The Macroscope:

An Engine for Pattern Discovery

AstroInformatics 2012
Microsoft Research, Inc.
Redmond, Washington
10-13 September 2012

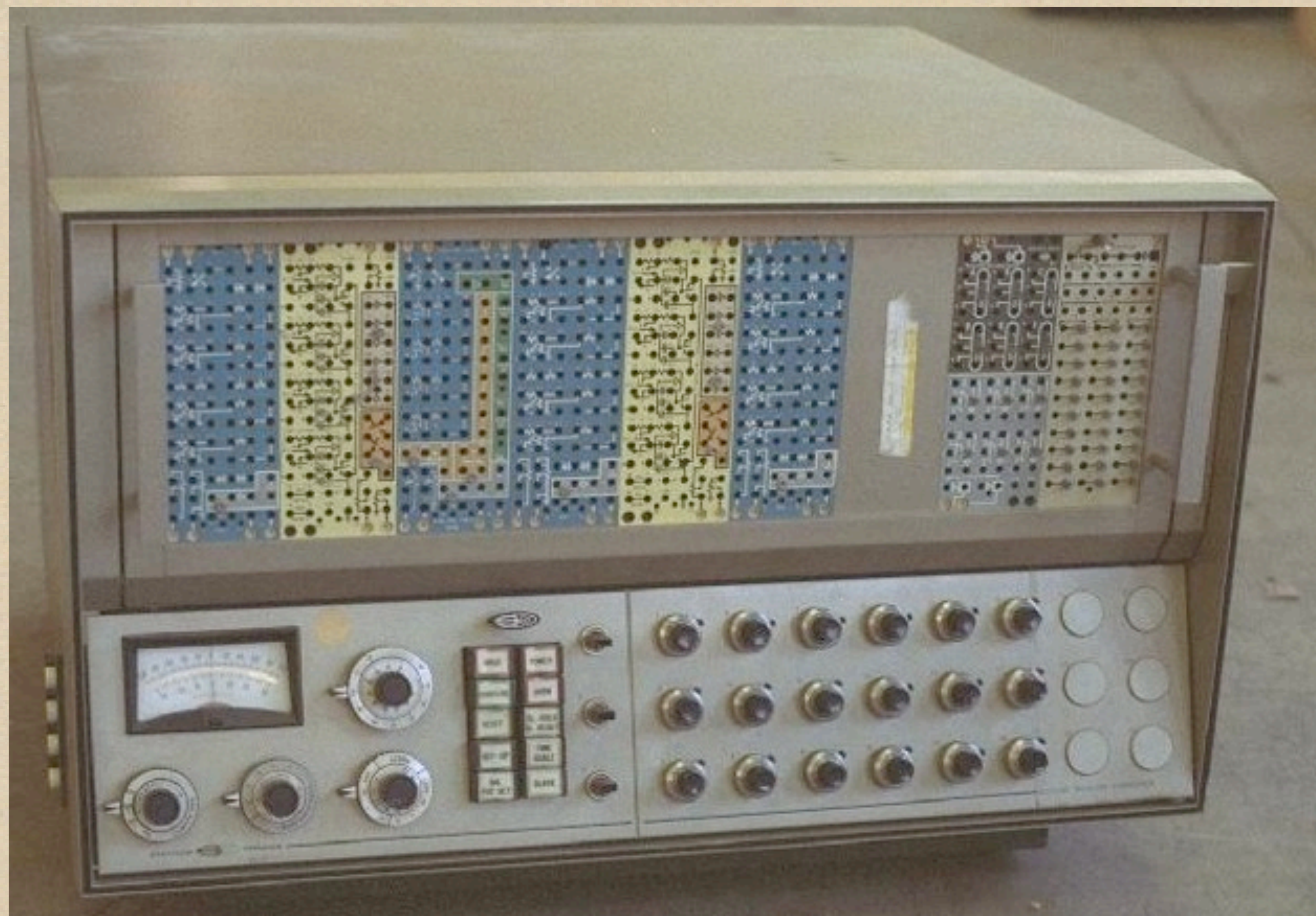
James P. Crutchfield
Complexity Sciences Center
Physics Department
University of California, Davis
csc.ucdavis.edu/~chaos

Big Theory

- Big Data is Not Enough
- Big Iron is Not Enough
- Theory's Role
- Automating Science?
- Pattern Discovery
- Sensory Immersive Mathematics

J. P. Crutchfield, Between Order and Chaos, Nature Physics 8 (January 2012) 17-24.

A beginning



A beginning

Systron Donner
SD 10/20

Why We Must Compute I

- Fundamental science:
  - Emergent structures
    - Nature spontaneously organizes

Emergent structures

Emergent structures

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

Consequence

- Each system needs its own function basis
- Emergent structures not given directly by the governing equations of motion

- We must compute to explore the possible

Why we must compute II

Pierre Simon de Laplace, Calculus of Probabilities (1776).

- Determinism: "... if we conceive of an intelligence which at a given instant comprehends all the relations of the entities of this universe, it could state the respective positions, motions, and general affects of all these entities at any time in the past or future."
- A paradigm: "Physical astronomy, the branch of knowledge which does the greatest honor to the human mind, gives us an idea, albeit imperfect, of what such an intelligence would be."
- Ignorance: "So it is that we owe to the weakness of the human mind one of the most delicate and ingenious of mathematical theories, the science of chance or probability."

Pierre Simon de Laplace, Calculus of Probabilities (1776).

- Determinism: "... if we conceive of an intelligence which at a given instant comprehends all the relations of the entities of this universe, it could state the respective positions, motions, and general affects of all these entities at any time in the past or future."
- A paradigm: "Physical astronomy, the branch of knowledge which does the greatest honor to the human mind, gives us an idea, albeit imperfect, of what such an intelligence would be."
- Ignorance: "So it is that we owe to the weakness of the human mind one of the most delicate and ingenious of mathematical theories, the science of chance or probability."

Deterministic chaos: "... it may happen that small differences in the initial conditions produce very great ones in the final phenomena. A small error in the former will produce an enormous error in the latter. Prediction becomes impossible ...

