PHY562 Final Presentation

Mihir Rao

mihirrao@princeton.edu

PHYSICAL REVIEW E **106**, 034102 (2022)

Editors' Suggestion

Emergence of local irreversibility in complex interacting systems

Christopher W. Lynn,[1,2] Caroline M. Holmes,[2] William Bialek,[1,2] and David J. Schwab[1]

[1] *Initiative for the Theoretical Sciences, Graduate Center, City University of New York, New York, New York 10016, USA*

[2] *Joseph Henry Laboratories of Physics and Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544, USA*

(Received 8 March 2022; accepted 24 June 2022; published 6 September 2022)

Living systems are fundamentally irreversible, breaking detailed balance and establishing an arrow of time. But how does the evident arrow of time for a whole system arise from the interactions among its multiple elements? We show that the local evidence for the arrow of time, which is the entropy production for thermodynamic systems, can be decomposed. First, it can be split into two components: an independent term reflecting the dynamics of individual elements and an interaction term driven by the dependencies among elements. Adapting tools from nonequilibrium physics, we further decompose the interaction term into contributions from pairs of elements, triplets, and higher-order terms. We illustrate our methods on models of cellular sensing and logical computations, as well as on patterns of neural activity in the retina as it responds to visual inputs. We find that neural activity can define the arrow of time even when the visual inputs do not, and that the dominant contribution to this breaking of detailed balance comes from interactions among pairs of neurons.

DOI: 10.1103/PhysRevE.106.034102

Princeton Physics

Outline

I. Time in living systems
II. Local irreversibility and multipartite dynamics
III. Decomposing irreversibility (independent + interacting)
IV. Irreversibility due to kth-order dynamics (k>2)
V. Decomposing irreversibility in
    A. Sensing systems
    B. Logical functions
    C. Neuronal populations

Outline

I. Time in living systems
II. Local irreversibility and multipartite dynamics
III. Decomposing irreversibility (independent + interacting)
IV. Irreversibility due to kth-order dynamics (k>2)
V. Decomposing irreversibility in
    A. Sensing systems
    B. Logical functions
    C. Neuronal populations

I. Time in living systems

Living systems consume energy in order to maintain order and function. Being away from equilibrium, we expect that their microscopic dynamics violate detailed balance. Macroscopically, their behaviors define an arrow of time. Despite recent progress in nonequilibrium statistical physics [1–3], there remain basic questions about how irreversibility at one scale emerges from collective dynamics at the scale below. To what extent does the irreversibility of a system arise from interactions between elements, rather than the independent dynamics of the elements themselves? Can simple dynamics involving pairs or triplets of elements build upon one another to generate large-scale irreversibility, thereby defining a macroscopic arrow of time, or do complex biological systems depend on higher-order combinatorial interactions?

Adapted from Introduction.

Metabolism
(ex: ATP consumption)

Development, Aging, Disease
(things we care about!)

Macroscopic emergence of AoT
(how does this emergence happen, mechanistically?)

Framing: **collective** dynamics = **individual** dynamics + **interactions**

Considering k-way combinatorial interactions of elements (k=1,2,3,...,N), which k are responsible for the emergence of irreversibility (and AoT)?

k=1 individual dynamics, k=2 pairs, k=3 triplets, k=4 quartets, etc.

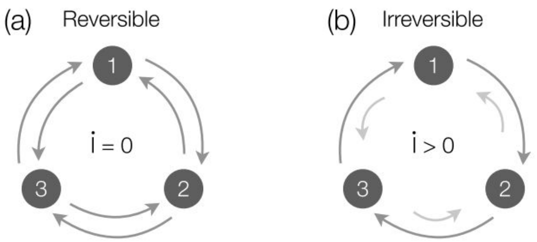
Outline

I. Time in living systems
II. Local irreversibility and multipartite dynamics
III. Decomposing irreversibility (independent + interacting)
IV. Irreversibility due to kth-order dynamics (k>2)
V. Decomposing irreversibility in
    A. Sensing systems
    B. Logical functions
    C. Neuronal populations

II. Local irreversibility and multipartite dynamics

Key idea #1:

Net fluxes ⇔ broken detailed balance ⇔ "irreversibility" (on avg.) ⇔ phenomenological AoT



Adapted from Fig. 1.

