An efficient algorithm to compute the minimum free energy of interacting nucleic acid strands

Ahmed Shalaby, Damien Woods

Final year PhD


ALFR· NOBEL
NAT· MDCCC XXXIII OB· MDCCC XCVI

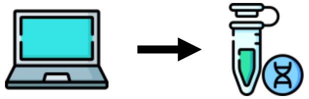

NAT.
MDCCC
XXXIII
OB.
MDCCC
XCVI
ALFR.
NOBEL

RNA/DNA folding problem

RNA/DNA folding problem

Aptamers: Movie by WuXi AppTec

RNA/DNA folding problem

 DNA Computing

Aptamers: Movie by WuXi AppTec

RNA/DNA folding problem

DNA Computing

Aptamers: Movie by WuXi AppTec

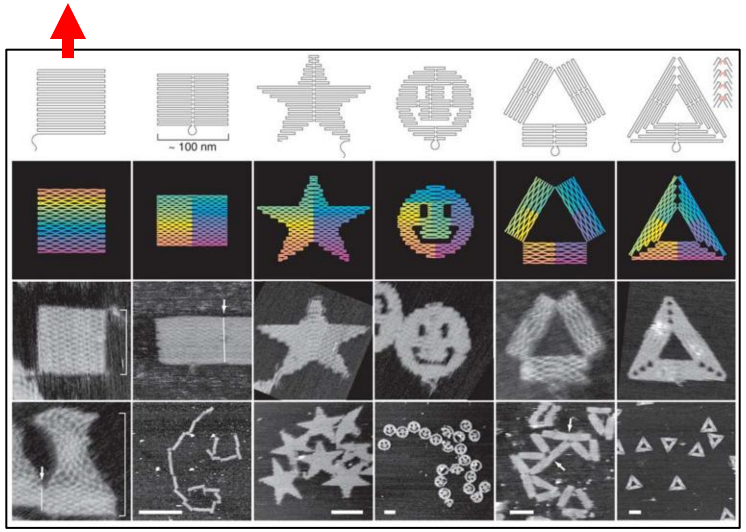
Movie by Shawn Douglas

RNA/DNA folding problem

DNA Computing

Aptamers: Movie by WuXi AppTec

Movie by Shawn Douglas



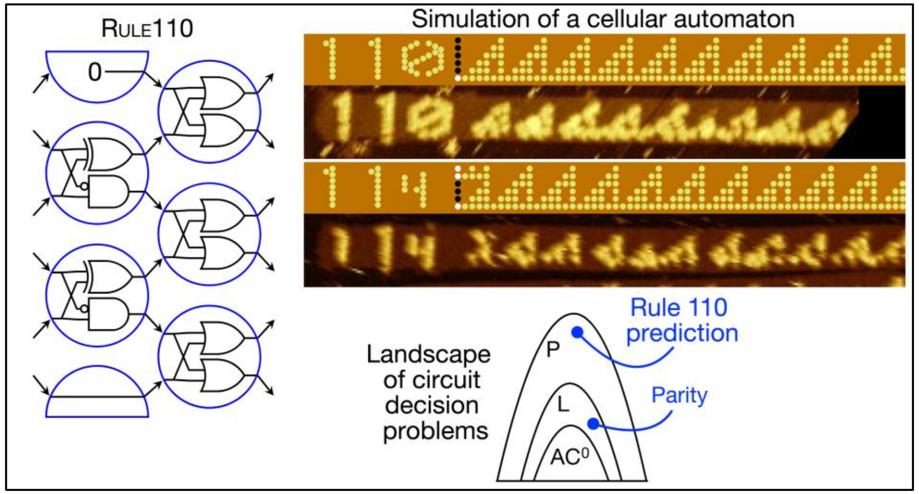
Rothemund 2006

RNA/DNA folding problem

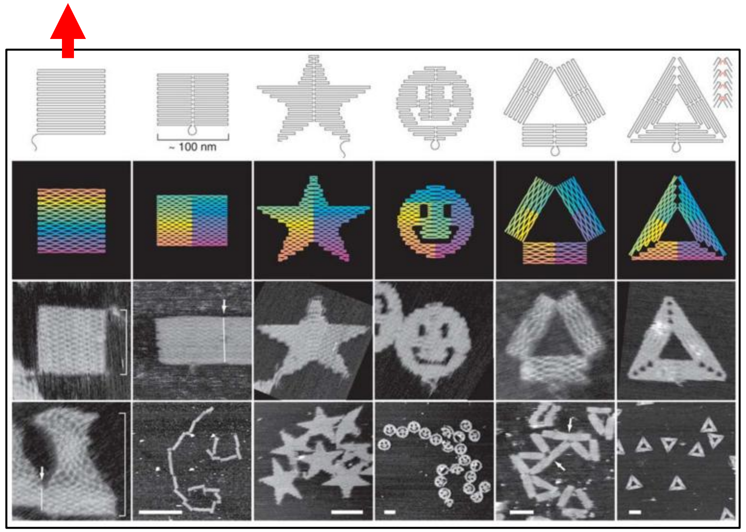
DNA Computing

Aptamers: Movie by WuXi AppTec

Movie by Shawn Douglas

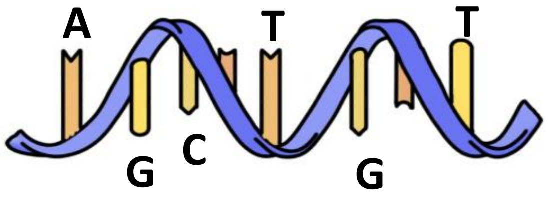


Doty, Woods et al. 2017

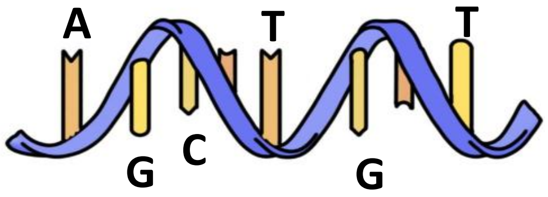


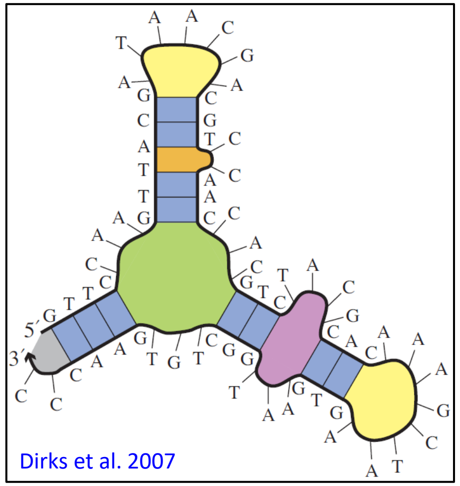
Rothemund 2006

RNA/DNA secondary structure



RNA/DNA secondary structure

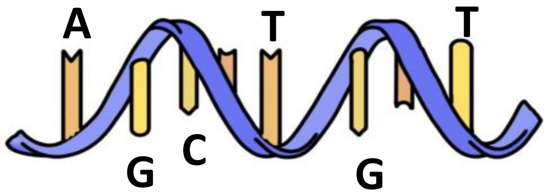


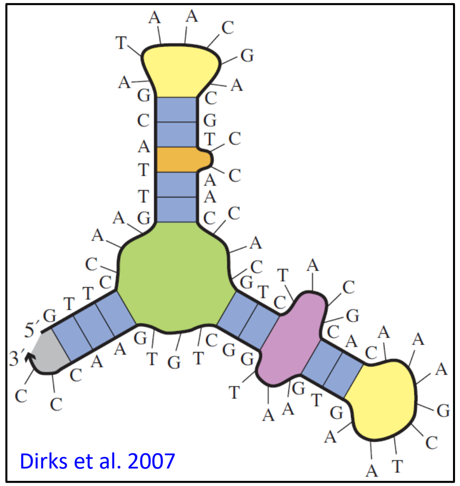


Dirks et al. 2007

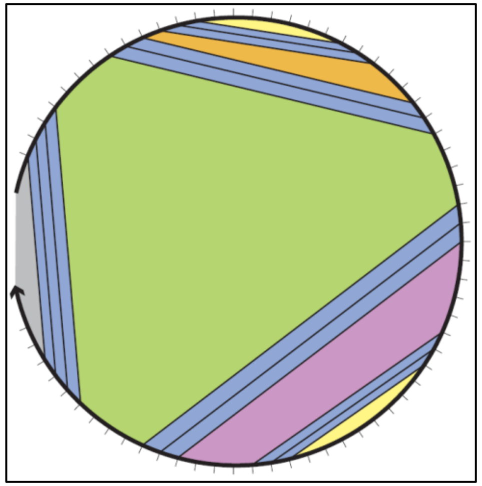
Secondary structure

RNA/DNA secondary structure



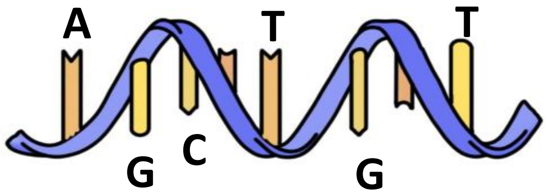


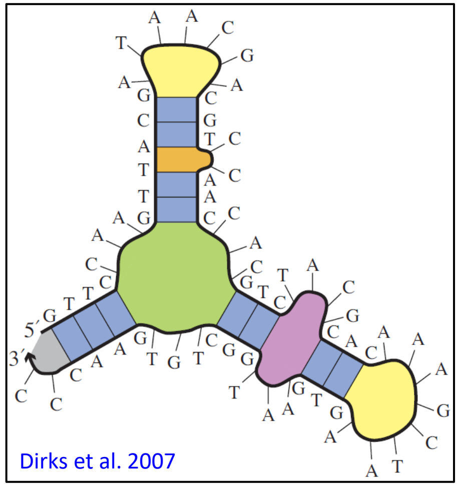
Secondary structure



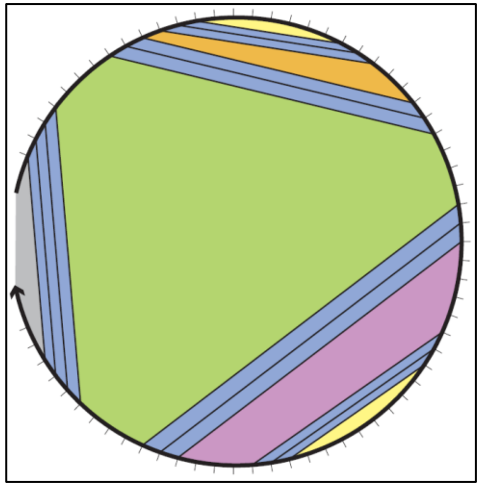
Polymer graph representation

RNA/DNA secondary structure





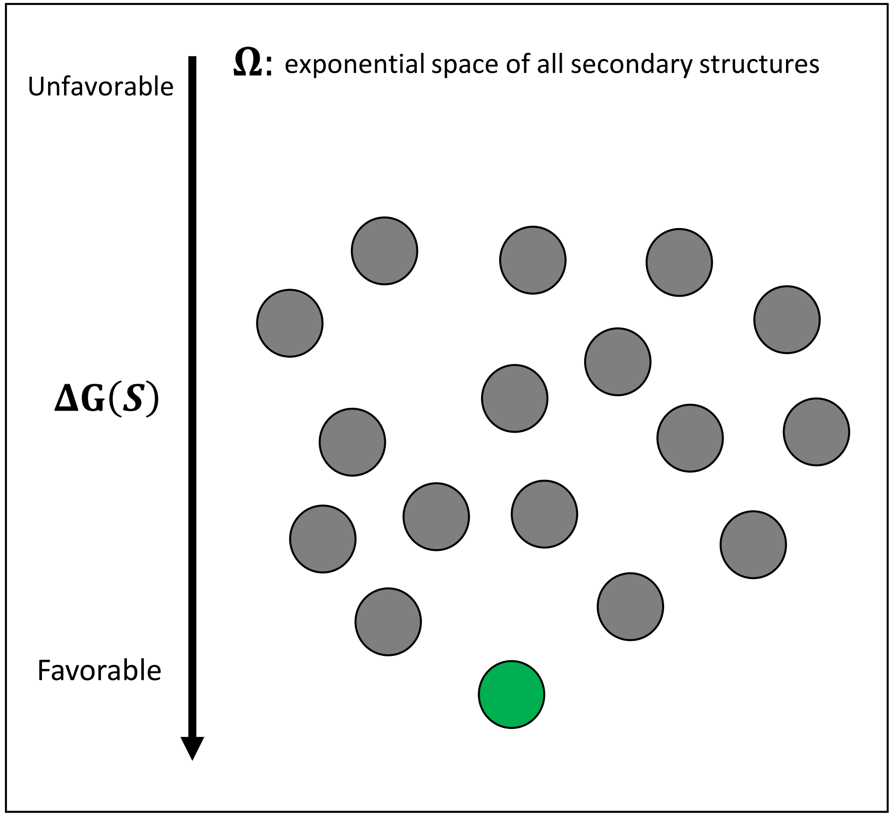
Secondary structure



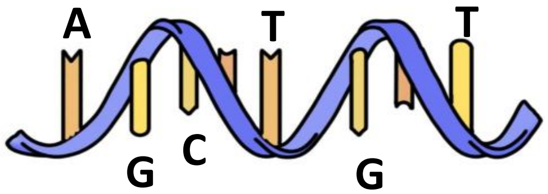
Polymer graph representation

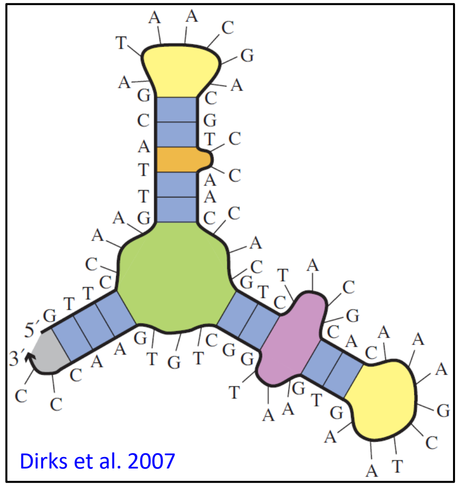
$\boldsymbol{\Delta G(S)}$

Energy model

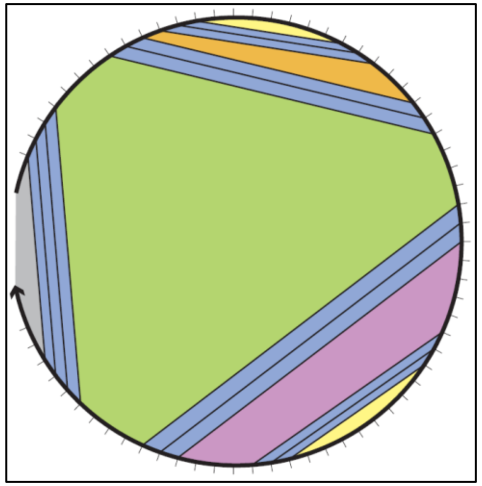


RNA/DNA secondary structure





Secondary structure

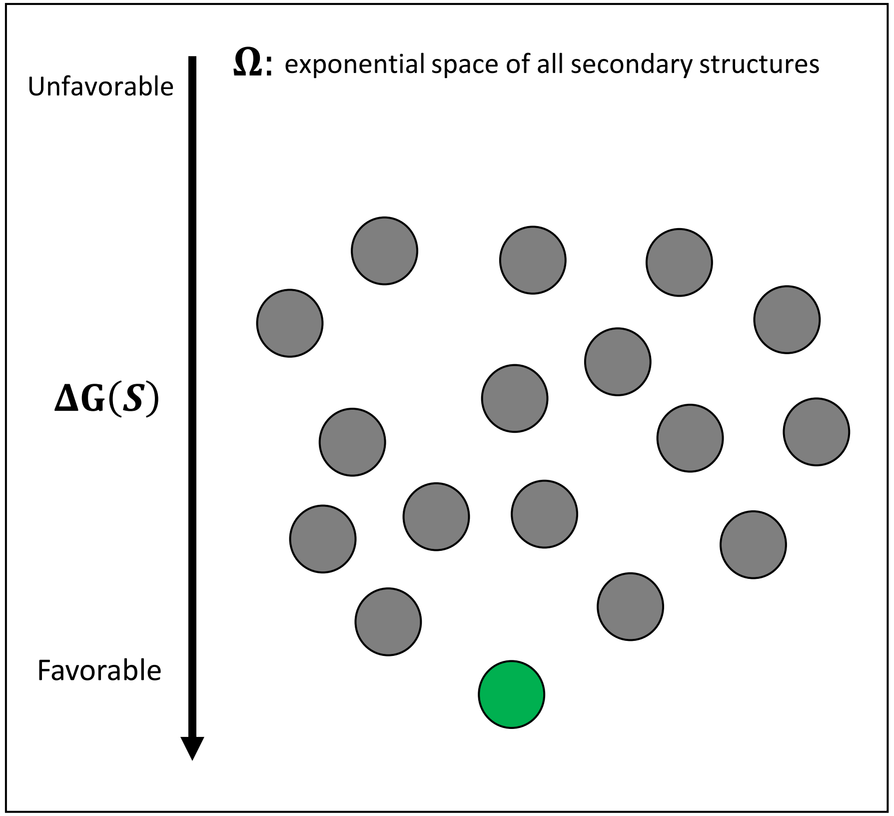


Polymer graph representation

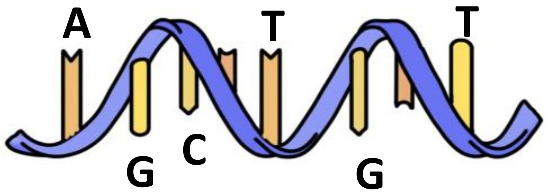
$$\Delta G(S)$$

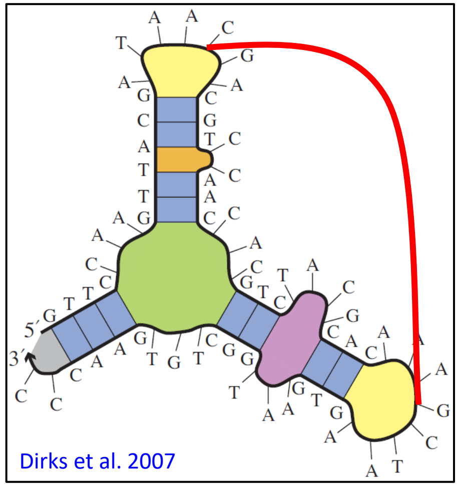
Energy model

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

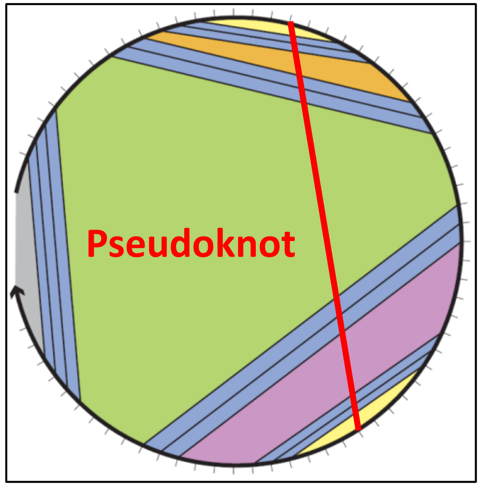


RNA/DNA secondary structure





Secondary structure



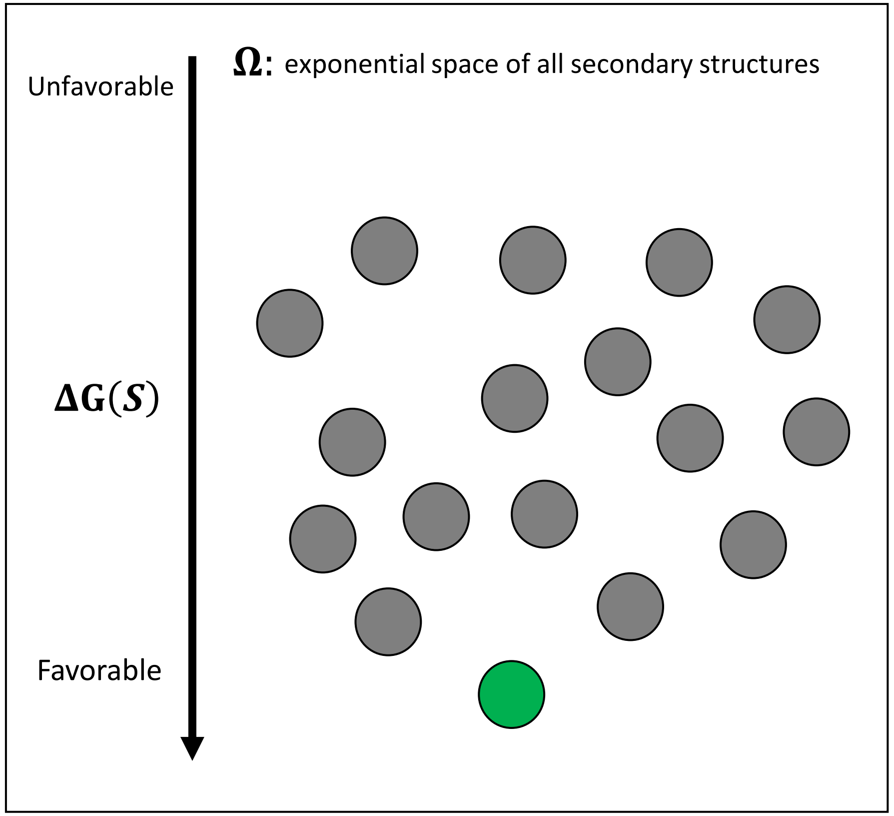
Polymer graph representation

$\mathbf{NP-Complete}$

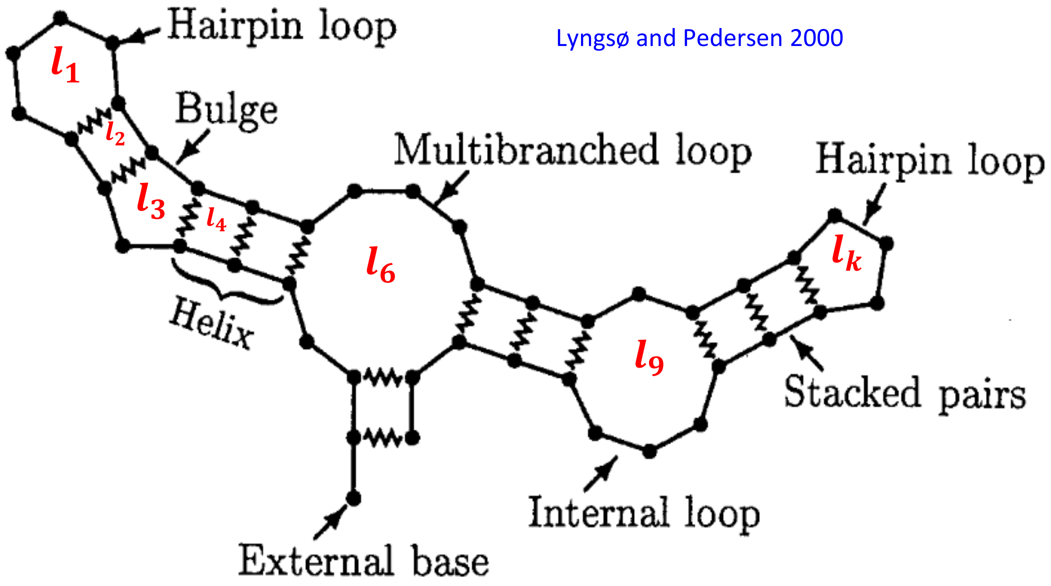
$$\boldsymbol{\Delta G(S)}$$

Energy model

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



Energy model: Nearest neighbour model

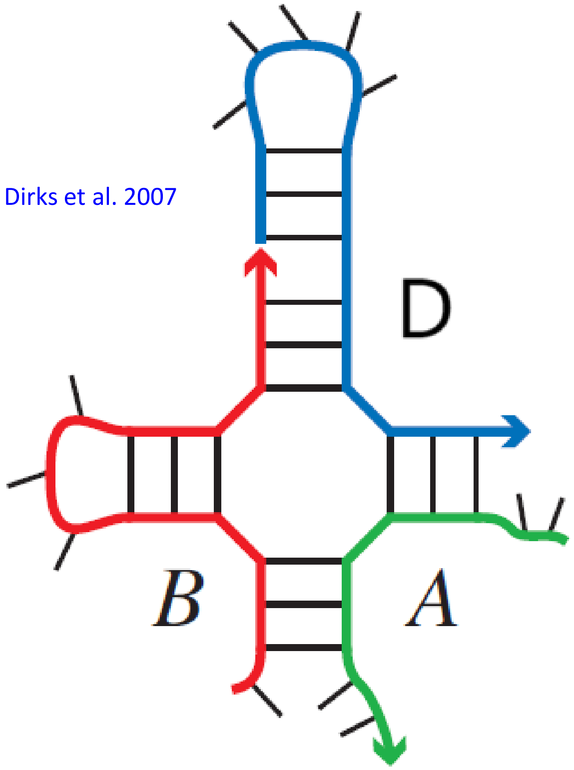


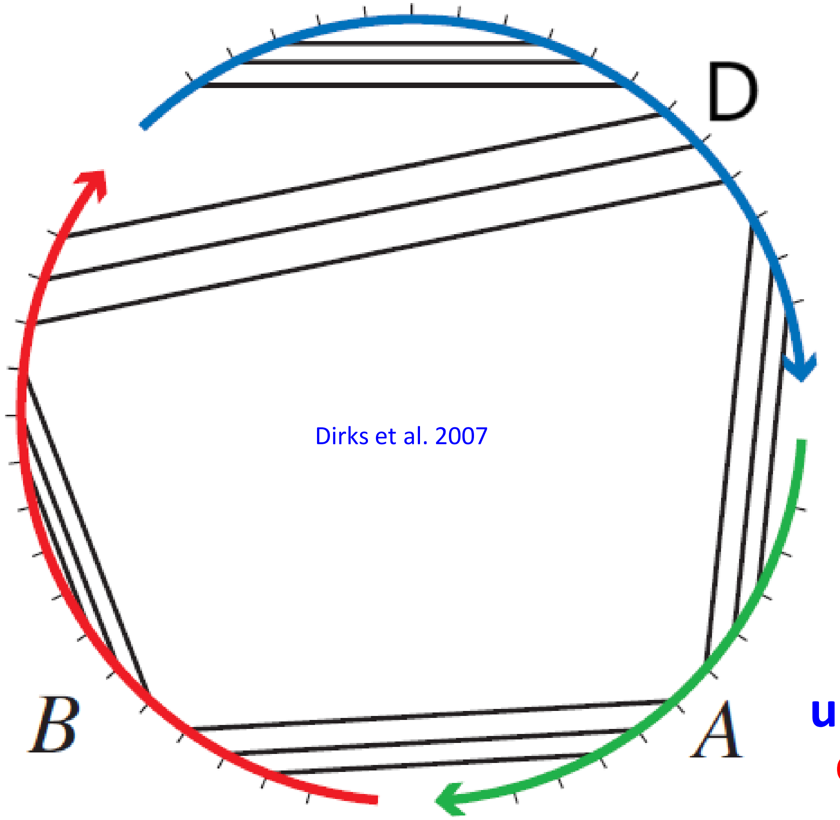
Lyngsø and Pedersen 2000

$$\boldsymbol{\Delta G(S) = \Delta G(l_1) + \Delta G(l_2) + \cdots + \Delta G(l_k)}$$

