Closed-loop interventions for Causal ID of neural circuits

Audience: Systems neuroscientists interested in making more rigorous conclusions in circuit ID problems

Problem: While tools for stim & rec. Advance, we are still unable to make the causal statements we need to understand and treat the brain. We need a framework to guide the application of intervention to uncover causal links, and we need to understand its boundaries, limitations.

Goal: Provide a **straightforward** and **practically useful** conceptual framework for applying closed-loop to circuit identification problems

- What's the value of closed-loop?
- What can i say about causal connections given the experiments i'm doing?
- How can we make experimental design decisions which improve the strength of our hypothesis testing?

Closed-loop interventions for Causal ID

“Intervention facilitates causal identification. Closed-loop intervention facilitates ID moreso, especially in reciprocally connected systems”

Results:

- restatement of the field of causal ID in neuro
- conceptual walkthrough of why and where intervention matters
- Simple numerical demonstration
- More detailed characterization of what factors matter when
- Summarizing remaining challenges / future work

Figure Outline

- **1a. Concept.** - closed-loop identification
- **1b. Concept.** - interventions in causal ID framework
- **2a. Methods** - Network models of the brain (simulation methods)
- **2b. Methods** - Reachability in neural circuits (theory methods)
- **3. Results** - case-study walkthrough
- **4. Methods/Results** - Empirical CL-CI pipeline + results
- **5. Discussion** - Enumerating remaining challenges / opportunities

Figure 1a: Conceptual overview - closed-loop identification

Context: Introduction

Takeaway: What is closed-loop control, what does it look like in neuroscience experiments?

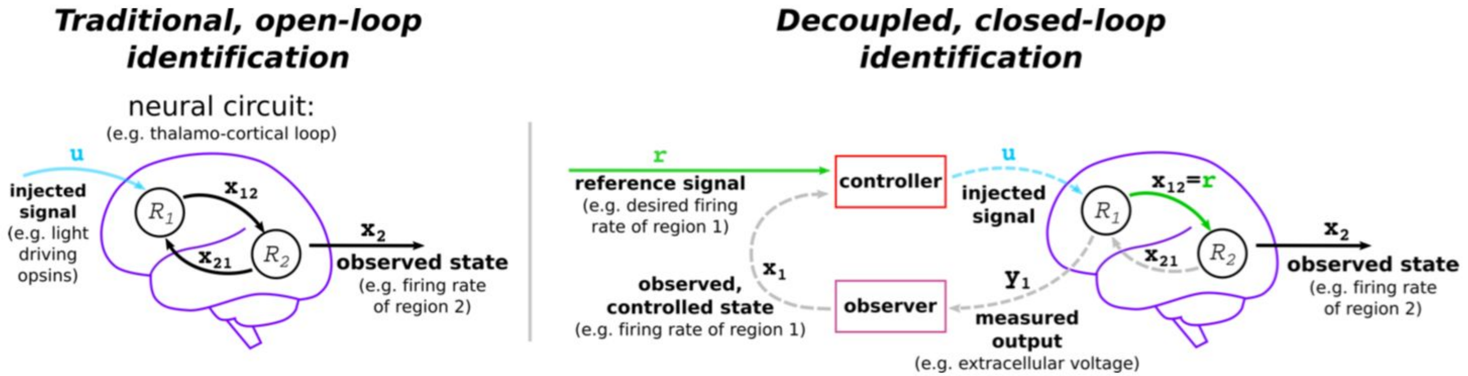
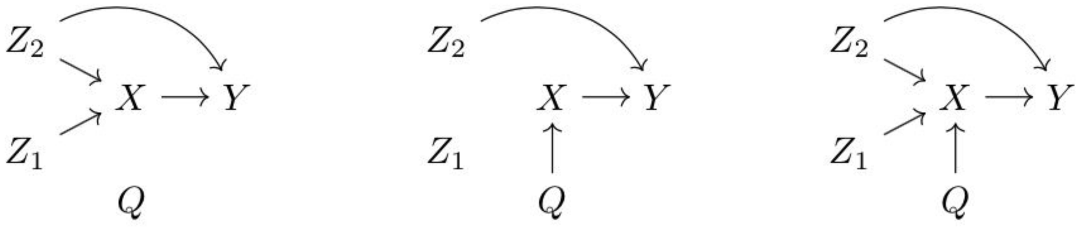


Figure 1b: Conceptual overview - interventions in causal ID framework

Context: Introduction

Takeaway: How do types of intervention relate to statements about dependent variables?

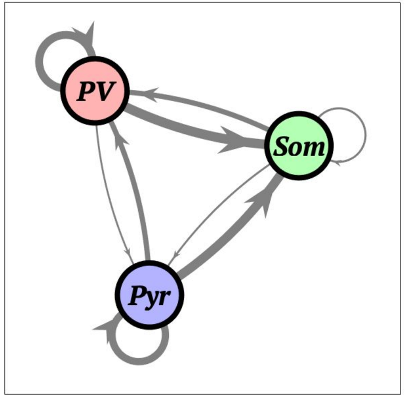


(a) Original circuit (b) CL control (c) OL control

Figure 2a: Methods overview - model

Context: Introduction

Takeaway: Here's the language of graphs, adjacency matrices, dynamics and interventions that will be used throughout the paper



$$\underbrace{\begin{bmatrix} \dot{x}_A \\ \dot{x}_B \\ \dot{x}_C \end{bmatrix}}_{\dot{x}} = \underbrace{\begin{bmatrix} w_{AA} & w_{AB} & w_{AC} \\ w_{BA} & w_{BB} & w_{BC} \\ w_{CA} & w_{CB} & w_{CC} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_A \\ x_B \\ x_C \end{bmatrix}}_{x}$$

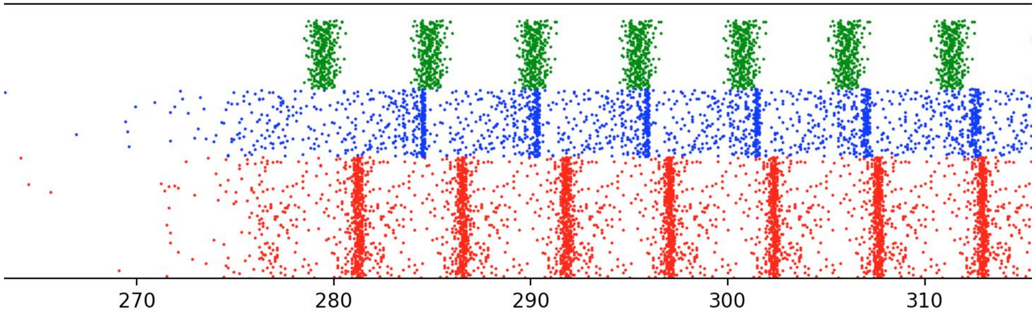
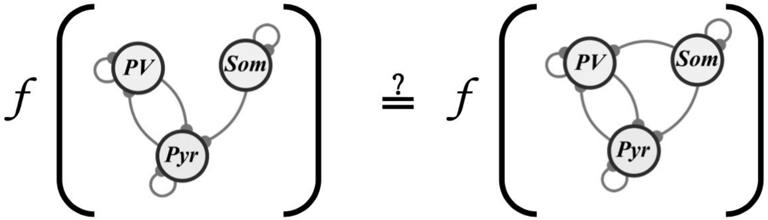


Figure 2b: Methods overview - reachability

Context: Introduction

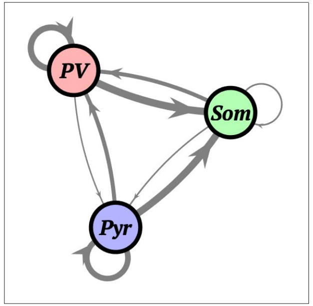
Takeaway: Here's the language of graphs, adjacency matrices, dynamics and interventions that will be used throughout the paper

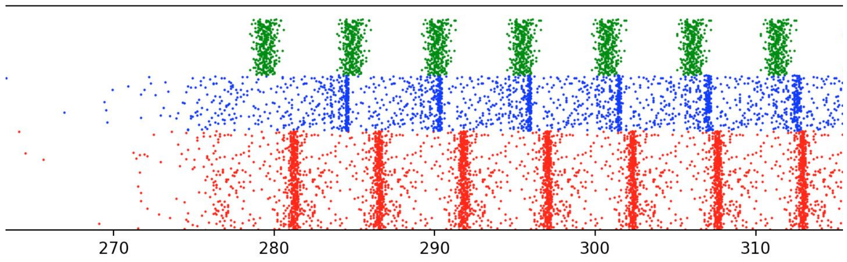


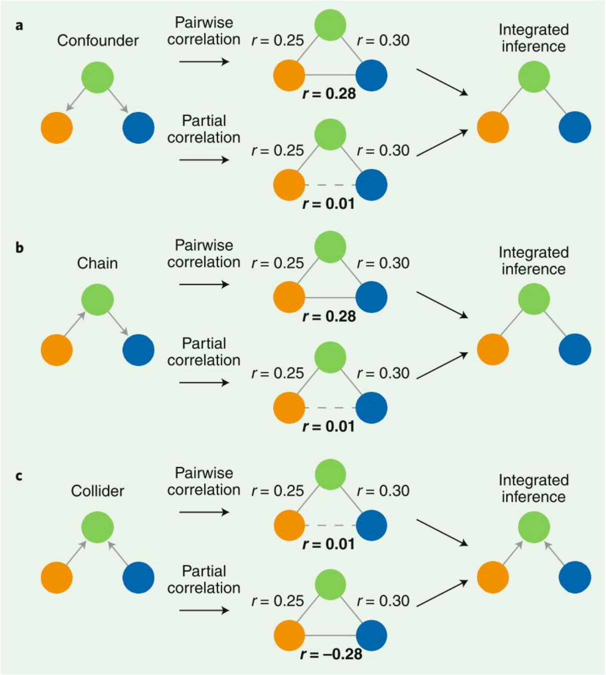
$$\underbrace{\begin{bmatrix} \dot{x}_A \\ \dot{x}_B \\ \dot{x}_C \end{bmatrix}}_{\dot{x}} = \underbrace{\begin{bmatrix} w_{AA} & w_{AB} & w_{AC} \\ w_{BA} & w_{BB} & w_{BC} \\ w_{CA} & w_{CB} & w_{CC} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_A \\ x_B \\ x_C \end{bmatrix}}_{x}$$

- Definitions of reachability, relationship to graphs
- "Rules of identification"
    - Statements about pairs of adjacency matrices which predict the success of identifying causal links

Figure 3: Case-study walkthrough





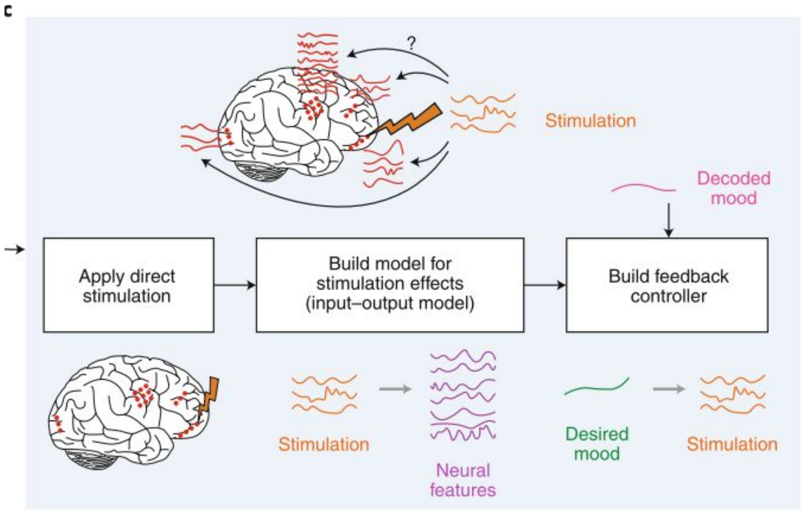


- May look a lot like my proposal slides
- 2 circuits, with outputs displayed
- looking at correlations that arise
- Demonstrating how closed-loop and causal tools can be brought to bear to disambiguate
    - Show the value of open-loop interventions along the way

Figure 4: Simulation pipeline + results

Example from Shanechi et al.

