SAND2020-8447C

Multifidelity Sequential Tempered Markov Chain Monte Carlo for Bayesian Inference

Thomas A. Catanach

MCQMC 2020

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND2020-1895 PE

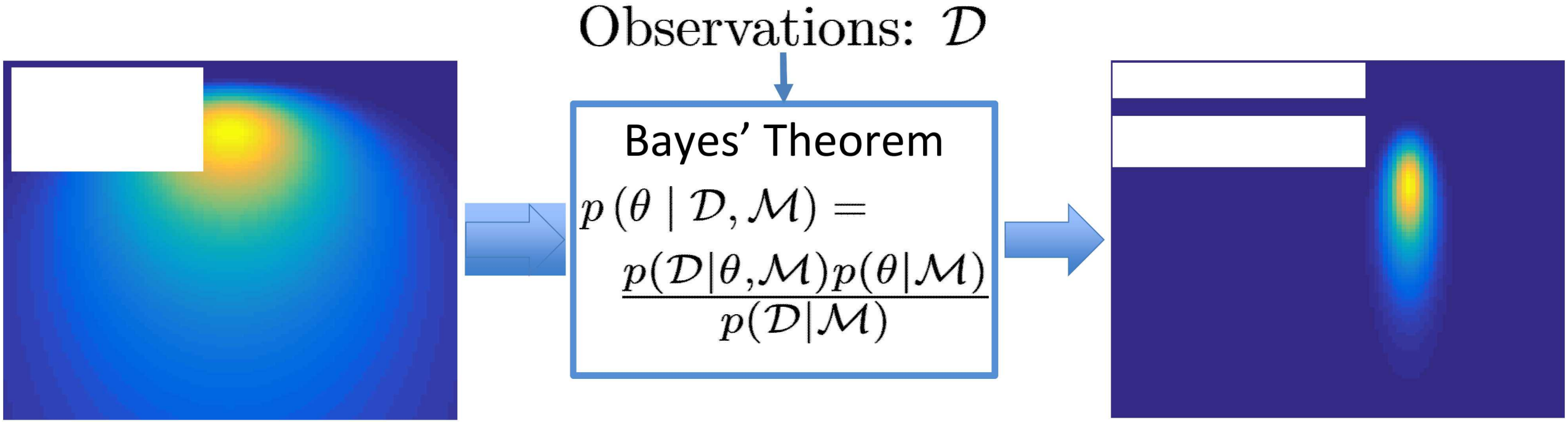
Overview

- Overview of Bayesian Inference and MCMC
- Sequential Tempered MCMC
- Multifidelity ST-MCMC
- Identifying gene regulatory networks
- Conclusion

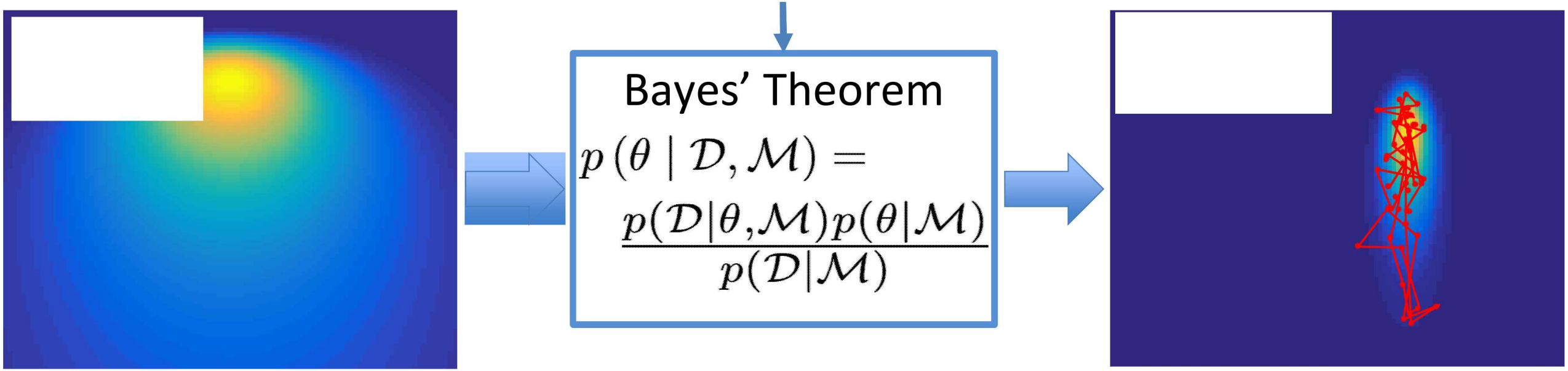
Motivation

- The Bayesian Perspective:
  - Probability distributions quantify uncertainty due to insufficient information
- Bayesian methods for identification and estimation are critical to the robust system analysis
- Yet, Bayesian methods like MCMC are often computationally very expensive since the require many model evaluations
- Often computational models exist with varying fidelities, which can be used to speed up MCMC
- How do we know when to use each model in a fidelity hierarchy?

The Bayesian Inference Problem



The Bayesian Inference Problem



Posterior Estimation: $$\mathbb{E}\left[g\left(\theta\right) \mid \mathcal{D}, \mathcal{M}\right] = \int g\left(\theta\right) p\left(\theta \mid \mathcal{D}, \mathcal{M}\right) d\theta \approx \frac{1}{N} \sum_{i=1}^{N} g\left(\theta_i\right)$$

Model Evidence: $$p\left(\mathcal{D} \mid \mathcal{M}\right) = \int p\left(\mathcal{D} \mid \theta, \mathcal{M}\right) p\left(\theta \mid \mathcal{M}\right) d\theta$$

Quantifying Learning

- Quantifying the accuracy of a quantity of interest
- Posterior Estimation:

$$\mathbb{E}\left[g\left(\theta\right) \mid \mathcal{D}, \mathcal{M}\right] = \int g\left(\theta\right) p\left(\theta \mid \mathcal{D}, \mathcal{M}\right) d\theta \approx \frac{1}{N} \sum_{i=1}^{N} g\left(\theta_i\right)$$

- Effective Number of Samples

$$ESS\left[g\left(\theta_{1:N}\right)\right] = \frac{var\left[g\left(\theta\right)\right]}{var\left[\frac{1}{N}\sum_{i=1}^{N} g\left(\theta_i\right)\right]}$$

- Information quantifies how belief $q(\theta)$ change to $p(\theta)$ with respect to a state of belief $r(\theta)$:

$$\mathcal{I}_{r(\theta)}\left[p(\theta) \mid\mid q(\theta)\right] = \int r(\theta) \log \frac{p(\theta)}{q(\theta)} d\theta$$

- Quantifying changes in belief due to inference:

$$\mathcal{I}_{p(\theta|\mathcal{D},\psi,\mathcal{M})}\left[p(\theta \mid \mathcal{D}, \psi, \mathcal{M}) \mid\mid p(\theta \mid \psi, \mathcal{M})\right] =$$

$$\mathrm{KL}\left[p(\theta \mid \mathcal{D}, \psi, \mathcal{M}) \mid\mid p(\theta \mid \psi, \mathcal{M})\right] = \int p(\theta \mid \mathcal{D}, \psi, \mathcal{M}) \log \frac{p(\theta \mid \mathcal{D}, \psi, \mathcal{M})}{p(\theta \mid \psi, \mathcal{M})} d\theta$$

J. A. Duersch and T. A. Catanach. Generalizing information to the evolution of rational belief. *Entropy*, 22(1):108, 2020.

