The GLM, experimental design and efficiency – part 1

Dr Steffen Bollmann

Postdoctoral Researcher and NIF Facility Fellow
Centre for Advanced Imaging, University of Queensland, Australia

(Re-)Sources

- SPM Course Slides from
    - Klaas Enno Stephan, Jean-Baptiste Poline, Rik Henson, Christian Ruff, Jakob Heinzle, Frederike Petzschner, Sandra Eglesias
- http://imaging.mrc-cbu.cam.ac.uk/imaging/CbuImaging
- http://www.fil.ion.ucl.ac.uk/spm/doc/books/hbf2/
- https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=spm
- http://www.fil.ion.ucl.ac.uk/spm/course/video/

Contents

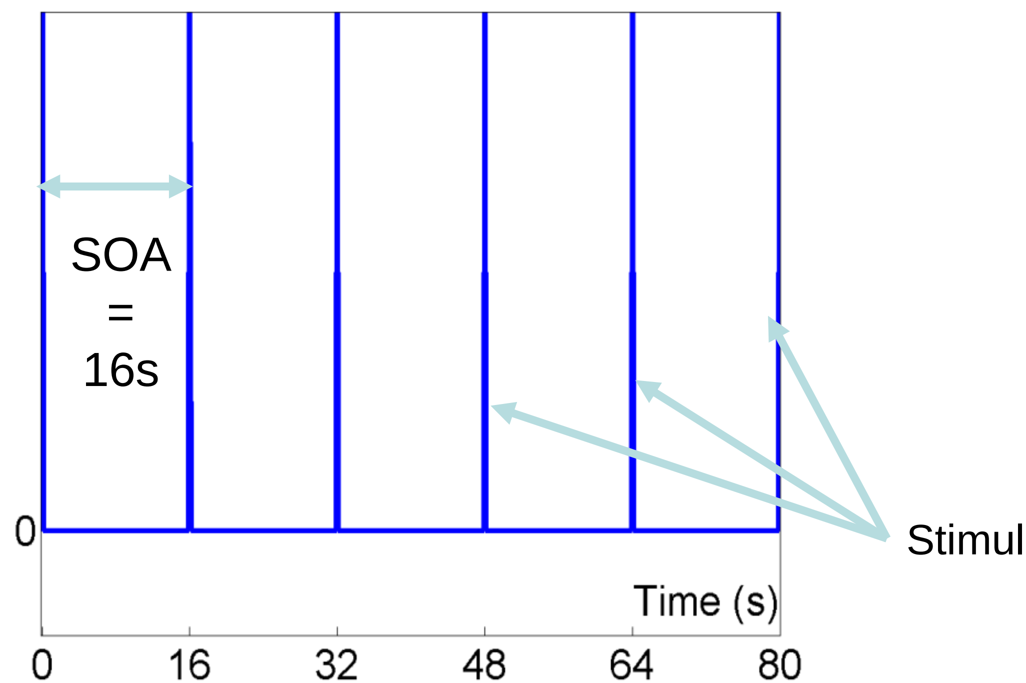
1. Definitions
2. The General Linear Model
3. Statistical Inference
4. How to estimate the efficiency of a design?

Contents

1. Definitions
2. The General Linear Model
3. Statistical Inference
4. How to estimate the efficiency of a design?

Terms & Definitions

SOA =
Stimulus
Onset
Asynchrony



Terms & Definitions

Epochs

- periods of sustained stimulation
- in SPM defined by duration > 0

Events

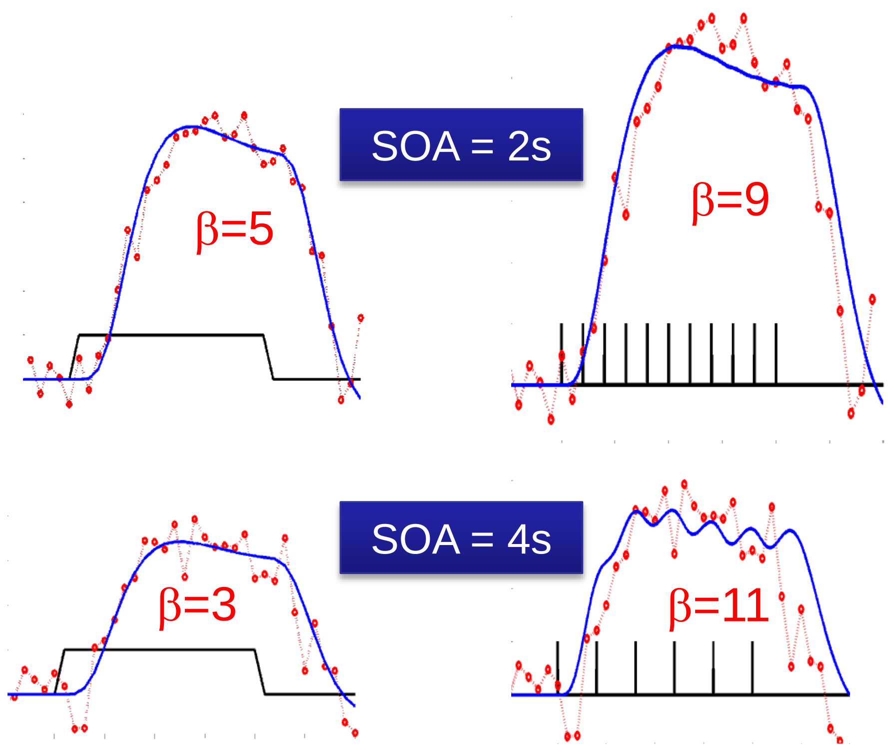
- impulses (delta-functions)
- in SPM defined by duration=0

Terms & Definitions

Near-identical regressors can be created by

1. **sustained epochs**
2. **rapid (SOAs<~3s) series of events**

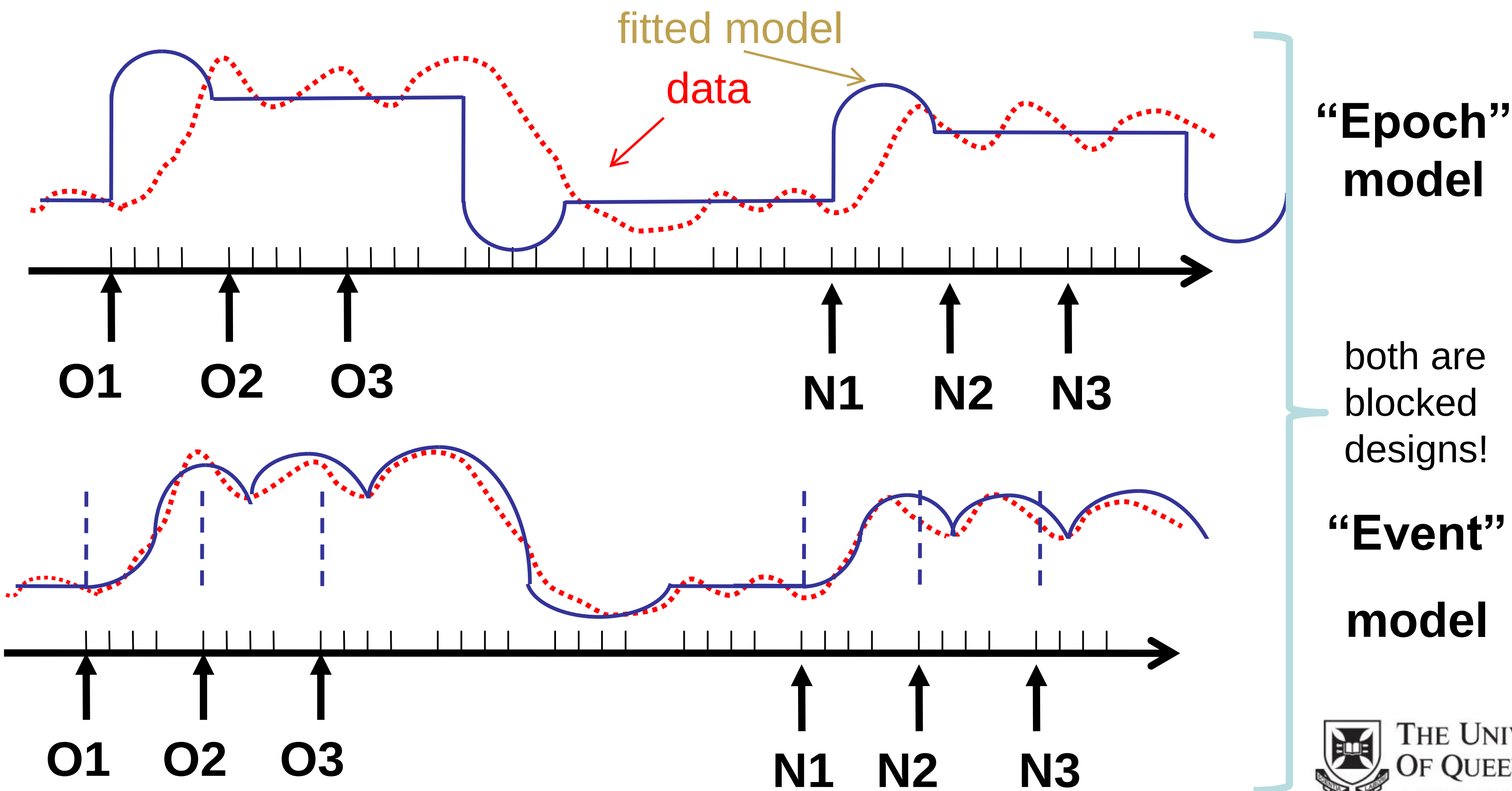
An “epoch” model will estimate parameters (β) that ***decrease*** with SOA



An “event” model will estimate parameters (β) that ***increase*** with SOA

Terms & Definitions

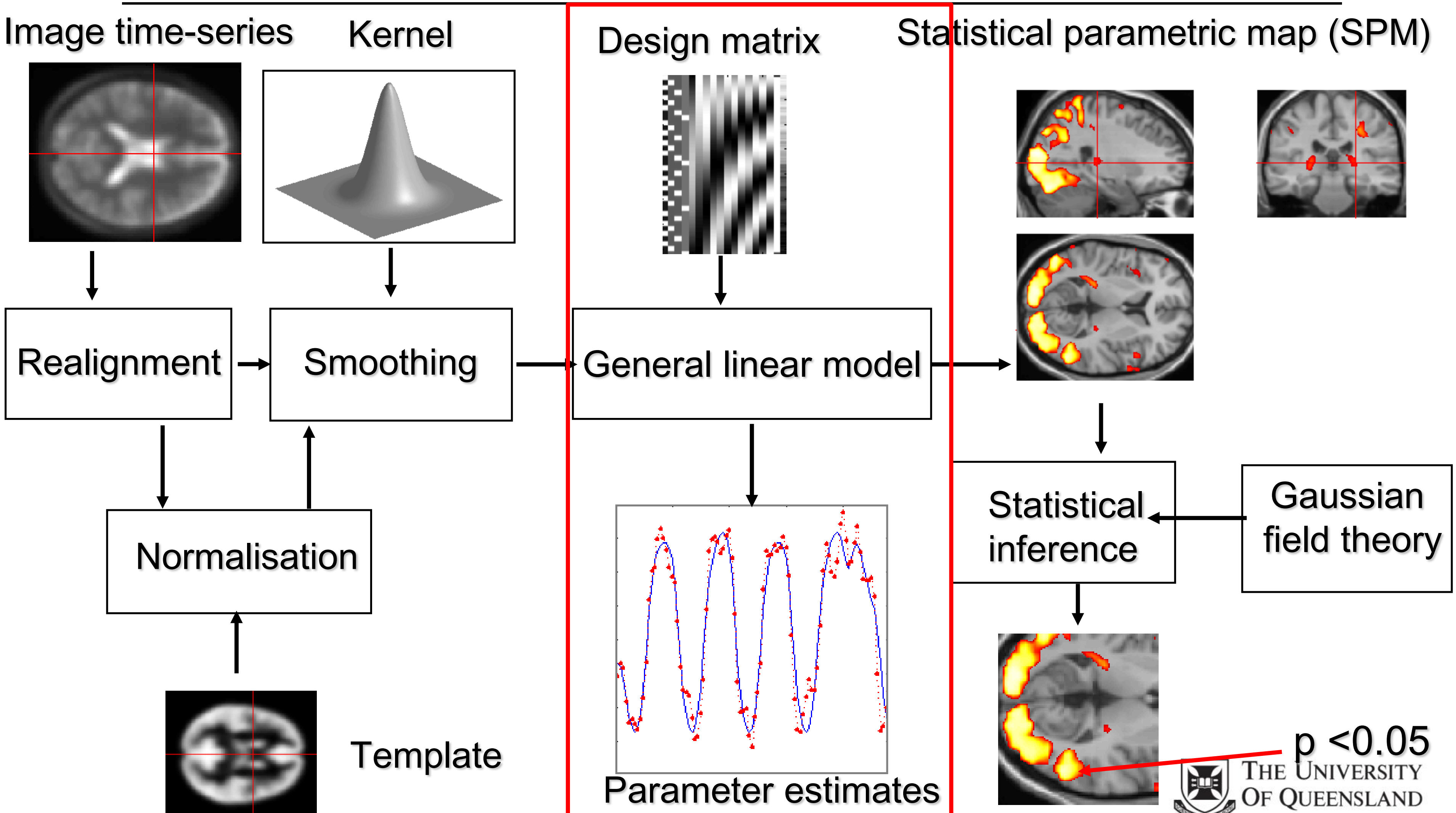
- **Designs can be: blocked or intermixed**
- **Models can be: epoch or event-related**



