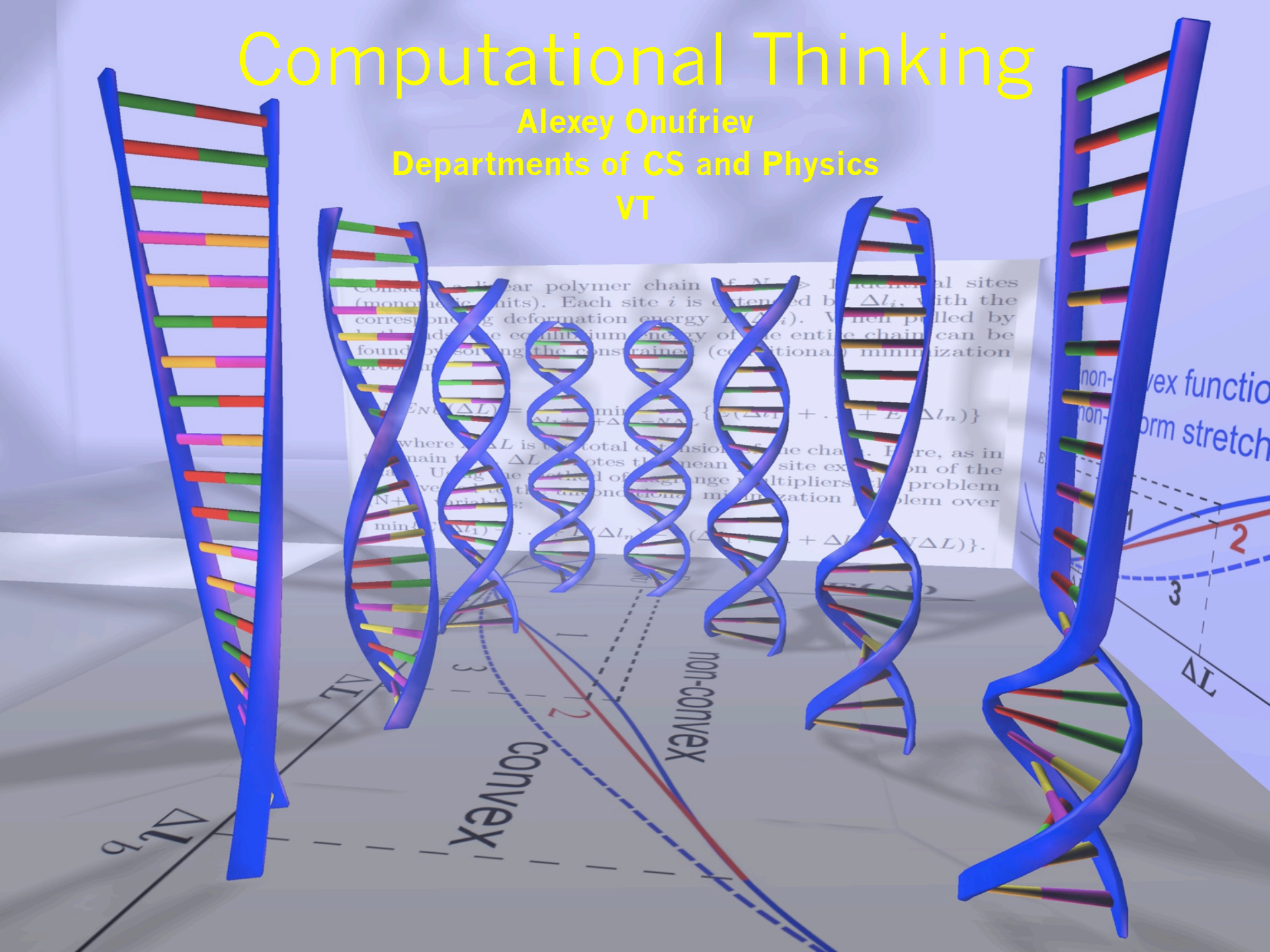
Computational Thinking

Alexey Onufriev
Departments of CS and Physics
VT



Outline

- The place of computational thinking. A question.
- The pervasive influence of computation on modern natural sciences.
- Example: the “grand challenge of computational science”: the protein folding problem. The Levinthal’s paradox
- Is the universe computable? A question with important consequences. The Laplace Demon.
- Ok, so where do we place KT?

The place of computational thinking in the context of human creativity. Where it stands?

Tools (e.g. paper, screwdriver, calculator)

LANGUAGE

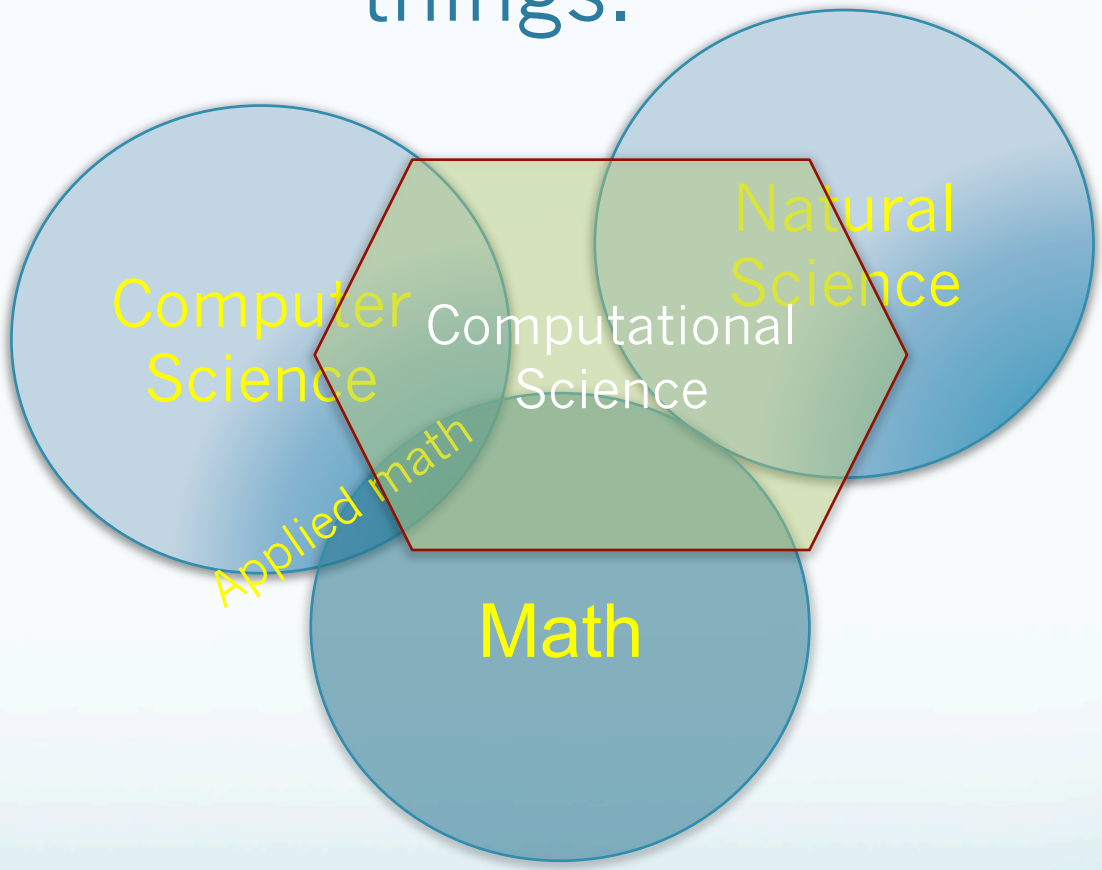
Brain

The UNIVRSE Itself

Degree of being external to creativity:

High Low

Computational Science/Scientific Computing in the grand scheme of things.



