Multi-CryoGAN:
Reconstruction of Continuous Conformations in Cryo-EM without pose nor conformation estimation

Dr. Harshit Gupta, Stanford LAC

Huy Thong

Dr. Jaejun Yoo

Dr. Michael T. McCann

Dr. Laurene Donati

Prof. Michael Unser, EPFL

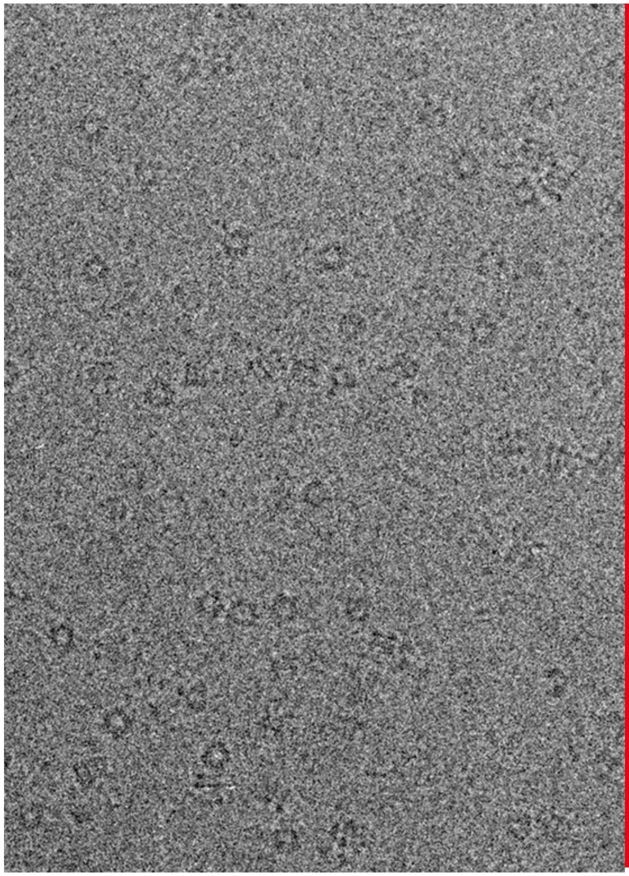
Work done at Biomedical Imaging Group

EPFL, Switzerland

April 6, 2022

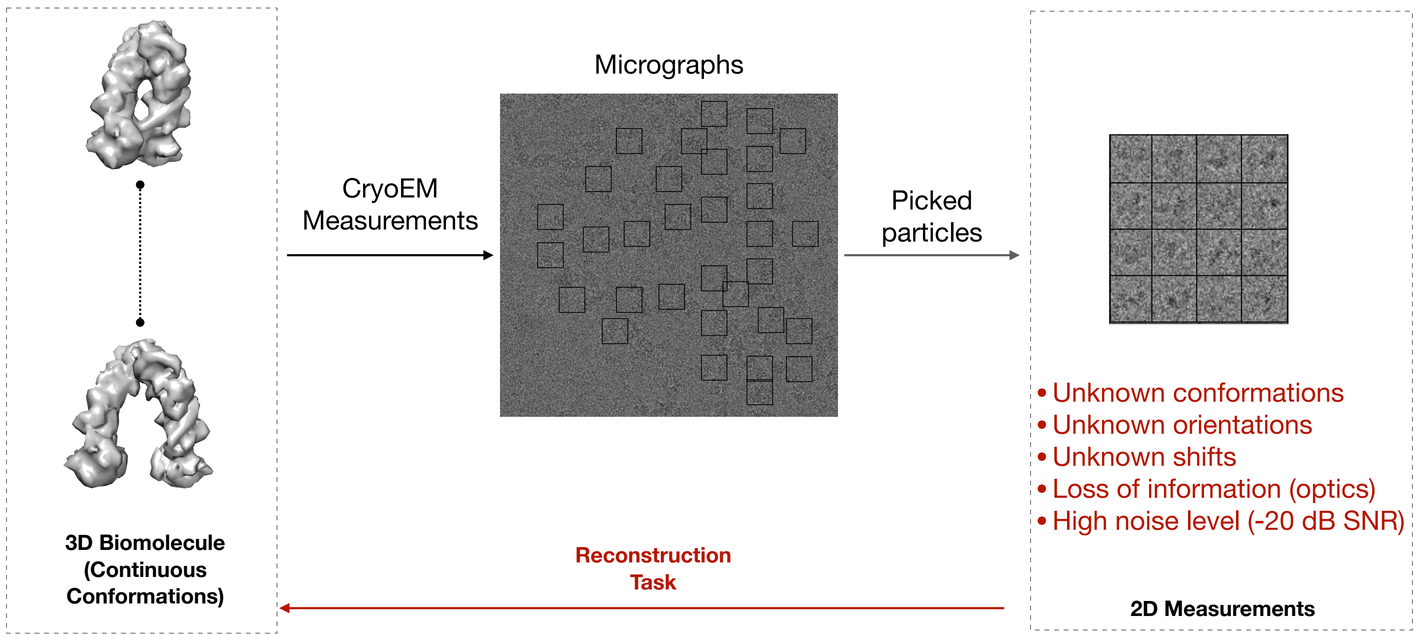
Outline

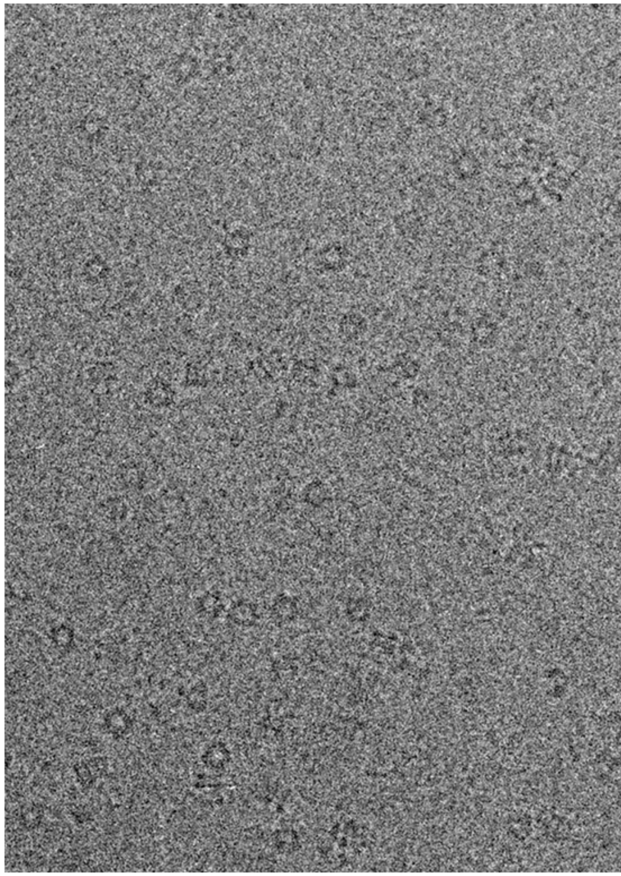
- Challenges
- Current Methods
- CryoGAN for single conformations
  - Intuition
  - Algorithm
  - Results
- Multi-CryoGAN for continuous conformations
  - Results



