TRamWAy: probabilistic pipelines approaches to single biomolecule dynamics

Decision and Bayesian Computation
Jean-Baptiste Masson
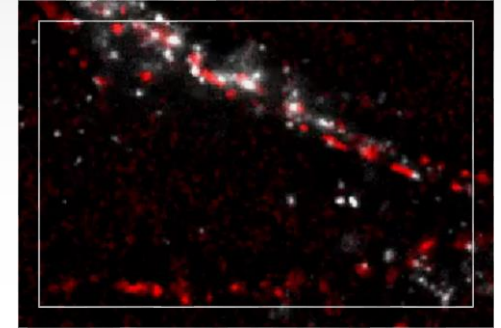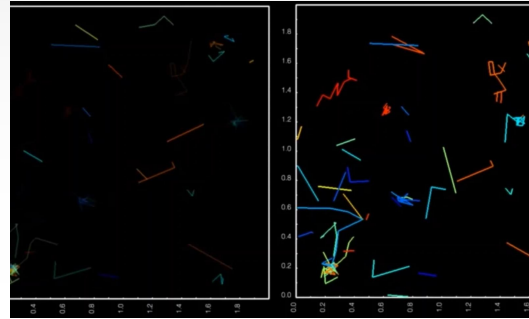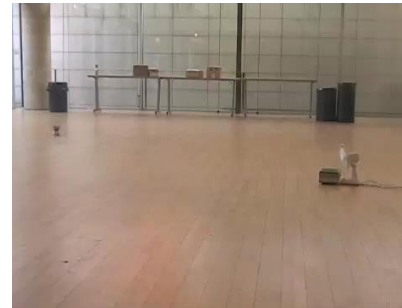**Institut Pasteur**
CNRS USR 3756 / CNRS UMR 3571
Computational Biology / NeuroScience Department
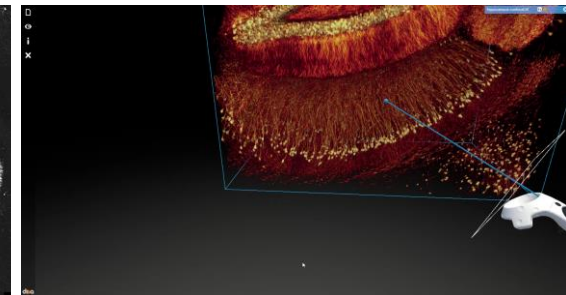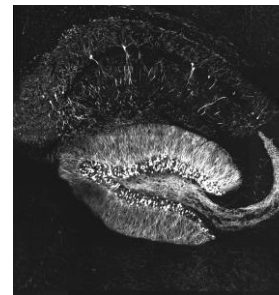
3 research axis of the lab

Probing cells with Random Walks
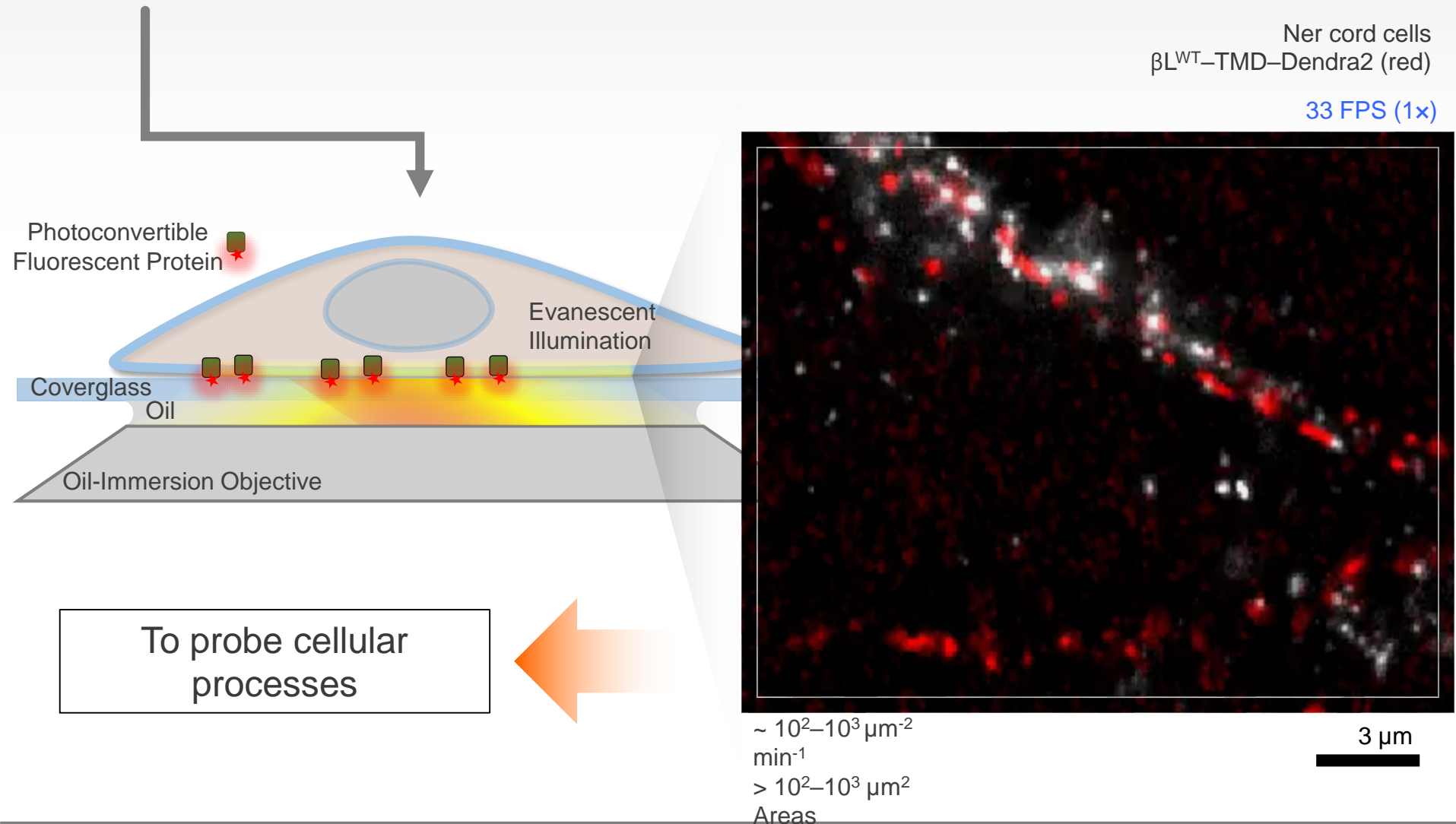

Deciding in complex environments

Data visualization and treatment in Virtual Reality

High-Density Single-Particle Tracking

1. uPAINT (Giannone, *et al.*, *Biophysical Journal*, 2010)
2. sptPALM (Manley, *et al.*, *Nature Methods*, 2008)

Ner cord cells
βL^{WT}–TMD–Dendra2 (red)

33 FPS (1×)

Photoconvertible
Fluorescent Protein

Evanescent
Illumination

Coverglass

Oil

Oil-Immersion Objective

To probe cellular processes

~ 10^2–10^3 µm^{-2} min^{-1}
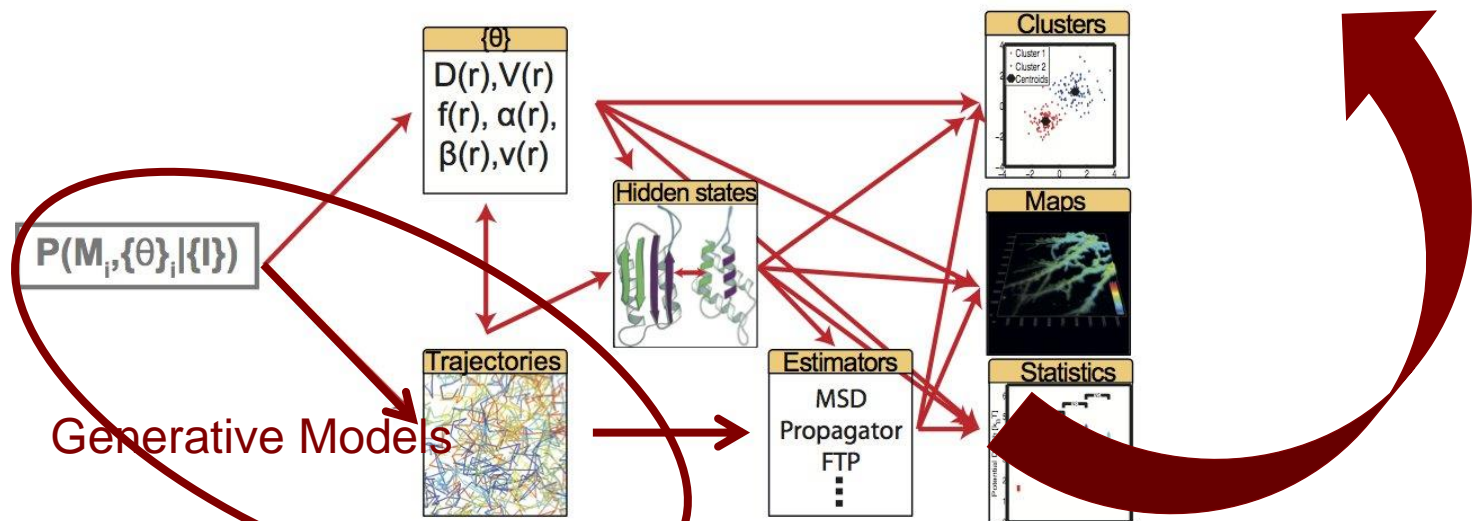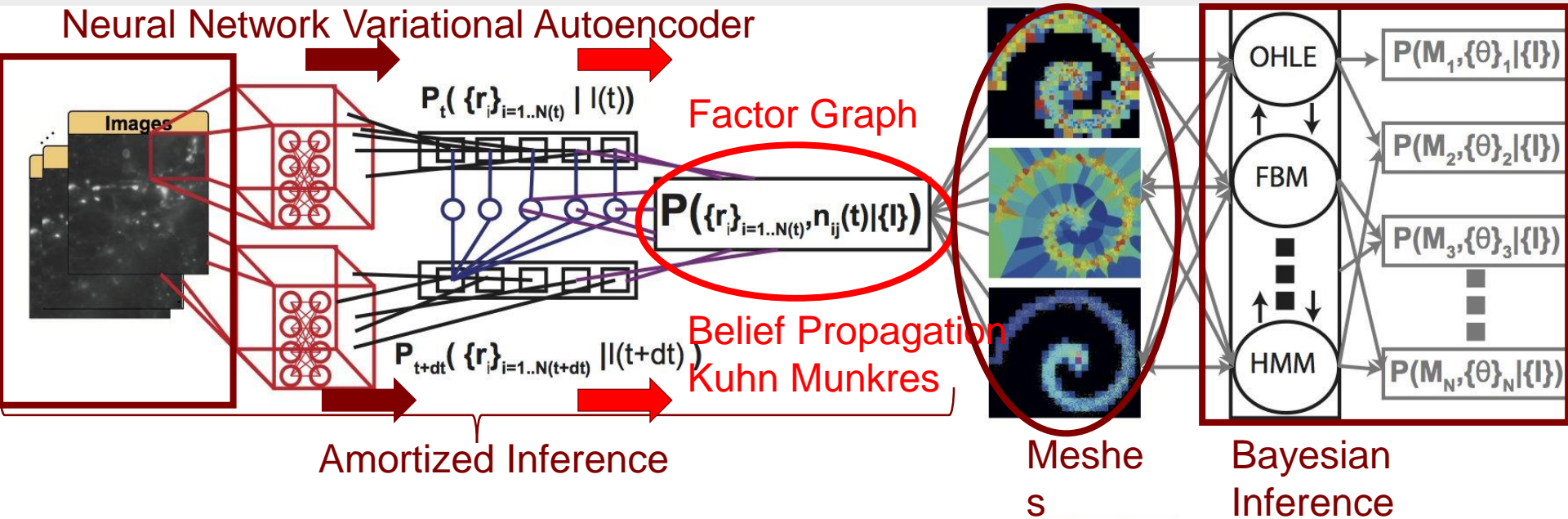> 10^2–10^3 µm^2 Areas

