Imaging and Localizing Neural Sources from MEG Data (A beamformer perspective)

John C. Mosher, Ph.D.
Epilepsy Center, Neurological Institute

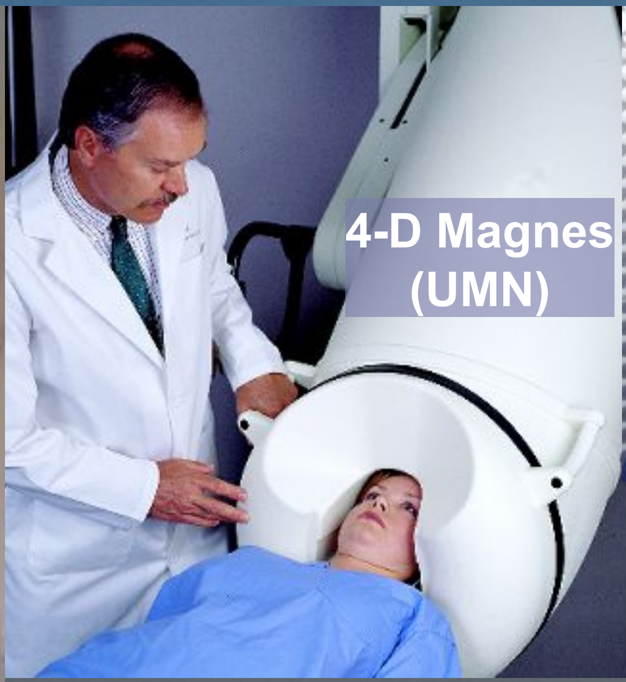
Cleveland Clinic
Cleveland, Ohio USA

Cleveland Clinic

Outline

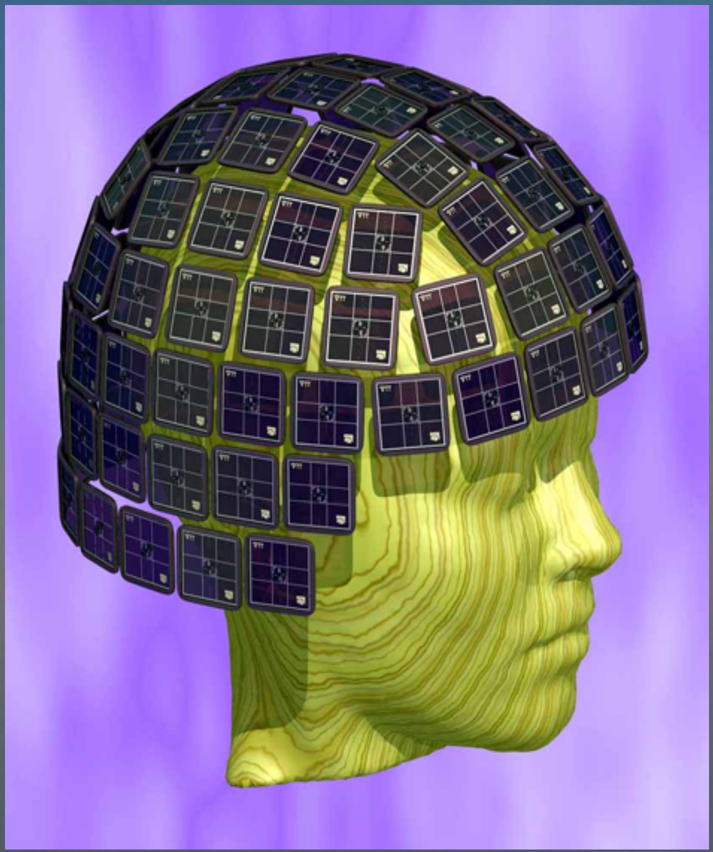
- Suggested in-depth review papers
  - See particularly review papers for the mathematical details
  - Some details in “hidden” slides of electronic copy
- General description of “beamformers”
  - A versatile approach to spatial filtering
- The “minimum norm” approach
  - Transform sensor data in more interpretable “images” of activity
- The recommended workflow

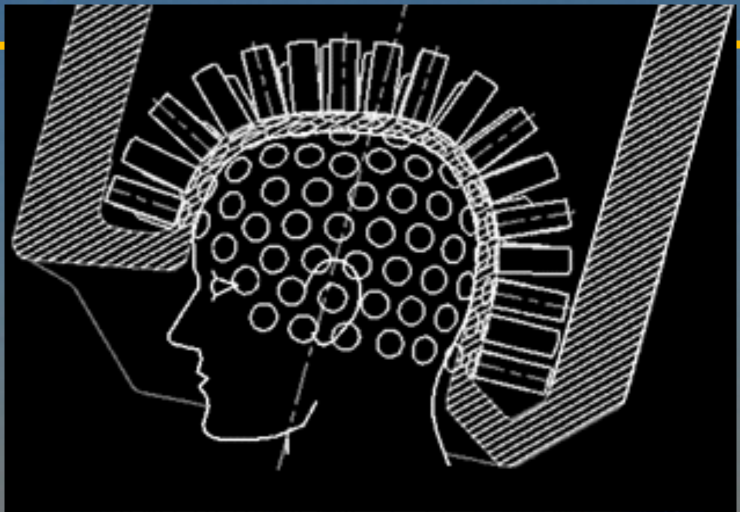
Example of MEG Arrays - MIND Institute Partner Sites: Mass General Hospital, U New Mexico, U Minn., LANL

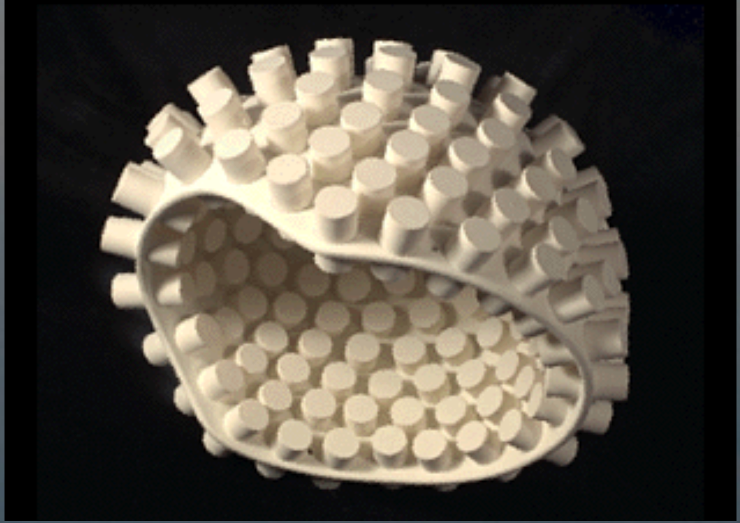


A goal is to demonstrate the removal of the effects of site and instrument from the estimation.

Magnetometer Arrays – (single coil, planar/axial gradiometers)







Generally hundreds of sensors

Biomagnetism Review Paper - 1981

- Williamson and Kaufman 1981 Journal of Magnetism and Magnetic Materials
  - 281 references, 73 pages
  - PDF available online from PUB-MED

Journal of Magnetism and Magnetic Materials 22 (1981) 129–201

BIOMAGNETISM *

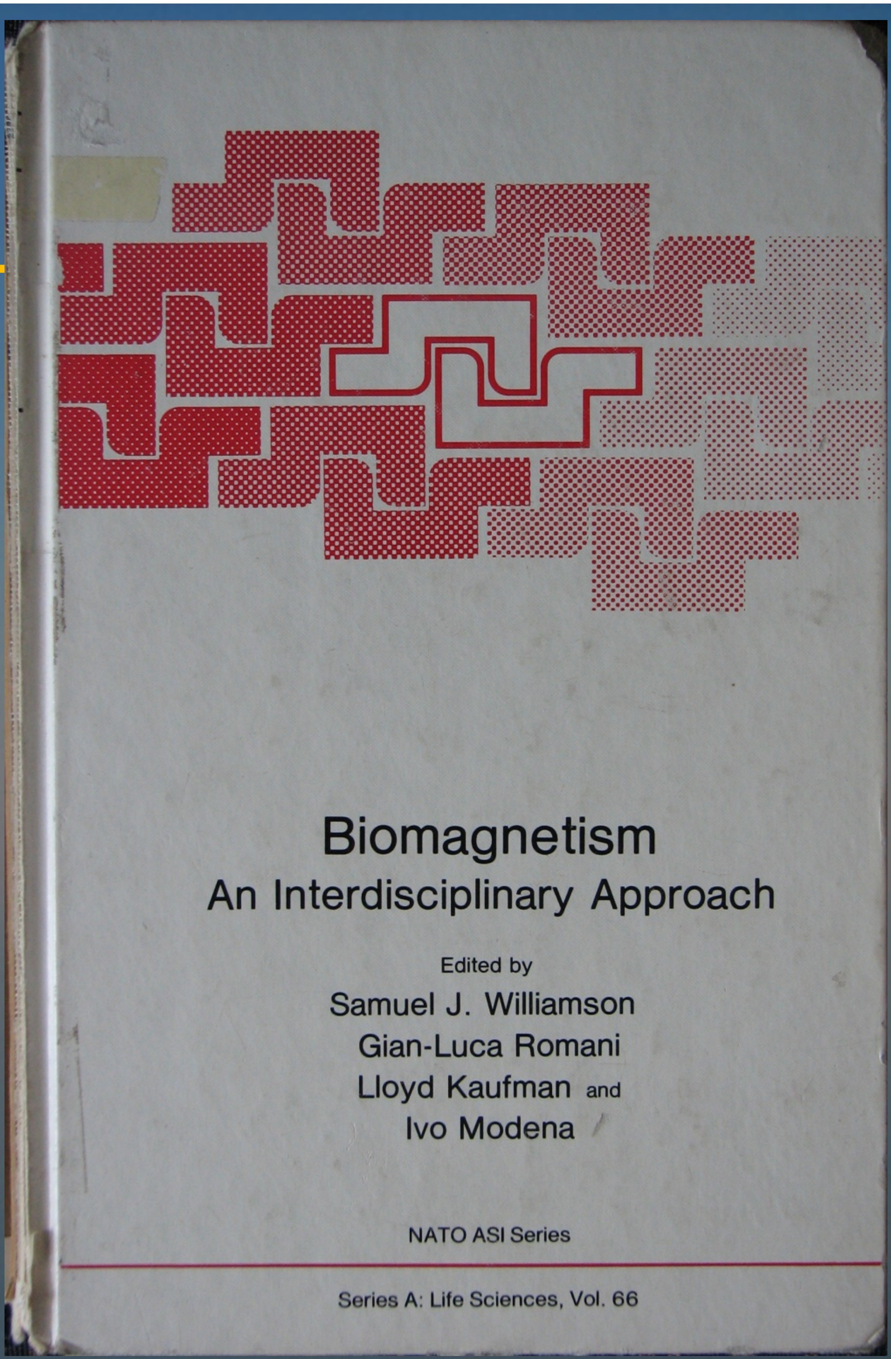
S.J. WILLIAMSON ** and L. KAUFMAN

Neuromagnetism Laboratory, Departments of Physics and Psychology, New York University, New York, NY 10003, USA

Received 12 May 1980; in revised form 18 August 1980

1982 NATO Proceedings

- "Proceedings of a NATO Advanced Study Institute on Biomagnetism" Sep 1-12, 1982, Rome conference
- Williamson, Romani, Kaufman, Modena, Eds.
- Excellent introduction by David Cohen.
- Excellent reference material on source modeling, forward modeling as related to dipoles.
- Amazon.com
  - ISBN: 0-306-41369-8



1993 Review

Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain

Matti Hämäläinen, Riitta Hari, Risto J. Ilmoniemi, Jukka Knuutila, and Olli V. Lounasmaa

Low Temperature Laboratory, Helsinki University of Technology, 02150 Espoo, Finland

- Reviews of Modern Physics, April 1993
  - 85 pages, hundreds of references of the primary contributions, available online as PDF
- Minimum norms, least-squares of multiple dipoles, MUSIC, Sarvas spherical model, boundary element solutions

Tutorial Overview

IEEE Signal Processing Magazine, Nov 2001

Baillet, Mosher, Leahy

See also web site at University of Southern California: neuroimage.usc.edu.

