frequencies.

on the specific application. For continuous target variables *y*, we simply replace the logistic link function (17) with a linear link function.

E. Real-Time Visualization

The proposed framework supports interactive real-time visualization of time series and estimated dynamical measures. This includes 2-D plots of raw and cleaned EEG channel or CSD time series, power spectra, as well as 4-D rendering of time-varying connectivity, graph-theoretic metrics, source distributions, power, etc., within a 3-D model of the head and brain. Pipeline elements can be enabled/disabled "in flight" using a graphical user interface.

F. Data Collection and Analysis Pipeline

Next, we describe validation of the above pipelines for 64-channel simulated EEG data as well as real 64-channel task data collected using Cognionics HD-72 hardware.

1) Simulated Data: To test the ability of our pipeline to accurately reconstruct source dynamics and connectivity in real time, we generated a 5-D VAR [3] system of coupled oscillators as described in [51, eq. 3.1]. This comprised the CSD time series of five sources positioned on a 3571-vertex cortical mesh. Each source had a Gaussian spatial distribution ($\sigma = 5\,\text{cm}$) with mean equal to the centroid of each of the following AAL ROIs (respectively): x_1: left middle cingulate gyrus, x_2: left middle occipital gyrus, x_3: right medial superior frontal gyrus, x_4: right precentral gyrus, x_5: left precentral gyrus. The system is depicted in Fig. 4. We generated 2 min of source time-series data (Sampling rate $= 300$ Hz) and projected this through the realistic forward model described in Section II-B to produce 64-channel EEG data. Gaussian i.i.d sensor noise was added with a signal to noise ratio of $\sigma_{\text{data}}/\sigma_{\text{noise}} = 5$. The simulated EEG data were streamed to an online BCILAB pipeline. cLORETA

was applied using a 32-sample block size. Median CSD was computed for the five ROIs $\{x_1 \ldots x_5\}$. A group-sparse order 3 VAR model was fit to normalized ROI time series via ADMM within a 1-s sliding window. We used an initial Group Lasso regularization parameter $\lambda = 0.1$ with online heuristic adaptation. Spectral density and PDC [52] were obtained from 1–65 Hz. Finally, the max operator was applied to PDC across frequency producing a 2-D connectivity matrix.

2) Real Data: To test real-world utility of our pipeline for BCI applications, we sought to detect behavioral response errors from single-trial cortical connectivity features. Univariate features, such as event-related potentials (ERPs), are known to perform well on this task, providing a competitive benchmark [53]. However, to our knowledge, effective connectivity features have not been used in this paper.

a) Data collection and task: Dry EEG data (Cognionics HD-72) were collected from nine right-handed male subjects ages 22–46, with no history of neurological disorders. Data were collected at the Swartz Center for Computational Neuroscience, UCSD under IRB approval. Each subject performed a modified Eriksen Flanker task with a 133 ms delay between flanker and target presentation [54]. Flanker tasks have been extensively studied and are known to produce error-related negativity (ERN, Ne) and error-related positivity (P300, Pe) ERPs following error commission [55], as illustrated in Fig. 9. The experimental session lasted on average (±std. dev.) 13.67 ± 0.54 min. The mean response time (following target presentation) was 179.3 ms ± 38.4 ms for error trials and 262.2 ms ± 21.6 ms for correct trials. To reduce risk of classification bias due to class imbalance, correct trials were subsampled uniformly at random to yield a 3/1 ratio to error trials. Across nine subjects, this yielded, on average, 51 ± 11.2 error trials and 153 ± 33.6 correct trials for a total average of 204 ± 44 trials.

3) Modeling Pipeline: Continuous EEG data were subjected to a BCILAB + SIFT pipeline, consisting of preprocessing, source reconstruction, neuronal system identification, and behavioral response classification. In this section, we outline each of these steps.

a) Preprocessing: Our online pipeline included the following preprocessing elements (in order of application): downsampling to 128 Hz, drift correction with 0.1–1 Hz transition high-pass filter, bad channel removal and ASR (cut-off parameter $c = 7$, sliding window length 0.5 s), common average referencing, and 45–50-Hz transition low-pass filtering. All filters were minimum-phase FIR. Single trial epochs centered at −0.6 to 1.6 s relative to button press events were then extracted for subsequent analysis.

b) Source Reconstruction: A distributed cortical inverse solution was obtained for each 2.2-s trial using (independently) cLORETA and LCMV. CSD was averaged within each of ten cortical ROIs constructed from AAL atlas parcels (see Fig. 5). ROIs were selected based on a literature review implicating them in visual sensory input, motor output, and error processing [56], and a prior study [57] indicating error-related connectivity changes in these regions. These consisted of Left+Right anterior cingulate cortex (ACC), Left+Right Middle Cingulate Cortex (MCC), Left+Right Pos-
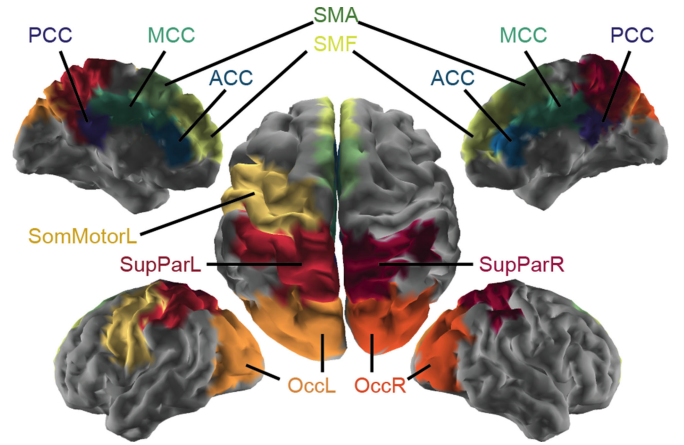

Fig. 5.   Ten cortical ROIs used for the real data analysis.

terior Cingulate Cortex (PCC), Left+Right Supplementary Motor Area (SMA), Left+Right Superior Medial Frontal, Left Precentral+Postcentral (SomMotorL), Left Mid+Sup+Inf Occipital (OccL), Right Mid+Sup+Inf Occipital (OccR), Left Superior+Mid Parietal (SupParL), Right Superior+Mid Parietal (SupParR).