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

Energy model: Nearest neighbour model



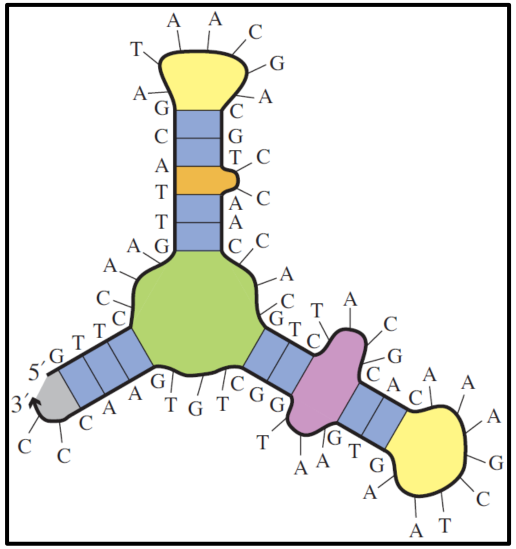


unique strands
c = 3 strands

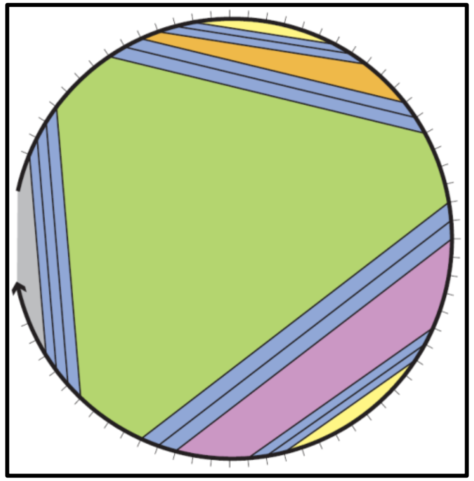
$$\Delta G(S) = \sum_{l \in S} \Delta G(l) + (c-1)\,\Delta G^{\mathrm{assoc}}$$

The Nearest Neighbour model for RNA/DNA

Single stranded system



Dirks et al. 2007

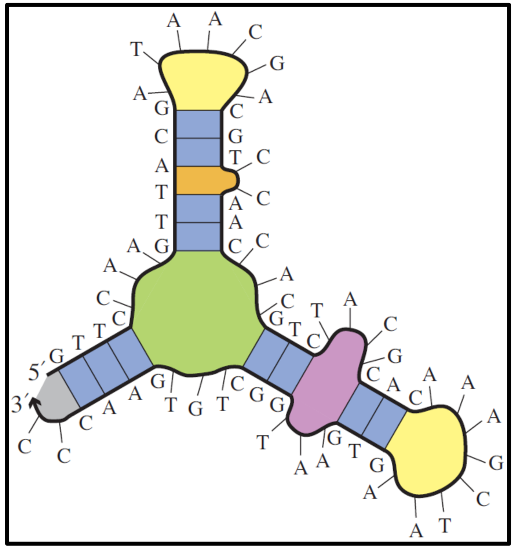


Hairpin loop
Bulge
Stack
Multiloop
Interior loop
Exterior loop

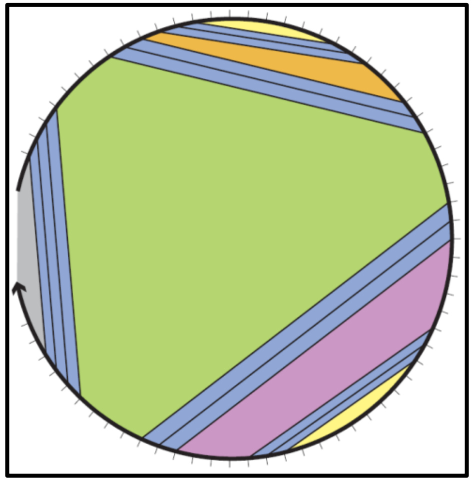
$$\Delta G(S) = \sum_{l \in S} \Delta G(l)$$

The Nearest Neighbour model for RNA/DNA

Single stranded system



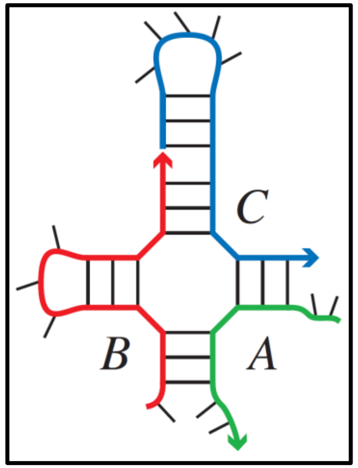
Dirks et al. 2007



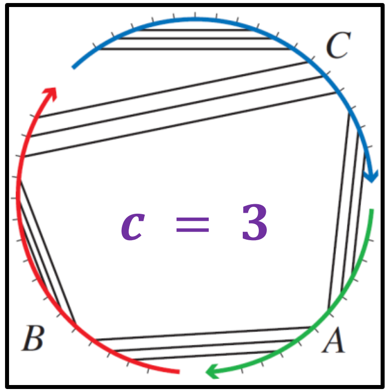
- Hairpin loop
- Bulge
- Stack
- Multiloop
- Interior loop
- Exterior loop

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

Multi-stranded system of c strands



Dirks et al. 2007

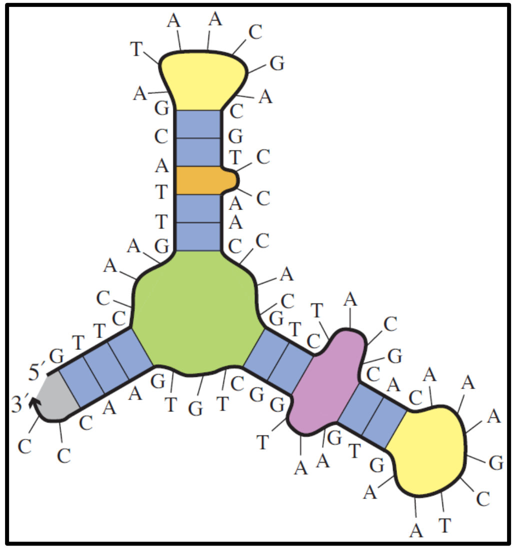


unique strands

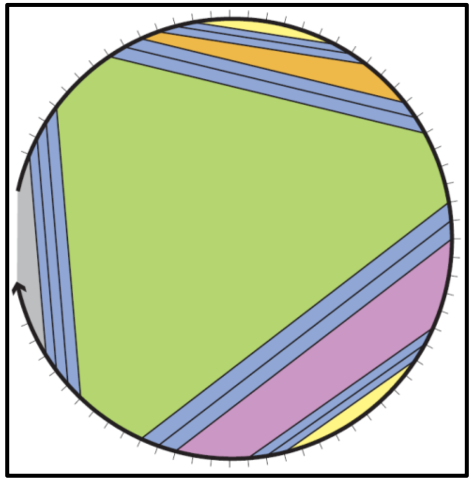
$$\Delta G(S) = \sum_{l} \Delta G(l) + \underbrace{(c-1)\,\Delta G^{\text{assoc}}}_{\text{Initiation penalty}}$$

The Nearest Neighbour model for RNA/DNA

Single stranded system



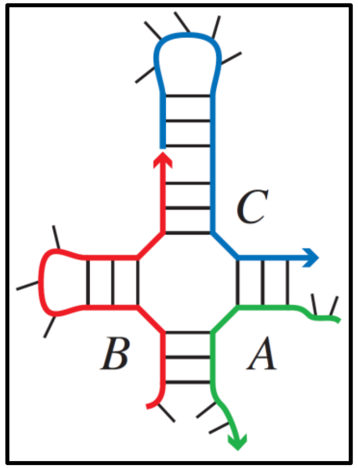
Dirks et al. 2007



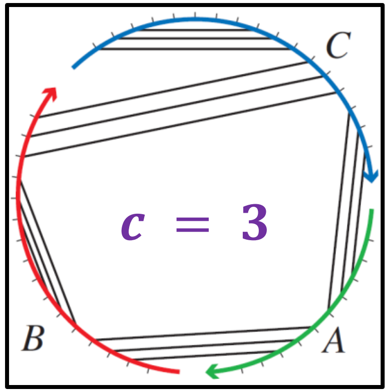
Hairpin loop | Bulge
Stack | Multiloop
Interior loop | Exterior loop

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

Multi-stranded system of c strands



Dirks et al. 2007

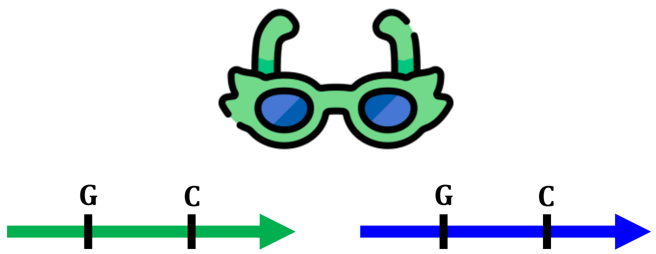


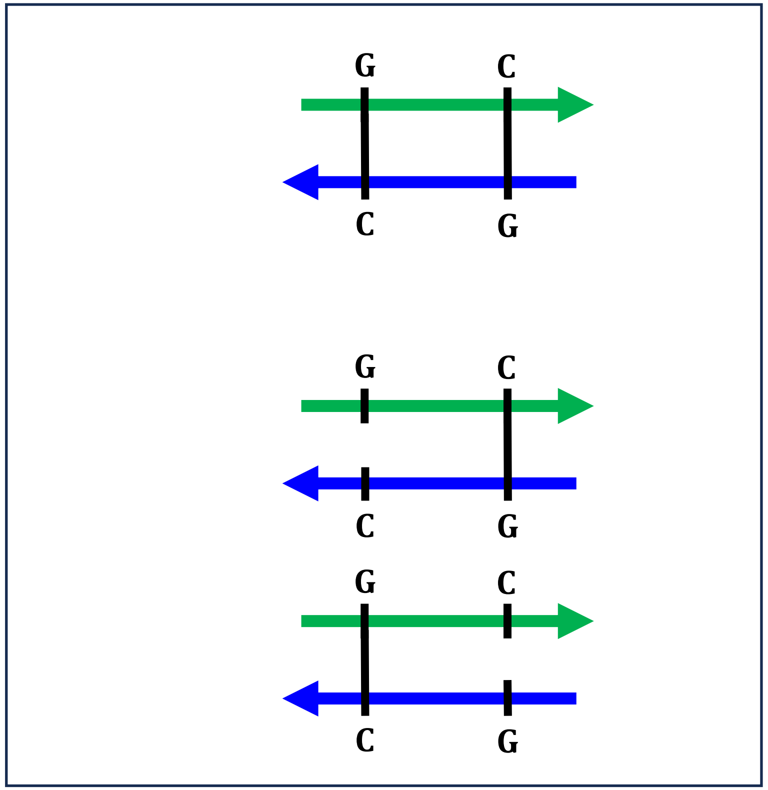
unique strands

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

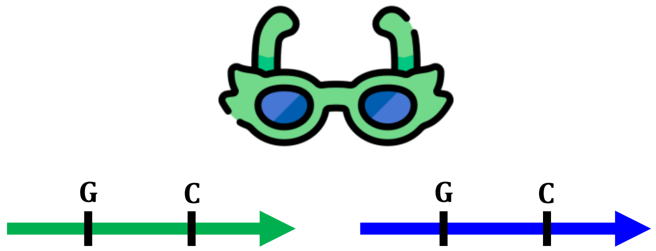
$$\text{MFE} = \min_{S \in \Omega} \Delta G(S)$$

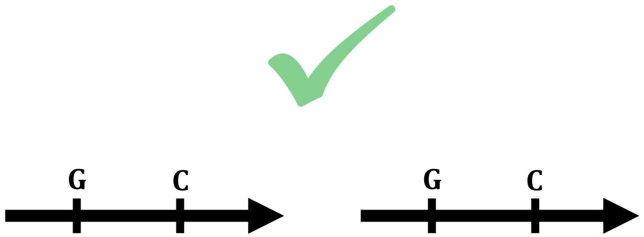
Ω: the set of all **connected <u>unpseudoknotted</u>** secondary structures

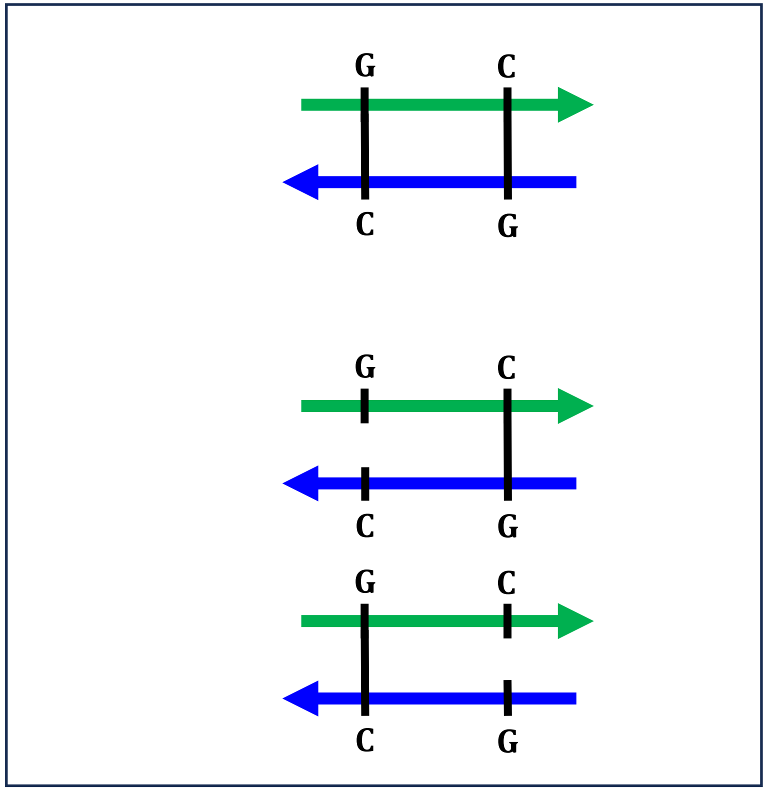

G
C
G
C



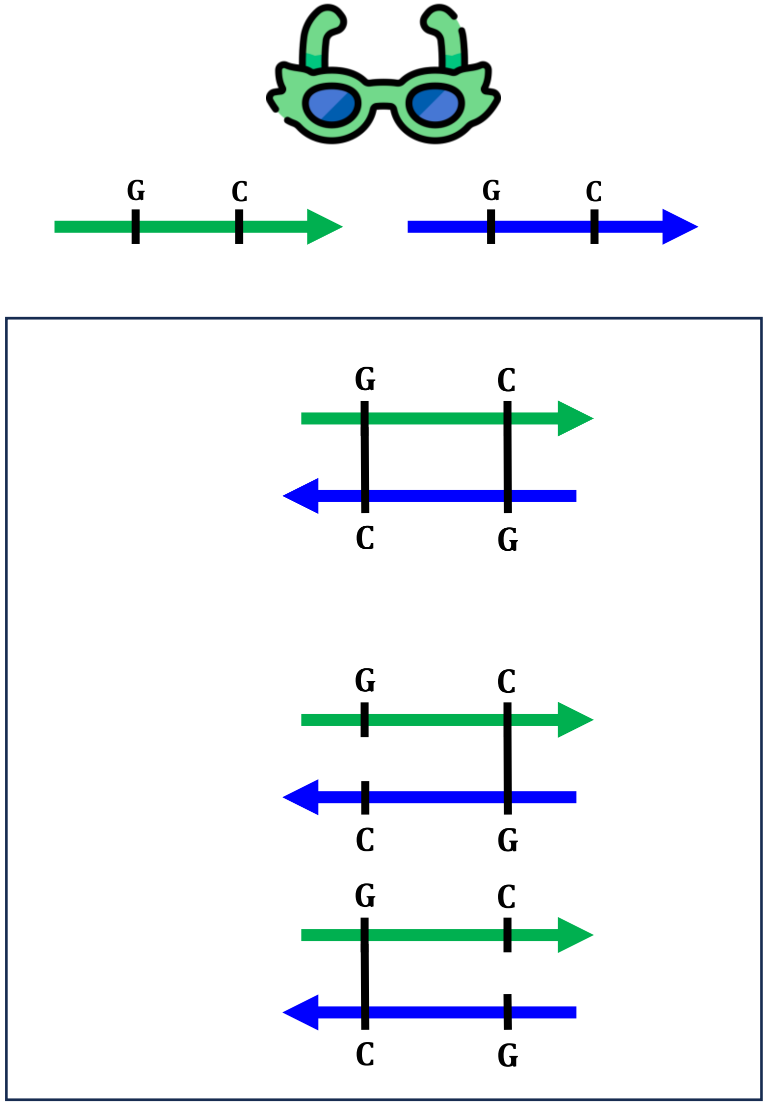
$$\Delta G(S) = \sum_{l} \Delta G(l) + (c-1)\,\Delta G^{\text{assoc}}$$