Solving Bayesian Inference and Uncertainty Quantification Problems

Sequential Tempered MCMC (ST-MCMC)

- ST-MCMC methods use parallel chains that interact with each other to speed up convergence
- ST-MCMC methods evolve to the posterior through a series of intermediate distributions
- ST-MCMC is able to solve Bayesian parameter estimation and model selection problems effectively

Sequential Tempered MCMC

- ST-MCMC methods combine:
  1) Annealing: Introduce intermediate distributions
  2) MCMC: Explore the intermediate distributions
  3) Importance Resampling: Discard unlikely chains and multiply likely chains while maintaining the distribution
- Examples: SMC[1,2,3], Subset Simulation[4], TMCMC[5], and AMSSA[6]

[1] Del Moral et al 2006
[2] Cérou et al 2012
[3] Kantas et al 2014
[4] S.K. Au and J.L. Beck 2001
[5] J. Ching and Y. C. Chen 2007
[6] E. Prudecio and S.H. Cheung 2012

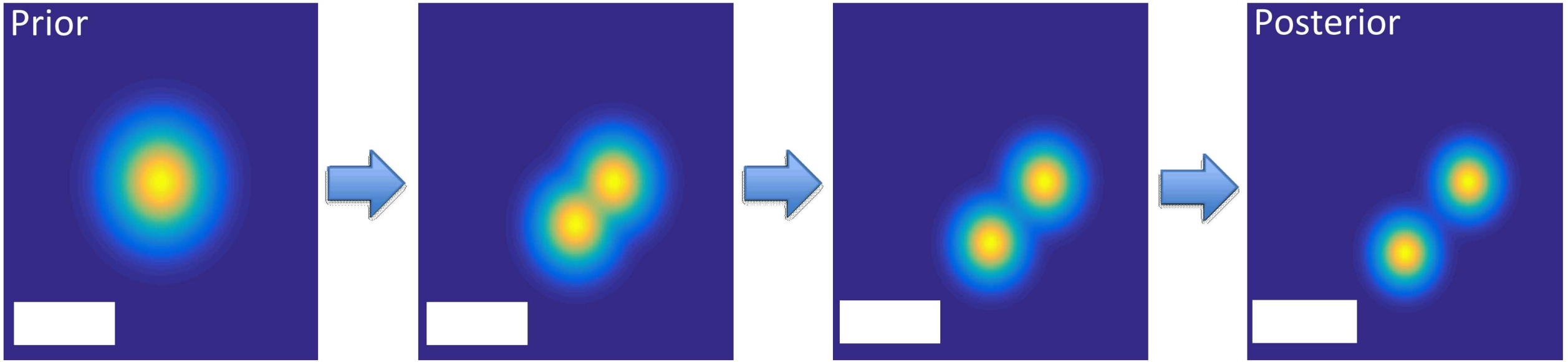
Annealing

β defines how much the data updates the intermediate distribution:

$$\pi_i(\theta) \propto p(\mathcal{D} \mid \theta, \mathcal{M})^{\beta_i} \, p(\theta \mid \mathcal{M}) \quad \beta_i \in [0, 1]$$

Intermediate distributions at different β levels

Level 0: $\beta_0 = 0$ | Level 1: $\beta_1 = \beta_0 + \Delta\beta_1$ | Level 2: $\beta_2 = \beta_1 + \Delta\beta_2$ | Level n: $\beta_n = 1$



Annealing: Finding Δβ

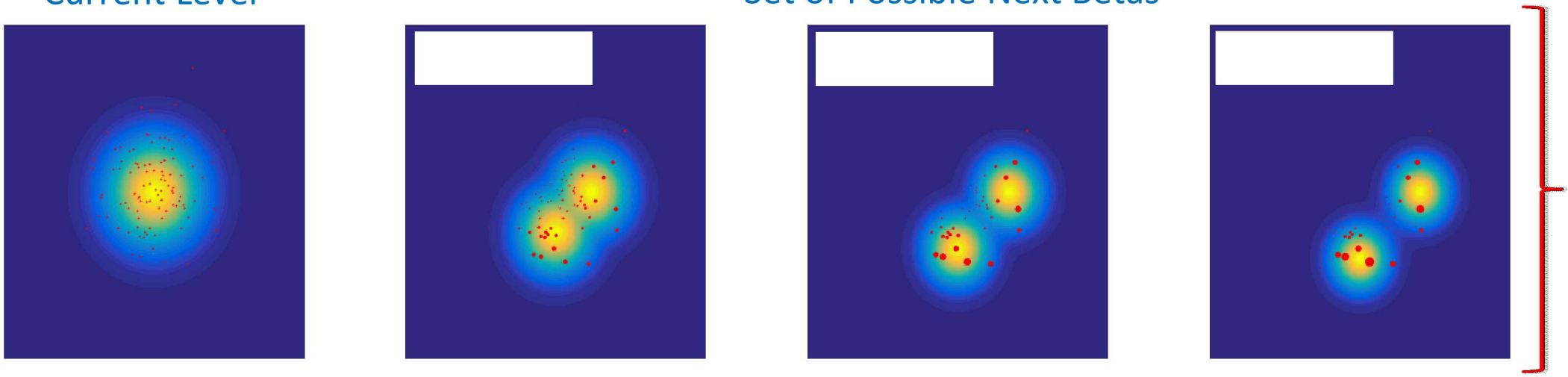
Find Δβ such that the coefficient of variation (κ) of the sample weights is 1

Sample weight: $w(\theta_j) \propto p(\mathcal{D} \mid \theta_j, \mathcal{M})^{\Delta\beta_i}$

Coefficient of variation: $\kappa(w) = \frac{\sigma(w)}{\bar{w}}$

Current Level

Set of Possible Next Betas



Weighted Sample Populations

Importance Resampling

- Resampling the population rebalances the weights as the distribution changes. This discards unlikely samples and replicates likely samples
- Multinomial Resampling from level i-1 to level i:

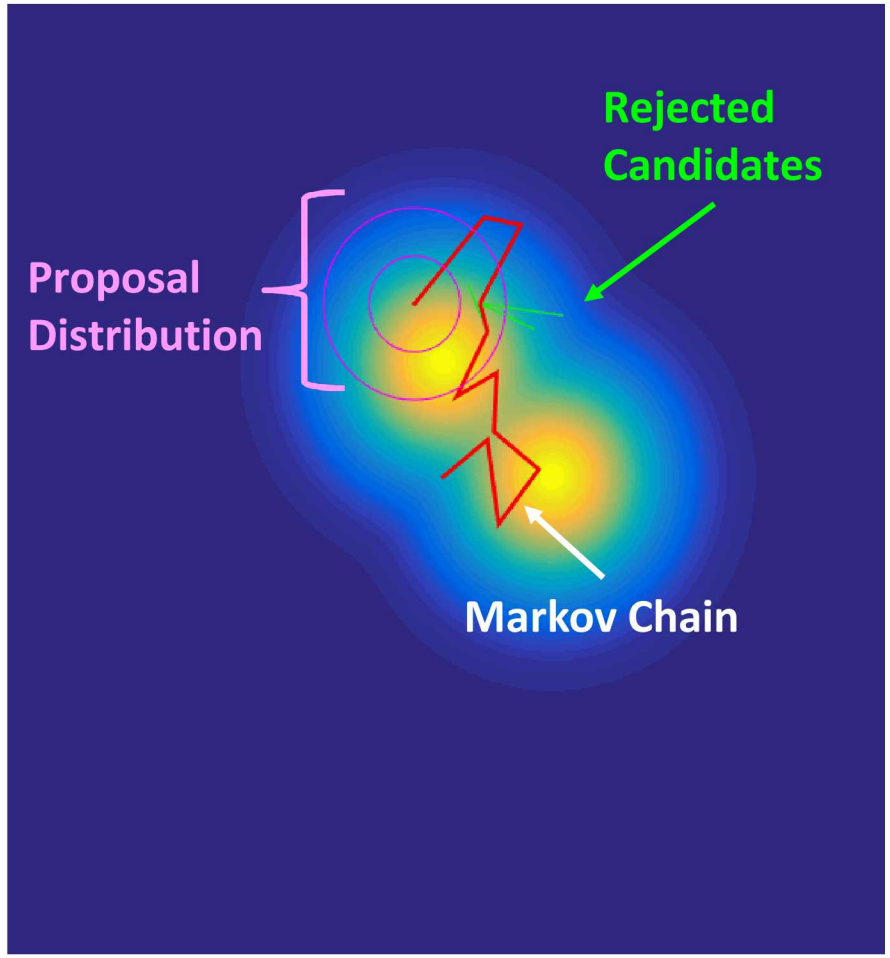
Probability of selecting sample k: $P\left(\theta_{i,j} = \theta_{i-1,k}\right) = w\left(\theta_{i-1,k}\right)$

Sample weight: $w\left(\theta_{i-1,j}\right) \propto p\left(\mathcal{D} \mid \theta_{i-1,j}, \mathcal{M}\right)^{\Delta\beta_i}$

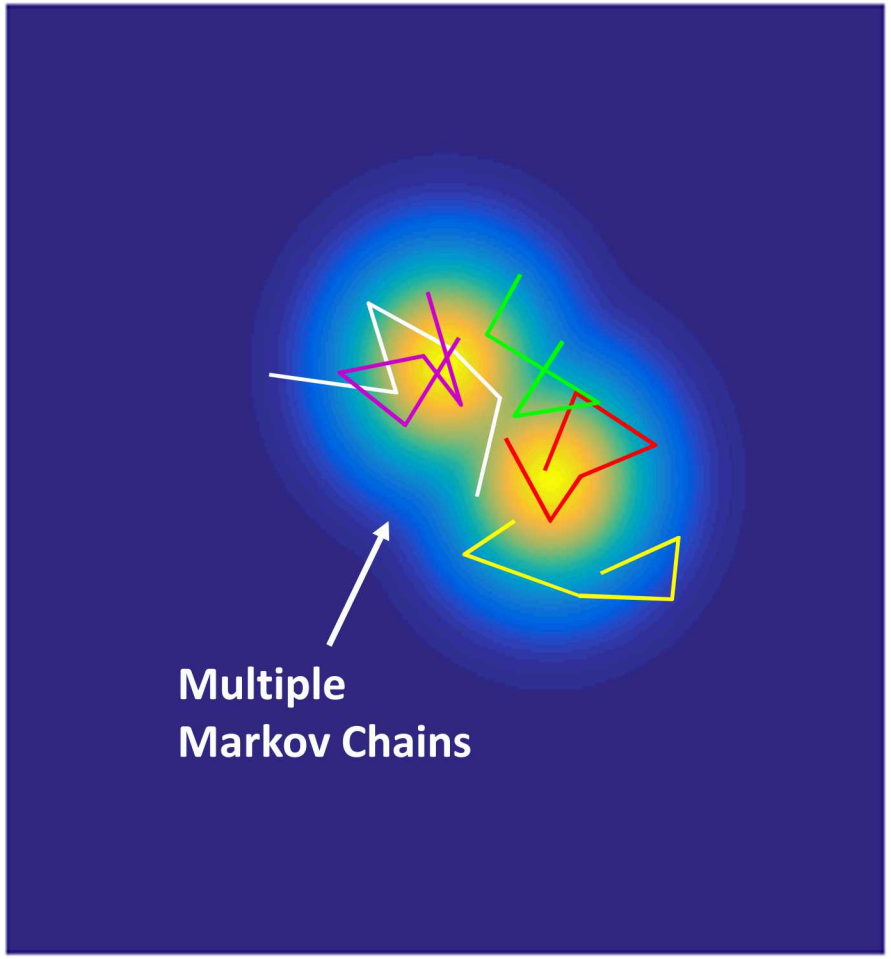
Metropolis Hastings MCMC with Parallel Chains

Sandia National Laboratories

Single MH Markov Chain



Parallel MH Markov Chain



Estimating Model Evidence using ST-MCMC

Model Class Probability:

$$p(\mathcal{M} \mid z) = \frac{p(z \mid \mathcal{M})\, p(\mathcal{M})}{p(z)} \propto \left(\int p(z \mid \theta, \mathcal{M})\, p(\theta \mid \mathcal{M})\, d\theta \right) p(\mathcal{M})$$

Model Class Evidence:

$$\int p(z \mid \theta, \mathcal{M})\, p(\theta \mid \mathcal{M})\, d\theta = \prod_{i=1}^{l} c_i$$

Decompose into evidence for intermediate distributions

ith Distribution Evidence Estimate:

$$c_i = \int p(z \mid \theta, \mathcal{M})^{\Delta\beta_i} \frac{p(z \mid \theta, \mathcal{M})^{\beta_{i-1}}\, p(\theta \mid \mathcal{M})}{\prod_{j=1}^{i-1} c_j} d\theta \approx \frac{1}{N} \sum_{k=1}^{N} p(z \mid \theta_{i-1,k}, \mathcal{M})^{\Delta\beta_i}$$

Level i Likelihood

Level i prior

Monte Carlo Estimate

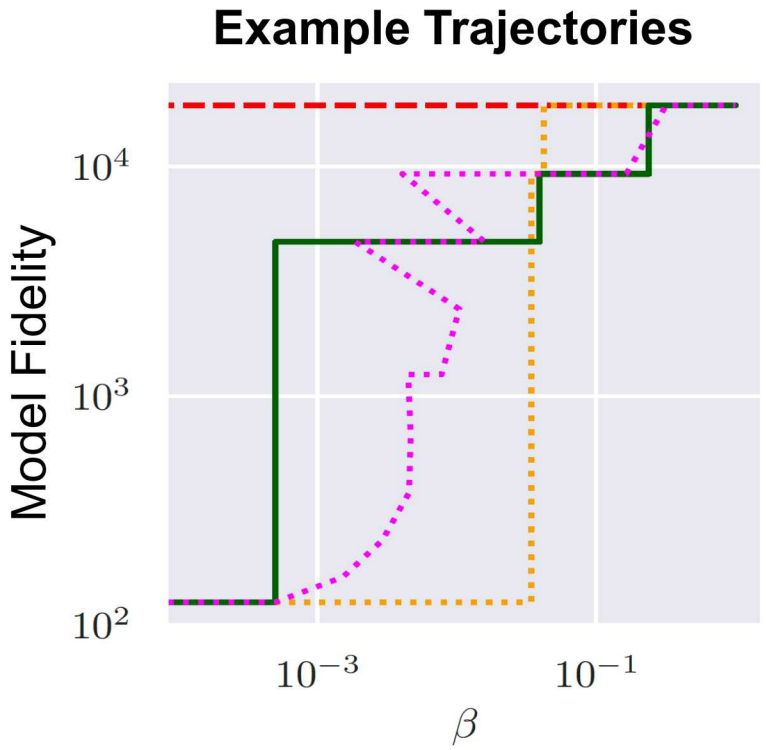
Multifidelity ST-MCMC

- Models in a multifidelity hierarchy can define intermediate levels just like the annealing factor[1]

$$\pi_{\beta_l}\left(\theta \mid M_{m_l}\right) = \frac{p\left(\mathcal{D} \mid \theta, M_{m_l}\right)^{\beta_l} p\left(\theta\right)}{p\left(\mathcal{D} \mid M_{m_l}, \beta_l\right)}$$