3 µm

Biology

- Dynamics dominated by noise

- Heterogeneity in space, time at all scales

- Heterogeneity in state: phosphorilation, methylation conformation

- Unknown number of interacting partners

- Single molecule experiments are single cell experiments

Analysis: probabilistic pipelines

- 3 steps
    - Localization
    - Tracking
    - Estimating

- Cumulative errors

- **Probabilistic pipelines**

- Exploratory analysis

- **Robustness** vs optimality

- Computation complexity

Probabilistic Pipeline: TRamWAy

Neural Network Variational Autoencoder

$P_t(\{r_i\}_{i=1..N(t)} | I(t))$

Factor Graph

$P(\{r_i\}_{i=1..N(t)}, n_{ij}(t)|\{I\})$

$P_{t+dt}(\{r_i\}_{i=1..N(t+dt)} | I(t+dt))$

Belief Propagation
Kuhn Munkres

Amortized Inference

Meshes

Bayesian Inference

OHLE

FBM

HMM

$P(M_1,\{\theta\}_1|\{I\})$

$P(M_2,\{\theta\}_2|\{I\})$

$P(M_3,\{\theta\}_3|\{I\})$

$P(M_N,\{\theta\}_N|\{I\})$

$\{\theta\}$
$D(r),V(r)$
$f(r), \alpha(r),$
$\beta(r),v(r)$

$P(M_i,\{\theta\}_i|\{I\})$

Generative Models

Hidden states

Trajectories

Estimators
MSD
Propagator
FTP

Clusters

Maps

Statistics

D. P. Kingma *et al*, arXiv. 2013.
W Ouyang *et al, Nat Biotech 2018* M. Mezard & A. Montanari 2012 Cherktkov *et al, PNAS* , 2016 C. Vestergaard et al (In Preparation

Deconvolving with Neural Nework

- Learning on numerically generated data
- Aiming for robustness
  - Trained on noise free data
  - Increase noise in second training
- Reconstitute Gaussians in High dimension
- From high dimension Gaussians to Pixels position
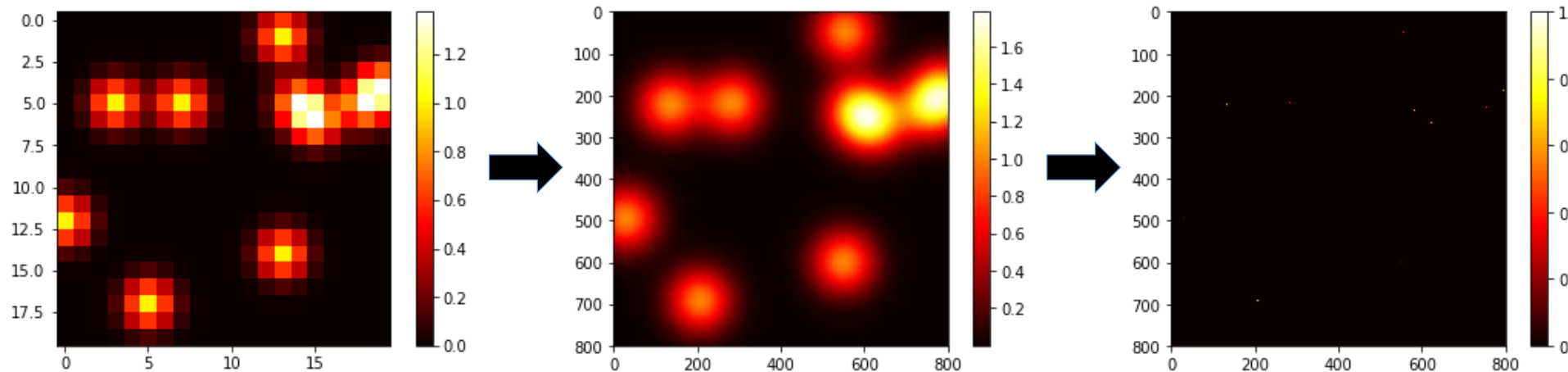
- Efficient in small sliding window
- Needs efficient space sampling
- Trained on full size image (ongoing)

Shi *et al,* arXiv:1609.05158v2 201
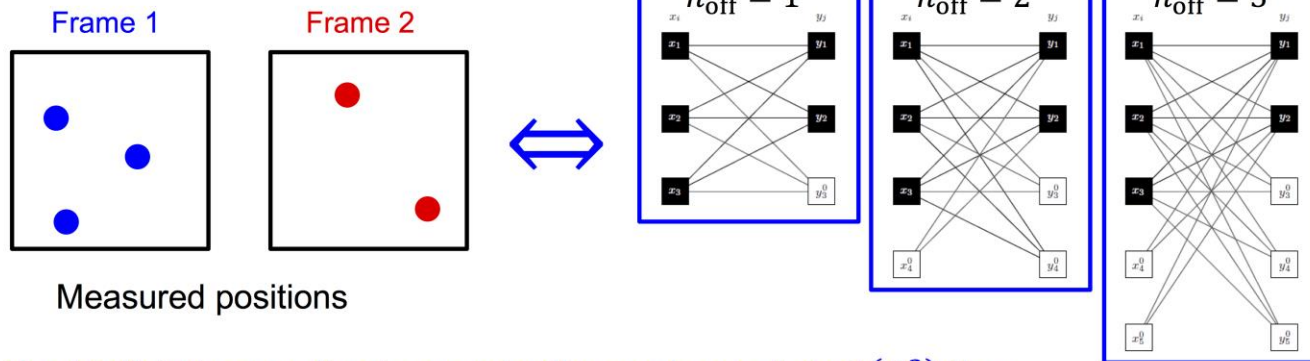W Ouyang *et al, Nat Biotech 2018*
Nehme *et al, optica 2018*

Non Tracking: Sum over ghost graphs

Calculating the marginal likelihood with photoswitching

Maximize marginalized likelihood:

sum over n_{off}

probability for n_{off} molecules to disappear

$$\mathcal{L}(\boldsymbol{\theta}|\mathbf{r}_1, \mathbf{r}_2) = \sum_{n_{\text{off}}} \sum_{\bar{\bar{\mathbf{s}}}} p(\mathbf{r}_1, \mathbf{r}_2, \bar{\bar{\mathbf{s}}}|n_{\text{off}}, \boldsymbol{\theta}) \, p(n_{\text{off}}) \, .$$

likelihood

probability of matching given n_{off}

Map to graph matching problems:

Frame 1   Frame 2

$\Longleftrightarrow$

$n_{\text{off}} = 1$ $n_{\text{off}} = 2$ $n_{\text{off}} = 3$

Measured positions

Use *Belief Propagation* to marginalize each graph in $O(N^2)$ time

Non-Tracking: BP on all the Graphs
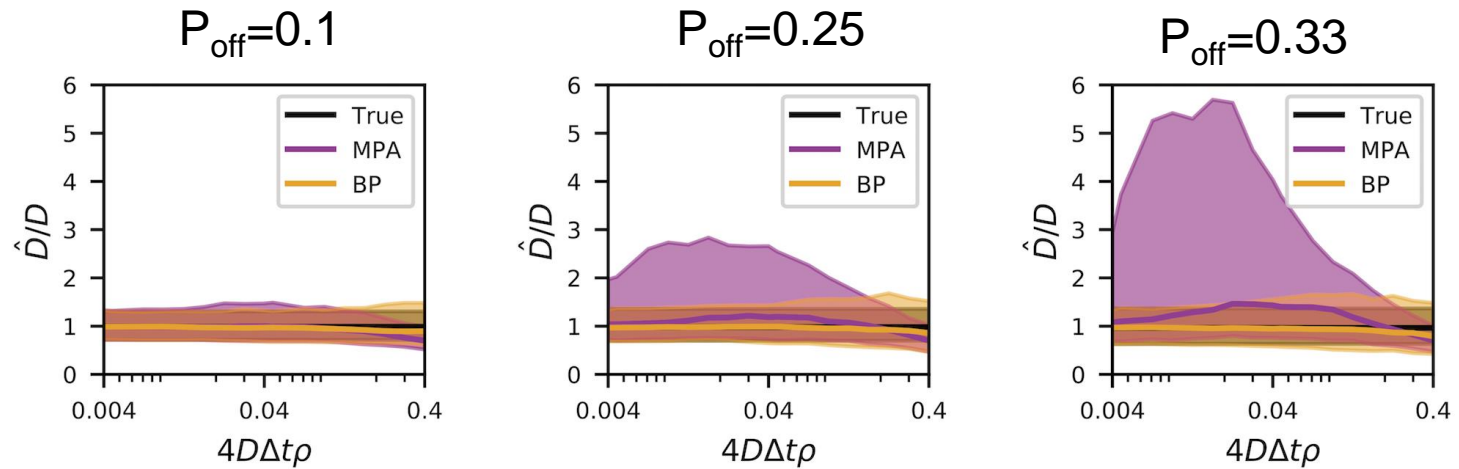
Approximating the likelihood using Belief Propagation

$P_{off}=0.1$ $P_{off}=0.25$ $P_{off}=0.33$

Equivalent to sending messages between images

Frame 1   Frame 2

$h^{1 \to 2}$

Chertkov *et al.*, PNAS (2010)

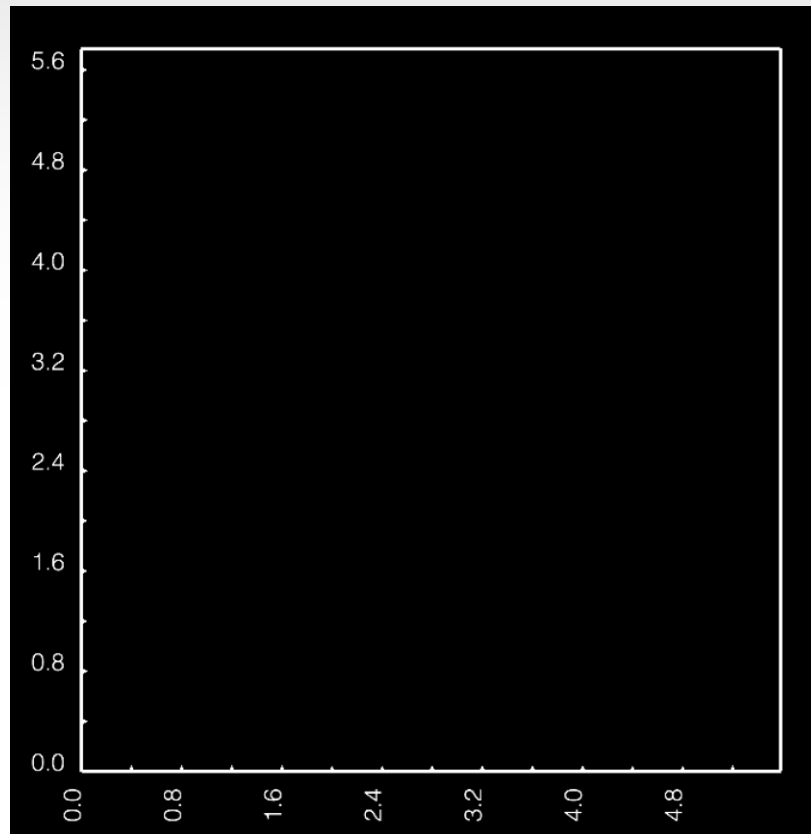**Messages:**

$$h^{i \to j} = -\ln \sum_{j' \neq j} p_{ij'} e^{h^{j' \to i}}$$

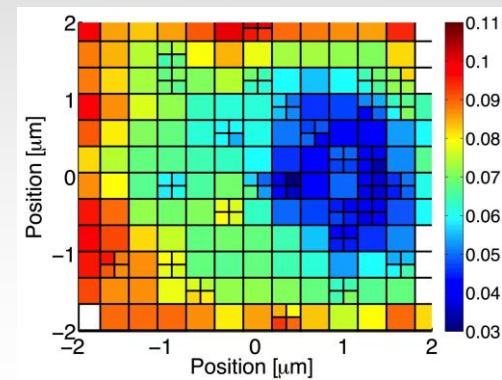$$h^{j \to i} = -\ln \sum_{i' \neq i} p_{i'j} e^{h^{i' \to j}}$$