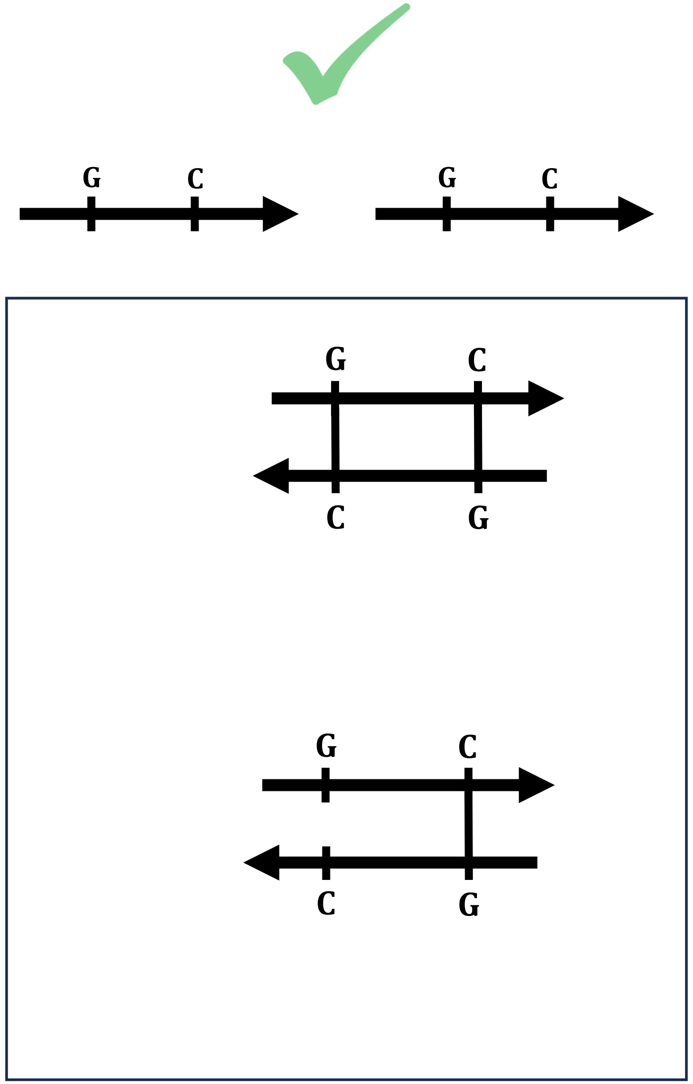




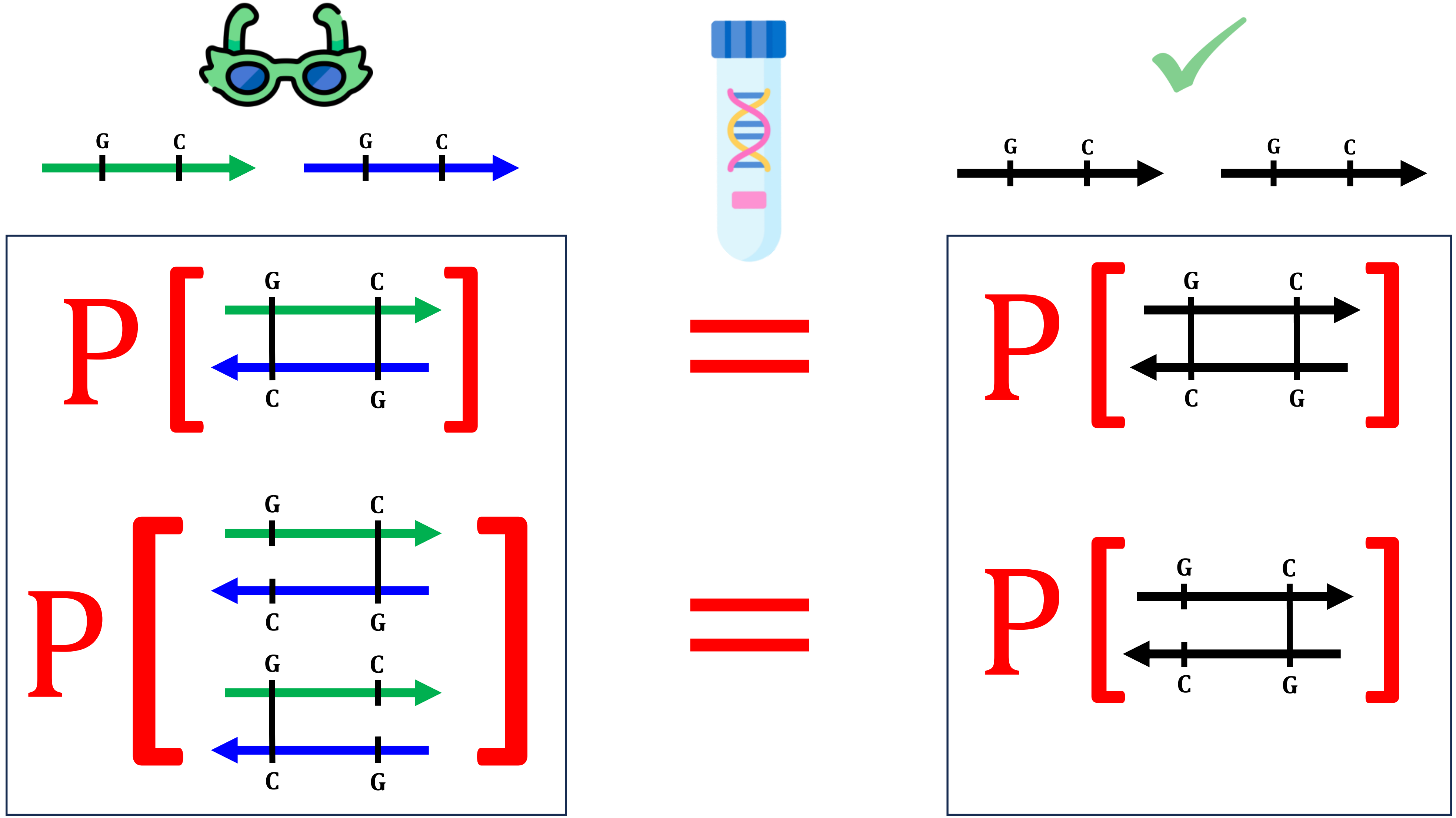


$$\Delta G(S) = \sum_{l} \Delta G(l) + (c-1)\,\Delta G^{\mathrm{assoc}}$$

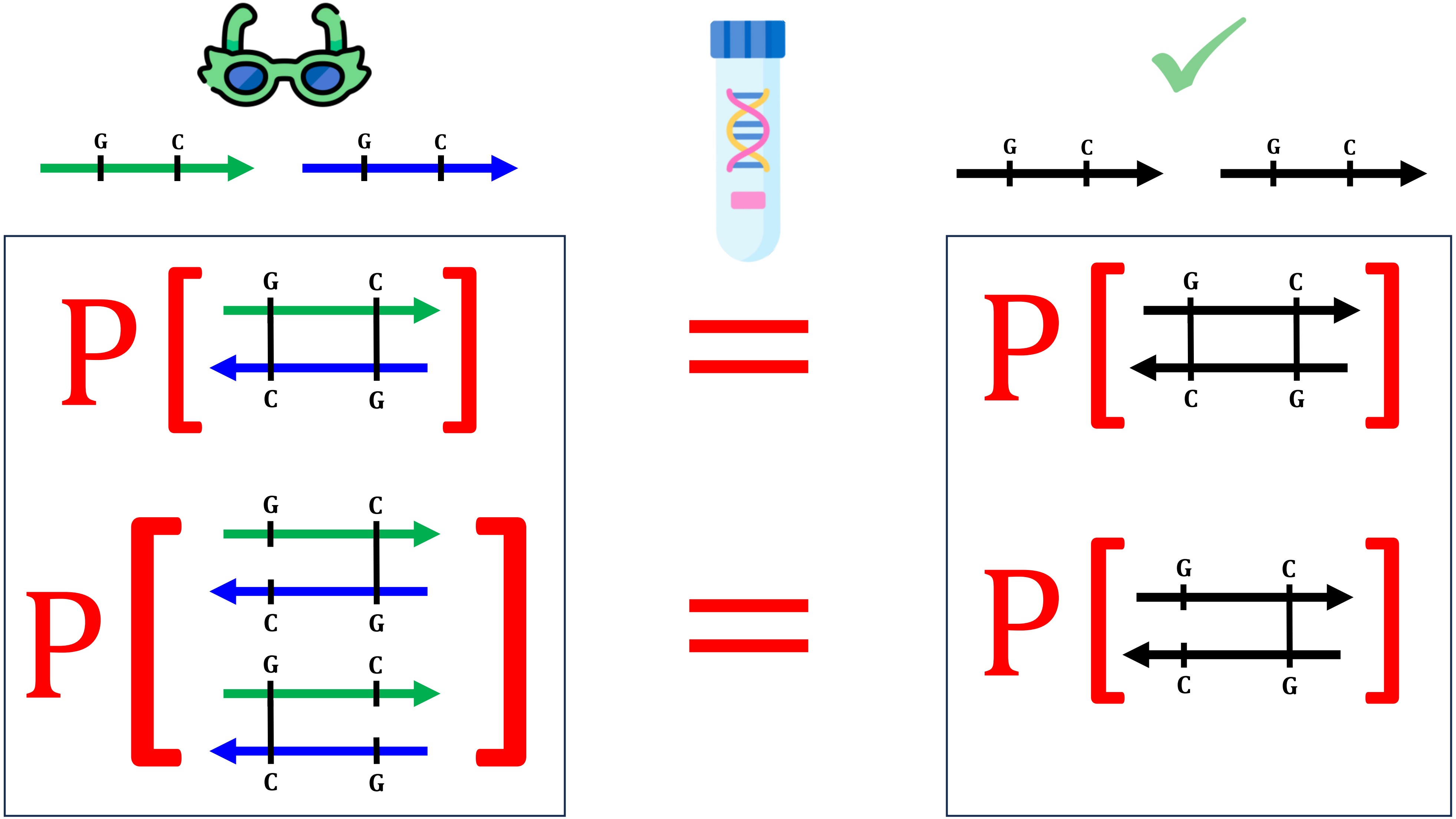




$$\Delta G(S) = \sum_{l} \Delta G(l) + (c-1)\, \Delta G^{\mathrm{assoc}}$$



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

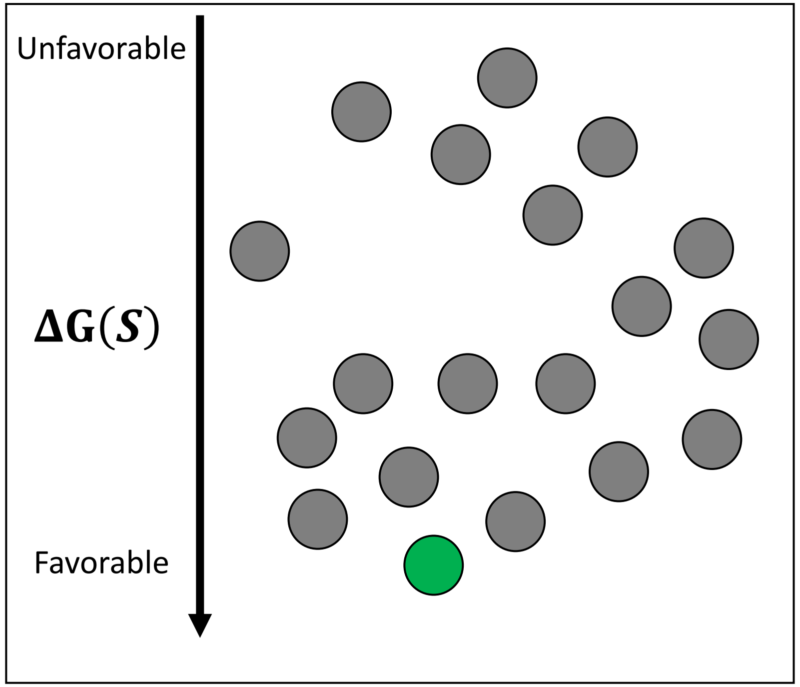


$$\Delta G(S) = \sum_{l} \Delta G(l) + (c-1)\,\Delta G^{\text{assoc}}$$

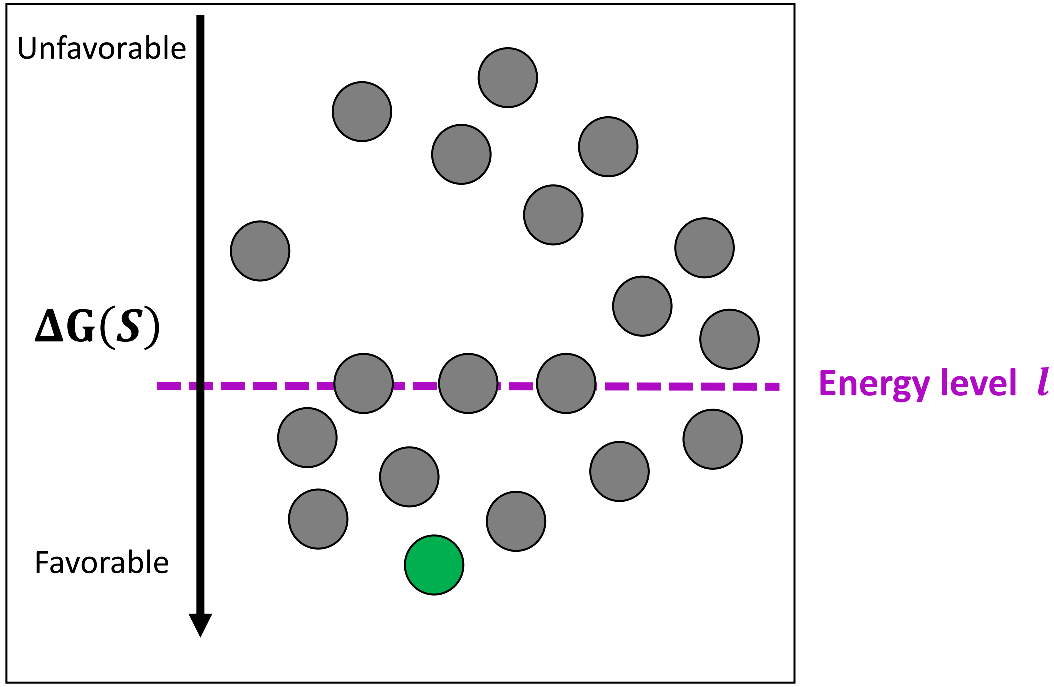
Entropic penalty

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

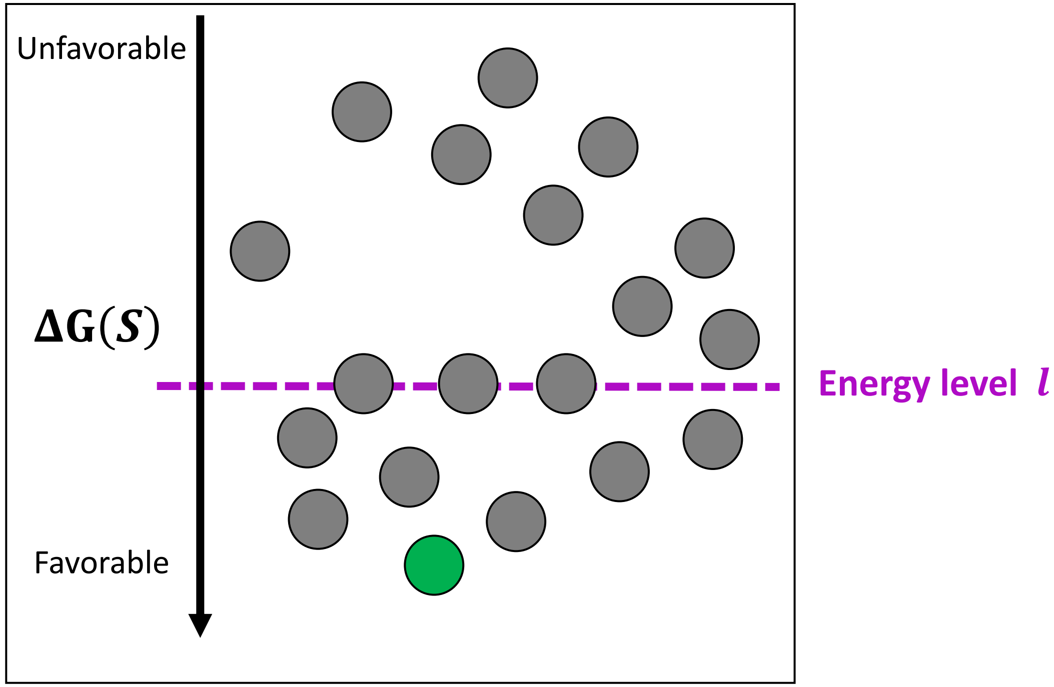
$R(S)$: degree of rotational symmetry of S



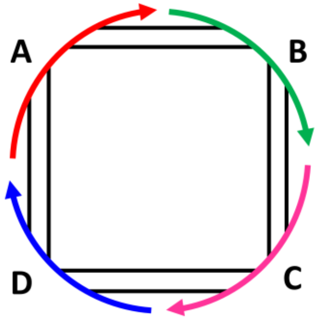
$$\Delta G(S) = \sum_{l} \Delta G(l) + (c-1)\,\Delta G^{\text{assoc}}$$

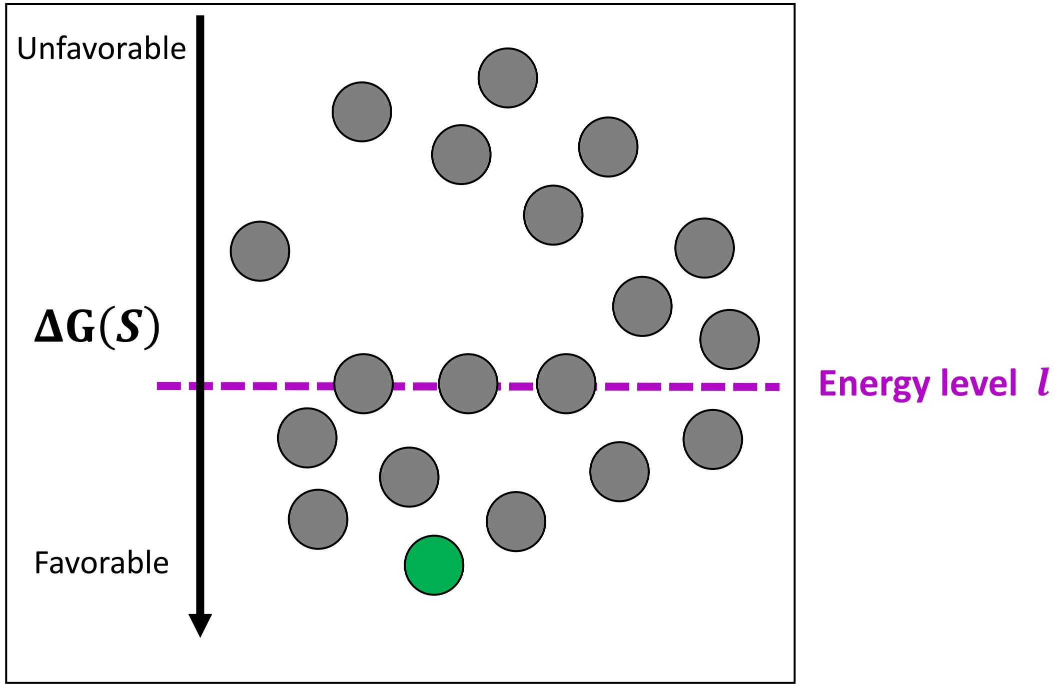


$$\Delta G(S) = \sum_{l} \Delta G(l) + (c-1)\,\Delta G^{\mathrm{assoc}}$$

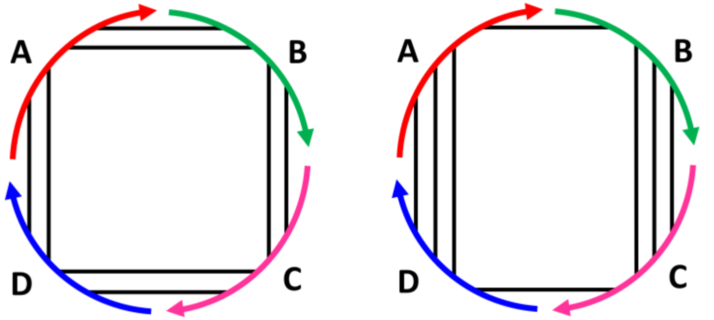


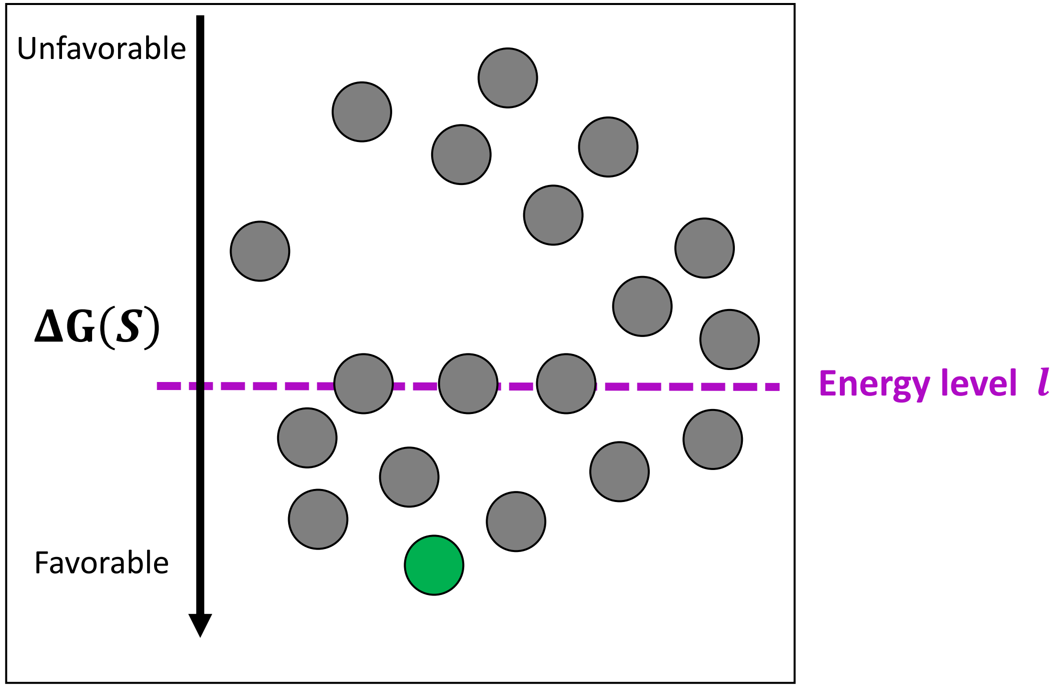
$$\Delta G(S) = \sum_{l} \Delta G(l) + (c-1)\,\Delta G^{\text{assoc}}$$



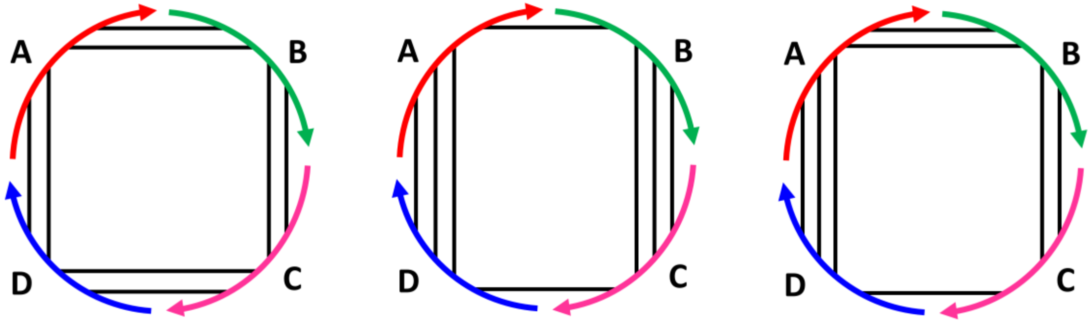


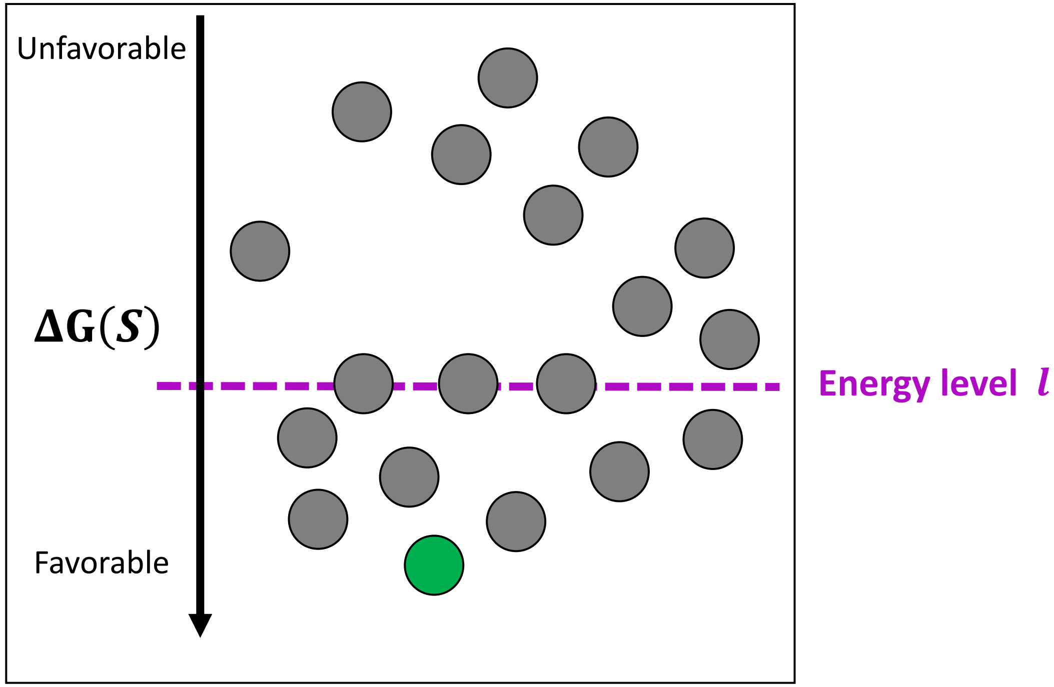
$$\Delta G(S) = \sum_{l} \Delta G(l) + (c-1)\,\Delta G^{\text{assoc}}$$



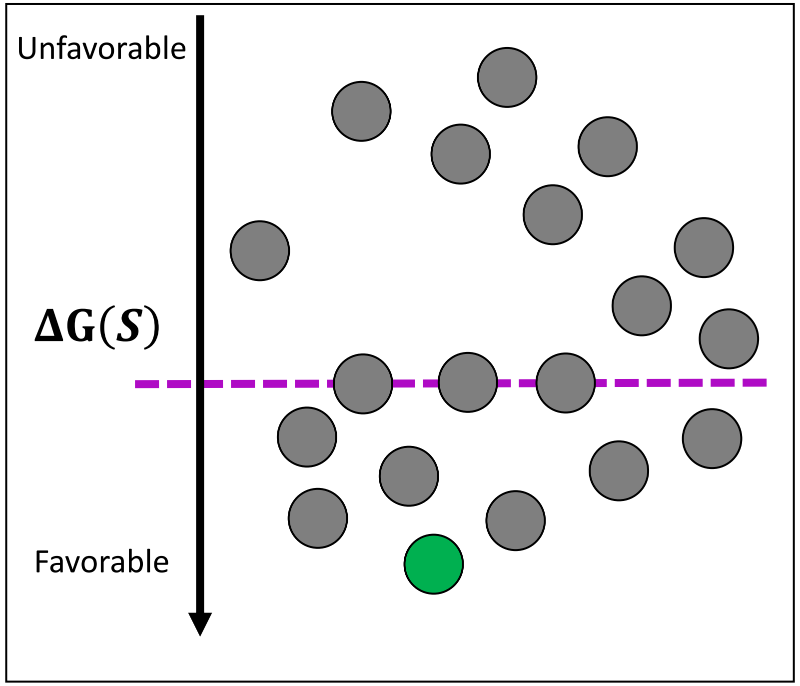


$$\Delta G(S) = \sum_{l} \Delta G(l) + (c-1)\,\Delta G^{\text{assoc}}$$



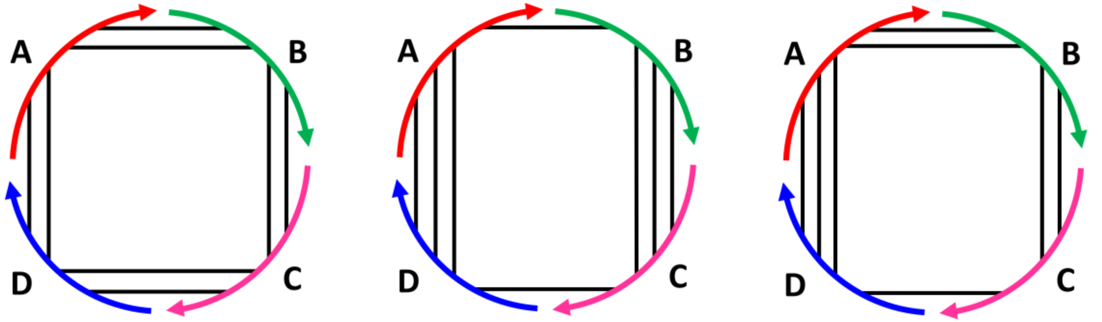


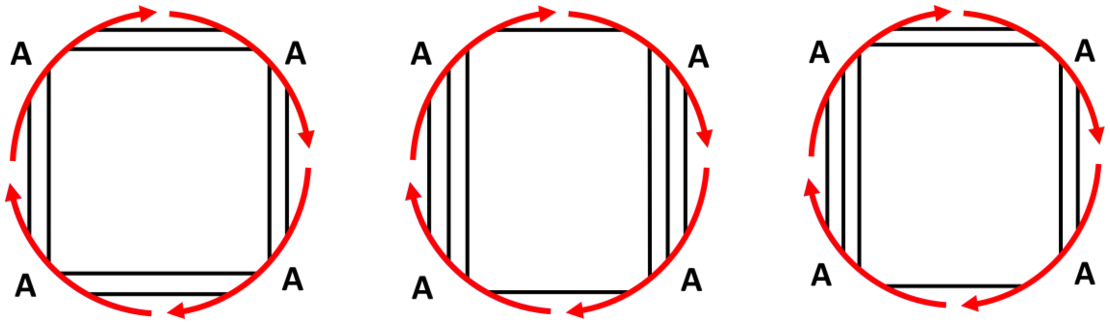
$$\Delta G(S) = \sum_{l} \Delta G(l) + (c-1)\, \Delta G^{\text{assoc}}$$

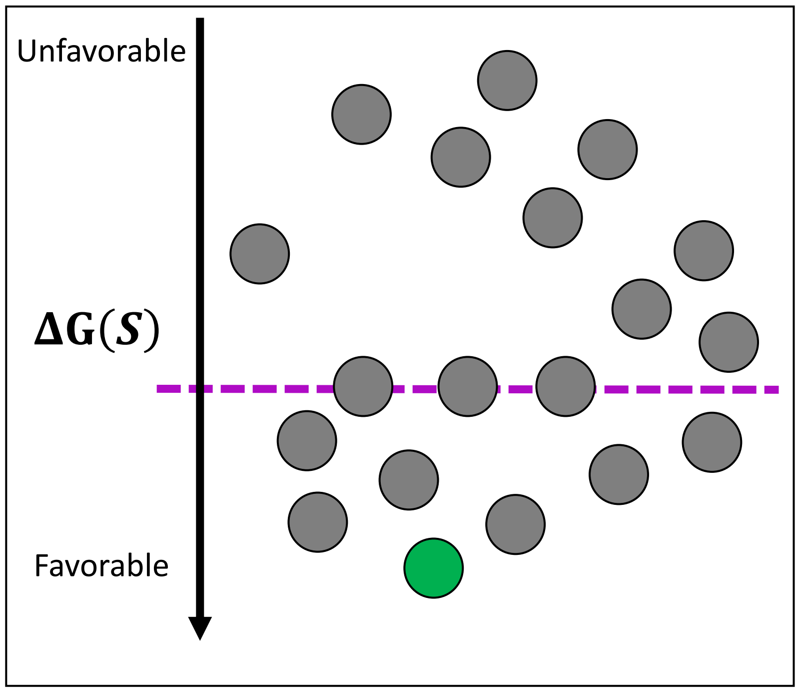


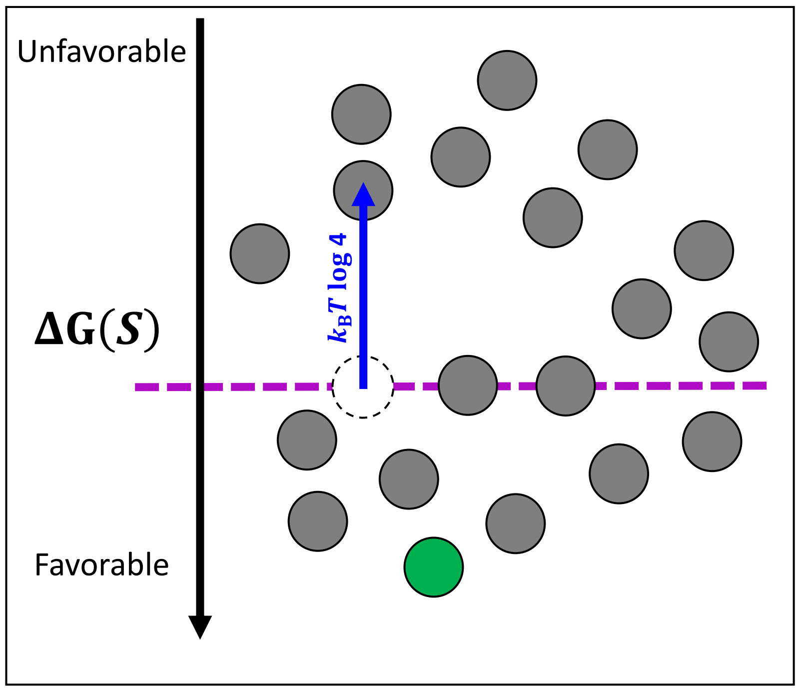
$$\Delta G(S) = \sum_{l} \Delta G(l) + (c-1)\Delta G^{\text{assoc}} + \boldsymbol{k_B T \log R(S)}$$

