Decomposed dynamics: toward discovering latent brain factors underlying depression and recovery

Christopher J. Rozell

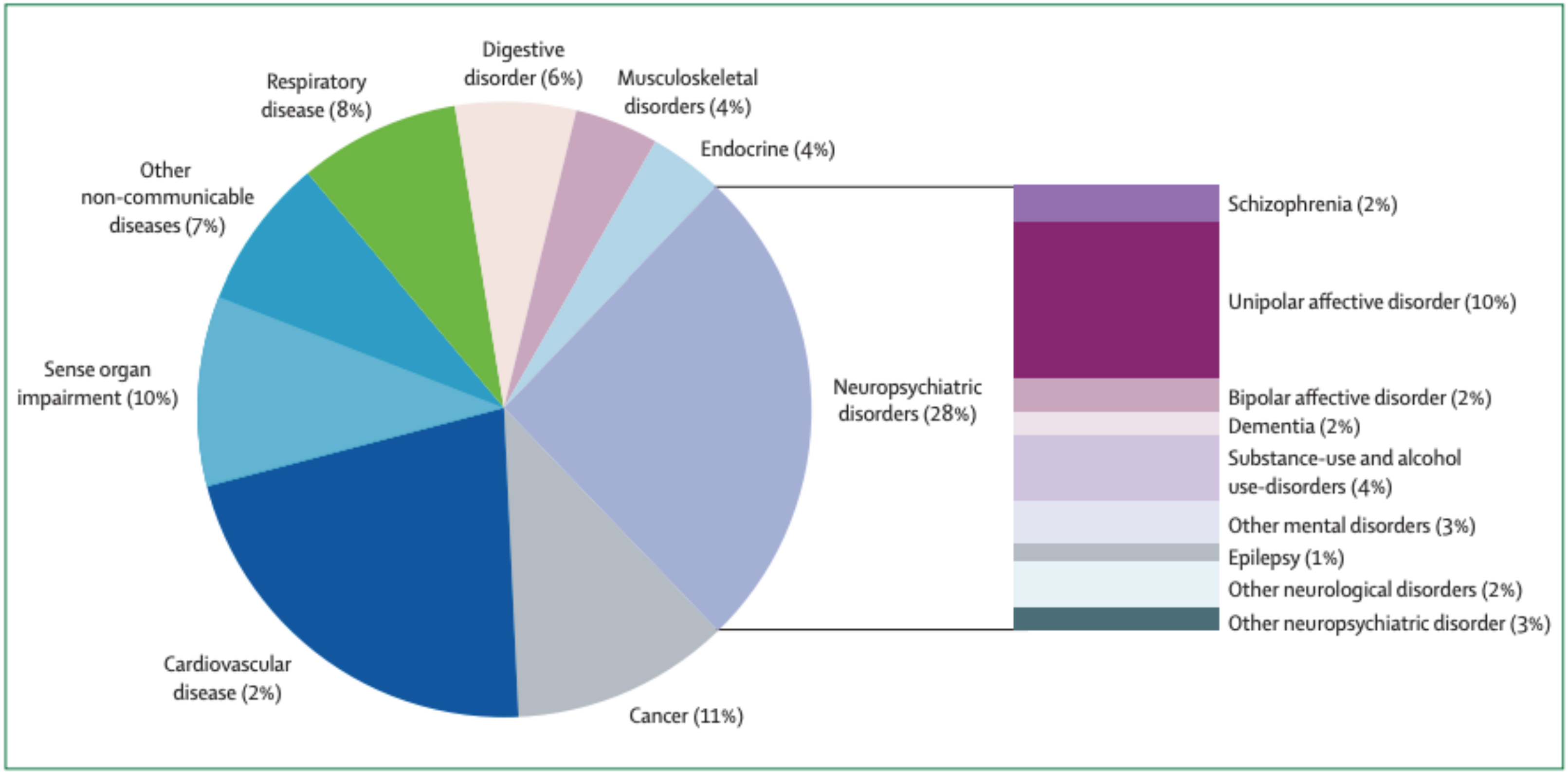
School of Electrical & Computer Engineering
Georgia Institute of Technology

Disclosures

- Off-Label use of devices:
  - Medtronic Activa PC+S and Summit RC+S donated devices
- Registered studies:
  - FDA IDE: G130107 (PI: Mayberg)
  - Clinicaltrials.gov ID#: NCT00367003, NCT01984710
- BRAIN Initiative funding:
  - NIH UH3NS103550 (Co-PIs Mayberg, Rozell and Riva-Posse)
- Potential conflicts of interest
  - Advisory Board and shareholder: Motif Neurotech, Inc.
  - Intellectual Property: relevant patent applications filed

Major depressive disorder (MDD)

Contribution to non-communicable disease burden



(Prince, et al. 2007)

300M+ people globally
20-30% treatment resistant (TRD)

https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf
https://www.dbsalliance.org/education/depression/statistics/

Defining and rating depression

Table 1. DSM-5 Diagnostic Criteria for Major Depressive Disorder.*

Five or more of the following symptoms must be present nearly every day during a 2-wk period:

Core symptoms (≥1 required for diagnosis)

- Depressed mood most of the day
- Anhedonia or markedly decreased interest or pleasure in almost all activities

Additional symptoms

- Clinically significant weight loss or increase or decrease in appetite
- Insomnia or hypersomnia
- Psychomotor agitation or retardation
- Fatigue or loss of energy
- Feelings of worthlessness or excessive or inappropriate guilt
- Diminished ability to think or concentrate, or indecisiveness
- Recurrent thoughts of death or suicidal ideation

* DSM-5 denotes *Diagnostic and Statistical Manual of Mental Disorders,* fifth edition.

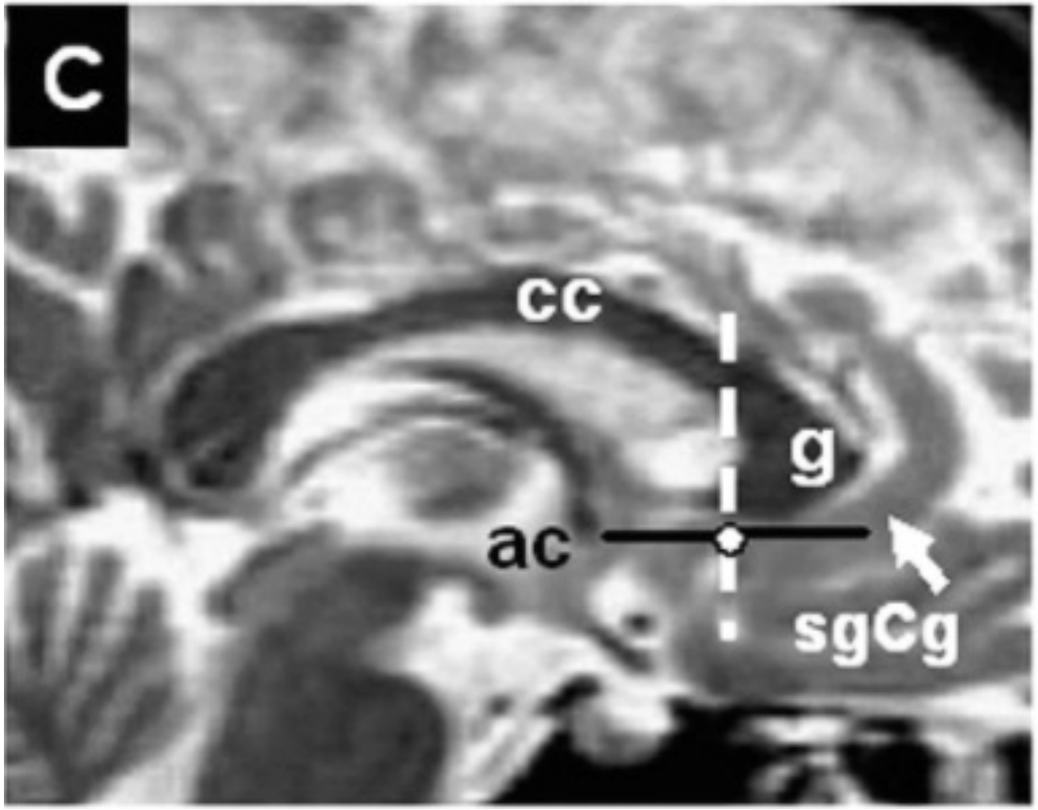
Hamilton depression rating scale (HDRS, Ham-D 17)

No.	Item <semantic label>	Examples of the criteria
1	Depressed mood <Depressed>	Feels sad, hopeless, helpless, and so on
2	Feelings of guilt <Guilty>	Feels self-reproach, a sense of having let people down
3	Suicide <Suicide>	Feels life is not worth living
4	Insomnia-early <Insomnia_E>	Complains of difficulty falling asleep
5	Insomnia-middle <Insomnia_M>	Complains of being restless and disturbed during the night
6	Insomnia-late <Insomnia_L>	Wakes in early morning hours
7	Work and activities <Work>	Loss of interest in activity, hobbies, or work
8	Retardation <Retardation>	Decreased motor activity
9	Agitation <Agitation>	Fidgeting, playing with hands, biting lips
10	Anxiety <Anxiety>	Worries about minor matters
11	Anxiety—somatic <Anxiety_S>	Flushing, sweating, tremor, and so on
12	Somatic symptoms (gastrointestinal) <Somatic_G>	Loss of appetite
13	Somatic symptoms <Somatic>	Backaches, headaches, muscle aches, and so on
14	Genital symptoms <Genital>	Loss of libido and menstrual disturbances
15	Hypochondriasis <Hypochondriasis>	Preoccupied with health
16	Weight loss <Weight>	Weight loss
17	Insight <Insight>	Acknowledges or denies being depressed

(Taylor 2014)

(Hamilton 1960; Wu et al. 2005; Bagby et al. 2004; Faries, et al. 2000)

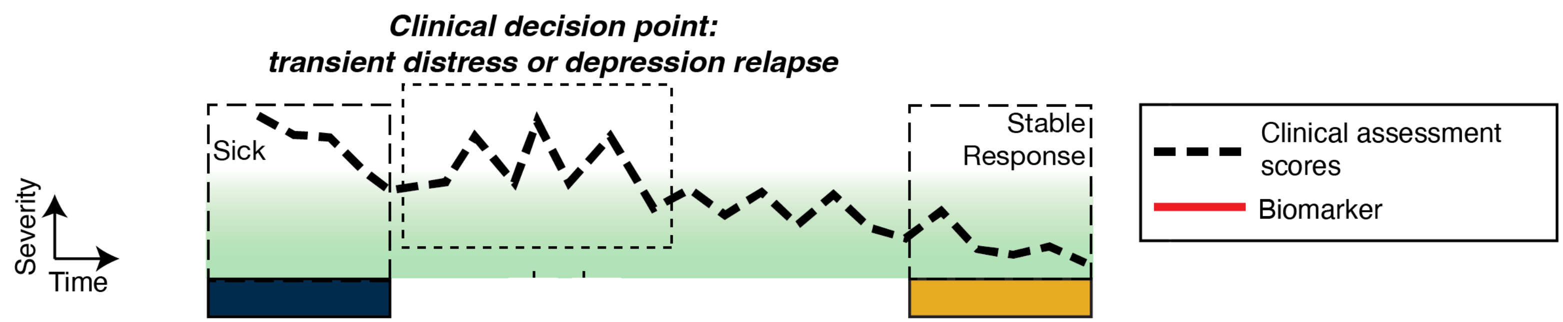
Deep brain stimulation



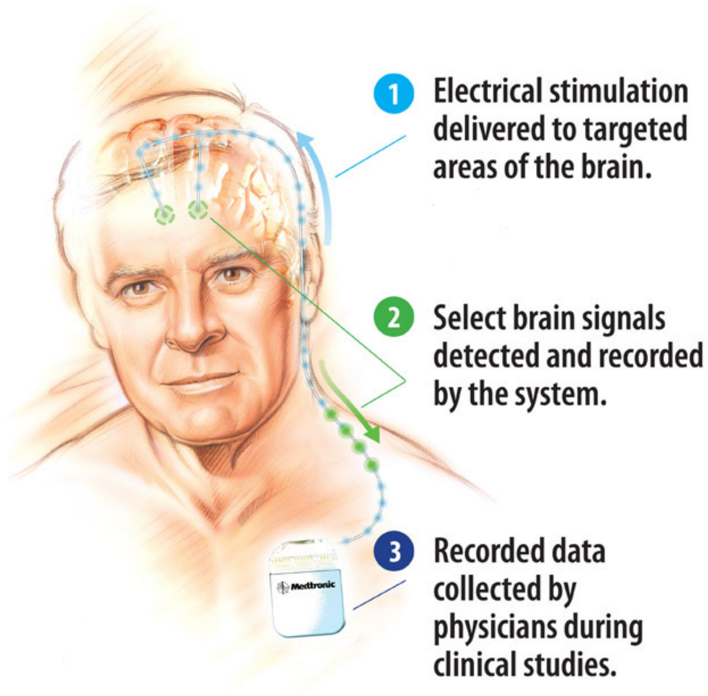
(Mayberg, et al., 2005)

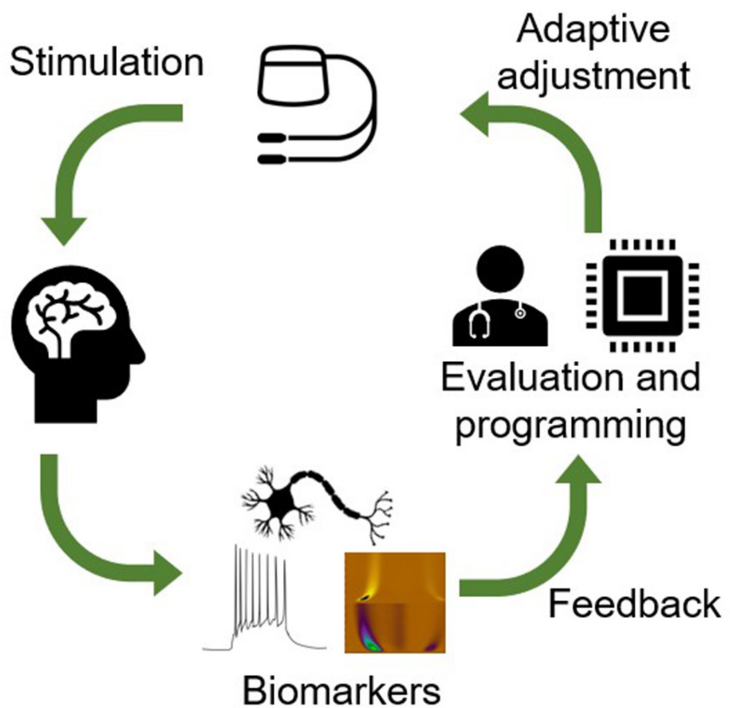
Disorder vs. distress

- Gold standard survey measures biased, non-specific, and confounded by transient distress
- Want objective biomarker that tracks disease state, reflects intervention, and is useful in decision support



Brain sensing-> useful biomarkers?