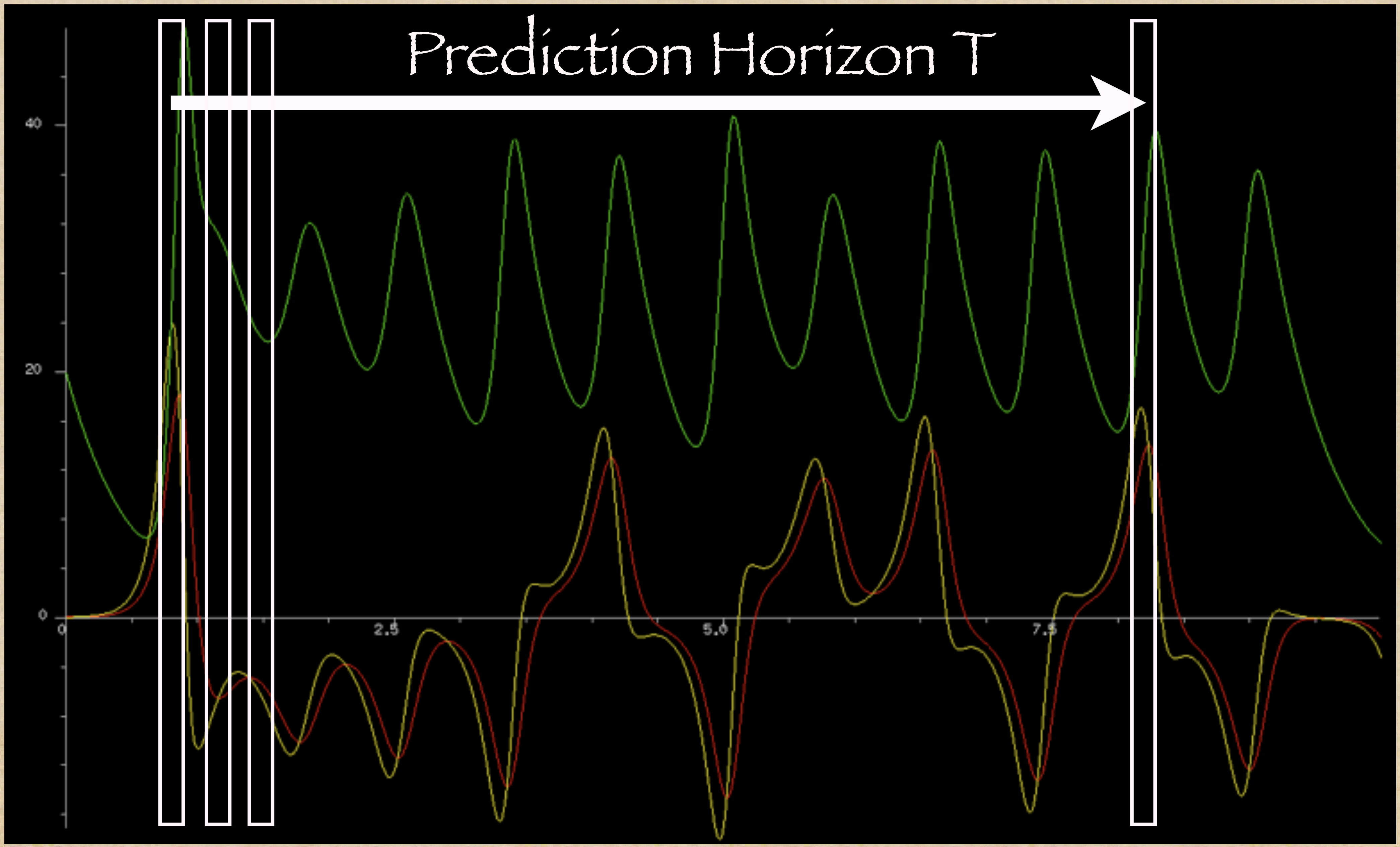
Henri Poincaré, Les Methodes Nouvelles de la Mecanique Celeste (1892).

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

- We must compute full trajectory

Why We Must Compute

- Computing is a response:
  - Emergent organization
  - Unpredictability
  - Limited epistemology

Consequence

- Computationalists will be employed

(We just proved guaranteed employment.)

The Big Iron Fallacy

- Build it and they will make progress
  - Wrong!
- Machines as powerful as many natural systems
  - Thus, using computing to "understand" natural processes begs the scientific question.
  - The simulations are as complex as reality.
- Need: Tools for understanding.

The Big Iron Fallacy

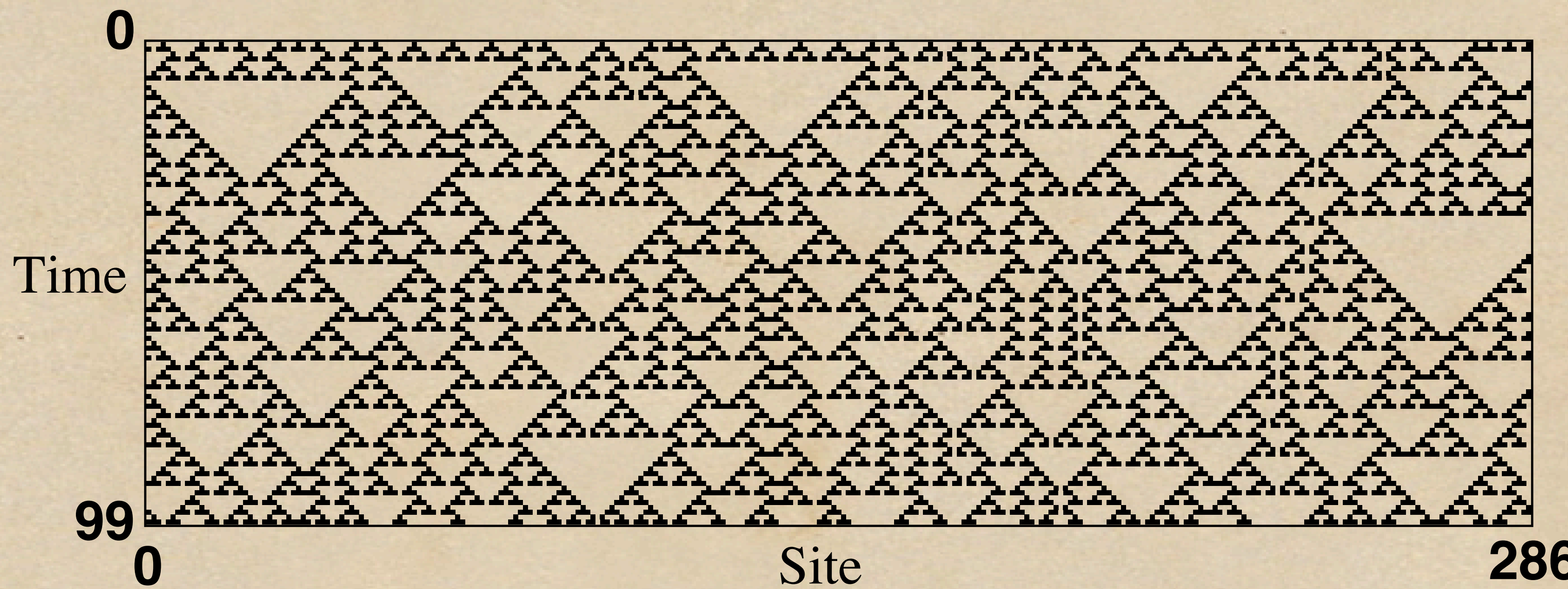
- Build it and they will make progress
  - Wrong!
- Machines as powerful as many natural systems
  - Thus, using computing to "understand" natural processes begs the scientific question.
  - The simulations are as complex as reality.
- Need: Tools for understanding.

