THÈSE DE DOCTORAT DE

L'UNIVERSITÉ DE RENNES 1

École Doctorale N° 601
Mathématiques et Sciences et Technologies
de l'Information et de la Communication
Spécialité : *Signal, Image, vision*

Par

Yunjiao LU

« Dynamiques intracellulaires et imagerie de super résolution : la paroi bacterienne sondée à l'echelle moléculaire »

Thèse présentée et soutenue EN VISIOCONFÉRENCE, le 7 Juin 2021
Unité de recherche : MaIAGE, Inra Jouy-en-Josas
Serpico, Inria Rennes – Bretagne Atlantique

Rapporteurs avant soutenance :

Xavier Descombes, Directeur de Recherche Inria, Inria Sophia Antipolis
Jean-Baptiste Masson, Chargé de Recherche, Institut Pasteur

Composition du jury :

Président :	Guy Carrault, Professeur, LTSI et Université de Rennes 1
Examinateurs :	Leila Muresan, Principal Investigator, Cambridge Advanced Imaging Center, Cambridge
Dir. de thèse :	Charles Kervrann, DR INRIA, Centre INRIA Rennes-Bretagne Atlantique
Co-dir. de thèse :	Rut Carballido-Lopez, DR INRAE, Unité Micalis Centre de Recherche de Jouy-en-Josas
Membre invité :	Alain Trubuil, IR INRAE, Unité MaIAGE, Centre de Recherche de Jouy-en-Josas

ACKNOWLEDGEMENT

Three years and eight months, I am thankful for this precious time that allows me to discover, experience, fail, think, learn, and better understand myself and the world of scientific research. There are many people to whom I owe great gratitude.

I want to thank my supervisors, Alain Trubuil and Charles Kervrann, for the favorable environment they offered me for my personal and professional development to become a responsible scientific researcher. I am thankful for the freedom that I owned to explore and for them being my support and guiding me with their professional expertise and rich experiences. My thankfulness goes to my supervisor Rut Carballido-Lopez and colleagues Cyrille Billaudeau and Arnaud Chastanet, who introduced me to their research field, microbiology. The collaboration and multiple discussions with them throughout the preparation of my manuscript make our work much more rich and valuable. I would like to thank my colleagues at MaIAGE, including my former collegue Pierre Hodara, as well as our administrators and IT support, for their help, friendship, good humor, and encouragement while I was writing the manuscript, and my colleagues at Serpico and the Institut Curie. I would like to express my gratitude to Vincent Fromion and Hugues Berry, members of the CSI, for their attention to the progress of the thesis, their precious suggestions. I also want to thank the members of the jury for their valuable time in reading my manuscript.

I am grateful for my beloved brothers and sisters in the Church of Christ in Paris, for their support at some impossible moments and for their living faith, which I share, which is the true rock of my life and gives meaning to all that I experienced. I am also grateful for the interesting encounters at the residence MIAA, in the beautiful CIUP, which became almost my whole living space (work and live) during the last year of my PhD, due to the Covid-19 sanitary crisis. Thank you Zejian for your help with several physical problems ; thank you Giacomo for being there. Thank you for your passions and talents, which brightened my life during that difficult time. I wish you all the best. Thanks to my dear parents, who waited for my return with long endurance far away in China.

TABLE OF CONTENTS

LIST OF FIGURES

RÉSUMÉ EN FRANÇAIS

Contexte et Motivations

Contexte scientifique et motivations

Au cours des deux dernières décennies, l'imagerie biologique a connu une révolution dans le développement de nouvelles techniques de microscopie qui permettent de visualiser les tissus, les cellules, les protéines et les structures macromoléculaires à tous les niveaux de résolution, ainsi que leurs états physiologiques, leurs compositions chimiques et leurs dynamiques. Grâce aux récents progrès réalisés en optique, en biologie moléculaire et en chimie, des capteurs numériques et les sondes de marquage (e.g. sondes GFP (Green Fluorescence Protein)), permettant désormais visualiser les composants et les organites infracellulaires à l'échelle de quelques dizaines de nanomètres. En conséquence, la microscopie fluorescente et l'imagerie multimodale (acquisition avec plusieurs sondes et longueurs d'onde variables) sont devenues le fer de lance de la biologie moderne. Toutes ces avancées technologiques ont fait émergé de nouveaux problèmes et des défis pour les chercheurs spécialistes du traitement et de l'analyse des images. Désormais les questions relatives au numérique sont davantage prises en compte lors de la définition des preuves de concept en bioimagerie.. Un brillant exemple de ce changement de paradigme est la microscopie de super-resolution basé sur la détection de molecules uniques (Single-molecule localization microscopy, PALM, STED), qui a reçu le prix Nobel de chimie en 2014.

Les défis en traitement des images biologiques et en microscopie quantitative

Dans la plupart des cas, la microscopie moderne en biologie est relativement complexe et multidimensionelle : deux ou trois dimensions spatiales, de la macro à la nano-échelle, et une dimension temporelle, parfois définie spectralement et correspondant souvent à une espèce biomoléculaire particulière. La microscopie dynamique est également caractérisée par la nature des objets (cellules, organites, molécules individuelles, etc.), par le grand nombre d'éléments petits et mobiles (chromosomes, vésicules, etc.), par la complexité des processus dynamiques qu'elle permet d'observer. Par ailleurs, le corpus de données à considérer pour toute analyse impliquant l'acquisition de séries d'images multiples est massif (jusqu'à quelques Gigaoctets par heure). Il devient donc nécessaire de faciliter et de rationaliser la production de ces données multidimensionnelles, d'améliorer l'analyse post-acquisition, et de favoriser l'organisation et l'interprétation des informations extraites de ce corpus de données. Ce constat a motivé le développement de méthodes et de concepts innovants pour la fusion de données, le recalage des images, la super-résolution, la fouille de données ("data mining"), etc. Plus important encore, la microscopie moderne a permis de réaliser des percées récentes, en mettant en évidence les corrélations et les interactions entre des molécules dans la cellule. Un objectif de recherche à long terme consiste désormais à établir les liens qui permettent d'associer les dynamiques des macromolécules aux fonctions cellulaires.

Afin de relever les défis associés, il est nécessaire de développer des approches et des paradigmes innovants pour reconstruire les images, suivre des biomolécules, estimer le mouvement des molécules en 3D et les paramètres biophysiques sous-jacents, tout en faisant face aux énormes volumes de données générés par les dispositifs de microscopie, toujours de plus en plus performants. Pour répondre à ces exigences, les mathématiques appliquées, le traitement et l'analyse des images et l'informatique ont un rôle essentiel à jouer pour accompagner et soutenir les recherches menées en biophysique et en biologie.

Contexte biologique

Identifier les mécanismes qui permettent de déterminer la forme des cellules est un enjeu essentiel en biologie cellulaire. Chez les bactéries, il est établi que la paroi cellulaire extra-cytoplasmique joue un rôle déterminant dans le contrôle de la forme des cellules. La paroi cellulaire est principalement composée d'un réseau de peptidoglycanes, qui n'est pas statique mais qui se développe en permanence en fonction de la croissance cellulaire. La composition de la paroi cellulaire et les voies impliquées dans la synthèse des composants moléculaires sont aujourd'hui bien établies. Néanmoins, son organisation tridimensionnelle et les mécanismes qui contrôlent son assemblage sont encore méconnus.

D'une manière générale, il est très important de bien comprendre comment une cellule bactérienne grandit, i.e., comment son enveloppe rigide externe est synthétisée, et comment certains processus cellulaires ont lieu à travers cette enveloppe (e.g., infection virale, sécrétion, compétence génétique). D'une manière plus spécifique, plus de la moitié des antibiotiques découverts (par exemple les beta-lactamines comme les pénicillines ou l'amoxicilline) ciblent directement ou indirectement les processus de synthèse de la paroi bactérienne. Comprendre comment la machinerie de synthèse de la paroi cellulaire est contrôlée spatio-temporellement ou encore comment la synthèse de la paroi est régulée pour permettre différents taux de croissance et tailles, permettra de mieux comprendre les mécanismes d'action des antimicrobiens, puis d'établir de nouvelles pistes thérapeutiques.

Au cours des dernières décennies, les acteurs moléculaires impliqués dans l'élongation de la paroi cellulaire ont été identifiés et analysés (Fig. 1). Il est désormais bien établi que la protéine MreB, avec sa structure filamenteuse, assure la médiation des activités Peptidoglycan

Elongation Machinery (PGEM) avec une dynamique permanente. Les trajectoires des agrégats de MreB ont récemment été analysées à partir d'images acquises en microscopie de fluorescence à haute résolution et des algorithmes de suivi de particules uniques. Ces analyses permettent de mieux répondre aux nombreuses questions posées par les microbiologistes : quels sont les différents rôles/fonctions des patchs de MreB au cours des différents régimes de diffusion ; quel est l'élément principal responsable de la dynamique circonférentielle dirigée du PGEM induit par MreB, plus spécifiquement le "rod complex" ; enfin, comment MreB organise les différentes enzymes afin qu'elles collaborent à l'expansion du PG.

Modèles de diffusion pour l'analyse de dynamiques en bioimagerie

Dans l'environnement subcellulaire, on observe beaucoup d'activités qui relèvent du transport de biomolécules et des échanges entre les différentes organelles. La diffusion moléculaire qui induit des fonctions cellulaires, a toujours fasciné les biophysiciens, les mathématiciens, et les statisticiens [BN13]. La diffusion est un processus très général qui peut concerner les protéines, les récepteurs dans les cellules neurones, et les vésicules dans les activités d'endo/exocytose. On observe néanmoins une grande hétérogénéité liée à des changements de régime de mouvement et modes de diffusion. Bien souvent le mouvement de référence est la diffusion libre, expliquée pour la première fois par A. Einstein et P. Langevin selon deux point de vue différents. Par la suite, d'autres processus de diffusion, principalement caractérisés par l'équation différentielle stochastique (SDE), ont été proposés pour décrire des mouvements moléculaires et des comportements dynamiques intracellulaires plus complexes.

Un important effort de recherche a été conduit pour caractériser précisément la mobilité des molécules ([Ver+21], [Ser+20]). On pense qu'il serait même possible de reconstruire les surfaces de certaines structures biologiques à partir des trajectoires, voir de démêler les processus physiques sous-jacents à ces trajectoires.

Entre-temps, la modélisation d'un ensemble de trajectoires statistique à partir de grandes quantités de données ont déjà permis de caractériser les propriétés physiques locales des biomolécules, telles que les modules et les orientations des vecteurs vitesse, le transport actif, les forces d'attraction, reflétant les interactions des particules avec leur environnement ([Lau+19], [HH15]).

Notons que les méthodes d'estimation sont généralement évaluées sur des données synthétiques, avant d'être testées sur des données expérimentales, bien souvent imparfaites. Dans ce contexte, un bruit de mesure est induit lors de l'acquisition des données et lors de la construction des trajectoires. La durée de vie du fluorophore est limitée et contraint la résolution temporelle de l'acquisition et la durée d'observation. Plusieurs travaux de recherche ont donc été menés pour caractériser l'impact de ces sources d'erreur sur l'estimation des paramètres et des processus étudiés.

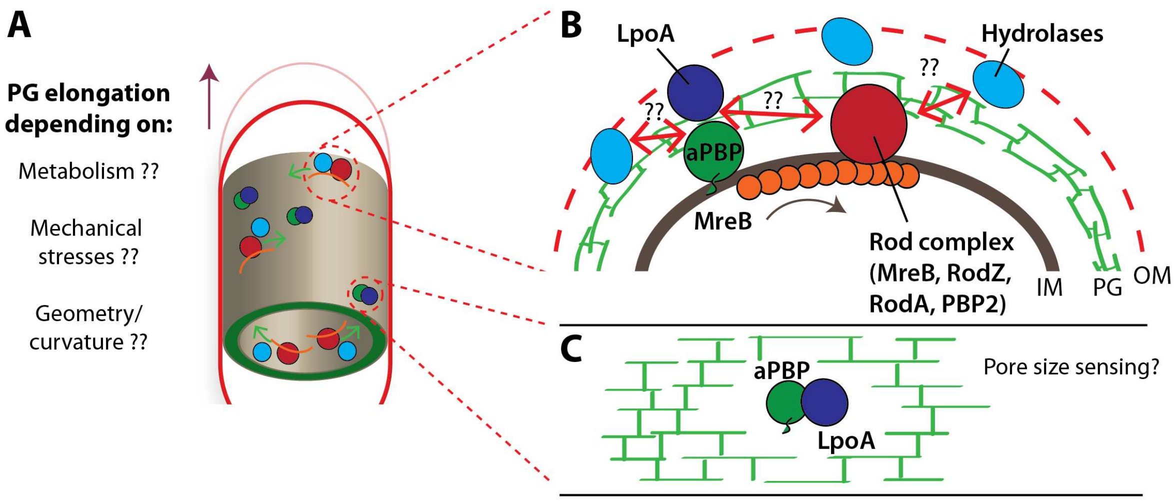


FIGURE 1 – **Représentation des mécanismes dominants de l'allongement de la paroi cellulaire chez les bactéries en forme de bâtonnet (Figure en provenance de [TR18])** (A) Illustration des mécanismes impliqués dans la croissance cellulaire, c'est-à-dire l'élongation du peptidoglycane (PG), dans une cellule en forme de bâtonnet : le "rod complex", les hydrolases et les protéines de liaison à la pénicilline de classe A (aPBP). Malgré des efforts importants pour comprendre l'allongement des bactéries, de nombreuses questions restent en suspens sur la manière dont l'assemblage des PG et la croissance cellulaire s'adaptent à différents facteurs, tels que le métabolisme, les contraintes mécaniques et la forme des cellules. (B) Encart illustrant les principaux mécanismes de synthèse de la paroi cellulaire qui fonctionnent de manière indépendante (voir, [Cho+16]) : (i) le "rod complexe" en mouvement contenant la transglycosylase RodA, la transpeptidase PBP2, et des filaments MreB orientés circonférentiellement le long de la membrane interne ; (ii) aPBPs bifonctionnels. Les hydrolases peuvent être activement engagées dans le clivage de la paroi cellulaire pendant la synthèse de la paroi cellulaire. Il reste à découvrir comment les hydrolases interagissent avec le "rod complexe" et les aPBPs, et comment le "rod complexe" interagit avec les aPBPs. (C) (dans les bactéries Gram-) les aPBPs sont activées par les lipoprotéines LpoA et LpoB de la membrane externe. Cette interaction est suggérée pour fournir un mécanisme permettant de détecter la taille des pores dans la paroi cellulaire afin de guider l'insertion des PG.

Contributions de cette thèse

Dans cette thèse, nous avons rédigé un état des lieux sur les modélisations physique, statistique et mathématique pour l'analyse des trajectoires de biomolécules. Cette présentation a été accompagné d'une évaluation expérimentale sur un problème spécifique, à savoir l'analyse de la dynamique des protéines MreB. Les données expérimentales relatives à cette dynamique, acquises en microscopie TIRF (Total Internal Reflection Fluorescent) présentent certaines limites et sont entachées d'erreur qui peuvent être classées selon trois groupes : les erreurs de mesure induites par une faible résolution spatiale et temporelle, la durée du film (limitées par la durée de vie du fluorophore) et l'observation partielle induite par le champ évanescent du dispositif TIRFM (Fig. 5). Dans le problème spécifique qui nous intéresse (MreB), nous avons également considérer les caractéristiques particulières de la dynamique de MreB, par exemple, le type de diffusion, la densité des particules, la forme des protéines MreB (i.e. sous la forme d'un filament de longeur $\sim 170nm$). A cet égard, la microscopie SIM (Structured Illumination Microscopy) apporte une amélioration sensible de la perception de la dynamique de MreB, et d'observer un filament MreB unique).

Dans ce contexte, nous avons considéré deux cadres de modélisation et d'estimation pour analyser les trajectoires de MreB basées sur des équations différentielles stochastiques. La première approche vise à analyser les trajectoires individuelles, c'est-à-dire à déterminer le mode de diffusion le plus pertinent en distinguant les trois régimes suivants : diffusion libre, super-diffusion ou sous-diffusion, et estimer les paramètres sous-jacents (i.e. dérive et coefficient de diffusion). Nous avons ainsi évalué la performance de plusieurs méthodes de classification de trajectoires sur des données MreB. Nous avons considéré ensuite deux estimateurs du coefficient de diffusion dans le cas de la diffusion libre. L'estimation du coefficient de diffusion a été affinée en incluant les erreurs de mesure. Nous avons par ailleurs étudié les estimateurs "optimaux" liés à la limite inférieure de Cramer Rao (CRLB) et nous avons établi les conditions à remplir pour estimer de manière fiable le coefficient de diffusion en fonction de la qualité des images (rapport signal-sur-bruit). La seconde approche vise à étudier la dynamique des biomolécules dans des voisinages locaux afin de construire des cartes de diffusion et de dérive. En supposant que la dynamiques est définie par un modèle de dynamique spécifique, et étant donné des trajectoires reposant sur une surface 3D, et observés sur un plan de projection 2D, nous avons établi la relation entre la dérive locale et le coefficient de diffusion local en fonction de la courbure locale de la surface. Ce résultat permet ensuite d'inférer la surface 3D et les dynamiques 3D.

Dans la partie centrale de la thèse, afin de traiter le problème de la vue incomplète relative aux acquisitions en microscopie TIRFM-2D, nous avons proposé un cadre mathématique pour aborder le problème de la reconstruction des trajectoires des biomolécules qui se déplacent sur

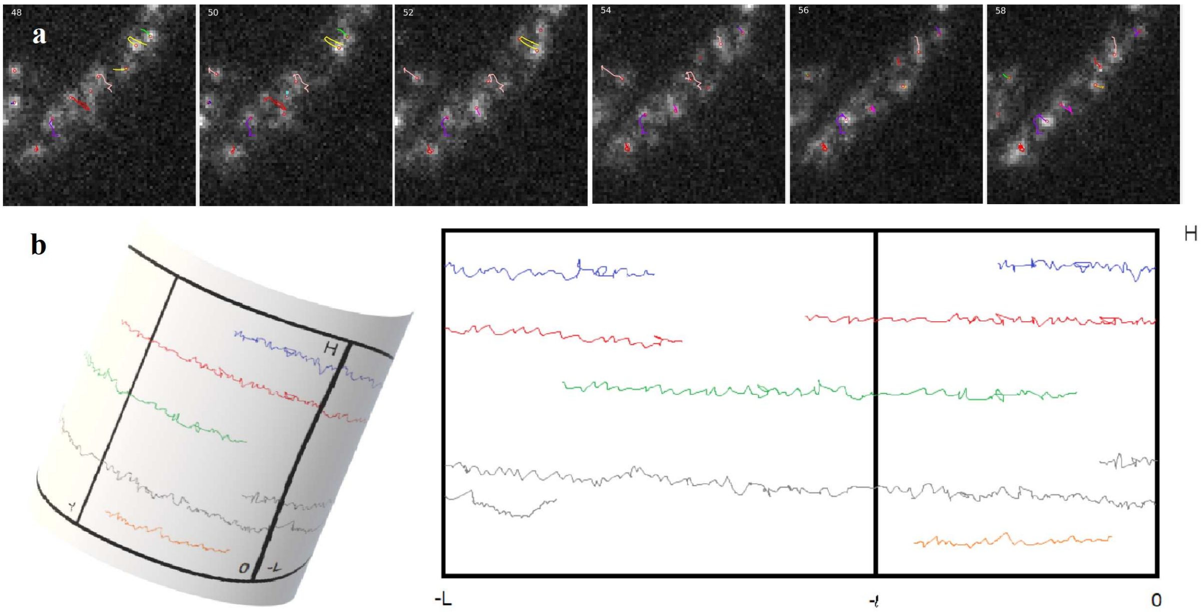


FIGURE 2 – (a) : Plusieurs images consécutives d'une vidéo réelle TIRFM [ABT84]. Les traces sont superposées sur les images. (b) Gauche : Illustration des trajectoires observées pendant le temps enregistré $[0, T_S]$ sur la surface d'un cylindre. Seuls les mouvements à l'intérieur du regions of observation (ROO) $]-l, 0[\times[0, H]$ peuvent être observés, même si la dynamique se produit sur toute la surface. Droite : L'objectif est de récupérer la dynamique sur toute la surface à partir des observations partielles, en coordonnant les entrées par $\{-l\} \times [0, H]$ et les sorties par $\{0\} \times [0, H]$ dans un film pendant T_S, en prenant en compte les événements de "naissance" et de "mort" des particules.

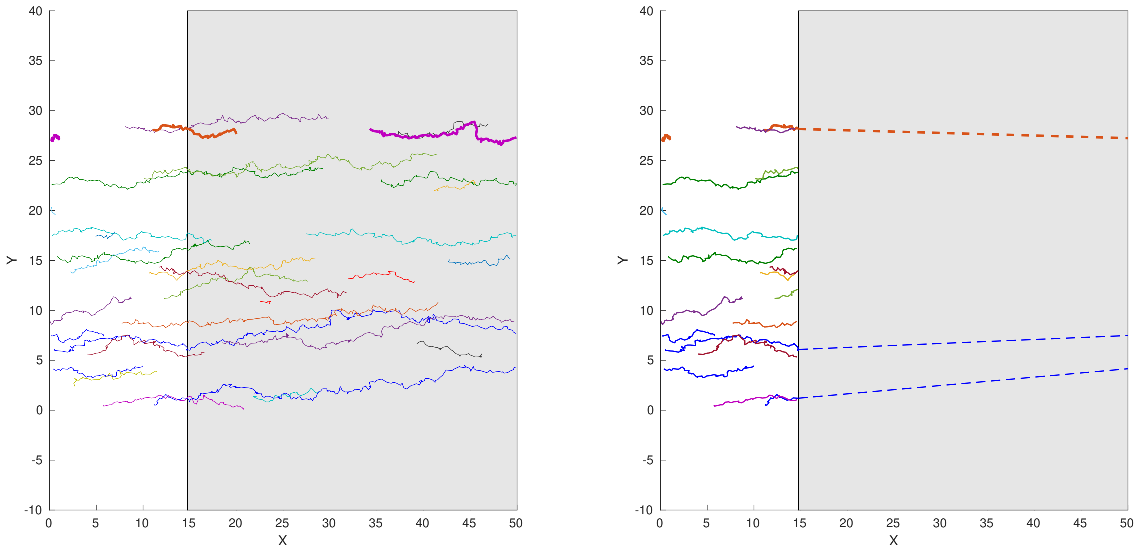


FIGURE 3 – Illustration des résultats de la reconstruction sur la surface des cellules approchée par un cylindre. La direction Y représente l'axe du cylindre et la direction X le côté cylindrique déroulé. La partie ombrée représente la région cachée et la partie claire correspond à la région observée par Total Internal Reflection Fluorescent (TIRF) microscopie. A gauche : Trajectoires simulées dans un film. A droite : Le résultat de la connexion.

des surfaces cylindriques 3D, à partir de bouts de trajectoire ("tracklets") observés dans des fenêtres 2D (voir la région d'observation (ROO)) échantillonnée sur la surface, Fig. 2). Dans cette étude, les particules sont supposées suivre un mouvement brownien avec dérive, et peuvent apparaître ou disparaître pendant la période d'observation. Les événements de "naissance" et de "mort" sont décrits par des processus de Poisson. Les événements de fission ou de fusion des agrégats ne sont pas pris en compte dans cette modélisation. Le problème de reconstruction des trajectoires est défini comme la maximisation de la fonction de vraisemblance étant donné les "tracklets" à l'intérieur de la ROO. Le problème d'optimisation revient à résoudre un problème de programmation linéaire entier. L'algorithme final est un algorithme piloté par les données, sans paramètre caché. Nous avons évalué la performance et la robustesse de notre méthode sur des données de simulation, en faisant varier le rapport entre la région observée et la région non observée, la dérive et la variance des particules simulées, ainsi que les taux de naissance et de mort des particules. Une illustration de la procédure de reconstruction est présentée sur la figure 3. La procédure a également été testée sur des données expérimentales de MreB. Cette approche statistique et computationnelle est la principale contribution de la thèse.

Organisation du manuscrit

Le document est organisé en quatre chapitres comme suit :

Chapitre 1

Dans le Chapitre 1, nous décrivons les composants de la paroi cellulaire, la modélisation de la structure 3D la plus populaire et l'expansion de la paroi pendant la croissance cellulaire, pour les bactéries en forme de bâtonnet. Au cours des dernières décennies, les acteurs moléculaires impliqués dans la synthèse du peptidoglycane ont été identifiés et analysés individuellement. Nous présentons ici les enzymes qui sont impliquées dans la machine d'élongation des peptidoglycanes (PGEM). Il a également été établi que la protéine MreB, avec sa structure filamenteuse, assure continuellement la médiation des activités de PGEM. Dans ce chapitre, nous décrivons brièvement l'orientation et la dynamique du filament MreB, et nous discutons plusieurs hypothèses relatives à la fonction de MreB. Nous évoquons plusieurs travaux qui traitent des interactions dynamiques de MreB avec d'autres enzymes. Ces travaux récents ouvrent des perspectives nouvelles pour comprendre les mécanismes d'organisation moléculaire et d'élongation des parois latérales.

Chapitre 2

Dans le Chapitre 2, nous présentons le principe de l'acquisition d'images avec les techniques Total Internal Reflection Fluorescent (TIRF) et Structured Illumination microscopy (SIM). Nous présentons les spécificités des images acquises pour observer MreB. Nous mettons en évidence les sources d'erreurs de mesure et les limites de ces deux microscopes. Les limitations de la microscopie à fluorescence sont principalement dues aux aberrations optiques, à la résolution du système microscopique et au bilan photonique disponible pour le spécimen biologique. En microscopie TIRFM, la résolution latérale est limitée par la limite de diffraction $\sim 250nm$. En imagerie SIM, cette résolution est doublée à $\sim 120nm$ grâce au principe de l'illumination structurée. Nous évoquons les différentes sources d'erreur impliquées : l'erreur de localisation due à la fonction d'étalement du point ("Point Spread Function (PSF)) et à l'émission stochastique inhérente du photon, l'approximation induite par la projection des dynamiques 3D sur un plan 2D, ainsi que le champ d'observation incomplet.

Chapitre 3

Dans le Chapitre 3, nous passons en revue les méthodes statistiques couramment utilisées pour analyser des trajectoires et nous évaluons leur performances sur des trajectoires de filament MreB extraits des images SIM-TIRF, tout en tenant compte des limitations (bruit de mesure et projection) présentées dans le Chapitre 2. Les méthodes concernées peuvent être classées en deux catégories. La première vise à analyser des trajectoires individuellement, c'est-à-dire à déterminer le modèle de diffusion le plus pertinent pour décrire la trajectoire, et estimer les caractéristiques biophysiques associées (i.e. dérive et le coefficient de diffusion). La seconde approche s'intéresse à l'estimation des champs de dérive et de diffusion en estimant ces caractéristiques dans des voisinages locaux.

Chapitre 4

Le Chapitre 4 vise à concevoir et évaluer un cadre mathématique pour aborder la question de la reconstruction des trajectoires des biomolécules en mouvement sur des surfaces cylindriques 3D, à partir de bouts de trajectoire observés dans des fenêtres échantillonnées sur la surface. L'approche a été développée pour analyser des trajectoires dirigées. Les biomolécules sont supposées suivre un mouvement brownien stochastique avec dérive et peuvent apparaître ou disparaître pendant la période d'observation. Dans le cadre probabiliste choisi, les processus de Poisson sont utilisés pour décrire les événements de "naissance" et de "mort". La procédure algorithmique développée relie les bouts de trajectoire appartenant à une même trajectoire, et permet ainsi d'inférer les trajectoires des biomolécules sur la surface fermée.

INTRODUCTION

Context and motivation

Scientific context and motivations

During the past two decades, biological imaging has undergone a revolution in the development of new microscopy techniques that allow visualization of tissues, cells, proteins and macromolecular structures at all levels of resolution, physiological states, chemical composition and dynamics. Thanks to recent advances in optics, digital sensors and labeling probes (e.g., Colored Fluorescence Protein), one can now visualize sub-cellular components and organelles at the scale of several hundreds of nanometers to a few dozen nanometers. As a result, fluorescent microscopy and multimodal imaging (fluorophores at various wavelengths) have become the spearhead of modern biology. All these technological advances in microscopy have created new issues and challenges for researchers in quantitative image processing and analysis. Since digital imaging is part of the imaging loop, image processing may even drive imaging. A brilliant example of this shift in paradigm is Single-molecule localization microscopy (PALM, STED), which was awarded the 2014 Nobel Prize in Chemistry.

Challenges in biological image processing and quantitative microscopy

In most cases, modern microscopy in biology is characterized by a large number of dimensions that fit perfectly with the complexity of biological features: two or three spatial dimensions, at macro to nano-scales, and one temporal dimension, sometimes spectrally defined and often corresponding to one particular biomolecular species. Dynamic microscopy is also characterized by the nature of the observable objects (cells, organelles, single molecules, etc.), by the large number of small and mobile elements (chromosomes, vesicles, etc.), by the complexity of the dynamic processes involving many entities or group of entities sometimes interacting, by particular phenomena of coalescence often linked to image resolution problems, finally by the association, dissociation, recomposition or constitution of those entities (such as membrane fusion and budding). Thus, the corpus of data to be considered for any analysis involving multiple image series acquisitions is massive (up to a few Gigabytes per hour). Therefore, it becomes necessary to facilitate and rationalize the production of those multidimensional data, to improve post-acquisition analysis, and to favor the organization and the interpretation of the information extracted from this data corpus. This motivates innovative methods and concepts for data fusion, image registration, super-resolution, data mining, etc. More importantly, modern microscopy led to recent breakthroughs, related to the potential correlations between interacting molecules in the cell. A long-term research aim now consists in inferring the relationships between the dynamics of macromolecules and their functions.

In order to tackle the aforementioned challenges, it is necessary to develop innovative approaches and paradigms for image reconstruction, 3D molecule tracking and motion estimation, and biophysical parameter estimation to face the huge data volumes acquired with cutting-edge microscopy set-ups. To this end, applied mathematics, image processing and analysis, and computer science have to be considered in association with biophysics and biology.

Biological context

Determining the mechanisms of cell shape establishment is one of the critical goals of cell biology. In bacteria, it is widely assumed that the extra-cytoplasmic cell wall (CW) is the primary physical determinant of cell shape. The CW is mainly composed of a peptidoglycan (PG) meshwork, which is not static but expands in permanent concert with cell growth. The composition of the cell wall and the pathways involved in the synthesis of the molecular components are well established. Still, its three-dimensional organization and the mechanism controlling its assembly remain unclear.

In general, it is very important to understand how a bacterial cell grows, i.e., how its outer rigid envelope is synthesized, and how certain cellular processes take place through this envelope (e.g., viral infection, secretion, and genetic competence). In a more focused way, more than half of the antibiotics discovered (e.g. beta-lactam antibiotics such as penicillins or amoxicillin) directly or indirectly target the bacterial wall synthesis processes and thus understanding how the CW synthesis machinery is spatio-temporally controlled or how the wall synthesis is regulated to allow different growth rates and sizes may lead to a better understanding of the mechanism of action of antimicrobials and open new therapeutic avenues.

During the past decades of research, the molecular actors implicated in CW elongation have been identified and analyzed (Fig. 4). It is well established that the protein MreB, with its filamentous structure, mediates the PGEM activities with permanent dynamics. The trajectories of MreB aggregates have been analyzed with the help of high-resolution fluorescent microscopy and single particle techniques. The analyses of the dynamics of MreB help us to addressing different questions which interest microbiologists: what the different roles/functions of MreB patches play during its different diffusion regimes; what drives the directed circumferential dynamics of the MreB-mediated PGEM, more specifically the Rod Complex; and finally, how MreB organizes different enzymes to work together on expanding the PG.

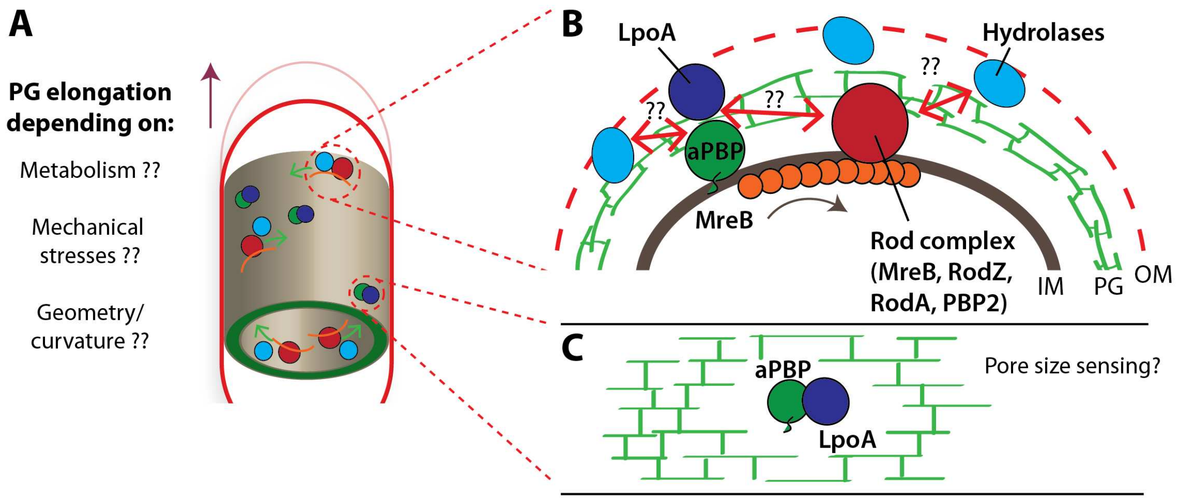


FIGURE 4 – **Representation of the prevailing mechanisms of CW elongation in rod-shaped bacteria (Figure from [TR18])** (A) Illustration of machinery implicated during cell growth, i.e., peptidoglycan (PG) elongation, in rod-shaped cell: the rod complex, hydrolases, and class A penicillin-binding proteins (aPBPs). Despite long-lasting efforts to understand bacterial CW elongation, many open questions still exist about how PG assembly and cell growth adapt to different factors, such as metabolism, mechanical stresses, and cell shape. (B) Insert illustrating the key CW synthesis machineries which work independently, according to a recent work [Cho+16]: (i) the processively moving rod complex containing the transglycosylase RodA, the transpeptidase PBP2, and circumferentially oriented MreB filaments along the inner membrane; (ii) the bi-functional aPBPs. Hydrolases may be actively engaged in CW cleavage during CW synthesis. It remains to be discovered how hydrolases interact with the rod complex and aPBPs, and how the rod complex interacts with aPBPs. (C) (in Gram- Bacteria) aPBPs are activated through outer membrane (OM) lipoproteins LpoA and LpoB. This interaction is suggested to provide a mechanism to sense pore sizes in the CW to guide PG insertion.

Diffusion models for dynamics analysis in bioimaging

In the subcellular environment, active life activities are constantly taking place. The molecular diffusion underlying the operation of the cell has always fascinated biophysicians, mathematicians, and statisticians to analyze them [BN13]. The diffusion may concern proteins, receptors in neuronal cells, and vesicles in the endo/exo-cytosis activities which depict great heterogeneity, in terms of motion regime. Free diffusion, which was first explained by A. Einstein and P. Langevin from two different perspectives, has the position of the cornerstone of the diffusion process. Later on, other diffusion processes, mostly characterized by the Stochastic Differential Equation (SDE), have been proposed to model more complicated intracellular biomolecular diffusion, related to different underlying mechanisms.

A large research effort in cell biology is focused on the inference on the modes of mobility of molecules ([Ver+21], [Ser+20]). It is believed that it may be possible to reconstruct some biological structures from the trajectories of a certain biomolecule, and to unravel the underlying physical processes corresponding to these trajectories.

In the mean time, statistical ensemble modeling from large quantities of data has also been developed to observe local physical properties of the particles, such as velocity, diffusion, confinement, or attracting forces, reflecting the interactions of the particles with their environment ([Lau+19], [HH15]).

Mathematical or statistical modeling is usually evaluated on synthetic data, before being applied on experimental biological data. It is worth noting that no perfect experimental data exist. The measurement noise is induced in data acquisition and trajectory construction. The lifetime of fluorophore set a constraint on the temporal resolution of acquisition and the duration of the observation. Therefore, some research work has been dedicated to the evaluation of noise and artefacts on the estimation of the parameters of interest or the processes under study.

Contribution of this thesis

In this thesis, physical, statistical and mathematical modeling methods for the analysis of single particle trajectories have been reviewed and improved, accompanied by experimental evaluation on a specific problem, concerning the dynamics of MreB. The experimental data related to MreB dynamics acquired by TIRF microscopy have some limitations, which can be categorized into three groups: measurement errors (due to the spatial resolution), the time resolution and the duration of the movie (limited by the lifetime of fluorophore) and the partial observation induced by TIRF evanescent field (Fig. 5). In the specific problem of MreB, we should also consider the particular features of the dynamics of MreB, e.g., the diffusion types (or modes), the density of the particles, the existence form of the MreB (MreB exists as a filament of $\sim 170nm$. SIM can therefore bring improvement in the perception of the MreB dynamics, as it allows us to observe a unique MreB filament).

In this context, we consider two modeling and estimation frameworks to analyze the trajectories of MreB based on diffusion characterized by Stochastic Differential Equations. The first approach consists in analyzing individual trajectories, that is determining their diffusion mode (free diffusion, super-diffusion or sub-diffusion), as well as estimating diffusion parameters (drift and diffusion coefficient). We evaluate the performance of several classification methods on MreB trajectories. Then we apply two estimators to obtain the diffusion coefficient for the class of free diffusion particles. The estimation of the diffusion coefficient is refined by including the measurement errors. We study the "optimal" estimators related to the Cramer Rao Lower Bound (CRLB) and try to answer the question of how well the diffusion coefficient can be estimated, given the image quality (Signal-to-Noise Ratio). The second approach involves investigating the dynamics of particles in the neighborhood of a position point, and further obtaining the drift field and the diffusion field. Assuming that the dynamics, defined by a prior dynamic model, take place on a 3D surface and are recorded as the planar projection, we derive the relation between the local drift and local diffusion coefficient, and the local curvature. Therefore, the resulting estimation method can lead to the recovery of the 3D surface and the reconstruction of the 3D dynamics, from the planar projection of the dynamics.

In the central part of the thesis, in order to deal with the incomplete view in TIRF microscopy, we focused on the design and evaluation of a self-contained mathematical framework to tackle the reconstruction of particle trajectories on cylindrical surfaces, given the tracklets observed in a small window (the region of observation (ROO)) sampled on the surface (Fig. 5). In our study, the particles are assumed to obey a Brownian motion with drift and may appear or disappear during the observation period. The birth and death events are described by Poisson processes. Split or merge events of aggregates are not considered in the modeling framework. The trajectory reconstruction problem is defined as the maximization of the likelihood function given tracklets inside the ROO. The optimization problem to be solved is formulated as an integer linear programming problem. The final algorithm is a data-driven algorithm with no hidden parameter to be set by the user. We evaluated the performance and robustness of our computational method on simulation data, by varying the ratio of observed to unobserved regions, the drift and variance of particles, as well as the rates of birth and death of particles. An illustration of the reconstruction procedure is shown in Fig. 6. The procedure is also tested on an experimental MreB dynamics movie. According to the procedure, some MreB particles having exited the ROO are identified when they re-enter into the ROO. This statistical and computational approach is the main contribution of the thesis.

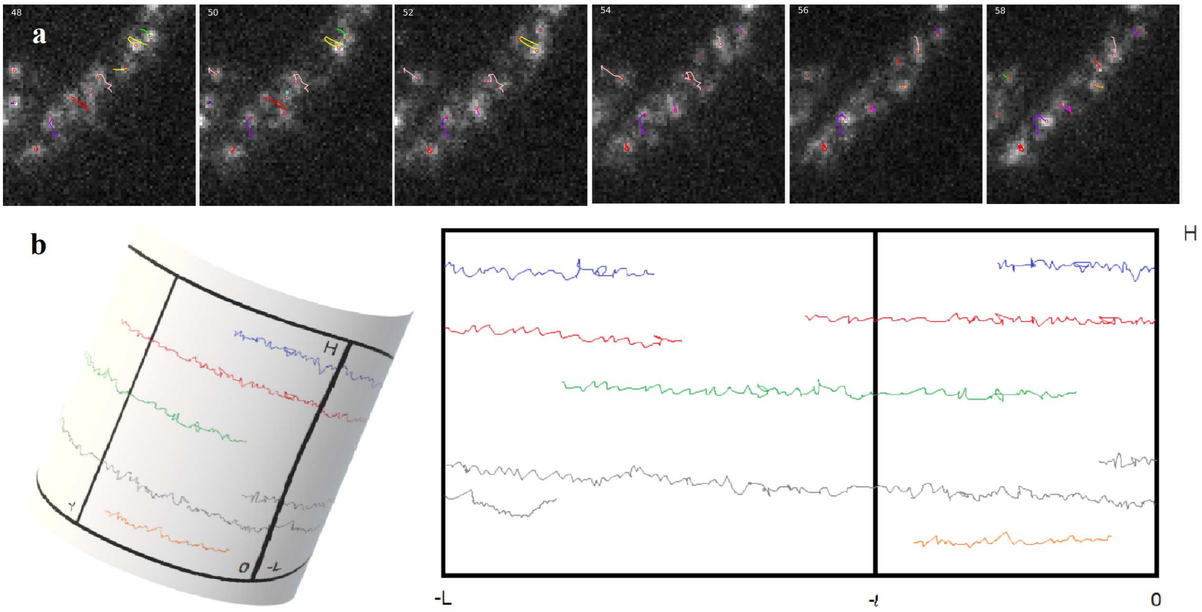


FIGURE 5 – (a): Several consecutive images from a real TIRFM movie [ABT84]. Tracks are superposed on the images. (b) Left: Illustration of trajectories observed during recorded time $[0, T_S]$ on the surface of a cylinder. Only the motions inside the regions of observation (ROO) $]-l, 0[\times[0, H]$ can be observed, even though the dynamics happen on the whole surface. Right: Representation of the dynamics on a two-dimensional (2D) unwrapped surface $]-L, 0[\times[0, H]$. The objective is to recover the dynamics on the whole surface from the partial observations, by coordinating the inputs through $\{-l\} \times [0, H]$ and the outputs through $\{0\} \times [0, H]$ in a movie during T_S, taking into account particle birth and death events.

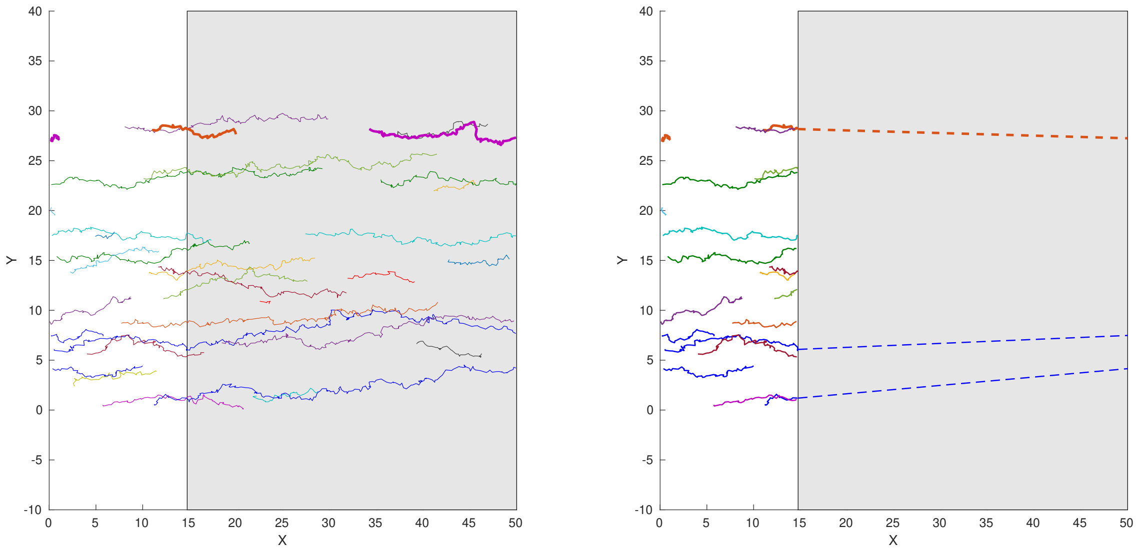


FIGURE 6 – Illustration of the reconstruction results on the cell surface approximated by a cylinder. $Y-$ direction represents the axis of the cylinder and $X-$ direction the unwrapped cylindrical side. The shaded part represents the hidden region and the light part corresponds to the region observed by TIRF microscopy. Left: Simulated trajectories in one movie. Right: The connection results.

The organization of the manuscript

This thesis is organized into four chapters as follows:

Chapter 1

In Chapter 1, we describe the cell wall's components, its most popular 3D structure modeling and its dynamic expansion during cell growth, for rod-shaped bacteria. During the last decades, the molecule actors involved in Peptidoglycan synthesis, have been identified and analyzed individually. We introduce the enzymes that have been identified in the Peptidoglycan elongation machinery (PGEM). It has been also established that the protein MreB, with its filamentous structure, mediates PGEM activities with permanent dynamics. In this chapter, we briefly describe the orientation and dynamics of the MreB filament, and we discuss several hypotheses about the function of MreB. Several works devoted to the analysis of the dynamic interaction of MreB with other enzymes are discussed, which open new perspectives for understanding the molecular organization and the mechanism of sidewall elongation.

Chapter 2

In Chapter 2, we explain the basics related to image acquisition with TIRF and SIM techniques. We present the modalities of acquisition of MreB dynamics images. We highlight the sources of measurement errors and the limitations using these two microscopes. Fluorescence microscopy limitations are due to optical aberrations, the resolution of the microscopy system, and the photon budget available for the biological specimen. In TIRFM, the lateral resolution is limited by the diffraction limit $\sim 250nm.$ In SIM this resolution is doubled to $\sim 120nm$ thanks to the structured illumination. We summarize different sources of error, the localization error due to the PSF and due to the inherent stochastic emission of photon, the approximation caused by the projection of 3D dynamics onto a 2D plane, and the incomplete observation field.

Chapter 3

In Chapter 3, we review commonly-used statistical methods for trajectory analysis and we evaluate the performance of these methods on the trajectories of MreB filaments extracted from SIM-TIRF images, while taking into account the limitations (measurement noise and projection) of the images introduced in chapter 2. The methods concerned can be divided into two categories. The first one comprises individual trajectory analysis, i.e. the classification of the diffusion regimes, and the estimation of the diffusion features (the drift and the diffusion coefficient). The second category concerns the estimation of the drift field and the diffusion field by estimating the diffusion features at each point position in the field.

Chapter 4

Chapter 4 focuses on the conception and evaluation of a self-contained mathematical framework to tackle the reconstruction of particle trajectories on cylindrical surfaces, given the tracklets observed in a small window sampled on the surface. The analysis is restricted to directed trajectories. The particles are assumed to obey a stochastic Brownian motion with drift and may appear or disappear during the observation period. In the probabilistic framework, Poisson processes are used to describe birth and death events. The developed computational procedure aims at connecting tracklets belonging to the same trajectory, and thus recovering the dynamics of particles on the whole closed surface.

CHAPTER 1

CELL BIOLOGY OF THE BACTERIA CELL WALL

Introduction

Determining the mechanisms of cell shape establishment is one of the critical goals of cell biology. In bacteria, it is widely assumed that the extra-cytoplasmic CW is the primary physical determinant of cell shape. The CW is mainly composed of a PG meshwork, which is not static but expanding in permanent concert with the cell growth. The composition of the cell wall and the pathways involved in the synthesis of the molecular components are well established [Höl98]. Still, its three-dimensional organization and the mechanism controlling its assembly remain unclear. PG is also known to bear the intracellular osmotic pressure and protect bacteria from the outside environment and, importantly, it is unique to bacterial cells. It is consequently the target of numerous antibiotics.

During decades of research, the molecular actors implicated in cell-wall elongation have been identified and analyzed. The PGEM is thought to contain the transmembrane proteins MreC and MreD, RodA and RodZ, PG hydrolases, and penicillin-binding proteins (PBPs), the enzymes that catalyze PG elongation and cross-linking ([CLF07], [Mar09]). It has been shown that the protein MreB controls the action of the PGEM ([Gar+11], [DE+11]). However, how these presumed individual PGEMs are coordinated in time and space and surmount the formidable task of weaving an intact mesh network three orders of magnitude larger than themselves, remains active research subject.

Over the last two decades, the rapid development of protein labeling and high-resolution fluorescent microscopy have made possible that we observe specifically targeted protein dynamics with unprecedented spatial and temporal resolution. TIRF microscopy is a technique that allows us to image events close to the cell surface. This property makes it possible to observe the events in the vicinity of the CW and cell membrane and avoid the noise coming from the unrelated cytoplasmic dynamics.

TIRFM makes it possible to unravel that MreB proteins form discrete, diffraction-limited (less than $250nm$) assemblies that exhibit different dynamic behaviors along the bacterial sidewalls, processive, circumferential motion around the rod-shaped cell's perimeter, free diffusion, or constrained motion [Bil+17]. Cells adapt the dynamics of MreB to distinct growth states. More recently, super-resolution SIM-TIRF (resolution $\sim 120nm$) showed that in the Gram-positive model organism *Bacillus subtilis* MreB forms $\sim 170nm$-long filaments during exponential growth period. On the contrary, the filaments' width is still inaccurate, as it is under the resolution limit [Bil+19].

The function of MreB filament is debated. It is well accepted that MreB moves in the direction of its orientation. The prevailing model states that MreB filaments intrinsically orient in the direction of the largest principal curvature ([Hus+18]) of the membrane. This property alone can promote the formation of rod-shape and reinforce the elongation of rod-shape. It has also been hypothesized that diffusive MreB may be in the process of recruiting other proteins to form the PGEM. Study of the dynamics of components of the PGEM can shed light on the role of MreB and the organization of the enzymes during CW expansion. By combining different microscopic techniques, TIRF, SIM, and single-particle Tracking (SPT)-Photoactivated Localization microscopy (PALM), it is possible to observe the dynamics of proteins in different spatio-temporal scales. Some mathematical modeling also exist to combine the enzymes organization and the mechanical 3D expansion model of the PG during cell growth ([Din+17] and [Ngu+15]).

In Section 1.1, the PG composition, its most popular 3D structure for both Gram-positive (Gram+) and Gram-negative (Gram-) bacteria, and the known molecular actors involved in PG synthesis are introduced. Next, in Section 1.2, the knowledge about MreB filaments, their existing form, orientation and dynamics observed by optical fluorescence microscopy are presented. Finally, in Section 1.3, the interaction of MreB with other PGEM proteins, e.g., RodZ, RodA, and PBPs is discussed.

1.1 The bacteria cell wall

The bacteria envelope, at the interface between the external medium and the intracellular medium, consists of several structures: the cytoplasmic membrane; the CW, a rigid sugar-based exoskeleton (or sacculus) primarily composed of a PG meshwork; and the outer membrane in Gram- bacteria (Fig. 1.1) [Liu+15]. In Gram+ bacteria, in addition to PG, the CW is also made of anionic polymers (teichoic and lipoteichoic acids). They constitute up to 60% of the dry weight of CW in *B. subtilis* and have essential cellular function [AS21]. The CW is a highly dynamic structure, constantly synthesized and remodeled as the cell cycle progresses and as the bacterium adapts to its environment. In this work, we were mostly interested in the bio-genesis of the cylindrical part of the CW in *B. subtilis*. We will describe the composition and architecture of the wall and the major molecular actors involved in the elongation of the CW.

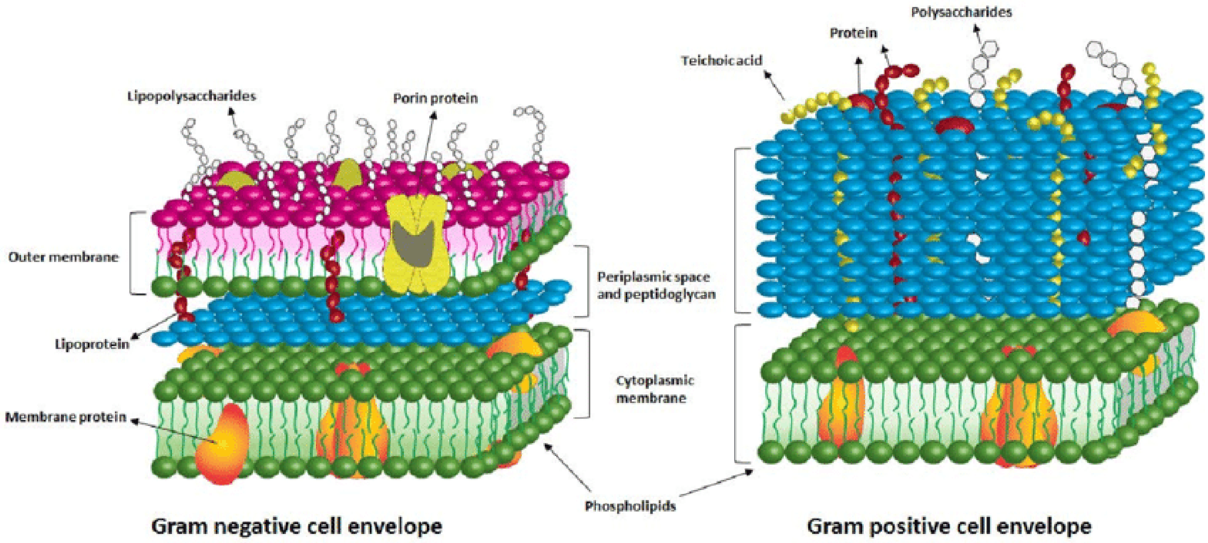


FIGURE 1.1 – Schematic illustration of the cell envelope in Gram- (left) and Gram+ (right) bacteria [Liu+15].

