FLATIRON INSTITUTE

APPLE Picking: Particles without Templates

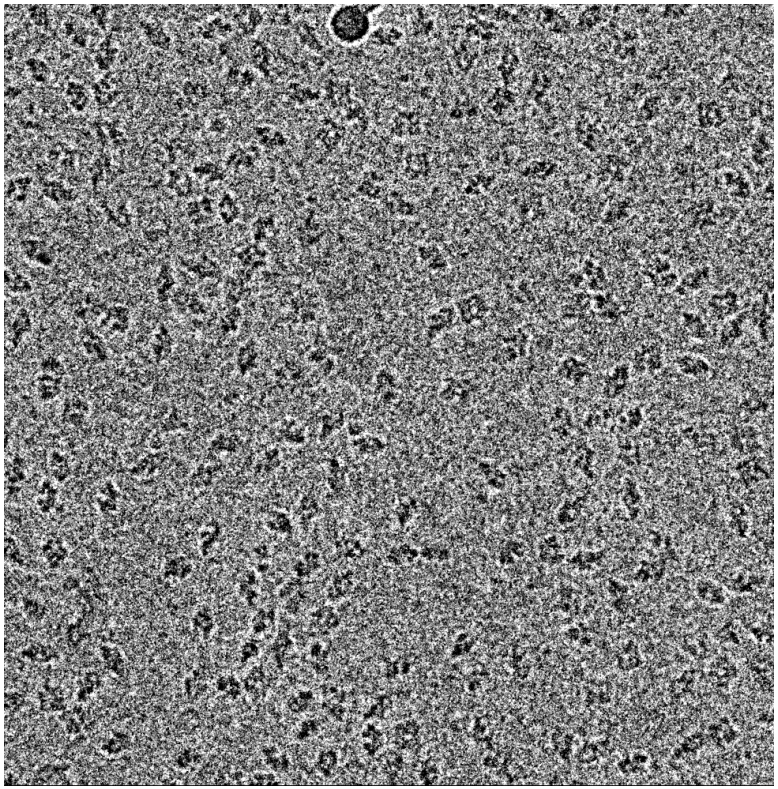
Joakim Andén[1], Ayelet Heimowitz[2], Amit Singer[2]

April 10, 2018

[1]Center for Computational Biology, Flatiron Institute, Simons Foundation

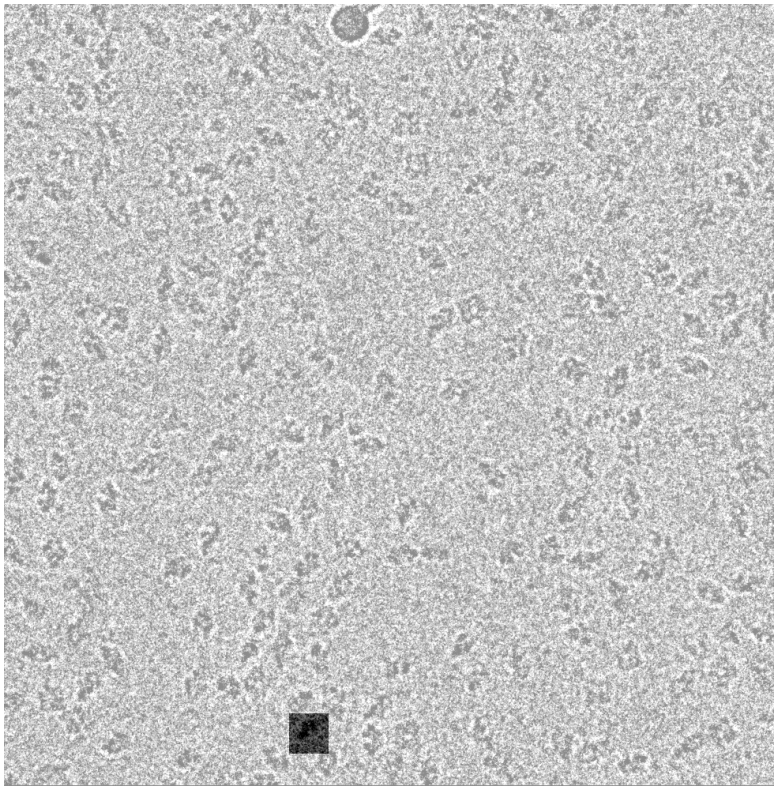
[2]Program in Applied and Computational Mathematics (PACM), Princeton University

Particle Picking in Cryo-EM



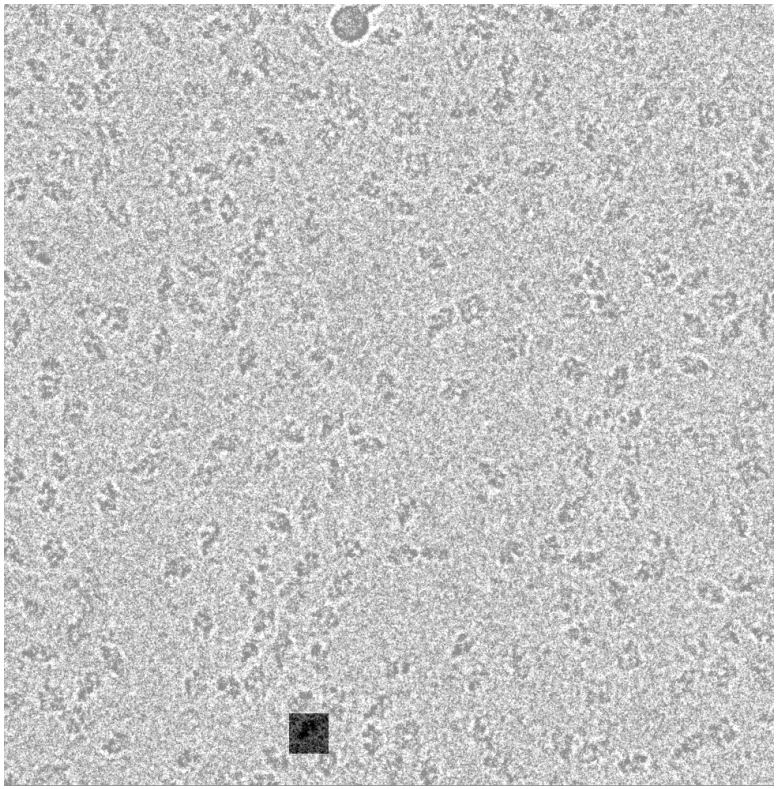
FLATIRON INSTITUTE

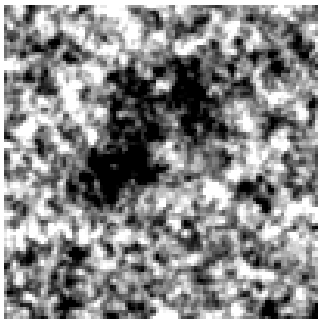
Particle Picking in Cryo-EM



FLATIRON INSTITUTE

Particle Picking in Cryo-EM





Template Matching

- Query window $f[n, m]$, template $g[n, m]$, cross-correlation:[1]

$$\max_{n,m} \sum_{n',m'} f[n - n', m - m']g[n', m']$$

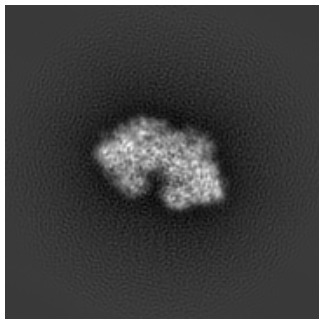
[1] Frank & Wagenknecht, 1983

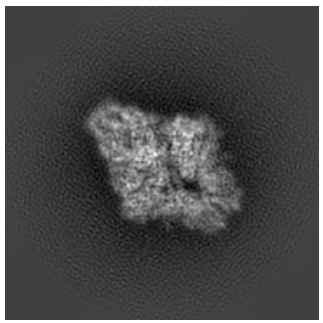
Template Matching

- Query window $f[n, m]$, template $g[n, m]$, cross-correlation:[1]