Contents

1. Definitions
2. The General Linear Model
3. Statistical Inference
4. How to estimate the efficiency of a design?

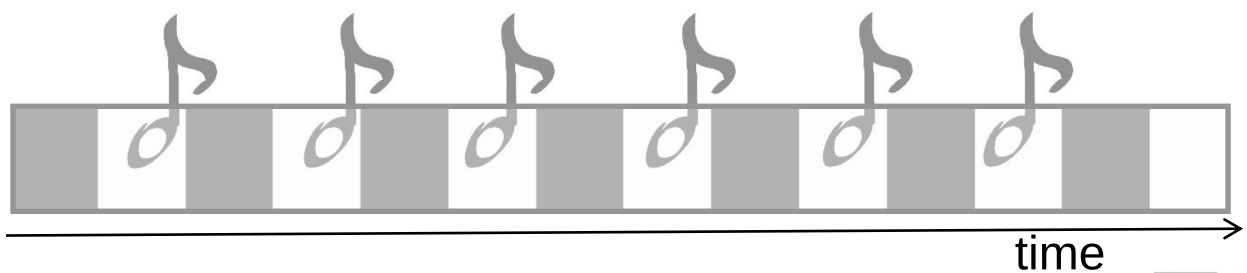
SPM Overview



The General Linear Model

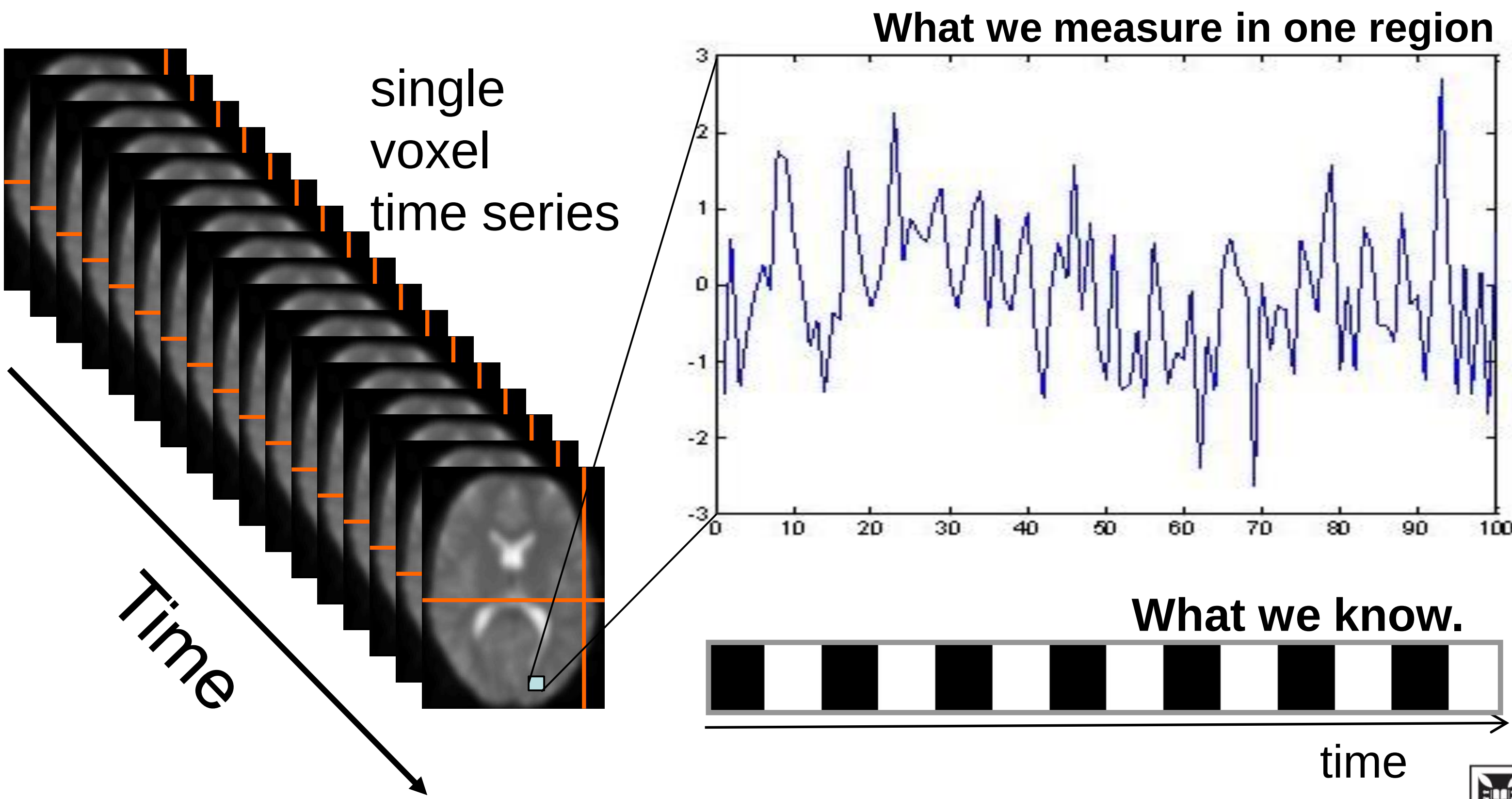
Let us do an fMRI experiment

7 cycles of rest and listening



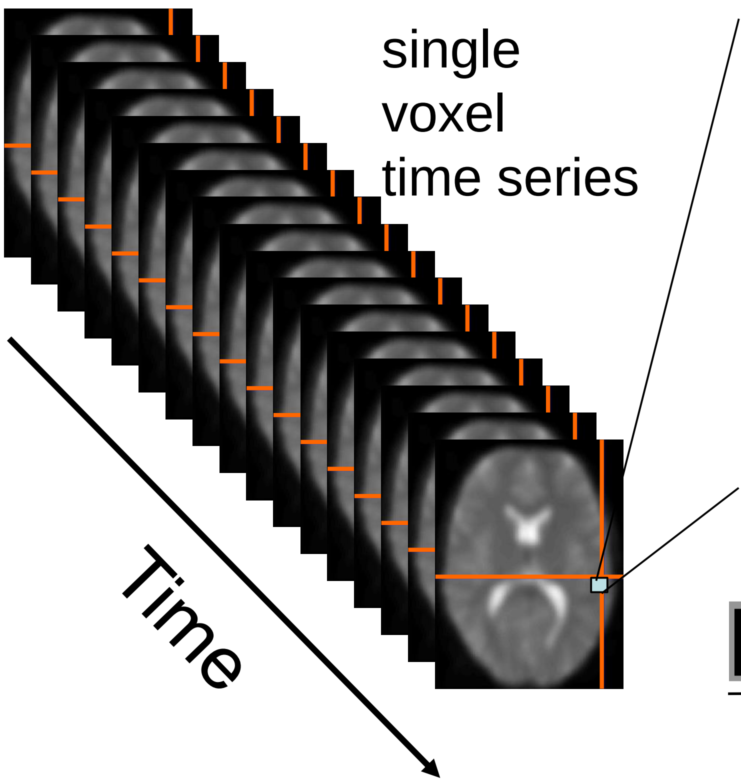
The General Linear Model

Question: Where in the brain do we see a change in BOLD activity comparing listening to rest?

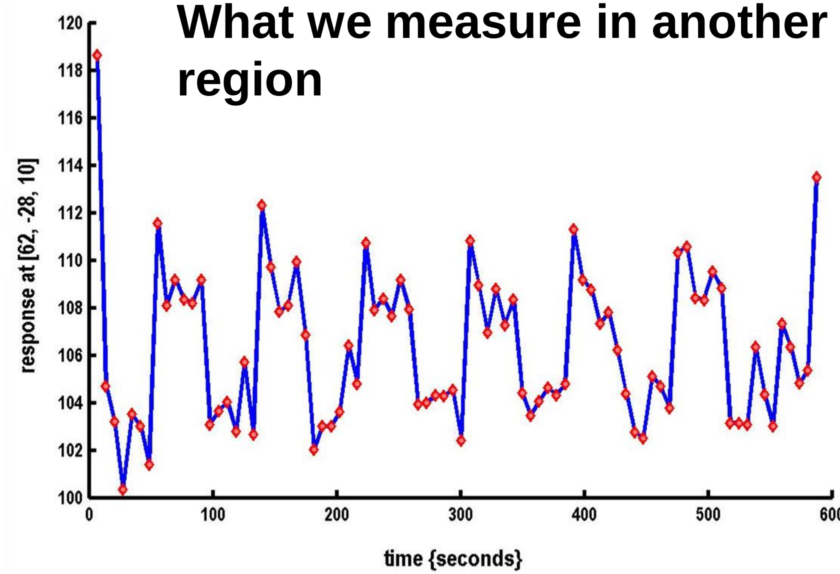


The General Linear Model

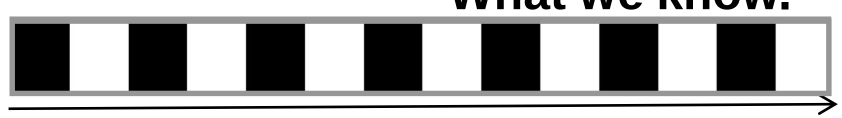
Question: Where in the brain do we see a change in BOLD activity comparing listening to rest?



What we measure in another region



What we know.

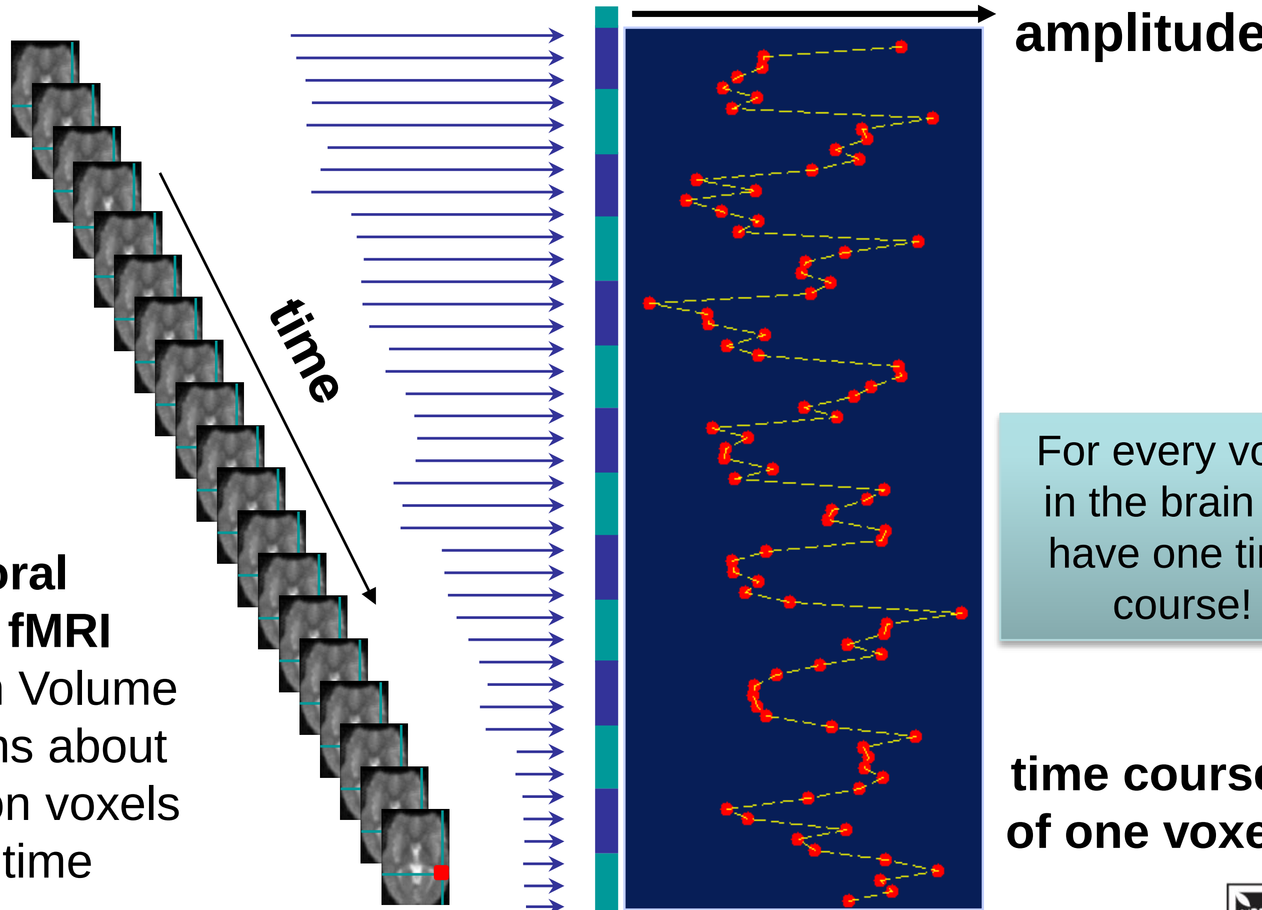


time

The University of Queensland
AUSTRALIA
Create change

The General Linear Model

We fit one model per voxel ! (= mass-univariate approach)



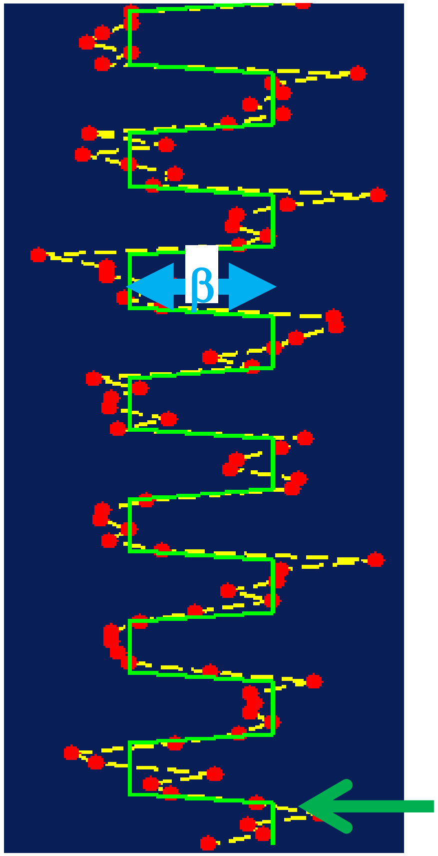
Temporal series fMRI
⇨each Volume contains about 2 million voxels at one time point

For every voxel in the brain we have one time course!

time course of one voxel

The General Linear Model

Fitting the model = finding the best estimate of the betas by minimising the error (often named residuals)



The „height“ of the fitted regressor is the β value

(–> you end up with 2 million betas in the brain - for each voxel there is one beta)

regressor is fitted to the data by height adjustment, the beta tells us the height of the regressor to get the smallest error!

The General Linear Model

- The computation of the betas is done by Ordinary Least Squares (OLS)
  - If we can assume that the noise is i.i.d.

$$\varepsilon \sim N(0, \sigma^2 I)$$

  - Then we can compute the betas and get the optimal solution, which minimizes the error between the design matrix X and our data y

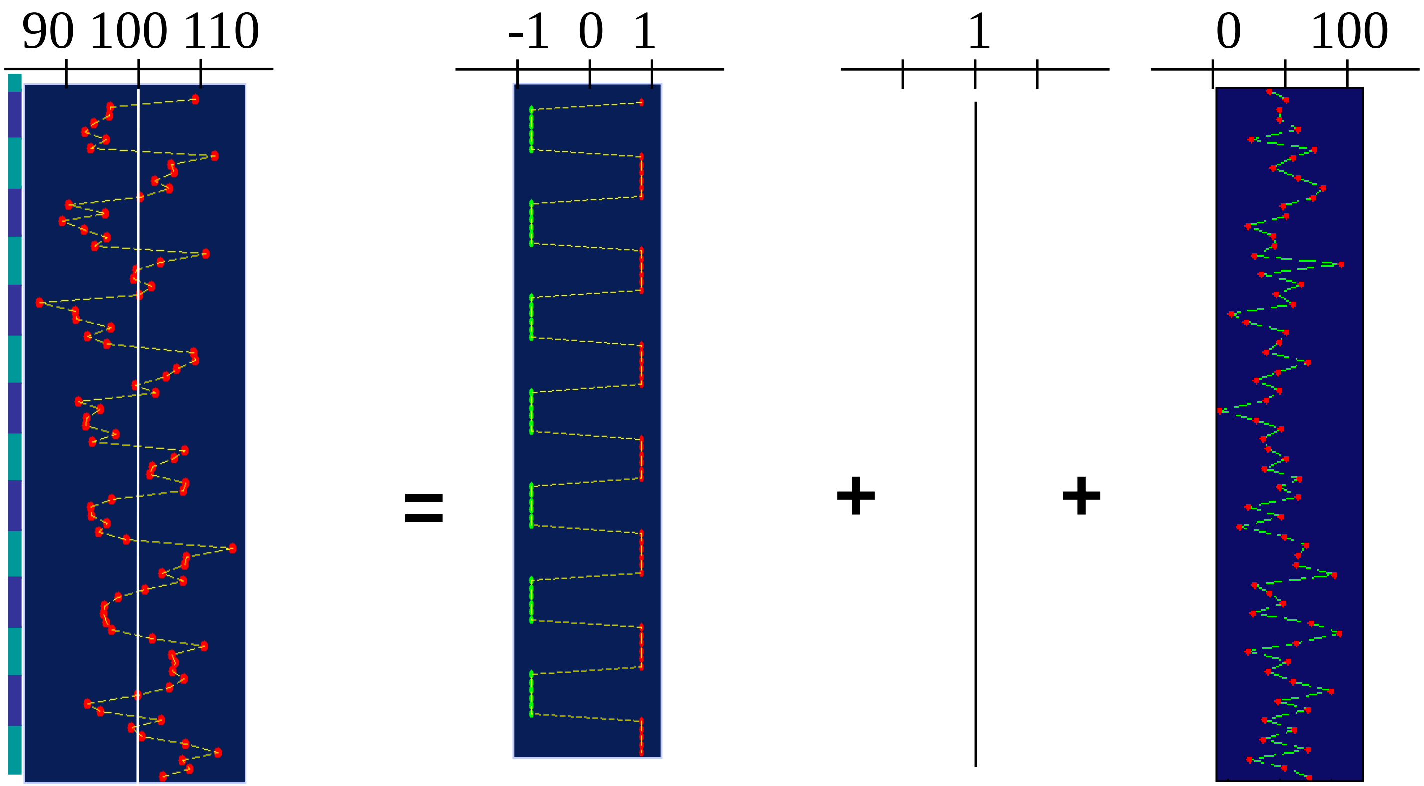
$$\hat{\beta} = (X^T X)^{-1} X^T y$$

The General Linear Model

- Why does OLS give us the optimal betas?
  - Our model should predict our data $\hat{y} = X\hat{\beta}$
  - The error between predicted and measured data $e = y - \hat{y}$
  - Our goal is to minimize the quadratic error by adjusting the betas $e = y - X\hat{\beta}$
  - The sum of squared residuals (RSS) is $e'e$
  - So we can write: $e'e = \left(y - X\hat{\beta}\right)'(y - X\hat{\beta})$
  - And some rewriting: $e'e = y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}$
  - Taking the derivative of this with respect to beta and solving for beta gives us the solution:

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

The General Linear Model



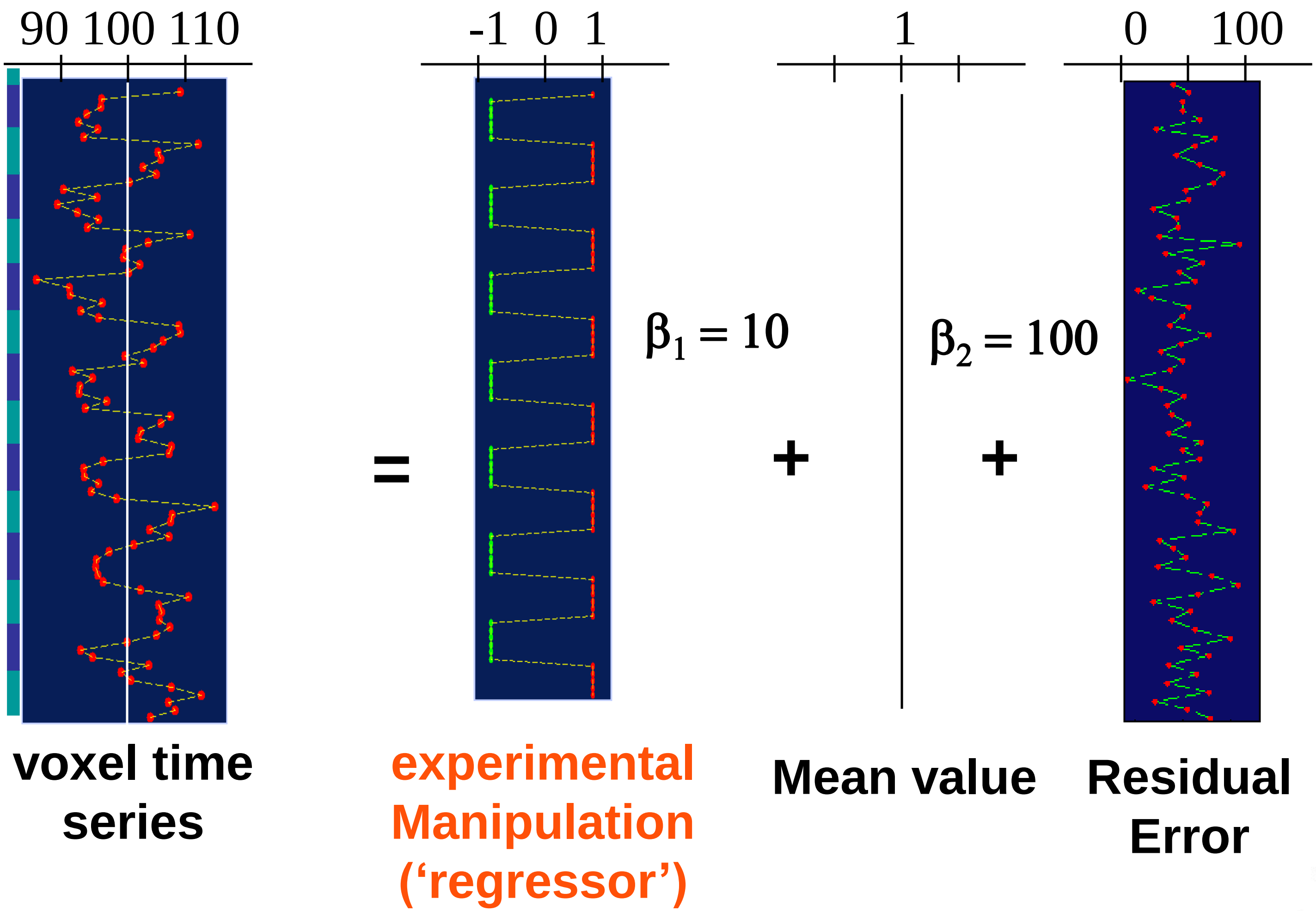
voxel time series

experimental Manipulation (‘regressor’)

Mean value

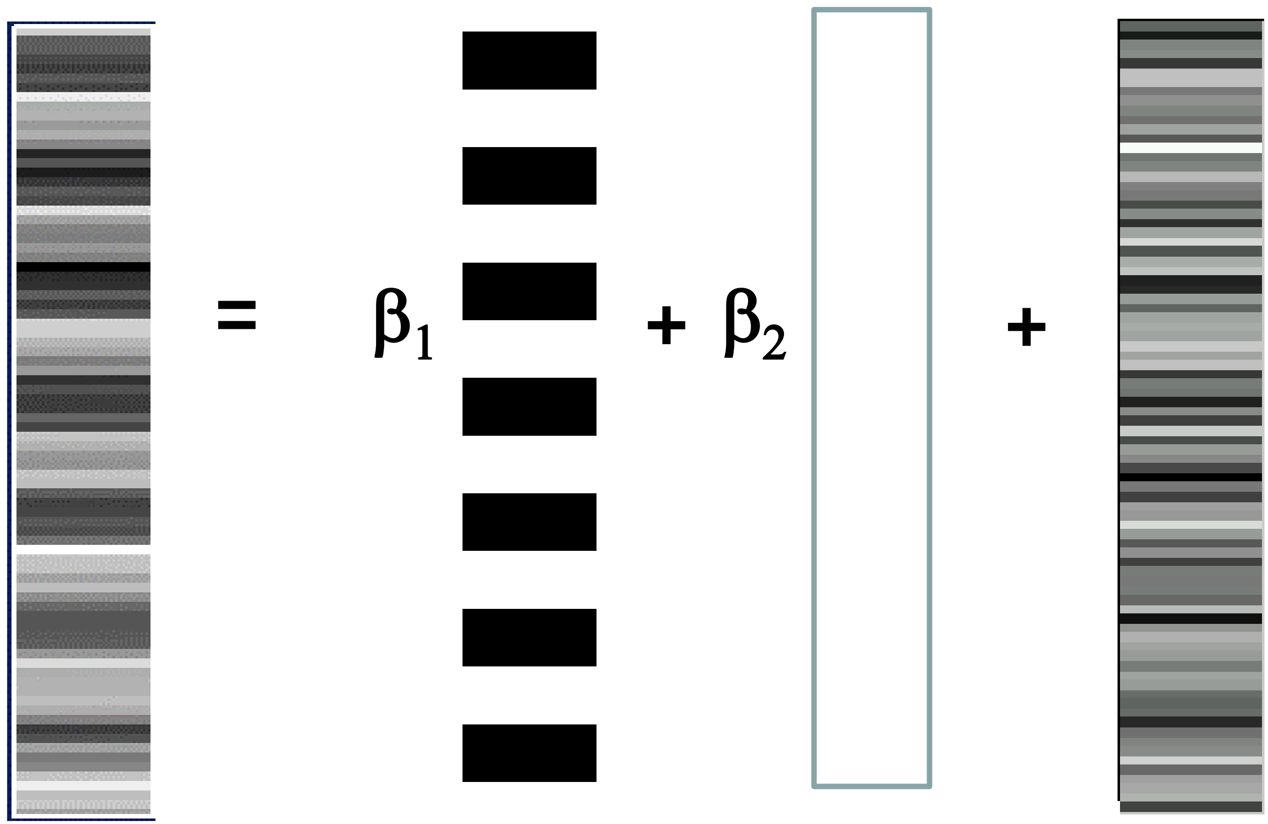
Residual Error

The General Linear Model



The General Linear Model

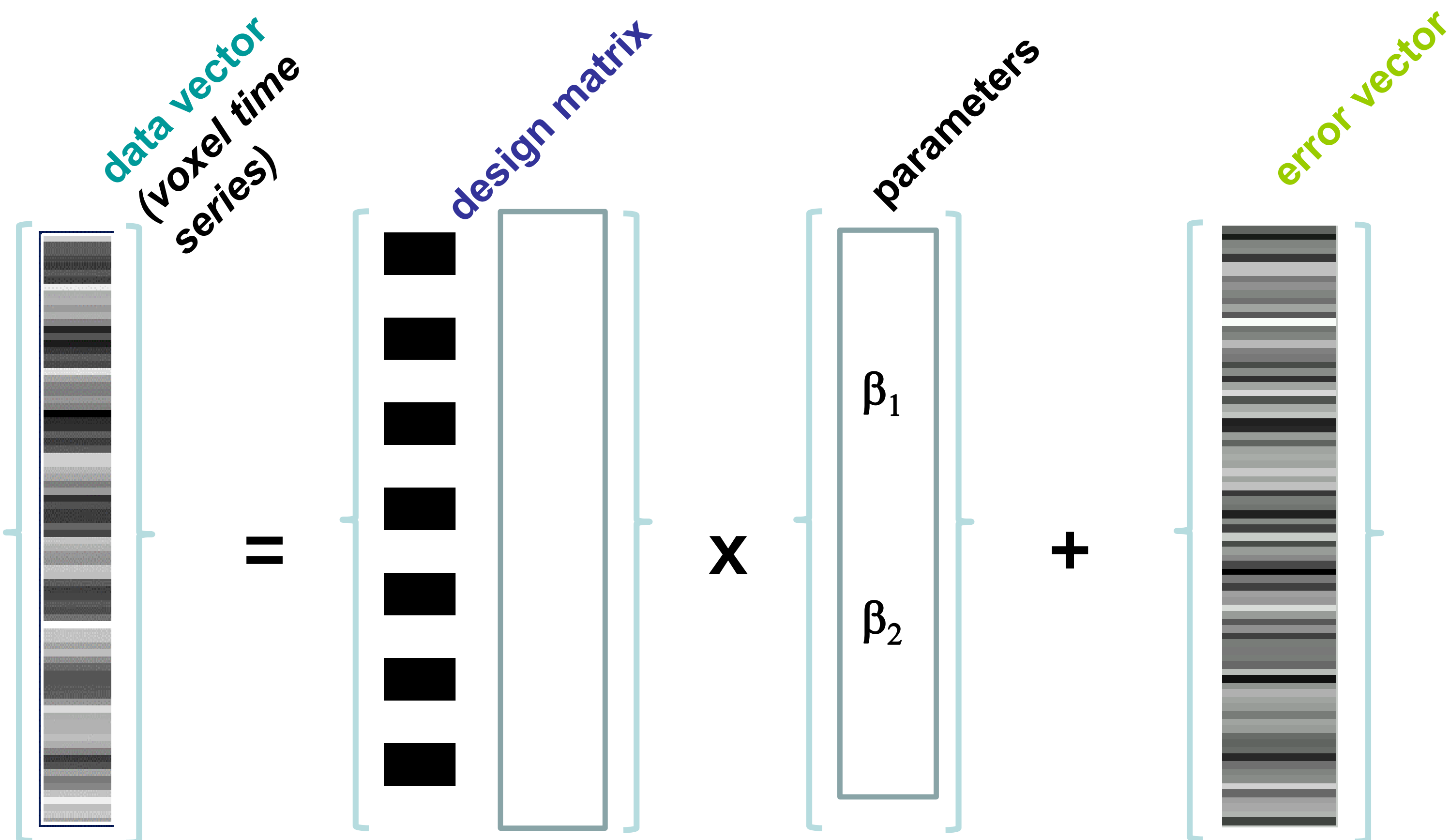
In SPM it looks like this:



$$Y = \beta_1 \times f(t) + \beta_2 \times 1 + \varepsilon_s$$

The General Linear Model

using a matrix notation we get:



$$\underline{Y} = X \times \underline{\beta} + \underline{\varepsilon}$$

The General Linear Model

- We have to solve 3 Problems to make it work in reality

1. The BOLD response is sluggish and we need to take the shape of the response into account
2. Our Scanner is not as stable as we wish – we need to handle low frequency drifts in the data
3. We have to deal with serial correlations in the data

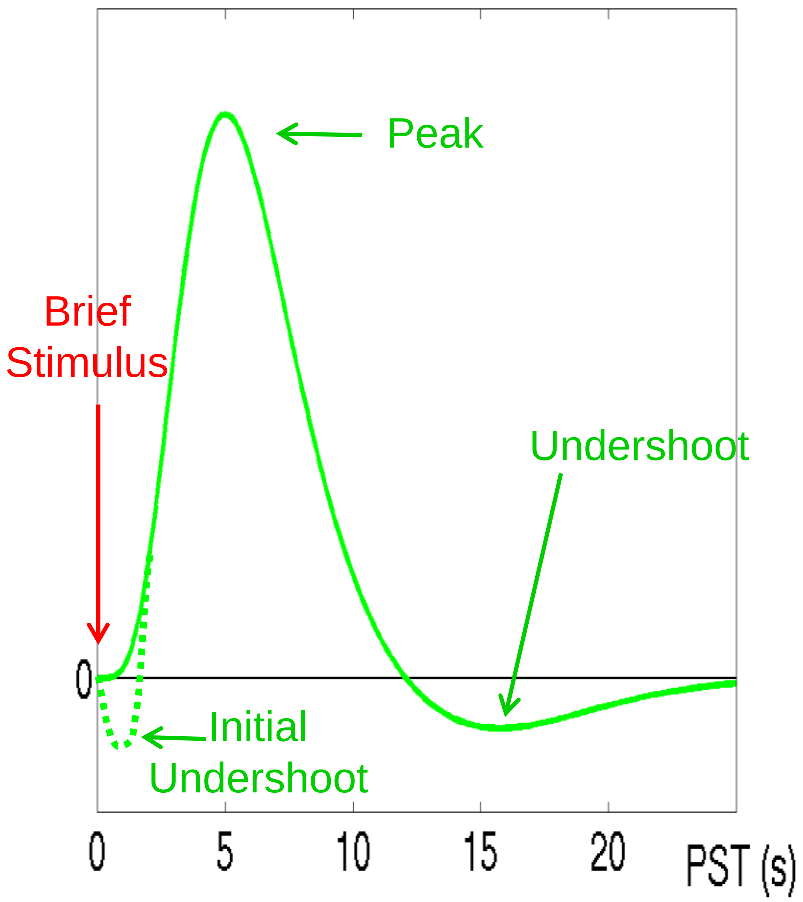
The General Linear Model

- We have to solve 3 Problems to make it work in reality

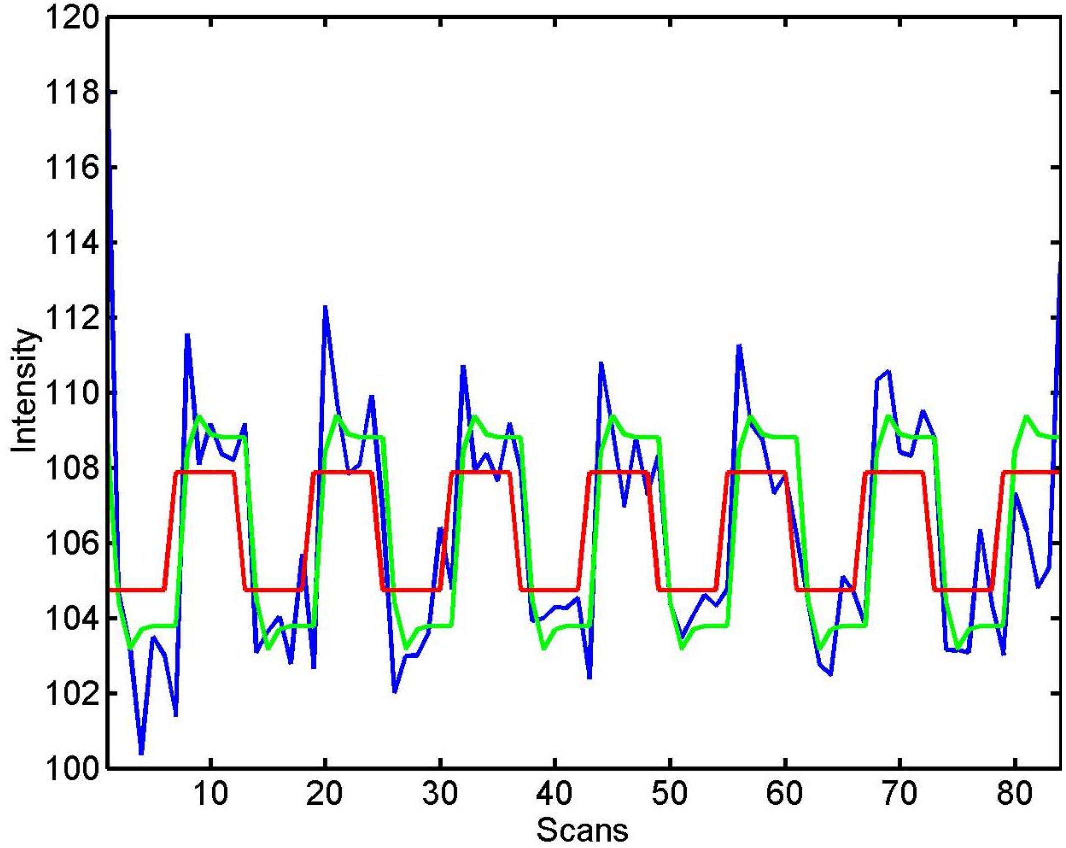
1. The BOLD response is sluggish and we need to take the shape of the response into account
2. Our Scanner is not as stable as we wish – we need to handle low frequency drifts in the data
3. We have to deal with serial correlations in the data

