Statistical mechanics of the maximum-average submatrix problem

Vittorio Erba, Florent Krzakala, Rodrigo Pérez, and Lenka Zdeborová

arXiv:2303.05237

Swiss National Science Foundation

EPFL

Maximum-average submatrix problem

Given $N \times N$ matrix J of real numbers find $k \times k$ submatrix with **largest** sum of entries

Maximum-average submatrix problem

Given $N \times N$ matrix J of real numbers find $k \times k$ submatrix with **largest** sum of entries

Submatrix: intersection of k rows and columns, not necessarily adjacent

$$k = 4, N = 8$$

$$\begin{bmatrix}
J_{1,1} & J_{1,2} & J_{1,3} & J_{1,4} & J_{1,5} & J_{1,6} & J_{1,7} & J_{1,8} \\
J_{2,1} & J_{2,2} & J_{2,3} & J_{2,4} & J_{2,5} & J_{2,6} & J_{2,7} & J_{2,8} \\
J_{3,1} & J_{3,2} & J_{3,3} & J_{3,4} & J_{3,5} & J_{3,6} & J_{3,7} & J_{3,8} \\
J_{4,1} & J_{4,2} & J_{4,3} & J_{4,4} & J_{4,5} & J_{4,6} & J_{4,7} & J_{4,8} \\
J_{5,1} & J_{5,2} & J_{5,3} & J_{5,4} & J_{5,5} & J_{5,6} & J_{5,7} & J_{5,8} \\
J_{6,1} & J_{6,2} & J_{6,3} & J_{6,4} & J_{6,5} & J_{6,6} & J_{6,7} & J_{6,8} \\
J_{7,1} & J_{7,2} & J_{7,3} & J_{7,4} & J_{7,5} & J_{7,6} & J_{7,7} & J_{7,8} \\
J_{8,1} & J_{8,2} & J_{8,3} & J_{8,4} & J_{8,5} & J_{8,6} & J_{8,7} & J_{8,8}
\end{bmatrix}$$

Maximum-average submatrix problem

Given $N \times N$ matrix J of real numbers find $k \times k$ submatrix with **largest** sum of entries

Submatrix: intersection of k rows and columns, not necessarily adjacent

Encoding: use two two N-dim $0/1$ vector σ, τ

$\sigma_i = 1$ iff i-th row chosen

$\tau_j = 1$ iff j-th column chosen

$$\text{Energy}(\sigma, \tau) = \sum_{i,j} J_{ij}\sigma_i \tau_j$$

Random + high dim setting:

J iid Gaussian, $N \to \infty$

$k = 4, N = 8$

$$\begin{array}{c|c}
\sigma \backslash \tau & \begin{array}{cccccccc} \downarrow & & \downarrow & \downarrow & & & \downarrow & \end{array} \\
\begin{array}{c} \\ \rightarrow \\ \rightarrow \\ \\ \\ \rightarrow \\ \\ \rightarrow \end{array} &
\begin{bmatrix}
J_{1,1} & J_{1,2} & J_{1,3} & J_{1,4} & J_{1,5} & J_{1,6} & J_{1,7} & J_{1,8} \\
J_{2,1} & J_{2,2} & J_{2,3} & J_{2,4} & J_{2,5} & J_{2,6} & J_{2,7} & J_{2,8} \\
J_{3,1} & J_{3,2} & J_{3,3} & J_{3,4} & J_{3,5} & J_{3,6} & J_{3,7} & J_{3,8} \\
J_{4,1} & J_{4,2} & J_{4,3} & J_{4,4} & J_{4,5} & J_{4,6} & J_{4,7} & J_{4,8} \\
J_{5,1} & J_{5,2} & J_{5,3} & J_{5,4} & J_{5,5} & J_{5,6} & J_{5,7} & J_{5,8} \\
J_{6,1} & J_{6,2} & J_{6,3} & J_{6,4} & J_{6,5} & J_{6,6} & J_{6,7} & J_{6,8} \\
J_{7,1} & J_{7,2} & J_{7,3} & J_{7,4} & J_{7,5} & J_{7,6} & J_{7,7} & J_{7,8} \\
J_{8,1} & J_{8,2} & J_{8,3} & J_{8,4} & J_{8,5} & J_{8,6} & J_{8,7} & J_{8,8}
\end{bmatrix}
\end{array}$$

Why is this model interesting

N matrix size
k submatrix size

1. Results for random Gaussian i.i.d. J, $N \to \infty$ and $k \ll N$ (small submatrices)

Missing: large submatrix regime $k \sim N$ + can we say more for $k \ll N$?

[1] Bhamidi, Dey, and Nobel. "Energy landscape for large average submatrix detection problems in Gaussian random matrices." (2017)
[2] Gamarnik, Li "Finding a large submatrix of a Gaussian random matrix," (2018)
[3] El Cheairi, Gamarnik. "Densest Subgraphs of a Dense Erdös-Rènyi Graph. Asymptotics, Landscape and Universality." (2022).

Why is this model interesting

N matrix size
k submatrix size

1. Results for random Gaussian i.i.d. J, $N \to \infty$ and $k \ll N$ (small submatrices)

Missing: large submatrix regime $k \sim N$ + can we say more for $k \ll N$?

[1] Bhamidi, Dey, and Nobel. "Energy landscape for large average submatrix detection problems in Gaussian random matrices." (2017)
[2] Gamarnik, Li "Finding a large submatrix of a Gaussian random matrix," (2018)
[3] El Cheairi, Gamarnik. "Densest Subgraphs of a Dense Erdös-Rènyi Graph. Asymptotics, Landscape and Universality." (2022).

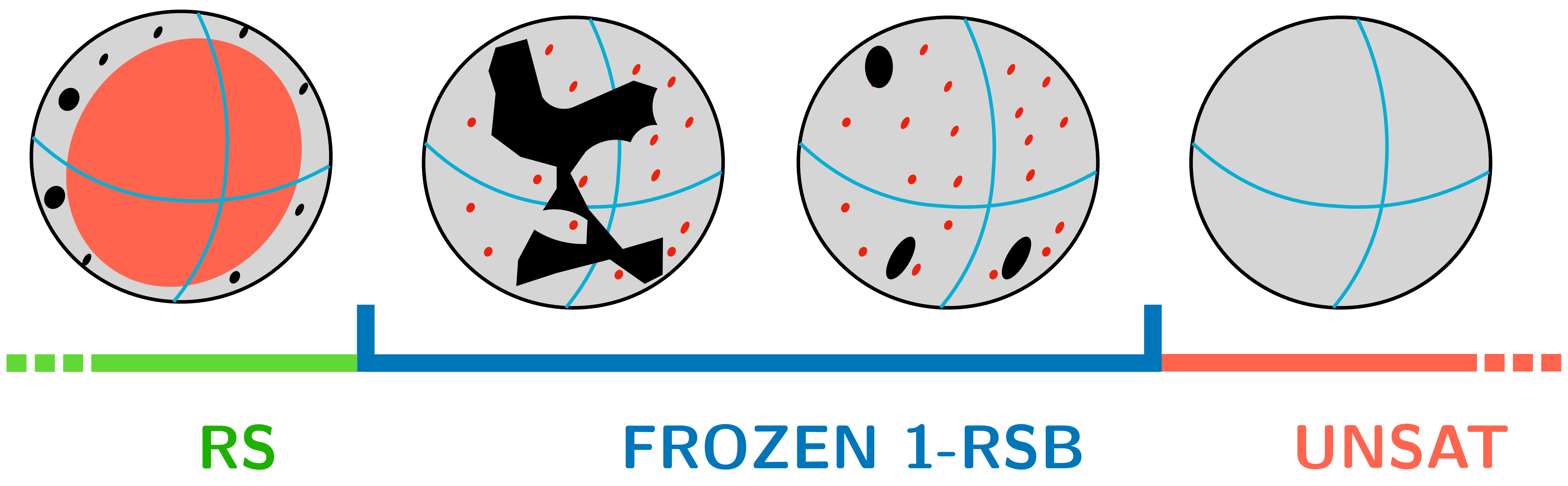
2. Unexpected link with rare cluster phenomenology of binary perceptron

Typical structure of energy level sets is clustered, but efficient algorithms work

Equilibrium vs

non-equilibrium vs

algorithmic properties?



Result: equilibrium phase diagram

N matrix size
k submatrix size
$m = k/N$

Symmetric case: $J = J^T$, same row/columns

Equilibrium/typical properties through

$$p(\sigma) \propto \exp \beta H(\sigma)$$

$$H(\sigma) = \frac{1}{2\sqrt{N}} \sum_{i<j} J_{ij} \sigma_i \sigma_j + h \sum_i \sigma_i$$

β: fix energy level

h: fix submatrix size

Variant of SK model computation ("classic" replicas)

- boolean spins + fixed magnetisation $m = k/N$
- small submatrix limit recovered for $m \to 0$
- rigorous [1]?

[1] Panchenko. "Free energy in the mixed p-spin models with vector spins." (2018)

Result: equilibrium phase diagram

N matrix size

k submatrix size

$m = k/N$

Symmetric case: $J = J^T$, same row/columns

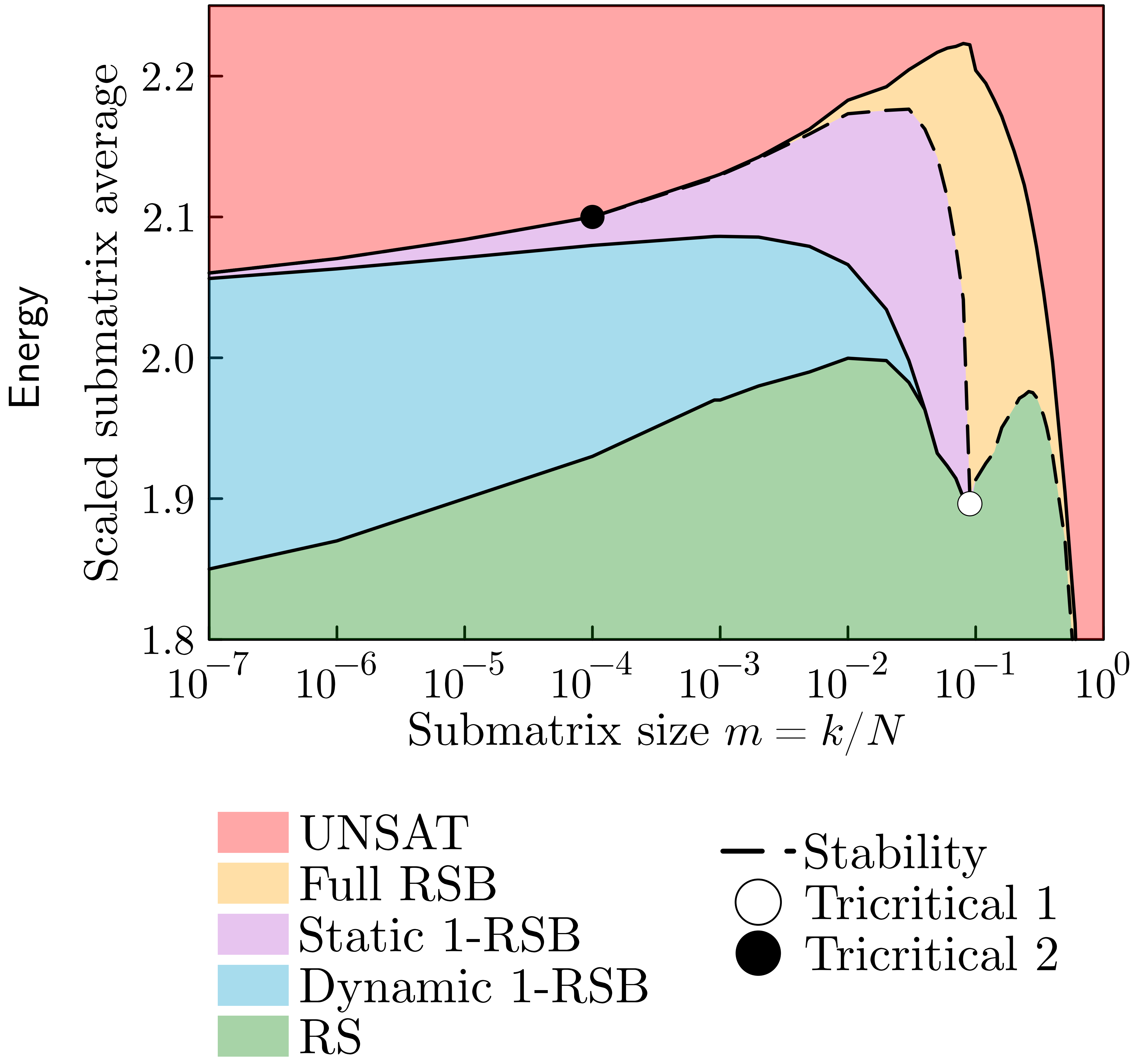
Equilibrium/typical properties through

$$p(\sigma) \propto \exp \beta H(\sigma)$$

$$H(\sigma) = \frac{1}{2\sqrt{N}} \sum_{i<j} J_{ij} \sigma_i \sigma_j + h \sum_i \sigma_i$$

β: fix energy level

h: fix submatrix size



Variant of SK model computation ("classic" replicas)

- boolean spins + fixed magnetisation $m = k/N$
- small submatrix limit recovered for $m \to 0$
- rigorous [1]?

[1] Panchenko. "Free energy in the mixed p-spin models with vector spins." (2018)

Result: equilibrium phase diagram

N matrix size
k submatrix size
$m = k/N$

Symmetric case: $J = J^T$, same row/columns

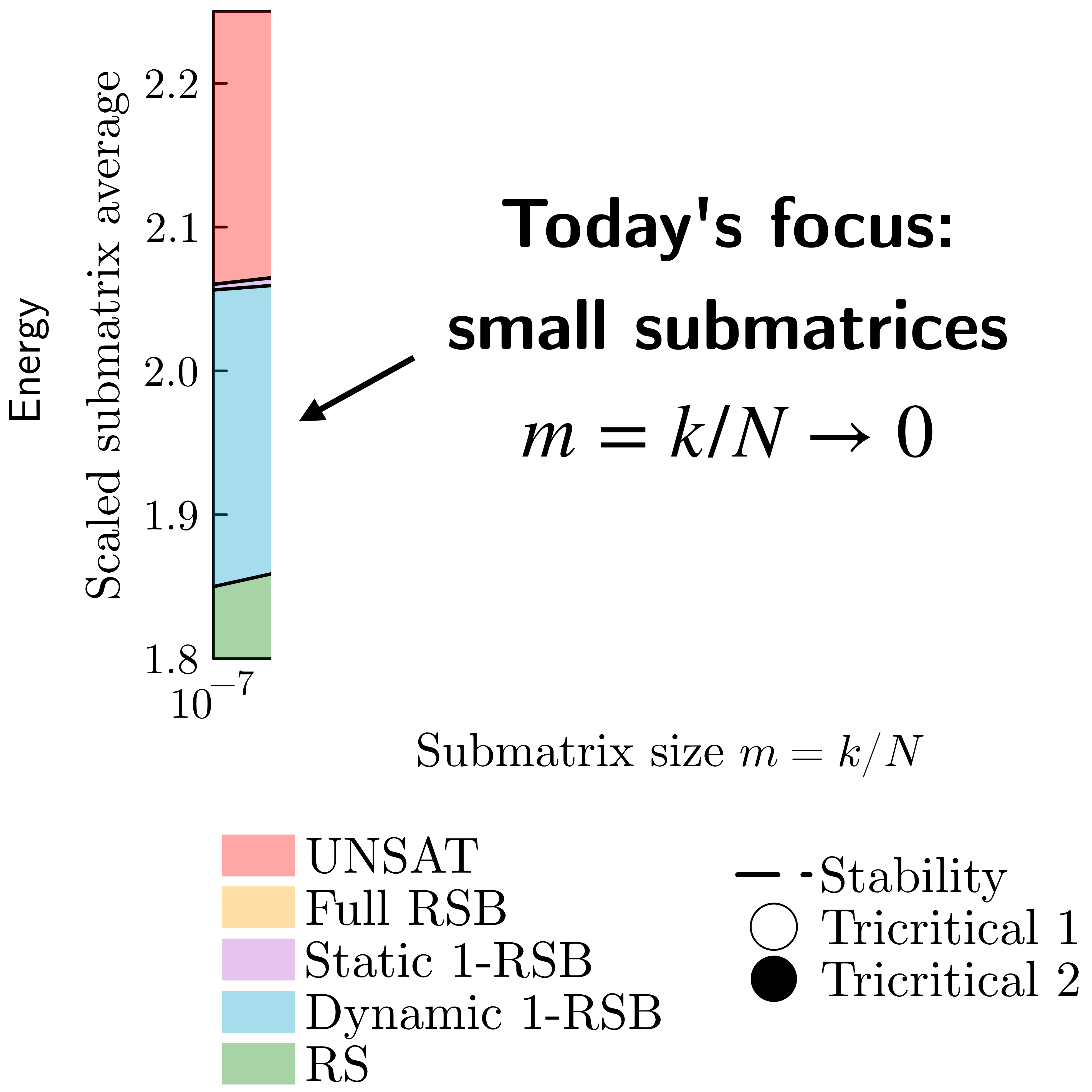
Equilibrium/typical properties through

$$p(\sigma) \propto \exp \beta H(\sigma)$$

$$H(\sigma) = \frac{1}{2\sqrt{N}} \sum_{i<j} J_{ij} \sigma_i \sigma_j + h \sum_i \sigma_i$$

β: fix energy level

h: fix submatrix size



Today's focus: small submatrices
$m = k/N \to 0$

Variant of SK model computation ("classic" replicas)
- boolean spins + fixed magnetisation $m = k/N$
- small submatrix limit recovered for $m \to 0$
- rigorous [1]?

[1] Panchenko. "Free energy in the mixed p-spin models with vector spins." (2018)

The phase diagram for small submatrices

N matrix size

k submatrix size

$m = k/N$

A submatrix avg

[1] Bhamidi, Dey, and Nobel. "Energy landscape for large average submatrix detection problems in Gaussian random matrices." (2017)
[2] Gamarnik, Li "Finding a large submatrix of a Gaussian random matrix," (2018)
[3] El Cheairi, Gamarnik. "Densest Subgraphs of a Dense Erdös-Rènyi Graph. Asymptotics, Landscape and Universality." (2022).

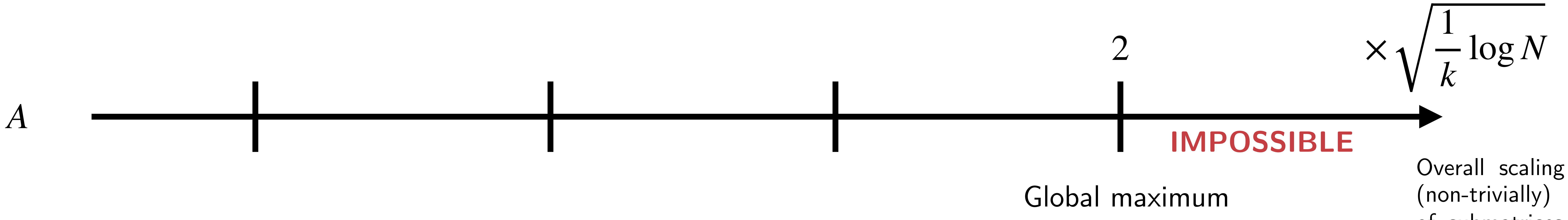
The phase diagram for small submatrices

N matrix size

k submatrix size

$m = k/N$

A submatrix avg



Overall scaling determined (non-trivially) by number of submatrices $\sim N^k$ and variance of submatrix-average A of $2k^{-2}$.

