Efficient Parameter Estimation for DNA Kinetics Modeled as Continuous-Time Markov Chains

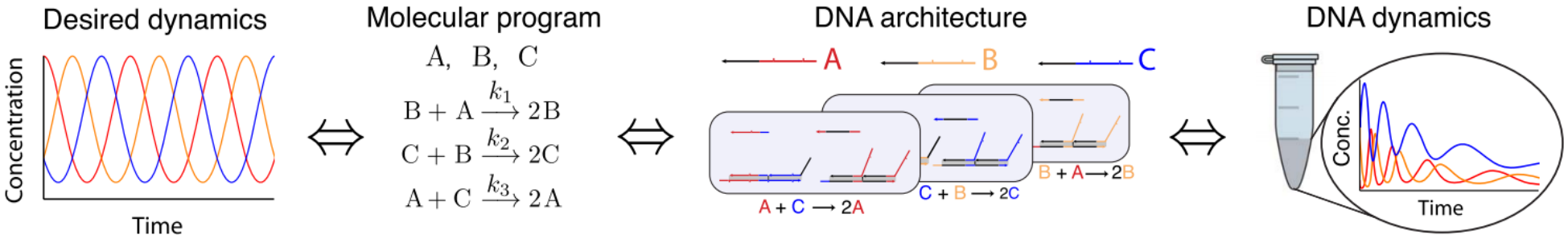
SEDIGHEH ZOLAKTAF[1], FRITS DANNENBERG[2], ERIK WINFREE[2],
ALEXANDRE BOUCHARD-COTE[1], MARK SCHMIDT[1], ANNE CONDON[1]

[1]UNIVERSITY OF BRITISH COLUMBIA, [2]CALIFORNIA INSTITUTE OF TECHNOLOGY

Nucleic Acid Kinetic Simulators

- Nucleic acid kinetic simulators aim to predict the kinetics of reactions involving interacting nucleic acid strands
  - e.g., rate of a reaction or sequence of interactions between the strands

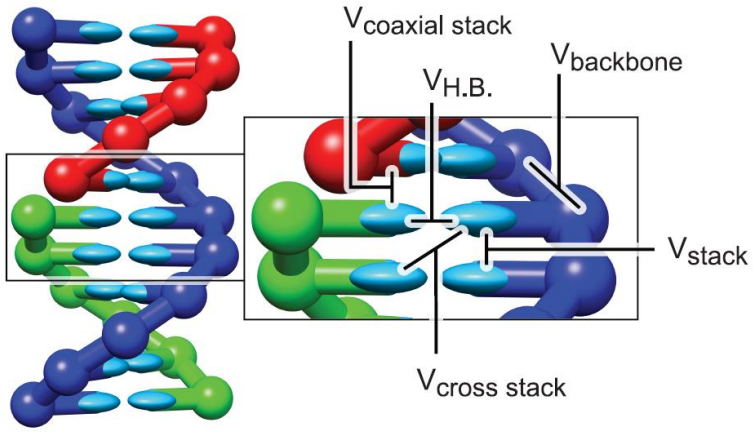
- Useful for nucleic acid-based devices



[Srinivas et al., 2017]

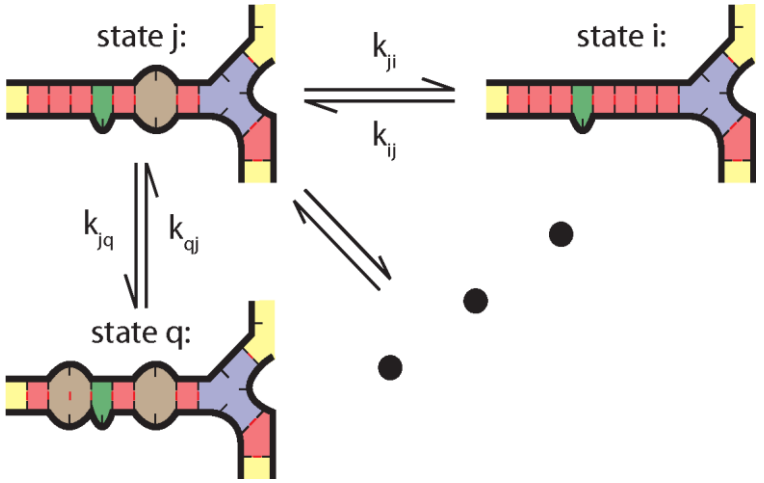
Nucleic Acid Kinetic Simulators

- Model kinetics at various levels of granularity



[Ouldridge et al., 2013]

Molecular dynamics models that follow the three-dimensional motion of the polymer chains

