

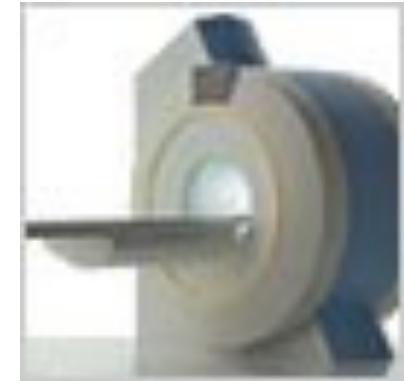
Lineare Regression: Das lineare Modell

- Problemstellung am Beispiel funktioneller Kernspintomographie (fMRI)
- Lineares Perceptron und Lineare Regression
- Parameterschätzung
- Varianzschätzung und Konfidenzintervalle
- Anwendungsbeispiel fMRI-Analyse:
Hirnaktivierung als Statistische Parameterkarten (SPM)

Anwendungsbeispiel: Analyse von fMRI-Daten

Technik: funktionelle Kernspintomographie

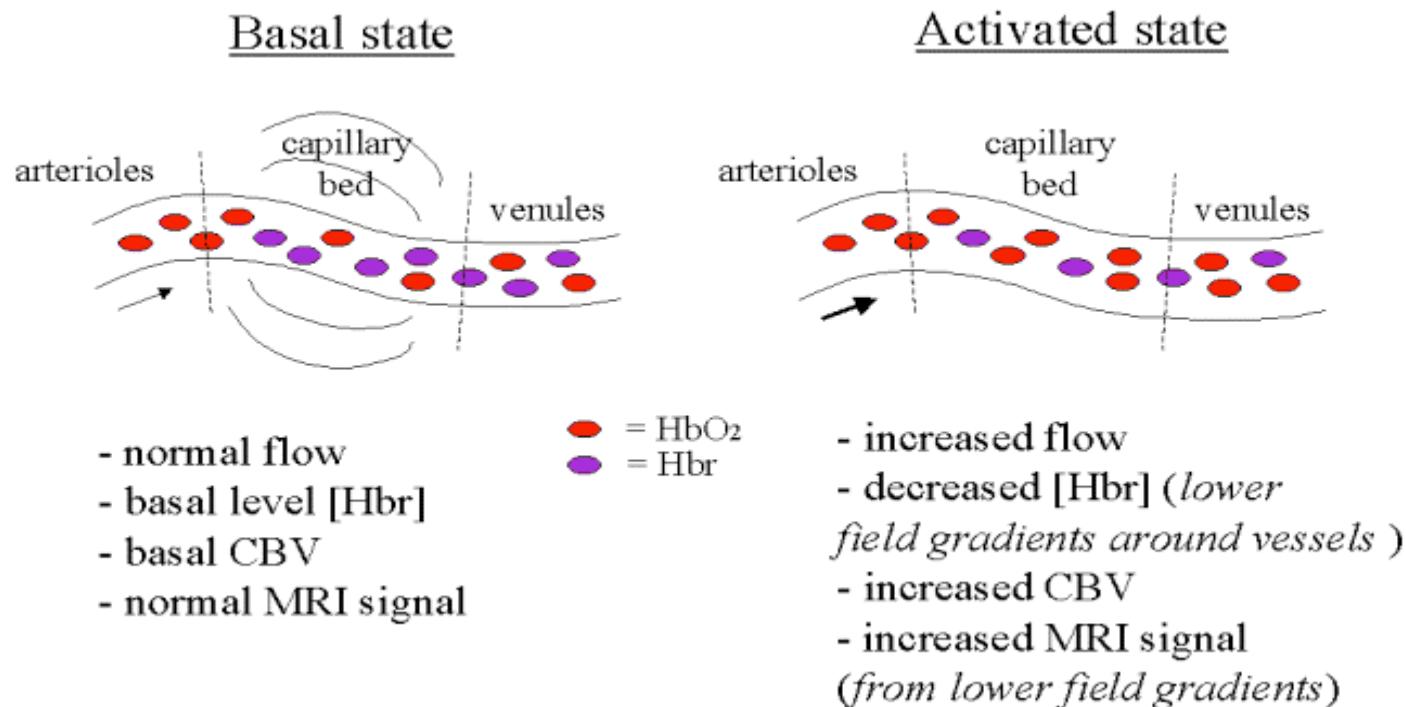
- „functional Magnetic Resonance Imaging“ (fMRI)
- Magnetmomente der Atome werden in einem starken Magnetfeld angeregt (Magnetresonanz)
- Sie relaxieren in den Grundzustand zurück
- Die Relaxation wird durch Gradientenfelder ortsabhängig gemacht (=> Imaging)
- Die Relaxation ist abhängig von der molekularen Umgebung
 - Wasserstoffkonzentration: strukturelle MRI
 - Paramagnetische Substanzen (Hbr): funktionelle MRI