For each trial, an order 15 time-varying sparse VAR model was fit, using ADMM, to the ten ROI CSD. We used a 660-ms sliding window with a step size of 50 ms. The sliding window length was chosen to span at least one cycle of our lowest frequency of interest (2 Hz). At a sampling rate of 128 Hz, this yielded 84 multivariate data samples for fitting $10^2 \times 15 = 1500$ VAR parameters within a window. We used an initial Group Lasso regularization parameter $\lambda = 0.1$ with online heuristic adaptation. From the model coefficients, we obtained the sdDTF [58], which can be regarded as a multivariate, frequency-domain analogue to Granger Causality. The measure at frequency f and time t is given by

$$\eta_{ij}^2(f,t) = \frac{|H_{ij}(f,t)|^2 \, |P_{ij}(f,t)|^2}{\sum_{klf\tau} |H_{kl}(f,\tau)|^2 \, |P_{kl}(f,\tau)|^2} \quad (19)$$

where $H(f,t)$ is the VAR transfer matrix and $P(f,t)$ is the partial coherence. We estimated sdDTF over the range 1–15 Hz. The frequency range was based on a prior study by the first author, which found significant sdDTF connectivity differences within this range between error and correct response conditions in a error-generating two-back task [57]. Additionally, evidence suggests that theta (4–7 Hz) and delta (2–3 Hz) medial-frontal cortical activity are related to error processing and conflict monitoring [55], [59], [60].

c) Behavioral response classification and performance evaluation: ProxConn regularized logistic regression models were trained on standardized log-transformed sdDTF time–frequency features (cross- and auto-connectivity) from labeled trials with label mapping Error $\rightarrow$ +1 and Correct $\rightarrow$ −1. Model evaluation and hyperparameter search was performed using a nested fivefold blockwise cross validation, with a five-trial margin between consecutive blocks to cleanly separate testing and training data. For each fold, we measured the area under the
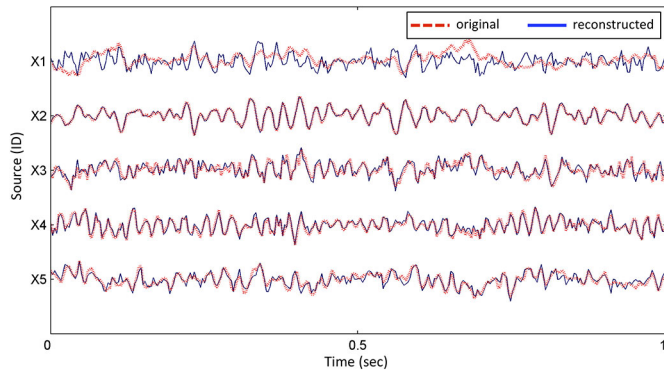
Fig. 6. Comparison of true (red, dashed) versus reconstructed (blue, solid) CSD (cLORETA) for a 1-s segment of our five simulated ROIs. Time series are normalized to unit variance.
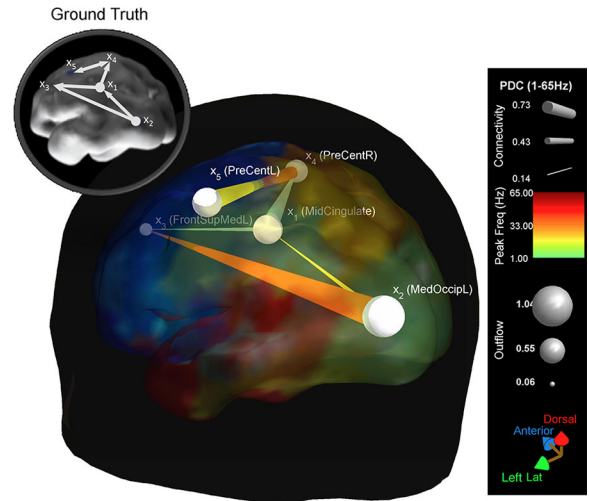

Fig. 7. BrainMovie3D frame showing source networks reconstructed online. Here, edge color denotes preferred coupling frequency, while edge size and tapering, respectively, denote coupling strength (PDC) and directionality at that frequency. PDC is thresholded at the common heuristic level of 0.1.
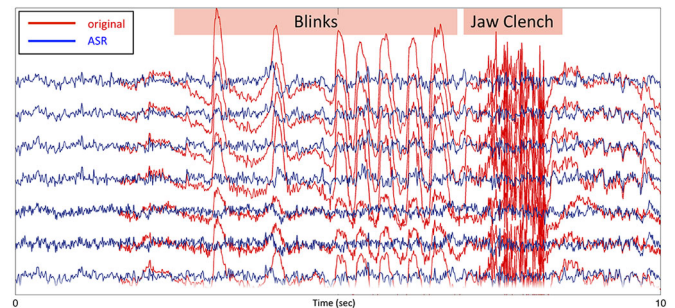

Fig. 8. 10 s of EEG data following ASR data cleaning (blue trace) superimposed on original data (red trace).

receiver operating characteristic curve (AUC) with respect to trial class predictions. The regularization hyperparameter for the ProxConn data term was searched via an inner (nested) five-fold blockwise cross validation over the range $2^{\{3, 2.34...-7.56\}}$. To reduce computation time, the weights of the additional regularization terms were set to 1. For further details on (nested) cross validation, we refer the reader to [61].