- Tempering: Adjust the annealing factor
- Bridging: Change model fidelity
- Reweighting

$$w_i = \frac{\pi_{\beta_{l+1}}\left(\theta_i^l \mid M_{m_{l+1}}\right)}{\pi_{\beta_l}\left(\theta_i^l \mid M_{m_l}\right)} = \frac{p\left(\mathcal{D} \mid \theta_i^l, M_{m_{l+1}}\right)^{\beta_{l+1}}}{p\left(\mathcal{D} \mid \theta_i^l, M_{m_l}\right)^{\beta_l}}$$



[1]Latz, Jonas, Iason Papaioannou, and Elisabeth Ullmann. "Multilevel sequential[2] Monte Carlo for Bayesian inverse problems." *Journal of Computational Physics* 368 (2018): 154-178.

General Multifidelity ST-MCMC Algorithm

Inputs : Prior distribution $p(\theta)$
Model fidelity hierarchy $\mathcal{M} = \{M_j : j = 1 \dots K\}$
Likelihood function $p(\mathcal{D} \mid \theta, M_j)$
Number of samples N
Output : Posterior samples $\theta_{1\dots N}$

begin
Initialize: Set level counter $l = 0$
Set annealing factor $\beta_0 = 0$
Set model level $m_0 = 1$
Define the first intermediate distribution $\pi_0(\theta) = p(\theta)$
Draw initial samples $\theta^0_{1\dots N} \sim \pi_0(\theta)$
while $\pi_{\beta_l}(\theta \mid M_{m_l}) \neq p(\theta \mid \mathcal{D}, M_K)$ **do**
1) Increment level counter $l = l + 1$
2) Choose the next β_l and m_l based on the previous level sample population $\theta^{l-1}_{1\dots N}$
3) Define the next intermediate distribution $\pi_{\beta_l}(\theta \mid M_{m_l}) \propto p(\mathcal{D} \mid \theta, M_{m_l})^{\beta_l} p(\theta)$
4) Compute the unnormalized importance weights for the population as $w_i = \frac{\pi_{\beta_l}(\theta_i^{l-1} \mid M_{m_l})}{\pi_{\beta_{l-1}}(\theta_i^{l-1} \mid M_{m_{l-1}})}$
5) Resample the population according to the normalized importance weights to initialize $\theta^l_{1\dots N}$
6) Evolve the samples $\theta^l_{1\dots N}$ using MCMC with stationary distribution $\pi_{\beta_l}(\theta \mid M_{m_l})$
end
return $\theta_{1\dots N} = \theta^l_{1\dots N}$
end

Define the model hierarchy

Identify the annealing factor and model fidelity for the next level of ST-MCMC

Transform the sample population

Bridging/Tempering Strategies

Multilevel Sequential^2 Monte Carlo (ESS Bridging)

1. Try a Bridging Step:

$$\beta_l, M_{m_l} \rightarrow \beta_{l+1} = \beta_l, M_{m_{l+1}} = M_{m_l+1}$$

2. Compute the population weights

$$w_i = \frac{p\left(\mathcal{D} \mid \theta_i^l, M_{m_l+1}\right)^{\beta_l}}{p\left(\mathcal{D} \mid \theta_i^l, M_{m_l}\right)^{\beta_l}}$$

3. Temper instead if ESS of the weighted population is too low

$$\beta_l, M_{m_l} \rightarrow \beta_{l+1} = \beta_l + \Delta\beta, M_{m_{l+1}} = M_{m_l}$$

Information Theoretic Bridging

1. Try a Tempering Step

$$\beta_l, M_{m_l} \rightarrow \beta_{l+1} = \beta_l + \Delta\beta, M_{m_{l+1}} = M_{m_l}$$

2. Determine if this step gains or looses information about the posterior

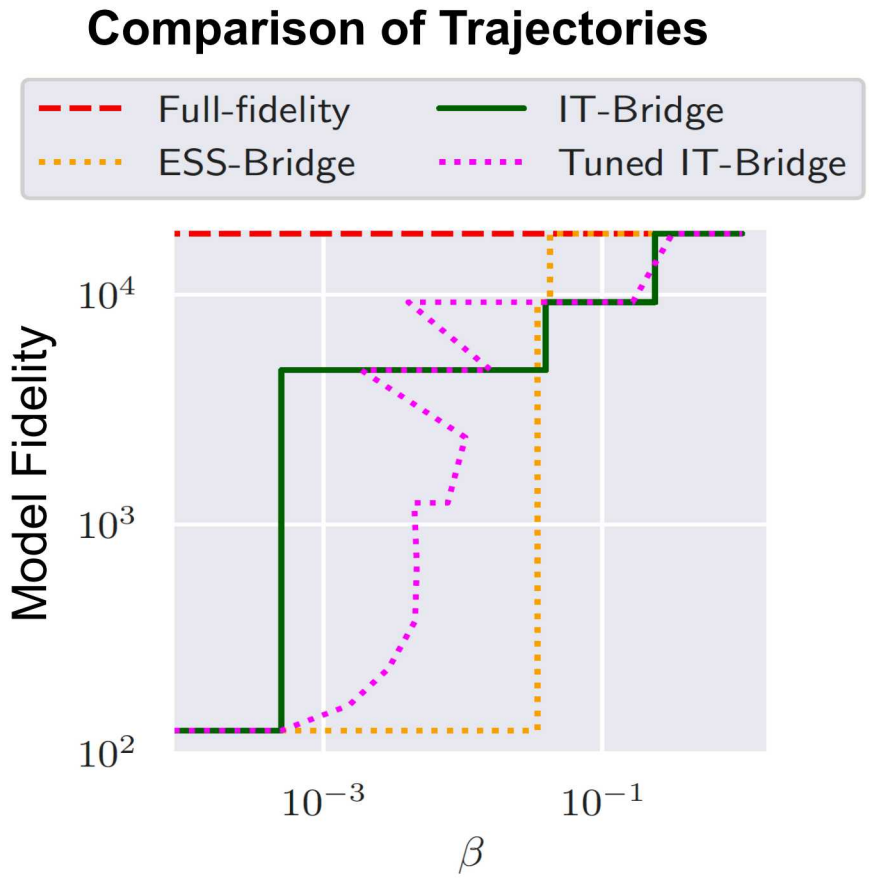
$$\mathcal{I}_{p(\theta|\mathcal{D},M_K)}\left[p\left(\theta \mid \mathcal{D}, M_{m_l}, \beta_{l+1}\right) || p\left(\theta \mid \mathcal{D}, M_{m_l}, \beta_l\right)\right]$$