No net fluxes	Net fluxes
$P(x \to x') = P(x' \to x)$	$P(x \to x') \neq P(x' \to x)$

This is precisely the statement of detailed balance.

$$P(x \to x') \equiv \mathrm{Prob}[x_t = x,\ x_{t+1} = x']$$

Joint transition probability
not conditional; bakes in P(x)

Key idea #2:

Quantifying irreversibility with *Kullback-Leibler divergence*

$$\dot{I} = \sum_{x,x'} P(x \to x') \log \left[\frac{P(x \to x')}{P(x' \to x)} \right]$$

KL divergence between forward- and reverse-time transition probabilities quantifies the evidence of AoT that the joint transition probability carries.

For Markovian systems:
- Joint transition probabilities completely define dynamics.
- $\dot{I}$ captures all available information about AoT
- If *x* and *x'* include all microscopic DoFs in a system, then, under reasonable assumptions, $\dot{I}$ defines the physical rate of entropy production by the system.

Key idea #3: In general, and in practice, when we don't observe all relevant DoFs, the dynamics of observable states *x* are **non-Markovian**, but $\dot{I}$ still precisely represents the *local evidence* for the AoT.

II. Local irreversibility and multipartite dynamics

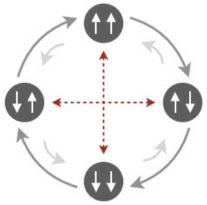
Local evidence of irreversibility

$$\dot{I} = \sum_{x,x'} P(x \to x') \log \left[\frac{P(x \to x')}{P(x' \to x)} \right]$$

In living systems, overall state x is composed of states { x_i } for many interacting composite elements, i=1,2,...,N.

Key idea #4: we consider the limit in which we have sufficient temporal resolution s.t. no two elements change state at exactly the same time.

(c) Multipartite



Adapted from Fig. 1.

In this limit, element dynamics are defined by the joint transition probability $P(x_i \to x_i', x_{-i})$ of
(1) element i transitioning from x_i to x_i' and
(2) the **rest of the elements not changing state** (denoted by x_{-i}).

Why is this useful? Multipartite assumption allows us to isolate the irreversibility associated with an individual element.

$$\begin{aligned}
\dot{I} &= \sum_{x,x'} P(x \to x') \log \left[\frac{P(x \to x')}{P(x' \to x)} \right] \\
&= \sum_{x} \sum_{i=1}^{N} \sum_{x_i'} P(x_i \to x_i', x_{-i}) \log \left[\frac{P(x_i \to x_i', x_{-i})}{P(x_i' \to x_i, x_{-i})} \right] \\
&= \sum_{i=1}^{N} \sum_{x_{-i}} \sum_{x_i, x_i'} P(x_i \to x_i', x_{-i}) \log \left[\frac{P(x_i \to x_i', x_{-i})}{P(x_i' \to x_i, x_{-i})} \right] \\
&= \sum_{i=1}^{N} \dot{I}_i,
\end{aligned}$$

Local irreversibility associated with element i

$$\dot{I}_i = \sum_{x_{-i}} \sum_{x_i, x_i'} P(x_i \to x_i', x_{-i}) \log \left[\frac{P(x_i \to x_i', x_{-i})}{P(x_i' \to x_i, x_{-i})} \right]$$

Mihir Rao

II. Local irreversibility and multipartite dynamics

Recap of key ideas

1) *Irreversibility* (on average) ⇔ phenomenological arrow of time

2) Quantify irreversibility with *KL divergence*

3) In practice, we don't observe all relevant DoFs, so dynamics of observable states are non-Markovian, but irreversibility is still *local (observable) evidence* for AoT.

4) Taking the limit of multipartite dynamics lets us isolate the irreversibility associated with an individual element in a system.