In order to benchmark the ProxConn classifier against a conventional approach, we also applied a state-of-the-art first-order ERP classification method: dual-spectral regularized logistic regression (DSLR) [62]. This was applied separately to single-trial evoked responses from the ten ROIs, as well as preprocessed data from 64 channels. The epoch window and ROI CSD estimates were identical for ProxConn and DSLR approaches. DSLR evaluation and regularization parameter selection (over the range $2^{\{-3, -2.75...4\}}$) was carried out using the aforementioned 5×5 nested blockwise cross-validation approach.

III. RESULTS

A. Simulation Data

Fig. 6 shows a 1-s segment of cLORETA estimated CSD superimposed on the true CSD. Superficial sources were accurately recovered, while the deep tangential source (X_1; midcingulum) was somewhat more noisily reconstructed.

Fig. 7 shows the reconstructed source network for a representative time window, using our BrainMovie3D visualizer. Ground truth is displayed in the inset. Over all time windows, the connectivity graph was recovered with high accuracy—the area under ROC curve (AUC), averaged over time windows, was $0.97 + / - 0.021$. Peak coupling frequency and relative strength were also correctly recovered.

B. Real Data

1) Data Quality and Artifact Rejection: Fig. 8 shows a representative segment of EEG data contaminated by blink and muscle artifacts, before and after ASR artifact removal. High-variance artifacts were removed.

Fig. 9 shows single-trial EEG data (subject 8) for response-locked error trials at electrode FCz. Trials are sorted by reaction time. Although acausal filters cannot be used online, for this plot alone, in order to accurately assess ERP latencies, all filters were zero phase (acausal). We ran the analysis with and without ASR (the latter shown here) and confirmed that ASR did not distort ERPs (see Fig. 9, red trace). Note that nearly every trial shows a visual evoked response to the stimulus as well as prominent Ne and Pe following the erroneous button press. The scalp topography of the Ne (upper left) has a frontocentral distribution centered at FCz, as expected for a mid/anterior cingulate or frontal midline generator. Encouragingly, the quality of the evoked responses is comparable to that reported using research-grade gel-based EEG systems.

2) Classification Results: Table I shows individual subject and group averaged fivefold CV performance for classifying erroneous versus correct responses using sdDTF connectivity features (ProxConn) and single-trial ERP (DSLR) features, using either LCMV or cLORETA source localization. Performance was measured using the area under the receiver operating characteristic curve (AUC). Chance AUC is 0.5.

Application of ProxConn to cLORETA sources yielded a group mean AUC of 0.74 ± 0.09 (max: 0.87 ± 0.08),
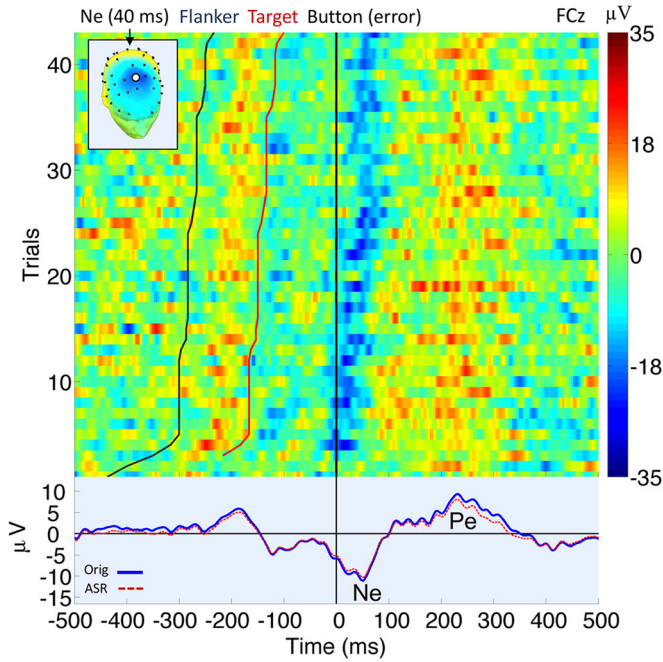
Fig. 9. Representative ERPImage (subject 8) showing single-trial EEG potentials (no smoothing) at FCz for response-locked error trials, sorted by latency of response to target onset (red sigmoidal trace). Responses occur at 0 ms (vertical line). The bottom panel shows the averaged ERP without ASR in blue, and the ERP with ASR enabled in red.