$$\max_{n,m} \sum_{n',m'} f[n - n', m - m']g[n', m']$$

$g[n, m]$





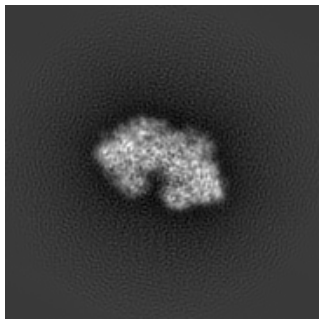
[1] Frank & Wagenknecht, 1983

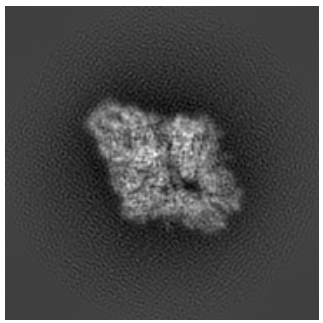
Template Matching

- Query window $f[n, m]$, template $g[n, m]$, cross-correlation:[1]

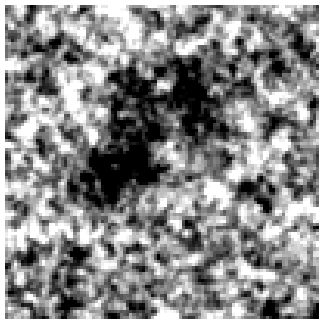
$$\max_{n,m} \sum_{n',m'} f[n - n', m - m']g[n', m']$$

$g[n, m]$





$f[n, m]$



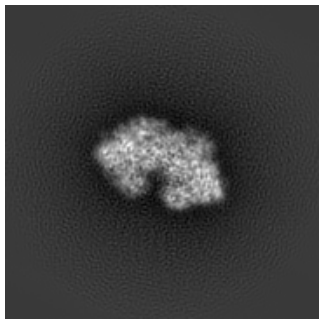
[1] Frank & Wagenknecht, 1983

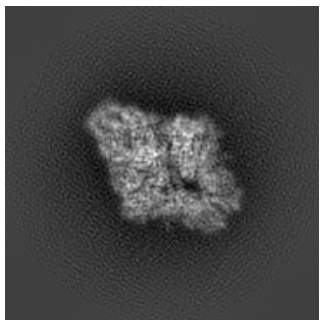
Template Matching

- Query window $f[n, m]$, template $g[n, m]$, cross-correlation:[1]

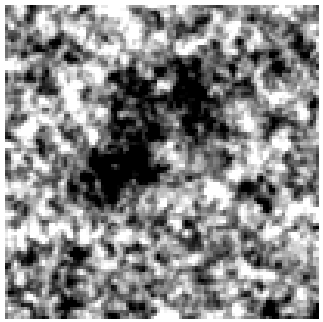
$$\max_{n,m} \sum_{n',m'} f[n - n', m - m']g[n', m']$$

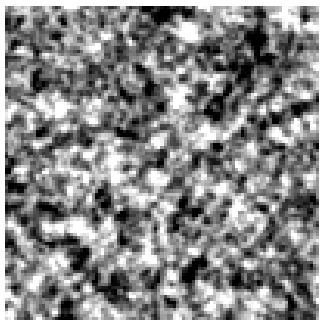
$g[n, m]$





$f[n, m]$



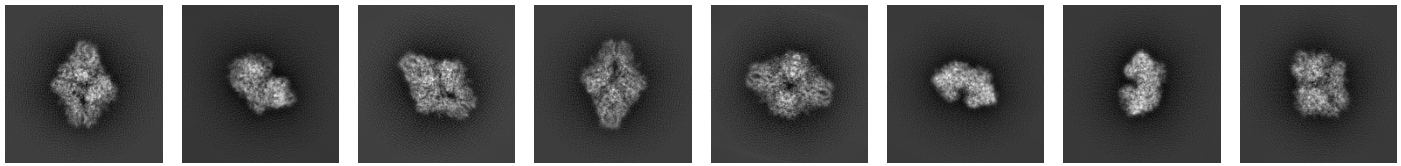


[1] Frank & Wagenknecht, 1983

FLATIRON INSTITUTE

What Templates?

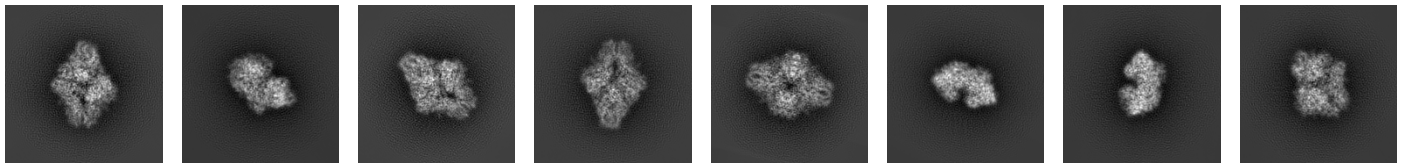
- Simulated projections of particles



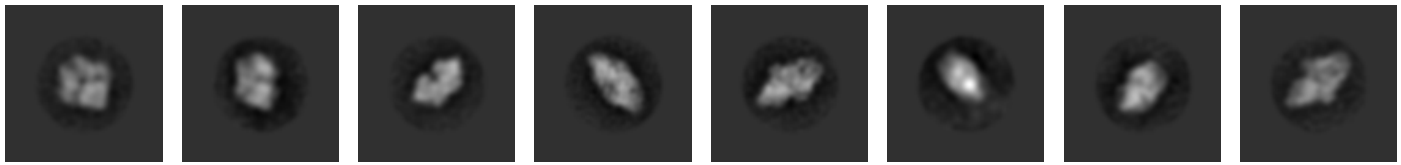
FLATIRON INSTITUTE

What Templates?

- Simulated projections of particles



- Class averages from micrograph[2]

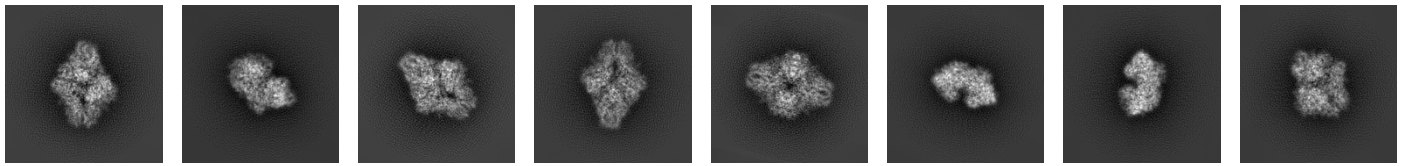


[2]Scheres, 2015

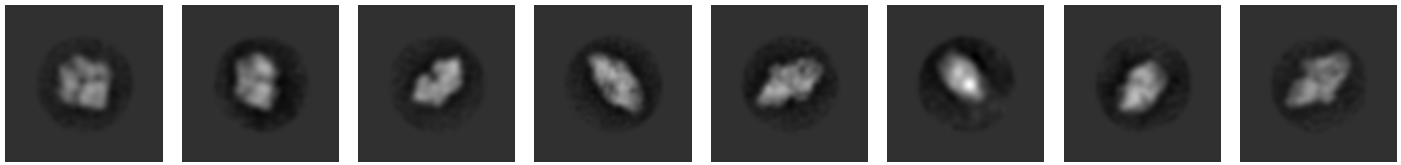
FLATIRON INSTITUTE

What Templates?

