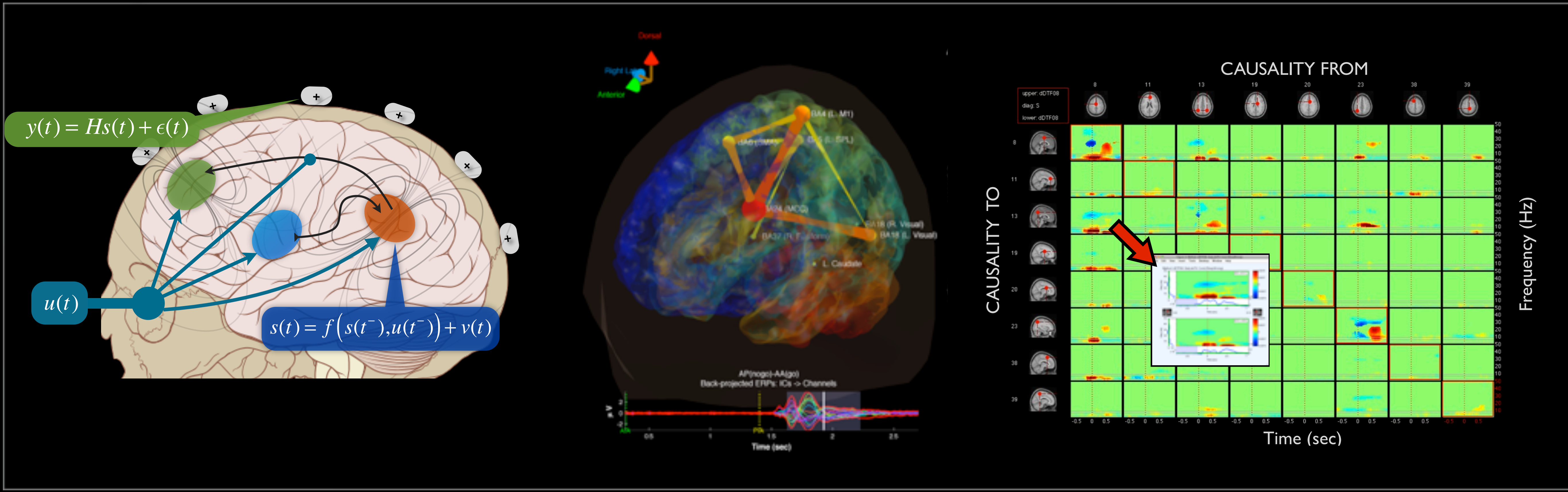
The Dynamic Brain: Modeling Neural Dynamics and Interactions from M/EEG



Tim Mullen, PhD

Intheon Labs

Swartz Center for Computational Neuroscience
Institute for Neural Computation
UC San Diego

Outline

Theoretical Foundations

- Introduction to Brain Connectivity Analysis
- Linear Dynamical Systems and the VAR model
- Granger Causality and Effective Connectivity Measures
- Scalp versus Source
- Adapting to Time-Varying Changes in Dynamics
- Statistics

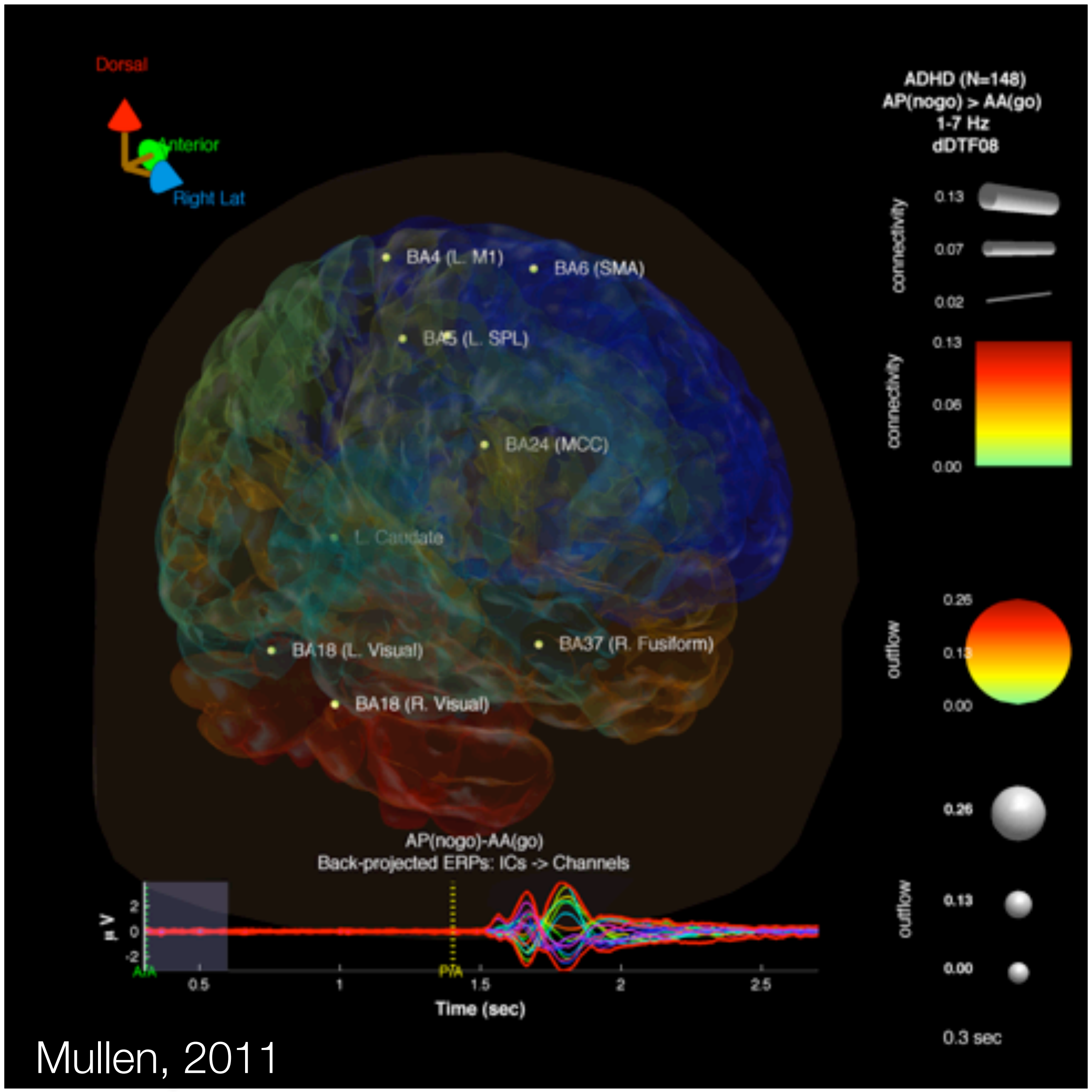
Break

- SIFT Walkthrough and Hands-on Practicum

The Dynamic Brain

- A key goal: To model temporal changes in neural dynamics and information flow that index and predict task-relevant changes in cognitive state and behavior

- **Open Challenges:**
  - Non-invasive measures (source inference)
  - Robustness and Validity (constraints & statistics)
  - Scalability (multivariate)
  - Temporal Specificity / Non-stationarity / Single-trial (dynamics)
  - Multi-subject Inference
  - Usability and Data Visualization (software)



Mullen, 2011

Modeling Brain Connectivity

- Model-based approaches mitigate the ‘curse of dimensionality’ by making some assumptions about the structure, dynamics, or statistics of the system under observation

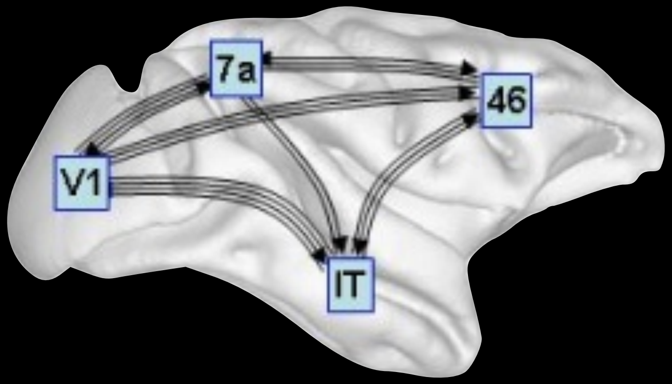
Box and Draper (1987):

> “Essentially, all models are wrong, but some are useful [...] the practical question is how wrong do they have to be to not be useful”

Categorizations of Large-Scale Brain Connectivity Analysis

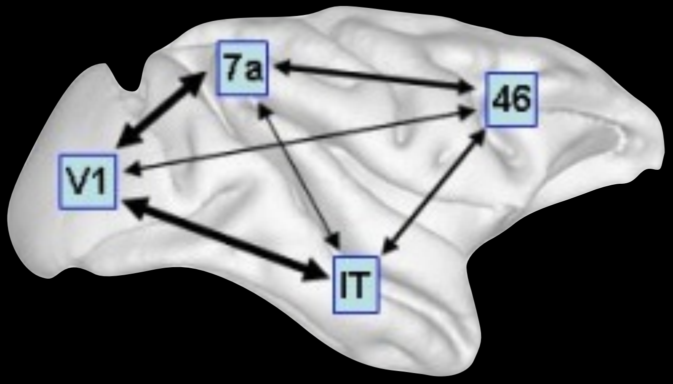
(Bullmore and Sporns, *Nature*, 2009)

Structural



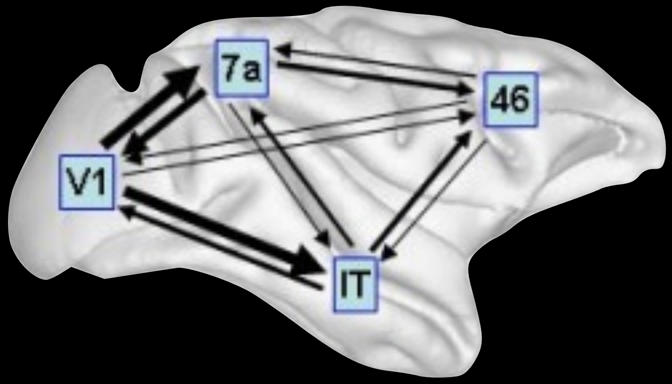
state-invariant, anatomical

Functional



dynamic, state-dependent, correlative, symmetric

Effective



dynamic, state-dependent, asymmetric, causal, information flow

Hours-Years **milliseconds-seconds**

Temporal Scale

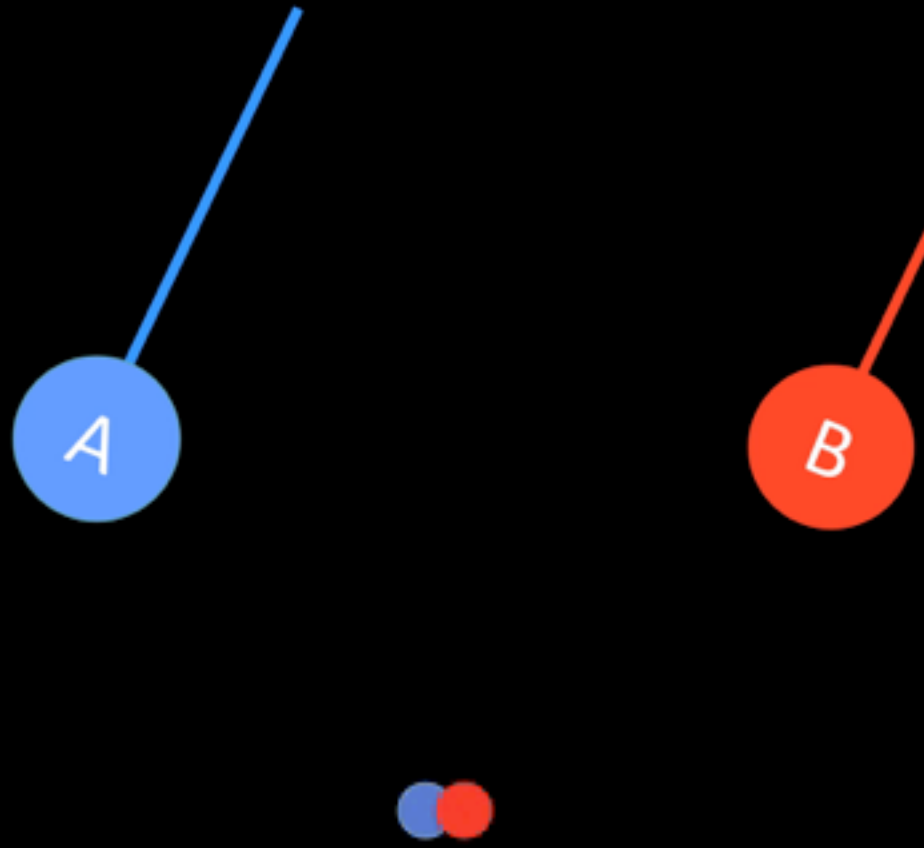
Estimating Functional Connectivity

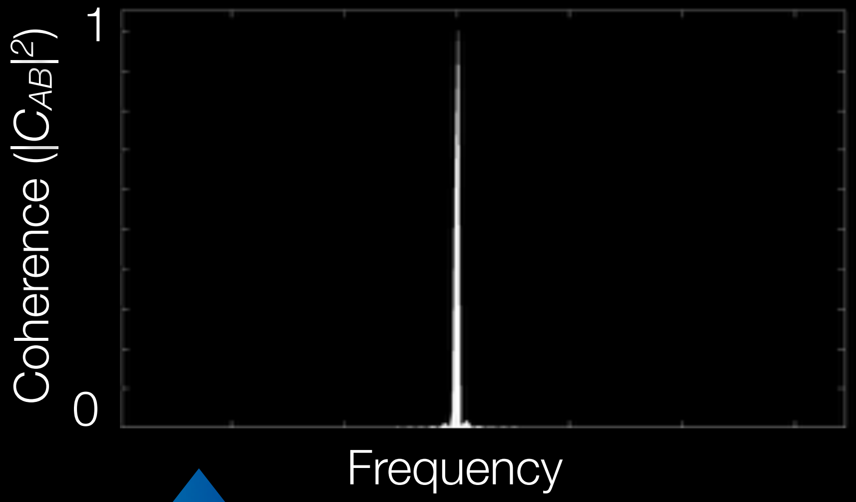
Popular measures

- Cross-Correlation
- Coherence
- Phase-Locking Value
- Phase-amplitude coupling

…

(Cross)-Correlation ≠ Causation

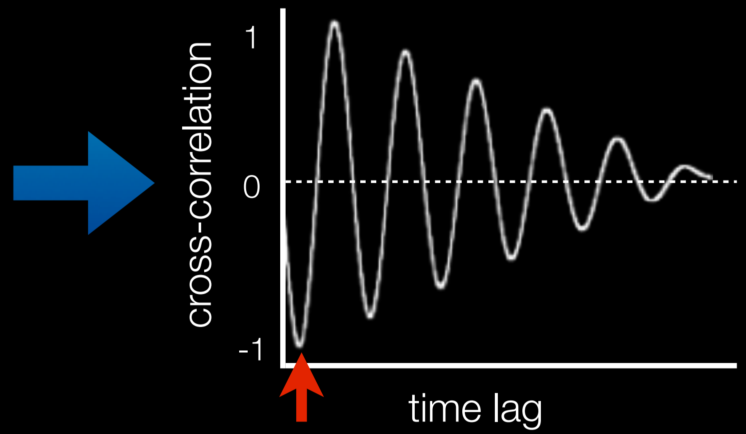




DFT

$$C_{AB}(f) = \sum_{k=0}^{p} r_{AB}(k) e^{-i2pfk}$$

$$= \frac{S_{AB}(f)}{\sqrt{S_A(f) S_B(f)}}$$



$$r_{AB}(k)$$

Coherence/CC/PLV indicate ***functional***, but not ***effective*** connectivity

Estimating Effective Connectivity

Non-Invasive

- *Post-hoc* analyses applied to measured neural activity
- Confirmatory
  - Dynamic Causal Models
  - Structural Equation Models
- Exploratory
  - Granger-Causal methods

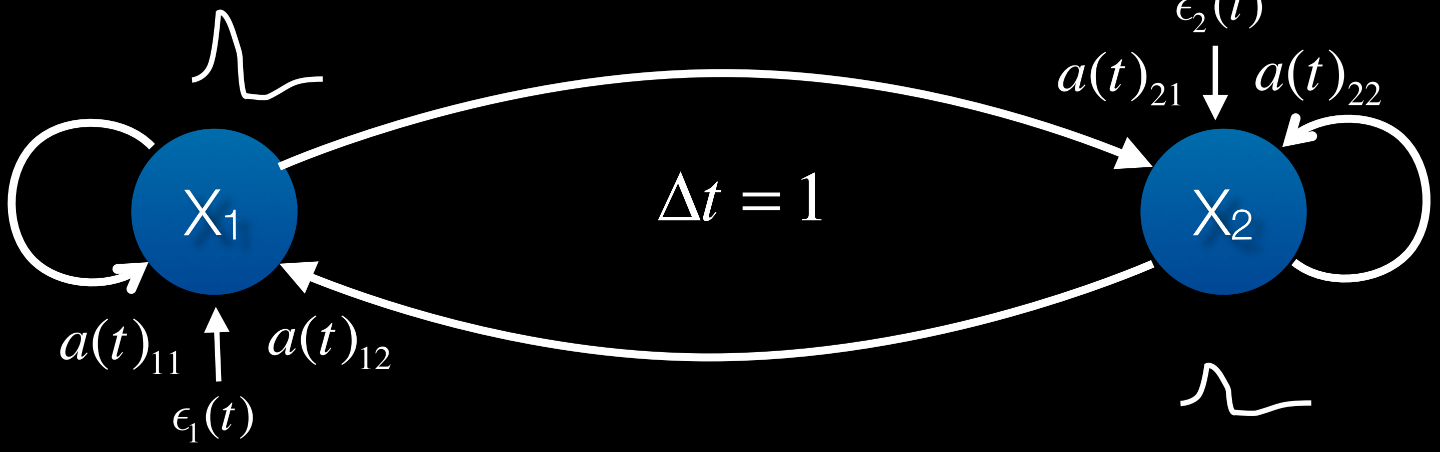
- Data-driven
- Rooted in *conditional predictability*
- Scalable (Valdes-Sosa, 2005)
- Extendable to nonlinear and/or non-stationary systems (Freiwald, 1999; Ding, 2001; Chen, 2004; Ge, 2009)
- Extendable to non-parametric representations (Dhamala, 2009a,b)
- Can be (partially) controlled for (unobserved) exogenous causes (Guo, 2008a,b; Ge, 2009)
- Equivalent to Transfer Entropy for Gaussian Variables (Seth, 2009)
- Flexibly allows us to examine **time-varying** (dynamic) multivariate causal relationships in either the **time** or **frequency** domain