Electromagnetic Brain Mapping

Sylvain Baillet, John C. Mosher, and Richard M. Leahy

The past 15 years have seen tremendous advances in our ability to produce images of human brain function. Applications of functional brain imaging extend from improving our understanding of the basic mechanisms of cognitive processes to better characterization of pathologies that impair normal function. Magnetoencephalography (MEG) and electroencephalography (EEG) (MEG/EEG) localize neural electrical activity using noninvasive measurements of external electromagnetic signals. Among the available functional imaging techniques, MEG and EEG uniquely have temporal resolutions below 100 ms. This temporal precision allows us to explore the timing of basic neural processes at the level of cell assemblies. MEG/EEG source localization draws on a wide range of signal processing techniques including digital filtering, three-dimensional image [redacted]

Software Packages Mentioned in This Talk

- See http://neuroimage.usc.edu
- MRI Processing
  - **BrainSuite** (University of Southern California / University of California Los Angeles)
  - **Freesurfer** (Massachusetts General Hospital)
- MEG Processing
  - **BrainStorm** (University of Southern California)
  - **MNE** (Massachusetts General Hospital)

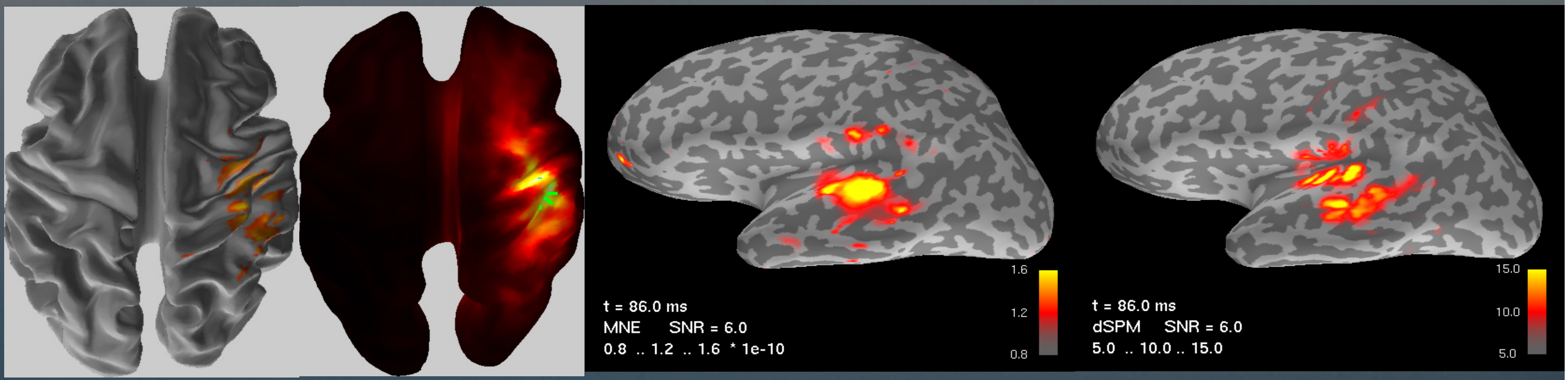
Additional Reading

- "Multiple dipole modeling and localization from spatio-temporal MEG data" IEEE Trans. BME 1992
- "Recursive MUSIC: A Framework for EEG and MEG Source Localization," IEEE Trans BME 1998
- "Electromagnetic Brain Mapping" IEEE SP Magazine 2001
- "Equivalence of linear approaches in bioelectromagnetic inverse solutions" IEEE Workshop in Statistical Signal Processing 2003
- See http://neuroimage.usc.edu for others

How do we interpret these images?

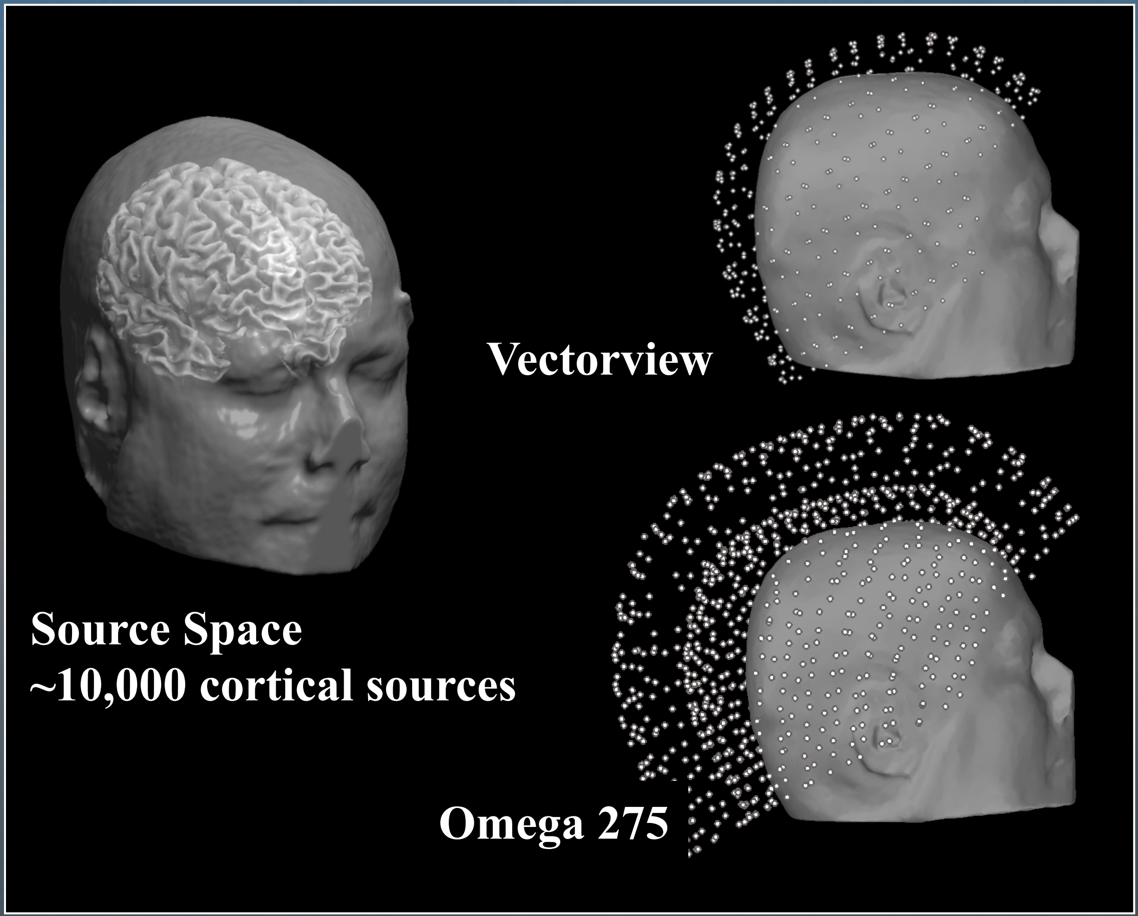
- "Reconstruction," "Significance," "Scan," "Detection"
- Particularly: What does the extent of the "blur" mean?
- Answers:
  - Could be evidence of a distributed source in the region
  - Could be evidence of the location uncertainty of a point source