How it works

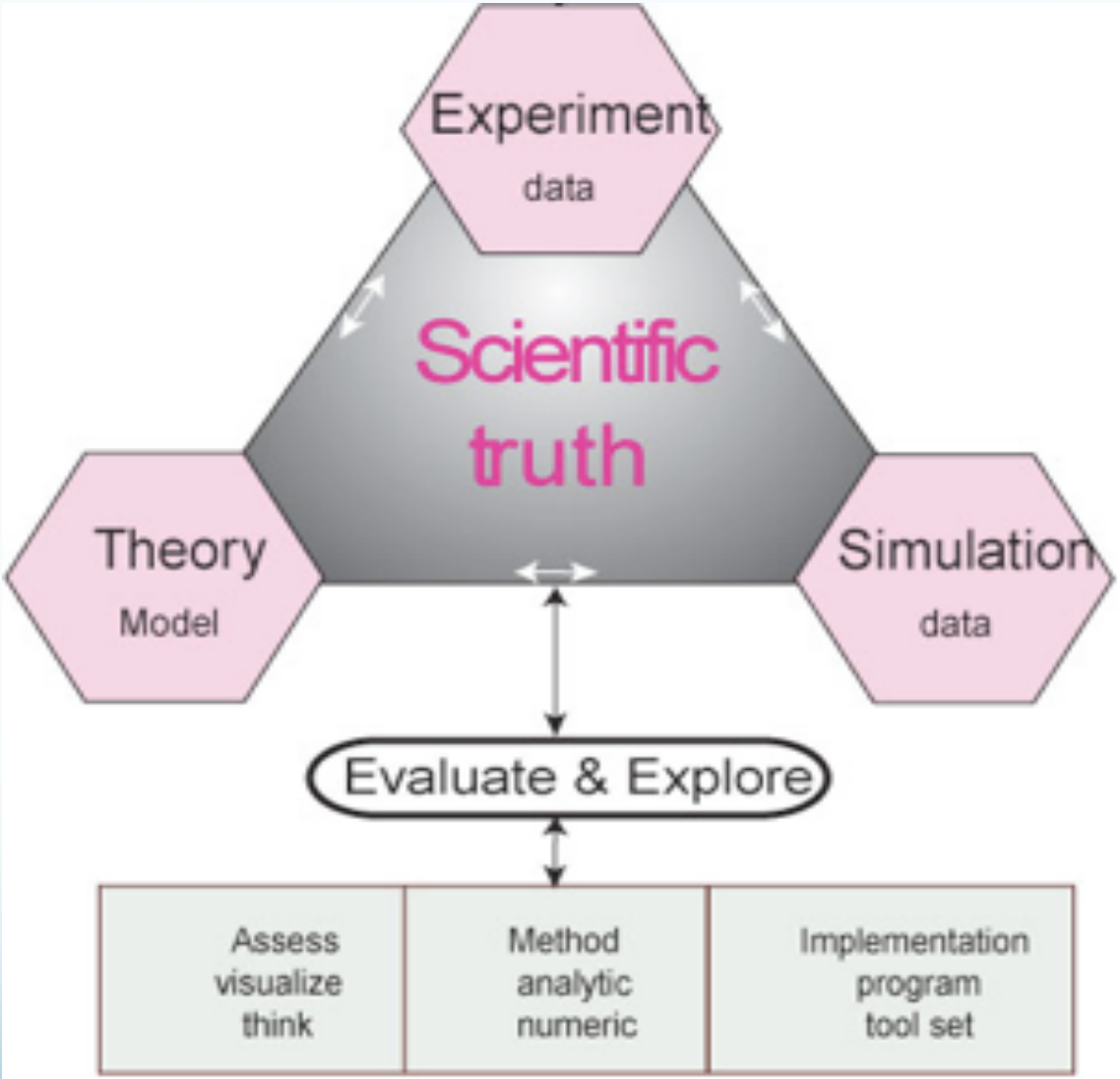
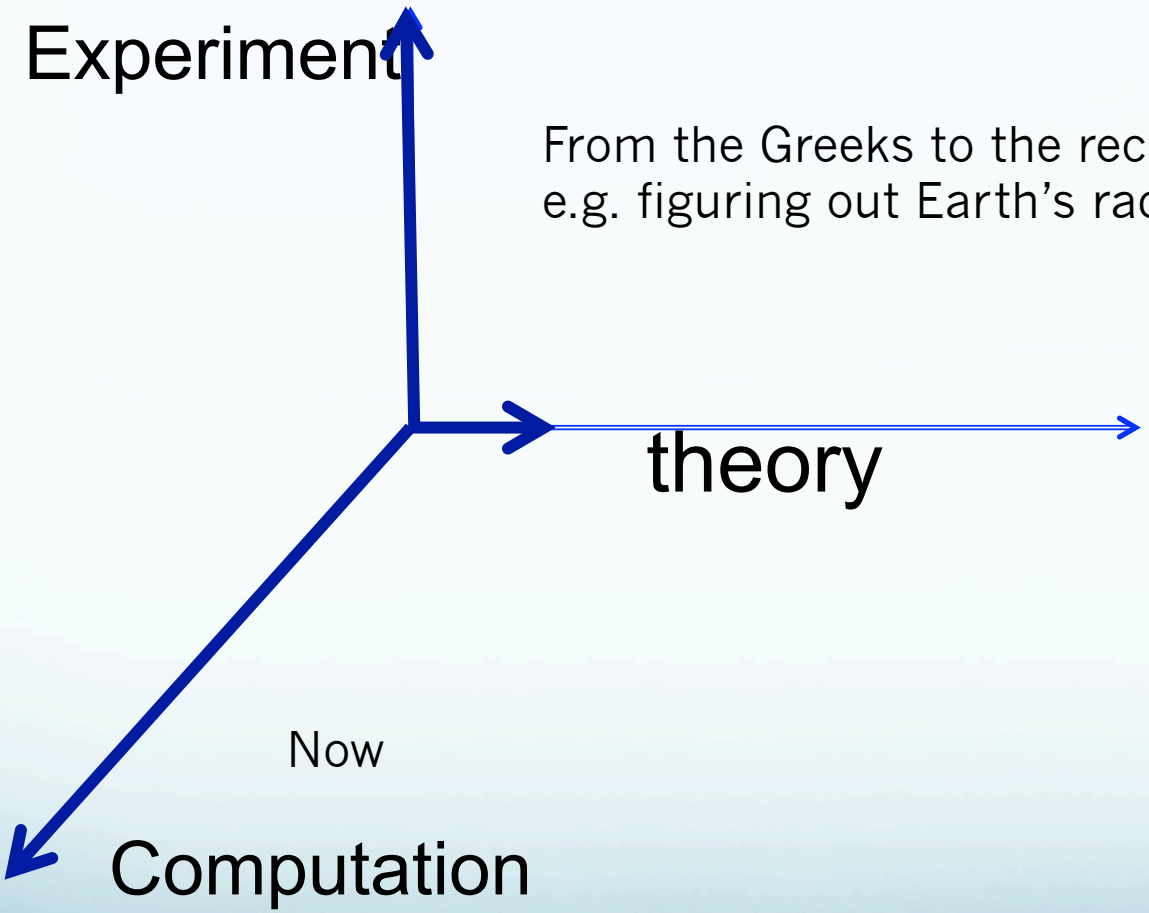


Image credit: http://www.physics.orst.edu/~rubin/INSTANCES/index.html

Role of computation in Natural Sciences now: a paradigm shift



From the Greeks to the recent past, e.g. figuring out Earth's radius

Now

What can computational science accomplish?

- Speed up finding a solution
- Find additional solutions
- Find solutions otherwise impossible to find
- Discover new fundamental laws???

My own observation:
Computationally Constrained Thinking.

- ~95% of modern Physics is computational, one way or another.
- ~50% of Chemistry
- Even some biology
- In all of the above, one always has to keep in mind what's feasible computationally when starting on a problem.

Examples

- Human Genome
- Stealth Fighter
- MRI
- LHS
- Weather Forecasts
- Protein Folding

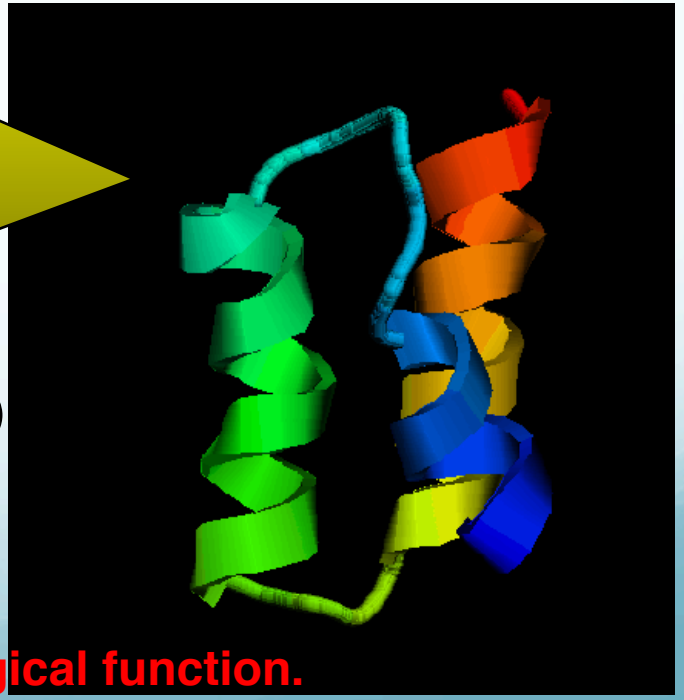
THEME II *The protein folding challenge.*

Nature does it all the time. Can we?

Amino-acid sequence – translated genetic code.

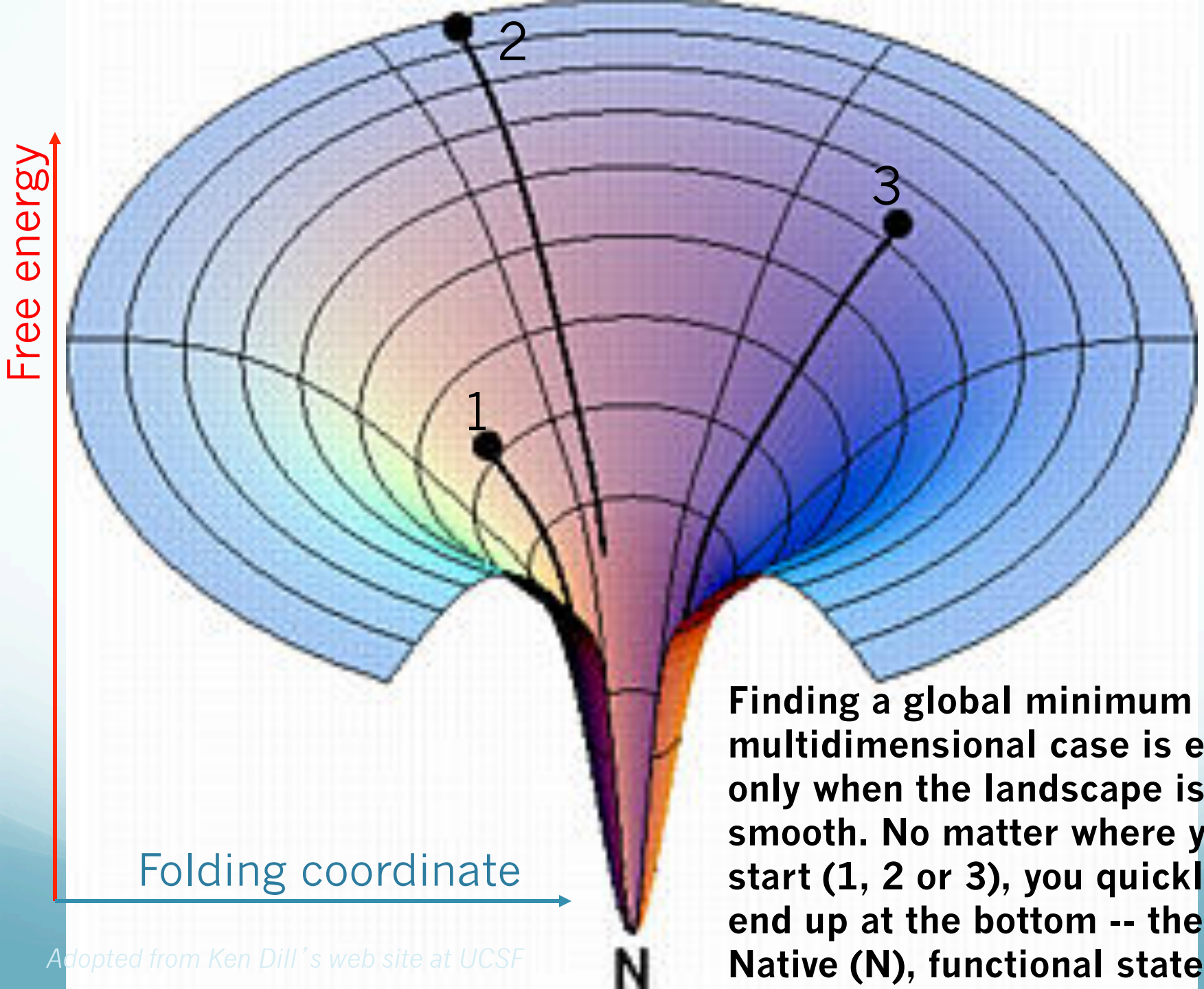
MET—ALA—ALA—ASP—GLU—GLU--….