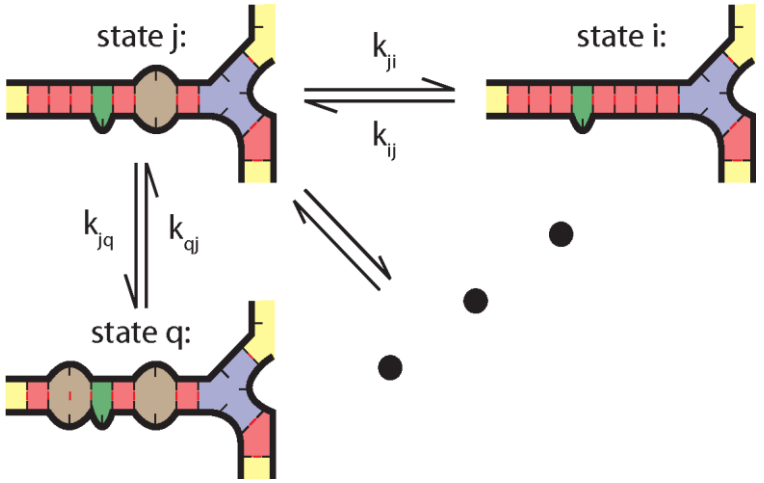


[Schaeffer et al., 2015]

Continuous-time Markov chain (CTMC) models with elementary steps that consider the forming or breaking of a single base pair

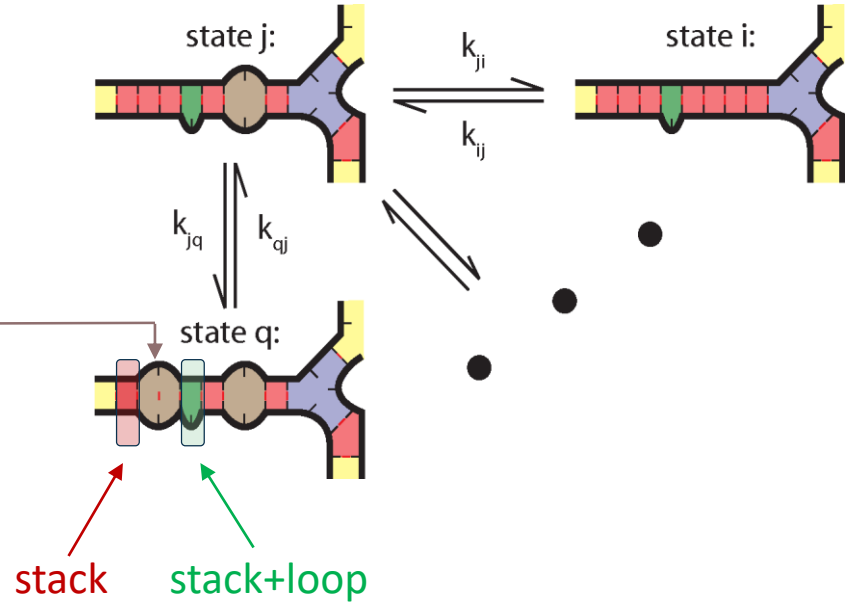
The Multistrand Kinetic Simulator [Schaeffer et al., 2015]

- Kinetics of (multiple) interacting strands are modelled as CTMCs with elementary steps.
- States represent a collection of non-pseudoknotted secondary structures
- Transitions between the states correspond to the forming or breaking of a base pair
- Transition rates determine the holding time of states and the transition probabilities between states



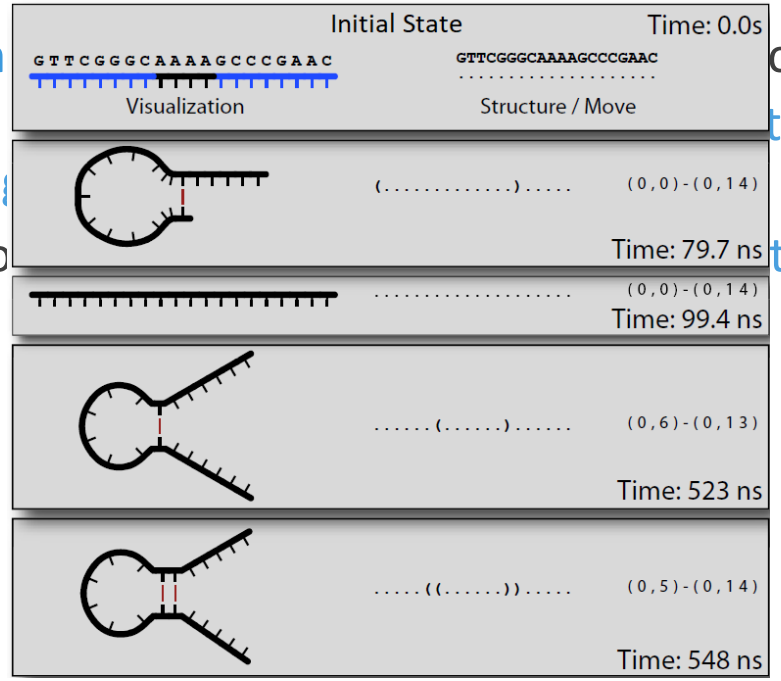
The Multistrand Kinetic Simulator [Schaeffer et al., 2015]