Pipeline



Characterization + examples

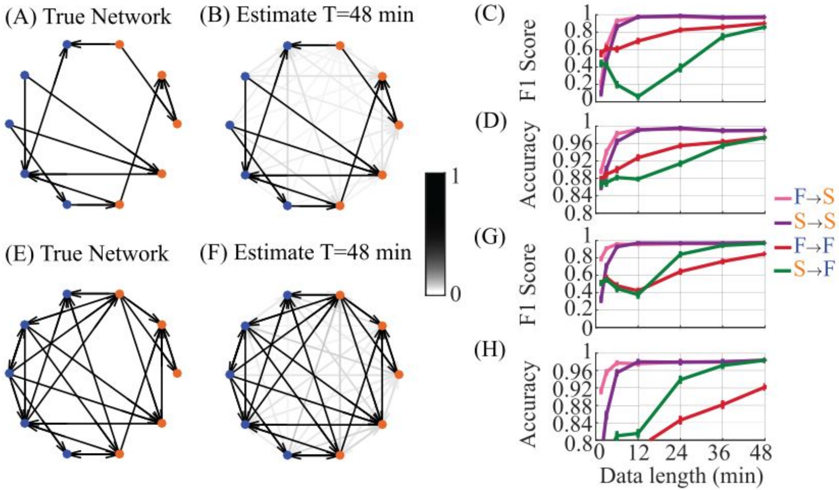
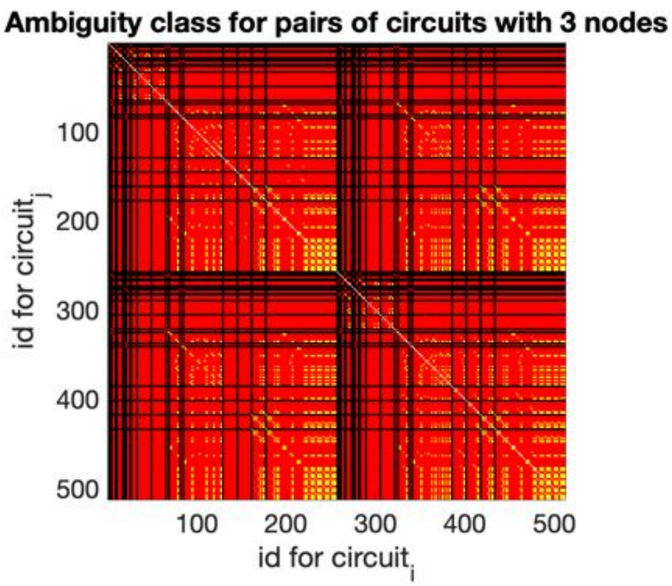


Figure 4: Simulation pipeline + results

<u>Binary reachability analysis:</u> <u>Impact of circuit adjacency on inference success</u>



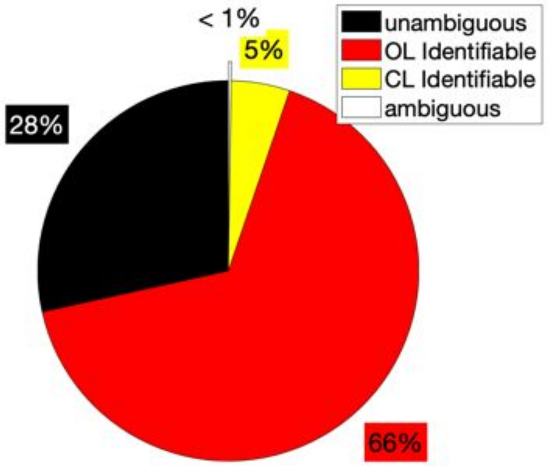
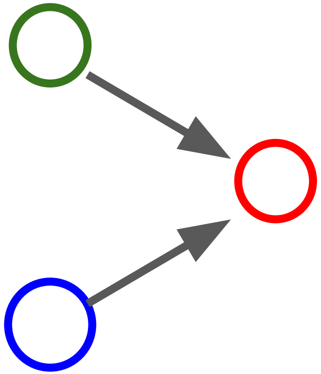


Figure 4: Simulation pipeline + results

Analysis of simulated circuits suggest open-loop intervention facilitates identification with less data

Network:
3 populations of integrate-and-fire neurons (simulated using Brian2)

Measure of causality:
Multivariate transfer entropy (estimated using IDTxl)

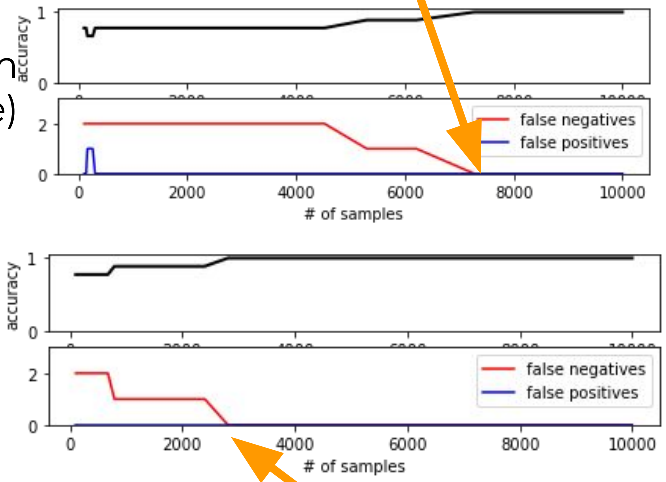


ground truth circuit

~7k samples need

No stim (passive)

w/stim (open-loop)



~2.5k samples need

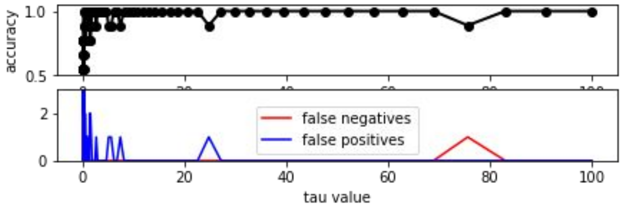
Stimberg, M., Brette, R. & Goodman, D. F. ***Brian 2, an intuitive and efficient neural simulator****. eLife (2019).*

Wollstadt P., Lizier et al ***IDTxl: The Information Dynamics Toolkit xl: a Python Package for the efficient analysis of multivariate information dynamics in networks***
PloS One (2013)

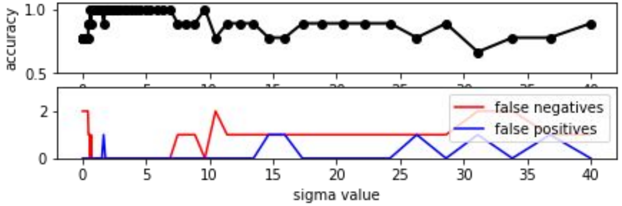
Figure 4: Simulation pipeline + results

<u>Impact of neuron parameters on identification success</u>

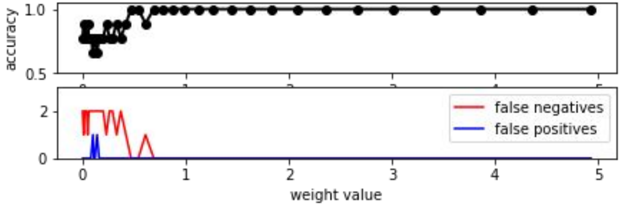
Tau:
Time-constant of internal noise
effective noise amplitude seems to increase with low tau



Sigma:
Std dev. of internal noise for neurons



Weight:
Connection strength,
Change in downstream voltage after and upstream spike



Next up:
Impact of intervention factors

- OL, CL
- Controller bandwidth vs network timescale
- Stimulus location relative to key nodes

Additional network factors

- Parameter heterogeneity
- Network size

Figure 5: Enumerating future work (?)

Challenges

	Examples
disturbances	additive noise, bias
data complexity / data length v.s. model parameterization	Models with large numbers of independent parameters require more data to be regressed effectively
limited observability	extracellular recordings currently sample from 1-10s of neurons simultaneously
limited actuation	optogenetic stimulation setups have few spatial degrees of freedom
unobserved confounds	**latent nodes with additive or multiplicative connections to observed nodes**
model uncertainty / mismatch	static nonlinearities, misspecified noise model

Spectrum of interventions

	Static	Time-varying
0: passive observation	recordings under highly controlled experimental conditions, i.e. from anesthetized animals	spontaneous recordings, esp. during awake behaving tasks
1: open-loop stimulation	**step-response characterization**	input frequency sweep, pseudo-random binary sequences
2: trial-adaptive stimulation	active learning, Bayesian experimental design	trial-adaptive stimulus design
3: closed-loop stimulation	Pre-determined stimulation triggered from behavior/neural activity, **sweeping across step-targets with various amplitudes**	sine-target sweep, naturalistic target characterization

Scope Decisions

- How much theory? What is its role?
  - Provide basic intuition?
  - Provide mathematical framework?
  - Translate results in causality literature to neuro?
  - New theoretical results?
- How far to push case studies?
  - Include laminar thalamocortical network models?

Theory: Interesting Questions

- Is there a graph/matrix representation which allows us to clearly state the effect of open and closed-loop intervention on the observable links in a circuit?
    - Adaptation of do-calculus to time series?
    - Empirical results with other causal inference techniques?
- What are the consequences of **imperfect control** for identification?
    - Can our conceptual framework predict these effects?
- How can we represent limited/budgeted interventions?

Core challenges to completing paper

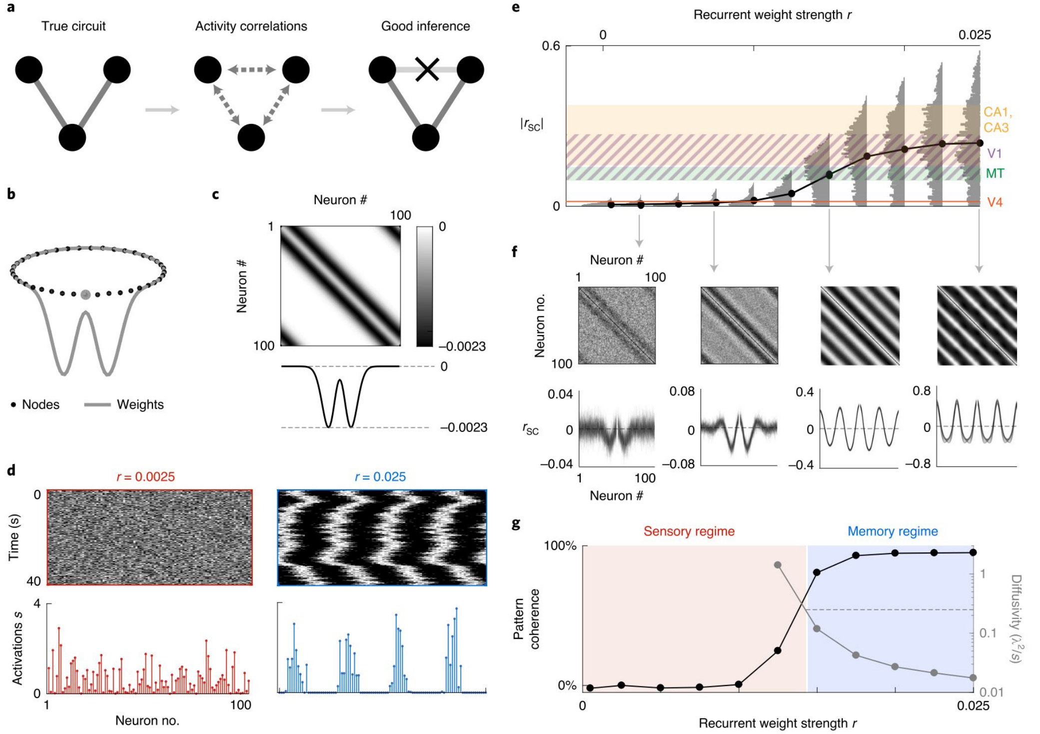
- Providing value in the niche between other similar papers without overextending the scope
- Our intuition is, for the most part, build around small (<10 node) networks, and we know some key aspects don’t generalize to larger (>100 node) networks
- Huge number of potential factors which influence successful ID

Next steps for Adam & Matt

- **Rescan literature**, compile succinct statement of **what the gap** in the field is
  - Likely centers around closed-loop intervention
  - **What to emulate for characterization?**
    -
- Identify what theoretical work is both feasible and useful to complement ongoing simulation experiments

