MEG Source Localization Using an MLP With Distributed Output Representation

Sung Chan Jun, Barak A. Pearlmutter, Guido Nolte

CBLL Meeting – October 5, 2005

Source localization

- Definition:
  Identification of brain regions that emit detectable electromagnetic signals
- Assumption of dipole
- Application: detection of regions of the brain that cause epilepsy (for further neurosurgery)

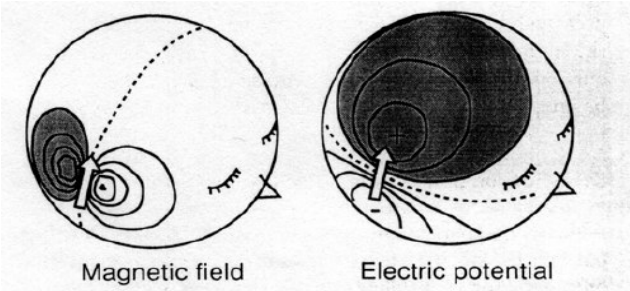


FIG. 5. Schematic illustration of idealized magnetic-field and electric-potential patterns produced by a tangential dipole (white arrow). The head was approximated with a four-compartment sphere consisting of the brain, the cerebrospinal fluid, the skull, and the scalp. From noninvasive measurements of the MEG or EEG field distributions, the active area in the brain can be determined by a least-squares fit to the data.

- Source: Hamalainen M. et al., “Reviews of Modern Physics”, 1993

Magnetoencephalography (MEG)

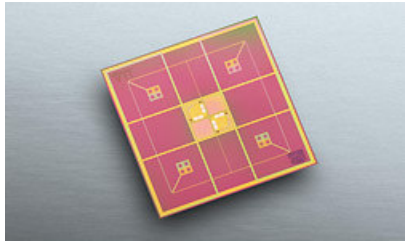
- Electrical activity produces magnetic fields
  Field 10^{-15} T (earth 0.5×10^{-4} T)
- Field recorded outside skull
  (Skull and tissue do not affect signal)
- SQUID = Superconductor Quantum Device, immersed in liquid He
- Eliminates impedance
- Allows for high sensitivity
- Magnetic shield room needed
- 4-D Neuroimag-122 as:
  122 pairs of sensors measuring signal over time (time resolution 1ms)

- Source: www.elekta.com

How MEG works?

- Electromagnetic induction: magnetic field causes electric current in a moving coil
- Property used by magnetometers
- Gradiometers:
  Measure value of magnetic field between two different points, i.e. its gradient

- Sources:
  - Gravity/Magnetic Glossary, http://www.igcworld.com/gm_glos.html,
  - Hamalainen M. et al., “122-Channel SQUID Instrument for Investigating Magnetic Signals from the Human Brain”, Physica Scripta, 1993



