Particle Classification

(Supervised and Unsupervised)

Steven Ludtke

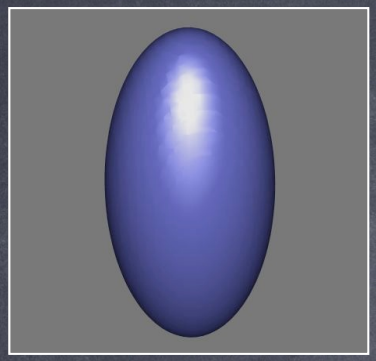
Professor, Biochemistry and Molecular Biology
Co-director, National Center for Macromolecular Imaging (P41)
Co-director, Center for Computational and Integrative Biomedical Research
Baylor College of Medicine

Baylor College of Medicine

VERNA & MARRS McLEAN DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

Supervised vs Unsupervised

- Supervised Classification (Reference-Based)
  - Used to determine particle orientations with respect to existing 3-D map
- Unsupervised Classification (Reference-Free)
  - Used to make 2-D class-averages without reference to 3-D map for:
    - Initial models
    - Conformational/Compositional Variability
    - Imagic -> 3-D



Initial 3D Model

Uniform Projections

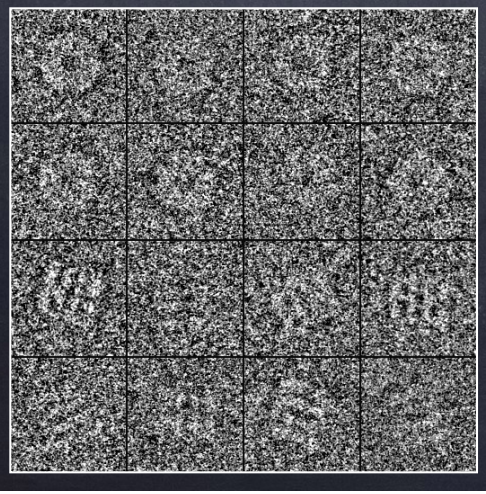
Build New 3D Model

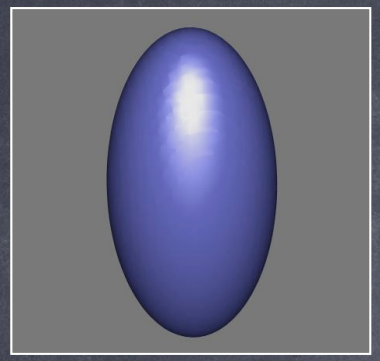
Final 3D Model

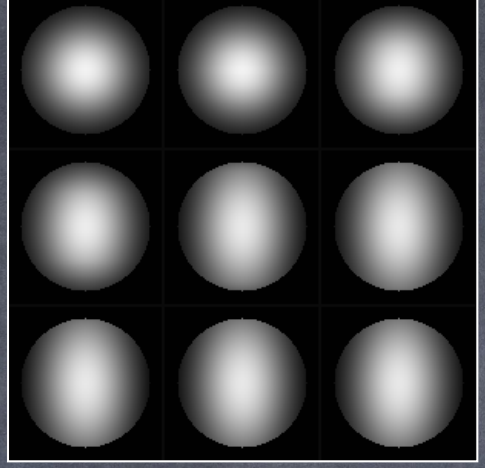
Particle Images

Classify Particles

Align and Average Classes







Initial 3D Model

Uniform Projections

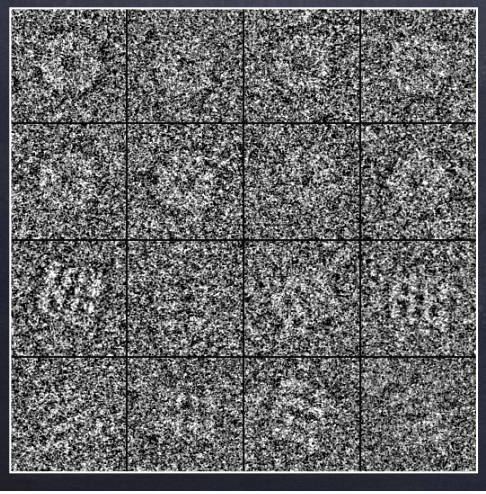