Challenges

Cryo-EM reconstruction problem

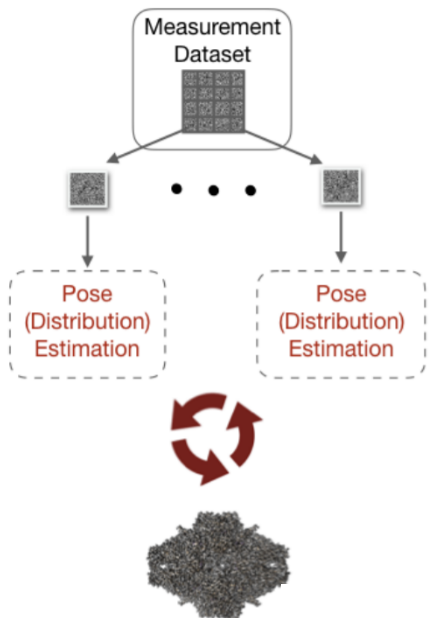




Current Methods

Current Methods

- Use complicated routines to estimate pose (or distribution) for each projection.

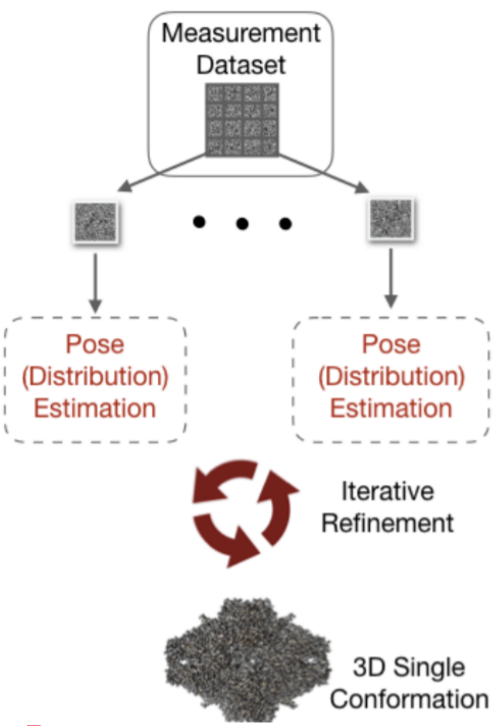


- Number of variables to estimate grow with the data size.

Standard reconstruction method in Cryo-EM

Standard techniques

(Likelihood-based)



Single Conformation

$$\mathbf{x}_{\text{rec}} = \arg\max_{\mathbf{x}} \sum_{n=1}^{N} \log p(\mathbf{y}_{\text{data}}^{n} | \mathbf{x})$$

Pose (distribution) estimation for each projection

RELION EMAN mipp cryoSPARC

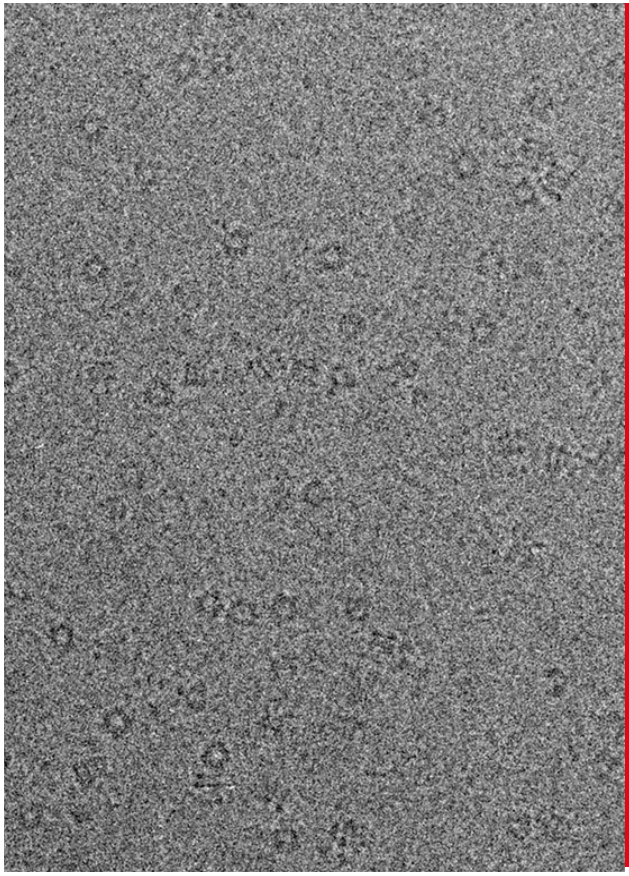
Multiple Conformation

Pose (distribution) estimation for each projection

Conformation estimation for each projection

[Dashti et al., 2014, Anden et al. 2015 Moscovich et al., 2020, Lederman et al. 2020, Seitz et al. 2019, Sorzano et al. 2020, Zhong et a. 2020]

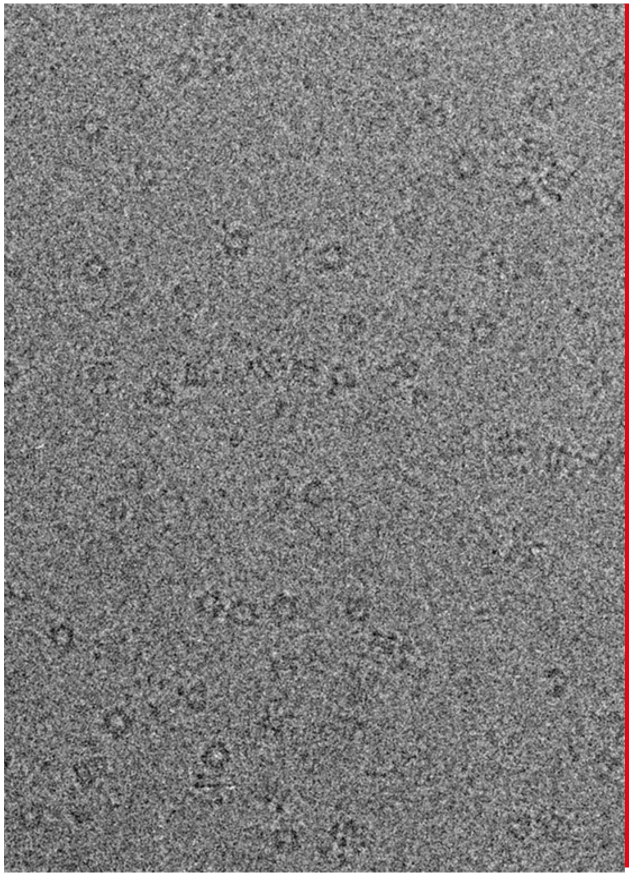
How about no pose or conformation estimation routine?