How?



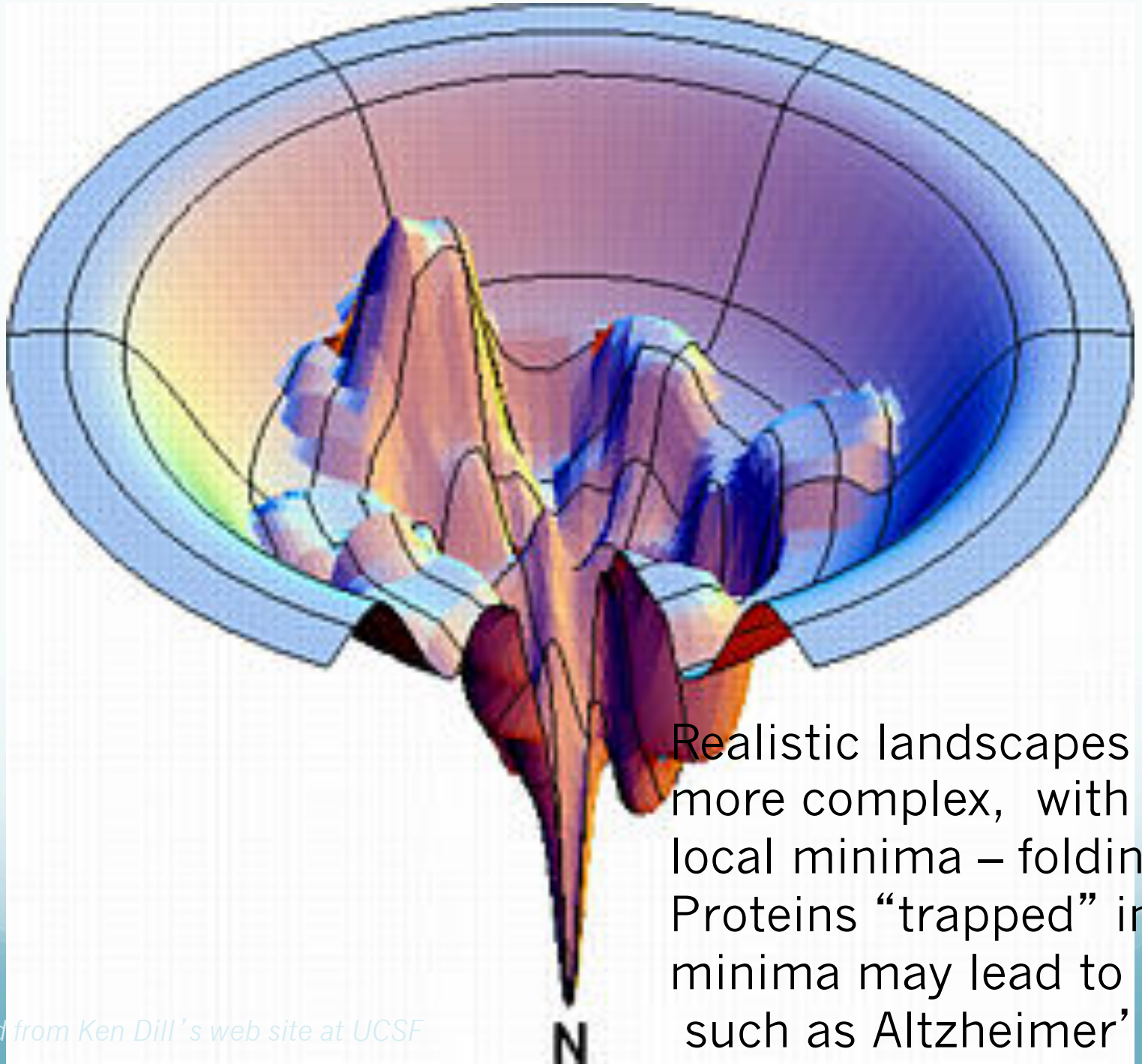
Experiment: amino acid sequence uniquely determines protein's 3D shape (ground state)

Why bother: protein's shape determines its biological function.



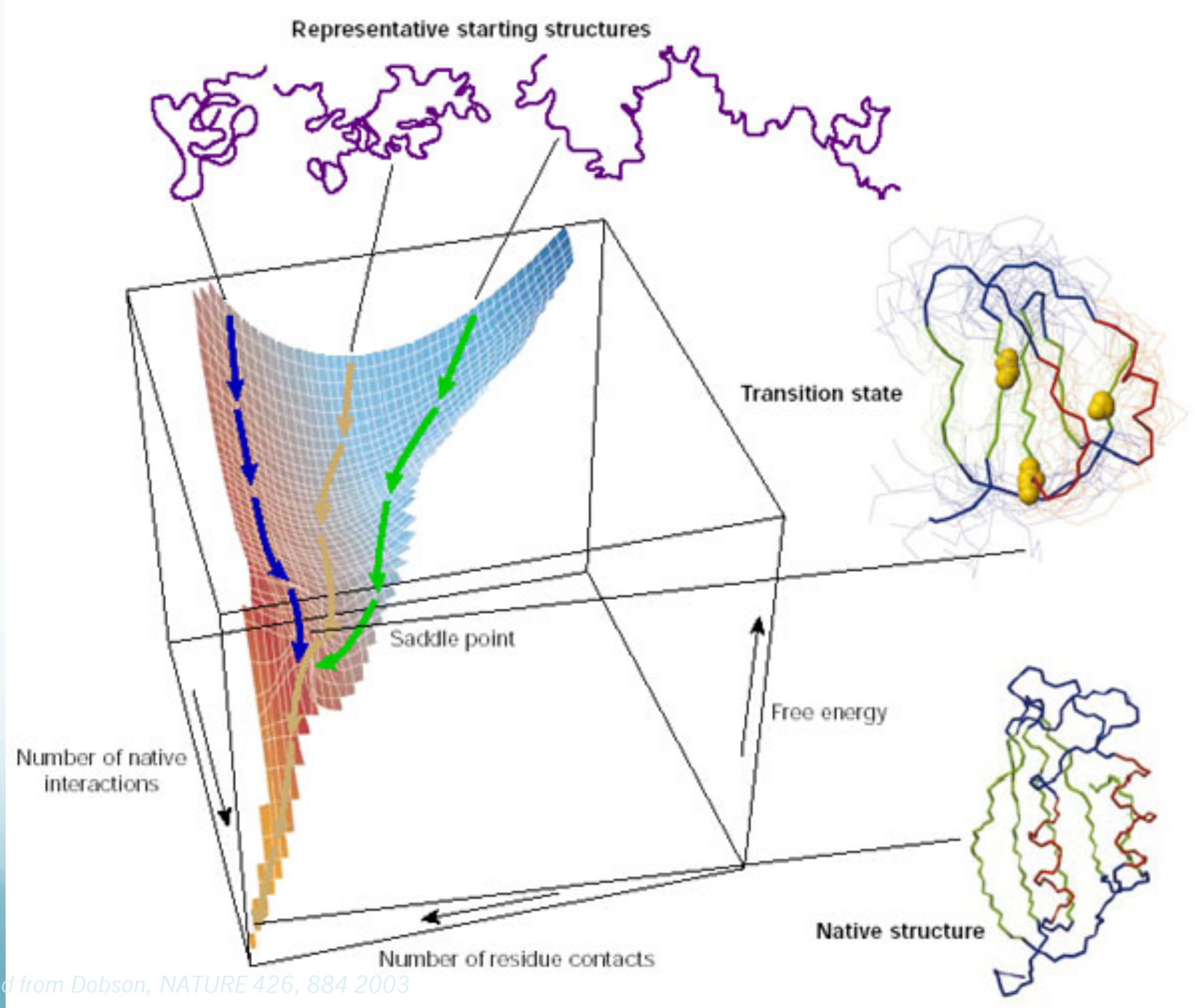
Finding a global minimum in a multidimensional case is easy only when the landscape is smooth. No matter where you start (1, 2 or 3), you quickly end up at the bottom -- the Native (N), functional state of

Adopted from Ken Dill's web site at UCSF



Realistic landscapes are mu
more complex, with multip
local minima – folding traps
Proteins “trapped” in those
minima may lead to diseas
such as Altzheimer’s

Adopted from Ken Dill’s web site at UCSF



Adopted from Dobson, NATURE 426, 884 2003

Since minimization won't work, choose an alternative.