Build New 3D Model

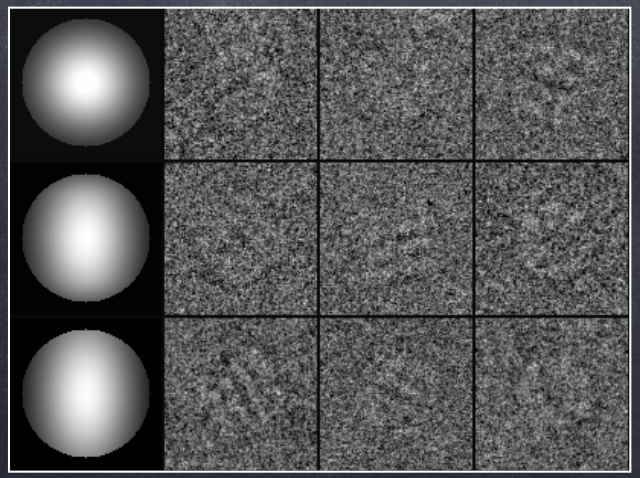
Final 3D Model

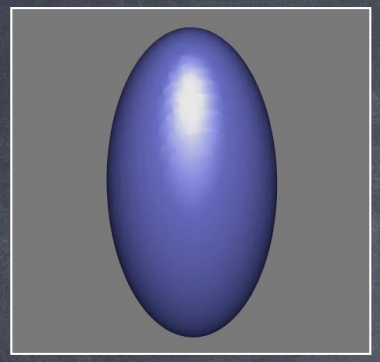
Particle Images

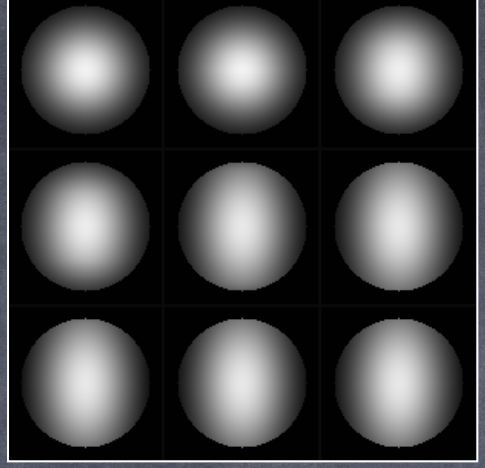
Classify Particles

Align and Average Classes









Initial 3D Model

Uniform Projections

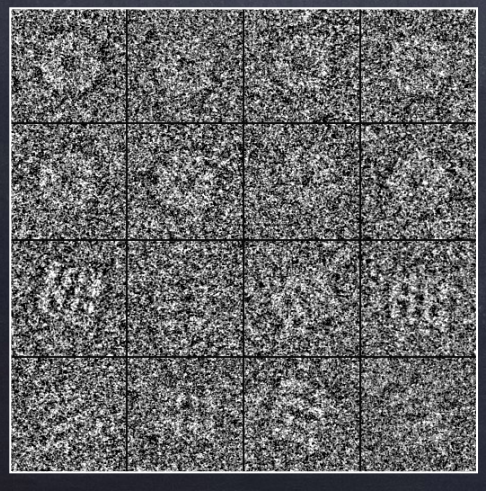
Build New 3D Model

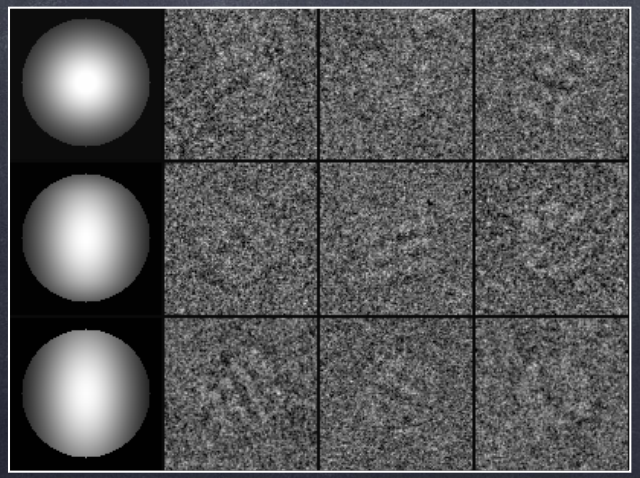
Final 3D Model

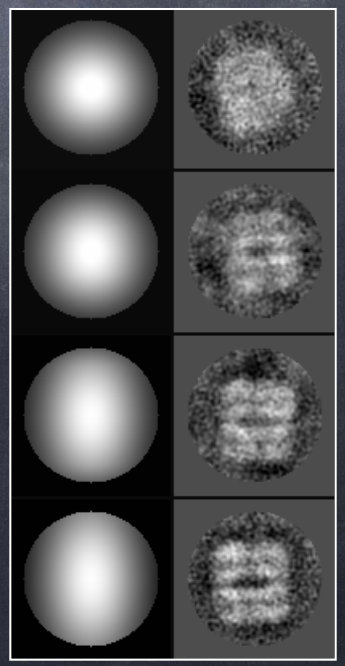
Particle Images

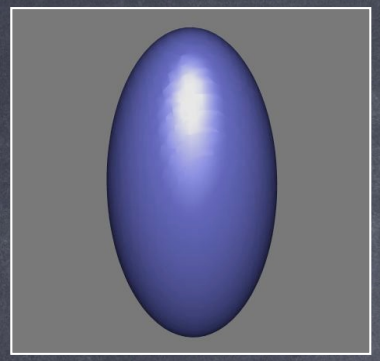
Classify Particles

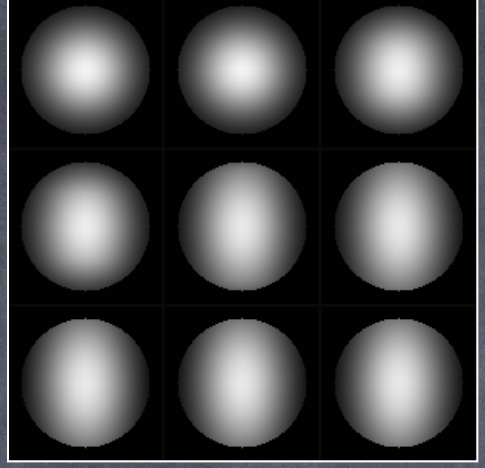
Align and Average Classes

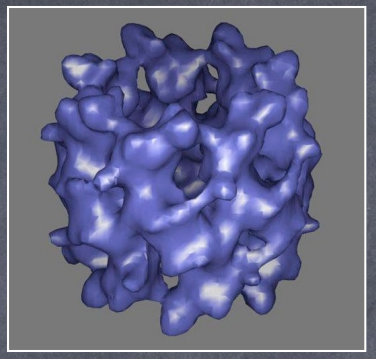












Initial 3D Model

Uniform Projections

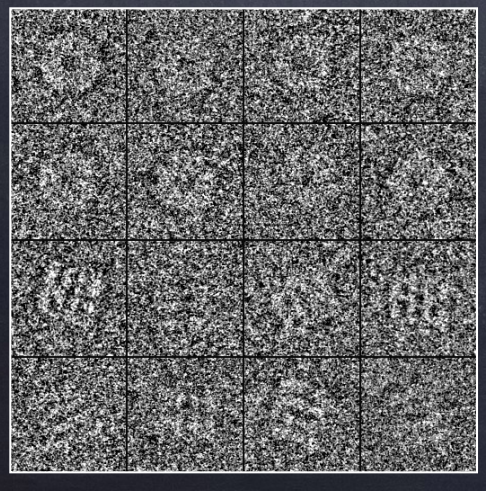
Build New 3D Model

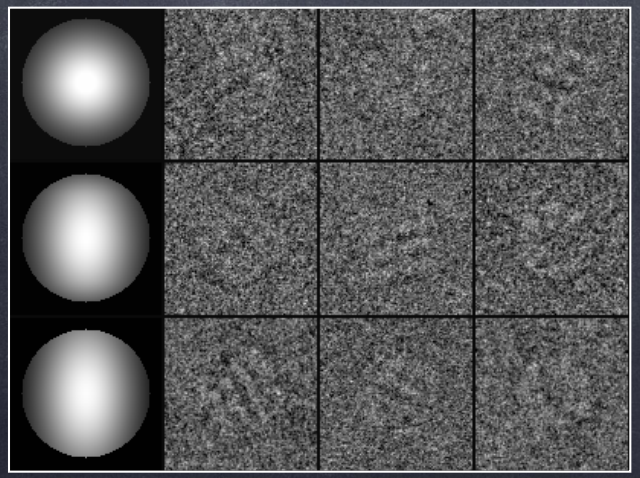
Final 3D Model

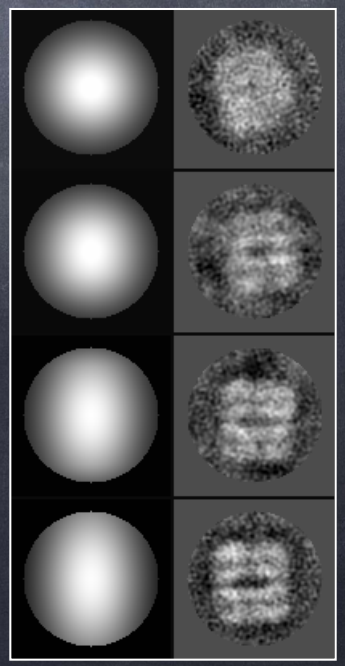
Particle Images

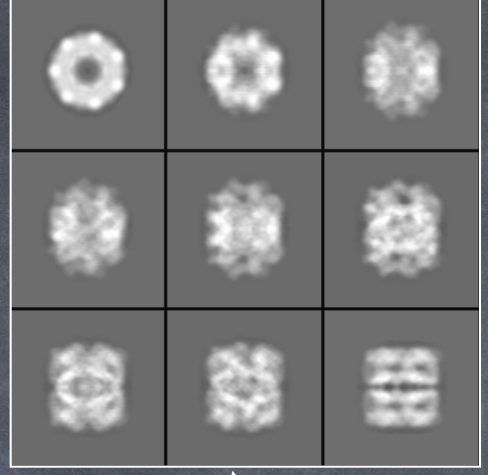
Classify Particles

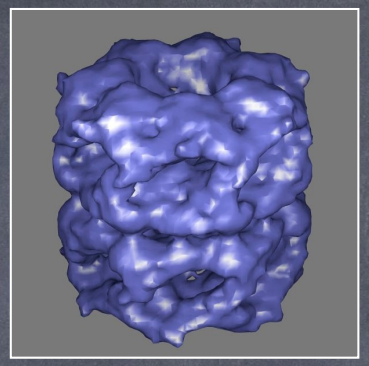
Align and Average Classes

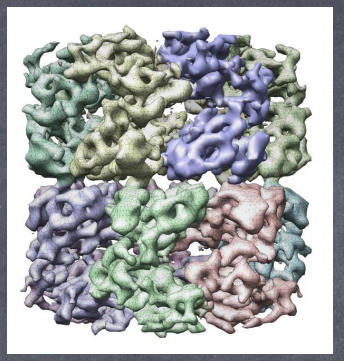












Initial 3D Model

Uniform Projections

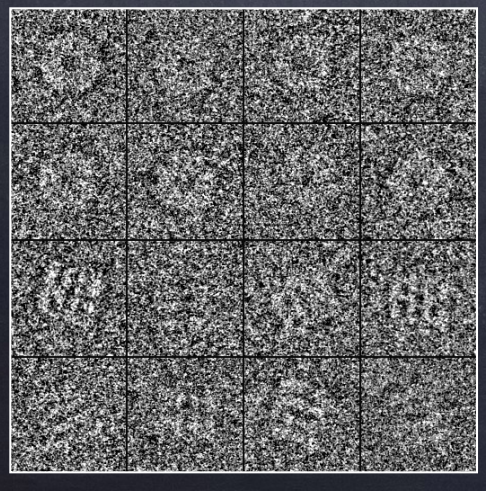
Build New 3D Model