Additional Slides

"Systematic errors in connectivity"
- Fiete, Das



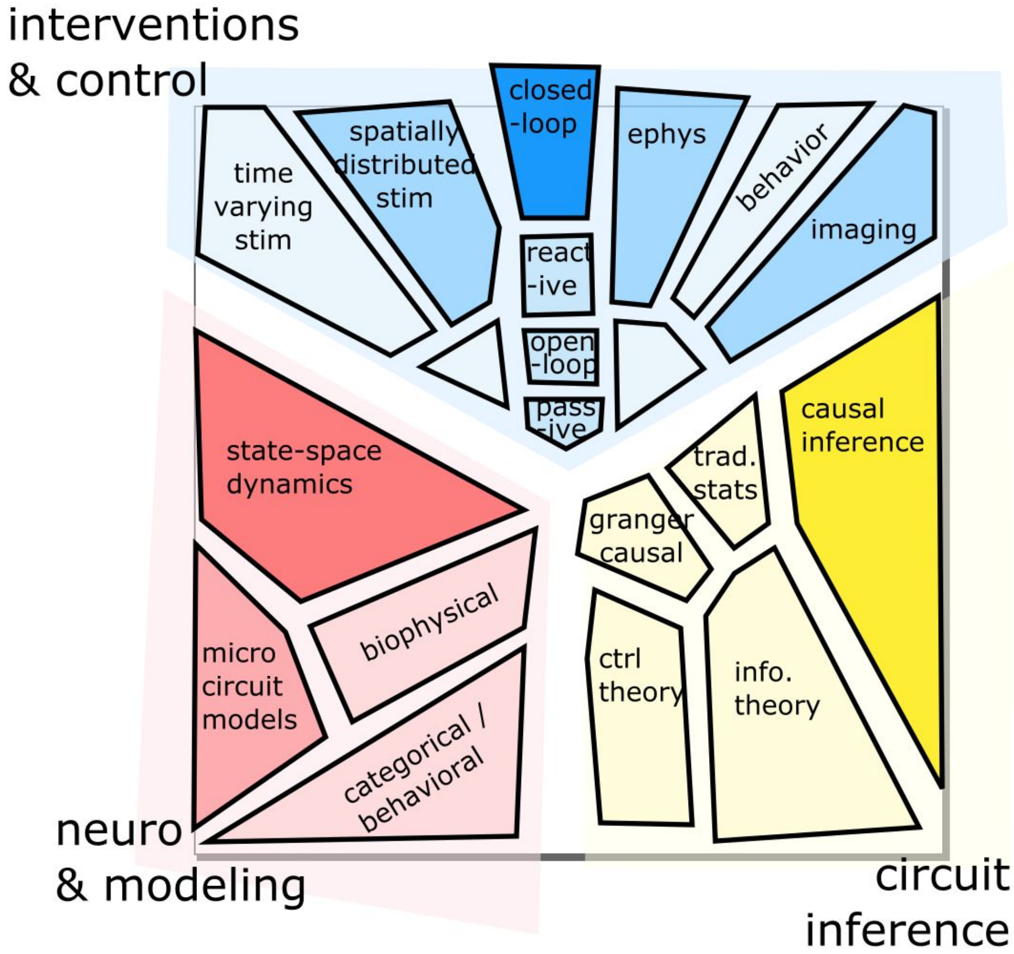
Hypothesis:

We hypothesized that inference models that are slightly mis-matched to the generative system cannot exactly capture (and thus explain away) all observed correlations derived from multi-hop interactions, and the residual unexplained correlations are then interpreted as excess direct connections.

Results:

- Sweeping recurrent weight strength mimics spectrum of connectivity across sensory and memory brain regions
- Bias - variance tradeoff leads to U-shaped inference success as a function of recurrent weight strength
- Intervention matters!

our paper's focus



Scope decision:
Contemp. vs time-lagged circuit representation

Start w/ contemporaneous examples -> move to time-lagged practical examples only in empirical section

Time-lagged throughout

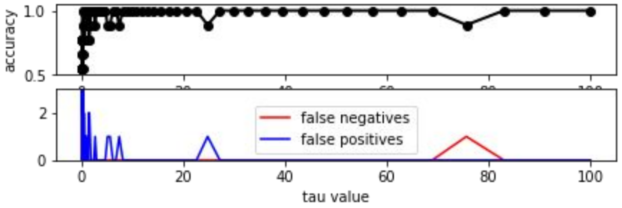
Or whole paper is contemp. Framework, and mention challenges/opportunities of time-lags

Start with time-lagged view, this perspective is central

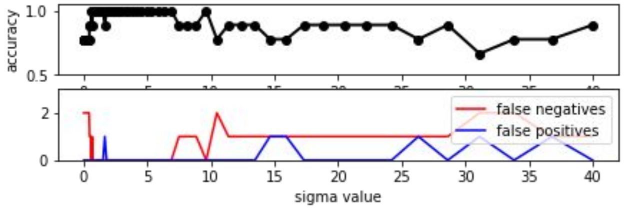
But simple intuitive tools from contemp. Circuits allow you to tackle that

Could also move from time-lagged to contemp. by using parameter domains where contemp. Is more appropriate

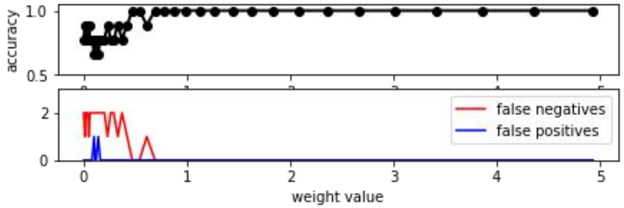
Tau:
Time-constant of internal noise
effective noise amplitude seems to increase with low tau



Sigma:
Std dev. of internal noise for neurons



Weight:
Connection strength,
Change in downstream voltage after and upstream spike



In summary:

1. Lower-frequency internal noise (higher tau) makes identification easier

2. For the amplitude of internal noise, there's a tradeoff between SnR and having enough noise to generate spontaneous spiking. This results in perfect identification only for a limited range of sigmas

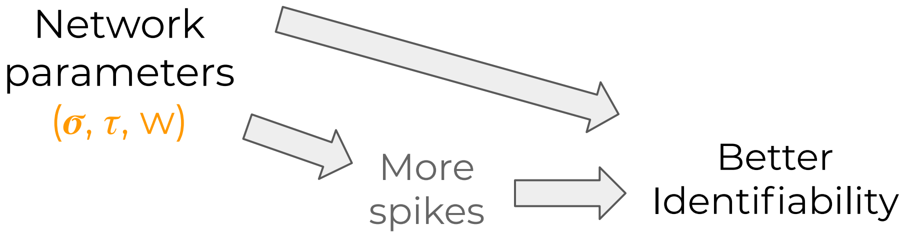
3. Identification gets easier with increasing synaptic weights

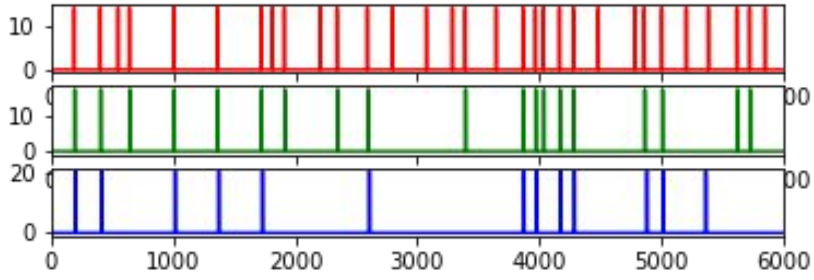
Next Steps

We understand how identifiability correlates with network parameters (under passive observation)

I suspect **total spike counts** are an important mediator in the success of identification

For future analysis we should control for overall spike counts and/or include spike counts as a predictive factor in analysis





Next Steps

We also have evidence of stimulation improving identifiability

While elevating firing rates are one important goal of OL & CL stimulation, it would be interesting to see if these interventions **improve identifiability beyond increasing spike counts.**

We also need to dig more into the interplay of **timing** for interventions relative to timescales in the network:

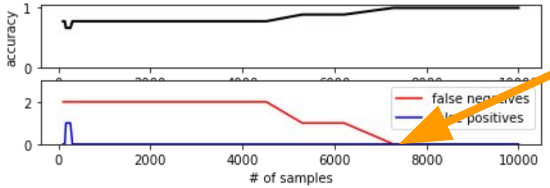
Sweep bandwidth and delay of controller against network noise timescale and synaptic delay

Intervention parameters (passive, OL, CL)

More spikes

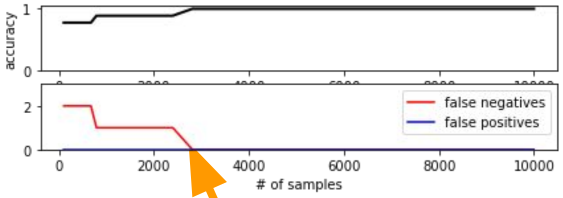
Better Identifiability

No stim (passive)



~7k samples need

w/stim (open-loop)



~2.5k samples need

Additional factors to dig into

Network

Role of negative weights

neurons per population

(multi-unit population firing rates versus multi-channel spiking)

Within-node heterogeneity

Circuit complexity & biophysics(EI, Laminar, sparse networks)
Compare spike-based versus voltage-based identifiability

Intervention

Does time-varying matter?

(square wave versus static for the same spike counts)

How much does stim location matter?

Simplest version: stim @ source versus terminal node should matter

Stim @ node with high out-degree also results in magnified impact

Theory

General: How do we link these sims to theory developed with matt?

Specific: Can our adjacency analysis *rank* which circuits should be more identifiable

Is there a simple shintani-esque formula for identifiability in N-node LIF circuits?

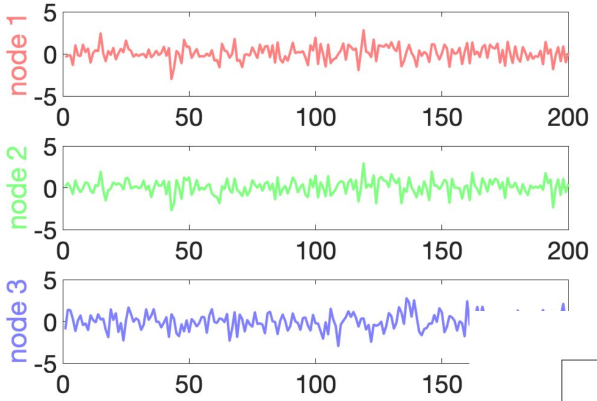
FR = exp(w*tau +mu - sigma^2), Spks = FR*duration

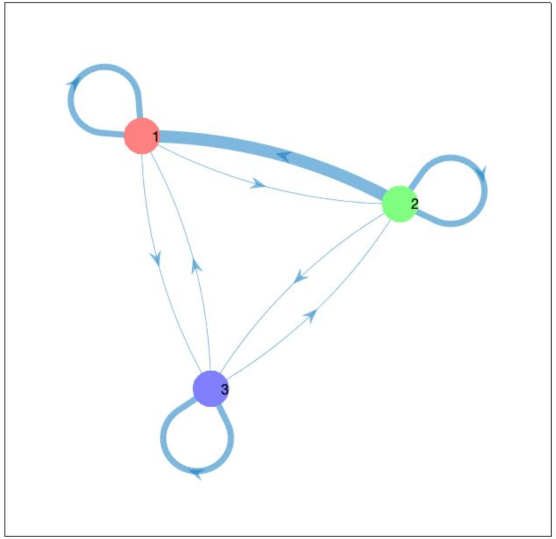
ID = sigmoid(Spks)