Ditto Big Data

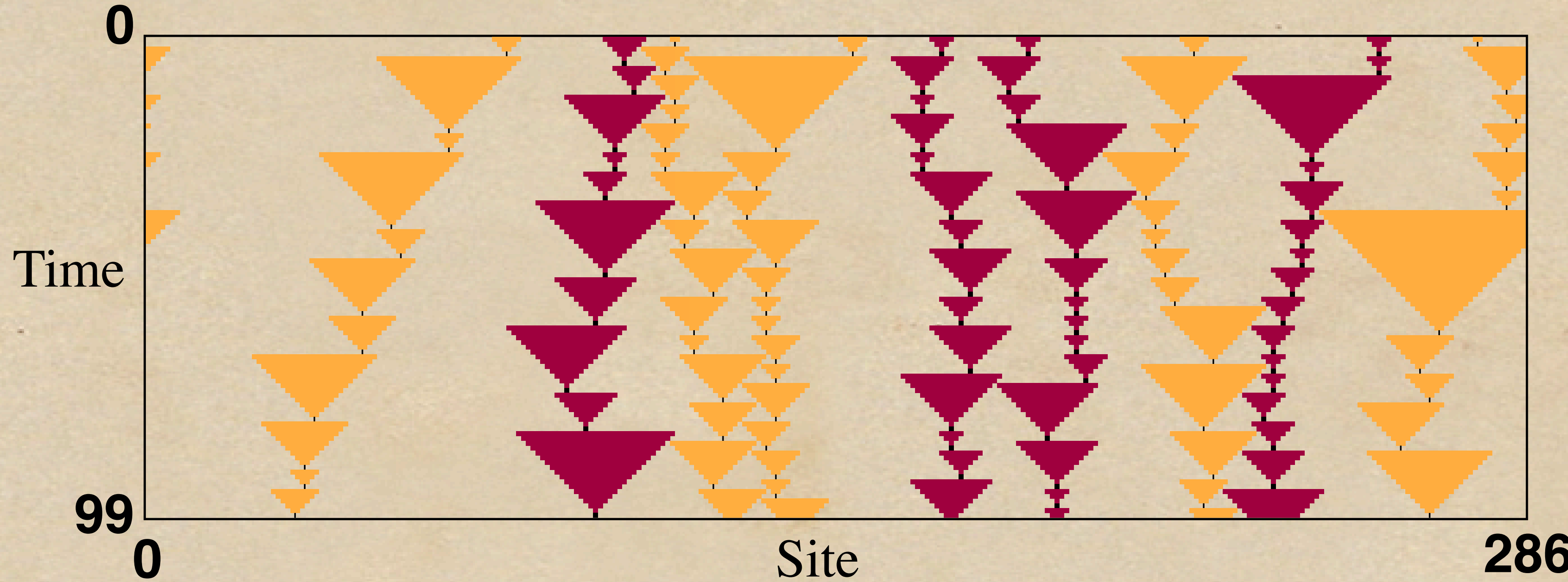
Frontiers I: Artificial Science

- A theory of theory building:
  - Process of modeling the order that lies between Laplacian determinism and statistics
- Direct computational support of theory building
- No alternative: Too nonlinear, too complex, too much data, too many probes

