It’s not about DNA, It’s all about Pizza

Ahmed Shalaby

2nd year PhD

Supervisor: Damien Woods

CS workshop [19/06/2024]

Hamilton Institute

Maynooth University
National University of Ireland Maynooth

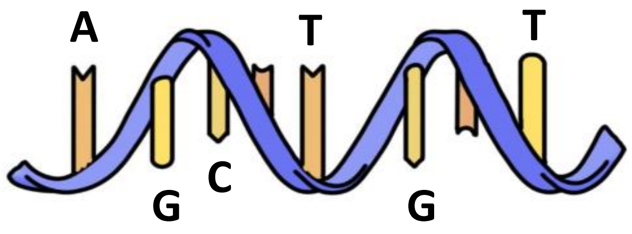
Fondúireacht Eolaíochta Éireann sfi Dá bhfuil romhainn

European Innovation Council

Funded by the European Union

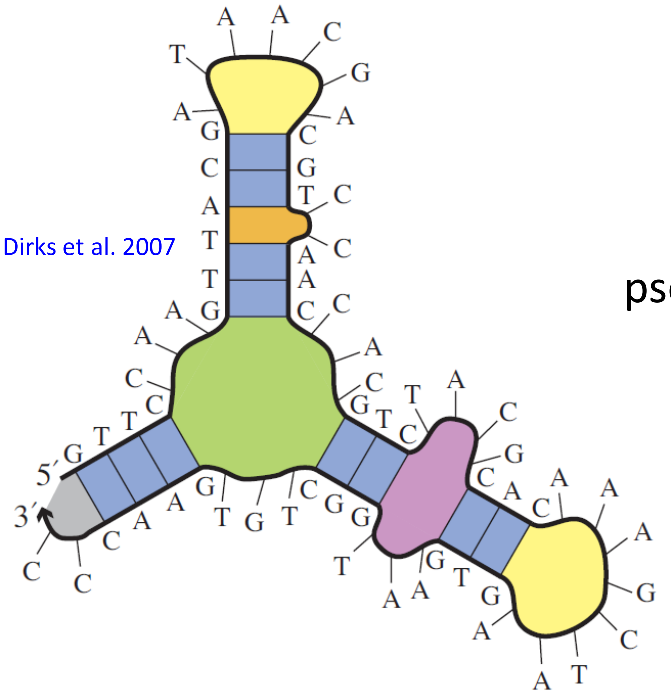
erc

Secondary structure



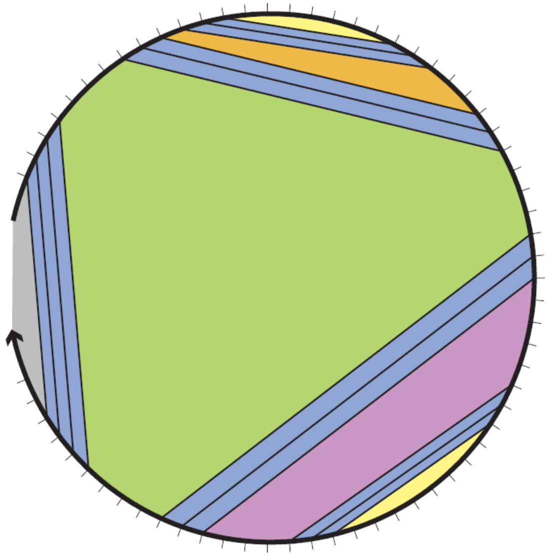
Single stranded DNA

Dirks et al. 2007



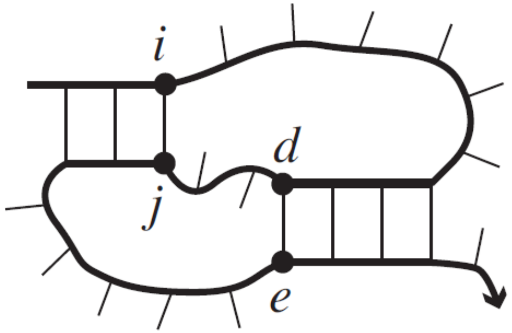
Secondary structure

pseudoknot-free



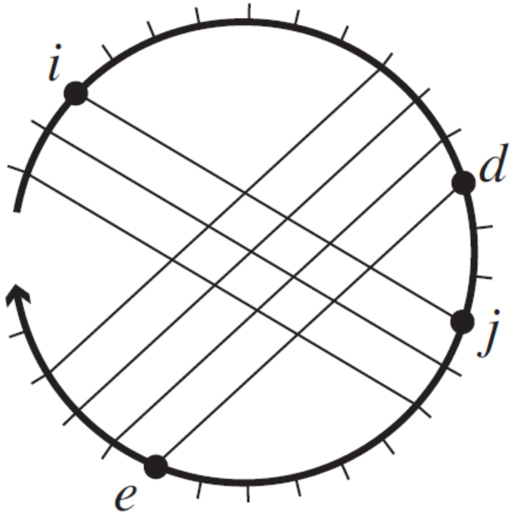
Dirks et al. 2007

Polymer graph representation



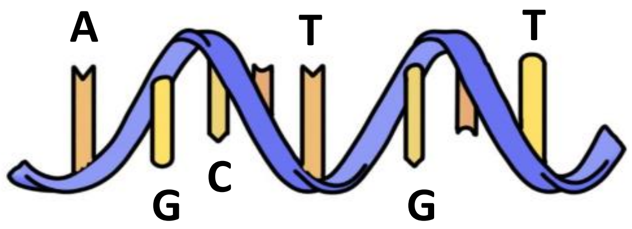
Dirks et al. 2007

pseudoknotted



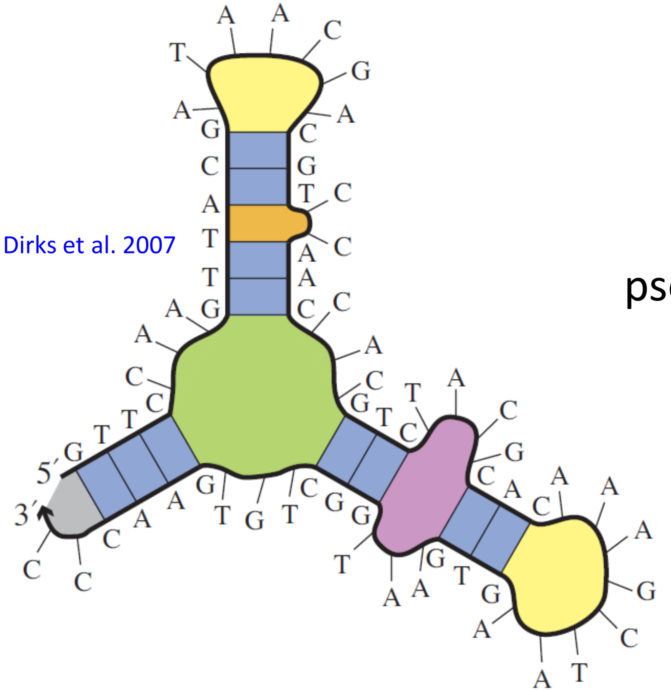
Dirks et al. 2007

Secondary structure



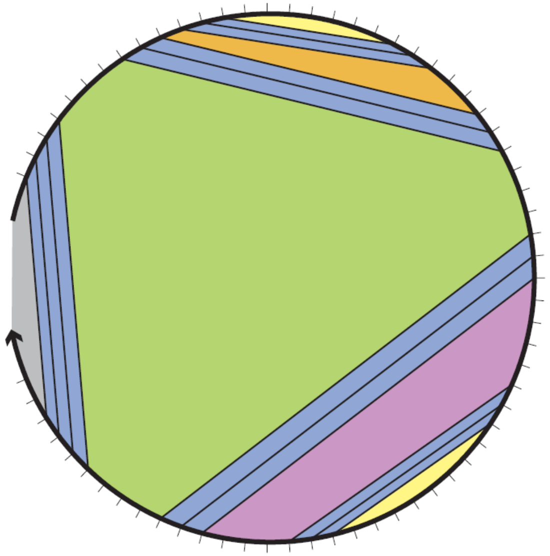
Single stranded DNA

Dirks et al. 2007



Secondary structure

