Why is there a macro-world?

Large-deviations scaling and its connection to macrostates and robustness

Purpose of this lecture

- The concepts of micro-world and macro-world in equilibrium thermodynamics are well known
- Less well-known is how to separate these notions from their historical roots in physical mechanics, and to treat them in a pure form
- Goal today: emphasize the concepts of micro and macro, and what we know about emergence in these terms
- We can then readily extend to non-equilibrium, problems of inference, and much more

The difference between "micro-worlds" and "macro-worlds"

- (For the purpose of this talk) . . .

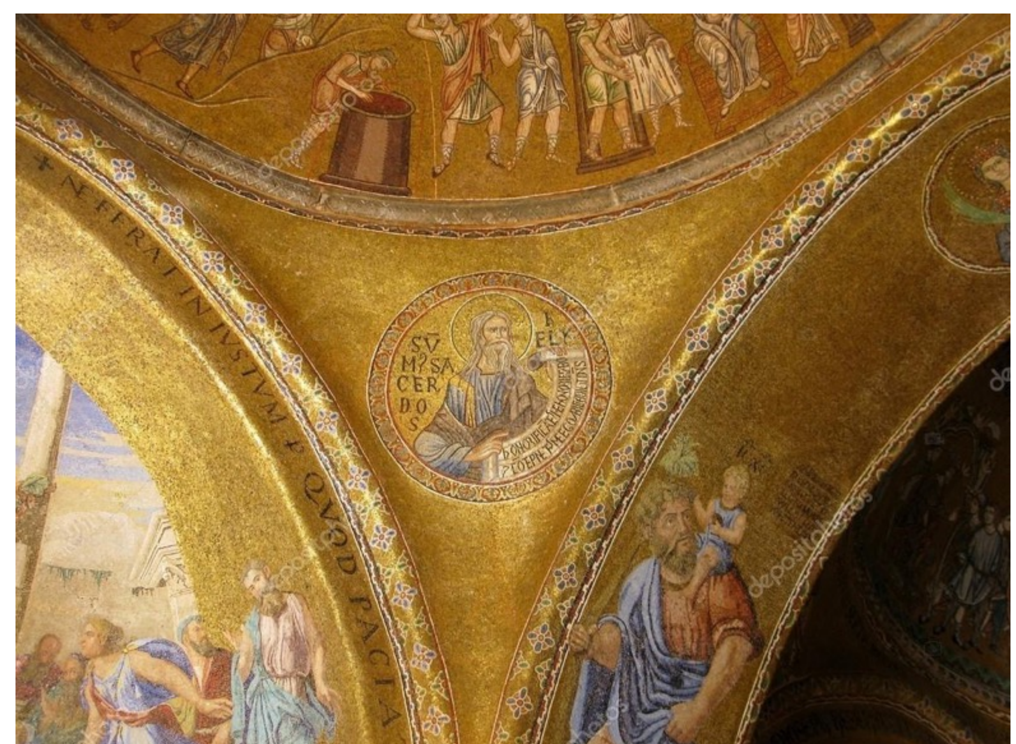
the concept of a micro-world isn't about being *small*,
it is about being ***particular***
in reference to the parts and their assembly

In Mechanics

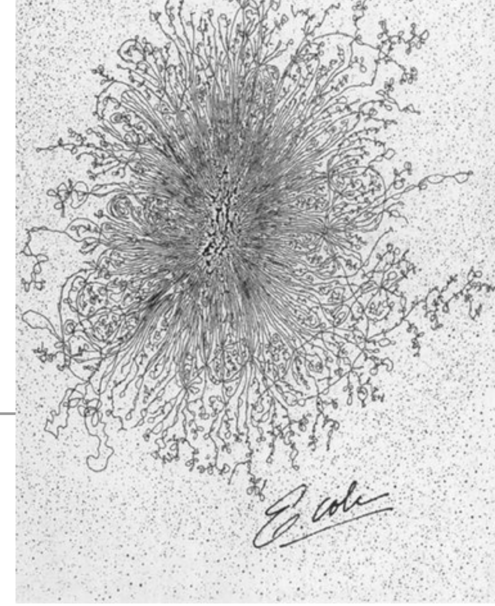
- "Newtonian Clockwork"

Super-complex design/debugging

- The Spandrels of San Marco



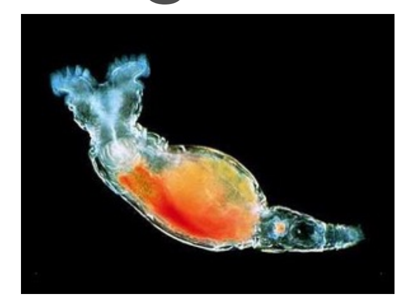
In Biology



microbewiki.kenyon.edu

- Bacterial cell:
  1 genome, small-numbers regulatory enzymes

- 1-2-4-6 ploidy qualitatively changes how organisms function and how they evolve



- C. elegans (herm) 959 cells (male) 1031

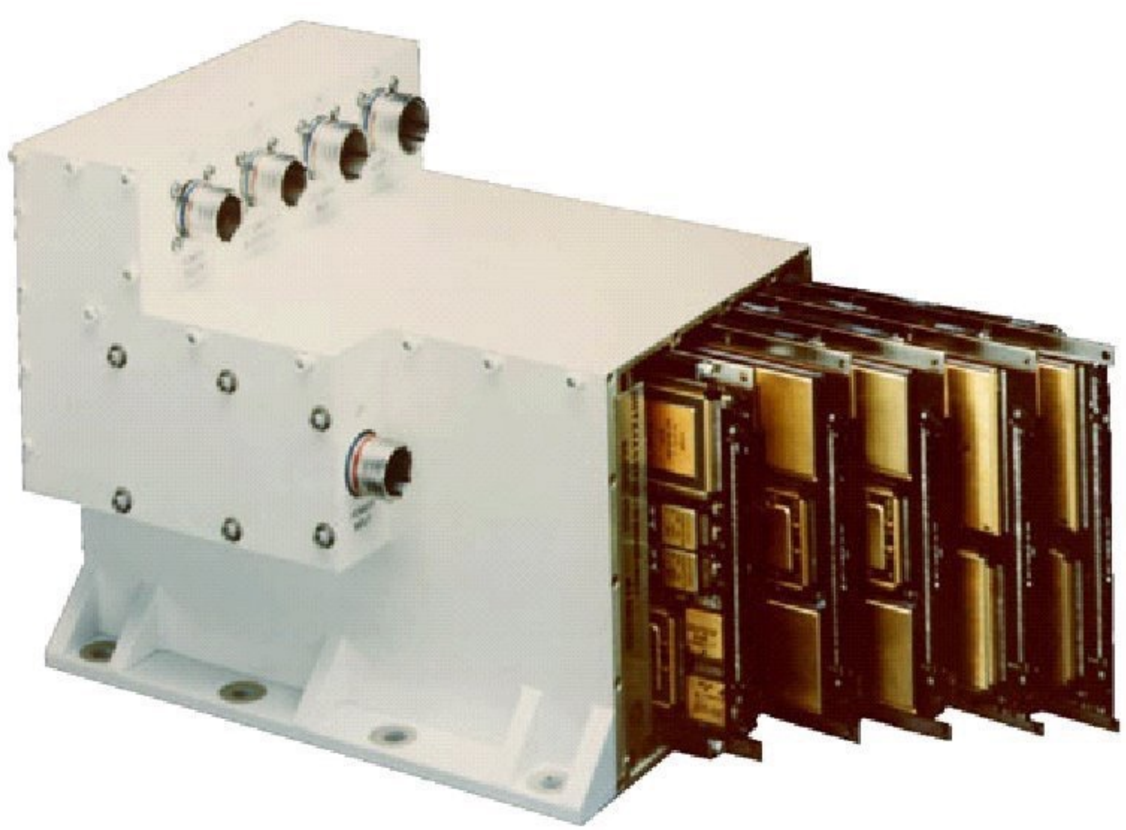
- n-pods

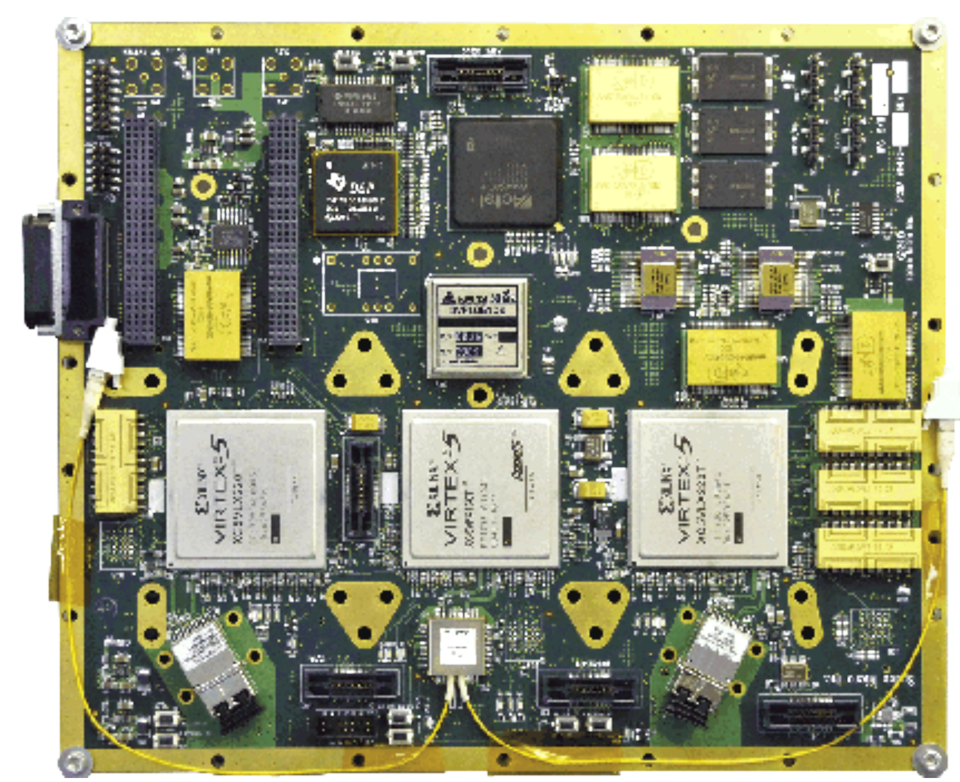
- Reliable translation (1 gene -> 1 enzyme)

In Communication and Computation

- Design errors in specific relations
- Price and energy devoted to correction against operations errors