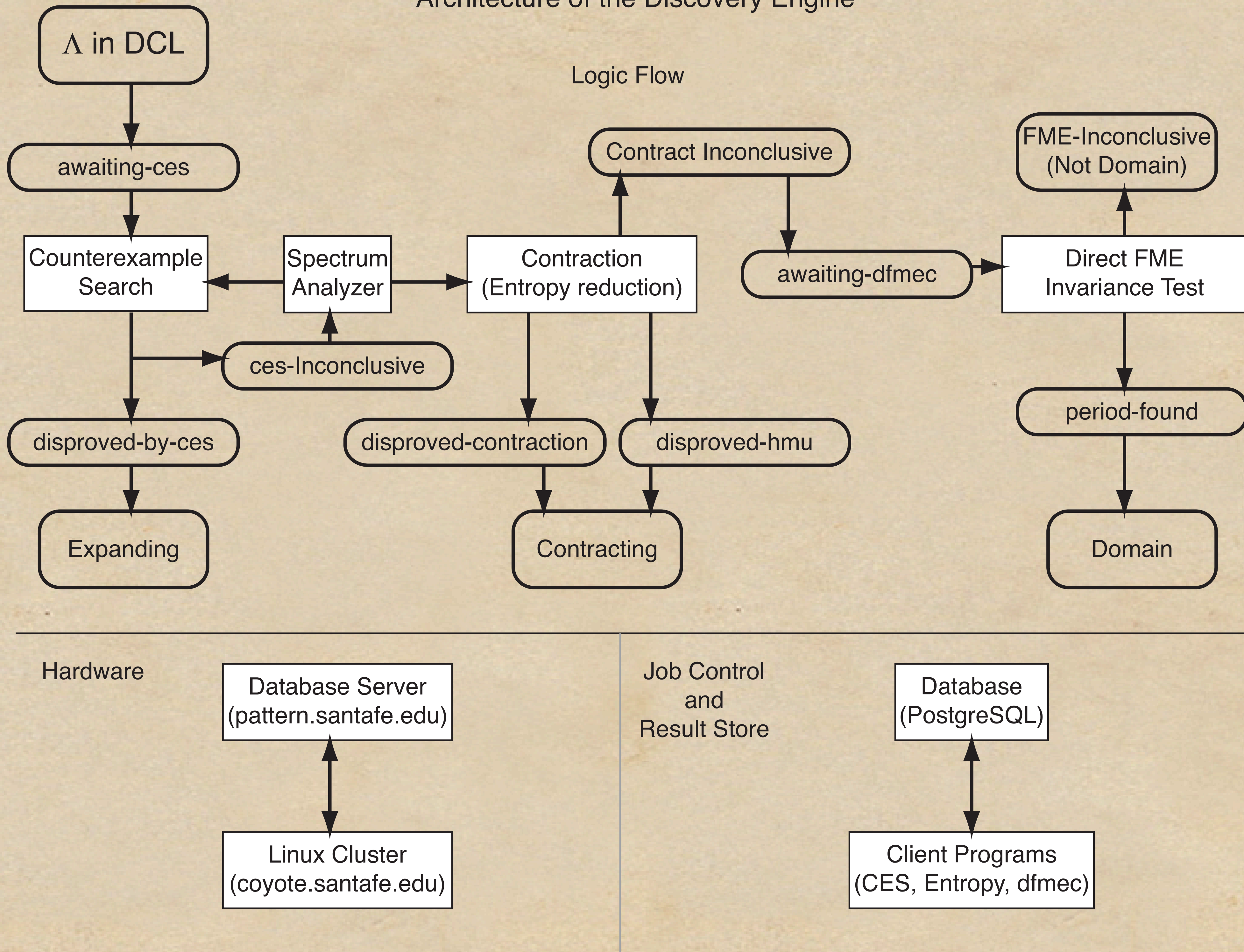
Artificial Particle Physics



Automated Discovery of Hidden Structure



Architecture of the Discovery Engine



Automated Discovery of Hidden Structure

ECA 22: Table of a Million Theorems

States	n-DCL	ECA 22 Nonexpanding n-DCL		
n	Size	Number	Contracting	Domains
1	3	2	Σ^*	$\Lambda^1 = 0^*$
2	7	2	$(0\Sigma)^*$	$\Lambda^2 = (01)^*$
3	78	0		
4	1,388	2		Λ^0_0, $\Lambda^3 = (0011)^*$
5	35,186	5	$((11)(11+01)^*(10)+00)^*$ $((101)(00+1)+(1(00+1)+0))^*$ $([(101)^+(1(00+1)+00)]+[1(00+1)+0])^*$ $([1^+(01)(00+1)]+[1^+(00)+0])^*$ $([(1^+01)^+(1^+00+00)]+[1^+00+0])^*$	
6	1,132,613	268	267	Λ^0_1

QED

Two years ago: Estimated compute time $\sim$ 5 Beowulf years!

One year ago, (first) proof took $\sim$ 1 week: speed $\sim$ 7000 tph.

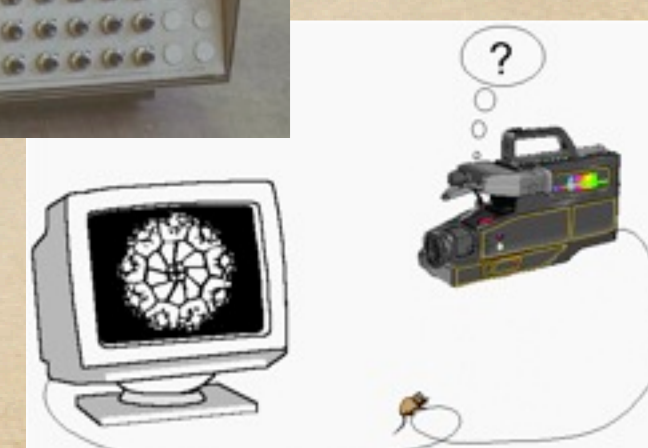
Proof now takes $\sim$ several hours: speed $\sim 10^6$ tph!

Automated Discovery of Hidden Structure

ECA 22's Particle Catalog	
Domains	Process Language
Λ^0	$\Lambda^0_0 = (000\Sigma)^*$ $\Lambda^0_1 = (1110 + 0000)^*$
Λ^1	0^*
Λ^2	$(01)^*$
Λ^3	$(0011)^*$
Particles	Wedges
α	$\Lambda^0_{0A} 001_A \Lambda^0_0$ $\Lambda^0_{1A'} 1111_{B'} \Lambda^0_1$
β	$\Lambda^0_{0A} 01_A \Lambda^0_0$ $\Lambda^0_{1A'} 110_{B'} \Lambda^0_1$ $\Lambda^0_{1A'} 001_{B'} \Lambda^0_1$
Interactions	
$\alpha + \alpha \rightarrow \beta$ $\beta + \beta \rightarrow \emptyset$ $\alpha + \beta \rightarrow \Lambda_{\alpha-\beta\text{gas}}$ $\beta + \alpha \rightarrow \Lambda_{\alpha-\beta\text{gas}}$	

Frontiers II:
Tools of Discovery

- Nature of complex systems & human perception ...
- How to design tools that aid pattern discovery?
- Towards a Cognitive Ergonomics



The Holodeck is Here!