[1,3] The best possible submatrix-average equals the largest out of N^k independent Gaussians with variance $2k^{-2}$ (REM-like).

$$\begin{bmatrix}
J_{1,1} & J_{1,2} & J_{1,3} & J_{1,4} & J_{1,5} & J_{1,6} & J_{1,7} & J_{1,8} & J_{1,9} & J_{1,10} & J_{1,11} & J_{1,12} \\
J_{2,1} & J_{2,2} & J_{2,3} & J_{2,4} & J_{2,5} & J_{2,6} & J_{2,7} & J_{2,8} & J_{2,9} & J_{2,10} & J_{2,11} & J_{2,12} \\
J_{3,1} & J_{3,2} & J_{3,3} & J_{3,4} & J_{3,5} & J_{3,6} & J_{3,7} & J_{3,8} & J_{3,9} & J_{3,10} & J_{3,11} & J_{3,12} \\
J_{4,1} & J_{4,2} & J_{4,3} & J_{4,4} & J_{4,5} & J_{4,6} & J_{4,7} & J_{4,8} & J_{4,9} & J_{4,10} & J_{4,11} & J_{4,12} \\
J_{5,1} & J_{5,2} & J_{5,3} & J_{5,4} & J_{5,5} & J_{5,6} & J_{5,7} & J_{5,8} & J_{5,9} & J_{5,10} & J_{5,11} & J_{5,12} \\
J_{6,1} & J_{6,2} & J_{6,3} & J_{6,4} & J_{6,5} & J_{6,6} & J_{6,7} & J_{6,8} & J_{6,9} & J_{6,10} & J_{6,11} & J_{6,12} \\
J_{7,1} & J_{7,2} & J_{7,3} & J_{7,4} & J_{7,5} & J_{7,6} & J_{7,7} & J_{7,8} & J_{7,9} & J_{7,10} & J_{7,11} & J_{7,12} \\
J_{8,1} & J_{8,2} & J_{8,3} & J_{8,4} & J_{8,5} & J_{8,6} & J_{8,7} & J_{8,8} & J_{8,9} & J_{8,10} & J_{8,11} & J_{8,12} \\
J_{9,1} & J_{9,2} & J_{9,3} & J_{9,4} & J_{9,5} & J_{9,6} & J_{9,7} & J_{9,8} & J_{9,9} & J_{9,10} & J_{9,11} & J_{9,12} \\
J_{10,1} & J_{10,2} & J_{10,3} & J_{10,4} & J_{10,5} & J_{10,6} & J_{10,7} & J_{10,8} & J_{10,9} & J_{10,10} & J_{10,11} & J_{10,12} \\
J_{11,1} & J_{11,2} & J_{11,3} & J_{11,4} & J_{11,5} & J_{11,6} & J_{11,7} & J_{11,8} & J_{11,9} & J_{11,10} & J_{11,11} & J_{11,12} \\
J_{12,1} & J_{12,2} & J_{12,3} & J_{12,4} & J_{12,5} & J_{12,6} & J_{12,7} & J_{12,8} & J_{12,9} & J_{12,10} & J_{12,11} & J_{12,12}
\end{bmatrix}$$

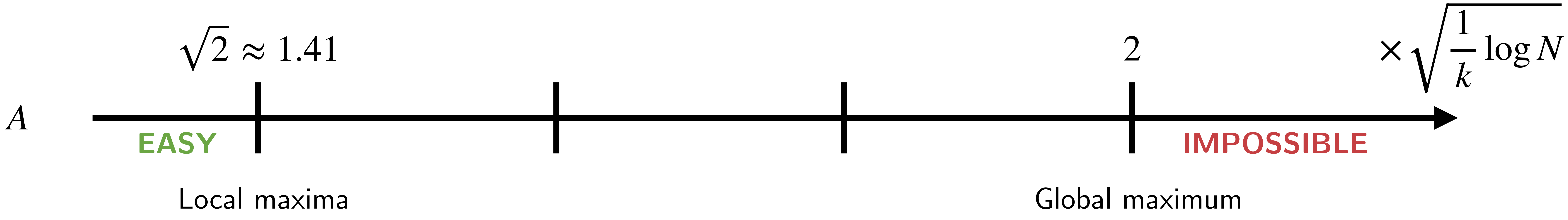
[1] Bhamidi, Dey, and Nobel. "Energy landscape for large average submatrix detection problems in Gaussian random matrices." (2017)

[2] Gamarnik, Li "Finding a large submatrix of a Gaussian random matrix," (2018)

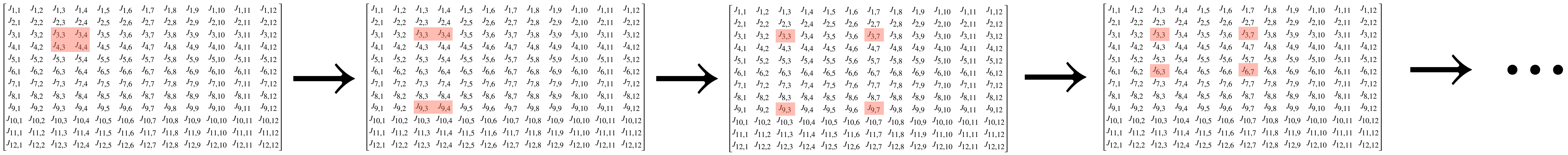
[3] El Cheairi, Gamarnik. "Densest Subgraphs of a Dense Erdös-Rènyi Graph. Asymptotics, Landscape and Universality." (2022).

The phase diagram for small submatrices

N matrix size
k submatrix size
$m = k/N$
A submatrix avg



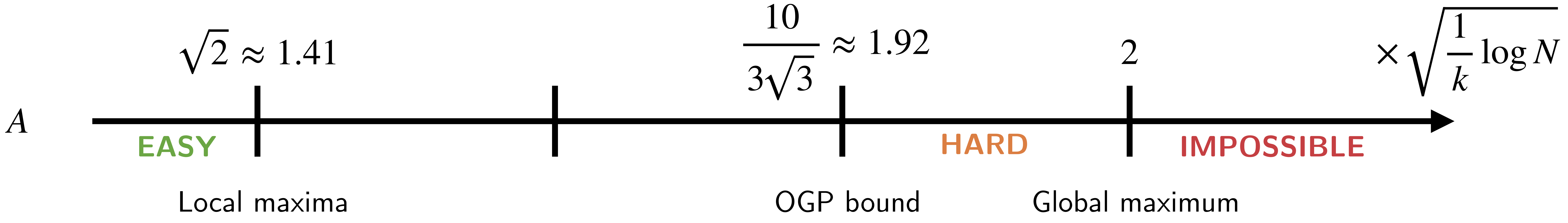
[1,2] Local maxima (stable among same-row or same-column) have lower submatrix-average than the best one, and can be found by an efficient iterative algorithm

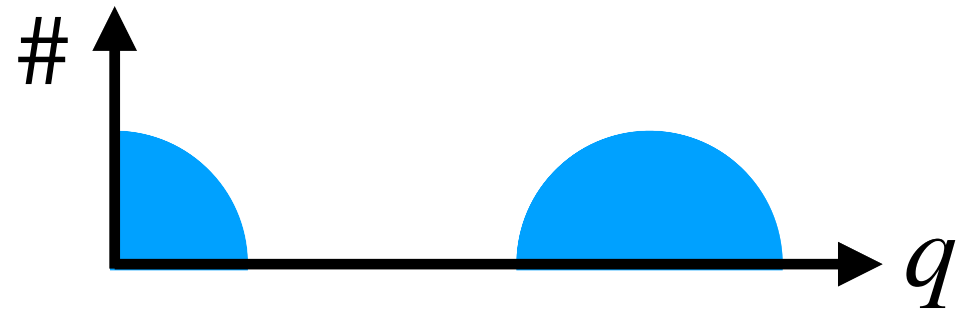


[1] Bhamidi, Dey, and Nobel. "Energy landscape for large average submatrix detection problems in Gaussian random matrices." (2017)
[2] Gamarnik, Li "Finding a large submatrix of a Gaussian random matrix," (2018)
[3] El Cheairi, Gamarnik. "Densest Subgraphs of a Dense Erdös-Rènyi Graph. Asymptotics, Landscape and Universality." (2022).

The phase diagram for small submatrices

N matrix size
k submatrix size
$m = k/N$
A submatrix avg





[2] Algorithmic hardness bound through Overlap Gap Property ⟶

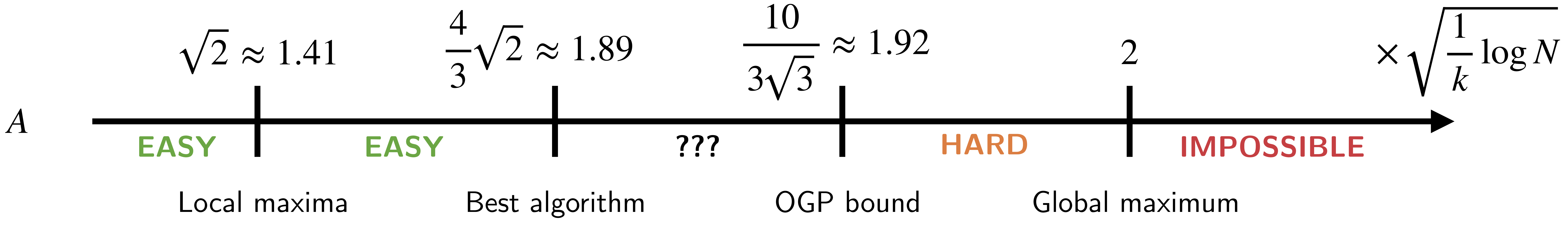
Overlap Gap Property: w.h.p. on the disorder, each pair of submatrices with submatrix-average at this energy is either very close (large overlap q) or very far (zero overlap q). No intermediate distance can be observed.

Gamarnik. "The overlap gap property: A topological barrier to optimizing over random structures." (2021)

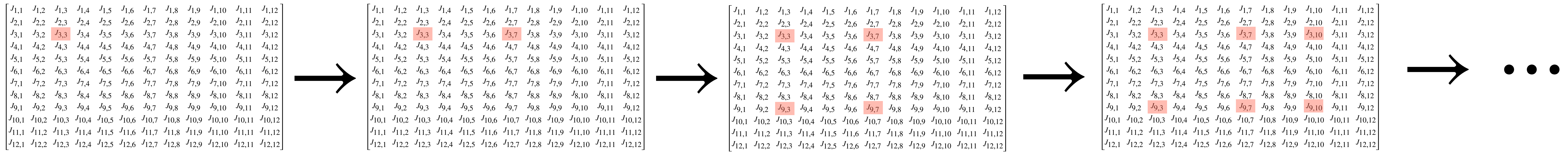
[1] Bhamidi, Dey, and Nobel. "Energy landscape for large average submatrix detection problems in Gaussian random matrices." (2017)
[2] Gamarnik, Li "Finding a large submatrix of a Gaussian random matrix," (2018)
[3] El Cheairi, Gamarnik. "Densest Subgraphs of a Dense Erdös-Rènyi Graph. Asymptotics, Landscape and Universality." (2022).

The phase diagram for small submatrices

N matrix size
k submatrix size
$m = k/N$
A submatrix avg



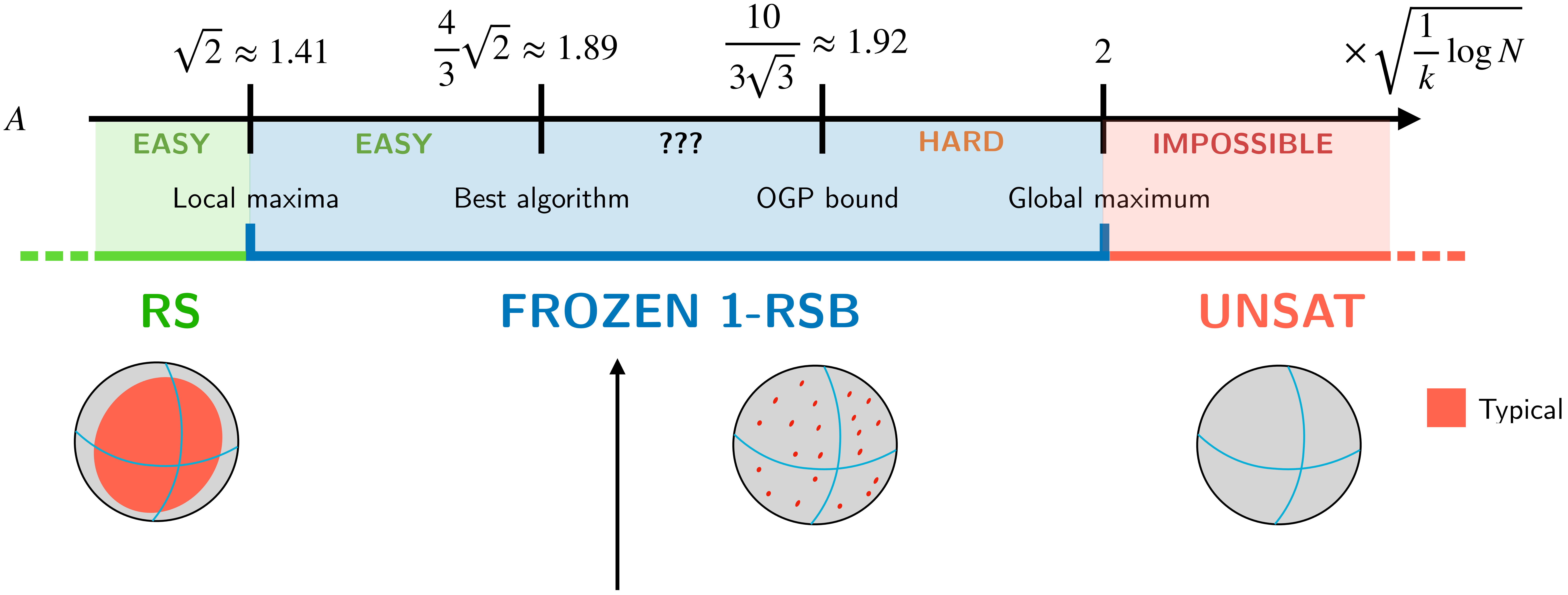
[2] Best efficient algorithm (greedy) performs better than typical local-maxima



[1] Bhamidi, Dey, and Nobel. "Energy landscape for large average submatrix detection problems in Gaussian random matrices." (2017)
[2] Gamarnik, Li "Finding a large submatrix of a Gaussian random matrix," (2018)
[3] El Cheairi, Gamarnik. "Densest Subgraphs of a Dense Erdös-Rènyi Graph. Asymptotics, Landscape and Universality." (2022).

Results: weak clustering phase

N matrix size
k submatrix size
$m = k/N$
A submatrix avg

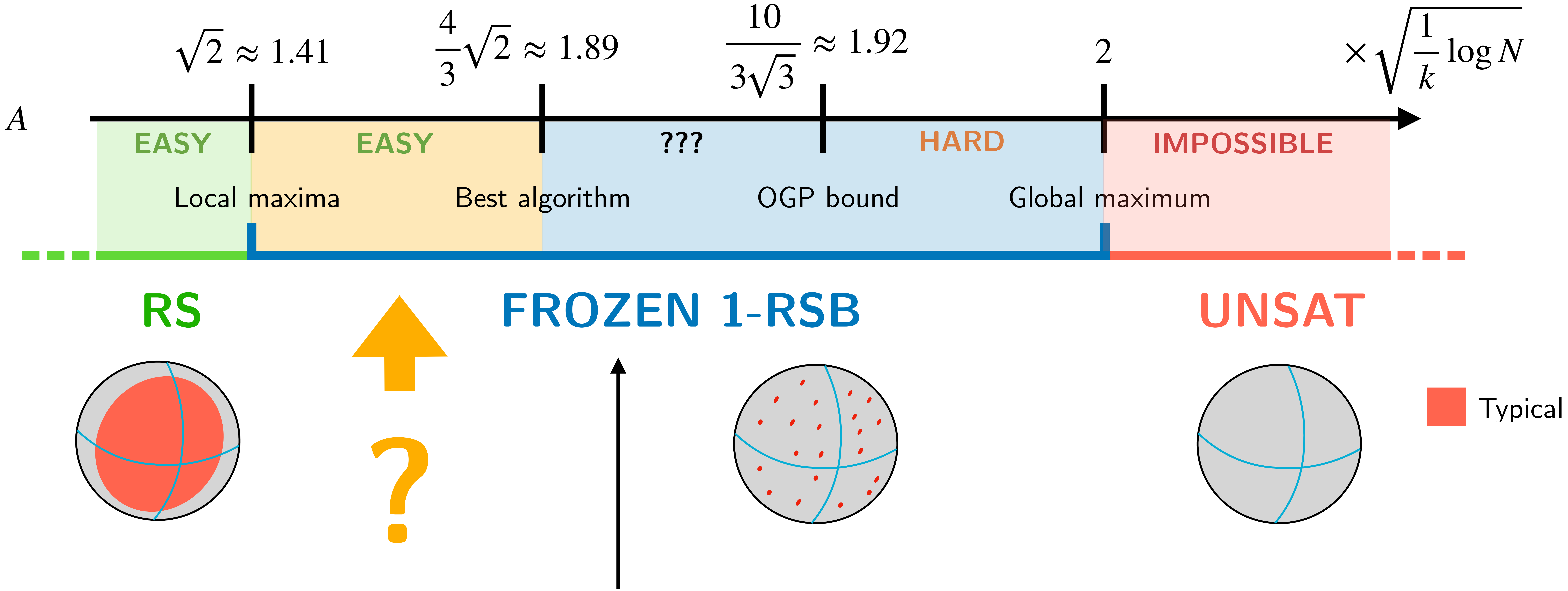


Clustering: exponential number of orthogonal clusters

Freezing: each cluster has zero entropy + internal overlap = 1

Algorithm working in frozen phase?

N matrix size
k submatrix size
$m = k/N$
A submatrix avg

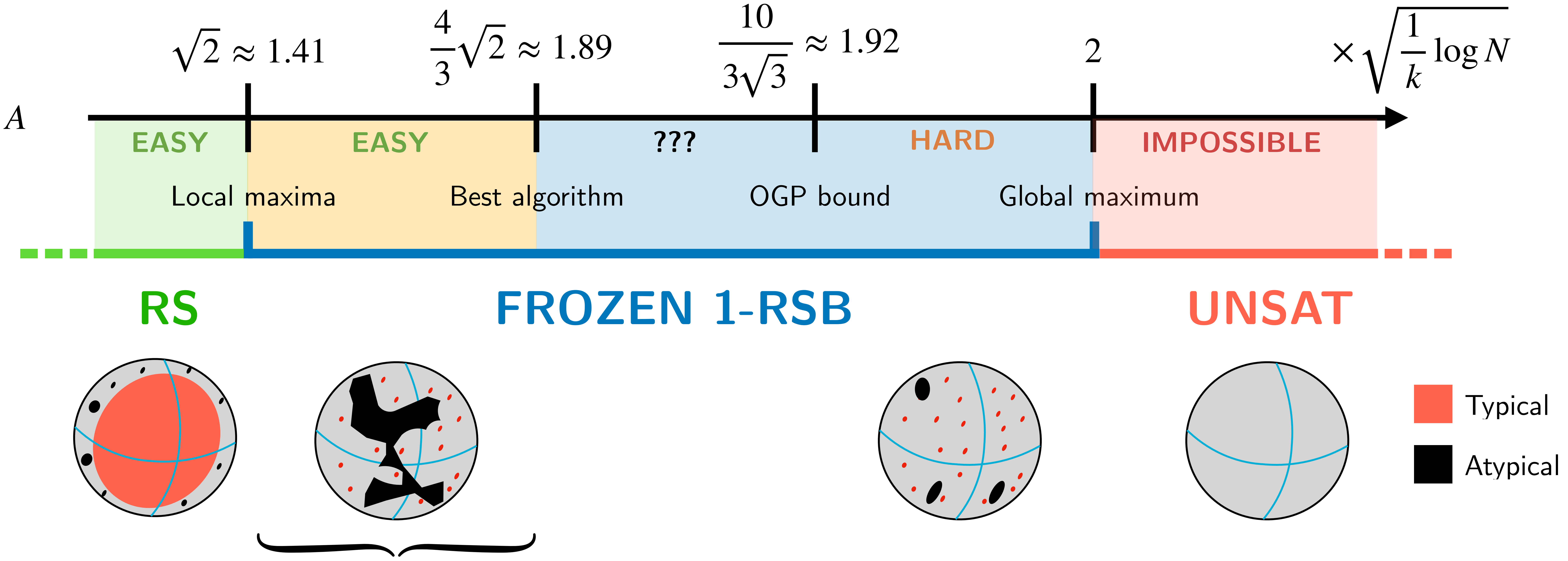


Clustering: exponential number of orthogonal clusters

Freezing: each cluster has zero entropy + internal overlap = 1

Atypical clusters?

