Title: ...

Authors: Adam Willats, Matt O'Shaughnessy

Table of Contents

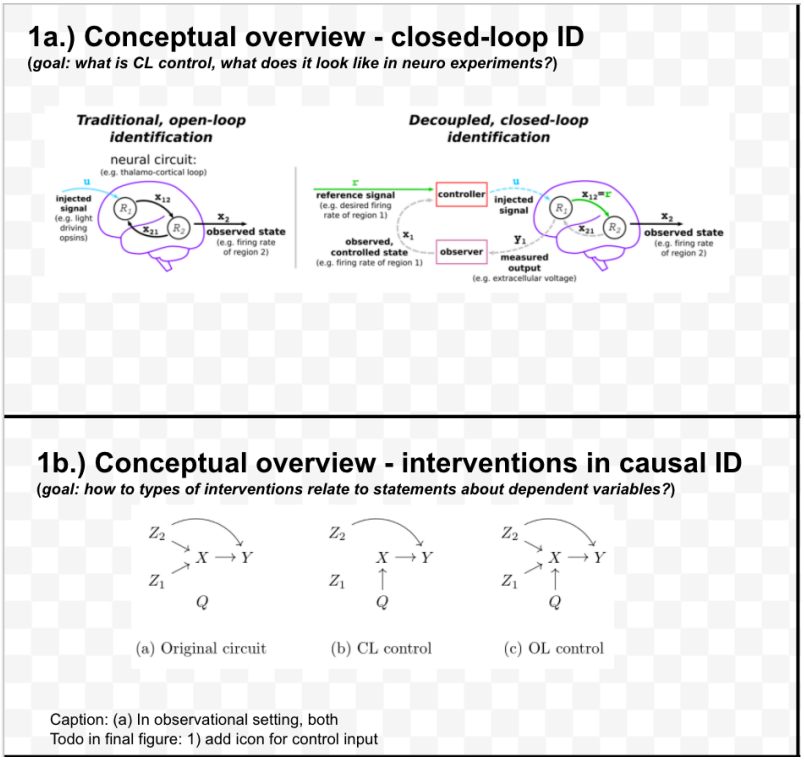
Abstract

The necessity of intervention in inferring cause has long been understood in neuroscience. Recent work has highlighted the limitations of passive observation and single-site lesion studies in accurately recovering causal circuit structure. The advent of optogenetics has facilitated increasingly precise forms of intervention including closed-loop control which may help eliminate confounding influences. However, it is not yet clear how best to apply closed-loop control to leverage this increased inferential power. In this paper, we use tools from causal inference, control theory, and neuroscience to show when and how closed-loop interventions can more effectively reveal causal relationships. We also examine the performance

of standard network inference procedures in simulated spiking networks under passive, open-loop and closed-loop conditions. We demonstrate a unique capacity of feedback control to distinguish competing circuit hypotheses by disrupting connections which would otherwise result in equivalent patterns of correlation. Our results build toward a practical framework to improve design of neuroscience experiments to answer causal questions about neural circuits.

Introduction

Estimating causal interactions in the brain



@ import "section_content/background_causal_network_id.md"

Interventions in neuroscience & causal inference

@ import "section_content/background_intervention_neuro.md"
@ import "section_content/background_intervention_causal_inf.md"

Representations & reachability

Representations

- The circuit view

        - (A) → (B) ↔ (C)
- The dynamical system view

$$\begin{cases} x' = Ax + Bu \\ y = Cx + \eta \end{cases}$$

- The connectivity (adjacency matrix) view

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

- why consider multiple perspectives

Reachability

- concept of **binary reachability** as a "best case scenario" for identification.
    - binary reachability describes which pairs of nodes we expect to have any correlation
    - can be used to predict "equivalence classes", i.e. circuits which may be indistinguishable under certain interventions
    - how binary reachability is computed
        - [...equations here...]
- **graded reachability** can help predict the influence of parameter values (e.g. edge weights, time-constants) on identifiability
    - quantifies impact of inputs, noise on outputs
    - easiest to describe/understand in linear-gaussian setting
    - [...equations here...]

