An information theoretic technique for harnessing attenuation of high spatial frequencies to design ultra-high-density EEG

Pulkit Grover, Jeffrey A Weldon, Shawn K Kelly, Praveen Venkatesh, Haewon Jeong
Electrical & Computer Engineering, Carnegie Mellon University
{pgrover, jweldon, skkelly, praveen1, haewonj}@andrew.cmu.edu

***Abstract*—It is widely believed in the clinical and biosciences community that Electroencephalography (EEG) is fundamentally limited in the spatial resolution achieved using a few hundred electrodes. This belief rests on the well known decay of high-spatial frequencies as the signal passes from the brain surface to the scalp surface. These high spatial frequencies carry high spatial resolution information about the source. However, recent experimental work as well as our theoretical and numerical analyses strongly suggest that EEG's resolution could be improved significantly through increased electrode density despite this decay. Somewhat counterintuitively, instead of viewing this decay of spatial frequencies as a detriment to signal quality (which it is), in this work we propose an information-theoretic strategy to *harness* this decay to reduce circuit area and energy needed for high-resolution signal acquisition. This is made possible by the observation that this spatial-low-pass filtering of the signal as it passes from the brain to the scalp induces large spatial correlations that can be exploited information-theoretically. The proposed techniques are shown in idealized head models to reduce requirements on energy required for sensing by 3x. These results are being applied towards an ongoing project on developing the "Neural Web," a 10,000 electrode portable EEG system at CMU.**

I. Introduction and motivation

Modern Electroencephalography (EEG) systems used to measure electric potentials generated by the brain on the scalp surface use at most 512 electrodes, and often much fewer, while aiming to understand activity in the brain in a noninvasive fashion. A competing noninvasive modality, Magnetoencephalography (MEG), relies on measuring magnetic fields and is much more commonly used for spatial localization. While MEG uses only a few hundred sensors (comparable in number to modern EEG systems), MEG "SQUIDS" are much more expensive[1], and restrict the subjects' movements. Despite these shortcomings, MEG is more commonly used because it is thought to provide higher resolution for brain-activity localization and imaging[2], especially for shallow sources. It is well known that high spatial frequencies of EEG signals are severely attenuated (relative to low spatial frequencies), and because these frequencies carry high spatial resolution information, this attenuation reduces the resolution of EEG signals significantly. This spatial low-pass filtering effect of the human head has been used to estimate the "spatial Nyquist rates" of EEG using idealized head models by Srinivasan et al. [2], [5][3]. It appears to us that this Nyquist-rate analysis, which estimates a Nyquist rate of a few hundred sensors[4], along with practitioners' observations of low-spatial resolution of existing EEG systems (which also use a few hundred electrodes), has lead to a widespread belief that the resolution provided by EEG saturates at a few hundred electrodes.

Our research program challenges this belief. Our accompanying works make an informational case that ultra-high-density EEG systems, with thousands (if not tens of thousands) of electrodes can help in improving EEG's spatial resolution significantly[5] [6], [7]. This paper shows that information theoretic techniques can enable us to better engineer these systems. This research is a part of an ongoing project for building the "Neural Web": a 10,000 electrode portable EEG system, and this paper brings out how information theory is a crucial enabler in conceptualizing and building this system.

In [6], [7], we build and improve the analysis in [2], [5], concluding that these works underestimate the spatial Nyquist rate of EEG signals. We estimate that between 500-800 electrodes are needed to achieve the Nyquist rate. However, recent experimental evidence strongly indicates that there is value to sampling signals at 3mm inter-sensor distances [8]–[11] (and potentially smaller), which would require thousands of electrodes.

Motivated by these experimental results, in [6], [7] we question[6] the EEG spatial Nyquist-rate formulation as a framework for understanding fundamental limits on required spatial sampling. Commonly, the goal of spatial sampling is not to recover the EEG signals themselves, but to use them to re-

[1]For example, the entire Greater Pittsburgh area has only one MEG system at the University of Pittsburgh Medical Center, and is shared for research and clinical purposes.

[2]There are studies that observe that the localization accuracy (e.g. [1] and references therein) for the the two modalities are comparable, or complementary [2, Ch. 2], and sometimes, EEG is noted to be better [3]. However, in practice, EEG is not often used for imaging, but instead for characteristic signatures such as Event-Related Potentials (ERPs) or emissions in temporal frequency bands that are well established biomarkers [4].

[3]See [2] for a great introduction to the models and techniques.

[4]We note that a careful reading of Srinivasan et al. [5] makes it clear that their goal was to *increase* the number of electrodes used in EEG by providing a *lower*, not upper, bound.

[5]Or, at the very minimum, existing theoretical arguments and experimental results on EEG's limitations do not rule out improvements in resolution with increase in number of sensors to the large values proposed here.

[6]We note that similar questions are raised in the conclusions section in work of Srinivasan [5].

construct the signal source inside the brain. The Nyquist-rate formulation attempts to minimize the mean-squared error in reconstructing the EEG signal, not the error in reconstructing the signal source. Thus, while the Nyquist-rate formulation provides lower bounds on the required spatial sampling, these lower bounds are likely quite loose. We illustrate this possible looseness by using standard source-localization algorithms on increasing number of electrodes: from a few hundred to a few thousand. With each increase, we observe that the source-localization accuracy also increases substantially, *suggesting that even a few thousand electrodes may not be sufficient for accurate source localization*.