- Kinetics of (multiple) interacting strands are modelled as CTMCs with elementary steps.
- Transition rates are determined by a kinetic model (along with a thermal stability model)
  - Metropolis kinetic model (2 free parameters)
  - Arrhenius kinetic model (15 free parameters)
    - Arrhenius rate constants and activation energies
    - Transitions depends on the local context of the base pair that is forming or breaking



The Multistrand Kinetic Simulator [Schaeffer et al., 2015]

- Stochastically samples trajectories for a reaction
  - Sequences of states from an initial to a target state, along with the holding times to transition between successive states
- Uses Gillespie's Stochastic Sim ch advances forward in two steps
  1. Samples the holding time tial distribution with a rate equal to the sum of the outgoin
  2. Samples the next state fro ties of the current state



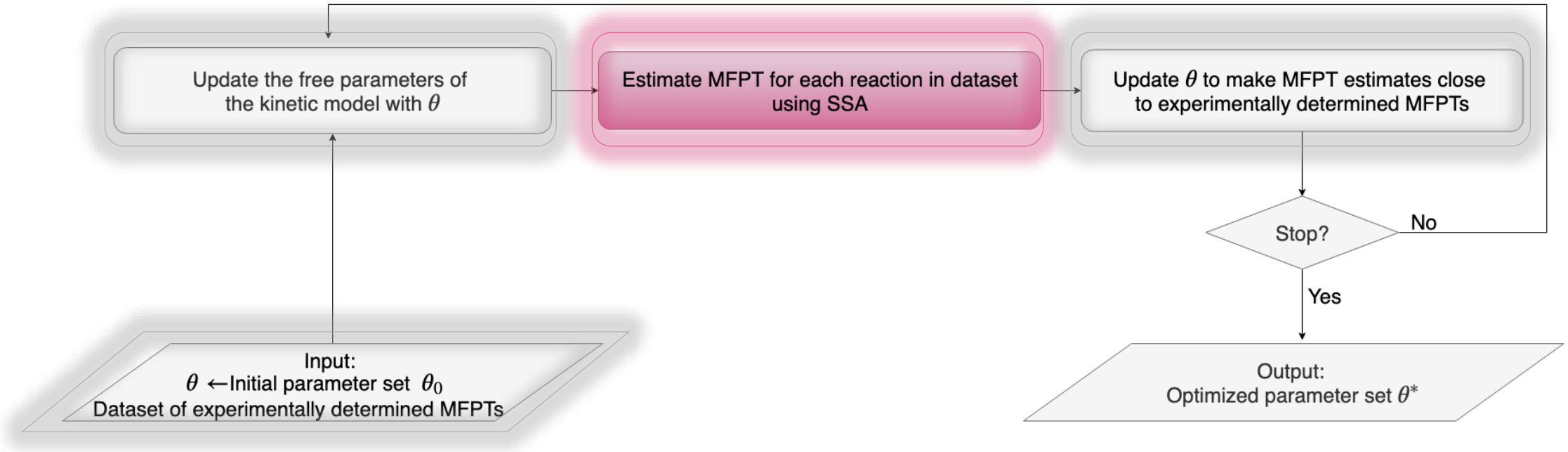
Adapted from [Schaeffer, 2013]

Mean First Passage Time (MFPT) Estimation

- The first passage time (FPT) of a trajectory, the first time that the trajectory occupies the target state, is the sum of the holding times of the states of the trajectory
- Estimate the mean first passage time (MFPT) from an initial state to a target state by using the FPT of independently sampled trajectories

→MFPT is useful to estimate kinetic rates, such as reaction rate constant

Parameter Estimation for DNA Kinetic Models



- To accurately estimate reaction kinetics, kinetic models should be calibrated

Challenges

- Two challenging tasks in accurately estimating reaction kinetics in the full state space of all non-pseudoknotted secondary structures

1. MFPT estimation
   - Sampling trajectories could be slow
   - Exact linear algebra methods is infeasible for CTMCs with large implicitly-represented state spaces
2. Parameter estimation for DNA kinetic models based on MFPT estimates
   - MFPT estimation for every parameter set variation could be slow

Contributions

- We address the following tasks in accurately estimating reaction kinetics in the full state space of all non-pseudoknotted secondary structures

1. MFPT estimation
   - We show how to use a reduced variance stochastic simulation algorithm (RVSSA)

2. Parameter estimation for DNA kinetic models based on MFPT estimates
   - We introduce a fixed path ensemble inference (FPEI) approach

Reduced Variance Stochastic Simulation Algorithm (RVSSA)

- RVSSA advances forward in two steps:
  1. Computes the expected holding time of a state equal to the inverse of the sum of the transition rates from the state
  2. Samples the next state from the outgoing transition probabilities of the current state

	Holding Time Variance	Path Variance
SSA	+	+
RVSSA		+

- RVSSA computes the MFPT of a path as the sum of the expected holding times of the states
- RVSSA estimates the MFPT from an initial state to a target state by using the MFPT of independently sampled paths

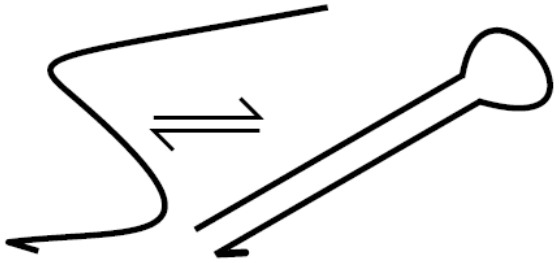
→ The estimator of the MFPT produced by RVSSA has a lower variance than the estimator produced by SSA
→RVSSA has smaller expected mean squared error and requires fewer sampled paths

How well do RVSSA and SSA Compare?

