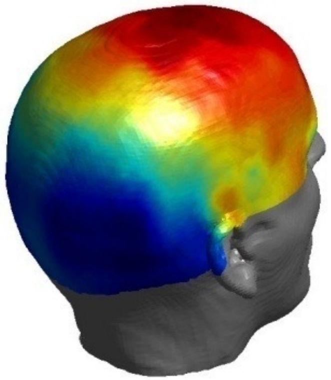
# Multiple Comparisons: Problems & Solutions

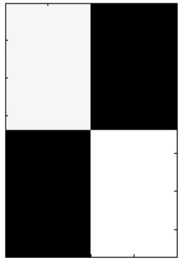
Andreea O. Diaconescu
Cognitive Network Modelling, Krembil Centre for Neuroinformatics,
Centre for Addiction and Mental Health (CAMH) & University of Toronto, Toronto, Canada

SPM M/EEG Course

**With many thanks for slides & images to:**

**FIL Methods group**

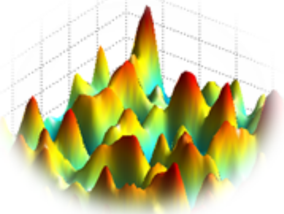


$y =$  $\beta + \varepsilon$

Contrast

$c$ 

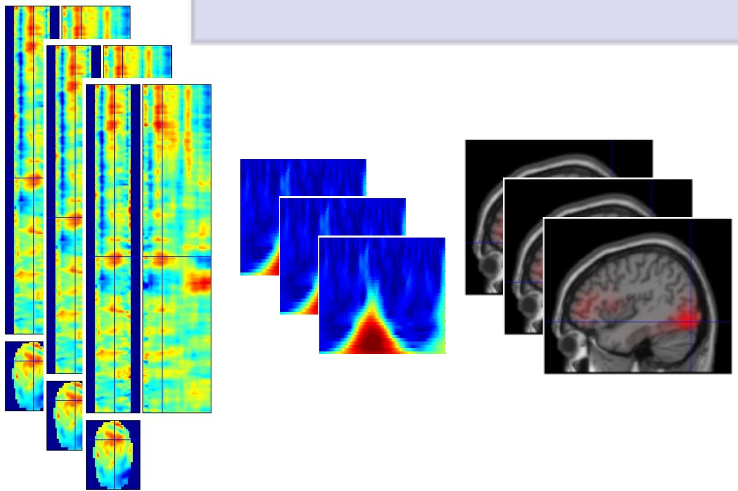
**Random Field Theory**



Pre-processings

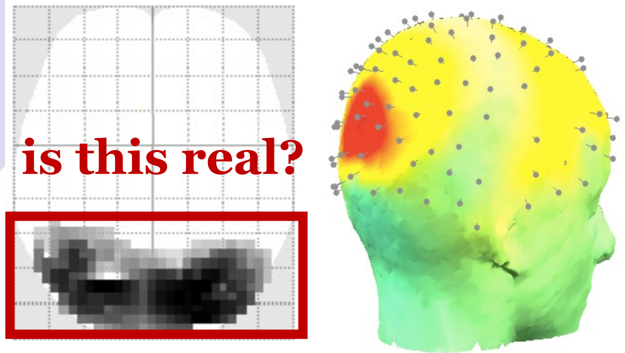
General Linear Model

Statistical Inference



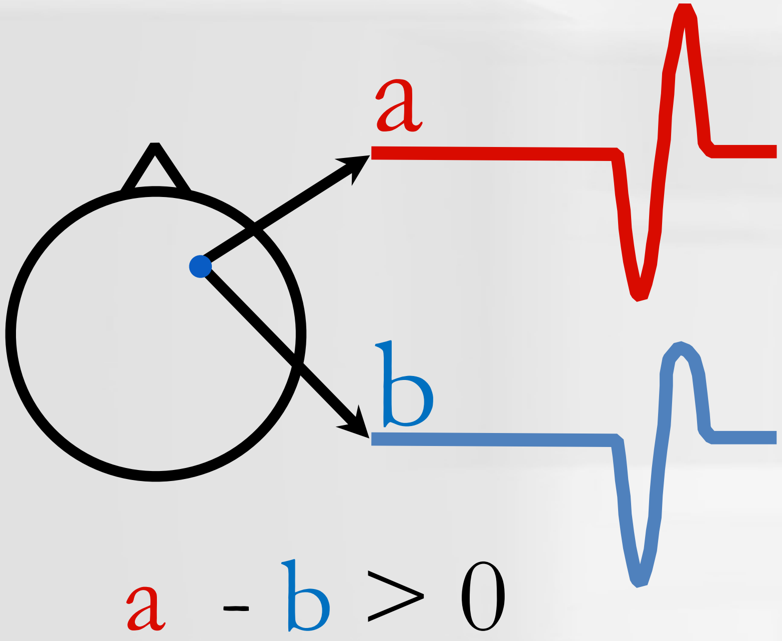
$$\hat{\beta} = (X^T X)^{-1} X^T y$$

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{rank(X)}$$




**Statistical image (SPM)**

# Statistical test on a Single Timepoint




To test a hypothesis, we construct a "test statistic".

- **"Null hypothesis" $H_0$ = "there is no effect" $\Rightarrow$ $c^T\beta = 0$**

  This is what we want to disprove.

  ⇨ The "alternative hypothesis" $H_1$ represents the outcome of interest.

- **The test statistic T**

  The test statistic summarises the evidence for $H_0$.

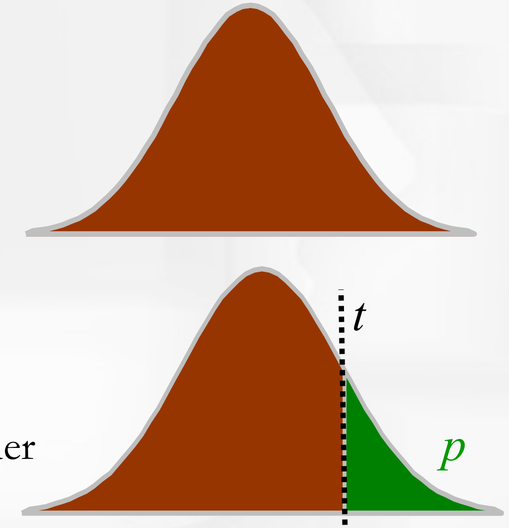
  ⇨ We need to know the distribution of T under the null hypothesis.

- **Observation of test statistic t, a realisation of T**

  A p-value summarises evidence against $H_0$.

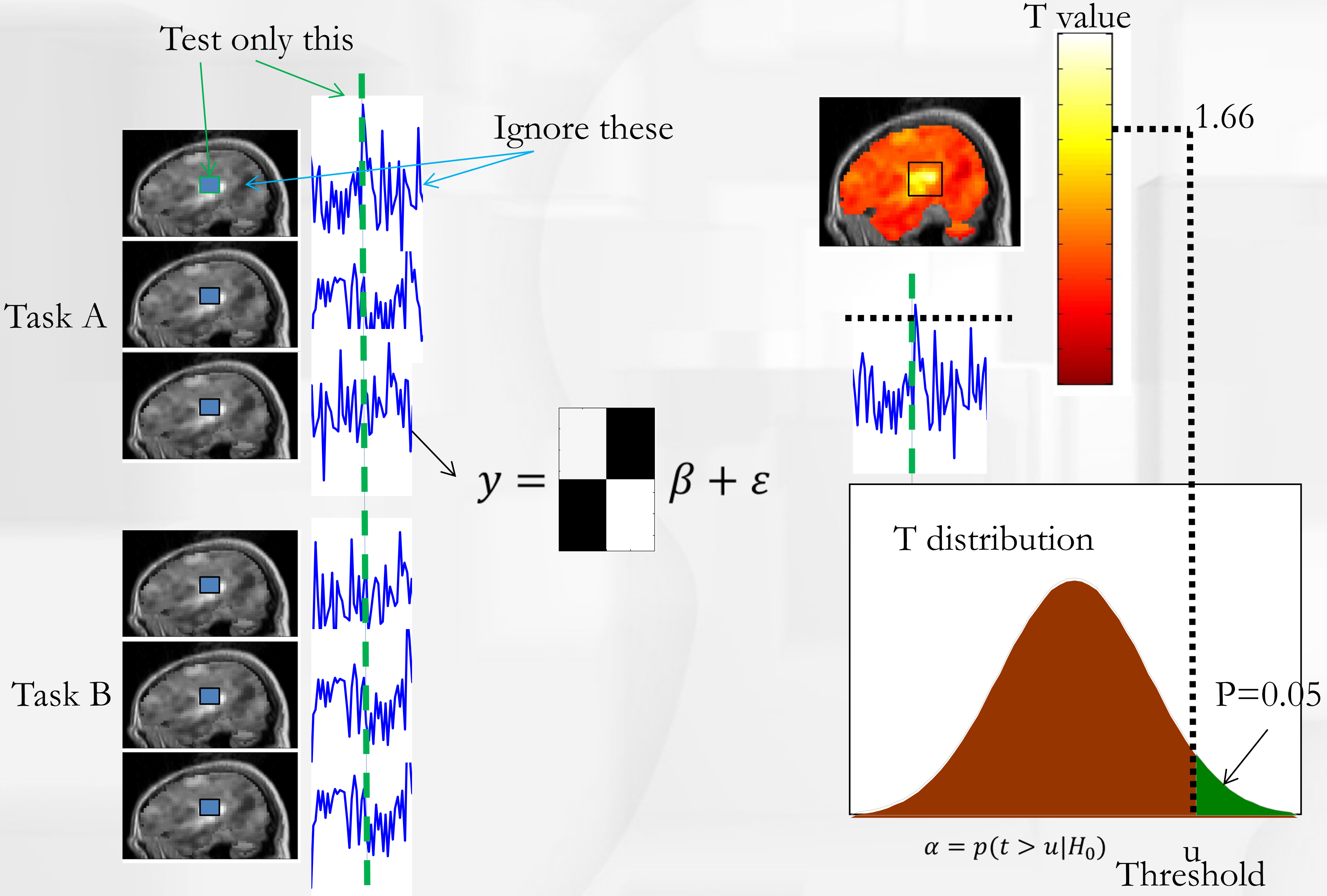
  This is the probability of observing t, or a more extreme value, under the null hypothesis:

$$p(T \geq t \mid H_0)$$



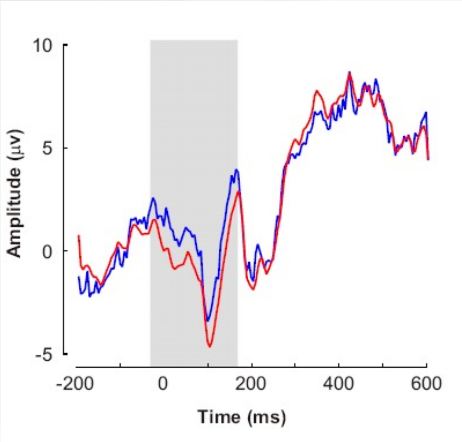

Null Distribution of T

# Statistical test at a single voxel/timepoint

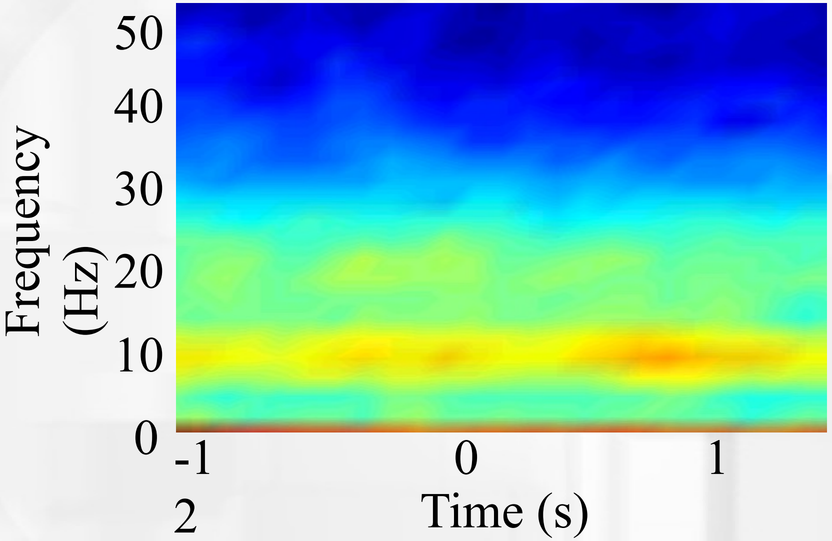


# Multidimensional Data

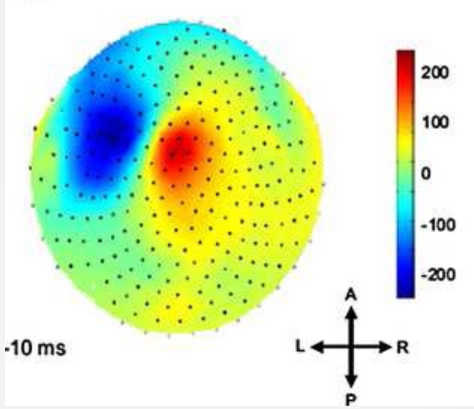
1-D



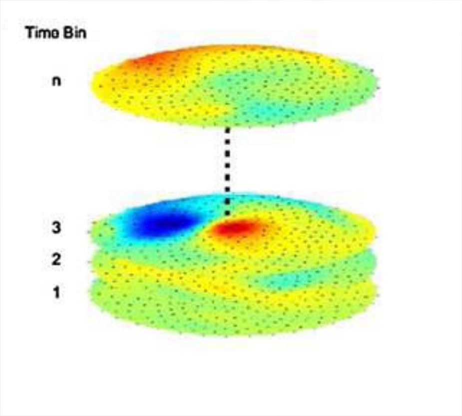

2-D



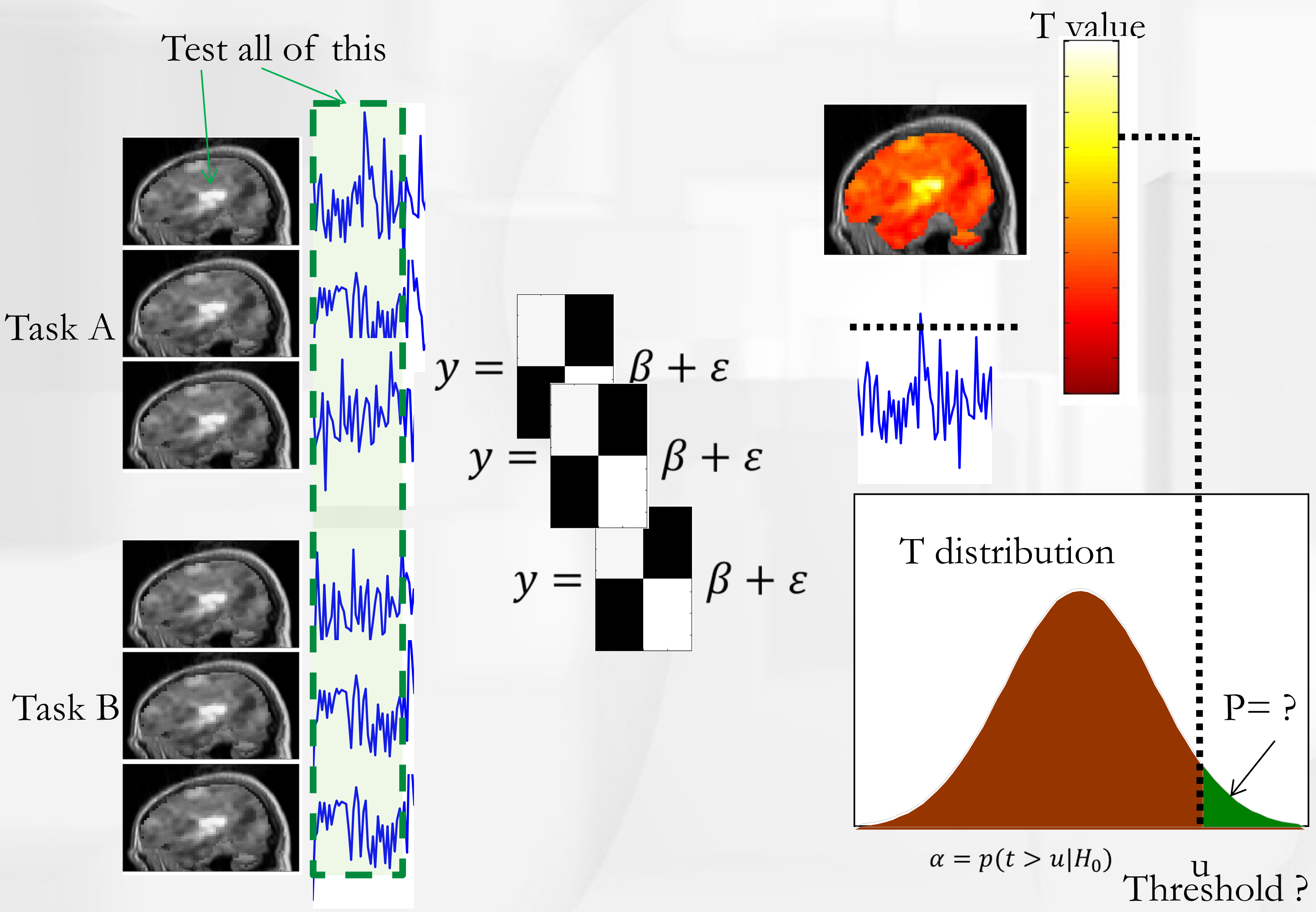

2-D

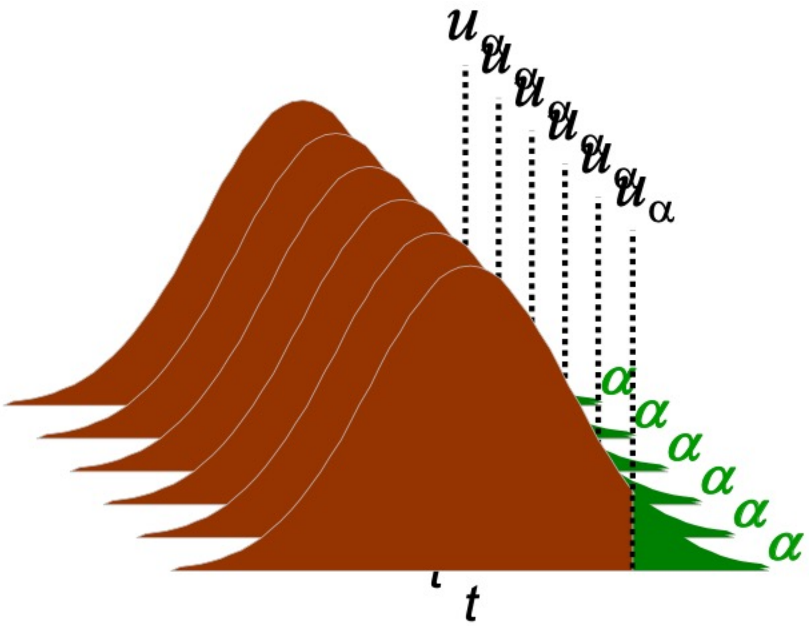



3-D



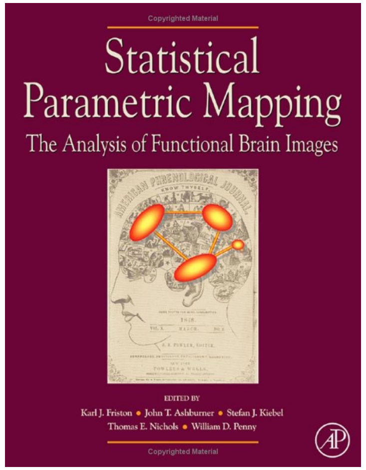
# Statistical test at multiple voxels/timepoints



