Just a quick, useless recap

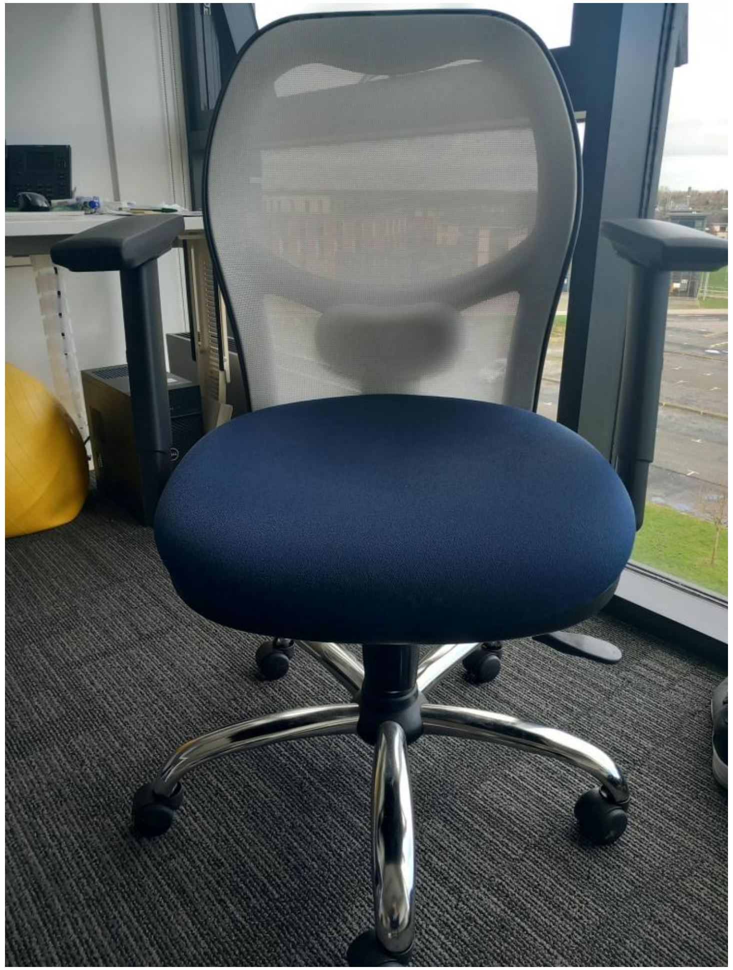
What happened last semester!

The Curse of Hamilton's Chairs

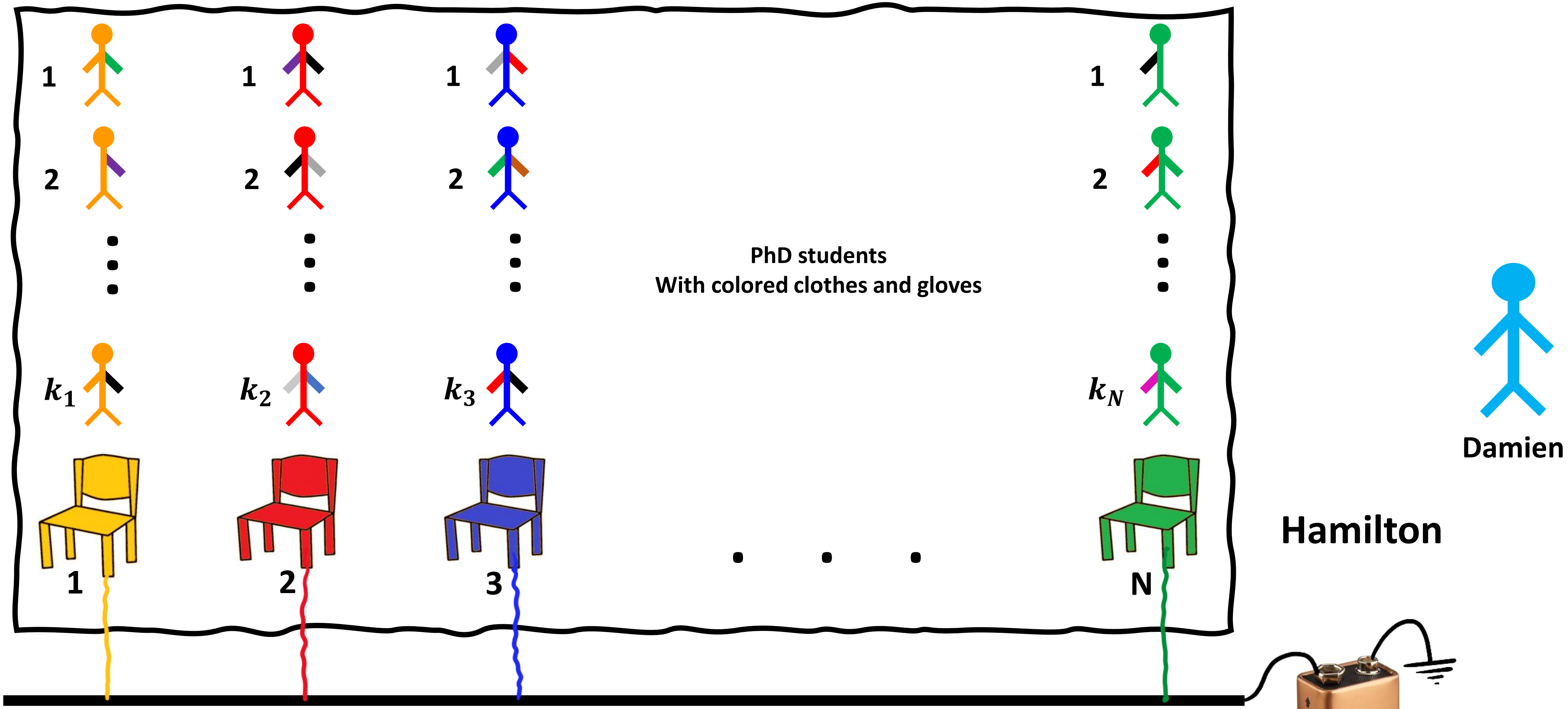
Ahmed Shalaby
2nd year PhD

Supervisor: Damien Woods

European Innovation Council
Funded by the European Union



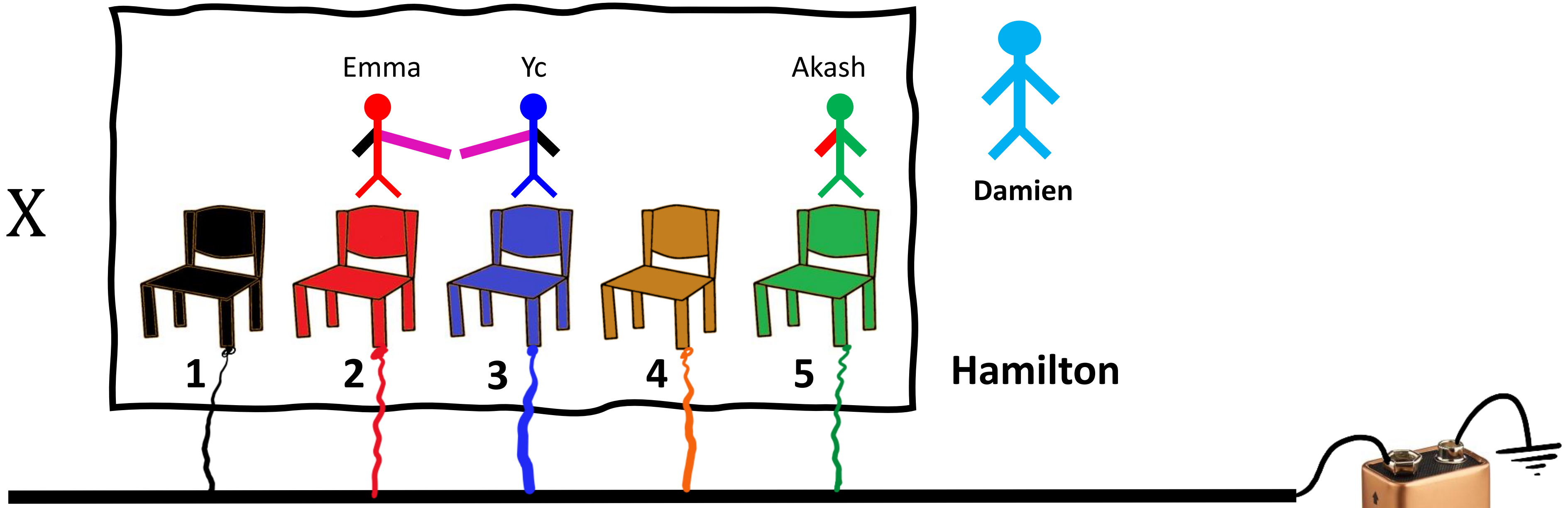
Let’s discover the rules of the game



- How many different configurations we will have ?

$$(K+1)^N$$

(Exponential in the # of chairs)



$$\mathrm{E(X)} = \mathrm{sit(Emma)} + \mathrm{sit(Yc)} + \mathrm{sit(Akash)} + \mathrm{handshake(Emma, Yc)} + 3 * \mathrm{sit\_convincing\_cost}.$$

\+ \+ \+ \+ -

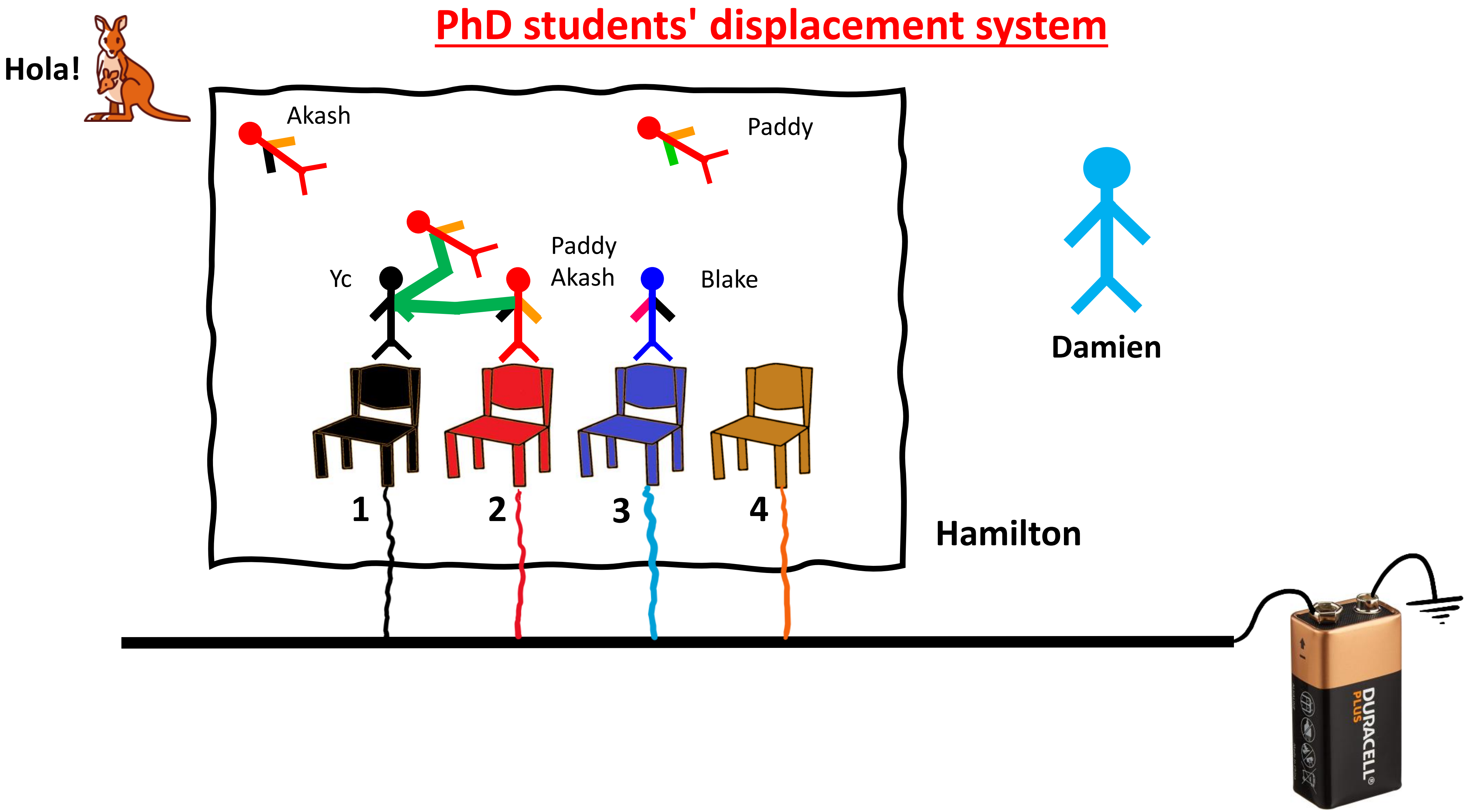
$$\mathrm{E}(X) = \sum_{p \in X} \mathrm{sit(p)} + \sum_{p_i, p_{i+1} \in X} \mathrm{handshake}(p_i, p_{i+1}) + l * \mathrm{sit\_convincing\_cost}.$$

configuration X of size l PhD students

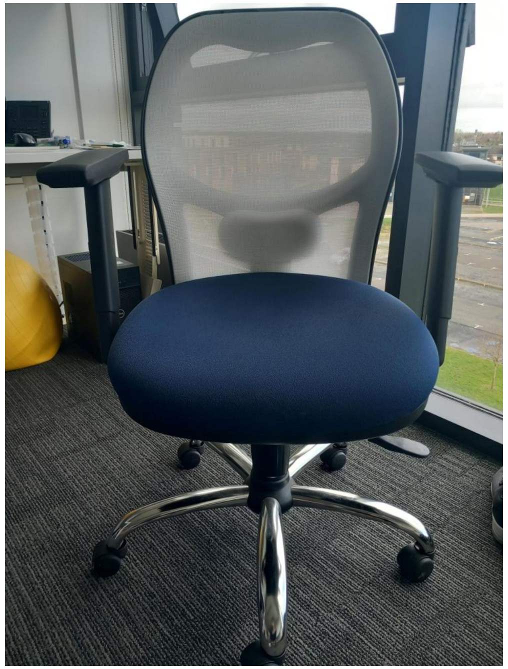
We further assume the following:

- $|\mathrm{sit}(p)| > |\mathrm{sit\_convincing\_cost}|$. (Damien always gains by convincing a PhD student to sit)

Built–in self improvement mechanism



When you don’t set your boundaries



Hamilton Institute

Maynooth University
National University of Ireland Maynooth

The Curse of Hamilton's Sofa

How I discovered that my supervisor is actually a VAMPIRE

An efficient minimum free energy algorithm for interacting nucleic acid strands

Ahmed Shalaby

2nd year PhD

Supervisor: Damien Woods

Fondúireacht Eolaíochta Éireann sfi Dá bhfuil romhainn

European Innovation Council

Funded by the European Union

erc

Let’s discover the rules of the game

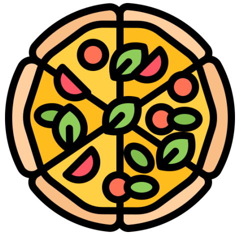
In a perfect world

- **Abstract Algebra**

- **Graph theory**

- **Algorithm analysis**

- **Number theory**

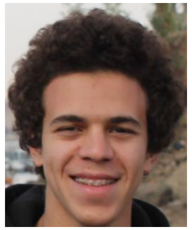


DUNNES STORES

ALDI

TESCO

SuperValu



Ahmed's goal

- What is his mindset? → **Modelling**
- What he prefers? → **Computation**

PI

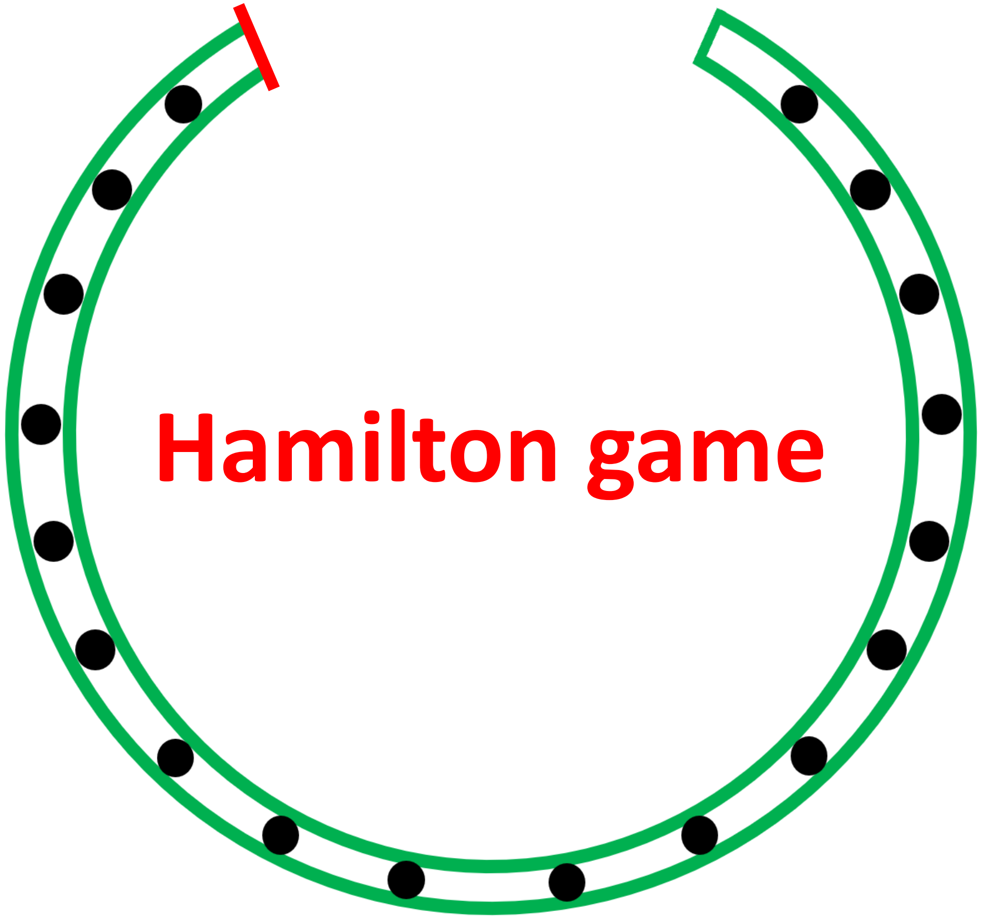
?

?

?

Once Upon a Time in Hamilton

Email From Rosemary Kate


Hamilton game

Level 1

Hamilton game

First year

Daniel
Augustina
.
.
.

Second year

Ahmed
Andre
Paddy
Cormac
.
.

Third year

Dara
Solmaz
Oluwayomi
.
.
.

fourth year

Akash
Yc
Emma
Darshana
.
.

Don’t trust our new lab

Akash
Akash
Akash
Akash
.

Dara
Dara
Dara
Dara
.

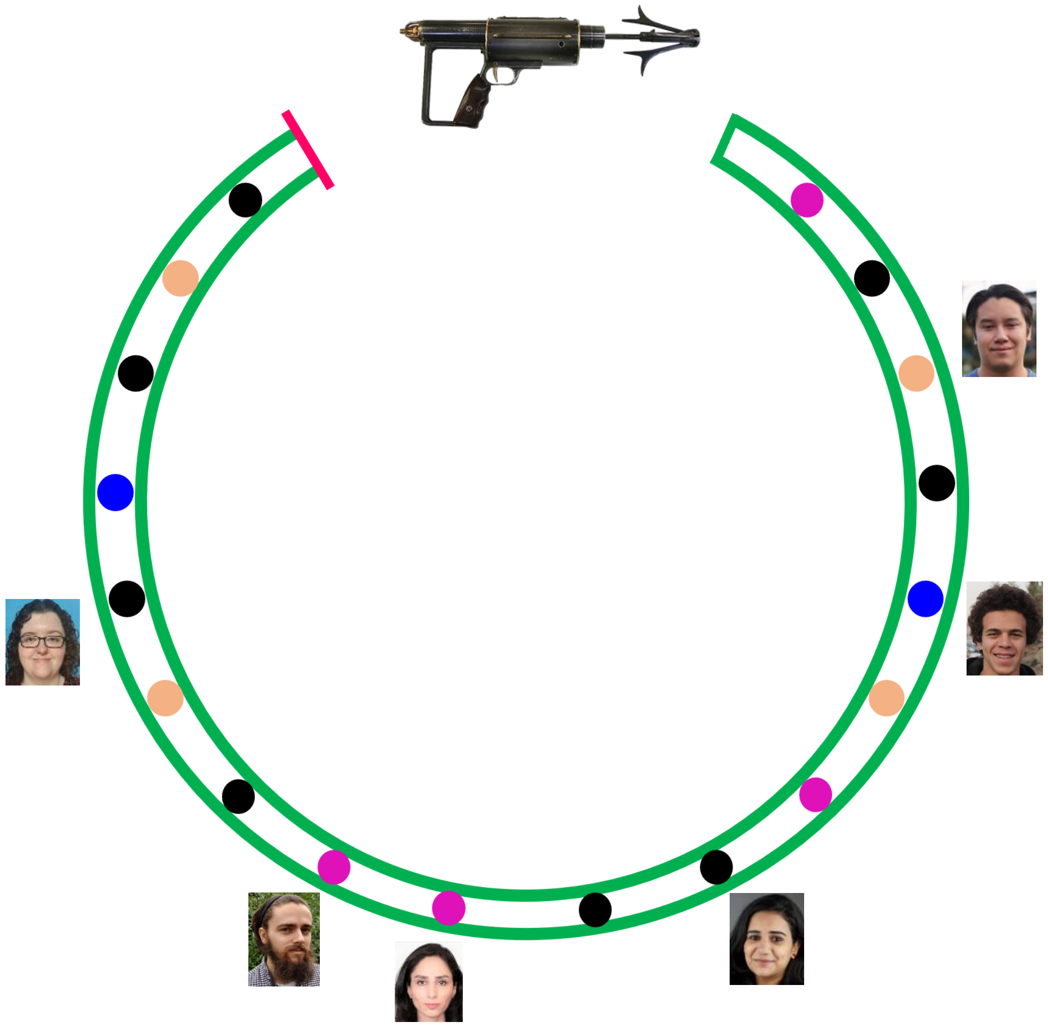
Darshana
Darshana
Darshana
Darshana
.

Fergal
Fergal
Fergal
Fergal
.

. . .