In this paper, we focus on the problem of building such high-density systems. The system design problem is nontrivial because state-of-the-art EEG recording systems require 5mW of power per electrode (e.g. [12]), which quickly scales to 50 W at 10,000 electrodes for just the sensing circuitry. Such power levels are (in best case) too large to allow the system to be portable, or even to keep the scalp perspiration-free, and in the worst case, harmful for moderate-term use. Towards lowering these energy costs, we observe that the decay of high spatial frequencies in EEG, while being a detriment to signal quality overall, also provides an opportunity for reducing circuit area and power in high-density sensing. This is because decay of these frequencies induces large spatial correlations in nearby electrodes (see Fig. 8) that can be harnessed through information-theoretic techniques. In particular, we provide a hierarchical referencing mechanism that reduces power consumption by about 3x or more in idealized head models with conservative assumptions on power savings in comparison with classical direct global referencing strategies, and provides significantly higher fidelity of measurements than a natural sequential referencing strategy.

It might appear that in information-theoretic terms, our problem is simply a distributed source coding problem of with correlated sources. However, the twist here is that *the "source" of information itself has an element of choice*: it is generated through taking difference of values of the underlying process at any two points[7]. Because the compression is performed through ADCs, it might be difficult to use sophisticated binning-type strategies such as Wyner-Ziv coding [13]. Therefore, it becomes important to "choose a source" (through choosing which values to take difference between) that is easily compressible with little coordination between sensors when compressing. The hierarchical-referencing approach provides one way to make this choice which outperforms conventional approaches. We conclude in Section V, where we also discuss the need for obtaining fundamental limits on referencing strategies, which we believe will require novel information-theoretic tools.

[7]This taking of differences is necessitated by measurement of the potential, which by definition is a relative quantity measured with respect to a reference.

II. Spherical harmonics, notation and preliminaries

We refer the reader to [14, Chapter 19] for an introduction to spherical harmonics. A suitably nice function $f(\theta,\phi)$ on (the surface of) a sphere, where θ is the polar angle, and ϕ is the azimuthal angle (in spherical coordinates), can be represented as a linear combination of *spherical harmonics* $Y_{l,m}$'s.

$$f(\theta,\phi)=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}c_{l,m}Y_{l,m}(\theta,\phi), \tag{1}$$

where

$$Y_{l,m}(\theta,\phi)=\sqrt{\frac{(2l+1)}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m(\cos\theta)e^{im\phi},$$

are orthogonal basis functions, and $P_l^m(\cdot)$ is the Legendre polynomial of the first kind [14].

Lemma 1 (Parseval's relation for spherical harmonics): For normalized spherical harmonics discussed here, (*i.e.*, $\int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi}|Y_{l,m}(\theta,\phi)|^2\sin\theta d\theta d\phi = 1$), Parseval's relation is given by:

$$\int_{\mathbf{x}\in\partial\mathcal{S}}|f(\mathbf{x})|^2d\mathbf{x}=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}|c_{l,m}|^2, \tag{2}$$

where $c_{l,m}$ are the spherical harmonics coefficients of $f(\cdot)$, and $\partial\mathcal{S}=\{\mathbf{x}:\|\mathbf{x}\|_2=r\}$ denotes the boundary of sphere $\mathcal{S}$ of radius r.

Because of the orthonormality of the spherical harmonics basis function, we can define the "energy" contained in harmonics $l=l_0,l_0+1,\ldots,\infty$ as

$$E_{l_0}^{\infty}=\sum_{l=l_0}^{\infty}\sum_{m=-l}^{l}|c_{l,m}|^2. \tag{3}$$

A. The "N-sphere model": computing potentials on the scalp surface for given potential on the cortical surface using an idealized model

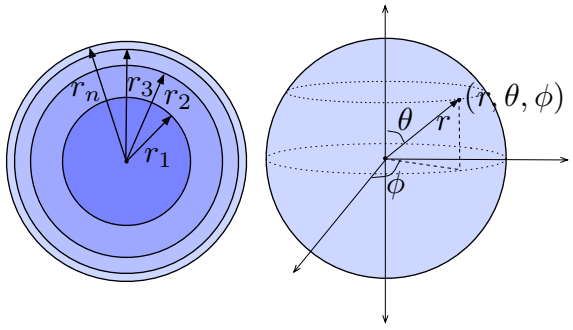

Fig. 1. The N-sphere model of the human head. Conductivities of different layers are different. Also shown are the polar angle θ and the azimuthal angle ϕ in the spherical coordinate system.

Fig. 1 shows the N-sphere model of the human head. The innermost sphere is the brain, and the outermost sphere is

the scalp. A common approximation is the 4-sphere model with brain, CSF, skull, and scalp layers. The layers have radii $r_1 = 7.8$ cm, $r_2 = 7.9$ cm, $r_3 = 8.6$ cm, and $r_4 = 9.2$ cm, and conductivities of CSF layer 5 times, the skull $1/15$ times, and the scalp equal to, the conductivity of the brain. The conductivity of the skull has had various estimates over the years. While Nunez and Srinivasan [2] use $1/80$ a the conductivity of the skull relative to the brain, we use $1/15$ based on more recent estimates that are now widely accepted [15].

Solving Laplace's equation in spherical coordinates for this model, the potential in the spherical annuli between $(s-1)$-th and s-th sphere (denoted generally by $\Phi^{(s)}$) is given by [2]:

$$\Phi^{(s)}(r,\theta,\phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\left(A_{lm}^{(s)}\left(\frac{r}{r_s}\right)^l + B_{lm}^{(s)}\left(\frac{r_s}{r}\right)^{l+1}\right)Y_{lm}(\theta,\phi), \quad r_{s-1}\le r\le r_s$$

so that at $r = r_s$, the field is:

$$\Phi^{(s)}(r_s,\theta,\phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\left(A_{lm}^{(s)} + B_{lm}^{(s)}\right)Y_{lm}(\theta,\phi). \quad (4)$$

As a reminder: here, $\Phi^{(s)}$ is the expression for the potential *in*side the s-th shell.

The following section builds on well-known solutions for Laplace's equation with spherical boundary conditions (see, e.g., [2], [14]).