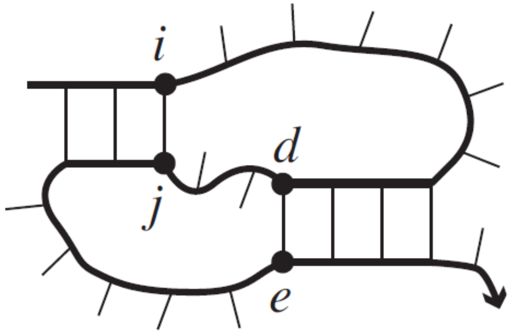
pseudoknot-free



Dirks et al. 2007

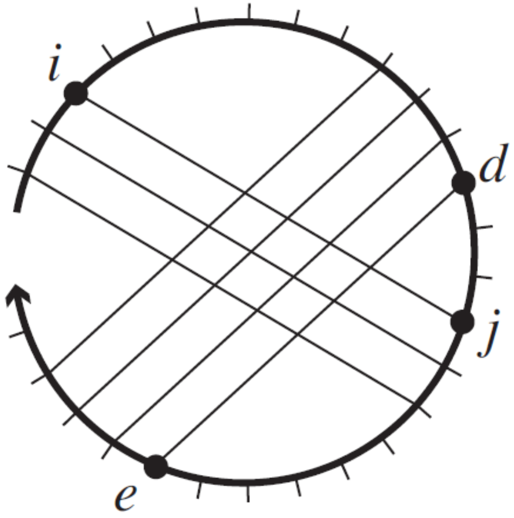
Polymer graph representation

NP – Complete



Dirks et al. 2007

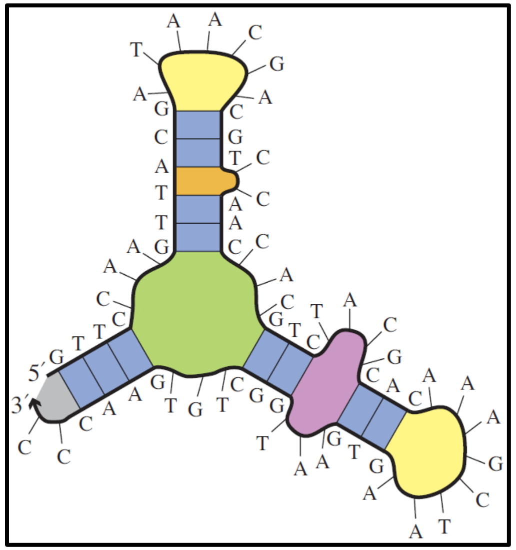
pseudoknotted

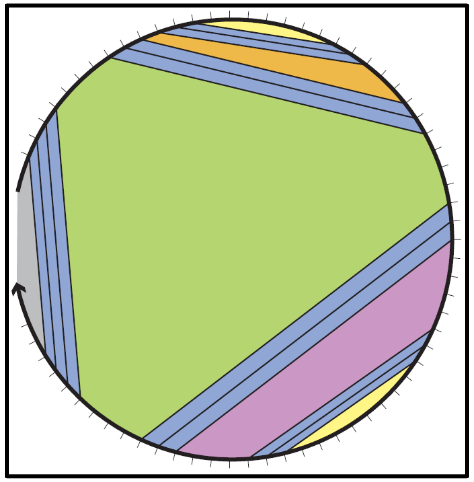


Dirks et al. 2007

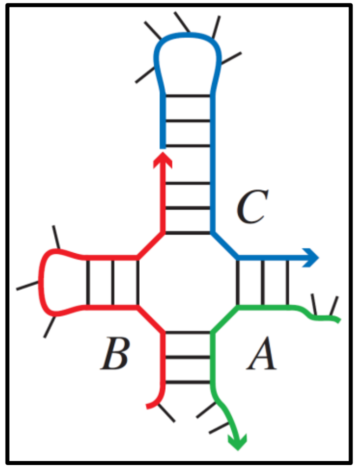
Energy models, Minimum Free Energy and Partition Function

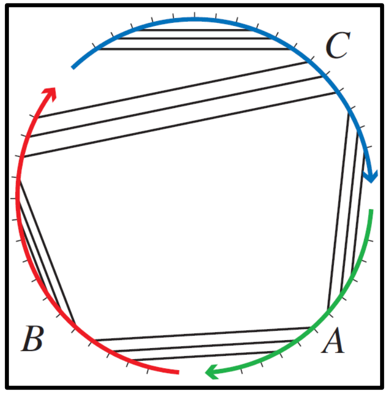
Single stranded system





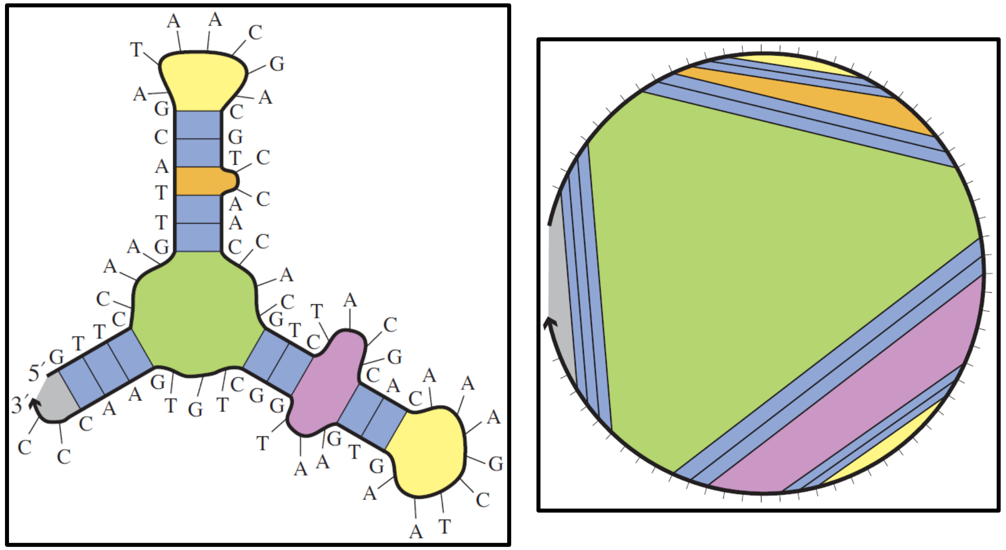
Multi stranded system of s strands





Energy models, Minimum Free Energy and Partition Function

Single stranded system

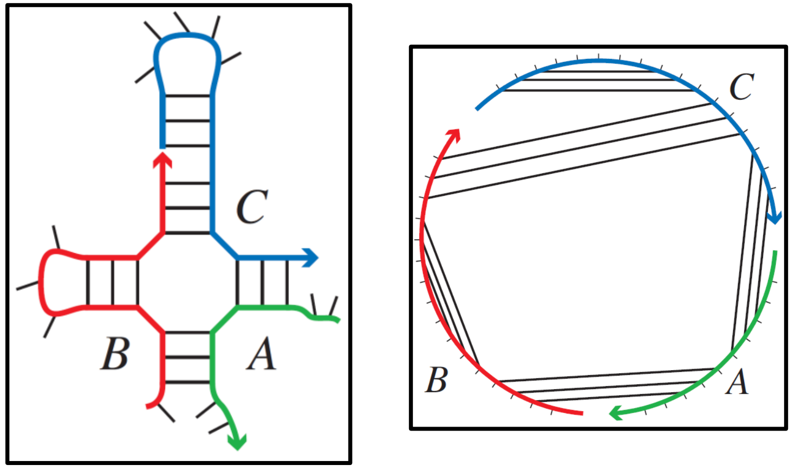


$\Delta G(S)$

Energy model

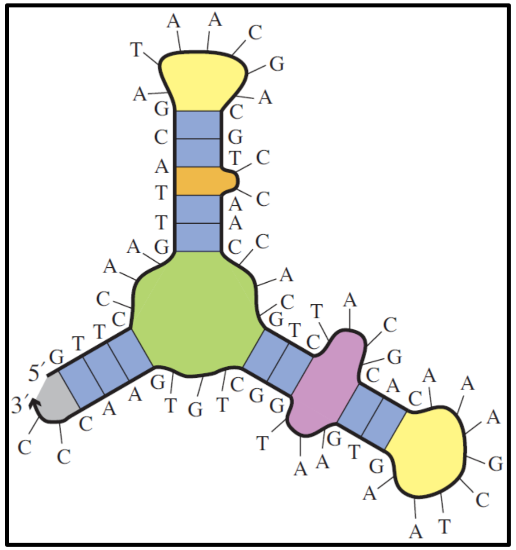
Capture the free energy of secondary structure

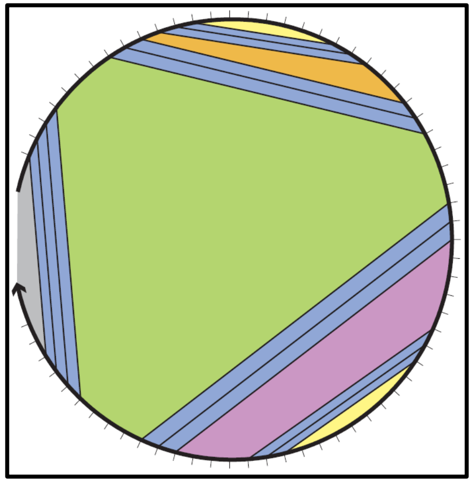
Multi stranded system of s strands



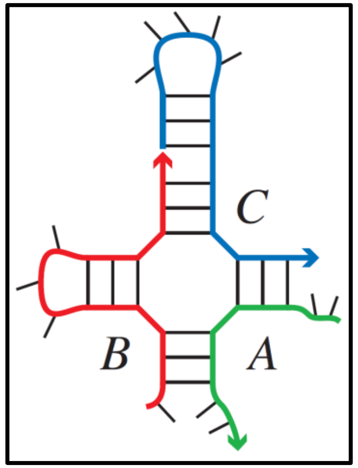
Energy models, Minimum Free Energy and Partition Function