TABLE I
FIVEFOLD CV AREA UNDER ROC CURVE (MEAN ± STD)

Subj	LCMV		cLORETA	
	DSLR	ProxConn	DSLR	ProxConn
1	0.9682 ± 0.02	0.7923 ± 0.14	0.8909 ± 0.09	0.8203 ± 0.16
2	0.8132 ± 0.11	0.6609 ± 0.10	0.6101 ± 0.18	0.6150 ± 0.14
3	0.8047 ± 0.06	0.7164 ± 0.09	0.6497 ± 0.10	0.7682 ± 0.10
4	0.7809 ± 0.13	0.7425 ± 0.07	0.5895 ± 0.08	0.7967 ± 0.05
5	0.5693 ± 0.17	0.5792 ± 0.12	0.5000 ± 0.00	0.6345 ± 0.08
6	0.9434 ± 0.03	0.7228 ± 0.08	0.8717 ± 0.08	0.6714 ± 0.09
7	0.8524 ± 0.04	0.7142 ± 0.10	0.8548 ± 0.08	0.8029 ± 0.04
8	0.8934 ± 0.03	0.8848 ± 0.08	0.9713 ± 0.02	0.8653 ± 0.08
9	0.7882 ± 0.14	0.6936 ± 0.18	0.7110 ± 0.16	0.6657 ± 0.06
Avg	**0.8237 ± 0.12**	**0.7229 ± 0.08**	**0.7388 ± 0.16**	**0.7378 ± 0.09**
pval	0.000033	0.000046	0.002316	0.000053

Mean ± standard deviation of Area under Receiver Operating Characteristic Curve (AUC) for individual subject 5-fold cross-validation, as well as group averages. Shaded cells denote results that did not significantly exceed chance AUC of 0.5. Pval denotes p-values for one-sided t-test against the null hypothesis that group mean does not differ from chance.

Mean ± standard deviation of area under receiver operating characteristic curve (AUC) for individual subject fivefold cross validation, as well as group averages. Shaded cells denote results that did not significantly exceed chance AUC of 0.5. Pval denotes *p*-values for one-sided *t*-test against the null hypothesis that group mean does not differ from chance.

significantly higher than chance. This did not significantly differ from group means obtained for ProxConn on LCMV sources, and for DSLR on cLORETA sources. However, substantial differences in within-subject performance across the methods tested were observed. Application of DSLR to LCMV sources yielded significantly better mean performance (AUC = 0.82 ± 0.12). Compared to DSLR, ProxConn showed less variance in

performance across subjects, and a greater proportion of subjects exceeded chance performance. DSLR applied to 64-channel sensor data yielded a group mean AUC of 0.88 ± 0.08.

Given the comparatively low dimensionality and saliency of error-related ERP features (c.f., Fig. 9), it is not surprising that the DSLR method can perform quite well. We note however, that time-domain evoked response methods can only be used to detect, not predict, events, and generally require reliable event indicators, around which to extract phase-locked ERP features. In many real-world applications, these requirements cannot be met, and alternative methods such as ProxConn may be attractive.

3) Real-Time Performance: Once a ProxConn model is trained, the presented system runs online with real-time performance on typical computing platforms. We simulated online application of the above ProxConn error-detection pipeline to streaming Flanker task data from subject 8 on a 4-core 2.4 GHz AMD Opteron PC. Compute time (including preprocessing, source localization, connectivity feature extraction, and classification) was 438 ms per second of data (2.26× real time). We have demonstrated parallelized acceleration of several components of this pipeline using graphical processing units [63]. This also allows higher dimensional models to be estimated with minimal increase in computation time. Note that for neuroimaging applications, pretraining of a classification model is not a requirement.

IV. DISCUSSION

The combination of wearable, mobile EEG, and real-time neuroimaging and cognitive state classification offers opportunities to study the human brain in action. As noted in Section I, and reviewed in [16], an increasing number of studies have applied source connectivity methods to EEG data. However, these typically leverage multitrial ensembles of data or other offline processing steps. In contrast, the pipelines presented in this paper focus on measuring brain dynamics at the level of single trials and are capable of online real-time operation. While such capabilities may not be a prerequisite for scientific study of the brain, they are required for many practical real-world neurotechnology applications. These range from clinical neuroimaging and BCI [16], [64]–[66], to neuroergonomics [67], [68], and extending to diverse general-purpose applications [4].