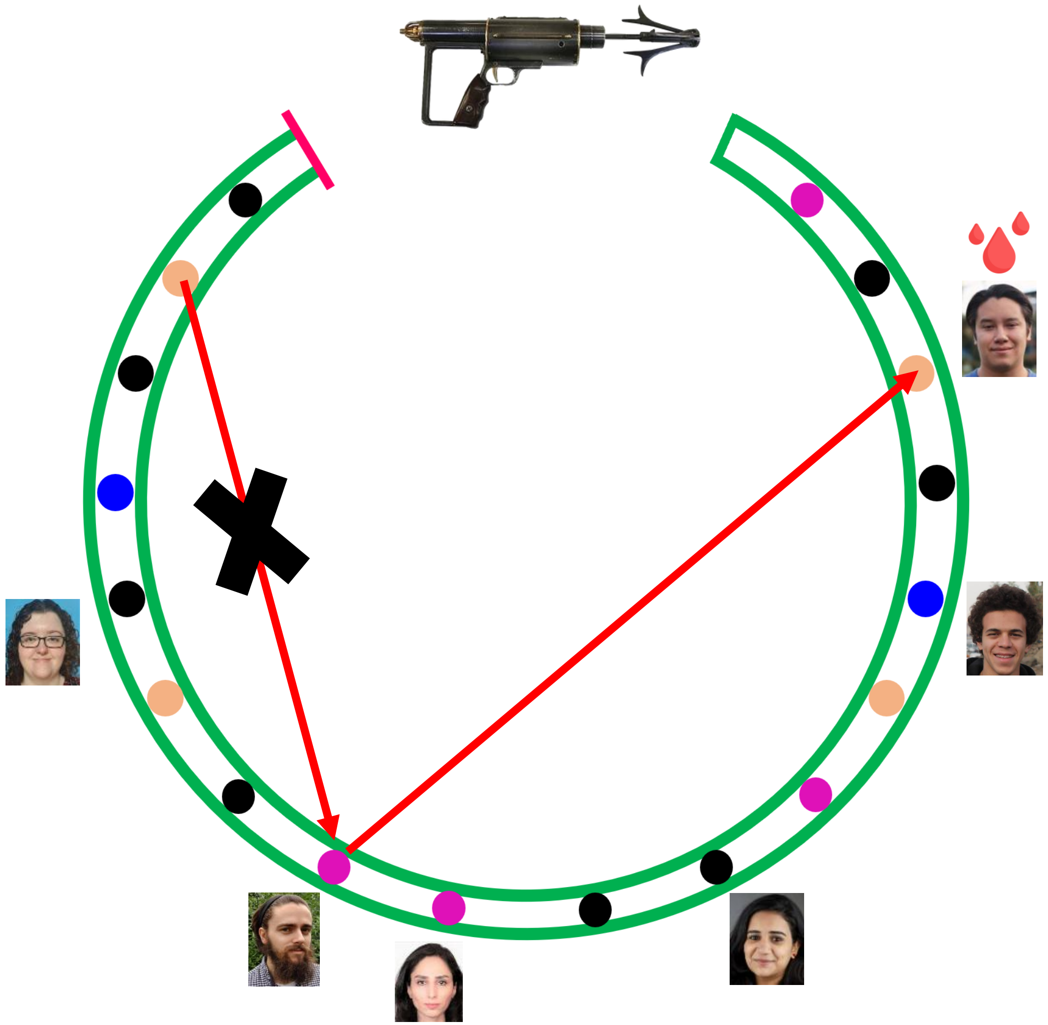
Ahmed

Rules of Hamilton game



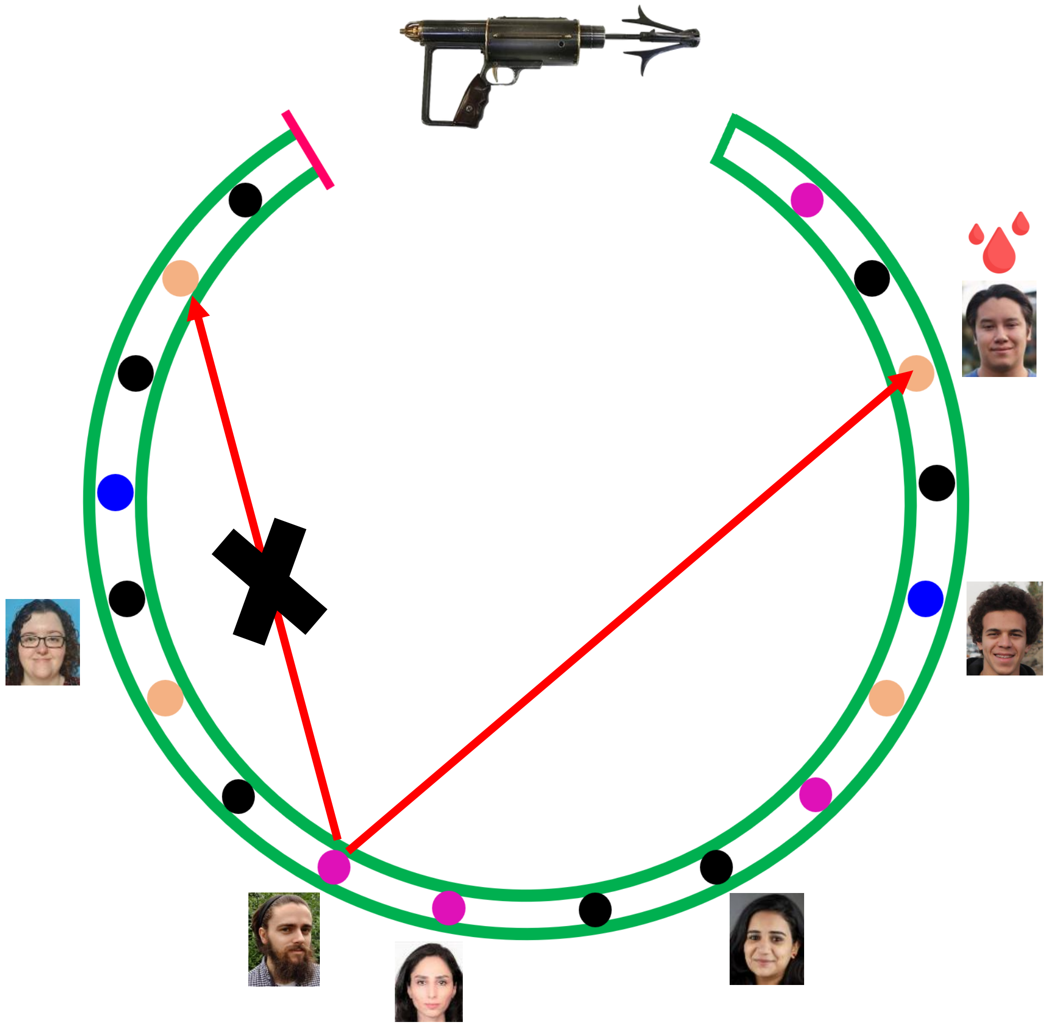
Rules of Hamilton game

- If you kill, you are safe



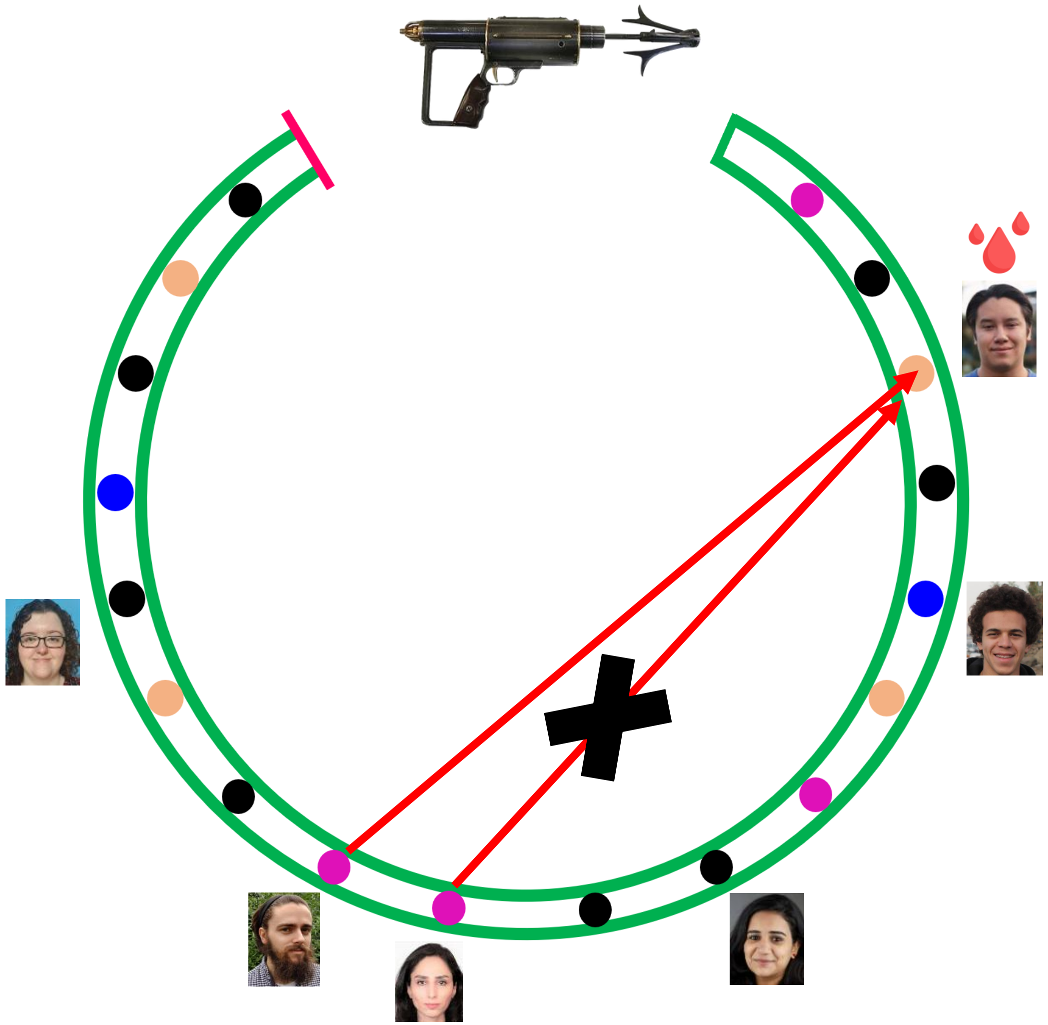
Rules of Hamilton game

- If you kill, you are safe
- You can only kill one student, you have only one bullet



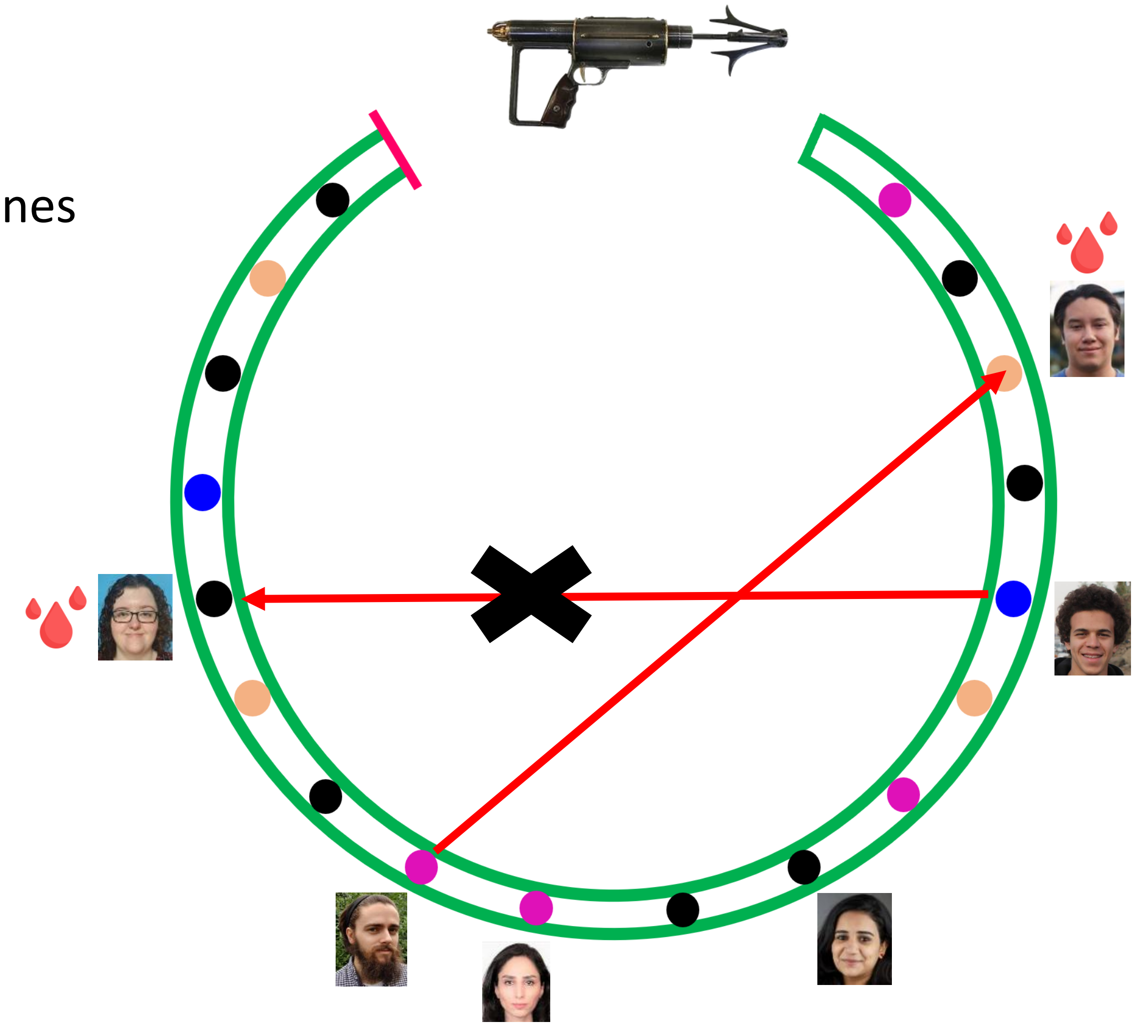
Rules of Hamilton game

- If you kill, you are safe
- You can only kill one student, you have only one bullet
- You can be killed once



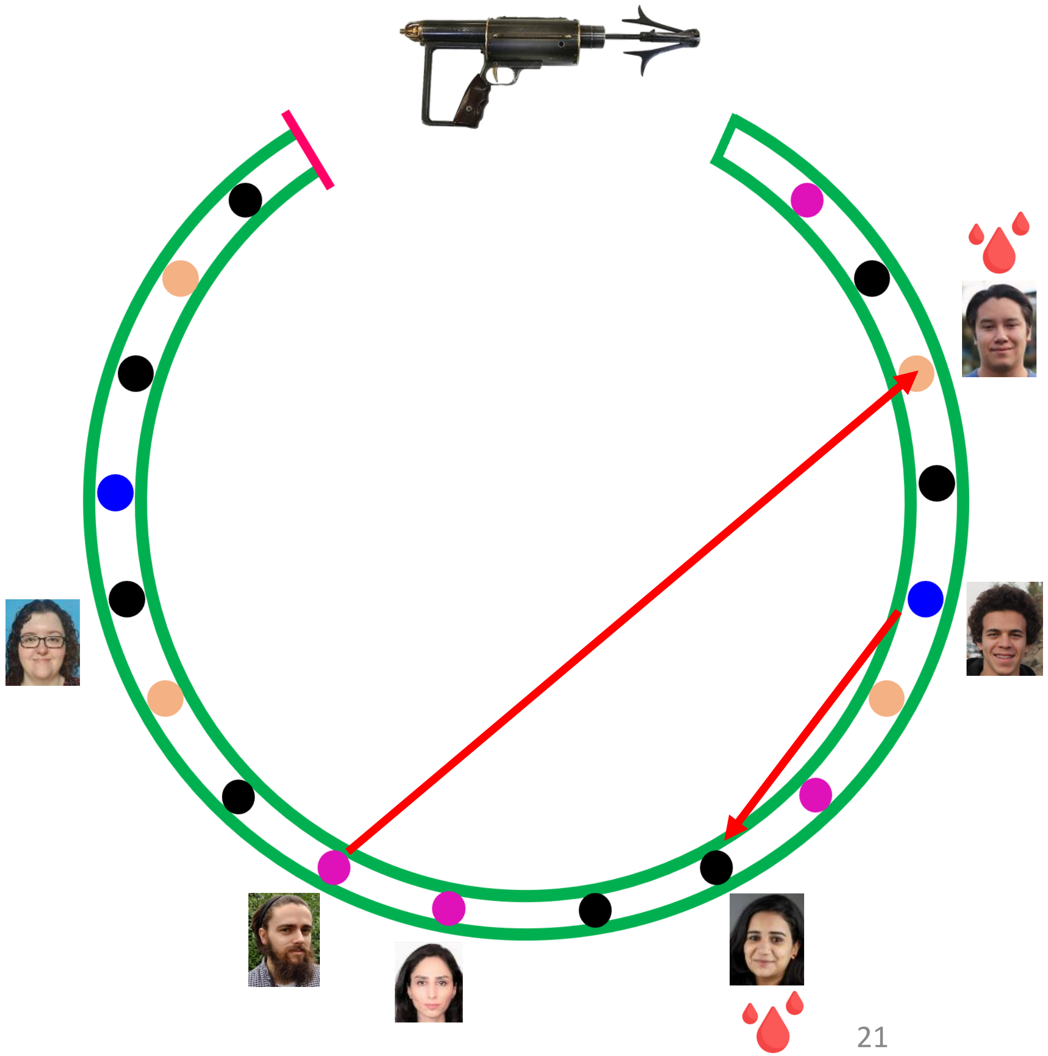
Rules of Hamilton game

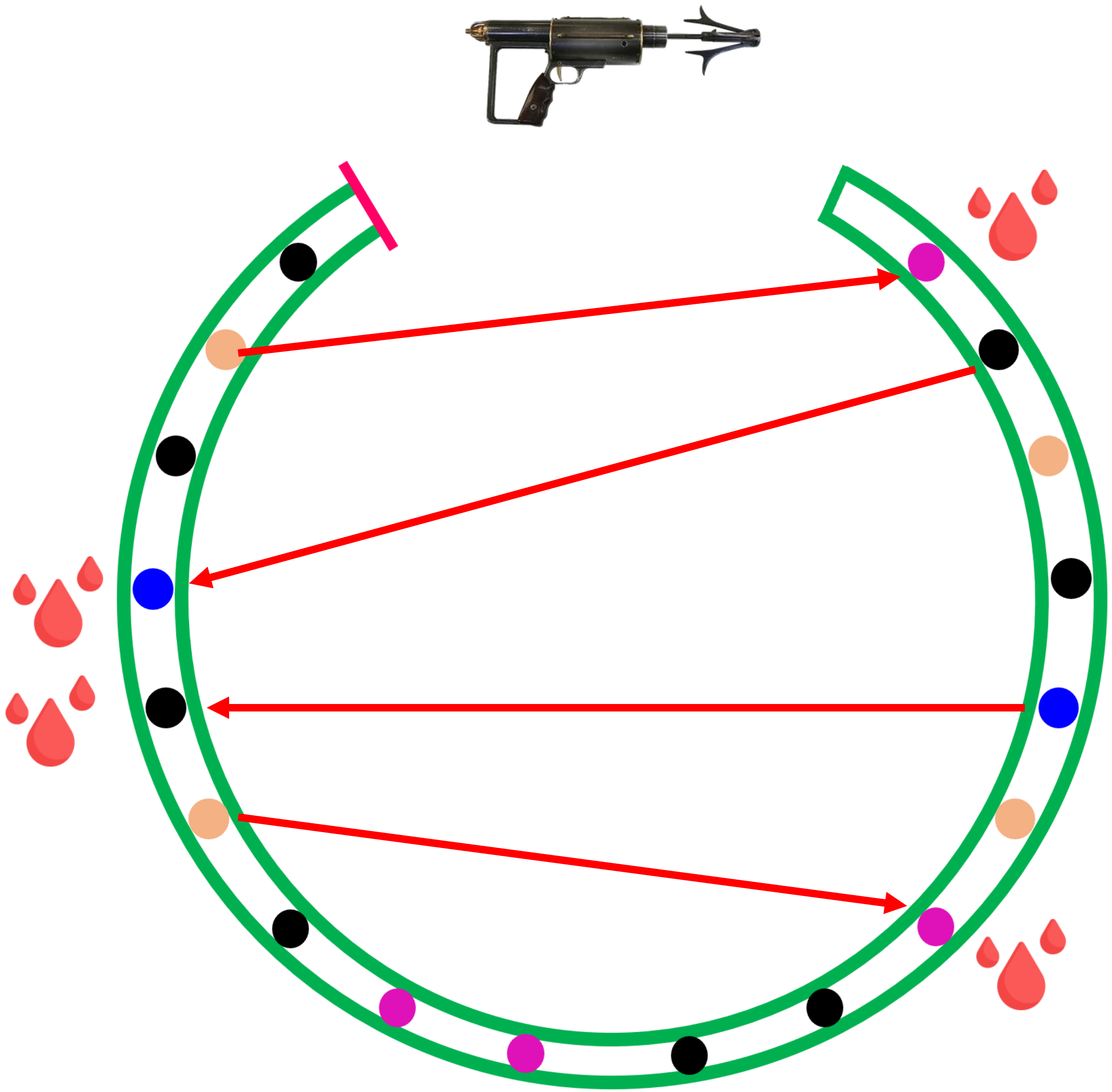
- If you kill, you are safe
- You can only kill one student, you have only one bullet
- You can be killed once
- You must respect other killers, you can't cross their killing lines



Rules of Hamilton game

- If you kill, you are safe
- You can only kill one student, you have only one bullet
- You can be killed once
- You must respect other killers, you can cross their killing line
- Ahmed can only kill, he's an immortal man





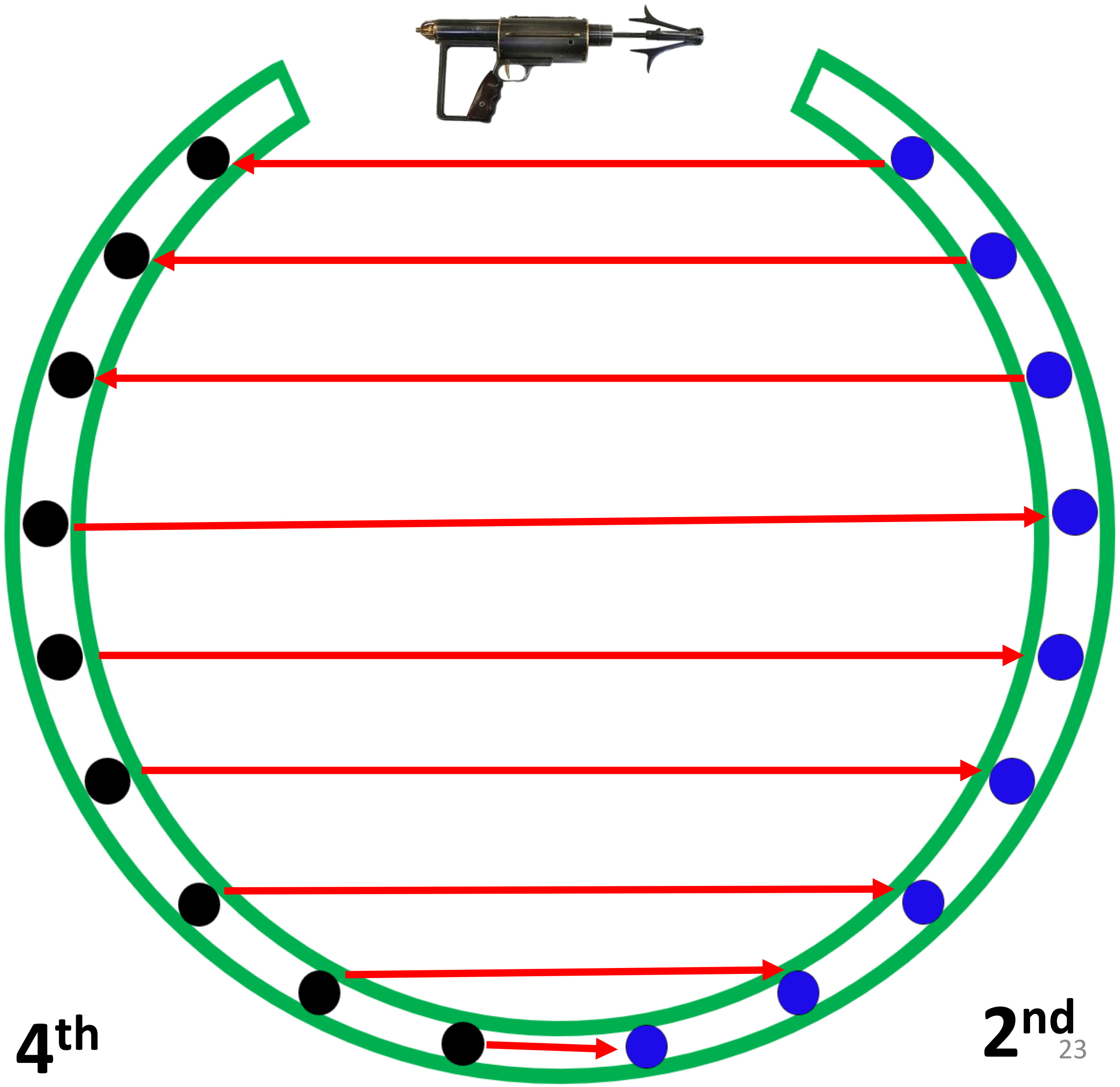
Structure S

$$\# = \binom{9}{0} + \binom{9}{1} + \dots + \binom{9}{9} = 2^9 = 2^{N/2}$$

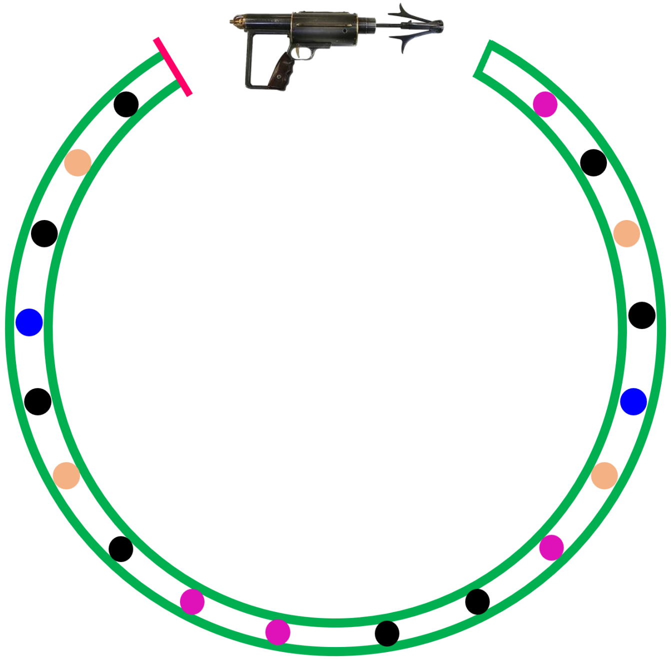
Ω: the set of all possible structures that respect the game rules