The General Linear Model – Problem 1

The regressors are convolved with the so called ‘‘haemodynamic response function (HRF)



The General Linear Model – Problem 1

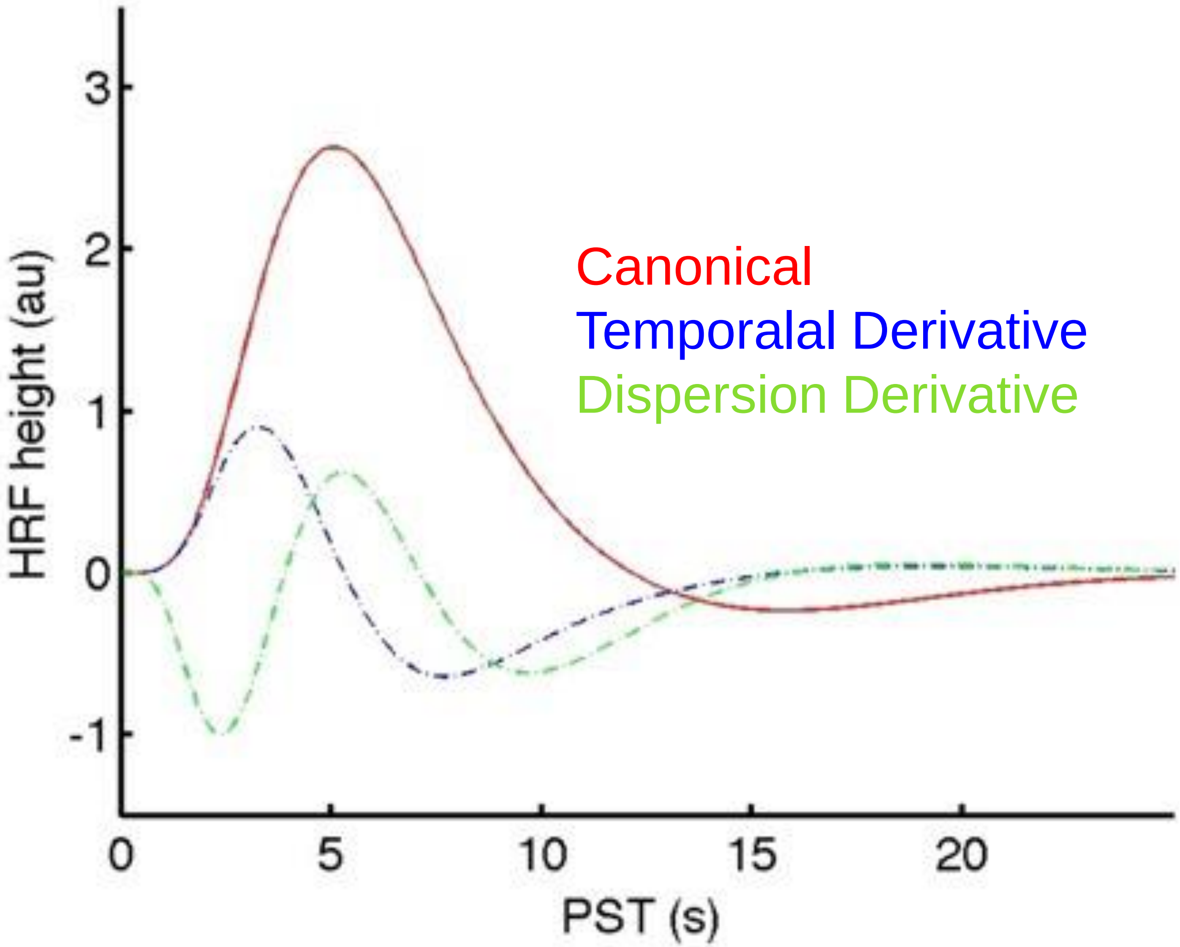


blue = data

red = predicted response, NOT taking into account the HRF

green = predicted response, convolved with HRF

The General Linear Model – Problem 1



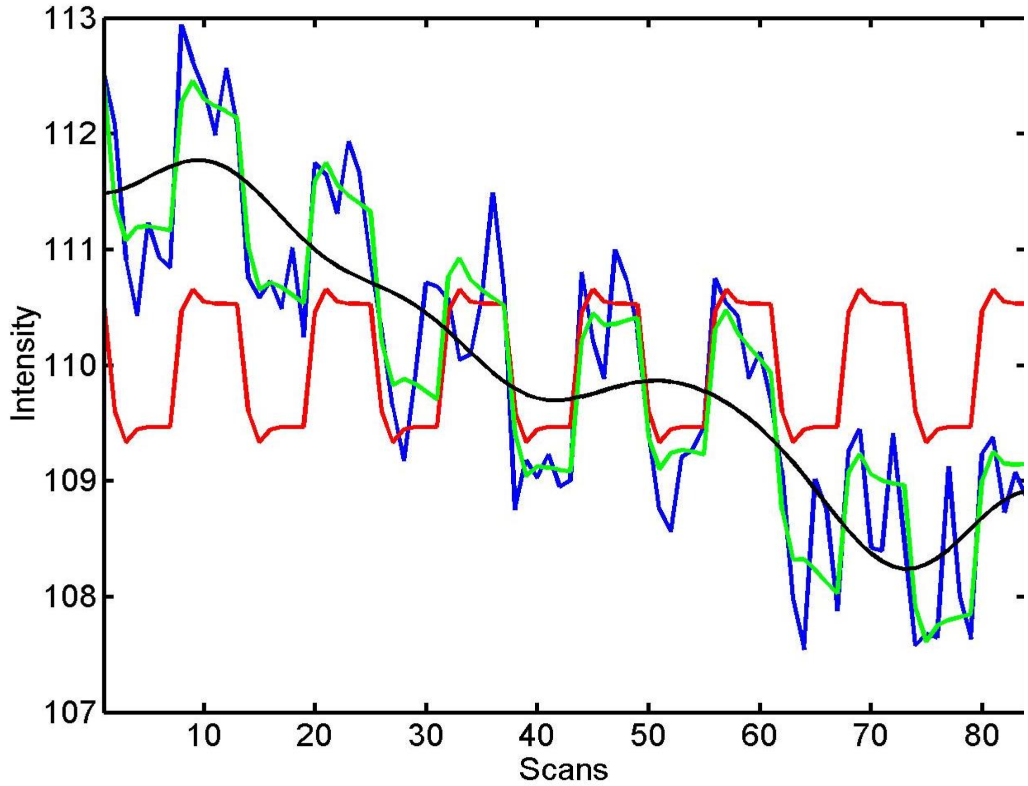
In SPM one usually uses an informed basis set to account for different shapes of the HRF

The General Linear Model

We have to solve 3 Problems to make it work in reality

1. **The BOLD response is sluggish and we need to take the shape of the response into account**
2. **Our Scanner is not as stable as we wish – we need to handle low frequency drifts in the data**
3. **We have to deal with serial correlations in the data**

The General Linear Model



blue = data

red = predicted response, NOT taking into account low-frequency drift

green = predicted response, taking into account low-frequency drift

black = mean + low-frequency drift

The General Linear Model

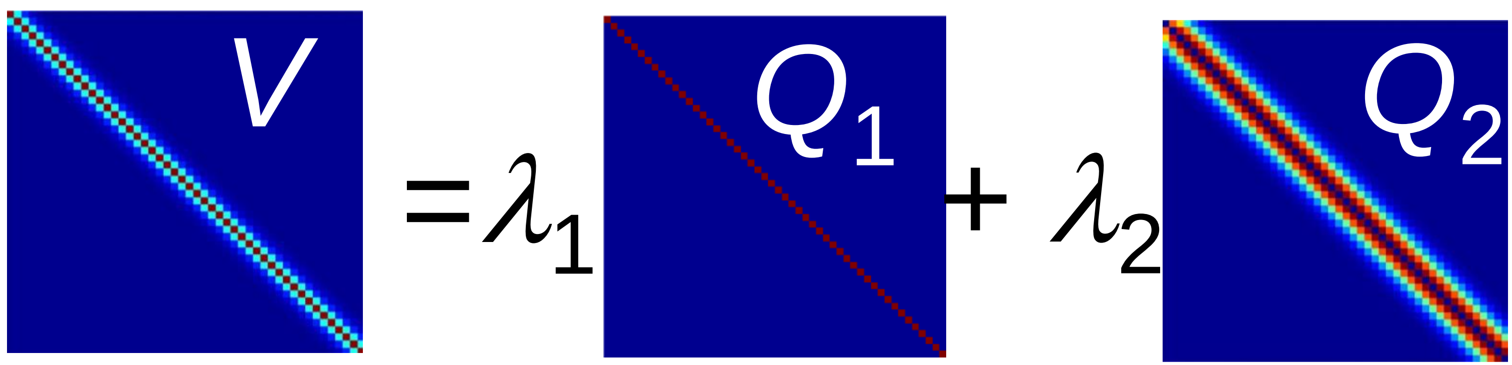
We have to solve 3 Problems to make it work in reality

1. **The BOLD response is sluggish and we need to take the shape of the response into account**
2. **Our Scanner is not as stable as we wish – we need to handle low frequency drifts in the data**
3. **We have to deal with serial correlations in the data**

The General Linear Model

We have to model serial correlations, e.g. by using an autoregressive model of the order one (takes one volume as history into account => AR(1))

We estimate so called hyper parameters during the estimation process using ReML (restricted maximum likelihood)



$$V = \lambda_1 Q_1 + \lambda_2 Q_2$$

The General Linear Model - Summary

•We put in our model regressors that represent how we think the signal is varying

•The regressors are convolved with the so called „haemodynamic response function“ (HRF) to account for the slow BOLD response

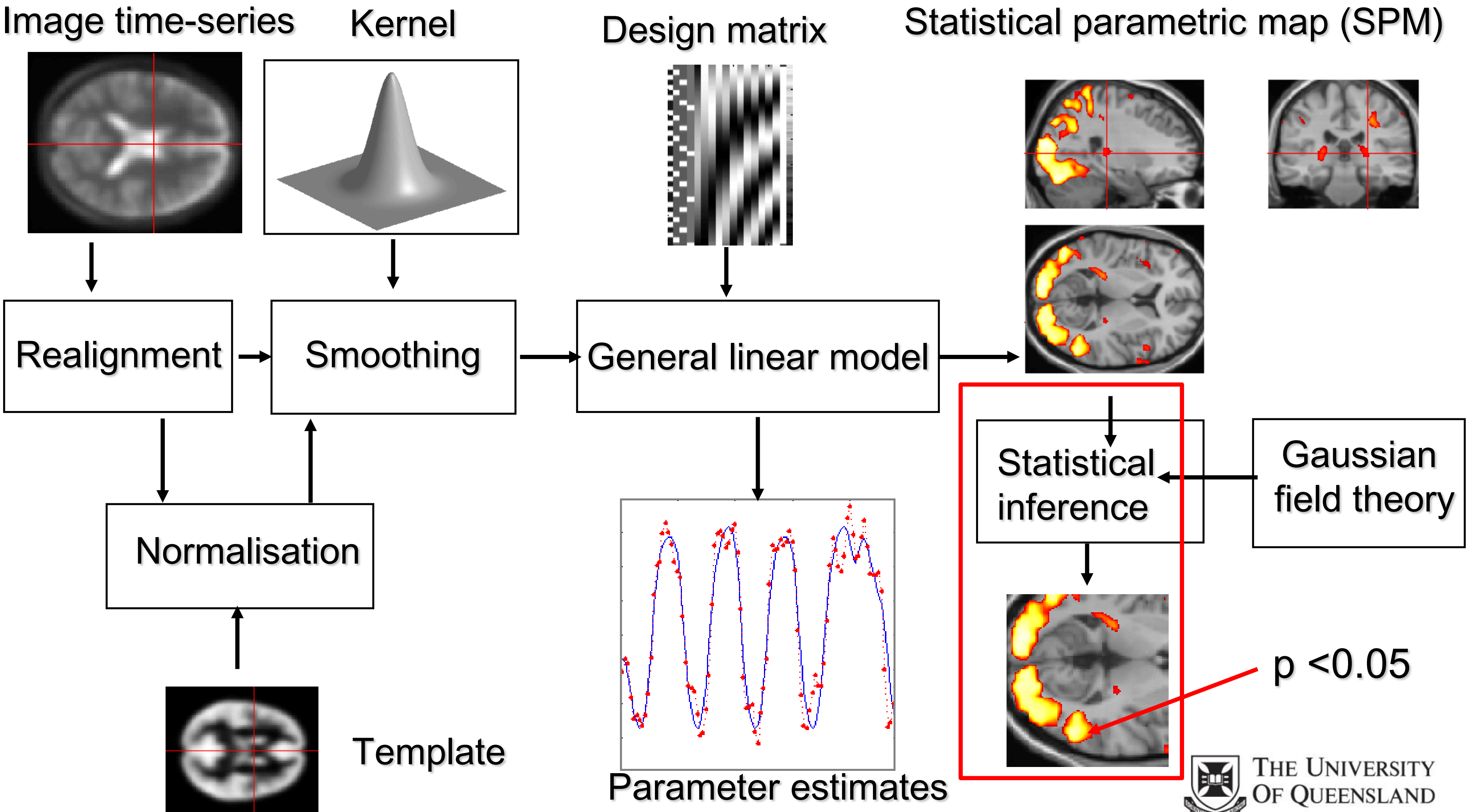
•Coefficients (= parameters or betas) are estimated by minimizing the residuals (= the error)

Part 2

Contents

1. Definitions
2. The General Linear Model
3. **Statistical Inference**
4. How to estimate the efficiency of a design?

SPM Overview



Statistical Inference

We want to test whether our experimental manipulation changed the data significantly -> Let's simply use a T test for that

$$T = \frac{\text{contrast of estimated parameters}}{\sqrt{\text{variance estimate}}}$$

Statistical Inference

A contrast vector *c* selects a specific effect of interest

$$T = \frac{\text{contrast of estimated parameters}}{\sqrt{\text{variance estimate}}} = \frac{c'\beta}{\sqrt{s^2 c' inv(X'X) c}}$$

Statistical Inference

The contrast vector **c** is just a vector with 1s and 0s and it selects the betas we want to investigate

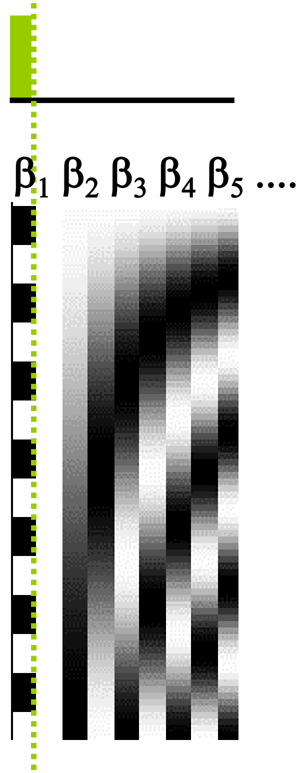
$$= 1\beta_1 + 0\beta_2 + 0\beta_3 + 0\beta_4 + \cdots$$

$$T = \frac{\text{contrast of estimated parameters}}{\sqrt{\text{variance estimate}}} = \frac{c'\beta}{\sqrt{s^2 c' inv(X'X) c}}$$

Statistical Inference

The contrast c is just a vector with 1s and 0s and it weights the betas we want to investigate

c' **=** **1 0 0 0 0 0 0 0**



$$T = \frac{\text{contrast of estimated parameters}}{\sqrt{\text{variance estimate}}} = \frac{c'\beta}{\sqrt{s^2 c' inv(X'X) c}}$$

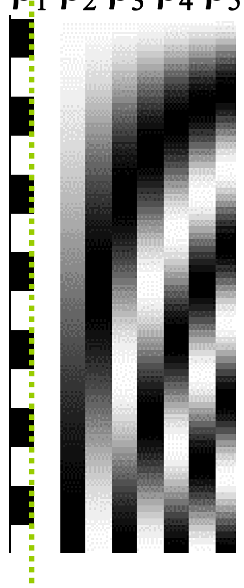
$$c'\beta = 1\beta_1 + 0\beta_2 + 0\beta_3 + 0\beta_4 + \cdots$$

Statistical Inference

$$= 1\beta_1 + 0\beta_2 + 0\beta_3 + 0\beta_4 + \cdots$$

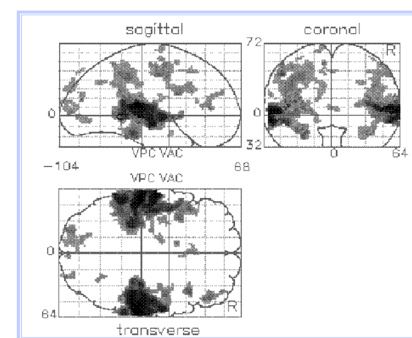
c' = 1 0 0 0 0 0 0 0

$\beta_1\ \beta_2\ \beta_3\ \beta_4\ \beta_5$ ....