Understanding identification through derived properties of circuits (reachability rules)

- connect **binary reachability** to classes of ambiguity
    - a pair of networks are ambiguous (given some intervention) if they are in the same markov equivalence class
    - ambiguity x intervention leads to the following classes
        - passively unambiguous
        - open-loop unambiguous
        - (single-site) closed-loop unambiguous

`(text from proposal)`

To perform this evaluation, I used a simple derived quantity of the circuit's adjacency matrix: its reachability (Skiena 2011). This quantity measures which nodes can be "reached" by direct and indirect connections starting from a target node. Moreover, this measure allows us to predict which node's signals will be correlated. If the reachability of two circuits are equal, they will have similar correlational structure and be difficult to distinguish with that level of intervention. For instance looking at hypotheses for cortical gain control in open-loop (Figure 12, left column), in both circuit 2a and 2b, PV cells are reachable from the Som cell node since Som activity can influence PV activity indirectly through the Pyr node. As such these

circuits cannot be distinguished in open loop.

If the reachability of two circuits are unequal for a given intervention, differences in correlation between observed regions will be sufficient to distinguish between the two hypotheses. Looking at these same circuits under closed-loop control of the pyramidal population (Figure 12, right column), dashed lines reveal that there is no longer an indirect functional connection from Som to PV cells. As such, in circuit 2a, PV cells are no longer reachable from the Som population, whereas they are reachable under circuit 2b. This difference in reachability corresponds to the difference in correlational structure that allows us to distinguish these two hypotheses under closed-loop control.

(see also `sketches_and_notation/notation0_reachability.md` , `walkthrough_EI_dissection.md`)