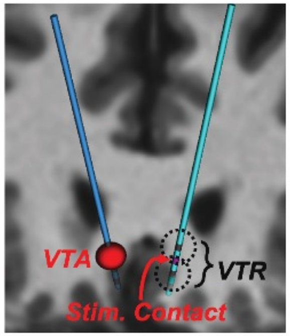
(Price, et al., 2020)

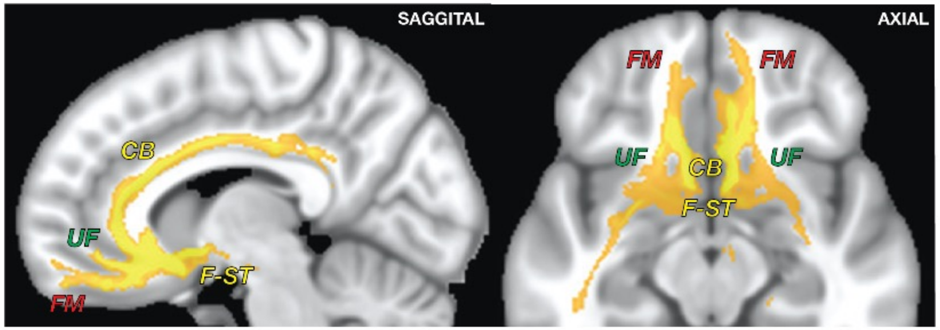
- Ten research prototype Medtronic Activa PC+S
- Records Local Field Potential (LFP) aggregated activity

Medtronic

Clinical population

- Cohort (**n=10**) implanted at Emory
  - Targeting using presurgical imaging





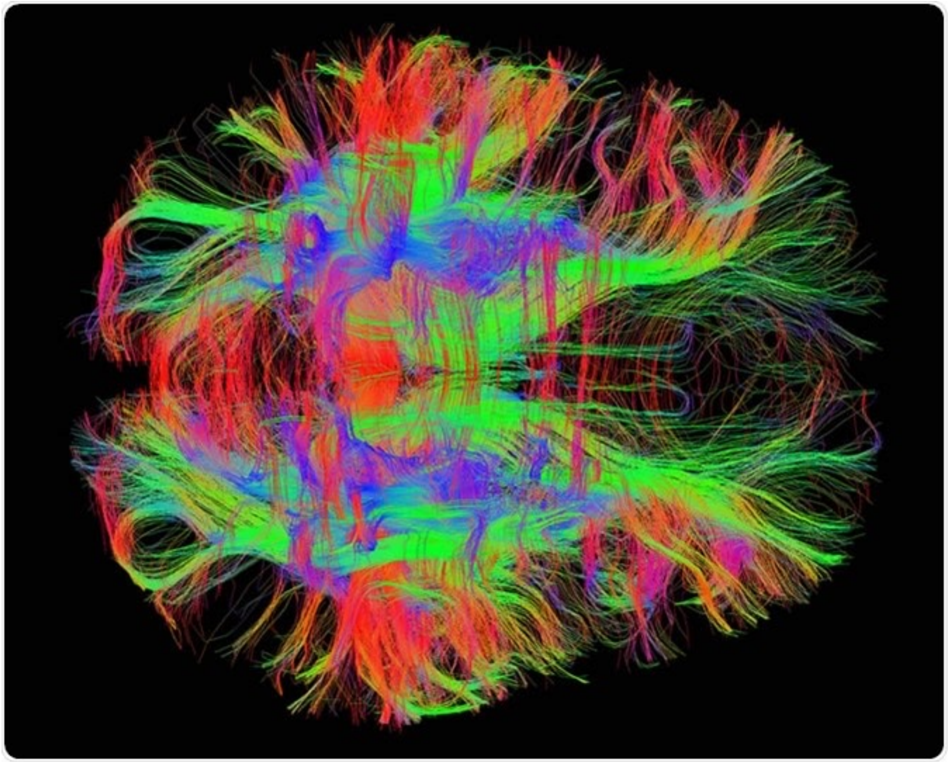
(Riva-Posse, et al., 2014)

- Clinical outcomes at predefined endpoint (24 weeks):
  - 90% response (HDRS<50%) and 70% remission (HDRS<8)

Characteristics	**Mean (S.D) or Count**
Age at surgery	49.40 (11.2)
Sex	Female (7), Male (4)
Employed at time of surgery	2 of 10
Baseline HDRS-17	22.3 (1.64)
Duration of current episode (months)	47.3 (44.03)
Age at first depressive episode (years)	26.6 (10.48)
Number of depressive episodes	3.3 (1.06)

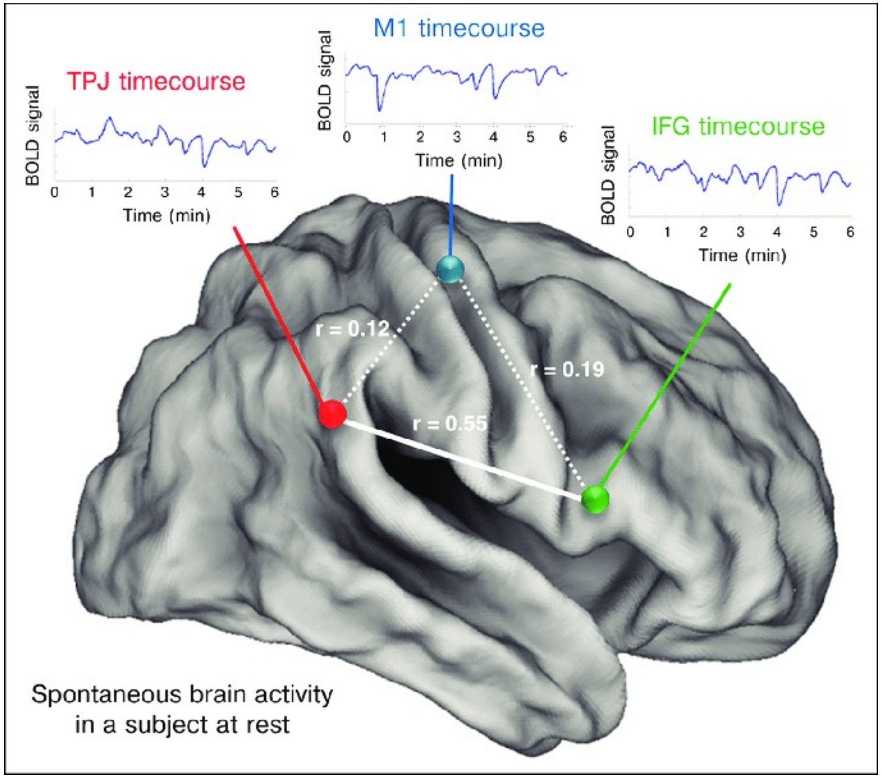
Structural and functional connectivity

Diffusion Tensor Imaging



(Zeynep Saygin)

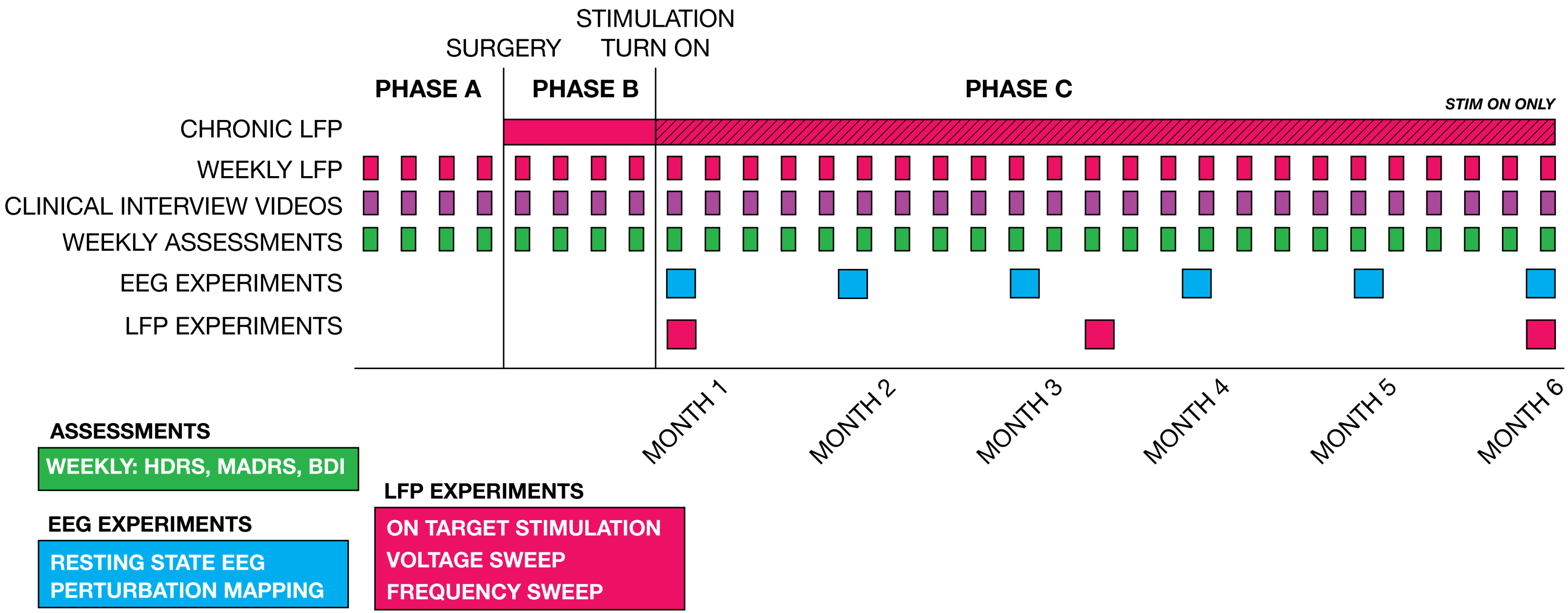
Functional MRI



(Gillebert & Mantini, 2012)

Longitudinal data collection (24 weeks)

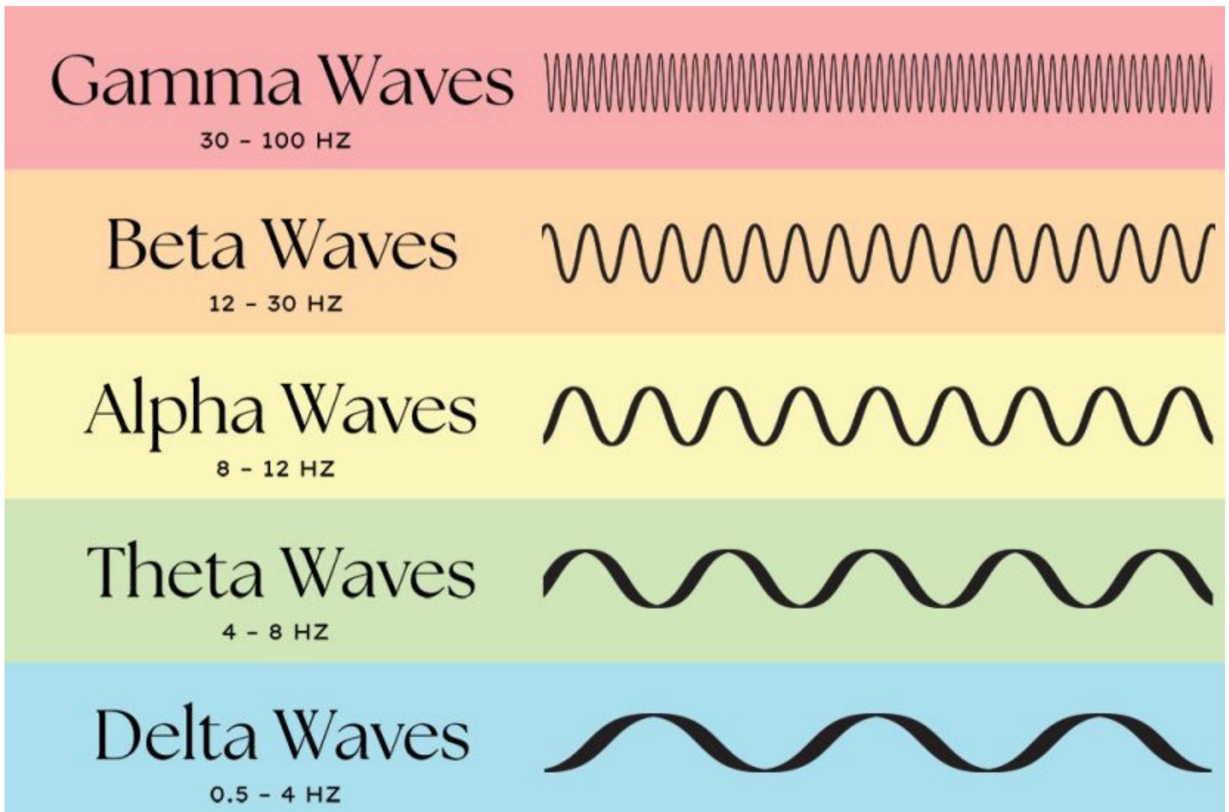
- Subset (**n=6**) with usable electrophysiology data
  - Weekly clinical visit with recorded clinical interviews



