Graphical Models for Stochastic Verification and Synthesis

Christopher James Langmead
Department of Computer Science &
Lane Center for Computational Biology
Carnegie Mellon University

Research Goals

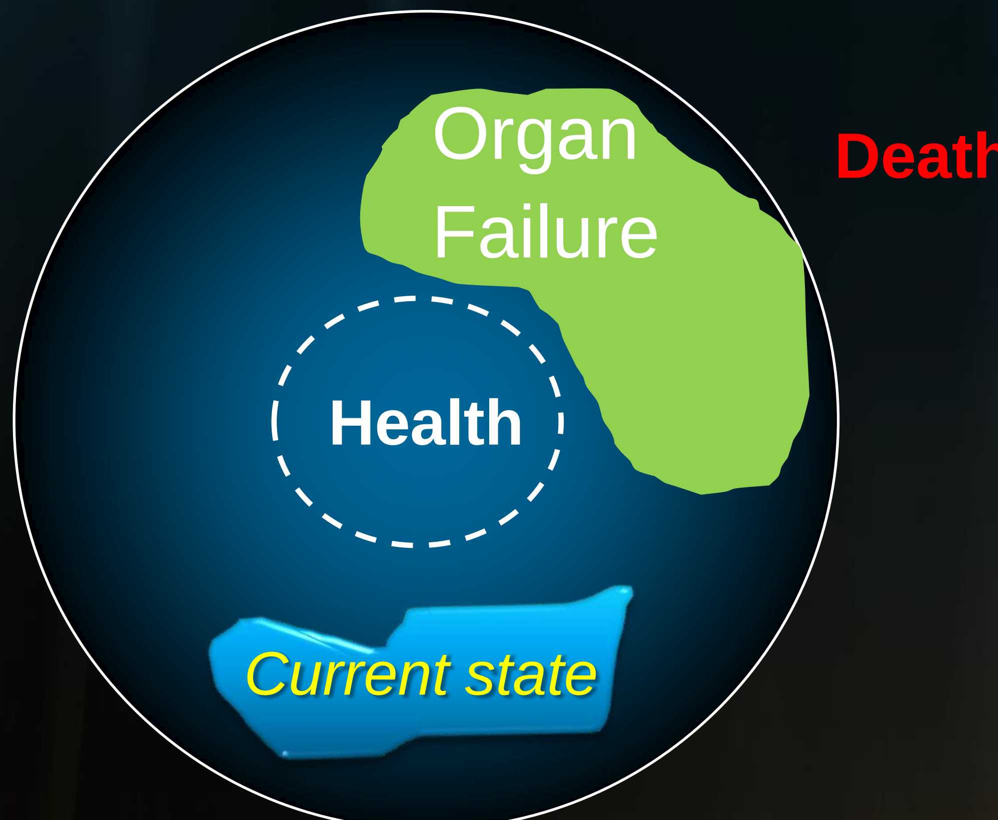
- Develop new methods for reasoning about stochastic processes by adapting and combining methods from Machine Learning and Formal Verification
  - Machine Learning addresses uncertainty
  - Verification addresses complexity and scalability
- Application Domains:
  - Computational Biology
  - Embedded Systems