N matrix size
k submatrix size
$m = k/N$
A submatrix avg

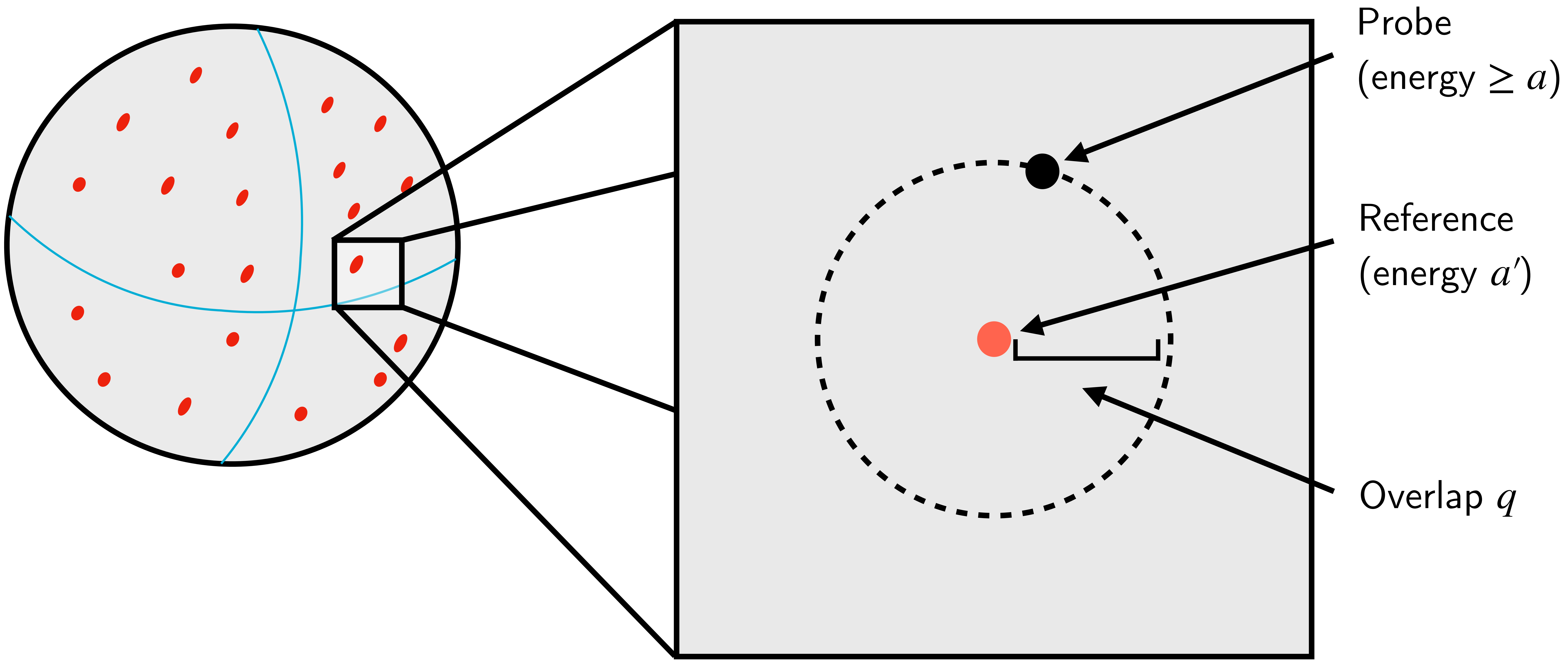


In binary perceptron: atypical clusters unfrozen + connected [1,2].

Is the same happening in this model? Need to look at non-equilibrium

[1] Baldassi et al. "Subdominant dense clusters allow for simple learning and high computational performance in neural networks with discrete synapses." (2015)
[2] Abbe, Li, Sly. "Binary perceptron: efficient algorithms can find solutions in a rare well-connected cluster." (2022)

Result: a first look at non-equilibrium properties

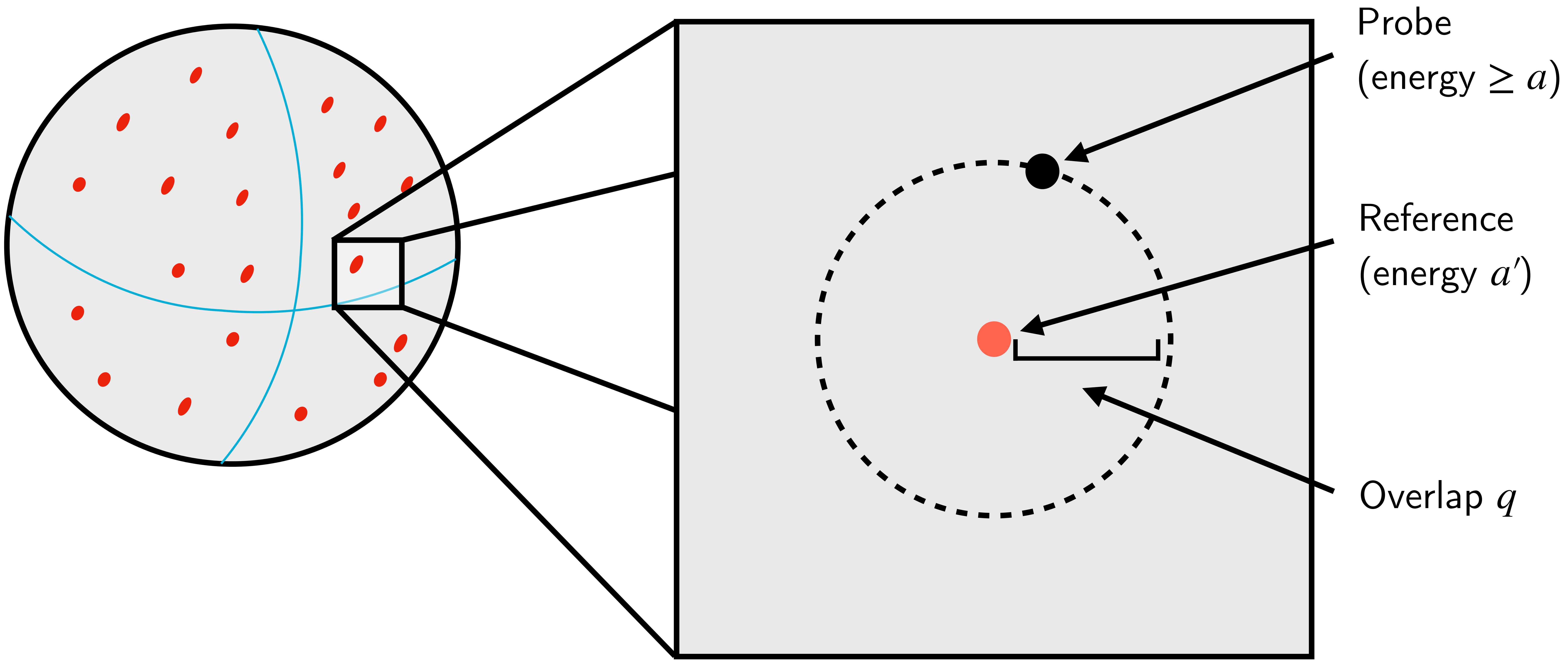


Franz-Parisi potential (For the curious: RS ansatz for probe, dynamical 1RSB ansatz for reference)

Entropy of probes (energy $\geq a$) at fixed overlap q with typical frozen states (energy a')

$\iff$ structure of super-level sets around typical frozen states

Result: a first look at non-equilibrium properties



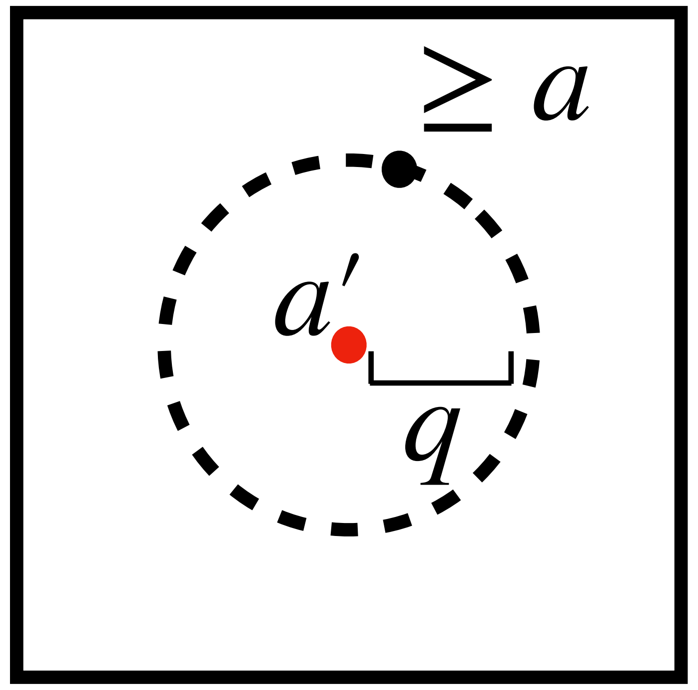
Only looks at surroundings of frozen clusters! Other atypical clusters could still exists!

Franz-Parisi potential (For the curious: RS ansatz for probe, dynamical 1RSB ansatz for reference)

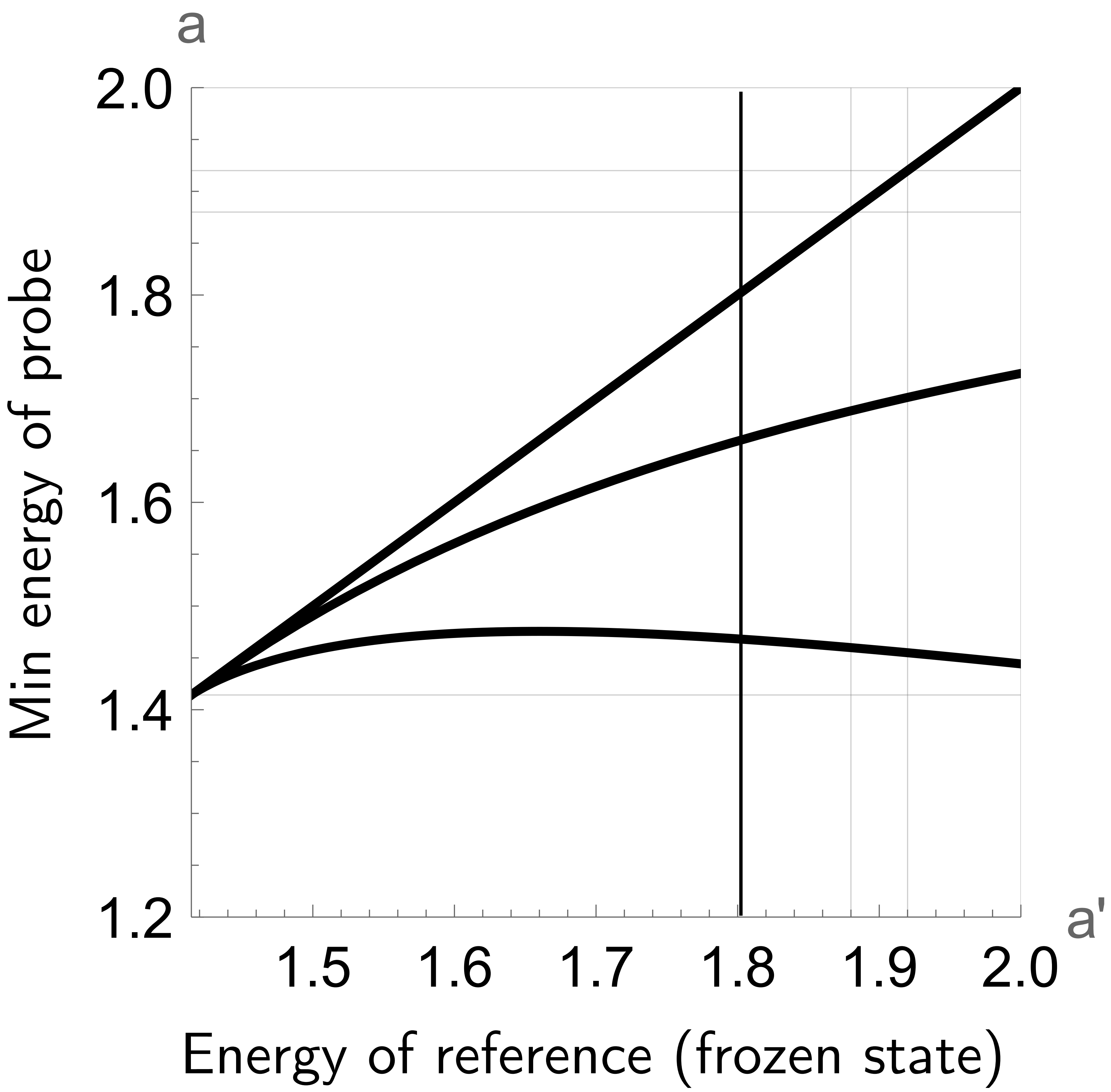
Entropy of probes (energy $\geq a$) at fixed overlap q with typical frozen states (energy a')

$\iff$ structure of super-level sets around typical frozen states

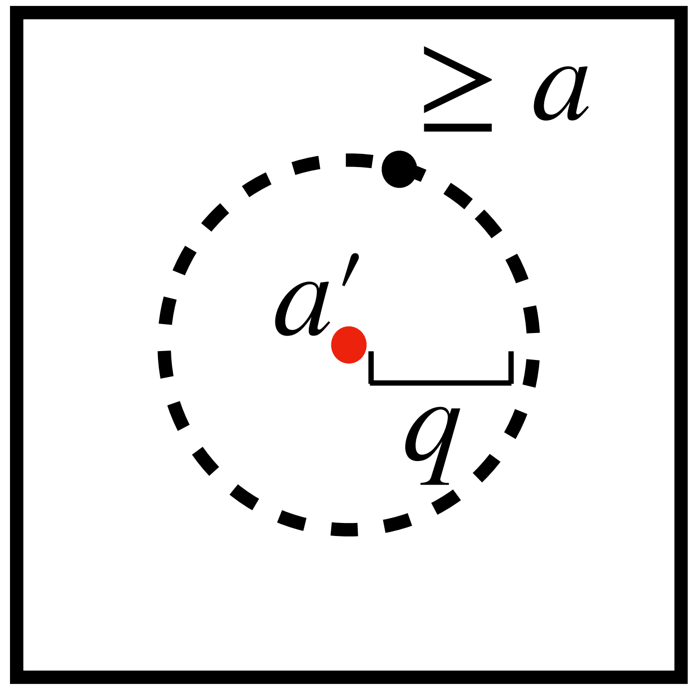
Result: Planted clusters in superlevel sets



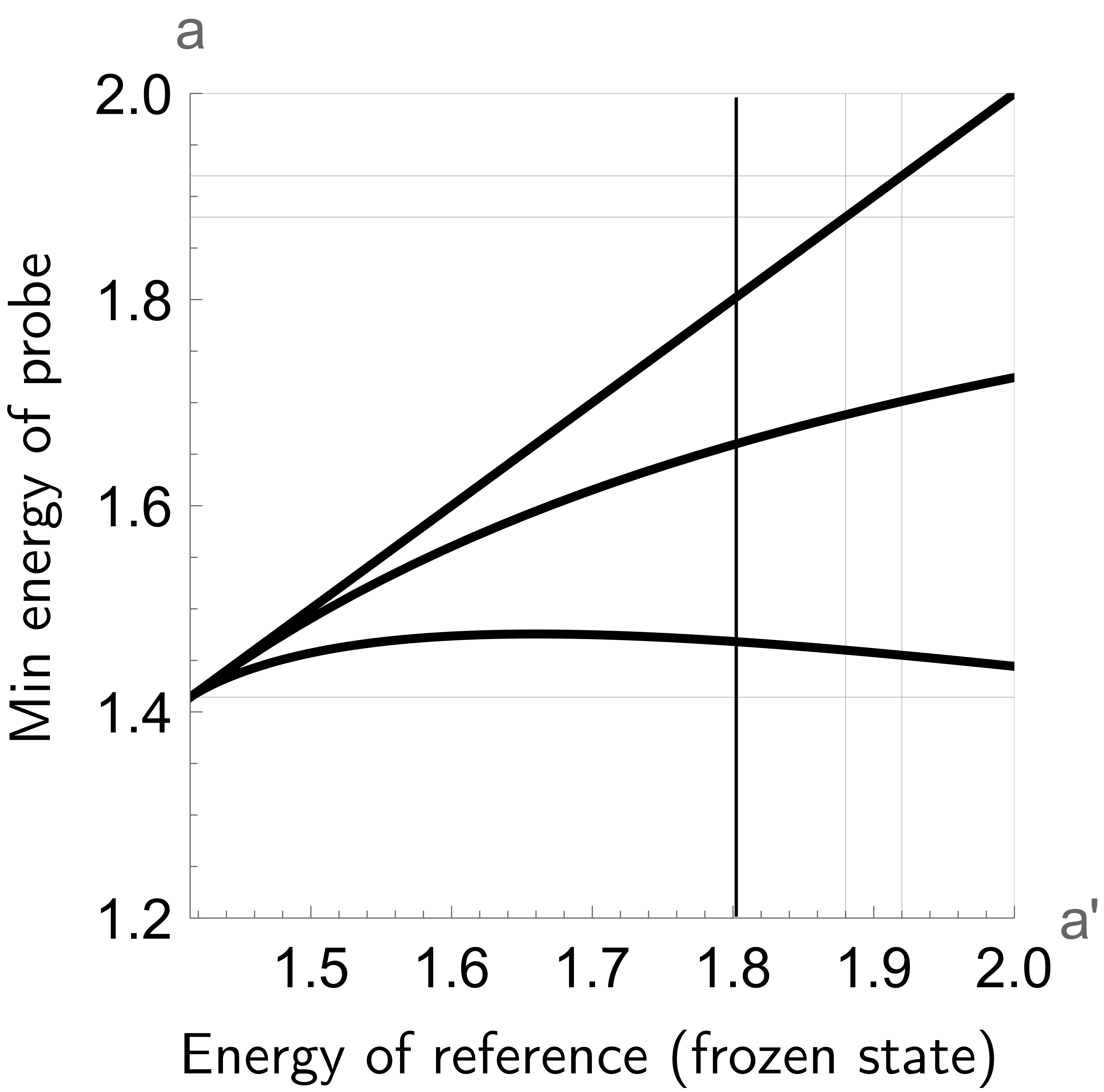
Fix one typical frozen state (energy a'). Compute entropy of probes at all energies $\geq a$ and at fixed overlap $q \implies$ planted cluster