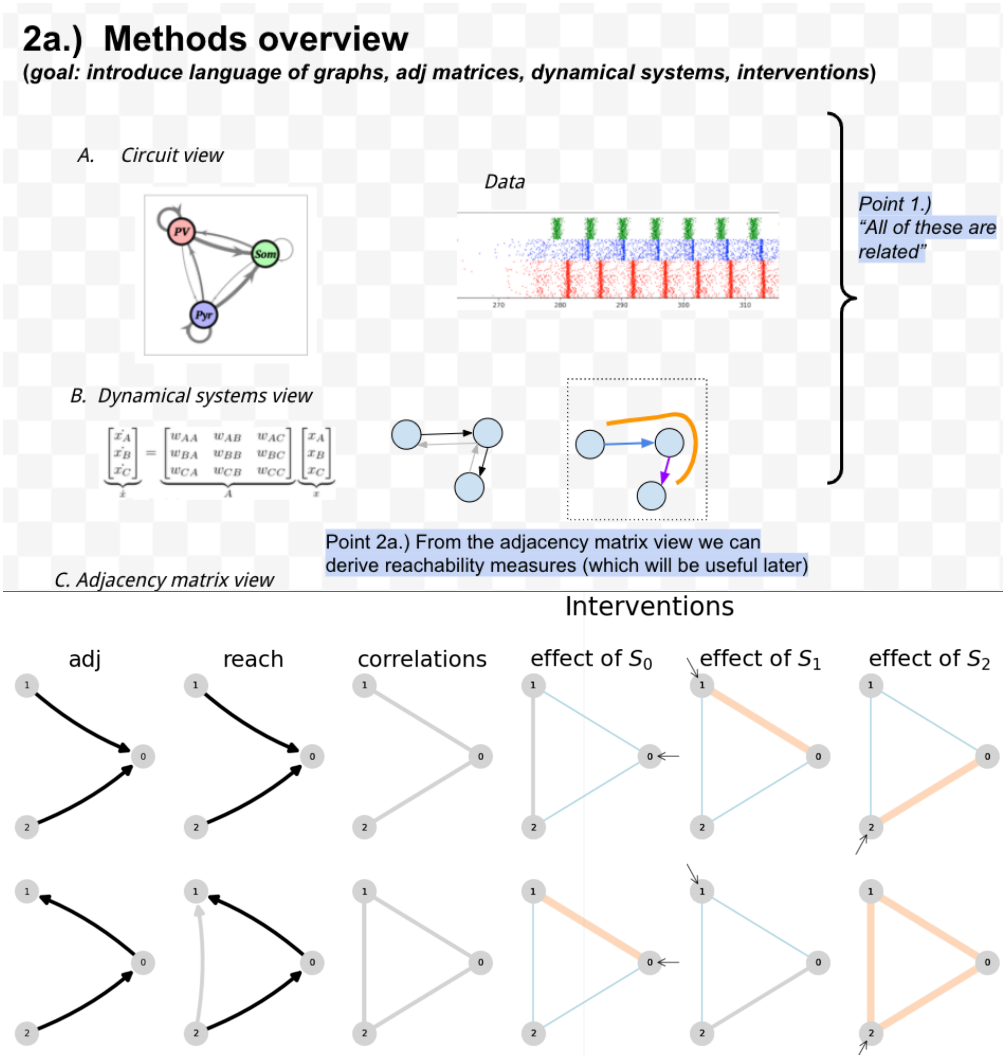
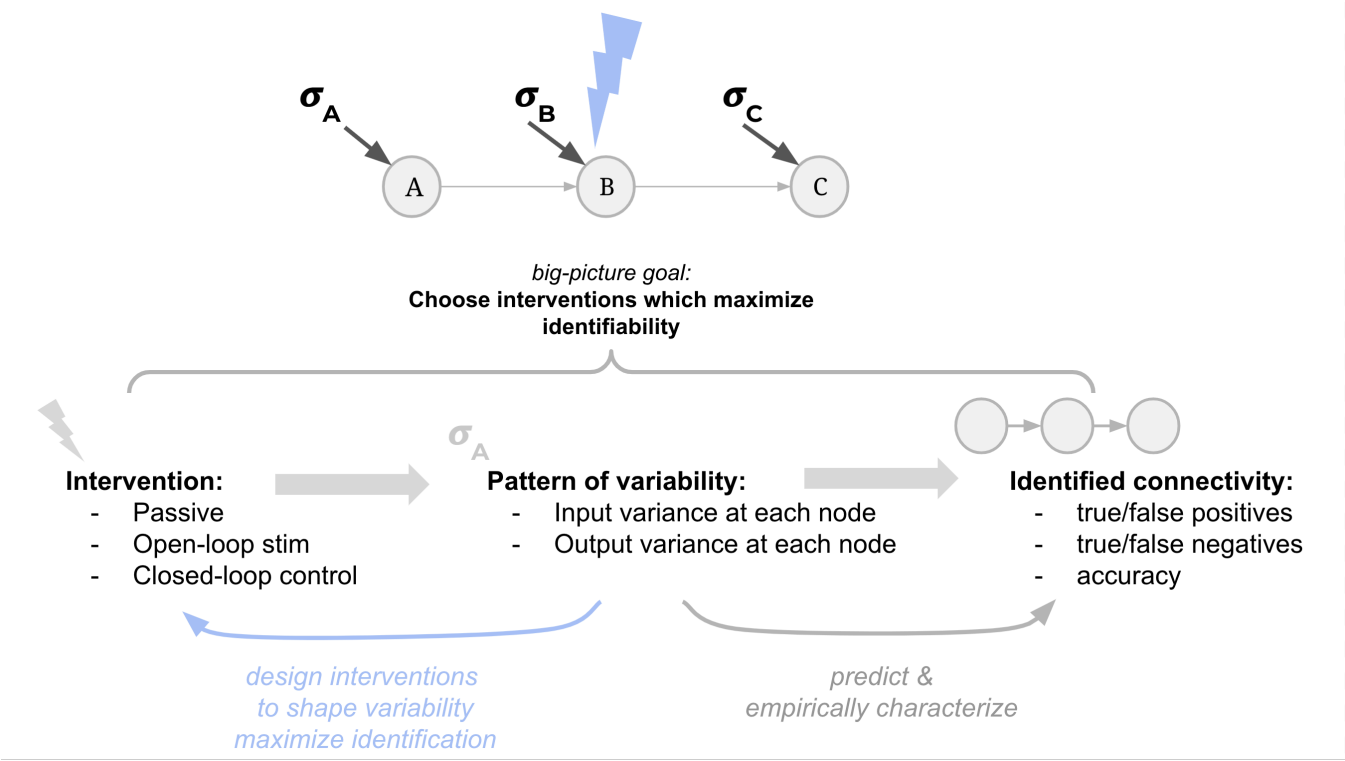


Figure DEMO: Applying CLINC to distinguish a pair of circuits

Theory / Prediction

Computing reachability (theory)

Predicting correlation structure (theory)



Simulation

Network simulations (simulation)

Implementing interventions (simulation)

Extracting circuit estimates (empirical)

Information-theoretic measures of hypothesis ambiguity

Results

Interaction of intervention on circuit estimation

going to assume these have already been discussed

- predicting correlation
- measuring dependence
- markov equivalence

Intervening provides (categorical) improvements in inference power beyond passive observation

> **Theme B.** Experiments for circuit inference can be thought of as **narrowing the set of plausible explanations**, refining a hypotheses space

