Computing Functional Brain Connectivity in Neurological Disorders: Efficient Processing and Retrieval of Electrophysiological Signal Data

Arthur Gershon, Ph.D.[1], Pramith Devulapalli[1], Bilal Zonjy, M.D.[2], Kaushik Ghosh, Ph.D.[3], Curtis Tatsuoka, Ph.D.[4], Satya S. Sahoo, Ph.D.[1,2]

[1]Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH
[2]Department of Neurology, School of Medicine, Case Western Reserve University, Cleveland, OH
[3]Department of Mathematical Sciences, College of Sciences, University of Nevada, Las Vegas, NV, U.S.A
[4]Neurological Institute, University Hospitals Cleveland Medical Center, Cleveland, OH

Abstract

Brain functional network connectivity is an important measure for characterizing changes in a variety of neurological disorders, for example Alzheimer's Disease, Parkinson Disease, and Epilepsy. Epilepsy is a serious neurological disorder affecting more than 50 million persons worldwide with severe impact on the quality of life of patients and their family members due to recurrent seizures. More than 30% of epilepsy patients are refractive to pharmacotherapy and are considered for resection to disrupt epilepsy seizure networks. However, 20-50% of these patients continue to have seizures after surgery. Therefore, there is a critical need to gain new insights into the characteristics of epilepsy seizure networks involving one of more brain regions and accurately delineate epileptogenic zone as a target for surgery. Although there is growing availability of large volume of high resolution stereotactic electroencephalogram (SEEG) data recorded from intracranial electrodes during presurgical evaluation of patients, there are significant informatics challenges associated with processing and analyzing this large signal dataset for characterizing epilepsy seizure networks. In this paper, we describe the development and application of a high-performance indexing structure for efficient retrieval of large-scale SEEG signal data to compute seizure network patterns corresponding to brain functional connectivity networks. This novel Neuro-Integrative Connectivity (NIC) search and retrieval method has been developed by extending the red-black tree index model together with an efficient lookup algorithm. We systematically perform a comparative evaluation of the proposed NIC index using de-identified SEEG data from a patient with temporal lobe epilepsy to retrieve segments of signal data corresponding to multiple seizure events and demonstrate the significant advantages of the NIC index as compared to existing methods. This new NIC Index enables faster computation of brain functional connectivity measures in epilepsy patients for large-scale network analysis and potentially provide new insights into the organization as well as evolution of seizure networks in epilepsy patients.

Introduction

The significant growth in the capabilities of neurotechnologies and computational methods, for example digital era 10KHz acquisition of high resolution Stereotactic Electroencephalogram (SEEG), has the potential to provide novel insights into brain activities that characterize human behavior and cognition (1). In particular, SEEG data together with diffusion-weighted Magnetic Resonance Imaging (DWI) data is being increasingly used to model both functional and structural brain networks in the context of a variety of neurological disorders (2). Structural brain network represents the brain white matter consisting of myelinated axons and functional network connectivity represents correlated brain activity (3, 4). Functional connectivity measures can be computed from multiple modalities of data, such as functional Magnetic Resonance Imaging (fMRI) and SEEG, which can be analyzed to understand the evolution of network connectivity patterns in neurological disorders, including Alzheimer's Disease, Parkinson's Disease, and epilepsy. Epilepsy is one the most common serious neurological disorder affecting more than 50 million persons worldwide and it is characterized by recurring seizures that severely impact the quality of life of patients and their family members. Approximately, 30% of epilepsy patients are refractory to pharmacotherapy and they are considered for surgery to resect brain structures called the seizure onset zone that are responsible for seizures. However, there are critical challenges in accurately delineating the epileptogenic zone, therefore surgery fails in 20-50% of the patients (5). Therefore, there is an urgent need to develop more efficient techniques that can process the growing volume of electrophysiological signal data for improving the characterization of the extent of seizure network and the associated seizure onset zone in patients (6).