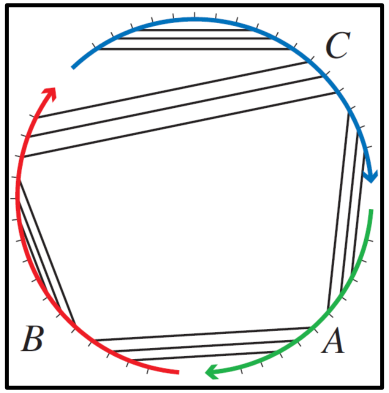
Single stranded system





Multi stranded system of s strands

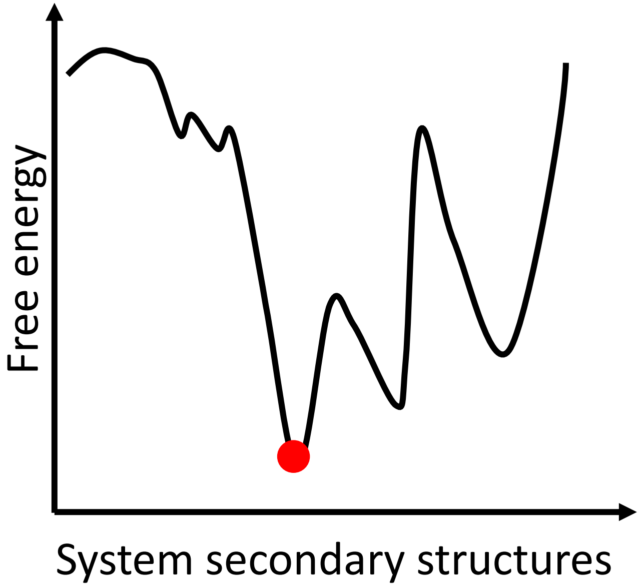




$\Delta G(S)$

Energy model

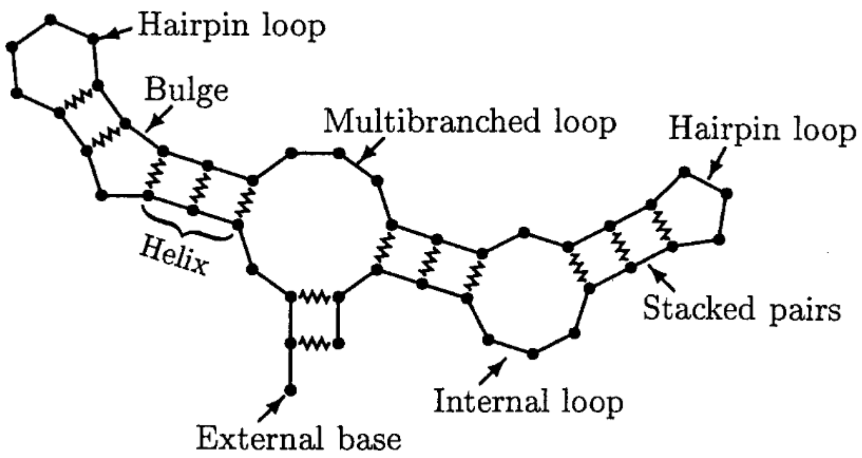
Capture the free energy of secondary structure

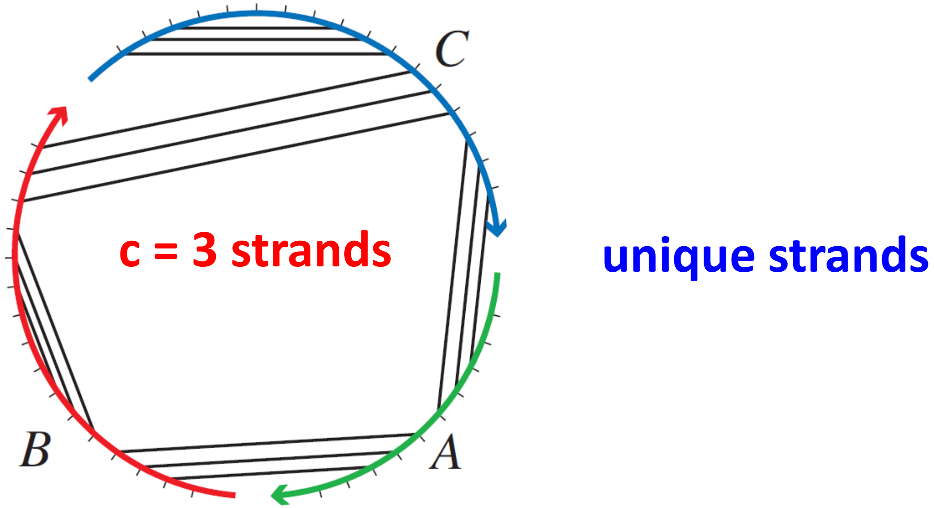


$$\text{MFE} = \min_{S\in\Omega} \Delta G(S)$$

Minimum Free Energy

Energy model: Loop model





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

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

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

Ω: the set of all secondary structures

Energy model: Loop model (allowing repeats)

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

$-$ $+$ $+$

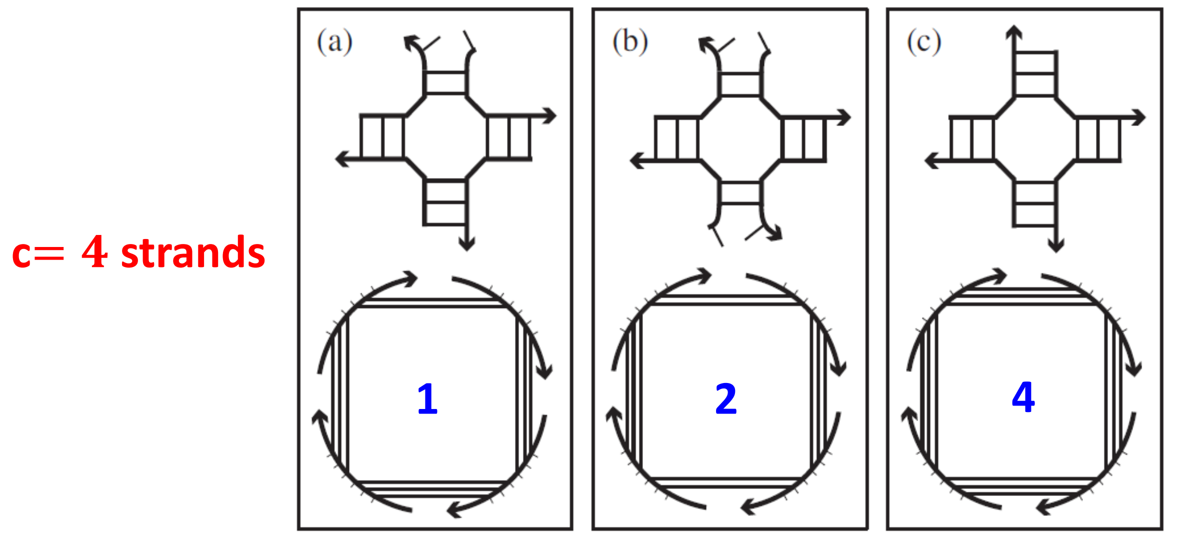


Fig. 2.2 *Sample secondary structures and polymer graphs for a complex of four indistinguishable strands. (a) 1-fold (i.e., no) rotational symmetry. (b) 2-fold rotational symmetry. (c) 4-fold rotational symmetry.*

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

Ω: the set of all connected structures

R: degree of rotational symmetry

Computational complexity of Minimum Free Energy algorithms

Input Type	MFE
Single Strand	$O(N^3)$
Multiple **unique** Strands, Bounded ($\leq$ c)	$O(N^3(\mathrm{c}-1)!)$
Multiple Strands **allowing repeats**, Bounded ($\leq$ c)	?
Multiple Strands, Unbounded	$NP-\mathrm{Complete}$

N bases, c strands

Computational complexity of Minimum Free Energy algorithms

Input Type	MFE
Single Strand	$O(N^3)$
Multiple **unique** Strands, Bounded $(\leq \text{c})$	$O(N^3(\text{c}-1)!)$
Multiple Strands **allowing repeats**, Bounded $(\leq \text{c})$	?
Multiple Strands, Unbounded	$NP-\text{Complete}$

N bases, c strands

Open problem for $\approx$ 20 years

Why symmetry makes it difficult?

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

N bases, c strands

Why symmetry makes it difficult?

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

} **Dynamic programming algorithms**

N bases, c strands

All of these are dynamic programming algorithms

Why symmetry makes it difficult?

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

} Dynamic programming algorithms

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

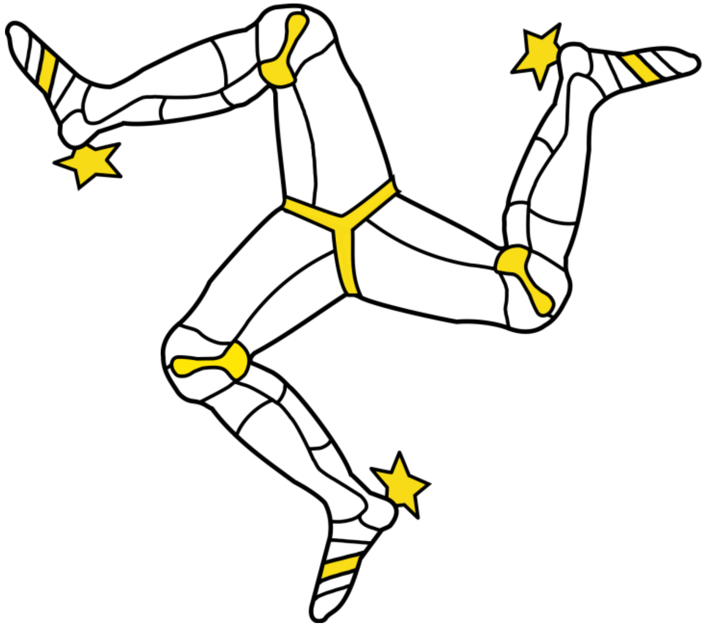
Subprograms → Big problem

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 ($\leq c$)	$O(N^3(c-1)!)$
4	Multiple Strands, Bounded ($\leq c$)	?