(Alagapan,…, Mayberg, R., 2023)

Local field potentials

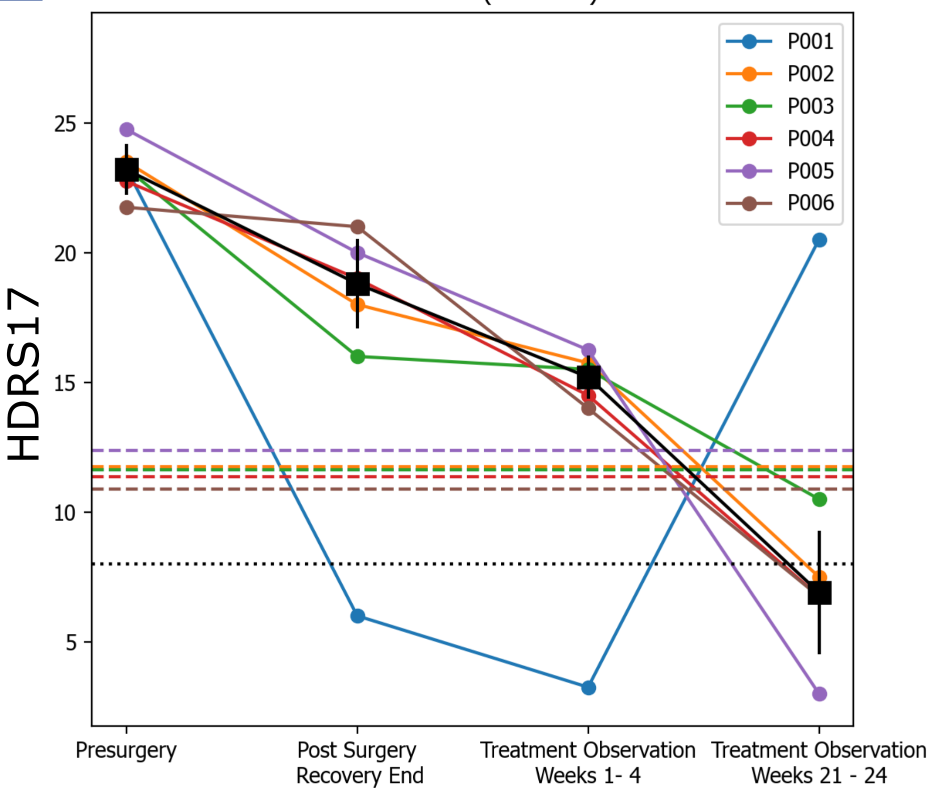
- Differential local field potential (LFP) recording *with stim off*
  - 7 min data broken into 10s blocks and and filtered <40Hz
  - Calculate standard features of spectral power, coherence, etc.



https://www.diygenius.com/the-5-types-of-brain-waves/

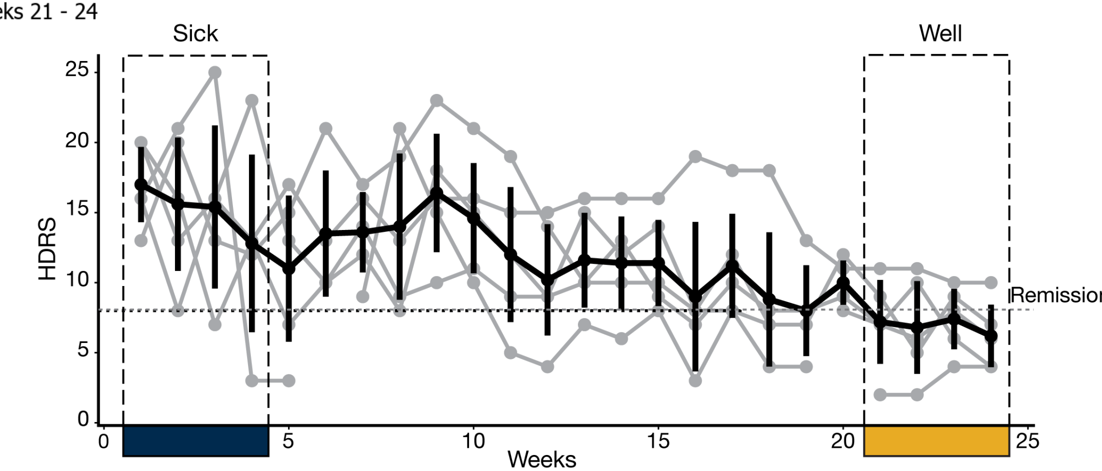
(Alagapan,…, Mayberg, R., 2023)

Clinical trajectory over 24 weeks



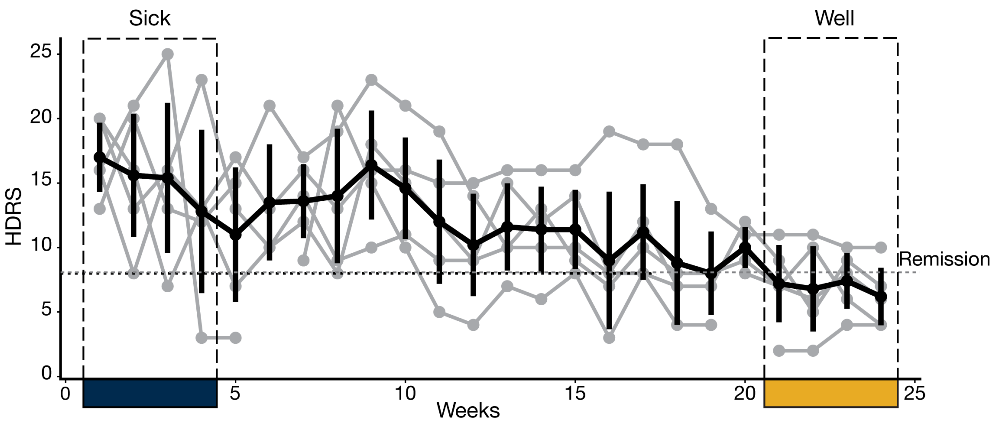
Patient responses

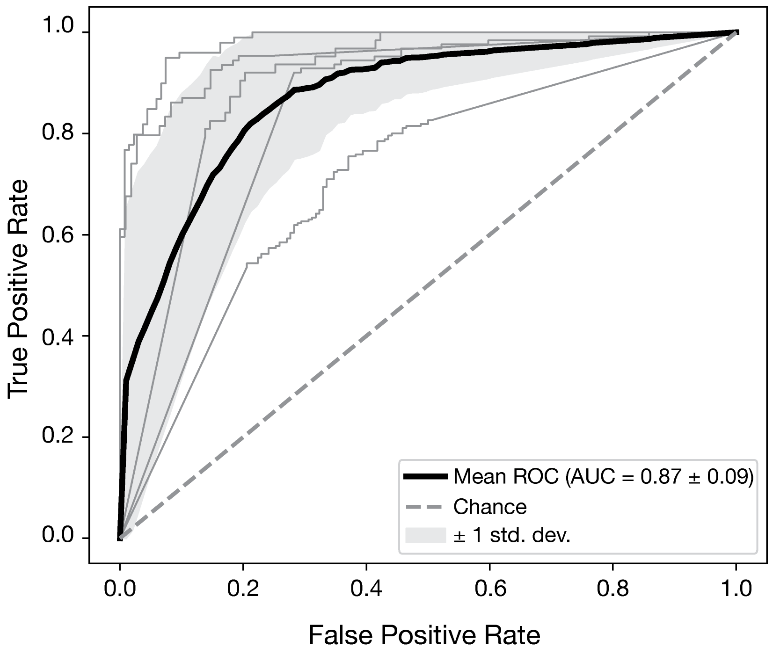
- 5/6 “typical” responders
- 4/6 achieve remission
- 1/6 “relapsed responder”
- Idiosyncratic trajectories



Can we see latent recovery dynamics?

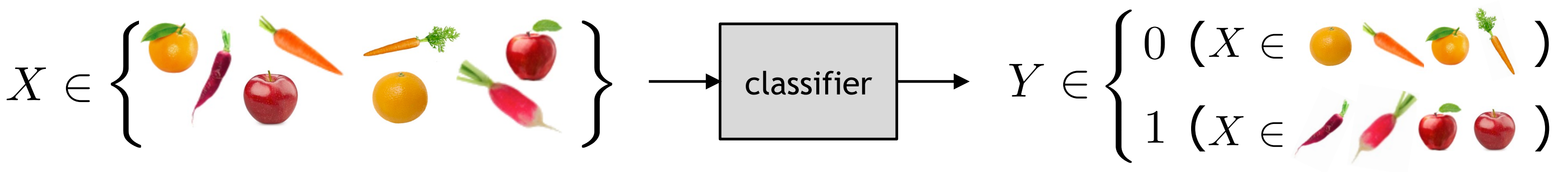
- Want biomarker of sick/well response instead of HDRS
  - Unknown if changes visible in SCC electrophysiology
  - Train classifier with month 1 (sick) and month 6 (well) data
  - Two layer MLP trained on ***group*** of typical responders with leave-one-out cross validation





What's changing in the brain?
Requires that we "explain" this classifier

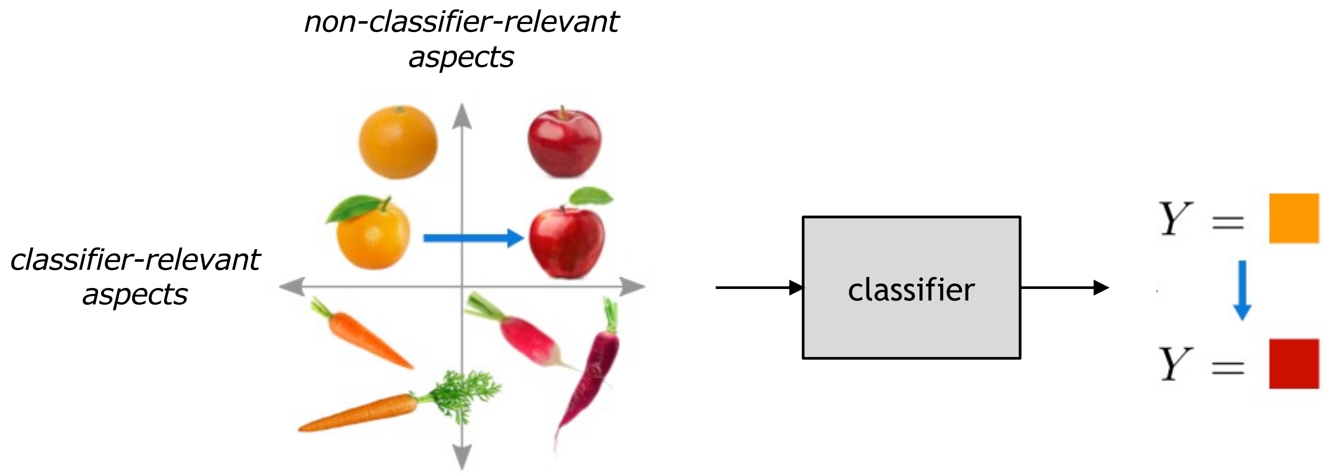
Causality as a direct explanation



data aspects: shape, color

Identify aspects of the data that *have a large causal effect on* the classifier

"when intervened on, produce a large change in the black-box output"

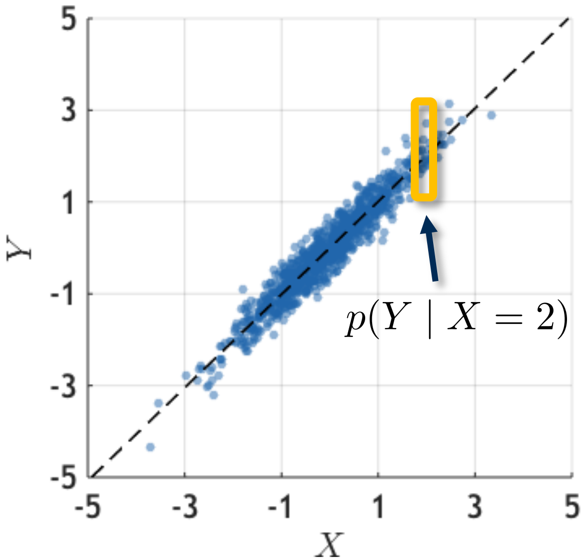


Requires ability to **generate realistic** data with **controlled** variations without **prior** knowledge of features to use with **any** classifier