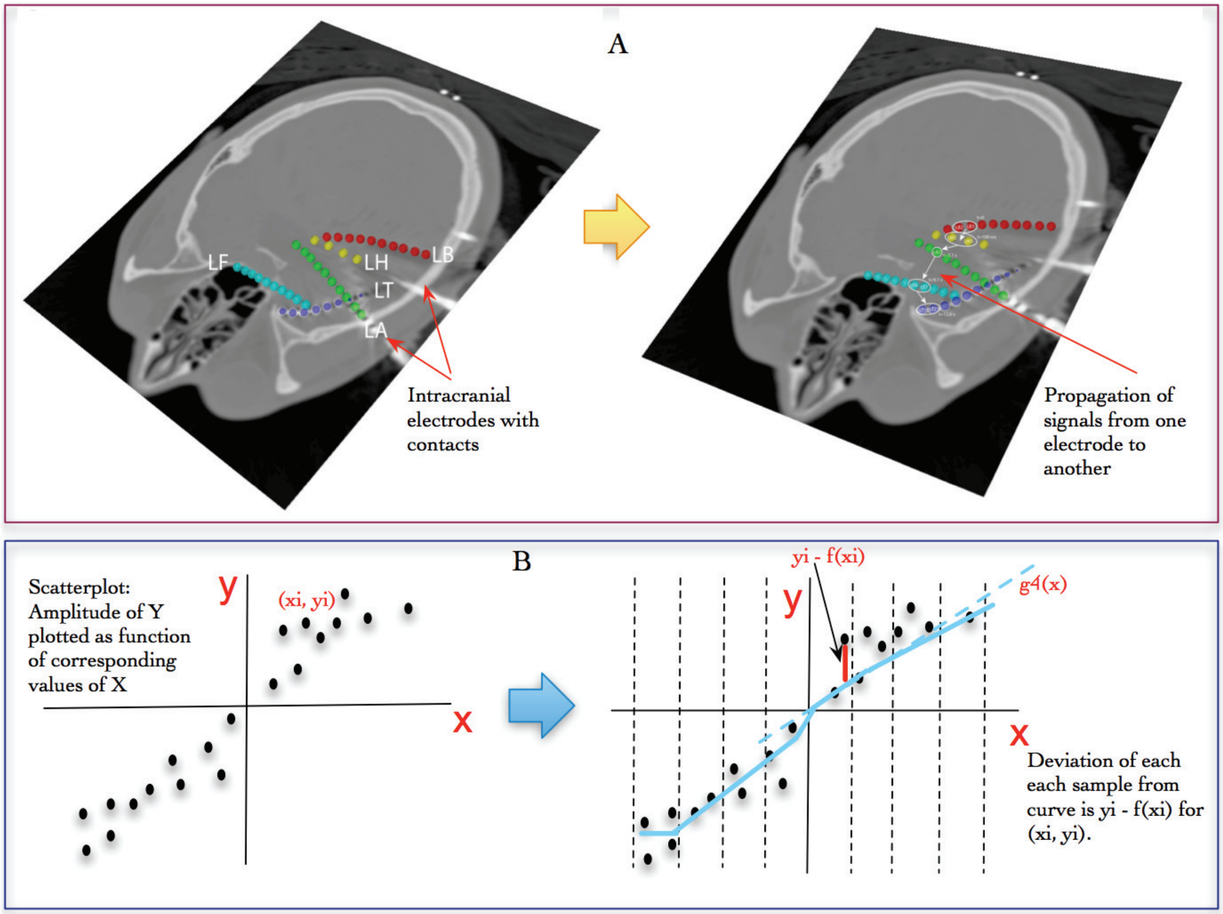

Figure 1: (A) Intracranial electrodes implanted in brain structures of an epilepsy patient and its electrode contacts are used to record high resolution EEG signal data. (B) The signal data are processed and analyzed to compute functional connectivity measures using methods such as nonlinear correlation coefficient measure, which provide correlation as well as directionality of seizure signals.

Deriving Brain Functional Connectivity Measures from Electrophysiological Signal Data. SEEG data recorded from intracranial electrodes implanted in brain structures provide a high-resolution "map" of seizure activities that can be used to model seizure networks and delineate the epileptogenic zone (7). The intracranial electrodes used to record SEEG signal data feature 10-15 contacts with length of 2mm, diameter of 0.8 mm with contacts that are 1.5 mm apart on the electrode, which allows them to record brain activity at a fine-level of granularity, therefore SEEG data is often used as "gold standard" in characterizing the extent of seizure networks (Figure 1(A) shows intracranial electrode contacts implanted in the insula of an epilepsy patient during pre-surgical evaluation). SEEG depth electrodes are used to identify the EZ and relevant components of the epileptogenic network (8). The potential of a brain region to generate high frequency oscillations (e.g. gamma range), also called rapid discharges, are used as an identifying feature for locating EZ in focal epilepsy (7, 9-12). The propagation characteristics of the seizure discharges, recorded by SEEG depth electrodes located at other brain regions, are used to identify the extent of the epileptogenic network (7). The epileptogenic zone in MTL epilepsy patients is multi-structural involving multiple temporal lobe structures in seizure onset (13), such as hippocampus, enterohinal cortex, and amygdala (14-16).

SEEG data recorded as multi-channel signals corresponding to electrical activity from different brain regions can be analyzed to identify whether the SEEG signals are correlated and also the specific type of correlation between signal channels. Seizure signals originate in one or more locations and involve additional brain regions, which together constitute a *seizure network* with spatial topology and temporal properties. This network is intuitively represented as a graph structure consisting of *nodes* representing single neuron or a brain region, and *edges* representing functional connectivity (17). In addition, the analysis of SEEG data can also provide a measure of directionality regarding the propagation of seizure signals from the region of onset to other brain regions (7).

The growing capabilities of SEEG recording systems has led to availability of large-scale, high-resolution data that can enable the neuroscience community to move away from small, underpowered studies to large-scale studies especially through multi-center research studies such as the National Institutes of Health (NIH)-funded Center Without Walls (CWW) for studying the mechanisms for Sudden Unexpected Death in Epilepsy (SUDEP) (18). The Center for SUDEP Research (CSR) involves 14 Epilepsy Monitoring Units (EMUs) across the U.S. and the United Kingdom. The CSR data repository is an excellent example of neuroscience signal big data, for example, the University Hospitals Cleveland Medical Center (UH-CMC) Epilepsy Center has collected an average of 321 megabytes (MB) of electrophysiological signal data per day per patient and about 30 terabytes (TB) of surface and intracranial signal data in the past 3 years. Therefore, it is important to leverage this high quality, signal big data to compute as well as analyze large-scale models of seizure networks and compute functional connectivity measures to characterize the seizure networks (19). There are multiple measures used to compute correlation between signal data recorded from different brain regions (20), for example non-linear correlation coefficient is used to derive functional connectivity measures with directionality from SEEG data (21). Nonlinear regression analysis of SEEG data quantify the degree of co-occurrence of signal values X(t) and Y(t) recorded from two brain locations G_x and G_y where t is the time of recording (22). Using the approach first described in (23), we define $h^2_{XY}(\tau)$ to be the non-linear measure of association between two signals at a time lag of τ in the direction of G_X to G_Y, where $h^2_{XY}(\tau) = 1 - (Var\ (Y(t+\tau) \mid X(t))) / (Var\ (Y(t+\tau))$.