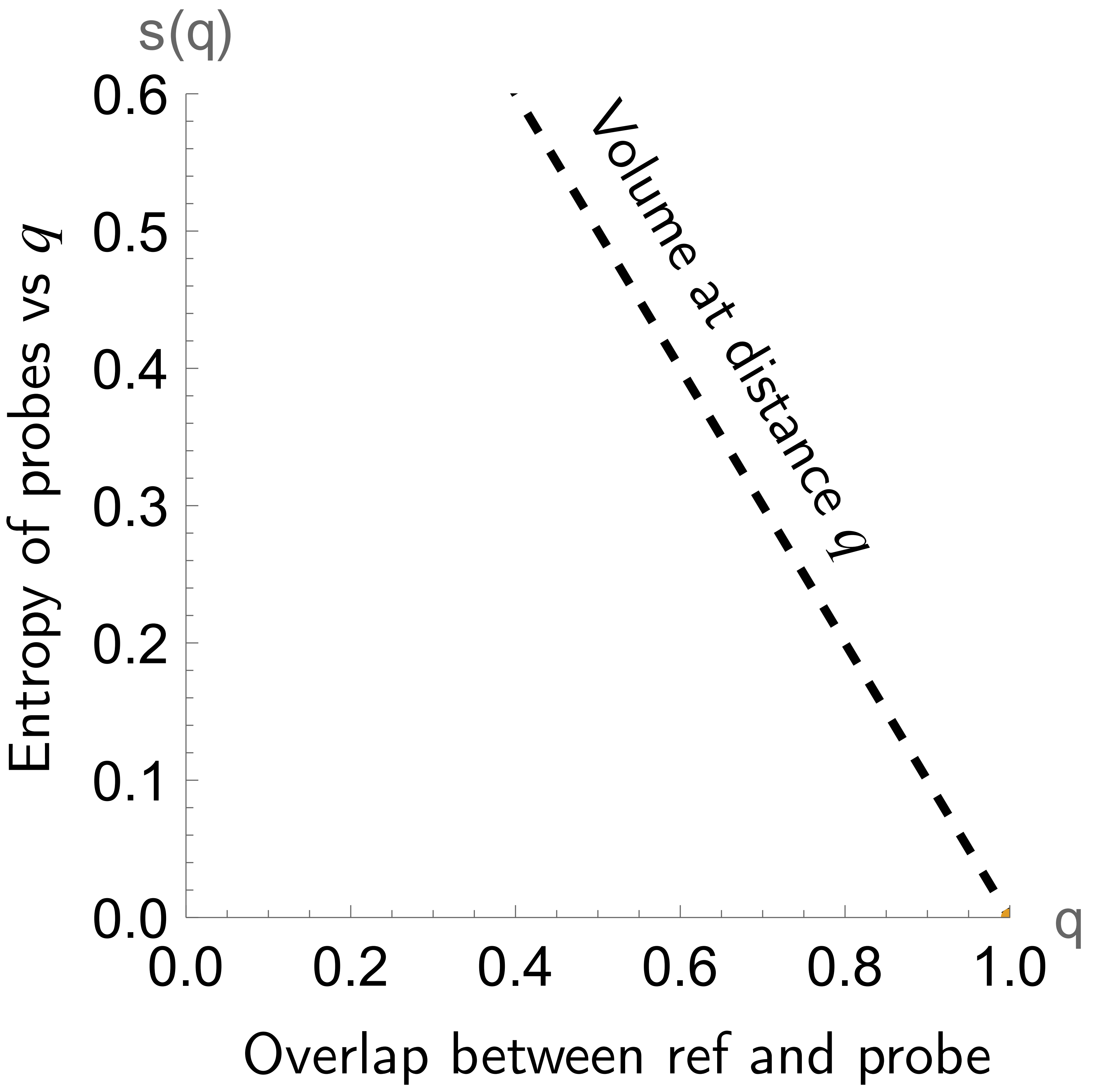


Result: Planted clusters in superlevel sets

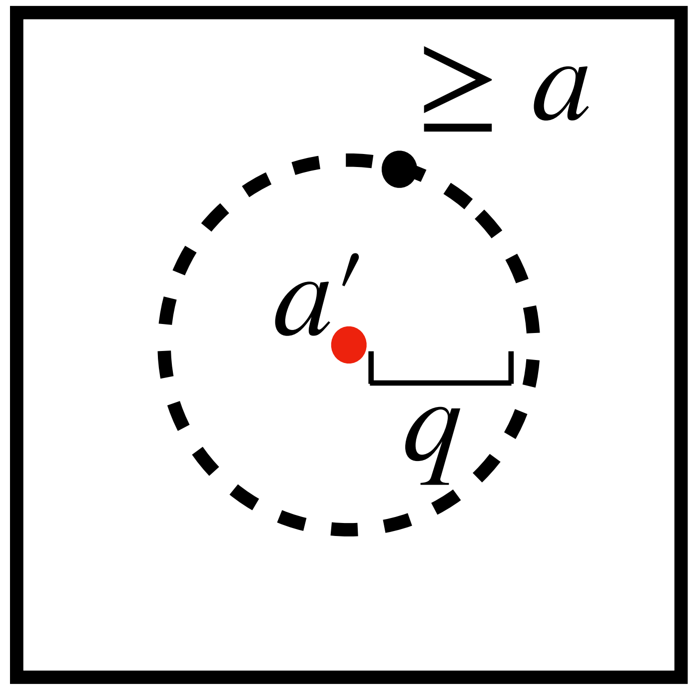


Fix one typical frozen state (energy a'). Compute entropy of probes at all energies $\geq a$ and at fixed overlap $q \Longrightarrow$ planted cluster

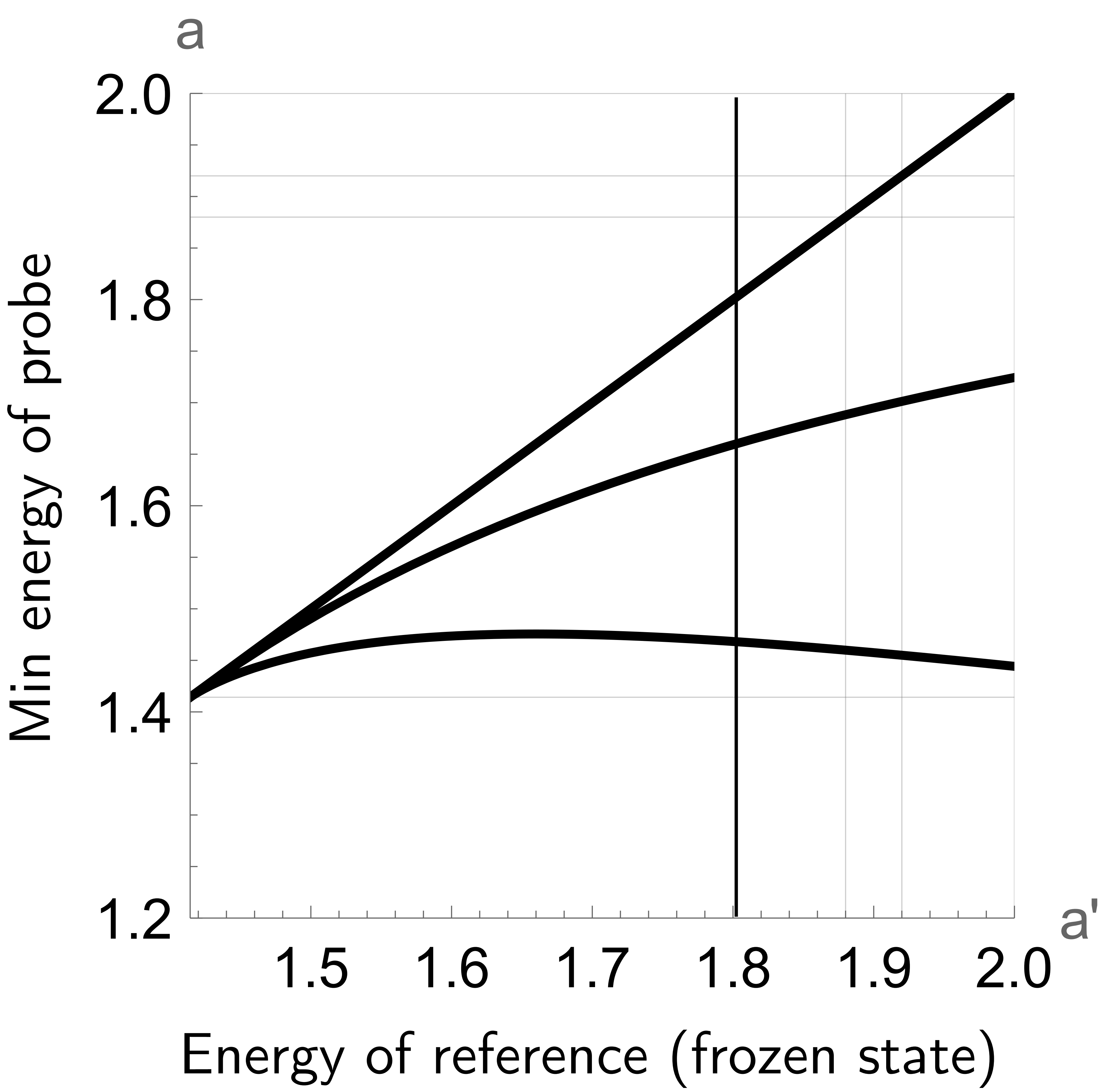


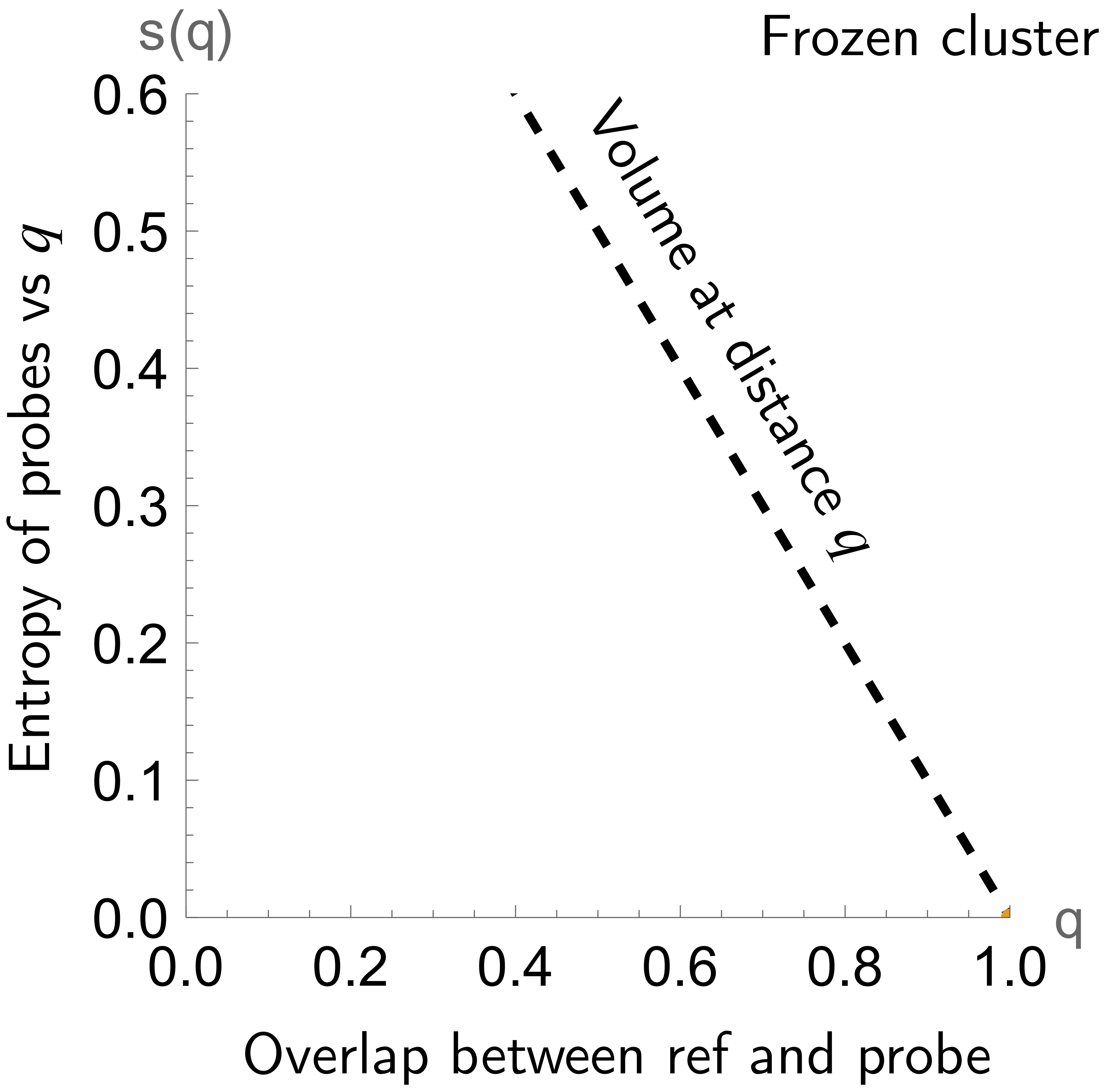


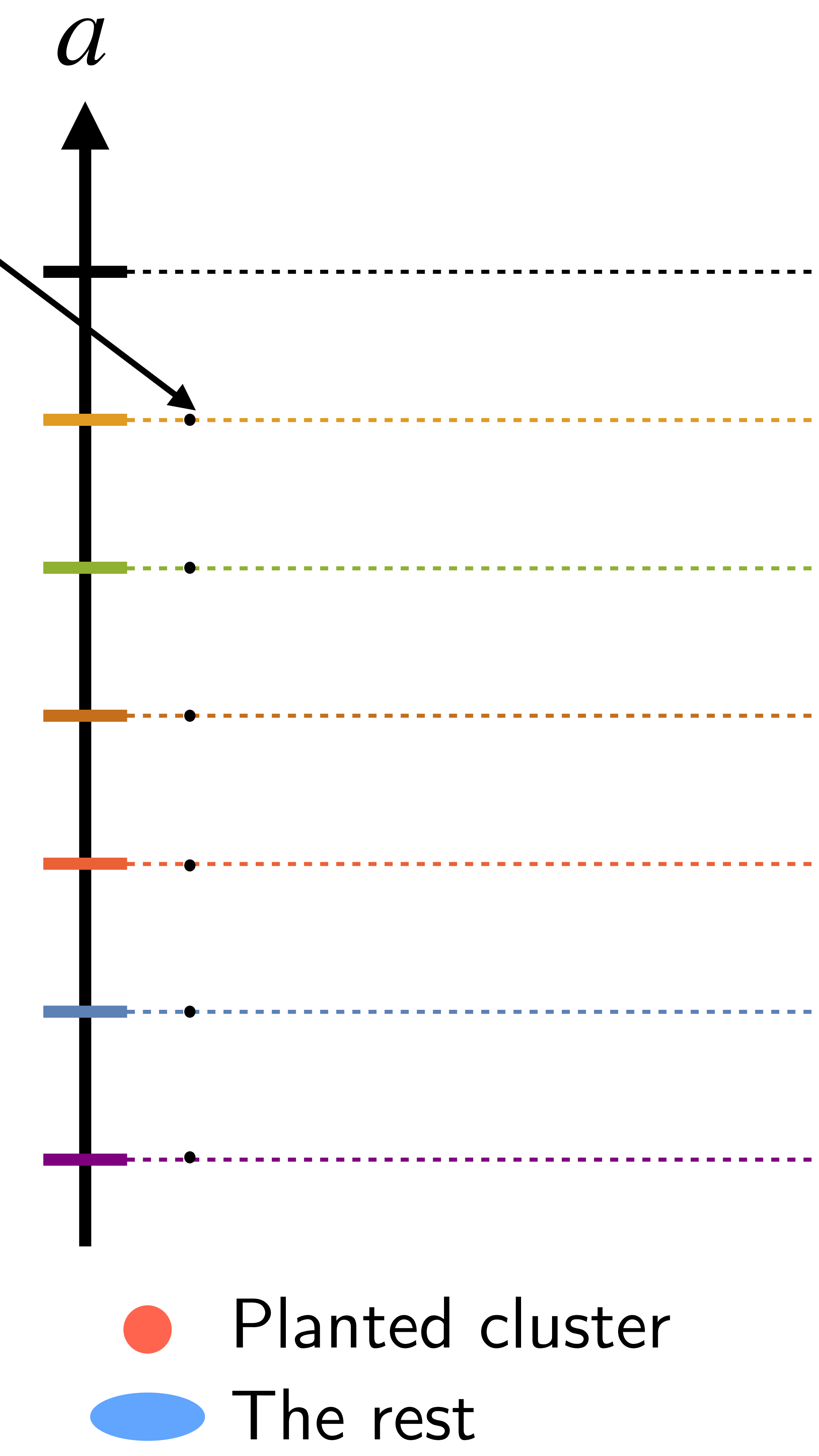
Result: Planted clusters in superlevel sets



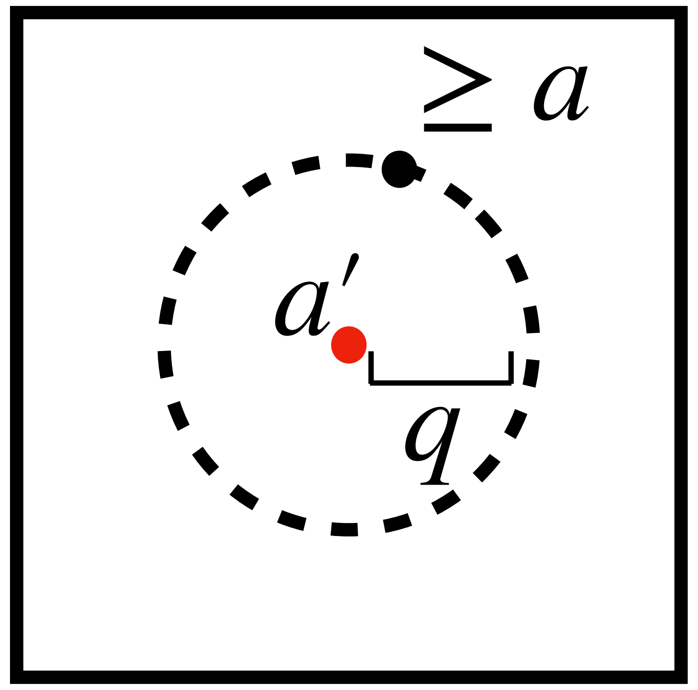
Fix one typical frozen state (energy a'). Compute entropy of probes at all energies $\geq a$ and at fixed overlap $q \Longrightarrow$ planted cluster