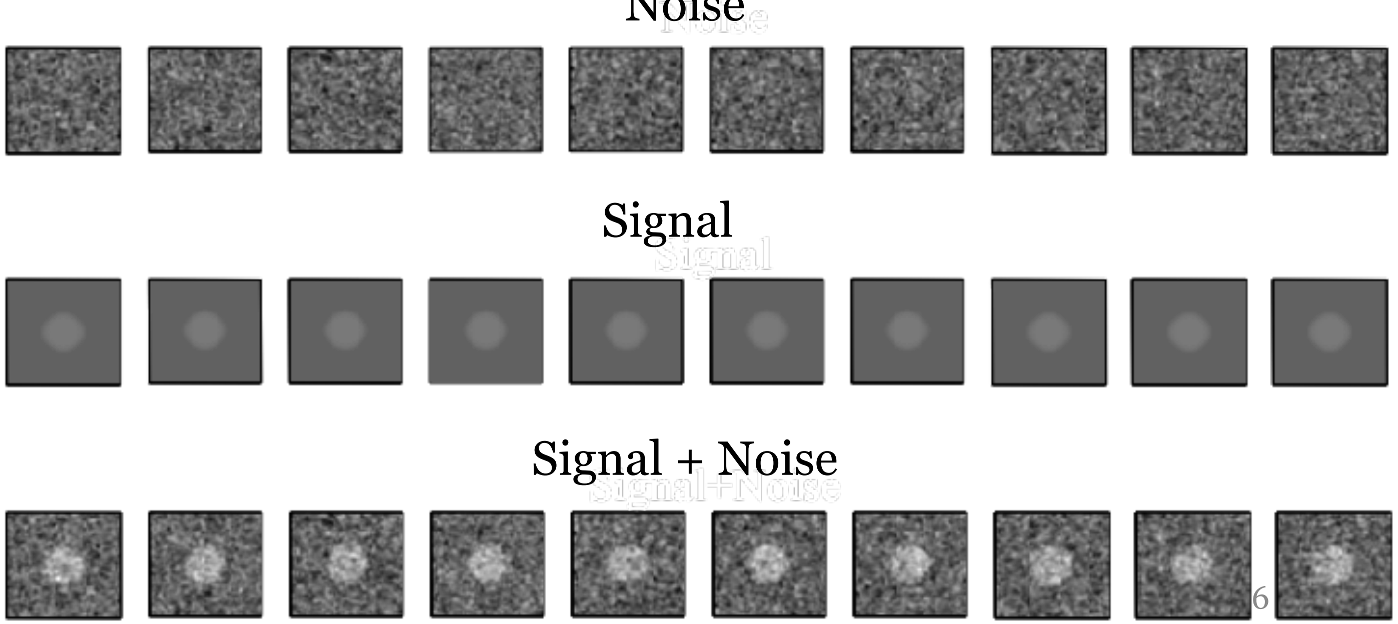


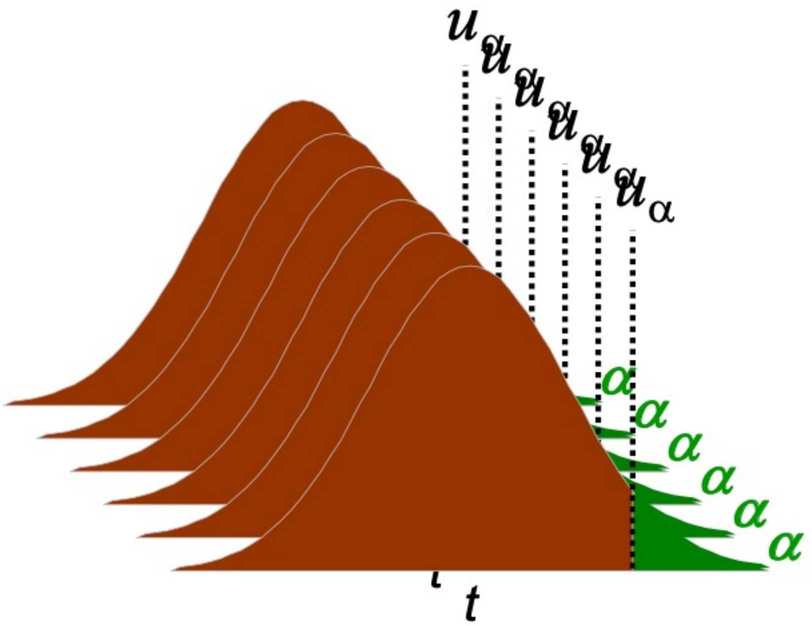

| | | Conclusion about null hypothesis from statistical test | |
|---|---|---|---|
| | | Accept Null | Reject Null |
| Truth about null hypothesis in population | True | Correct | Type I error<br>Observe difference when none exists |
| | False | Type II error<br>Fail to observe difference when one exists | Correct |

If we have 100,000 voxels,
$\alpha$=0.05 -> 5,000 false positive voxels.



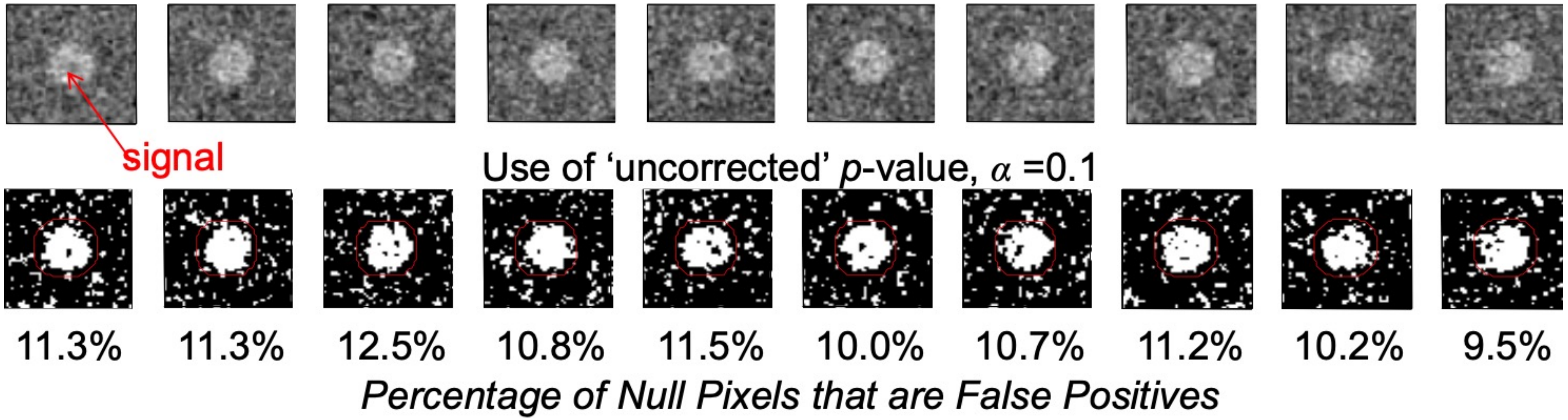

*Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier, 2007.






| | | Conclusion about null hypothesis from statistical test | |
|---|---|---|---|
| | | Accept Null | Reject Null |
| Truth about null hypothesis in population | True | Correct | Type I error<br>Observe difference when none exists |
| | False | Type II error<br>Fail to observe difference when one exists | Correct |

If we have 100,000 voxels,
$\alpha$=0.05 -> 5,000 false positive voxels.



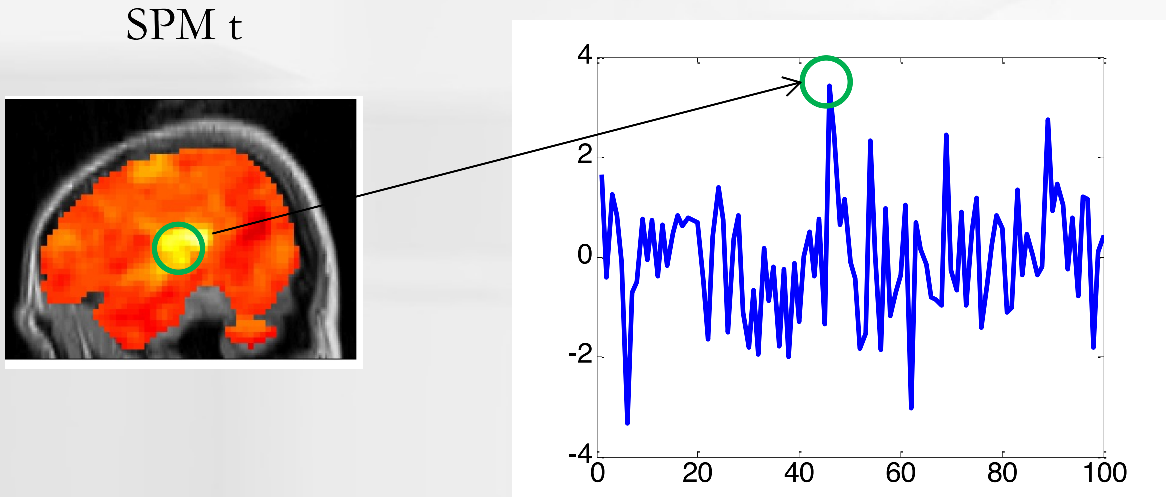
# Common Solutions: 1. Averaging

One solution is to reduce the multi-dimensional data to zero-dimensional data by averaging over a window of interest

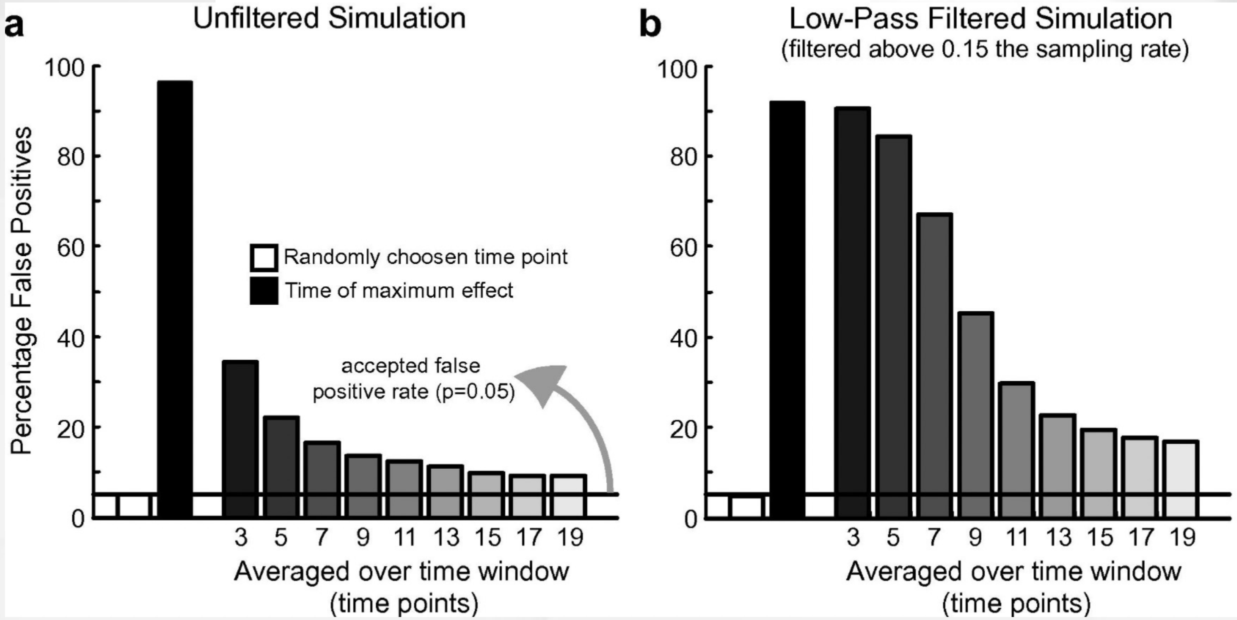
This must be specified *a priori* or derived from an independent contrast.

One cannot base this window on where the effect size is largest!

# Don’t do this!




With no prior hypothesis. Test whole volume. Identify SPM peak.
Then make a test assuming a single voxel/time of interest at that peak.




**James Kilner.** Clinical Neurophysiology 2013.

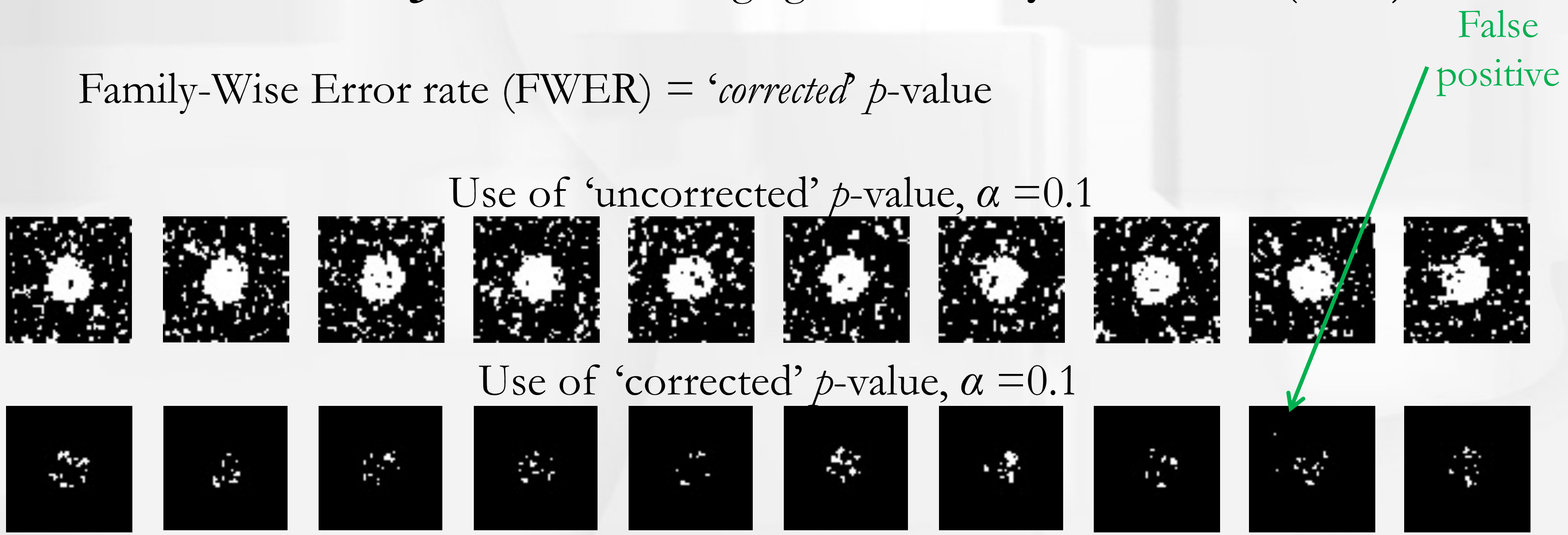
# 2. Family-Wise Null Hypothesis

***Family-Wise Null Hypothesis:***
*Activation is zero everywhere*

If we reject a voxel null hypothesis at *any* voxel, we reject the family-wise Null hypothesis