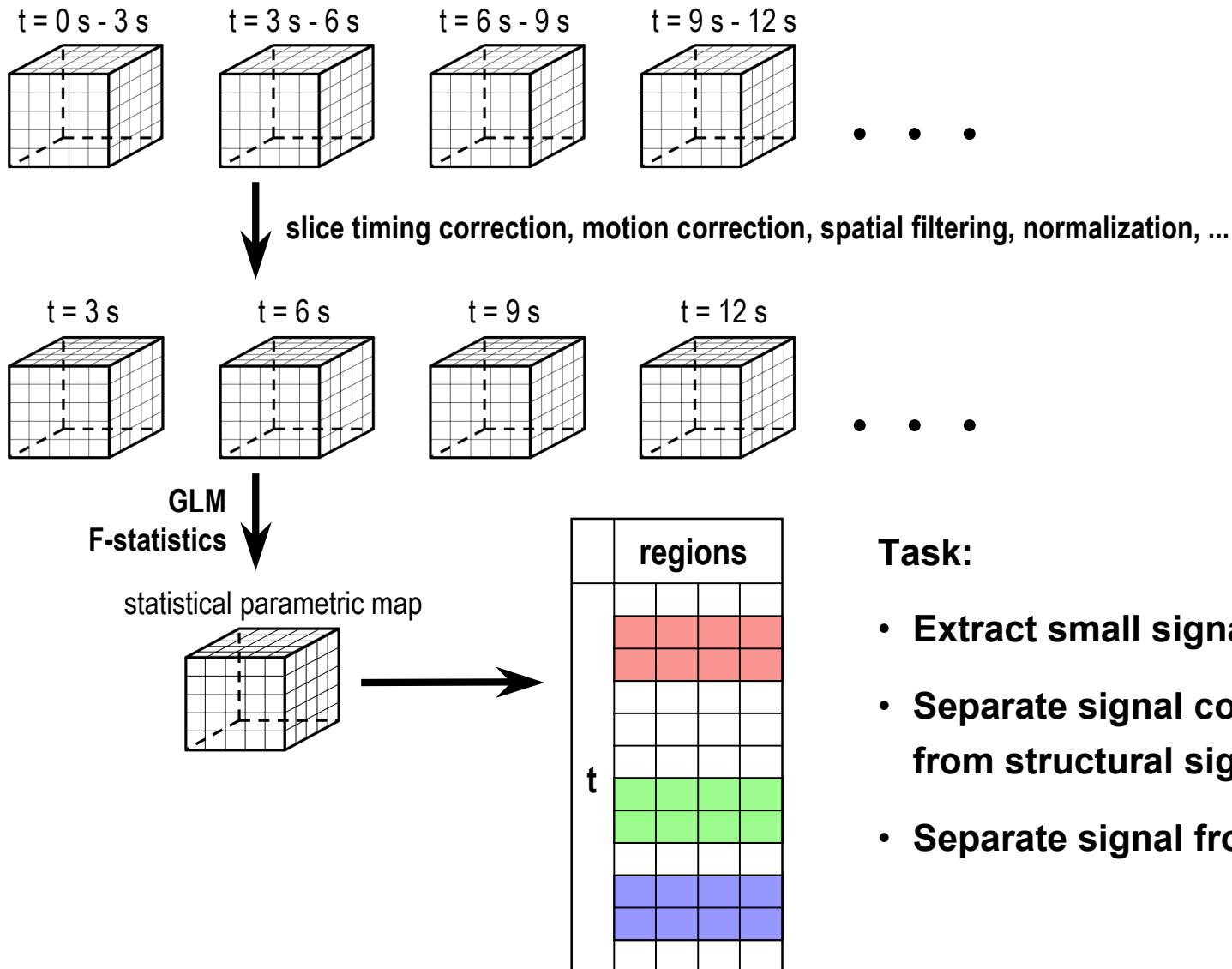
BOLD Signal

in fMRI experiments the BOLD signal (blood oxygen level dependent) is measured:

- At least two components: local de-oxygenation and local increase in flow



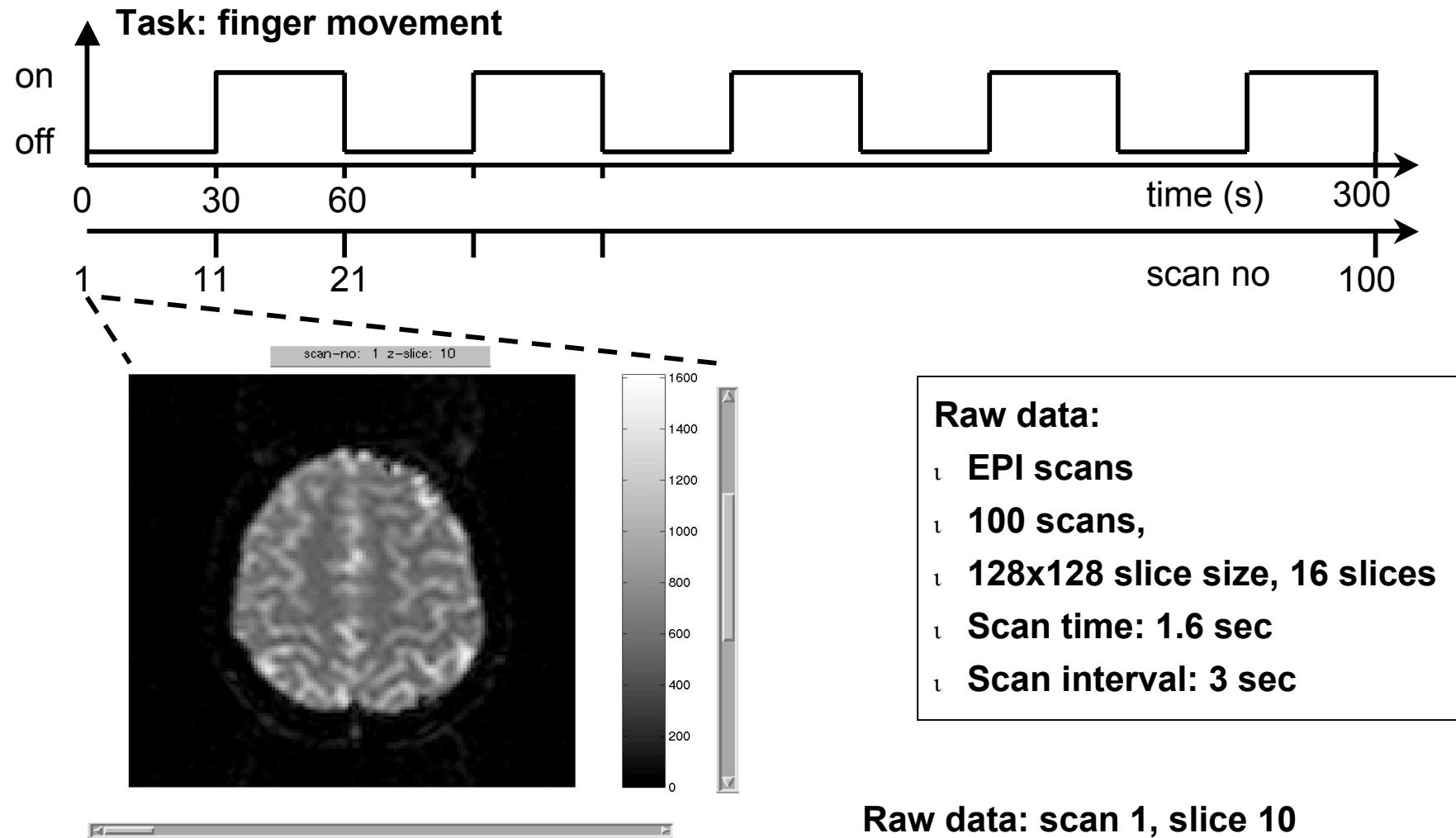
fMRI Data Format



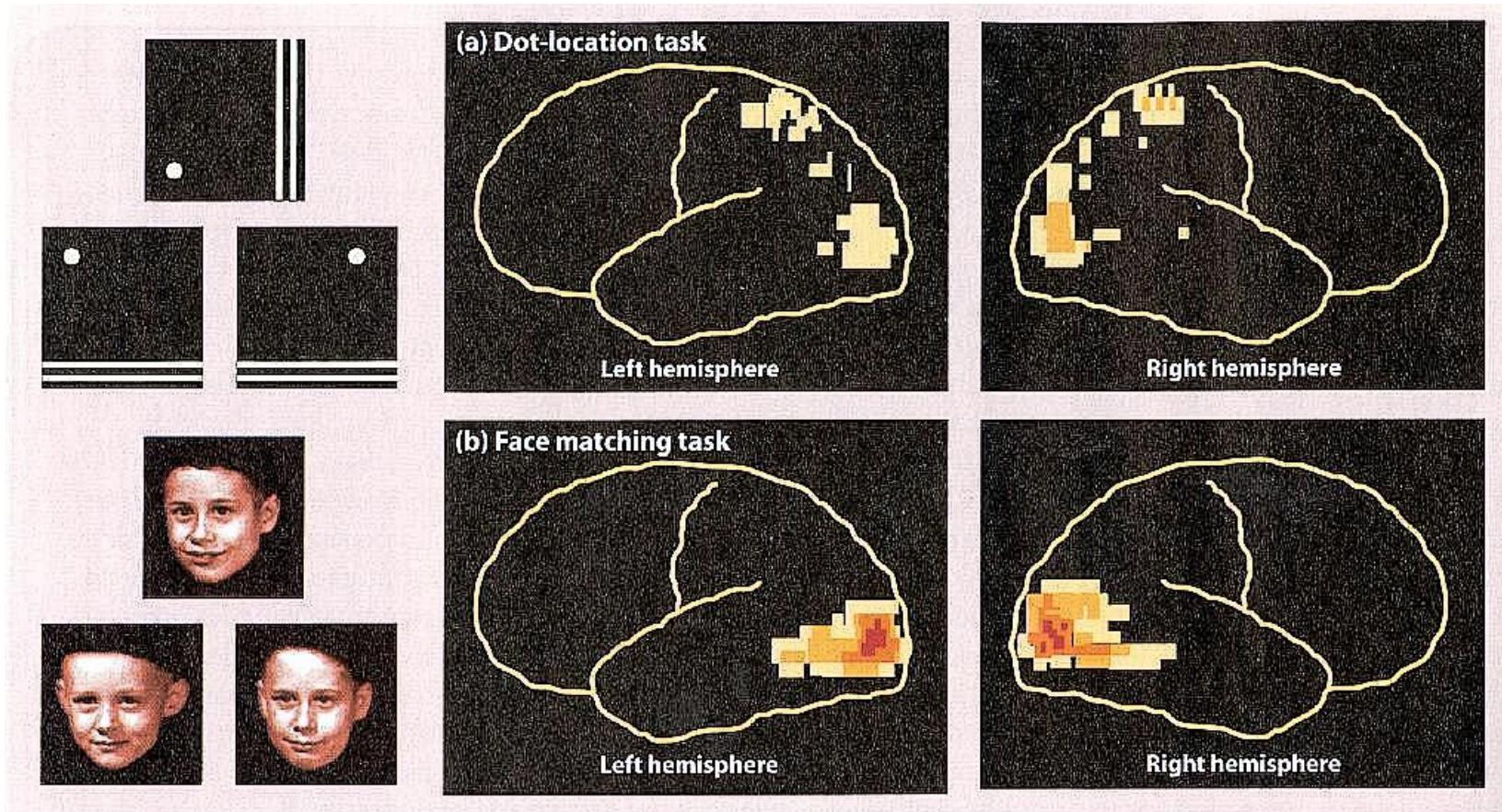
Task:

- Extract small signal
- Separate signal components (BOLD from structural signal, background)
- Separate signal from noise