Example: 8-Slot Flight Chassis



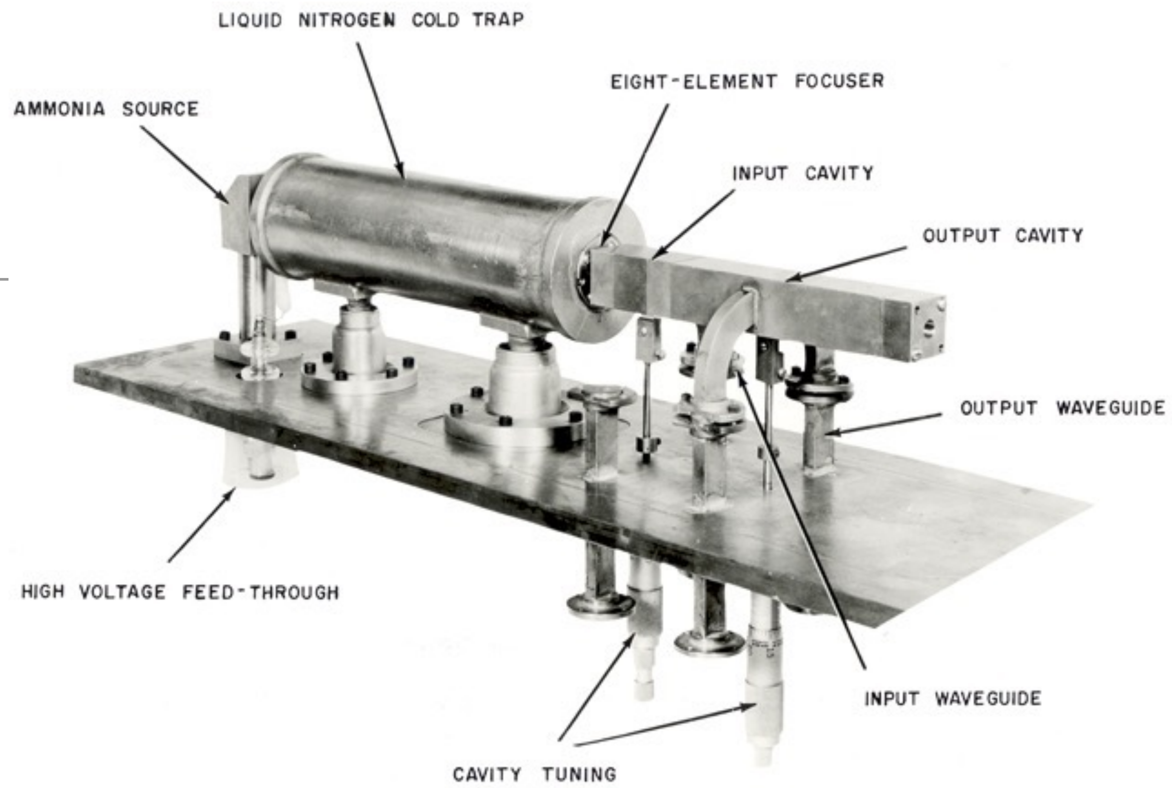


Macro-worlds come from indefiniteness

- The thing that makes a macro-world “macro” is indefiniteness: the ability to take well-defined structure, dynamics, or properties with an indefinite composition of either counts or detailed relations among components
- We should not take it for granted that macro-ness is even possible
- With some experience, it should become striking that the actually-realized macro-worlds are rare compared to their possibilities, even moreso than for micro-worlds

In Mechanics or Materials

- Ammonia MASER



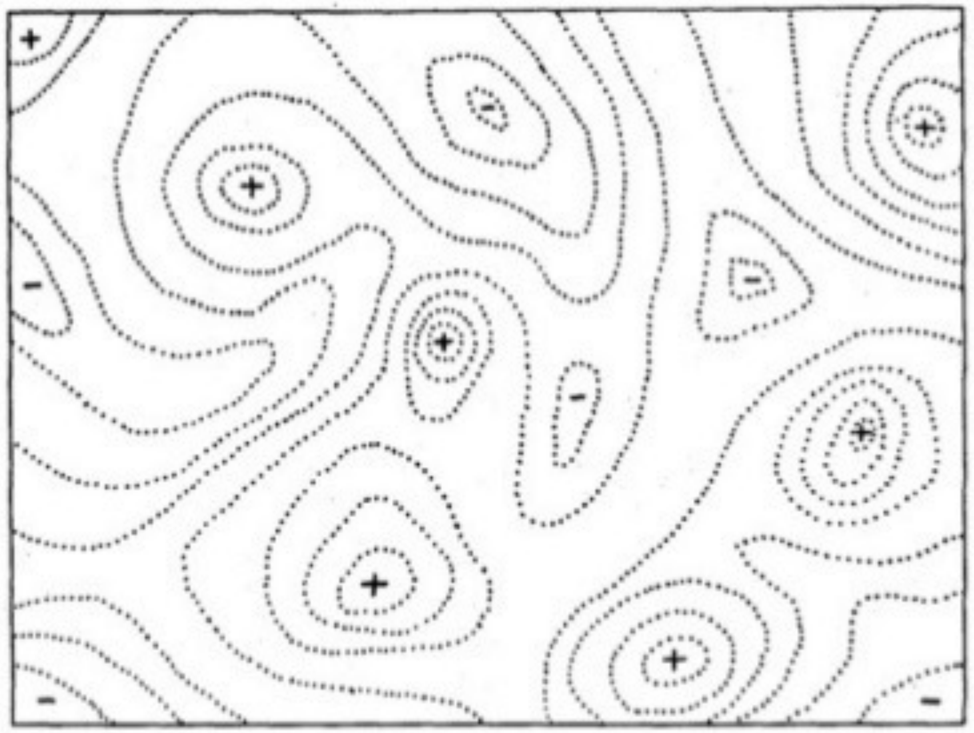
- Materials properties

- Working fluids etc.

In Evolutionary Dynamics

- Genetics concept of an adapted population is inherently a statistical one

www.slideshare.net/eserrelli

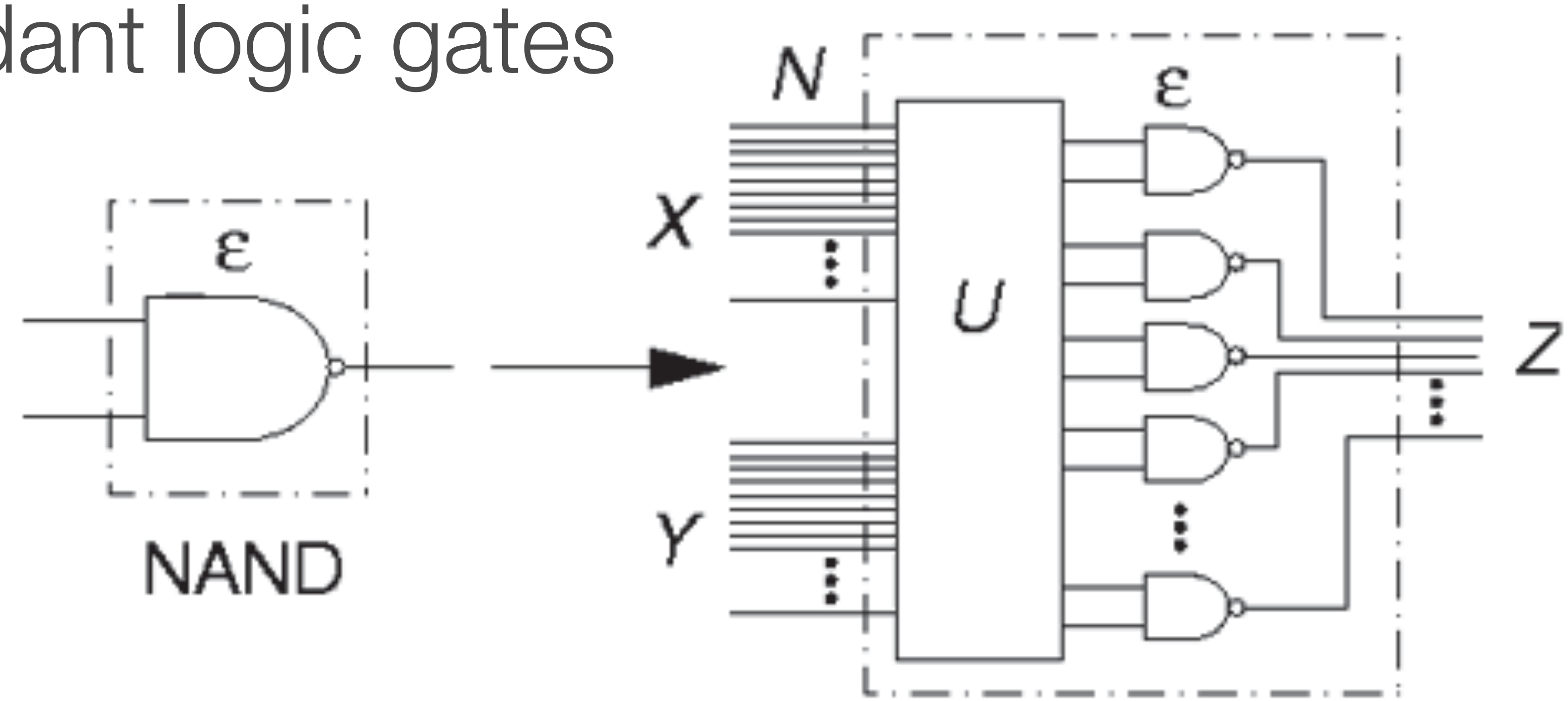


First appearance of the adaptive landscape: Sewall Wright (1932) "The roles of mutation, inbreeding, crossbreeding and selection in evolution", communication at 6th International Congress of Genetics, Cornell University

In Communication and Computation

- Error-correcting codes
  (*e.g.* Reed-Solomon on CD/DVD)

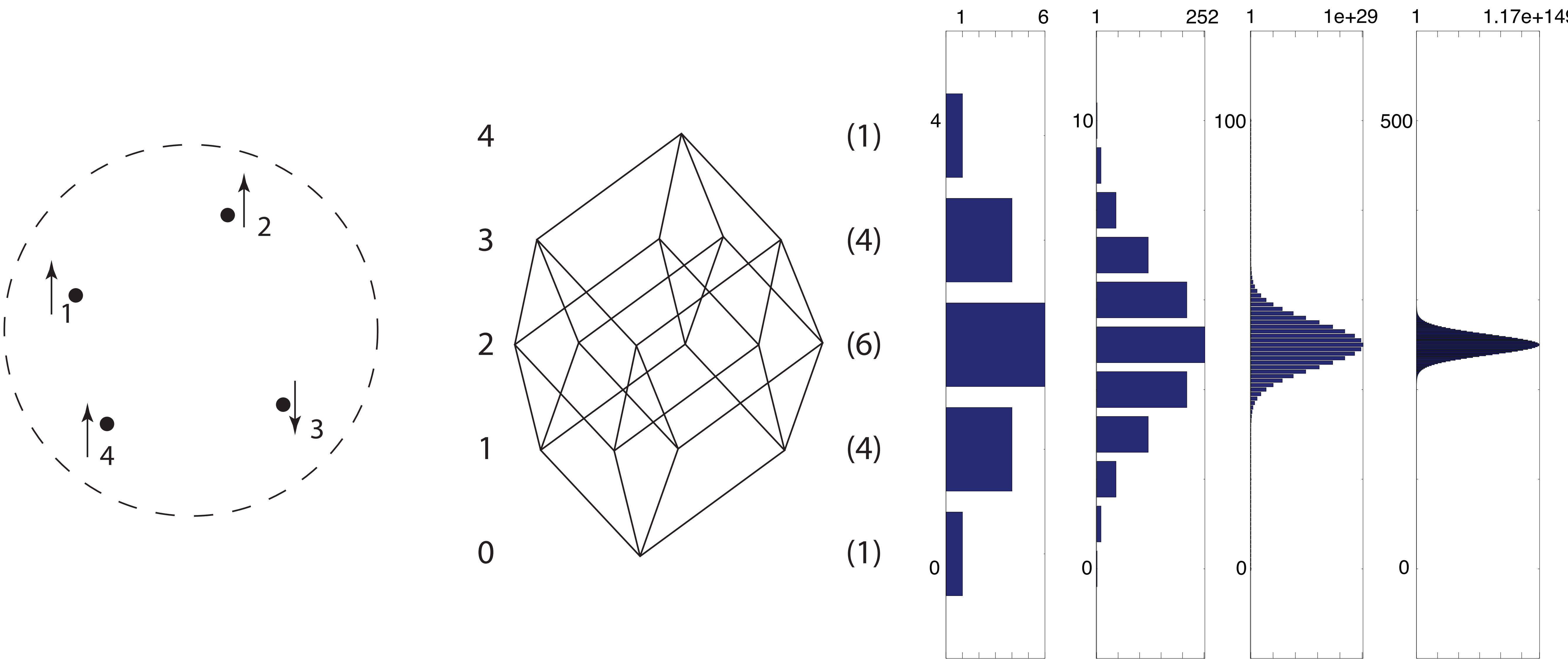
- von Neumann redundant logic gates



Introduction to Large-deviations scaling and the entropy

- Some examples
- How the entropy comes from counting and large-deviations scaling
- The general characteristics shared by large-deviations limits