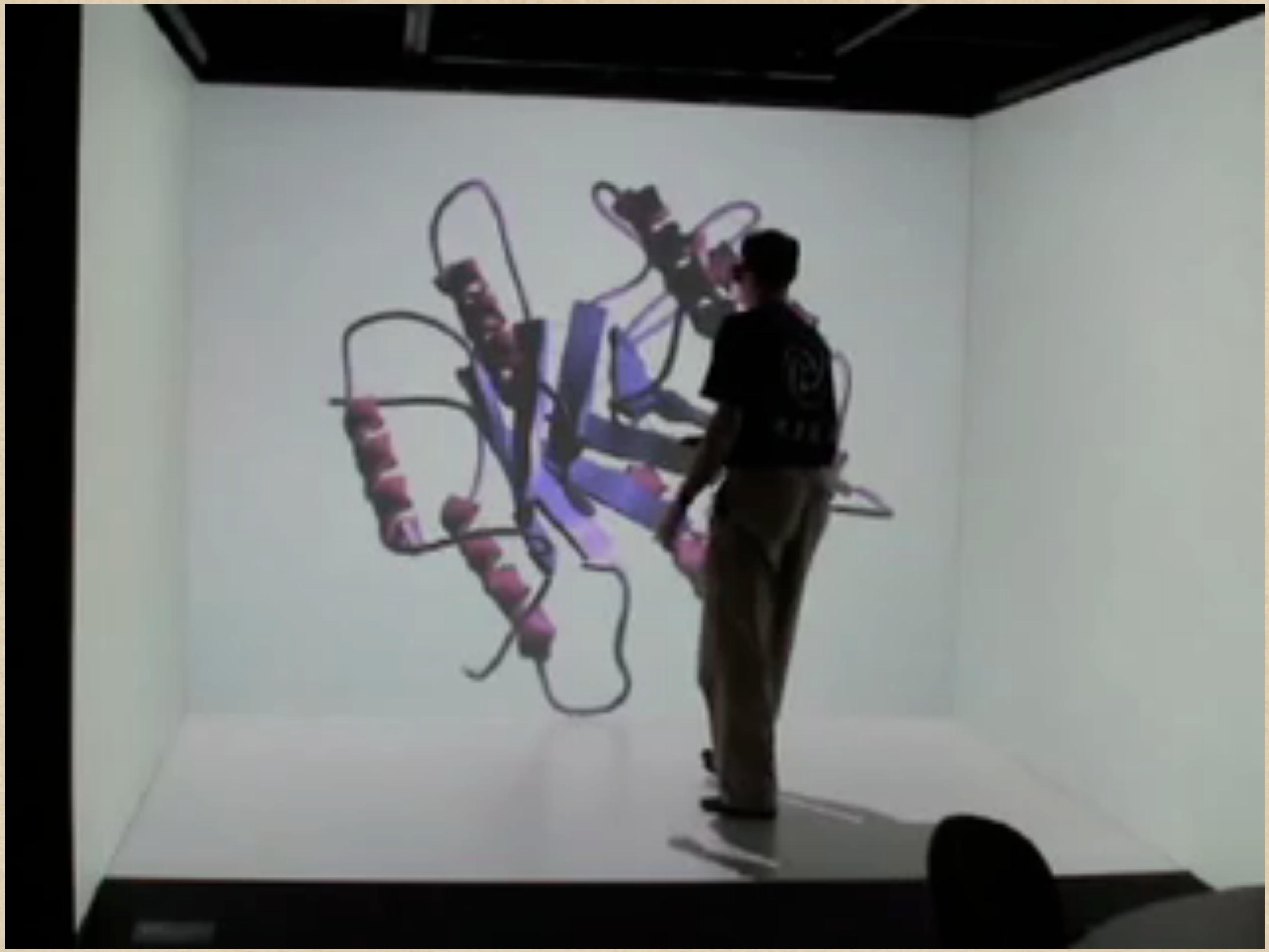
KeckCAVES

- Sensory Immersive Environment
- 10' x 10' x 8' Room
- Three Walls + Floor:
  Each a stereoscopic
  projection screen