3. Bridge if looses information

Bridging/Tempering Strategies

- ESS-Bridging requires only evaluating models at the current and next level of the hierarchy
- IT-Bridging additionally requires some full model evaluations
- After bridging, the annealing factor can also be tuned to an ESS target (Tuned IT-Bridge)

$$\beta_l, M_{m_l} \rightarrow \beta_{l+1} = \beta_l + \Delta\beta, M_{m_{l+1}} = M_{m_l+1}$$



Information Theoretic Criteria

- To determine if information is lost, we need to estimate the sign of

$$\mathcal{I}_{p(\theta|\mathcal{D},M_K)}\left[p\left(\theta \mid \mathcal{D}, M_{m_l}, \beta_{l+1}\right) || p\left(\theta \mid \mathcal{D}, M_{m_l}, \beta_l\right)\right]$$

$$= \int p\left(\theta \mid \mathcal{D}, M_K\right) \log p\left(\mathcal{D} \mid \theta, M_{m_l}\right)^{\Delta\beta} \frac{p\left(\mathcal{D} \mid M_{m_l}, \beta_l\right)}{p\left(\mathcal{D} \mid M_{m_l}, \beta_{l+1}\right)} d\theta$$

$$\propto \int p\left(\theta \mid \mathcal{D}, M_{m_l}, \beta_l\right) \frac{p\left(\mathcal{D} \mid \theta, M_k\right)}{p\left(\mathcal{D} \mid \theta, M_{m_l}\right)^{\beta_l}} \log \frac{p\left(\mathcal{D} \mid \theta, M_{m_l}\right)^{\Delta\beta}}{\mathrm{E}_{\theta|\mathcal{D},M_{m_l},\beta_l}\left[p\left(\mathcal{D} \mid \theta, M_{m_l}\right)^{\Delta\beta}\right]} d\theta$$

- This can be estimate using samples from $\pi_{\beta_l}\left(\theta \mid M_{m_l}\right)$

$$\mathcal{I}_{criteria} = \mathrm{E}_{\theta|\mathcal{D},M_{m_l},\beta_l}\left[\frac{p\left(\mathcal{D} \mid \theta, M_K\right)}{p\left(\mathcal{D} \mid \theta, M_{m_l}\right)^{\beta_l}} \log p\left(\mathcal{D} \mid \theta, M_{m_l}\right)^{\Delta\beta}\right] -$$

$$\mathrm{E}_{\theta|\mathcal{D},M_{m_l},\beta_l}\left[\frac{p\left(\mathcal{D} \mid \theta, M_K\right)}{p\left(\mathcal{D} \mid \theta, M_{m_l}\right)^{\beta_l}}\right] \log \mathrm{E}_{\theta|\mathcal{D},M_{m_l},\beta_l}\left[p\left(\mathcal{D} \mid \theta, M_{m_l}\right)^{\Delta\beta}\right]$$

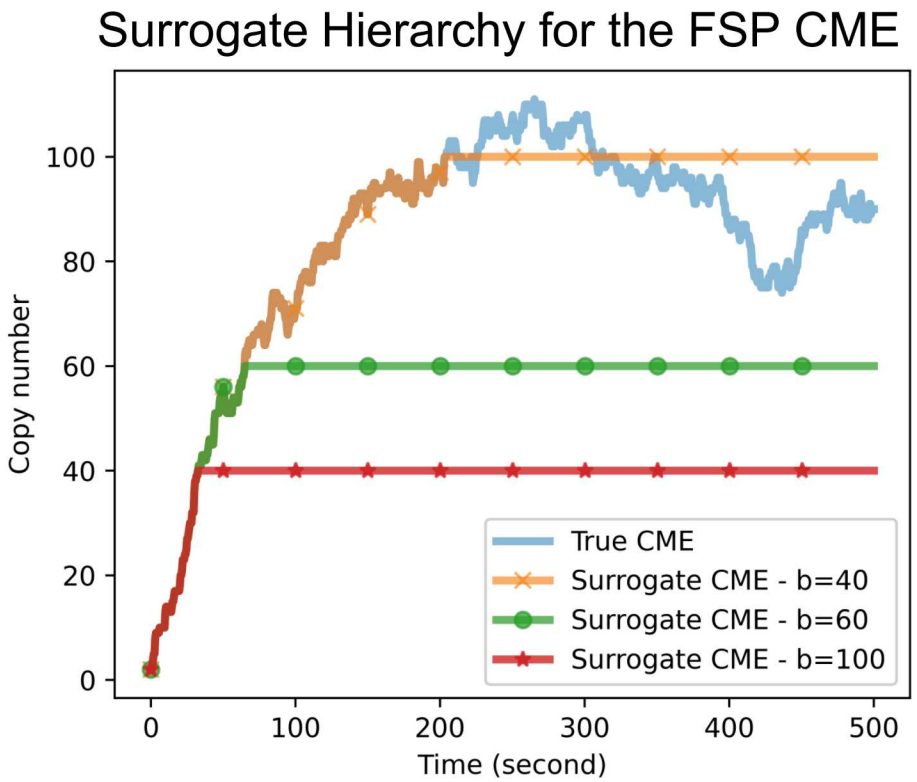
Bayesian Inference of Stochastic Reaction Networks

with Huy Vo and Brian Munsky at Colorado State University

Catanach, Thomas A., Huy D. Vo, and Brian Munsky. "Bayesian inference of Stochastic reaction networks using Multifidelity Sequential Tempered Markov Chain Monte Carlo." arXiv preprint arXiv:2001.01373 (2020). [Accepted IJUQ]

Chemical Master Equation Model

	Stochastic Reaction Networks
Dynamics	$\dot{p} = A(\theta)\, p$
Observations ($\mathcal{D}$)	Chemical Species Copy Number (Molecule Count)
States (p)	Probability of a cell having a certain number of each species (p)
Parameters (θ)	Propensities for certain reactions to occur
Model Hierarchy ($\mathcal{M}$)	Maximum copy number in Finite State Projection CME algorithm



- Limiting the maximum copy number significantly reduces the size of the matrix A(θ) making the simulations faster.