1.1.1 The composition and architecture of the cell wall

Along decades of study, the chemical composition of PG has been revealed with great details, as well as the biosynthetic pathway leading to the construction of the PG scaffold [AS21]. The PG is a unique large macromolecule that envelops the entire bacterial cell. It is made of a covalently bound meshwork of glycan strands interconnected by peptide bonds. Glycan strands are made of an alternation of two families of sugar molecules, N-Acetylmuramic acid (NAM) and N-acetylglucosamine (NAG) (Fig. 1.2). Two strands are linked by peptidic bridges between two NAM molecules. Supported by Atomic force microscopy (AFM), Electron cryotomography (CryoET) imaging, and live-cell imaging of cell-wall insertion, the current model is that the glycan strands are oriented around the cell circumference in rod-shaped bacteria, either Gram- or Gram+([Bee+13], [GCJ08]). In Gram-, the PG layers is $\sim 5nm$ thick and the PG structure is compatible with a monolayed model of sacculus (Fig. 1.1, left). In Gram+ bacteria, the sacculus structure is harder to observe due to the thickness of the material ($25nm \sim 50nm$) [MB05]. The CW is composed of several (10 to 30) layers of PG strands (Fig. 1.1 (right)).

1.1.2 Different models of biogenesis of the cell wall (Remodeling of the cell wall during cell expansion)

Although the composition of PG is well known, the dynamics and the 3D structure of Gram- and Gram+ sacculi are poorly understood. Even though this thesis is primarily concerned by the CW of Gram+ bacteria using the *B. subtilis* as model, the remodeling of CW is discussed in this

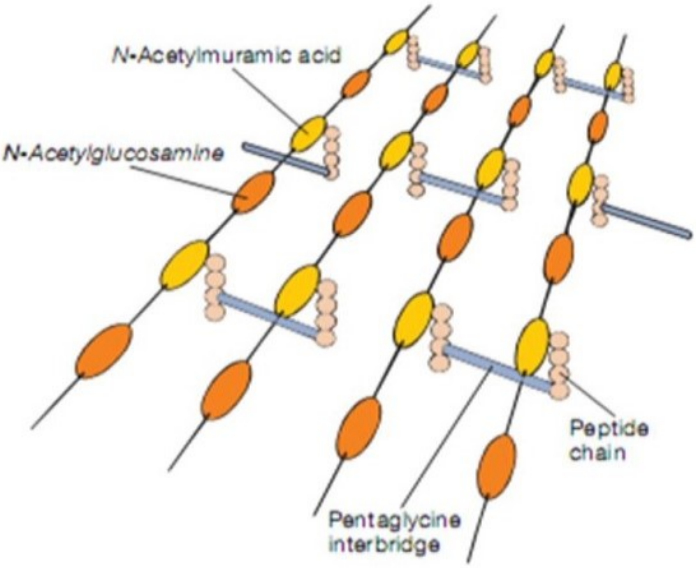


FIGURE 1.2 – A schematic diagram of one model of PG. Shown are the polysaccharide chains made of alternating N-Acetylmuramic acid and N-Acetylglucosamine, the tetrapeptide side chains and peptide interbridges [WSW08].

section with relation to both Gram+ and Gram- bacteria. The common features and differences between the two models can be informative of each other.

The series of events involved in the PG synthesis can be summarized as follows: the basic bricks, made of a disaccharide (N-Acetylmuramic acid (NAM) plus N-acetylglucosamine (NAG)), bearing a pentapeptide, are synthesized in the cytosol and flipped across the membrane before being assembled into glycan chains by transglycosylases, and finally cross-linked into the existing mesh by transpeptidation of their stem peptides (see [Höl98] for a review).

In Gram- bacteria, CW expansion implies expansion of the PG monolayer. New insertion happens at the same region as the hydrolysis. The prevailing strategies of CW expansion are the "cut and insertion" from [BP84] and the "3-for-1" from [Höl98]. In a "cut-and-insertion" strategy (Fig. 1.3b left), peptide bridges are cleaved, allowing new glycan strands to be inserted. However, odd numbers of strands (e.g., single strand insertion) impose a stress on the cross-links as the peptides on the left and on the right of a strand are not aligned but staggered, while even numbers (e.g., double-strand insertion) do not. In the "3-for-1" insertion model (Fig. 1.3b right), a triplet of glycan strands is inserted to replace an existing strand of the meshwork. The strand number increases by two, therefore no stress is introduced. Moreover, in the "3-for-1" model, 50$\%$ of PG is recycled per cell cycle, which coincide with the observed high rate of PG recycling per generation in Gram- [PU08].

For the multilayered Gram+ sacculi, the "3-under-2" model was recently proposed by [Bil+17]. Due to the thickness of the cell wall in Gram+ bacteria ($\sim 25-50nm$), the detailed ultrastructure of the sacculus is not yet been established [VH04]. In the "3-under-2" model, the assumption was made that the CW is composed of several concentric layers of PG. In each layer, parallel circumferential glycan strands cross-linked by peptide bridges run perpendicular to the long axis of the cell [Koc98] (Fig. 1.3a). The "3-under-2" model is based on the previous work of "inside-to-out" mode of growth, where new layers of PG are added to the innermost face of the sacculus, pushing outwards the previous layers which will be eventually degraded by hydrolases, accounting for the observed PG turnover ([KD85], [Poo76], [DBKW81]). In the "3-under-2" model, other two assumptions were made. Firstly, a single new PG layer is added per cell cycle and the cell has to double its length. Secondly, the stress-bearing layer is the innermost PG layer. According to the "3-under-2" model, three new glycan strands are added underneath two strands of the innermost PG layer (Fig. 1.3c). It worth noting that the addition of new PG strands will not allow expansion, until peptide cross-links on the upper layers are cleaved in a somewhat synchronized way and the new layer becomes stress bearing.

Other possibilities of stress-bearing are explored (Fig. 1.3c). If the stress is borne by the penultimate $n-1$ layer (Fig. 1.3c right), the number of glycan chains inserted to double the length of the sidewall is two-fold bigger than when stress is borne on the innermost layer. Accordingly, the energetic cost will increase as well.

The above mentioned principles are tested respectively by mathematical modeling, "3-for-1" principle for Gram- [Ngu+15] and the "3-under-2" for Gram+ [Din+17]. The mathematical models make it possible to couple global aspects of the morphogenesis with more local molecular activities related to the wall synthesis. These modeling approaches have proven useful to test predictions on the fine mechanisms underlying the insertion of strands during growth, such as the composition of the synthetic machinery, orientation of the enzymes, the sequence of enzymatic reactions, etc.

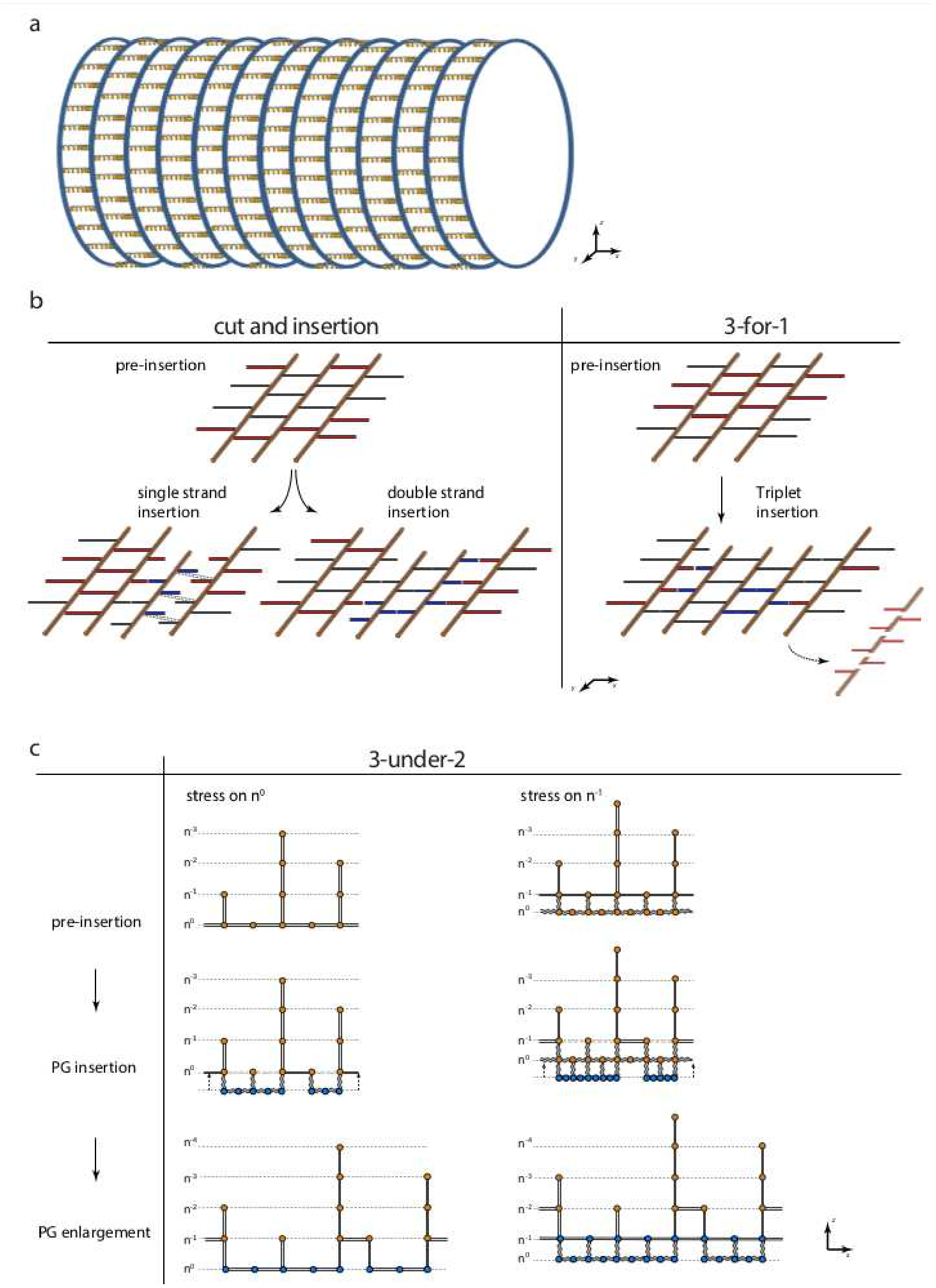


FIGURE 1.3 – **Schematic illustration of mechanistic models of PG synthesis in rod-shaped bacteria [Bil+17]** (a): Idealized and simplified 3D representation of one layer of PG. Glycan strands (blue rings) run orthogonal to cell's long axis, cross-linked by stem peptides (yellow

springs). (b): Strategies of glycan strand insertion in a 2D (x, y) PG layer in Gram- bacteria (monolayered PG). Red lines represent peptide bridges, brown tubes for disaccharide strands. In a "cut-and-insertion" strategy (b left), peptide bridges are cleaved, allowing new strands to be inserted. Odd numbers of strands (e.g., single strand insertion) impose a stress on the cross-links as the peptides are staggered rather than aligned (dotted line), while even numbers (e.g., double-strand insertion) do not. In the "3-for-1" insertion (b right), a triplet of glycan strands is replacing an existing strand of the meshwork. The net increase in strand number being even, no stress is introduced. (c): "3-under-2" glycan strand insertion modes in a multilayered PG meshwork in 2D side view (x, z) in Gram+ bacteria. Peptide bridges are shown in gray, in a relaxed conformation (broken lines), in full extension (straight lines), or hydrolyzed (dashed lines). Circles indicate cross-sections of glycan chains newly (blue) or previously inserted (orange). Two possibilities are explored in which stress either is borne by the innermost (newest) layer alone (left) or by the penultimate $n-1$ layer (right). Arbitrarily, only five PG layers are presented (numbered n 0 to n-4 from the innermost to the outermost).

1.1.3 Molecules involved in the construction of the cell wall and the special role of MreB

The following section concerns the two best studies bacterial models, the Gram+ *B. subtilis* and the Gram- *Escherichia coli*. displaying both dispersed insertion of new PG along the cell-wall, without preferred regions of insertion (e.g., at the poles).

Proper cell-wall expansion requires the coordination of multiple cytoplasmic and extracellular or periplasmic steps: (1) precursor synthesis and flipping from the cytoplasm to the extracellular/periplasmic side; (2) cell-wall cleavage by DD-endopeptidases and possibly other hydrolases for cell-wall expansion along the long axis of the cell; (3) glycan-strand polymerization (transglycosylation,RodA); (4) cross-linking to neighboring strands (transpeptidation PBP2A and PBPH in *B. subtilis*).

The most recent evidence suggests that two different machineries contribute to sidewall synthesis somewhat in an independent manner [Cho+16] (see Fig. 1.4, B, C): firstly, the rod complex consisting of mono-functional transpeptidases (PBP2A and PBPH in *B. subtilis*), the transglycosylase RodA, and the transmembrane protein RodZ, in concert with the cytoskeletal MreB; secondly, the bi-functional class A penicillin-binding proteins (aPBPs) (importantly, PBP1a and PBP1b in *E. coli* [Cho+16] and PBP1, PBP2c, and PBP4 in *B. subtilis* [Mee+16] and [KDE09]). The two machineries exhibit different dynamics in the cell membrane, which have been proposed to reflect distinct aspects of PG synthesis. The rod complex exhibits circumferential, processive movement around the cell diameter. It has been suggested that this corresponds to the processive insertion of long glycan strands. The aPBPs move seemingly diffusively in the membrane, with intermittent pauses, and have been associated with localized cell-wall insertion events, to repair PG synthesis errors or holes in the sacculus [Cho+16]. An

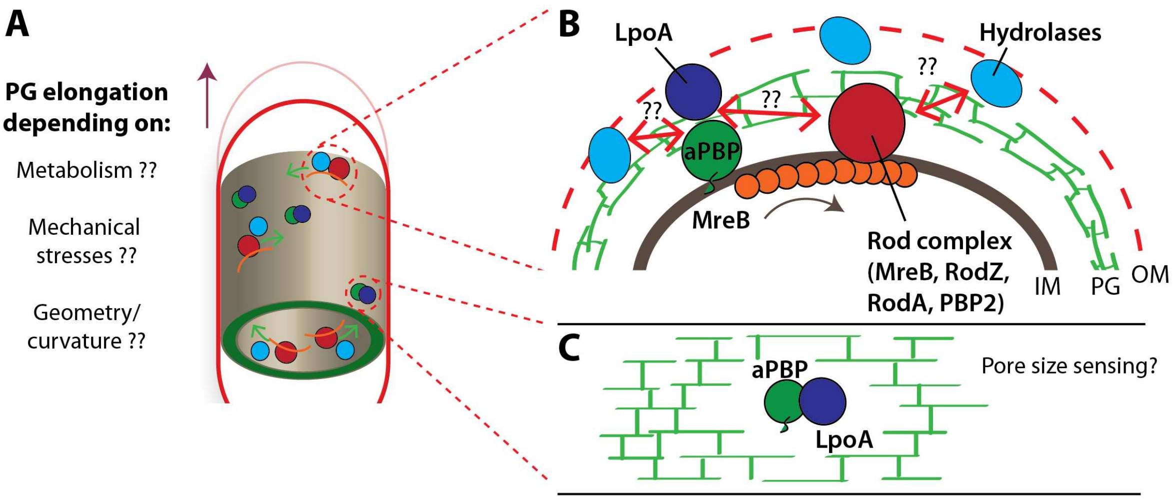


FIGURE 1.4 – **Representation of the prevailing mechanisms of CW elongation in rod-shaped bacteria [TR18]** (A) Illustration of machinery implicated during cell growth, i.e., peptidoglycan (PG) elongation, in rod-shaped cell: the rod complex, hydrolases, and class A penicillin-binding proteins (aPBPs). Despite long-lasting efforts to understand bacterial CW elongation, many open questions still exist about how PG assembly and cell growth adapt to different factors, such as metabolism, mechanical stresses, and cell shape. (B) Insert illustrating the key CW synthesis machineries which work independently, according to a recent work. [Cho+16]: (i) the processively moving rod complex containing the transglycosylase RodA, the transpeptidase PBP2, and circumferentially oriented MreB filaments along the inner membrane; (ii) the bi-functional aPBPs. Hydrolases may be actively engaged in CW cleavage during CW synthesis. It remains to be discovered how hydrolases interact with the rod complex and aPBPs, and how the rod complex interacts with aPBPs. (C) (in Gram- Bacteria) aPBPs are activated through outer membrane (OM) lipoproteins LpoA and LpoB. This interaction is suggested to provide a mechanism to sense pore sizes in the CW to guide PG insertion.

attractive hypothesis for the role of aPBPs is that in Gram- bacteria, aPBPs together with their cognate outer-membrane lipoprotein activators LpoA and LpoB ([Typ+10], [PB+10])(Fig. 1.4 C) sense the pore size on the PG meshwork and insert more PG where the pore size is increased by the mechanical pressure.

Physically, cleavage is indispensable for CW expansion. Endopeptidase activity has been found to be particularly important, as it cuts the bonds between neighboring glycan strands, thus allowing new PG to be inserted and the CW to be expanded along the long axis of the cell.

How these groups of machinery depend and interact with each other remains to be elucidated. Hypotheses are made that they may work together and even form transient joint complexes [Cho+16]. Elucidating the strengths and duration of these interactions and spatio-temporal organization of all these actors will be important to fully understand the important factors to maintain the integrity and rod shape of cell-wall.

1.2 The structure of MreB and the dynamics of MreB

The actin-like MreB protein is a crucial player of the machinery controlling the elongation and maintenance of the cell shape in most rod-shaped bacteria. MreB motion is shown to be dependent on PG synthesis ([DE+11], [Gar+11]), indicating that it parallels the action of the PGEM and thus reflects sidewall synthesis. Although the ultra-structure of MreB assemblies *in vivo* remains controversial, the current model proposes that they are membrane-associated scaffolds that spatially coordinate extra- and intra-cellular PG-synthesizing enzymes to ensure controlled cylindrical expansion of the sacculus ([Gar+11], [DE+11], [Ols+13], [Rue+14]).

It is well accepted that MreB forms filaments. Its structure, length, orientation, and conditions of establishment and its dynamics have been actively discussed ([Bil+17], [Bil+19], [Hus+18]).

MreB protein is known to be highly dynamic. A portion of the MreB filaments move along the short axis of cells, which is assumed to be driven by the PG synthesis. More recently, it has been debated that the curved structure of MreB filaments could be a determinant of the orientation of the filament and the orientation of the circumferential movement along the short axis of the cells ([Hus+18], [WGA19]).

1.2.1 The existing form of MreB

In vitro, MreB proteins self-assemble into filaments that bind directly to membranes ([Sal+11], [Ent+14]). *In vivo*, the structure of MreB first observed using immunofluorescence microscopy and Green fluorescence protein (GFP) fusion showed micron-long helical polymers spanning the length of the cell ([EWT06], [EAL01]). However, later *in vivo* studies, using higher-resolution light microscopy (TIRF and confocal microscopy), it is shown that MreB forms discrete patches along the plasmic membrane during active growth. MreB patches appeared as spherical or elliptical light spots close to the diffraction limit ($\sim 250-300nm$) ([VT+11], [Gar+11] and [DE+11], [Bil+17]). More recently, combining TIRF with super-resolution SIM, it was found that MreB assemblies have a mean length of $172nm \pm 41nm$ in *B. subtilis* during exponential growth [Bil+19]. As the length of MreB filaments was above the lateral resolution of the SIM-TIRF ($\sim 110nm$), the measure is reliable. However, the width of the filament is remained under the resolution limit of SIM-TIRF, calling for other super-resolution techniques to be elucidated.

1.2.2 The orientation of MreB filaments in the membrane and what it reveals

It has been known that the MreB filaments move in the direction along which they are oriented ([Urs+14], [Bil+14]). In [Hus+18], new insights were obtained into possible mechanisms by which MreB guides CW synthesis to create rod shape. First, the curved ultra-structure of MreB filaments causes them to self-orient and move along the direction of the greatest principal curvature of the membrane, coordinating the insertion of new PG strands in that direction. Second, both the formation and propagation of rod shape occur by a local, self-reinforcing process: once a local region of rod shape forms, it promotes self-alignment of MreB filaments along the highest curvature, which in turn propagates more rod shape by pushing the neighboring strands in the axial direction of the rod.

As MreB filaments align along the direction of the largest principal curvature, they are supposed to serve as a curvature sensor. This conclusion can be supported by the fact that MreB filaments move in the direction perpendicular to the axis in rod-shaped bacteria cells, even when in wall-less cells in which the rod shape is imposed outwardly [Hus+18]. Another observation can also support the conclusion. In spherical cells, the direction of MreB motion is isotropic. However, rod shape can be *de novo* established, in the region of local bulges, flanked by negative Gaussian curvature. Once rod shape is reestablished, the rotation of MreB filament and the new formed rod-shaped PG can reinforce and maintain the rod shape propagation [Hus+18].

The principle that MreB filaments align in the largest principal curvature can also explain the enrichment of MreB in regions with negative Gaussian curvature, observed in several reports ([Urs+14], [Won+17]) and the removal of MreB filament from cell poles. Wong et al. [WGA19] modeled MreB motion by minimizing the energy needed for inwardly curved MreB to bind into the less curved cell membrane. The model explained why MreB prefers to align the largest principal curvature. According to the same model, as the curved filaments need more energy to escape the regions with negative Gaussian curvature, it causes the enriched localization of MreB filaments in these regions.

Based on the mechanical binding model of MreB filaments to cell membrane [WGA19], a model for the dynamics of translocation of MreB filaments has also been proposed. The model for directed motion, called "biased random walk," suggests that MreB filaments move at the direction of its orientation θ, i.e., the direction of the largest principle curvature, with a slight bias $\eta \sim \mathcal{N}(0, \sigma^2)$ due to stochastic factors (equation (3) in [WGA19]). It is easy to demonstrate that this model is equivalent to the directed Brownian motion model (Bmd) ((3.11) in chapter 3). The authors acclaimed that the "biased random walk" model of filament trajectories leads to predictions of MreB localization.

The orientation of MreB filaments has revealed the role of local sensing of MreB ([WGA19], [Hus+18]). Consequently, MreB, together with the shape-reinforcing CW synthesis, could provide a robust, self-organizing mechanism for the stable maintenance and rapid reestablishment of rod shape. We gained one step further in the understandings of how the local activity of short MreB filaments ($\sim 170nm$) guide the emergence of a shape many times their size (cell diameter ($\sim 900nm$)).

Even though MreB filaments play a crucial role in the circumferential insertion of the CW, it should be noticed that other factors are also needed for the maintenance of rod shape. The interaction of MreB with other proteins will be discussed in the following sections. As the curvature of membrane-bound MreB filaments is much greater than the curvature of the cell membrane [WGA19], MreB filaments do not define a specific cell radius. A discussion about how MreB filaments intervene in the regulation of cell radius can be found in [Dio+19].

1.2.3 MreB exhibits distinct types of motion

As MreB assemblies are regarded as the proxy of PGEM, its localization and dynamics during growth and dynamic perturbations such as nutrient shift can reveal how bacteria cells regulate the expansion of CW and cell growth. Sub-populations of MreB with distinct behaviors were found by analyzing the Mean Squared Displacement (MSD) of the trajectories of MreB patches, imaged by TIRF microscopy. It is observed that MreB patches perform circumferential directed motion, random diffusion and constrained motion. Due to the limit of the classification method by MSD, one fraction of the trajectories of MreB is unclassified, which we call from now on the anomalous diffusion. This heterogeneity of the dynamics of MreB patches may reflect distinct physiological states (Fig. 1.5). The fraction of MreB performing circumferential motion is believed to be related to the PG insertion, with a speed $v = 30 - 60nm/s$ under different nutrient conditions. The circumferential motion, mostly orthogonal to the axis of the bacteria, is driven by the PG synthesis ([Gar+11], [Bil+17]). The intrinsic curvature of the MreB filament orients the motion in the direction of the maximal principal curvature (section 1.2.2). PG polymerization by aPBPs occurs in a diffusive manner, outside MreB-associated PGEMs ([KDE09], [SSS97], [Cho+16], [Lee+16]"). It is plausible that the fraction of MreB exhibiting random diffusion is associated with aPBPs-dependent diffusion PG synthesis. The anomalous motion of MreB may be caused by the interaction with other components of the PGEM. One hypothesis is that constraint motions reflect MreB recruiting missing components of the PGEM to initiate local PG insertion.

Another independent study of the dynamics of MreB in *B. subtilis* was carried out using Single Molecule Tracking (SMT) [Rös+18], which has a temporal resolution ~ 25 fold higher than in TIRF microscopy [Der+20]. Instead of analysis by MSD, which calculates one averaged diffusion coefficient for a trajectory, a Gaussian Mixture Model (GMM) was used to fit the displacement Δx to estimate the diffusion coefficient of single-molecule MreB. It was found that MreB diffusion could be best described by at least two populations: The slow moving population with diffusion coefficient of $D = 0.044 \mu m^2 s^{-1} \pm 0.005$ [Luc+18], and the fast diffusive population, with $D = 0.53 \mu m^2 s^{-1} \pm 0.08$, which lies in the range of a freely diffusing cytosolic protein (e.g., DnaA, [Sch+17]).

1.3 The interaction of MreB with other PGEM components

The machineries contributing to sidewall synthesis have been described above (section 1.1.3). To understand how MreB orchestrates PG synthases (transglycosylases and transpeptidases) and PG hydrolases to allow coordinated CW expansion, several studies have addressed the dynamic interaction of MreB with other enzymes in the rod complex, especially the transmembrane protein RodZ, the transglycosylase RodA and the transpeptidase PbpH, in *B. subtilis*. Based on the knowledge of the function of the individual enzymes, the interaction between MreB and these enzymes is expected to reveal the spatio-temporal organization of their activity during CW expansion.

1.3.1 The transmembrane RodZ is required for cell-shape maintenance in *B. subtilis*

RodZ is one of the numerous players in the CW biosynthesis game (see [AS19] for a mini-review of *E. coli*). It was discovered more than ten years ago ([SSN08], [Ben+09], [Aly+09]) and has been investigated mostly in *E. coli*. However, it is also found in numerous bacterial species, both Gram+ and Gram-, either rod-shaped bacteria or cocci. It is a bipartite membrane protein with a cytoplasmic domain and an extracellular domain. The structure of the extracellular domain of RodZ of *B. subtilis* was predicted by Nuclear magnetic resonance (NMR) [Per+15]. RodZ may interact with cytoplasmic proteins in the vicinity of the internal membrane and with CW synthesis enzymes in the periplasm.

RodZ interacts with MreB and several components of the Rod complex including MreC, MreD, PB2, RodA as shown using Bimolecular fluorescence complementation, Fluorescence Resonance Energy Transfer (FRET), and bacterial two-hybrid assays [Mor+15]. It also performs

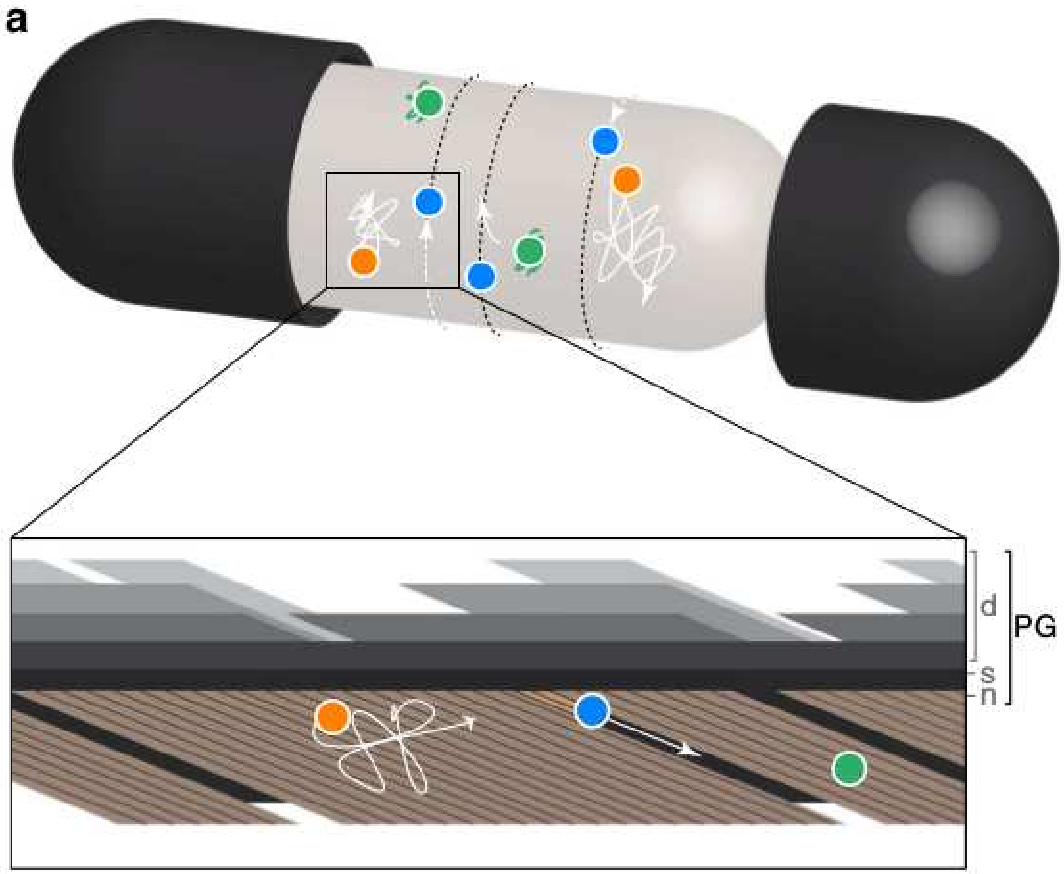


FIGURE 1.5 – **Schematic illustration of a Gram+ bacterium of MreB patches exhibiting different dynamic behaviors in the cytoplasmic membrane [Bil+17]:** Orange: randomly diffusing MreB filaments; blue: circumferentially moving MreB filaments; green: constrained MreB filaments. The black cover represents the PG CW. The zoom-in view shows different PG layers, d: discontinuous outermost PG layers; s: stress-bearing PG layer; n: newly constructed, uncompleted innermost PG layer.

circumferentially oriented motion and thus is assumed to belong to the Rod complex.

RodZ has also been observed in the vicinity of LipidII in the Gram- *Caulobacter crescentus*. It is supposed to be implicated in the regulation of peptidoglycan synthesis [Aly+09]. Several works have proposed that RodZ is not necessary but promotes MreB rotation [Der+20]. It could also regulate the MreB polymer number by stimulating the production of new polymers [Bra+18]. In the *B. subtilis* model, it is supposed that RodZ is required for cell-shape maintenance [MCB13].

To summarize, RodZ and MreB have intricate feedback with each other. MreB may influence RodZ localization and dynamics, and RodZ may influence the density and rotation of MreB polymers. However, RodZ is not the motor for the rotation of MreB. RodZ may be a link between MreB and proteins RodA and PBP2 from the Rod complex.

1.3.2 MreB, RodA, and PbpH do not form a stable complex that lasts over many minutes but rather perform constant dynamic binding and off-binding

It has been shown that RodA, the transglycosydases, and PbpH, the transpeptidase, circumferentially move around the cell wall, and their motion stronly colocalizes with MreB and its homologues (Mbl and MreBH) [DE+11]. The velocity and the angle of the trajectories of MreB, RodA and PbpH show strong similarities by comparing the maximum projections of two movies of two protein species, and by kymograph quantification.

In [Der+20], dynamics of RodA, PbpH and MreB are visualized with the same imaging settings. Two types of microscopy were employed, SIM-TIRF and SMT [Rös+18] to investigate the dynamics in different spatial and temporal scales. SIM-TIRF revealed that MreB, RodA, and PbpH all move in the direction perpendicular to the long axis of the CW ($\theta = 88{\pm}8°, 88{\pm}10°$, and $92 \pm 14°$, respectively) with similar velocity ($v = 37 \pm 16, 33 \pm 16$, and $38 \pm 22 nm/s$, respectively), which coincides with the findings in [Gar+11] and [DE+11]. More precisely, it is found that the distribution of the length of the tracks of the three proteins differs evidently, $8 \pm 3s, 3.8 \pm 2.1s$, and $2.4 \pm 1.1s$, respectively. These statistical findings support the fact that the three proteins share some tracks, because of the similarity of the velocity and the direction among the tracks. At the same time, it argues against the existence of a multi-protein complex that is stable over many minutes.

Using SMT [Rös+18], which offers temporal resolution 25-folds higher than SIM-TIRF, two populations of single-molecule dynamics of MreB are revealed, one freely diffusing ($D \sim 0.53 \mu m^2 s^{-1} \pm 0.08$) and another with slow/static mobility ($D \sim 0.044 \mu m^2 s^{-1} \pm 0.005$). The findings of different behaviors of MreB motion are comparable to a previous study, [Bil+17], where the motion of MreB is categorized into three groups, the directed motion (the slow motion), free diffusion (the fast motion) and constrained motion (the slow/static motion). The ratio of these two populations is around half-half. Similarly, two populations, half free diffusive and half slow/static moving are discovered for RodA and PbpH. It leads to assuming free diffusion for molecules not involved in active cell wall synthesis and slow-motion for enzymes associated with activities in synthesis or filament formation in the case of MreB. Interestingly, RodA, PbpH and MreB react remarkably differently when cells are challenged by an outward stress (e.g. osmotic stress). MreB responds to the stress condition by increasing the free diffusion fraction. In contrast, the movement of RodA and PbpH was hardly affected. During the stress adaptation period ($\sim 2h$), cells continued to grow, but in a way less coordinated.

Conclusions

Despite decades of research, understanding bacteria cell morphogenesis, i.e., how cells maintain their shape and regulate their size during cell growth remains an active research domain. The CW, as a rigid sacculus, plays a vital role in processes of bacteria cell morphogenesis. As the osmotic pressure bearer, the cell wall protects the physiology and metabolism inside the cell.

In this chapter, we reviewed several mechanical CW expansion models, e.g., "3-for-1" for Gram- and "3-under-2" for Gram+. Moreover, we presented the molecular actors (enzymes) discovered during the past years implicated in the cell wall elongation process. More specifically, two machineries were found to work independently, one is the processively and circumferentially moving rod complex, controlled by the MreB proteins, and the other is the bi-functional aPBPs. Biologists are interested in knowing how molecular actors are involved in the regulation of CW expansion and in the maintenance of cell shape. By mathematical modeling, different mechanistic hypotheses combined with the enzymes dynamics introducing new material into an existing PG network have been simulated. For Gram+ bacteria, due to the complexity of multilayer PG, the existing model remains preliminary and needs to be developed further.

To further explore the question of how bacteria cells maintain their shape, a much-studied key player is MreB, a cytoplasmic filamentous protein, which is known to control the PGEM working outside the cytoplasmic membrane to insert new PG in order to complete the task of building the PG meshwork. We summarized the known dynamic characteristics of MreB, e.g., its different modes of motion (circumferentially directed motion, free diffusion and constrained motion), its speed and diffusion coefficient, and the orientation of its directed motion. We also presented several works focused on depicting the interaction of the MreB with other crucial enzymes involved in cell wall elongation processes (e.g. RodA and PbpH) by quantitatively analyzing the simultaneously recorded fluorescent images of different species of molecules.

Over several decades, research in the field focused on the function of individual genes and proteins implicated in CW synthesis in isolation. Nowadays, thanks to the development of fluorescence labeling and super-resolution microscopy, it is a highly active domain to study proteins behavior by analyzing their high-resolution fluorescence microscopy images, to deduce their functionality or co-functionality with other proteins.

CHAPTER 2

IMAGING OF THE MEMBRANE PROTEIN MREB

Introduction

Optical fluorescent microscopy has the advantages of sample preservation, imaging flexibility and target specificity. It allows the observations of specifically targeted proteins or vesicles, performing dynamics in an ideally living specimen. It becomes a widespread tool used by biologists in understanding the cellular function and organization in a molecular scale.

In this chapter, we present several techniques related to the observations of the dynamics of MreB in model bacteria *Bacillus subtilis*. Total Internal Reflection Fluorescent (TIRF) microscopy is used to observe the dynamics of MreB protein in [Bil+17]. Compared to Wide-Field Fluorescence microscopy (WFFM), TIRF imaging provides improved axial resolution (limited by the diffraction of light, presented in Section 2.1) with thin optical sectioning ($< 200nm$). Moreover, it allows us to follow rapid dynamics of molecules by acquiring fluorescent images at high temporal resolution. Therefore, it is specifically suitable for observing the protein MreB which is known to be highly dynamic moving near the membrane. In addition, because that the evanescent wave avoids the illumination of the whole cell but only illuminate a thin section near the slides, TIRF provides high signal-to-noise and low phototoxicity to the living specimen.

However, in TIRF images, MreB proteins appear as patches of the same size, with diameter $\sim 250nm$ which is around the diffraction limit ([Bil+17], [Bil+19]). It suggests that the actual size of a unique MreB filament should be smaller than the diffraction limit. SIM is a super-resolution technique which goes beyond but is still essentially limited by the diffraction limit. By structured illumination, 2D-SIM can achieve at best a lateral resolution equal to half of the diffraction limit ($\sim 120nm$). However, SIM methods are suited for live-cell imaging with conventional fluorophores. Furthermore, it can be combined with TIRF to benefit the high axial resolution. In [Bil+19], the SIM-TIRF images are used to observe the dynamics of MreB and it brings very important information on the length of a unique MreB filament ($\sim 170nm$), which is hidden in conventional TIRF images.

The microscopic images of the dynamics of MreB during the active growth phase acquired by SIM-TIRF and TIRF are further analyzed, using advanced statistical and mathematical modeling in Chapter 3 and Chapter 4. It worth noting that during the image acquisition, different sources of noise are introduced, due to the stochastic nature of the photon emission and the imaging system. It can make the inference of the underlying biophysical processes (e.g., the mode of motion, the parameters of motion) more accurate, when we involve the noise into the modeling.

This chapter is organized as follows: an easy-to-understand introduction of the diffraction limit is presented in Section 2.1; In Section 2.2, we summarize the conventional and super-resolution fluorescent microscopy. The physical principles for the conception of TIRF and SIM are respectively presented in Sections 2.3 and 2.4. In Section 2.5, we summarize the common sources of noise that could be produced during the fluorescent image acquisition. Finally, in Section 2.6, we present the practical application to observe MreB filaments using TIRF and SIM-TIRF.

2.1 The diffraction limit

As stated previously, the resolution of conventional wide field microscopy is limited to $\sim 200 - 300nm$ laterally and $\sim 500 - 700nm$ axially due to the diffraction limit of light microscopy, as characterized by Abbe and Rayleigh [Ray80]. The image of an infinitely small self-luminous object in a microscopy is a finite-size diffraction pattern created by the interference in the image plane. The image consists of a central spot surrounded by concentric rings of decreasing intensity. The central spot which contains $\sim 84\%$ of the photons from the emitter, is called the Airy disk (Fig. 2.1 left). According to the Rayleigh criterion[Ray80], two point sources are just resolved if the central maximum of the intensity diffraction pattern produced by one point source coincides with the first zero of the intensity diffraction pattern produced by the other. Note the Airy disk radius as d, given by the equation $d = 0.61\lambda/\text{NA}$, where λ is the emission wavelength and NA is the numerical aperture of the objective, then optical microscope cannot distinguish two objects that are closer than d. This means that all tagged fluorescent proteins or protein complexes ($1 - 10nm$) show up as blured focal spots of diameter $\sim 250nm$ (of length $\sim 600nm$ in the axial dimension), and thus that two molecules within this distance cannot be separately identified.

The Point-Spread Function (PSF) describes the response of an imaging system to a point source or point object (Fig. 2.1). In 3D, the focal spot referred as the PSF appears as an ellipse stretched along the optical axis. In space-invariant system, i.e. the PSF is the same everywhere in the imaging space, the fluorescent images are the convolution of the PSF and the intensity of the image, related to the number of photons. The PSF is commonly approximated by a Gaussian distribution [ZZOM07]. The diffraction limit can be broken when SNR is high enough.

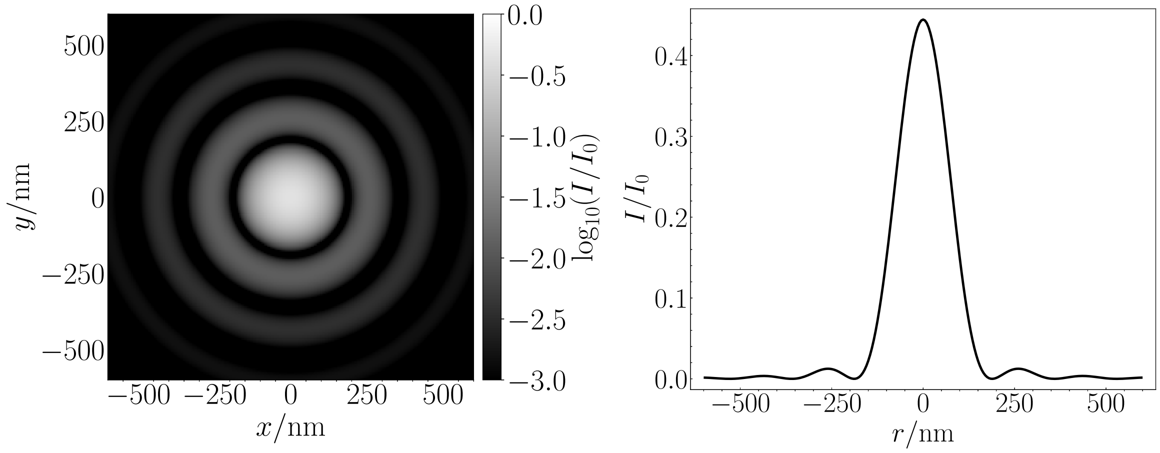


FIGURE 2.1 – Left: the Airy disk; Right: The Point-Spread Function (PSF) when setting $\lambda = 509nm$ (GFP emission wavelength) and $NA = 1.5$ (theoretical maximum value). Plotted according to the Point-Spread Function (PSF) equation in [ZZOM07]

If the good model of PSF is chosen and the images are recorded without noise, then any wanted resolution can be reached through modeling. However, in real, the PSF is unknown and the noises introduced during the acquisition are the reasons of the existence of a limit of resolution. One category of the recently widely used and still in fast development Super-Resolution microscopy (SRM) is based on the modeling of the PSF by localizing the position of a single molecule ([Hua+08], [SWB08]).

2.2 A summary of the conventional and super-resolution microscopy

In this section, we give a brief summary of the existing conventional and super-resolution microscopy techniques. We keep in mind the objective of observing the dynamics of MreB in model bacteria *Bacillus subtilis*. The bacteria cells are small, $\sim 1\mu m$ in diameter and several micrometer in length. The MreB patches are relatively sparse and perform at least two different modalities of motion (slow and fast) [DE+11].

The conventional techniques of light microscopy involve WFFM, TIRF, and Confocal microscopy (Fig. 2.2 [Sch+19]). In WFFM (or epi-fluorescence) setups, the incident light is parallel to the optical axis (i.e., normal to the coverslip/sample surface) and thus travels across the entire sample, simultaneously exciting all fluorophores in the cell. The detected fluorescence from out-of-focus plane generates a high background noise and thus a low signal-to-noise ratio (contrast).

On the contrary, in confocal microscopy, point illumination is used to scan across the sample and a pinhole in front of the detector keep only the in-focal signal. Therefore, confocal microscopy produces images with enhanced contrast compared to WFFM. However, the improved contrast is at the expense of fluorescence intensity (much of the light is blocked at the pinhole) and the temporal resolution (scanning is required as only one point is illuminated at one time). Given the diameter of a bacterium cell $\sim 1\mu m$ and the focal plane thickness $\sim 0.5\mu m$, confocal microscopy brings little improvement in the contrast.

In TIRF microscopy, an exponentially decaying evanescent wave is generated at the interface between the coverslip (glass) and the sample/water (aqueous medium), when a laser beam with an incident angle larger than the critical angle is completely reflected at the surface. TIRF excites the fluorophores at the proximity $(100 - 200nm)$ of the coverslip, smaller than the conventional axial diffraction-limited resolution. As only a small section of the cellular volume is illuminated, it can be much less phototoxic than WFFM, and can image longer periods if new fluorescent molecules enter into the evanescent field. Less detected out-of-focus fluorescence enables a high contrast. Due to the low phototoxicity, high contrast, high axial resolution and high temporal resolution, and its relatively low cost set-up, TIRF stands as a method of choice for studying membrane-associated events in bacterial cells, as in our case, the dynamics of the membrane-associated filamentous protein MreB.

Super-resolution techniques can be sorted into two main strategies, patterned light illumination techniques with Stimulated Emission Depletion (STED) and SIM, and localization-based techniques with Stochastic Optical Reconstruction microscopy (STORM) and PALM. Two recent reviews of the SRM are available, see: [LLB15] and [Sch+19], where an overview of commercially available and emerging SRM techniques, together with a balanced assessment of their strengths and weaknesses with examples of biological applications are provided.

Compared to other diffraction-unlimited SRM, SIM methods are considered rather ‘gentle’, and are particularly suited for live-cell imaging with conventional (nonswitchable) fluorophores (e.g., GFP), and for higher throughput applications. Classic interference-based SIM utilizes frequency shifting upon patterned wide-field illumination and mathematical reconstruction. It can best double the spatial diffraction-limited resolution in lateral (x, y) and axial (z), reaching $\sim 100nm$ lateral and $\sim 300nm$ axial resolution, equivalent to an ~ 8-fold volumetric improvement. However, the mathematical post-processing of classic interference-based SIM requires significant knowledge to detect and avoid the reconstruction artefacts [Dem+17].

In addition to TIRF and SIM, there exist other techniques which can bring improvement on the observation of the dynamics of MreB. However, due to technical and practical reasons,

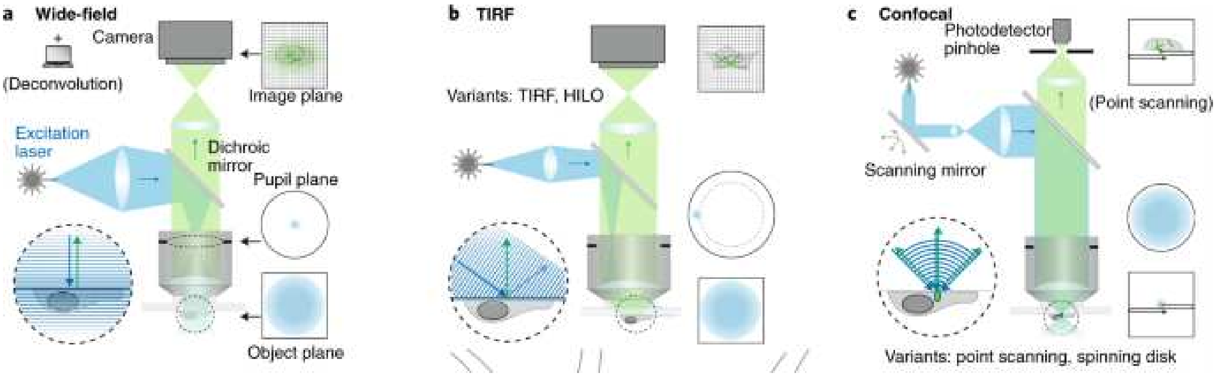


FIGURE 2.2 – **Illustration of the light paths of common conventional microscopy techniques. From [Sch+19]** (a) Wide-field illumination is achieved by concentrating the excitation light to a single point in the centre of the pupil plane, (b) In TIRF the spot is shifted to the edge of the pupil plane so that the light beam hits the cover slip interface at a super-critical angle to the optical axis, generating a rapidly decaying excitation beam close to the cover slip surface (the upright set-up is shown, however the inverted set-up for TIRF is more common), (c): In confocal microscopy, the pupil plane is filled, generating a focused spot in the image plane which is to be scanned across the field-of-view.

these techniques has not yet been realized on our model bacterium. They are further addressed in the discussion section. In the next two sections, we focus on the physical principles of TIRF and SIM.

2.3 Basic principles of Total Internal Reflection Fluorescence microscopy

The phenomenon of total internal reflection was first introduced to the domain of optical fluorescence microscopy by [ABT84]. Here we give a rapid recap of the principles of TIRF and its virtues compared to the conventional WFFM and the confocal FM. More details of the theories used in the conception of TIRF can be find in [MFTW13].

When a light beam hits the interface of two isotropic media with refractive indices n_i and n_t, then **Snell's law** describes the relation between the incidence angle θ_i and the refractive angle θ_t as follows:

$$n_i \sin\theta_i = n_t \sin\theta_t. \tag{2.1}$$

When the light passing from the medium with higher refractive index into that with lower refractive index (*i.e.* $n_i > n_t$), and as the incident angle increases, it reaches one point called the **critical angle** θ_c. The critical angle $\theta_c = \sin^{-1}(\frac{n_t}{n_i})$, where the refraction angle $\theta_t = 90°$ and the light is totally reflected.

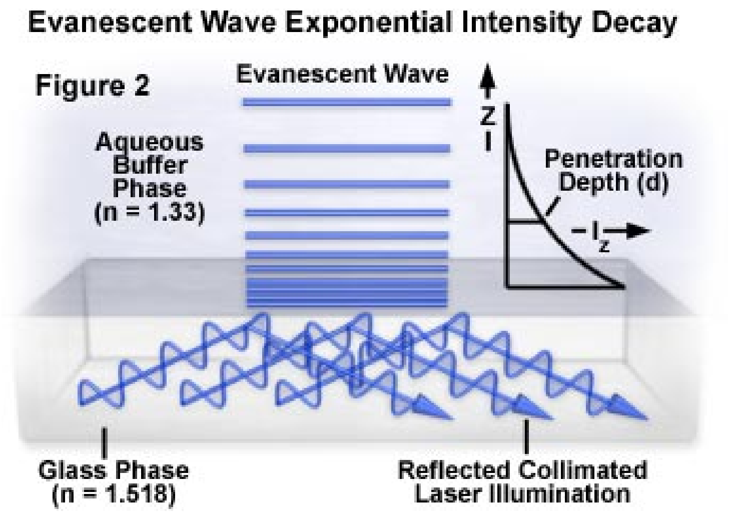


FIGURE 2.3 – The Exponential Intensity Decay in the evanescent field [AWD]

When the $\theta_i > \theta_c$, the large majority of the light is reflected and only a small portion of the reflected light penetrates through the interface, and propagates parallel to the plane of incidence creating an electromagnetic field in the vicinity of the interface. This electromagnetic field is named evanescent field and can be described by (2.2), where the real term $E_{ot}e^{-z\xi}$ describes the wave amplitude which decreases with z and the complex term describes the propagation of the electromagnetic wave along the interaction surface in the $x-$ direction.

$$E_t = E_{ot}e^{-z\xi}e^{i[k_t n_i x \sin\theta_i / n_i - \omega_t t]} \tag{2.2}$$

where $\xi = \frac{k_t}{n_t}(n_i^2 \sin^2\theta_i - n_t^2)^{1/2}$.

As the intensity of the electromagnetic I is proportional to E_t^2, from (2.2), I decays along $z-$ direction normal to the interface:

$$I = I_0 e^{-z/d} \tag{2.3}$$

where d, termed as penetration depth, is functional to the wavelength of incident light λ_0 and the refractive indices n_i and n_t:

$$d = \frac{\lambda_0}{4\pi}(n_i^2 \sin^2\theta_i - n_i^2)^{-1/2}. \tag{2.4}$$

It can be noticed from (2.3) that at $z = d$, the intensity $I = I_0e^{-1}$ (right side of Fig. 2.3), in other words, the intensity at d decays to e^{-1} times as the intensity at the interface (where $z = 0$).

The polarization properties of the evanescent field can be deduced using Fresnel's equations (see details in [MFTW13]). It can be shown that the intensity of the evanescent field at the interface, I_0, can be several times stronger than the intensity of the incident wave (Fig. 2.4).

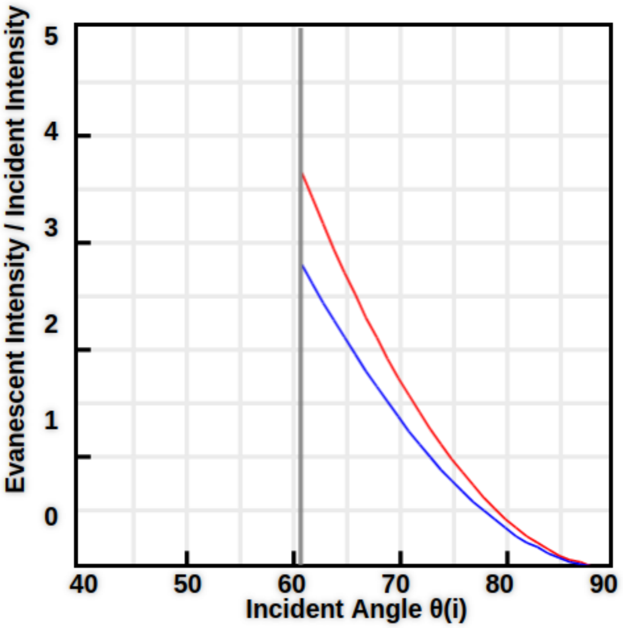


FIGURE 2.4 – Polarized Light Evanescent Intensities, where the critical angle $\theta_c = 60°$. [AWD]

When the incident angle increases from $\theta_c \sim 60°$ to $\sim 75°$, I_0 can be 4 to 1 times stronger than the intensity of the incident wave.

2.4 Basic principles of Structured Illumination microscopy

SIM uses spatial structured illumination light in a wide field microscope. By the inference of the excitation light with the emission light from the illuminated sample, higher frequency information is encoded, by the produced analogous Moiré fringes effect, and is recorded in the images. It improves the diffraction limit by a factor of two. In the following, we introduce firstly the diffraction limit explained on the reciprocal space. Then we briefly present the principles of SIM.