500K$

http://keckcaves.org/

Interactive Protein Design

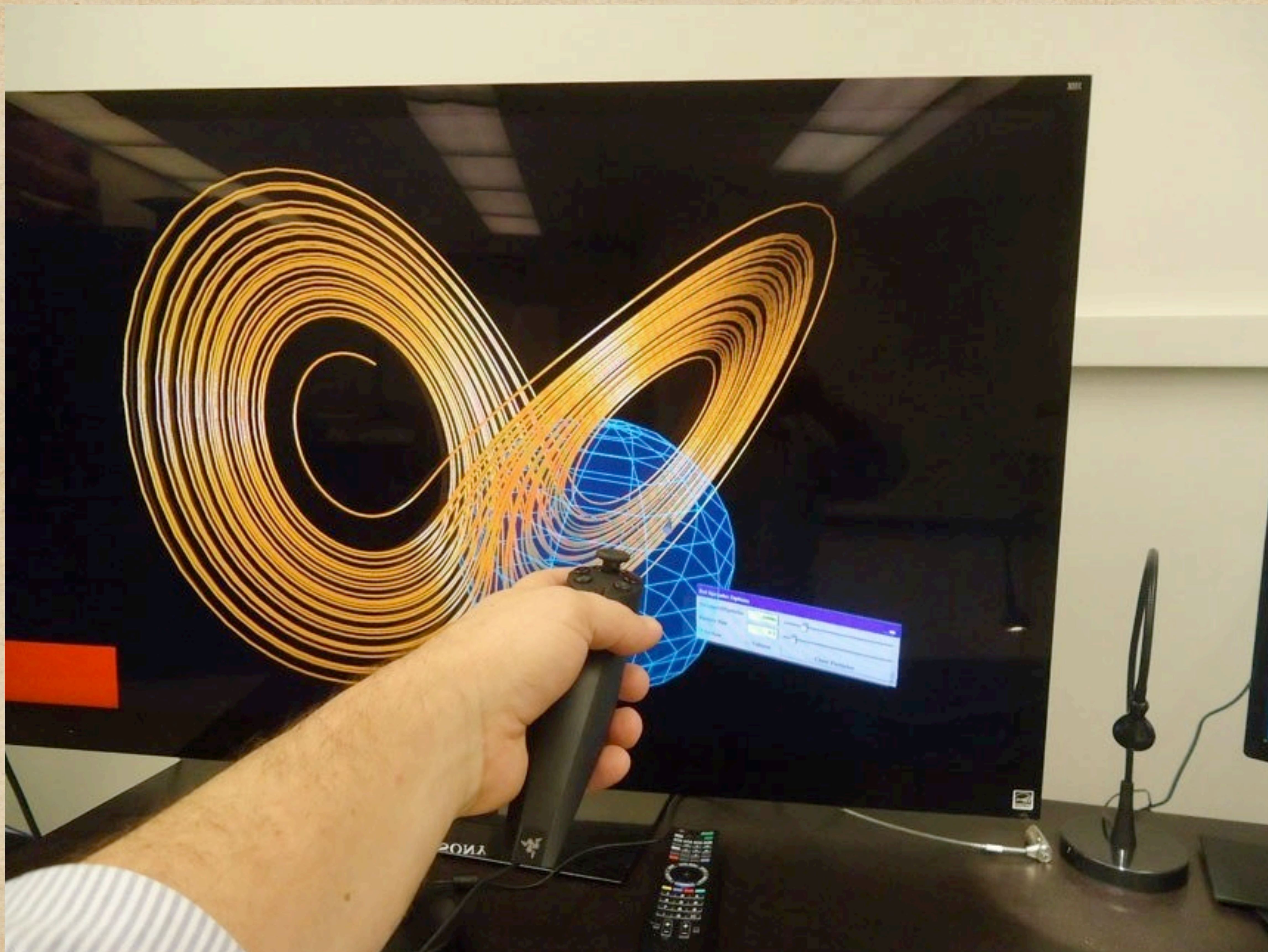


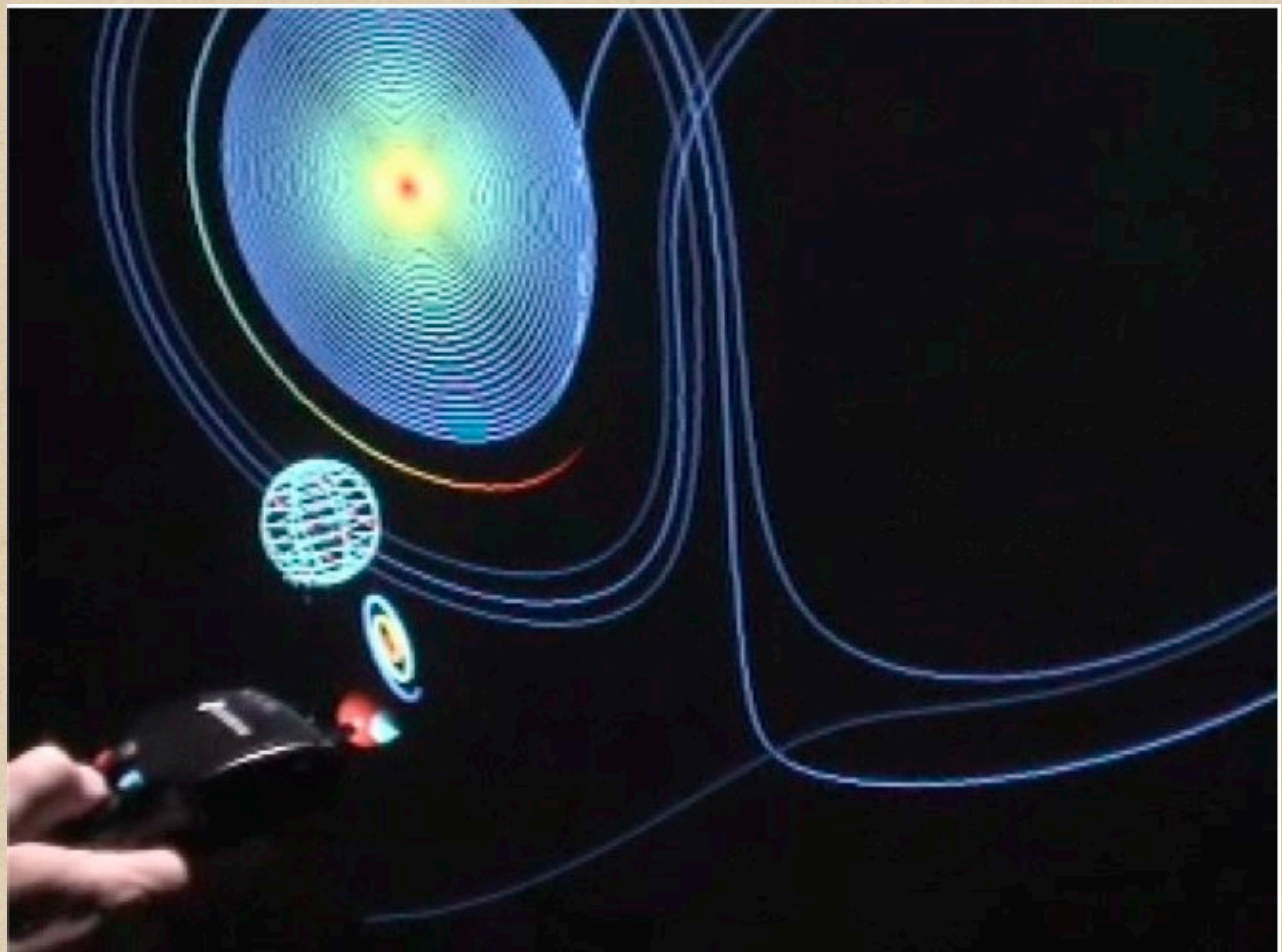
Deterministic Chaos

The Macroscope

- Open Source/DIY
- COTS
- 5K$


SONY

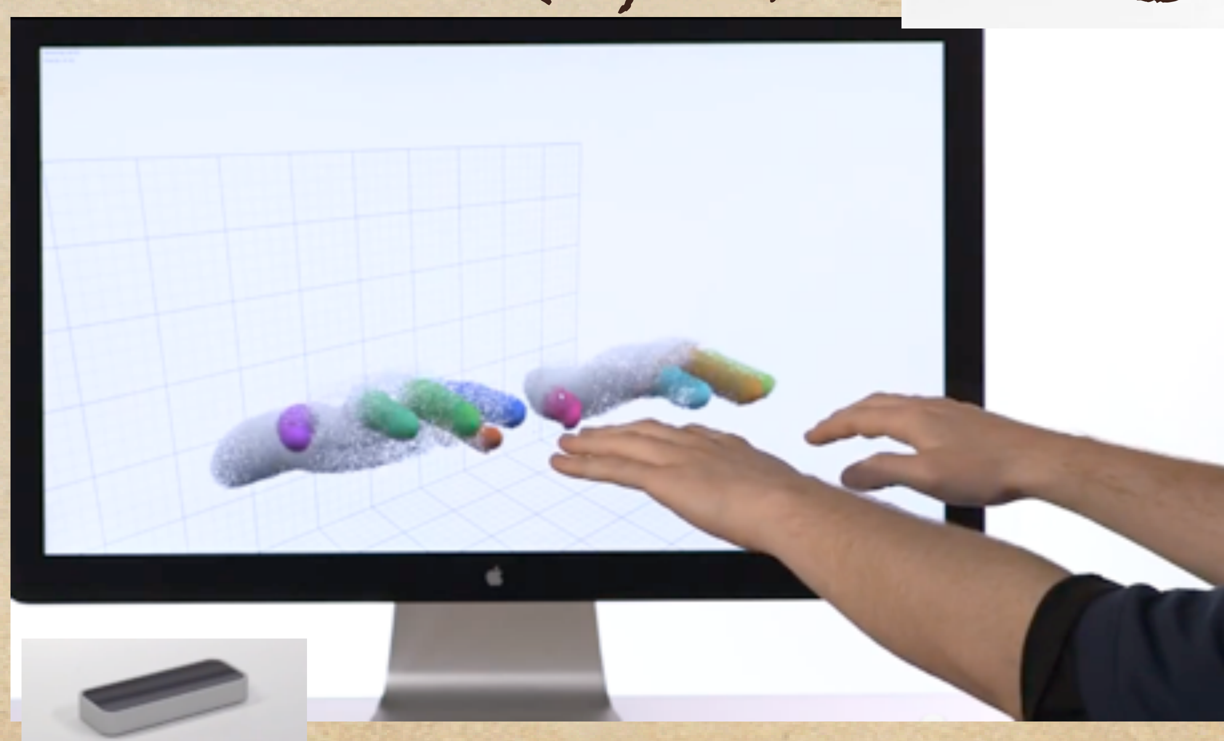




Asymptopia

Leap Motion ($70)

Oculus Rift ($500)



Closing Remarks

- Nature and our designed world are complex
- Computing has been and is a consequence
- Our (collective) view of modern scientific practice is out of date: Positively 19th century