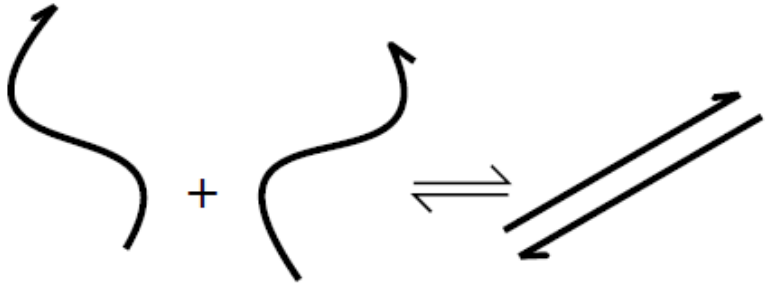
- We conduct computational experiments on real data

Dataset

- 21 reactions

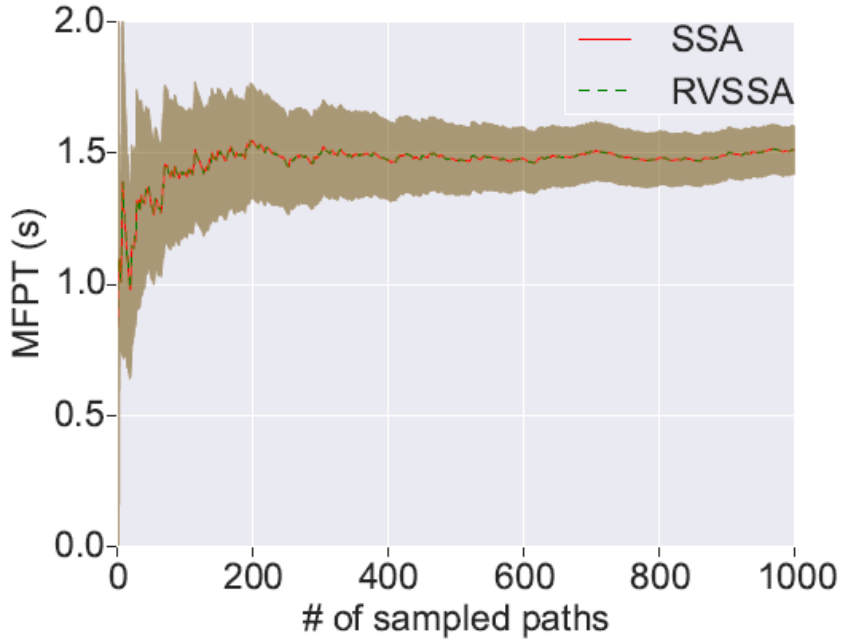


Hairpin closing and opening [Bonnet et al. 1998]

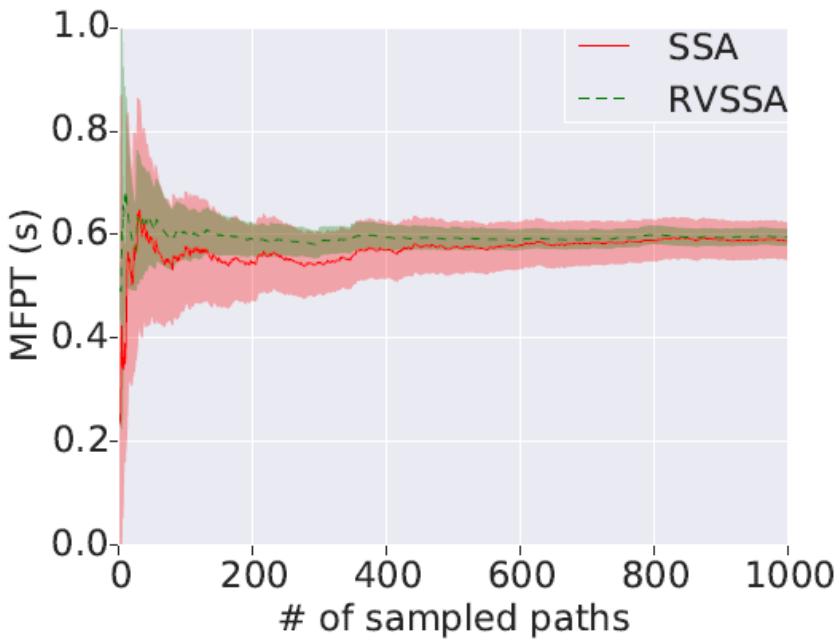


Helix association [Hata et al., 2017] [Wetmur, 1976] and helix dissociation (with mismatches) [Cisse et al., 2012]