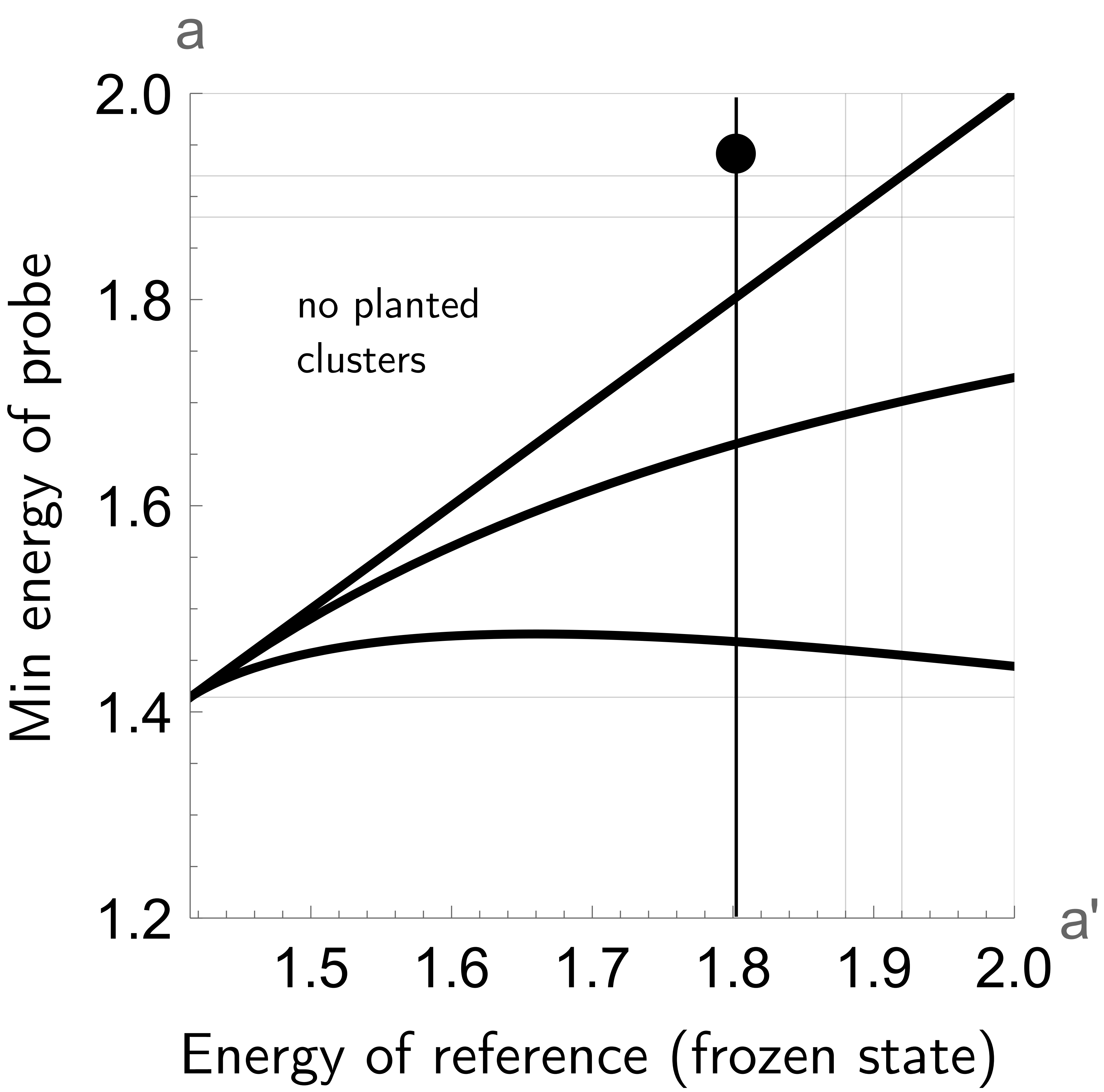


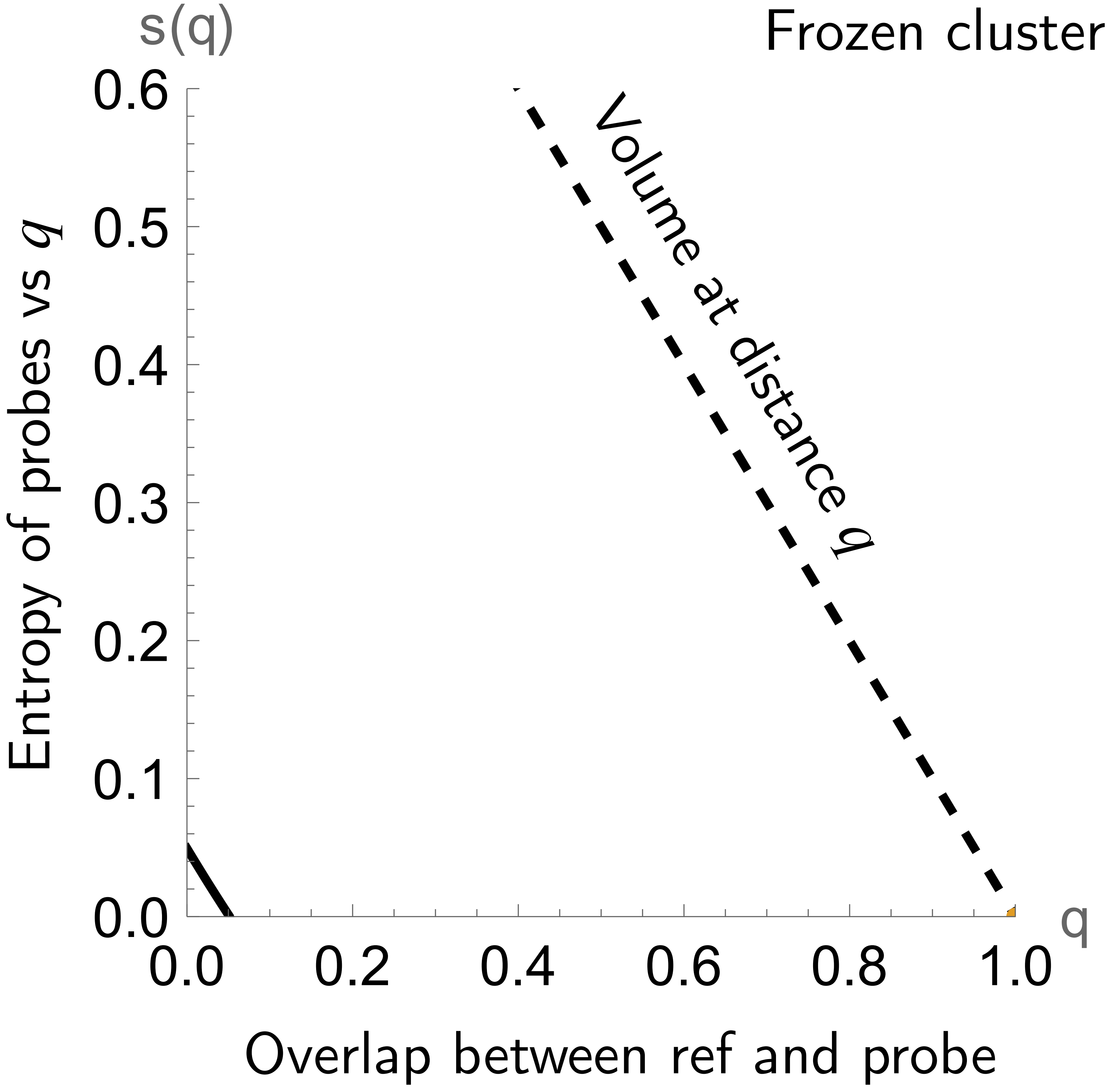


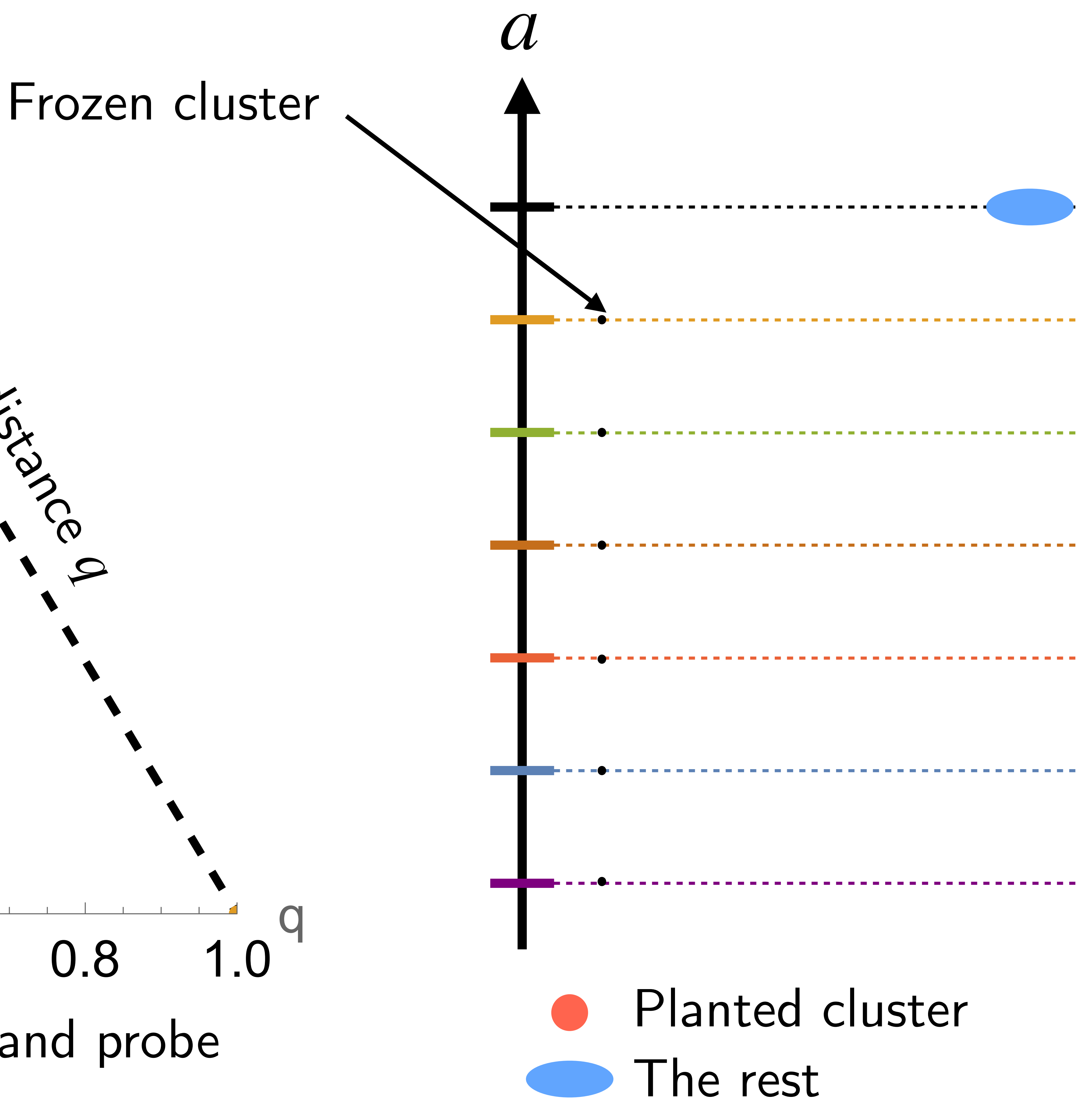
Result: Planted clusters in superlevel sets



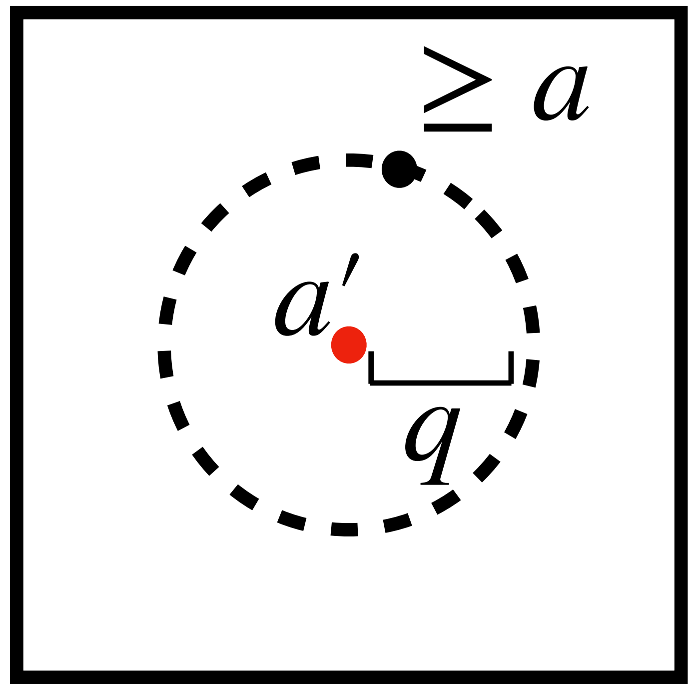
Fix one typical frozen state (energy a'). Compute entropy of probes at all energies $\geq a$ and at fixed overlap $q \implies$ planted cluster