- Simulated projections of particles



- Class averages from micrograph[2]

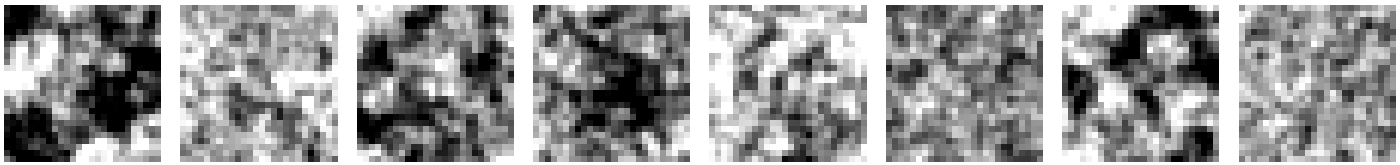


- Difference of Gaussians[3], disks[4]

[2]Scheres, 2015
[3]Voss et al., 2009
[4]Langlois et al., 2014

Template-Free Matching

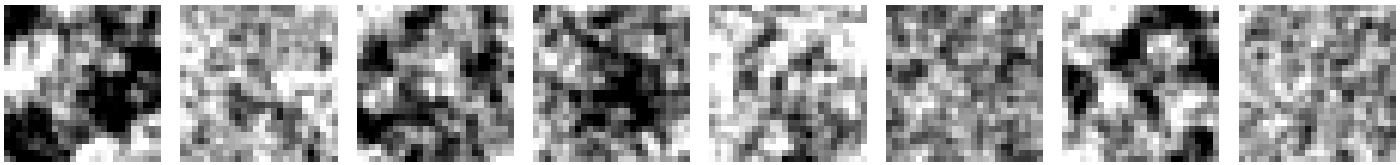
- Randomly select *reference* windows from micrograph



Some will contain particles, others not

Template-Free Matching

- Randomly select *reference* windows from micrograph

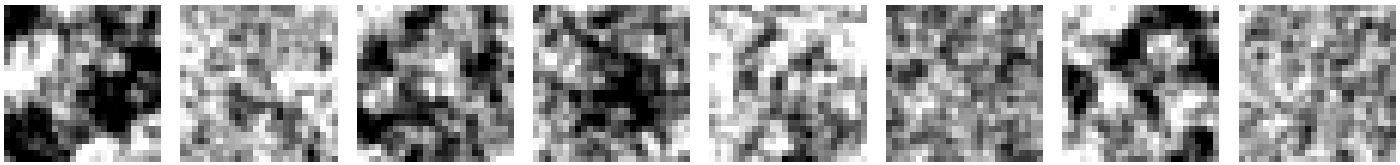


Some will contain particles, others not

- What are correlations of query window with references?

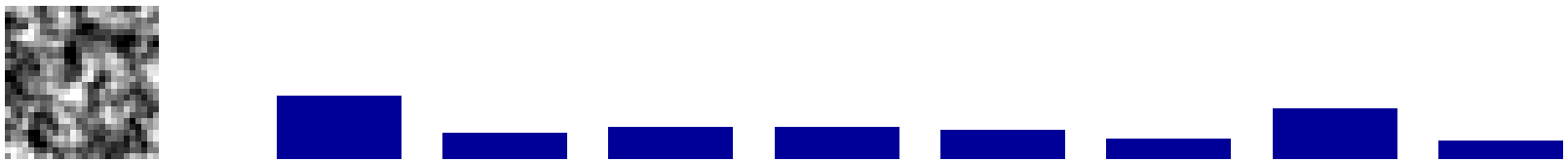
Template-Free Matching

- Randomly select *reference* windows from micrograph



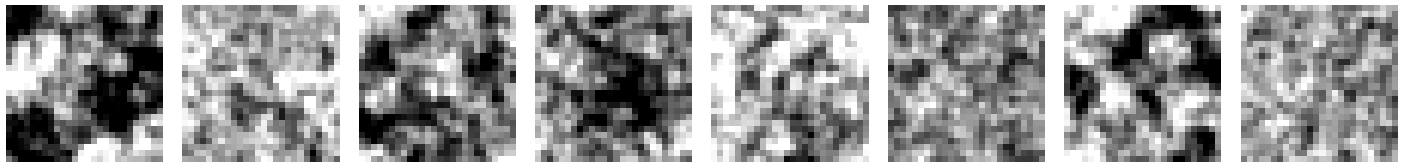
Some will contain particles, others not

- What are correlations of query window with references?



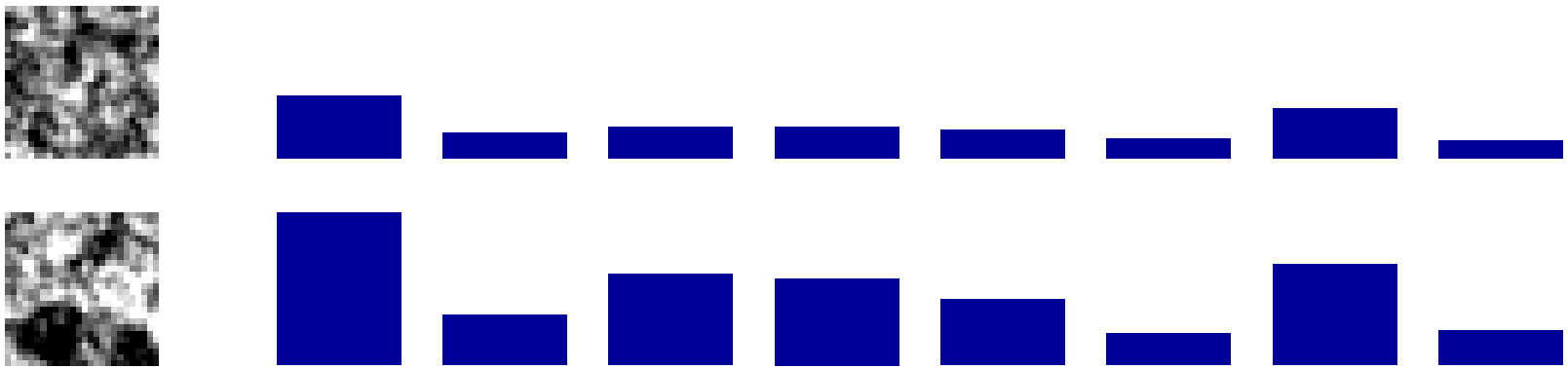
Template-Free Matching

- Randomly select *reference* windows from micrograph



Some will contain particles, others not

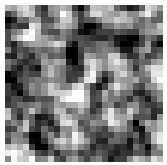
- What are correlations of query window with references?

