Experimental Realizations of Favorable-Contact Model

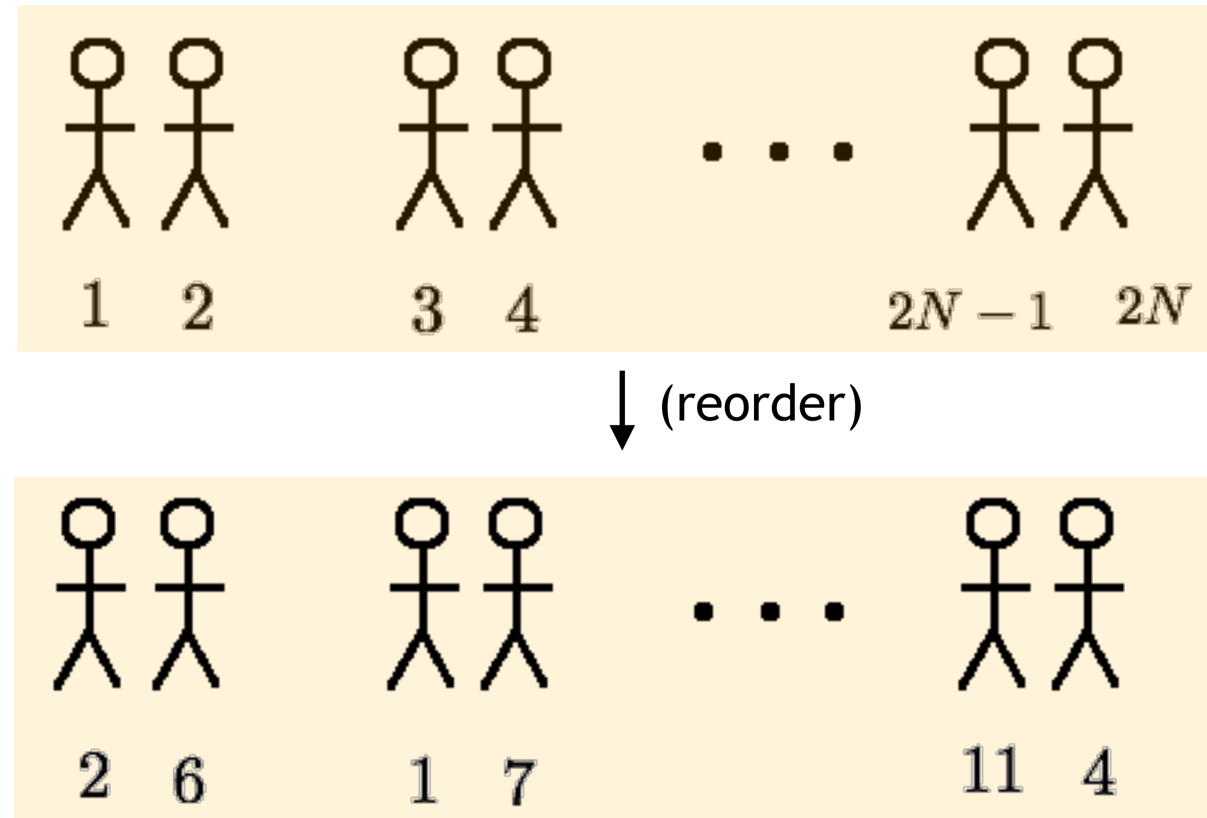
Mobolaji Williams — Shakhnovich Group Meeting— Aug. 15, 2017

(Thanks to Evgeny Serebryany)

Recall: "Dancing Partners Problem"

(B. H. Margolius,"Avoiding your spouse at a bridge party, 2001)

- *N* partners (i.e., *2N* total people) arrive at a dancing party
- Each dance partner pair is re-ordered such that each person may or may not be with their original dance partner.
- In how many ways can this happen? (What is the size of the state space?)



$$|S_N| = \frac{(2N)!}{2^N N!}$$

Main Question: Can we develop an experimentally testable physical model from these statistics?