Context & Motivation

- Personalized Medicine
  - Developing patient-specific treatment plans

Disease Phase Space

Intensive Care Unit



Context & Motivation

- Personalized Medicine
  - Developing patient-specific treatment plans

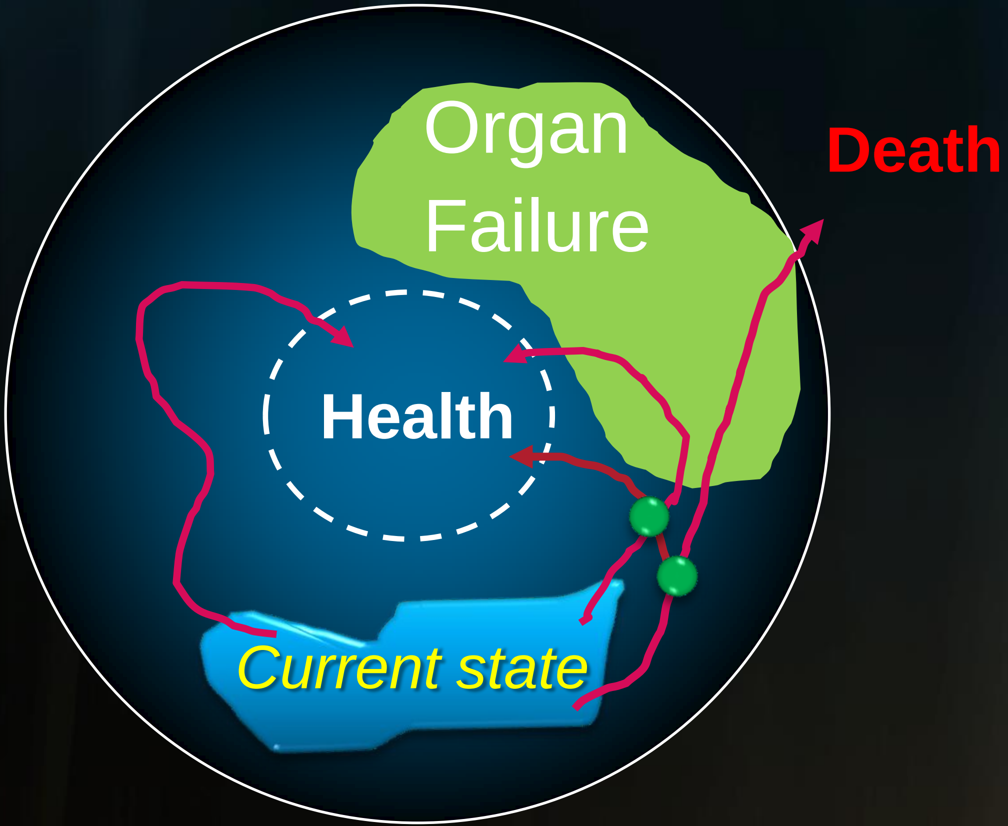
Disease Phase Space

Primary Tasks:

(1) Determine where the patient is now (approximately)

(2) Characterize the patient's trajectory (approximately)

(3) Select interventions based on (1) and (2)



Context & Motivation

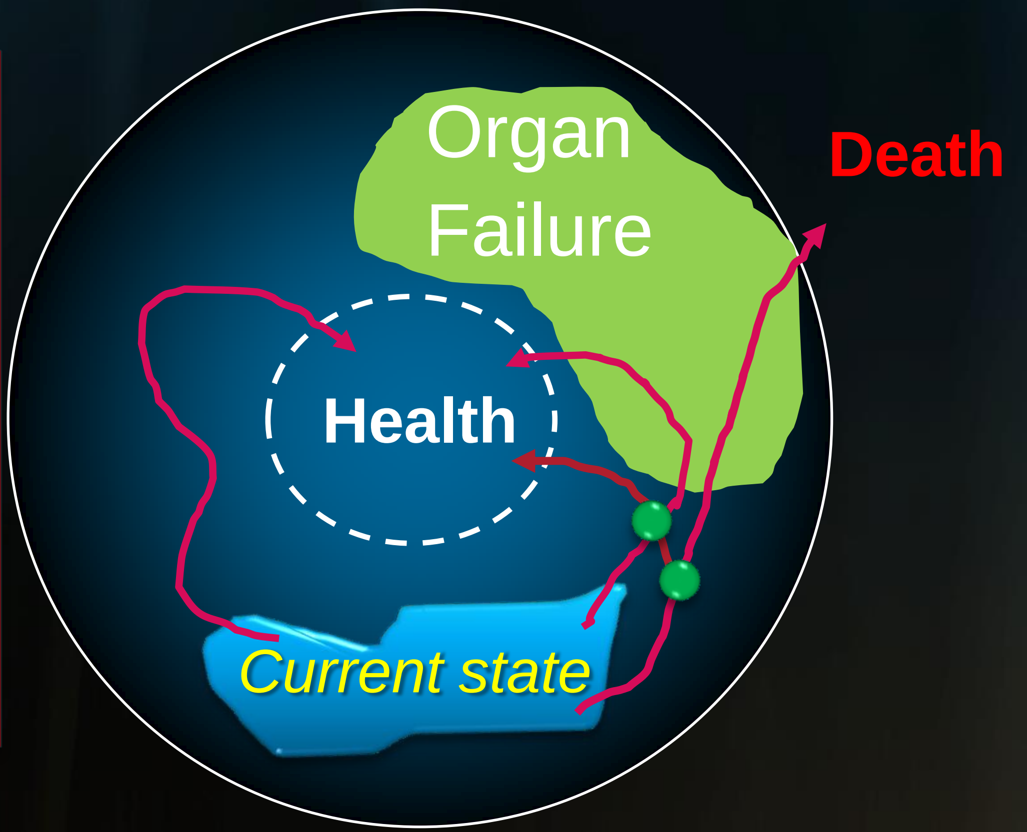
- Physicians routinely use simple (non-dynamic) models in these tasks
  - We believe that dynamic models will be more useful