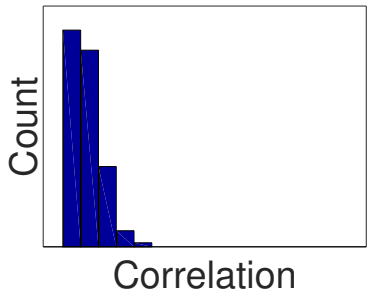


FLATIRON INSTITUTE

Correlation Distribution

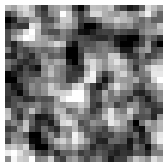
- Empty window correlates badly with all

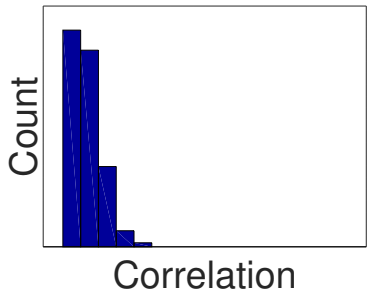




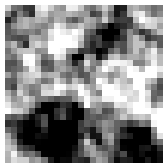
Correlation Distribution

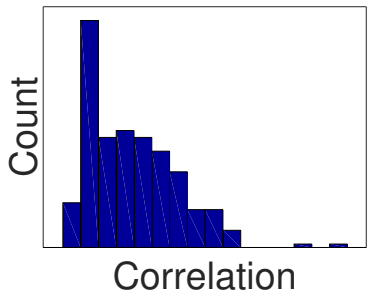
- Empty window correlates badly with all