$$T = \frac{\text{contrast of estimated parameters}}{\sqrt{\text{variance estimate}}} = \frac{c'\beta}{\sqrt{s^2 c' inv(X'X) c}}$$

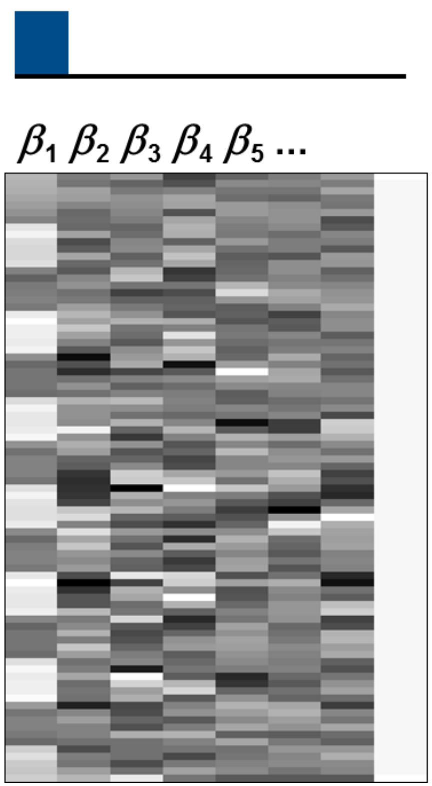
- Doing this for every single voxel in the brain will result in an image
- in this image, voxels are darker if our model fits the data very well (explains a lot of the data)
- We will do ***a lot*** of tests and we need to correct for that later

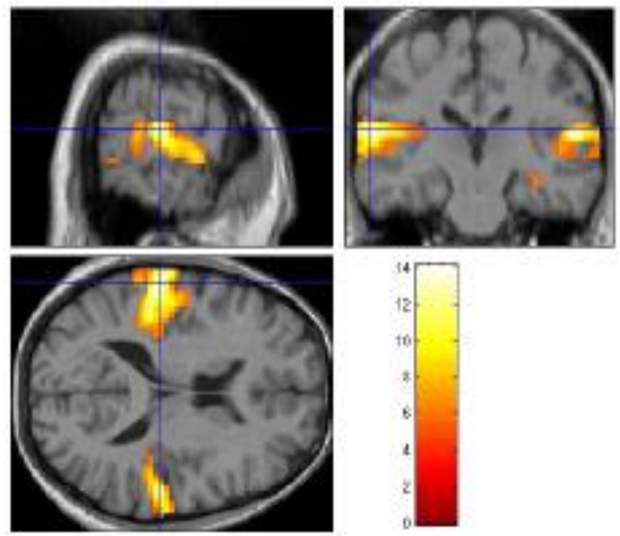


Statistical Inference – an example

We want to know in which voxels of the brain we cause an increase in BOLD signal when our subject listens to words

$c^T = 1\ 0\ 0\ 0\ 0\ 0\ 0$





Threshold T = 3.2057 {p<0.001}

	($Z_{\equiv}$)	p uncorrected	Mm	mm	mm
13.94	**Inf**	**0.000**	**-63**	**-27**	**15**
12.04	Inf	0.000	-48	-33	12
11.82	Inf	0.000	-66	-21	6
13.72	**Inf**	**0.000**	**57**	**-21**	**12**
12.29	Inf	0.000	63	-12	-3
9.89	7.83	0.000	57	-39	6
7.39	**6.36**	**0.000**	**36**	**-30**	**-15**
6.84	**5.99**	**0.000**	**51**	**0**	**48**
6.36	**5.65**	**0.000**	**-63**	**-54**	**-3**

Statistical Inference - Summary

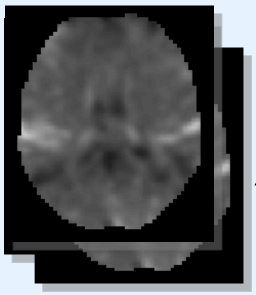
- **Contrast c = linear combination of parameters: c' β**

- **this means we select the betas we want to look at**

$$T = \frac{\text{contrast of estimated parameters}}{\sqrt{\text{variance estimate}}} = \frac{c'\beta}{\sqrt{s^2 c' inv(X'X) c}}$$

Statistical Inference – Summary

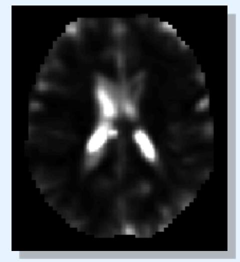
This is how the output looks in SPM



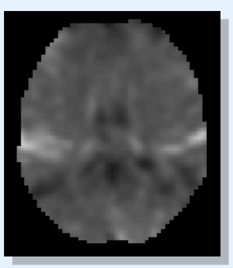
beta_???? images

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

ResMS image

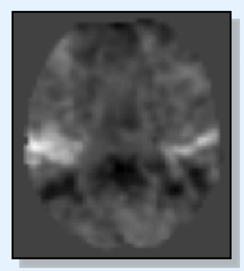


$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{N - p}$$



con_???? image

$$c^T \hat{\beta}$$



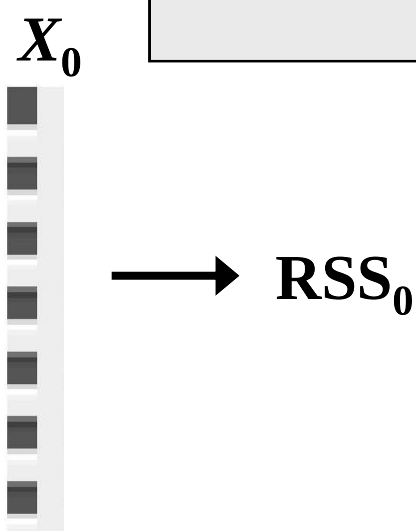
spmT_???? image

SPM{t}

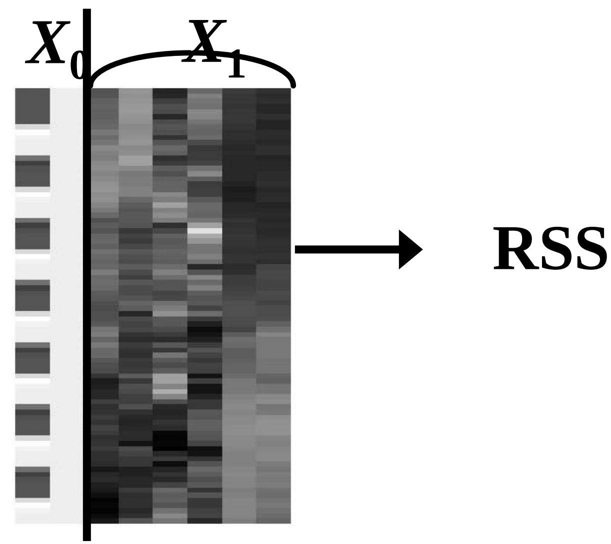
Statistical Inference – F-tests

- **F tests can be used to compare different models**
- **The test statistic is the ratio of explained variability and unexplained variability (error)**

$$F \propto \frac{RSS_0 - RSS}{RSS}$$



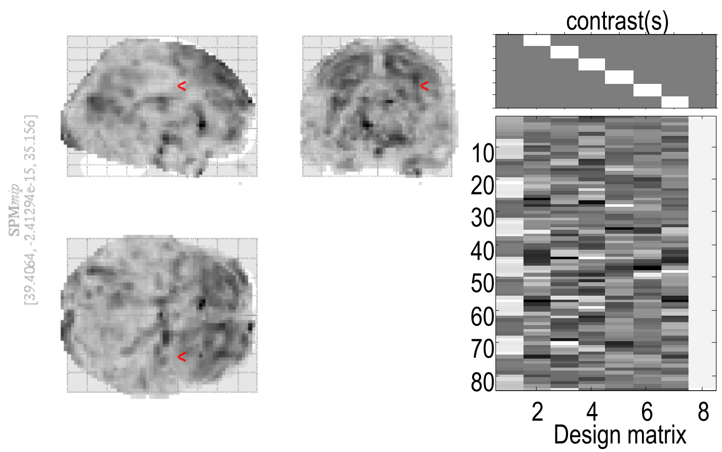
Reduced model?

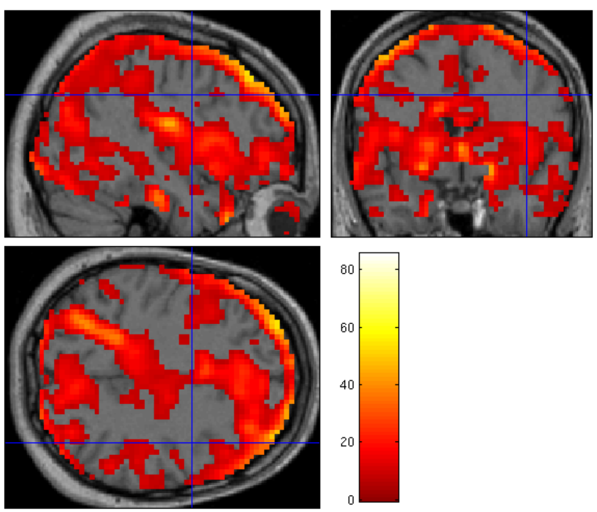


Or Full model ?

Statistical Inference – F-tests

Example: Should we include the realignment parameters in our model (full model) or can we ignore them (reduced model)





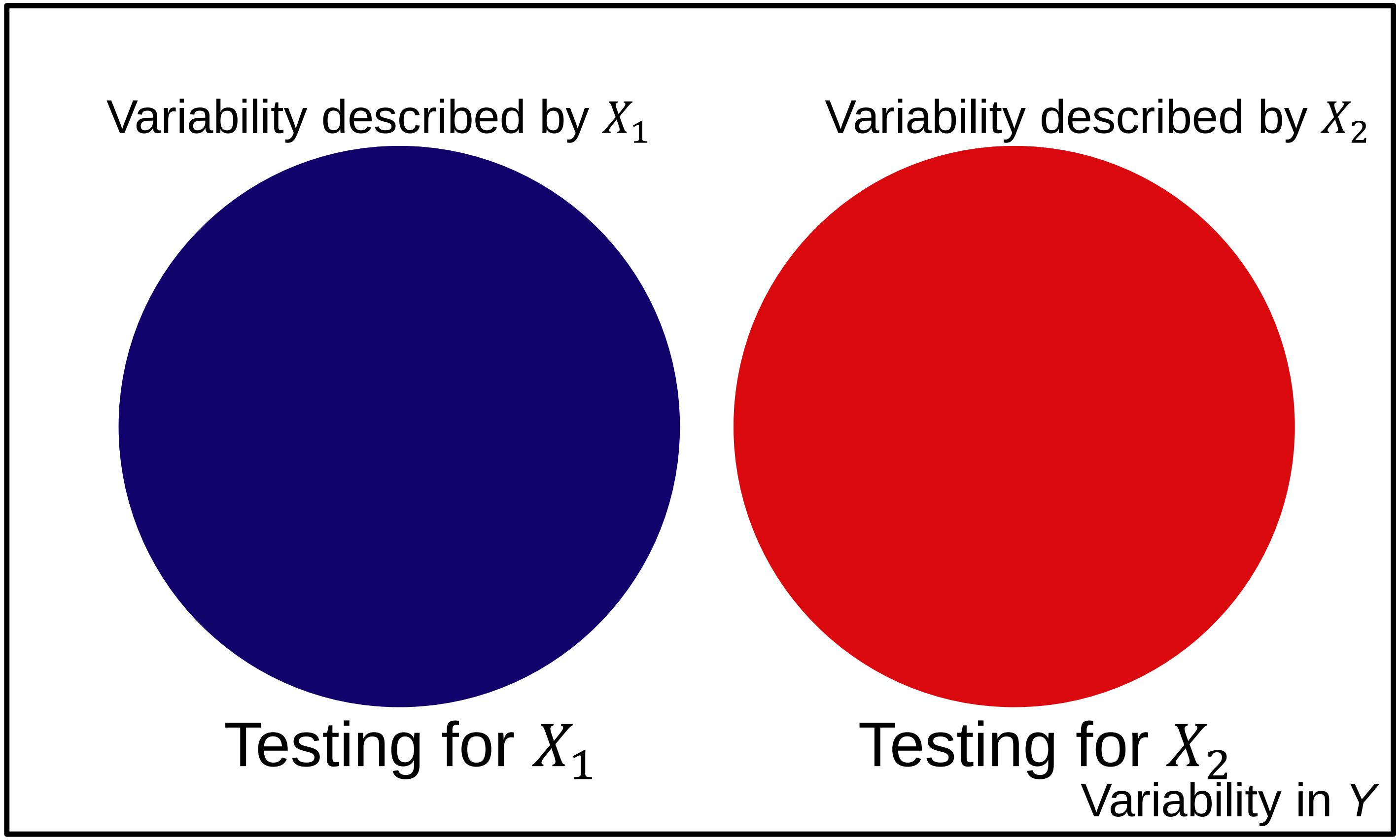
⇨ Yes, we should use the full model, because the regressors explain noise in regions we are interested in

Orthogonality of regressors

What happens if my regressors are partly explaining the same and are not orthogonal to each other?

Orthogonality of regressors

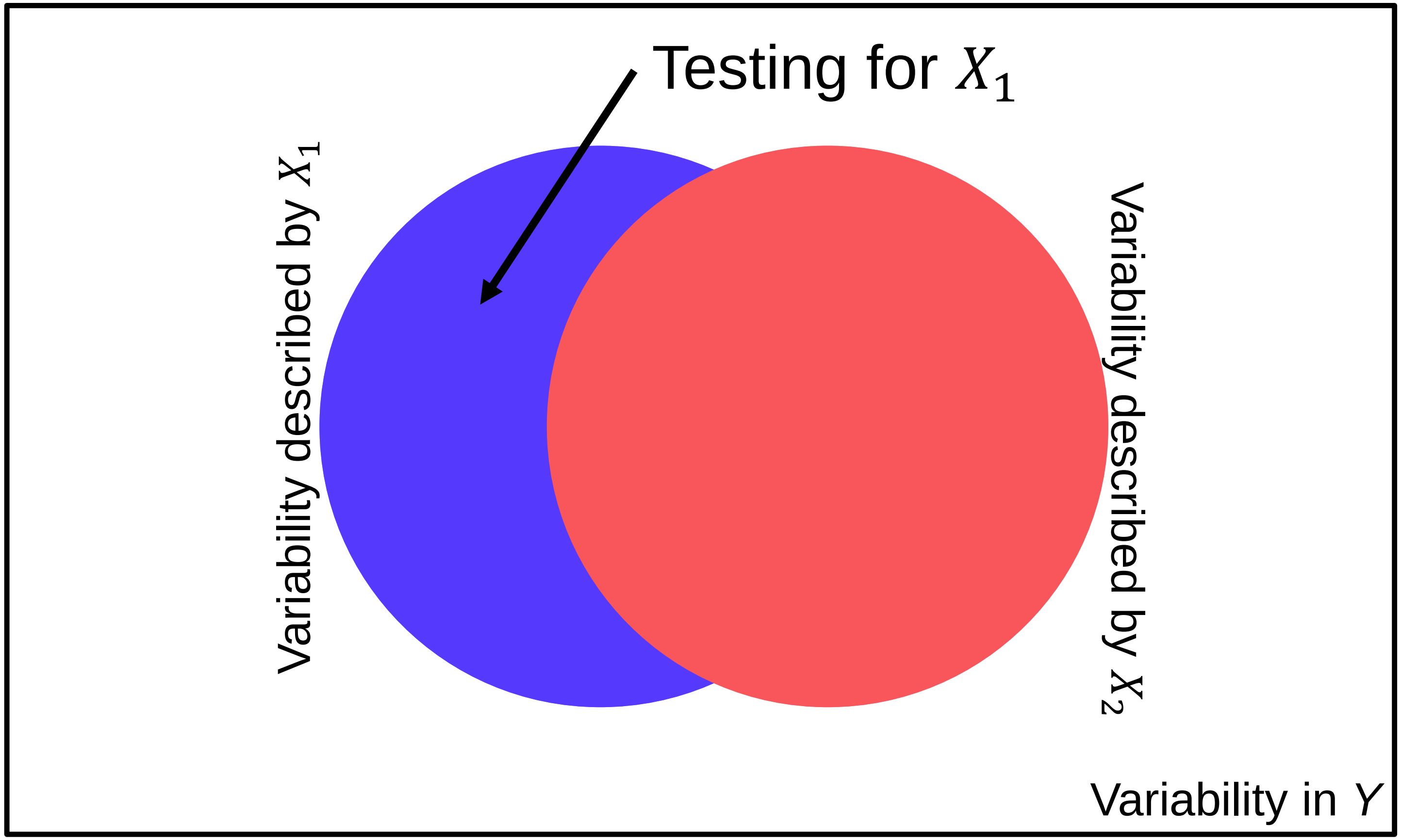
- **These regressors are orthogonal to each other**
- **No shared variance**