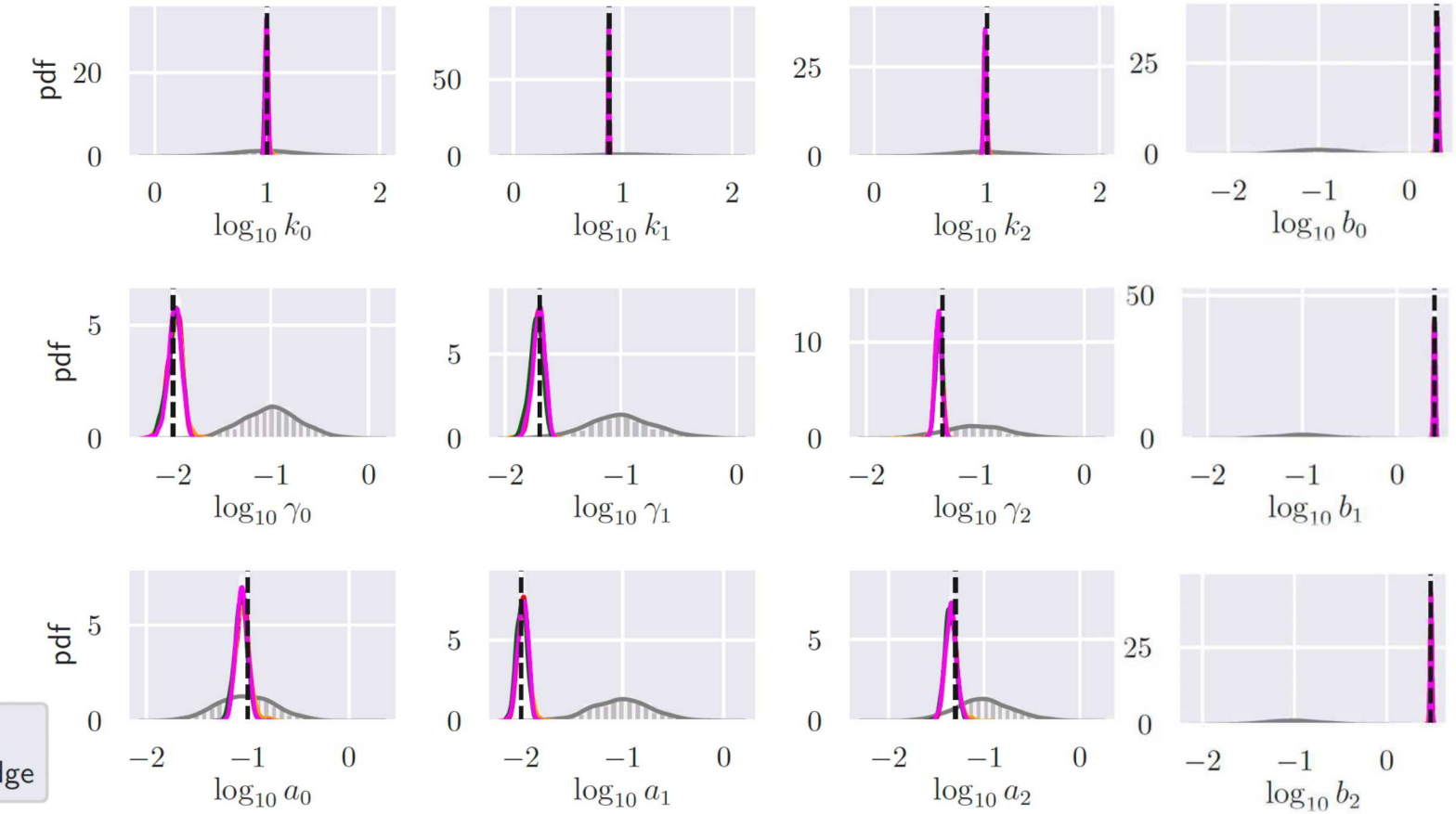
Example 1: Repressilator Gene Circuit

- Theoretical three species gene circuit with negative feedback

Chemical Reactions

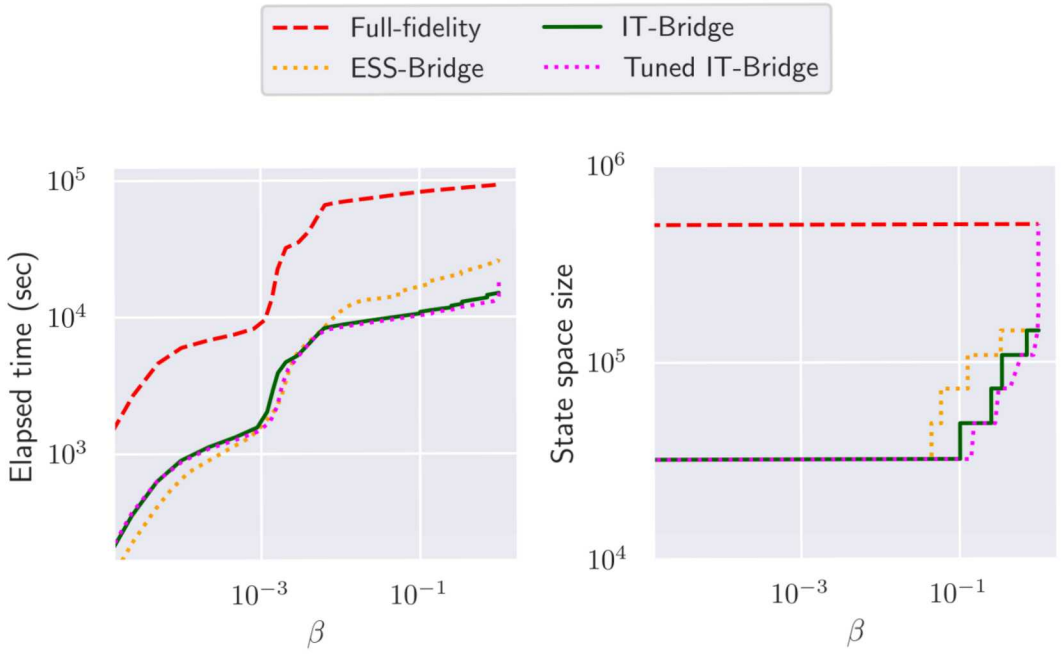
	reaction	propensity
1.	$\emptyset \rightarrow \mathrm{TetR}$	$k_0/(1+a_0[\mathrm{LacI}]^{b_0})$
2.	$\mathrm{TetR} \rightarrow \emptyset$	$\gamma_0[\mathrm{TetR}]$
3.	$\emptyset \rightarrow \lambda\mathrm{cI}$	$k_1/(1+a_1[\mathrm{TetR}]^{b_1})$
4.	$\lambda\mathrm{cI} \rightarrow \emptyset$	$\gamma_1[\lambda\mathrm{cI}]$
5.	$\emptyset \rightarrow \mathrm{LacI}$	$k_2/(1+a_2[\lambda\mathrm{cI}]^{b_2})$
6.	$\mathrm{LacI} \rightarrow \emptyset$	$\gamma_2[\mathrm{LacI}]$

Parameter ID with ST-MCMC

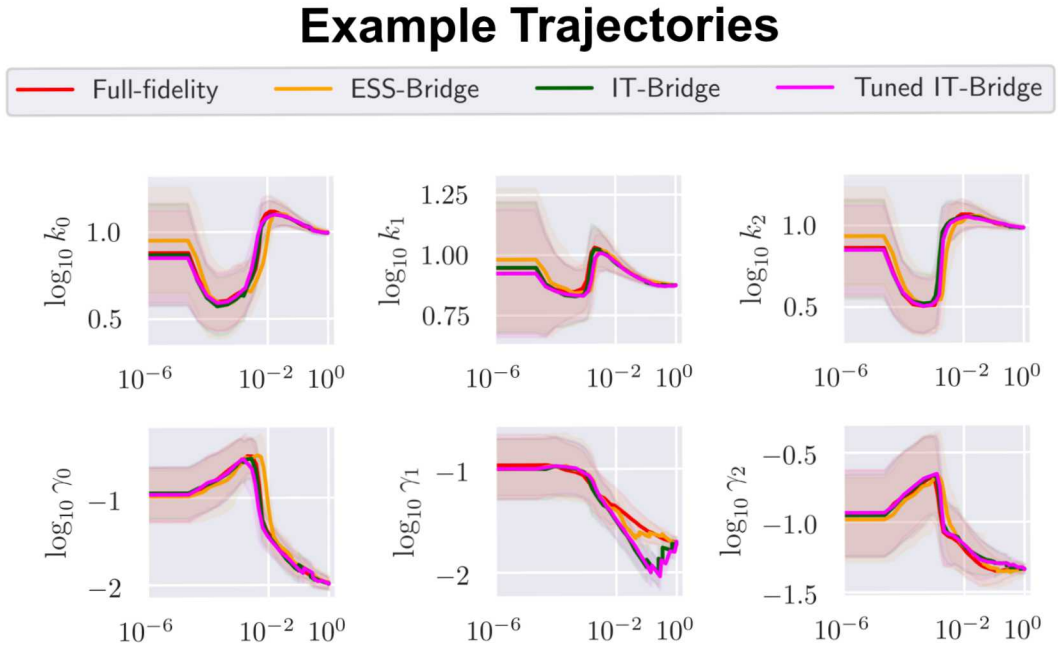


Example 1: Repressilator Gene Circuit

- Multifidelity ST-MCMC provides speed ups of 3.5 (ESS-Bridge), 4.8 (Tuned IT), 6.2 (IT-Bridge)