Typical Experimental Task and Data Format



Funktionelle Kernspintomographie: Ein Beispiel



Das lineare Modell

- Ausgangspunkt: Lineares Perceptron

$$\hat{y}(\mathbf{x}) = \sum_{i=1}^d w_i x_i = \mathbf{w} \cdot \mathbf{x}$$

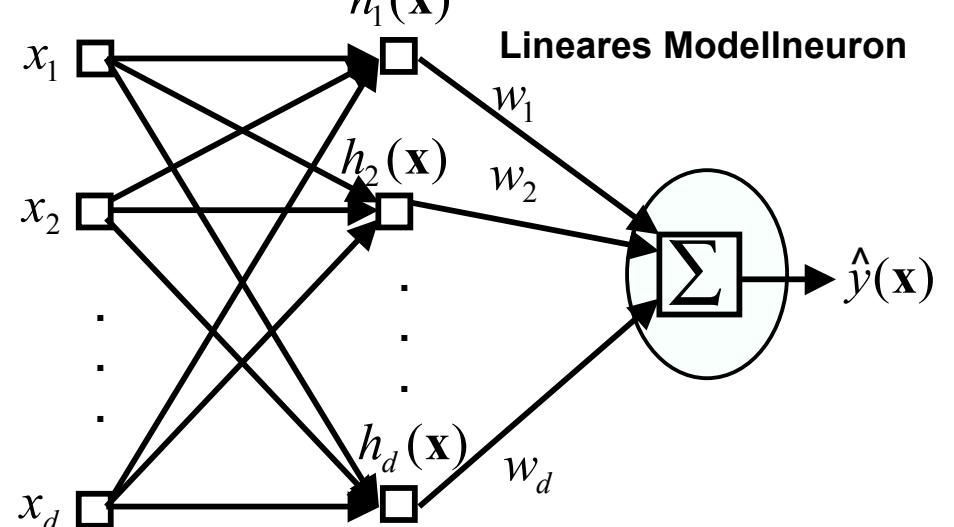
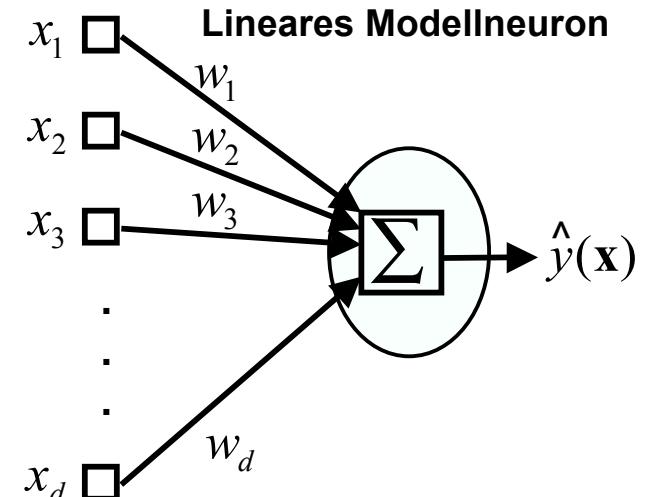
- Fittet Ebene (mit Gradient \mathbf{w})

Interessanter:

- Lineares Perceptron mit Vorverarbeitung

$$\begin{aligned}\hat{y}(\mathbf{x}) &= h_1(\mathbf{x})w_1 + h_2(\mathbf{x})w_2 + \dots + h_d(\mathbf{x})w_d \\ &= \sum_{i=1}^d h_i(\mathbf{x})w_i\end{aligned}$$

- Perceptron mit VV ist linear in \mathbf{w}
- Aber: Nichtlinear in \mathbf{x}
- Funktionen $h(\mathbf{x})$ sind von Hand vorgegeben, werden nicht gelernt



- **Geschrieben als Regressionsmodell**

Allgemein: $y = \hat{y} + n = f(\mathbf{x} | \mathbf{w}) + n$

Lineares Modell: $y = \sum_{i=1}^d h_i(\mathbf{x}) w_i + n$

Bsp: Polynom-Fit:

$$\hat{y} = w_0 + w_1 x + w_2 x^2 + \dots + w_n x^n$$

- **Empirische Formulierung, Design-Matrix:**

Betrachte Datensatz $D = \{(\mathbf{x}^{(m)}, y^{(m)}), m = 1, \dots, M\}$

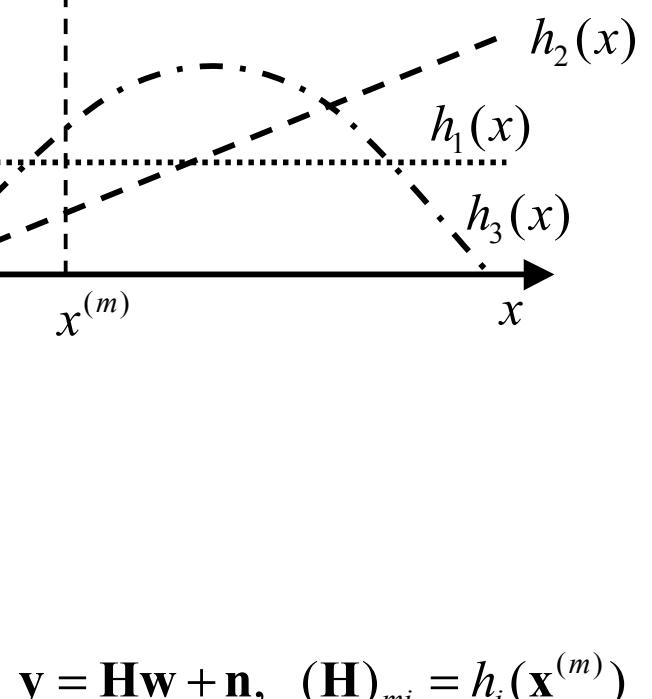
$$\begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(M)} \end{pmatrix} = \begin{pmatrix} h_1(\mathbf{x}^{(1)}) & \dots & h_d(\mathbf{x}^{(1)}) \\ h_1(\mathbf{x}^{(2)}) & \ddots & h_d(\mathbf{x}^{(2)}) \\ \vdots & & \vdots \\ h_1(\mathbf{x}^{(M)}) & \dots & h_d(\mathbf{x}^{(M)}) \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{pmatrix} + \begin{pmatrix} n^{(1)} \\ n^{(2)} \\ \vdots \\ n^{(M)} \end{pmatrix}$$

Output

Design-Matrix

Parameter

Rauschen



$$\mathbf{y} = \mathbf{H}\mathbf{w} + \mathbf{n}, \quad (\mathbf{H})_{mi} = h_i(\mathbf{x}^{(m)})$$

- „Lernen“: ML-Parameterschätzung des besten lin. Modells

Likelihood: $p(D | \mathbf{w}) \equiv p(\mathbf{y} | \mathbf{w}) = p_{\mathbf{n}}(\mathbf{y} - \mathbf{H}\mathbf{w})$

