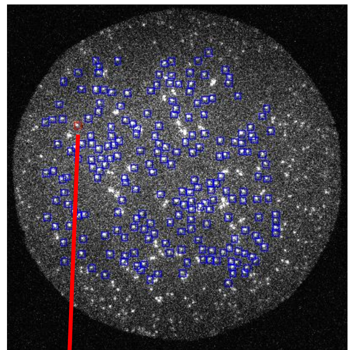
Bayesian classification and modeling of single-molecule fluorescence images using Pyro

Yerdos A. Ordabayev, Larry Friedman, Douglas L. Theobald, Jeff Gelles
Department of Biochemistry, Brandeis University

Colocalization Single-Molecule Spectroscopy

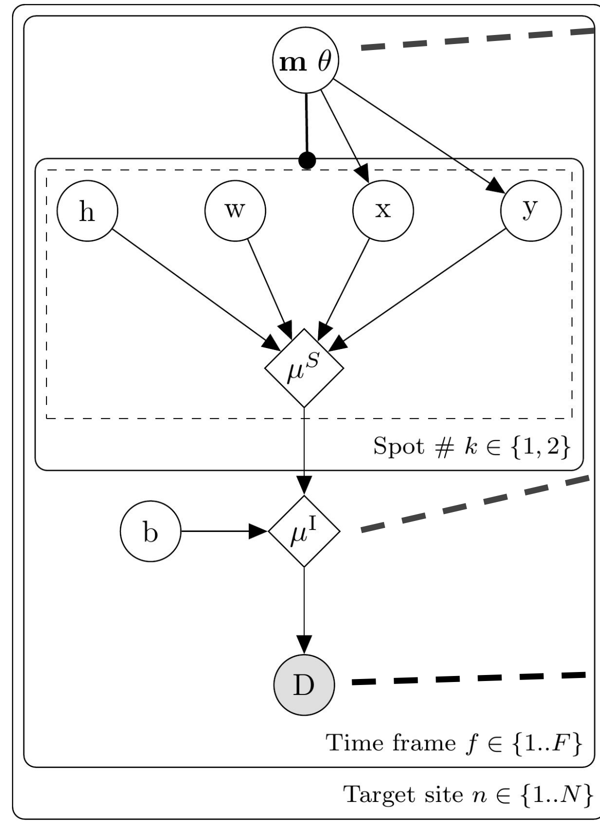
- CoSMoS is used to study the assembly of dynamic molecular complexes in real time
- The individual dye molecules can be detected as discrete spots and the locations visualized
- "**Target**" molecule is tethered to the surface of a microscope slide and localized
- "**Binder**" molecule labeled with a different fluorescent dye is added
- The association and dissociation of "**binder**" molecules with "**the target**" molecule can be monitored as **co-localization** events

Field of view of the microscope



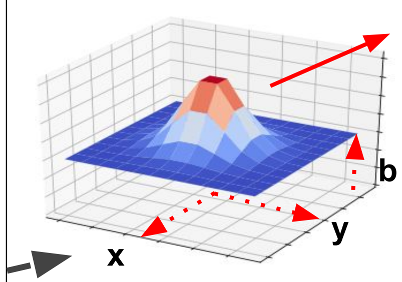
Target site
(Dye labeled DNA molecule)

Graphical Model



m_1, m_2 ∈ {0,1} - spot existence indicator
θ ∈ {0,1,2} - on-target spot index

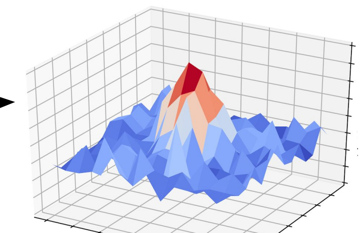
Image Model ($\boldsymbol{\mu}^I$)



2D Gaussian ($\boldsymbol{\mu}^S$)
- Volume (**h**)
- Width (**w**)
- Position (**x**, **y**)

Local background intensity (**b**)