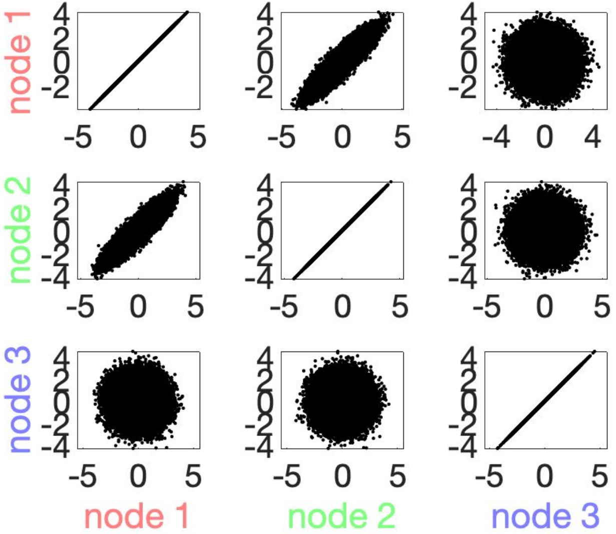
Tools for ID

Component functions in identifiability

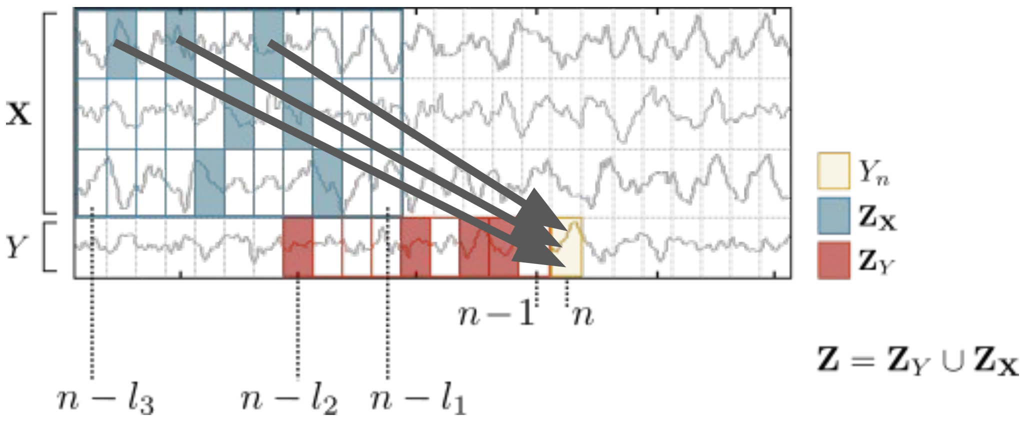
`reachability(A)`	
`passiveAmbiguous(A,B)`	Reach(undirect(A)) == Reach(undirect(B))
`openloopAmbiguous(A,B)`	Reach(A) == Reach (B)
`closedloopAmbiguous_i(A,B,i)`	Reach(control(A, i)) == Reach(control(B, i))
`closedloopAmbiguous(A,B)`	{Reach(ctrl(A,i))== Reach(ctrl(B,i)) } for all i


node 1
node 2
node 3
5
0
-5
0
50
100
150
200


1
2
3


node 1
node 2
node 3
node 1
node 2
node 3

The IDTxl toolbox!

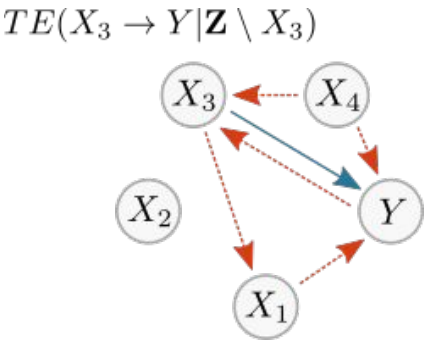


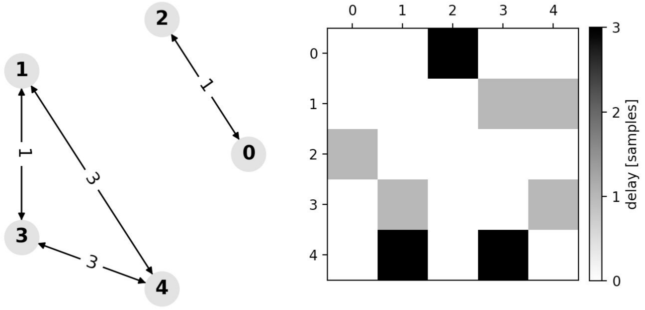
Might set:

L_1 = underestimate of synaptic delay

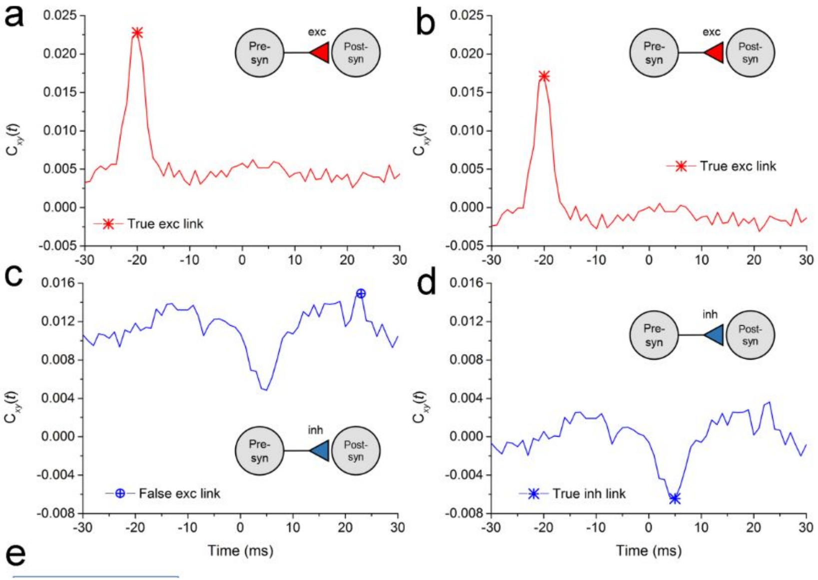
L_3 = overestimate of n*syn delay

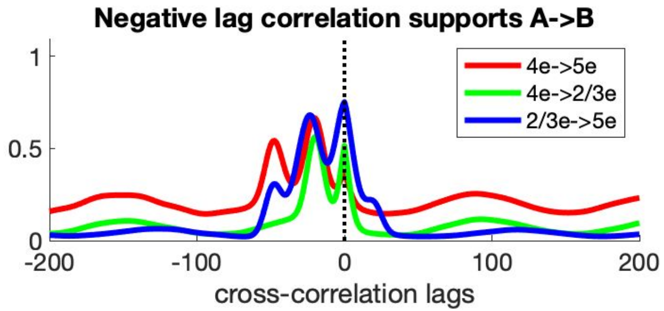
L_2 = autoregressive time constant





Identification of excitatory-inhibitory links and network topology in large-scale neuronal assemblies from multi-electrode recordings





Comparing cross-correlation to TE

Evaluation of the performance of information theory-based methods and cross-correlation to estimate the functional connectivity in cortical networks

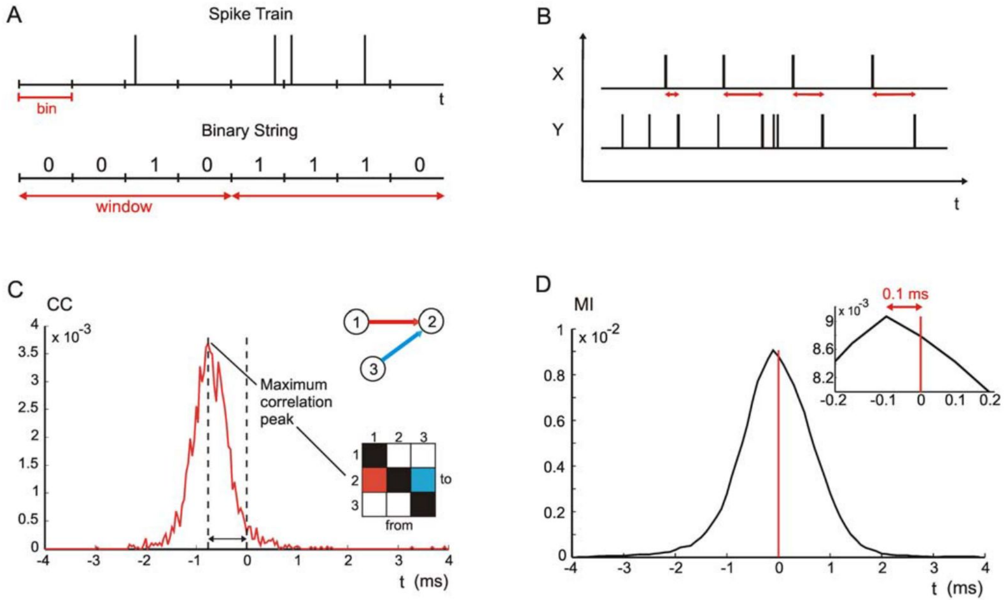


Figure 2. Schematic overview of the considered connectivity methods. (A) Binary string is created starting from the spike train. A window is selected to evaluate the TE and to define the MI symbols (window = 0.3 ms). (B) Cross-inter-spike-intervals (cISI) between neurons *X* and *Y* are highlighted by the red arrows. (C) Cross-correlation function between neuron 1 and 2. The directionality of the connection is evaluated considering the peak latency from zero. (D) Mutual information (spike count approach) function related to a pair of nodes of the network model. The inset shows that the MI peak value falls close to the zero time shift (value −0.1 ms).
doi:10.1371/journal.pone.0006482.g002

Example figures from similar papers

Results

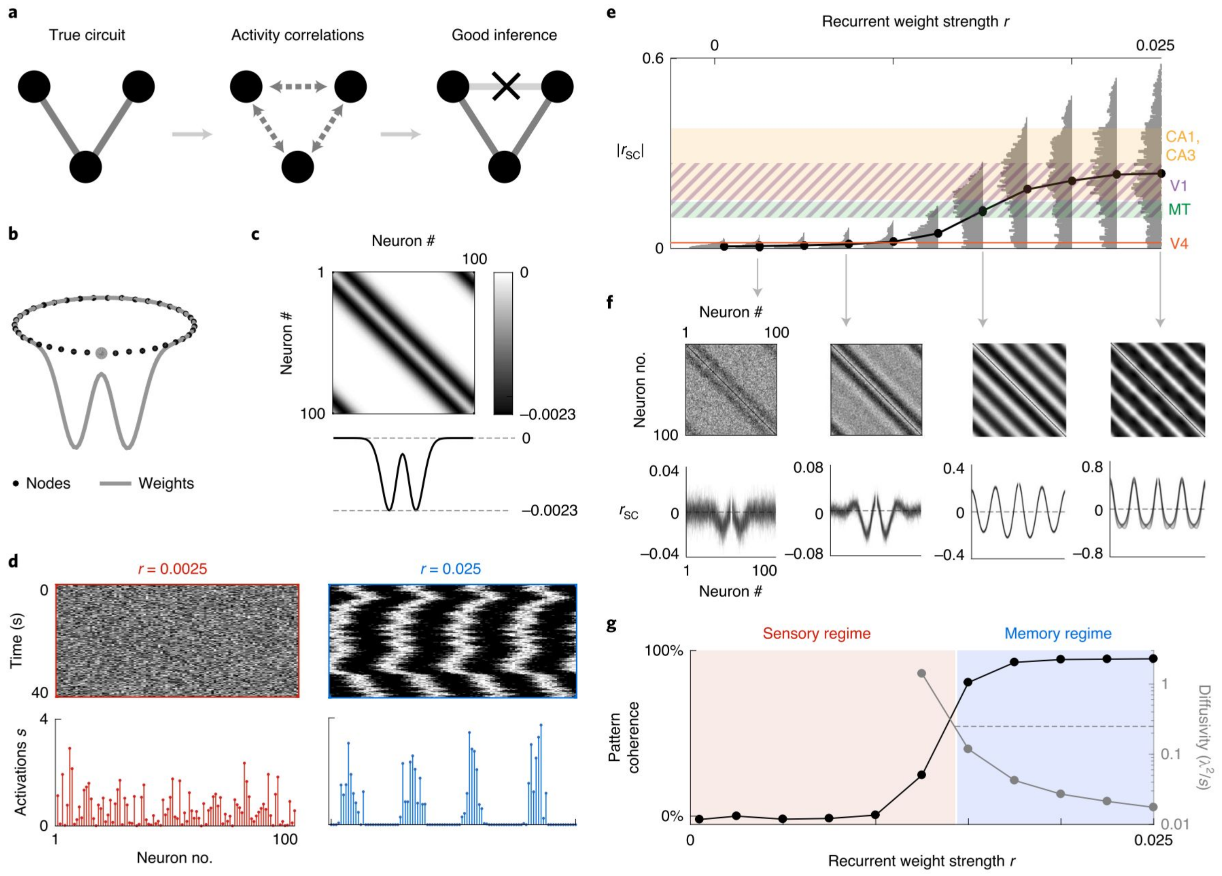


Fig. 1 | Structure and dynamics of the generative network. a, Left: schematic three-neuron circuit with two connections. Center: all neurons are correlated.