We reiterate that the presented BCILAB + SIFT system is not the first or only software solution for single-trial source connectivity analysis and/or cognitive state classification. For instance, eConnectome [23] offers routines for adaptive connectivity estimation and visualization from continuous data. Billinger *et al.* [19] presented a system for single-trial connectivity analysis and state classification, subsequently made available in the SCoT toolbox [24]. As with SIFT, these toolboxes leverage a VAR representation of system dynamics and offer a range of connectivity measures. However, there are also many significant differences with the presented system ranging from software design, to the breadth and type of methods offered, to online and/or real-time capabilities (to our knowledge not available in other systems). For instance, SCoT focuses on ICA-based source separation (not localization) using pretrained
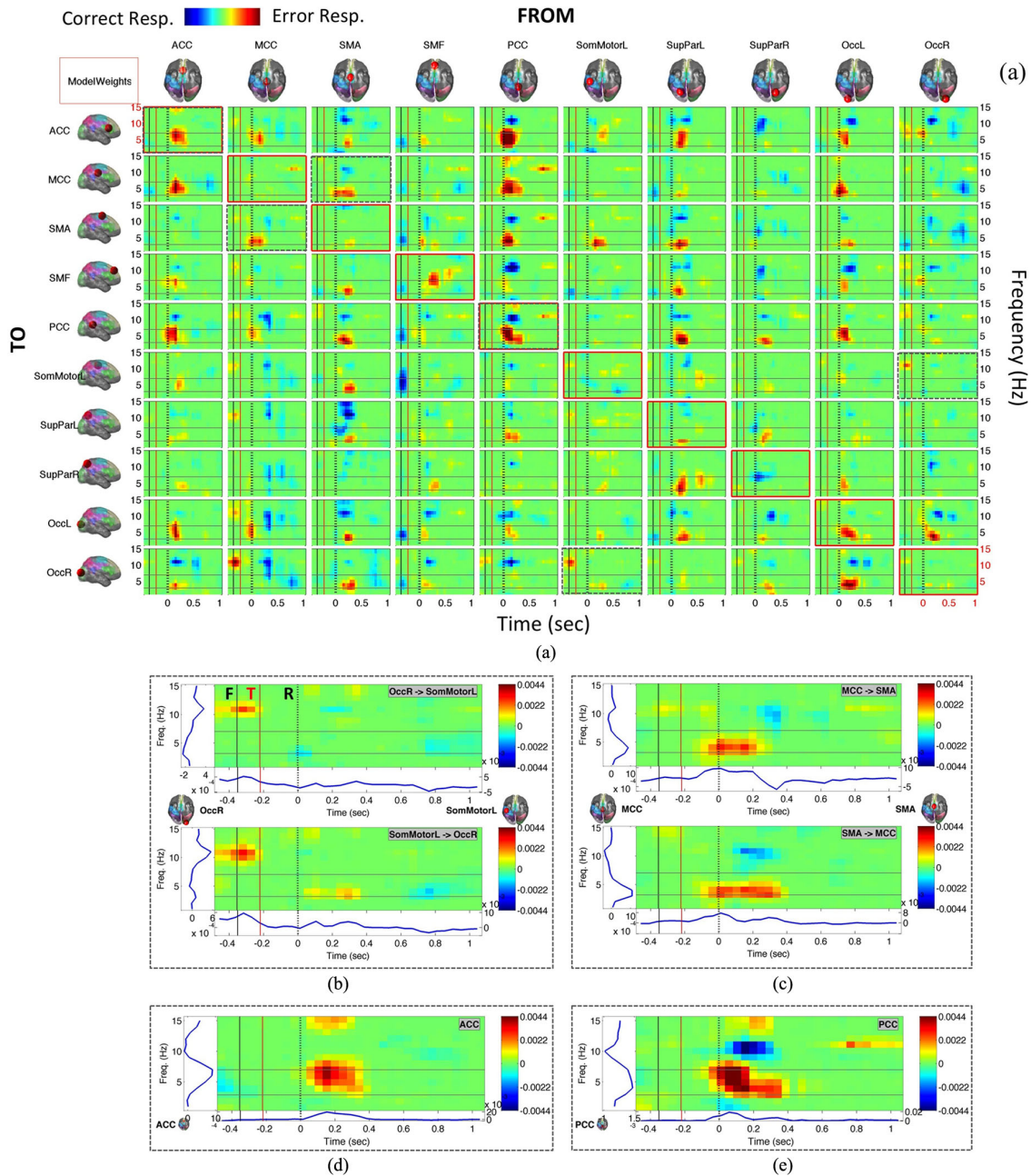
Fig. 10. (a) Time–frequency grid representation of cLORETA + sdDTF ProxConn classifier weights for subject 8. Each cell shows sdDTF from the respective column ROI to row ROI across time (*x*-axis) and frequency (*y*-axis). Cortical surfaces for Colin27 template brain are shown on row and column headers with color-coded ROI spatial extent and ROI centroid (red dot). Warm (cool)-colored pixels indicate that pairwise time–frequency sdDTF contributed to classification of error (correct) behavioral responses. Markers F (black solid) and T (red solid) denote mean latency of Flanker and Target presentation, respectively. Marker R (black dashed) denotes mean behavioral response latency. Time axis reflects VAR sliding window centers, corrected to account for online pipeline delay (ASR, causal filters) of ∼263 ms. Horizontal markers are placed at 3 and 7 Hz. Panels (b) and (c) detail sdDTF interactions between representative cortical ROI pairs: Response errors are associated with (b) preresponse alpha-band connectivity between OccR and SomMotorL and (c) early peri- and postresponse theta-band sdDTF between MCC and SMA. Panels (d) and (e) detail sdDTF autoconnectivity within representative ROIs: theta-band sdDTF within ACC and PCC are associated with errors. (e) Postresponse alpha sdDTF in PCC is associated with correct responses.

spatial filters, while this paper presents online distributed corti-cal localization methods. While a detailed comparison is beyond the scope of this paper, we encourage the reader to explore these and other software solutions.

A recent trend in the neurosciences is the biological interpre-tation of weight vectors or corresponding pattern vectors from classifier models trained, using neuronal data features, to dis-

criminate between experimental conditions [69], [70]. While caution should be exercised in overinterpreting such weights [70], the ProxConn regression approach may likewise yield insight into source-level networks predictive of cognitive and behavioral states.

As a demonstrative example, Fig. 10(a) depicts a "time–frequency grid" plot of ProxConn classifier weights for

subject 8. To obtain a single weight vector, ProxConn was applied to all single trials (no outer CV) following application of the cLORETA + sdDTF pipeline variant reported for Table I. Here, the ProxConn regularization terms of (18) were searched via fivefold blockwise CV over 25 parameter combinations sampled uniformly from the distribution $2^{N(1,\sqrt{2})}$. For each sdDTF time–frequency-pair estimate, ProxConn yields a real-valued weight. Its amplitude and sign can be interpreted as that feature's fractional contribution in discriminating between classes (e.g., error versus correct response). ProxConn's ℓ_1 regularization promotes shrinkage to zero of weights for uninformative features.