Annahme: Gaussisches weißes Rauschen

$$p_{\mathbf{n}}(\mathbf{n}) = \prod_m p_n(n^{(m)}) = \prod_m \frac{1}{(2\pi\sigma_n^2)^{1/2}} \exp\left(-\frac{n^{(m)2}}{2\sigma_n^2}\right)$$

$$= \frac{1}{(2\pi\sigma_n^2)^{M/2}} \exp\left(-\frac{\mathbf{n}^T \mathbf{n}}{2\sigma_n^2}\right)$$

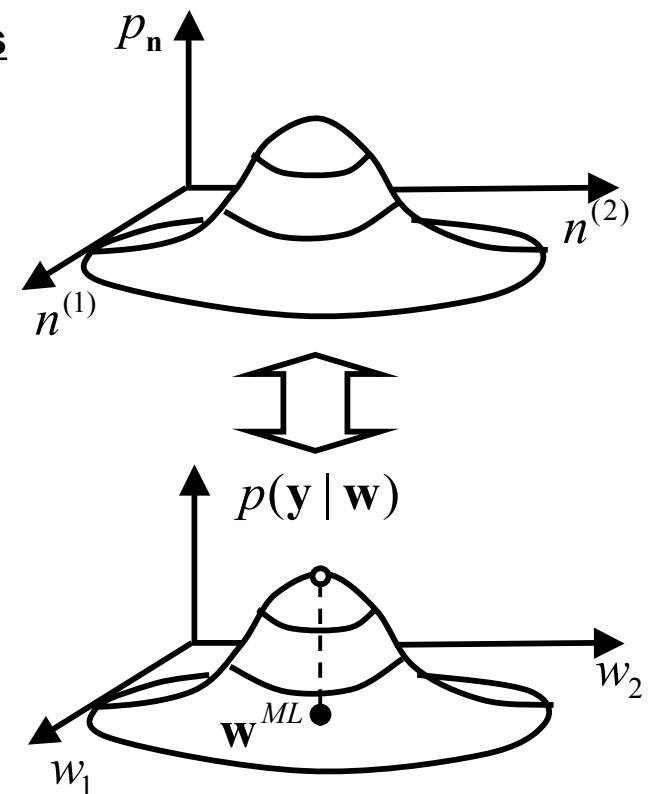
$$p(\mathbf{y} | \mathbf{w}) \propto \exp\left(-\frac{(\mathbf{y} - \mathbf{H}\mathbf{w})^T (\mathbf{y} - \mathbf{H}\mathbf{w})}{2\sigma_n^2}\right)$$

- Maximum-Likelihood Parameter

$$\begin{aligned} 0! = \nabla_{\mathbf{w}} \ln p(\mathbf{y} | \mathbf{w}) &= -\frac{1}{2\sigma_n^2} \nabla_{\mathbf{w}} (\mathbf{y}^T \mathbf{y} - 2\mathbf{w}^T \mathbf{H}^T \mathbf{y} + \mathbf{w}^T \mathbf{H}^T \mathbf{H} \mathbf{w}) \\ &= \frac{1}{\sigma_n^2} (\mathbf{H}^T \mathbf{y} - \mathbf{H}^T \mathbf{H} \mathbf{w}) \\ &= \frac{(\mathbf{H}^T \mathbf{H})}{\sigma_n^2} ((\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y} - \mathbf{w}) \end{aligned}$$

$\Rightarrow \hat{\mathbf{w}} = \mathbf{w}^{ML} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}$

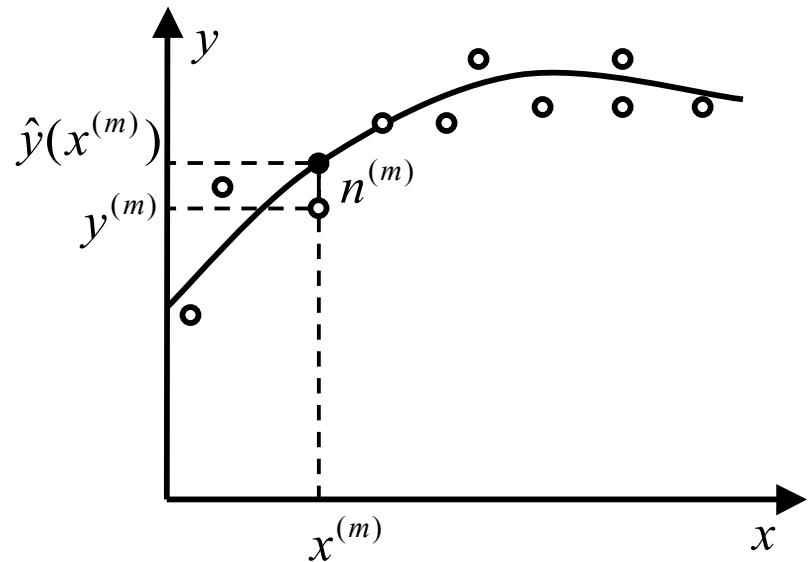
analytisch berechenbar!



- **Schätzung der Rauschvarianz :**

Schätzer des Rauschvektors:

$$\begin{aligned}\hat{\mathbf{n}} &= \mathbf{y} - \hat{\mathbf{y}} \\ &= \mathbf{y} - \mathbf{H}\hat{\mathbf{w}} \\ &= \mathbf{y} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H} \mathbf{y} \\ &=: \mathbf{R}\mathbf{y}\end{aligned}$$



Residuum-erzeugende Matrix: $\mathbf{R} = \mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}$

Schätzer Rauschvarianz: $\hat{\sigma}_n^2 = \frac{\sum_{m=1}^M \hat{n}^{(m)2}}{M - (\approx \text{Anz. geschätzter Parameter})}$

$$\boxed{\hat{\sigma}_n^2 = \frac{\hat{\mathbf{n}}^T \hat{\mathbf{n}}}{\text{tr}(\mathbf{R})}}$$