2.4.1 The translation of the diffraction limit in frequency space

In the reciprocal space (or the frequency space), the resolution limit is translated to the observable region (Fig. 2.5 left). The Optical Transfer Function (OTF) is the Fourier transformation of the PSF (Fig. 2.5 right). The observable region corresponds to all frequencies where the OTF has non-zero values, thus in the support of the OTF. Note k_0 the maximum non-zero value of the OTF, then k_0 is the limited observable spatial frequency.

High-frequency information residing outside of the support of the OTF can not be observed.

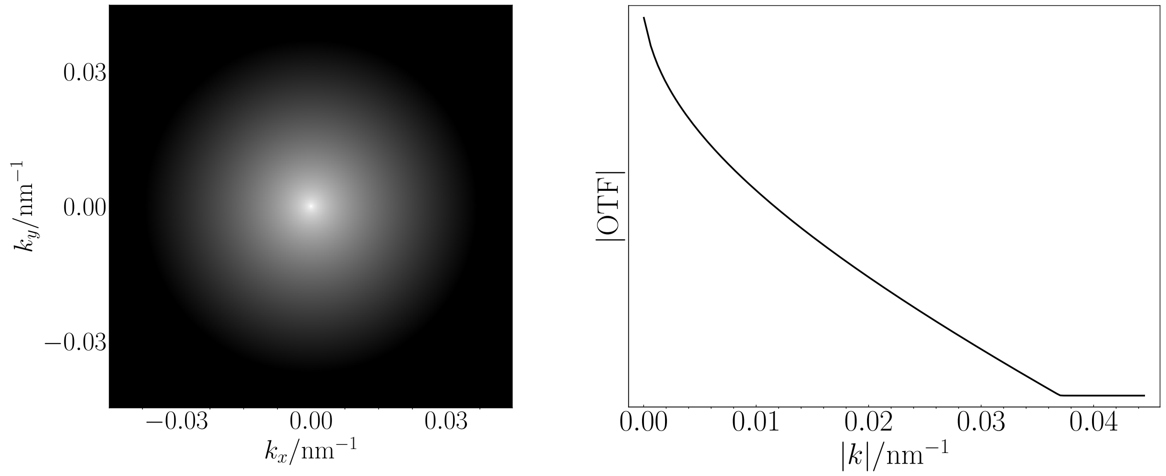


FIGURE 2.5 – Left: Airy disk in the frequency space; Right: The OTF, when setting $\lambda = 509nm$ (GFP emission wavelength) and $NA = 1.5$ (theoretical maximum value). Plotted according to the PSF equation in [ZZOM07]

Therefore the microscopy acts as a low-pass filter, only allowing the observation of frequencies with a magnitude less than or equal to that of k_0 as illustrated in Fig. 2.6 (b).

The classical limit of resolution in the microscopy translates into frequency space, defining a maximum observable spatial frequency k_0 given by:

$$k_0 = 2\mathrm{NA}/\lambda_{\mathrm{em}}, \tag{2.5}$$

where NA is the numerical aperture of the objective, and λ_{em} is the average observed emission wavelength.

2.4.2 The principle of SIM

As stated in the previous section, the observable region of spatial frequencies is within k_0 in conventional microscope. When structured illumination is used, it interacts with the sample and produce an effect similar to the generation of Moiré fringes (Fig. 2.6 (a)), which is the multiplication of the illumination light and the sample. It can be seen that such Moiré fringes can be coarser than either of the original patterns. The higher spatial frequencies in the sample which are beyond the diffraction limit can therefore be shifted inside the observable region.

Let us see the theory in the way of Fourier transform. The observed fluorescent emission, $E(\mathbf{r})$, is primarily determined by three factors, the local density of the fluorophore $D(\mathbf{r})$, multiplied by the local intensity of excitation illumination $I(\mathbf{r})$, which is then convolved with the PSF

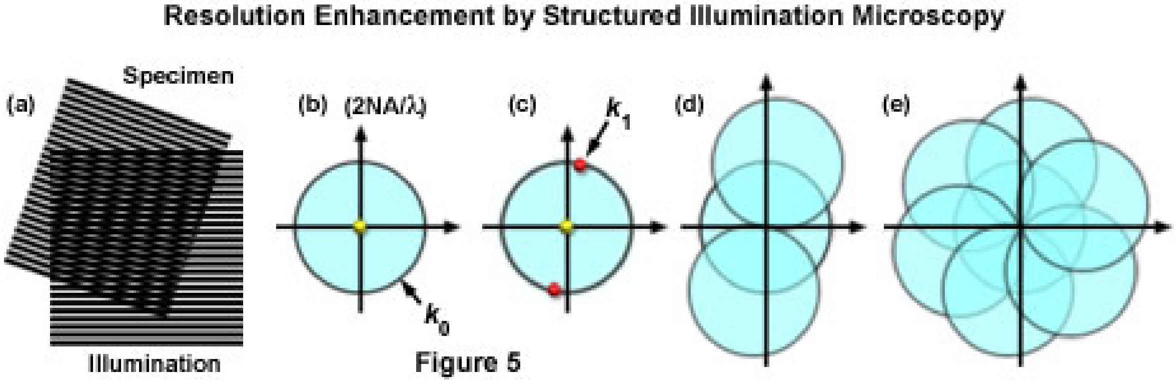


FIGURE 2.6 – **The principle of SIM. From [Gus00]** (a): The interference between the illumination and the specimen produce the analogous Moiré fringes. The high spatial frequency information cannot be observed due to diffraction limit is encoded in the Moiré fringes; (b): The spatial frequency can be observed only if it falls into the circle region with raduis k_0 (2.5); (c):A sinusoidal structured light with parallel stripes has three non-zero points in the reciprocal space; (d): Three circle regions, centred by three points shown in (c) and with radius k_0, can be observable (2.10); (e): Repeat (d) for two more times by shifting the direction of the stripes, normally $120°$ each time, an observable region twice as large as (b) can be obtained.

of the optical system PSF$(\mathbf{r})$,

$$E(\mathbf{r}) = [D(\mathbf{r}) \cdot I(\mathbf{r})] \circledast \text{PSF}(\mathbf{r}). \tag{2.6}$$

The Fourier transform of (2.6) is given by:

$$\tilde{E}(k) = [\tilde{D}(k) \circledast \tilde{I}(k)] \cdot \text{OTF}(k), \tag{2.7}$$

where $\tilde{f}(\cdot)$ denotes the Fourier transform of $f(\cdot)$ and OTF(k) is the Fourier transform of PSF$(\mathbf{r})$. As it is explained before, $\tilde{D}(k) \circledast \tilde{I}(k)$ can only be observable within the non-zero support of OTF(k) (Fig. 2.5), that is a circle region of radius k_0 (Fig. 2.6 (b)). The same circle defines the set of patterns that it is possible to create in the illumination light. Structured illumination does not alter this physically observable region, but it moves information into the region from the outside, and thereby makes that information observable.

If $I(\mathbf{r})$ is uniform, therefore $\tilde{I}(k)$ has trivial value, then the observed fluorescence frequency $\tilde{E}(k)$ depends directly on $\tilde{D}(k)$. However if the illumination $I(\mathbf{r})$ is structured, *e.g.* sinusoidal pattern of parallel stripes used in the original of SIM, given by

$$I(\mathbf{r}) = I_0[1 + \cos(k_0 \cdot \mathbf{r} + \phi)], \tag{2.8}$$

its Fourier transform has three non-zero points (Fig. 2.6 (c)):

$$\tilde{I}(k) = I_0[\delta(0) + \delta(k + k_0)e^{i\phi} + \delta(k - k_0)e^{-i\phi}]. \tag{2.9}$$

The observed frequency in images following:

$$\tilde{E}(k) = I_0[\tilde{D}(k) + 0.5\tilde{D}(k + k_0)e^{i\phi} + 0.5\tilde{D}(k - k_0)e^{-i\phi}] \cdot \text{OTF}(k) \tag{2.10}$$

Therefore, the image that is seen through the microscope contains, not only the normal image as in conventional microscope, but also higher frequency information, which, in the reciprocal space, is moved from the outside into the observable region (Fig. 2.6 (d)).

If the frequency of the structured illumination is chosen as high as possible, i.e., k_0 as given in (2.8), then it is possible to extract information out to double the resolution limited by the diffraction of light, in the direction of the stripe (2.6 (d)). By repeating this three times with the pattern orientated in different directions, one can gather essentially all the information within a circle twice as large as the physically observable region (Fig. 2.6 (e)).

2.5 Sources of noise

We introduce here the common sources of noise during the image acquisition, in order to include them in the image processing and dynamics analysis. The noise sources in the microscope imaging system are listed in [ZZOM07]. The principle instruments used for fluorescence detection in TIRF is Charge-Coupled Device (CCD). The noise model can be approximated by a mixed-Poisson-Gaussian (MPG) model, combining the Fluorescence photon noise, dark noise, and readout noise.

The final detected signal S can be given as follows

$$S = \alpha(N_p + N_d) + N_r, \tag{2.11}$$

where α is the overall gain of the camera, N_p is the number of photon which follows a Poisson distribution with intensity parameter λ, $N_p \sim \mathcal{P}(\lambda)$, N_d is the dark noise due to the kinetic vibration of silicon atoms in the CCD substrate which will liberate electrons or holes even when no incident fluorescence photon is present and $N_d \sim \mathcal{P}(\lambda_d)$, N_r is the readout noise primarily originates from the imperfectness of the output amplifier during the process of converting charge into a voltage signal, which is described by a normal distribution, $N_r \sim \mathcal{N}(\mu, \sigma^2)$.

The fluorescence photon noise, due to the stochastic nature of the photon emission, is inherent in all optical signals. If we dispose of a high photon flux, which is the case for WFFM and TIRF, N_p will be asymptotically normally distributed with both the mean and the variance equal to λ.

In [TLW02], the precision of localization under two limit conditions, the photon shot noise-limited case and the background noise-limited case is discussed. The photon shot noise-limited case occurs when the noise in each pixel is dominated by photons originating from the particle being localized. The background noise-limited case occurs when the noise is caused by other sources, e.g. readout noise, dark noise, extraneous fluorescence in the microscope (caused for example by dust), and cellular auto-fluorescence.

TIRF gives high signal-to-noise with little recorded out-of-focus fluorescence and it has a rather high flux of photon. Therefore, we consider the photon noise-limited case.

Each photon collected in the image gives a measure of the position of the object, and the position error of each measurement is identical as the Standard Deviation (SD) of the PSF of the microscope. The estimated position is the average of the positions of the individual detected photons, and the variance of localization is given by

$$\mathbb{E}((\Delta x)^2) = \frac{s^2}{N}; \tag{2.12}$$

where s^2 is the variance of the PSF and N the number of photons N. This noise model will further be used in the next chapter in the modeling of the dynamics by diffusion processes.

2.6 Observation of the dynamics of the protein MreB during active bacteria cell growth

In this section, we discuss the practical issues concerning the observation of the dynamics of the MreB during active bacteria cell growth.

A detailed guidebook for MreB observations in TIRF microscopy is available [Cor+20]. Materials and Methods for Bacteria cell growth, Agarose pad preparation and slide mounting, and TIRF imaging are presented. The challenges and practical issues are also discussed. The TIRF images used in this thesis are acquired in the ProCeD team, Micalis Unit (INRAE, Jouy-en-josas, France). The modalities for TIRF setting is described in [Bil+17]. MreB dynamics movies are acquired at $\Delta t = 1s$ during 2 minutes, with an exposure time $t_E = 0.1s$. The final pixel size was $64nm$.

SIM-TIRF imaging was performed at the Advanced Imaging Center (AIC) (Janelia Research Campus, VA, USA), as described in [Li+15]. SIM reconstructed images and average TIRF images were processed from nine raw acquisitions (three translations × three rotations). Each image is exposed during $10ms$ (total exposure time $90ms$) and the time interval $\Delta t = 1s$. The duration of movies is $0.5 - 1$ minute. The experimental resolution limit at $488nm$ was estimated to be $114.37 \pm 16nm$ (n = 198 beads), using $40nm$ fluorescent beads.

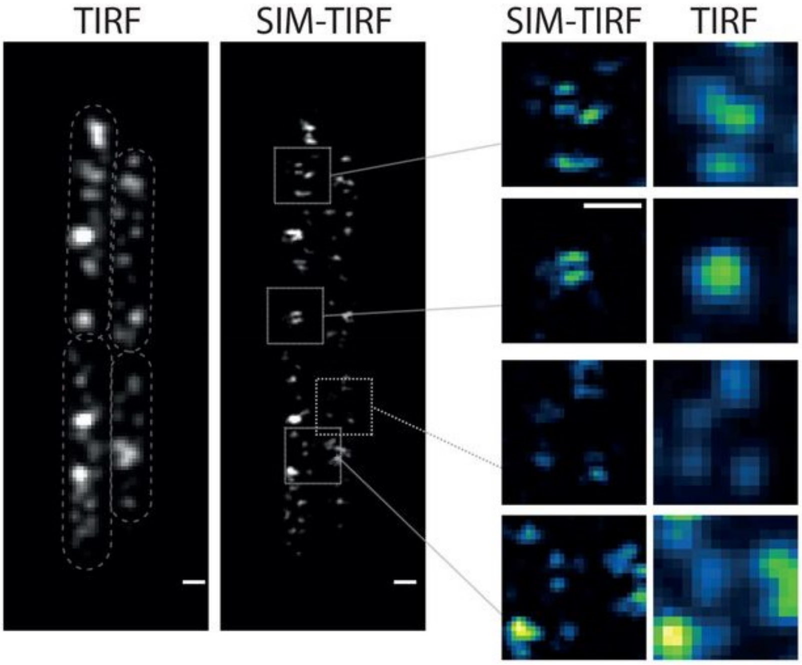


FIGURE 2.7 – A comparison of images aquired by TIRFM and SIM-TIRFM [Bil+19].

The images of MreB, acquired by TIRFM and SIM-TIRFM, respectively, are presented in Fig. 2.7. It can be noticed that, in TIRFM the MreB proteins shows up as spherical spots, due to the diffraction limit. In SIM-TIRF instead, several separated elliptical spots are observed.

There exist other possible microscopy methods suitable for MreB imaging, which have the potential to break certain limits of the current acquired MreB image by TIRFM and SIM. Especially, we would like to have a complete view to observe the dynamics of MreB on the whole membrane, at high spatial (lateral and axial) and temporal resolution.

TIRF imaging, thanks to the illumination in a small section, benefits from the low phototoxicity and photobleaching. At the same time, it sets an important limit of TIRFM, the incomplete observation view. TIRF measurements can be further exploited using the relationship between the incident angle and the penetration depth of the evanescent wave (2.4), thus adding a third dimension to TIRF imaging. A reconstruction algorithm of high-resolution 3D volume from multiangle TIRF images stacks has been developed by Boulanger et al. [Bou+14]. The Multi-angle TIRF (3D) is able to achieve $50nm$ axial resolution over a range of $800nm$ above the coverslip. At the same time, it has reduced phototoxicity and photobleaching compared to WFFM.

The Lattice Light Sheet microscopy (LLSM) can also be considered for 3D MreB dynamics observation, as it is characterized by very high imaging speed, high contrast, and good optical penetration depth. The recent developed LLSM is improved to reach a quasi-isotropic resolution $230nm \times 230nm \times 370nm$, thus increasing the volumetric resolution [Che+14a]. It can be combined to SIM to further increase the resolution.

In order to analyse the dynamics of MreB, to test whether it performs free diffusion, the temporal scale is also critical. The 3D-SIM even can achieve super resolution volume acquisition [Sch+08], and therefore allows the observation of the whole membrane surface of *B. subtilis*, has rather low temporal resolution of $(\Delta t \sim 1s - 1m)$ [Sch+19].

Multi-angle TIRF microscopy and LLSM devices installed in a collaborative lab, PICT - IBISA, Institut Curie, have been tested to acquire the dynamics of MreB. However, due to practical challenges of microscope and obstacles related to the sanitary (COVID-19) crisis, the Multi-angle TIRF and LLSM acquisitions was suspended. The 3D-SIM device is recently installed in the ProCeD team, Micalis Unit (INRAE, Jouy-en-josas, France), yet, some technical issues (e.g., photobleaching and sensitivity) prevent us from obtaining 3D-SIM images with enough temporal frequency and observation duration.

Conclusion

In this chapter, the principles of TIRFM and SIM are reviewed. Several information should be kept in mind to continue the reading of the following chapters:

Due to the total reflection and the evanescent field of around $100nm \sim 200nm$, only around one third of the bacteria surface can be observed. The recorded trajectories are the projection of the trajectories along the surface of the 3D cell membrane.

The resolution of SIM is around $\sim 110nm$, which allows to observe unique MreB filament, whose average length is around $\sim 170nm$ [Bil+19]. Therefore, the gain in resolution by combining SIM with TIRF sheds light on the understanding of the dynamics and spatio-temporal distribution of MreB. However, as several raw images are needed to reconstruct an image, the movie duration is limited to $0.5 - 1$ minute due to the photobleaching of the fluorophores.

Several sources of noises in fluorescent imaging system are introduced. The localization errors in the shot noise-limited case is given by a simple formula (2.12), which allows the quantification of errors and is further included in trajectories analysis procedures presented in Chapter 3.

CHAPTER 3

STATISTICAL ANALYSIS OF THE DYNAMICS OF THE PROTEIN MREB DURING CELL WALL SYNTHESIS

Motivation

The environment inside living cells is very complex. Experimental observations have revealed that biomolecules and molecular aggregates inside cells perform different types of movements. Several of them differ significantly from pure diffusion. The study of the trajectories of molecular aggregates can help to decipher different biophysical processes such as free diffusion, constrained dynamics in crowded environments, and dynamics mediated by molecular motors. Several types (classes) of dynamic biophysical processes have been identified over the past years. These processes are often characterized by a few interpretable features (e.g. diffusion coefficient, velocity). These features allow us to compare the individual interaction among particles and their environment, within the same cell or between different cells.

In this chapter, we aim to gain a detailed description of the heterogeneous dynamics of a target protein, MreB, which is known to be moving along the surface of the inner membrane and performing different regimes of motions during the growth of the rod-shaped bacteria cell. Two different points of view are considered to study the dynamics. The first one consists in analyzing individual trajectories of particle, including the mode of motion and the biophysical parameters of the motion, e.g., the drift and the diffusion coefficient. The second one aims at studying the dynamics of particles in the neighborhood of a position point and further characterizing dynamical events, e.g., the shape and distribution of potential wells or confinements areas.

In the first section of this chapter, we review some classical diffusion models and measurement errors, including static localization error induced by the limit of spatial-resolution of optical microscopy and the motion blur due to the finite exposure time during image acquisition. **Section 2** is dedicated to the analysis methods of individual trajectories. We first review the main approaches for trajectories classification into several categories of diffusion: free, confined, and directed diffusion. Furthermore, we present the advantages and limitations of competitive methods used to estimate the underlying biophysical parameters. **Section 3** is devoted to analyzing the dynamics of MreB aggregates analyzed with the aforementioned methods. The dynamics of MreB, which take place in the vicinity of the inner membrane surface, are acquired with SIM-TIRFM. It is worth noting that the analysis is performed on partially observed dynamics, as only the particles of the evanescent field can be observed with this imaging technology. It is also noted that a significant bias is induced by the planar projection of the 3D dynamics onto a 2D plane. In **Section 4**, the local properties of the dynamics field are examined. We focus on the local drift and the particle diffusion coefficient. Assuming that the dynamics follow a certain diffusion model, a method for reconstructing the 3D surface (in our specific case, the recovery of the radius of the bacteria cell) from a planar projection of the 3D trajectories, is proposed.

3.1 Diffusion models and sources of measurement errors

Thanks to the development of fluorescence labeling, super-resolution microscopes, and single-particle tracking techniques, research scientists are equipped with many advanced tools to take a glimpse into the very complex environment in living cells. It is well established that the characteristics of the motions of biomolecules present a great heterogeneity in living cells. In this section, firstly, we describe the main diffusion types usually considered in biophysics and related stochastic models. Beforehand, it is important to know that the input data (i.e., images, trajectories) are corrupted by several sources of errors during the acquisition and analysis of fluorescence images. We discuss here two main sources of errors: the limited localization precision due to light diffraction and motion blur caused by the exposure time and particle diffusion.

3.1.1 Diffusion models

From a probabilistic point of view, a diffusion process X_t is a continuous time process that possesses the Markov property and for which the sample paths are continuous. The Markov property states that if we know the present state of the process, the future behavior of the process is independent of its past, described by the σ-field $\mathcal{F}_t$:

$$P(X_{t+s} \in A|\mathcal{F}_t) = P(X_{t+s} \in A|X_t). \tag{3.1}$$

Diffusion process satisfies three key conditions [KT81]:

$$\lim_{\Delta\to 0}\frac{1}{\Delta}P(||X_{t+\Delta}-X_t||>\epsilon|X_t=x)=0,\forall\epsilon>0, \tag{3.2}$$

$$\lim_{\Delta\to 0}\frac{1}{\Delta}\mathbb{E}(X_{t+\Delta}-X_t|X_t=x)=\mu(x,t), \tag{3.3}$$

$$\lim_{\Delta\to 0}\frac{1}{\Delta}\mathbb{E}(||X_{t+\Delta}-X_t||^2|X_t=x)=\sigma^2(x,t). \tag{3.4}$$

The first condition (3.2) states that large displacements are very unlikely over infinitesimal time intervals, while conditions (3.3) and (3.4) characterize the mean and the variance of the infinitesimal displacements and confirm the existence of the limits.

Brownian motion or free diffusion is a fundamental diffusion process in biophysics. It was first observed in 1827 by the British botanist, who observed the motion of particles in pollen grains in water. It was explained many decades later by A. Einstein in [Ein05], who demonstrated that apparent pollen motion was induced by water molecule displacement.The related probability theory was derived by N. Wiener in 1923. In the subsequent years, it has been established that Brownian motion was equivalent to a Wiener process defined as follows:

Definition 1 *A stochastic process B_t is a* ***Brownian motion*** *if it satisfies the following properties:*

- *B_t is a process with independent increments: for $t>s$, B_t-B_s is independent of the field $\mathcal{F}_s$ generated by the history of the process $B_m, m\in[0,s]$ until time s.*
- *For all $t>s$, $B_t-B_s\sim\mathcal{N}(0,t-s)$.*
- *The paths of B_t are almost surely continuous and nondifferentiable.*

Moreover, Langevin justified with another modeling approach the motion of suspended particles [Lan08]. Particle motion is characterized by the $d-$dimensional so-called Langevin equation:

$$m\frac{dv(t)}{dt}=-\zeta v(t)+L(t), \tag{3.5}$$

where v is the velocity of the particle, m denotes the mass, $\zeta>0$ is the friction coefficient and L is a random force resulting from the collisions with the surrounding particles. The process received a more thorough mathematical examination several decades later by Uhlenbeck and Ornstein [UO30] and Chandrasekhar [Cha43].

A recent presentation of Einstein's approach and Langevin's approach in bioimaging is given in [BVK20]. In summary, it is established that the solution of the overdamped Langevin equation defined as follows

$$\frac{dX_t}{dt}=-\frac{1}{\zeta}\nabla U(X)+\sqrt{2D}R(t) \tag{3.6}$$

where $U(X)$ is the particle interaction potential and $R(t)$ is a standard Gaussian variable, is also known as Brownian motion. Note that Brownian motion by Einstein (equivalently driftless

Wiener process) is nowhere differentiable. However, Brownian motion characterized by the overdamped Langevin equation is differentiable. Nevertheless, both definitions are consistent in the sense that the MSD is asymptotically linear with time:

$$\rho(t) = \mathbb{E}(||X_t - X_0||^2) = 2dDt, \tag{3.7}$$

where $D = \frac{k_B T}{\zeta}$ denotes the diffusion coefficient, k_B is the Bolzmann constant, T is the temperature and ζ is the friction constant as introduced in (3.5).

In [BVK20], other diffusion processes are reviewed, well-grounded in Einstein's and Langevin's theories, for modeling respectively sub-diffusion and super-diffusion in biophysics. The most intuitive way to distinguish sub- and super- diffusions from free diffusion is still performed via MSD analysis (see details in Sections 3.2.1.2 and 3.2.2.1). If the MSD is sub-linear with time t then the diffusion is regarded as sub-diffusion. On the contrary, if the MSD is over-linear with time t, then the diffusion is accordingly super-diffusive. Among existing models, the most popular sub-diffusion models are shortly described below: Ornstein-Uhlenbeck (OU) process [UO30], fractional Brownian motion (fBm) (with $0 < H < 1/2$ where H is the Hurst index), recommended to characterize dynamics in crowded cellular environments [MVN68], Continuous Time Random Walk (CTRW), for analyzing trapping areas and nanodomains within the cell ([SM75]) or modeling the axon growth cone [DZK17], Generalized Langevin Equation (GLE) ([Zwa01], [Kou+08]), for taking into account the viscosity properties of the environment. Super-diffusions are mainly defined as Brownian motion with drift and fBm (with 1/2<H<1).

Ornstein-Uhlenbeck (OU) process

A unified approach for representing diffusion processes is to consider that they are solutions of Stochastic Differential Equation (SDE). For example, the OU velocity process is the solution of the SDE:

$$mdv(t) = -\zeta v(t)dt + \sigma dB_t, \tag{3.8}$$

where m, v and ζ represent the mass and the velocity of the particle and the friction coefficient, respectively, as in (3.5) and B_t denotes Brownian motion defined in Def. 1.

In mathematics, the OU process is defined by the following stochastic differential equation:

$$dX_t = -\lambda(X_t - \theta)dt + \sigma dB_t \tag{3.9}$$

where X_t is the position of the process, and $\lambda > 0$ reflects a restoring force directed towards the long term average θ.

Fractional Brownian motion (fBm)

fBm is defined by the covariance function of $\{B_t^H\}$, a continuous zero-mean Gaussian process with stationary increments:

$$cov(B_t^H, B_s^H) = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}), \tag{3.10}$$

where $0 < H < 1$ denotes the Hurst index. It is noticed that in biophysics, although fBm with $1/2 < H < 1$ may represent super-diffusion, fBm is mainly used to characterize sub-diffusion (i.e., for $0 < H < 1/2$).

Brownian motion with drift (Bmd)

The Brownian motion with drift (Bmd) is solution of the d-dimensional SDE:

$$dX_t = vdt + \sigma dB_t, \tag{3.11}$$

where v is the drift and σ the standard deviation, B_t is the standard Brownian motion such that the diffusion coefficient $D = 2\sigma^2$.

For the past 15-20 years, we have the opportunity to observe and analyze the dynamics of biomolecules in live cells with advanced imaging technologies. They can produce high spatial resolution image sequences and particle tracks, but different sources of noise and errors inherently corrupt the data. In the next section, we quantify the errors of measurements due to the light diffraction limit and exposure time duration.

3.1.2 The sources of errors inherent to fluorescence microscopic images acquisition

Localization error due to the diffraction limit

The main source of uncertainty is the localization noise. Usually, the localization uncertainty of a particle located at (x, y) in a 2D plane can be characterized by a Gaussian probability distribution, with standard deviation σ_x (resp. σ_y) depending on the standard deviation s_0 of PSF and the number of photons N. If we assume $\sigma_{\text{loc}} = \sigma_x = \sigma_y$, we have

$$\sigma_{\text{loc}} = \frac{s_0}{\sqrt{N}} \tag{3.12}$$

which represents the *static localization uncertainty*. In other words, σ_{loc} represents the localization uncertainty for an immobile particle [TLW02]. Consequently, the measured position of the particle at time t is written $X_t = \tilde{X}_t + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma^2_{\text{loc}})$ and $\tilde{X}_t$ is the physical real position of the particle.

Motion blur

The second source of uncertainty is due to "*Motion blur*," where the average position of a particle is estimated over the exposure time t_E. We suppose that a particle moving in one dimension according to Brownian motion with diffusion coefficient D is captured by a camera. During the time interval Δt, the camera shutter may be opened during duration t_E, $t_E \leq \Delta t$. In [Ber10], Berglund defines the ***motion blur coefficient*** as:

$$M = \frac{1}{\Delta t} \int_0^{\Delta t} S(t)[1 - S(t)]dt, \tag{3.13}$$

where $S(t)$ is the illumination percentage occurring before time t (between the beginning and the end of a frame), during the interval of time Δt. $S(t) = \int_0^t s(t')dt'$, where $s(t)$ is the shutter function. The value of M lies between 0 and 1/4, depending on the camera exposure mode and duration. M equals zero (no motion blur) if and only if the shutter function $s(t)$ consists of a single vanishing peak at any point within the frame. The coefficient M reaches its maximum value $1/4$ when there is a double-pulse sequence at the start and the end of the frame. In the common experimental case of a uniform exposure of duration $t_E \leq \Delta t$, setting $S(t) = \frac{t}{t_E}$ in (3.13), it yields to

$$M = \frac{1}{6} \frac{t_E}{\Delta t}. \tag{3.14}$$

It is shown in Section 3.2.2 that the localization noise and motion blur induce correlations between the observed displacements of the moving particles. Both static and dynamic localization errors σ^2_{loc} and M are used further to elaborate several different estimators (Optimal Least Squared Fitting (OLSF), Maximum Likelihood Estimation (MLE), Covariance-based Estimator (CVE)) of particle dynamics.

3.1.3 Discussion

In addition to the two main sources of errors under concern, the errors induced by the spot detectors ([Des17], [DG+19]) or particle trackers ([Jaq+08], [Yan+12]) should not be ignored either. The light diffraction makes that the particle in the image appears not as a dot but rather as a blurred spot. During the spot detection step, when two particles are located too close to each other (i.e., when the distance is less than the size of Airy disk [Air35]), two fluorescent spots overlap. This induces confusion for the algorithm to detect the two spots. The detection errors are false negatives (the algorithm fails to detect an existing particle) or false positives (the algorithm detects a fake spot while there is no particle). Falsely detected particles usually yield short trajectories. However, particles along a trajectory may not be detected in all frames, generating undesirable several sub-tracks and gaps. To re-connect these sub-tracks (or tracklets), several tracking algorithms include a gap-closing procedure ([Jaq+08], [Che+14b]).

The success rate of trajectory estimation is closely related to noise level (or Signal to Noise Ratio (SNR)) in images. The *peak SNR* value is generally used to assess the amount of noise of images and requires no information about the acquisition device. It is defined as follows [Che+14b]:

$$pSNR(x,y,z) = \frac{I(x,y,z)}{\sigma_n}, \tag{3.15}$$

where $I(x,y,z)$ is the intensity at point (x,y,z) and σ_n the empirical estimation of the standard deviation of the noise. The impact of the noise can be evaluated by simulating blurred data, applying the spot detection and the tracking algorithm before applying the trajectories analysis procedures.

3.2 Methods for individual trajectories analysis

In this section, we review several competitive methods developed for the analysis of individual trajectories. Several features, including length and direction of particle trajectories, diffusion types, parameters (diffusion coefficient, drift), and diffusion regime transitions, help decipher the mechanisms of various cellular processes and the roles of proteins involved in these processes. The problem amounts to identifying the right diffusion model and the optimal estimated parameters for each trajectory. We introduce several methods for classifying different diffusion motions and estimating the parameters of models under concern (diffusion coefficients, drift...). MSD is probably the baseline method used to address this challenge. More sophisticated methods based on MLE, Bayesian estimation, and learning-based approaches have been proposed to overcome the difficulties faced by the MSD approach.

In the classification task, the parameters, e.g., the diffusion coefficient D, must be preliminarily estimated. On the flip side, to estimate the parameters, especially when using parametric methods (e.g., MLE), knowing the diffusion mode is required. In practice, one needs to keep in mind that those two problems are interwoven.

The presence of measurement errors adds more difficulties to the problem. On the one hand, the errors degrade the estimators of D; On the other hand, the quantification of the errors due to measurement depends on the diffusion coefficient D. Finding a way out of this interwoven problem requires different shortcuts and ingenuity, depending on the experimental conditions and prior knowledge. In the case of the dynamics of the MreB observed with SIM-TIRFM, we present a dedicated solution in Section 3.3.

In what follows, firstly, the classification methods, using the canonical MSD method and a new statistical hypothesis test, called Testing HypOtheses for diffusion TricHotomy (THOTH), are presented (Section 3.2.1). Secondly, several parameter estimation methods, including the MSD based OLSF, MLE, and CVE, are presented in Section 3.2.2. The measurement errors is taken into account in the definition of estimators.

3.2.1 Comparison of classification methods

In this section, we provide an overview of existing methods used to determine modes of motion for single-particle trajectories. Then, in the case of MreB trajectories, we focus on three non-parametric classification methods, two of them are based on MSD and another one is THOTH.

3.2.1.1 Bayesian methods and Deep Learning

In [TM13], Türkcan and Masson proposed a computational approach able to determine the motion model that best fits the observed trajectories. The method is based on Bayesian inference to calculate the *a posteriori* probability given an observed trajectory. Information theory criteria, such as the Bayesian information criterion (BIC), are used for selecting the most appropriate model taken in a set of pre-specified models, including free Brownian motion, confined motion in 2nd or 4th order potentials, and more complex motion models able to represent for instance phenomena like active transport and hopping. The main limitation of the method is related to the number of trajectory points required for reliable estimation. At least 20 points are actually needed to discriminate free diffusion from confined motions, and at least 50 points are needed to determine confined motions in 2nd order (even more points with higher order).

The Bayesian approach in [TM13] is based on the exhaustive comparison of several parametric models, and assess the goodness of fit of the observed trajectory to each parametric model. Unlike [TM13], the statistical test approach called THOTH and proposed by [BKV18] is a non-parametric method. It can discriminate at once free diffusion from sub- and super- diffusion. It is a non-parametric method, therefore it does not determine to which sub- (or super-) diffusive model the data fit the best, knowing that there are different models for sub- (or super-) diffusion due to different physical mechanisms. One of the advantage of the hypothesis test is that it performs well for short trajectories from 10 points and beyond. About the calculation complexity, it needs to be mentioned that the optimization algorithm to find the Maximum *A Posteriori* in Bayesian approach is very time consuming compared to the simplicity of the method proposed by [BKV18].

Meanwhile, machine learning methods have been investigated to classify particles trajectories into categories of motion. Deep learning method, to our knowlege, was first introduced by [Gra+19a] for single-particle-diffusion analysis and classification of single-particle trajectories into three sub-groups: Brownian motion, fractional Brownian motion and Continuous Time Random Walk. A network architecture is also used to estimate the Hurst exponent for fractional Brownian motion and the diffusion coefficient for Brownian motion, on both simulated and experimental data. It turns out that these networks achieve greater accuracy than time-averaged MSD analysis on simulated trajectories while only requiring as few as 25 time points. A set of multi-tracks networks were also designed to tackle a problem of high practical importance of experiments in which only numerous very short trajectories are available (i.e., $\sim$10 time points) rather than a single long trajectory. Therefore deep learning is an appealing approach with high potential for identifying complex models in large quantities of data, with a relative simplicity of implementation (once trained), inference speed, and robustness. In addition, graph models have been developed to describe single-particle trajectories or other sub-cellular objects ([Ver+21], [Gra+19b]). Recently, combining with deep learning networks, the Graph Neural Networks (GNN) have been widely used to infer the underlying biophysical properties of the studied objects ([ZCZ20], [Art+19]).

In the scenario of MreB protein dynamics observed in TIRFM, the average number of points in a trajectories is around 10. The length of trajectories limits the use of Bayesian method and makes it difficult to have information of goodness of fit of trajectories to a specific parametric model. As investigated in [Bil+17], the MreB trajectories can be clustered into three groups: free diffusion, super-diffusion and sub-diffusion. Our first objective is then to classify trajectories into these three categories with non parametric model selection. Therefore, we focused on the MSD method and the THOTH to address the problem.

3.2.1.2 Classification methods based on MSD

The MSD is a widely used approach (e.g. [Fed+96], [Bil+17]) to discriminate free diffusion, sub, and super diffusion. Given a lag n, MSD (3.7) can be estimated from the average over all independent pairs of points with time lag n, or all pairs of points with time lag n:

$$\hat{\rho}_n = \hat{\rho}(n\Delta t) = \frac{1}{N-n} \sum_{i=1}^{N-n} ||X_{n+i} - X_i||^2. \tag{3.16}$$

According to diffusion regimes, the MSD curve has different shapes [SJ97]. The analytical forms of the curves of MSD versus time lag n for different modes of motion are at the origin of various classification methods. In summary, three main modes are characterized as follows in the 2D spatial domain:

$$\begin{cases} \rho(t) = 4Dt & \text{(1) free diffusion,} \\ \rho(t) = 4Dt^{\beta}, \beta < 1 & \text{(2) anomalous diffusion,} \\ \rho(t) = 4Dt + (vt)^2 & \text{(3) directed motion with diffusion,} \end{cases} \tag{3.17}$$

where $\rho(t)$ is the theoretical MSD at time lag t, v is the velocity, and D is the diffusion coefficient.

The most intuitive and the first MSD-based classification method was proposed by [Fed+96]. The main mechanism lies in the setting of a threshold on the coefficient of the exponent β, which is obtained by linear fitting $\log(\hat{\rho}_n) = \beta_0 + \beta \log(n\Delta t)$. If $0.9 < \beta < 1.1$, the particle motion is classified as Brownian diffusion; if $0.1 < \beta < 0.9$, the particle motion is classified as anomalous sub-diffusion; if $\beta > 1.1$, the particle motion is classified as super-diffusion. The particle is considered immobile if $\beta < 0.1$. However, the critical values are chosen empirically. We call hereafter this classification method by "β-MSD."

In [Bil+17], Billaudeau performed the classification of trajectories of MreB aggregates by adjusting linear and quadratic curves to MSD data points of MSD curves, that is $\hat{\rho}_n = a_0 + a_1 n\Delta t$ and $\hat{\rho}_n = b_0 + b_2(n\Delta t)^2$. The idea is based on the known result that the MSD of Brownian diffusion is linear to time and the MSD of Brownian motion with drift is parabolic to time. Classification is performed according to the goodness of fit measured by the coefficient of determination, R-squared. R^2_{diff} (resp., R^2_{dir}) is the coefficient of determination to measure the goodness-of-fit w.r.t. the diffusion (resp., directed) model. The MSD curves with a maximal value <0.05μm^2 are classified as constrained patches. The MSD curves with a maximal value >0.05μm^2, providing that the fitting of the MSD curve with a quadratic model provides a goodness-of-fit $R^2_{\text{dir}} > 0.8$ and $R^2_{\text{dir}} > R^2_{\text{diff}}$, are classified as directed motion. If the fitting of MSD data points with a linear model provides a goodness-of-fit $R^2_{\text{diff}} > 0.8$ and $R^2_{\text{diff}} > R^2_{\text{dir}}$, the MSD curve is classified as random diffusive motion. Otherwise, they are unclassified. The unclassified trajectories will be labeled "anomalous diffusion" in our simulations and experiments.

It is worth noting the MSD-based classification is controlled by parameters arbitrarily chosen, including the number of points to be used in the MSD curves and the threshold value for the coefficient β and the R-squared. Also, it is established that the variance of MSD increases with higher values of t. This property was used in [Pis+15] to design a more properly weighted least square estimation method, but with moderate improvement on experimental data. One of the noticeable drawbacks of classification methods based on MSD is the lack of precision. They have major difficulties in distinguishing free diffusion from super- and sub- diffusions.

3.2.1.3 Classification by statistical hypothesis test, the THOTH method

To overcome the limitations of MSD, a recent statistical test procedure for motion classification, based on the normalized maximal distance covered by a particle, has recently been proposed by [BKV18]. The three-decision procedure includes two complementary tests:

— Test 1. H_0 : "X_t is a Brownian motion" versus H_1: "X_t is a sub-diffusion",

— Test 2. H_0 : "X_t is a Brownian motion" versus H_2: "X_t is a super-diffusion".

and results in three possible decisions:

$$\begin{cases} \text{do not reject } H_0; \\ \text{accept that } H_1 \text{ is true;} \\ \text{accept that } H_2 \text{ is true.} \end{cases} \tag{3.18}$$

The test statistic denoted T_n is the standardized maximal distance covered by the dynamical process from its starting point:

$$T_n = \frac{S_n}{\sqrt{(t_n - t_0)\hat{\sigma}_n^2}} \tag{3.19}$$

where S_n is the maximal distance of the process with respect to its starting point X_{t_0}, defined as

$$S_n = \max_{i=1,\cdots,n} \|X_{t_i} - X_{t_0}\|, \tag{3.20}$$

with $\hat{\sigma}_n$ a consistent estimator of σ and n the number of time steps of the trajectory. If T_n is small, it means that the process stays close to its initial position during the period $[t_0, t_n]$.

The asymptotic distribution of the test statistic T_n when $n \to \infty$ under the null hypothesis is given in Theorem 3.1 in [BKV18]. Furthermore, the $p-$value of the test H_0 vs H_1 is defined as:

$$p_{1,n} = F_n(T_n).$$

Similarly, the $p-$value of the test H_0 vs H_2 is defined as:

$$p_{2,n} = 1 - F_n(T_n),$$

where F_n denotes the cumulative distribution function of T_n under H_0. However, in practice, n may be small, the asymptotic approximation of quantiles of T_n may not be accurate. A Monte Carlo estimate of the quantile $q_n(\alpha)$ was then proposed by drawing a sample from the distribution of T_n under H_0, where α is the significance level of the test. Therefore, the cumulative distribution function of T_n under H_0 can be estimated by Monte Carlo simulation,

$$\hat{F}_n(x) = N^{-1} \sum_{i=1}^{N} 1(T_n^{(i)} \leq x).$$

In [BKV18], two estimators (first and second orders) of σ are considered:

$$\hat{\sigma}^2_{1,n} = \frac{1}{dn\Delta} \sum_{j=1}^{n} ||X_{t_j} - X_{t_{j-1}}||^2, \tag{3.21}$$

$$\hat{\sigma}^2_{2,n} = \frac{1}{2dn\Delta} \sum_{j=1}^{n} ||(X_{t_{j+1}} - X_{t_j}) - (X_{t_j} - X_{t_{j-1}})||^2. \tag{3.22}$$

Note that these two estimators exploit the information of the first and second lag of the MSD $\hat{\rho}_1$ and $\hat{\rho}_2$, respectively. At this stage, the localization errors have not yet been taken into account in the calculation of estimators.

3.2.1.4 Illustration of the performance of MSD-based and THOTH classification methods on simulated trajectories

In this section, we compare the performance of the MSD-based classification methods and THOTH on synthetic data.

Data generation

Our motivation is to evaluate the potential of the aforementioned methods on synthetic data that mimic MreB dynamics.To test whether the classification methods can reliably identify free diffusion, sub-, and super- diffusion, three datasets depicting different diffusion processes are respectively simulated with Algorithm 1-3. Dataset of Brownian motion (Algo. 1, $v = 0$); The super-diffusion dataset consists of 50% Bmd (Algo. 1, $v = 0.7, \sigma = 0.3$) and 50% fBm (Algo. 2, $H = 0.85$); The sub-diffusion dataset consists of 50% OU (Algo. 3 $\lambda = 0.53$) and 50% fBm (Algo. 2, $H = 0.13$). Each dataset contains 40 trajectories. The length of the trajectories is set to $N = 30$. The time interval $\Delta t = 1s$. Note that the unit used for simulations is expressed in pixel size instead of micrometer. We manually tuned the parameters of the two MSD-based methods ([Fed+96], [Bil+17]) in order to get the best possible results.

Experimental results on simulated data

Trajectories in the Brownian motion dataset have been simulated similarly: the identical parameters $\sigma = 0.3$, length $N = 30$, $\Delta t = 1$. Nevertheless, as illustrated in (Fig. (3.1) (b)), the MSD curves calculated from trajectories depict high heterogeneity, which is due to the stochasticity of the process. The red line represents the theoretical result $\rho(t) = 4Dt$. It can be noticed that the standard deviation grows with high values of t. The coefficient β, therefore,

Algorithm 1 Simulation of a trajectory following Brownian Motion with drift

```
Input: $v, \sigma, X_{\text{ini}}, \Delta t, N$
    //$v$ the speed of the drift term, when $v = \mathbf{0}$, it corresponds to pure diffusion
    //$\sigma$ the standard deviation
    //$X_{\text{ini}}$ the initial position of a trajectory
    //$\Delta t$ time interval between two points
    //$N$ the number of points in a path
Output: the path of a trajectory of Bmd: $X$
for i=1 to N do
    if i=1 then
        $X(i) = X_{\text{ini}}$
    else
        Draw $\epsilon \sim \mathcal{N}(0, 1)$
        drift = $v\Delta t$
        variance = $\sigma\sqrt{\Delta t}\epsilon$
        increment = drift + variance
        $X(i) = X(i-1)+$ increment
    end if
end for
```

also has a higher variance as t increases. The $\beta-$MSD method provides unsatisfactory results, as it wrongly classifies Brownian motion into either sub-diffusion or super-diffusion (Fig. (3.1) (c)). The R^2-MSD method based on linear and quadratic fitting of MSD ([Bil+17]) has similar drawbacks (see Fig. (3.1) (d)). Several MSD curves depicting Brownian motion fit better the quadratic model (red dashed line) or fit neither of the two models (green dashed line) and are therefore not identified as Brownian motion. On the contrary, the THOTH achieves an accuracy of 95$\%$ (Fig. (3.1) (e)), which is consistent with the theoretical result, as the significance level α was set to 0.05.

All three methods perform correctly for the super-diffusion dataset (Fig. 3.2). In the second dataset dedicated to super-diffusion analysis, the two processes have different mechanisms: fBm ($H = 0.85$) is caused by positive correlation, and the Bmd ($v = 0.7$) results from a deterministic force and a random force induced by the thermal motion of molecules. The two motion types have different shapes (Fig. 3.2 (a)). Given the parameters of the two processes, the MSD curves of fBm are generally smaller than Bmd curves, which is illustrated by the dark red curves under the red curves (Fig. 3.2 (b)). Similarly, the test statistics T_n of fBm are generally smaller than that of the Bmd (Fig. 3.2 (e)). The majority of the T_ns is above the 0.975 quantiles, which means that the corresponding trajectories are classified as super-diffusion. The statistic summary of the classification results shown in Fig. 3.2 (f) shows that the three methods have high accuracy $\sim 95\%$ on the simulated dataset.

For the third dataset depicting sub-diffusion, obtained by simulating OU ($\lambda = 0.53$) and fBm

Algorithm 2 Simulation of a fBm trajectory using Cholesky decomposition

```
Input: $H, \sigma, X_{\text{ini}}, \Delta t, n$
    //$H$ the Hurst index
    //$\sigma$ the standard deviation
    //$X_{\text{ini}}$ the initial position of a trajectory
    //$\Delta t$ time interval between two points
    //$n$ the number of points in a path
Output: the path of a trajectory of fBm: $X$
Generate the covariance matrix $C$ following (3.23).
    $C(i,j) = \frac{\sigma^2 \Delta t}{2}(|i|^{2H} + |j|^{2H} - |i-j|^{2H}), i,j = 1, \cdots, n-1$    (3.23)
Do Cholesky decomposition $C = LL^T$, where $L$ is a $n-1$ dimensional lower triangular matrix with positive diagonal entries, $L^T$ is the transpose of L;
for i=1 to n-1 do
    Draw V(i)~ $\mathcal{N}(0,1)$
end for
Generate $Y = LV$, then the co-variance matrix of Y is C.
for i=1 to n do
    if i=1 then
        $X(i) = X_{\text{ini}}$
    else
        $X(i) = Y(i-1) + X_{i-1}$
    end if
end for
```

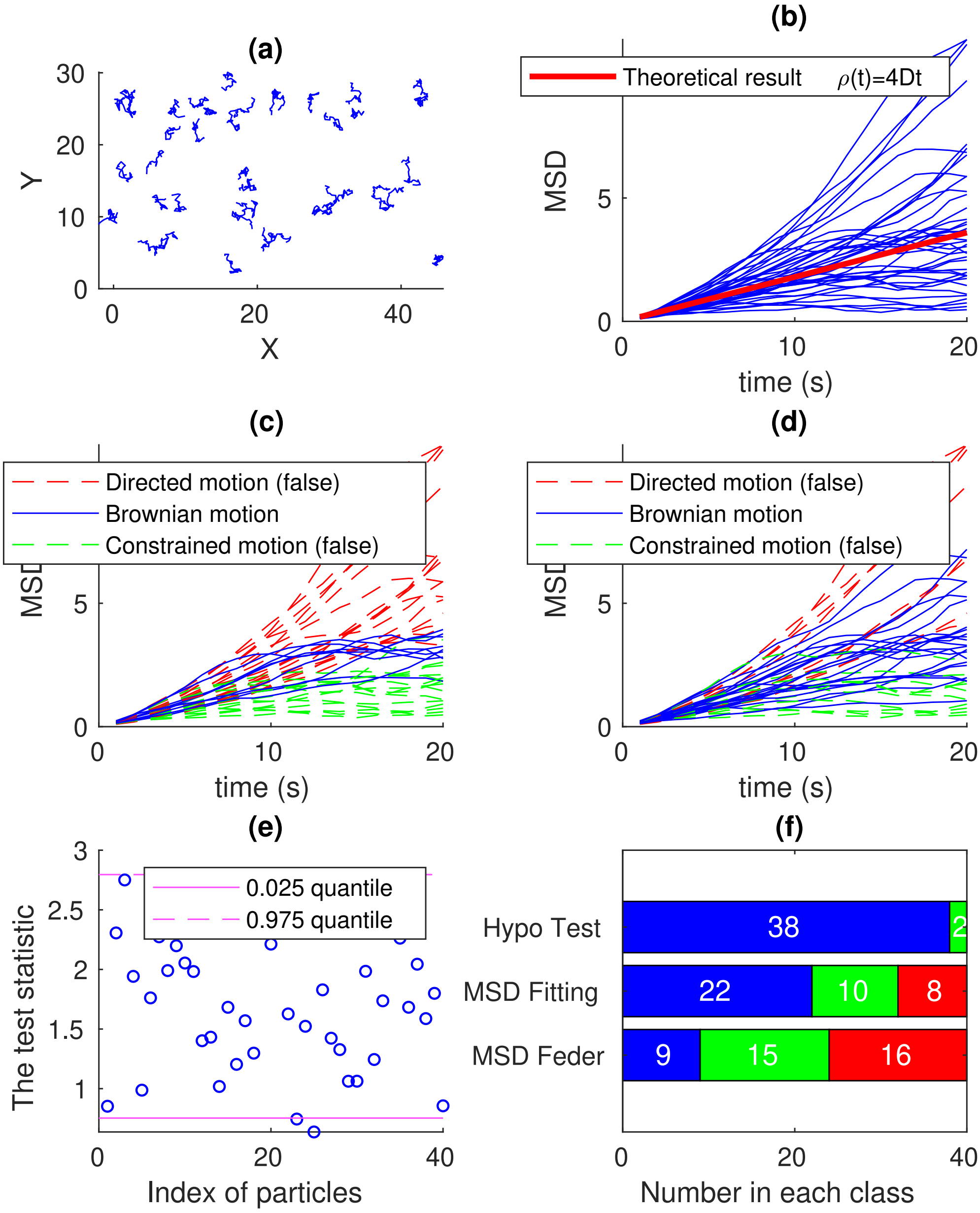


FIGURE 3.1 – **Simulation and classification for Brownian motion.** (a): Simulated Brownian trajectories, with $\sigma = 0.3$. (b): The MSD curves of trajectories. The red line represents the theoretical expectation of MSD, $\rho(t) = 4Dt$. (c): Classification result by $\beta-$MSD. (d): The result by R^2-MSD. In both (c) and (d), the dashed red (resp. green) lines correspond to the trajectories wrongly classified as super- (resp. sub-) diffusion. (e) The test statistics T_n for all trajectories and the 0.025 and 0.975 quantiles; (f) A statistic summary of the classification results. Blue color represents the number classified as Brownian motion, and the red (resp. green) colors represent the numbers classified as super- (resp. sub-) diffusion.

Algorithm 3 Simulation of a multivariate O-U process
(Notations are the same as in [Meu09]: $dX_t = -\Theta(X_t - \mu)dt + SdB_t$)

```
Input: $\Theta, \mu, S, x_0, \Delta t, n$
    n: the number of time steps
    $\Delta t$: the time step
    $x_0$: the initial position of a trajectory
Output: the path of a trajectory of a multivariate OU process $x_{courant}$
- Diagonalization of $\Theta = A\Gamma A^{-1} \Rightarrow A$, $\Gamma = (\lambda_1, \ldots, \lambda_K)$
- change of variable : $z = A^{-1}(x - \mu) \Rightarrow dZ_t = -\Gamma Z_t dt + VdB_t$
- Calculation of $V = A^{-1}S$
- Calculation of $\Phi = VV'$
- Calculation of the time varying covariance matrix $\Phi_t$ for $z$
- Choleski's decomposition : $(\Phi_t = LL^T)$
Initialisation of $z$ variable : $z_0 = A^{-1}(x_0 - \mu)$
for n-1 time steps do
    - Calculate $z_t$
    - Simulate a unit variance centered i.i.d. Gaussian vector $\tau$
    - Calculate increment $u_{courant} = esp_u + L\tau$ in the $z$ coordinates system ($esp_{u,k} = e^{-\lambda_k \Delta t} z_{k-1}$)
    - Coming back to the original coordinates system: $x_{courant} = Au_{courant} + \mu$)
end for
```

($H = 0.13, \sigma = 0.3$), the three classification methods perform equally well (Fig. 3.3). Given the parameters, the MSD curves of OU are generally under that of fBm (Fig. 3.3 (b)). Note that the $\beta-$MSD method is not always capable of distinguishing sub-diffusion from "immobile" motion, especially in the case of very small coefficients β (see Fig. 3.3 (c)). As to THOTH, the test statistic T_n are mostly under the 0.025 quantile and are consequently classified as sub-diffusion. However, the H_0 has not been rejected for several trajectories (more fBm than OU).