Correlation and causation

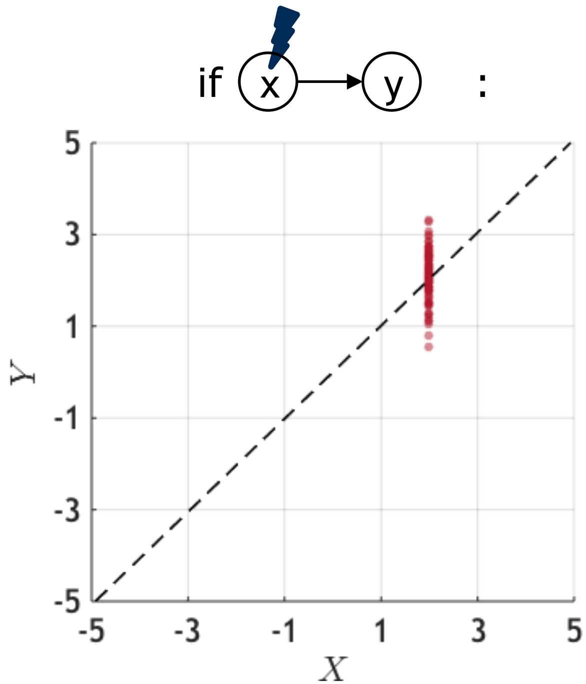
Observational

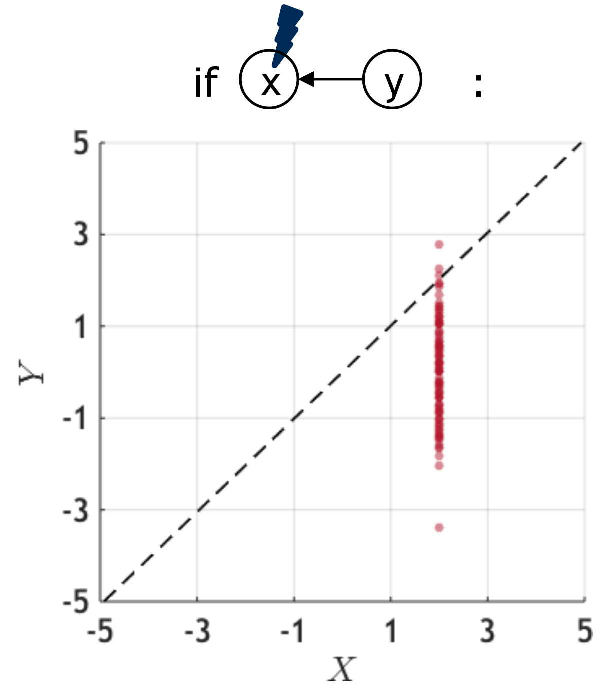
$p(X, Y)$



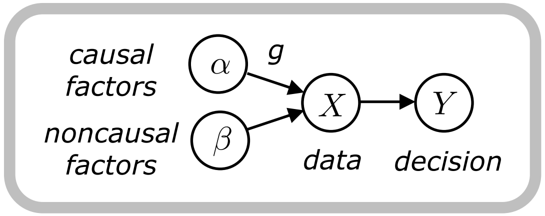
Interventional

$p(Y \mid do(X = 2))$





Generative causal explainer (GCE)



- Structural causal model is DAG with independent factors (α,β)
- Optimize model *g* so that α has large causal effect on *Y*

$$\arg\max_{g \in G} \; \mathcal{C}(\alpha, Y) + \lambda \, \mathcal{D}(p(g(\alpha, \beta)), p(X))$$

  – Metric of causal influence *C*: *information flow* (Ay & Polani 2008)

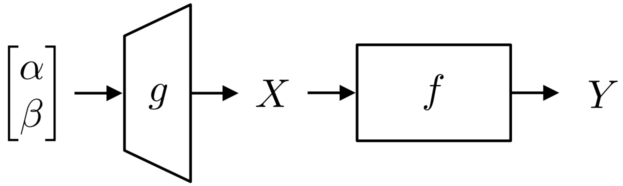
$$I(U \to V) := \int_U p(u) \int_V p(v \mid do(u)) \log \frac{p(v \mid do(u))}{\int_{u'} p(u') p(v \mid do(u'))} dV \, dU$$

  – Intervention *do*(*u*) fixes the value of *u* regardless of parent node
  – In this model, $I(\alpha \to Y)$ reduces to mutual information $I(\alpha;Y)$

(O'Shaugnessy, Canal, Connor, Davenport, R., 2020)

Analysis in simple model

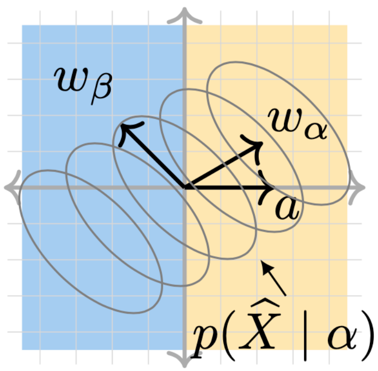
Objective: $\arg\max_g \; I(\alpha \to Y) - \lambda \cdot \mathrm{D}_{KL}\left(p(g(\alpha,\beta)),\; p(X)\right)$



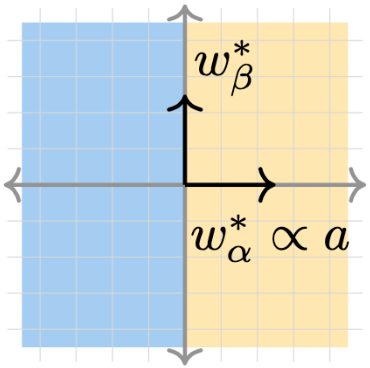
linear-Gaussian generative model:
$g(\alpha,\beta) = W_\alpha \alpha + W_\beta \beta + \mathcal{N}(0,\gamma I)$

Proposition (O'Shaughnessy et al., 2020): *Let* $g(\alpha,\beta) = \mathcal{N}(w_\alpha \alpha + W_\beta \beta, \gamma)$ *and let* $p(Y=1|x) = \sigma(a^T x)$ *where* σ *is the normal CDF. Under normalization and using KL for* $\mathcal{D}$*, the objective is maximized when* $w_\alpha \propto a$ *and* $W_\beta^T a = 0$.

Linear classifier

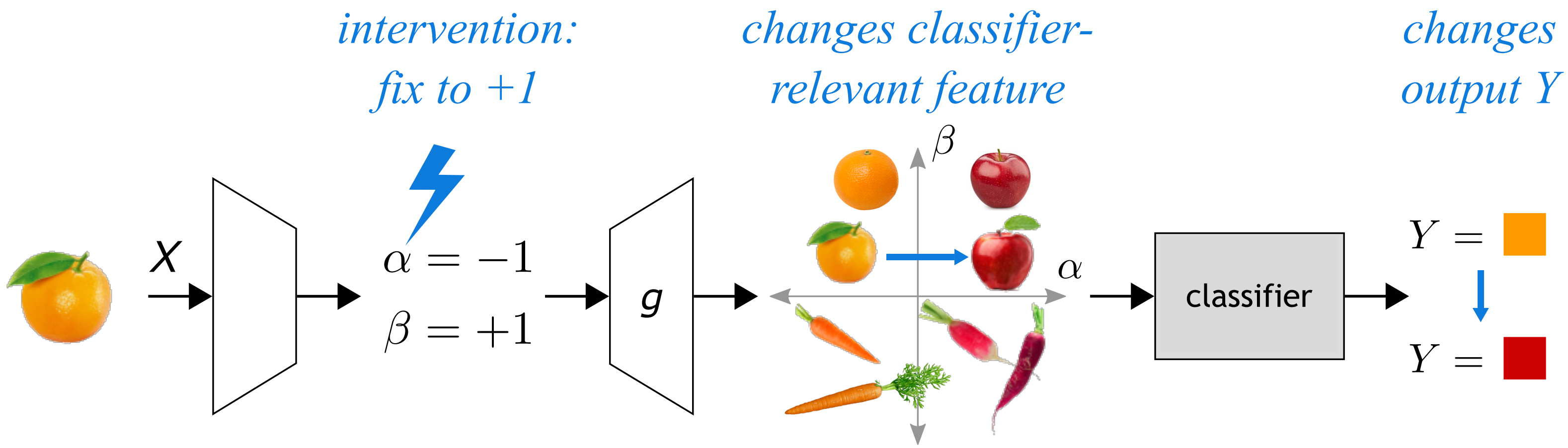


Linear classifier



GCE in practice

- Learn a generative model using a Variational Autoencoder (VAE)
- Low-dimensional latent space divided into (α,β)
- Similar to nonlinear Principal Component Analysis (PCA)



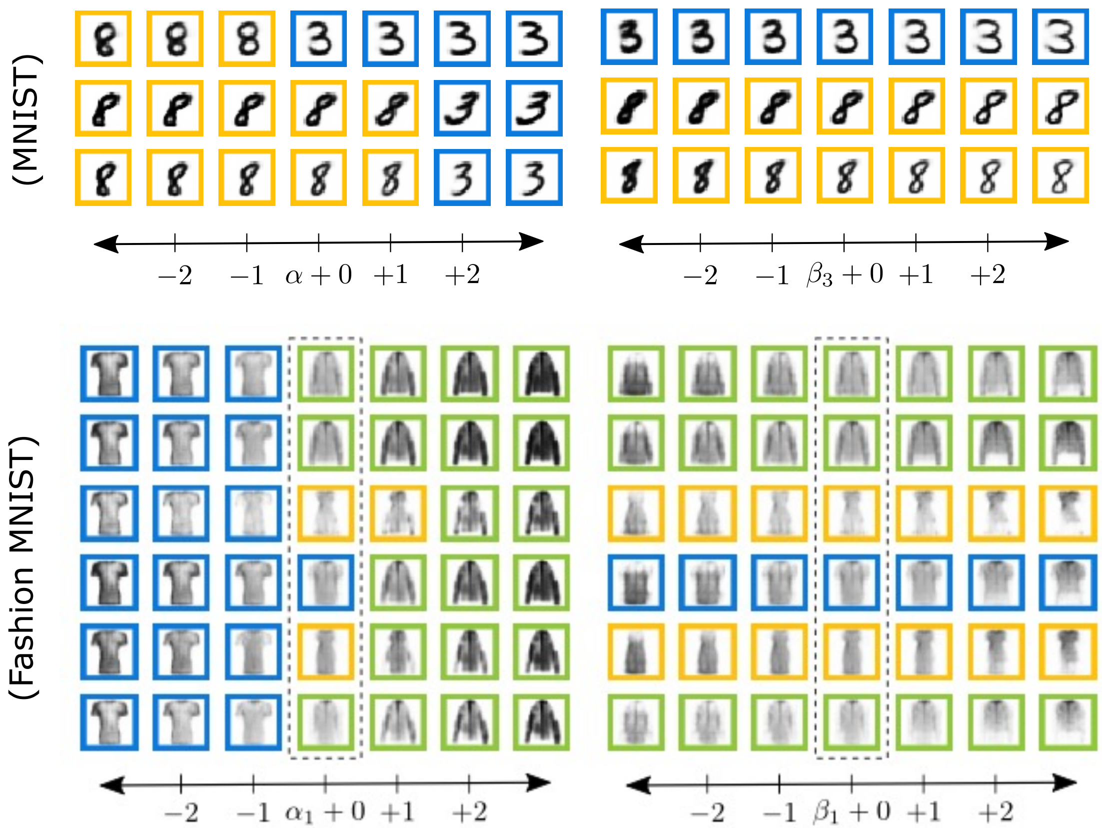
VAE objective

$$\arg\max_{g:\,(\alpha,\beta)\to X} \underbrace{\mathrm{CausalEffect}(\alpha \to Y)}_{\alpha\text{ has a large causal effect on } Y} \quad \text{s.t.} \quad \underbrace{p(g(\alpha,\beta)) \approx p(X)}_{p(\alpha,\beta)\text{ is a valid representation of } p(X)}$$

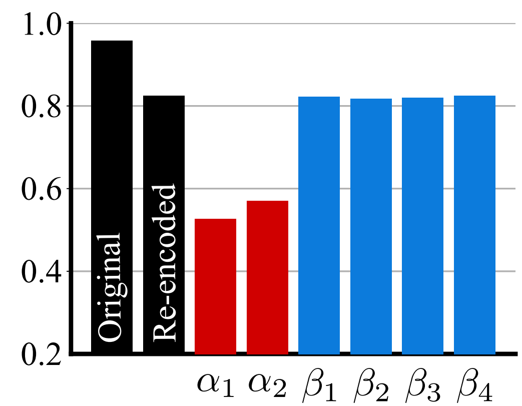
Sample results

- Latent variable α controls classifier-relevant features
- Latent variable β controls within-class variations