MFPT Estimation



(a) Association of CCCACACTCTTACTTATCGACT with complement [Hata et al., 2017]



(b) Association of 25-mer poly(A) with 25-mer poly(T) [Wetmur, 1976]

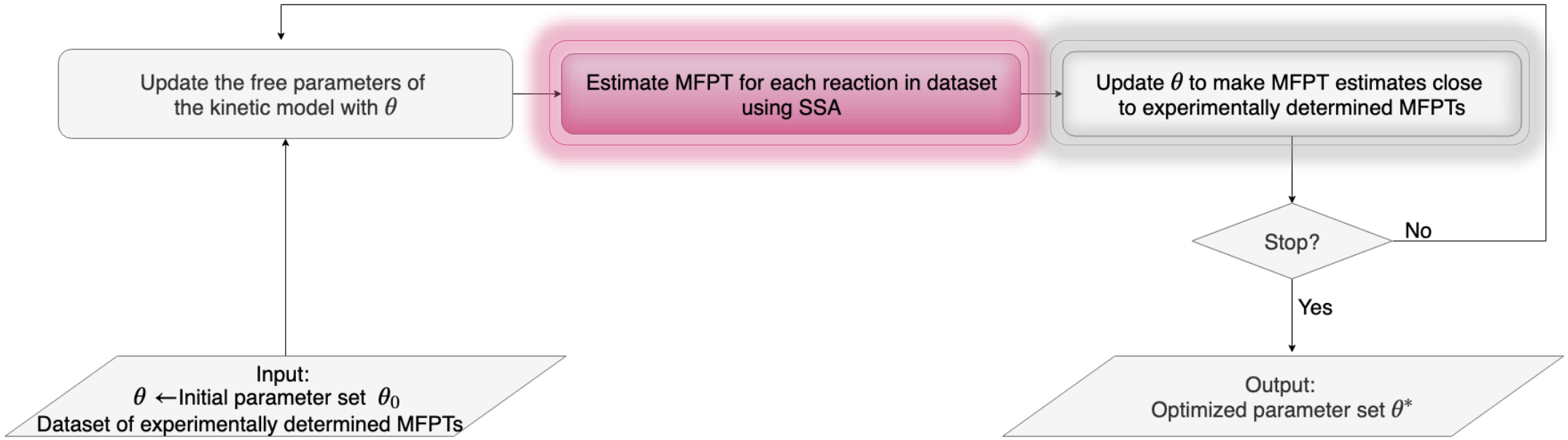
Contributions

- We address the following tasks in accurately estimating reaction kinetics in the full state space of all non-pseudoknotted secondary structures

1. MFPT estimation
   - We show how to use a reduced variance stochastic simulation algorithm (RVSSA)

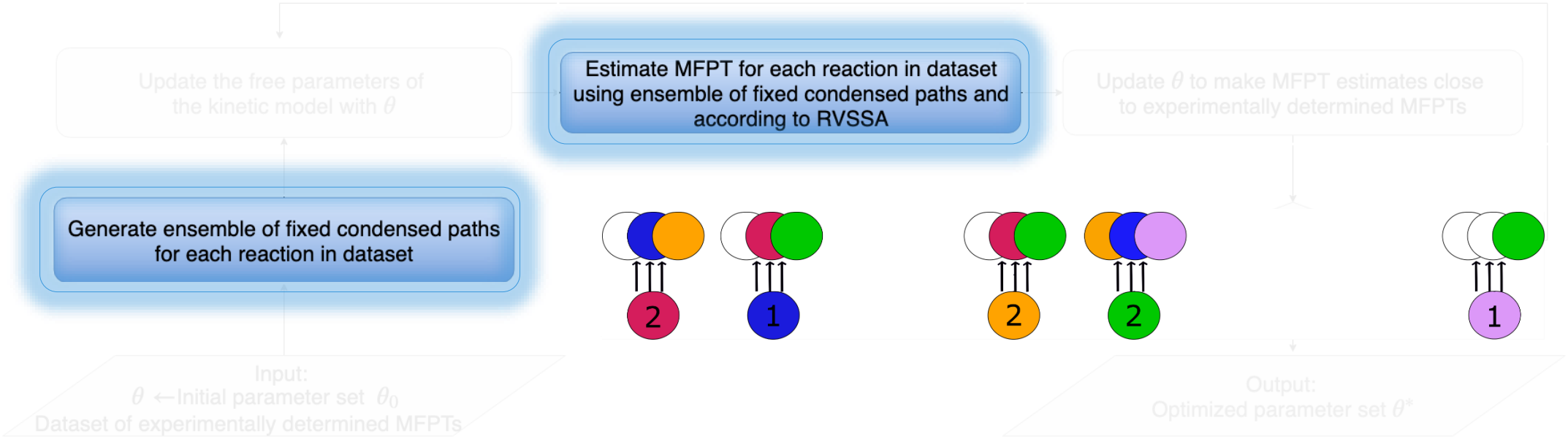
2. Parameter estimation for DNA kinetic models based on MFPT estimates
   - We introduce a fixed path ensemble inference (FPEI) approach

