USC–SIPI REPORT #294

MEG-Based Imaging of Focal Neuronal Current Sources

by

James W. Phillips, Richard M. Leahy
and John C. Mosher
December 1995

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Room 404
Los Angeles, CA 90089-2564 U.S.A.

Abstract

We describe a new approach to imaging neuronal current sources from measurements of the magnetoencephalogram (MEG) associated with sensory, motor, or cognitive brain activation. Previous approaches to this problem have concentrated on the use of weighted minimum norm inverse methods. While these methods ensure a unique solution, they do not introduce information specific to the MEG inverse problem, often producing overly smoothed solutions and exhibiting severe sensitivity to noise. We describe a Bayesian formulation of the inverse problem in which a Gibbs prior is constructed to reflect the sparse focal nature of neuronal current sources associated with evoked response data. We demonstrate the method with simulated and experimental phantom data, comparing its performance with several weighted minimum norm methods.

Key words: magnetoencephalogram, minimum norm, mean field annealing, Bayesian imaging.

I. INTRODUCTION

The magnetoencephalogram (MEG), and the associated electroencephalogram (EEG), provide unique insights into the dynamic behavior of the human brain, as they are able to follow changes in neural activity on a millisecond time scale. In comparison, the other functional neuroimaging modalities (functional magnetic resonance imaging, and single photon and positron emission computed tomography) are limited in temporal resolution to at best one second, due to physiological time constants and signal to noise considerations. Furthermore, MEG signals are produced directly by electrical brain activity through which the brain communicates, rather than the indirect correlates, such as regional cerebral blood flow or glucose metabolism, that are imaged by the alternative functional modalities. In this paper we address the problem of producing an image of the three dimensional distribution of neural current sources that produced an observed MEG signal. The inherent ill-posedness of the associated quasistatic electromagnetic inverse problem, compounded by the limited number of spatial measurements available with current MEG systems, presents a difficult challenge for standard methods for image reconstruction. In the following we will describe an alternative Bayesian approach that uses a physiologically based model for the source to estimate a plausible solution from the available data.

An array of superconducting quantum interference device (SQUID) biomagnetometers may be used to measure the spatio-temporal MEG signal produced by the brain. Physiological models for the MEG assume primary sources are constrained to the cortex with current flow ori-

ented normal to the local surface [20]. The image can therefore be constrained to the cortical surface, which can be extracted from a registered volume MR image of the subject's head. By tessellating the cortex with N disjoint regions and representing the sources in each region by an equivalent constrained current dipole [7] oriented normal to the surface with amplitude y_i, the MEG inverse problem can be expressed in terms of a linear model. The linear forward model relating the N sources $\mathbf{y}$ (Nx1) and the M MEG measurements $\mathbf{b}$ (Mx1) can be written,

$$\mathbf{b} = \mathbf{G}\mathbf{y} + \mathbf{n} \tag{1}$$

where the i'th row of the MxN system matrix $\mathbf{G}$ is a discrete representation of the lead field (sensitivity) of the i'th sensor. The j'th column of $\mathbf{G}$ specifies the gain vector for the j'th constrained dipole component. The Mx1 vector $\mathbf{n}$ represents noise generated within the sensor and by unwanted electromagnetic sources (power lines, the heart, background brain activity, etc.), or from modeling errors.

We assume a quasistatic approximation, since the time-derivatives of the electric and magnetic field are small compared to the ohmic current [26]. If the head is modeled as a spherically symmetric conductor, the radial component of the magnetic field outside the head is entirely due to the primary current; consideration of the volume or return currents can be neglected. In this case, the relationship between the j'th source dipole and the radial magnetometer measurement at the i'th sensor is [24],

$$G_{ij} = \frac{\mu_0}{4\pi} \frac{\mathbf{r}_i \times \mathbf{r}_j}{|\mathbf{r}_j - \mathbf{r}_i|^3 |\mathbf{r}_i|} \cdot \mathbf{e}_j \tag{2}$$

where μ_0 is the permittivity of free space, $\mathbf{r}_j$ and $\mathbf{r}_i$ are the locations of the dipole j and sensor i respectively, and $\mathbf{e}_j$ is a unit magnitude vector with the j'th constrained source's orientation.

In practice of course, the head is not spherical and the sensors do not make point measurements of the radial component of the magnetic field. To reduce sensitivity to distant sources and the earth's magnetic field, a gradiometer is used to measure the difference in the magnetic flux through two adjacent coils. The forward model is easily modified to include the effects of the gradiometer and of non-radial sensors [17]. Finite and boundary elements can be used to further extend the model to include non-spherical head geometries [5]. While the forward model in these cases becomes increasingly complex, the problem remains linear and the reconstruction methods described below can still be used. In the experimental results presented below we use a spherical

head geometry but specifically model the non-radial, planar gradiometer configuration of the Neuromag122 system used in our work [24].

A linear relationship also exists between the primary current density and the potential difference between two electrodes measured by the electroencephalogram (EEG). Mosher *et al.* [19] show quantitatively that MEG and EEG provide complementary information. In the spherical head model, MEG is sensitive only to the tangential component of the primary current density, whereas EEG is sensitive to all components, and the MEG lead field magnitude falls off more quickly than EEG near the center of the sphere. For radial sensors, EEG is affected by the skull and scalp much more than MEG, however. To obtain the best conditioned data set, therefore, the two may be concatenated, forming a new $(M+L)$xN gain matrix where L is the number of EEG leads available. In the following however, we will concentrate on MEG only.

Many inverse procedures for EEG/MEG can be classified as either multiple dipole or imaging. In multiple dipole methods, a small set of current dipoles are used to model neural activity at a small number of distinct sites. The unknown dipole locations and moments form a set of unknown parameters which are typically found using a non-linear least squares fit to the observed data. A more recent method, which avoids some of the problems of mis-locating sources due to the non-convexity of the error surface with respect to the location parameters, is based on the MUltiple SIgnal Classification (MUSIC) algorithm introduced earlier in another context [18]. Multiple dipole methods are optimal for point sources but may perform poorly if sources are distributed over a significant area. While multipole expansions could be used to represent these more distributed sources, this approach has not received much recent attention in the MEG literature. Imaging is also well suited for distributed source representation, and has the advantage that the inverse problem is linear and that anatomical constraints can be easily implemented. The major disadvantage of the imaging approach over multiple dipole methods is that the number of unknowns can rise dramatically and the problem is typically highly under-determined.

The search for an appropriate imaging method is primarily concerned with finding a way to choose within a set of images that produces essentially the same fit to the data. Weighted minimum norm inverse methods [22,24,28] typically find solutions which match the data while minimizing a weighted l_2-norm on the solution vector. These techniques tend to smear sources over the entire reconstruction region and are generally unstable due to ill-conditioning of the system matrix. The instability can be overcome using Tikhonov regularization [13] but the reconstructions remain smeared. The iteratively reweighted minimum norm method [10][11] is a nonlinear

approach to overcoming the problem of smeared sources in which the norm weighting is updated at each iteration based on the result of the previous iteration. This method uses a weighting matrix which, as the iterations proceed, reinforces strong sources and reduces weak ones. This results in very sparse solutions, but again the method is extremely sensitive to noise and highly dependent on the initial estimate.

Here we propose an alternative approach to the inverse problem based on a Bayesian formulation. Rather than use an arbitrary weighted l_2-norm to select the solution, we introduce a prior distribution on the source which is used to resolve the ambiguities inherent in the inverse problem. This prior is constructed to favor the reconstruction of physiologically plausible solutions. Basic studies of functional activation, such as somatotopic or retinotopic mapping using fMRI or PET, reveal the sparse and highly localized nature of activation in the cerebral cortex [1]. Our prior is therefore specifically designed to reflect the expectation that current sources tend to be sparse and focal. This prior is combined with a Gaussian likelihood model for the data which is based on the linear model (1) and an assumption of additive white Gaussian noise. Maximization over the resulting posterior probability results in a maximum *a posteriori* (MAP) estimate of the primary current sources.

In the following, we briefly describe the regularized, weighted minimum norm methods. We then develop our Bayesian approach and describe a numerical procedure for computing a MAP image estimate. We compare the performance of the different methods for simulated data and experimental phantom data from a clinical Neuromag 122 channel MEG system.