We note a pattern of preresponse alpha/mu (8–12 Hz) and postresponse theta-band (3–7 Hz) connectivity being associated with errors (warm colors) while postresponse alpha/mu and beta (13–15 Hz) connectivity were associated with correct responses (cool colors). Error-related theta connectivity was prominent within and between a number of ROIs, including ACC, MCC, and SMA [see Fig. 10(c)–(d)]. Theta power and connectivity modulation in these regions has been linked to error processing and conflict monitoring [5], [55], [59], [60]. PCC [see Fig. 10(e)], SupParL, and OccL also showed significant error-related theta bursts. Preresponse alpha connectivity between OccR and several ROIs, including somatomotor cortex [see Fig. 10(b)] was also associated with errors. In prior studies, prestimulus alpha in occipital and sensorimotor regions has been shown to predict subsequent response errors [71].

These results demonstrate the feasibility of recovering meaningful single-trial source connectivity features from dry-electrode EEG, which can be used to detect or predict cognitive state and behavior. To our knowledge, this is the first demonstration of single-trial behavioral error detection using cortical effective connectivity measures. However, since event locking is not required for VAR-based feature extraction, they may have greater use where traditional event-locked analyses (e.g., ERP or ERD/ERS) cannot be applied; for instance, to predict future behavior from ongoing EEG activity. Finally, we note these methods have broad applications outside cognitive monitoring, including detection or prediction of neuropathologies, such as epileptic seizures [64], [66].

V. CONCLUSION

In this paper, we presented and evaluated a wearable high-density (64-channel) dry-electrode EEG system and an open-source software framework for real-time neuroimaging and user state classification in the dynamic environment of the wearable setting. We first presented details on the wearable EEG form factor, compact electronics, and wireless triggering system. Dry-electrode signal quality was comparable to simultaneously recorded wet electrodes for average evoked responses (AEP, P300 corr. > 0.9) and single-trial data. We next described the software framework in detail, which included automated artifact rejection; neuronal system identification (cortical source localization and multivariate effective connectivity); prediction of behavior using spatiotemporal connectivity features; and interactive 2-D and 3-D data visualization. We presented mathematical details of several recent methods including the ASR

technique for online artifact removal, the use of ADMM for efficient small-sample sparse VAR model fitting and power and connectivity estimation, and the ProxConn constrained regression technique for connectivity-based classification.

We evaluated our framework on simulated high-density EEG data and on single-trial classification of Flanker-task response error commission from cortical multivariate effective connectivity (sdDTF) features using two source localization methods, cLORETA, and regularized LCMV Beamforming. Classification performance with cLORETA and LCMV was significantly above chance (mean AUC = 0.74 ± 0.09 and 0.72 ± 0.08, respectively). cLORETA performance did not differ when using a state-of-the-art ERP method (DSLR). However, application of DSLR to LCMV sources yielded significantly higher mean performance (AUC = 0.82 ± 0.12). To our knowledge, this is the first demonstration of a neuronal system identification and cognitive state classification using 64-channel dry EEG. We hope this will encourage new applications of wearable EEG to the study and monitoring of cognition and behavior in mobile real-world environments.

REFERENCES

[1] S. Ha et al., "Integrated circuits and electrode interfaces for noninvasive physiological monitoring," IEEE Trans. Biomed. Eng., vol. 61, no. 5, pp. 1522–1537, May 2014.
[2] L. D. Liao et al., "Biosensor technologies for augmented brain-computer interfaces in the next decades," Proc. IEEE, vol. 100, pp. 1553–1566, May 2012.
[3] S. H. Fairclough, "Fundamentals of physiological computing," Interacting Comput., vol. 21, nos. 1/2, pp. 133–145, Jan. 2009.
[4] B. Lance, "Brain-computer interaction technologies in the coming decades," Proc. IEEE, vol. 100, pp. 1585–1599, May 2012.
[5] T. Mullen, "The dynamic brain: Modeling neural dynamics and interactions from human electrophysiological recordings," Ph.D. dissertation, Dept. Cognitive Science, University of California San Diego, La Jolla, CA, USA, 2014.
[6] S. Debener et al., "How about taking a low-cost, small, and wireless EEG for a walk?" Psychophysiology, vol. 49, no. 11, pp. 1617–1621, Nov. 2012.
[7] A. Stopczynski et al., "The smartphone brain scanner: A portable real-time neuroimaging system," PloS One, vol. 9, no. 2, p. e86733, 2014.
[8] F. Varela et al., "The brainweb: Phase synchronization and large-scale integration," Nature Rev. Neurosci., vol. 2, no. 4, pp. 229–239, Apr. 2001.
[9] G. Tononi and G. M. Edelman, "Consciousness and complexity," Science, vol. 282, no. 5395, pp. 1846–1851, Dec. 4, 1998.
[10] K. Friston, "Beyond phrenology: What can neuroimaging tell us about distributed circuitry?" Annu. Rev. Neurosci., vol. 25, pp. 221–250, 2002.
[11] J. F. Kalaska and D. J. Crammond, "Cerebral cortical mechanisms of reaching movements," Science, vol. 255, no. 5051, pp. 1517–1523, Mar. 20, 1992.
[12] M. Hassan et al., "EEG source connectivity analysis: From dense array recordings to brain networks," PloS One, vol. 9, no. 8, p. e105041, Aug. 12, 2014.
[13] J. M. Schoffelen and J. Gross, "Source connectivity analysis with MEG and EEG," Hum. Brain Mapping, vol. 30, no. 6, pp. 1857–1865, Jun. 2009.
[14] S. Haufe et al., "A critical assessment of connectivity measures for EEG data: A simulation study," NeuroImage, vol. 64, pp. 120–133, Jan. 1, 2013.
[15] S. Makeig et al., "Evolving signal processing for brain-computer interfaces," Proc. IEEE, vol. 100, pp. 1567–1584, May 2012.
[16] B. He et al., "Electrophysiological imaging of brain activity and connectivity-challenges and opportunities," IEEE Trans. Biomed. Eng., vol. 58, no. 7, pp. 1918–1931, Jul. 2011.
[17] F. Babiloni et al., "Estimation of the cortical functional connectivity with the multimodal integration of high-resolution EEG and fMRI data by directed transfer function," NeuroImage, vol. 24, no. 1, pp. 118–131, Jan. 1, 2005.
[18] L. Astolfi et al., "Comparison of different cortical connectivity estimators for high-resolution EEG recordings," Hum. Brain Mapping, vol. 28, no. 2, pp. 143–157, Feb. 2007.