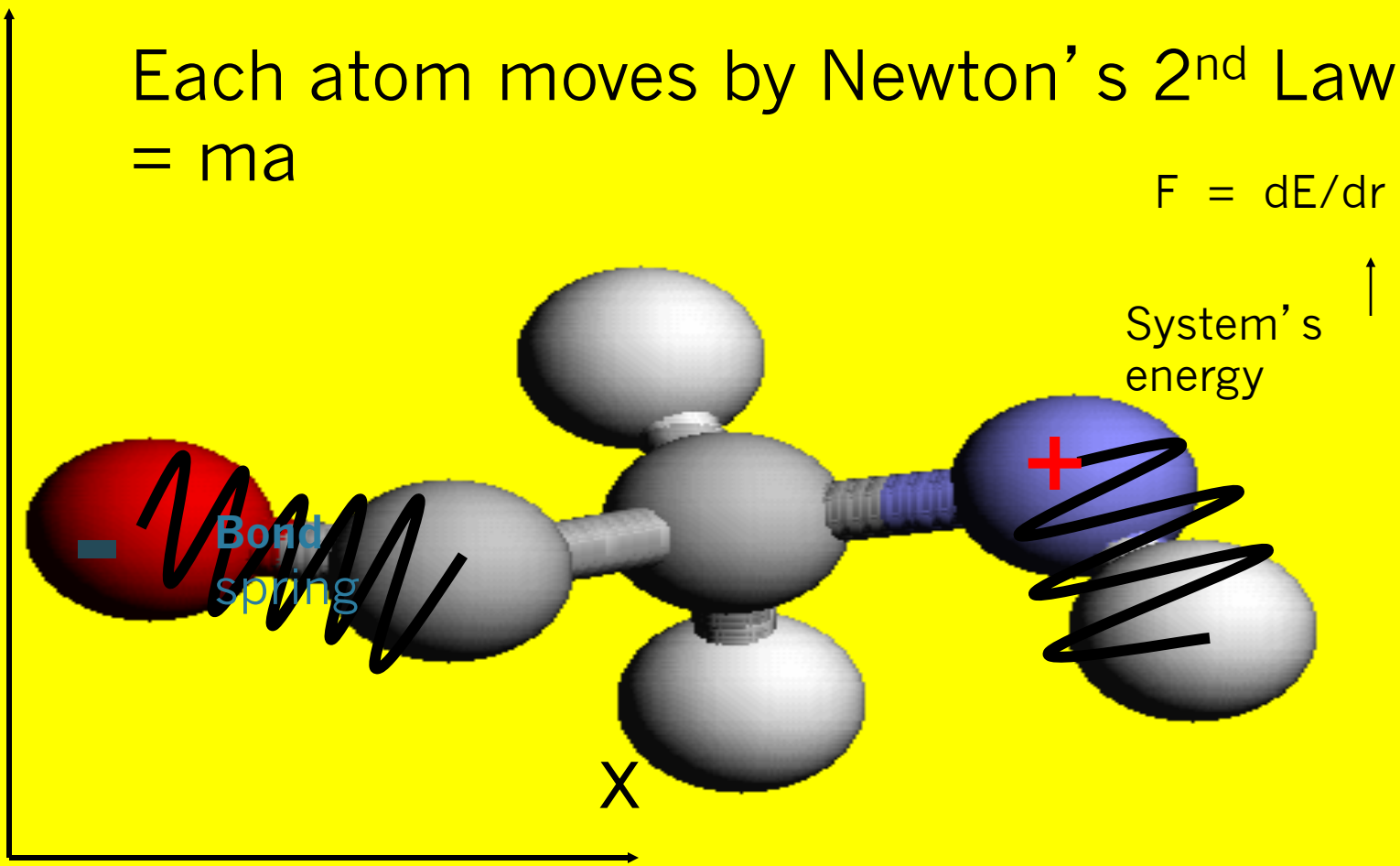
Do what Nature does: just let it fold on its own, at normal temperature.

Method: Molecular Dynamics

Principles of Molecular Dynamics (MD):

Each atom moves by Newton's 2nd Law: F = ma

$F = dE/dr$



$E = Kr^2 + A/r^{12} - B/r^6 + Q_1Q_2/r + \ldots$

Bond stretching

VDW interaction

Electrostatic forces

Molecular Dynamics:

PRICIPLE:

Given positions of each atom x(t) at time t, its position at nex
t + Δt is given by:

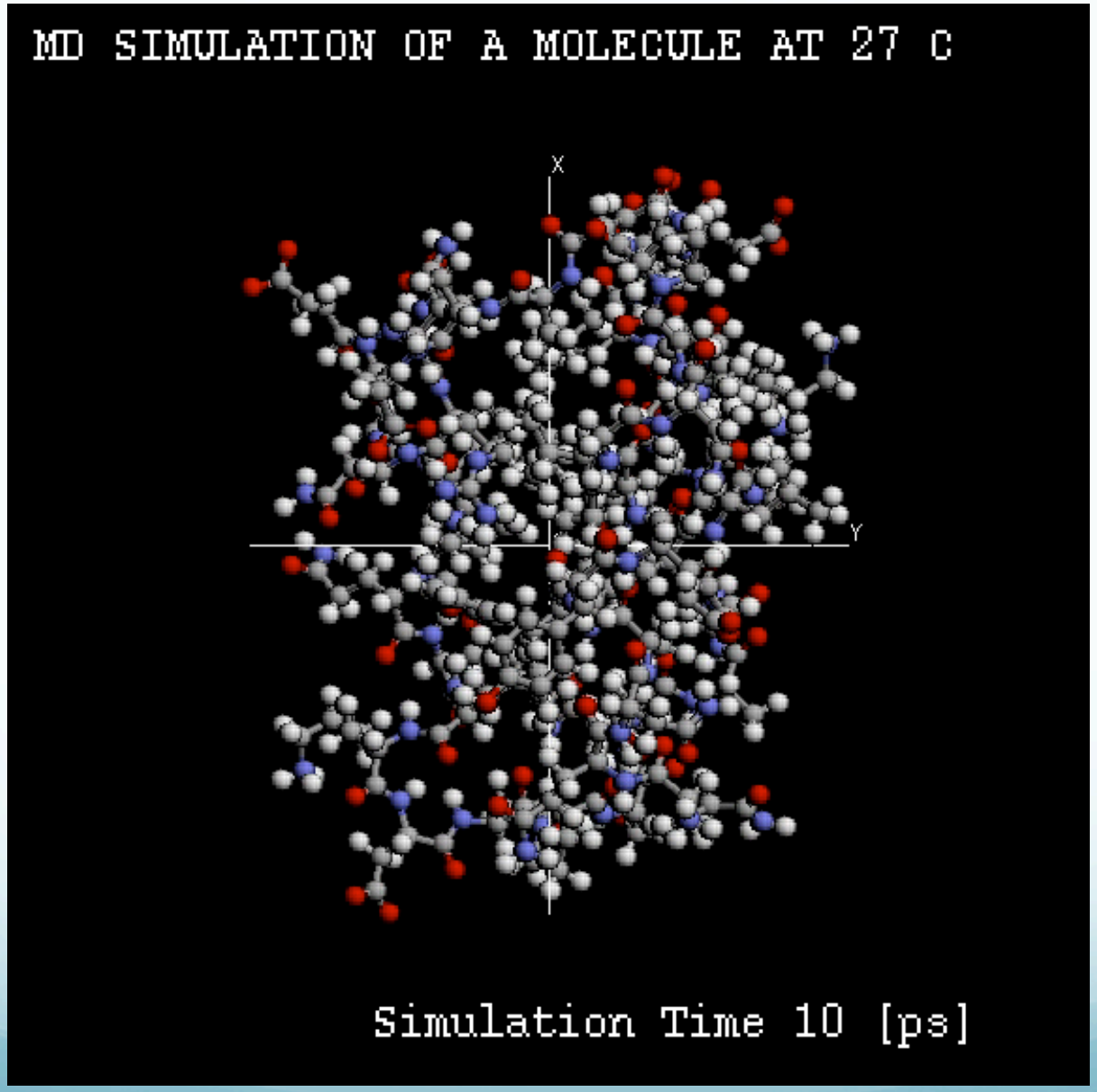
force

$$x(t + \Delta t) \rightarrow x(t) + v(t)\,\Delta t + \tfrac{1}{2} * F/m * (\Delta t)^2$$

Key parameter: *integration time step Δt*. Controls accuracy and
of numerical integration routines.

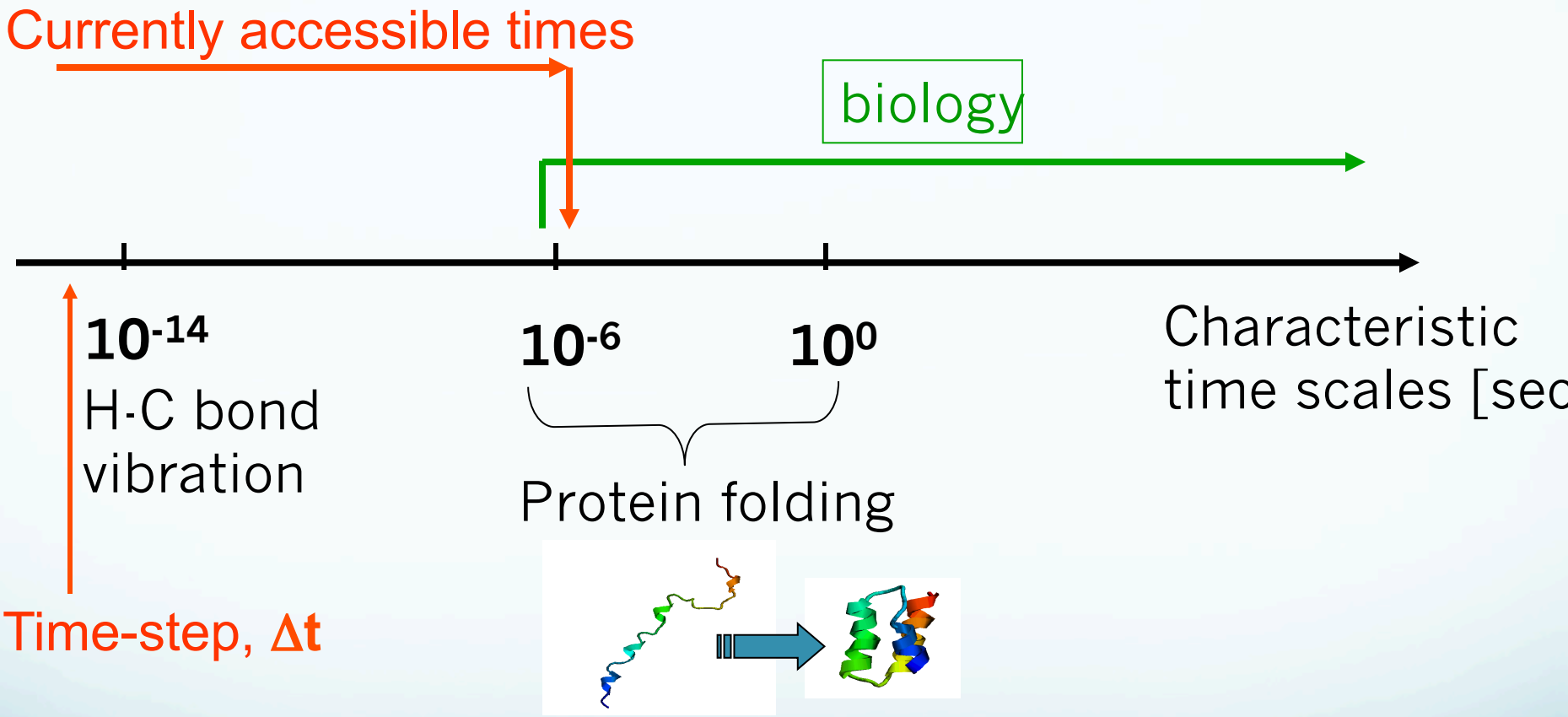
Smaller Δt – more accurate, but need more steps.
How many needed to simulate biology? How many can one afford?

Now we have positions of all atoms as a function of time.



Can compute statistical averages, fluctuations; Analyze side chain movements, Cavity dynamics, Domain motion, Etc.

As a result, we can not quite get into the "biological" time scales.



