TRamWAy: mapping physical properties of individual biomolecule random motion in large-scale single-particle tracking experiments

François Laurent, Hippolyte Verdier, Maxime Duval, Alexander Serov, Christian L Vestergaard, Jean-Baptiste Masson

TRamWAy: Mapping physical properties of individual biomolecule random motion in large-scale single-particle tracking experiments

François Laurent [1,2*], Hippolyte Verdier [1,3,4], Maxime Duval [1], Alexander Serov [1], Christian L. Vestergaard [1] and Jean-Baptiste Masson [1,*]

[1] *Decision and Bayesian Computation, Computational Biology Department, Neuroscience Department, CNRS USR 3756, CNRS UMR 3571, Institut Pasteur, Paris 75015, France*

[2] *Institut Pasteur, Université Paris Cité, Bioinformatics and Biostatistics Hub, Paris F-75015, France*

[3] *Histopathology and Bio-Imaging Group, Sanofi R&D, Vitry-Sur-Seine 94400, France*

[4] *Université Paris Cité, UFR de Physique, Paris 75013, France*

[*] To whom correspondence should be addressed.

Motivation: Single-molecule localization microscopy allows studying the dynamics of biomolecules in cells and resolving the biophysical properties of the molecules and their environment underlying cellular function. With the continuously growing amount of data produced by individual experiments, the computational cost of quantifying these properties is increasingly becoming the bottleneck of single-molecule analysis. Mining these data requires an integrated and efficient analysis toolbox.

Results: We introduce TRamWAy, a modular Python library that features: (i) a conservative tracking procedure for localization data, (ii) a range of sampling techniques for meshing the spatio-temporal support of the data, (iii) computationally efficient solvers for inverse models, with the option of plugging in user-defined functions and (iv) a collection of visualisation tools and a simple web-based interface.

Availability and implementation: TRamWAy is a Pythonlibrary and can be installed with pip and conda. The source code is available at *https://github.com/DecBayComp/TRamWAy*.

Contact: *flaurent@pasteur.fr* & *jbmasson@pasteur.fr*

I. INTRODUCTION

Rapid progress in single-molecule localization microscopy (SMLM Betzig *et al.* 2006, Manley *et al.* 2008) has allowed the recording of moving biomolecules in cells with spatio-temporal precision of tens of nanometre-milliseconds. The biomolecules' random walks allow inferring local biophysical and biochemical properties of their cellular environments (Fazel *et al.* 2020, Sgouralis and Pressé 2017). Solving this task is a challenging inverse problem, notably due to spatio-temporal heterogeneities of both the environments and biomolecule populations, as well as the inherent stochasticity of biomolecule motion, experimental noises and short durations of most trajectories. Bayesian inference (Beheiry *et al.* 2016, Bryan *et al.* 2020) provides a principled framework to address this inverse problem. Continuous increase in both the number of trajectories and particle densities has recently paved the way to global cellular mapping (Beheiry *et al.* 2016, Briane *et al.* 2019, Hoze *et al.* 2012, Salomon *et al.* 2020). The massive increase in data throughput induces new computational challenges however. Common experiments can yield millions of recorded points and hundreds of thousands of trajectories (Giannone *et al.* 2012), and models can have hundreds of thousands of parameters (Laurent *et al.* 2019). This imposes significant constraints on algorithmic design and computing infrastructures. In this article, we introduce TRamWAy, The Automatic Random Walk Analyzer, a Python library for Bayesian spatio-temporal analysis of dynamic localization microscopy data. TRamWAy implements the full processing chain from raw SMLM data to biophysical landscapes ready to be visually or programmatically explored. This includes simplified and robust particle tracking, tessellation and segmentation of the spatio-temporal support of the data, and inference of the parameters of the local dynamics such as diffusivity, drift, force and effective interaction energy (Fig. 1(a)).