Taking the maximum of this measure over all possible lag values τ, we get the overall measure of nonlinear association from node G_X to node G_Y as $h^2_{XY} = \max\ (\tau_{min} < \tau < \tau_{max})\ h^2_{XY}(\tau)$. The resulting quantity h^2_{XY} is called the nonlinear correlation coefficient in the direction G_X to G_Y with values between 0 (for no association) and 1 (for perfect association). In an analogous manner, interchanging the roles of X and Y can be used to derive h^2_{YX} representing nonlinear correlation coefficient from G_Y to G_X. We note that due to the nonlinear relationship between *X(t)* and *Y(t)*, h^2_{XY} and h^2_{YX} may not be equal. If there are k nodes in the seizure network graph, then the above computation will result in $k(k-1)$ pairwise directional nonlinear correlation coefficient values, which can be used to define the graph edges SP_F from the SEEG electrode contact locations (Figure 1(B) shows the method used to compute the nonlinear correlation measures from two SEEG channels). The functional connectivity measures are analyzed to address fundamental issues in characterizing the properties of seizure networks in epilepsy, for example:

1. *Can we predict the spatio-temporal properties of epilepsy seizure networks based on the previous seizure events in an individual patient or patients in a cohort?*
2. *Can we predict future seizure events based on network analysis of previous epilepsy seizure networks data of a patient?*

Therefore, there is a clear need to leverage the growing volume of high resolution SEEG data to derive functional connectivity measures during various seizure events and use comprehensive data mining techniques to gain new insights into the characteristics of epilepsy seizure networks in individual patients as well as in patient cohort studies. However, the processing and analysis of large-scale signal data stored using traditional data formats, such as the European Data Format (EDF) (24) (25, 26), is extremely challenging and requires development of new informatics methods that can efficiently derive functional connectivity measures. Although EDF is widely used as "de-facto standard" for signal data, it was not designed to support efficient signal data processing and analysis using new Big Data technologies, for example distributed computing infrastructure, parallelized algorithms, and indexing of partitioned data segments for faster data retrieval (27). To address the limitations of the EDF specifications, we developed a new flexible representation format for signal data called Cloudwave Signal Format (CSF), which extends the Javascript Object Notation (JSON) (28) with semantic annotation of signal data with domain ontology annotations (29). Using CSF as a common data model for SEEG data, we developed the Neuro-Integrative Connect (NIC) framework consisting of: (1) a new indexing structure called **NIC-Index**, which extends the red-black tree data structure to index signal data stored in CSF files for efficient processing and retrieval of specific segments of signal data corresponding to seizure events. In the following sections, we provide additional details regarding the NIC framework and in particular the NIC-Index as well as work related to the use of signal data retrieval techniques for computing brain functional connectivity in epilepsy.

Background

Brain Functional Connectivity using SEEG data. SEEG signal data is the predominant approach for recording high-resolution electrical activity for both neuroscience research and clinical purposes using precisely placed intracranial electrodes. EEG records post-synaptic activity in neuronal cells that are spatially oriented in the same direction (7). SEEG addresses the limitations of other EEG recording approaches, such as scalp EEG. SEEG records signals in three dimensional topography that allows accurate correlation between anatomical, electrical and clinical aspects of functional networks, for example occurrence of clinical signs with seizure discharges (7). SEEG is used as gold standard for evaluating the topology and extent of seizure network based on high frequency electrical activity (e.g., gamma frequency, 30-80Hz) and the locations participating in seizure propagation (9). Information about seizure network is essential for making decisions regarding surgical removal of brain tissue responsible for onset of seizure signals while protecting important brain functions, such as speech center (30). There are multiple approaches used to compute functional connectivity, such as phase synchronization (using Hilbert phase entropy, wavelet phase entropy etc.), generalized synchronization, and regression methods (we refer to (23) for detailed discussion). In particular, nonlinear regression analysis has been found to be more effective in analysis of EEG data in comparison with linear regression analysis and mutual information for analyzing interdependence (31).

Cloudwave Signal Format. The CSF format supports significant improvements over the EDF format in terms of semantic annotation of signal data, accessibility, and interoperability. In particular, the extraction of specific segments of signal data from EDF files requires several steps; each involving some computation of byte offset values in order to access the data. In contrast to EDF files, CSF files can be easily processed using dedicated "getter" functions supported by programming languages such as Java with the associated key string as the function's input. The CSF files also support the interoperability of signal data generated from different sources through the use of ontology terms for data annotation. In particular, the CSF files storing SEEG data are annotated with ontology terms described in the Epilepsy and Seizure Ontology (EpSO) (32). This semantic annotation of signal data allows easier reconciliation of seizure related event annotation in signal data from various sources, for example multi-center research studies like CSR (18). We demonstrate the effectiveness of CSF as a common data model for signal data access and processing by developing the NIC Index for use in computation of functional connectivity measures in epilepsy.