Binomial distributions and the convergence to Gaussian limiting distributions



$$\mathrm{count}(\mathrm{n} \mid N) = \binom{N}{\mathrm{n}} \equiv \frac{N!}{\mathrm{n}!\,(N-\mathrm{n})!} \approx 2^N \sqrt{\frac{2}{\pi N}}\, e^{-2N\left(\frac{\mathrm{n}}{N}-\frac{1}{2}\right)^2}$$

The multinomial, Stirling's formula, and the Gibbs-Shannon entropy

- Stirling approximation for logs: $\log n! \approx n\left(\log n - 1\right)$
- Multinomial for counting:

$$\begin{pmatrix} N \\ n_1 \ldots n_P \end{pmatrix} \equiv \frac{N!}{n_1! \ldots n_P!} \approx e^{NS\left(\frac{n_1}{N},\ldots\frac{n_P}{N}\right)}$$

- Shannon entropy: $S\left(\frac{n_1}{N}, \ldots \frac{n_P}{N}\right) = -\sum_{p=1}^{P} \frac{n_p}{N} \log \frac{n_p}{N}$

notice: $\max\left(S\left(\frac{n_1}{N}, \ldots \frac{n_P}{N}\right)\right) = \log P$

so:

$$\frac{1}{P^N} \begin{pmatrix} N \\ n_1 \ldots n_P \end{pmatrix} \approx e^{-N\left[S_{\max} - S\left(\frac{n_1}{N},\ldots\frac{n_P}{N}\right)\right]}$$

The general characterization of Large-Deviations scaling

- The separation of scale from structure in fluctuation probabilities under some process of aggregation

scale structure

$$P_{\text{fluct}} \sim e^{-NS}$$

A good introduction:
H. Touchette, The large deviation approach to statistical mechanics, Phys. Rep. 478, 1–69, 2009.
arxiv:0804.0327

- Recall the binomial and its Gaussian limit

$$\text{count}(n \mid N) = \begin{pmatrix} N \\ n \end{pmatrix} \approx 2^N \sqrt{\frac{2}{\pi N}} e^{-2N\left(\frac{n}{N} - \frac{1}{2}\right)^2}$$

scale structure

The two main claims for this lecture:
Large-Deviations Scaling (LDS) and macro-worlds

- The ***kind*** of separation of scale from structure seen in LDS is the concept we need to characterize macro-worlds

- Many *(all?)* real instances of the emergence of macro-worlds are formalized as cases of LDS

How aggregate fluctuation laws give rise to classical thermodynamics

- Extensive state variables: origin, roles, and meaning
- The equation of state starting from the right end
- How fluctuation theorems recover classical thermodynamic laws
- Why historically we started with conservation of energy, but why conceptually it is not the best way

Extensive State Variables: scalable arguments to the Entropy function

- The values of limiting quantities determine what “volume” or “measure” of states can be reached by a system

- When the states are achieved as combinations of elements made available by ***U, V, N,*** then state space volume will scale exponentially in these constraints, and entropy will scale in the same way as they do, defining extensivity

- We write, then: $S(U, V, N)$

Roles of Extensive State Variables: Constraints of possibility and boundary interfaces

- Suppose multiple systems must share something:

$$U = U_1 + U_2$$
$$V = V_1 + V_2$$
$$N = N_1 + N_2, \quad etc.$$

- Since, for each system, $S_i = S(U_i, V_i, N_i)$

- The extensive state variables that are constraints also become interface properties at system boundaries

Intensive State Variables: the gradients of the Entropy

- The most-likely fluctuation is (in this language, by construction) the one to a state that maximizes total entropy: $P_{\text{fluct}} \sim e^{-NS}$

- In multi-part systems, what matters is how the constraints of sharing impose trade-offs in entropy; for each *i*, we simply define these gradients to have names:

$$\left.\frac{\partial S}{\partial U}\right|_{V,N} \equiv \frac{1}{T}$$

$$\left.\frac{\partial S}{\partial V}\right|_{U,N} \equiv \frac{p}{T}$$

$$\left.\frac{\partial S}{\partial N}\right|_{U,V} \equiv -\frac{\mu}{T}$$

- The dependence of S is then just a notation:

$$\delta S \equiv \frac{1}{T}\delta U + \frac{p}{T}\delta V - \frac{\mu}{T}\delta N$$

Entropy maximization across a shared quantity equalizes intensive state variables

- Max total entropy: $\delta\left(S_1 + S_2\right) = 0$
- Each entropy is constrained: $= \frac{\partial S_1}{\partial U_1}\delta U_1 + \frac{\partial S_2}{\partial U_2}\delta U_2 + \ldots$
- But they must share: $= \left[\frac{\partial S_1}{\partial U_1} - \frac{\partial S_2}{\partial U_2}\right]\delta U_1 + \ldots$
- And these have names: $= \left[\frac{1}{T_1} - \frac{1}{T_2}\right]\delta U_1 + \ldots$

All this is standard and familiar, but it sets up a frame . . .

From constraints to independence, in 2 steps:
I: maximizing against a constraint

- Maximizing entropy subject to a constraint (general):

$$\mathcal{L} = S - \beta \left(\sum_{i=1}^{P} p_i H_i - U \right) - \eta \left(\sum_{i=1}^{P} p_i - 1 \right)$$

- Resulting probabilities are just exponential:

$$\bar{p}_i = e^{-\left[1+\bar{\eta}+\bar{\beta} H_i\right]} = \frac{1}{\bar{Z}} e^{-\bar{\beta} H_i}$$

- The sum *(Z)* is related to the entropy and the constraint:

$$\bar{Z} = -\bar{\beta}\bar{F} = \bar{S} - \bar{\beta} U$$

From constraints to independence, in 2 steps: II: independence of internal variables given the bdry

- A marginal large-deviations function only for internal fluctuations of probability is obtained by fixing boundary values of the extensive state variables

$$\mathcal{L} - \bar{\beta} U - \log \bar{Z} = - \sum_{i=1}^{P} p_i \log \left(\frac{p_i}{\bar{p}_i} \right) - \left(\sum_{i=1}^{P} p_i - 1 \right)$$

- But remember Stirling's formula and the factorials: This says internal fluctuations behave like random samplex conditioned only on the boundary constraint

$$P(\{n_i\}) \sim e^{-N\left(\mathcal{L} - \bar{S}\right)} \approx \left(\prod_{i=1}^{P} \bar{p}_i^{n_i} \right) \binom{N}{n_1 \dots n_p}$$

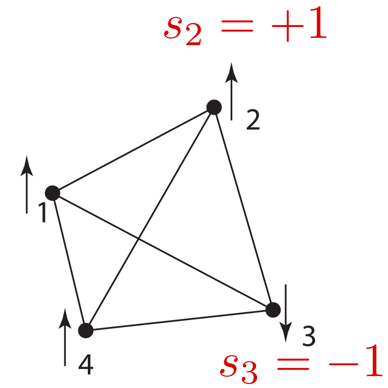
$$\sum_{i=1}^{P} p_i = \sum_{i=1}^{P} \frac{n_i}{N} = 1$$

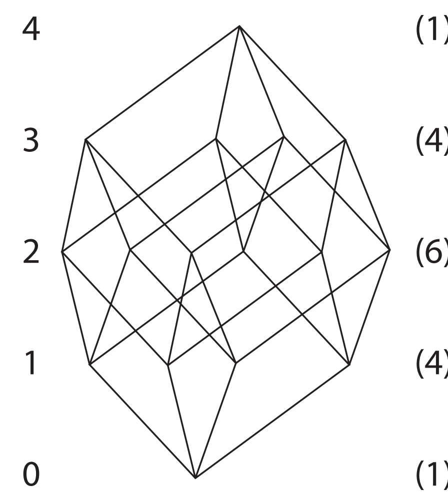
The concept of phase transitions

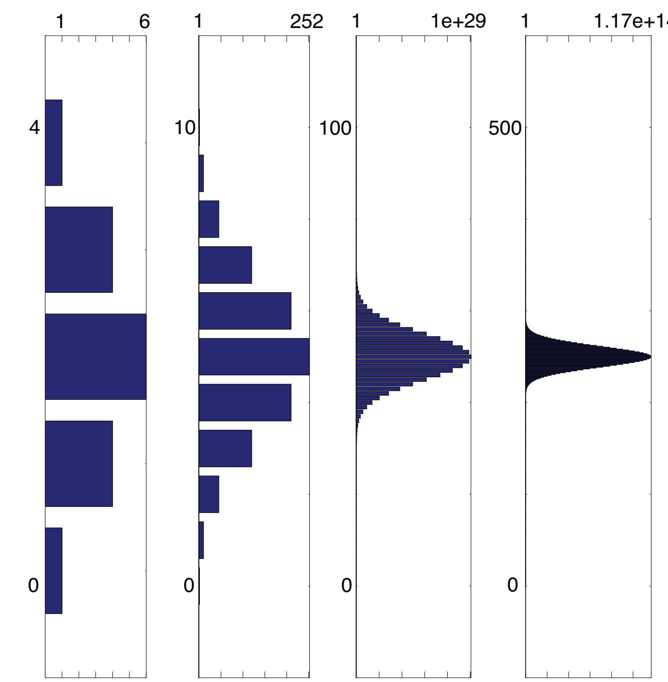
- Changes in the part of the “structure” factor in LD scaling that controls macrostates
- Competitions for entropy induce phase transitions
- The idea of the order parameter and its role with respect to information

Example: a simple model of a magnet

- The counting (combinatorial) factor comes only from spin configurations







- Each configuration requires energy to reach; aligned configurations require less

$$H = -\frac{\epsilon}{2} - \frac{\epsilon}{N} \sum_{\langle i,j \rangle} s_i s_j - h \sum_{i=1}^{N} s_i$$

- Configurations requiring more energy are exponentially more improbable (we will say why in a moment)

$$p(\vec{s}) \propto e^{-H(\vec{s})/k_B T}$$

Entropy all the way down: thermal probability weight factors are just Large-Deviation probabilities