Primary Tasks:

(1) Determine where the patient is now (approximately)

(2) Characterize the patient's trajectory (approximately)

(3) Select interventions based on (1) and (2)



Context and Motivation

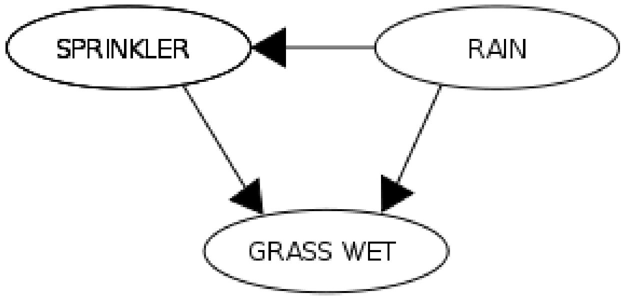
- Dynamic models of disease processes
  - (Stochastic) ODEs/PDEs
  - Graphical models

Graphical Models

- Let $\mathbf{X} = \{x_1, \ldots, x_n\}$ be a set of random variables
  - Each X_i can be continuous or discrete
- A probabilistic graphical model (PGM) is a factored encoding of $P(\mathbf{X})$
  - $\mathcal{M} = (G, \Psi, \Theta)$
    - $G = (V,E)$ is a graph over the random variables
      - V_i corresponds to x_i
      - Edges reveal conditional independencies
    - Ψ is a set of functions over V and E
    - Θ is a vector of parameters

Example

	SPRINKLER	
RAIN	T	F
F	0.4	0.6
T	0.01	0.99



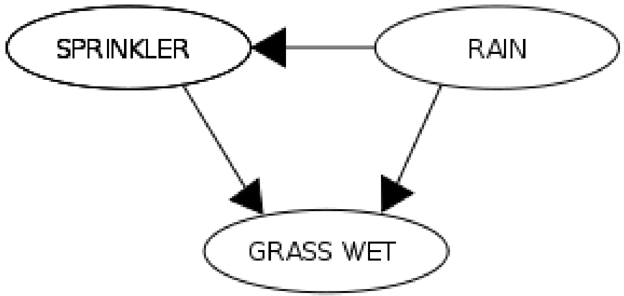
RAIN	
T	F
0.2	0.8

		GRASS WET	
SPRINKLER	RAIN	T	F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

- This models P(Sprinkler, Rain, Grass Wet)
  - V = {Sprinkler, Rain, Grass Wet}
  - Ψ = { P(rain), P(Sprinkler | Rain), P(Grass Wet | Sprinkler, Rain) }
  - Θ = The elements in the 3 tables

Example

RAIN	SPRINKLER T	SPRINKLER F
F	0.4	0.6
T	0.01	0.99



RAIN T	RAIN F
0.2	0.8

SPRINKLER	RAIN	GRASS WET T	GRASS WET F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