Min Norm Dipole Scan Min Norm dSPM

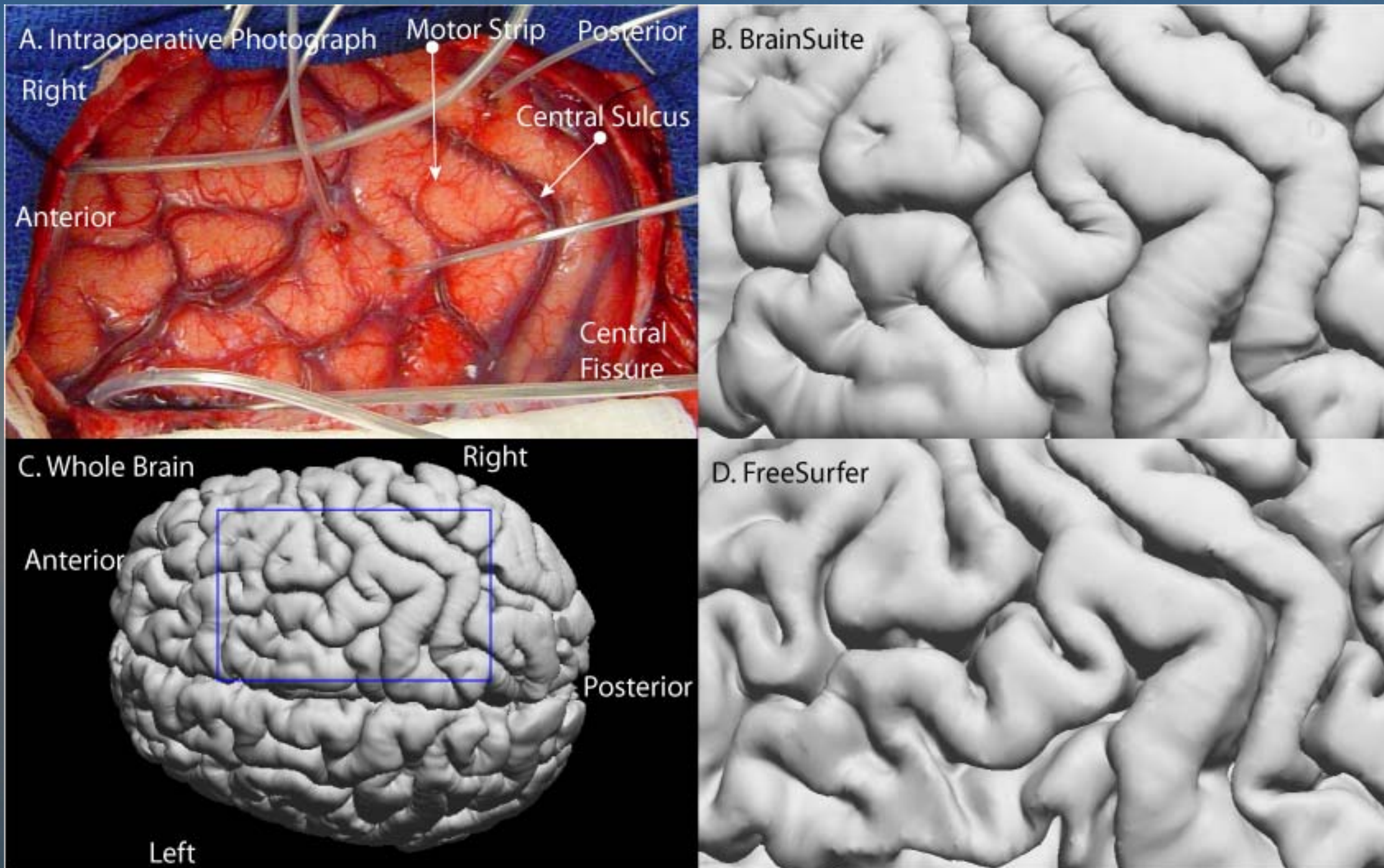


The Acquisition Challenge Today

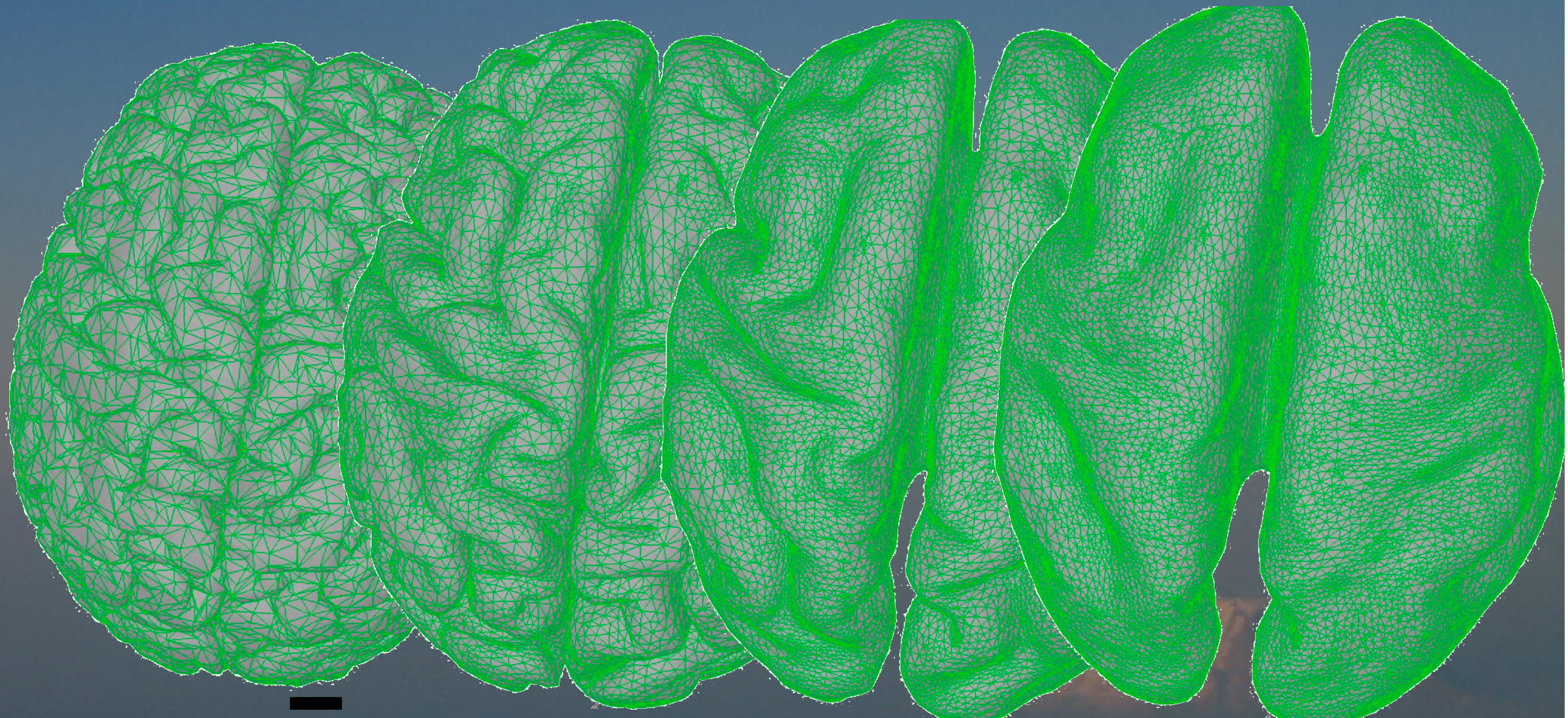
- Hundreds of MEG and EEG sensors simultaneously collecting thousands of samples per second
- Detailed high-resolution MRI routinely available
- Segmentation and surface extraction tools automatically generate 500K tesselated surfaces
  - Reduction to 10K for easier computing



Comparison of Surface Extraction Algorithms



Smoothing of Pial Surface: Visualization into Depths of Sulcal Folds



The Recommended Workflow

- Noise rejection schemes
  - Spatial filters, beamformers
- Imaging estimators
  - Statistical parametric maps
- Adaptive beamformers
  - Statistical parametric maps
- Low order parametric models
  - Scanning for solutions

Progressively harder steps!

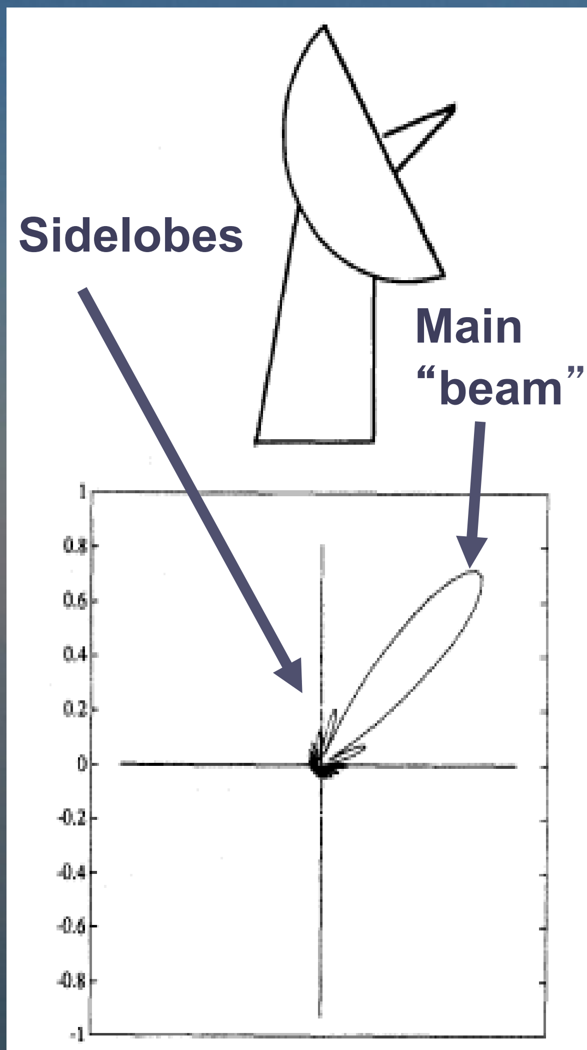
Beamformer Signal Processing Reference

Beamforming: A Versatile Approach to Spatial Filtering

Barry D. Van Veen and Kevin M. Buckley

- IEEE ASSP Magazine 1988
- Algorithmic details and references

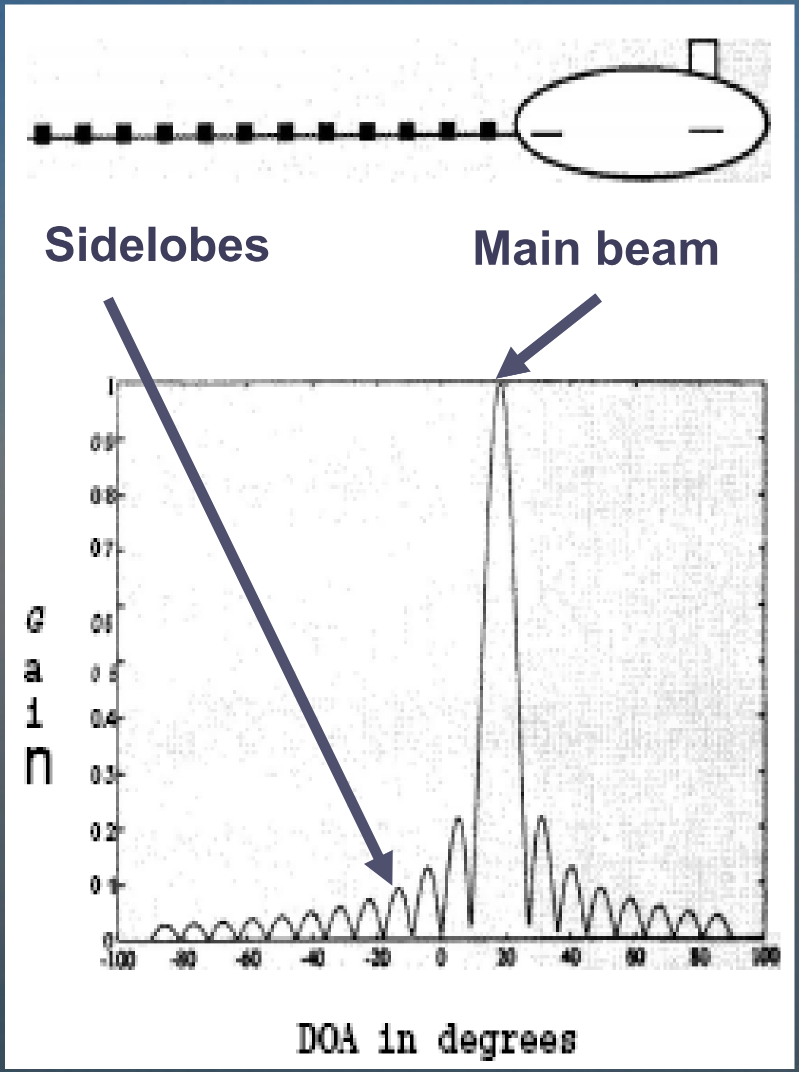
What is a “beamformer”?



van Veen and Buckley 88

- Microwave antenna example
- The size of the dish is the “spatial aperture” or “spatial extent” of the receiver
  - Bigger is better
- The primary reception is along the direction of the main “beam”
- Note the presence of smaller “sidelobes” of the receiver