For stability, Δt must be at least an order of magnitude less than the fastest motion, *i.e* **Δt ~ 10^{-15} s.** **Example: to simulate folding of the fastest folding protein**, at least $10^{-6}/10^{-15}$ = **10^{9} steps will be needed** .

The bottleneck of the methodology:

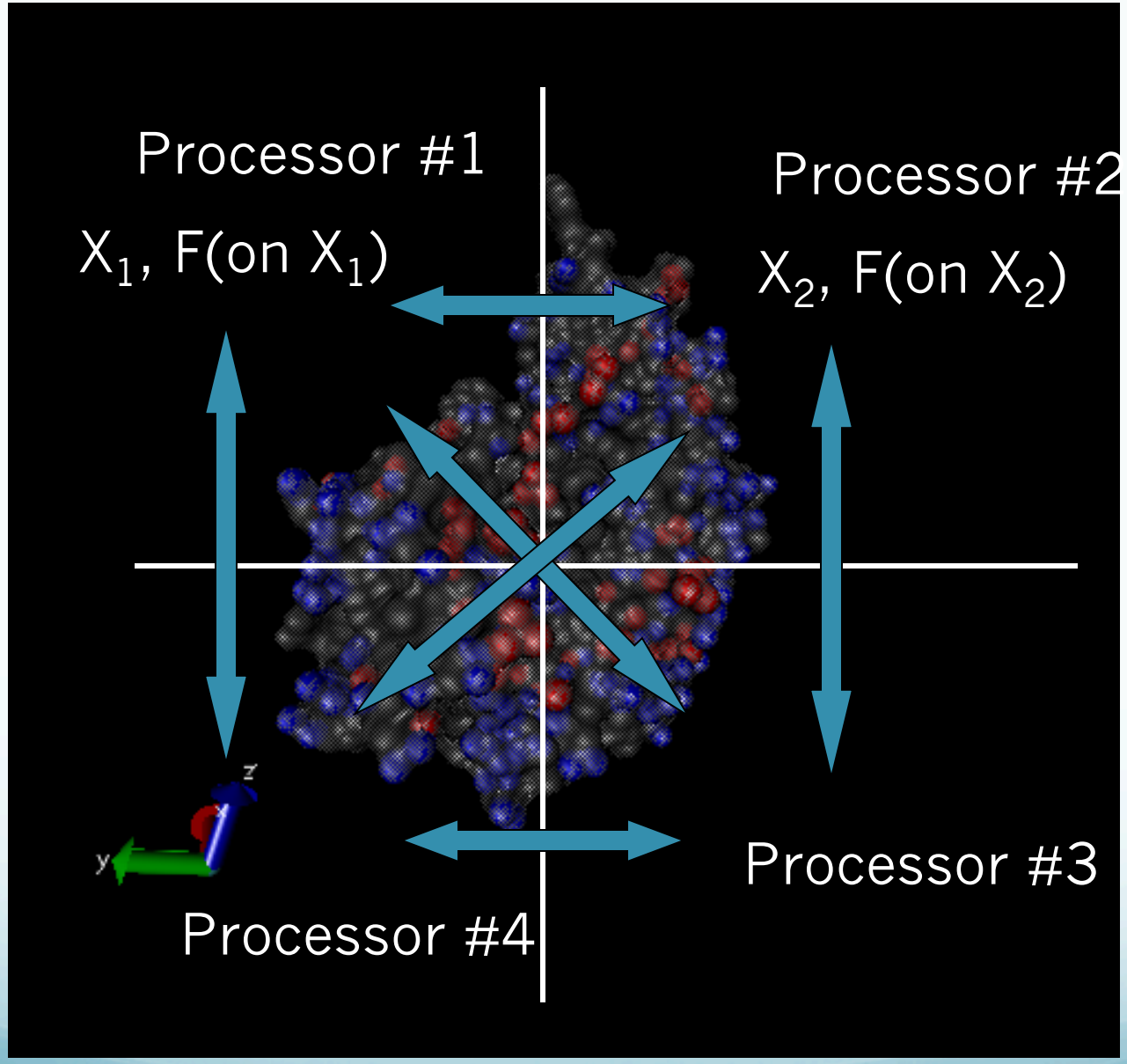
computation of long-range interactions.

Electrostatic interactions fall of as inverse distance between atoms. Too strong to neglect. Need to account for all of them. Very expensive. Up to 99% of total cost for a protein.

Massive parallel machines help.

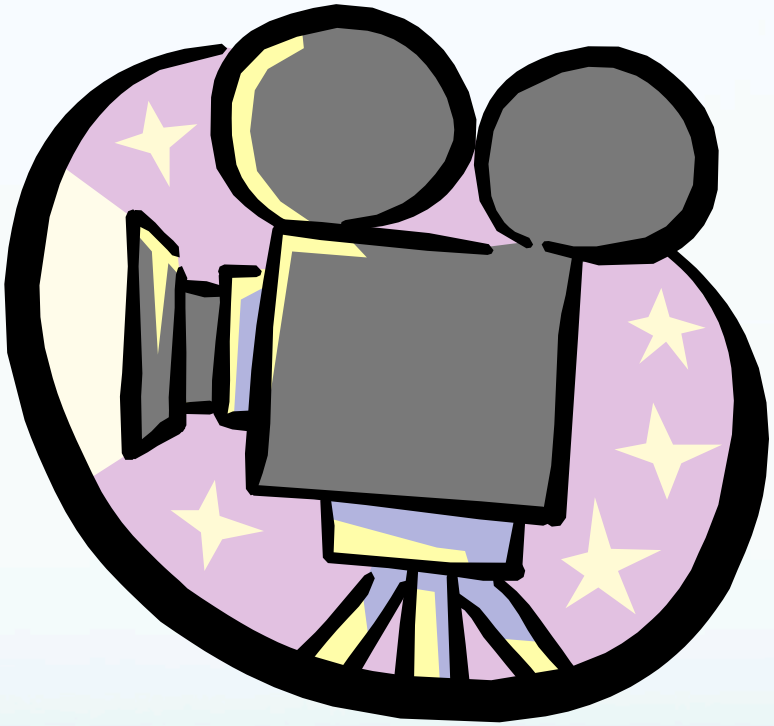
Virginia Tech's supercomputer, System-X

The “worst” problem for parallel computations:



Force acting on each ato depends up positions of every other atom in the system.

Computed coordinates have to be communicat between all processors at each step



Folding Proteins *"in virtuo"*

Starring: Protein A from ***staphylococcus aureus.***

Director: Fine balance between favorable energy and unfavorable entropy.

Screenplay: Molecular Dynamics, implicit solvation.

Producer: NIH Grant 57513

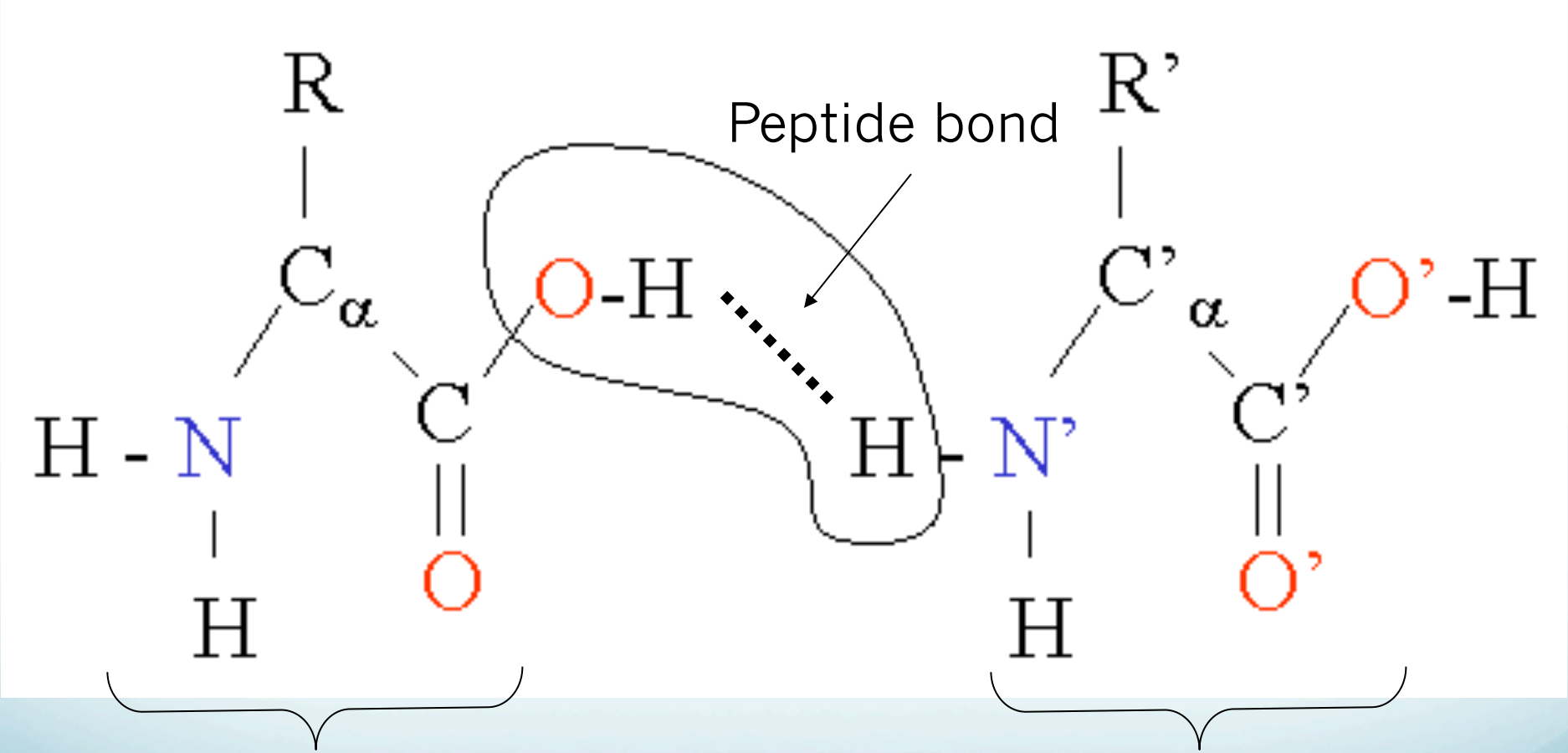
Technical consultant: Alexey Onufriev

The Levinthal's paradox (circa 1969)

- One of the most famous examples of (wrong) application o KT that has had a profound (positive) influence on the field.