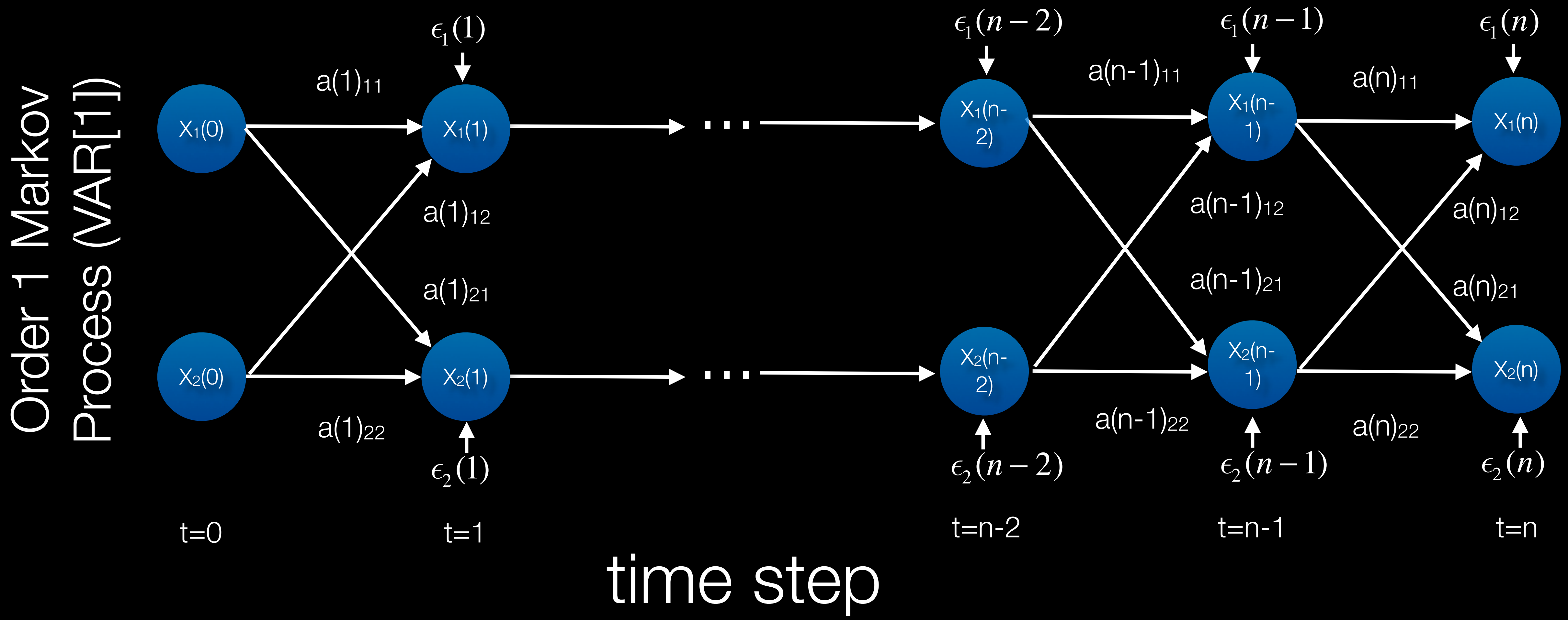
Linear Dynamical Systems

Stochastic Linear Dynamical System

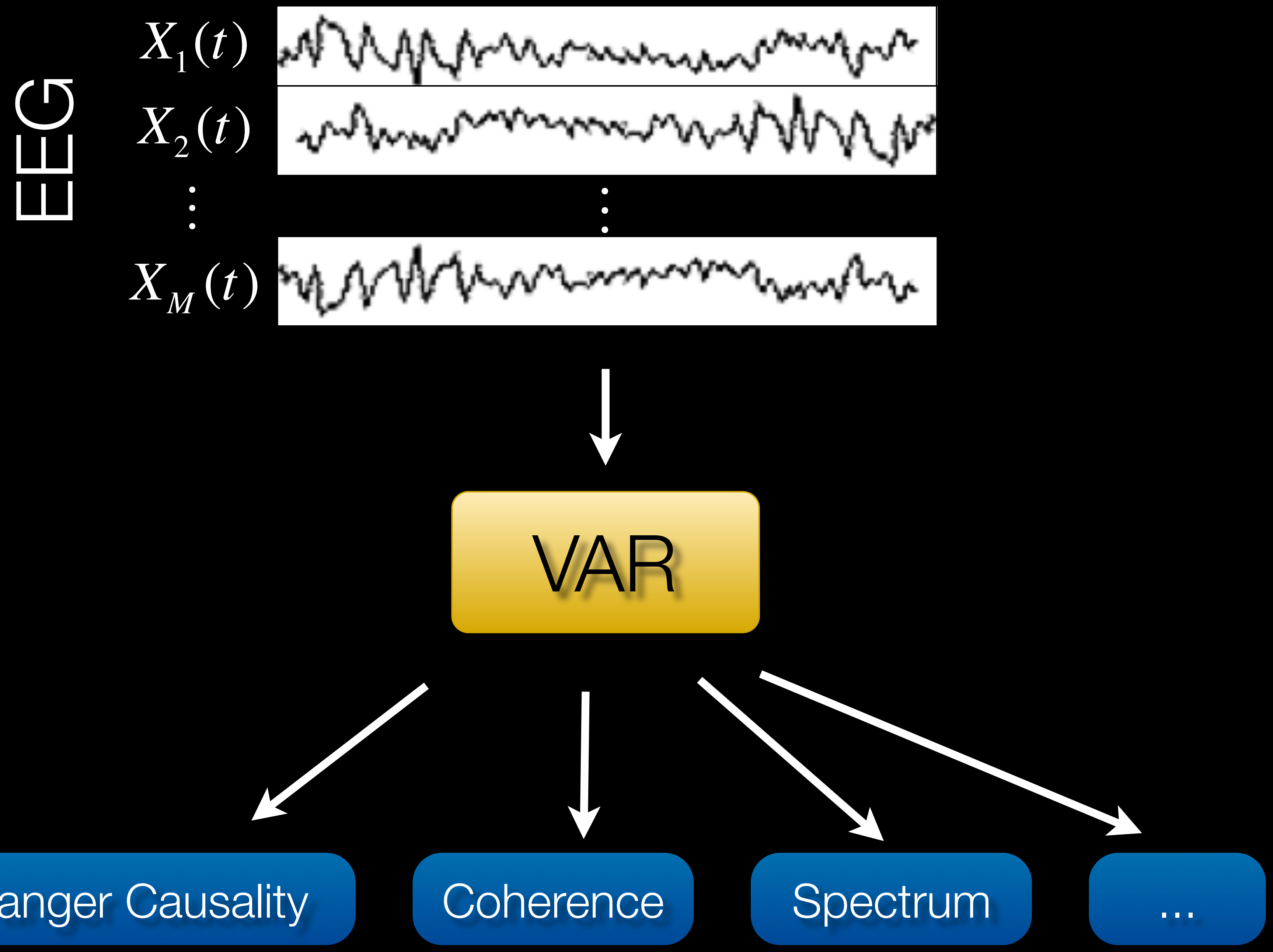
$$X_1(t) = a(t)_{11} X_1(t-1) + a(t)_{12} X_2(t-1) + \epsilon_1(t)$$

$$X_2(t) = a(t)_{22} X_2(t-1) + a(t)_{21} X_1(t-1) + \epsilon_2(t)$$





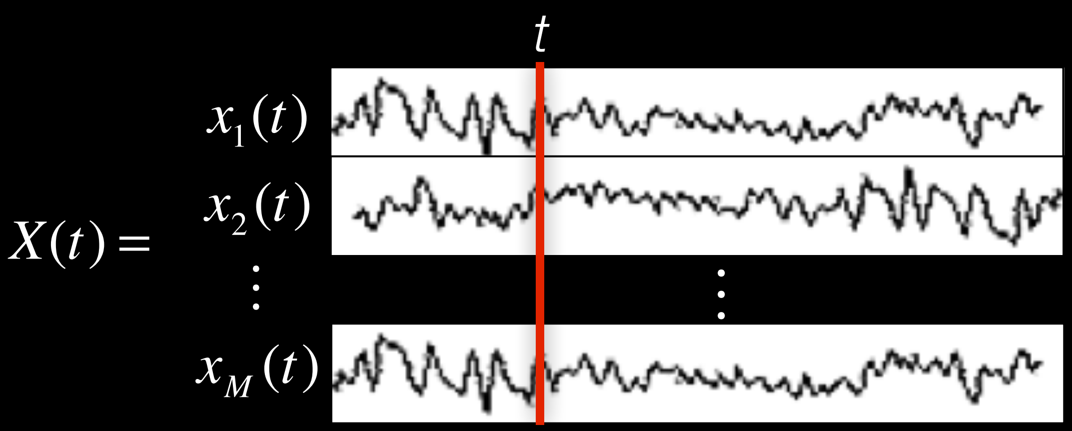
Vector Autoregressive (VAR / MAR / MVAR) Modeling



VAR Modeling: Assumptions

- **“Weak” stationarity of the data**
  - mean and variance do not change with time
  - An EEG trace containing prominent evoked potentials is a classic example of a non-stationary time-series
- **Stability**
  - All eigenvalues of the system matrix are ≤1
  - A stable process will not “blow up” (diverge to infinity)
  - A stable model is always a stationary model (however, the converse is not necessarily true). If a stable model adequately fits the data (white residuals), then the data is likewise stationary

The Linear VAR Model



Ordinary Least-Squares
Lattice Filters
Kalman Filtering
Bayesian Methods
Sparse methods
…

VAR[*p*] model

model order

random noise process

$$\mathbf{X}(t) = \sum_{k=1}^{p} \mathbf{A}^{(k)}(t)\mathbf{X}(t-k) + \mathbf{E}(t)$$

M-channel data vector at current time *t*

M x M matrix of (possibly time-varying) model coefficients indicating variable dependencies at lag *k*

multichannel data *k* samples in the past

$$\mathbf{A}^{(k)}(t) = \begin{pmatrix} a_{11}^{(k)}(t) & \dots & a_{1M}^{(k)}(t) \\ \vdots & \ddots & \vdots \\ a_{M1}^{(k)}(t) & \cdots & a_{MM}^{(k)}(t) \end{pmatrix}$$

$$\mathbf{E}(t) = N(0, \mathbf{V})$$

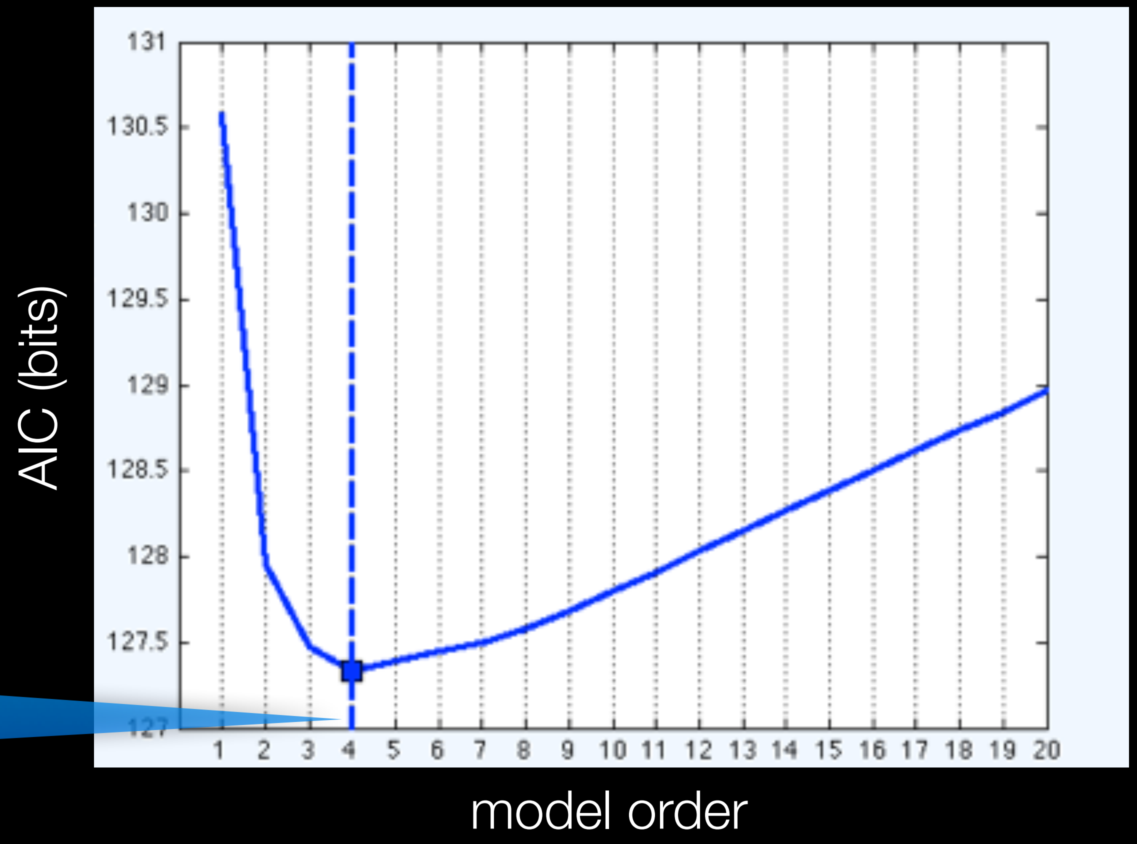
Selecting a VAR Model Order

- Model order is typically determined by minimizing information criteria such as Akaike Information Criterion (AIC) for varying model order (p):

$$\mathrm{AIC}(p) = 2\log(\det(\mathbf{V})) + M^2p/N$$

Penalizes high model orders (parsimony)

entropy rate (amount of prediction error)