Beamformer from Arrays of Sensors



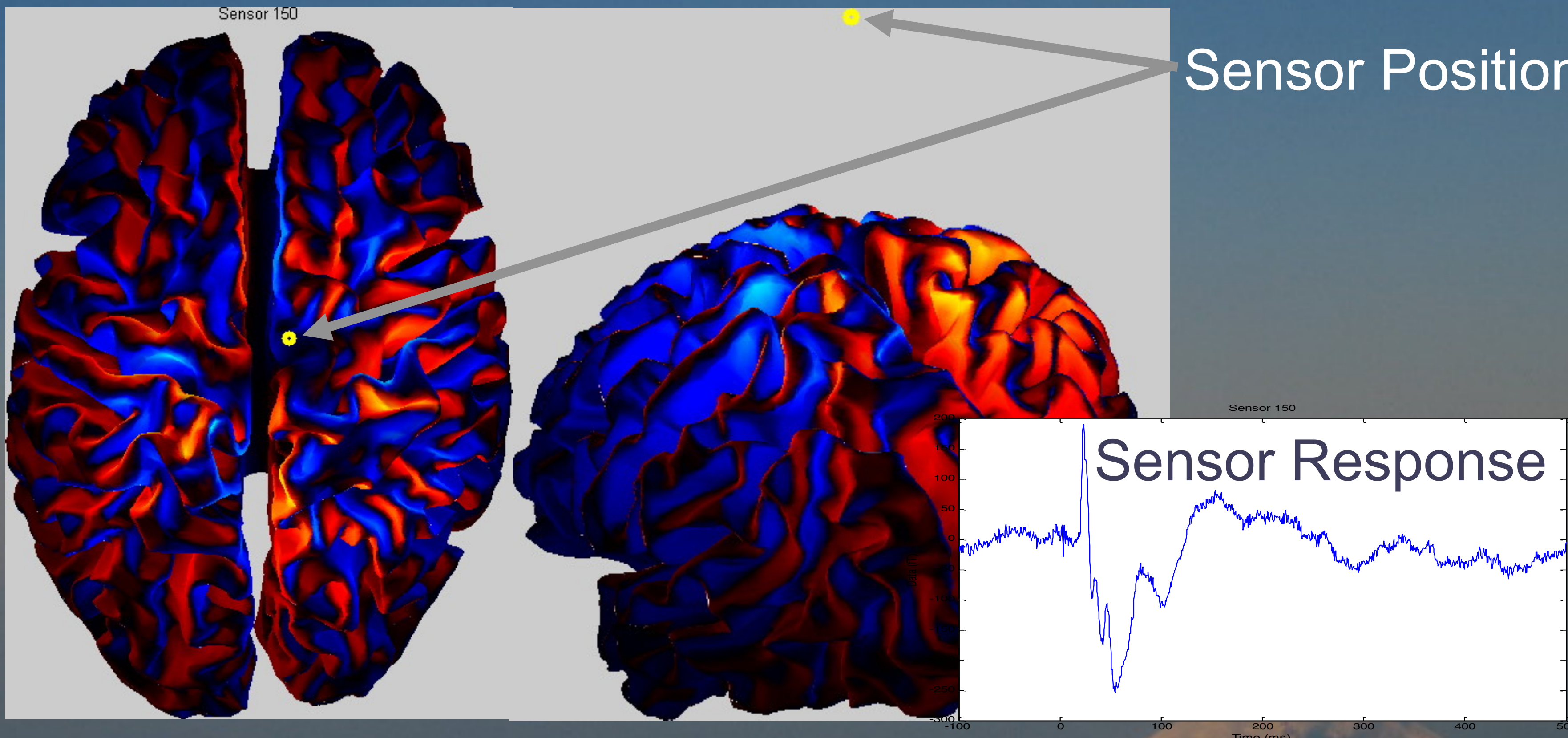
van Veen and Buckley 88

- Sonar towed array example
- Each of the hydrophones is isotropic and identical to the others
- The output of each hydrophone is weighted and delayed relative to the others, then all are summed together into a single output
- The result is high sensitivity to a "direction of arrival" (e.g. 20 degrees here) to the array, and diminished sensitivity to the other directions
- Again, note "sidelobes"

The Linear MEG Beamformer

- An array of sensors
  - 151 or 275 1st-order MEG gradiometers (VSM Medtech Omega)
  - 204 1st-order planar + 102 magnetometers (Elekta Neuromag)
- The measurement vector "**d**(t)"
  - Arrange the measurements at a single time instance
- A set of weights "**w**"
  - To be derived, as many weights as there are sensors
- The beamformer output: $x(t) = \mathbf{w}^t * \mathbf{d}(t)$
  - A linear combination of the data into a scalar output
- Goal: Point the beam at a tiny spot on the cortex

The MEG Sensor as a Beamformer



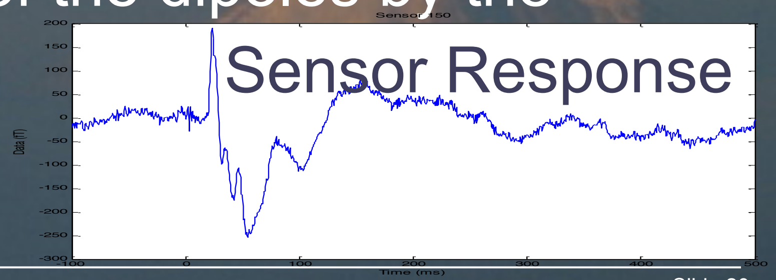
- 10,000 cortical sources, each individually generates an external magnetic field
- The MEG sensor sees the linear combination of ALL of these sources
  - Red is positive weight, blue is negative weight

The Simple Math of the Linear Beamformer

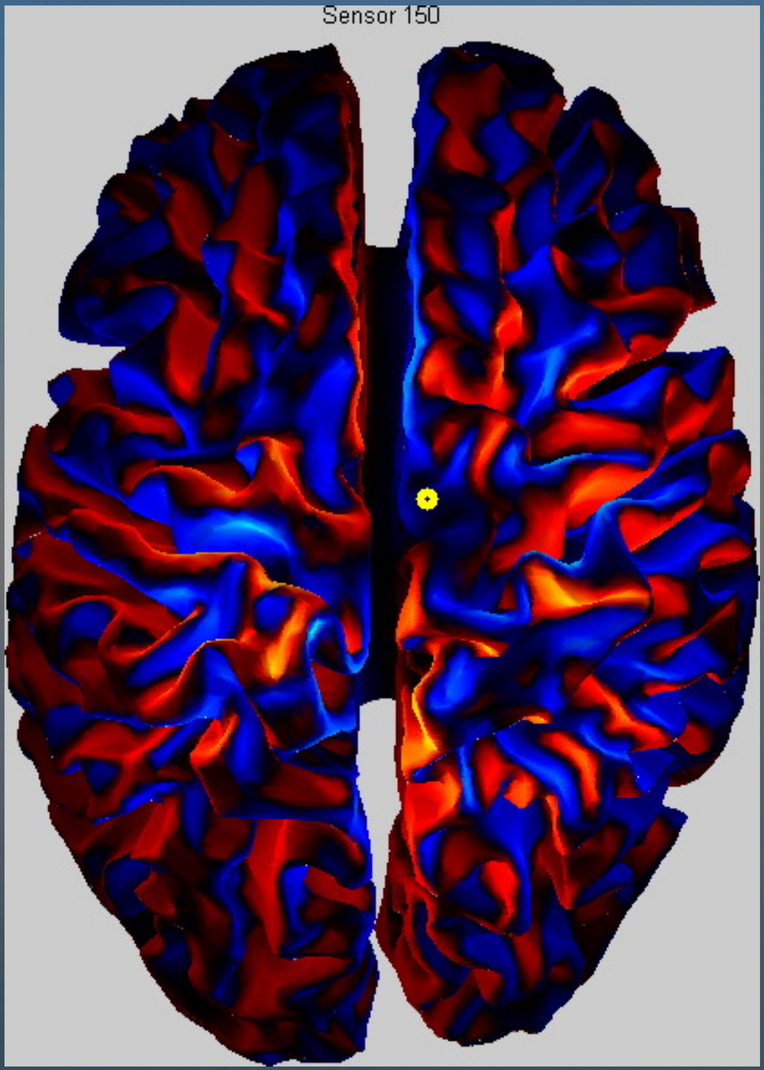
- We have 10,000 cortical dipoles
  - Make a vector "j(t)" 10,000 x 1, as a function of time
- We know the forward model for each dipole to the sensor point
  - Make a vector "l" 10,000 x 1, aka the "lead field model"
- The sensor response, by electromagnetic superposition and quasistatics, is simply

$$d(t) = l^t * j(t)$$

- In other words, we weight each of the dipoles by the correct weight, then sum

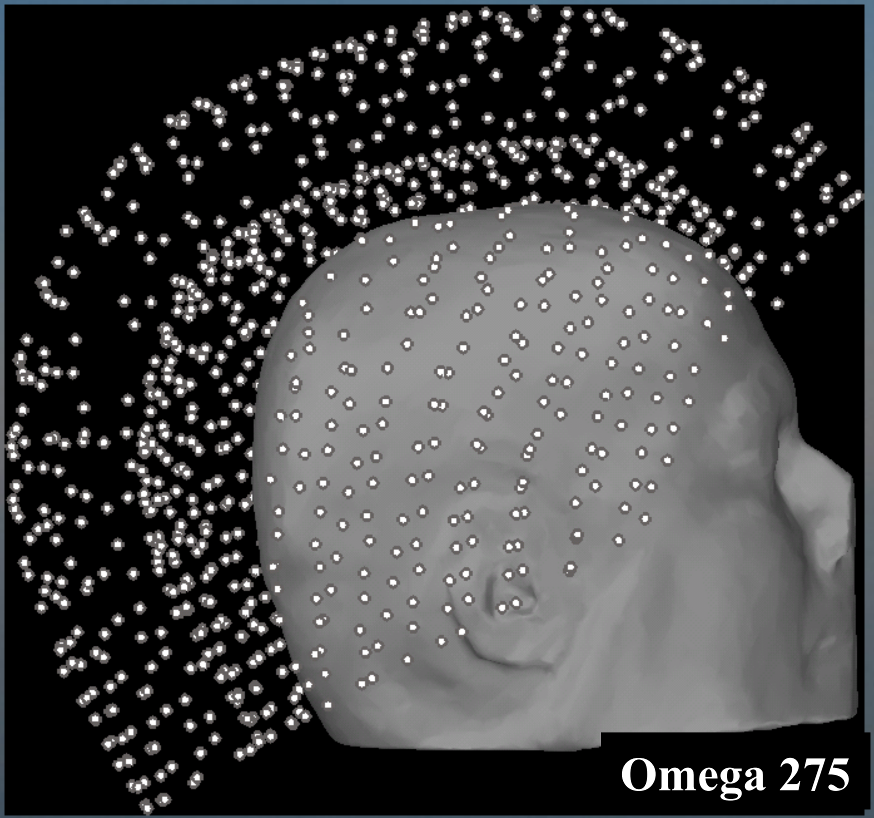


Good News, Bad News



- Good: If activity in the cortex happens anywhere in the colored region, we see it in the sensor
  - "Synoptic", i.e. global view
- Bad: If activity in the cortex happens anywhere in the colored region, we see it in the sensor
  - Ambiguous
- Goal: We would like to know WHERE the activity occurred