   (taking this limit is trivial analytically, intractable experimentally)

Outline

I. Time in living systems
II. Local irreversibility and multipartite dynamics
III. Decomposing irreversibility (independent + interacting)
IV. Irreversibility due to kth-order dynamics (k>2)
V. Decomposing irreversibility in
    A. Sensing systems
    B. Logical functions
    C. Neuronal populations

III. Decomposing irreversibility (independent + interacting)

Local evidence of irreversibility

$$\dot{I} = \sum_{x,x'} P(x \to x') \log \left[\frac{P(x \to x')}{P(x' \to x)} \right]$$

internal dynamics of all elements and all element interactions

Local irreversibility associated with element i

$$\dot{I}_i = \sum_{x_{-i}} \sum_{x_i, x_i'} P(x_i \to x_i', x_{-i}) \log \left[\frac{P(x_i \to x_i', x_{-i})}{P(x_i' \to x_i, x_{-i})} \right]$$

internal dynamics of element i and all interactions involving element i

Q: For the irreversibility associated with element i, how can we **separate** the contribution of element i's **independent** internal dynamics from that of element i's **interactions** with other elements?

Consider a hypothetical system in which the elements do not interact.

$$P(x_i \to x_i') = \sum_{x_{-i}} P(x_i \to x_i', x_{-i})$$

Transitions of each element i are completely defined by the marginal transition probabilities.

Independent irreversibility of element i

$$\dot{I}_i^{\text{ind}} = \sum_{x_i, x_i'} P(x_i \to x_i') \log \left[\frac{P(x_i \to x_i')}{P(x_i' \to x_i)} \right]$$

only internal dynamics of element i

We now have the contribution of element i's internal dynamics. How do we obtain the contribution of element i's **interactions**?

Mihir Rao

III. Decomposing irreversibility (independent + interacting)

Local irreversibility associated with element i

$$\dot{I}_i = \sum_{x_{-i}} \sum_{x_i, x_i'} P(x_i \to x_i', x_{-i}) \log \left[\frac{P(x_i \to x_i', x_{-i})}{P(x_i' \to x_i, x_{-i})} \right]$$

Independent irreversibility of element i

$$\dot{I}_i^{\text{ind}} = \sum_{x_i, x_i'} P(x_i \to x_i') \log \left[\frac{P(x_i \to x_i')}{P(x_i' \to x_i)} \right]$$

Define: *Interaction* irreversibility of element i

$$\dot{I}_i^{\text{int}} = \dot{I}_i - \dot{I}_i^{\text{ind}} = \sum_{x_{-i}} \sum_{x_i, x_i'} P(x_i \to x_i', x_{-i}) \left(\log \left[\frac{P(x_i \to x_i', x_{-i})}{P(x_i' \to x_i, x_{-i})} \right] - \log \left[\frac{P(x_i \to x_i')}{P(x_i' \to x_i)} \right] \right)$$

$$= \sum_{x_{-i}} \sum_{x_i, x_i'} P(x_i \to x_i', x_{-i}) \log \left[\frac{P(x_{-i} \mid x_i \to x_i')}{P(x_{-i} \mid x_i' \to x_i)} \right] = \sum_{x_i, x_i'} P(x_i \to x_i') D_{KL}[P(x_{-i} \mid x_i \to x_i') || P(x_{-i} \mid x_i' \to x_i)]$$

only interactions of element i

Notice: $\dot{I}_i^{\text{int}} \geqslant 0$
⇔ the presence of interactions can only increase the local irreversibility of a system.

Interpretation: $\dot{I}_i^{\text{int}}$ is the amount of information one gains about the state x_{-i} of the rest of the system by observing the forward-time dynamics of element i rather than its reverse-time dynamics.

Mihir Rao

III. Decomposing irreversibility (independent + interacting)

Local irreversibility associated with element i

$$\dot{I}_i = \sum_{x_{-i}} \sum_{x_i, x_i'} P(x_i \to x_i', x_{-i}) \log \left[\frac{P(x_i \to x_i', x_{-i})}{P(x_i' \to x_i, x_{-i})} \right]$$

Independent irreversibility of element i

$$\dot{I}_i^{\text{ind}} = \sum_{x_i, x_i'} P(x_i \to x_i') \log \left[\frac{P(x_i \to x_i')}{P(x_i' \to x_i)} \right]$$

Interaction irreversibility of element i

$$\dot{I}_i^{\text{int}} = \dot{I}_i - \dot{I}_i^{\text{ind}}$$

Decomposition result: local irreversibility of a system can be split into two non-negative components:

$$\boxed{\dot{I} = \dot{I}^{\text{ind}} + \dot{I}^{\text{int}}}$$

where $\dot{I}^{\text{ind}} = \sum_{i=1}^{N} \dot{I}_i^{\text{ind}}$ and $\dot{I}^{\text{int}} = \sum_{i=1}^{N} \dot{I}_i^{\text{int}}$

System's local irreversibility due to individual elements' internal dynamics

System's local irreversibility due to interactions between elements

Can we decompose further?