II. Minimum Norm Approaches

The weighted minimum l_2-norm approach to linear inverse problems involves solving the constrained optimization problem:

$$\mathbf{y}_{\text{wmn}} = \arg\min_{\mathbf{y}} \quad \mathbf{y}^{\mathrm{T}}\mathbf{C}_{\mathbf{y}}^{-1}\mathbf{y} \quad \text{such that} \quad |\mathbf{b} - \mathbf{G}\mathbf{y}|^2 = 0 \tag{3}$$

where $\mathbf{C}_{\mathbf{y}}$ is an arbitrary symmetric positive definite matrix. Writing $\mathbf{C}_{\mathbf{y}} = \mathbf{W}\mathbf{W}^{\mathrm{T}}$, we can form the solution as,

$$\begin{aligned} \mathbf{y}_{\text{wmn}} &= \mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}(\mathbf{G}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}})^{-1}\mathbf{b} \\ &= \mathbf{W}(\mathbf{G}\mathbf{W})^{\dagger}\mathbf{b} \end{aligned} \tag{4}$$

where $(\mathbf{GW})^{\dagger}$ denotes the pseudoinverse of $\mathbf{GW}$. The weighted minimum norm solution can also be formed from the singular value decomposition (SVD) of $\mathbf{GW}$:

$$\mathbf{y}_{\text{wmn}} = \mathbf{W}\sum_{i=1}^{M}\left(\frac{\mathbf{u}_i^{\mathrm{T}}\mathbf{b}}{s_i}\right)\mathbf{v}_i \tag{5}$$

where s_i, $\mathbf{v}_i$, and $\mathbf{u}_i$ are the i'th singular value and corresponding right and left singular vectors of $\mathbf{GW}$, respectively.

Several forms of $\mathbf{W}$ have been proposed for MEG imaging applications. In [24] and [28], the weight matrix is implicitly the identity matrix ($\mathbf{W}=\mathbf{I}$). In the normalized minimum norm method [16], $\mathbf{W}_{\text{norm}} = \text{diag}(1/\|\mathbf{g}_1\|, 1/\|\mathbf{g}_2\| \ldots, 1/\|\mathbf{g}_N\|)$, where $\|\mathbf{g}_i\|$ is the Euclidean norm of the i'th column of $\mathbf{G}$. This weighting is designed to compensate for the reduced sensitivity of MEG to deep sources which results in a preference for superficial distributions when $\mathbf{W}=\mathbf{I}$. The NxN Laplacian operator $\mathbf{B}$ is commonly selected to smooth the reconstruction of minimum norm techniques such that $\mathbf{W} = \mathbf{B}^{-1}$ [9](pg. 562)[21]. The LORETA technique [22] uses a weighting matrix $\mathbf{W}_{\text{L}} = \mathbf{W}_{\text{norm}}\mathbf{B}^{-1}$. The Laplacian operator tends to select smoother reconstructions than either of the previous two weight functions. Dale and Sereno [6] use a weighting based on estimates of signal power at each location. These estimates can come from other modalities or from physiological knowledge.

Exact matching of the data in (3) results in ill-conditioning and high sensitivity to noise. Regularized forms of weighted minimum norm methods lower this noise sensitivity. One popular technique is to simply truncate the summation in (5) at a threshold index, known as truncated SVD [25]. An alternative method, Tikhonov regularization [13], replaces the original problem (3) with the unconstrained minimization of a combination of the residual error norm and weighted l_2-norm of the solution vector:

$$y_{\lambda} = \arg\min_{\mathbf{y}} \quad |\mathbf{b}-\mathbf{G}\mathbf{y}|^2 + \lambda \mathbf{y}^{\mathrm{T}}\mathbf{C}_{\mathbf{y}}^{-1}\mathbf{y}. \tag{6}$$

For any value of the regularization parameter λ, the solution can be found as:

$$\mathbf{y}_{\lambda} = \mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\left(\mathbf{G}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}} + \lambda\mathbf{I}\right)^{-1}\mathbf{b} = \mathbf{W}\sum_{i=1}^{M} f_i\left(\frac{\mathbf{u}_i^{\mathrm{T}}\mathbf{b}}{s_i}\right)\mathbf{v}_i \tag{7}$$

where the filter parameters $f_i = s_i^2/(s_i^2+\lambda)$. Note that the filter coefficients decrease as s_i decreases. Therefore, the contributions of $(\mathbf{u}_i^{\mathrm{T}}\mathbf{b}/s_i)\,\mathbf{v}_i$ to the solution from the smaller s_i are

effectively filtered out. In our implementation of the regularized weighted minimum norm methods we use a *regularization toolbox* [13] which selects an optimal value for λ using the L-curve method [14].

The iteratively reweighted minimum norm approach [11], also known as focal underdetermined system solution (FOCUSS)[10], is a novel inverse method which iteratively updates the weight matrix using the result of the previous iteration. The algorithm first provides an initial estimate of $\mathbf{y}(0)$ using (4) with $\mathbf{W}_{\text{norm}}$ as the weighting matrix. At each step thereafter, a separate weight matrix $\mathbf{W}(k) = \text{diag}(\mathbf{y}(k-1))$ is formed. At each iteration, we update the solution using,

$$\mathbf{y}(k) = (\mathbf{W}_{\text{norm}}\mathbf{W}(k))\,(\mathbf{G}\mathbf{W}_{\text{norm}}\mathbf{W}(k))^{\dagger}\mathbf{b}. \tag{8}$$

For any starting point $\mathbf{y}(0)$ asymptotic convergence to a fixed point is guaranteed [11]. The fixed points are, however, unknown and highly dependent on the initial estimate. Also, the final error in fitting the data may be large, even for the noiseless case. In our implementation we use Tikhonov regularization and the L-curve method at each iteration to avoid ill-conditioning.

III. Bayesian Interpretation of Minimum Norm

The minimum norm methods can be re-interpreted in a Bayesian format. If we have additional statistical knowledge of the structure of the image $\mathbf{y}$, then we can incorporate this information to pick a single reconstruction from the set of primary current distributions consistent with the measurements. By Bayes rule,

$$p(\mathbf{y}|\mathbf{b}) = \frac{p(\mathbf{b}|\mathbf{y})p(\mathbf{y})}{p(\mathbf{b})} \tag{9}$$

Maximizing over the log posterior to find the Maximum APosteriori (MAP) estimate,

$$\mathbf{y}^{*} = \underset{\mathbf{y}}{\arg\max} \ln\{p(\mathbf{b}|\mathbf{y})\} + \ln\{p(\mathbf{y})\}. \tag{10}$$

If $\mathbf{y}$ and $\mathbf{n}$ are independent and normally distributed with zero mean and covariance $\mathbf{C}_{\mathbf{y}}$ and $\mathbf{C}_{\mathbf{n}}$, then the MAP estimate is,

$$\mathbf{y}^{*} = \underset{\mathbf{y}}{\arg\max}\left(-(\mathbf{G}\mathbf{y}-\mathbf{b})^{\mathrm{T}}\mathbf{C}_{\mathbf{n}}^{-1}(\mathbf{G}\mathbf{y}-\mathbf{b}) - \mathbf{y}^{\mathrm{T}}\mathbf{C}_{\mathbf{y}}^{-1}\mathbf{y}\right) \tag{11}$$

That is,

$$\mathbf{y}^{*} = \mathbf{C}_{\mathbf{y}}\mathbf{G}^{\mathrm{T}}(\mathbf{G}\mathbf{C}_{\mathbf{y}}\mathbf{G}^{\mathrm{T}} + \mathbf{C}_{\mathbf{n}})^{-1}\mathbf{b}. \tag{12}$$

For the noiseless case ($\mathbf{C}_n = \mathbf{0}$), this result is identical to (4), the weighted minimum norm solution. Therefore, the various weight matrices can be interpreted in terms of the covariances of a Gaussian prior with $\mathbf{C}_y = \mathbf{W}\mathbf{W}^T$. If we assume $\mathbf{C}_n = \lambda\mathbf{I}$, we arrive at the general Tikhonov regularized form as given in (7).

Viewed in this way, the choice of weight matrices used in weighted minimum norm methods are equivalent to a somewhat arbitrary selection of the covariance matrix of a Gaussian prior. The alternative formulation presented below replaces this arbitrary Gaussian prior with one chosen to better reflect the expected attributes of the current sources that produced the observed data.