Stochastic Simulation Algorithm Inference (SSAI)



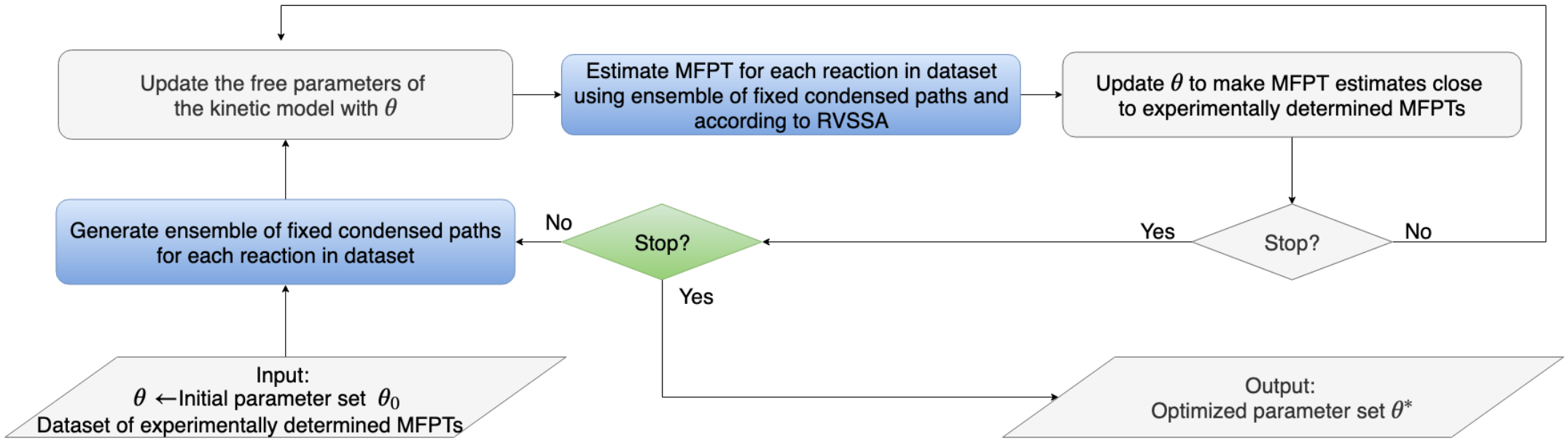
- We minimize the mean squared error (MSE) of the log of the MFPT estimates and the experimentally determined MFPTs
- Accurately estimating MFPTs for every parameter set variation could be slow

Fixed Path Ensemble Inference (FPEI)



- We condense a path by computing the set of states and the number of times each state is visited
  - We also require information about outgoing transitions of each state of the path, since we compute the holding time of a state in a path as if the path is regenerated in the full state space

Fixed Path Ensemble Inference (FPEI)



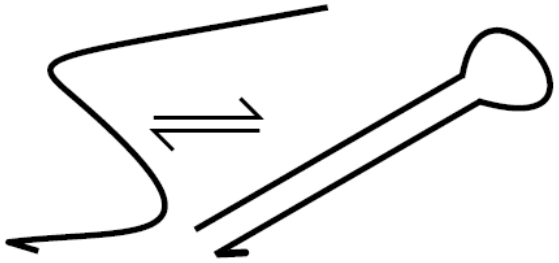
We occasionally generate a new ensemble of fixed condensed paths

How do FPEI and SSAI Compare?

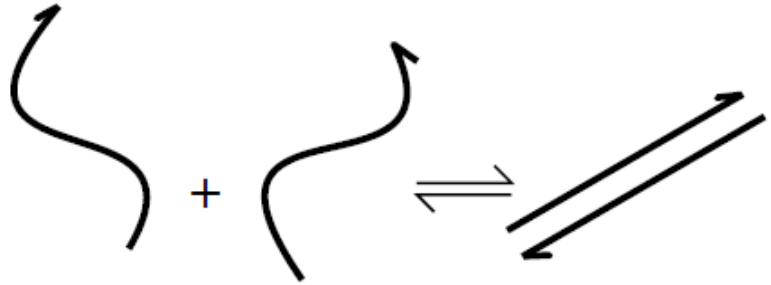
- We conduct computational experiments on real data