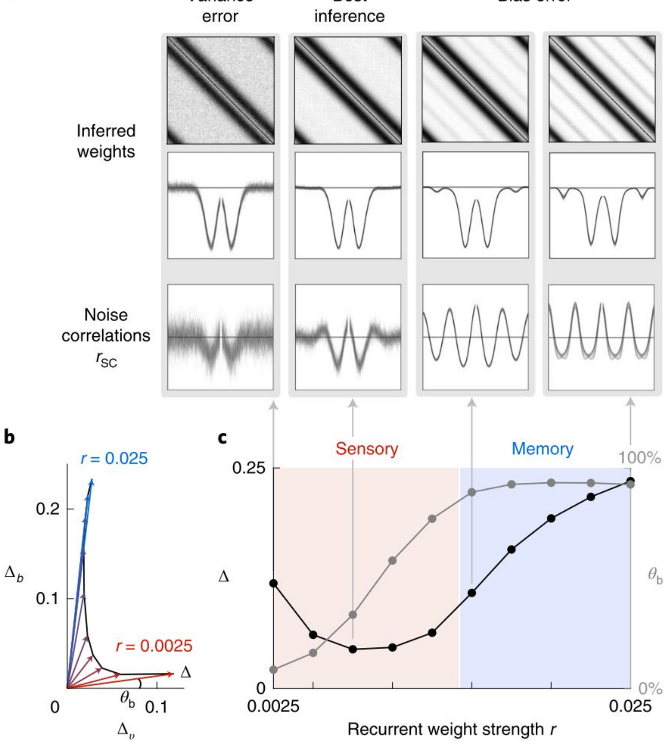


Fig. 2 | Quality of circuit inference using 10^8 spikes from a fully observed ring network, as a function of weight strength. a, Inferred weight matrices $\hat{\mathbf{W}}$ (top row), superposed rows of the weight matrix (middle row; line marks zero) and raw noise correlations (bottom row), at different weight strengths. **b**, Total inference error (arrows) as a vector of independent variance (Δ_v) and bias (Δ_b) error components (see Methods), at different weights. The vector magnitude Δ is the total inference error (as a fraction of the magnitude of the true weights vector), and the normalized angle θ_b is the fraction of bias error in this total error. **c**, Total inference error Δ and fraction of bias θ_b, against weight strength.

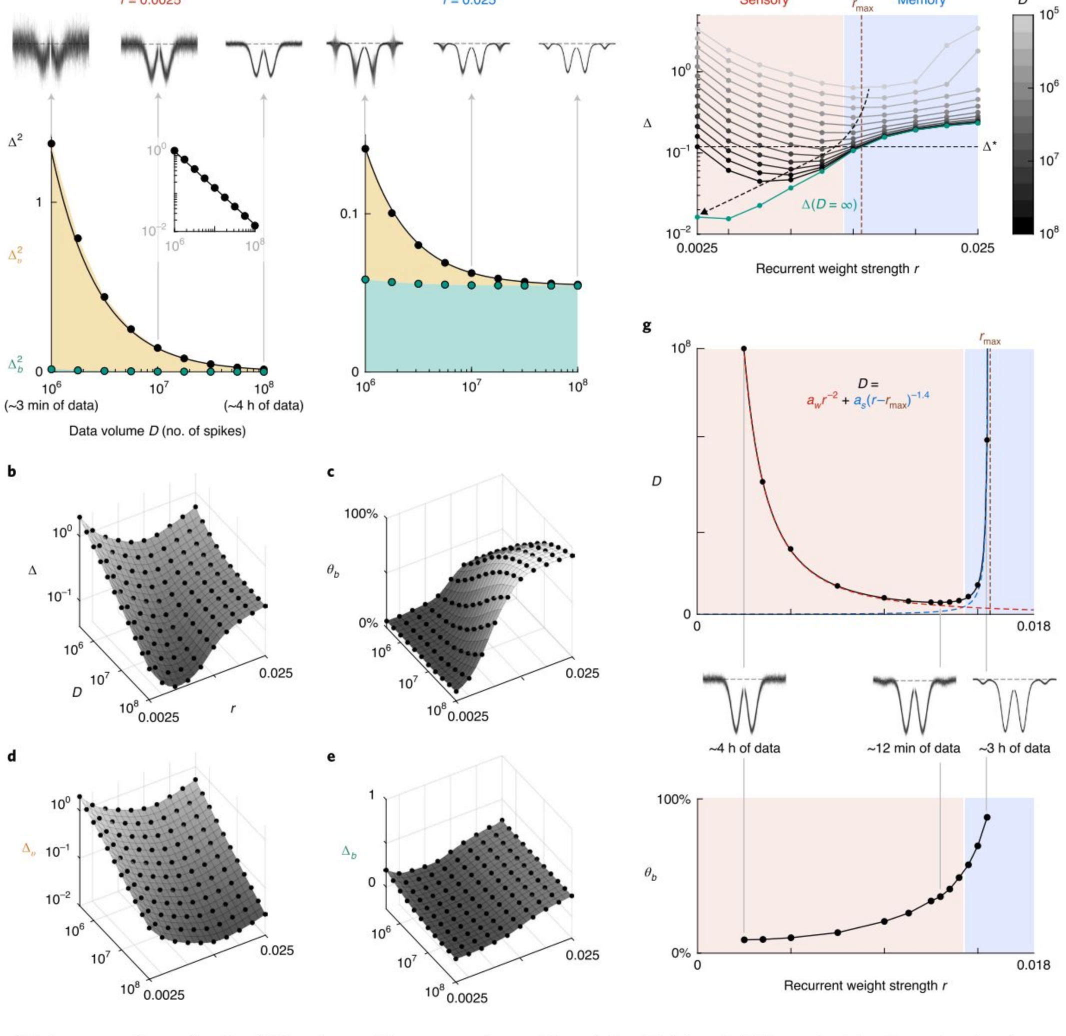


Fig. 3 | Inference quality as a function of data volume. a, Top: superposed rows of the weight matrix inferred with increasing data volume, at weak and

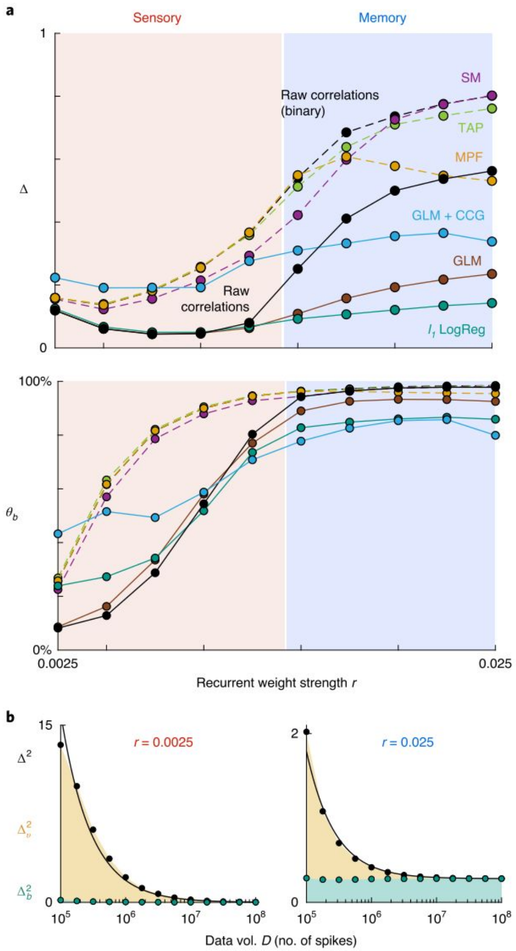


Fig. 4 | Results extend to different inference methods. a, Inference error

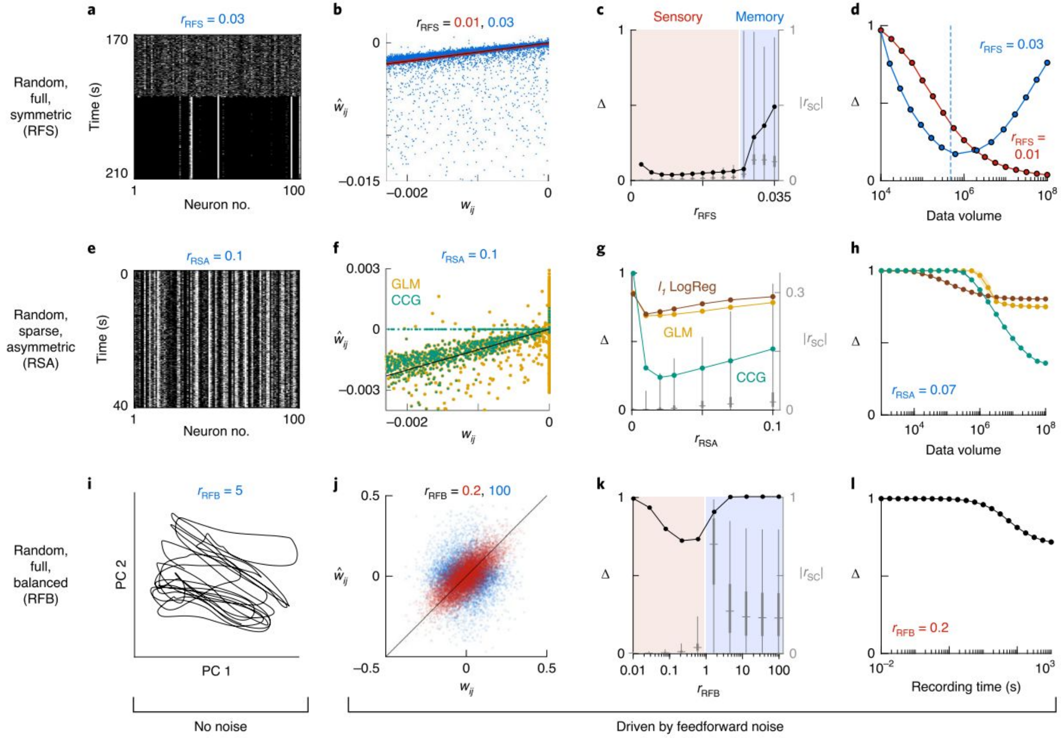


Fig. 5 | Results extend to different networks. a, Spike raster from a random, fully connected, symmetric recurrent network with strong weights, initialized

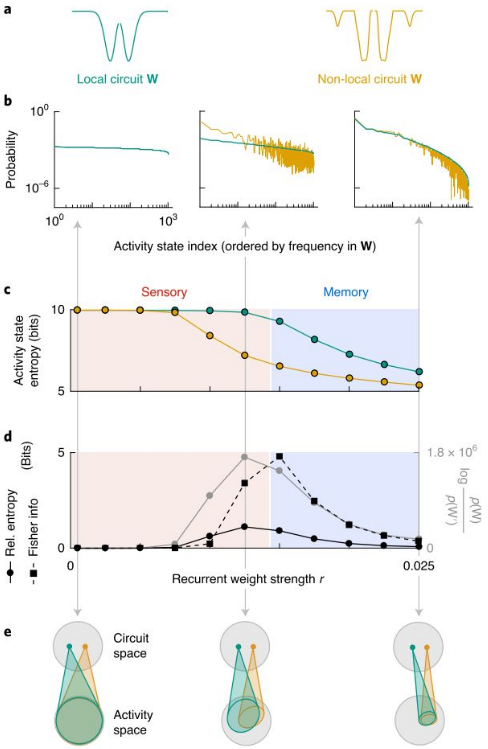


Fig. 6 | Circuit-to-activity map is inherently less invertible when correlations are strong. a, True weight profile for the ring circuit (with only local weights) and an alternative circuit (with a different weight profile and extra non-local weights). **b**, Probabilities of activity states (binary

