Mixture of Neural Fields for Heterogeneous Reconstruction in Cryo-EM

A. Levy*, R. Raghu*, D. Shustin*, A. R.-Y. Peng, H. Li, O. B. Clarke, G. Wetzstein, E. D. Zhong

*equal contribution

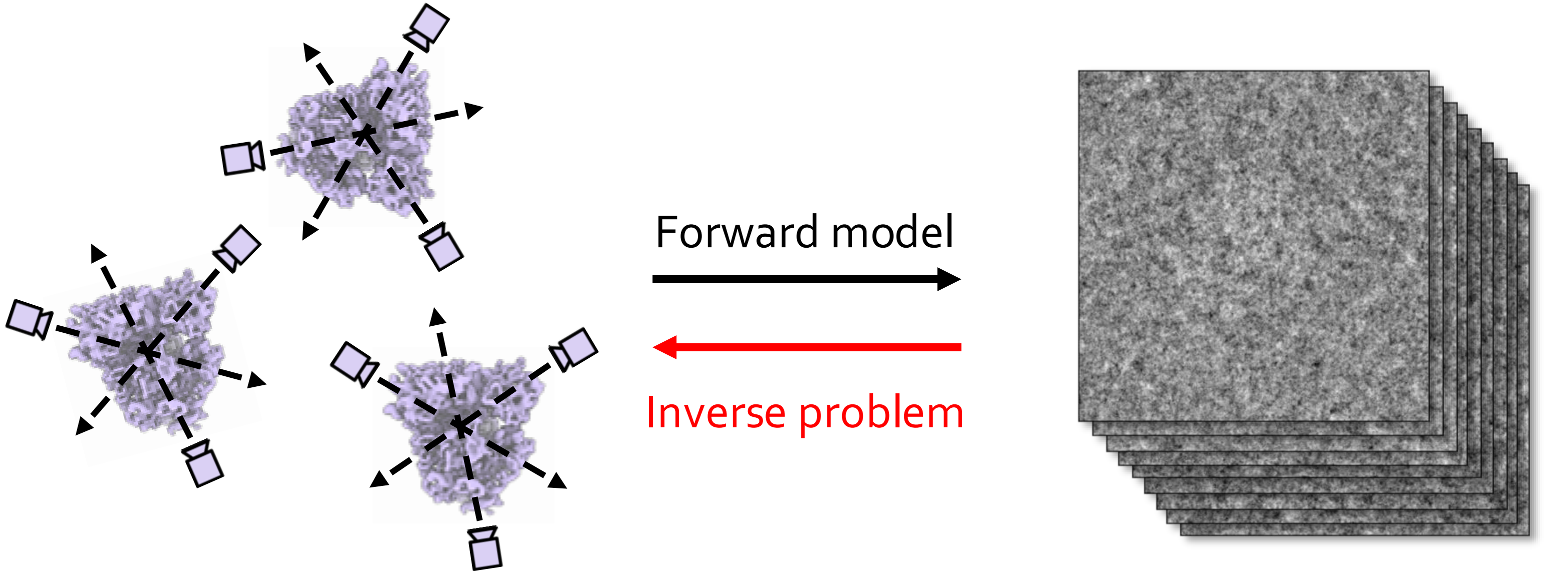
Stanford University

PRINCETON UNIVERSITY

SLAC

COLUMBIA UNIVERSITY

The Challenges of Cryo-EM



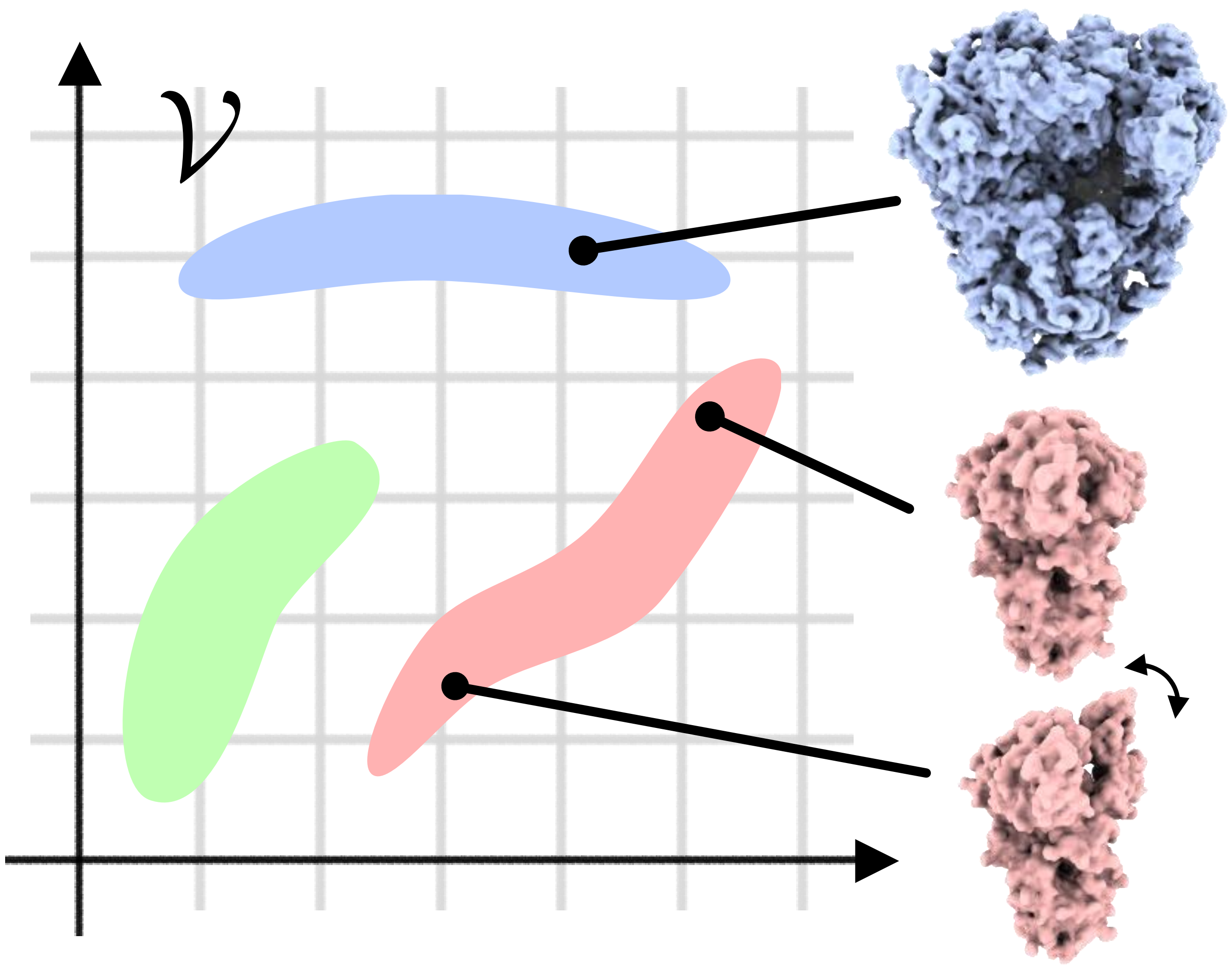
Ensemble of 3D structures

Noisy, **randomly-oriented** views

SNR = -10 to -20dB → 10^5 to 10^6 images

- ➢ Challenge #1: poses (i.e., viewing directions) are unknown
- ➢ Challenge #2: images are extremely noisy
- ➢ Challenge #3: the structure can change from one image to the other

The Conformational Landscape



3D Variability

=

"Compositional" variability

+

"Conformational" variability

Existing Approaches and Limitations

- **Discrete** approaches: RELION (Scheres, 2012), CryoSPARC (Punjani et al., 2017)
  - ✗ Do not handle conformational (i.e., continuous) variability
- **Continuous** methods:
  - Linear combination of voxel arrays (Punjani et al., 2021)
  - Nonlinear (e.g., neural-based) methods (Frank and Ourmazad, 2016; Lederman and Singer, 2017; Maji et al., 2020; Moscovich et al., 2020 Gupta et al., 2020; Zhong et al., 2021; Levy et al., 2022, 2024)
  - Gaussian mixture models (Chen and Ludtke, 2021)
  - Flow fields (Punjani and Fleet, 2023)
  - ✗ Do not handle strong compositional heterogeneity (e.g. different types of proteins)

Hydra – The Working Principles

- *Ab initio* reconstruction method (does not need poses from upstream reconstruction)
- Handles compositional **and** conformational variability using a **mixture of neural fields**