Huge

$$N = 18$$

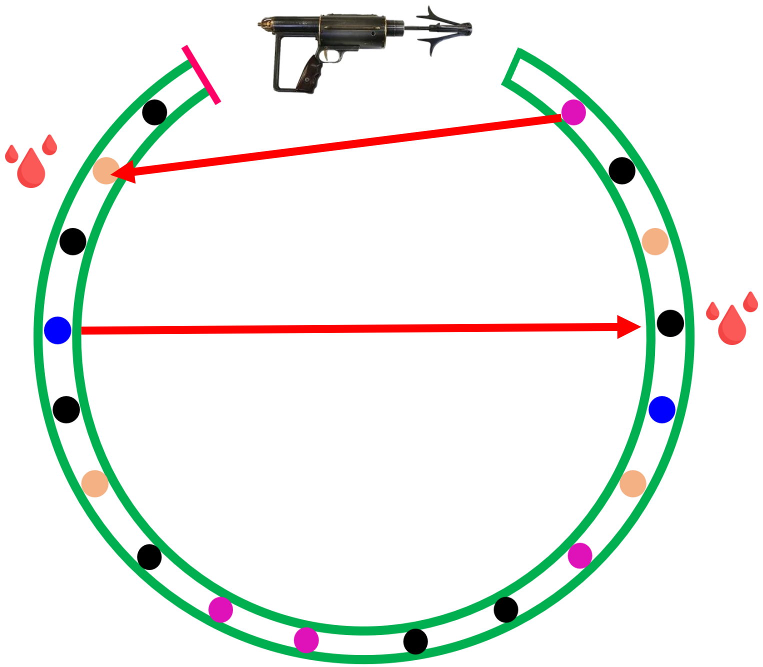


Possible scenarios



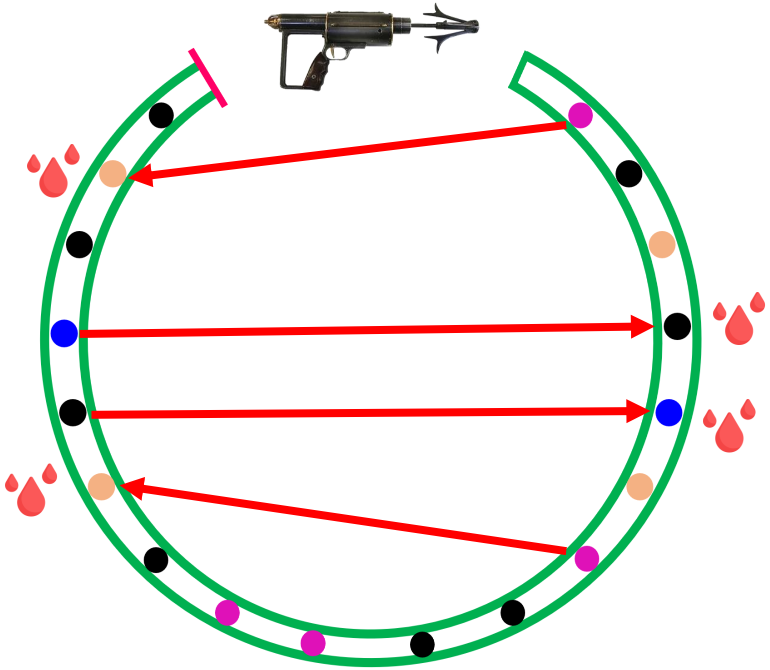
S_1

Not cool



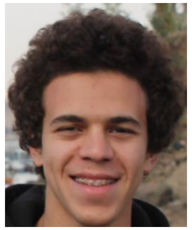
S_2

Getting better



S_3

Much better



Ahmed's goal

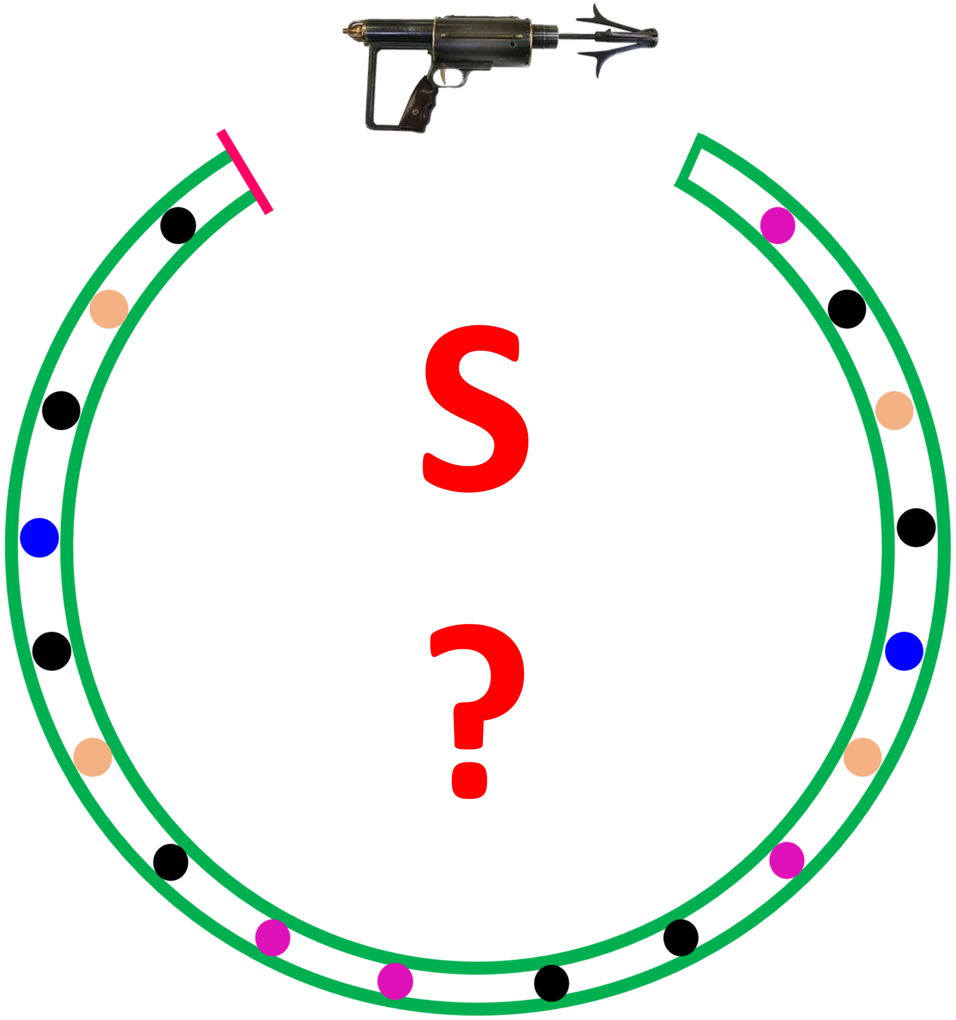
PI

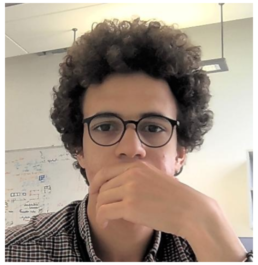
Loves more blood

?

?

?





Some Criteria/Model ?

$$B(S) = \#\text{killed PhDs}$$

$$\max_{S \in \Omega} B(S)$$

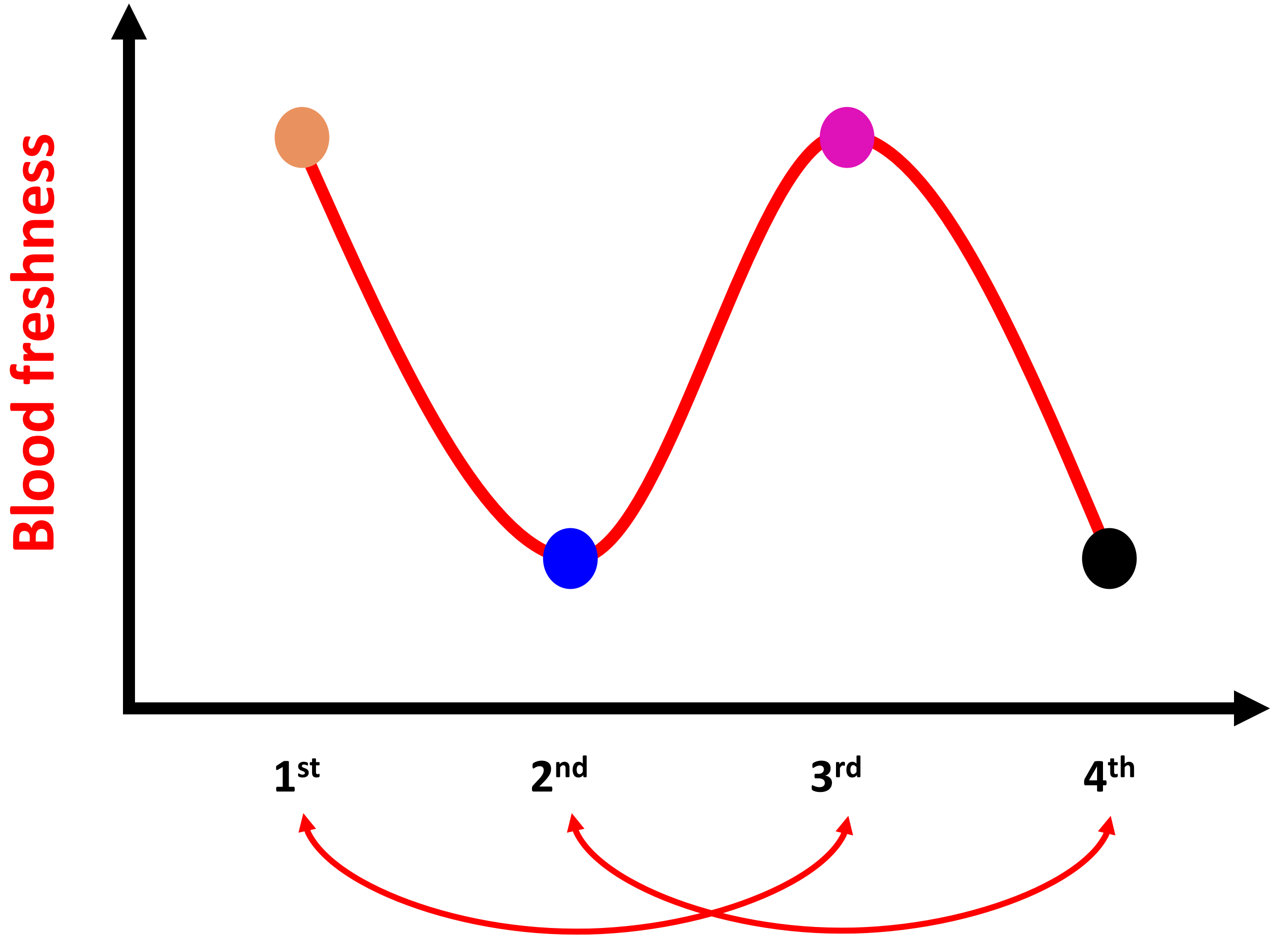
Ω is the set of all possible structures that respect the game rules

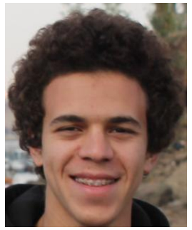
How to compute this fast?

That is so cool!

Level 2

After playing that game over and over





Ahmed’s goal

PI

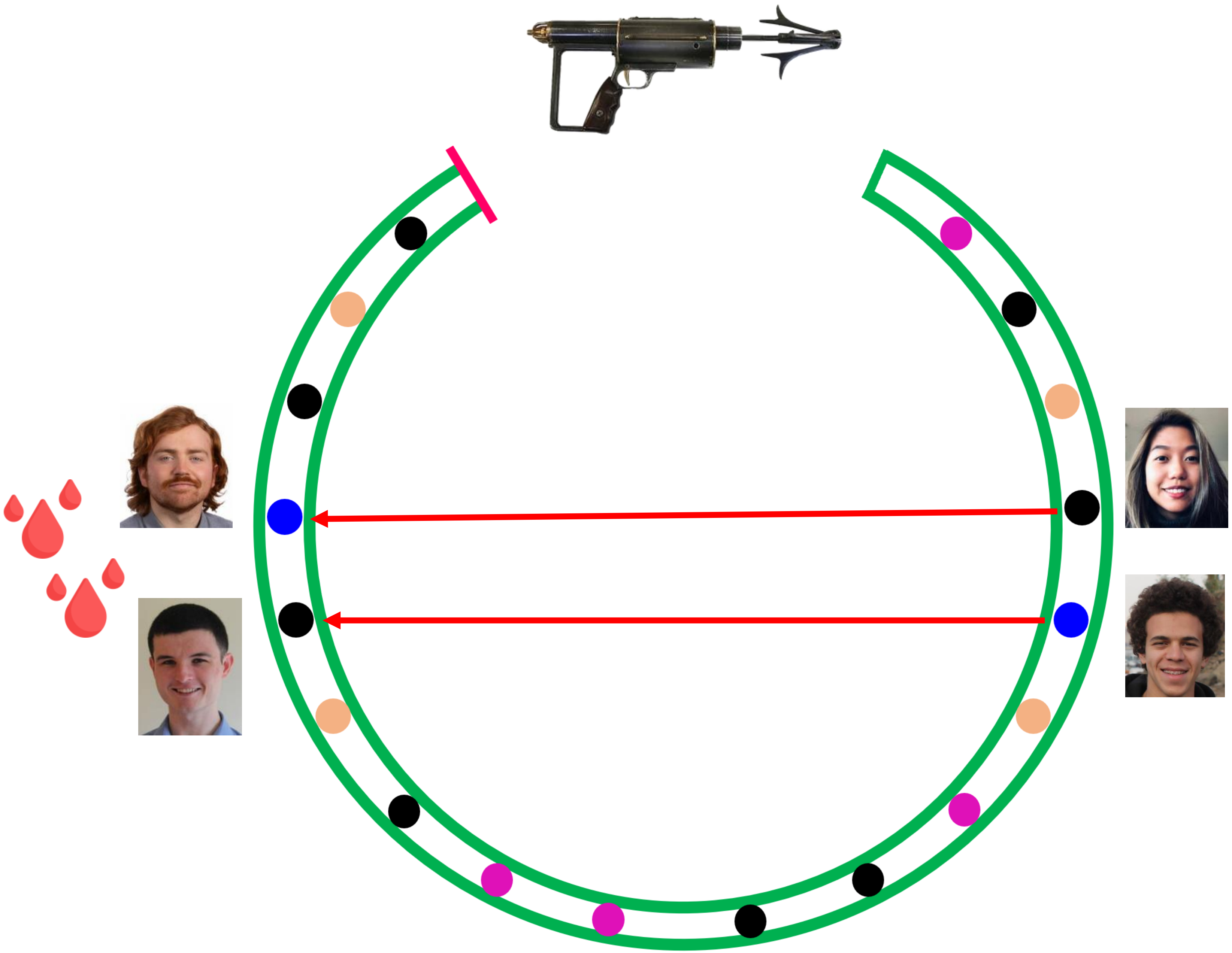
Loves high quality blood

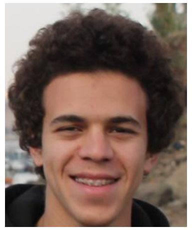
?

?

?

mixed blood





Ahmed’s goal

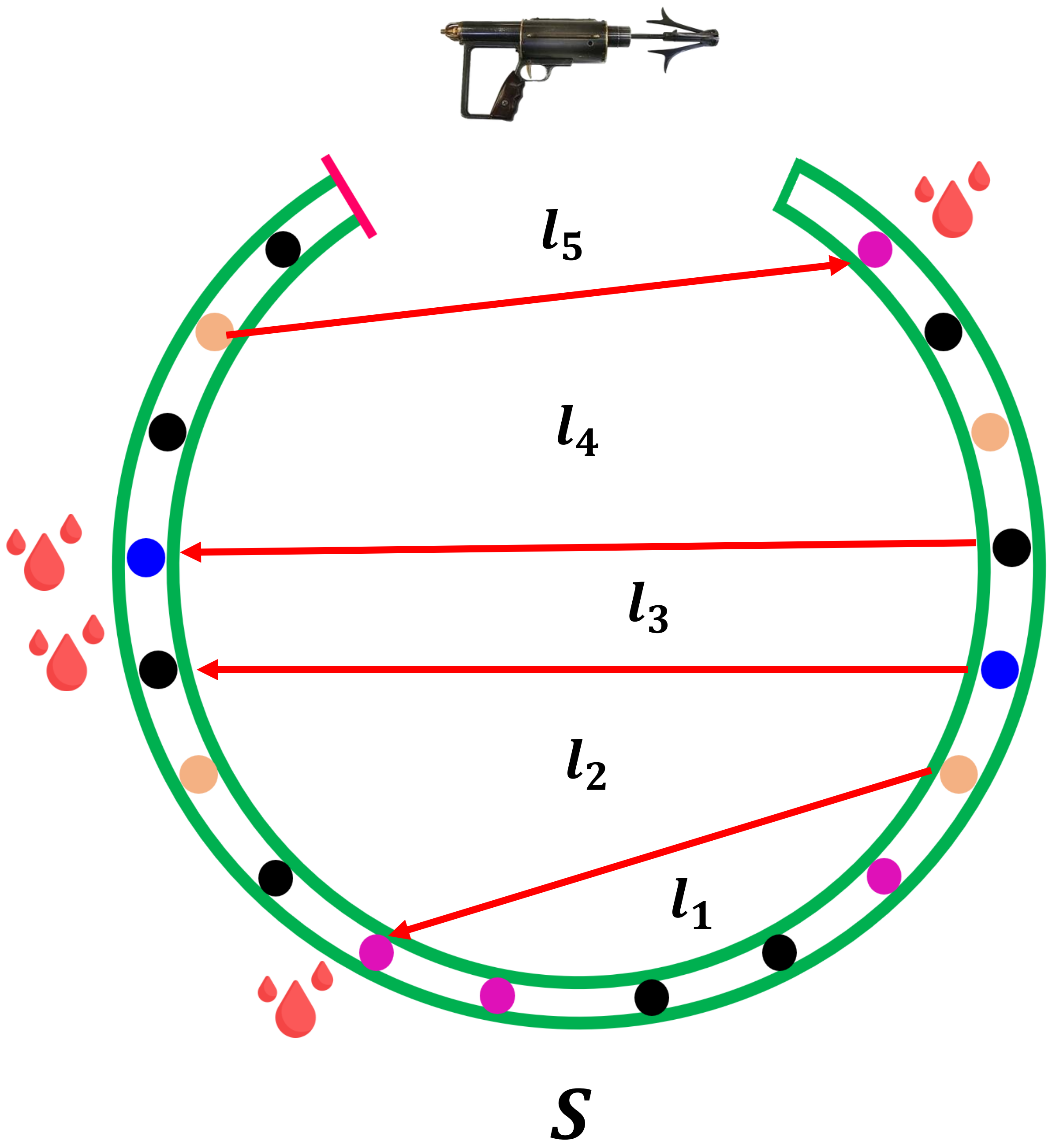
PI

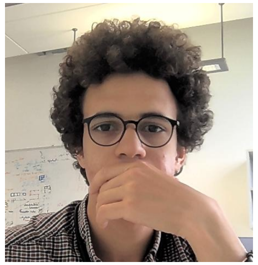
Loves high quality blood

Loves mixed blood

?

?





Some Criteria/Model ?

$$B(S) = \sum_{l} B(l)$$

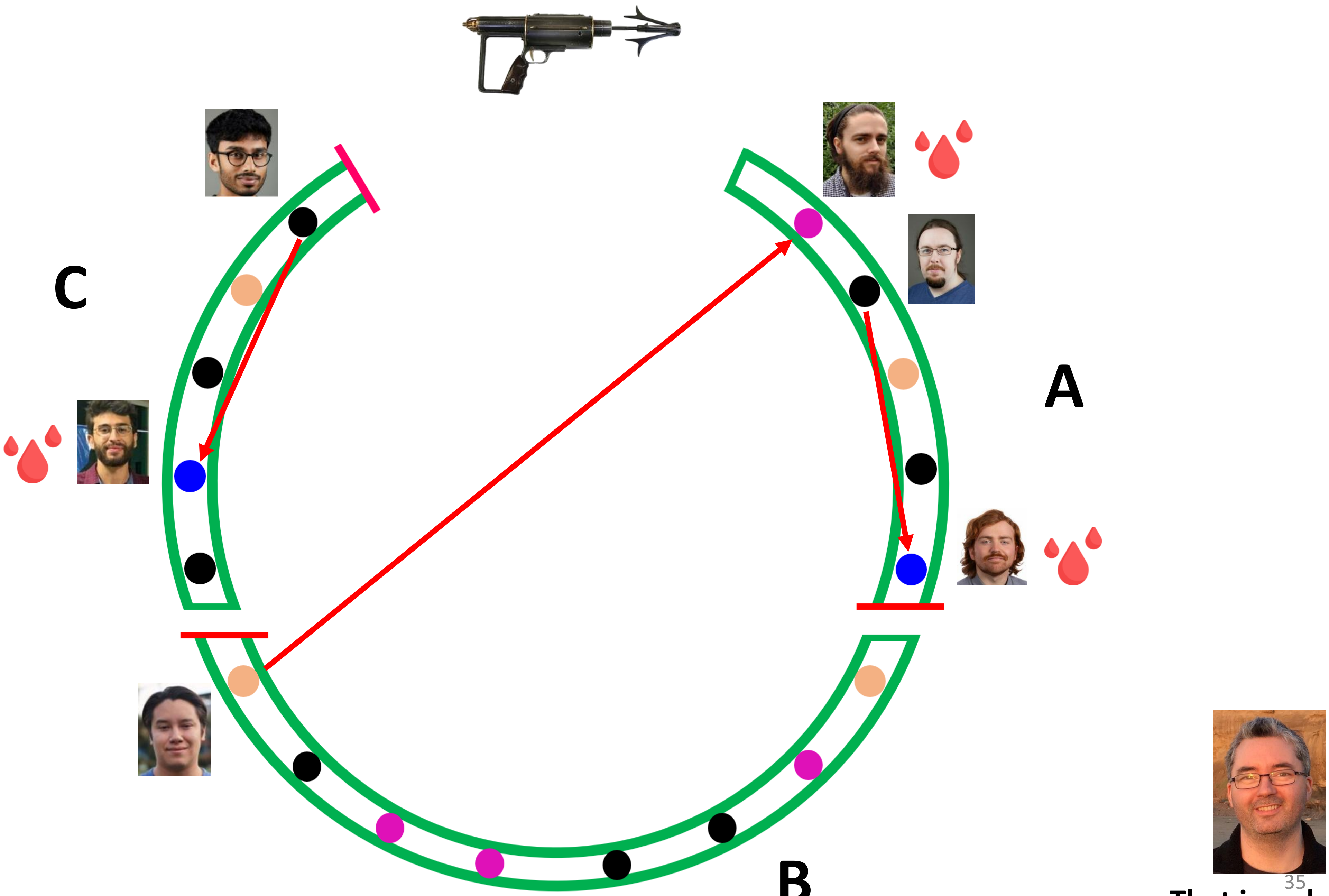
$$\max_{S \in \Omega} B(S)$$

How to compute this fast?

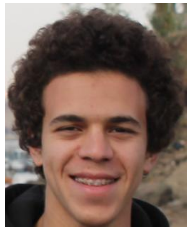
Level 3


FIRE POINT

c = 3 sofas



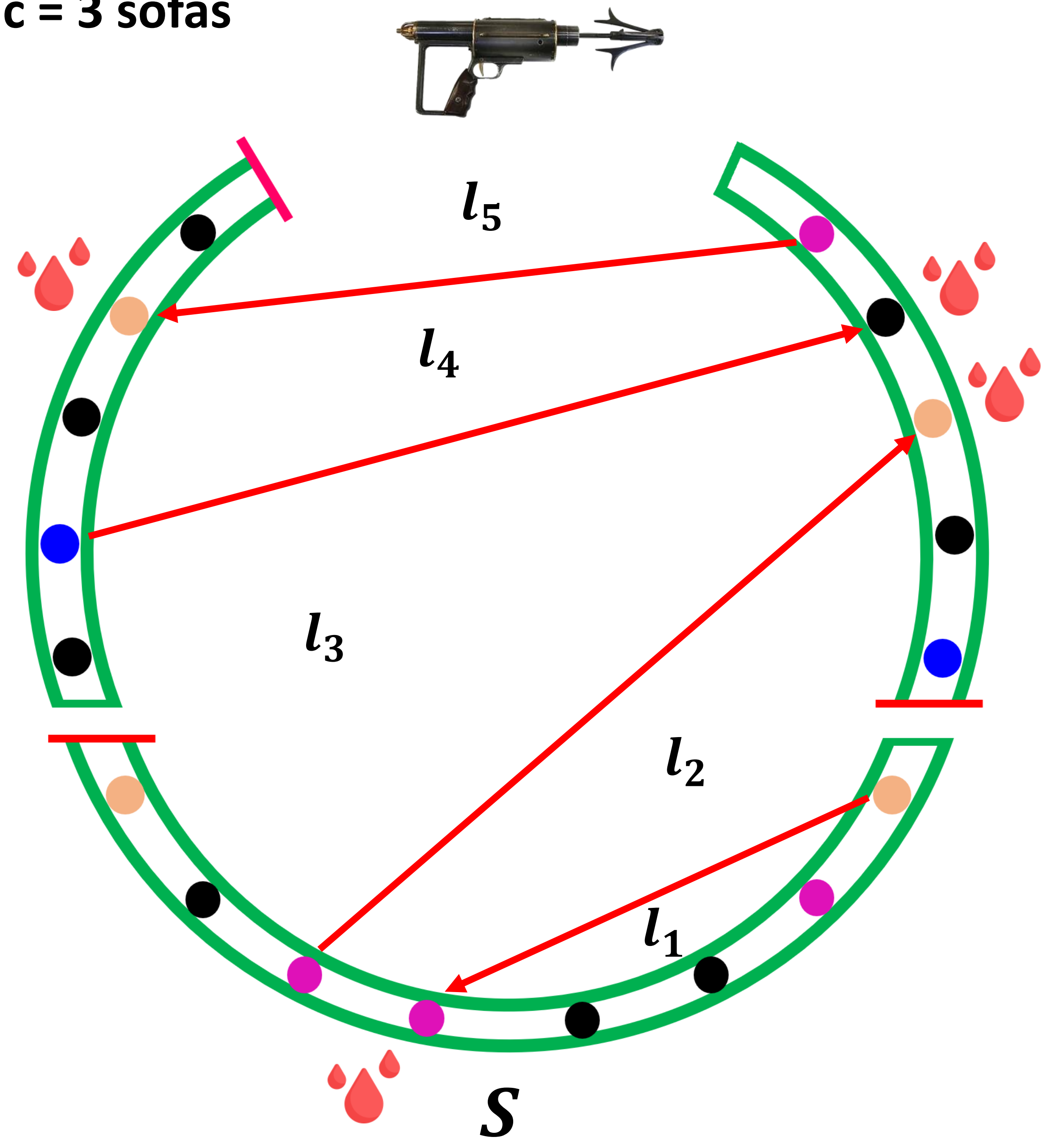
That is so bad!