- Window with particle correlates well with some





FLATIRON INSTITUTE

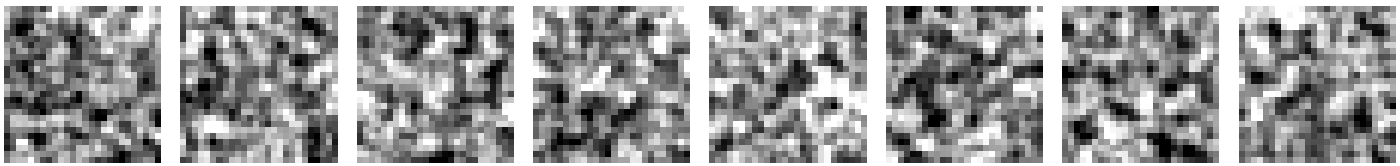
Window Score

- Count number k of correlations above threshold

Window Score

- Count number k of correlations above threshold

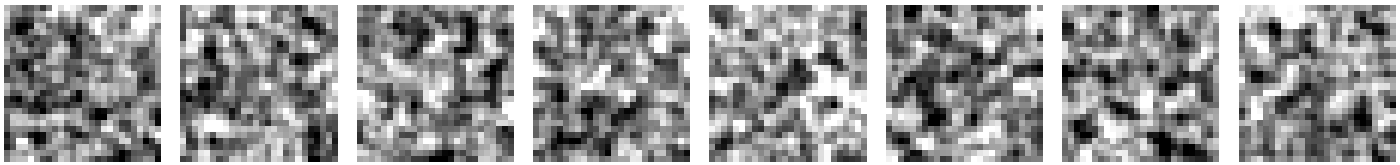
$k = 30$

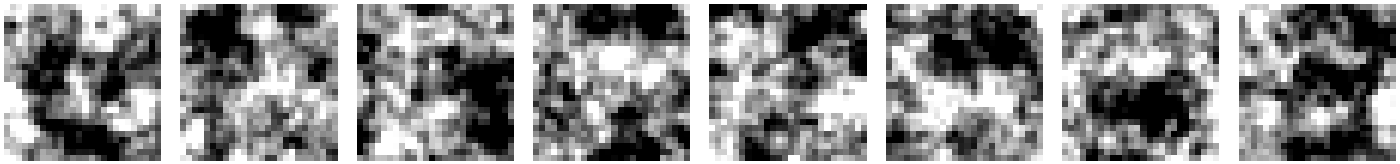


FLATIRON INSTITUTE

Window Score

- Count number k of correlations above threshold

$k = 30$ 

$k = 168$ 

FLATIRON INSTITUTE

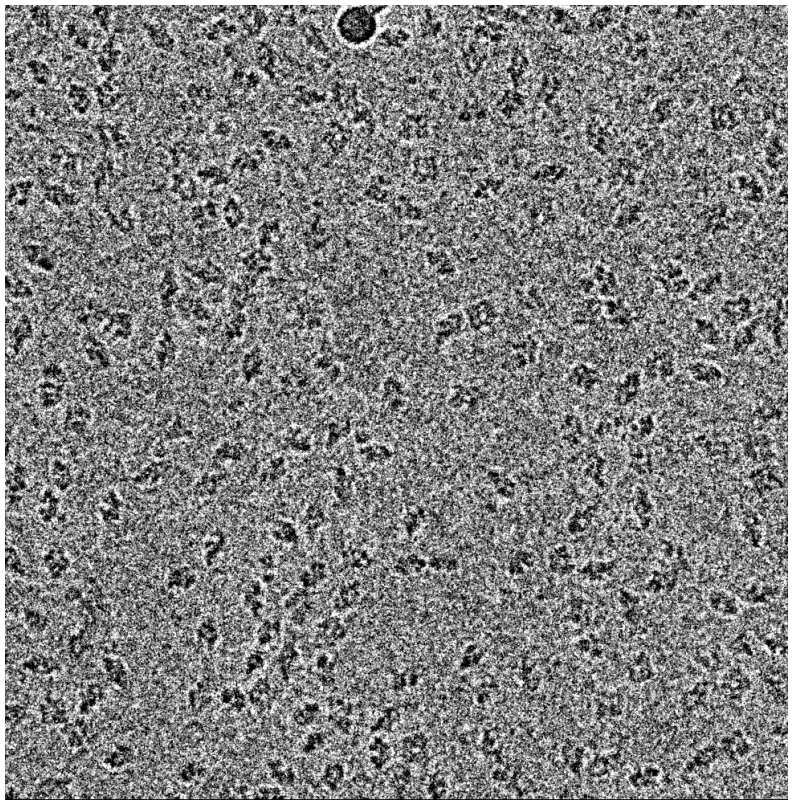
Results

- Good indicators, but expensive

Results

- Good indicators, but expensive

Micrograph

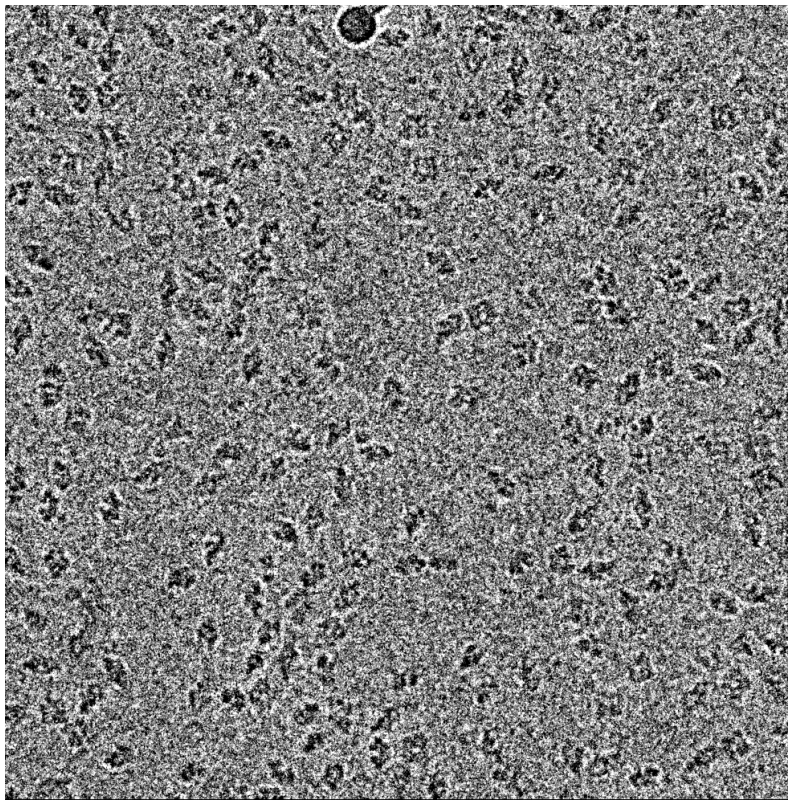


FLATIRON INSTITUTE

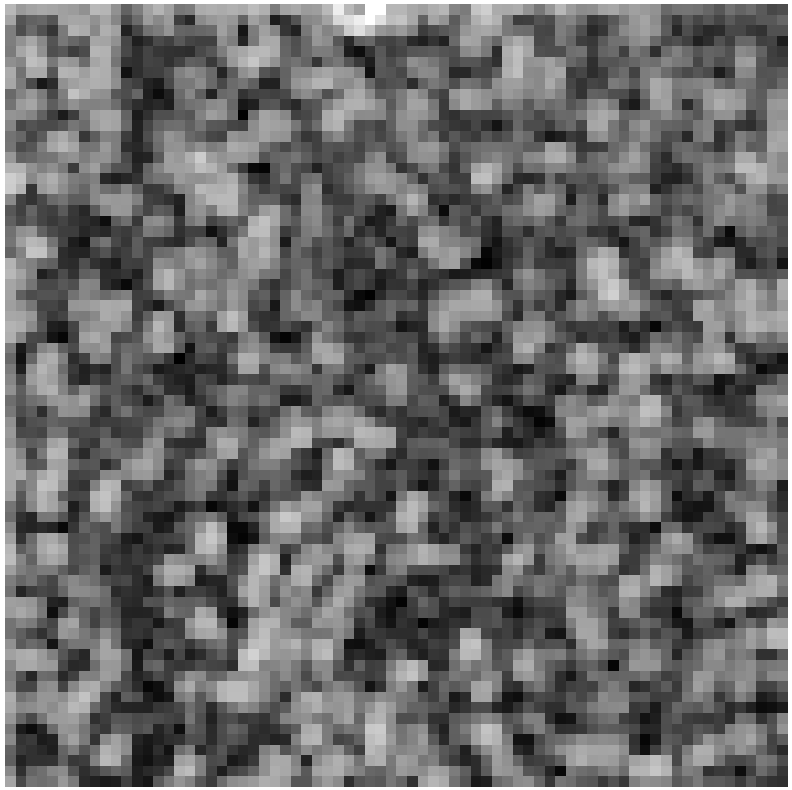
Results

- Good indicators, but expensive

Micrograph



k values

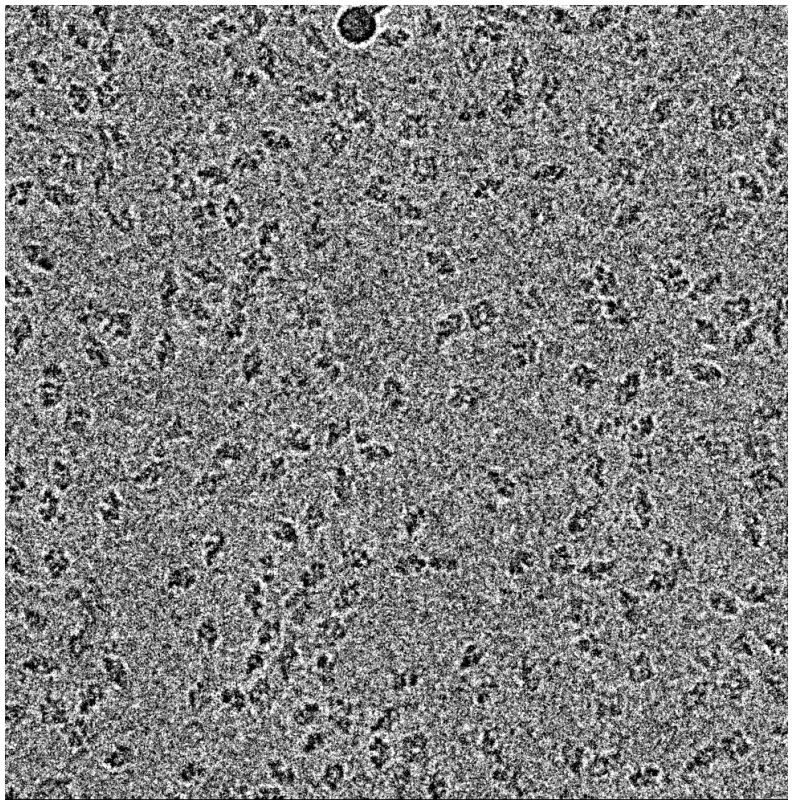


FLATIRON INSTITUTE

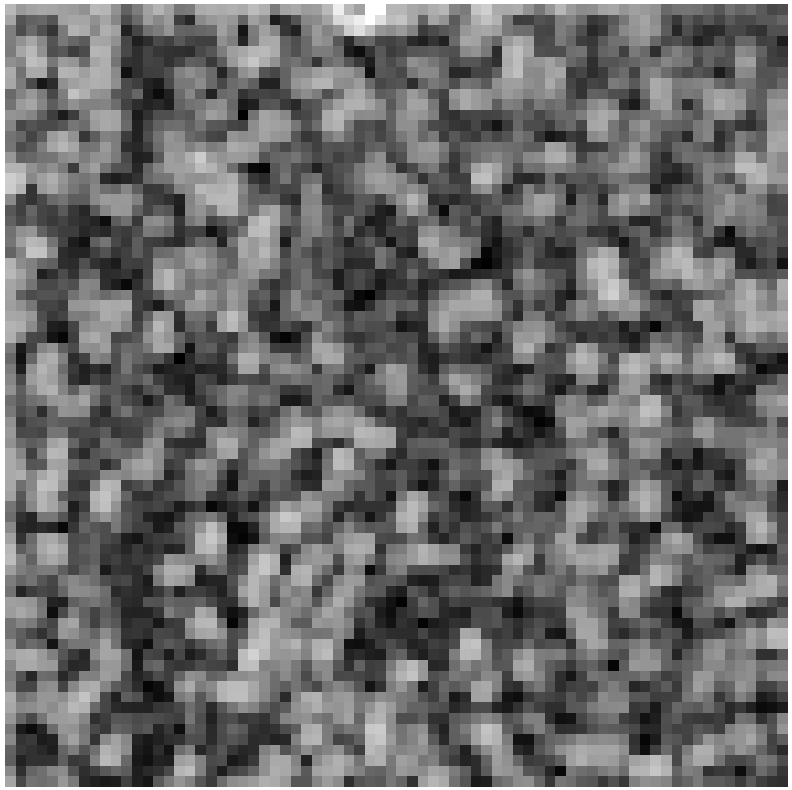
Results

- Good indicators, but expensive

Micrograph



k values



- Need denser sampling

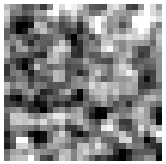
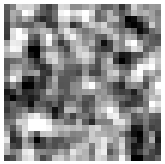
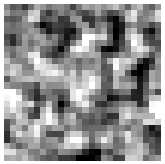
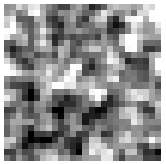
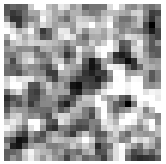
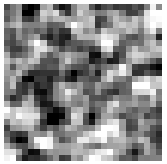
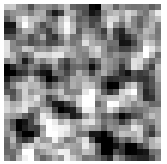
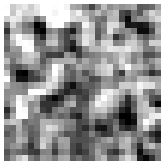
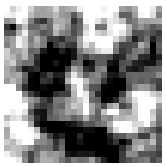
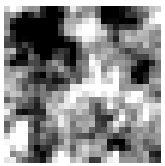
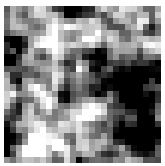
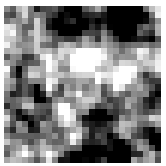
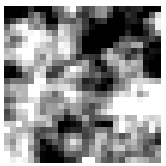
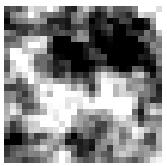
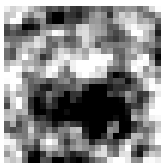
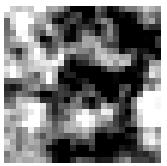
FLATIRON INSTITUTE