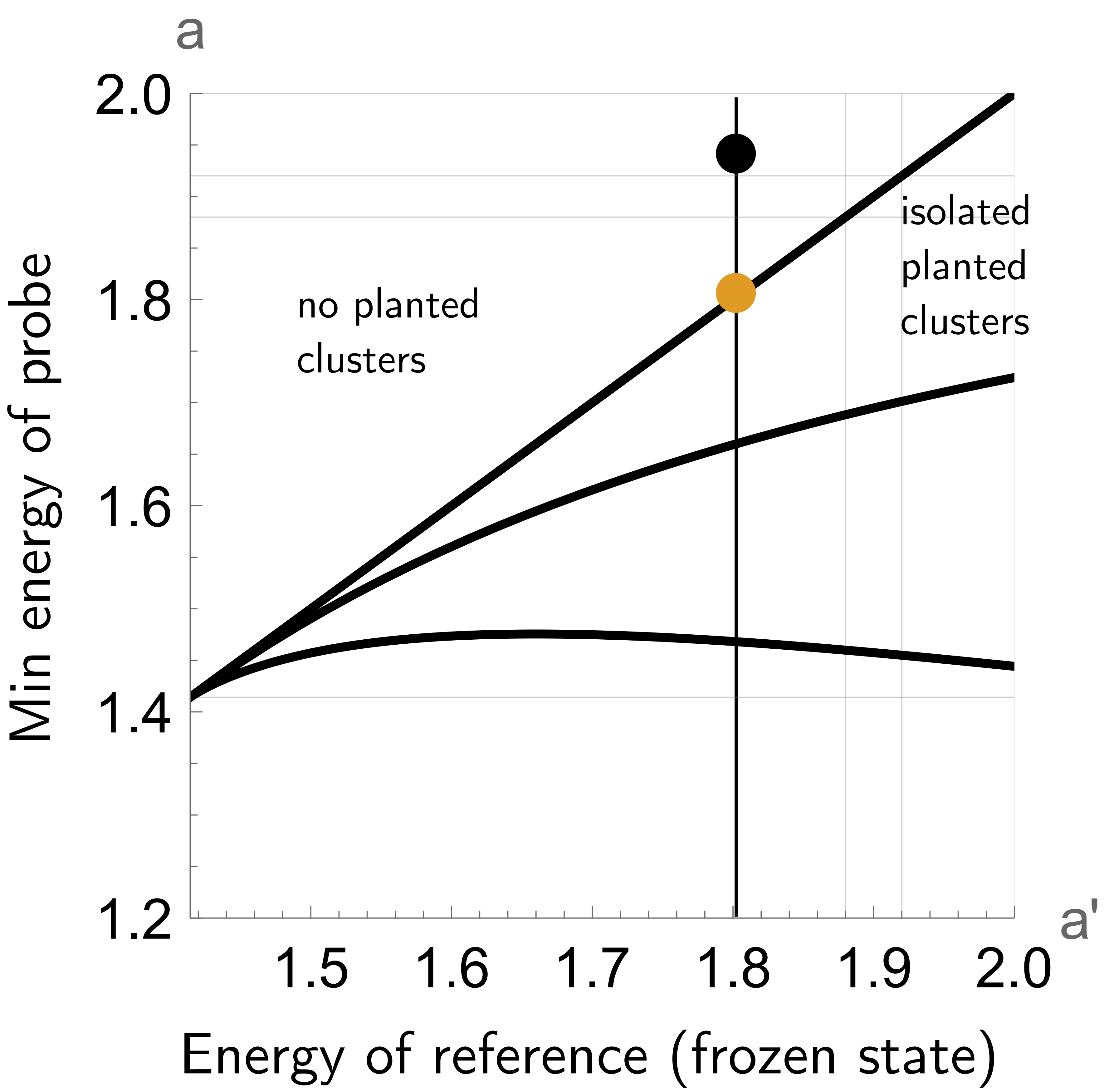


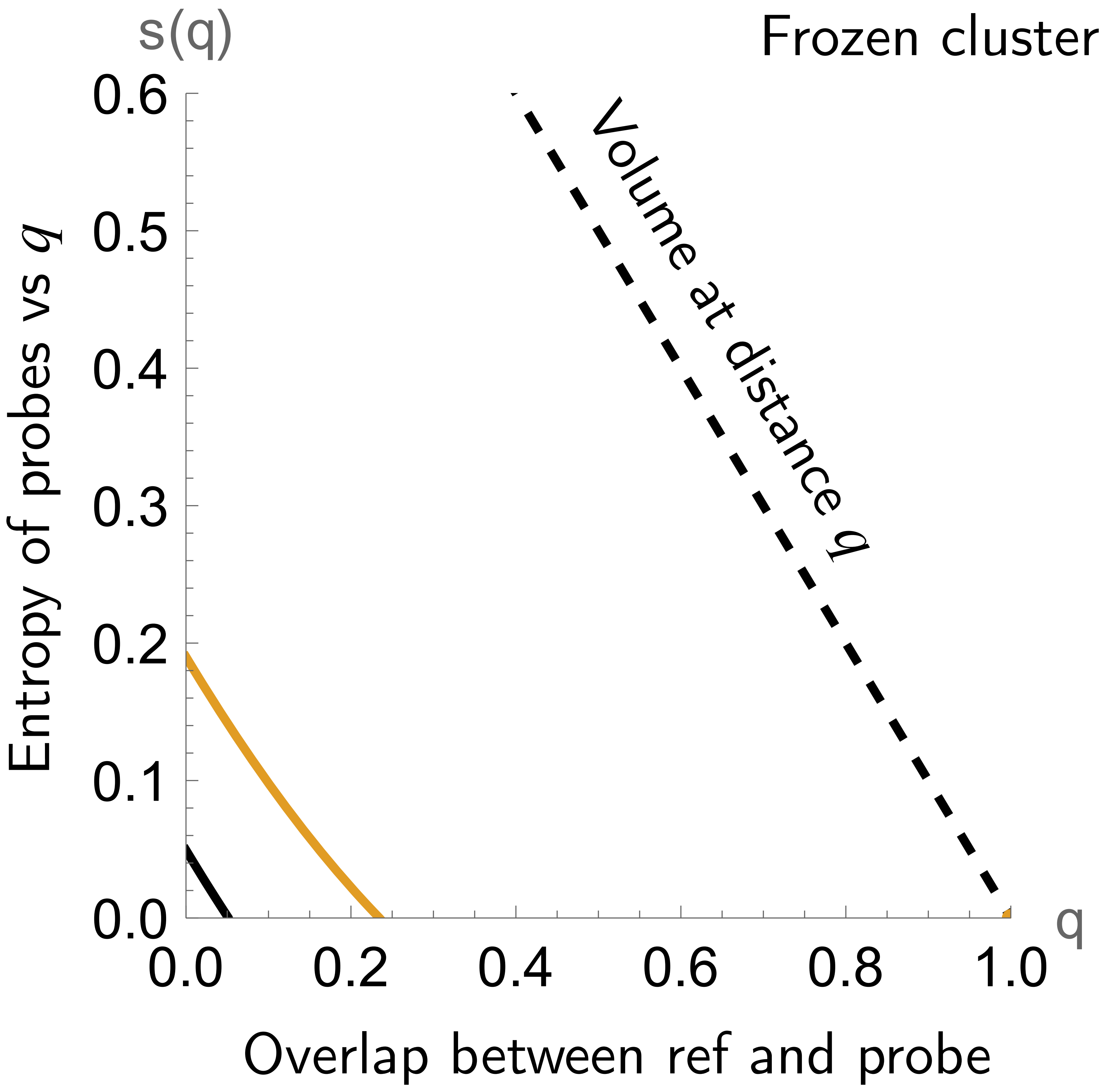


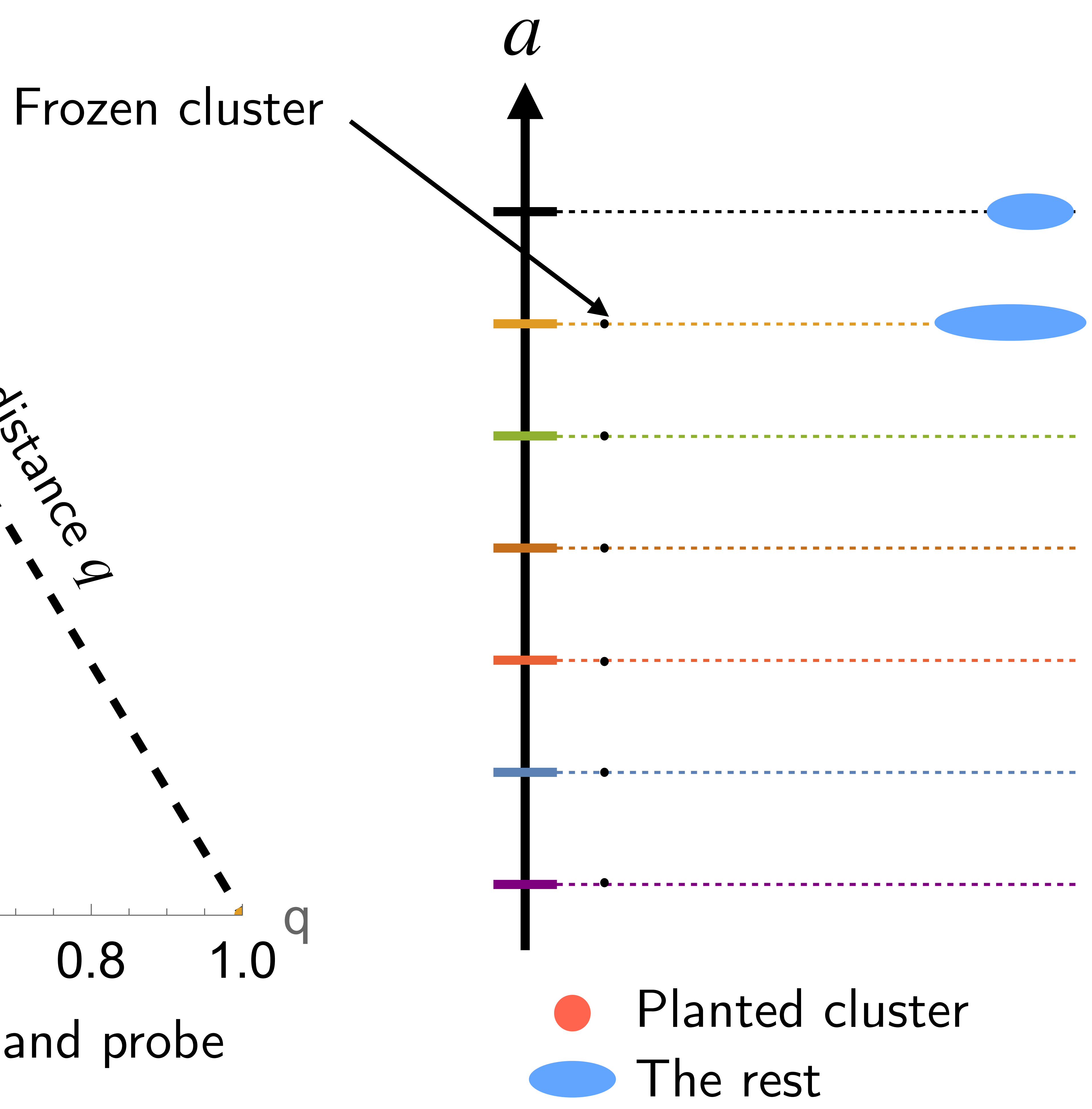
Result: Planted clusters in superlevel sets



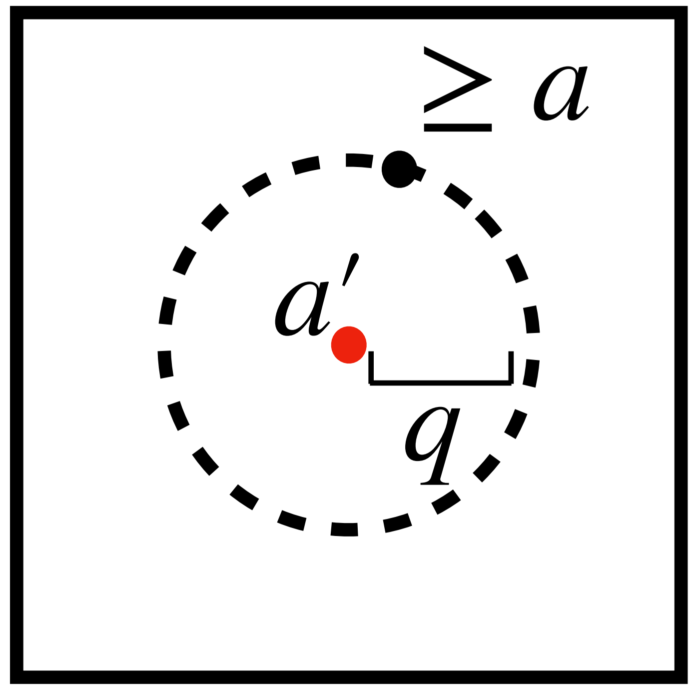
Fix one typical frozen state (energy a'). Compute entropy of probes at all energies $\geq a$ and at fixed overlap $q \implies$ planted cluster



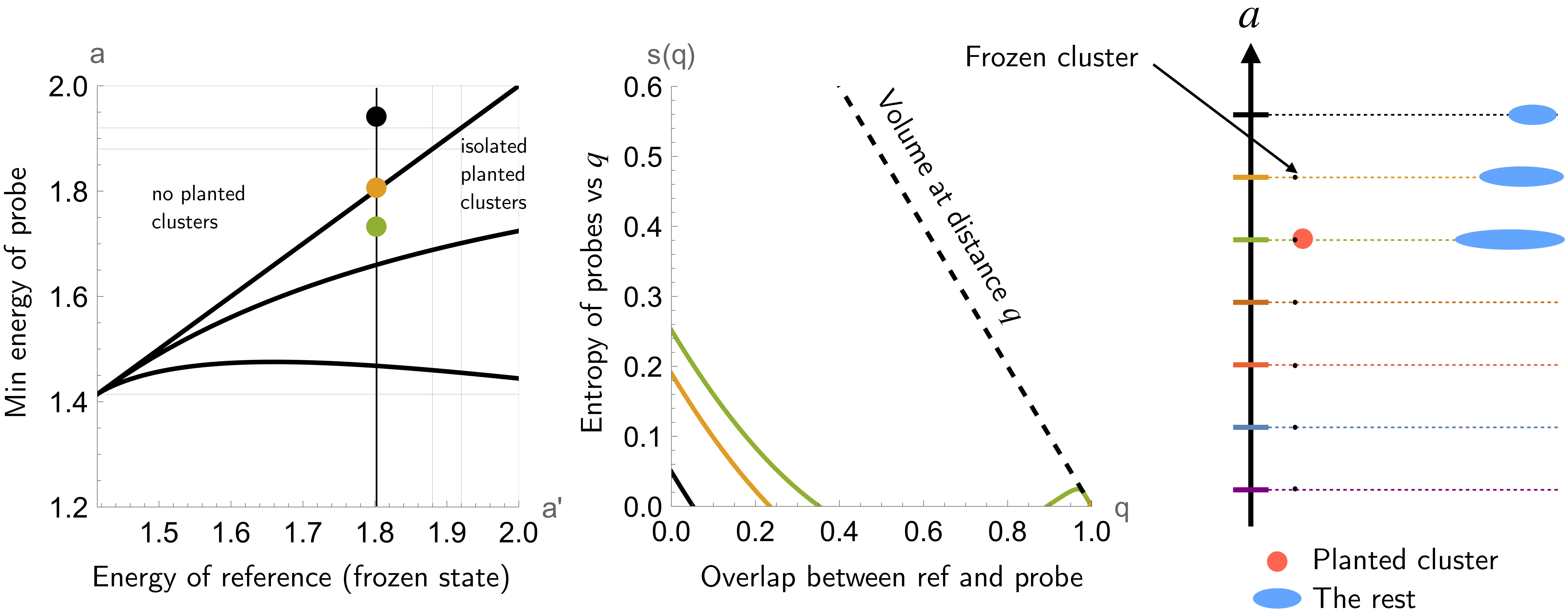




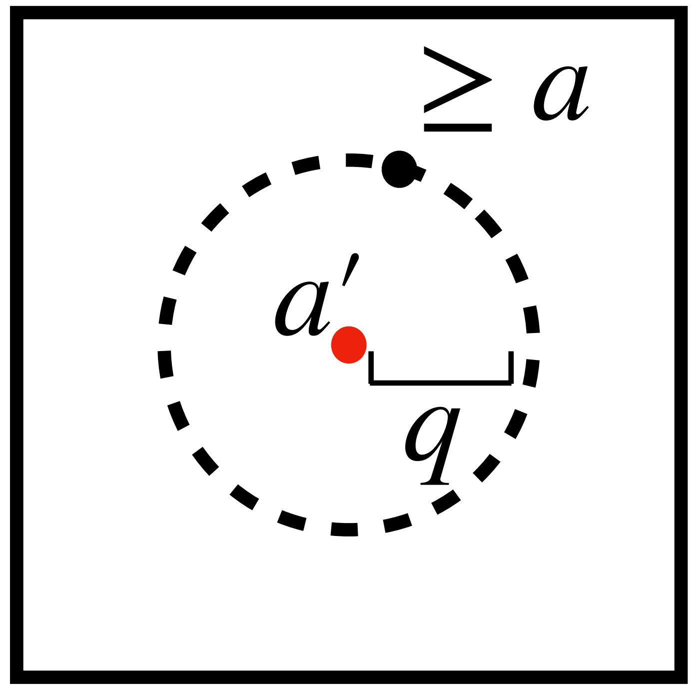
Result: Planted clusters in superlevel sets



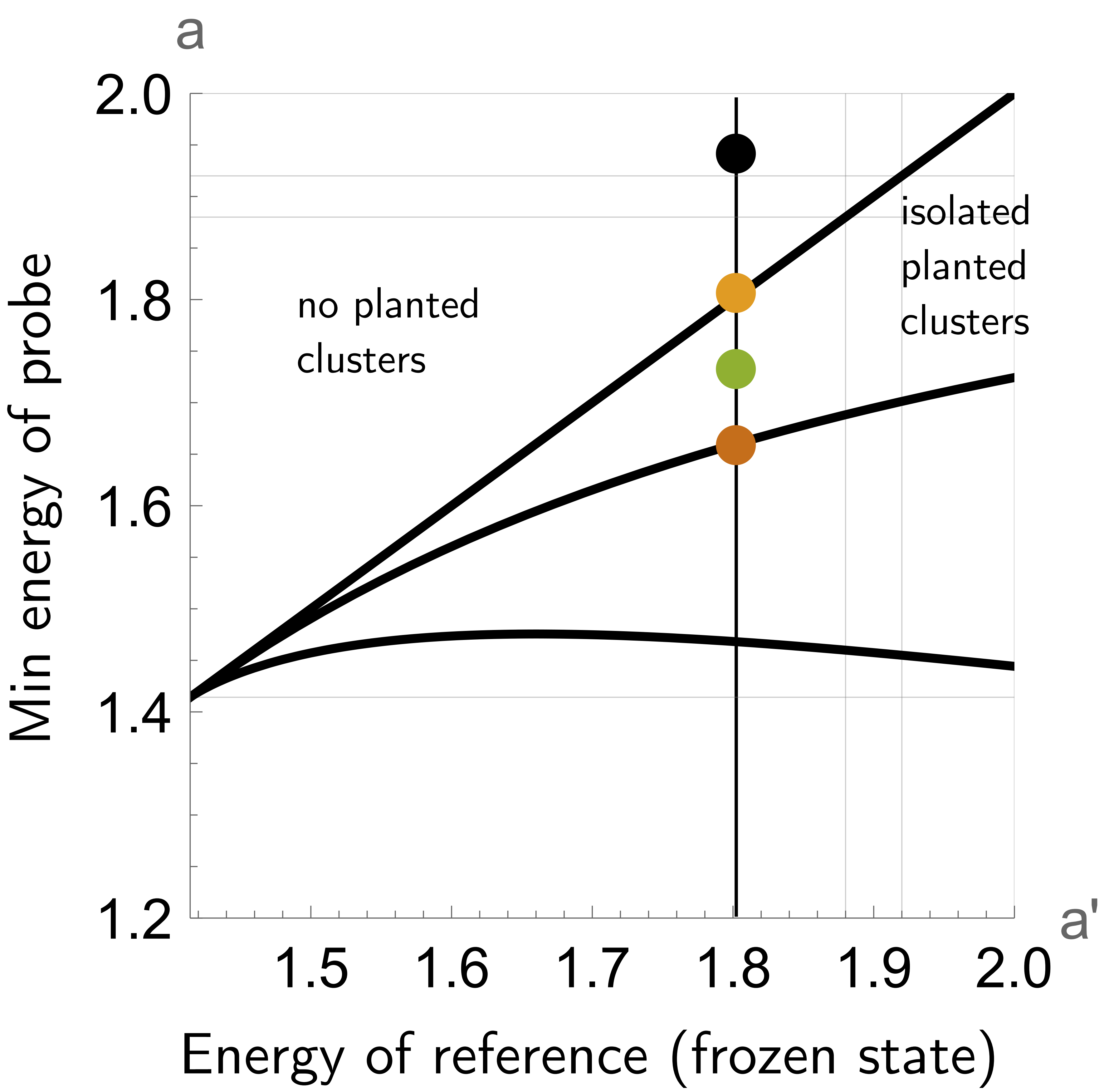
Fix one typical frozen state (energy a'). Compute entropy of probes at all energies $\geq a$ and at fixed overlap $q \implies$ planted cluster

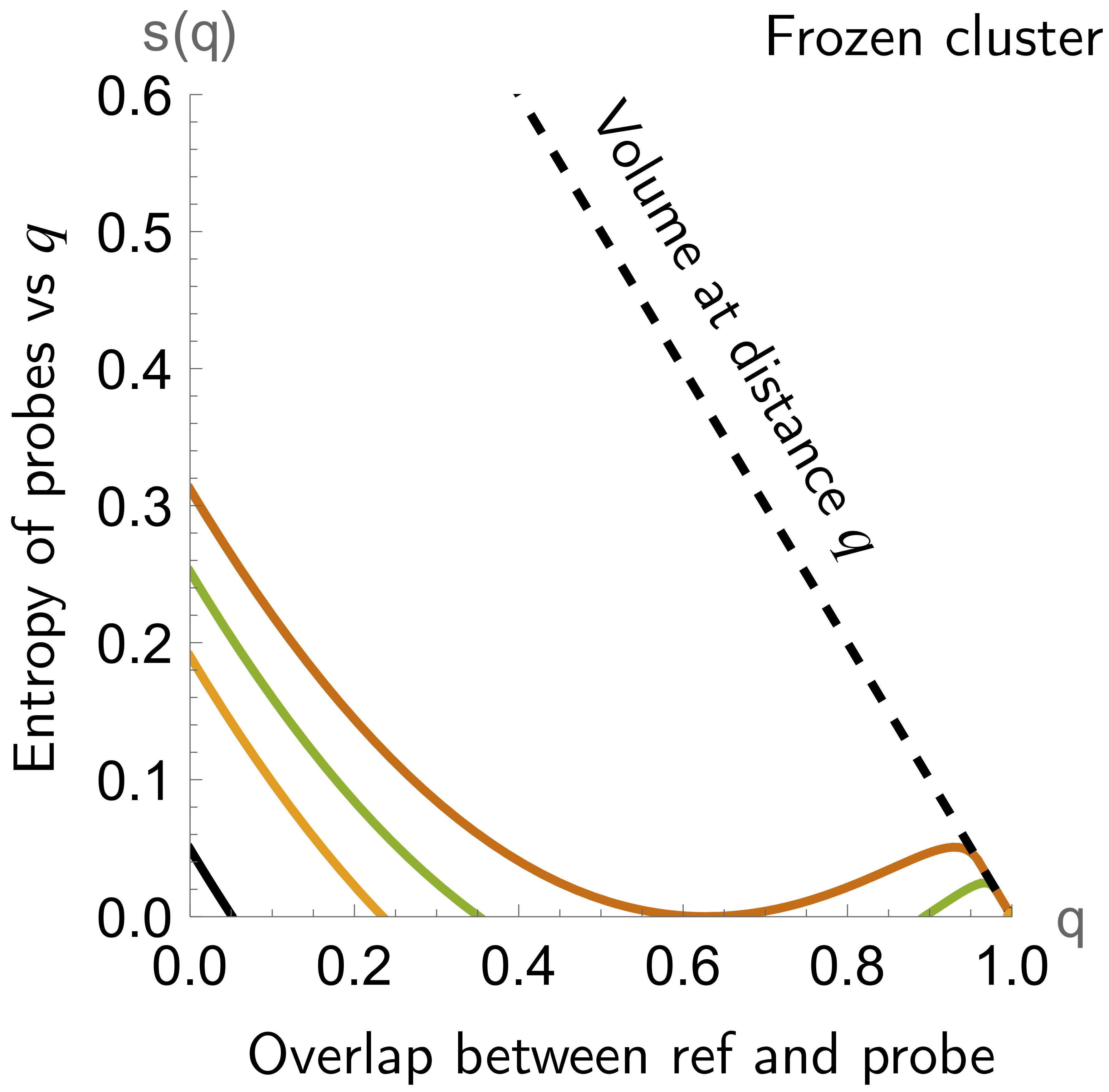


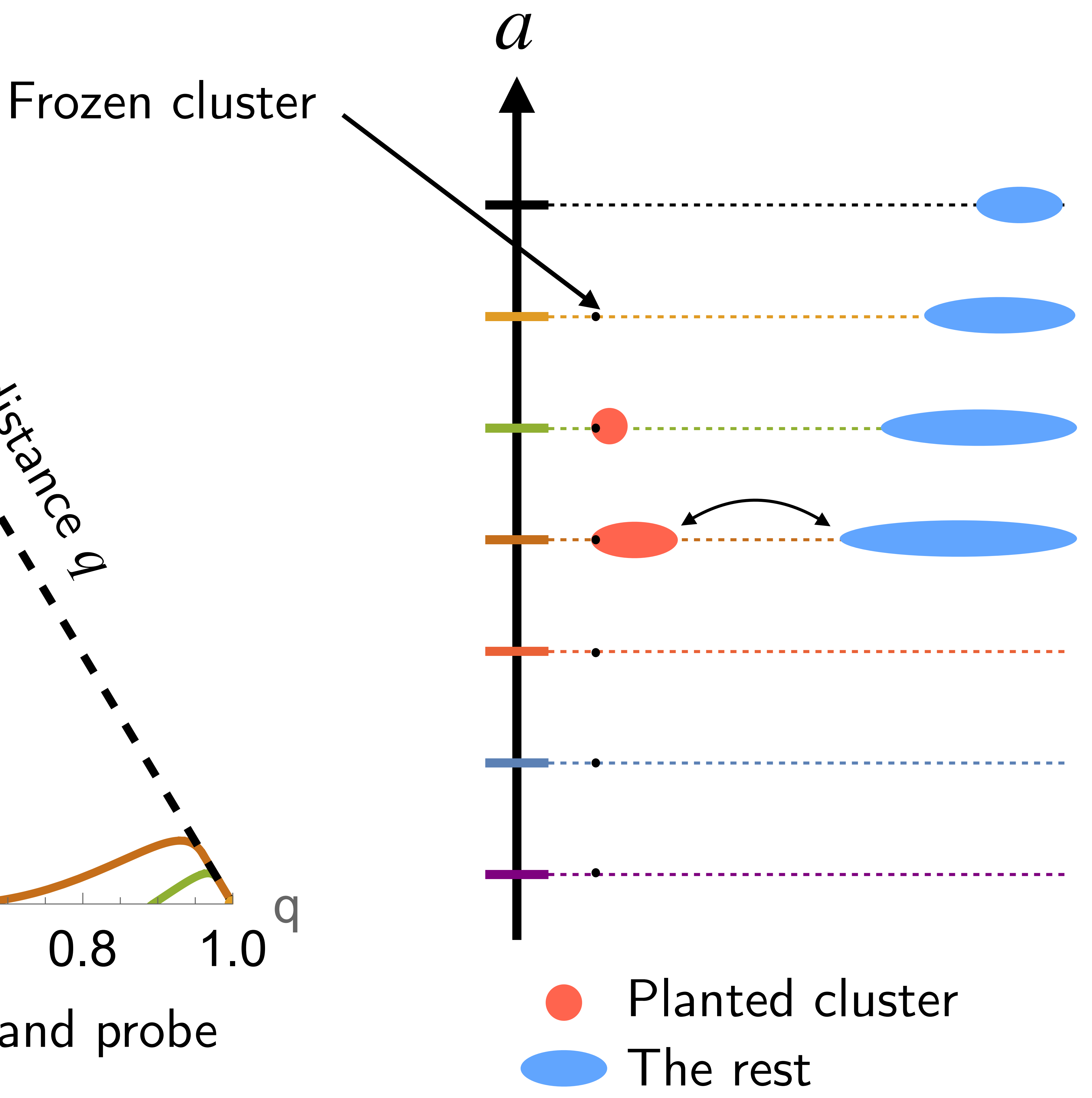
Result: Planted clusters in superlevel sets