- Response: Pattern Discovery

Thanks!

Why we must compute II

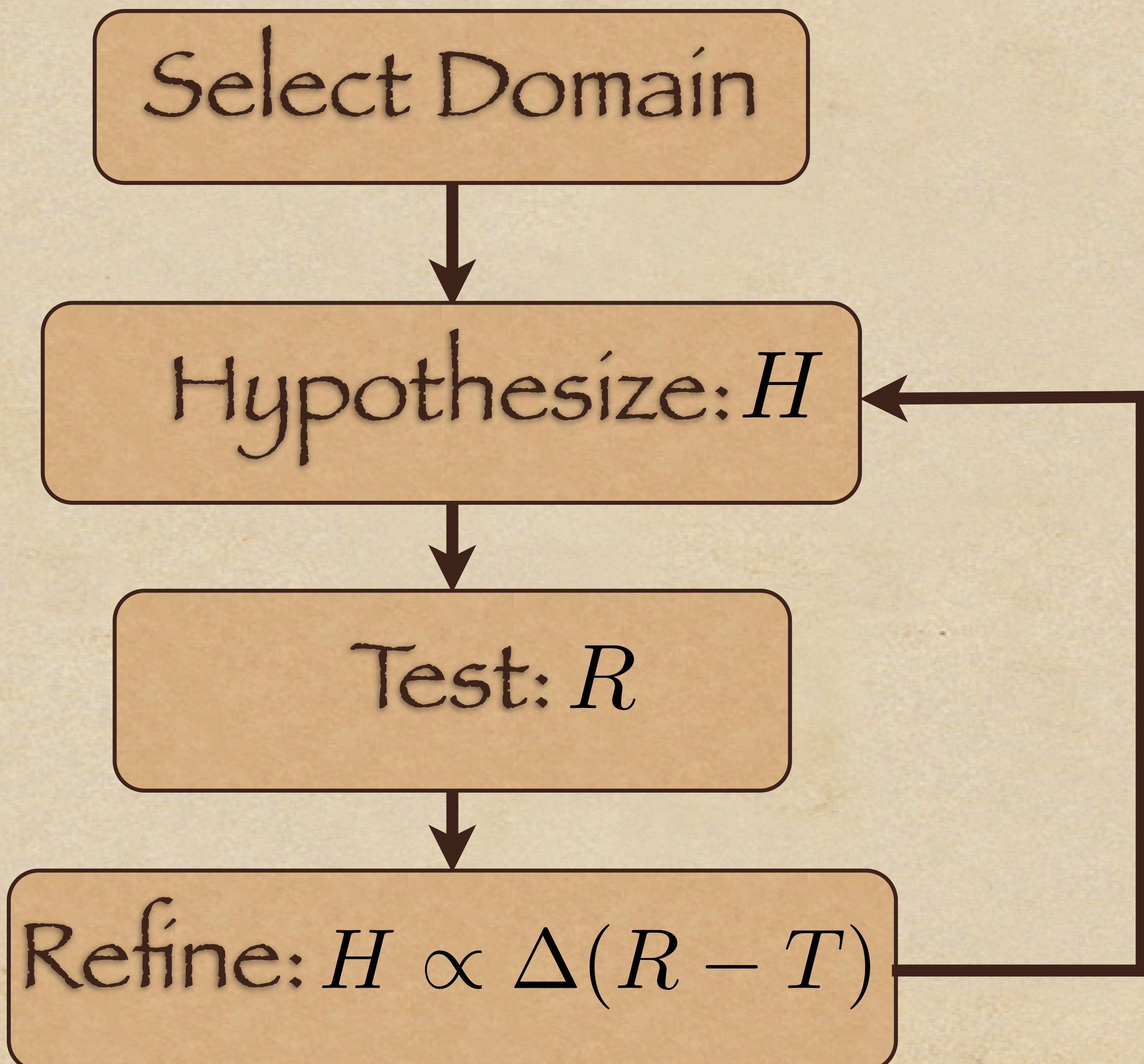
- Fundamental Mathematics: Intrinsic Randomness
  - Nonlinear dynamical systems [Kolmogorov 1958]:
    - Chaotic systems: Shannon entropy $h_\mu > 0$
  - Kolmogorov-Chaitin [1963] complexity of Data:
    - Size of shortest Turing Machine Program to predict Data
  - KC complexity = Shannon entropy [Brudno 1978] :