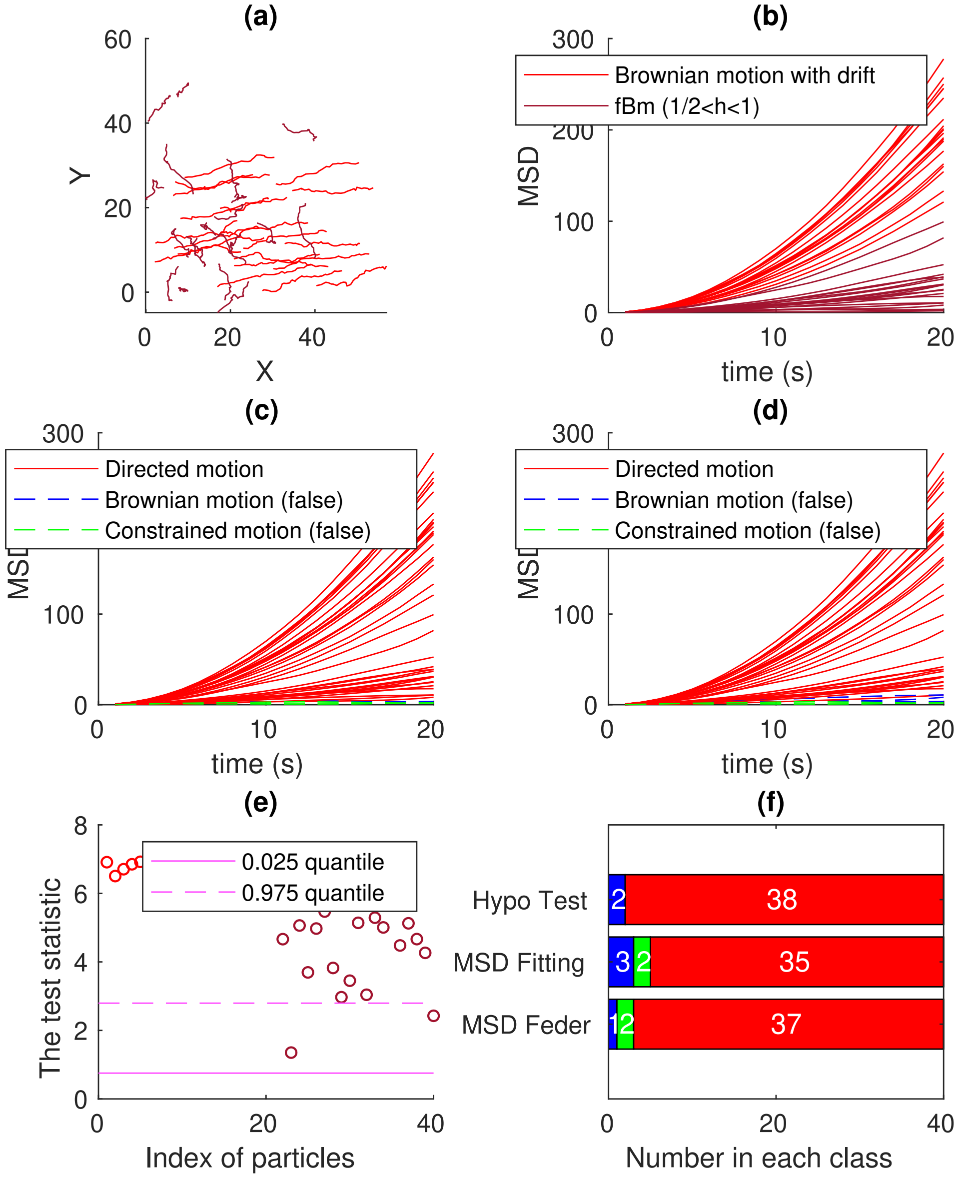


FIGURE 3.2 – **Simulation and classification for super-diffusion.** (a): The simulation of 40 trajectories, half performing Bmd ($v = 0.7$) (red) and another half fBm ($H = 0.85$) (dark red). The two kinds of motions have different shapes. (b) MSD curves of the simulated trajectories. Given the parameters of two processes, the MSD curves of trajectories performing fBm are under the MSD curves performing Bmd. (c): The classification result by $\beta-$MSD method. (d): The classification result by R^2-MSD method. In both (c) and (d), the dashed blue (resp. green) lines correspond to the trajectories wrongly classified as free (resp. sub-) diffusion. (e) The test statistics T_n for all trajectories and the 0.025 and 0.975 quantiles. (f) A statistic summary of the classification results. Blue color (resp. red and green) represents the number classified as Brownian motion (resp. super- and sub- diffusion).

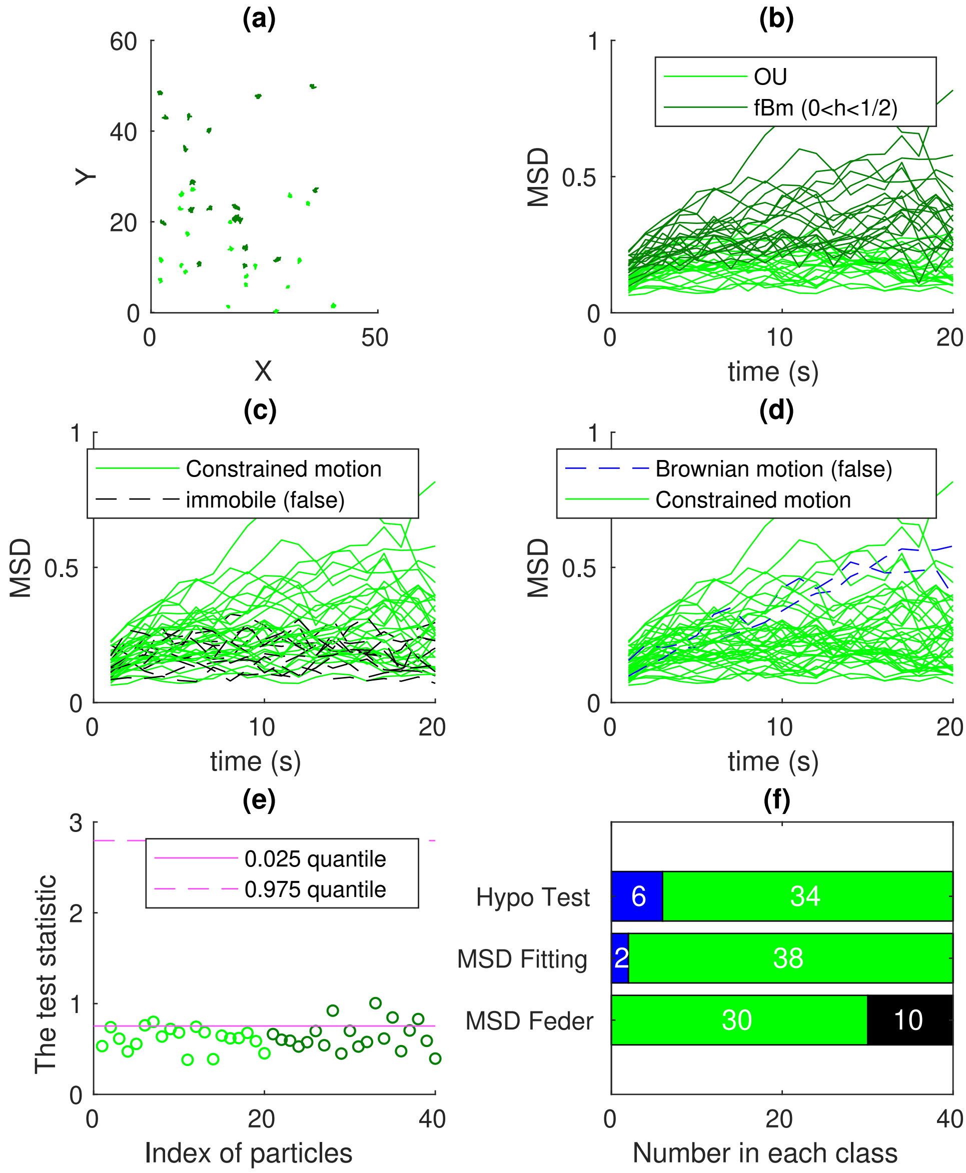


FIGURE 3.3 – **Simulation and classification for sub-diffusion.** (a): The simulation of 40 trajectories, half performing fBm ($h = 0.13$) (dark green) and another half OU ($\lambda = 0.53$) (green). (b) The MSD curves for the simulated trajectories. Given the parameters, the MSD curves of trajectories performing OU are generally under the fBm. (c): The classification result $\beta-$MSD. (d): The classification result by R^2-MSD. In both (c) and (d), the dashed blue (resp. black) lines correspond to the trajectories wrongly classified as free diffusion (resp. immobile/confined). (e) The test statistics T_n for all trajectories and the 0.025 and 0.975 quantiles; If T_n is under the 0.025 quantiles, the corresponding trajectory is classified as sub-diffusion. (f) A statistic summary of the classification results. Blue color (resp. red and green) represents the number classified as Brownian motion (resp. super- and sub- diffusion).

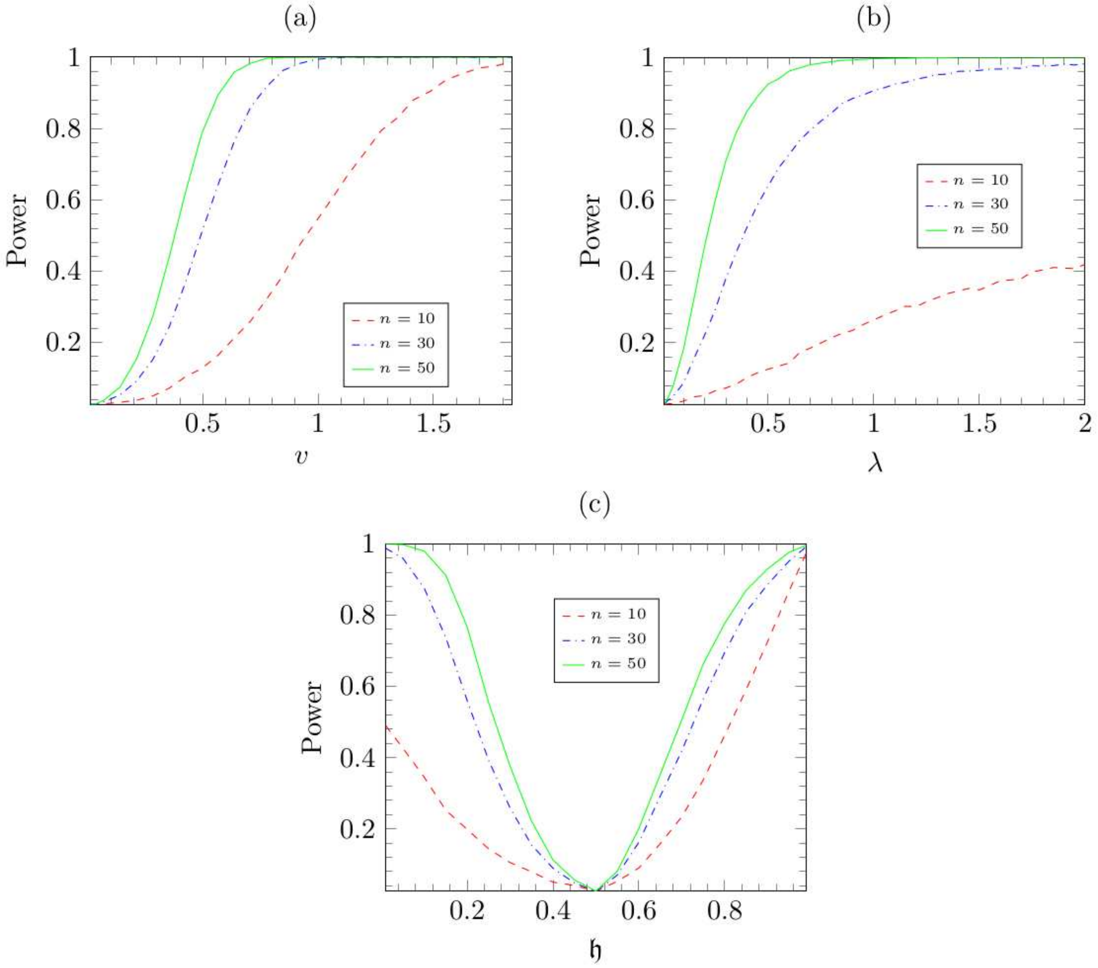


FIGURE 3.4 – The power of test of THOTH performed by Monte Carlo sampling when $\alpha = 0.05, n = 10, 30$ and 50 time points. The parameters v, λ, and H correspond to the drift in Bmd, the power in OU, and the Hurst coefficient in fBm, from [BKV18].

Discussion

From our experimental results, THOTH controls the type I error as confirmed in theory. The rate of wrongly rejected H_0 is around 5%. When particles performing super- or sub- diffusion, the power of test, i.e., the probability that the test rejects the null hypothesis H_0 when a specific alternative hypothesis H_1 is true, is shown in Fig. 3.4 ([BKV18]). The larger v, λ are, and the farther H moves away from 0.5, the higher the power of test is. In other words, the more the trajectories differ from the free diffusion, the better THOTH performs. Therefore, in applications, the values of dynamics parameters (e.g. v, H, λ) are very helpful in predicting THOTH's power of test. We can notice the low power of the test when the number of time steps is reduced to 10.

Because of the heterogeneity induced by stochasticity, free diffusion is frequently wrongly classified as super- or sub- diffusion when the two MSD-based methods are applied. Moreover, the number of points used to fit the MSD curves is arbitrarily chosen by different practitioners. Most importantly, the accuracy of the classification rate cannot theoretically be predicted.

When applying THOTH, two estimators of variance σ^2 are given in (3.21) (first-order) and (3.22) (second-order), respectively. The first order estimator (3.21) can generate bias induced by the velocity term when the motion is Bmd. Moreover, these two estimators are not optimal in the presence of measurement errors.

Simulation of trajectories with measurement errors could give a complementary understanding of how the classification algorithms perform in the presence of measurement errors.

In the next section, we will discuss the optimal coefficient estimation for individual trajectories in the presence of measurement errors. We will focus on the simplest case, i.e., Brownian motion.

3.2.2 Estimation of the diffusion coefficient in single-particle tracking

In the literature, several approaches have been developed to robustly estimate the diffusion coefficient with respect to the localization noise and the motion blur effect. These approaches have also been evaluated using challenging data. In this section, we describe Optimal Least Square Fitting (OLSF, a MSD-based estimator), Maximum Likelihood Estimation (MLE), and Covariance-based Estimation (CVE). It has been shown that the sources of error (see Section 3.1.2), together with the finite length of trajectory N, determine the limit of the best theoretically achievable precision in estimating D.

This theoretical limit is quantified by the Cramer-Rao Lower Bound (CRLB), given in Appendix A in [MB12]. It is proved that CRLB is related to the length of trajectory, N, and the reduced square localization error ξ,

$$\xi = \frac{\sigma_{\text{loc}}^2}{D\Delta t} - 2M, \tag{3.24}$$

where M is the motion blur coefficient defined in (3.13). The CRLB is an information-theoretic quantity derived from the likelihood function and used to determine the lower limit on the variances of unbiased estimators. In other words, no practical algorithm used to estimate the parameters D and σ_{loc} can achieve a smaller uncertainty than the CRLB.

It follows that an estimator is said to be optimal if it reaches the CRLB. We examine here several estimators which reach CRLB in certain conditions. We illustrate the performance of the estimators for different values of ξ and N.

3.2.2.1 Diffusion coefficient estimation by MSD analysis

MSD and covariance analysis

MSD is a measure of the fluctuations in the position of a particle over specified time intervals. Its definition and empirical estimator are given respectively by (3.7) and (3.16). For the sake of simplicity, we use ρ_n to denote $\rho(n\Delta t)$, which is the expectation of $\hat{\rho}_n$, written:

$$\rho_n = 2dDn\Delta t. \tag{3.25}$$

However, if we simulate particles undergoing Brownian motion, one can notice that the MSD curves deviate from the linear model defined in (3.25), especially for large time lag n. The empirical MSD $\hat{\rho}_n$ is generally defined as the average over $N-n+1$ elements (3.16), which are $\Delta_1^2, \cdots, \Delta_{N-n+1}^2$, where $\Delta_k = X((k+n)\Delta t) - X(k\Delta t), k = 0, \cdots, N-n$. For large time lags n, the MSD estimation $\hat{\rho}_n$ is computed from a few elements as the corresponding number of displacements decreases. Consequently, the variance C_{nn} of $\hat{\rho}_n$ is calculated as proposed in [QSE91]. Finally, it is worth noting that the successive pairs of points with time lag n are not independent.

In recent years, [Mic10] and [Ber10], with two different perspectives, provided similar expressions for ρ_n and for the variance and covariance C_{mn} of the MSD in the presence of measurement errors. Given σ_{loc} (3.12) and M (3.13) as the quantified localization error and motion blur effect, respectively, the MSD gives:

$$\rho(n\Delta t) = 2d(\sigma_{\text{loc}}^2 - 2MD\Delta t) + 2dDn\Delta t, \quad 0 \leq M \leq \frac{1}{4}. \tag{3.26}$$

If (3.26) is normalized by the average single frame displacement $2dD\Delta t$, we obtain:

$$\frac{\rho(n\Delta t)}{2dD\Delta t} = (\frac{\sigma_{\text{loc}}^2}{D\Delta t} - 2M) + n = \xi + n,$$

and the definition of the reduced localization error ξ is given in (3.24).

The explicit expressions for the variance and covariance of MSD, in the presence of measurement errors, are given in Appendix C in [MB12]. It raises the question of how many data points in the MSD curve are required to estimate the diffusion coefficient optimally. It was earlier established that at large n, the variance increases due to the reduced number of samples. This may lead us to choose the first 3-5 points in the MSD curves. However, measurement errors can make the first few points of MSD very noisy, especially when $\xi > 1$. Consequently, there exists an optimal number of points to ensure reliable estimation of the diffusion coefficient. This number was previously arbitrarily chosen by different authors. Careful studies showed that the choice of this number has sensible influence on the accuracy of the estimation [Mic10]. If another number other than the optimal number is chosen, the fitted D can be several orders of magnitude larger than the real diffusion coefficient, i.e., the estimation can degrade several times inferior to the optimal estimator.

The optimal number of points to be used for MSD fitting, p_{min}

The diffusion coefficient is obtained by linear regression given the MSD curve defined as:

$$\hat{\rho}(n\Delta t) = a + bn\Delta t. \tag{3.27}$$

Combining the MSD in (3.26) with the linear model in (3.27), we obtain $\hat{a} = 2d\sigma^2_{\text{loc}} - 4dDM\Delta t$ and $\hat{b} = 2dD$. Here p_{min} denotes the optimal number of points that minimizes the fitting error on the fit parameters, a or b. The fitting MSD data points using the optimum number is called the Optimal Least Squared Fitting (OLSF). The explicit expression of p_{min} is given in Appendix B in [MB12]. It showed that p_{min} is dependent on the total number of points (i.e., the length) in a trajectory N and the reduced localization error ξ. To guide users in practices, the authors provide a straightforward representation of p_{min} as a function of N and ξ (Fig. 7 in [MB12]).

A practical problem arises when one tries to find the optimal number of points in experimental data, especially when ξ is not known in advance due to the lack of knowledge of D and σ^2_{loc}. An algorithm of iterative search was designed ([MB12] Appendix B), given a start point as $p_0 = N/10$, to jointly find p_{min} and estimate the parameters D and σ^2_{loc}.

Summarizing the current knowledge of the MSD, it is found that a complete analysis has been carried out, started by [QSE91], complemented by the more recent work of [Ber10],[Mic10], and [MB12]. The complexity, the challenge, the optimal estimators, and the errors of the estimators have been well understood. However, some information, e.g., the covariance in higher time lag n, cannot be used easily to improve the accuracy of estimators due to its complexity.

3.2.2.2 Alternative methods for estimating diffusion coefficients

Other methods such as MLE ([Ber10]), CVE ([VBF14]), and Motion Blur Filter (MBF) ([Cal16]) have been proposed to estimate the coefficients reliably. We present below the main ideas and concepts under concern.

The concept of MLE

We recall that the one-step displacement is denoted as $\Delta_k = X((k+1)\Delta t) - X(k\Delta t), k = 1, \cdots, N-1$. In [Ber10], it is established that the measured displacements Δ_k are distributed according to a multivariate Gaussian distribution, i.e.:

$$\mathbb{E}(\Delta_k) = 0,$$
$$\mathbb{E}(\Delta_j \Delta_k) = \begin{cases} 2D\Delta t - 2(2DM\Delta t - \sigma_{\text{loc}}^2), & j = k \\ 2DM\Delta t - \sigma_{\text{loc}}^2, & j = k \pm 1 \\ 0, & \text{otherwise.} \end{cases} \tag{3.28}$$

Although a particle undergoing Brownian motion has independent increments, the presence of measurement errors induces a non-diagonal covariance matrix. Let $\boldsymbol{\Delta}$ be the $N-1$ component column vector of observed displacements $\Delta_k, k = 1, \cdots, N-1$, and let $\boldsymbol{\Sigma}$ be the $N-1 \times N-1$ covariance matrix defined by (3.28). The log-likelihood function of the data $\boldsymbol{\Delta}$ is written:

$$L_{(\sigma_{\text{loc}}^2, D)}(\boldsymbol{\Delta}) = -\frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\boldsymbol{\Delta}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta} + const. \tag{3.29}$$

The D and σ_{loc} parameters are obtained by maximizing $L(\Delta)$ numerically, which requires calculating the determinant and inverse of the $N \times N$ matrix Σ, or by a standard approximation of Σ [Gra06].

The concept of CVE

The CVE is based on (3.28) and is defined as follows:

$$\mathbb{E}(\Delta_k^2) = 2D\Delta t - 2(2DM\Delta t - \sigma_{\text{loc}}^2), \tag{3.30}$$
$$\mathbb{E}(\Delta_k \Delta_{k+1}) = 2DM\Delta t - \sigma_{\text{loc}}^2.$$

By replacing the expectation by the empirical unbiased estimators, the unbiased CVEs of D and σ_{loc}^2 write:

$$\hat{D} = \frac{< \Delta_k^2 >}{2\Delta t} + \frac{< \Delta_k \Delta_{k+1} >}{\Delta t}, \tag{3.31}$$
$$\hat{\sigma_{\text{loc}}^2} = M < \Delta_k^2 > + (2M-1) < \Delta_k \Delta_{k+1} >, \tag{3.32}$$

where $< \cdot >$ denotes the average over $\Delta_1, \cdots, \Delta_{N-1}$.

In summary, the OLSF method relies on not only the first-step displacement $\rho_1 =< \Delta_k^2 >$ but also on higher orders $\rho_2, \cdots, \rho_m$ where m is the optimal numbers used in MSD curves fitting. However, OLSF does not take into account the variance of ρ_n except when the weighted regression fitting is used. The variance and covariance of higher orders ρ_n have very complicated forms and cannot be used easily. MLE and CVE are regression-free methods, essentially exploiting the first-step displacement Δ_k and the full distribution of Δ_k. The variance and covariance of Δ_k are both used in the construction of the estimators.

3.2.2.3 The performance of OLSF, MLE, and CVE compared to CRLB

As a reminder, the CRLB characterizes the smallest uncertainty that an estimator can achieve. In [MB12], the relative standard deviation (the standard deviation of the parameter normalized by its expectation) of the estimators is obtained from the CRLB as a function of trajectory length N and the reduced localization error ξ. As shown by a simulation study in Fig. 9 in [MB12], when $\xi \sim 10^0$, for a trajectory length $N = 1000$ (blue curve), the normalized standard error of σ^2_{loc} (resp. D) is around 0.08 (resp. 0.08), while for $N = 10$, the normalized standard error of σ^2_{loc} (resp. D) is around 1 (resp. 1). Therefore, the best possible precision depends on the length of the trajectory N and the reduced localization error. It can be noticed that greater N enables better accuracy of both D and σ^2_{loc}.

Among the three proposed estimators, it is claimed that the OLSF and MLE estimators can reach approximately the CRLB for all the settings of N from 5 to 1000 and ξ from 10^{-2} to 10^5 (Fig. 9 in [MB12]). For CVE, it can reach CRLB when $\xi \leq 1$ but has greater variance than the CRLB for $\xi > 1$.

3.2.3 Perspectives

In this section, we analyzed the influence of correlations and localization errors on the estimating of diffusion coefficients. We focused on Brownian motion, but similar analyses are required in the case of sub- and super-diffusion.

In addition to the mentioned estimators MSD, MLE, and CVE, several other approaches have been proposed, like MBF (modified Kalman Filter) [Cal16] and neural networks [Gra+19a]. In both cases, a unified framework was developed to handle several motion modes at once.

MBF [Cal16] models particle motion as an OU process:

$$dX_t = (v - \kappa X_t)dt + \sqrt{2D}dB_t \tag{3.33}$$

$$\psi_{t_i} = \frac{1}{t_E}\int_{t_i - t_E}^{t_i} X_t ds + \epsilon^{\text{loc}}_{t_i}, \tag{3.34}$$

where v and κ characterize the instantaneous velocity of the particle. When $v \neq 0$ and $\kappa = 0$, (3.33) is nothing else than Brownian motion with drift given in (3.11). When $v = 0$ and $\kappa > 0$, (3.33) is the conventional model used to represent confined motion. MBF can separate the motion blur effect from other localization errors and reliably estimates the model parameters and coefficients. Equation (3.34) is used to model the errors of measurements (as in [Ber10]). The MBF can handle the varying localization error and take into account phenomenon such as photobleaching, and the time order of the measurement into account. It is useful to estimate the temporal-varing parameter changes instantaneously, rather than giving only an average over the trajectory. Therefore, the stationarity of the process is not required. The author tested the performance of MBF for D ranging from 10^{-3} to $10^{-1}\mu m^2/s$, the time interval Δt ranging from $25ms - 100ms$, and the total number of points $N \sim 100 - 400$, $\sigma_{\text{loc}} = 30nm$. However, not all experimental data up to now can reach such quality of information. The limit of MBF needs to be tested further in real experimental conditions.

Another model which unifies super-, free, and sub- diffusion is fBm. The Hurst index H is model-specific, where $0 < H < 1/2, H = 1/2$, and $1/2 < H < 1$ represent sub-, free, and super-diffusion, respectively. In [Gra+19a], the neural network is trained to estimate the Hurst index in the presence of localization errors (without the effect of motion blur). The author assumed that the dynamics of the observed particles are consistent with the fBm.

In the next section, we apply the previously mentioned methods, which are suited for MreB dynamics analysis, to characterize dynamics of MreB aggregates observed in 2D image sequences acquired with fluorescence microscopy techniques.

3.3 MreB dynamics analysis

As is observed in [Bil+17], the MreB aggregates present different regimes of motion. The proportion of each regime may vary during different cell growth periods. In this section, we focus on the problem of classification of MreB dynamics and the optimal estimation of diffusion coefficient D and localization error σ_{loc}.

3.3.1 MreB dynamics data

The images of the MreB dynamics are acquired by combining SIM with TIRFM (see Section 2.3). The structured illumination technique improves the spatial resolution by a factor of two ($\sim 120nm$) compared to the diffraction limit ($\sim 250nm$). Consequently, SIM offers better localization precision of the MreB aggregates and makes it possible to observe the unique filament of MreB ($\sim 170nm$ in length). In TIRF microscopy, the evanescent field penetrates $\sim 200nm$ into the sample, which makes it suitable to observe the MreB aggregates moving along the surface of the 3D inner membrane of rod-shaped bacteria cells. However, data acquired with TIRF is biased as it produces temporal series of 2D images depicting the projection of the 3D dynamics onto the support plane.

In this section, the projection bias is corrected before applying the analysis methods to the dynamics of MreB. As the membrane shape in 3D is approximated by a cylinder, provided that the cell radius is estimated in a separate way (e.g., by phase-contrast microscopy), the reconstruction of the dynamics on 3D is straightforward by applying a geometrical transform. If the radius is unknown, we propose a computational method in Section 3.4 to assess the radius from the planar projection of trajectories [HSH13].

3.3.2 Classification of MreB trajectories

The first problem we deal with is the classification of trajectories into different modes of diffusion, namely directed motion, Brownian motion, anomalous sub-diffusion, and "immobile." The performance of three classification methods described in Section 3.2.1, $\beta-$MSD, R^2-MSD, and the Hypothesis Test THOTH, are compared. To illustrate the performance of the three methods, the MSD and the trajectories are plotted with classification results of the three methods (Fig. 3.5). The two MSD methods, $\beta-$MSD and R^2-MSD, show similar results. The MSD curves with larger slopes are classified as directed motion (Fig. 3.5a and b), which correspond mainly to particles crossing the field (Fig. 3.5e and f). The THOTH also provides consistent results with the MSD methods. In Fig. 3.5c, points with hot colors (from red to yellow), representing T_n above the 0.975 quantiles line, are classified as super-diffusion. Points in dark blue, with T_n lower than the 0.025 quantiles line, are classified as sub-diffusion. In Fig. 3.5d (MSD curves plotted with the same color as in Fig. 3.5c), hot colors correspond to MSD curves with large slopes and correspondingly cold colors to MSD curves with small slopes.

The two-by-two statistical comparison among the three methods is presented in Fig. 3.6 with the help of confusion matrices. The confusion matrix between the two MSD methods is presented in the 1^{st} matrix. Accordingly, the confusion matrices between R^2-MSD and THOTH, between $\beta-$MSD and THOTH are presented in the 2^{nd} and 3^{rd} matrices. It turns out that there exist large discrepancies among the results of the three methods. The results should be analyzed in special cases. Relatively speaking, the R^2-MSD and THOTH give similar results (the 2^{nd} matrix), especially in the categories of BM and super-diffusion. Between β-MSD and THOTH (the 3^{rd} matrix), among those classified as BM by β-MSD, 95.3% are classified identically by THOTH. On the contrary, among those classified as BM by THOTH, only 29.1% are classified identically by β-MSD. Among those classified as super-diffusion by β-MSD, only 56.5% are classified identically by THOTH. On the contrary, among those classified as super-diffusion by THOTH, 94.1% are in agreement with β-MSD.

How to choose the classification method in real experimentation ? Depending on the objective of the classification, one or a combination of two methods can be privileged. For example, if the goal of the classification is to obtain an accurate percentage of the different classes, i.e. to maximize the overall accuracy, then the THOTH method can be preferred. If the goal is to select all super-diffusive trajectories to further analyze the characteristics of super-diffusion, then MSD may be preferable.

It is worth noting that if the aggregate moves by switching between different diffusion types during the observed period, the estimated diffusion coefficient value D will be averaged. Several

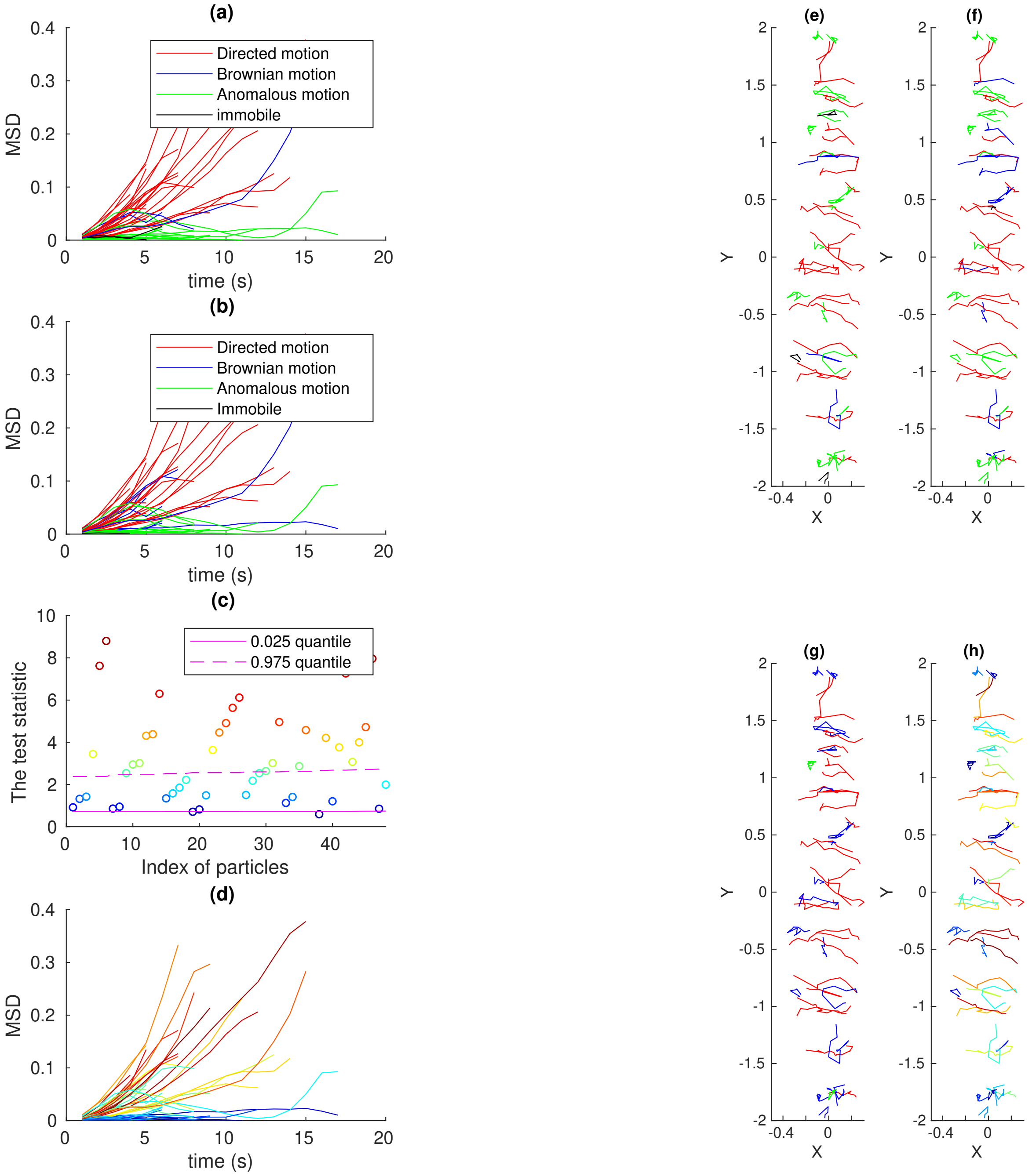


FIGURE 3.5 – Illustration of the performance of the classification methods on one bacterium cell. (a) and (e): The MSD and trajectories by $\beta-$MSD. (b) and (f): The MSD and trajectories by R^2-MSD. (c) and (g): the test statistic T_n and trajectories classified by THOTH. (d) and (h): The MSD and trajectories plotted with the same color as the T_n in (c).

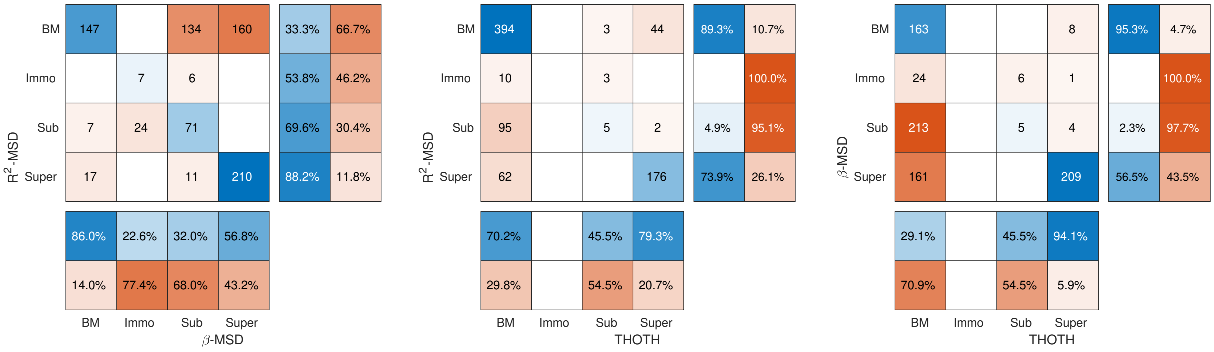


FIGURE 3.6 – Comparison of three classification methods on MreB trajectories. The 1^{st} matrix represents the comparison between R^2-MSD and $\beta-$MSD. The 2^{nd} matrix represents the comparison between R^2-MSD and THOTH. The 3^{rd} matrix represents the comparison between $\beta-$MSD and THOTH. Blue (resp. orange) cells contain the number of trajectories identically (resp. differently) classified by two methods. For example, in the first table, 147 trajectories are classified as BM by both R^2-MSD and β-MSD methods. However, among the trajectories classified as BM by R^2-MSD, only 33.3% are in agreement with β-MSD, others are classified otherwise by the latter method. On the contrary, among the trajectories classified as BM by β-MSD, 86.0% are in agreement with R^2-MSD.

methods have been proposed to detect diffusion switches between different diffusion regimes [Bri+20]. The estimation of diffusion coefficient D is then performed on sub-trajectories obtained by segmenting the long trajectory into homogeneous motion regions.

3.3.3 The estimation of diffusion coefficient D and localization error σ^2_{loc}

As stated in Section 3.2.2, the CRLB depends on σ_{loc} and D. In this section, we apply two methods, the Optimal Least Squared Fitting (OLSF) and the Maximum Likelihood Estimation (MLE), to estimate D and σ^2_{loc}. The final results are influenced by the signal-to-noise level (the reduced localization error (3.24)), as discussed below.

Before applying these two estimators, we mention a simple way to estimate the localization error with an "immobile" dataset.

Estimation of the localization uncertainty σ_{loc} with the "immobile" dataset

In this particular dataset, the "trajectories" of immobilized probes are recorded in the same conditions for diffusing probes (same exposure time and illumination power). The particles detected as "immobile" particles with the MSD methods, that is, the particles for which $\rho(n\Delta t) < 0.01\mu m^2$ for $n = 1, \cdots, N$, are used to constitute the dataset (Fig. 3.7). Thanks to the class of "immobile" particles in the dataset, there exist simple and effective ways to estimate the static localization uncertainty σ_{loc}. Two estimators of σ_{loc} are suggested in [Mic10]. The first estimator

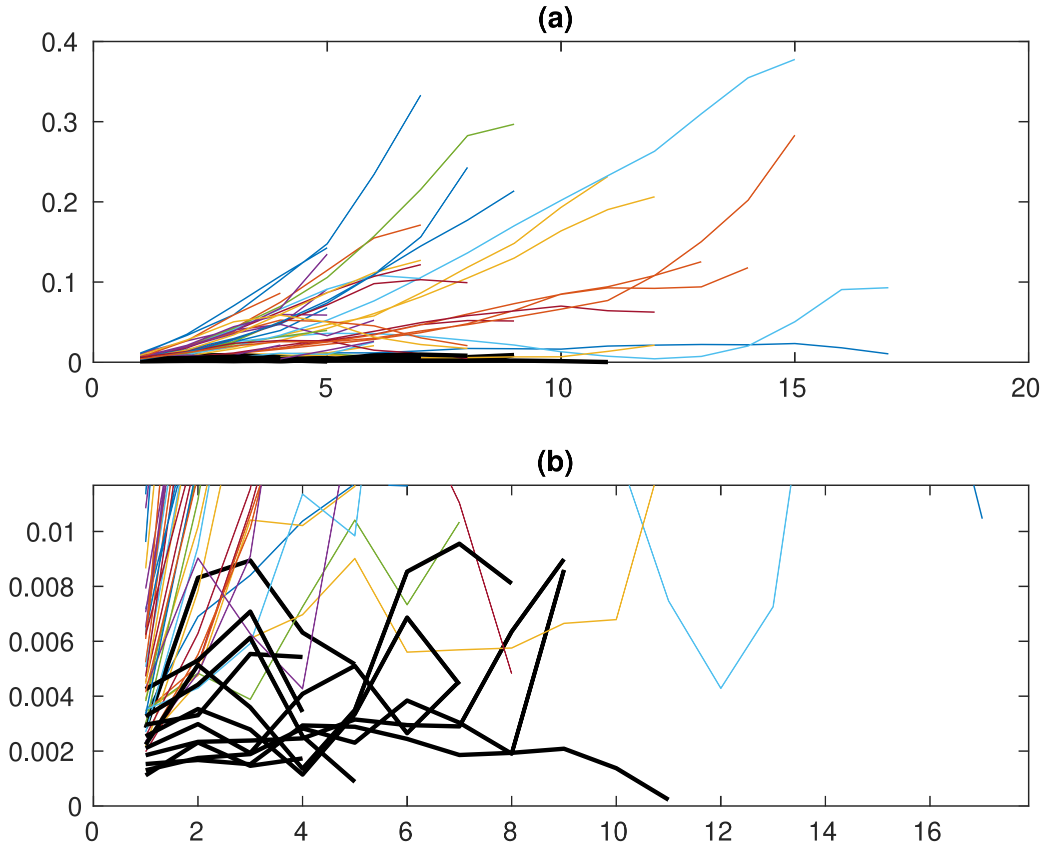


FIGURE 3.7 – **The MSD curves of immobile particles.** (a) MSD for trajectories in one bacteria cell. Black color represents the MSD curves of immobile particles ($\rho_n < 0.01\mu m^2, n = 1, \cdots, N$). The randomly colored MSD curves correspond to moving particles; (b): A zoom for immobile particles.

is through the first-order MSD:

$$\hat{\sigma}^2_{1,\text{loc}} = \rho_1/4.$$

The second estimator concerns an unweighted linear fit of the two (or few) points of the MSD curve. The second estimator gives

$$\hat{\sigma}^2_{2,\text{loc}} = \hat{a}/4,$$

where $\hat{a}$ is the fitted intercept by the linear model (3.27). Applying both estimators of σ^2_{loc} on a set of trajectories extracted from a given cell, the order of magnitude of both estimators is $\sim 10^{-4}\mu m^2$ (Fig. 3.8).

In what follows, we apply OLSF and MLE to estimate diffusion coefficient D and localization error σ^2_{loc} simultaneously for the Brownian motion dataset. In this section, images used to extract

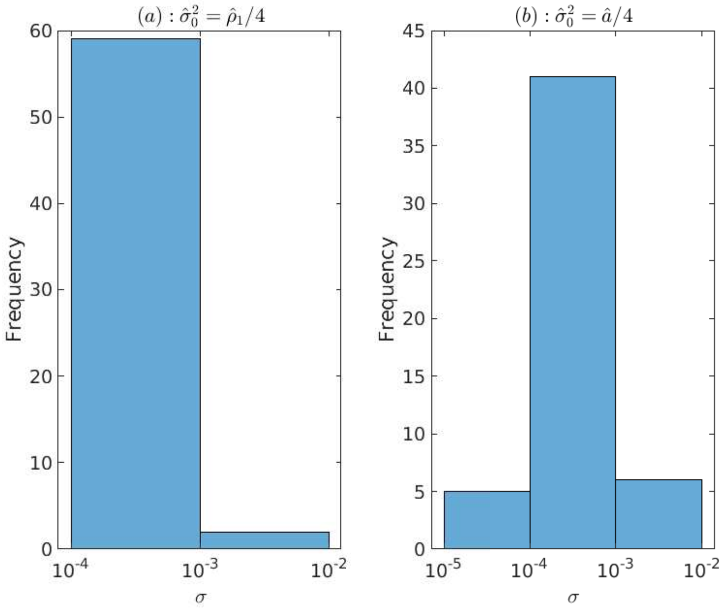


FIGURE 3.8 – **The estimation of static localization error.** (a): σ^2_{loc} is estimated by $\rho_1 = 4\sigma^2_{\text{loc}}$; (b): by linear regression $\rho(t) = \hat{a} + \hat{b}t$, $\hat{a} = 4\sigma^2_{\text{loc}}$.

trajectories are recorded by SIM-TIRFM to benefit from the doubled resolution compared to the diffraction limit. However, the complexity of the image reconstruction procedure makes it difficult to quantify the motion blur effect. Moreover, in the image acquisition setting, the exposure time ($t_E = 10ms$ for one raw image, 9 raw images required to reconstruct one SIM image) is much shorter than the time interval ($\Delta t = 1s$). Therefore, the motion blur effect caused by particle diffusion and finite exposure time are not considered on these datasets.

Estimation of D and σ^2_{loc} for particles undergoing Brownian motion

Firstly, we extract the free diffusion dataset by trajectory classification. In the previous section, we performed a simulation study with Brownian trajectories to evaluate the performance of MSD-based methods and THOTH. We confirmed the results presented in [BKV2018] that more than 50% of Brownian trajectories are misclassified either as super-diffusion or sub-diffusion. In what follows, we consider THOTH as it provided more robust results: 95% of trajectories were correctly classified as free diffusion.

Next, the diffusion coefficient D and the localization error σ^2_{loc} are estimated by OLSF and MLE, respectively, using estimation algorithms given in the supplementary materials in [MB12]. It is shown in [MB12] that both methods achieve nearly optimal performance according to CRLB, as the algorithms proposed by the authors have an excellent agreement with the theory at different settings of (ξ, N).

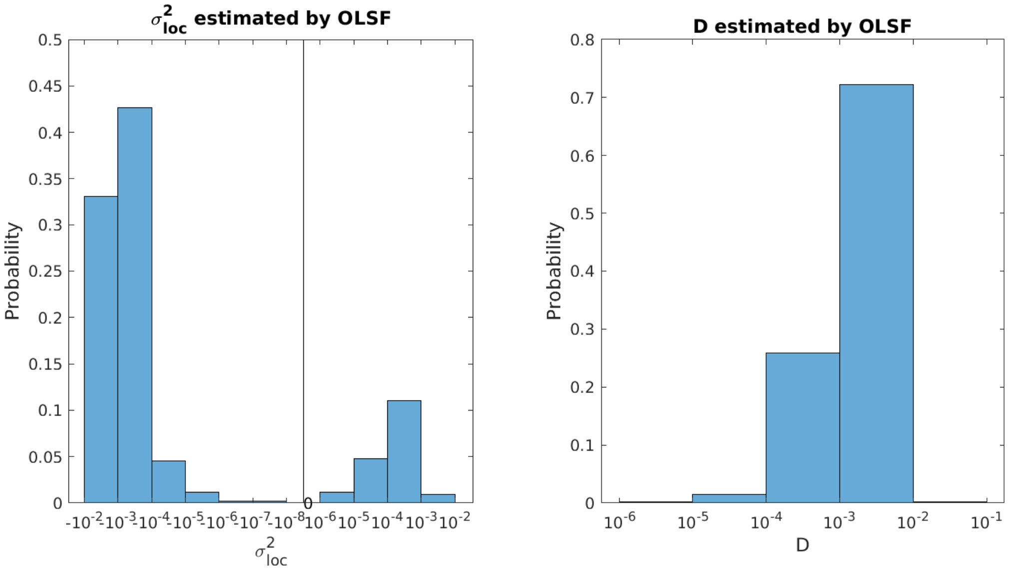


FIGURE 3.9 – The estimation of D and σ^2_{loc} by OLSF.

It is recalled that by OLSF, the MSD curve is fitted to a linear model (3.27). Consequently, $\hat{D} = \hat{b}/4$ and $\hat{\sigma}^2_{\text{loc}} = \hat{a}/4$. Although negative values for the estimated σ^2_{loc} do not have physical meaning, the OLSF method cannot guarantee to provide positive values (Fig. (3.9) (a)). The presence of negative values of σ^2_{loc} may be caused by the non-free diffusion regime of the particles due to errors in tracking or classification steps. It may also be caused by variance in MSD curves calculation. In a word, the regression method for parameters estimation is very sensible to noise and to variance of the MSD, therefore is not easy to use in real experimentation.

With MLE, we obtain $\hat{D} \sim 10^{-4}\mu m^2$, $\hat{\sigma}^2_{\text{loc}} \sim 10^{-4}\mu m^2$, and $\hat{\xi} \sim 10^0$. (Fig. 3.10). Unlike OLSF, MLE returns non-negative values for D or σ^2_{loc}. Although it seems initially a desirable attribute, it tends to have an asymmetric distribution resulting in "piling up" at $D = 0$ and $\sigma^2_{\text{loc}} = 0$, as discussed in [Mic10]. That could explain why two peaks appear in the distribution of estimated σ^2_{loc} (Fig. 3.10a) and D (Fig. 3.10b). We consider that the right side peaks of both sub-figures correspond to the estimators for free diffusion. In contrast, the left side small peaks may correspond to particles wrongly constructed or wrongly classified as free diffusion.

In summary, by applying the OLSF and MLE, we estimate the diffusion coefficient D and the localization error σ^2_{loc} for particles performing free diffusion. The regression-free MLE over-performs the OLSF and gives more reliable estimators of $D \sim 10^{-4}\mu m^2$ and $\sigma^2_{\text{loc}} \sim 10^{-4}\mu m^2$. Consistently, by another independent approach, the analysis of "immobile" particles, the localization error σ^2_{loc} is also estimated to be $\sim 10^{-4}\mu m^2$. Consequently, the order of magnitude of the reduced localization error ξ should lie in the range $1 - 10$.

The information on ξ and N allows one to see how well the parameters can be estimated in a given experimental scenario. As stated in Section 3.2.2, with the knowledge of ξ and the

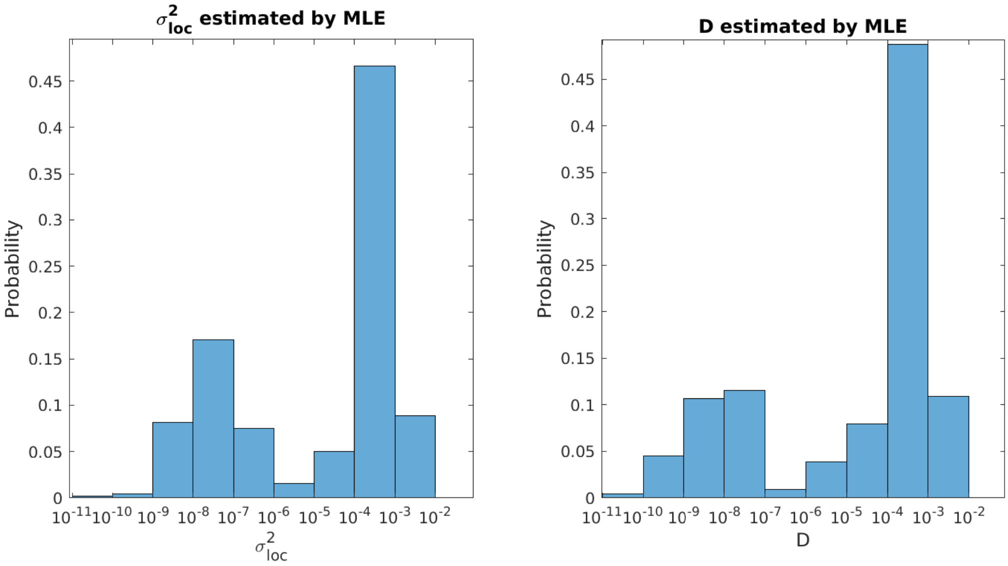


FIGURE 3.10 – The estimation of D and σ^2_{loc} by MLE.

average length of the trajectories, the CRLB of the estimators of D and σ^2_{loc} can be calculated [MB12]. From MreB trajectories obtained with SIM-TIRF, the reduced localization noise $\xi \sim 1$ and the average length of trajectories $N \sim 10$. Consequently, the optimal estimation procedure (i.e., the best attainable accuracy of the estimators of parameters D and σ^2_{loc}) gives $S(D)/D \sim 1$ and $S(\sigma^2_{\text{loc}})/\sigma^2_{\text{loc}} \sim 1$.

3.3.4 Discussion

In this section, three non-parametric classification methods were applied to analyze the MreB trajectories, recorded with SIM-TIRF. We derived estimators of the diffusion coefficient D and the localization error σ_{loc} obtain by the OLSF and the MLE approaches.

More sophisticated methods could be investigated (e.g., Bayesian approach), but they would probably fail in the present study for the following reasons. First of all, the average length of trajectories is ~ 10. Secondly, the diffusion coefficient $D \sim 10^{-4} \mu m^2/s$ is small. Finally, the localization error σ_{loc} has the same order of magnitude as D, which degrades the estimation of D.

3.4 Estimation of the radius of the cylinder from projected trajectories

Unlike previous methods established in the Lagrangian setting, we focus here on an Euler paradigm suggesting that the observer stands in one position and observes all particles passing through this position. The approach inspired by the ideas of Euler aims at estimating the local features of the trajectories, e.g., the local drift and the diffusion coefficient. It is assumed that the trajectories are independent. Unlike MSD methods that require long trajectories and average the diffusion parameters along the trajectory, the approach concerned is suitable for analyzing short trajectories and detecting the appearance or disappearance of certain dynamical events depending on time.

In what follows, the position of particles is modeled by the overdamped Langevin equation, where the diffusion is generated by the thermal agitation of the ambient molecules and a field of force. Let us consider the following SDE ([Hoz+12])

$$dX_t = \mu(X_t)dt + \sqrt{2}B(X_t)dW_t, \tag{3.35}$$

where $\mu(X_t)$ is the drift field, $D(X_t) = \frac{1}{2}B(X)B^T(X)$ is the diffusion tensor, and W_t is the Wiener process. The difference between Bmd in (3.11) and (3.35) is that the drift and diffusion coefficients of a particle depend on its position X_t in (3.35).