🚧 figure request: flowchart for steps of intervention experiment 🚧

We envision the structure of an experiment[1] to include the following broad stages:

1. First, explicitly **enumerate the set of hypothesized circuits.** Hypotheses about the structure of the circuit would be based on multiple sources of information including prior recordings, anatomical constraints revealed by `experiments where you look at the fiber bundles connecting regions`, or commonly observed connectivity patterns in other systems `[🚧 add other sources of priors for circuit hypotheses]`[2][3] These hypotheses should be expressed as a set of circuits (adjacency matrices) each with a probability representing the prior belief about the relative likelihood of these options.
2. Second, *in silico*, **forecast patterns of correlation** which could result from applying candidate interventions. A detailed forecast of the observed outputs could be achieved by simulating detailed biophysical networks across both candidate interventions, and hypothesized ground-truth circuits. However, for large networks or large hypothesis sets this may be infeasible. Instead, using the reachability representation of a linear (linearized) network we can succinctly and efficiently predict the observed correlations[4] across nodes[5]. The methods described in `[ref. prediction methods]` allow us to anticipate how open and closed-loop interventions across nodes in the network might increase, decrease, or sever dependencies between node outputs.

3. `{Survey / analyze / compare / summarize}` `{diversity / equivalence / distinguishability of}` patterns of correlation across each hypothesized circuit.

Our goal is to reduce the set of [...]

- if there are many distinct patterns of correlation for a given intervention depending on the ground-truth circuit structure then an experimenter is likely to learn a lot
    - then we can consider these hypotheses to be "distinguishable" under that intervention
    - this situation would allow

However, since the ground-truth circuit is not known *a priori*, it is useful to measure the expected information gained across the set of possible circuits

- uniqueness / diversity of observed patterns can be summarized across a hypothesis set with (Shannon) entropy, a scalar which quantifies the expected information gained `[...]`

4. *(in expt. or detailed biophysical simulation)* Apply intervention, collect data**
5. **Given the observed dependency pattern, form a posterior belief over hypotheses**

- remaining entropy quantifies the "realized" information (which may be larger or smaller than the expectation), and equivalently the remaining size and uncertainty of the posterior belief over the hypothesis set
- Can choose the most likely (MAP) circuit amongst this posterior hypotheses
    - *(optionally, this posterior distribution can be used as an updated prior for the next iteration)*

Next, we apply (steps 1-3 of) this circuit search procedure to a collection of closely related hypotheses for 3 interacting nodes[5:1] to illustrate the impact of intervention.

Figure DISAMBIG: Stronger intervention facilitates disambiguating equivalent hypotheses

Why does closed-loop control provide a categorical advantage? Because it severs indirect links

- this is especially relevant in recurrently connected networks where the reachability matrix becomes more dense.
- more stuff is connected to other stuff, so there are more indirect connections, and the resulting correlations look more similar (more circuits in the equivalence class)

Where you intervene strongly determines the inference power of your experiment.

secondary: having (binary) prediction helps capture this relationship

shows a dataset with many correlations, multiple plausible circuit hypotheses

- patterns of correlation become more specific with increasing intervention strength
- in aggregate: focuses on reduced bias, higher accuracy for "infinite" data limit
- closed-loop > open-loop > passive

Stronger intervention results in more efficient, accuracy inference

> - quantitative
> - "here's additional nuance"
> -

Figure VAR: Stronger intervention allows better control of covariance

shaping covariance is the focus of this paper.