CryoGAN for single conformation

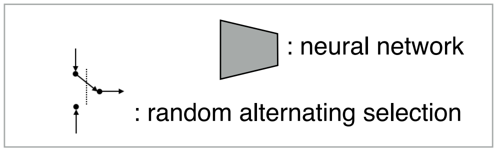
Intuition

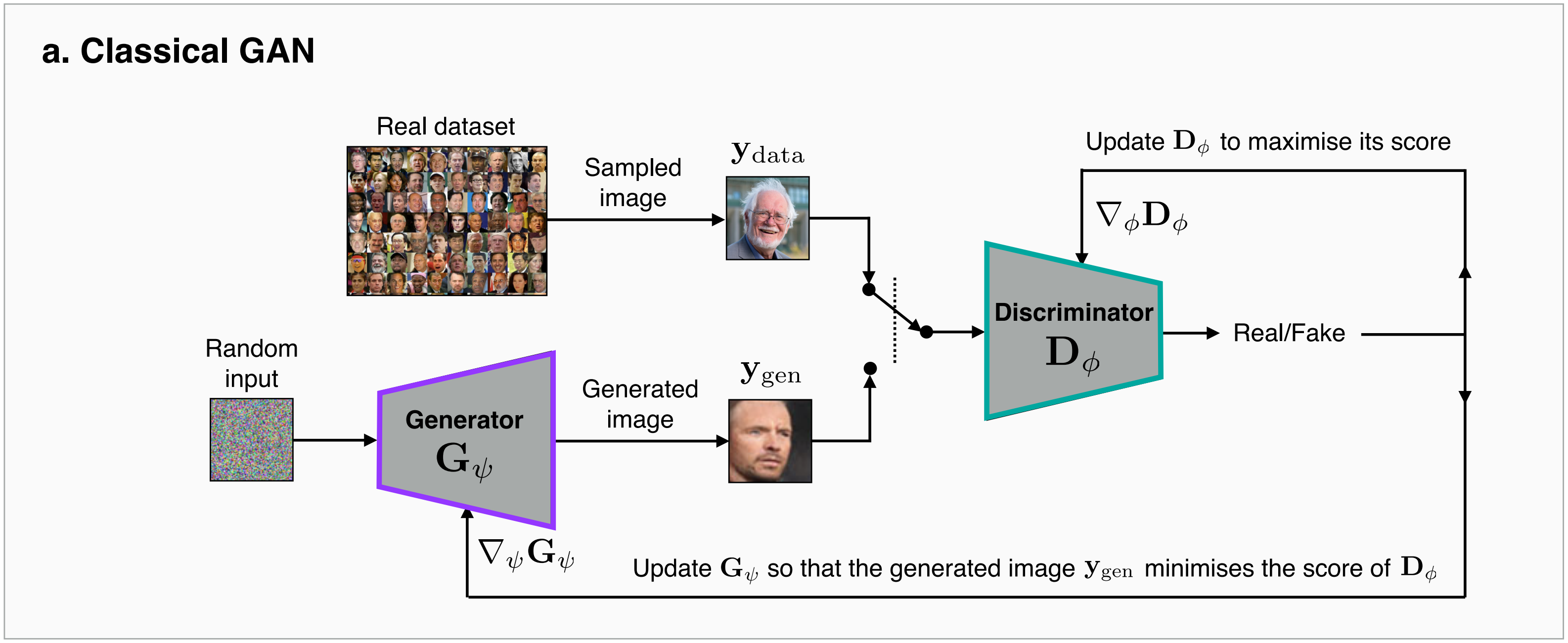
Reconstruct the structure whose
“set of projections from random poses”
looks similar to
“acquired data”



