Extending computational complexity theory to include thermodynamic resource costs

David H. Wolpert (Santa Fe Institute)

with

Gülce Kardes, Jan Korbel, Tom Ouldridge, Farita Tasnim

SANTA FE INSTITUTE

COMPLEXITY SCIENCE HUB VIENNA

ASU ARIZONA STATE UNIVERSITY

ICTP The Abdus Salam International Centre for Theoretical Physics

For continuous-time Markov chains sending p(0) to $p(1) = \sum_j P(i \mid j)\, p_j(0)$

Van denBroeck and Esposito, *Physica A*, 2015

For many non-Markvonian chains sending p(0) to $p(1) = \sum_j P(i \mid j)\, p_j(0)$

Ptaszynski and Esposito, *PRL*, 2019

$$-\Delta Q = \Delta\Sigma - \Delta S$$

- $\Delta S = S(p_1) - S(p_0)$ is gain in Shannon entropy of p
- $-\Delta Q$ is (Shannon) entropy flow from system between t = 0 and t = 1
- $\Delta\Sigma$ is total entropy production in system between t = 0 and t = 1
  - ***cannot be negative***

  (I.e., the second law of thermodynamics)

GENERALIZED LANDAUER BOUND

- System connected to multiple reservoirs, e.g., heat baths at different temperatures. (So "$k_B T$" not defined.)
- Arbitrary number of states
- Arbitrary initial distribution p_0
- Arbitrary dynamics $P(x_1 \mid x_0)$

$$-\Delta Q = \Delta\Sigma + S(p_0) - S(p_1)$$

Entropy Production ($\Delta\Sigma$) is non-negative. So:

"*Generalized Landauer's bound*"

$$-\Delta Q \geq S(p_0) - S(p_1)$$

BOOLEAN CIRCUITS

- Currently, all mass-produced computers are implemented with circuits.

- The simplest circuit is one without loops or branches (a "straight-line program")

- If set of allowed gates are a universal basis (e.g., NAND gates), then can build a circuit with them to implement any desired Boolean function.

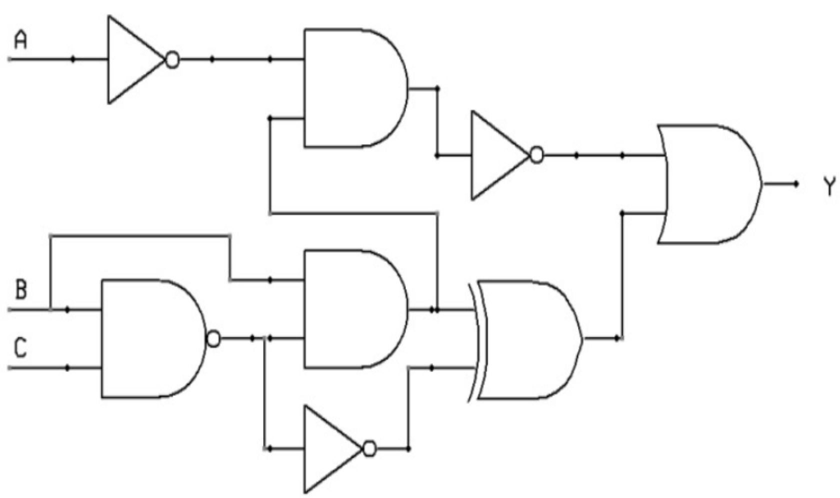
$$-\Delta Q = \Delta\Sigma + S(p_0) - S(p_1)$$

- For fixed $P(x_1 \mid x_0)$, changing p_0 changes Landauer cost, $S(p_0) - S(p_1)$

- N.b., the same $P(x_1 \mid x_0)$ - e.g., same AND gate - has different p_0, depending on where it is in a circuit.

- So *even for a thermo. reversible gate* ($\Delta\Sigma(p_0) = 0$), ***changing the gate's location in a circuit*** (changes $S(p_0) - S(p_1)$ and so) ***changes*** $-\Delta Q(p_0)$

A	B	C	Y
0	0	0	
0	0	1	
0	1	0	
0	1	1	
1	0	0	
1	0	1	
1	1	0	
1	1	1	



- Changing a gate's location in a circuit changes $S(p_0) - S(p_1)$, and so changes the heat it produces, $-\Delta Q(p_0)$

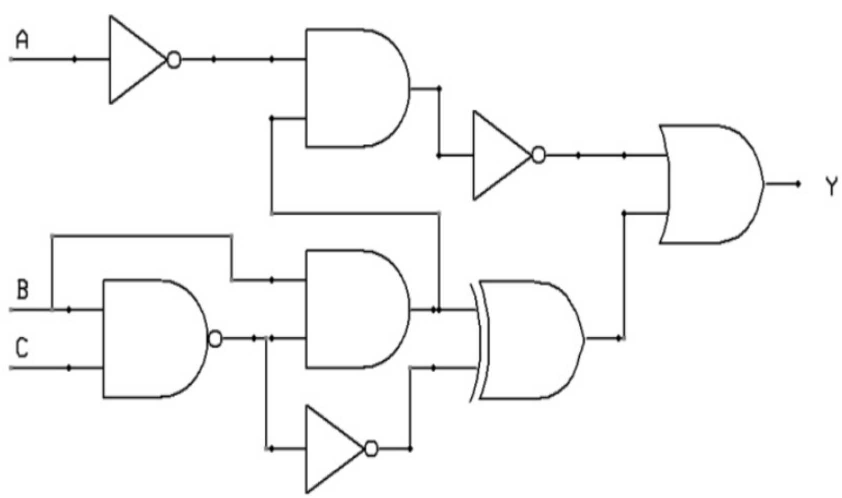
- Sum those heats over all gates to get minimal heat flow of that circuit

Different circuits implementing *same* Boolean function on *same* input distribution have *different* minimal heat

- Formally, those differences in minimal heat of the circuits are differences in EPs of the circuits, arising due to modularity of gates

➢ A new circuit design optimization problem

A	B	C	Y
0	0	0	
0	0	1	
0	1	0	
0	1	1	
1	0	0	
1	0	1	
1	1	0	
1	1	1	



Demaine, E., et al., *Comm. ACM*, 2016

- Considers a similar problem - but incorrectly sets Landauer cost at each gate to same value, KT ln(2).

WHAT IS *REALLY* IMPORTANT THERMODYNAMICALLY?

$$-\Delta Q = \Delta\Sigma + S(p_0) - S(p_1)$$

- System evolves while connected to *single* heat bath at temperature *T*
  - Then heat flow into environment = $-k_B T\, \Delta Q$

- At scale of real computers and brains, $k_B T[S(p_0) - S(p_1)]$ is small

- At scale of real computers and brains, $\Delta\Sigma$ **is dominant cost**
  - Generalized Landauer's bound often irrelevant (it assumes $\Delta\Sigma = 0$!)

What determines $\Delta\Sigma$?

BEYOND GENERALIZED LANDAUER

- System connected to multiple reservoirs, e.g., heat baths at different temperatures. (So "$k_B T$" not defined.)
- Arbitrary number of states
- Arbitrary initial distribution p_0
- Arbitrary dynamics $P(x_1 \mid x_0)$

$$-\Delta Q = \Delta\Sigma + S(p_0) - S(p_1)$$

Entropy Production ($\Delta\Sigma$) is non-negative.

Are there broadly applicable non-negative lower bounds on $\Delta\Sigma$, to complement Landauer's bound?

BEYOND GENERALIZED LANDAUER

- System connected to multiple reservoirs, e.g., heat baths at different temperatures. (So "$k_B T$" not defined.)
- Arbitrary number of states
- Arbitrary initial distribution p_0
- Arbitrary dynamics $P(x_1 \mid x_0)$

$$-\Delta Q = \Delta\Sigma + S(p_0) - S(p_1)$$

Entropy Production ($\Delta\Sigma$) is non-negative.

Are there broadly applicable non-negative lower bounds on $\Delta\Sigma$, to complement Landauer's bound?

- Yes.

BEYOND GENERALIZED LANDAUER

$$-\Delta Q = \Delta\Sigma + S(p_0) - S(p_1)$$

Entropy Production ($\Delta\Sigma$) is non-negative.

Are there broadly applicable non-negative lower bounds on $\Delta\Sigma$, to add to the lower bound $-\Delta Q \geq S(p_0) - S(p_1)$?

- Yes.

  - Focus on two: ***Speed limit theorem*** (SLT) and ***Mismatch cost***

Use them to investigate
the (thermodynamic) resource costs of
computational machines

Original speed limit theorem (SLT): $$\Delta\Sigma \geq \frac{L(p(0), p(1))}{A_{0,1}}$$

- $L(p(0), p(1))$: L_1 distance from distribution p(0) to distribution p(1)
- $A_{0,1}$: total number of (stochastic) state jumps from t = 0 to t = 1

Since introduced, SLT has been strengthened several ways

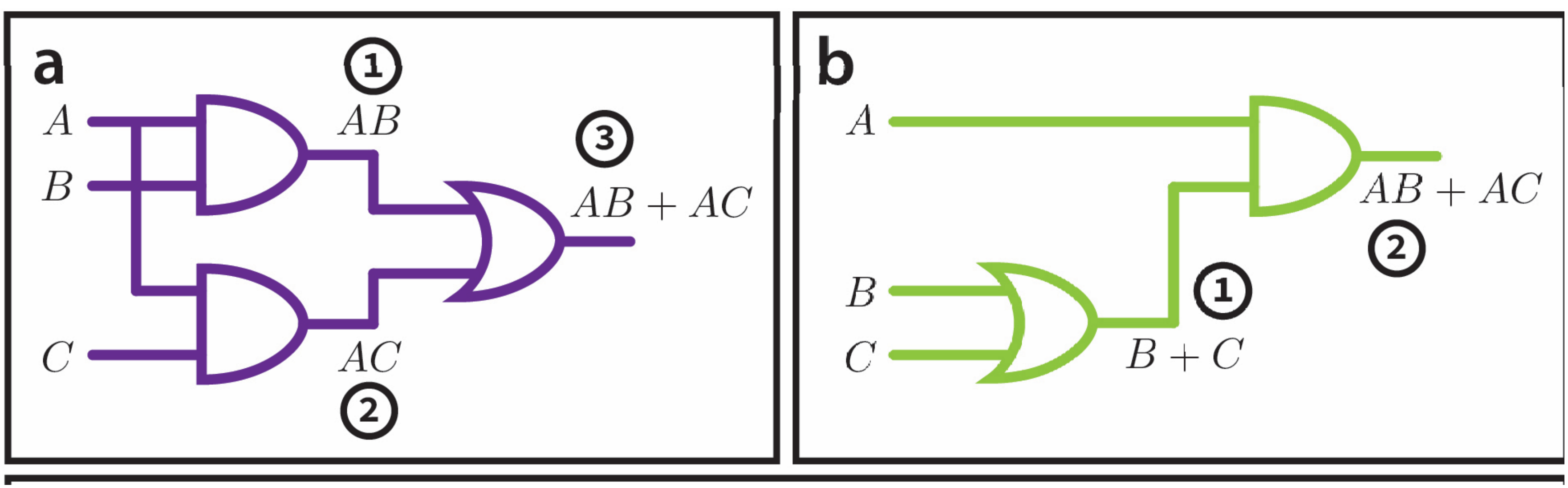
(more complicated formulas).