N bases, c strands

All of these are **dynamic programming** algorithms

Subproblems → **Big problem**

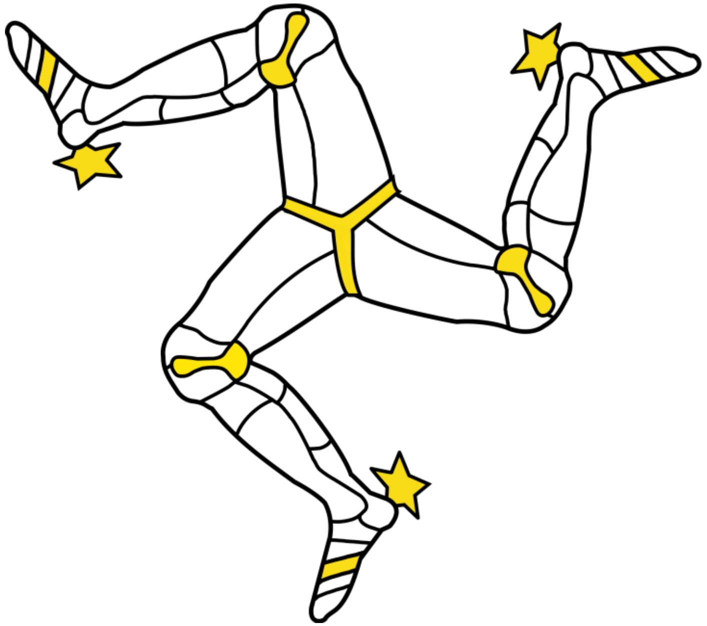


Level	Input Type	MFE
1	Single Strand (Maximum matching)	$O(N^3)$
2	Single Strand (Loop model)	$O(N^3)$
3	Multiple unique Strands, Bounded ($\leq$ c)	$O\left(N^3(c-1)!\right)$
4	Multiple Strands, Bounded ($\leq$ c)	?

N bases, c strands

All of these are dynamic programming algorithms

Subproblems → **Big problem**



Global property

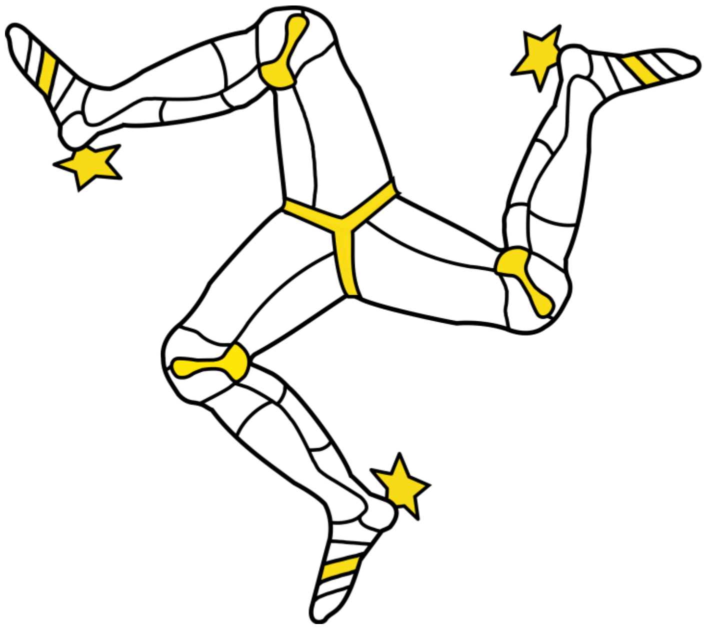
Level	Input Type	MFE
1	Single Strand (Maximum matching)	$O(N^3)$
2	Single Strand (Loop model)	$O(N^3)$
3	Multiple unique Strands, Bounded ($\leq$ c)	$O\left(N^3(c-1)!\right)$
4	Multiple Strands, Bounded ($\leq$ c)	?

N bases, c strands

All of these are dynamic programming algorithms

Subprojects → Big problem

Local point of view



Global property

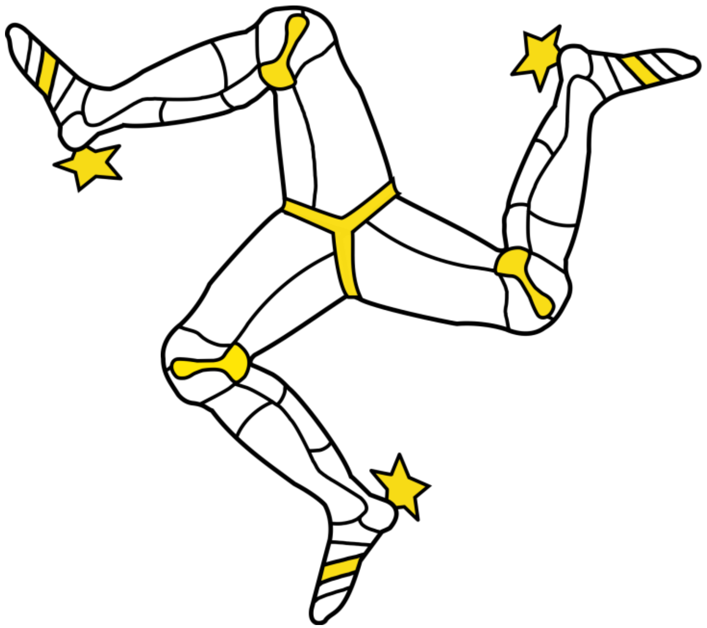
Level	Input Type	MFE
1	Single Strand (Maximum matching)	$O(N^3)$
2	Single Strand (Loop model)	$O(N^3)$
3	Multiple unique Strands, Bounded ($\leq$ c)	$O(N^3(c-1)!)$
4	Multiple Strands, Bounded ($\leq$ c)	?

N bases, c strands

All of these are dynamic programming algorithms

Subprblems → Big problem

Local point of view → ? → Global property



Possible approach

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

N bases, c strands

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

Possible approach

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

N bases, c strands

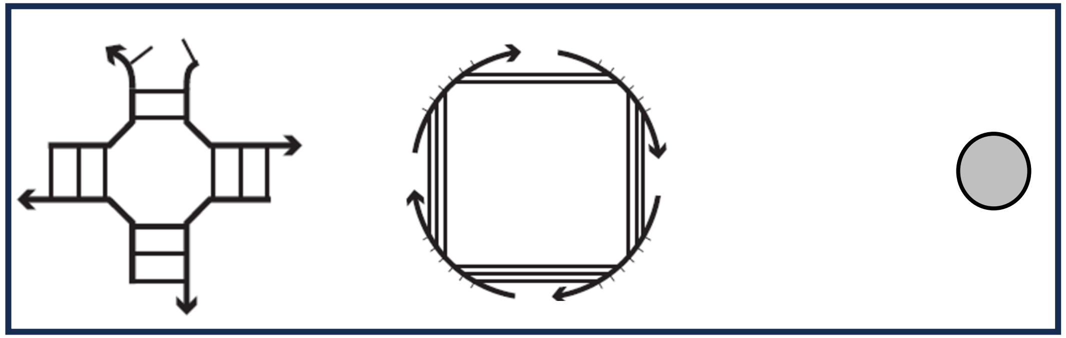
$$\Delta G(S) = \sum_l \Delta G(l) + (c-1)\Delta G^{\text{assoc}} + \boldsymbol{k_B T * \log R}$$

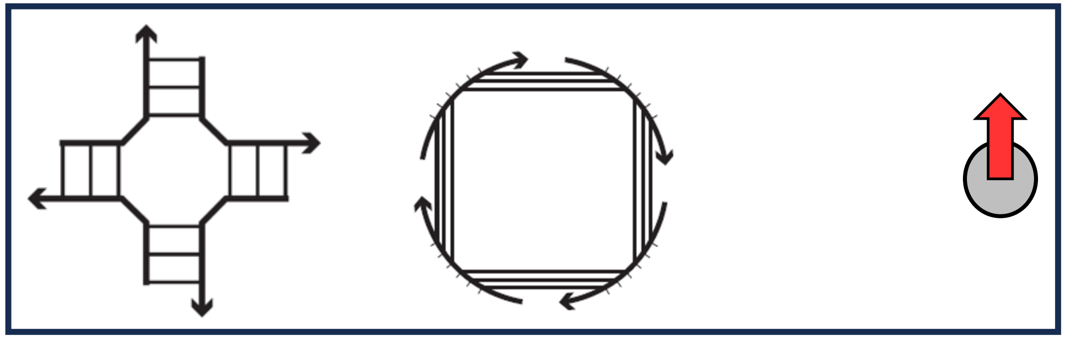
$$\Delta G'(S) = \sum_l \Delta G(l) + (c-1)\,\Delta G^{\text{assoc}}$$