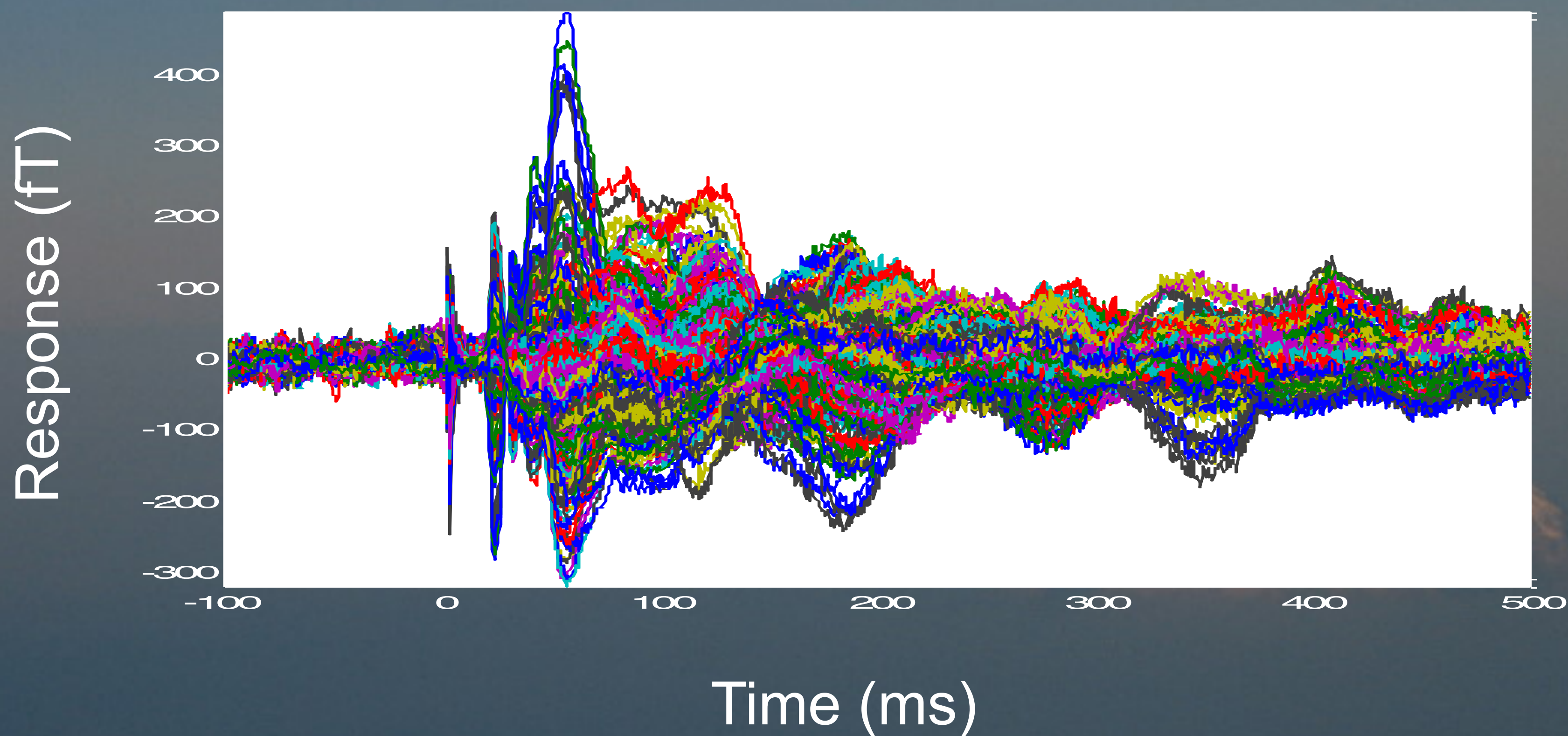
The Dense Sensor Array



- Modern EEG and MEG sensor arrays now comprise 100's of locations
- Each sensor has its own unique "lead field" that relates how each of the 10,000 dipoles generates a field at that sensor
- 1st-order gradiometers (shown here) have a second coil further from the head to reduce environmental noise
  - It's effect is automatically built into the lead field model

The Data "Butterfly" Plot

- Overlay all 275 sensor outputs as a function of time
- The "spatio-temporal" data matrix, here 275 x 1,441 samples
  - (MEG response to left median nerve stimulation, ~100 averages)



The Lead Field Matrix

- For each of the 275 sensors, calculate the lead field for the 10,000 dipoles
  - The "forward problem", a separate topic. Use spheres, boundary elements, finite elements, etc, include gradiometer effects.
- Put all of these lead field vectors into a matrix
  - Rows are the **sampled lead field**, columns are the **sampled dipolar field**

$$\mathbf{d}(t) = \begin{bmatrix} \mathbf{l}_1^t \\ \cdots \\ \mathbf{l}_{275}^t \end{bmatrix} \mathbf{j}(t) = \left[\mathbf{a}_1, \ldots, \mathbf{a}_{10,000}\right] \mathbf{j}(t) = \mathbf{L}\mathbf{j}(t)$$

- The 275 sensor outputs d(t) can now be related to the 10,000 dipoles j(t) by a 275 x 10,000 matrix, the "lead field matrix" L

The Matrix Model

- Spatio-temporal Data Matrix **"D"** (275 x 1,441)
- The Lead Field Matrix **"L"** (275 x 10,000)
  - the forward model relating source to sensor
- The Source Matrix **"J"** (10,000 x 1,441)
- The Noise Matrix **"V"** (275 x 1,441)
  - "other stuff"
- By electromagnetic superposition and noise assumption, the data are linear in the sources and additive in the noise

$$\mathbf{D} = \mathbf{L}\mathbf{J} + \mathbf{V} \qquad \mathbf{d}(t) = \mathbf{L}\mathbf{j}(t) + \mathbf{v}(t)$$

The Ideal (Impossible) Estimator

- Given the 275 channels of data, reconstruct exactly the 10,000 dipole signals

$$\hat{\mathbf{j}}(t) = f(\mathbf{d}(t)) = f(\mathbf{L}\mathbf{j}(t))$$

- Generally impossible, not enough sensors
  - Indeed, not enough information, so more sensors won’t help

- Primary physical limitations:
  Distance to the quasistatic source,
  - limited by skull thickness, scalp, CSF, Dewar, helmet shape, etc.

The Further Complication – “Noise”

- Environmental Noise – cars, elevators, power lines
- Sensor Noise – thermal, electronics, vibrations
- Biologic Noise – heartbeat, eyeblink, respiration
- “Brain” Noise – brain activity “not of interest”
  - “One researcher’s signal is another researcher’s noise.”
- For simplicity, we generally assume noise is additive:

$$\mathbf{d}(t) = \mathbf{L}\mathbf{j}(t) + \mathbf{v}(t)$$

Common Statistical Assumptions

- We generally assume some convenient 2nd order statistics about these vectors
  - Also, for convenience, we assume zero-mean
- The "noise covariance" is the expected cross-correlation of the noise vector
  - Since noise is "random," we can only rely on its statistics
- We assume some sort of source covariance among the dipoles
  - **Key Question**: where does this assumption come from?
- Finally, we assume that the noise is independent of the dipoles, so that the data covariance has a simple relationship

$$\mathbf{C_j} = E\{\mathbf{j}(t)\mathbf{j}(t)^t\} \qquad \mathbf{C_v} = E\{\mathbf{v}(t)\mathbf{v}(t)^t\}$$

$$\boxed{\mathbf{C_d} = E\{\mathbf{d}(t)\mathbf{d}(t)^t\} = \mathbf{L}\mathbf{C_j}\mathbf{L}^t + \mathbf{C_v}}$$

The Simple Min Norm Estimator

- Because our data model in linear in the lead-field, to estimate **j**(t), simply “pseudo-invert” L
- The 275 channel measurements are simultaneously converted into 10,000 separate little beamformers, each representing one dipole

(**BUT**, this is a generally terrible estimator, since noise is mishandled)

$$\mathbf{d}(t) = \mathbf{L}\mathbf{j}(t) + \mathbf{v}(t)$$

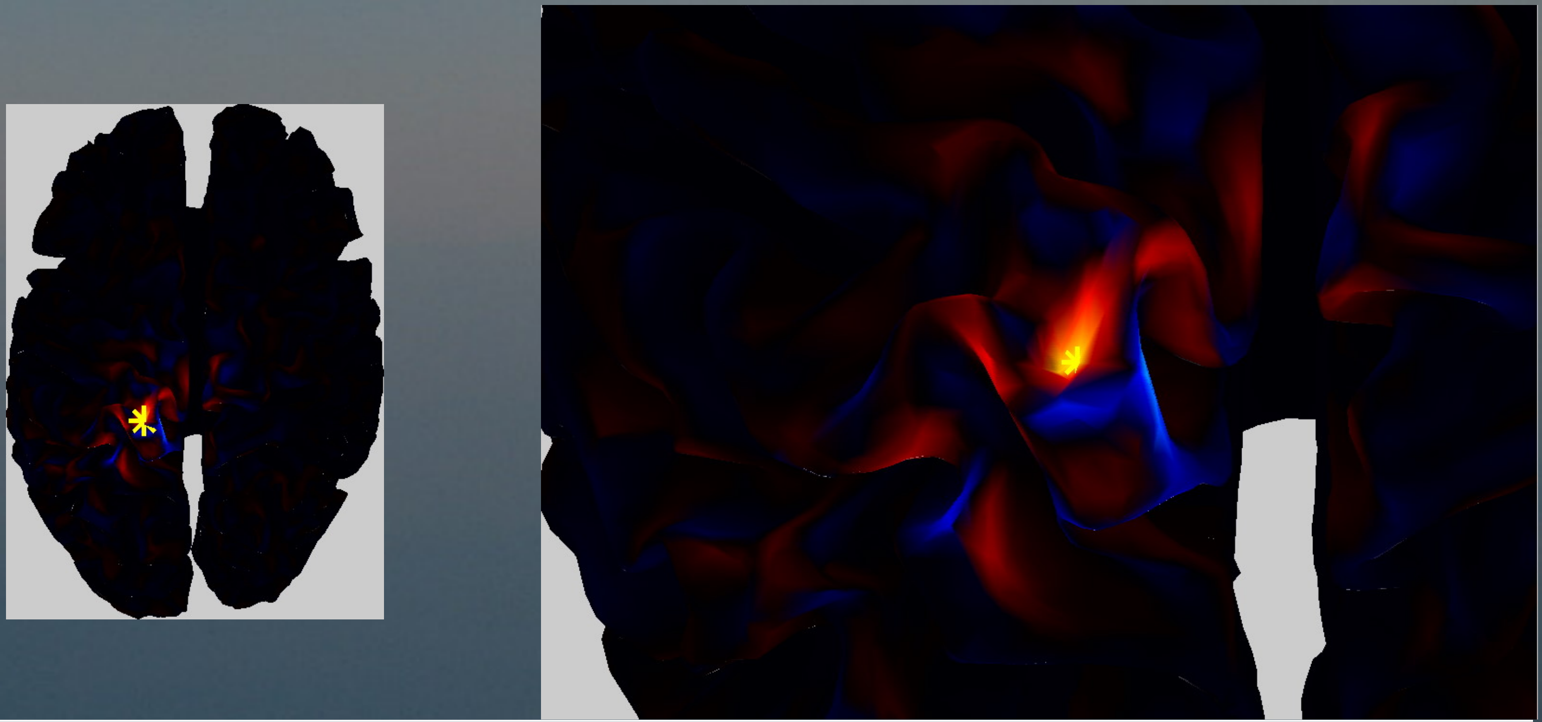
$$\hat{\mathbf{j}}(t) = \mathbf{L}^{\dagger}\mathbf{d}(t) = \mathbf{L}^{\dagger}\mathbf{L}\mathbf{j}(t) + \mathbf{L}^{\dagger}\mathbf{v}(t)$$

$$\mathbf{L}^{\dagger} = \mathbf{L}^{\mathrm{t}}(\mathbf{L}\mathbf{L}^{\mathrm{t}})^{-1}$$

The Min Norm as a Static Beamformer

- Look at the "*i*th" estimated dipole, find the weight vector for just it.
- The weights tell us how the original 10,000 dipoles were linearly combined to make the output seen at the *i*th dipole in the estimate

$$\hat{j}_i(t) = \mathbf{a}_i^t(\mathbf{L}\mathbf{L}^t)^{-1}\mathbf{L}\mathbf{j}(t) = \mathbf{w}_i^t\mathbf{j}(t)$$



What’s Wrong with That?