IV. The New Bayesian Approach

Since the primary sources of the MEG are widely accepted to be sparse and focal [23], we suggest that this information be used in reconstructing the image. This information can be naturally introduced into the problem using the Bayesian paradigm in which the source is modeled as a random field. Since we assume that sources are sparse, the large majority of source pixels will have zero amplitude. We therefore use a binary indicator process $\mathbf{x}$ to model whether each source dipole is *on* ($x_i = 1$) or *off* ($x_i = 0$). Those sites that are active are assumed to have a Gaussian amplitude, z_i. We can then write the source image vector $\mathbf{y}$ as,

$$\mathbf{y} = \mathbf{x}.*\mathbf{z} = \mathbf{X}\mathbf{z} \tag{13}$$

where ".*" signifies the Schur product (element by element pair-wise multiplication) and $\mathbf{X} = \mathrm{diag}(\mathbf{x})$. Assuming independence of the indicator and amplitude processes, we can write the posterior probability for $\mathbf{x}$ and $\mathbf{z}$ given the MEG data $\mathbf{b}$ as,

$$p(\mathbf{x}, \mathbf{z}|\mathbf{b}) = \frac{p(\mathbf{b}|\mathbf{x}, \mathbf{z})p(\mathbf{x})p(\mathbf{z})}{p(\mathbf{b})}. \tag{14}$$

We find a MAP estimate $\mathbf{y}^*$ of $\mathbf{y}$ as $\mathbf{y}^* = \mathbf{x}^*.*\,\mathbf{z}^*$ where,

$$\mathbf{x}^*, \mathbf{z}^* = \underset{\mathbf{x}, \mathbf{z}}{\arg\max}\ p(\mathbf{x}, \mathbf{z}|\mathbf{b}). \tag{15}$$

The joint probability $p(\mathbf{x})$ is chosen to reflect the expectation that the sources are sparse and focal. To achieve this goal, we use a Markov Random Field (MRF) [8] model for which

sparse focal sources have a higher probability of occurring than more distributed sources. We define $p(\mathbf{x})$ to be a Gibbs distribution,

$$p(\mathbf{x}) = (1/K) \exp\{-V(\mathbf{x})\} \tag{16}$$

where K is a proportionality constant and the energy function $V(\mathbf{x})$ is given by,

$$V(\mathbf{x}) = \sum_i \left[\underbrace{\alpha_i x_i}_{\text{Sparseness Term}} + \underbrace{\beta_i C_i\{x_i, x_j \;\; j \in \xi_i\}}_{\text{Clustering Term}} \right] \tag{17}$$

where the parameters $\alpha_i > 0$ and $\beta_i > 0$ determine the relative weights of the sparseness and clustering terms. The potential function $C_i\{x_i, x_j \;\; j \in \xi_i\}$ is defined in terms of each pixel and its neighbors ξ_i as,

$$C_i\{x_i, x_j \;\; j \in \xi_i\} = \left[\sum_{j \in \xi_i} (x_i - x_j)^2 \right]^Q . \tag{18}$$

This clustering term is small if neighboring pixels are of the same magnitude. The exponential parameter Q determines the strength of the clustering. As Q increases, the size of the clusters tends to increase. In the results presented below we constrain sources to a plane and use a four nearest-neighbor interaction. Some examples of binary images, produced by sampling from this prior using a Gibbs Sampler [8], are shown in Fig. 1. The source amplitude process, $\mathbf{z}$, is assumed to be a set of independent zero mean Gaussian random variables with covariance $\mathbf{C_z}$.

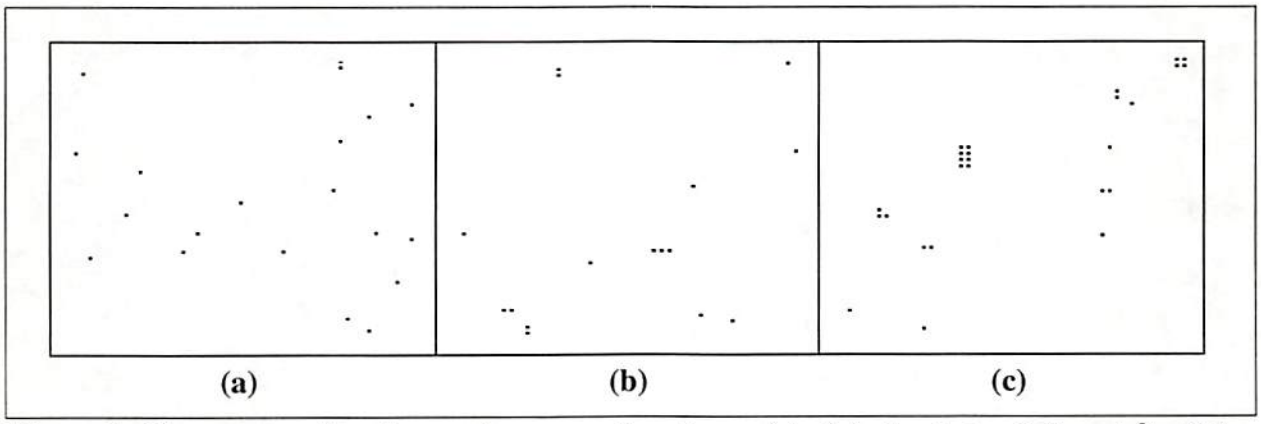

Figure 1. Three examples of a random sampling from $p(\mathbf{x})$. (a) $Q = 1$, $\alpha = 1.0$, and $\beta = 0.4$ (b) $Q = 2$, $\alpha = 1.0$, and $\beta = 0.2$ (c) $Q = 3$, $\alpha = 1.0$, and $\beta = 0.1$.

Using the definitions above and assuming the noise process in (1) is zero mean Gaussian with covariance $\mathbf{C_n}$, we can write,

$$p(\mathbf{x}, \mathbf{z}|\mathbf{b}) = \frac{1}{Z}\exp\{-U(\mathbf{x}, \mathbf{z}|\mathbf{b})\} \tag{19}$$

where Z is the posterior partition function. The posterior energy function is given by:

$$U(\mathbf{x}, \mathbf{z}|\mathbf{b}) = \frac{1}{2}[\mathbf{b} - \mathbf{GXz}]^{\mathrm{T}}\mathbf{C}_{\mathbf{n}}^{-1}[\mathbf{b} - \mathbf{GXz}] + \frac{1}{2}\mathbf{z}^{\mathrm{T}}\mathbf{C}_{\mathbf{z}}^{-1}\mathbf{z} + V(\mathbf{x}) \; . \tag{20}$$

The MAP estimate is found by maximizing over the log-posterior, or equivalently,

$$\mathbf{x}^*, \mathbf{z}^* = \underset{\mathbf{x}, \mathbf{z}}{\arg\min} \quad U(\mathbf{x}, \mathbf{z}|\mathbf{b}) . \tag{21}$$

The solution to the optimization problem (21) provides our estimate of the neural current sources. The solution is clearly a function of the parameters of the likelihood function and the prior probabilities. Methods for selecting these parameters will be addressed in a future publication. However, we note that since the parameters α_i represent the relative probabilities that each source pixel is active, it is straightforward to include pixel-wise probability weightings determined from fMRI or PET activation studies to influence the formation of these sparse images.

V. Mean Field Annealing

Minimization of $U(\mathbf{x}, \mathbf{z}|\mathbf{b})$ is difficult since the optimization must be performed over a mixture of discrete and continuous variables. Since this function is quadratic in the continuous variables, $\mathbf{z}$, we can derive a closed form expression for the optimal $\mathbf{z}$ as a function of any particular indicator process $\mathbf{x}$. i.e. the vector $\mathbf{z}^*(\mathbf{x})$ which minimizes (20) given the binary vector $\mathbf{x}$ is given by,

$$\mathbf{z}^*(\mathbf{x}) = \mathbf{C_z}\mathbf{X}\mathbf{G}^{\mathrm{T}}(\mathbf{GXC_z}\mathbf{X}\mathbf{G}^{\mathrm{T}} + \mathbf{C_n})^{-1}\mathbf{b} . \tag{22}$$

Substituting $\mathbf{z}^*(\mathbf{x})$ into $U(\mathbf{x}, \mathbf{z}|\mathbf{b})$ results in,

$$\tilde{U}(\mathbf{x}|\mathbf{b}) = U(\mathbf{x}, \mathbf{z}|\mathbf{b})\big|_{\mathbf{z} = \mathbf{z}^*(\mathbf{x})} \tag{23}$$

which is a Gibbs energy function for the binary density $\tilde{p}(\mathbf{x}|\mathbf{b}) = (1/\tilde{K})\exp\{-\tilde{U}(\mathbf{x}|\mathbf{b})\}$. We can therefore first find the optimal indicator process by minimizing $\tilde{U}(\mathbf{x}|\mathbf{b})$, then substitute this result in (22) to obtain the optimal amplitude process.

Coordinate-wise optimization with respect to a collection of binary variables using, for example, iterated conditional modes (ICM), tends to produce rapid convergence to an undesirable local minimum. Here we use, instead, a continuation method based on mean field annealing (MFA) [3].