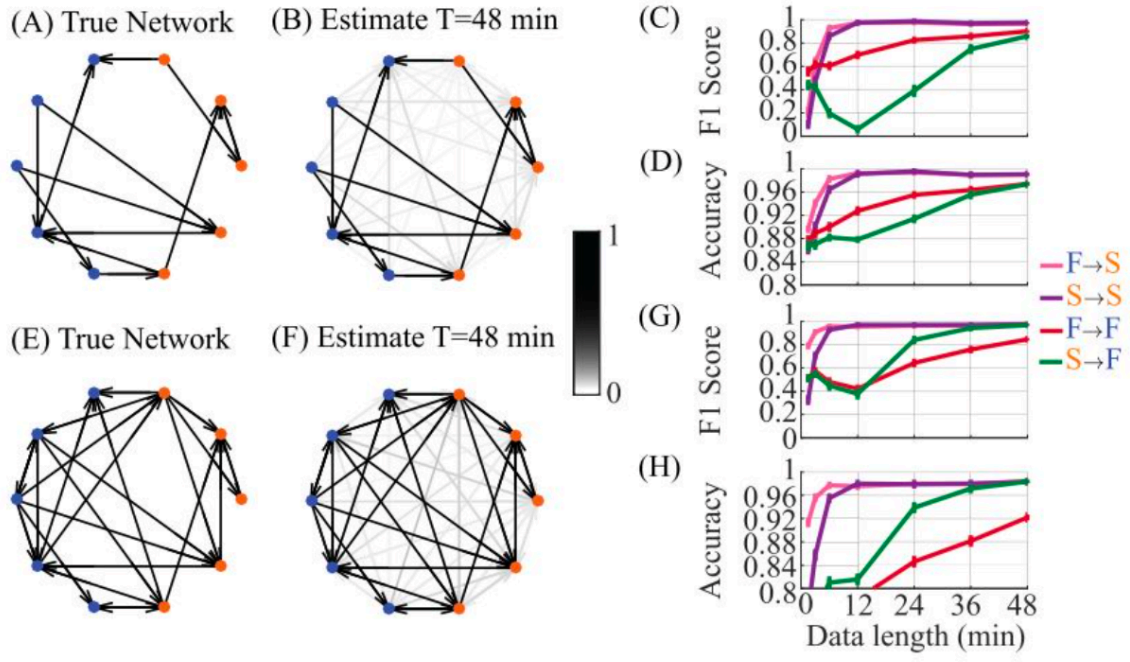
- this is a key advantage of closed-loop control
    - which can have bidirectional influence over variance

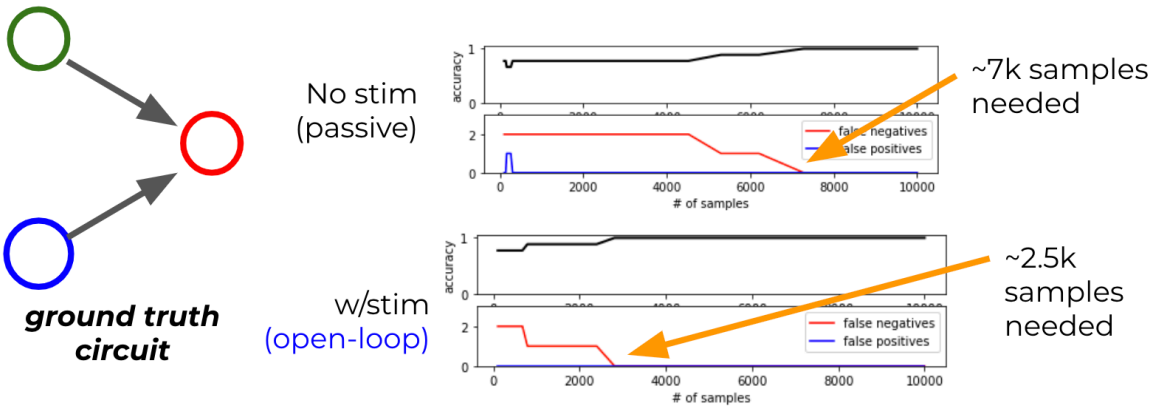
> while you can deliver open-loop inputs with titrated amounts of variance, you're often only able to add variance rather than subtract it, and the amount of variance you would add to the system is hard to predict a priori

- having (quantitative) prediction helps capture this relationship
    - **Figure PREDICT: Comparing predicted and empirical correlation, identification performance**

Related sections

- Quantitative advantages -- higher accuracy & data efficiency
  Figure DATA: Analysis of simulated circuits suggest stronger intervention facilitates identification with less data





- **impact of circuit structure**
    - Figure MOTIF: Interaction of network structure and intervention location on identifiability

Interaction of intervention & circuit structure

Figure MOTIF: Interaction of network structure and intervention location on identifiability

---

Discussion

References

Supplement

---

1. more than just an experiment, this is a "hypothesis search." Is this procedure what we're going to brand as the "CLINC" process? ↩
2. **[future work]** use causality + graph theory to find "lurking look-alikes" i.e. Markov-equivalent circuits ↩
3. should also enumerate assumptions about the dynamics of the network, signs of network weights, approximate timescales of interaction. ↩
4. using binary reachability, we can be more general above predicting the "direction" of other measures of bivariate dependence like transfer entropy ↩
5. nodes in such a graphical model may represent populations of neurons, distinct cell-types, different regions within the brain, or components of a latent variable represented in the brain. ↩ ↩