The model parameters of (3.35), $\mu(X_t)$ and $D(X_t)$, are recovered from the conditional moments of the trajectory increments $\Delta X = X(t+\Delta t) - X(t)$ ([Sch09])

$$\mu(x) = \lim_{\Delta t \to 0} \frac{\mathbb{E}[\Delta X(t)|X(t) = x]}{\Delta t}, \quad D(x) = \lim_{\Delta t \to 0} \frac{\mathbb{E}[\Delta X(t)^T \Delta X(t)|X(t) = x]}{2\Delta t}, \tag{3.36}$$

where $\mathbb{E}[\cdot|X(t) = x]$ denotes the expectation of the particle increments at point x. In practice, the expectation is estimated with a finite sample size, and at the same time, Δt, the time-resolution of the recording cannot be infinitely small.

3.4.1 The estimation of the effective velocity and standard error

In this section, we briefly summarize the approach proposed by Hoze and Holcman [HH17] to estimate the local drift and diffusion coefficients. First, the image is partitioned into square bins (or blocks) noted as $B(\mathbf{x}_k, r)$ with size r and square center $\mathbf{x}_k = (x_k, y_k)$. The local drift and diffusion coefficients are then estimated from trajectory points falling into each square (see details in [HSH13]). Note that the time-resolution Δt generates bias on the calculation of diffusion coefficient. Therefore, based on the estimators proposed in [HSH13], we proposed an updated version for the estimator of diffusion coefficient, consistent with the SDE (3.35) (A.1):

$$\mu_x(\mathbf{x}_k) \approx \frac{1}{N_k \Delta t} \sum_{j=1}^{N_t} \sum_{\tilde{\mathbf{x}}_i^j \in B(\mathbf{x}_k, \Delta x/2)} (x_{i+1}^j - x_i^j), \tag{3.37}$$

$$\mu_y(\mathbf{x}_k) \approx \frac{1}{N_k \Delta t} \sum_{j=1}^{N_t} \sum_{\tilde{\mathbf{x}}_i^j \in B(\mathbf{x}_k, \Delta x/2)} (\frac{y_{i+1}^j - y_i^j}{\Delta t});$$

$$\tilde{\sigma}_{xx}^2(\mathbf{x}_k) \approx \frac{1}{2N_k \Delta t} \sum_{j=1}^{N_t} \sum_{\tilde{\mathbf{x}}_i^j \in B(\mathbf{x}_k, \Delta x/2)} (x_{i+1}^j - x_i^j)^2 - a_x^2(x_k)\Delta t, \tag{3.38}$$

$$\tilde{\sigma}_{yy}^2(\mathbf{x}_k) \approx \frac{1}{2N_k \Delta t} \sum_{j=1}^{N_t} \sum_{\tilde{\mathbf{y}}_i^j \in B(\mathbf{y}_k, \Delta y/2)} (y_{i+1}^j - y_i^j)^2 - a_y^2(y_k)\Delta t.$$

where $B(\mathbf{x}_k, \Delta\mathbf{x}/2)$ is a square bin $\{x \in [x_k - \Delta x/2, x_k + \Delta x/2], y \in [y_k - \Delta y/2, y_k + \Delta y/2]\}$, N_k is the number of points in $B(\mathbf{x}_k, \Delta\mathbf{x}/2)$.

Based on the extracted local information, an algorithm is established for reconstructing the 3D surface and the 3D stochastic dynamics from the planar projection of trajectories [HSH13]. For particles diffusing on a 3D membrane surface, the common 2D microscope captures images depicting the projection of their dynamics onto the 2D planar cover-slip. This projection causes bias on the estimations of drift and diffusion coefficients. In [HSH13], the explicit expressions of the effective (or projected) drift and diffusion coefficient are given as a function of the local surface properties, such as the mean curvature. It is shown that for free diffusion, the projection generates a drift term in the projected dynamics. As for diffusion already containing a drift, the observed drift is the composite of the true drift in 3D and another term due to curvature effects. In what follows, we recap the principles of the algorithm. Furthermore, we propose an application of this method to recover the radius of the 3D cylindrical surface.

3.4.2 Projection of a diffusion process onto a 2D plane

Assume that the surface under study has an explicit representation $z = f(x, y)$, where $f(x, y)$ is a sufficiently smooth function defined in the (x, y) plane. The 2D plane is assumed to be tangent to the surface at the origin $\mathbf{0}$. Let $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ be the orthonormal unit vector, where $\mathbf{k}$ is in the direction of the $z-$axis. It follows that

$$\mathbf{X}_t = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}, \quad z(t) = f(x(t), y(t)).$$

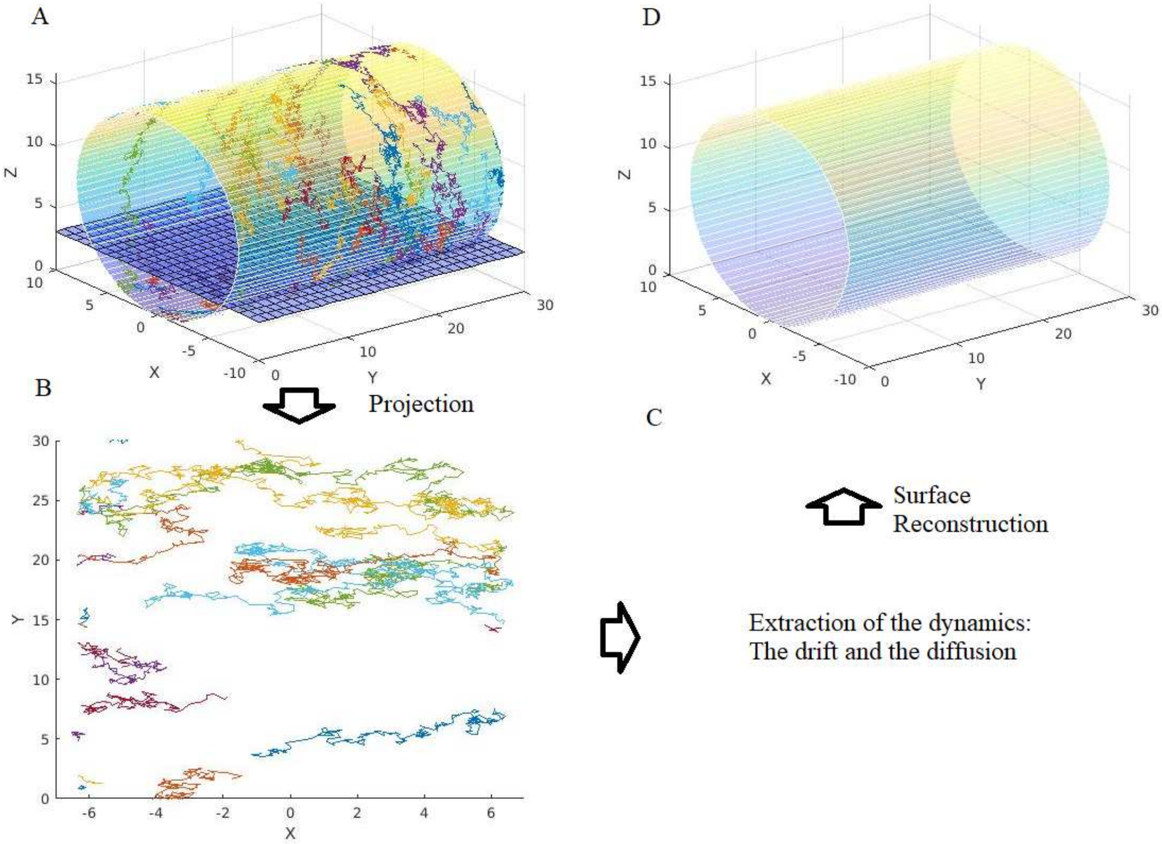


FIGURE 3.11 – The schema of 3D surface reconstruction. The dynamics taking place on a 3D surface (A) are acquired as 2D projected trajectories by 2D microscope (B). Through the extracted drift and diffusion of the projected dynamics (C), the original surface can be reconstructed (D).

It is shown in [HSH13] that when a Brownian motion on the surface $\mathcal{S}$ is projected onto a 2D plane, the observed motion is no longer pure diffusion, but is composed of two terms, the drift term and the diffusion term. The projected process on the (x, y) plane, $\mathbf{x}_t = (x_t, y_t)$ is

$$d\mathbf{x}_t = \mathbf{a}_{\text{geometric}}(\mathbf{x})dt + \tilde{B}(\mathbf{x}, f(\mathbf{x}))d\mathbf{w}, \tag{3.39}$$

where $\mathbf{a}_{\text{geometric}}$ the effective drift and $\tilde{\sigma} = 1/2\tilde{B}\tilde{B}^T$ the effective diffusion tensor are dependent on the local properties of the surface, e.g. the mean curvature. Given the dynamics on the surface as defined by the stochastic process in (3.35), the motion projected onto the plane is also a Brownian motion with drift. The observed drift term is composed of the projection of the original drift and the projection of diffusion, while the observed diffusion is the projection of the original diffusion. In the simplified case, where the diffusion coefficient is constant and isotropic, the effective diffusion tensor has the following form ((43) in [HSH13]):

$$\tilde{\sigma}_{x,x} = D\frac{1+f_y^2}{1+f_x^2+f_y^2},\ \tilde{\sigma}_{x,y} = D\frac{f_x f_y}{1+f_x^2+f_y^2},\ \tilde{\sigma}_{y,y} = D\frac{1+f_x^2}{1+f_x^2+f_y^2}, \tag{3.40}$$

where D is the physical diffusion coefficient, and f_x and f_y are the spatial partial derivatives of f in the direction of x and y, respectively.

3.4.3 Estimation of the cylinder radius

It is well known that MreB aggregates diffuse on the inner membrane of the rod-shaped bacteria. The dynamics of MreB can be considered as occurring on the surface of the cylinder due to osmotic pressure, with the inner membrane usually adhering to the cell wall. On the surface of the cylinder, the curvature in the direction of the axis of the cylinder is null. Whereas in the circumferential direction, it is constant $1/R$ where R is the radius of the cylinder. If the radius of the bacteria cell is not provided, assuming that MreB aggregates undergo Brownian motion with drift (3.35), it is possible to estimate the radius and therefore correct the bias of coefficients estimations, as explained below.

Let us define a coordinate system for a cylinder object (Fig. 3.11a). The $x - y$ plane is parallel to the cylinder, and the $x - y$ plane passes through the origin $(0, 0, 0)$. The axis y is along the direction of the axis of the cylinder, while the axis x is orthogonal to it. The axis z is orthogonal to the $x - y$ plane. We suppose that the dynamics are isotropic and constant on the surface of the membrane, that is, $B(\mathbf{X}) = \sqrt{2D}\mathbf{I}$. From (3.40), it follows that

$$\tilde{\sigma}_{xx}(\mathbf{x}) = D\frac{R^2 - x^2}{R^2}, \qquad \tilde{\sigma}_{yy}(\mathbf{x}) = D. \tag{3.41}$$

Using (3.41), we aim to estimating the radius of the cylinder by linear fitting by solving the equation

$$1 - \frac{\tilde{\sigma}_{xx}}{\tilde{\sigma}_{yy}} = \frac{1}{R^2} x^2. \tag{3.42}$$

By variable substitution, note $T = 1 - \frac{\tilde{\sigma}_{xx}}{\tilde{\sigma}_{yy}}$, $S = x^2$, then it gives $T = \frac{1}{R^2} S$. By the ordinary least square fitting, $R = \sqrt{1/\hat{b}}$, where $\hat{b}$ is the estimated slope of the linear model. Although there is no intercept term in the model, we can consider including one into it due to stochastic factors. In the simulation experiments, linear models with and without intercept term are both evaluated.

3.4.4 Procedure for estimating the cylinder radius and evaluation

By using simulated data that mimics the dynamics of MreB on the cell membrane surface captured by TIRF microscopy, we evaluate the procedure dedicated to estimating the cylinder radius. In our experiments, during time T_S, particles are assumed to appear and disappear on the surface with constant rates λ and τ_d, respectively. Particles move according to the stochastic process (3.35). The drift μ is set to 0.7, and the diffusion coefficient D is set to a constant value in one movie during T_S. Note that, without specification, the unit used in the simulation is given in pixel size(1pixel $= 64nm$). Finally, the radius of the cylinder is ~ 8 ($\sim 510nm$). Dynamics on the surface of the cylinder, having a distance to the support plane $< 200nm$, are projected onto the plane and are captured with TIRFM.

In the estimation procedure, the size of the square r is set to 0.5. We aim to evaluate how the number of points N_k in each square $B(\mathbf{x}_k, r)$ and the dynamic feature D influence the estimation accuracy. It is not difficult to find that N_k depends highly on the time-resolution Δt and total observation time T_S. Therefore in what follows, the procedure is evaluated with different $\Delta t, T_S$, and D values.

The effective drift and diffusion coefficient are estimated using (3.37) and (3.38) from the simulated images. According to (3.41), $\tilde{\sigma}_{yy}$ should be constant and $\tilde{\sigma}_{xx}$ parabolic with respect to x, as confirmed by the simulation results (Fig. 3.12a-d). The radius is then estimated by the linear regression (3.42) with and without an intercept term. By simulation, it turns out that allowing the intercept term in fitting can make the estimation more stable (Fig. 3.12e-h, Fig. 3.13). In Fig. 3.12, $\Delta t = 0.1$, $D = 0.04$, and T_S varies is set to $0.5, 1, 2$, and 5 minutes, respectively. It can be noticed that when $T_S = 0.5min$, accordingly, the average number of points in a square $N_k \sim 1.5 \times 10^3$, the curves are noisy, and the points are scattered (a and e). When $T_S = 5min$, accordingly, $N_k \sim 1.5 \times 10^4$, the curves are much more smooth and the scatter points are more regular and aligned (d and h).

Next, for each set of $\Delta t, T_S$, and D values, the experiment is replicated 100 times. The bar plot represents the mean and standard deviation of the estimator of the radius. If we set $T_S = 0.5$ min, $D = 0.04$ (Fig. 3.13). When Δt varies from 0.1 to 0.25, the average number of points in each square decreases from 1300 to 500 (b). At $\Delta t = 0.1$, the standard deviation of

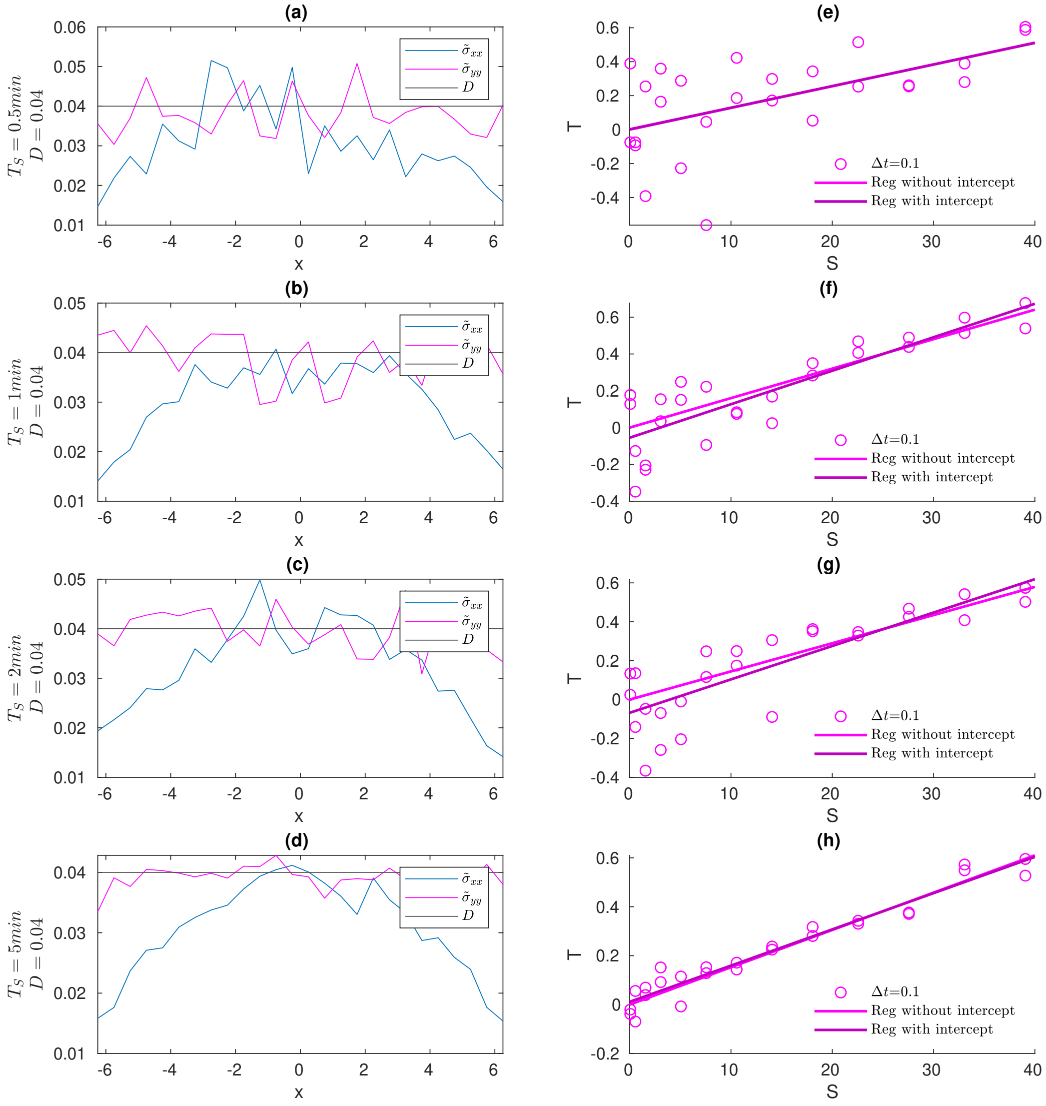


FIGURE 3.12 – The estimated effective diffusion coefficients as a function of x (a)-(d) and the linear regression without and with intercept(e)-(h). $\Delta t = 0.1, D = 0.04$. Corresponding to each line, the total observation time $T_S = 0.5, 1, 2,$ and 5 min, respectively. (a): As $T_S = 0.5$ min, $\tilde{\sigma}_{yy}$ (magenta curve) oscillates around the horizontal line $y = D = 0.04$, and $\tilde{\sigma}_{xx}$ (blue curve) oscillates around a parabolic curve, which corresponds to (3.40). (c)-(d): As T_S increases, the curves become smoother and fit better (3.40). (e): The scatter plot of T vs. S is fitted by linear models, without (magenta) and with (purple) the intercept term, respectively. The slope of straight lines is equal to $1/R^2$ (3.42). (f)-(h): As T_S increases, the scattered points become more regular and aligned. For experiments results with $\Delta t = 0.25$, $D = 0.04, 0.09$ and 0.16, see A.2. The value of the slope of the fitted lines, i.e., the estimated R in different parameter settings, can be found in Figs. 3.13 and 3.14.

R is around $50nm$. However, at $\Delta t = 0.25$, the standard deviation of R is increased to around $200nm$, while the true radius is $510nm$. The performance of the estimator of R decreases when Δt varies from 0.1 to 0.25, which is related to the decrease of the number of points N_k in the squares, but not Δt itself.

If we fix $T_S = 5$ min, varying Δt from 0.1 to 0.25, the order of magnitude of the number of points in each square varies from 1.3×10^4 to 5×10^3. For all the evaluated values of D, $0.04, 0.09$, and 0.16, and the drift $\mu = 0.7$, the estimation of R is satisfying (Fig. 3.14), with relative error $< 5\%(25nm)$.

It can be concluded that to obtain a reliable estimator of the radius, the number of points in each square bin should be $\sim 10^3$. The requirement of this number is remarkably relaxed compared to the recovery of the sphere in the original paper [HSH13] ($\sim 2.25 \times 10^4$). Moreover, $\Delta t = 0.1$ to 0.25 is possible with nowadays commonly used fluorescent microscopy. The requirement on the time interval Δt is also eased compared to the infinitely small $\Delta t = 5 \times 10^{-5}$ in [HSH13].

In Fig. 3.14, the estimator without an intercept has a smaller standard deviation ($< 10nm$) than the estimator with intercept (standard deviation$< 25nm$). However, when there are fewer points, in Fig. 3.13a, when the intercept is deactivated, a bias is introduced to the estimator. On the contrary, using linear fit with intercept, the estimator becomes non-biased and more stable.

3.4.5 Discussion

In this section, we proposed consistent estimators of the drift and the diffusion coefficient given the dynamics following the Langevin equation (3.35). The estimators correct the biases induced by the finite time-resolution Δt, as investigated in [HH15]. Secondly, we applied the method of surface reconstruction proposed in [HSH13] to estimate the radius of the rod-shaped bacteria cell.

Again, we studied the dynamics of particles on an Euler paradigm. The local information of the drift and diffusion is probed instead of the average information over long trajectories. Consequently, the method is appropriate to deal with short trajectories and to handle dynamics with switching episodes.

Assuming the dynamics follow the Langevin equation (3.35) on a cylindrical surface, a linear fitting procedure has been designed to recover the radius of the cylinder from the projected dynamics. The model is simplified by approximating the surface as cylindrical, whose curvature is constant ($1/R$ and 0 along the x and y directions, respectively) and assuming that the drift and diffusion coefficient are constant. In order to derive reliable estimators of the effective (or projected) drift and diffusion coefficients, the number of points in the neighborhood of a point $\mathbf{x}_k$ (e.g., a square $B(\mathbf{x}_k, r)$) must be in the order of $\sim 10^3$. The time interval is not required to be infinitely small. The time step $\Delta t = 0.1$ to 0.25, which is generally possible with fluorescent

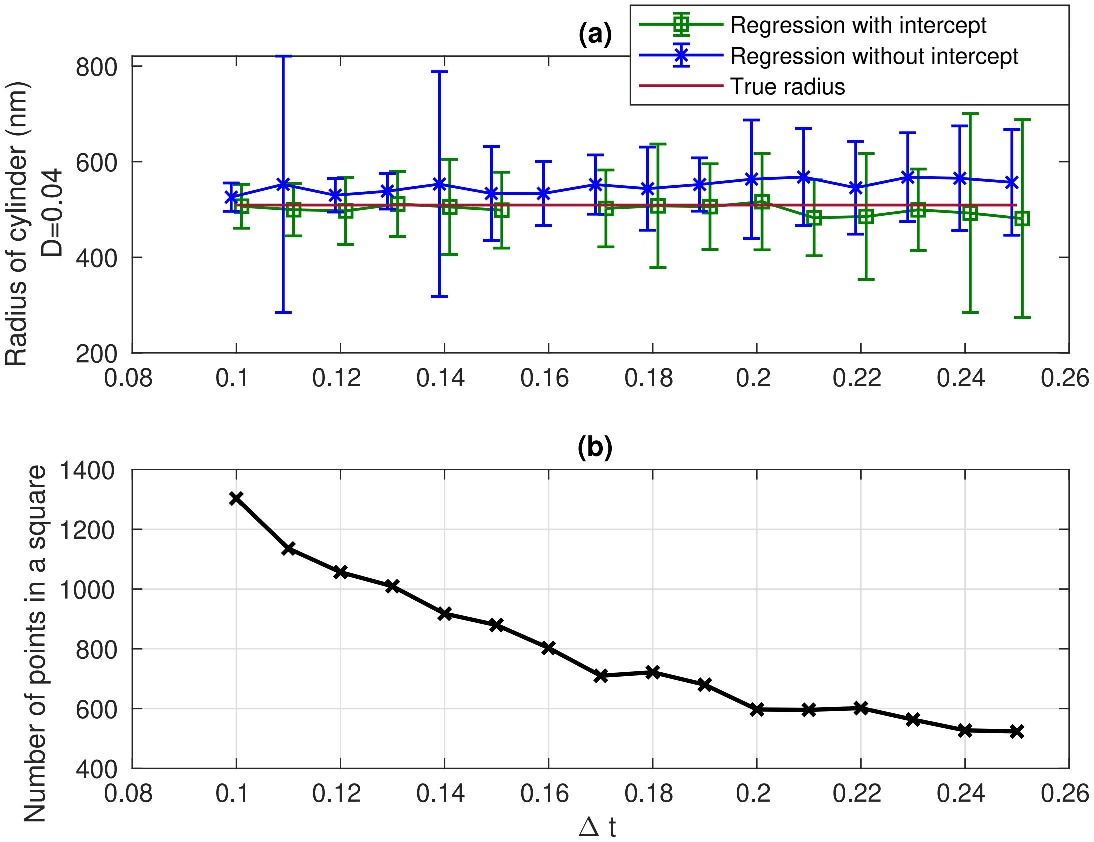


FIGURE 3.13 – **The estimation of R at different Δt.** The diffusion coefficient $D = 0.04$ and the total observation time $T_S = 0.5$ min. (a): The mean and standard deviation over 100 replications of the estimator of the radius R, by linear fitting with (green) and without (blue) intercept term. The red line represents the true radius $510nm$. Without intercept (blue), the estimator has a smaller standard deviation but is biased and less robust. On the contrary, the estimator with intercept (green) is almost unbiased and is more robust. (b): The average number of points N_k in each square at Δt from 0.1 to 0.25, over the 100 replications. N_k decrease almost linearly with respect to Δt.

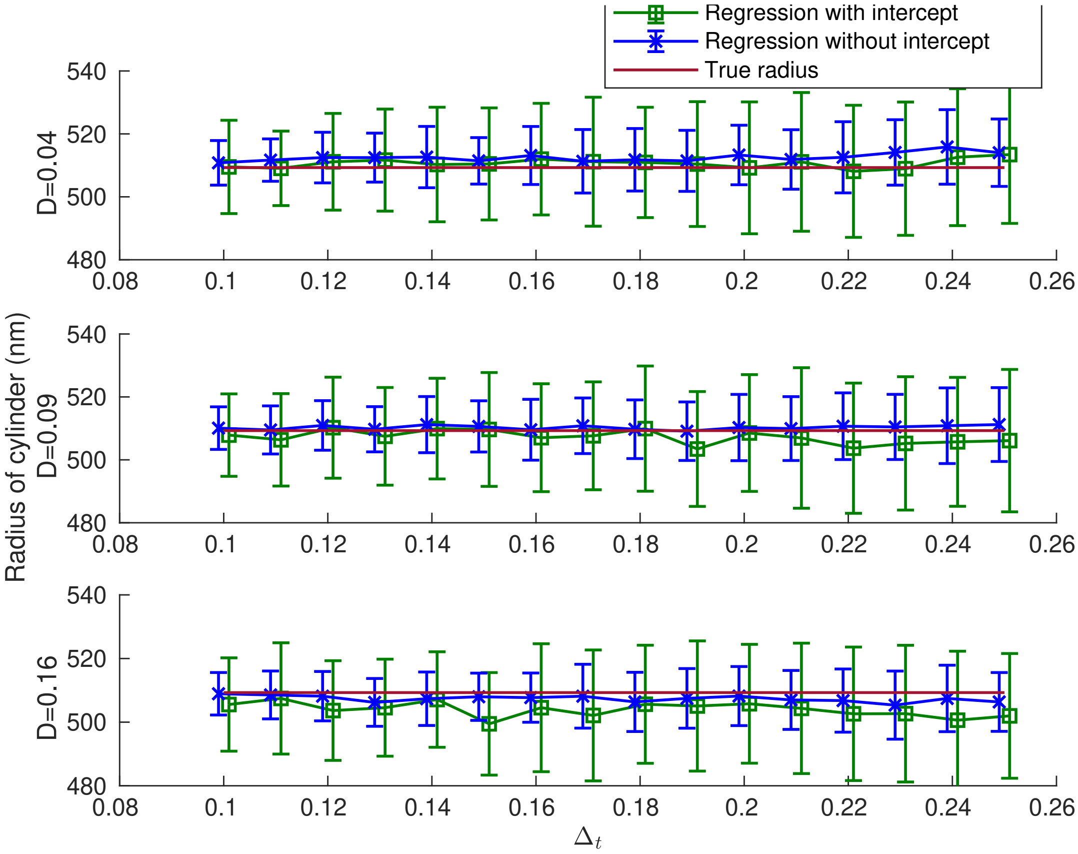


FIGURE 3.14 – The estimation of R with different values of D and different Δt. The total observation time is $T_S = 5$min. The red line is the true value of the radius $(510nm)$. The estimations of R using linear regression, without intercept and with intercept, are represented in blue and green, respectively. The vertical bars represent the standard error of 100 replications of the experiments.

microscopy, provides reliable estimators of R, provided that the number of points in the square is large enough.

The procedure to estimate the cylinder radius R was evaluated on synthetic datasets. Unfortunately, the experimental data of MreB dynamics recorded with TIRF and SIM-TIRF do not provide enough density of points to confirm our results obtained on synthetic datasets. In order to ensure good performance of the procedure, single-particle tracking super-resolution microscopy (e.g., spt-PALM [Bet+06]) will be investigated in future works since it will provide a high densities of particles.

3.5 Conclusion and Perspectives

In this chapter, we reviewed a few methods for dynamics analysis in cell imaging. The ability and limits of these methods were discussed. In our experimental data obtained by fluorescence microscopy, several sources of error cannot be ignored and should be included in the algorithms in order to better interpret the results.

We addressed two main problems, that is the classification of trajectories (adjusting data to the most appropriate model) and the estimation of coefficients. We considered different diffusion types, including free diffusion, sub-diffusion (fBm (with $0 < H < 1/2$) and OU) and super-diffusion (Brownian motion with drift and fBm ($1/2 < H < 1$)).

To address the problem of model classification and parameters estimation, we evaluated several approaches:

- MSD is the most used method for single-particle trajectories analysis;
- Hypothesis Test (i.e. THOTH algorithm) [BKV18] using maximum excursion distance;
- The Bayesian approach, which amounts to both estimating parameters and selecting the best motion model among a set a pre-defined collection of models (e.g. [TM13]);
- The machine learning approach recently applied to SPT and used to discriminate free diffusion from fBm and CTRW (see [Granik et al., 2019]). This approach is actually dedicated to anomalous diffusion and the comparison of competing models but it can be extended to any diffusion process.

In the case of MreB dynamics analysis, the trajectories are short (~ 10 points, $\Delta t = 1$) and all the aforementioned methods cannot be safely applied. We focused on MSD and THOTH which are non-parametric test method. THOTH considers Brownian motion as the reference motion (Hypothesis H_0) and the alternative (H_1) is split into two parts: sub-diffusion and super-diffusion. MSD follows the same idea since the reference model is free diffusion. Simulations were carried out to assess the precision of MSD classification. Despite the lack of information (limited SNR, short trajectories and low density of particles), we applied the approach of [MB12] to trajectories classified as free diffusion, to extract diffusion coefficient D, localization error σ^2_{loc}.

The analysis of individual trajectories in single-particle tracking is generally formulated in the Lagrangian setting. Another approach consists in estimating diffusion at a given position from several trajectories passing through a local neighborhood. This Euler setting can reveal the 3D space or 3D surface properties, for example the potential wells, the shape of the surface etc, provided that the density of trajectories is high enough, which is not the case with MreB dynamics.

For future work, we plan to investigate MSD analysis by including measurement errors for directed motion and confined motion as a supplement to free diffusion [Cal16]. In the case of MreB, it should be valuable to aggregate the trajectories in similar cells to have more significant information. Furthermore, the deep learning method showed great potential for analysis in case of noise, with multiple short trajectories [Gra+19a], [Art+19]. It can be extended to other modes of motion and even distinguish the change of regimes along a given trajectory.

As we know that the dynamics are partially observed by TIRF microscopy, in order to obtain a global view of the particle dynamics from the partial observations, a probabilistic procedure, taking into account the particles exiting and re-entering the field of view is discussed in Chapter 4, to reconstruct the MreB dynamics over the whole membrane surface.

CHAPTER 4

PROBABILISTIC RECONSTRUCTION OF TRUNCATED PARTICLE TRAJECTORIES ON A CLOSED SURFACE

Introduction

In two-dimensional (2D) and three-dimensional (3D) live-cell imaging, spatiotemporal events and biomolecule dynamics are frequently observed with an incomplete field of view. Very often these observations are related to regions of observation (ROO) inside a tissue, a cell, or in the neighborhood of membranes. Nevertheless, it is quite unusual to analyze 3D dynamics of biomolecules or events occurring on a closed surface and observed on a 2D plane. Our work is motivated by the study of dynamics of MreB proteins, moving close to the inner membrane during cell wall construction in rod-shaped bacteria ([Bil+17], [TR18]). Its dynamics can only be observed in a small region and are recorded as 2D time-lapse movies (Fig. 4.1a). As for 3D image acquisition, it can even solve the problem of partial observation, but is not always appropriate, especially if the objective is to capture fast and temporally short events as described in [Bil+17]. The frame rate adapted to the scale of dynamics may be too high when compared to the period of time to acquire temporal series of 3D volume ([Bou+14] and [Cor+20]).

To the best of our knowledge, identifying re-entrance events of the same entities inside the ROO is not addressed in the literature. In experimental data, when the unobserved region represents a significant part of the entire surface, a complete description of the dynamics on these closed surfaces becomes of paramount importance for deciphering the mechanisms of some processes. In our study of the regulation of the dynamics of MreB protein, as inputs, we consider a set of trajectories estimated by tracking algorithms (e.g., [Jaq+08], [CBOM13], [Yan+12]). These tracking algorithms are very sophisticated and allow us to handle large sets of particles, different stochastic dynamical models [BBS88], [BN13], and observation models [Gen+06], [LL01]. They take into account birth/death events, and/or split/merge events. Particles may be unobserved or undetected for short periods of time, especially in 2D+time microscopy. However, no computational or statistical method manages the situation corresponding

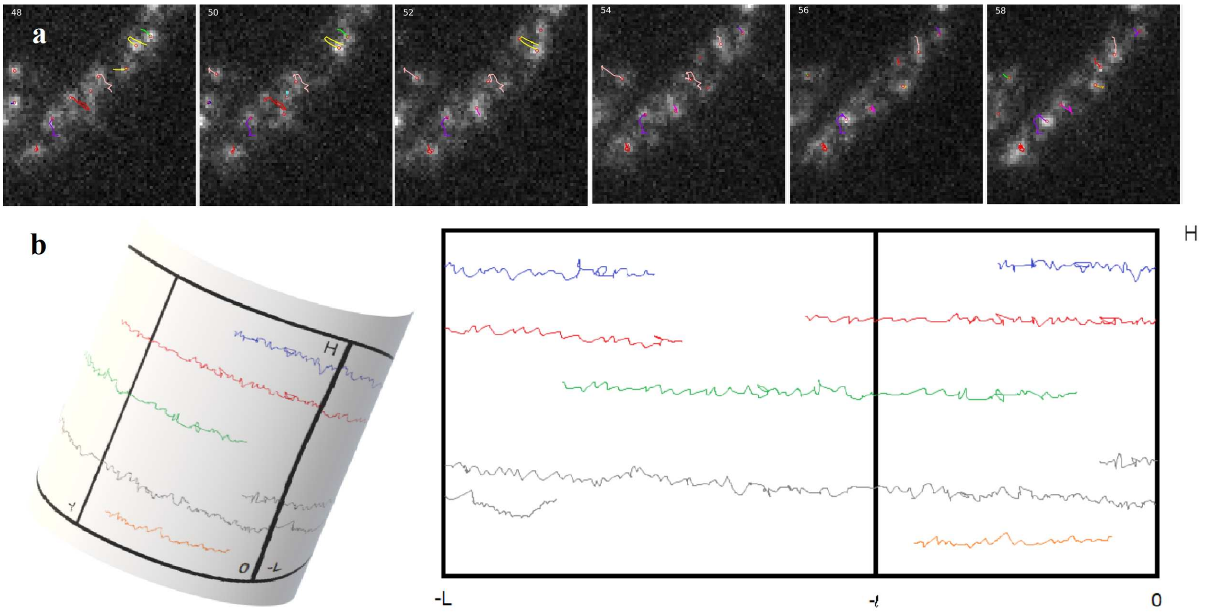


FIGURE 4.1 – **Illustration of the incomplete view in TIRFM.** (a): Several consecutive images from a real TIRFM movie [ABT84]. Tracks are superposed on the images.(b) Left: Illustration of trajectories observed during recorded time $[0, T_S]$ on the surface of a cylinder. Only the motions inside the ROO $]{-l}, 0[\times[0, H]$ can be observed, even though the dynamics happen on the whole surface. Right: Representation of the dynamics on a 2D unwrapped surface $]-L, 0[\times[0, H]$. The objective is to recover the dynamics on the whole surface from the partial observations, by coordinating the inputs through $\{-l\} \times [0, H]$ and the outputs through $\{0\} \times [0, H]$ in a movie during T_S, taking into account particles birth and death events.

to a large hidden region inside the region of interest. Also, the identification of particles leaving the ROO through one border of the domain and re-entering from a far border has not been addressed. Our objective is then to provide a generic approach to tackle the problem of the reconstruction of particle trajectories observed on a small part of a closed surface as illustrated in Fig. 4.1b.

In this chapter, we focus on the design and evaluation of a self-contained mathematical framework to tackle the reconstruction of particle trajectories on cylindrical surfaces, given the tracklets observed in a small window sampled on the surface. In our study, the particles are assumed to obey a stochastic Brownian motion with drift and may appear or disappear during the observation period. Split or merge events are not considered in the modeling framework. The trajectory reconstruction problem is defined as the maximization of the likelihood function given tracklets inside the ROO. The optimization problem to be solved is formulated as an integer linear programming problem. The final algorithm is a data-driven algorithm with no hidden parameter to be set by the user. We demonstrate the performance and robustness of our computational method on simulation data, by varying the ratio of observed to unobserved region, the drift and variance of particles, as well as the rates of birth and death of particles.

The remainder of this chapter is organized as follows. In Section 2, we present the problem formally and introduce notation. In Section 3, we describe the probabilistic framework, including Poisson processes used to describe birth and death events, and Brownian motion with drift to represent particle motion. We also describe the computational procedure aiming at connecting tracklets belonging to the same trajectory, and then recovering the dynamics of particles on the whole closed surface. Note that we suppose that the curvature of the cylinder is known so that the movements are represented on a 2D unwrapped surface. In Section 4, the performance of our algorithm is evaluated on simulated data. Finally, we conclude and propose some future work. A summary of notation useful for the evaluation of the likelihood is given in Section 4.2.1.

4.1 Problem statement and notation

We consider a probabilistic model to represent particles that are born, move and die on a cylindric membrane.

4.1.1 Models description

Formally, let us denote H and L to be the height and perimeter of the cylinder respectively (see Fig. 4.1). We associate 2D coordinates $(x, y) \in [-L, 0] \times [0, H]$ to each point of the underlying cylindric manifold. The particles are "born" with a constant rate λ and appear uniformly at random on the membrane surface. We consider a Poisson process with intensity λ to statistically represent the birth events. Each particle is assumed to have the same constant rate of death τ_d such that life duration T_d of a particle follows an exponential law of parameter τ_d. During its lifetime, a particle k born at time t_0 and located at $\mathbf{Z}_0^k = (X_0^k, Y_0^k)$ moves according to Brownian motion with drift. On the set $]-L, 0[\times[0, H]$, the position of the particle at time $t \geq t_0$ prior to its death time is given by

$$\mathbf{Z}_t^k = \mathbf{Z}_0^k + \mathbf{v}(t - t_0) + \mathbf{\Sigma}\mathbf{B}_{t-t_0}^k \tag{4.1}$$

where

$$\mathbf{Z}_t^k = (X_t^k, Y_t^k), \mathbf{v} = (v_x, v_y), \mathbf{\Sigma} = \left[\begin{array}{cc} \sigma_x & 0 \\ 0 & \sigma_y \end{array} \right],$$

and $\mathbf{B}_t^k$ is a two-dimensional Wiener process.

4.1.2 Modeling hypothesis and MreB dynamics

The study of the dynamics of MreB patches or assemblies in the vicinity of the internal membrane of *Bacillus subtilis* bacteria reveals several subpopulations undergoing constrained, randomly or directionally moving dynamics [Bil+17]. Herein we are interested in the directionally moving subpopulation dynamics. This subpopulation moves possessively around the cell diameter [Gar+11 ; DE+11]. Following Hussain et al, [Hus+18], Billaudeau et al [Bil+19] confirmed that directionally moving filaments travel in a direction close to their main axis, perpendicularly to the long axis of the cell (angle $\gamma = 89.9° \pm 37.0°$). Hence, for some filaments, the speed vector may have a component in the main direction of the bacteria.

According to [WGA19], a motion model (named "biased random walk") reproduces the dynamics patterns of MreB filaments. In their simulations, the speed is constant and the noise variance between several time steps depends on the duration and, possibly on the local curvature of the surface. These properties are shared with the Brownian motion model with constant drift we consider.

4.1.3 Some practical consideration

In order to model the topology of the cylinder as illustrated in Fig. 4.1, we impose deterministic jumps when the process reaches one of the two borders $\{-L\} \times [0, H]$ or $\{0\} \times [0, H]$. For any $y \in [0, H]$, the process reaching position $(-L, y)$ jumps to position $(0, y)$ and vice versa. In y direction the initial position of a particle lies between $[0, H]$. When a particle hits the vertical borders, its following trajectory is no longer considered. Finally, we assume that each particle behaves independently from the others and that there is no fission or fusion of particles.

In what follows, we observe the dynamics at discrete times $\Delta t, 2\Delta t, 3\Delta t \ldots$ We denote Δt the time step on the subset $[-l, 0] \times [0, H]$ with $l < L$. The observations are recorded during a time interval $[0, T_S]$. As we suppose that a particle does not change its drift direction along its trajectory, we assume that $v_x > 0$, even though particles can actually move in both directions, which requires a classification to separate them into two groups. We consider that an observed tracklet of a given trajectory is an output if the last observed point of the segment is within a neighborhood of $\{0\} \times [0, H]$. Meanwhile, we consider that it is an input if the first observed point is within a neighborhood of $\{-l\} \times [0, H]$. Our main objective is then to associate the set of tracklets exiting the observed set $[-l, 0] \times [0, H]$ with the set of tracklets entering this observation set. The challenge is to correctly match the outputs and the inputs associated to particles (see Fig. 4.1).

4.2 Probabilistic models and methods

Let us consider a given sample S, the observation set of all the trajectories. We define the sets $O_S = \{o_1, ..., o_p\}$ and $I_S = \{i_1, ..., i_q\}$ of p outputs and q inputs. Each output $o = (t_o, y_o) \in O_S$ is characterized by its output time t_o and its position $y_o \in [0, H]$, where the particle left the observed region. Similarly, each input $i = (t_i, y_i) \in I_S$ is characterized by its input time t_i and its position $y_i \in [0, H]$, where it entered the observed region. A particle "involved" in an output $o \in O_S$ either died after time t_o in the unobserved region, or is "involved" in a given input $i \in I_S$ with $t_i > t_o$. We will denote this event by $\{o \to i\}$. Similarly, a particle "involved" in an input $i \in I_S$ was either born before time t_i in the unobserved region, or is "involved" in a given output $o \in O_S$ with $t_i > t_o$, which corresponds to the event $\{o \to i\}$.

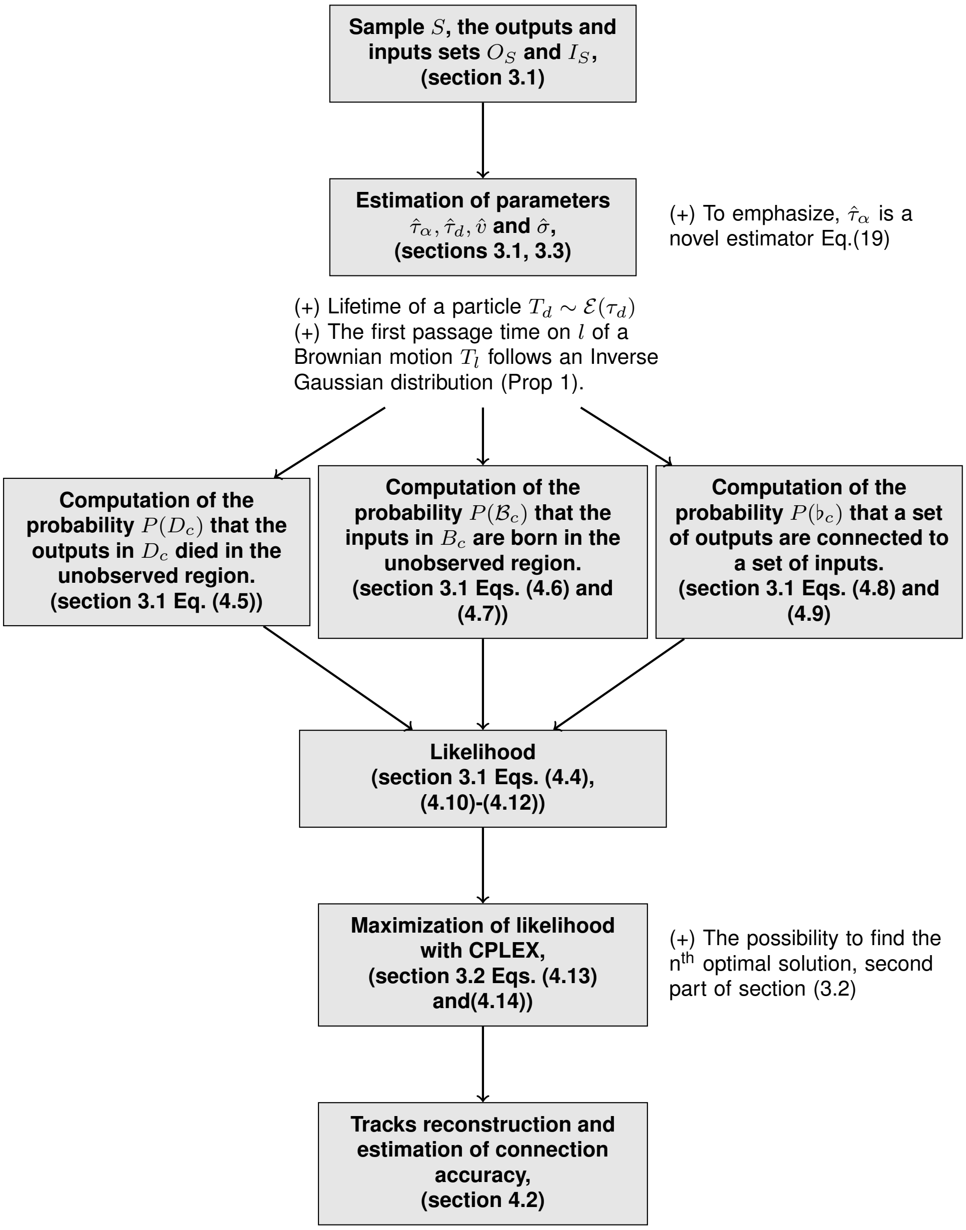
Define $c = (D_c, B_c, b_c)$ with $D_c \subset O_S$, $B_c \subset I_S$, and b_c a bijection from $O_S \setminus D_c$ to $I_S \setminus B_c$ in order to describe the configuration for which all outputs in D_c died in the unobserved region, all inputs in B_c are born in the unobserved region, and the event

$$\bigcap_{o \in O_s \setminus D_c} \{o \to b_c(o)\}$$

was realized. Our aim is to determine the maximum likelihood configuration c given the sample S. The outline of the connection procedure is given in Fig. 4.2 to facilitate the understanding of the modeling steps.

4.2.1 Summary of notation useful for the evaluation of the likelihood

- H: the length of the cylinder,
- L: the length of the circumference of the cylinder,
- l: the length of the part of the of circumference that is observed,
- l_u: the length of the part of the of circumference that is not observed,
- l_e: the difference between the length of the unobserved region and the observed region,
- B_c: for a given configuration (reconstruction) of trajectories c, the subset of trajectories born in the unobserved region and seen on the border $\{-l\} \times [0, H]$,
- D_c: for a given configuration (reconstruction) of trajectories c, the subset of trajectories seen on the border $\{0\} \times [0, H]$ and died in the unobserved region,
- Δt: time stepsize between two consecutive observations,
- N_l: the number of particles born in the observed region and reaching the border $\{0\} \times [0, H]$,
- p_x: the probability of birth of a particle in a strip of width x, to the left side of the border $\{-l\} \times [0, H]$,
- $\hat{p}_x$: an estimator of p_x,



(+) To emphasize, $\hat{\tau}_\alpha$ is a novel estimator Eq.(19)

(+) Lifetime of a particle $T_d \sim \mathcal{E}(\tau_d)$
(+) The first passage time on l of a Brownian motion T_l follows an Inverse Gaussian distribution (Prop 1).

(+) The possibility to find the nth optimal solution, second part of section (3.2)

FIGURE 4.2 – An outline of the connection procedure: from the estimation of the parameters to connection accuracy measurement, including likelihood formulation. All notation are defined in the corresponding sections.

- S: the observation set of all trajectories,
- S_o: the set of tracklets having an output in $\{0\} \times [0, H]$,
- S_l^*: the set of tracklets having an input in $\{-l\} \times [0, H]$ and an ouput in $\{0\} \times [0, H]$, that is crossing the observed region,
- S_r: sample of points inside a restricted region inside the observed region. This region should allow to decide if a particle died or is just moving outside the observed region.
- τ_α: arrival rate at border $\{l\} \times [0, H]$ of particles born in the unobserved region $]-L, l[$,
- $\hat{\tau}_\alpha$: estimator of the arrival rate at border $\{l\} \times [0, H]$,
- τ_d: death rate of particles,
- $\hat{\tau}_d$: estimator of the death rate,
- T_S: time duration of observation.

4.2.2 Likelihood of a configuration

In this section, our objective is to derive an analytic expression of the likelihood $Q(c)$ of a configuration c. The aim is to find, for a given sample S, the configuration $\hat{c}$ such that $P(\hat{c}/S)$ is maximal. It is difficult to calculate directly $P(\hat{c}/S)$. Since $c \subset S \subset O_S$, we can compute $P(\hat{c}/S)$ working conditionally on O_S.

However, since the model is in continuous time and involves random variables with continuous densities with respect to the Lebesgue measure, the conditional probability $P(c/O_S)$ is equal to 0. This prevents us from computing $P(\hat{c}/S)$ directly with the classical conditional formula

$$P\left(c/S\right) = \frac{P\left(c/O_S\right)}{P(S/O_S)},$$

because it gives $P(S/O_S) = \sum_{c \in \mathcal{C}_S} P(c/O_S) = 0$.

Therefore, for each input $i = (t_i, y_i) \in I_S$, we consider a spatiotemporal neighborhood $V_i^\epsilon = T_i^\epsilon \times H_i^\epsilon$ with $T_i^\epsilon = [t_i - \frac{\epsilon}{2}, t_i + \frac{\epsilon}{2}]$ and $H_i^\epsilon = [y_i - \frac{\epsilon}{2}, y_i + \frac{\epsilon}{2}]$ for some $\epsilon > 0$.

The idea is to replace a given configuration c by a set $\mathcal{C}_c^\epsilon$ of configurations where each element $c^* \in \mathcal{C}_c^\epsilon$ is similar to c but each input $i \in I_S$ is replaced by an input in V_i^ϵ. Formally, for each configuration c leading to the input set I_S, $\mathcal{C}_c^\epsilon$ is the set of configurations defined as follows: $c^* = (D_{c^*}, B_{c^*}, b_{c^*}) \in \mathcal{C}_c^\epsilon$ if and only if for each $i \in I_S$, there exist $i_\epsilon^* \in V_i^\epsilon$ satisfying

$$\begin{cases} D_{c^*} = D_c, \\ B_{c^*} = \{i_\epsilon^*, i \in B_c\}, \\ \text{For each } i \in I_S \setminus B_c,\ b_{c^*}\left(b_c^{-1}(i)\right) = i_\epsilon^*. \end{cases}$$

With this definition, we have

$$P(c/S) = \lim_{\epsilon \to 0} P\left(\mathcal{C}_c^\epsilon / S\right) = \lim_{\epsilon \to 0} \frac{P\left(\mathcal{C}_c^\epsilon / O_S\right)}{\sum_{c' \in \mathcal{C}_S} P\left(\mathcal{C}_{c'}^\epsilon / O_S\right)}. \tag{4.2}$$

In what follows, we study the behavior of $P(\mathcal{C}_c^\epsilon / O_S)$ when ϵ goes to 0. We will always work conditionally on the realization of the output set O_S but we will keep this conditioning implicit and write $P(\mathcal{C}_c^\epsilon)$ instead of $P(\mathcal{C}_c^\epsilon / O_S)$ in order to simplify the notation. The study of $P(\mathcal{C}_c^\epsilon)$ will involve the probability for a particle to die in the unobserved region but also the probability that a particle born in this unobserved region enters the observed one in a given spatiotemporal neighborhood V_i^ϵ.

Furthermore, we assume that the particles born in the unobserved region, enter the observed one with a constant rate τ_α and with a uniform distribution on $\{-l\} \times [0, H]$. This is consistent with the fact that the particles are born with constant rate λ and appear uniformly at random on the membrane surface. Therefore, denote by N_α the Poisson process of intensity τ_α counting the number of inputs involved by particles born in the unobserved region.

Consider an output $o \in O_S$ and the possibility for the particle involved in o to die in the unobserved region. We have the following proposition (see [Sch15], [Twe45], [Wal73]).

Proposition 1 *Given the particle motion model as Brownian motion with drift as described in equation 4.1, the first passage time noted as* T_l *on the entrance line* $\{-l\} \times [0, H]$ *of a particle starting at position* $z_0 = (0, y_0)$ *for some* $y_0 \in [0, H]$ *follows a law of inverse Gaussian, that is,* $T_l \sim IG\left(\frac{l_u}{\mathbf{v}_x}, (\frac{l_u}{\sigma_x})^2\right)$ *where* $l_u := L - l$ *is the length of the unobserved region.*

Recall that if $X \sim IG(\mu, \lambda)$*, then* $X \geq 0$ *almost surely, and for each* $x \geq 0$*,*

$$P(X \leq x) = \int_0^x \sqrt{\frac{\lambda}{2\pi y^3}} \exp\left(-\frac{\lambda (y - \mu)^2}{2\mu^2 y}\right) dy. \tag{4.3}$$

In our framework, the event corresponding to the death of a particle with life duration T_d following an exponential law of parameter τ_d in the unobserved region is precisely $\{T_d < T_l\}$. Hence, we can derive an explicit expression of $P(\mathcal{C}_c^\epsilon)$.

