Structure or Noise?

NSF Workshop on
Uncertainty in Complex Interacting Systems

University of Southern California, Los Angeles

James P. Crutchfield
Complexity Sciences Center, Director
Physics Department
University of California, Davis
cse.ucdavis.edu/~chaos
13 April 2009

Joint work with

- Susanne Still:

  Information and Computer Sciences

  University of Hawaii at Manoa

- Chris Ellison

  Complexity Sciences Center

  Physics Department, UC Davis

Why We Must Model I

- Nature spontaneously organizes

Emergent structures

Why We Must Model 2

- Engineered systems spontaneously organize
  - Internet route flapping
  - Power-law Internet organization
  - Financial markets crash
  - Power grids fail spectacularly
  - Social pattern formation on the web
  - ...

And so ...

- Problem:

  Emergent structures not given directly by the
  system coordinates or
  governing equations of motion

- Consequence:

  Each system needs its own explanatory basis

Why we must Model 3

Fundamental: Mathematics of Intrinsic Randomness

- Nonlinear dynamical systems [Kolmogorov 1958]:

  Chaotic systems: Shannon entropy $h_\mu > 0$

- Kolmogorov-Chaitin complexity of Data [1963]:

Size of shortest Turing Machine Program to predict Data

- KC complexity = f(Shannon entropy) [Brudno 1978] :

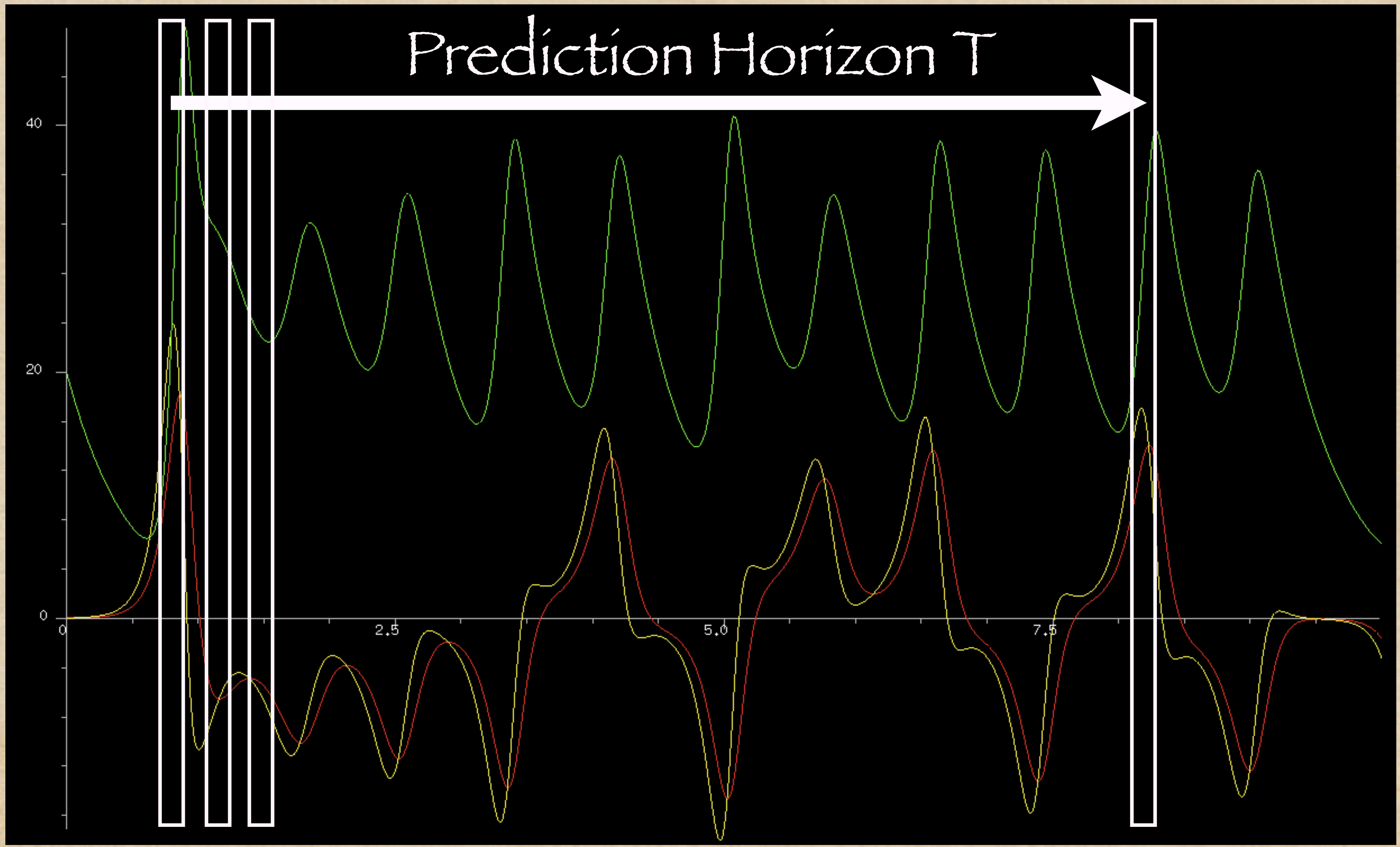
$$|\mathrm{Program}| \propto e^{h_\mu |\mathrm{Data}|}$$