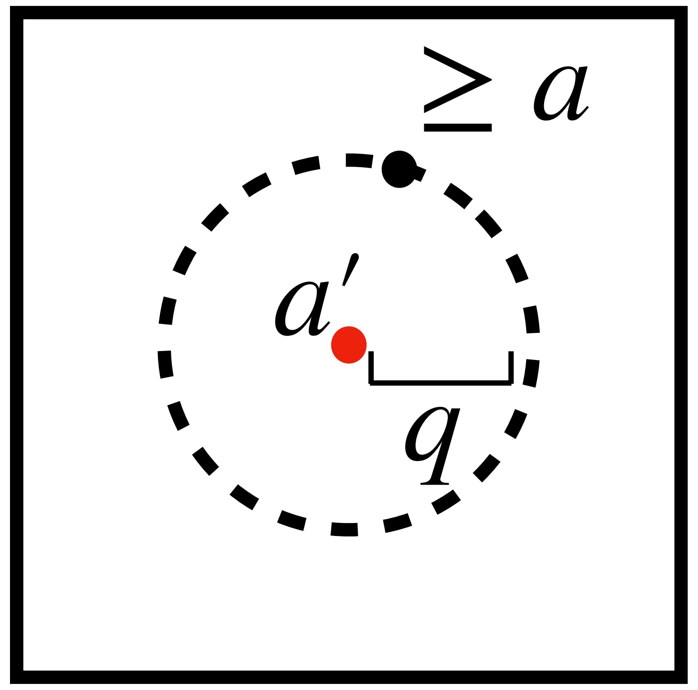
Fix one typical frozen state (energy a'). Compute entropy of probes at all energies $\geq a$ and at fixed overlap $q \implies$ planted cluster



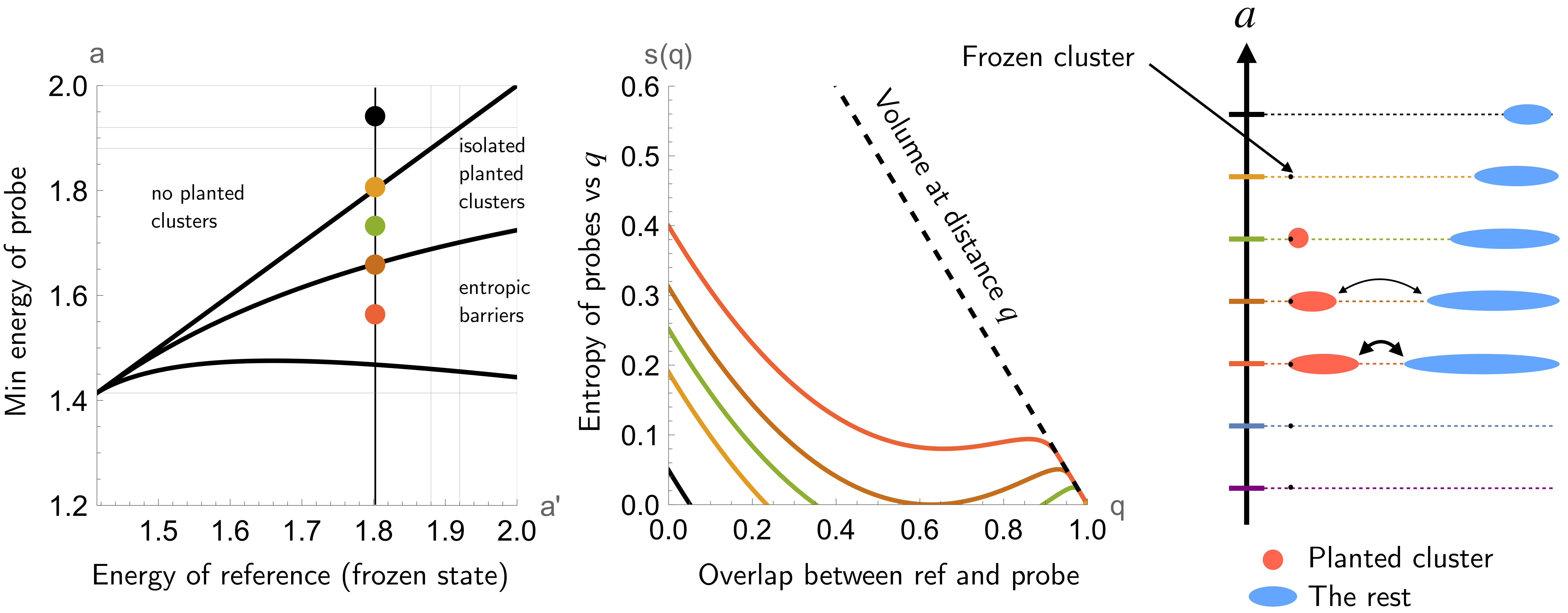




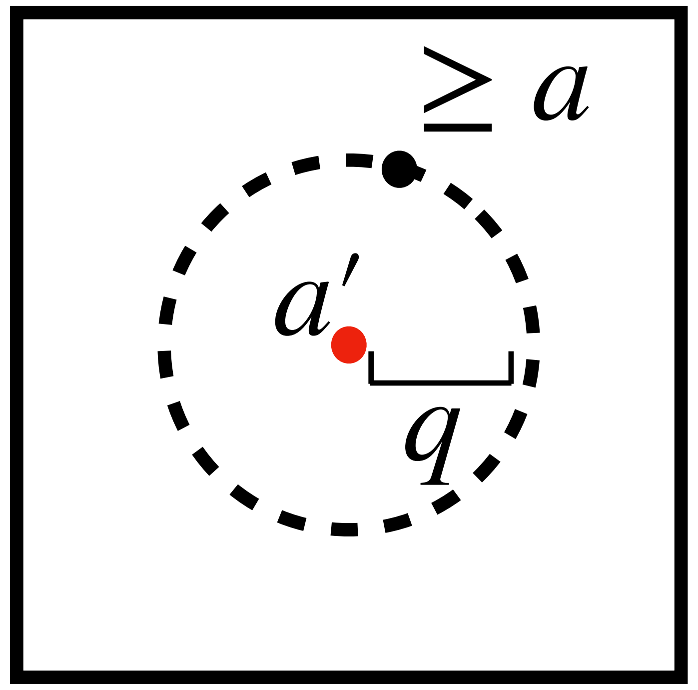
Result: Planted clusters in superlevel sets



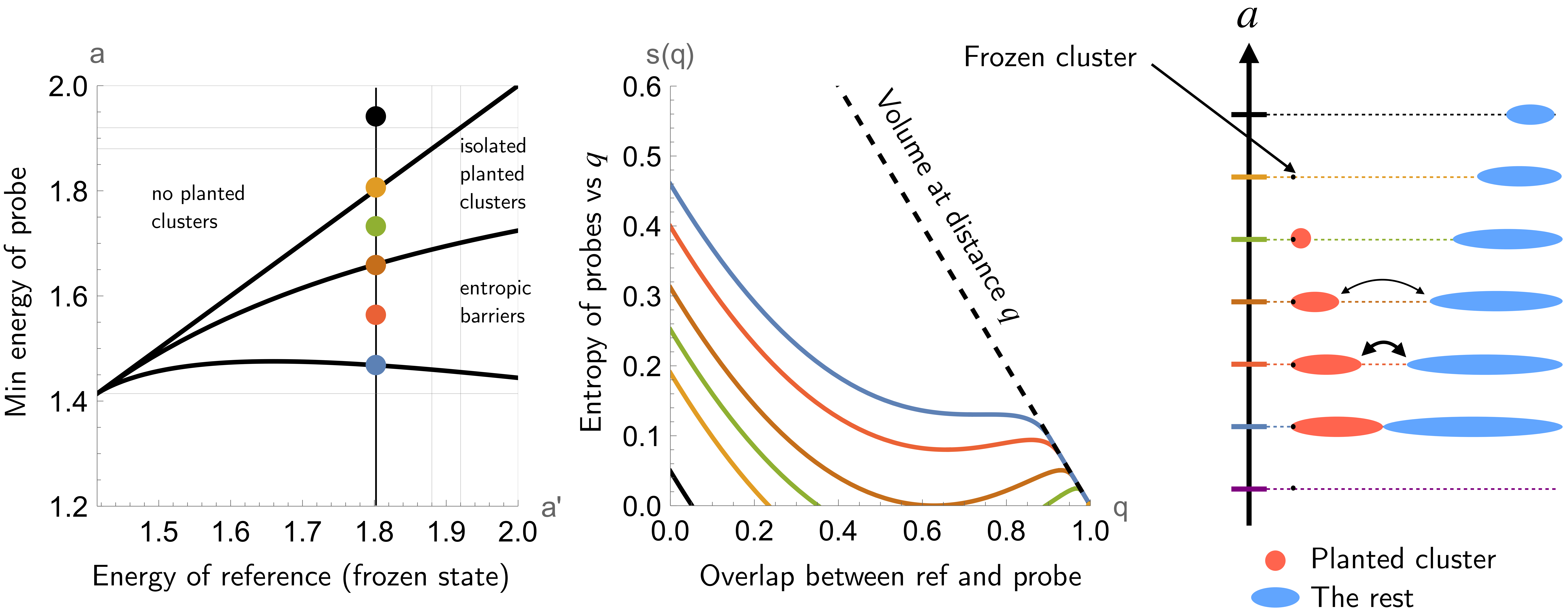
Fix one typical frozen state (energy a'). Compute entropy of probes at all energies $\geq a$ and at fixed overlap $q \implies$ planted cluster



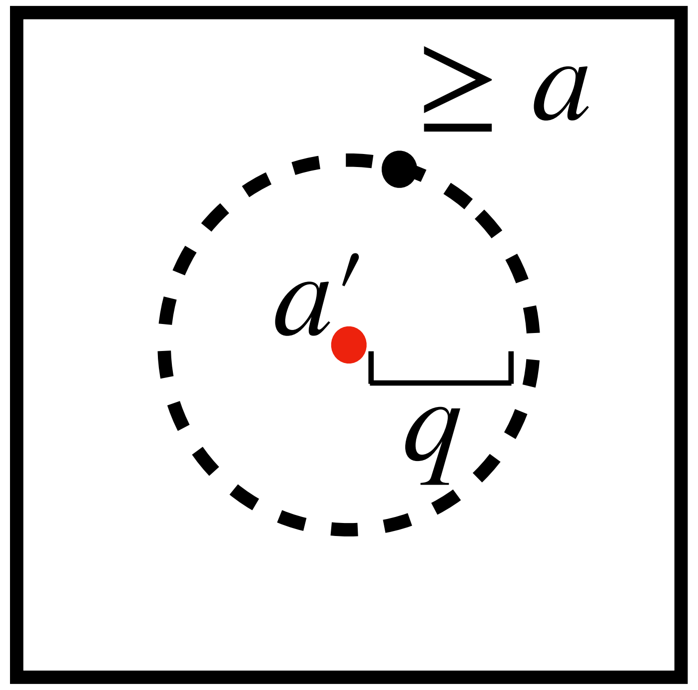
Result: Planted clusters in superlevel sets



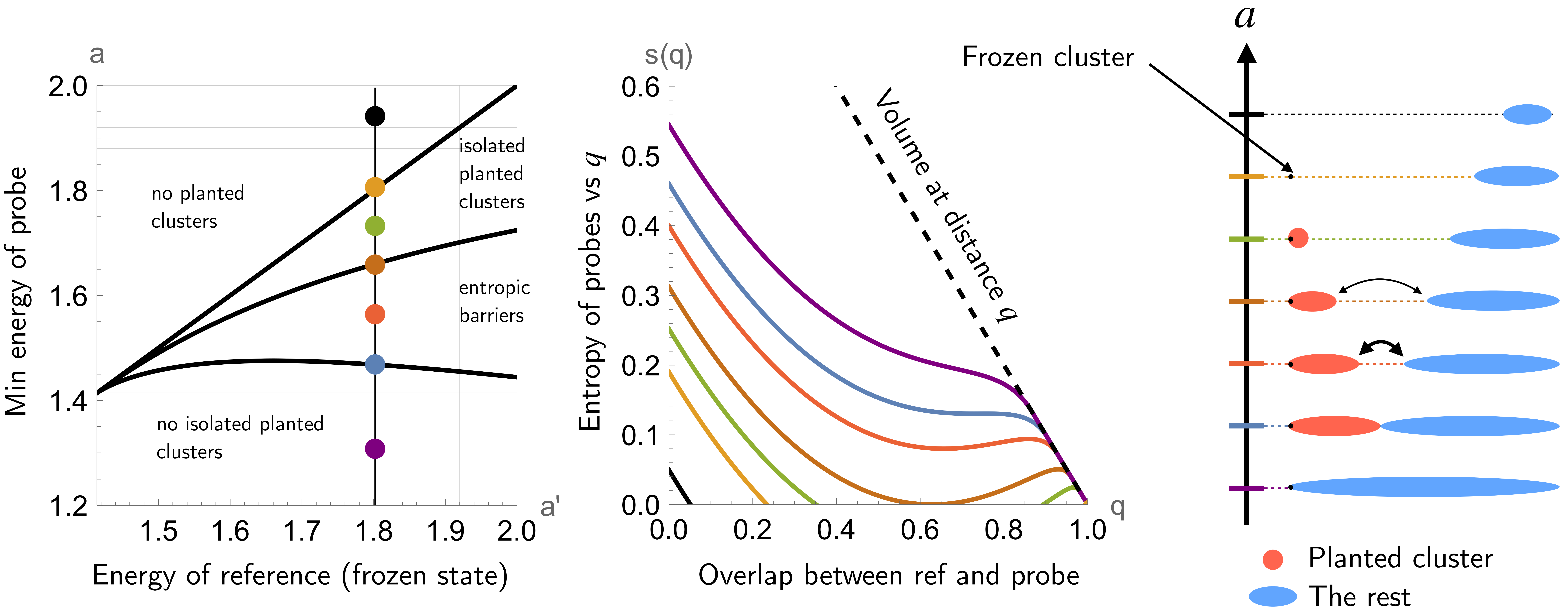
Fix one typical frozen state (energy a'). Compute entropy of probes at all energies $\geq a$ and at fixed overlap $q \implies$ planted cluster



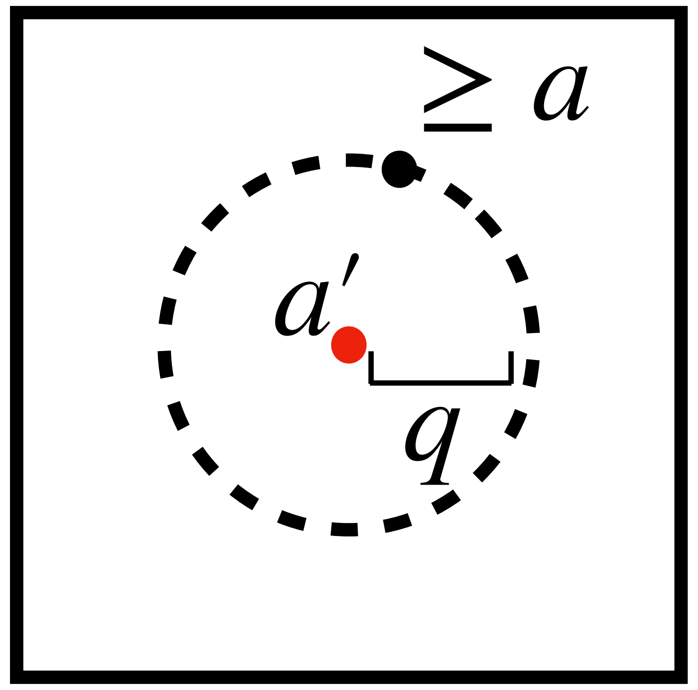
Result: Planted clusters in superlevel sets



Fix one typical frozen state (energy a'). Compute entropy of probes at all energies $\geq a$ and at fixed overlap $q \implies$ planted cluster

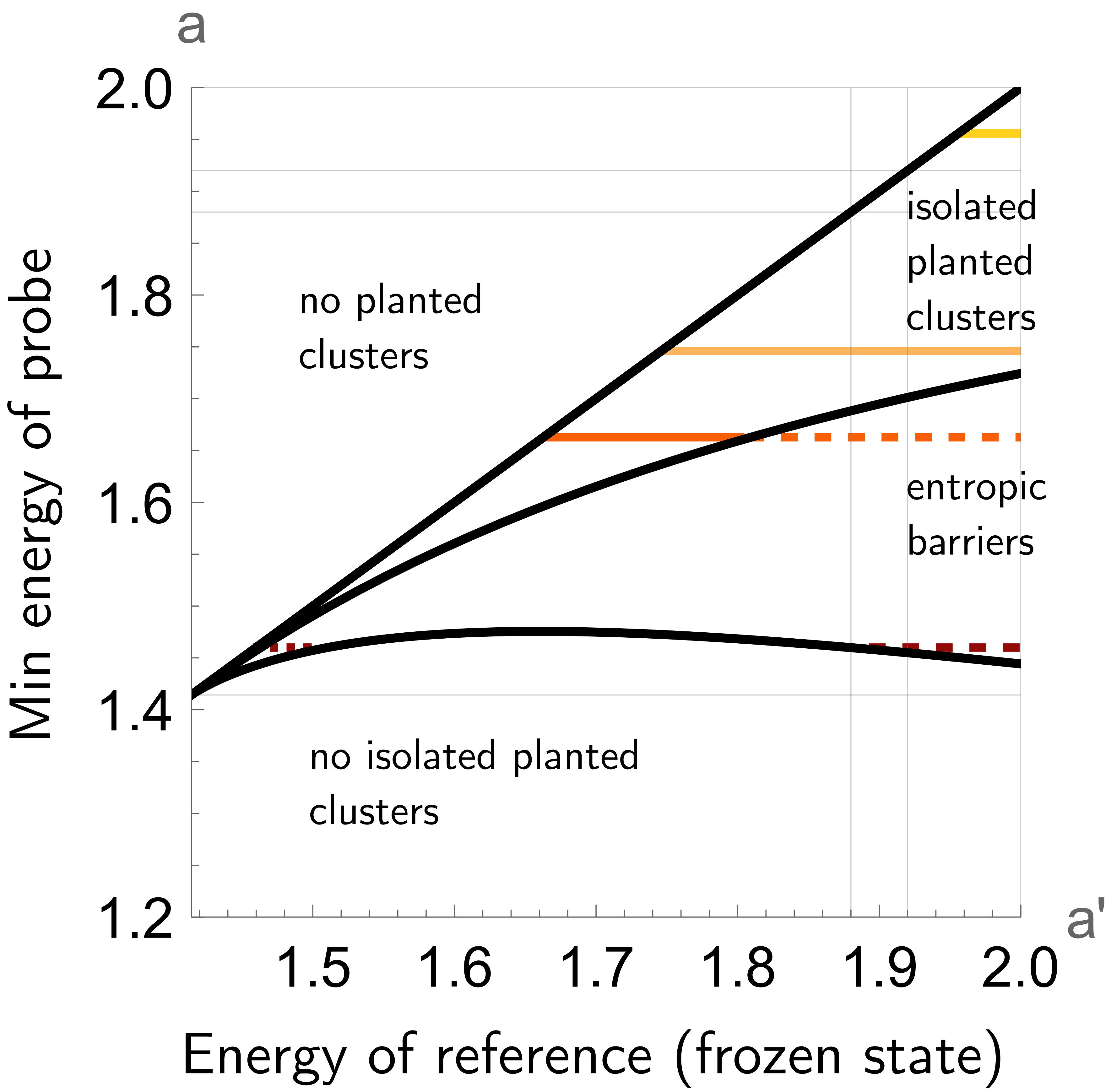


Result: Complexity vs Entropy

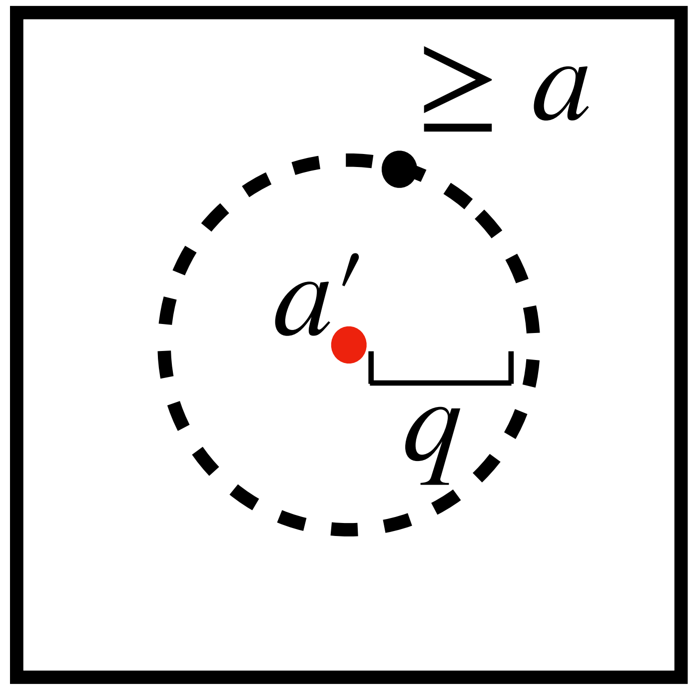


Fix energy super-level of probes (energy $\geq a$)

How many planted clusters are there in the super-level, and how large?