Top-Bottom View

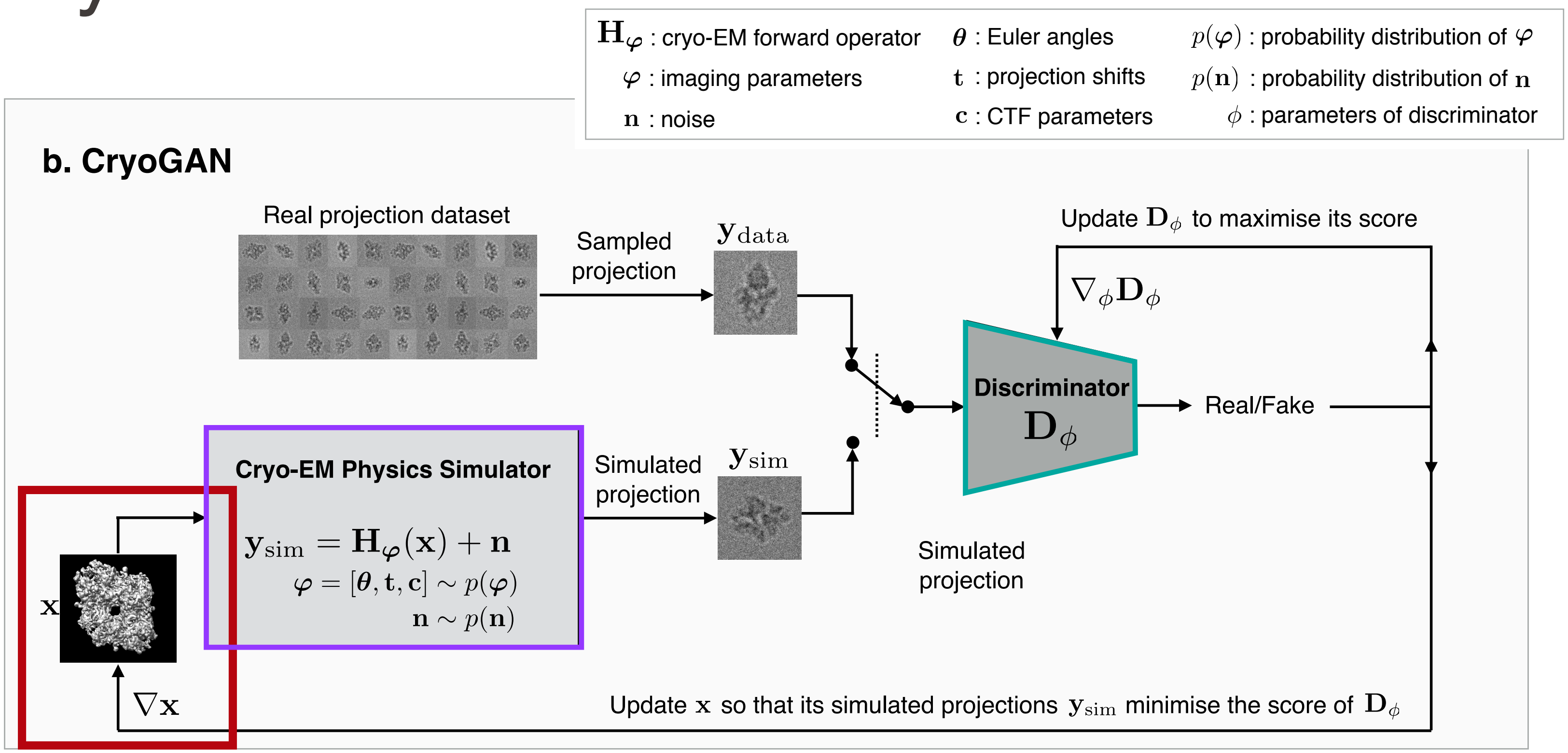
Classical GAN



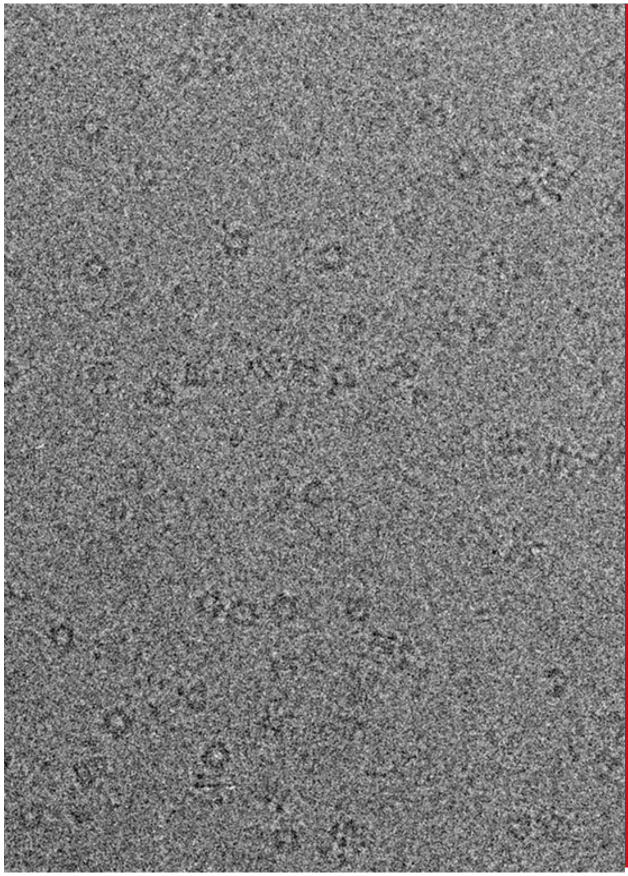


- **Discriminator** Learns to **discriminate** between real and fake (=synthetic) faces
- **Generator** Learns to **generate** fake (=synthetic) faces that fool the discriminator
- **Adversarial** **Captures the distribution of the real dataset** https://thispersondoesnotexist.com/

CryoGAN

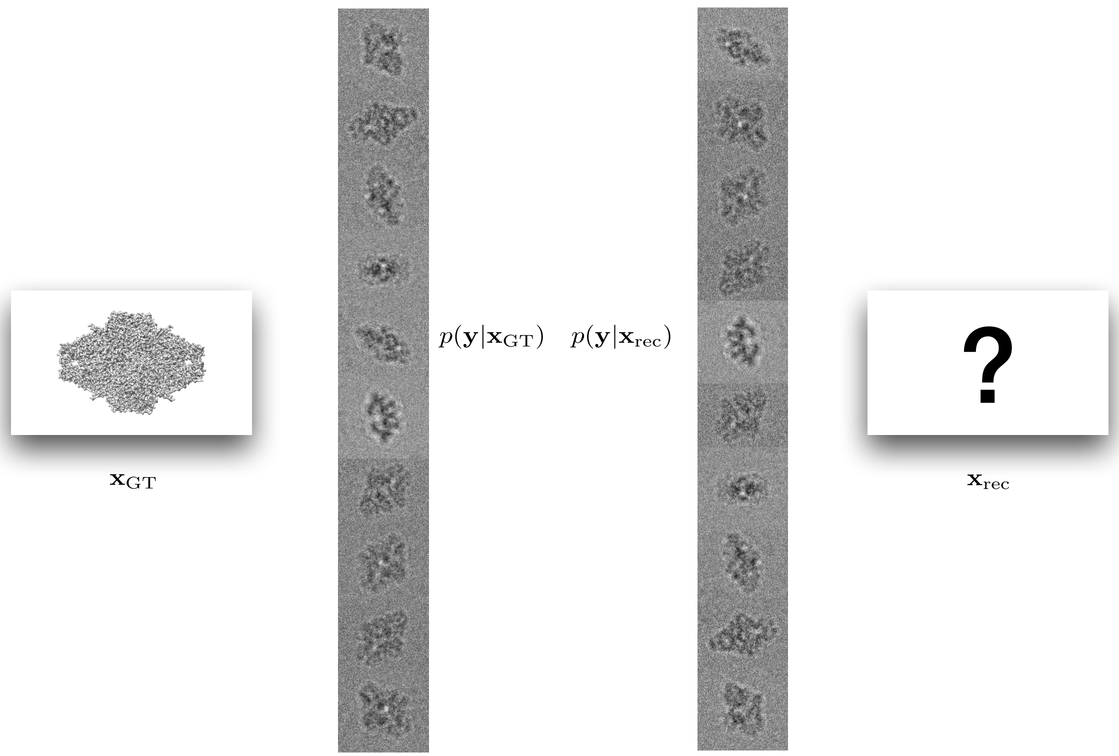


- **Discriminator** — Learns to **discriminate** between real and fake (=synthetic) projections
- **Cryo-EM Simulator** — **Generate** fake (=synthetic) projections from a given 3D volume
- **Adversarial** — **Learns a volume whose fake projections fool the Discriminator**



Bottom-Top View

Distribution Matching



Distribution Matching

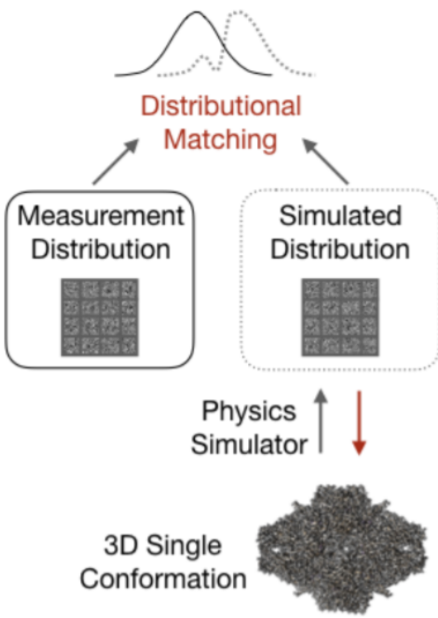
- Why?

- With some assumptions on forward model.