III. The brain-to-scalp spatial transfer function

We refer the reader to [14] for a quick review of spherical harmonics, and [2] for propagation of potential in conducting media. The following theorem provides the spatial transfer function for a known potential at the surface of the innermost sphere (specified by coefficients $c_{l,m}^{(1)}$ of spatial harmonics $Y_{l,m}$'s) to a potential at the surface of the outermost sphere, given the conductivities and the radii of the intermediate shells. This can be used, for instance, to compute the potential at an idealized scalp given the potential at the idealized cortical surface.

Theorem 1 (Spatial transfer function): For the N-sphere model described in Section II-A, the following recursive equations describe the spatial transfer function connecting $c_{l,m}^{(N)}$ and $c_{l,m}^{(1)}$,

$$c^{(N)}(l,m) = A_{lm}^{(N)} + B_{lm}^{(N)}, \quad (5)$$

and

$$c^{(1)}(l,m) = A_{lm}^{(1)} + B_{lm}^{(1)}, \quad (6)$$

where,

$$A_{lm}^{(s)} = \frac{A_{lm}^{(s-1)} + B_{lm}^{(s-1)}}{\left(\left(\frac{r_{s-1}}{r_s}\right)^l + \gamma_l^{(s)}\left(\frac{r_s}{r_{s-1}}\right)^{l+1}\right)} \quad (7)$$

for $s = 2,3,\ldots,N$, and

$$B_{lm}^{(s)} = \gamma_l^{(s)} A_{lm}^{(s)}, \quad (8)$$

for $s = 1,2,\ldots,N$. Also, $\gamma_l^{(N)} = \frac{l}{l+1}$, and for $s = 2,3,\ldots,N-1$,

$$\gamma_l^{(s)} = \frac{\frac{l}{l+1}\sigma_{s,s+1} - \frac{\zeta_{l,s+1}}{l+1}}{\sigma_{s,s+1} + \frac{\zeta_{l,s+1}}{l+1}}, \quad (9)$$

where $\zeta_{l,s} = \frac{l\left(\frac{r_{s-1}}{r_s}\right)^l - (l+1)\gamma_l^{(s)}\left(\frac{r_s}{r_{s-1}}\right)^{l+1}}{\left(\frac{r_{s-1}}{r_s}\right)^l + \gamma_l^{(s)}\left(\frac{r_s}{r_{s-1}}\right)^{l+1}}$, and $\sigma_{s,s+1} = \frac{\sigma_s}{\sigma_{s+1}}$.

Proof: This result corrects and extends results obtained by Srinivasan et al. (see [2], [5])[8]. See [6] for a detailed proof. ∎

In [2], the authors provide numerical approximations to fields generated by radial dipoles that are convenient to use in practice [2, Pg. 565]. Here, we instead simplify the expressions in Theorem 1 to provide different approximations that hold in the limit of large l (as it turns out, l does not need to be very large for the approximations to be quite precise).

Corollary 1 (Exponential decay with spatial frequency): For large values of l, the transfer function can be approximated as:

$$H^{(N)}(l,m) = \frac{c^{(N)}(l,m)}{c^{(1)}(l,m)} \approx 2\left(\prod_{s=2}^{N-1}\frac{2\sigma_{s+1}}{\sigma_{s+1}-\sigma_s}\right)\frac{1}{\left(\frac{r_N}{r_1}\right)^l}, \quad (10)$$

where $\xi_l^{(s)} = \frac{\sigma_{s,s+1}+1}{\sigma_{s,s+1}-1}$, and $\sigma_{s,s+1} = \frac{\sigma_s}{\sigma_{s+1}}$.

Remark: Notice that higher frequencies decay primarily because of the term $\left(\frac{r_N}{r_1}\right)^l$ in the denominator, which is the reason why the decay is exponential in l. This means that the low-pass filtering is not as much influenced by low-conductivity of the skull as simply by distance from the brain surface. Further notice that the transfer function does not depend on σ_1. This might seem surprising. The reason σ_1 does not appear in the expression is because by specifying $c_{l,m}^{(1)}$ for all l,m, we specify the potential ϕ on the sphere of radius r_1. With this boundary condition, the conductivity σ_1 does not matter. However, to compute potential due to a spread of dipoles inside the brain will require computation of $c_{l,m}^{(1)}$ themselves, which will require use of σ_1.

A. Nyquist rates for a 4-sphere model: Analysis and numerical results

While the fundamental goal is source reconstruction, what if we simply want to reconstruct the EEG signal on the scalp? Because the transfer function never becomes zero, one can think of measuring the spatial bandwidth at, say, the threshold that contains 99% of the signal's energy. If so, the aliased signal will contribute no more than 2% errors. This philosophy (though not quite the techniques) seems to have been implicitly adopted in the work of Srinivasan et al. [5]. Observe that the number of sensors required to reconstruct scalp EEG faithfully should be a *lower bound* on the required

[8] We observed that some of the mathematical expressions provided on pg. 563 in [2], [5] are erroneous. This theorem corrects and generalizes these expressions. The code for computing the transfer function derived in our work is available at [16].

number of sensors to do any reasonable whole-brain imaging: if one can faithfully recover the sources, one can faithfully recover the EEG as well. Using numerical results, we will argue in Section III-B that the converse of this statement is not true, which implies that Nyquist rate for EEG signals is only a loose lower bound on the required number of sensors. Nevertheless, it is useful to have an estimate of number of sensors needed for recovering EEG for practical systems. In our numerical results (see [6], [7]) we observe that this lower bound is 578 sensors, which is more than any system currently in use.