Exponential Increase in Prediction Resources

$$\text{Accuracy} \propto e^{-T}$$

$$|\text{Measurements}| \propto e^{T}$$

$$|\text{Compute time}| \propto e^{T}$$



Consequence

- No short cuts!
  - No closed-form solutions
  - No computational speed-ups
  - Must compute full trajectory

- Right representation is critical for reducing the prediction error as far as possible (but no further!)
- Cannot estimate uncertainty unless you know the right model.

Fundamental in Complex Systems!

- Each nonlinear system requires its own representation
- Selecting balance between ascribing structure or noise to a measurement depends on representation

- Fundamental issue: Theory building
- Subsidiary issue:

  Statistical fluctuations due to finite data sample

  (This is NOT a talk about machine learning.)

Learning a Model

- Problem: Learn from Observations
  - The world behaves: $\overleftrightarrow{X} = \underset{\text{past}}{\overleftarrow{X}} \; \underset{\text{future}}{\overrightarrow{X}}$
    - Given $\Pr(\overleftarrow{X}, \overrightarrow{X})$, agent learns model:

      States $\mathcal{R}$ and Dynamic

- Pattern Discovery:

  Learn the world's hidden states $\Pr(\mathcal{R} | \overleftarrow{X})$

JPC & K. Young, Inferring Statistical Complexity, Physical Review Letters 63 (1989) 105-108.
C. R. Shalizi & JPC, Journal Statistical Physics 104 (2001) 817-879.

Learning a Model ...

- Causal shielding:

$$\Pr(\overleftarrow{X}\overrightarrow{X}) = \Pr(\overleftarrow{X}\,|\mathcal{R})\Pr(\overrightarrow{X}\,|\mathcal{R})$$

- Dynamics of learning:

  Search in the space of models: $\mathcal{R} \in \mathcal{M}$

Learning a Model ...

(Susanne Still, Chris Ellison, & JPC)

- Causal shielding objective function

$$\min_{\Pr(\mathcal{R}|\overleftarrow{X})} \left(I[\overleftarrow{X};\mathcal{R}] + \beta I[\overleftarrow{X};\overrightarrow{X}\,|\mathcal{R}] \right)$$

Model: Map from histories to states	Info states contain about histories	Reduce info history has about future

- Structure/Noise Design Control: $\beta \sim 1/T$

arxiv.org: 0708.0654 [physics.gen-ph] & 0708.1580[cs.IT]

Learning a Model ...

- Optimal states $\Pr(\mathcal{R}|\overleftarrow{X})$ are Gibbs states:

$$\Pr_{\text{opt}}(\mathcal{R}|\overleftarrow{X}) = \frac{\Pr(\mathcal{R})}{Z(\overleftarrow{X},\beta)} e^{-\beta E(\mathcal{R},\overleftarrow{X})}$$

where

$$E(\mathcal{R},\overleftarrow{X}) = \mathcal{D}\left(\Pr(\overrightarrow{X} \mid \overleftarrow{X}) || \Pr(\overrightarrow{X} | \mathcal{R})\right)$$

$$\Pr(\overrightarrow{X} | \mathcal{R}) = \frac{1}{\Pr(\mathcal{R})} \sum_{\overleftarrow{X}} \Pr(\overrightarrow{X} \mid \overleftarrow{X}) \Pr(\mathcal{R}|\overleftarrow{X}) \Pr(\overleftarrow{X})$$

$$\Pr(\mathcal{R}) = \sum_{\overleftarrow{X}} \Pr(\mathcal{R}|\overleftarrow{X}) \Pr(\overleftarrow{X})$$

Learning a Model ...