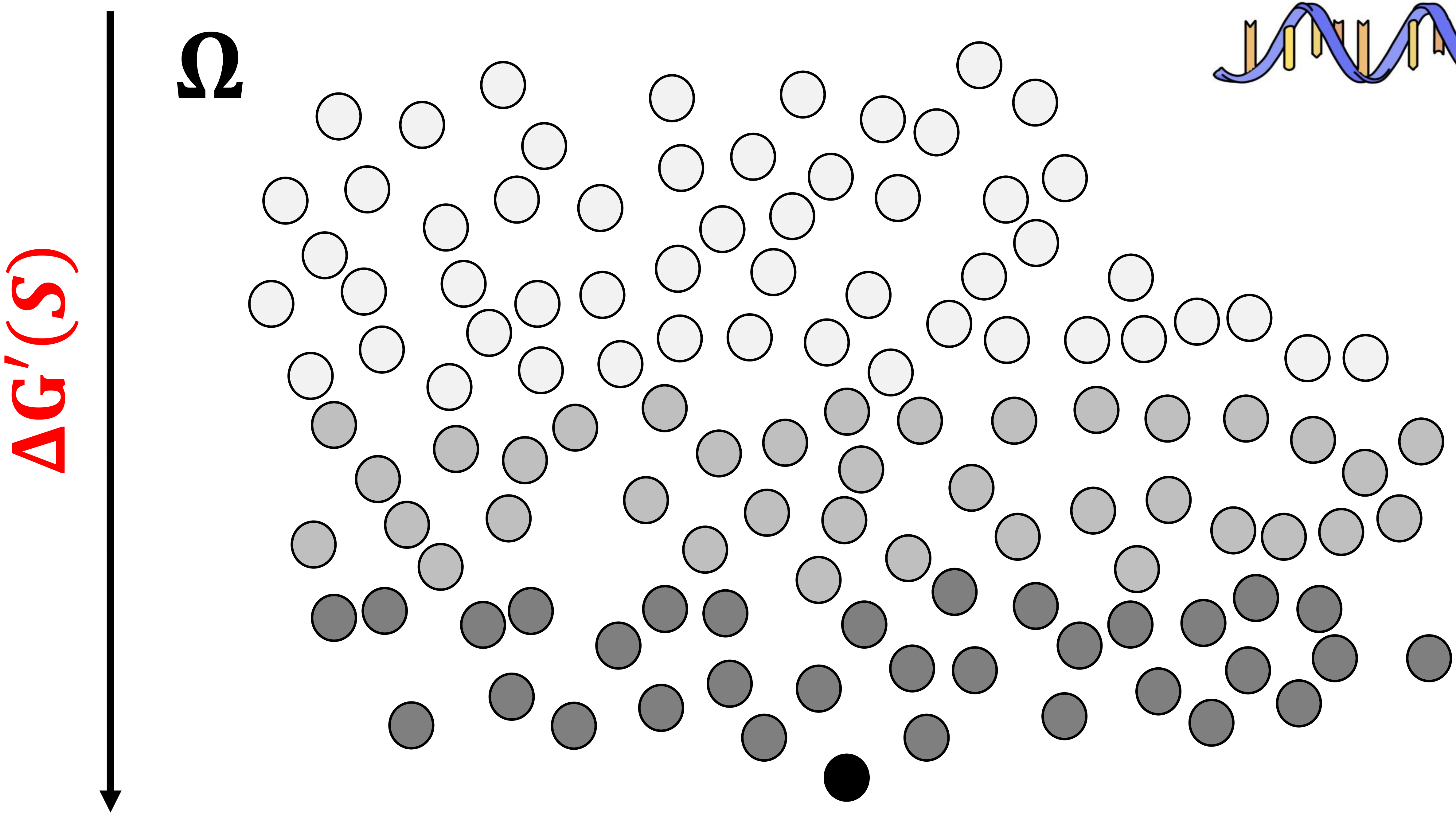
Ignore symmetry

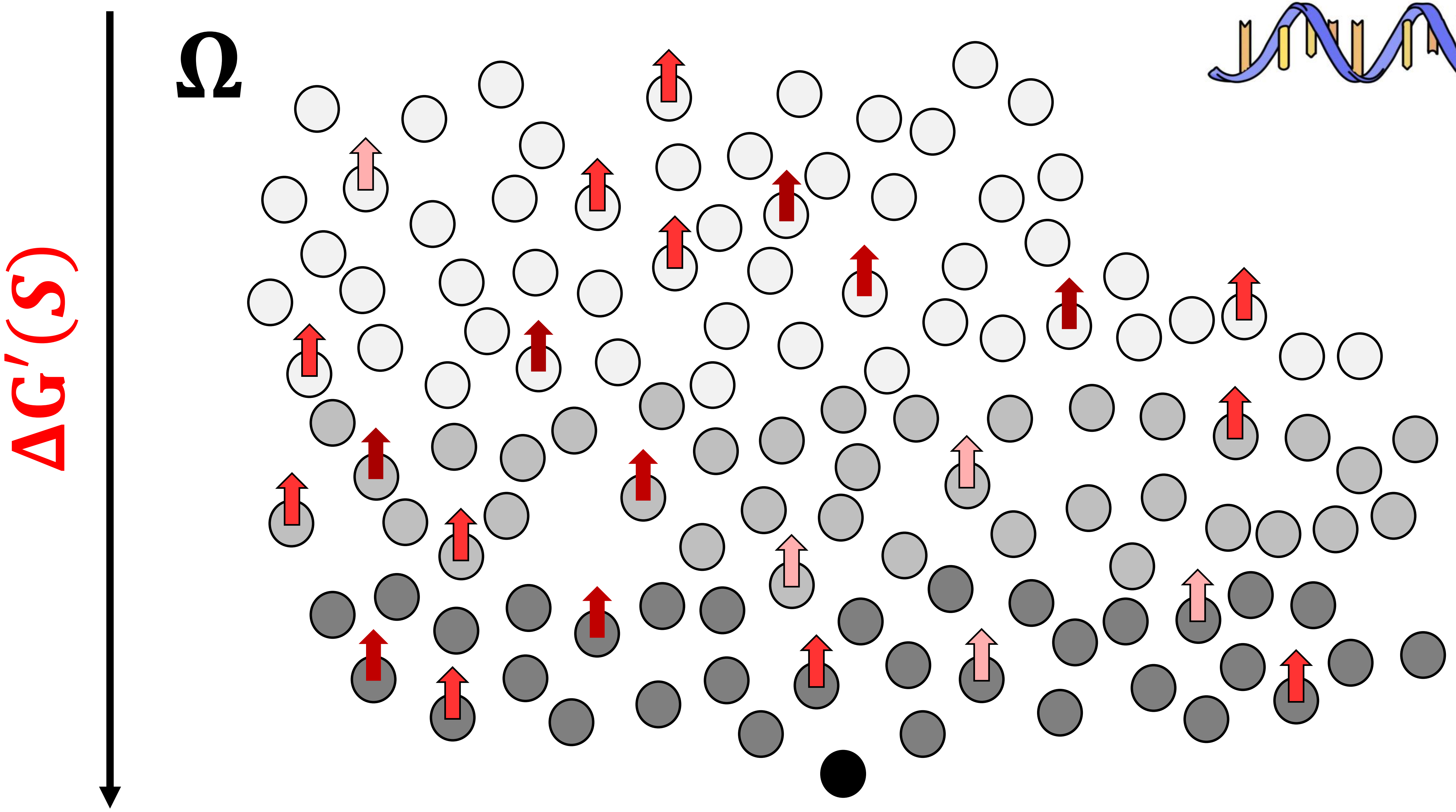
$\Delta G'(S)$

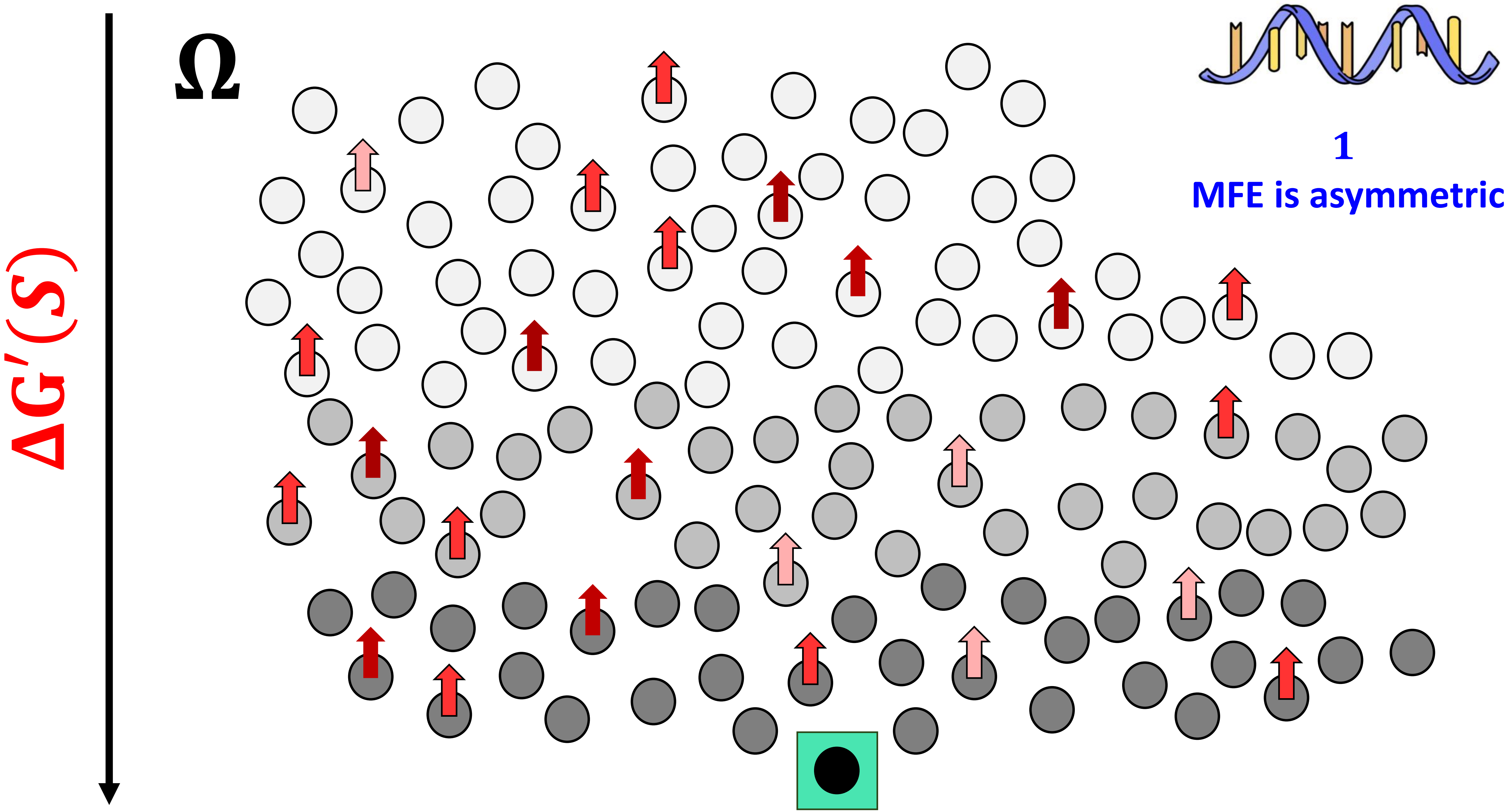