(more granularity = improved experimental fidelity)

Outline

I. Time in living systems
II. Local irreversibility and multipartite dynamics
III. Decomposing irreversibility (independent + interacting)
IV. Irreversibility due to kth-order dynamics (k>2)
V. Decomposing irreversibility in
    A. Sensing systems
    B. Logical functions
    C. Neuronal populations

IV. Irreversibility due to kth-order dynamics (k>2)

Start by considering pairs of elements {i,j} (k=2).

$$P(x_i \to x_i', x_j) = \sum_{x_{-\{i,j\}}} P(x_i \to x_i', x_{-i}),$$
$$P(x_j \to x_j', x_i) = \sum_{x_{-\{i,j\}}} P(x_j \to x_j', x_{-j}).$$

Now, imagine a hypothetical system which matches these marginal dynamics for all pairs {i,j}, and contains **minimal** information about the AoT so that the dynamics are **maximally** reversible.

This corresponds to a minimal irreversibility $\dot{I}^{(2)}$.

$\dot{I}$ is convex, so this is an optimization problem (search for global minimum) with efficient algorithmic solutions.

In general: one can compute the minimum irreversibility $I^{(k)}$ consistent with dynamics of k elements at a time.

kth-order dynamics contain all information about smaller groups of size 1,2,...,k-1:

$$0 \leqslant \dot{I}^{(1)} \leqslant \dot{I}^{(2)} \leqslant \cdots \leqslant \dot{I}^{(N-1)} \leqslant \dot{I}^{(N)} = \dot{I}$$

Local irreversibility due to kth-order dynamics:

$$\dot{I}^{(k)}_{\mathrm{int}} = \dot{I}^{(k)} - \dot{I}^{(k-1)} \geqslant 0$$

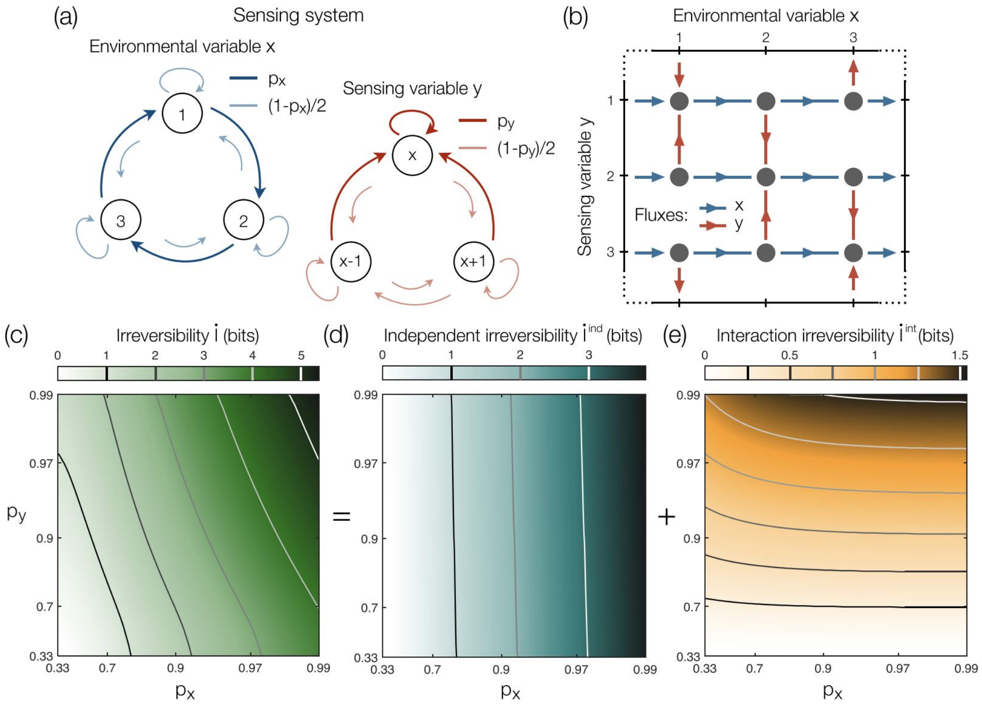
Combining contributions from $\dot{I}^{(1)}_{\mathrm{int}} = \dot{I}^{(1)} = \dot{I}^{\mathrm{ind}}$ to $\dot{I}^{(N)}_{\mathrm{int}}$, we get a **full decomposition result:**

