Cryo-EM Principles

Single-Particle Reconstruction

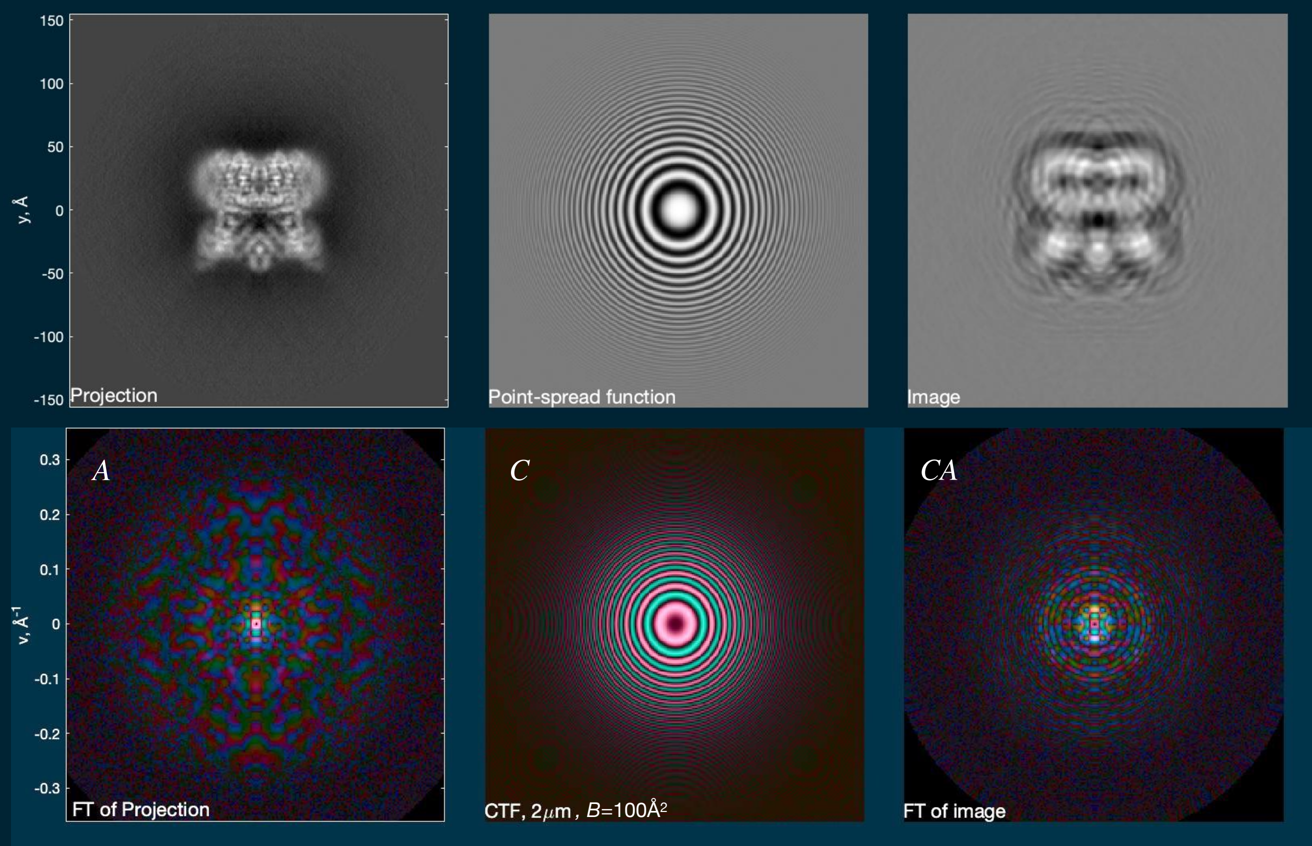
Fred Sigworth
Yale University

Modeling an image

$$X = CA + N$$

Deconvolution:

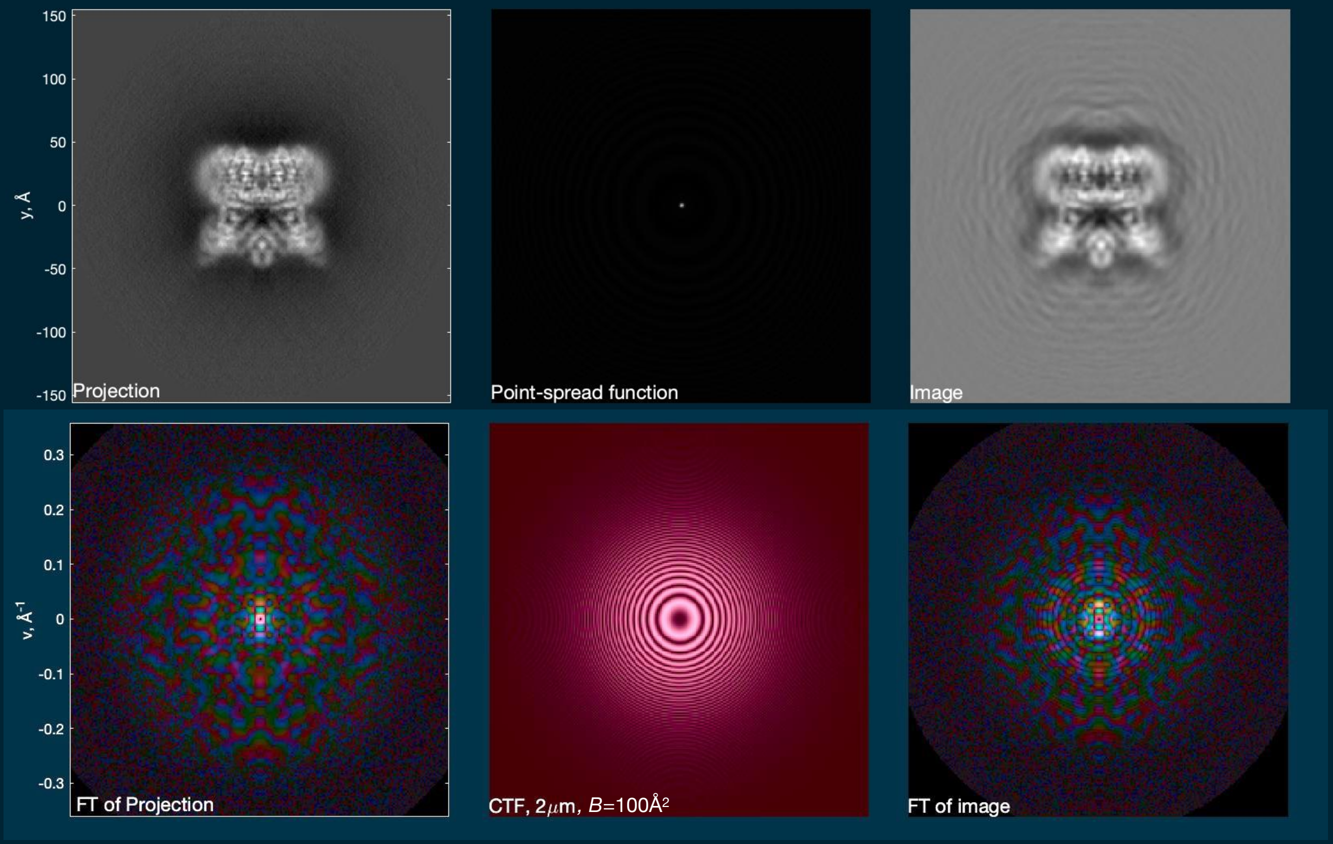
$$\tilde{A} = X/C\ ??$$



How to undo the CTF effects?

1. **Phase flipping**

$$\tilde{A} = \mathrm{sgn}(C)X$$



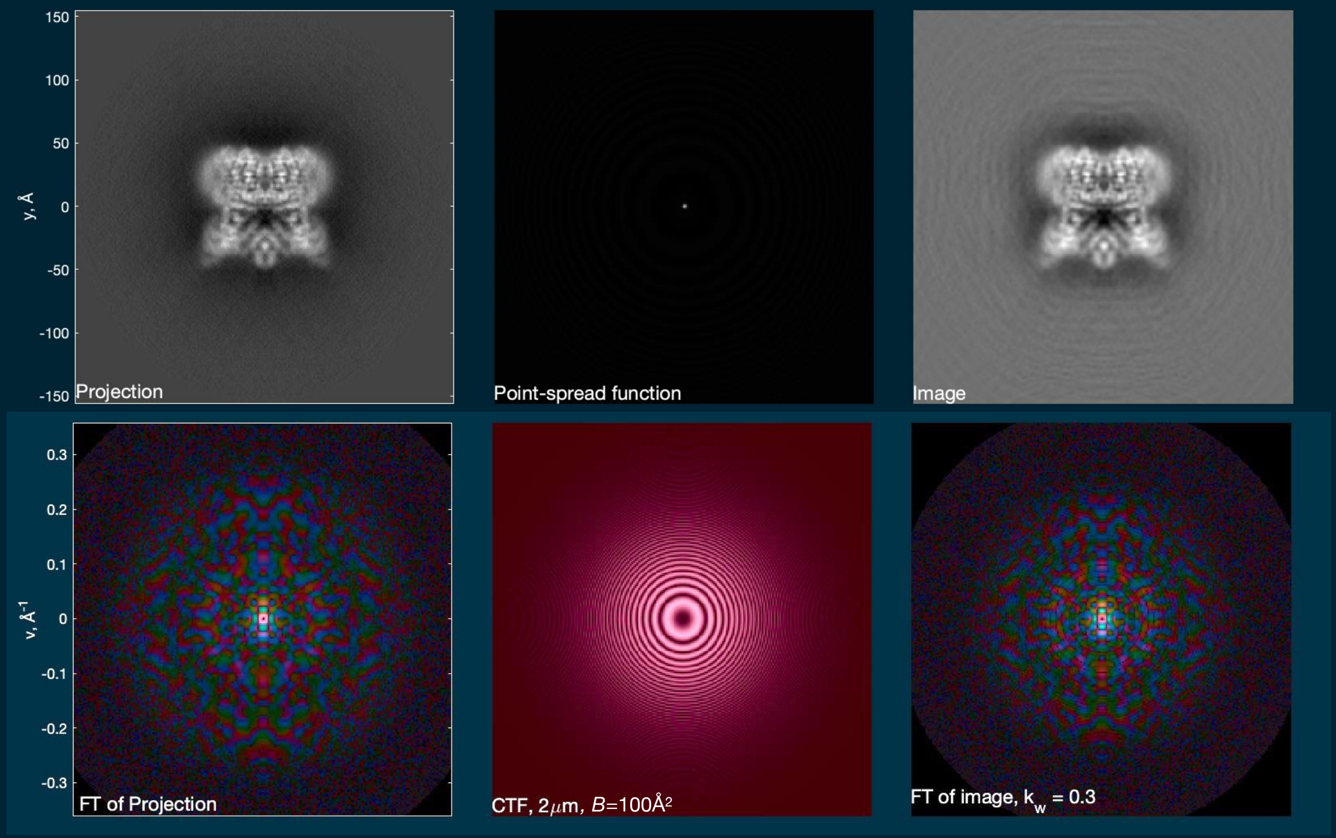
How to undo the CTF effects?