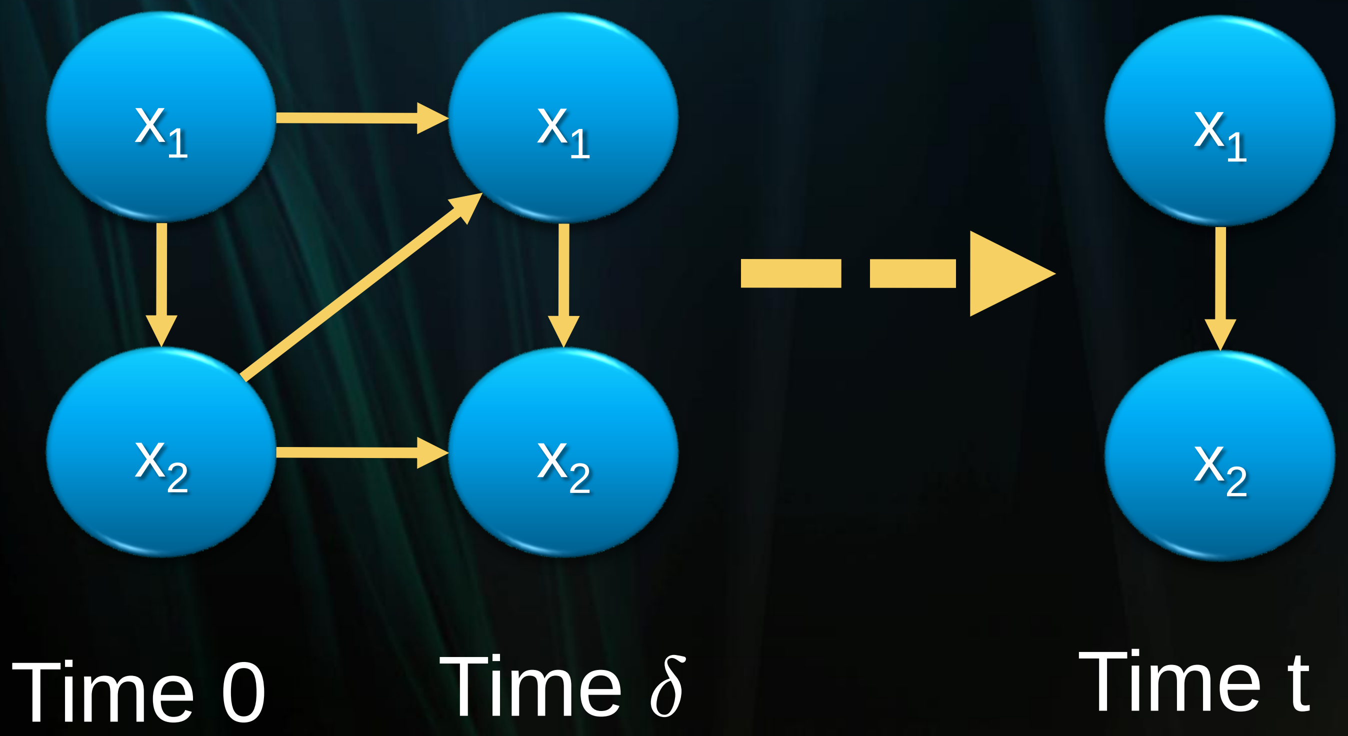
What is the probability that it is raining, given that the grass is wet?

$$\mathrm{P}(R=T \mid G=T) = \frac{\mathrm{P}(G=T, R=T)}{\mathrm{P}(G=T)} = \frac{\sum_{S\in\{T,F\}} \mathrm{P}(G=T, S, R=T)}{\sum_{S,R\in\{T,F\}} \mathrm{P}(G=T, S, R)}$$

$$= \frac{(0.99 \times 0.01 \times 0.2 = 0.00198_{TTT}) + (0.8 \times 0.99 \times 0.2 = 0.1584_{TFT})}{0.00198_{TTT} + 0.288_{TTF} + 0.1584_{TFT} + 0_{TFF}} \approx 35.77\%.$$

PGMs for Stochastic Processes

- Models distributions over time series $P(X_{0:t})$
  - Time can be continuous or discrete



