UNIVERSITY OF CALIFORNIA
RIVERSIDE

Modeling, Simulation, and High-Performance Implementation of High-Dimensional Micro-Macro Biophysical Models

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Computer Science

by

Steven Cook

December 2017

Dissertation Committee:

Dr. Tamar Shinar, Chairperson
Dr. Zizhong Chen
Dr. Stefano Lonardi
Dr. Monica Martinez-Wilhelmus

The Dissertation of Steven Cook is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I am grateful to my advisor, Dr. Tamar Shinar, for the many forms of guidance provided on this challenging journey. I thank Dr. Christel Hohenegger for invaluable mathematical and modeling assistance. I have fond memories of the hours spent brainstorming together with both.

Chapter 2 has been published in the SIAM Journal on Applied Mathematics, 74.5 (2014), pages 1338-1353. Chapter 3 has been accepted to the SIAM Journal on Multiscale Modeling and Simulation.

To Kelsey -

My love is always just as she seems

A force of nature all her own

to be reckoned with

Whatever's wrong with me

her kiss redeems

And it's all there

in my dreams

ABSTRACT OF THE DISSERTATION

Modeling, Simulation, and High-Performance Implementation of High-Dimensional Micro-Macro Biophysical Models

by

Steven Cook

Doctor of Philosophy, Graduate Program in Computer Science
University of California, Riverside, December 2017
Dr. Tamar Shinar, Chairperson

Gliding assays, reduced in vitro model systems where motor proteins adsorbed onto a planar substrate bind to and move laments, exhibit large-scale dynamic patterns like coherent swarming motion and density waves. These systems are sensitive to microscopic behavior such as the motor protein binding and unbinding dynamics, which take place on a faster timescale than the direct and uid-mediated lament interactions. A micro-macro multiscale modeling and simulation framework for gliding assays is developed, allowing for detailed microscopic motor modeling as well as both steric and hydrodynamic interactions between laments. A scalable hybrid CPU and multi-GPU implementation alleviates the cost associated with tracking the high-dimensional microstructure, achieving performance over 500 times greater than a single-threaded implementation.

Contents

List of Figures **ix**

List of Tables **x**

1 Introduction **1**

2 Dimensional reduction of a multiscale continuum model of microtubule gliding assays **3**
2.1 Introduction 4
2.2 Model 7
2.2.1 Filament conservation equation 7
2.2.2 Fluid equations 9
2.2.3 Bound and free motor equations 10
2.3 Nondimensionalized Equations 12
2.3.1 Fluid equations 12
2.3.2 Filament equations 13
2.3.3 Bound and free motor equations 13
2.4 Two-dimensional Reduction 14
2.4.1 Fluid equations 15
2.4.2 Filament equations 17
2.4.3 Bound and free motor equations 19
2.5 Numerical method 20
2.6 Results 22
2.7 Conclusions 25

3 A Micro-Macro Framework for Analyzing Steric and Hydrodynamic Interactions in Gliding Assays **30**
3.1 Introduction 31
3.2 Modeling framework 35
3.2.1 Microscale model 35
3.2.2 Two-dimensional reduction 43
3.2.3 Nondimensionalization 47

3.3 Implementation . . . 52
3.3.1 Microtubule density . . . 53
3.3.2 Motor distributions . . . 55
3.3.3 Fluid . . . 56
3.4 Results . . . 58
3.4.1 Evolved motor model . . . 59
3.4.2 Evolved motor model with self-propulsion . . . 66
3.4.3 Simplified motor model . . . 69
3.5 Conclusions . . . 70

4 Enabling Simulation of High-Dimensional Micro-Macro Biophysical Models through Hybrid CPU and Multi-GPU Parallelism 78
4.1 Introduction . . . 79
4.2 Motility Assay model . . . 81
4.3 Acceleration Methodology . . . 86
4.3.1 Holistic Algorithmic and Data Flow Restructuring . . . 86
4.3.2 GPU Optimizations . . . 89
4.3.3 Scaling to Multiple GPUs . . . 93
4.4 Results . . . 94
4.5 Conclusions . . . 96

5 Conclusions 102

Bibliography 105

List of Figures

2.1 Microtubule gliding assay 7
2.2 Filament clump evolution 23
2.3 Emergence of ordered subregions 29

3.1 Microtubule gliding assay 36
3.2 Motor brotein binding geometry 39
3.3 Nonuniform z grid 57
3.4 Single microtubule clump, bound motor density evolution 60
3.5 Single microtubule clump, microtubule density evolution 61
3.6 Perpendicular colliding clumps 62
3.7 Antiparallel colliding clumps 63
3.8 Vortex lattice experiment 65
3.9 Perturbation experiment 66
3.10 Colliding clump experiment with self-propulsion 67
3.11 Vortex ring experiment with self-propulsion 75
3.12 Perturbation experiment with self-propulsion 76
3.13 Vortex comparison with different motor models 76
3.14 Colliding clump experiment with simplified motor model 77
3.15 Perturbation experiment with and without motor-based fluid forces 77

4.1 Microtubule gliding assay 81
4.2 Data residency and computation 87
4.3 New time step evolution 89
4.4 MPI control flow for multiple GPUs and nodes 94
4.5 GPU optimization effects 100
4.6 Multi-GPU scaling of bound motor density and motor force computations . 101
4.7 Sample simulation output 101

List of Tables

2.1 Physical parameters and values 27
2.2 Summary of equations 28

3.1 Summary of steric equations 73
3.2 Model parameters and variables. 74

4.1 Summary of model equations 98
4.2 Single-node, single-GPU optimizations 99
4.3 Overall speedup for multi-GPU approach 99

Chapter 1

Introduction

This thesis is concerned with the development of a modeling and simulation framework for gliding motility assays of filaments and bound motor proteins. Of interest in the study of such assays is the relationship between their complex emergent macroscale behavior (vortex lattices, density waves, spatiotemporally persistent clumps) and the underlying microscopic interactions (direct, motor-mediated, and hydrodynamic) between the constituent filaments and motor proteins. A broader impact of understanding such systems is shedding new light on the organizational principles of the mitotic spindle, a dynamic yet persistent network of filaments and motors. In addition to being vital to the cell lifecycle, mitotic spindles are a prominent anti-cancer target. A useful review of the relevant field of active matter physics is found in [43], and details of the mechanics of the underlying motor proteins and filaments can be found in [28].

Direct simulation of such systems and their interactions grows infeasible at larger scales. This work instead uses a micro-macro approach, in which kinetic theory is used to

describe the configuration of the microstructural elements (motor proteins and filaments) as a density function, and this density is then coupled to the macroscopic fluid equations via extra stress terms. These methods enable detailed modeling of the microscopic interactions, and thus provide a valuable tool for testing hypotheses regarding the aforementioned relationship between macroscale and microscale phenomena. This microscopic fidelity comes at the expense of tracking the microstructural density, the dimensionality of which grows with the degrees of freedom of the microstructure. To overcome this computational challenge, a variety of algorithmic and modeling approaches are used to develop a hybrid CPU and multi-GPU implementation. The interested reader is referred to [31] for a review of micro-macro modeling.

The thesis is structured as follows. Chapter 2 details the development and implementation of the initial model. Chapter 3 expands the modeling and simulation framework with an additional motor behavior model, steric interactions, a passive filament advection velocity to overcome resolution limitations, and a more general three-dimensional fluid grid. Chapter 4 features a holistic restructuring of the algorithm and dataflow that breaks dependencies and parallelizes the system for multi-GPU. Chapter 5 summarizes findings and outlines limitations and directions for future investigation.

Chapter 2

Dimensional reduction of a multiscale continuum model of microtubule gliding assays

Microtubule gliding assays, in which molecular motors anchored to a plate drive the gliding motion of filaments in a quasi-two-dimensional fluid layer, have been shown to organize into a variety of large-scale patterns. We derive a fully three-dimensional multiscale coarse-grained model of a gliding assay including the evolution of densities of rigid filaments, bound motors, and free motors, coupled to fluid equations. Our model combines continuum theories of polymeric liquids with the force spreading approach of the immersed boundary method. We use dimensional and asymptotic analysis to derive a reduced two-dimensional model and show that, to leading order, the filaments evolve in a plane, similar to what is experimentally observed. We simulate our model numerically with a GPU-based

implementation and observe the same qualitative behavior as in experimental work.

2.1 Introduction

Active gels form the basic structural network of the cell and can be synthesized by mixing cytoskeletal filaments and small motor proteins. In such mixtures, the collective behavior is driven by the microscopic ATPase-generated motion of the motor proteins along the filaments. A molecular motor typically consists of a tail, a flexible stalk, and a head with two binding sites. Through the hydrolysis of ATP at the binding sites, the motor grabs, pulls and releases the filament yielding a processive directional motion and generating forces inside the cell. Molecular motors have been the subject of many experimental [16, 14, 26, 28, 22, 10] and theoretical [44, 35, 33] studies, which have resulted in a broader understanding of the molecular stepping mechanism and the collaboration of many motors when carrying cargo along filamentous tracks. On a larger scale, mixtures of cytoskeletal filaments and small motor proteins have been shown to self-organize into a variety of macroscopic patterns from asters to vortices to swarms [47, 50, 61]. We study gliding assays, where single motors are adsorbed onto a substrate and filaments, stabilized so that their lengths are fixed, glide over them in a quasi-two-dimensional fluid layer (Fig. 3.1). This reduced system, where filaments do not crosslink, is used to study gliding velocities of microtubules and motor directionality [55] and has been observed to form large-scale patterns [61].

Systems of fluids, filaments, and motors are inherently multiscale in space and time, making them a challenge to accurately model. Moreover, [67] showed that the macroscopic behavior of these active mixtures is sensitive to the microscopic model. A variety

of theoretical models attempt to bridge the understanding of the microscopic mechanisms with the emergent macroscopic phenomena. One computationally intensive approach is to model collections of explicitly represented filaments and motors without coarse-graining [46, 50, 48]. In contrast, other modeling standpoints rely on a generic hydrodynamic theory that is inherently macroscopic [40, 58, 37, 38, 21]. Considering gliding assays specifically, [36, 32] developed a Langevin dynamics model coupling the dynamics of the filament, the dynamics of the motor head and the elasticity of the motor stalks. Other approaches have exhibited pattern formation through phenomenological particle-based steric interactions [61, 65], while [21] demonstrated the importance of hydrodynamic effects in the collective behavior of large-scale systems of filaments. A final set of approaches starts with a simplified microscopic model and coarse-grains the system via *a priori* closure approximations to attain a macroscopic description [67, 42]. We take such an approach while additionally considering fluctuations in motor and filament densities, the background flow, and the interactions of the immersed structure and the fluid. This allows us to directly incorporate knowledge of the microscale mechanisms while maintaining the efficiency of coarse-grained approaches over particle-based approaches in simulating hydrodynamic interactions.

As a first step towards a multiscale model of gliding assays, we develop a three-dimensional model, including hydrodynamics effects via the coupling of the filaments and the fluid, but assume a dilute suspension of filaments thus neglecting steric effects. In contrast to previous modeling approaches, we do not assume that the filaments move in a two-dimensional plane, but instead we show that this results from an asymptotic analysis and Hele-Shaw reduction of the fluid-coupled continuum model. Furthermore, our modeling

approach couples the microscale modeling of the attachment and detachment of molecular motors to the mesoscale motion of the filaments, and finally to the macroscale fluid motion. This explicit coupling is obtained by combining continuum theories of polymeric liquids [5] with the force spreading approach of the immersed boundary method [53]. There are three components of our continuum model: a filament conservation equation, conservation equations for bound and free motors, and fluid equations. These coupled equations evolve on three length scales: L, the length of the side of the cover slip, l, the length of the filament, and L_m, the length of the motor stalk. We assume that $L_m \ll H = \varepsilon L \ll L$, where H is the distance between the cover slip and the plate.

The remainder of the paper is organized as follows. In Section 4.2, we derive the full three-dimensional continuum model focusing on the fluid coupling. Next, in Section 3.2.3, we nondimensionalize the model and neglect small terms. In Section 2.4, we obtain reduced two-dimensional gap-averaged equations through an asymptotic analysis and depth averaging, and we show that in this reduced formulation the filaments move in a two-dimensional plane parallel to the bottom plate. Section 2.5 is devoted to the numerical implementation, which is spectral for the fluid equations and second-order in time and space for the evolution equations. Difficulties resulting from the high dimensionality are resolved by exploiting data sparsity and using GPU parallelism. Finally, Section 2.6 illustrates the reduced model and numerical implementation for a set of parameters chosen to be experimentally relevant. In this case, our simulation shows the emergence of ordered subregions of filaments and motors as observed experimentally. We end with a short conclusion in Section 2.7.

2.2 Model

In a gliding assay, motor protein tails are anchored to a horizontal plate while their heads are free to bind and pull filaments suspended in a liquid. Figure 3.1 illustrates the experimental setup. Since there are no free motor complexes in suspension and a motor can only attach to a single filament, there is no cross-linking of filaments via motor complexes. In Table 3.2, we summarize the physical parameters relevant for the gliding assay geometry,

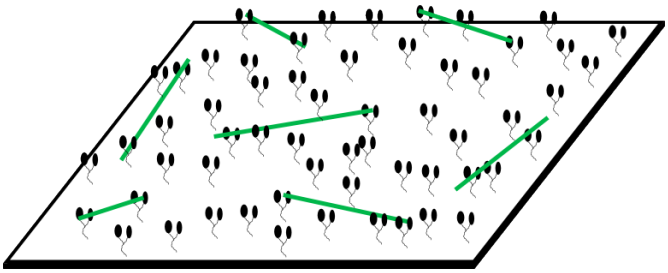


Figure 2.1: Microtubule gliding assay setup. Motor protein (black) tails are anchored to a fixed plate, while their heads bind and pull microtubule filaments (green) suspended in a liquid.

based on values reported in the literature.

2.2.1 Filament conservation equation

For a dilute suspension of filaments, we denote by $\Psi(\mathbf{x}, \mathbf{p}, t)$ the three-dimensional probability density function of the filament center-of-mass position $\mathbf{x}$ and orientation $\mathbf{p}$ with $|\mathbf{p}| = 1$. Here, we parametrize a *rigid, fixed length* filament by $\mathbf{x}+s\mathbf{p}$, where $s \in [-l, l]$ is the arc length parameter and we neglect bending, growth or shrinkage of filaments. The total number of filaments $N = \iint \Psi(\mathbf{x}, \mathbf{p}, t) d\mathbf{x} d\mathbf{p}$ is conserved and the Smoluchowski equation or

the conservation of mass in probability for Ψ is [4, 3, 11]

$$\partial_t \Psi + \nabla_x \cdot (\dot{\mathbf{x}}\Psi) + \nabla_p \cdot (\dot{\mathbf{p}}\Psi) = 0. \tag{2.1a}$$

To close Eq. (2.1a), we use equations for the $\mathbf{x}, \mathbf{p}$ dynamics derived from slender-body theory as in [57, 56] for active swimmers

$$\dot{\mathbf{x}}_2 = \mathbf{u}_2(\mathbf{x}) - D_{\mathrm{t},||} \nabla_{x_2} \ln \Psi \tag{2.1b}$$

$$\dot{z} = w(\mathbf{x}) - D_{\mathrm{t},\perp} \partial_z \ln \Psi \tag{2.1c}$$

$$\dot{\mathbf{p}} = (\mathbf{I} - \mathbf{pp}) \nabla_x \mathbf{u}(\mathbf{x}) \mathbf{p} - D_{\mathrm{r}} \nabla_p \ln \Psi. \tag{2.1d}$$

Here $\nabla_p = (\mathbf{I} - \mathbf{pp})\partial_{\mathbf{p}}$ is the gradient on the unit sphere, $\mathbf{u} = (\mathbf{u}_2, w)$ is the fluid velocity, and the subscript 2 denotes the in-plane coordinates x, y. To obtain Eq. (2.1b)-(2.1d), we assumed that the background flow $\mathbf{u}$ is linear along the slender and rigid filaments. Thus $\mathbf{u}$ can be written as $\mathbf{u}(\mathbf{x} + s\mathbf{p}) = \mathbf{u}(\mathbf{x}) + s\nabla \mathbf{u}(\mathbf{x})\mathbf{p}$. While in the theory of active swimmers, a swimming velocity appears in Eq. (2.1b), in our model the motor force acts directly on the fluid and indirectly on the filaments as they move passively with the fluid. This is similar to the approach taken in the immersed boundary method, which was originially developed for flexible passive fibers and membranes immersed in fluids [53]. In our case, the filaments are rigid, resulting in an additional stress on the fluid due to the filaments' resistance to deformation. We discuss this term in Section 2.2.2. In the present model, we do not include thermal fluctuations, instead we consider generalized diffusive terms intended to phenomenologically model small-scale effects due to the surrounding fluid [56]. $D_{\mathrm{t},||}$, $D_{\mathrm{t},\perp}$ and D_{r} are the in-plane translational, out-of-plane translational, and rotational diffusion coefficients, respectively. Physically, if $D_{\mathrm{t},||} = D_{\mathrm{t},\perp} = D_{\mathrm{r}} = 0$, Eq. (2.1b) says that the

center-of-mass passively follows the background velocity $\mathbf{u}(\mathbf{x})$ and Eq. (2.1d) says that the rate of change in orientation is given by $\nabla\mathbf{u}(\mathbf{x})\mathbf{p}$. The projection operator $(\mathbf{I}-\mathbf{pp})$ in Eq. (2.1d) ensures that the length of the filament is constant, i.e. $|\mathbf{p}|=1$ or $\mathbf{p}\cdot\dot{\mathbf{p}}=0$.

2.2.2 Fluid equations

Using the characteristic numbers given in Table 3.2 for the fluid velocity and characteristic length yields a small Reynolds number and hence the convective acceleration is negligible. Furthermore, the frequency parameter given by the time scale of the molecular force is small compared to the Reynolds number and inertia is negligible. Thus, the fluid equations are the incompressible Stokes equations with extra stress and an external force:

$$-\mu\Delta_x\mathbf{u}(\mathbf{x},t)+\nabla_x q(\mathbf{x},t)=\nabla_{\mathbf{x}}\cdot\boldsymbol{\sigma}^{\mathrm{p}}(\mathbf{x})+\mathbf{f}_{\mathrm{m}}(\mathbf{x},t),\quad \nabla_x\cdot\mathbf{u}=0. \tag{2.2}$$

Here, $\mathbf{f}_{\mathrm{m}}$ is the force density due to the motors acting at $\mathbf{x}$ on the immersed filaments, q is the pressure, and $\boldsymbol{\sigma}^{\mathrm{p}}$ is the extra stress arising from the microstructure [4]. For passive filaments, $\boldsymbol{\sigma}^{\mathrm{p}}$ can be decomposed into two contributions: $\boldsymbol{\sigma}^{\mathrm{p}}=\boldsymbol{\sigma}^{\mathrm{p}}_B+\boldsymbol{\sigma}^{\mathrm{p}}_F$ [15]. $\boldsymbol{\sigma}^{\mathrm{p}}_B$ arises from Brownian rotations and is modeled as $\boldsymbol{\sigma}^{\mathrm{p}}_B=3kT\mathbf{D}$, where $\mathbf{D}=\int\Psi\left(\mathbf{pp}-\mathbf{I}/3\right)d\mathbf{p}$. This term is small relative to the motor-based force [6] and we neglect it. $\boldsymbol{\sigma}^{\mathrm{p}}_F$ represents the resistance of the filaments to deformation and is modeled as $\boldsymbol{\sigma}^{\mathrm{p}}_F=\sigma_f\mathbf{S}:\mathbf{E}$, where $\mathbf{S}=\int\Psi\left(\mathbf{pppp}-\mathbf{Ipp}/3\right)d\mathbf{p}$, $\mathbf{E}=\left(\nabla_x\mathbf{u}+\nabla_x\mathbf{u}^T\right)/2$ is the rate-of-strain tensor, and $\sigma_f=l^3/(3c)$ is a coefficient depending on the aspect ratio r of the filament ($c=\log(2/r)/(4\pi\mu)>0$). The boundary conditions are doubly periodic in x,y and no-slip at $z=\pm H/2$.

2.2.3 Bound and free motor equations

A motor that walks in a direction $\mathbf{p}$ under a load $\mathbf{F}$ will step with a speed V given by the single motor load-velocity relationship [26]. Here, the load on the motor is the force exerted on the motor head due to the motor's intrinsic motion. Since we do not model that length scale directly, we make the simplifying assumption that the motor is subject to a load of one half of its stall force and walks at one half of its maximum velocity giving the simplified equation $\mathbf{F} = -F_{\mathrm{st}}\mathbf{p}/2$. The motor force, $\mathbf{F}$, gives the force magnitude and direction generated by a motor with tail at $\mathbf{r}_0$ and head on the filament at $\mathbf{r}_1 = \mathbf{y} + s\mathbf{p}$. The motor-based force density at a spatial point $\mathbf{x}$ is then

$$\mathbf{f}_{\mathrm{m}}(\mathbf{x},t) = -\frac{F_{\mathrm{st}}}{2}\iiiint \mathbf{p}\delta(\mathbf{y}+s\mathbf{p}-\mathbf{x})\Psi(\mathbf{y},\mathbf{p},t)\mathcal{M}_{\mathrm{b}}(s,\mathbf{r}_0|\mathbf{y},\mathbf{p})ds\,d\mathbf{r}_0\,d\mathbf{y}\,d\mathbf{p}. \tag{2.3}$$

The convolution with the $\delta-$Dirac function converts from the center-of-mass-based description of Ψ to the spatial description of the force density. Here, $\mathcal{M}_{\mathrm{b}}(s,\mathbf{r}_0|\mathbf{x},\mathbf{p};t)$ is the *conditional* probability density of bound motors with tail $\mathbf{r}_0$ on the plate and head at s on the given filament $\mathbf{x},\mathbf{p}$. The total number of bound motors is given by $N_{\mathrm{b}} = \iiiint \mathcal{M}_{\mathrm{b}}\Psi ds\,d\mathbf{r}_0\,d\mathbf{x}\,d\mathbf{p}$. We also introduce $\mathcal{M}_{\mathrm{f}}(\mathbf{r}_0,t)$, the density function of motors with tail at $\mathbf{r}_0$ and free head. The total number of free motors is $N_{\mathrm{f}} = \int \mathcal{M}_{\mathrm{f}} d\mathbf{r}_0$. In the entire system, the total number of motors $N_{\mathrm{m}} = N_{\mathrm{f}} + N_{\mathrm{b}}$ is conserved.

In general, the free and bound motor populations evolve according to a reaction-diffusion-advection equation. In a gliding assay, motor tails are fixed to a plate and cannot diffuse or advect with the flow. Hence, we consider only the conversion between the free and bound populations, and the advection and procession of the bound motor heads. Further, we assume that if a motor is close enough to the filament, it can bind to the filament with

an attachment rate per filament k_{on} and that a head can detach with a detachment rate k_{off}. We derive the evolution equation for $\mathcal{M}_{\text{b}}$ by considering the conservation law for $\mathcal{M}_{\text{b}}\Psi$, which evolves as

$$\begin{aligned}&\partial_t(\mathcal{M}_{\text{b}}\Psi) + \frac{V_{\max}}{2}\partial_s(\mathcal{M}_{\text{b}}\Psi) + \nabla_x\cdot(\dot{\mathbf{x}}\mathcal{M}_{\text{b}}\Psi) + \nabla_p\cdot(\dot{\mathbf{p}}\mathcal{M}_{\text{b}}\Psi)\\ &\quad= \frac{k_{\text{on}}}{2l\iint\limits_{B_{r_{\text{c}}}}\Psi d\mathbf{p}d\mathbf{x}}\mathcal{M}_{\text{f}}\Psi\mathbb{1}_{B_{r_{\text{c}}}} - k_{\text{off}}\mathcal{M}_{\text{b}}\Psi.\end{aligned} \tag{2.4}$$

Here $\mathbb{1}_{B_{r_{\text{c}}}}$ is the characteristic function on the ball $B_{r_{\text{c}}}$ of (capture) radius r_{c} centered at the attachment point $\mathbf{r}_1$. The three advection terms on the left hand side express the procession of the motor along the filament with speed $V = V_{\max}/2$ and the motion of the motor-filament complex with the background flow. The source term on the right hand side express the attachment of a free motor at s to the filament $\mathbf{x}, \mathbf{p}$, where the factor $2l\iint\limits_{B_{r_{\text{c}}}}\Psi d\mathbf{p}d\mathbf{x}$ measures the approximate available length of filaments in $B_{r_{\text{c}}}$. Since we are not tracking the position of the head $\mathbf{r}_1$, then, for a given filament $\mathbf{x}, \mathbf{p}$, a head attaches at s with uniform probability in $B_{r_{\text{c}}}$. The sink term represents the detachment of a bound motor. Using Eq. (2.1a) to eliminate Ψ, Eq. (3.10) becomes

$$\begin{aligned}&\partial_t\mathcal{M}_{\text{b}} + \frac{V_{\max}}{2}\partial_s\mathcal{M}_{\text{b}} + \dot{\mathbf{x}}\cdot\nabla_x\mathcal{M}_{\text{b}} + \dot{\mathbf{p}}\cdot\nabla_p\mathcal{M}_{\text{b}}\\ &\quad= \frac{k_{\text{on}}}{2l\iint\limits_{B_{r_{\text{c}}}}\Psi d\mathbf{p}d\mathbf{x}}\mathcal{M}_{\text{f}}\mathbb{1}_{B_{r_{\text{c}}}} - k_{\text{off}}\mathcal{M}_{\text{b}}.\end{aligned} \tag{2.5a}$$

We impose an additional condition on detachment. In particular, a head detaches if the motor reaches the end of the filament (i.e. $|s| > l$). We also note that a motor should detach if the motor force exceeds a threshold. Since we assume a constant motor force, we do not encounter this case.

Finally, since the total number of motors N_{m} is conserved, we define the free motor

density $\mathcal{M}_{\rm f}$ implicitly as

$$\mathcal{M}_{\rm f} = \mathcal{M} - \iiint \mathcal{M}_{\rm b}\, ds\, d\mathbf{x}\, d\mathbf{p}, \tag{2.5b}$$

where $\mathcal{M}$ is the total density of motors with tails at $\mathbf{r}_0$.

2.3 Nondimensionalized Equations

In order to simplify the model described in Section 4.2, we start by nondimensionalizing the equations and determining characteristic scales.

2.3.1 Fluid equations

We first nondimensionalize the incompressible Stokes equations (3.13). Let the characteristic length be L in the x and y directions and $H = \varepsilon L$ in the z direction ($\varepsilon \ll 1$), the characteristic velocity be U in the x and y directions and W in the z direction, and the characteristic fluid time scale be $T = L/U$. Since the flow time scale in the z-direction is the same as in the x, y plane, we must have $W = \varepsilon U$. We denote with primes nondimensional quantities. Substituting the corresponding equations, we note that the nondimensional gradient of $\mathbf{u}$ is

$$\nabla_x \mathbf{u} = \frac{U}{L}\begin{pmatrix} \nabla_{x_2'}\mathbf{u}_2' & \frac{1}{\varepsilon}\partial_{z'}\mathbf{u}_2' \\ \varepsilon\nabla_{x_2'}w'^T & \partial_{z'}w' \end{pmatrix} := \frac{1}{T}\nabla_{x'}\mathbf{u}'. \tag{2.6}$$

From equation (2.6), it follows that the extra stress due to the resistance to deformation $\boldsymbol{\sigma}_F^{\rm p}$ has the form $\sigma_f n/T \approx 10^{-4}$ for the dilute microtubule suspensions considered here, so we neglect it. The pressure and force density are nondimesionalized with characteristic pressure Q and force density F. Denoting the components of the force density $\mathbf{f}_{\rm m} = (\mathbf{f}_{\rm m,2}, h) =$

$F(\mathbf{f}'_{\mathrm{m},2}, h')$ and neglecting the extra stress contributions, the Stokes equations (3.13) become

$$\nabla_{x'_2}\mathbf{u}'_2 + \partial_{z'}w' = 0 \tag{2.7a}$$

$$-\Delta_{x'_2}\mathbf{u}'_2 - \frac{1}{\varepsilon^2}\partial_{z'z'}\mathbf{u}'_2 + \frac{LQ}{\mu U}\nabla_{x'_2}q' = \frac{FL^2}{\mu U}\mathbf{f}'_{\mathrm{m},2} \tag{2.7b}$$

$$-\Delta_{x'_2}w' - \frac{1}{\varepsilon^2}\partial_{z'z'}w' + \frac{LQ}{\varepsilon^2\mu U}\partial_{z'}q' = \frac{FL^2}{\varepsilon\mu U}h'. \tag{2.7c}$$

2.3.2 Filament equations

We nondimensionalize the filament equation (2.1) with the same scales as the fluid equations and we nondimensionalize Ψ by the total concentration of filaments $n = N/(\varepsilon L^3)$, setting $n\Psi'(\mathbf{x}', \mathbf{p}, t') = \Psi(L\mathbf{x}'_2, \varepsilon Lz, \mathbf{p}, Tt')$. We find

$$\partial_{t'}\Psi' + \nabla_{x'}\cdot(\mathring{\mathbf{x}}'\Psi') + \nabla_p\cdot(\mathring{\mathbf{p}}\Psi') = 0, \tag{2.8a}$$

$$\mathring{\mathbf{x}}'_2 = \mathbf{u}'_2 - \frac{D_{\mathrm{t},||}T}{L^2}\nabla_{x'_2}\ln\Psi', \quad \mathring{z}' = w' - \frac{D_{\mathrm{t},\perp}T}{\varepsilon^2L^2}\partial_{z'}\ln\Psi' \tag{2.8b}$$

$$\mathring{\mathbf{p}} = (\mathbf{I} - \mathbf{pp})\nabla_{x'}\mathbf{u}'\mathbf{p} - D_{\mathrm{r}}T\nabla_p\ln\Psi', \tag{2.8c}$$

where $\mathring{\cdot}$ indicates the time derivative with respect to t' and $\nabla_{x'}\mathbf{u}'$ refers to the tensor given explicitly in Eq. (2.6).