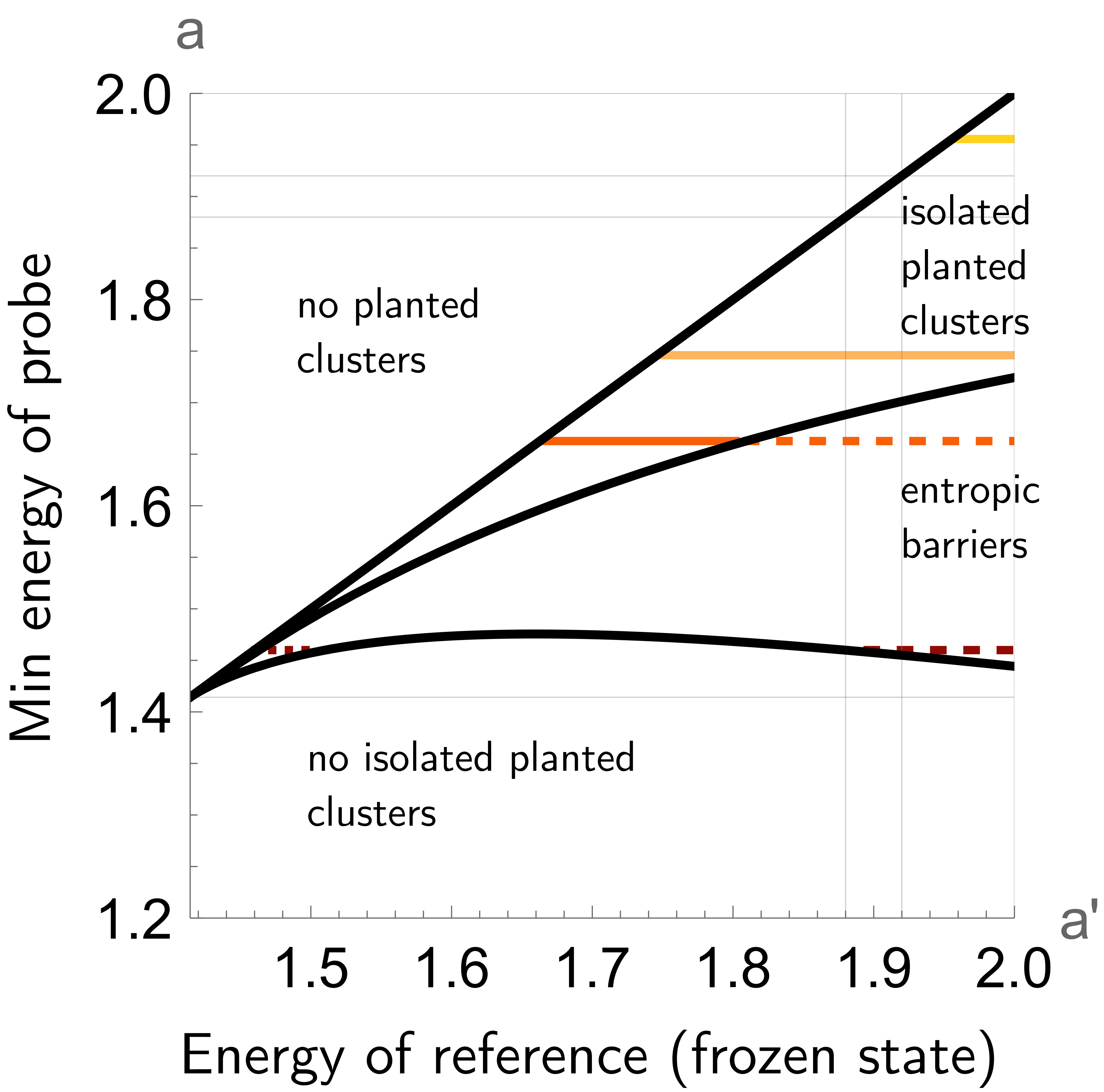


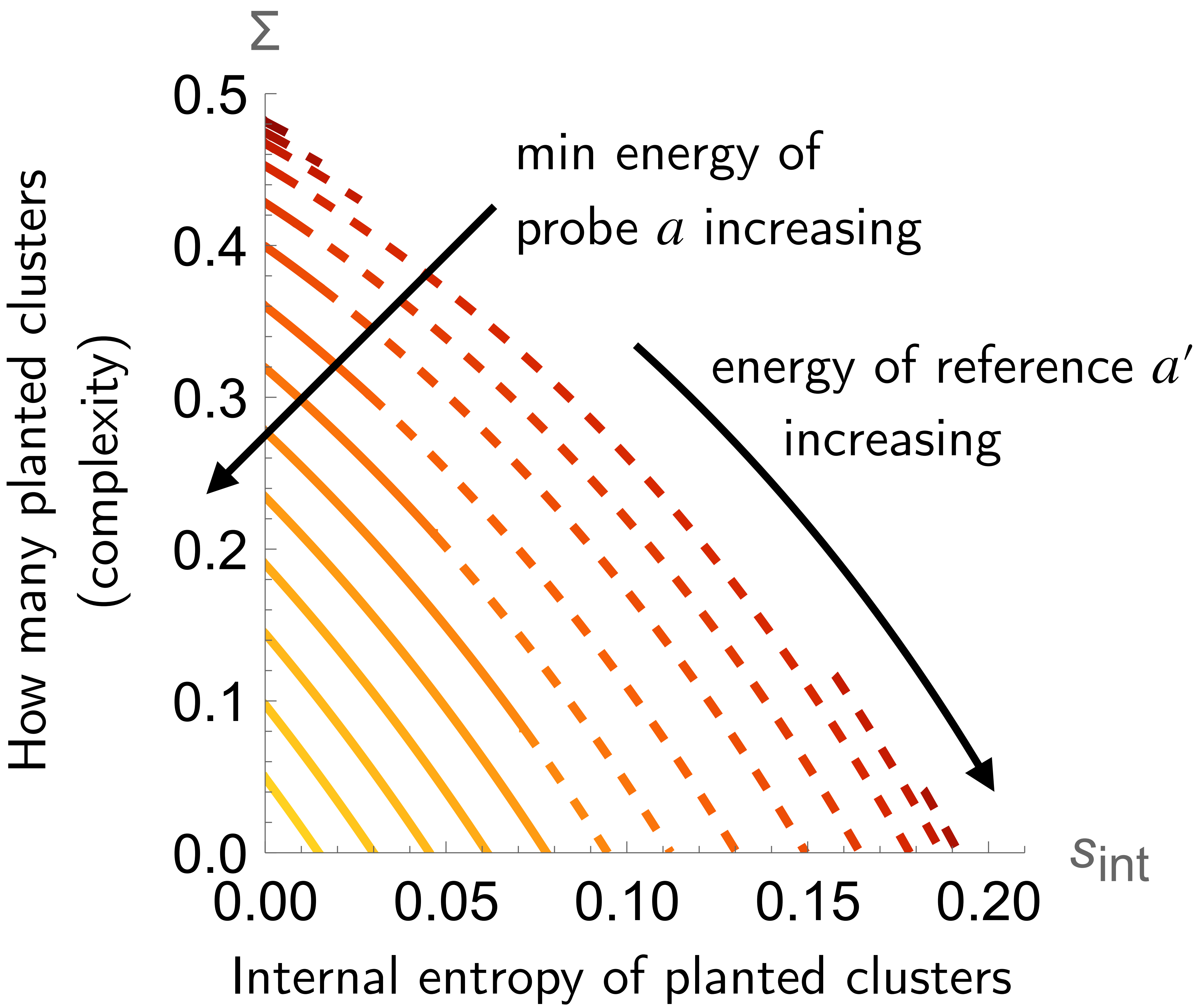
Result: Complexity vs Entropy



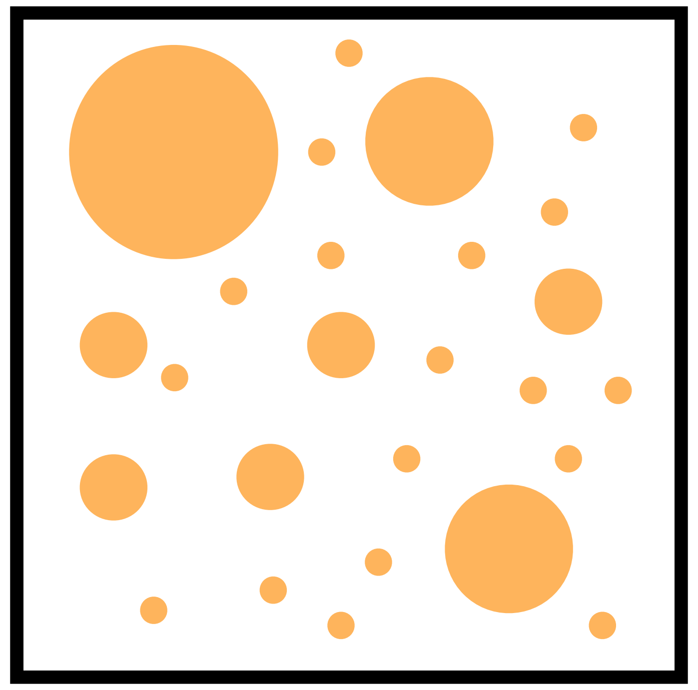
Fix energy super-level of probes (energy $\geq a$)

How many planted clusters are there in the super-level, and how large?

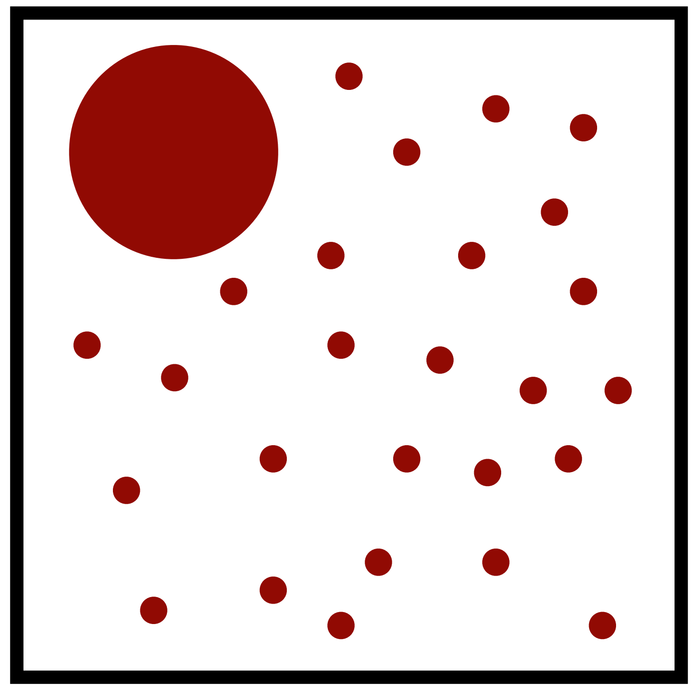




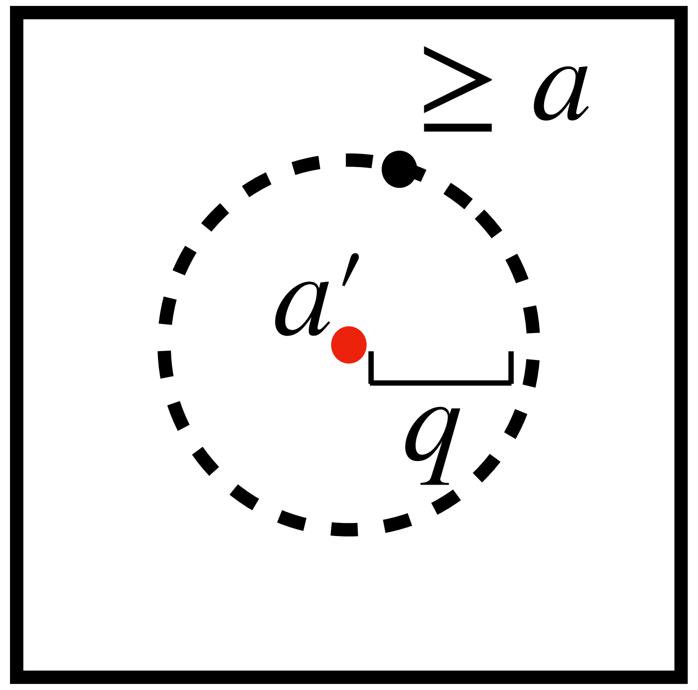
a close to global max

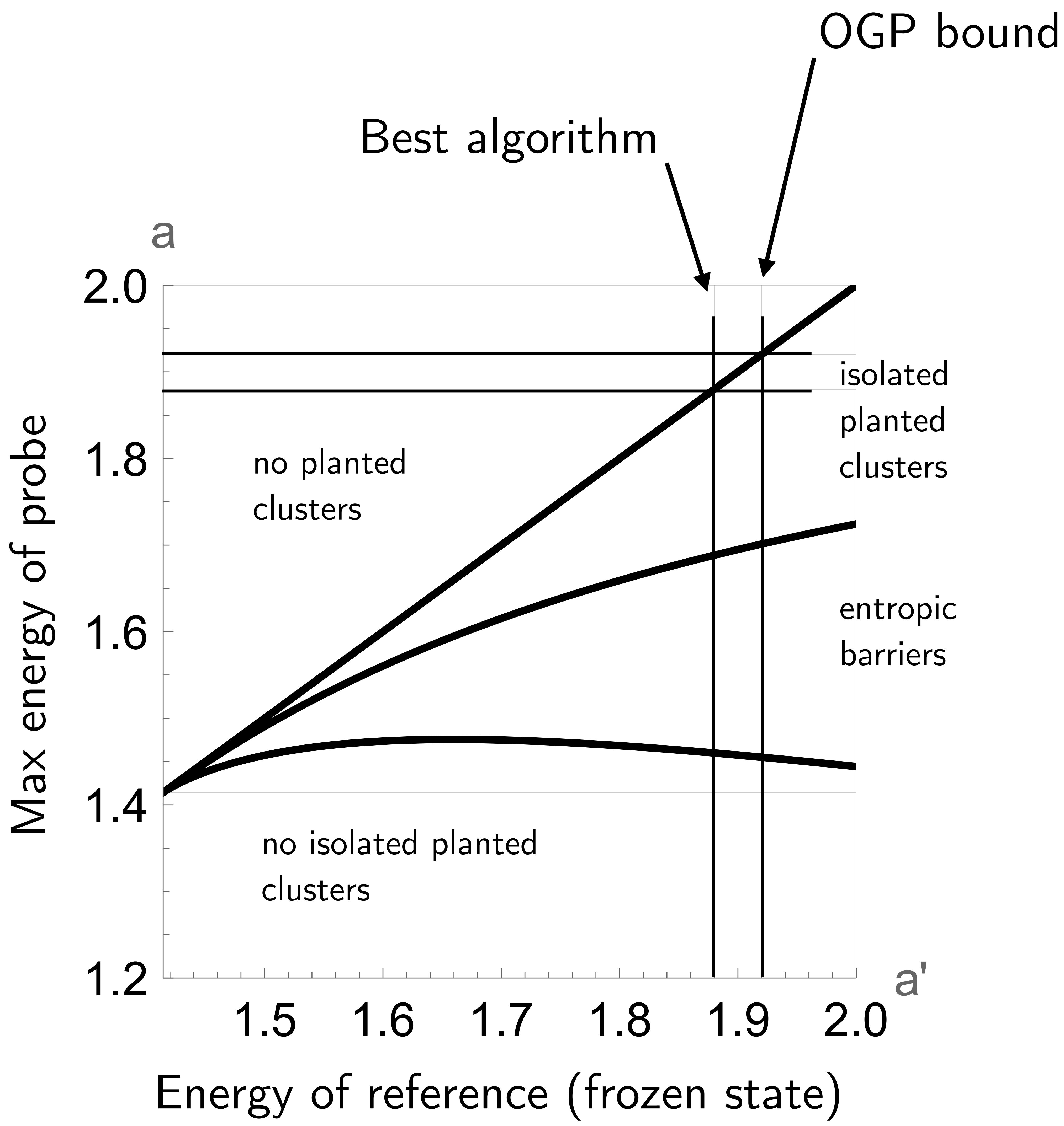


a close to freezing



A first look at non-equilibrium properties





There seems to be **no correlation** between planted clusters and algorithmic/OGP thresholds!

We learn structural informations on level sets + that planted clusters are not responsible for algorithmic properties

Only looks at surroundings of frozen clusters! Other atypical clusters could still exists

Results and open questions

N matrix size
k submatrix size
$m = k/N$

Characterised phase diagram in large ($0 < m < 1$) and small ($m \to 0$) submatrix regimes

Found frozen 1-RSB phase where efficient algorithms provably work for $m \to 0$

Studied planted clusters around frozen configurations for $m \to 0$
Found no link with algorithmic properties

Technical: very simple and explicit formulas (SP equations solved) for $m \to 0$!!

Thank you!

Paper on arXiv:2303.05237