PGMs for Stochastic Processes

- Traditional Tasks
  - Inference
    - Computing $P(A_{0:t+\delta} \mid B_{0:t}; \mathcal{M})$
      - $A \cap B = \varnothing, \; A \cup B = \mathbf{X}$
  - Learning
    - Computing $\text{argmax}_{\theta} P(\mathcal{D}; \mathcal{M})$
      - $\mathcal{D}$ is a set of observations over $\mathbf{Y}_{0:t} \subseteq \mathbf{X}_{0:t}$
  - Structure Learning
    - Computing $\text{argmax}_{G,\theta} P(\mathcal{D}; \mathcal{M})$
      - I.e., simultaneously learning graph topology and parameters

PGMs for Stochastic Processes

- We introduced the following generalizations:
  - Inference over temporal logic formulas*
    - Computing P(ϕ_1 | ϕ_2; $\mathcal{M}$)
  - Learning over temporal logic formulas*
    - Computing argmax_θ P(ϕ ; $\mathcal{M}$)
  - Structure Learning over temporal logic formulas*
    - Computing $\text{argmax}_{G,\theta}$ P(ϕ ; $\mathcal{M}$)

*Formulas are in bounded LTL

PGMs for Stochastic Processes

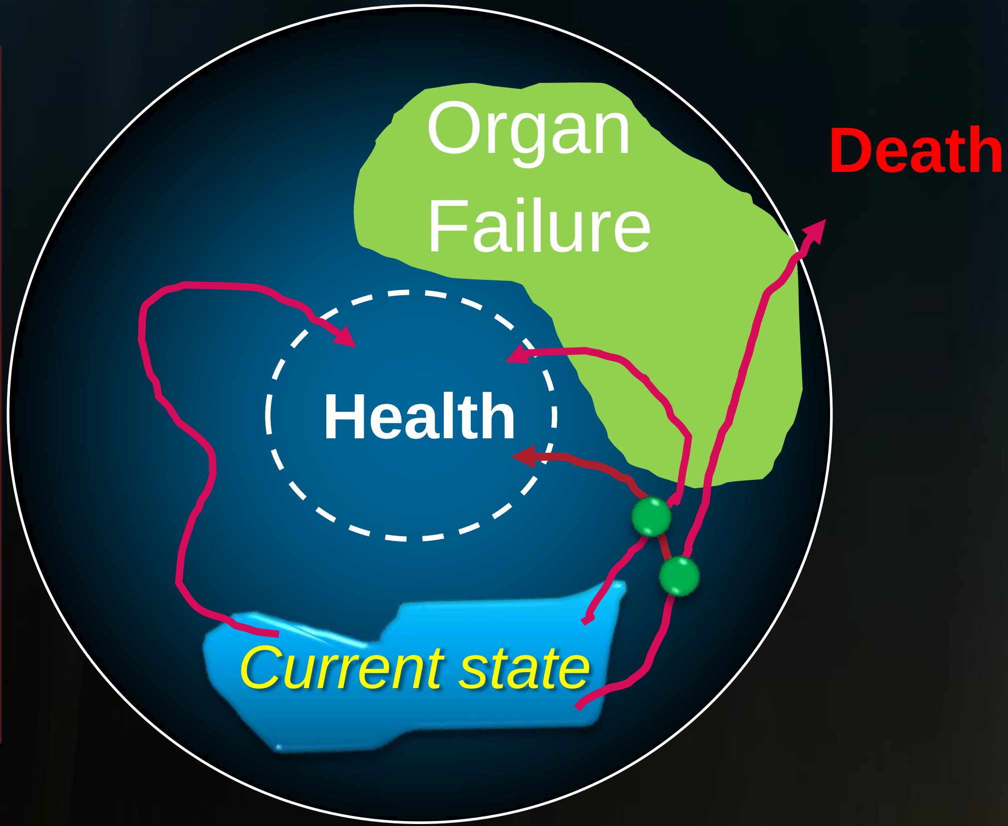
- We introduced the following generalizations:
  - Inference over temporal logic formulas*
    - Computing P(ϕ_1 | ϕ_2; $\mathcal{M}$)
  - Learning over temporal logic formulas*
    - Computing argmax$_\theta$ P(ϕ ; $\mathcal{M}$)
  - Structure Learning over temporal logic formulas*
    - Computing argmax$_{G,\theta}$ P(ϕ ; $\mathcal{M}$)