1. **Phase flipping**

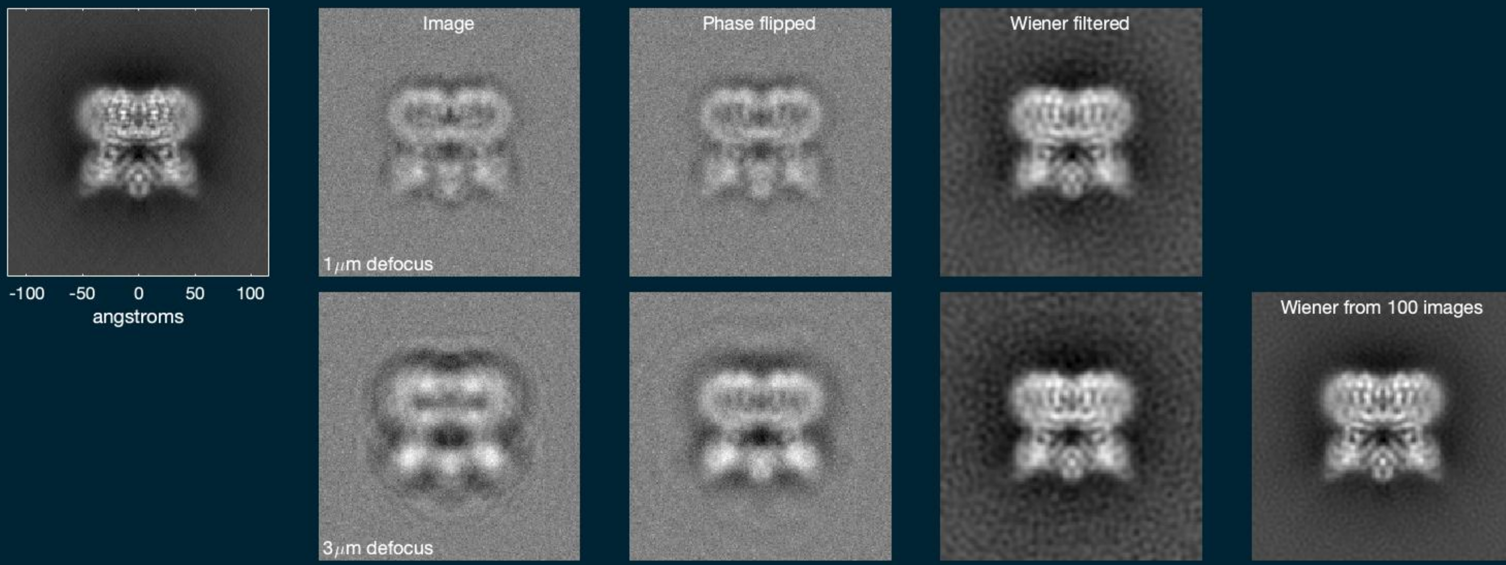
$$\tilde{A} = \operatorname{sgn}(C)X$$

2. **Wiener filter**

$$\tilde{A} = \frac{CX}{C^2 + k}$$



How to undo the CTF effects in noisy images?