MFA changes an optimization problem into a series of expected value problems. We form a new density function based on a temperature parameter T,

$$\tilde{p}_T(\mathbf{x}|\mathbf{b}) = \frac{1}{K_T}\tilde{p}^{\frac{1}{T}}(\mathbf{x}|\mathbf{b}) = \frac{1}{\tilde{K}_T}\exp\{-\tilde{U}(\mathbf{x}|\mathbf{b})/T\} \tag{24}$$

where $\tilde{K}_T$ is a temperature dependent normalization parameter. The mean of the newly defined process $\tilde{p}_T(\mathbf{x}|\mathbf{b})$ approaches the mode of $\tilde{p}(\mathbf{x}|\mathbf{b})$ as $T \to 0$ in the following sense.

Theorem. If $\mathbf{x}^*$ is the unique global maximizer of $\tilde{p}(\mathbf{x}|\mathbf{b})$, then,

$$\lim_{T \to 0} E_T\{\mathbf{x}|\mathbf{b}\} = \mathbf{x}^*. \tag{25}$$

where $E_T\{\mathbf{x}|\mathbf{b}\}$ denotes the expected value of $\mathbf{x}$ with respect to $\tilde{p}_T(\mathbf{x}|\mathbf{b})$.

Proof. We can find $E_T\{\mathbf{x}|\mathbf{b}\}$ by the relation,

$$E_T\{\mathbf{x}|\mathbf{b}\} = \frac{1}{\tilde{K}_T}\sum_{\mathbf{x}} \mathbf{x}\exp\{-\tilde{U}(\mathbf{x}|\mathbf{b})/T\}. \tag{26}$$

If we separate $\mathbf{x}^*$ from the summation and multiply the numerator and denominator by $\exp\{\tilde{U}(\mathbf{x}^*|\mathbf{b})/T\}$, we can rewrite this as,

$$E_T\{\mathbf{x}|\mathbf{b}\} = \frac{\left(\sum_{\mathbf{x}\neq\mathbf{x}^*} \mathbf{x}\exp\{[\tilde{U}(\mathbf{x}^*|\mathbf{b}) - \tilde{U}(\mathbf{x}|\mathbf{b})]/T\}\right) + \mathbf{x}^*}{\left(\sum_{\mathbf{x}\neq\mathbf{x}^*} \exp\{[\tilde{U}(\mathbf{x}^*|\mathbf{b}) - \tilde{U}(\mathbf{x}|\mathbf{b})]/T\}\right) + 1}. \tag{27}$$

Since $\mathbf{x}^*$ is the unique global maximizer, $\tilde{U}(\mathbf{x}^*|\mathbf{b}) < \tilde{U}(\mathbf{x}|\mathbf{b}) \quad \forall\ \mathbf{x} \neq \mathbf{x}^*$. Therefore the summations go to zero as $T \to 0$ and (25) results.

Q.E.D.

Using (26) to determine $E_T\{\mathbf{x}|\mathbf{b}\}$ is infeasible due to the large dimensionality and coupling in $\mathbf{x}$. Instead we use a mean field approximation [4] to simplify the computation of the (approximate) mean. Let S represent all elements of the field. For each pixel i we create a univariate energy function $U_i^{\text{mf}}(x_i|\mathbf{b})$, written as,

$$U_i^{\text{mf}}(x_i|\mathbf{b}) = \tilde{U}(\mathbf{x}|\mathbf{b})\Big|_{\mathbf{x}_{S-i} = \bar{\mathbf{x}}_{S-i}} \tag{28}$$

where $\bar{\mathbf{x}}_{S-i}$ denotes the mean of the field restricted on S-i. We define $K_{i,T}^{\mathrm{mf}}$ to be the corresponding mean field local partition function,

$$K_{i,T}^{\mathrm{mf}} = \sum_{x_i} \exp\{-U_i^{\mathrm{mf}}(x_i|\mathbf{b})/T\} . \tag{29}$$

We then approximate the joint probability, $\tilde{p}_T(\mathbf{x}|\mathbf{b})$, as the product of independent mean field approximated marginal probabilities:

$$p_T^{\mathrm{mf}}(\mathbf{x}|\mathbf{b}) = \prod_i p_T^{\mathrm{mf}}(x_i|\mathbf{b}) = \frac{1}{K_T^{\mathrm{mf}}} \exp\left\{-\left(\sum_i U_i^{\mathrm{mf}}(x_i|\mathbf{b})\right)/T\right\} . \tag{30}$$

where the overall mean field partition function K_T^{mf} is the product of all local mean field partition functions. This mean field approximation effectively replaces the influence of the statistical fluctuations of neighboring sites by their mean values, where the means are computed with respect to the approximated probability $p_T^{\mathrm{mf}}(\mathbf{x}|\mathbf{b})$.

Using this mean field approximation, we can compute $\bar{x}_{i|\mathbf{b}}(T)$ at a given pixel site as:

$$\bar{x}_{i|\mathbf{b}}(T) = \left(\frac{1}{K_{i,T}^{\mathrm{mf}}}\right) \sum_{x_i \in \{0,1\}} x_i \exp\{-U_i^{\mathrm{mf}}(x_i|\mathbf{b})/T\} = \frac{1}{K_{i,T}^{\mathrm{mf}}} \exp\{-U_i^{\mathrm{mf}}(x_i = 1|\mathbf{b})/T\} . \tag{31}$$

Note that the mean is computed with respect to the approximation in (30) to the joint density $\tilde{p}_T(\mathbf{x}|\mathbf{b})$ in (24). Substituting $K_{i,T}^{\mathrm{mf}}$ from (29) and dividing numerator and denominator by $\exp\{-U_i^{\mathrm{mf}}(x_i = 1|\mathbf{b})/T\}$, we can simplify this to,

$$\bar{x}_{i|\mathbf{b}}(T) = \frac{1}{1 + \exp\left\{\dfrac{\Delta U_{i|\mathbf{b}}^{\mathrm{mf}}}{T}\right\}} \tag{32}$$

where the difference in energies is given by,

$$\Delta U_{i|\mathbf{b}}^{\mathrm{mf}} = U_i^{\mathrm{mf}}(x_i = 1|\mathbf{b}) - U_i^{\mathrm{mf}}(x_i = 0|\mathbf{b}) . \tag{33}$$

Note that to compute the mean of pixel i, we need the mean of all other pixels in the image since the statistical coupling between sites in the original probability is replaced with a coupling through their means. Therefore, the mean field must be computed iteratively. In our work, the algorithm cycles several times through all pixels using (32) to update their mean values.

The temperature parameter T is slowly reduced as the iterations proceed. As $T \to 0$, coordinate-wise updating using the conditional mean reduces to the method of Iterated Conditional Modes (ICM) [2]. Therefore, the asymptotic local convergence behavior of mean field annealing

is identical to that of ICM. It is easy to show that ICM will converge to a local minimum $\mathbf{x}^*$, i.e., to a point $\mathbf{x}^*$ where $\tilde{U}(\mathbf{x}^*|\mathbf{b}) < \tilde{U}(\mathbf{x}|\mathbf{b}) \quad \forall \quad \mathbf{x} \in \{0, 1\}^N$ such that $\mathbf{x}^*$ and $\mathbf{x}$ differ at no more than one site. The potential advantage of mean field annealing over direct application of ICM, is that by using the mean, in combination with an annealing schedule, we can slowly approach the solution, and hopefully avoid undesirable local minima.

Our algorithm to perform MFA is summarized as follows:

1. Set $T(0)$ and find an arbitrary initialization for $\bar{x}_{j|\mathbf{b}}[0]$, $j = 1 \ldots N$. Set the location of interest $i = 1$ and iteration number $k = 0$.
2. Use (22) to determine $\mathbf{z}^*(\mathbf{x}[k])$ for $x_i[k] = 0$ and $x_i[k] = 1$, all other $x_j[k]$ set to $\bar{x}_{j|\mathbf{b}}[k] \quad j \neq i$.
3. Use the z^* vectors to find $U_i^{\mathrm{mf}}(x_i = 1|\mathbf{b})$ and $U_i^{\mathrm{mf}}(x_i = 0|\mathbf{b})$.
4. Use (32) to calculate $\bar{x}_{i|\mathbf{b}}[k]$.
5. Increment i. If $i \leq N$, go back to #2, else go to #6.
6. Increment the iteration number k and decrease $T(k)$, set $i = 1$, and if not converged go to #2.