These generalizations provide a more expressive framework for using graphical models

In Context

- These generalizations are also relevant to our three tasks
  - Eg., $\phi := \neg$"Organ Failure" U^t "Health"

Primary Tasks:
(1) Determine where the patient is now (approximately)
(2) Characterize the patient's trajectory (approximately)
(3) Select interventions based on (1) and (2)



New Algorithms (1)

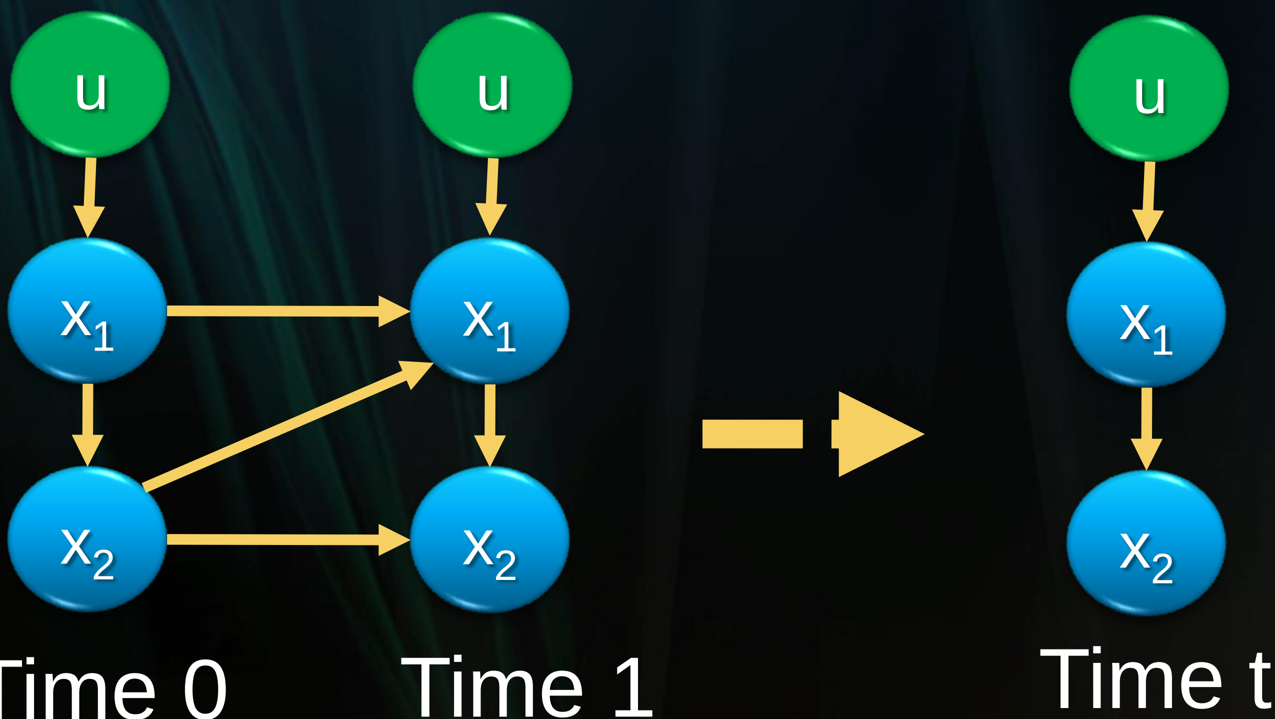
- Inference over temporal logic formulas [L09]

Dynamics	Distribution	Time	Inference
Linear	Gaussian	Discrete	Exact
Non-linear	Gaussian	Discrete	Accurate to 2^{nd} order
Non-linear	Non-Gaussian	Discrete	Approximate
Non-linear	Multinomial	Discrete	Exact

  - Also: new sampling algorithms for rare events

New Algorithms (2)

- Control policy synthesis and structure learning for synchronous or asynchronous Boolean networks [LJ08,LJ09]
  - Relies on symbolic model checking



New Algorithms (3)

- Learning over temporal logic formulas [L08]