$R(S)$: degree of rotational symmetry of S





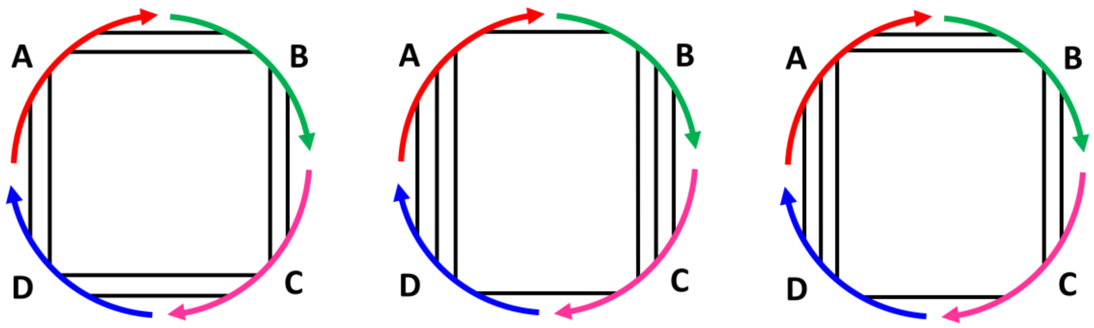


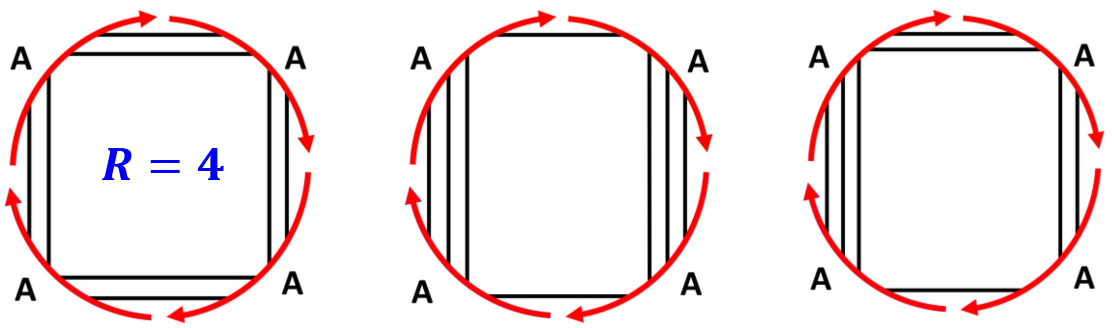


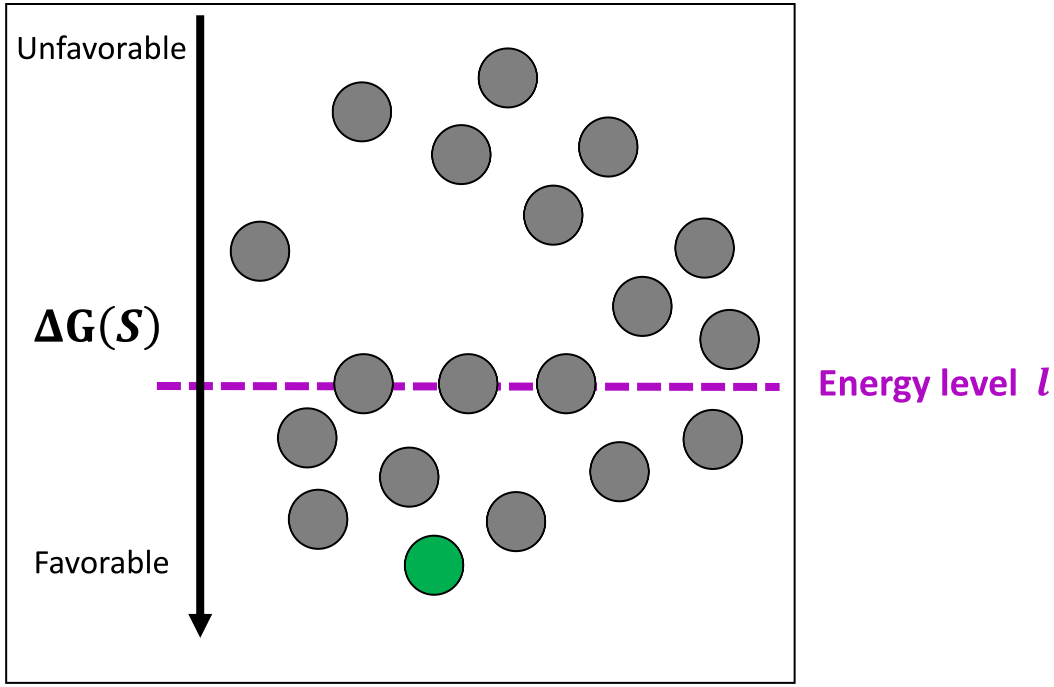
$$\Delta G(S) = \sum_{l} \Delta G(l) + (c-1)\,\Delta G^{\mathrm{assoc}}$$

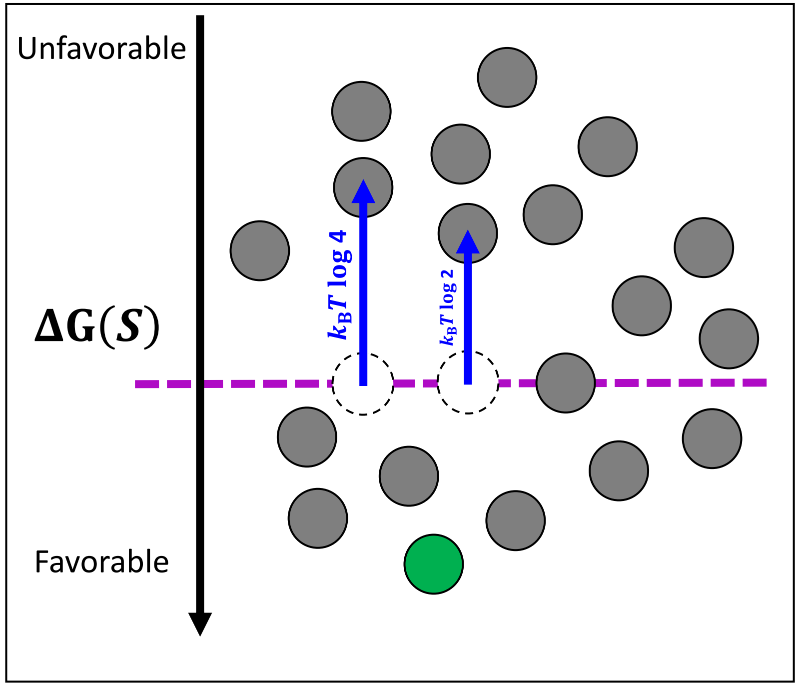
$$\Delta G(S) = \sum_{l} \Delta G(l) + (c-1)\Delta G^{\mathrm{assoc}} + \boldsymbol{k_B T \log R(S)}$$

$\boldsymbol{R(S)}$: **degree of rotational symmetry of** $\boldsymbol{S}$





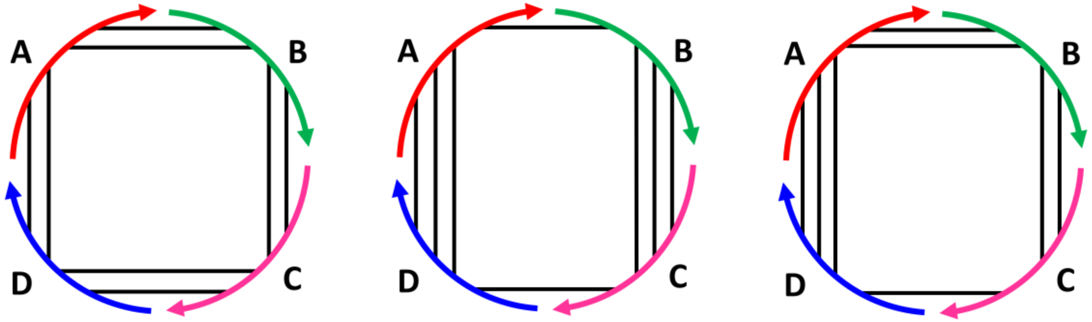


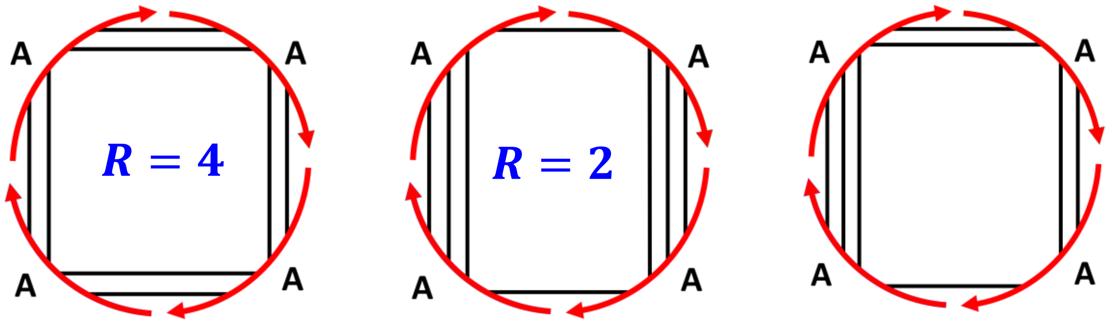


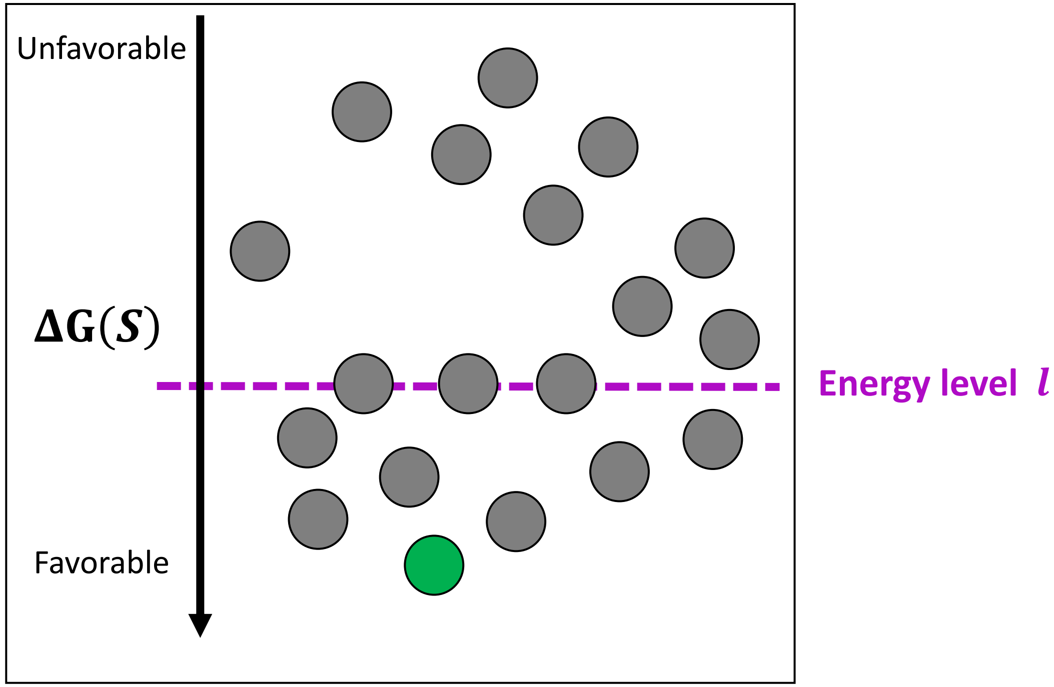
$$\Delta G(S) = \sum_{l} \Delta G(l) + (c-1)\,\Delta G^{\text{assoc}}$$

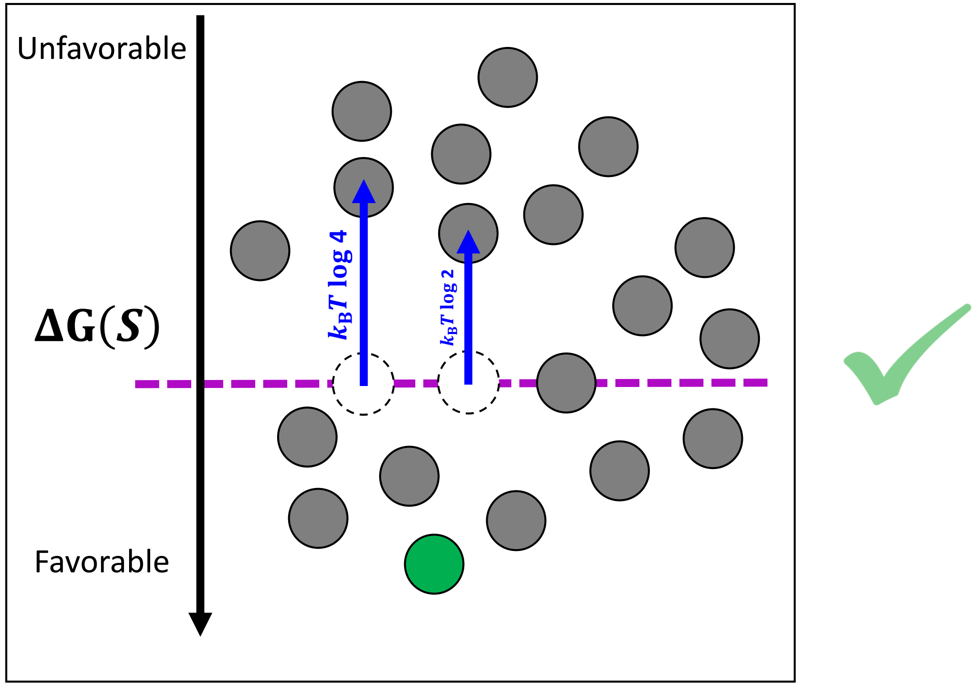
$$\Delta G(S) = \sum_{l} \Delta G(l) + (c-1)\Delta G^{\text{assoc}} + \boldsymbol{k_B T \log R(S)}$$

$\boldsymbol{R(S)}$: **degree of rotational symmetry of** $\boldsymbol{S}$





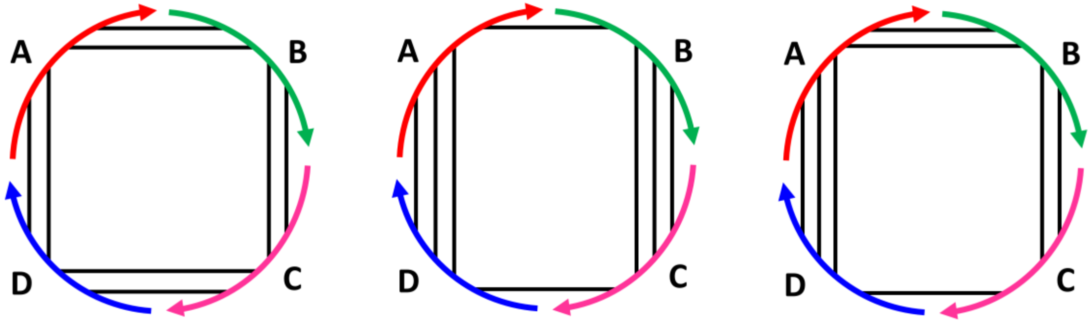


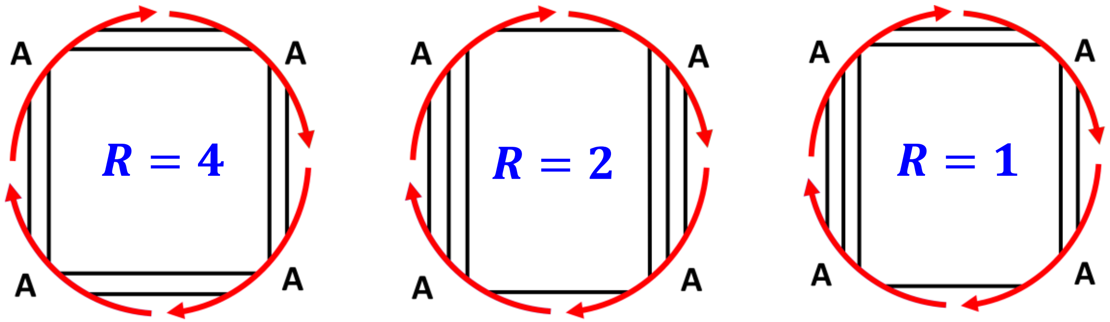


$$\Delta G(S) = \sum_{l} \Delta G(l) + (c-1)\,\Delta G^{\text{assoc}}$$

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

$\boldsymbol{R(S)}$**: degree of rotational symmetry of** $\boldsymbol{S}$





Computational complexity of Minimum Free Energy

Input Type	MFE
Single strand	$O(N^3)$ Nussinov et al. 1978
$c = O(1)$ **unique** strands	$O(N^3(c-1)!)$ Dirks et al. 2007
$c = O(1)$ **allowing repeated** strands	? Open problem for ≈ **20** years
Multiple strands, unbounded, $O(N)$ strands	NP − Complete Condon, Hajiaghayi, Thachuk 2021

N bases, c strands

Computational complexity of Minimum Free Energy

Input Type	MFE
Single strand	$O(N^3)$ Nussinov et al. 1978
$c = O(1)$ **unique** strands	$O(N^3(c-1)!)$ Dirks et al. 2007
$c = O(1)$ **allowing repeated** strands	? Open problem for ≈ **20** years
Multiple strands, unbounded, $O(N)$ strands	$\mathrm{NP-Complete}$ Condon, Hajiaghayi, Thachuk 2021

} **Dynamic programming algorithms**

N bases, c strands

Computational complexity of Minimum Free Energy

Input Type	MFE
Single strand	$O(N^3)$ Nussinov et al. 1978
$c = O(1)$ **unique** strands	$O(N^3(c-1)!)$ Dirks et al. 2007
$c = O(1)$ **allowing repeated** strands	? Open problem for ≈ **20** years
Multiple strands, unbounded, $O(N)$ strands	$\text{NP} - \text{Complete}$ Condon, Hajiaghayi, Thachuk 2021

Dynamic programming algorithms

N bases, c strands

Subproblems → **Big problem**

Computational complexity of Minimum Free Energy

Input Type	MFE
Single strand	$O(N^3)$ Nussinov et al. 1978
$c = O(1)$ **unique** strands	$O(N^3(c-1)!)$ Dirks et al. 2007
$c = O(1)$ **allowing repeated** strands	? Open problem for ≈ **20** years
Multiple strands, unbounded, $O(N)$ strands	$\mathrm{NP-Complete}$ Condon, Hajiaghayi, Thachuk 2021

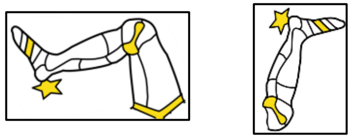
} **Dynamic programming algorithms**

N bases, c strands

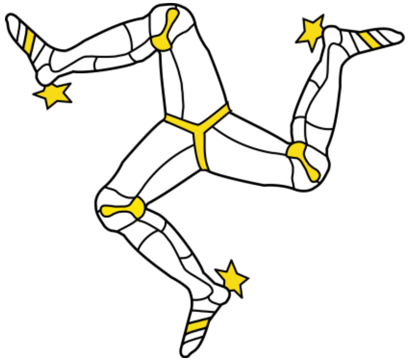
Subproblems → **Big problem**

↓

Local point of view



?



Global property

Current state of art

Input Type	MFE
Single strand	$O(N^3)$
$c = O(1)$ **unique** strands	$O(N^3(c-1)!)$
$c = O(1)$ **allowing repeated** strands	?
Multiple strands, unbounded, $O(N)$ strands	$\mathrm{NP-Complete}$

N bases, c strands

Current state of art

Input Type	MFE
Single strand	$O(N^3)$
$c = O(1)$ unique strands	$O(N^3(c-1)!)$
$c = O(1)$ allowing repeated strands	?
Multiple strands, unbounded, $O(N)$ strands	$\mathrm{NP-Complete}$

N bases, c strands

Goal

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

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

Current state of art

Input Type	MFE
Single strand	$O(N^3)$
$c = O(1)$ unique strands	$O(N^3(c-1)!)$
$c = O(1)$ allowing repeated strands	?
Multiple strands, unbounded, $O(N)$ strands	NP − Complete

NUPACK

ViennaRNA

N bases, c strands

Goal

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

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

Current state of art

Input Type	MFE
Single strand	$O(N^3)$
$c = O(1)$ **unique** strands	$O(N^3(c-1)!)$
$c = O(1)$ **allowing repeated** strands	?
Multiple strands, unbounded, $O(N)$ strands	$\mathrm{NP-Complete}$

NUPACK

ViennaRNA

N bases, c strands

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

Symmetry naive energy function

$$\mathbf{snMFE} = \min_{S \in \Omega} \overline{\Delta G}(S)$$

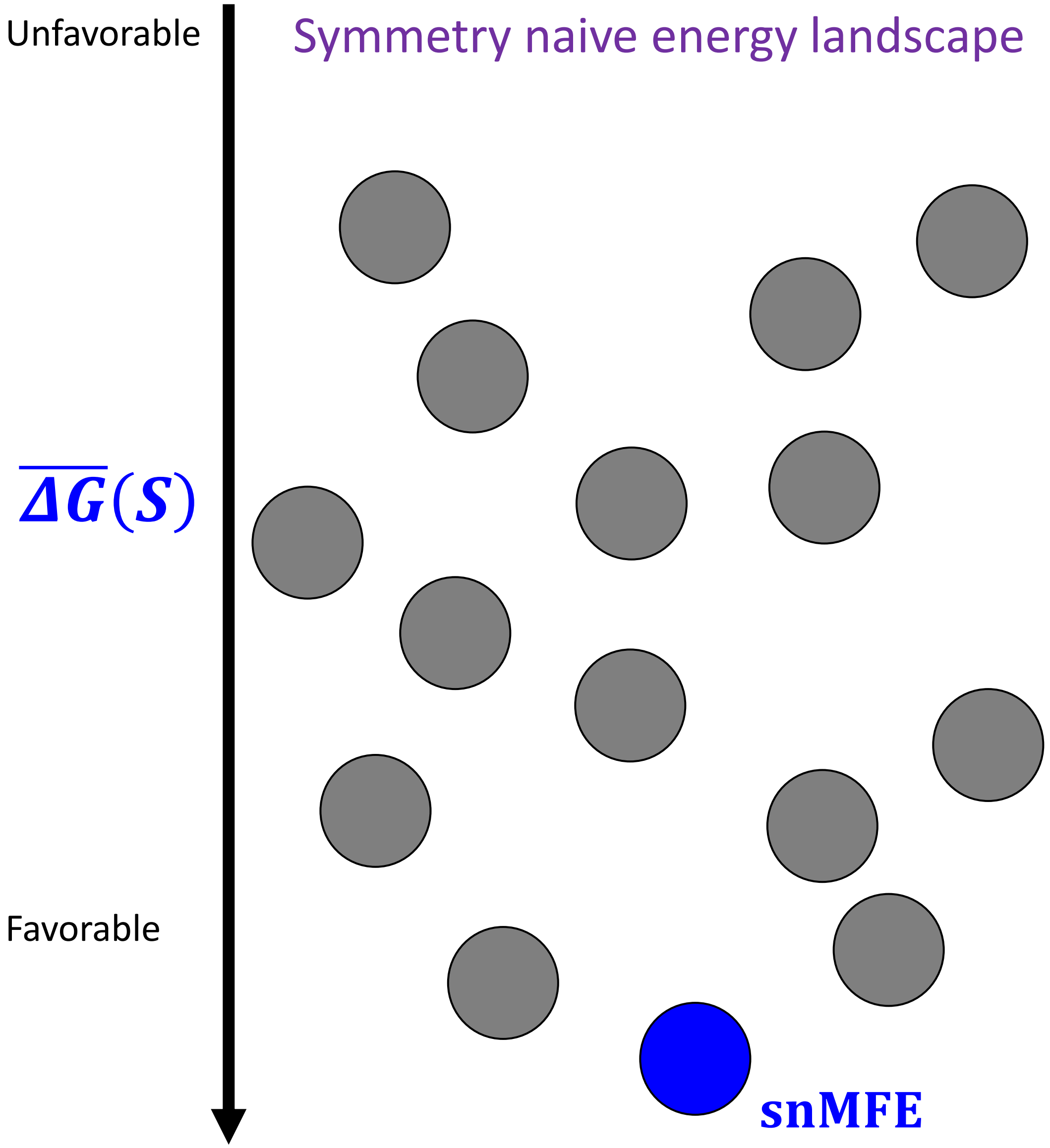
Symmetry naive MFE

Goal

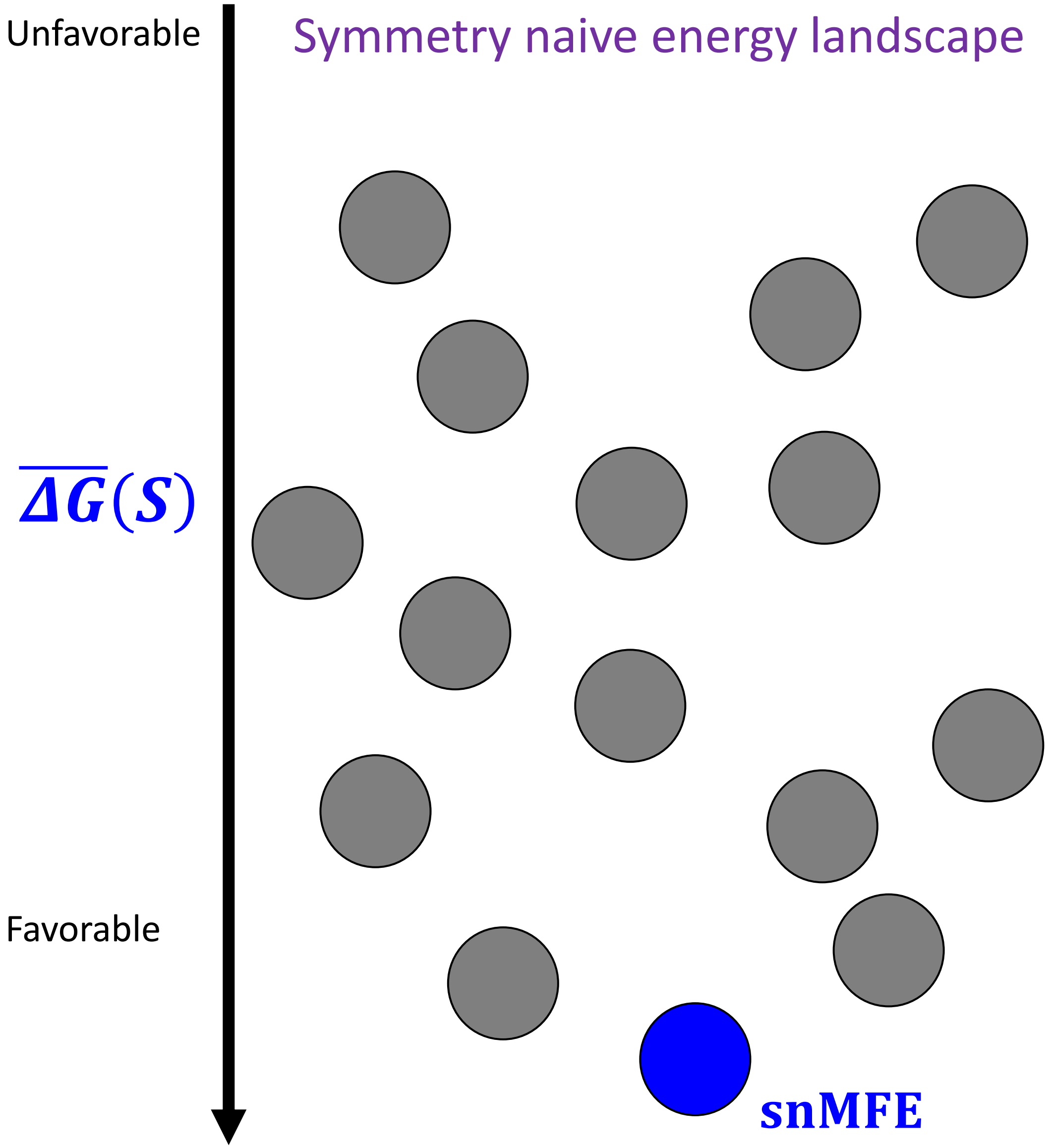
$$\Delta G(S) = \sum_{l} \Delta G(l) + (c-1)\Delta G^{\mathrm{assoc}} + k_B T \log R(S)$$