- Schätzung der Parametervarianz

Beobachtung:

$$\nabla_{\mathbf{w}} \ln p(\mathbf{y} | \mathbf{w}) = \frac{(\mathbf{H}^T \mathbf{H})}{\sigma_n^2} ((\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y} - \mathbf{w}) = \frac{(\mathbf{H}^T \mathbf{H})}{\sigma_n^2} (\hat{\mathbf{w}} - \mathbf{w}) = -\frac{(\mathbf{H}^T \mathbf{H})}{\sigma_n^2} (\mathbf{w} - \hat{\mathbf{w}})$$

$$\Rightarrow \ln p(\mathbf{y} | \mathbf{w}) = -\frac{1}{2} (\mathbf{w} - \hat{\mathbf{w}})^T \frac{(\mathbf{H}^T \mathbf{H})}{\sigma_n^2} (\mathbf{w} - \hat{\mathbf{w}}) \equiv -\frac{1}{2} (\mathbf{w} - \hat{\mathbf{w}})^T \Sigma_{\mathbf{w}}^{-1} (\mathbf{w} - \hat{\mathbf{w}})$$

$$\Rightarrow p(\mathbf{y} | \mathbf{w}) \propto \exp\left(-\frac{1}{2} (\mathbf{w} - \hat{\mathbf{w}})^T \Sigma_{\mathbf{w}}^{-1} (\mathbf{w} - \hat{\mathbf{w}})\right)$$

Parametervektor folgt einer multivariaten Gaußverteilung

- Kovarianzmatrix der Parameter:

$$\Sigma_{\mathbf{w}} = \sigma_n^2 (\mathbf{H}^T \mathbf{H})^{-1}$$

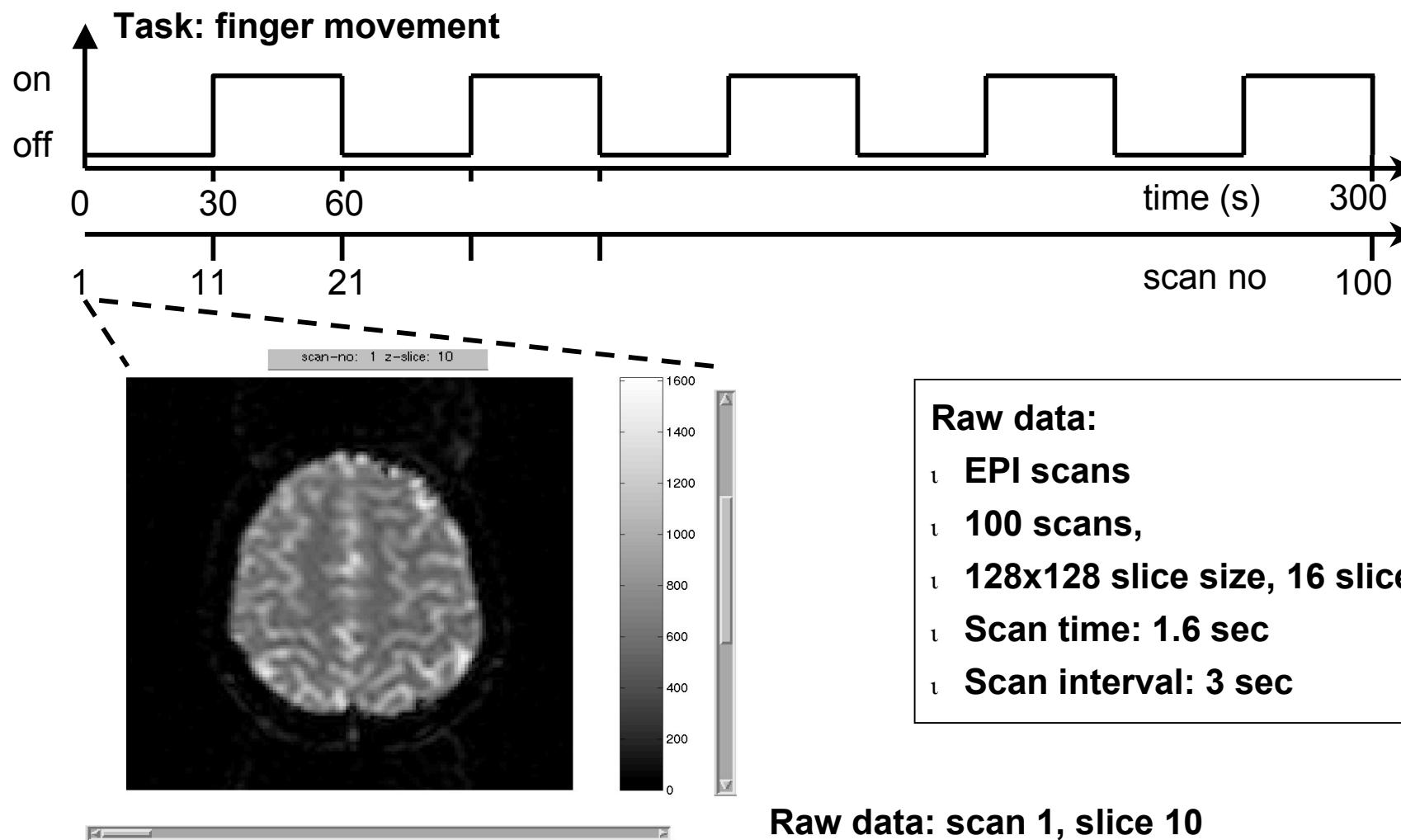
- Varianz des i -ten Parameters:

$$\sigma_i^2 = (\sigma_n^2 (\mathbf{H}^T \mathbf{H})^{-1})_{ii}$$

- Geschätztes Signal-Rausch-Verhältnis: Der Z-score

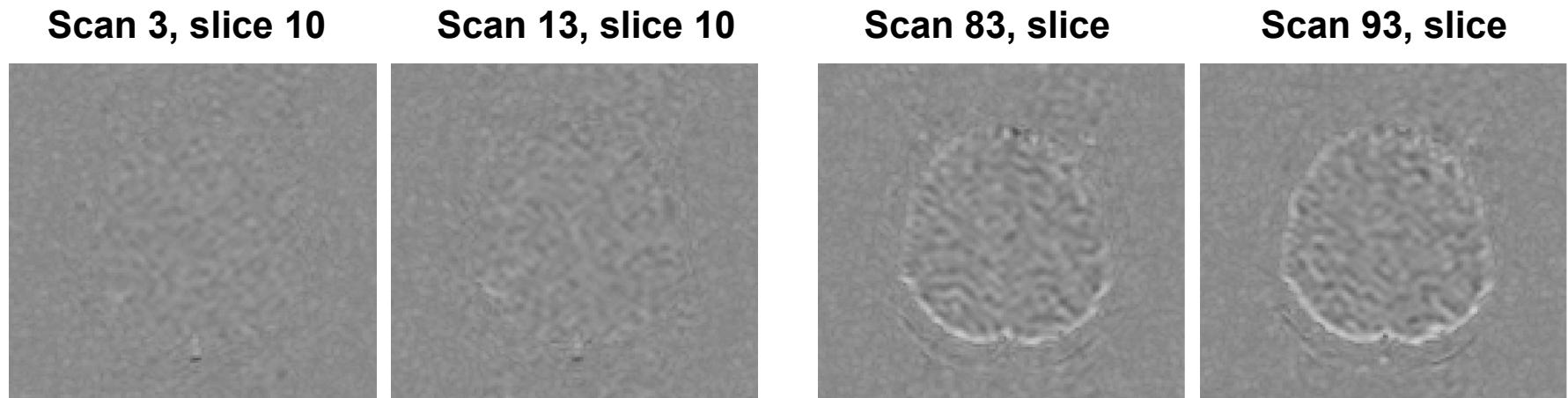
$$Z_i =: \frac{\hat{w}_i}{\hat{\sigma}_i} = \frac{w_i}{(\Sigma_{\mathbf{w}})_{ii}^{1/2}} = \frac{w_i}{\left(\hat{\sigma}_n^2 (\mathbf{H}^T \mathbf{H})_{ii}^{-1} \right)^{1/2}}$$

Linear Model Analysis of fMRI Time Series



fMRI Raw Images: Visual Inspection

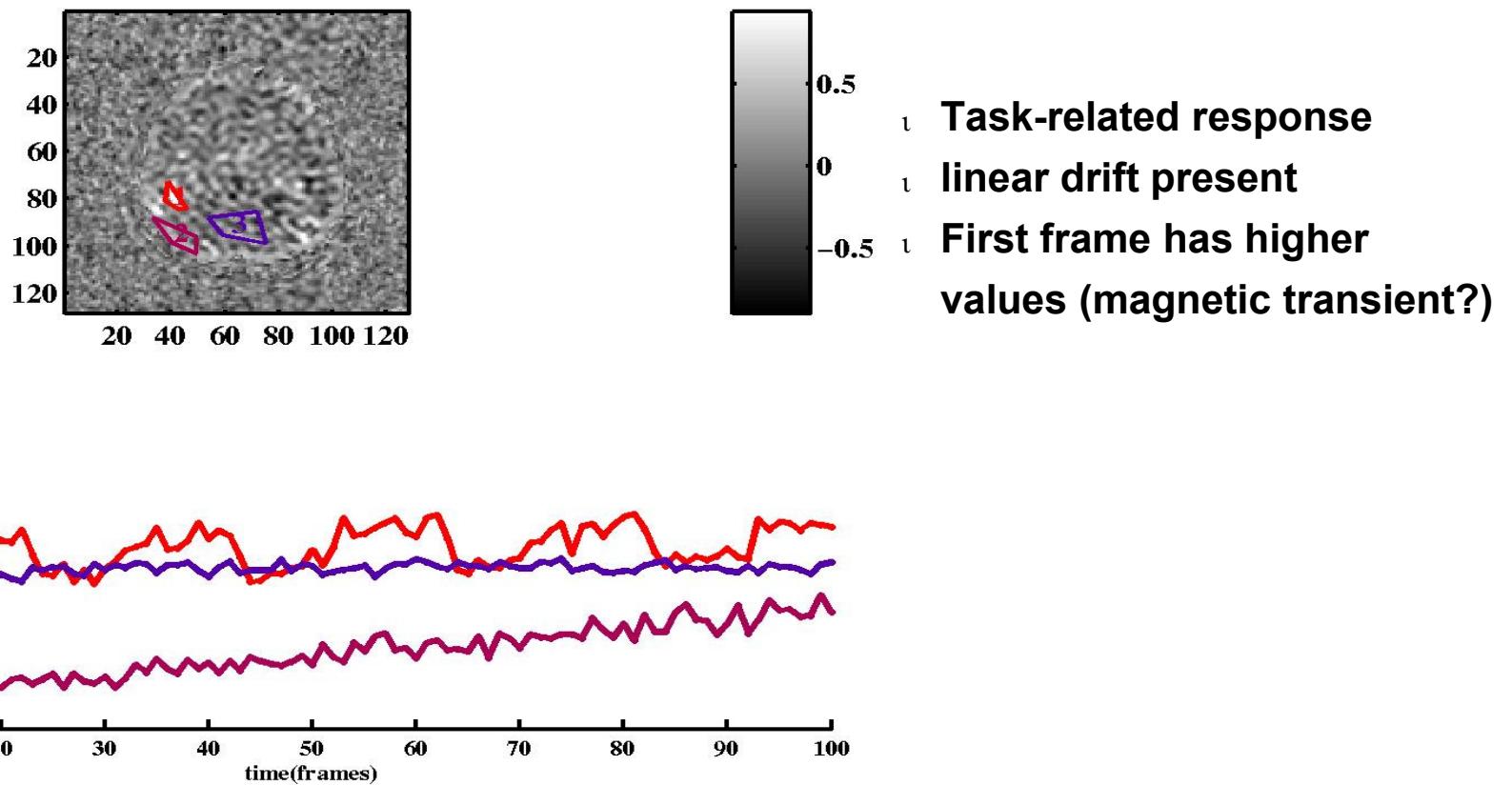
After subtraction of second scan from the others (first frame has higher intensity)