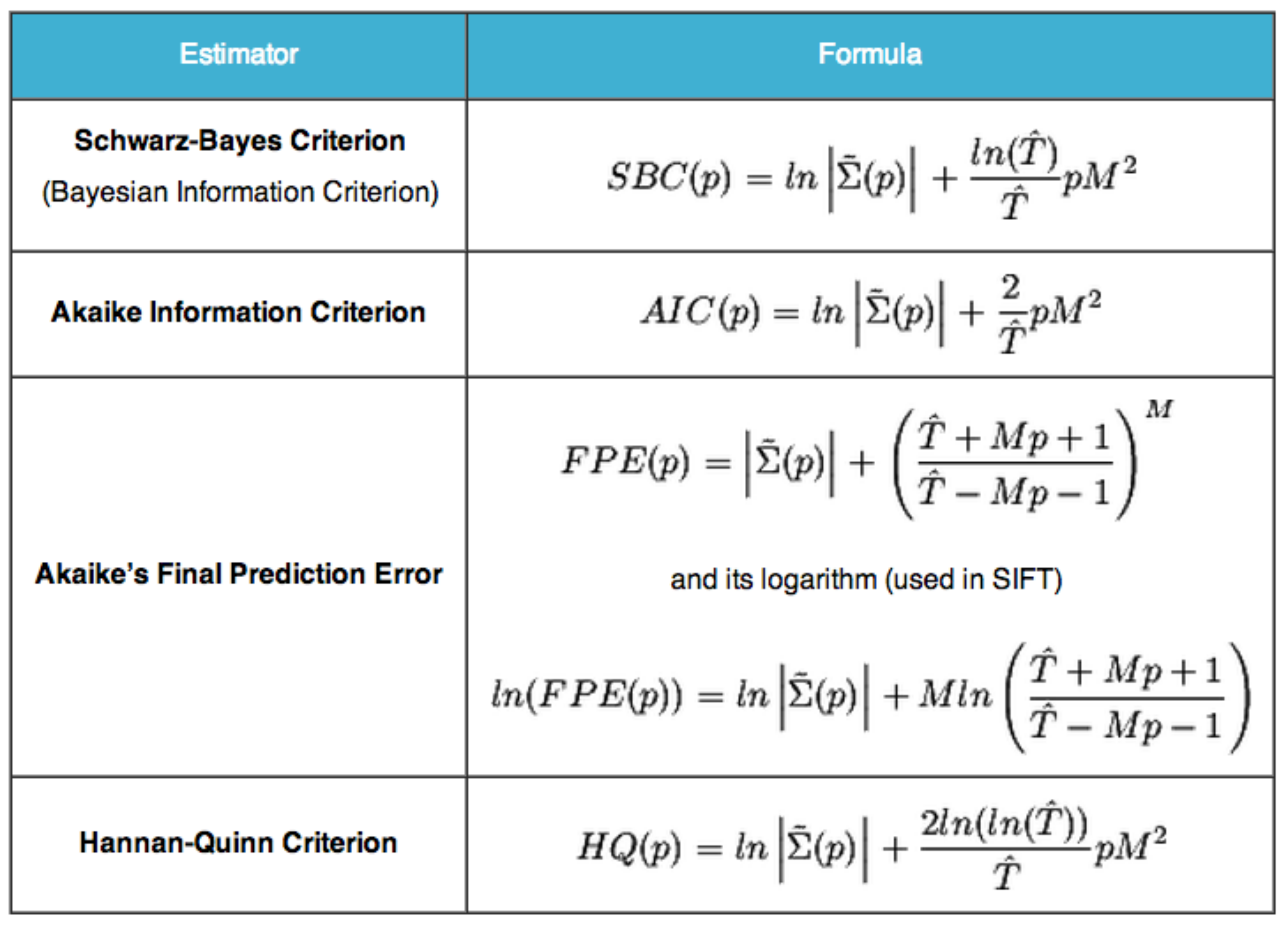
optimal order

Model Order Selection Criteria

More Conservative

Less Conservative

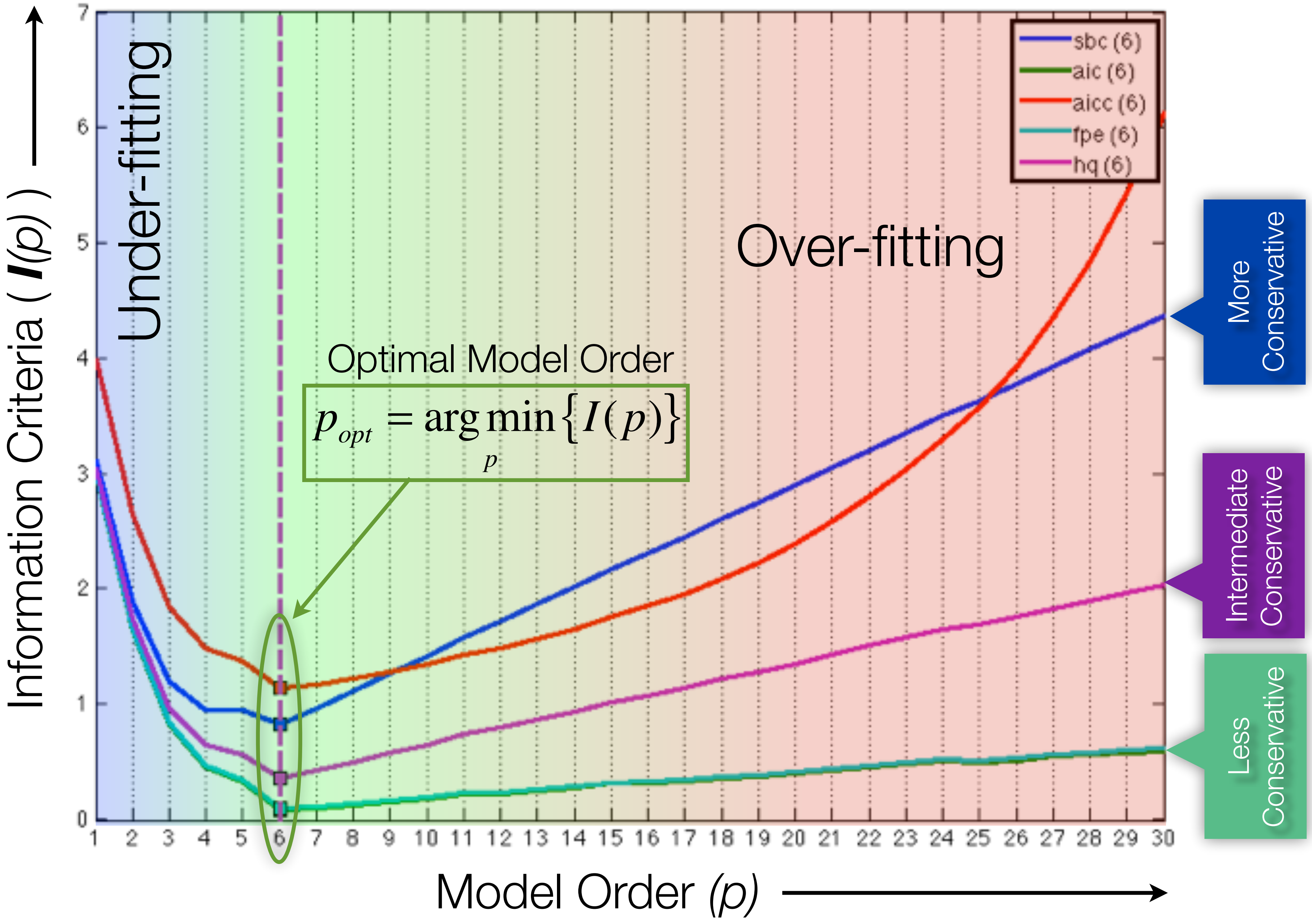
Intermediate Conservative



Estimator	Formula
Schwarz-Bayes Criterion (Bayesian Information Criterion)	$SBC(p) = ln\left\|\tilde{\Sigma}(p)\right\| + \frac{ln(\hat{T})}{\hat{T}}pM^2$
Akaike Information Criterion	$AIC(p) = ln\left\|\tilde{\Sigma}(p)\right\| + \frac{2}{\hat{T}}pM^2$
Akaike's Final Prediction Error	$FPE(p) = \left\|\tilde{\Sigma}(p)\right\| + \left(\frac{\hat{T}+Mp+1}{\hat{T}-Mp-1}\right)^M$ and its logarithm (used in SIFT) $ln(FPE(p)) = ln\left\|\tilde{\Sigma}(p)\right\| + M ln\left(\frac{\hat{T}+Mp+1}{\hat{T}-Mp-1}\right)$
Hannan-Quinn Criterion	$HQ(p) = ln\left\|\tilde{\Sigma}(p)\right\| + \frac{2ln(ln(\hat{T}))}{\hat{T}}pM^2$

Model Order Selection Criteria

$I(p)$ = [Prediction Error] + [Overfitting Penalty]

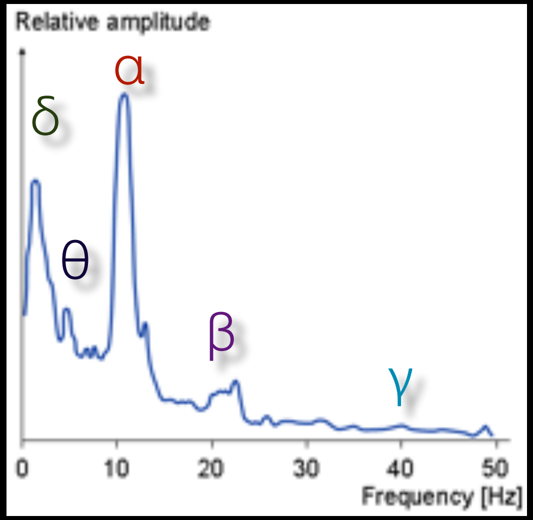


Optimal Model Order

$$p_{opt} = \arg\min_{p} \{ I(p) \}$$

Selecting a VAR Model Order

- Other considerations:
  - A M-dimensional VAR model of order *p* has at most *Mp/2* spectral peaks distributed amongst the *M* variables. This means we can observe at most *p/2* peaks in each variables' spectrum (or in the cross spectrum between each pair of variables)



  - Optimal model order depends on sampling rate. Higher sampling rate often requires higher model orders.

Model Validation

- If a model is poorly fit to data, then few, if any, inferences can be validly drawn from the model. There a number of criteria which we can use to determine whether we have appropriately fit our VAR model. Here are three commonly used categories of tests:
- **Whiteness Tests:** checking the residuals of the model for serial and cross-correlation
- **Consistency Test:** testing whether the model generates data with same correlation structure as the real data
- **Stability Test:** checking the stability/stationarity of the model.

Granger Causality

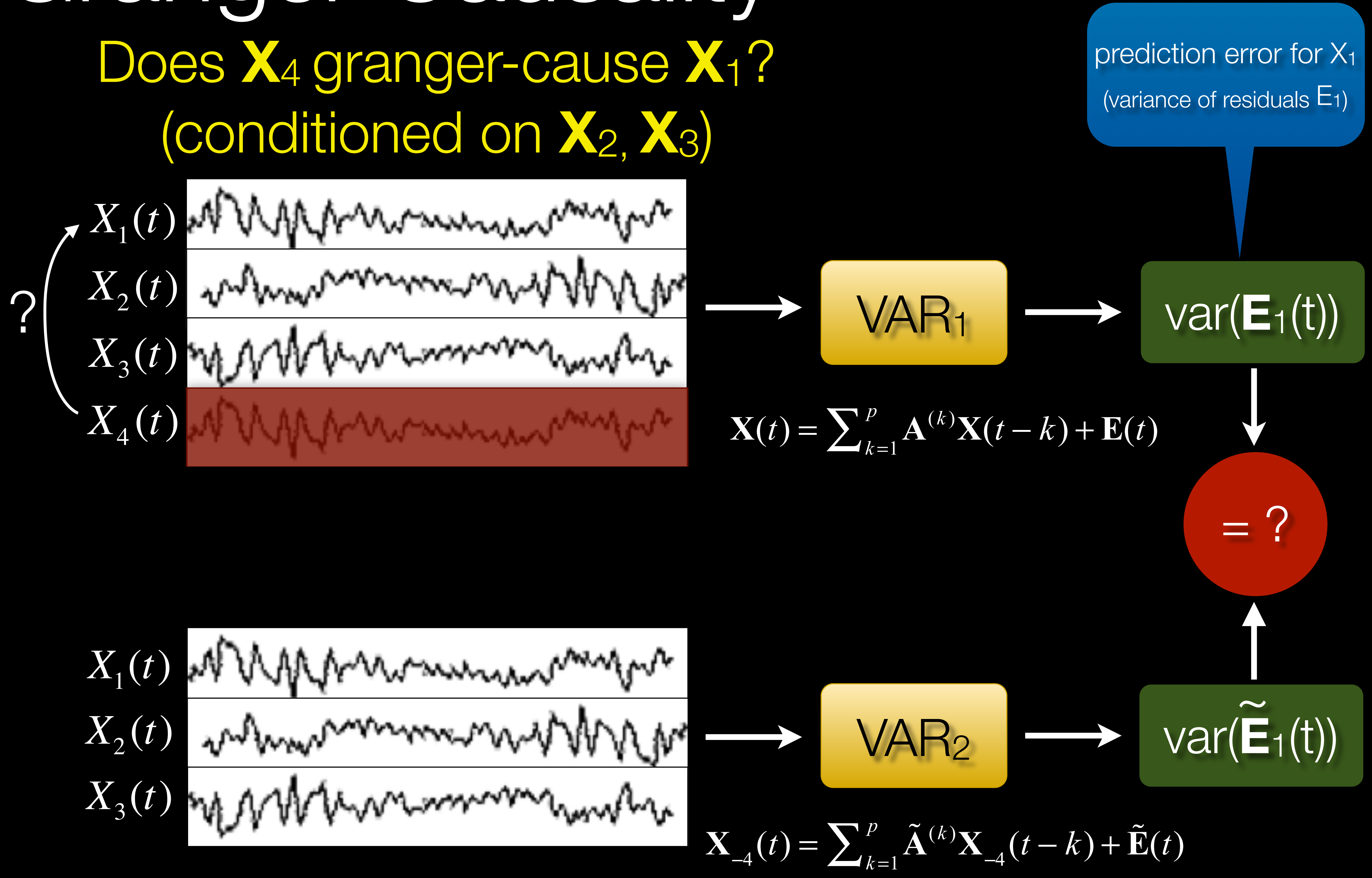
- First introduced by Wiener (1958). Later reformulated by Granger (1969) in the context of linear stochastic autoregressive models
- Relies on two assumptions:

Granger Causality Axioms
1. Causes should precede their effects in time (Temporal Precedence)
2. Information in a cause's past should improve the prediction of the effect, above and beyond the information contained in past of the effect (and other measured variables)

Granger Causality

Does $\mathbf{X}_4$ granger-cause $\mathbf{X}_1$?

(conditioned on $\mathbf{X}_2$, $\mathbf{X}_3$)



prediction error for X_1

(variance of residuals E_1)

VAR$_1$

var($\mathbf{E}_1$(t))

= ?

VAR$_2$

var($\tilde{\mathbf{E}}_1$(t))

$$\mathbf{X}(t) = \sum_{k=1}^{p} \mathbf{A}^{(k)} \mathbf{X}(t-k) + \mathbf{E}(t)$$

$$\mathbf{X}_{-4}(t) = \sum_{k=1}^{p} \tilde{\mathbf{A}}^{(k)} \mathbf{X}_{-4}(t-k) + \tilde{\mathbf{E}}(t)$$

Granger Causality

- Granger (1969) quantified this definition for **bivariate** processes in the form of an F-ratio:

reduced model

$$F_{X_1 \leftarrow X_2} = \ln\left(\frac{var(\tilde{E}_1)}{var(E_1)}\right) = \ln\left(\frac{var(X_1(t) \mid X_1(\cdot))}{var(X_1(t) \mid X_1(\cdot), X_2(\cdot))}\right)$$

full model

- Alternately, for a **multivariate interpretation** we can fit a single VAR model to all channels and apply the following definition:

Definition 1

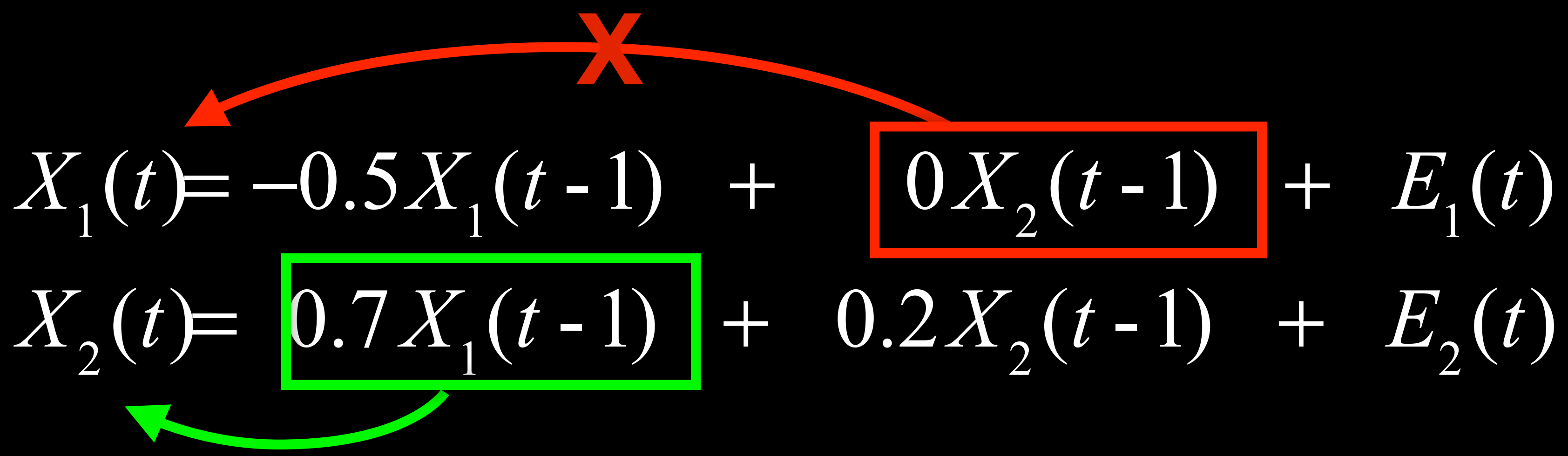
X_j granger-causes X_i conditioned on all other variables in **X**

if and only if $\mathbf{A}_{ij}(k) >> 0$ for some lag $k \in \{1, \ldots, p\}$

Granger Causality Quiz

- Example: 2-channel VAR process of order 1

$$\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} -0.5 & \enclose{circle}{0} \\ \enclose{circle}{0.7} & 0.2 \end{pmatrix} \begin{pmatrix} X_1(t-1) \\ X_2(t-1) \end{pmatrix} + \begin{pmatrix} E_1(t) \\ E_2(t) \end{pmatrix}$$



$$\begin{aligned} X_1(t) &= -0.5 X_1(t-1) + \boxed{0 X_2(t-1)} + E_1(t) \\ X_2(t) &= \boxed{0.7 X_1(t-1)} + 0.2 X_2(t-1) + E_2(t) \end{aligned}$$

Which causal structure does this model correspond to?

a) ① → ②   b) ① ← ②   c) ① ↔ ②

Granger Causality – Frequency Domain

$$\mathbf{X}(t) = \sum_{k=1}^{p} \mathbf{A}^{(k)} \mathbf{X}(t-k) + \mathbf{E}(t)$$

Fourier-transforming $\mathbf{A}^{(k)}$ we obtain

Likewise, **X**(*f*) and **E**(*f*) correspond to the fourier transforms of the data and residuals, respectively

$$\mathbf{A}(f) = -\sum_{k=0}^{p} \mathbf{A}^{(k)} e^{-i2p fk}; \mathbf{A}^{(0)} = I$$

We can then define the spectral matrix **X**(*f*) as follows:

$$\mathbf{X}(f) = \mathbf{A}(f)^{-1}\mathbf{E}(f) = \mathbf{H}(f)\mathbf{E}(f)$$

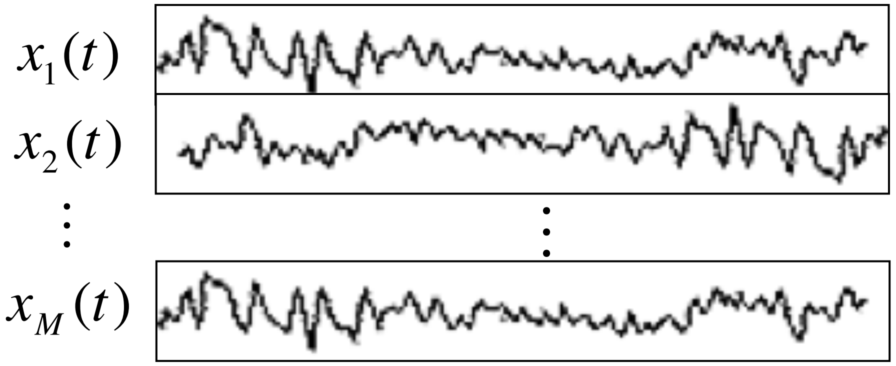
Where **H**(*f*) is the *transfer matrix* of the system.