Dataset of Experimental Reaction Rate Constants

- 19 reactions

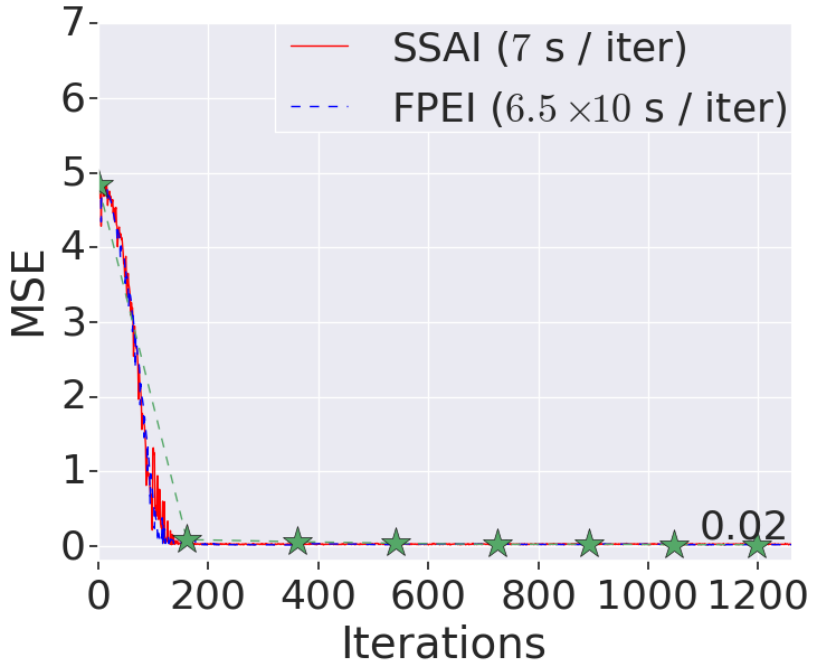


Hairpin closing and opening [Bonnet et al. 1998]

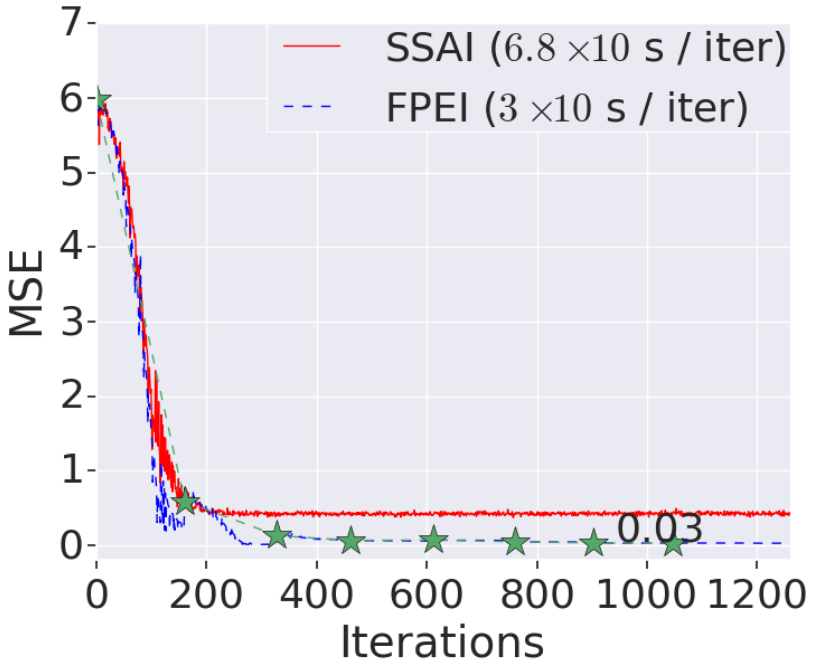


Helix association [Hata et al., 2017] and helix dissociation (with mismatches) [Cisse et al., 2012]

Parameter Estimation

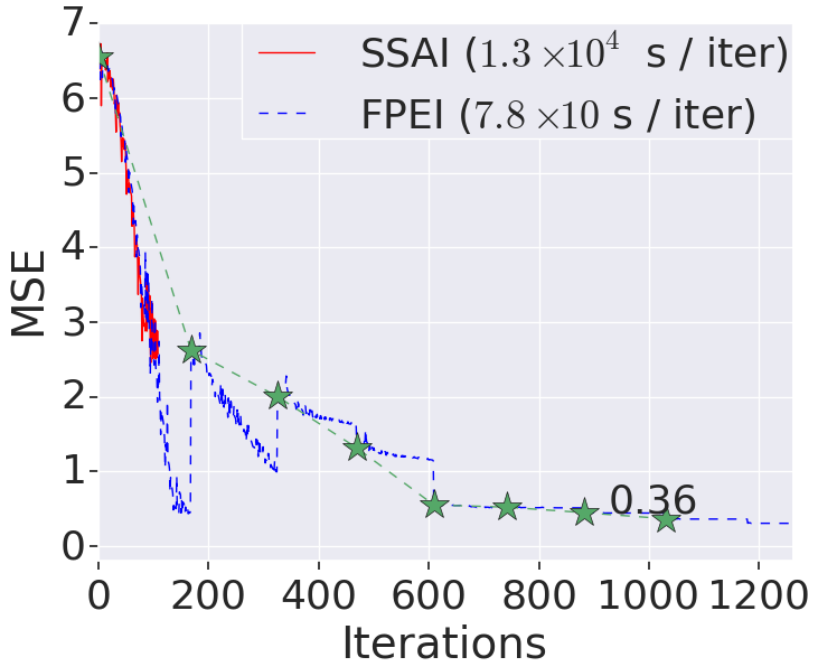


(a) 5 hairpin closing reactions [Bonnet et al., 1998]