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

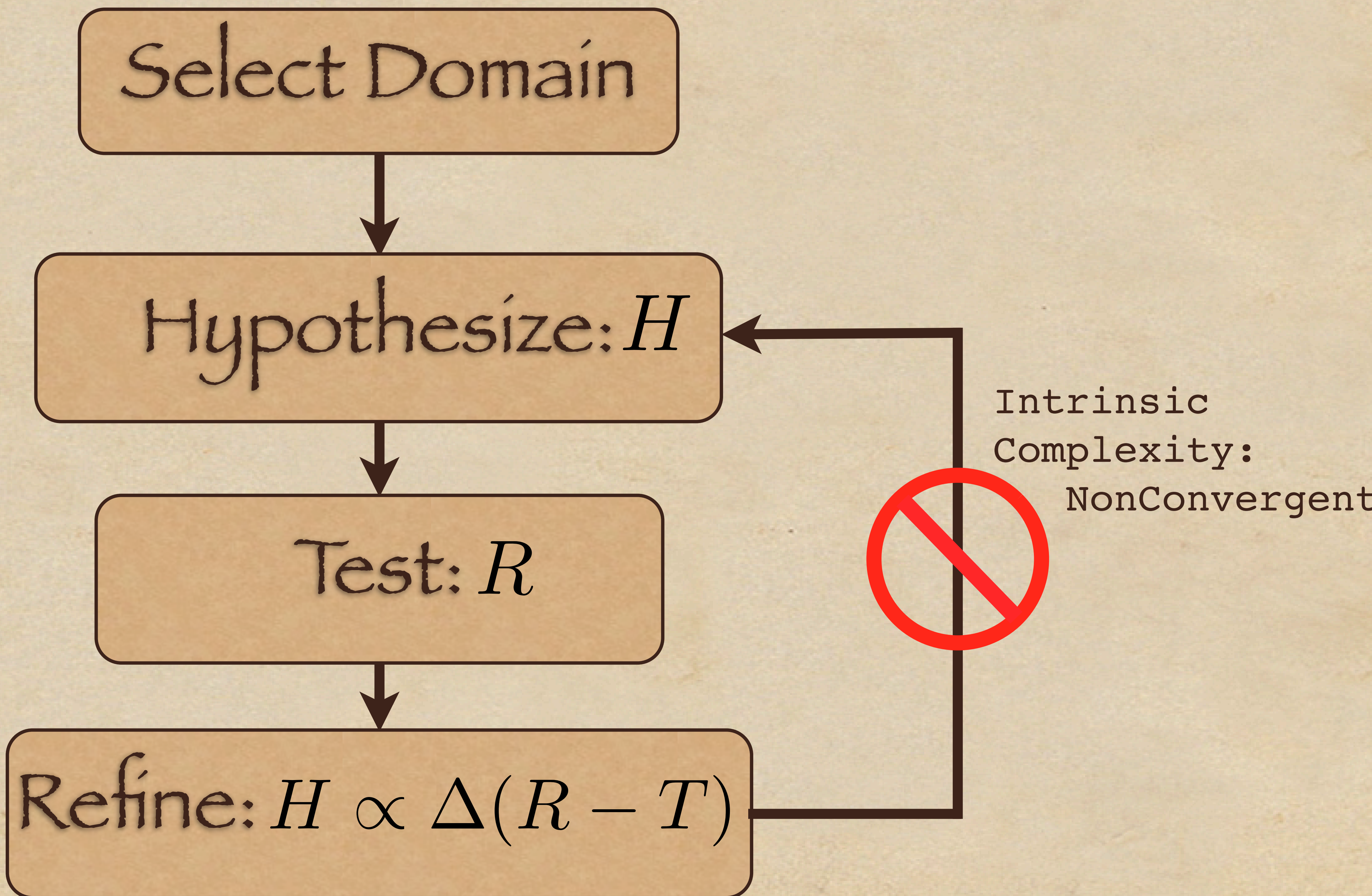
Frontiers II: Pattern Discovery

- Human perception as a communication channel
- Match vast output to human perceptual bandwidth
- Study human (multimodal) perception and conceptualization to match bandwidth
- Computing psychophysics: E.g., browser brought constellation of existing capabilities within human attention span.
- Beyond presentation to interaction: Feedback channel

Baconian Science Algorithm



Baconian Science Algorithm



Consequence

- Symbolic mathematics is fundamentally limited [Chaitin]

- We must compute to know

Complex Systems **1** (1987) 417–452

Equations of Motion from a Data Series*

James P. Crutchfield
Bruce S. McNamara†
Physics Department, University of California,
Berkeley, CA 94720, USA

Abstract. Temporal pattern learning, control and prediction, and chaotic data analysis share a common problem: deducing optimal equations of motion from observations of time-dependent behavior. Each desires to obtain models of the physical world from limited information. We describe a method to reconstruct the deterministic portion of the equations of motion directly from a data series. These equations of motion represent a vast reduction of a chaotic data set's observed complexity to a compact, algorithmic specification. This approach employs an informational measure of model optimality to