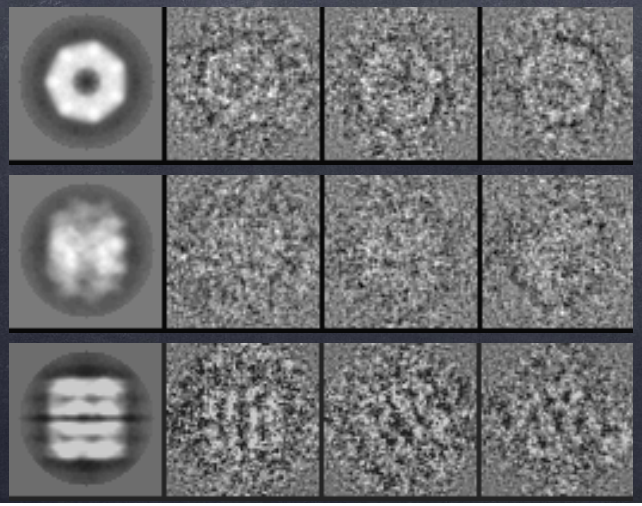
Final 3D Model

Particle Images

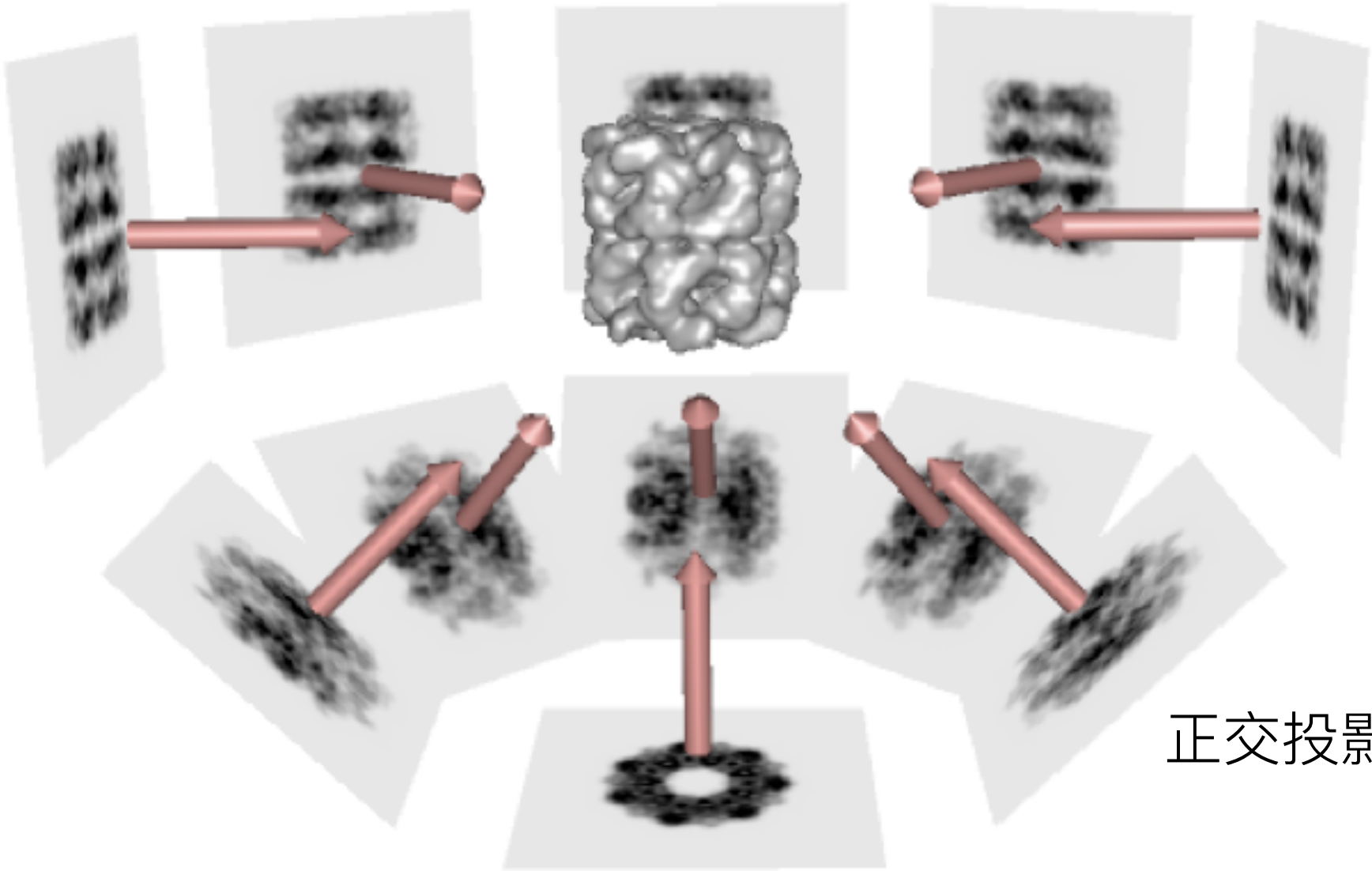
Classify Particles

Align and Average Classes



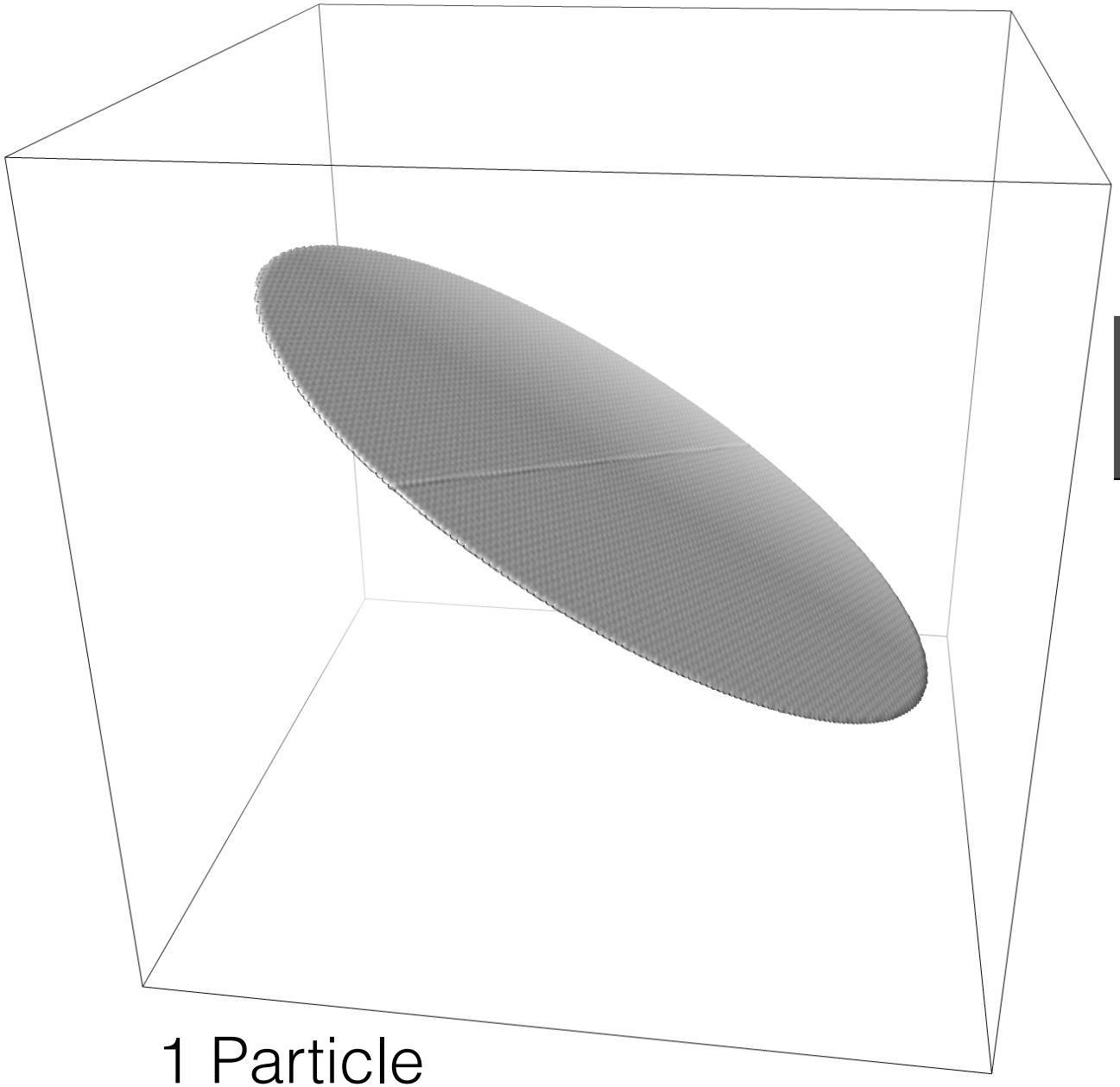


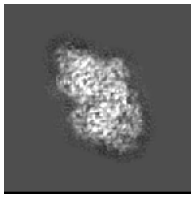
TEM Produces Projections



正交投影

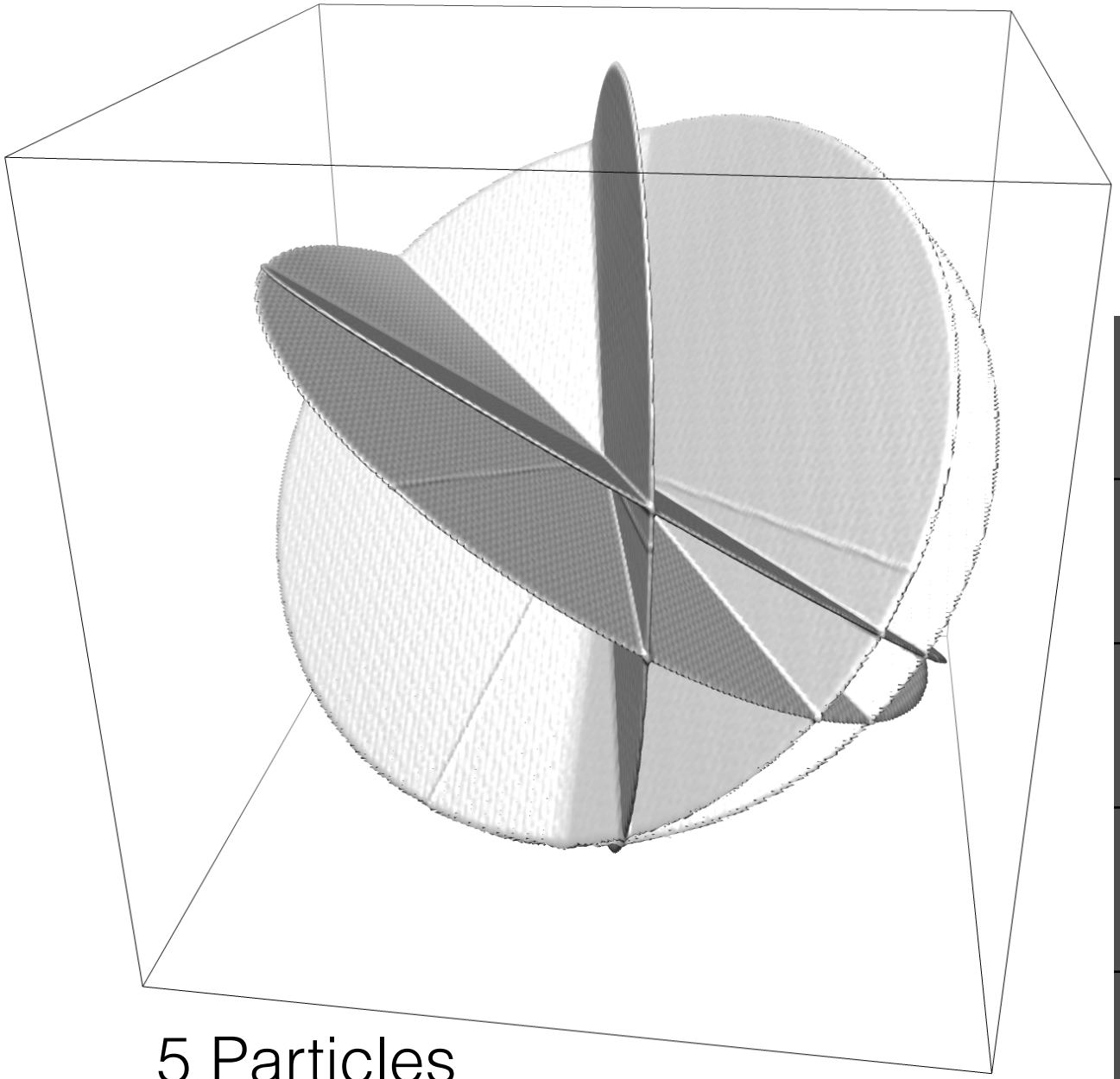
Fourier Space Reconstruction



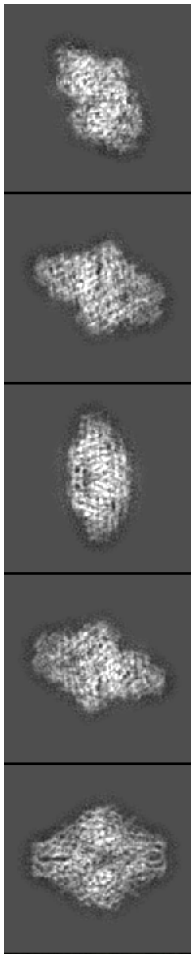


1 Particle

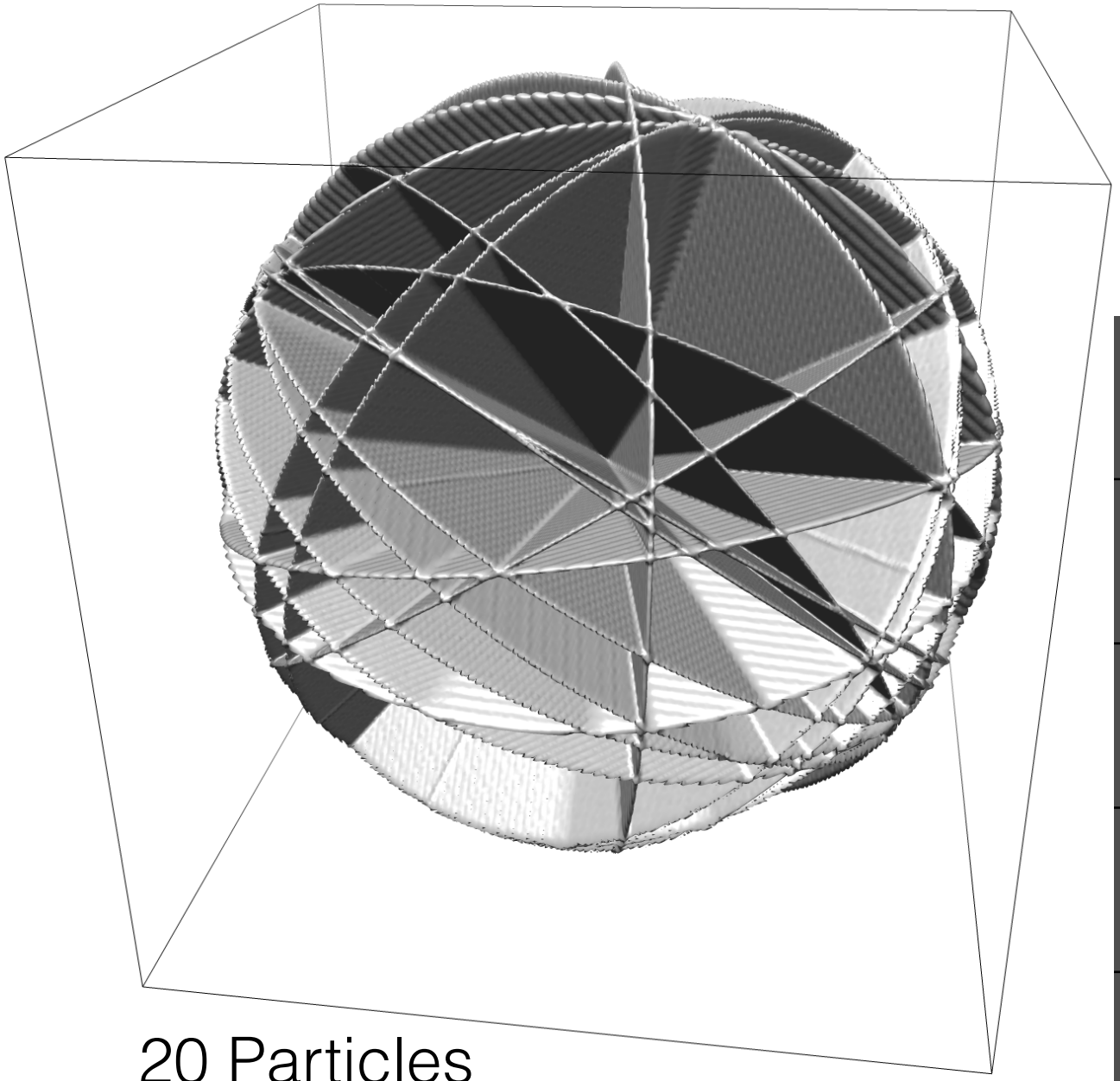
Fourier Space Reconstruction



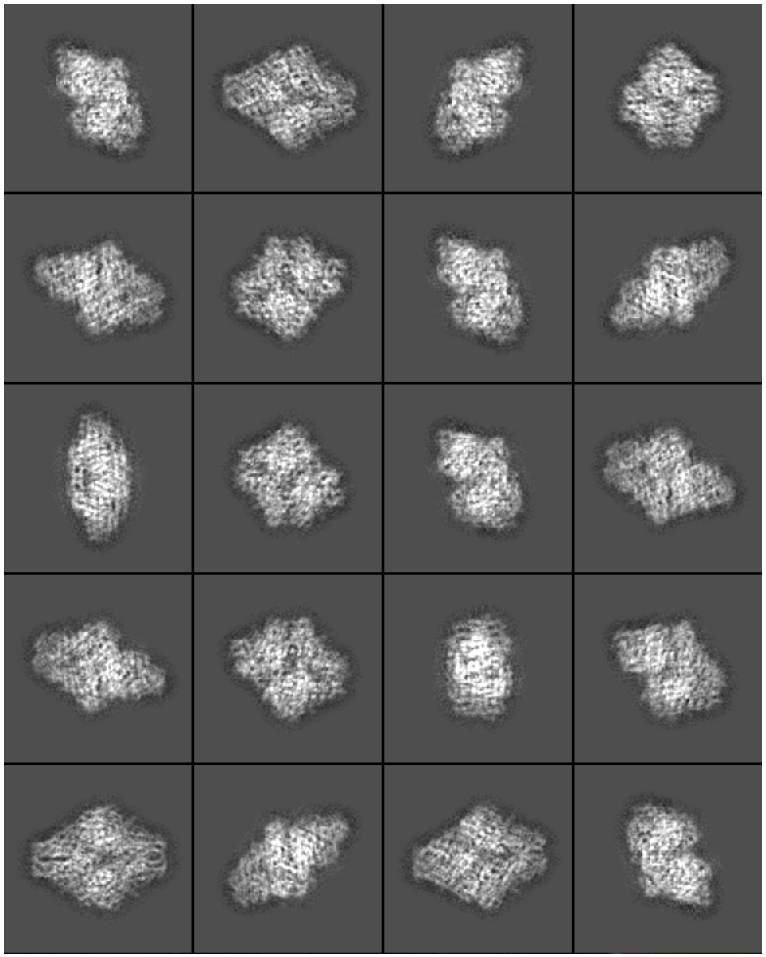
5 Particles



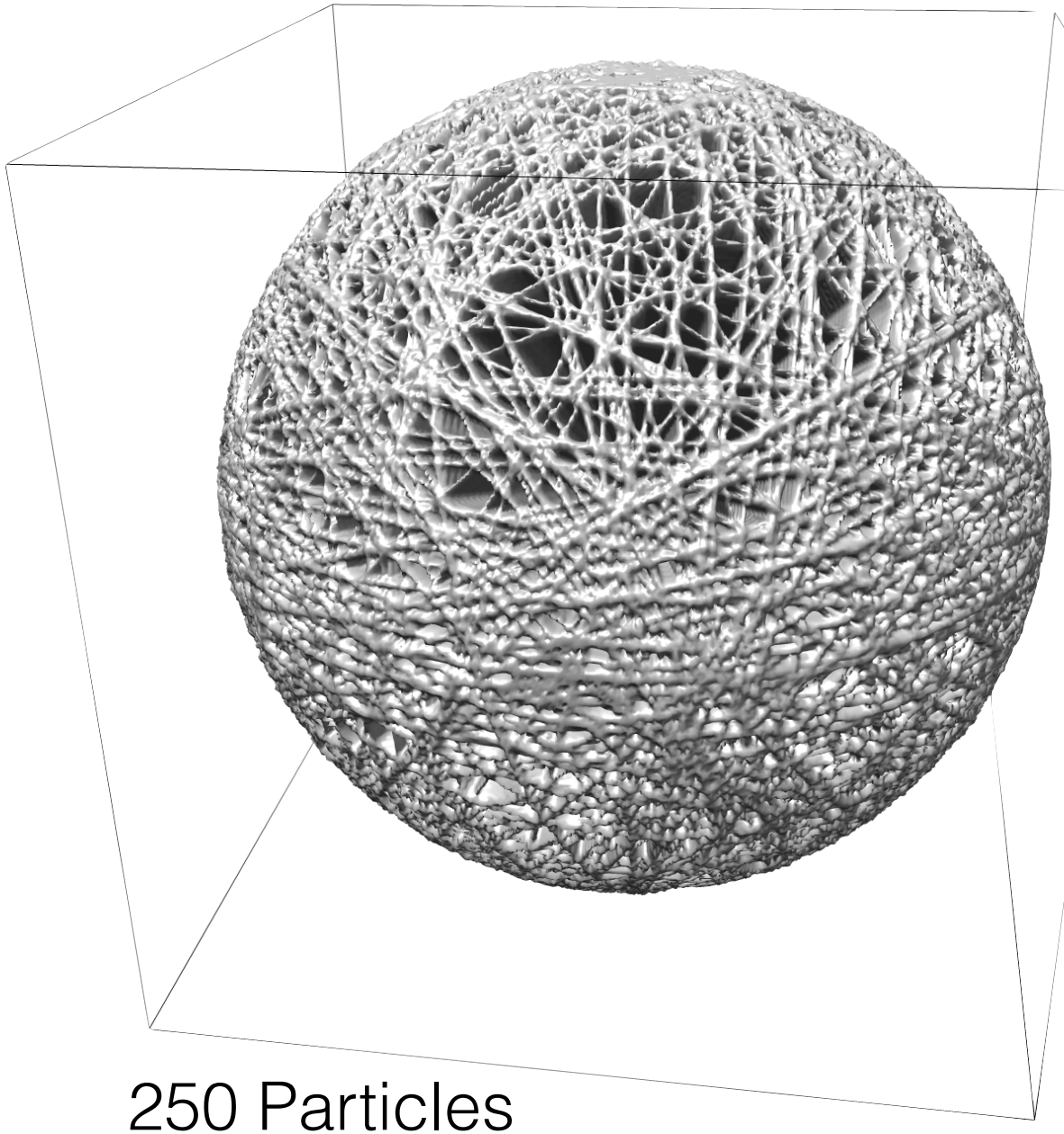
Fourier Space Reconstruction



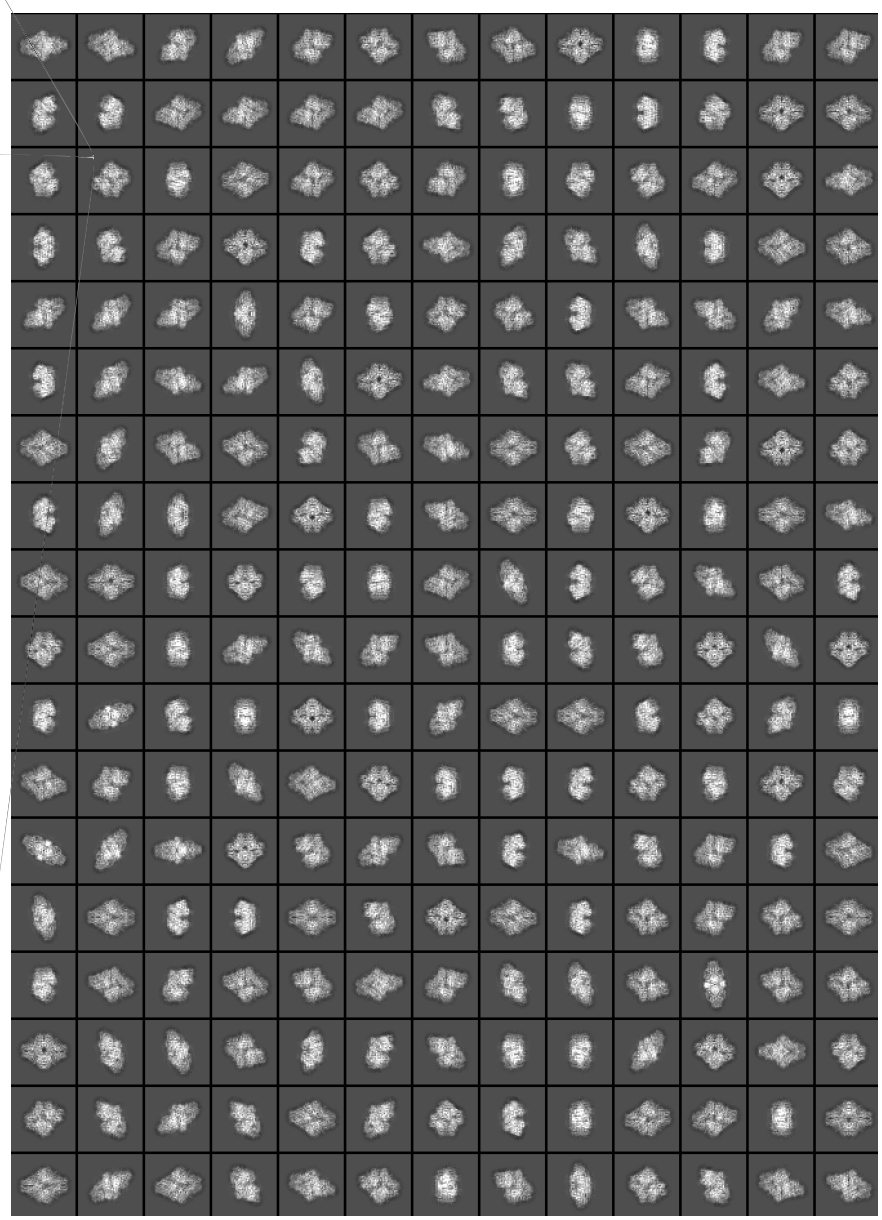
20 Particles

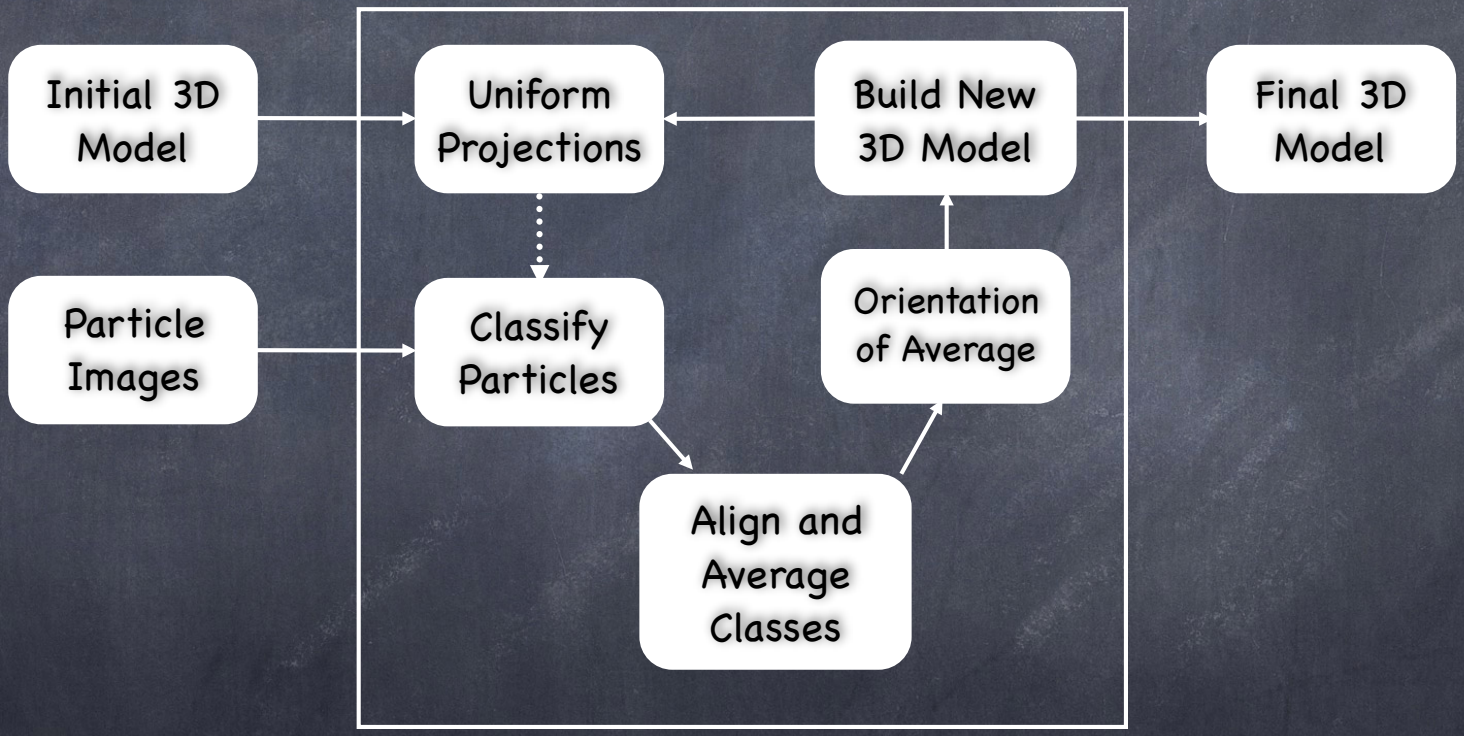


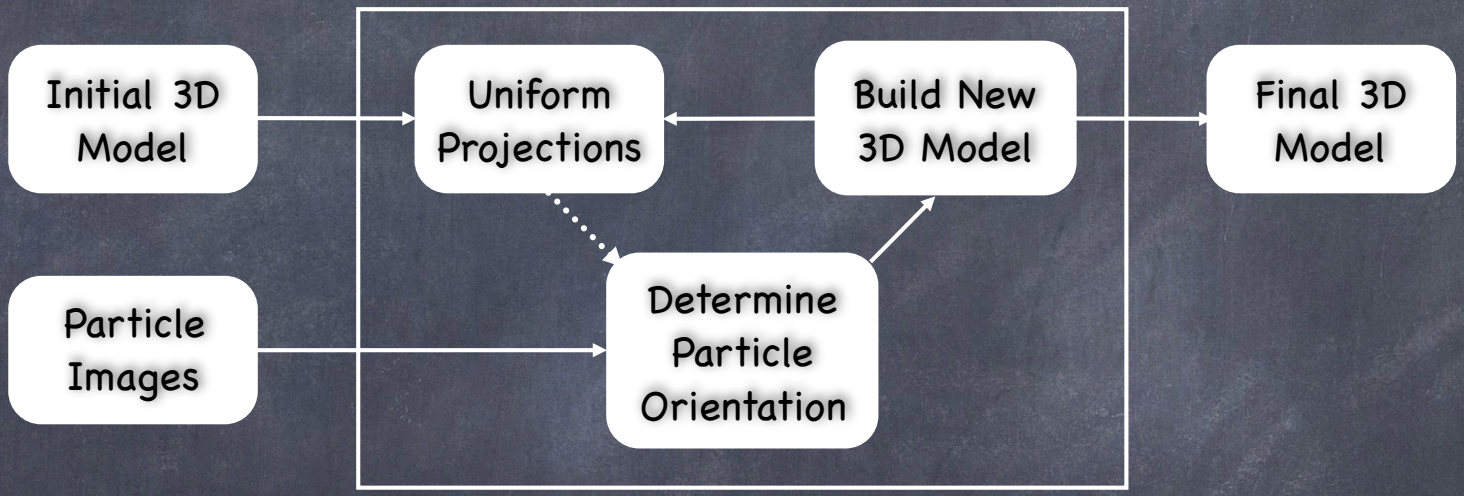
Fourier Space Reconstruction



250 Particles




Initial 3D Model
Uniform Projections
Build New 3D Model
Final 3D Model
Particle Images
Classify Particles
Orientation of Average
Align and Average Classes

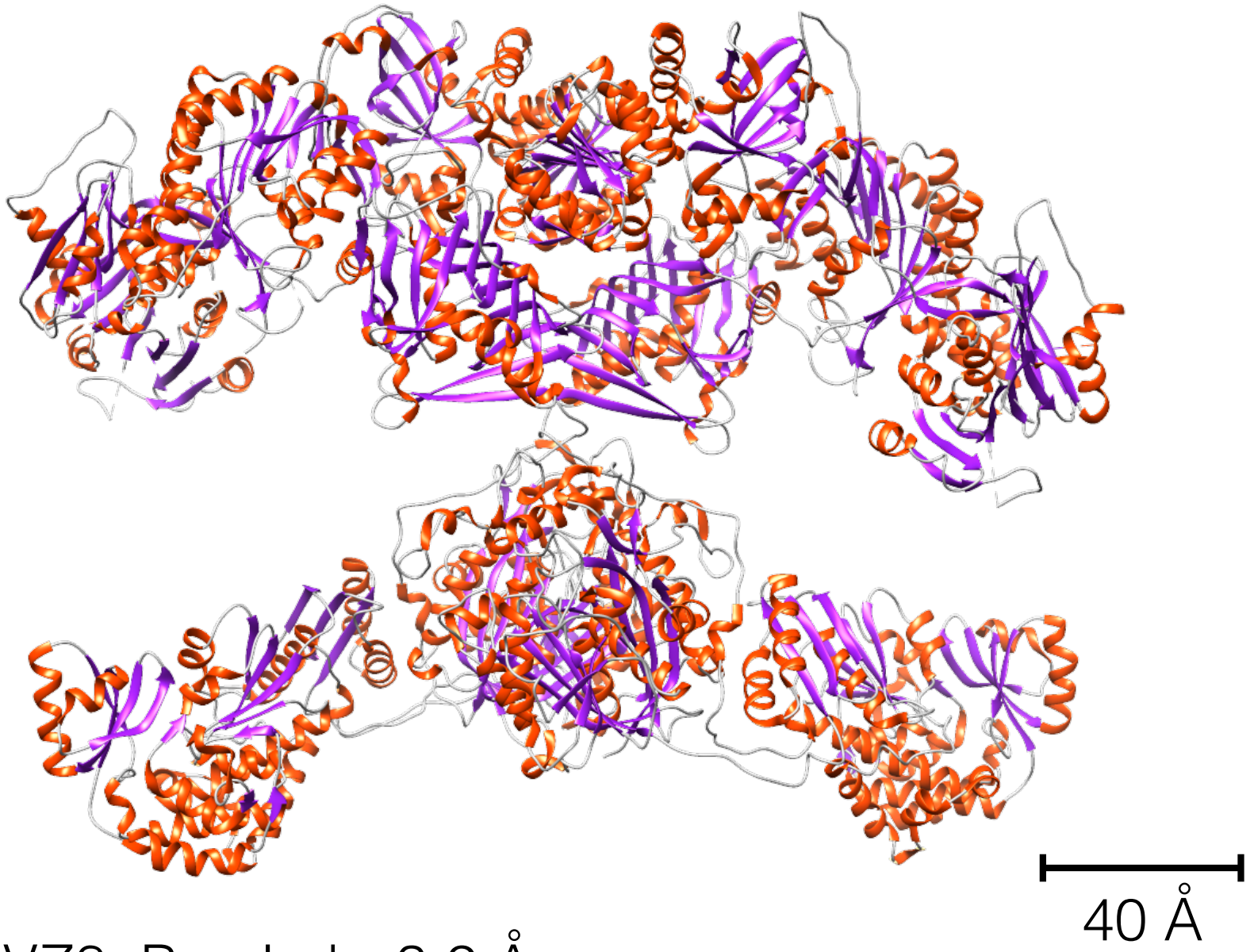

Initial 3D Model
Particle Images
Uniform Projections
Determine Particle Orientation
Build New 3D Model
Final 3D Model