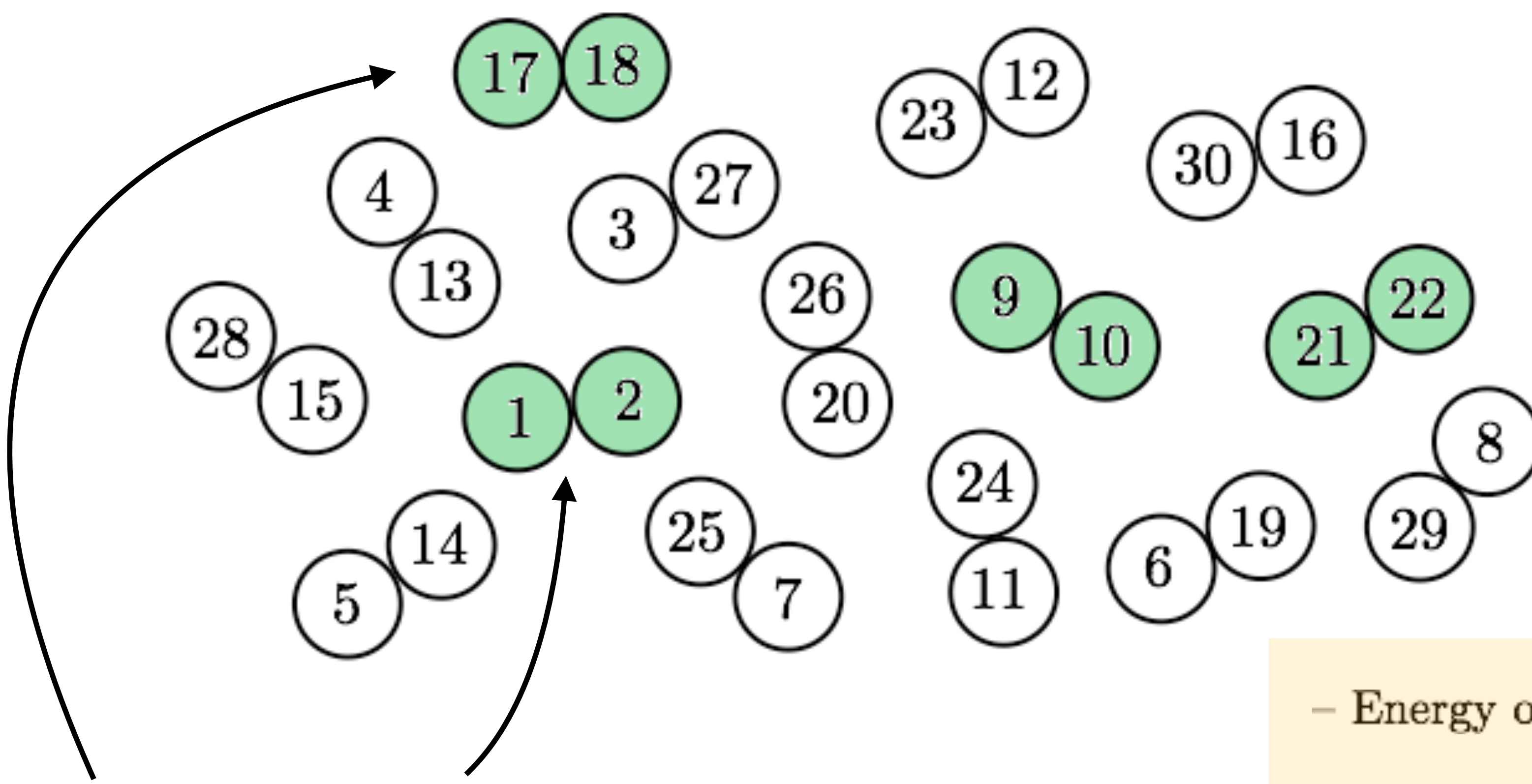
Basics of Favorable Contact Model

Say we have a system of $2N$ distinguishable elements which are organized into pairs.

We denote the elements as $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_{2N}$, and each element can be paired with any other element and the ordering within the pair is not important.

Example Microstate:

– A possible microstate of the $2N = 30$ system



– Energy of microstate is $E = 11\gamma_0$

We also take there to be a single zero-energy collection of pairings. We call these pairings "**favorable contacts**", and we write this zero-energy microstate as $\{(\alpha_1, \alpha_2), (\alpha_3, \alpha_4), \ldots, (\alpha_{2N-1}, \alpha_{2N})\}$.

For other microstates, whenever a pairing of elements does not pair α_{2k-1} with α_{2k}, (i.e., for each unfavorable contact) there is an energy costs of γ_0.

Equilibrium Thermodynamics of Model

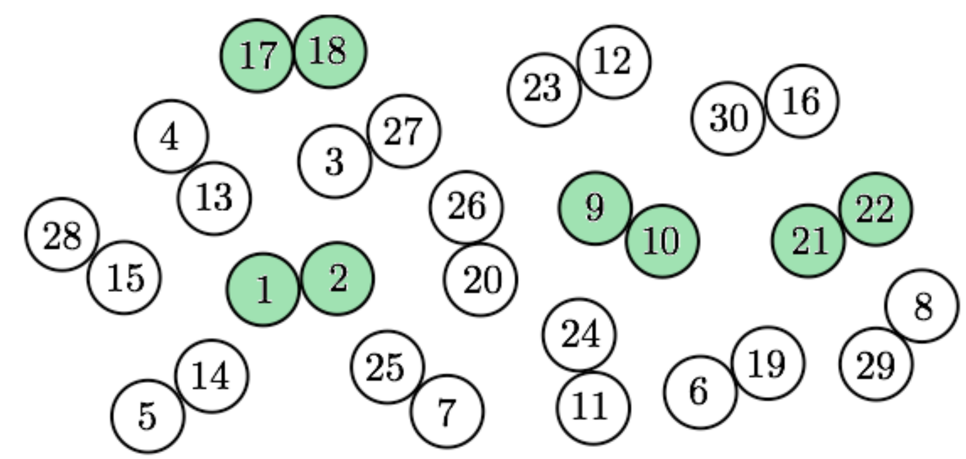
The partition function for this system is then

$$Z_N(\beta\gamma_0) = \sum_{\{(\theta_1^{(i)},\theta_2^{(i)})\}} \exp\left(-\beta\gamma_0 \sum_{j=1}^{N} I(\theta_1^{(j)},\theta_2^{(j)})\right),$$

where $(\theta_1^{(i)},\theta_2^{(i)})$ denotes the 1st and 2nd components (where the internal ordering of the components is not important) of the ith interaction pair, and where

$$I(\theta_1^{(j)},\theta_2^{(j)}) = \begin{cases} 0 & (\theta_1^{(i)},\theta_2^{(i)}) = (\alpha_1,\alpha_2),\ldots,\text{ or } (\alpha_{2N-1},\alpha_{2N}) \\ 1 & \text{otherwise.} \end{cases}$$

What are the equilibrium thermodynamics of this model?