Orthogonality of regressors

These regressors are correlated to each other

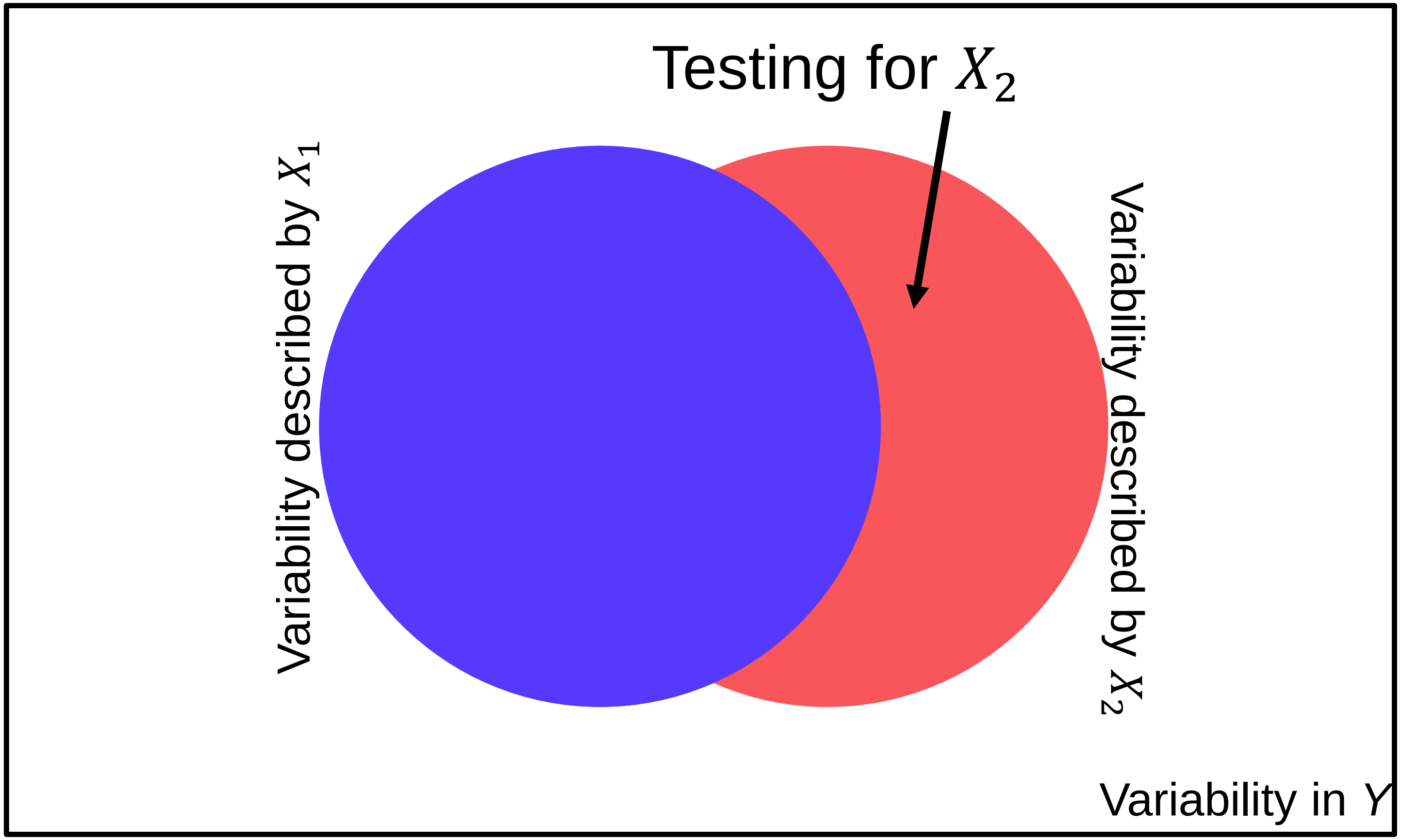
We only explain the variability which is not shared between them!



Orthogonality of regressors

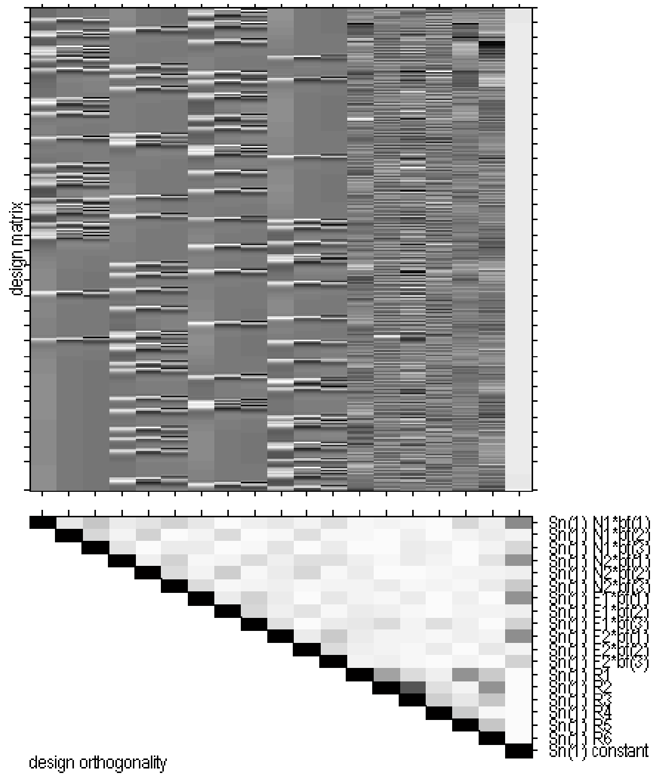
These regressors are correlated to each other

We only explain the variability which is not shared between them!



Orthogonality of regressors

The degree of correlation is plotted in SPM



Measure : abs. value of cosine of angle between columns of design matrix
Scale : black - colinear (cos=+1/-1)
white - orthogonal (cos=0)
gray - not orthogonal or colinear

- The more overlap the regressors have the less variance can be explained
- In a contrast the regressors of interest should be as little correlated as possible!

Contents

1. Definitions
2. The General Linear Model
3. Statistical Inference
4. How to estimate the efficiency of a design?

How to estimate the efficiency of a design?

•This question can only be answered concerning a specific contrast!

⇨ The same design can be efficient for one contrast and inefficient for another!!!

•How?

- The aim is to minimize the standard error of a t-contrast
- This can be calculated using the Design Matrix X, and a contrast vector c
- Design Efficiency = 1/(c*inv(X‘*X)*c‘);

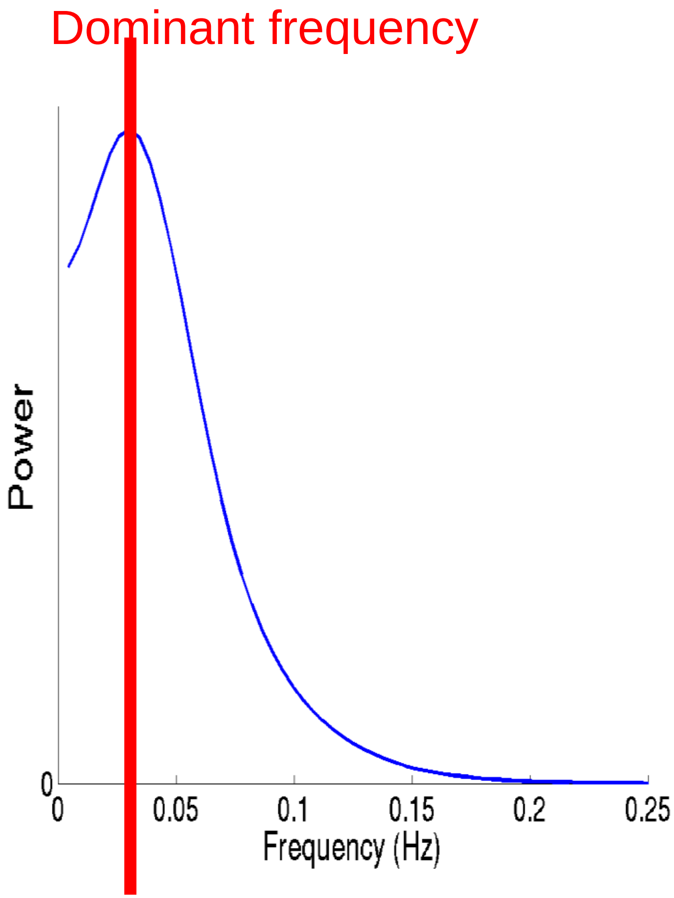
How can we optimize a design?

$$T = \frac{\text{contrast of estimated parameters}}{\sqrt{\text{variance estimate}}} = \frac{c'\beta}{var(c'\beta)} = \frac{c'\beta}{\sqrt{s^2 c' inv(X'X)c}}$$

- for maximal T we want minimal contrast variability (Friston 1999)
- This can be calculated using the Design Matrix X, and a contrast vector c (we assume noise variance ***s*²** is unaffected by change in Design Matrix X)
- Design Efficiency = 1/(c * inv(X‘ * X) * c‘);
- **The design efficiency values are relative and not absolute values and can only be compared in similar designs (e.g. same experimental length)!**

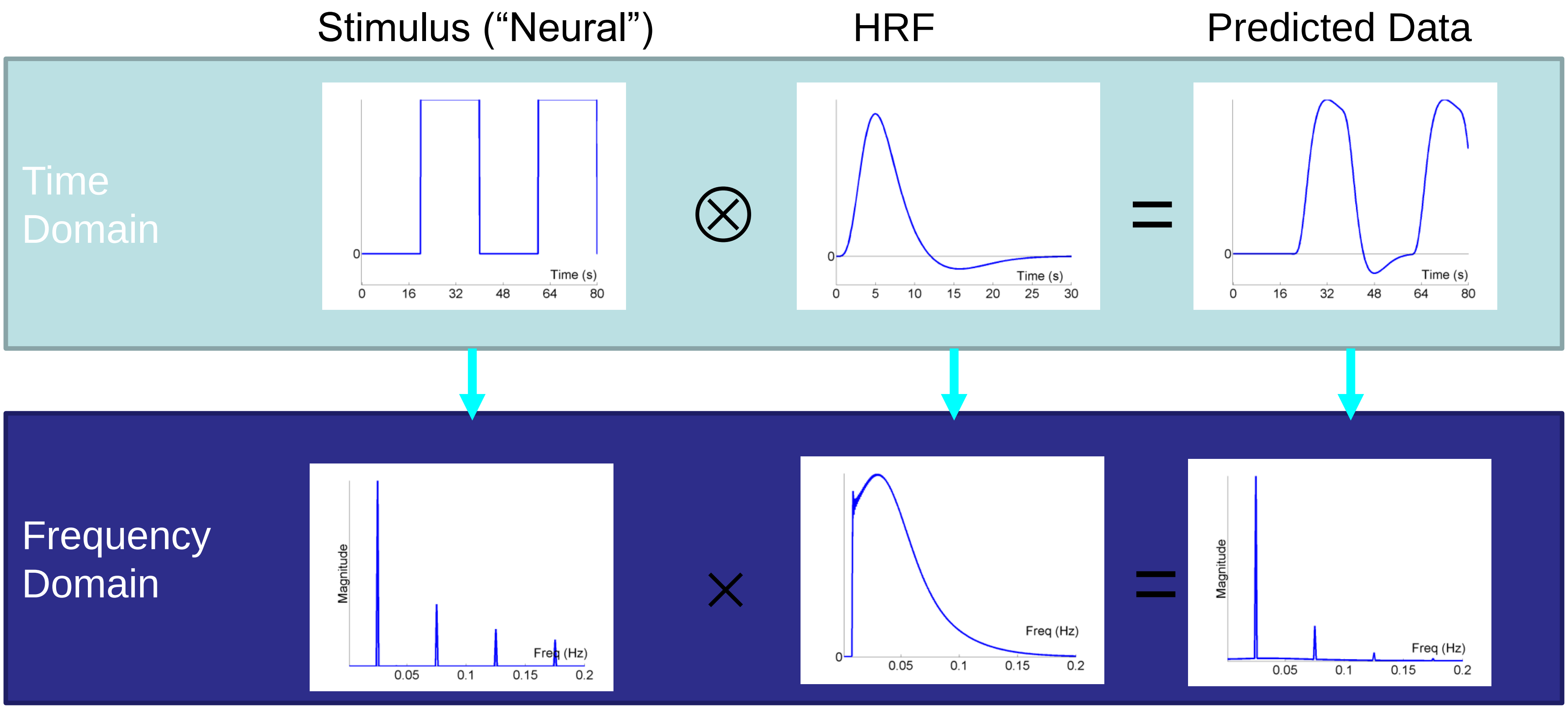
How can we optimize a design?

- Convolving regressors with the HRF can be seen as a filter (Josephs & Henson, 1999)
- We want to maximise the signal passed by this filter
- Dominant frequency of canonical HRF is ~0.04 Hz
- The most efficient design is a sinusoidal modulation of neural activity with period ~25s (e.g., boxcar with 12.5s on/ 12.5s off)



How can we optimize a design?

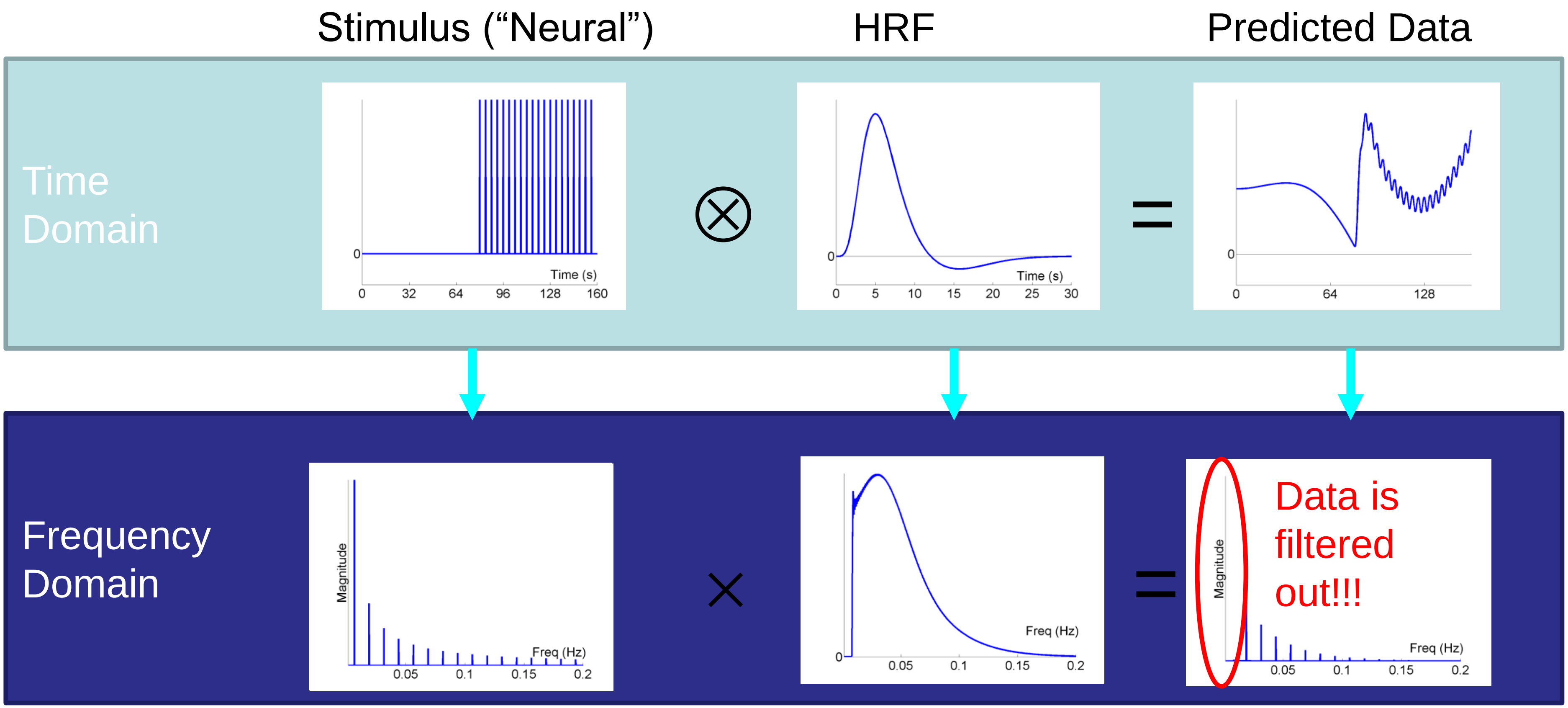
Example – Blocked (20s); SOA = small



quite efficient!