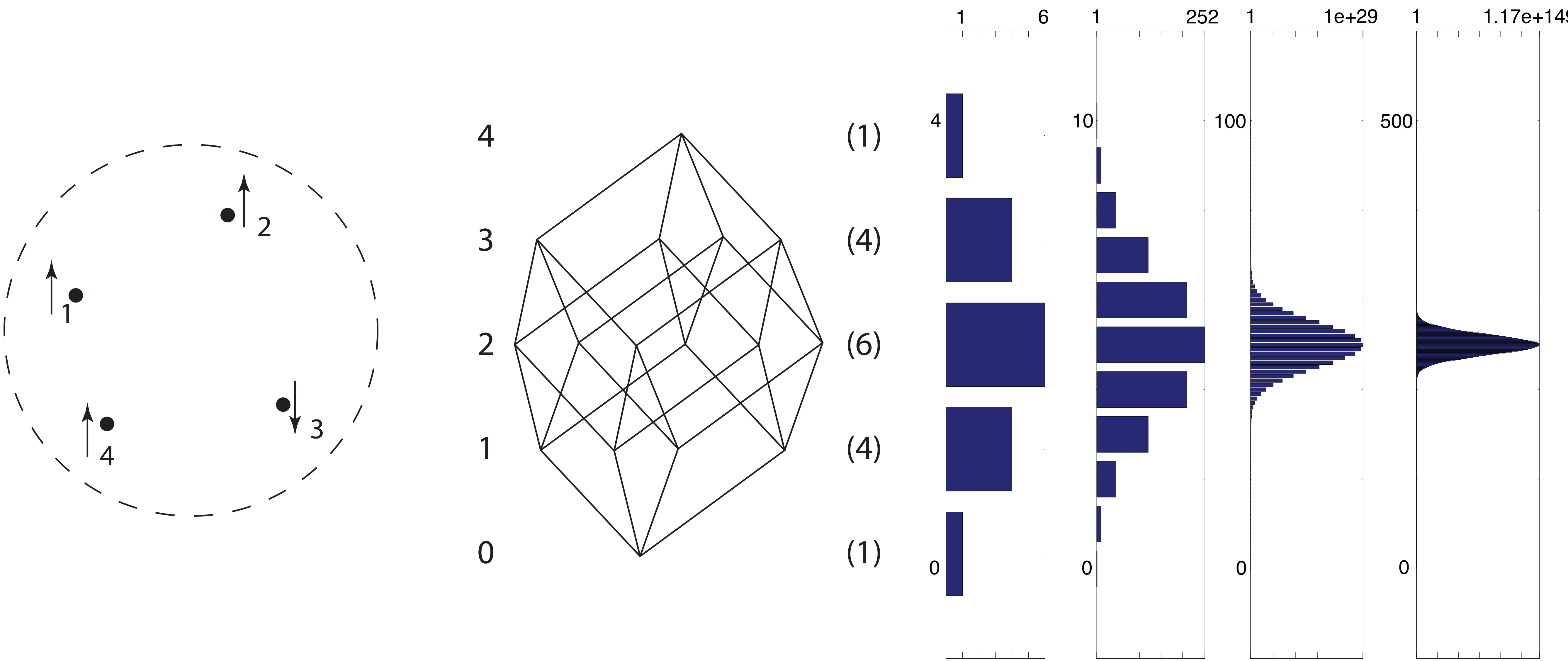
- The exponential weight function (called a “Boltzmann factor”) itself is just a large-deviation function for fluctuations of the surroundings to give up enough energy that the system can reach a particular state

$$p(\vec{s}) \propto e^{-H(\vec{s})/k_B T}$$

- Resulting probability density competes combinatorics in the system’s and the environment’s states, connected by the constraint of energy conservation

Will walk you through how that happens, and introduce the idea of thermodynamic effective potentials

Remember the large-deviations form we computed for the combinatoric part of the two-state system



I will call this the "system" Large-Deviation function in the next slide

$$\mathrm{count}(n \mid N) = \binom{N}{n} \approx 2^N \sqrt{\frac{2}{\pi N}}\, e^{-2N\left(\frac{n}{N}-\frac{1}{2}\right)^2}$$

scale **structure**

First notice how we separate scale from structure in defining aggregate variables

- Remember that the microscopic Hamiltonian was:

$$H = -\frac{\epsilon}{2} - \frac{\epsilon}{N} \sum_{\langle i,j \rangle} s_i s_j - h \sum_{i=1}^{N} s_i$$

- Introduce descaled counting variables:

(structure)

$$x \equiv 2\frac{n}{N} - 1$$

magnetization (relative to system size)

- Then the Hamiltonian separates scale/structure

$$H(\vec{s}) = -N \left(hx + \frac{\epsilon}{2} x^2 \right)$$

scale **structure**

Structure changes are governed by a Large-Deviation function called the Effective Potential Φ

$$x \equiv 2\frac{n}{N} - 1$$

(structure)

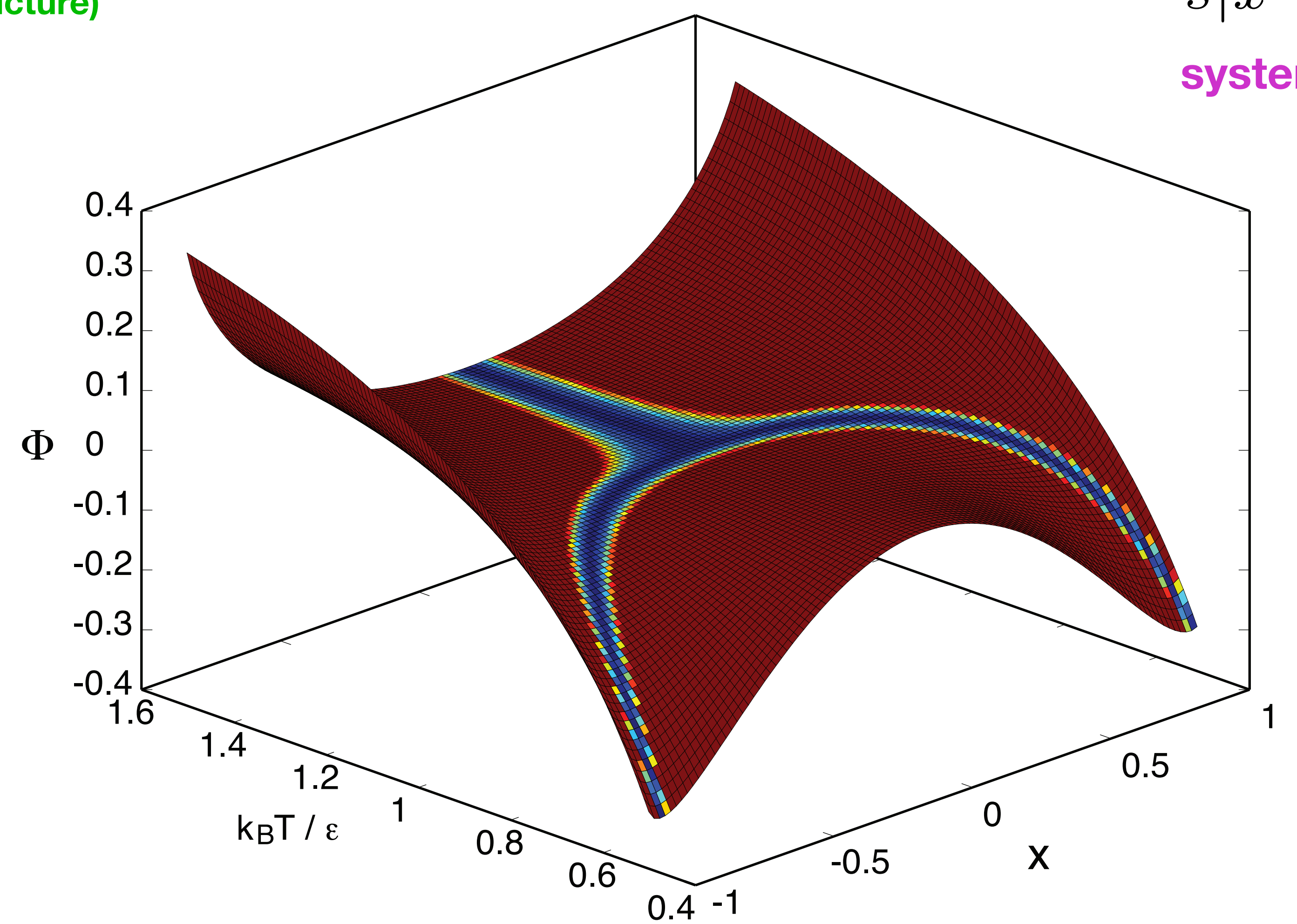
$$H(\vec{s}) = -N\left(hx + \frac{\epsilon}{2}x^2\right)$$

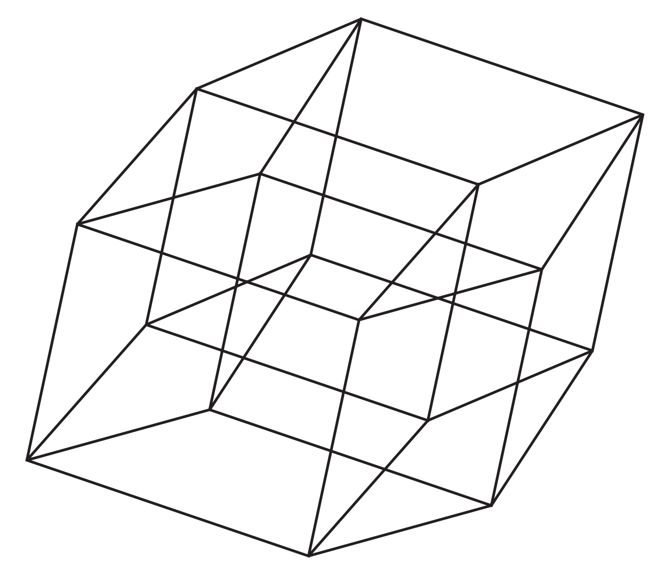
scale structure

$$\sum_{\vec{s}|x} p_0(\vec{s})\, e^{-H(\vec{s})/k_BT} \equiv e^{-N\Phi(x)}$$

scale structure

system LDP environment LDP eff. potential LDP





- Competition between system and environment combinatorics for the distribution of energy

$$\Phi(x) \approx \frac{1}{2}\left[-2\frac{hx}{k_BT} + \left(1 - \frac{\epsilon}{k_BT}\right)x^2 + \frac{x^4}{6}\right]$$

system LDP environment LDP

Order Parameters, Sufficient Statistics, and Inference

- An *order parameter* is a property of structure in a macrostate brought into existence by a phase transition
- The macro-world is built from the order parameters of robust states
- The “systems” in which order parameters are organized (groups, geometries, etc.) characterize kinds of order
- A suggestion: the emergence of individuality in biogenesis was (1 or more) phase transition(s) for which the order parameters brought into existence are *proper names*

(Fun: Bertrand Russell, “On Denoting” 1905 *Mind* 14:479–493)

An example: what new kind of information exists when the magnet goes into an ordered phase?

Probability for some particular spin to have value +1

$$P(s_1 = 1 \mid n) = \binom{N-1}{n-1} / \binom{N}{n} = \frac{n}{N}$$

Probability for two chosen spins each to have value +1

$$P(s_1 = 1, s_2 = 1 \mid n) = \binom{N-2}{n-2} / \binom{N}{n} = \frac{n\,(n-1)}{N\,(N-1)}$$

$$= P(s_1 = 1, \mid n)\, P(s_2 = 1, \mid n) \left\{ 1 + \mathcal{O}\left(\frac{1}{N}\right) \right\}$$

Magnetization as a measure of the state of order: $n = (\mu + N)/2$

Details become independent of each other beyond their dependence on the mean

$$P(s_1 = 1, s_2 = 1 \mid \mu) \rightarrow P(s_1 = 1, \mid \mu)\, P(s_2 = 1, \mid \mu) \left\{ 1 + \mathcal{O}\left(\frac{1}{N}\right) \right\}$$

Asymptotically optimal error correction is a large-deviations result

- Concepts from block encoding: error probability, message expansion factor, and correlation length

- The Shannon proof of asymptotically reliable message transmission is a large-deviation result

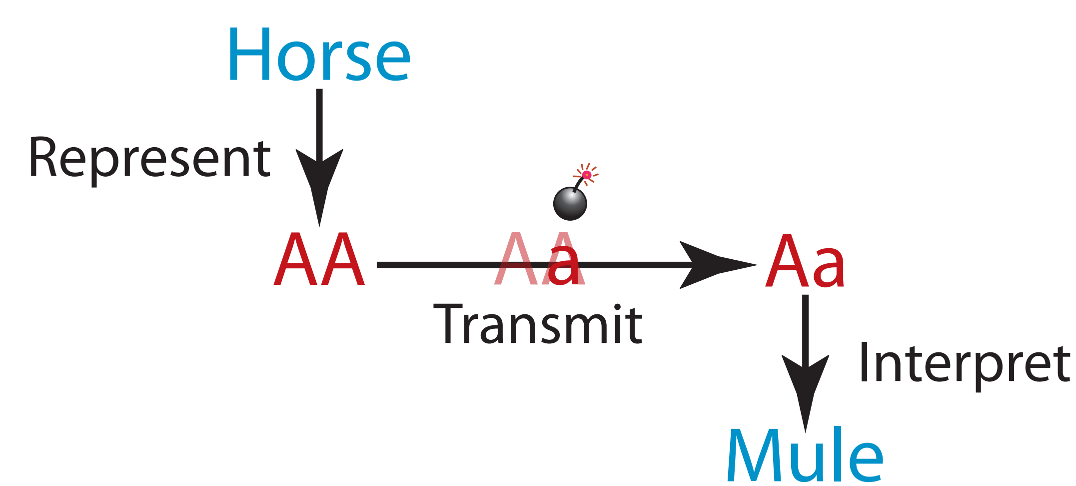
Being ordered & Correcting errors: Two names for the same thing