Redefine partition function in terms of ℓ, the number of unfavorable contacts.

$$Z_N(\beta\gamma_0) = \sum_{\ell=0}^{N} \binom{N}{\ell} a_\ell \, e^{-\beta\gamma_0\ell}$$

ℓ : number of unfavorable contacts

(Definition of a_ℓ): 2ℓ objects are initially paired in some way and then are re-paired so that no object is with its original partner. a_ℓ is the number of ways this re-pairing can occur.

combinatorics manipulations

$$Z_N(\beta\gamma_0) = \frac{1}{\sqrt{\pi}} \int_0^\infty dt\, t^{-1/2} e^{-t} \left[1 + (2t-1)e^{-\beta\gamma_0}\right]^N$$

$$a_\ell = (-1)^\ell \sum_{k=0}^{\ell} (-1)^k \binom{\ell}{k} \prod_{i=1}^{k} (2i-1)$$

Physics Predictions: Order Parameter and Tc

$$Z_N(\beta\gamma_0) = \frac{1}{\sqrt{\pi}} \int_0^\infty dt\, t^{-1/2} e^{-t} \left[1 + (2t-1)e^{-\beta\gamma_0}\right]^N$$

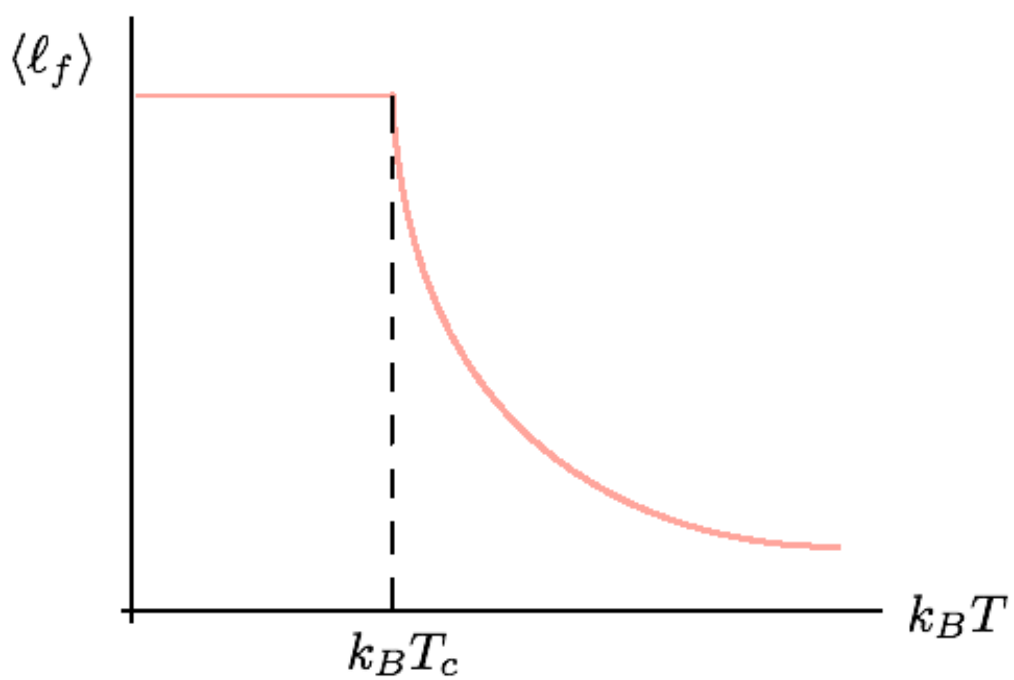
What physical predictions can we obtain from the model?

$$\langle \ell_f \rangle = N + \frac{\partial}{\partial(\beta\gamma_0)} \ln Z_N(\beta\gamma_0)$$
$$\simeq \frac{1}{4(1-e^{-\beta\gamma_0})} \left[2(N-1) + e^{\beta\gamma_0} - \sqrt{(2N - e^{\beta\gamma_0})^2 - 4(e^{\beta\gamma_0} - 1)}\right]$$

Order Parameter: $\langle \ell \rangle$, average number of unfavorable contacts

* $\langle \ell_f \rangle = N - \langle \ell \rangle$, average number of favorable contacts

Schematic plot of average number of favorable contacts



We can compute this critical temperature.

Temperature where $\langle \ell_f \rangle \in \mathbb{R} \to \langle \ell_f \rangle \in \mathbb{C}$:

$$k_B T_c = \frac{\gamma_0}{\ln(2N)} \left(1 + \sqrt{\frac{2}{N} \frac{1}{\ln(2N)}}\right) + \mathcal{O}\left(N^{-1}\right)$$

Checking Limiting cases

- $\gamma_0 \to \infty, T_c \to \infty$: Favorable contacts persist at higher temperature
- $N \to \infty, T_c \to 0$: Larger entropic contributions require a lower T_c