How can we optimize a design?

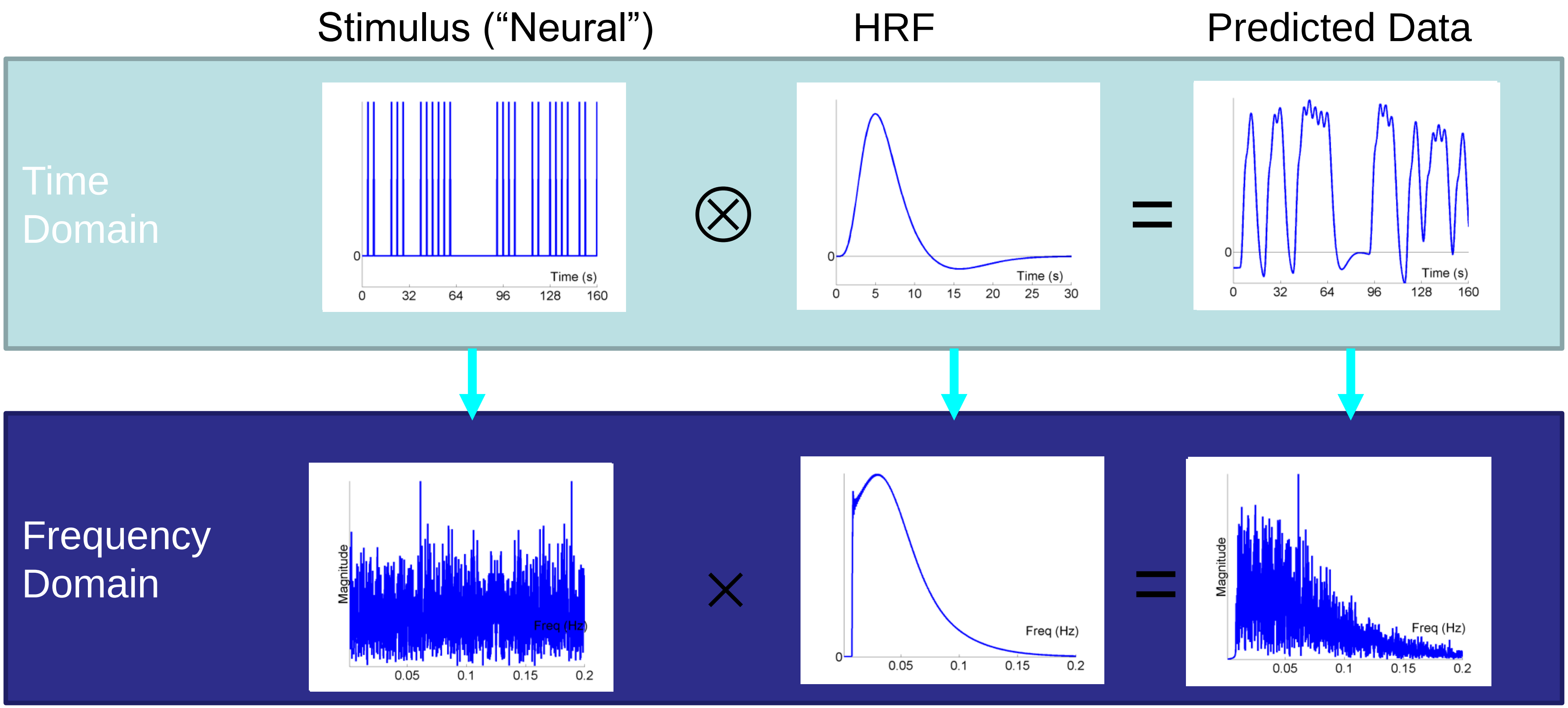
Example – how not to do it: Blocked (80s); SOA = 4s



Never have too long blocks!

How can we optimize a design?

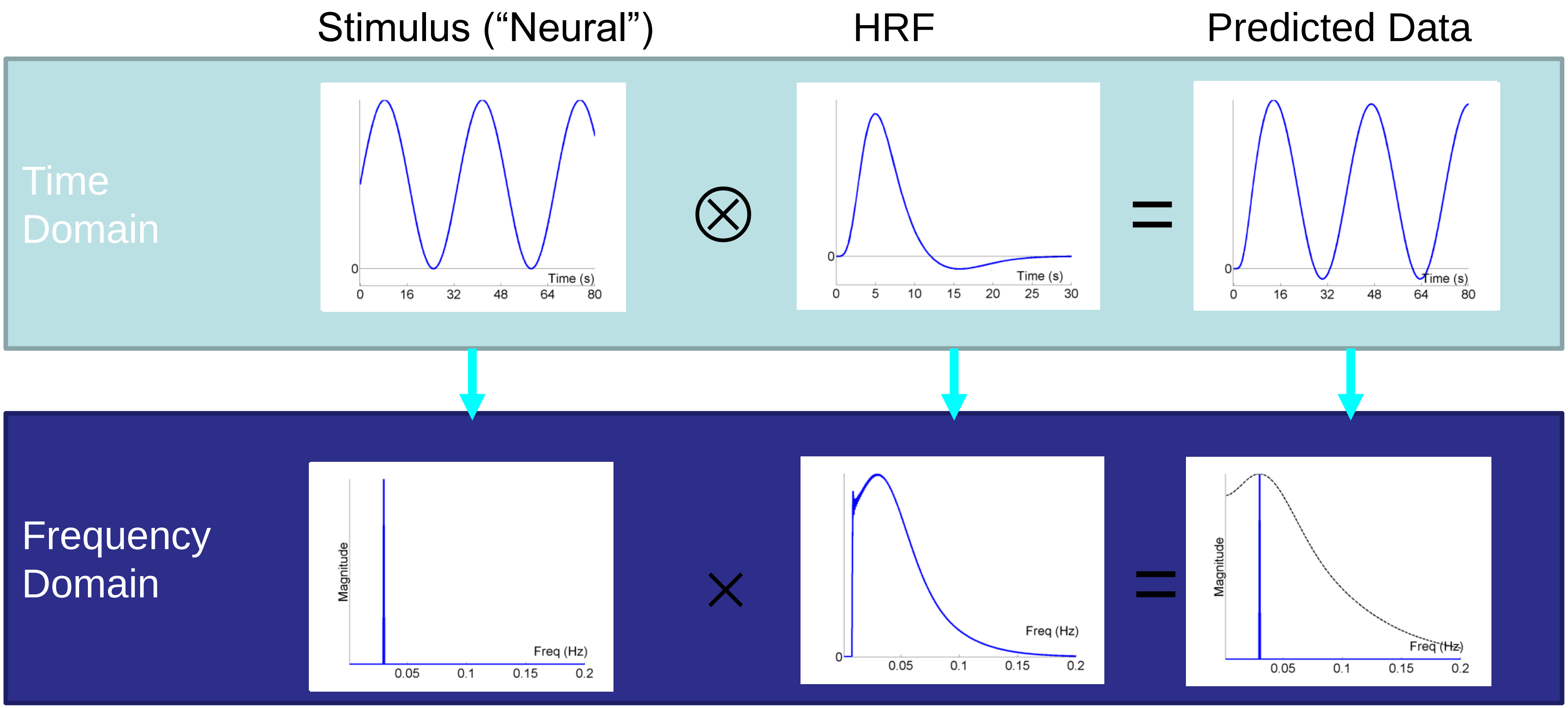
Example – Randomized: SOAmin = 4s



Randomised design spreads power over frequencies

How can we optimize a design?

Example – The perfect design: sinusoidal with f = 1/33s



The sinusoidal places the energy in the frequency domain at exactly the right position

How can we optimize a design?

Blocked designs are generally most efficient with short Stimulus Onset Asynchronys (SOAs)

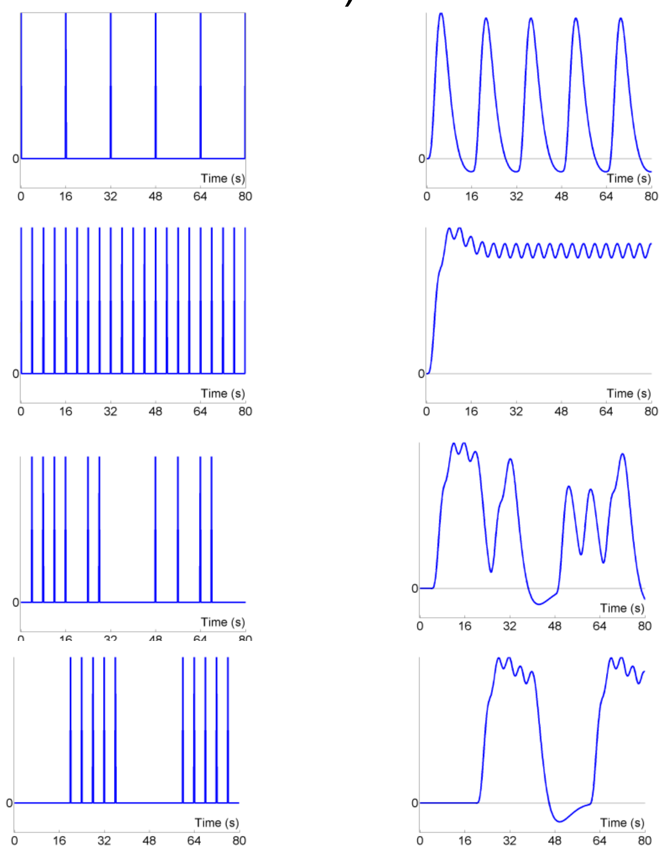
Stimulus (“Neural”) Predicted Data (after convolution with HRF)

Fixed SOA 16s → *Not very efficient…*

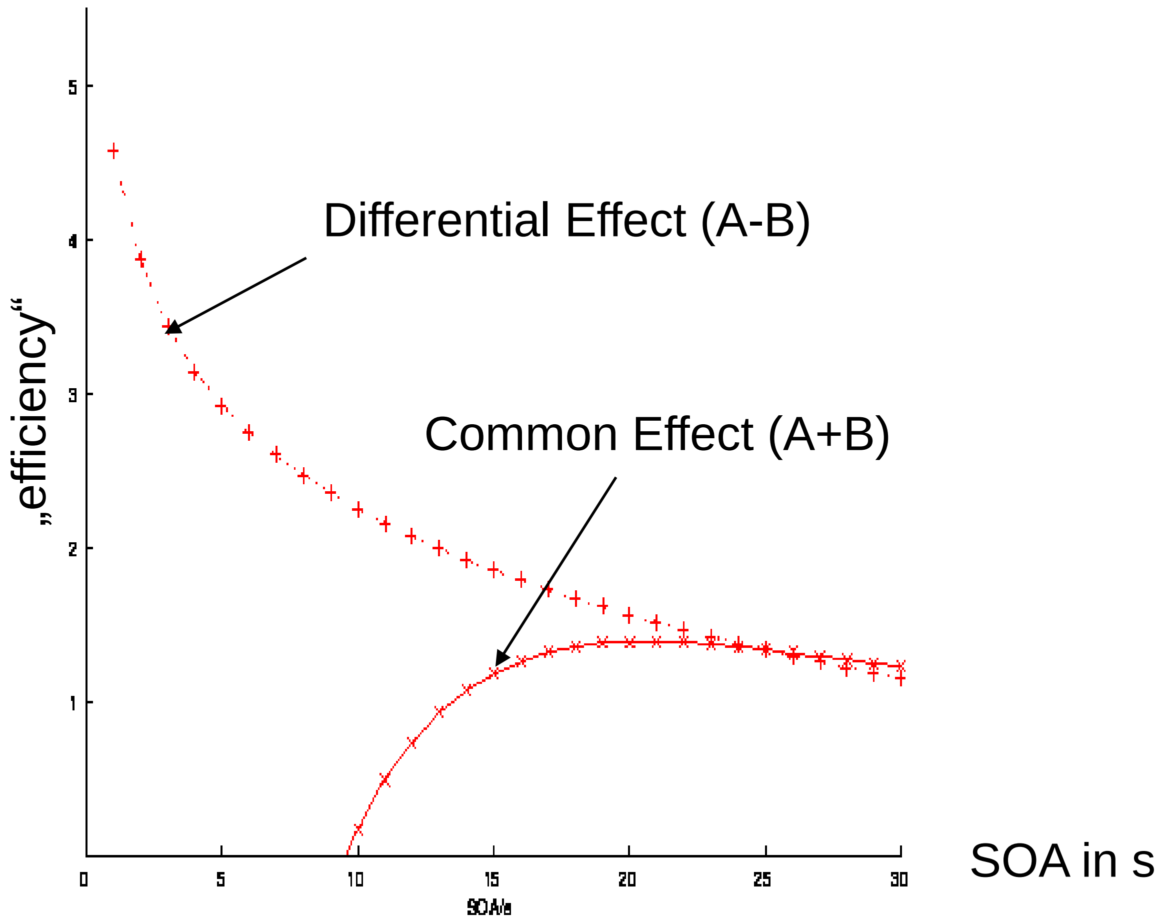
Fixed SOA 4s → *Very Inefficient…*

Randomized SOA → *More Efficient*

Blocked SOA 4s → *Even More Efficient*



Efficiency for multiple event types



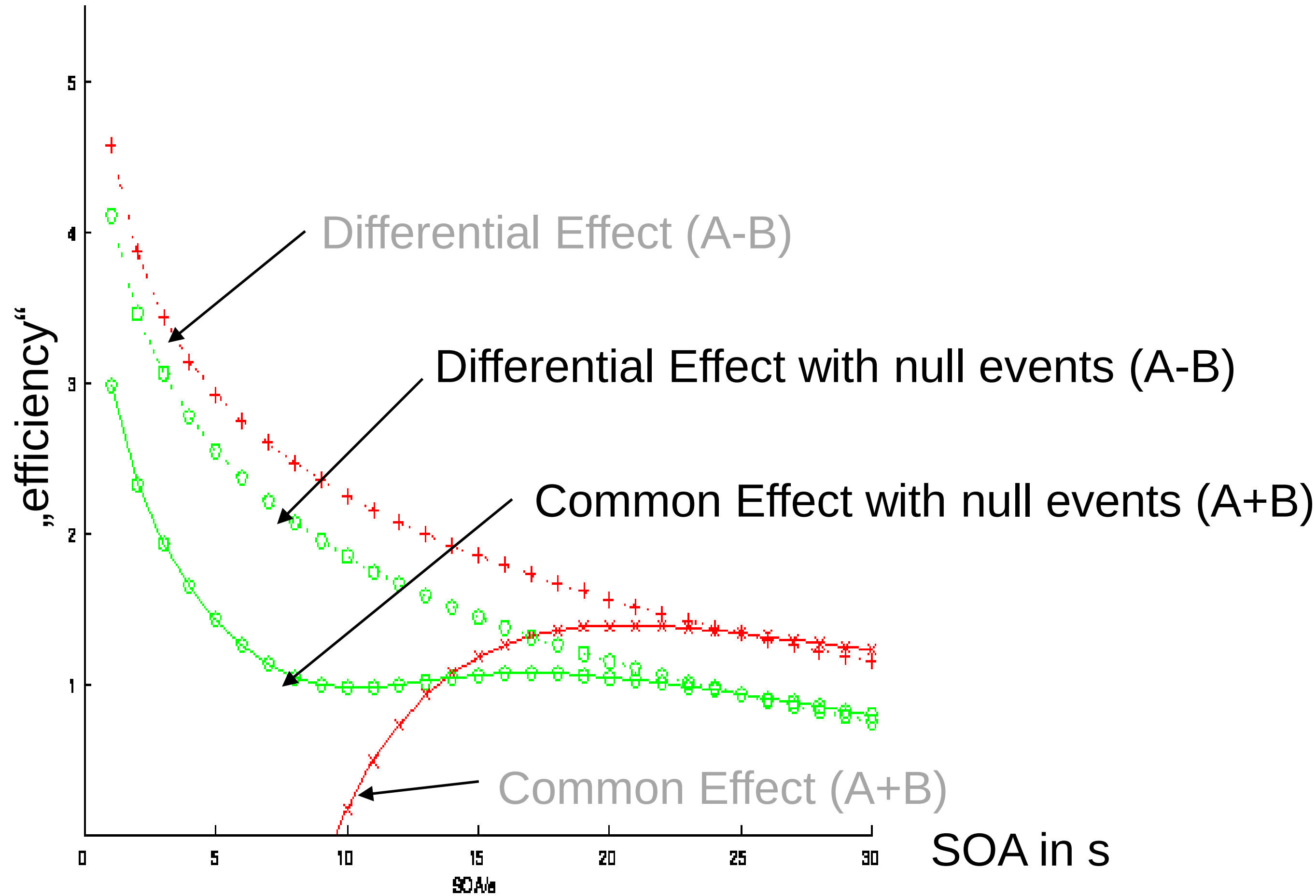
If you are interested in differential effects only, it is ok to use short SOAs