- The evolution of parameter estimates are very similar



Example 2: Compartmental Model of Gene Expression

- Theoretical Bayesian model selection for a 2, 3, and 4 gene model for the transcription, transportation, and degradation of nuclear mRNA
- Tuned-IT bridging was used because of its robustness at estimating the evidence.
- Speed ups of 1.6, 3.8, and 3.2 for the 2, 3, and 4 gene models respectively

Chemical Reactions

reaction index	reaction	propensity
$1, \ldots, n_G$	$G_{i-1} \to G_i, i = 1, \ldots, n_G - 1$	$k^+_{i-1}[G_{i-1}]$
$n_G + 1, \ldots, 2n_G$	$G_i \to G_{i-1}, i = 1, \ldots, n_G - 1$	$k^-_i[G_i]$
$2n_G + 1, \ldots, 3n_G - 1$	$G_i \to G_i + \mathrm{RNA}_{nuc}, i = 1, \ldots, n_G - 1$	$r_i[G_i]$
$3n_G$	$\mathrm{RNA}_{nuc} \to \mathrm{RNA}_{cyt}$	$k_{trans}[\mathrm{RNA}_{nuc}]$
$3n_G + 1$	$\mathrm{RNA}_{cyt} \to \emptyset$	$\gamma[\mathrm{RNA}_{cyt}]$

Model Evidence

Model	Full-fidelity		Tuned IT Bridge	
	Log Evidence	Time (Sec)	Log Evidence	Time (Sec)
2 Gene	-20108.8 ± 5.4	1244	-20112.6 ± 2.0	758
3 Gene	-20111.9 ± 5.6	67496	-20115.7 ± 2.0	17511
4 Gene	-20113.0 ± 5.7	76546	-20117.5 ± 1.8	23777

Example 3: Inflammation response gene IL1beta

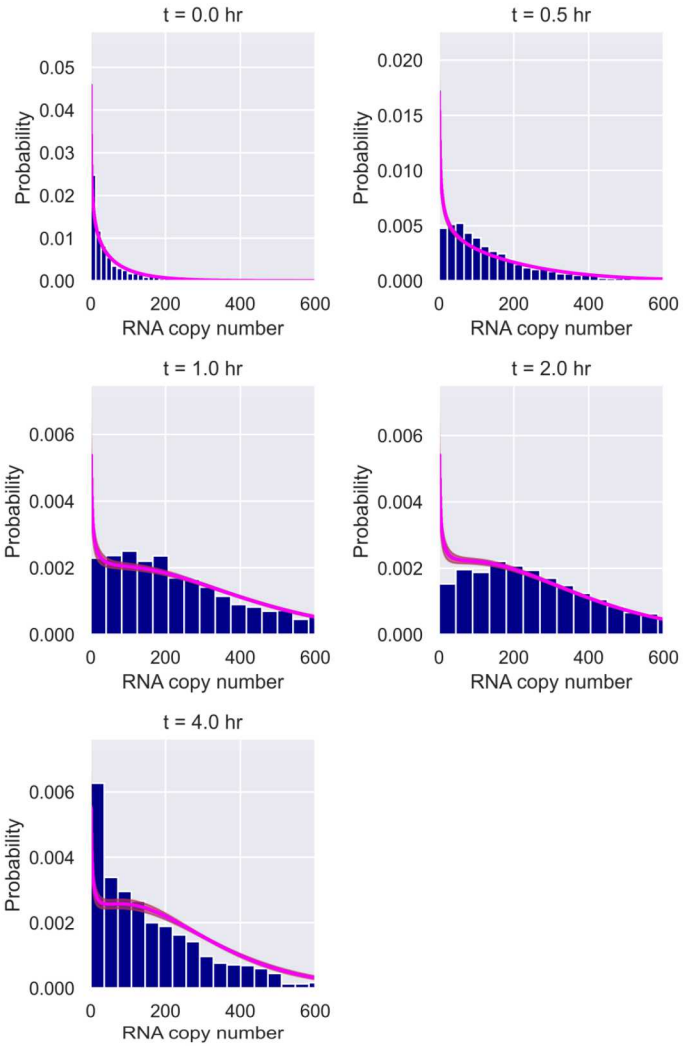
- Three state gene expression model for inflammation response IL1beta dataset[1]

Chemical Reactions

	reaction	propensity
1.	$G_0 \rightarrow G_1$	$k_{01}[G_0]$
2.	$G_1 \rightarrow G_2$	$k_{12}[G_1]$
3.	$G_2 \rightarrow G_1$	$k_{21}[G_2]$
4.	$G_1 \rightarrow G_0$	$k_{10}(t) = \max\{0, a_{10} - b_{10}S(t)\}$
5.	$\emptyset \rightarrow \text{RNA}$	$\alpha_1[G_1] + \alpha_2[G_2]$
6.	$\text{RNA} \rightarrow \emptyset$	$\gamma[\text{RNA}]$

$$S(t) = \max\{0, \exp(-r_1(t - T_0))(1 - \exp(-r_2(t - T0)))\}$$

Data Fits



[1]Kalb, Daniel M., et al. "Single-cell correlations of mRNA and protein content in a human monocytic cell line after LPS stimulation." *PloS one* 14.4 (2019): e0215602.