Shiraishi, N., Funo, K.; Saito, K., *PRL* (2018)
Delvenne, J., Falasco, G.; arXiv:2110.13050
Lee, J., et al.; *PRL* (2022)
Van Vu, T., Saito, K.; *PRL* (2023)

Original speed limit theorem (SLT): $\Delta\Sigma \geq \frac{L(p(0), p(1))}{A_{0,1}}$

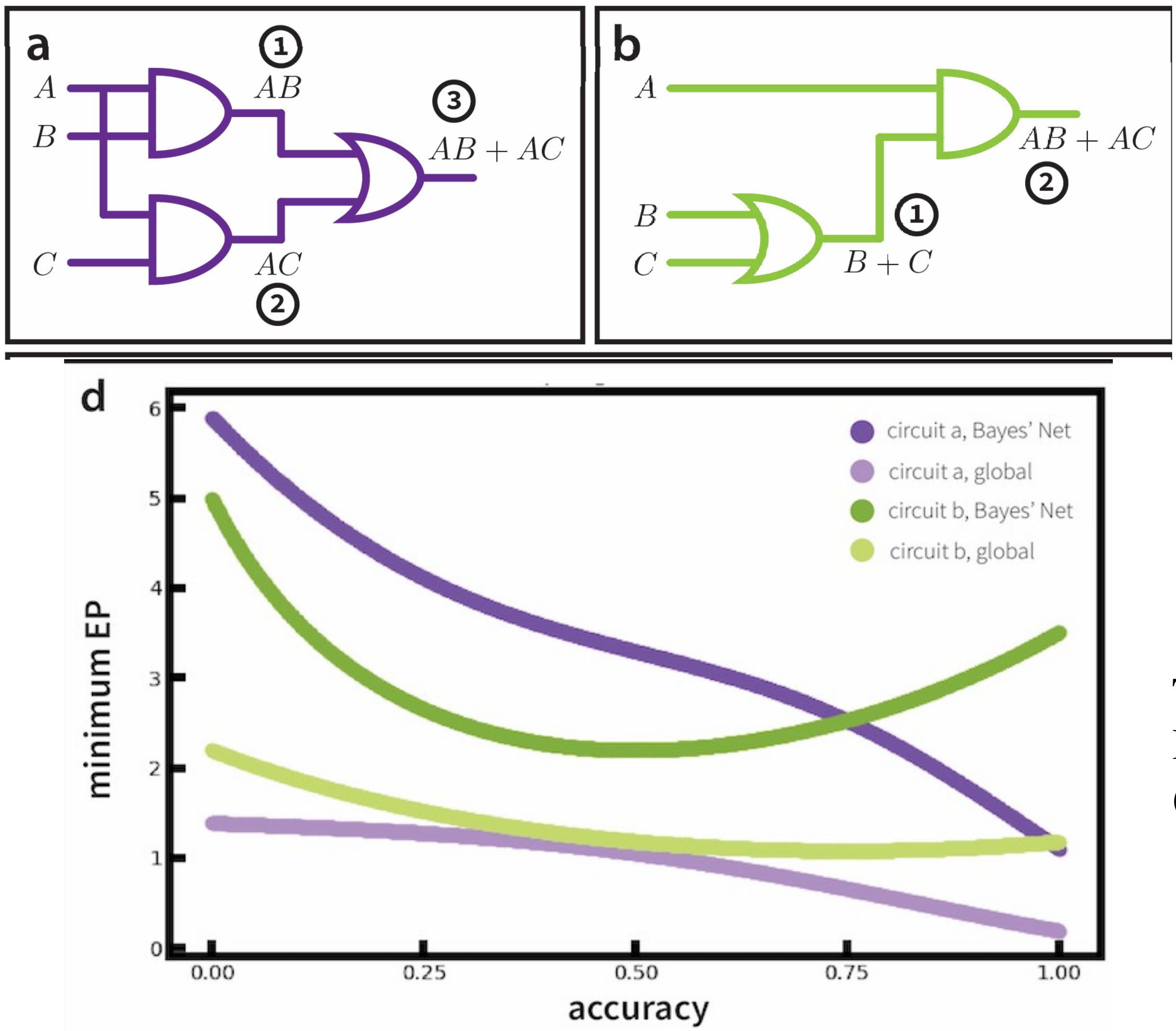
- *L(p(0), p(1))*: L_1 distance from distribution p(0) to distribution p(1)
- $A_{0,1}$: total number of (stochastic) state jumps from t = 0 to t = 1

- Suppose uniform initial distribution over all gates and input bits;
- How does the (Lee et al.) SLT bound vary with error rate of gates, for two logically equivalent circuits?



Original speed limit theorem (SLT): $$\Delta\Sigma \geq \frac{L(p(0), p(1))}{A_{0,1}}$$

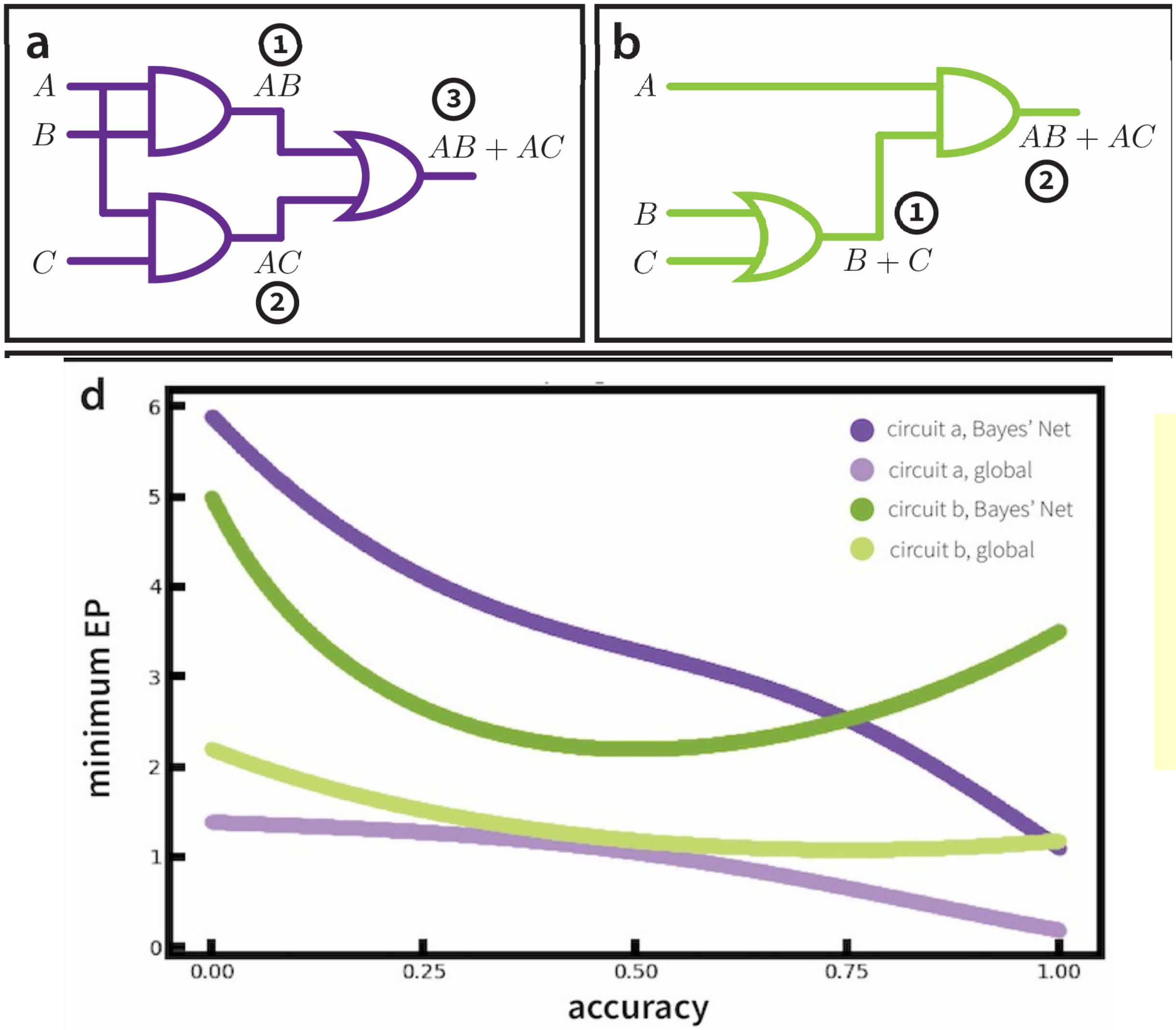
- *L(p(0), p(1)): L_1* distance from distribution p(0) to distribution p(1)
- *$A_{0,1}$*: total number of (stochastic) state jumps from t = 0 to t = 1



Tasnim, F., Wolpert, D., Korbel J., Lynn, C., et al. (2023)

Original speed limit theorem (SLT): $$\Delta\Sigma \geq \frac{L(p(0), p(1))}{A_{0,1}}$$

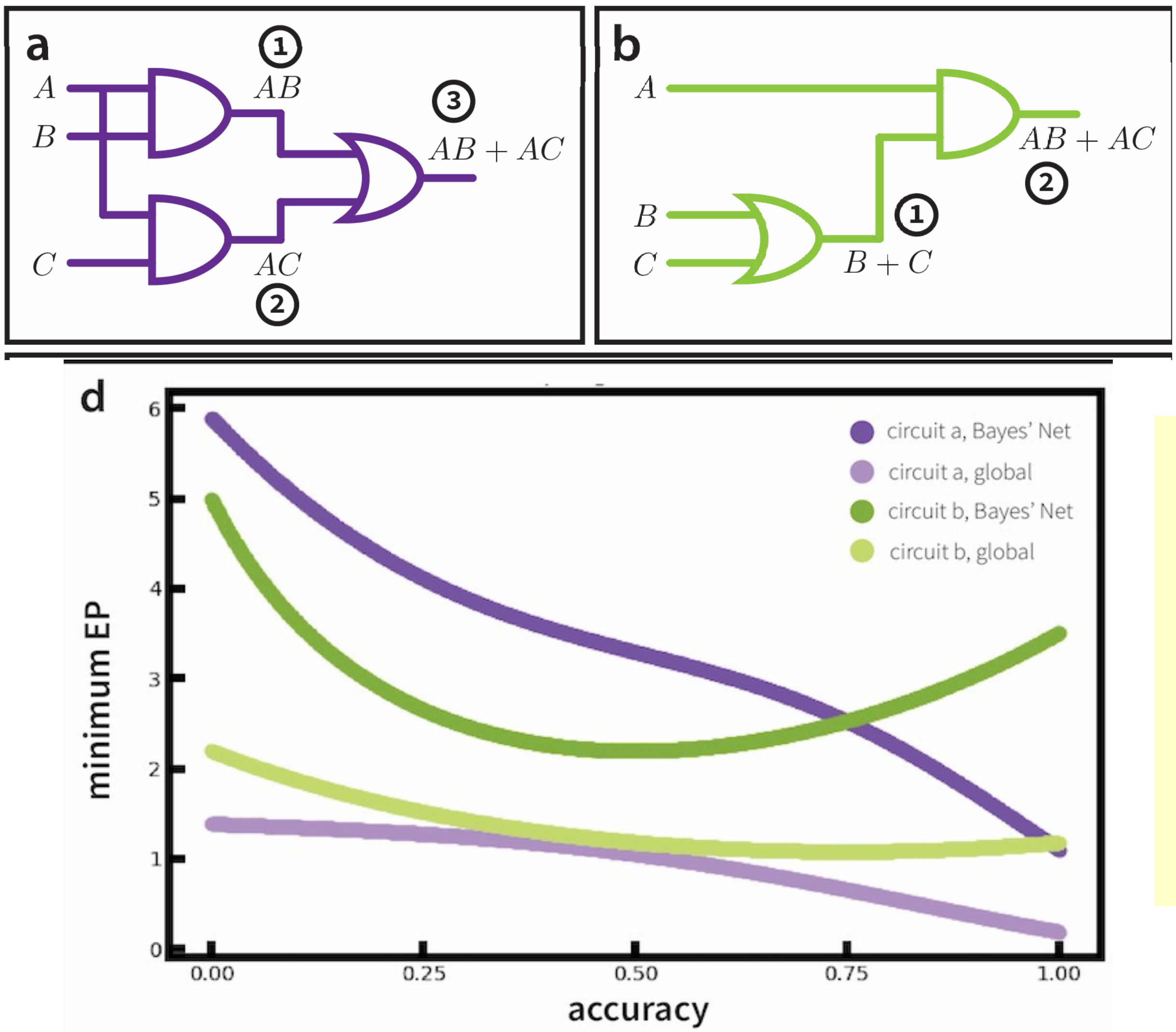
- *L(p(0), p(1))*: L_1 distance from distribution p(0) to distribution p(1)
- $A_{0,1}$: total number of (stochastic) state jumps from t = 0 to t = 1



What causes the curves to have these shapes? What are curves for other circuits?

Original speed limit theorem (SLT): $\Delta\Sigma \geq \frac{L(p(0), p(1))}{A_{0,1}}$

- *L(p(0), p(1)): L_1* distance from distribution p(0) to distribution p(1)
- $A_{0,1}$: total number of (stochastic) state jumps from t = 0 to t = 1



What causes the curves to have these shapes? What are curves for other circuits?

A: Who knows!