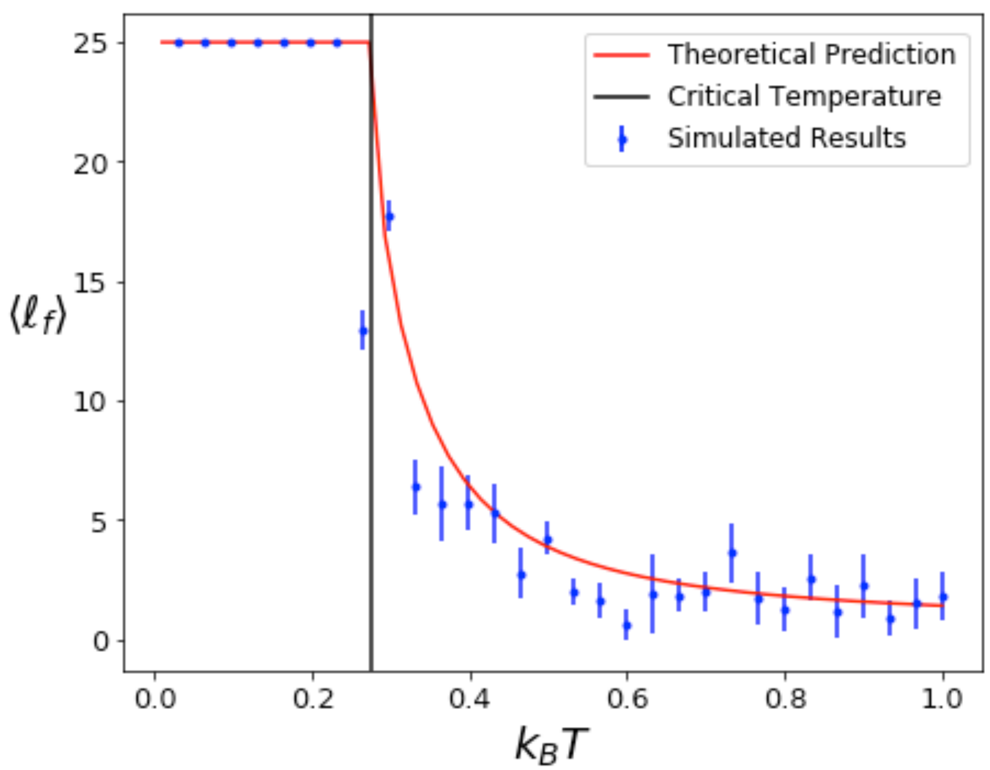
Comparison with Simulations

$$\langle \ell_f \rangle \simeq \begin{cases} N & \text{for } T < T_c, \\ \dfrac{1}{4(1-e^{-\beta\gamma_0})}\left[2(N-1) + e^{\beta\gamma_0} - \sqrt{(2N-e^{\beta\gamma_0})^2 - 4(e^{\beta\gamma_0}-1)}\right] & \text{for } T \geq T_c. \end{cases}$$

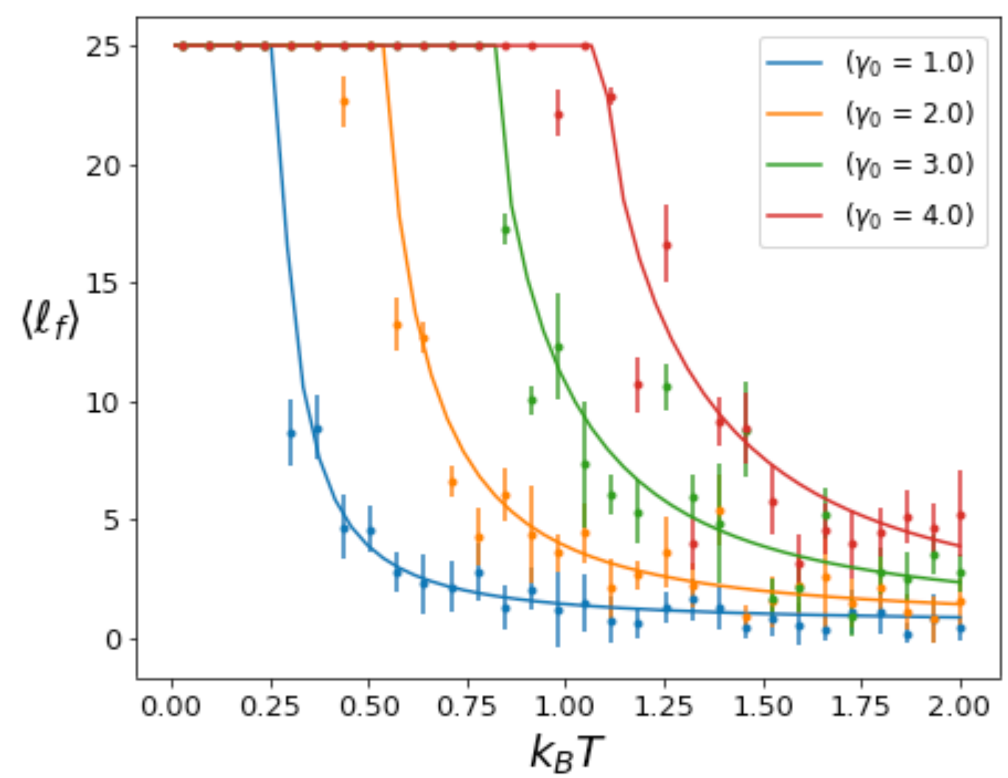
($\langle \ell_f \rangle = N - \langle \ell \rangle$, average number of favorable contacts)

$$k_B T_c = \frac{\gamma_0}{\ln(2N)}\left(1 + \sqrt{\frac{2}{N}}\frac{1}{\ln(2N)}\right) + \mathcal{O}\left(N^{-1}\right)$$

(critical temperature; i.e., T where $\langle \ell_f \rangle \in \mathbb{R} \to \langle \ell_f \rangle \in \mathbb{C}$)