We have found empirically that a good choice for the annealing schedule $T(k)$ is:

$$T(k) = T_0 \exp\{-k/K_c\} \tag{34}$$

where K_c is a constant controlling the rate of $T(k)$'s decrease and T_0 is the initial temperature. Bilbro *et al.* [3] find that in MFA, after a certain "critical temperature", all locations in a binary problem move to the binary values without any direction change. They suggest at least 40 iterations to reach this temperature. Below this temperature, the function equivalently reduces to a smoothing of the binary problem. Hiriyannaiah *et al.* [15] have found the optimal initial temperature is noise dependent and of $O(\sigma^2)$, were σ^2 is the noise variance. They also find that a decrement of 5% or less in the temperature works well as a general annealing schedule. We find these numbers to be appropriate as well.

Note that this algorithm requires that we take the inverse of an MxM matrix twice in step #2. This can be computationally prohibitive, especially if we have a large number of sensors. To increase the computational speed, we recognize that when we change only one element from its current value to either a zero or a one, we effectively make a rank one update of the matrix $\mathbf{L} = \mathbf{G}\mathbf{X}\mathbf{C}_\mathbf{z}\mathbf{X}\mathbf{G}^\mathrm{T} + \mathbf{C}_\mathbf{n}$ in (22). A fast method for performing the inverse of a rank one updated matrix was proposed by Sherman and Morrison [9]. The method is based on the property that the

new inverse is a rank one update of the previous inverse. This method requires storing the previous $\mathbf{L}^{-1}$, but it reduces the order of computation from M^3 to M^2.

VI. Simulations

We have conducted extensive simulations based on a simplified 2D source model with a 1D array of 64 sensors. All dipolar sources are constrained to the annular segment of the x-y plane shown in Fig. 2. The distance between source locations was set to 2.8 mm providing an isotropically sampled image grid of 560 pixels. All dipoles were constrained in orientation perpendicular to the x-y plane. The source images were chosen using stochastic sampling from $p(\mathbf{x})$ and $p(\mathbf{z})$ with $Q = 3$, $\alpha_i = 1.15$, $\beta_i = 0.075$, and $\sigma_i = 10$ nAm $i = 1...N$. The first three parameters were chosen heuristically to provide good sparseness with a small degree of clustering. The fourth parameter, the standard deviation of the dipole moments, reflects the typical activity seen in an evoked response study [12]. We set $\mathbf{C_n}$ to $\nu^2 \cdot \mathbf{I}$, where ν^2 is the added noise variance which we assume known. For the study with noiseless data, we used a value in the algorithm of non-zero ν^2, comparable to a SNR of 50 dB, to avoid numerical instabilities.

The results of a representative simulation are shown in Fig. 3 for three scenarios, one noiseless, and two with added white Gaussian noise. The percent residual error printed on these images, is defined as,

$$\left[\frac{(\mathbf{Gy} - \mathbf{b})^{\mathrm{T}} (\mathbf{Gy} - \mathbf{b})}{\mathbf{b}^{\mathrm{T}} \mathbf{b}} \right] \cdot 100\% \tag{35}$$

where **y** is the solution shown. We compared the MAP technique to the four methods discussed in Section II: standard minimum l_2 norm technique, the normalized minimum norm technique, the LORETA method, and the iteratively reweighted minimum norm technique.

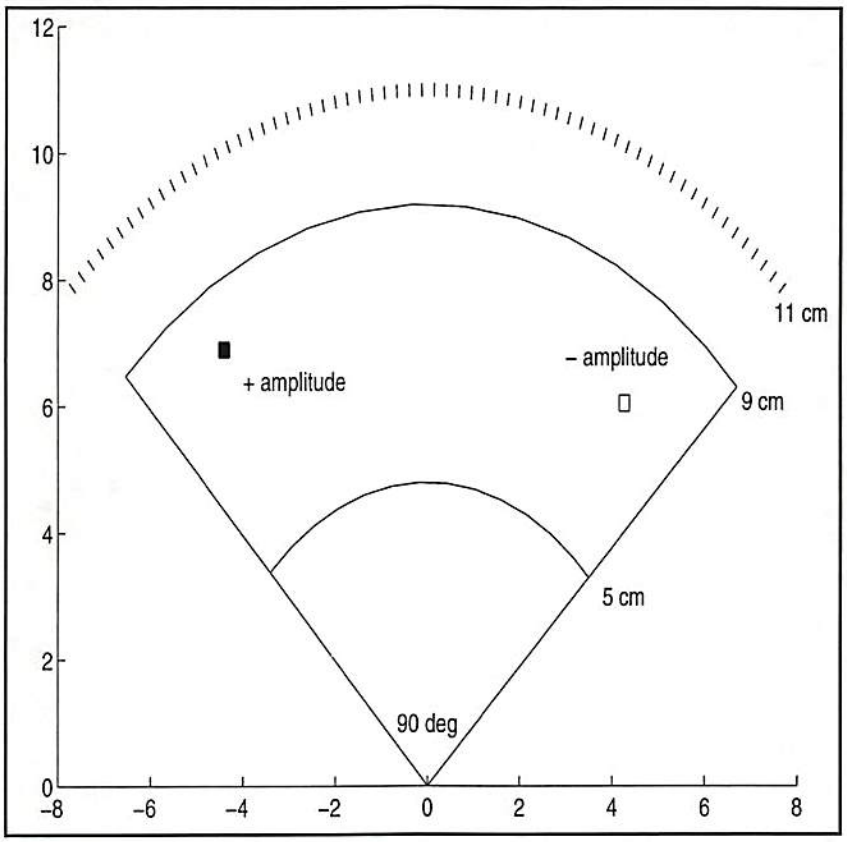

Figure 2. Sample plot showing the support of the source region (a 90 degree annulus from 5cm to 9cm) and the locations of the 64 MEG sensors (radial orientation, point measurement). Sources are shown as full boxes (positive amplitude) and empty boxes (negative amplitude), with the area of the box proportional to the magnitude of the source. In the following reconstructions, sources with magnitude less than one percent of the maximum are not included to improve the clarity of the figures.