B. Is Nyquist rate the fundamental limit on number of sensors? A source-localization perspective

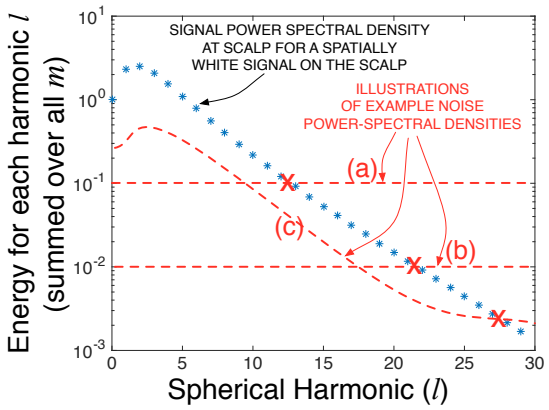

Fig. 2. A plot of energy in a spherical harmonic on log-scale, illustrating why the Nyquist-rate analysis might be conservative in estimating the required number of sensors for achieving close-to-maximum number of sensors. Without an understanding of the noise spatial spectrum, it is unclear where signal power is superseded by noise power. Intuitively, that is the point where gains of increasing number of sensors start diminishing.

The Nyquist-rate analysis effectively provides a limit on the number of sensors needed to reconstruct the EEG signal with small mean-square error: by ignoring 1% of energy in spherical-harmonics domain above, we obtain an estimate of signal in space such that the error has at most 2% of the energy of the original signal (the error has twice the energy because of aliasing it causes). It is worth noting that both Nyquist-rate analysis and our simulations below are noiseless. The number of required sensors are only expected to increase when realistic noise sources are accounted for in order to overcome loss of accuracy due to noise. Nevertheless, for quick estimates, intuitively, the number of sensors should be increased until the point when the signal power is superseded by noise power.

However, mean-squared error in reconstructing the *EEG* signal is not a good metric to evaluate the goodness of EEG signal: it is important to estimate the variable that EEG signal is being used for. For instance, for source-localization, where the goal is to image the brain activity, accuracy of source localization is of importance, not the accuracy of EEG signal reconstruction. In Fig. 3, using simulations and standard algorithms, we estimate improvements in accuracy as number

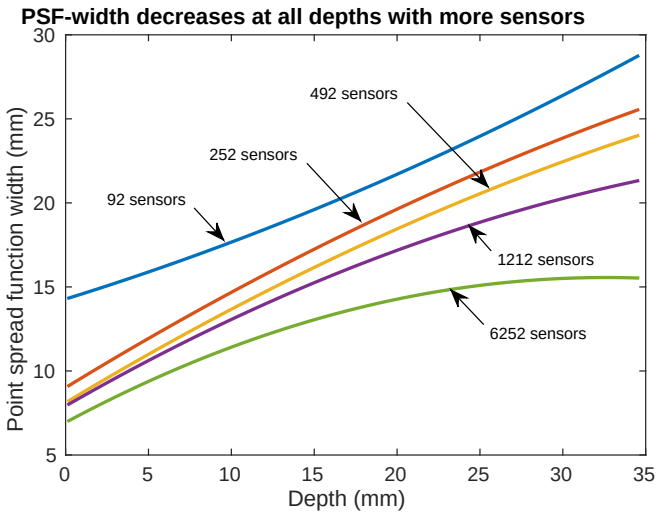

Fig. 3. The width of the "point-spread-function" (PSF) is used as a measure of accuracy of reconstruction. The PSF is the spread of dipoles in the reconstruction of the source, a dipole. All points at distances greater than the PSF width (from the true source) are guaranteed to have PSF amplitude less than 1/e times the max PSF amplitude. It is plotted here vs depth from cortical surface for different number of sensors for noiseless measurements. The algorithm assumes that the sources can be recovered using elastic net, which minimizes a linear combination of an L_2 and an L_1 cost on the source, and the error in signals estimated (vs measured signal) at the EEG sensor using the predicted sources. There is significant improvement in resolution as the number of electrodes increases.

of sensors increase from currently used few hundred sensors to a few thousand sensors. Our main conclusion is that there is significant improvement in resolution at all depths as the number of electrodes increases. Thus, the Nyquist rate analysis likely provides only a very loose lower bound. However, in absence of fundamental limits on the required number of sensors for source localization (a topic of our current study, which appears nontrivial because these limits require knowing noise power-spectral-density), it is hard to be certain that these high-density systems are fundamentally needed.

Why should we focus on source localization? In effect, source localization for EEG captures the spirit of the more classical Nyquist-sampling of 1-D signals: if the source can be reconstructed, the EEG signal can be reconstructed at any point simply by using standard forward models [17].

IV. Referencing mechanisms for lossy compression

The spatial power-spectral-density of EEG as it passes through the CSF, the skull, and the scalp (in a 4-sphere model) is shown[9] in Fig. 2. Because most energy is concentrated at low frequencies, the induced correlation is quite high (see Fig. 8). Further, noise due to EMG and EOG (muscle and eye artifacts) is also spatially highly correlated [8]. Thus, the sensors are sensing highly correlated signals.

One might be tempted to ask: since the sensors are sensing highly correlated data, do we need to have high-density

[9]The induced correlations will be computed and used in Fig. 8 in Section IV-B to estimate energy savings using our proposed strategy.

