Single Filament Length‑Control Mechanisms: Theory, Simulation, and SINDy Model Comparison

Abstract

We investigate three distinct length control mechanisms for a single filament: (1) constant disassembly under a limited monomer pool, (2) severing in an unlimited (free) pool, and (3) severing under a limited pool. For each mechanism, we perform stochastic simulations using the Gillespie algorithm across numerous trajectories. We then compute the ensemble-averaged filament length, apply a Savitzky–Golay filter to smooth the dynamics, and use Sparse Identification of Nonlinear Dynamics (SINDy) for model inference. By fitting models that include only a constant term and the lowest-order length-dependent term predicted by theory, SINDy accurately recovers the theoretical coefficients within approximately 10%. These data-driven results validate the underlying rate equations and suggest that high-throughput single-filament experiments could reliably distinguish among the three control mechanisms.

1 Introduction

We begin by analyzing the assembly of bare filaments in a monomer pool. The polymerization rate depends on the concentration of available monomers, which can be either finite (limited pool) or effectively infinite (free pool). Filament disassembly occurs either through monomer loss at the ends—returning monomers to the pool—or through severing events, where the detached segments re-enter the pool as free subunits

Objective and Approach:

Filaments grow and shrink through the addition and removal of monomers, respectively, with their length determined by the balance of these processes. Given a set of rate parameters, we use Gillespie stochastic simulations to model the time evolution of filament length. By running these simulations over many independent trajectories, we obtain statistically mean length dynamics.

Step 1: Constant Disassembly in a Limited Pool

We first consider the case of filaments regulated by constant disassembly in a limited monomer pool. For this mechanism, the theoretical equation governing filament dynamics is already known. Using simulation data, we extract the ensemble-averaged filament length over time and apply the Sparse Identification of Nonlinear Dynamics (SINDy) algorithm to infer the underlying dynamical equation. Remarkably, SINDy is able to reconstruct the expected equation from the simulation data, validating both the simulation approach and the theoretical model.

Step 2: Severing Mechanisms in Free and Limited Pools

Building on this, we next extend the analysis to include two additional regulatory mechanisms: severing in a free (unlimited) monomer pool, and severing in a limited pool. For both cases, we follow the same methodology—running Gillespie simulations to generate filament length trajectories, computing the ensemble-averaged dynamics, and applying SINDy to infer the underlying length regulation equations.

1. **Resource limitation (pool-depletion feedback).**

When a filament grows, it literally steals subunits from the pool.
Fewer free monomers ⇒ slower addition ⇒ growth rate drops automatically.

2. **Length-dependent severing (geometry feedback).**

Longer filaments present more “cut sites.”

Every sever event removes a chunk of the filament, so the loss rate rises faster than length itself.

A single bare filament is simple enough to capture these ideas mathematically yet rich enough to illustrate how different feedbacks leave distinct signatures in a length-versus-time trace.
Our goal is two-fold:

- Forward problem: write down the mean-field ordinary-differential equation (ODE) for the ensemble-average length ⟨L(t)⟩ under each feedback motif.
- Inverse problem: show that a data-driven method—Sparse Identification of Non-linear Dynamics (SINDy)—can rediscover the correct ODE directly from simulated trajectories.

2 Theoretical Framework

Before diving into theory, we fix notation:

Notation	Units
L(t)	filament length
N_{tot}	total number of monomers in system
r or k_+' (k plus prime)	Assembly rate constant
k_-	Constant disassembly rate
s	severing rate

To simulate the stochastic dynamics of filament length, we use the Gillespie algorithm, a Monte Carlo method that generates statistically exact trajectories for systems governed by discrete chemical reactions. The algorithm calculates when the next reaction will occur and which specific reaction it will be, based on the current state of the system and the rates of all possible events. In the context of filament dynamics, these events include monomer addition, monomer dissociation, and filament severing.

At each step, the algorithm performs two key tasks:

1. **Time selection:** It samples the time interval until the next reaction using an exponential distribution derived from the sum of all reaction propensities.
2. **Event selection:** It randomly selects which reaction occurs next, with probabilities weighted by each reaction's rate.

This process is repeated iteratively, allowing the system to evolve in time. By applying this to many independent simulation runs, we capture the stochastic variability of filament growth and decay. The Gillespie algorithm is particularly valuable here because it accurately represents the noise and fluctuations inherent to single-filament behavior—features that are often averaged out or missed in deterministic models.

$\frac{d<L>}{dt}$= (assembly term) - (disassembly term).