Theorem 1. $p(\mathbf{y}|\mathbf{x}_1) = p(\mathbf{y}|\mathbf{x}_2) \Leftrightarrow \mathbf{x}_1 = G(\mathbf{x}_2)$,
where G is some rotation-reflection operation.

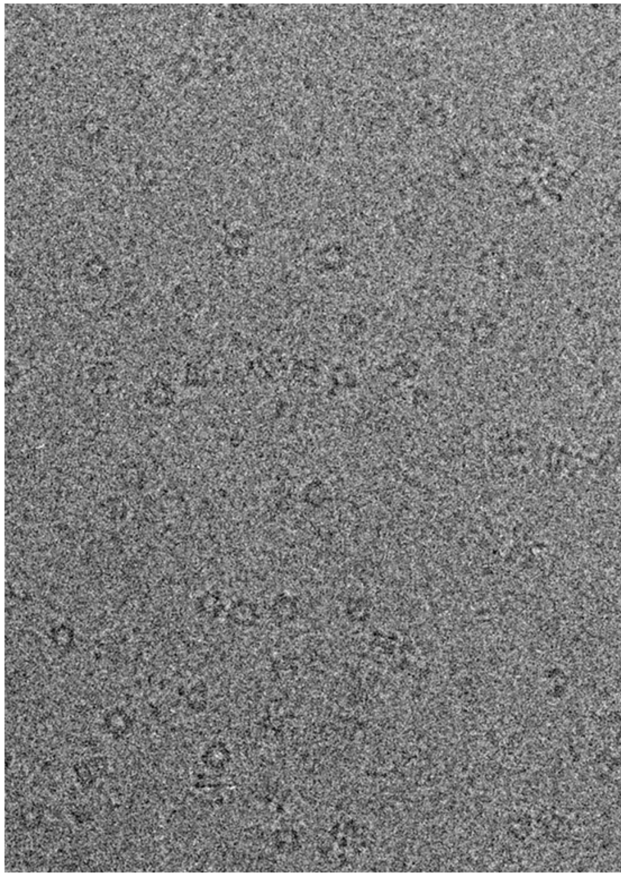
[G., 2020]

Distribution Matching



- Learn a 3D volume whose **simulated projection distribution** matches **data distribution**

- Use GAN
- No pose estimation
- No required good initial volume
- End-to-end



Algorithm

Distances

- Wasserstein Distance

$$WD(p_1, p_2) = \inf_{\gamma \in \Pi(p_1, p_2)} \mathbb{E}_{(\mathbf{y}_1, \mathbf{y}_2) \sim \gamma} [\|\mathbf{y}_1 - \mathbf{y}_2\|]$$

$$\mathbf{x}_{\text{rec}} = \underset{\mathbf{x}}{\operatorname{argmin}} \inf_{\gamma \in \Pi(p_{\mathbf{x}}, p_{\text{data}})} \mathbb{E}_{(\mathbf{y}_1, \mathbf{y}_2) \sim \gamma} [\|\mathbf{y}_1 - \mathbf{y}_2\|]$$

- Dual form

$$\mathbf{x}_{\text{rec}} = \underset{\mathbf{x}}{\operatorname{argmin}} \max_{s : \|s\|_L < 1} \left(\mathbb{E}_{\mathbf{y} \sim p_{\text{data}}} [s(\mathbf{y})] - \mathbb{E}_{\mathbf{y} \sim p_{\mathbf{x}}} [s(\mathbf{y})] \right)$$

- Using Neural Network : universal approximation property

$$\mathbf{x}_{\text{rec}} = \underset{\mathbf{x}}{\operatorname{argmin}} \max_{\mathbf{D}_\phi : \|\mathbf{D}_\phi\|_L < 1} \left(\mathbb{E}_{\boxed{\mathbf{y} \sim p_{\text{data}}}} [\mathbf{D}_\phi(\mathbf{y})] - \mathbb{E}_{\boxed{\mathbf{y} \sim p_{\mathbf{x}}}} [\mathbf{D}_\phi(\mathbf{y})] \right)$$

Acquired Measurements

Simulated measurements

Optimization

- Min max loss

$$\mathbf{x}_{\mathrm{rec}} = \operatorname*{argmin}_{\mathbf{x}} \max_{\mathbf{D}_\phi} \left(\mathbb{E}_{\mathbf{y}\sim p_{\mathrm{data}}}[\mathbf{D}_\phi(\mathbf{y})] - \mathbb{E}_{\mathbf{y}\sim p_{\mathbf{x}}}[\mathbf{D}_\phi(\mathbf{y})] + \lambda \cdot \mathbb{E}_{\mathbf{y}\sim p_{\mathrm{int}}}[(\|\nabla_{\mathbf{y}}\mathbf{D}_\phi(\mathbf{y})\| - 1)^2] \right)$$

- Sample a batch

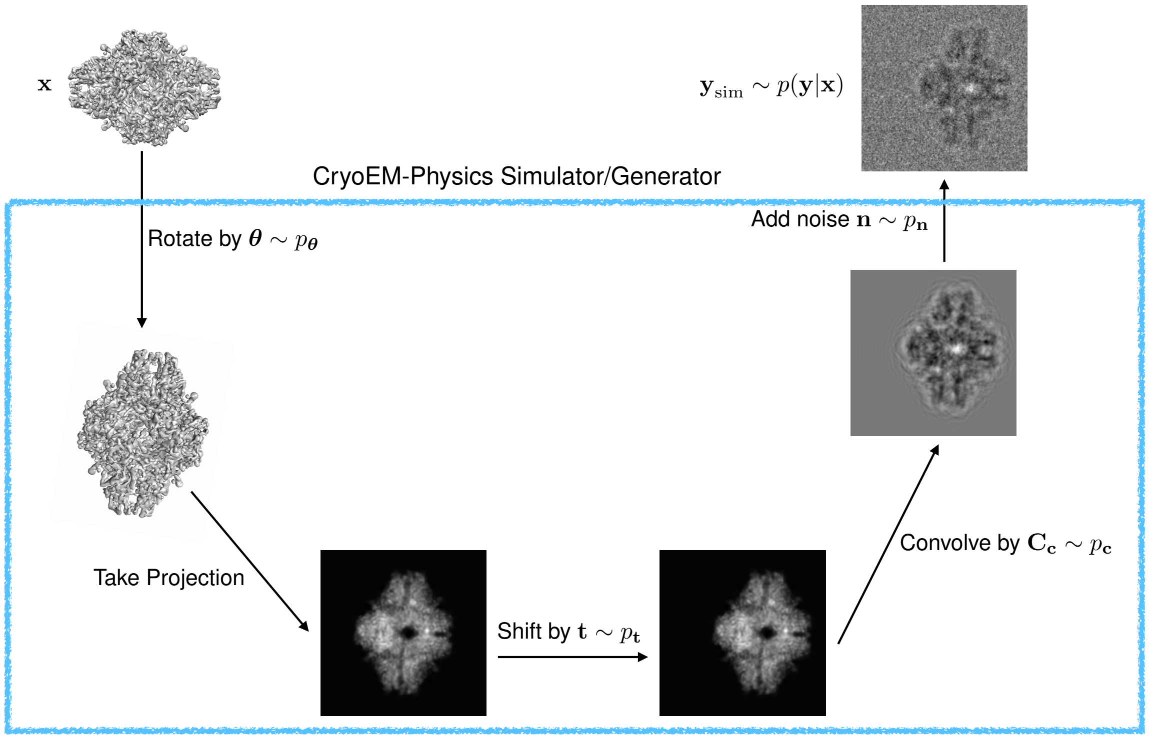
$$\mathbf{y}^b_{\mathrm{data}} \sim p_{\mathrm{data}} \qquad \mathbf{y}^b_{\mathrm{sim}} \sim p_{\mathbf{x}} \qquad \mathbf{y}^b_{\mathrm{int}} \sim p_{\mathrm{int}}$$