Block encoding and asymptotically reliable communication in the presence of noise

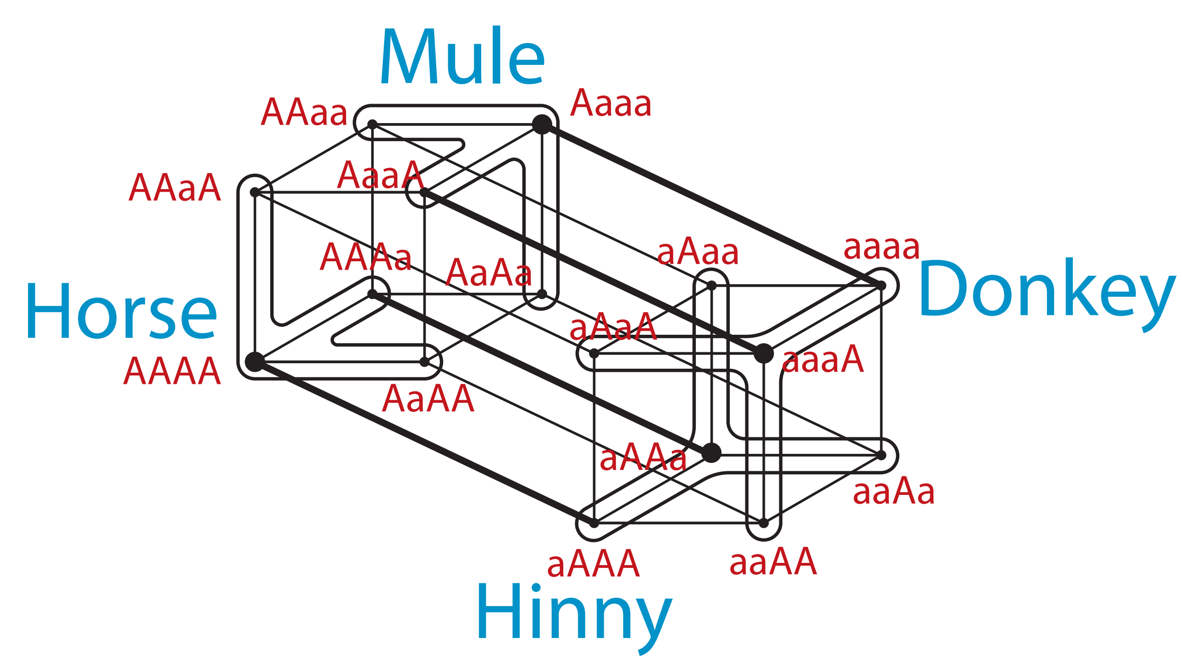
Suppose we want to send any of 4 messages:

Horse:	AA
Mule:	Aa
Hinny:	aA
Donkey:	aa

- Without error correction:
    - Can encode them as two "bits" sent in a sequence
    - Under transmission with noise, bit errors produce message errors



- Adding redundancy to reduce errors:
    - Same messages, *encoded* by an expansion of 2 bits to 4



The Intensive and Extensive state variables of a block code

- Two properties of our block code can take the same form for any block length:
    - The flip probabiilty per symbol
    - The expansion factor (how many bits used per letter)
- One property defines a "scale" of the code:
    - The length of encoding blocks
- These are respectively **intensive** and **extensive** variables

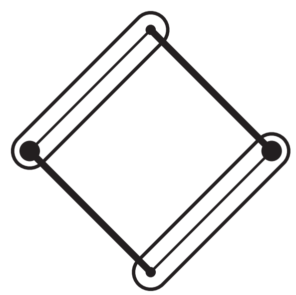
The Shannon sphere-packing proof of Asymptotically Reliable Error Correction (AREC) at finite noise

- Two steps to understand:
  - Why increasing block length at fixed expansion factor can decrease the probability of error
  - Why this has a limit of perfect reliability at infinite block length but *fixed* expansion factor

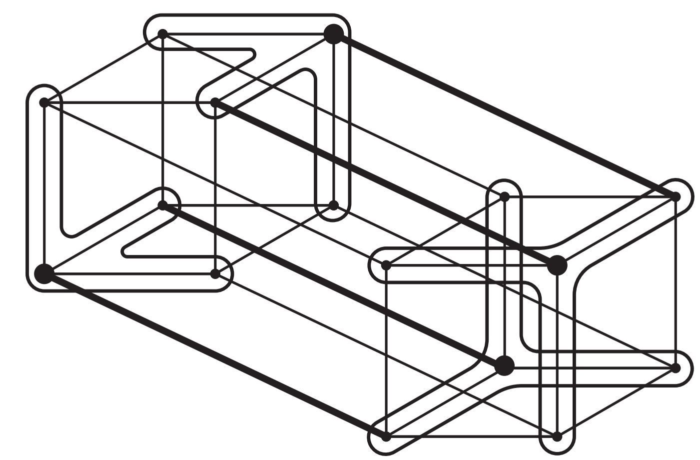
Concept of correlation length in a code: filtering un-caught residuals in longer blocks

- In a message represented as a sequence of letters, I can send a 2x-expanded message for bit-blocks of increasing length

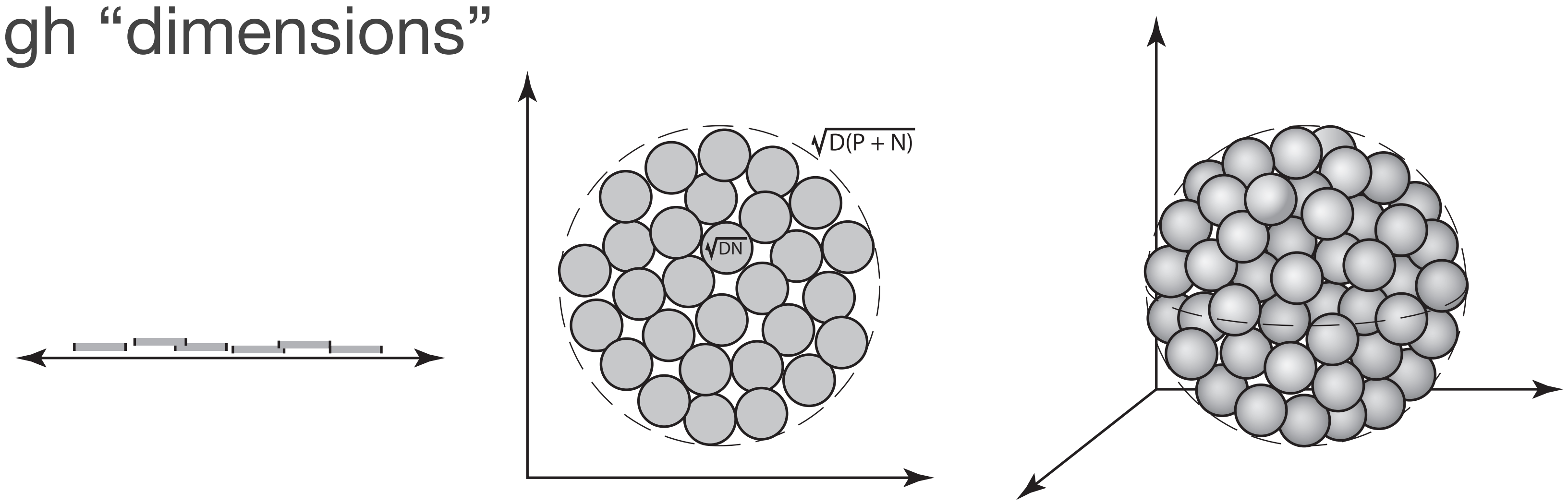
$1 \rightarrow 2$



$2 \rightarrow 4$



- Continuous-valued versions of the same idea exist, using sphere-packing in increasingly high "dimensions"



D = 1   D = 2   D = 3   . . .

The concept of the "capacity" of a channel: Example using the Shannon sphere-packing code

- How many message spheres can you fit within the power budget of your signal, leaving room for one noise-ball around each message?

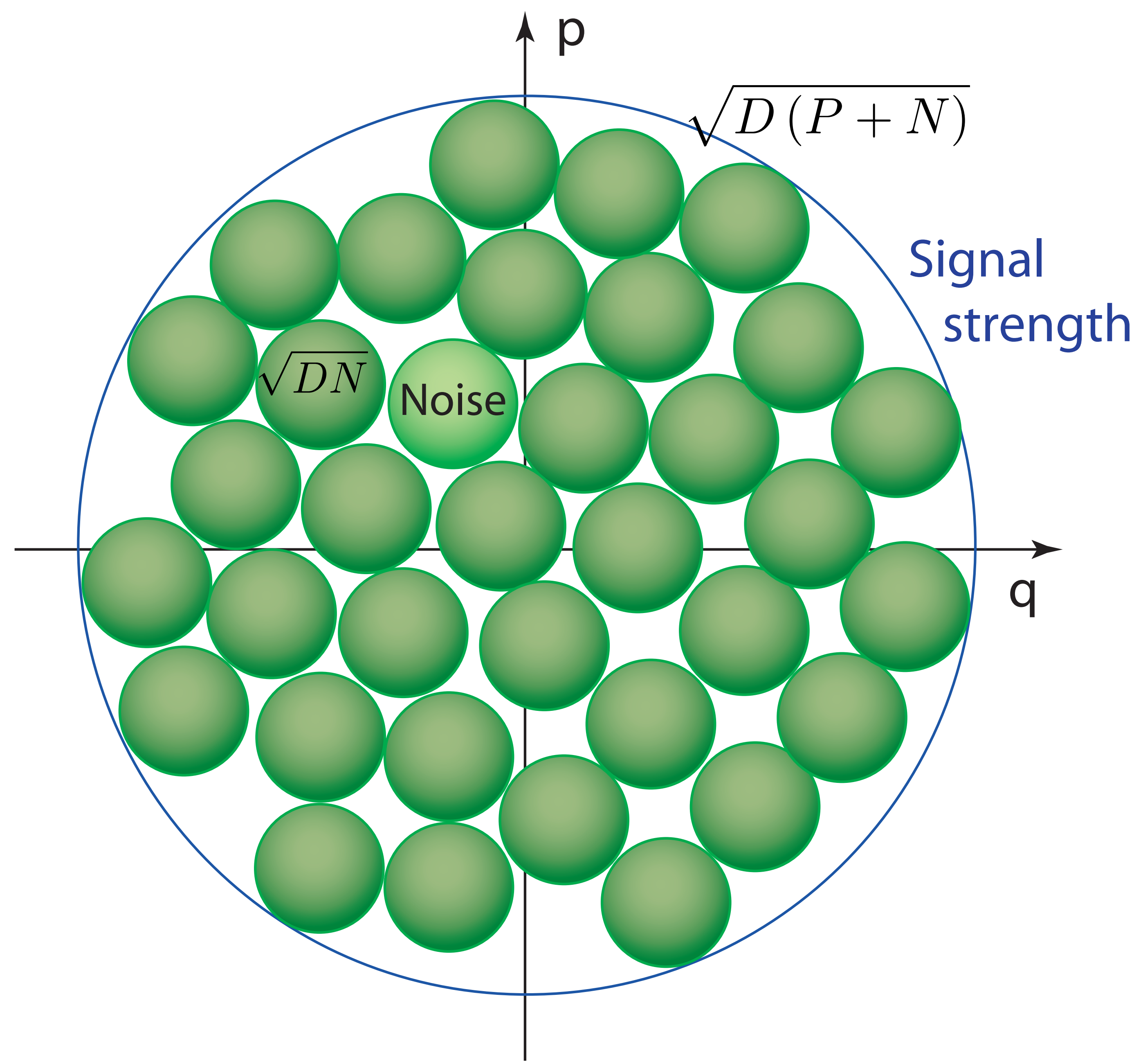
Fill D-bit code space with maximally distant spheres