2.3.3 Bound and free motor equations

Since the motors evolve on a smaller scale than the fluid, we introduce new characteristic scales. We take the filament half-length l as the length scale for motor evolution. We also define a new time scale, $\tau = l/V_{\max}$. We denote with stars the new nondimensional quantities and keep the prime notation for quantities that depend on both scales. We

choose to nondimensionalize the density of bound and free motors to the fraction of bound, $N_{\rm b}/N_{\rm m}$, and free $N_{\rm f}/N_{\rm m}$ motors respectively. In other words, we set

$$\mathcal{M}_{\rm b}(s,\mathbf{r}_0|\mathbf{x},\mathbf{p};t) = \frac{N_{\rm m}}{NL^2 l}\mathcal{M}'_{\rm b}(ls^*, L\mathbf{r}'_0|L\mathbf{x}'_2, \varepsilon L z', \mathbf{p}; \tau t^*),$$
$$\mathcal{M}_{\rm f}(\mathbf{r}_0,t) = \frac{N_{\rm m}}{L^2}\mathcal{M}'_{\rm f}(L\mathbf{r}_0, \tau t^*).$$

Substituting the appropriate definitions into Eqs. (2.5), the nondimensional evolution equation for $\mathcal{M}'_{\rm b}$ and $\mathcal{M}'_{\rm f}$ are

$$\begin{aligned}\partial_{t^*}\mathcal{M}'_{\rm b} + \frac{1}{2}\partial_{s'}(\mathcal{M}'_{\rm b}) + \frac{\tau}{T}\mathring{\mathbf{x}}'\cdot\nabla_{\mathbf{x}'}\mathcal{M}'_{\rm b} + \frac{\tau}{T}\mathring{\mathbf{p}}\cdot\nabla_p\mathcal{M}'_{\rm b} \\ = \frac{k'_{\rm on}}{\iint_{B_{r'_{\rm c}}}\Psi' d\mathbf{x}' d\mathbf{p}}\mathcal{M}'_{\rm f}\mathbb{1}_{B_{r'_{\rm c}}} - k'_{\rm off}\mathcal{M}'_{\rm b}\end{aligned} \tag{2.9a}$$

$$\mathcal{M}'_{\rm f} = \mathcal{M}' - \iiint \mathcal{M}'_{\rm b}\Psi' ds'\, d\mathbf{x}'\, d\mathbf{p}, \tag{2.9b}$$

with $k'_{\rm on} = k_{\rm on}\tau/2$, $k'_{\rm off} = k_{\rm off}\tau$, $B_{r'_{\rm c}}$ is the nondimensional capture ball and $\mathcal{M} = N_{\rm m}/(L^2)\mathcal{M}'$. Finally, nondimensionalizing the force density (3.15) and integrating over y_z, we have

$$\begin{aligned}\mathbf{f}_{\rm m}(\mathbf{x},t) &= F\mathbf{f}'_{\rm m}(L\mathbf{x}'_2, \varepsilon L x_z, Tt') \\ &= -\frac{F_{\rm st}N_{\rm m}}{2\varepsilon L^3}\iiiint \mathbf{p}\delta\left(\mathbf{y}'_2 + \frac{l\sin\phi}{L}s^*\mathbf{p}_2 - \mathbf{x}'_2\right)\Psi'\left(\mathbf{y}'_2, x'_z + \frac{l}{\varepsilon L}\cos\phi, \mathbf{p}, t'\right) \\ &\quad \mathcal{M}_{\rm b}\left(s^*, \mathbf{r}'_0|\mathbf{y}'_2, x'_z + \frac{l}{\varepsilon L}\cos\phi, \mathbf{p}; t^*\right) ds^* d\mathbf{r}'_0 d\mathbf{y}'_2 d\mathbf{p},\end{aligned} \tag{2.10}$$

where (ϕ,θ) are the polar and azimuthal angles, respectively and $\mathbf{p}_2 = (\cos\theta, \sin\theta)$. The above equation yields the characteristic force $F = F_0/\varepsilon$ with $F_0 = F_{\rm st}N_{\rm m}/L^3$.

2.4 Two-dimensional Reduction

In this section, we reduce the spatial dimension of our system of equations by performing an asymptotic analysis of the fluid equations in $\varepsilon \ll 1$. For clarity of exposition,

in the remainder of the paper we drop the primes, replace the open circles with dots and s^* by s.

2.4.1 Fluid equations

We reduce the fluid equations following the standard Hele-Shaw approach [34]. For the solution to the leading order equation in Eq. (2.7b) to be nontrivial, the pressure term and/or the forcing term have to balance the last viscous term. If the forcing term is not one of the dominating terms, then the fluid equations reduce to those of a Newtonian Hele-Shaw cell and are not driven by the motor forces, in contrast to the gliding assay. Therefore, using the characteristic force derived at the end of Section 3.2.3, we have

$$\frac{FL^2}{\mu U} = \frac{F_0 L^2}{\varepsilon \mu U} = O\left(\frac{1}{\varepsilon^2}\right).$$

From the above relationship to be asymptotically valid, we have $U = O(\varepsilon)$ and we set $U = \varepsilon U_0$. From the definition of the characteristic time, we have $T = L/U = T_0/\varepsilon$ with $T_0 = L/U_0$.

We now discuss the order of the pressure term. If the pressure term is of lower order than the force term, then there are two cases to consider: $Q = O(1)$ and $Q = O(\varepsilon^m)$ with $m \geq 1$. In the latter case, the only leading order term in the z-momentum equation (2.7c) is the force term which is therefore unbalanced and this case is impossible. In the first case, the pressure term balances the force term in the z-momentum equation (2.7c). But, since the pressure term drops out of the x, y-momentum equations (2.7b), the continuity equation (2.7a) cannot be satisfied without imposing a condition on the forcing term and

this case is also impossible. Therefore, we have

$$\frac{LQ}{\mu U} = \frac{LQ}{\varepsilon \mu U_0} = O\left(\frac{1}{\varepsilon^2}\right)$$

as well. Consequently, Q scales like $1/\varepsilon$ and we set $Q = Q_0/\varepsilon$. As a result from the previous the discussion, the leading order terms in Eqs. (2.7) are $O(1)$ in Eq. (2.7a), $O(1/\varepsilon^2)$ in Eq. (2.7b), and $O(1/\varepsilon^4)$ in Eq. (2.7c).

Next, we perform the asymptotic analysis and consider the leading order terms in the momentum equations (2.7b). Setting $A_0 = LQ_0/(\mu U_0)$ and $B_0 = F_0 L^2/(\mu U_0)$, we have for the leading $O(1/\varepsilon^2)$ terms

$$-\partial_{zz}\mathbf{u}_2 + A_0 \nabla_2 q = B_0 \mathbf{f}_{\mathrm{m},2}. \tag{2.11}$$

In the z-momentum equation (2.7c), the pressure term is the only leading $O(1/\varepsilon^4)$ term and Eq. (2.7c) reduces to $\partial_z q = 0$, which implies that q is independent of z, in other words $q(\mathbf{x}_2, z) = q(\mathbf{x}_2)$. Integrating Eq. (2.11) twice with respect to z we find

$$-\mathbf{u}_2(\mathbf{x}_2, z) + \boldsymbol{\alpha_0} z + \boldsymbol{\alpha_1} = -\frac{A_0}{2} z^2 \nabla_2 q(\mathbf{x}_2) + B_0 \mathbb{I}(z),$$

where $\mathbb{I}(z) = \int_{-1/2}^{z} \int_{-1/2}^{z'} \mathbf{f}_{\mathrm{m},2}(\mathbf{x}_2, z'') dz'' dz'$. Using the no-slip boundary conditions at $z = \pm 1/2$, we obtain the constants of integration

$$\boldsymbol{\alpha}_0 = B_0 \mathbb{I}(1/2), \text{ and } \boldsymbol{\alpha}_1 = -\frac{A_0}{8} \nabla_2 q(\mathbf{x}_2) + \frac{B_0}{2} \mathbb{I}(1/2).$$

Because $\varepsilon \ll 1$, we now define a gap-averaged velocity $\bar{\mathbf{u}}_2(\mathbf{x}_2) = \int \mathbf{u}_2(\mathbf{x}_2, z) dz$. Averaging the continuity equation (2.7a) over the gap and using the no-slip boundary conditions yields

$$\nabla_2 \cdot \bar{\mathbf{u}}_2 = 0. \tag{2.12a}$$

Further, substituting the values of the integration constants and integrating we obtain the gap-averaged two-dimensional velocity field $\bar{\mathbf{u}}_2(\mathbf{x}_2)$ driven by the x, y-component of the force density $\mathbf{f}_{\mathrm{m},2}$

$$\bar{\mathbf{u}}_2(\mathbf{x}_2) = -\frac{A_0}{12}\nabla_2 q(\mathbf{x}_2) + \frac{B_0}{2}\mathbb{I}\left(\frac{1}{2}\right) - B_0\int \mathbb{I}(z)dz. \tag{2.12b}$$

Equations (2.12) are the gap-averaged two-dimensional forced Stokes equations. These equations are very similar to those obtained for non-Newtonian Hele-Shaw flows [34]. We also remark that the integral operator and its mean can be expressed as moments of the force density by reversing the order of integration

$$\mathbb{I}(1/2) = \int\left(\frac{1}{2} - z\right)\mathbf{f}_{\mathrm{m},2}(\mathbf{x}_2, z)dz, \text{ and } \int \mathbb{I}(z)dz = \frac{1}{2}\int\left(\frac{1}{2} - z\right)^2 \mathbf{f}_{\mathrm{m},2}(\mathbf{x}_2, z)dz.$$

2.4.2 Filament equations

To reduce the evolution equations for the distribution functions Ψ we again rely on leading order analysis and depth-averaging. Since the diffusion terms represent phenomenologically observed small-scale motions, the diffusion constants are chosen so that these terms are of comparable magnitude with the other terms in Eq. (2.8b)-(2.8c). Integrating Eq. (2.8a) over the gap yields

$$\partial_t\left(\int \Psi dz\right) + \nabla_{x,2}\cdot\left(\int \dot{\mathbf{x}}_2\Psi dz\right) + (\dot{z}\Psi)\Big|_{-1/2}^{1/2} + \nabla_p\cdot\left(\int \dot{\mathbf{p}}\Psi dz\right) = 0. \tag{2.13}$$

In order to express the previous equation in terms of the gap-averaged distribution of filaments $\bar{\Psi}$, we will make use of the general closure approximation $\int f(z)\Psi dz = \bar{f}\bar{\Psi}$, for any function $f(z)$. For a velocity field $\mathbf{u}$ that is approximately parabolic in z, the accuracy of this approximation will depend on the maximum value of ψ_{zz} in the gap. This

approximation may degrade if all the filaments are bound, as all the filaments will then be within a $\sim$50nm band relative to the $\sim 100\mu$m gap. However, this is not the case in general. With this closure approximation and the no-slip boundary conditions, Eq. (2.13) becomes

$$\partial_t \bar{\Psi} + \nabla_{x,2} \cdot \left(\dot{\bar{\mathbf{x}}}_2 \bar{\Psi}\right) + \nabla_p \cdot \left(\int \dot{\mathbf{p}} \Psi dz \right) = 0 \quad \text{with} \quad \dot{\bar{\mathbf{x}}}_2 = \bar{\mathbf{u}}_2 - \frac{D^0_{\mathrm{t},||} T_0}{L^2} \nabla_{x,2} \ln \bar{\Psi}, \tag{2.14}$$

where $D_{\mathrm{t},||} = D^0_{\mathrm{t},||}\varepsilon$. Next, we reduce the rotational flux to two dimensions. To do so, we denote the unit vectors in spherical coordinates by $\hat{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\theta}}$. Then, expressing $\dot{\mathbf{p}}$ in spherical coordinates as $\dot{\mathbf{p}} = \dot{\phi}\hat{\boldsymbol{\phi}} + \sin\phi\dot{\theta}\hat{\boldsymbol{\theta}}$, taking the dot product of Eq. (2.8c) with $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\phi}}$, and defining $D_{\mathrm{r}} = D^0_{\mathrm{r}}\varepsilon$, we arrive at

$$\dot{\theta}\sin\phi = \hat{\boldsymbol{\theta}}^T \nabla_x \mathbf{u}\mathbf{p} - \frac{D^0_{\mathrm{r}} T_0}{\sin\phi} \partial_\theta \ln \Psi \quad \text{and} \quad \dot{\phi} = \hat{\boldsymbol{\phi}}^T \nabla_x \mathbf{u}\mathbf{p} - D^0_{\mathrm{r}} T_0 \partial_\phi \ln \Psi. \tag{2.15}$$

Recalling the form of $\nabla_x \mathbf{u}$ in Eq. (2.6), we let $\mathbf{p} = (\sin\phi\mathbf{p}_2, \cos\phi)$, $\hat{\boldsymbol{\theta}} = (\mathbf{p}_2^{\perp}, 0)$, and $\hat{\boldsymbol{\phi}} = (\cos\phi\mathbf{p}_2, -\sin\phi)$ with $\mathbf{p}_2^{\perp} = (-\sin\theta, \cos\theta)$. The leading order term in Eq. (2.15) is of order $1/\varepsilon$ and carrying out the matrix multiplications, we obtain

$$\cos\phi \left(\mathbf{p}_2^{\perp T} \partial_z \mathbf{u}_2\right) = 0, \text{ and } \cos^2\phi \left(\mathbf{p}_2^T \partial_z \mathbf{u}_2\right) = 0.$$

The only non-trivial solution to the previous equations is $\cos\phi = 0$, that is $\phi = \frac{\pi}{2}$. In other words, to leading order the filaments lie in a horizontal plane, which has been observed experimentally [61]. In this case, $\nabla_p \cdot (\int \dot{\mathbf{p}}\Psi) = \partial_\theta(\int \dot{\theta}\Psi)$ and $\dot{\theta} = \mathbf{p}_2^{\perp T} \nabla_{x,2} \mathbf{u}_2 \mathbf{p}_2 - D^0_{\mathrm{r}} T_0 \partial_\theta \ln \Psi$, which allows us to reduce the last term in Eq. (2.14) to a single θ derivative. Combining everything and applying the closure approximation on the velocity gradient tensor, we arrive at the two-dimensional gap-averaged filament evolution equation

$$\partial_t \bar{\Psi} + \nabla_{x,2} \cdot (\dot{\bar{\mathbf{x}}}_2 \bar{\Psi}) + \partial_\theta(\dot{\bar{\theta}}\bar{\Psi}) = 0 \tag{2.16a}$$

$$\dot{\bar{\mathbf{x}}}_2 = \bar{\mathbf{u}}_2 - \frac{D^0_{\mathrm{t},||} T_0}{L^2} \nabla_2 \ln \bar{\Psi}, \quad \dot{\bar{\theta}} = \mathbf{p}_2^{\perp T} \nabla_{x,2} \bar{\mathbf{u}}_2 \mathbf{p}_2 - D^0_{\mathrm{r}} T_0 \partial_\theta \ln \bar{\Psi}. \tag{2.16b}$$

2.4.3 Bound and free motor equations

Since $T = T_0/\varepsilon$, the last two advective terms on the left of Eq. (2.9a) scale like ε and they can be neglected. Further, because no other term in Eq. (2.9a) depends on ε, the $O(1)$ leading order equation for $\mathcal{M}_\mathrm{b}$ is

$$\partial_{t^*}\mathcal{M}_\mathrm{b} + \frac{1}{2}\partial_s\mathcal{M}_\mathrm{b} = \frac{k_\mathrm{on}}{\iint_{B_{r_\mathrm{c}}}\Psi d\mathbf{x}d\theta}\mathcal{M}_\mathrm{f}\mathbb{1}_{B_{r_\mathrm{c}}} - k_\mathrm{off}\mathcal{M}_\mathrm{b}, \tag{2.17}$$

Integrating over z and using the closure approximation, we obtain the gap-averaged equation for the density of bound and free motors

$$\partial_{t^*}\bar{\mathcal{M}}_\mathrm{b} + \frac{1}{2}\partial_s(\bar{\mathcal{M}}_\mathrm{b}) = \frac{k_\mathrm{on}}{\iint_B \bar{\Psi} d\mathbf{x}_2 d\theta}\mathcal{M}_\mathrm{f}\mathbb{1}_B - k_\mathrm{off}\bar{\mathcal{M}}_\mathrm{b}, \tag{2.18a}$$

$$\mathcal{M}_\mathrm{f} = \mathcal{M} - \iiint \bar{\mathcal{M}}_\mathrm{b}\bar{\Psi} ds\, d\mathbf{x}_2\, d\theta. \tag{2.18b}$$

Here B is the two-dimensional projection of B_{r_c}.

Next, we reduce the motor force density (2.10). Integrating over z, using the closure approximation, and $\phi = \pi/2$, we obtain the gap-averaged force density

$$\bar{\mathbf{f}}_m(\mathbf{x}_2, t) = -\frac{1}{2}\iiiint \mathbf{p}_2\delta\left(\mathbf{y}_2 + \frac{l}{L}s\mathbf{p}_2 - \mathbf{x}_2\right)\bar{\Psi}(\mathbf{y}_2, \theta, t)\bar{\mathcal{M}}_b(s, \mathbf{r}_0|\mathbf{y}_2, \theta) ds\, d\mathbf{r}_0\, d\mathbf{y}_2\, d\theta. \tag{2.19}$$

The closure approximation and Eq. (2.19) allow us to solve for the constants of integration in Eq. (2.12b): $\mathbb{I}(1/2) = \bar{\mathbf{f}}_{\mathrm{m},2}/2$ and $\int \mathbb{I}(z)dz = \bar{\mathbf{f}}_{m,2}/6$. Plugging these expressions into Eq. (2.12b) yields the final fluid gap-averaged fluid equations

$$\nabla_{x,2}\cdot\bar{\mathbf{u}}_2(\mathbf{x}_2) = 0 \tag{2.20a}$$

$$\bar{\mathbf{u}}_2(\mathbf{x}_2) = -\frac{A_0}{12}\nabla_2 q(\mathbf{x}_2) + \frac{B_0}{12}\bar{\mathbf{f}}_m(\mathbf{x}_2, t). \tag{2.20b}$$

This concludes the model reduction to a system of equations depending on the two-dimensional configuration variables $\mathbf{r}_0, \mathbf{x}_2$ and one-dimensional configuration variables s, θ. Table 3.1 summarizes the two-dimensional model equations. For simplicity, we drop the subscript 2 and bars in the notation and let $D_{\mathrm{t},||} = D^0_{\mathrm{t},||} T^0/L^2$ and $D_{\mathrm{r}} = D^0_{\mathrm{r}} T$.

2.5 Numerical method

In this section, we discuss the discretization of the nondimensionalized equations summarized in Table 3.1 and the development of a stable algorithm. First, we note that $\mathcal{M}_{\mathrm{b}}$ is high dimensional, with six variables in 2D. However, since a head detaches if the elongation of the motor's stalk exceeds a certain threshold, $\mathcal{M}_{\mathrm{b}}$ is sparse in the $\mathbf{x} - \mathbf{r}_0$ hyperplane, hence we store a small $\mathbf{x}_2$-grid, whose size depends on (l/L), around each $\mathbf{r}_0$.

At $t = 0$, we initialize the filament distribution Ψ, the motor distribution $\mathcal{M}$, and the bound motor distribution $\mathcal{M}_{\mathrm{b}}$. The free motor distribution can then be computed with (M2) and the fluid equations (U1)-(U2) solved (see below). The algorithm steps and substeps to advance from time t^n to time t^{n+1} are described next. Since $\tau < T$, we denote by k the time index on which $\mathcal{M}_{\mathrm{b}}$ is solved.

1. Compute Ψ^{n+1} by solving (F1) together with (F2). We use second-order Crank-Nicolson for the diffusion terms, and Adams Bashforth 2 with an adaptive timestep for the time discretization of the advection terms. The advection terms are discretized in space using a high-resolution total variation diminishing upwind scheme with superbee flux limiter [13].

2. Compute $\mathcal{M}_{\mathrm{b}}^{k+1}$ by evaluating (M2) and solving (M1). The integrals in (M2) are evaluated using the midpoint rule using data at time n. We again use upwinding with superbee flux limiter for the advection terms, and Adams Bashforth 2 with an adaptive timestep for the time discretization. We repeat this process until $\mathcal{M}_{\mathrm{b}}^{n+1}$ is obtained.

3. Compute the force density $\mathbf{f}_{\mathrm{m}}^{n+1}$ from (U2) using $\mathcal{M}_{\mathrm{b}}^{n+1}$ and Ψ^{n+1}. The integrals are evaluated using the midpoint rule and an approximate Dirac delta function, for which we use a hat function spanning two grid cells in each direction. This approximation is C_0 and satisfies a first moment condition, resulting in conservation of angular momentum. For a discussion of the implications of different choices of numerical delta functions, see [53].

4. Solve (U1) spectrally for $\mathbf{u}^{n+1}$. Because of the periodic boundary conditions, we transform (U1) in Fourier space ($\mathbf{k}$ wave vector):

$$\hat{\mathbf{u}}_{\mathbf{k}} = -\frac{A_0}{12} i\mathbf{k}\hat{p} + \frac{B_0}{12}\hat{\mathbf{f}}_{\mathrm{m}_{\mathbf{k}}} \quad \mathbf{k}\cdot\hat{\mathbf{u}}_{\mathbf{k}} = 0.$$

Taking the dot product with $\mathbf{k}$, the pressure can be eliminated yielding

$$\begin{cases} \hat{\mathbf{u}}_{\mathbf{k}} = \frac{B_0}{12}\left(\mathbf{I} - \frac{\mathbf{k}\mathbf{k}^T}{k^2}\right)\hat{\mathbf{f}}_{\mathrm{m}_{\mathbf{k}}} \text{ if } k = \|\mathbf{k}\| \neq 0 \\ \hat{\mathbf{u}}_{\mathbf{0}} = \frac{B_0}{12}\hat{\mathbf{f}}_{m_{\mathbf{0}}} \text{ if } k = 0 \end{cases} . \tag{2.21}$$

We remark that the constants A_0 and Q_0 drop out and Eq. (2.21) depends on B_0 only. At this stage, we also compute the rate-of-strain tensor spectrally.

The two most time-intensive portions of the algorithm are the bound motor density evolution and the motor force calculation. For feasibility, we implemented these using

Nvidia's CUDA C language and ran them on Nvidia Tesla GPU accelerators obtaining speedup factors of nearly 50 for all parameters tested. Since there is no interaction between motors in neighboring $\mathbf{r}_o$ cells, we update each $\mathbf{r}_0$ cell independently.

2.6 Results

We illustrate the behavior of the system with two examples, using the parameters $L = 125\mu\text{m}$, $l = 1\mu\text{m}$, $\varepsilon = 5 \cdot 10^{-4}$, $N = 2 \cdot 10^4$, $N_\text{m} = 3 \cdot 10^6$, $k_\text{on} = 1\text{s}^{-1}$, $k_\text{off} = 0.5\text{s}^{-1}$. Here we demonstrate that the proposed model captures qualitative features of microtubule gliding assays.

First, we consider a clump of locally aligned filaments in the center of the domain with an initially uniform bound motor density. The initial alignment drives the fluid motion, and consequently the filaments, in that direction. The clump advects and diffuses through the domain as illustrated in Fig. 2.2 (see supplementary movie S1). This motion is qualitatively similar to the moving clusters observed in [61]. The resulting disturbance in the velocity field propagates out from the leading edge of the clump in a wave-like motion. Eventually, diffusion causes the filament density to relax to the uniform isotropic density and the velocity field to decay to zero. We set the dimensionless diffusion parameters to $D_{\text{t},||} = 2.5 \cdot 10^{-3}$ and $D_\text{r} = 2.5 \cdot 10^{-5}$.

Second, we consider the experimentally motivated example of a gliding assay of non-localized filaments and motors (Fig. 2.3 and supplementary movies S2-S6). If the initial density of filaments is uniform, our model does not generate disturbance flows different than

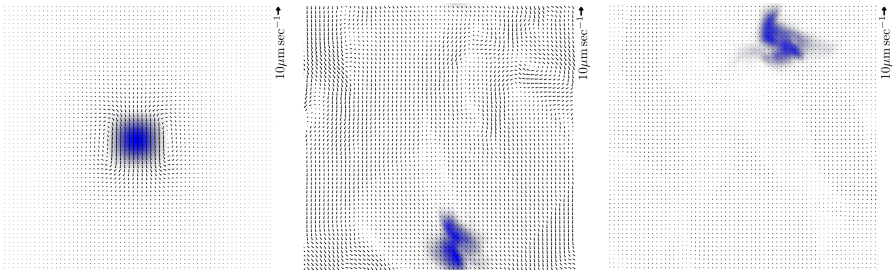


Figure 2.2: Evolution of the velocity field and filament density for a clump of locally aligned filaments with uniform bound motor density, depicted at $t = 0$s, $t = 50$s, and $t = 150$s. The parameters are $L = 125\mu$m, $l = 1\mu$m, $\varepsilon = 5 \cdot 10^{-4}$, $N = 2 \cdot 10^4$, $N_{\mathrm{m}} = 3 \cdot 10^6$, $k_{\mathrm{on}} = 0.5\mathrm{s}^{-1}$, $k_{\mathrm{off}} = 0.5\mathrm{s}^{-1}$. The dimensionless diffusion parameters to $D_{\mathrm{t},||} = 2.5 \cdot 10^{-3}$ and $D_{\mathrm{r}} = 2.5 \cdot 10^{-5}$.

the imposed mean flow. Therefore, the filament density is perturbed away from uniformity in both space and orientation with

$$\frac{1}{a} \sum_{i,j=1}^{8} \epsilon_{ij} \cos(\pi i x + \xi_{ij}) \cos(\pi j y + \xi_{ij}) P_{ij}(\theta), \tag{2.22}$$

where ϵ_{ij} is a uniform random number in $[-.01, .01]$, ξ_{ij} is a uniform random number in $[0, 2\pi]$, a is a normalization constant and $P_{ij}(\theta)$ are third order polynomials in $\cos(\theta)$ and $\sin(\theta)$ with random coefficients in $[-1, 1]$ as in [56].

The typical state of the system is illustrated in Fig. 2.3 at times $t = 0, 10, 20$ min. The first row depicts the evolution of the velocity and vorticity fields (supplementary movie S2), and in the second row, tracer particles highlight fluid mixing (supplementary movie S3). After the transient velocity field disappears, the flow organizes into distinct subregions separated by boundaries of non-zero vorticity. These subregions span tens of microns, compared with the filament length of 1 micron, and can persist for several minutes. The third row shows the spatial filament density defined as

$$\Psi_{\mathrm{spatial}}(\mathbf{x}, t) = \int \Psi(\mathbf{x}, \theta, t) d\theta \tag{2.23}$$

(supplementary movie S4). As t increases, the filaments form varying patterns, concentrating into bands at flow subregion boundaries and clumps in flow subregion interiors. These filament concentration patterns largely mantain their structure as they traverse the domain following the flow field. Some regions collide resulting in the formation of larger fronts or bands, while others break apart. The concentrated regions contain more than twice the number of filaments as the sparse regions. Such migrating cohesive structures have been observed in experiment [61, 65]. The fourth row depicts the filament orientation field (supplementary movie S5). We compute the orientation matrix

$$\mathbf{N}(\mathbf{x},t) = \frac{\int \mathbf{p}\mathbf{p}^T \Psi(\mathbf{x},\theta,t) d\theta}{\int \Psi(\mathbf{x},\theta,t) d\theta} \tag{2.24}$$

and draw its eigenvectors scaled by their associated eigenvalues. We remark that Ψ maintains antipodal symmetry, i.e $\Psi(\mathbf{x},\theta,t) = \Psi(\mathbf{x},\theta+\pi,t)$, because Eqs. (2.16a)-(2.16b) are invariant under this transformation. The two-dimensional nematic order parameter is

$$S(\mathbf{x},t) = \frac{\int \left(2(\mathbf{p}\cdot\mathbf{n})^2 - 1\right) \Psi(\mathbf{x},\theta,t) d\theta}{\int \Psi(\mathbf{x},\theta,t) d\theta}, \tag{2.25}$$

where $\mathbf{n}$ is the eigenvector associated with the largest eigenvalue of $\mathbf{N}$. We observe increased local nematic order at the boundaries of the flow subregions, where filaments tend to align tangentially to the boundaries, while the interiors of the flow subregions contain areas of both high and low nematic order. Finally, the fifth row depicts the bound motor concentration with tail at $\mathbf{r}_0$

$$\mathcal{M}_{\mathrm{b}}^{\mathrm{tail}}(\mathbf{r}_0,t) = \iiint \mathcal{M}_{\mathrm{b}}(s,\mathbf{r}_0,t|\mathbf{x},\theta)\Psi(\mathbf{x},\theta,t) ds d\mathbf{x} d\theta \tag{2.26}$$

(supplementary movie S6). Our simulation shows a strong positive correlation between high bound motor density and high filament concentration. We note that motors are not directly

advected by the flow, as their tails are fixed to the substrate. Instead, the bound motor concentration increases as more filaments become available to bind to.

2.7 Conclusions

Starting from conservation equations, we developed a multiscale model describing the complex interplay between molecular motors, filaments, and fluid in a gliding assay. Our asymptotic and dimensional analysis agrees with experimental results [61] and shows that, for a small channel height, the filaments evolve in a two-dimensional plane parallel to the bottom plate upon which molecular motors are anchored. The coupling between the motors, filaments, and surrounding fluid is achieved by directly spreading the motor force onto the fluid and passively advecting the filaments with the local fluid velocity. As a result, the depth-averaged fluid equations ressemble the non-Newtonian Hele Shaw equations. In the present model, we only consider a dilute suspension of filaments and therefore neglect effects due to steric interactions between filaments. The only interactions considered are therefore purely hydrodynamic interactions between the fluid and the filaments. Nevertheless, our simulations show behavior that is qualitatively similar to that seen in [61]. We observe swirls, moving high-density fronts, and cluster movements.

An in-depth study of the parameter set should reveal the transition between different states as described in [36, 32, 59]. While our model and numerical experiments demonstrate the importance of hydrodynamic interactions, they do not address the relative importance of steric and hydrodynamic interactions for pattern formation in gliding assays. To be able to ascertain the role of both effects as well as observe additional patterns, the

density of filaments must be increased. As a result, the continuum model should be extended to include both steric interactions between the filaments and $\boldsymbol{\sigma}_F^{\mathrm{p}}$, the extra stress due to the filament resistance to deformation [15, 11].