Assume ϵ small enough so that for each $i, i' \in I_S$, $T_i^\epsilon \cap T_{i'}^\epsilon = \emptyset$. For a given configuration c and a given $\epsilon > 0$, we will write $\mathcal{C}_c^\epsilon = (D_c, \mathcal{B}_c^\epsilon, \flat_c^\epsilon)$ with $\mathcal{B}_c^\epsilon = \{B_{c^*}, c^* \in \mathcal{C}_c^\epsilon\}$ and $\flat_c^\epsilon = \{b_{c^*}, c^* \in \mathcal{C}_c^\epsilon\}$.

Due to the independent behavior of the particles, we have the following decomposition:

$$P(\mathcal{C}_c^\epsilon) = P(D_c) P(\mathcal{B}_c^\epsilon) P(\flat_c^\epsilon). \tag{4.4}$$

We can then compute separately the probabilities of events D_c, $\mathcal{B}_c^\epsilon$, and $\flat_c^\epsilon$. First, note that we can assume without loss of generality that each output $o \in D_c$ starts at time $t_o = 0$ and that only the position $y_o \in [0, H]$ fluctuates with o, but with no influence on T_d or T_l. Moreover, the loss of memory property of the exponential law ensures that the life duration T_d of the particle after the output o still follows an exponential law of parameter τ_d.

Since all outputs behave identically and independently, we have $P(D_c) = P(T_d < T_l)^{|D_c|}$,where $|D_c|$ stands for the cardinal of D_c. According to Proposition 1, and since T_d and T_l are independent, we have

$$\begin{aligned} P(T_d < T_l) &= \int_0^{+\infty} \int_0^{t_l} f_{T_d}(t_d) f_{T_l}(t_l) dt_d \, dt_l, \\ &= \int_0^{+\infty} \int_0^{t_l} \tau_d e^{-\tau_d t_d} \frac{l_u}{\sigma_x \sqrt{2\pi t_l^3}} \exp\left(-\frac{(\mathbf{v}_x t_l - l_u)^2}{2\sigma_x^2 t_l}\right) dt_d \, dt_l, \\ &= \int_0^{+\infty} \frac{l_u \left(1 - e^{-\tau_d t_l}\right)}{\sigma_x \sqrt{2\pi t_l^3}} \exp\left(-\frac{(\mathbf{v}_x t_l - l_u)^2}{2\sigma_x^2 t_l}\right) dt_l, \end{aligned} \tag{4.5}$$

where f_{T_d} and f_{T_l}, respectively stand for the density functions of T_d and T_l.

Now, consider the event $\mathcal{B}_c^\epsilon$. We call "spontaneous input" an input related to a particle born in the unobserved region that has never been observed. The set $\mathcal{B}_c^\epsilon$ is defined so that, for each input $i \in B_c$, we have exactly one "spontaneous input" appearing during the time interval T_i^ϵ, with a position in H_i^ϵ. Moreover, outside $\cup_{i \in B_c} T_i^\epsilon$, there is no "spontaneous input." Formally, we have

$$\mathcal{B}_c^\epsilon = \left\{ N_\alpha \left([0, T_S] \setminus \bigcup_{i \in B_c} T_i^\epsilon \right) = 0 \right\} \bigcap \left(\bigcap_{i \in B_c} \left(\{ N_\alpha(T_i^\epsilon) = 1 \} \cap H_i^\epsilon \right) \right), \tag{4.6}$$

where N_α is a Poisson process of intensity τ_α associated to the counting of inputs involved by particles born in the unobserved region on the time interval $[0, T_S]$. In order to simplify the notation, H_i^ϵ denotes also the event of "spontaneous" appearance of an input i in H_i^ϵ. This event is independent of the process N_α, and since the "spontaneous inputs" appear uniformly on $[0, H]$, we have $P(H_i^\epsilon) = \frac{\epsilon}{H}$.

Meanwhile, for any time interval I, $N_\alpha(I)$ follows a Poisson law of parameter $\tau_\alpha |I|$ where $|I|$ denotes the length of the interval I. Since ϵ is small enough so that for each $i, i' \in I_S$, $T_i^\epsilon \cap T_{i'}^\epsilon = \emptyset$, $N_\alpha(T_i^\epsilon)$ and $N_\alpha(T_{i'}^\epsilon)$ are independent. Consequently, we can compute $P(\mathcal{B}_c^\epsilon)$ as follows:

$$P(\mathcal{B}_c^\epsilon) = e^{-\tau_\alpha (T_S - |B_c| \epsilon)} \left(\epsilon \tau_\alpha e^{-\epsilon \tau_\alpha} \frac{\epsilon}{H} \right)^{|B_c|} = \left(\frac{\epsilon^2 \tau_\alpha}{H} \right)^{|B_c|} e^{-\tau_\alpha T_S}. \tag{4.7}$$

Finally, consider the event $\flat_c^\epsilon$. For each input $i \in I_S \setminus B_c$, we denote by $\{o_c^i \to V_i^\epsilon\}$ the survival event of the particle involved in the output $o_c^i = b_c^{-1}(i)$ in the unobserved region which appears in the spatiotemporal neighborhood V_i^ϵ. Since the particles behave independently, we have

$$P(\flat_c^\epsilon) = \prod_{i \in I_S \setminus B_c} P\Big(\{ o_c^i \to V_i^\epsilon \} \Big). \tag{4.8}$$

In what follows, we consider a given input $i \in I_S \setminus B_c$ and its related output $o = b_c^{-1}(i)$. Defining $s_i = t_i - t_o$ and $h_i = y_i - y_o$ allows us to center the situation around the output o in the following way. A particle born at time 0 in position $z_0 = (0,0)$ has a life duration T_d following an exponential law of parameter τ_d. During its lifetime, the position of the particle is driven by a Brownian motion with drift $\mathbf{Z}_t = (X_t, Y_t)$: $\mathbf{Z}_t = \mathbf{v}t + \mathbf{\Sigma}\mathbf{B}_t$, where $\mathbf{B}_t$ is a 2D Wiener process and $\mathbf{v}$ and $\mathbf{\Sigma}$ are given in (4.1). Define T_l to be the first reaching time of $l_u = L - l$ of the process X_t. The event $\{o \to V_i^\epsilon\}$ can now be written as follows:

$$\{o_c^i \to V_i^\epsilon\} = \{T_d > T_l\} \bigcap \left\{T_l \in \left[s_i - \frac{\epsilon}{2}, s_i + \frac{\epsilon}{2}\right]\right\} \bigcap \left\{Y_{T_l} \in \left[h_i - \frac{\epsilon}{2}, h_i + \frac{\epsilon}{2}\right]\right\}.$$

This expression corresponds exactly to the fact that in order to realize $\{o_c^i \to V_i^\epsilon\}$ the particle needs to have a life duration longer than its first reaching time of l_u and to appear in the spatiotemporal neighborhood $\left[s_i - \frac{\epsilon}{2}, s_i + \frac{\epsilon}{2}\right] \times \left[h_i - \frac{\epsilon}{2}, h_i + \frac{\epsilon}{2}\right]$. Furthermore, T_d follows an exponential law of parameter τ_d, Y_t follows a Gaussian law of parameters $\mathbf{v}_y t$ and $\sigma_y^2 t$ and $T_l \sim IG\left(\frac{l_u}{\mathbf{v}_x}, (\frac{l_u}{\sigma_x})^2\right)$. Moreover, due to the fact that $\mathbf{\Sigma}$ is diagonal, the process Y_t is not only independent of T_d but also of T_l. This allows us to write

$$P\left(\{o_c^i \to V_i^\epsilon\}\right) = \int_{s_i - \frac{\epsilon}{2}}^{s_i + \frac{\epsilon}{2}} f_{T_l}(t_l) \left(\int_{t_l}^{+\infty} f_{T_d}(t_d) \left(\int_{h_i - \frac{\epsilon}{2}}^{h_i + \frac{\epsilon}{2}} f_{Y_{t_l}}(y)dy\right) dt_d\right) dt_l.$$

As the two integrals involve a small domain of size ϵ, $P\left(\{o_c^i \to V_i^\epsilon\}\right) \sim \mathcal{O}(\epsilon^2)$, and

$$\lim_{\epsilon \to 0} \frac{P\left(\{o_c^i \to V_i^\epsilon\}\right)}{\epsilon^2} = f_{T_l}(s_i) f_{Y_{s_i}}(h_i) \int_{s_i}^{+\infty} f_{T_d}(u)du \tag{4.9}$$

$$= \frac{l_u}{\sigma_x \sqrt{2\pi s_i^3}} \exp\left(-\frac{(\mathbf{v}_x s_i - l_u)^2}{2\sigma_x^2 s_i}\right) \frac{1}{\sigma_y \sqrt{2\pi s_i}} \exp\left(-\frac{(h_i - \mathbf{v}_y s_i)^2}{2\sigma_y^2 s_i}\right) e^{-\tau_d s_i}$$

$$= \frac{l_u}{2\pi \sigma_x \sigma_y s_i^2} \exp\left(-\frac{(\mathbf{v}_x s_i - l_u)^2}{2\sigma_x^2 s_i} - \frac{(h_i - \mathbf{v}_y s_i)^2}{2\sigma_y^2 s_i} - \tau_d s_i\right).$$

For each configuration c, we calculate the likelihood $Q(c)$ of the configuration c as follows:

$$Q(c) := \lim_{\epsilon \to 0} \frac{P(\mathcal{C}_c^\epsilon)}{\epsilon^{2|I_S|}}.$$

From (4.4), (4.5), (4.7), (4.8), and 4.9, we finally obtain the likelihood

$$\begin{aligned} Q(c) \; = \; & \left(\frac{\tau_\alpha}{H}\right)^{|B_c|} e^{-\tau_\alpha T_S} \left(\int_0^{+\infty} \frac{l_u \left(1 - e^{-\tau_d t_l}\right)}{\sigma_x \sqrt{2\pi t_l^3}} \exp\left(- \frac{(\mathbf{v}_x t_l - l_u)^2}{2\sigma_x^2 t_l} \right) dt_l \right)^{|D_c|} \\ & \times \prod_{i \in I_S \setminus B_c} \left[\frac{l_u}{2\pi \sigma_x \sigma_y s_i^2} \exp\left(- \frac{(\mathbf{v}_x s_i - l_u)^2}{2\sigma_x^2 s_i} - \frac{(h_i - \mathbf{v}_y s_i)^2}{2\sigma_y^2 s_i} - \tau_d s_i \right) \right]. \end{aligned} \tag{4.10}$$

Note that the limit when ϵ goes to 0 of $\frac{P(\mathcal{C}_c^\epsilon)}{\epsilon^{2|I_S|}}$ is well defined, strictly positive, and that the exponent $2|I_S|$ does not depend on the configuration c.

Recalling (4.2), this allows us to write

$$P(c/S) = \frac{Q(c)}{\sum_{c' \in \mathcal{C}_S} Q(c')} \tag{4.11}$$

and as a consequence, we have

$$\hat{c} = \underset{c \in \mathcal{C}_S}{\operatorname{argmax}} \{Q(c)\}. \tag{4.12}$$

4.2.3 Maximum likelihood and optimal configuration

The aim of this section is to identify the configuration c corresponding to the maximal likelihood $Q(c)$ (see (4.10)). Define

$$\beta := -\log\left(\frac{\tau_\alpha}{H}\right),$$

$$\delta := -\log\left(\int_0^{+\infty} \frac{l_u \left(1 - e^{-\tau_d t_l}\right)}{\sigma_x \sqrt{2\pi t_l^3}} \exp\left(- \frac{(\mathbf{v}_x t_l - l_u)^2}{2\sigma_x^2 t_l} \right) dt_l \right)$$

and for each configuration c and each $i \in I_S \setminus B_c$

$$\gamma_c^i := -\log\left[\frac{l_u}{2\pi \sigma_x \sigma_y s_i^2} \exp\left(- \frac{(\mathbf{v}_x s_i - l_u)^2}{2\sigma_x^2 s_i} - \frac{(h_i - \mathbf{v}_y s_i)^2}{2\sigma_y^2 s_i} - \tau_d s_i \right) \right].$$

It follows that

$$\begin{aligned} \hat{c} \; = \; \underset{c \in C}{\operatorname{argmax}}\, Q(c) \; & = \; \underset{c \in C}{\operatorname{argmin}} - \log\left(Q(c)\right) \\ & = \; \underset{c \in C}{\operatorname{argmin}} \left(\beta |B_c| + \delta |D_c| + \sum_{i \in I_S \setminus B_c} \gamma_c^i \right). \end{aligned} \tag{4.13}$$

This decomposition allows us to consider a linear optimization problem where β represents the cost of the spontaneous birth of an input, δ the cost of the death of an output and γ_c^i the cost of the connection between the output $b_c^{-1}(i)$ and the input i. The cost of connection can be defined for any couple $(o,i) \in O_S \times I_S$ as

$$\gamma_o^i := -\log\left[\frac{l_u}{2\pi\sigma_x\sigma_y s_{o,i}^2}\exp\left(-\frac{(\mathbf{v}_x s_{o,i} - l_u)^2}{2\sigma_x^2 s_{o,i}} - \frac{(h_{o,i} - \mathbf{v}_y s_{o,i})^2}{2\sigma_y^2 s_{o,i}} - \tau_d s_{o,i}\right)\right],$$

where $s_{o,i} := t_i - t_o$, $h_{o,i} = y_i - y_o$ and the convention $\gamma_o^i = +\infty$ if $t_i \leq t_o$.

In order to write in a canonical way this linear optimization problem, we associate to each configuration c a family of coefficients $(c^{o,i})_{(o,i)\in O_S\times I_S}$ such that $c^{o,i} = 1$ if $b_c(o) = i$ and $c^{o,i} = 0$ if $b_c(o) \neq i$. Since an output can be connected to at most one input, for each $o \in O_S$, $\sum_{i\in I_S} c^{o,i} \in \{0,1\}$ and $\sum_{i\in I_S} c^{o,i} = 0$ corresponds to the death of the output o. Similarly, for each $i \in I_S$, $\sum_{o\in O_S} c^{o,i} \in \{0,1\}$ and $\sum_{o\in O_S} c^{o,i} = 0$ corresponds to the fact that the input i is a "spontaneous input."

Our optimization problem is then equivalent to finding the family of coefficients $(c^{o,i})_{(o,i)\in O_S\times I_S}$ that minimizes the quantity

$$\beta\left(\sum_{i\in I_S}\left(1 - \sum_{o\in O_S} c^{o,i}\right)\right) + \delta\left(\sum_{o\in O_S}\left(1 - \sum_{i\in I_S} c^{o,i}\right)\right) + \sum_{o\in O_S}\sum_{i\in I_S}\gamma_o^i c^{o,i}$$

or, equivalently

$$K(c) := \sum_{o\in O_S}\sum_{i\in I_S}\left(\gamma_o^i - \beta - \delta\right)c^{o,i} \text{ s.t. } \begin{cases} \forall o \in O_S,\, \forall i \in I_S,\, c^{o,i} \in \{0,1\}, \\ \forall o \in O_S,\, \sum_{i\in I_S} c^{o,i} \in \{0,1\}, \\ \forall i \in I_S,\, \sum_{o\in O_S} c^{o,i} \in \{0,1\}. \end{cases} \tag{4.14}$$

In order to avoid having infinite costs γ_o^i when $t_i \leq t_o$,, we can also impose $c^{o,i} = 0$ if $t_i \leq t_o$. Actually the problem (4.14) is a conventional linear optimization problem which can be solved by applying the CPLEX Linear Programming solver https://www.ibm.com/analytics/cplex-optimizer.

The configuration $\hat{c}$ is then the solution of the optimization problem (4.14) and corresponds to the most likely configuration given the sample S. In order to complete the study, we propose computing the following most likely configurations in a recurrent way by solving (4.14) with additional constraints ensuring that the solution is different from the previous ones. In other words we define recursively the sequence $(c_n)_{n\in\mathbb{N}}$ in the following way:

— $c_1 := \hat{c}$,

— $\forall n \geq 2$, c_n solves (4.14) with the $n-1$ additional constraints

$$\forall k \in \{1,\ldots,n-1\},\ \sum_{o\in O_S}\sum_{i\in I_S}\left[c_n^{o,i}(1 - c_k^{o,i}) + (1 - c_n^{o,i})c_k^{o,i}\right] \geq 1. \tag{4.15}$$

With this definition, c_n is then the nth most likely configuration. When n is greater than the number n_S of configurations compatible with the sample S, the constraints are impossible to satisfy. In other words this sequence is well defined up to n_S.

4.2.4 Estimation of parameters

Several parameters are involved in our computational approach. In this section, we propose clues to set these parameters. First, the parameters $\mathbf{v}$ and $\boldsymbol{\Sigma}$ can be estimated with classical maximum likelihood estimation procedures.

Second, we propose an estimator $\hat{\tau}_d$ of τ_d as explained below. The sample S can be considered as a set of points $p = (t_p, \mathbf{Z}_p)$ observed at time t_p and position $\mathbf{Z}_p = (X_p, Y_p)$ grouped in clusters s corresponding to tracklets of trajectories. The death of a particle in the observed region is detected in S for each point $p \in S$ for which the associated tracklet s_p has no successor point at time $t_p + \Delta t$. In order to be sure that the absence of successor is effectively due to the death of a particle and not to a particle leaving the observed region, we restrict the analysis to a region excluding a neighborhood of the border. However, in this neighborhood we can check the existence of successors for points in the restricted region. We denote by $S_r \subset S$ the sample of points in the restricted region. For each point $p \in S_r$, we denote by D_p the event corresponding to the absence of successor for p. This corresponds to the fact that the particle involved in p died during the time interval $[t_p, t_p + \Delta t]$. Since the life duration T_d of a particle follows an exponential law of parameter τ_d, and the absence of memory property of the exponential law, we have

$$P(D_p) = P(T_d \in [0, \Delta t]) = 1 - e^{-\tau_d \Delta t}. \tag{4.16}$$

Hence, we define our estimator $\hat{\tau}_d$ as

$$\hat{\tau}_d = \frac{1}{\Delta t |S_r|} \sum_{p \in S_r} 1[D_p], \tag{4.17}$$

where $|S_r|$ stands for the number of points in S_r and $1[\cdot]$ denotes the indicator function. Due to the absence of memory property of the exponential law, the random variables $1[D_p]$ are i.i.d. As $|S_r|$ goes to $+\infty$, the strong law of large numbers yields to

$$\lim_{|S_r| \to \infty} \hat{\tau}_d = \frac{1 - e^{-\tau_d \Delta t}}{\Delta t} \qquad a.s.$$

The justification of this choice for $\hat{\tau}_d$ relies in the following almost sure convergence:

$$\lim_{\Delta t \to 0} \lim_{|S_r| \to \infty} \hat{\tau}_d = \tau_d \qquad a.s. \tag{4.18}$$

Our estimator $\hat{\tau}_d$ is then consistent as Δt is small enough. Moreover, since the variables $1[D_p]$ are i.i.d Bernoulli random variables, we can calculate the related confidence interval. If q_α denotes the α-quantile of the standard normal distribution, we have the following confidence interval of level α for $\frac{1-e^{-\tau_d \Delta t}}{\Delta t}$:

$$CI_\alpha = \left[\hat{\tau}_d - q_\alpha \sqrt{\frac{\hat{\tau}_d \left(\frac{1}{\Delta t} - \hat{\tau}_d \right)}{|S_r|}}, \hat{\tau}_d + q_\alpha \sqrt{\frac{\hat{\tau}_d \left(\frac{1}{\Delta t} - \hat{\tau}_d \right)}{|S_r|}} \right]. \tag{4.19}$$

If Δt is small enough, we get a good approximation of a confidence interval of level α for τ_d since

$$\lim_{\Delta t \to 0} \frac{1 - e^{-\tau_d \Delta t}}{\Delta t} = \tau_d.$$

Now, we describe the estimation procedure for the rate τ_α of "spontaneous inputs" induced by particles born in the unobserved region $[-L, -l] \times [0, H]$ and reached the border $\{-l\} \times [0, H]$. We assume here that the parameters $\mathbf{v}$, $\mathbf{\Sigma}$ and τ_d are known, keeping in mind that in practice estimators are used instead. As introduced earlier, L is the perimeter of the cylinder, l is the length of the observed region, and $l_u = L - l$ is the length of the unobserved region. For a given length x, we denote by N_x the number of tracklets born in the region $] - x, 0] \times [0, H]$ and reached the border $\{0\} \times [0, H]$. Accordingly, $\frac{N_{l_u}}{T_S}$ is a consistent estimator of τ_α since the dynamics are assumed to be homogeneous on the surface of the cylinder. Our aim is actually to build an estimator for τ_α in the case where $l_u > l$ which prevents us from computing directly N_{l_u}. Therefore, we compute N_l by taking the whole observed region into account, and denote by S_l^* the set of tracklets having an input in $\{-l\} \times [0, H]$ and an output in $\{0\} \times [0, H]$. For each tracklet $s \in S_l^*$ and each length $x \in [0, l_u]$, we denote by B_s^x the event corresponding to the birth of s within $] - l - x, -l] \times [0, H]$. Let $l_e = l_u - l$ be the length of the extended zone $[-l_u, -l] \times [0, H]$. We are now interested in the realization of the events $B_s^{l_e}$.

In Fig. 4.3, $N_l = 2$ corresponds to tracks #1 and #4, $S_l^* = \{2, 5\}$, and the event $B_2^{l_e}$ is realized while $B_5^{l_e}$ is not.

Note that since the particles have the same independent dynamics, $P\left(B_s^x\right)$ does not depend on s. For $x < l$, this probability can easily be estimated as follows:

$$\hat{p}_x = \frac{N_x}{|S_o|},$$

where S_o is the set of tracklets having an output in $\{0\} \times [0, H]$. The strong law of large numbers yields a consistent estimator and allows us, in the case where $l_e < l$, to define our estimator $\hat{\tau}_\alpha$ as follows:

$$\hat{\tau}_\alpha = \frac{N_l + \hat{p}_{l_e} |S_l^*|}{T_S}. \tag{4.20}$$

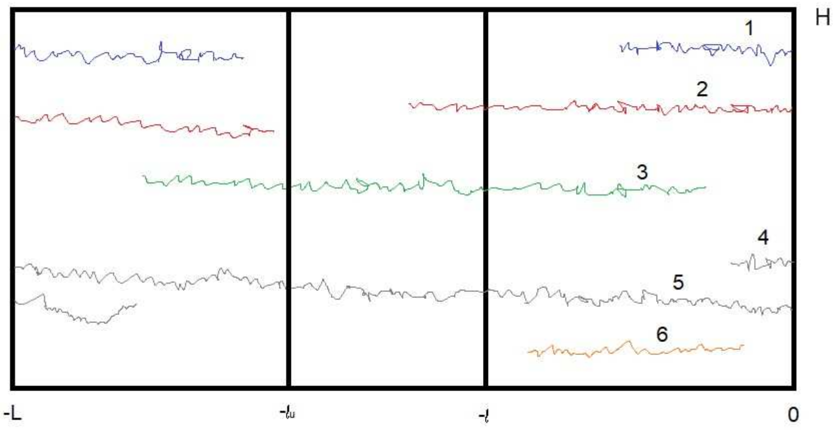


FIGURE 4.3 – An artificially constructed zone $]-l_u, 0] \times [0, H]$ having the same size as the unobserved region $]-L, -l] \times [0, H]$. The observed region is $]-l, 0] \times [0, H]$; as the width of the invisible part is l_u, the extended zone has width $l_e = l_u - l$.

Intuitively, this estimator amounts to counting the number of particles reaching $\{0\} \times [0, H]$ with weight 1 for each tracklet that we actually saw being born in the observed region and with weight $\hat{p}_{l_e}$ for each spontaneous input that appeared in $\{-l\} \times [0, H]$. Note that, as $N_{l_u} = N_l + \sum_{s \in S_l^*} 1[B_s^{l_e}]$, $\hat{\tau}_\alpha$ is an unbiased estimator of τ_α. Moreover, if we assume that the number of observed tracklets grows linearly with the observation time T_S, this estimator is consistent when T_S goes to $+\infty$.

Now, we consider the case $l < l_e < 2l$ which can easily be extended to the general case $l < l_e$. Consider $s \in S_l^*$ and denote for each interval $J \subset [-L, 0]$ the event B_s^J where the tracklet s is born in the region $J \times [0, H]$. The event $B_s^{l_e}$ can be decomposed as follows:

$$B_s^{l_e} = B_s^{[-2l,-l]} \bigcup \left(\overline{B_s^{[-2l,-l]}} \cap B_s^{[-l_u,-2l]} \right).$$

The loss of memory and homogeneity properties of the dynamics lead to the following estimator $\hat{p}_{l_e}$:

$$\hat{p}_{l_e} := \hat{p}_l + (1 - \hat{p}_l)\, \hat{p}_{l_e - l}.$$

4.2.5 Limits of the model

The main assumptions in this work are homogeneity in time and space, induced by the constant death and birth rates, as well as constant speed and noise. While these assumptions lead to a simple model and allow a reasonably technical study, it is natural to question it. The main reason for this choice is that it corresponds to uniform laws when we have no reason to prioritize one specific behavior in particular.

Note that a similar study can be made with different speeds among trajectories. This can be done by classifying the trajectories according to their speeds and applying the present procedure to each class. This would lead to the same estimation procedure with smaller datasets but theoretical results will still hold.

We then discuss the homogeneity in time, for which the most questionable assumption is the constant death rate that could possibly depend on the position or on the age of the particle. Concerning the dependence in space, this modification would lead to the estimation of a function of the position instead of the simple constant τ_d. From a practical point of view, this would increase the dimension of the parameter to estimate, with the same size of dataset. From a theoretical point of view a more technical study can be made as long as we assume the death function rate (depending on the position) constant on each tracklet $\{y\}\times]-L,0]$ in order to overcome the issue of partial observation.

Concerning the dependence in time, the assumption that the death rate depends on the age of the particle prevents us from proposing a similar study. Indeed, due to partial observation, the age of each particle entering the observed region is unknown and cannot be estimated.

4.3 Simulation study

In this section, we present a series of experiments performed on synthetic datasets. These experiments aim to evaluate and analyze the sensitivity of the reconstruction procedure when the characteristics of the dynamics as well as the spatio-temporal sampling resolution of observations vary. In addition, to demonstrate the potential of our procedure, these experiments might also be useful for the design of the experimental setting for images acquisition. The reconstruction procedure has been implemented in MATLAB ver. R2018b. The codes are available on Github `https://github.com/atrubuil/ReconstructionOfTruncatedTrajectories`.

4.3.1 Generation of trajectories

Trajectories are generated on a rectangular unwrapped cylindrical surface of size $[0, L] \times [0, H]$ (Fig. 4.4). In our experiments, we set $L = 50, H = 30$. The initial position of each trajectory is drawn from uniform distribution on the surface. Time duration T between two births follows an exponential law with birth rate parameter λ. At each birth, the intrinsic properties of a trajectory i are given, such as velocity $\mathbf{v}_i$, variance $\mathbf{\Sigma}_i$, and lifetime T_d^i. The lifetime T_d follows an exponential law, with the same death rate τ_d for all trajectories in the whole simulated image sequence. The

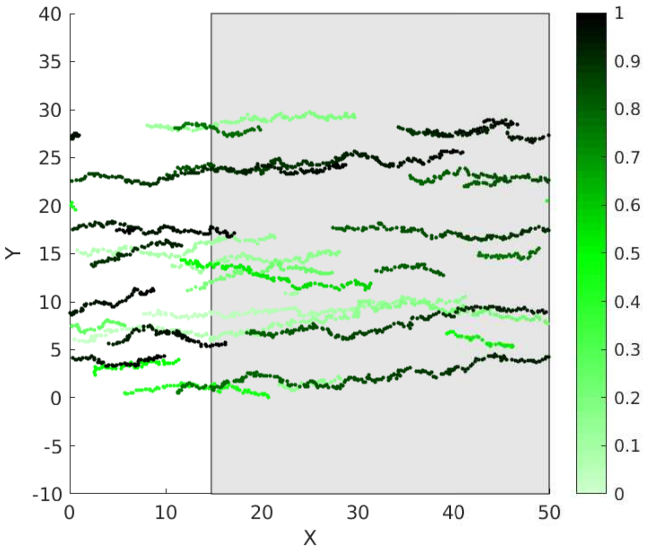


FIGURE 4.4 – **A set of simulated trajectories during 2.5 minutes (in stationary regime).** X(resp., Y) axis represents the unfolded circumferential (resp., main) direction of the cylinder. Colors from light to dark green represent time evolution. The shadowed area corresponds to the unobserved region and the white area corresponds to the ROO.

drift $\mathbf{v}_i = (v_{xi}, v_{yi})$ and noise

$$\mathbf{\Sigma}_i = \begin{bmatrix} \sigma_{xi} & 0 \\ 0 & \sigma_{yi} \end{bmatrix}$$

are set to be constant along one given trajectory.

According to the assumptions made on real biological context, unless otherwise stated, it is set by default, $\theta = 0.01 (\approx 0.6°)$ is the angle between the direction of motion of particle and the X direction, $v_y = \tan(\theta) v_x$, $\sigma_x = \sigma_y = \sigma$, $v_x = 0.6$, $\sigma = 0.2$, $\lambda = 0.03$, $\tau_d = 0.005$. The time interval between two images is $\Delta t = 0.25$. As known, the theoretical depth of the observation field of TIRFM is $200nm$ and the diameter of the bacteria cell is $1\mu m$. Therefore the width of the ROO l is set to 14.76 and that of the unobserved region $l_u = 35.24$ (unit in pixel, note that in TIRF images 1 pixel $\approx 64nm$).

As there is no particle on the surface at the beginning, the simulated set of trajectories needs some warm-up time to reach the stationary regime, where the law of the number of trajectories does not depend on time. The assumed dynamic process is a special case of birth and death process. As a known result [Kar14], the expectation of the trajectories number N during stationary regime is $E(N) = \frac{\lambda}{\tau_d}$. To ensure that the dynamics are in a stationary regime, the images sequence is simulated for around 2 hours (Fig. 4.5).

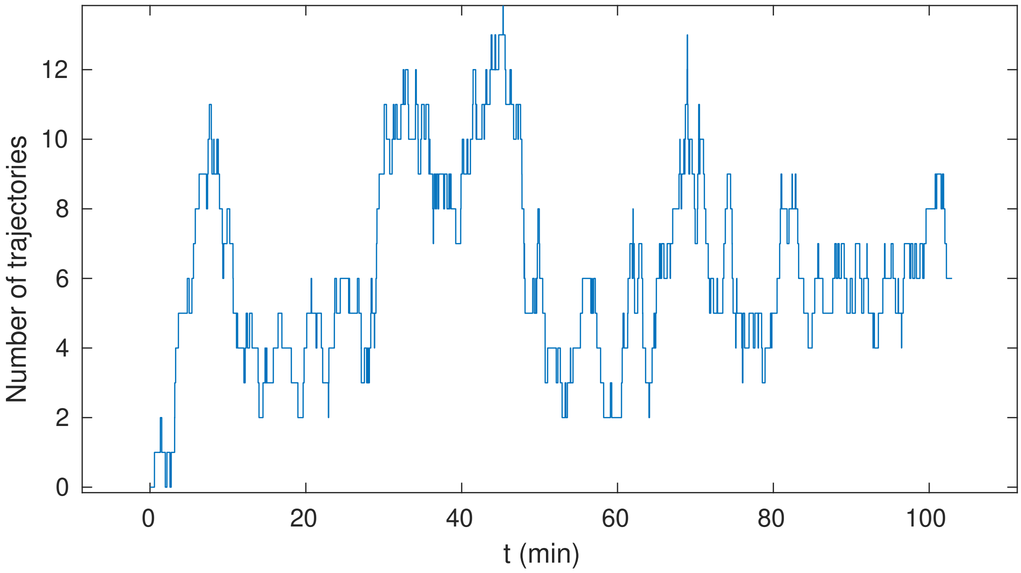


FIGURE 4.5 – Fluctuations of the number of trajectories w.r.t. time. At around $t = 20$ minutes, the trajectories number fluctuates around the theoretical expectation value 6.

4.3.2 The "adjusted Rand index" for the evaluation of connection results

Given the true and estimated class assignments, we compute the so-called adjusted Rand Index (ARI) to evaluate similarity or consensus between the two sets. The ARI is the corrected-for-chance version of the Rand index (RI). It is scored exactly 1 when the two sets are identical, close to 0 for random labeling. It could be negative when the index is lower than the expectation under random labeling. More precisely, let G and K be the true and estimated assignments, respectively. Let us define a and b as: a the number of pairs of elements that are in the same class in G and in the same class in K, b the number of pairs of elements that are in different classes in G and in different classes in K. The raw (unadjusted) RI is then given by

$$RI = \frac{a+b}{C_2^M}, \tag{4.21}$$

where C_2^M is the total number of possible pairs in the dataset (without ordering) of size M. The RI score does not guarantee that random assignments will get a value close to zero. This is especially true if the number of clusters has the same order of magnitude as the number of samples. To overcome this difficulty, we prefer to consider the ARI defined by [Ran71]:

$$ARI = \frac{RI - E(RI)}{1 - E(RI)}. \tag{4.22}$$

Here $E(RI)$ denotes the expectation of the RI where the estimated assignment K is replaced by an assignment chosen uniformly at random. This means that the assignment procedure does not do better than random assignment if the ARI score is zero, and that it does worse than random if $ARI < 0$.

4.3.3 Experimental results

The good performance of the connection procedure relies on the estimation of the characteristics of the dynamics: the speed, $\mathbf{v}$, the diffusion variance, $\mathbf{\Sigma}$, the arrival rate τ_α and the death rate τ_d, as these quantities are used in the calculation of the likelihood (4.13). Here we evaluated the impact of spatio-temporal sampling $(l/l_u, T_S)$ on the estimators and the impact of parameters of the dynamics $(\mathbf{v}, \mathbf{\Sigma}, \tau_\alpha, \tau_d)$ on the accuracy of the reconstruction.

4.3.3.1 The estimator $\hat{\tau}_\alpha$ performs well, in the case of realistic 2D-TIRF, where $\frac{l}{l_u} \approx 0.42$

The estimator $\hat{\tau}_\alpha$ is proposed in 4.20. Here we test how it performs with different spatio-temporal sampling $(l/l_u, T_S)$, and different birth rate λ and death rate τ_d.

By its definition in section 4.2.2, τ_α, the rate of "spontaneous input" induced by particles born in the unobserved region and reach the border of the ROO, is not a preset parameter. A reference of the "true" value of τ_α is given by $\frac{N_{l_u}}{T_s}$, where N_{l_u} denotes the number of tracklets born in the region $]-l_u, 0] \times [0, H]$ and reached the border $\{0\} \times [0, H]$, l_u is the width of the unobserved region.

Next, we test the robustness of the estimator $\hat{\tau}_\alpha$ w.r.t. l/l_u (Fig. 4.6). To avoid the influence of T_S on the consistency of the estimator, T_S is set to be 30 minutes. We can conclude that, naturally, the larger the observed area is, the better is the performance of the estimator $\hat{\tau}_\alpha$ is. In the case of the simulation of the real situation, where $l/l_u = 0.42$, the estimator works reasonably well.

Next, we test the robustness of $\hat{\tau}_\alpha$ w.r.t. T_S (Fig. 4.7). This test is essential because in reality it is impossible to use a 30-minute movie, because of technical issues like photobleaching of fluorophores and natural growth in living samples. At this stage, the proportion of observed and unobserved region l/l_u is set to 0.42. T_S varies from 2.5 to 30 minutes. In Fig. 4.7, it can be noticed that the reference "ground truth" of τ_α (blue boxes) decreases as T_S lengthens. Actually, the reference is only a pseudo "ground truth". It is sensible to T_S when T_S is small and it converges as $T_S \to \infty$. The distributions of counted "ground truth" and estimator become close to each other for $T_S \geq 10$ minutes.

The absolute value of τ_α depends on λ and τ_d. Fig. 4.8 displays for different combinations of λ and τ_d, the estimations of $\hat{\tau}_\alpha$ by 5-minute movies (magenta) and 30-minute movies (blue).

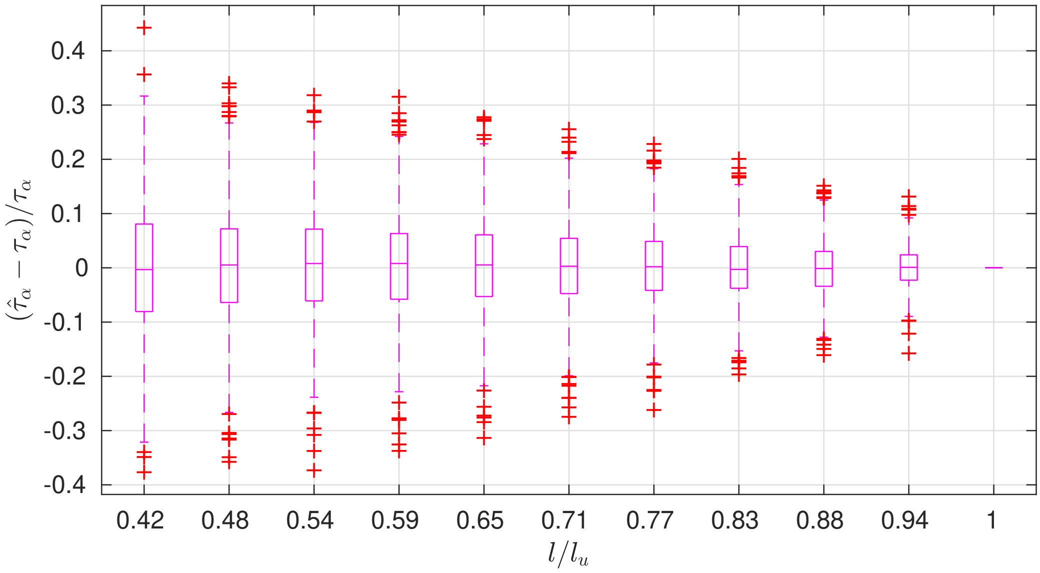


FIGURE 4.6 – At $l/l_u = 0.42$, which corresponds to the realistic situation in 2D-TIRF, more than half of the trials presents a relative error smaller than 10%. The proposed estimator $\hat{\tau}_\alpha$ is unbiased and the variance decreases as l/l_u increases.

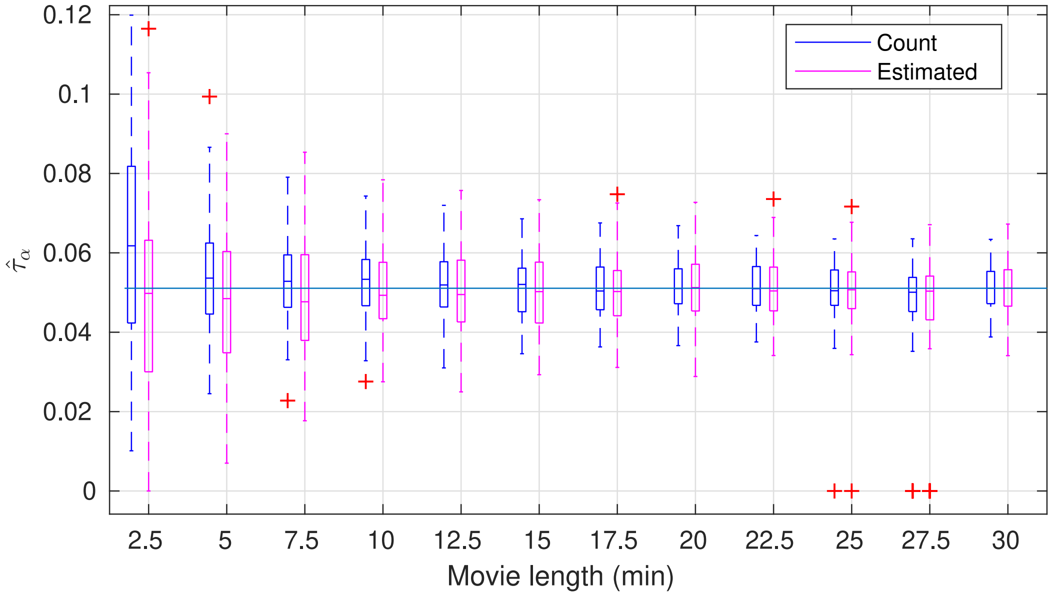


FIGURE 4.7 – The counted "ground truth" τ_α and the estimated $\hat{\tau}_\alpha$ obtained by movies of different duration, varying from 2.5 to 30 minutes. Blue boxes (resp., magenta boxes) correspond to the "ground truth", i.e. the counted value (resp., the estimator $\hat{\tau}_\alpha$). The blue horizontal line represents the "ground truth" value when $T_S = 30$ minutes.

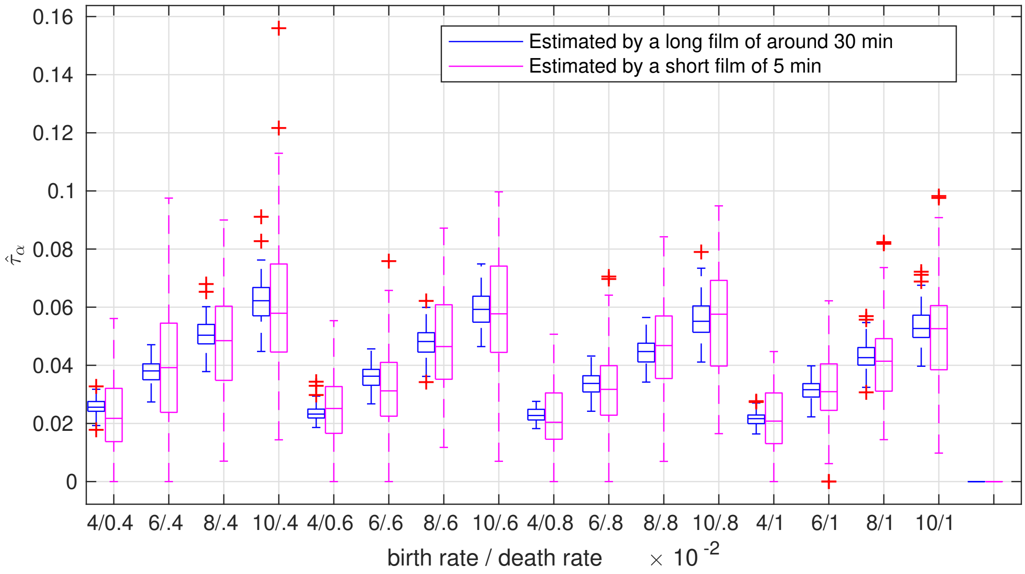


FIGURE 4.8 – The estimation of arrival rate τ_α w.r.t. different λ and τ_d. For example, when $\lambda = 0.04, \tau_d = 0.004$, the median value of $\hat{\tau}_\alpha$ is around 0.025, which means that at each moment, the probability that a particle born in the invisible zone arrives at $\{-l\} \times [0, H]$ is around 0.025.

It shows that τ_α increases linearly as the birth rate λ increases, and decreases slightly linearly as the death rate τ_d increases.

4.3.3.2 The estimator $\hat{\tau}_d$ is unbiased and performs reasonably well with 5-minute movies

As explained in section 4.2.4, $\hat{\tau}_d$ is a rather classical estimator. Fig. 4.9 shows the estimator with 5-minute movies (magenta) and 30-minute movies (blue) respectively. It confirms that the estimator is unbiased. Black horizontal lines represent the true value of τ_d. Naturally, the variance is bigger with shorter movies.

4.3.3.3 The choice of T_S

According to Figs. 4.8 and 4.9, when $T_S = 30$ min, the estimators of τ_α and τ_d perform well, being converged with small variance. As 30-minute movie acquisition is almost infeasible under the situation of fluorescence microscopy, we need to find a compromise with smaller T_S and reasonably good estimators. We especially tested T_S=2.5 minutes and T_S=5 minutes. Comparing the estimation results with 2.5-minute movies, we found that $T_S = 5$ minutes is a good choice to limit the estimation error of τ_d and to ensure a good connection performance.

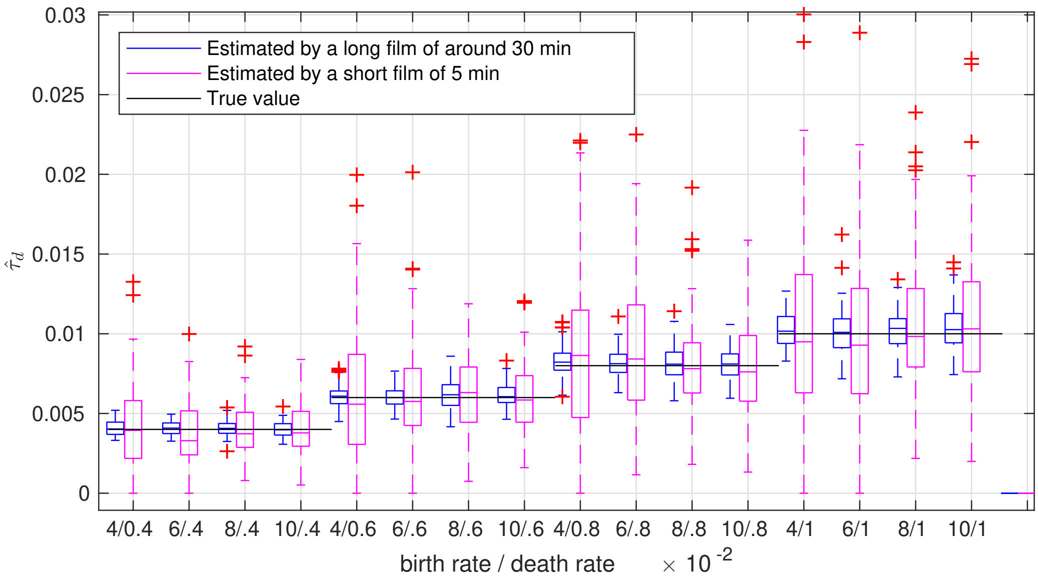


FIGURE 4.9 – $\hat{\tau}_d$ with different λ and τ_d. The estimator is unbiased. The variance of the estimator is larger with shorter movies (magenta). For a given τ_d, when birth rate λ increases (e.g., the first four boxes), then the number of particles also increases. Consequently, the variance decreases.

When the length of movies equals to 2.5 min, the estimation errors of $\hat{\tau}_\alpha$ and $\hat{\tau}_d$ cause the failure of the connection procedure as shown below.

Estimation of birth rate τ_α and of death rate τ_d

The arrival rate τ_α is estimated from 2.5 min movies and the result is presented in Fig. 4.10. Compared to Fig. 4.8 where 5-min movies are used, the estimations with 2.5-min movies have bigger variance. With 2.5-min movies, a considerable proportion of the estimations equal to zero. Similar conclusion can be made for $\hat{\tau}_d$ comparing Fig. 4.9 and Fig. 4.11.

It is shown that both in Figs 4.10 and 4.11, the estimators $\hat{\tau}_\alpha$ and $\hat{\tau}_d$ have many zero values. This phenomenon will deny the birth and death of particles, and force the connection of observed tracklets. The result of connection is presented in the next section and it shows that these biases of parameter estimation have severe influence on the tracklets connection performance.

Evaluation of the connection procedure

Figs. 4.12 and 4.13 show the results of tracklets connection measured by ARI, according to different settings of birth rate λ and death rate τ_d. To be specified, $\tilde{\tau}_\alpha$ and $\tilde{\tau}_d$ are estimators when $T_S = 30$ min, on the contrary, the estimators obtained with 2.5-min movies are denoted as $\hat{\tau}_\alpha$ and $\hat{\tau}_d$. Concerning the experiment with true parameters (blue boxes in both figures),

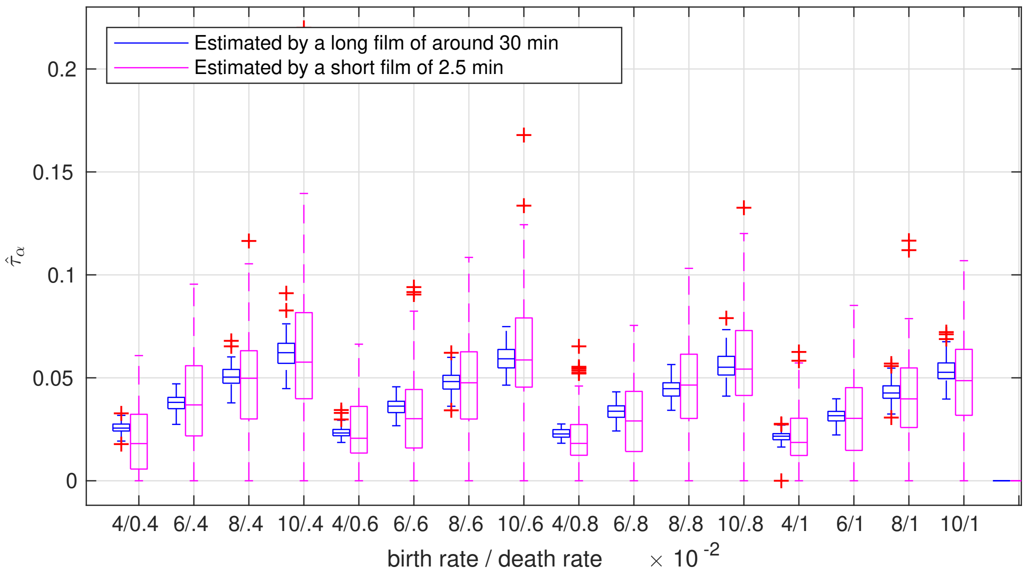


FIGURE 4.10 – The estimation of arrival rate τ_α with different λ and τ_d. Magenta boxes represent $\hat{\tau}_\alpha$ estimated by 2.5-min movies and blue boxes by 30-min movies.

for all settings of λ and τ_d, the connection results are satisfying. However, for the experiments with estimated parameters (magenta boxes in both figures), we see clearly the failure of the connection represented by boxes with very large variance and low ARI values.

To identify which estimators are the main cause of this failure, many intermediate experiments are designed. To remind, $\tilde{\tau}_\alpha$ and $\tilde{\tau}_d$ represent the estimators of τ_α and τ_d from 30-min movies, while $\hat{\tau}_d$ and $\hat{\tau}_\alpha$ from 2.5-min movies. Through Fig. 4.12 and 4.13, it can be concluded that the error of estimators $\hat{\tau}_d, \hat{\tau}_\alpha$ is the main cause of the dramatic decrease of ARI when all the estimators $\hat{\tau}_d, \hat{\tau}_\alpha, \hat{v}$ and $\hat{\sigma}$ are used.

The accuracy of $\hat{\tau}_\alpha$ and of $\hat{\tau}_d$ increases as the total observed time T_S increases (see Fig. 4.7). Therefore, in the study, we chose movies of 5 min to estimate τ_α and τ_d and then to evaluate the connection performance.

4.3.3.4 The connection procedure works well, even when true parameters are unknown

In this part, we assess the performance of the connection algorithm with different parameters λ and τ_d. We evaluate as well the impact of the error of the estimator, by using in the connection procedure true parameters $\tau_\alpha, \tau_d, v, \sigma$ and their estimators $\hat{\tau}_\alpha, \hat{\tau}_d, \hat{v}, \hat{\sigma}$, respectively. The duration of movies T_S is set to 5 minutes. The connection results measured by ARI are presented in Fig. 4.14.

Each pair of blue and magenta boxes represents the connection result of a setting of λ and

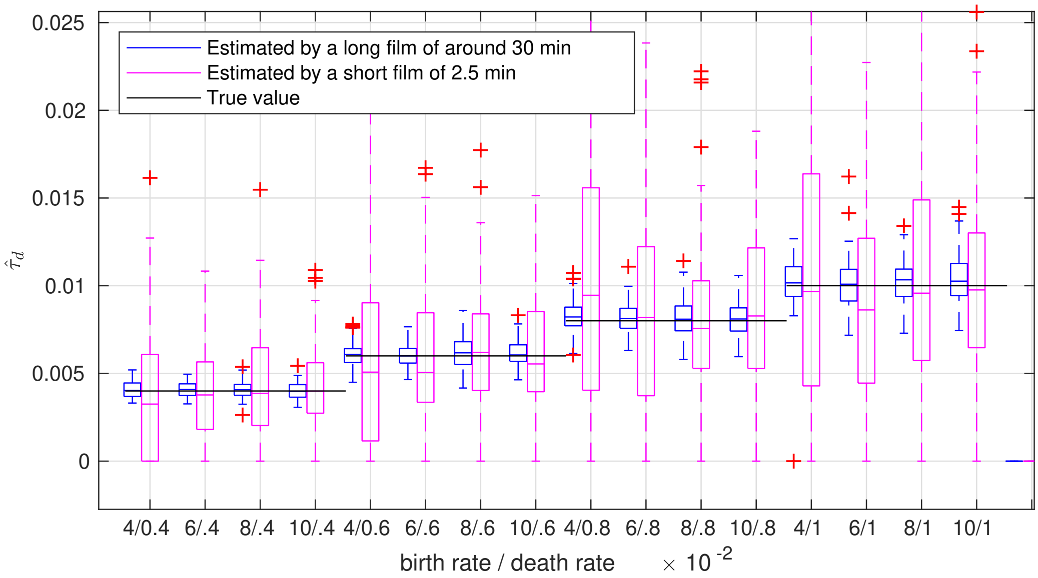


FIGURE 4.11 – The estimation of death rate τ_d with different λ and τ_d. Magenta boxes represent $\hat{\tau}_d$ estimated by 2.5-min movies and blue boxes by 30-min movies. Black horizontal lines represent the true value of τ_d.