2.1 Limited pool with Disassembly

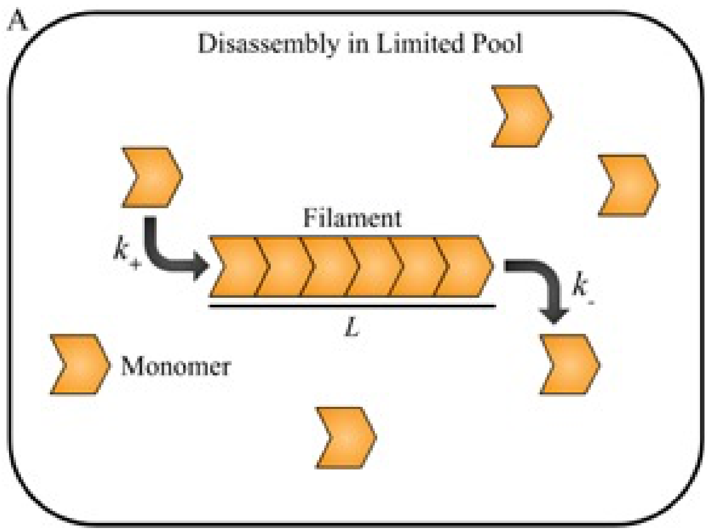

Fig 2.1.1 Disassembly in Limited Pool

Parameters used:

N = 1000, $k'_+ = 0.3$, $k_- = 225$

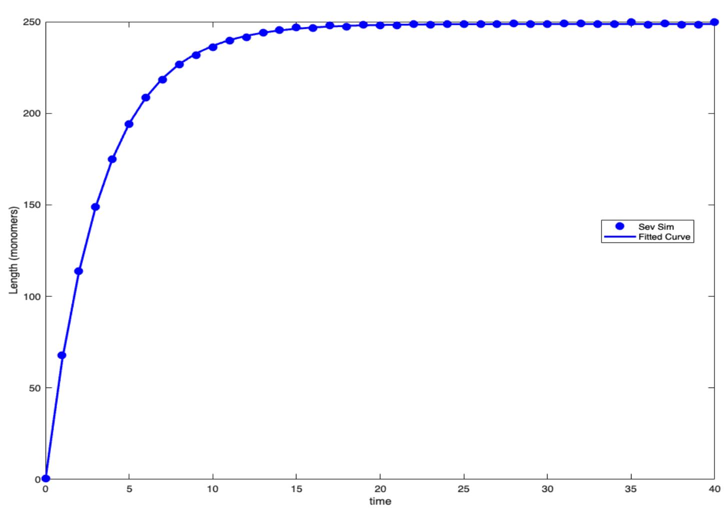

Fig 2.1.1 Ensemble-average filament length (blue dots) approaching its theoretical steady state with an exponential fit (solid line).

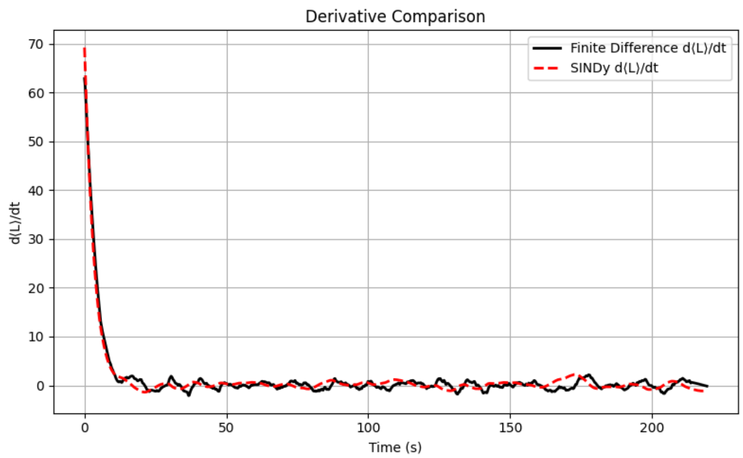

Fig 2.1.2 Finite-difference estimate of d⟨L⟩/dt (black) overlaid with SINDy prediction (red dashed) for the limited-pool case.