Explanations



Classifier accuracy after "removing" aspects

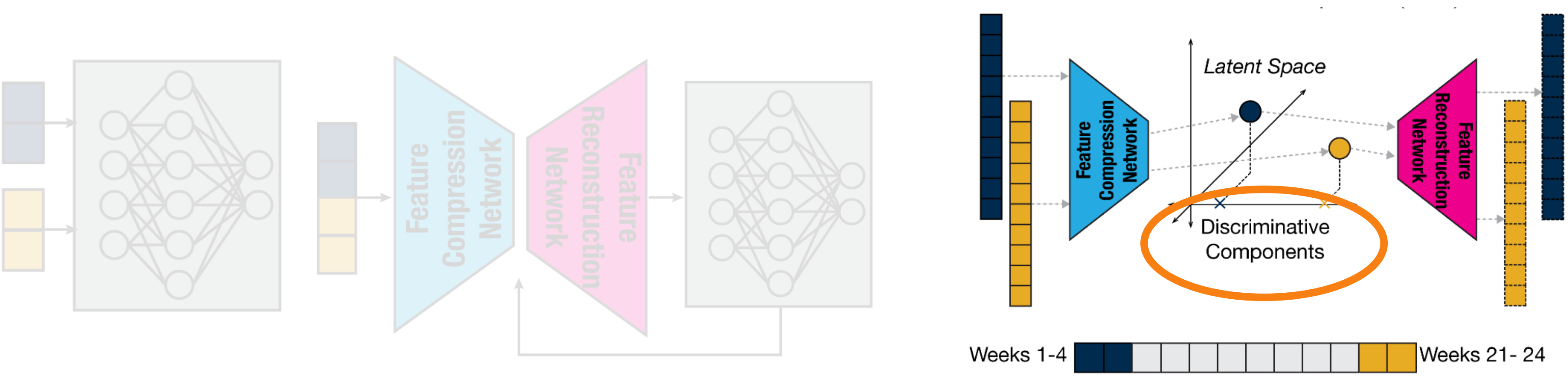


(Fashion MNIST)

(O'Shaugnessy, Canal, Connor, Davenport, R., 2020)

LFP analysis workflow

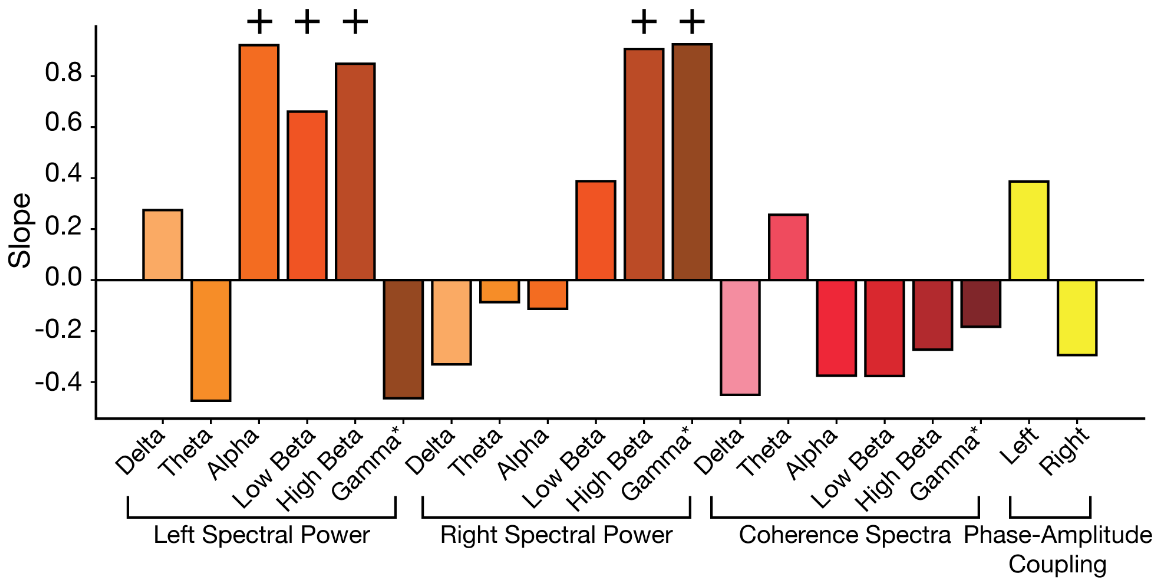
- Novel XAI methods to discover low-dimensional latent factors that distinguish sick/well LFP signal

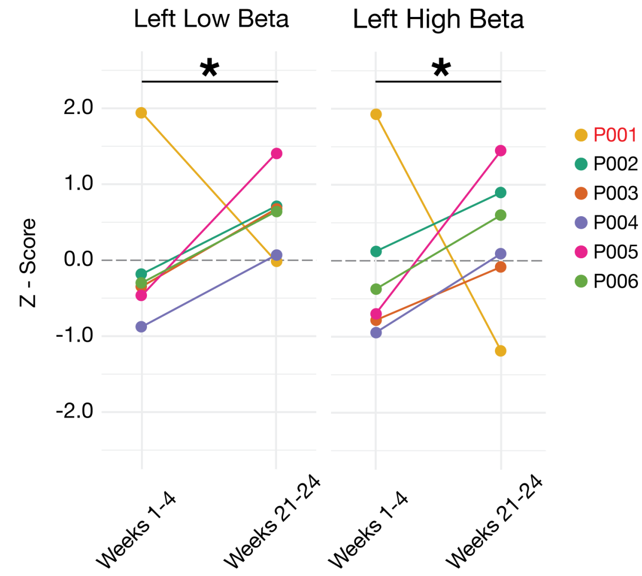


- Proposed biomarker: Spectral Discriminative Component (SDC)
- SDC is a *collection* of LFP features that jointly change to indicate sick/well

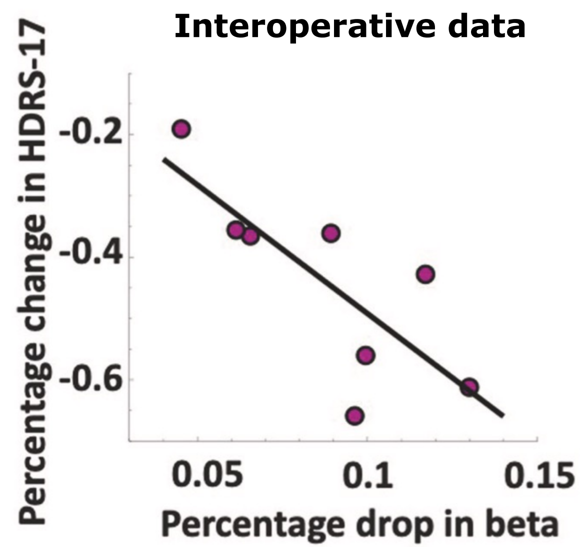
SDC composition

- SDC identifies spectral feature combinations indicating recovery
- Left beta has consistent acute decrease followed by facilitation



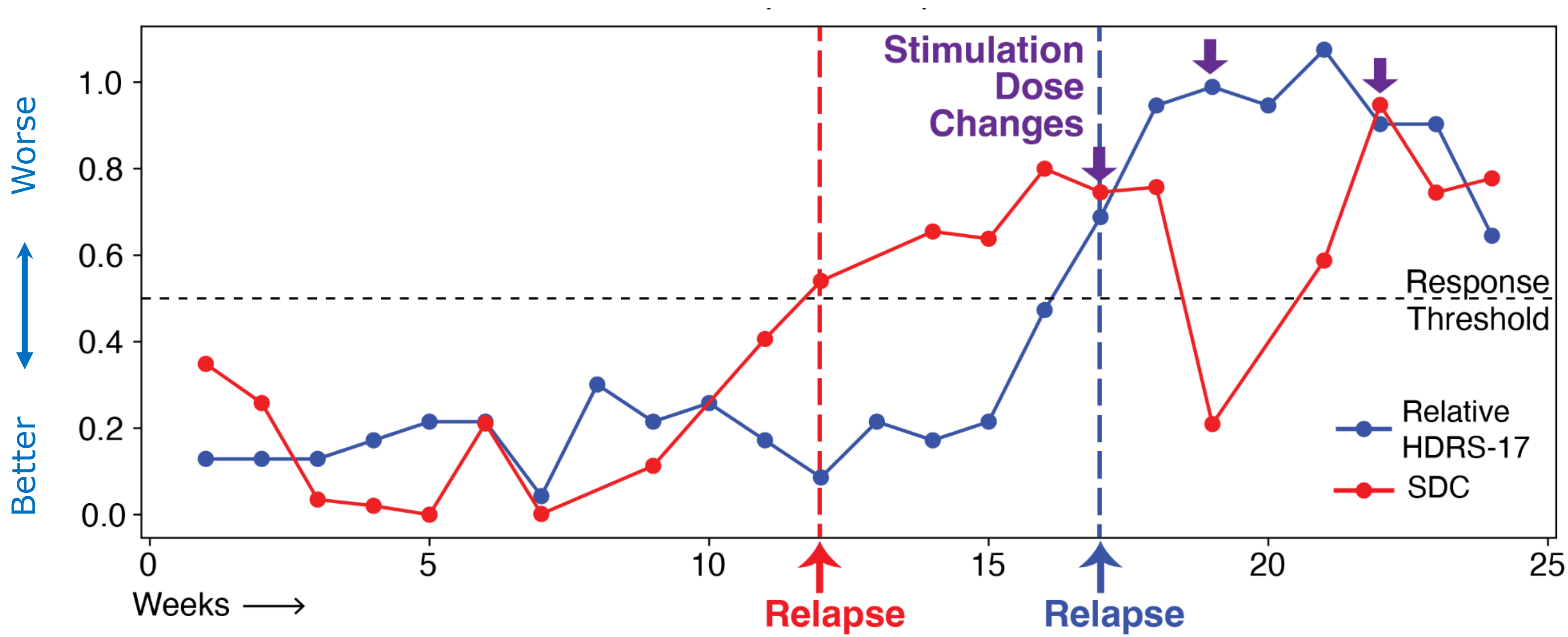


Evidence supporting stable depression recovery is an adaptation mediated effect



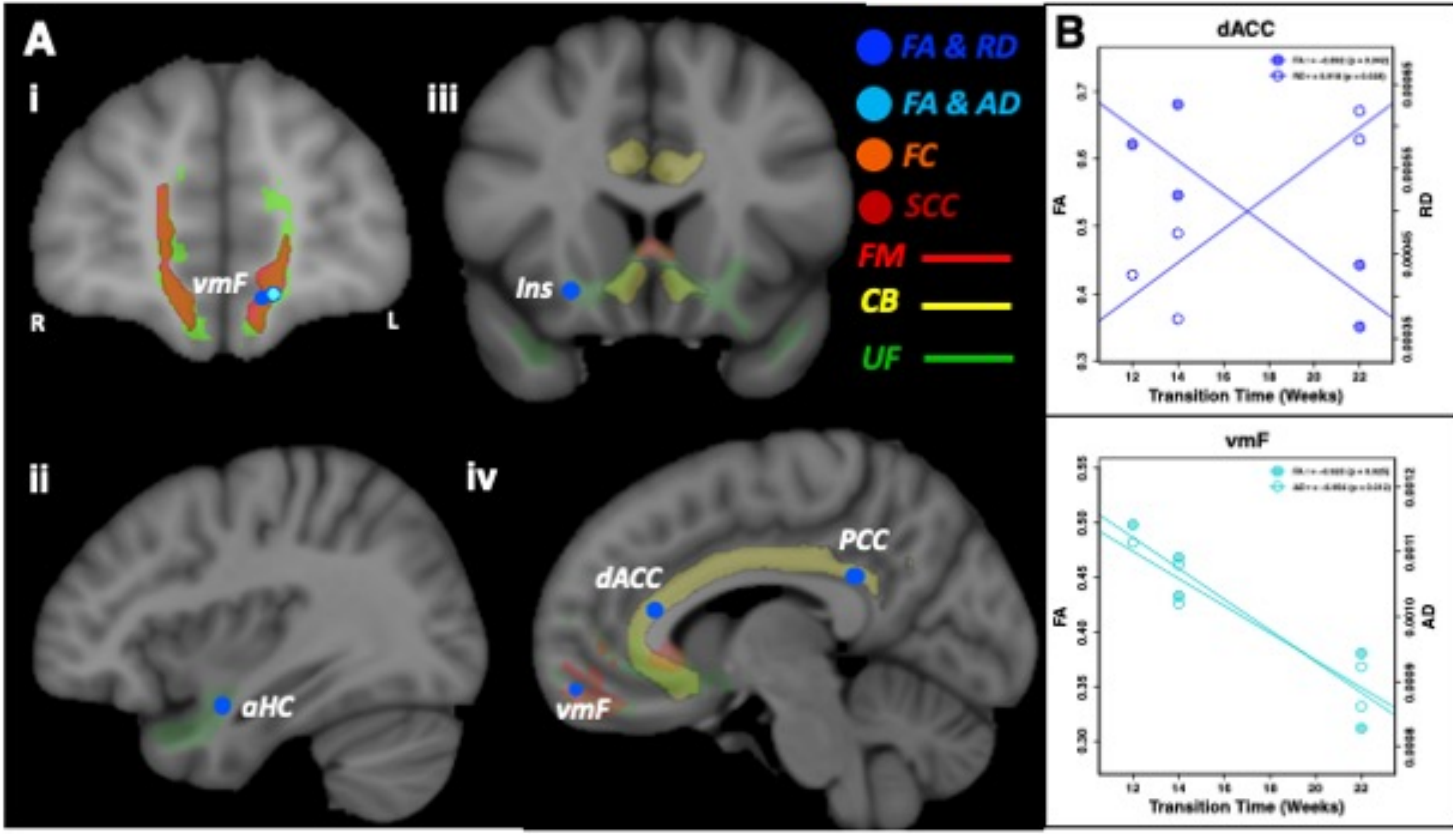
(Sendi et al. 2021; Smart et al. 2018)

Case study (relapsed responder)



Biomarker predicted relapse and need for treatment adjustment ***five weeks*** *before clinical deterioration*