τ_d. The black line represents the mean value of the number of tracklets fluctuating with different settings of λ and τ_d. The performance of connection is affected by the number of tracklets in each movie to be connected. The higher the density of tracklets is, the more difficult it is to find the right ones.

It can be noticed that the ARI value when we use the estimators $\hat{\tau}_d, \hat{\tau}_\alpha, \hat{v}$, and $\hat{\sigma}$ is almost as good as when we use true values for all the parameters. This is an encouraging result as it means that it is feasible to apply the algorithm in real image sequences. When the number of tracklets is around 20 (e.g., $\lambda = 0.04$ and $\tau_d = 0.008$), the median values of ARI are higher than 0.9, showing a promising connection performance. Even for the case with the highest particle density, when the average number of tracklets reaches 100 ($\lambda = 0.1$ and $\tau_d = 0.004$), the median value of ARI is still higher than 0.7.

4.3.3.5 The connection procedure is robust even when each particle moves at different speed (but with constant speed along a trajectory). However v and σ should be well estimated

In the previous experiments, all trajectories are generated with the same speed v and standard error σ. In this section, we design experiments to test the performance of the connection algorithm, when the drift v varies from particle to particle, $v_x \sim \mathcal{U}(0.5, 0.9)$. In one movie, as

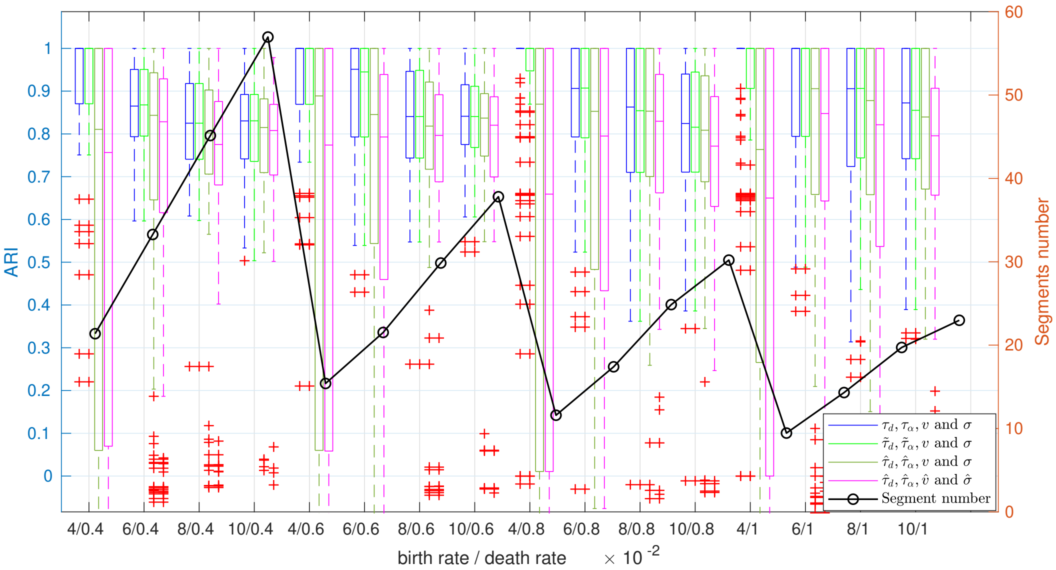


FIGURE 4.12 – Connection performance comparison. Blue (resp. green, dark green and magenta) boxes represent ARI values obtained with parameters $(\tau_d, \tau_\alpha, \mathbf{v}, \sigma)$ (resp. $(\tilde{\tau}_d, \tilde{\tau}_\alpha, \mathbf{v}, \sigma)$, $(\hat{\tau}_d, \hat{\tau}_\alpha, \mathbf{v}, \sigma)$ and $(\hat{\tau}_d, \hat{\tau}_\alpha, \hat{\mathbf{v}}, \hat{\sigma})$). Light green boxes, representing the result when $\tilde{\tau}_\alpha, \tilde{\tau}_d$ (estimators with 30-min movies) and true value v, σ are used, show performance as good as the blue boxes, where true parameters values are used. However, the dark green boxes, where $\hat{\tau}_\alpha, \hat{\tau}_d$ (with 2.5-min movies) and v, σ are used, show much degraded results.

all particles are in the same environment, there is no obvious reason for different particles to have different σ . Therefore, the standard error σ is set to be constant for particles in one movie. However, we test in independent movies, when $\sigma = 0.2, 0.3$, or 0.4, the influence of σ on the performance of connection procedure. Other parameters to be specified are the angle between the direction of the motion and the circumferential direction of the cylinder, $\theta = 0.15 (\approx 8.6^\circ)$, $v_y = \tan(\theta) v_x$, $\sigma_y = \tan(\theta)\sigma_x$, $\sigma_x = \sigma$, birth rate and death rate are fixed, with $\lambda = 0.08$ and $\tau_d = 0.02$.

The normalized error (NE) of an estimator is defined by the error of the estimator normalized by its ground truth. For example, the NE of v_x equals $\frac{v_x - \hat{v}_x}{v_x}$. In Fig. 4.15, the NEs of $\hat{v}_x, \hat{v}_y, \hat{\sigma}_x$ and $\hat{\sigma}_y$ when σ takes different values are presented. It shows that when σ increases, the variance of $\hat{v}_x$ and $\hat{v}_y$ increases.

For tracklets connection, we compare the results when true values of v and σ or when the estimated value $\hat{v}$ and $\hat{\sigma}$, respectively, are taken by the connection procedure. The experiments are carried under three situations, when $\sigma = 0.2, 0.3,$ and 0.4. The results in Fig. 4.16 show that whether using true v and σ or estimated value $\hat{v}$ and $\hat{\sigma}$, the performance measured by ARI degrades when σ increases. When the standard error $\sigma = 0.2$, using true v and σ, the

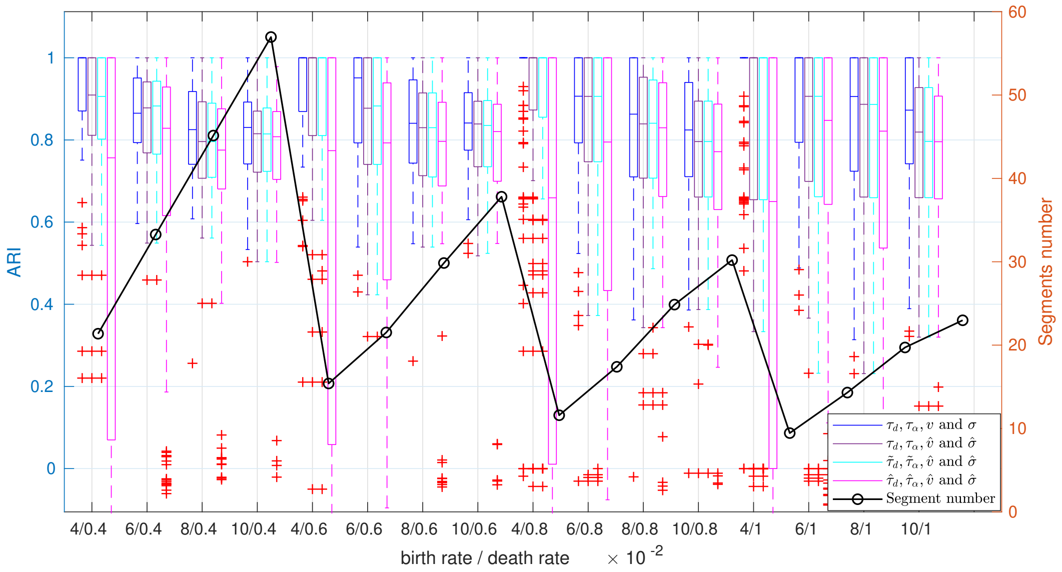


FIGURE 4.13 – **Connection performance comparison.** Blue (resp. violet, cyan and magenta) boxes represent ARI values obtained with parameters $(\tau_d, \tau_\alpha, \mathbf{v}, \sigma)$ (resp. $(\tau_d, \tau_\alpha, \hat{\mathbf{v}}, \hat{\sigma})$, $(\tilde{\tau}_d, \tilde{\tau}_\alpha, \hat{\mathbf{v}}, \hat{\sigma})$, and $(\hat{\tau}_d, \hat{\tau}_\alpha, \hat{\mathbf{v}}, \hat{\sigma})$). Boxes with blue, violet, and cyan colors are similar, which means that the estimators $\hat{v}, \hat{\sigma}, \tilde{\tau}_\alpha$ and $\tilde{\tau}_d$ do not cause degradation of the connection. Only when $\hat{\tau}_\alpha$ and $\hat{\tau}_d$ are used, shown in magenta boxes, the connection results degrades.

median value of ARI reaches 1. When using the estimated $\hat{v}$ and $\hat{\sigma}$, the median value of ARI is approximately 0.75. It can be concluded that the estimation of v and σ has an impact on the performance of the algorithm.

4.3.4 Analysis of the connection results

4.3.4.1 An example of tracklets connection

Figure 4.17 shows, on the left, trajectories in a movie and on the right, the results of tracklets connection. The path from an output to the matched input is represented by the dashed straight line, as we don't know how exactly the particle went through the hidden zone. The only wrong connection corresponds to the bold line. Compared with the figure on the left, we can find the realization of these two tracklets. In reality, the orange bold tracklet disappeared at the hidden region and the bold purple tracklet appeared nearby and entered into the observed zone.

In fact, not only the optimal configuration can be calculated, but also the most likely alternative configurations in decreasing order of probability (Fig. 4.18). It should be noticed that the optimization algorithm tries to minimize $K(c) = -\log Q(c)$, instead of finding the c^* maximizing $P(c)$. It costs too much to obtain the probability $P(c^*)$, as it requires the enumeration of all the

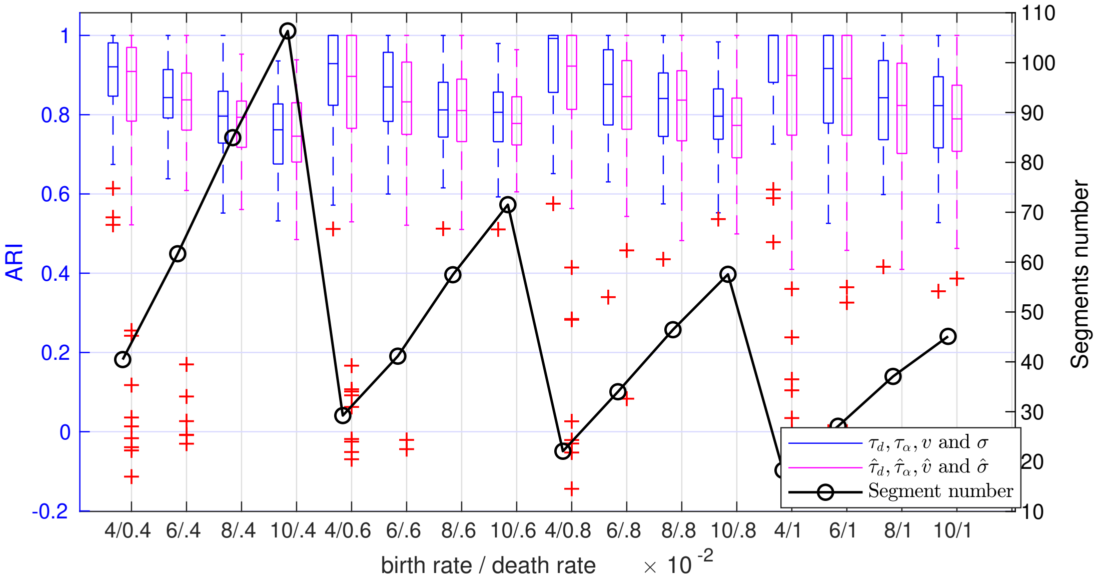


FIGURE 4.14 – Connection performance comparison for different λ and τ_d when $T_S = 5$ minutes. Blue (resp., magenta) boxes represents ARI values obtained with true parameters τ_d, τ_α, v, and σ (resp., with estimators $\hat{\tau}_d, \hat{\tau}_\alpha, \hat{v}$, and $\hat{\sigma}$). The black line represents the mean number of tracklets in a movie.

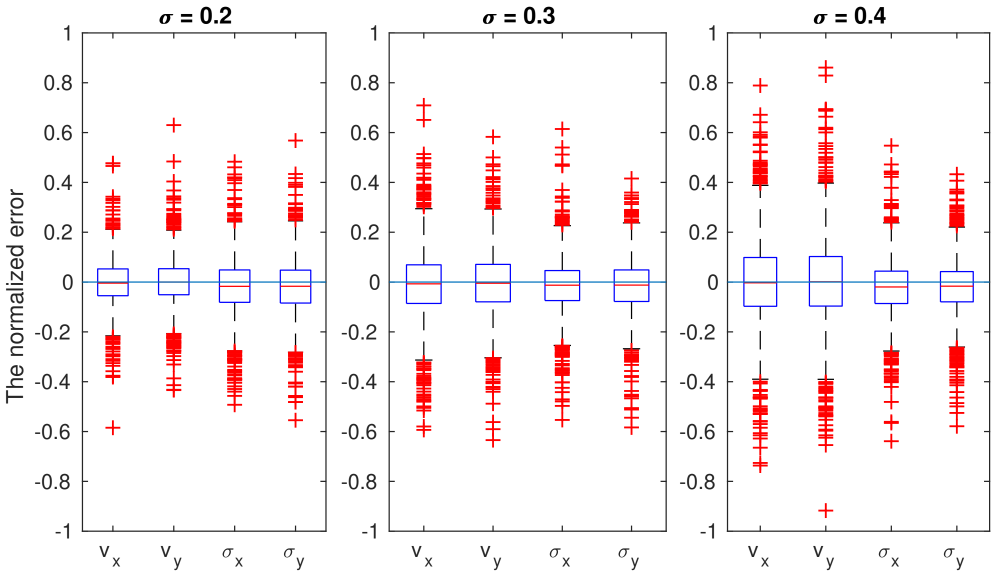


FIGURE 4.15 – The NEs of v_x, v_y, σ_x, and σ_y in cases where $\sigma = 0.2, 0.3$, and 0.4.

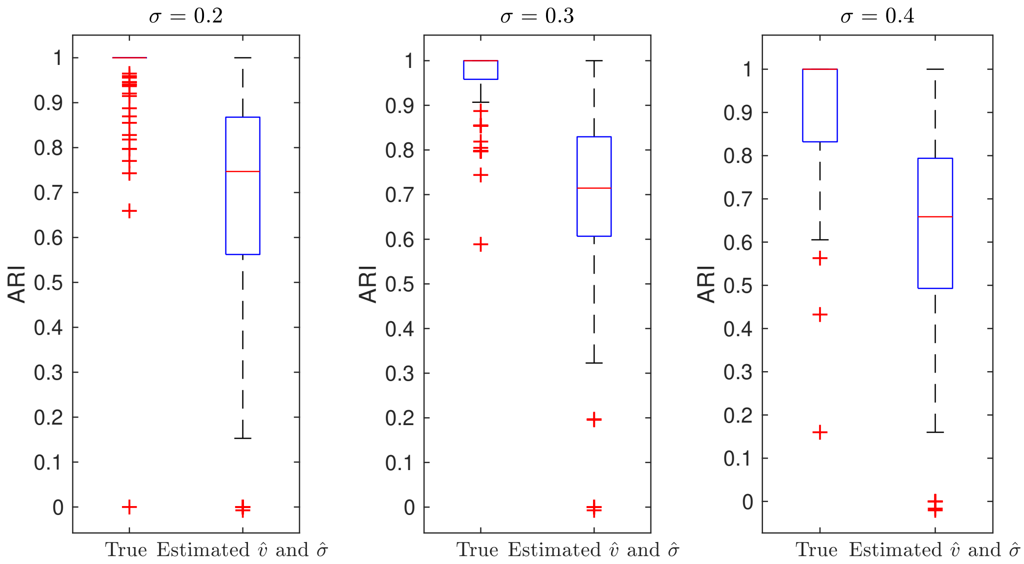


FIGURE 4.16 – Comparison of connection performance measured by ARI when the procedure takes true v and σ or estimated $\hat{v}$ and $\hat{\sigma}$ in three experiments where $\sigma = 0.2, 0.3$, and 0.4 respectively.

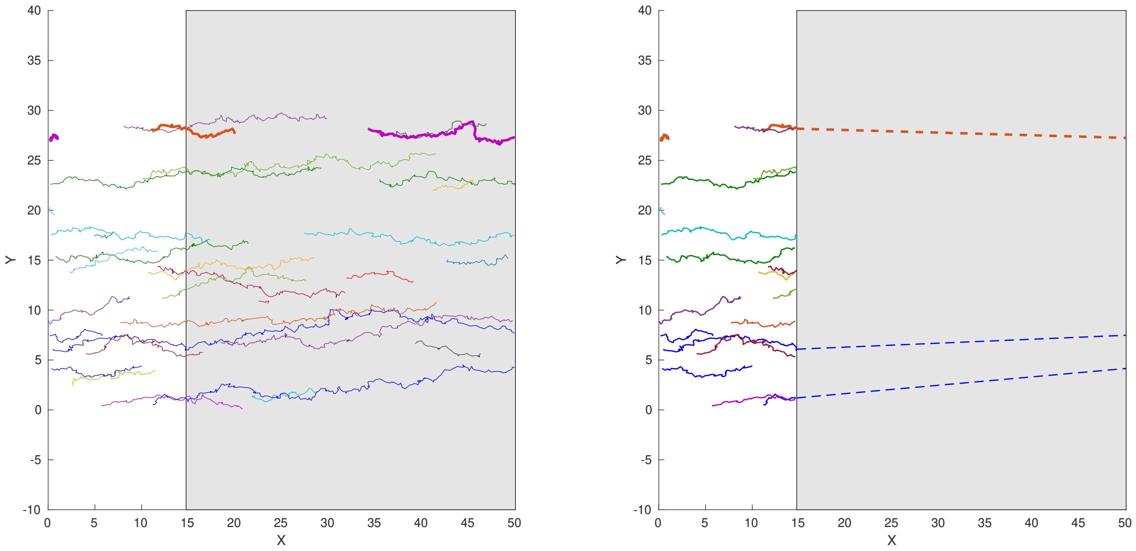


FIGURE 4.17 – Illustration of the reconstruction results on the cell surface approximated by a cylinder. $Y-$ direction represents the axis of the cylinder and $X-$ direction the unwrapped cylindrical side. The shaded part represents the hidden region and the light part corresponds to the region observed by TIRF microscopy. Left: Simulated trajectories in one movie. Right: The connection results.

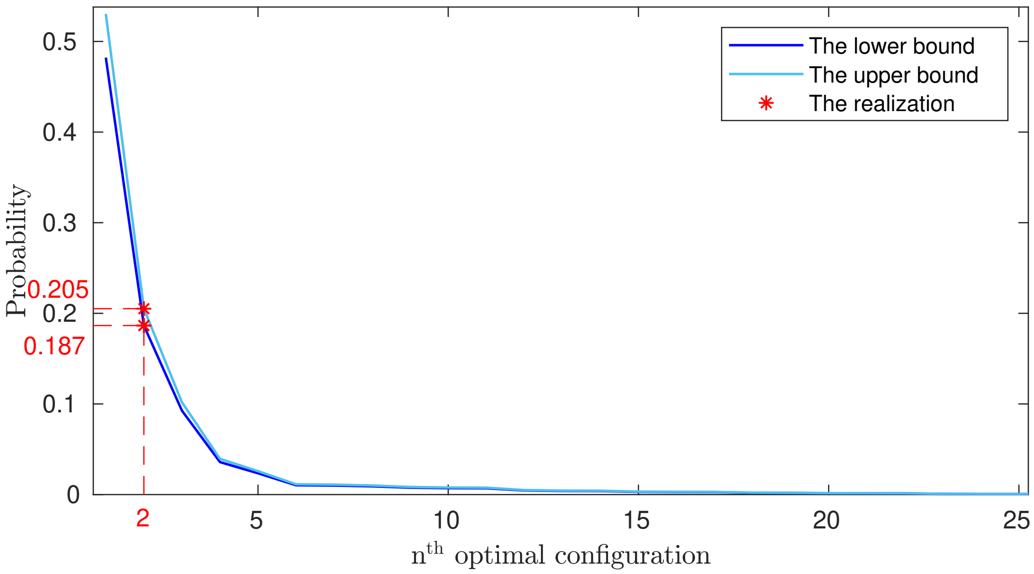


FIGURE 4.18 – The probability of n^{th} optimal configuration and the probability of the realization.

possible configurations $c \in \mathcal{C}$ (4.11). However, the number of configurations can be determined in order to guarantee that the sum of the probability of these configurations will be greater than a given threshold (see A.4). As a result, we can obtain lower and upper bounds for the probability.

For this example, we see that the second most likely configuration corresponds to the realization of trajectories, $0.187 < P(c^t) < 0.205$ according to the algorithm (Fig. 4.18). Combining with Fig. 4.17, the optimal configuration found by the algorithm, committing one connection error, does not correspond to the realization. In section A.3, we evaluated the connection error caused by randomness.

4.3.4.2 The number of rotations around the cylinder

Once the connection procedure is achieved, we can address the question of the number of rotations of a particle around the cylinder. In the context of simulation, the death rate τ_d and the dynamic velocity v_x are known. Accordingly, the value of the number of rotations is known to be equal to $\frac{v_x T_d}{L}$, where $T_d \sim \mathcal{E}(\tau_d)$ ensures a theoretical expectation value of $\frac{v_x}{\tau_d L}$. By counting the tracklets for each trajectory, we can obtain a proxy of the number of rotations around the cylinder.

In Fig. 4.19, λ is set to 0.04 and different values of τ_d between 0.004 and 0.01 are evaluated. Blue bars represent the distribution of the number of rotations of true connections. The magenta bars display the distribution of the number of rotations estimated by the connection

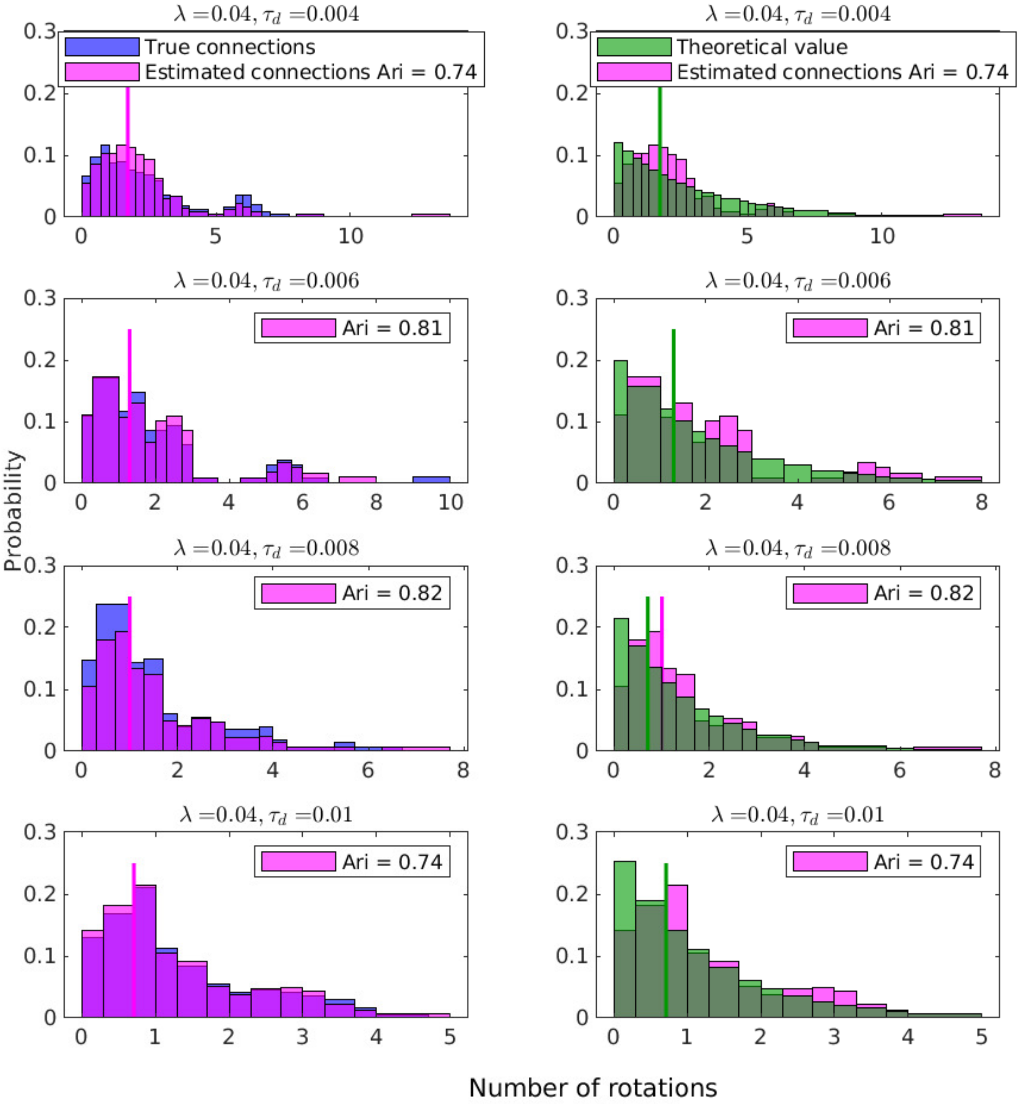


FIGURE 4.19 – The density distribution of "number of rotations." Each row represents the result for a different τ_d value.

procedure. The corresponding ARI values, indicating the connection accuracy, are given as well. The density of the theoretical values of the number of rotations is presented in green. The vertical lines represent the median values of the corresponding distribution. Overall, when τ_d is small, the median value of number of rotations is higher and the distribution has a heavier tail. In general, the distributions with all three colors are similar to each other.

4.4 An illustration of the connection algorithm applied to real MreB dynamics

Data obtained using TIRF microscopy of MreB aggregates in *Bacillus subtilis* [Bil+17] are considered. A typical movie from this dataset shows several MreB aggregates moving inside one or several cells (see Fig. 4.1 and Supplementary Materials 3). The pixel size, frame rate and duration are, respectively, $\Delta x = \Delta y = 64nm$, $\Delta t = 1s$, $T = 2$ minutes. Hereafter, we selected one cell to illustrate the application of our algorithm. First, tracklets exhibiting directed motion should be extracted from the movie data, then tracklets should be projected back on the cylinder shape of the cell and unwrapped, eventually the connection algorithm is applied and a list of likelihood decreasing ordered configurations of trajectories connections is presented to the user.

4.4.1 Construction of the local cell referential

Once MreB aggregates pixels are separated from the background inside each image of the movie, a bounding box is drawn around a given cell and a local **x-y** referential is estimated using Principal Component Analysis (PCA) on the coordinates of pixels belonging to aggregates (Fig. 4.20). The **z** coordinate of an aggregate is inferred using as a prior the cylinder shape of the bacteria and its radius, R, so $z(x, y) = R - \sqrt{R^2 - x^2}$.

4.4.2 Tracking and selection of aggregates in the observed region

Using U-Track [Jaq+08], MreB aggregates are tracked and constitute a set of tracklets. The automatic classification of these tracklets in three classes, respectively Brownian, subdiffusive, and directed motion is done using two algorithms: the classical MSD algorithm and a recent algorithm by statistical test [BKV18]. The tracklets classified as directed motion by either one of the two algorithms are selected for the application of our connection algorithm (Fig. 4.21). The tracklets were projected back on the cylinder and unwrapped. As we can see, only a few aggregates crossed the borders of the visible region. Other aggregates, according to our definitions are born or die in the visible region, which is not true. When an aggregate approaches

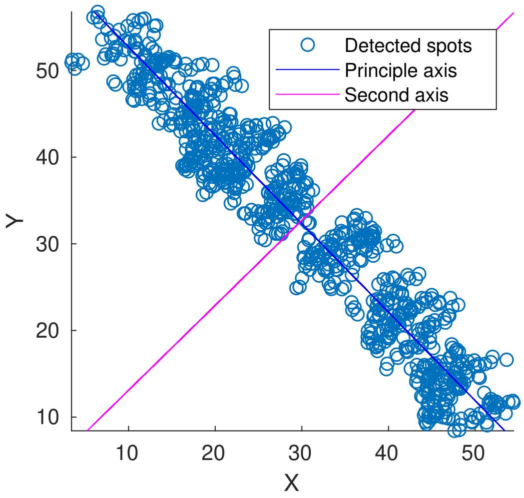


FIGURE 4.20 – The estimation of the local $x-y$ referential for a cell.

the borders, its intensity becomes weak as it is farther from the support plane, and less excitation light penetrates higher z-position in TIRF microscopy settings. As a result, the detection algorithm fails to detect the aggregates when they approach the borders.

4.4.3 The connection of tracklets

All the selected tracklets crossing the magenta lines in Fig. 4.21 are considered as having crossed the visible region. On the contrary, all those who appeared or disappeared inside the magenta lines are considered as birth or death events inside the visible region.

First, the speed and diffusion are estimated (Fig. 4.22) for each tracklet, respectively. Two populations of tracklets evolving in opposite x directions are identified. These two populations are considered one after the other in the connection procedure. Tracklets corresponding to speed lower than 0.4 are filtered out.

For the population of tracklets associated with positive (resp., negative) v_x, death rate τ_d is estimated as 0.0691 (resp., 0.0756). The arrival rate τ_α is estimated as 0.0310 (resp., 0.0220).

Tracklets of positive speed v_x

The first, fifth, seventh, and eighth optimal configuration suggests one connection. The second suggests that there is no connection. The third, fourth, and sixth configurations suggest two connections. Some of these configurations are shown in Fig. 4.23.

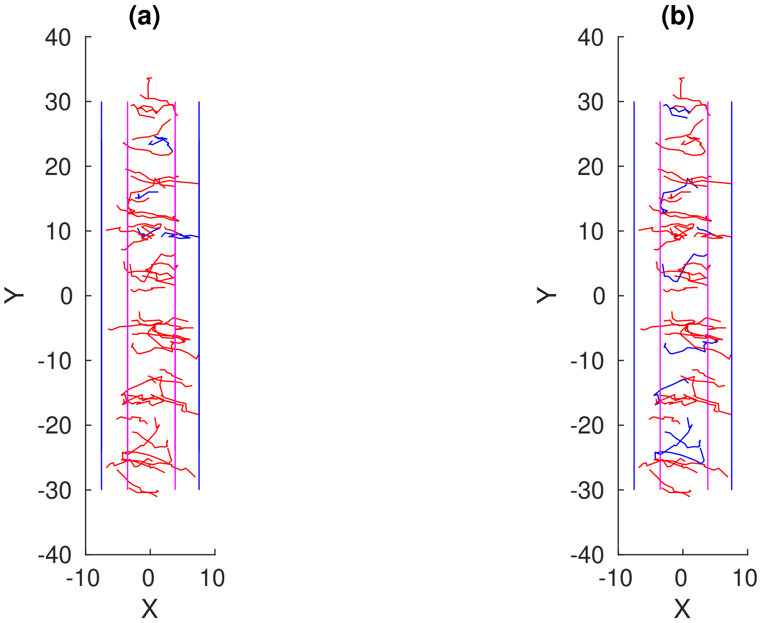


FIGURE 4.21 – The MreB tracklets classification. (a) MSD classification. (b) STP classification. Brownian tracklets are in blue and directed tracklets are in red. Blue lines represent the border of the visible region. Magenta lines represent the 0.1 quantile and the 0.9 quantile of *x* coordinate values.

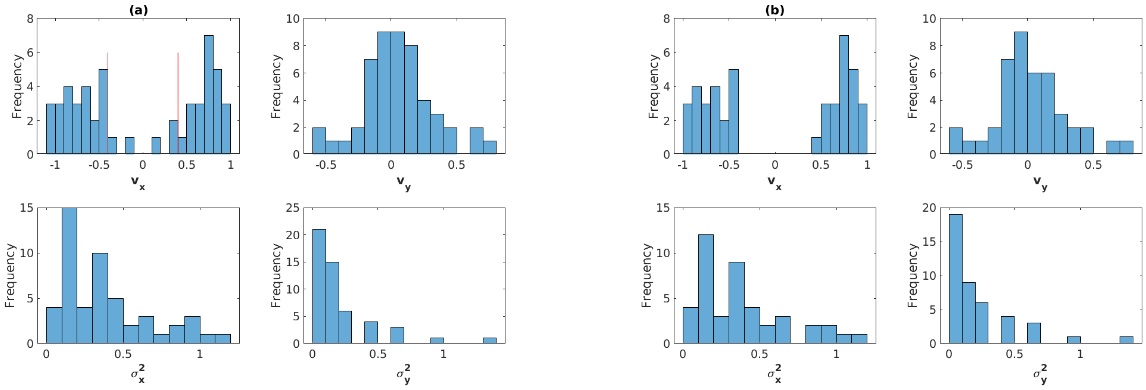


FIGURE 4.22 – The distribution of drift and variance in the selected tracklets population: (a) without filtering, (b) after filtering.

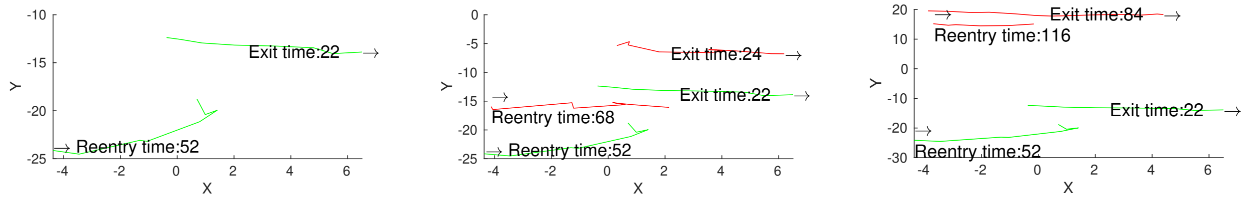


FIGURE 4.23 – Some optimal configurations for the population of positive speed v_x. From left to right, the first, third, and sixth better configurations. Connected tracklets are drawn with the same color.

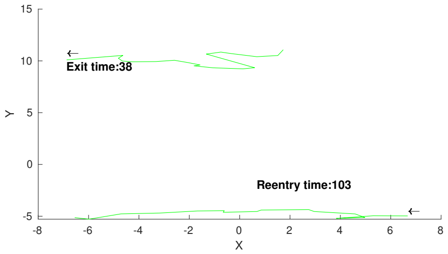


FIGURE 4.24 – Second optimal configuration for the population of negative speed v_x.

Tracklets of negative speed v_x

The first optimal configuration suggests no connection. The second one suggests one connection, Fig. 4.24.

The 3D reconstruction of the dynamics

In Fig. 4.25 we show a 3D reconstruction of the aggregates and two tracks that could correspond to aggregates doing more than one loop around the cylinder surface of the cell. For the positive (resp., negative) speed set of tracklets, the eighth (resp., second) optimal solution was selected.

Conclusion

In this chapter, we proposed a probabilistic framework and a computational approach with no hidden parameter to connect tracklets from 2D partial observations. We provided several consistent estimators of parameters to automatically drive the connection procedure. The performance of our procedure is satisfying if we consider the ARI criterion. Moreover, an ordered set of the best reconstructions could also be proposed. The robustness of the procedure has been tested for different drifts, diffusion of the dynamics, and trajectory densities. Our computational approach can be extended to the case when the drift/speed is not the same for all

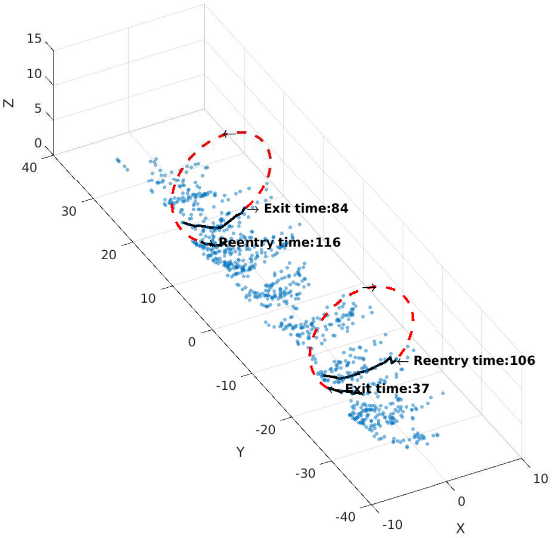


FIGURE 4.25 – The 3D reconstruction of tracks. The centroids of aggregates are represented as blue squares.The arrows indicates the direction of motion. The full black lines represent tracklets crossing the observed region. The dotted red lines represent the simplified extrapolation of aggregate motion in the invisible region.

particles but remains constant along time. In that case, it is straightforward to estimate and classify the drifts before applying our connection procedure to each class of drift since the tracklets with different speeds are not likely to be connected.

After recovering the whole trajectory on the surface of the cylinder, we can have a better understanding of the average duration of a particle, and more accurate statistics about the spatio-temporal organization of particles. The simulation study can also serve as a guideline for the design of experiments.

The connection procedure is tested with a real TIRFM dataset. The experimental results are illustrated in Section 4.4. For future works, we plan to further investigate on real TIRFM datasets. In real experimental data, the observed region corresponds approximately to one-third of the total surface, which is rather small. However, we have shown that we are able to cope with the hidden region of such size. Nevertheless, several assumptions and approximations need to be further investigated. For instance, we assumed spatial homogeneity, suggesting that the particles are born or die uniformly on the membrane surface. Moreover, we assumed a memoryless lifetime while dependency with respect to particle "age" could be more realistic.

CONCLUSION AND PERSPECTIVES

Conclusion

In this thesis, we reviewed the advanced statistical and mathematical modeling tools for single particle trajectory analysis. We deepened the analysis of the trajectories of MreB filaments with the available statistical methods and algorithms and developed new probabilistic and computational tools, to better exploit the information from trajectory data.

In the first part, we proposed two approaches to analyze the MreB trajectories based on diffusion models characterized by Stochastic Differential Equations. The first approach consists in analyzing individual trajectories, classifying the tracks into different diffusion modes (free diffusion, super-diffusion or sub-diffusion) and estimating motion features (drift and diffusion coefficient). We compared three non-parametric classification methods, two MSD-based methods and another method based on statistical test, namely THOTH. We found that the three methods provided consistent results in classifying super-diffusion, while THOTH is more robust than the others in identifying free diffusion. We then evaluated two diffusion coefficient estimators for free diffusive trajectories, taking into account the measurement errors observed in high-resolution fluorescent microscopy. The two estimators, Optimal Least Squared Fitting (OLSF) and maximum likelihood estimator (MLE), provided consistent results. From MreB trajectories recorded by SIM-TIRFM and classified as free diffusion, the diffusion coefficient $D \sim 10^{-3} \mu m^2/s$ while the measurement error approximated by a Gaussian distribution has a variance $\sigma^2_{\text{loc}} \sim 10^{-3} \mu m^2/s$. The measurement error is estimated at the same time as the diffusion coefficient from the trajectories and cannot be ignored, as the result shows that the measurement error has the same order of magnitude as the the diffusion coefficient D.

It was found that the temporal resolution, the length of trajectories, and the ratio of the signal to noise (the ratio between the diffusion coefficient and localization error) set a fundamental limit on the optimal estimators, which can be assessed by the Cramer-Rao Lower Bound (CRLB). Overall, the analysis of diffusion modes and the estimation of the diffusion coefficients by investigating individual trajectories (in the Langrangian setting), is very helpful to guide the image acquisition by experts, that is balancing the settings of exposure time, record duration, and the frequency of acquisition, and then to achieve the best possible precision on the target estimators. As a short-term objective, we plan to evaluate the procedure of classification and coefficient estimation on a larger dataset and confirm the preliminary experimental results.

The second approach involves estimating the dynamics of particles in the neighborhood of a position point, and further obtaining drift and diffusion fields. The dynamics is modeled by Langevin equation. By approximating the rod-shaped cell by a cylinder, we proposed an original procedure to estimate the cylinder radius from the planar projection of trajectories taking place on the 3D cylindrical surface. The procedure was evaluated on synthetic data. Unfortunately, it turned out that the current density of the trajectories of MreB aggregates is insufficient for this procedure to work reliably on experimental data. This will be accomplished in future work from single-molecule localization and photo-activated microscopy.

In the second part, we dealt with the problem of the incomplete (i.e., partial) observation field in TIRMF, where only around one third of the cell membrane can be observed. Our objective was to recover the dynamics of the directed MreB particles on the whole surface of the cylindrical membrane surface. Therefore, we proposed a probabilistic procedure, modeling the birth and death process of the particles on the surface and particles' exit from the region of observation (ROO) and re-entrance into the ROO. Based on the observations of MreB dynamics in the ROO, the parameters were estimated using the proposed consistent estimators. The performance of our procedure was evaluated on synthetic data, by varying the diffusion features and the spatio-temporal sampling rate. It is worth noting that the procedure provides not only the optimal configuration of the reconstruction but also the first n optimal configurations with descending probability. This ranking can be exploited by experts to select several reconstructions for further examination. Once the dynamics over the whole surface is recovered, it becomes possible to address new questions. For instance, the average duration of the directed motion of MreB filament can be further investigated.

The reconstruction procedure will be further applied to a larger dataset of cells in similar physiological states, to gain more significant statistical information. Some assumptions imposed in the conception of the probabilistic model could be revisited. For example, we assumed spatial homogeneity, suggesting that the particles are born or die uniformly on the membrane surface, independently of the existing particles in the field of view. Another assumption supposes a constant death rate of particles, without dependency on particle "age". Furthermore, in order to give more realistic reconstruction, once an exit is connected to a re-entrance, the path in the hidden region could be simulated by *Brownian bridge* instead of being represented by a directed motion (straight trajectory).

Perspectives

We suggest some avenues of research for future work. To reconstruct the trajectories in the hidden region, from a large volume of repetitive observations in the ROO, the use of autoencoders and Generative Adversary Networks (GAN) is under investigation. The challenge arises from the large size of the hidden area, which is almost twice as large as the ROO. The learning method will consist in analyzing the trajectories inside the ROOs from large quantities of movies of a collection of different bacteria. The main idea would be to assume that, in some way, the main characteristics of the trajectories to be reconstructed are already present inside the training set. The first reason for this is that there is no preference for observing a specific part of the surface with respect to the imaging support. Secondly, the biogenesis process is mostly identically reproduced by the bacteria set under study.

Moreover, we have assumed that the particles are born or die uniformly on the membrane surface, independently of the existing particles in the field of view. This assumption can be formally tested by considering the recent spatial birth-death-move process [LG20]. The spatial birth-death processes are generalizations of simple birth-death processes, where the birth and death dynamics depend on the spatial locations of individuals. Spatial birth-death-move processes further generalize the spatial birth-death processes by including a move during the life time of a particle, according to a continuous Markov process. A non-parametric estimation of the birth and death intensity functions β and δ has also been proposed in [LG20]. If we assume that the regularity of the total intensity $\alpha = \beta + \delta$, e.g., that is $\alpha(x)$ depends only on the cardinality of x, where x is a certain point process, the proposed estimator then depends only on the distance between the cardinalify of x and the cardinalities of the observed process (X_t) for $t \in [0, T]$. The non-parametric estimation of the birth and death intensity function makes it possible to test the hypothesis made on the homogeneity of birth and death of MreB particles on the membrane surface. The challenge will be to apply this concept and the non-parametric estimator on partial observations on a closed surface with periodicity.

As studied in the thesis, the MreB aggregates perform different types of motion. The directed circumferential movement along the perimeter is supposed to coincide with the active cell wall elongation activities. The class of free diffusion is supposed to coincide with MreB aggregates recruiting other necessary enzymes to form the machinery. Therefore, a birth event of a free diffusion may correspond to the formation of a MreB aggregate, and a death event of a free diffusion corresponds to the disappearance of a MreB aggregate or the starting of a directed motion. Similarly, in the directed motion class, a birth event may represent a diffusive particle that gets ready to circulate, and a death event may mean the end of a cell wall elongation task. The regulation of each class proportion and the spatial-temporal organization of different classes need to be further investigated. The spatial birth-death-move process provides a promising approach to explore the joint spatio-temporal dynamics of two groups of proteins that seem to interact during the physiological mechanisms.

ANNEXE A

APPENDIX

A.1 Itô integral

The true velocity and standard error on the cylinder surface is unknown. But firstly we estimate the effective velocity and standard error of the dynamics projected on the plane. Given the stochastic process (3.35), for some Δt we can integrate from t to Δt

$$\int_t^{t+\Delta t} dX_t = \int_t^{t+\Delta t} \mu(X_t, t)dt + \int_t^{t+\Delta t} \sigma(X_t, t)dW_t.$$

We suppose that Δt is small enough, that the drift and diffusion coefficient can be considered constant during t and $t + \Delta t$. Then it gives

$$X_{t+\Delta t} - X_t = \mu\Delta t + \sigma \int_t^{t+\Delta t} dW_t.$$

Then let us calculate the expectation of $X_{t+\Delta t} - X_t$ and of $(X_{t+\Delta t} - X_t)^2$,

$$\mathbb{E}(X_{t+\Delta t} - X_t) = \mu\Delta t + \sigma\mathbb{E}\left[\int_t^{t+\Delta t} dW_t\right] = \mu\Delta t, \tag{A.1}$$

$$\mathbb{E}[(X_{t+\Delta t} - X_t)^2] = (\mu\Delta t)^2 + \sigma^2\mathbb{E}\left[\int_t^{t+\Delta t} dW_t\right]^2 + 2\mu\sigma\mathbb{E}\left[\int_t^{t+\Delta t} dW_t\right] = (\mu\Delta t)^2 + \sigma^2\Delta t. \tag{A.2}$$

According to Itô Isometry Thm. 1,take $\phi(t, w) = 1$, the second term in the last step $\mathbb{E}\left[\int_t^{t+\Delta t} dW_t\right]^2 = \Delta t$.

Theorem 1 ***(Itô Isometry** P. 26 [Oks13]) If $\phi(t, w)$ is bounded and elementary, then*

$$\mathbb{E}\left[\left(\int_S^T \phi(t, w)dB_t(w)\right)^2\right] = \mathbb{E}\left[\int_S^T \phi(t, w)^2 dt\right]. \tag{A.3}$$

To estimate $\mu(X)$ and $\sigma(X)$, we divide the field into a grid and we consider that in each cell of the grid, the two coefficients are constant.The empirical version for (A.1) and (A.2) can be written as:

$$\frac{1}{N_k \Delta t} \sum_{j=1}^{N_t} \sum_{\tilde{\mathbf{x}}_i^j \in B(\mathbf{x}_k, \Delta x/2)} (x_{i+1}^j - x_i^j) \approx a_x(\mathbf{x}_k), \tag{A.4}$$

$$\frac{1}{N_k \Delta t} \sum_{j=1}^{N_t} \sum_{\tilde{\mathbf{x}}_i^j \in B(\mathbf{x}_k, \Delta x/2)} (x_{i+1}^j - x_i^j)^2 \approx \tilde{\sigma}_{xx}^2(\mathbf{x}_k) + a_x^2(x_k)\Delta t, \tag{A.5}$$

where $B(\mathbf{x}_k, \Delta\mathbf{x}/2)$ is a square bin $\{x \in [x_k - \Delta x/2, x_k + \Delta x/2], y \in [y_k - \Delta y/2, y_k + \Delta y/2]\}$, N_k is the number of points in $B(\mathbf{x}_k, \Delta\mathbf{x}/2)$.

A.2 Estimation of the radius of cylinder by linear fitting (supplementary experiments)

As a complement to the results of the estimation of radius R presented in Section 3.4.4, experiments with different parameters are presented in the following, with $D = 0.04, \Delta t = 0.25$ (Fig. A.1), $D = 0.09, \Delta t = 0.1$ (Fig. A.2), and $D = 0.16, \Delta t = 0.1$ (Fig. A.3).

A.3 Analysis of errors

In this section, we evaluated the connection error caused by randomness. We display in Fig. A.4 the scatter plots of ARI value vs $K(c^t) - K(c^*)$, where c^* denotes the optimal configuration calculated by the "Tracklets Connection Algorithm" while c^t is the true configuration. Each scatter plot displays the results of 100 simulations for a given combination of λ and τ_d.

The difference between $K(c^t)$ and $K(c^*)$ is always positive or null, showing the optimization procedure works correctly to find the optimal solution. When $K(c^t) > K(c^*)$, it means that the configuration c^* has a higher probability than the true realization c^t, which can be due to randomness. We can notice that ARI decreases as soon as $K(c^t) - K(c^*)$ increases. This error occurs when the realization is significantly different from the optimal configuration. Overall, ARI values are generally above 0.7.

Overall, we observe an increment of ARI when the difference decreases. The point clouds are diagonally shaped from top left to bottom right, showing a continuity that bigger the difference between $K(c^t)$ and $K(c^*)$, lower are the values of ARI.

Therefore, to improve the performance of the connection model, it is needed to find some characteristics of different trajectories to distinguish the true realization. To investigate the connection errors, we display in Fig. A.4 the scatter plots of ARI value vs. the difference bet-

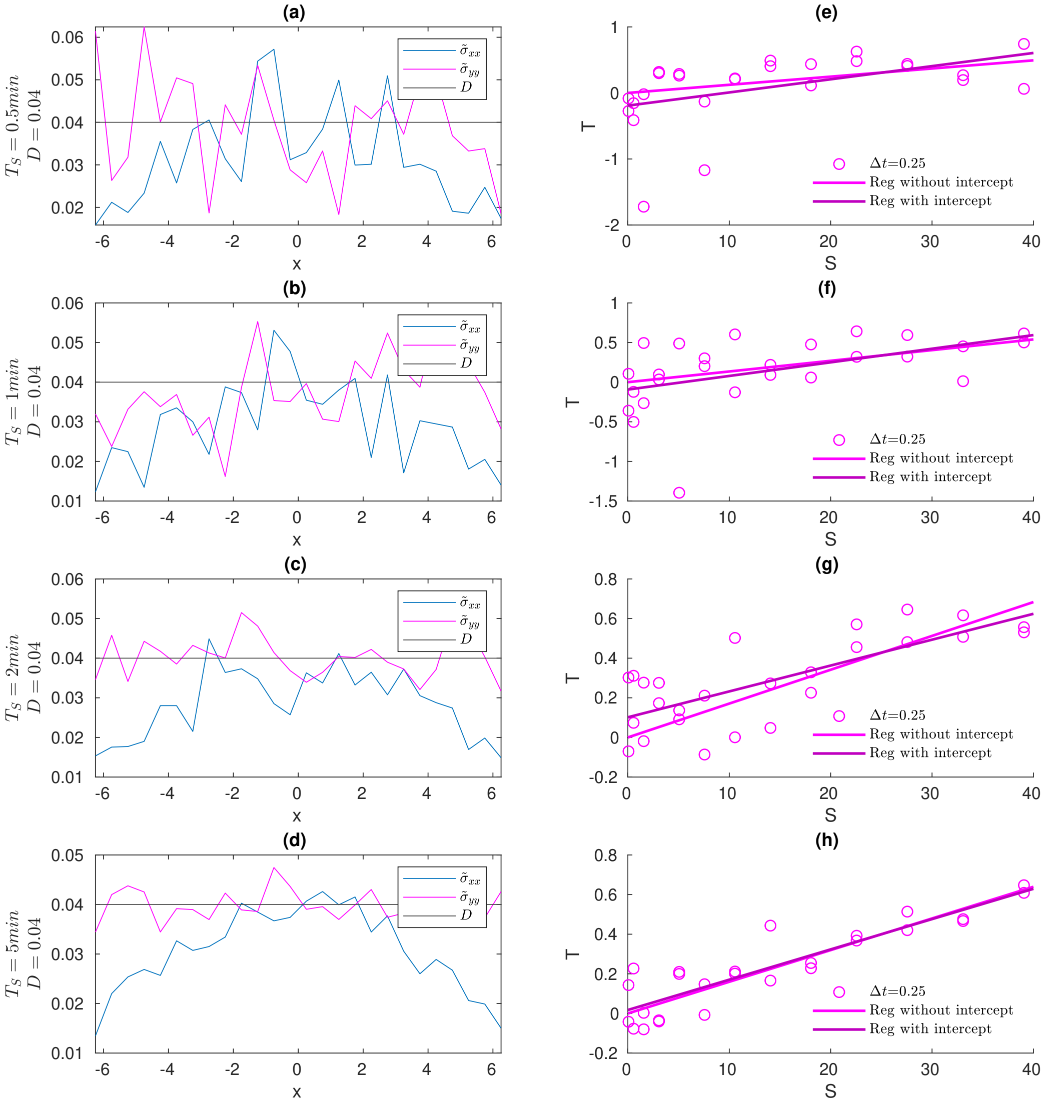


FIGURE A.1 – The estimated effective diffusion coefficients as a function of x (a)-(d), and the linear regression without and with intercept(e)-(h). $\Delta t = 0.25, D = 0.04$. Corresponding to each line, the total observation time $T_S = 0.5, 1, 2$, and 5 min, respectively. (a): As $T_S = 0.5$ min, $\tilde{\sigma}_{yy}$ (magenta curve) oscillates around the horizontal line $y = D = 0.04$, and $\tilde{\sigma}_{xx}$ (blue curve) oscillates around a parabolic curve, which correspond to (3.40). (c)-(d): As T_S increases, the curves become more smooth and better fit to (3.40). (e): The scatter plot of T vs S is fitted by linear model, without (magenta) and with (purple) the intercept term, respectively. The slope of straight lines is equal to $1/R^2$ (3.42). (f)-(h): As T_S increases, the scattered points become more regular and aligned.