$$\frac{[D(P+N)]^D}{[DN]^D} \sim \left(\frac{P+N}{N}\right)^D$$

$$= e^{D\log\left(\frac{P+N}{N}\right)}$$

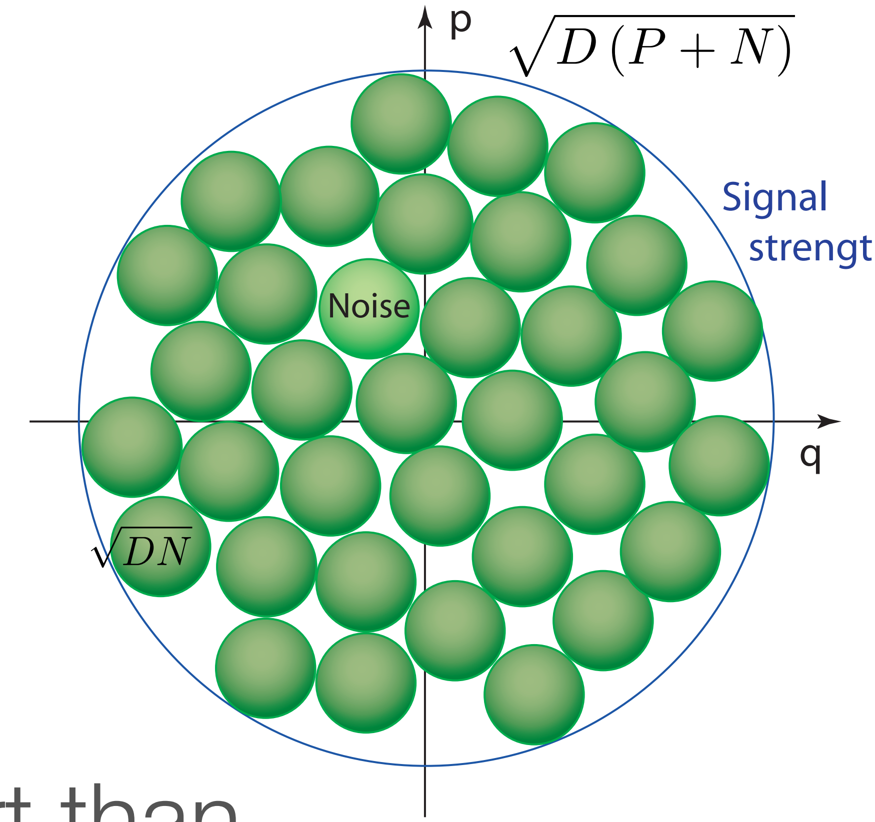
Channel capacity per symbol transmitted

$$C = \frac{1}{2}\log\left(\frac{P+N}{N}\right)$$



AREC is a Large-Deviations result, with block length as scale and capacity as rate function

- To show that error probability is bounded, send slightly fewer words, spaced at $\sqrt{D\,(N+\epsilon)}$



- Rate of transmission is then given by:

$$\mathcal{R}_\epsilon = \frac{1}{2}\log\left(\frac{N+P}{N+\epsilon}\right) \approx \mathcal{C} - \frac{\epsilon}{2N}$$

- Probability of any mis-decode is probability a signal is driven further apart than

$$P\left(z_1^2 + \cdots + z_D^2 > D\,(N+\epsilon)\right) \approx \frac{1}{\Gamma\,(D/2)} \int_{\frac{D}{2}\left(1+\frac{\epsilon}{N}\right)}^{\infty} du\, u^{(D/2-1)} e^{-u}$$

$$\left(u \equiv (1/2N) \sum_{i=1}^{D} z_i^2\right) \qquad \sim e^{-D\epsilon/2N} = e^{-D(\mathcal{C}-\mathcal{R}_\epsilon)}$$

scale structure

Large-deviations formulae for non-equilibrium systems

- The shift from ensembles of states to ensembles of histories
- A simple example with a path entropy that is not the equilibrium entropy
- Some illustrative examples from population processes

From States to Histories:
The Entropy concept is still the same one

- The most important message:
  Entropy is defined relative to distributions and aggregation
  We haven't said which entropy until we have said which distribution

- This doesn't change if we pass from equilibrium to dynamical systems

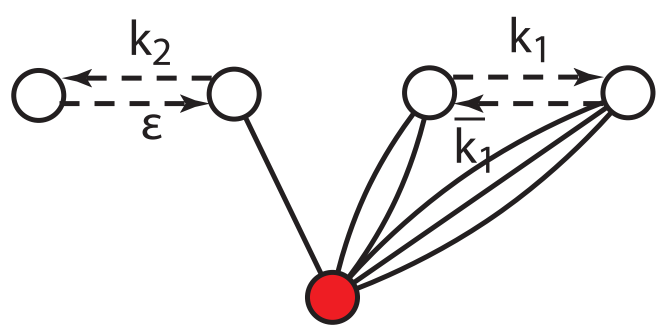
  - Equilibrium <-> ensembles of states

  - Dynamics <-> ensembles of histories

A model of a non-equilibrium phase transition mimicing metabolism

$$\varnothing \underset{k_2}{\overset{\epsilon}{\rightleftharpoons}} A$$

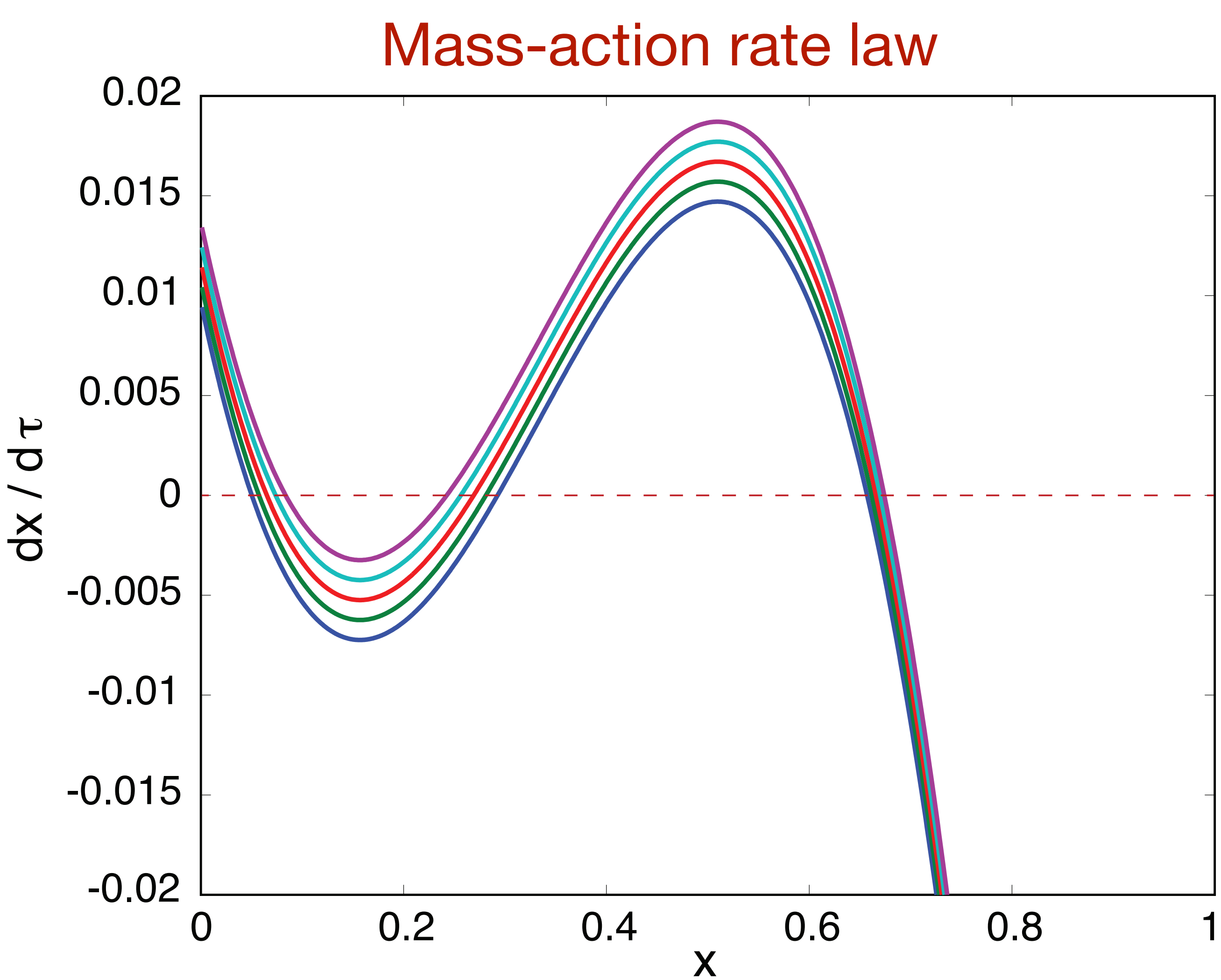
$$2A \underset{\bar{k}_1}{\overset{k_1}{\rightleftharpoons}} 3A$$