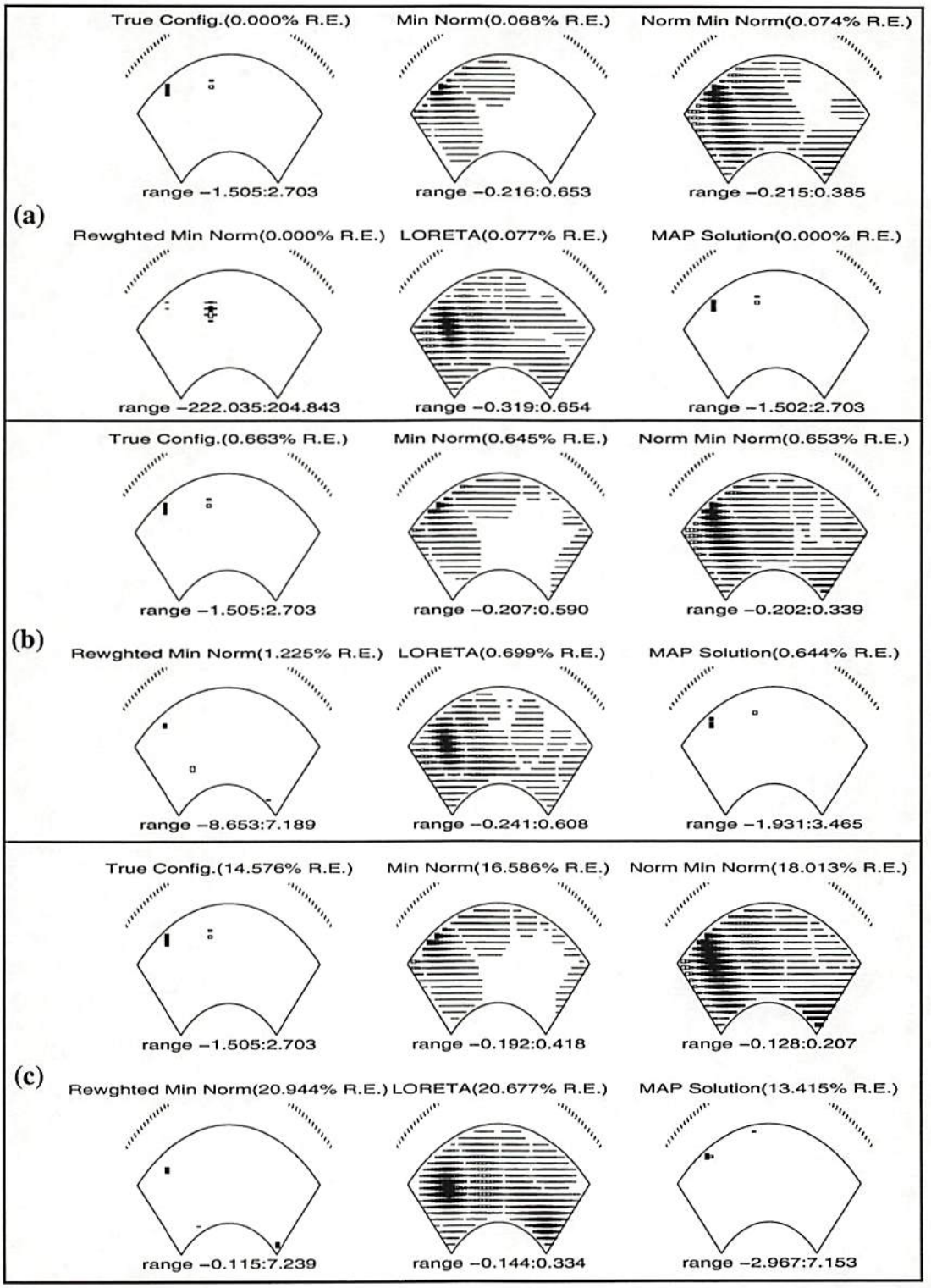

Figure 3. Sample 2D simulation with a 1D sensor array. 64 point measurements, 560 point image grid. The percent residual error (% R.E.) is shown above and the range of source values (in nAm) is shown below each image. Three scenarios are presented. (a) No noise added to the computed data (b) White Gaussian noise added to the data to achieve a SNR of 20 dB (c) White Gaussian noise added to the data to achieve a SNR of 8 dB.

These results show a wide variation in the characteristics of the solution obtained using different weighting functions. All of the linear minimum norm methods produce results exhibiting a relatively large degree of smoothness. In comparison, both the iteratively re-weighted minimum norm and the new Bayesian method produce very sparse solutions. For each scenario, all five methods generally give similar residual errors in the fit to the data. In some instances, the iteratively reweighted minimum norm method gives larger errors since there is nothing inherent in the method to limit increases in the error from one iteration to the next. All results can be considered 'correct' in the sense that they are configurations that could have produced the observed data. This observation emphasizes the severely under-determined nature of this problem. Clearly, in order to select between these feasible solutions we must use additional information concerning the expected nature of the source. We see that as the SNR decreases, performance of all methods deteriorates. In the 8dB case, the MAP solution clearly misplaces one of the source clusters, indicating that at this SNR, it is not possible to resolve between the true source configuration and the MAP reconstruction since both give similar fits to the data and both exhibit the sparse, clustered property preferred by our prior - in fact the MAP reconstruction gives this preference to the true solution in the sense of producing a lower posterior energy.

VII. PHANTOM EXPERIMENTS

We have applied the same techniques used in Section VI to experimental phantom data collected with a Neuromag-122 system [12] using the 28 dipole phantom supplied by the manufacturer. The Neuromag-122 system employs 61 dual-channel planar first-order gradiometer units in a helmet-shaped configuration at a radius of 10-11 cm, measuring the magnetic field gradient in two orthogonal tangential directions, for a total of 122 individual sensor measurements. The helmet for a Neuromag-122 is shown in Fig. 4.

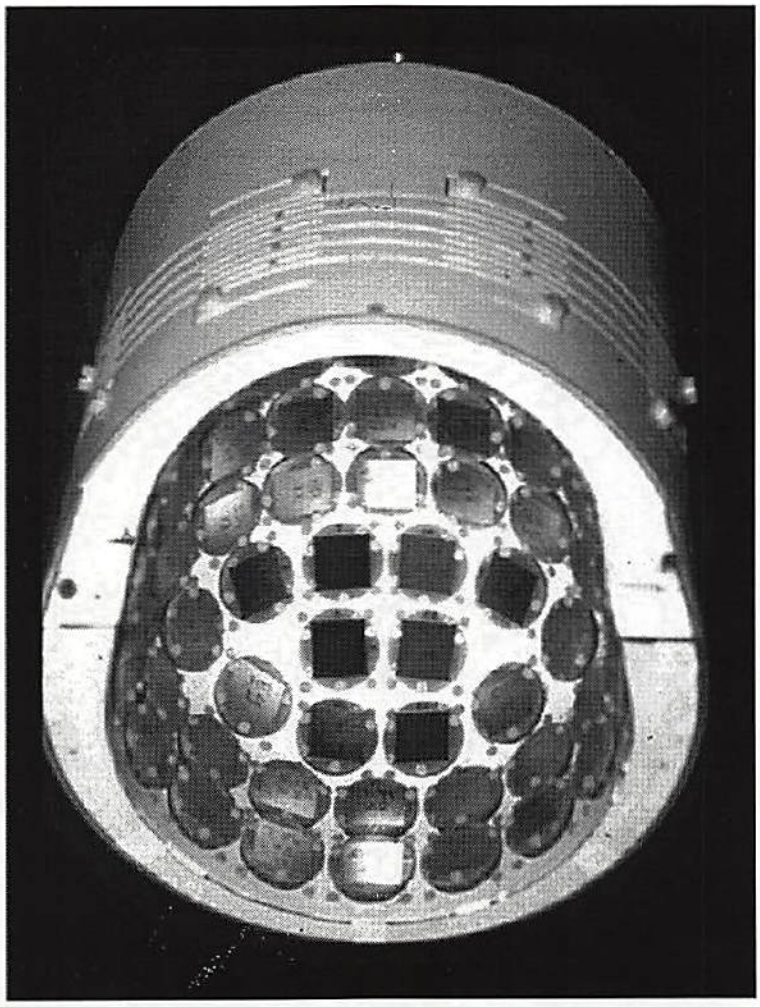

Figure 4. A view from below of the helmet of the Neuromag-122, showing the configuration of the 61 dual-channel planar first order gradiometers (Neuromag Ltd., Helsinki, Finland).

The phantom consists of two half circles with a 7 cm radius in the x-z plane and y-z plane, with dipoles in fixed positions in these planes oriented tangential to the outer edge. The image reconstruction grid consisted of 768 locations spaced 4 mm apart on two 180 degree annuli, with

an inner radius of 3 cm and an outer radius of 7 cm. A schematic of the imaging surface with the surrounding sensor elements is shown in Fig. 5, and a photograph of the phantom in Fig. 6. The

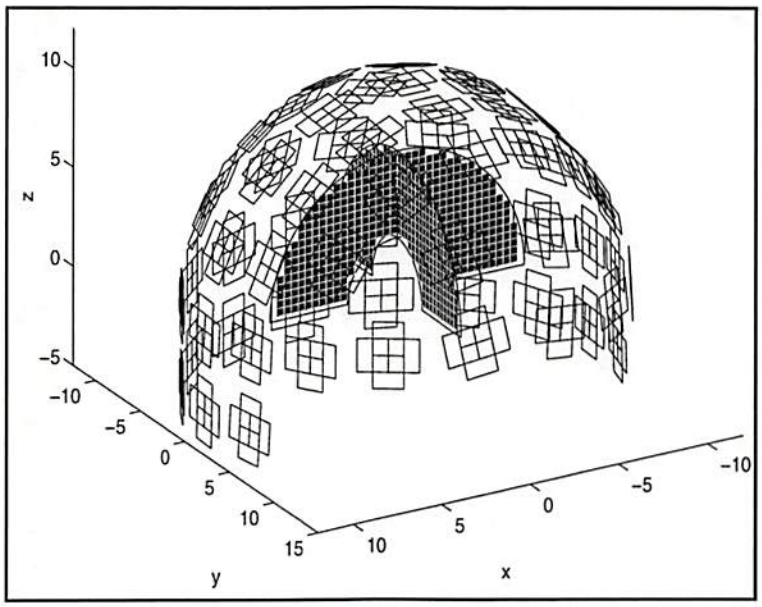

Figure 5. Schematic representation of the phantom reconstruction regions (two orthogonal, 180 degree annular regions) with the locations of the planar gradiometers of the neuromag-122 system superimposed.