Indexing Structure for Efficient Data Access. Indexing techniques are widely used in data management and information retrieval applications for fast access to relevant data in a large data repository, for example database management system and documents. Binary search trees are specialized data structure consisting of a root node and two child nodes and each child node recursively has two nodes as children to form an overall tree structure. Binary search tree nodes consist of sorted "keys" for lookup and their corresponding values, which support fast access to specific data items and addition or deletion operations, are widely used in large scale data management systems (33). In particular, Binary search tree implementations, such as red-black tree and Adelson-Velsky and Landis (AVL) tree, support guaranteed search time. For example, red-black tree has a search time of O (log n), where n is the total number of elements in the tree (33). Therefore, red-black tree are suitable data structures for supporting efficient search and retrieval of large volume of complex biomedical data such as electrophysiological signal data. We demonstrate the development of the new NIC indexing structure based on red-black tree structure, which allows fast access to specific segments of signal data stored in CSF files corresponding to various seizure events, for example seizure onset, ictal events related to spread of seizures to other brain regions. This new NIC indexing structure for signal data allows faster computation of functional connectivity measures, which can be analyzed for characterizing epileptogenic seizure networks.

Graph theoretical approaches have been used to identify and describe network characteristics of brain regions participating in seizure events of epilepsy patients (34, 35). To the best of our knowledge, the NIC framework is the first project to define a flexible data format CSF for electrophysiological signal data and develop NIC Index as a new high-performance indexing structure to process and retrieve specific segments of signal data for computation of brain functional connectivity in epilepsy.

Method

The NIC framework has been developed to support comprehensive network analysis of epilepsy seizure networks using large-scale signal data to accurately characterize epileptogenic zone. To support large-scale signal data analysis, the NIC index has two primary objectives:

1. Development of a highly scalable index model for signal data stored in multiple CSF files corresponding to a single patient or multiple patients in a cohort study.
2. Development of efficient signal data retrieval algorithm that is implemented on the index model that can support faster processing of signal data to derive correlation values and subsequently generation of graph network models of epilepsy seizure network for network analysis.

These two objectives and the corresponding intermediate steps are shown in Figure 2, which the generation of directed graph network model representing the spatial and temporal properties of epilepsy seizure network.

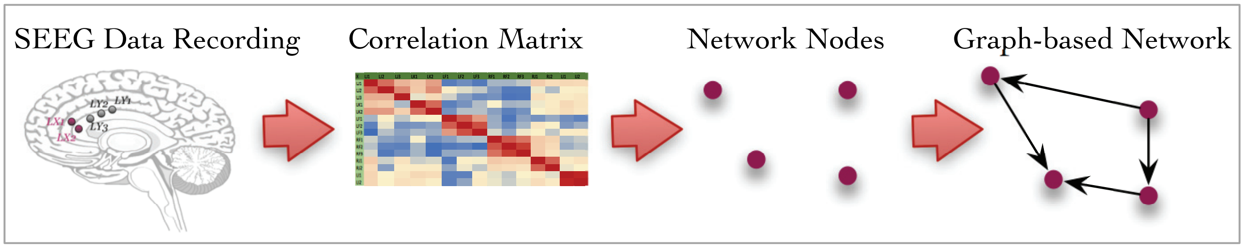

Figure 2: The core components of the NIC framework for computing graph-based model of seizure networks

Phase I: Search and Retrieval Algorithm for Specific Segments of Signal Data.

In this section, we describe the challenges associated with fast retrieval of specific segments of signal that correspond to the clinical events in epilepsy patients, such as seizure onset, spread of seizure to other brain structures (also called Ictal events), and end of seizure. To address the challenges associated with processing signal data stored in existing file formats such as EDF, we store signal data in CSF files that supports scalable storage of signal data as smaller segments with semantic annotations using standardized vocabulary modeled in an epilepsy ontology (32). Using the layout and storage structure of CSF file (please see detailed description of the CSF in our previous publication (29), we have developed an efficient algorithm for fast retrieval of specific segments of signal data from CSF files corresponding to clinical events. Figure 3 shows the input, output, and different steps of the algorithm to retrieve specific segments of signal data from CSF file. A naïve approach to retrieve specific segment of signal data from a CSF file is to take advantage of the underlying JSON "key-value" structure of the CSF file and use either the JSON Reader or JSON parser Application Programming Interface (API) for accessing file data.

```
Input: Start and End time of signal segment, List of Channels
Output: Segment of signal data as HashMap
For each CSF file in chronological order
    Create a blank HashMap<String, String> for segments
    Create a boolean foundData
    for each segment
        Get start time and end time for the segment
        if start time or end time is within the time interval
            add segment to HashMap
    if map is empty
        if foundData == false
            continue to the next CSF file
        else
            break from for loop
    else
        get sampling rates for each channel
        for each segment in the list
            for each channel
                get data as a CSV string
                convert CSV string to ArrayList<Double>
                use sampling rates to get start index and end index
                    ((# of seconds / record duration) * sampling rate)
                addAll of the sub-ArrayList from start index to end index to HashMap<String, ArrayList<Double>>
    return HashMap<String, ArrayList<Double>>
```

Figure 3: A new signal data search and retrieval algorithm developed as part of the NIC framework to retrieve specific segment of signal data from CSF files corresponding to specific seizure events.