Whereas the SINDy derived equation is as follows:

(x0)' = 70.194 1 + -0.279 x0

Where (x0)' = $\frac{d<L>}{dt}$, (x0) = L

Substituting the coefficients we get,

$$\frac{d<L>}{dt} = 70.194 - 0.279 < L >$$

$$\frac{d<L>}{dt} = c0 + c1 < L >$$

Expected theoretical equation is as follows:

$$\frac{d<L>}{dt} = k_{+}'(N - < L >) - k_{-}$$

$$\frac{d<L>}{dt} = (k_{+}'N - k_{-}) + (-k_{+}') < L >$$

SINDy	c0 = 70.194	c1 = -0.279
Theory	$(k_{+}'N - k_{-})$ = 75	$-k_{+}'$ = -0.3

2.2 Free pool with length-dependent severing

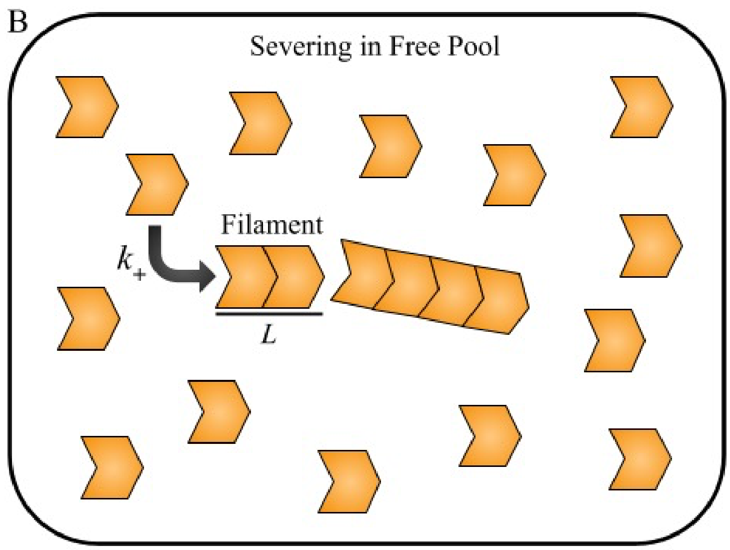

Fig 2.2.1 Filament Length Severing in Free Pool

Parameters used:

N = 1000, $k'_+ = 0.3$, s = 0.0075

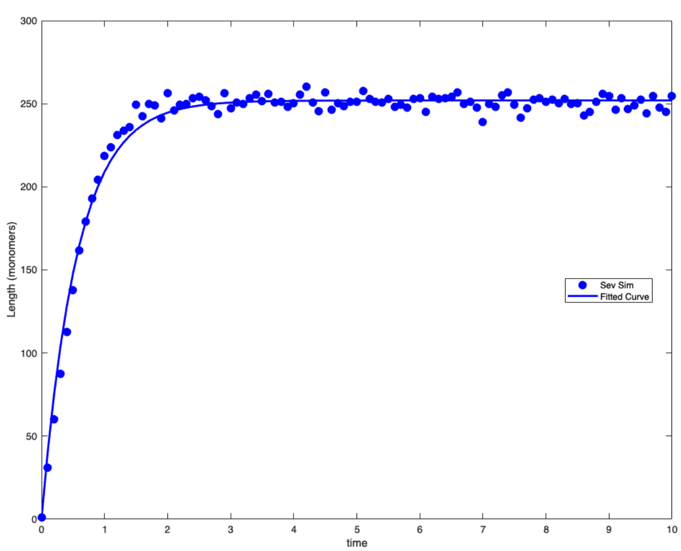

Fig 2.2.1 Ensemble-average filament length (blue dots) approaching its theoretical steady state with an exponential fit (solid line).

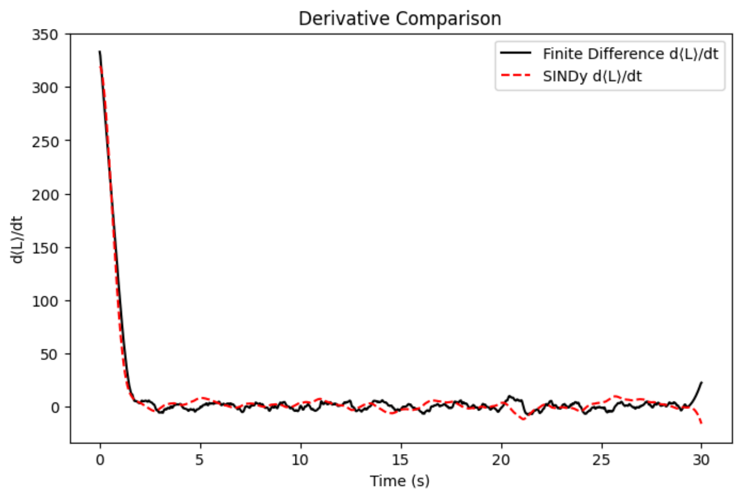

Fig 2.2.2 Finite-difference estimate of d⟨L⟩/dt (black) overlaid with SINDy prediction (red dashed) for the limited-pool case.