Let's ignore the symmetry for a while



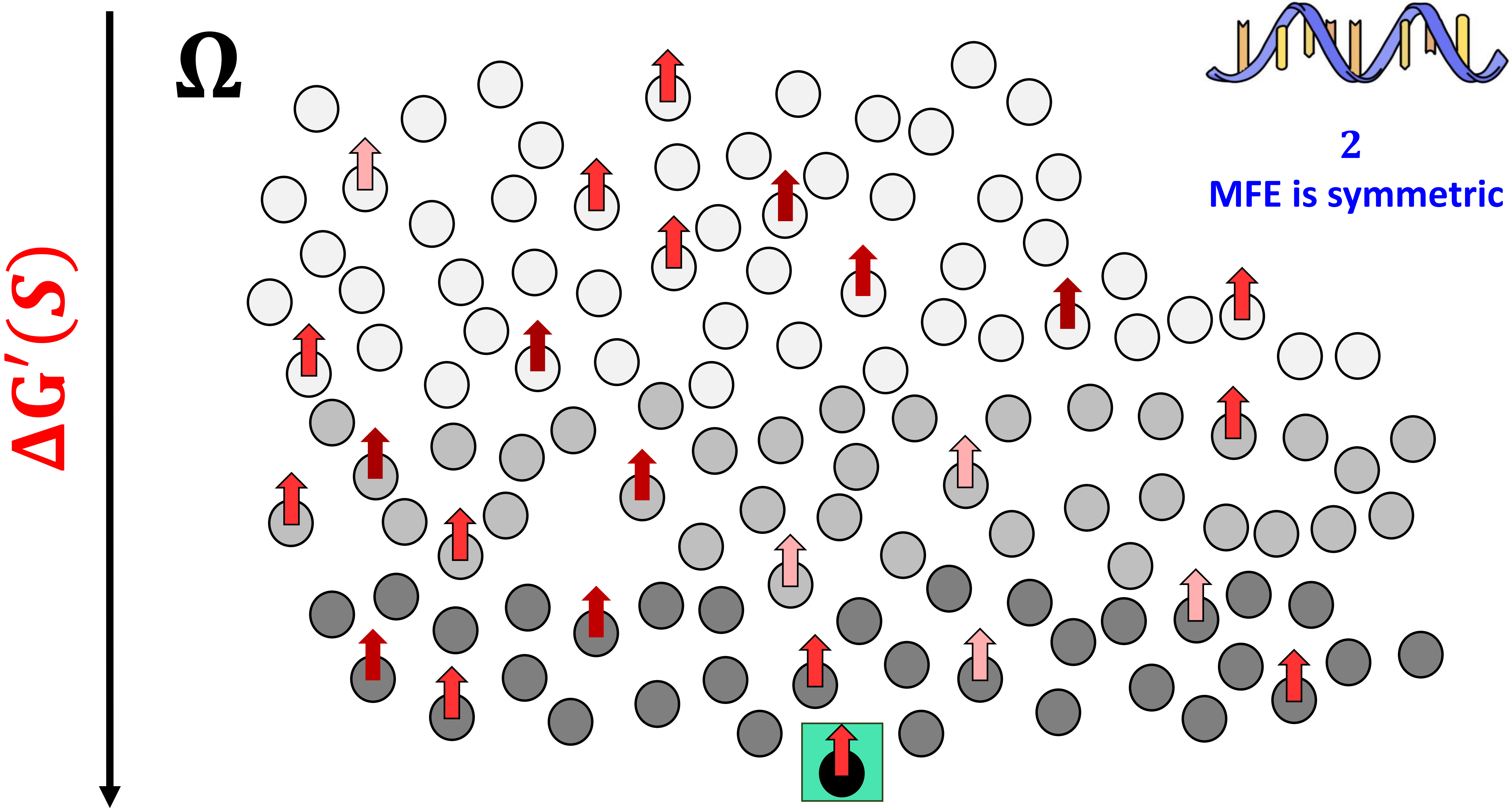


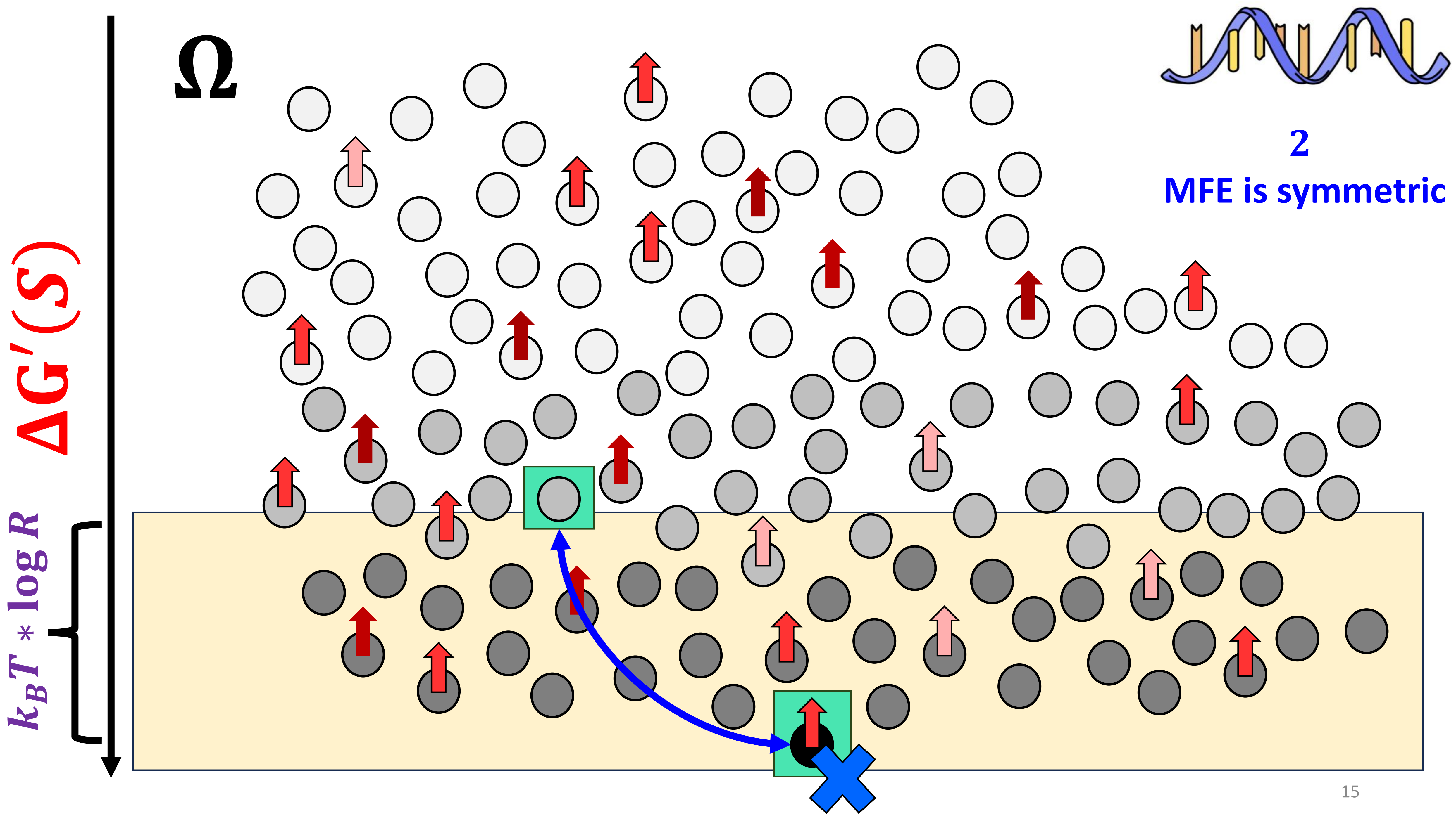

Ω
ΔG′(S)