Retrieval of Signal Segments using JSON API. The two standard JSON APIs provide different approaches for processing JSON files that are similar to the eXtensible Markup Language (XML) Document Object Model (DOM) for the *JsonReader* and Simple API for XML (SAX) for *JsonParser*. However, there are several limitations associated with the use of both these APIs to retrieve specific segments of signal data from CSF files, including significant overhead associated with the use of JSON APIs to open, read, and process the CSF files. In particular, the large size of CSF files (e.g., 120MB or more) makes it extremely difficult to use the DOM-based *JsonReader* API that loads the complete CSF file into the main memory. Therefore, to address these limitations of the JSON APIs, we developed a new tree-based approach to accessing and retrieving signal data segments from CSF files called **NIC Index**.

Phase II: A Tree-based Indexing Structure for Fast Retrieval of Large-Scale Signal Data.

As we discussed earlier in Background section, binary search trees are widely used to index large scale data for fast retrieval. Therefore, we selected a binary search tree-based indexing approach to address the limitation of JSON APIs to retrieve specific segments of signal data. However, the first version of the index for CSF was only able to handle range queries within the timestamps of single CSF file, which limited the usability of the index to the duration of signal data in a CSF file, which by default stores signal data recorded for 60 seconds. We note that seizure events often span minutes, therefore many of signal data search and retrieval queries require access to multiple CSF files, which requires the development of scalable indexing techniques for multiple CSF files. To support indexing of multiple CSF files, the initial version of the index was re-designed to support search and retrieval of signal data segments from multiple CSF files. In addition, the search and retrieval performance of the re-designed index (version 2) was optimized for multiple CSF files.

Re-Design of the NIC Index (version 2). We faced two primary challenges in the implementation of the revised NIC index, which allows search queries to span multiple CSF files with a comparable performance to version 1 of the NIC index:

1. Challenge 1*: Given multiple implementations of the binary search tree index structure, which data structure will meet the requirements of indexing multiple CSF files (potentially hundreds of files) and quickly identify the relevant CSF files for a given query?*
2. Challenge 2: *What data processing techniques can be used to allow fast access to signal data as numeric values (in contrast to string values) in a CSF file (please see our previous publication describing the details of signal data storage in a CSF* (29)*)?*

A naïve approach to address the first challenge is to search through every CSF file in a data repository until the search algorithm identifies and locates all relevant CSF files that match the start and end timestamps corresponding to the given seizure event. The time complexity of such an operation is characterized as O(N), where N is the number of CSF files in a data repository, which is clearly unacceptable for the large volume of signal data generated in epilepsy centers. Therefore, we explored all implementation of BST and identified the red-black tree, which a self-balancing binary search tree for extension and implementation as the NIC Index. We describe the implementation of this new index (version 2 of the NIC index in the following section).

Addressing Challenge 1: Implementation of the Red-Black Tree-based NIC Index. To address challenge 1 described above, we selected a red-black tree index based on its efficient look up process, which has a guaranteed time complexity of O (log n), where n is the total number of elements, and addition as well as deletion operations (33). In this section, we describe the details of the implementation of the NIC index. Using the property of the CSF file structure, which consists of two or more "fragments" of signal data with metadata information describing the epoch duration, start date, and start time, the version 2 of the NIC index stores the starting period and end timestamp of each individual fragment in the CSF file. Additionally, fragments in a CSF file can have different epoch durations, but no two fragments can have overlapping time periods as defined in the CSF specification (29). Given that all signal fragments of a CSF file have a unique timestamp, a binary search tree index structure can be implemented such that each node in the tree will be unique and there won't be two nodes with overlapping time segments. Now, it is possible for a binary search tree to be a simple linear chain of nodes in the worst-case scenario, which will not result in an improvement over the version 1 of the NIC index, therefore implementing a data structure that forces a balanced binary search tree is critical. As discussed earlier in the Background section, the two common approaches to implementing balanced binary search trees are AVL trees and red and black trees. Red and black trees were chosen because of the specific properties of CSF files. For example, CSF files continue to be added over a period of time as new patients are evaluated in the epilepsy center. Therefore, insertions or deletions operations (if a CSF file is updated) in a red-black tree are computationally less expensive procedures as compared to AVL trees (due to the inherent properties of each tree). While AVL trees can be more balanced than red-black trees, the time complexity of a query on both trees is still O (log n), where n is the number of elements in the tree. Therefore, we implemented a red-black tree to index all the different fragments in each CSF file based on the chronological order of the time segments.

Addressing Challenge 2: Pre-processing of CSF files using Object Serialization. The complexity of Challenge 2 is higher as compared to Challenge 1 primarily due to the time required to convert signal data stored as characters in a CSF file into numeric values, which can be analyzed to derive functional connectivity measures. We note that it required 18 seconds to load one CSF file in the version 1 of NIC index, which is clearly not a practical approach with hundreds of CSF files that need to processed for a single patient. Further, the contents of a CSF file can be split into five different categories: header, study metadata, channel data, clinical annotations, and data records. The search and retrieval queries for deriving functional connectivity measures access only the header and data records sections of a CSF file. We note that the maximum space in a CSF file is taken up by the data records as all the signal data is stored in the data records section. Further, a CSF file can have up to 200 channels, but most search and retrieval queries require access to signal data corresponding to ten channels.