[19] M. Billinger *et al.*, "Single-trial connectivity estimation for classification of motor imagery data," *J. Neural Eng.*, vol. 10, no. 4, art. no. 046006, Aug. 2013.

[20] T. Mullen *et al.*, "Real-time modeling and 3D visualization of source dynamics and connectivity using wearable EEG," in *Proc. IEEE Eng. Med. Biol. Soc. Conf.*, 2013, pp. 2184–2187.

[21] C. A. Kothe and S. Makeig, "BCILAB: A platform for brain-computer interface development," *J. Neural Eng.*, vol. 10, no. 5, p. 056014, Oct. 2013.

[22] T. Mullen, "The source information flow toolbox (SIFT): An electrophysiological information flow toolbox for EEGLAB," Ph.D. dissertation, Dept. Cognitive Science, University of California San Diego, La Jolla, CA, USA, 2010.

[23] B. He *et al.*, "Econnectome: A MATLAB toolbox for mapping and imaging of brain functional connectivity," *J. Neurosci. Methods*, vol. 195, no. 2, pp. 261–269, Feb. 15, 2011.

[24] M. Billinger *et al.*, "SCoT: A python toolbox for EEG source connectivity," *Front. Neuroinformat.*, vol. 8, Mar. 11, 2014, doi: 10.3389/fninf.2014.00022.

[25] Q. V. Le *et al.*, "ICA with reconstruction cost for efficient overcomplete feature learning," in *Proc. Adv. Neural Inf. Process. Syst. Conf.*, 2011, pp. 1017–1025.

[26] M. T. Akhtar *et al.*, "Recursive independent component analysis for online blind source separation," *IEEE Int. Symp. Circuits Syst.*, 2012, vol. 6, pp. 2813–2816.

[27] S.-H. Hsu *et al.*, "Online recursive independent component analysis for real-time source separation of high-density EEG," in *Proc. IEEE Eng. Med. Biol. Soc. Conf.*, 2014, pp. 3845–3848.

[28] H. P. Lopuhaa and P. J. Rousseeuw, "Breakdown points of affine equivariant estimators of multivariate location and covariance matrices," *Ann. Statist.*, vol. 19, no. 1, pp. 229–248, Mar. 1991.

[29] N. Bigdely-Shamlo *et al.*, "The PREP pipeline: Standardized preprocessing for large-scale EEG analysis," *Front. Neuroinformat.*, vol. 9, p. 16, 2015.

[30] A. Gramfort *et al.*, "OpenMEEG: Opensource software for quasistatic bioelectromagnetics," *Biomed. Eng. Online*, vol. 9, p. 45, 2010.

[31] N. Tzourio-Mazoyer *et al.*, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *NeuroImage*, vol. 15, no. 1, pp. 273–289, Jan. 2002.

[32] A. Ojeda *et al.*, "MoBILAB: An open source toolbox for analysis and visualization of mobile brain/body imaging data," *Front. Hum. Neurosci.*, vol. 8, no. 121, Mar. 5, 2014, doi: 10.3389/fnhum.2014.00121.

[33] R. D. Pascual-Marqui, "Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: Exact, zero error localization," arXiv:0710.3341 [math-ph]. 2007a.

[34] G. H. Golub *et al.*, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.

[35] B. D. VanVeen *et al.*, "Localization of brain electrical activity via linearly constrained minimum variance spatial filtering," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 9, pp. 867–880, Sep. 1997.

[36] A. Delorme *et al.*, "EEGLAB, SIFT, NFT, BCILAB, and ERICA: New tools for advanced EEG processing," *Comput. Intel. Neurosci.*, vol. 2011, pp. 130–714, 2011.

[37] W. Hesse *et al.*, "The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies," *J. Neurosci. Methods*, vol. 124, pp. 27–44, 2003.

[38] M. Dhamala *et al.*, "Analyzing information flow in brain networks with nonparametric Granger causality," *NeuroImage*, vol. 41, pp. 354–362, 2008.

[39] C. M. Bishop, *Pattern Recognition Machine Learning*. New York, NY, USA: Springer, 2006.

[40] Z. L. Zhang and B. D. Rao, "Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2009–2015, Apr. 2013.

[41] P. A. Valdes-Sosa *et al.*, "Estimating brain functional connectivity with sparse multivariate autoregression," *Philos. Trans. Roy. Soc. Lond., B, Biol. Sci.*, vol. 360, no. 1457, pp. 969–981, May 29, 2005.

[42] S. Hauf K. R. Müller, G. Nolte, and N. Krämer, "Sparse Causal Discovery in Multivariate Time Series" in *Causality: Objectives and Assessment*, I. Guyon, D. Janzing, and B. Schölkopf, Eds., JMLR W&CP. vol. 6, pp. 97–106.