SINDy derived equation is as follows:

(x0)' = 319.541 f0(x0) + -0.005 f1(x0)

Where (x0)' = $\frac{d<L>}{dt}$, f0(x0) = 1, f1(x0) = L^2

Substituting the coefficients we get,

$$\frac{d<L>}{dt} = 319.541 - 0.005 < L^2 >$$

$$\frac{d<L>}{dt} = c0 + c2 < L^2 >$$

Expected theoretical equation is as follows:

$$\frac{d<L>}{dt} = k_+'N - \frac{s}{f} < L^2 >$$

Where we need to find out the value of f

SINDy	c0 = 319.541	c2 = -0.005
Theory	$k_+'N_{tot}$ = 300	$-\frac{s}{f} = -\frac{2s}{\pi}$ = -0.00478

Where, f = π / 2
So the equation will be

$$\frac{d<L>}{dt} = k_{+}'N_{tot} - \frac{2}{\pi}s< L >^{2}$$

Assembly: constant $k_{+}'N_{tot}$ (pool effectively infinite).

Disassembly: each cut returns, on average, half the filament, so expected subunits lost per cut is L/2. Cut sites are chosen uniformly ⇒ propensity ∝ sL.

2.3 Limited pool and severing

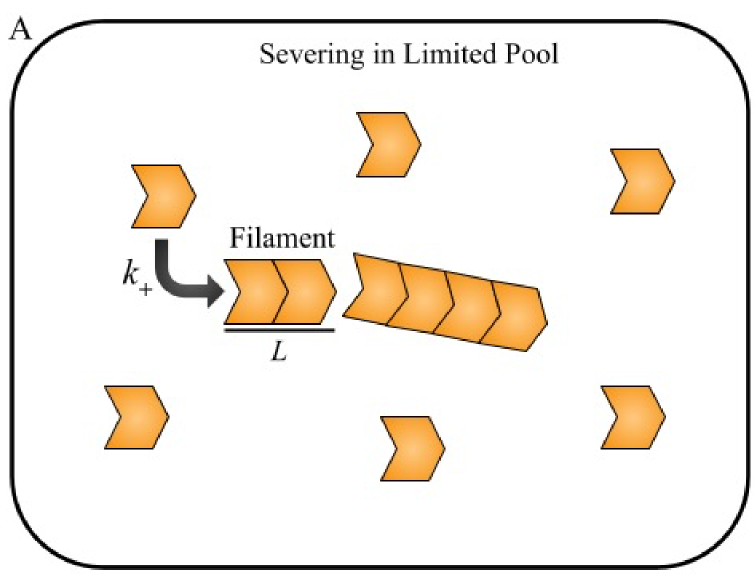

Fig 2.3.1 Severing in Limited Pool where k_{+} is the assembly rate constant and L is the Filament Length

Parameters used:

N = 1000, k_+'= 0.3, s = 0.0075

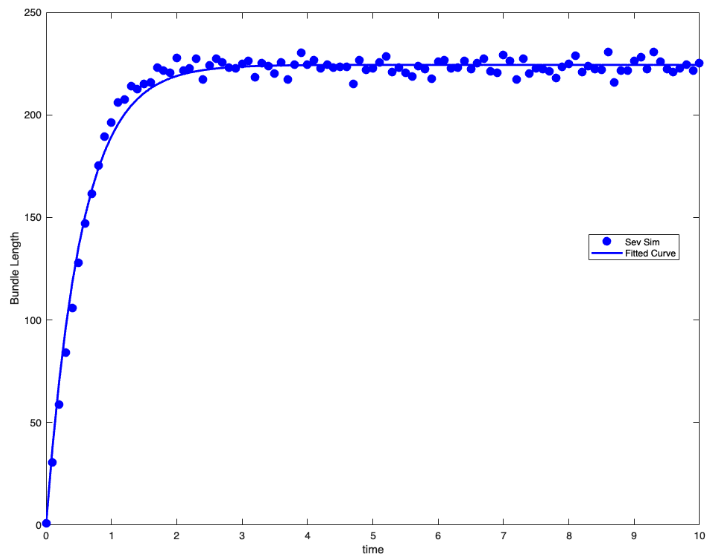

Fig 2.3.2 Ensemble-average filament length (blue dots) approaching its theoretical steady state with an exponential fit (solid line).