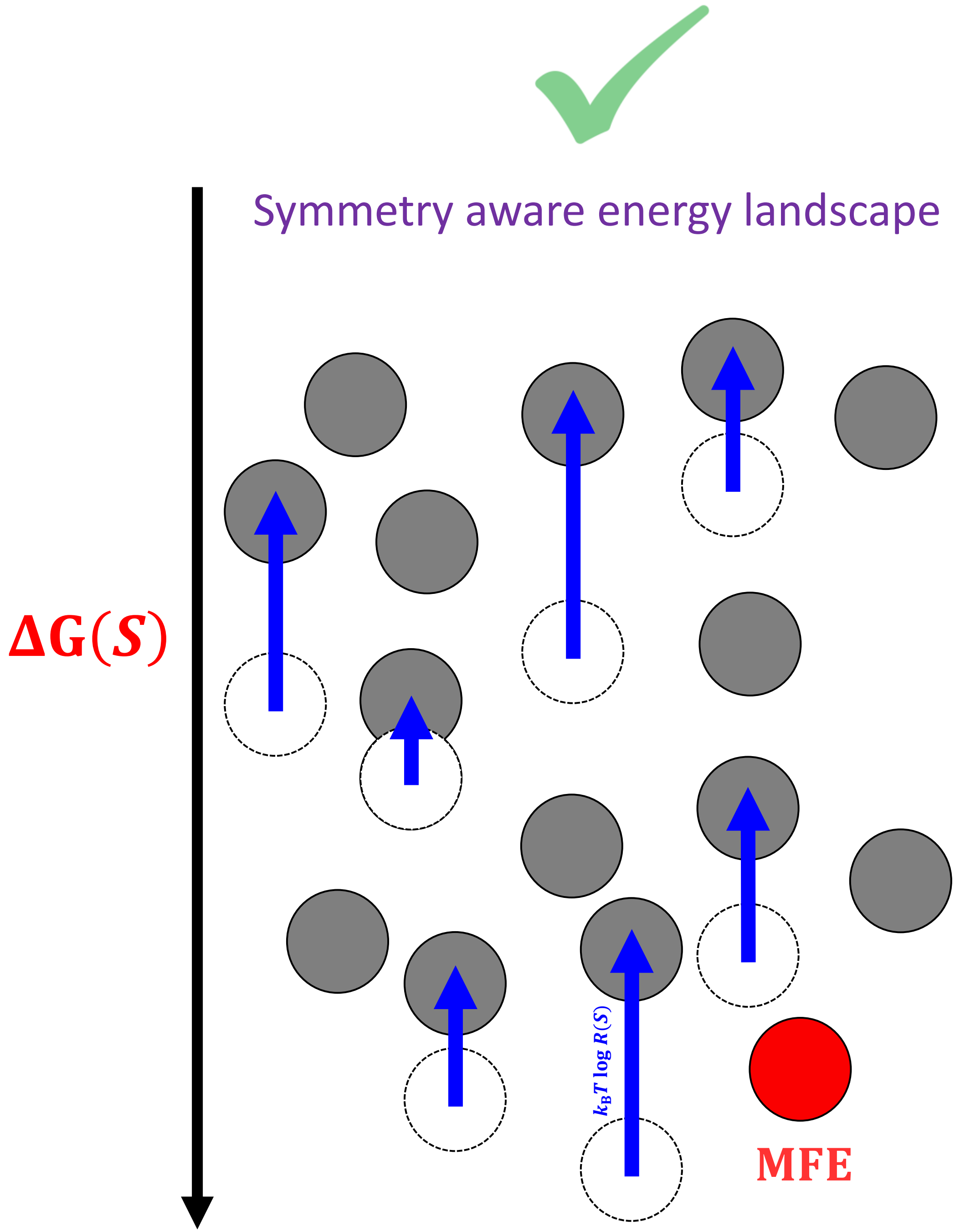
$$\mathbf{MFE} = \min_{S \in \Omega} \Delta G(S)$$

Current state of art

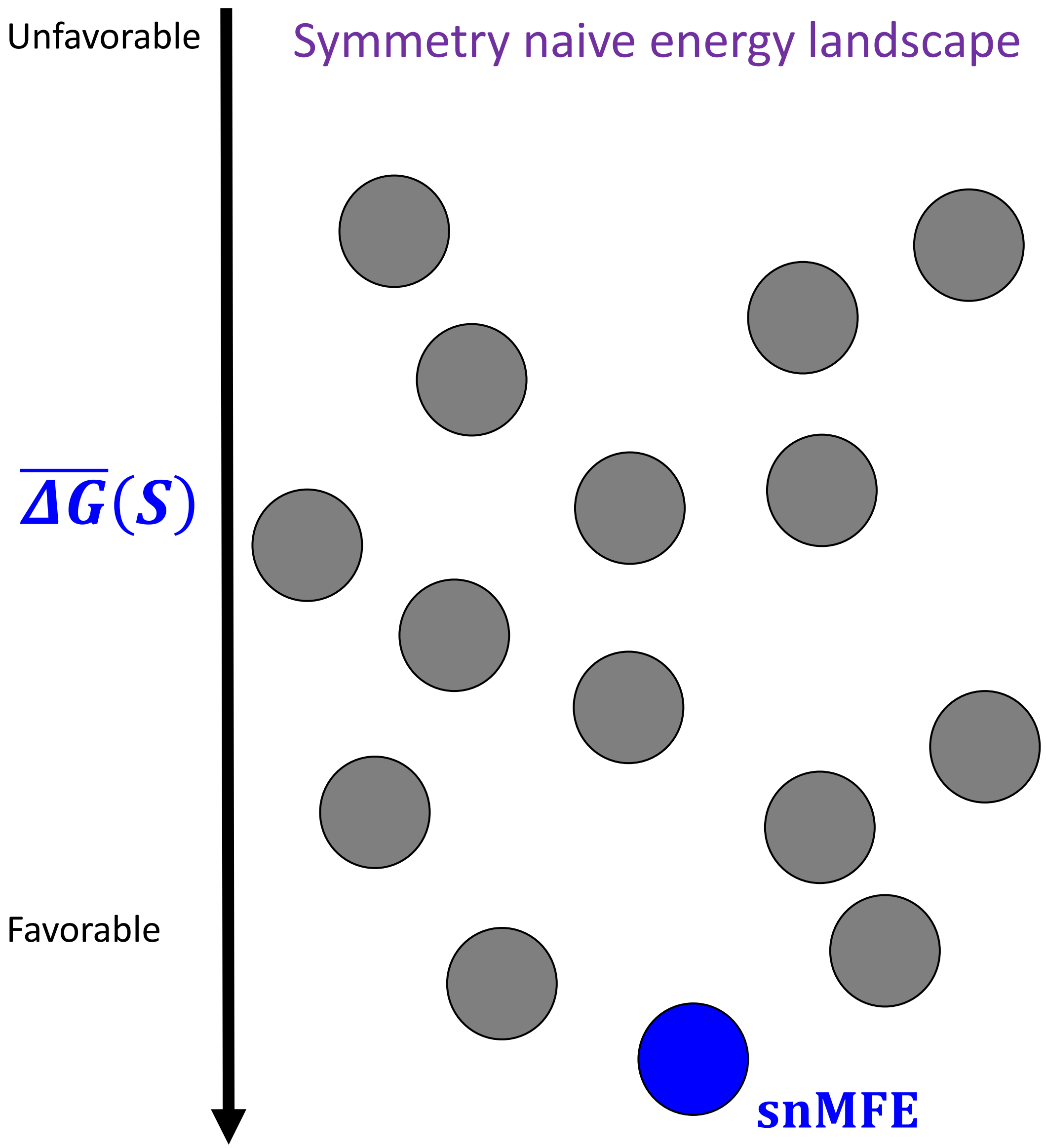


Current state of art

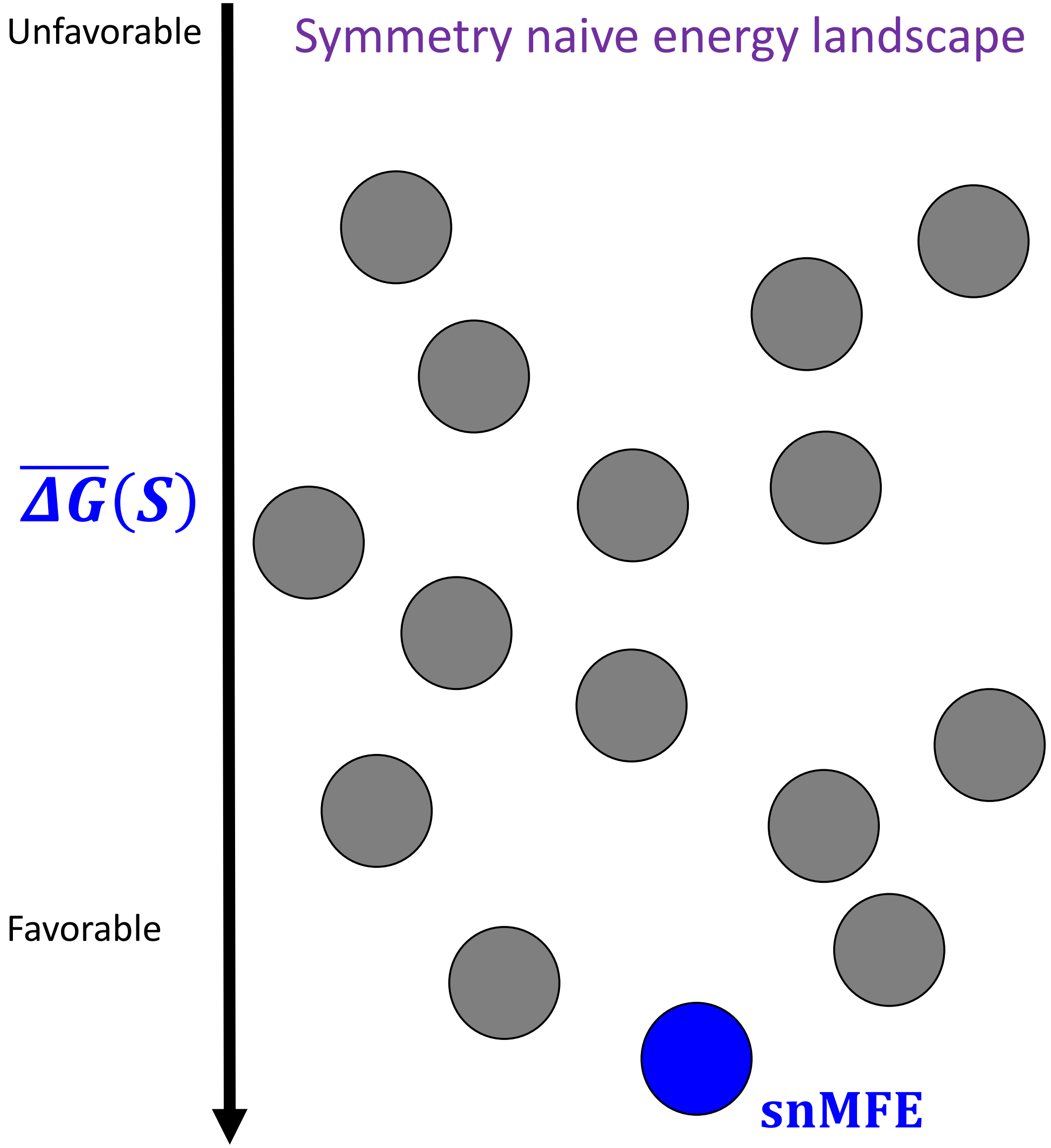


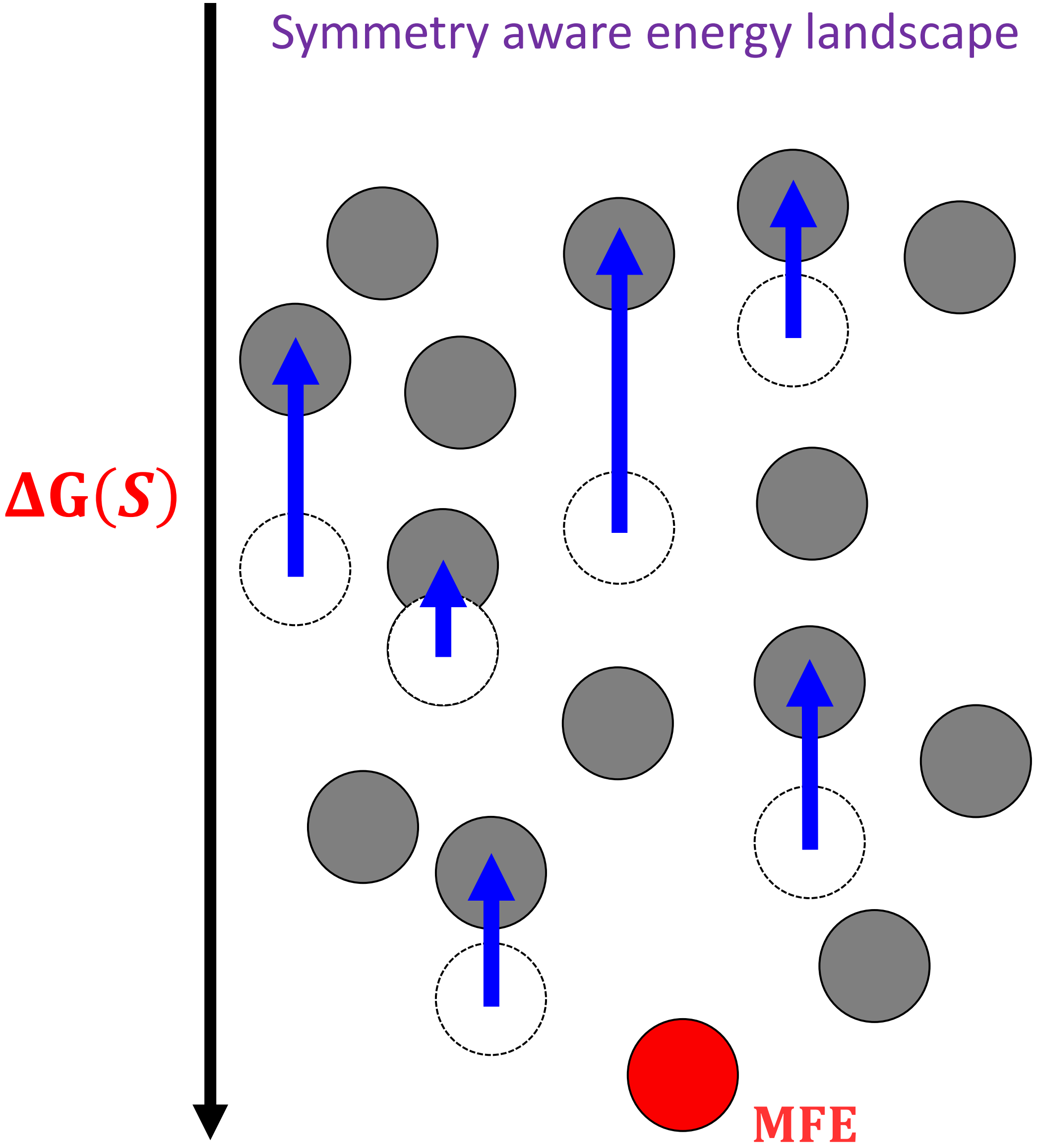


If snMFE structure is Asymmetric $(R = 1)$

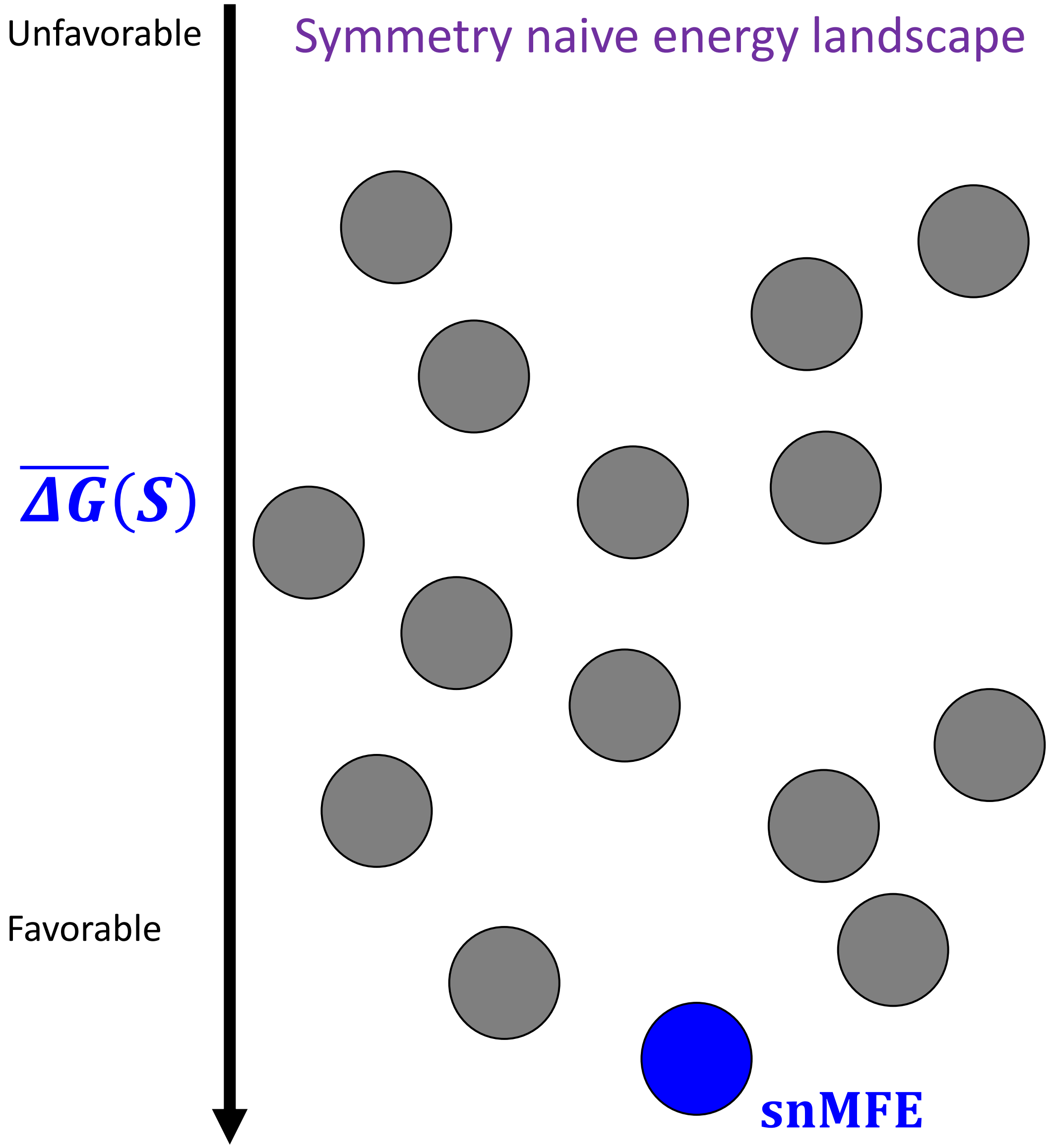


If snMFE structure is Asymmetric $(R = 1)$

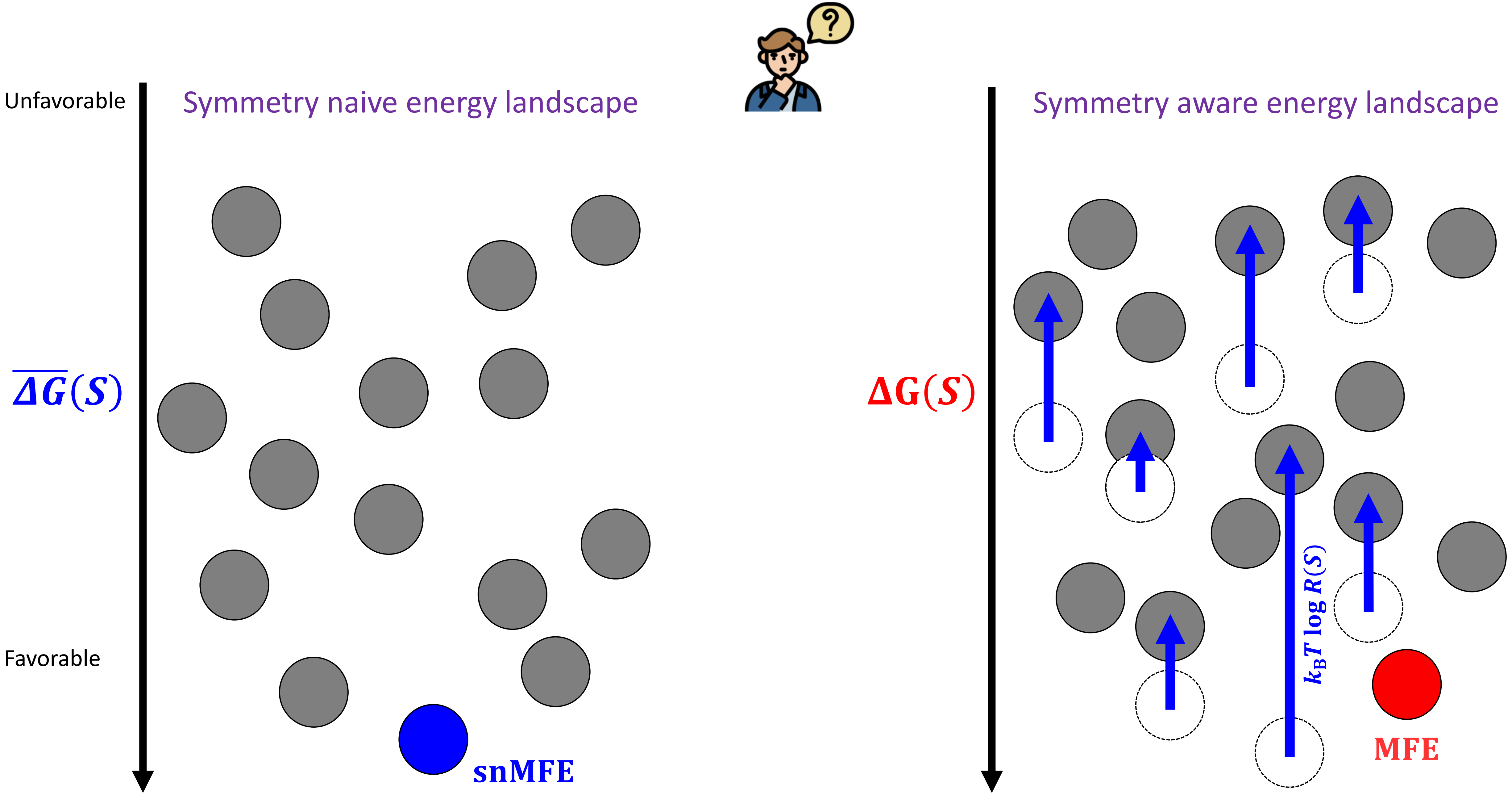




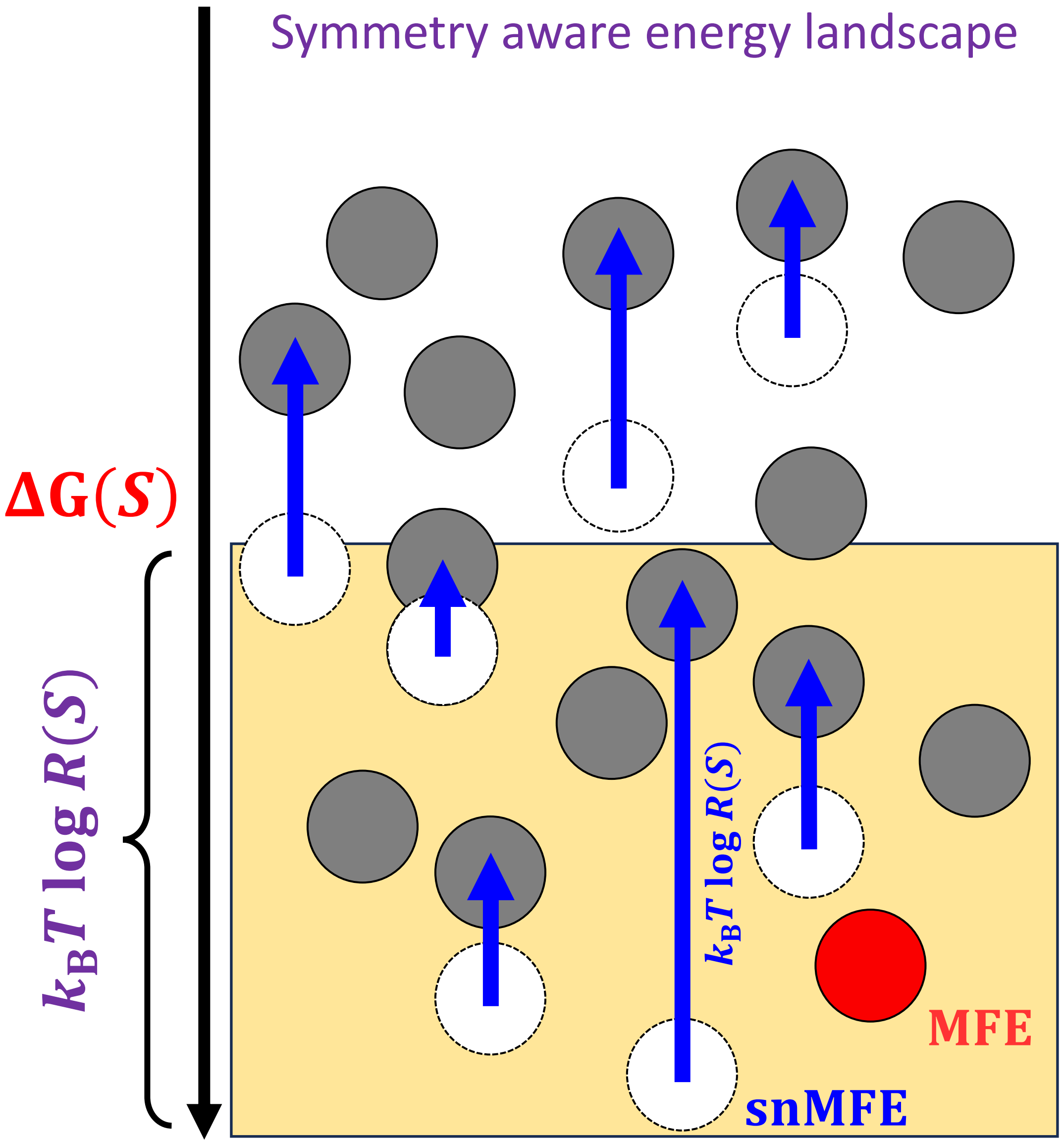
If snMFE structure is Symmetric $(R > 1)$