- The min norm beamformer certainly looks good, relatively focal around the true location
- BUT, weights are also applied to the NOISE:

$$\hat{j}_i(t) = \mathbf{w}_i^t \mathbf{j}(t) + \mathbf{a}_i^t (\mathbf{L}\mathbf{L}^{\mathrm{t}})^{-1} \mathbf{v}(t)$$

- **The Short Answer** on how to fix: Include the noise and source statistics (see “Equivalence Paper” by Mosher, ref’d in hidden notes)

$$\hat{\mathbf{j}}(t) = \mathbf{L}^t (\mathbf{L}\mathbf{L}^t)^{-1} \mathbf{d}(t)$$

becomes

$$\hat{\mathbf{j}}(t) = \mathbf{C}_{\mathrm{j}} \mathbf{L}^t (\mathbf{L}\mathbf{C}_{\mathrm{j}}\mathbf{L}^t + \mathbf{C}_{\mathrm{v}})^{-1} \mathbf{d}(t)$$

Statistically Optimum Beamformer

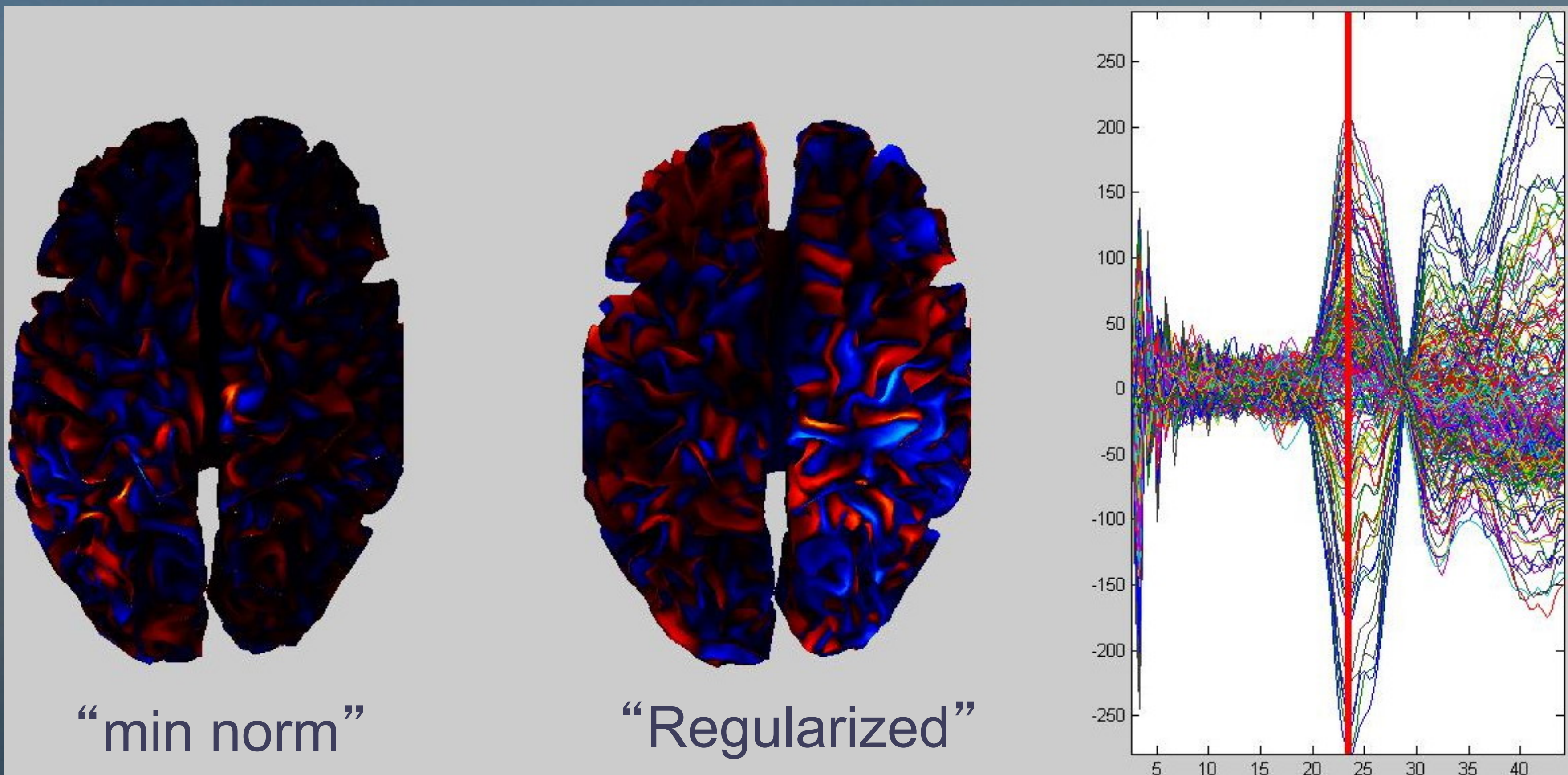
- Our goal is to minimize the means squared error in our estimate, given just the second order statistics, subject to the constraint we want a linear solution:
  - Minimize for $\hat{\mathbf{j}}(t)$ $\quad E\{(\mathbf{j}(t)-\hat{\mathbf{j}}(t))^T(\mathbf{j}(t)-\hat{\mathbf{j}}(t))\}$
- The Linear Minimum Mean-Square (LMMS) solution is

$$\hat{\mathbf{j}}(t)=\mathbf{C}_{\mathrm{j}}\mathbf{L}^t(\mathbf{L}\mathbf{C}_{\mathrm{j}}\mathbf{L}^t+\mathbf{C}_{\mathrm{v}})^{-1}\mathbf{d}(t)$$

- If variables are multivariate Gaussian: LMMS = MAP

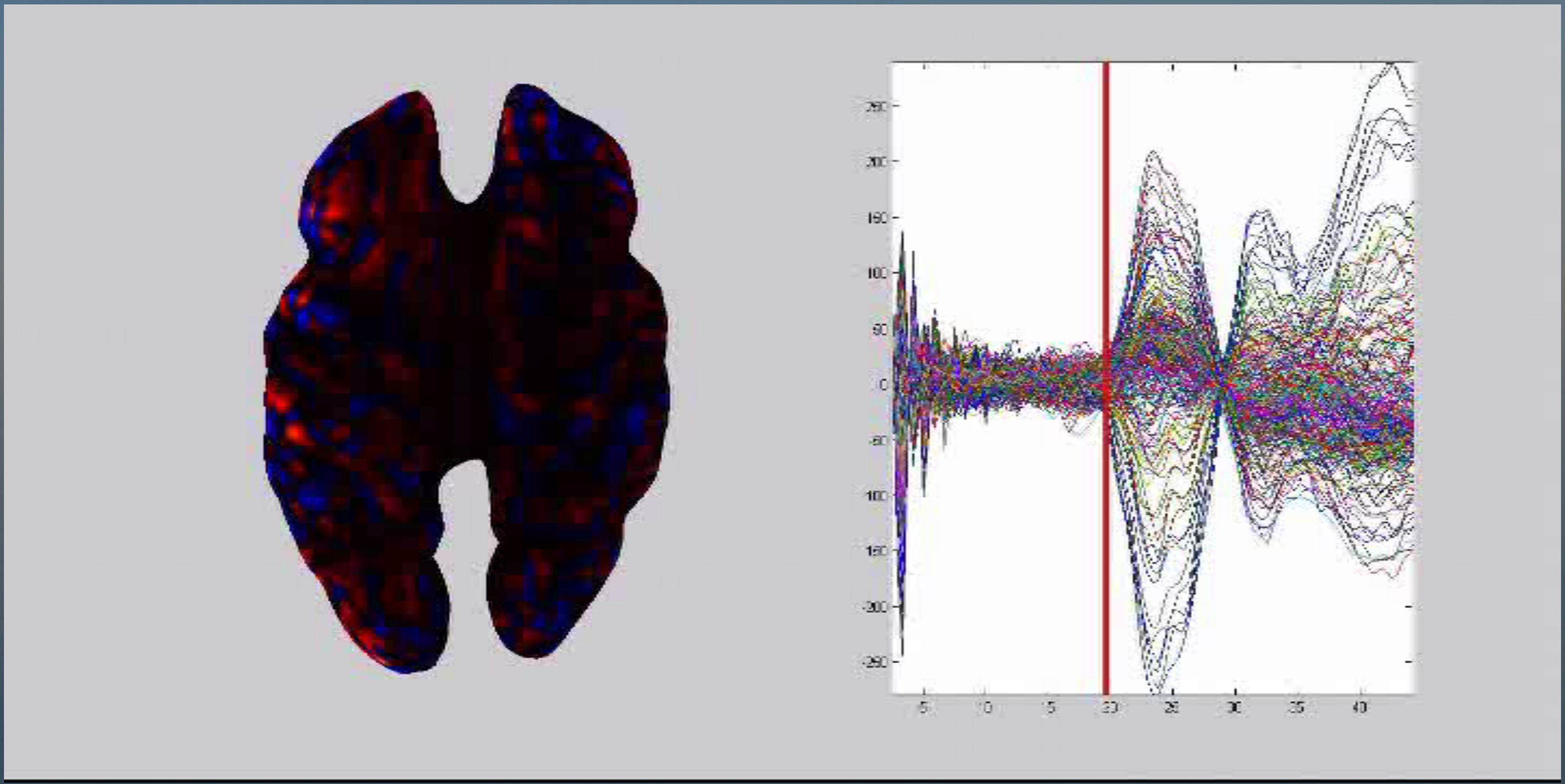
AKA: Weighted Regularized Minimum Norm

- The noise stabilization helps suppress noise boosting.
- Example: Left median nerve stimulation, N20 response