Unsupervised Classification

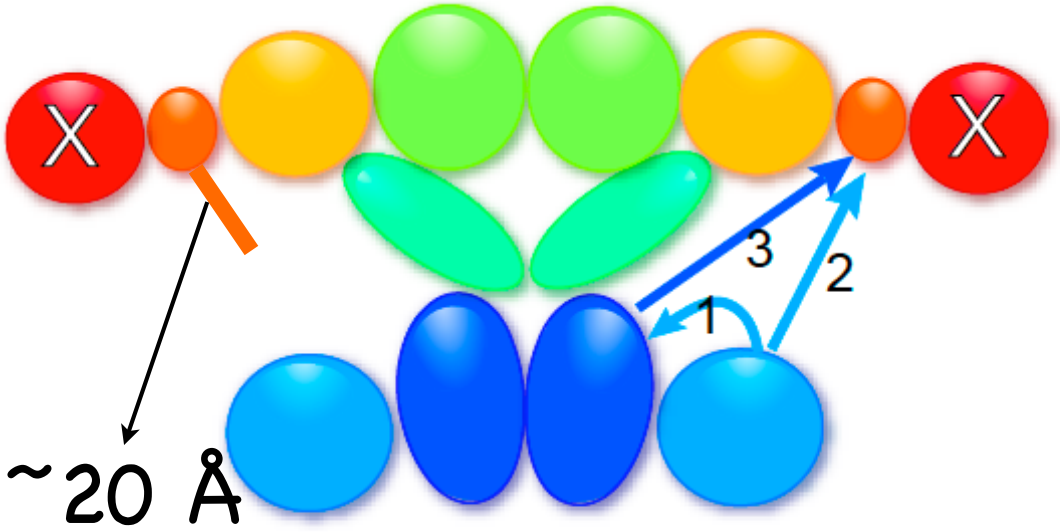
Fatty Acid Synthase

- Synthesizes saturated hydrocarbon chains
- Homo-dimer, ~500 kDa
- 7 Enzymatic activities
- Obesity
- Cancer
  - Highly expressed in most tumors
  - Particularly important in breast and prostate cancer
  - FAS inhibitors can prevent tumor growth

Fatty Acid Synthase

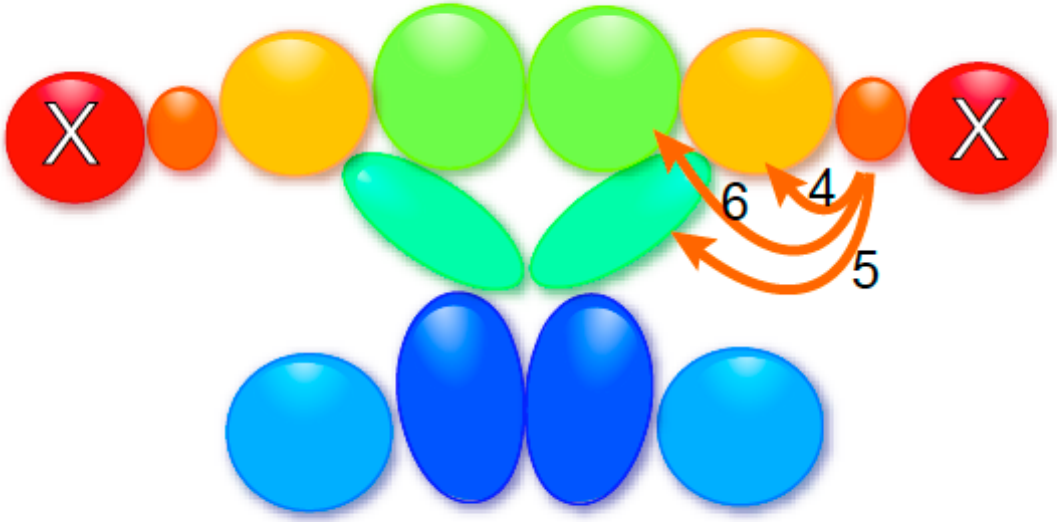


PDB: 2VZ8, Ban Lab, 3.2 Å



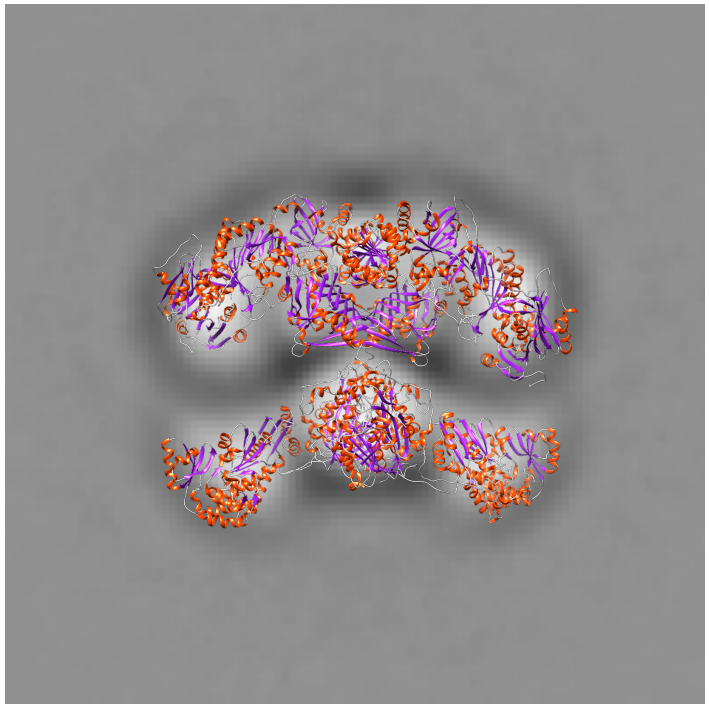
1. Loading Acetyl-CoA
2. Loading Malonyl-CoA
3. Condensation

~80 Å !?