$$\dot{I} = \underbrace{\dot{I}^{(1)}_{\mathrm{int}}}_{\dot{I}^{\mathrm{ind}}} + \underbrace{\dot{I}^{(2)}_{\mathrm{int}} + \dot{I}^{(3)}_{\mathrm{int}} + \cdots + \dot{I}^{(N)}_{\mathrm{int}}}_{\dot{I}^{\mathrm{int}}}$$

Outline

I. Time in living systems
II. Local irreversibility and multipartite dynamics
III. Decomposing irreversibility (independent + interacting)
IV. Irreversibility due to kth-order dynamics (k>2)
V. Decomposing irreversibility in
    A. Sensing systems
    B. Logical functions
    C. Neuronal populations

V.A Sensing systems



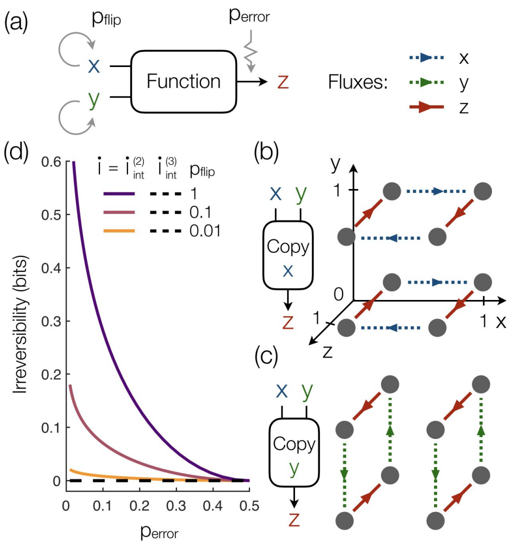
Adapted from Fig. 2.

Mihir Rao

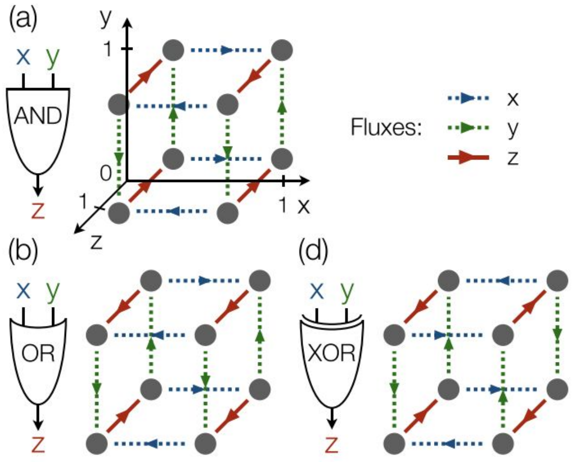
Outline

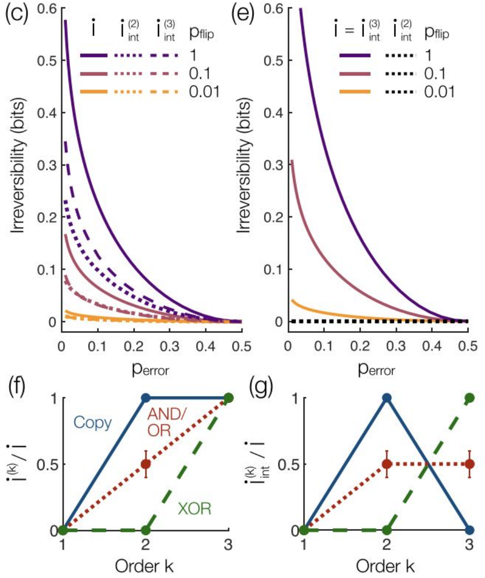
I. Time in living systems
II. Local irreversibility and multipartite dynamics
III. Decomposing irreversibility (independent + interacting)
IV. Irreversibility due to kth-order dynamics (k>2)
V. Decomposing irreversibility in
   A. Sensing systems
   B. Logical functions
   C. Neuronal populations

V.B Logical functions



Adapted from Fig. 3.