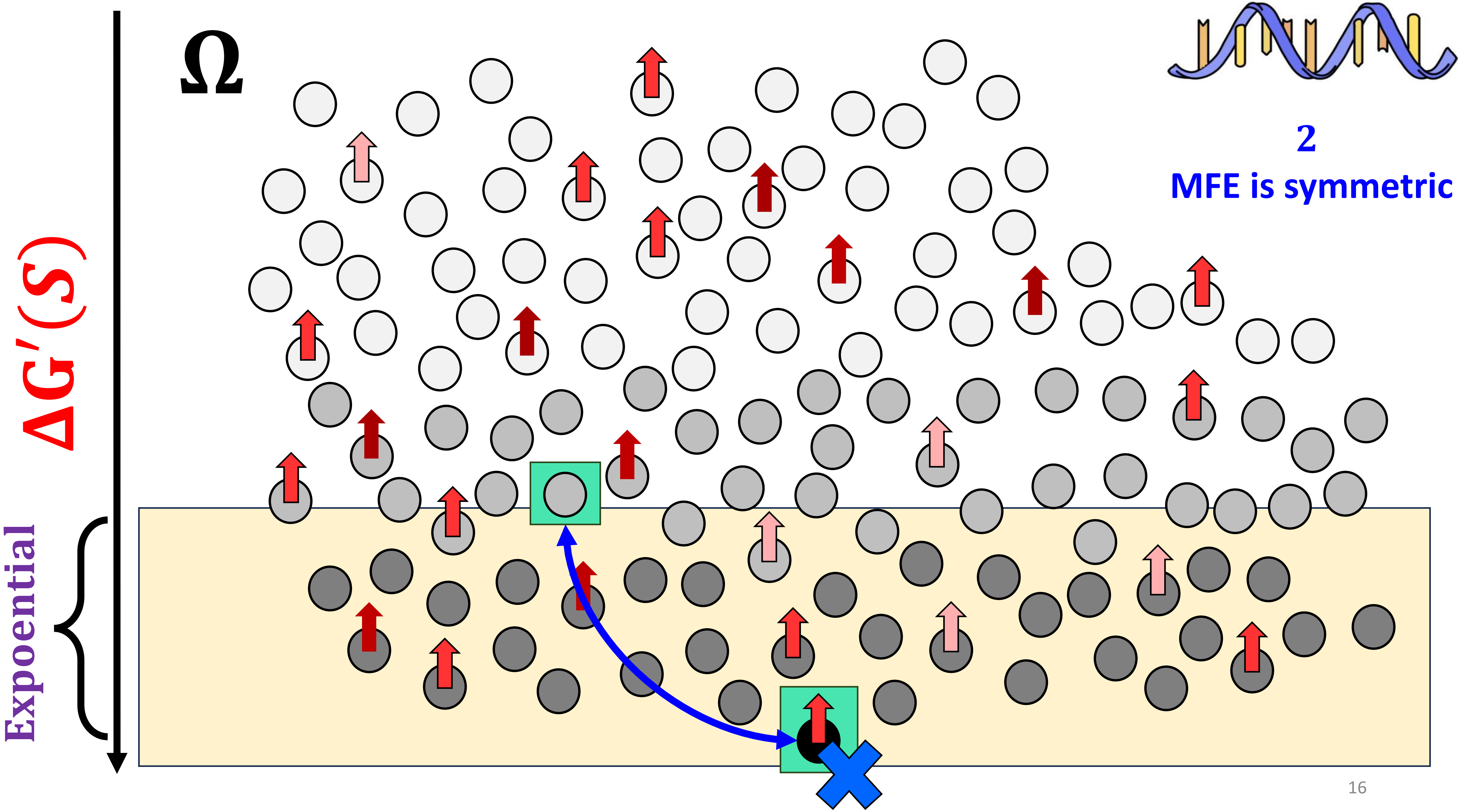

Ω
ΔG′(S)



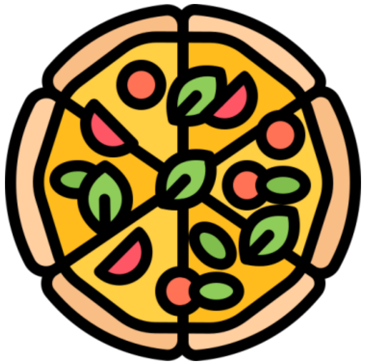
1
MFE is asymmetric

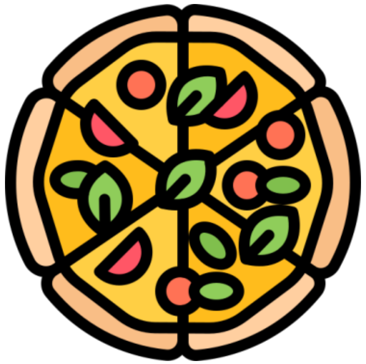

Ω
ΔG′(S)
2
MFE is symmetric


Ω
$\Delta G'(S)$
$k_B T * \log R$
2
MFE is symmetric

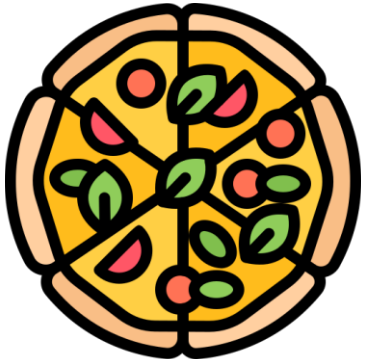

Ω
ΔG′(S)
Exponential
2
MFE is symmetric

Our solution





S_x

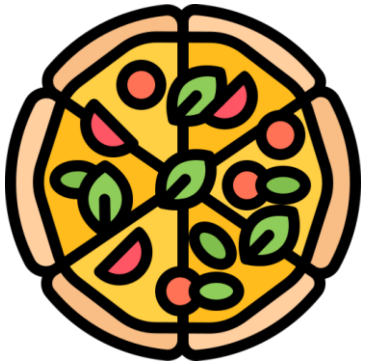


S_y

$\Delta G'(S)$

$\Delta G'(S_y) \leq \Delta G'(S_x)$

S_x

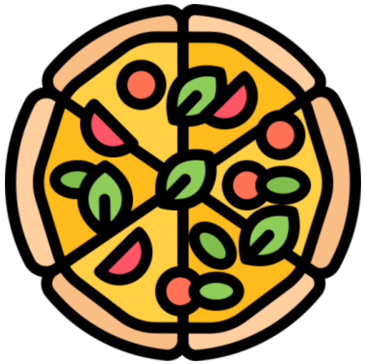


S_y

$\Delta G'(S)$

$\Delta G'(S_y) \leq \Delta G'(S_x)$

S_x **Symmetric**



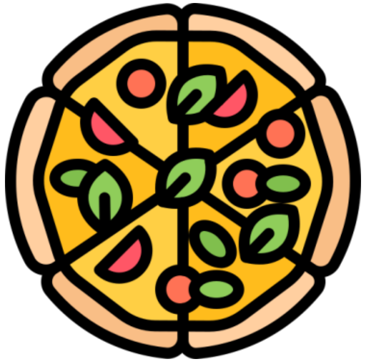
S_y **Symmetric**

$\Delta G'(S)$

$\Delta G'(S_y) \leq \Delta G'(S_x)$

S_x **Symmetric**

S_x and S_y **Admissible cut**



S_y **Symmetric**

$\Delta G'(S)$

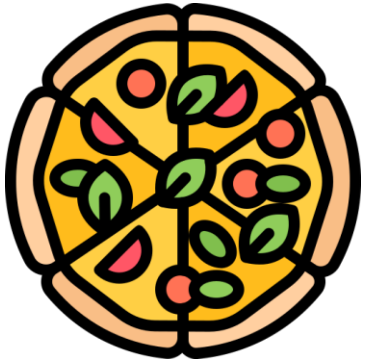
$\Delta G'(S_y) \leq \Delta G'(S_x)$

S_x Symmetric

X

S_z Asymmetric

S_x and S_y Admissible cut



S_y Symmetric

$\Delta G'(S)$

$\Delta G'(S_y) \leq \Delta G'(S_x)$

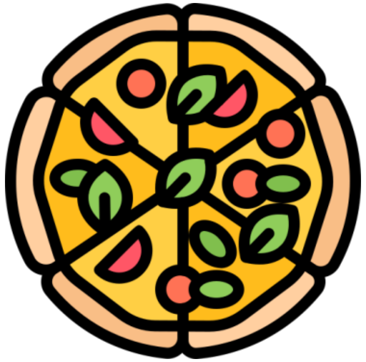
$\Delta G'(S_y) \leq G(S_z) \leq \Delta G'(S_x)$