Distribution	Time	Learning
Multinomial	Discrete	Approximate*
Continuous	Discrete	Approximate

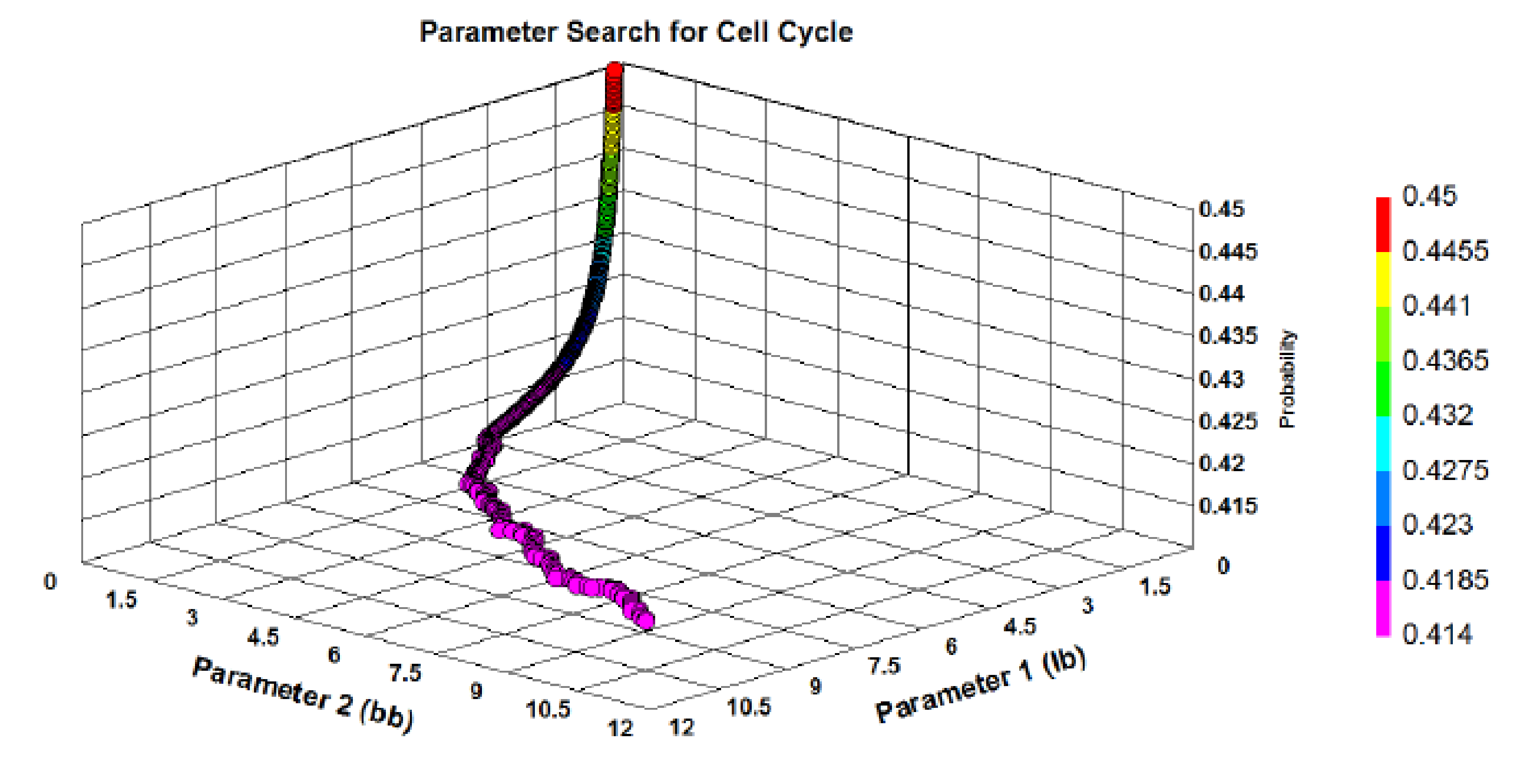
  - * Uses bit-vector decision procedures
- Parameter Synthesis [DCL09,JL10]