- Empirical

$$L_S(\mathbf{x}, \mathbf{D}_\phi) = \sum_{b=1}^{B} \mathbf{D}_\phi(\mathbf{y}^b_{\mathrm{data}}) - \sum_{b=1}^{B} \mathbf{D}_\phi(\mathbf{y}^b_{\mathrm{sim}}) + \lambda \sum_{b=1}^{B} (\|\nabla_{\mathbf{y}}\mathbf{D}_\phi(\mathbf{y}^b_{\mathrm{int}})\| - 1)^2)$$

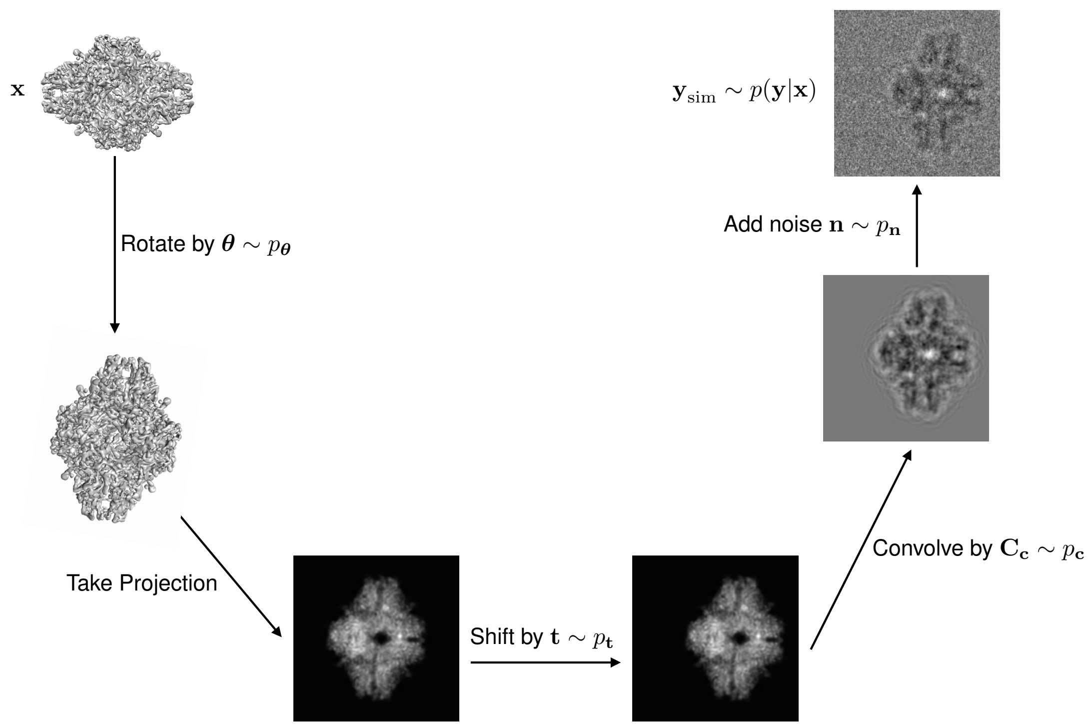
Sampler/Simulator/Generator/

- CryoEM physics simulator



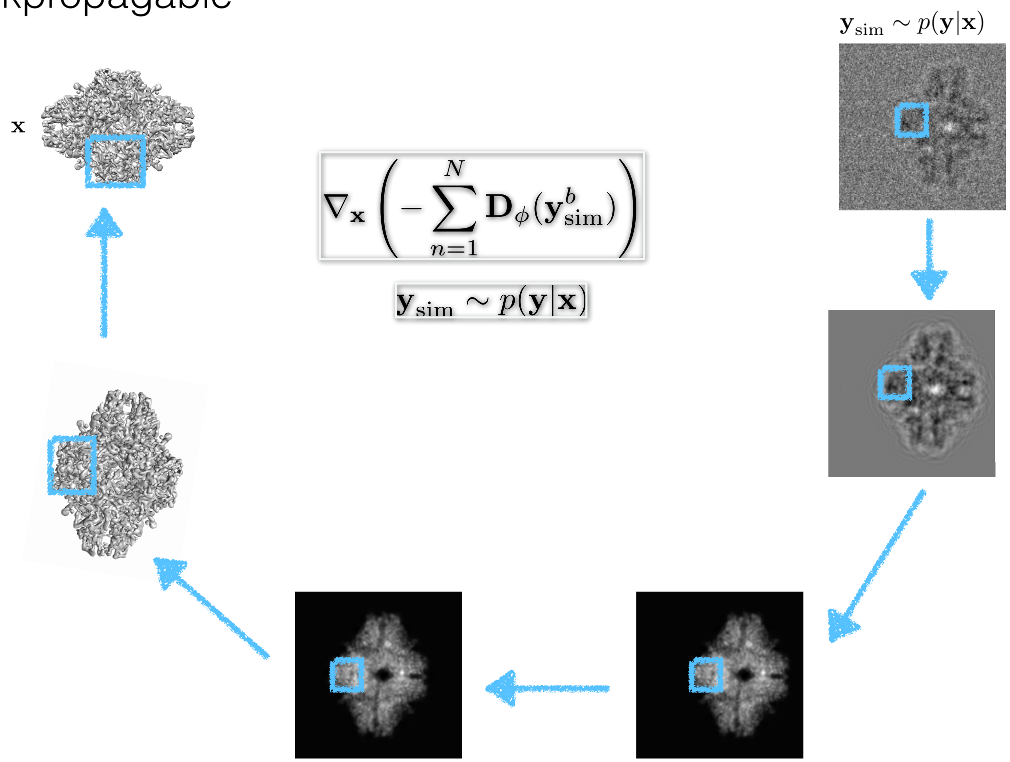