Parallel implementation

Bethe free energy:

$$\mathcal{F}_B = +\sum_{i,j} \ln \left(1 + p_{ij} e^{h^{i \to j} + h^{j \to i}} \right)$$

$$- \sum_i \ln \left(\sum_j p_{ij} e^{h^{j \to i}} \right)$$

$$- \sum_j \ln \left(\sum_i p_{ij} e^{h^{i \to j}} \right).$$

C. Vestergaard et al, PRX (in preparation)
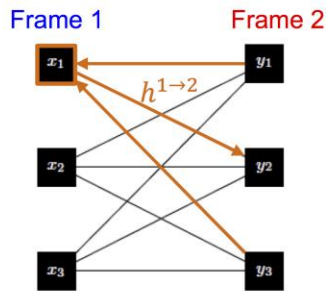
Temporal evolution of diffusive trap

t=0s

t=30s

t=60s

Numerical Simulations
Scaffolding proteins evolution
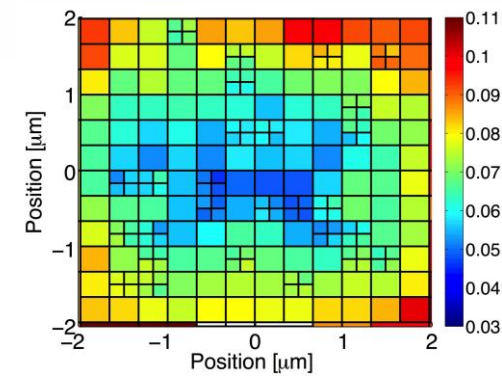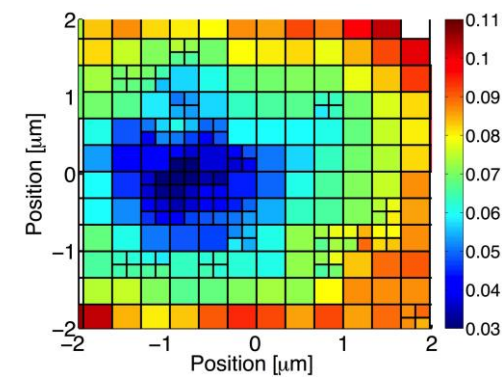
The Bayesian Framework

$$\text{Bayes' Law} \qquad \underset{\text{Posterior}}{P(U \mid T)} = \frac{\overset{\text{Likelihood}}{P(T \mid U)}\overset{\text{Prior}}{P(U)}}{\underset{\text{Evidence}}{P(T)}}$$

Bayes' Law
For Inference

$$P(U \mid T) \propto P(T \mid U)P(U)$$

Physical Parameters:
Diffusion
Effective Interaction
Energy

Physical Model

Beforehand Knowledge

Smoothing

Maximum *a posteriori* (MAP)

Posterior

U
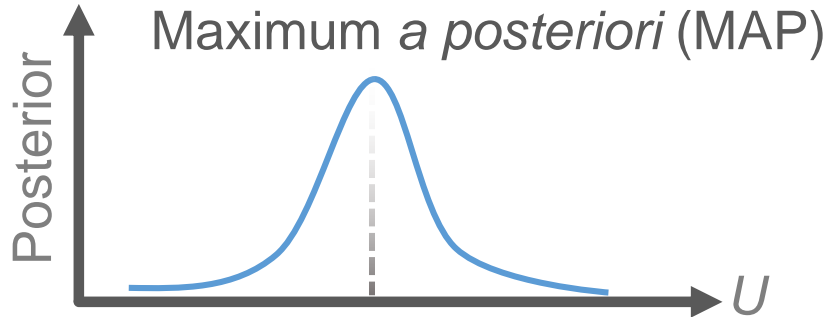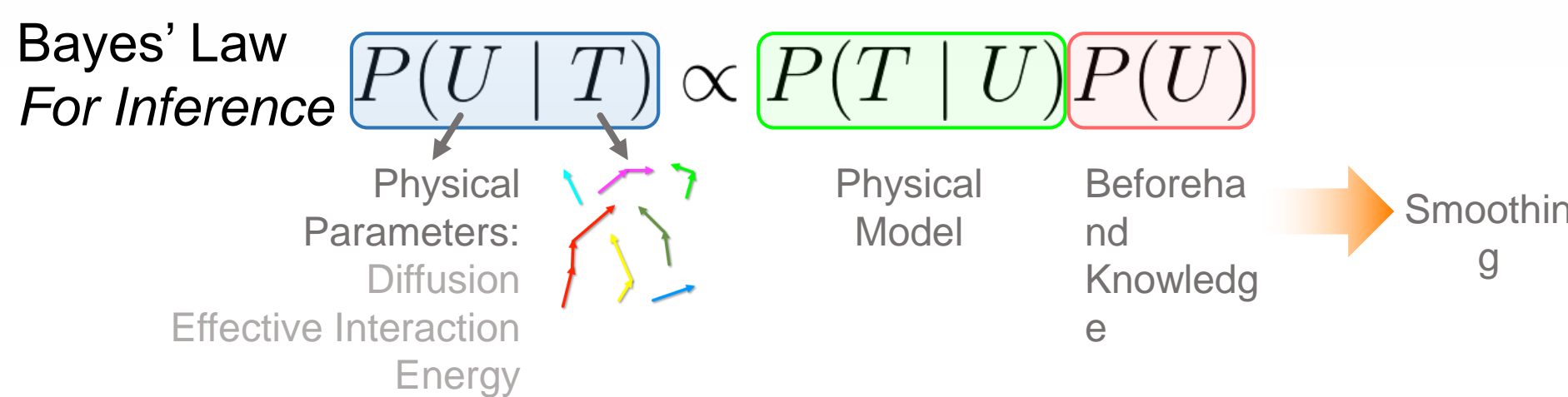
$$U^{\text{MAP}} = \arg\max_{U} P(T \mid U)P(U)$$

Meshes
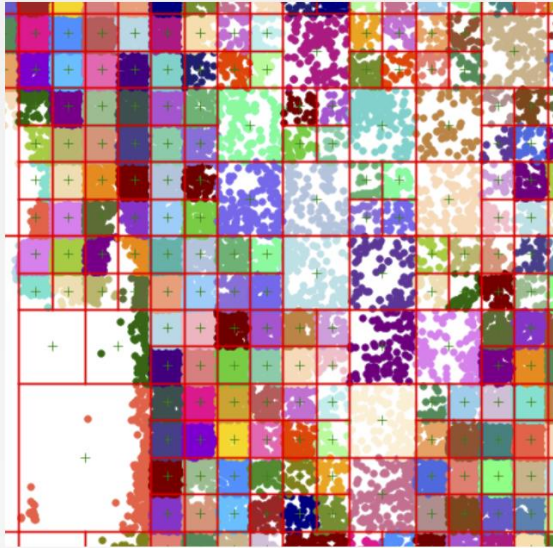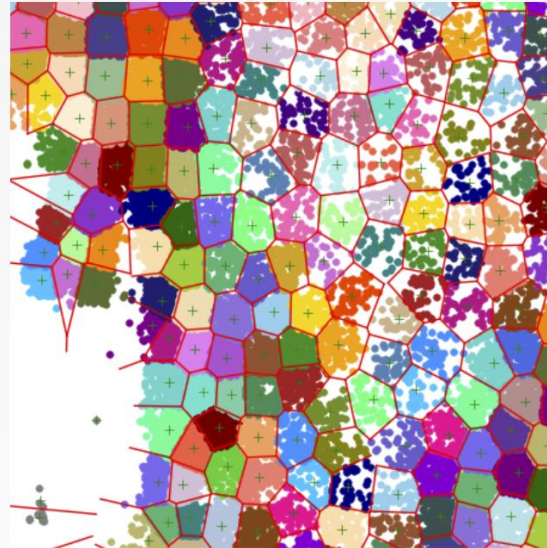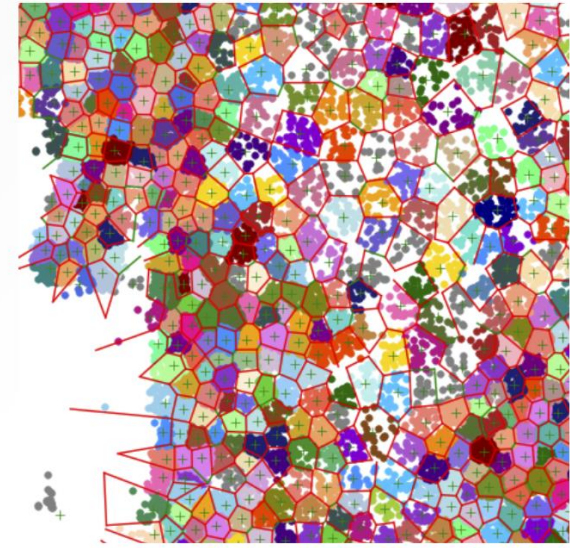
kd-tree   k-means   GWR gas

- Multiple scales
- Multiple conditions on number of points and neighbors domain
- Multiple time scales

Heterogenous Langevin Equation

$$\frac{d\vec{r}}{dt} = D_t(\vec{r})\left[\vec{f}(\vec{r}) + \lambda \frac{\nabla D_t(\vec{r})}{D_t(\vec{r})}\right] + \sqrt{2D_t(\vec{r})}\vec{\xi}(t)$$

- Likelihood of the model is the local solution of the Fokker-Planck

$$\frac{\partial P(\mathrm{r},t|\mathrm{r}_0,t_0)}{\partial t} = -\nabla \cdot \left[\left(\frac{f_t(\mathrm{r})}{g_t(\mathrm{r})} + l\nabla D_t(\mathrm{r})\right)P(\mathrm{r},t|\mathrm{r}_0,t_0)\right] + \nabla \cdot \left[D_t(\mathrm{r})\nabla P(\mathrm{r},t|\mathrm{r}_0,t_0)\right]$$

- Prior $\quad P\big(D(\vec{r}),V(\vec{r})\big) \propto \exp\left(-\mu \iint \left\|\nabla D(\vec{r})\right\|^2 d^2r - \psi \iint \left\|\nabla V(\vec{r})\right\|^2 d^2r\right)$

- Optimize (λ,ψ) at the full dataset scale

JBM *et al.* *PNAS* , 2013.   AI and Physics 2019   M. Sarris, JBM *et al.* *Curr. Biol* , 201

13

3 interpretations

- ## Statistics
  - Low dimensional representation of the dynamics
  - Statistics of value encode partial information of the undergoing process
  - Statistical test between conditions

- ## Physics
  - Physical meaning of the underlying process
  - Physical assumptions in the model
  - Generative model: predicting other properties

- ## Empirical
  - Test numerous models
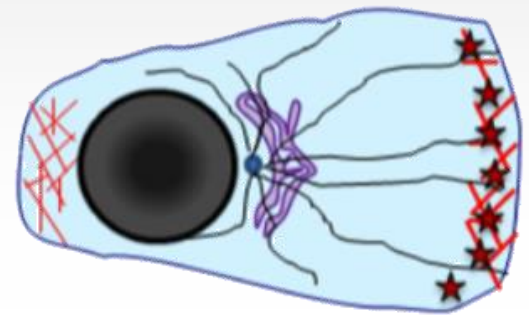 - Sample all parameters and feature to search for information

Mapping as a Readout to Biomolecules
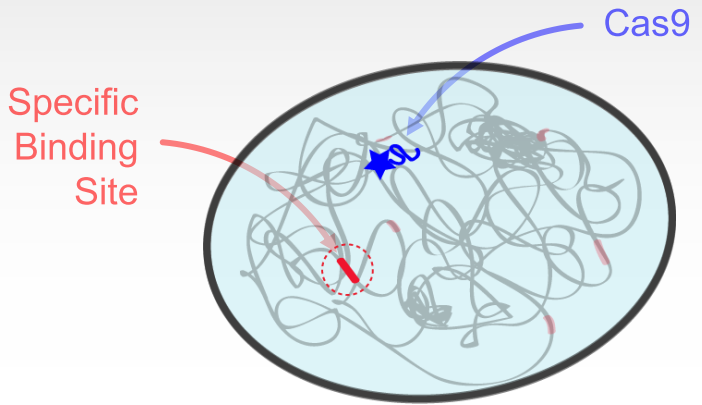
CRISPR-Cas9 genome interrogation