3. Wiener from multiple images

$$\tilde{A} = \frac{\sum_i^N C_i X_i}{k + \sum_i^N C_i^2}$$

$$k = 1/\mathrm{SNR}$$
$$= \frac{|N|^2}{|A|^2}$$

Notation

Two notations for a single pixel in the image X:

X_j —the j^{th} pixel (out of J pixels total)

$X(x, y)$ —the pixel at position x, y

Ways to compare images

Define the “reference”

$R = CA$

Squared difference

$$(X - R)^2 = \sum_j (X_j - R_j)^2$$

[note that

$(X - R)^2 = X^2 + R^2 - 2X \cdot R$]

Correlation

$$\mathrm{Cor} = X \cdot R$$

Correlation coefficient

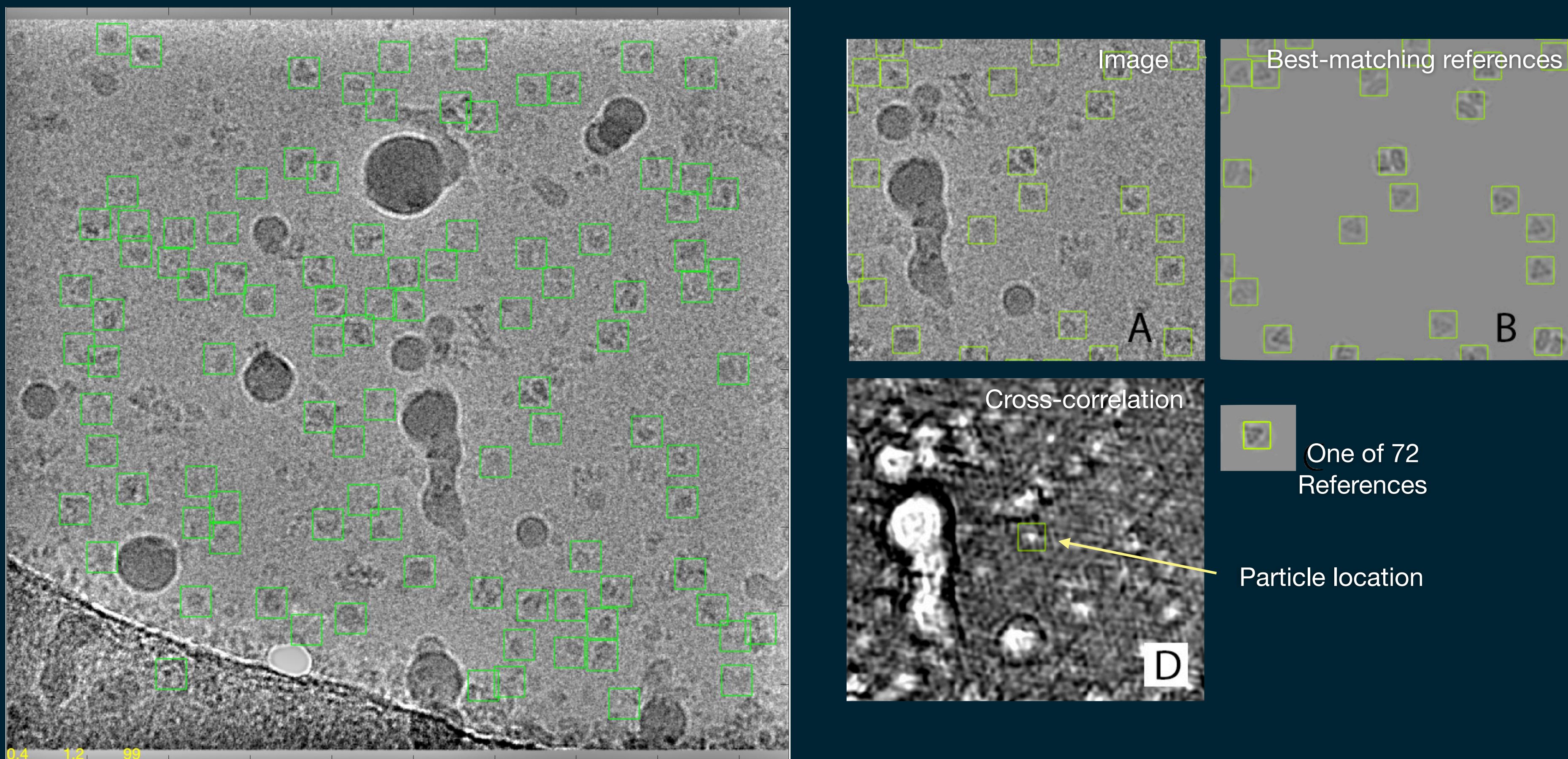
$$\mathrm{CC} = \frac{X \cdot R}{|X||R|}$$