Definition 2

X_j granger-causes X_i conditioned on all other variables in **X**

if and only if $|\mathbf{A}_{ij}(f)| >> 0$ for some frequency *f*

leads to PDC

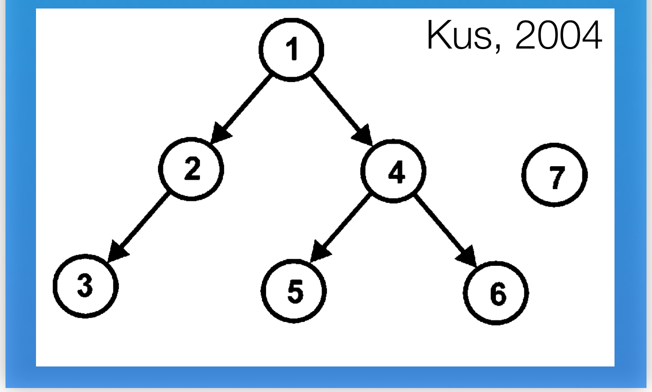


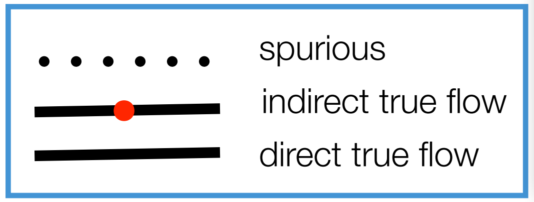
$$\mathbf{X}(t) = \sum_{k=1}^{p} \mathbf{A}^{(k)}(t)\mathbf{X}(t-k) + \mathbf{E}(t)$$

$$\mathbf{A}(f,t) = -\sum_{k=0}^{p} \mathbf{A}^{(k)}(t)e^{-i2pfk}; \quad \mathbf{A}^{(0)} = I$$

$$\mathbf{X}(f,t) = \mathbf{A}(f,t)^{-1}\mathbf{E}(f,t) = \mathbf{H}(f,t)\mathbf{E}(f,t)$$

Ground Truth

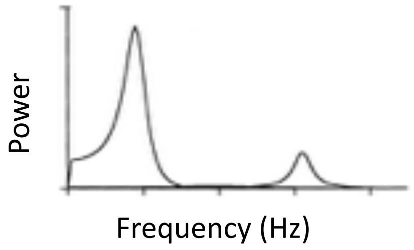




Spectrum

$$S(f) = \mathbf{X}(f)\mathbf{X}(f)^{*}$$
$$= \mathbf{H}(f)\Sigma\mathbf{H}(f)^{*}$$

(Brillinger, 2001)



NOTE: time index (t) dropped for convenience

Functional

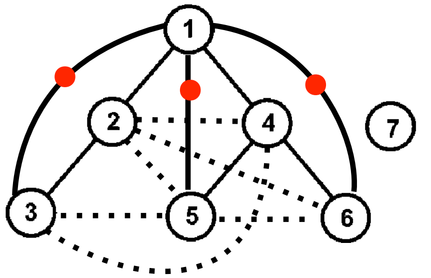
Effective

Bivariate

Coherency

$$C_{ij}(f) = \frac{S_{ij}(f)}{\sqrt{S_{ii}(f)S_{jj}(f)}}$$

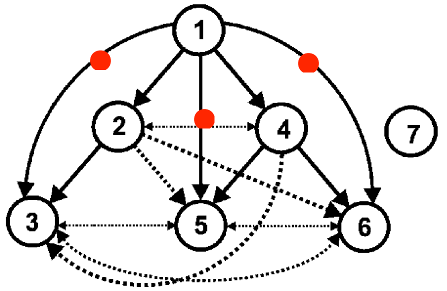
(Bendat and Piersol, 1986)



Granger-Geweke Causality

$$F_{ij}(f) = \frac{\Sigma_{jj} - (\Sigma_{ij}^{2} / \Sigma_{ii}))\,|\,H_{ij}(f)\,|^{2}}{S_{ii}(f)}$$

(Geweke, 1982; Bressler et al., 2007)

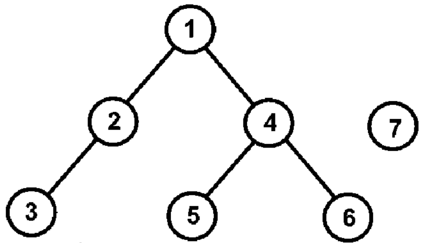


Multivariate

Partial Coherence

$$P_{ij}(f) = \frac{S_{ij}^{-1}(f)}{\sqrt{S_{ii}^{-1}(f)S_{jj}^{-1}(f)}}$$

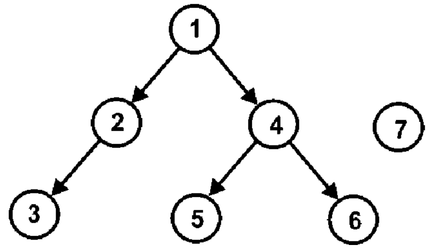
(Bendat and Piersol, 1986; Dalhaus, 2000)



Partial Directed Coherence

$$p_{ij}^{2}(f) = \frac{|\,A_{ij}(f)\,|^{2}}{\sum_{k=1}^{M}\left|A_{kj}(f)\right|^{2}}$$

(Baccalá and Sameshima, 2001)



	Estimator	Formula
Spectral M.	Spectral Density Matrix	$S(f)=X(f)X(f)^*$ $=H(f)\Sigma H(f)^*$
Coherence Measures	Coherency	$C_{ij}(f)=\frac{S_{ij}(f)}{\sqrt{S_{ii}(f)S_{jj}(f)}}$ $0\le\lvert C_{ij}(f)\rvert^2\le1$
	Imaginary Coherence (iCoh)	$iCoh_{ij}(f)=\mathrm{Im}(C_{ij}(f))$
	Partial Coherence (pCoh)	$P_{ij}(f)=\frac{\hat{S}_{ij}(f)}{\sqrt{\hat{S}_{ii}(f)\hat{S}_{jj}(f)}}$ $\hat{S}(f)=S(f)^{-1}$ $0\le\lvert P_{ij}(f)\rvert^2\le1$
	Multiple Coherence (mCoh)	$G_i(f)=\sqrt{1-\frac{\det(S(f))}{S_{ii}(f)\mathbf{M}_{ii}(f)}}$ $\mathbf{M}_{ii}(f)$ is the **minor** of *S*(*f*) obtained by removing the ith row and column of *S*(*f*) and returning the determinant.

	Estimator	Formula
Partial Directed Coherence Measures	Normalized Partial Directed Coherence (PDC)	$\pi_{ij}(f)=\frac{A_{ij}(f)}{\sqrt{\sum_{k=1}^{M}\lvert A_{kj}(f)\rvert^2}}$ $0\le\lvert\pi_{ij}(f)\rvert^2\le1$ $\sum_{j=1}^{M}\lvert\pi_{ij}(f)\rvert^2=1$
	Generalized PDC (GPDC)	$\bar{\pi}_{ij}(f)=\frac{\frac{1}{\Sigma_{ii}}A_{ij}(f)}{\sqrt{\sum_{k=1}^{M}\frac{1}{\Sigma_{ii}^2}\lvert A_{kj}(f)\rvert^2}}$ $0\le\lvert\bar{\pi}_{ij}(f)\rvert^2\le1$ $\sum_{j=1}^{M}\lvert\bar{\pi}_{ij}(f)\rvert^2=1$
	Renormalized PDC (rPDC)	$\lambda_{ij}(f)=Q_{ij}(f)^*V_{ij}(f)^{-1}Q_{ij}(f)$ where $Q_{ij}(f)=\begin{pmatrix}\mathrm{Re}[A_{ij}(f)]\\ \mathrm{Im}[A_{ij}(f)]\end{pmatrix}$ and $V_{ij}(f)=\sum_{k,l=1}^{p}R_{jj}^{-1}(k,l)\Sigma_{ii}Z(2\pi f,k,l)$ $Z(\omega,k,l)$ $=\begin{pmatrix}\cos(\omega k)\cos(\omega l) & \cos(\omega k)\sin(\omega l)\\ \sin(\omega k)\cos(\omega l) & \sin(\omega k)\sin(\omega l)\end{pmatrix}$ ***R*** is the [(*Mp*)² x (*Mp*)²] covariance matrix of the VAR[*p*] process (Lütkepohl, 2006)
Granger-Geweke	Granger-Geweke Causality (GGC)	$F_{ij}(f)=\frac{\left(\Sigma_{jj}-(\Sigma_{ij}^2/\Sigma_{ii})\right)\lvert H_{ij}(f)\rvert^2}{S_{ii}(f)}$

	Estimator	Formula
Directed Transfer Function Measures	Normalized Directed Transfer Function (DTF)	$\gamma_{ij}(f)=\frac{H_{ij}(f)}{\sqrt{\sum_{k=1}^{M}\lvert H_{ik}(f)\rvert^2}}$ $0\le\lvert\gamma_{ij}(f)\rvert^2\le1$ $\sum_{j=1}^{M}\lvert\gamma_{ij}(f)\rvert^2=1$
	Full-Frequency DTF (ffDTF)	$\eta_{ij}^2(f)=\frac{\lvert H_{ij}(f)\rvert^2}{\sum_f\sum_{k=1}^{M}\lvert H_{ik}(f)\rvert^2}$
	Direct DTF (dDTF)	$\delta_{ij}^2(f)=\eta_{ij}^2(f)P_{ij}^2(f)$

$$\mathbf{X}(t)=\sum_{k=1}^{p}\mathbf{A}^{(k)}(t)\mathbf{X}(t-k)+\mathbf{E}(t)$$

$$\mathbf{A}(f,t)=-\sum_{k=0}^{p}\mathbf{A}^{(k)}(t)e^{-i2pfk};\ \ \mathbf{A}^{(0)}=I$$

$$\mathbf{X}(f,t)=\mathbf{A}(f,t)^{-1}\mathbf{E}(f,t)=\mathbf{H}(f,t)\mathbf{E}(f,t)$$

$H(f)$ Transfer Function

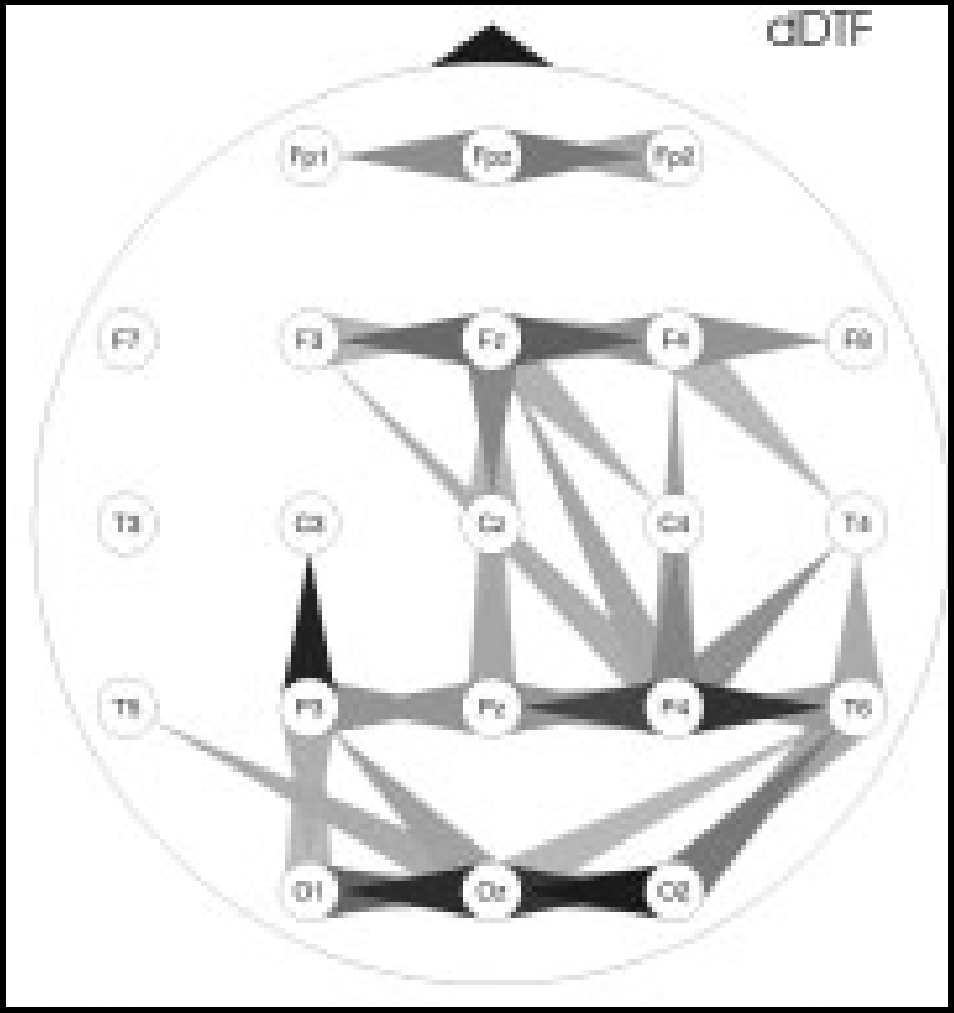
$A(f)$ System Matrix

Σ Noise Covariance Matrix

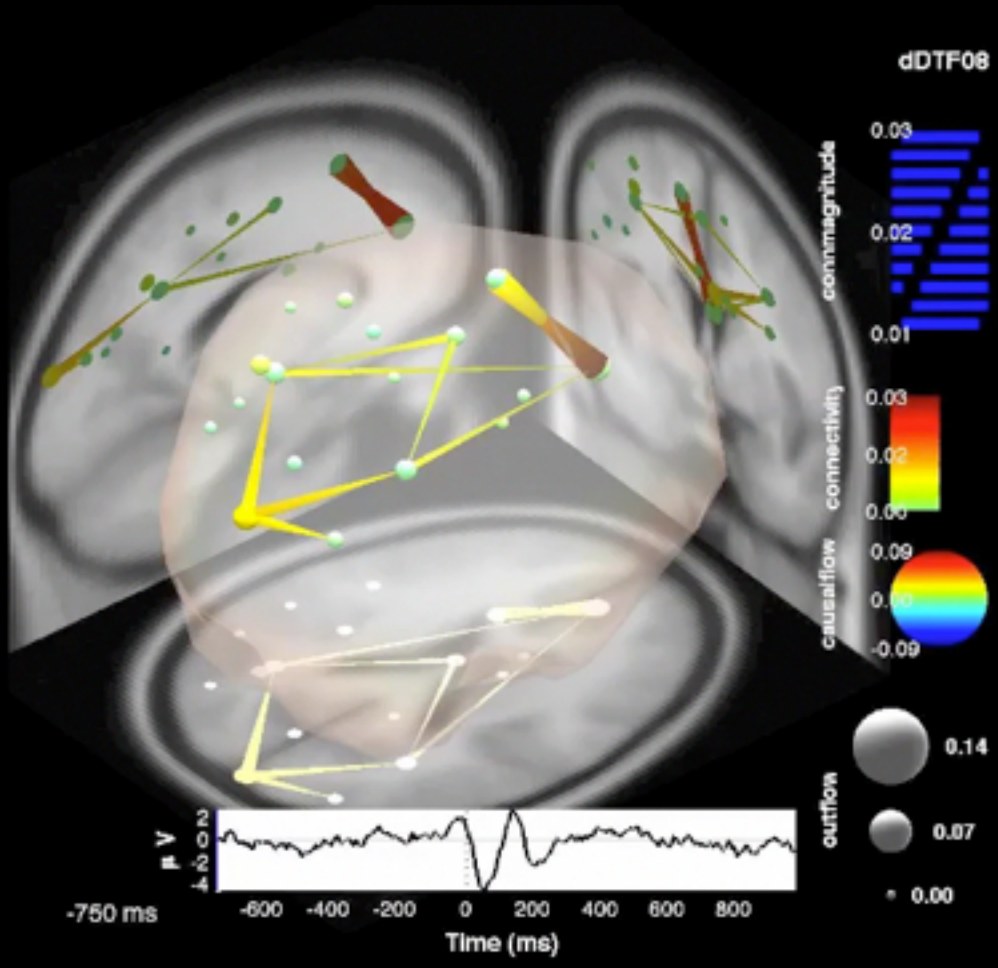
🟨 Variance Stabilization

For additional details, see SIFT Handbook (sccn.ucsd.edu/wiki/SIFT)

Scalp or Source?