II. SOFTWARE IMPLEMENTATION AND METHODS

In this section, we introduce a selection of key methods provided in TRamWAy. The full list of features, tutorials and notebooks can be found at *https://tramway-tour.readthedocs.io*

(i) Particle tracking: TRamWAy can take as input both trajectories or time-stamped localizations. For treating localization data, a *non-tracking* graph-based algorithm is featured to extract an optimal pairing between molecules in successive frames. Trajectories can be reconstructed by linking consecutive elementary displacements. The procedure,

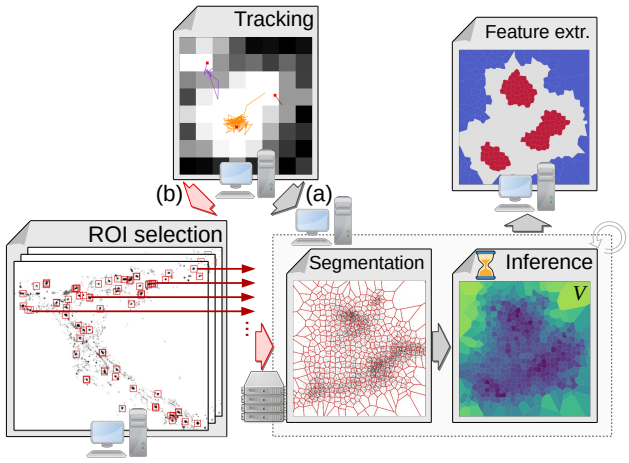

FIG. 1. Local (a) and cloud (b) processing chains for SMLM data: **Tracking:** Graph-based sequential linking between consecutive frames with tracking if needed (see §2i). **Segmentation:** Multi-constrained spatio-temporal meshing of trajectories. **Inference:** Parameter inference is performed and physical parameters are associated to the mesh. **Feature extraction:** Post-processing of the spatio-temporal maps to extract relevant features, e.g. depth of effective confining potential, heterogeneity of diffusion at the cell membrane, etc. (b) **Computational scalability:** The inference is the main computational bottleneck. Most analysis are focused in ROIs. TRamWAy features building blocks to help the user parallelize the analysis and run it automatically on a remote HPC host with the resulting files retrieved back to the local computer.

implemented following Jaqaman *et al.* (2008) (*no gap closing* variant), ensures minimal mislinking and is computationally optimized. **(ii) Spatio-temporal segmentation:** The temporal support of the data can be segmented using a sliding time window or any user-defined series of time segments (Laurent *et al.* 2019). Several modalities for tessellating space are provided, including regular tessellations of square or hexagonal bins, adaptive methods such as the k-means algorithm (Beheiry *et al.* 2016) and a growing-when-required neural gas. **(iii) Model inference:** To infer dynamics, TRamWAy inherits models from InferenceMAP (Beheiry *et al.* 2016) and uses a Bayesian framework based on the overdamped Langevin equation to compute the posterior probability of spatio-temporal parameter maps. TRamWAy allows maximizing this posterior with respect to multiple combinations of physical parameters: diffusion, drift, force and effective interaction energy with recently developed stochastic optimization methods (Laurent *et al.* 2019). Parameter values are inferred onto the mesh. TRamWAy allows regularizing the inferred parameter maps using both physics-informed priors and classic Bayesian priors (Beheiry *et al.* 2016, Laurent *et al.* 2019). **(iv) Post-processing:** Due to the large variety of possible post-analyses, TRamWAy focuses on making the outcome of these analyses robust to various sources of noise and bias. TRamWAy stores all data, inferences and analyses in HDF5-based files to be manipulated in Python with the RWA-python library, or in text files. Inferred parameter landscapes can be rendered in many ways, in Python or using a minimalist web interface. The graphical output options include movies for displaying dynamic landscapes. Furthermore TRamWAy features a Bayesian evidence test for force/drift detection in heterogeneous diffusion landscapes (Serov *et al.* 2020).