Cross-correlation function

$$\begin{aligned}\mathrm{Cor}(x, y) &= X \otimes R \\ &= \sum_{s,t} X(s, t)R(x + s, y + t)\end{aligned}$$

[Correlation is like convolution.
The FT pair is: $g \otimes h \rightarrow GH^*$]

Example: cross-correlation particle picker



Probabilities, another way to compare images

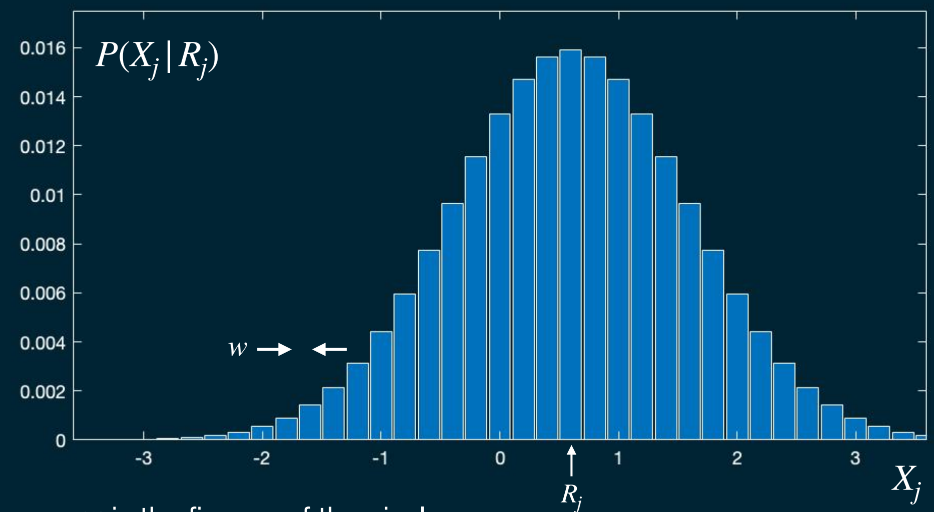
$$X = R + N$$

Probability of a pixel value:

$$P(X_j \,|\, R_j) = \frac{\cancel{w}\ 1}{\sqrt{2\pi\sigma^2}}\, e^{-(X_j - R_j)^2/2\sigma^2}$$

Probability of observing an image that comes from R:

$$P(X \,|\, R) = \frac{\cancel{w^J}\ 1}{(2\pi\sigma^2)^{J/2}}\, e^{-(X-R)^2/2\sigma^2}$$



w is the finesse of the pixel intensity measurements. We'll ignore it (set it to 1).

Probabilities, another way to compare images

A projection

$$A = \mathbf{P}_\phi V$$

Probability of observing an image

$$P(X \mid V, \phi) = (2\pi\sigma^2)^{-J/2} \; e^{-(X - \mathbf{C}\mathbf{P}_\phi V)/2\sigma^2}$$

Probability of a projection direction

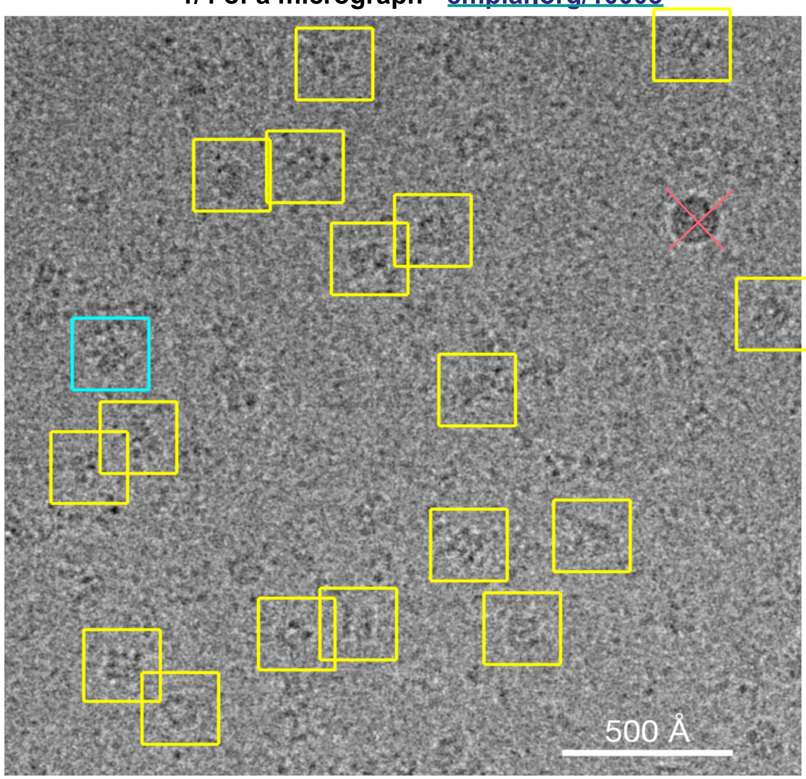
$$P(\phi \mid X, V) = \frac{P(X \mid V, \phi)P(\phi)}{\int P(X \mid V, \phi)P(\phi)d\phi}$$