Parameter	Symbol	Value or Range	Source
Motor attachment rate	k_{on}	$.005 - .05\text{s}^{-1}$ (sp)	[58]
		$.04 - 50\text{s}^{-1}$ (sp)	[46]
		50s^{-1}	[66]
Motor detachment rate	k_{off}	$.083 - .167\text{s}^{-1}$	[27]
		$.005 - .05\text{s}^{-1}$ (sp)	[58]
		$.04 - 50\text{s}^{-1}$ (sp)	[46]
		$.5\text{s}^{-1}$	[66]
Motor maximum speed	$V_{\max}$	$500 - 750\text{nm/s}$	[27]
		$1\mu\text{m/s}$ (sp)	[46, 32, 58]
Motor stall force	F_{stall}	$0.5 - 2\text{pN}$	[46]
		5pN	[32]
Motor length (fully stretched)	L_m	50nm	[32]
Motor capture radius	r_{c}	$10 - 22\text{nm}$	[32]
Chamber length	L	$5 - 18\text{mm}$	[27]
		3-15mm	[45]
Chamber height	H	$110\mu\text{m}$	[27]
		$100\mu\text{m}$	[45]
Filament length	l	$.7 - 7\mu\text{m}$	[27]
		$50\mu\text{m}$	[30]
		$1\mu\text{m}$	[32]
Fluid viscosity	μ	$0.5\text{pNs}/\mu\text{m}^2$	[32]
Fluid velocity	U	$10\mu\text{m}/s$	[65]

Table 2.1: List of relevant physical parameters and their values as reported in the literature. (sp) denotes simulation parameters used in the referenced source.

Model equations

Filaments	$\partial_t \Psi + \nabla_x \cdot (\dot{\mathbf{x}}\Psi) + \partial_\theta(\dot{\theta}\Psi) = 0$	(F1)
	$\dot{\mathbf{x}} = \mathbf{u} - D_{\mathrm{t},\|\|}\nabla_x \ln\Psi, \quad \dot{\theta} = \mathbf{p}_\perp^T \nabla_x \mathbf{u}\mathbf{p} - D_\mathrm{r}\partial_\theta \ln\Psi$	(F2)
Motors	$\partial_{t^*}\mathcal{M}_\mathrm{b} + \frac{1}{2}\partial_s \mathcal{M}_\mathrm{b} = \frac{k_\mathrm{on}}{\iint_B \Psi d\mathbf{x}_2 d\theta}\mathcal{M}_\mathrm{f}\mathbb{1}_B - k_\mathrm{off}\mathcal{M}_\mathrm{b}$	(M1)
	$\mathcal{M}_\mathrm{f} = \mathcal{M} - \iiint \mathcal{M}_\mathrm{b}\Psi ds\, d\mathbf{y}\, d\theta$	(M2)
Fluid	$\mathbf{u} = -\frac{A_0}{12}\nabla_2 q + \frac{B_0}{12}\mathbf{f}_m, \quad \nabla\cdot\mathbf{u} = \mathbf{0}$	(U1)
	$\mathbf{f}_\mathrm{m} = -\frac{1}{2}\iiint \mathbf{p}\ \delta(\mathbf{y} + \frac{l}{L}s\mathbf{p} - \mathbf{x})\Psi\mathcal{M}_\mathrm{b} ds\, d\mathbf{r}_0\, d\mathbf{y}\, d\theta$	(U2)

Table 2.2: Summary of the two-dimensional, gap-averaged, nondimensional model equations for the evolution of the filament distribution $\Psi(\mathbf{x},\theta,t)$, bound motor distribution $\mathcal{M}_\mathrm{b}(\mathbf{r}_0, s|\mathbf{x},\theta;t)$, free motor distribution $\mathcal{M}_\mathrm{f}(\mathbf{r}_0)$ and fluid $\mathbf{u}(\mathbf{x})$ in a motility assay.

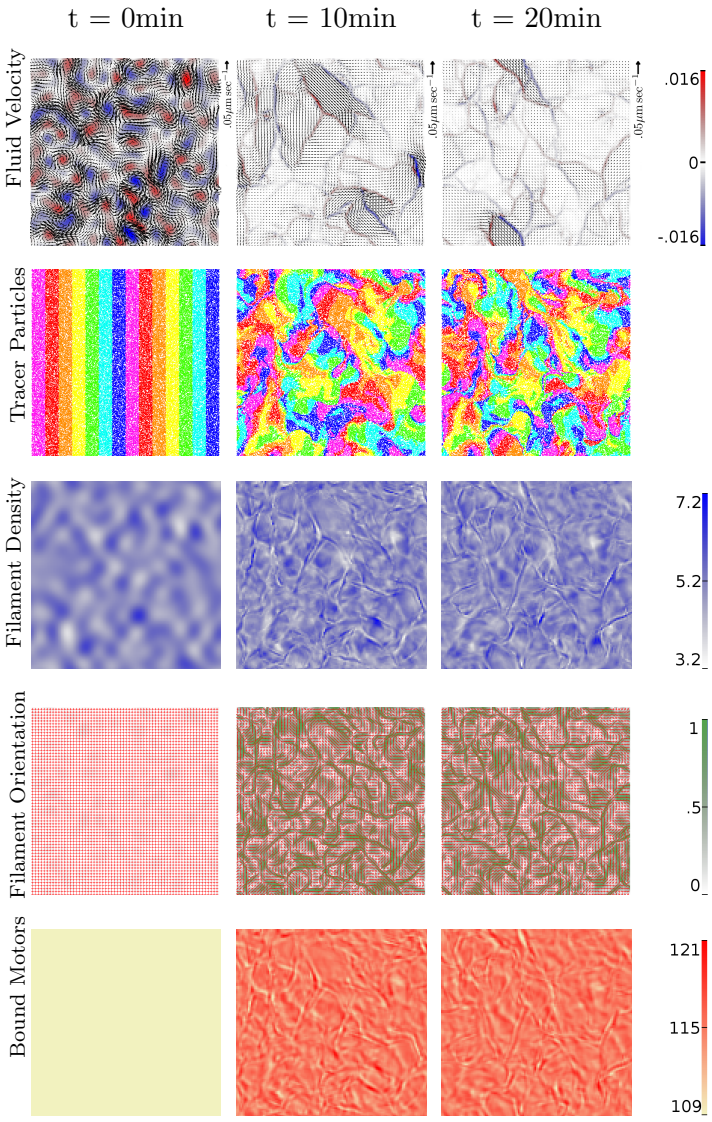


Figure 2.3: Emergence of ordered subregions in the filament-motor-fluid system, depicted at $t = 0$ (first column), $t = 10\,\text{min}$ (second column), and $t = 20\,\text{min}$ (third column). First row: $\mathbf{u}(\mathbf{x}, t)$, with vorticity plotted in color (units of s^{-1}). As the simulation proceeds, the fluid subregions form vorticity bands at their boundaries with zero vorticity in their interiors. Second row: tracer particles show fluid mixing. Third row: $\Psi_{\text{spatial}}(\mathbf{x}, t)$ in μm^{-2} defined in Eq. (2.23). Fourth row: eigenvectors of $\mathbf{N}(\mathbf{x}, t)$ (red lines) and $S(\mathbf{x}, t)$ (green field) given in Eqs. (3.52)-(2.25). Fifth row: $\mathcal{M}_{\text{b}}^{\text{tail}}(\mathbf{r}_0, t)$ in μm^{-2} defined in Eq. (3.51). At $t = 0$, the density of filaments is perturbed in space and orientation according to Eq. (3.53), while the bound motor density is uniform. The parameters are $L = 125\mu\text{m}$, $l = 1\mu\text{m}$, $\mu = 0.12\text{pNs}/\mu\text{m}^2$, $\varepsilon = 5 \cdot 10^{-4}$, $N = 2 \cdot 10^4$, $N_{\text{m}} = 3 \cdot 10^6$, $k_{on} = 1\text{s}^{-1}$, $k_{\text{off}} = 0.5\text{s}^{-1}$,$D_{\text{t},||} = 0\mu\text{m}^2\text{s}^{-1}$, $D_{\text{r}} = 0\text{s}^{-1}$.

Chapter 3

A Micro-Macro Framework for Analyzing Steric and Hydrodynamic Interactions in Gliding Assays

Macroscopic flows of filament-motor mixtures, driven by the hydrolysis of ATP, are important to many cellular processes such as cytoplasmic streaming in *Drosophila* oocytes and cortical flow in the first cell division of *C. elegans*. Gliding assays, reduced *in vitro* model systems where motor proteins adsorbed onto a planar substrate bind to and move filaments, recreate large-scale dynamic patterns like coherent swarming motion and density waves. These systems are sensitive to the microscopic behavior such as the motor protein binding and unbinding dynamics, which take place on a faster timescale than the direct and

fluid-mediated filament interactions. In this work, we present a multiscale modeling and simulation framework for gliding assays that allows detailed microscopic motor modeling as well as both steric and hydrodynamic interactions between filaments. Our model is based on continuum kinetic theory, and our implementation utilizes CPU and GPU parallelism to track the sparse but high-dimensional state space arising from the microscopic motor protein configurations. We find that steric interactions play a role in the formation of spatiotemporally coherent flow structures, and qualitatively reproduce experimentally observed behaviors including filament crossover and alignment, and clump formation, merging, and splitting.

3.1 Introduction

Actin and tubulin filaments working in concert with motor proteins play a central role in cell functions including mitosis and pronuclear centering [64]. Gliding assays, in which stabilized filaments are propelled by anchored motor proteins powered by the hydrolysis of ATP in a thin quasi-two-dimensional chamber, are commonly used to study the behavior of these cellular components *in vitro* (Figure 3.1). Large-scale pattern formation is observed in such experiments, including clump formation, merging, and splitting, and density waves [61], and the emergence of a lattice of microtubule vortices [65].

The physics of the filament-motor-fluid system are inherently multiscale in space and time, with nanoscale motors with fast binding/unbinding kinetics coupled to microscale filaments interacting in a macroscopic fluid domain. These systems have been studied with a variety of theoretical and computational approaches. Models that track explicit repre-

sentations of filaments with Langevin dynamics underscore the sensitivity of the system to the motor behavior; [36, 32] included a load-dependent force-velocity relationship of motor proteins in a gliding assay and found that the motor activity increases nematic ordering, and [46, 50, 48] found that the time spent by individual motors at the end of a microtubule before falling off plays a central role in the emergence of coherent structures. These models neglect fluid-mediated filament-filament interactions, and are therefore not suitable to address such behaviors. Moreover, the computations have scaled to only hundreds of filaments, while many systems of interest such as a mitotic spindle are estimated to have tens or hundreds of thousands of filaments. A model based on macroscopic configuration fields, and thus more suitable to a large system and large length and time scales, was proposed by [39]. This was expanded upon by [60] in a hydrodynamic theory incorporating explicit tracking of bound and unbound motor populations. Both [60, 39] rely on phenomenologically motivated constitutive equations in the model derivation and neglect filament density fluctuations. Another class of modeling approaches starts with a microscopic model and coarse-grains the system to attain a macroscopic description [42, 67, 1]. For example, [42] assume a constant motor density and demonstrate that inhomogeneities in motor stepping rate are necessary to drive bundle formations, and [67], without considering fluctuations in motors or filament densities, show that the order of the isotropic-nematic transition depends on the force-dependent motor detachment. However, both of [67, 1] neglect fluid-mediated filament-filament interactions, although they could be coupled to the fluid equation using a configurational average of an expression involving the distribution function to include the contribution of the particles to the fluid stress [11, 5]. This approach has been widely

applied to nanorods [69], and more recently to active gels [41] and to suspensions of active swimmers [57, 56], and was used in our previous work [24]. Such methods have the benefit of flexibly allowing detailed microscopic modeling. However, it is not always possible to avoid tracking the microscopic variables, which can incur significant computational cost.

As shown experimentally by, among others, [61, 32, 65], the filament density and steric interactions play a critical role in the formation of coherent structures. In this paper we consider dense suspensions of filaments, and build upon our previous model [24] to include steric interactions between filaments. A widespread model of steric interaction is the excluded volume potential [11]. This model, widely used in liquid crystal theory [69, 42], has been adapted to active suspensions of self-propelled pushers and pullers in three dimensions [15]. The latter includes a rotational steric alignment term, but neglects the linear steric contribution which can prevent unphysical "piling up" amongst the microtubules in a gliding assay. We follow this approach to modeling steric interactions, including both rotational and linear steric interaction terms.

Characterizing the interplay of steric versus hydrodynamic effects requires an exploration of different filament densities and motor systems. At a microtubule density of $.05\mu m^{-2}$, Sumino et al. [65] are able to model their experimentally observed vortex lattice with a phenomenological agent-based method neglecting hydrodynamics. In contrast, at actomyosin surface densities ranging between $2\mu m^{-2}$ to $21\mu m^{-2}$, Schaller et al. [61] demonstrate filament clump and density wave persistence and scale that cannot be explained through purely steric interactions. Additionally, Schaller et al. [62] demonstrate evidence of hydrodynamic effects in the formation of depletion layers between clump-clump

or clump-wall collisions that cause reorientation before physical contact. The particulars of the microscopic motors may also significantly influence the collective motion. For example, [65] reports that using kinesin motor proteins instead of dynein results in a higher rate of microtubule crossover events, limiting steric interactions and preventing the formation of the vortex lattice. The importance of fluid effects can also be seen further in the theoretical study of filaments in a quasi-two-dimensional chamber [21].

In this work, we present a novel micro-macro model and computational framework to simulate both steric and hydrodynamic interactions in a microtubule gliding assay. Our new framework supports different motor protein activity models, as well as the option to treat the filaments as self-propelled. Rather than use phenomenological steric interaction rules, we model fluid stresses due to microtubule inextensibility, rotational and translational steric interactions, and self-propulsion if applicable.

We base our approach on our previous work [24], where we developed a continuum model coupling the motion of the fluid to the motion of the motors and microtubules. In that work, we used closure approximations to reduce the fluid equations to depth-averaged two-dimensional equations, and restricted ourselves to the dilute limit, ignoring steric interactions. Here, we solve the fluid equations in three dimensions and avoid making closure approximations. As in [24], we track distributions of microtubules and kinesin motor proteins, with behavior governed by conservation equations.

The paper is organized as follows. Our framework is presented in Section 3.2, the implementation and numerical methods are presented in Section 3.3, simulation results are presented in Section 4.4, and we conclude in Section 4.5.

3.2 Modeling framework

This section reviews the setup of a gliding motility assay, describes our continuous representation, and details the individual components of our modeling framework. Two distributions are tracked: one for the microtubules, and one for the bound motors. Each distribution satisfies a conservation equation. We present one such equation for the microtubule distribution taking into account hydrodynamic and steric effects, and two for the bound motor distribution. Bulk fluid forces and steric stresses are calculated from the bound motor proteins and the microtubule distribution respectively, and included in the fluid equations.

3.2.1 Microscale model

Figure 3.1 illustrates the experimental setup. A microtubule gliding assay consists of two plates separated by a small distance. A fluid fills the gap, with fluid flow characterized by low Reynolds number. Motor proteins (kinesin in our case) are anchored to the bottom plate with their heads free to bind to microtubules, walk along them, and detach. The microtubules, in turn, glide along the motor protein heads, effectively constrained within a single plane. They are stabilized to prevent growth or depolymerization, and have an orientation defined in terms of a plus and minus end. Upon binding, kinesin motor proteins walk towards the plus end of the microtubule, propelling the microtubule in the direction of its minus end. We assume ATP saturation so the motor proteins are continuously active. We refer the reader to [24] for a list of values of physical parameters found in the literature.

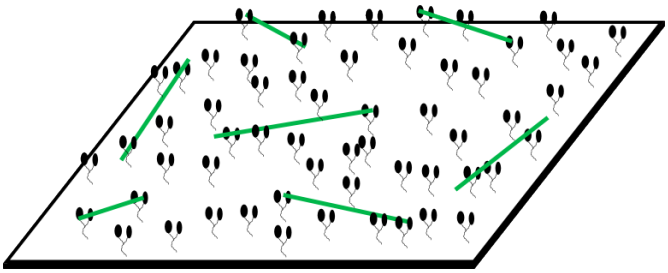


Figure 3.1: Microtubule gliding assay setup. Motor protein (black) tails are anchored to a fixed plate, while their heads bind and pull microtubules (green).

Microtubule distribution

In this model, we assume that the microtubules of length $2l$ and diameter b are inextensible and rigid, an appropriate approximation for microtubules of $l \leq 2\mu m$ [63]. We describe the microtubules by the position of their center-of-mass $\mathbf{x}$ and a vector $\mathbf{p}$ pointed towards their plus end. Let $\Psi(\mathbf{x}, \mathbf{p}, t)$ be the microtubule distribution function. Ψ evolves according to the Smoluchowski equation (see also [25, 24])

$$\partial_t \Psi + \nabla_x \cdot (\dot{\mathbf{x}}\Psi) + \nabla_p \cdot (\dot{\mathbf{p}}\Psi) = 0 \tag{3.1}$$

$$\dot{\mathbf{x}}_2 = -V_{\mathrm{sp}}\mathbf{p}_2 + \mathbf{u}_2(\mathbf{x}) - \nabla_{x_2} U_t - D_{\mathrm{t},||}\nabla_{x_2} \ln \Psi, \tag{3.2}$$

$$\dot{z} = w(\mathbf{x}) - \partial_z U_t - D_{\mathrm{t},\perp}\partial_z \ln \Psi \tag{3.3}$$

$$\dot{\mathbf{p}} = (\mathbf{I} - \mathbf{pp})\nabla_x \mathbf{u}(\mathbf{x})\mathbf{p} - \nabla_p U_r - D_{\mathrm{r}}\nabla_p \ln \Psi. \tag{3.4}$$

Here the subscript 2 denotes the in-plane x, y components and derivatives with respect to these variables. The first two and last terms in the equations for $\dot{\mathbf{x}} = (\dot{\mathbf{x}}_2\, \dot{z})^T$ and $\dot{\mathbf{p}}$ are similar to those of the active bacteria swimming model of [56]. V_{sp} is a propulsion speed, analogous to the self-propulsion term in active swimmer models. In our model, motor forces spread to the fluid grid (described below) can only be resolved on the scale of the fluid grid.

We allow a phenomenological self-propulsion velocity to compensate for this reduction in resolution of microtubule motion. Since microtubules do not propel themselves through the fluid as a bacteria does [62], but are propelled by motors distributed along the filament, the resulting force should act like a monopole (see Eqs. (3.15)-(3.16)), similar to sedimenting particles [12]. Therefore, we include both a self-propulsion velocity and a passive force in our model. Furthermore, $\mathbf{u}(\mathbf{x}) = (\mathbf{u}_2(\mathbf{x})\, w(\mathbf{x}))^T$ is the velocity of the surrounding fluid at $\mathbf{x}$ with which the microtubules are advected. Because the kinesin motors walk towards the plus end, i.e. in the direction of $\mathbf{p}$, the microtubule will move in the direction $-\mathbf{p}_2$ in the plane of motion where it is restricted to leading order as shown in [24]. $D_{\mathrm{t},||}$, $D_{\mathrm{t},\perp}$ and D_{r} are the translational and rotational diffusion coefficients, respectively. Because of the channel geometry and the experimental observation that microtubules move in a z-plane, we expect $D_{\mathrm{t},\perp}$, the diffusion in the z direction, to be smaller than the in-plane diffusion $D_{\mathrm{t},||}$. We use zero diffusion in all of our examples. We do not include thermal fluctuations in the present model. The third term in $\dot{\mathbf{x}}_2$ and $\dot{\mathbf{p}}$ and the second term in $\dot{z}$ describes the effect of steric interactions through a translational and rotational potential, $U_{t/r}$, respectively. We model the steric potential using the Maier-Saupe potential $K_{t/r}(\mathbf{p}, \mathbf{p}') = -U^0_{t/r}(\mathbf{p} \cdot \mathbf{p}')^2$ with

$$U_{t/r}(\mathbf{x}, \mathbf{p}, t) = \int \Psi(\mathbf{x}, \mathbf{p}', t) K_{t/r}(\mathbf{p}, \mathbf{p}') d\mathbf{p}'.$$

The above form of U_r is identical to the one proposed by [15] for active suspension, but we also keep the translational steric potential U_t from [42, 2] to prevent interpenetration in the plane of the microtubules. Neglecting this term was less consequential for [15] as their model is three-dimensional; our two-dimensional rod reduction below (Section 3.2.2) leads to a more highly constrained geometry where this term is important. With this term

included, we see behavior closer to the experiments of Schaller et al. [62] where clumps grow in extent rather than increasingly concentrating into a small area. We refer the reader to [2] for details on how the translational term arises in the Smoluchowski equation for the microtubule density. With the previous definitions of the steric potential the translational and rotational fluxes become

$$\dot{\mathbf{x}}_2 = -V_{\mathrm{sp}}\mathbf{p}_2 + \mathbf{u}_2(\mathbf{x}) + U^0_{t,||}\mathbf{pp} : \nabla_2\mathbf{D}(\mathbf{x},t) - D_{\mathrm{t},||}\nabla_2 \ln\Psi \tag{3.5}$$

$$\dot{z} = w(\mathbf{x}) + U^0_{t,\perp}\mathbf{pp} : \partial_z\mathbf{D}(\mathbf{x},t) - D_{\mathrm{t},\perp}\partial_z \ln\Psi \tag{3.6}$$

$$\dot{\mathbf{p}} = (\mathbf{I}-\mathbf{pp})(\nabla_x\mathbf{u}(\mathbf{x}) + 2U^0_r\mathbf{D}(\mathbf{x},t))\mathbf{p} - D_{\mathrm{r}}\nabla_p \ln\Psi, \tag{3.7}$$

where $\mathbf{D}(\mathbf{x},t) = \int \Psi(\mathbf{x},\mathbf{p},t)\mathbf{pp}d\mathbf{p}$ is the second moment of Ψ with respect to $\mathbf{p}$. The total number of microtubules is given by $N = \iint \Psi d\mathbf{x}d\mathbf{p}$.

Motor distributions

In general, the free and bound motor populations evolve according to a reaction-diffusion-advection equation. In a gliding assay, motor tails are fixed to a plate and cannot diffuse or advect with the flow. Hence, we consider only the conversion between the free and bound populations, and the advection and procession of the bound motor heads. We represent the free motor density as $\mathcal{M}_f(\mathbf{r}_0)$, the density of motors with free heads and tails anchored at position $\mathbf{r}_0$. We do not track the position of free motor heads. We represent the bound motor density per microtubule as $\mathcal{M}_b(\mathbf{r}_0, s|(\mathbf{x},\mathbf{p}),t)$, with $\mathbf{r}_0$ the position where the motor tail is anchored, $\mathbf{x}, s, \mathbf{p}$ the center of mass, arclength parameter, and orientation of the microtubule the motor head is bound to, and t the time. The notation $|(\mathbf{x},\mathbf{p})$ denotes that the probability is conditional on the distribution of microtubules $\Psi(\mathbf{x},\mathbf{p},t)$. Finally,

we let $\mathcal{M}(\mathbf{r}_0)$ be the total (bound + free) motor density at $\mathbf{r}_0$. We define the free motor density $\mathcal{M}_f$ as

$$\mathcal{M}_f(\mathbf{r}_0, t) = \mathcal{M}(\mathbf{r}_0) - \iiint \mathcal{M}_b(\mathbf{r}_0, s|(\mathbf{x}, \mathbf{p}), t)\Psi(\mathbf{x}, \mathbf{p}, t)ds\, d\mathbf{x}\, d\mathbf{p}, \tag{3.8}$$

that is, for every motor with tail anchored at $\mathbf{r}_0$ the head is either free or bound. The total number of bound motors is given by $N_b = \iiiint \mathcal{M}_b \Psi d\mathbf{r}_0 d\mathbf{x} d\mathbf{p} ds$. In the entire system, the total number of motors $N_m = N_f + N_b$ is constant.

The possible configurations of bound motor heads face constraints. First, a head detaches if it walks off the plus end of the microtubule ($|s| > l$). Second, the motor head detaches due to stretching of the motor stalk, which happens beyond a critical distance r_c. We do not model the spring force of the motor stalk extension. We enforce these two constraints by requiring that $\mathcal{M}_b$ is zero when $|s| > l$ and by taking $\mathcal{M}_b$ to be zero when $|\mathbf{x} + s\mathbf{p} - \mathbf{r}_0| \geq r_c$.

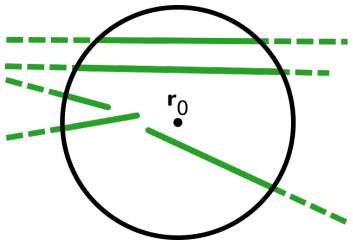


Figure 3.2: Set $B_{r_c}(\mathbf{r}_0)$ of microtubule segments accessible to motor head for motor tail anchored at $\mathbf{r}_0$ in solid green (illustrated in 2D for clarity).

Equivalently, the second condition says that for a given tail $\mathbf{r}_0$ there is only a small subset of $\mathbf{x} + s\mathbf{p}$ available for attachment. This crucial locality restriction effectively reduces the dimensionality of $\mathcal{M}_b$. We represent the allowable local configurations as a ball of radius r_c illustrated in Figure 3.2,

$$B_{r_c}(\mathbf{r}_0) = \{(\mathbf{x}', s', \mathbf{p}') : |\mathbf{x}' + s'\mathbf{p}' - \mathbf{r_0}| < r_c\}. \tag{3.9}$$

We consider two equations for the bound motor distribution. The first, hereafter referred to as the "evolved" model, tracks $\mathcal{M}_b$ through the evolution of a full conserva-

tion equation. It models motor head stepping along microtubules as well as attachment proportional to the available number and length of microtubules and detachment. Since motors bind and unbind quickly relative to the speed of the microtubules, this conservation equation has its own smaller timescale. The second model, hereafter referred to as the "simplified" model, assumes that motors bind to any reachable position s along a microtubule with equal probability, and that the distribution of bound motors $\mathcal{M}_b\Psi$ with tails anchored at $\mathbf{r}_0$ is proportional to the density of microtubules to bind to up until all available motors are bound. The fidelity of the smaller timescale behavior from the evolved motor model is lost. Instead of solving a conservation equation at a separate timescale, $\mathcal{M}_b$ is updated from Ψ on its timescale.

In the evolved motor model, following [49], the number of binding events per second is proportional to the local density of free motors times the available length of microtubules (as an approximation for the available binding sites) with the constant of proportionality, denoted by $\overline{k_{\text{on}}}$, that has units of $\mu m^2 s^{-1}$. Defining $\bar{B}(\mathbf{r}_0) = \iiint_{B_{r_c}} \Psi \mathcal{M}_b ds d\mathbf{x} d\mathbf{p}$, then $\bar{B}(\mathbf{r}_0)$ represents the number of bound motors in B_{r_c} per unit area. Since $\bar{B}$ and $\mathcal{M}_f$ have units of number per unit area (and not per unit volume), we divide $\overline{k_{\text{on}}}$ by the capture radius r_c to obtain a constant of proportionality that has units of $\mu m\ s^{-1}$, before repeating the argument of [49] for $\bar{B}$. In order to convert from a number per area to a number, we multiply $\bar{B}(\mathbf{r}_0)$by the area of the disk D_{r_c} of radius r_c centered at $\mathbf{r}_0$. Therefore, after dividing through by $|D_{r_c}|$ and neglecting advection terms, we have a relationship of the form

$$\partial_t \bar{B} = -k_{\text{off}} \bar{B} + \frac{\overline{k_{\text{on}}}}{r_c |D_{r_c}|} \mathcal{M}_f \iiint_{B_{r_c}} \Psi d\mathbf{x} d\mathbf{p} ds,$$

where the integral of Ψ over B_{r_c} represents the total length of available microtubules and k_{off} (units of s^{-1}) is the detachment rate. Setting $k_{\text{on}} = \overline{k_{\text{on}}}/(r_c|D_{r_c}|)$ with units of $(\mu m s)^{-1}$ and including advection terms, the conservation equation for $\mathcal{M}_b\Psi$ is

$$\begin{aligned}\partial_t(\mathcal{M}_b\Psi) + \partial_s(V_{\text{m}}\mathcal{M}_b\Psi) + \nabla_x\cdot(\dot{\mathbf{x}}\mathcal{M}_b\Psi) + \nabla_p\cdot(\dot{\mathbf{p}}\mathcal{M}_b\Psi)\\ = k_{\text{on}}\Psi\mathcal{M}_{\text{f}}\mathbb{1}_{B_{r_c}(\mathbf{r}_0)} - k_{\text{off}}\mathcal{M}_b\Psi.\end{aligned} \tag{3.10}$$

Here $\mathbb{1}_{B_{r_c}}(\mathbf{r}_0)$ is the indicator function for $B_{r_c}(\mathbf{r}_0)$. The flux terms on the left hand side express the procession of the motor along the microtubule with speed V_{m} and the motion of the motor-microtubule complex with the background flow. The source terms on the right hand side express the attachment of a free motor at s to the microtubule $\mathbf{x}, \mathbf{p}$ at a rate per length k_{on}, and the detachment of a bound motor at a rate k_{off}. Using Eq. (3.1) to eliminate Ψ, Eq. (3.10) simplifies to

$$\left[\partial_t\mathcal{M}_b + \partial_s(V_{\text{m}}\mathcal{M}_b) + \dot{\mathbf{x}}\cdot\nabla_x\mathcal{M}_b + \dot{\mathbf{p}}\cdot\nabla_p\mathcal{M}_b\right]\Psi = k_{\text{on}}\mathcal{M}_f\Psi\mathbb{1}_{B_{r_c}} - k_{\text{off}}\mathcal{M}_b\Psi. \tag{3.11}$$

We note that if $\Psi \neq 0$ for all $(\mathbf{x}, \mathbf{p}, t)$ we can divide by Ψ, but we will refrain from doing so until Section 3.2.2.

In this paper, we also consider a simplified heuristic motor model where all motor heads are located in the same plane at height z_0 and the number of bound motors is proportional to the number of microtubules available within the binding range. Therefore, we let $\mathcal{M}_b$ be the piecewise function

$$\mathcal{M}_b(\mathbf{r}_0, s|(\mathbf{x},\mathbf{p}), t) = \begin{cases} 0 & \text{if } |\mathbf{x}+s\mathbf{p}-\mathbf{r}_0| \geq r_c \\ \min\left(C, \frac{\mathcal{M}(\mathbf{r}_0)}{\iiint_{B_{r_c}}\Psi d\mathbf{x}d\mathbf{p}ds}\right) & \text{if } |\mathbf{x}+s\mathbf{p}-\mathbf{r}_0| < r_c \end{cases}. \tag{3.12}$$

In the above, the cutoff constant C has the same units as $\mathcal{M}_b$. The second term in the minimum effectively caps $\mathcal{M}_b$ so that $\iiint_{B_{r_c}}\mathcal{M}_b\Psi d\mathbf{x}dsd\mathbf{p} \leq \mathcal{M}$, the total available motors

at $\mathbf{r}_0$. Above the threshold value C, all local motor heads are bound. As C increases, so do the number of bound motors at $\mathbf{r}_0$ for a fixed value of $\iiint_{B_{r_c}} \Psi d\mathbf{x} ds d\mathbf{p}$, so larger C values decrease the minimum rod density needed to bind all local motors.