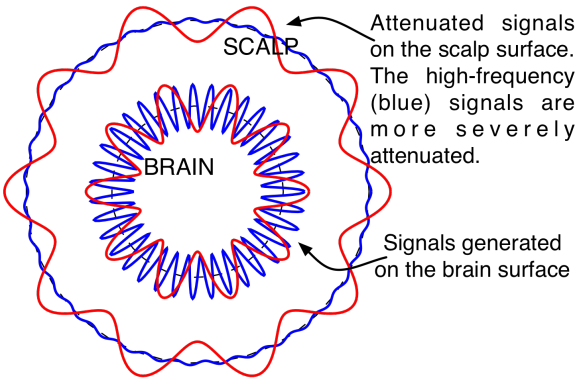

Fig. 4. An illustration of how the signal attenuates as it travels from the brain surface to the scalp surface. The inner circle represents the brain surface, and the outer the scalp surface. The low-frequency signal (denoted in red) has a much smaller decay than the high frequency signal. Electrodes sense the sum of the two signals, and thus relatively small difference of the signals at recorded at nearby locations on the scalp captures the high frequency signal.

sensing in the first place? For EEG, the answer is illustrated in Fig. 4: there is important information buried even in the "less significant" bits of each observation. Because of decay of high spatial frequencies (see Fig. 2), these lower order bits are the ones carrying high-resolution information. Perhaps as an acknowledgement of this fact, many state-of-the-art systems record each electrode at a very high precision (e.g. 24 bit ADCs [12]) with 256 or more sensors. Small differences between recordings of nearby sensors matter because these differences capture events that generate small variations in potential either due to their shallow depth or small spatial extent. It is unclear to us, however, if this is the actual motivation behind the use of 24-bit ADCs.

A. *Referencing mechanisms as trees*

We first present a simple observation: every referencing mechanism can be represented as an associated tree. Each electrode corresponds to one node in this tree. The global reference electrode is the root node. All nodes referenced directly against the global reference are the nodes at level 1. All nodes referenced directly against level 1 nodes are level 2 nodes, and so on. In the tree, a node's parent is the node the corresponding electrode is referenced against. Because each electrode is referenced against exactly one electrode, each node has exactly one parent, and thus the corresponding graph for any referencing strategy is a tree.

Let us consider two simple strategies to illustrate this association. In the first case, shown in Fig. 5, is a commonly used system where all electrodes are referenced directly against a global reference (sometimes called "unipolar" or "direct global referencing"). The corresponding tree, also shown in Fig. 5, has one root node with all other nodes as its children. In the second case, that we will call "sequential bipolar referencing," each electrode is referenced to an electrode prior to it in a chain of electrodes (see Fig. 6) leading to a tree where each node has 1 child. We now discuss another possible mechanism: hierarchical referencing.

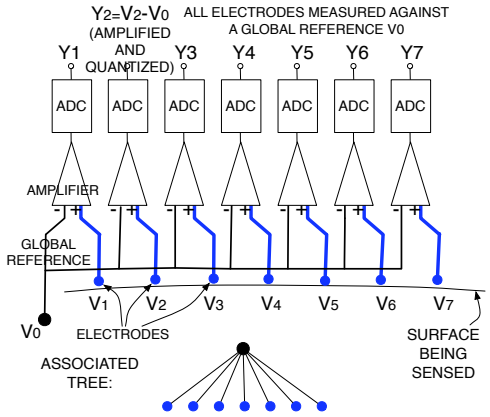

Fig. 5. Direct global referencing, and the associated tree, where every electrode is referenced directly against a global electrode.

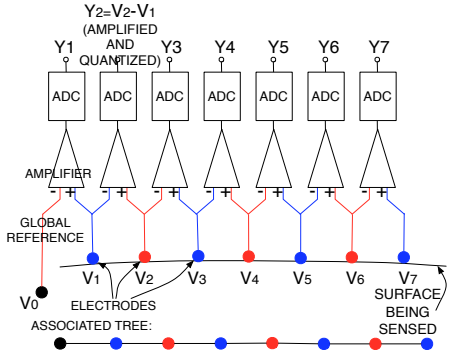

Fig. 6. Sequential bipolar biopotential referencing and the associated tree.

1) The hierarchical referencing mechanism, and comparison with sequential bipolar and direct global referencing: The mechanism proposed in this work is illustrated in Fig. 7. Instead of using sequential bipolar referencing, we arrange the electrodes in a tree pattern with number of levels that can be optimized based on spatial correlations. The mechanism is easy to design. For instance, for placement of electrodes along a grid, one can simply think of the grid as a hierarchical subdivision of squares. In case of placement of electrodes on a spherical surface, commonly used techniques are inherently recursive (e.g. icosahedron bisections used in EEG grids) and lend themselves naturally to hierarchical constructions of unequal but roughly constant degree per level. We now discuss why this strategy improves on the sequential bipolar strategy and the direct global referencing (unipolar) strategy.

Comparison with sequential bipolar strategy: Let us first assume that the input-referred noise variance for all electrodes is the same. To recover potential $V_i - V_0$ from measurements $Y_1 = V_1 - V_0 + Z_1, Y_2 = V_2 - V_1 + Z_2, \ldots, Y_i = V_i - V_{i-1} + Z_i$, one can simply add these potentials:

$$\sum_{j=1}^{i} Y_j = \sum_{j=1}^{i} (V_j - V_{j-1} + Z_j) = V_i - V_0 + \sum_{j=1}^{i} Z_j,$$

and thus noise variance increases linearly with the number of electrodes.

On the other hand, the hierarchical referencing strategy has a

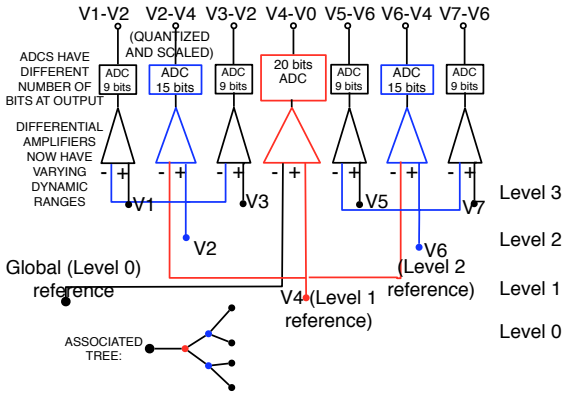

Fig. 7. The hierarchical referencing strategy. For clarity, we indicate on top the differences of voltages being measured, and ignore quantization and thermal noise added by the circuitry. If the circuits were noiseless, and actual differences of voltages were being measured, all voltages can be recovered with respect to the global reference in just three steps. For instance, $V_3 - V_0 = (V_3 - V_2) + (V_2 - V_4) + (V_4 - V_0)$. Of course, what are actually measured are $Y_i = V_i - V_{\mathrm{ref}(i)} + Z_i$, where Z_i is the noise term. Thus what is recovered is the sum of these voltages added to a noise whose variance is the sum of the variance of noise at each electrode.

slower increase in noise. For instance, consider a tree where every node has D children. Then, to reach every node, one only requires to sum of $O(\log(n))$ terms, and the noise also grows only logarithmically.