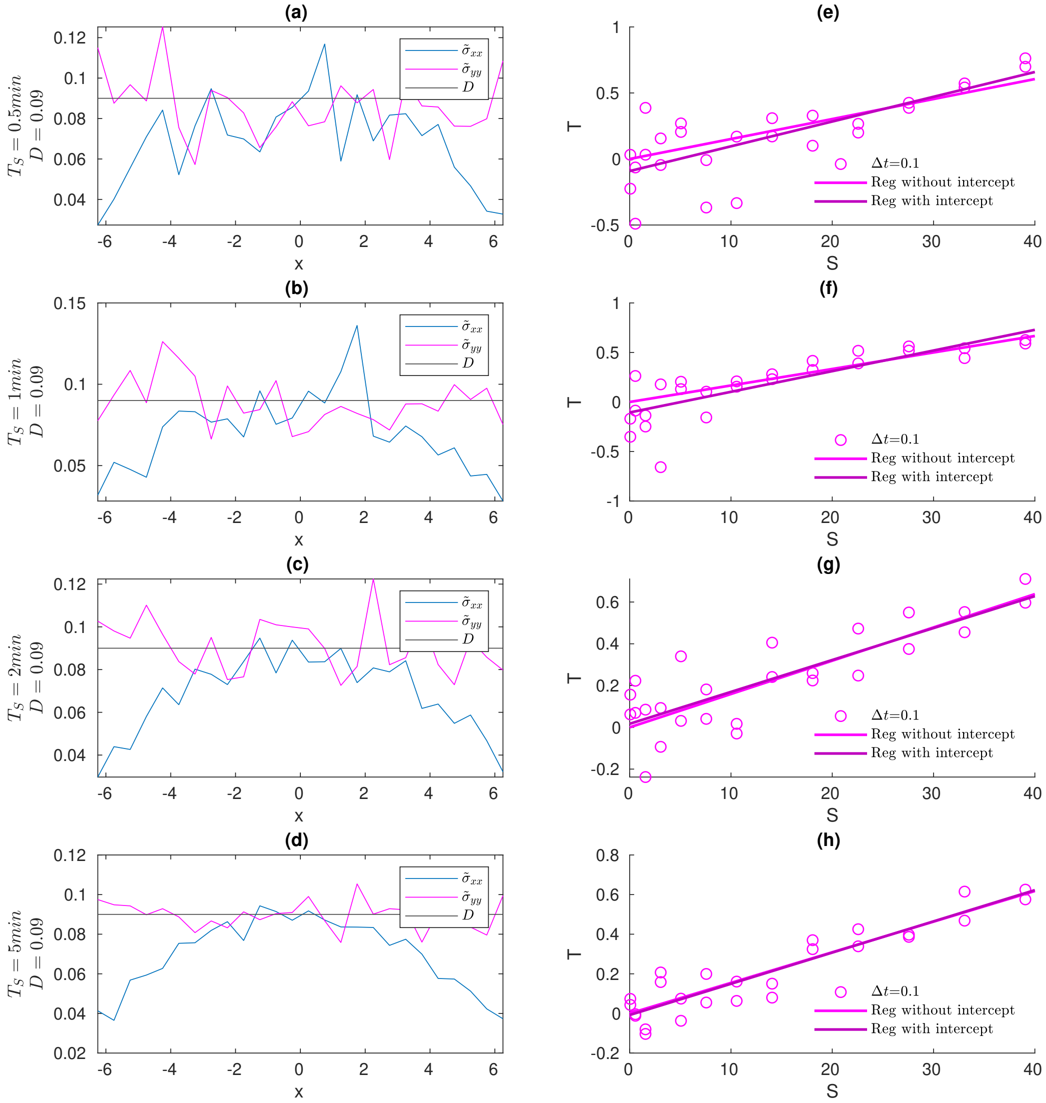


FIGURE A.2 – The estimated effective diffusion coefficients as a function of x (a)-(d), and the linear regression without and with intercept(e)-(h). $\Delta t = 0.1, D = 0.09$. Corresponding to each line, the total observation time $T_S = 0.5, 1, 2$, and 5 min, respectively. (a): As $T_S = 0.5$ min, $\tilde{\sigma}_{yy}$ (magenta curve) oscillates around the horizontal line $y = D = 0.09$, and $\tilde{\sigma}_{xx}$ (blue curve) oscillates around a parabolic curve, which correspond to (3.40). (c)-(d): As T_S increases, the curves become more smooth and better fit to (3.40). (e): The scatter plot of T vs S is fitted by linear model, without (magenta) and with (purple) the intercept term, respectively. The slope of straight lines is equal to $1/R^2$ (3.42). (f)-(h): As T_S increases, the scattered points become more regular and aligned.

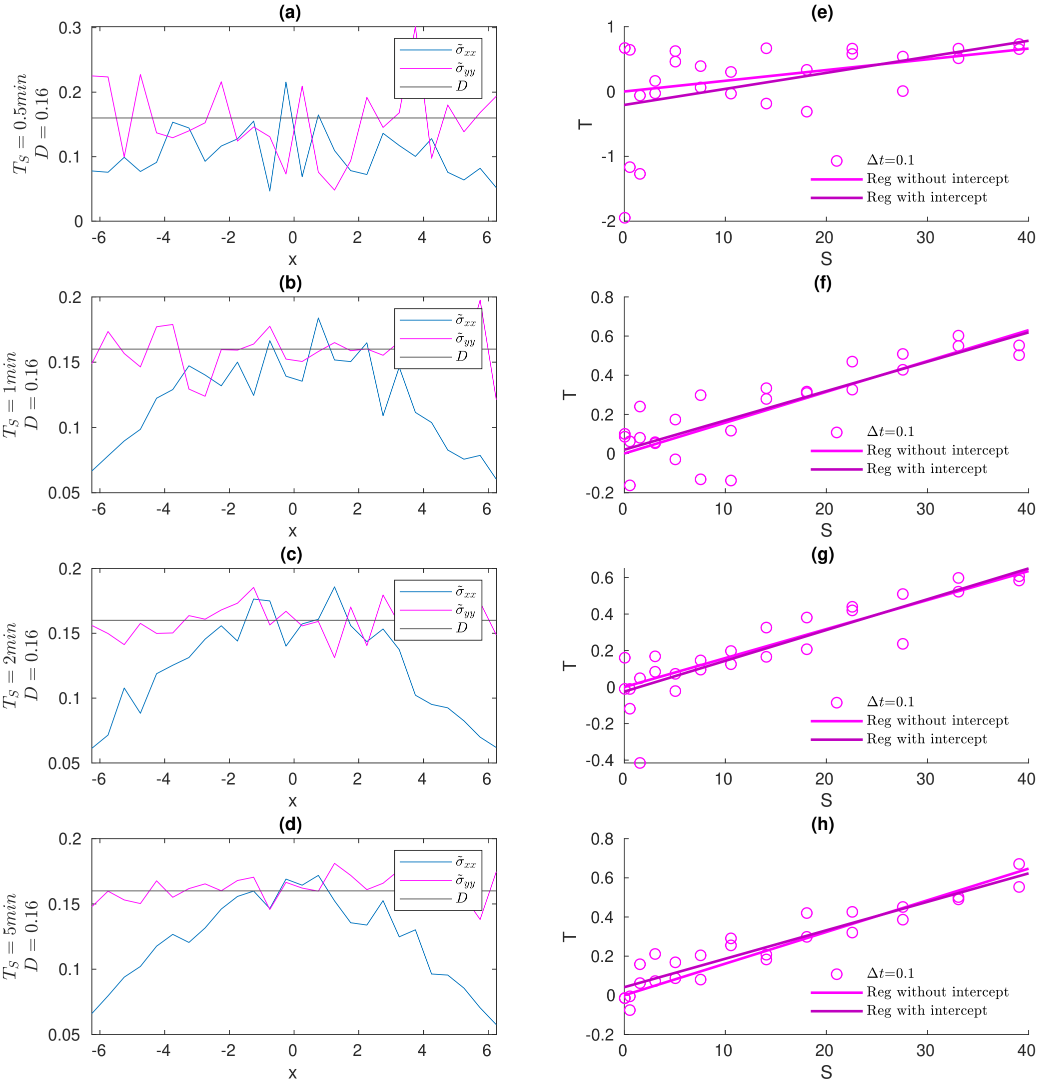


FIGURE A.3 – The estimated effective diffusion coefficients as a function of x (a)-(d), and the linear regression without and with intercept(e)-(h). $\Delta t = 0.1, D = 0.16$. Corresponding to each line, the total observation time $T_S = 0.5, 1, 2,$ and 5 min, respectively. (a): As $T_S = 0.5$ min, $\tilde{\sigma}_{yy}$ (magenta curve) oscillates around the horizontal line $y = D = 0.16$, and $\tilde{\sigma}_{xx}$ (blue curve) oscillates around a parabolic curve, which correspond to (3.40). (c)-(d): As T_S increases, the curves become more smooth and better fit to (3.40). (e): The scatter plot of T vs S is fitted by linear model, without (magenta) and with (purple) the intercept term, respectively. The slope of straight lines is equal to $1/R^2$ (3.42). (f)-(h): As T_S increases, the scattered points become more regular and aligned.

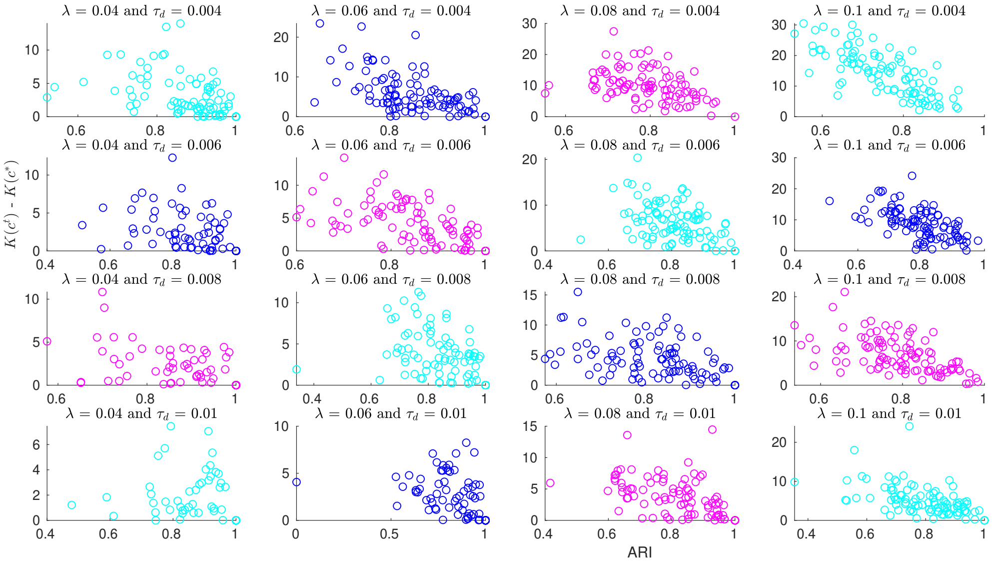


FIGURE A.4 – The difference between $K(c^t)$ and $K(c^*)$ versus ARI for different values of birth rate and death rate.

ween $K(c^t)$ and $K(c^*)$, where c^* denotes the optimal configuration calculated by the "tracklets Connection Algorithm" while c^t is the true configuration. Each scatter plot is associated with a given combination of λ and τ_d. We considered 100 replications for this experiment. On the left column, the connection is perfect in most case. Many points overlapped and located at coordinates (1,0). Overall, we observe an increment of ARI when the difference decreases. This can be due to low values of ARI obtained when the ground truth does not correspond to the optimal configuration.

A.4 Calculation of the boundary of $P(c)$

With the optimization algorithm (Eq. 4.15), we note c_i the i$^{\text{th}}$ optimal solution. Note that N_c is the number of all possible configurations given S. For any $1 \leq n \leq N_c$, we have

$$\sum_{i=1}^{n} Q(c_i) \leq \sum_{i=1}^{N_c} Q(c_i) \leq \sum_{i=1}^{n} Q(c_i) + (N_c - n)Q(c_n)$$

Using Eq. 4.11, it gives

$$\frac{Q(c)}{\sum_{i=1}^{n} Q(c_i) + (\tilde{N}_c - n)Q(c_n)} \leq P(c) \leq \frac{Q(c)}{\sum_{i=1}^{n} Q(c_i)},$$

where $\tilde{N}_c$ is the number of possible configurations calculated through the number of inputs and outputs. We know that $\tilde{N}_c$ is bigger than N_c because some configurations counted in $\tilde{N}_c$ are not compatible according to the time of inputs and outputs. Unfortunately we don't know the exact N_c.

Note $l(c,n) = \frac{Q(c)}{\sum_{i=1}^{n} Q(c_i)+(\tilde{N}_c-n)Q(c_n)}$ and $u(c,n) = \frac{Q(c)}{\sum_{i=1}^{n} Q(c_i)}$. In order to guarantee the precision of the probability, we choose $n^*, 1 \leq n^* \leq N_c$, big enough to satisfy, for $\alpha > 0$,

$$\frac{(\tilde{N}_c - n^*)Q(c_{n^*})}{\sum_{i=1}^{n^*} Q(c_i)} < \alpha. \tag{A.6}$$

This gives us

$$\frac{u(c,n^*)}{l(c,n^*)} = \frac{\sum_{i=1}^{n^*} Q(c_i) + (\tilde{N}_c - n^*)Q(c_{n^*})}{\sum_{i=1}^{n^*} Q(c_i)} = 1 + \frac{(\tilde{N}_c - n^*)Q(c_{n^*})}{\sum_{i=1}^{n^*} Q(c_i)} < 1 + \alpha$$

which ensures that the upper and lower bounds are close from each other. As a result, it gives

$$\sum_{i=1}^{n^*} P(c_i) > \sum_{i=1}^{n^*} l(c_i, n^*) > \frac{1}{1+\alpha}.$$

In other words, this means that the set of configurations c_i up to n^* correspond to an highly likely event for $\frac{1}{1+\alpha}$ close to 1.

BIBLIOGRAPHY

[ABT84] Daniel AXELROD, Thomas P BURGHARDT et Nancy L THOMPSON, « Total internal reflection fluorescence », in : *Annual Review of Biophysics and BioEngineering* 13.*1* (1984), p. 247-268.

[Air35] George Biddell AIRY, « On the diffraction of an object-glass with circular aperture », in : *Transactions of the Cambridge Philosophical Society* 5 (1835), p. 283.

[Aly+09] S Anisah ALYAHYA, Roger ALEXANDER, Teresa COSTA, Adriano O HENRIQUES, Thierry EMONET et Christine JACOBS-WAGNER, « RodZ, a component of the bacterial core morphogenic apparatus », in : *Proceedings of the National Academy of Sciences* 106.*4* (2009), p. 1239-1244.

[Art+19] Marloes ARTS, Ihor SMAL, Maarten W PAUL, Claire WYMAN et Erik MEIJERING, « Particle mobility analysis using deep learning and the moment scaling spectrum », in : *Scientific Reports* 9.*1* (2019), p. 1-10.

[AS19] Risa AGO et Daisuke SHIOMI, « RodZ : a key-player in cell elongation and cell division in Escherichia coli », in : *AIMS Microbiology* 5.*4* (2019), p. 358.

[AS21] Danae Morales ANGELES et Dirk-Jan SCHEFFERS, « The cell wall of Bacillus subtilis », in : *Curr. Issues Mol. Biol* 41 (2021), p. 539-596.

[AWD] Daniel AXELROD et Michael W. DAVIDSON, *Introduction and Theoretical Aspects*, URL : `https://www.olympus-lifescience.com/fr/microscope-resource/primer/techniques/fluorescence/tirf/tirfintro/`.

[BBS88] Henk AP BLOM et Yaakov BAR-SHALOM, « The interacting multiple model algorithm for systems with Markovian switching coefficients », in : *IEEE Transactions on Automatic Control* 33.*8* (1988), p. 780-783.

[Bee+13] Morgan BEEBY, James C GUMBART, Benoît ROUX et Grant J JENSEN, « Architecture and assembly of the Gram-positive cell wall », in : *Molecular Microbiology* 88.*4* (2013), p. 664-672.

[Ben+09] Felipe O BENDEZÚ, Cynthia A HALE, Thomas G BERNHARDT et Piet AJ DE BOER, « RodZ (YfgA) is required for proper assembly of the MreB actin cytoskeleton and cell shape in E. coli », in : *The EMBO Journal* 28.*3* (2009), p. 193-204.

[Ber10] Andrew J BERGLUND, « Statistics of camera-based single-particle tracking », in : *Physical Review E* 82.*1* (2010), p. 011917.

[Bet+06] Eric BETZIG, George H PATTERSON, Rachid SOUGRAT, O Wolf LINDWASSER, Scott OLENYCH, Juan S BONIFACINO, Michael W DAVIDSON, Jennifer LIPPINCOTT-SCHWARTZ et Harald F HESS, « Imaging intracellular fluorescent proteins at nanometer resolution », in : *Science* 313.*5793* (2006), p. 1642-1645.

[Bil+14] Gabriel BILLINGS, Nikolay OUZOUNOV, Tristan URSELL, Samantha M DESMARAIS, Joshua SHAEVITZ, Zemer GITAI et Kerwyn Casey HUANG, « De novo morphogenesis in L-forms via geometric control of cell growth », in : *Molecular Microbiology* 93.*5* (2014), p. 883-896.

[Bil+17] Cyrille BILLAUDEAU, Arnaud CHASTANET, Zhizhong YAO, Charlène CORNILLEAU, Nicolas MIROUZE, Vincent FROMION et Rut CARBALLIDO-LÓPEZ, « Contrasting mechanisms of growth in two model rod-shaped bacteria », in : *Nature Communications* 8 (2017), p. 15370.

[Bil+19] Cyrille BILLAUDEAU, Zhizhong YAO, Charlène CORNILLEAU, Rut CARBALLIDO-LÓPEZ et Arnaud CHASTANET, « MreB Forms Subdiffraction Nanofilaments during Active Growth in Bacillus subtilis », in : *MBio* 10.*1* (2019), e01879-18.

[BKV18] Vincent BRIANE, Charles KERVRANN et Myriam VIMOND, « Statistical analysis of particle trajectories in living cells », in : *Physical Review E* 97.*6* (2018), p. 062121.

[BN13] Paul C BRESSLOFF et Jay M NEWBY, « Stochastic models of intracellular transport », in : *Reviews of Modern Physics* 85.*1* (2013), p. 135.

[Bou+14] Jérôme BOULANGER, Charles GUEUDRY, Daniel MÜNCH, Bertrand CINQUIN, Perrine PAUL-GILLOTEAUX, Sabine BARDIN, Christophe GUÉRIN, Fabrice SENGER, Laurent BLANCHOIN et Jean SALAMERO, « Fast high-resolution 3D total internal reflection fluorescence microscopy by incidence angle scanning and azimuthal averaging », in : *Proceedings of the National Academy of Sciences* 111.*48* (2014), p. 17164-17169.

[BP84] Lars G BURMAN et James T PARK, « Molecular model for elongation of the murein sacculus of Escherichia coli », in : *Proceedings of the National Academy of Sciences* 81.*6* (1984), p. 1844-1848.

[Bra+18] Benjamin P BRATTON, Joshua W SHAEVITZ, Zemer GITAI et Randy M MORGENSTEIN, « MreB polymers and curvature localization are enhanced by RodZ and predict E. coli's cylindrical uniformity », in : *Nature Communications* 9.*1* (2018), p. 1-11.

[Bri+20] Vincent BRIANE, Myriam VIMOND, Cesar Augusto VALADES-CRUZ, Antoine SALOMON, Christian WUNDER et Charles KERVRANN, « A sequential algorithm to detect diffusion switching along intracellular particle trajectories », in : *Bioinformatics* 36.*1* (2020), p. 317-329.

[BVK20] Vincent BRIANE, Myriam VIMOND et Charles KERVRANN, « An overview of diffusion models for intracellular dynamics analysis », in : *Briefings in Bioinformatics* 21.*4* (2020), p. 1136-1150.

[Cal16] Christopher P CALDERON, « Motion blur filtering : a statistical approach for extracting confinement forces and diffusivity from a single blurred trajectory », in : *Physical Review E* 93.*5* (2016), p. 053303.

[CBOM13] Nicolas CHENOUARD, Isabelle BLOCH et Jean-Christophe OLIVO-MARIN, « Multiple hypothesis tracking for cluttered biological image sequences », in : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.*11* (2013), p. 2736-3750.

[Cha43] Subrahmanyan CHANDRASEKHAR, « Stochastic problems in physics and astronomy », in : *Reviews of Modern Physics* 15.*1* (1943), p. 1.

[Che+14a] Bi-Chang CHEN, Wesley R LEGANT, Kai WANG, Lin SHAO, Daniel E MILKIE, Michael W DAVIDSON, Chris JANETOPOULOS, Xufeng S WU, John A HAMMER, Zhe LIU et al., « Lattice light-sheet microscopy : imaging molecules to embryos at high spatiotemporal resolution », in : *Science* 346.*6208* (2014).

[Che+14b] Nicolas CHENOUARD, Ihor SMAL, Fabrice DE CHAUMONT, Martin MAŠKA, Ivo F SBALZARINI, Yuanhao GONG, Janick CARDINALE, Craig CARTHEL, Stefano CORALUPPI, Mark WINTER et al., « Objective comparison of particle tracking methods », in : *Nature Methods* 11.*3* (2014), p. 281-289.

[Cho+16] Hongbaek CHO, Carl N WIVAGG, Mrinal KAPOOR, Zachary BARRY, Patricia DA ROHS, Hyunsuk SUH, Jarrod A MARTO, Ethan C GARNER et Thomas G BERNHARDT, « Bacterial cell wall biogenesis is mediated by SEDS and PBP polymerase families functioning semi-autonomously », in : *Nature microbiology* 1.*10* (2016), p. 1-8.

[CLF07] Rut CARBALLIDO-LÓPEZ et Alex FORMSTONE, « Shape determination in Bacillus subtilis », in : *Current opinion in microbiology* 10.*6* (2007), p. 611-616.

[Cor+20] Charlene CORNILLEAU, Arnaud CHASTANET, Cyrille BILLAUDEAU et Rut CARBALLIDO-LÓPEZ, « Methods for Studying Membrane-Associated Bacterial Cytoskeleton Proteins In Vivo by TIRF Microscopy », in : *Cytoskeleton Dynamics*, Springer, 2020, p. 123-133.

[DBKW81] WR DE BOER, FJ KRUYSSEN et JT WOUTERS, « Cell wall turnover in batch and chemostat cultures of Bacillus subtilis. », in : *Journal of Bacteriology* 145.*1* (1981), p. 50-60.

[DE+11] Julia DOMÍNGUEZ-ESCOBAR, Arnaud CHASTANET, Alvaro H CREVENNA, Vincent FROMION, Roland WEDLICH-SÖLDNER et Rut CARBALLIDO-LÓPEZ, « Processive movement of MreB-associated cell wall biosynthetic complexes in bacteria », in : *Science* 333.*6039* (2011), p. 225-228.

[Dem+17] Justin DEMMERLE, Cassandravictoria INNOCENT, Alison J NORTH, Graeme BALL, Marcel MÜLLER, Ezequiel MIRON, Atsushi MATSUDA, Ian M DOBBIE, Yolanda MARKAKI et Lothar SCHERMELLEH, « Strategic and practical guidelines for successful structured illumination microscopy », in : *Nature Protocols* 12.*5* (2017), p. 988-1010.

[Der+20] Simon DERSCH, Johanna MEHL, Lisa STUCKENSCHNEIDER, Benjamin MAYER, Julian ROTH, Alexander ROHRBACH et Peter L GRAUMANN, « Super-Resolution Microscopy and Single-Molecule Tracking Reveal Distinct Adaptive Dynamics of MreB and of Cell Wall-Synthesis Enzymes », in : *Frontiers in Microbiology* 11 (2020), p. 1946.

[Des17] Xavier DESCOMBES, « Multiple objects detection in biological images using a marked point process framework », in : *Methods* 115 (2017), p. 2-8.

[DG+19] Fabienne DE GRAEVE, Eric DEBREUVE, Somia RAHMOUN, Szilvia ECSEDI, Alia BAHRI, Arnaud HUBSTENBERGER, Xavier DESCOMBES et Florence BESSE, « Detecting and quantifying stress granules in tissues of multicellular organisms with the Obj. MPP analysis tool », in : *Traffic* 20.*9* (2019), p. 697-711.

[Din+17] Marc DINH, Loic STRAFELLA, Pierre FLORES, Arnaud CHASTANET, R CARBALLIDO-LÓPEZ et Vincent FROMION, « Validation of the 3-under-2 principle of cell wall growth in Gram-positive bacteria by simulation of a simple coarse-grained model », in : *arXiv preprint arXiv :1702.02747* (2017).

[Dio+19] Michael F DION, Mrinal KAPOOR, Yingjie SUN, Sean WILSON, Joel RYAN, Antoine VIGOUROUX, Sven VAN TEEFFELEN, Rudolf OLDENBOURG et Ethan C GARNER, « Bacillus subtilis cell diameter is determined by the opposing actions of two distinct cell wall synthetic systems », in : *Nature Microbiology* 4.*8* (2019), p. 1294-1305.

[DZK17] Xavier DESCOMBES, Elena ZHIZHINA et Sergey KOMECH, « View of mathematicians on biological data : modeling axon growth using CTRW », in : *Miždisciplinarni doslidžennja skladnih sistem 10-11* (2017), p. 35-42.

[EAL01] Fusinita van den ENT, Linda A AMOS et Jan LOÈWE, « Prokaryotic origin of the actin cytoskeleton », in : *Nature* 413.*6851* (2001), p. 39-44.

[Ein05] Albert EINSTEIN, « On the electrodynamics of moving bodies », in : *Annalen der Physik* 17.*10* (1905), p. 891-921.

[Ent+14] Fusinita Van den ENT, Thierry IZORÉ, Tanmay AM BHARAT, Christopher M JOHNSON et Jan LÖWE, « Bacterial actin MreB forms antiparallel double filaments », in : *Elife* 3 (2014), e02634.

[EWT06] Osigwe ESUE, Denis WIRTZ et Yiider TSENG, « GTPase activity, structure, and mechanical properties of filaments assembled from bacterial cytoskeleton protein MreB », in : *Journal of Bacteriology* 188.*3* (2006), p. 968-976.

[Fed+96] Toni J FEDER, Ingrid BRUST-MASCHER, James P SLATTERY, Barbara BAIRD et Watt W WEBB, « Constrained diffusion or immobile fraction on cell surfaces : a new interpretation », in : *Biophysical Journal* 70.*6* (1996), p. 2767-2773.

[Gar+11] Ethan C GARNER, Remi BERNARD, Wenqin WANG, Xiaowei ZHUANG, David Z RUDNER et Tim MITCHISON, « Coupled, circumferential motions of the cell wall synthesis machinery and MreB filaments in B. subtilis », in : *Science* 333.*6039* (2011), p. 222-225.

[GCJ08] Lu GAN, Songye CHEN et Grant J JENSEN, « Molecular organization of Gram-negative peptidoglycan », in : *Proceedings of the National Academy of Sciences* 105.*48* (2008), p. 18953-18957.

[Gen+06] Auguste GENOVESIO, Tim LIEDL, Valentina EMILIANI, Wolfgang J PARAK, Maité COPPEY-MOISAN et J-C OLIVO-MARIN, « Multiple particle tracking in 3-D+ t microscopy : method and application to the tracking of endocytosed quantum dots », in : *IEEE Transactions on Image Processing* 15.*5* (2006), p. 1062-1070.

[Gra06] Robert M GRAY, *Toeplitz and circulant matrices : A review*, now publishers inc, 2006.

[Gra+19a] Naor GRANIK, Lucien E WEISS, Elias NEHME, Maayan LEVIN, Michael CHEIN, Eran PERLSON, Yael ROICHMAN et Yoav SHECHTMAN, « Single-particle diffusion characterization by deep learning », in : *Biophysical Journal* 117.*2* (2019), p. 185-192.

[Gra+19b] Anca-Ioana GRAPA, Laure BLANC-FÉRAUD, Ellen van OBBERGHEN-SCHILLING et Xavier DESCOMBES, « Optimal Transport vs Many-to-many assignment for Graph Matching », in : *GRETSI 2019-XXVIIème Colloque Francophone de Traitement du Signal et des Images*, 2019.

[Gus00] Mats GL GUSTAFSSON, « Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy », in : *Journal of Microscopy* 198.*2* (2000), p. 82-87.

[HH15] Nathanael HOZE et David HOLCMAN, « Recovering a stochastic process from noisy ensembles of many single particle trajectories », in : *arXiv preprint arXiv :1509.02312* (2015).

[HH17] Nathanaël HOZÉ et David HOLCMAN, « Statistical methods for large ensembles of super-resolution stochastic single particle trajectories in cell biology », in : *Annual Review of Statistics and Its Application* 4 (2017), p. 189-223.

[Höl98] Joachim-Volker HÖLTJE, « Growth of the stress-bearing and shape-maintaining murein sacculus of Escherichia coli », in : *Microbiology and Molecular Biology Reviews* 62.*1* (1998), p. 181-203.

[Hoz+12] Nathanael HOZE, Deepak NAIR, Eric HOSY, Christian SIEBEN, Suliana MANLEY, Andreas HERRMANN, Jean-Baptiste SIBARITA, Daniel CHOQUET et David HOLCMAN, « Heterogeneity of AMPA receptor trafficking and molecular interactions revealed by superresolution analysis of live cell imaging », in : *Proceedings of the National Academy of Sciences* 109.*42* (2012), p. 17052-17057.

[HSH13] N. HOZE, Z. SCHUSS et D. HOLCMAN, « Reconstruction of surface and stochastic dynamics from a planar projection of trajectories », in : *SIAM Journal on Imaging Sciences* 6.*4* (2013), p. 2430-2449.

[Hua+08] Bo HUANG, Wenqin WANG, Mark BATES et Xiaowei ZHUANG, « Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy », in : *Science* 319.*5864* (2008), p. 810-813.

[Hus+18] Saman HUSSAIN, Carl N WIVAGG, Piotr SZWEDZIAK, Felix WONG, Kaitlin SCHAEFER, Thierry IZORÉ, Lars D RENNER, Matthew J HOLMES, Yingjie SUN, Alexandre W BISSON-FILHO et al., « MreB filaments align along greatest principal membrane curvature to orient cell wall synthesis », in : *Elife* 7 (2018), e32471.

[Jaq+08] K. JAQAMAN, D. LOERKE, M. METTLEN, H. KUWATA, S. GRINSTEIN, S. L. SCHMID et G. DANUSER, « Robust single-particle tracking in live-cell time-lapse sequences », in : *Nature Methods* 5.*8* (2008), p. 695.

[Kar14] Samuel KARLIN, *A first course in stochastic processes*, Academic press, 2014.

[KD85] Arthur L KOCH et Ronald J DOYLE, « Inside-to-outside growth and turnover of the wall of gram-positive rods », in : *Journal of Theoretical Biology* 117.*1* (1985), p. 137-157.

[KDE09] Yoshikazu KAWAI, Richard A DANIEL et Jeffery ERRINGTON, « Regulation of cell wall morphogenesis in Bacillus subtilis by recruitment of PBP1 to the MreB helix », in : *Molecular Microbiology* 71.*5* (2009), p. 1131-1144.

[Koc98] AL KOCH, « Orientation of the peptidoglycan chains in the sacculus of Escherichia coli », in : *Research in microbiology* 149.*10* (1998), p. 689-701.

[Kou+08] Samuel C KOU et al., « Stochastic modeling in nanoscale biophysics : subdiffusion within proteins », in : *The Annals of Applied Statistics* 2.*2* (2008), p. 501-535.

[KT81] Samuel KARLIN et Howard E TAYLOR, *A second course in stochastic processes*, Elsevier, 1981.

[Lan08] P. LANGEVIN, « Sur la théorie du mouvement brownien », in : *C. R. Acad. Sci. (Paris)* 146 (1908).

[Lau+19] François LAURENT, Charlotte FLODERER, Cyril FAVARD, Delphine MURIAUX, Jean-Baptiste MASSON et Christian L VESTERGAARD, « Mapping spatio-temporal dynamics of single biomolecules in living cells », in : *Physical Biology* 17.*1* (2019), p. 015003.

[Lee+16] Timothy K LEE, Kevin MENG, Handuo SHI et Kerwyn Casey HUANG, « Single-molecule imaging reveals modulation of cell wall synthesis dynamics in live bacterial cells », in : *Nature Communications* 7.*1* (2016), p. 1-9.

[LG20] Frédéric LAVANCIER et Ronan Le GUÉVEL, « Spatial birth-death-move processes : basic properties and estimation of their intensity functions », in : *arXiv preprint arXiv :2002.05423* (2020).

[Li+15] Dong LI, Lin SHAO, Bi-Chang CHEN, Xi ZHANG, Mingshu ZHANG, Brian MOSES, Daniel E MILKIE, Jordan R BEACH, John A HAMMER, Mithun PASHAM et al., « Extended-resolution structured illumination imaging of endocytic and cytoskeletal dynamics », in : *Science* 349.*6251* (2015).

[Liu+15] Yao LIU, Rong QIN, Sebastian AJ ZAAT, Eefjan BREUKINK et Michal HEGER, « Antibacterial photodynamic therapy : overview of a promising approach to fight antibiotic-resistant bacterial infections », in : *Journal of Clinical and Translational Research* 1.*3* (2015), p. 140.

[LL01] Ning LI et X Rong LI, « Target perceivability and its applications », in : *IEEE Transactions on Signal Processing* 49.*11* (2001), p. 2588-2604.

[LLB15] Zhe LIU, Luke D LAVIS et Eric BETZIG, « Imaging live-cell dynamics and structure at the single-molecule level », in : *Molecular Cell* 58.*4* (2015), p. 644-659.

[Luc+18] Daniella LUCENA, Marco MAURI, Felix SCHMIDT, Bruno ECKHARDT et Peter L GRAUMANN, « Microdomain formation is a general property of bacterial membrane proteins and induces heterogeneity of diffusion patterns », in : *BMC Biology* 16.*1* (2018), p. 1-17.

[Mar09] William MARGOLIN, « Sculpting the bacterial cell », in : *Current Biology* 19.*17* (2009), R812-R822.

[MB05] Valério RF MATIAS et Terry J BEVERIDGE, « Cryo-electron microscopy reveals native polymeric cell wall structure in Bacillus subtilis 168 and the existence of a periplasmic space », in : *Molecular Microbiology* 56.*1* (2005), p. 240-251.

[MB12] Xavier MICHALET et Andrew J BERGLUND, « Optimal diffusion coefficient estimation in single-particle tracking », in : *Physical Review E* 85.*6* (2012), p. 061916.

[MCB13] Katarína MUCHOVÁ, Zuzana CHROMIKOVÁ et Imrich BARÁK, « Control of Bacillus subtilis cell shape by RodZ », in : *Environmental Microbiology* 15.*12* (2013), p. 3259-3271.

[Mee+16] Alexander J MEESKE, Eammon P RILEY, William P ROBINS, Tsuyoshi UEHARA, John J MEKALANOS, Daniel KAHNE, Suzanne WALKER, Andrew C KRUSE, Thomas G BERNHARDT et David Z RUDNER, « SEDS proteins are a widespread family of bacterial cell wall polymerases », in : *Nature* 537.*7622* (2016), p. 634-638.

[Meu09] Attilio MEUCCI, « Review of statistical arbitrage, cointegration, and multivariate Ornstein-Uhlenbeck », in : *Cointegration, and Multivariate Ornstein-Uhlenbeck (May 14, 2009)* (2009).

[MFTW13] ML MARTIN-FERNANDEZ, CJ TYNAN et SED WEBB, « A 'pocket guide'to total internal reflection fluorescence », in : *Journal of Microscopy* 252.*1* (2013), p. 16-22.

[Mic10] Xavier MICHALET, « Mean square displacement analysis of single-particle trajectories with localization error : Brownian motion in an isotropic medium », in : *Physical Review E* 82.*4* (2010), p. 041914.

[Mor+15] Randy M MORGENSTEIN, Benjamin P BRATTON, Jeffrey P NGUYEN, Nikolay OUZOUNOV, Joshua W SHAEVITZ et Zemer GITAI, « RodZ links MreB to cell wall synthesis to mediate MreB rotation and robust morphogenesis », in : *Proceedings of the National Academy of Sciences* 112.*40* (2015), p. 12510-12515.

[MVN68] Benoit B MANDELBROT et John W VAN NESS, « Fractional Brownian motions, fractional noises and applications », in : *SIAM Review* 10.*4* (1968), p. 422-437.

[Ngu+15] Lam T NGUYEN, James C GUMBART, Morgan BEEBY et Grant J JENSEN, « Coarse-grained simulations of bacterial cell wall growth reveal that local coordination alone can be sufficient to maintain rod shape », in : *Proceedings of the National Academy of Sciences* 112.*28* (2015), E3689-E3698.

[Oks13] Bernt OKSENDAL, *Stochastic differential equations : an introduction with applications*, Springer Science & Business Media, 2013.

[Ols+13] Philipp v OLSHAUSEN, Hervé Joël Defeu SOUFO, Kai WICKER, Rainer HEINTZMANN, Peter L GRAUMANN et Alexander ROHRBACH, « Superresolution imaging of dynamic MreB filaments in B. subtilis—a multiple-motor-driven transport ? », in : *Biophysical Journal* 105.*5* (2013), p. 1171-1181.

[PB+10] Catherine PARADIS-BLEAU, Monica MARKOVSKI, Tsuyoshi UEHARA, Tania J LUPOLI, Suzanne WALKER, Daniel E KAHNE et Thomas G BERNHARDT, « Lipoprotein cofactors located in the outer membrane activate bacterial cell wall polymerases », in : *Cell* 143.*7* (2010), p. 1110-1120.

[Per+15] Ana C PEREIRA, Ana PAIVA, Ivo H SARAIVA, Teresa COSTA, Adriano O HENRIQUES et Manolis MATZAPETAKIS, « Chemical shift assignments and secondary structure determination of the ectodomain of Bacillus subtilis morphogenic protein RodZ », in : *Biomolecular NMR Assignments* 9.*2* (2015), p. 285-288.

[Pis+15] AS PISAREV, SA RUKOLAINE, AM SAMSONOV et MG SAMSONOVA, « Numerical analysis of particle trajectories in living cells under uncertainty conditions », in : *Biophysics* 60.*5* (2015), p. 810-817.

[Poo76] HM POOLEY, « Turnover and spreading of old wall during surface growth of Bacillus subtilis. », in : *Journal of Bacteriology* 125.*3* (1976), p. 1127-1138.

[PU08] James T PARK et Tsuyoshi UEHARA, « How bacteria consume their own exoskeletons (turnover and recycling of cell wall peptidoglycan) », in : *Microbiology and Molecular Biology Reviews* 72.*2* (2008), p. 211-227.

[QSE91] Hong QIAN, Michael P SHEETZ et Elliot L ELSON, « Single particle tracking. Analysis of diffusion and flow in two-dimensional systems », in : *Biophysical Journal* 60.*4* (1991), p. 910-921.

[Ran71] William M RAND, « Objective criteria for the evaluation of clustering methods », in : *Journal of the American Statistical association* 66.*336* (1971), p. 846-850.

[Ray80] Lord RAYLEIGH, « Investigations in optics, with special reference to the spectroscope », in : *Monthly Notices of the Royal Astronomical Society* 40 (1880), p. 254.

[Rös+18] Thomas C RÖSCH, Luis M OVIEDO-BOCANEGRA, Georg FRITZ et Peter L GRAUMANN, « SMTracker : a tool for quantitative analysis, exploration and visualization of single-molecule tracking data reveals highly dynamic binding of B. subtilis global repressor AbrB throughout the genome », in : *Scientific Reports* 8.*1* (2018), p. 1-12.

[Rue+14] Anne-Stéphanie RUEFF, Arnaud CHASTANET, Julia DOMÍNGUEZ-ESCOBAR, Zhizhong YAO, James YATES, Maria-Victoria PREJEAN, Olivier DELUMEAU, Philippe NOIROT, Roland WEDLICH-SÖLDNER, Sergio R FILIPE et al., « An early cytoplasmic step of peptidoglycan synthesis is associated to MreB in B acillus subtilis », in : *Molecular Microbiology* 91.*2* (2014), p. 348-362.

[Sal+11] Jeanne SALJE, Fusinita van den ENT, Piet de BOER et Jan LÖWE, « Direct membrane binding by bacterial actin MreB », in : *Molecular Cell* 43.*3* (2011), p. 478-487.

[Sch+08] Lothar SCHERMELLEH, Peter M CARLTON, Sebastian HAASE, Lin SHAO, Lukman WINOTO, Peter KNER, Brian BURKE, M Cristina CARDOSO, David A AGARD, Mats GL GUSTAFSSON et al., « Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy », in : *Science* 320.*5881* (2008), p. 1332-1336.

[Sch09] Zeev SCHUSS, *Theory and applications of stochastic processes : an analytical approach*, t. 170, Springer Science & Business Media, 2009.

[Sch15] E. SCHRODINGER, « Zur Theorie der Fall-und Steigversuche an Teilchen mit Brownscher Bewegung », in : *Physikalische Zeitschrift* 16 (1915), 289–295.

[Sch+17] Katrin SCHENK, Ana B HERVÁS, Thomas C RÖSCH, Marc EISEMANN, Bernhard A SCHMITT, Stephan DAHLKE, Luise KLEINE-BORGMANN, Seán M MURRAY et Peter L GRAUMANN, « Rapid turnover of DnaA at replication origin regions contributes to initiation control of DNA replication », in : *PLoS Genetics* 13.*2* (2017), e1006561.

[Sch+19] Lothar SCHERMELLEH, Alexia FERRAND, Thomas HUSER, Christian EGGELING, Markus SAUER, Oliver BIEHLMAIER et Gregor PC DRUMMEN, « Super-resolution microscopy demystified », in : *Nature Cell Biology* 21.*1* (2019), p. 72-84.

[Ser+20] Alexander S SEROV, François LAURENT, Charlotte FLODERER, Karen PERRONET, Cyril FAVARD, Delphine MURIAUX, Nathalie WESTBROOK, Christian L VESTERGAARD et Jean-Baptiste MASSON, « Statistical tests for force inference in heterogeneous environments », in : *Scientific Reports* 10.*1* (2020), p. 1-12.

[SJ97] Michael J SAXTON et Ken JACOBSON, « Single-particle tracking : applications to membrane dynamics », in : *Annual Review of Biophysics and Biomolecular Structure* 26.*1* (1997), p. 373-399.

[SM75] Harvey SCHER et Elliott W MONTROLL, « Anomalous transit-time dispersion in amorphous solids », in : *Physical Review B* 12.*6* (1975), p. 2455.

[SSN08] Daisuke SHIOMI, Masako SAKAI et Hironori NIKI, « Determination of bacterial rod shape by a novel cytoskeletal membrane protein », in : *The EMBO Journal* 27.*23* (2008), p. 3081-3091.

[SSS97] Gerhard J SCHÜTZ, Hansgeorg SCHINDLER et Thomas SCHMIDT, « Single-molecule microscopy on model membranes reveals anomalous diffusion. », in : *Biophysical Journal* 73.*2* (1997), p. 1073.

[SWB08] Hari SHROFF, Helen WHITE et Eric BETZIG, « Photoactivated localization microscopy (PALM) of adhesion complexes », in : *Current Protocols in Cell Biology* 41.*1* (2008), p. 4-21.

[TLW02] Russell E THOMPSON, Daniel R LARSON et Watt W WEBB, « Precise nanometer localization analysis for individual fluorescent probes », in : *Biophysical Journal* 82.*5* (2002), p. 2775-2783.

[TM13] Silvan TÜRKCAN et Jean-Baptiste MASSON, « Bayesian decision tree for the classification of the mode of motion in single-molecule trajectories », in : *PloS ONE* 8.*12* (2013), e82799.

[TR18] Sven van TEEFFELEN et Lars D RENNER, « Recent advances in understanding how rod-like bacteria stably maintain their cell shapes », in : *F1000Research* 7 (2018).

[Twe45] Maurice CK TWEEDIE, « Inverse statistical variates », in : *Nature* 155.*3937* (1945), p. 453.

[Typ+10] Athanasios TYPAS, Manuel BANZHAF, Bart van den Berg van SAPAROEA, Jolanda VERHEUL, Jacob BIBOY, Robert J NICHOLS, Matylda ZIETEK, Katrin BEILHARZ, Kai KANNENBERG, Moritz von RECHENBERG et al., « Regulation of peptidoglycan synthesis by outer-membrane proteins », in : *Cell* 143.*7* (2010), p. 1097-1109.

[UO30] George E UHLENBECK et Leonard S ORNSTEIN, « On the theory of the Brownian motion », in : *Physical Review* 36.*5* (1930), p. 823.

[Urs+14] Tristan S URSELL, Jeffrey NGUYEN, Russell D MONDS, Alexandre COLAVIN, Gabriel BILLINGS, Nikolay OUZOUNOV, Zemer GITAI, Joshua W SHAEVITZ et Kerwyn Casey HUANG, « Rod-like bacterial shape is maintained by feedback between cell curvature and cytoskeletal localization », in : *Proceedings of the National Academy of Sciences* 111.*11* (2014), E1025-E1034.

[VBF14] Christian L VESTERGAARD, Paul C BLAINEY et Henrik FLYVBJERG, « Optimal estimation of diffusion coefficients from single-particle trajectories », in : *Physical Review E* 89.*2* (2014), p. 022726.

[Ver+21] Hippolyte VERDIER, Maxime DUVAL, François LAURENT, Alhassan CASSÉ, Christian VESTERGAARD et Jean-Baptiste MASSON, « Learning physical properties of anomalous random walks using graph neural networks », preprint, url=https ://hal-pasteur.archives-ouvertes.fr/pasteur-03150190/document, 2021.

[VH04] Waldemar VOLLMER et Joachim-Volker HÖLTJE, « The architecture of the murein (peptidoglycan) in gram-negative bacteria : vertical scaffold or horizontal layer (s) ? », in : *Journal of Bacteriology* 186.*18* (2004), p. 5978-5987.

[VT+11] Sven VAN TEEFFELEN, Siyuan WANG, Leon FURCHTGOTT, Kerwyn Casey HUANG, Ned S WINGREEN, Joshua W SHAEVITZ et Zemer GITAI, « The bacterial actin MreB rotates, and rotation depends on cell-wall assembly », in : *Proceedings of the National Academy of Sciences* 108.*38* (2011), p. 15822-15827.

[Wal73] A. WALD, *Sequential Analysis*, Courier Corporation, 1973.

[WGA19] Felix WONG, Ethan C GARNER et Ariel AMIR, « Mechanics and dynamics of translocating MreB filaments on curved membranes », in : *Elife* 8 (2019), e40472.

[Won+17] Felix WONG, Lars D RENNER, Gizem ÖZBAYKAL, Jayson PAULOSE, Douglas B WEIBEL, Sven VAN TEEFFELEN et Ariel AMIR, « Mechanical strain sensing implicated in cell shape recovery in Escherichia coli », in : *Nature microbiology* 2.*9* (2017), p. 1-8.

[WSW08] JM WILLEY, LM SHERWOOD et CJ WOOLVERTON, « Bacteria : the low G+ C Gram positives », in : *Prescott, Harley, and Klein's Microbiology, 7th Ed. McGraw-Hill Higher Education* (2008), p. 571-588.

[Yan+12] Huei-Fang YANG, Xavier DESCOMBES, Charles KERVRANN, Caroline MEDIONI et Florence BESSE, « Tracking growing axons by particle filtering in 3D+ t fluorescent two-photon microscopy images », in : *Asian Conference on Computer Vision*, Springer, 2012, p. 272-283.

[ZCZ20] Ziwei ZHANG, Peng CUI et Wenwu ZHU, « Deep learning on graphs : A survey », in : *IEEE Transactions on Knowledge and Data Engineering* (2020).

[Zwa01] Robert ZWANZIG, *Nonequilibrium statistical mechanics*, Oxford University Press, 2001.

[ZZOM07] Bo ZHANG, Josiane ZERUBIA et Jean-Christophe OLIVO-MARIN, « Gaussian approximations of fluorescence microscope point-spread function models », in : *Applied Optics* 46.*10* (2007), p. 1819-1829.

ABBREVIATIONS

2D two-dimensional

3D three-dimensional

AFM Atomic force microscopy

ARI adjusted Rand Index

Bmd Brownian motion with drift

CCD Charge-Coupled Device

CRLB Cramer-Rao Lower Bound

CryoET Electron cryotomography

CTRW Continuous Time Random Walk

CVE Covariance-based Estimator

CW cell wall

fBm fractional Brownian motion

GFP Green fluorescence protein

GMM Gaussian Mixture Model

Gram+ Gram-positive

Gram- Gram-negative

LLSM Lattice Light Sheet microscopy

MBF Motion Blur Filter

MLE Maximum Likelihood Estimation

MPG mixed-Poisson-Gaussian

MSD Mean Squared Displacement

NAG N-acetylglucosamine

NAM N-Acetylmuramic acid

OLSF Optimal Least Squared Fitting

OU Ornstein-Uhlenbeck

PALM Photoactivated Localization microscopy

PG peptidoglycan

PGEM Peptidoglycan Elongation Machinery

PSF Point-Spread Function

ROO regions of observation

SD Standard Deviation

SDE Stochastic Differential Equation

SIM Structured Illumination microscopy

SMT Single Molecule Tracking

SNR Signal to Noise Ratio

SPT single-particle Tracking

SRM Super-Resolution microscopy

STED Stimulated Emission Depletion

STORM Stochastic Optical Reconstruction microscopy

THOTH Testing HypOtheses for diffusion TricHotomy

TIRF Total Internal Reflection Fluorescent

WFFM Wide-Field Fluorescence microscopy

List of publications

Journal paper

Y. Lu, P. Hodara, C. Kervrann, A. Trubuil, Probabilistic reconstruction of truncated particle trajectories on a closed surface, *Multiscale Modeling and Simulation* 19.1 (2021), p. 87-112

Communications in national seminars

Seminar in MaIAGE unit, INRAE, Jouy-en-Josas, Oct 2019

Seminar in Micalis unit, INRAE, Jouy-en-Josas, Nov 2019

Communications in international conferences and workshops

Y. Lu, P. Hodara, A. Trubuil, C. Kervrann, Probabilistic overall reconstruction of membrane-associated molecular dynamics from partial observations in rod-shaped bacteria, Selected talk in *Quantitative BioImaging*, Oxford, United Kingdom, 6th Jan 2020

Y. Lu, C. Billaudeau, R. Carballidoz-Lopez, A. Trubuil, C. Kervrann, 3D stochastic process simulation for better interpretation of molecular dynamics related to cell wall biogenesis observed with TIRF microscopy, Poster presentation in *Quantitative BioImaging*, Rennes, France, Jan 2019

Titre : Dynamiques intracellulaires et imagerie de super résolution : la paroi bacterienne sondée à l’echelle moléculaire

Mot clés : Analyse des trajectoires de particules uniques, processus de diffusion, vidéomicroscopie de haute résolution, reconstruction/ traitement des images, modélisation probabilistique, estimation paramétrique

Resumé : Dans cette thèse, les méthodes de modélisation physique, statistique, et mathématique pour l’analyse des trajectoires de particules uniques ont été étudiées, accompagnées d’une évaluation expérimentale sur un problème spécifique, concernant la dynamique d’agrégats de MreB. Dans un premier temps, nous avons proposé deux approches pour analyser les trajectoires de MreB à l’aide de processus de diffusion caractérisés par des équations différentielles stochastiques. La première approche consiste à analyser les trajectoires individuelles, à déterminer leur mode de diffusion (diffusion libre, super-diffusion ou sous-diffusion) et à estimer les caractéristiques de diffusion (dérive et coefficient de diffusion). La seconde approche consiste à étudier la dynamique des particules au voisinage d’un point dans le champ, et à obtenir ensuite le champ de dérive et le champ de diffusion. Dans la seconde partie de la thèse, une modélisation probabiliste est proposée afin de reconstruire des dynamiques de particules en mouvement sur une surface cylindrique à partir d’observations partielles de ces dernières. La modélisation s’appuie d’une part, sur un processus de naissance et mort et d’autre part un processus Brownien avec dérive. Elle a donné lieu à une procédure de reconstruction autonome par maximisation de vraisemblance. Cette procédure est évaluée sur des données de synthèse ainsi que des données réelles de dynamiques de MreB obtenue en microscopie TIRFM.

Title: Intracellular dynamics and super-resolution imaging: the bacterial wall probed at the molecular scale

Keywords : Single-particle trajectory analysis, diffusion process, time-lapse high-resolution microscopy, image reconstruction/ processing, probabilistic modeling, parametric estimation

Abstract: In this thesis, physical, statistical and mathematical modeling methods for the analysis of single particle trajectories have been reviewed, accompanied by experimental evaluation on a specific problem, concerning the dynamics of MreB. Firstly, we proposed two approaches to analyze the trajectories of MreB based on diffusion characterized by Stochastic Differential Equations. The first approach consists in analyzing individual trajectories, determining their diffusion nature (free diffusion, super-diffusion or sub-diffusion) and estimating diffusion features (drift and diffusion coefficient). The second approach amounts to investigating the dynamics of particles in the neighborhood of a position point, and further obtaining the drift field and the diffusion field. In the second part of the thesis, a probabilistic modeling has been proposed in order to reconstruct the dynamics of moving particles on a cylindrical surface from partial observations of the dynamics. The modeling was based on a birth and death process on the one hand, and a Brownian process with drift on the other hand. It gave rise to an autonomous reconstruction procedure by maximizing likelihood. This procedure was evaluated on synthetic data as well as real data of MreB dynamics obtained in TIRFM microscopy.