If snMFE structure is Symmetric $(R > 1)$

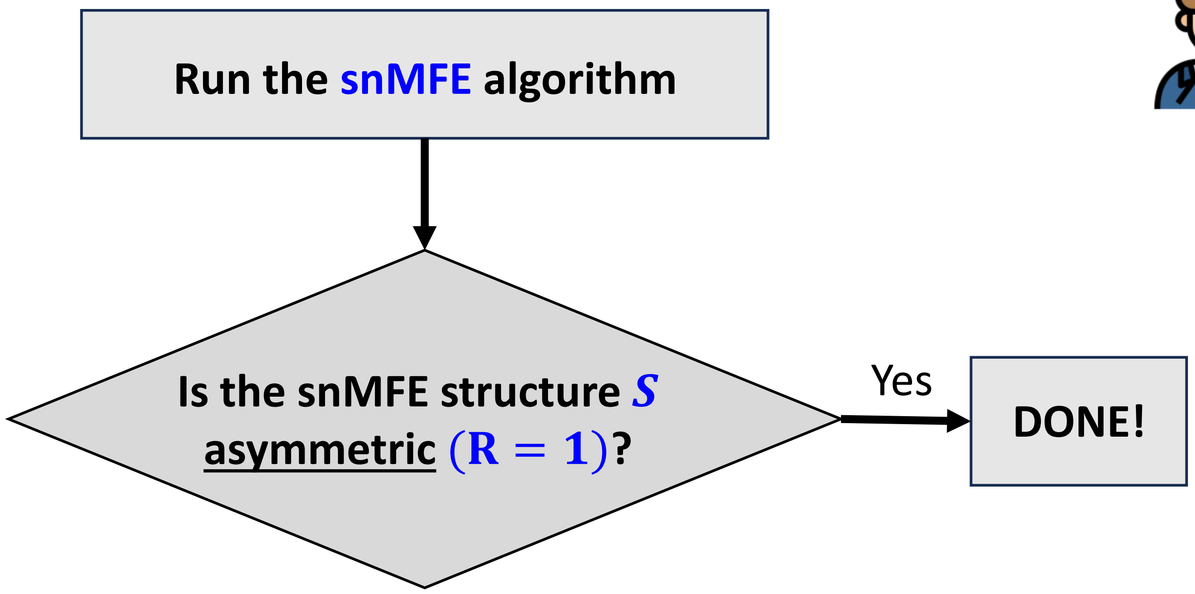


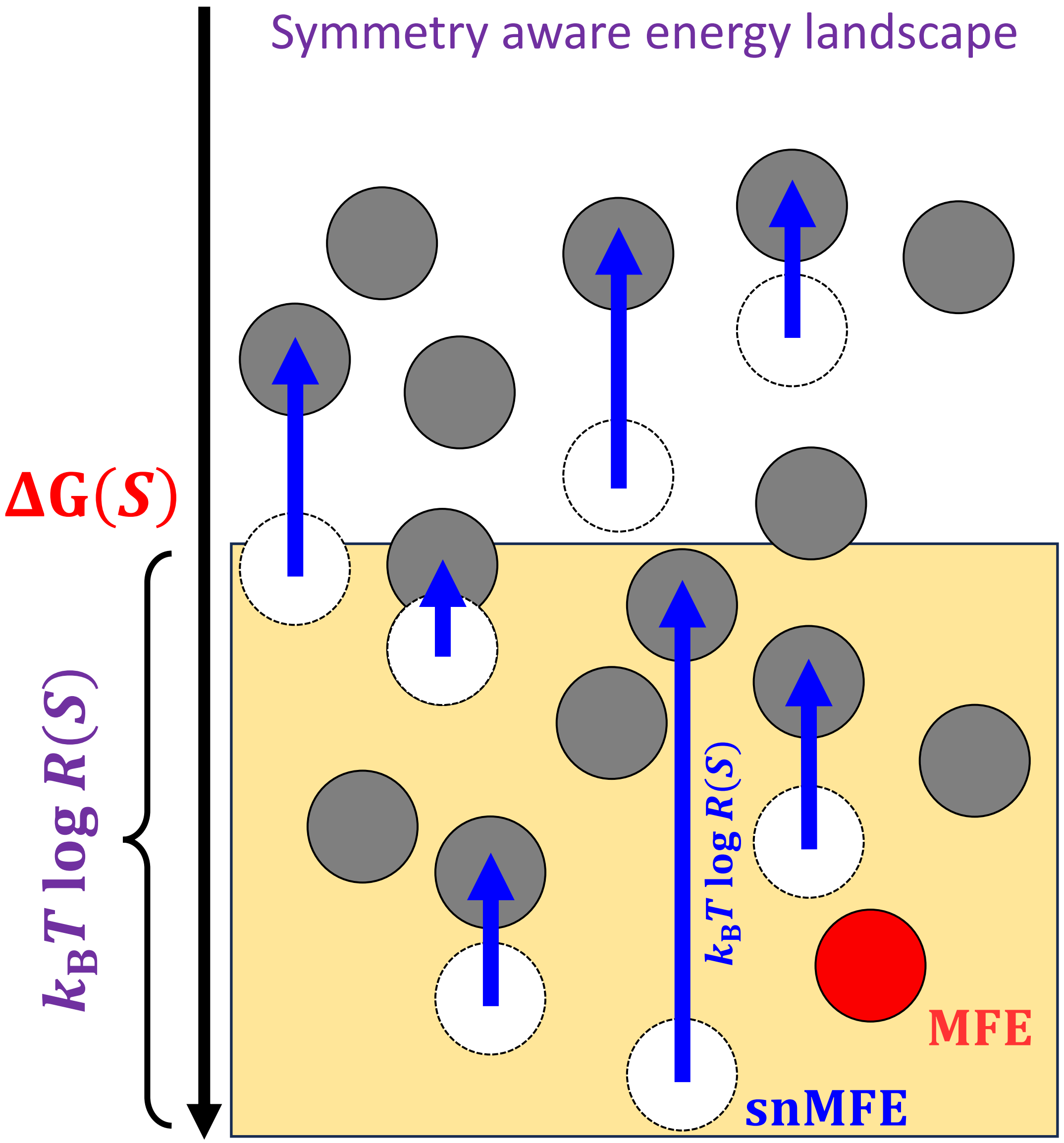
If snMFE structure is Symmetric $(R > 1)$



If snMFE structure is Symmetric $(R > 1)$

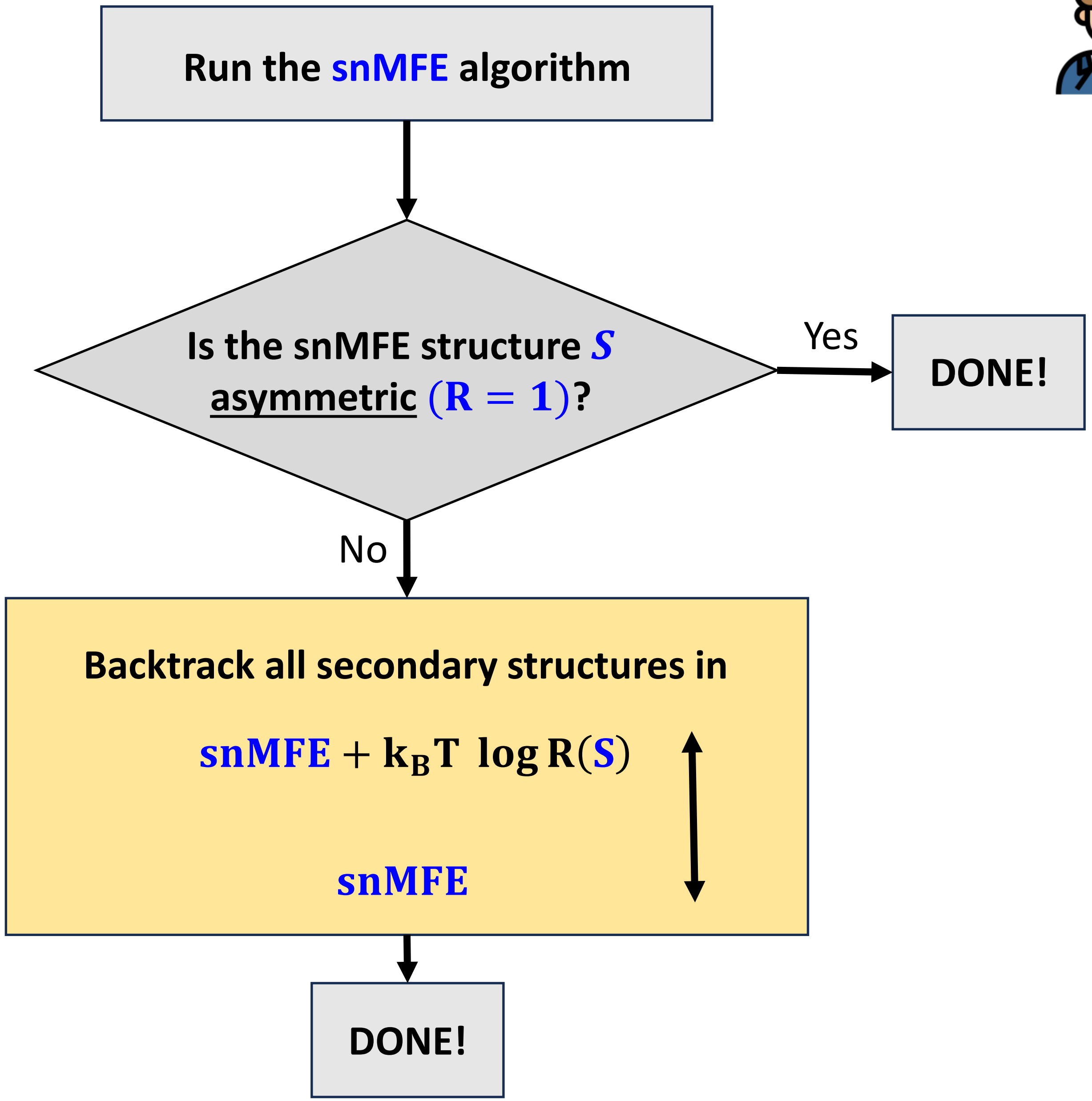
Dirks et al. 2007 approach

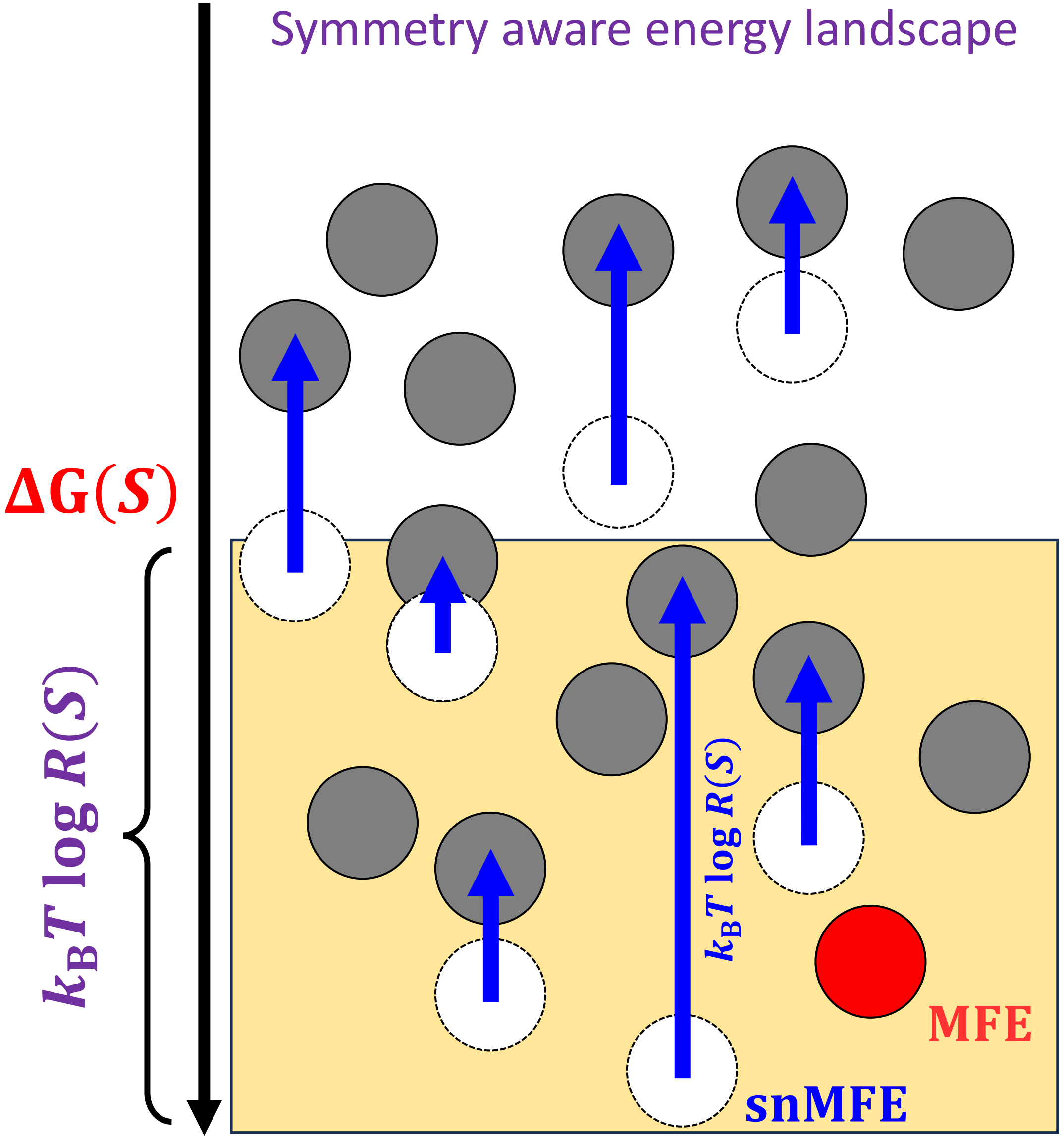




If snMFE structure is Symmetric $(R > 1)$

Dirks et al. 2007 approach





If snMFE structure is Symmetric $(R > 1)$

Dirks et al. 2007 approach

Run the snMFE algorithm

↓

Is the snMFE structure S asymmetric $(\mathrm{R} = 1)$?

Yes → **DONE!**

No ↓

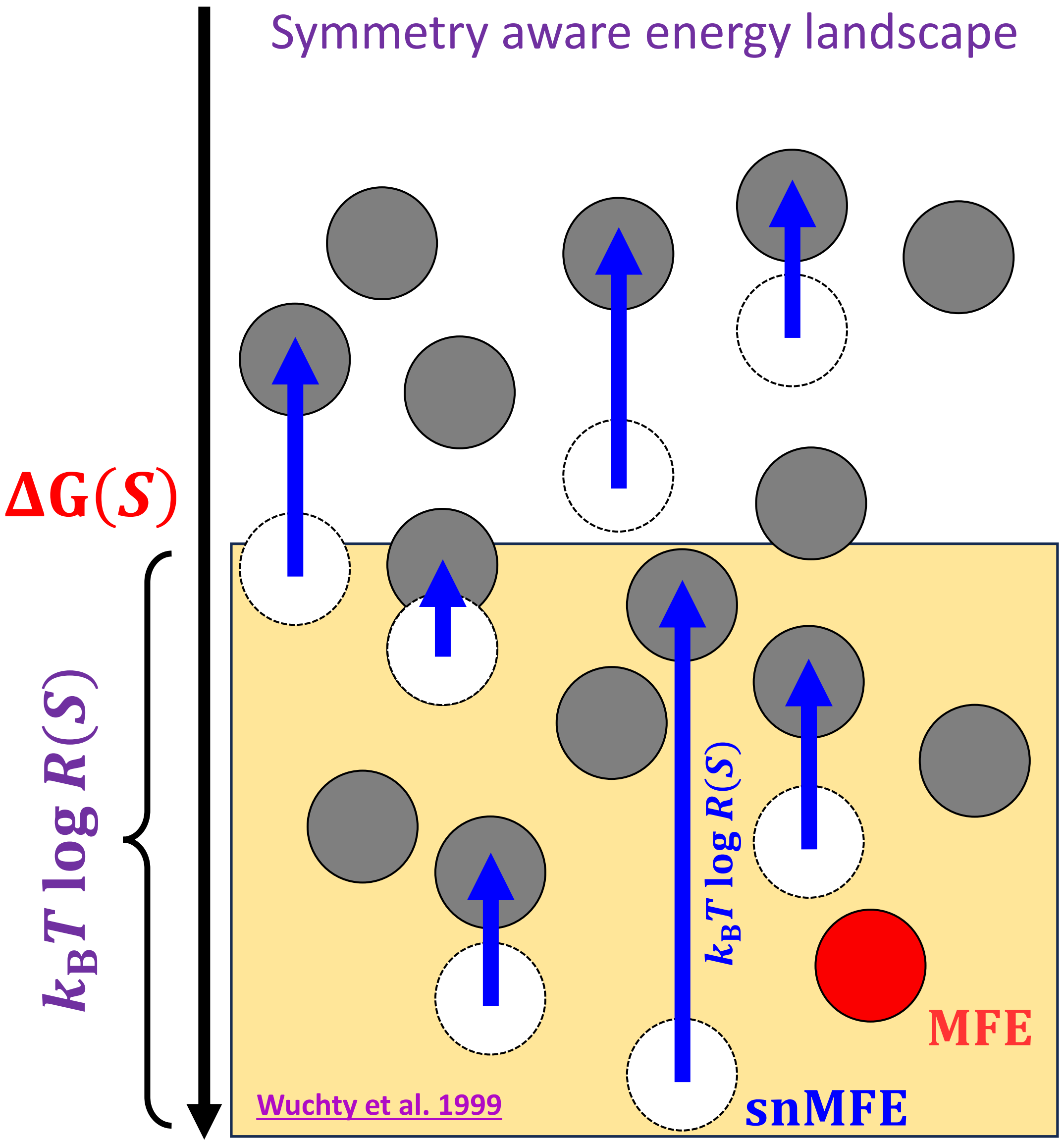
Backtrack all secondary structures in

$$\mathrm{snMFE} + \mathrm{k_B T}\ \log \mathrm{R}(S)$$

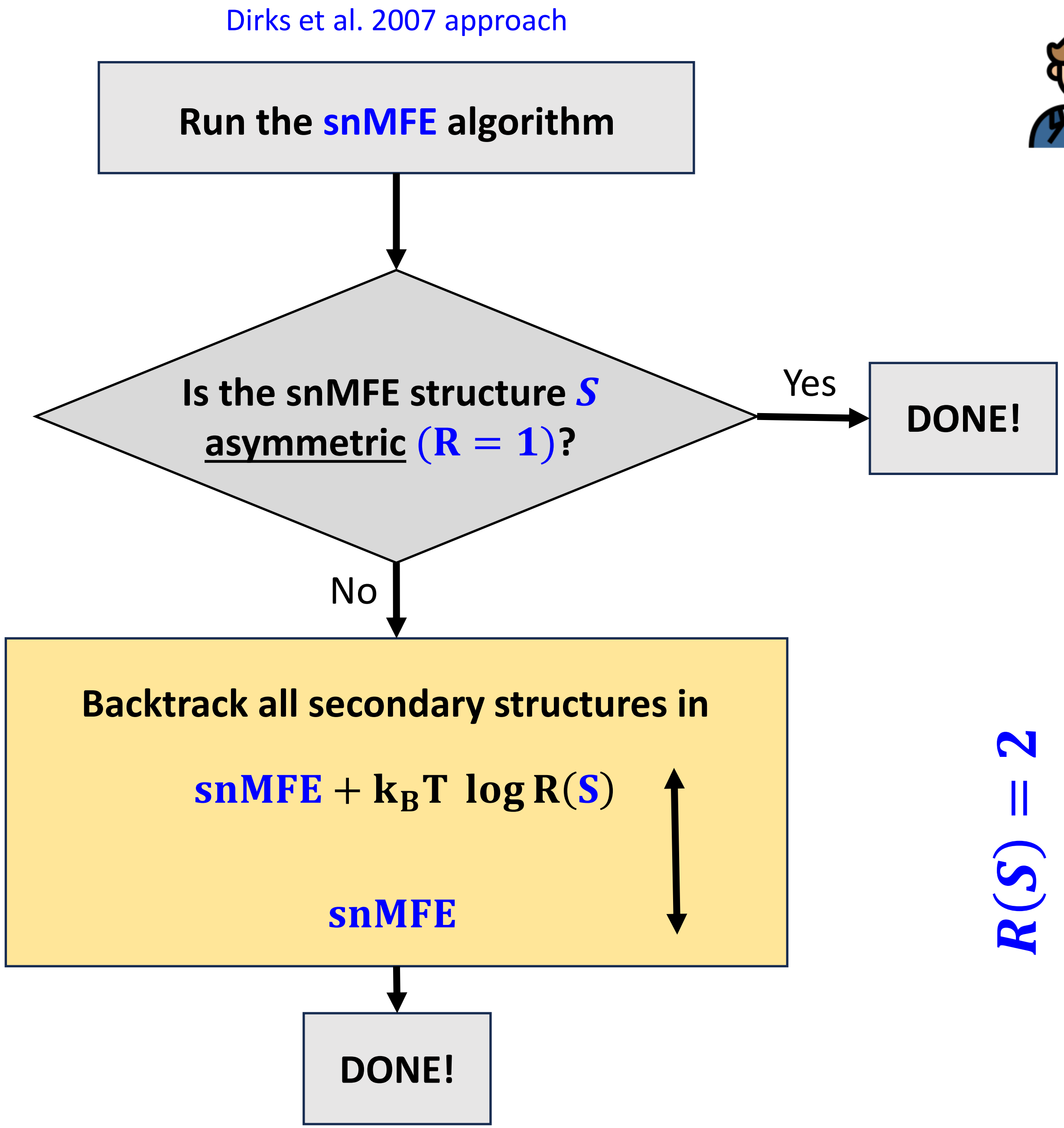
$$\mathrm{snMFE}$$

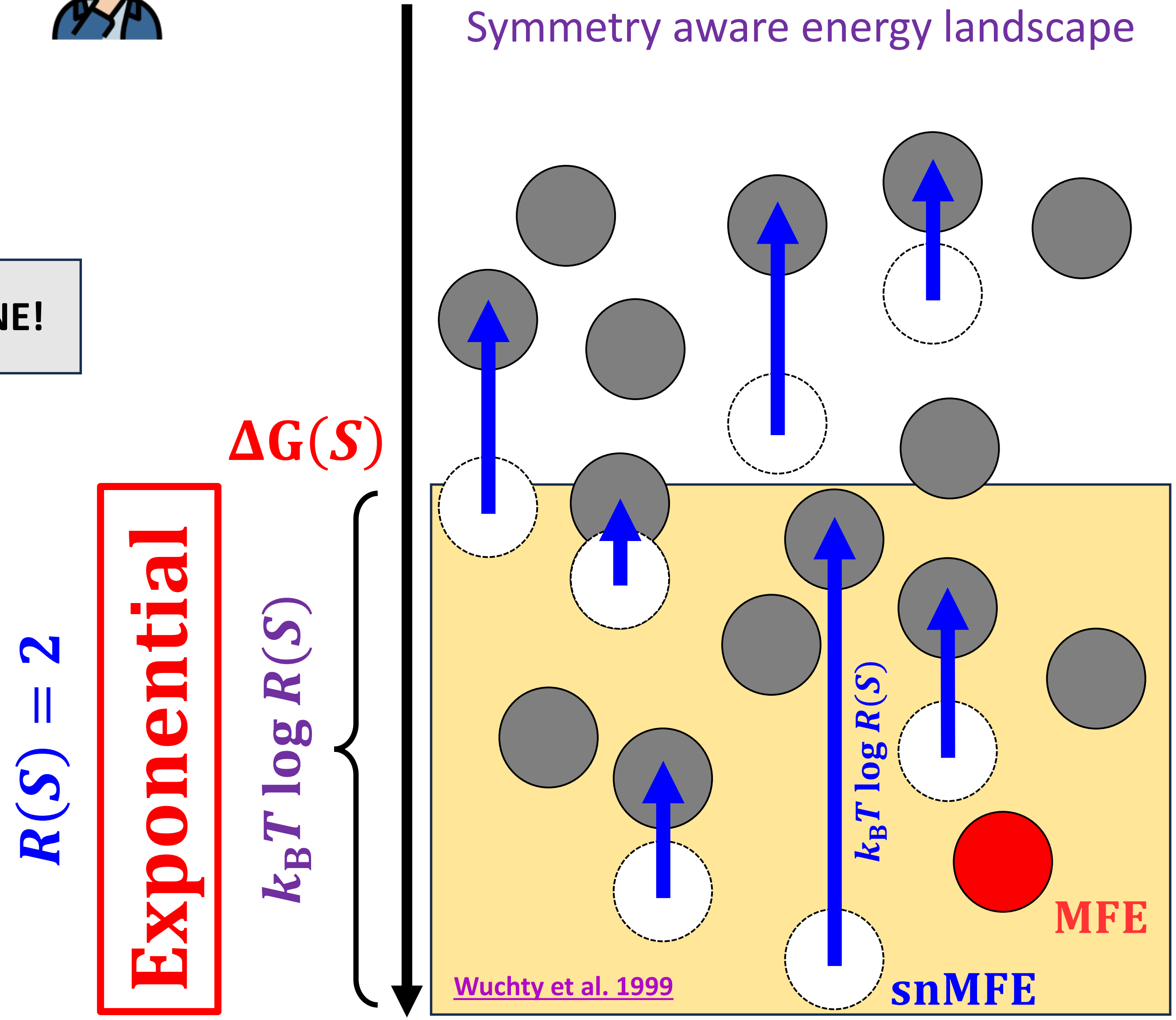
↓

DONE!