Fluid

The bulk fluid motion is described by the incompressible Stokes equations for low Reynolds number flows with suspended microtubule and motor microstructure. As is customary [4], the total stress in the fluid can be divided into a Newtonian stress and an extra stress arising from the microstructure, leading to

$$-\mu \nabla_{\mathbf{x}}^2 \mathbf{u}(\mathbf{x}) + \nabla_{\mathbf{x}} q(\mathbf{x}) = \nabla_{\mathbf{x}} \cdot \boldsymbol{\sigma}^{\mathrm{p}}(\mathbf{x}) + \mathbf{f}_{\mathrm{m}}(\mathbf{x}), \quad \nabla_{\mathbf{x}} \cdot \mathbf{u}(\mathbf{x}) = 0. \tag{3.13}$$

In the above, q is the pressure, μ is the dynamic viscosity, $\boldsymbol{\sigma}^{\mathrm{p}}$ is the extra stress, and $\mathbf{f}_{\mathrm{m}}$ is the force density due to the motors acting at $\mathbf{x}$ on the immersed microtubules. We take the fluid domain to be doubly periodic in x, y with no-slip conditions at the plate $z = -H/2$ and at the cover slip $z = H/2$.

We define the extra stress as $\boldsymbol{\sigma}^{\mathrm{p}} = \boldsymbol{\sigma}^{\mathrm{f}} + \boldsymbol{\sigma}^{\mathrm{t}}$ similar to [15], where $\boldsymbol{\sigma}^{\mathrm{f}}$ arises from microtubule inextensibility and $\boldsymbol{\sigma}^{\mathrm{t}}$ arises from steric interaction. These extra stresses are

$$\boldsymbol{\sigma}^{\mathrm{f}} = \sigma_f \mathbf{S} : \mathbf{E}, \quad \boldsymbol{\sigma}^{\mathrm{t}} = -\sigma_t [\mathbf{D} \cdot \mathbf{D} - \mathbf{S} : \mathbf{D}], \tag{3.14}$$

where $\mathbf{E}(\mathbf{x}, t) = \frac{1}{2}\left(\nabla \mathbf{u} + \nabla \mathbf{u}^T\right)$ is the rate-of-strain tensor and $\mathbf{S}$ is the fourth order moment of Ψ, $\mathbf{S}(\mathbf{x}, t) = \int \Psi \mathbf{pppp} d\mathbf{p}$. The coefficients are $\sigma_f = \pi \mu 4 l^3 / 3 \ln(2r)$ and $\sigma_t = \pi \mu 8 l^3 U_r^0 / 3 \ln(2r)$, which can be derived using slender body theory, with r the microtubule aspect ratio and l the microtubule half-length [15, 56]. While the derivation in [15, 56] is for rods in free space, and our model has walls in the z dimension, our two-dimensional

reduction below (Section 3.2.2) restricts these stress terms to the x, y plane, for which we use periodic boundary conditions. Therefore we do not derive these coefficients for the case of a nearby boundary. We remark that steric interaction in space does not lead to extra stress terms in the slender body framework as the resulting force is constant along the microtubule.

In this model, we apply the force spreading approach of the immersed boundary method (see [53]) to our polymeric fluid to obtain the motor force [5]

$$\mathbf{f}_{\mathrm{m}}(\mathbf{x},t) = \iiiint \mathbf{F}(\mathbf{y},s,\mathbf{p},\mathbf{r}_0)\delta(\mathbf{y}+s\mathbf{p}-\mathbf{x})\Psi(\mathbf{y},\mathbf{p},t)\mathcal{M}_b(\mathbf{r}_0,s|(\mathbf{y},\mathbf{p}),t)ds\,d\mathbf{r}_0\,d\mathbf{y}\,d\mathbf{p}, \quad (3.15)$$

where $\mathbf{F}$ is the force associated with a single motor. Further, we note that the convolution with the $\delta-$Dirac function converts from the center-of-mass based description of Ψ to the spatial description of the force density. The force generated by all bound motor heads acting at $\mathbf{y}+s\mathbf{p}$ is spread to $\mathbf{x}$ by integrating over all possible motor configurations with head at $\mathbf{y}+s\mathbf{p}$. In general motor stepping speed is load-dependent [26]. However, here we assume that the motor is stepping at a constant speed V_{m}, where its max stepping speed is $V_{\max}$, and thus exerts a constant force of magnitude $F_{\mathrm{st}}(1-\frac{V_{\mathrm{m}}}{V_{\max}})$ in $-\mathbf{p}$, with F_{st} the motor stall force. This gives the simplified expression for the motor force

$$\mathbf{F}(\mathbf{y},s,\mathbf{p},\mathbf{r}_0) = \mathbf{F}(\mathbf{p}) = -F_{\mathrm{st}}\left(1-\frac{V_{\mathrm{m}}}{V_{\max}}\right)\mathbf{p}. \quad (3.16)$$

3.2.2 Two-dimensional reduction

In [24] we showed that to leading order the microtubule orientations are restricted to the x, y-plane. Moreover, experiments have shown that the microtubules and bound motor

heads are concentrated about a two-dimensional plane of motion [54]. Taking advantage of this fact obviates the need to track Ψ and $\mathcal{M}_b$ in the z-dimension, which provides critical memory and computation savings when storing and solving for the two distributions. We also restrict the forcing term in the fluid equations to be localized to that plane.

Microtubule distribution

The microtubules are centered around a plane $z = z_0$, where z_0 is about the length of the motor protein, above the bottom plate

$$\Psi(\mathbf{x}, \mathbf{p}, t) = \Psi_{z_0}(\mathbf{x}_2, \mathbf{p}, t)\delta_a(z - z_0). \tag{3.17}$$

Here δ_a is a smooth delta function, chosen to be

$$\delta_a(z - z_0) = \begin{cases} \frac{1}{2a}(1 + \cos(\frac{\pi(z-z_0)}{a})) & |z - z_0| \leq a \\ 0 & |z - z_0| > a \end{cases}. \tag{3.18}$$

We remark that to prevent the presence of microtubules or motors at the top or bottom plates of the assay, we further require that $z_0 - a > \frac{-H}{2}$ and $z_0 + a < \frac{H}{2}$. We also introduce the notation $A_i = \int \delta_a^i(z - z_0)dz$ to denote the moments of δ_a. By construction, we have $A_1 = 1$, $A_2 = 3/(4a)$ and $A_3 = 5/(6a)$. Next, the microtubules are constrained to the plane given the geometry of the assay, so $\mathbf{p} = (\cos\theta, \sin\theta, 0)^T$. Defining $\mathbf{p_2} = (\cos\theta, \sin\theta)^T$, we have the decomposition $\Psi_{z_0}(\mathbf{x}_2, \mathbf{p}, t) = \Psi_{z_0}(\mathbf{x}_2, \mathbf{p}_2, t)$.

Using Eq. (3.17), the microtubule reduction proceeds by integrating Eq. (3.1) with respect to z. We use a bar to denote the integral over z of a quantity weighted by the smooth delta function, for example $\bar{\mathbf{u}}(\mathbf{x}_2) = \int \mathbf{u}(\mathbf{x}_2, z)\delta_a(z)dz$. Plugging Eq. (3.17) into Eqs. (3.1) and (3.5)-(3.7), integrating over z, and using the facts that $\nabla_p = \mathbf{p}_2^{\perp}\partial_\theta$ and that both w

and δ_a vanish at the top and bottom plate, we obtain

$$\partial_t \Psi_{z_0} + \nabla_2 \cdot (\dot{\bar{\mathbf{x}}}_2 \Psi_{z_0}) + \partial_\theta (\dot{\bar{\theta}} \Psi_{z_0}) = 0, \tag{3.19}$$

where we have defined the quantities $\dot{\bar{\mathbf{x}}}_2$ and $\dot{\bar{\theta}}$ as

$$\dot{\bar{\mathbf{x}}}_2 = -V_{\mathrm{sp}} \mathbf{p}_2 + \bar{\mathbf{u}}_2 + A_2 U^0_{t,||} \mathbf{p}_2 \mathbf{p}_2 : \nabla_2 \mathbf{D}_{2,z_0} - D_{\mathrm{t},||} \nabla_2 \ln \Psi_{z_0} \tag{3.20}$$

$$\dot{\bar{\theta}} = (\nabla_2 \bar{\mathbf{u}}_2 + 2U^0_r A_2 \mathbf{D}_{2,z_0}) : \mathbf{p}_2^{\perp} \mathbf{p}_2 - D_{\mathrm{r}} \partial_\theta \ln \Psi_{z_0}. \tag{3.21}$$

Bound motor distribution

We make the same assumptions for the bound motor distribution, since the bound motor heads must be in plane with the microtubules they are bound to and write $\mathcal{M}_b$ analogously to (3.17) as

$$\mathcal{M}_b(\mathbf{r}_0, s|(\mathbf{x}, \mathbf{p}), t) = \mathcal{M}_{b,z_0}(s, \mathbf{r}_0|\mathbf{x}_2, \theta, t)\delta_a(z - z_0). \tag{3.22}$$

To derive a reduced equation for the evolved bound motor distribution, we plug in the assumptions (3.22) and (3.17) into equation (3.11), integrate with respect to z and divide by Ψ_{z_0}. Noting that the set $B_{r_c}(\mathbf{r}_0)$ can be approximated as

$$B_{r_c}(\mathbf{r}_0) \approx \left\{ (\mathbf{x}, s, \mathbf{p}) : (\mathbf{x}_2, s, \mathbf{p}_2) \in D_{r_c}(\mathbf{r}_0) \text{ and } -\frac{H}{2} \leq z \leq -\frac{H}{2} + d_{r_c}(\mathbf{x}_2, s, \theta) \right\},$$

where $D_{r_c}(\mathbf{r}_0) = \{(\mathbf{x}_2, s, \theta) : |\mathbf{x}_2 + s\mathbf{p}_2 - \mathbf{r_0}| < r_c\}$ is the disk of capture radius r_c and $d_{r_c}(\mathbf{x}_2, s, \theta) = \sqrt{r_c^2 - |\mathbf{x}_2 + s\mathbf{p}_2 - \mathbf{r}_0|^2}$, we find

$$\begin{aligned} &\partial_t(\mathcal{M}_{b,z_0}) + \partial_s(V_{\mathrm{m}} \mathcal{M}_{b,z_0}) + \dot{\bar{\mathbf{x}}}_2 \cdot \nabla_2 \mathcal{M}_{b,z_0} - \dot{\bar{\zeta}} \mathcal{M}_{b,z_0} + \dot{\bar{\theta}} \partial_\theta \mathcal{M}_{b,z_0} \\ &= -k_{\mathrm{off}} \mathcal{M}_{b,z_0} + \frac{k_{\mathrm{on}}}{A_2} \mathcal{M}_f B_2 \mathbb{1}_{D_{r_c}}. \end{aligned} \tag{3.23}$$

In (3.23), we defined the tilde quantities similarly to the bar quantities in (3.20)-(3.21), but with respect to δ_a^2 as opposed to δ_a. We have

$$\dot{\tilde{x}}_2 = -V_{\text{sp}}\mathbf{p}_2 + \frac{1}{A_2}\tilde{\mathbf{u}}_2 + \frac{A_3}{A_2}U_{t,\|}^0\mathbf{p}_2\mathbf{p}_2 : \nabla_2\mathbf{D}_{2,z_0} - D_{t,\|}\nabla_2 \ln \Psi_{z_0} \tag{3.24}$$

$$\dot{\tilde{\zeta}} = \frac{1}{2A_2}\widetilde{\partial_z w} - \frac{B_1}{A_2}U_{t,\perp}^0\mathbf{p}_2\mathbf{p}_2 : \mathbf{D}_{2,z_0} \tag{3.25}$$

$$\dot{\tilde{\theta}} = \left(\frac{1}{A_2}\nabla_2\tilde{\mathbf{u}}_2 + 2U_r^0\frac{A_3}{A_2}\mathbf{D}_{2,z_0}\right) : \mathbf{p}_2^{\perp}\mathbf{p}_2 - D_r\partial_\theta \ln \Psi_{z_0}. \tag{3.26}$$

The constants B_1 and B_2 are

$$B_1 = \frac{1}{2}\int \delta_a^2(z-z_0)\partial_{zz}\delta_a(z-z_0)dz = -\frac{\pi^2}{4a^4} \quad B_2 = \int_{-H/2}^{-\frac{H}{2}+d_{r_c}(\mathbf{x}_2,s,\theta)} \delta_a(z-z_0)dz.$$

While the quantity B_2 is a function of $\mathbf{x}_2, s, \theta$, plugging Eq. (3.18) for δ_a into B_2, integrating and using a Taylor series expansion of sine, yield $0 \leq B_2 \leq r_c/a$. Therefore, for the remainder of this paper, we let $B_2 = r_c/a$.

Following the same steps for the simplified motor model, we have

$$\mathcal{M}_{b,z_0}(\mathbf{r}_0, s|(\mathbf{x}_2,\theta),t) = \begin{cases} 0 & \text{if } (\mathbf{x}_2,s,\theta) \notin D_{r_c} \\ H\min\left(C, \frac{\mathcal{M}(\mathbf{r}_0)}{\iiint_{D_{r_c}}\Psi_{z_0}d\mathbf{x}_2 d\theta ds}\right) & \text{if } (\mathbf{x}_2,s,\theta) \in D_{r_c} \end{cases}. \tag{3.27}$$

Fluid

While we do not average the fluid equations over z, some of the stress and force components are zero as a result of the two-dimensional reduction of $\mathcal{M}_b$ and Ψ. As the stresses are defined in terms of moments of Ψ with respect to $\mathbf{p}$, the implications of Eq. (3.17) for the stress tensors in (3.14) are

$$\boldsymbol{\sigma}^{\text{f}} = \sigma_f\mathbf{S}_{z_0}(\mathbf{x}_2,t) : \mathbf{E}(\mathbf{x},t)\delta_a(z-z_0) \tag{3.28}$$

$$\boldsymbol{\sigma}^{\text{t}} = -\sigma_t\left(\mathbf{D}_{z_0}(\mathbf{x}_2,t)\cdot\mathbf{D}_{z_0}(\mathbf{x}_2,t) - \mathbf{S}_{z_0}(\mathbf{x}_2,t) : \mathbf{D}_{z_0}(\mathbf{x}_2,t)\right)\delta_a^2(z-z_0), \tag{3.29}$$

where we defined

$$\mathbf{D}_{z_0}(\mathbf{x}_2,t)=\int \Psi_{z_0}(\mathbf{x}_2,\theta,t)\mathbf{pp}d\theta \quad \text{and} \quad \mathbf{S}_{z_0}(\mathbf{x}_2,t)=\int \Psi_{z_0}(\mathbf{x}_2,\theta,t)\mathbf{pppp}d\theta.$$

We note that the third row and column of $\boldsymbol{\sigma}^{\mathrm{f}},\boldsymbol{\sigma}^{\mathrm{t}}$ are identically zero because the z-component of $\mathbf{p}$ is zero. Therefore, we use the subscript 2 to denote the upper 2x2 block of each tensor, such as $\mathbf{D}_{2,z_0}$. Since the motor force $\mathbf{F}(\mathbf{p})$ in Eq. (3.16) is in the direction $-\mathbf{p}$, the z-component of $\mathbf{f}_{\mathrm{m}}$ is zero. As a result, the fluid equations (3.13) take the form

$$-\mu\nabla_2^2\mathbf{u_2}(\mathbf{x})-\mu\partial_{zz}\mathbf{u_2}(\mathbf{x})+\nabla_2 q(\mathbf{x})=\nabla_2\cdot\boldsymbol{\sigma}_2^{\mathrm{p}}(\mathbf{x})+\mathbf{f}_2(\mathbf{x}) \tag{3.30}$$

$$-\mu\nabla_2^2 w(\mathbf{x})-\mu\partial_{zz}w(\mathbf{x})+\partial_z q(\mathbf{x})=0 \tag{3.31}$$

$$\nabla_2\cdot\mathbf{u_2}(\mathbf{x})+\partial_z w(\mathbf{x})=0 \tag{3.32}$$

with

$$\mathbf{f}_2(\mathbf{x})=-F_{\mathrm{st}}\left(1-\frac{V_{\mathrm{m}}}{V_{\max}}\right)\delta_a(z-z_0)^2\iiiint \mathbf{p}_2\delta(\mathbf{y}_2+s\mathbf{p}_2-\mathbf{x}_2)\Psi_{z_0}M_{b,z_0}dsd\mathbf{r}_0d\mathbf{y}_2d\theta.$$

3.2.3 Nondimensionalization

In this section, we nondimensionalize the set of reduced equations introduced in Section 3.2.2. To do so, we first introduce the characteristic scales. Let the characteristic length be L in the x,y dimensions and $H=\varepsilon L$ in the z dimension ($\varepsilon\ll 1$), let the characteristic velocity be U in the x,y dimensions and W in the z dimension, and let the characteristic fluid timescale be $T=L/U$. We set

$$\mathbf{x}_2=L\mathbf{x}',\quad z=\varepsilon Lz'=Hz',\quad \mathbf{u}_2=U\mathbf{u}_2',\quad w=Ww',\quad t=Tt',$$

where $'$ denotes dimensionless quantities. By a similarity argument, we have that $W=\varepsilon U$.

Since the motors evolve on a smaller scale than the fluid, we introduce new characteristic scales. We take the microtubule half-length l as the length scale, the motor speed V_{m} as the velocity scale for motor evolution, and obtain a new timescale, $\tau = l/V_{\mathrm{m}}$. We set

$$s = ls', \quad V = V_{\mathrm{m}}V', \quad t = \tau t^*.$$

Microtubule distribution

To nondimensionalize Ψ we recall that it integrates to N, the number of microtubules. Furthermore, since the smooth delta function satisfies $\int \delta_a(z-z_0)dz = 1$, we define Ψ'_{z_0} as

$$\Psi_{z_0}(\mathbf{x}_2, \theta, t) = \frac{N}{L^2}\Psi'_{z_0}(\mathbf{x}'_2, \theta, t') \tag{3.33}$$

so that $\iint \Psi'_{z_0} d\mathbf{x}'_2 d\theta = 1$. Alternatively, if we introduce the nondimensional smooth delta function $\delta'_{a'}(z' - z'_0)$ as $\delta_a(z - z_0) = \frac{1}{H}\delta'_{a'}(z' - z'_0)$ with $a' = a/H$ to mimic the behavior of a Dirac delta function, then we have for the full nondimensional density of microtubules $\Psi'(\mathbf{x}', \mathbf{p}, t') = \Psi'_{z_0}(\mathbf{x}'_2, \theta, t')\delta'_{a'}(z' - z'_0)$ and $\iiint \Psi' d\mathbf{x}' d\mathbf{p} = 1$. The microtubule evolution equation (3.19) is nondimensionalized on the same scale as the fluid equations. Therefore, plugging the definition of the rescaled quantities into Eqs. (3.19), (3.20) and (3.21) yields

$$\partial_{t'}\Psi'_{z_0} + \nabla'_2 \cdot (\dot{\bar{\mathbf{x}}}'_2 \Psi'_{z_0}) + \partial_\theta(\dot{\bar{\theta}}'\Psi'_{z_0}) = 0 \tag{3.34}$$

$$\dot{\bar{\mathbf{x}}}'_2 = -V'_{\mathrm{sp}}\mathbf{p}_2 + \bar{\mathbf{u}}'_2 + A'_2 U^{0'}_{t,\|}\mathbf{p}_2\mathbf{p}_2 : \nabla'_2\mathbf{D}'_{2,z_0} - D'_{\mathrm{t},||}\nabla'_2 \ln \Psi'_{z_0} \tag{3.35}$$

$$\dot{\bar{\theta}}' = (\nabla'_2\bar{\mathbf{u}}'_2 + A'_2 U^{0'}_r \mathbf{D}'_{2,z_0}) : \mathbf{p}_2^{\perp}\mathbf{p}_2 - D'_{\mathrm{r}}\partial_\theta \ln \Psi'_{z_0} \tag{3.36}$$

with constants $D'_{\mathrm{t},||} = \frac{D_{\mathrm{t},||}T}{L^2}$, $D'_{\mathrm{r}} = D_{\mathrm{r}}T$, $U^{0'}_r = \frac{2U^0_r NT}{HL^2}$, $U^{0'}_t = \frac{U^0_{t,\|}NT}{HL^4}$, $V'_{\mathrm{sp}} = \frac{V_{\mathrm{sp}}}{U}$, and $A'_2 = \int \delta'^2_{a'}(z' - z'_0)dz' = HA_2$. We let $\tilde{U^0_{t,\|}} = A'_2 U^{0'}_{t,\|}$ to simplify notation. The nondimensional

form of the moment tensors are

$$\mathbf{S}'_{z_0} = \frac{N}{L^2}\mathbf{S}_{z_0}, \quad \mathbf{D}_{z_0} = \frac{N}{L^2}\mathbf{D}'_{z_0}.$$

Motor distributions

We nondimensionalize the bound motor distribution $\mathcal{M}_b$ so that it integrates to the ratio of bound to total motors:

$$\iiiint \mathcal{M}'_b \Psi' ds'\, d\mathbf{r}'_0\, d\mathbf{x}'\, d\mathbf{p} = \frac{N_b}{N_m} \quad \text{and} \quad \iiiint \mathcal{M}'_{b,z_0} \Psi_{z'_0} ds' d\mathbf{r}'_0 d\mathbf{x}'_2 d\theta = \frac{N_b}{N_m}.$$

Recalling that $\iiiint \mathcal{M}_b \Psi ds\, d\mathbf{r}_0\, d\mathbf{x}\, d\mathbf{p} = N_b$, plugging in the two-dimensional reductions (3.17) and (3.22) and comparing to the above, we obtain

$$\mathcal{M}_{b,z_0}(s, \mathbf{r}_0|(\mathbf{x}_2, \theta), t) = \frac{N_m H}{l L^2 N A'_2} \mathcal{M}'_{b,z_0}(s', \mathbf{r}'_0|(\mathbf{x}'_2, \theta), t^*) \tag{3.37}$$

and similarly for the full nondimensional distribution

$$\mathcal{M}'_b(s, \mathbf{r}_0|(\mathbf{x}, \mathbf{p}), t^*) = \frac{1}{A'_2} \mathcal{M}'_{b,z_0}(s, \mathbf{r}_0|(\mathbf{x}_2, \theta), t^*) \delta'_{a'}(z' - z'_0).$$

We rescale the distribution of free motors to the fraction of free motors, setting $\mathcal{M}_f = \frac{N_m}{L^2}\mathcal{M}_f$ and $\mathcal{M} = \frac{N_m}{L^2}\mathcal{M}'$ to obtain

$$\mathcal{M}'_f(\mathbf{r}'_0, t^*) = \mathcal{M}'(\mathbf{r}'_0) - \iiint \mathcal{M}'_{b,z_0} \Psi'_{z_0}\, ds'\, d\mathbf{x}'_2\, d\theta. \tag{3.38}$$

Using these definitions and the second set of nondimensional variables, we have

$$\begin{aligned} \partial_{t*}\mathcal{M}'_{b,z_0} + \partial_{s'}\mathcal{M}'_{b,z_0} + \frac{\tau}{T}\dot{\tilde{\mathbf{x}}}'_2 \cdot \nabla_{2'}\mathcal{M}'_{b,z_0} - \frac{\tau}{T}\dot{\tilde{\zeta}}\mathcal{M}'_{b,z_0} + \frac{\tau}{T}\dot{\tilde{\theta}}\partial_\theta \mathcal{M}'_{b,z_0} \\ = -k'_{\text{off}}\mathcal{M}'_{b,z_0} + k'_{\text{on}}\mathcal{M}'_f \mathbb{1}_{D_{r'_c}}, \end{aligned} \tag{3.39}$$

together with the dimensionless fluxes

$$\dot{\tilde{\mathbf{x}}}'_2 = -V'_{\mathrm{sp}}\mathbf{p}_2 + \frac{1}{A'_2}\widetilde{\mathbf{u}}'_2 + \frac{A'_3}{A'_2}U^{0'}_{t,\parallel}\mathbf{p}_2\mathbf{p}_2 : \nabla_2\mathbf{D}'_{2,z_0} - D'_{t,\parallel}\nabla'_2 \ln \Psi'_{z_0}$$

$$\dot{\tilde{\zeta}}' = \frac{1}{2A'_2}\widetilde{\partial_{z'}w'} - \frac{B'_1U^{0'}_{t,\perp}}{A'_2}\mathbf{p}_2\mathbf{p}_2 : \mathbf{D}'_{2,z_0}$$

$$\dot{\tilde{\theta}} = \left(\frac{1}{A'_2}\nabla'_2\widetilde{\mathbf{u}}'_2 + \frac{A'_3}{A'_2}U^{0'}_r\mathbf{D}'_{2,z_0}\right) : \mathbf{p}_2^{\perp}\mathbf{p}_2 - D'_r\partial_\theta \ln \Psi'_{z_0}$$

and constants $A'_3 = H^2A_3$, $U^0_{t,\perp} = \frac{U^0_{t,\perp}NT}{H^3L^2}$, $B'_1 = H^4B_1$, $k'_{\mathrm{off}} = k_{\mathrm{off}}\tau$, $k'_{\mathrm{on}} = k_{\mathrm{on}}\tau lNr_c/a$. We note that the dot in the above equations refers to a time derivative with respect to t^*. For completeness, we write the definition of the dimensionless disk of radius r'_c centered at $\mathbf{r}'_0$ as

$$D_{r'_c}(\mathbf{r}'_0) = \left\{(\mathbf{x}'_2, s', \theta) : \left|\mathbf{x}'_2 + \frac{l}{L}s'\mathbf{p}_2 - \mathbf{r}'_0\right|^2 \leq \frac{l^2}{L^2}r'^2_c\right\}.$$

As the bound motor density timescale is approximately a thousand times smaller than the microtubule timescale, we drop most of the terms with $\frac{\tau}{T}$ in Eq. (3.39), except the terms involving the steric parameters $U^{0'}_{t,\parallel}$ and $U^{0'}_r$ as their product with τ/T could end up being order one. We drop the term with $U^{0'}_{t,\perp}$, since our assumptions that the motion of the microtubule is constrained to a plane makes it a small number. Eliminating these terms we have

$$\begin{aligned}\partial_{t*}\mathcal{M}'_{b,z_0} + \partial_{s'}\mathcal{M}'_{b,z_0} + \frac{A'_3}{A'_2}\frac{\tau}{T}U^{0'}_{t,\parallel}\mathbf{p}_2\mathbf{p}_2 : \nabla'_2\mathbf{D}'_{2,z_0}\cdot\nabla_{2'}\mathcal{M}'_{b,z_0} \\ + \frac{A'_3}{A'_2}\frac{\tau}{T}U^{0'}_r\mathbf{D}'_{2,z_0} : \mathbf{p}_2^{\perp}\mathbf{p}_2\partial_\theta\mathcal{M}'_{b,z_0} = -k'_{\mathrm{off}}\mathcal{M}'_{b,z_0} + k'_{\mathrm{on}}\mathcal{M}'_f\mathbb{1}_{D_{r'_c}}.\end{aligned} \tag{3.40}$$

Again, for simplicity, we introduce $\widetilde{U^0_{t,\parallel}} = A'_3\tau U^{0'}_{t,\parallel}/(A'_2T)$ and $\widetilde{U^0_r} = A'_3\tau U^{0'}_r/(A'_2T)$.

For the simplified motor model, it is straightforward to see that the nondimensional

version of Eq. (3.27) is

$$\mathcal{M}'_{b,z_0}(\mathbf{r}'_0, s'|(\mathbf{x}'_2, \theta), t^*) = \begin{cases} 0 & \text{if } (\mathbf{x}'_2, s', \theta) \notin D'_{r'_c} \\ \min\left(C', \frac{\mathcal{M}'(\mathbf{r}'_0)}{\iiint_{D'_{r'_c}} \Psi'_{z'_0} d\mathbf{x}'_2 d\theta ds'}\right) & \text{if } (\mathbf{x}'_2, s', \theta) \in D'_{r'_c} \end{cases}, \tag{3.41}$$

where C' is an independent problem specific parameter.