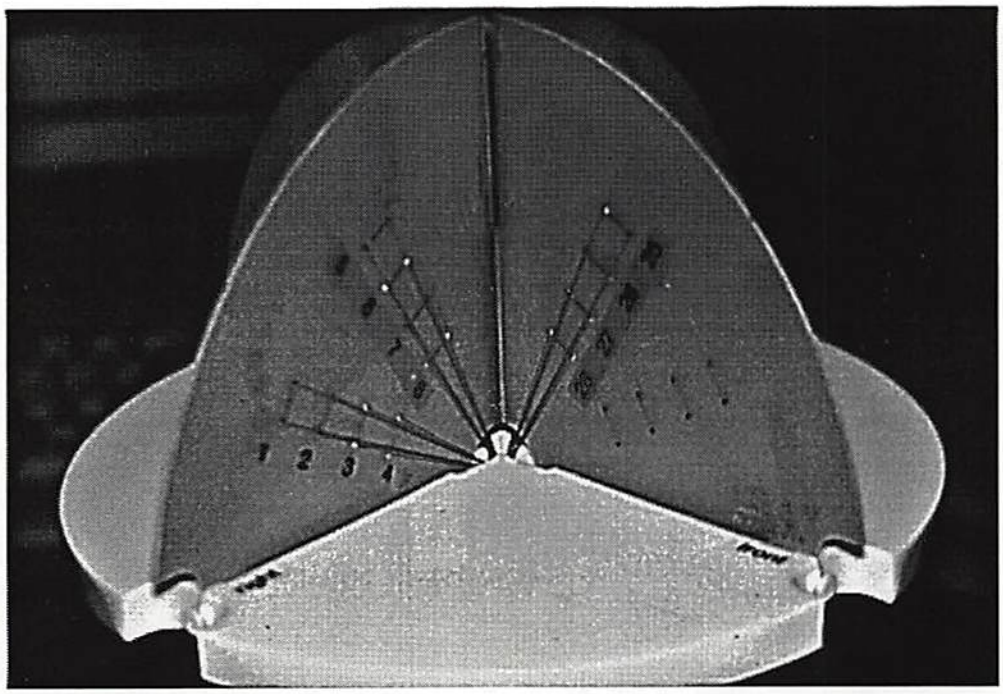

Figure 6. The phantom used in the Neuromag-122 experiments consists of 28 tangentially oriented dipoles on two orthogonal planes.

distance between sources and sensors creates a gain matrix which is more poorly conditioned than the one in the previous simulations. This increases the sensitivity to noise of all inverse methods. In all cases, the gain matrix was constructed to include gradiometer effects and non-radial sensor orientations assuming a spherical source volume [24].

The phantom data was scaled to reflect a reasonable evoked field response. We then added data collected in the same system from a passive human subject (100 averages of a pre-stimulus interval from an evoked response paradigm). This background was added to the phantom data to obtain a specified SNR. Fig. 7 shows the true configuration and the MAP solution plotted as a 2D projection of the three dimensional source distribution. In this example, noise was added to obtain

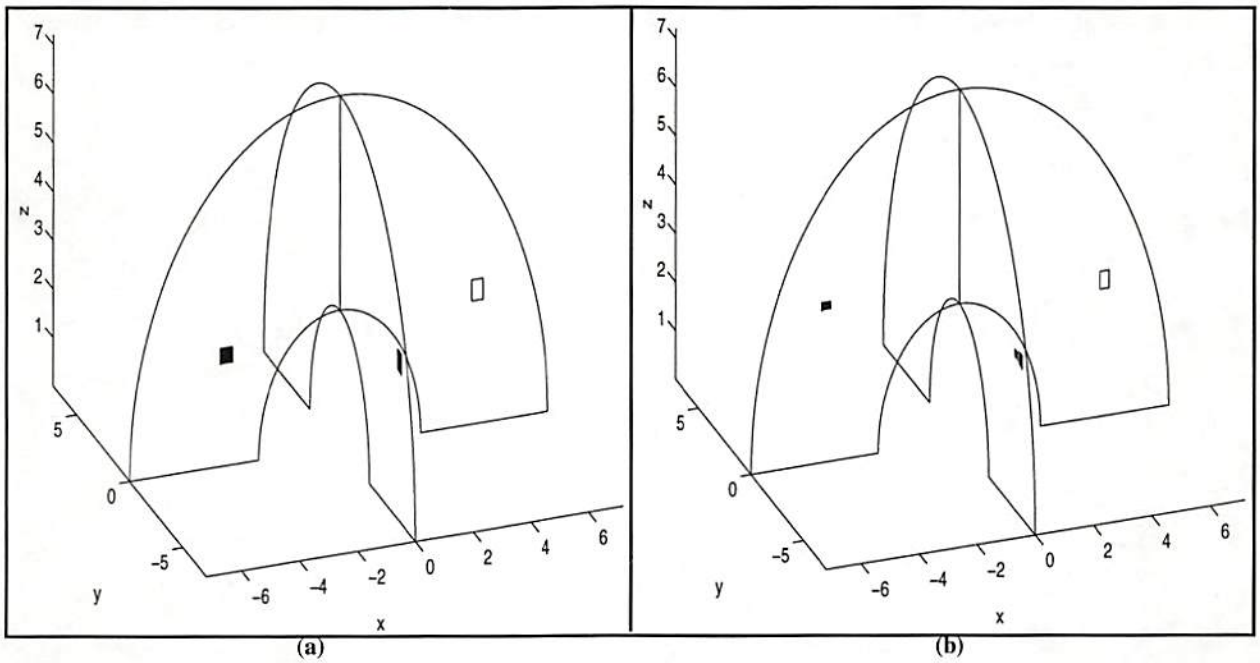

Figure 7. Reconstructions from a phantom experiment with three active dipoles. 122 point measurements, 768 point image grid. Colored subject noise was added to the data to achieve a SNR of 15 dB. (a) True configuration of phantom dipoles. (b) MAP Solution.

a SNR of 15 dB. The minimum norm techniques tested on phantom data performed comparably to the simulations in Section VI. The comparison plots are shown in Fig. 9 with high and low SNR. The phantom experiments show that our forward model is accurate and that the MAP technique can also be used effectively to reconstruct sparse sources from real data. We found that colored noise from the human subject was slightly more damaging than white Gaussian noise, since human subject noise tends to be smoother across the sensor array and therefore more likely to resemble spurious sources.