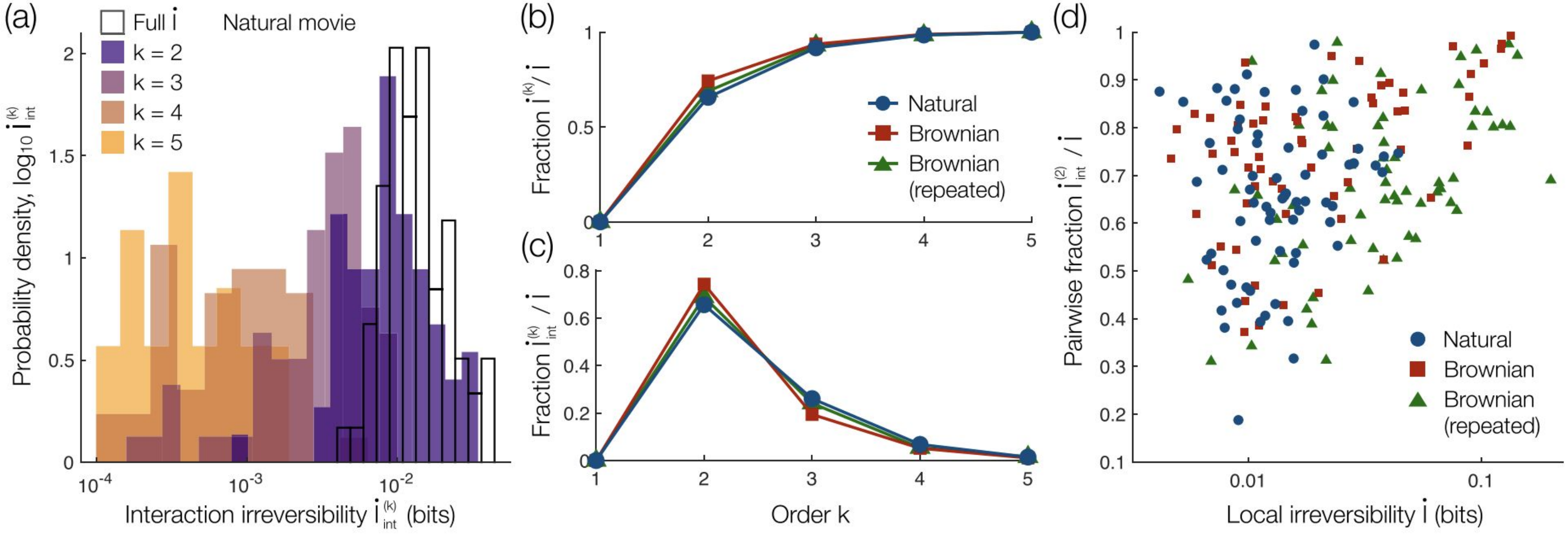
Adapted from Fig. 4.

Mihir Rao

Outline

I. Time in living systems
II. Local irreversibility and multipartite dynamics
III. Decomposing irreversibility (independent + interacting)
IV. Irreversibility due to kth-order dynamics (k>2)
V. Decomposing irreversibility in
   A. Sensing systems
   B. Logical functions
   C. Neuronal populations

V.C Neuronal populations



Adapted from Fig. 7.

Takeaways

- The arrow of time is critically relevant to living systems and phenomenon like development, aging, and disease.
- Irreversibility offers a mechanistic explanation, and detailed balance + information theory offer a sufficiently expressive modeling framework.
- Irreversibility is decomposable into internal and interaction dynamics of system elements, and interaction dynamics can be further decomposed into higher order interaction terms.
- Synthetic (sensing system, logic function) and real (neuronal population) data are compatible with this paper's framing.

Next: explore combinatorial and hypergraphs structure, implement (even approximately) for real biological data modalities to empirically identify AoT "hotspots".