Determining the orientation angles: example from the TRPV1 dataset

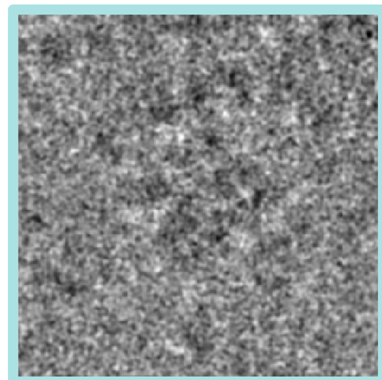
Structure of the TRPV1 ion channel determined by electron cryo-microscopy

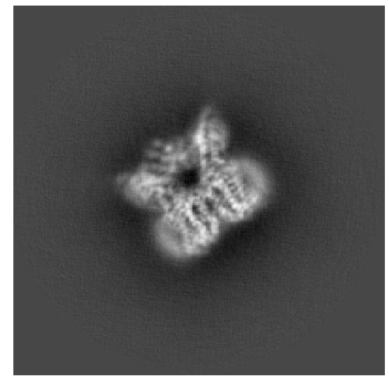
Maofu Liao[1]*, Erhu Cao[2]*, David Julius[2] & Yifan Cheng[1]

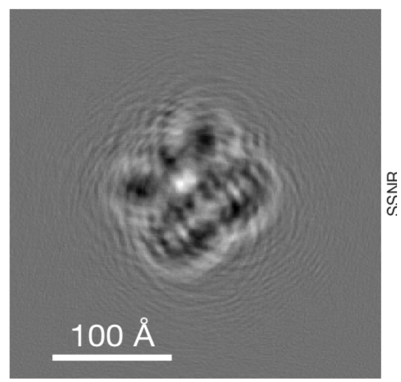
1/4 of a micrograph - empiar.org/10005

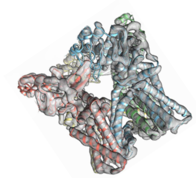


One particle image

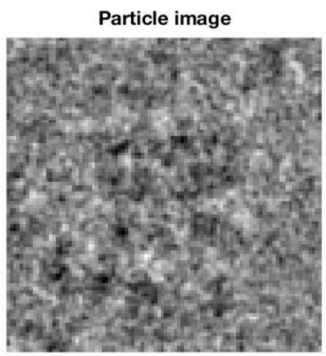


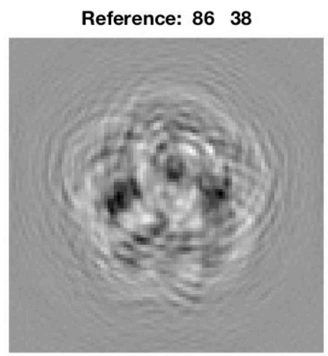


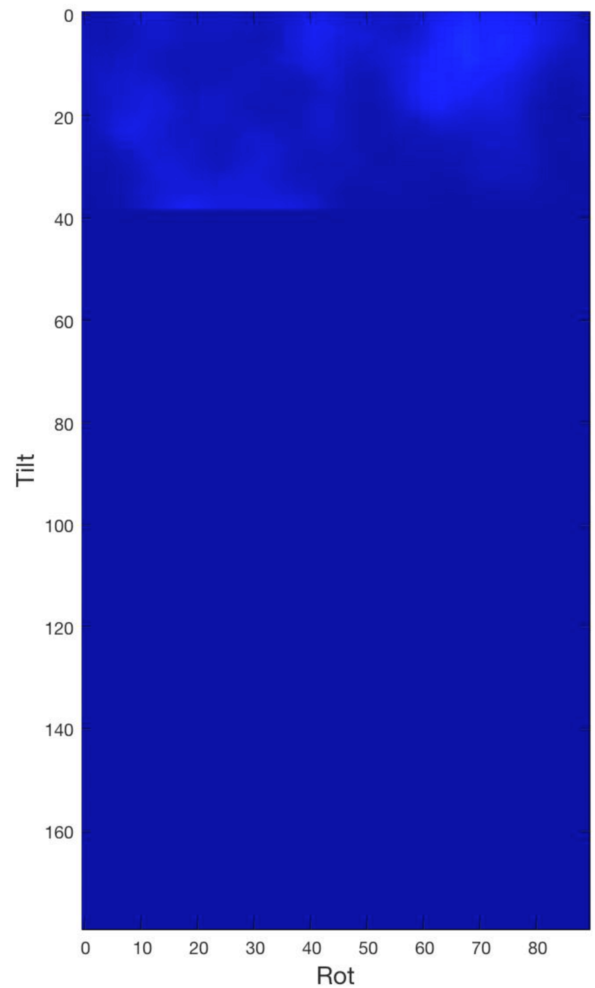




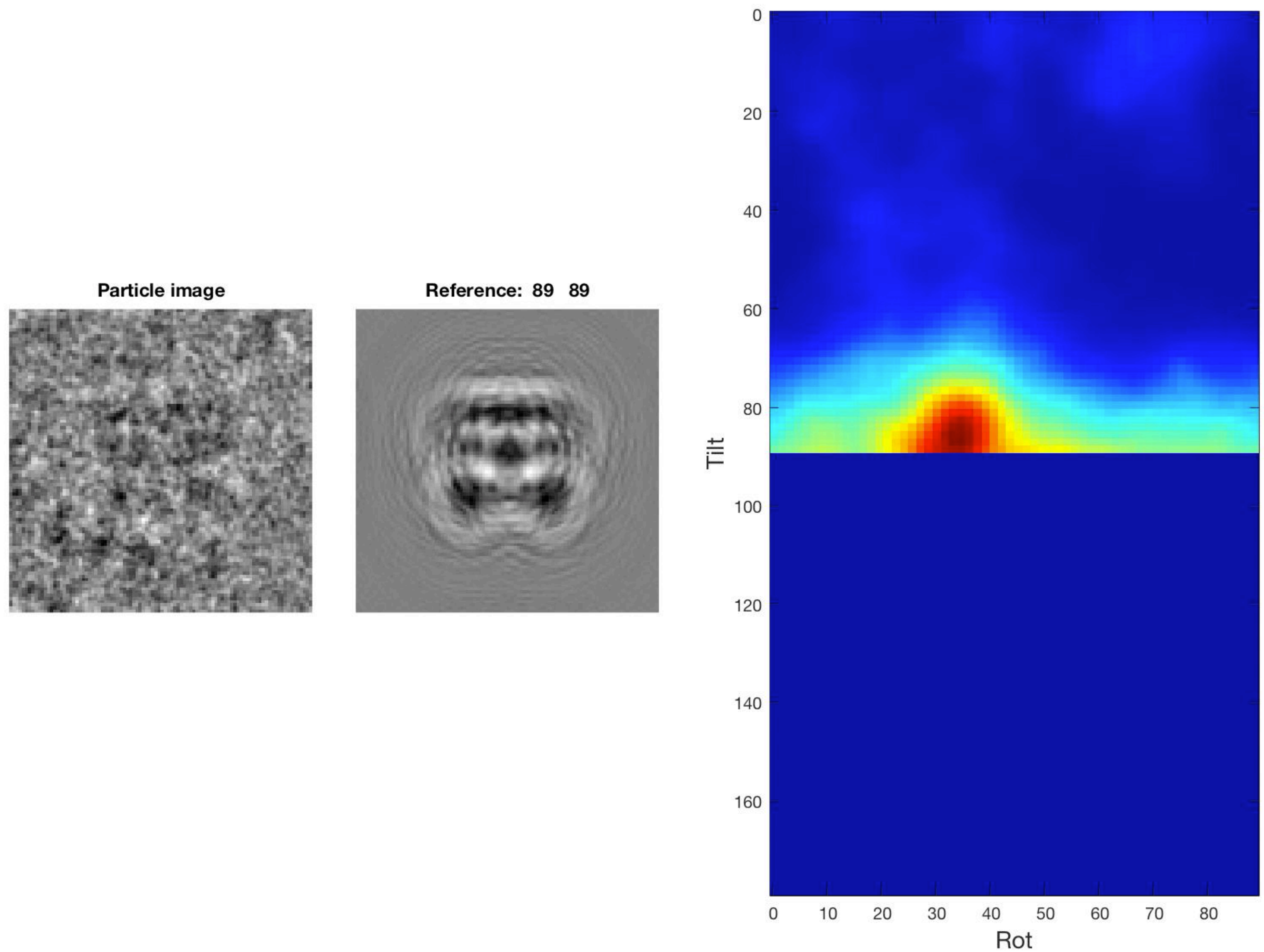
Probability of orientations $P(\phi \mid X, V)$







Probability of orientations $P(\phi \mid X, V)$



FREALIGN (N. Grigorieff) is like a Wiener filter

A Frealign iteration, refining $V^{(n)}$ to $V^{(n+1)}$, consists of two steps:

1. Find the projection image $R_i = C_i \mathbf{P}_{\phi_i} V^{(n)}$ that maximizes the correlation coefficient for each image X_i,

$$\mathrm{CC} = \frac{X_i \cdot R_i}{|X_i||R_i|}.$$

2. Update the volume according to

$$V^{(n+1)} = \frac{\sum_i^N \mathbf{P}_{\phi_i}^{\mathbf{T}} C_i X_i}{k + \sum_i^N \mathbf{P}_{\phi_i}^{\mathbf{T}} C_i^2}$$

Notes

1. C_i is the CTF corresponding to the image X_i.
2. The projection operator $\mathbf{P}_\phi$ also includes translations. So ϕ consists of five variables: $\phi = \{\alpha, \beta, \gamma, t_x, t_y\}$.
3. $\mathbf{P}_{\phi_i}^{\mathbf{T}}$ is the corresponding back projection operator. It operates on a 2D image and returns a 3D volume. In Fourier space it yields a volume that is all zeros except for values along a slice.