Knight, *et al.*, **Science**, 2015

Masson, *et al.*, **Biophysical Journal**, 2014

Molecular Mechanisms of Cell Motility

Remorino, *et al.*, **Cell reports**, 2017

Floderer et al, **Sci. rep** . 2018
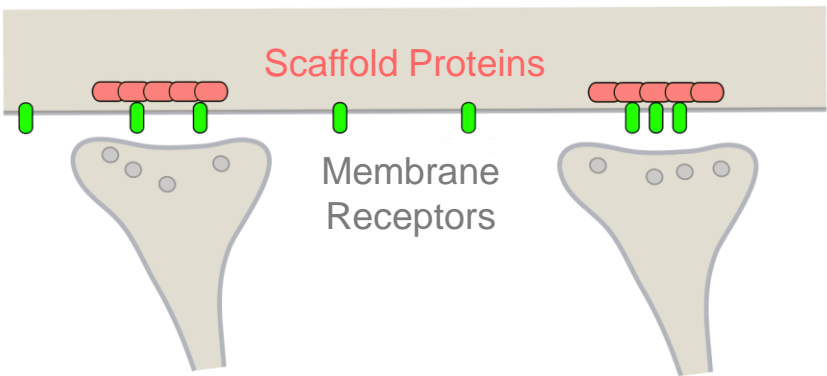
Receptor-Scaffold Interactions
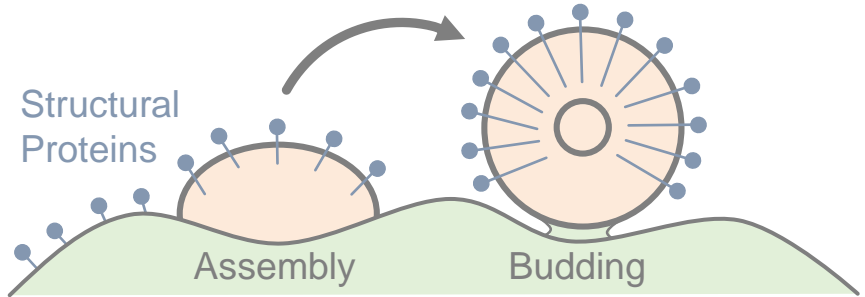
Dynamics of Virion Assembly

Virion formation: Gag dynamics

- How are single viral proteins recruited to virus budding sites once at the cell plasma membrane?

- PALM nanoscopy on Gag

- Time evolving dynamics

- 20 minutes recording

- First use of TramWAy

HIV-1 structural protein Gag assembling at membrane (TEM)

Cellular Factors in HIV-1 Assembly and Budding

Budding

Assembly

Tsg101

?Ub Vps28

Vps4 & ?other factors

Gag

HP68

Vif

?other factors

Jaisri Lingappa Lab, University of Washington

Gag as the VLP assembles

M. El Beheiry *et al*, Nat. Methods (2015)
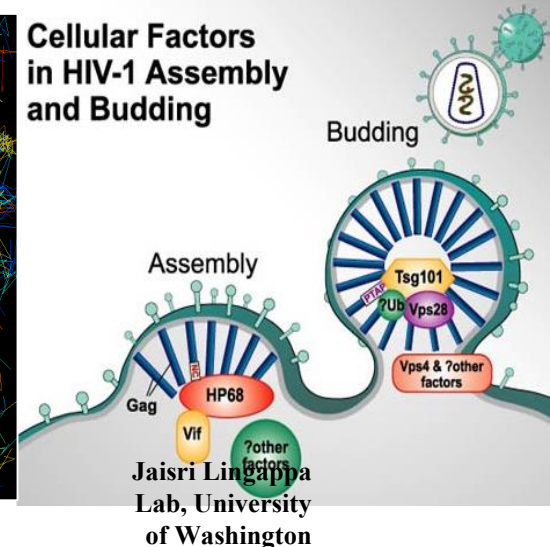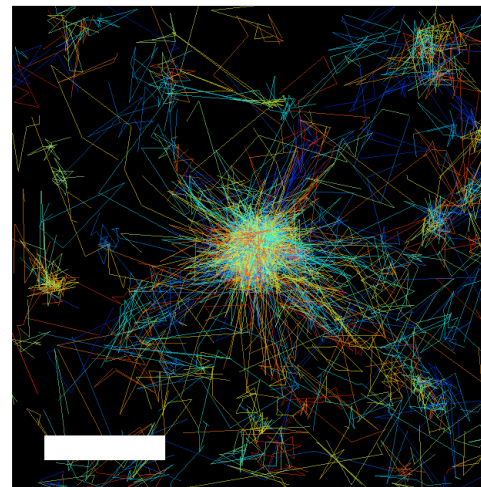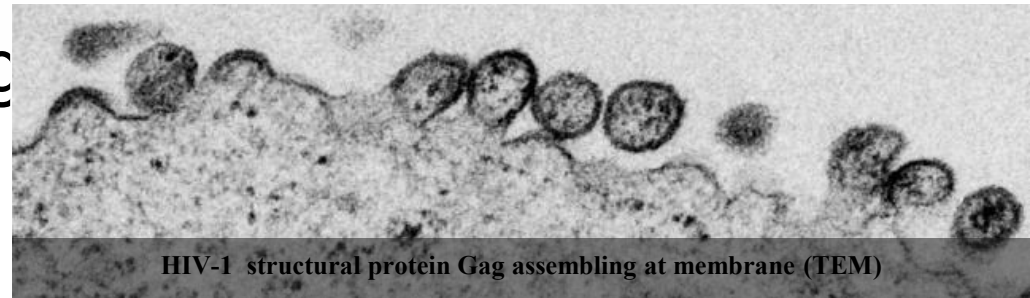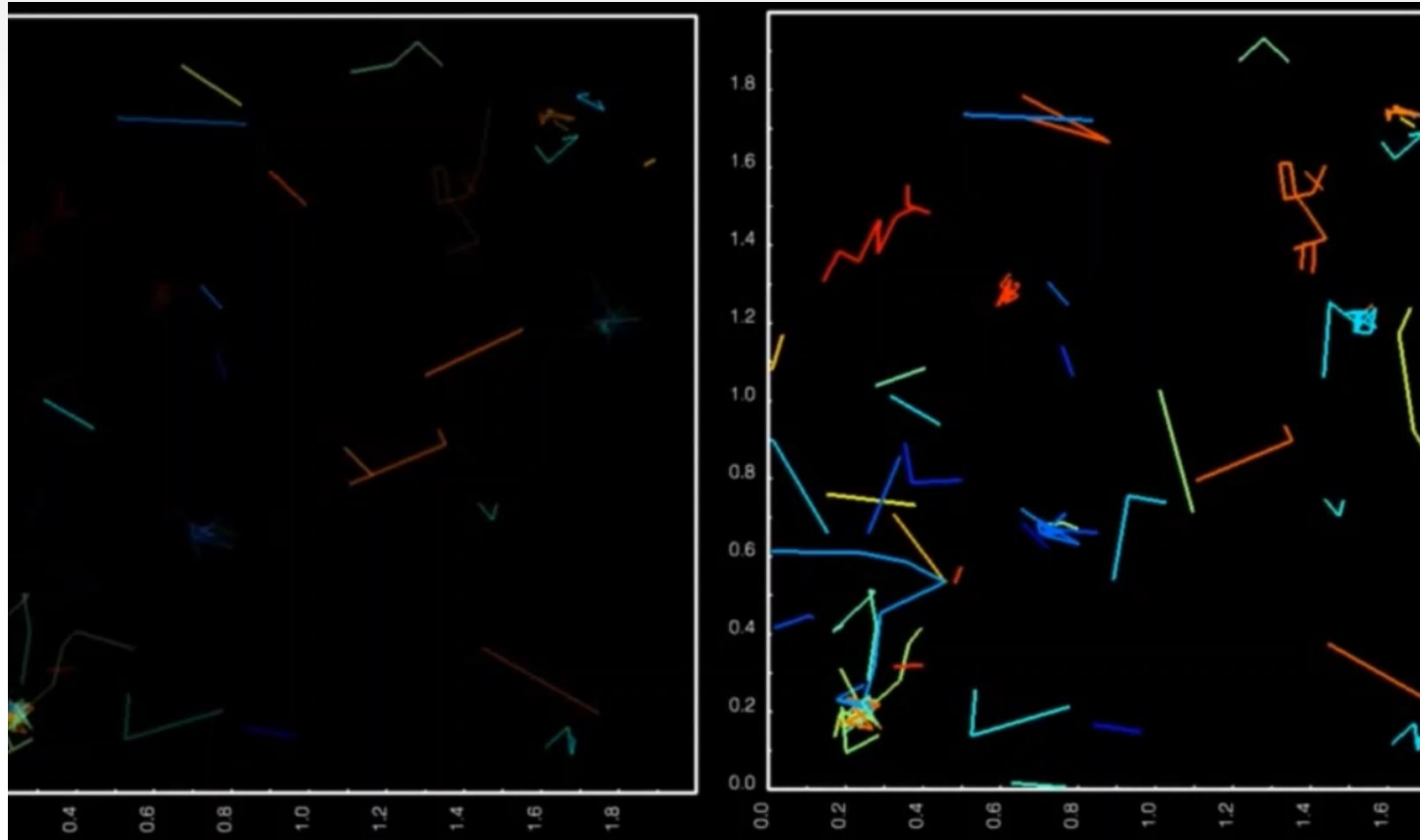
Finding Area of Interest: Bayesian Evidence

- Original Problem: Ito-Stratonovitch dilemma and Bayesian Inference

$$dr\vec{} = \frac{f(\vec{r})}{\gamma(\vec{r})} + \lambda\nabla D(\vec{r}) + \sqrt{2D(\vec{r})}\cdot d\vec{w}(t)$$

- Statistical test : Bayes factor = Ratio of Bayesian Evidence
  - Null Model : rand⬚
  - Active interaction⬚

$$\mathcal{B}^{\mathrm{M}} = \eta^{\mathrm{d}} \frac{\int_0^1 d\lambda \left[v + \eta^2(\zeta_{\mathrm{t}} - \lambda\zeta_{\mathrm{sp}})^2 \right]^{-\kappa(\mathrm{d})}}{\int_0^1 d\lambda \left[v + (\zeta_{\mathrm{t}} - \lambda\zeta_{\mathrm{sp}})^2 \right]^{-\kappa(\mathrm{d})}}$$

$$\zeta_{\mathrm{t}} \equiv \overline{\Delta\mathbf{r}}/\sqrt{V} \quad \text{form} \quad \mathbf{g} \equiv \nabla\mathbf{b}$$

$$\zeta_{\mathrm{sp}} \equiv \hat{\mathbf{g}}\Delta\mathrm{t}/\sqrt{V} \quad V \equiv \sum_{i=1}^{n} |\Delta\mathbf{r}_i - \overline{\Delta\mathbf{r}}|^2/n$$

Finding Areas of interest: Bayesian factor

Detected directional drift:
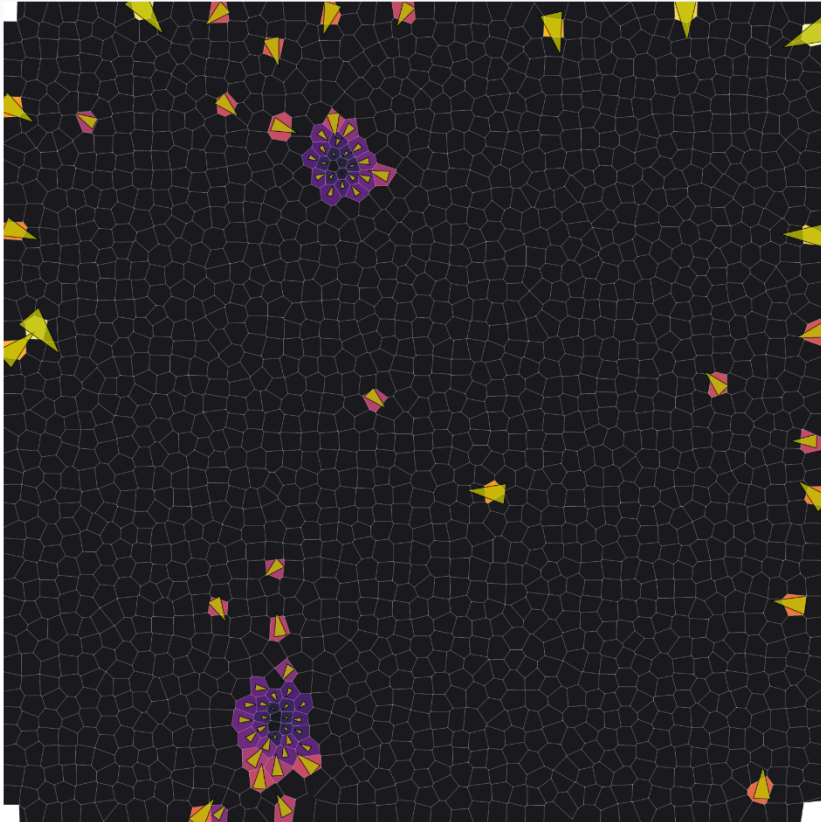
Thresholded Bayes factor:

A. Serov *et al (J. Chem. Ph*
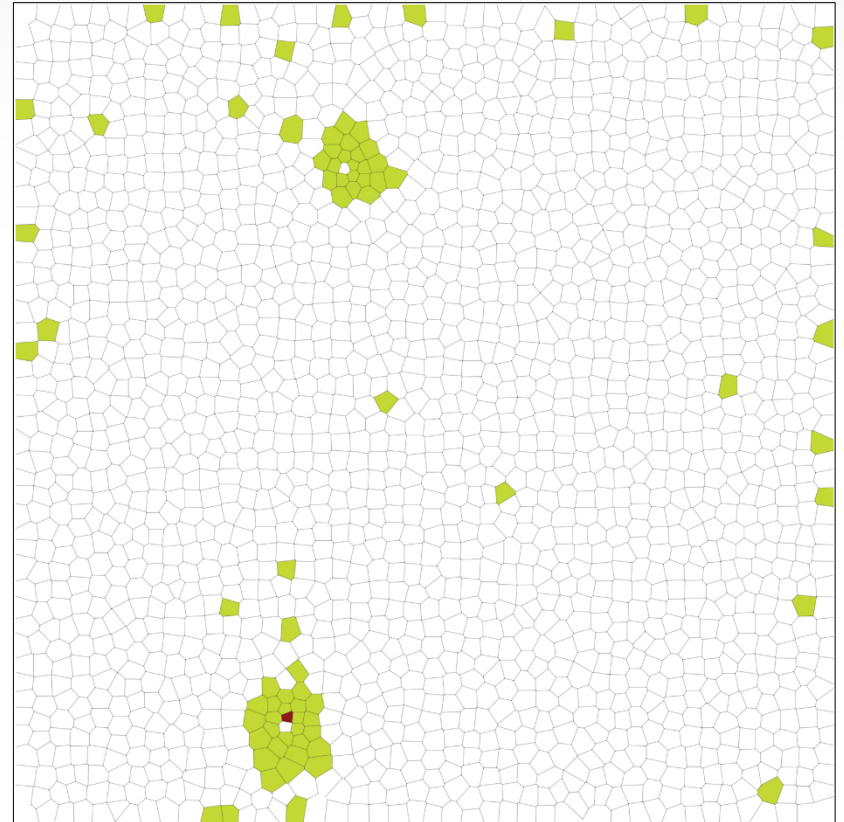
Evolving maps as the VLP assembles:
~1To – 300 000 inferences – 10h

Density

Diffusion

Effective Interaction
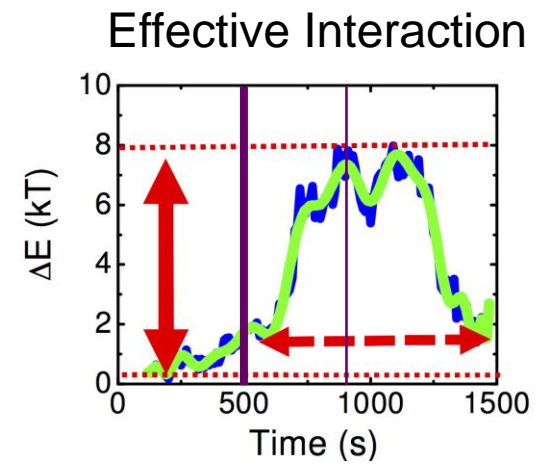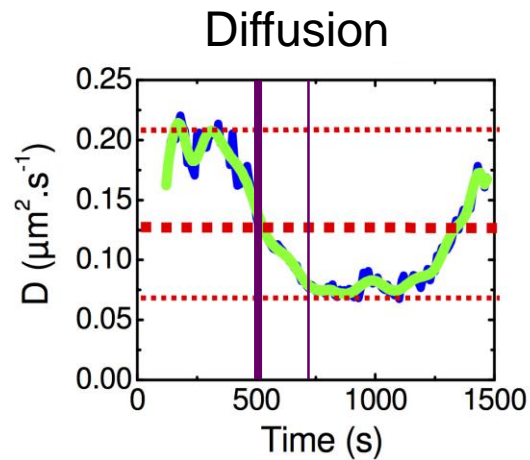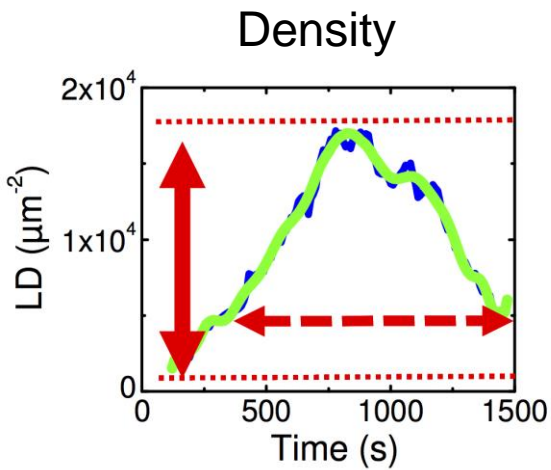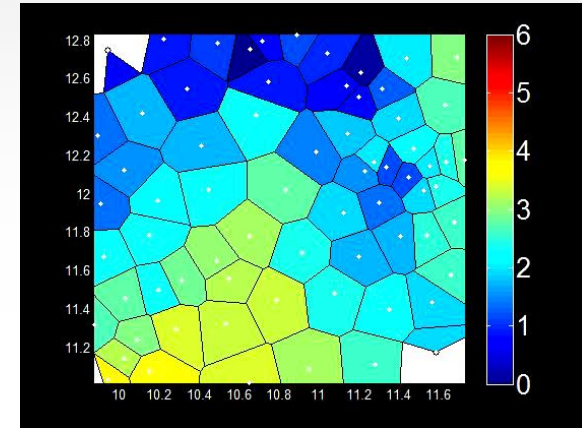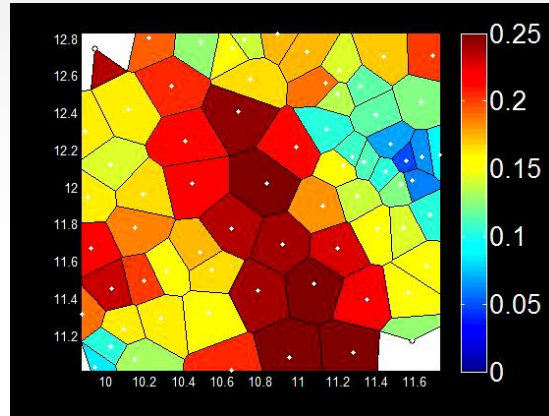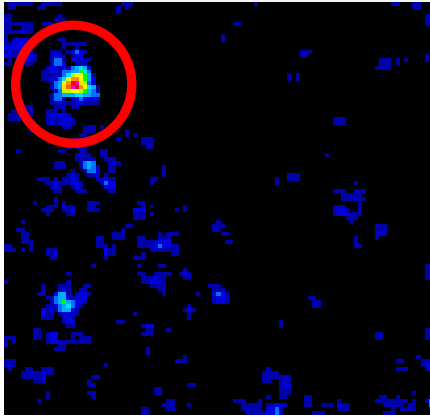
Modifying Gag
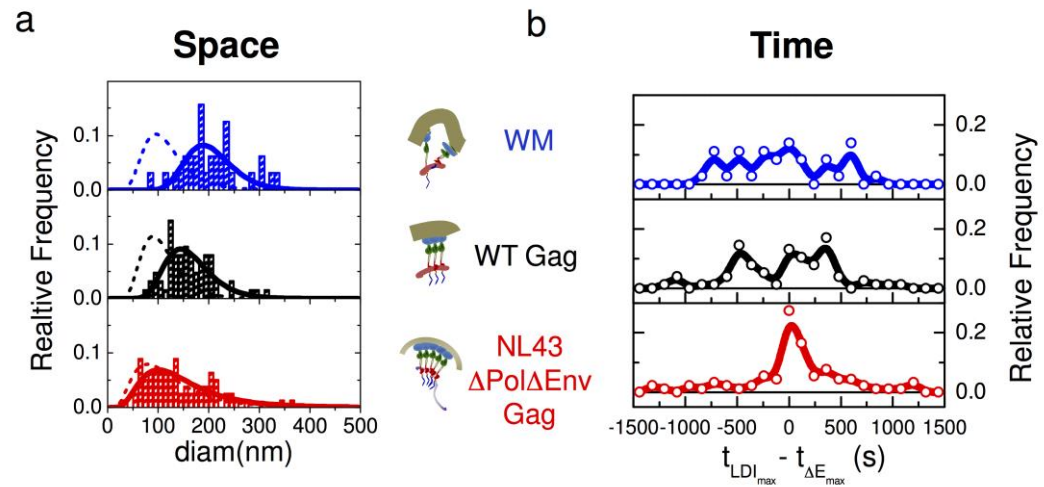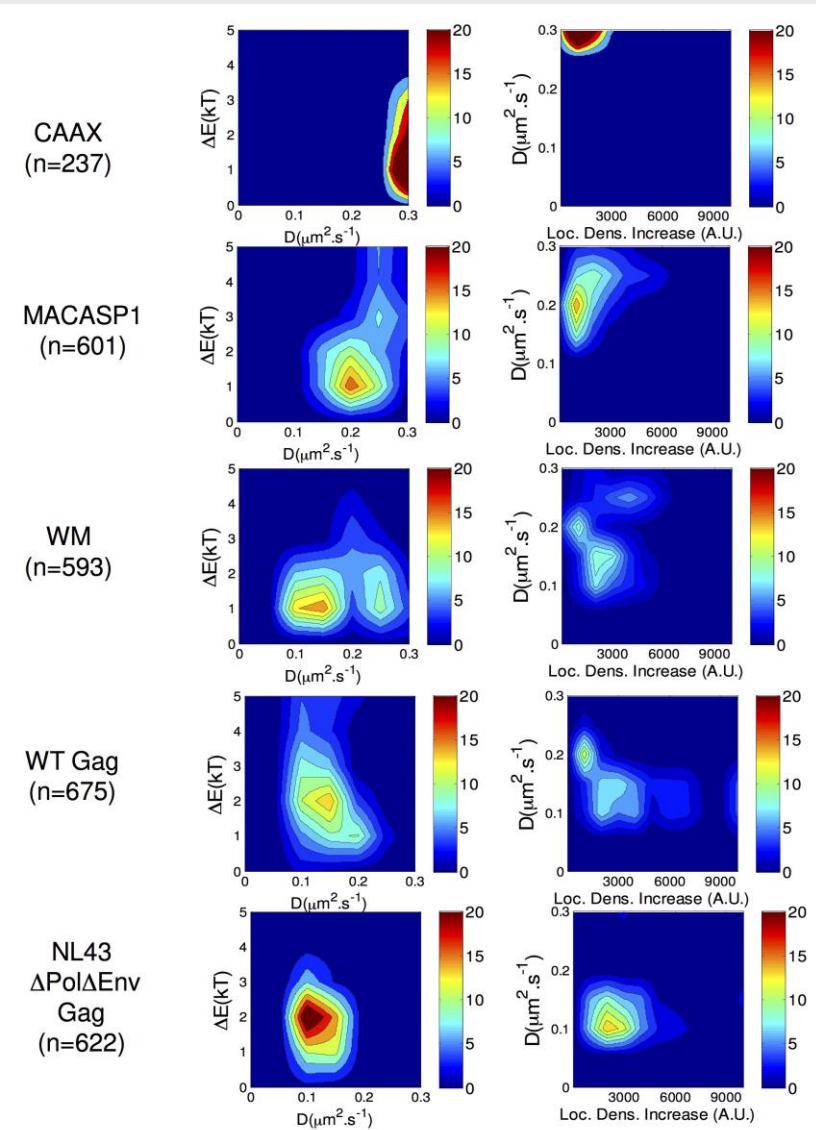
Conclusion: TRamWAy