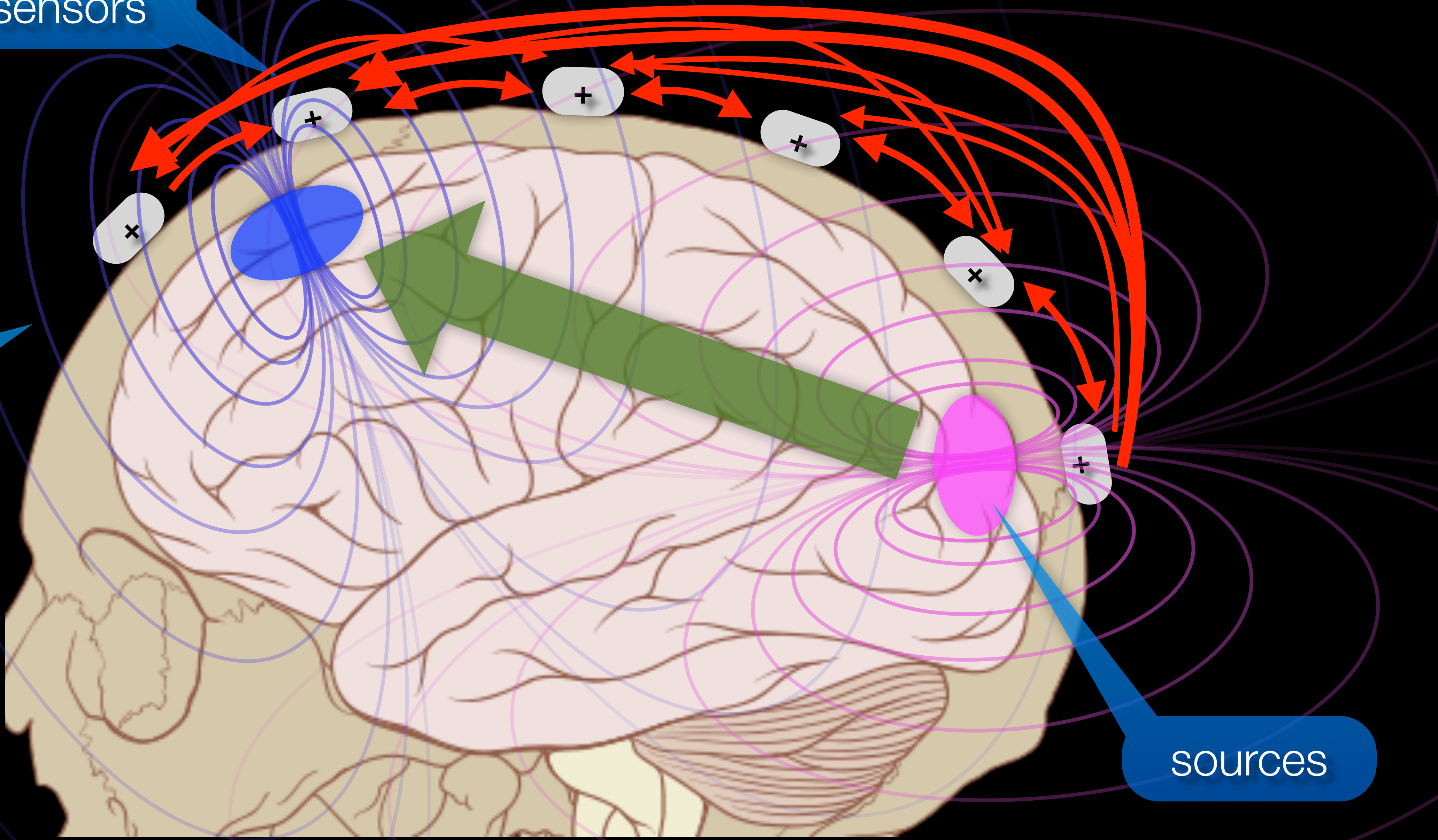


or



Scalp or Source?

$$X(t) = HS(t) = \sum_{k=1}^{p} \boxed{HA^{(k)}(t)H^{-1}} X(t-k) + HE(t)$$



sensors

$$X(t) = \overset{H^{-1}}{\boxed{H}}\boxed{S(t)}$$

Volume Conduction

ICA
SBL
Beamforming
Minimum-norm
...

sources

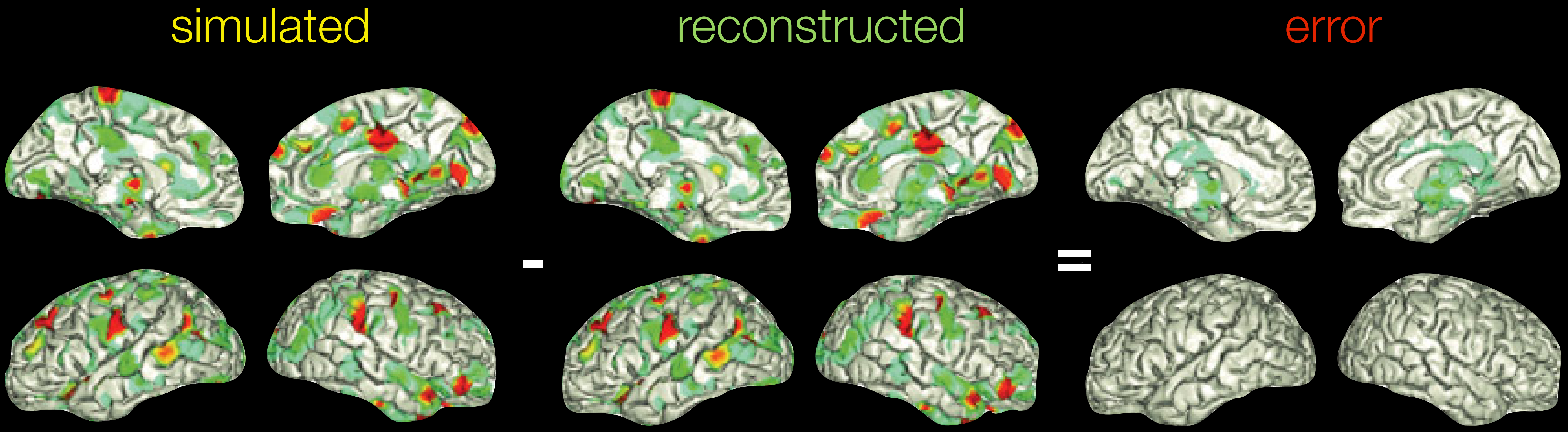
Solution? Source Separation

$$S(t) = \sum_{k=1}^{p} \boxed{A^{(k)}(t)} S(t-k) + E(t)$$

Forward/Inverse Modeling

Method	Smoothness	Sparsity	Independence/Orthogonality
MNE	X		
LORETA	X		
dSPM	X		
Beamforming			X
Sparse Bayesian Learning	X	X	
S-FLEX	X	X	
FOCUSS		X	
ICA/PCA/SOBI			X

Source reconstruction with ICA+SBL



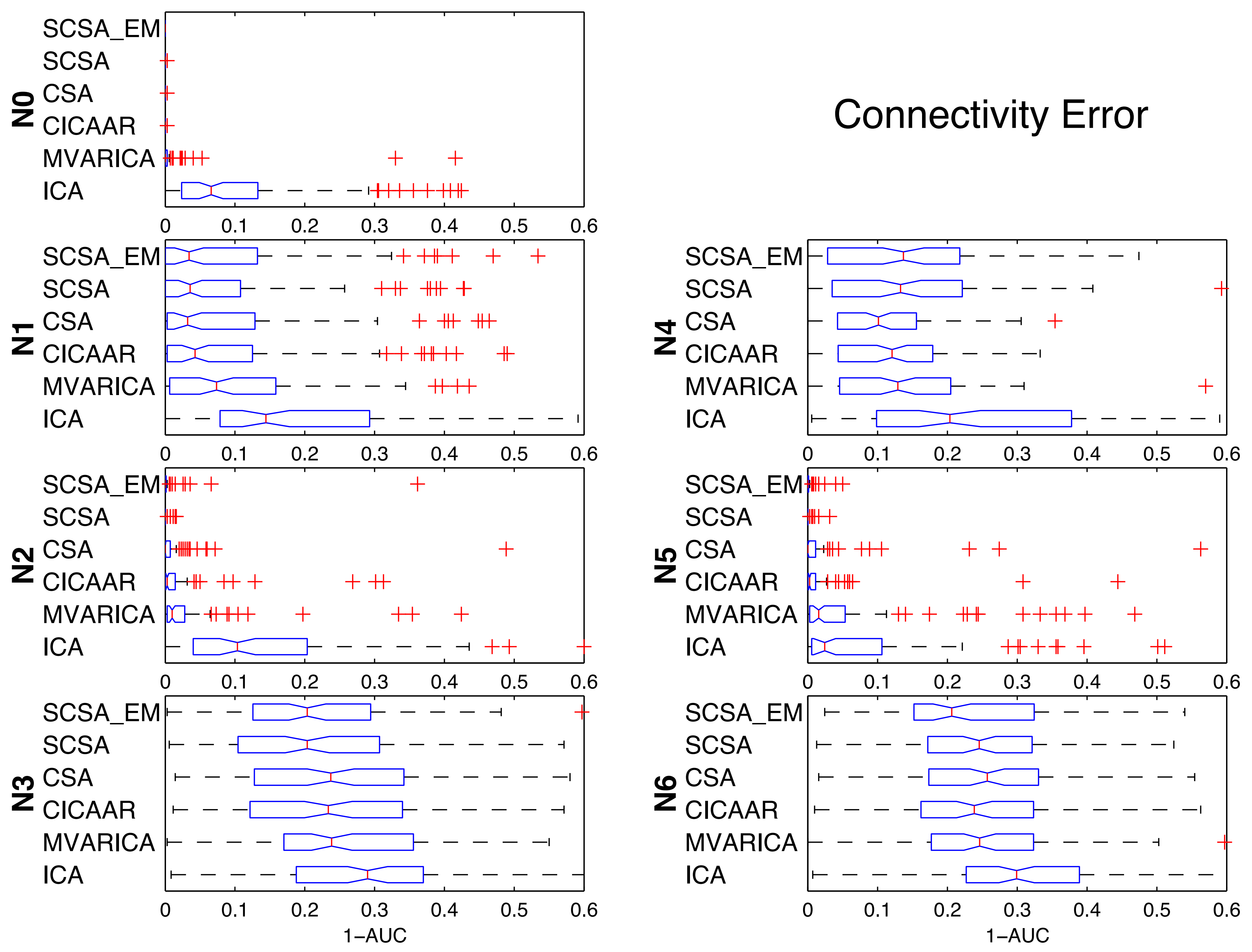
Makeig, Ramirez, Weber, Wipf, Dale, Simpson, *15th Inter. Conf on Biomagnetism* (2006)

Estimating Dependency of Independent Components ?

- Isn’t it a contradiction to examine dependence between Independent/ Uncorrelated Components?
- Instantaneous (e.g., Infomax) ICA only explicitly seeks to maximize *instantaneous* independence. Time-delayed dependencies may be preserved.
- Infomax ICA seeks to maximize *global* independence (over entire recording session), transient dependencies may be preserved.
- Independence is a very strict criterion that cannot be achieved *in general* by a linear transformation (such as ICA). Instead, dependent variables will form a **dependent subspace**.

However, the *best* approach is to use an inverse model that explicitly preserves time-delayed dependencies or *jointly* estimates sources (de-mixing matrix) and connectivity (VAR parameters). See Haufe, 2008 IEEE TBME for a good treatment (implemented as mvar_scsa in SIFT 2.0).

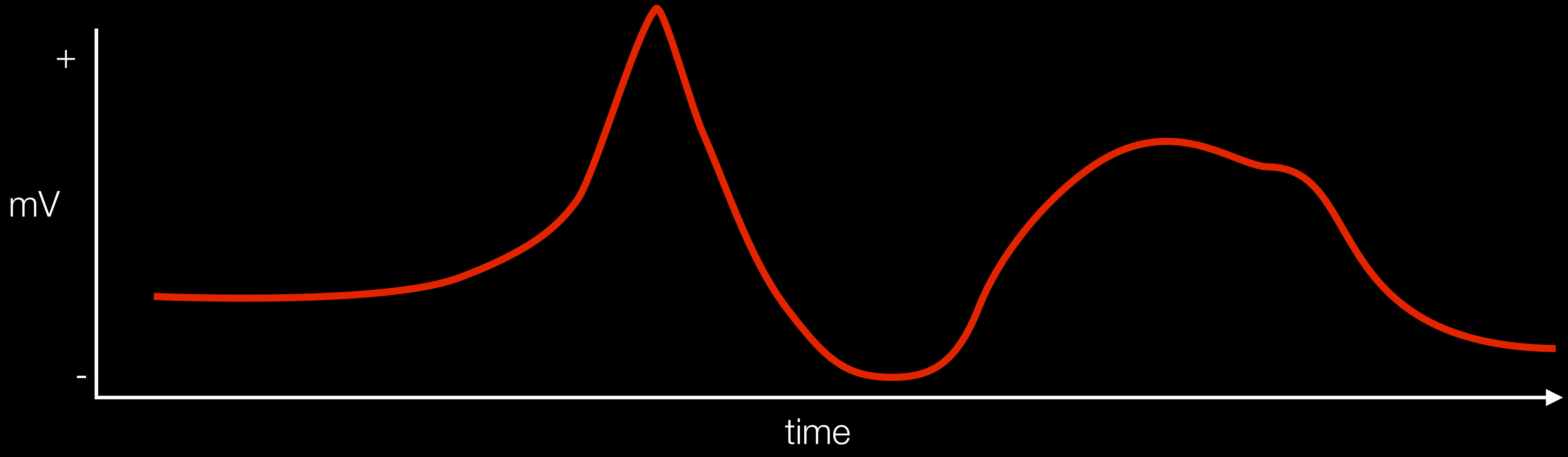
Estimating Dependency of Independent Components ?



Haufe et al, IEEE TBME 2008

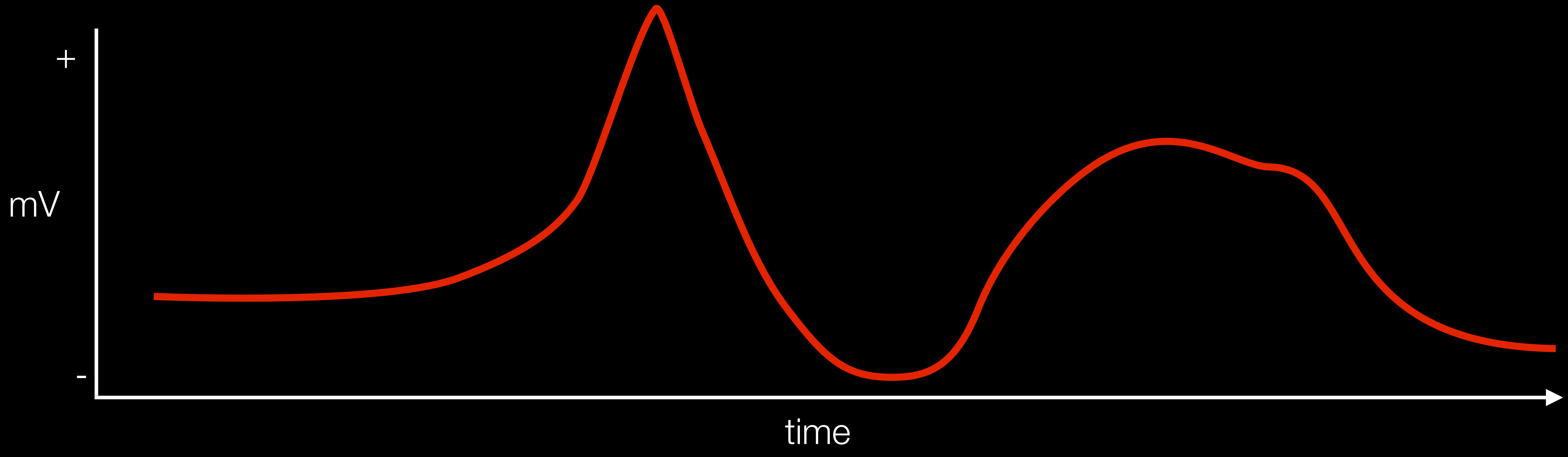
Adapting to Non-Stationarity

- The brain is a **dynamic system** and measured brain activity and coupling can change rapidly with time (non-stationarity)
  - event-related perturbations (ERSP, ERP, etc)
  - structural changes due to learning/feedback
- How can we adapt to non-stationarity?



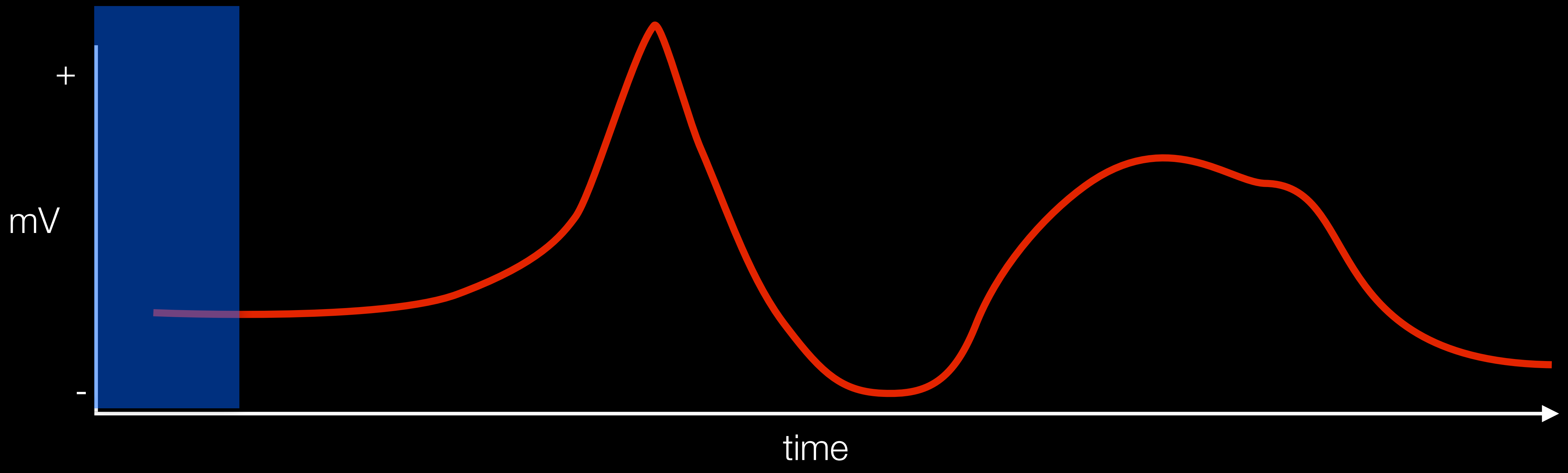
Adapting to Non-Stationarity

- **Many ways to do adaptive VAR estimation**
- Two popular approaches (adopted in SIFT):
  - Segmentation-based adaptive VAR estimation (assumes local stationarity)
  - State-Space Modeling