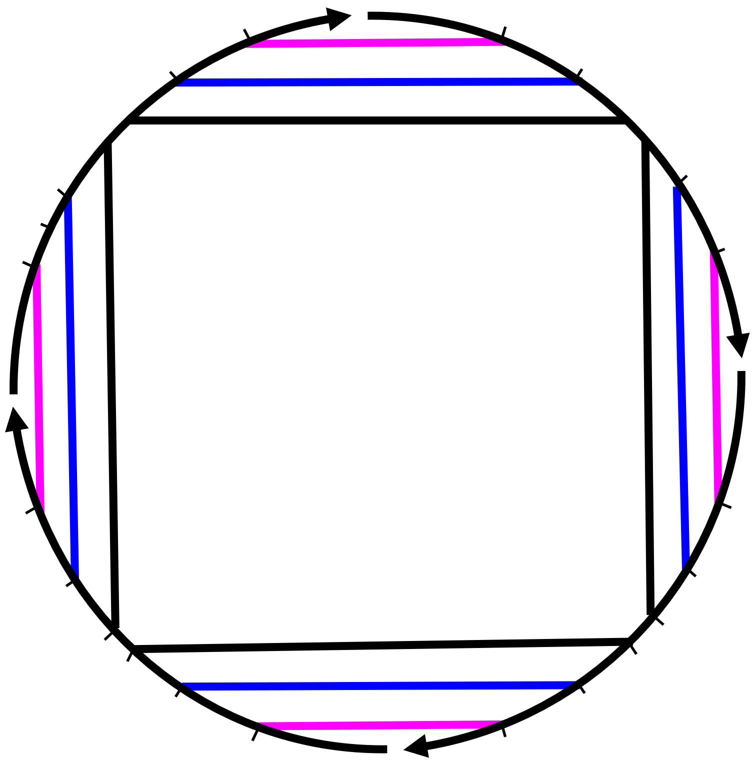
If snMFE structure is Symmetric $(R > 1)$

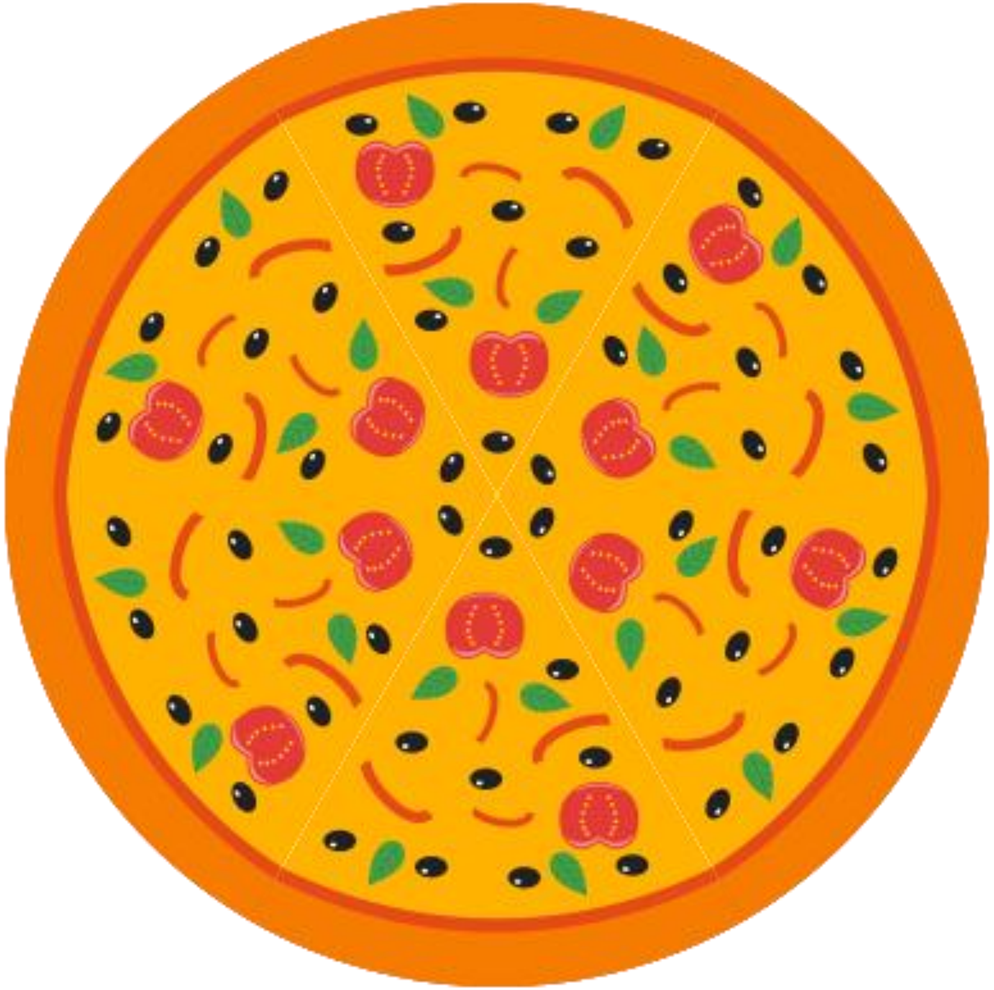




Our Approach

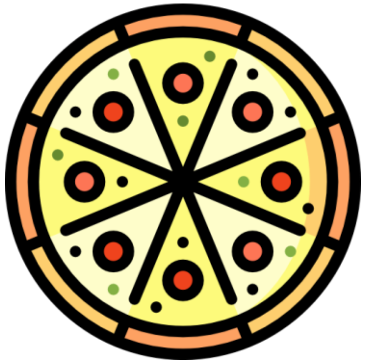
What if we could break symmetry with symmetry?





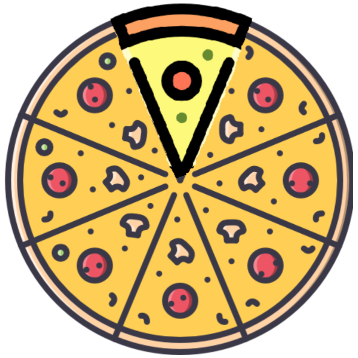
What if we could break symmetry with symmetry?

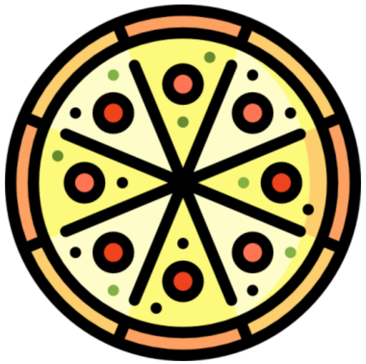
Slicing and swapping strategy



What if we could break symmetry with symmetry?

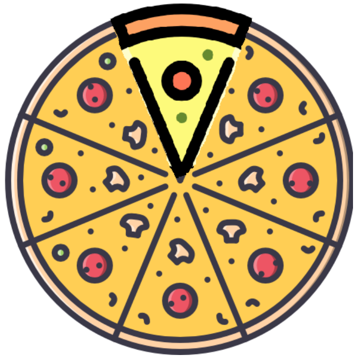
Slicing and swapping strategy

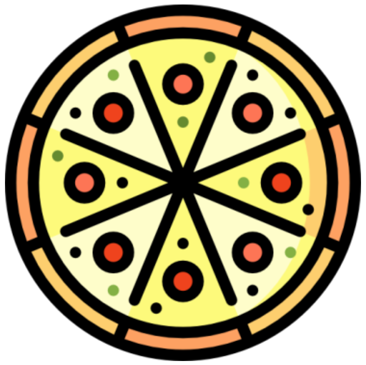




What if we could break symmetry with symmetry?

Slicing and swapping strategy





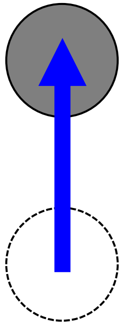
Sharing the same
Admissible pizza cut

Unfavorable

$\overline{\Delta G}(S)$

$\overline{\Delta G}(S_y) \leq \overline{\Delta G}(S_x)$

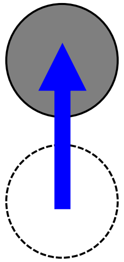
Favorable

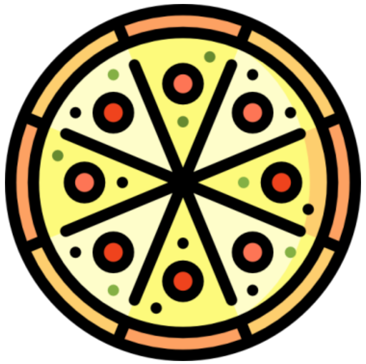


S_x Symmetric

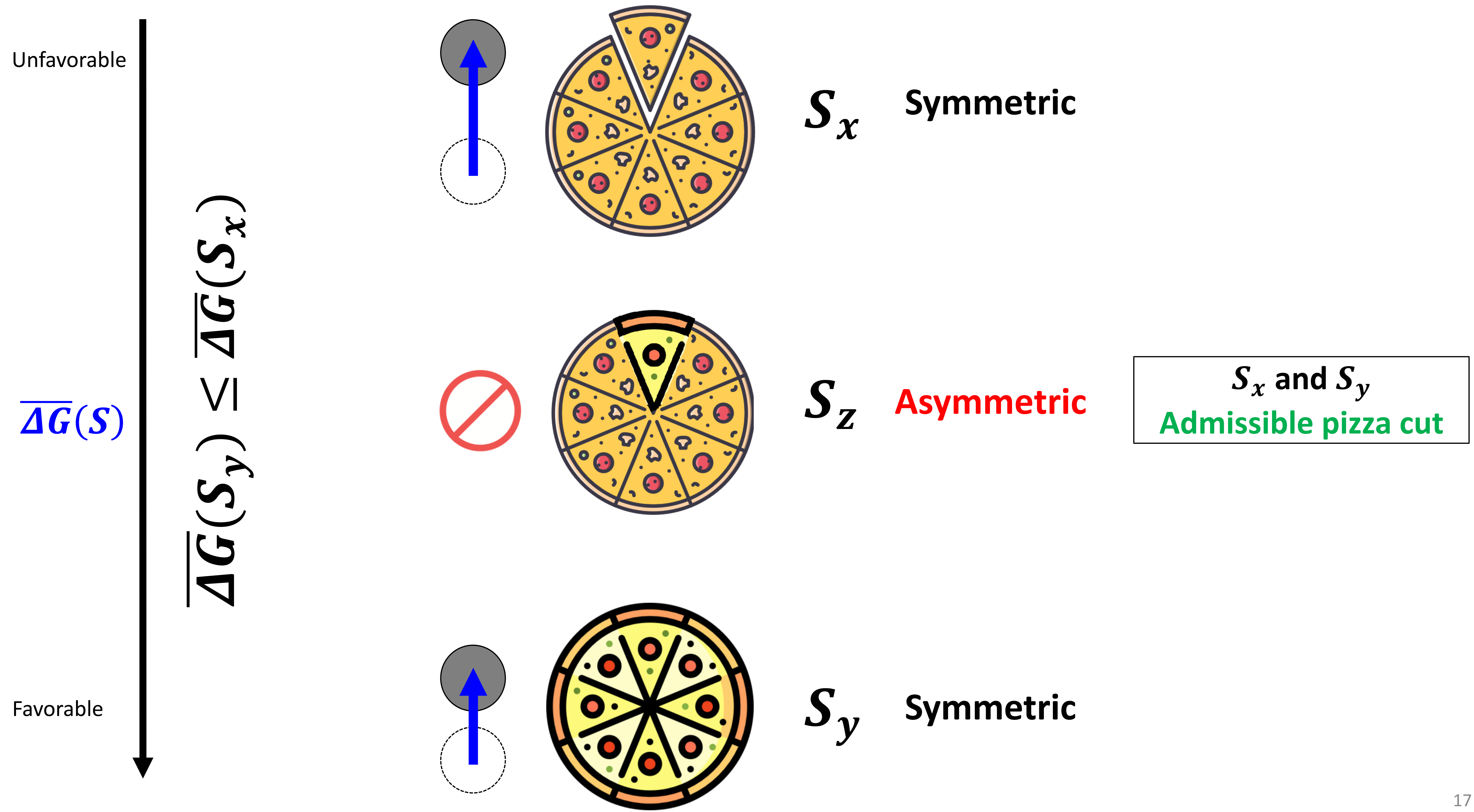
S_x and S_y

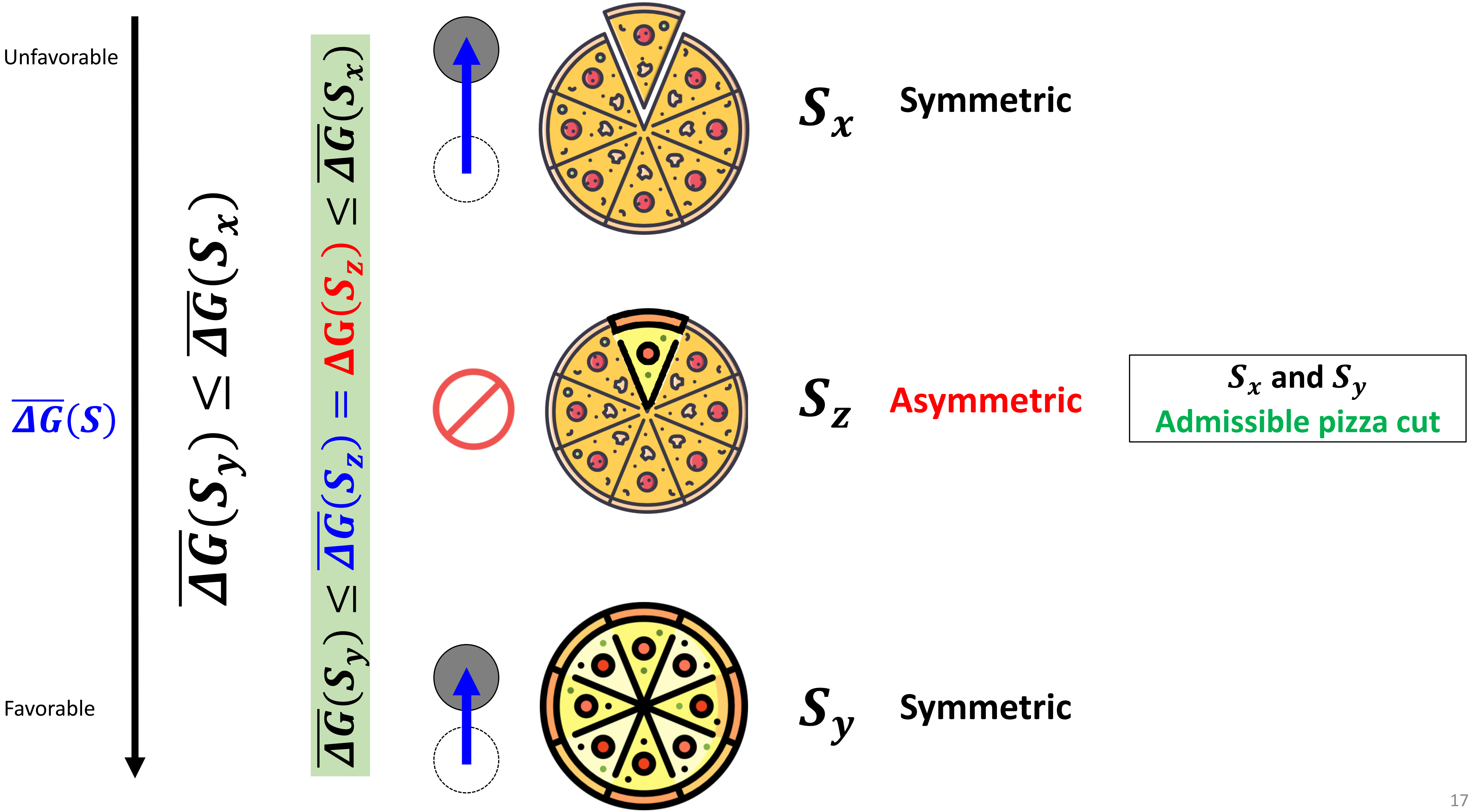
Admissible pizza cut

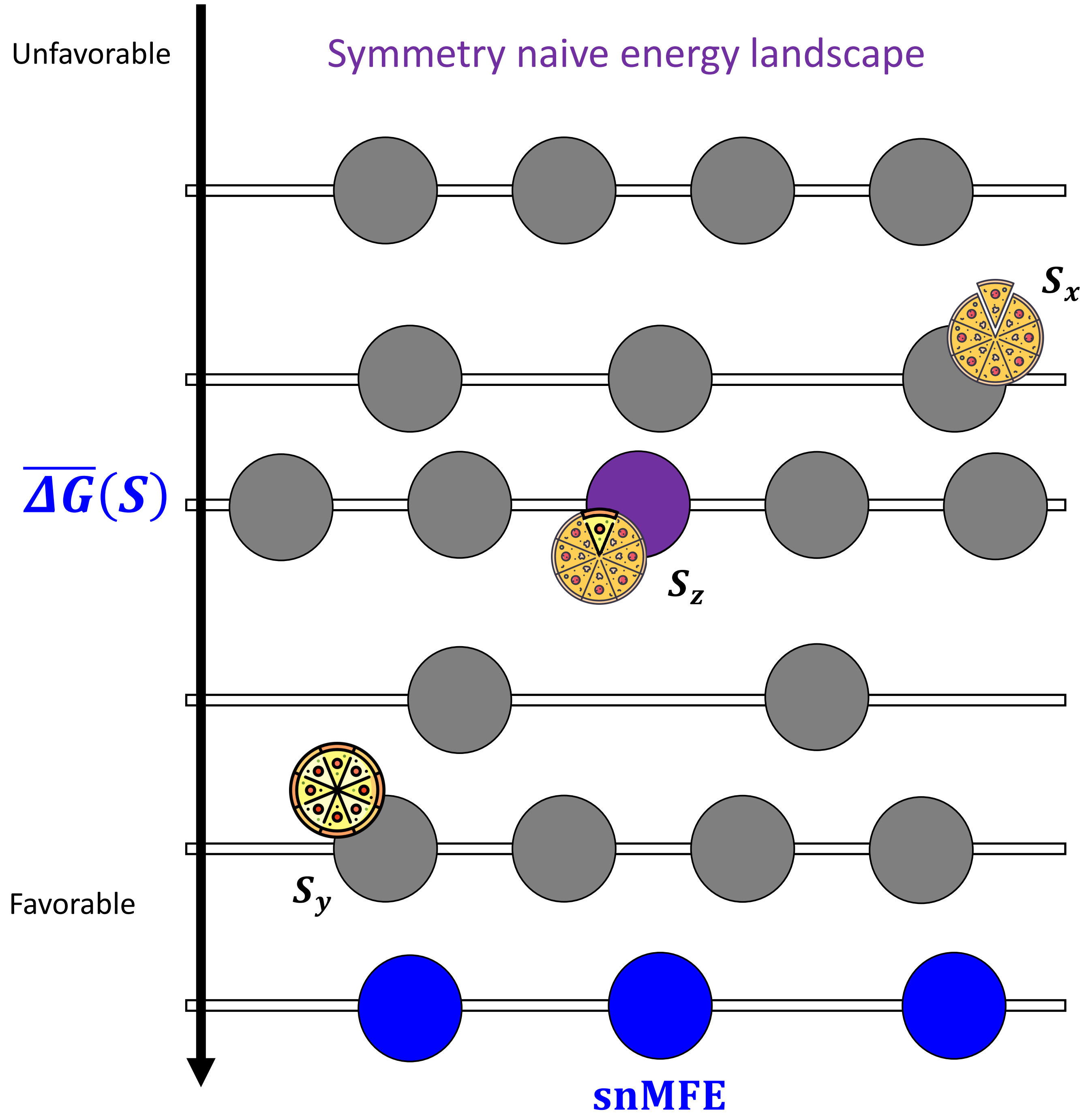




S_y Symmetric

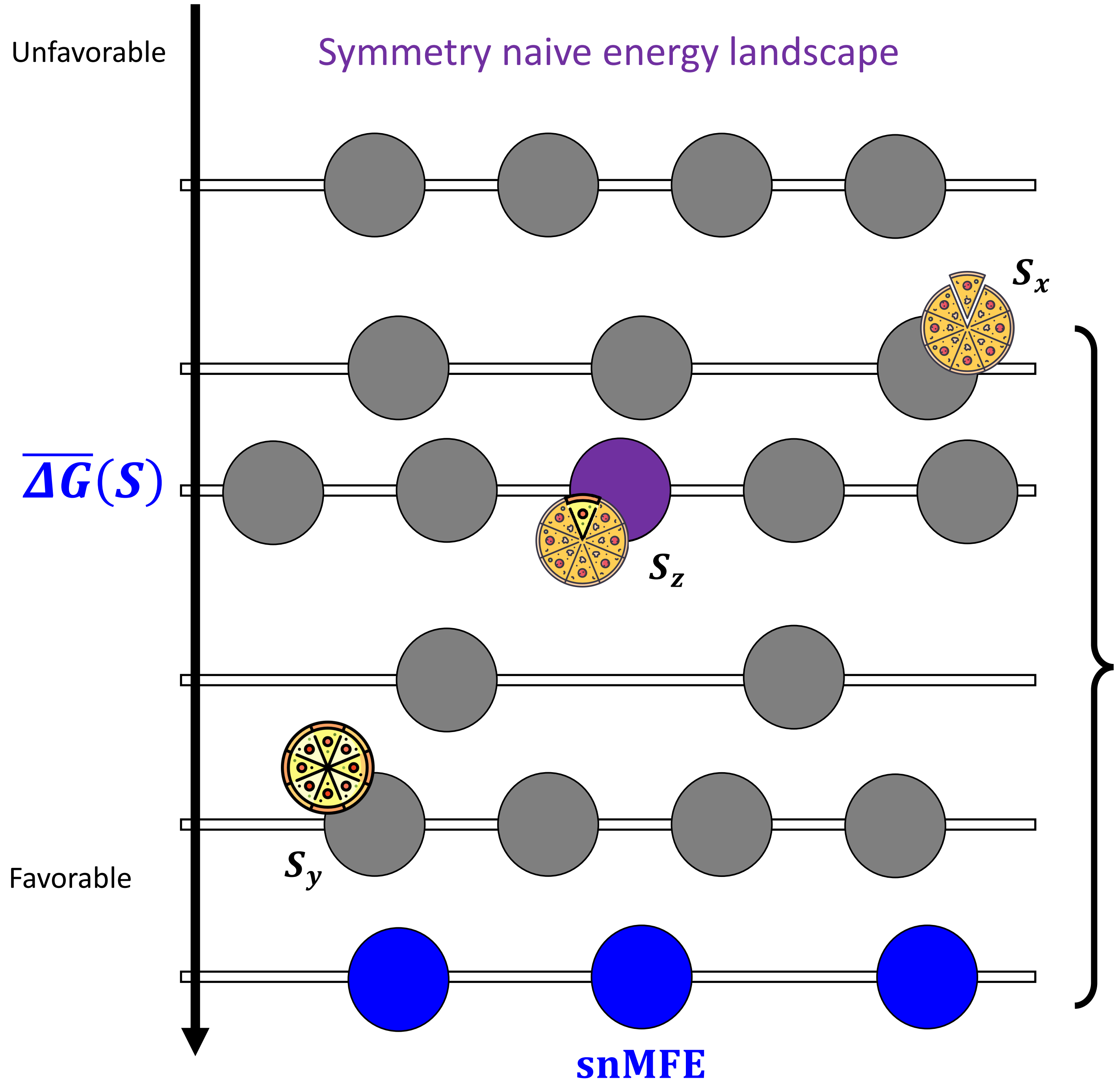

Unfavorable
$\overline{\Delta G}(S)$
Favorable
$\overline{\Delta G}(S_y) \leq \overline{\Delta G}(S_x)$
S_x Symmetric
S_z Asymmetric
S_y Symmetric
S_x and S_y
Admissible pizza cut


Unfavorable
$\overline{\Delta G}(S)$
Favorable
$\overline{\Delta G}(S_y) \leq \overline{\Delta G}(S_x)$
$\overline{\Delta G}(S_y) \leq \overline{\Delta G}(S_z) = \Delta G(S_z) \leq \overline{\Delta G}(S_x)$
S_x Symmetric
S_z Asymmetric
S_y Symmetric
S_x and S_y
Admissible pizza cut