Monte Carlo Simulation for $N = 25$ pairs, $\gamma_0 = 1.0$



Monte Carlo Simulation for $N = 25$ pairs, for various γ

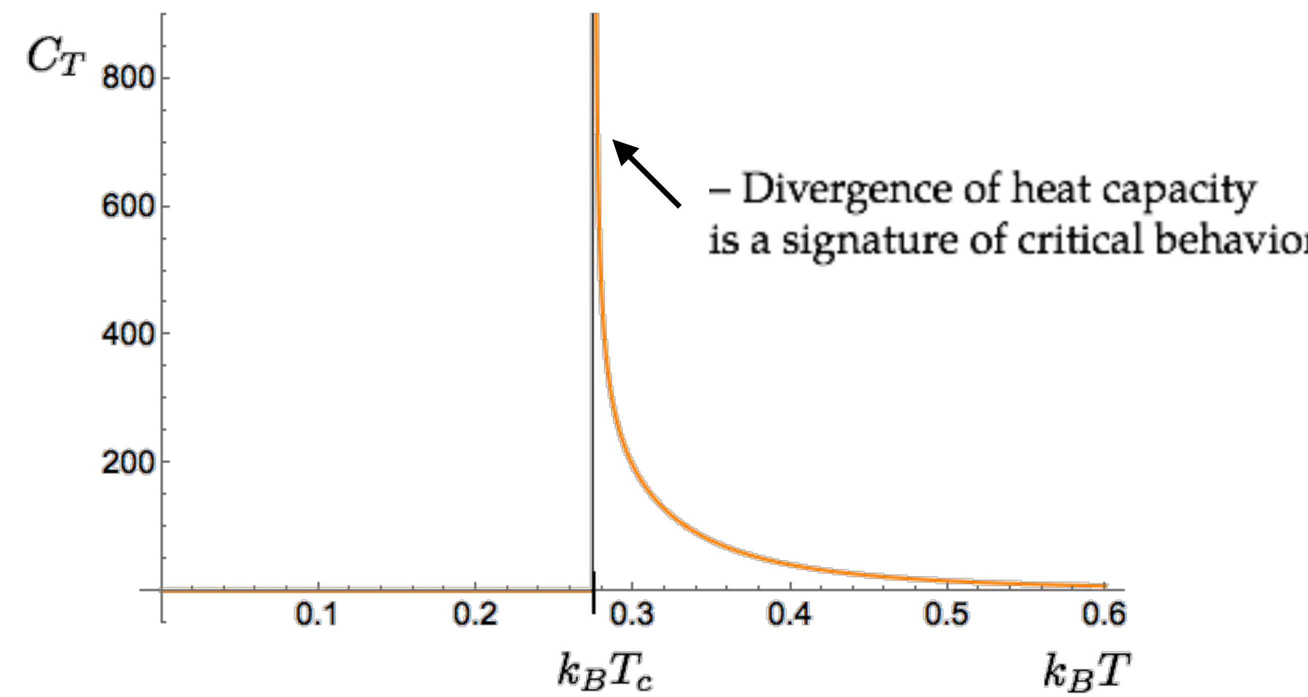
Physics Predictions: Heat Capacity and Critical Behavior

Does the model exhibit critical behavior?

(Definition of universality):
When the behavior of a system near a critical transition is independent of the microscopic details of the system.

– critical exponents indicate the type of universality a system exhibits

Heat Capacity for $N = 25$ and $\gamma_0 = 1.0$



– Critical behavior is important because of what is suggests about universality

Determining critical exponents

– How do order parameter and heat capacity vary with T near critical temperature (i.e., for $|T - T_c| \ll T_c$)?

$$\langle \ell_f \rangle \sim N - (T - T_c)^{\beta}, \qquad C_T \sim (T - T_c)^{-\alpha}.$$

→ $$\langle \ell_f \rangle \simeq N - \frac{1}{2} (2N)^{1/4} |f'(T_c)|^{1/2} (T - T_c)^{1/2}$$

$$C_T \simeq \frac{\gamma_0}{4} (2N)^{1/4} |f'(T_c)|^{1/2} (T - T_c)^{-1/2}$$

$^* f(T) = e^{\gamma_0 / k_B T}$

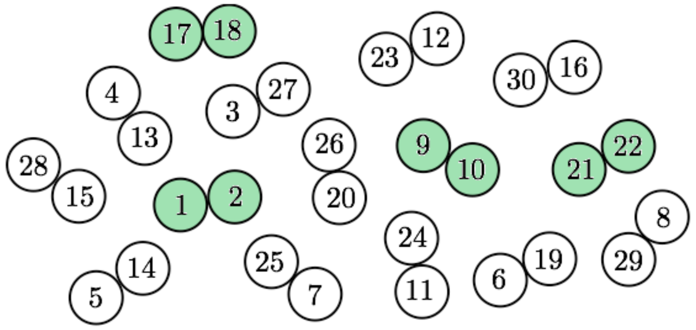
$\alpha = \beta = 1/2$

Table of critical exponents for canonical systems

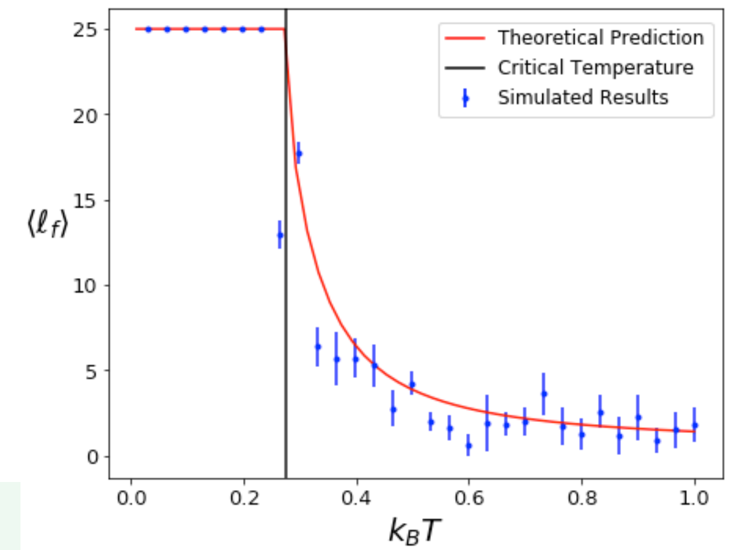
Theory	α	β
Mean Field Theory Ising	0	1/2
Liquid Gas Transition	0	1/2
2D Ising Model	0	1/8
Favorable contact Model	1/2	1/2

Critical behavior, but no existing universality class!