- Solve these equations self-consistently

  (Analytical in special cases; numerical generally)

- Parametrized family of models:

  $$R_\beta \colon \Pr(\mathcal{R} | \overleftarrow{X})$$

- Structure or Noise?

  β trades-off model size against prediction error

  Structure versus Noise

What Do Solutions Mean?

Causal Models

- Causal architecture given by ϵ-Machine M:
  - Optimal predictor:

$$h_\mu(M) \leq h_\mu(\mathcal{R})$$

  - Minimal size (within optimal predictors $\widehat{\mathcal{R}}$):

$$C_\mu(M) \leq C_\mu(\widehat{\mathcal{R}})$$

  - Unique (within min, opt predictors)

JPC & K. Young, Inferring Statistical Complexity, Physical Review Letters 63 (1989) 105-108.
C. R. Shalizi & JPC, Journal Statistical Physics 104 (2001) 817-879.

Learning a Model ...

- Theorem: Low-temperature limit

$$\beta \rightarrow \infty$$

Recover ϵ-Machine:

$$R_\beta \rightarrow M$$

- Conclusion:

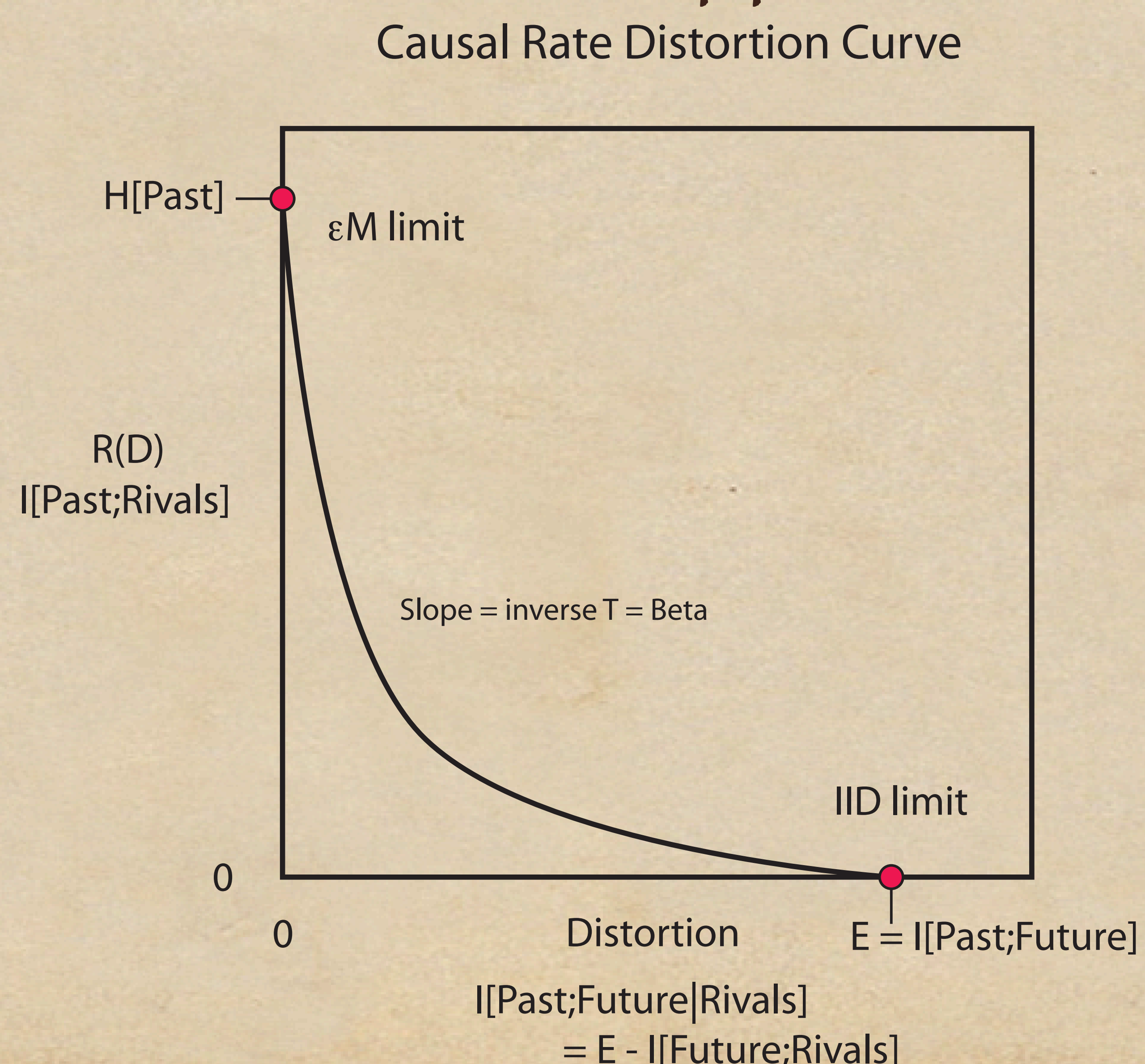
At given prediction error

R_β is best causal approximate.