Example 3: Inflammation response gene IL1beta

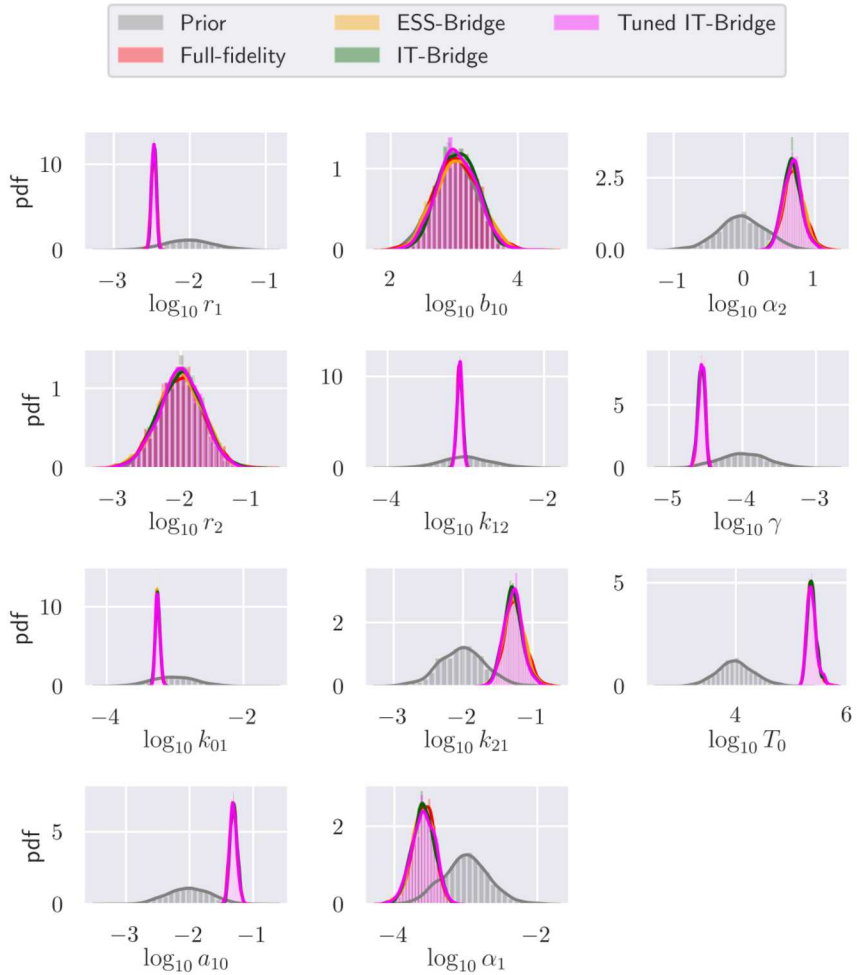
- Three state gene expression model for inflammation response IL1beta dataset[1]

Chemical Reactions

	reaction	propensity
1.	$G_0 \rightarrow G_1$	$k_{01}[G_0]$
2.	$G_1 \rightarrow G_2$	$k_{12}[G_1]$
3.	$G_2 \rightarrow G_1$	$k_{21}[G_2]$
4.	$G_1 \rightarrow G_0$	$k_{10}(t) = \max\{0, a_{10} - b_{10}S(t)\}$
5.	$\emptyset \rightarrow \text{RNA}$	$\alpha_1[G_1] + \alpha_2[G_2]$
6.	$\text{RNA} \rightarrow \emptyset$	$\gamma[\text{RNA}]$

$$S(t) = \max\{0, \exp(-r_1(t - T_0))(1 - \exp(-r_2(t - T0)))\}$$

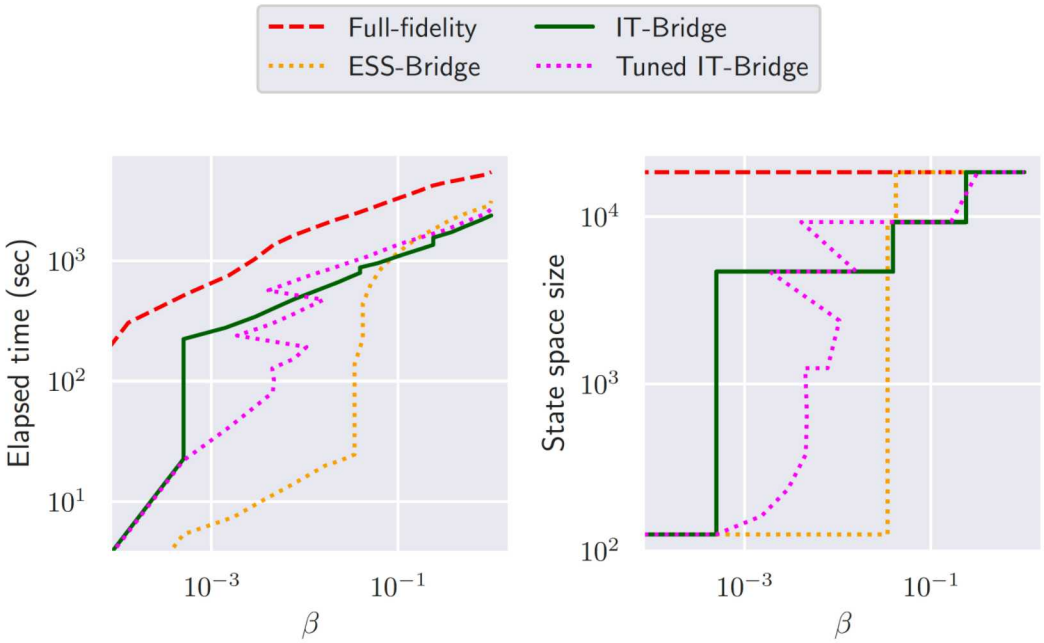
Parameter ID with ST-MCMC



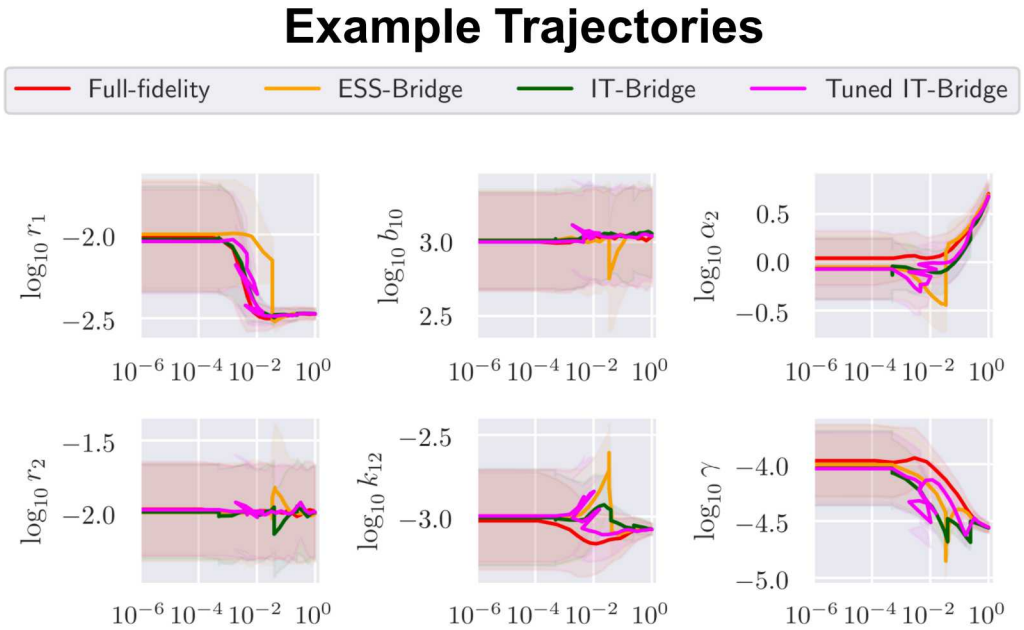
[1]Kalb, Daniel M., et al. "Single-cell correlations of mRNA and protein content in a human monocytic cell line after LPS stimulation." *PloS one* 14.4 (2019): e0215602.

Example 3: Inflammation response gene IL1beta

- Multifidelity ST-MCMC provides speed ups of 1.7 (ESS-Bridge), 2.0 (Tuned IT), 2.2 (IT-Bridge)



- There is more variation in the trajectories but they still follow similar trends



Conclusion

- Multifidelity Sequential Tempered MCMC methods are able to solving Bayesian system Identification and model selection.

- Information Theoretic approaches can be used to identify when a model in the multifidelity hierarchy no longer is informative in ST-MCMC

- Multifidelity CME models can be used to accelerate identification of gene regulatory networks