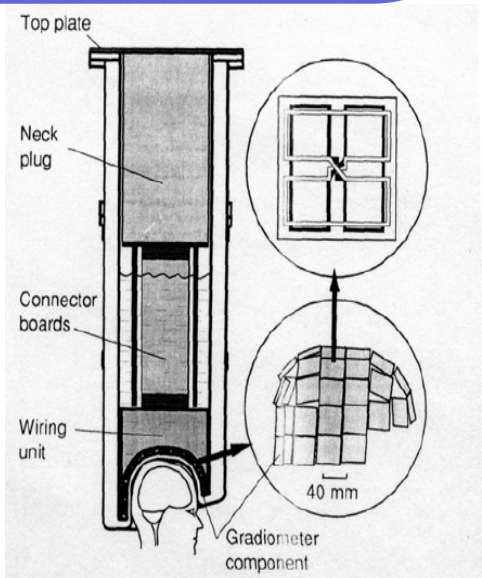
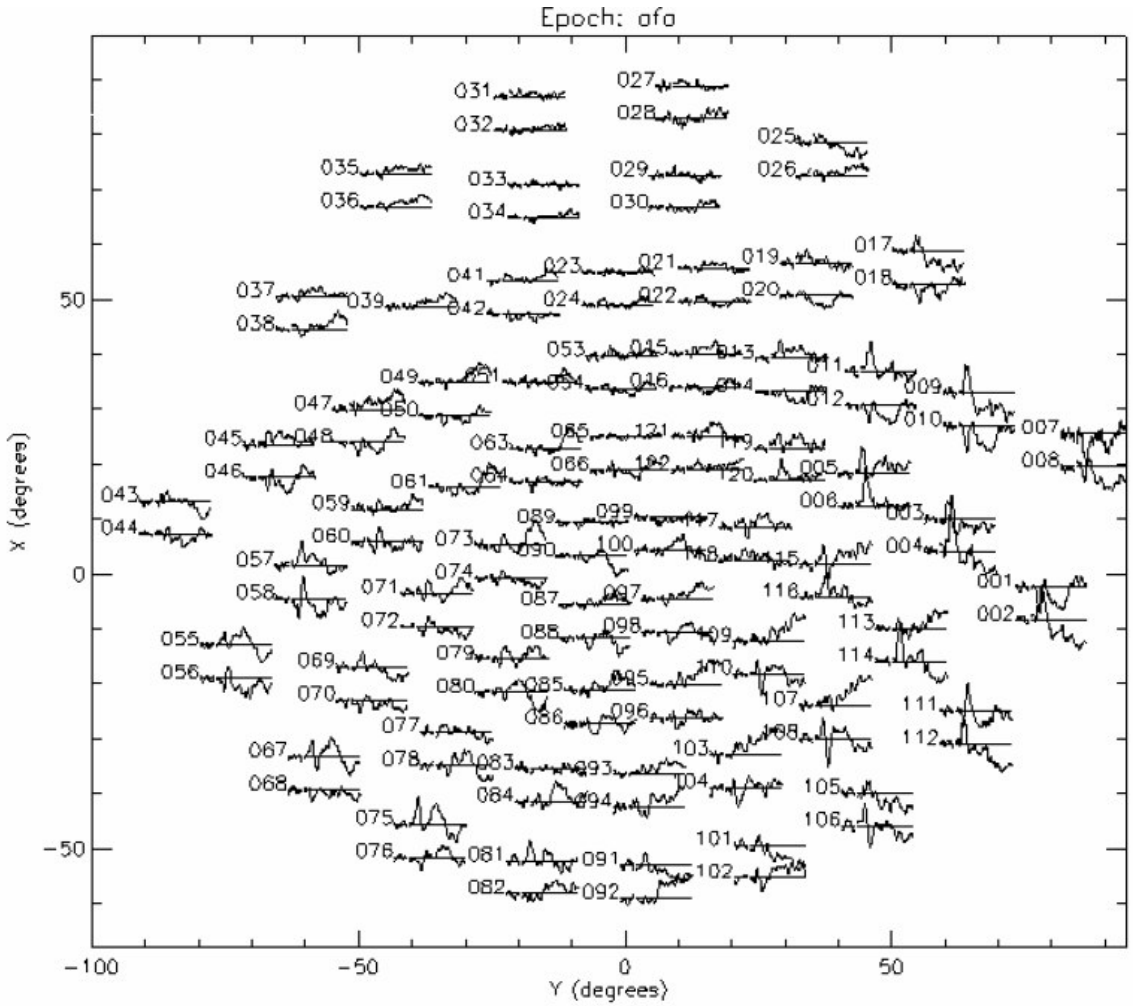


FIG. 41. The 122-channel gradiometer (“Neuromag-122”) built in Helsinki. The main modules inside the Dewar are the top plate made of fiber glass and a printed circuit board, the neck plug of fiber-glass-coated foam plastic, the four connector boards, and the wiring unit with 61 two-channel plug-in gradiometer components. Multipair shielded cables are laminated into the surface layer of the neck plug and lead from the top plate to the connector boards where the components for the SQUID gain control are located. On the right, the distribution of the sensors (below) and the gradiometer chip (above) with the two orthogonal figure-of-eight coils are illustrated schematically.