Learning a Model ...

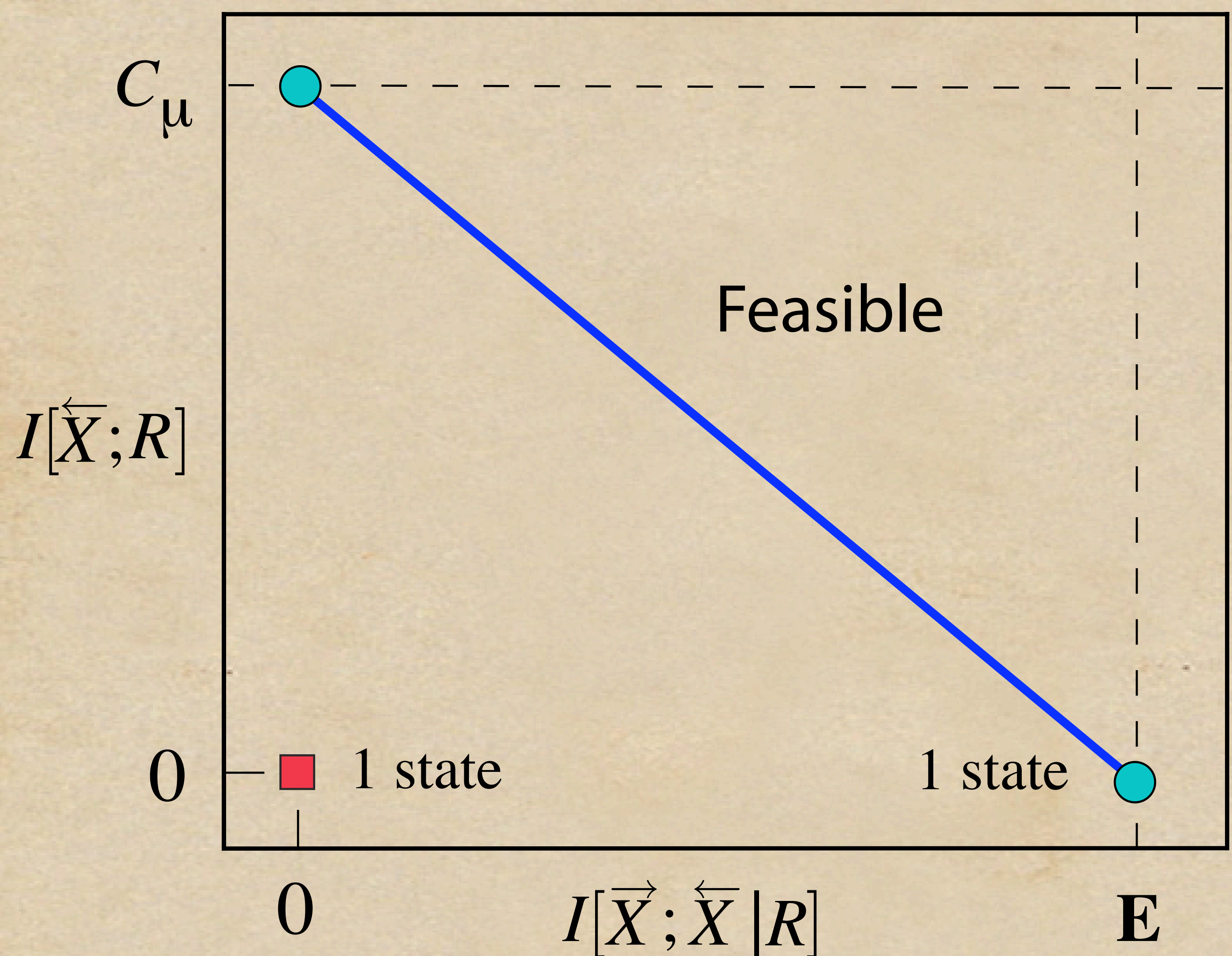
Optimally balance structure & error
At each level β of approximation

In theory



Learning a Model ...