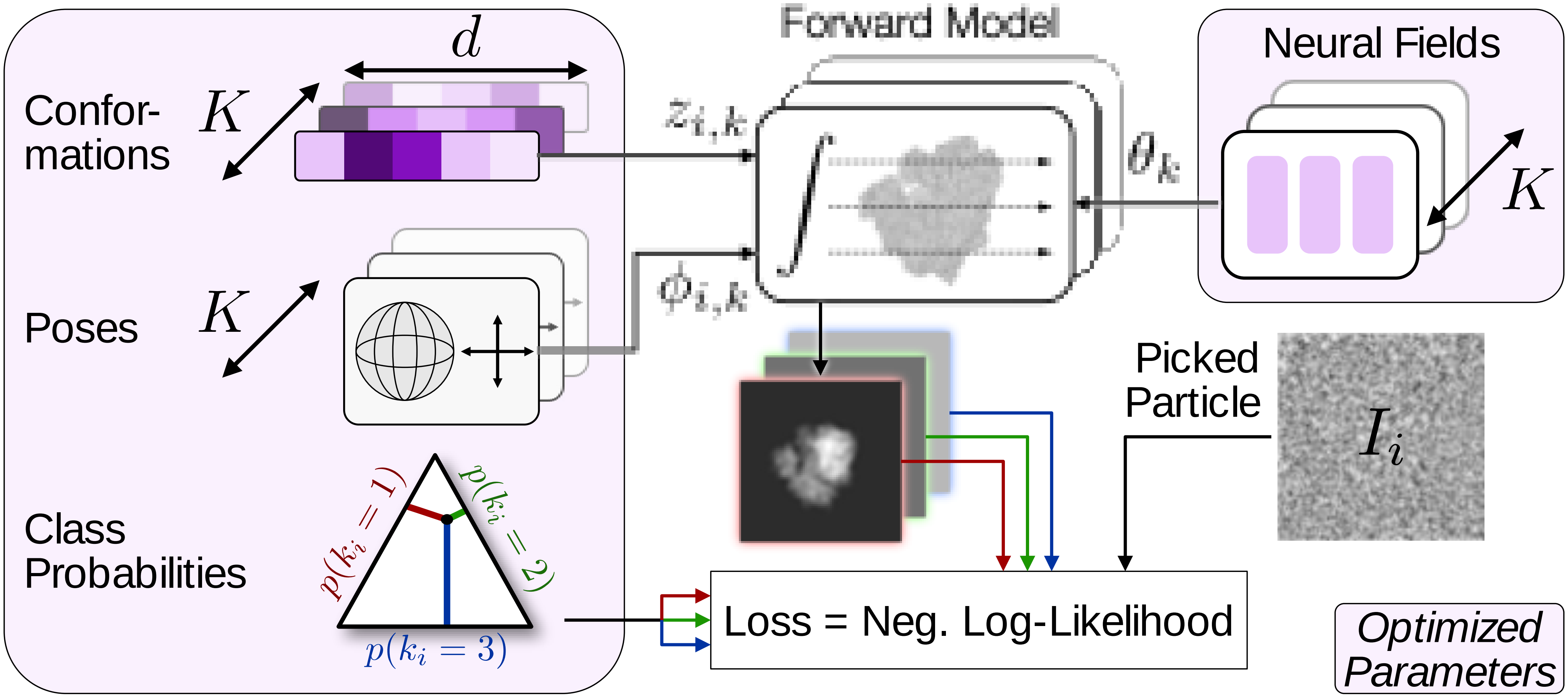
$$I_i = \textsc{Forward}(\mathcal{V}_{\theta_{k_i}}, z_i, \phi_i) + \eta_i$$

Gaussian noise

$\mathcal{V}_\theta : \mathbb{R}^3 \to \mathbb{R}$	Density map (neural network)
$k_i \in \{1.\dots, K\}$	Discrete state variable
$z_i \in \mathbb{R}^d$	Continuous conformational variable
$\phi_i \in \mathrm{SO}(3) \times \mathbb{R}^2$	Pose (i.e., viewing direction)