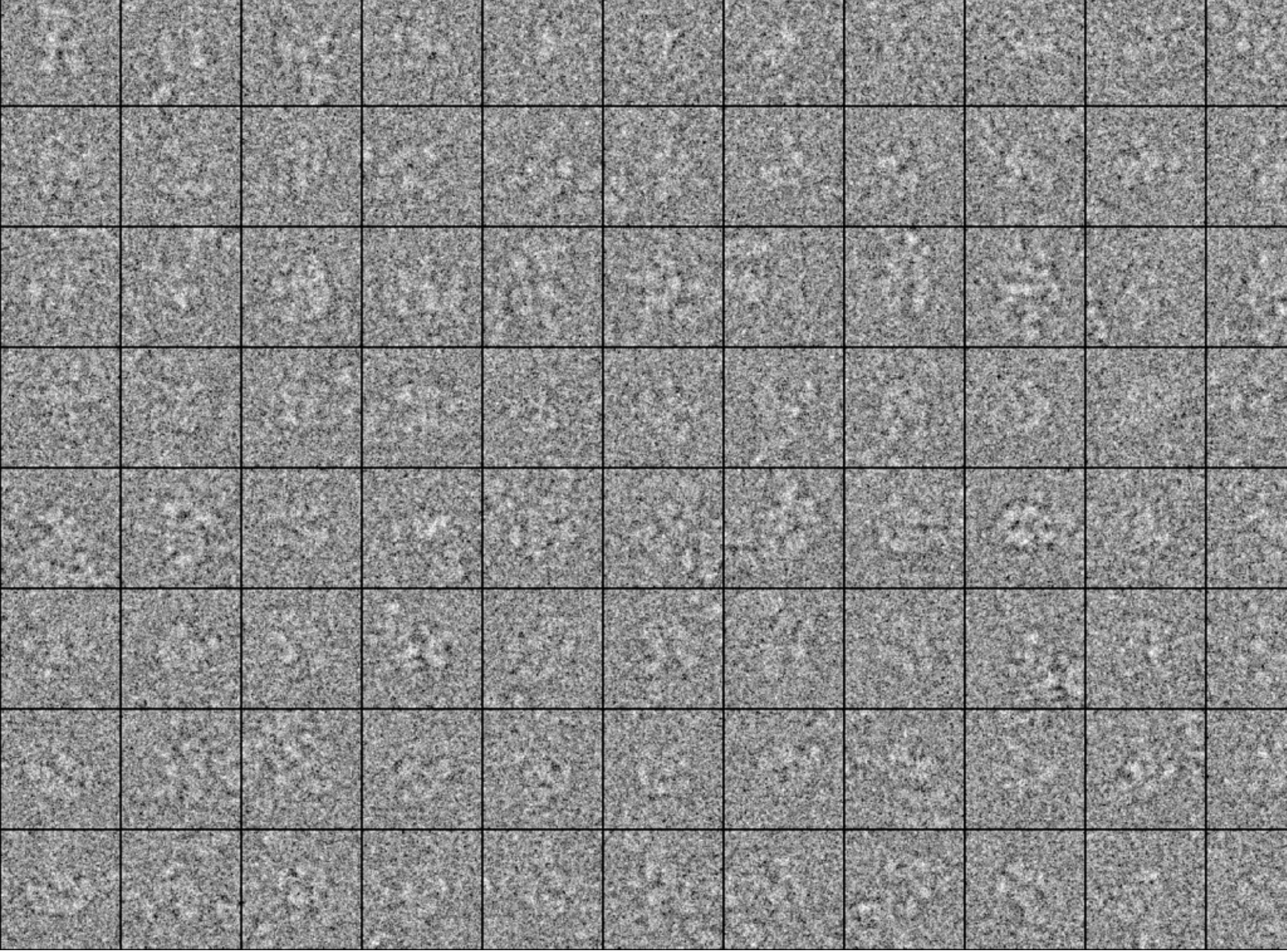


4. Keto reduction
5. Dehydration
6. Enoyl reduction

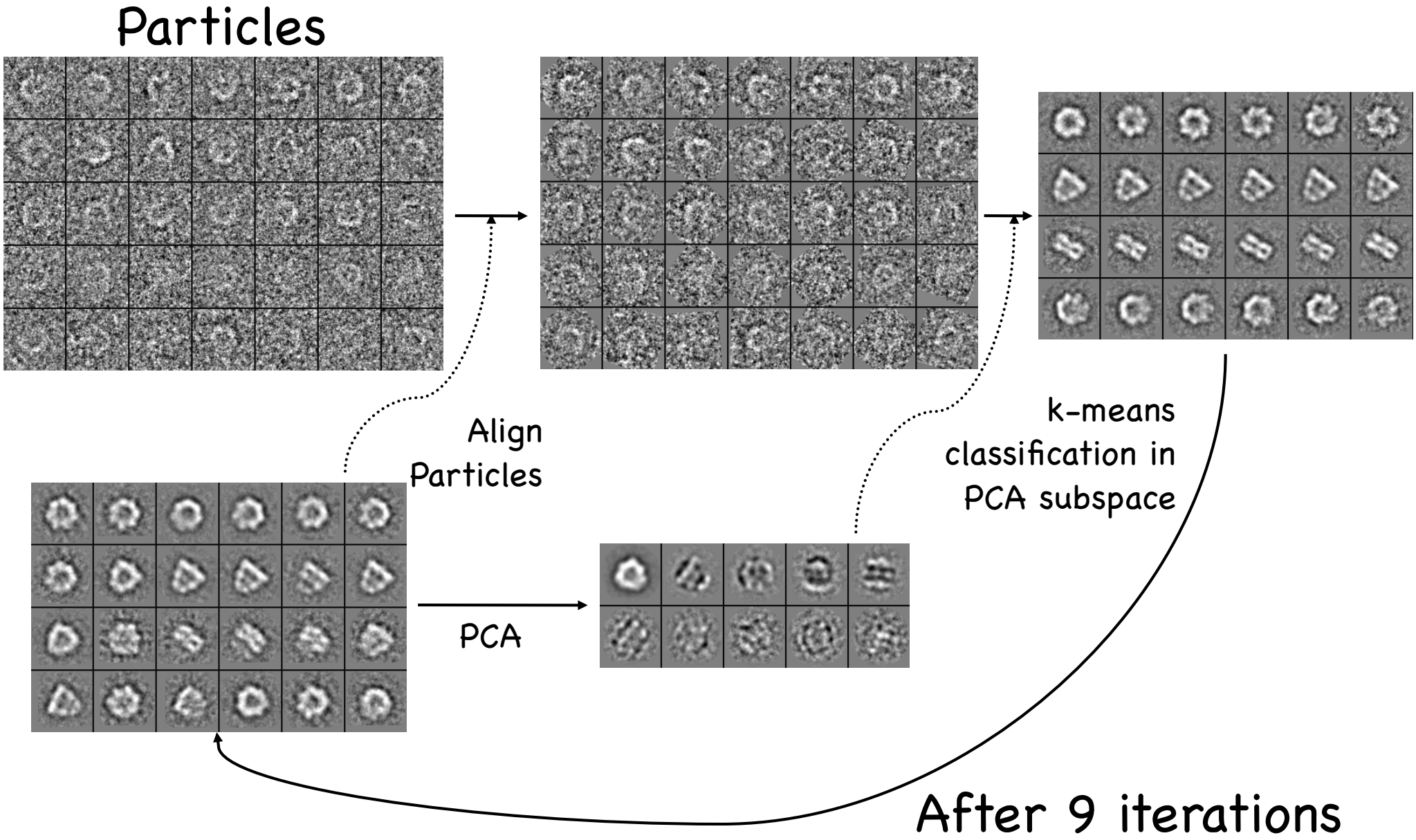
Single View CryoEM Average

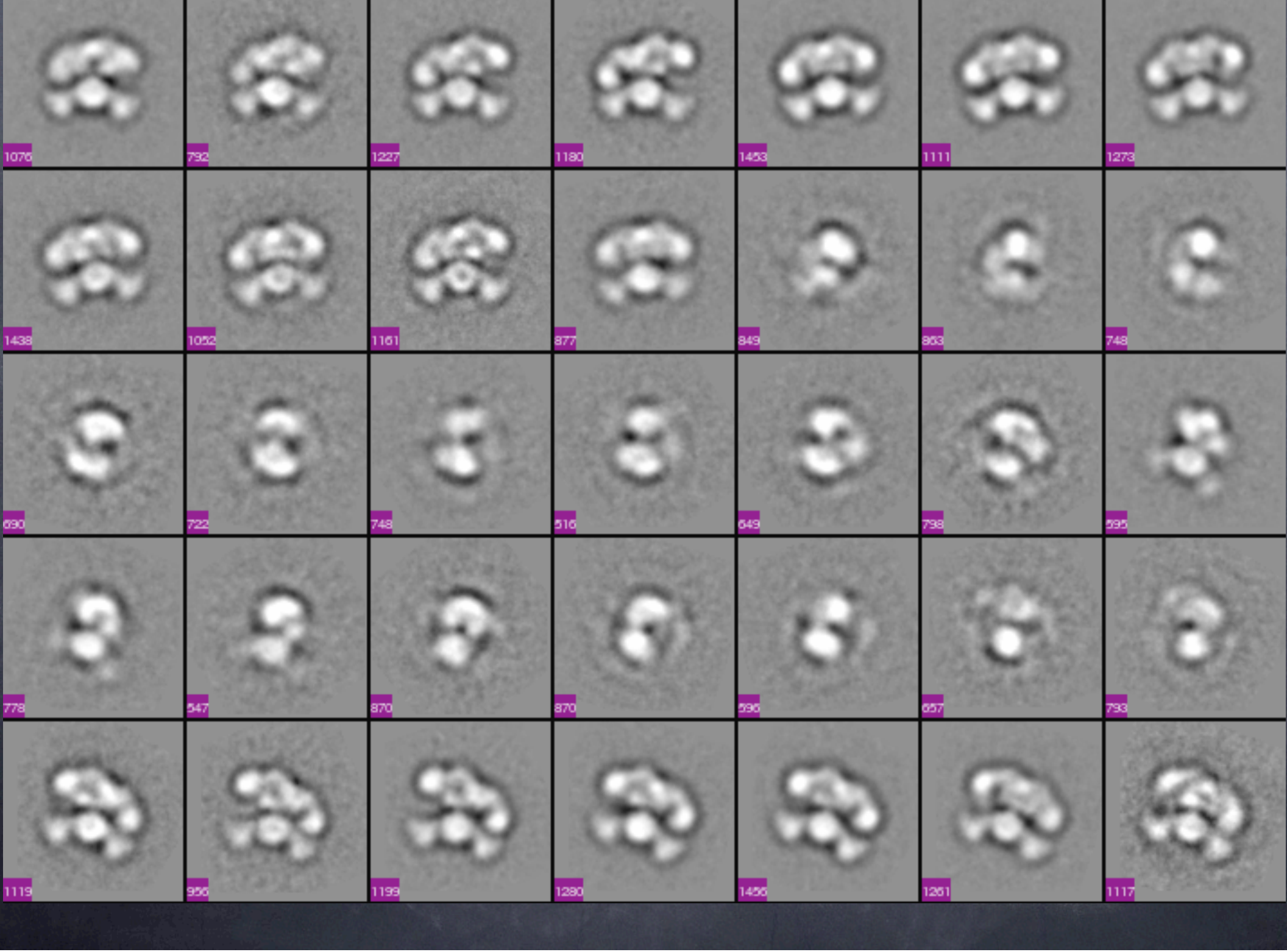


150 Å



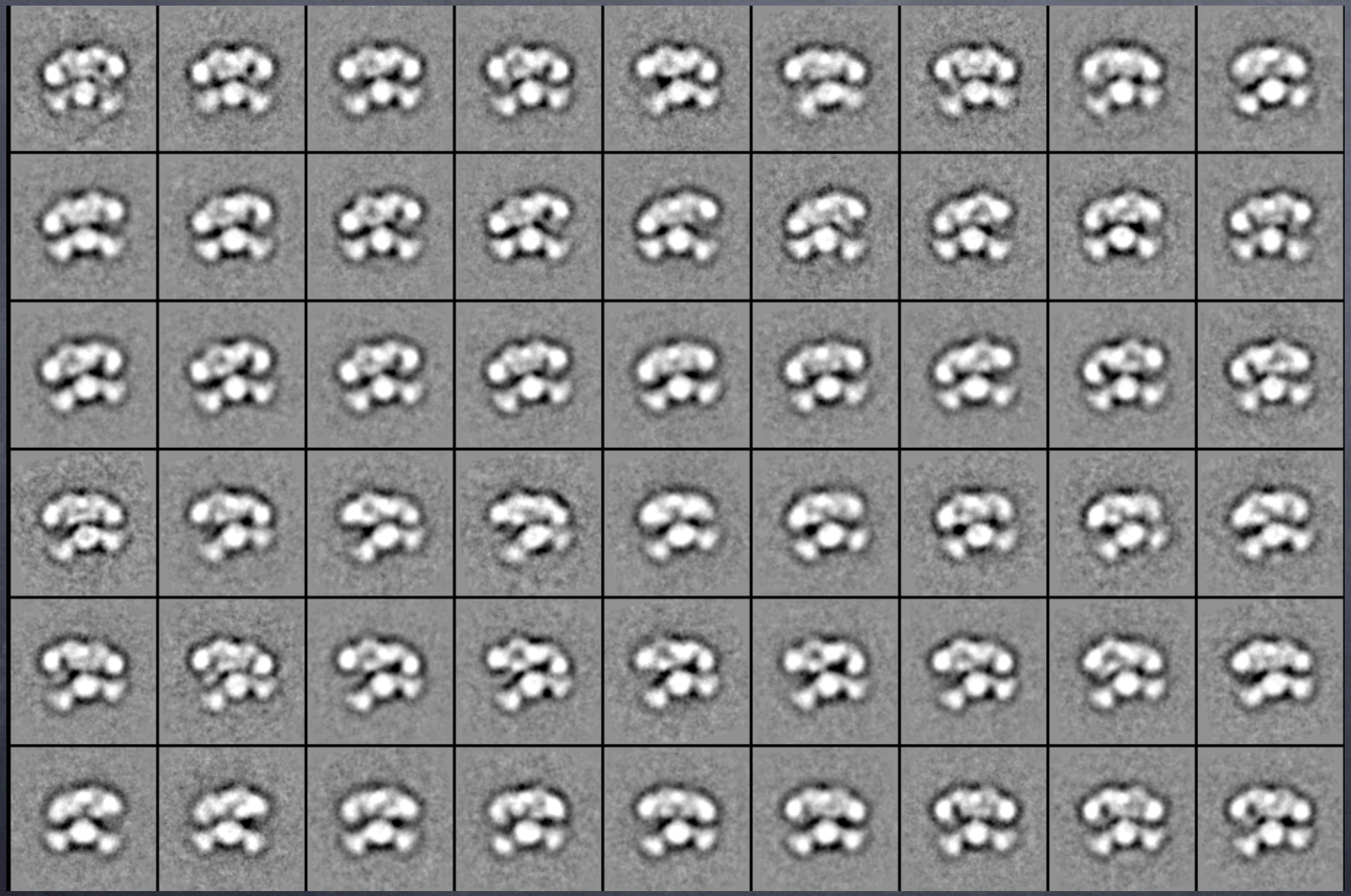
2-D Refinement



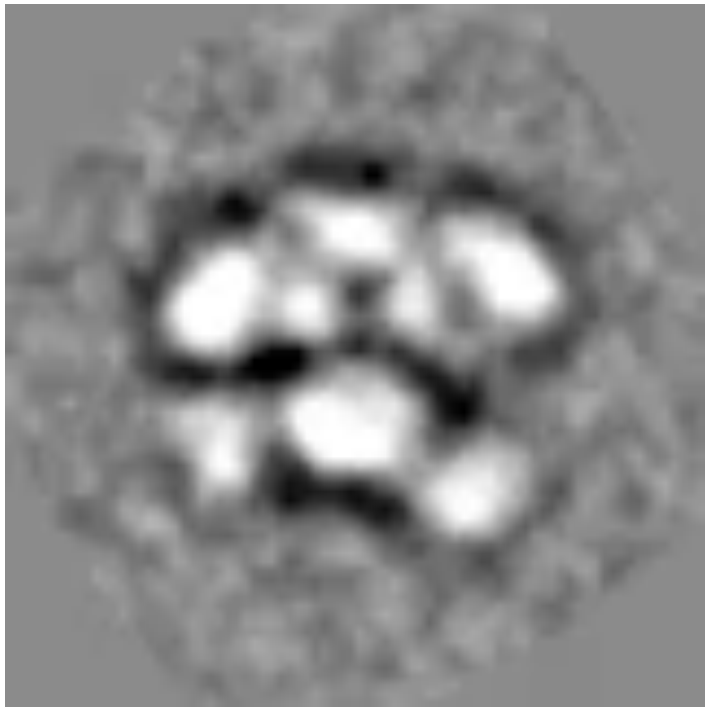

1076
792
1227
1180
1453
1111
1273
1438
1052
1161
877
849
803
748
690
722
748
516
649
798
595
778
547
870
870
596
657
793
1119
956
1199
1280
1456
1261
1117

Overall Process

- 2-D Reference-Free Classification
- Select particles in similar view
- 2-D Reference-Free Classification on subset
- Sort class-averages into order by similarity
- Make a movie

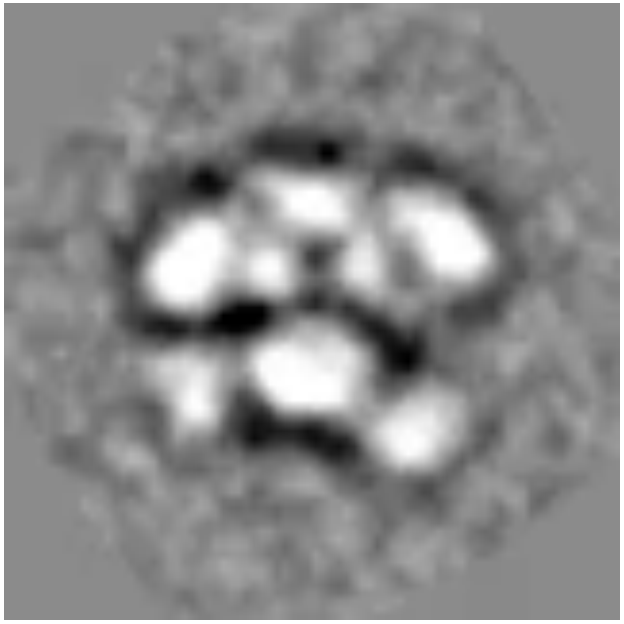


Fatty Acid Synthase, ~30 Å motion

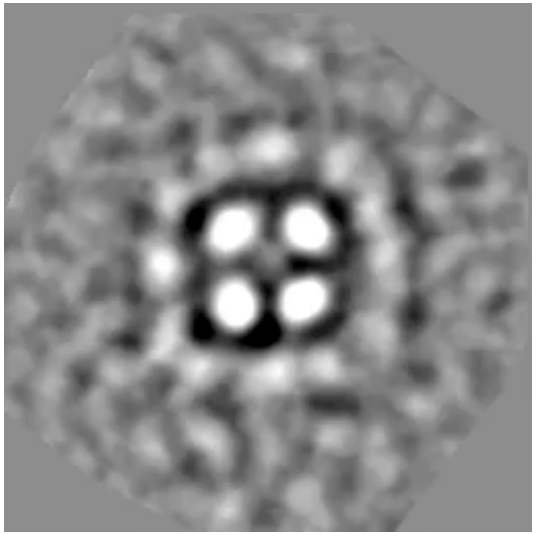


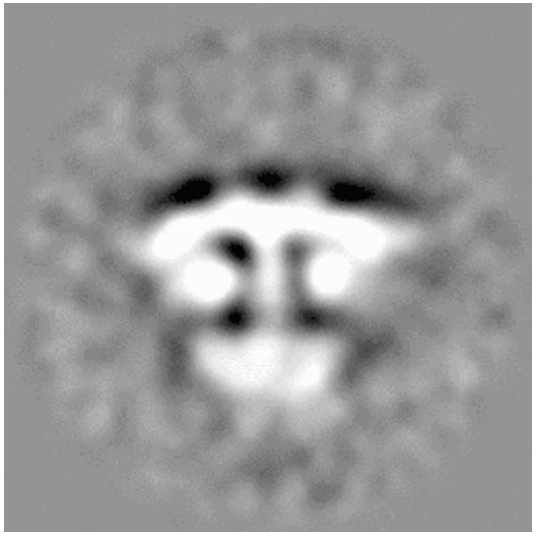
150 Å

Fatty Acid Synthase
~30 Å motion

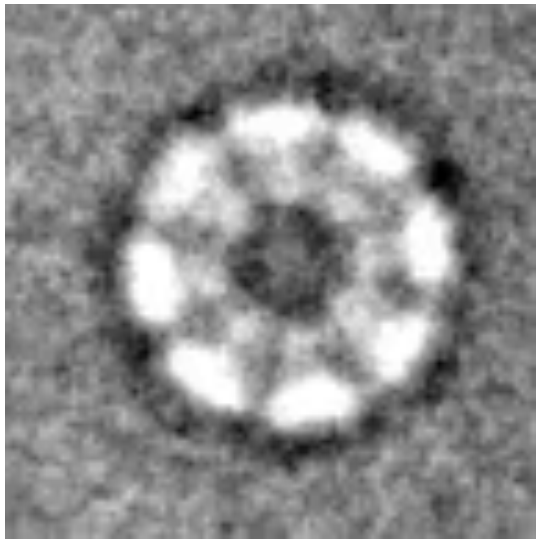


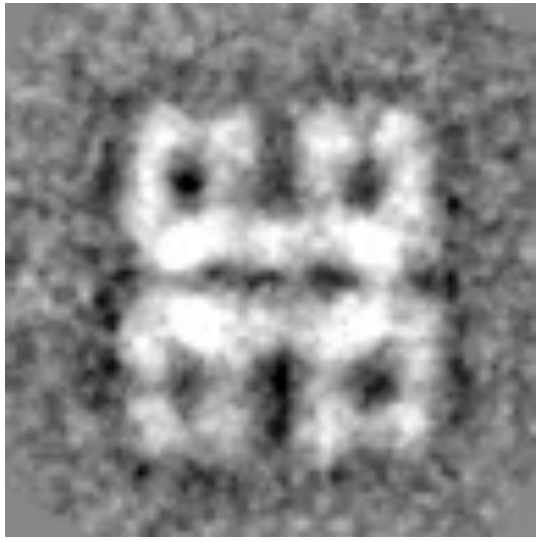
IP3 Receptor
~15 Å motion





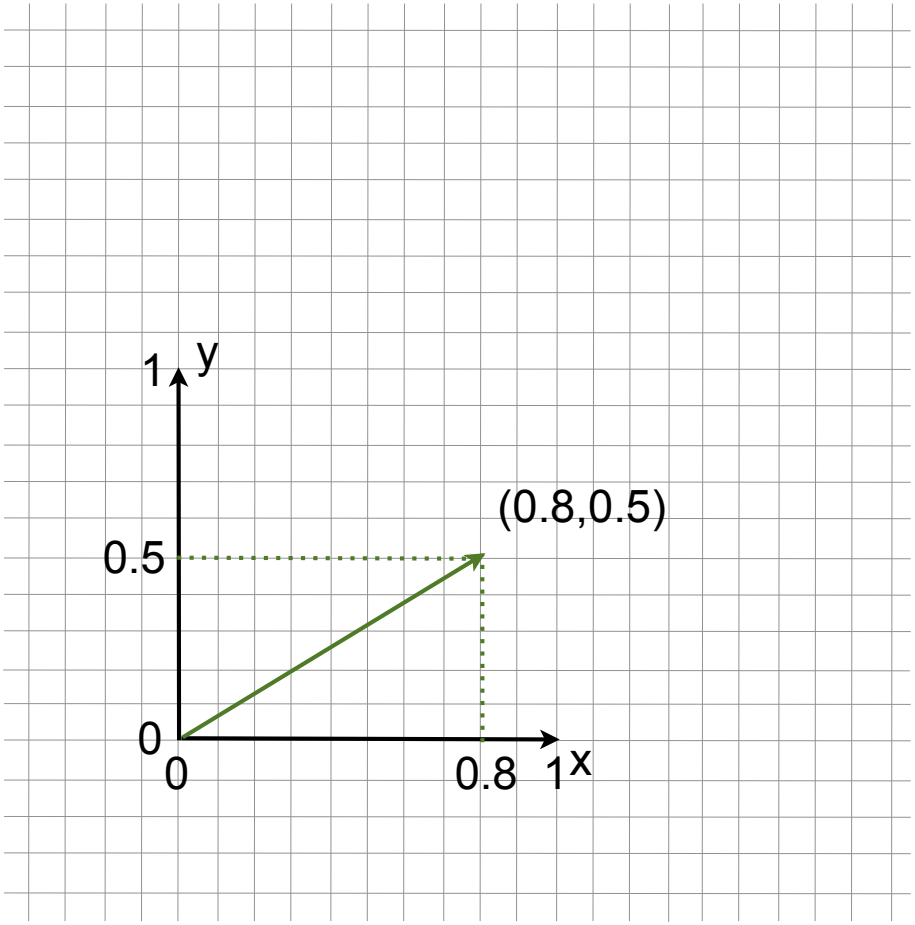
TriC (open)
~12 Å motion



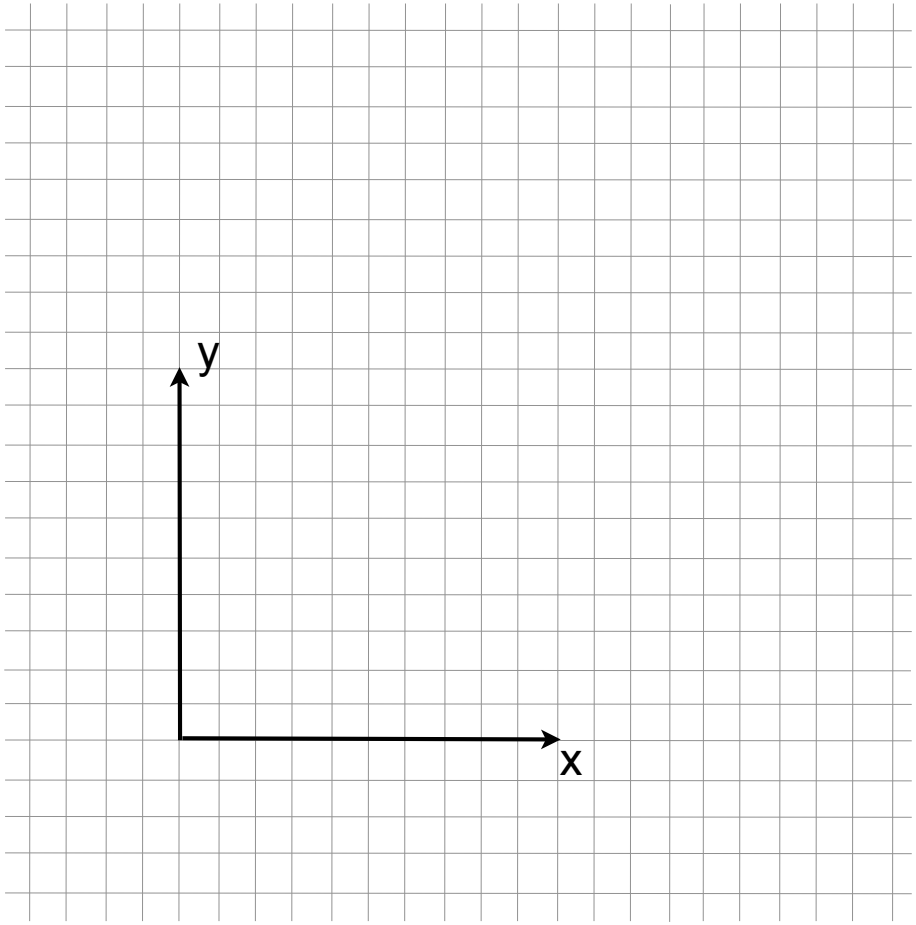


PCA

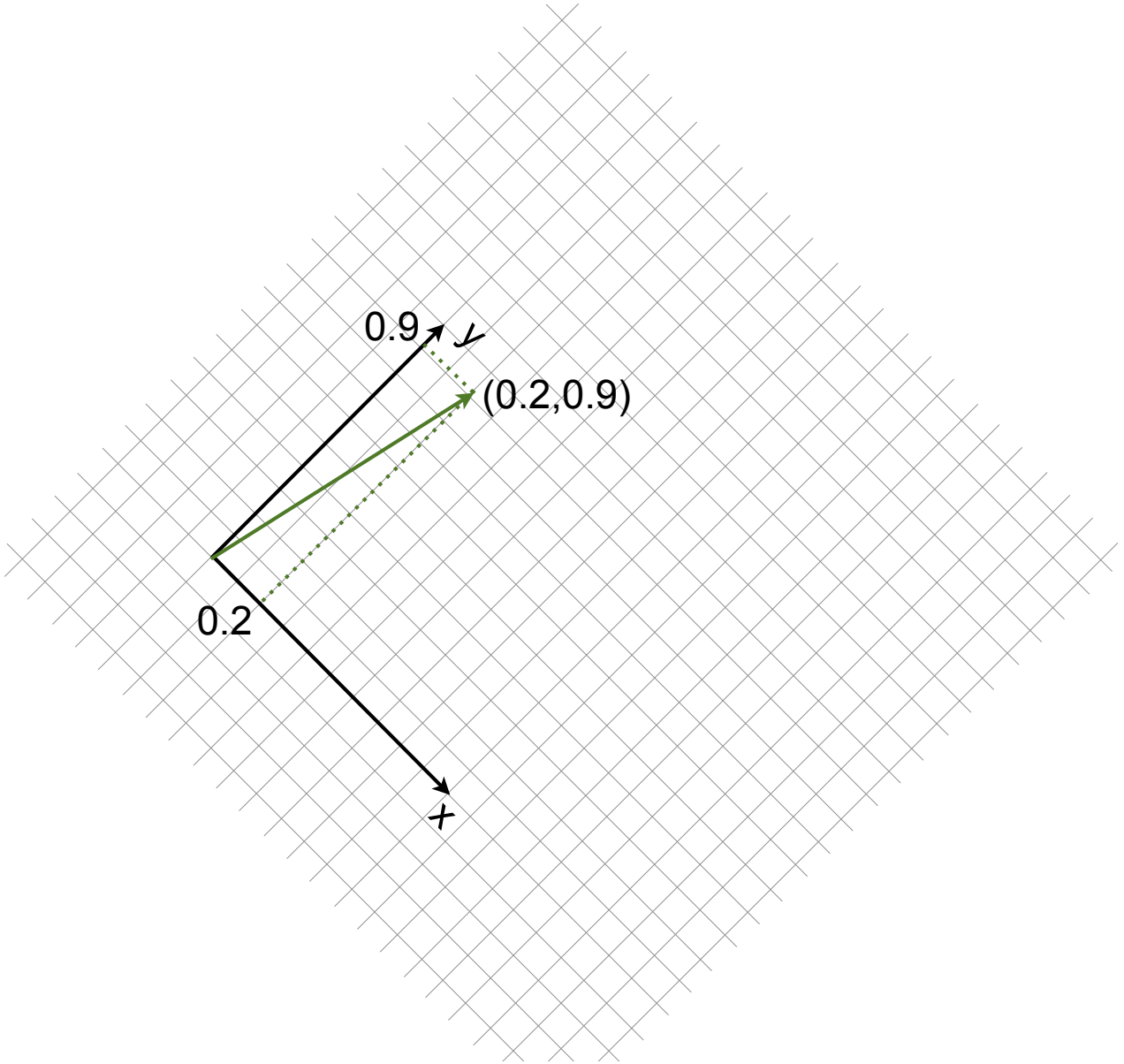
Orthogonality



Orthogonality



Orthogonality



2D Rotation Matrix

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

Discrete Data and N-space

- 2 independent values → point in 2-D
  - (x_0,x_1)
- 3 independent values → point in 3-D
  - (x_0,x_1,x_2)
- n independent values → point in n-D space
  - $(x_0,x_1,x_2,\ldots,x_n)$