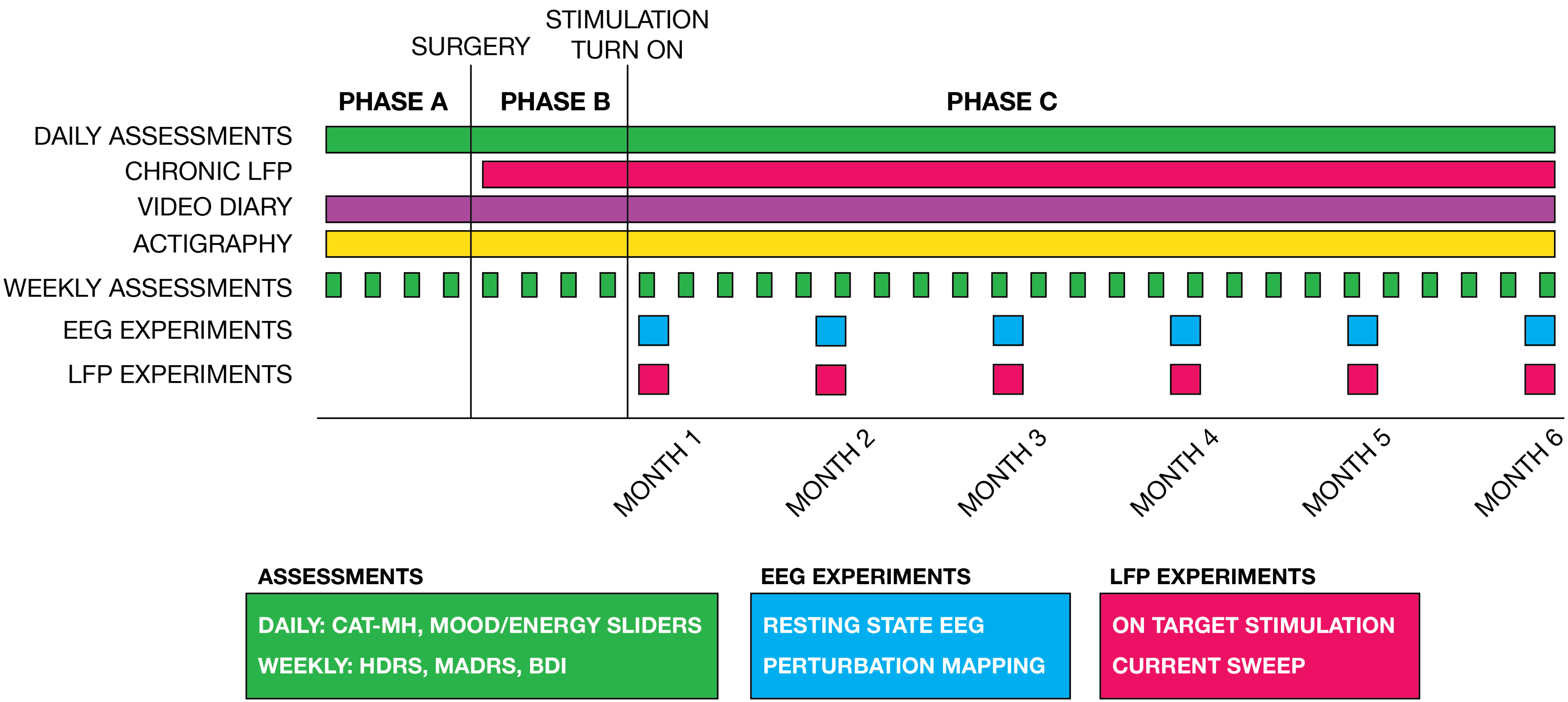
WM deficits correlate to transition time



- Most consistent microstructure result is a lesion (low FA) in dACC consistent with demyelination (high RD)

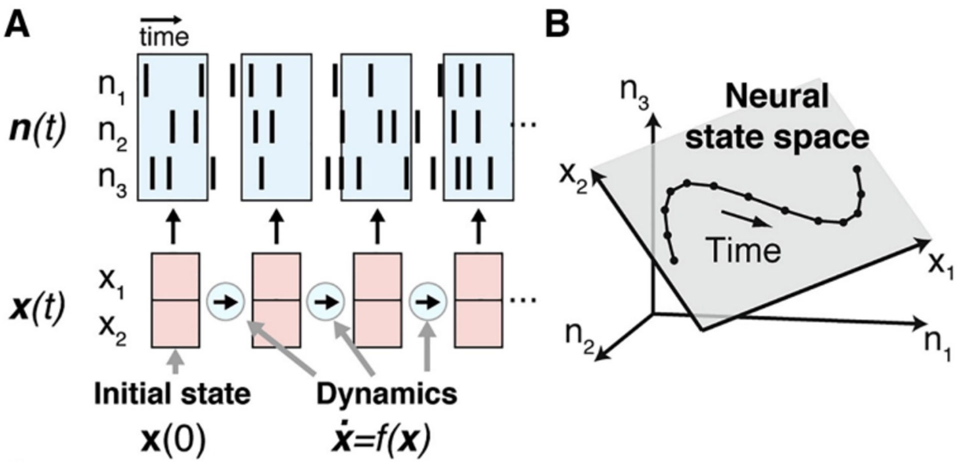
Next cohort data collection

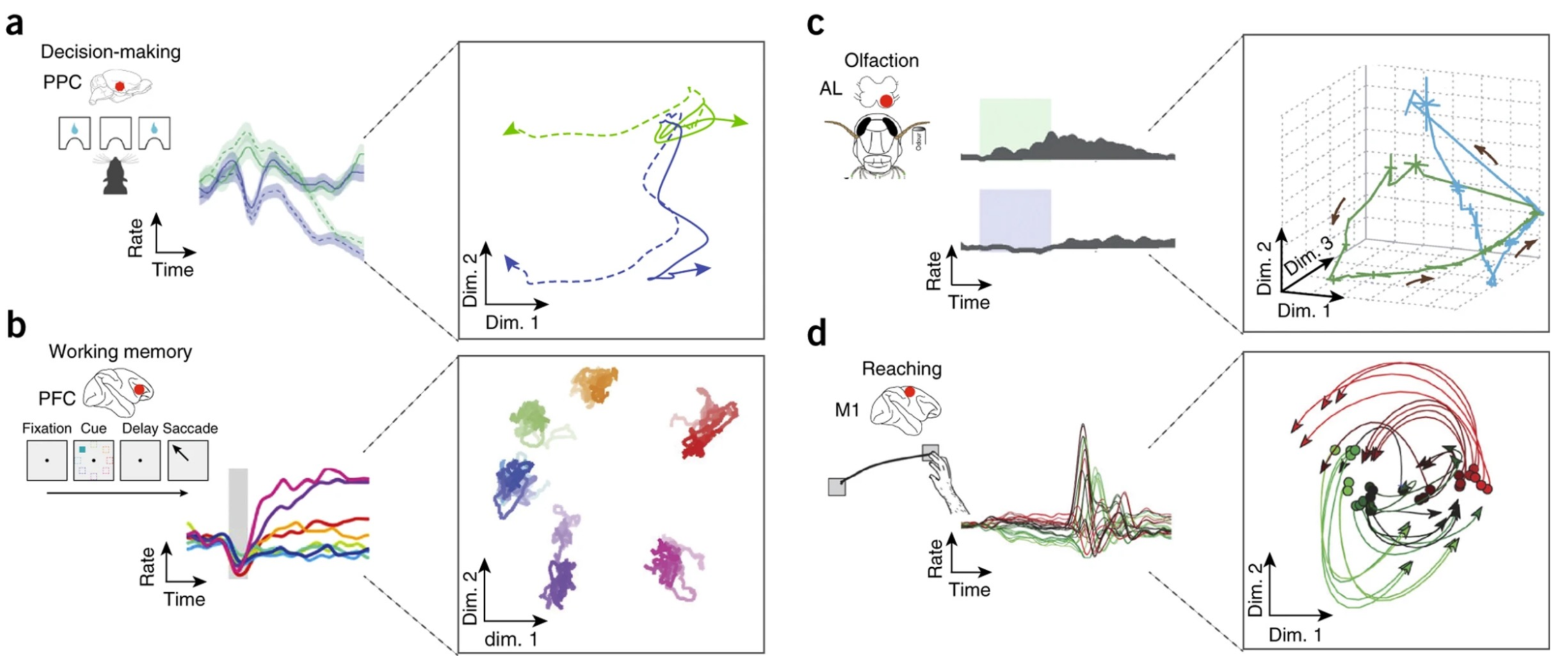
- Cohort (**n=10**) implanted at Mt. Sinai with Summit RC+S
  - 70% response and 40% remission
  - Daily stim off LFP recording at home (morning)
- Heterogeneous multimodal nonstationary time series data with presumed common latent variables



Dynamical Systems view of population activity

Pandarinath et al. • Dynamics in Motor Cortex with Application to BMIs

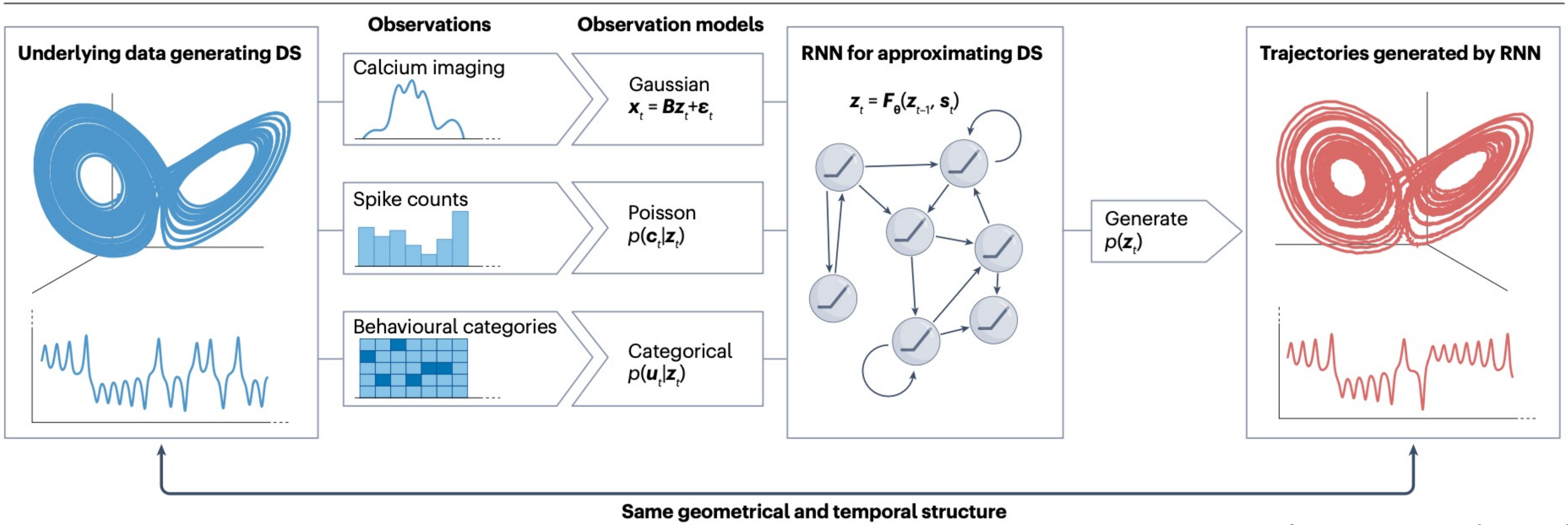




(Elsayed & Cunningham, 2017)

Latent dynamics of neural activity

- Movement toward conceptualizing brain activity as dynamical systems and disorders as aberrant dynamics



(Durstewitz, et al., 2023)

Linear Gaussian State Space Models

State Space Equations

$$x_{t+1} = Ax_t + \epsilon_x \qquad \epsilon_x \sim \mathcal{N}(0, \Sigma_x)$$

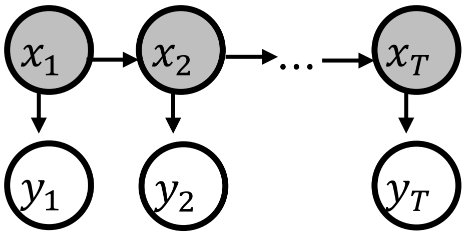
$$y_t = Dx_t + \epsilon_y \qquad \epsilon_y \sim \mathcal{N}(0, \Sigma_y)$$

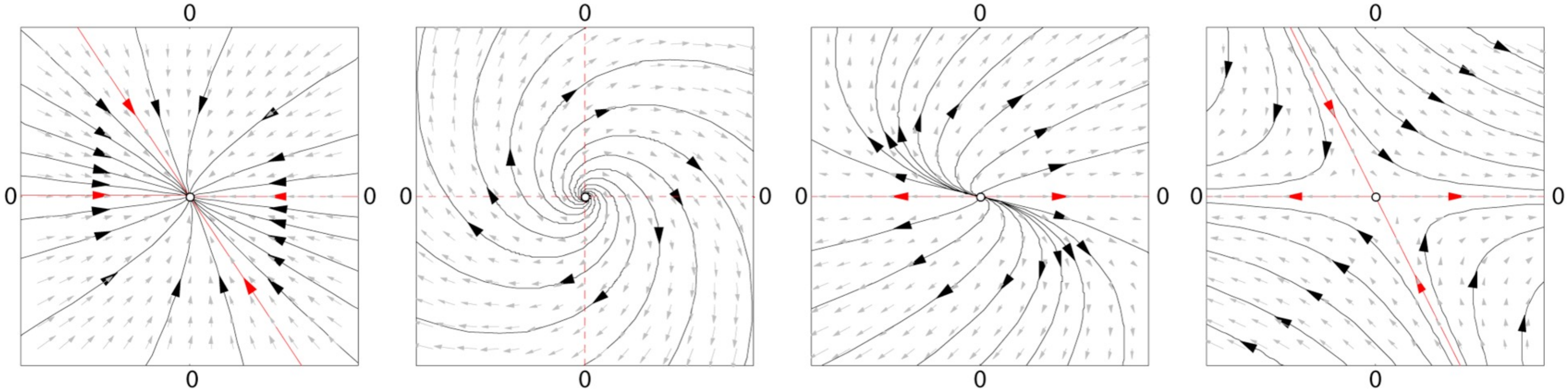
Probabilistic Model

$$p(x_{1:T}, y_{1:T}) = p(x_1) \prod_{t=1}^{T} p(y_t|x_t) \left[\prod_{t=2}^{T} p(x_t|x_{t-1}) \right]$$