Therefore, to address need for fast access to signal data stored in a CSF file in numeric form, we use "object serialization" technique available in the Java programming language. To implement object serialization, we preprocess every CSF file in the data repository, extract the header and individual channel data for every fragment, and store the signal data with the associated fragment identifier as serialized bin objects. The names of the individual bin objects correspond to the channel names. The processing of serialized objects (called "reading" in Java) is much faster than reading and preprocessing a CSF file in its original JSON format. Another reason for faster performance is the inherent size of the individual bin files as compared to the CSF file. A CSF file can be as large as 400 MB whereas the bin files for each channel varies around 500 KB. Reading in smaller files gives the NIC index significant advantage in terms of performance with respect to reading data from CSF files.

We now describe the step-wise procedure to perform the search and retrieval algorithm in the red-black tree-based NIC index:

1. Step 1: Read in the user defined start and end timestamps, the number of channels, and the specific channel names in a CSF file as part of the preprocessing phase.
2. Step 2: Access the red-black tree and perform an in-order traversal on the tree from the start timestamp to the end timestamp of a seizure event (e.g., seizure onset or Ictal event 1).
3. Step 3: Each node in the NIC index contains the canonical file path to a specific signal fragment. The in-order traversal reaches a node, get the nodes file path, extract the data from the specified channels by accessing only the necessary bin files, and proceed to the next node.
4. Step 4: Once the in-order traversal is complete, the range query is finished and the data is stored as a text file.

In the next section, we describe the results of our comparative evaluation of the two JSON APIs and the two versions of the NIC index (version 1 and version 2) using de-identified SEEG data from a patient with temporal lobe epilepsy.

Results and Evaluation

We used de-identified SEEG data from the Epilepsy Center at the University Hospital Cleveland Medical Center to evaluate the new signal indexing methods developed as part of the NIC framework.

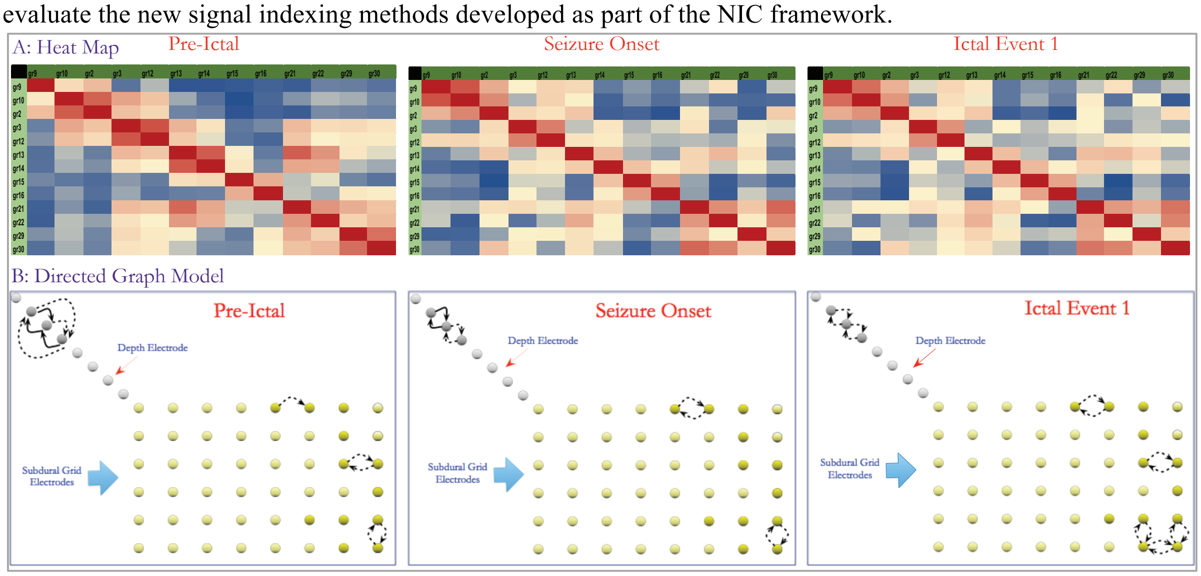

Figure 4: Derivation of graph network model of three clinical events using SEEG data by leveraging the NIC-Index

Temporal Lobe Epilepsy Patient. The SEEG data was recorded from a patient with epileptic paroxysmal episodes with epileptogenic zone in the left and right temporal lobes. The patient was refractory to anti-epileptic medications and was considered for surgery. The patient was implanted with three intracranial electrodes and a subdural grid covering the left fronto-temporal area. The SEEG data was de-identified before use in the project through removal of all protected health information (PHI) data. The raw stored in EDF files were processed and 596 CSF files were generated from specific EDF files with data corresponding to seizure events. In the following sections, we discuss the directed graph networks derived from the SEEG data and a comparative evaluation of the JSON APIs with NIC Index.

Functional Connectivity Measures for Seizure Events. Figure 4 shows the results of computation of nonlinear correlation coefficient measure to compute functional connectivity using SEEG data corresponding to three seizure events: (a) before onset of seizure, (b) seizure onset, and (c) first ictal event. The first part of the figure (A), shows the nonlinear correlation coefficient h^2_{XY} values represented as "heat map" matrices with red representing high correlation values and blue representing low correlation values. The second part of the figure (B) shows the directed graph model of the seizure activities between contacts on depth electrodes and subdural grid electrodes.

The graph network representation of the nonlinear correlation coefficient measure for functional connectivity in Figure 4 shows the formation of interesting network motifs over the course of time, especially during Ictal event 1 after onset of seizure. We use 8 clinical events associated with this data to perform a comparative evaluation of the NIC-Index with the three other approaches to demonstrate the significant improvement in the performance for retrieval of signal segments to compute functional connectivity measures from SEEG data stored in CSF files.