2D Rotation Matrix

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

1 angle to specify

3D Rotation Matrix

$$\begin{pmatrix} \cos(\theta)\cos(\omega) - \cos(\phi)\sin(\theta)\sin(\omega) & -\cos(\phi)\cos(\omega)\sin(\theta) - \cos(\theta)\sin(\omega) & \sin(\theta)\sin(\phi) \\ \cos(\omega)\sin(\theta) + \cos(\theta)\cos(\phi)\sin(\omega) & \cos(\theta)\cos(\phi)\cos(\omega) - \sin(\theta)\sin(\omega) & -\cos(\theta)\sin(\phi) \\ \sin(\phi)\sin(\omega) & \cos(\omega)\sin(\phi) & \cos(\phi) \end{pmatrix}$$

3 angles to specify (Euler angles or quaternions in most cases)

4D Rotation Matrices

$$rotXY = \begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 & 0 \\ -\sin(\theta) & \cos(\theta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$rotYZ = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & \sin(\theta) & 0 \\ 0 & -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$rotXZ = \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) & 0 \\ 0 & 1 & 0 & 0 \\ \sin(\theta) & 0 & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$rotXU = \begin{bmatrix} \cos(\theta) & 0 & 0 & \sin(\theta) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\sin(\theta) & 0 & 0 & \cos(\theta) \end{bmatrix}$$

$$rotYU = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 0 & 1 & 0 \\ 0 & \sin(\theta) & 0 & \cos(\theta) \end{bmatrix}$$

$$rotZU = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos(\theta) & -\sin(\theta) \\ 0 & 0 & \sin(\theta) & \cos(\theta) \end{bmatrix}$$

6 angles to specify in 4D space
(in general n(n-1)/2)

Real Space Basis

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

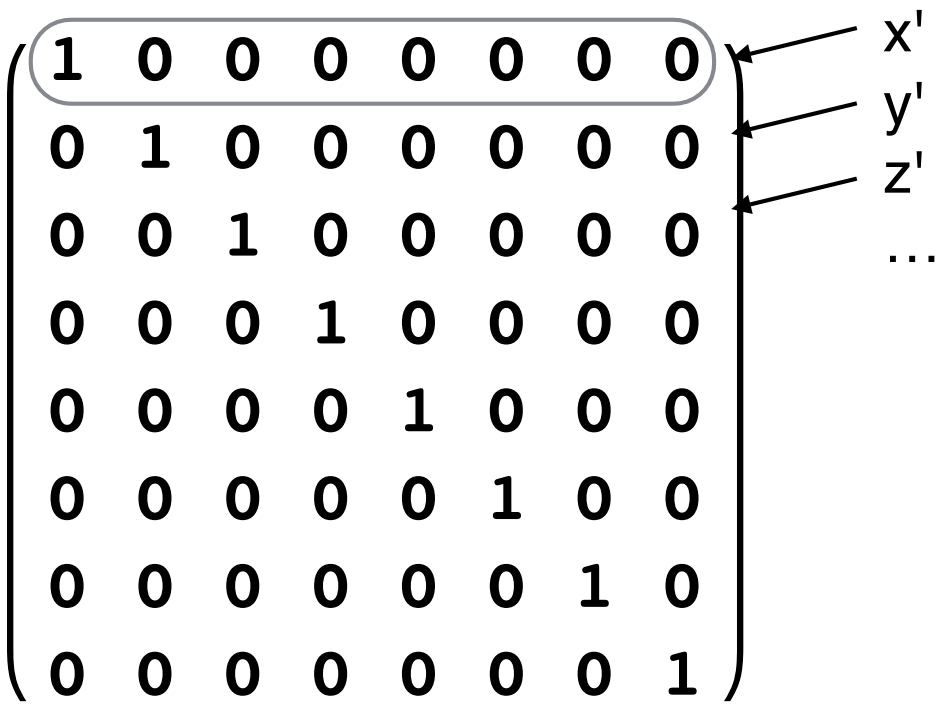
Orthonormal Transforms

Preserve orthogonality and normality:

- Rotation
- Reflection (mirroring)
- Permutation of Axes

- Inverse = Transpose!
- Vectors have unit length, otherwise just orthogonal

Real Space Basis



Non-orthogonal Matrix

$$\frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

Non-orthogonal Matrix Inverse

$$\frac{1}{\sqrt{3}} \begin{pmatrix} -1 & 2 & -1 & -1 & 2 & -1 & -1 & 2 \\ 2 & -1 & 2 & -1 & -1 & 2 & -1 & -1 \\ -1 & 2 & -1 & 2 & -1 & -1 & 2 & -1 \\ -1 & -1 & 2 & -1 & 2 & -1 & -1 & 2 \\ 2 & -1 & -1 & 2 & -1 & 2 & -1 & -1 \\ -1 & 2 & -1 & -1 & 2 & -1 & 2 & -1 \\ -1 & -1 & 2 & -1 & -1 & 2 & -1 & 2 \\ 2 & -1 & -1 & 2 & -1 & -1 & 2 & -1 \end{pmatrix}$$

FFT

(finite discrete)

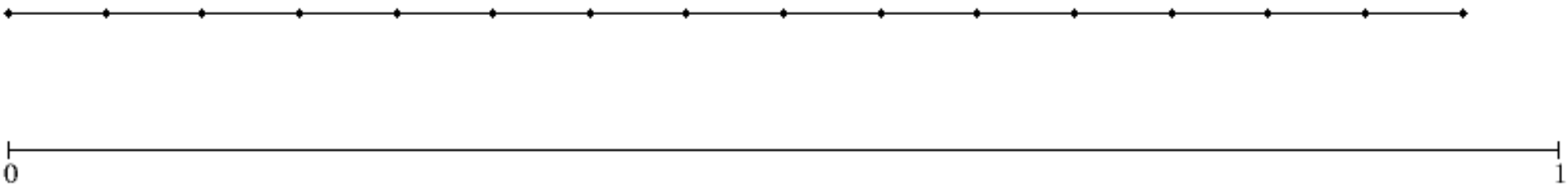
$$\hat{F}_k = \frac{1}{\sqrt{n}} \sum_{x=0}^{n-1} f_x \, e^{-i \, 2 \, \pi \, k \, x/n}$$

k, x and n are integers!

Orthonormality

- All of the basis vectors in the transformation matrix must be orthogonal (and unit length)!

FFT Basis



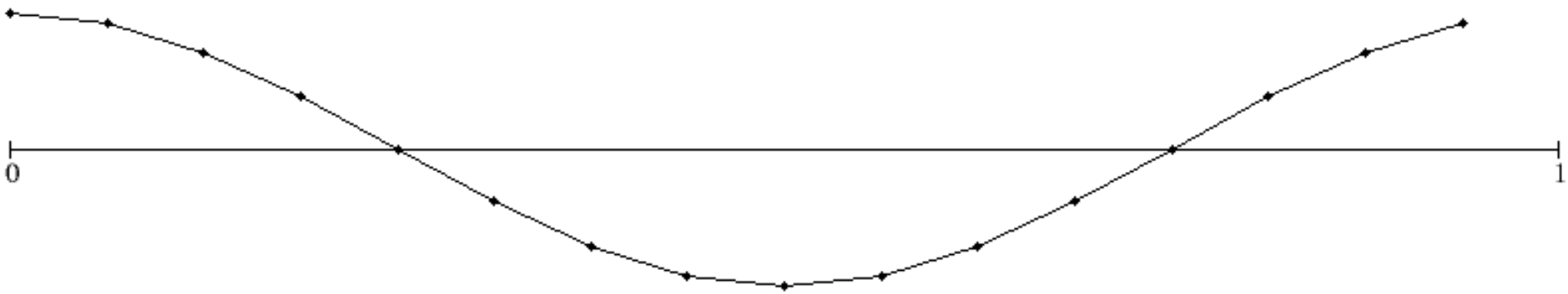

```
1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
```

FFT of curve above

```
R I R I R I R I R I R I R I R I R I
0 0 1 1 2 2 3 3 4 4 5 5 6 6 7 7 8 8
```

cos(kx) k=0 -> 1.0

FFT Basis

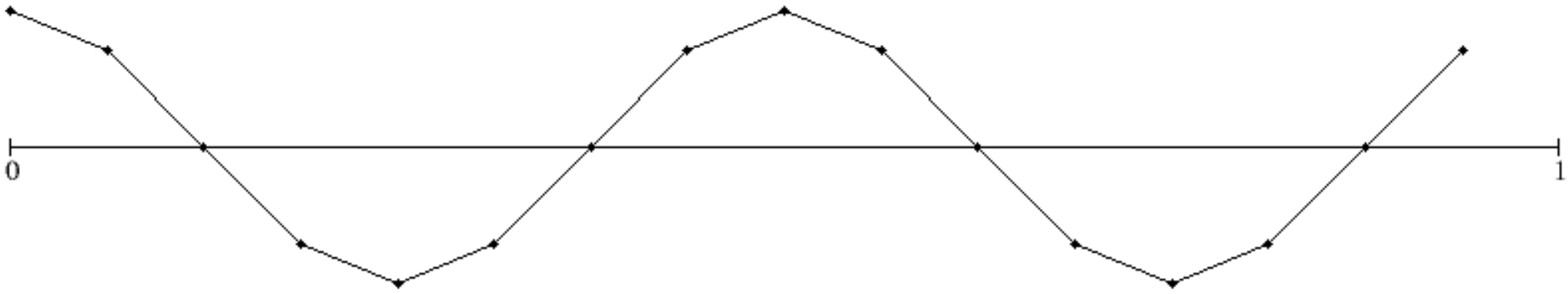


F(k) -> 0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0

FFT of curve above

	R	I	R	I	R	I	R	I	R	I	R	I	R	I	R	I	R	I
k ->	0	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8

FFT Basis

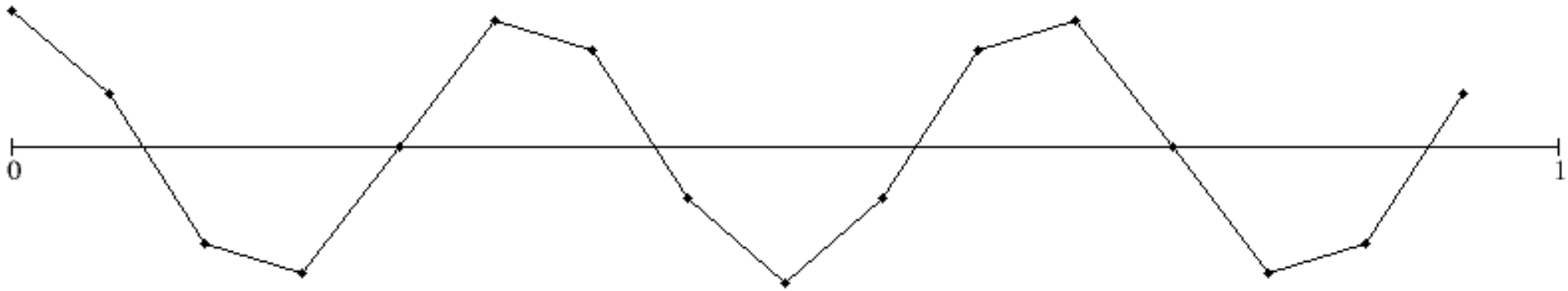



```
0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0

R I R I R I R I R I R I R I R I R I
0 0 1 1 2 2 3 3 4 4 5 5 6 6 7 7 8 8
```

FFT of curve above

FFT Basis

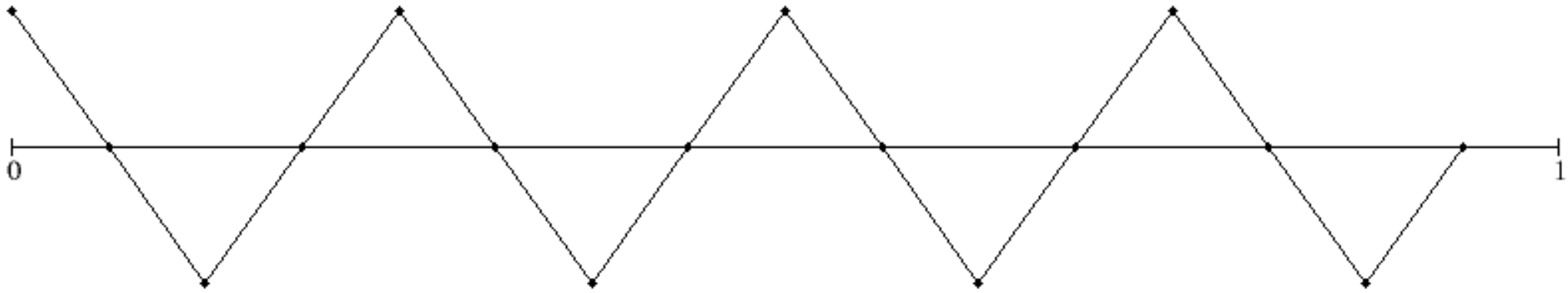



```
0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0

R I R I R I R I R I R I R I R I R I
0 0 1 1 2 2 3 3 4 4 5 5 6 6 7 7 8 8
```

FFT of curve above

FFT Basis

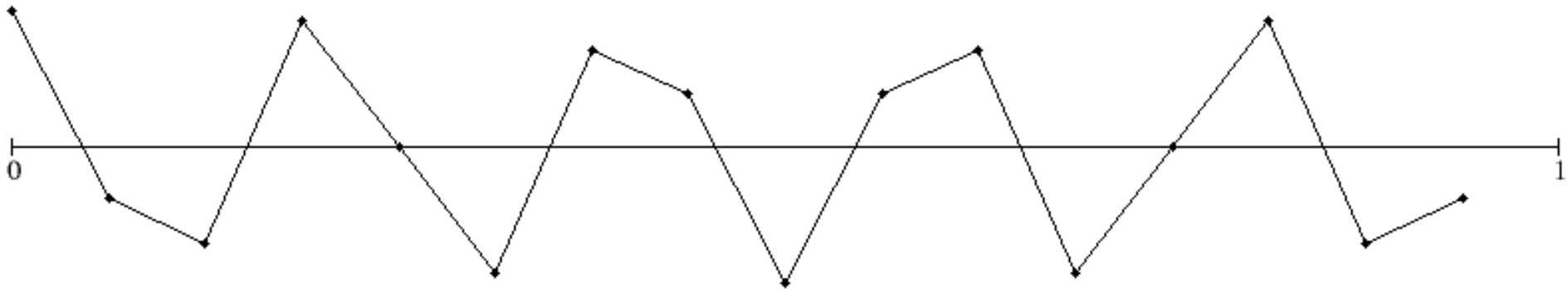



```
0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0
```

FFT of curve above

```
R I R I R I R I R I R I R I R I R I
0 0 1 1 2 2 3 3 4 4 5 5 6 6 7 7 8 8
```

FFT Basis

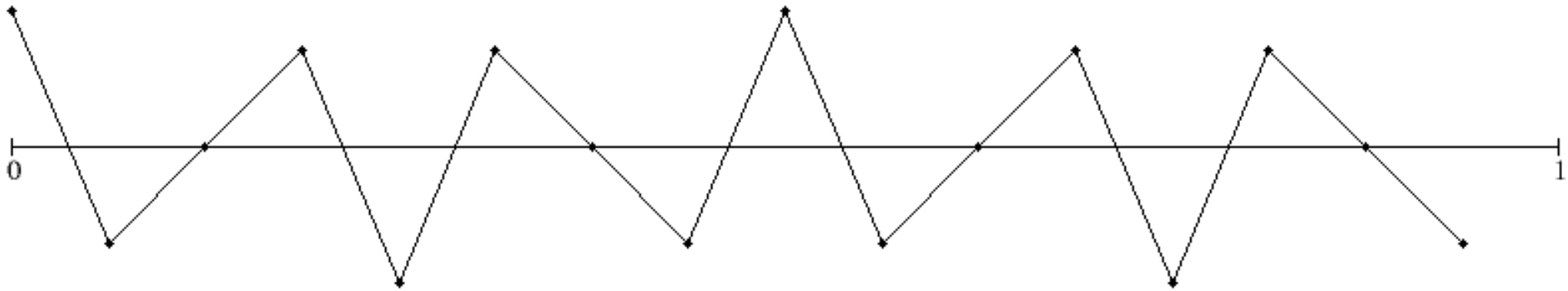



```
0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0

R I R I R I R I R I R I R I R I R I
0 0 1 1 2 2 3 3 4 4 5 5 6 6 7 7 8 8
```