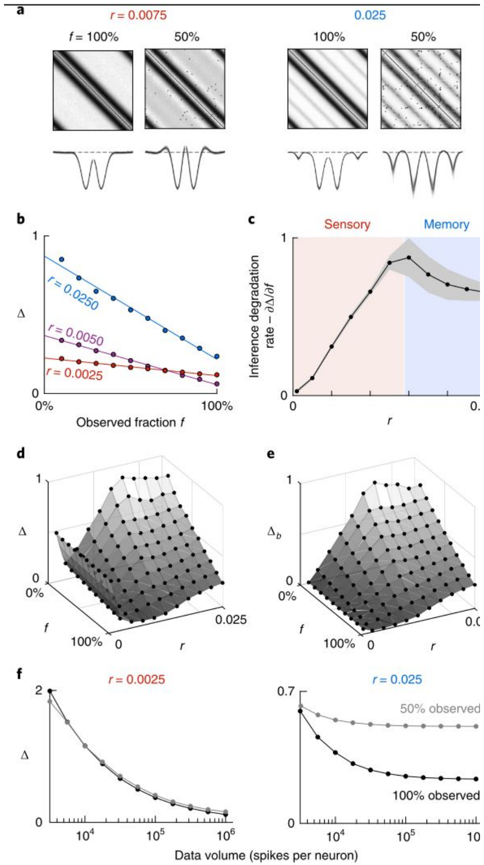


Fig. 7 | Inference bias due to unobserved neurons is exacerbated at strong weights. a, Weight matrices (top) and superposed rows (bottom)

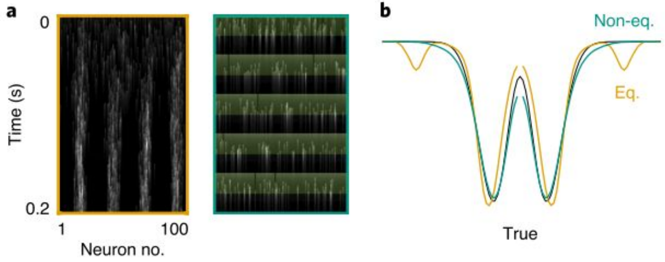


Fig. 8 | Sampling non-equilibrium data mitigates inference bias. a, Left: slow synaptic activations, *s*, of the ring network neurons at strong weight ($r = 0.025$). Right: the same when the feed-forward drive is pulsed on and off (green overlays mark 'on' periods). **b**, Average inferred weight profiles using data collected at equilibrium versus out of equilibrium.

I really like this figure.

I think it encompasses what we're trying to do.

I think we're just pushing the **observational level** to include closed-loop, and then taking another look at the **inferential level.**

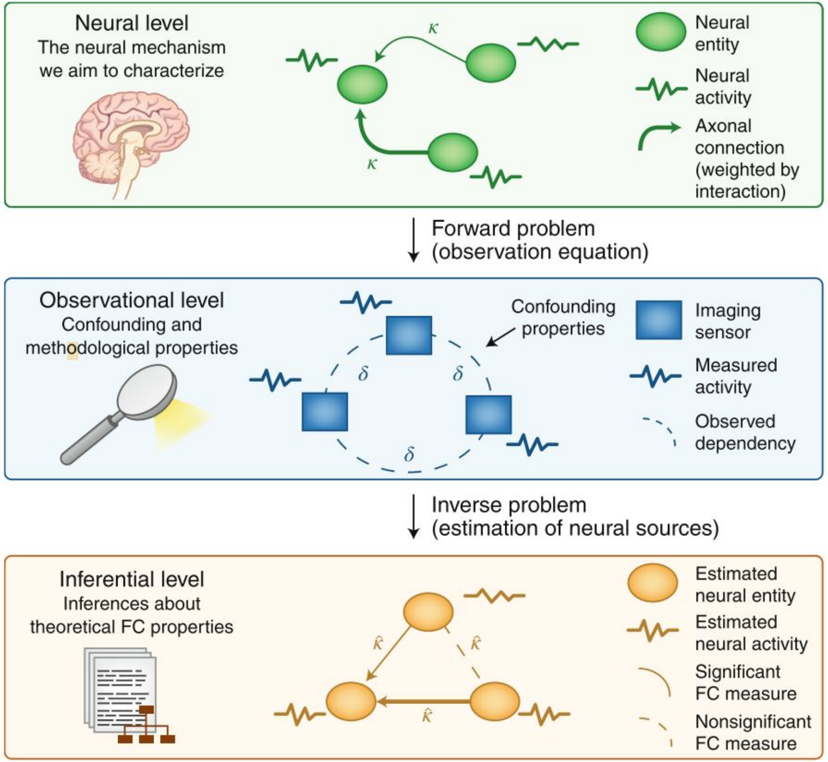


Fig. 1 | Ontological levels relevant to mechanistic interpretation of FC, defining the pathway from neural mechanisms (neural level) to imaging measurements (observational level) to inferences about target theoretical properties (inferential level). At the neural level, physical

Advancing functional connectivity research from association to causation

Examples of conceptual diagrams

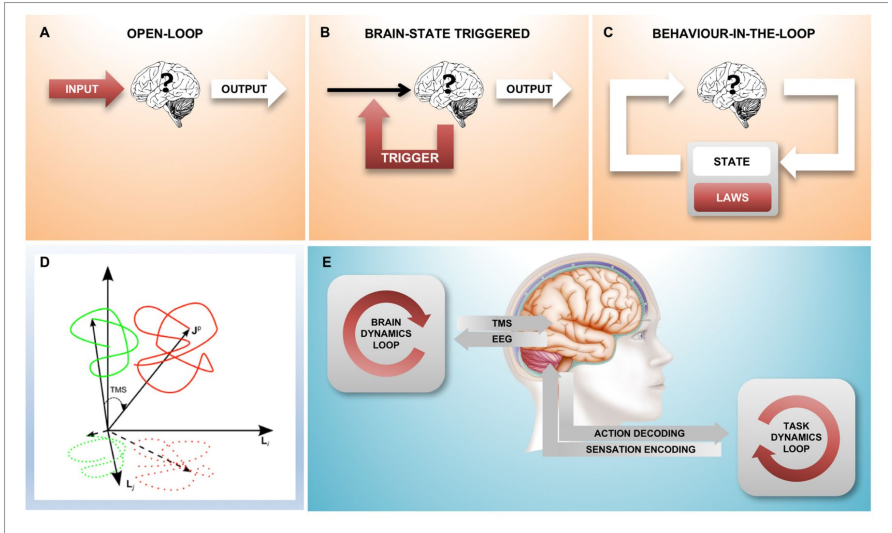


FIGURE 1 | A family of different closed-loop designs to couple with brain state dynamics. (A) Traditional "black box" experiment where there is no

Closed-Loop Neuroscience and Non-Invasive Brain Stimulation: A Tale of Two Loops

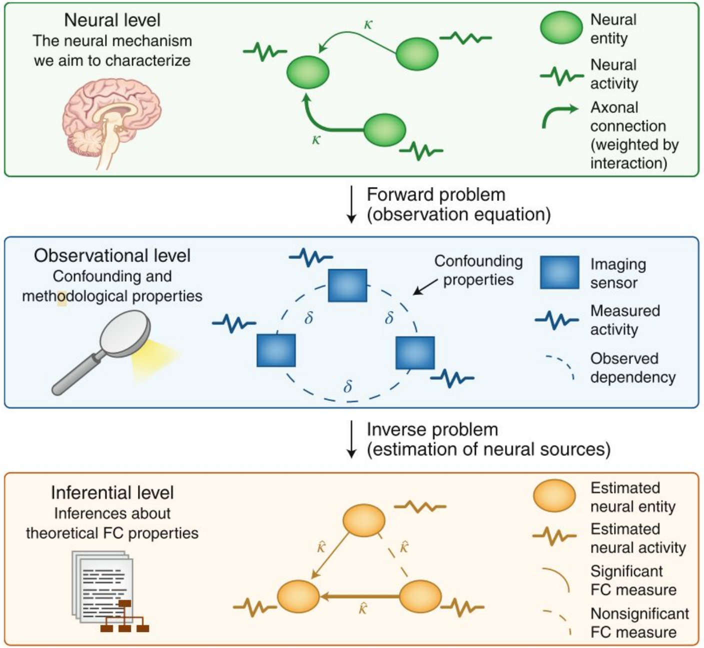
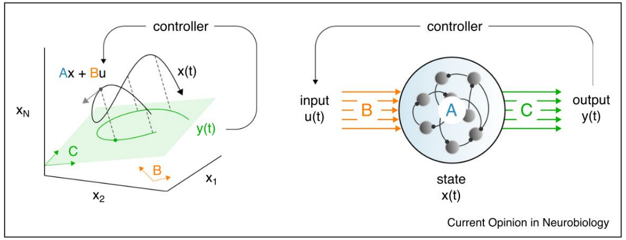


Fig. 1 | Ontological levels relevant to mechanistic interpretation of FC, defining the pathway from neural mechanisms (neural level) to imaging measurements (observational level) to inferences about target theoretical properties (inferential level). At the neural level, physical

Advancing functional connectivity research from association to causation

Examples of conceptual diagrams

Figure 1



State space & control perspective on neural dynamics. The population activity vector **x**(*t*) traces out trajectories in state space (solid black, left diagram), following a flow (gray arrow) determined by the state matrix **A** as well as external inputs **u**(*t*) (right diagram; cf. Equation 1). In a standard feedback control scenario, inputs are computed based on some measurements **y**(*t*) = **Cx**(*t*) (green) of the state vector, modifying the flow of activity along a few select 'input channels' **B** (orange).

Brain–machine interfaces from motor to mood

Neuroscience out of control: control-theoretic perspectives on neural circuit dynamics

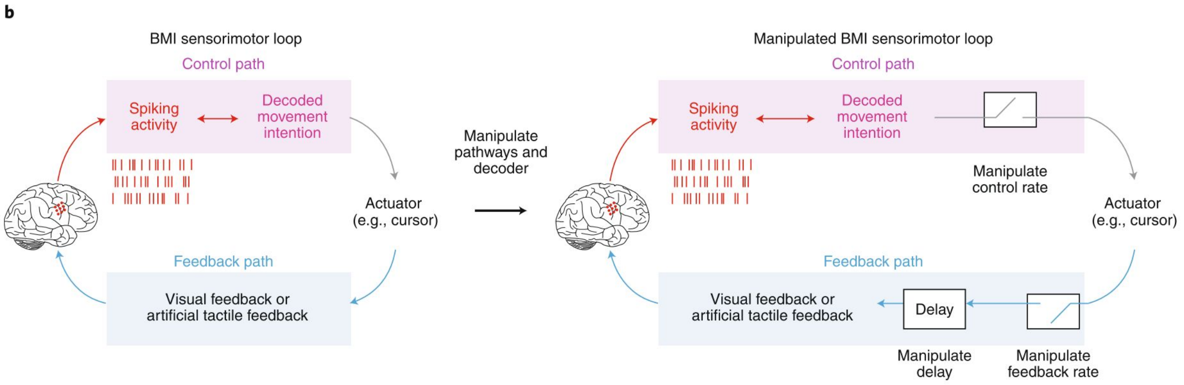


Fig. 2 | Motor BMIs for functional restoration and scientific discovery. a, The closed-loop control view of motor BMIs has substantially advanced their