Simulator/Generator

- Sample $\mathbf{y}_{\text{sim}} \sim p(\mathbf{y}|\mathbf{x})$



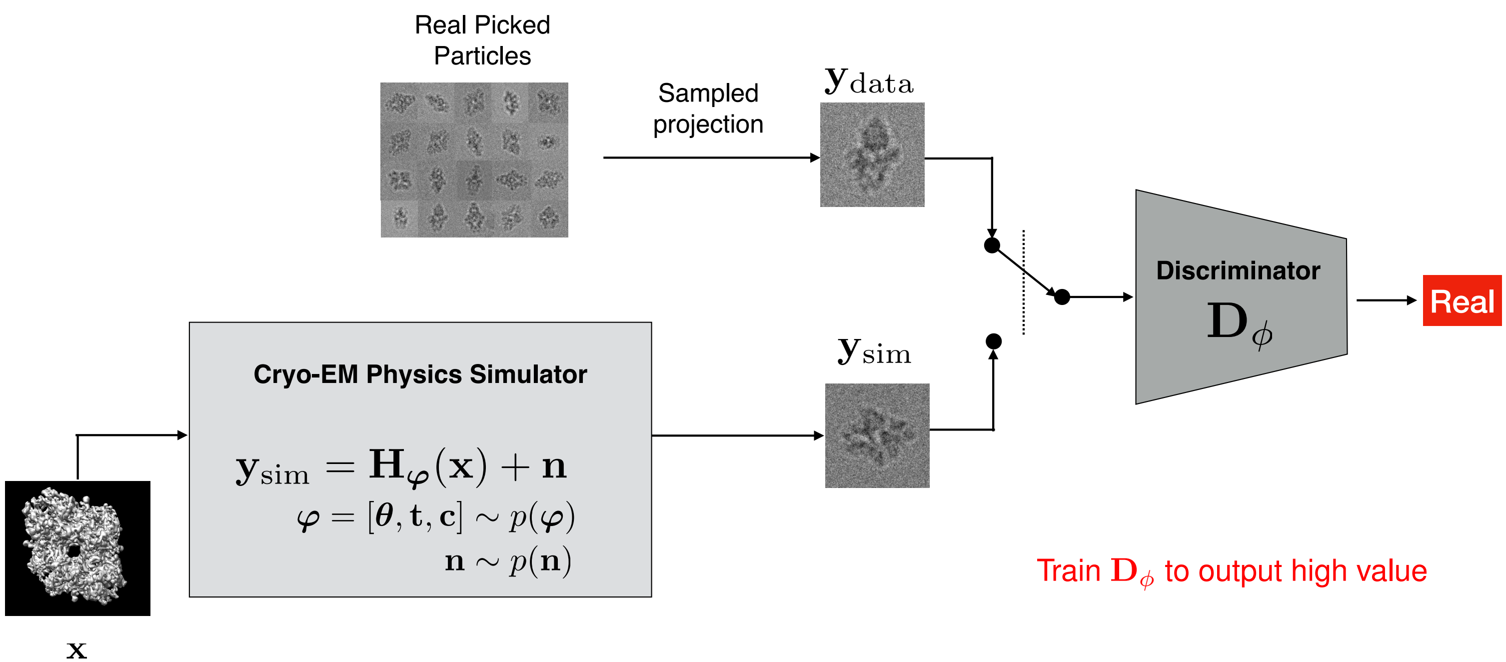
Simulator/Generator

- Backpropagable



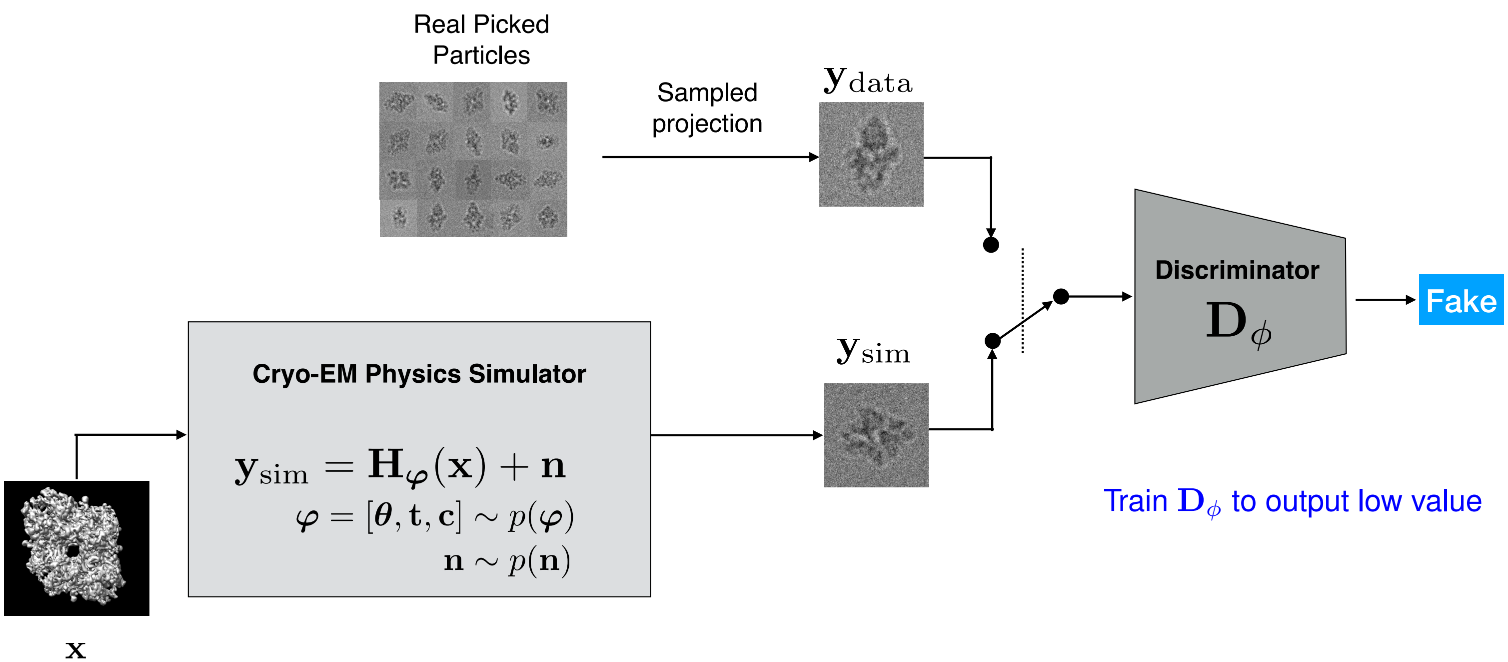
CryoGAN

- Discriminator



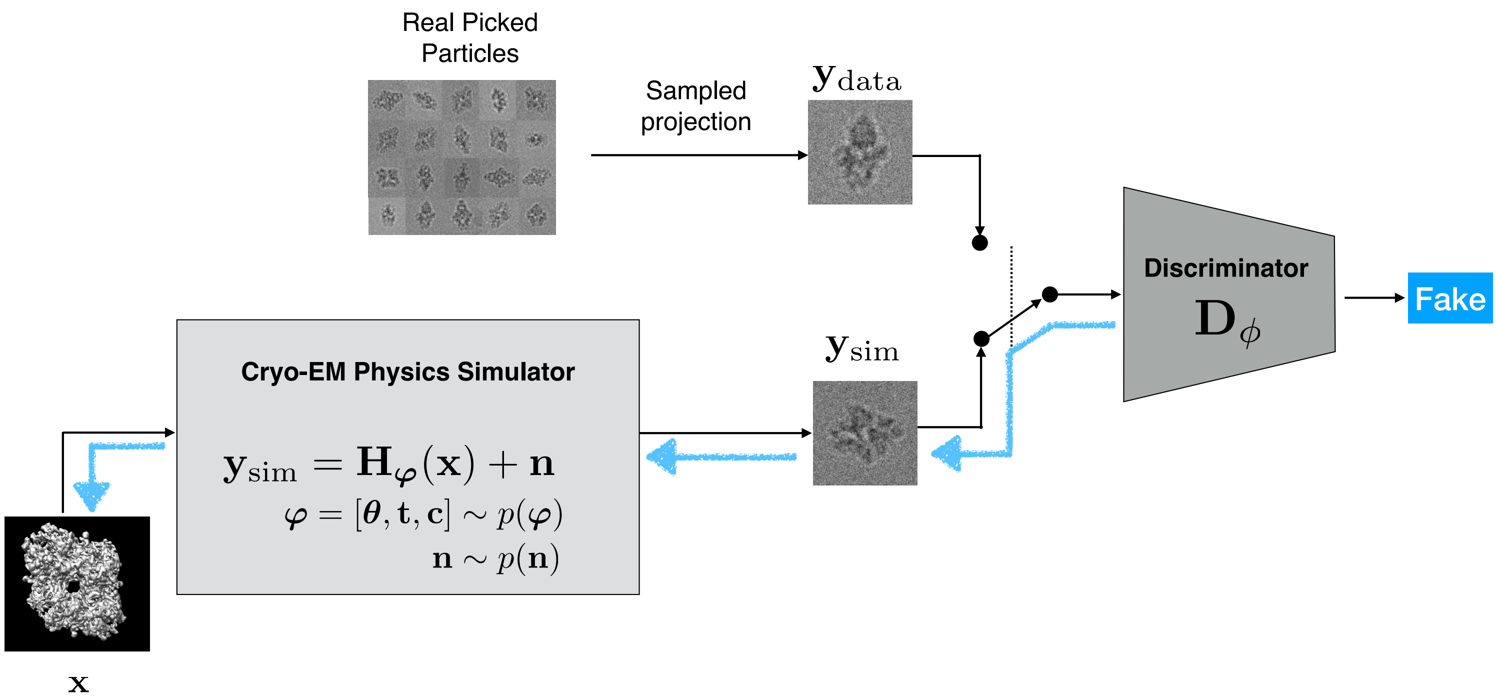
CryoGAN

- Discriminator

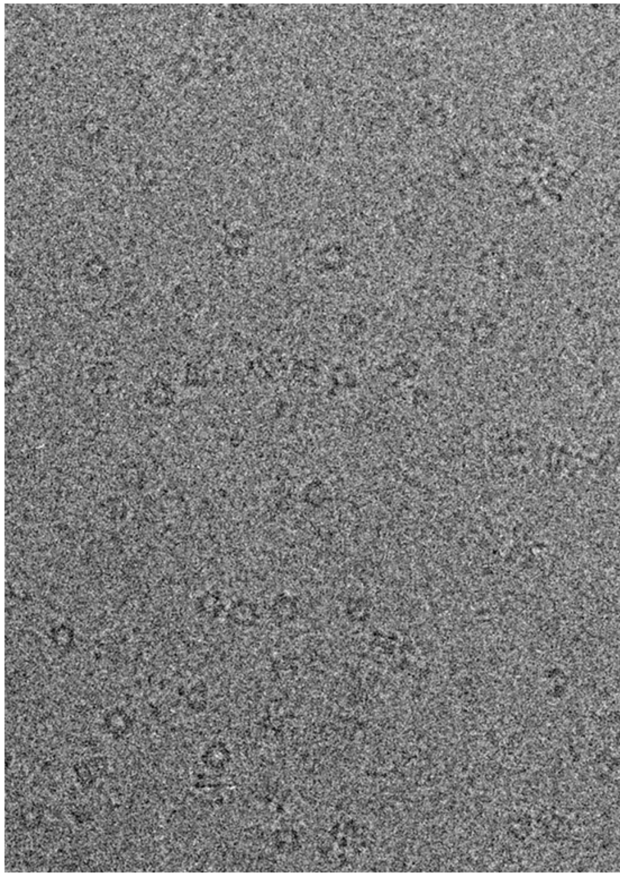


CryoGAN

- Generator



Change $\mathbf{x}$ to output high value