Analytical cases



Predictively Reversible:

$$P(\overrightarrow{x} \mid \overleftarrow{x}) = \delta_{\overrightarrow{x}, f(\overleftarrow{x})}$$

(e.g., periodic)

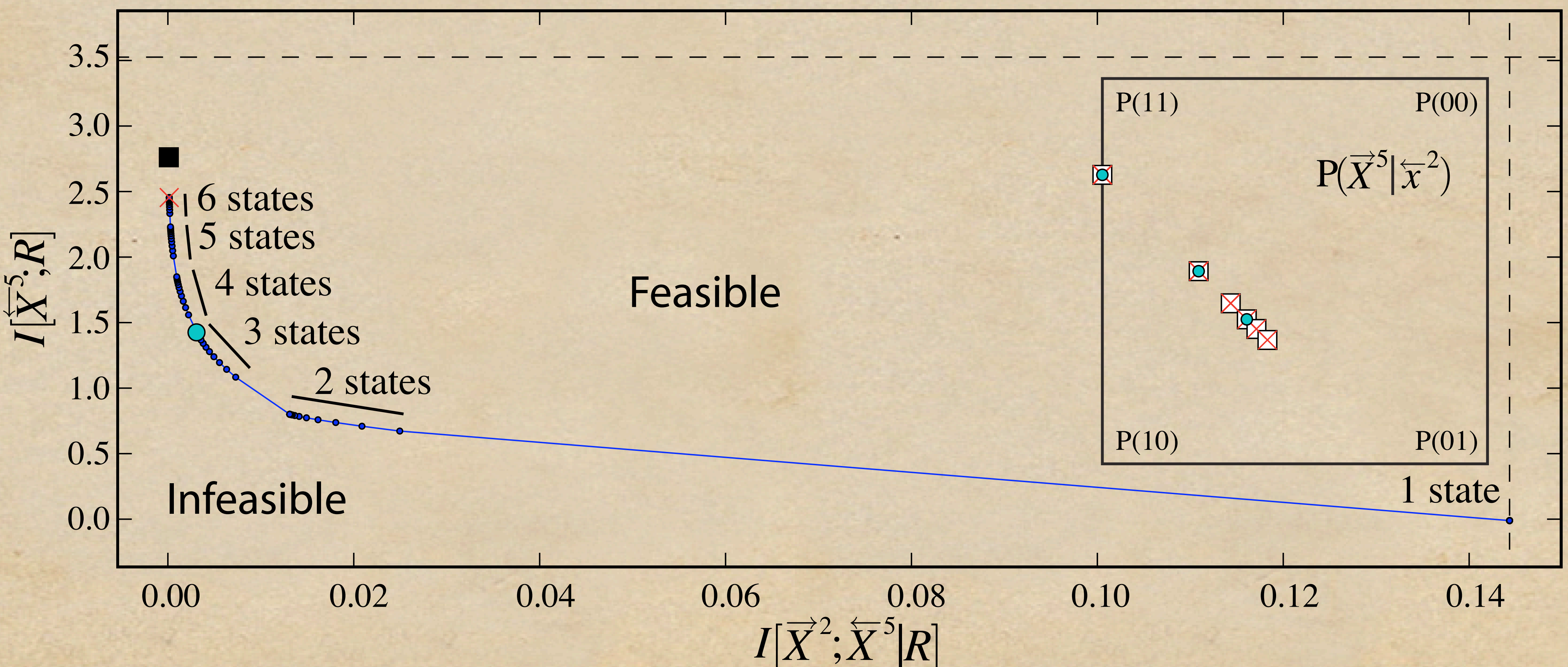
All IID processes:

$$P(\overrightarrow{x} \mid \overleftarrow{x}) = P(\overrightarrow{x})$$

Learning a Model ...

Optimally balance structure & error
At each level β of approximation

In practice: Learn an oo-state world (SNS: "simple nondeterminstic source")



Learning a Model ...

- Causal compressibility: Shape of RD curve
  - Benefit of choosing smaller model for loss in predictability
  - Deviation from straight-line RD curve
- IID: No
- Predictively reversible: No
- SNS: Yes

Conclusions

- Causal shielding principle leads to
  - Process's organization: ϵ-Machine
  - Family of best approximations to ϵ-M
- Structure or Noise?
  - Cannot estimate uncertainty until you know structure.
  - First-principle distinction
  - Principled trade-off

Thanks!

All papers online at the
Computational Mechanics Archive
http://cse.ucdavis.edu/~cmg/