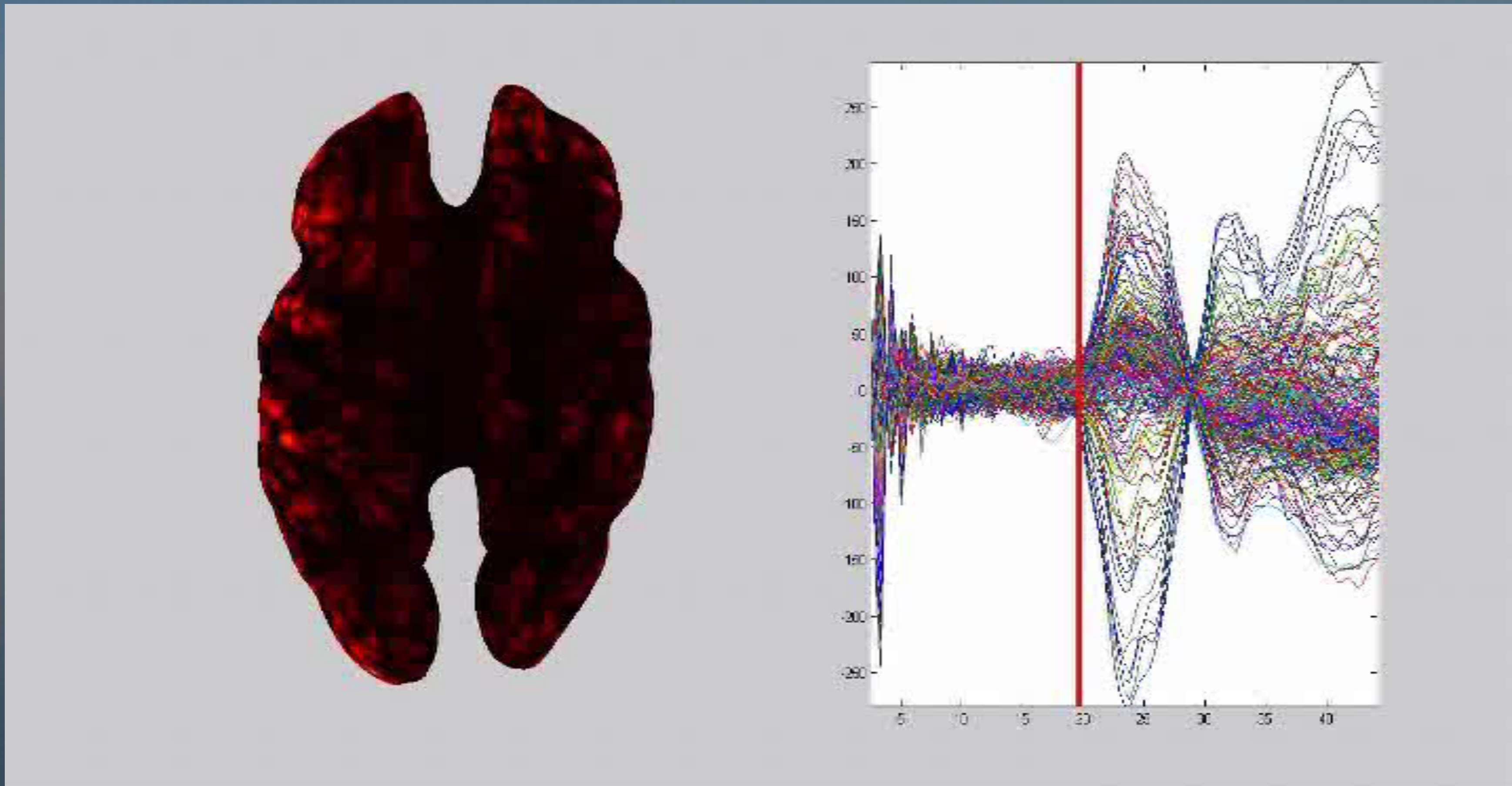
The Min Norm "Movie" - 1

- View a portion of the estimation as a sequence of images
  - But what about those annoying color changes?



The Min Norm "Movie" - 2

- Use only the absolute values, ignore sign
  - But what about those small values?

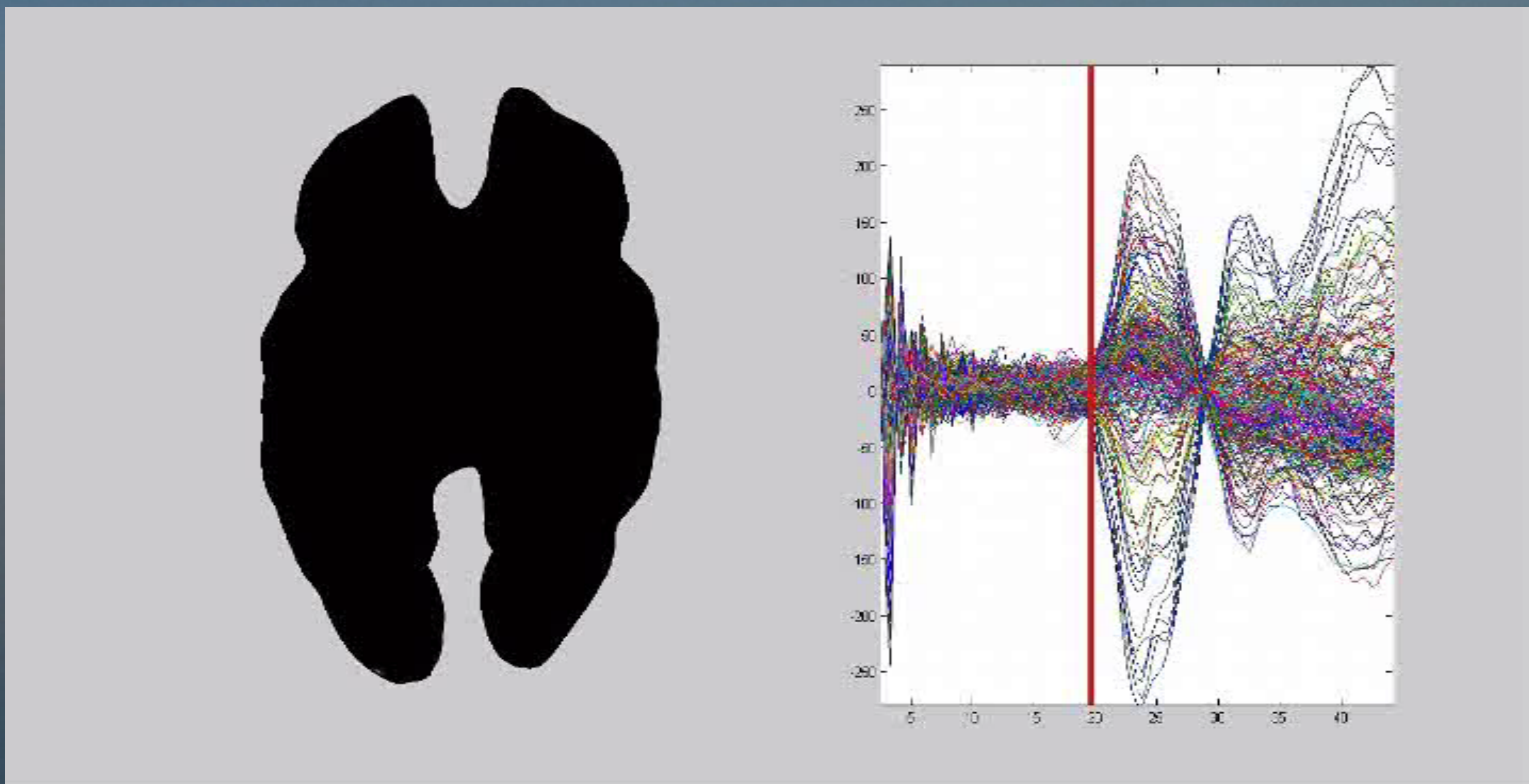


Looking for Peaks in the Min Norm

- Simply slide the colorbar level to truncate the small values

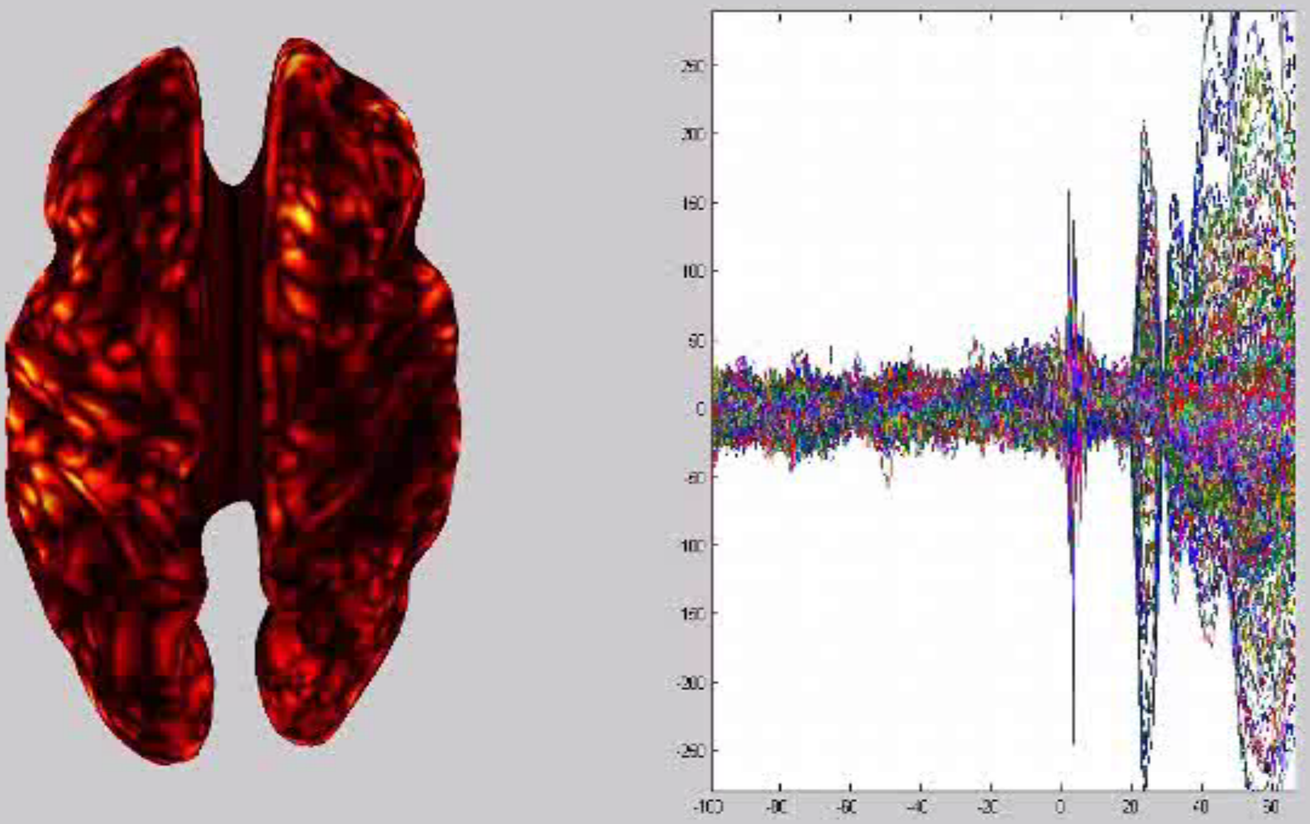
The Min Norm “Movie” - 3

- Truncate smaller values in the presentation
  - But the region is fractured across several sulci