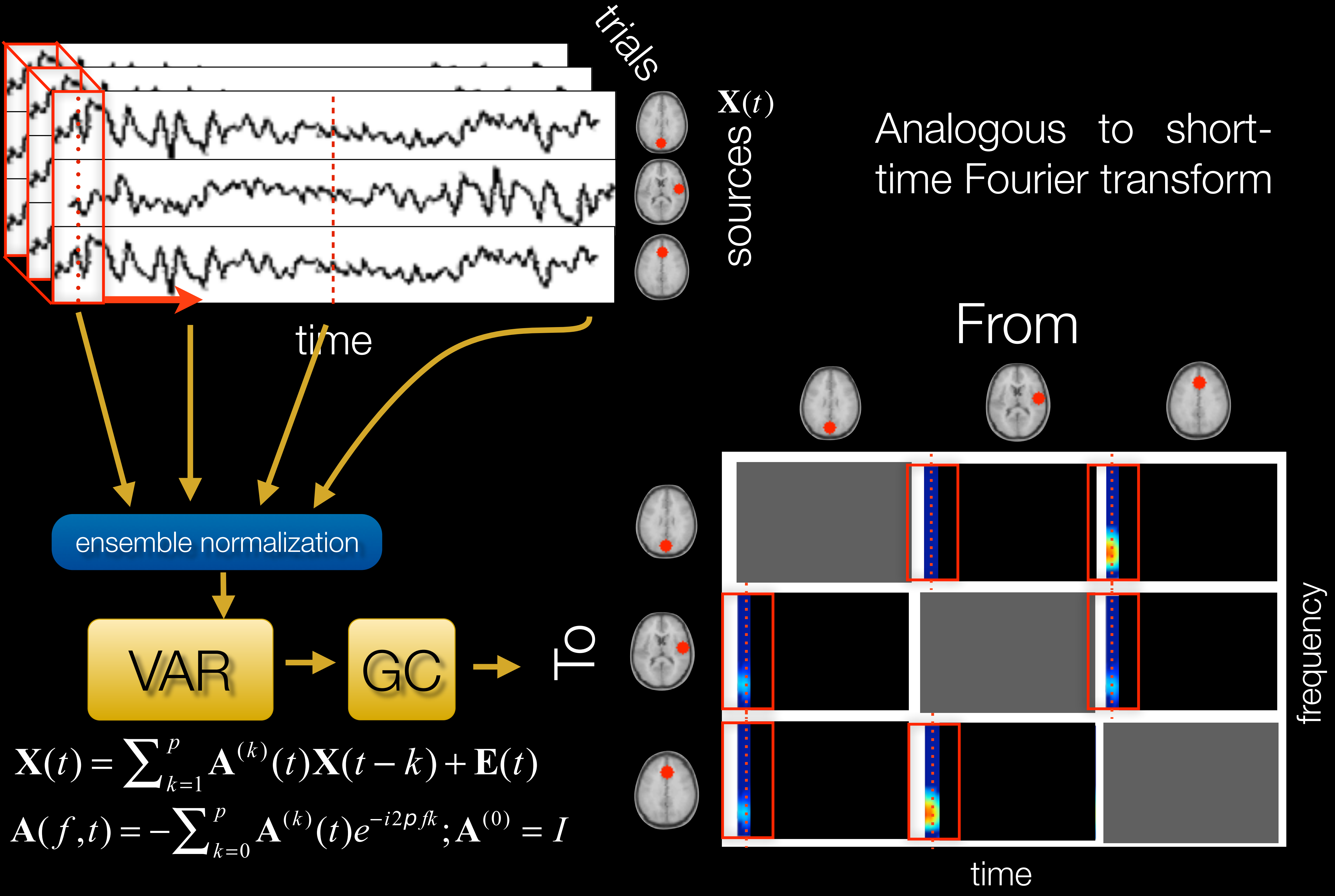
Adapting to Non-Stationarity

- **Many ways to do adaptive VAR estimation**
- Two popular approaches (adopted in SIFT):
  - **Segmentation-based adaptive VAR estimation (assumes local stationarity)**
  - State-Space Modeling



Segmentation-based VAR

(Jansen et al., 1981; Florian and Pfurtscheller, 1995; Ding et al, 2000)



Analogous to short-time Fourier transform

$$\mathbf{X}(t) = \sum_{k=1}^{p} \mathbf{A}^{(k)}(t)\mathbf{X}(t-k) + \mathbf{E}(t)$$

$$\mathbf{A}(f,t) = -\sum_{k=0}^{p} \mathbf{A}^{(k)}(t)e^{-i2pfk}; \mathbf{A}^{(0)} = I$$

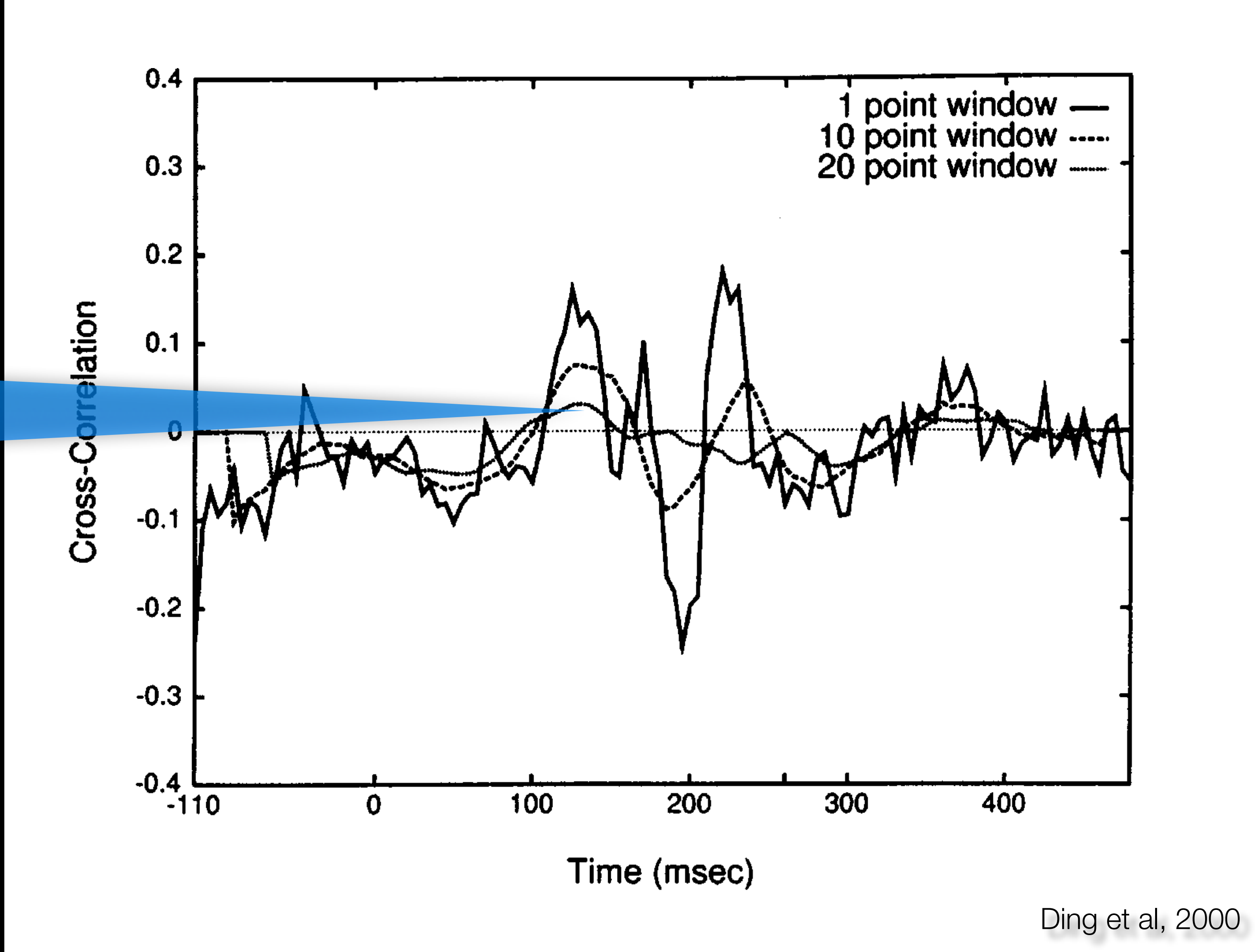
Adapting to Non-Stationarity

- **What is a good window length?**
- Considerations:
  - Temporal smoothing
  - Local stationarity
  - Sufficient amount of data
  - Process dynamics

Adapting to Non-Stationarity

Consideration: Temporal Smoothness

Too-large windows may smooth out interesting transient dynamic features.

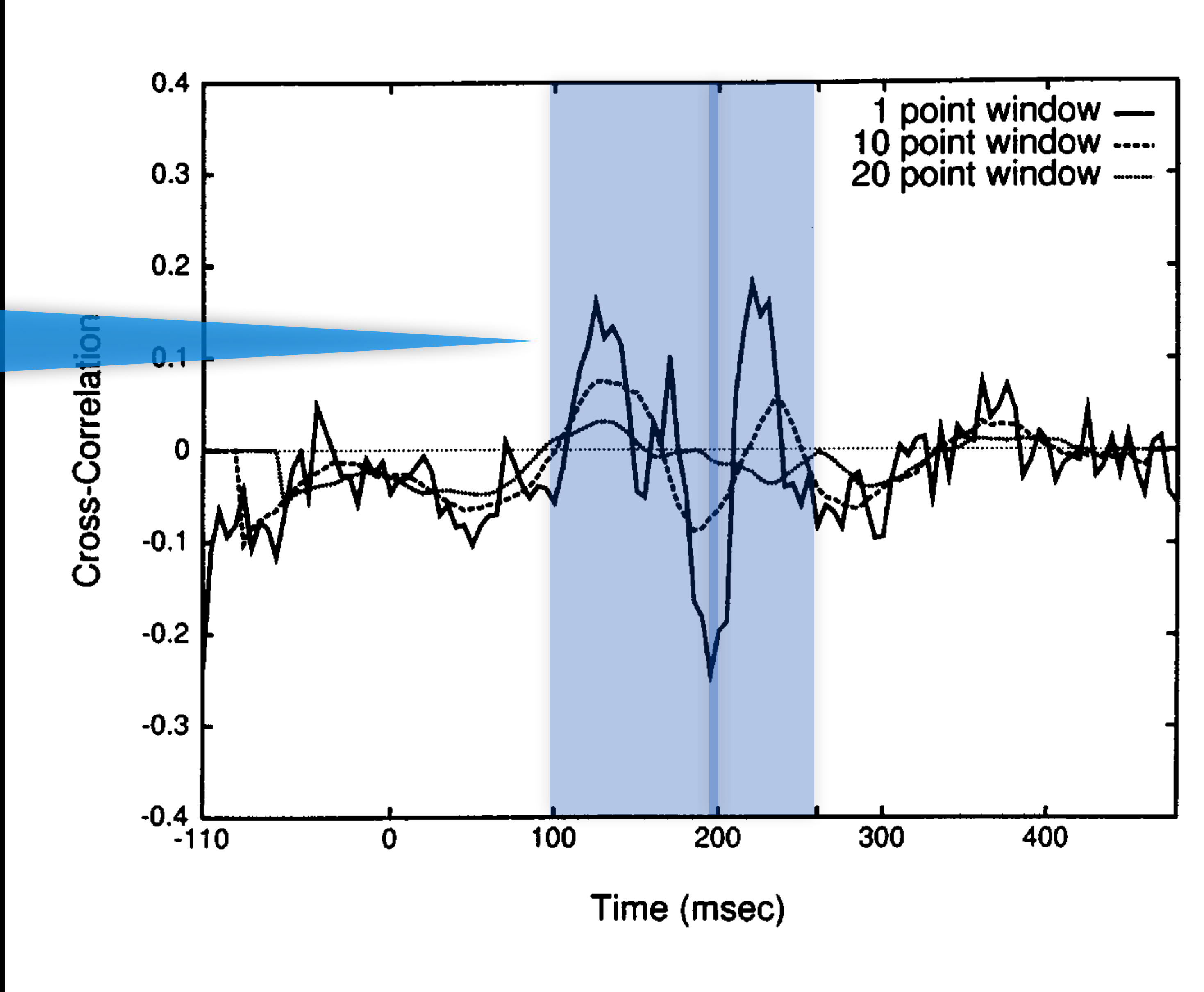


Ding et al, 2000

Adapting to Non-Stationarity

Consideration: Local Stationarity

Too-large windows may not be locally-stationary



Adapting to Non-Stationarity

Consideration: Sufficient data

M = number of variables
p = model order
N_{tr} = number of trials
W = length of each window (sample points)

We have M^2p model coefficients to estimate. This requires a minimum of M^2p independent samples.
So we have the constraint $M^2p <= N_{tr} W$.
In practice, however, a better heuristic is $M^2p <= (1/10)N_{tr} W$.

Or: **$W >= 10(M^2p/N_{tr})$**

10x more data points than parameters to estimate

SIFT will let you know if your window length is not optimal

Regularization

- But what if $W < (M^2p/N_{tr})$?
  - single/few trials or continuous data
  - short time window
  - large number of model variables (channels/sources, high model order)
- There are insufficient observations to uniquely determine a solution to the system of equations defining our model and the problem becomes *ill-posed* or *under-determined*.

Regularization

Solutions?

Make assumptions (impose constraints)

We want to *a priori* restrict the range of allowable values for our parameters -- transforming the problem from one with infinite number of solutions in the original parameter space to one with a unique (“best”) solution in the new parameter space

In a Bayesian context, this corresponds to making assumptions about the *prior distribution* of the parameters (Gaussian, Laplacian, ...)

Regularization

Solutions?

Make assumptions (impose constraints)

- Smoothness Constraints (Gaussian prior)
  - e.g. *Ridge Regression*

- Sparsity Constraints (Laplacian prior or mixed prior)
  - e.g. *Group Lasso*

Constraints Improve Estimation
(if prior assumptions are correct)

- Significant improvements using smoothness or sparsity assumptions
- (e.g. Haufe et al, 2009, Valdez-Sosa et al, 2009)

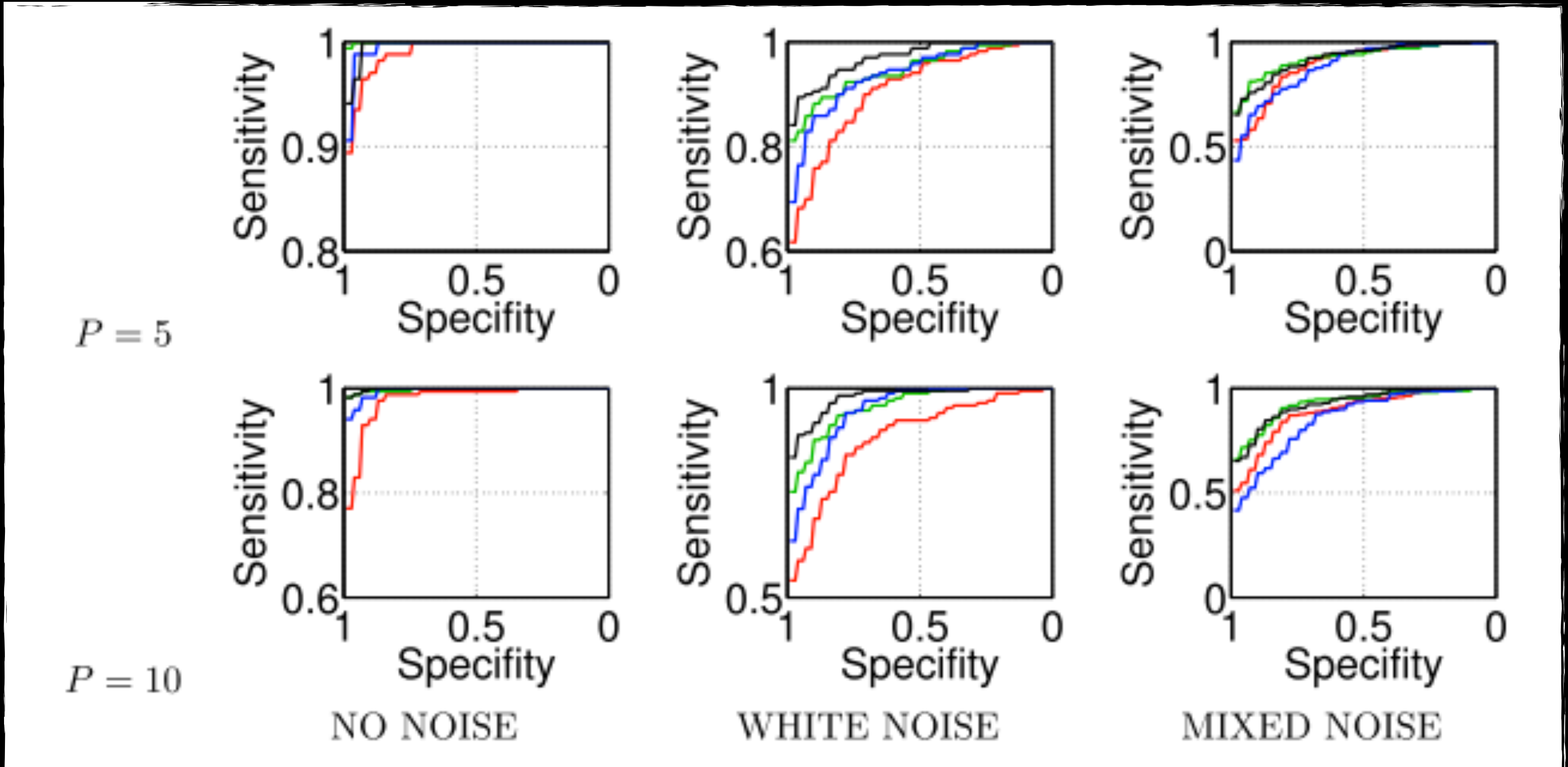


Figure 2: Average ROC curves of Granger Causality (red), Ridge Regression (green), Lasso (blue) and Group Lasso (black) in three different noise conditions and for two different model orders.