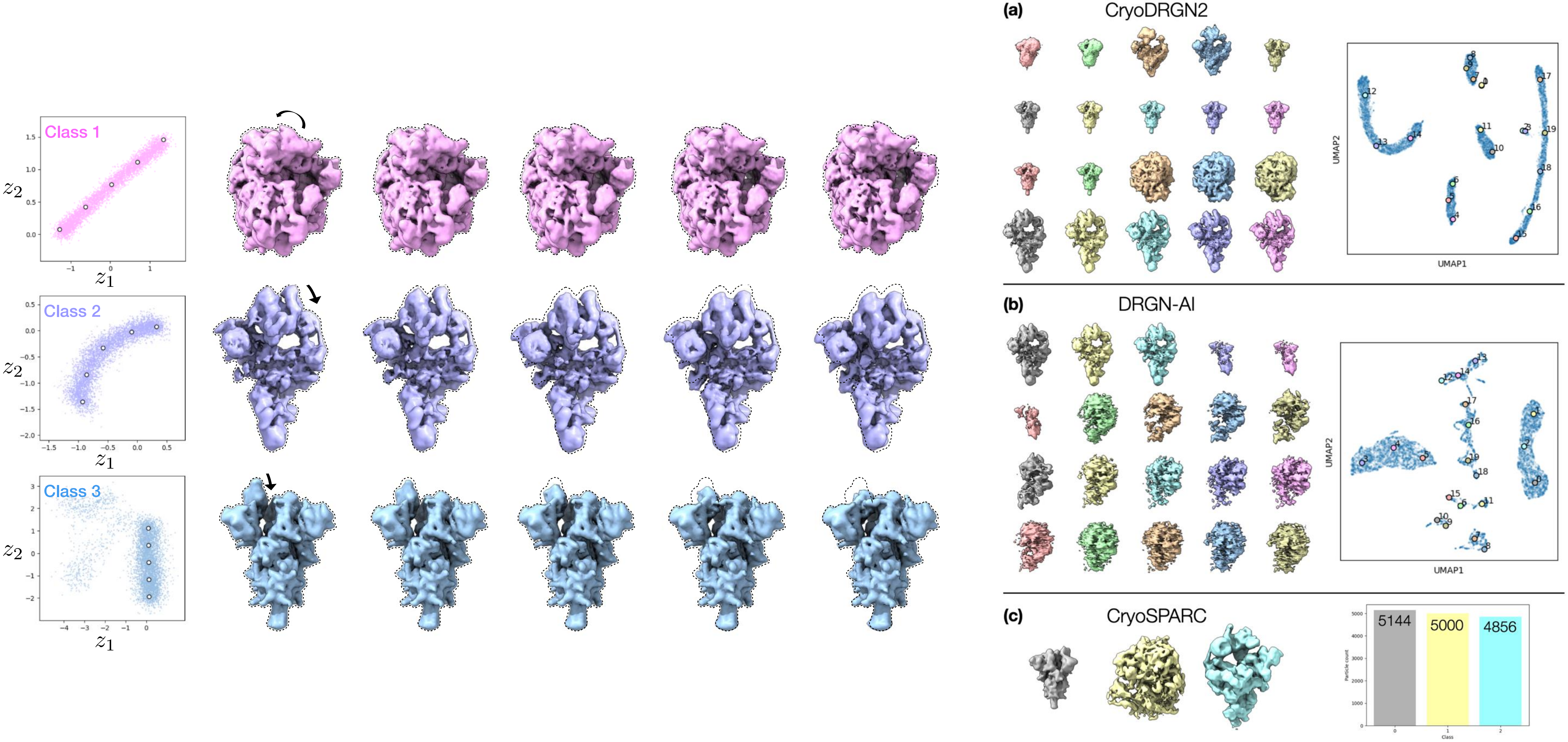
- All the variables are optimized such as to **maximize the likelihood** of observed images