Physics Predictions: Summary



Model has a number of physical predictions (which can be checked by simulations)



1. **Order Parameter:**

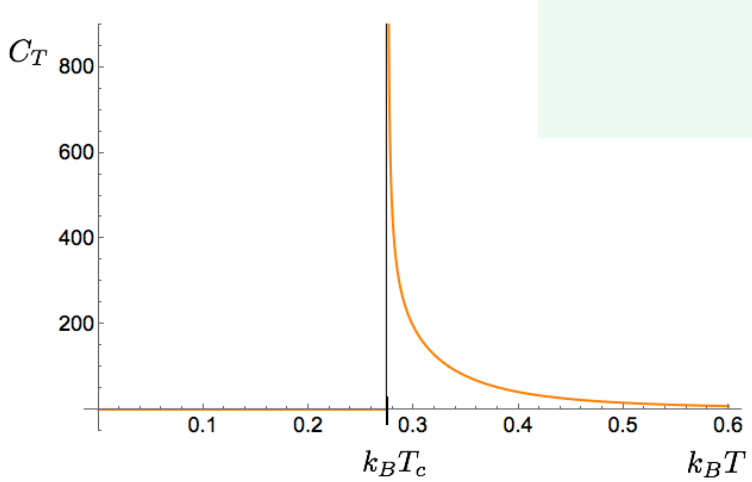
$$\langle \ell_f \rangle \simeq \frac{1}{4(1-e^{-\beta\gamma_0})}\left[2(N-1)+e^{\beta\gamma_0}-\sqrt{(2N-e^{\beta\gamma_0})^2-4(e^{\beta\gamma_0}-1)}\right].$$

2. **Critical Temperature:**

$$k_B T_c = \frac{\gamma_0}{\ln(2N)}\left(1+\sqrt{\frac{2}{N}}\frac{1}{\ln(2N)}\right)+\mathcal{O}(N^{-1}).$$

3. **Critical Exponents:**

$$C_T \sim (T-T_c)^{-1/2} \text{ and } \langle \ell \rangle \sim (T-T_c)^{1/2}.$$



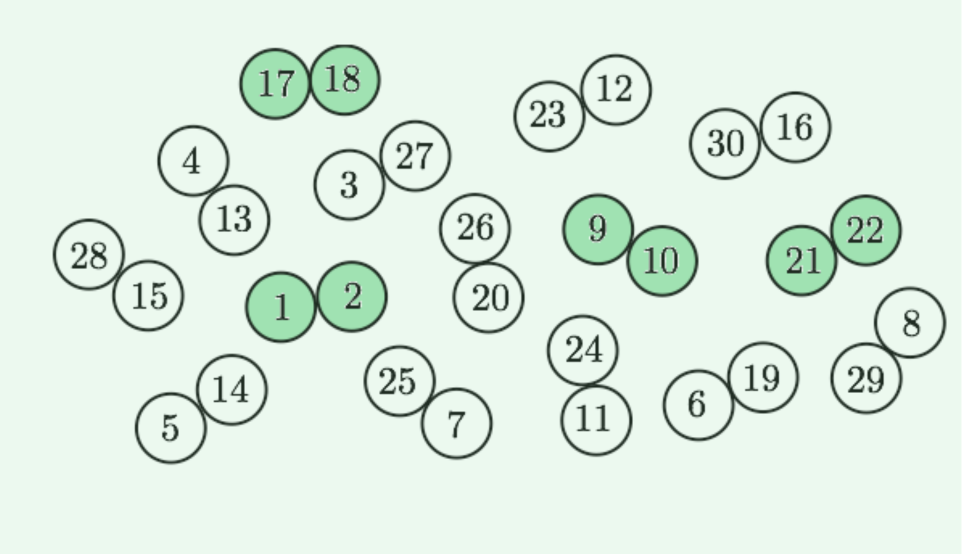
Are there any experimental systems where we can test these ideas?

Needed Elements

Is it possible to prepare a system where:

1. there is an even number of distinguishable subunits
2. subunits exist primarily subunit-subunit pairs (energetically unfavorable to exist otherwise)
3. there is a single optimal collection of pairings
4. there is a measurable energy cost for deviations from these optimal pairings.

$2N$ subunits

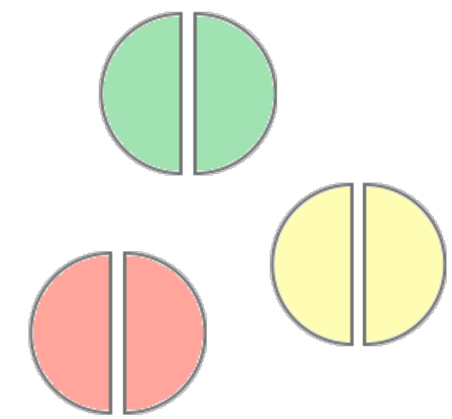


$$\mathcal{H} = \gamma_0 \sum_{j=1}^{N} I(\theta_1^{(j)}, \theta_2^{(j)})$$

Optimal Pairings: $\{(\alpha_1, \alpha_2), (\alpha_3, \alpha_4), \ldots, (\alpha_{2N-1}, \alpha_{2N})\}$