A False Positive ***anywhere*** in the image gives a **Family Wise Error** (FWE)

Family-Wise Error rate (FWER) = '*corrected*' *p*-value



# Bonferroni Correction

The Family-wise Error rate (FWE) $\alpha$ for a family of N independent voxels is:

$$\alpha = Nv$$

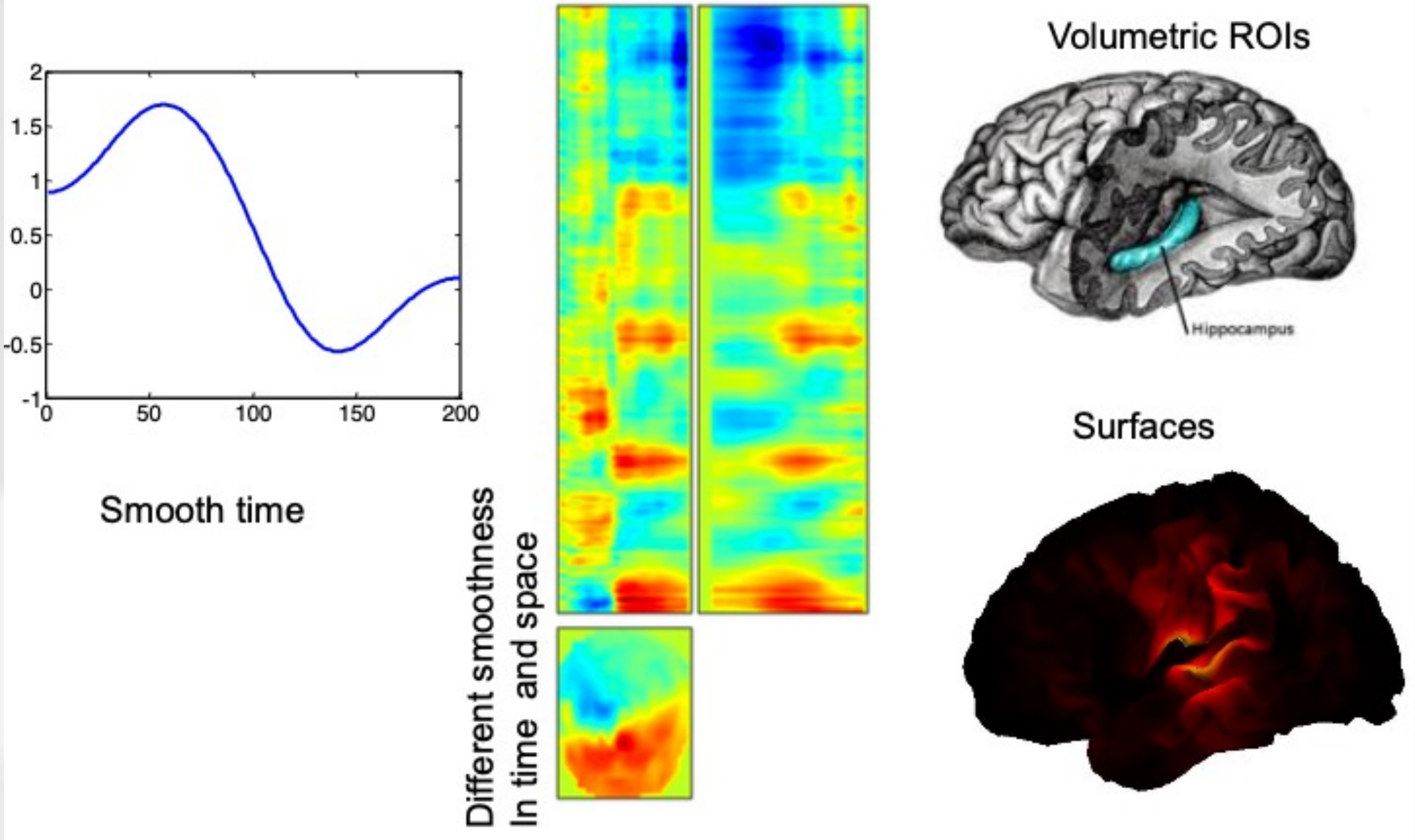
where $v$ is the voxel-wise error rate. Therefore, to ensure a particular FWE, we set

$$v = \frac{\alpha}{N}$$

*However, Bonferroni correction assumes independence*

*M/EEG data are correlated either temporally, spatially or in frequency space*

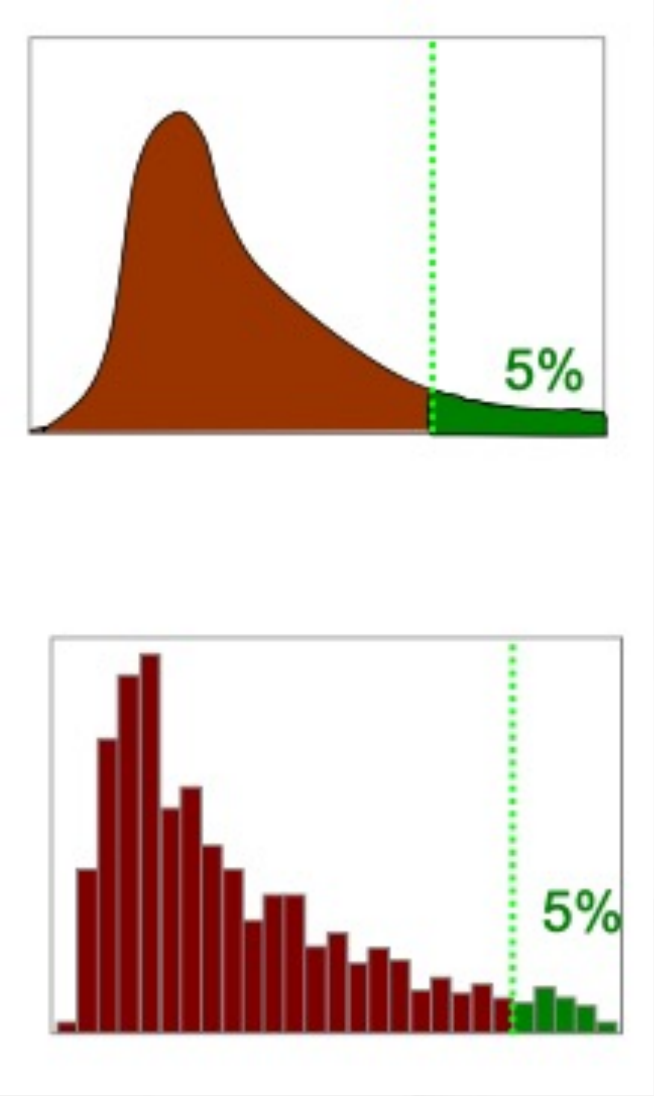
# Bonferroni correction is too conservative for data with different topologies and smoothness



# Nonparametric inference: Permutation tests to control for FWE rates

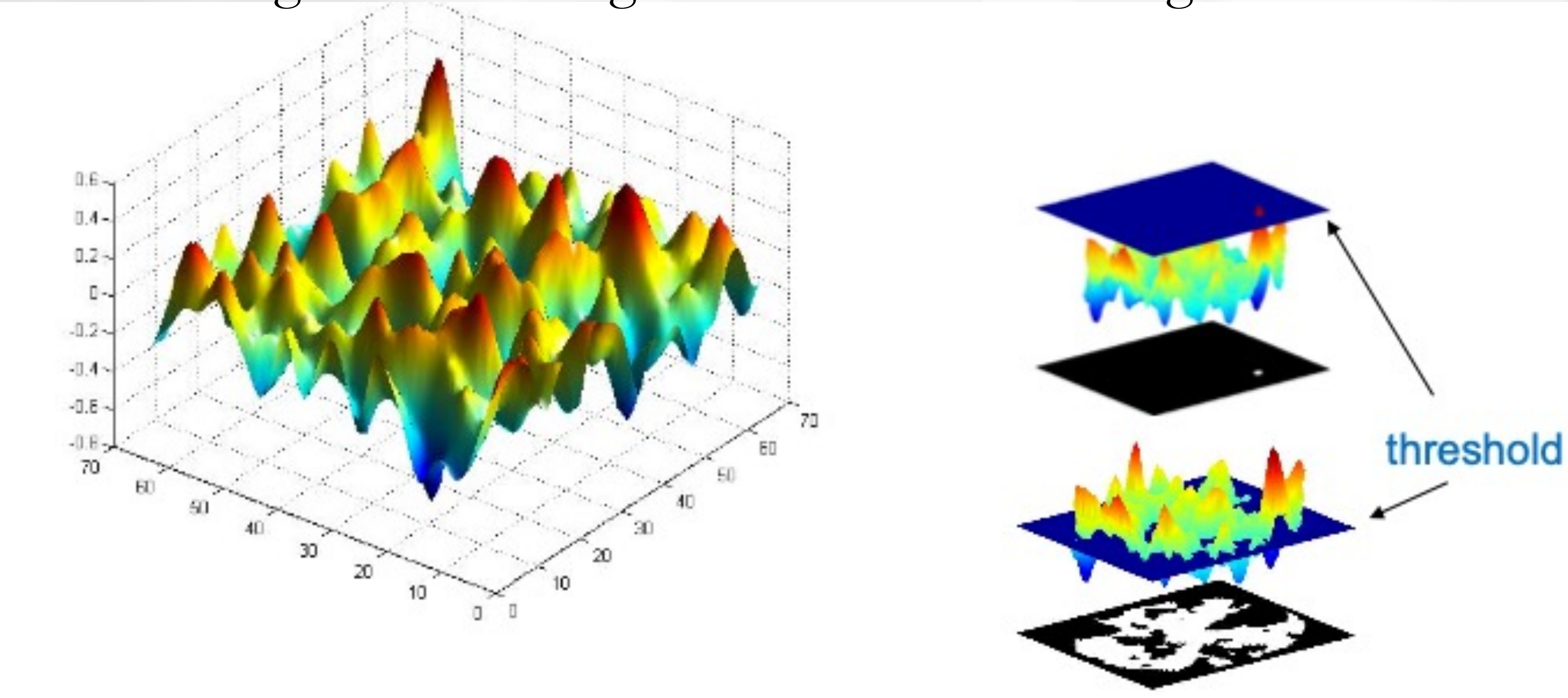
- Parametric methods
  - Assume distribution of *max* statistic under null hypothesis