Example of MEG recordings



Why MEG?

- EEG = Electroencephalography
  - Time to place and calibrate EEG sensors
  - Skull has low conductivity to electric current but is transparent to magnetic fields
  - MEG not invasive, compared to placing EEG under skull
- fMRI = functional Magnetic Resonance Imaging
  - Time resolution 1s instead of 1ms
- Sources:
  - Barkley G.L et al., “MEG and EEG in Epilepsy”, Journal of Clinical Neurophysiology, May-June 2003
  - Malmuivo J. et al., “Sensitivity Distribution of EEG and MEG Measurements”, IEEE Transactions on Biomedical Engineering, March 1997

Inverse problem

- Localize source dipole **x** given sensor activation $\mathbf{B}_m$
- General approach uses forward modeling:
  1) Compute $\mathbf{B}_c(\mathbf{x})$ from **x**
  2) Compute cost $c(\mathbf{x}) = |\mathbf{B}_c(\mathbf{x}) - \mathbf{B}_m|^2$
  Iterate on **x** to minimize $c(\mathbf{x})$
- Used algorithms: Simplex, LM, ANN

(Jun et al., 2002)

Forward model

- 122 pairs of sensors
- Source dipole location **x** and moment **Q** (**Q** can be deduced from **x** and B_s)
- Sensor *s* has activation $B_s(\mathbf{x}, \mathbf{Q})$

$$B_s(\mathbf{x}, \mathbf{Q}) = \frac{\left[\mathbf{M}(\mathbf{x}, \mathbf{Q}; \mathbf{t})|_{\mathbf{t}=\mathbf{x}_s^1} - \mathbf{M}(\mathbf{x}, \mathbf{Q}; \mathbf{t})|_{\mathbf{t}=\mathbf{x}_s^2}\right] \cdot \mathbf{r}_s}{|\mathbf{x}_s^1 - \mathbf{x}_s^2|}, \quad s = 1, \cdots, 122,$$

$$\mathbf{M}(\mathbf{x}, \mathbf{Q}; \mathbf{x}_s) = \frac{\mu_0}{4\pi} \frac{F\mathbf{Q} \times \mathbf{x} - (\mathbf{Q} \times \mathbf{x} \cdot \mathbf{x}_s)\nabla F_{\mathbf{x}}}{F^2},$$

$$F(\mathbf{x}, \mathbf{x}_s) = d(x_s d + {x_s}^2 - (\mathbf{x}_s \cdot \mathbf{x}))$$

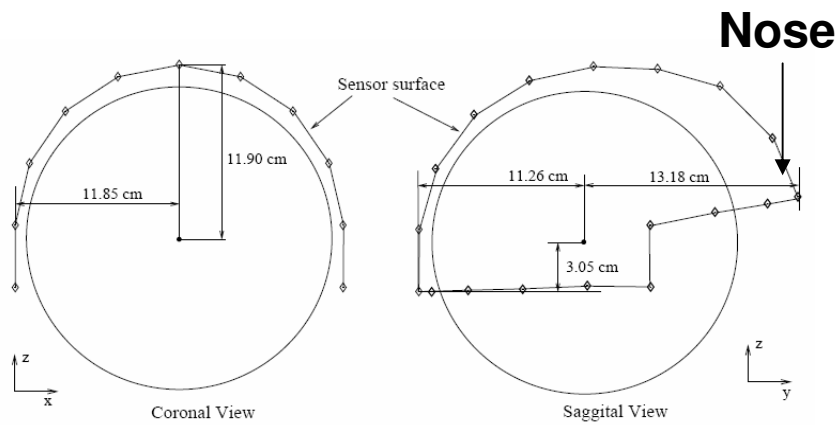
$$\nabla F_{\mathbf{x}}(\mathbf{x}, \mathbf{x}_s) = \left(\frac{d^2}{x_s} + \frac{(\mathbf{d} \cdot \mathbf{x}_s)}{d} + 2d + 2x_s\right)\mathbf{x}_s - \left(d + 2x_s + \frac{(\mathbf{d} \cdot \mathbf{x}_s)}{d}\right)\mathbf{x}$$