DEPENDENCE OF EP ON INITIAL DISTRIBUTION

- Arbitrary dynamics $P(x_1 \mid x_0)$
- ***Assume system is thermo. reversible for initial distribution q_0***

  I.e., $\Delta\Sigma(q_0) = 0$

- Run that system with initial distribution $p_0 \neq q_0$ instead:

$$\Delta\Sigma(p_0) = D(p_0 \,||\, q_0) - D(p_1 \,||\, q_1) \geq 0$$

where D(. || .) is relative entropy (KL divergence)

Kolchinsky, A, Wolpert D., *J. Stat. Mech.* (2017)
Wolpert, D., Kolchinsky, A., *New J. Phys.* (2020)
Riechers, P., Gu, M., *Phys. Rev. E* (2021)
Kolchinsky, A., Wolpert D., *arxiv:2103.05734*

DEPENDENCE OF EP ON INITIAL DISTRIBUTION

- Arbitrary dynamics $P(x_1 \mid x_0)$
- ***Assume system is thermo. reversible for initial distribution q_0***

I.e., $\Delta\Sigma(q_0) = 0$

- Run that system with initial distribution $p_0 \neq q_0$ instead:

$$\Delta\Sigma(p_0) = D(p_0 \,||\, q_0) - D(p_1 \,||\, q_1) \geq 0$$

where D(. || .) is relative entropy (KL divergence)

$D(p_0 \,||\, q_0) - D(p_1 \,||\, q_1)$ is called ***mismatch cost***

Wolpert, D., Kolchinsky, A., *New J. Phys.* (2020)
Riechers, P., Gu, M., *Phys. Rev. E* (2021)
Kolchinsky, A., Wolpert D., *arxiv:2103.05734*

DEPENDENCE OF EP ON INITIAL DISTRIBUTION

- Arbitrary dynamics $P(x_1 \mid x_0)$
- ***Assume system is thermo. reversible for initial distribution q_0***

  I.e., $\Delta\Sigma(q_0) = 0$

- Run that system with initial distribution $p_0 \neq q_0$ instead:

$$\Delta\Sigma(p_0) = D(p_0 \,||\, q_0) - D(p_1 \,||\, q_1) \geq 0$$

where D(. || .) is relative entropy (KL divergence)

Any nontrivial process that is thermodynamically reversible for one initial distribution ***will be costly*** for any other initial distribution

DEPENDENCE OF EP ON INITIAL DISTRIBUTION

- Arbitrary dynamics $P(x_1 \mid x_0)$
- ***Assume system is thermo. reversible for initial distribution q_0***

  I.e., $\Delta\Sigma(q_0) = 0$

- Run that system with initial distribution $p_0 \neq q_0$ instead:

$$\Delta\Sigma(p_0) = D(p_0 \,||\, q_0) - D(p_1 \,||\, q_1) \geq 0$$

where D(. || .) is relative entropy (KL divergence)

Holds for master equations, Langevin dynamics, (open) quantum thermodynamics, inclusive (Hamiltonian) dynamics – pretty much everything.

DEPENDENCE OF EP ON INITIAL DISTRIBUTION

- Arbitrary dynamics $P(x_1 \mid x_0)$
- ***Assume system is thermo. reversible for initial distribution q_0***

  I.e., $\Delta\Sigma(q_0) = 0$

- Run that system with initial distribution $p_0 \neq q_0$ instead:

$$\Delta\Sigma(p_0) = D(p_0 \,||\, q_0) - D(p_1 \,||\, q_1) \geq 0$$

where D(. || .) is relative entropy (KL divergence)

Same formula (just different prior q) for nonadiabatic EP, net lost free energy, etc.

EXAMPLE: Mismatch cost for parallel bit erasure

- ***Two*** distinct bit-erasing gates, each with thermo. rev. initial distribution q_0
- Run gates in parallel, on bits x^A and x^B, with initial distribution $p_0(x^A, x^B)$
- Assume $p_0(x^A) = q_0(x^A)$ and $p_0(x^B) = q_0(x^B)$.

- So each gate, by itself, generates zero EP. But:

If $p_0(x^A, x^B)$ statistically couples the bits, then full system is **not** *thermo. reversible, and generates nonzero EP*

- **Formally**: Since gates are distinct, the thermo. rev. *joint* distribution is

$$q_0(x^A, x^B) = q_0(x^A)q_0(x^B)$$

EXAMPLE: Mismatch cost for parallel bit erasure

- ***Two*** distinct bit-erasing gates, each with thermo. rev. initial distribution q_0
- Run gates in parallel, on bits x^A and x^B, with initial distribution $p_0(x^A, x^B)$
- Assume $p_0(x^A) = q_0(x^A)$ and $p_0(x^B) = q_0(x^B)$.

- So each gate, by itself, generates zero EP. But:

If $p_0(x^A, x^B)$ statistically couples the bits, then
full system is **not** *thermo. reversible,*
and generates nonzero EP

- **Formally**: Since gates are distinct, the thermo. rev. *joint* distribution is

$q_0(x^A, x^B) = q_0(x^A)q_0(x^B)$. So $D(p_0 \,||\, q_0) - D(p_1 \,||\, q_1) \neq 0$

EXAMPLE: Mismatch cost for parallel bit erasure

- ***Two*** distinct bit-erasing gates, each with thermo. rev. initial distribution q_0
- Run gates in parallel, on bits x^A and x^B, with initial distribution $p_0(x^A, x^B)$
- Assume $p_0(x^A) = q_0(x^A)$ and $p_0(x^B) = q_0(x^B)$.

- So each gate, by itself, generates zero EP. But:

If $p_0(x^A, x^B)$ statistically couples the bits, then full system is **not** *thermo. reversible, and generates nonzero EP*

- Intuition: **Running two thermo. reversible gates in parallel loses information in their initial coupling, and so is not thermo. reversible**.

EXAMPLE: Mismatch cost for parallel bit erasure

- ***Two*** distinct bit-erasing gates, each with thermo. rev. initial distribution q_0
- Run gates in parallel, on bits x^A and x^B, with initial distribution $p_0(x^A, x^B)$
- Assume $p_0(x^A) = q_0(x^A)$ and $p_0(x^B) = q_0(x^B)$.

- So each gate, by itself, generates zero EP. But:

If $p_0(x^A, x^B)$ statistically couples the bits, then full system is **not** *thermo. reversible, and generates nonzero EP*

- **Broader lesson**: ***Modularity increases EP***

EXAMPLE: Mismatch cost for parallel bit erasure

- ***Two*** distinct bit-erasing gates, each with thermo. rev. initial distribution q_0
- Run gates in parallel, on bits x^A and x^B, with initial distribution $p_0(x^A, x^B)$
- Assume $p_0(x^A) = q_0(x^A)$ and $p_0(x^B) = q_0(x^B)$.

- So each gate, by itself, generates zero EP. But:

If $p_0(x^A, x^B)$ statistically couples the bits, then full system is **not** *thermo. reversible, and generates nonzero EP*

- **Broader lesson**: Whatever its practical benefits might be,

modularity is thermodynamically costly (!)

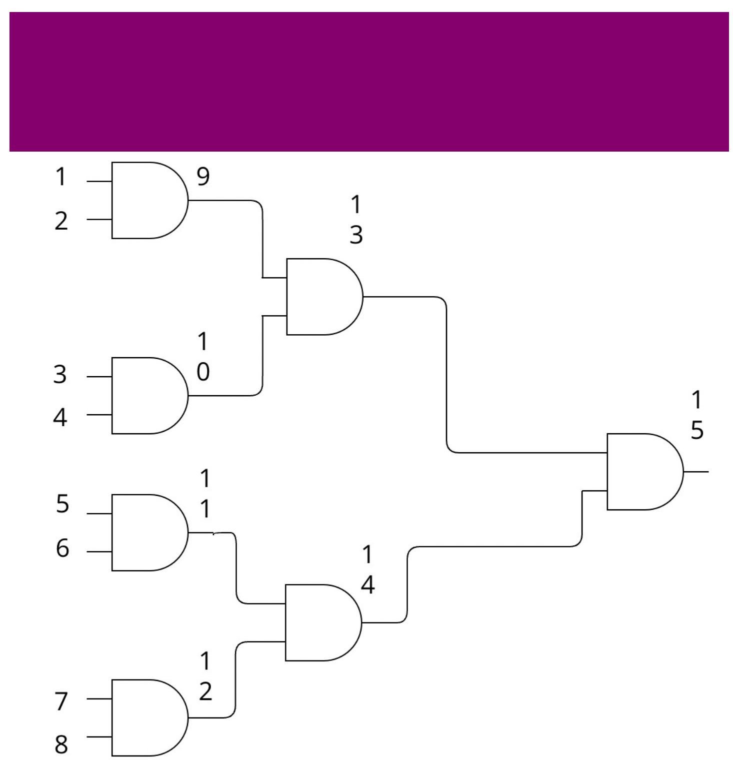
EXAMPLE: Mismatch cost for reusing a Boolean circuit

Over N periods, the sum-total mismatch cost (lower bound on EP) is:

$$\sum_{t=0}^{N-1} \left[D(P^t p_0 \,||\, q) - D(P^{t+1} p_0 \,||\, Pq) \right]$$

Gates are not reinitialized after being run; have old values when next run.
So assuming IID generation of input, $R(x_0)$, initially, joint distribution is

$$R(x) = R(x_0)\, R(x_{1,2,3} \mid x_0)$$



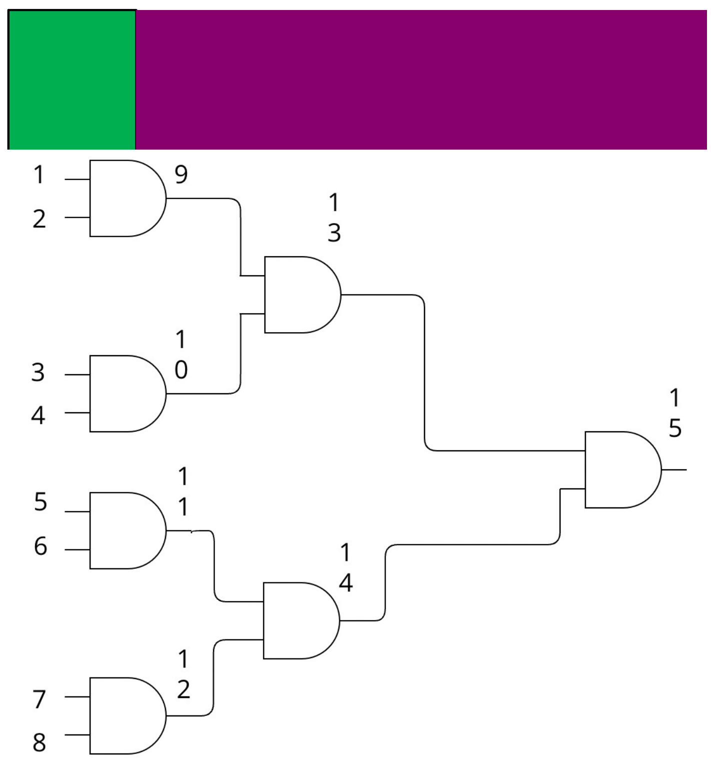
EXAMPLE: Mismatch cost for reusing a Boolean circuit

Over N periods, the sum-total mismatch cost (lower bound on EP) is:

$$\sum_{t=0}^{N-1} \left[D(P^t p_0 \,||\, q) - D(P^{t+1} p_0 \,||\, Pq) \right]$$

So assuming IID generation of input, $R(x_0)$, after running 1st layer, joint distribution is

$$R(x) \;=\; R(x_0)\, R(x_{1,2,3})$$



EXAMPLE: Mismatch cost for reusing a Boolean circuit

Over N periods, the sum-total mismatch cost (lower bound on EP) is:

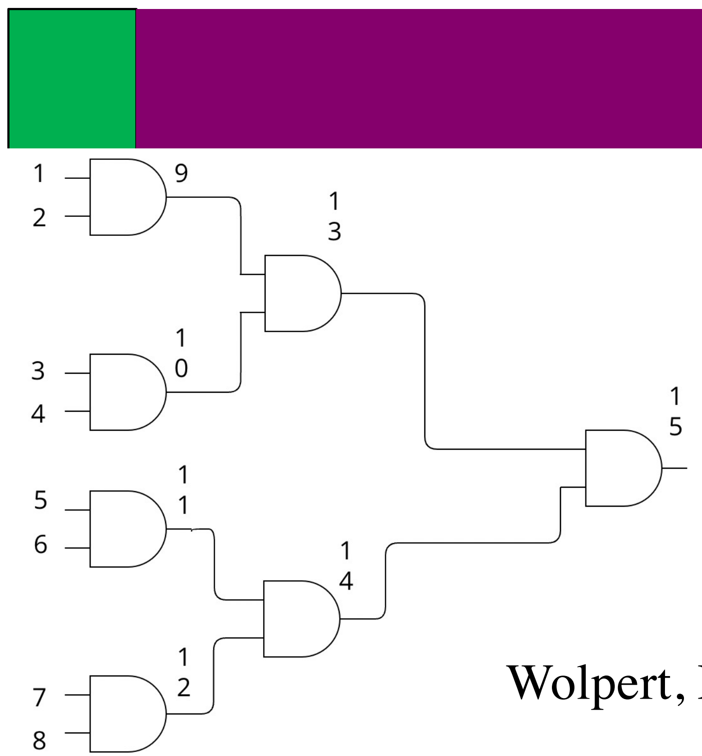
$$\sum_{t=0}^{N-1} \left[D(P^t p_0 \,||\, q) - D(P^{t+1} p_0 \,||\, Pq) \right]$$