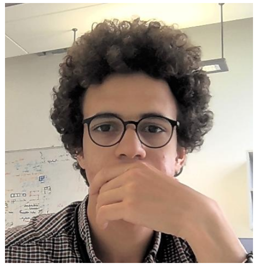


Ahmed's goal

PI
Loves high quality blood
Loves mixed blood
Hates disconnectedness
?

c = 3 sofas





Some Criteria/Model ?

$$B(S) = \sum_{l} B(l) - (c-1)\, B^{\text{assoc}}$$

$$\max_{S \in \Omega} B(S)$$

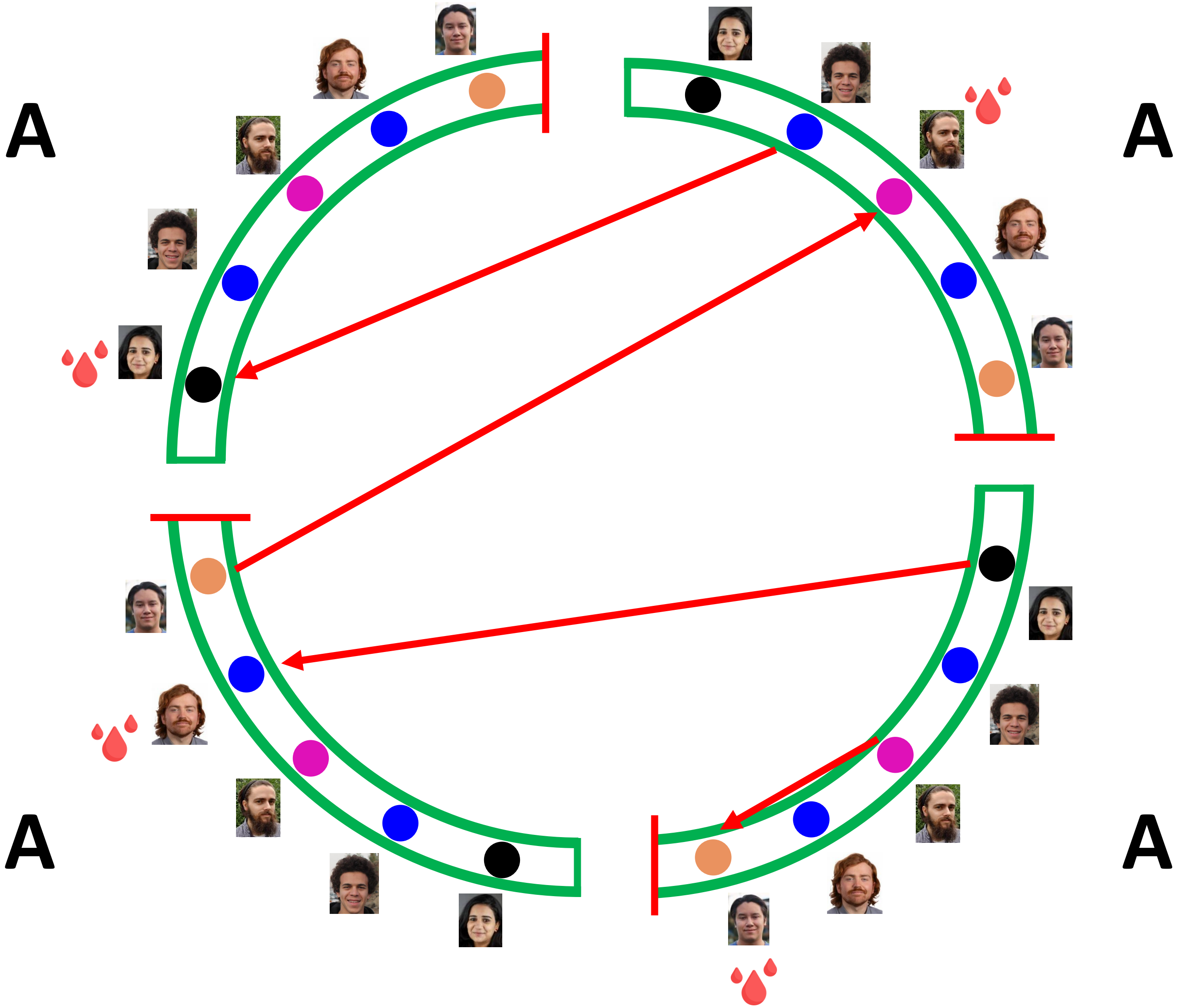
Ω: the set of all connected structures that respect the game rules

How to compute this fast?

Level 4

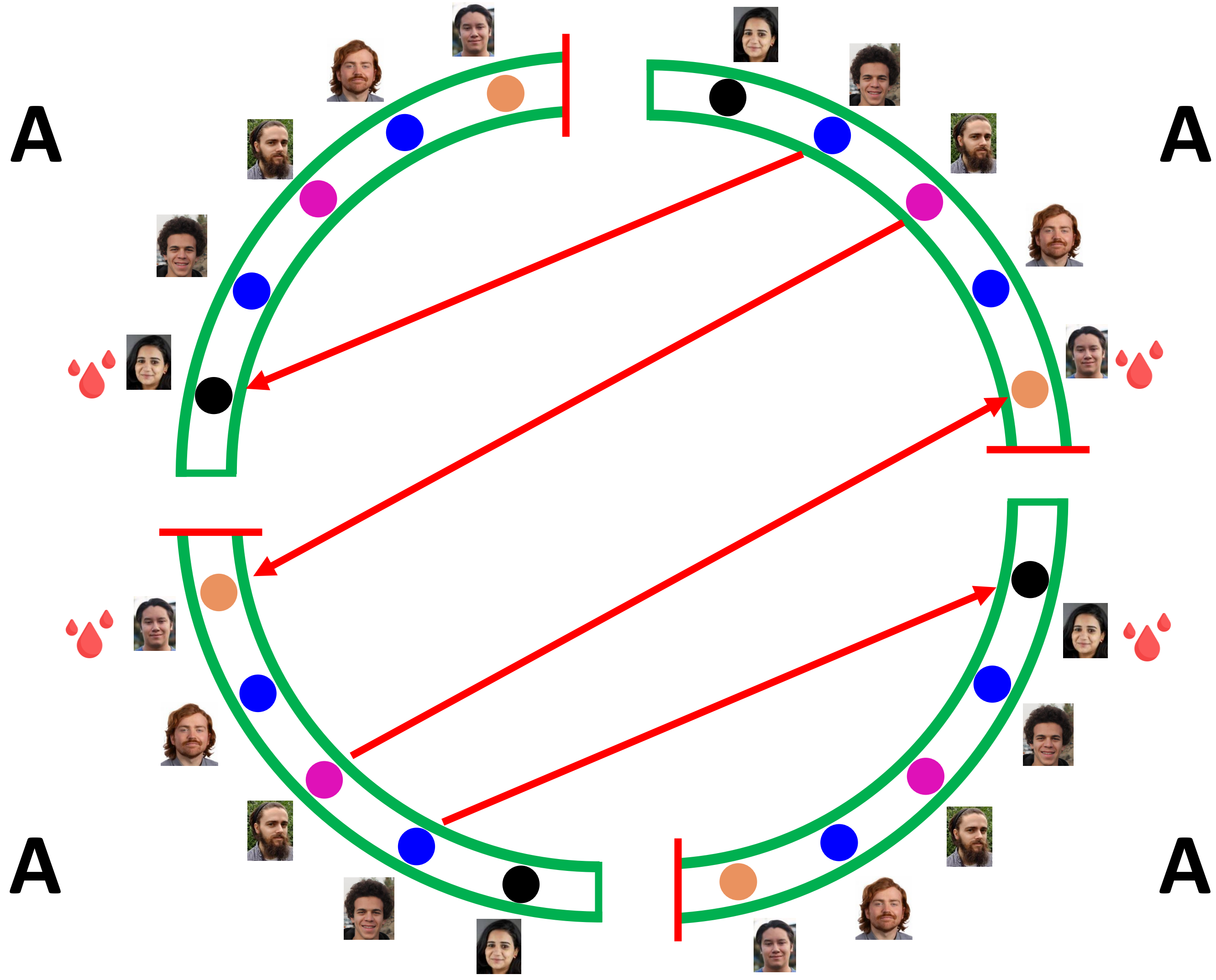

FIRE POINT

c = 4 sofas



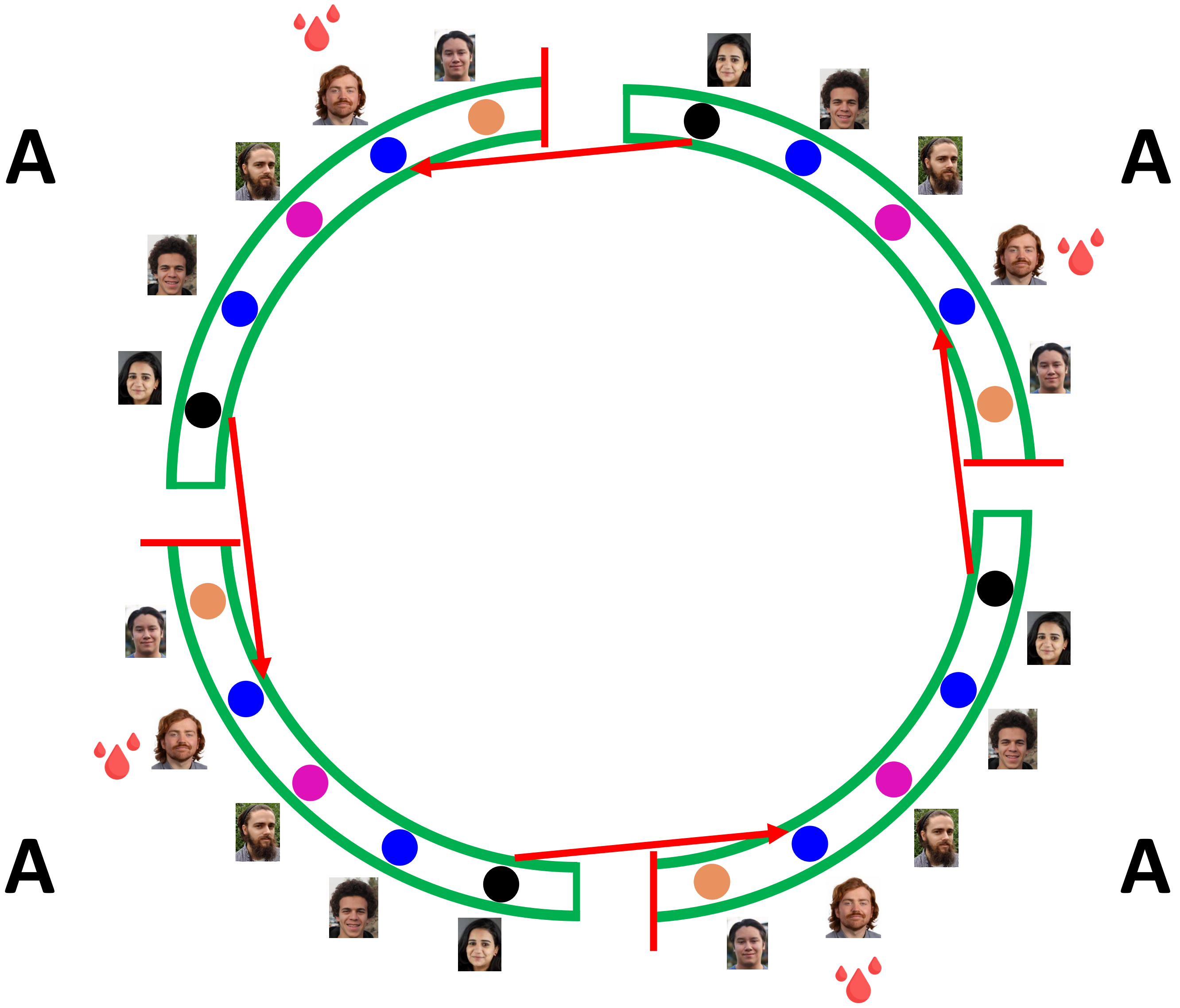
That is ok!

c = 4 sofas



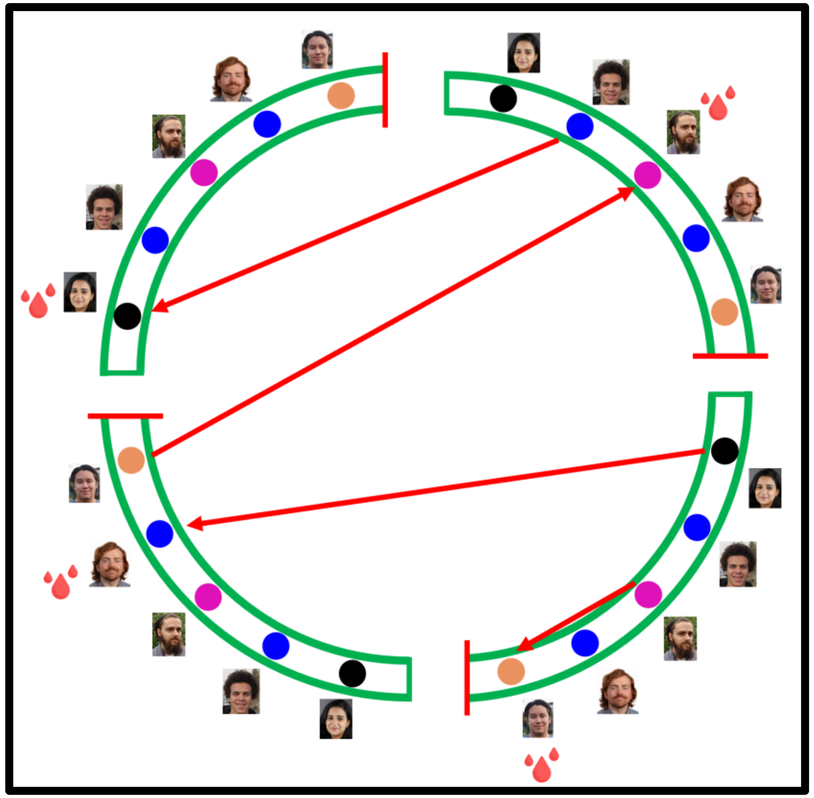
That is ugh!

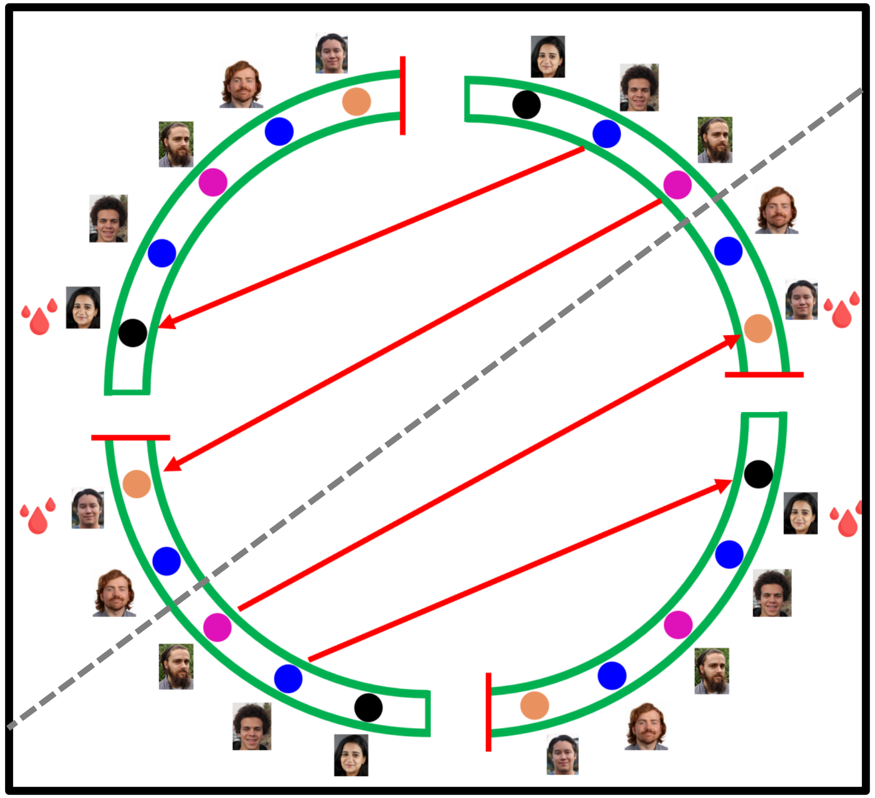
c = 4 sofas



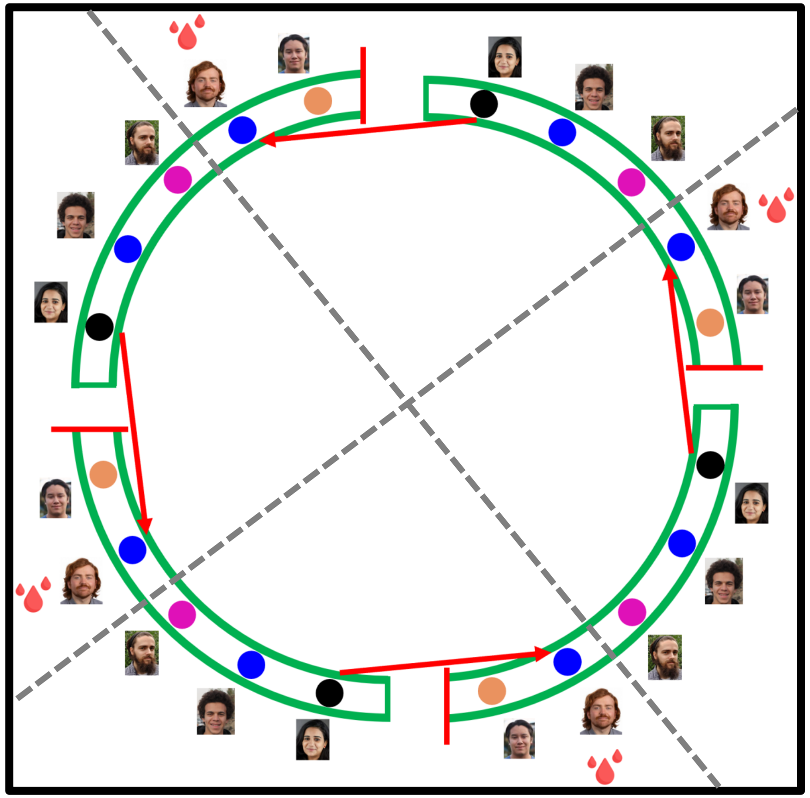
That is ugh ugh!

Let’s analyse this





Rotate by 180 degrees



Rotate by 90 degrees

$R = 1$

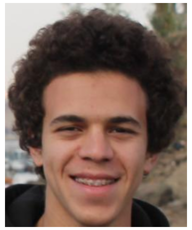
Doesn’t penalize

$R = 2$

Penalize a little bit

$R = 4$

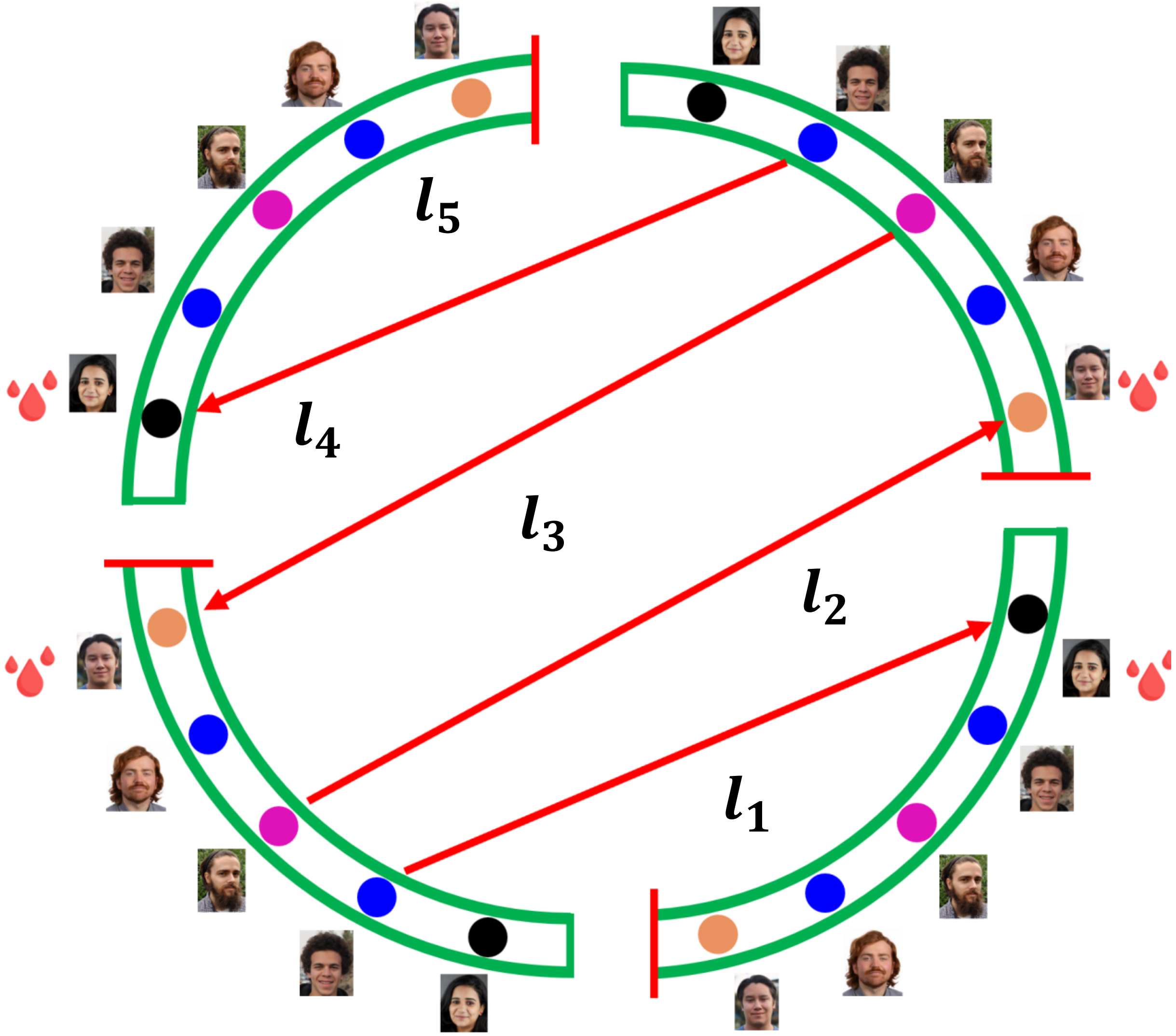
Penalize more

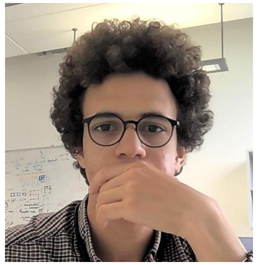


Ahmed’s goal

PI
Loves high quality blood
Loves mixed blood
Hates disconnectedness
Hates rotational symmetry

c = 4 sofas





Some Criteria/Model ?

$$B(S) = \sum_{l} B(l) - (c-1)B^{\text{assoc}} - k_B T * \log R$$

$$\max_{S \in \Omega} B(S)$$

How to compute this fast?

We are done

Why, Ahmed?

Hamilton game

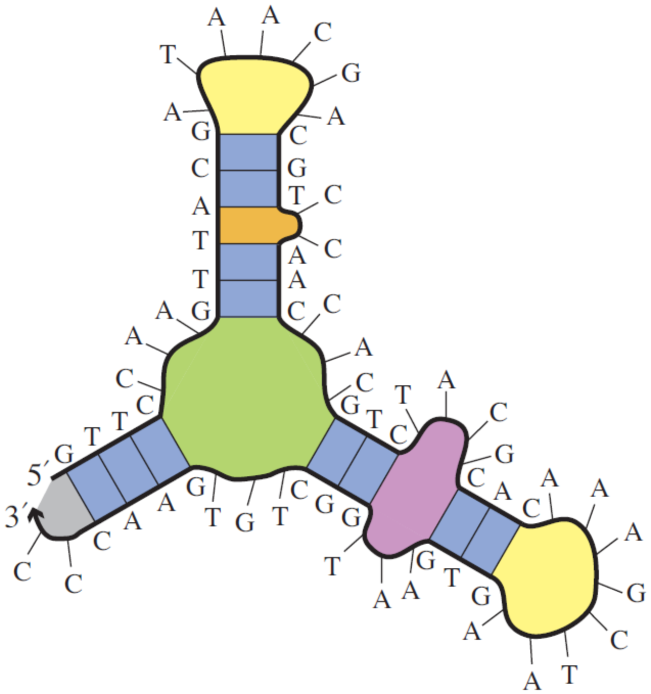
DNA secondary structures

First year	Second year	Third year	fourth year
G	A	C	T
Daniel Augustina . . .	Ahmed Andre Paddy Cormac .	Dara Solmaz Oluwayomi .	Akash Emma Darshana . .