Haufe, 2009

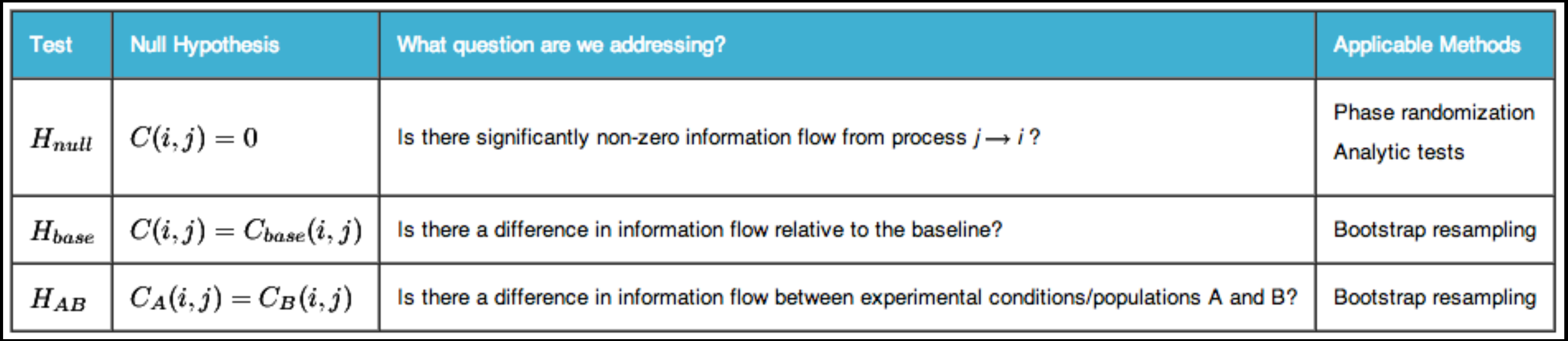
Adapting to Non-Stationarity

Consideration: Process dynamics

- Your window must be larger than the maximum expected interaction time lag between any two processes.
- Your window should be large enough to span ~1 cycle of the lowest frequency of interest (remember the Heisenberg uncertainty principle: increased time resolution —> reduced frequency resolution)

Statistics

- Different ways to do statistics in SIFT
  - Phase Randomization
  - Bootstrapping
  - Analytic Tests



Test	Null Hypothesis	What question are we addressing?	Applicable Methods
H_{null}	$C(i,j) = 0$	Is there significantly non-zero information flow from process $j \longrightarrow i$?	Phase randomization Analytic tests
H_{base}	$C(i,j) = C_{base}(i,j)$	Is there a difference in information flow relative to the baseline?	Bootstrap resampling
H_{AB}	$C_A(i,j) = C_B(i,j)$	Is there a difference in information flow between experimental conditions/populations A and B?	Bootstrap resampling

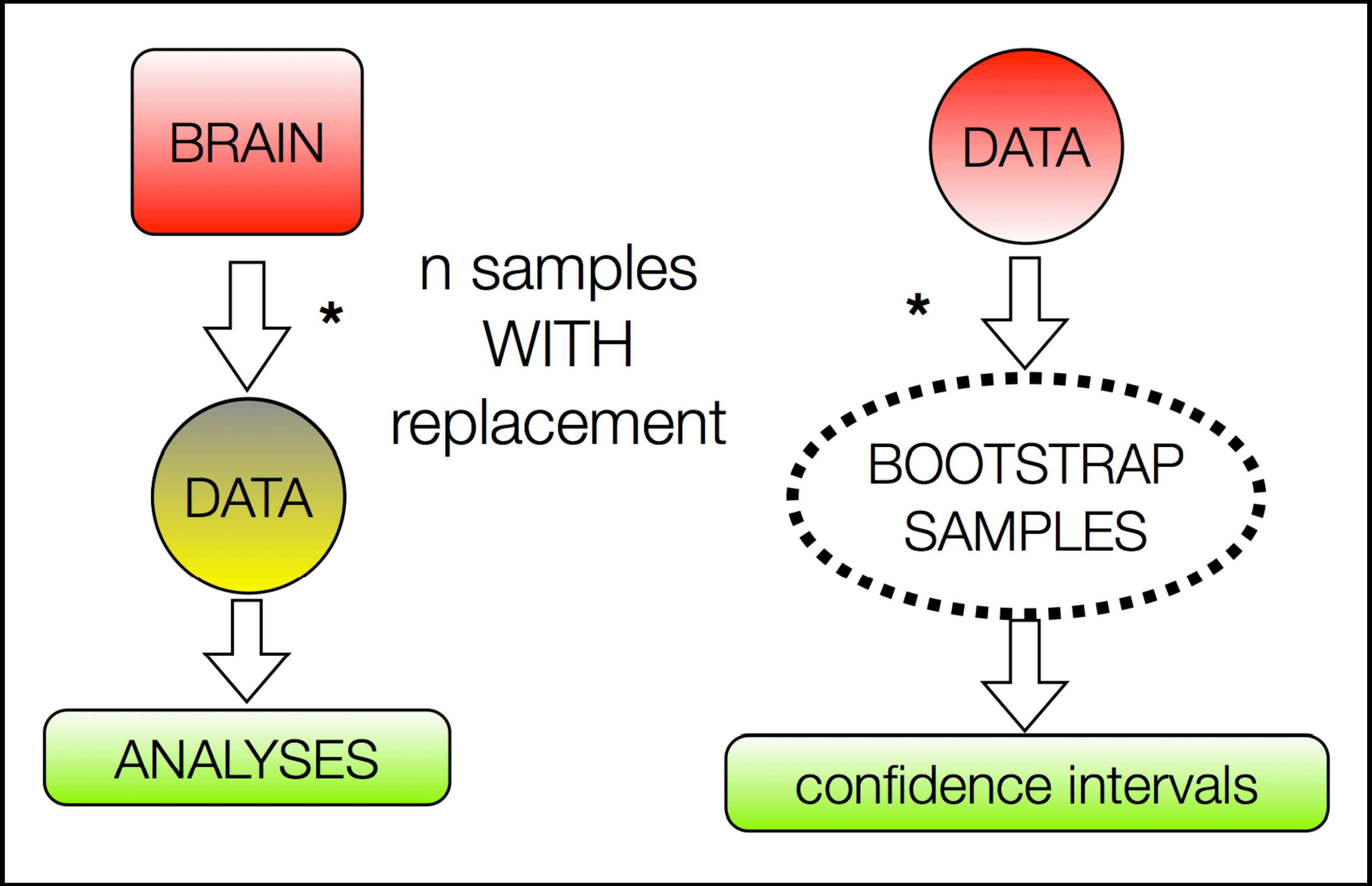
C(i,j) is the measured information flow from process j --> i.
C_{null} is the expected measured information flow when there is no true information flow.
C_{base} is the expected information flow in some baseline period.

Statistics

Statistical Approach	Test	Parametric	Nonparam.
Asymptotic analytic estimates of confidence intervals. Applies to: PDC, nPDC, DTF, nDTF, rPDC	H_{null}, H_{base}, H_{AB}	☑	
Theiler phase randomization Applies to: all	H_{null}		☑
Bootstrap, Jacknife, Cross-Validation Applies to: all	H_{AB}, H_{base}		☑
Confidence intervals using Bayesian smoothing splines Applies to: all	H_{base}, H_{AB}	☑	☑

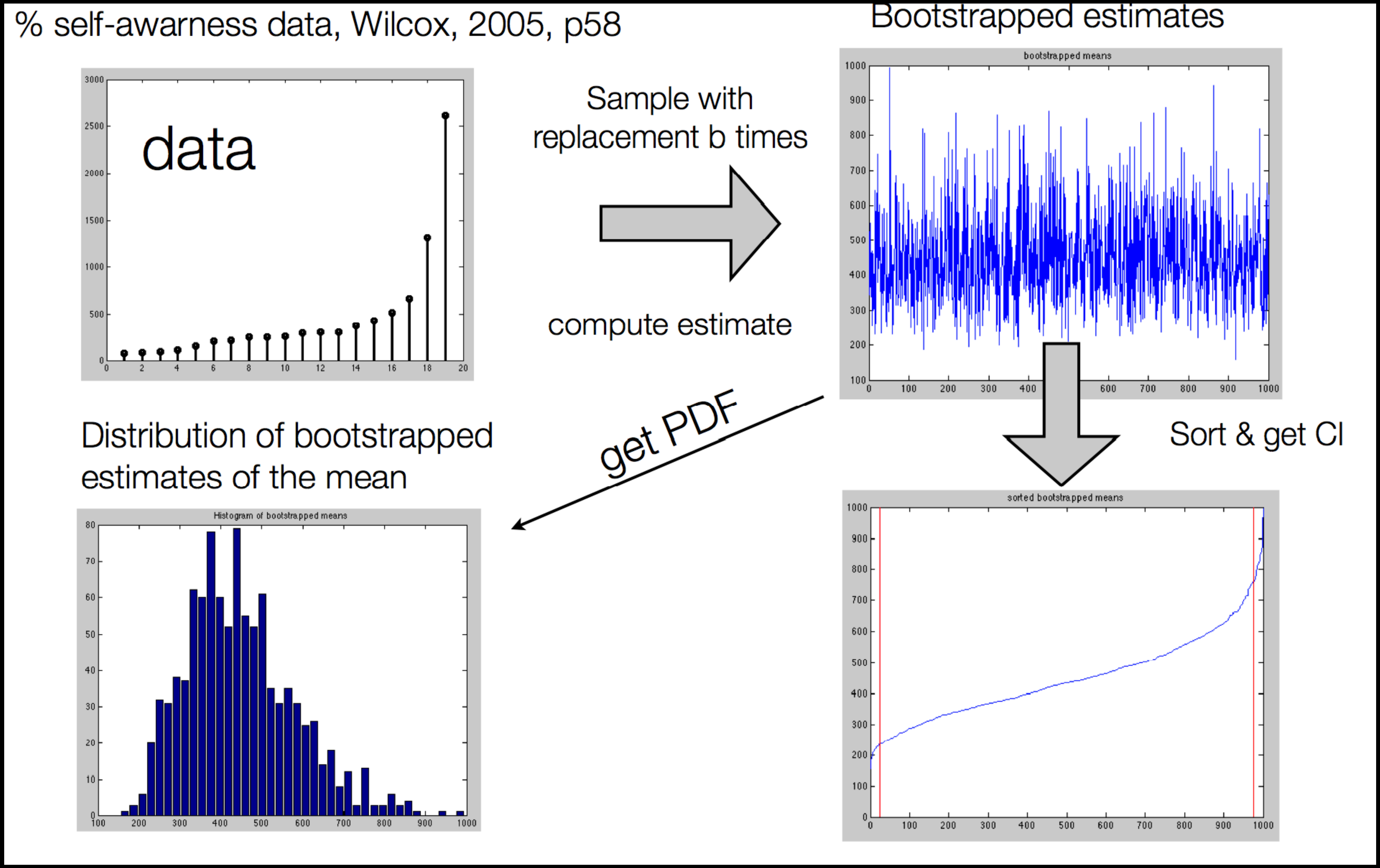
$H_{null} : \mathbf{C}_{ij} = 0$ $H_{base}: \mathbf{C}_{ij} = \mathbf{C}_{baseline}$ $H_{AB}: \mathbf{C}^{A}_{ij} = \mathbf{C}^{B}_{ij}$

Bootstrap Statistics



- sample = X1, ..., Xn
- for k=1:R (number of bootstrap resamples/iterations)
  - resample n observations (trials) with replacement $X^* = \{X^*_{1, ...} X^*_n\}$
  - compute estimator E_k (fit model, obtain connectivity) based on X^*
  - repeat
- with R large enough $P_E = \{E_1, ..., E_R\}$ provides a good approximation to the true distribution of the estimator (connectivity, power, etc)

Bootstrap Statistics



% self-awarness data, Wilcox, 2005, p58

Bootstrapped estimates

Sample with replacement b times

compute estimate

get PDF

Sort & get CI

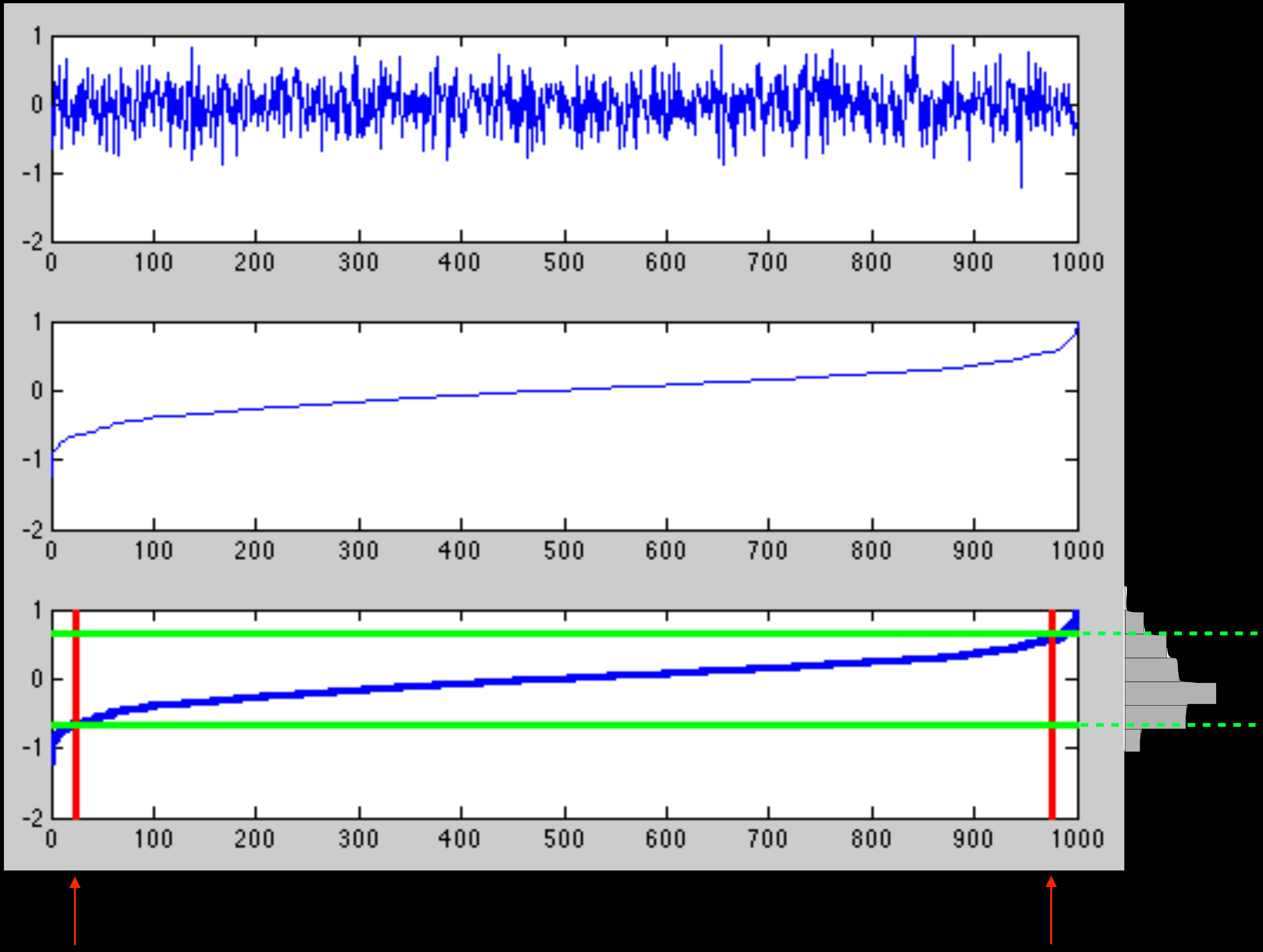
Distribution of bootstrapped estimates of the mean

Bootstrap Statistics

bootstrap

sorted values (cdf)

thresholds (ci)