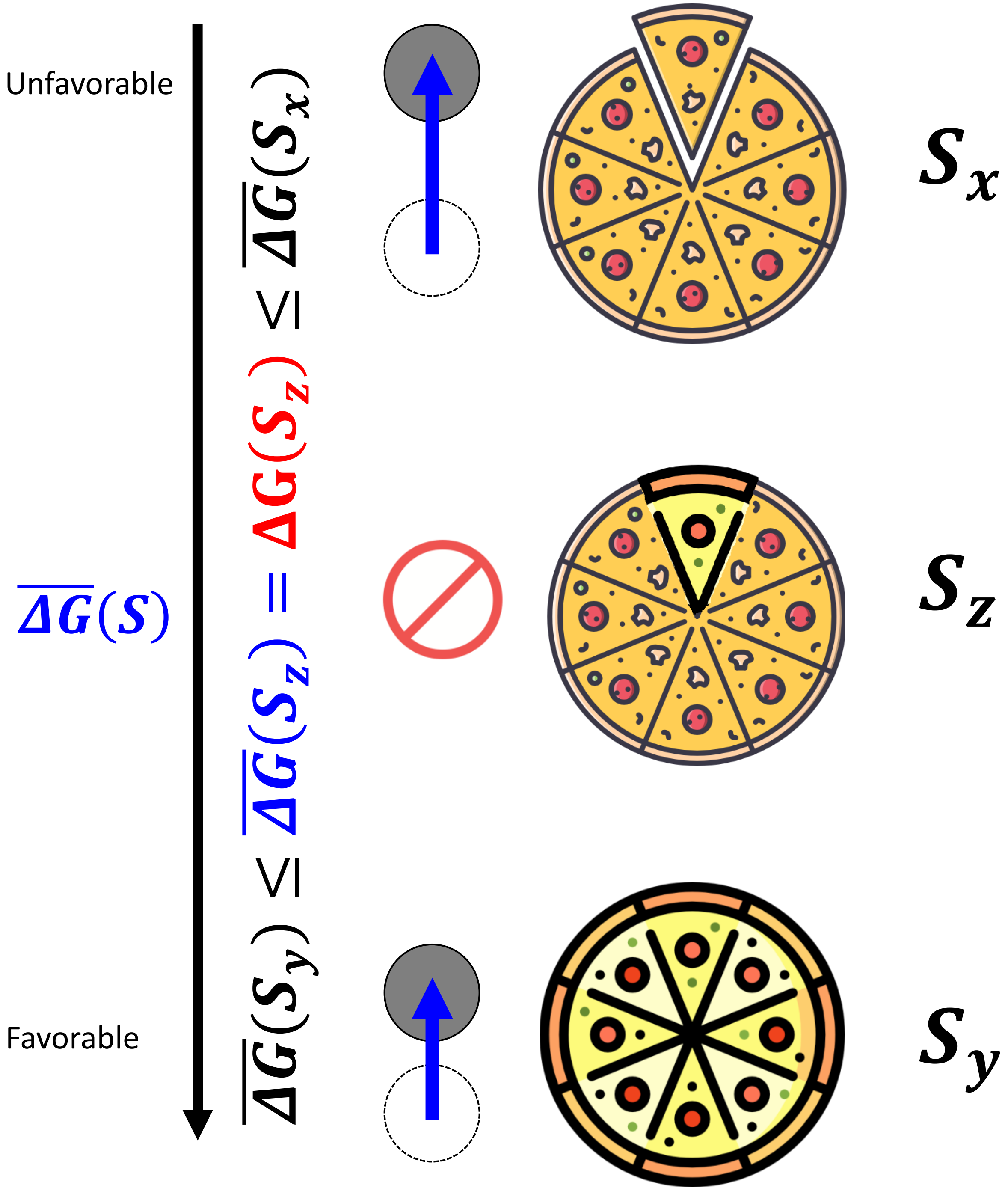
S_x and S_y

Admissible pizza cut

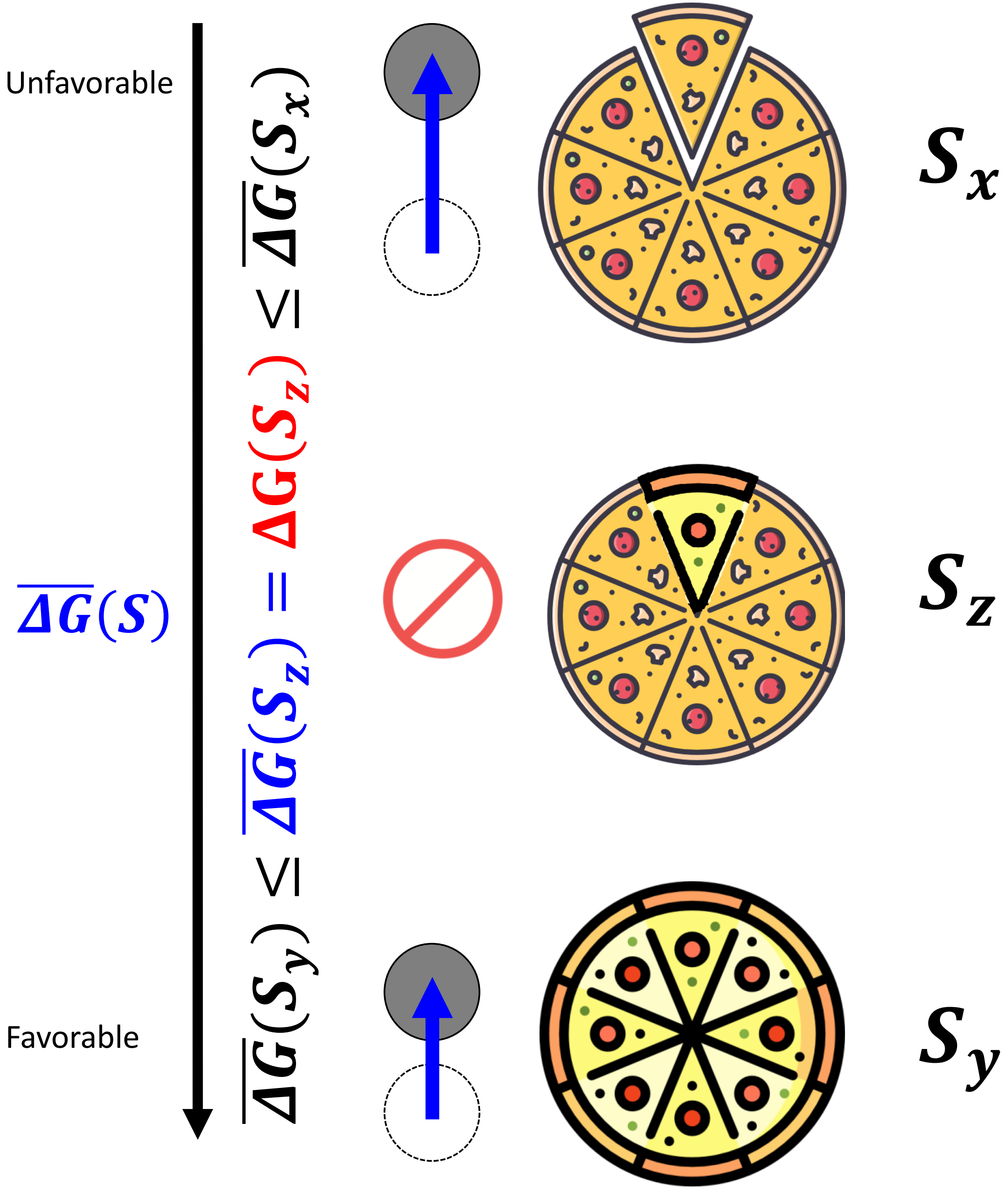


S_x and S_y
Admissible pizza cut

What is the maximum number of structures we need to backtrack until scanning S_x?



What is the maximum number of structures we need to backtrack until scanning S_x?



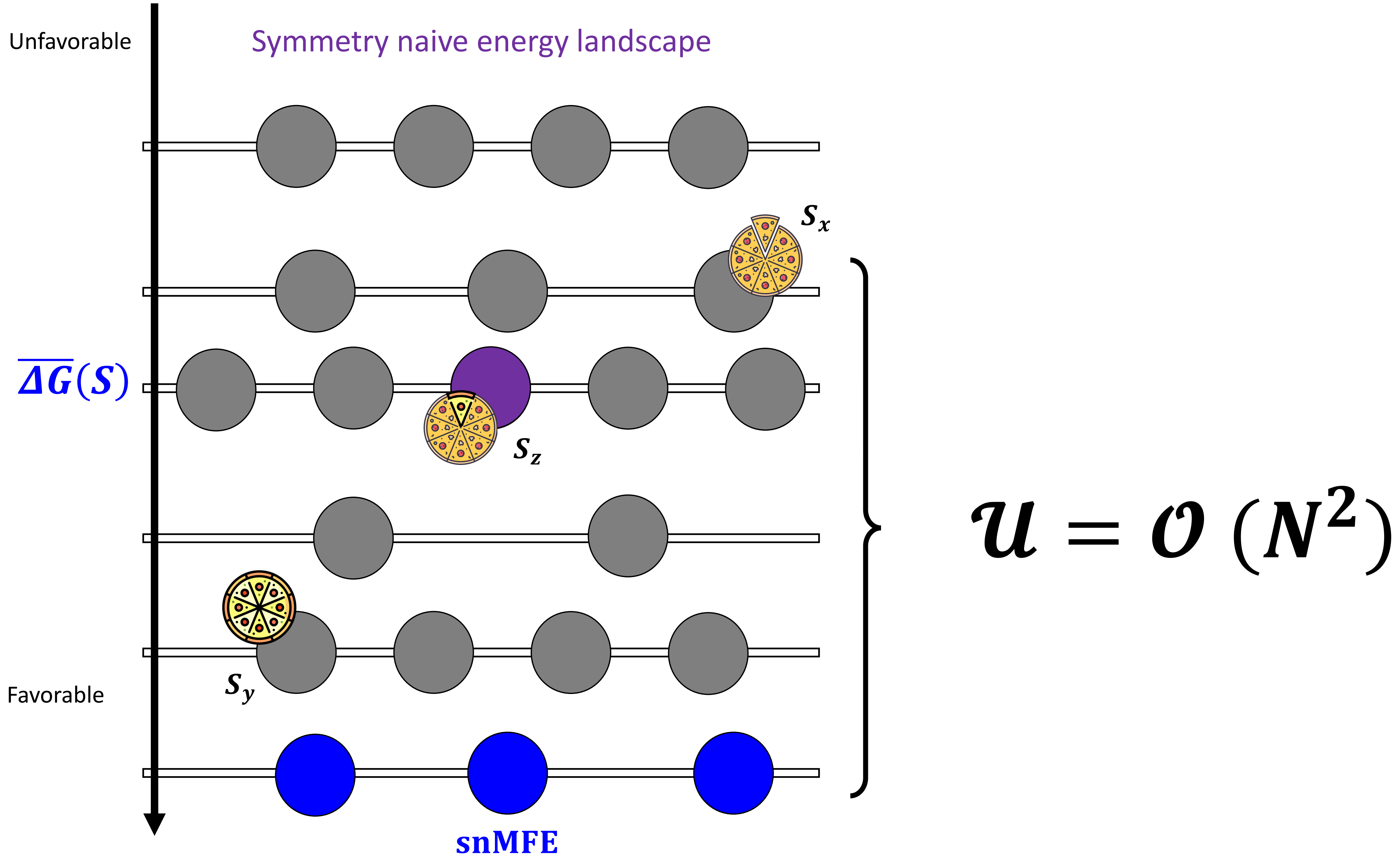
What is the maximum number of structures we need to backtrack until scanning S_x?

In worst case

Upper bound $\mathcal{U}$

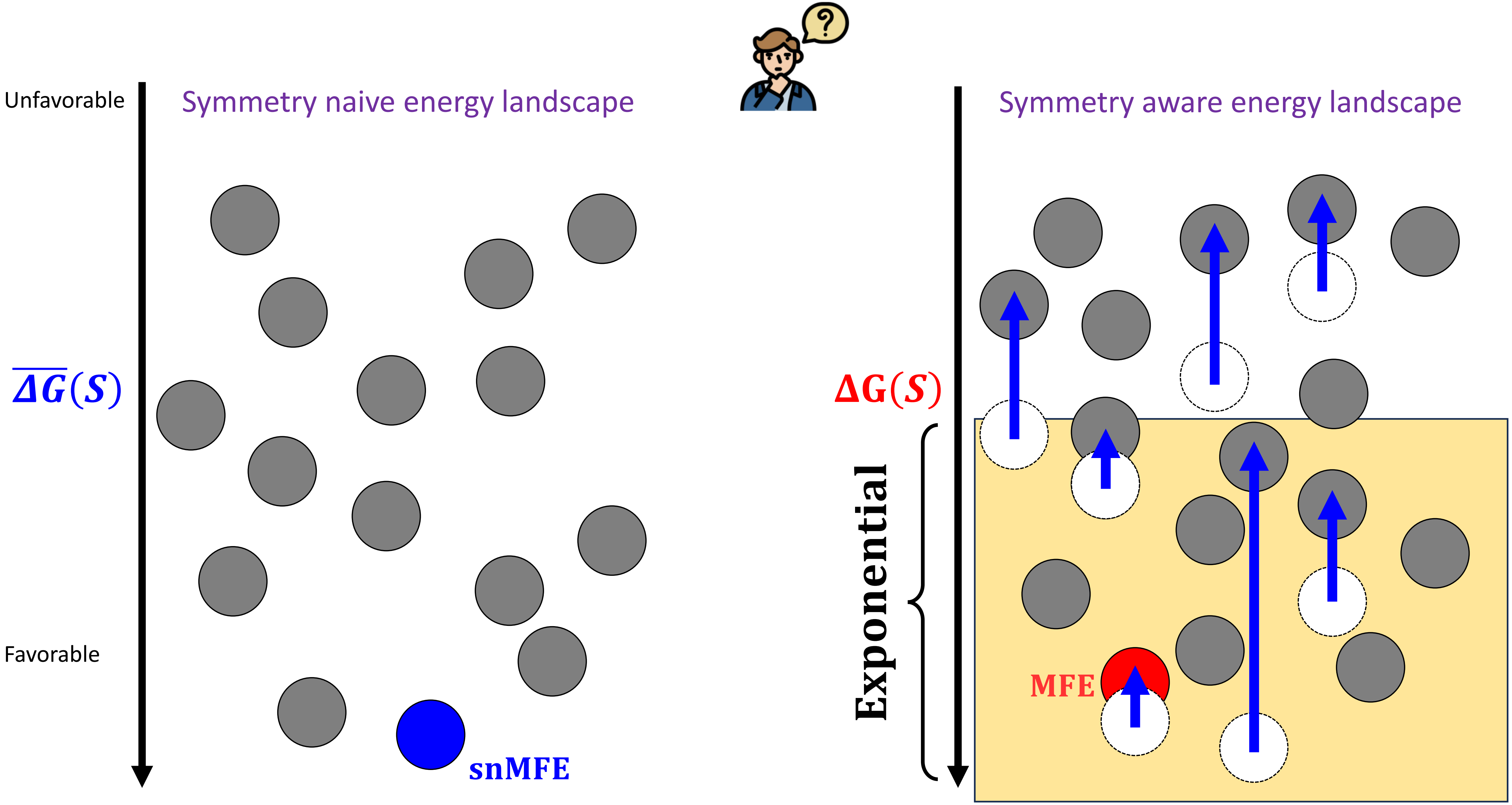
$$\frac{N-c}{v(\pi)}\left[\sigma(v(\pi)) - v(\pi)\right] + \frac{N^2}{16} = \mathcal{O}(N^2)$$

N: Number of bases
c: Number of strands
$v(\pi)$: Maximum degree of symmetry given the ordering π
$\sigma(.)$: Sum of divisors function

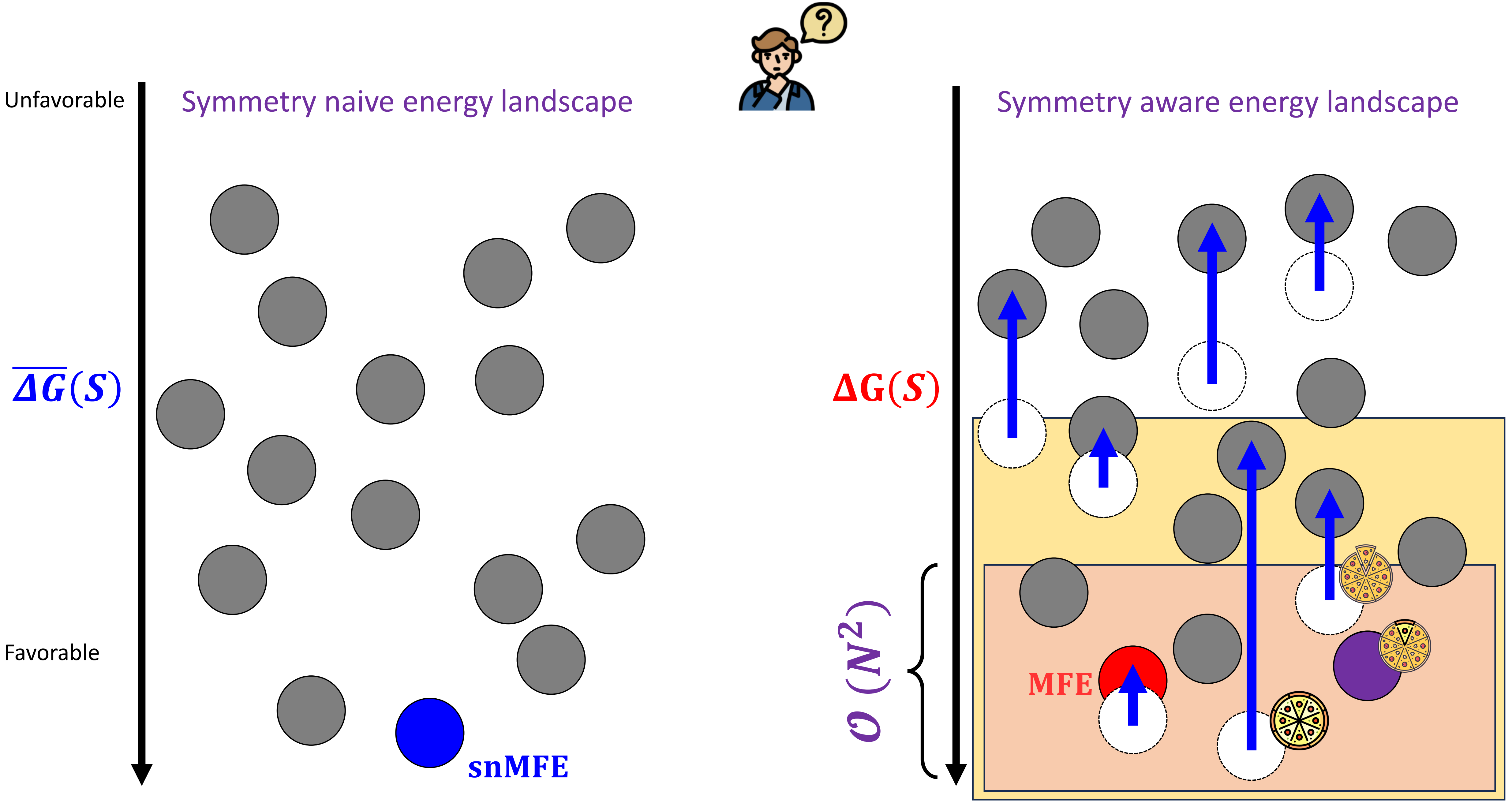


Unfavorable
Symmetry naive energy landscape
S_x
$\overline{\Delta G}(S)$
S_z
$\mathcal{U} = \mathcal{O}\left(N^2\right)$
S_y
Favorable
snMFE

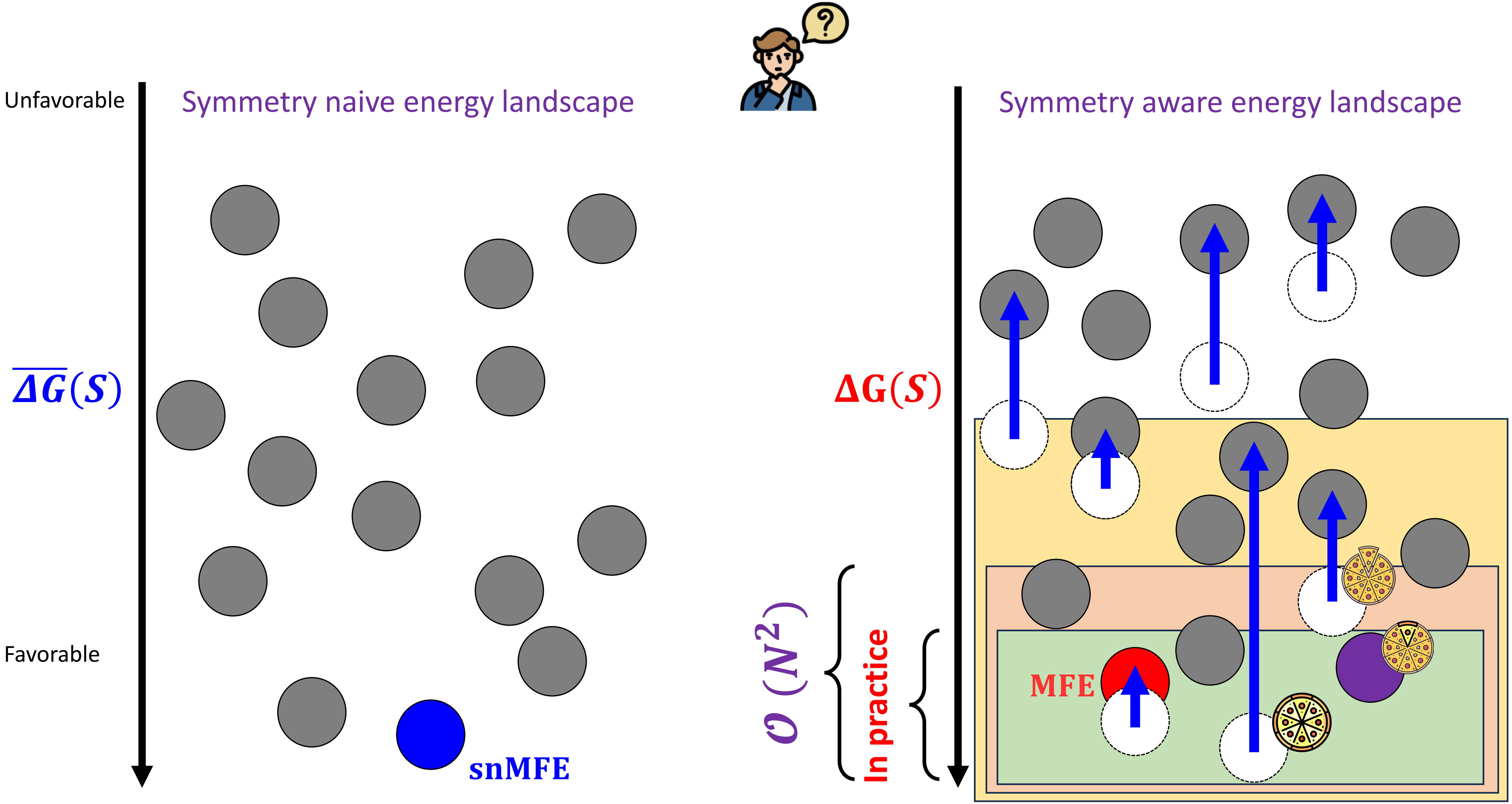
If snMFE structure is Symmetric $(R > 1)$



If snMFE structure is Symmetric $(R > 1)$



If snMFE structure is Symmetric $(R > 1)$



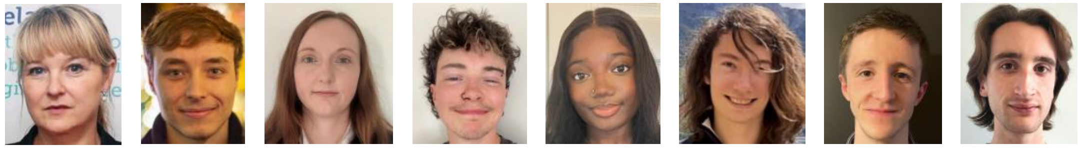
Computational complexity of MFE algorithms

Input Type	MFE
Single strand	$O(N^3)$
$c = O(1)$ **unique** strands	$O(N^3(c-1)!)$
$c = O(1)$ **allowing repeated** strands	$\boldsymbol{O(N^4(c-1)!)}$
Multiple strands, unbounded, $O(N)$ strands	$\mathrm{NP-Complete}$

ICALP 2025

N bases, c strands

Thanks



Maynooth University
National University of Ireland Maynooth

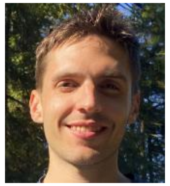
Hamilton Institute

Anne Condon

Dave Doty

Erik Winfree

Niles Pierce



Mark Fornace

David Soloveichik

Matthew Patitz

Sergiu Ivanov

dna.hamilton.ie shalabycave.com

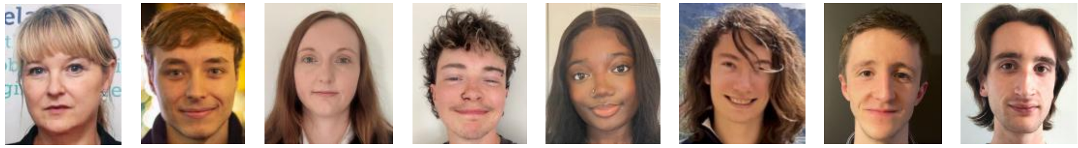
European Innovation Council

Funded by the European Union

erc

Taighde Éireann
Research Ireland

Thanks



Maynooth University
National University of Ireland Maynooth

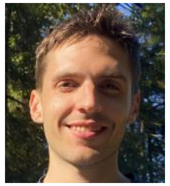
Hamilton Institute

Anne Condon

Dave Doty

Erik Winfree

Niles Pierce



Mark Fornace

David Soloveichik

Matthew Patitz

Sergiu Ivanov

dna.hamilton.ie

shalabycave.com

European Innovation Council

Funded by the European Union

erc

Taighde Éireann
Research Ireland