Distribution	Time	Learning
Continuous	Continuous	Approximate
Multinomial	Discrete	Approximate*

  - * Uses abstraction-refinement

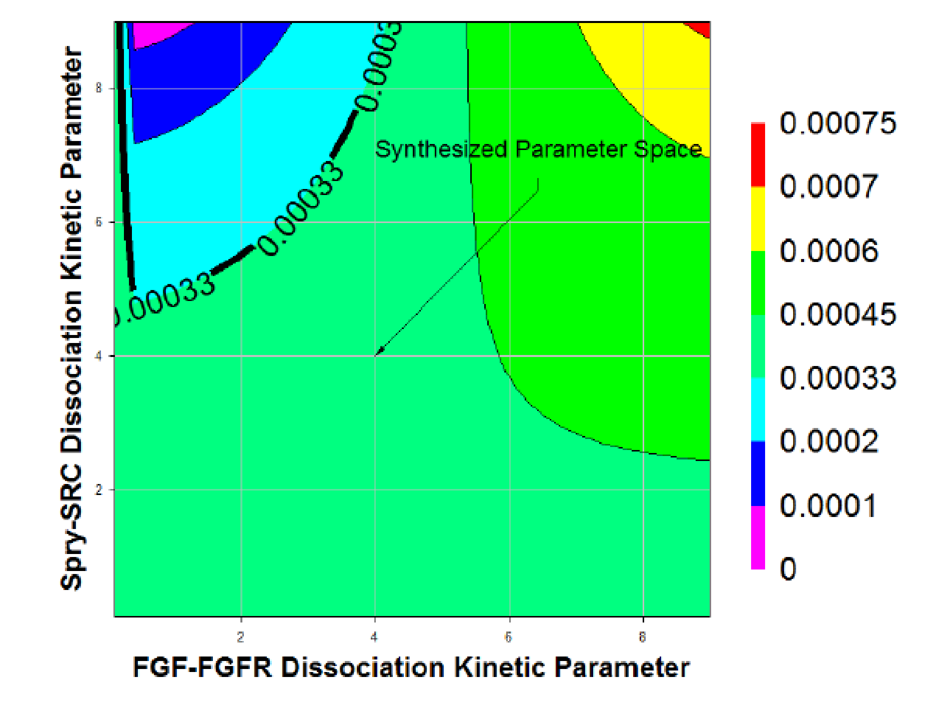
Example: Parameter Estimation

- Cell-cycle model
  - ϕ := F^t "Cyclin bound = 0"



Example: Synthesis

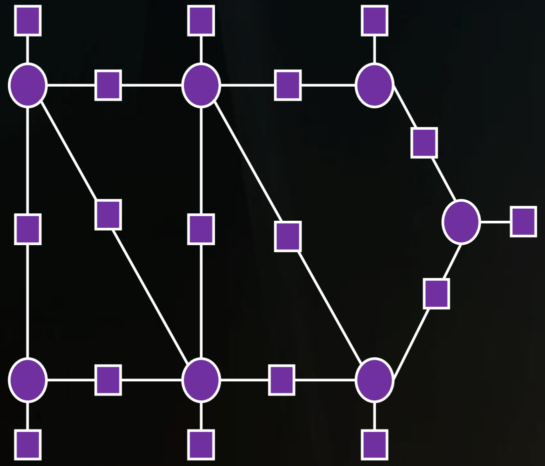
- Fibroblast growth receptor pathway model
  - $\phi := F^t (A>0 \& B = 0)$
  - 2D synthesis



Note: our method has been used to synthesize up to 6 parameters, simultaneously

Applications and Extensions

- Medical
  - Sepsis
  - Pancreatitis
  - Chronic Myeloid Leukemia
- Biological
  - Embryogenesis
  - Signaling Pathways
- Spatial Models
  - Markov Random Fields



Potential Collaborations

- Jim Faeder
  - Our algorithm for parameter synthesis for stochastic systems was developed with BioNetGen in mind
- Bud Mishra
  - Applications to GOALIE
- Atrial Fibrillation
  - Markov Random Fields
- Rance Cleaveland
  - Our rare event sampling procedure might be relevant to Reactis©