Candidate Biophysical Systems

Possible Systems

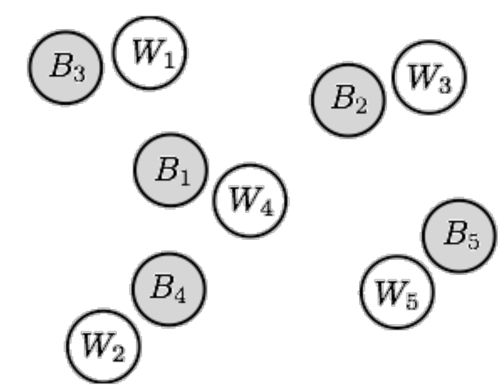


1. Protein-protein interactions

1. need to engineer/collect proteins prone to forming dimers
2. avoid formation of oligomers greater than 2
3. need to determine interaction energies for pairs

2. Transcription Factor-single stranded DNA (ssDNA) interactions

1. **Advantage:** ability to tune sequence allows for precise binding
2. **Problem:** This is a “gendered” favorable contact model
   - ➤ Analyze using previous permutation model



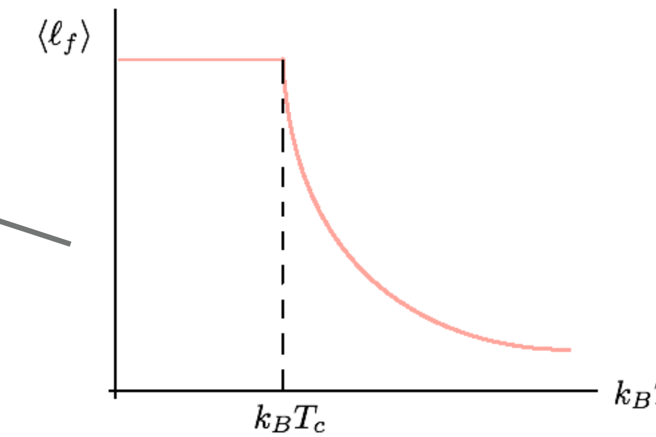
3. ssDNA-ssDNA interactions

1. Don’t have to deal with complications of tertiary structure
2. sequences provide simple biophysical estimate of model parameters
3. (supposedly) easy to order/prepare

e.g., Unfavorable contact	e.g., Favorable Contact
5’-AGCTAACGTA-3’	5’-AGCTAACGTA-3’
3’-GGACTACTTT-5’	3’-TCGATTGCAT-5’

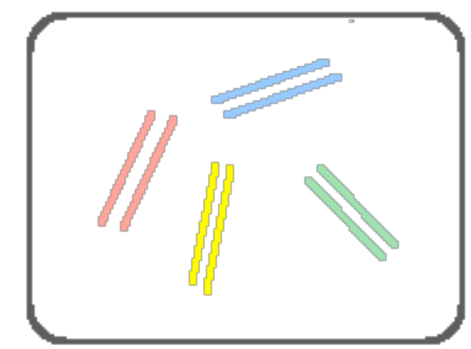
Setup with Single Stranded DNA (ssDNA)

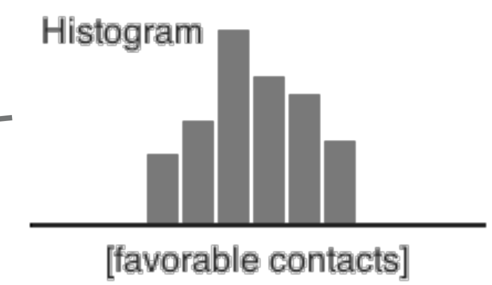
Experimental Goal: Measure number of ssDNA bonded to their complementary strands across a range of temperatures.

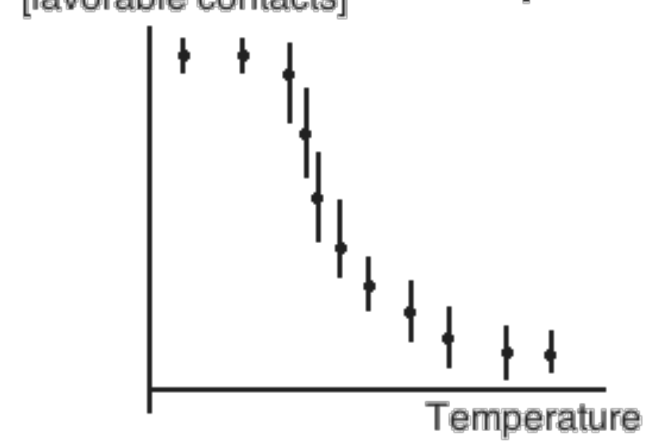


1. Prepare *N* different ssDNA *n*-mers
    1. Not clear how long they should be, but the shorter the better. (in order to prevent partial alignment of sequences, and formation of hairpins)
2. Prepare *N* complementary ssDNA fragments (all in an aqueous solution)
3. Begin at high temperature such that ssDNA are unbound
4. Rapidly lower temperature to desired T_i
5. Measure concentration or fraction of favorable contacts*
6. Repeat 1—5 at specific T_i to accumulate statistics
7. Repeat 1—6 for range of temperatures {T_1, T_2, T_3, . . .}

5’-AGCTAACGTA-3’