S_x **Symmetric**

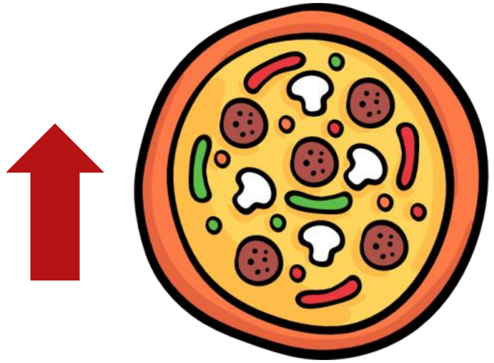
X

S_z **Asymmetric**

S_x and S_y Admissible cut



S_y **Symmetric**



S_x

Symmetric

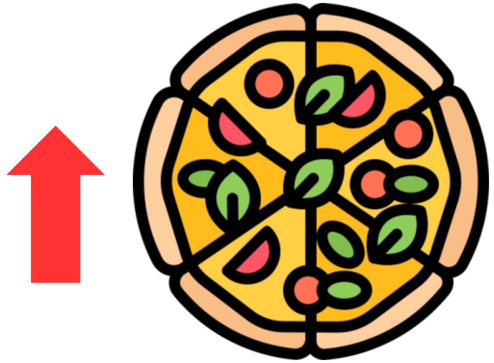
X

S_z

Asymmetric

S_x and S_y

Admissible cut



S_y

Symmetric

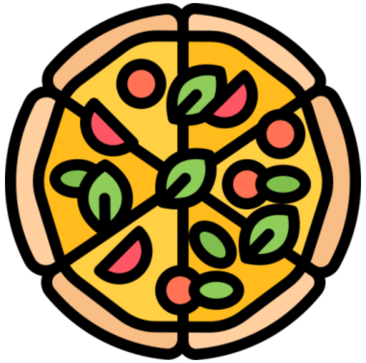
S_x

Symmetric

X 

S_z

Asymmetric



S_y

Symmetric

S_x and S_y

Admissible cut

S_x

Symmetric

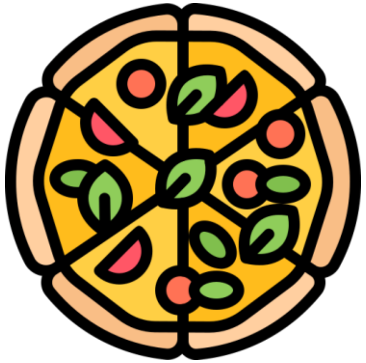
X

S_z

Asymmetric

S_x and S_y

Admissible cut



S_y

Symmetric

Upper bound

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

$$+$$

$$\boxed{N^2/16}$$

S_x

Symmetric

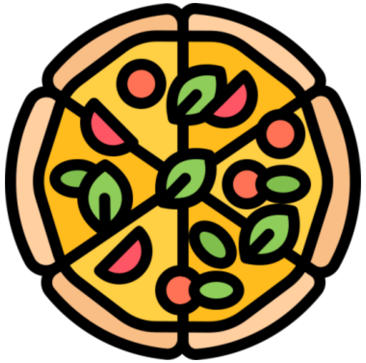
X

S_z

Asymmetric

S_x **and** S_y

Admissible cut



S_y

Symmetric

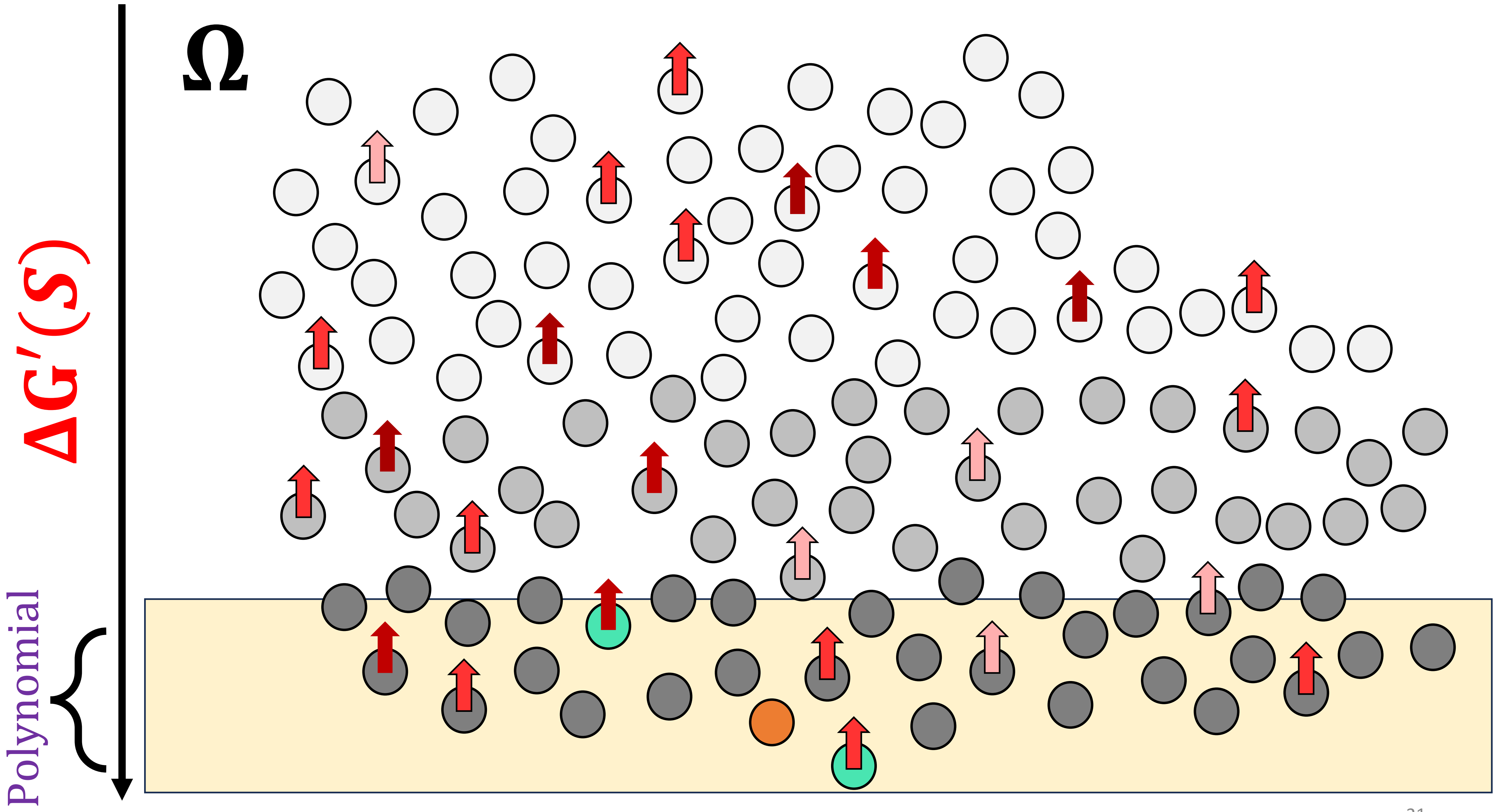
Upper bound

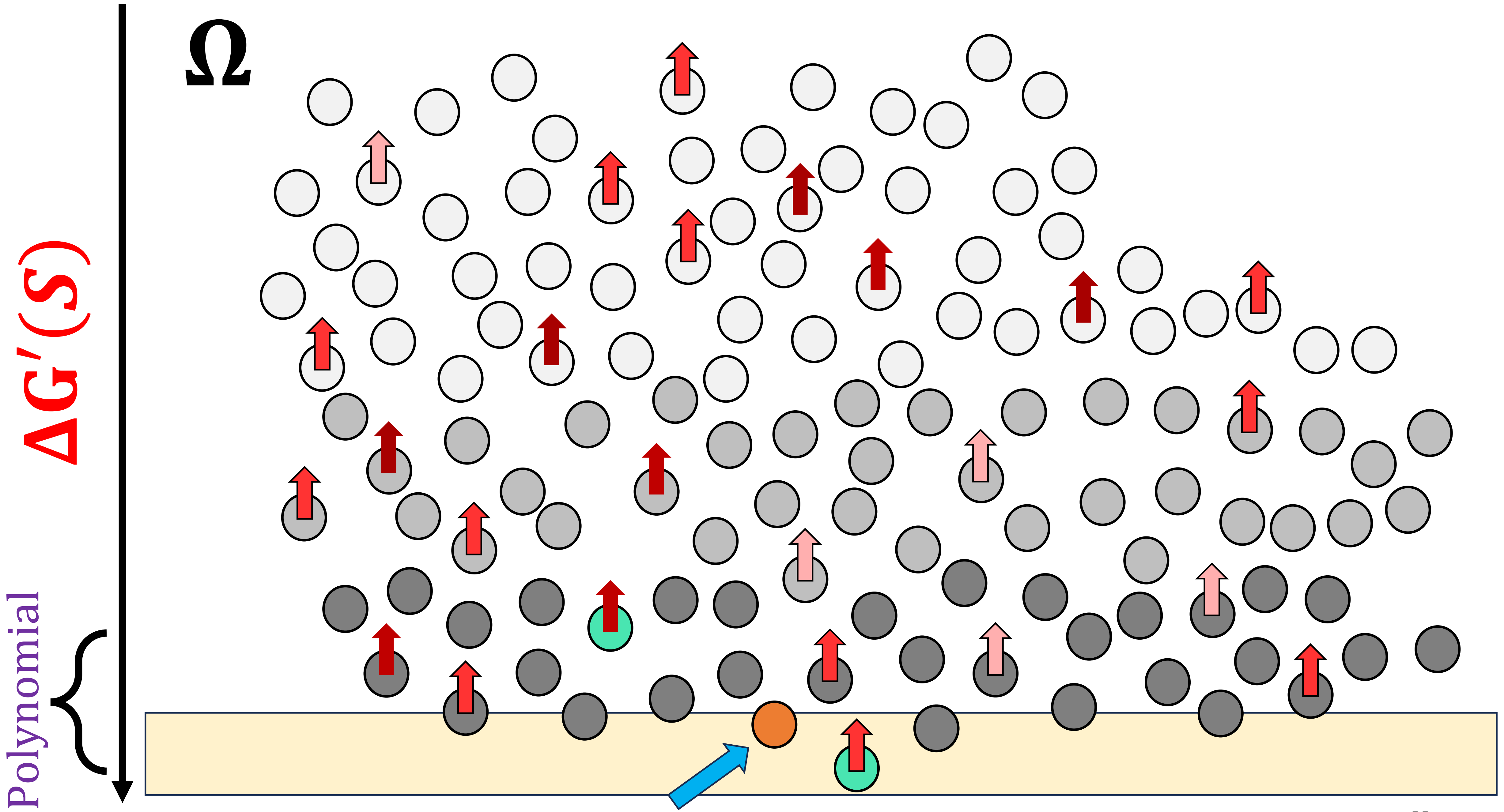
$$\boxed{\frac{N-c}{v(\pi)}\left(\sigma(v(\pi)) - v(\pi)\right)}$$

$$+$$

$$\boxed{N^2/16}$$

Adjusting the backtracking algorithm to go through energy levels sequentially starting from the MFE level.


Ω
ΔG′(S)
Polynomial


Ω
ΔG′(S)
Polynomial

Computational complexity of Minimum Free Energy algorithms

Input Type	MFE
Single Strand (Loop model)	$O(N^3)$
Multiple **unique** Strands, Bounded ($\leq$ c)	$O(N^3(c-1)!)$
Multiple Strands, Bounded ($\leq$ c)	$\mathbf{O\left(N^4(c-1)!\right)}$
Multiple Strands, Unbounded	NP – Complete

N bases, c strands

Thanks

Hamilton Institute

Maynooth University
National University of Ireland Maynooth

dna.hamilton.ie/shalaby

sfi Science Foundation Ireland

erc

European Innovation Council

Funded by the European Union