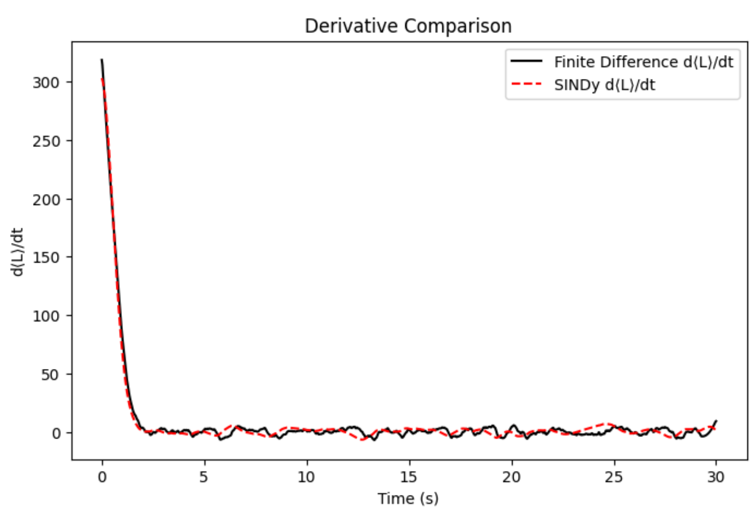

Fig 2.3.3 Finite-difference estimate of d⟨L⟩/dt (black) overlaid with SINDy prediction (red dashed) for the limited-pool case.

SINDy derived equation is as follows:

(x0)' = 338.750 1 + -0.829 x0 + -0.003 x0^2

Where (x0)' = $\frac{d<L>}{dt}$, x0 = L, $x0^2 = L^2$

Substituting the coefficients we get,

$$\frac{d<L>}{dt} = 338.750 - 0.829 < L > - 0.003 < L^2 >$$

$$\frac{d<L>}{dt} = c0 + c1 < L > + c2 < L^2 >$$

Expected theoretical equation is as follows:

$$\frac{d<L>}{dt} = k'_+(N - < L >) - \frac{s}{f} < L^2 >$$

$$\frac{d<L>}{dt} = k'_+ N - k'_+ < L > - \frac{s}{f} < L^2 >$$

Where we need to find out the value of f

SINDy	c0 = 319.541	c1 = 0.829	c2 = -0.003
Theory	$k'_+ N$ = 300	k'_+ = 0.3	$\frac{s}{f} = \frac{2s}{\pi}$ = 0.00478

Where, f = π / 2

So the equation will be

$$\frac{d<L>}{dt} = k'_+ N - k'_+ < L > - \frac{2s}{\pi} < L^2 >$$

3 Conclusion

In this study, we explored filament length dynamics under three regulatory mechanisms: constant disassembly in a limited pool, severing in a free pool, and severing in a limited pool. For each case, we performed Gillespie stochastic simulations to generate multiple filament length trajectories and extracted the mean length dynamics over time.

We then applied the Sparse Identification of Nonlinear Dynamics (SINDy) framework to the simulation data in order to infer the underlying dynamical equations governing the average filament length $\langle L \rangle$. Starting with the constant disassembly mechanism in a limited pool—where the theoretical rate equation is known—we used SINDy to recover the corresponding dynamical form $d\langle L \rangle/dt$. The coefficients obtained through SINDy closely matched the theoretical values, validating both the simulation framework and the accuracy of the data-driven approach.

We next applied the same methodology to the two severing-based mechanisms: one in a free monomer pool and the other in a limited pool. In both cases, SINDy successfully identified the expected form of the length dynamics and returned coefficient values that were in close agreement with theoretical predictions.

These results demonstrate that data-driven modeling using SINDy, when combined with high-throughput stochastic simulations, provides a powerful and accurate method for uncovering the governing equations of filament length regulation. Importantly, this approach suggests that experimentally measurable mean length trajectories could be sufficient to distinguish between different filament regulation mechanisms.

4 Next Goals

1. **Extend Beyond Polynomial Representations:**

While our current SINDy implementation relies on polynomial basis functions, future work should focus on developing or incorporating methods that can identify more complex functional forms. This is particularly important for cases where theoretical equations are unknown, and the governing dynamics may involve non-polynomial or nonlinear interactions.

2. **Increase Simulation Depth:**

To ensure statistical robustness and explore the system's long-term behavior, we aim to scale up the simulations to over 10,000 trajectories. This will help identify potential saturation points or steady-state regimes in filament length dynamics across different regulatory mechanisms.

3. **Refine Model Selection via Regression Analysis:**

We will either enhance the model discovery process by integrating domain-specific candidate variables (as in Step 1) or continue with general polynomial-based models and apply regression techniques. This will help isolate the most relevant terms in the governing equations and improve model interpretability and accuracy.