So running 1st layer gives mismatch cost

$$D\big(R(x_0)\, R(x_{1,2,3} \mid x_0) \,||\, q(x_0)\, q(x_{1,2,3} \mid x_0)\big) \; - \; D\big(R(x_0)\, R(x_{1,2,3}) \,||\, R(x_0) q(x_{1,2,3})\big)$$

$$\geq$$

$$I_R(X_0\, ; X_1, X_2, X_3)$$



Wolpert, D.H. *Phys. Rev. Letters*, **125**, (2020)

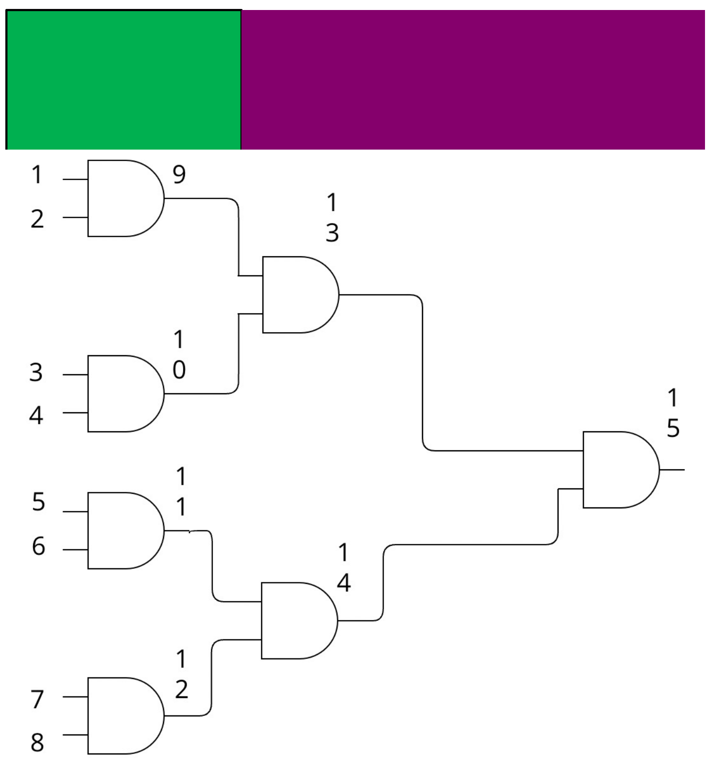
EXAMPLE: Mismatch cost for reusing a Boolean circuit

Over N periods, the sum-total mismatch cost (lower bound on EP) is:

$$\sum_{t=0}^{N-1} \left[D(P^t p_0 \,||\, q) - D(P^{t+1} p_0 \,||\, Pq) \right]$$

So assuming IID generation of input, $R(x_0)$, after running 2nd layer, joint distribution is

$$R(x) = R(x_{0,1})\, R(x_{2,3})$$



EXAMPLE: Mismatch cost for reusing a Boolean circuit

Over N periods, the sum-total mismatch cost (lower bound on EP) is:

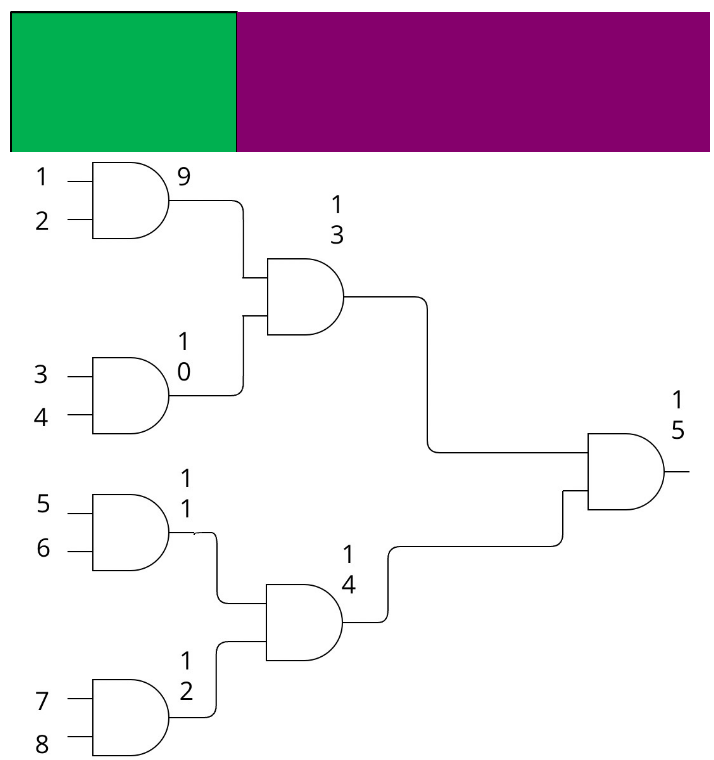
$$\sum_{t=0}^{N-1} \left[D(P^t p_0 \,||\, q) - D(P^{t+1} p_0 \,||\, Pq) \right]$$

So running 2nd layer gives mismatch cost

$$D(R(x_0)\, R(x_{1,2,3}) \,||\, R(x_0)\, q(x_{1,2,3})) \;-\; D(R(x_{0,1}) R(x_{2,3}) \,||\, R(x_{0,1}) q(x_{2,3}))$$

$$\geq$$

$$I_R(X_1\,;\, X_2, X_3)$$



EXAMPLE: Mismatch cost for reusing a Boolean circuit

Over N periods, the sum-total mismatch cost (lower bound on EP) is:

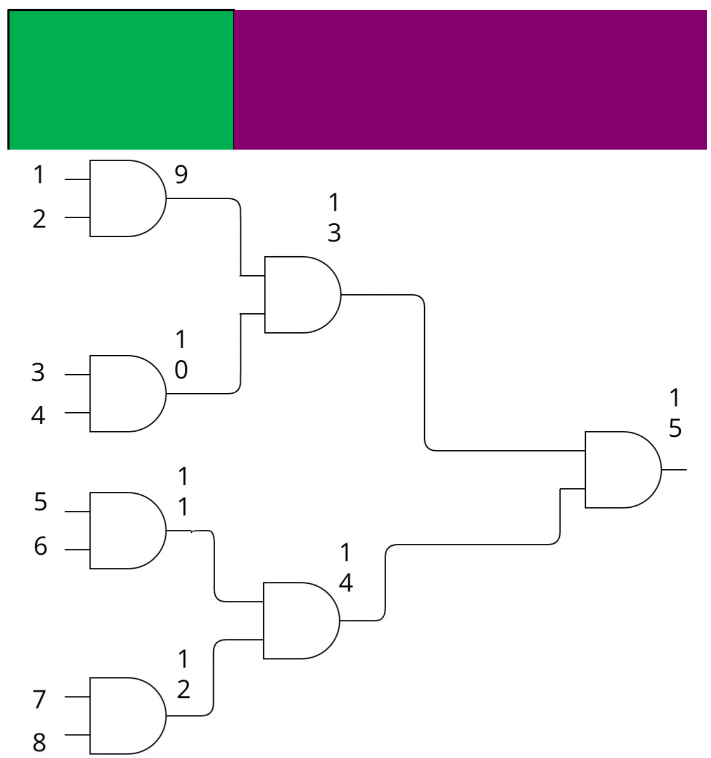
$$\sum_{t=0}^{N-1} \left[D(P^t p_0 \,||\, q) - D(P^{t+1} p_0 \,||\, Pq) \right]$$

So running 2nd layer gives mismatch cost

$$D(R(x_0)\, R(x_{1,2,3}) \,||\, R(x_0)\, q(x_{1,2,3})) \;-\; D(R(x_{0,1}) R(x_{2,3}) \,||\, R(x_{0,1}) q(x_{2,3}))$$

$$\geq$$

$$I_R(X_1 \,;\, X_2, X_3)$$



Technical detail: Also is a variable saying which layer is currently being updated, which increments in each iteration

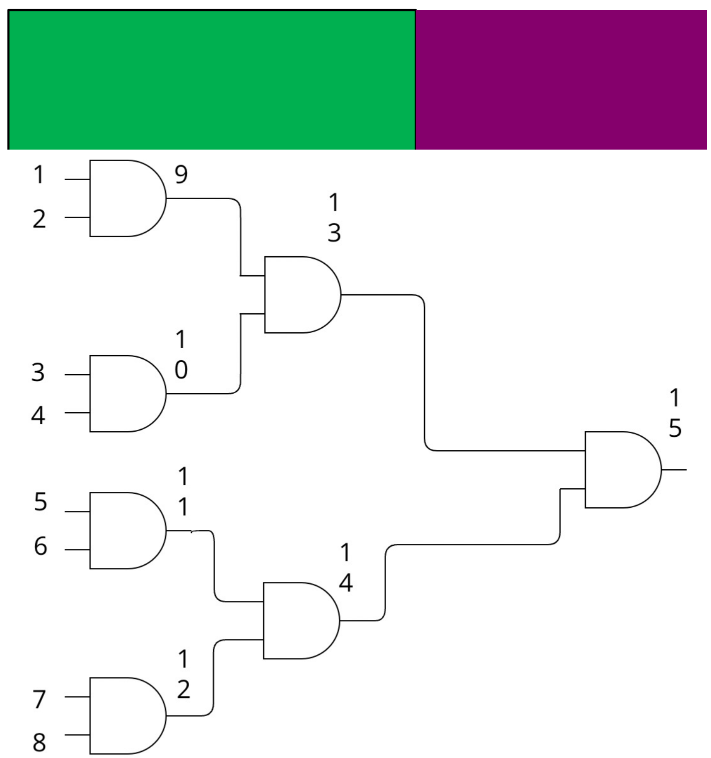
EXAMPLE: Mismatch cost for reusing a Boolean circuit

Over N periods, the sum-total mismatch cost (lower bound on EP) is:

$$\sum_{t=0}^{N-1} \left[D(P^t p_0 \,||\, q) - D(P^{t+1} p_0 \,||\, Pq) \right]$$

So assuming IID generation of input, $R(x_0)$, after running 3rd layer, joint distribution is

$$R(x) = R(x_{0,1,2})\, R(x_3)$$



EXAMPLE: Mismatch cost for reusing a Boolean circuit

Over N periods, the sum-total mismatch cost (lower bound on EP) is:

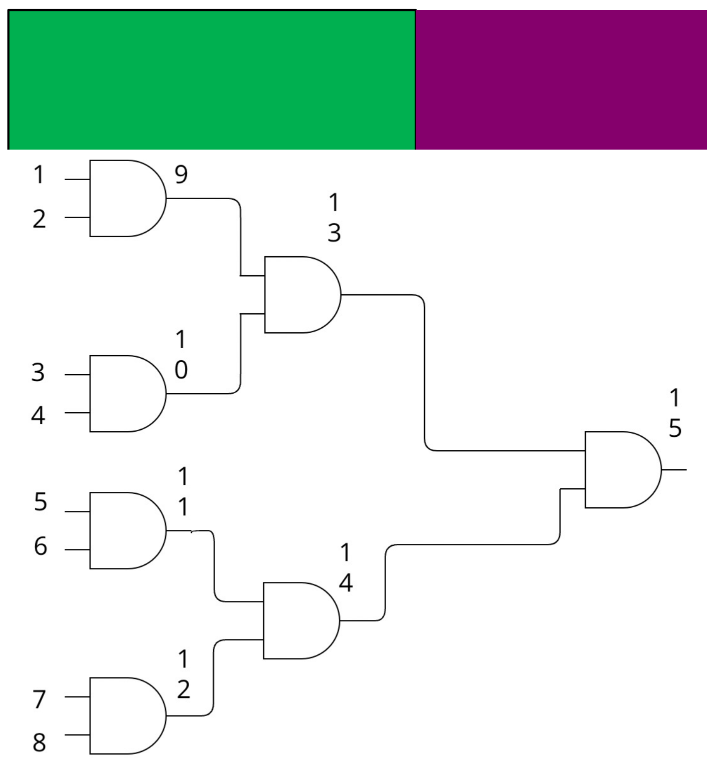
$$\sum_{t=0}^{N-1} \left[D(P^t p_0 \,||\, q) - D(P^{t+1} p_0 \,||\, Pq) \right]$$