- Nonparametric methods
  - Use *data* to find distribution of *max* statistic under null hypothesis
  - any max statistic



# 3. Random Field Theory

**A random field:** an array of smoothly varying test statistics. e.g. a slice through a t-statistic brain image.




Keith Worsley, Karl Friston, Jonathan Taylor, Robert Adler and colleagues

# Euler Characteristic

**Euler Characteristic $\chi_u$:**

- Topological measure

  $\chi_u$ = # blobs - # holes

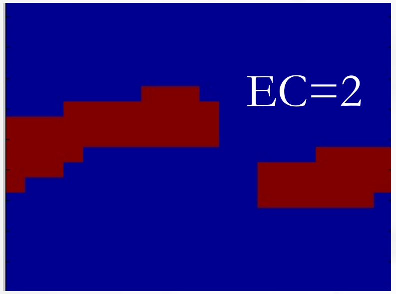
- at high threshold $u$:

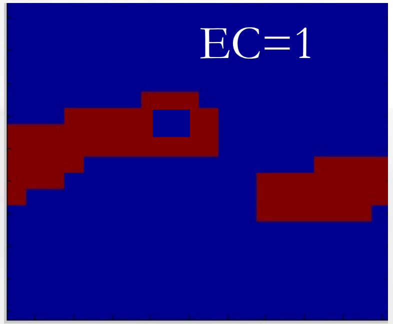
  $\chi_u$ = # blobs

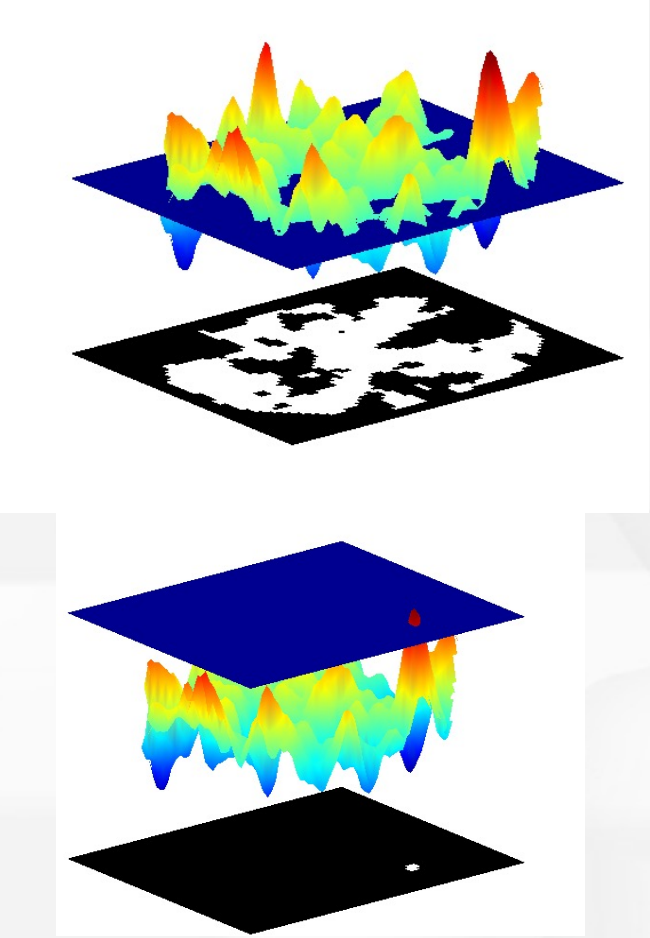
$$
\begin{aligned}
FWER &= p(FWE) \\
&= p\left(\bigcup_i \{T_i \geq u\} \middle| H_0\right) \\
&= p\left(\max_i T_i \geq u \middle| H_0\right) \\
&= p(one\ or\ more\ blobs\ |H_0) \\
&\approx p(\chi_u \geq 1|H_0) \\
&\approx E[\chi_u|H_0] \approx \boldsymbol{\alpha_{FWE}}
\end{aligned}
$$

*No holes*

*Zero or one blob*

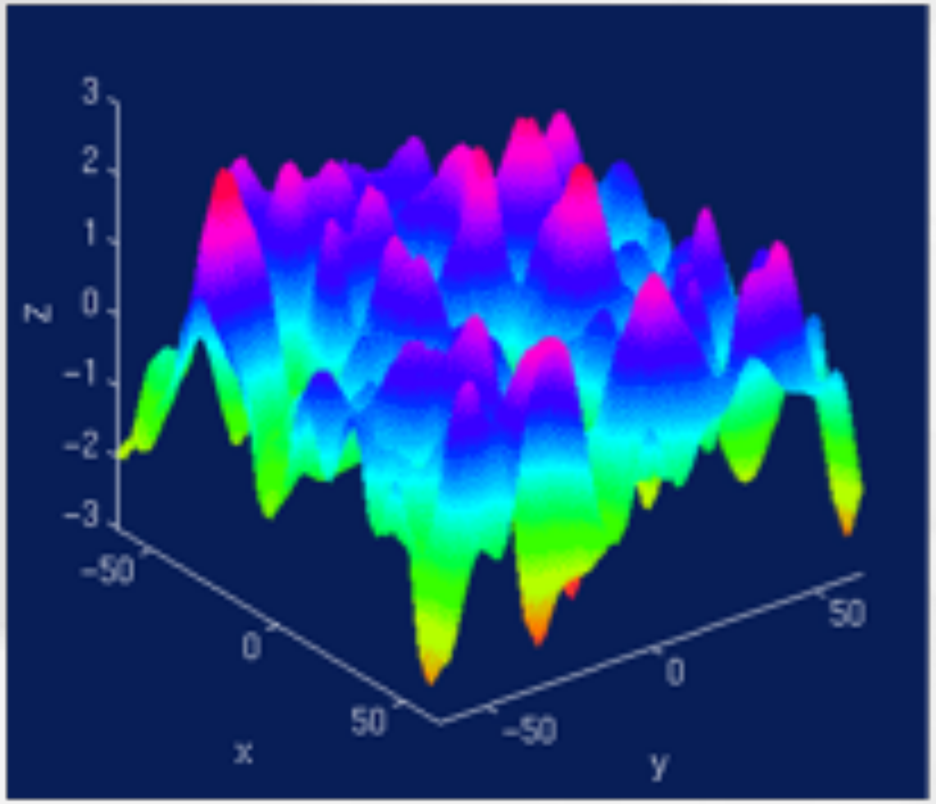


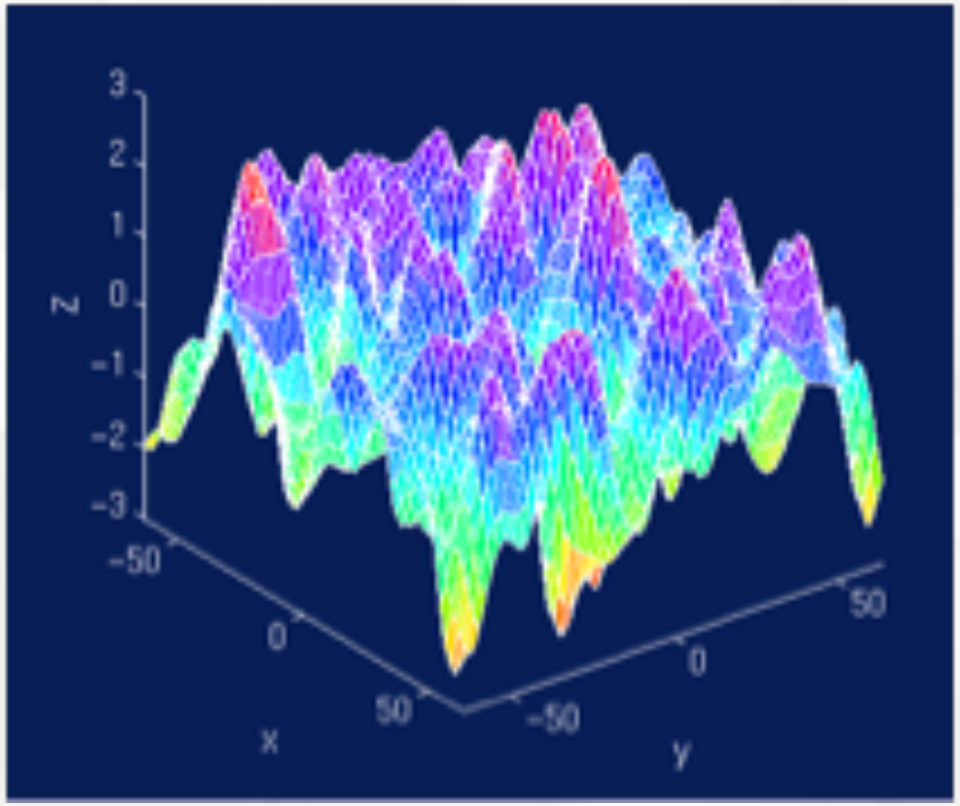



At high thresholds there are no holes so EC= number blobs

# Good lattice approximation?




**Only true for high density recordings**

# Euler characteristic is given by the Gaussian Kinematic Formula

**Search region: Small volume correction**

$$E[\chi_u(\Omega)] = \sum_{d=0}^{D} L_d(\Omega) p_d(u)$$

Expected Euler Characteristic

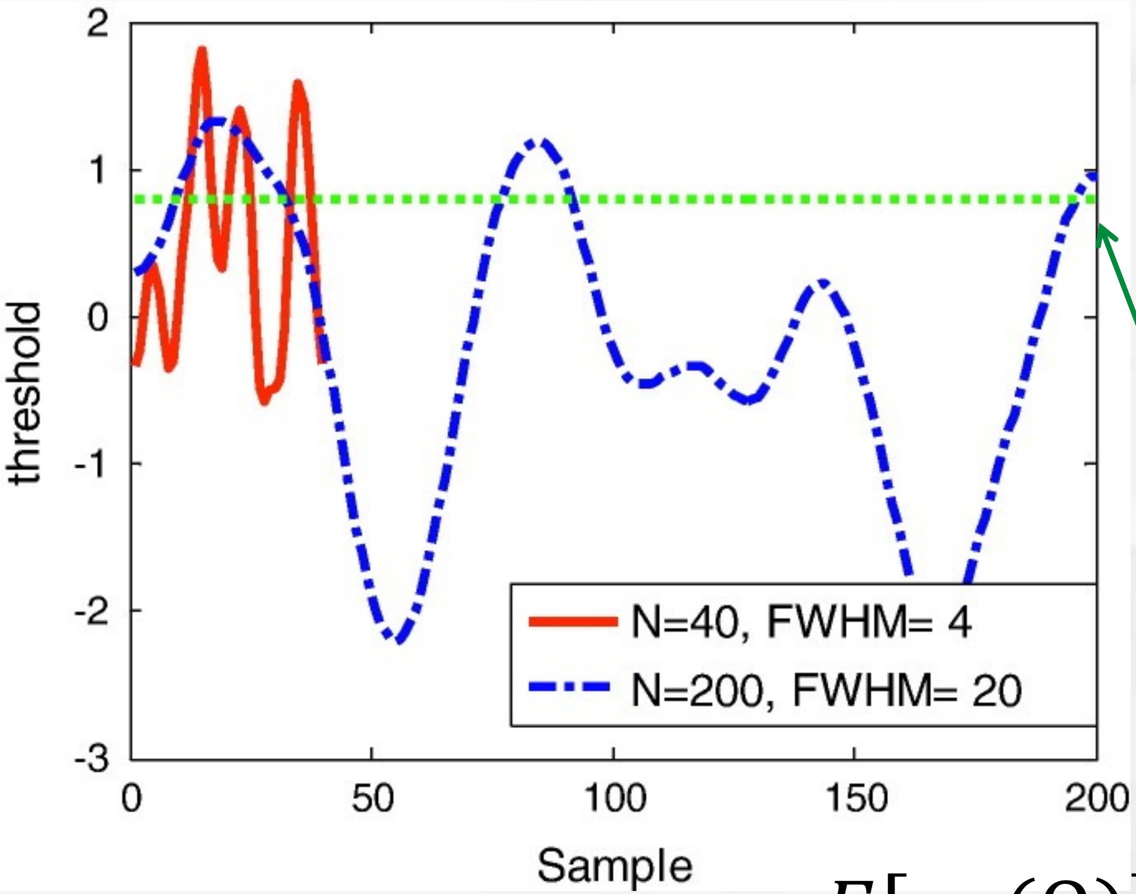
Intrinsic volume (depends on shape and smoothness of space)