Fluid

To nondimensionalize the bulk fluid equations, we first rescale the extra stresses as

$$\boldsymbol{\sigma}^{\mathrm{f}} = \sigma'_f \mathbf{S}'_{z_0}(\mathbf{x}_2, t) : \mathbf{E}'(\mathbf{x}, t)\delta'_a(z - z_0) \tag{3.42}$$

$$\boldsymbol{\sigma}^{\mathrm{t}} = -\sigma'_t \left(\mathbf{D}'_{z_0}(\mathbf{x}_2, t) \cdot \mathbf{D}'_{z_0}(\mathbf{x}_2, t) - \mathbf{S}'_{z_0}(\mathbf{x}_2, t) : \mathbf{D}'_{z_0}(\mathbf{x}_2, t)\right) \delta'^2_a(z - z_0), \tag{3.43}$$

and $\mathbf{E}(\mathbf{x}, t) = \frac{1}{T}\mathbf{E}'$, where $\mathbf{E}'$ is dimensionless rate-of-strain tensor. Here, we note that the dimensionless gradient of the velocity field has the form

$$\nabla' \mathbf{u}'(\mathbf{x}, t) = \begin{pmatrix} \nabla'_2 \mathbf{u_2}' & \frac{1}{\varepsilon}\partial_{z'} \mathbf{u}'_2 \\ \varepsilon(\nabla'_2 w')^T & \partial_{z'} w' \end{pmatrix}.$$

In Eqs. (3.42)-(3.43) the constants are $\sigma'_f = \frac{\sigma_f N}{THL^2}$, and $\sigma'_t = \frac{\sigma t N^2}{H^2 L^4}$. Plugging in Eq. (3.33) for Ψ and Eq. (3.37) for $\mathcal{M}_b$ into the force density (3.15), changing variables and setting $F = F_{\mathrm{st}}\left(1 - \frac{V_{\mathrm{m}}}{V_{\max}}\right)\frac{N_m}{L^2 H A'_2}$, we obtain $\mathbf{f}_2(\mathbf{x}, t) = -F\mathbf{f}'_2(\mathbf{x}', t')$ where

$$\mathbf{f}'_2(\mathbf{x}', t') = \delta'^2_{a'}(z' - z'_0) \iiiint \mathbf{p}_2 \delta\left(\mathbf{y}'_2 + \frac{l}{L}s'\mathbf{p}_2 - \mathbf{x}'_2\right) \Psi'_{z_0} \mathcal{M}'_{b,z_0} ds' d\mathbf{r}'_0 \, d\mathbf{y}'_2 \, d\theta.$$

Finally, we plug the nondimensional stresses and forces into the incompressible reduced Stokes equation (3.30)-(3.32) and we let the characteristic pressure be P_0 to find

($\varepsilon = H/L$)

$$-\nabla_2'^2 \mathbf{u}_2' - \frac{1}{\varepsilon^2}\partial_{z'z'}\mathbf{u}_2' + P_0'\nabla_2' q' = F'\mathbf{f}_2' + \tilde{\sigma}_f \nabla_2' \cdot \boldsymbol{\sigma}^{\mathrm{f'}} - \tilde{\sigma}_t \nabla_2' \cdot \boldsymbol{\sigma}^{\mathrm{t'}} \tag{3.44}$$

$$-\nabla_2'^2 w' - \frac{1}{\varepsilon^2}\partial_{z'z'}w' + P_0'\partial_{z'} q' = 0 \tag{3.45}$$

$$\nabla_2' \cdot \mathbf{u}_2' + \partial_{z'} w' = 0. \tag{3.46}$$

In the above, the constants are $P_0' = \frac{P_0 L}{\mu U}$, $F' = \frac{FL^2}{\mu U}$, $\tilde{\sigma}_f = \frac{L}{\mu U}\sigma_f'$, and $\tilde{\sigma}_t = \frac{L}{\mu U}\sigma_t'$. In the remainder of this paper and the supplemental movies, we drop all prime, star, tilde and bar notation and numerically solve the complete set of nondimensional equations which are summarized in Table 3.1. For reference, Table 3.2 lists all parameters and variables used in the model.

3.3 Implementation

In this section, we discuss the discretization of the nondimensionalized equations summarized in Table 3.1 and the development of a stable algorithm. Because Ψ_{z_0} and $\mathcal{M}_{b,z_0}$ evolve on two different timescales, we discretize Ψ_{z_0} at time t^n, $n = 0, \ldots, N_T$ with adaptive time step dt and $\mathcal{M}_{b,z_0}$ at time t^m, $m = 0, \ldots, N_{T*}$ with smaller adaptive time step dt^* such that $t^n \leq t^m \leq t^{n+1}$. After initializing Ψ_{z_0} and $\mathcal{M}_{b,z_0}$, we calculate the initial time step dt and, if the evolved motor model is being used, the initial time step dt^* as well. The motor forces and steric stresses are computed next, and used to solve the fluid equations. The new fluid velocities are used to update Ψ_{z_0} to time $t + dt$, and finally M_{b,z_0} is updated to time $t + dt$ based on the updated Ψ_{z_0}. New time steps are computed, and the simulation

continues. The procedure is summarized in algorithm 3 and details are given below. Our numerical scheme is similar to that of our previous scheme in [24]. The primary differences are that the fluid equations are now solved in three dimensions, and that extra stress terms resulting from the steric interactions are included.

To compute the time step dt, we calculate the maximum of the angular and linear advection velocities in Eq. (3.33) and adjust the time step according to the CFL condition. To find the small time step for $\mathcal{M}_b$, we limit the fraction of available motors that can bind or unbind in any given time step. To calculate dt^*, we compare the change due to s-advection with the greatest change due to binding and unbinding, and use the more restrictive of the two to clamp dt^*.

3.3.1 Microtubule density

We discretize $\Psi_{z_0}(\mathbf{x}_2, \theta)$ over the domain $(\mathbf{x}_2, \theta) \in [-1, 1]^2 \times [0, 2\pi]$ at the plane of motion $z = z_0$ with a triply periodic grid of size $N_x \times N_y \times N_\theta$, with $N_x = N_y$. The advection terms in (MT1) (Table 3.1) are discretized spatially with an upwinding scheme and Superbee flux limiter [13]. The equation (MT1) is integrated in time using second order Adams-Bashforth for the advective terms and Crank-Nicolson for the diffusive terms. The resulting system of equations for Ψ_{z_0} is solved using the Conjugate Gradient method with Incomplete Cholesky factorization used as a preconditioner.

Algorithm 1 Numerical evolution scheme for the coupled microtubule density, motor protein distribution, and fluid velocity equations.

Initialize Ψ_{z_0} and $\mathcal{M}_{b,z_0}$.

Precompute LU-decomposition of semi-spectral matrices for all frequency pairs.

while $t < t_{\text{end}}$ **do**

Compute adaptive dt.

Compute $\Psi_{z_0}(t + dt)$ by solving (MT1)-(MT3) using second order Crank-Nicolson for the diffusive terms and Adams-Bashforth 2 for the advection terms.

if (using EM) **then**

set $t^*_{\text{end}} = t + dt$.

while $t^* < t^*_{\text{end}}$ **do**

Compute adaptive dt^*.

Compute $\mathcal{M}_{b,z_0}(t^* + dt^*)$ by solving (EM) with Adams-Bashforth 2.

Update $\mathcal{M}_f$ from M_{b,z_0} with (MF).

end while

else if (using simplified motor model) **then**

Solve (SM)

Update $\mathcal{M}_f$ from M_{b,z_0} with (MF).

end if

Calculate extra stresses.

Calculate motor force (F1) using trapezoidal rule and a local grid.

Solve semi-spectral (U1)-(U3).

end while

3.3.2 Motor distributions

To advance $\mathcal{M}_{b,z_0}$ forward in time according to the evolved motor model (EM) in Table 3.1, two-step Adams-Bashforth with variable time step dt^* is used to discretize the s-advection term and the binding and unbinding terms. We clamp $\mathcal{M}_{b,z_0}(\mathbf{r}_0, s|\mathbf{x}_2, \theta)$ so that $\iiint \mathcal{M}_{b,z_0}\Psi_{z_0}\, ds\, d\mathbf{x}_2\, d\theta \leq \mathcal{M}(\mathbf{r}_0)$.

The bound motor density $\mathcal{M}_{b,z_0}$ is high-dimensional as it tracks motor tail position, the filament arclength parameter, and the center of mass and orientation of the filament the bound motor head is attached to. However, since a head detaches if the elongation of the motor stalk exceeds a certain threshold, $\mathcal{M}_{b,z_0}$ can be computed and stored sparsely in a local grid around $\mathbf{r}_0$. The specific condition $\mathbf{x}_2 + \frac{l}{L}s\mathbf{p}_2 - \mathbf{r}_0 \leq r_c$ allows further pruning of this localized configuration space. In our formulation, the activity of the motors anchored at $\mathbf{r}_0$ is independent of motors anchored elsewhere. In discrete form each cell $\mathbf{y}$ stores the local grid over $\mathbf{x}, \theta, s$ for bound motors whose tails are anchored anywhere within the boundaries of cell $\mathbf{y}$. Each cell's motor distribution is updated in parallel. $\mathcal{M}_{b,z_0}$ is stored as a two-dimensional array over $\mathbf{r}_0$, each containing an unrolled flat array for $\mathbf{x}_2, s, \theta$. We solve the evolved motor density equation on the GPU, where each $\mathbf{r}_0$ is updated in SIMD fashion by several threads. Another advantage to this layout is that $\mathcal{M}_{b,z_0}$ independent outermost two-dimensional array can be split up and sent to multiple GPUs, or solved in batches on a single GPU if the shared memory is exceeded. As grid resolution increases, the three copies of $\mathcal{M}_{b,z_0}$ at the current and two previous times required by the two-step Adams-Bashforth time integration scheme may not all fit onto the GPU on-board memory simultaneously and instead need to be solved a few rows at a time.

Since we do not track free motor heads, we can discretize the distribution of free motors $\mathcal{M}_f$ and total motors $\mathcal{M}$ over a uniform grid of size $N_x \times N_y$. Updating $\mathcal{M}_f$ from $\mathcal{M}$ and $\mathcal{M}_{b,z_0}$ is straightforward and parallelizable over $\mathbf{x}_2$ by evaluating the discretized form of (MF) in Table 3.1.

3.3.3 Fluid

We discretize the domain into $N_x \times N_y \times N_z$ grid cells, where $N_x = N_y$, and solve for the fluid state at each discrete time t^n. $\mathbf{u}_2$ and q are sampled at cell centers, while w is sampled at the z faces. As we have periodic boundary conditions in the $\mathbf{x}_2-$plane, we use a semi-spectral approach and take the Fourier transform in $\mathbf{x}_2$ of (U1)-(U3), giving for each frequency pair $\mathbf{k} = (k_x\, k_y)^T$

$$(|\mathbf{k}|^2 - \frac{1}{\varepsilon^2}\partial_{zz})\hat{\mathbf{u}}_2^n + iP_0\hat{q}^n\mathbf{k} = F\hat{\mathbf{f}}_2^n + i\sigma_f\hat{\boldsymbol{\sigma}^{\mathrm{f}}}^n\mathbf{k} + i\sigma_t\hat{\boldsymbol{\sigma}^{\mathrm{t}}}^n\mathbf{k} \tag{3.47}$$

$$(|\mathbf{k}|^2 - \frac{1}{\varepsilon^2}\partial_{zz})\hat{w}^n + P_0\partial_z\hat{q}^n = 0 \tag{3.48}$$

$$i\mathbf{k}\cdot\hat{\mathbf{u}}_2^n + \partial_z\hat{w}^n = 0. \tag{3.49}$$

Equations (3.47)-(3.49) yield an independent $(4N_z - 1) \times (4N_z - 1)$ linear system for each frequency pair. This formulation is computationally advantageous for several reasons. First, the equations for each $\mathbf{k}$ can be solved independently, allowing simple parallelization. Second, the coefficient matrix of each linear system is constant in time, and an LU-factorization for each can be precomputed and stored. We can reasonably store $N_x \times \frac{N_y}{2}$ separate $(4N_z - 1) \times (4N_z - 1)$ matrices, and use them to solve for multiple right-hand sides. The FFTW library [18] is used with precomputed transformation mappings to efficiently perform the FFT and inverse FFT.

As the microtubules are concentrated around the $z = z_0$ plane, it is desirable to have more accuracy there and the thin δ_a-width region around it than in the distant assay regions above and below it. Given the aforementioned scaling of each frequency pair fluid solve matrix with N_z^2, we use a nonuniform grid with variable spacing in the z dimension. We store $\hat{\mathbf{u}}_2$, and $\hat{q}$ at the $z-$cell centers, and $\hat{w}$ at the $z-$cell faces. A schematic of the z-grid is shown in Figure 3.3.

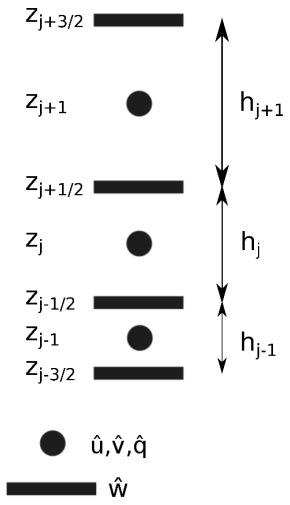


Figure 3.3: Schematic of the nonuniform staggered grid in z used to store the spectral values of $\hat{\mathbf{u}}, \hat{w}, \hat{p}$.

To compute second derivatives with respect to z at a $z-$cell j, we construct a fourth order Lagrange interpolating polynomial using z_{j-2}, z_{j-1}, z_j, z_{j+1}, z_{j+2}, and differentiate twice. Near the boundaries, we use boundary data and the no slip boundary condition for the extreme samples and drop to third order interpolation for the bottom-most and top-most equations. For first derivatives at a $z-$face $j + 1/2$, we construct a third order Lagrange interpolating polynomial using z_{j-1}, z_j, z_{j+1}, z_{j+2}, and differentiate once. Near the boundaries, we use the nearest four samples to construct the interpolating polynomial.

We find that our semi-spectral fluid solver is second-order accurate.

The motor force calculation is the single most computationally intense portion of the algorithm because the force spreading dictates that nearby forces be calculated in order to determine the total force at $\mathbf{x}$. In terms of implementation, this effectively increases the already high dimensionality of the bound motor distribution, whether it is approximated with the simple motor model or the evolved motor model. To account for the motor force's highly parallel but computationally intensive nature, it is calculated on a GPU using a similar scheme to the bound motor solve described above. A speedup of roughly forty times is observed versus a single-core implementation.

Computation of the stress tensors is straightforward and parallelizable. $\boldsymbol{\sigma}^{\mathrm{f}}$ depends on the rate-of-strain tensor $\mathbf{E}$, which we have only for the previous time step since we compute the extra stresses before the fluid solve. We therefore linearly extrapolate $\mathbf{E}$ at the new time $t + dt$ using the current and previous values, as in [15].

3.4 Results

In this section, we present results for various experiments with the following parameters held constant: 2μm-length microtubules, $k_{\mathrm{on}} = 25$, $k_{\mathrm{off}} = .1$, $U_t = -.01$, $125 \times 125\mu\mathrm{m}^2$ assay, $N_m = 3 \times 10^6$ motors, $V_{\max} = 1\mu\mathrm{m}s^{-1}$, and $N = 22300$ microtubules. Our nonuniform z-grid has 30 evenly sized fine z-cells covering the range $\{-.5, -.4\}$, with $z_0 = \frac{-h}{2} + .05 = -.45$ in the middle. Above $z = -.4$, the height of each cell doubles consecutively until the cell size is sixteen times greater than the fine z-cells at the bottom. Our final 3D grid dimensions are $128 \times 128 \times 49$, with 32 cells in s and θ, at which we find

the qualitative features to be well-resolved.

In the figures, we plot the nondimensionalized spatial microtubule distribution

$$\Psi_{\text{spatial},z_0}(\mathbf{x}_2,t) = \int \Psi_{z_0}(\mathbf{x}_2,\theta,t)d\theta \tag{3.50}$$

with color ranging from white (low) to blue (high). We plot the nondimensional spatial bound motor distribution

$$\mathcal{M}_{\text{b,spatial},z_0}(\mathbf{r}_0,t) = \iiint \mathcal{M}_{b,z_0}(s,\mathbf{r}_0,t|\mathbf{x}_2,\theta)\Psi_{z_0}(\mathbf{x}_2,\theta,t)dsd\mathbf{x}_2d\theta \tag{3.51}$$

with color ranging from tan (low) to red (high). The colorbars are annotated with the corresponding percentage of the total available motors in the bound configuration, at the low and high ranges of each normalization. Finally, we compute the orientation matrix

$$\mathbf{N}(\mathbf{x}_2,t) = \frac{\int \mathbf{p}_2\mathbf{p}_2^T\Psi_{z_0}(\mathbf{x}_2,\theta,t)d\theta}{\int \Psi_{z_0}(\mathbf{x}_2,\theta,t)d\theta} \tag{3.52}$$

and draw its eigenvectors in red scaled by their associated eigenvalues. When present, velocity vectors (black) and orientation eigenvectors are plotted for every fifth cell for clarity.

3.4.1 Evolved motor model

Single clump. We first examine the processive behavior of a clump of aligned filaments as the steric alignment parameter is varied between $U_r = 0$ and .01. We present results for $U_r = 0, .01$ and the evolved motor model in Figure 3.4. We observe that as U_r increases to .01 the clump better maintains its shape, whereas at $U_r = 0$ microtubules become concentrated along the leading edge of the clump, which assumes a widening crescent-like shape. The bottom row of Figure 3.5 shows the microtubule orientation field at the final

frame displayed in the upper rows with $U_r = 0$ (left) and $U_r = .01$ (right). For higher values of U_r, the microtubule orientation field is uniformly aligned. For smaller values of U_r, the microtubule orientation field at the clump's leading edge becomes tangential to the leading edge, as the rods rotate to avoid compression or extension by the steep velocity gradient, clearly visible in the $U_r = 0$ case. For $U_r = .01$, the steric resistance to rotate relative to neighboring microtubules counteracts this effect and the orientation field remains more uniform. *In vitro* experiments have shown shape persistence in aligned clumps [62], qualitatively similar to the $U_r = .01$ case.

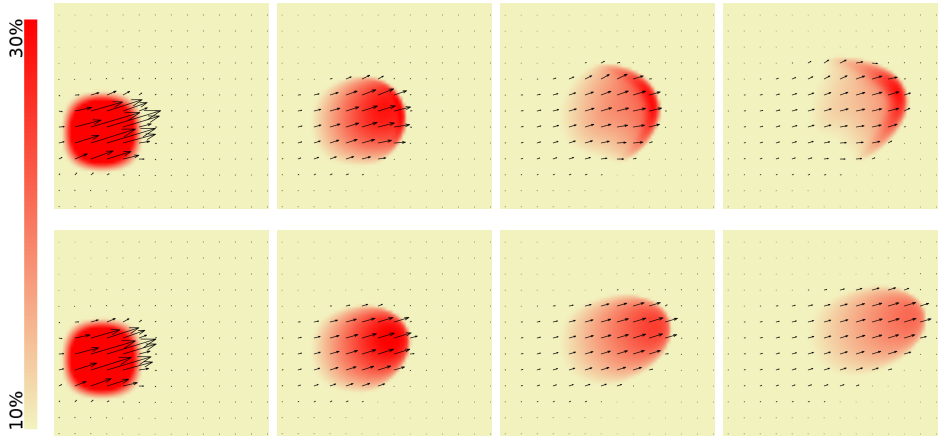


Figure 3.4: Single aligned microtubule clump simulation with evolved motor model. Bound motor protein density visualized with gradient from tan (low) to red (high). Fluid velocity plotted as a black arrow every 5th cell. First row: $U_r = 0$. Second row: $U_r = .01$. Images are at times t=0, 15, 30, and 45 sec.

Colliding clumps. We examined the behavior of colliding clumps for three values of the steric alignment parameter $U_r = 0, .001, .01$ and both head-on and perpendicular collisions. When clumps collide, the behavior depends on the angle between the microtubule orientations of the clumps.

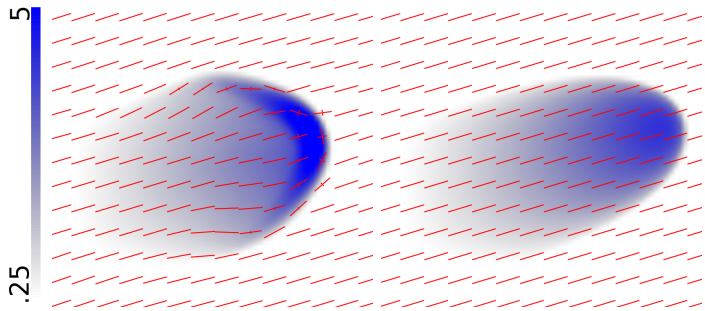


Figure 3.5: Single aligned microtubule clump simulation with evolved motor model. Microtubule density visualized with gradient from white (low) to blue (high). Microtubule orientation eigenvectors plotted as red vectors every 5th cell. Simulation shown at time t=45 sec for $U_r = 0$ (left) and $U_r = .01$ (right).

A nearly perpendicular collision as in Figure 3.6 results in the clumps merging and moving as a single clump for all U_r tested. While the U_r term drives local alignment, alignment also occurs in the $U_r = 0$ case as follows. When the self-propulsion velocity is zero, microtubules move passively with the flow. As motor forces act directly on the fluid, motor forces acting in opposite directions cancel out. This cancellation occurs in the example depicted in Figure 3.6, where the resultant force points in the average direction of the colliding microtubule orientations, in this case, upwards. These two mechanisms give different qualitative results as illustrated in Figure 3.6. In the $U_r = 0$ case (third row), the microtubule distribution remains isotropic as long as local fluid flow remains negligible, whereas in the $U_r = .01$ case (fourth row), we observe steric alignment of the microtubules throughout the entire domain. Higher U_r results in steeper gradients in microtubule orientation and density at the midline. The higher concentration in turn leads to stronger motor forces and higher fluid velocities. These U_r-dependent collision phenomena are observed wherever two regions of dense microtubules collide.

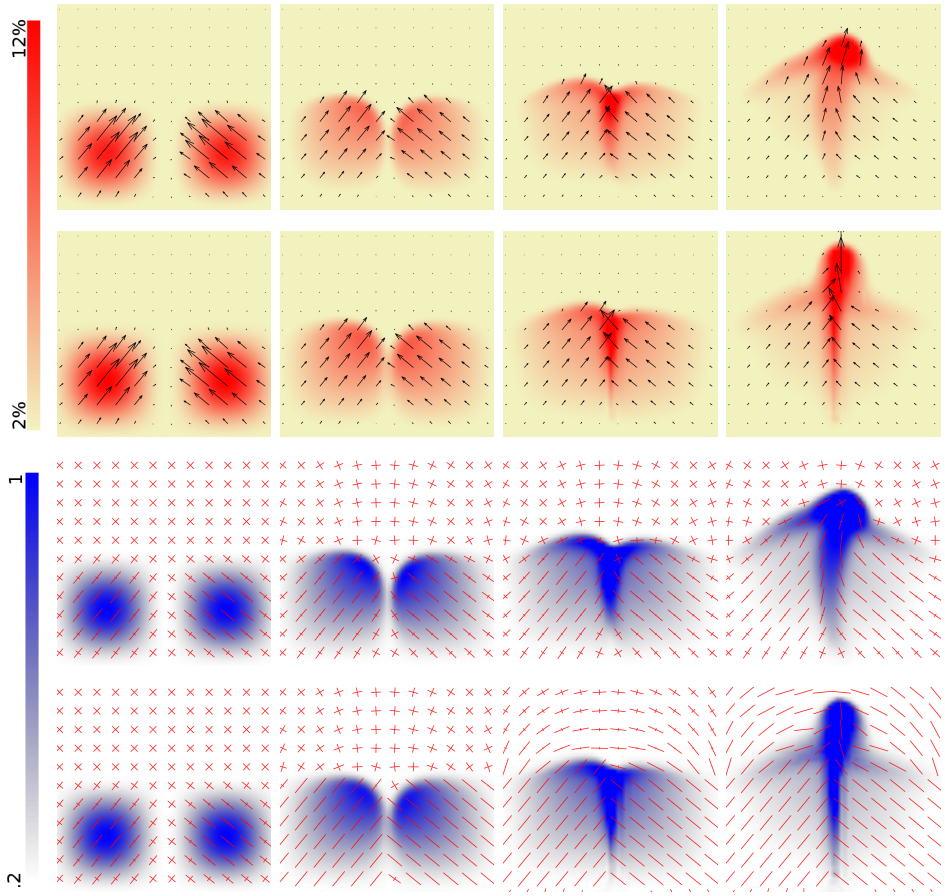


Figure 3.6: Nearly perpendicular microtubule clumps driven by the evolved motor model colliding. First row: $U_r = 0$ velocity field. Second row: $U_r = .01$ velocity field. Third row: $U_r = 0$ orientation field. Fourth row: $U_r = .01$ orientation field. Images are at times t=0, 45, 90, and 180 sec.

In the case of two clumps with antiparallel orientations colliding close to head-on (Figure 3.7, supplemental movie 1 first example), significant differences are observed for $U_r = 0$ versus $U_r = .01$. In the $U_r = 0$ case, the motor forces drive an extensional fluid flow on either side of the collision centerline, creating two clumps moving in opposite directions. In the case $U_r = .01$, the steric force prevents alignment with the extensional flow and the microtubules of each clump slide past each other. As a result of slight differences in the

original clump position, the clumps break down after collision, and smaller clumps pass through each other and continue along the initial clump trajectories. The orientation field (Figure 3.7, fourth row) shows that the microtubules do not rotate during the initial collision and aftermath.

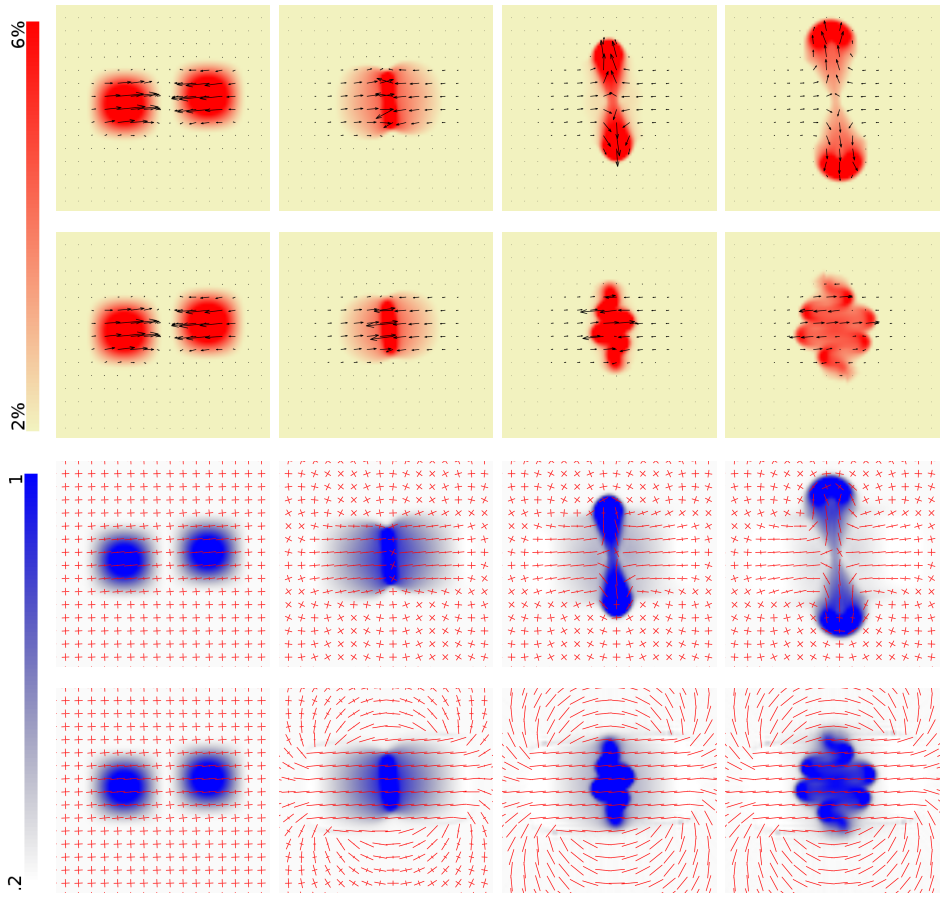


Figure 3.7: Antiparallel microtubule clumps driven by the evolved motor model colliding. First row: $U_r = 0$ velocity field. Second row: $U_r = .01$ velocity field. Third row: $U_r = 0$ orientation field. Fourth row: $U_r = .01$ orientation field. Images are at times t=0, 90, 180, and 270 sec. See supplemental movie 1 first example.

Vortex Lattice. To test our model's ability to reproduce characteristics of the lattice of vortices observed in [65], we simulate four overlapping rings of microtubules oriented

in clockwise fashion as shown in Figure 3.8 and the first example in supplemental movie 2. In the overlapping regions, the microtubules from adjacent rings are oriented opposite each other. We observe extensional flow in the dense overlapping regions combined with counterclockwise rotation driven by the initial orientations. With $U_r = 0$ the rotational flow develops four vortices rotating clockwise centered about the spaces between the initial four vortices. The new vortices contract then expand outward until they develop overlapping regions moving in opposite directions, similar to the initial condition. The process repeats itself; extensional flow and rotation forms again in the overlapping regions, leading to the formation of four vortices rotating counterclockwise at the original four vortex locations. Due to diffusion, the maximum concentration and hence velocity decreases on average throughout the process. Due to symmetry breaking, the transition from vortices with overlapping regions to new vortices with overlapping regions and opposite rotation repeats a few times at most, depending on parameters, until the original structure is lost. Increasing U_r from 0 to .01 increases the maximum microtubule density and flow velocity and gives steeper gradients in microtubule concentration and orientation, as seen in previous examples. It also affects the degree to which the initial dense overlapping regions break down with the rotational forcing from the motor proteins. In particular, for $U_r = .01$ (Figure 3.8, second and fourth rows), the dense overlapping regions extend but do not separate and thus preserve much of the original four vortex structure. With the inclusion of the steric interaction term, our results are more consistent with the experiments of [65], which demonstrate a temporally persistent lattice of vortices.