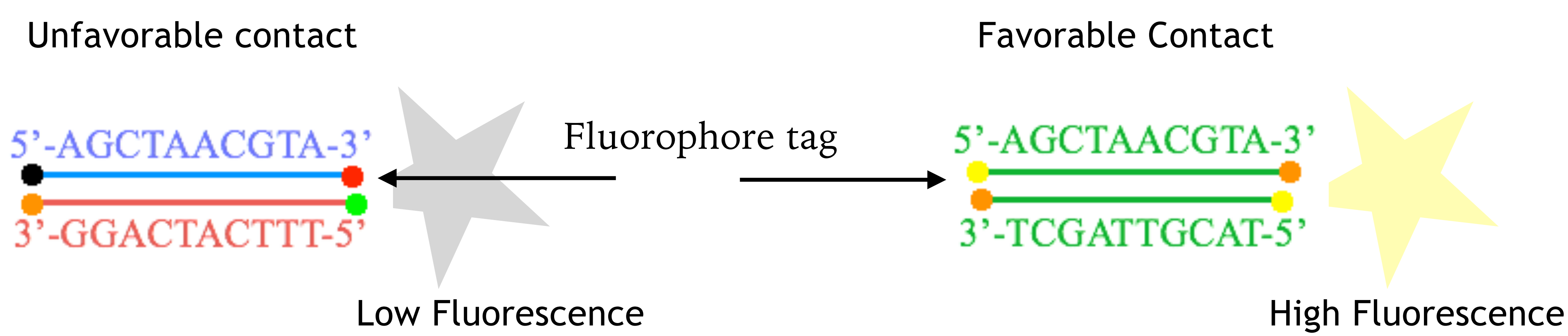
Measuring Number of Favorable Contacts

Pivotal Question: How do we measure number of favorable contacts?

➤ Attaching fluorophores to ends of ssDNA can allow us to measure the amount of DNA that has formed double strands

(Quantifying DNA concentrations using fluorometry: A comparison of fluorophores, Rengarajan et. al. (2002))

➤ Certain combinations of fluorophore tags lead to quenching or amplification of light produced by fluorophore



Fluorescence of Solution ~ Concentration of Favorable Contacts

Short list of possible issues

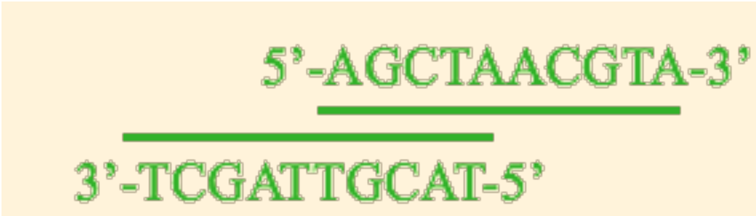
$$Z_N^{\mathrm{FCM}}(\{\beta\gamma_i\}) = \frac{1}{\sqrt{\pi}} \int_0^\infty dt\, t^{-1/2} e^{-t} \prod_{k=1}^{N} \left(1 + (2t-1)e^{-\beta\gamma_k}\right)$$

1. **Parameter Space**: The model only has two free parameters (energy cost and the number of subunits). Might not be enough freedom to represent an experimental situation.

   ➤ Would have to consider the generalized Favorable Contact Model

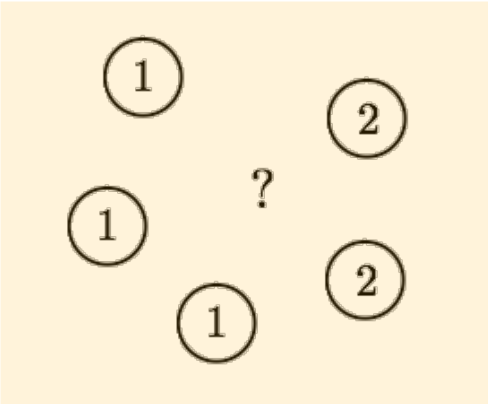
Free parameters:
γ_0 and N

2. **Fragment Overlap:** It's possible for complementary fragments to overlap in a way so that the nucleotide bases are not coupled to complementary base. Model doesn't allow for such partial contacts.



3. **Multiple Copies of Each ssDNA:** Experimentally, it is not feasible to make a single copy of each strand, but in the model there is only a single copy of each subunit.

   ➤ Could model by modified to include multiple subunits of each type? (Or can the model encompass this situation?)



4. **Biophysical Issues:** The pivotal physical prediction of the model is the critical temperature. Is this temperature in a physically relevant range?

$$k_B T_c = \frac{\gamma_0}{\ln(2N)} \left(1 + \sqrt{\frac{2}{N}} \frac{1}{\ln(2N)}\right) + \mathcal{O}\left(N^{-1}\right)$$

“Order of Magnitude” Sanity Check

Question: Does an estimate of the critical temperature yield a reasonable temperature (e.g.,

1. below melting temperature of DNA and
2. above freezing temperature of water?)

$$k_B T_c \simeq \frac{\gamma_0}{\ln(2N)}$$

Energy cost for unfavorable contact

N: number of favorable contact pairings

- Say we have a 30-mer oligonucleotide.

e.g., 5’-ACT GTG TCG TAT GAA TCG-3’

- For its double stranded form, we can approximate a melting temperature of:

$$T_\mathrm{m} \approx 60^\circ\,\mathrm{C}$$

- Let’s approximate the energy cost as:

$$\gamma_0 = k_B T_m \approx 3\ \mathrm{kJ/mol}$$

- Assume there are *N* = 25 favorable pairings

$$N = 25$$

- We then find a critical temperature estimate of

$$T_c \simeq \frac{\gamma_0}{k_B \ln(2N)} \approx 92\,\mathrm{K} = -181\,^\circ\mathrm{C}$$

Rough prediction of critical temperature (in Kelvin) is **off by a factor of ~3** from a physical regime.

END