Depends only on type of test and dimension

Number peaks= intrinsic volume * peak density

**Taylor & Worsley.**
*Journal of the American Statistical Association*. 2007.

# Example: one dimensional statistical field




The intrinsic volume (or the number of resels or the Lipschitz-Killing Curvature) of the two fields is identical

**Barnes et al.,** *NeuroImage*. 2013.

Threshold u

$$E[\chi_u(\Omega)] = \sum_{d=0}^{D} L_d(\Omega) p_d(u)$$

= 2.9 (in this example)

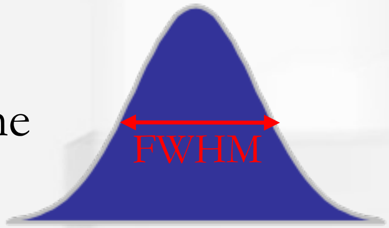
Expected Euler Characteristic

Intrinsic volume (depends on shape and smoothness of space)

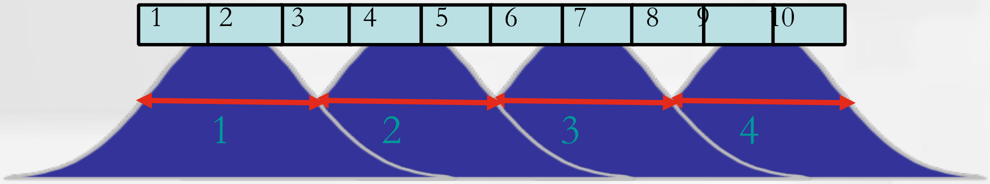
Depends only on type of test and dimension

# What determines the smoothness?

**Smoothness parameterised in terms of FWHM:**
Size of Gaussian kernel required to smooth i.i.d. noise to have same smoothness as observed null (standardized) data.



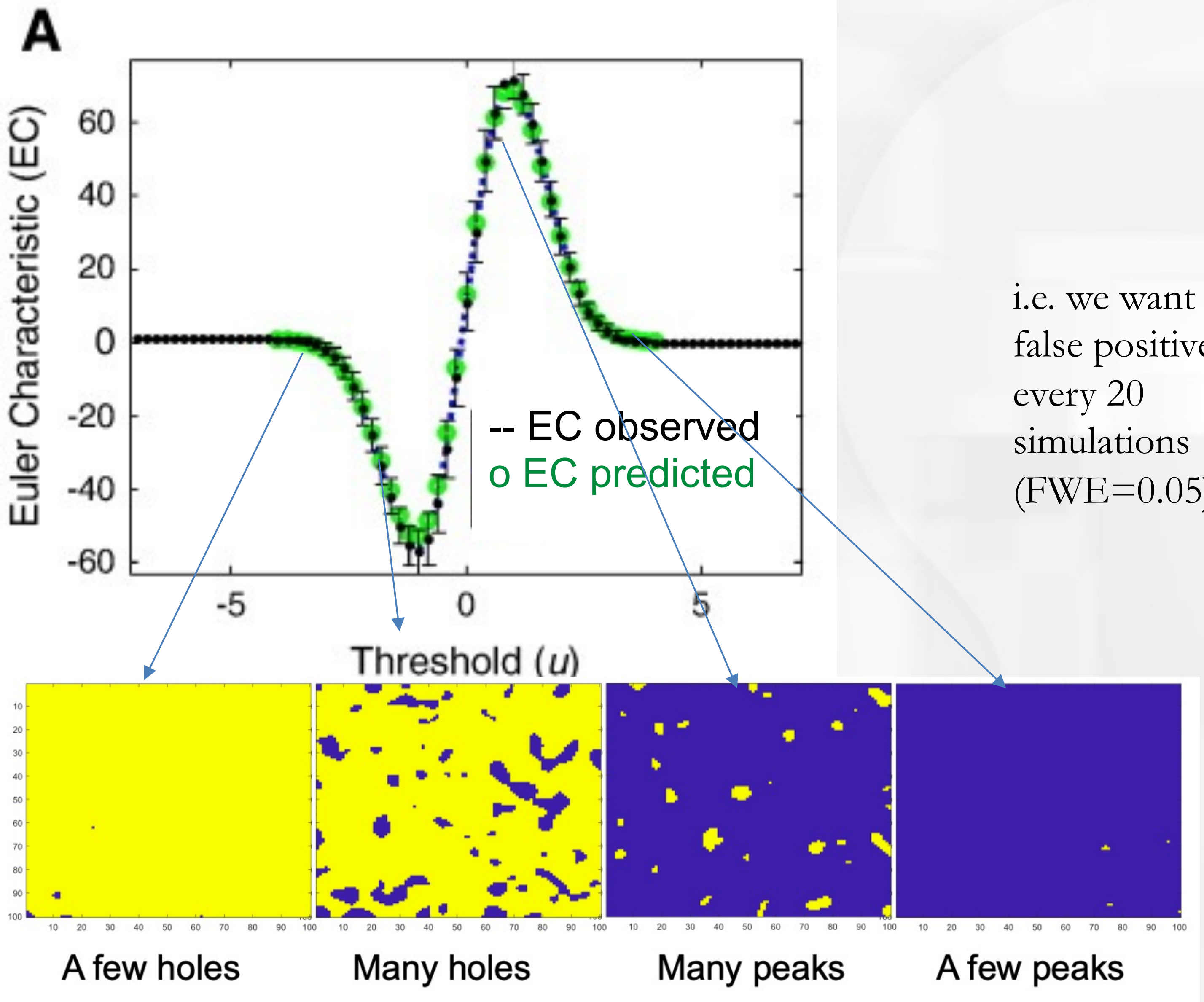

**RESELS (Resolution Elements):**
1 RESEL = $FWHM_x FWHM_y FWHM_z$
RESEL Count $R$ = volume of search region in units of smoothness



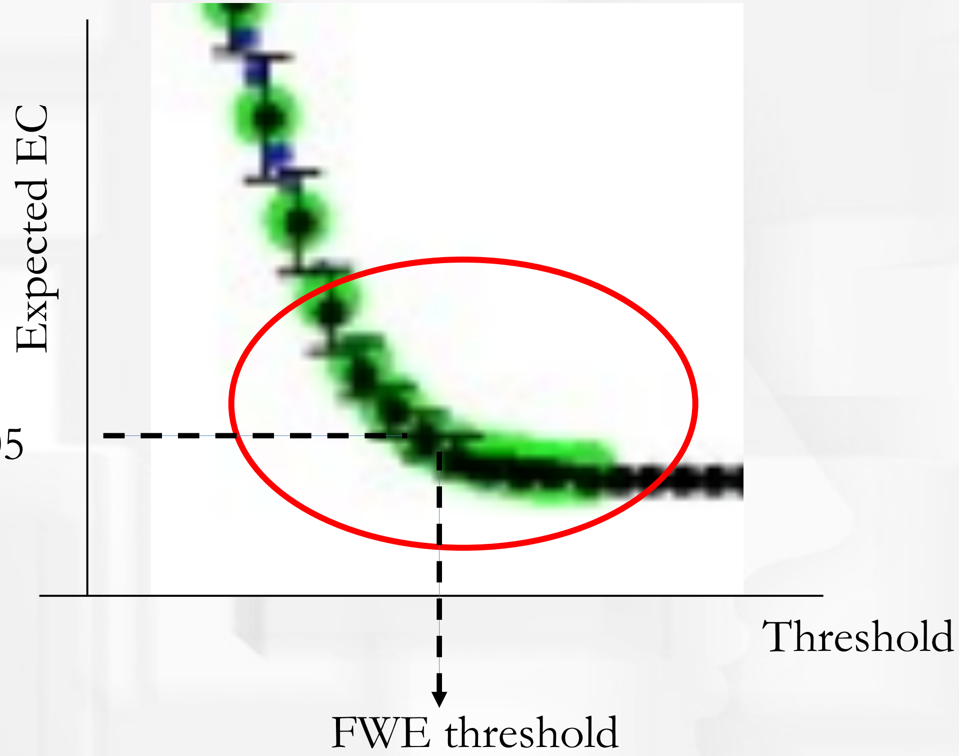

Eg: 10 voxels, 2.5 FWHM, 4 RESELS

The number of resels is similar, but not identical to the number independent observations.

# Euler Characteristic as a Function of Threshold



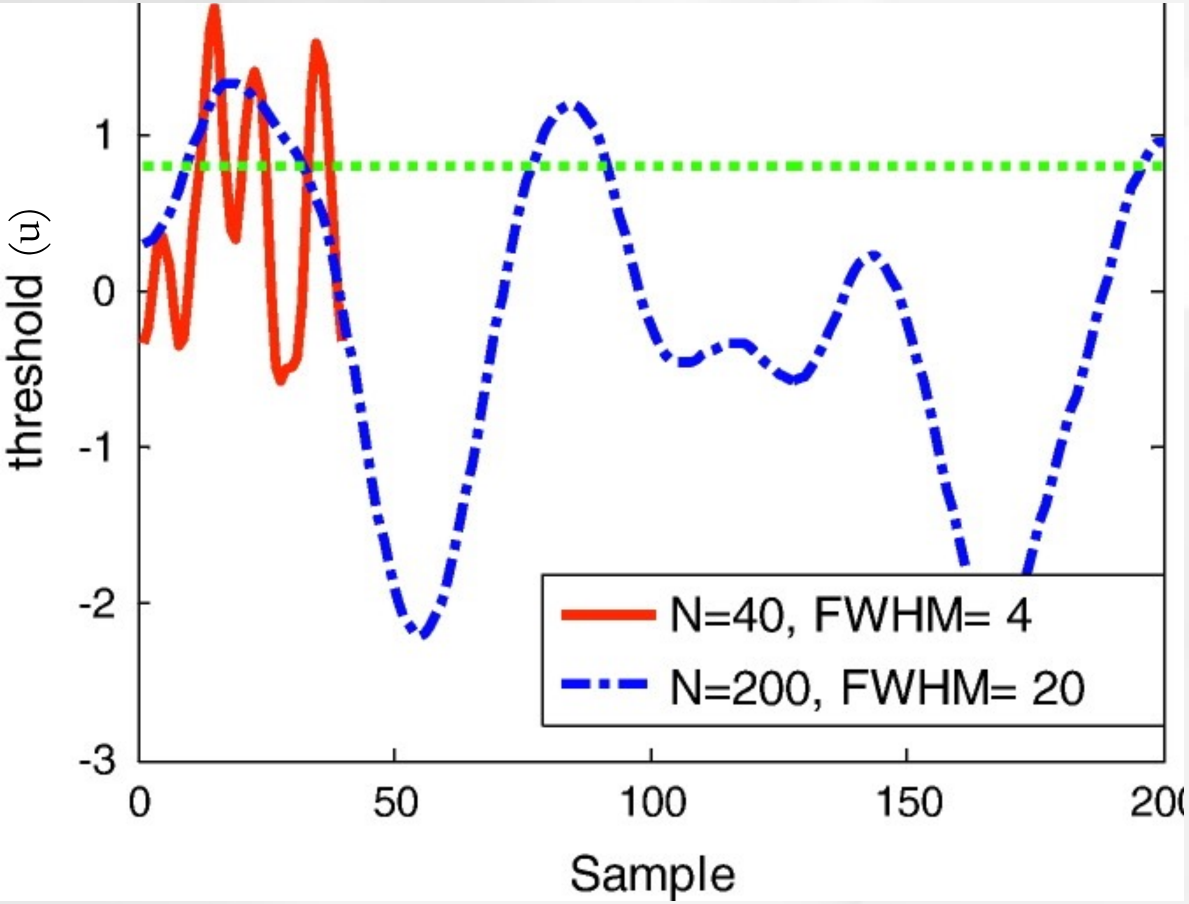

i.e. we want one false positive every 20 simulations (FWE=0.05)