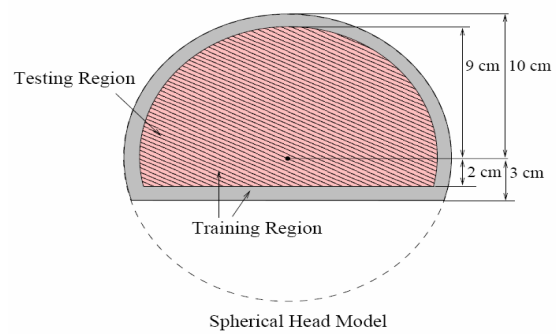
$$\mathbf{d} = \mathbf{x}_s - \mathbf{x}, \quad d = |\mathbf{x}_s - \mathbf{x}|, \quad x_s = |\mathbf{x}_s|,$$

(Jun et al., 2002)

About the dataset

- Geometry of sensors:
- Dataset consists in:
  - Pairs ($\mathbf{x}$, $\mathbf{B}_s$)
  - Training dataset
  - Testing dataset
- Sensor activations = forward model + noise model
- Examples of noise model: abrupt visual stimulation followed by brief motor output and audio feedback, measured far from the stimulus or response





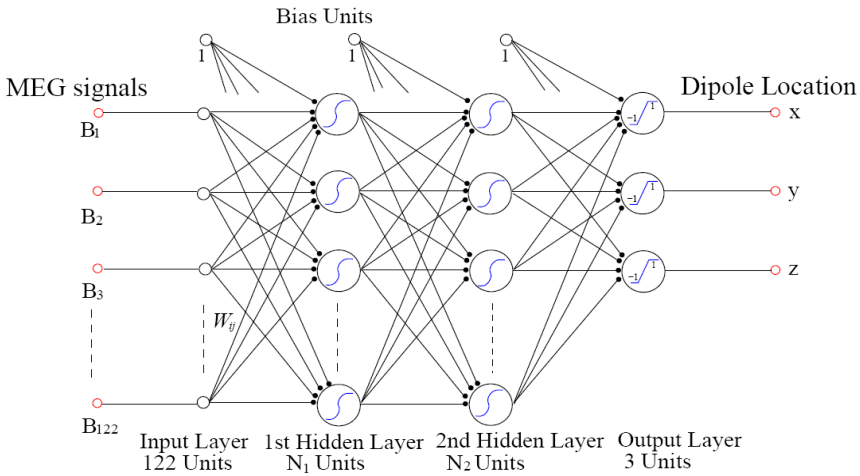
(Jun et al., 2002)

Distributed representation

- Usual architecture, "Cartesian-MLP" (Jun et al., 2002 and 2005) gives 3 outputs for dipole location

- Proposed approach in Jun et al., 2003: "distributed representation" Dipole location **x** deduced from **G**(**x**), vector of Gaussian receptive fields
- Receptive field *i* located at $\mathbf{x}_i$ (regularly distributed) is:

$$G_i(\mathbf{x}) = \exp\left(-\left\|\mathbf{x}-\mathbf{x}_i\right\|^2\right)/2\sigma^2$$



"Cartesian-MLP"