Perturbation. We perturb a uniform isotropic microtubule density in both space

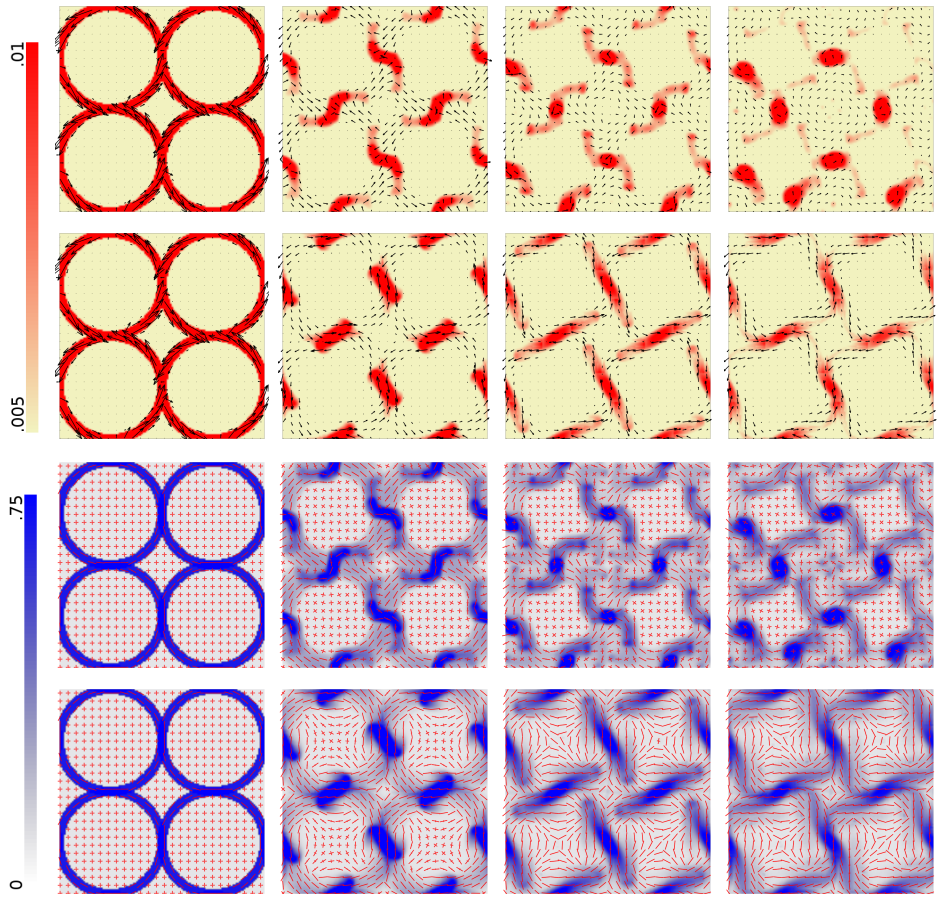


Figure 3.8: Vortex lattice experiment with evolved motor model. First row: $U_r = 0$ velocity field. Second row: $U_r = .01$ velocity field. Third row: $U_r = 0$ orientation field. Fourth row: $U_r = .01$ orientation field. Images are at times t=0, 240, 480, and 720 sec. See supplemental movie 2 first example.

and orientation by adding

$$\frac{1}{a}\sum_{i,j=1}^{8} \epsilon_{ij}\cos(\pi i x + \xi_{ij})\cos(\pi j y + \xi_{ij})P_{ij}(\theta), \tag{3.53}$$

where ϵ_{ij} is a uniform random number in $[-.001, .001]$, ξ_{ij} is a uniform random number in $[0, 2\pi]$, a is a normalization constant and $P_{ij}(\theta)$ are third order polynomials in $\cos(\theta)$ and $\sin(\theta)$ with random coefficients in $[-1, 1]$. The $U_r = 0$ case results in a spatiotemporally stable Ψ_{z_0} density (Figure 3.9, supplemental movie 3). In the $U_r = .01$ case, continuous

narrow tracks of stationary microtubules form and remain stable.

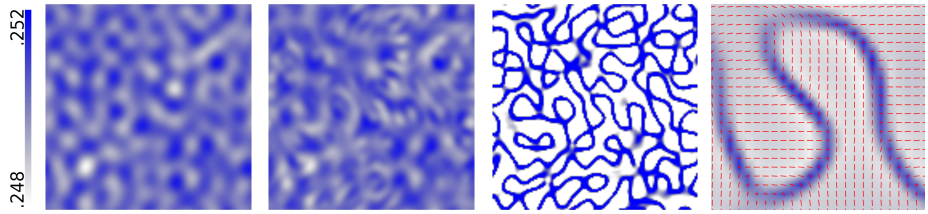


Figure 3.9: Perturbation in x and θ giving rise to stationary concentrated pattern with $U_r = .01$ for evolved motor model. First three images: evolution of microtubule density in time. Fourth image: magnified section of final top row image with orientation eigenvectors in red. The microtubule density concentrates along steep gradients in the microtubule orientation field. Images are at times t=0, 75, and 113 sec. See supplemental movie 3.

3.4.2 Evolved motor model with self-propulsion

Colliding clumps. We repeat the antiparallel colliding clumps experiment with the addition of a phenomenological self-propulsion velocity V_{sp} (Eq. (3.2), (MT2)) in Figure 3.10 and the second two examples in supplemental movie 1. In the case of $V_{\text{sp}} = 0$, illustrated in Figure 3.7, the clumps break up as they collide. At $V_{\text{sp}} = 1$ and $U_r = .01$, the clumps pass through each other largely intact. For $V_{\text{sp}} = 1$ and $U_r = 0$, we see a combination of both effects, with some passthrough and some spreading of microtubules with the extensional flow formed in the collision. In general, varying the value of V_{sp} between 0 and 1 leads to a corresponding combination of the extreme $V_{\text{sp}} = 0$ and $V_{\text{sp}} = 1$ behaviors. The experiments of [62] demonstrate a combination of passthrough and breakup when clumps collide. Experiments of the behaviors of microtubules undergoing collisions [65] show that colliding microtubules can merge and realign or pass through depending on the angle of

collision. With the addition of a self-propulsion term, our numerical experiments reproduce such behaviors.

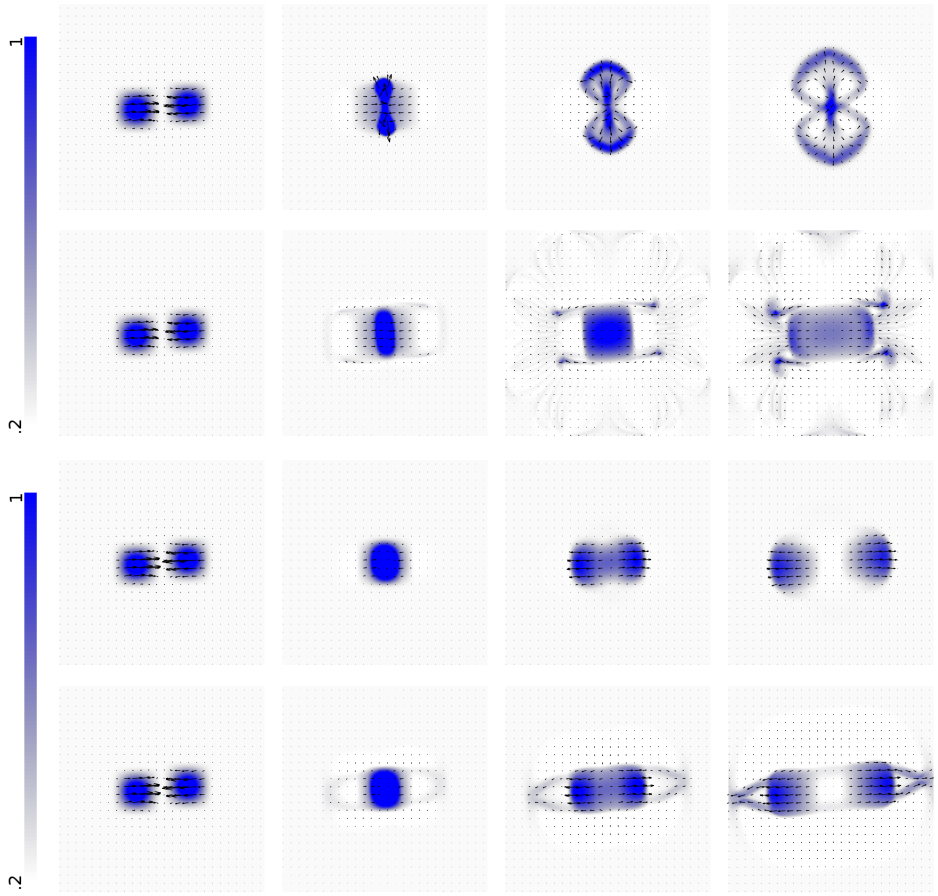


Figure 3.10: Colliding clump experiment with added self-propulsion velocity. First row: $U_r = 0$ and $V_{\text{sp}} = .1$. Second row: $U_r = .01$ and $V_{\text{sp}} = .1$. Third row: $U_r = 0$ and $V_{\text{sp}} = 1$. Fourth row: $U_r = .01$ and $V_{\text{sp}} = 1$. Images are at times t=0, 105, 210, 315 sec in rows one and two, and at t=0, 24, 48, 72 sec in rows three and four. See second and third examples in supplemental movie 1.

Vortex rings. We repeat the four ring vortex experiment with the addition of a self-propulsion velocity V_{sp} (Figure 3.11, second and third examples in supplemental movie 2). With $V_{\text{sp}} = .1$ and $U_r = 0$, depicted in the first row, motor forces at the overlapping re-

gions of the initial rings create a shear flow that separates these regions and, in conjunction with the self-propulsion, creates counterclockwise vortices at the separatrix between clockwise vortices as seen in the second image of the first row. This separates the dense bands of microtubules into two connected bands that translate and rotate away from each other, eventually meeting other bands at the centers of the original rings in a cross-like pattern (third image). The microtubules gather at the centers of the crosses, then reverse direction and expand outward in a nonsymmetric way (fourth image), similar to the switching behaviors observed when the experiment is run without self-propulsion (Figure 3.8). With $V_{\text{sp}} = .1$ and $U_r = .01$, depicted in the second row, the steric alignment prevents the shear flow from separating the initial overlapping regions, and the self-propulsion drives antiparallel sliding that stretches the dense microtubule regions into long cohesive bands (second image, second row). The bands break down into smaller clumps (third image), but the steric alignment term keeps the new clumps following the paths of the initial bands, which roughly correspond with the initial four vortex structure (fourth image) as observed without self-propulsion (Figure 3.8). Increasing the self-propulsion velocity to $V_{\text{sp}} = 1$ causes the self-propulsion to dominate the effects of the motor forces, so microtubule passthrough (with alignment if $U_r > 0$) becomes dominant. With $U_r = 0$ (third row), switching events occur continuously and the four quadrants of the assay are symmetric. We observe that with $U_r = .01$, passthrough rapidly breaks up the ring structure (fourth row). As a result, no switching events occur.

Perturbation We repeat the perturbation experiment with the addition of a self-propulsion velocity V_{sp} (Figure 3.12, second and third examples in supplemental movie 3).

At $V_{\text{sp}} = 1$ or .1 and $U_r = 0$ the clumps translate but simply pass through one another without increasing in density or aligning. With $V_{\text{sp}} = .1$ and $U_r = .01$ (first row), we get fast translational microtubule bands as opposed to the stationary continuous tracks in the $V_{\text{sp}} = 0$ case shown in Figure 3.9. At $V_{\text{sp}} = 1$ and $U_r = .01$ (second row), the bands form faster and are denser than in the $V_{\text{sp}} = .1$ case.

3.4.3 Simplified motor model

Vortex rings. For the parameters, experiments, and timescales presented here, the differences in density and feature shape and location are observed between the microtubule distribution fields generated by the simplified and evolved motor models are minor. One notable exception is that in the vortex ring experiment, the evolved motor model drives clockwise rotation in the four central clumps whereas the simplified motor model drives counterclockwise rotation (Figure 3.13). This effect is due to a slight difference in the motor force pattern around each clump. On timescales longer than those presented in this work, simulations may eventually show significant divergence.

Colliding clumps. Results from the antiparallel colliding clump simulation driven by the simplified motor model are presented in Figure 3.14 (supplemental movie 4) for values of the parameter $C = 10, 50, 250$. Increasing C not only increases the fluid velocities in the simulation by increasing value of $\mathcal{M}_{b,spatial,z_0}$ for a given Ψ_{z_0}, but also changes the flow features that emerge over time. We observe that for the highest tested value $C = 250$, any cell with a $\Psi_{spatial,z_0}$ value over a threshold results in fully bound motors, exercising the

second argument to the minimum function in the definition of the simplified motor model (SM). Therefore two cells with distinct $\Psi_{spatial,z_0}$ values above the threshold will produce motor forces of equal magnitude, changing the emergent behavior within the assay.

Perturbation with and without motor-based fluid forces. In Figure 3.15 (supplemental movie 5), we repeat the perturbation experiment with self-propulsion in the presence and absence of the hydrodynamic forces generated by the motor proteins. This allows us to test the observation of [62] that the stability and size of the observed filament patterns depend on long-range hydrodynamic interactions. Consistent with [62], we observe larger flow structures forming in a shorter amount of time in the presence of the fluid flows driven by the motor proteins. We used the simplified motor model, $C = 100$, $U_r = .01$, and $V_{\text{sp}} = .1$.

3.5 Conclusions

We have developed a modeling and simulation framework coupling multiple microscopic models of propulsion to macroscopic steric and hydrodynamic interactions in a quasi-two-dimensional assay. Populations of bound and free motor proteins and microtubules are represented as continuum distributions. The framework facilitates study of the relative effects of hydrodynamic and steric interactions on emergent phenomena. Stress tensors arising from rotational and translational steric interactions and self-propulsion are supported in addition to body forces from active motor proteins. Experimentation is needed to empirically determine the steric interaction parameters U_r and U_x. We avoid closure approximations in the z dimension, and high precision around a z-plane of interest is achieved

without incurring significant computational overhead. Results demonstrate our framework's ability to replicate some of the behavior of individual and colliding clumps of filaments including crossovers, alignment, merging, and splitting [61], and support observations of [62] regarding hydrodynamic effects.

We present two motor protein models, the evolved motor model which incorporates motor head procession and binding/unbinding dynamics, and the simplified motor model which determines the bound motor distribution instantaneously as a function of the microtubule distribution, and therefore eliminates the high-dimensional and computationally expensive motor evolution at the smaller timescale t^*. While the different models may yield visually similar motor distributions, they can result in qualitatively different dynamics as illustrated in Figure 3.13. Additional motor models could be investigated within our framework, for example, models accounting for cooperativity or competition between motor proteins.

One limitation of our model is the way we model the motor stalk force. The motor stalk force has a component parallel to the filament, perpendicular to the filament and in the x, y-plane, and in the z-direction. The component parallel to the filament is equivalent to the motor force we apply. We ignore the perpendicular component in the x, y-plane, which would act to translate and re-orient the filament to align the filament director with the in-plane motor stalk force. We also ignore the component in the z-direction, which would act to move filaments in the z direction, since we phenomenologically model the filament distribution in the z-direction. In future work it would be interesting to refine our motor model to capture the second component of the force as well.

Our model applies to a microtubule gliding assay, where the motors generate force monopoles. If multi-headed motor complexes bind two filaments, this leads to force dipoles as in [19]. While our framework could be extended to such cases, in practice this leads to a very high-dimensional motor complex state space which may require additional work to make feasible.

Motor forces on the fluid compose flow features in the microtubule density by advecting all local microtubules with the same velocity. Even without any steric interaction terms, two colliding clumps will proceed in a direction roughly equal to the average of their orientations. However, the motor forces acting on the fluid are prone to cancellation in isotropic or anti-aligned microtubule configurations. Combining either motor model with a self-propulsion term in the microtubule advective flux provides a mechanism for anti-parallel sliding resulting in persistent motion of the microtubules. Addition of the self-propulsion term enables the passthrough of colliding clumps, consistent with the simulations of [61].

Model Summary

Microtubules

$$\partial_t \Psi_{z_0} + \nabla_2 \cdot (\dot{\mathbf{x}}_2 \Psi_{z_0}) + \partial_\theta (\dot{\theta} \Psi_{z_0}) = 0 \tag{MT1}$$

$$\dot{\mathbf{x}}_2 = -V_{\mathrm{sp}} \mathbf{p}_2 + \mathbf{u}_2 + U^0_{t,\parallel} \mathbf{p}_2 \mathbf{p}_2 : \nabla_2 \mathbf{D}_{2,z_0} - D_{t,\parallel} \nabla_2 \ln \Psi_{z_0} \tag{MT2}$$

$$\dot{\theta} = \nabla_2 \mathbf{u}_2 + U^0_r \mathbf{D}_{2,z_0} : \mathbf{p}_2^\perp \mathbf{p}_2 - D_r \partial_\theta \ln \Psi_{z_0} \tag{MT1}$$

Motors

$$\begin{aligned} &\partial_t \mathcal{M}_{b,z_0} + \partial_s \mathcal{M}_{b,z_0} + U^0_{t,\parallel} \mathbf{p}_2 \mathbf{p}_2 : \nabla_2 \mathbf{D}_{2,z_0} \cdot \nabla_2 \mathcal{M}_{b,z_0} \\ &\quad + U^0_r \mathbf{D}_{2,z_0} : \mathbf{p}_2^\perp \mathbf{p}_2 \partial_\theta \mathcal{M}_{b,z_0} = -k_{\mathrm{off}} \mathcal{M}_{b,z_0} + k_{\mathrm{on}} \mathcal{M}_f \mathbb{1}_{D_{r_c}} \end{aligned} \tag{EM}$$

$$\mathcal{M}_{b,z_0} = \begin{cases} 0 & \text{if } |\mathbf{x}_2 + \frac{l}{L} s \mathbf{p}_2 - \mathbf{r}_0| \geq r_c \\ \min\left(C, \frac{\mathcal{M}(\mathbf{r}_0)}{\iiint_{D_{r_c}} \Psi_{z_0} d\mathbf{x}_2 d\mathbf{p}_2 ds} \right) & \text{if } |\mathbf{x}_2 + \frac{l}{L} s \mathbf{p}_2 - \mathbf{r}_0| < r_c \end{cases} \tag{SM}$$

$$\mathcal{M}_f = \mathcal{M} - \iiint \mathcal{M}_b \Psi_{z_0} \, ds \, d\mathbf{x}_2 \, d\theta \tag{MF}$$

Fluid

$$-\nabla_2^2 \mathbf{u}_2 - \tfrac{1}{\varepsilon^2} \partial_{zz} \mathbf{u}_2 + P_0 \nabla_2 q = \sigma_f \nabla_2 \cdot \boldsymbol{\sigma}^{\mathrm{f}} - \sigma_t \nabla_2 \cdot \boldsymbol{\sigma}^{\mathrm{t}} + F \mathbf{f}_2 \tag{U1}$$

$$-\nabla_2^2 w - \tfrac{1}{\varepsilon^2} \partial_{zz} w + P_0 \partial_z q = 0 \tag{U2}$$

$$\nabla_2 \cdot \mathbf{u}_2 + \partial_z w = 0 \tag{U3}$$

Force

$$\mathbf{f}_2 = F \delta_a^2 (z - z_0) \iiint\!\!\iint \mathbf{p}_2 \delta(\mathbf{y}_2 + \tfrac{l}{L} s \mathbf{p}_2 - \mathbf{x}_2) \Psi_{z_0} \mathcal{M}_{b,z_0} \, ds \, d\mathbf{r}_0 \, d\mathbf{y}_2 \, d\theta \tag{F1}$$

Table 3.1: Summary of the nondimensionalized model equations for the evolution of microtubules, motors and fluid in a gliding assay. The primes and nondimensionalizing constants have been dropped for simplicity.

Model Parameters and Variables

Ψ	microtubule density
$\mathbf{x}, \mathbf{p}$	three-dimensional position and orientation
θ	in-plane orientation angle
$\dot{\mathbf{x}}, \dot{p}$	microtubule linear and angular velocity
V_{sp}	microtubule self-propulsion velocity
N	total number of microtubules
$\boldsymbol{\sigma}^{\mathrm{f}}, \boldsymbol{\sigma}^{\mathrm{t}}, \boldsymbol{\sigma}^{\mathrm{p}}$	microtubule inextensibility stress, steric interaction stress, and combined total stress
σ_f, σ_t	inextensibility and steric interaction stress coefficients
$\mathbf{u}, q, \mu$	fluid velocity, pressure, and viscosity
U_t, U_r	translational and rotational steric potentials
K_t, K_r	Maier-Saupe translational and rotational steric potentials
$D_{\mathrm{t},\perp}, D_{\mathrm{t},\|\|}, D_{\mathrm{r}}$	linear perpendicular diffusion, linear parallel diffusion, and rotational diffusion
$\mathbf{D}, \mathbf{S}$	second and fourth moments of Ψ with respect to $\mathbf{p}$
$\mathbf{E}$	rate-of-strain tensor
$\mathcal{M}, \mathcal{M}_f, \mathcal{M}_b$	total motor density, free motor density, and bound motor density per microtubule
r_c	motor protein capture radius
$\mathbf{r}_0$	motor protein tail position
z_0	height above bottom of assay at which microtubules and bound motor protein heads are centered
$B_{r_c}(\mathbf{r}_0)$	allowable configurations $\mathbf{x}, \mathbf{p}, s$ of bound motors in a ball of radius r_c around $\mathbf{r}_0$
$D_{r_c'}(\mathbf{r}_0')$	dimensionless disk of radius r_c' centered at $\mathbf{r}_0'$
$k_{\mathrm{on}}, k_{\mathrm{off}}$	motor protein binding and unbinding rates
C	simplified motor model binding coefficient
δ, δ_a	Dirac delta and smooth Dirac delta function
$\mathbf{f}$	bound motor force on fluid
$\mathbf{F}$	force associated with a single motor
F_{st}	motor protein stall force magnitude
L, H, U, P_0, T	macroscopic characteristic scales: assay length and height, fluid velocity, pressure, and time
ε	ratio between assay height and length ($H = \varepsilon L$)
l, V_m, τ	microscopic characteristic scales: microtubule half-length, motor stepping speed, and motor evolution time

Table 3.2: Model parameters and variables.

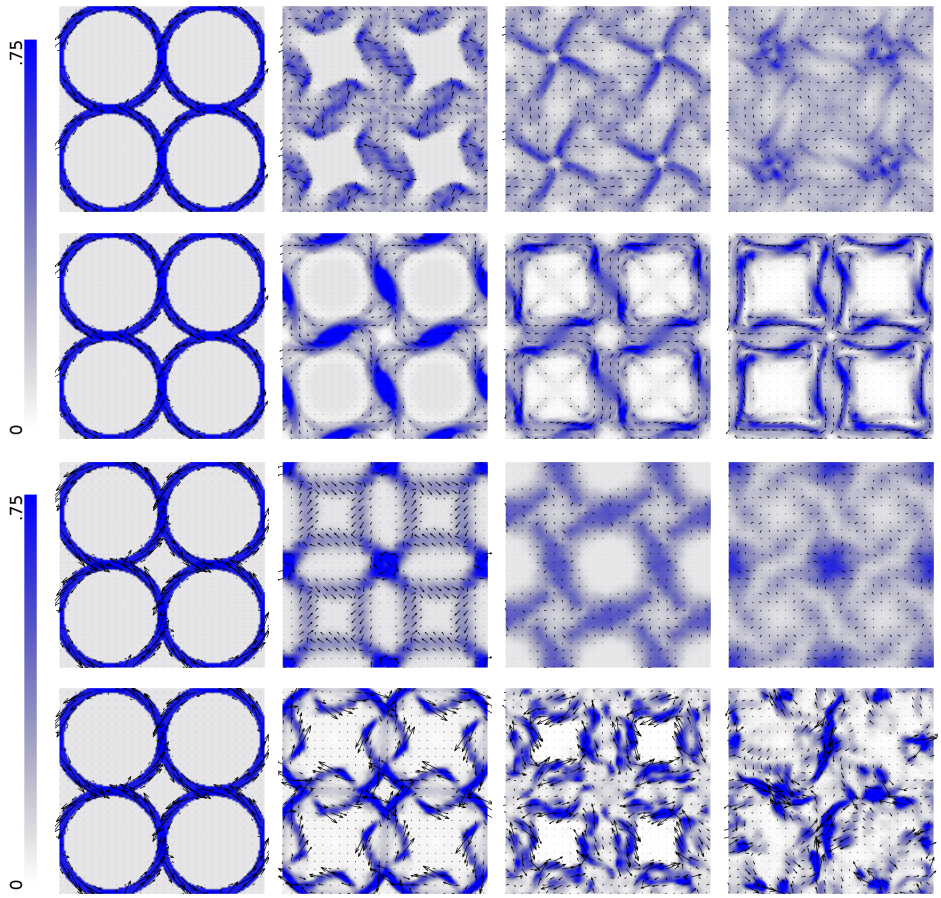


Figure 3.11: Vortex ring experiment with added self-propulsion velocity. Images are illustrative of behavior and thus are not necessarily taken at the same simulation time between rows. First row: $U_r = 0$ and $V_{\text{sp}} = .1$. Second row: $U_r = .01$ and $V_{\text{sp}} = .1$. Third row: $U_r = 0$ and $V_{\text{sp}} = 1$. Fourth row: $U_r = .01$ and $V_{\text{sp}} = 1$. Images are at times t=0, 225, 450, 675 sec in rows one and two, and at t=0, 60, 120, 180 sec in rows three and four. See second and third examples in supplemental movie 2.

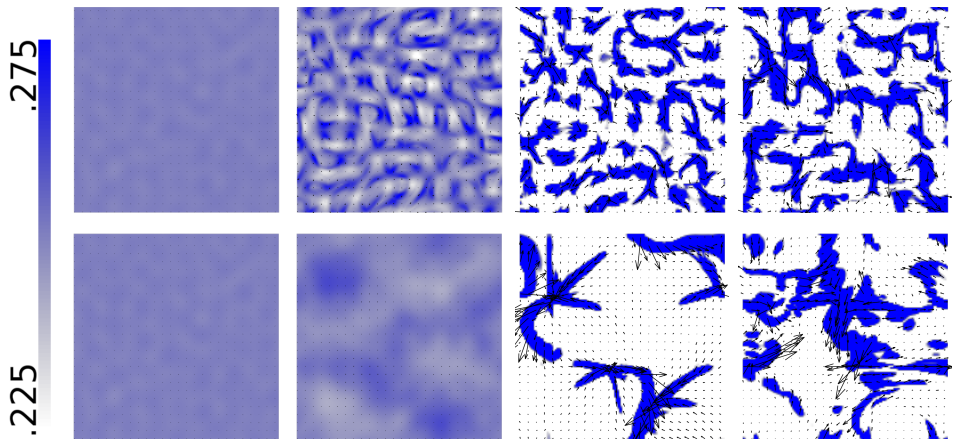


Figure 3.12: Perturbation experiment with added self-propulsion velocity. First row: U_r = .01 and V_{sp} = .1. Second row: U_r = .01 and V_{sp} = 1. Images are at times t=0, 105, 210, 315 sec. See second and third examples in supplemental movie 3.

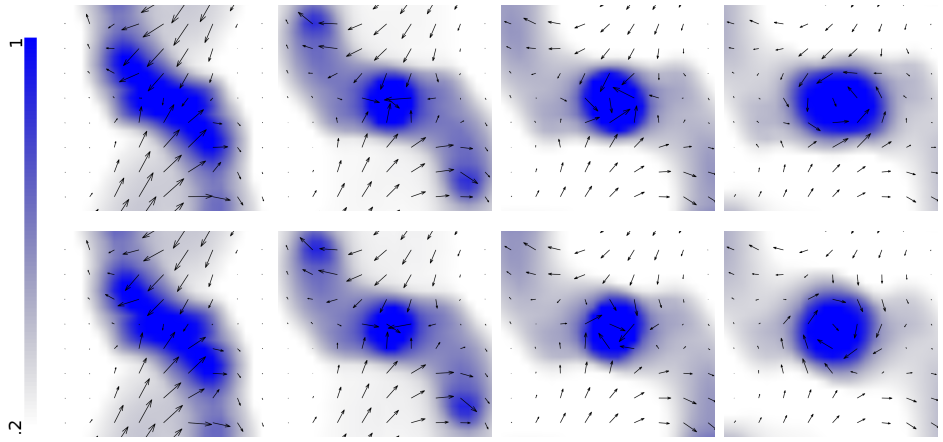


Figure 3.13: Magnified view of vortex ring experiment with evolved motor model (top row) and simplified motor model (bottom row), $U_r = 0$ and $V_{\text{sp}} = 0$. A counter-clockwise velocity field forms with the evolved motor model, whereas a clockwise velocity field forms with the simplified motor model. Images are at times t=150, 300, 450, 600 sec.

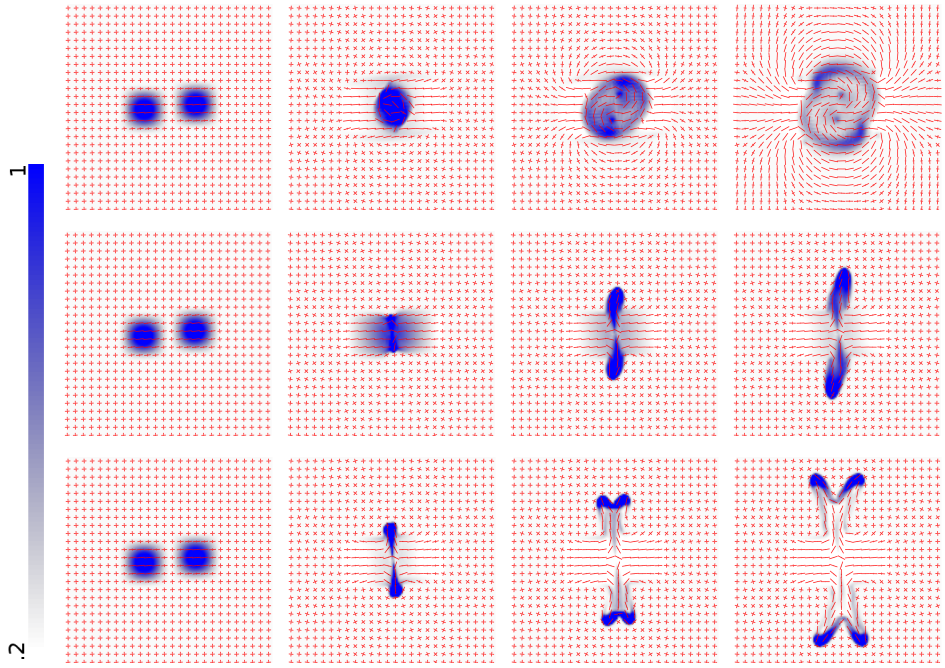


Figure 3.14: Antiparallel colliding clump experiment with simplified motor model, $U_r = .001$ and $V_{\text{sp}} = 0$. First row: $C = 10$. Second row: $C = 50$. Third row: $C = 250$. First row images are at times t=0, 300, 600, 900 sec. Second row images are at times t=0, 21, 42, 63 sec. Third row images are at times t=0, 15, 30, 45 sec. See supplemental movie 4.

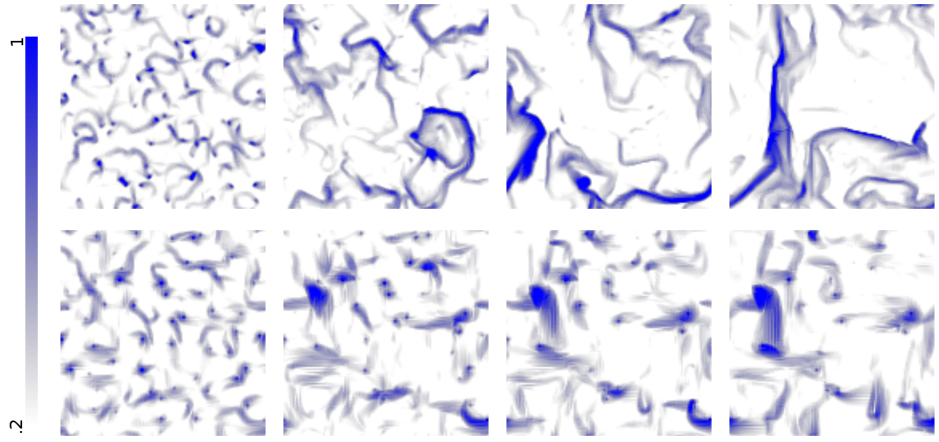


Figure 3.15: Perturbation experiment with (first row) and without (second row) motor-based fluid forces, showing faster formation of larger-scale features in the former case. First row images are at times t=150, 300, 450, 600 sec. Second row images are at times t=300, 600, 900, 1200 sec. See supplemental movie 5.