- Probabilistic approach to single molecule biology
  - Incorporating all Sources of noises
  - Automating biological activity detection
  - Probing the underlying physical process
  - Supervised/unsupervised analysis of random walks
  - Computation adapted to scale

- Releases:
  - Stable release accessible : **pip install --user tramway**
  - Documentation : **https://tramway.readthedocs.io/en/latest/**
  - Developing (no insurance):
    https://github.com/DecBayComp/TRamWAy
    - Non-tracking (MAP): ~Mars-April 2019
    - Non-Tracking (BP): ~Summer 2019
    - Unsupervised all Random Walk: ~Summer 2019
    - Unsupervised along time: ~Autumn 2019
    - Stable Image Deconvolution: ~Winter 2019
    - Jupyter notebooks on github: ~Summer 2019
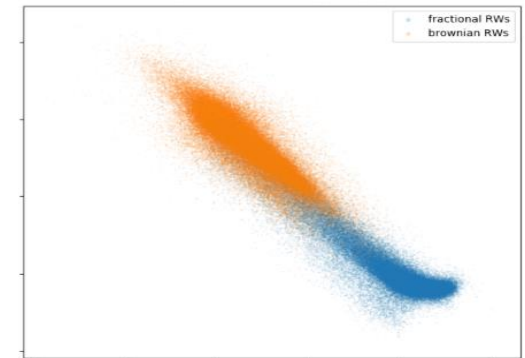
Decision and Bayesian Computation
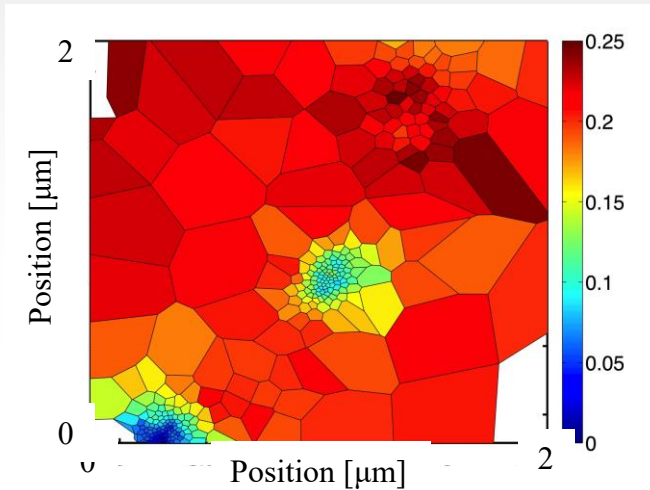

Physicists

Software engineers

AI and Physics 2019

Lab Assistant

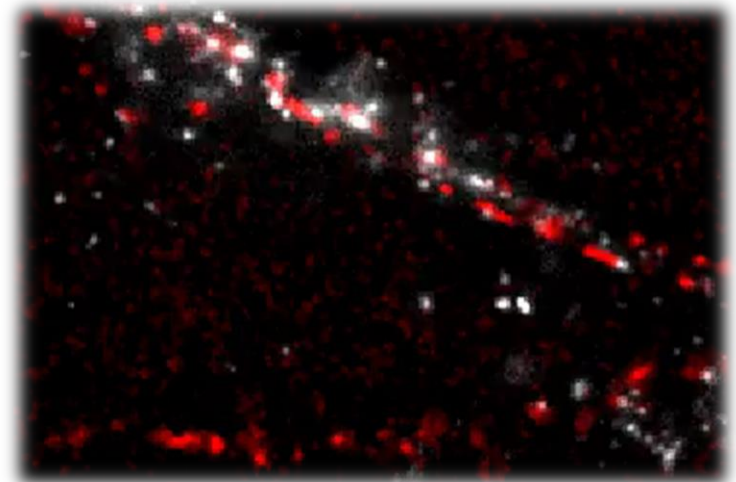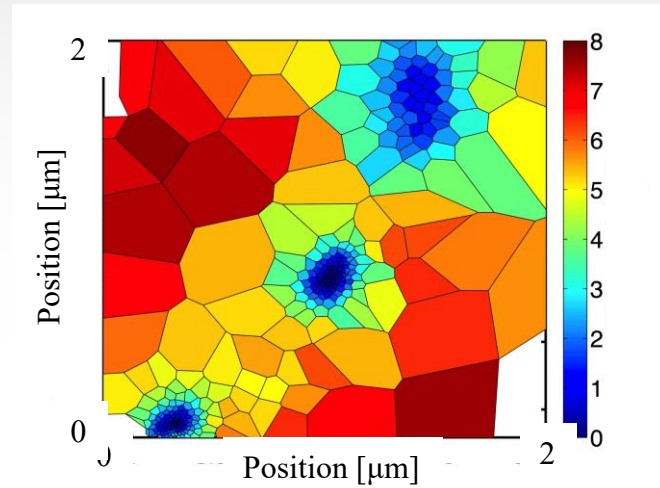# Receptor-Scaffold Interactions

Glycine Receptors (GlyR)

Gephyrin
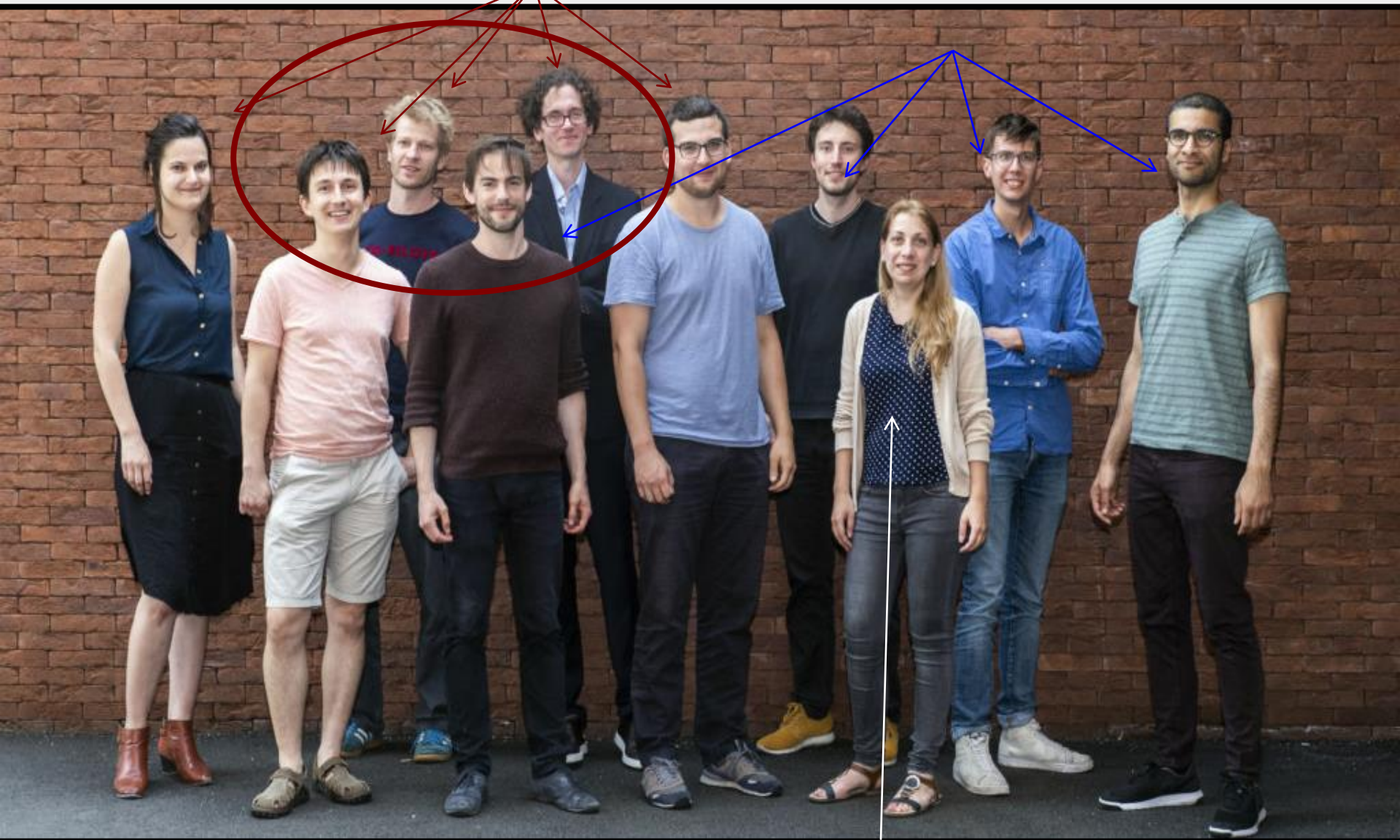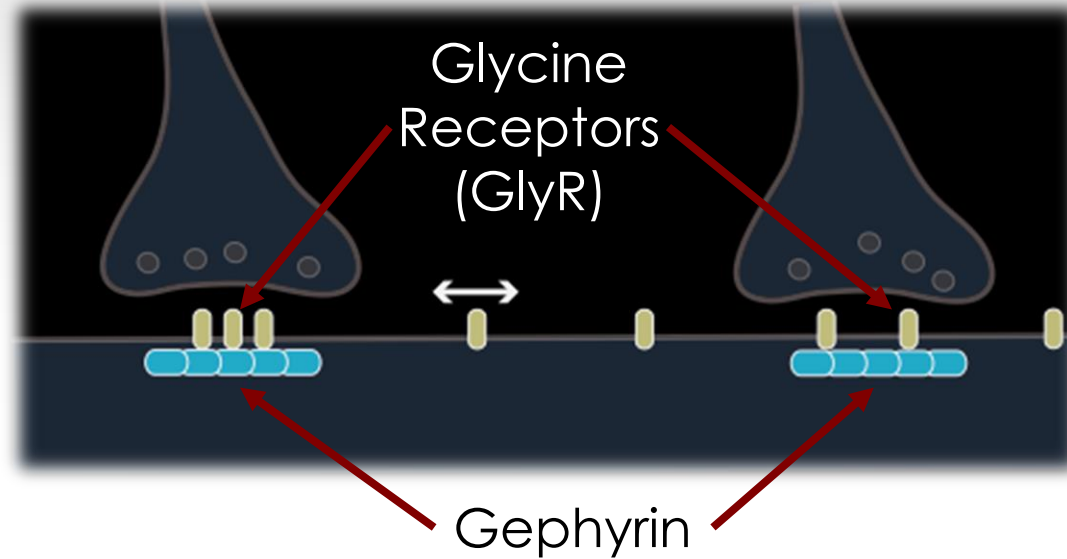
- Bottleneck of Neural Information Processing

- Scale ~100nm (lateral) + Perisynaptic (200-300nm)

- Dynamics

Structures – Dynamics – Functions

Diffusion Map [μm^2.s^{-1}]

Potential Map [k$_B$T]

$$\frac{d\vec{r}}{dt} = D(\vec{r})\left[-\nabla V(\vec{r}) + \frac{\nabla D(\vec{r})}{D(\vec{r})}\right] + \sqrt{2D(\vec{r})}\vec{\xi}(t)$$
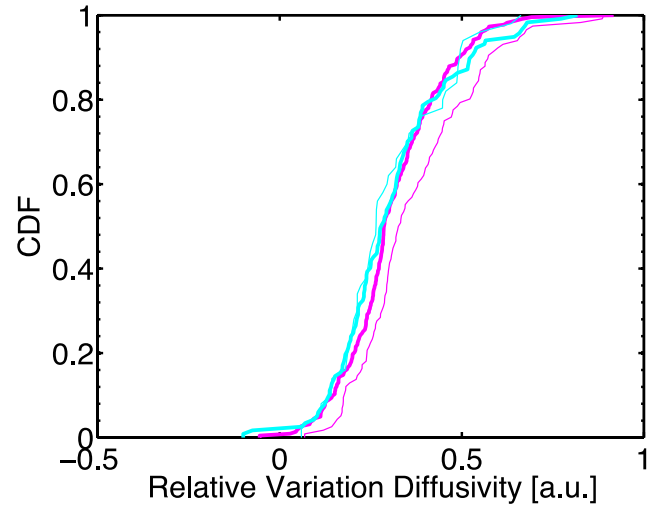
Masson, *et al.*, Biophysical Journal, 2014
M. El Beheiry *et al*, Nat. Methods (2015)