Table 1: A Comparative Evaluation of four approaches to retrieve segments of signal data from CSF file corresponding to seizure events

Seizure, Event	JSON Reader	JSON Parser	NIC Index (version 1)	NIC Index (version 2)
	Time in ms **(for single CSF file)**	*Time in ms* **(for single CSF file)**	*Time in ms* **(for single CSF file)**	*Time in ms* **(for multiple CSF files)**
Seizure 2, Seizure Onset	74210.98	389.44	0.74	13.29
Seizure 2, Ictal Event 1	71935.60	337.71	0.35	10.26
Seizure 3, Ictal Event 1	129377.02	407.40	2.54	15.70
Seizure 4, Seizure Onset	159917.88	356.73	0.43	13.48
Seizure 4, Ictal Event 1	159325.75	348.11	0.44	14.19
Seizure 4, Ictal Event 2	158283.59	425.16	2.13	15.20
Seizure 5, Seizure Onset	144972.16	341.56	0.77	15.7
Seizure 5, Ictal Event 2	144684.76	413.45	2.37	16.54

Comparative Evaluation of JSON APIs and NIC Index. Table 1 shows that there is a significant performance improvement for both the tree-based index and NIC-index as compared to the two JSON APIs, which validates our hypothesis for the development of dedicated index structures for CSF files. As expected, the DOM-based JsonReader API performs takes up a significant amount of time as compared to both the index structures. We note that time reported for version 1 of the NIC index is for single file, while version 2 of the NIC index is for multiple CSF files (596 files for this particular epilepsy patient). In particular, we note version 1 of the NIC index (for single CSF file) is more than 5 orders magnitude faster than the JSON reader API and 3 orders of magnitude faster than the JSON Parser API (for the seizure onset event). Similarly, we note that version 2 of the NIC index is 3 orders of magnitude faster than the JSON reader and at least 1 order of magnitude faster than JSON parser API (for the seizure onset event). We note that both the NIC Index versions are consistently faster as compared to the JSON APIs across all seizure events, which demonstrate the effectiveness of the new indexing structure for search and retrieval of specific signal data segments for computing functional connectivity measures. We now discuss the implication of the NIC index on the computation of functional connectivity measure for characterizing epilepsy seizure network.

Discussion and Conclusion

As part of our ongoing work in the NIC framework, we are integrating the NIC-Index in a new multi-step workflow system to compute multiple types of functional connectivity measures from large-scale SEEG data and generate corresponding adjacency matric representation of seizure networks. The NIC framework is expected to significantly improve the rate of processing and analyzing large volumes of SEEG data to facilitate leveraging high resolution signal data to study seizure networks. This effective use of neuroscience Big Data in clinical research has the potential to provide new insights into the underlying mechanisms of epilepsy seizures and advance treatment methods. In conclusion, the NIC framework described in this paper is one of the first comprehensive platforms to process and analyze SEEG data using a variety of functional connectivity measures. The NIC-Index is a novel approach to address a critical challenge in efficient processing and retrieval of signal data corresponding to a variety of clinical events, which can be used to generate graph-based seizure network models. We successfully demonstrated the performance and features of the NIC-index as compared to JSON APIs and a simple tree-based index using de-identified SEEG data corresponding to 8 clinical events.

Acknowledgement. This work is supported in part by the NIH Big Data to Knowledge (BD2K) grant (1U01EB020955) and NSF grant# 1636850.

References

1. Bargmann C, Newsome, W., Anderson, D., et al. BRAIN 2025: a scientific vision. NIH Report, 2014.
2. Swann NC, Cai, W., Conner, C.R., Pieters, T.A., Claffey, M.P., George, J.S., Aron, A.R., Tandon, N. Roles for the pre-supplementary motor area and the right inferior frontal gyrus in stopping action: electrophysiological responses and functional and structural connectivity. Neuroimage. 2012;59(3):2860-70.
3. Mukherjee P, Berman, J.I., Chung, S.W., Hess, C.P., Henry, R.G. Diffusion tensor MR imaging and fiber tractography: theoretic underpinnings. American Journal of Neuroradiology (AJNR). 2008;29(4):632-41.
4. Rogers BP, Morgan, V.L., Newton, A.T., Gore, J.C. Assessing functional connectivity in the human brain by fMRI. Magnetic Resonance Imaging. 2007;25(10):1347-57.
5. Englot DJ, Hinkley, L.B., Kort, N.S., Imber, B.S., Mizuiri, D., Honma, S.M., Findlay, A.M., Garrett, C., Cheung, P.L., Mantle, M. and Tarapore, P.E. Global and regional functional connectivity maps of neural oscillations in focal epilepsy. . Brain. 2015;138(8):2249-62.
6. Englot DJ, Lee, A.T., Tsai, C., et al., Seizure types and frequency in patients who "fail" temporal lobectomy for intractable epilepsy. Neurosurgery. 2013;73(5):834-44.
7. Bartolomei F, Chauvel, P., Wendling, F. Epileptogenicity of brain structures in human temporal lobe epilepsy: a quantified study from intracerebral EEG. Brain. 2008;131(Pt.7):1818-30.
8. Thornton R, Laufs, H., Rodionov, R., et al., EEG correlated functional MRI and postoperative outcome in focal epilepsy. Journal of Neurology, Neurosurgery, and Psychiatry. 2010;81(8):922-7.
9. Kahane P, Landre, E.,Minotti, L.,Francione, S.,Ryvlin, P. The Bancaud and Talairach view on the epileptogenic zone: A working hypothesis. Epileptic Disorders. 2006;8(Suppl2):S16-S26.
10. Wendling F, Bartolomei, F., Senhadji, L. Spatial analysis of intracerebral electroencephalographic signals in the time and frequency domain: identification of epileptogenic networks in partial epilepsy. Philosophical Tansactions Maths Physics Engineering Science. 2009;367(1887):297-316.
11. Wendling F, Chauvel, P., Biraben, A., Bartolomei, F. From intracerebral EEG signals to brain connectivity: identification of epileptogenic networks in partial epilepsy. Frontiers in systems neuroscience. 2010;4(154).
12. Jefferys JG, Menendez de la Prida, L., Wendling, F., Bragin, A., Avoli, M., Timofeev, I., Lopes da Silva, F.H. Mechanisms of physiological and epileptic HFO generation. Prog Neurobiol 2012;98(3):250-64.
13. Bragin A, Wislon, C.L., Engel, J. Jr. Chronic epileptogenesis requires development of a network of pathologically interconnected neuron clusters: a hypothesis. Epilepsia. 2000;1:144-52.
14. Wieser H. Electroclinical features of the psychomotor seizures. London: Butterworths; 1983.
15. Chabardes S, Kahane, P., Hoffman, D., Munari, C., Benabid, A.L. Role of the temporo-polar region in the genesis of temporal lobe seizures. Epilepsia. 1999;40(Sppl 7):78.
16. Wendling F, Hernandez, A., Bellanger, J.J., Chauvel, P., Bartolomei, F. . Interictal to ictal transition in human temporal lobe epilepsy: insights from a computational model of intracerebral EEG. J Clin Neurophysiol. 2005;22(5):343-56.