Hydra

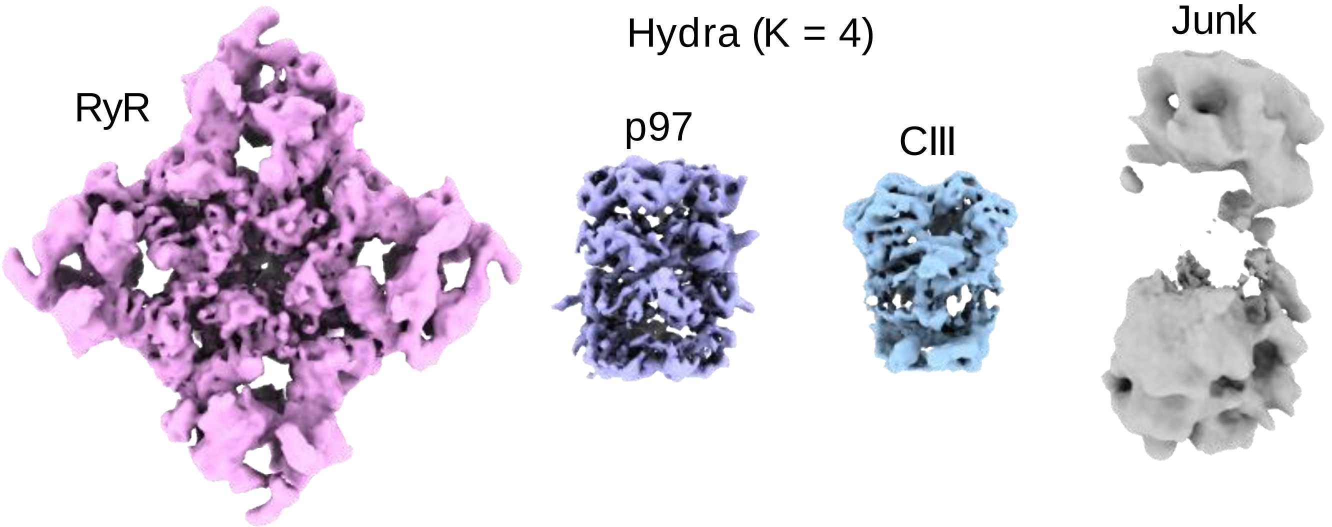


Confor-
mations
K
d
Poses
K
Class
Probabilities
$p(k_i = 1)$
$p(k_i = 2)$
$p(k_i = 3)$
Forward Model
$z_{i,k}$
$\phi_{i,k}$
θ_k
Neural Fields
K
Picked
Particle
I_i
Loss = Neg. Log-Likelihood
Optimized
Parameters

Synthetic Dataset - Ribosplike

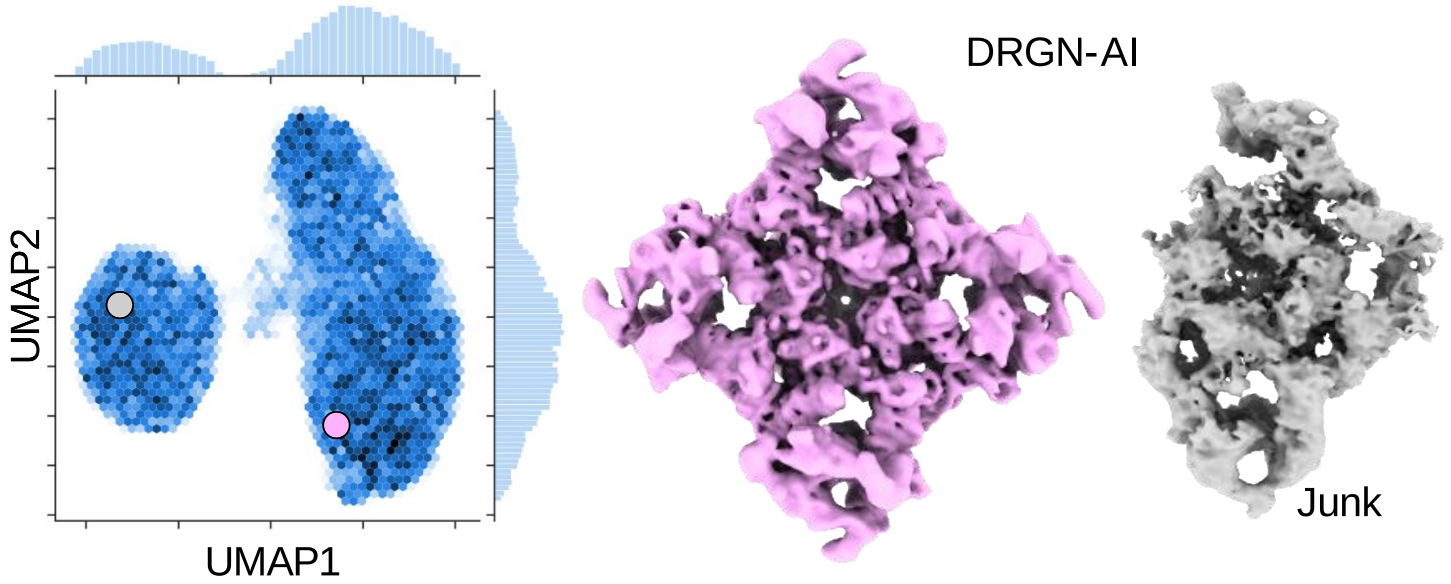


Real Dataset – RyR Mixture



CryoSPARC Labels

Hydra Labels	RyR	p97	CIII	Junk	Total
RyR	60,007	34	98	167	60,306
p97	21	7,073	192	312	7,598
CIII	16	100	7,717	2,553	10,386
Junk	523	641	2,446	3,756	7,366
Total	60,567	7,848	10,453	6,788	85,656



Thank You!