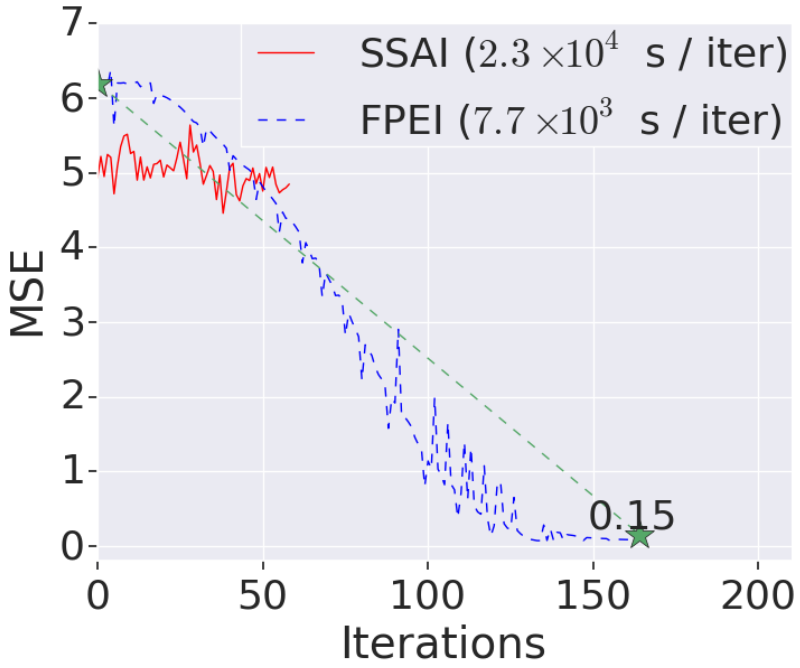


(b) 5 hairpin opening reactions [Bonnet et al., 1998]

Parameter Estimation

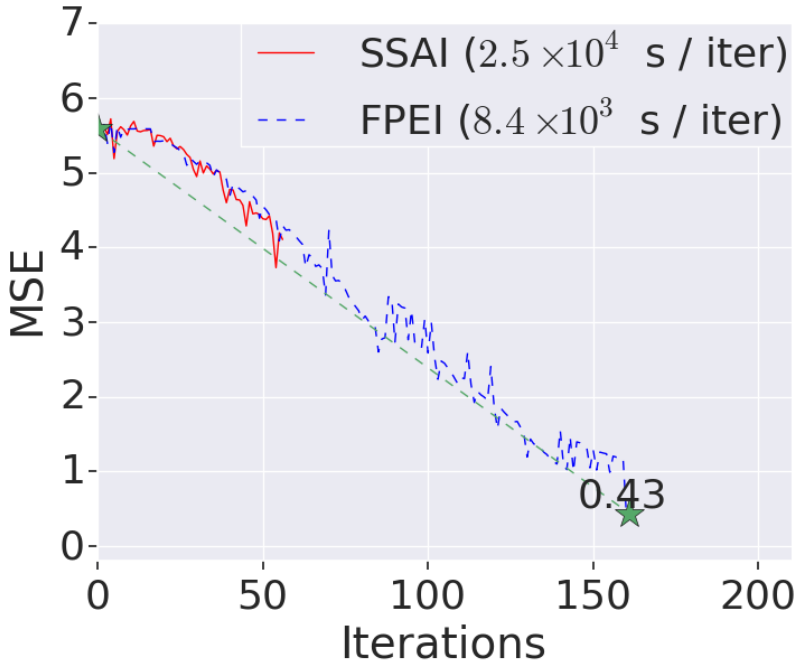


(a) 5 helix dissociation reactions (with mismatches) [Cisse et al., 2012]



(b) 4 helix association reactions [Hata et al., 2017]

Parameter Estimation



19 reactions: hairpin closing [Bonnet et al. 1998], hairpin opening [Bonnet et al. 1998], helix association [Hata et al., 2017] and helix dissociation (with mismatches) [Cisse et al., 2012]

Summary

- We address the following tasks in accurately estimating reaction kinetics in the full state space of all non-pseudoknotted secondary structures
- MFPT estimation
  - We show how to use a reduced variance stochastic simulation algorithm (RVSSA)
- Parameter estimation for DNA kinetic models based on MFPT estimates
  - We introduce a fixed path ensemble inference (FPEI) approach
- Computational experiments on real data by augmenting the Multistrand kinetic simulator
  - RVSSA is useful when states of paths between initial and final states have large expected holding times
  - FPEI is useful when the number of unique states of the fixed paths between initial and final states is significantly smaller than the length of the paths, and is applicable if the fixed condensed paths can be generated in a timely manner