Chemical bonds

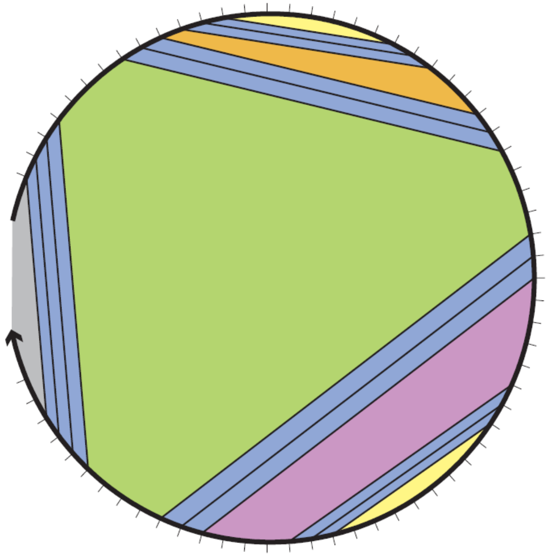
DNA secondary structure

Single stranded DNA



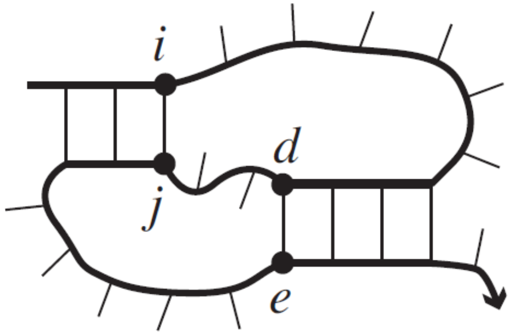
Secondary structure
=
A list of base pairs

pseudoknot-free

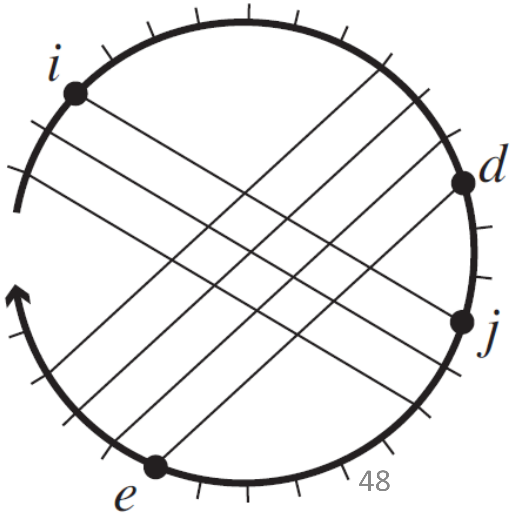


Polymer graph representation

NP – Hard

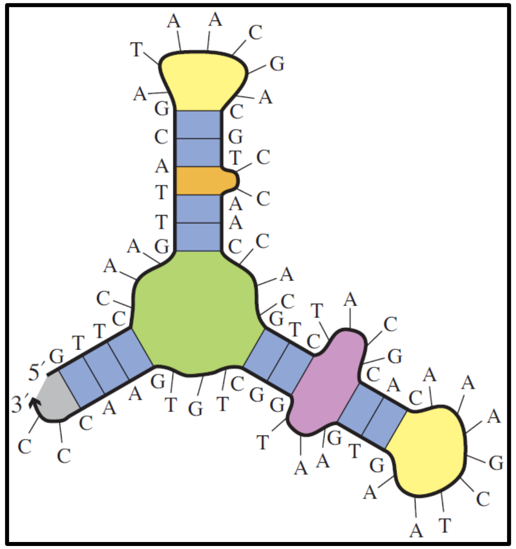


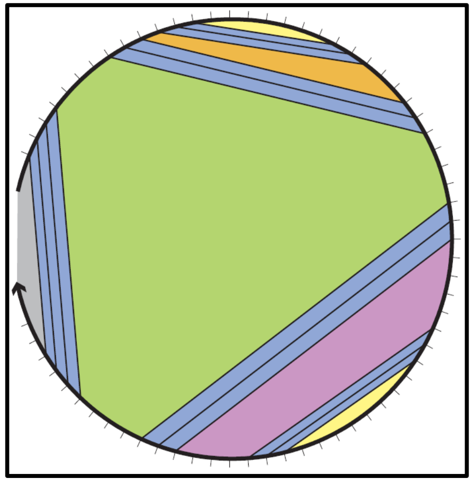
pseudoknotted



Energy models and Minimum Free Energy

Single stranded system



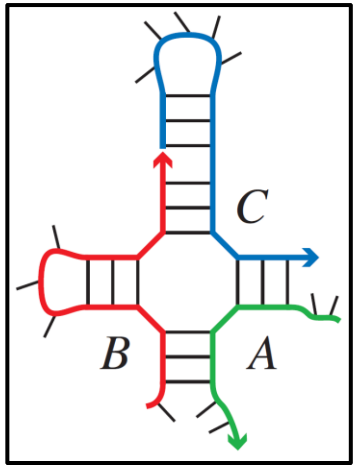


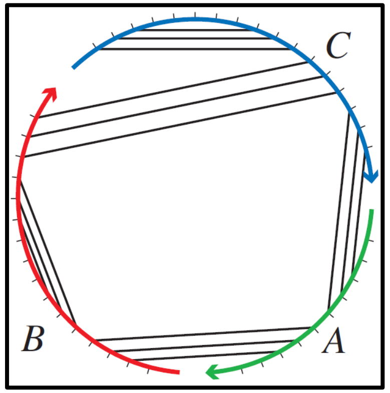
$\Delta G(S)$

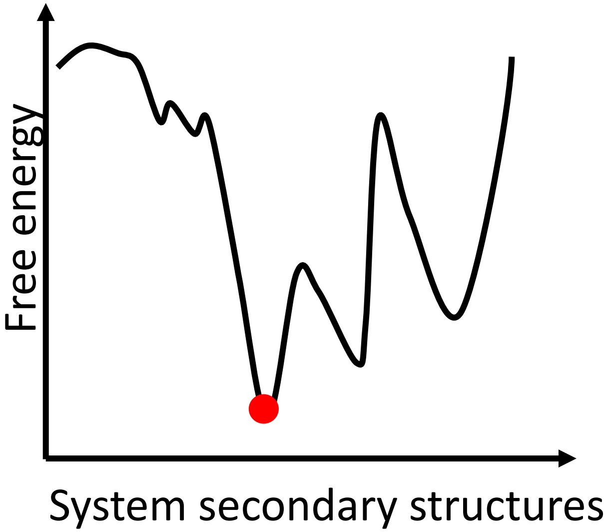
Energy model

Capture the free energy of secondary structure

Multi stranded system of s strands

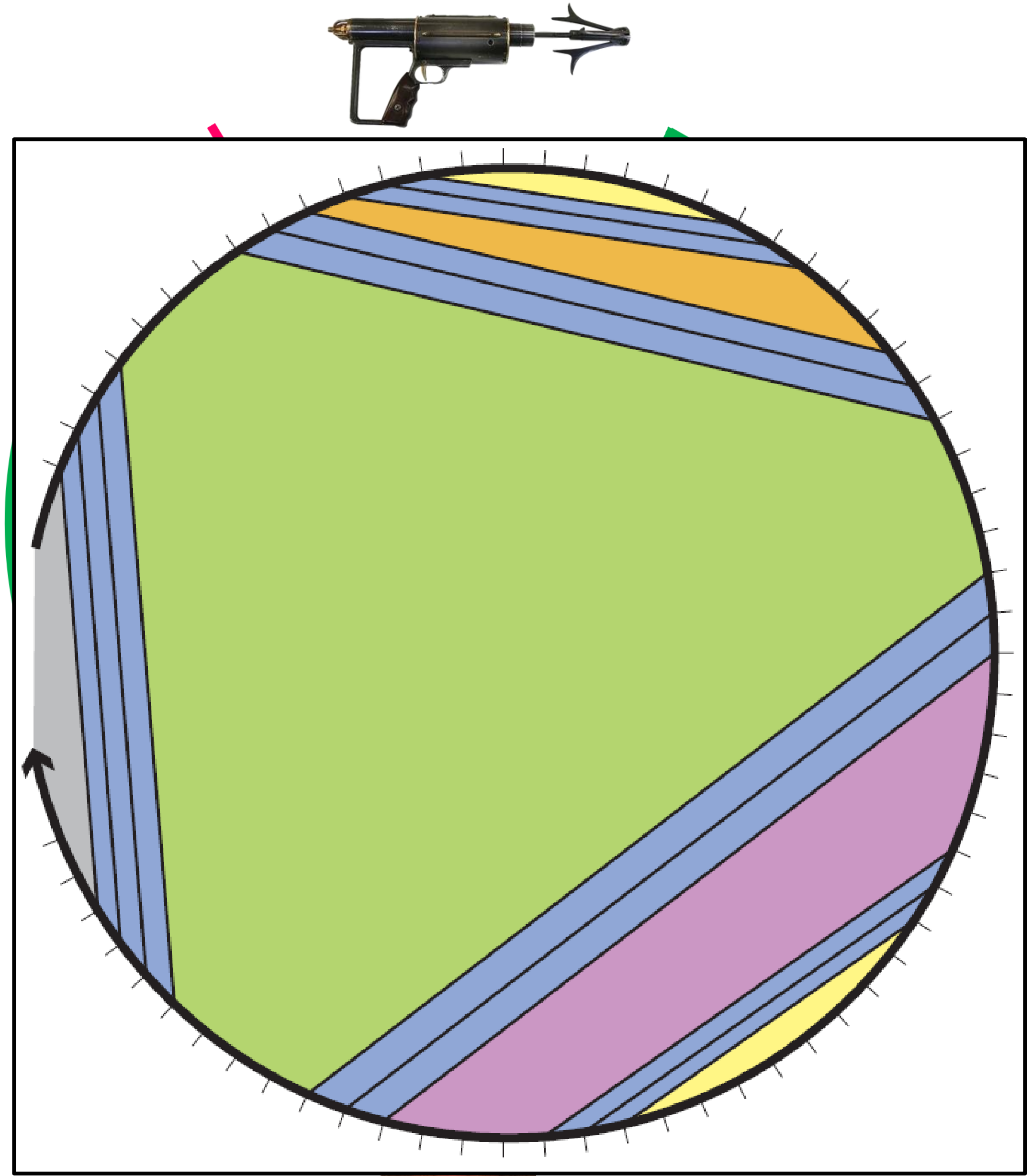






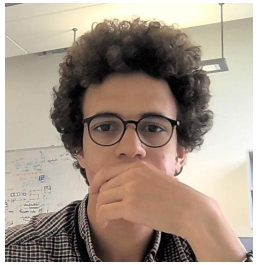
$$\text{MFE} = \min_{S \in \Omega} \Delta G(S)$$

Minimum Free Energy



That is so cool!

1



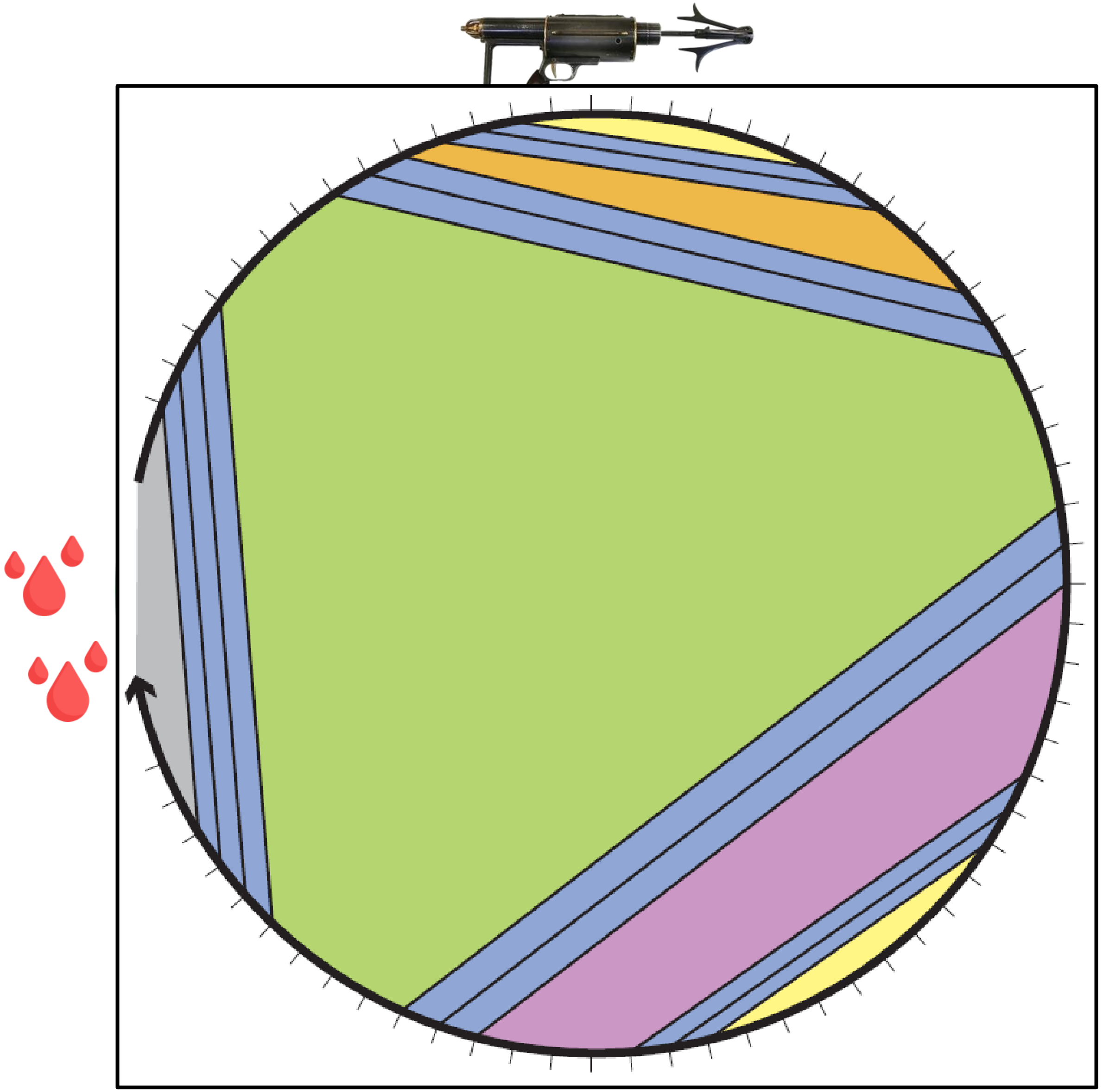
Some Criteria/Model ?

$$\Delta G(S) = -\#\text{base pairs}$$

$$\min_{S \in \Omega} \Delta G(S)$$

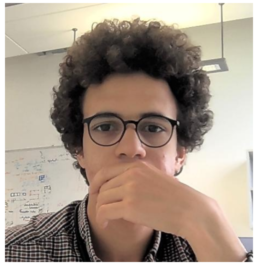
Ω is the set of all possible structures that respect the game rules

How to compute this fast? **Yes**



S

2



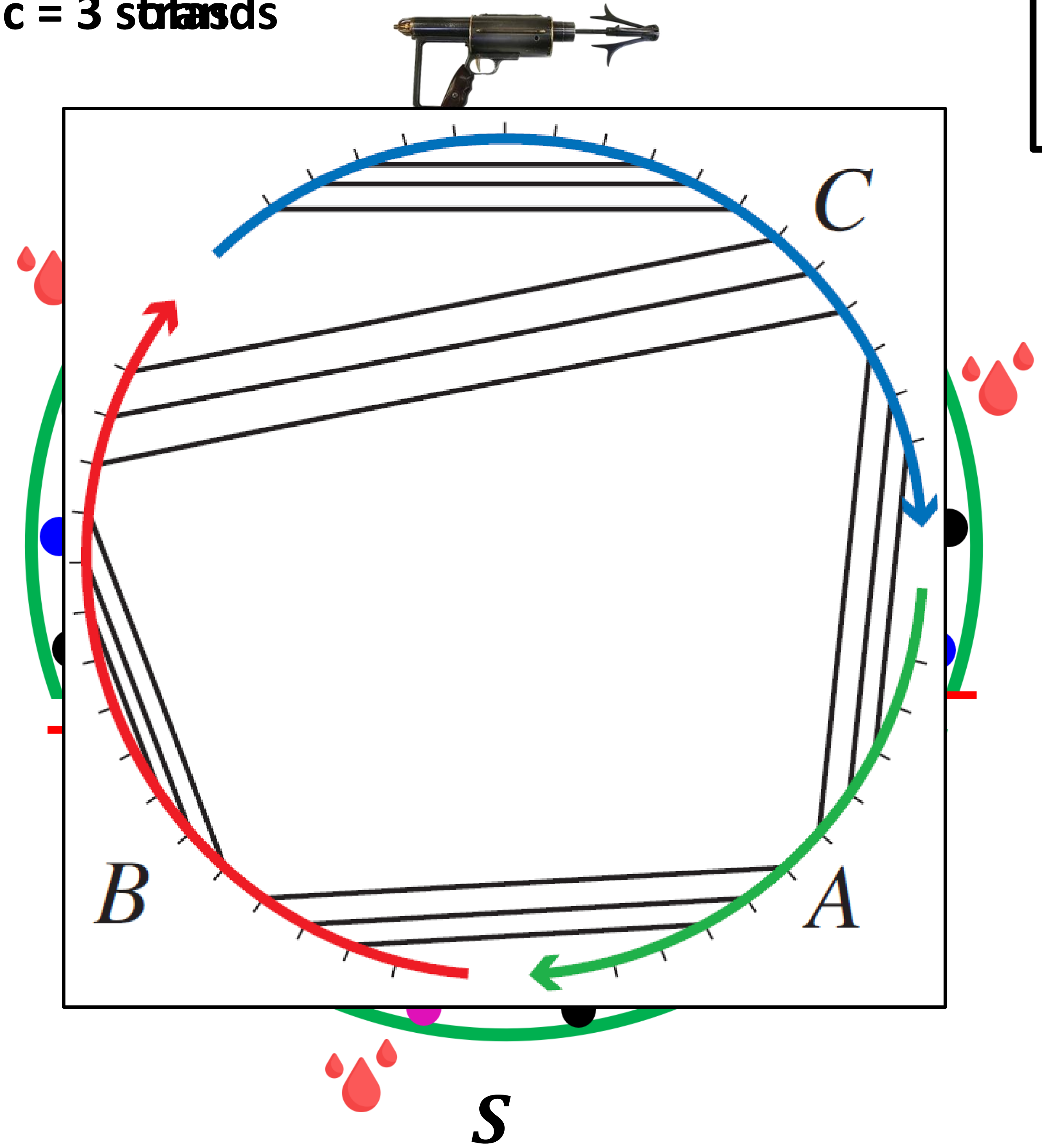
Some Criteria/Model ?

$$\Delta G(S) = \sum_{l} \Delta G(l)$$

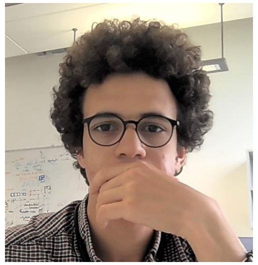
$$\min_{S \in \Omega} \Delta G(S)$$

How to compute this fast? Yes

c = 3 strands



3



Some Criteria/Model ?

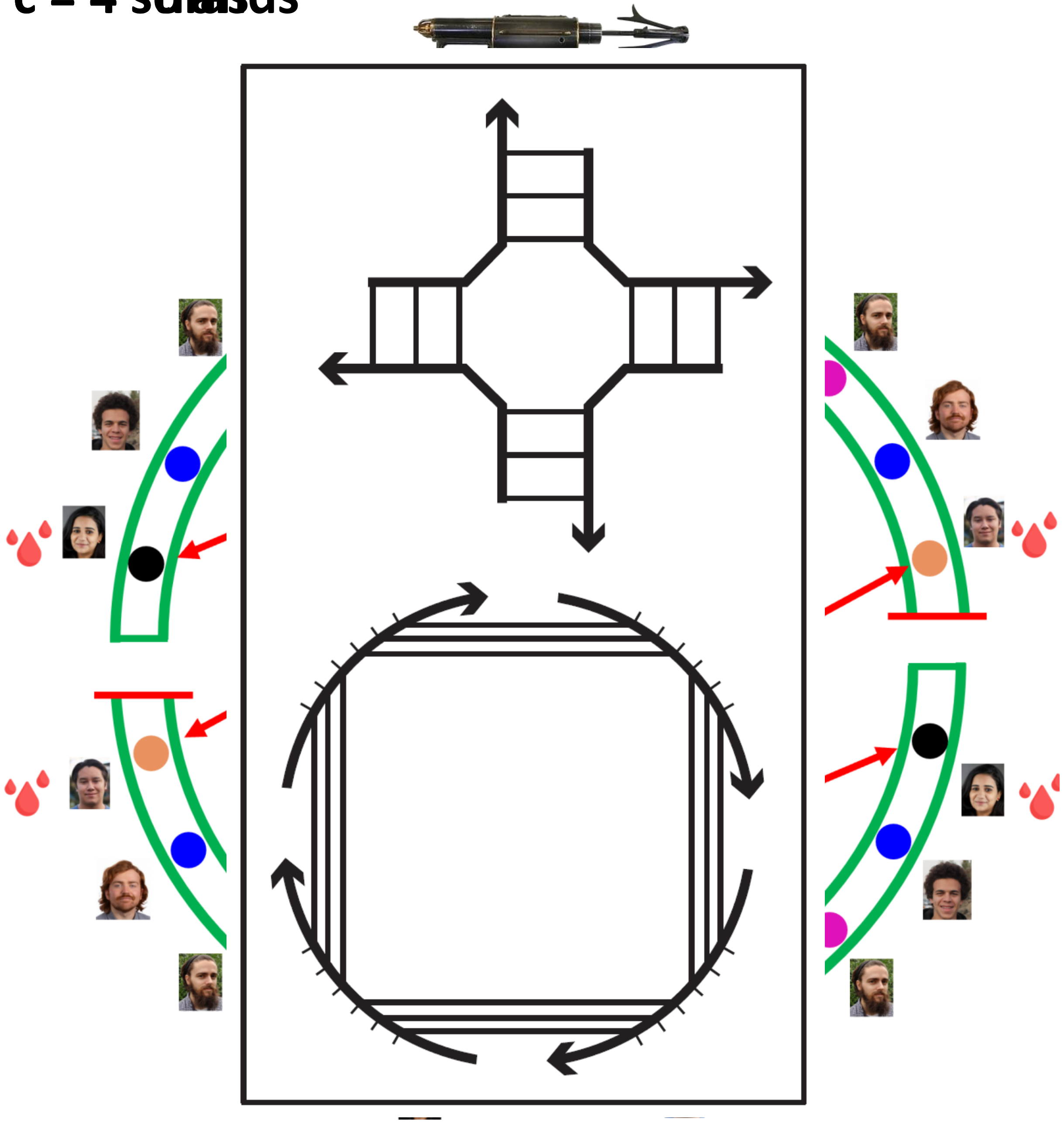
$$\Delta G(S) = \sum_{l} \Delta G(l) + (c-1)\,\Delta G^{\text{assoc}}$$

$$\min_{S \in \Omega} \Delta G(S)$$

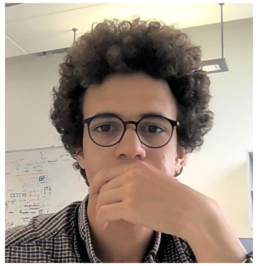
Ω: the set of all connected structures that respect the game rules

How to compute this fast? Yes

c = 4 strands



4



Some Criteria/Model ?

$$\Delta G(S) = \sum_{l} \Delta G(l) + (c-1)\Delta G^{\text{assoc}} + k_B T * \log R$$

$$\min_{S \in \Omega} \Delta G(S)$$

How to compute this fast?