$$E[\chi_u(\Omega)] = \sum_{d=0}^{D} L_d(\Omega) p_d(u)$$

**Barnes et al.,** *NeuroImage.* 2013.

# How to determine the FWE threshold




Know test (t) and dimension (1) so can get threshold u

$$E[\,0.05\,] = \sum_{d=0}^{D} L_d(\Omega)\, p_d(u)$$

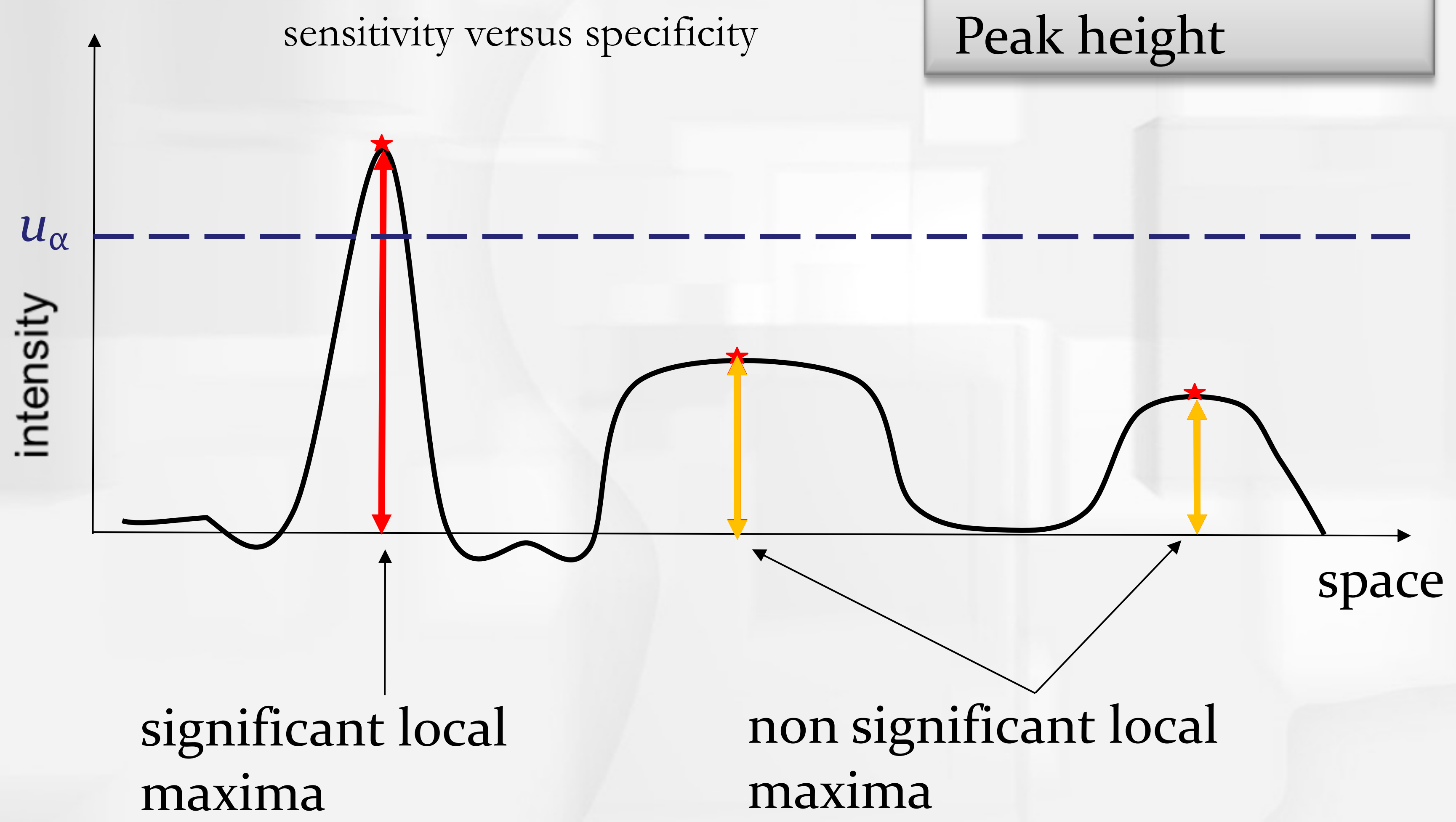
Know intrinsic volume (10 resels)

Want only a 1 in 20
Chance of a false positive

**Barnes et al.,** *NeuroImage*. 2013.

# Topological Inference

**Topological feature:**
Peak height

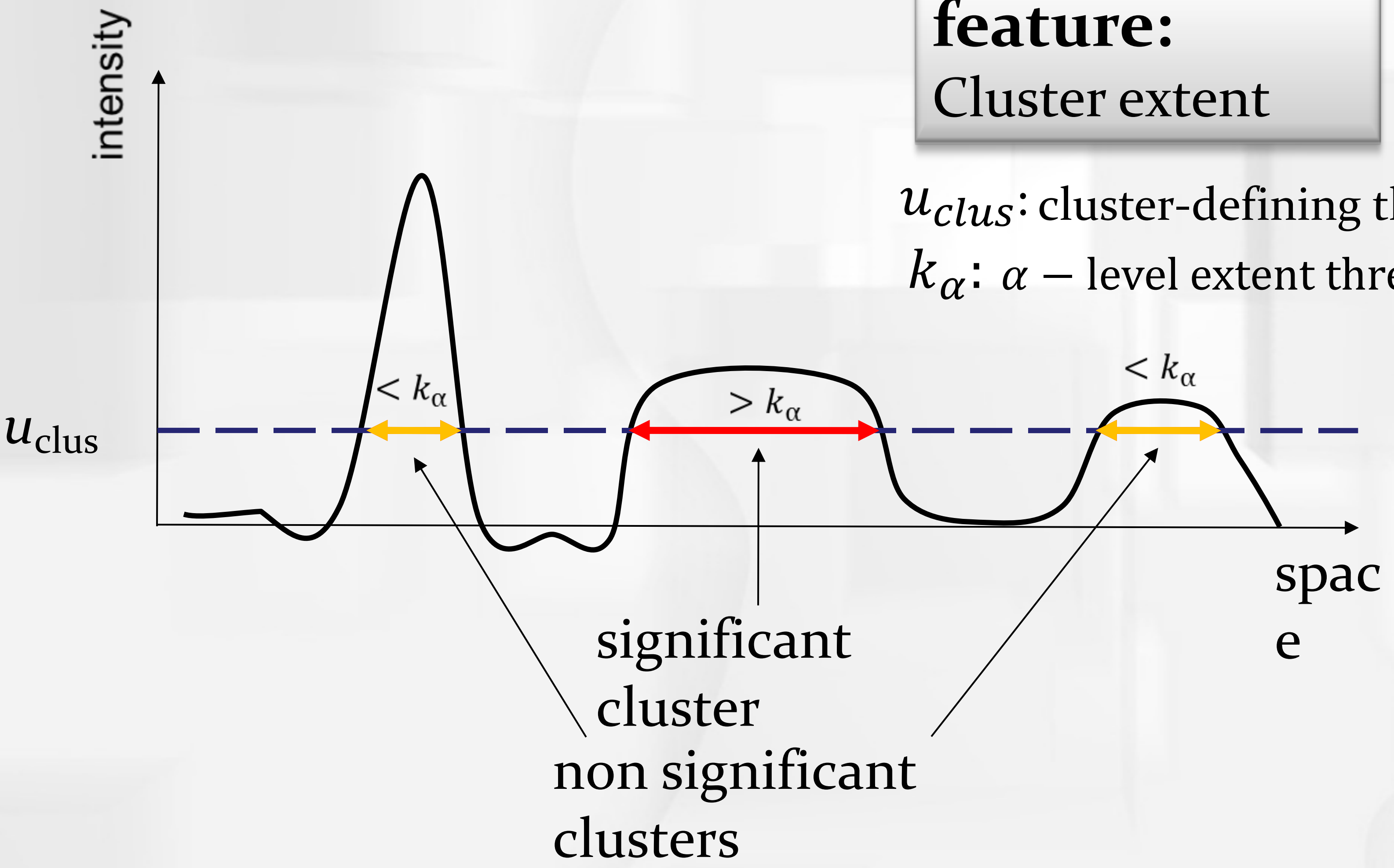


# Topological Inference: Cluster-level

**Topological feature:**
Cluster extent

$u_{clus}$: cluster-defining threshold (p = 0.001)