So running 3rd layer gives mismatch cost

$$D(R(x_{0,1})R(x_{2,3}) \,||\, R(x_{0,1})q(x_{2,3})) \;-\; D(R(x_{0,1,2})R(x_3) \,||\, R(x_{0,1,2})q(x_3))$$

$$\geq$$

$$I_R(X_2 \,;\, X_3)$$



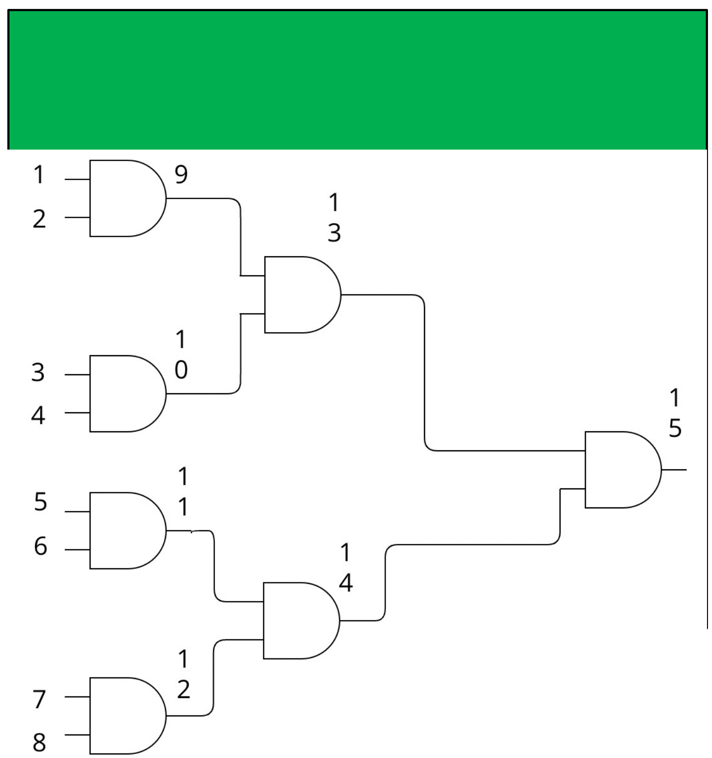
EXAMPLE: Mismatch cost for reusing a Boolean circuit

Over N periods, the sum-total mismatch cost (lower bound on EP) is:

$$\sum_{t=0}^{N-1} \left[D(P^t p_0 \,||\, q) - D(P^{t+1} p_0 \,||\, Pq) \right]$$

So assuming IID generation of input, $R(x_0)$, after running 4th layer, joint distribution is

$$R(x) \;=\; R(x_{0,1,2,3}) \;=\; R(x_{0,1,2,3})$$



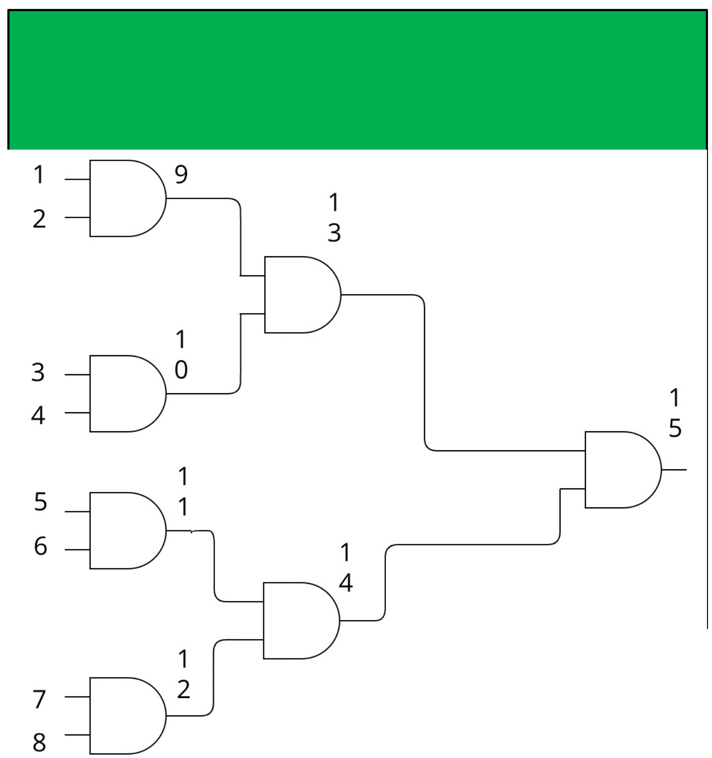
EXAMPLE: Mismatch cost for reusing a Boolean circuit

Over N periods, the sum-total mismatch cost (lower bound on EP) is:

$$\sum_{t=0}^{N-1} \left[D(P^t p_0 \,||\, q) - D(P^{t+1} p_0 \,||\, Pq) \right]$$

So running 4th layer gives mismatch cost

$$D(R(x_{0,1,2})R(x_3) \,||\, R(x_{0,1,2})q(x_3)) \; - D(R(x_{0,1,2,3}) \,||\, R(x_{0,1,2,3}))$$



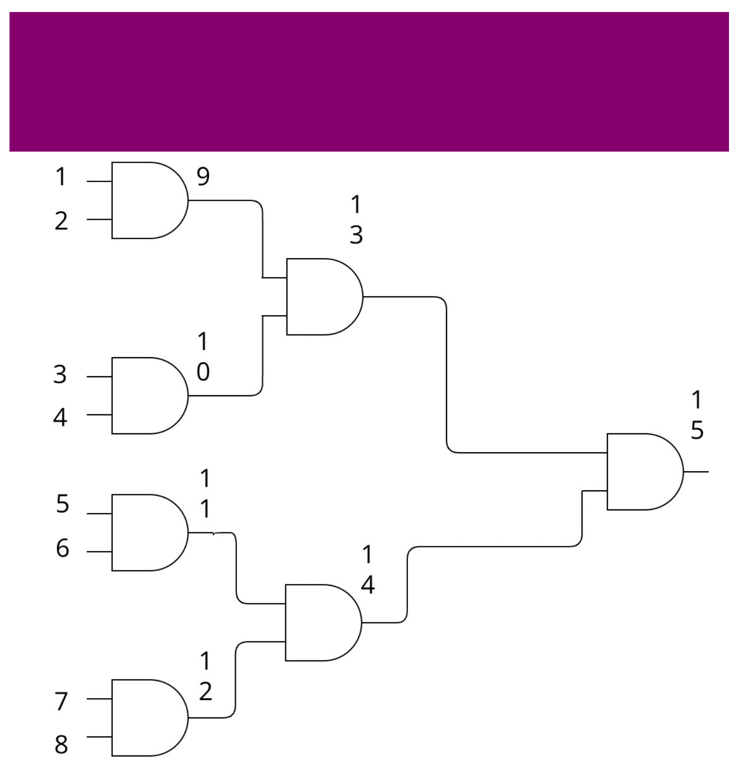
EXAMPLE: Mismatch cost for reusing a Boolean circuit

Over N periods, the sum-total mismatch cost (lower bound on EP) is:

$$\sum_{t=0}^{N-1} \left[D(P^t p_0 \,||\, q) - D(P^{t+1} p_0 \,||\, Pq) \right]$$

So assuming IID generation of input, $R(x_0)$, after 4th layer, joint distribution is

$$R(x) \;=\; R(x_0)R(x_{1,2,3} \mid x_0) = R(x)$$



MISMATCH COST IN PERIODIC PROCESSES

A process over a space X that is periodic, with period **λ**. So for all n,

$$P\left(x(n\lambda) \mid x((n-1)\lambda)\right) = P\left(x(\lambda) \,|\, x(0)\right)$$

So over N periods, the sum-total mismatch cost (lower bound on EP) is:

$$\sigma(N\lambda) \geq \inf_{q \in \Delta_X} \sum_{t=0}^{N-1} \left[D(P^t p_0 \,||\, q) - D(P^{t+1} p_0 \,||\, Pq)\right]$$

MISMATCH COST IN PERIODIC PROCESSES

A process over a space X that is periodic, with period **λ**. So for all n,

$$P\left(x(n\lambda) \mid x((n-1)\lambda)\right) = P\left(x(\lambda) \mid x(0)\right)$$

So over N periods, the sum-total mismatch cost (lower bound on EP) is:

$$\sigma(N\lambda) \geq \inf_{q \in \Delta_X} \sum_{t=0}^{N-1} \left[D(P^t p_0 \,||\, q) - D(P^{t+1} p_0 \,||\, Pq) \right]$$

KEY POINT: Since the process is periodic, q is the same in each period.
However, $P^t p_0$ will differ over periods.

Therefore

At most one mismatch cost in the sum can equal 0 in general

A process over a space X that is periodic, with period **λ**. So for all n,

$$P\left(x(n\lambda) \mid x((n-1)\lambda)\right) = P\left(x(\lambda) \mid x(0)\right)$$

So over N periods, the sum-total mismatch cost (lower bound on EP) is:

$$\sigma(N\lambda) \geq \inf_{q \in \Delta_X} \sum_{t=0}^{N-1} \left[D(P^t p_0 \,||\, q) - D(P^{t+1} p_0 \,||\, Pq)\right]$$

KEY POINT: Since the process is periodic, q is the same in each period.
However, $P^t p_0$ will differ over periods.

Therefore *At most one mismatch cost in the sum can equal 0 in general*

A new strictly positive lower bound on EP for *any* periodic process

A process over a space X that is periodic, with period **λ**. So for all n,

$$P\left(x(n\lambda) \mid x((n-1)\lambda)\right) = P\left(x(\lambda) \mid x(0)\right)$$

So over N periods, the sum-total mismatch cost (lower bound on EP) is:

$$\sigma(N\lambda) \geq N\Delta JS(\{P^t p_0\})$$

- JS is Jensen-Shannon divergence for uniform distribution over N periods

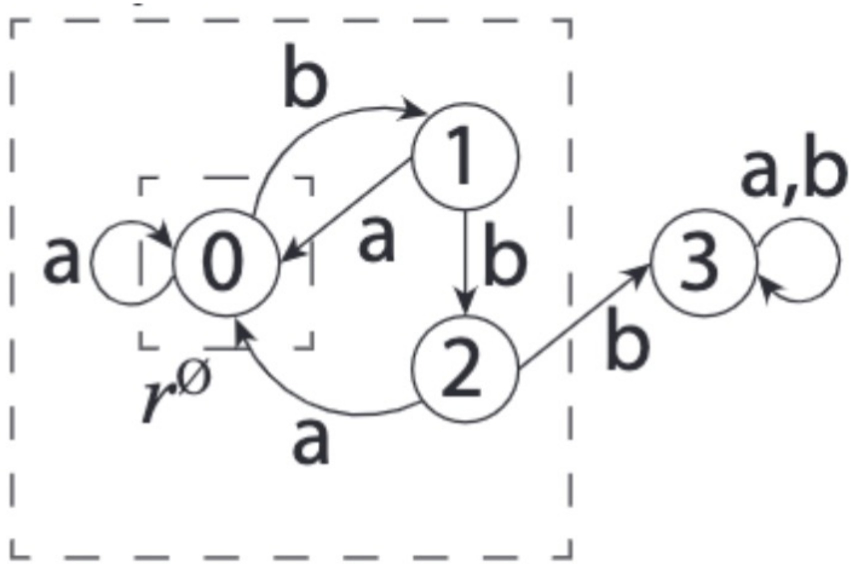
A new strictly positive lower bound on EP for *any* periodic process

Wolpert, D.H. *Phys. Rev. Letters*, **125**, (2020)
Ouldridge, T.; Wolpert, D.H., arxiv:2208.06895 (2022)
Tasnim, F.; Wolpert, D.H. *Entropy*, (2023)
Manzano, G., Kardes, G.; Roldan, E.; Wolpert, D.H., arxiv: 2307.05713 (2023)

DETERMINISTIC FINITE AUTOMATA (DFA)

- **Simplest computational machine in Chomsky hierarchy**
  - Finite number of states; one initial state, multiple "accept states"
  - Feed in a finite string of bits;
  - Each (bit, state) pair maps to a new state, after which next bit is read
  - A DFA "accepts" a string if it causes the DFA to end in an accept state
  - "Language" of a DFA is all input strings that it accepts
  - Many languages that are not accepted by *any* DFA

- **Example**: DFA that accepts any string with no more than two successive 'b' bits:



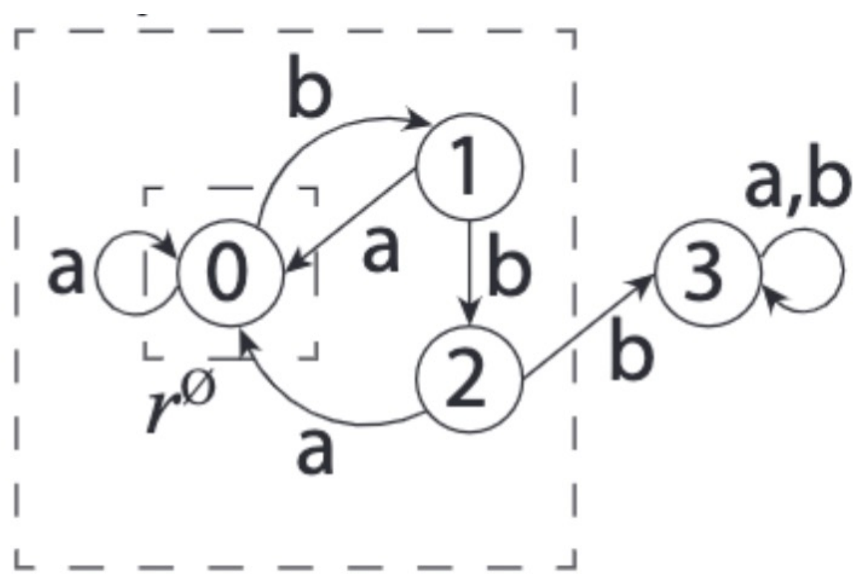
- *Every* digital computer is a **sequence of solitary processes**
    - Only part of the memory is physically to any processor at any time
    - So evolving subsystem is processor and current part of memory

- Results in **modularity (mismatch) cost** – just like parallel bit erasure
    - State space = {joint state of *full* memory and processor}

- *Every* digital computer is a **sequence of solitary processes**
    - Only part of the memory is physically to any processor at any time
    - So evolving subsystem is processor and current part of memory

- Results in **modularity (mismatch) cost** – just like parallel bit erasure
    - State space = {joint state of *full* memory and processor}

Example: In a DFA, memory contains *entire* string of input symbols
However, DFA state only physically coupled to *current* input symbol, **not** earlier or later symbols in the string

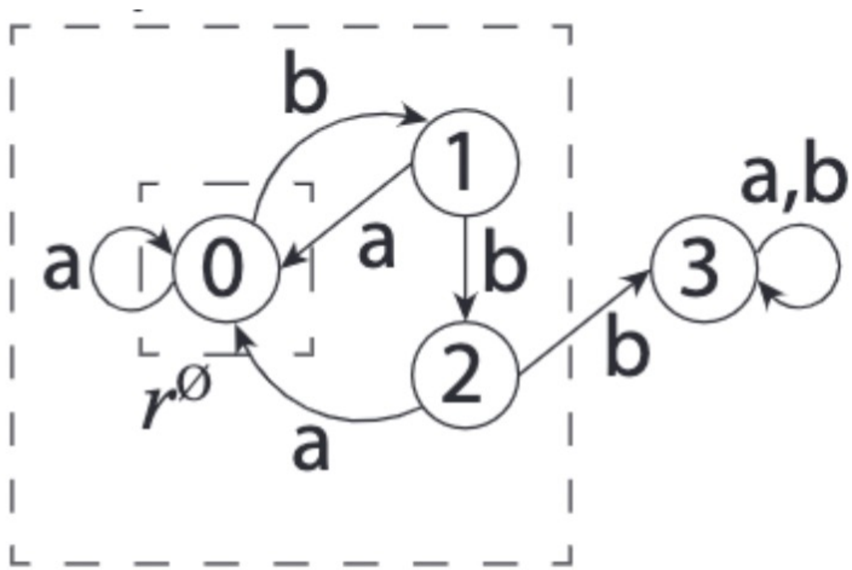


- *Every* (synchronous) digital computer is "**periodic**"
  - Every successive iteration is the same physical process.
  And so every iteration **has the same prior**

- E.g., in a DFA, every iteration has same prior

- So if prior distribution = actual distribution for iteration i, there is zero mismatch cost for iteration i .... ***But***:
  The distributions will differ for iteration i + 1 in general

- Results in "**periodicty (mismatch) cost**"



- *Every* (synchronous) digital computer is "**periodic**"
  - Every successive iteration is the same physical process.
    And so every iteration **has the same prior**

- E.g., in a DFA, every iteration has same prior

- So if prior distribution = actual distribution for iteration i, there is
  zero mismatch cost for iteration i .... ***But***:
  The distributions will differ for iteration i + 1 in general

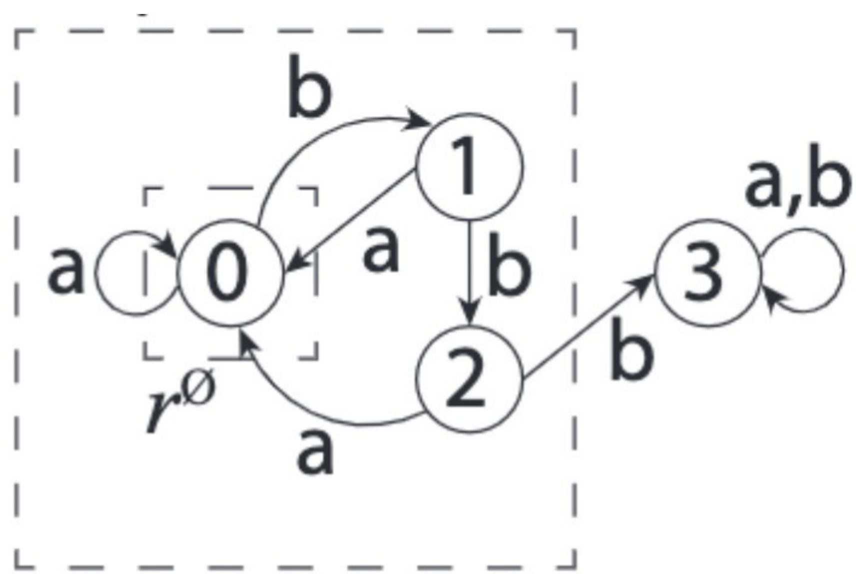
- Results in "**periodicity (mismatch) cost**"

- So, for any p_0 and any update conditional distribution P,

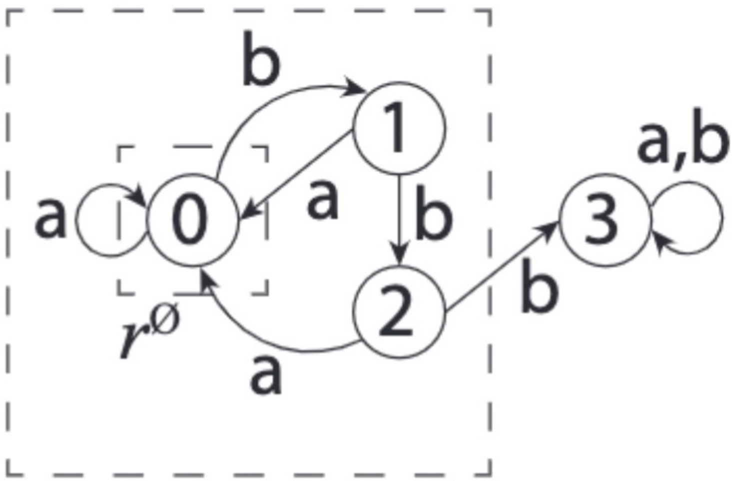
$$EP \geq \operatorname{argmin}_q \sum_{i=0}^{n} \left[D(P^i p_0 || q) - D(P^{i+1} p_0 || Pq) \right]$$

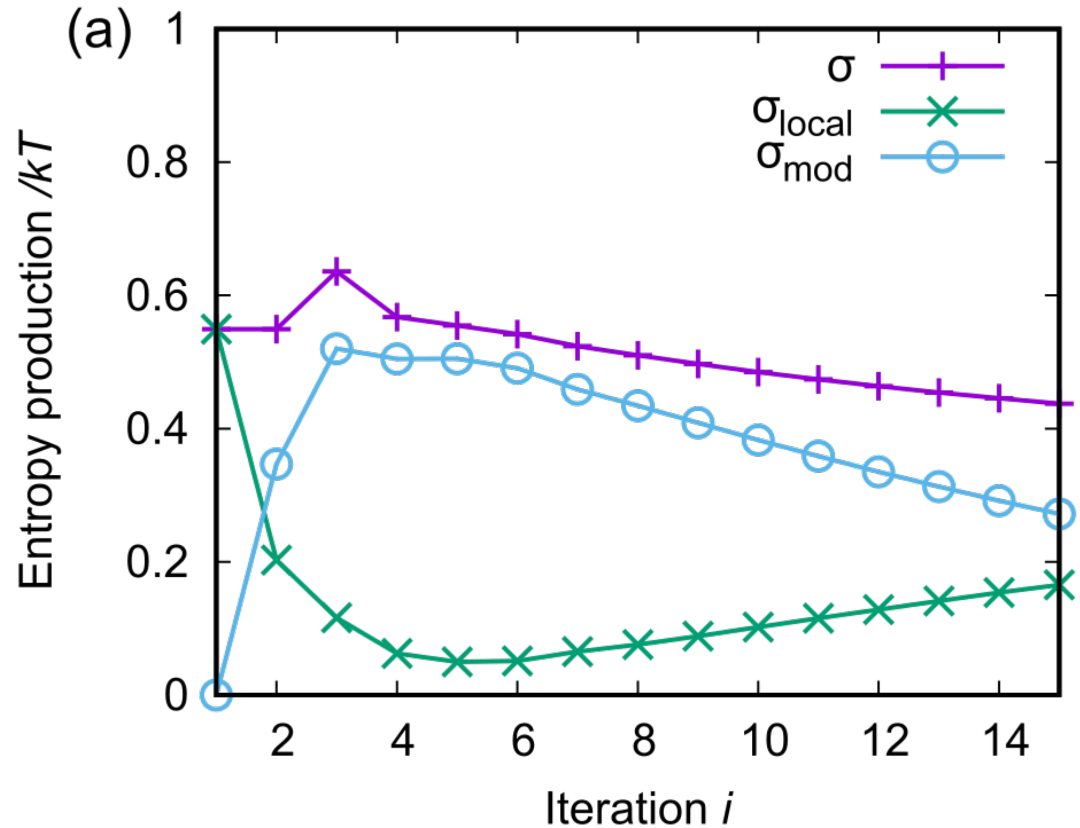
  - *Independent of the physical details of the underlying process*
    (just like generalized Landauer bound)
  - In general, RHS is strictly positive

- Total mismatch cost = modularity cost + periodicity cost



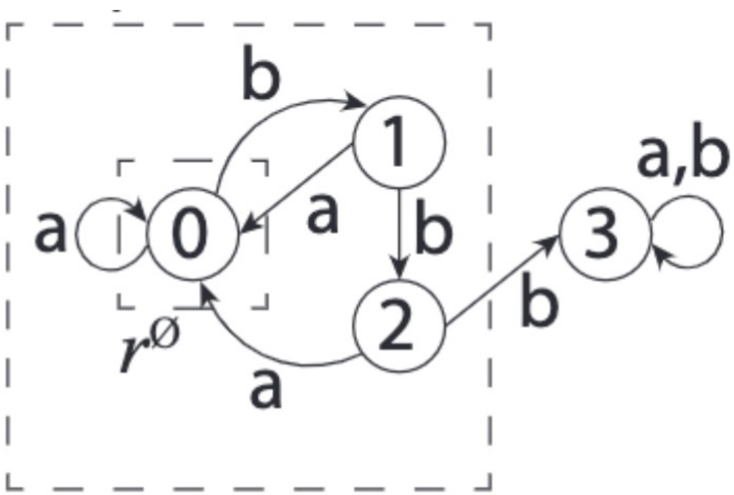
EXAMPLE

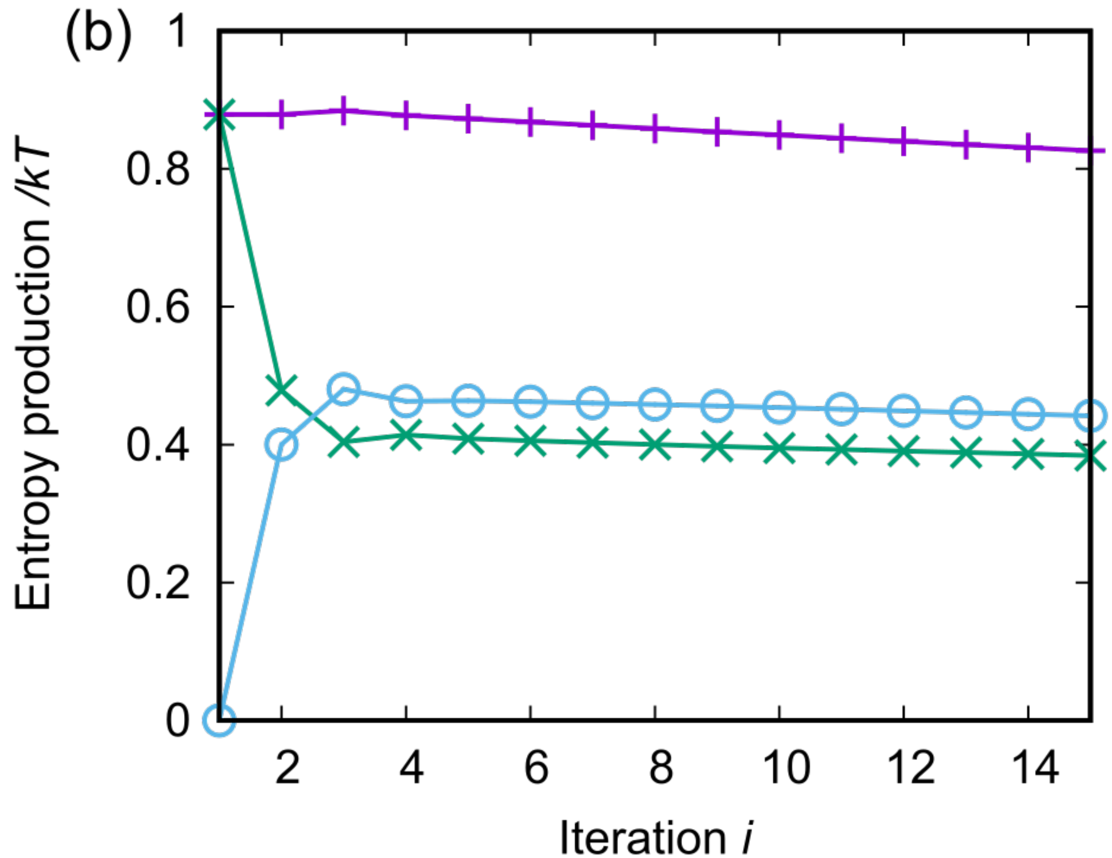




- Input strings have IID symbols with equal probability of a and b
- Uniform prior