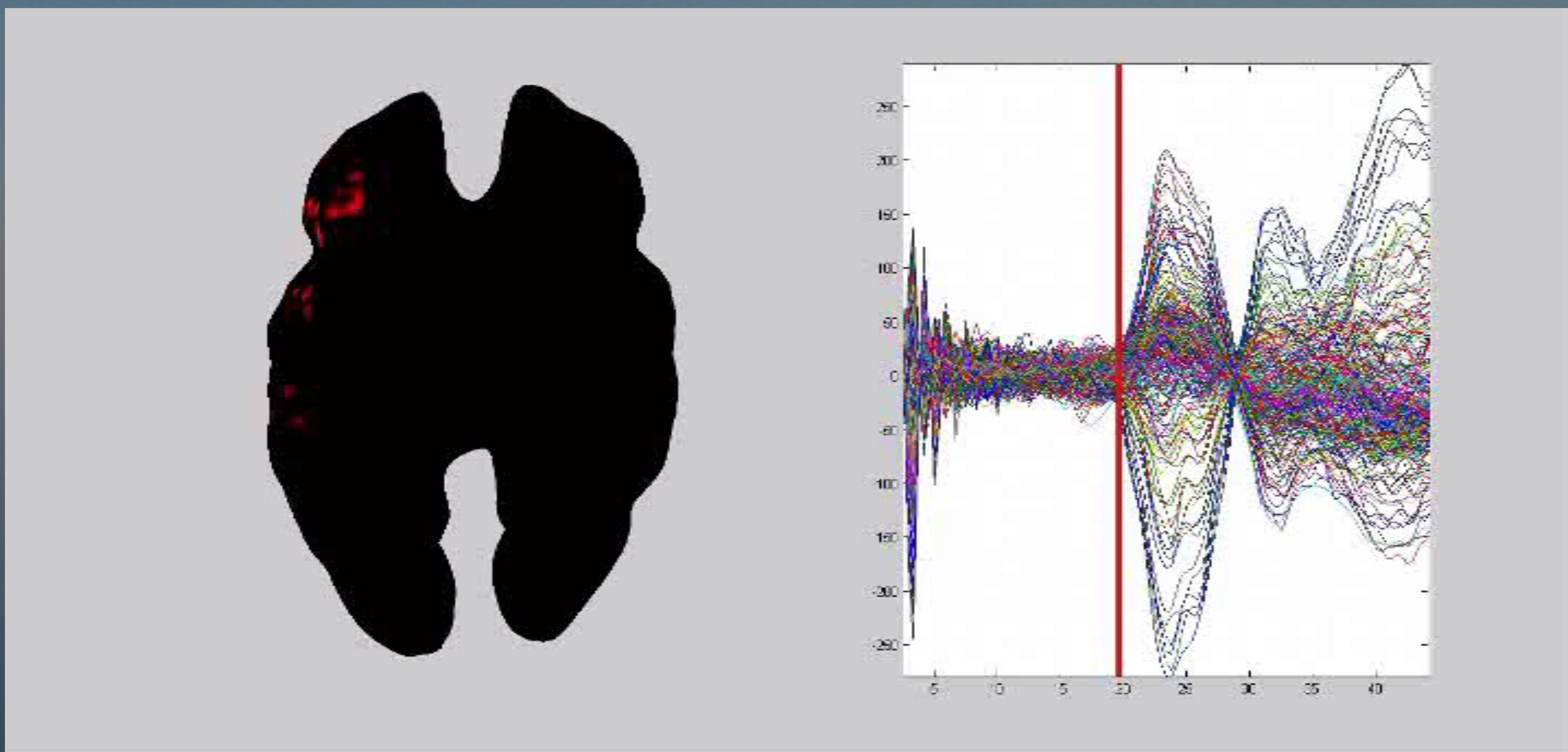
White Noise Gain Power

- Run the prestim or other “noise-only” regions of data through the exact same estimator



The Min Norm "Movie" - 3

- Highlight the "Significant" portions
  - Use the prestim noise to rebalance all peaks ("Z-Score") by their "white noise gain power",
  - "dSPM", "sLORETA" use somewhat different rebalancers



Specific White-noise Gain Power Calculation

- Recall that the ith pixel in the image is simply a linear weight (beamformer) of the data,

$$\hat{\mathbf{j}}(t) = \mathbf{C}_{\mathbf{j}}\mathbf{L}^t(\mathbf{L}\mathbf{C}_{\mathbf{j}}\mathbf{L}^t + \mathbf{C}_{\mathbf{v}})^{-1}\mathbf{d}(t)$$

$$\hat{j}_i(t) == \mathbf{w}_i^t\mathbf{d}(t)$$

- Consider if this *i*th source saw only noise, not data

$$\sigma_{iv}^2(t) == E\{\mathbf{w}_i^t\mathbf{v}(t)\mathbf{v}^t(t)\mathbf{w}_i\} == \mathbf{w}_i^t\mathbf{C}_{\mathbf{v}}\mathbf{w}_i$$

- Weight (score) by same function with noise instead

$$\hat{z}_i(t) == \mathbf{w}_i^t\mathbf{d}(t) / \sqrt{\mathbf{w}_i^t\mathbf{C}_{\mathbf{v}}\mathbf{w}_i}$$

Independent Sources Assumption

- Common assumption in community is that $\mathbf{C_j}$ is diagonal
  - Easier storage, simpler prior that sources are independent (cf. Box in 2003 paper)

$$\hat{j}_i == \mathbf{w}_i^t \mathbf{d} = \sigma_i^2 \mathbf{a}_i^t \mathbf{C}_\mathbf{d}^{-1} \mathbf{d}$$

- Average z^2 is therefore

$$E\{\hat{z}_i^2(t)\} == \frac{\mathbf{a}_i^t \mathbf{C}_\mathbf{d}^{-1} \mathbf{a}_i}{\mathbf{a}_i^t \mathbf{C}_\mathbf{d}^{-1} \mathbf{C}_\mathbf{v} \mathbf{C}_\mathbf{d}^{-1} \mathbf{a}_i}$$

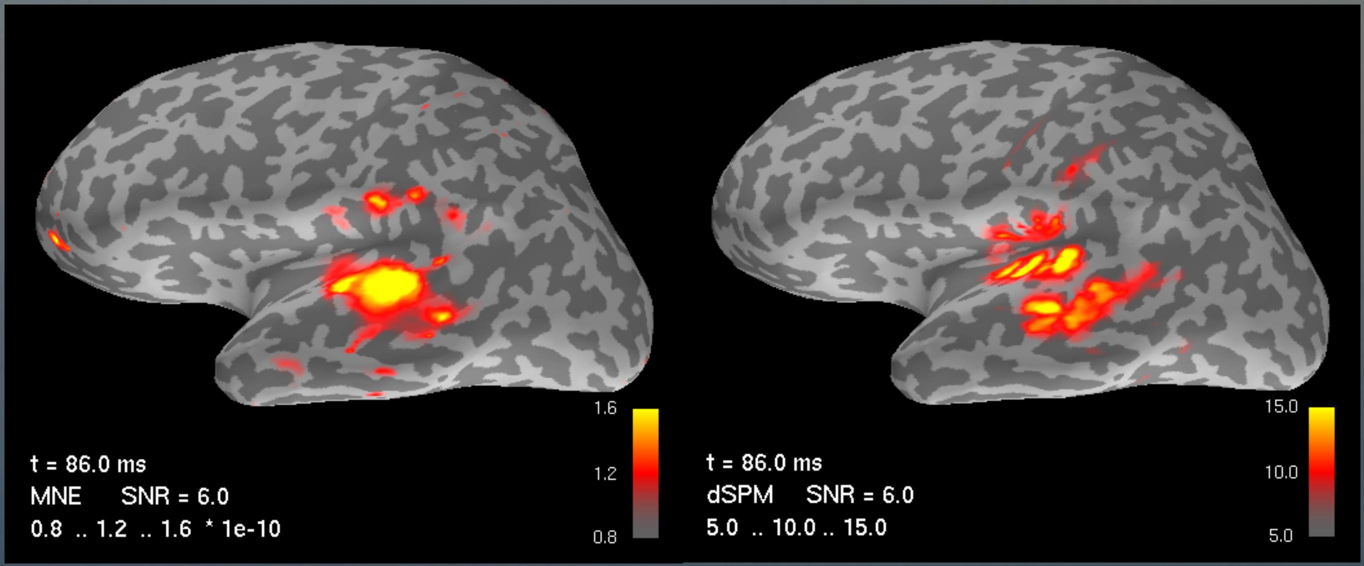
- NOTE: Diagonal covariance term gone!

Statistical Parametric Maps

- Determine what sources in the image are "significant."
- First step towards identifying "regions of interest"
- Z-score, dSPM and sLORETA can produce very similar results with real data

- Auditory MEG data
  Source locations constrained to the cortex.
  -No orientation constraint

"Reconstruction Image" "Significance Map"



Freesurfer / MNE Processing

Different Linear Imaging Estimator Viewpoints

- **Also known as:**
  - Minimum norm least squares with weighting and regularization
  - Generalized least-squares with priors
  - Maximum a Posteriori (MAP) or Bayesian estimation with Gaussian priors
  - Linear minimum mean square
  - LORETA, LAURA, other variations
- cf. Mosher, Baillet, Leahy, "Equivalence of Linear Approaches in Bioelectromagnetic Inverse Solutions", *2003 IEEE Workshop on Statistical Signal Processing*, St. Louis, Missouri, Sep 28 - Oct 1, 2003.

Other Imaging Estimators

- Can impose more difficult constraints/priors yielding generally nonlinear estimators
- Iterative source covariance adjustments
  - MFT (one), FOCUSS (multi)
- Exponential prior instead of Gaussian (“L1” estimation)
  - MCE, VESTAL
- Other Non-Gaussian priors
  - Phillips 1997, Baillet 1997, Schmidt 1999

Strong Prior Example Viewpoint

- **Primary issue among all of these imaging estimators**: The strong prior required to estimate ~10,000 parameters from 200 measurements
- From a least-squares perspective, we have ~200 measurements, but 10,000 unknowns.
  - Problem is unsolvable for unique solution
- Solution: Add 10,000 virtual measurements!
  - Problem now solvable from a least-squares point of view
  - 10,200 measurements and 10,000 unknowns
- **But only 2% of our data came from the laboratory! The rest came from our theory**.
  - Careful what you wish for: Imaging vs. Imagining.

So Why “Image” if the Prior is so Strong!

- EEG Example:
  Ten-Twenty array is a standardized adaptable array
  - 10%, 20% of distance of subjects aricular and inion/nasion lines
  - Common ground
  - “Cz” “Fp” etc universally understood among all researchers
- Today’s arrays are denser, rigid (MEG), more differential pairs (EEG), gradiometric, with adaptive noise schemes
- Imaging allows us to transform sensor data into a more interpretable format in order to “peak under the data and gain insight into its processes”
  - Tukey 1976 Exploratory Data Analysis
- Reversible Linear Transformations preserve information

The Recommended Workflow

- Noise rejection schemes
  - Spatial filters, beamformers
- Imaging estimators
  - Statistical parametric maps
- Adaptive beamformers
  - Statistical parametric maps
- Low order parametric models
  - Scanning for solutions

Progressively harder steps!

ACKNOWLEDGEMENTS

This material is based upon work supported by the Epilepsy Center of the Cleveland Clinic, and by the National Institutes of Health under grant R01-EB009048, DP2-OD006469, and R01-EB002010.

Cleveland Clinic