Diagrams for correlation tests

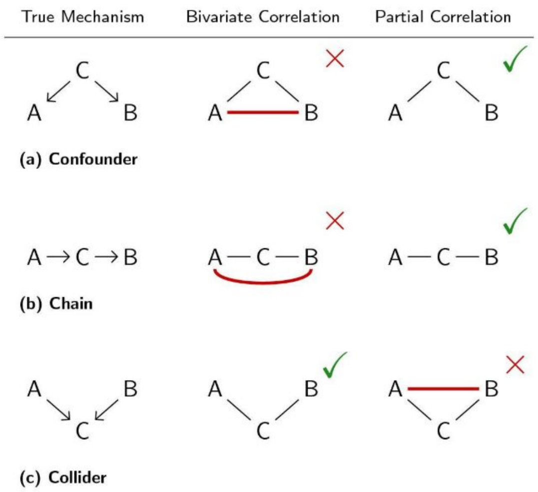
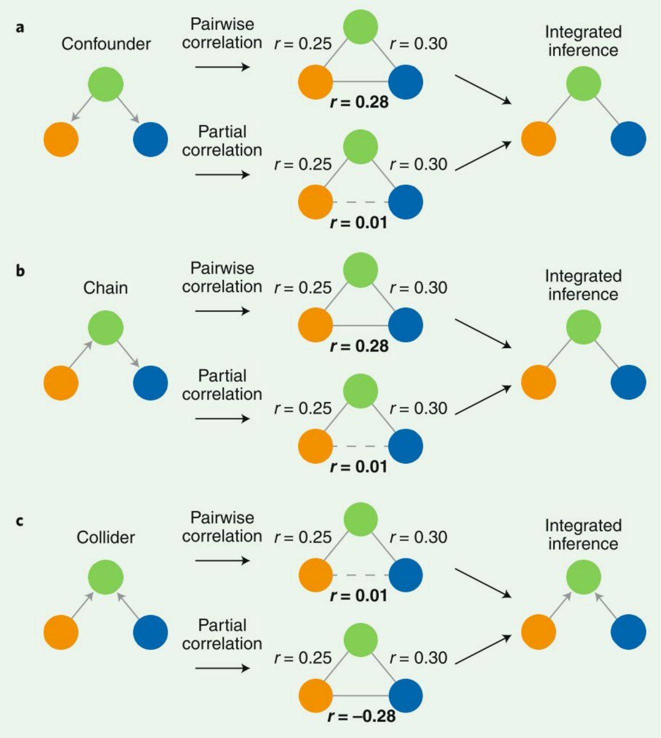


Figure 1. The pattern of spurious causal inferences for bivariate and partial correlations. Switching from correlation to partial correlation improves causal inference (but is not perfect). We propose integrating inferences from both correlation and partial correlation, which we predict will produce further improvements to causal inferences. Red lines indicate spurious causal inferences. Note that, in the case of a collider, when A → C and B → C are positive then the spurious A – B connection induced by partial correlation will be negative (this becomes relevant in the Results).

Combining multiple functional connectivity methods to improve causal inferences



Advancing functional connectivity research from association to causation

“Systematic errors in connectivity”
 - Fiete, Das

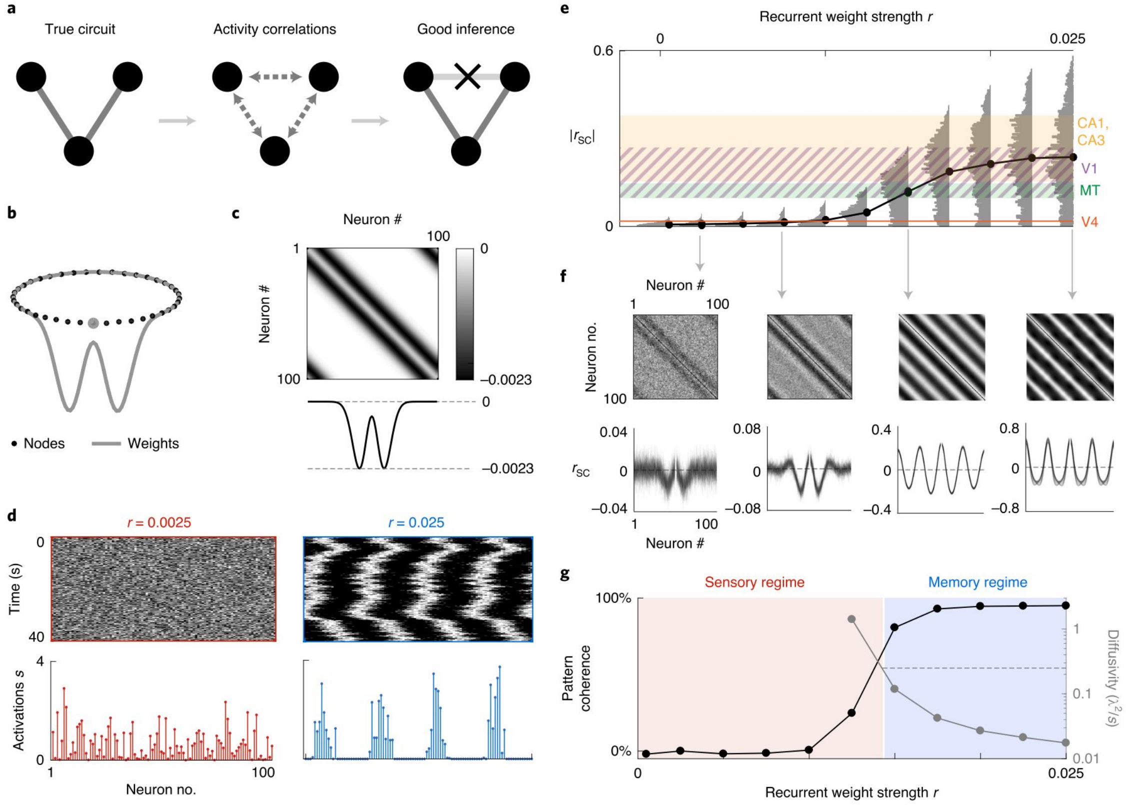


Fig. 1 | Structure and dynamics of the generative network. a, Left: schematic three-neuron circuit with two connections. Center: all neurons are correlated.

Examples of “methods overview” diagrams

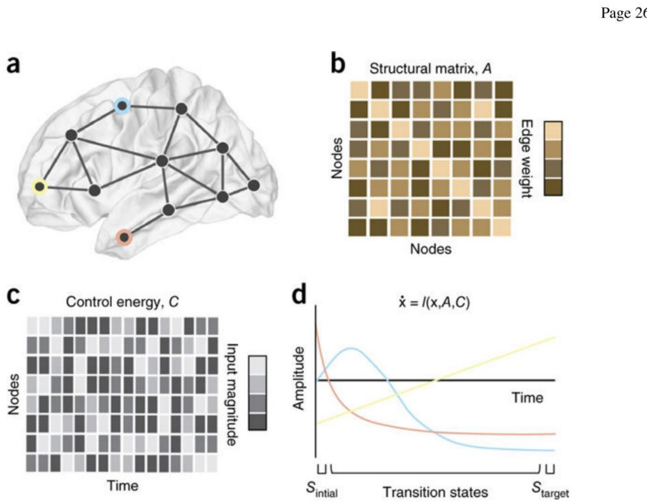
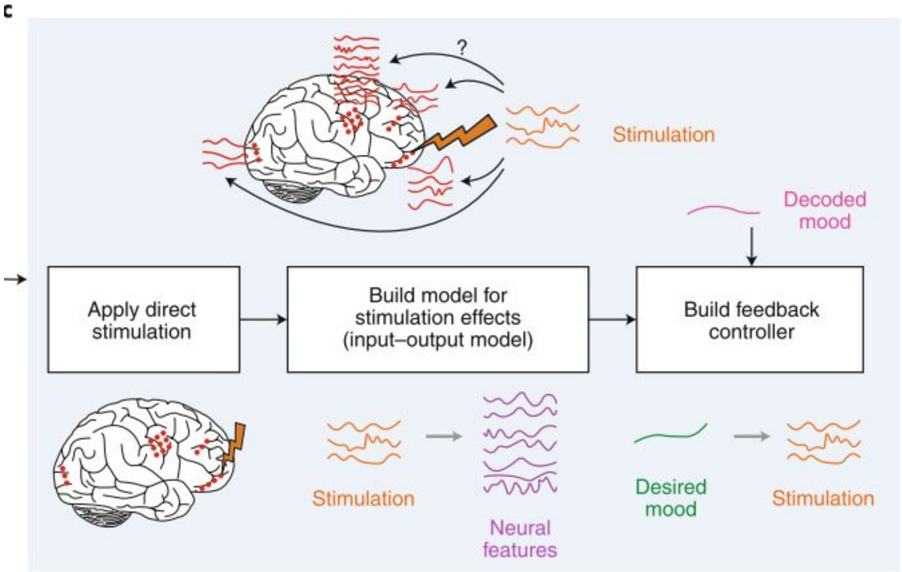


Figure 5.
Controlling brain networks. Following a careful description of the network properties of the brain, we might wish to intervene: to push a diseased brain toward health or to enhance the

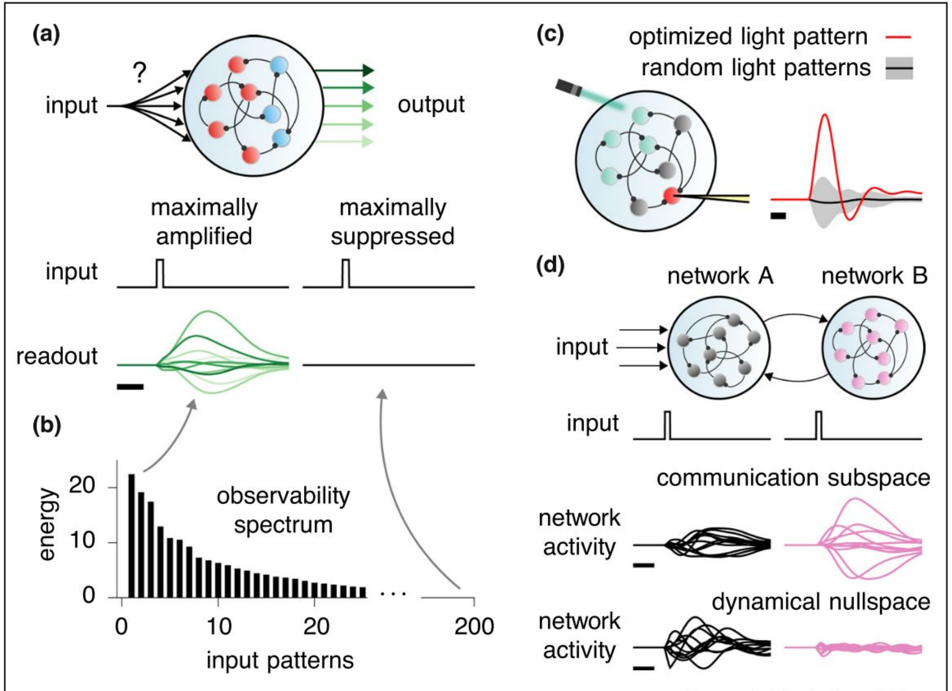
Network neuroscience



Brain–machine interfaces from motor to mood

Examples of “methods overview” diagrams

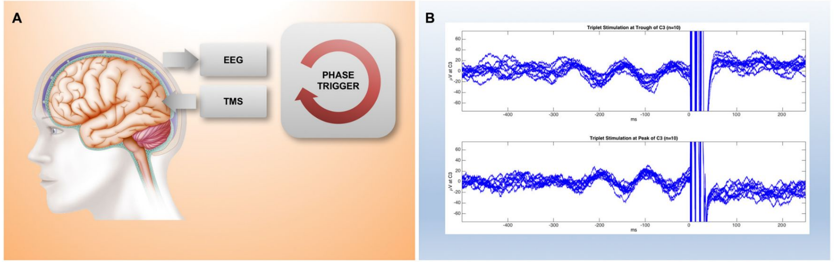
Figure 2

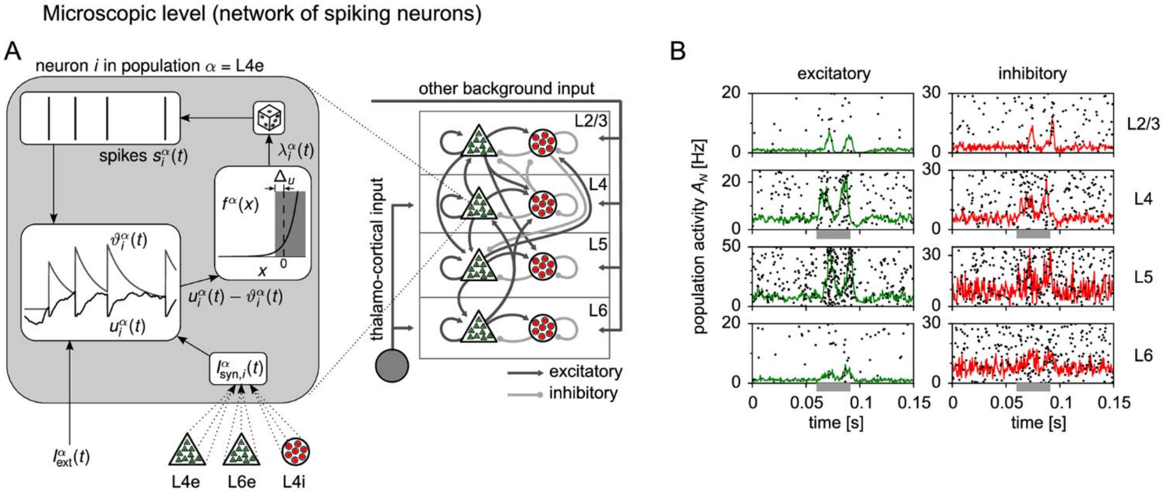


Current Opinion in Neurobiology

Neuroscience out of control: control-theoretic perspectives on neural circuit dynamics