Structures – Dynamics – Functions
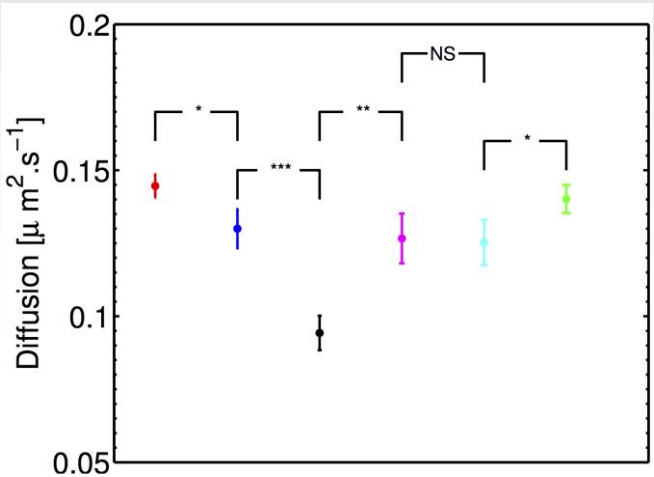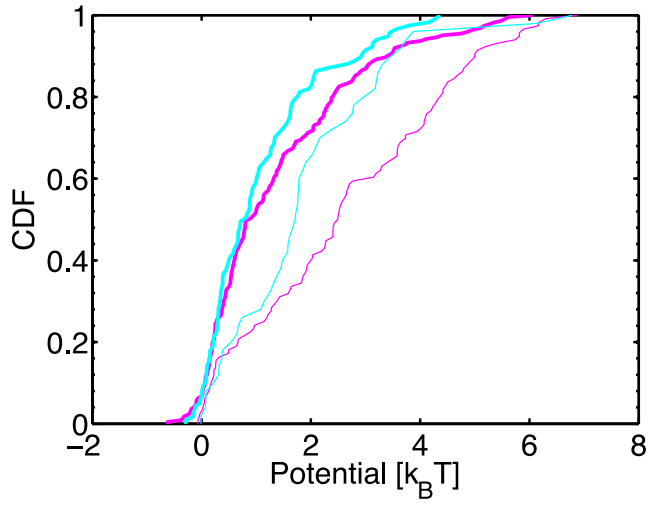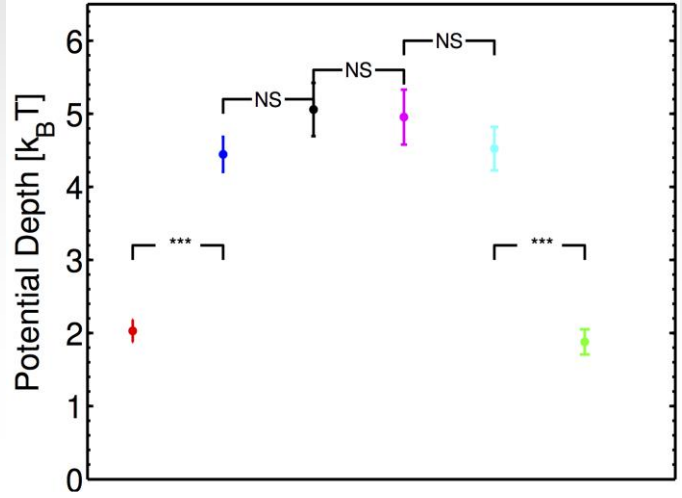
200 synapses
Drug: PMA

Y413D: cyan   Y413F: magenta

Generative Model: Simulation in The Maps

- Fokker Planck $\longrightarrow$ Master Equation

- Masters Equations

$$\frac{dP_{(i,j)}(t)}{dt} = \mathring{a}_{(i',j') \hat{I} \, N(i,j)} W_{(i,j),(i',j')} P_{(i',j')}(t) - \mathring{a}_{(i',j') \hat{I} \, N(i,j)} W_{(i',j'),(i,j)} P_{(i,j)}(t) \qquad a_v = W_{(i,j),(i',j')}$$

$$W_{(i,j),(i',j')} = \frac{D(i',j')}{\Delta x^2} \exp\left(-\frac{\Delta x F^x_{(i,j),(i',j')}}{2\gamma(i',j') D_{(i',j')}}\right) \qquad a_0 = \mathring{a}_v \, a_v$$

- Multi-scale Space Structure

- Time to wait

$$t = \frac{1}{a_0} \log\left(\frac{1}{r_1}\right) \qquad r_1 = [0..1]$$

- Site to chose, k,

$$\mathring{a}_{u=0}^{k-1} a_u \, \pounds \, r_2 a_0 < \mathring{a}_{u=0}^{k} a_u \qquad r_2 = [0..1]$$

Gillespie, D. ,Exact Stochastic Simulation of Coupled Chemical Reactions. J. of Phys. Chem. 81, 2340-2361 (1977)
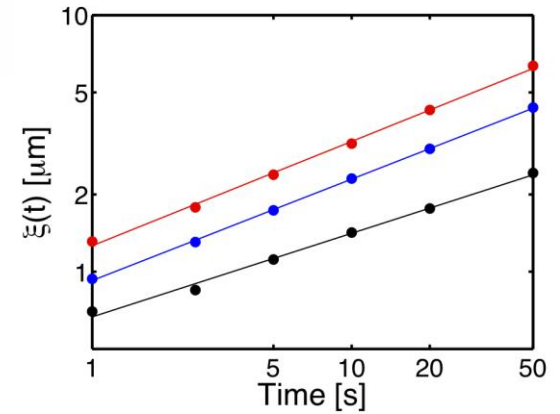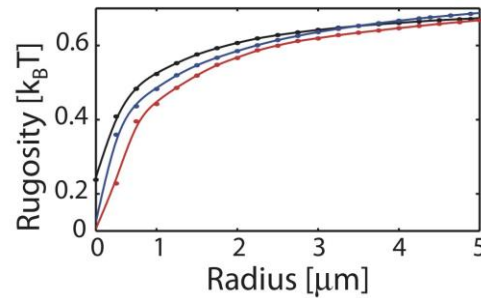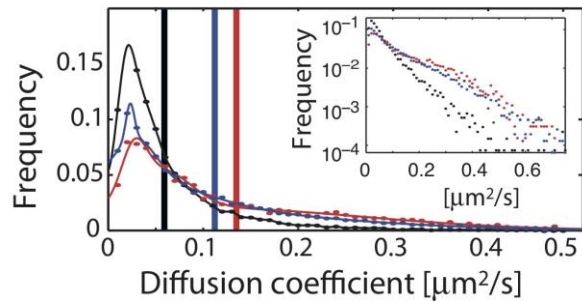
Few Results

In the synapses

$$P(r) = \frac{1}{\rho x^2(t)} e^{-\frac{r^2}{x^2(t)}}$$

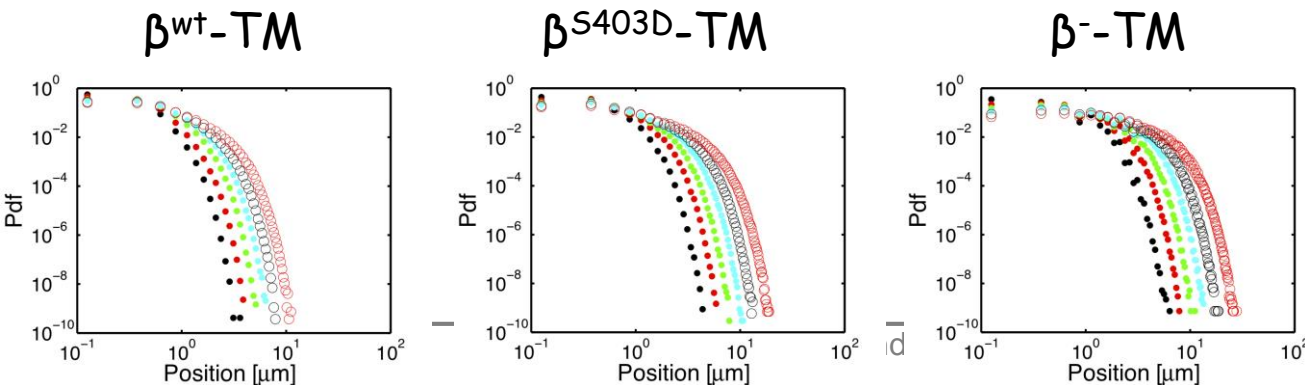$$r = \frac{r}{x(t)}$$

β^{WT}-TM, β^{S403D}-TM, β$^-$-TM

β^{wt}-TM   β^{S403D}-TM   β$^-$-TM

$$x(t) \propto t^g$$

β^{--TM}: γ=0.41
β^{S403D}-TM:
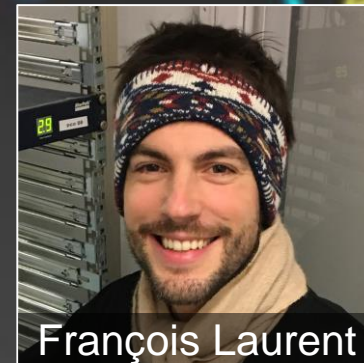γ=0.39
β^{wt}-TM: γ=0.33

20 nm beads in nucleoplasm of U2OS

Hajj, Wisniewski, El Beheiry, Chen, Revyakin, Wu & Dahan, *PNAS,* 2014

InferenceMAP → **TRamWAy (Github)**

François Laurent

M. El Beheiry, M. Dahan & J.-B. Masson,
"InferenceMAP: mapping of single-molecule dynamics with Bayesian inference,"
Nature Methods, 12 (7), 594–595 (2015)

Non-Tracking: BP on all the Graphs

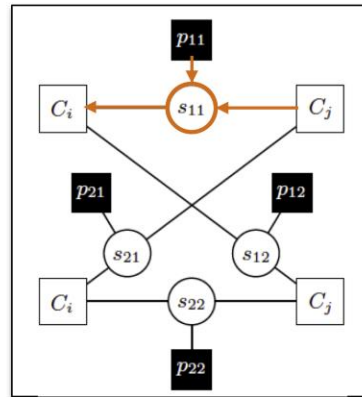## Approximating the likelihood using Belief Propagation

Likelihood as a factor graph:

$$\mathcal{L}(\theta|\mathbf{r}_1, \mathbf{r}_2) \propto \sum_{\mathbf{s}} p(\mathbf{s}, \mathbf{r}_1, \mathbf{r}_2 | \theta)$$

$$p(\mathbf{s}, \mathbf{r}_1, \mathbf{r}_2 | \theta) = \prod_i C_i(\vec{s}_i) \prod_i C_j(\vec{s}_j) \prod_{i,j} p_{ij}^{s_{ij}}$$

$$C_i(\vec{s}_i) = \delta \left(\sum_j s_{ij}, 1 \right)$$

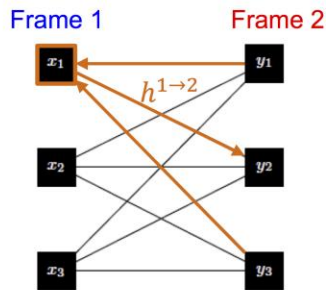$$C_j(\vec{s}_j) = \delta \left(\sum_i s_{ij}, 1 \right)$$

Corresponding factor graph for two particles

Belief propagation sends messages between nodes to update their beliefs.

Iterate until convergence.

Takes $O(N^2)$ operations.