No, till now

Free energy — Loop energy — Entropic association cost — Symmetry penalty

$$\underbrace{\Delta G(S)}_{\text{Free energy}} = \underbrace{\sum_{l \in S} \Delta G(l)}_{\text{Loop energy}} + \underbrace{(c-1) * \Delta G^{\text{assoc}}}_{\text{Entropic association cost}} + \underbrace{k_B T * \log R}_{\text{Symmetry penalty}}$$

− + +

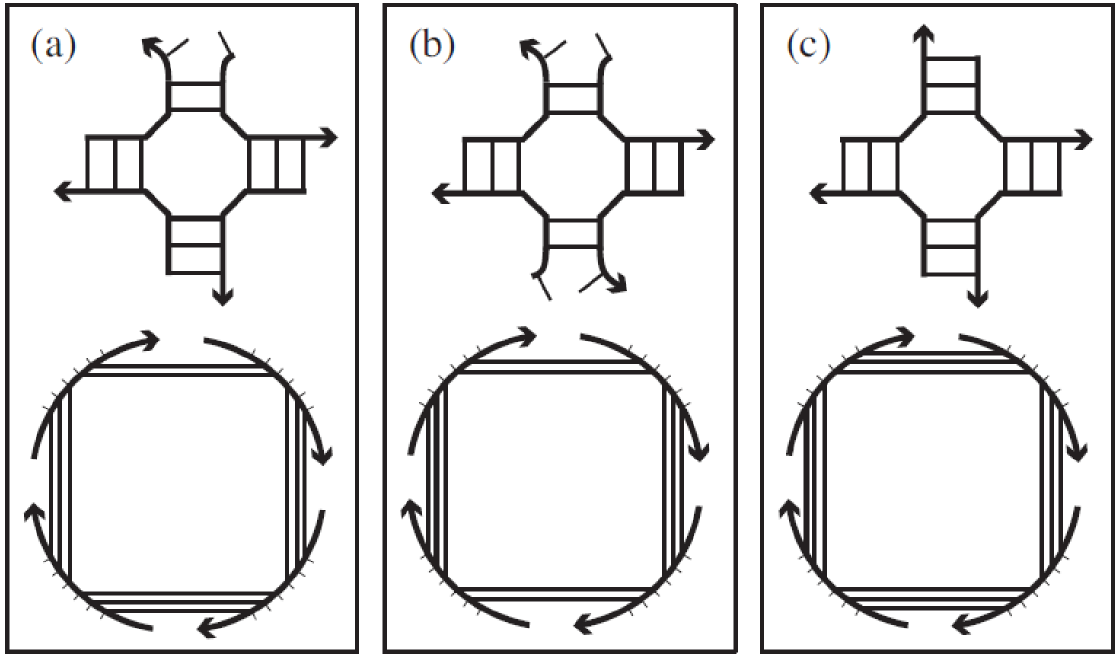


Fig. 2.2 *Sample secondary structures and polymer graphs for a complex of four indistinguishable strands.* (a) *1-fold (i.e., no) rotational symmetry.* (b) *2-fold rotational symmetry.* (c) *4-fold rotational symmetry.*

Computational complexity of Minimum Free Energy algorithms

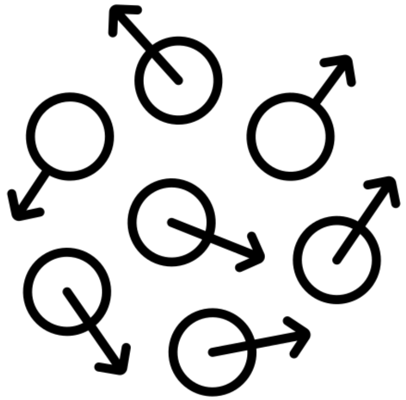
Level	Input Type	MFE
1	Single Strand (Maximum matching)	$O(N^3)$
2	Single Strand (Loop model)	$O(N^3)$
3	Multiple **unique** Strands, Bounded ($\leq$ c)	$O(N^3(c-1)!)$
4	Multiple Strands, Bounded ($\leq$ c)	?

N bases, c strands

Open problem for $\approx$ 20 years

Why symmetry makes that difference?

Entropy

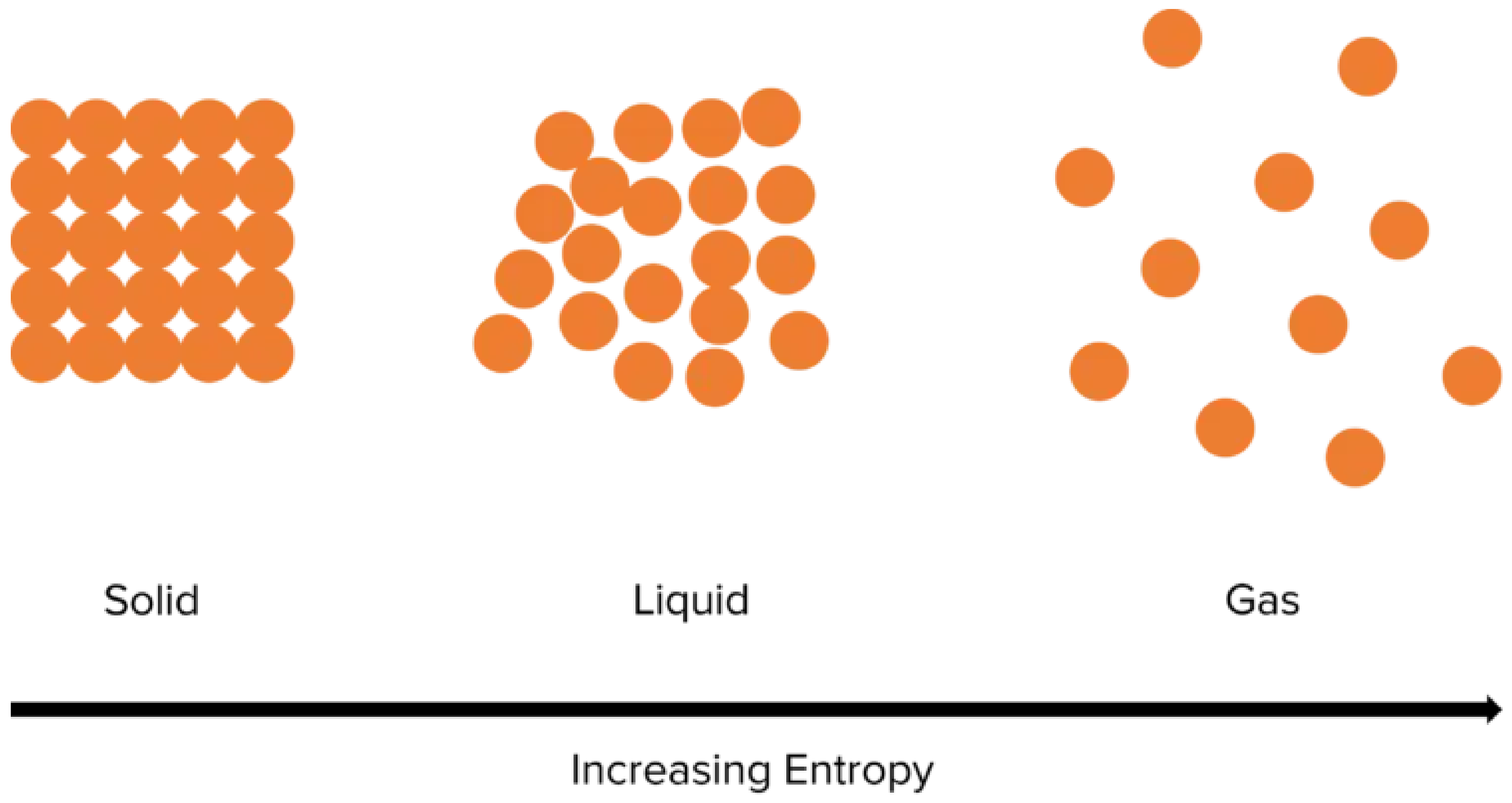


ΔG

Free energy

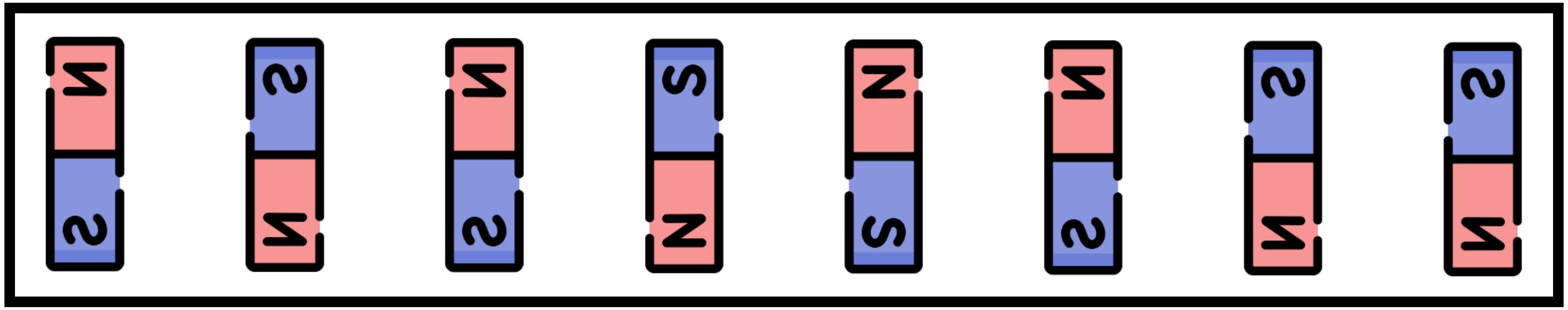
Enthalpy — **H**

Entropy — **S**


Solid
Liquid
Gas
Increasing Entropy

$$S = k_B \log \Pi$$

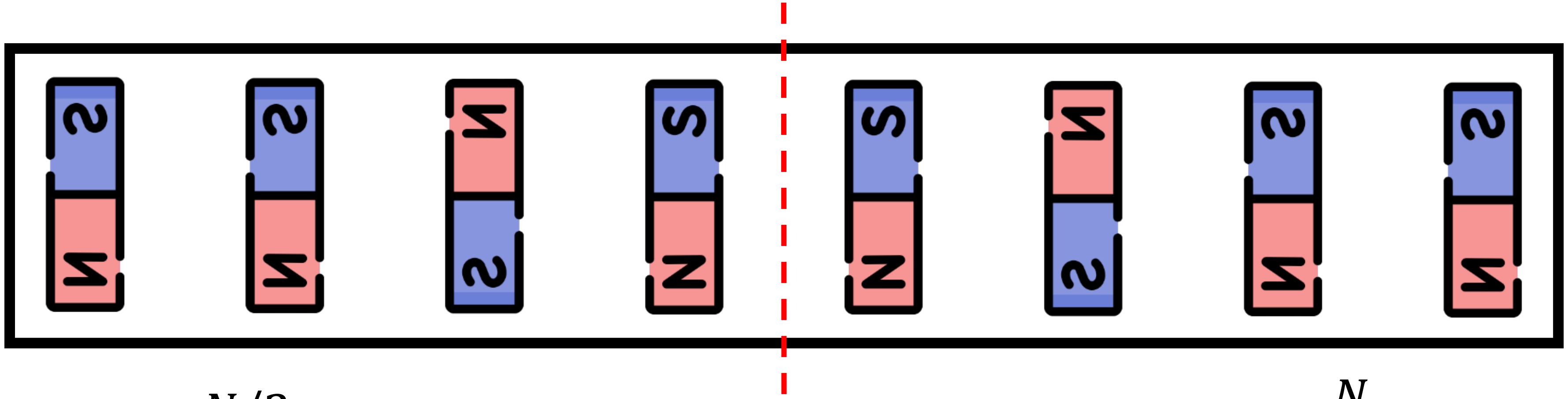
$\boldsymbol{N}$



The total number of states of the N magnets is $\Pi = 2^N$

$$S = k_B N \log 2$$

$\boldsymbol{\frac{N}{2}}$

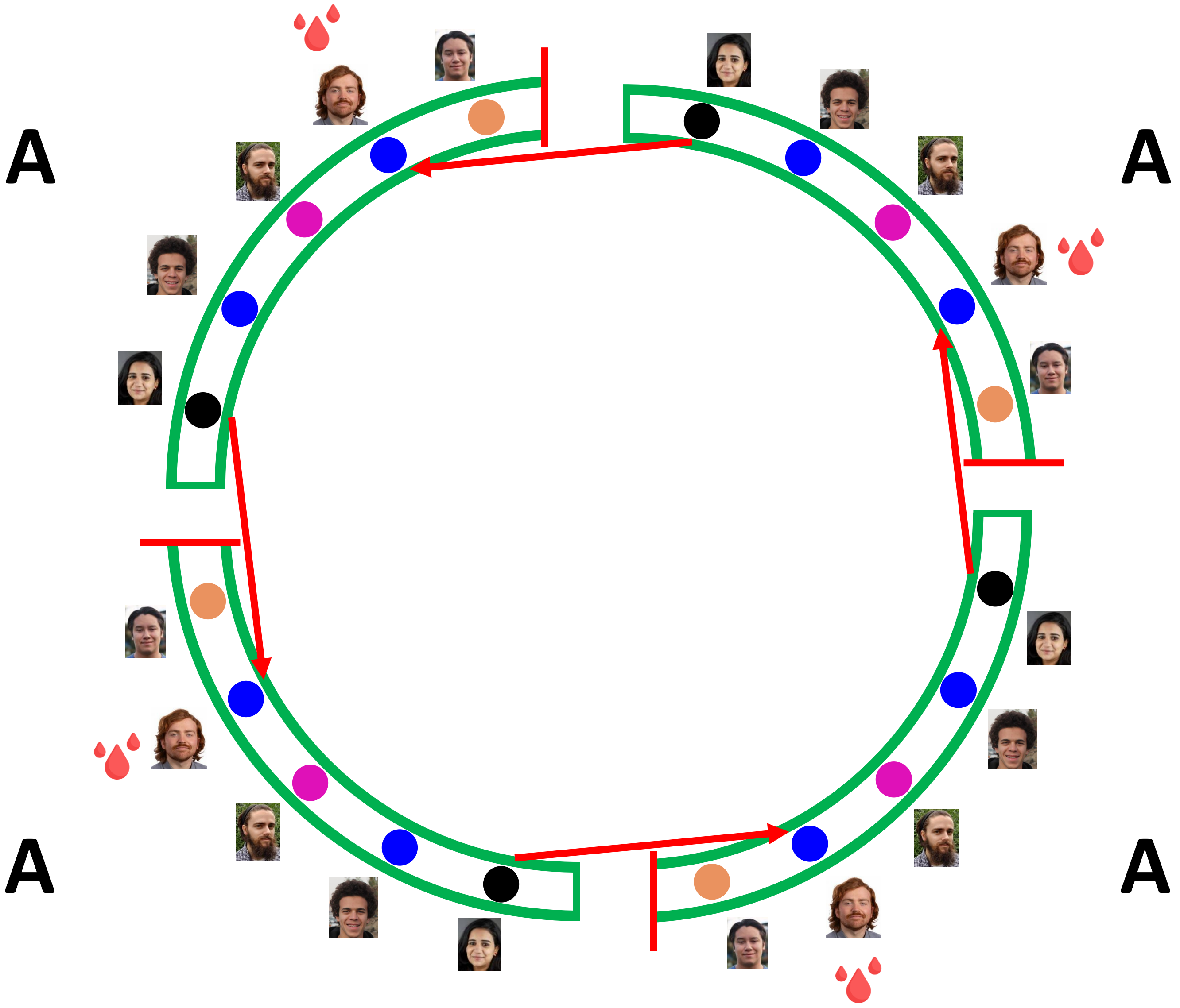


$\boldsymbol{\frac{N}{2}}$

$\Pi = 2^{N/2}$

$$S = k_B \frac{N}{2} N \log 2$$

c = 4 sofas



That is ugh ugh!

$$\underbrace{\Delta G(S)}_{\text{Free energy}} = \underbrace{\sum_{l \in S} \Delta G(l)}_{\text{Loop energy}} + \underbrace{(c-1) * \Delta G^{\text{assoc}}}_{\text{Entropic association cost}} + \underbrace{k_B T * \log R}_{\text{Symmetry penalty}}$$

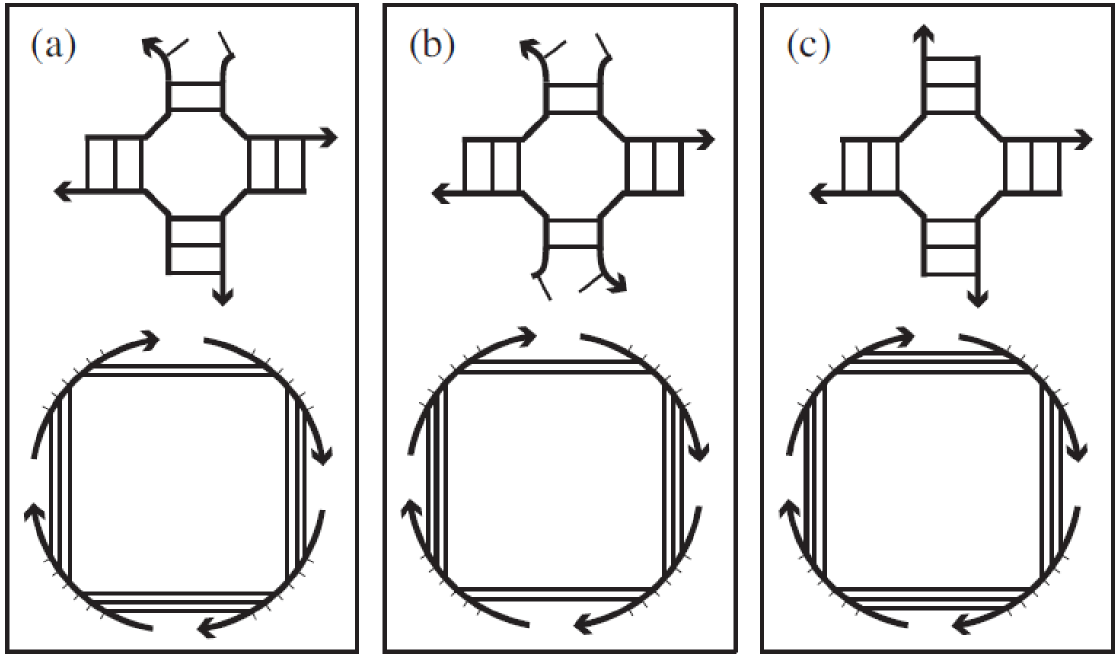


Fig. 2.2 *Sample secondary structures and polymer graphs for a complex of four indistinguishable strands.* (a) *1-fold (i.e., no) rotational symmetry.* (b) *2-fold rotational symmetry.* (c) *4-fold rotational symmetry.*

Why is this difficult?

Computational complexity of Minimum Free Energy algorithms

Level	Input Type	MFE
1	Single Strand (Maximum matching)	$O(N^3)$
2	Single Strand (Loop model)	$O(N^3)$
3	Multiple **unique** Strands, Bounded ($\leq$ c)	$O(N^3(c-1)!)$
4	Multiple Strands, Bounded ($\leq$ c)	?

N bases, c strands

All of these are dynamic programming algorithms

Subproblems → **Big problem**

Level	Input Type	MFE
1	Single Strand (Maximum matching)	$O(N^3)$
2	Single Strand (Loop model)	$O(N^3)$
3	Multiple unique Strands, Bounded ($\leq$ c)	$O\left(N^3(c-1)!\right)$
4	Multiple Strands, Bounded ($\leq$ c)	?

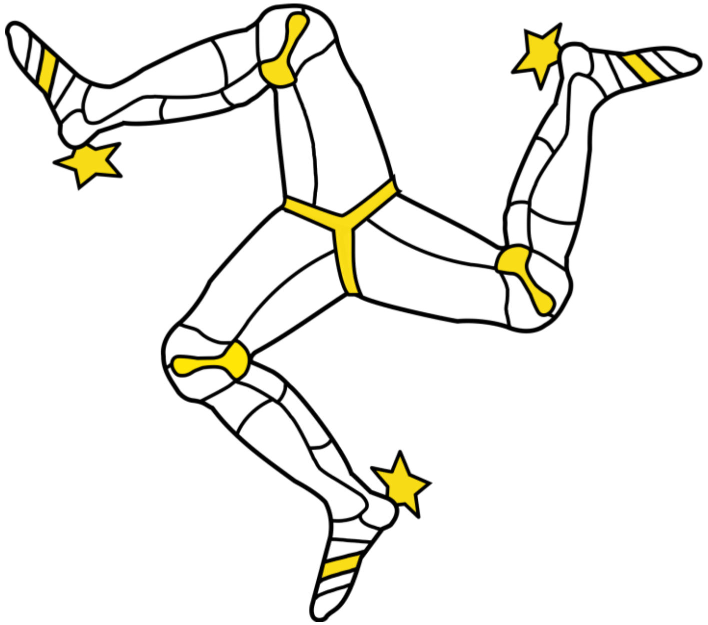
N bases, c strands

All of these are dynamic programming algorithms

Subprblems → Big problem

↓

Local point of view → ? → Global property



Computational complexity of Minimum Free Energy algorithms

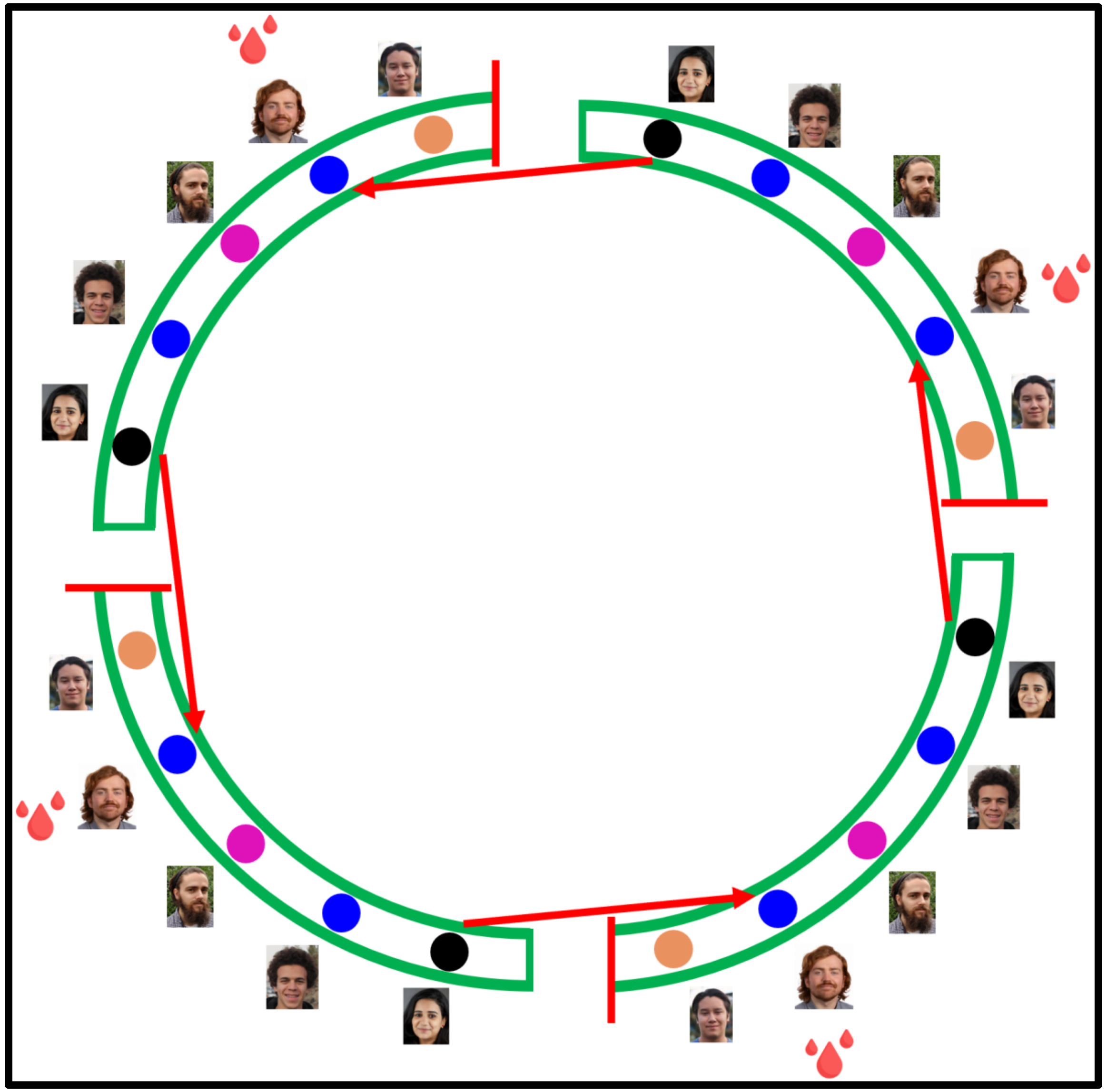
Level	Input Type	MFE
1	Single Strand (Maximum matching)	$O(N^3)$
2	Single Strand (Loop model)	$O(N^3)$
3	Multiple **unique** Strands, Bounded ($\leq$ c)	$O(N^3(c-1)!)$
4	Multiple Strands, Bounded ($\leq$ c)	?