If you are interested in differential **and** common effects then you could use long SOAs or …

e.g.: ABABABABABAB

Efficiency for multiple event types

Using null events you get a design which is efficient for differential **and** common effects at **short** SOAs



e.g.: AB-BAA--B---ABB

Efficiency - Detection versus Estimation

- Detection power
  - = Detect a response
  - maximal in blocked designs

- Estimation efficiency
  - = Estimate the shape of a response
  - maximal in randomised designs

Summary

•An optimal design for one contrast may not be optimal for another (it is crucial to know your hypotheses BEFORE you design the experiment)

•With randomized designs, optimal SOA for differential effect (A-B) is minimal SOA (assuming no saturation), whereas optimal SOA for main effect (A+B) is 16-20s

•Inclusion of null events improves efficiency for main effect at short SOAs (at cost of efficiency for differential effects)

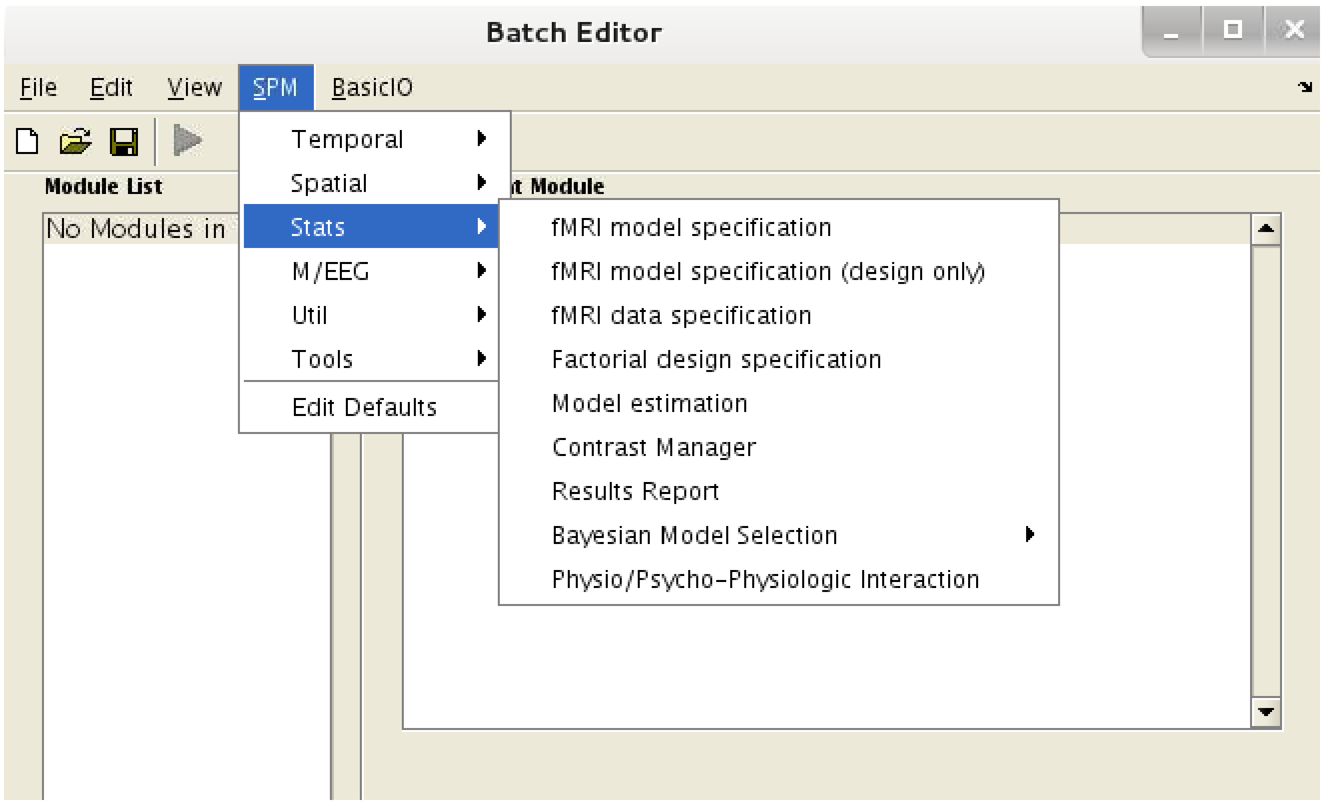
•If order constrained, intermediate SOAs (5-20s) can be optimal

•If SOA constrained, pseudorandomised designs can be optimal

•General advice: Keep the subject as busy as possible with your task

Hands on / Homework ☺

1. **Open the Batch Editor in SPM and Select**
    - **SPM -> Stats -> fMRI model specification (design only)**

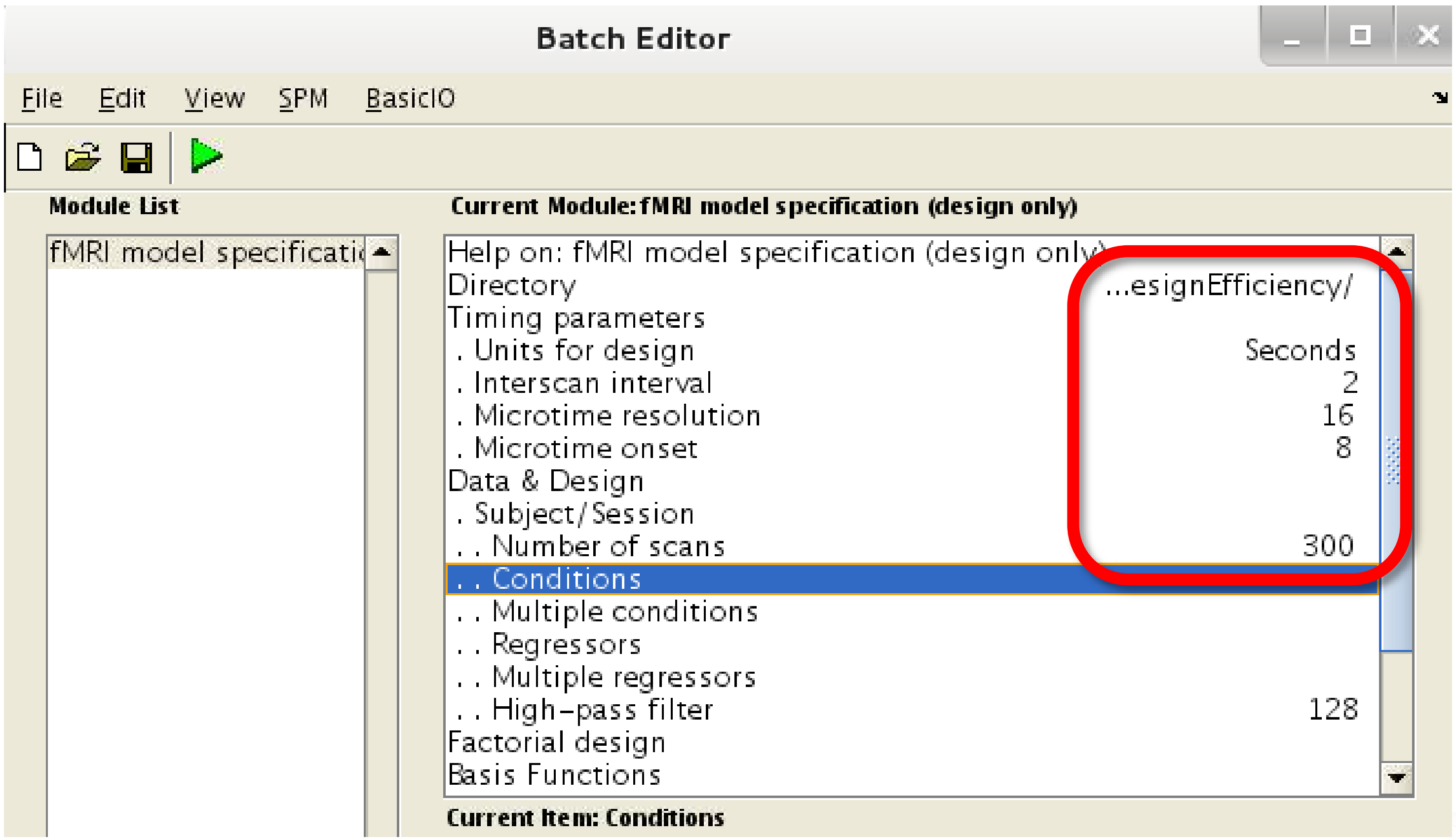


Hands on

- We want to investigate a simple On / Off paradigm
- 20 s on
- 20 s off

Hands on

1. Select a directory where to store the SPM.mat file
2. Enter parameters like shown in the image:



Hands on

1. Enter Parameters for condition A
- Name: A
- Onsets: 0:40:560 (This creates a vector from 0 to 560 in steps of 40)
- Durations: ones(15,1) * 20 (This creates a vector of 15 ones and multiplies it by 20 -> we end up with a vector of 15 twenties)

2. Enter Parameters for condition B
- Name: B
- Onsets: 20:40:600
- Durations: ones(15,1) * 20

- Check that what you have entered makes sense
- Save your design
- Hit the run button ☺

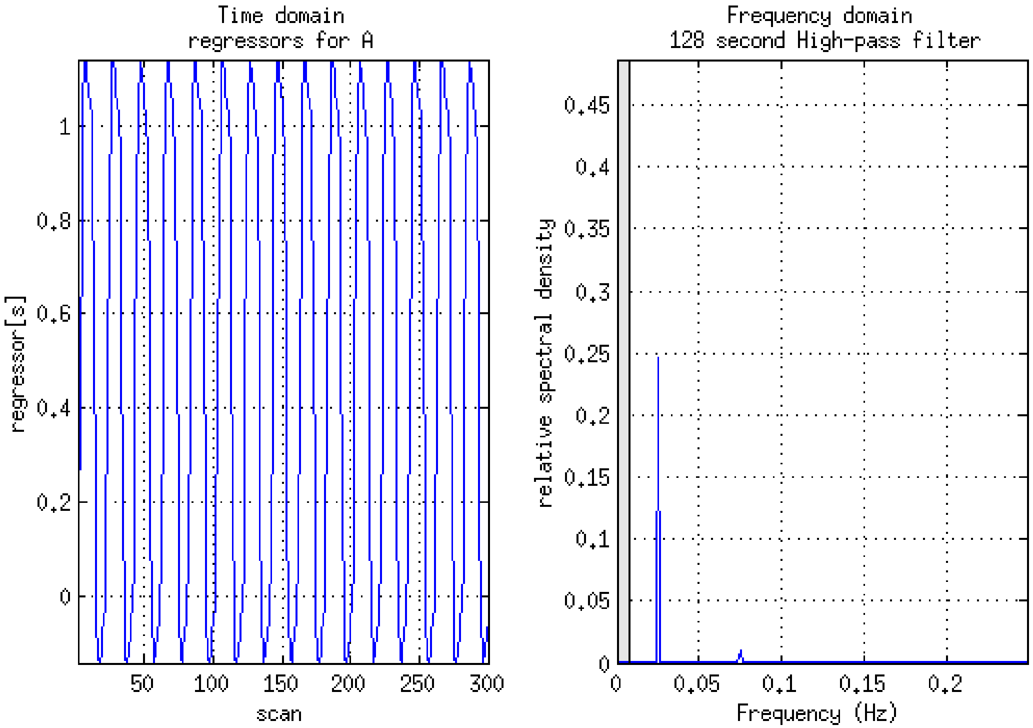
Hands on

1. Click on Review in the SPM main menu and select the SPM.mat we just created
2. Hit Design -> Design Orthogonality

Looks good ☺
No correlated regressors
The breaks look reasonable

Hands on

1. **In the SPM Menu click on Review and load the SPM.mat file you just created.**
2. **Click on Design -> Explore -> Session 1 -> A**



Looks good ☺
Our energy is at the right spot and not filtered out – yippie

Hands on

1. **Go to your matlab command line and load the design matrix:**
   - **X=tmp.xX.X;**
2. **Define your contrast of interest:**
   - **c=[1 -1 0]**

Compute the design efficiency

- **varEtaHat = c*inv(X'*X)*c';**
- **DesignEfficiency = 1/varEtaHat;**

- Our Design Efficiency for this design is
  - c=[1 -1 0]: 79.5
  - c=[-1 1 0]: 79.5
  - c=[1 1 0]: 0.56 -> oh … we are 142 times more inefficient for the common effect than for the difference effect ☹

Hands on

Lets make our design better for the common effect
lets insert null trials, where the subject is not engaged in task A or B

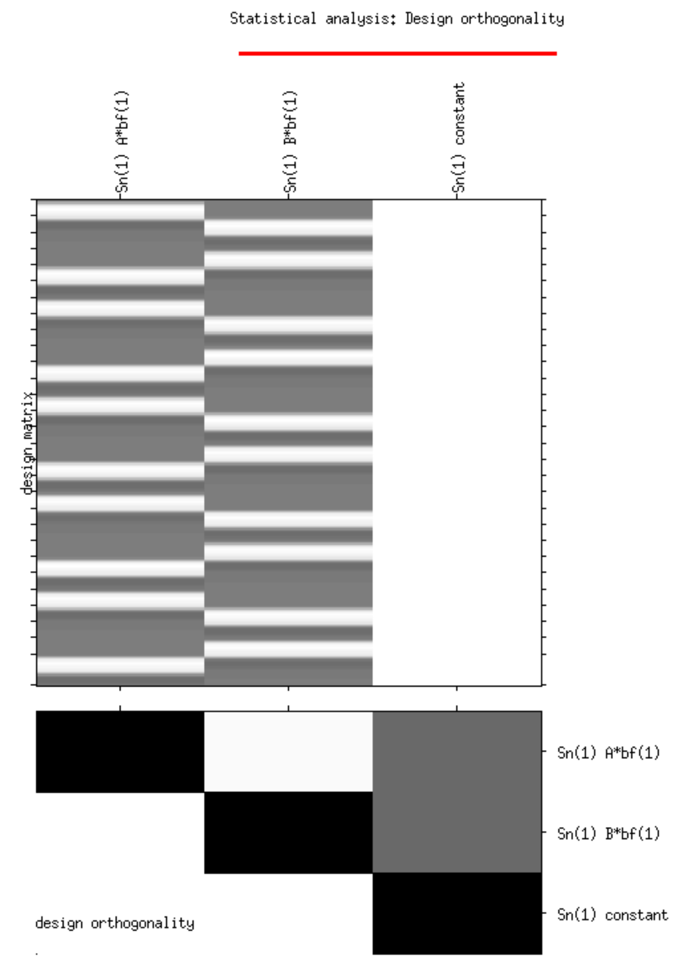
1. **Enter Parameters for condition A**
   - **Name: A**
   - **Onsets: 0 80 120 200 240 320 360 440 480 560**
   - **Durations: ones(10,1) * 20**
2. **Enter Parameters for condition B**
   - **Name: B**
   - **Onsets: 20 60 140 180 260 300 380 420 500 540**
   - **Durations: ones(10,1) * 20**

Save your design
Hit the run button ☺

Hands on

1. Click on Review and select the SPM.mat we just created
2. Hit Design -> Design Orthogonality

Looks good ☺
No correlated regressors
The breaks look reasonable



Hands on

1. Go to your matlab command line and load the design matrix:
   - X=tmp.xX.X;
2. Define your contrast of interest:
   - c=[1 -1 0]

Compute the design efficiency

- varEtaHat = c*inv(X'*X)*c';
- DesignEfficiency = 1/varEtaHat;

- Our Design Efficiency for this design is
  - c=[1 -1 0]: 52.8 (previous: 79.5)
  - c=[-1 1 0]: 52.8 (previous: 79.5) -> cool, we are still efficient for the difference effect
  - c=[1 1 0]: 17.6 (previous: 0.56) -> and we are only 3 times less efficient for the common effect ☺

Hands on

•Create a design with a very long block length and see what happens

•Create a design with very short block length and see what happens

•Create a design where your regressors are correlated and see what happens (**hint**: you create correlating regressors by overlapping your regressor slightly in time, then they get a shared variance because they explain the same thing)