Localization

- Mean (μ) and standard deviation (σ) can discriminate

Localization

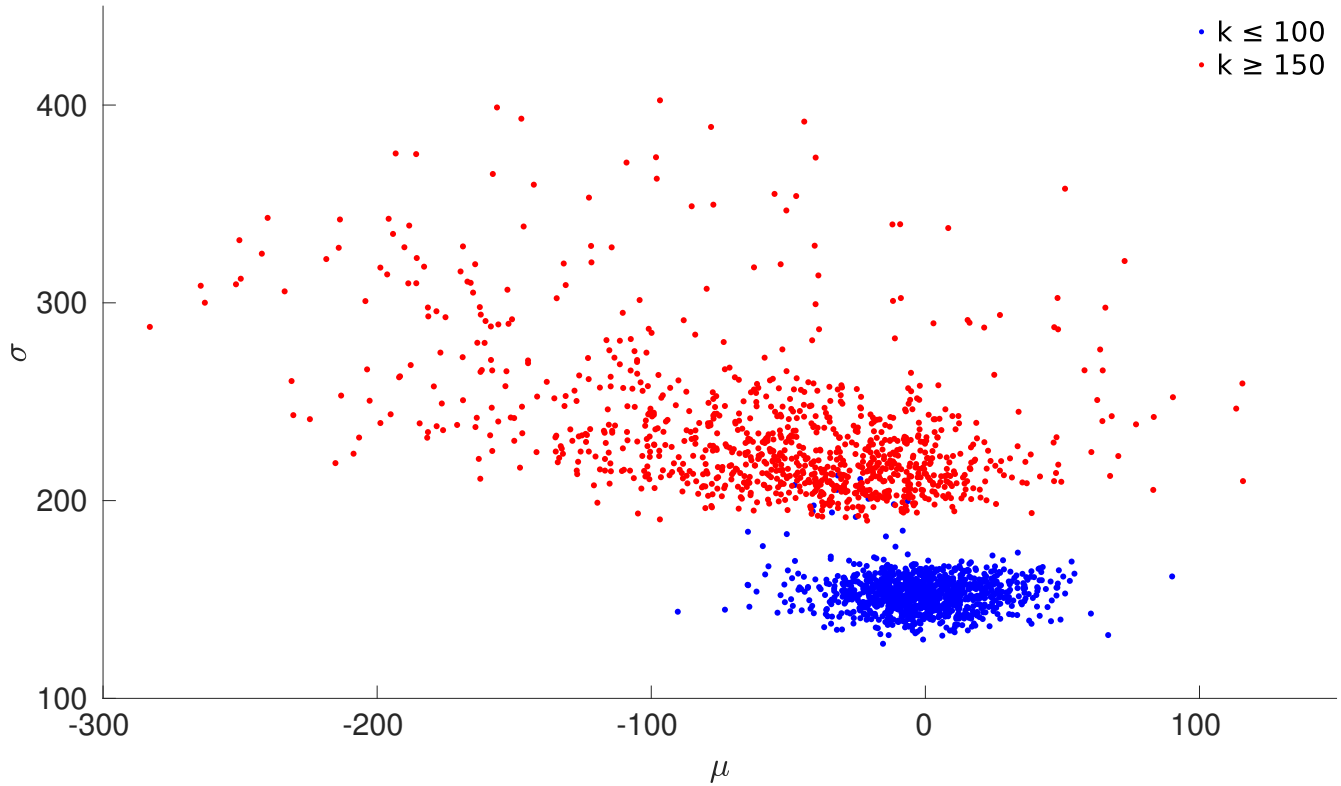
- Mean (μ) and standard deviation (σ) can discriminate

$k = 30$								
μ	-6.2	10	11	10	57	2.0	6.2	25
σ	141	140	140	140	140	150	150	150
$k = 168$								
μ	-10	-22	-35	-41	1.6	-18	-23	-61
σ	230	210	220	220	220	220	220	220

FLATIRON INSTITUTE

Localization

- Mean (μ) and standard deviation (σ) can discriminate



FLATIRON INSTITUTE

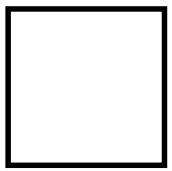
Localization (cont.)

- Train support vector machine[5] (SVM) on μ and σ from high and low k windows

[5] Cortes & Vapnik, 1995

Localization (cont.)

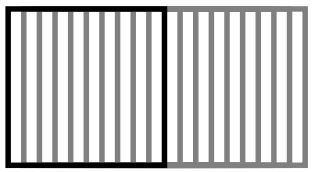
- Train support vector machine[5] (SVM) on μ and σ from high and low k windows
- Both μ and σ may be calculated fast



[5] Cortes & Vapnik, 1995

Localization (cont.)

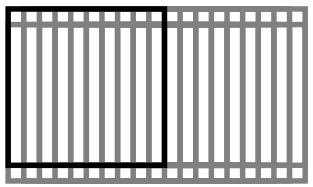
- Train support vector machine[5] (SVM) on μ and σ from high and low k windows
- Both μ and σ may be calculated fast



[5] Cortes & Vapnik, 1995

Localization (cont.)

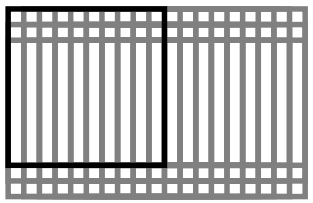
- Train support vector machine[5] (SVM) on μ and σ from high and low k windows
- Both μ and σ may be calculated fast



[5] Cortes & Vapnik, 1995

Localization (cont.)

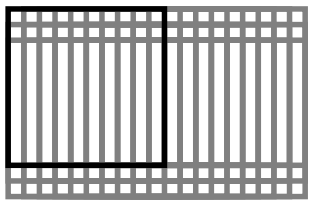
- Train support vector machine[5] (SVM) on μ and σ from high and low k windows
- Both μ and σ may be calculated fast



[5] Cortes & Vapnik, 1995

Localization (cont.)

- Train support vector machine[5] (SVM) on μ and σ from high and low k windows
- Both μ and σ may be calculated fast

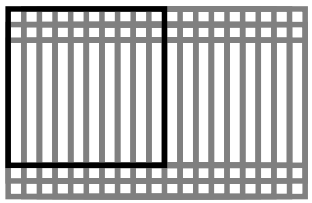


- Calculate μ and σ for dense sampling of windows

[5] Cortes & Vapnik, 1995

Localization (cont.)