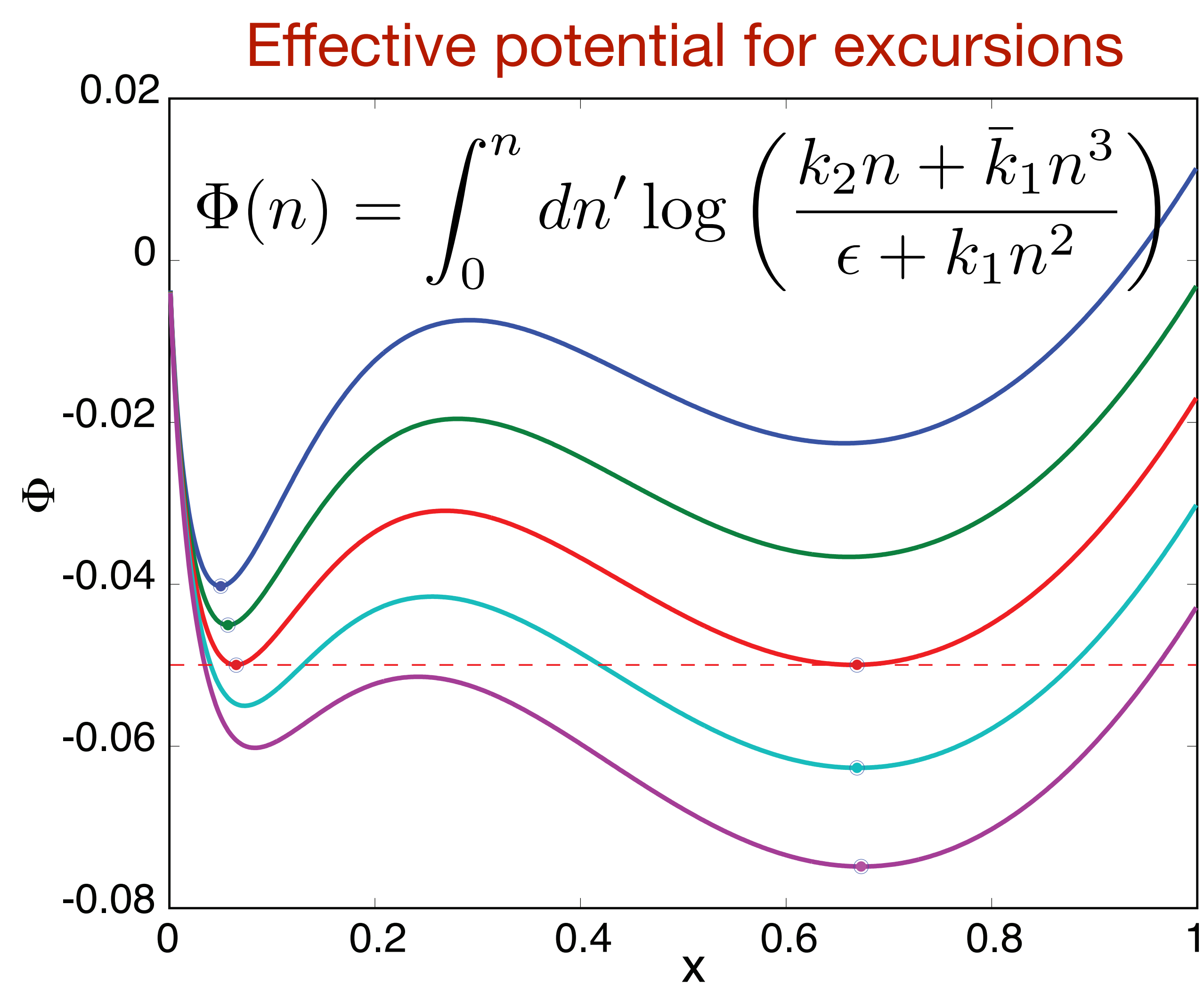


$$\frac{dn}{dt} = \epsilon - k_2 n + k_1 n^2 - \bar{k}_1 n^3$$

$$P_{\mathrm{escape}}(\bar{x} \mid \bar{x}_{\pm}) \sim e^{-n_0[\Phi(\bar{x}) - \Phi(\bar{x}_{\pm})]}$$

scale structure





Population processes: examples from Evolutionary Game Theory where collective effects matter

- A whole class of problems in neutral evolution and economic theory (Hahn paradox) where the order of limits seems to produce ambiguous results

- These are all “fragile” in the mechanical (microscopic) interpretation

- The right answer is that fluctuation effects, which remain robust, take the place of naive regulators as the naive regulators become weak

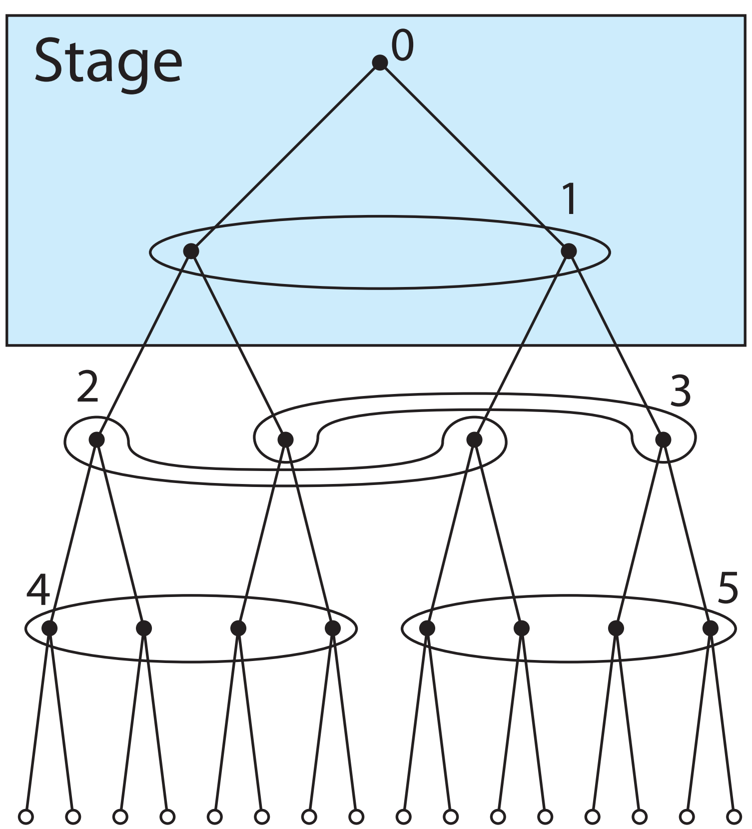
Repeated PD: entropy corrections and "free fitness" when fitness has w/ neutral directions

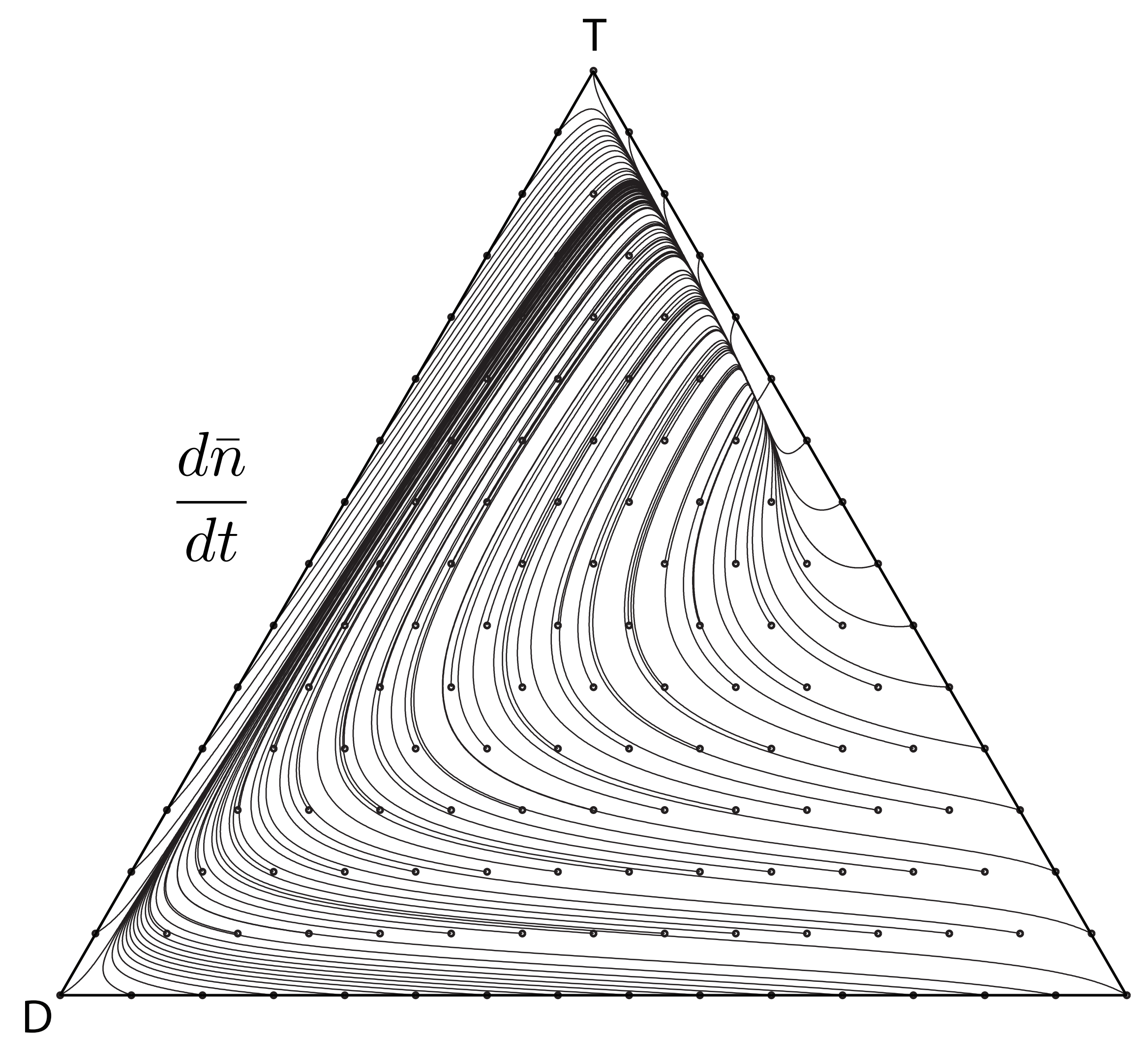
Classical repeated prisoners' dilemma:
Move "C" strictly dominated in 1 round

$$[a] = \begin{bmatrix} R & S \\ T & P \end{bmatrix} \qquad T > R > P > S$$

Repeated game;
3 strategies, all**C**, all**D**, **T**it-for-tat;
payoffs take on a neutral boundary axis

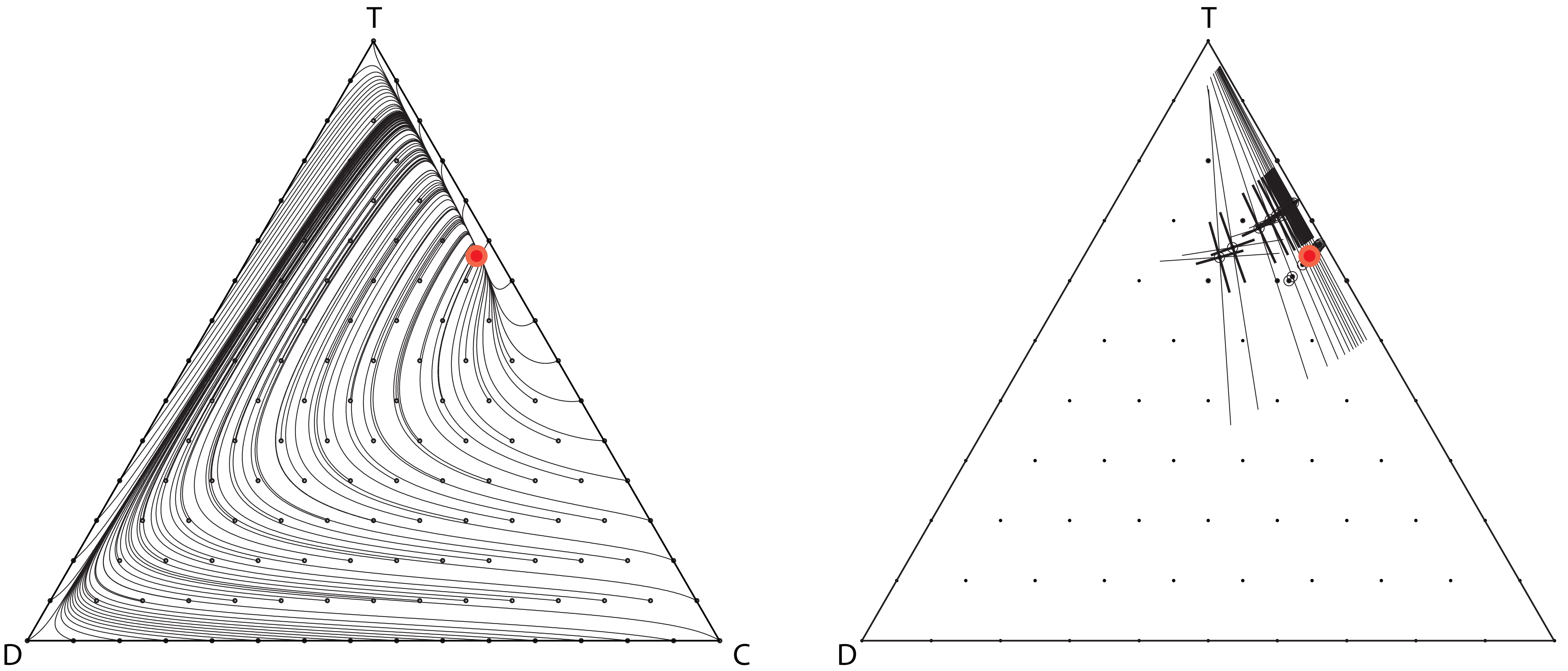
$$[a] = \begin{bmatrix} R & S & R \\ T & P & P + \epsilon_{TP} \\ R & P + \epsilon_{SP} & R \end{bmatrix}$$