FFT of curve above

FFT Basis

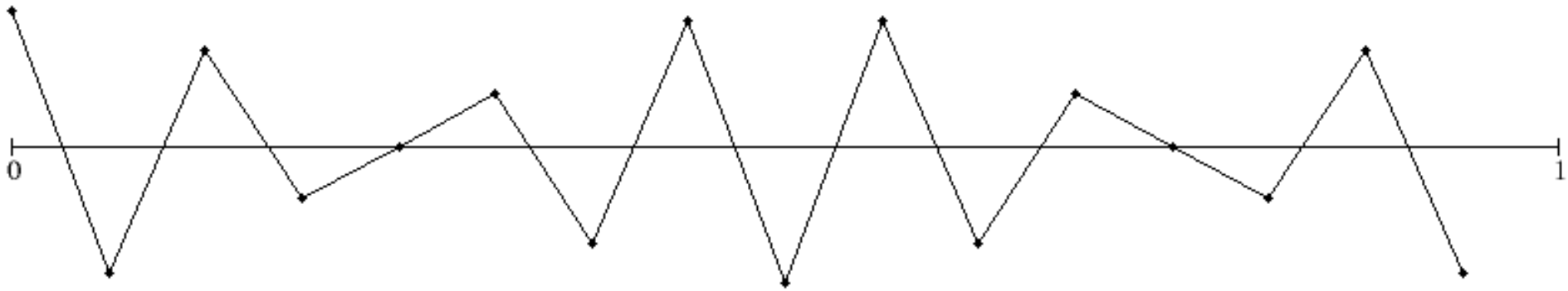


0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0

FFT of curve above

R I R I R I R I R I R I R I R I R I
0 0 1 1 2 2 3 3 4 4 5 5 6 6 7 7 8 8

FFT Basis

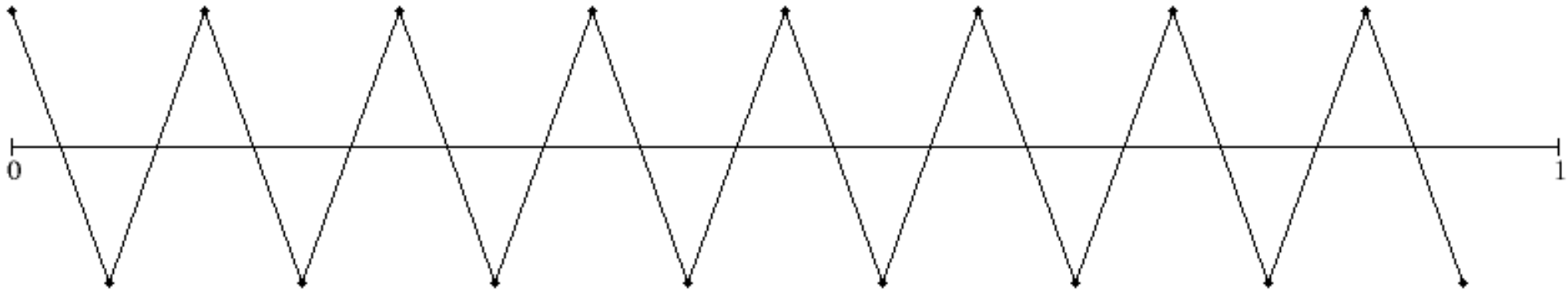



```
0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0

R I R I R I R I R I R I R I R I R I
0 0 1 1 2 2 3 3 4 4 5 5 6 6 7 7 8 8
```

FFT of curve above

FFT Basis




```
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0

R I R I R I R I R I R I R I R I R I
0 0 1 1 2 2 3 3 4 4 5 5 6 6 7 7 8 8
```

FFT of curve above

Orthogonality

$$\sum_{x=0}^{w-1} \sin\left(\frac{2\pi k_1 x}{w}\right) \sin\left(\frac{2\pi k_2 x}{w}\right) = 0 \qquad k_1 \neq k_2$$

$$\sum_{x=0}^{w-1} \cos\left(\frac{2\pi k_1 x}{w}\right) \cos\left(\frac{2\pi k_2 x}{w}\right) = 0 \qquad k_1 \neq k_2$$

$$\sum_{x=0}^{w-1} \sin\left(\frac{2\pi k_1 x}{w}\right) \cos\left(\frac{2\pi k_2 x}{w}\right) = 0$$

FFT

(finite discrete)

$$\hat{F}_k = \frac{1}{\sqrt{n}} \sum_{x=0}^{n-1} f_x \, e^{-i \, 2 \, \pi \, k \, x/n}$$

k, x and n are integers!

16 Element FFT Matrix

$$\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & \cos\left[\frac{\pi}{8}\right] & \frac{1}{\sqrt{2}} & \sin\left[\frac{\pi}{8}\right] & 0 & -\sin\left[\frac{\pi}{8}\right] & -\frac{1}{\sqrt{2}} & -\cos\left[\frac{\pi}{8}\right] & -1 & -\cos\left[\frac{\pi}{8}\right] & -\frac{1}{\sqrt{2}} & -\sin\left[\frac{\pi}{8}\right] & 0 & \sin\left[\frac{\pi}{8}\right] & \frac{1}{\sqrt{2}} & \cos\left[\frac{\pi}{8}\right] \\
0 & \sin\left[\frac{\pi}{8}\right] & \frac{1}{\sqrt{2}} & \cos\left[\frac{\pi}{8}\right] & 1 & \cos\left[\frac{\pi}{8}\right] & \frac{1}{\sqrt{2}} & \sin\left[\frac{\pi}{8}\right] & 0 & -\sin\left[\frac{\pi}{8}\right] & -\frac{1}{\sqrt{2}} & -\cos\left[\frac{\pi}{8}\right] & -1 & -\cos\left[\frac{\pi}{8}\right] & -\frac{1}{\sqrt{2}} & -\sin\left[\frac{\pi}{8}\right] \\
1 & \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} & -1 & -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 1 & \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} & -1 & -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\
0 & \frac{1}{\sqrt{2}} & 1 & \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} & -1 & -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 1 & \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} & -1 & -\frac{1}{\sqrt{2}} \\
1 & \sin\left[\frac{\pi}{8}\right] & -\frac{1}{\sqrt{2}} & -\cos\left[\frac{\pi}{8}\right] & 0 & \cos\left[\frac{\pi}{8}\right] & \frac{1}{\sqrt{2}} & -\sin\left[\frac{\pi}{8}\right] & -1 & -\sin\left[\frac{\pi}{8}\right] & \frac{1}{\sqrt{2}} & \cos\left[\frac{\pi}{8}\right] & 0 & -\cos\left[\frac{\pi}{8}\right] & -\frac{1}{\sqrt{2}} & \sin\left[\frac{\pi}{8}\right] \\
0 & \cos\left[\frac{\pi}{8}\right] & \frac{1}{\sqrt{2}} & -\sin\left[\frac{\pi}{8}\right] & -1 & -\sin\left[\frac{\pi}{8}\right] & \frac{1}{\sqrt{2}} & \cos\left[\frac{\pi}{8}\right] & 0 & -\cos\left[\frac{\pi}{8}\right] & -\frac{1}{\sqrt{2}} & \sin\left[\frac{\pi}{8}\right] & 1 & \sin\left[\frac{\pi}{8}\right] & -\frac{1}{\sqrt{2}} & -\cos\left[\frac{\pi}{8}\right] \\
1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 \\
0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 \\
1 & -\sin\left[\frac{\pi}{8}\right] & -\frac{1}{\sqrt{2}} & \cos\left[\frac{\pi}{8}\right] & 0 & -\cos\left[\frac{\pi}{8}\right] & \frac{1}{\sqrt{2}} & \sin\left[\frac{\pi}{8}\right] & -1 & \sin\left[\frac{\pi}{8}\right] & \frac{1}{\sqrt{2}} & -\cos\left[\frac{\pi}{8}\right] & 0 & \cos\left[\frac{\pi}{8}\right] & -\frac{1}{\sqrt{2}} & -\sin\left[\frac{\pi}{8}\right] \\
0 & \cos\left[\frac{\pi}{8}\right] & -\frac{1}{\sqrt{2}} & -\sin\left[\frac{\pi}{8}\right] & 1 & -\sin\left[\frac{\pi}{8}\right] & -\frac{1}{\sqrt{2}} & \cos\left[\frac{\pi}{8}\right] & 0 & -\cos\left[\frac{\pi}{8}\right] & \frac{1}{\sqrt{2}} & \sin\left[\frac{\pi}{8}\right] & -1 & \sin\left[\frac{\pi}{8}\right] & \frac{1}{\sqrt{2}} & -\cos\left[\frac{\pi}{8}\right] \\
1 & -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & -1 & \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} & 1 & -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & -1 & \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\
0 & \frac{1}{\sqrt{2}} & -1 & \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} & 1 & -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & -1 & \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} & 1 & -\frac{1}{\sqrt{2}} \\
1 & -\cos\left[\frac{\pi}{8}\right] & \frac{1}{\sqrt{2}} & -\sin\left[\frac{\pi}{8}\right] & 0 & \sin\left[\frac{\pi}{8}\right] & -\frac{1}{\sqrt{2}} & \cos\left[\frac{\pi}{8}\right] & -1 & \cos\left[\frac{\pi}{8}\right] & -\frac{1}{\sqrt{2}} & \sin\left[\frac{\pi}{8}\right] & 0 & -\sin\left[\frac{\pi}{8}\right] & \frac{1}{\sqrt{2}} & -\cos\left[\frac{\pi}{8}\right] \\
0 & \sin\left[\frac{\pi}{8}\right] & -\frac{1}{\sqrt{2}} & \cos\left[\frac{\pi}{8}\right] & -1 & \cos\left[\frac{\pi}{8}\right] & -\frac{1}{\sqrt{2}} & \sin\left[\frac{\pi}{8}\right] & 0 & -\sin\left[\frac{\pi}{8}\right] & \frac{1}{\sqrt{2}} & -\cos\left[\frac{\pi}{8}\right] & 1 & -\cos\left[\frac{\pi}{8}\right] & \frac{1}{\sqrt{2}} & -\sin\left[\frac{\pi}{8}\right] \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1
\end{pmatrix}$$

120 different angles in 16 dim space rotation matrix

Note: this matrix ostensibly excludes 2 rows of 0's. After the first row would be a sin(0)=0 row. Similarly, before the last row, another row of 0's should exist.

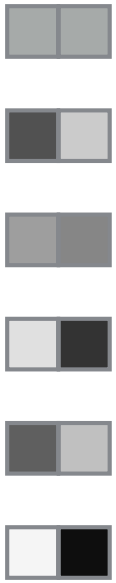
Other Basis Functions ?

- Real Space –
  - accurate position
  - no spectral information
- FFT
  - no positional information
  - accurate spectral information
- Wavelets
  - mix of spectra and position
- PCA/SVD
  - data-based basis (usually intentionally incomplete)
- ...

MSA

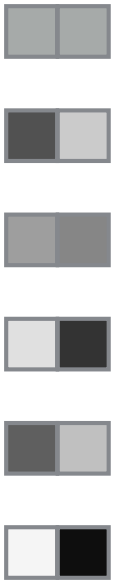
- MSA : Multivariate Statistical Analysis
- PCA : Principal Component Analysis
    - a type of MSA
- SVD : Singular Value Decomposition
    - One way to compute PCA