Inference: Find posterior $p(x_t|y_{1:T}, \Theta)$

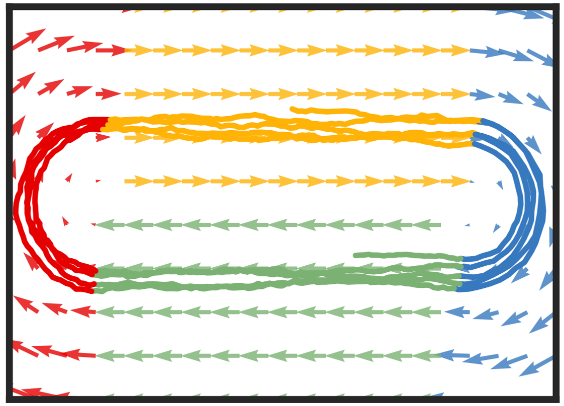
Learning: Find parameters $\Theta = \mathrm{argmax}_\Theta \, p(y_{1:T}|x_{1:T}, \Theta)$





LGSSMs are tractable and interpretable, but not expressive

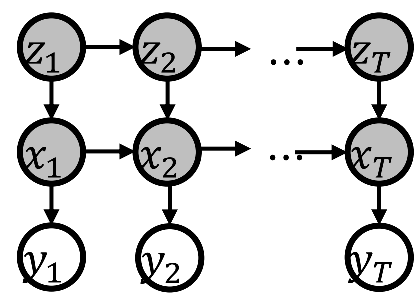
Switching Linear Dynamical Systems



Observations $y_t = D_{z_t} x_t + d_{z_t} + \epsilon_y$

Continuous state $x_{t+1} = A_{z_t} x_t + b_{z_t} + \epsilon_x$

Discrete state $z_{t+1} \sim \pi(z_{t+1} | z_t)$



Decomposed dynamics

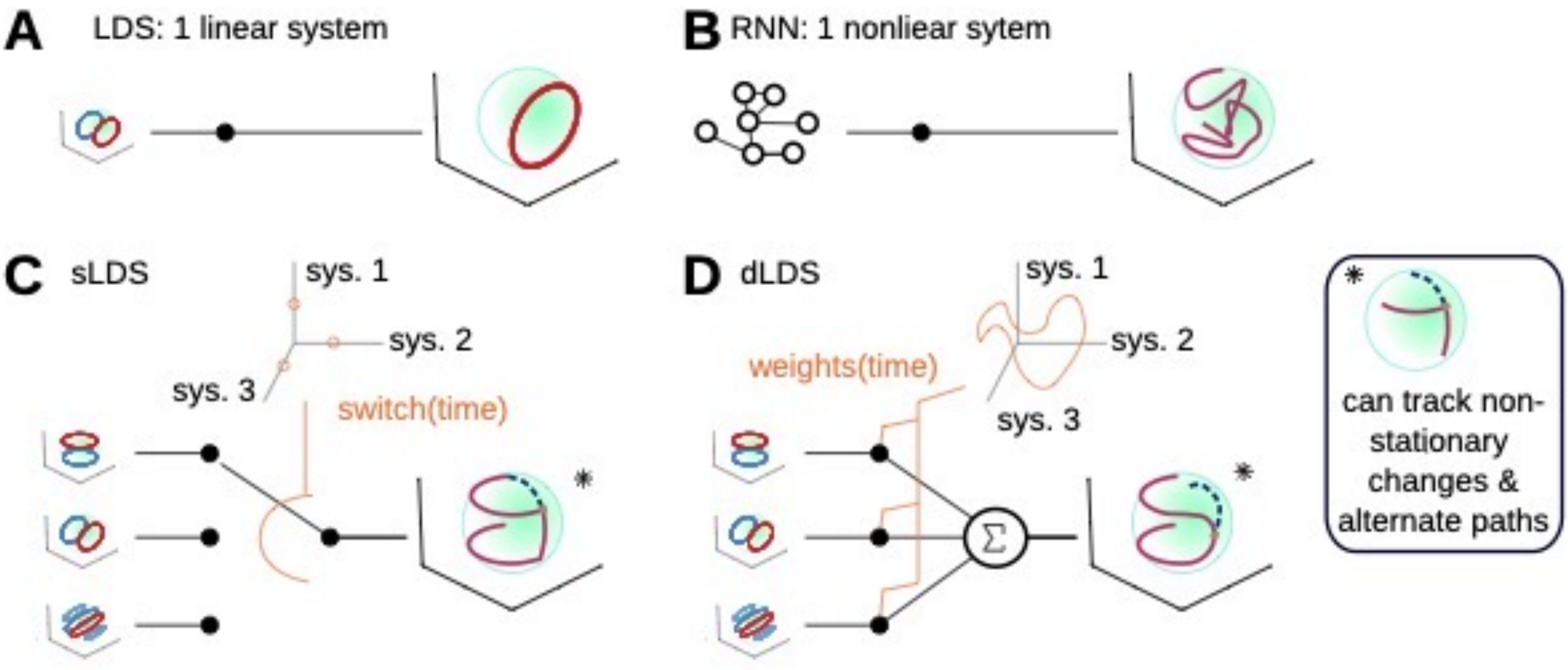
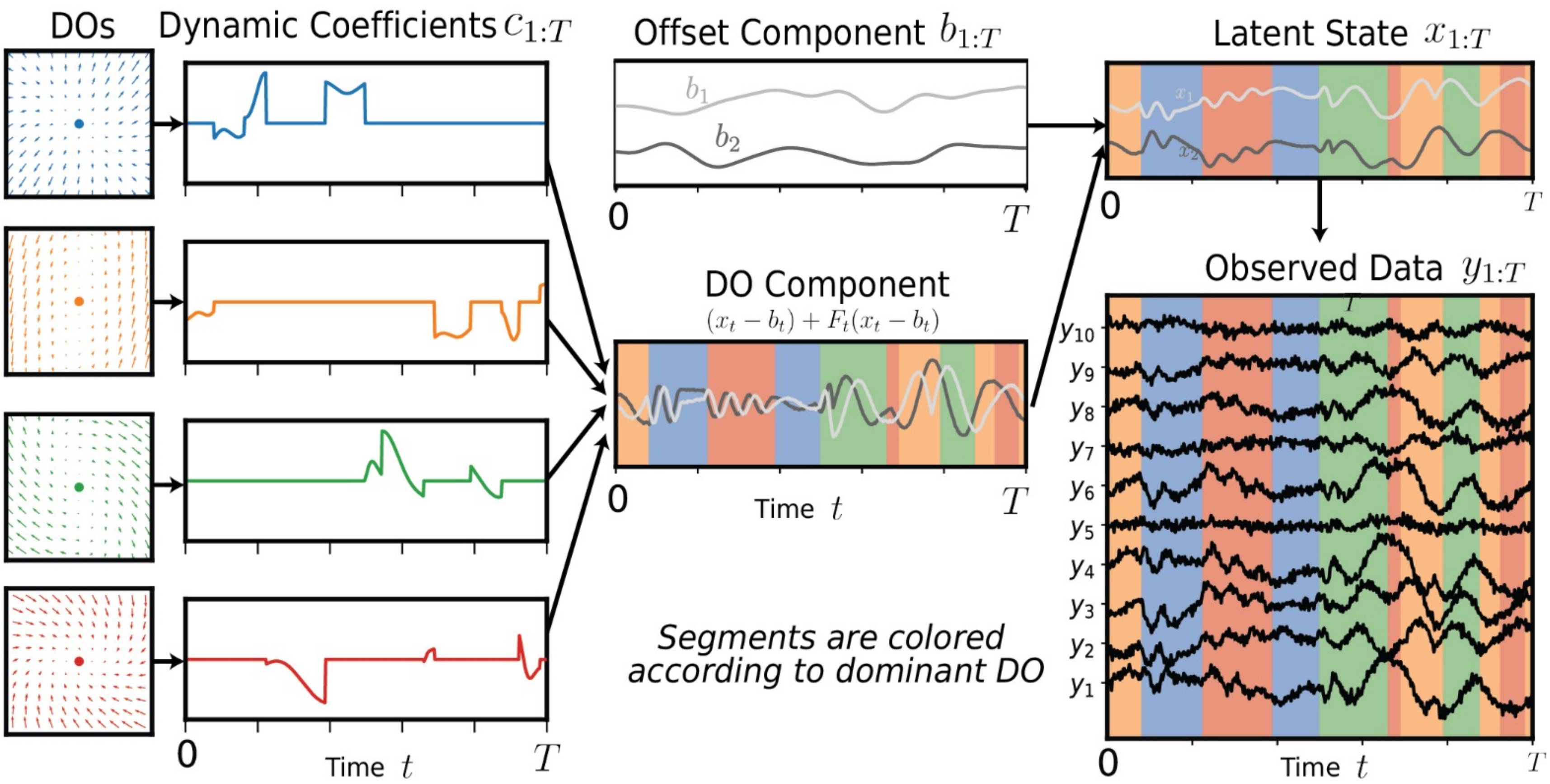


Illustration of decoposed LDS



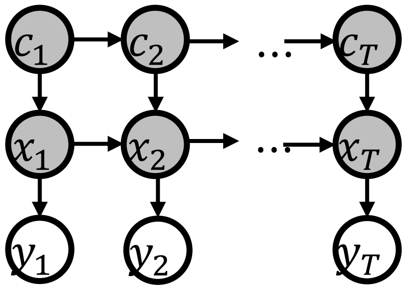
Decomposed LDS (dLDS)

Observations $y_t = Dx_t + \epsilon_y$

Continuous State $x_{t+1} = A_t x_t + b_t + \epsilon_x$

Dynamics Decomposition $A_t = \sum_{i=1}^{K} a_i c_{i,t}$

Coefficient Transition $c_{i,t+1} = c_{i,t} + \epsilon_c$



Graphical Model

$$\hat{x}_t, \hat{c}_t = \min_{x_t, c_t} \frac{1}{2} \|y_t - Dx_t\|_2^2 + \lambda_1 \left\| x_t - \left(\tilde{A}\hat{x}_{t-1}\right) c_t \right\|_2^2 + \lambda_2 \|\mathrm{c_t}\|_1 + \lambda_3 \|\mathrm{c_t} - \hat{c}_{t-1}\|_2^2$$

Problem: 1. Many interacting hyperparameters

2. Point estimate of latent variable for each time slice can propagate errors

(Mudrik, Chen, et al., arXiv)

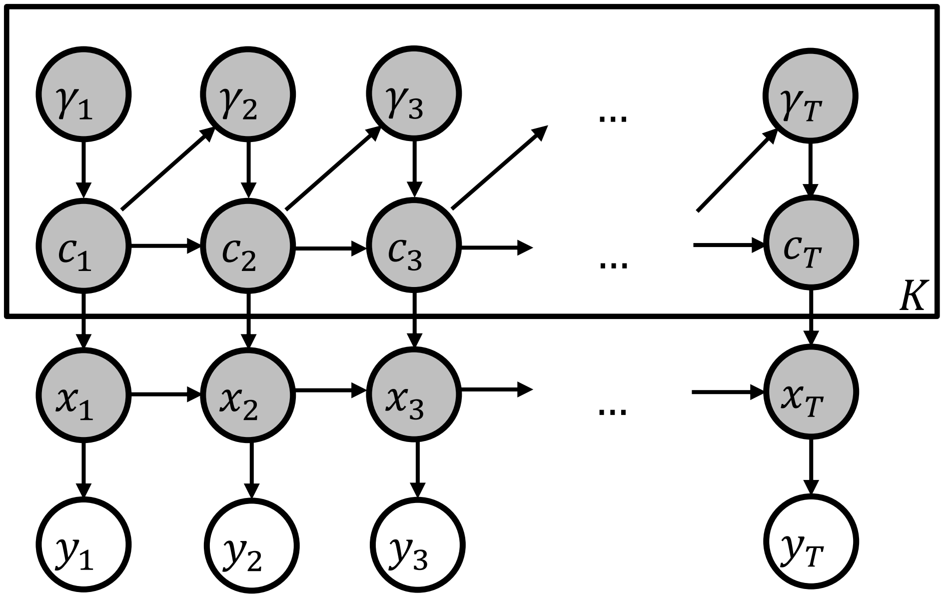
Probabilistic approach to dLDS

$$P(\gamma_t|c_t) = IG(\xi, \xi c_{t-1}^2)$$

$$P(c_t|\gamma_t, c_{t-1}) = N(0, \gamma_t)N(c_{t-1}, \Sigma_c)$$

$$P(x_t|x_{t-1}, c_t, b_t) = N(A_t x_{t-1} + b_t, \Sigma_x)$$

$$P(y_t|x_t) = N\left(D x_t, \Sigma_y\right)$$

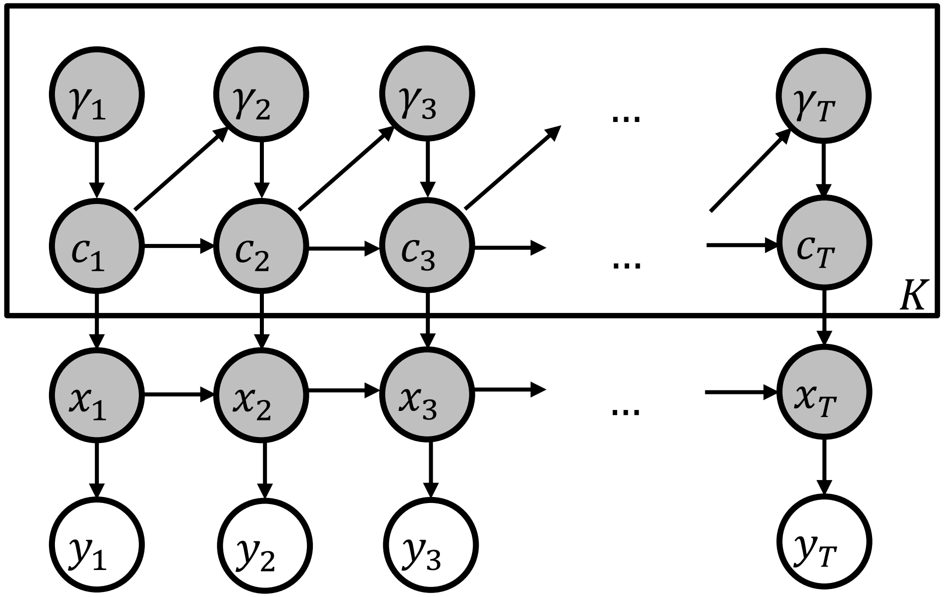


pdLDS Graph

(Chen, et al., in preparation)