[43] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Stat. Soc. B. Stat. Method*, vol. 68, pp. 49–67, 2006.

[44] E. Bullmore and O. Sporns, "Complex brain networks: Graph theoretical analysis of structural and functional systems," *Nature Rev. Neurosci.*, vol. 10, no. 3, pp. 186–198, Mar. 2009.

[45] O. Sporns, "The non-random brain: Efficiency, economy, and complex dynamics," *Front. Comput. Neurosci.*, vol. 5, p. 5, 2011.

[46] O. Sporns and C. J. Honey, "Small worlds inside big brains," *Proc. Nat. Acad. Sci. USA*, vol. 103, no. 51, pp. 19219–19220, Dec. 19, 2006.

[47] R. Tomioka and M. Sugiyama, "Dual-augmented Lagrangian method for efficient sparse reconstruction," *IEEE Signal Process. Lett.*, vol. 16, no. 12, pp. 1067–1070, Dec. 2009.

[48] S. Boyd *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[49] A. K. Seth, "Measuring autonomy and emergence via granger causality," *Artif. Life*, vol. 16, no. 2, pp. 179–196, 2010.

[50] F. D. Fallani *et al.*, "Graph analysis of functional brain networks: Practical issues in translational neuroscience," *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, vol. 369, no. 1653, Sept. 1 2014, DOI: 10.1098/rstb.2013.0521.

[51] B. Schelter *et al.*, "Assessing the strength of directed influences among neural signals using renormalized partial directed coherence," *J. Neurosci. Methods*, vol. 179, pp. 121–130, 2009.

[52] L. A. Baccalá and K. Sameshima, "Partial directed coherence: A new concept in neural structure determination," *Biol. Cybern.*, vol. 84, pp. 463–474, 2001.

[53] P. W. Ferrez and R. M. J. Del, "Error-related EEG potentials generated during simulated brain-computer interaction," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 923–929, Mar. 2008.

[54] G. McLoughlin *et al.*, "Performance monitoring is altered in adult ADHD: A familial event-related potential investigation," *Neuropsychologia*, vol. 47, no. 14, pp. 3134–3142, Dec. 2009.

[55] P. Luu *et al.*, "Frontal midline theta and the error-related negativity: Neurophysiological mechanisms of action regulation," *Clin. Neurophys.*, vol. 115, pp. 1821–1835, 2004.

[56] C. S. Yu *et al.*, "Functional segregation of the human cingulate cortex is confirmed by functional connectivity based neuroanatomical parcellation," *NeuroImage*, vol. 54, no. 4, pp. 2571–2581, Feb. 14, 2011.

[57] T. Mullen *et al.*, "Analysis and visualization of theta-band information flow dynamics in an ERN-producing task," presented at the Human Brain Mapping Conf., Barcelona, Spain, 2010.

[58] A. Korzeniewska *et al.*, "Dynamics of event-related causality in brain electrical activity," *Hum. Brain Mapping*, vol. 29, pp. 1170–1192, 2008.

[59] L. T. Trujillo and J. J. B. Allen, "Theta EEG dynamics of the error-related negativity," *Clin. Neurophys.*, vol. 118, no. 3, pp. 645–668, Mar. 2007.

[60] J. Yordanova *et al.*, "Parallel systems of error processing in the brain," *NeuroImage*, vol. 22, no. 2, pp. 590–602, Jun. 2004.

[61] S. Lemm *et al.*, "Introduction to machine learning for brain imaging," *NeuroImage*, vol. 56, no. 2, pp. 387–399, May 15, 2011.

[62] R. Tomioka and K. R. Muller, "A regularized discriminative framework for EEG analysis with application to brain-computer interface," *NeuroImage*, vol. 49, no. 1, pp. 415–432, Jan. 1, 2010.

[63] T. Mullen *et al.*, "Real-time functional brain imaging: How GPU acceleration redefines each stage," in *Proc. GPU Technol. Conf. NVIDIA Corporation*, 2014.

[64] T. Mullen *et al.*, "Modeling cortical source dynamics and interactions during seizure," in *Proc. IEEE Eng. Med. Biol. Conf.*, Boston, MA, USA, 2011, pp. 1411–1414.

[65] F. D. Broccard *et al.*, "Closed-loop brain-machine-body interfaces for noninvasive rehabilitation of movement disorders," *Ann. Biomed. Eng.*, vol. 42, no. 8, pp. 1573–1593, Aug. 2014.

[66] C. Wilke *et al.*, "Estimation of time-varying connectivity patterns through the use of an adaptive directed transfer function," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 11, pp. 2557–2564, Nov. 2008.

[67] D. Schmorrow *et al.*, "Foundations of augmented cognition neuroergonomics and operational neuroscience," presented at 5th Int. Conference, FAC 2009, San Diego, CA, USA, Jul. 19–24, 2009,

[68] P. R. Davidson *et al.*, "EEG-based lapse detection with high temporal resolution," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 5, pp. 832–839, May 2007.

[69] T. M. Mitchell *et al.*, "Predicting human brain activity associated with the meanings of nouns," *Science*, vol. 320, no. 5880, pp. 1191–1195, May 30, 2008.

[70] S. Haufe *et al.*, "On the interpretation of weight vectors of linear models in multivariate neuroimaging," *NeuroImage*, vol. 87, pp. 96–110, Feb. 15, 2014.

[71] A. Mazaheri *et al.*, "Prestimulus alpha and mu activity predicts failure to inhibit motor responses," *Hum. Brain Mapping*, vol. 30, no. 6, pp. 1791–1800, Jun. 2009.