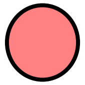
N bases, c strands

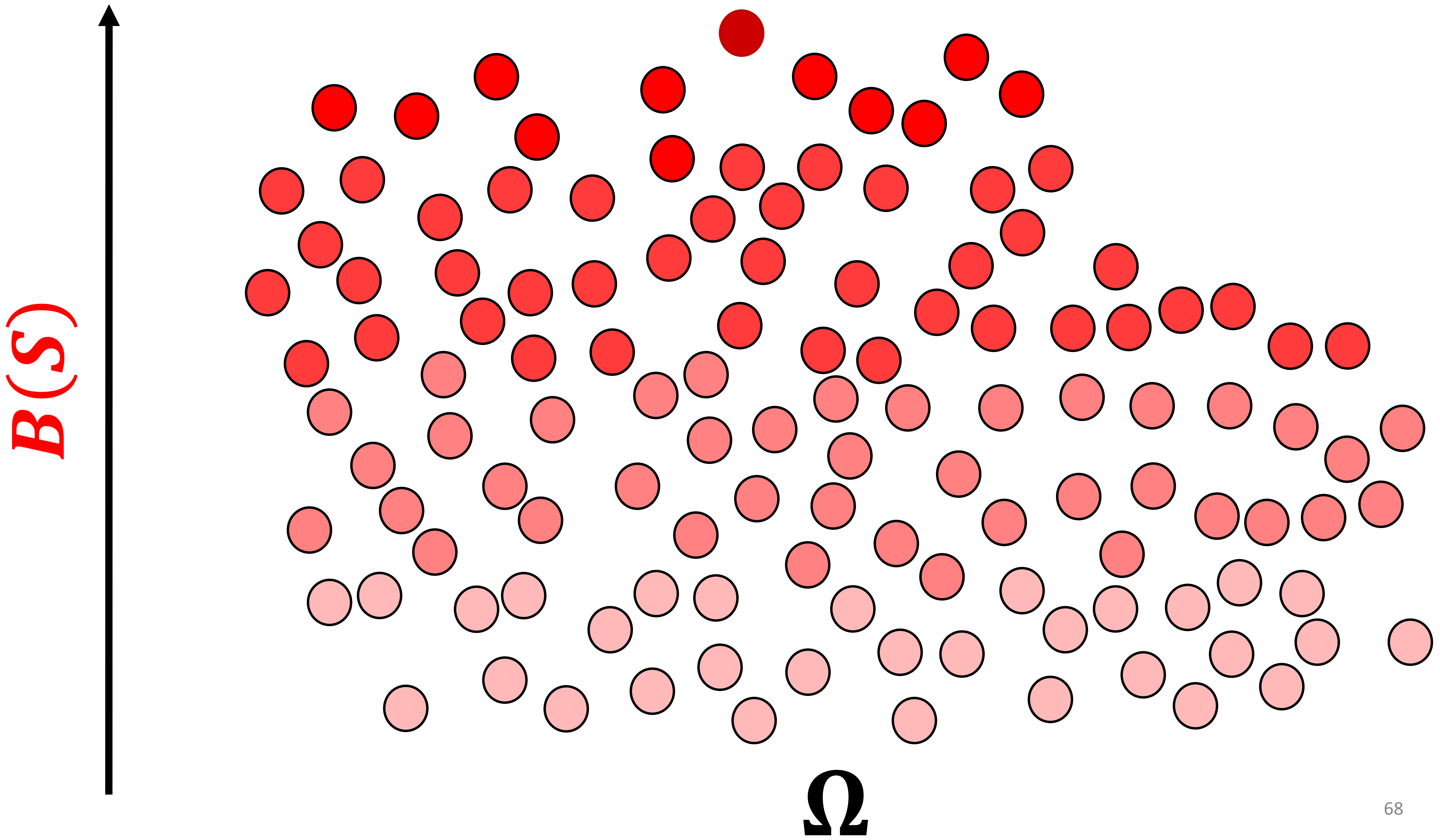
$$B(S) = \sum_{l} B(l) - (c-1)\, B^{\text{assoc}}$$

$$B(S) = \sum_{l} B(l) - (c-1) B^{\text{assoc}} - \boldsymbol{k_B T * \log R}$$

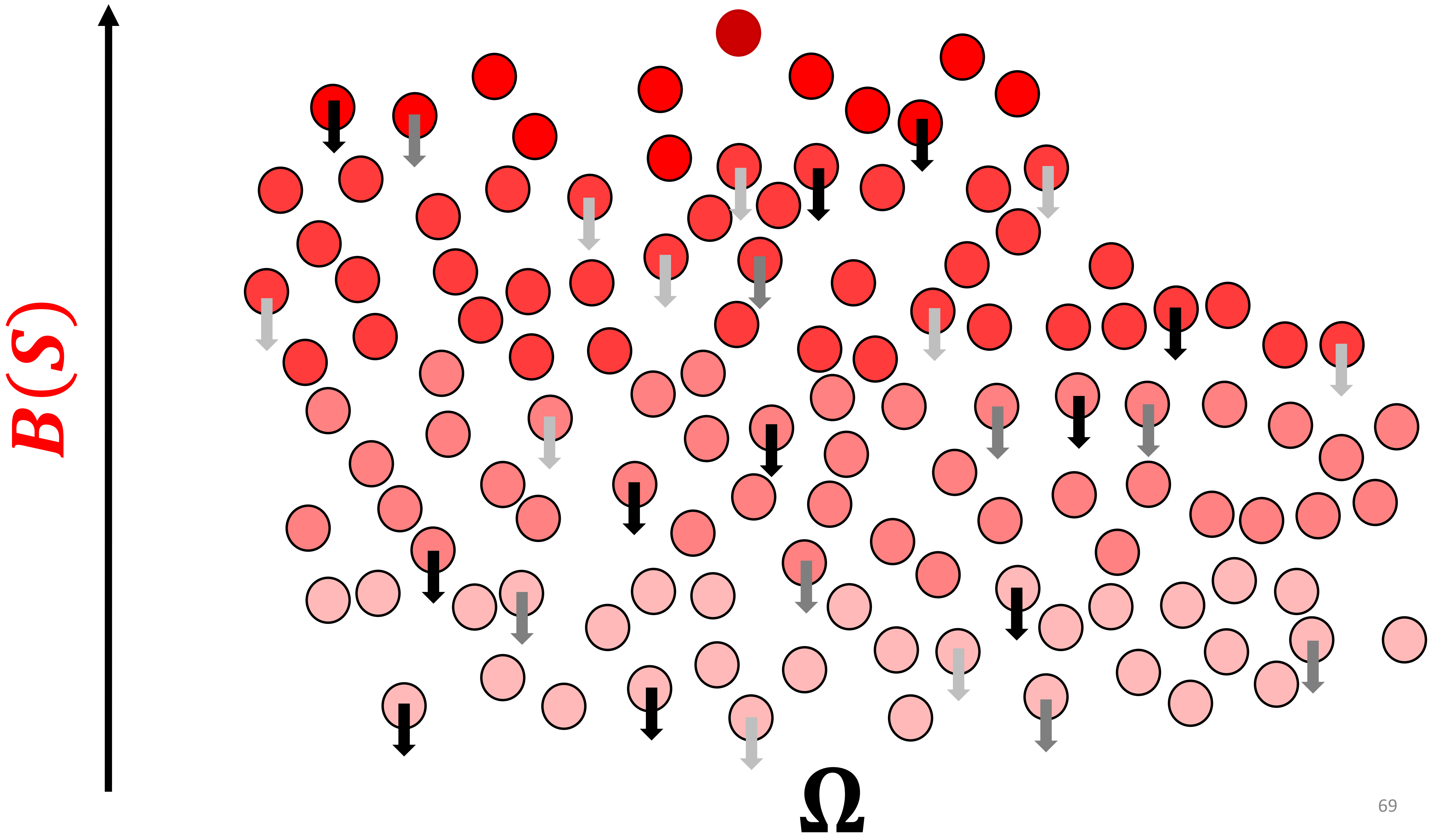
Let’s ignore the symmetry for a while

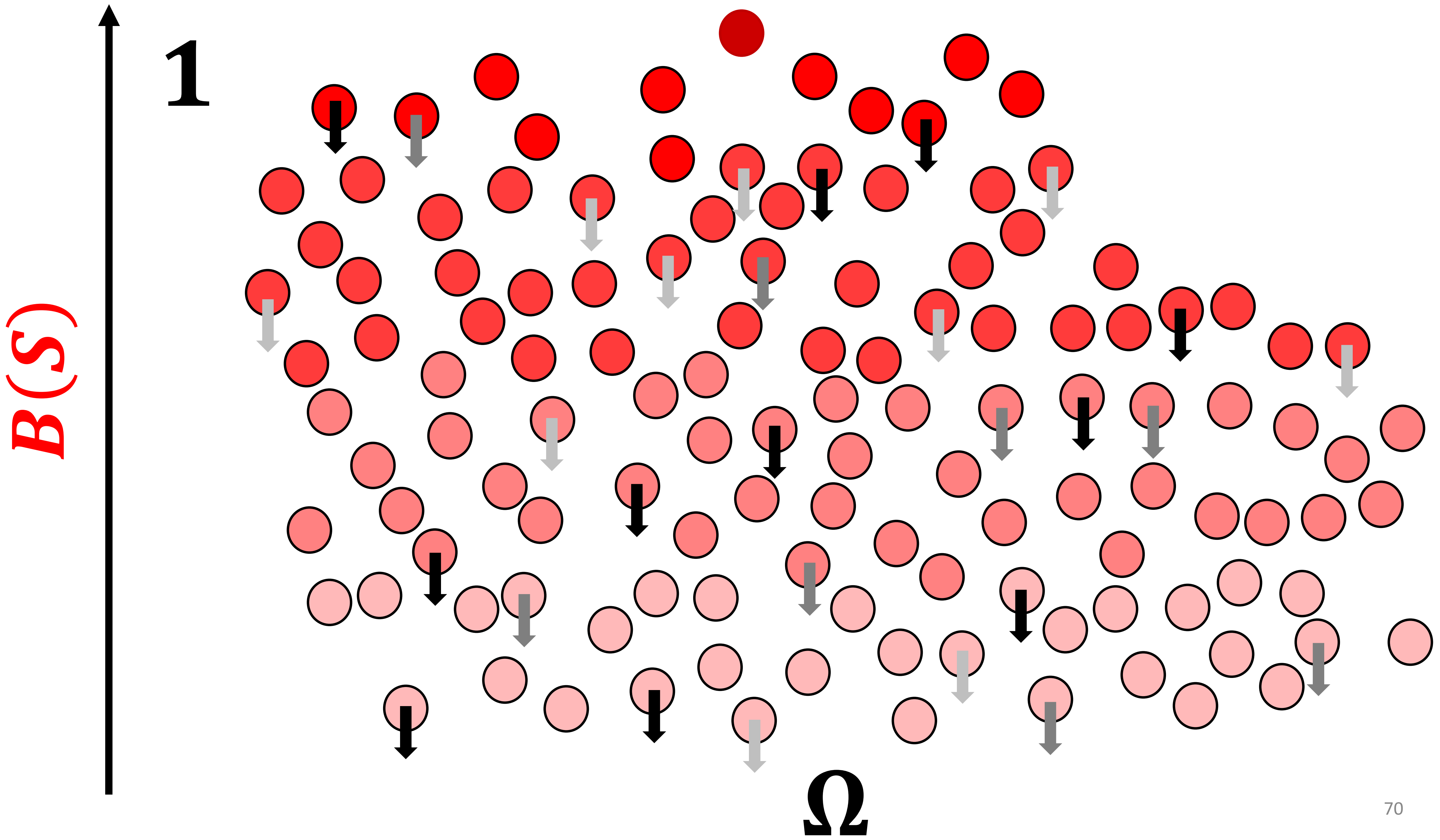


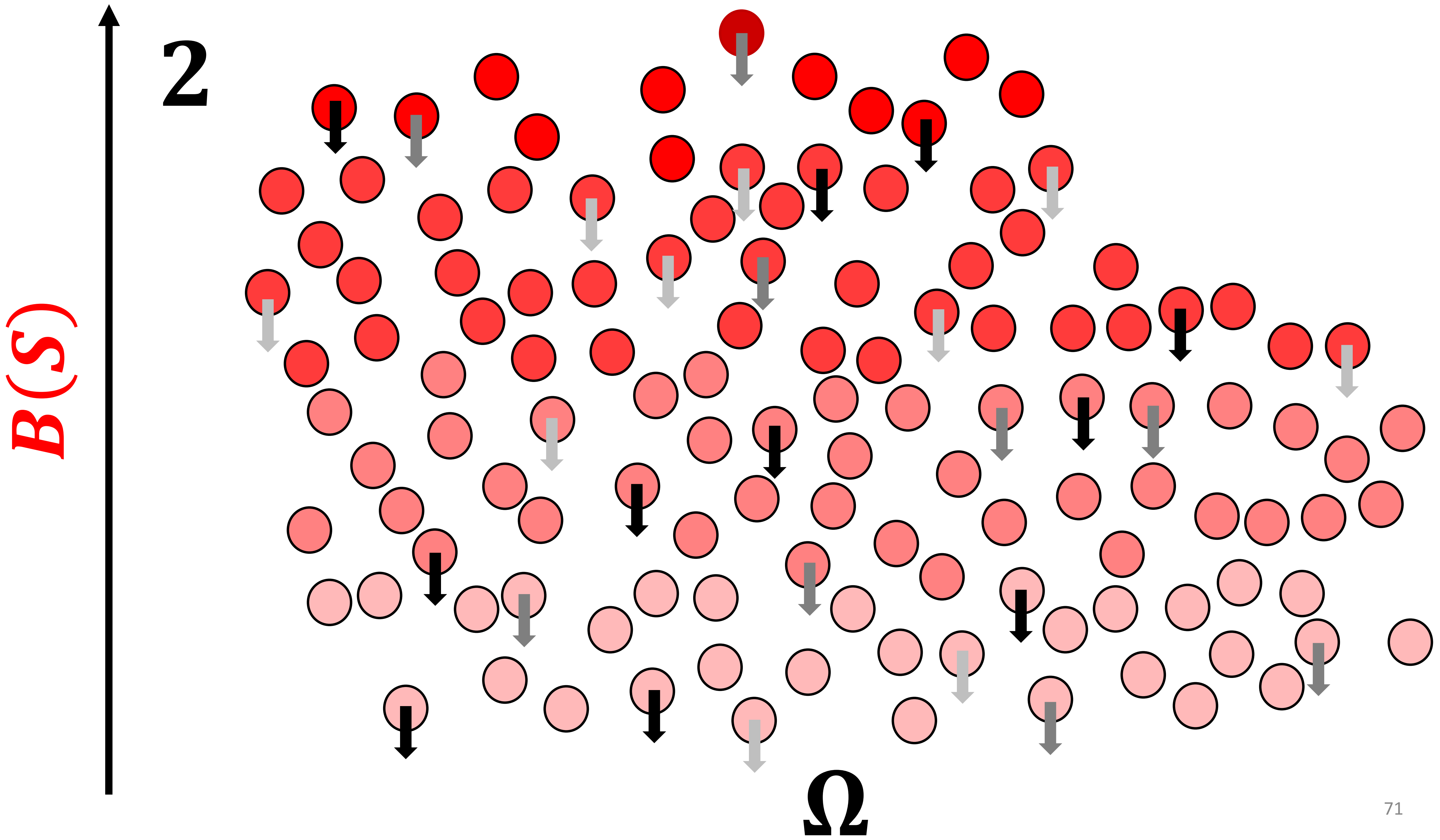




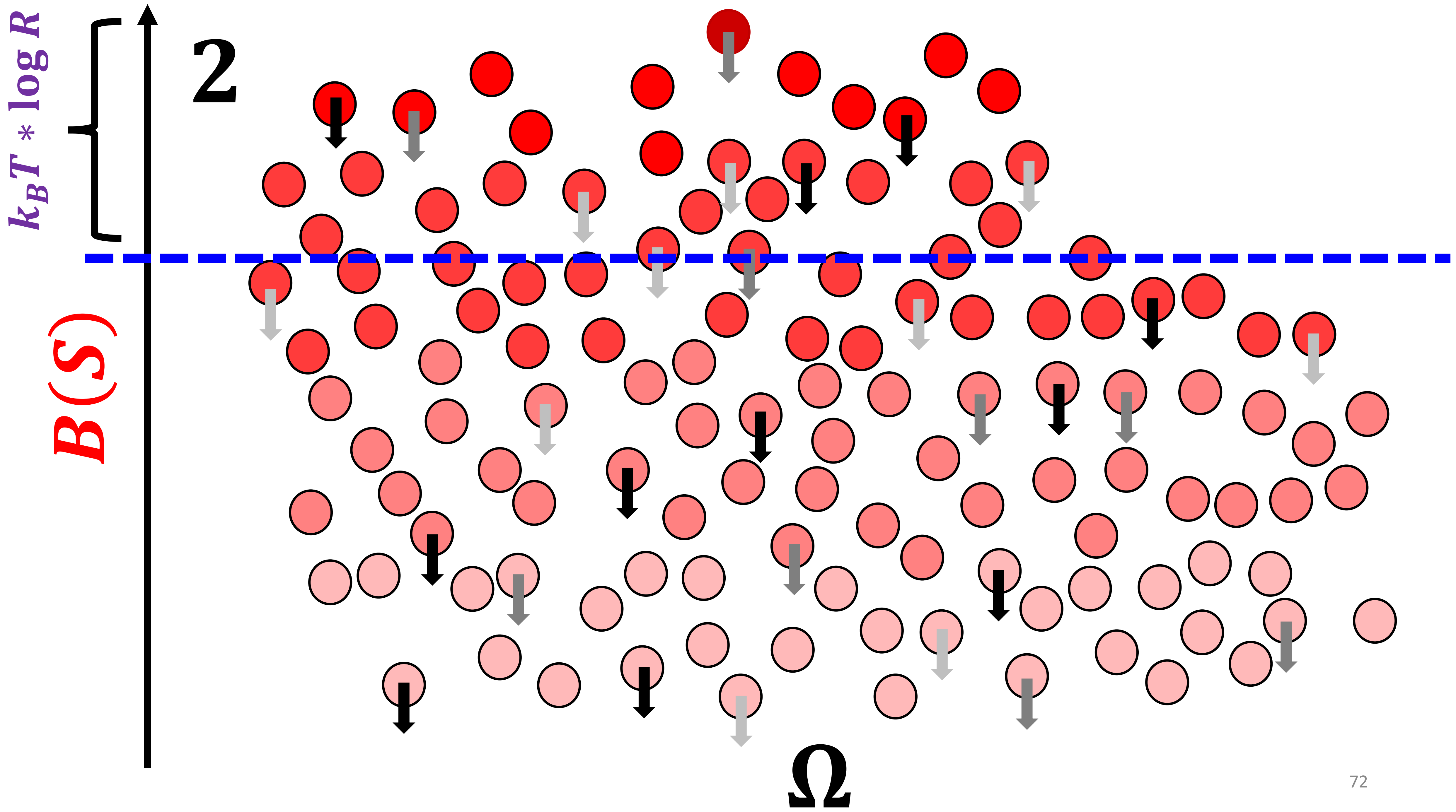
B(S)
Ω

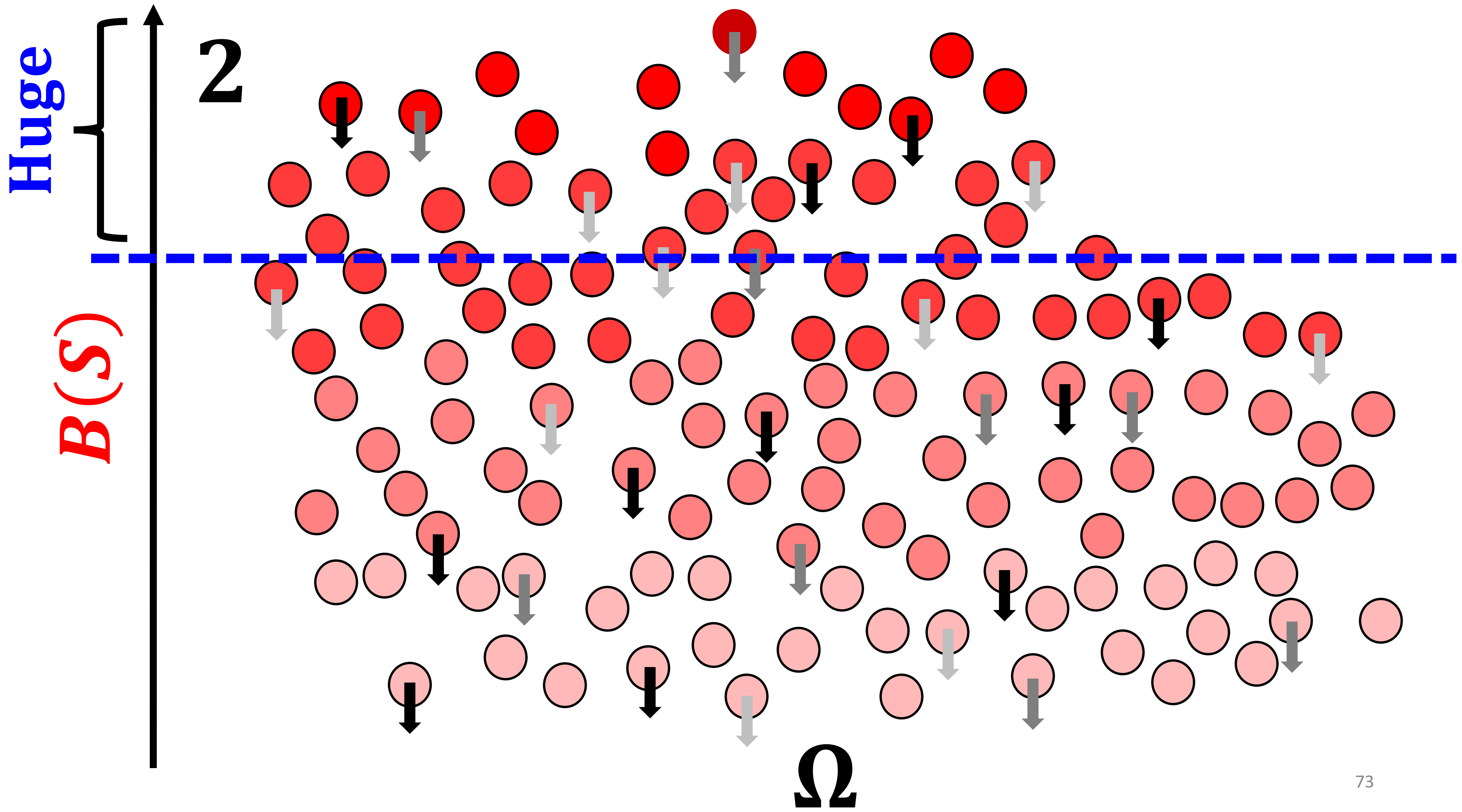

B(S)
Ω


B(S)
1
Ω



2
B(S)
Ω


$k_B T * \log R$
2
$B(S)$
Ω

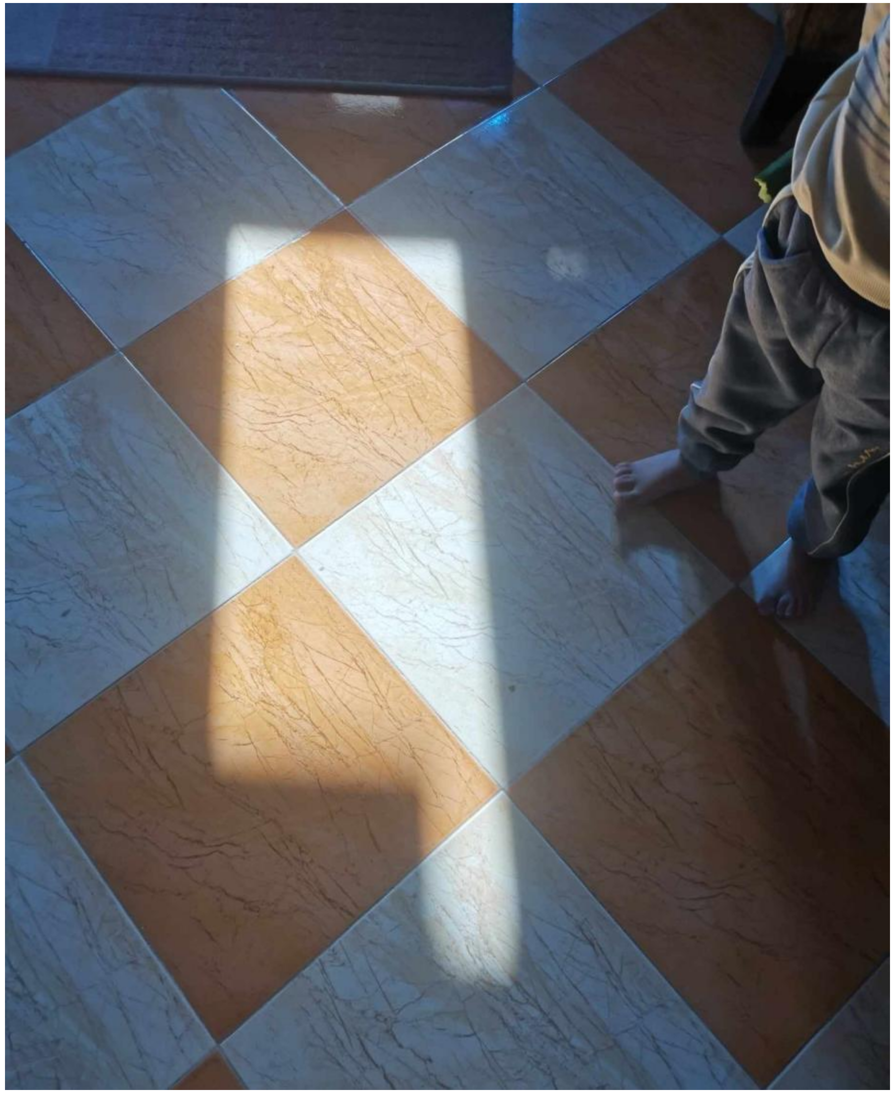

Huge
2
B(S)
Ω

Is there any hope?