- „ Strong correlated noise; correlations extend over several voxels
- „ From comparison of single scans with / without task no signal visible
- „ Slow drift in the head position (movement artifact in later scans)

fMRI Time Series: Visual Inspection

Mean time series over three regions: Signal, drift, and background



General Linear Model: Design Matrix

Principle: $\mathbf{X} = \mathbf{H}\mathbf{w} + \mathbf{n}$

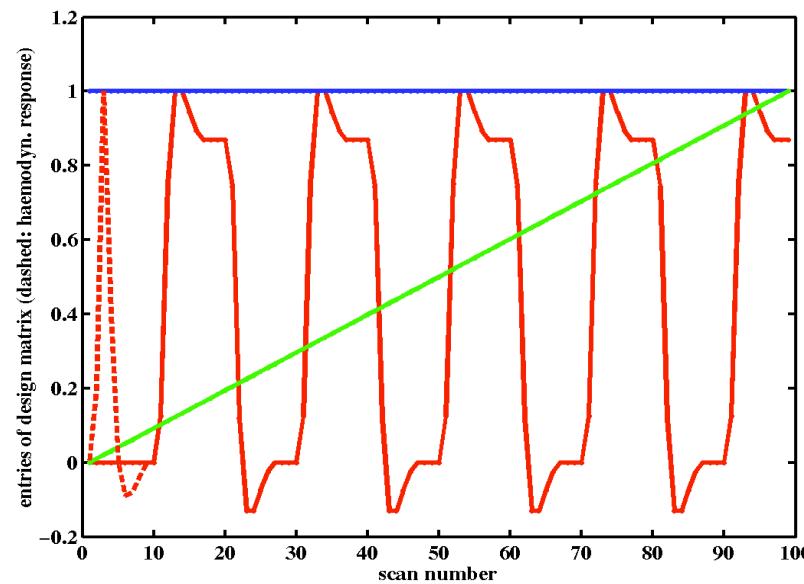
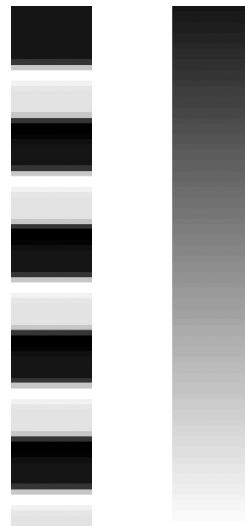
\mathbf{X} = matrix of time series

\mathbf{H} = Design matrix

\mathbf{w} = parameter (estimate used for t-test)

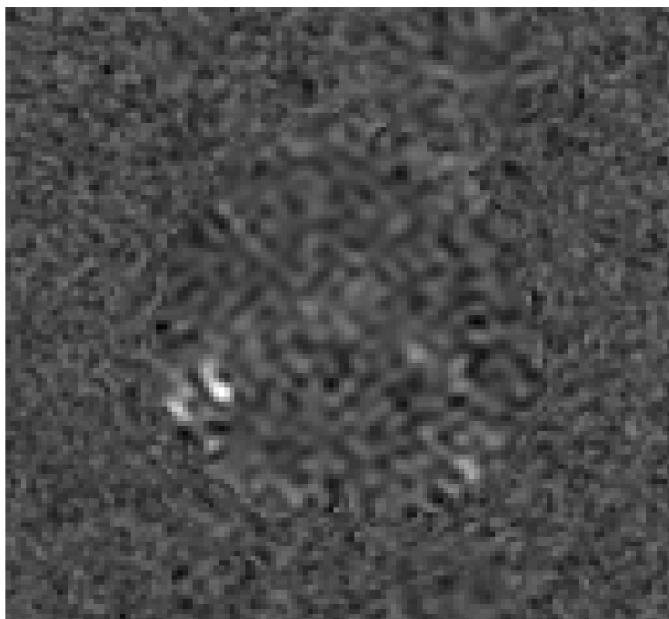
\mathbf{n} = noise (estimate used for t-test)

Entries of design matrix used:



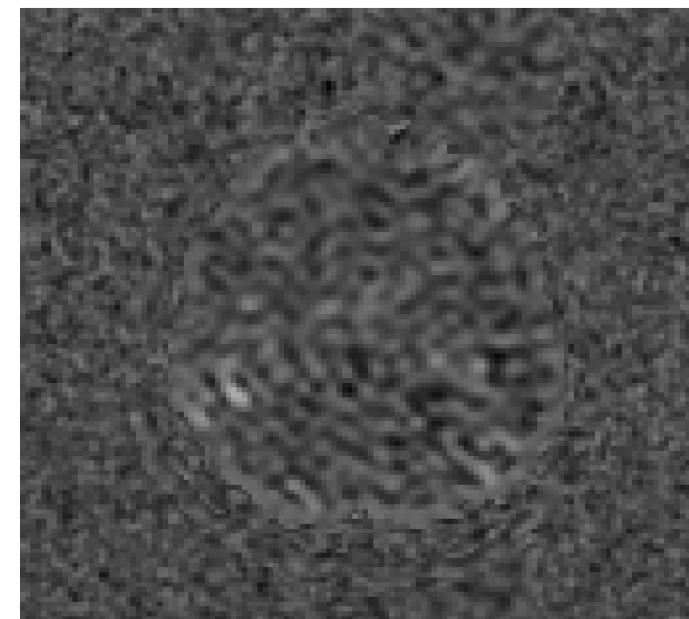
GLM: Statistical Parametric Maps (SPM)

3 component GLM on raw data



$$w_1(x, y)$$

Raw t-test on raw data



$$t(x, y)$$

- „ Higher absolute t-value in GLM (due to convolution with hemodynamic response)
- „ t-values along head border are suppressed in GLM (because drift is modeled)