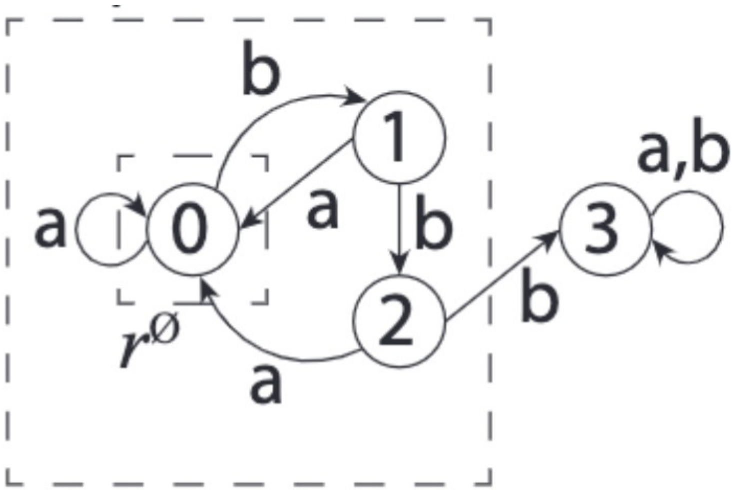
EXAMPLE

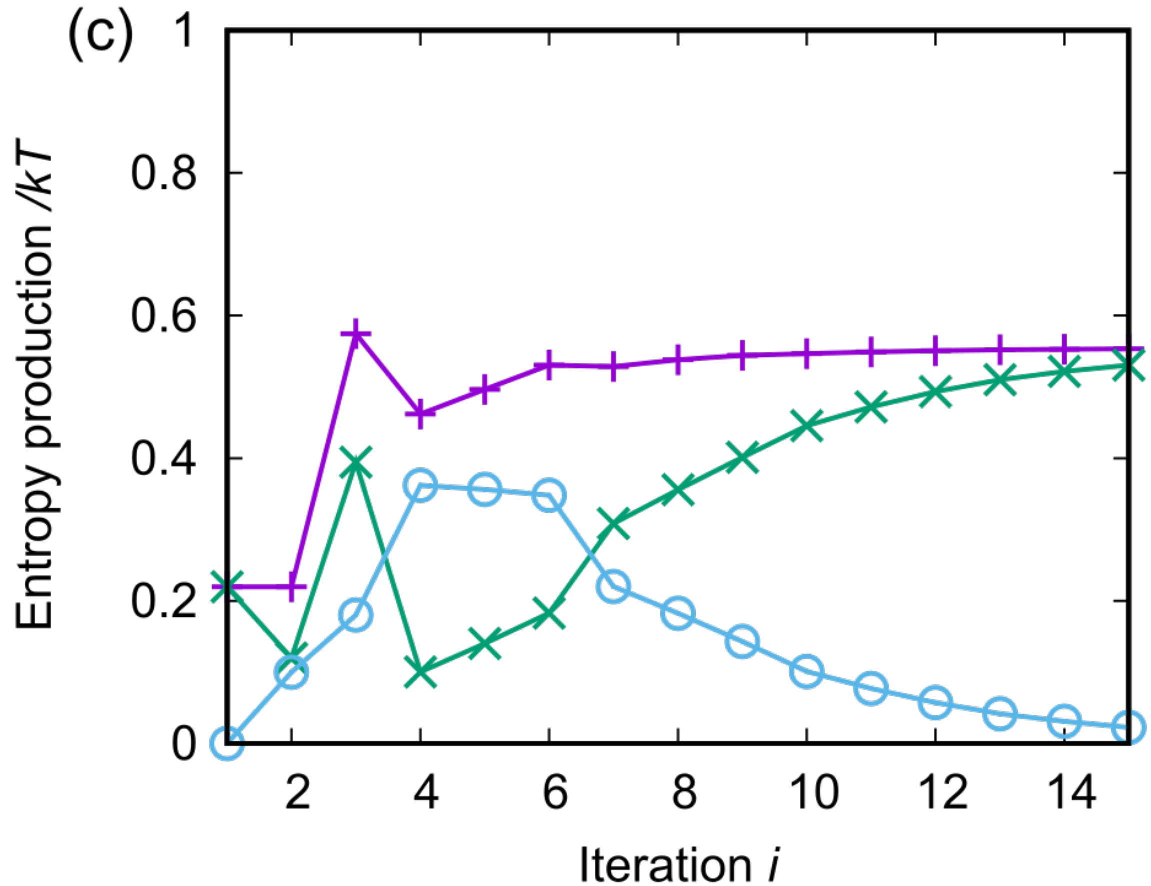




- Input strings have IID symbols with probability of a = 0.8

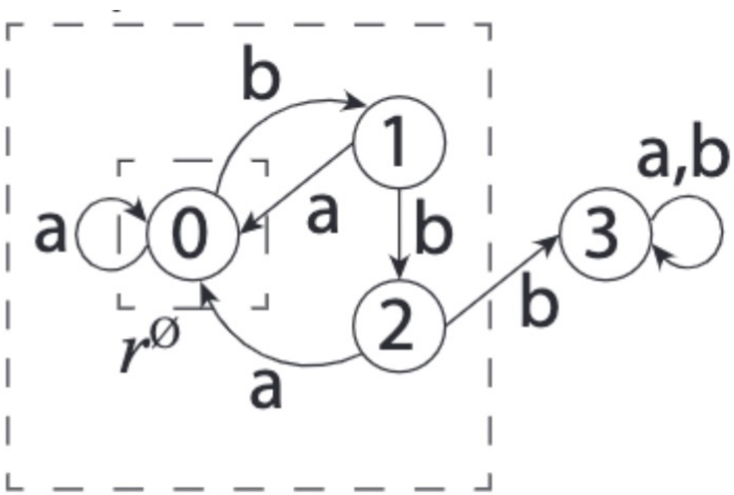
EXAMPLE

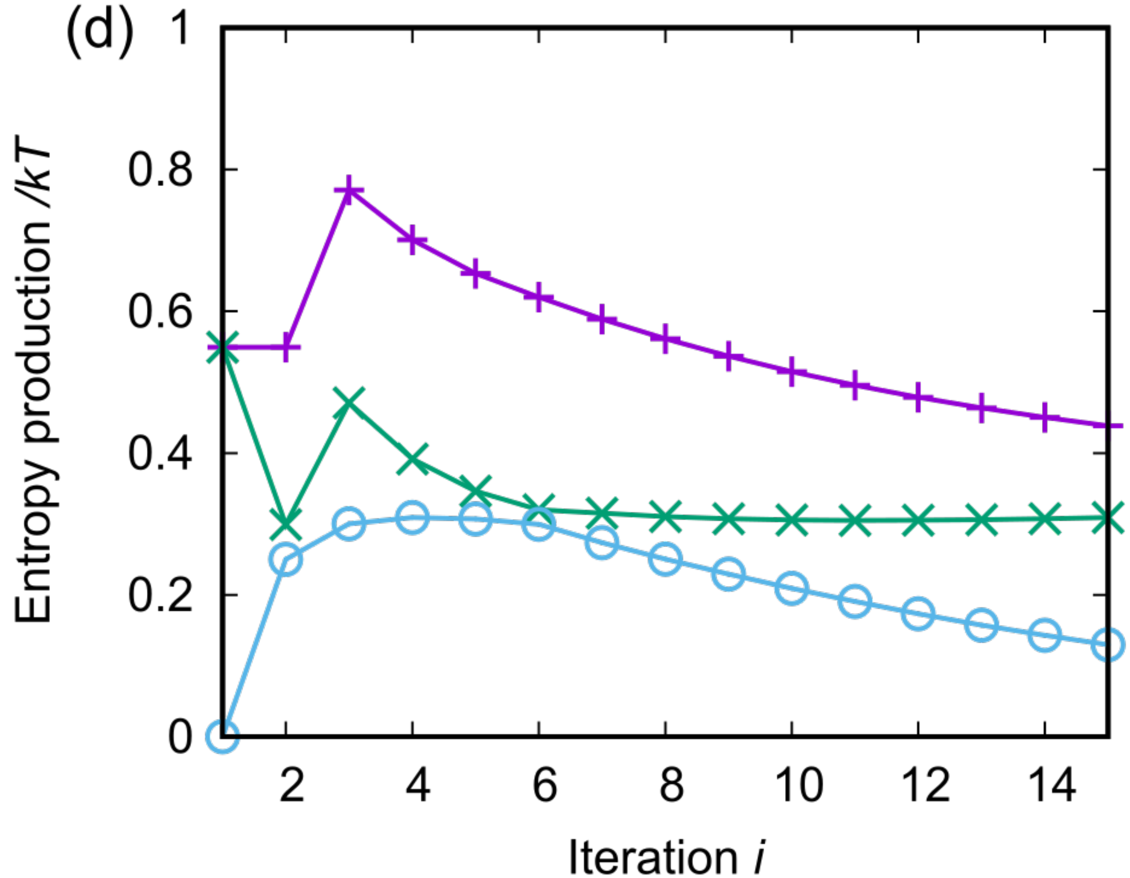




- Input strings have IID symbols with probability of a = 0.2

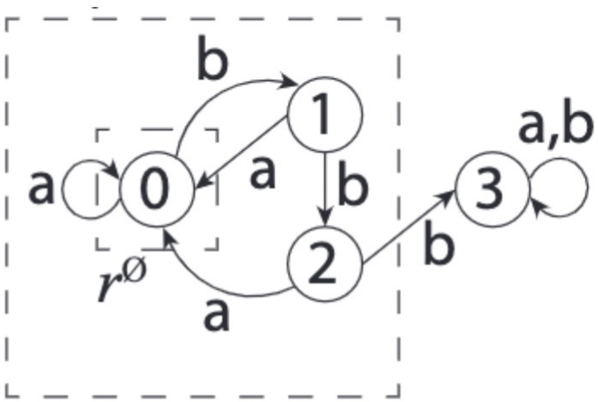
EXAMPLE

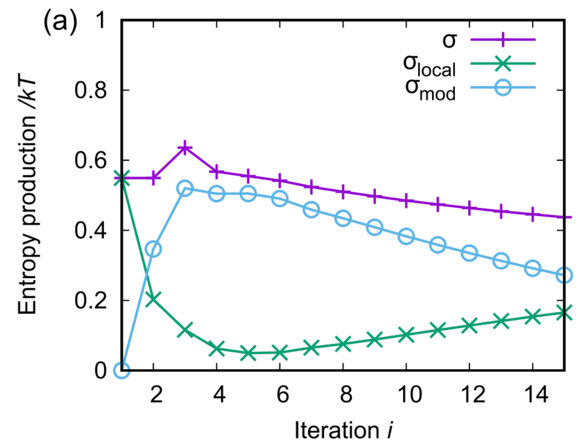


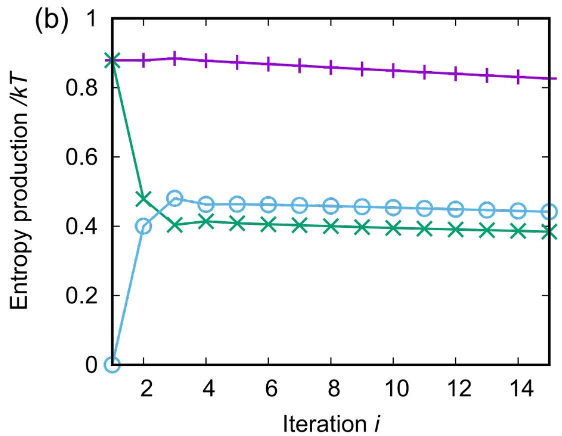


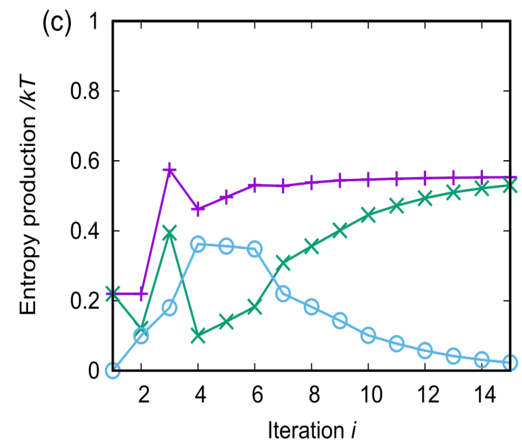
- Input strings are first order Markov chains (starting from uniform probability)