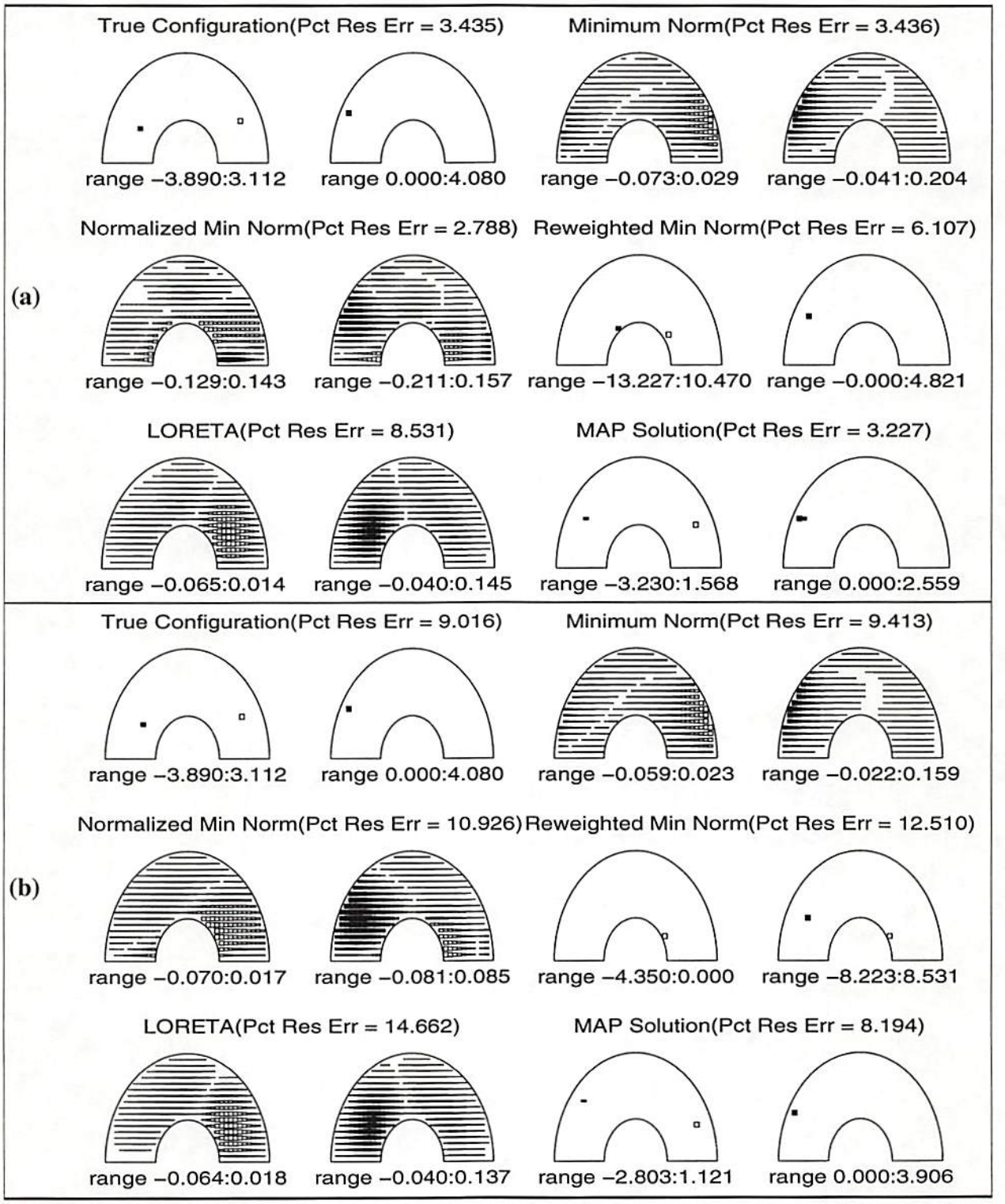

Figure 8. Sample phantom results with three active dipoles on two orthogonal planes. 122 point measurements, 768 point image grid. Percent residual error (Pct Res Err) is shown above and the range of source values (in nAm) is shown below each image. Two scenarios are presented. (a) Colored subject noise added to the data to achieve a SNR of 15 dB. (b) Colored subject noise added to the data to achieve a SNR of 10 dB.

VIII. CONCLUSIONS

In the simulations and phantom studies we have conducted, our MAP solution was generally superior to those obtained using minimum norm methods. However, this is true only when the sources exhibit the sparse focal characteristics on which our method is based. We stress that all of the methods provide good fits to the data, and hence are physically (if not physiologically) plausible. This ambiguity is inherent in attempting to infer spatial information from on the order of 100 external sensor measurements. Specific prior information is essential if useful spatial information is to be extracted from the data.

In this work we have developed a Bayesian framework for image estimation from MEG data. This approach can be extended to include information from other modalities (fMRI or PET) as well as using anatomical MR images to constrain sources to the cortex. The method can also be combined with more accurate forward models as discussed above, and also used for combined MEG/EEG data. We can also directly extend the model for dynamic imaging by simply replacing each of the amplitude processes, z_i, in our model with a time series model, $z_i(t)$.

The results that we have presented assume sources are constrained to 2D planes. Future research will focus on sources constrained to a realistic cortical surface. Only then can be we begin to establish realistic limits on the ability of MEG to usefully image neural activity.

We would like to thank Dr. Jeffrey Lewine of the New Mexico Institute of Neuroimaging, Albuquerque, for providing access to the Neuromag-122 system and data. We also wish to thank Dr. Cheryl Aine of the Biophysics group of The Los Alamos National Laboratory for her pre-stimulus subject data.

IX. REFERENCES

[1] C.J. Aine, "A Conceptual Overview and Critique of Functional Neuroimaging Techniques in Humans: 1. MRI/fMRI and PET," *Critical Reviews in Neurobiology*, vol. 9, nos. 2 and 3, pp. 229-309, 1995.

[2] J. Besag, "On the statistical analysis of dirty pictures (with discussion)," *J. of the Royal Statist. Soc., Series B*, vol. 48, pp. 259-302, 1986.

[3] G. Bilbro., *et al.* "Optimization by Mean Field Annealing," in: Touretzky D.S., *Advances in Neural Information Processing Systems*, San Mateo, CA: Morgan-Kauffman, pp. 91-98, 1989.

[4] D. Chandler, *Introduction to Modern Statistical Mechanics*, Oxford University Press, 1987, pp. 131-138.

[5] B.N. Cuffin, "A Method for Localizing EEG Sources in Realistic Head Models," *IEEE Trans on Bio. Eng.,* vol. 42, no. 1, pp. 68-71, 1995.

[6] A.M. Dale and M.I. Sereno, "Improved Localization of Cortical Activity by Combining EEG and MEG with MRI Cortical Surface Reconstruction: A Linear Approach," *Journal of Cognitive Neuroscience*, vol. 5, no. 2, pp. 162-176, 1993.

[7] J.C. De Munck , B.W. Van Dijk, and H. Spekreijse, "Mathematical Dipoles are Adequate to Describe Realistic Generators of Human Brain Activity," *IEEE Trans on Bio. Eng.,* vol. 35, no. 11, pp. 960-965, 1988.

[8] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Trans on Patt. Anal. and Mach. Int.*, vol. 6, no. 6, pp. 721-741, 1984.

[9] G.H. Golub and C.F. Van Loan, *Matrix Computation*, 2nd Ed., Johns Hopkins University Press, 1989, p. 51 and pp. 250-251.

[10]I.F. Gorodnitsky, J.S. George, and B.D. Rao, "Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm," *EEG and clinical Neurophysiol.*, 95, pp. 231-251, 1995.

[11]I.F. Gorodnitsky, B.D. Rao, and J. George, "Source Localization in Magnetoencephalography using an Iterative Weighted Minimum Norm Algorithm," *Proc. 26th Asilomar Conf.*, pp. 167-171,1992.

[12]M. Hämäläinen, R. Hari, *et al.*, "Magnetoencephalography - theory, instrumentation, and applications to noninvasive studies of the working human brain," *Rev Mod Phys*, vol. 65, no. 2, pp. 413-497, 1993.

[13]P.C. Hansen, *Regularization Tools, A Matlab Package for Analysis and Solution of Discrete Ill-Posed Problems*, available in Postscript from http://www.mathworks.com/, March 1993.

[14]P.C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," *SIAM Review*, vol. 34, pp. 561-580, 1992.

[15]H.P. Hiriyannaiah, G. Bilbro, and W.E. Snyder, "Restoration of piecewise-constant images by mean-field annealing," *Opt. Soc. of Amer. A*, vol. 6, no. 12, pp. 1901-1912, 1989.

[16]B. Jeffs, R.M. Leahy, and M. Singh, "An evaluation of methods for neuromagnetic image reconstruction," *IEEE Trans on Bio. Eng*, vol. 34, pp. 713-723, 1987.

[17]J.C. Mosher, R.M. Leahy, and P.S. Lewis, "Matrix Kernels for MEG and EEG Source Localization Imaging," *Proceedings, ICASSP*, pp. 2943-2946, 1995.

[18]J.C. Mosher, P.S. Lewis, R. Leahy, and M. Singh, "Multiple Dipole Modeling and Localization from Spatio-Temporal MEG Data," *IEEE Trans on Bio. Eng.,* vol. 39, no. 6, pp. 541-557, 1992.

[19]J.C. Mosher, M.E. Spencer, R.M. Leahy, and P.S. Lewis, "Error bounds for EEG and MEG dipole source localization," *EEG and clinical Neurophysiol.*, vol. 86, pp. 303-321, 1993.

[20]Y. Okada, "Neurogenesis of Evoked Magnetic Fields," in: S.J. Williamson, G.L. Romani, L. Kaufman, and I. Midena, *Biomagnetism, an Interdisciplinary Approach*, Plenum Press, N.Y., pp. 399-408, 1983.

[21]R.L. Parker, *Geophysical Inverse Theory*, Princeton University Press, New Jersey, 1994.

[22]R.D. Pascual-Marqui, C.M. Michel, D. Lehmann, "Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain," *Int. Journ. of Psyc.*, vol. 18, pp. 49-65, 1994.

[23]P.E. Roland, *Brain Activation*, Wiley-Liss, 1993.

[24]J. Sarvas, "Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem," *Physics in Medicine and Biology,* vol. 32, pp. 11-22, 1987.

[25]Y.S. Shim and Z.H. Cho, "SVD Pseudoinversion Image Reconstruction," *IEEE Trans. on Acoustics, Speech, and Sig. Proc.*, vol. ASSP-29, no. 4, pp. 904-909, 1981.

[26]J. Tripp, "Physical Concepts and Mathematical Models," in *Biomagnetism: An Interdisciplinary Approach*, Plenum Press, New York, pp. 101-139, 1983.

[27]D.E. Van den Bout and T.K. Miller, "Graph Partitioning Using Annealed Neural Networks," *International Conference on Neural Networks*, vol. 1, pp. 521-528, 1989.

[28]J.Z. Wang, S.J. Williamson, and L. Kaufman, "Magnetic Source Images Determined by a Lead-Field Analysis: The Unique Minimum-Norm Least-Squares Estimation," *IEEE Trans. on Bio. Eng.,* vol. 39, no. 7, pp. 665-675, 1992.