TRamWAy matches the current demand for automated pipelines and high flexibility to programmatically handle multiple and large datasets. Furthermore, observed molecules are often scattered into distinct spatial domains, which may be studied independently and in parallel as regions of interest (ROIs). While exploratory data analysis in single ROIs can be done with the InferenceMAP software (Beheiry *et al.* 2016) in 2D and in 3D with Genuage (Blanc *et al.* 2020), TRamWAy features sets of building blocks to automate parallel processing of large-scale data with large amounts of ROIs (Fig. 1(b)). The workflow is specified in a declarative way, *i.e.* in arbitrary order. TRamWAy can parallelize it on the local host and onto a remote host such as a high-performance computing (HPC) cluster. On remote hosts, TRamWAy automatically dispatches user scripts after minimal parsing for execution as job arrays in a container, which is also automatically fetched if missing. Input data are not transferred; only outputs are retrieved once the jobs are complete. User-defined functions can be added to the inference.

III. SUMMARY

The TRamWAy library allows spatio-temporal mapping of biomolecules physical properties in the cell. Examples of application to Virion assembly in HIV can be found in Laurent *et al.* (2019) and Floderer *et al.* (2018). The library provides tools to post-process and analyse the biophyscal maps in order to link them to biological results. TRamWAy is modular and can be rapidly repurposed and is associated to numerous notebooks to ease future users'

developments.

FUNDING

This study was funded by the French *Agence Nationale de la Recherche*, grants ANR-17-CE23-0016, PIA/ANR-16-CONV-0005 & ANR-19-P3IA-0001.

DATA AVAILABILITY

Manual and tutorials are available at *https://tramway-tour.readthedocs.io*.

Beheiry, M. E. et al. (2016). A primer on the bayesian approach to high-density single-molecule trajectories analysis. *Biophys J*, **110**(6), 1209–1215.

Betzig, E. et al. (2006). Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science*, **313**(5793), 1642–1645.

Blanc, T. et al. (2020). Genuage: visualize and analyze multidimensional single-molecule point cloud data in virtual reality. *Nat Methods*, **17**(11), 1100–1102.

Briane, V. et al. (2019). An overview of diffusion models for intracellular dynamics analysis. *Briefings in Bioinformatics*, **21**(4), 1136–1150.

Bryan, J. S. et al. (2020). Inferring effective forces for langevin dynamics using gaussian processes. *The Journal of Chemical Physics*, **152**(12), 124106.

Fazel, M. et al. (2020). Bayesian grouping of localizations, sub-nanometer precision, counting and resolution doubling. *Biophysical Journal*, **118**(3), 313a.

Floderer, C. et al. (2018). Single Molecule Localisation Microscopy Reveals How HIV-1 Gag Proteins Sense Membrane Virus Assembly Sites in Living Host CD4 T Cells. *Sci. Rep.*, **8**(1), 16283.

Giannone, G. et al. (2012). High-content super-resolution imaging of live cell by uPAINT. In A. Sousa et al., editors, *Nanoimaging*, volume 950, pages 95–110. Humana Press, Totowa, NJ.

Hoze, N. et al. (2012). Heterogeneity of AMPA receptor trafficking and molecular interactions revealed by superresolution analysis of live cell imaging. *Proc Natl Acad Sci*, **109**(42), 17052–17057.

Jaqaman, K. et al. (2008). Robust single-particle tracking in live-cell time-lapse sequences. *Nature Methods*, **5**(8), 695–702.

Laurent, F. et al. (2019). Mapping spatio-temporal dynamics of single biomolecules in living cells. *Physical Biology*, **17**(1), 015003.

Manley, S. et al. (2008). High-Density Mapping of Single-Molecule Trajectories with Photoactivated Localization Microscopy. *Nat Methods*, **5**(2), 155–157.

Salomon, A. et al. (2020). Dense mapping of intracellular diffusion and drift from single-particle tracking data analysis. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE.

Serov, A. S. et al. (2020). Statistical Tests for Force Inference in Heterogeneous Environments. *Scientific Reports*, **10**(1).

Sgouralis, I. et al. (2017). An Introduction to Infinite HMMs for Single-Molecule Data Analysis. *Biophysical Journal*, **112**(10), 2021–2029.