Decoding **G**(**x**) to **x**

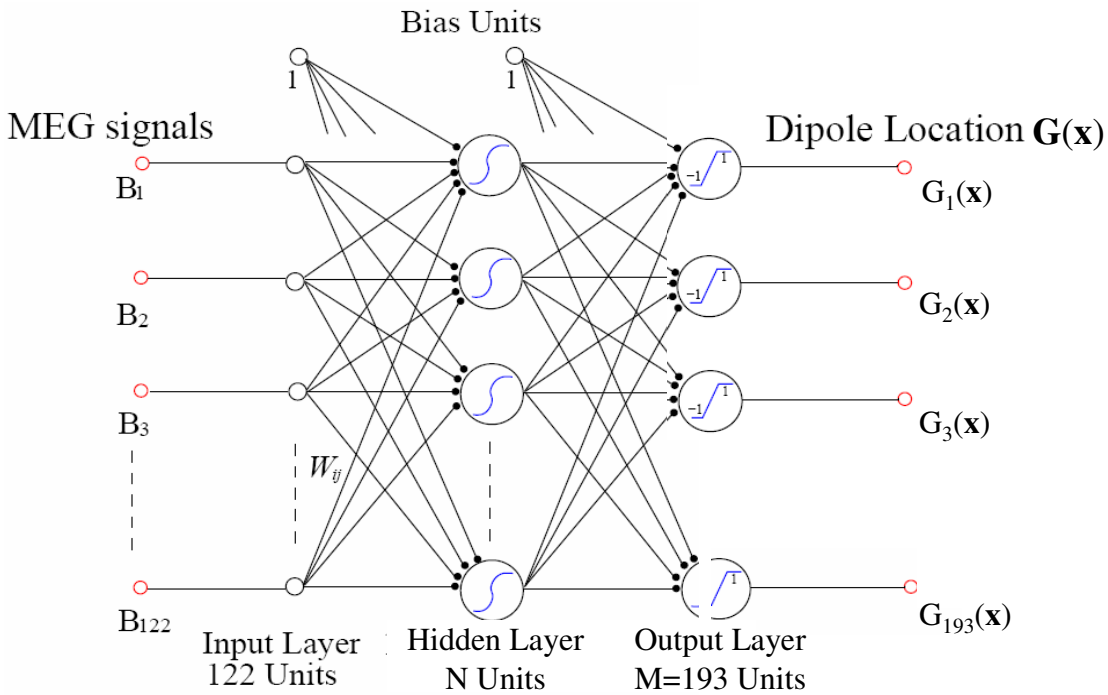
- Strategy 1:
  Find $i^* = \arg\max_i G_i(\mathbf{x})$
  Interpolate between centers $\mathbf{x}_i$ inside of a ball B_{i^*} centered on $\mathbf{x}_{i^*}$ and of radius 6cm
  $\sum_{\mathbf{x_i} \in B_{i^*}} G_i(\mathbf{x})\, \mathbf{x}_i / \sum_{\mathbf{x_i} \in B_{i^*}} G_i(\mathbf{x})$

  6cm = twice the inter-center distance

- Strategy 2:
  For each receptive field center $\mathbf{x}_i$, compute within ball B_i
  $c_i = G_i(\mathbf{x}) + \sum_{j \neq i, \mathbf{x}_j \in B_i} G_i(\mathbf{x}) / \|\mathbf{x}_i - \mathbf{x}_j\|$
  Find $i^* = \arg\max_i c_i$
  Interpolate between centers $\mathbf{x}_i$ inside of a ball B_{i^*}
  $\sum_{\mathbf{x_i} \in B_{i^*}} G_i(\mathbf{x})\, \mathbf{x}_i / \sum_{\mathbf{x_i} \in B_{i^*}} G_i(\mathbf{x})$

  (i.e. take into account neighborhood influence)

Soft-MLP architecture



Receptive field centers $\mathbf{x}_i$ evenly distributed every 3cm and cover training region

$$G_i(\mathbf{x}) = \exp\left(-\left\|\mathbf{x}-\mathbf{x}_i\right\|^2\right)/2\sigma^2 \longleftarrow \text{1.8cm}$$

"Hyperbolic activation units to accelerate training", c.f. LeCun 1991

Soft-MLP architecture

- Input data preprocessing:
  MEG sensor activations scaled to
  RMS=0.5
- Training algorithm:
  Backpropagation with online stochastic gradient descent
  Learning rate η empirically chosen