Alternatively, in sequential bipolar referencing, one could allocate different levels of resolution to different electrodes, for e.g., by allocating increasing number of bits of ADCs as the electrode gets farther from the reference electrode. To compare with hierarchical referencing, let us assume that the noise in the worst case is $\log(n)$ as well. In that case, one could use $\sigma_{zi}^2 = \sigma_{z0}^2/i$, because $\sum_{i=1}^{n} \frac{1}{i} \sim \log(n)$. However, this requires at least half the electrodes to have noise variance $\sigma_{zi}^2 \leq 2\sigma_{z0}^2/n$, which reduces to zero as n increases (while this noise variance can be kept constant in the hierarchical strategy for $\log(n)$ increase in overall noise).

The hierarchical strategy also makes it easier to ensure good contact: as long as (the fewer) electrodes in lower layers have good contact, the larger number of electrodes in the higher layers will be sensed accurately. Having poor contact in the highest layer hurts only the particular electrode with poor contact (because there are no further electrodes that reference against the highest layer).

Comparison with direct global referencing strategy: The savings are in energy and area because of savings in ADC bits. How many bits are saved? For a sensor that is sensing a signal of variance σ^2, the required ADC resolution for q_e^2 quantization noise scales as $\frac{1}{2}\log\frac{\sigma^2}{q_e^2}$. What happens if a signal X_2 is sensed with X_1 as reference? Assuming that the individual variance of the two signals is the same, σ^2, the variance of $X_2 - X_1$ is $2\sigma^2(1-\rho)$, where ρ is the correlation between the two signals. The required ADC resolution is now smaller, $\frac{1}{2}\log\frac{2\sigma^2(1-\rho)}{q_e^2}$, as long as $\rho > 0.5$. How large is ρ? This depends on sensor spacing and location. For idealized (4-sphere) head models of EEG, we obtain an estimate in Section IV-B, where we also obtain estimates of energy savings (using simplistic ADC power-consumption models) using this strategy for the same level of overall quantization noise.

B. A case study: ultra high density EEG for idealized spherical head models

In this section, we obtain estimates of power savings using hierarchical referencing for ultra high density EEG sensing (with up to 9331 electrodes; conventional systems use only up to 512 electrodes, typically much smaller). The code for this case study is available online at [16].

Algorithm for optimal tree search: To state the problem of optimal tree-search formally, there are n sensors sensing voltages $V_i, i = 1, \ldots, n$. A goal can be to find the referencing tree $\mathcal{T}$ that solves the following problem:

$$\min_{\mathcal{T}\ s.t. E_{\mathcal{T}} \leq E_{\max}} \max_{i=1,\ldots,n} \mathbb{E}[(\hat{V}_i - V_i)^2], \tag{11}$$

where V_i is the potential at electrode i with respect to the global reference, $\hat{V}_i$ is its estimate, $E_{\mathcal{T}}$ is the energy required by measurements corresponding to the measurement tree, and $E_{\max}$ is the total available energy. By taking maximum over $i = 1, \ldots, n$, we ensure that the maximum distortion over all electrodes is minimized. The expectation in (11) is taken over the (random) values of input V_i.

The search for the optimal referencing tree can be hard: the number of possible trees for n nodes is given by Cayley's formula, n^{n-2}, which is super-exponential in n. Therefore, we limit ourselves to obtaining the optimal hierarchical strategy, and comparing it to the global referencing strategy. Assuming that the spatial correlation between sensors as a function of distance between them is known in advance, the following brute-force algorithm provides a solution that is still polynomial time in number of sensors (simply because of the symmetry of the hierarchical tree: each node at a level has the same degree) when the signal is spatially stationary (as is the case in our idealized model for EEG signals). We believe that the algorithm can be improved using dynamic-programming, and are currently investigating this approach.

In Algorithm 1, QNVar is the variable estimating the aggregate quantization noise, and Energy counts the total energy of the ADCs in the hierarchy. An ADC at level i has Bi bits, and there are α^i such ADCs (because at level i, there are α^i nodes). E_ADC(Bi) is the energy required by an ADC that quantizes to Bi bits. rho, the correlation between i-th electrode and its reference ref(i), is computed as discussed below.

Computing correlations: To perform this case study, one final piece of the puzzle is computing correlations. Because of lack of such high-density systems, empirical measurements of correlations at small distances are lacking. Therefore, we use idealized 4-sphere models to obtain theoretical estimates (see Fig. 8). We observe that we underestimate the correlations somewhat (and therefore, underestimate the energy savings using our strategy) by assuming that the signal at the surface of the brain is white. One expects the signal to have more

Algorithm 1 Computing energy and noise with hierarchical referencing

FOR B1 = [1:BMAX]
...
FOR BL = [1:BMAX]
QN(B1,B2,...,BL) = 0; % Initializing noise variance
Energy(B1,B2,...,BL) = 0; % Initializing energy
FOR i=1:L

rho = calc_rho(i,ref(i));
QNVar(B1,...,BL) = QNVar(B1,...,BL) + $2\sigma^2$(1-rho) 2^{-2Bi};
Energy(B1,...,BL) = Energy(B1,...,BL) + α^i E_ADC(Bi);