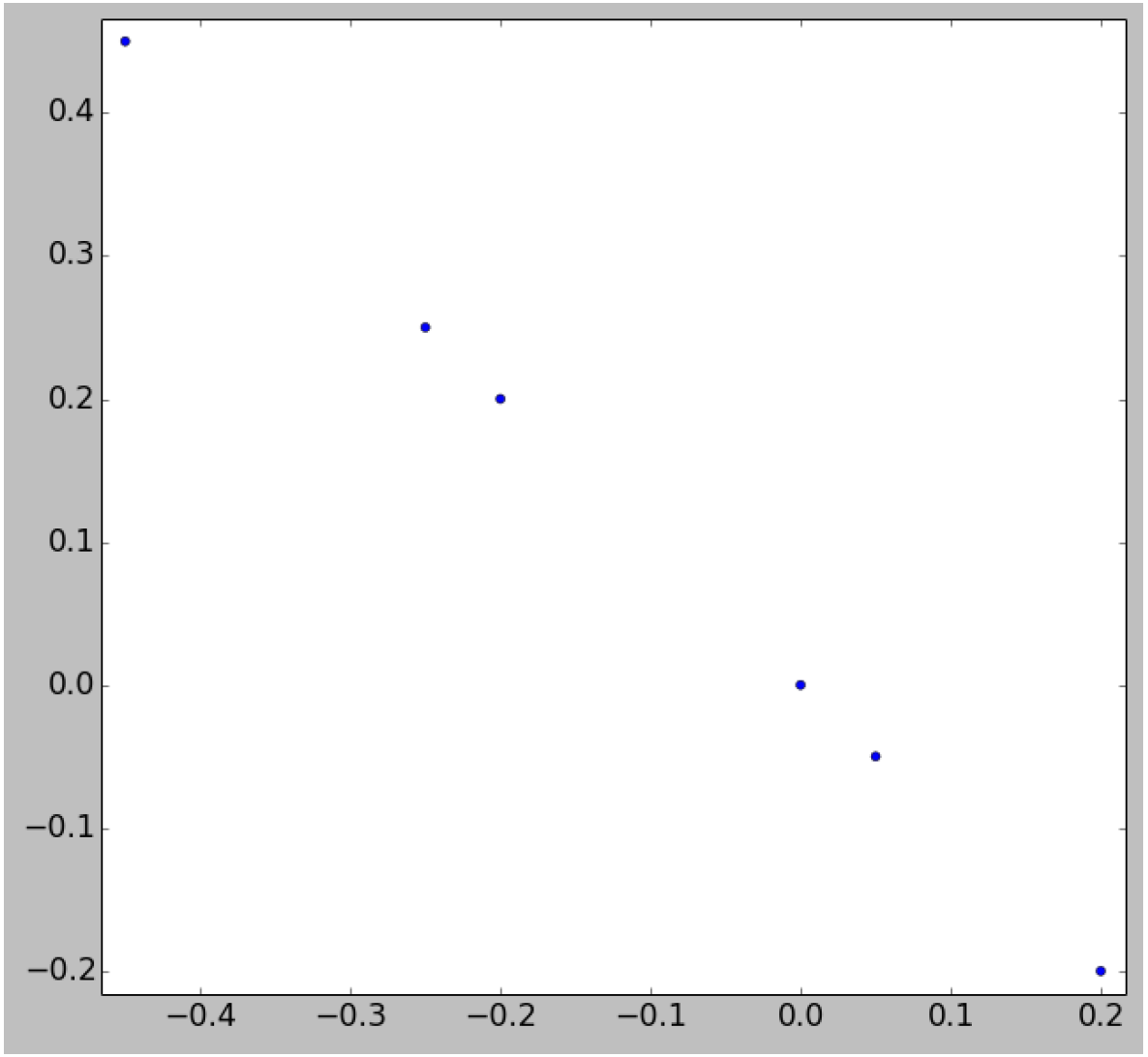
2-pixel Images



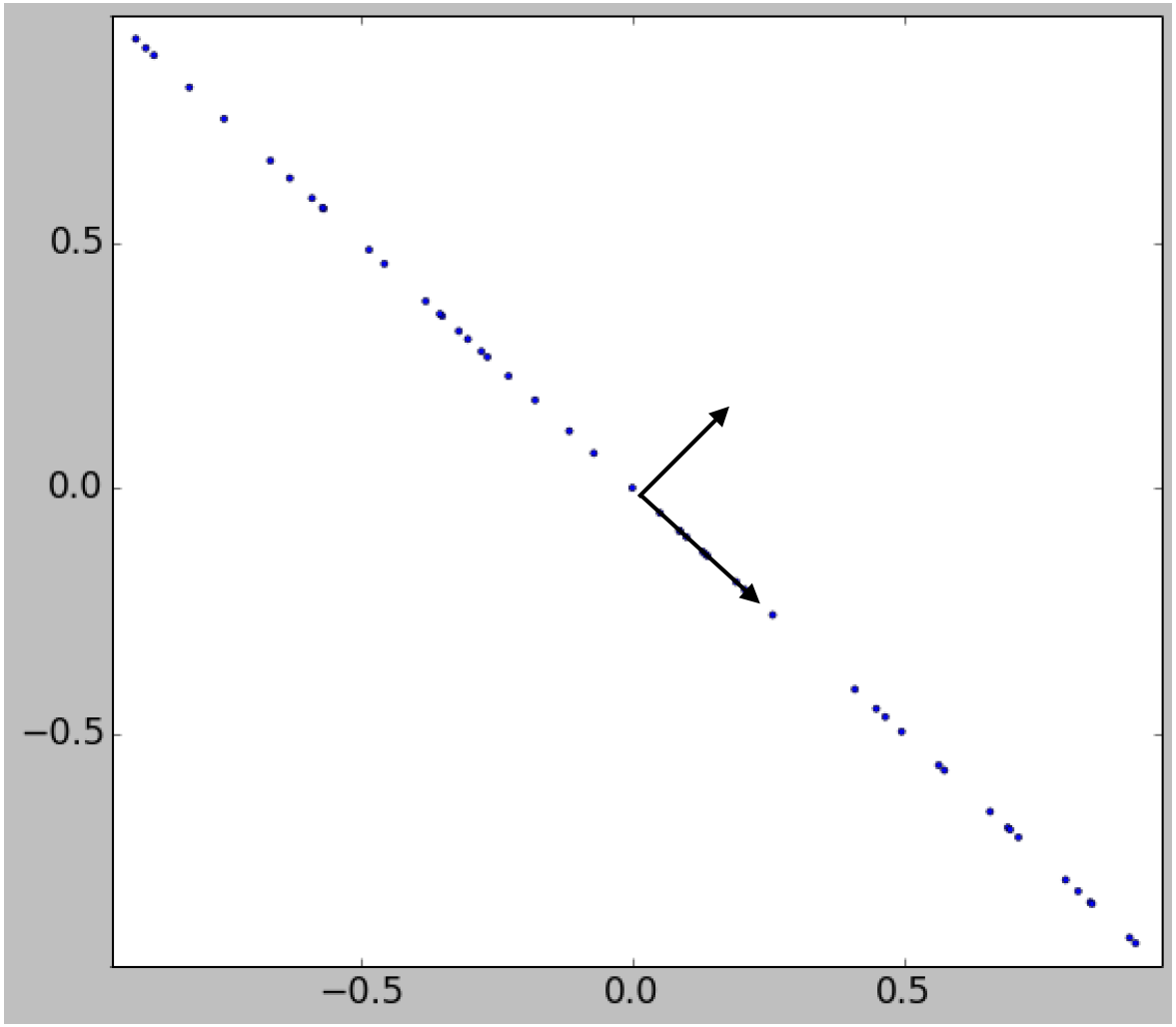
2-pixel Images

x y



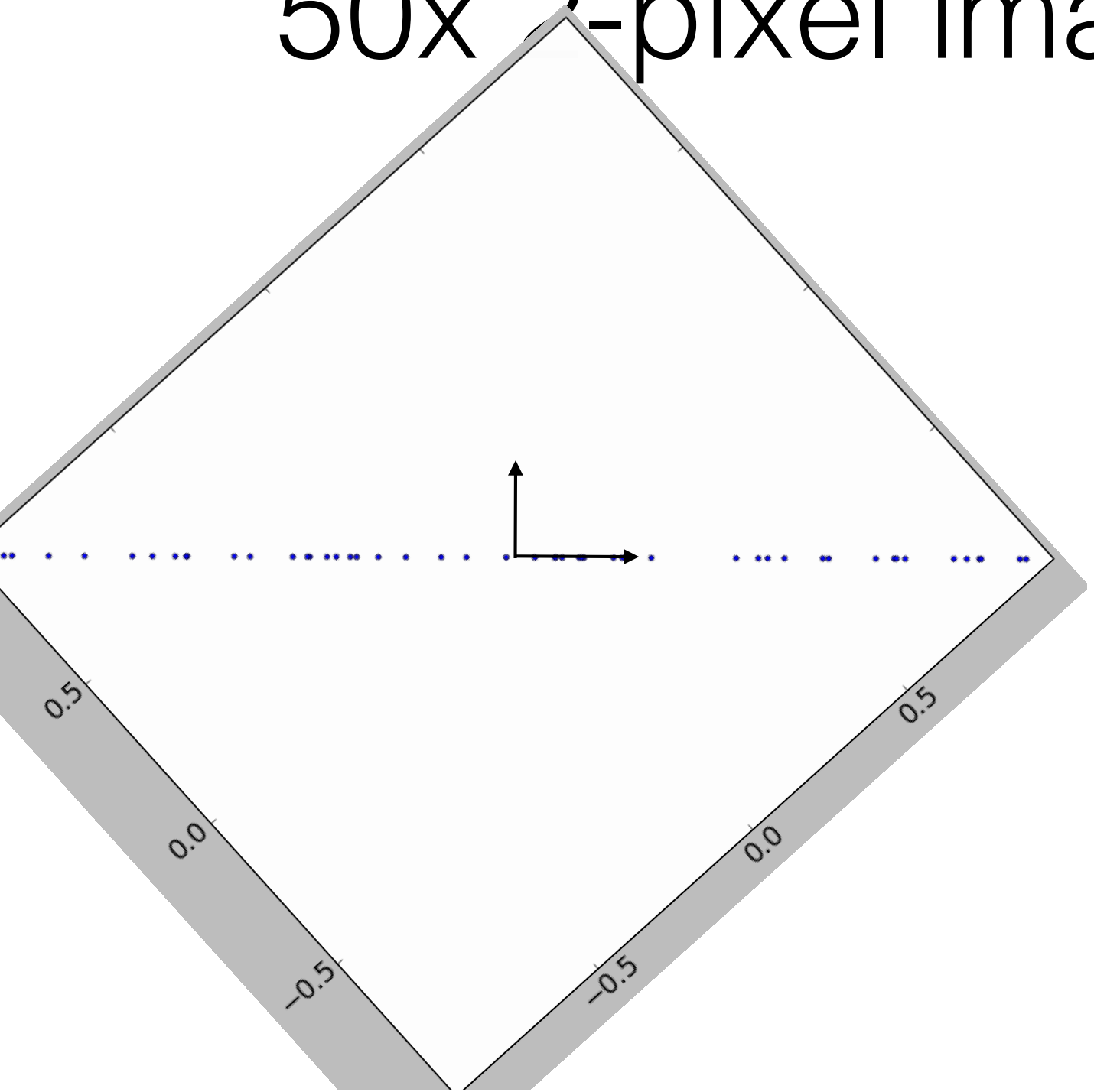


50x 2-pixel images



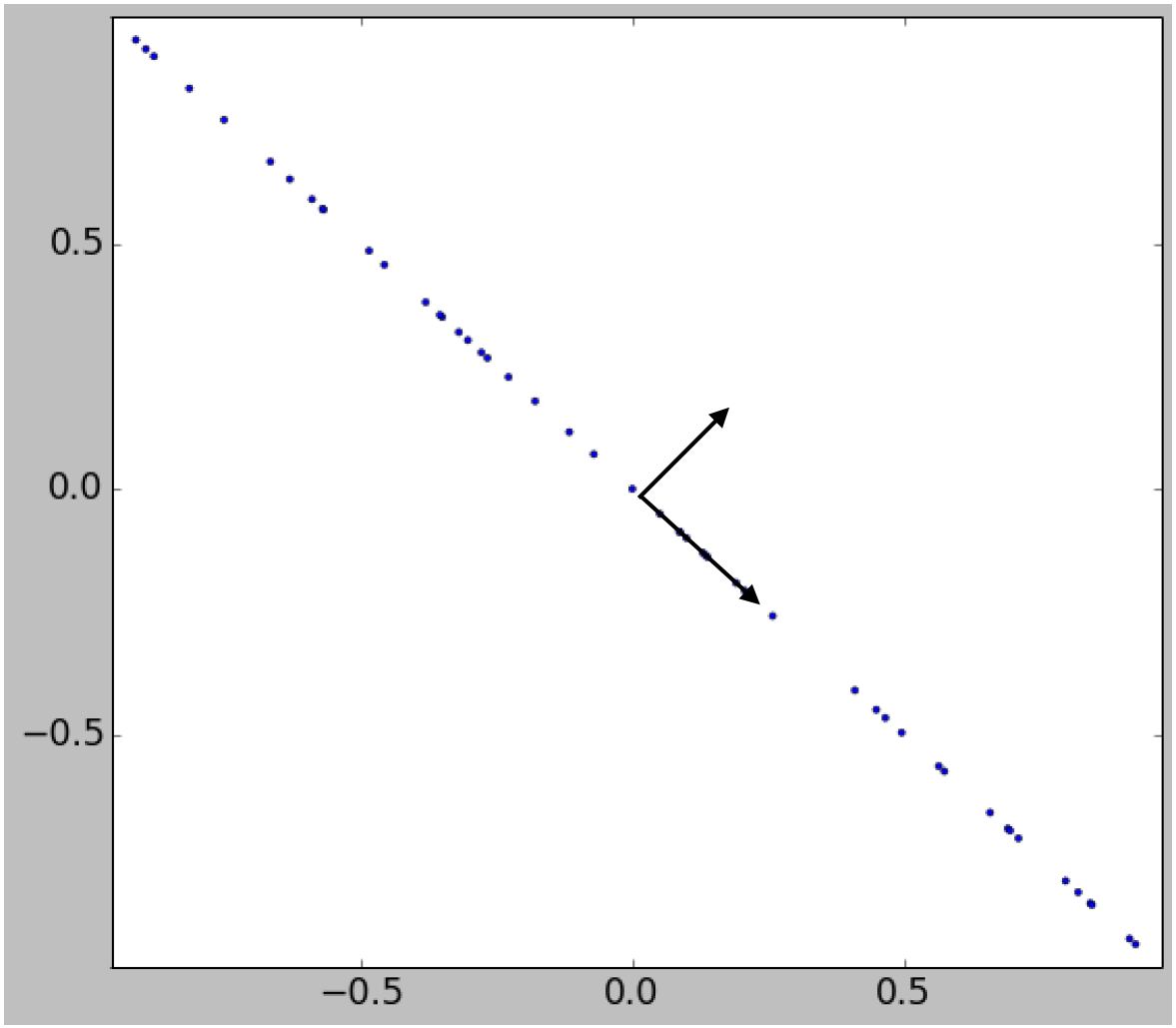
$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

50x 2-pixel images



$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

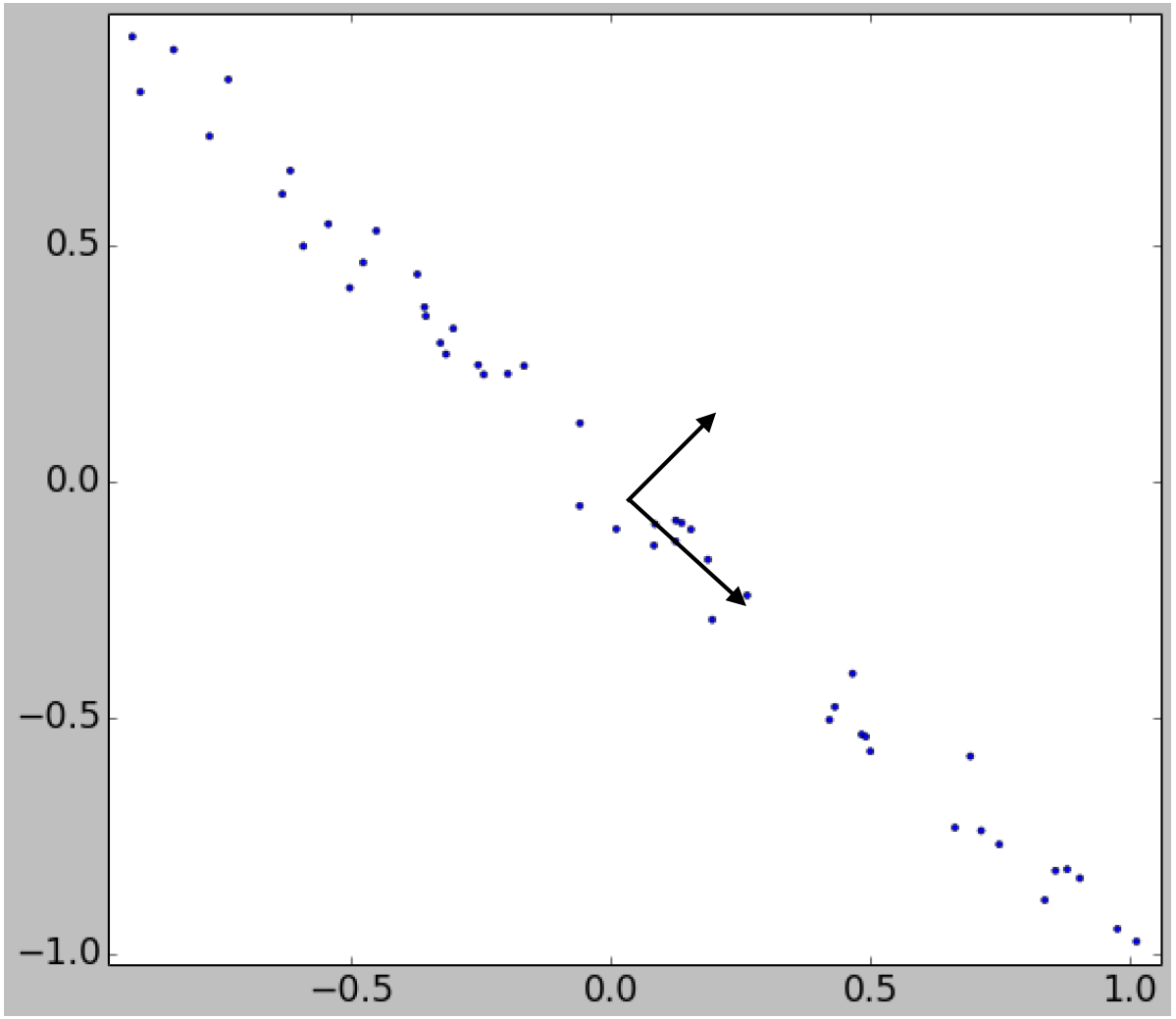
50x 2-pixel images



$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

Vectors:
(1,-1)
and
(1,1)
form an orthonormal basis
in 2 dimensions

Add Noise



What if we had 3-pixel images?

PCA

- 100x100 image →
  - 10,000 pixels →
    - 10,000 dimensional vector →
      - PCA is 10,000 vectors with 10,000 components each

4D Rotation Matrices

$$rotXY = \begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 & 0 \\ -\sin(\theta) & \cos(\theta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$rotYZ = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & \sin(\theta) & 0 \\ 0 & -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$rotXZ = \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) & 0 \\ 0 & 1 & 0 & 0 \\ \sin(\theta) & 0 & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$rotXU = \begin{bmatrix} \cos(\theta) & 0 & 0 & \sin(\theta) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\sin(\theta) & 0 & 0 & \cos(\theta) \end{bmatrix}$$

$$rotYU = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 0 & 1 & 0 \\ 0 & \sin(\theta) & 0 & \cos(\theta) \end{bmatrix}$$

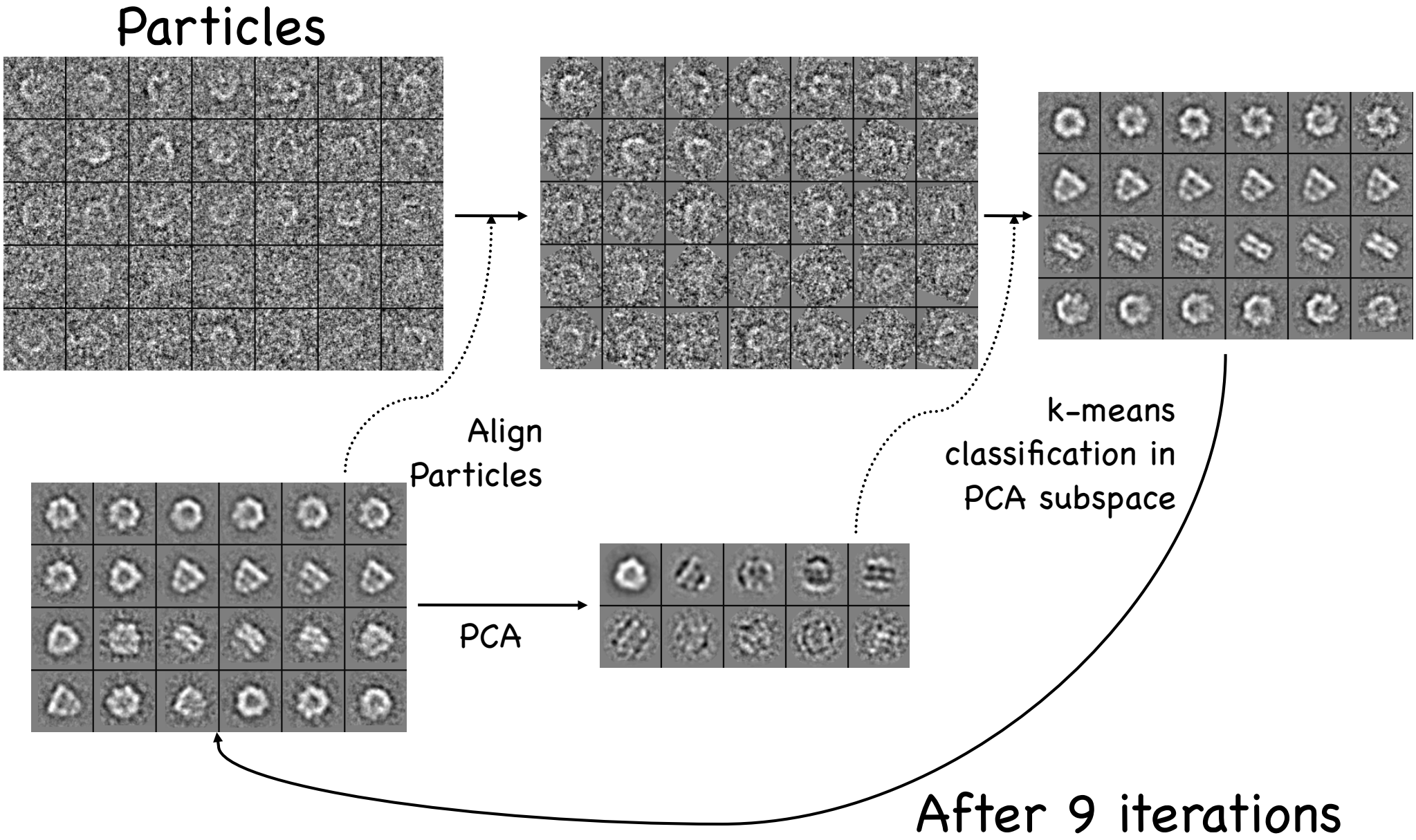
$$rotZU = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos(\theta) & -\sin(\theta) \\ 0 & 0 & \sin(\theta) & \cos(\theta) \end{bmatrix}$$

6 angles to specify in 4D space
(in general n(n-1)/2)

However

- If we have only 50 input images, then we can fully represent them with at most 50x 10,000 pixel basis vectors.

2-D Refinement



Rotational/ Translational Invariants