Chapter 4

Enabling Simulation of High-Dimensional Micro-Macro Biophysical Models through Hybrid CPU and Multi-GPU Parallelism

Micro-macro models provide a powerful tool to study the relationship between microscale mechanisms and emergent macroscopic behavior. However, the detailed microscopic modeling may require tracking and evolving a high-dimensional configuration space at high computational cost. In this work, we present a parallel algorithm for simulation a high-dimensional micro-macro model of a gliding motility assay. We utilize a holistic

approach aligning the data residency and simulation scales with the hybrid CPU and multi-GPU hardware. With a combination of algorithmic modifications, GPU optimizations, and scaling to multiple GPUs, we achieve speedup factors of up to 27 over our previous hybrid CPU-GPU implementation and up to 540 over our single-threaded implementation. This approach enables micro-macro simulations of higher complexity and resolution than would otherwise be feasible.

4.1 Introduction

Active gels exhibit macroscopic flow structures driven by the detailed microscopic interactions of constituent elements. Pronuclear centering and migration and cytoplasmic streaming are two such examples, both being critical cellular processes driven by filament-motor mixtures. Reduced-component studies have found these systems to be highly sensitive to the microscopic interactions between motors and filaments; for instance, the detachment time of a motor protein at a filament end affects whether filaments form networks of asters or vortices [46, 50, 48]. Additionally, the tens-of-nanometers sized motor proteins bind, walk along, and detach from micrometer-length filaments on a faster timescale than the filament network evolution. Simulating even a millimeter-sized system with such disparate length and time scales and sensitivity to detailed interactions thus poses a challenging computational problem. Tracking interacting Lagrangian particles can become infeasible with large quantities of microstructural elements.

A promising approach lies in micro-macro methods, which couple a kinetic theory model of the microstructure (here, the configuration of the motors and filaments in the

active gel) to the macroscale continuum mechanical representation of a viscoelastic fluid [31]. Kinetic theory models have been applied in the study of biological active matter [25], self-propelled particles [56], and networks of neurons [7]. They enable detailed microscale modeling that would otherwise be lost via closure approximations in macroscopic modeling approaches, and are particularly useful at scales where tracking individual particles and their interactions would be prohibitive. Compared to purely macroscopic methods, micro-macro methods are more computationally demanding, as they require evolving the microstructure density in a potentially high-dimensional configuration space.

[24] and [8] developed a micro-macro model for a gliding motility assay, consisting of immersed rigid filaments that glide along motor proteins anchored to the substrate of a chamber immersed in viscous fluid. This model includes hydrodynamic and steric interactions between the filaments. A high-dimensional kinetic theory describes the evolution of the filaments and motors. To make this model computationally feasible, parts of the microscale computation were ported to the GPU using Nvidia's CUDA C language [52]. In this work, we enable faster and significantly more detailed simulations through holistic restructuring of this algorithm, aligning the computation and data flow with the underlying heterogenous computational resources. Moreover, these changes facilitate scaling to multiple GPUs across separate machines with MPI. We further utilize a variety of CPU and GPU optimizations. Our work expands the range of micro-macro models which can be simulated by direct solution of the kinetic theory and coupling equations to models with higher dimensional configuration spaces, at higher resolutions [31]. To our knowledge, [24] is the first GPU-accelerated micro-macro kinetic theory-based simulation. [24] achieved up

to 20x speedups over a single-threaded CPU implementation, while the algorithm presented here achieves a further 27x speedup over [24] and [8]. Key to our approach is moving the microscopic scale and related tasks, which are smaller scale in both space and time, to the GPU and limiting CPU-GPU communications to the longer timescale of the filaments and fluid. Such holistic approaches are recommended to achieve scalability in heterogeneous environments [51], [20]. We note that our method does not suffer from common GPU simulation challenges encountered in various other approaches such as building adjacency lists [29], reordering storage based on cell location [70], dynamic, irregular data accesses [71], thread divergence [17], or neighbor exchanges of halo regions.

The paper is organized as follows. A description of the model and implementation is presented in Section 4.2, algorithmic, data flow, GPU, and MPI modifications are discussed in Section 4.3, results are discussed in Section 4.4, and we conclude in Section 4.5.

4.2 Motility Assay model

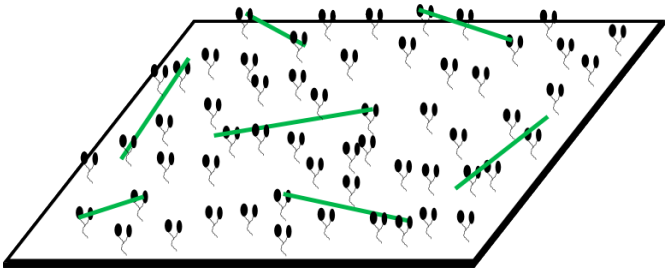


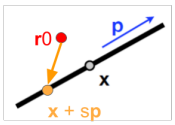
Figure 4.1: Gliding motility assay. Motor proteins (black) anchored to the substrate bind to filaments (green), walk along them and exert forces, then detach.

Figure 4.1 illustrates a gliding motility assay. The geometry consists of top and

bottom plates separated by a narrow vertical gap, which is filled with a viscous fluid containing ATP fuel. At the bottom plate, motor protein tails are anchored to a substrate. The motor protein heads diffuse in solution, tethered to their tail by a flexible stalk. When a filament enters the capture radius of a motor protein head, the head may bind to the filament. As the bound motor head walks toward the filament plus end, it exerts force, causing the filament to glide in the opposite direction, until the motor head detaches. When many filaments are present, the underlying microscopic mechanism coupled with hydrodynamic and steric interactions give rise to a variety of emergent macroscopic behaviors such as a lattice of vortices [65]. Through our modeling and simulation, we aim to better understand the relationship between the microscale interactions and the macroscopic phenomena. Related problems of emergent self-organization from simple interactions include flocking and swarming of birds, fish, and bacteria.

The model equations are presented in simplified, nondimensionalized form in Table 4.1. The filament density is parameterized by center-of-mass location $\mathbf{x}$, orientation $\mathbf{p}$, and time t as $\Psi(\mathbf{x}, \mathbf{p}, t)$.

The configuration space of bound motors is higher dimensional, as we need to track the center-of-mass position $\mathbf{x}$ and orientation $\mathbf{p}$ of the filament a bound motor with tail anchored at $\mathbf{r}_0$ is bound to, along with its arclength parameter along that filament s. This yields the high-dimensional density $\mathcal{M}_b(\mathbf{r}_0, \mathbf{x}, \mathbf{p}, s, t)$ of bound motors per filament, illustrated in the figure to the right. A key observation is that motors with tail anchored at position $\mathbf{r}_0$ on the assay substrate can only bind to filament sections that are within the capture radius of the motor



stalk r_c. This greatly reduces the feasible configurations a motor protein head may be bound in, and obviates the need to track configurations $|\mathbf{x} + s\mathbf{p} - \mathbf{r}_0| > r_c$. We denote by $B_{rc}(\mathbf{r}_0)$ all feasible $\mathbf{x}, \mathbf{p}, s$ configurations such that $|\mathbf{x} + s\mathbf{p} - \mathbf{r}_0| \leq r_c$. We do not track unbound (free) motor heads, only their tail position $\mathbf{r}_0$, so the density of free motors $\mathcal{M}_f(\mathbf{r}_0, t)$ is two-dimensional. We model the filament and bound motor protein densities as distributed by a smooth Dirac delta function in z about a plane a small distance z_0 above the bottom plate, i.e., $\Psi(\mathbf{x}, \ldots) = \Psi_{z_0}(\mathbf{x}_2, \ldots)\delta(z)$ and $\mathcal{M}_b(\mathbf{r}_0, \mathbf{x}, \cdots) = \mathcal{M}_{b,z_0}(\mathbf{r}_0, \mathbf{x}_2, \cdots)\delta(z)$. We thus evolve the lower-dimensional $\mathcal{M}_{b,z_0}$ and Ψ_{z_0} in our simulation. We drop the z_0 from Ψ and $\mathcal{M}_b$ in the remainder of the paper for brevity.

We represent the fluid velocity in three dimensions, with periodic boundary conditions in the x and y dimensions and no-slip conditions in the z dimension at the top and bottom plates. The system evolves on two timescales; the motors bind to, walk along, and unbind from the filaments on a faster timescale than the filaments and fluid evolve. Bound motor heads generate forces that are spread onto the fluid in an immersed boundary method fashion [53]. Together with stress terms arising from filament inextensibility and steric interactions [15], the motor forces (Eq. (4.10)) couple the densities Ψ and $\mathcal{M}_b$ to the fluid velocity (Eq. (4.7)). The x and y dimensions are discretized over a regular square grid, and the z dimension is discretized over an adaptive grid that is finely resolved around $z = z_0$ near the bottom plate and becomes coarser moving toward the upper plate. Allowable filament orientations are constrained to the (x, y) plane, so we can represent $\mathbf{p} = (\cos\theta, \sin\theta, 0)^T$. Orientation θ and arclength parameter s are discretized uniformly with the same resolution.

Algorithm 2 summarizes the process for evolving the filaments, motor proteins,

and fluid velocity as in [24], [8]. First, we compute the adaptive time steps based on their stability conditions, with outer time step dt restricted by the advective fluxes in Eq. (4.1), and the inner time step dt^* restricted by the motor speed and binding/unbinding rates in Eq. (4.4). Next, $\Psi(t + dt)$ is solved using Crank-Nicolson for the diffusive terms and Adams-Bashforth 2 with upwinding for the advective terms (line 4). The bound motor density evolution routine in lines 7-9 performs the motor protein advection (bound motor heads walking along the filaments toward their plus-ends) and applies a Superbee flux limiter, as well as simulates the binding and unbinding of free and bound motor proteins respectively. After every configuration of $\mathcal{M}_b(\mathbf{r}_0\mathbf{x}, \theta, s, t)$ for a particular $\mathbf{r}_0$ has been updated, the integral $\mathcal{M}_{b,coarse}(\mathbf{r}_0, t) = \iiint \mathcal{M}_b(\mathbf{r}0, \mathbf{x}, \theta, s, t)\Psi(\mathbf{x}, \theta, t)d\mathbf{x}d\theta ds$ is calculated at the same $\mathbf{r}_0$ to ensure that the number of bound motors does not exceed the total number of motors at $\mathbf{r}_0$. If so, all $\mathcal{M}_b$ configurations with that $\mathbf{r}_0$ are scaled down to conserve the total number of motors before $\mathcal{M}_f$ is calculated. We next update the free motor density in line 9. The extra stress terms $\boldsymbol{\sigma}^{\mathrm{f}}, \boldsymbol{\sigma}^{\mathrm{t}}$ arising from filament inextensibility and steric interactions [8] are computed as moments of Ψ in line 13. We perform a two-dimensional FFT and solve the transformed system of fluid equations for the three velocity components and pressure $\hat{u}, \hat{v}, \hat{w}, \hat{q}$ at every position on the z grid, then perform an inverse FFT to obtain the three-dimensional fluid velocity $\mathbf{u}$ in line 13.

In a single-threaded implementation, the high dimensionality of $\mathcal{M}_b$ makes the computations in lines 7-9 and line 11 prohibitively expensive for even moderate grid resolutions and experiment times. Fortunately, $\mathcal{M}_b$ can be computed explicitly and easily parallelized over $\mathbf{r}_0$. Thus in [24, 8], the $\mathcal{M}_b$ and $\mathbf{F}$ computations are ported to the GPU.

Algorithm 2 Evolution scheme for the coupled microtubule density, motor protein density, and fluid velocity equations.

Initialize Ψ and $\mathcal{M}_b$
while $t < t_{end}$ **do**
    Compute adaptive time steps dt, dt^*
    Compute filament density $\Psi(t + dt)$ (Eqs. (4.1)-(4.3))
    set $t^*_{\text{end}} = t + dt$
    while $t^* < t^*_{end}$ **do**
        Compute bound motor density $\mathcal{M}_b(t^* + dt^*)$ (Eq. (4.4))
        Update coarsened density $M_{b,coarse}$ (Eq. (4.5))
        Update free motor density $\mathcal{M}_f$ (Eq. (4.6))
    end
    Calculate motor force $\mathbf{F}_2$ (Eq. (4.10))
    Calculate extra stresses $\boldsymbol{\sigma}^{\text{f}}, \boldsymbol{\sigma}^{\text{t}}$ (Eq. (4.7))
    Calculate fluid velocity $\mathbf{u}$ (Eqs. (4.7)-(4.9))
end

On the other hand, the Ψ equation is stiff due to the diffusion terms and is computed semi-implicitly on the CPU. This decomposition of work is similar to several hybrid reactive flow solvers [51].

To perform the integral in line 11, [24, 8] use independent GPU threads to compute the integral at each $\mathbf{r}_0$ accumulating the partial results in thread-local storage, limiting the use of atomic operations to the final reduction over nearby $\mathbf{r}_0$ at each $\mathbf{x}$. Extra stress tensor calculation, forward and reverse fast Fourier transforms, and computation of the independent semispectral systems are all multithreaded on the CPU. In this work, we expand upon this hybrid computational approach as described below.

4.3 Acceleration Methodology

Our primary focus in this work is significant performance improvement through targeted algorithmic design enabling a multi-GPU decomposition, as well as single-GPU optimizations, described in detail below. Through these efforts we are able to scale to higher resolutions than previously possible and achieve up to 27x total simulation acceleration in a four GPU configuration. Use of additional GPUs is supported and should provide further speedup with similarly excellent scaling, although we did not test this in this work.

4.3.1 Holistic Algorithmic and Data Flow Restructuring

A primary goal of our approach is to restructure the algorithm so that the memory-intensive microscale motor protein data resides solely on the GPU, and only the smaller, coarsened data is transferred to/from main memory. A secondary goal is to support a multi-GPU decomposition. Additionally, we remove synchronization barriers and reduce GPU memory consumption by two-thirds. Figure 4.2 summarizes the changes and details follow.

Independent Time Steps. The original algorithm calculated a global dt^* and updated all $\mathcal{M}_b$ configurations by this fixed time step to time $t + dt$, hindering performance in several ways. First, it artificially limits the inherently independent per-cell update operations, some of which may be able to complete in fewer steps as their local configuration and stability restrictions allow. Second, it requires an expensive reduction operation over the entire bound motor density microstructure. Third, if $\mathcal{M}_b$ is distributed over multiple GPUs as desired, the reduction creates an unnecessary synchronization barrier. We instead com-

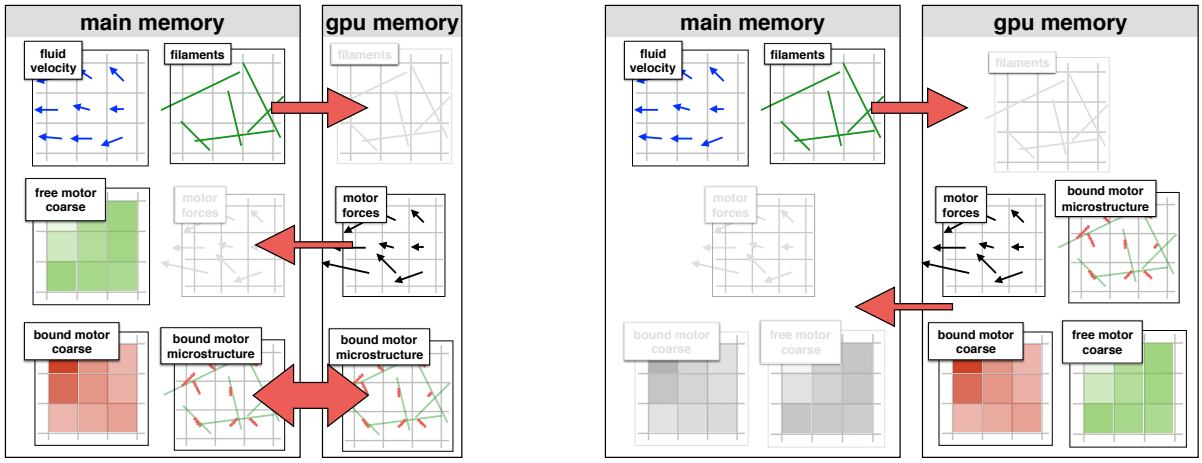


Figure 4.2: Data residency and computation before (left) and after (right) restructuring. Left, transfer of the microstructure creates significant communication overhead. Right, the bound motor microstructure now fully resides on the GPU eliminating expensive transfers, and the coarse bound and free motor densities are calculated on GPU and transferred back to main memory.

pute a local dt^* for each $\mathbf{r}_0$ at the beginning of each inner time step and update $\mathcal{M}_b$ at each $\mathbf{r}_0$ asynchronously. The most significant benefit of this change is enabling the multi-GPU implementation. Stability and accuracy were not adversely affected.

Numerical Integration Scheme. The algorithm in [8] used Adams-Bashforth 2 for time integration of the motor densities, whhich maintains the $\mathcal{M}_b$ array at three distinct time points (t^{n+1}, t^n, t^{n-1}). We instead use Runge-Kutta 2, which only requires the $\mathcal{M}_b$ array at t^{n+1} and t^n. This change reduces GPU memory requirements by one-third while causing negligible impact on computation time. With these improvements, higher-resolution $\mathcal{M}_b$ density representations may reside in scarce GPU memory.

Mixed Precision. We developed a mixed precision approach whereby we store and update $\mathcal{M}_b$ and $\mathcal{M}_f$ in single precision floating point while keeping the rest of the simulation as double precision. This saves space and improves performance without causing appreciable

change in simulation behavior.

Data Residency. The algorithm in [8] updated $\mathcal{M}_b$ one piece at a time due to GPU memory constraints, then transferred the complete updated $\mathcal{M}_b$ to the GPU for the motor force calculation. With the new numerical integration scheme and the use of mixed precision, we have enough GPU memory to store the high-dimensional microstructure data $\mathcal{M}_b$ solely on the GPU. This eliminates the overhead of transferring copies of $\mathcal{M}_b$ before, during, and after the $\mathcal{M}_b$ update. Since $\mathcal{M}_b$ is required in order to calculate $\mathcal{M}_{b,coarse}$ and $\mathbf{F}$, we also do those calculations on the GPU, and transfer results to the CPU. $\mathcal{M}_{b,coarse}$ and $\mathbf{F}$ are both macroscale data structures, and hence incur lower communication overhead. Finally, Ψ, which is also stored on the macroscale, is transferred as before. The updated data flow is shown in Figure 4.3.

Pseudocode describing the new GPU kernel is presented in Algorithm 3. The result is one large kernel that fully updates $\mathcal{M}_b$ at each independent $\mathbf{r}_0$ value to $t + dt$ in as many steps as needed, using a local adaptive time step. The new memory access pattern is more amenable to caching as well, as each running block of threads on each GPU reads the same contiguous memory for all $\mathcal{M}_b$ configurations at a fixed $\mathbf{r}_0$ location repeatedly until those configurations are fully updated before moving on.

We use CPU parallelism via OpenMP to further accelerate the simulation, specifically in the calculations of the fluxes and stress tensors for the fluid solves, the outer global time step dt calculation, and construction of the Ψ advection matrix. After moving the dt^* calculation into the motor force update kernel and multithreading the dt calculation, time step calculation becomes a negligible component of the total computation time.

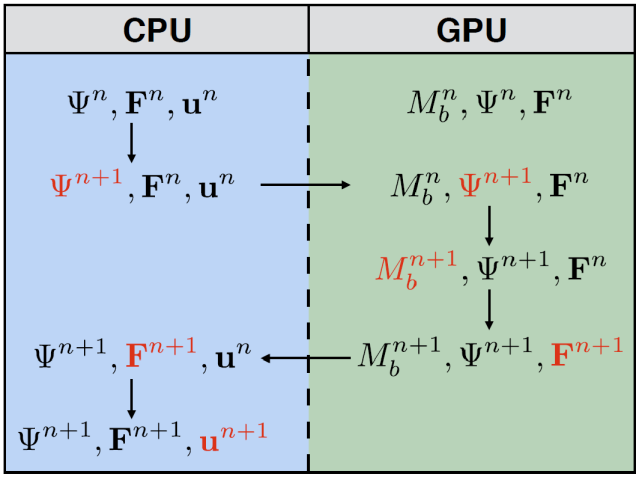


Figure 4.3: Residency and evolution of state from time step n to time step $n+1$ on CPU and GPUs in new algorithm. Red indicates the quantity updated through computation or data transfer.

4.3.2 GPU Optimizations

This section describes various optimizations of the GPU kernel shown in Algorithm 3. Combined, these single-GPU optimizations yield an average improvement of 4.7-7.5x depending on resolution. The optimizations are described below and the individual effect of each is listed in Table 4.5.

Mixed precision. As previously detailed, switching $\mathcal{M}_b$, $\mathcal{M}_{b,coarse}$, and $\mathcal{M}_f$ to single precision halves the GPU memory requirement. In addition, it provides a 4.3x to 5.8x speedup in our bound motor density evolution routine. This improvement will depend on the clock cycle ratio between single and double precision arithmetic for a given GPU family.

Fast Math. Compiling with CUDA's fast math library provides additional savings without noticeable change in simulation behavior. Accelerations of 1.35x were typical.

Launch Bounds. The *launch_bounds* macro in CUDA may be used to instruct the

compiler to ensure a user-specified maximum number of concurrent threads and threads per block running on each GPU Streaming Multiprocessor (SM). Using a *launch_bounds* configuration of 128 threads/block and 8 simultaneous blocks per SM gives the best performance of all configurations tested. Register spilling to global memory does occur at this configuration as each thread is limited to 32 registers. Newer architectures with more registers per SM will likely see immediate improvement by both reducing register spilling and enabling more threads per block. Accelerations of 1.2x were typical at the higher inner resolution and negligible at the lower inner resolution.

Dimension Mapping. CUDA threads are executed in simultaneous warps of 32 threads each, grouped first by their x-index then by their y-index. Since coalesced memory accesses are desirable for performance, the bound motor density evolution kernel was modified so that a thread's x-index maps to the s-index and the y-index maps to the θ-index. With this mapping threads executing in a warp will access $\mathcal{M}_b$ storage in a coalesced fashion since sequential s-indices are contiguous as the innermost array indices. Accelerations of 1.16x to 1.57x were observed.

Reordering Storage. The Superbee flux limiter operates in the arclength s dimension. Since threads in a warp operate on subsequent arclength indices, and the flux limiter has a neighborhood access pattern of $(s-2, s-1, s, s+1, s+2)$, this gives coalesced memory accesses and pulls adjacent arclength data into the cache for subsequent iterations. The layout of memory in $\mathcal{M}_b$ was modified to make s the innermost variable instead of θ in the storage of $\mathcal{M}_b(\mathbf{r}_0, \mathbf{x}, \theta, s)$, where $\mathbf{x}, \theta, s$ are represented as sequential flat four-dimensional arrays within a flat two-dimensional array over $\mathbf{r}_0$. This prevents strides between subsequent

s accesses. Another benefit to making s the innermost variable is that the value of $\Psi(\mathbf{x}, \theta)$ can be read after the θ loop instead of in the innermost loop. Reordering the loops in this fashion in the access-heavy motor force code resulted in a 1.54x acceleration. For the bound motor density update, accelerations of 1.1x were typical at the higher inner resolution and negligible at the lower inner resolution.

Unrolling Reductions. For the reduction step, [23] recommends manually unrolling a reduction when the number of remaining threads is less than the warp size (32 for our Tesla M2075), and performing part of a large reduction independently within each thread to reduce synchronization. We already follow the latter suggestion as each thread accumulates its contribution to $\mathcal{M}_{b,coarse}$ before storing this running sum in a shared memory array sized to the number of threads for the reduction step. We did not find meaningful performance improvements for the manual reduction.

Block Shaping. To update $\mathcal{M}_b(\mathbf{r}_0, \mathbf{x}, \theta, s)$ we assign one block of threads to each $\mathbf{r}_0$ position and map those threads to the $\mathbf{x}, \theta, s$ variables. For each $\mathbf{x}$, a two-dimensional block of threads is launched, with the threads' x and y indices corresponding to the innermost s and θ indices. Experimentation has shown 128 threads to be the optimal number in our implementation. The "block shaping" row of the optimizations table compares against running 256 threads in a 16x16 configuration. How these 128 threads are configured is important: $x = 8, y = 16$ runs faster than $x = 16, y = 8$. Both caching effects and memory coalescing play a role, and from our experience it is worthwhile to experiment with various configurations. Accelerations of 1.15x were typical at the higher inner resolution and accelerations of 1.65x were typical at the lower inner resolution.

Algorithm 3 Bound & free motor density update GPU kernel

Precondition: $\Psi(t + dt)$ and dt are loaded into GPU memory.
Set $t^*_{\text{end}} = t + dt$
for $\mathbf{r}_0 \in grid$ **do**
  while $t^* < t^*_{end}$ **do**
    Compute adaptive dt^*
    for $\mathbf{x} \in B_{rc}(\mathbf{r}_0)$ **do**
      for $\theta \in B_{rc}(\mathbf{r}_0)$ **do**
        for $s \in B_{rc}(\mathbf{r}_0)$ **do**
          Compute $\mathcal{M}_b(\mathbf{r}_0, \mathbf{x}, \theta, s, t^* + \frac{dt^*}{2})$ (Eq. (4.4))
          Compute $\mathcal{M}_b(\mathbf{r}_0, \mathbf{x}, \theta, s, t^* + dt^*)$ (Eq. (4.4))
        end
      end
    end
    Compute $\mathcal{M}_{\text{b,coarse}}(\mathbf{r}_0, t^* + dt^*)$ (Reduction) (Eq. (4.5))
    if $\mathcal{M}_{total}(\mathbf{r}_0) < \mathcal{M}_{b,coarse}(\mathbf{r}_0, t^* + dt^*)$ **then**
      for $\mathbf{x}, \theta, s \in B_{rc}(\mathbf{r}_0)$ **do**
        Scale $\mathcal{M}_b(\mathbf{r}_0, \mathbf{x}, \theta, s, t^* + dt^*)$
        Set $\mathcal{M}_{\text{b,coarse}}(\mathbf{r}_0, t^* + dt^*) = \mathcal{M}_{\text{total}}(\mathbf{r}_0)$
      end
    end
    Compute $\mathcal{M}_f(\mathbf{r}_0, t^* + dt^*)$ (Eq. (4.6))
  end
end

To analyze performance of the $\mathcal{M}_b$ and $\mathbf{F}$ GPU kernels, the Nvidia Visual Profiler v8.0 [9] was used. According to its output, arithmetic operations constitute the largest share of operations. No functional unit (load/store, arithmetic, control flow) is a bottleneck because of the balance of operations. We run the maximum possible number of simultaneous blocks per SM (8), but cannot run more threads per block without exhausting the available registersper SM. The result is a GPU occupancy of 66%, for which the profiler's heuristics report that increasing occupancy is unlikely to improve execution time. Our experience confirms this, as attempts to launch more threads per block to increase occupancy means decreasing registers per thread to keep the simultaneous blocks per SM maximized at 8,

resulting in longer execution times. GPU occupancy is one of many factors that contributes to kernel performance, and it is possible to obtain high throughput at low occupancy levels [68].

4.3.3 Scaling to Multiple GPUs

As the spatial resolution of the $\mathbf{r}_0$ grid increases, two factors limit the performance of a single GPU. The first is that the number of blocks (each updating an independent $\mathbf{r}_0$) that can run concurrently on the GPU is limited by the number of SMs on the card, as we are running the maximum 8 simultaneous blocks per SM. Using two equivalent cards simultaneously doubles the throughput at which we can update the motor densities and calculate the motor force. The second factor is that once GPU memory is exhausted by the bound motor density (and scratch space for the intermediate values needed for numerical routines), additional large memory transfers to and from main memory become necessary every time step.

We expand our implementation to multiple GPUs using simultaneous CUDA streams and to multiple machines using MPI. After $\mathcal{M}_b$ is initialized at the beginning of the simulation, subsections of $\mathcal{M}_b$ are transferred to the memory of each GPU. The outer two-dimensional array of $\mathcal{M}_b$ (over $\mathbf{r}_0$) is distributed among available GPUs by rows, which are contiguous in memory. Before invocation of the $\mathcal{M}_b$ and $\mathbf{F}$ kernels on the GPUs, the newly updated Ψ is broadcast to each with an MPI_BCAST from process rank 0. After the bound motor density update and motor force kernels complete, The $\mathcal{M}_f$ and $\mathcal{M}_{b,coarse}$ values computed by each GPU are collected by process rank 0 using an MPI_GATHER operation. The motor force output $\mathbf{F}$ from each GPU contains overlapping force vectors

that need to be summed together, so an MPI_REDUCE operation is used to combine them in process rank 0. From here on the fluid velocity update proceeds as normal. The process is summarized in Figure 4.4. OpenMPI 2.1 was used for this work.