END
...
END

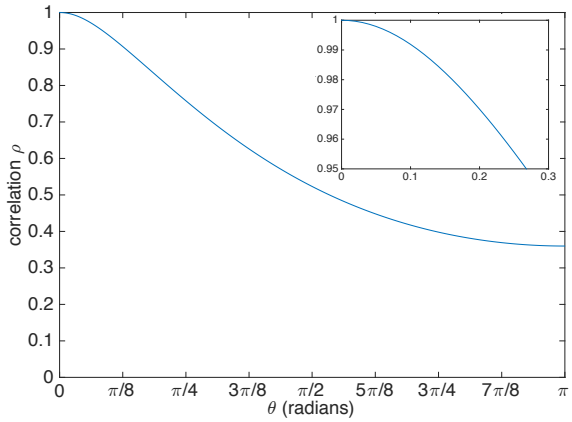

Fig. 8. Computed EEG spatial correlations between two points separated by θ angular separation on the scalp. The correlations are computed using the 4-sphere model with parameters $r_1 = 7.8$ cm, $r_2 = 7.9$ cm, $r_3 = 8.6$ cm, and $r_4 = 9.5$ cm, and conductivities of CSF layer 5 times, the skull $1/80$ times, and the scalp equal to, the conductivity of the brain, *assuming a white signal on the surface of the brain*. Correlations estimated using more realistic models such as fibre-tracks will likely only be larger because the signals on the surface of the brain are likely to be more slowly spatially changing than white signals. Note that parameters are slightly different from [2] to align with more recent studies in variability of these parameters.

content in lower frequencies because of spatial correlation in neuronal activity, especially in the cortex.

ADC energy model: How can we model the energy consumption of an ADC as a function of the number of bits? High precision ADCs (16 bits or larger) are often implemented using the sigma-delta architecture (see, e.g., [18]), frequently in biopotential measurements as well [12]. For these ADCs, the power consumption can be modeled as long as we know what order of filter is used to implement them. For instance, for second order ADCs, the oversampling needs to increase by a factor of 2 for every 15 dB increase in SNR [19]. Because one bit is approximately 6 dB of SNR, one obtains that r bits of saving leads to $2^{\frac{6r}{15}}$ savings in oversampling factor. Assuming a simple (if somewhat conservative) linear model of energy as a function of oversampling ratio yields the model we are using (see [16] for details).

Using correlations in Fig. 8 and ADC energy model above, the energy savings for a 9331-electrode EEG (obtained by allowing a maximum of 5 levels of hierarchy, with each reference electrode being used as a reference for $\alpha = 6$ electrodes in the lower layer) are:

TABLE I
ENERGY SAVINGS AND OPTIMAL BIT ALLOCATION FOR HIERARCHICAL REFERENCING (HR) IN COMPARISON WITH GLOBAL REFERENCING (GR)

GR bits	Optimal HR bits	energy savings factor
24	29, 25, 23, 22, 21	2.788
22	24, 23, 21, 19, 18	2.790
20	22, 21, 19, 17, 16	2.790

As shown in Table I, the savings are consistently close to a factor of 3 (and remarkably close to each other as well). We note that the bit-allocation of hierarchical strategy (HR) is chosen so that the overall standard deviation of noise for HR is smaller than the nonhierarchical (global) counterpart. The search for the optimal hierarchical strategy was performed assuming ADCs of 16, 17, ..., 30 bits are available. Savings will only be larger if a larger set of ADCs is allowed. Note that this architecture can lead to area-savings as well because lower-order/lower-area ADCs could be used at the levels farther away from the tree-root.

We believe that this analysis underestimates the benefits of our strategy because reducing the number of bits in ADCs can simplify the design itself: a lower order sigma-delta ADC could be used, or alternatively, one might be able to use a simpler SAR ADC. The precise amount of power saved will depend on how much resolution is needed, and most of all, on how much correlation is there between the signals in the first place. As density of electrodes increases, this correlation, and hence the power savings, are expected to increase as well.

V. Discussions and Conclusions

It is also of interest to obtain fundamental limits on source-localization accuracy (see, e.g., [20]). The difficulty in using these bounds is that the noise-spectral density is extremely ill understood. An information-science guided structured approach that mixes theory and experiments is needed to better estimate noise spectrum. Current techniques rely on "empty-room measurements" (see, e.g., [21]) to estimate noise covariance, ignoring the fact that much of the noise is generated by the subject themselves. A good estimate of noise spectrum can yield upper and lower bounds on source-localization accuracy — in spirit of traditional information theory — helping experimentalists provide confidence intervals on their results. Beyond source-localization, an exciting problem of current interest is estimating functional connectivity of brain networks, and obtaining directions of information flows [22]–[24], and thus it is also of interest to understand fundamental limits on required number of sensors for such connectivity-estimation problems.

On the implementation side, to make an ultra-high-density EEG system portable will likely require reducing measurement energy below that achieved by hierarchical referencing strategy proposed here. We believe that techniques that involve turning on only relevant sensors (e.g. [25] and references therein) will prove extremely relevant in this context. These techniques enable data compression at sensors by simply turning off less relevant sensors, or quantizing to just the required amount of precision, saving substantial circuit energy, offering another avenue for information-theoretic analyses.

It is also important to experimentally validate the hierarchical referencing technique. Recently we were able to obtain initial experimental verification which is quite promising [26].

Finally, while we provide an achievable strategy for the problem of distributed lossy compression considered here that outperforms conventional strategies in both theoretical and experimental analyses, it is unclear if it is close to optimal. Developing converse techniques for this distributed sensing setup is an exciting new information-theoretic problem. The problem could require novel information-theoretic tools because, as discussed in the intro, this requires choosing the optimal among all possible referencing trees.