Probabilities and Likelihoods

Let $\mathbf{X} = \{X_1 \,..\, X_N\}$ be our "stack" of particle images. We'd like to find the best 3D volume consistent with these data, say maximizing

$$P(V \mid \mathbf{X}).$$

According to Bayes' theorem,

$$P(V \mid \mathbf{X}) = P(\mathbf{X} \mid V)\,\frac{P(V)}{P(\mathbf{X})}.$$

prior $\rightarrow$ **Experiment** $\rightarrow$ ***posterior***

1. $P(\mathbf{X})$ doesn't depend on V so we can ignore it.
2. $P(\mathbf{X} \mid V)$ is something we can calculate. It's called the likelihood.
3. $P(V)$ is called the prior probability.
4. The product $P(\mathbf{X} \mid V)\,P(V)$ is called the posterior probability.

We know how to compute the likelihood

We already know that

$$P(X \mid V, \phi) = (2\pi\sigma^2)^{-J/2}\; e^{-(X - \mathbf{CP}_\phi V)/2\sigma^2}$$

To get the likelihood for one image we just integrate over all the ϕ's:

$$P(X \mid V) = \int P(X \mid V, \phi)\, P(\phi)\, d\phi$$

To get the likelihood for the whole dataset we compute the product over all the images,

$$P(\mathbf{X} \mid V) = \prod_i^N \int P(X_i \mid V, \phi)\, P(\phi)\, d\phi,$$

or for numerical sanity, we compute the log likelihood,

$$L = \sum_i^N \ln\left(\int P(X_i \mid V, \phi)\, P(\phi)\, d\phi \right).$$

Relion and CryoSPARC both do this to estimate V

$$V = \arg\max_{V} \left[\ln P(V) + \sum_{i}^{N} \ln \left(\int P(X_i \,|\, V, \phi)\, P(\phi)\, d\phi \right) \right]$$

.

It's not too hard using the E-M algorithm

The Expectation-Maximization (E-M) algorithm has this iteration, guaranteed to increase the likelihood:

$$V^{(n+1)} = \frac{\sum_i \int \Gamma_i^{(n)}(\phi) \mathbf{P}_\phi^\mathbf{T} C_i X_i \, d\phi}{\frac{\sigma^2}{T\tau^2} + \sum_i \int \Gamma_i^{(n)}(\phi) \mathbf{P}_\phi^\mathbf{T} C_i^2 \, d\phi}$$

where

$$\Gamma_i^{(n)}(\phi) = P(\phi \,|\, X_i, V^{(n)})$$

$$= \frac{P(X_i \,|\, V^{(n)}, \phi) P(\phi)}{\int P(X_i \,|\, V^{(n)}, \phi) P(\phi) \, d\phi}$$

Finally, recall,

$$P(X_i \,|\, V^{(n)}, \phi) = (2\pi\sigma^2)^{-J/2} \, e^{-(X_i - \mathbf{C}\mathbf{P}_\phi V^{(n)})/2\sigma^2}.$$

…Relion's "E step" is basically the evaluation of $\Gamma(\phi)$ for each image

3D Classification

To estimate multiple volumes, we start with the model that a given image results from one of K different volumes.

$X_i = C_i \mathbf{P}_\phi V_{k_i} + N_i, \;\; k = 1 \,..\, K$

Then to estimate those volumes, we expand the Γ function to include the probability that an image comes from volume k. Then the update of each volume becomes

$$V_k^{(n+1)} = \frac{\sum_i \int \Gamma_i^{(n)}(k, \phi) \mathbf{P}_\phi^{\mathbf{T}} C_i X_i \, d\phi}{\frac{\sigma^2}{T\tau^2} + \sum_i \int \Gamma_i^{(n)}(k, \phi) \mathbf{P}_\phi^{\mathbf{T}} C_i^2 \, d\phi}$$

.

2D Classification uses the same algorithm

Quantity	Meaning in 3D classification	Meaning in 2D classification
V_k	Class volume	Class average image
ϕ	3 Euler angles of orientation + 2 translations	1 angle of rotation + 2 translations
$\mathbf{P}_\phi$	Projection operator 3D$\longrightarrow$2D	Image rotation and shift
$\mathbf{P}_\phi^\mathbf{T}$	Back-projection operator 2D$\longrightarrow$3D	Reverse shift and rotation