Inference and Learning

Idea: Structured Variational EM

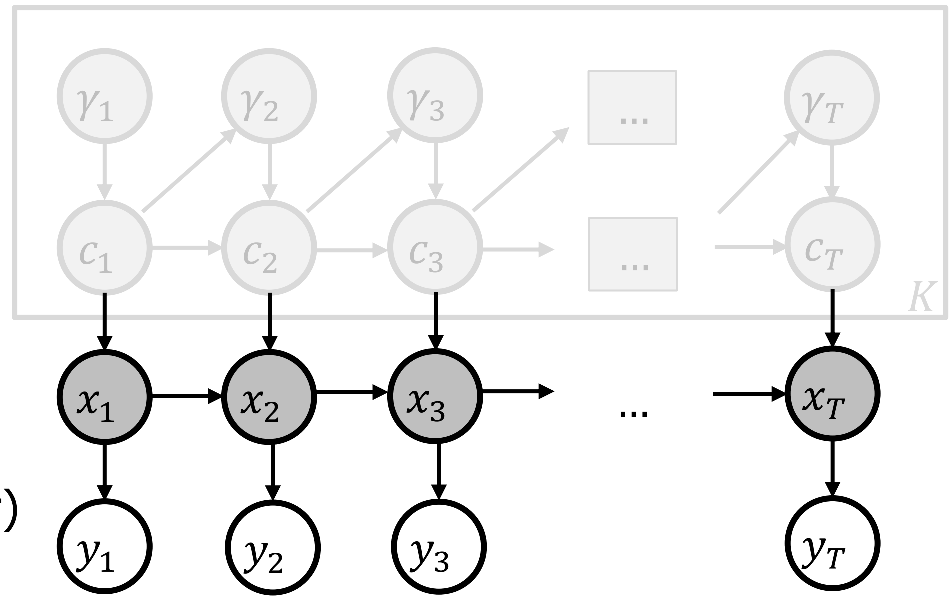


Inference and Learning

Idea: Structured Variational EM

Time-varying LDS

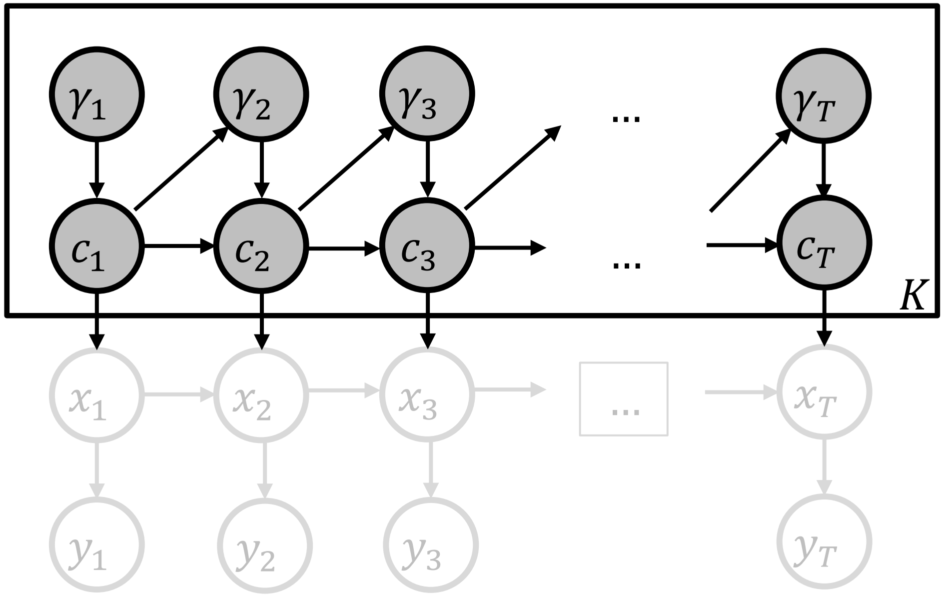
(Kalman Filter/Smoother)



Inference and Learning

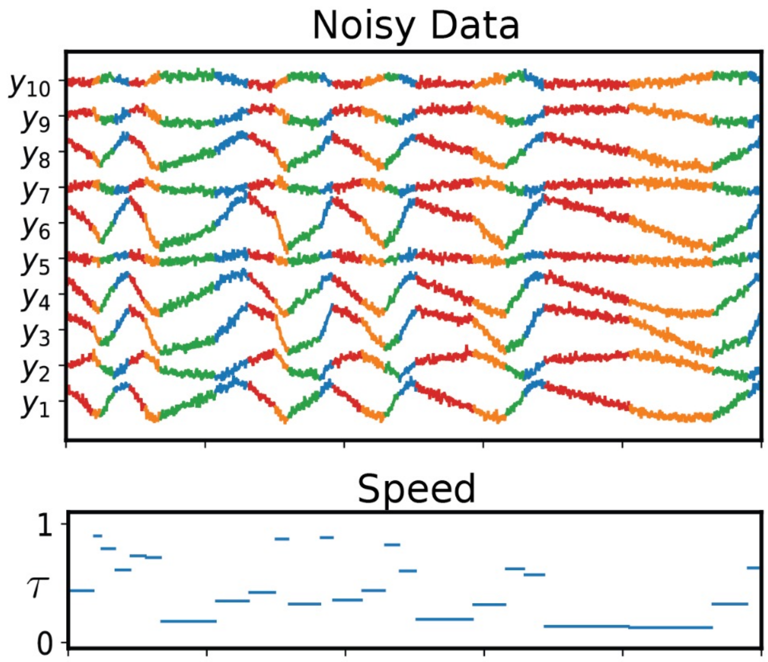
Idea: Structured Variational EM

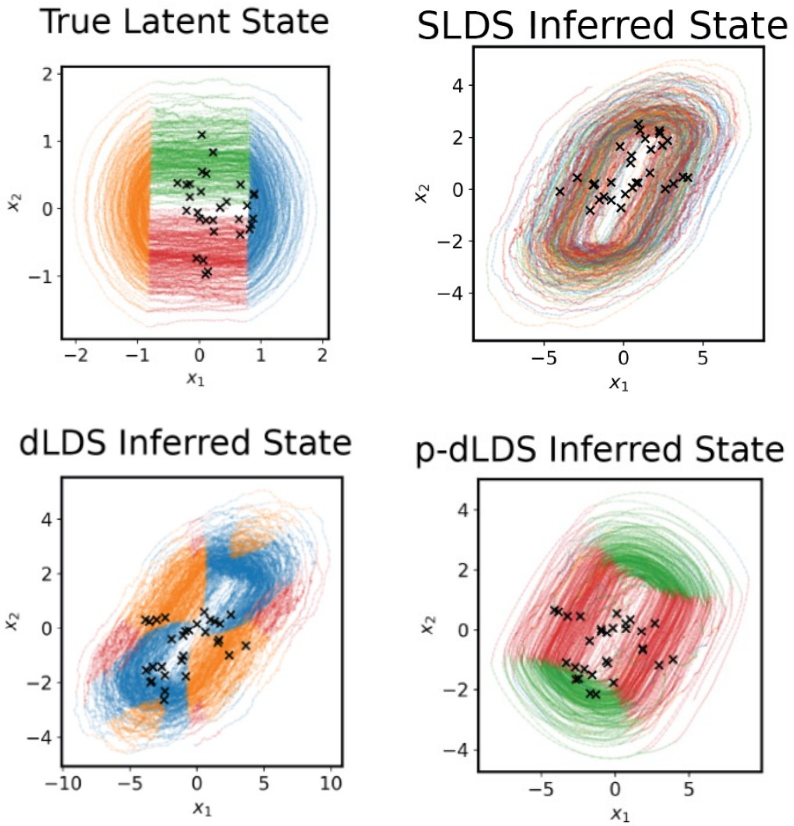
Sparse tracking
& smoothing
(efficient approximate inference)



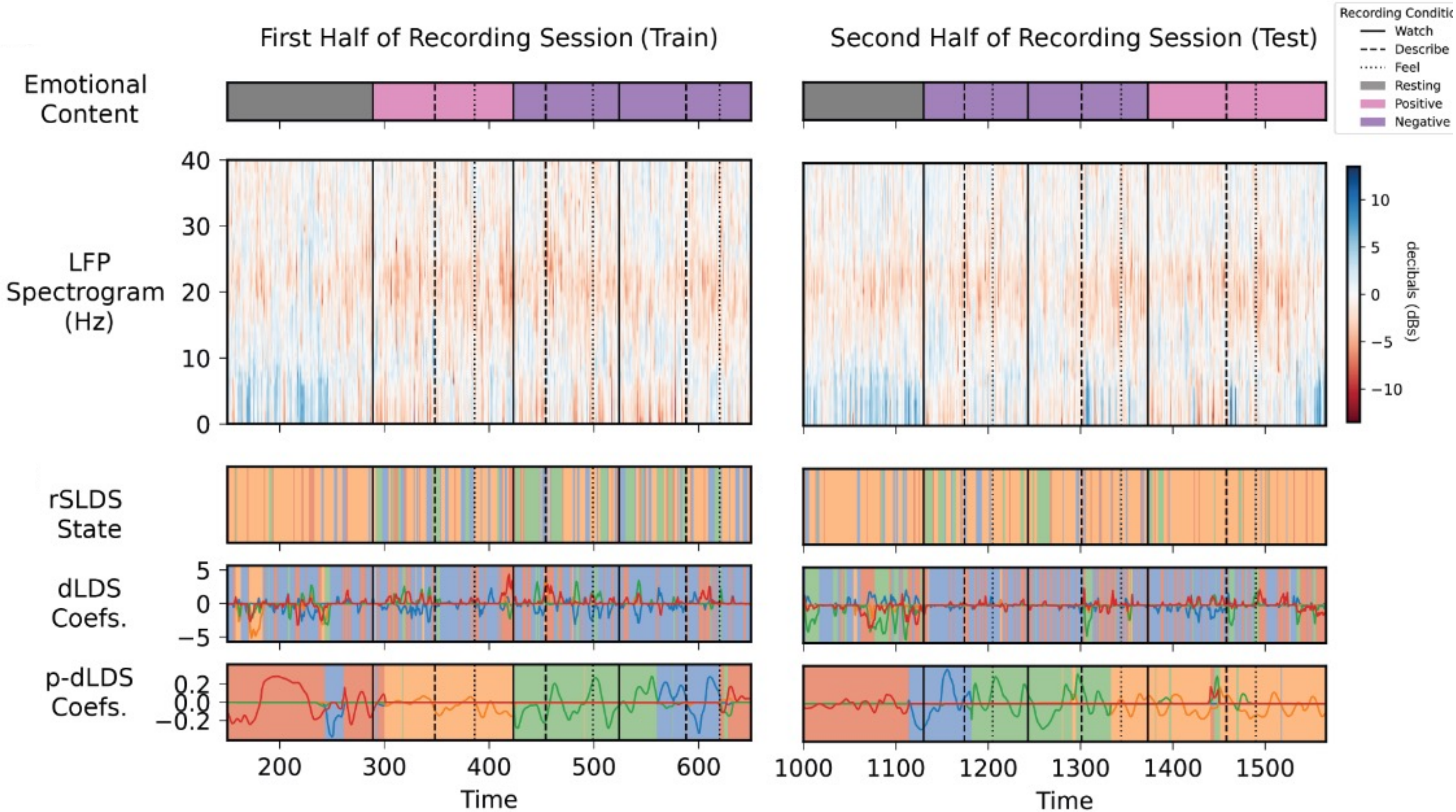
Iterate until Convergence

Noisy NASCAR Dataset





Clinical Neural Data



http://siplab.gatech.edu

crozell@gatech.edu

@crozSciTech