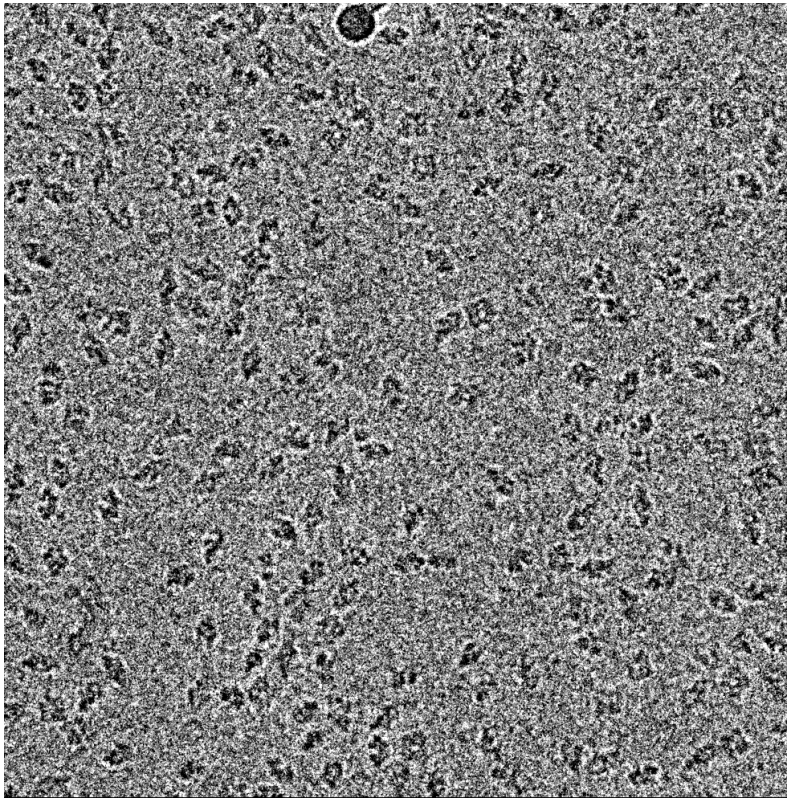
- Train support vector machine[5] (SVM) on μ and σ from high and low k windows
- Both μ and σ may be calculated fast



- Calculate μ and σ for dense sampling of windows
- Use SVM to predict

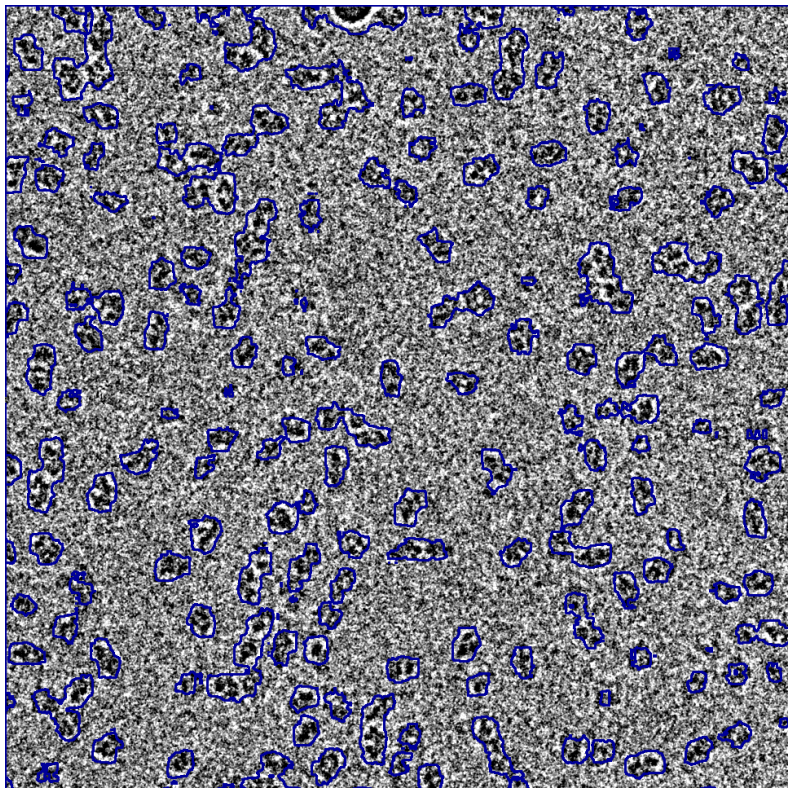
[5] Cortes & Vapnik, 1995

Results



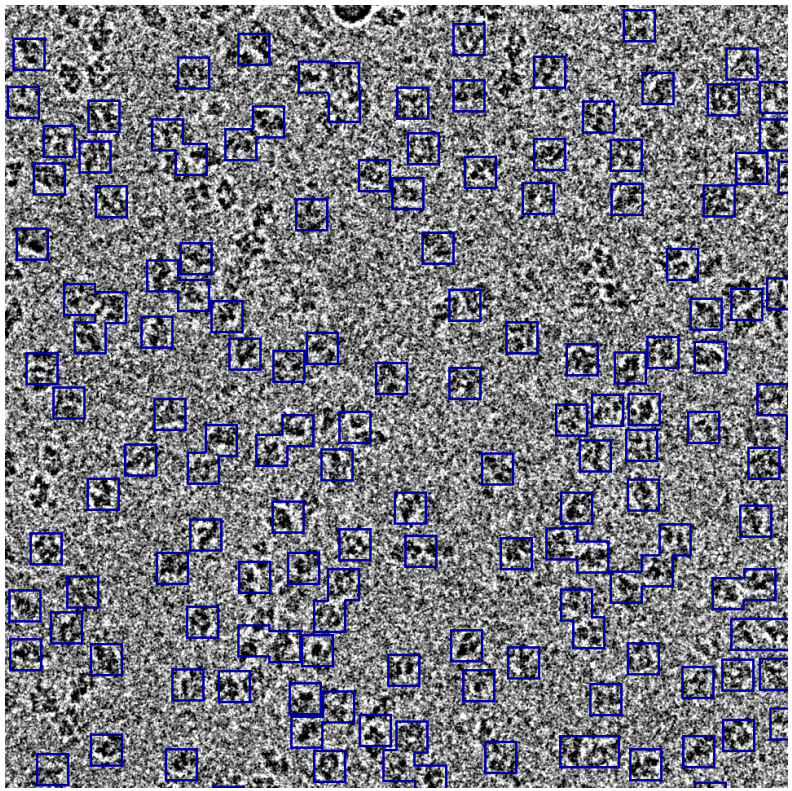
FLATIRON INSTITUTE

Results



FLATIRON INSTITUTE

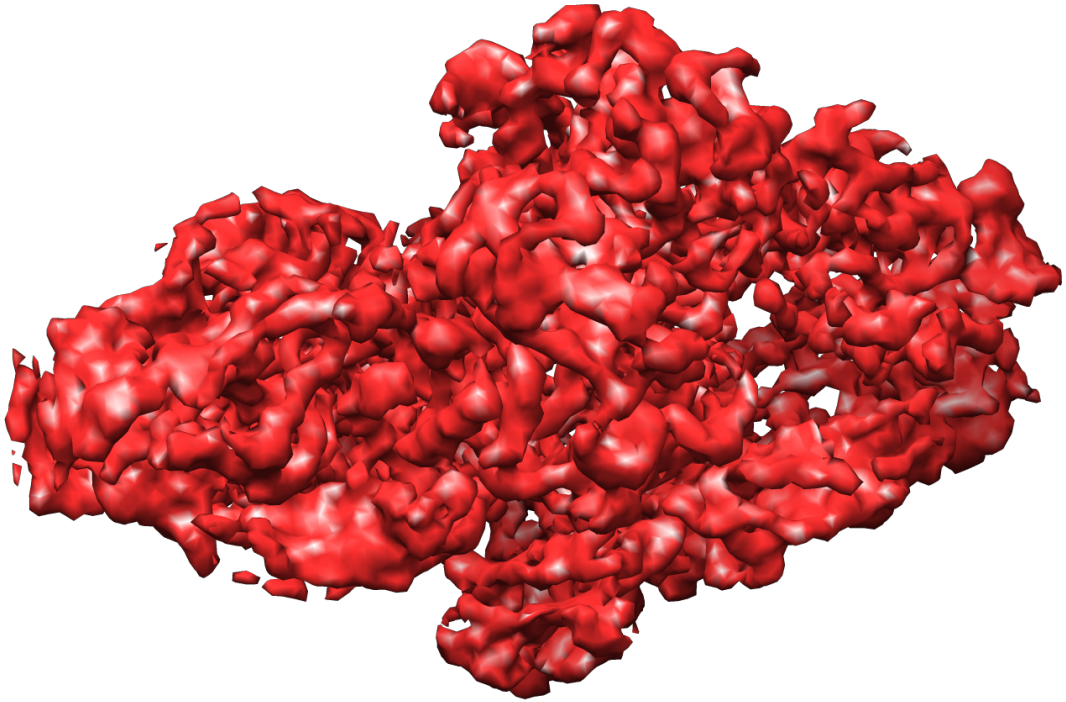
Results



FLATIRON INSTITUTE

Results (cont.)