Choice of MLP architecture

- Number of nodes in hidden layer
  N=20, 40, 60, 80, 120, 160
- 500 training epochs
- Training (noise free) dataset size:
  500, 1000, 2000, 4000, 8000, 16000, 32000
- Testing (noise free) dataset size:
  5000
- Generalization error averaged over 5 runs
- ➢ Asymptote in generalization error:
  N=80 hidden units, **8000 training samples**

Dataset

- Training: 20000 samples (with brain noise)
- Testing: 4500 samples (with brain noise)
- 500 epochs
- 12 hours per training!?

- Signal to noise ratio in data:
  SNR = $20\log_{10} P_s/P_n$

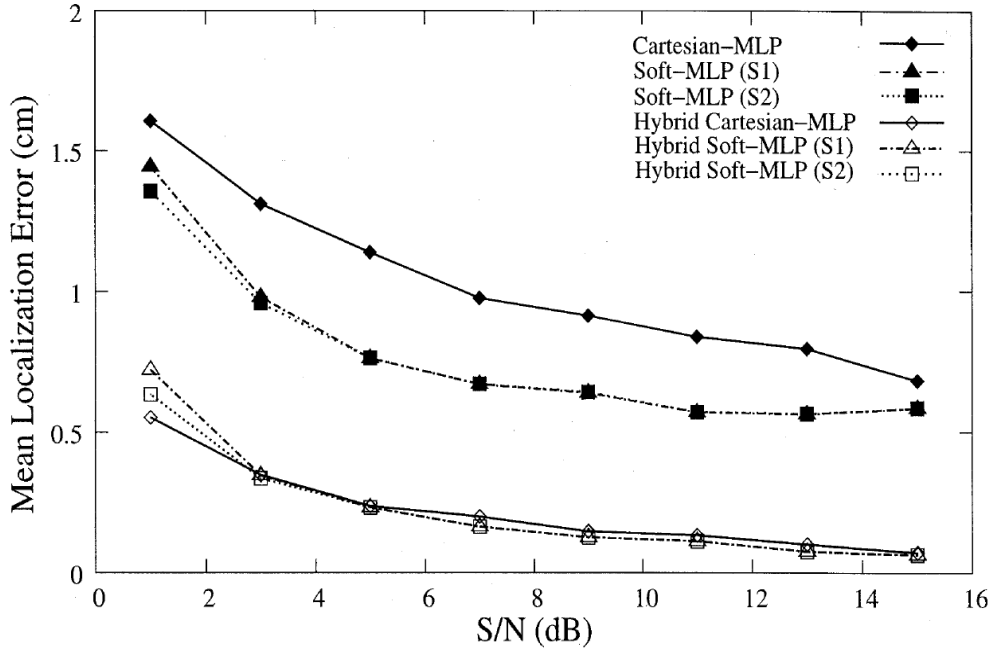
  (Noise has RMS:
  P_n = 50 to 100 fT/cm)

DISTRIBUTION OF SNR FOR THE 4500 TESTING PATTERNS

SNR (dB)	# of Patterns	Frequency (%)
0–2	892	19.82
2–4	806	17.91
4–6	824	18.31
6–8	627	13.93
8–10	538	11.96
10–12	381	8.47
12–14	229	5.09
> 14	203	4.51

Results

- Soft-MLP faster and more accurate than Cartesian-MLP
- But hybrid methods (when MLP provides initial guess for LM, Levenberg-Marquard optimization) are still better