ACKNOWLEDGMENTS

We thank Ted Huppert and Jay Deep Sau for helpful discussions, and Marlene Behrmann, Lori Holt, Mark Richardson and Michael Tarr for motivating the problem from a neuroscientific and clinical viewpoint. We also thank the CMU BrainHUB initiative for facilitating discussions between engineers and neuroscientists that led to the problems formulated and addressed here, and for support through the Proseed-BrainHUB seed funds. Pulkit Grover was supported through a CMU startup grant, and Praveen Venkatesh through a CMU CIT Dean's fellowship and a CMU Presidential Fellowship.

REFERENCES

[1] S. P. Ahlfors, J. Han, J. W. Belliveau, and M. S. Hämäläinen, "Sensitivity of MEG and EEG to source orientation," *Brain topography*, vol. 23, no. 3, pp. 227–232, 2010.

[2] P. L. Nunez and R. Srinivasan, *Electric fields of the brain: the neurophysics of EEG*. Oxford university press, 2006.

[3] S. Klamer, A. Elshahabi, H. Lerche, C. Braun, M. Erb, K. Scheffler, and N. K. Focke, "Differences Between MEG and High-Density EEG Source Localizations Using a Distributed Source Model in Comparison to fMRI," *Brain topography*, vol. 28, no. 1, pp. 87–94, 2015.

[4] S. J. Luck, *An introduction to the event-related potential technique*. MIT press, 2014.

[5] R. Srinivasan, D. M. Tucker, and M. Murias, "Estimating the spatial Nyquist of the human EEG," *Behavior Research Methods, Instruments, & Computers*, vol. 30, no. 1, pp. 8–19, 1998.

[6] P. Grover, "Revisiting the nyquist-rate of human eeg," in *Full version of the paper submitted to ICASSP 2016*. [Online]. Available: http://tinyurl.com/pulkitgrover/files/NyquistRateHumanEEGFull.pdf

[7] ——, "Ultra-high-density EEG systems: why and how?" in *Asilomar Conference on Signals, Systems, and Computers*, 2015, to appear.

[8] W. J. Freeman, M. D. Holmes, B. C. Burke, and S. Vanhatalo, "Spatial spectra of scalp EEG and EMG from awake humans," *Clinical Neurophysiology*, vol. 114, no. 6, pp. 1053–1068, 2003.

[9] M. Odabaee, W. J. Freeman, P. B. Colditz, C. Ramon, and S. Vanhatalo, "Spatial patterning of the neonatal EEG suggests a need for a high number of electrodes," *NeuroImage*, vol. 68, pp. 229–235, 2013.

[10] C. Ramon, M. Holmes, W. J. Freeman, M. Gratkowski, K. Eriksen, and J. Haueisen, "Power spectral density changes and language lateralization during covert object naming tasks measured with high-density EEG recordings," *Epilepsy & Behavior*, vol. 14, no. 1, pp. 54–59, 2009.

[11] Y. Petrov, J. Nador, C. Hughes, S. Tran, O. Yavuzcetin, and S. Sridhar, "Ultra-dense EEG sampling results in two-fold increase of functional brain information," *NeuroImage*, vol. 90, pp. 140–145, 2014.

[12] Texas Instruments, "ADS 1299: Low-Noise, 8-Channel, 24-Bit Analog Front-End for Biopotential Measurements," 2015. [Online]. Available: http://www.ti.com/product/ads1299

[13] A. D. Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, p. 1, 1976.

[14] K. F. Riley and M. P. Hobson, *Essential mathematical methods for the physical sciences*. Cambridge University Press, 2011.

[15] Y. Zhang, W. van Drongelen, and B. He, "Estimation of in vivo brain-to-skull conductivity ratio in humans," *Applied physics letters*, vol. 89, no. 22, 2006.

[16] "Code for a case study on power savings for high-density EEG using a hierarchical measurement technique." [Online]. Available: http://www.tinyurl.com/pulkitgrover/EEGCode.html

[17] R. R. Ramrez, "Source localization," *Scholarpedia*, vol. 3, no. 11, p. 1733, 2008.

[18] R. Schreier and G. C. Temes, *Understanding delta-sigma data converters*. IEEE press Piscataway, NJ, 2005, vol. 74.

[19] Maxim Integrated, "Demystifying Delta-Sigma ADCs," 2015. [Online]. Available: http://www.maximintegrated.com/en/app-notes/index.mvp/id/1870

[20] J. C. Mosher, M. E. Spencer, R. M. Leahy, and P. S. Lewis, "Error bounds for eeg and meg dipole source localization," *Electroencephalography and clinical Neurophysiology*, vol. 86, no. 5, pp. 303–321, 1993.

[21] R. N. Henson, E. Mouchlianitis, and K. J. Friston, "Meg and eeg data fusion: simultaneous localisation of face-evoked responses," *Neuroimage*, vol. 47, no. 2, pp. 581–589, 2009.

[22] S. Kim, C. J. Quinn, N. Kiyavash, and T. P. Coleman, "Dynamic and succinct statistical analysis of neuroscience data," *Proceedings of the IEEE*, vol. 102, no. 5, pp. 683–698, 2014.

[23] P. Venkatesh and P. Grover, "Is the direction of greater Granger causal influence same as the direction of information flow? ," in *Proceedings of the Allerton Conference on Communication, Control, and Computing, to appear*, Monticello, IL, Oct. 2015.

[24] ——, "Is the direction of Granger causality same as the direction of information flow? ," in *The Annual Meeting of the Society for Neuroscience (SfN)*, Chicago, IL, Oct. 2015.

[25] M. Mahzoon, H. Albalawi, X. Li, and P. Grover, "Using relative-relevance of data pieces for efficient communication, with an application to neural data acquisition," in *52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2014, pp. 160–166.

[26] M. J. Boring, S. Kelly, J. Weldon, A. Robinson, M. Behrmann, and P. Grover, "Experimentally challenging theoretical eeg correlations: Does a hierarchical-referencing strategy lead to bit savings?" *Submitted*, 2015.