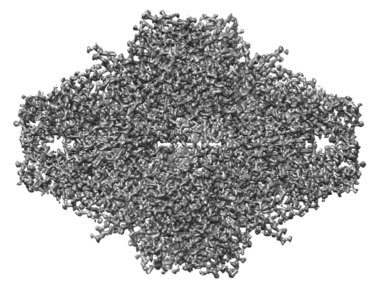
Experiments & Results

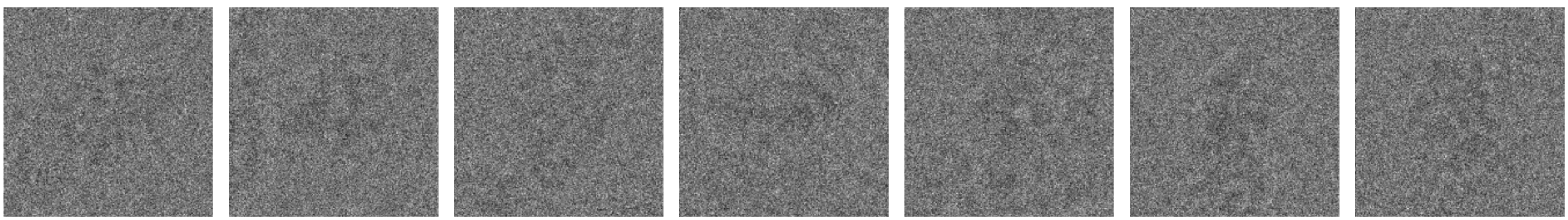
Synthetic experiments

Real experiments

Synthetic Experiment (-20 dB Noise)

- Synthetic beta-gal. volume
- Density map fitted on atomic model (2.5 A)
- 41,000 projections, uniformly distributed
- Realistic CTF and noise conditions
- Noise extracted from experimental micrographs





41000 projections

Results (Structure)