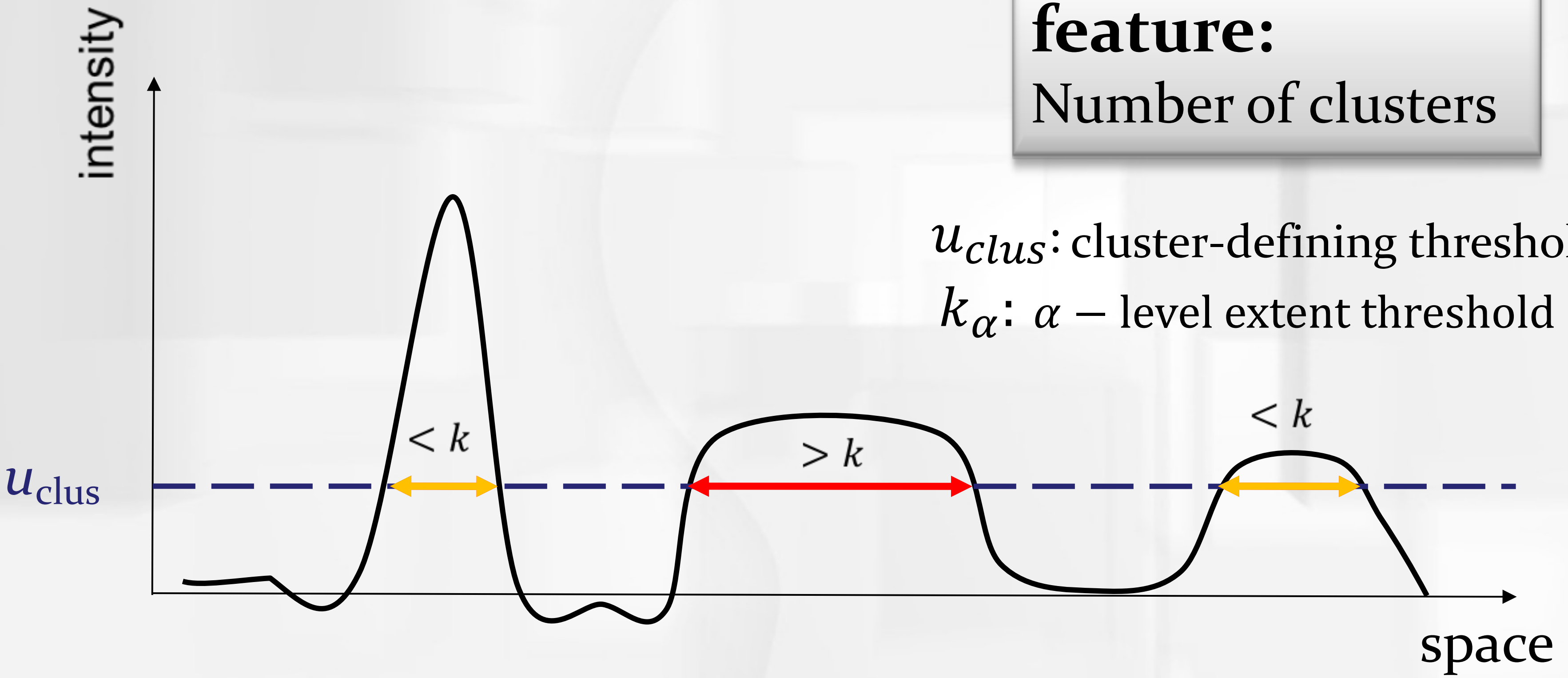
$k_{\alpha}$: $\alpha -$ level extent threshold



# Topological Inference: Set-level

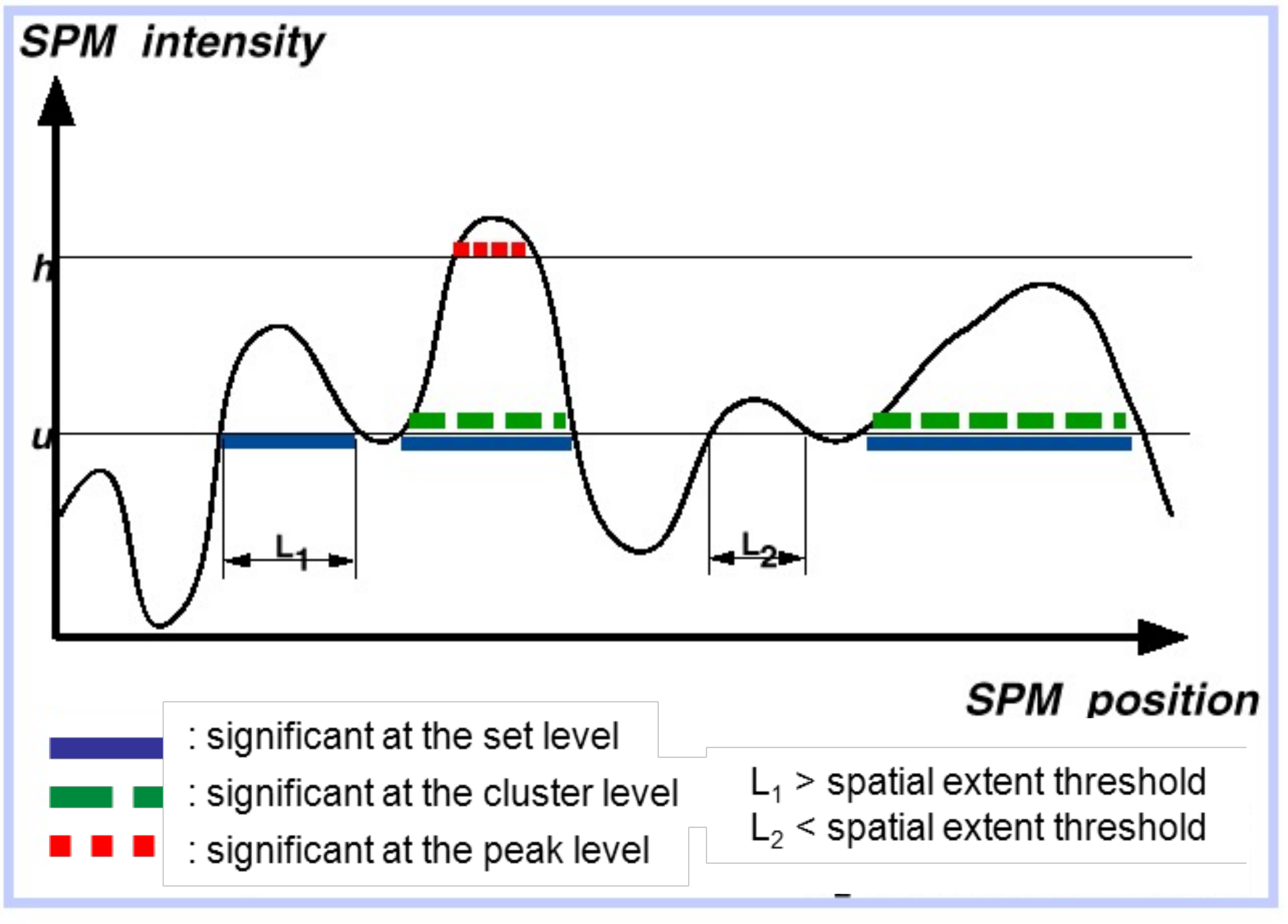
**Topological feature:**
Number of clusters

$u_{clus}$: cluster-defining threshold (p = 0.001)
$k_\alpha$: $\alpha$ – level extent threshold




Here, *c*=1, only one cluster larger than *k*.

# Peak, cluster, and set level inference




Sensitivity

Regional specificity

Peak level test:
height of local maxima

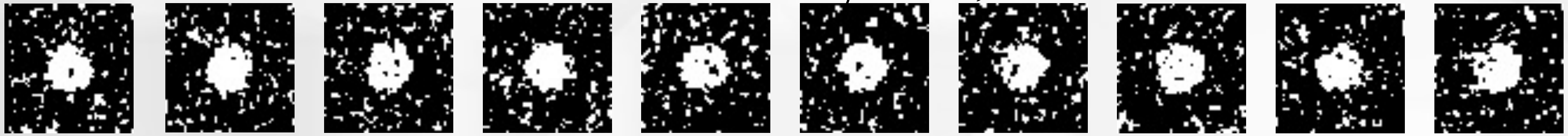
Cluster level test:
spatial extent above u

Set level test:
number of clusters above u

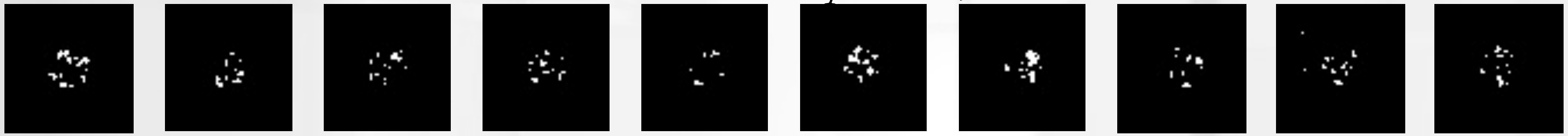
# False Discovery Rate

Family-wise Error Rate

Use of ‘uncorrected’ $p$-value, $\alpha = 0.1$



Use of ‘corrected’ $p$-value, $\alpha = 0.1$

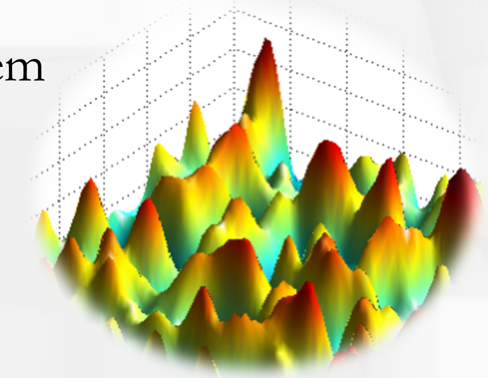


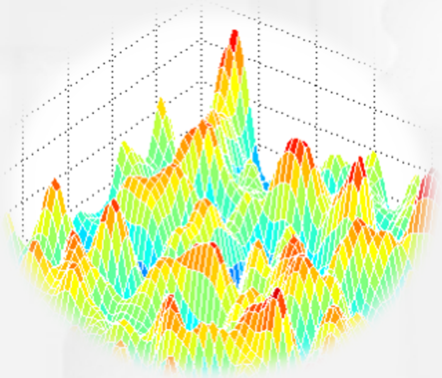
Control of False Discovery Rate as 10%

6.7% 10.4% 14.9% 9.3% 16.2% 13.8% 14.0% 10.5% 12.2% 8.7%

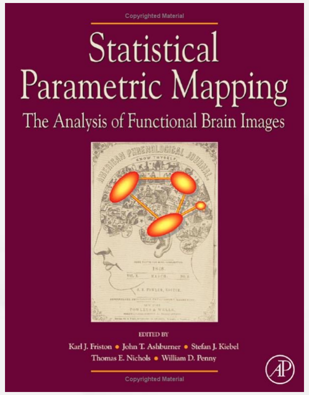
# Summary

- Bonferroni correction for multidimensional data is too conservative as it assumes independence of the samples
- Random Field Theory can be used to resolve the multiple comparisons problem that occurs when making inferences over the search-space
- The statistic image is assumed to be a good lattice representation of an underlying continuous stationary random field. Typically, FWHM > 3 voxels
- Smoothness of the data is unknown and estimated: very precise estimate by pooling over voxels ⇒ stationarity assumptions (esp. relevant for cluster size results).
- *A priori* hypothesis about where an activation should be, reduce search volume ⇒ Small Volume Correction:





# References:





- Friston KJ, Frith CD, Liddle PF, Frackowiak RS. *Comparing functional (PET) images: the assessment of significant change*. J Cereb Blood Flow Metab. 1991 Jul;11(4):690-9.
- Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. *A unified statistical approach for determining significant signals in images of cerebral activation*. Human Brain Mapping 1996;4:58-73.
- Chumbley J, Worsley KJ , Flandin G, and Friston KJ. *Topological FDR for neuroimaging*. NeuroImage, 49(4):3057-3064, 2010.
- Chumbley J and Friston KJ. *False Discovery Rate Revisited: FDR and Topological Inference Using Gaussian Random Fields*. NeuroImage, 2008.
- Kilner J and Friston KJ. *Topological inference for EEG and MEG data*. Annals of Applied Statistics, 2010.
- Kilner J. *Bias in a common EEG and MEG statistical analysis and how to avoid it*. Clinical Neurophysiology 2013.
- Barnes, GR, Ridgway, GR, Flandin, G, Woolrich, M and Friston, KJ. *Set-level threshold-free tests on the intrinsic volumes of SPMs*. Clinical Neurophysiology 2013.