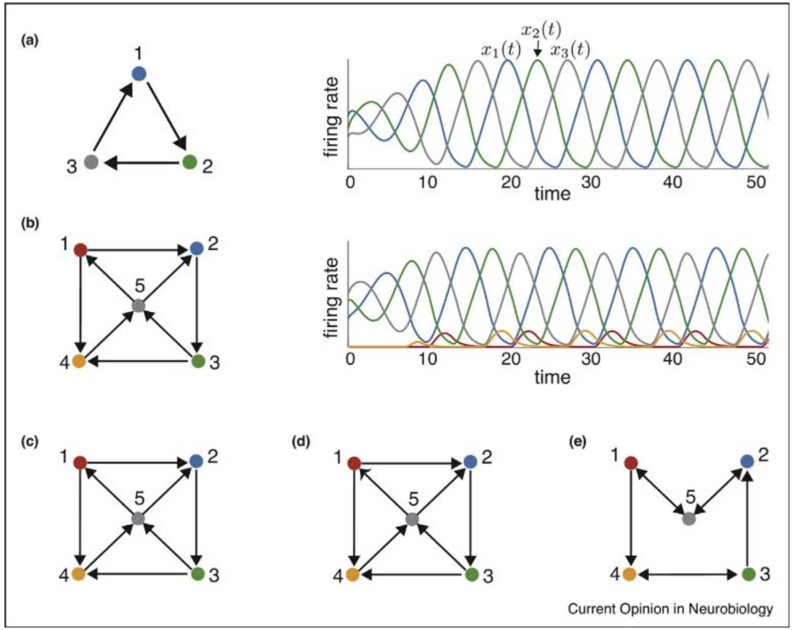
Examples of “model-data pair” diagrams





Towards a theory of cortical columns: From spiking neurons to interacting neural populations of finite size

Examples of “model-data pair” diagrams



Relating network connectivity to dynamics: opportunities and challenges for theoretical neuroscience

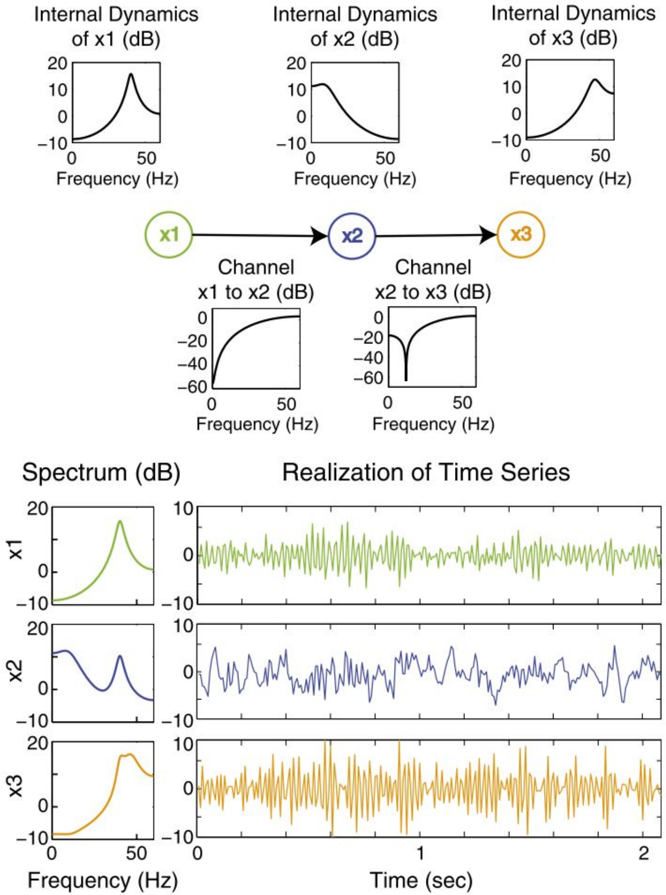
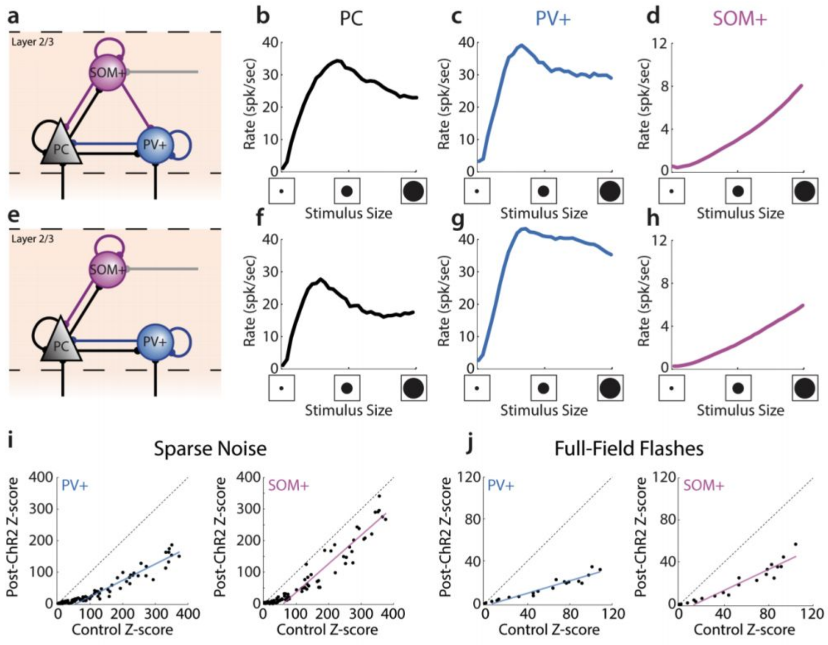


Fig. 1. VAR(3) three-node series system of example 1. *Top* shows the net-

Examples of results summaries



Supplementary Figure 12

Examples of results summaries

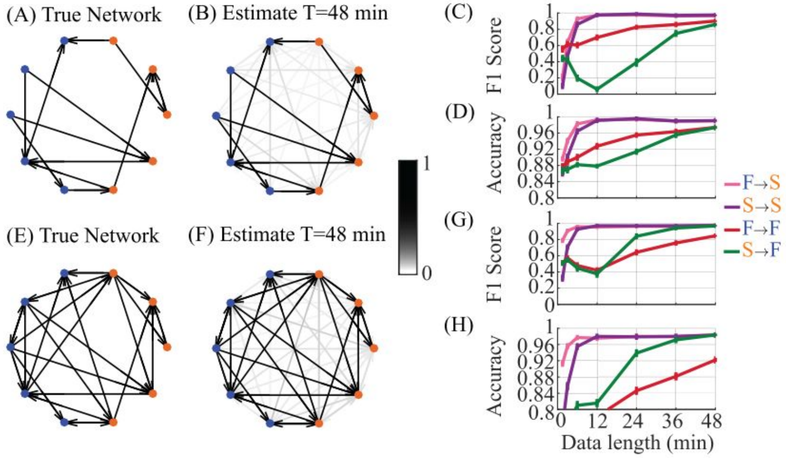


Fig. 5. Multiscale causality estimation algorithm accurately estimates the causality graph of two random 10-node networks. (A,E) True network graphs. The blue nodes indicate field nodes and the orange nodes indicate spike nodes. (B,F) Estimated network graphs. The color of arrows indicates the detection rate (between 0 to 1 as shown by the colorbar) of causality between each pair calculated across 100 simulated trials, with a darker color indicating a larger rate of detection. T indicates the data length used for estimation. (C,D,G,H) F1-score and accuracy for the 4 possible connection types using our method.

Estimating Multiscale Direct Causality Graphs in Neural Spike-Field Networks

Examples of “enumerating the field”

Bassett and Sporns Page 20

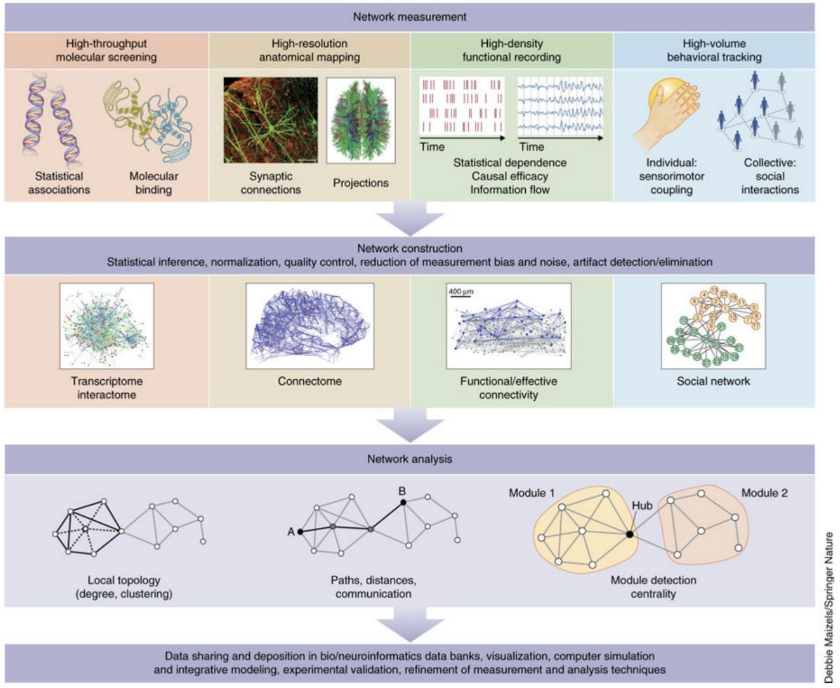


Figure 2.
Network measurement, construction and analysis. Top, network neuroscience begins with

Network Neuroscience

Figure 2

Graph Structures

Degree distribution
- Erdős-Rényi random (Poisson distribution)
- Scale-free (power law distribution)
- Exponential
- Truncated log normal
- Sparsity

Geometric/regular organization
- Ring or lattice connectivity
- Probability of connection drops off with distance
- Watts-Strogatz small world: locally geometric with random long range connections

Characteristic path length, Clustering coefficient
Centrality measures

Hubs, Rich Clubs, Modularity
Clusters & Communities (structural vs. functional)

Spectral properties of
- Adjacency matrix
- Weighted connectivity matrix
- Graph Laplacian

Connectivity statistics
- Probability of unidirectional vs. bidirectional connections
- Probability of connection between cell types
- Distribution of synaptic weights

Motif statistics
- Overrepresentation with respect to null model
- Motifs within/across cell types

Dynamic Models

Linear/Network flow models
- Random walk
- Percolation
- Diffusion
- Disease transmission
- Markov chain
- Information flow -- broadcast vs. routing
- Transport network
- Electric network flow

Nonlinear/Neuro-inspired models
- Spiking models, e.g., LIF, Hodgkin-Huxley
- Morris-Lecar, Fitzhugh-Nagumo, Izhikevich
- Coupled oscillators
- Firing rate models, e.g. Threshold-linear
- Discrete networks, e.g. Hopfield, SER models
- Linear conductance-based models

Dynamic Properties

Global population activity
- Average/mean field activity
- Synchronization vs. asynchronous firing
- Spontaneous activity
- UP/DOWN states
- State transitions
- Criticality/edge of chaos
- Oscillations and rhythmicity
- Traveling waves

Specific patterns of neural activity
- Attractor structure -- number and type
- Cell assembly structure
- Persistent activity
- Sequence generation

Spiking properties
- Correlation structure/variance of spike trains
- Bursting
- Neural avanlanches, cascades

Current Opinion in Neurobiology

Relating network connectivity to dynamics: opportunities and challenges for theoretical neuroscience