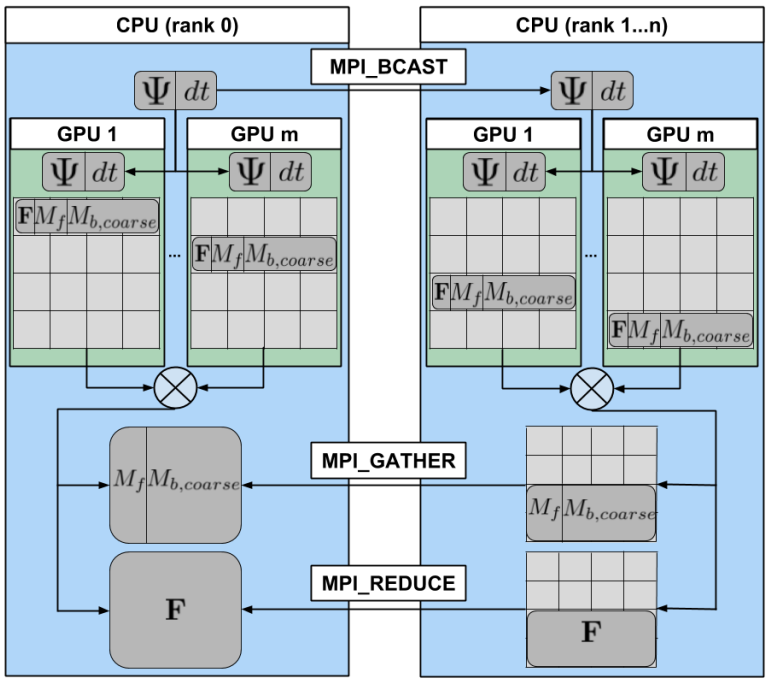


Figure 4.4: MPI control flow for multiple GPUs across multiple nodes. We acheive nearly ideal scaling of our GPU computation across multiple GPUs, indicating that the MPI overhead is negligible.

4.4 Results

We present results at different resolutions, scaling both the outer resolution of the x, y variables and the inner resolution of the θ, s variables. Increasing the outer resolution (x, y) affects the two-dimensional grid over which $\Psi, \mathcal{M}_f, \mathcal{M}_b$, and $\mathbf{u}$ are defined, and thus increases the workload across all steps of the simulation. Doubling the resolution of x and

y increases by a factor of four the total workload of the $\mathcal{M}_b$ and $\mathbf{F}$ kernels. Increasing the inner resolution (θ, s) affects Ψ and $\mathcal{M}_b$ through their dependence on θ and $\mathcal{M}_b$ through its dependence on s. Doubling the resolution of θ and s increases by a factor of four the number of update tasks per thread in the motor density and force kernels, and adds two more iterations to the reduction step in the motor density kernel.

Simulations were run on one or more servers configured with 2 Tesla M2075 GPUs, 64 GB RAM, and dual AMD Opteron 6272 processors. Speedup factors for the algorithmic modifications and GPU optimizations vs. the original implementation in a single-node, single-GPU configuration are shown in Table 4.2. An average performance increase of between 5.75x and 9.98x per full simulation step is observed versus the original implementation. The bulk of the improvement comes in the bound motor update and motor force computations. Additionally, the dt computation is accelerated, the dt^* computation is moved onto the GPU, and the Ψ solve noticeably benefits from CPU acceleration.

We individually disable each GPU optimization and compare the running time for a single invocation of the $\mathcal{M}_b$ update kernel in Table 4.5. We see the largest performance improvement from switching from double precision to single precision, which affects both floating point arithmetic performance as well as cache and memory demand. The $256^2 \times 32^2$ resolution could not be tested with double precision on a single GPU as $\mathcal{M}_b$ exceeded GPU memory. The launch_bounds and storage reordering optimizations see their biggest impact when the inner variable resolution is increased. When this optimization was originally applied it showed a small improvement, but when it alone is removed from the final implementation, no discernable impact is observed.

Figure 4.6 shows that scaling the $\mathcal{M}_b$ and $\mathbf{F}$ kernels to multiple GPUs and across nodes is effective, with nearly ideal linear acceleration at the higher resolutions where acceleration is most needed. This demonstrates that the overhead to merge output between GPUs on the same machine plus the MPI overhead among multiple machines is small compared to execution time. As the inter-node communication consists of MPI broadcast, reduction, and gather operations, it is expected that scaling to 8 or more GPUs would likewise involve minimal overhead cost.

Table 4.3 summarizes overall performance of our optimized implementation using one, two, and four GPUs vs. the original single-GPU only implementation. We obtain higher accelerations for higher inner resolution sizes, which is desirable as we find the inner resolution of 16^2 too coarse at outer resolutions over 64^2. Our maximum speedup factor over the original implementation was over 27x, obtained at the highest resolution. The simulation was previously limited to the $128^2 \times 32^2$ configuration given the running times involved. Reducing a day's worth of computation to less than one hour greatly facilitates the iterative exploration of the model's parameter space. Sample simulation results at the previously infeasible $256^2 \times 32^2$ resolution are shown in Figure 4.7.

4.5 Conclusions

Mapping the different scales of a simulation to different computational hardware, minimizing data transfers, and removing synchronization points like a global time step calculation allows us to explore the parameter space of our high-dimensional micro-macro

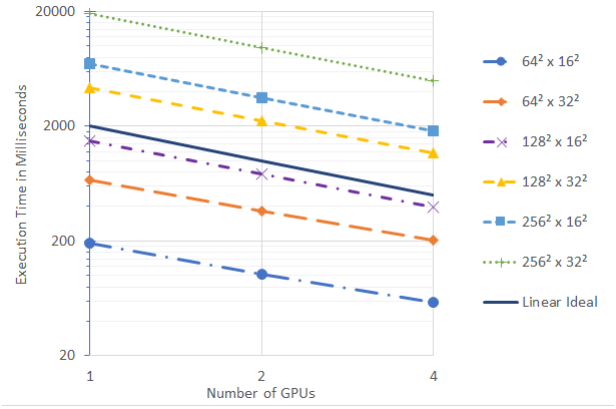


Figure 4.5: Multi-GPU scaling of the bound motor density and motor force computations is nearly ideal. The overhead of inter-node communication via MPI arising in the 4 GPU configuration does not have an appreciable effect.

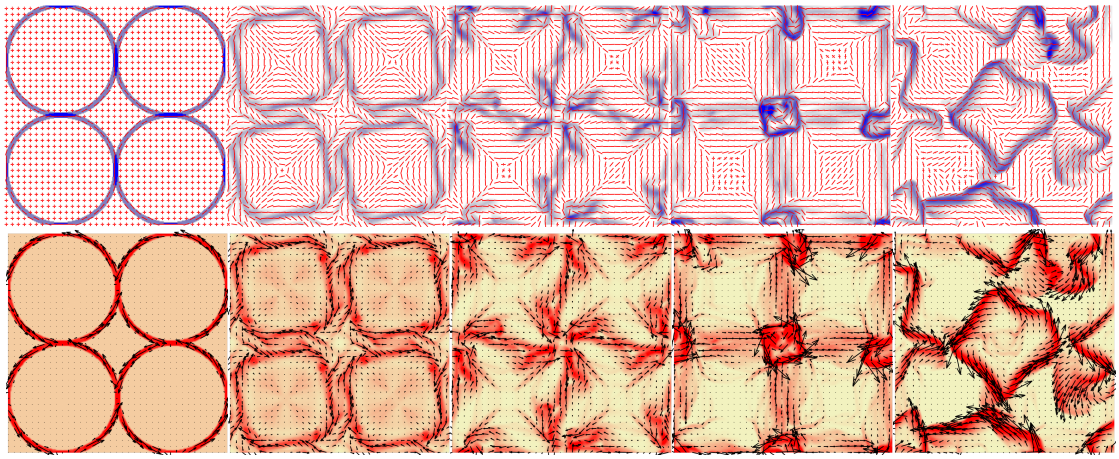


Figure 4.6: Sample simulation output at equally spaced timing intervals of the evolution of a lattice of overlapping filament rings from an overhead view of the assay, with periodic boundary conditions. Top row: filament density Ψ in blue with predominant orientation vector plotted every 8th cell in red. Bottom row: coarse bound motor density plotted in red with fluid velocity plotted as a black arrow every 8th cell.

simulation up to 540 times faster than a single-threaded implementation when using four GPUs simultaneously. This holistic approach significantly outperforms the commonly employed approach of accelerating individual functions in isolation [24]. Multi-node, multi-GPU overhead is minimal and the approach is expected to scale well to a greater number of GPU accelerators. This approach capitalizes on the increasing prevalence of GPUs in high performance computing.

As the number of GPUs used increases the semispectral fluid solve and filament evolution update steps will become the next bottlenecks. Further adjustment of simulation flow to offload more of the fluid solve computations onto the otherwise idle CPU cores of non-root processes may then become cost effective.

It is our hope that our algorithmic design and breakdown of the various CPU and GPU optimizations will provide a useful reference for prioritizing optimizations in HPC software development and in porting of existing applications, where there is often an expectation that porting time should be recovered by faster runtimes. While the specific improvement will vary for different programs, quantifying the improvements corresponding to various optimizations contributes to the growing information in the literature regarding their efficacy [51].

Filament equations

$$\partial_t \Psi + \nabla_2 \cdot (\dot{\mathbf{x}}_2 \Psi) + \partial_\theta (\dot{\theta} \Psi) = 0 \tag{4.1}$$

$$\dot{\mathbf{x}}_2 = -V_{\mathrm{sp}} \mathbf{p}_2 + \mathbf{u}_2 + U^0_{t,\parallel} \mathbf{p}_2 \mathbf{p}_2 : \nabla_2 \mathbf{D}_{2,z_0} - D_{t,\parallel} \nabla_2 \ln \Psi \tag{4.2}$$

$$\dot{\theta} = \nabla_2 \mathbf{u}_2 + U^0_r \mathbf{D}_{2,z_0} : \mathbf{p}_2^\perp \mathbf{p}_2 - D_r \partial_\theta \ln \Psi \tag{4.3}$$

Motor equations

$$\partial_t \mathcal{M}_b + \partial_s \mathcal{M}_b = -k_{\mathrm{off}} \mathcal{M}_b + k_{\mathrm{on}} \mathcal{M}_f \mathbb{1}_{D_{r_c}} \tag{4.4}$$

$$\mathcal{M}_{b,coarse} = \iiint \mathcal{M}_b \Psi \, ds \, d\mathbf{x}_2 \, d\theta \tag{4.5}$$

$$\mathcal{M}_f = \mathcal{M} - \mathcal{M}_{b,coarse} \tag{4.6}$$

Fluid equations

$$-\nabla_2^2 \mathbf{u}_2 - \frac{1}{\varepsilon^2} \partial_{zz} \mathbf{u}_2 + P_0 \nabla_2 q = \sigma_f \nabla_2 \cdot \boldsymbol{\sigma}^{\mathrm{f}} - \sigma_t \nabla_2 \cdot \boldsymbol{\sigma}^{\mathrm{t}} + \mathbf{F}_2 \tag{4.7}$$

$$-\nabla_2^2 w - \frac{1}{\varepsilon^2} \partial_{zz} w + P_0 \partial_z q = 0 \tag{4.8}$$

$$\nabla_2 \mathbf{u}_2 + \partial_z w = 0 \tag{4.9}$$

Motor force

$$\mathbf{F}_2 = F \iiiint \mathbf{p}_2 \delta(\mathbf{y}_2 + \frac{l}{L} s \mathbf{p}_2 - \mathbf{x}_2) \Psi \mathcal{M}_b \, ds \, d\mathbf{r}_0 \, d\mathbf{y}_2 \, d\theta \tag{4.10}$$

Table 4.1: Summary of model equations for the filament and motor protein densities, the macroscopic fluid equations, and the motor force that couples them.

res.	64^2x16^2	64^2x32^2	128^2x16^2	128^2x32^2	256^2x16^2	256^2x32^2
total	332(**7.16**)	854(**8.76**)	2054(**6.37**)	5026(**9.55**)	8980(**7.36**)	21408(**9.25**)
Ψ	24(**6.17**)	49(**5.37**)	88(**6.44**)	171(**7.88**)	389(**7.28**)	722(**9.19**)
$\mathcal{M}_b$&**F**	193(**6.26**)	682(**6.42**)	1492(**6.03**)	4366(**7.51**)	7050(**6.65**)	19125(**4.72**)
fluid	107(**2.50**)	112(**2.21**)	447(**1.63**)	456(**1.96**)	1437(**2.00**)	1438(**2.42**)

Table 4.2: Effect of single-node, single-GPU optimizations at various inner and outer resolutions. Average time per full outer time step (milliseconds) and speedup factors (bold) for original simulation vs. optimized simulation on a single machine with one GPU. Simulations with varied resolutions run to a fixed end time. The expensive dt^* calculation in the original simulation is now negligible, contributing to the increased overall speedup reported in the total time step row.

outer × inner resolution	$64^2 \times 16^2$	$64^2 \times 32^2$	$128^2 \times 16^2$	$128^2 \times 32^2$	$256^2 \times 16^2$	$256^2 \times 32^2$
optimization removed:						
none	350	710	1490	3560	5960	15660
mixed precision	1520 (**4.34**)	4140 (**5.83**)	7390 (**4.96**)	19190 (**5.39**)	30170 (**5.06**)	X (**X**)
-fast-math	470 (**1.34**)	870 (**1.23**)	2010 (**1.35**)	4530 (**1.27**)	8050 (**1.35**)	19710 (**1.26**)
launch_bounds	340 (**1.97**)	860 (**1.21**)	1490 (**1**)	4270 (**1.2**)	5890 (**99**)	18200 (**1.16**)
dimension mapping	440 (**1.26**)	1120 (**1.58**)	1820 (**1.22**)	5210 (**1.46**)	8200 (**1.38**)	22420 (**1.43**)
reorder storage	350 (**1**)	820 (**1.15**)	1490 (**1**)	3960 (**1.11**)	6010 (**1.01**)	16950 (**1.08**)
unroll reduction	370 (**1.06**)	700 (**0.99**)	1510 (**1.01**)	3630 (**1.02**)	5920 (**0.99**)	15350 (**0.98**)
block shaping	550 (**1.57**)	820 (**1.15**)	2460 (**1.65**)	4330 (**1.22**)	9800 (**1.64**)	18070 (**1.15**)

Table 4.3: GPU optimizations. Time (milliseconds) and slowdown factors (bold) for the $\mathcal{M}_b$ evolution kernel at different resolutions with various optimizations individually disabled.

resolution	64^2x16^2	64^2x32^2	128^2x16^2	128^2x32^2	256^2x16^2	256^2x32^2
1 node,1 GPU	7.16	8.76	6.37	9.55	7.36	9.25
1 node, 2 GPU	9.70	14.04	9.73	16.50	11.98	16.60
2 node, 4 GPU	11.98	20.23	13.59	26.69	17.69	27.41

Table 4.4: Overall speedup resulting from our optimized, multi-GPU approach, as compared with original single-GPU simulation, as the number of nodes & GPUs is increased.

Chapter 5

Conclusions

Before the framework designed herein can be used to make quantitative predictions, a validation effort needs to be undertaken with assistance from experimentalists. Tasks will include determining appropriate translational and rotational steric interaction potentials, measuring the viscous fluid flow, testing the hydrodynamics of individual filaments, and ensuring that motor-mediated and filament-filament interactions are well-resolved. No studies of the three-dimensional flow within gliding motility assays are known as of the time of writing.

Taking advantage of both sparsity and parallelism allows the application of micro-macro methods to this high-dimensional system. Extension to generic mixtures of filaments and freely flowing motor complexes capable of binding to multiple filaments would necessitate a further increase in dimensionality. In such scenarios, each multiply-bound motor complex would need to track the center-of-mass positions, orientations, and arclength binding locations along multiple filaments. The relatively small size of the motor complexes

compared to the filaments creates the same kinds of sparsity employed for dimensional reduction in this work; it is unclear whether this would be sufficient to make a similar micro-macro formulation viable. Monte Carlo simulations have been employed in very high-dimensional micro-macro models, and might be more appropriate.

Further computational improvements may be realized by porting the filament update and fluid solve to the GPU, as both steps comprise multiple parallel tasks. The filament solve and related advection terms are already computed via CPU parallelism and may exhibit further acceleration on the GPU. The upwinding scheme used in the advection matrix involves heavy conditional branching, and may thus exhibit poor GPU performance compared to a weighting-based approach. The fluid solve steps of FFT, semispectral system solves, and inverse FFT likewise already utilize CPU parallelism. The current approach of storing each frequency pair's lower and upper triangular matrix factorizations in dense format is very expensive in terms of memory. As the original semispectral matrices are sparse, a GPU-accelerated conjugate gradient method could calculate the sparse matrix-vector products directly and thus obviate the need for any matrix storage.

While such changes might further accelerate the filament and fluid solve steps on the root node, further work needs to be done to develop scalable multi-node implementations. Distributed memory multidimensional FFT and inverse FFT operations with the FFTW library require each node to send a unique message to all other nodes [18]. The semispectral solve might thus need to be replaced with a more scalable alternative. For the filament solve, distributed conservative schemes need to be investigated.

It is hoped that the methods described herein provide useful guidance for any who

find themselves facing similar computational bottlenecks. The successful acceleration and scaling results obtained in this work demonstrate the suitability of micro-macro methods for current state-of-the-art heterogenous high-performance computing paradigms.

Bibliography

[1] Igor S. Aranson and Lev S. Tsimring. Pattern formation of microtubules and motors: Inelastic interaction of polar rods. *Physical Review E*, 71:050901, May 2005.

[2] Aparna Baskaran and M Cristina Marchetti. Hydrodynamics of self-propelled hard rods. *Physical Review E*, 77(1):011920, 2008.

[3] G. K. Batchelor. Slender-body theory for particles of arbitrary cross-section in Stokes flow. *Journal of Fluid Mechanics*, 44(03):419–440, 1970.

[4] G. K. Batchelor. The stress system in a suspension of force-free particles. *Journal of Fluid Mechanics*, 41(03):545–570, 1970.

[5] R. Byron Bird, Charles F. Curtiss, Robert C. Armstrong, and Ole Hassager. *Dynamics of Polymeric Liquids, Kinetic Theory*. Wiley-Interscience, volume 2 edition, June 1987.

[6] Clifford P. Brangwynne, F. C. MacKintosh, and David A. Weitz. Force fluctuations and polymerization dynamics of intracellular microtubules. *Proceedings of the National Academy of Sciences*, 104(41):16128–16133, 2007.

[7] David Cai, Louis Tao, Michael Shelley, and David W McLaughlin. An effective kinetic representation of fluctuation-driven neuronal networks with application to simple and complex cells in visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 101(20):7757–7762, 2004.

[8] S.C. Cook, C. Hohenegger, and T. Shinar. A Micro-Macro Framework for Analyzing Steric and Hydrodynamic Interactions in Gliding Assays. *SIAM Journal on Multiscale Modeling and Simulation*, 2017.

[9] Nvidia Corporation. Nvidia visual profiler. `https://developer.nvidia.com/nvidia-visual-profiler`, 2017.

[10] David L. Coy, Michael Wagenbach, and Jonathon Howard. Kinesin takes one 8-nm step for each atp that it hydrolyzes. *The Journal of Biological Chemistry*, 274(6), 1999.

[11] Masao Doi and Sam F. Edwards. *The Theory of Polymer Dynamics*. Oxford University Press, Oxford, 1986.

[12] Knut Drescher, Raymond E Goldstein, Nicolas Michel, Marco Polin, and Idan Tuval. Direct measurement of the flow field around swimming microorganisms. *Physical Review Letters*, 105(16):168101, 2010.

[13] C.P. Dullemond. Lecture on numerical fluid dynamics. `http://www.mpia-hd.mpg.de/~dullemon/lectures/fluiddynamics08`, 2008.

[14] S.A. Endow and D.S. Barker. Processive and nonprocessive models of kinesin movement. *Annual review of physiology*, 65(1):161–175, 2003.

[15] Barath Ezhilan, Michael J Shelley, and David Saintillan. Instabilities and nonlinear dynamics of concentrated active suspensions. *Physics of Fluids*, 25:070607, 2013.

[16] T. L. Fallesen, J. C. Macosko, and G. Holzwarth. Force-velocity relationship for multiple kinesin motors pulling a magnetic bead. *European biophysics journal : EBJ*, 40(9):1071–1079, September 2011.

[17] Aldo Frezzotti, Gian Pietro Ghiroldi, and Livio Gibelli. Solving kinetic equations on gpu's. Technical report, DTIC Document, 2011.

[18] M. Frigo and S. G. Johnson. Fftw: an adaptive software architecture for the fft. In *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, volume 3, pages 1381–1384 vol.3, May 1998.

[19] Tong Gao, Robert Blackwell, Matthew A Glaser, MD Betterton, and Michael J Shelley. Multiscale modeling and simulation of microtubule–motor-protein assemblies. *Physical Review E*, 92(6):062709, 2015.

[20] Wei Ge, Ji Xu, Qingang Xiong, Xiaowei Wang, Feiguo Chen, Limin Wang, Chaofeng Hou, Ming Xu, and Jinghai Li. Multi-scale continuum-particle simulation on cpu–gpu hybrid supercomputer. In *GPU Solutions to Multi-scale Problems in Science and Engineering*, pages 143–161. Springer, 2013.

[21] L. Giomi, L. Mahadevan, B. Chakraborty, and M. F. Hagan. Excitable patterns in active nematics. *Phys. Rev. Lett.*, 106:218101, May 2011.

[22] W. O. Hancock and J. Howard. Kinesin's processivity results from mechanical and chemical coordination between the ATP hydrolysis cycles of the two motor domains. *Proceedings of the National Academy of Sciences*, 96(23):13147–13152, November 1999.

[23] Mark Harris. Optimizing cuda. *SC07: High Performance Computing With CUDA*, 2007.

[24] Christel Hohenegger, Steve Cook, and Tamar Shinar. Dimensional Reduction of a Multiscale Continuum Model of Microtubule Gliding Assays. *SIAM Journal on Applied Mathematics*, 74(5):1338–1353, January 2014.

[25] Christel Hohenegger and Michael J. Shelley. Dynamics of complex biofluids. In M. Ben Amar, A. Goriely, M. M. Müller, and L. F. Cugliandolo, editors, *New Trends*

in the Physics and Mechanics of Biological Systems, Ecole de Physique des Houches Session XCII, 6–31 July 2009, chapter 3, pages 65–92. Oxford University Press, 2011.

[26] J. Howard. The movement of kinesin along microtubules. *Annual review of physiology*, 58(1):703–729, 1996.

[27] J. Howard, A. J. Hudspeth, and R. D. Vale. Movement of microtubules by single kinesin molecules. *Nature*, 342(6246):154–158, November 1989.

[28] Jonathon Howard. *Mechanics of Motor Proteins & the Cytoskeleton*. Sinauer Associates, February 2001.

[29] Hanyu Jiang, Morisa Manzella, Luka Djapic, and Narayan Ganesan. Computational framework for in-silico study of virtual cell biology via process simulation and multiscale modeling. In *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pages 384–393. ACM, 2016.

[30] Eric Karsenti, François J. Nédélec, and Thomas Surrey. Modelling microtubule patterns. *Nature Cell Biology*, 8(11):1204–1211, November 2006.

[31] Roland Keunings. Micro-macro methods for the multiscale simulation of viscoelastic flow using molecular models of kinetic theory. *Rheology reviews*, pages 67–98, 2004.

[32] J. Kierfeld, K. Frentzel, P. Kraikivski, and R. Lipowsky. Active dynamics of filaments in motility assays. *The European Physical Journal-Special Topics*, 157(1):123–133, 2008.

[33] Stefan Klumpp and Reinhard Lipowsky. Cooperative cargo transport by several molecular motors. *PNAS*, 102(48):17284–17289, November 2005.

[34] Ljubinko Kondic, Peter Palffy-Muhoray, and Michael J. Shelley. Models of non-Newtonian Hele-Shaw flow. *Physical Review E*, 54(5):R4536–R4539, November 1996.

[35] Christian B. Korn, Stefan Klumpp, Reinhard Lipowsky, and Ulrich S. Schwarz. Stochastic simualtions of cargo transport by processive molecular motors. *The journal of chemical physics*, 131:245107, 2009.

[36] Pavel Kraikivski, Reinhard Lipowsky, and Jan Kierfeld. Enhanced ordering of interacting filaments by molecular motors. *Phys. Rev. Lett.*, 96(25):258103, Jun 2006.

[37] Karsten Kruse, Jean-François Joanny, Frank Jülicher, Jacques Prost, and Ken Sekimoto. Asters,vortices, and rotating spirals in active gels of polar filaments. *Physical Review Letters*, 92(7):078101–1:4, February 2004.

[38] Karsten Kruse, Jean-François Joanny, Frank Jülicher, Jacques Prost, and Ken Sekimoto. Generic theory of active polar gels: a paradigm for cytoskeletal dynamics. *The European Physical Journal E*, 16:5–16, 2005.

[39] H. Y. Lee and Mehran Kardar. Macroscopic equations for pattern formation in mixtures of microtubules and molecular motors. *Physical Review E*, 64(5):056113, October 2001.

[40] Ha Youn Lee and Mehran Kardar. Macroscopic equations for pattern formation in mixtures of microtubules and molecular motors. *Physical Review E*, 64(5):056113, 2001.

[41] Tanniemola B Liverpool. Active gels: where polymer physics meets cytoskeletal dynamics. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 364(1849):3335–3355, 2006.

[42] Tanniemola B. Liverpool and M. Cristina Marchetti. Bridging the microscopic and the hydrodynamic in active filament solutions. *Europhysics Letters*, 69(5):846–852, March 2005.

[43] MC Marchetti, J-F Joanny, S Ramaswamy, TB Liverpool, J Prost, Madan Rao, and R Aditi Simha. Hydrodynamics of soft active matter. *arXiv preprint arXiv:1207.2929*, 2012.

[44] Scott A. McKinley, Avanti Athreya, John Fricks, and Peter R. Kramer. Asymptotic analysis of microtubule-based transport by multiple identical molecular motors. *Journal of Theoretical Biology*, 305:54–69, July 2012.

[45] Daisuke Mizuno, Catherine Tardin, Christoph F. Schmidt, and Frederik C. MacKintosh. Nonequilibrium mechanics of active cytoskeletal networks. *Science*, 315:370–373, January 2007.

[46] F. Nédélec. Computer simulations reveal motor properties generating stable antiparallel microtubule interactions. *The Journal of Cell Biology*, 158(6):1005–1015, September 2002.

[47] F. J. Nédélec, T. Surrey, A. C. Maggs, and S. Leibler. Self-organization of microtubules and motors. *Nature*, 389(6648):305–308, September 1997.

[48] François Nédélec and Dietrich Foethke. Collective langevin dynamics of flexible cytoskeletal fibers. *New Journal of Physics*, 9(11):427, 2007.

[49] François Nédélec, Thomas Surrey, and A. C. Maggs. Dynamic concentration of motors in microtubule arrays. *Physical Review Letters*, 86:3192–3195, April 2001.

[50] François J. Nédélec and Thomas Surrey. Dynamics of microtubule aster formation by motor complexes. *Comptes-Rendus de l'Académie des Sciences Paris*, 4(2):841–847, May 2001.

[51] Kyle E Niemeyer and Chih-Jen Sung. Recent progress and challenges in exploiting graphics processors in computational fluid dynamics. *The Journal of Supercomputing*, 67(2):528–564, 2014.

[52] Nvidia. Nvidia cuda c programming guide. *Nvidia Corporation*, 2011.

[53] C. Peskin. The immersed boundary method. *Acta Numerica*, 11:479–517, July 2003.

[54] Stephen L. Rogers and Jonathan M. Scholey. Motility assays for microtubule motor proteins. *Nature Encyclopedia of Life Sciences*, 2001.

[55] Stephen L Rogers and Jonathan M Scholey. Motility assays for microtubule motor proteins. *eLS*, 2004.

[56] D. Saintillan and M. J. Shelley. Instabilities, pattern formation, and mixing in active suspensions. *Physics of Fluids*, 20(12):16, 2008.

[57] David Saintillan and Michael J. Shelley. Instabilities and pattern formation in active particle suspensions: Kinetic theory and continuum simulations. *Physical review letters*, 100(17):178103, May 2008.

[58] S. Sankararaman, G. I. Menon, and P. B. SunilKumar. Self-organized pattern formation in motor-microtubule mixtures. *Physical Review E*, 70:031905+, September 2004.

[59] Sumithra Sankararaman, Gautam I. Menon, and P. B. SunilKumar. Modelling pattern formation in motor-microtubule mixtures. *Physica Scripta*, T106:26–31, April 2003.

[60] Sumithra Sankararaman, Gautam I. Menon, and P. B. SunilKumar. Self-organized pattern formation in motor-microtubule mixtures. *Physical Review E*, 70:031905, September 2004.

[61] V. Schaller, C. Weber, C. Semmrich, W. Frey, and A. R. Bausch. Polar patterns of driven filaments. *Nature*, 467(7311):73–77, September 2010.

[62] Volker Schaller, Christoph Weber, Erwin Frey, and Andreas R. Bausch. Polar pattern formation: hydrodynamic coupling of driven filaments. *Soft Matter*, 7:3213–3218, 2011.

[63] Michael J Shelley. The dynamics of microtubule/motor-protein assemblies in biology and physics. *Annual Review of Fluid Mechanics*, 48:487–506, 2016.

[64] Tamar Shinar, Miyeko Mana, Fabio Piano, and Michael J Shelley. A model of cytoplasmically driven microtubule-based motion in the single-celled caenorhabditis elegans embryo. *Proceedings of the National Academy of Sciences*, 108(26):10508–10513, 2011.

[65] Yutaka Sumino, Ken H. Nagai, Yuji Shitaka, Dan Tanaka, Kenichi Yoshikawa, Hugues Chate, and Kazuhiro Oiwa. Large-scale vortex lattice emerging from collectively moving microtubules. *Nature*, 483(7390):448–452, March 2012.

[66] Thomas Surrey, François J. Nédélec, Stanislas Leibler, and Eric Karsenti. Physical properties determining self-organization of motors and microtubules. *Science*, 292:1167–1171, March 2001.

[67] Sumanth Swaminathan, Falko Ziebert, Dmitry Karpeev, and Igor S. Aranson. Motor-mediated alignment of microtubules in semidilute mixtures. *Physical Review E*, 79(3):036207, March 2009.

[68] Vasily Volkov. Better performance at lower occupancy. In *Proceedings of the GPU technology conference, GTC*, volume 10, page 16. San Jose, CA, 2010.

[69] Qi Wang. A hydrodynamic theory for solutions of nonhomogeneous nematic liquid crystalline polymers of different configurations. *The Journal of chemical physics*, 116(20):9120–9136, 2002.

[70] E Westphal, SP Singh, C-C Huang, Gerhard Gompper, and Roland G Winkler. Multiparticle collision dynamics: Gpu accelerated particle-based mesoscale hydrodynamic simulations. *Computer Physics Communications*, 185(2):495–503, 2014.

[71] Sergey Zabelok, Robert Arslanbekov, and Vladimir Kolobov. Adaptive kinetic-fluid solvers for heterogeneous computing architectures. *Journal of Computational Physics*, 303:455–469, 2015.