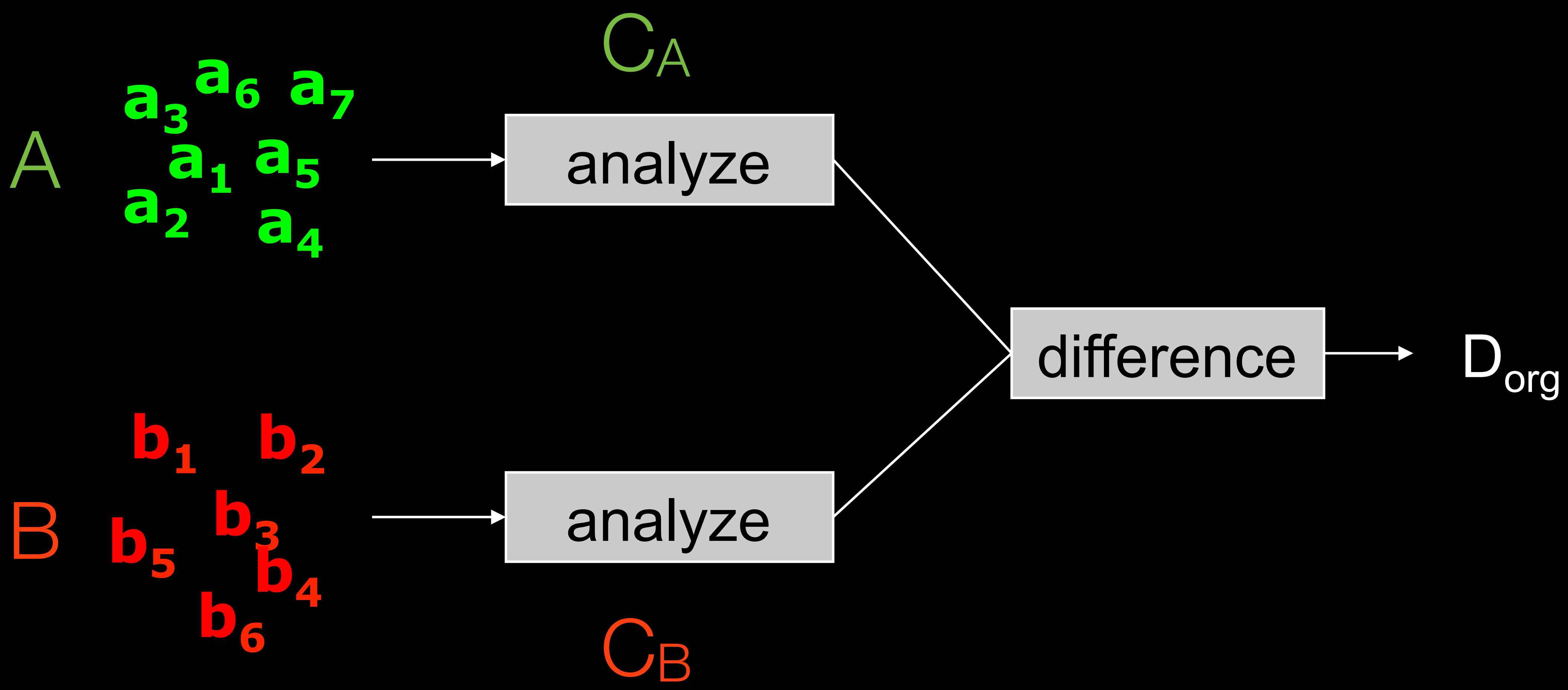


2.5%

97.5%

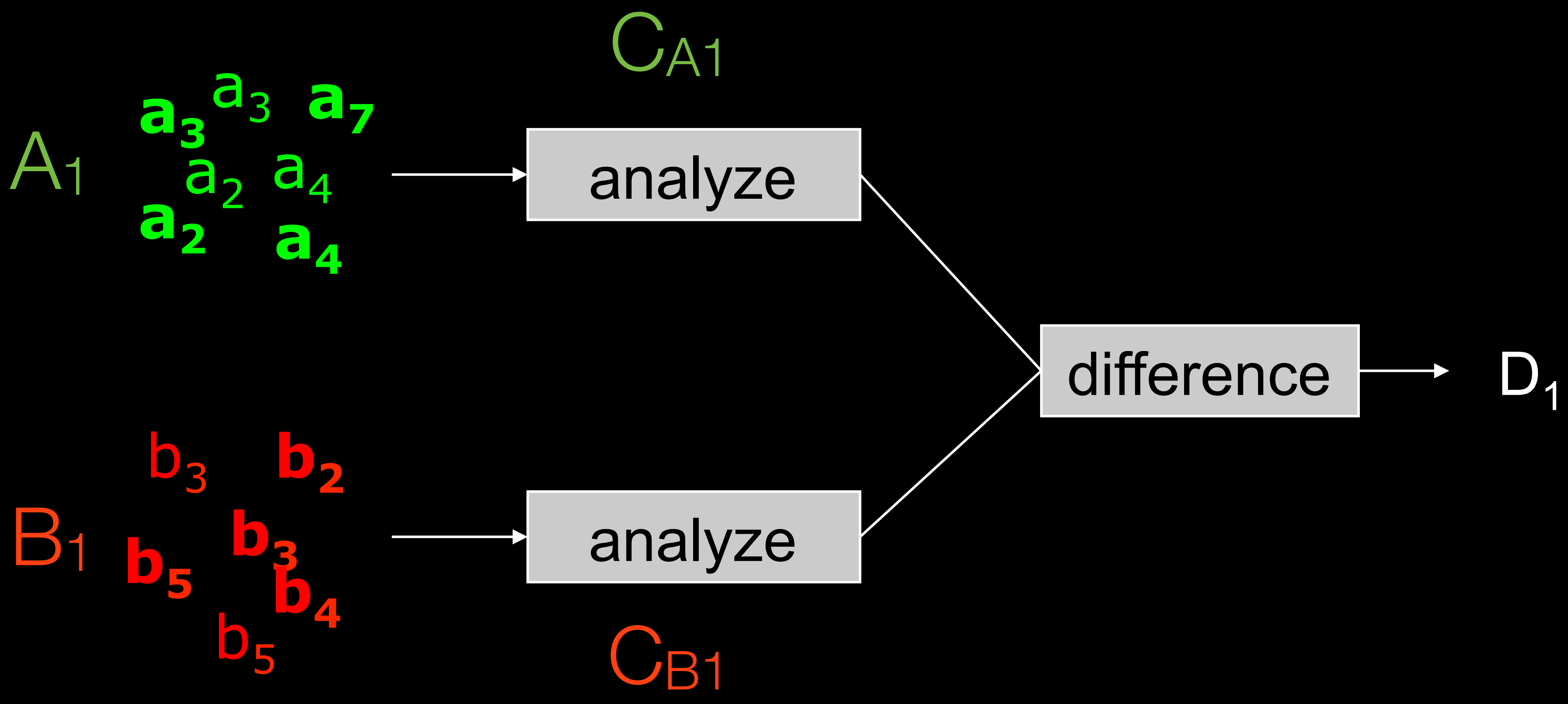
Bootstrap Differences

- Suppose we have two conditions
  A = {a1,...a7}
- B = {b1,...,b6}
- We want to estimate the distributions of connectivity estimator applied to A and B separately, as well as the difference distribution (for testing H0: A=B)



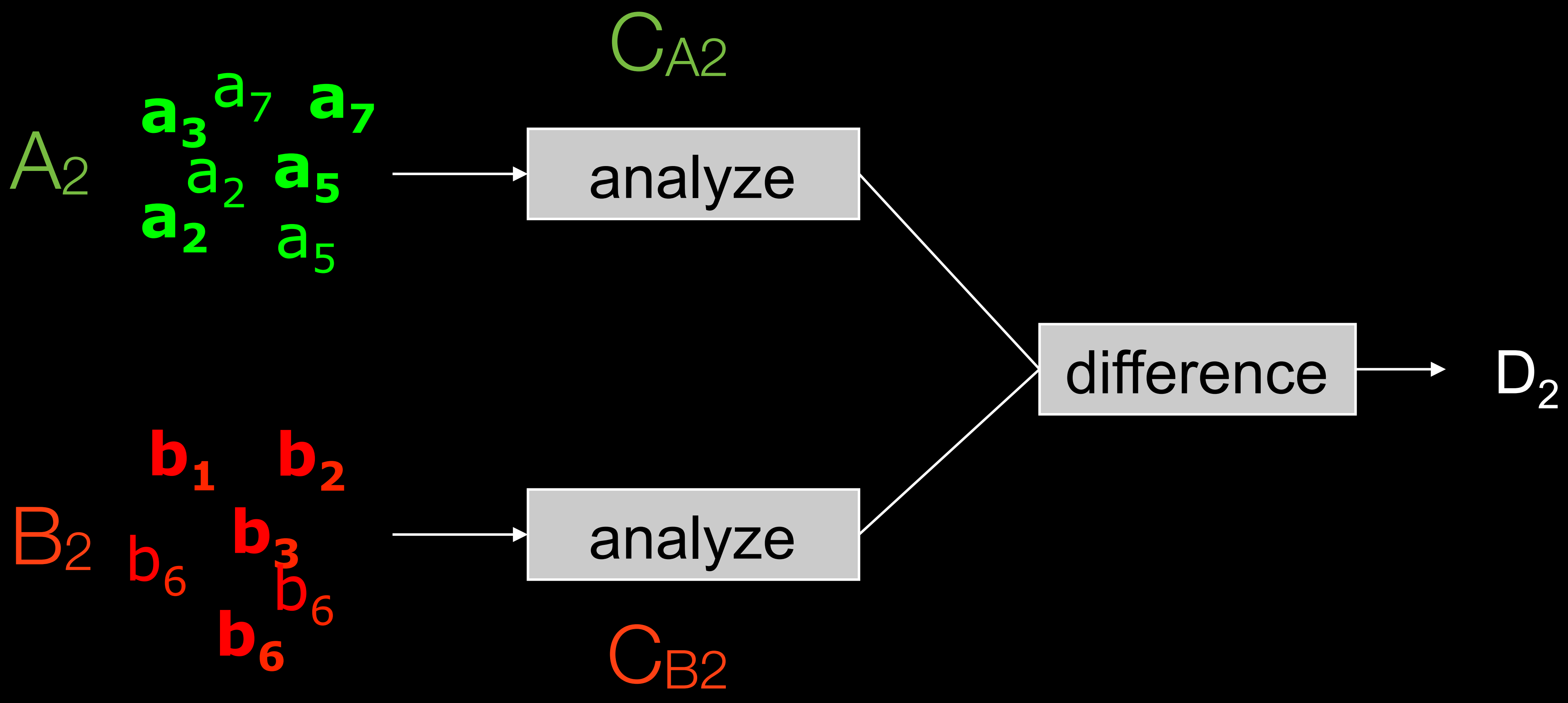
Bootstrap Differences

- For k=1:R (number of bootstrap iterations)
  - Resample with replacement from both groups to get A_k and B_k
  - Fit models and obtain connectivity C_{Ak}, C_{Bk}
  - Compute difference $D_k = C_{Ak}-C_{Bk}$
  - Repeat



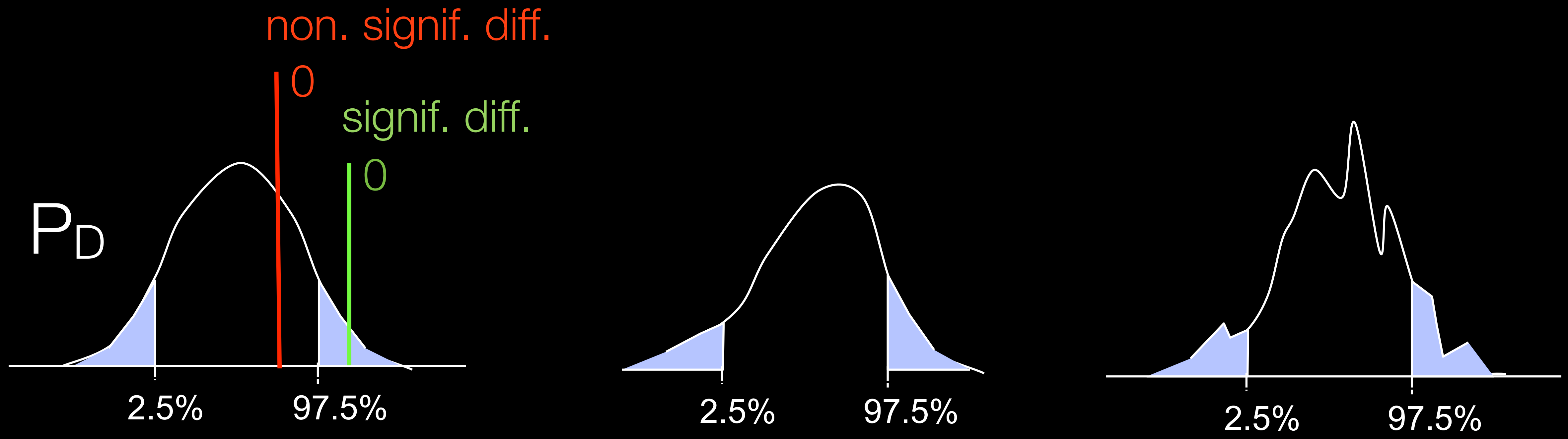
Bootstrap Differences

- For k=1:R (number of bootstrap iterations)
    - Resample with replacement from both groups to get A_k and B_k
    - Fit models and obtain connectivity C_{Ak}, C_{Bk}
    - Compute difference $D_k = C_{Ak} - C_{Bk}$
    - Repeat



Bootstrap Statistics

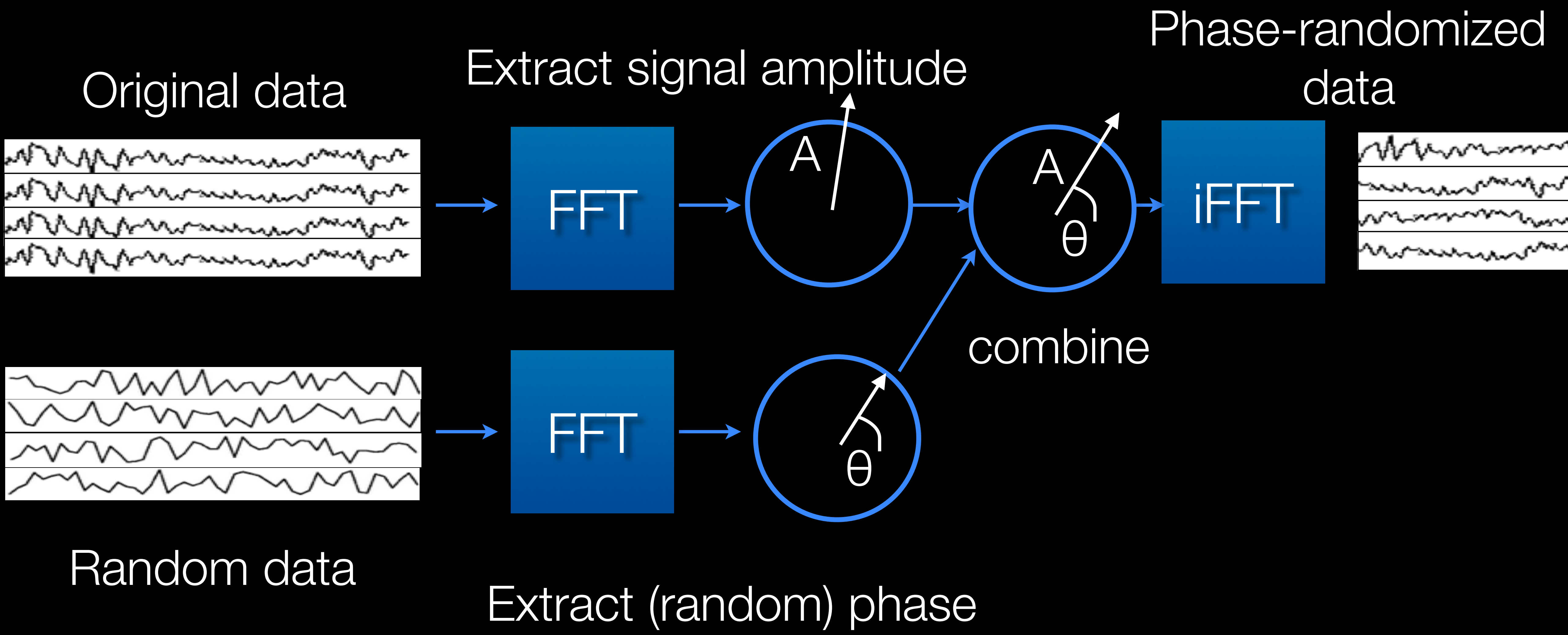
- The procedure yields a distribution $P_D = \{D_1, ... D_R\}$
- If 0 lies in the right (or left) tail of this "difference distribution", then we **reject** the null hypothesis that A=B at the chosen confidence level (below: alpha=0.05 for a two-sided test)



- Difference distribution can take any shape
- The procedure above also provides estimates of the individual distributions of C_A and C_B yielding confidence intervals for H1

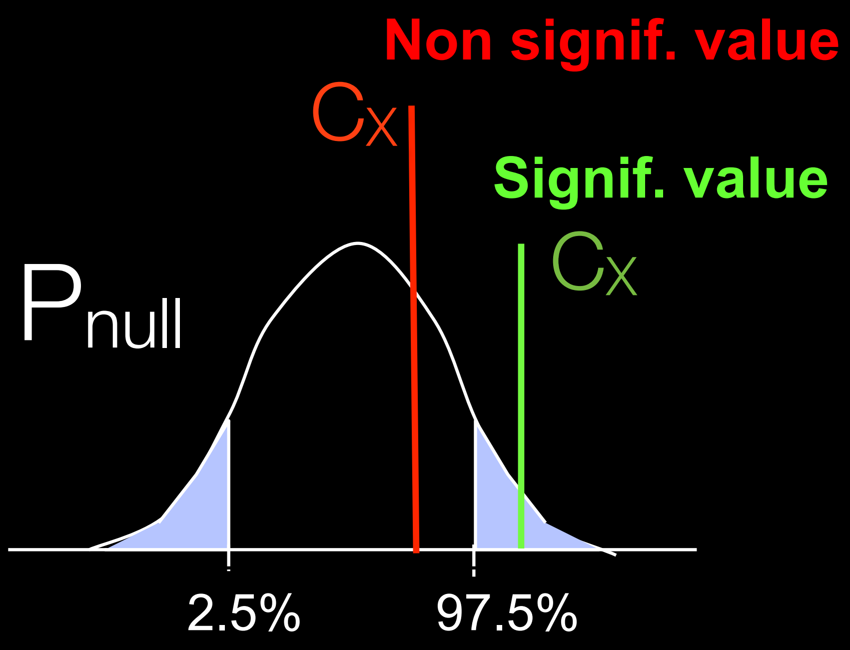
Phase-Randomization

- Phase Randomization Procedure (Theiler, 1992)
  - Method for testing whether there is non-zero information flow (H_{null})



Phase-Randomization

- Start with an n-trial sample: X = {X1, ..., Xn}
- for k=1:R (number of resamples)
  - randomize phases for all trials
  - compute connectivity estimate C_k
  - repeat
- With B large enough the B estimates provide a good approximation of the **null** distribution of the connectivity estimator
- Compare connectivity C_X from original (non-randomized) samples X to quantiles of $P_{null} = \{C_1, ..., C_R\}$



Group Source Statistics

- An alternative approach:
  For each subject…

  1. Perform distributed source localization (possibly after separating a subspace of brain components using ICA)

  2. Select *M* regions of interest (ROIs) e.g. from a standardized anatomical atlas (e.g. Desikan-Killiany, Destrieux, etc) and integrate current density within each ROI. This yields *M* source time-series for each subject

  3. Store results in EEG.srcpot

  4. Obtain connectivity estimates for sources using SIFT with the 'Sources' option set in pre-processing. Resulting [M x M x N_{freq} x N_{times}] connectivity matrices are stored in EEG.CAT.Conn.

5. Apply your favorite mass-univariate or multivariate statistical approach (e.g. GLM, t-test, (M)ANOVA, etc) to the collection of connectivity estimates from all subjects to obtain desired statistics. See LIMO-EEG Toolbox and EEGLAB's statcond(). Beware of multiple comparisons issues! FDR may not be suitable.

Group Source Statistics

- Also see Group-SIFT plugin by Makoto Miyakoshi

INTERMISSION