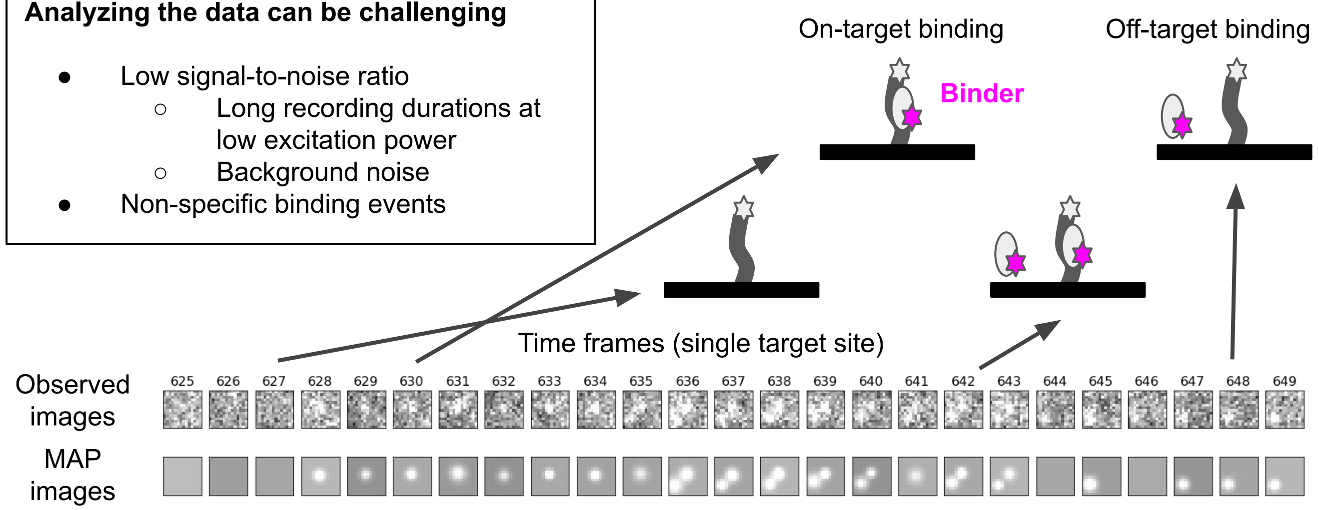
Observed Image (**D**)



Intensity Model
- Shot noise
- Gain amplification
- Offset noise

Analyzing the data can be challenging

- Low signal-to-noise ratio
  - Long recording durations at low excitation power
  - Background noise
- Non-specific binding events



Bayesian (probabilistic) analysis of the image data

- Discriminate authentic fluorescence spots from noise in a probabilistic manner
- Embed prior knowledge about the experiment (e.g., the likely positions of an on-target and an off-target spots)
- Maximize extraction of useful information from data by analyzing 2D images
- Use a realistic pixel intensity model that accounts for background, spot, and camera noise sources
- Explicitly model non-specific interactions of binder molecules with the surface

Inference

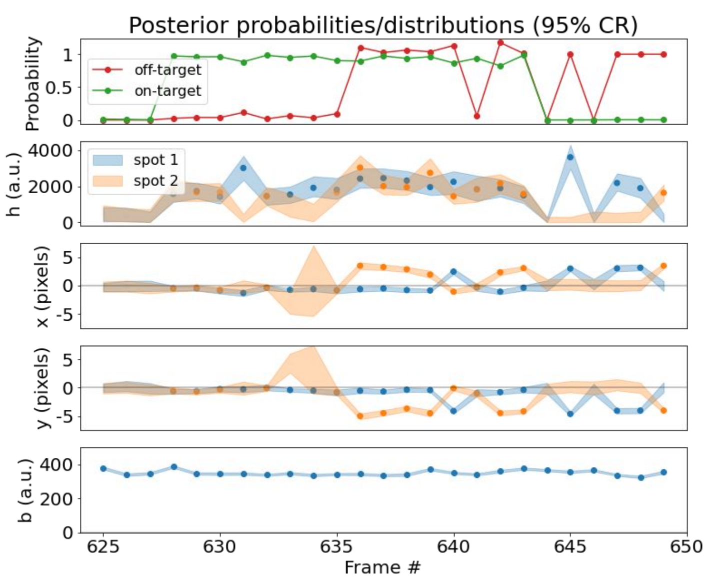
$$p(\mathbf{D}|\mathbf{b},\mathbf{m},\mathbf{h},\mathbf{w},\mathbf{x},\mathbf{y})p(\mathbf{b})p(\mathbf{h})^{\mathbf{m}}p(\mathbf{w})^{\mathbf{m}}p(\mathbf{x}|\theta)^{\mathbf{m}}p(\mathbf{y}|\theta)^{\mathbf{m}}p(\mathbf{m},\theta)$$

Joint Distribution Factorization

$$p(\mathbf{b},\mathbf{m},\theta,\mathbf{h},\mathbf{w},\mathbf{x},\mathbf{y}|\mathbf{D}) \simeq q(\mathbf{b})q(\mathbf{m},\theta)q(\mathbf{h})^{\mathbf{m}}q(\mathbf{w})^{\mathbf{m}}q(\mathbf{x})^{\mathbf{m}}q(\mathbf{y})^{\mathbf{m}}$$

Stochastic Variational Inference (**Pyro**)

Posterior probabilities/distributions (95% CR)



Our new method is

- Fast and scalable to large data sets (GPU computing)
- Requires minimal parameter tweaking
- Has a flexible framework that can naturally be extended to kinetic models and multi-wavelength analysis

Acknowledgments: This work is supported by grant R01GM121384 from the National Institute of General Medical Sciences, NIH.