Equivalent to sending messages between images

Frame 1   Frame 2

$h^{1 \to 2}$

Chertkov *et al.*, PNAS (2010)

Messages:

$$h^{i \to j} = -\ln \sum_{j' \neq j} p_{ij'} e^{h^{j' \to i}}$$

$$h^{j \to i} = -\ln \sum_{i' \neq i} p_{i'j} e^{h^{i' \to j}}$$

Parallel implementation

Bethe free energy:

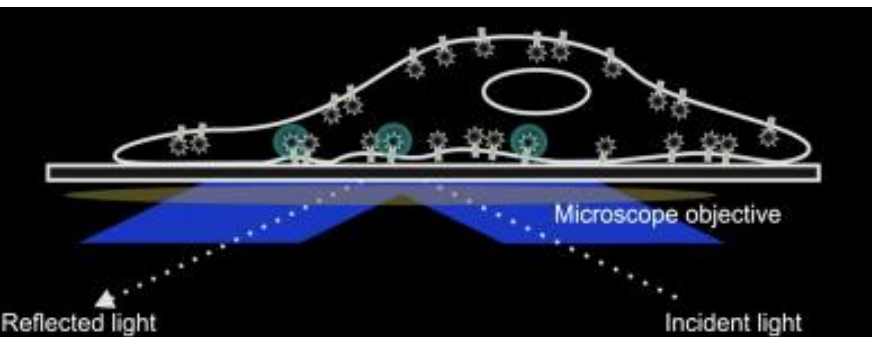$$\mathcal{F}_{\mathrm{B}} = + \sum_{i,j} \ln \left(1 + p_{ij} e^{h^{i \to j} + h^{j \to i}} \right)$$

$$- \sum_i \ln \left(\sum_j p_{ij} e^{h^{j \to i}} \right)$$

$$- \sum_j \ln \left(\sum_i p_{ij} e^{h^{i \to j}} \right).$$

SPT Microscopy Methods

sptPALM
Stochastic activation (photo-activatible proteins)
Short trajectories
Transfected fluorescent proteins
TIRF

S. Manley, *et al.* "High-density mapping of single-molecule trajectories with photoactivated localization microscopy," *Nature Methods* **5** (2), 155–157 (2008)

uPAINT
Stochastic binding (fluorescent ligands)
Long trajectories
Membrane-limited
Nanobody-compatible
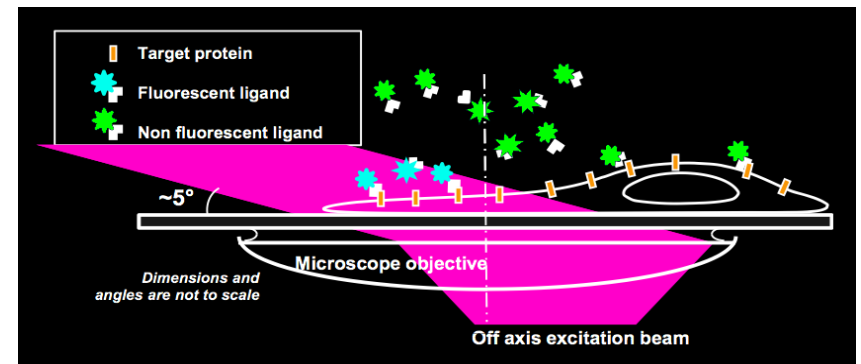
G. Giannone, *et al.* "Dynamic Superresolution Imaging of Endogenous Proteins on Living Cells at Ultra-High Density," *Biophysical Journal* **99**, 1303–1310 (2010).

Multi-Plane Microscopy
3D High-Density Single-Particle Tracking

10 µm

50 FPS

20 nm beads in nucleoplasm of U2OS Cell
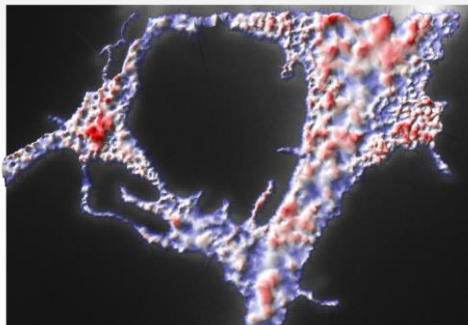
1.8

µm²/s

0.01
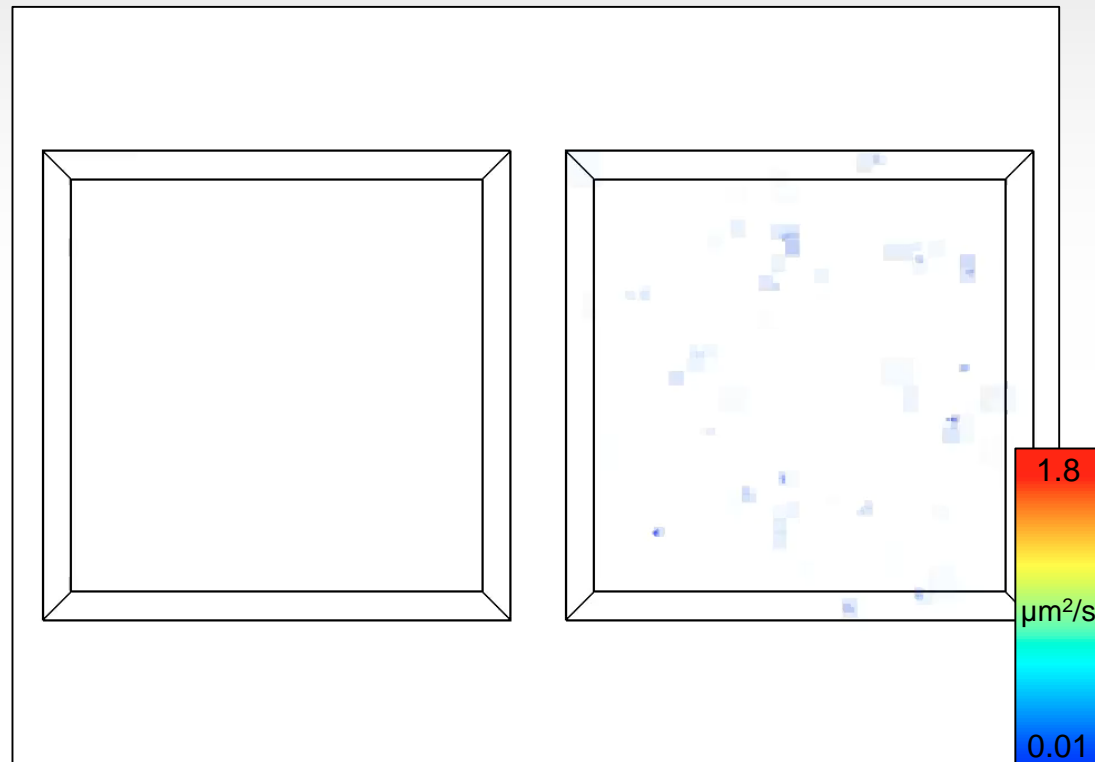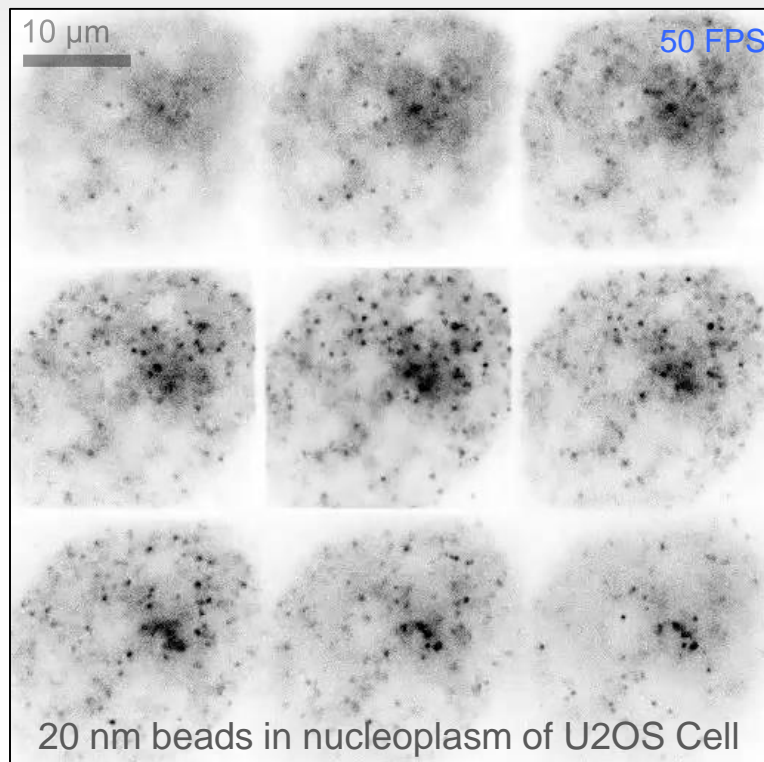
Localisation Density

Probing Depth

Multifocus microscopy achieves a workable balance for 3D parameter mapping

Bassam Hajj
Institut Curie

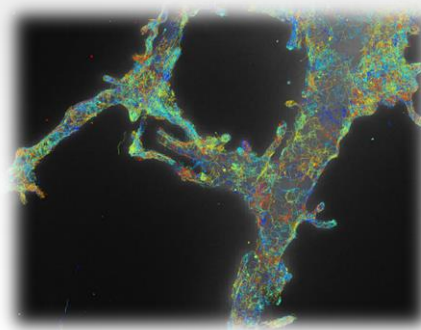# InferenceMAP
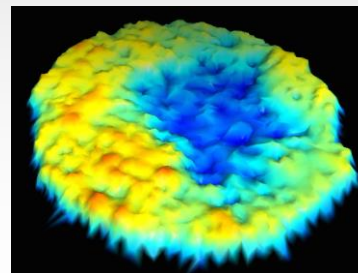
Transmembrane Protein Diffusion in Neurons

P. Dionne, Ecole Normale Supérieure

Rac$_1$ dynamics

A. Remorino, Insitut Curie

Membrane Confinement Zones

S. Türkcan, Ecole Polytechnique

Tailored for Big Data

Adaptive Multi-Scale Meshing

Nuclear Protein Diffusion

J. Liu, Janelia Research Campus

S. P. Knight *et al, Science 350, 6262 p823 (2015)*

M. El Beheiry *et al,* **InferenceMAP**. Nat. Methods 12, 594, (2015)
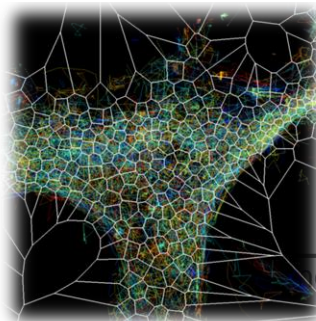
Energy of Virion Formation
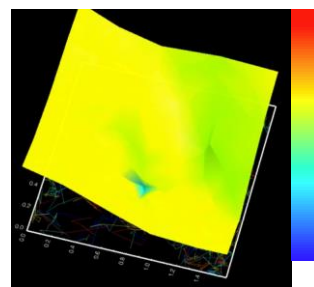
8.0 kT

4.0 kT

0.0 kT

C. Floderer, CNRS CPBS

Calculation Flexibility

Membrane Protein Interaction Energy

M. El Beheiry, Institut Curie

User-Friendly Interface

and Physics 2019
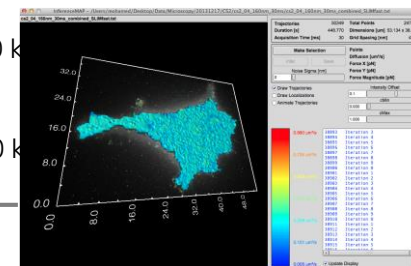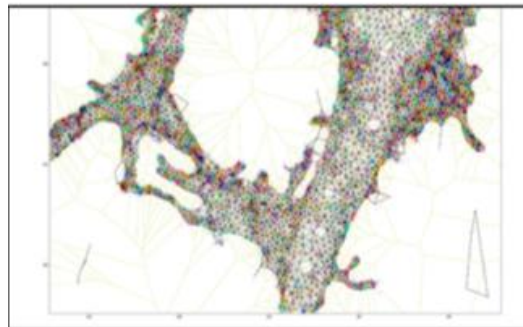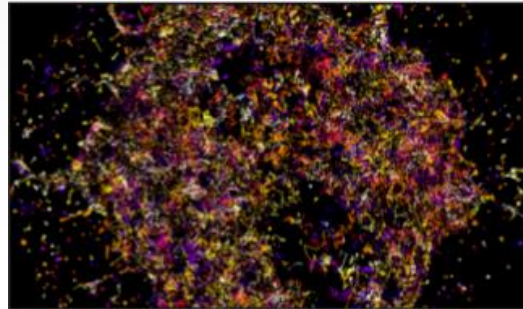
INSERM workshop 250
Intracellular dynamics of molecules:
analysis and models
Hugues Berry, Cyril Favard , JBM

Phase 1 : workshop June 24-26, 2019 – Bordeaux
Phase II : practical July 1-4, 2019 – Lyon

Information and Registration: ateliers@inserm.fr