17. Kramer MA, Cash S.S. Epilepsy as a disorder of cortical network organization. Neuroscientist. 2012;18(4):360-72.
18. Lhatoo SD. The Center for SUDEP Research (CSR). NINDS; 2014.
19. Wendling F, Benquet, P., Bartolomei, F., Jirsa, V. Computational models of epileptiform activity. Journal of Neuroscience Methods. 2015;S0165-0270(15):122-3.
20. Bartolomei F, Lagarde, S., Wendling, F., McGonigal, A., Jirsa, V., Guye, M. and Bénar, C. Defining epileptogenic networks: contribution of SEEG and signal analysis. Epilepsia. 2017;58(7):1131-47.
21. Pijn JP, da Silva, F.L. Propagation of electrical activity: nonlinear associations and time delays between EEG signals. Basic Mechanisms of the EEG. Boston, MA: Birkhäuser; 1993. p. 41-61.
22. Sertkaya B. A Survey on how Description Logic Ontologies Benefit from Formal Concept Analysis. CoRR. 2011;abs/1107.2822.
23. Wendling F, Ansari-Asl, K., Bartolomei, F., Senhadji, L. From EEG signals to brain connectivity: a model-based evaluation of interdependence measures. Journal of Neuroscience Methods. 2009;183(1):9-18.
24. Kemp B, Värri, A., Rosa, A.C., Nielsen, K.D., Gade, J. A simple format for exchange of digitized polygraphic recordings. Electroencephalography and Clinical Neurophysiology. 1992;82(5):391-3.
25. Kemp B, Olivan, J. European data format 'plus' (EDF+), an EDF alike standard format for the exchange of physiological data. Clinical Neurophysiology. 2003;114(9):1755-61.
26. Kemp B, Roessen, M. European Data Format Now Supports Video. Sleep. 2013;36(7):1111.
27. Agrawal D, Bernstein, P., et al., Challenges and Opportunities with Big Data. Purdue University, 2011.
28. Crockford D. Introducing JSON 1999. Available from: http://www.json.org/.
29. Jayapandian C, Wei, A., Ramesh, P., et al., A Scalable Neuroinformatics Data Flow for Electrophysiological Signals using MapReduce. Frontiers in Neuroinformatics. 2015;9(4).
30. Lüders HO, Amina, S., Baumgartner, C., et al., Modern technology calls for a modern approach to classification of epileptic seizures and the epilepsies. Epilepsia. 2012;53(3):405-11.
31. Pijn JP. Quantitative evaluation of EEG signals in epilepsy, nonlinear associations, time delays and nonlinear dynamics. Amsterdam: University of Amsterdam; 1990.
32. Sahoo SS, Lhatoo, S.D., Gupta, D.K., Cui, L., Zhao, M., Jayapandian, C., Bozorgi, A., Zhang, GQ. Epilepsy and seizure ontology: towards an epilepsy informatics infrastructure for clinical research and patient care. Journal of American Medical Informatics Association. 2014;21(1):82-9.
33. Cormen TH, Leiserson, C.E., Rivest, R.L., Stein, C. Introduction to Algorithms. 2nd ed: MIT press; 2004.
34. Franaszczuk PJ, Bergey, G.K., Kaminski, M.J. Analysis of mesial temporal seizure onset and propagation using the directed transfer function method. Electroencephalogr Clin Neurophysiol. 1994;91(413-427).
35. Baccala LA, Alvarenga, M.Y., Sameshima, K., Jorge, C.L., Castro, L.H. Graph theoretical characterization and tracking of the effective neural connectivity during episodes of mesial temporal epileptic seizure. J Integr Neurosci. 2004;3:379-95.