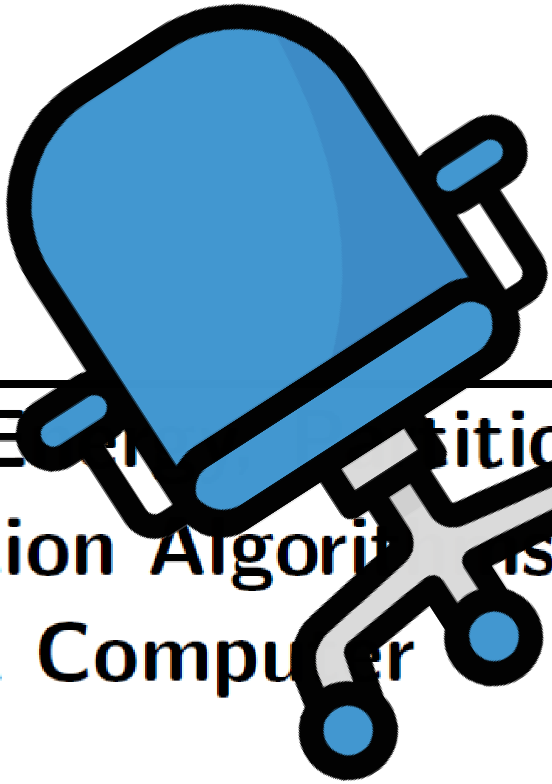
TAKE A BREAK

Yasso


BelFone
Digital CH 1
Zone 1
TX: 410.97500
RX: 410.97500
Menu
SMS



Last summer, we went to Japan



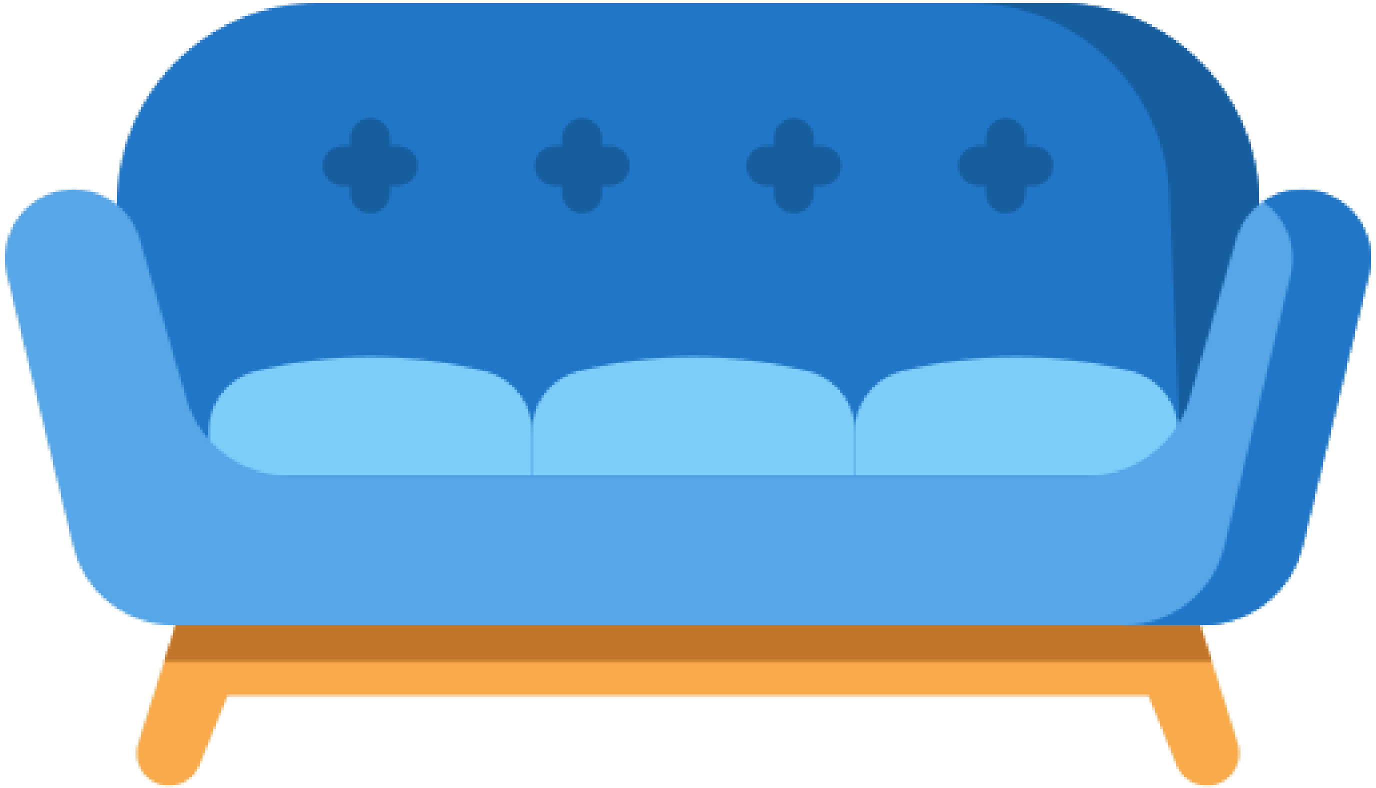
Minimum Free E… …ition Function and Kinetics Simulation Algori… …a Multistranded Scaffolded DNA Compu… …er

Ahmed Shalaby
Hamilton Institute, Department of Computer Science, Maynooth University, Ireland

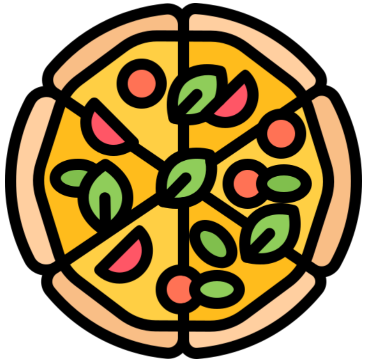
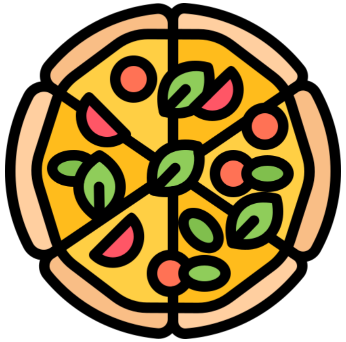
Chris Thachuk
Paul G. Allen School of Computer Science & Engineering, University of Washington, Seattle, WA, USA

Damien Woods
Hamilton Institute, Department of Computer Science, Maynooth University, Ireland

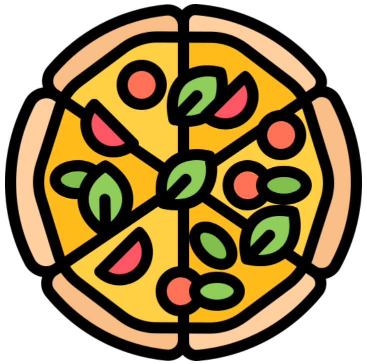
Yasso





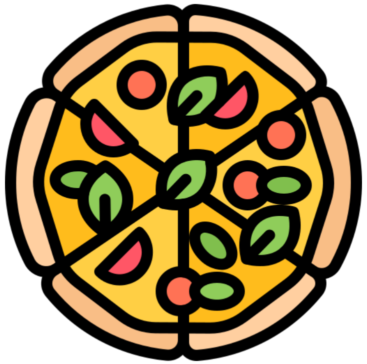





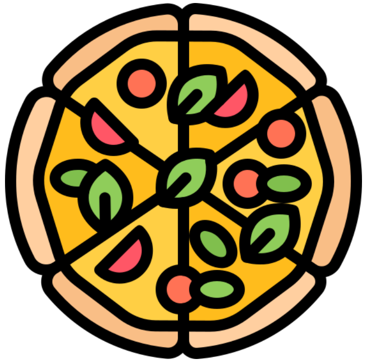


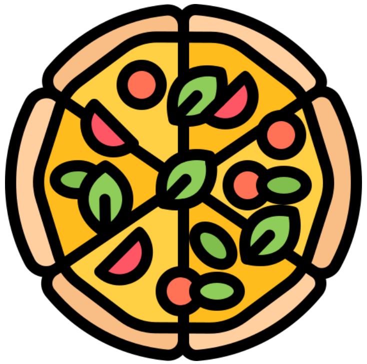




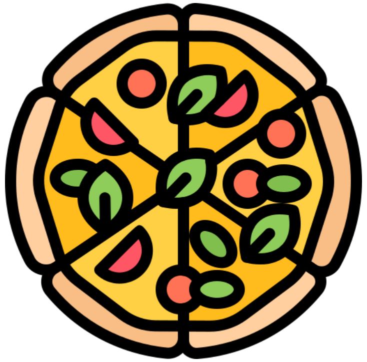




كله تمام يا خالو أحمد!



$B(S)$

$B(S_y) \leq B(S_x)$

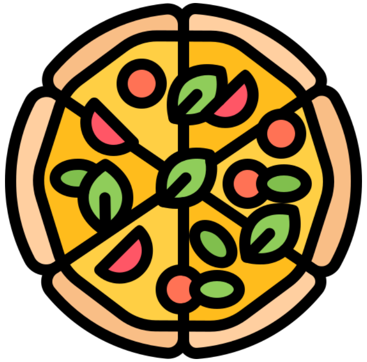
$B(S_y) \leq B(S_z) \leq B(S_x)$

S_x **Symmetric**

X

S_z **Asymmetric**

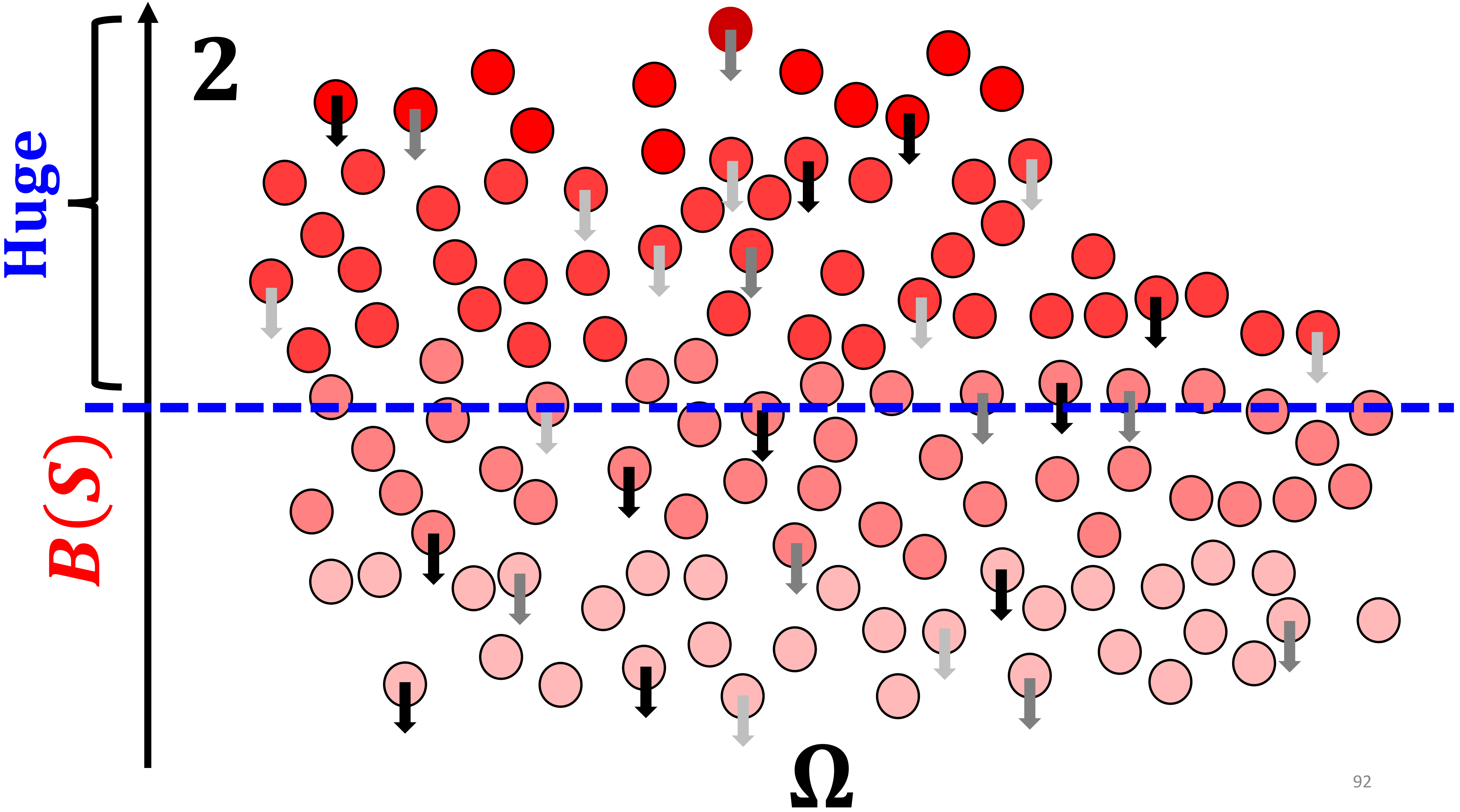
S_x and S_y
Admissible cut



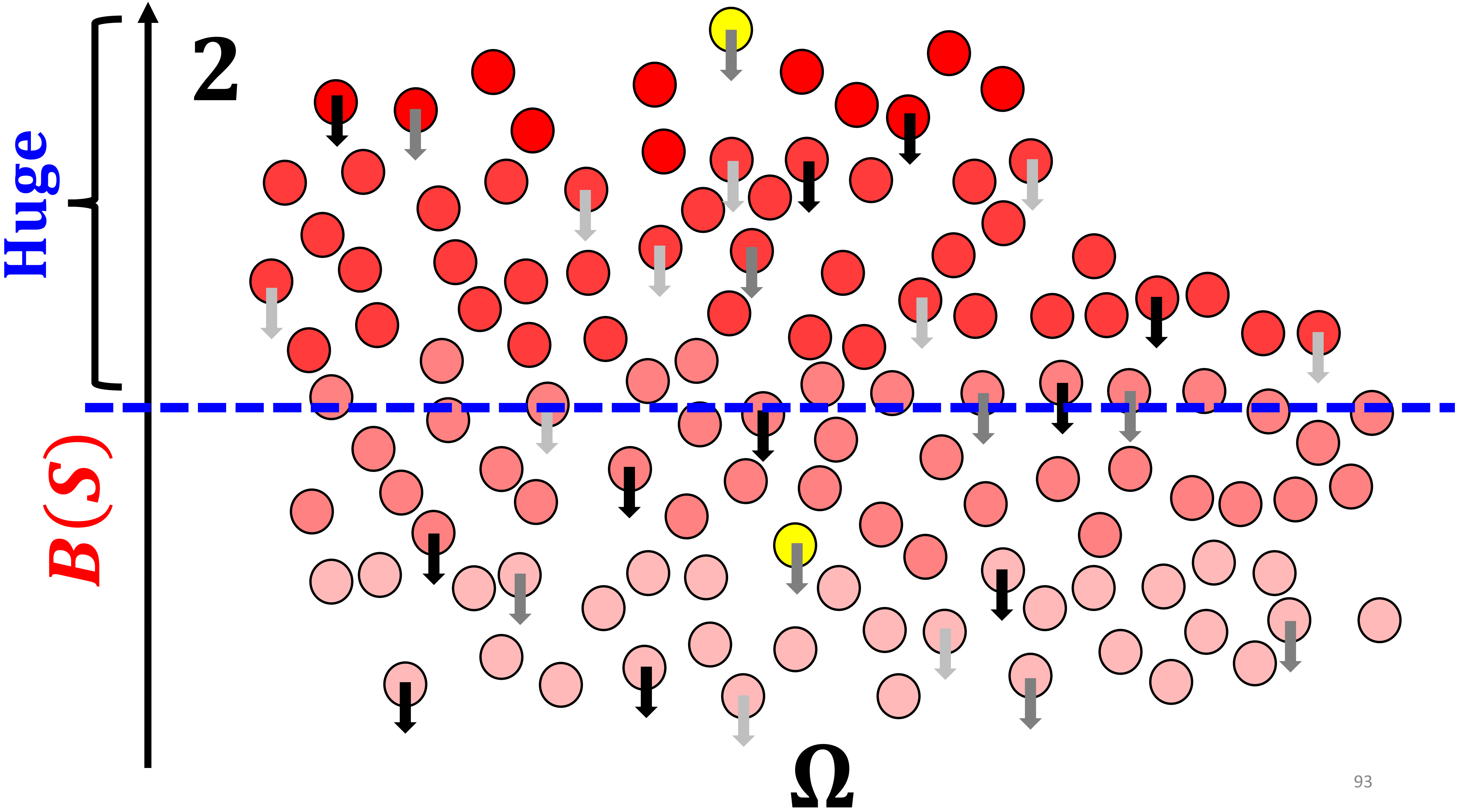
S_y **Symmetric**

The sandwich theorem of secondary structures

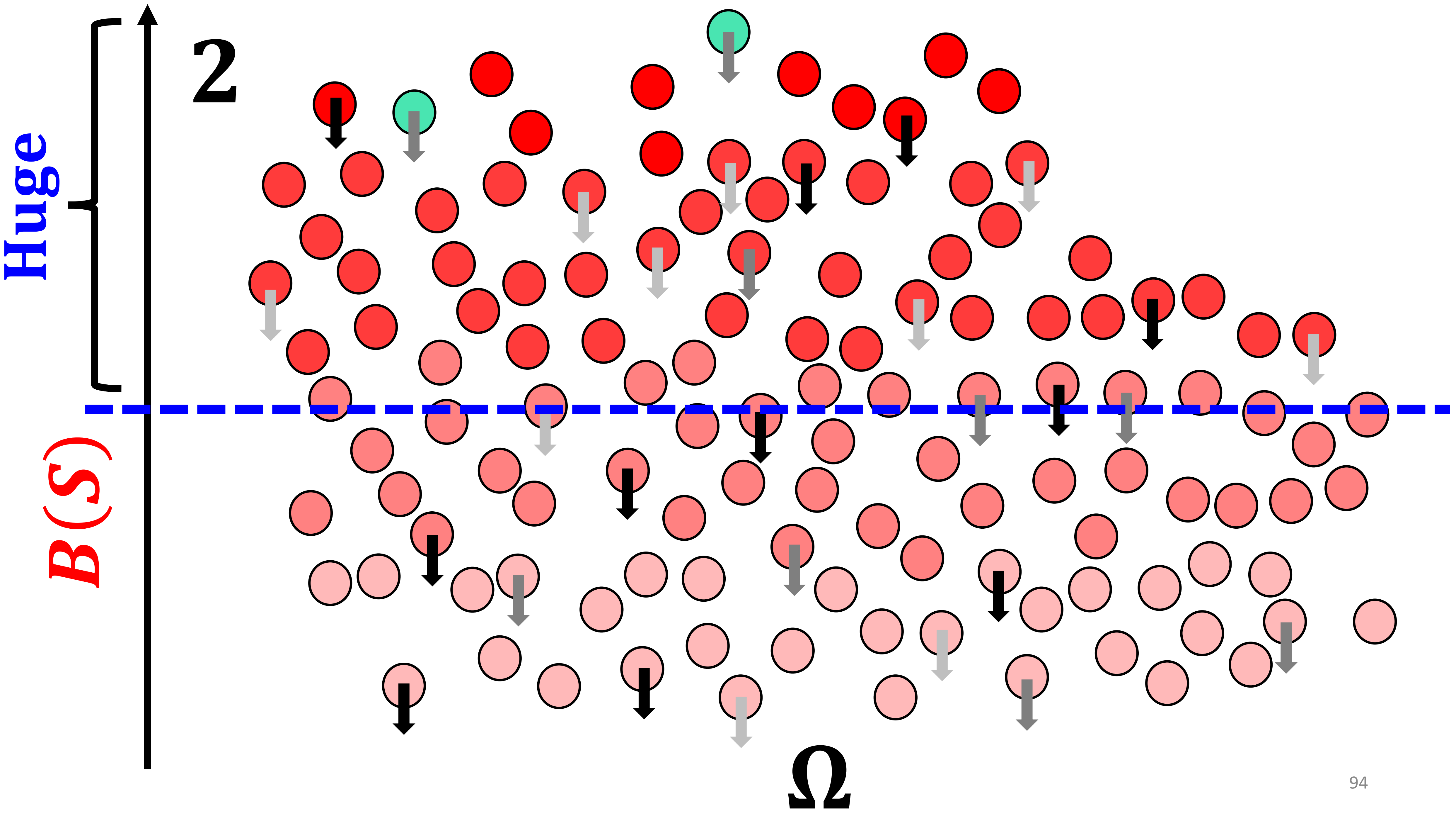
Does this solve the problem?



2
Huge
B(S)
Ω



2
Huge
$B(S)$
Ω


Huge
2
B(S)
Ω

$B(S)$

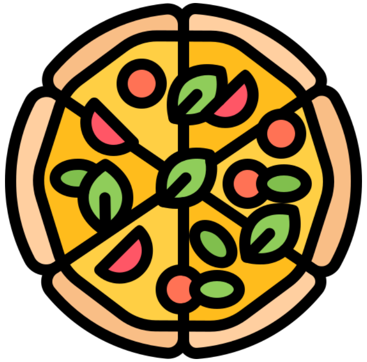
$B(S_y) \leq B(S_x)$

$B(S_y) \leq B(S_z) \leq B(S_x)$

S_x

X

S_z



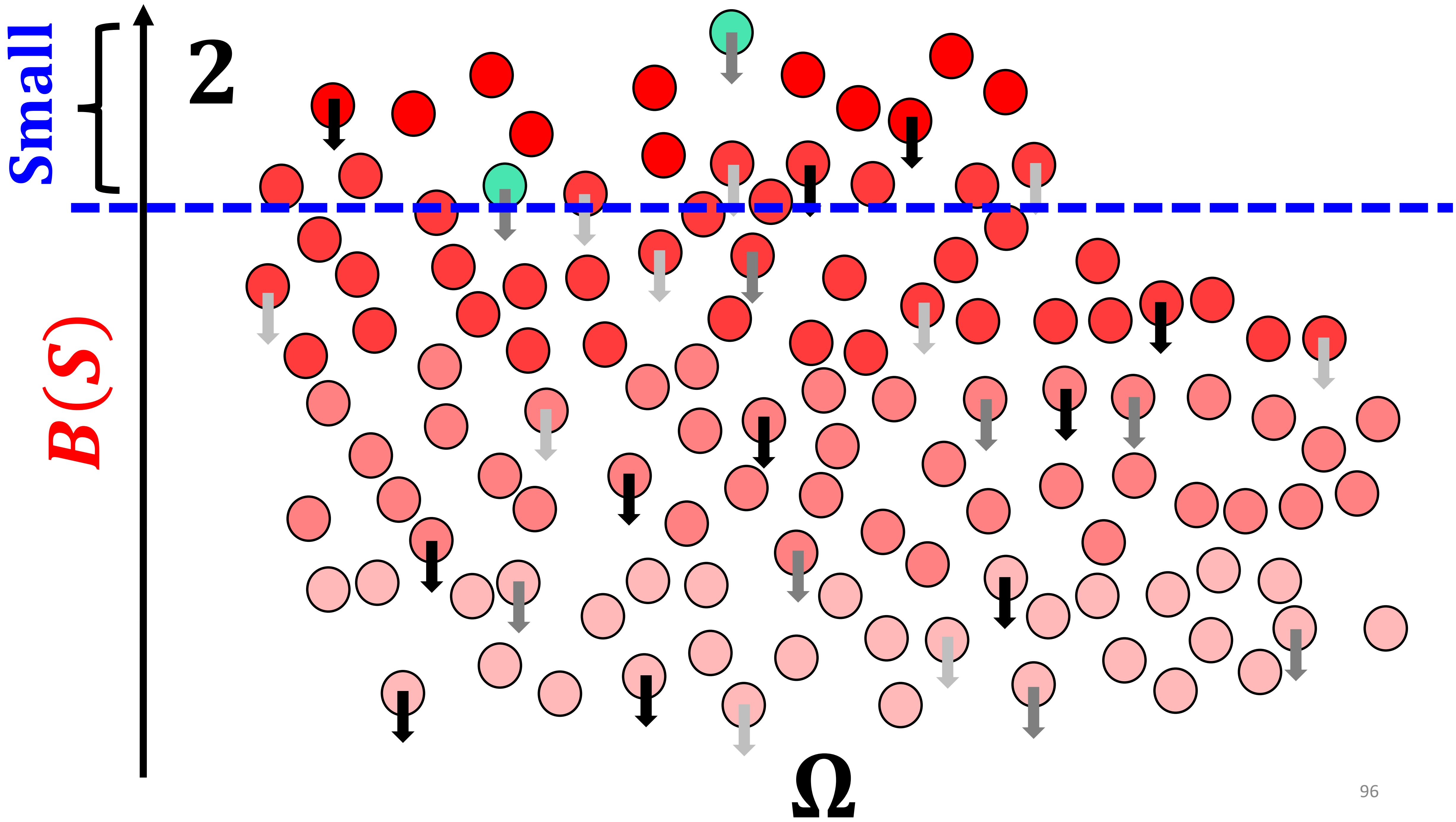
S_y

Upper bound

$$\frac{N-c}{v(\pi)}\left(\sigma(v(\pi)) - v(\pi)\right)$$

$$+$$

$$N^2/16$$


Small
2
B(S)
Ω

▶ **Lemma 28.** *For any two 2-fold rotational symmetric secondary structures, the maximum number of all distinct central internal loops is* $\sum_{s \in y} (\|A\|_s \|T\|_s + \|G\|_s \|C\|_s - \mathcal{I}_s) \leq N^2/16$,

where $\pi = y^2$, *and* $\mathcal{I}_s$ *is an indicator function such that* $\mathcal{I}_s = \begin{cases} 1 & c > 2 \text{ and } s(1) = \overline{s(|s|)}. \\ 0 & \text{otherwise} \end{cases}$

Computational complexity of Minimum Free Energy algorithms

Level	Input Type	MFE
1	Single Strand (Maximum matching)	$O(N^3)$
2	Single Strand (Loop model)	$O(N^3)$
3	Multiple **unique** Strands, Bounded ($\leq$ c)	$O(N^3(\mathrm{c}-1)!)$
4	Multiple Strands, Bounded ($\leq$ c)	?

N bases, c strands

Open problem for $\approx$ 20 years

Computational complexity of Minimum Free Energy algorithms

Level	Input Type	MFE
1	Single Strand (Maximum matching)	$O(N^3)$
2	Single Strand (Loop model)	$O(N^3)$
3	Multiple **unique** Strands, Bounded ($\leq$ c)	$O(N^3(c-1)!)$
4	Multiple Strands, Bounded ($\leq$ c)	$\mathbf{O(N^4(c-1)!)}$

N bases, c strands

Thanks

Maynooth University
National University of Ireland Maynooth

Hamilton Institute

dna.hamilton.ie

dna.hamilton.ie/shalaby

European Innovation Council

Funded by the European Union

erc

sfi Science Foundation Ireland