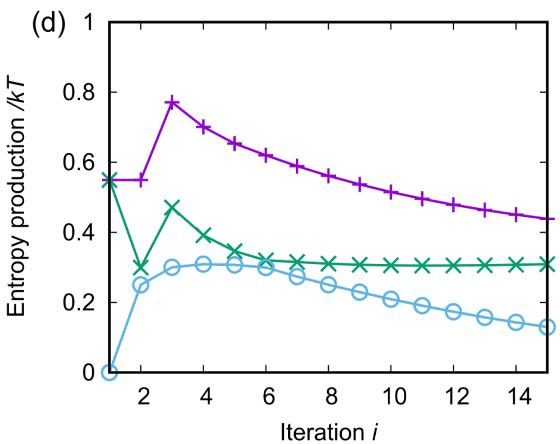
EXAMPLE





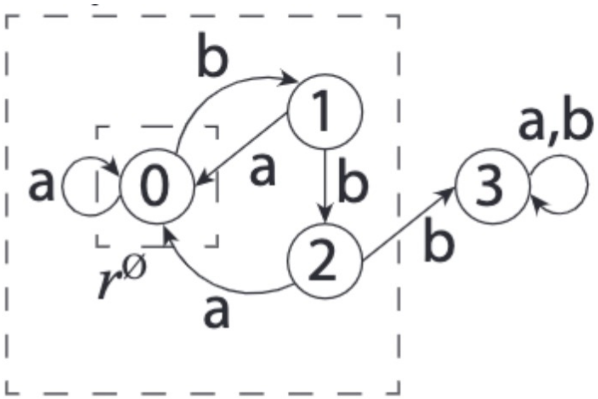


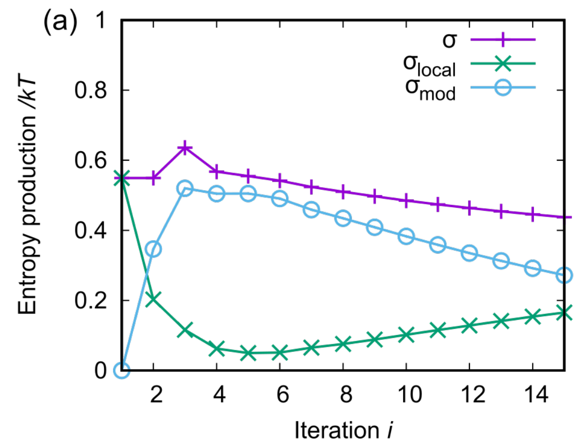


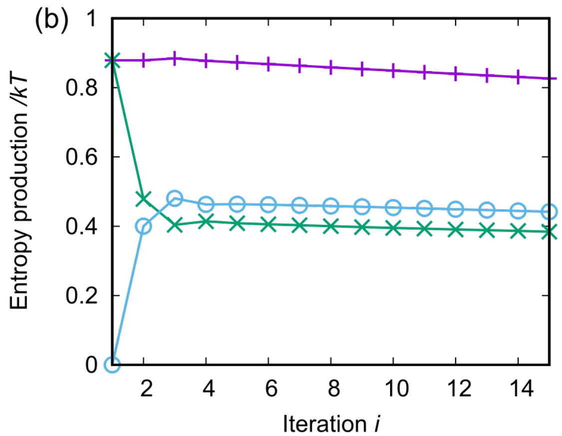


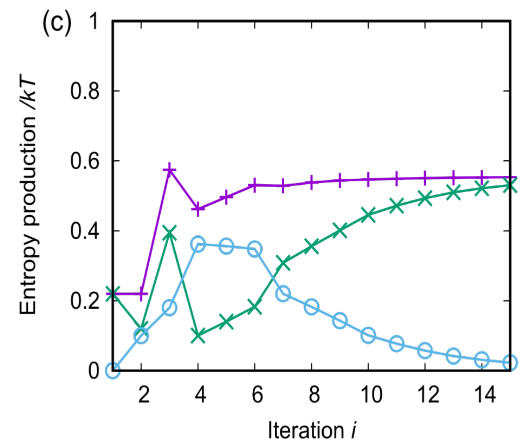
Ouldridge, T., Wolpert, D., arxiv:2208.06895 (2022)

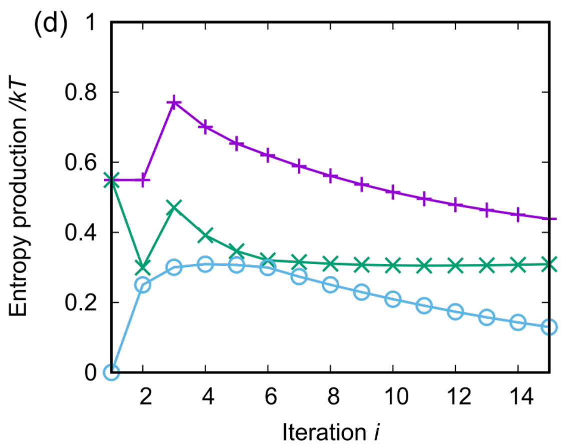
EXAMPLE











What causes curves to have these shapes?
What are curves for other DFAs?

A: Who knows!

BIBLIOGRAPHY (a bit out of date)

Analysis of electronic components used in digital computers

- Riechers, P., in "*The Energetics of Computing in Life and Machines*", Wolpert, D. et al. (Ed.'s), SFI Press (2019)
- Freitas, N., Delvenne, J., Esposito, M., *arxiv:2008.10578* (2021)
- Gao, C., Limmer, D., *arxiv:2102.13067* (2021)
- Boyd, A. Riechers, P., Wimsatt, G., Crutchfield, J., Gu, M., *arxiv:2104.12072* (2021)

Analysis of Turing machines - *based on stochastic thermodynamics*

- Strasberg, P., Cerrillo, J., Schaller, G., Brandes, T., *Phys. Rev. E* (2015)
- Wolpert, D., *J. Phys. A* (2019)
- Kolchinsky, A., Wolpert, D., *Phys. Rev. R* (2020)
- Brittain, R., Jones, N., Ouldridge, T., *arxiv:2102.03388*

Analysis of Turing machines - *not based on stochastic thermodynamics*

- Zurek, W., *Phys. Rev. A* (1989)
  - translation into stochastic thermodynamics in Wolpert, D., *J. Phys. A* (2019)
- Bennett, C., *IBM J. Res. Dev.* (1973)

BIBLIOGRAPHY (a bit out of date)

Analysis of straight-line programs (including Bayes nets)

- Ito, S., Sagawa, T., *Phys. Rev. Letters* (2013)
- Ito, S., Sagawa, T., in "*Mathematical Foundations and Applications of Graph Theory*", Dehmer M., et al. (Ed.'s), Wiley (2015)
- Wolpert, D., *J. Phys. A* (2019)
- Wolpert, D., Kolchinsky, A., *New J. Phys.* (2020)
- Wolpert, D., *Phys. Rev. Letters* (2020)

Analysis of finite state automata (including Mealy machines)

- Ganesh N., Anderson N., *Phys. Lett. A* (2013)
- Chu D., Spinney R., arXiv:1806.04875 (2018)
- Garner A., Thompson J., Vedral V., Gu M., *Phys. Rev. E* (2017)
- Boyd A., Mandal D., Crutchfield J., *New J. Phys.* (2016)
- Boyd A., Mandal D., Crutchfield J., *Phys. Rev. E* (2017)
- Boyd A., Mandal D., Riechers P., Crutchfield J., *Phys. Rev. Lett.* (2017)
- Boyd A., Mandal D., Crutchfield J., *J. Stat. Phys.* (2017)

BIBLIOGRAPHY (a bit out of date)

Analysis of arbitrary asynchronous information processing systems

- Sagawa, T., Ueda, M., *Phys. Rev. Letters* (2009)
- Sagawa, T., Ueda, M., *Phys. Rev. Letters* (2012)
- Sagawa, T., Ueda, M., *New. J. Phys.* (2013)
- Horowitz, J., Esposito, M., *Phys. Rev. X* (2014)
- Barato, A., Hartich, D., Seifert, U., *New. J. Phys.* (2014)
- Horowitz, J., *J. Stat. Mech.: Th. and Exp.* (2015)
- Barato, A., Seifert, U., *New. J. Phys.* (2017)
- Hartich, D., Barato, A., Seifert, U., *Phys. Rev. E* (2016)
- Brittain, R., Jones, N., Ouldridge, T., *J. Stat. Mech.: Th. and Exp.* (2017)
- Kardeş, G., Wolpert, D., *arxiv:2102:01610* (2020)
- Wolpert, D., *arxiv:2003:11144* (2020)
- Wolpert, D., *New J. Phys.* (2020)

BIBLIOGRAPHY (a bit out of date)

Thermodynamic (Ir)relevance of logical reversibility

- Maroney, O., *Phys. Rev. E* (2009)
- Sagawa, T., *J. Stat. Mech.: Th. and Exp.* (2014)
- Wolpert, D., *J. Phys. A* (2019)

Miscellaneous

- Parrondo., J., Horowitz, J., Sagawa, T., *Nature Physics* (2015)
- Sheng, S., Herpich, T. Diana, G., Esposito, M., *Entropy* (2019)
- Wolpert, D., Kempes, C., Stadler, P., Grochow, J., "*The Energetics of Computation in Life and Machines*", SFI Press (2018)
- Grochow, J., Wolpert, D., *ACM SIGACT News* (2018)
- Wolpert, D., Kolchinsky, A., Owen, J., *Nat. Comm.* (2019)
- Owen, J., Kolchinsky, A., Wolpert, D., *New J. Phys.* (2019)
- Riechers, P.. Gu, M., *Phys. Rev. E* (2021)
- Kolchinsky, A., Wolpert D., *arxiv:2103.05734*

BIBLIOGRAPHY (a bit out of date)

Relevant computer science; comp. in biology; comp. in foundations of physics (**very** partial)

- Nielsen, M., *Phys. Rev. Letters* (1997)
- Pour-El, M., Richards, J., *"Computability in Analysis and Physics"* (1997)
- Tegmark, M., *Annals of Phys.* (1998)
- Hut, P., Alford, M., Tegmark, M., *Found. Phys.* (2006)
- Li, M., Vitanyi, P., "An introduction to Kolmogorov Complexity and its applications", Springer (2008)
- Soloveichik D., Cook M., Winfree E., and Bruck J., *Nat. Computing* (2008)
- Tegmark, M., *Found. Phys.* (2008)
- Arora, S., Barak, B., "*Computational Complexity: A modern approach*", CUP (2009)
- Prohoska, S., Stadler, P., Krakauer, D., *J. Theor. Bio.* (2010)
- Barrow, J., in "*Kurt Godel and the Foundations of Mathematics*", Baaz M., et al. (Ed.'s), CUP (2011)
- Aaronson, S., in "*Electronic Colloquium on Computational Complexity"* (2011)
- Qian, L, Winfree, E., *Science* (2011)
- Benenson, Y., *Nat. Rev. Genetics* (2012)
- Cubitt, T., Garcia-Perez, D., Wolf, M., *Nature* (2015)
- Allaghi, A., Hayes, J., *IEEE Trans. CAD of ICs and Systems* (2015)
- Barua, B., Mondal, C.*; J. Inst. Engineers: Series B*, (2019)
- Shiraishi, N., Matsumoto, K., *arxiv:2012.13890*