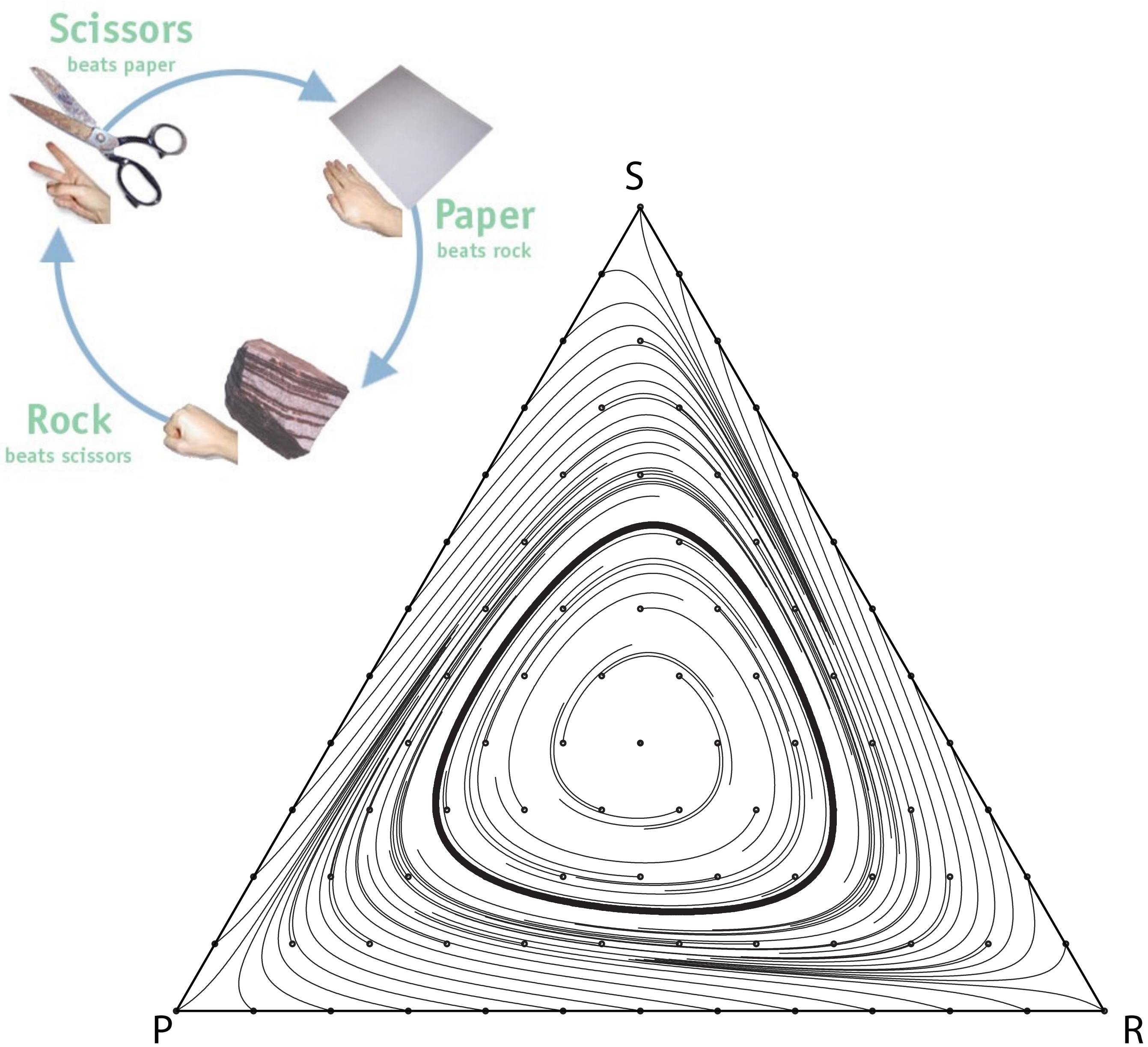
The naive infinite-population limit gives the wrong answer

- Large-population with limited mutation introduces a few “police” to disginguish naive from defensive cooperators

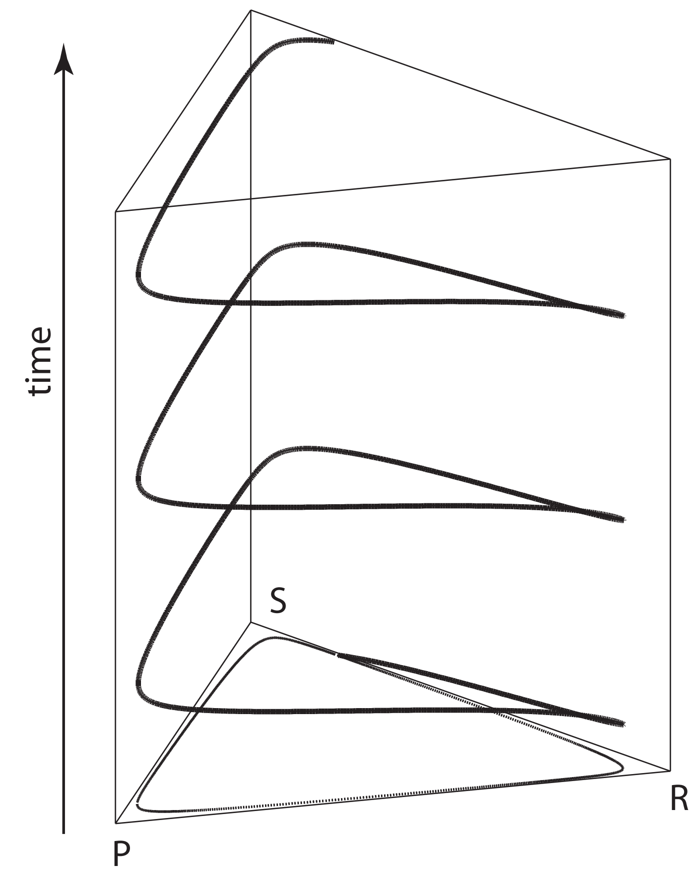


- This differs from the naive answer at O(1)

Rock-Paper-Scissors: continuous symmetry breaking in a discrete-state system



What protects the limit cycle? Spontaneously broken symmetry is **time translation**, not a spatial symmetry



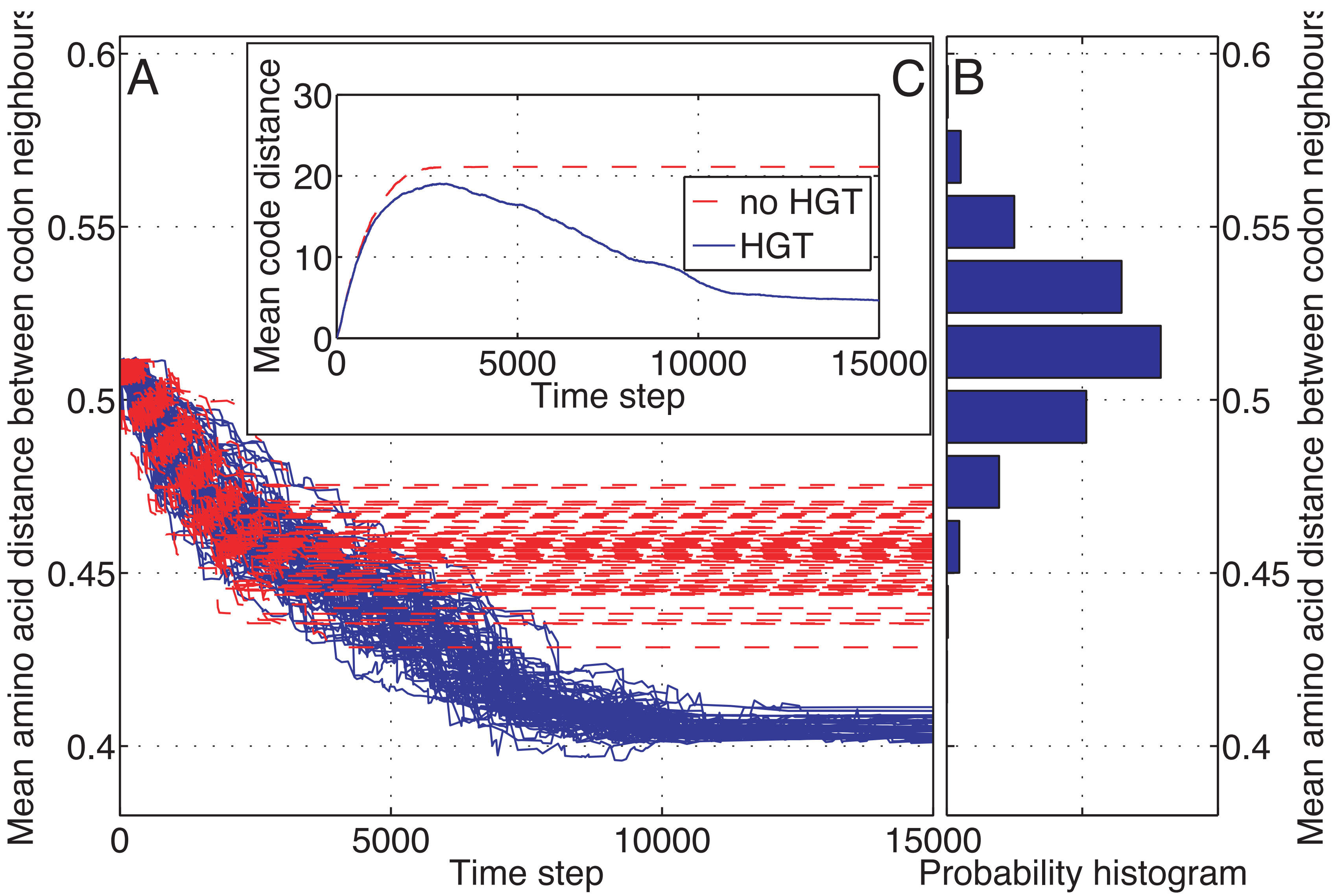
State space has only 3-cycle discrete symmetry, but dynamics takes continuous limit cycle

Quantitative argument that translation of proteins must not have been 1-to-1

Collective evolution and the genetic code

Kalin Vetsigian*, Carl Woese†‡§, and Nigel Goldenfeld*‡¶

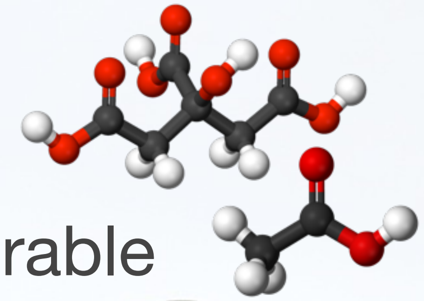
10696–10701 | PNAS | **July 11, 2006** | vol. 103 | no. 28 www.pnas.org/cgi/doi/10.1073/pnas.0603780103



Does understanding the origin and persistence of life depend on these ideas?

A fundamental question & a seeming paradox of the robust Biosphere

- **Matter continually cycles; entities always break down**
  Yet the defining patterns of life have persisted for ~4 Ga



- The most ephemeral entities (small metabolites) carry the most durable patterns (core-metabolic network roles)

- Intermediate scales are intermediate:
  Cells and organisms (10^3-10^9s) carry organizational architecture
  Populations of genomes carry species identities (~10^6 years)

- This order spontaneously displaced the order of a non-living Earth

- First emergence was quick (<10^5 years?), and seems “irreversible”

Take-home thoughts

- Scale/structure separation (LDS) gives us a formal way and quantitative tools to think about the transition from particularity to indefiniteness

- Thermodynamics is not about heat — it is an application of combinatorics originally to problems in mechanics

- The extension beyond equilibrium and beyond mechanics involves technical challenges and new kinds of organization, but not philosophical problems