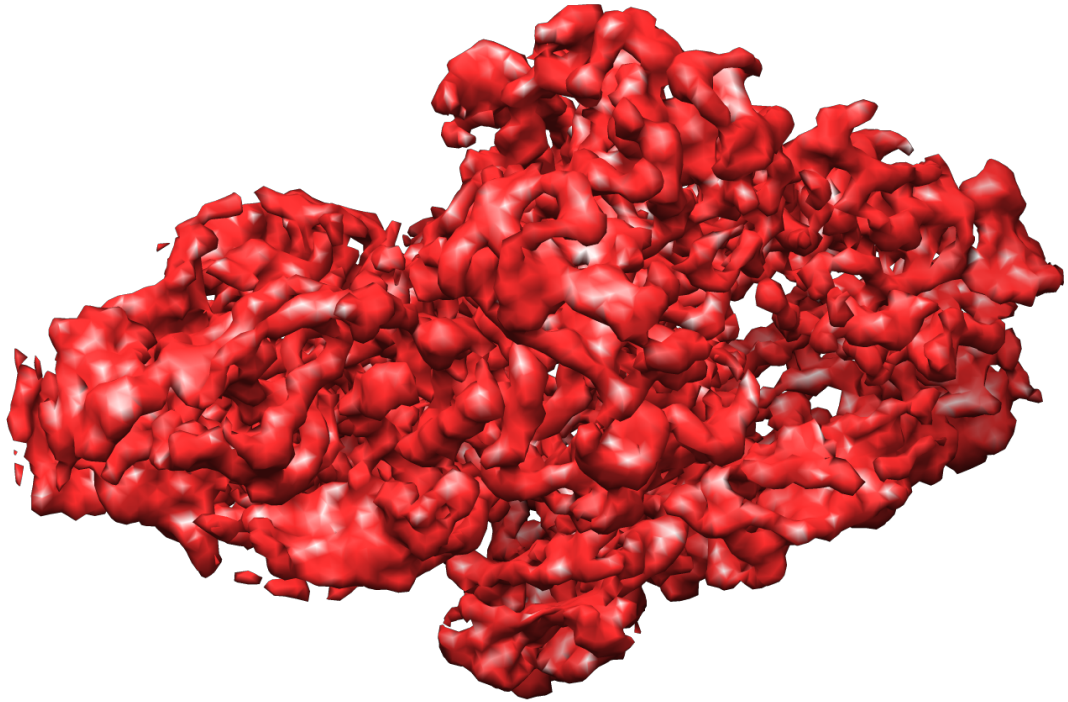
- Pick particles from 84 micrographs and reconstruct using RELION



FLATIRON INSTITUTE

Results (cont.)

- Pick particles from 84 micrographs and reconstruct using RELION

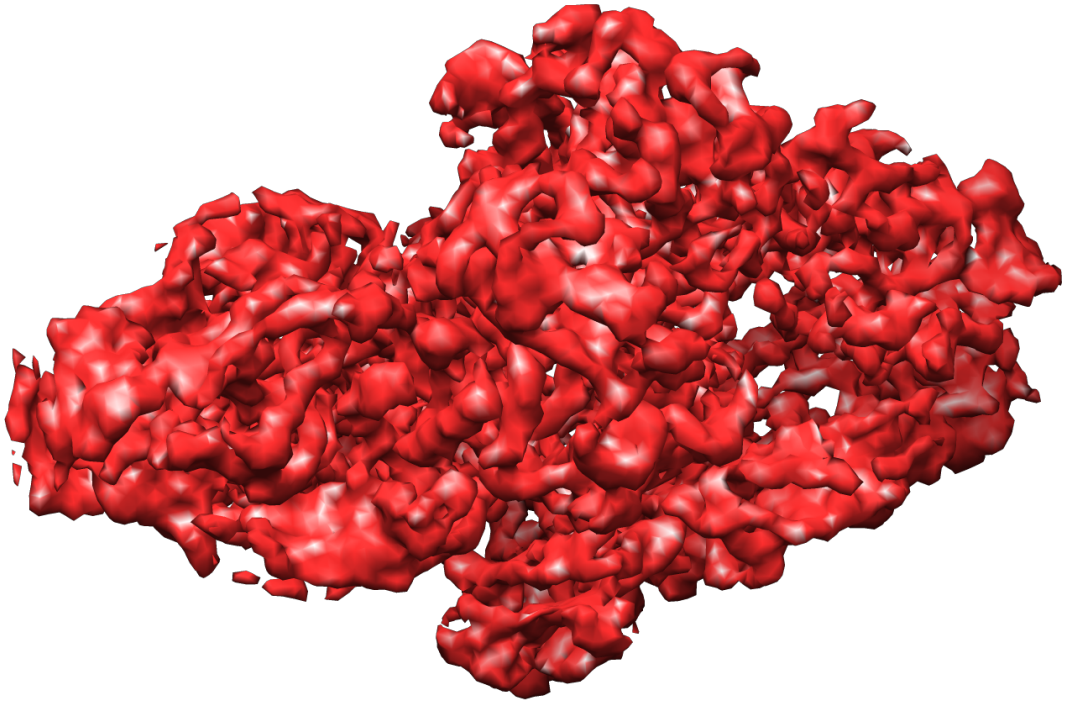


- Agrees up to 6.7 Å with published EMD-2824

FLATIRON INSTITUTE

Results (cont.)

- Pick particles from 84 micrographs and reconstruct using RELION



- Agrees up to 6.7 $\mathrm{\AA}$ with published EMD-2824
- Fully automatic, template-free particle picking, 15 $\mathrm{minutes}$

FLATIRON INSTITUTE

Conclusion

- Cross-correlation within micrograph strong signal

Conclusion

- Cross-correlation within micrograph strong signal
- Unsupervised classification using cross-correlation histogram

Conclusion

- Cross-correlation within micrograph strong signal
- Unsupervised classification using cross-correlation histogram
- Leverage initial classification to train SVM on mean and standard deviation

Conclusion

- Cross-correlation within micrograph strong signal
- Unsupervised classification using cross-correlation histogram
- Leverage initial classification to train SVM on mean and standard deviation
- Apply SVM to densely spaced windows for prediction

Future Work

- How to extract more information from cross-correlation? Distance on histograms? Moments of distributions? Separate identification of non-particles?

Future Work

- How to extract more information from cross-correlation? Distance on histograms? Moments of distributions? Separate identification of non-particles?

- Score function on densely spaced windows. Fast calculation? Smoothness? Efficient interpolation?

Future Work

- How to extract more information from cross-correlation? Distance on histograms? Moments of distributions? Separate identification of non-particles?

- Score function on densely spaced windows. Fast calculation? Smoothness? Efficient interpolation?

- More discriminative features for SVM? Large-scale gradients? Wavelet moments? Other classifiers? Training for deep networks?

Thank you!

Paper:
`https://arxiv.org/abs/1802.00469`

Software:
`https://github.com/PrincetonUniversity/APPLEpicker`

Funding:
NIGMS R01GM090200, BSF 2014401, AFOSR FA9550-12-1-0317, Simons Investigator Award, Simons Collaboration on Algorithms and Geometry, and Moore Data-Driven Discovery Investigator Award

FLATIRON INSTITUTE