Comparison vs. LM alone

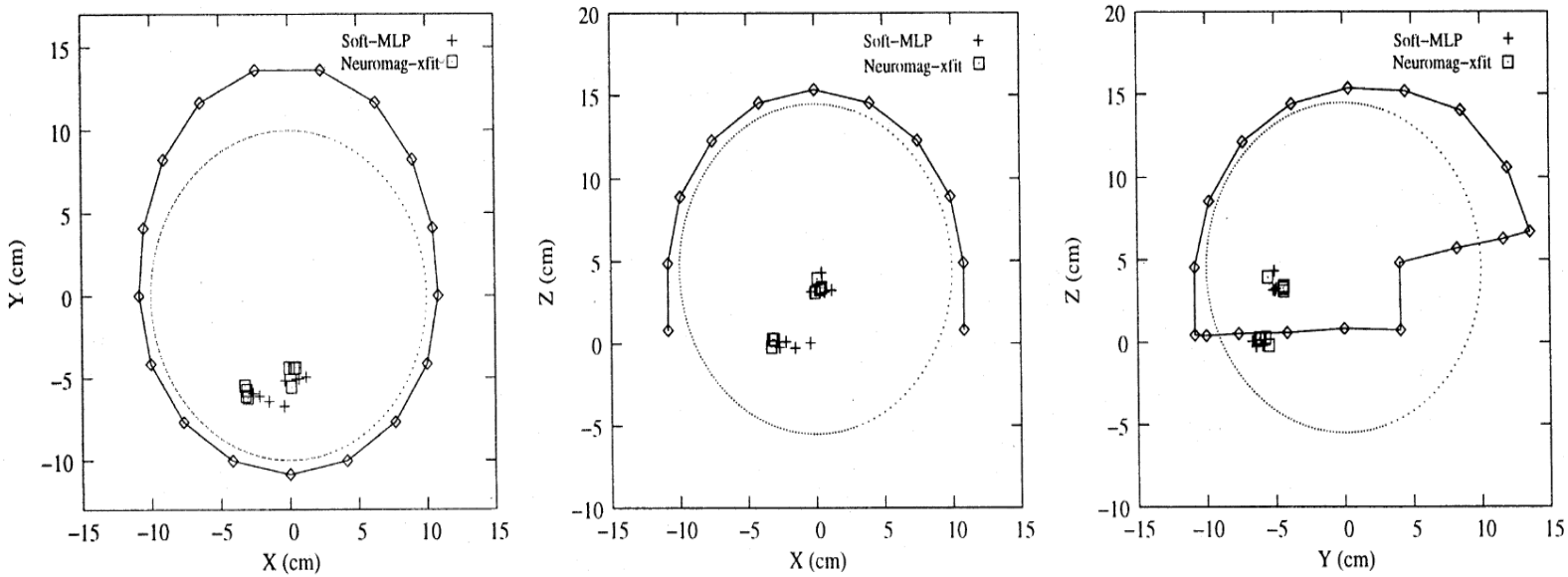
(Jun et al. 2002)

algorithm	trained noise	computation time (ms)	localization error (cm)
fixed-4-start-LM	—	449	1.16
random-20-start-LM	—	2 175	0.31
optimal-1-start-LM	—	22	0.23
MLP	N	0.3	2.70
	W	0.3	1.64
	C	0.3	2.06
	B	0.3	1.15
MLP-start-LM	N	53	0.84
	W	41	0.44
	C	49	0.67
	B	36	0.28

(Cartesian-MLP)

"Real brain noise"

Comparison with commercial software in MEG Neuroimag-122



xfit = commercial software
Soft-MLP used with Strategy 2

Good initial guess

Possible extensions

- Cartesian-MLP can localize only 1 dipole
- Soft-MLP could localize more…
- Better than global search algorithm

Area of research since 1991

- Abeyratne U.R. et al., “Artificial Neural Networks for Source Localization in the Human Brain”, Brain Topography, vol. 4, 1991
- Use of MLP with backpropagation on EEG signals
- Other article cited: Jun S.C., Pearlmutter B.A., Nolte G., “Fast Accurate MEG Source Localization Using a MLP Trained with Real Brain Noise”, Physics in Medicine and Biology, June 2002

Questions…

- Why use only 1 sample of Bs recording at a time? Does the shape of the MEG curves play a role?
- …