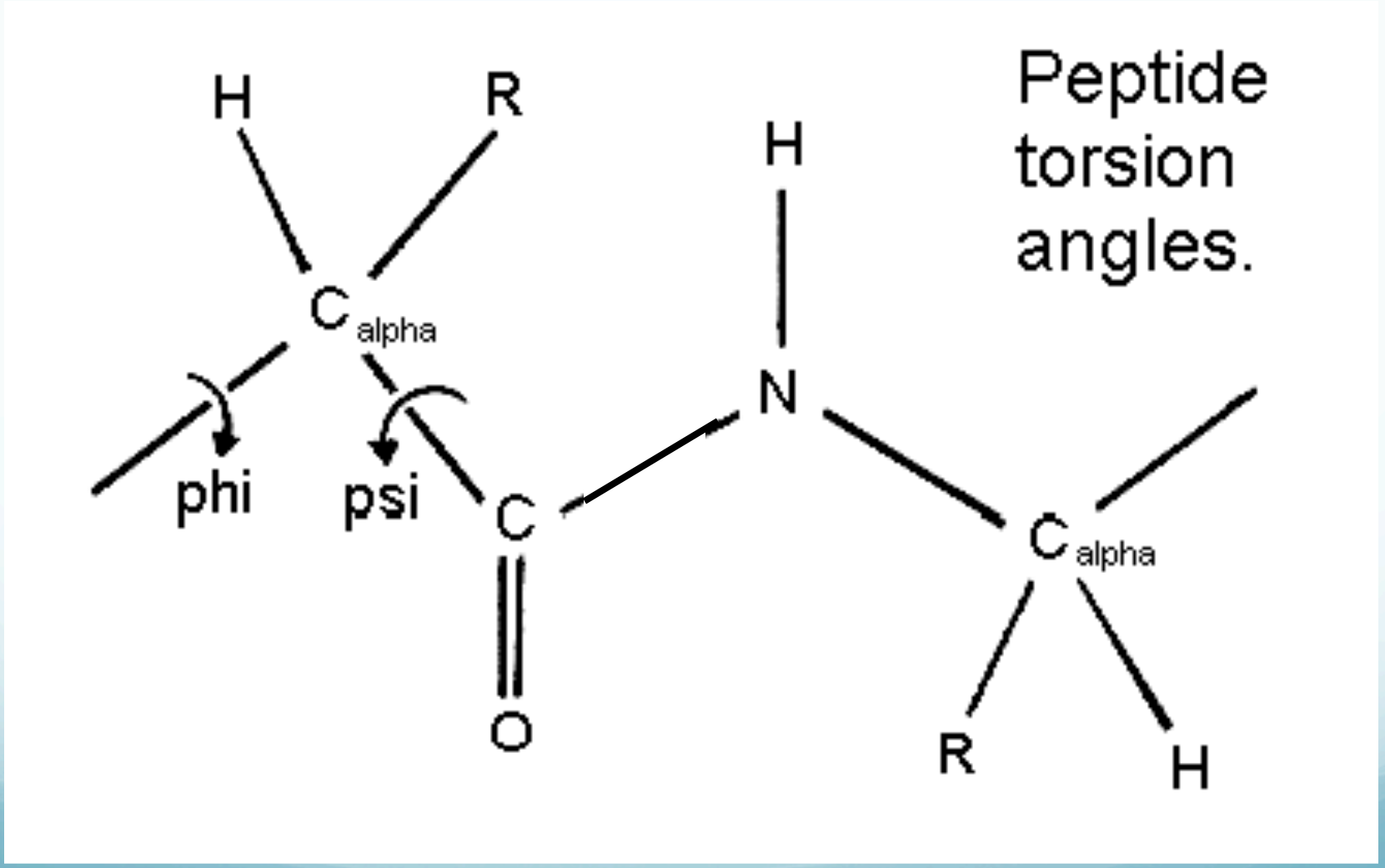
Protein Structure in 3 steps.

Step 1. Two amino-acids together (di-pep

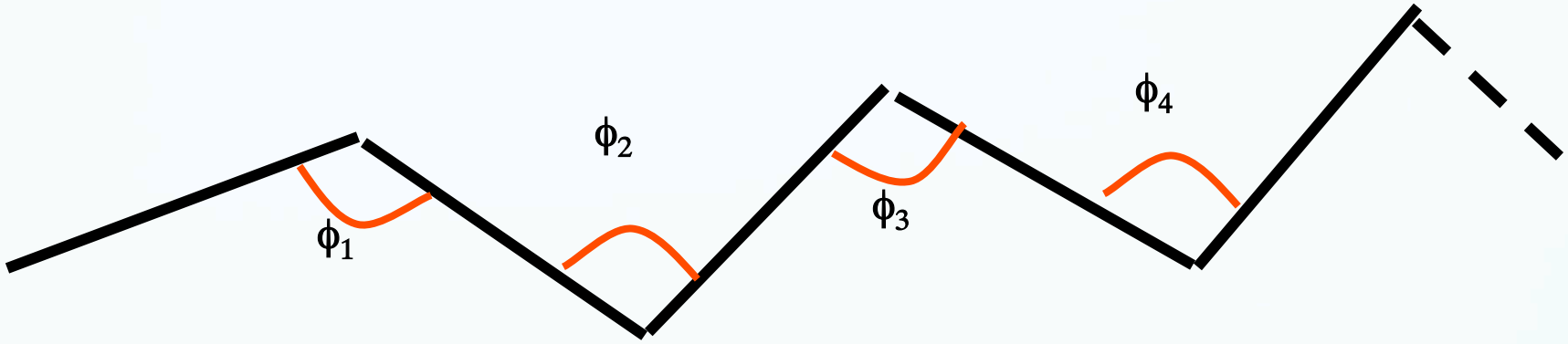


Protein Structure in 3 steps.

Step 2: Most flexible degrees of freedom:



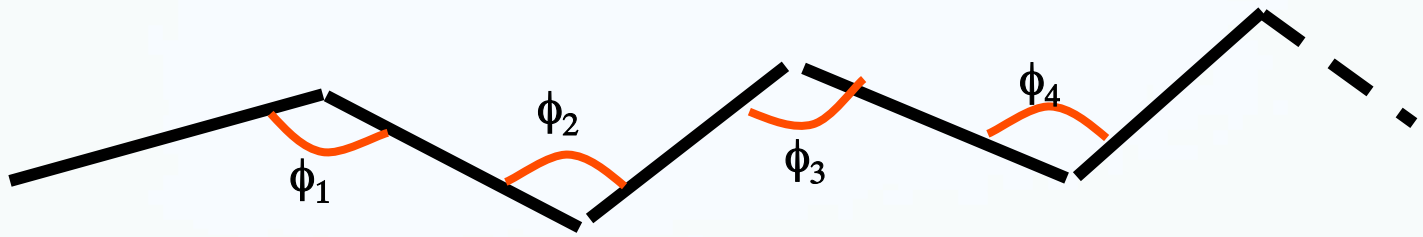
A protein is simply a chain of amino-acids:



Each configuration $\{\Phi 1, \Phi 2, \ldots \Phi \kappa\}$ has some energy. The folded (biologically functional) protein has the lowest possible energy - global minimum. So just find this conformation by some kind of a minimization algorithm... what's the big deal?

The magnitude of the protein folding challenge:

Enormous number of the possible conformations of the polyp



A typical protein is a chain of ~ 100 amino acids.

Assume that each amino acid can take up only 10 conformations (vast underestimation)

Total number of possible conformations: 10^{100}

Suppose Nature "makes" each energy estimate in just 10^{-15} seconds (which is about right, from how long it takes for these things to move).
In today's language, just 1 float point operation on a Penta-Flop supercomputer.

An exhaustive search for the global minimum would take 10^{85} seconds ~ $3*10^{78}$ years. Age of the Universe ~ $2*10^{10}$ years.

Levinthal's paradox: Proteins CAN NOT fold!

Intrigued?

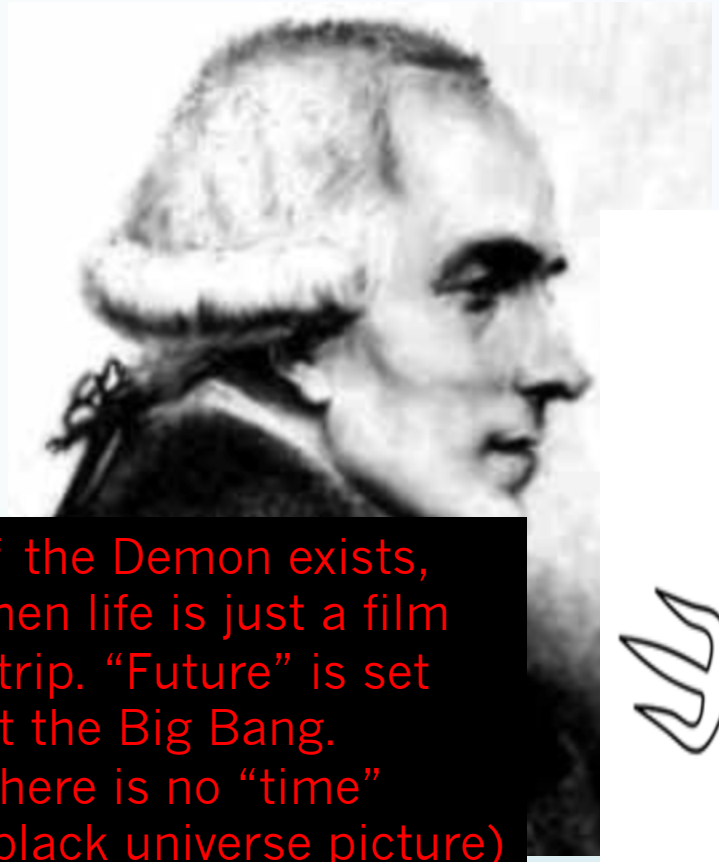
Suggested articles:

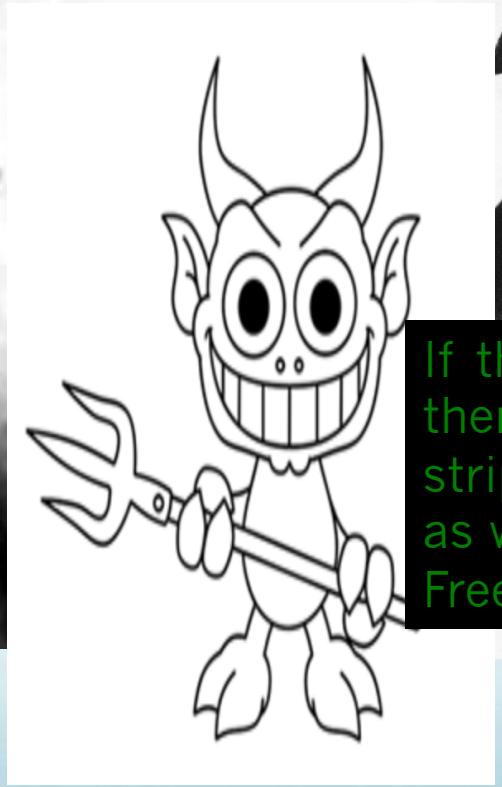
1. "Protein Folding and Misfolding", C. Dobson, *Nature* 426, 884 (2003).
2. "Design of a Novel Globular Protein Fold with Atomic-level Accuracy, Kuhlman et al. , *Science*, 302, 1364, (2003)

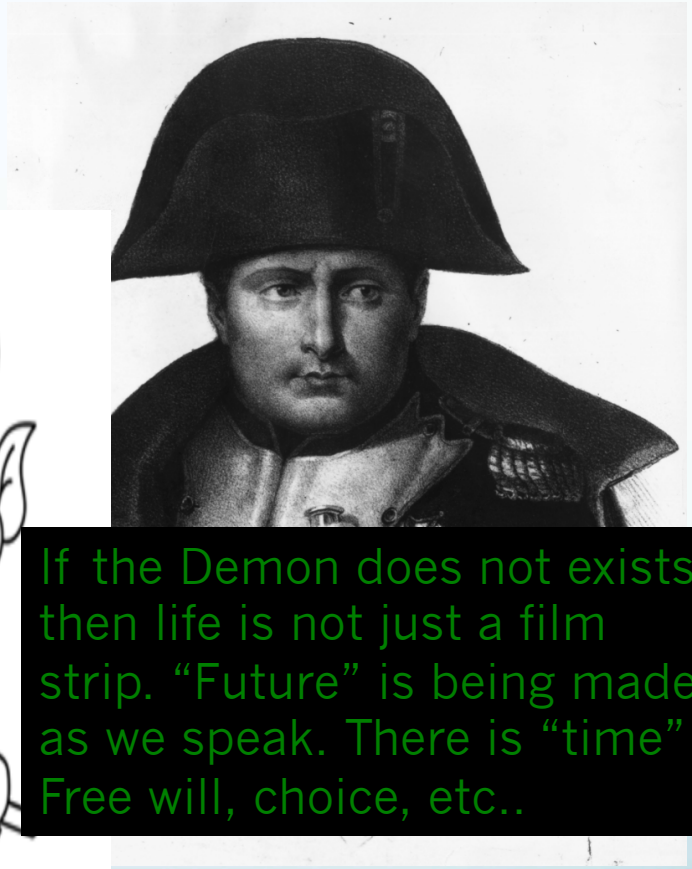
+ references therein.

Laplace's Demon

Laplace's Demon



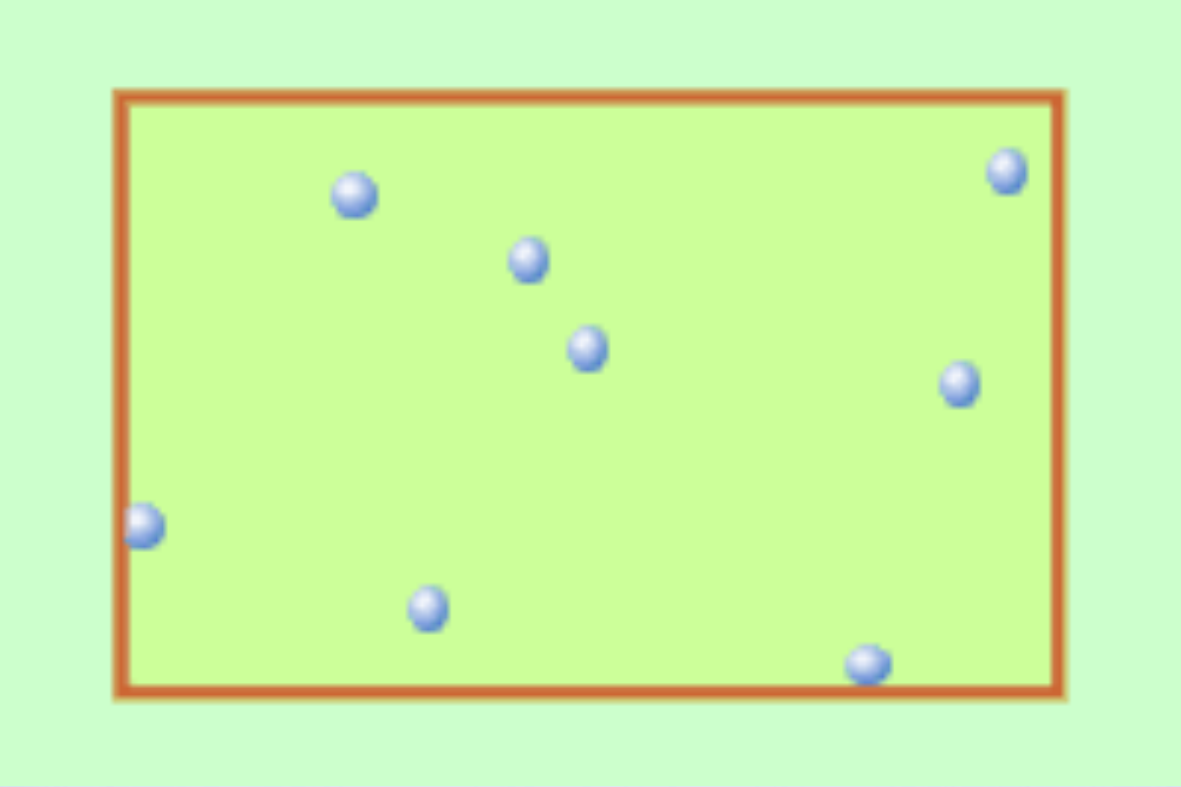




If the Demon exists, then life is just a film strip. "Future" is set at the Big Bang. There is no "time" (black universe picture) No free will.

If the Demon does not exists, then life is not just a film strip. "Future" is being made as we speak. There is "time" Free will, choice, etc..

Explore dynamic stability and instability using simple models

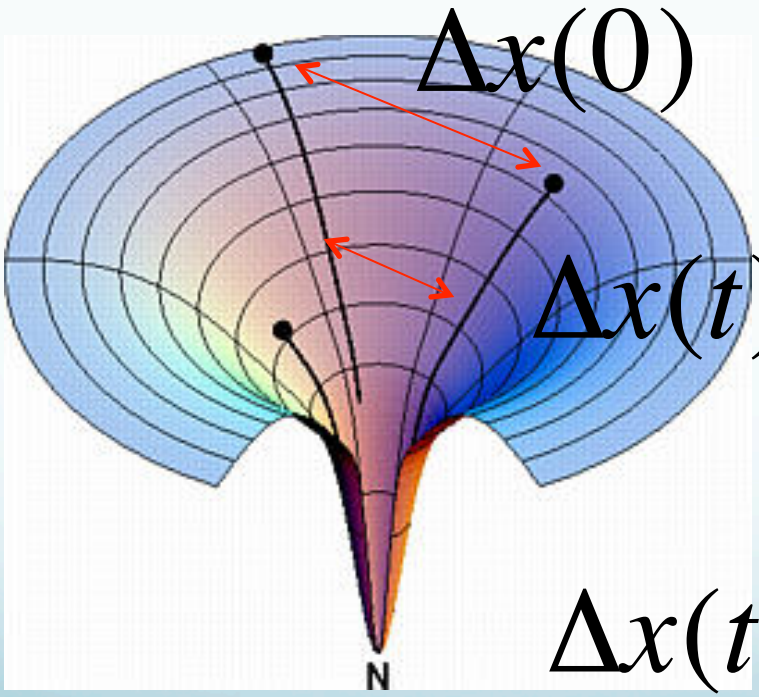


The three possibilities for how the World works:

1. Trajectories always converge
2. Trajectories stay parallel to each other
3. Trajectories diverge.

Possibility #1 for how the World works:

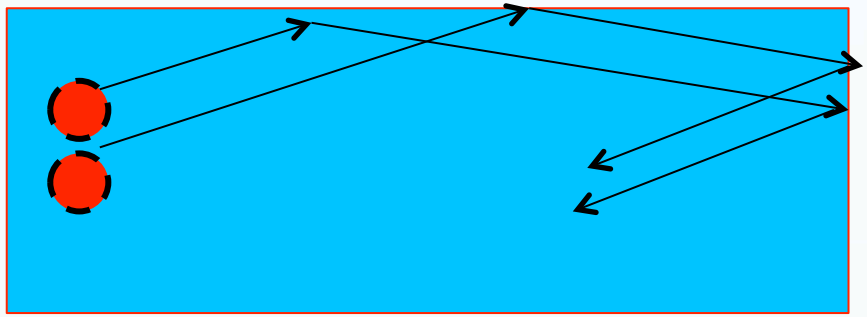
1. Trajectories converge $\lambda < 0$



$$\Delta x(t) = \Delta x(0)\exp(\lambda t)$$

The three possibilities for how the Universe works:

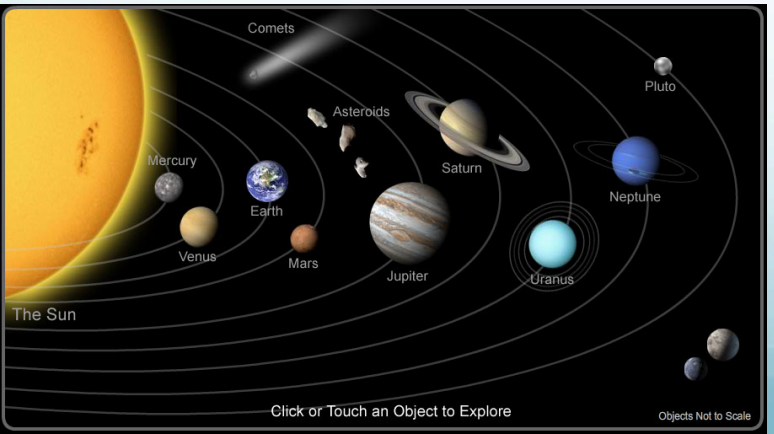
#2 :Trajectories always stay parallel to each other



$$\Delta x(t) = \Delta x(0)\exp(\lambda t)$$

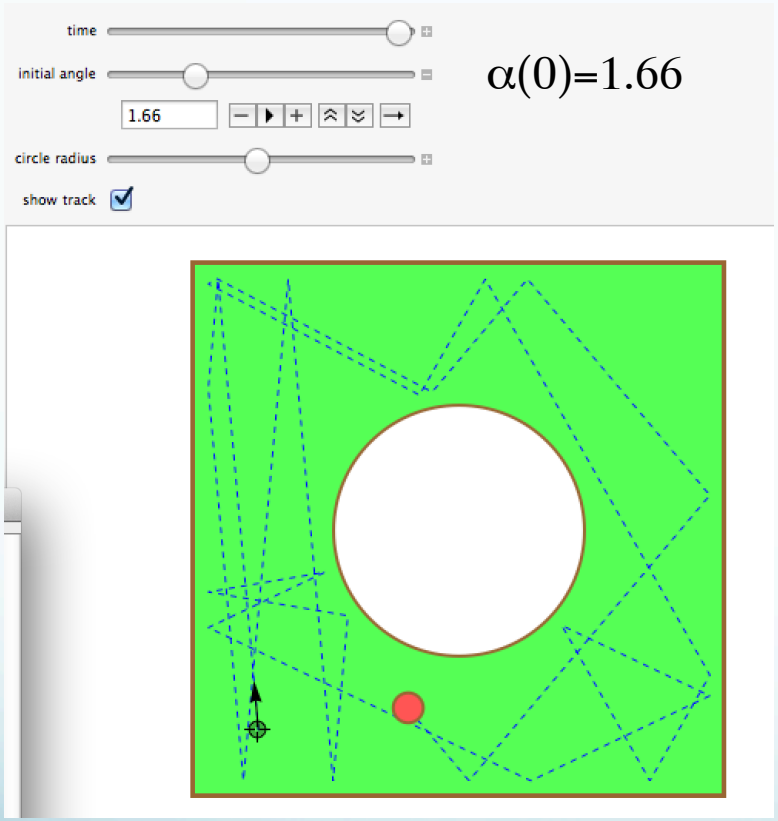
$$\lambda = 0$$

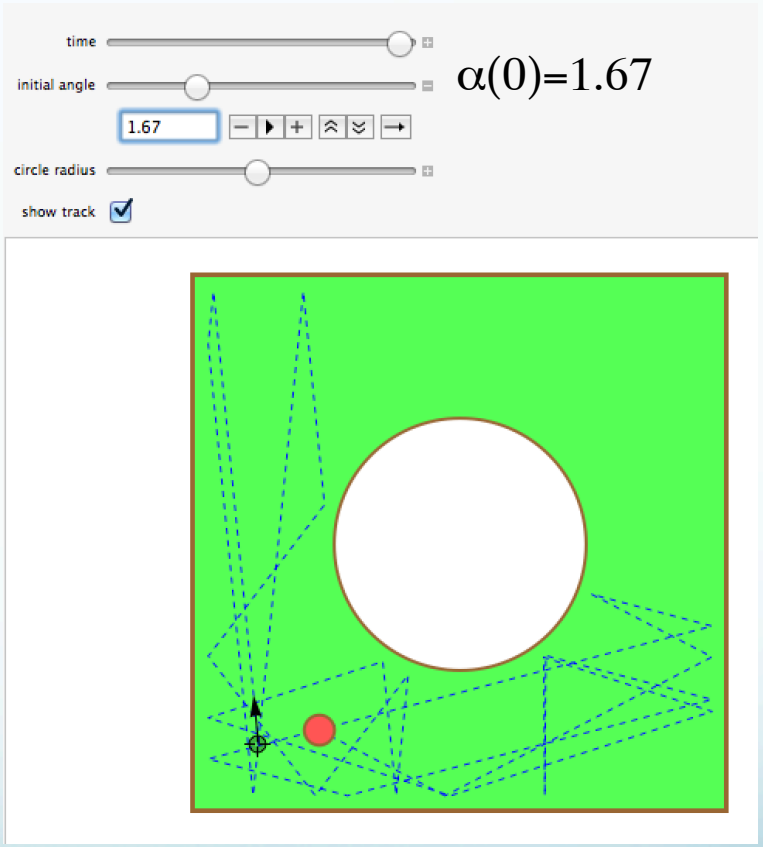
Other familiar systems with stable trajectories?



The three possibilities for how the World works:

#3 Trajectories diverge.





$$\Delta x(t) = \Delta x(0)\exp(\lambda t) \quad \lambda > 0$$

Which one is our World?

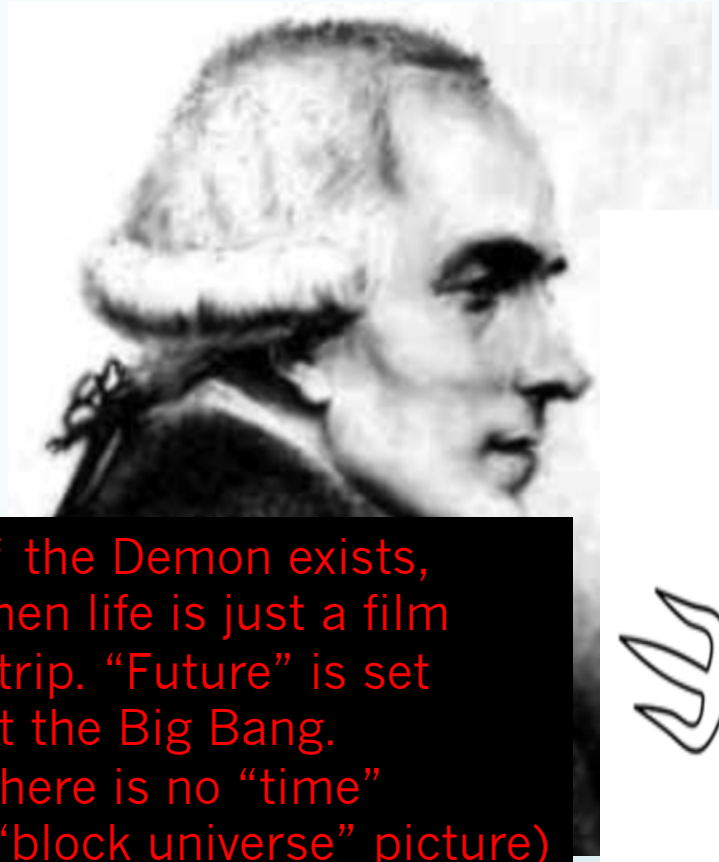
$\lambda<0$ [future super easy to predict],

$\lambda=0$ [future easy to predict],

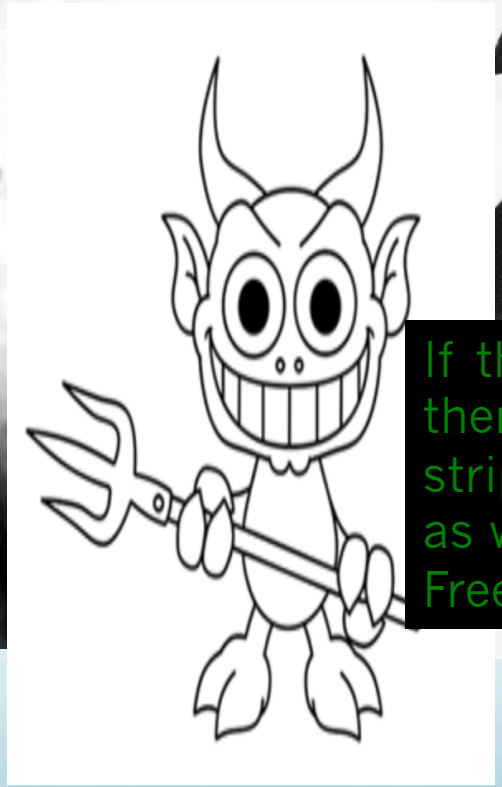
$\lambda>0$ [future is hard or impossible to predict]

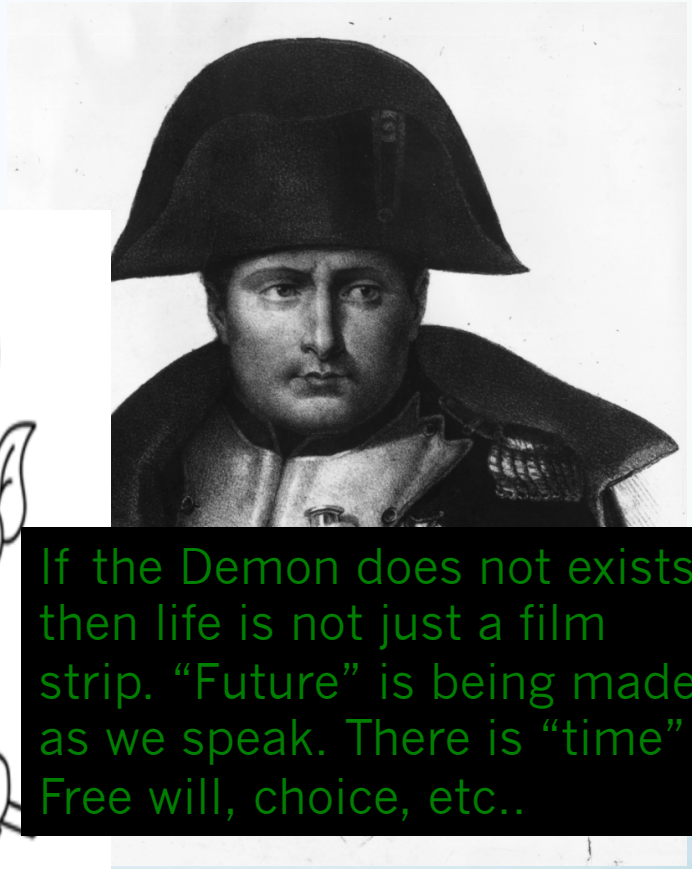
Profound implications for physics, philosophy, judicial system, religion, ...

Laplace's Demon



If the Demon exists, then life is just a film strip. "Future" is set at the Big Bang. There is no "time" ("block universe" picture) No free will.





If the Demon does not exists, then life is not just a film strip. "Future" is being made as we speak. There is "time" Free will, choice, etc..

Is our Universe computable?

$\lambda<0$ (yes),

$\lambda=0$ (most likely),

$\lambda>0$ (Most likely no, but) ?

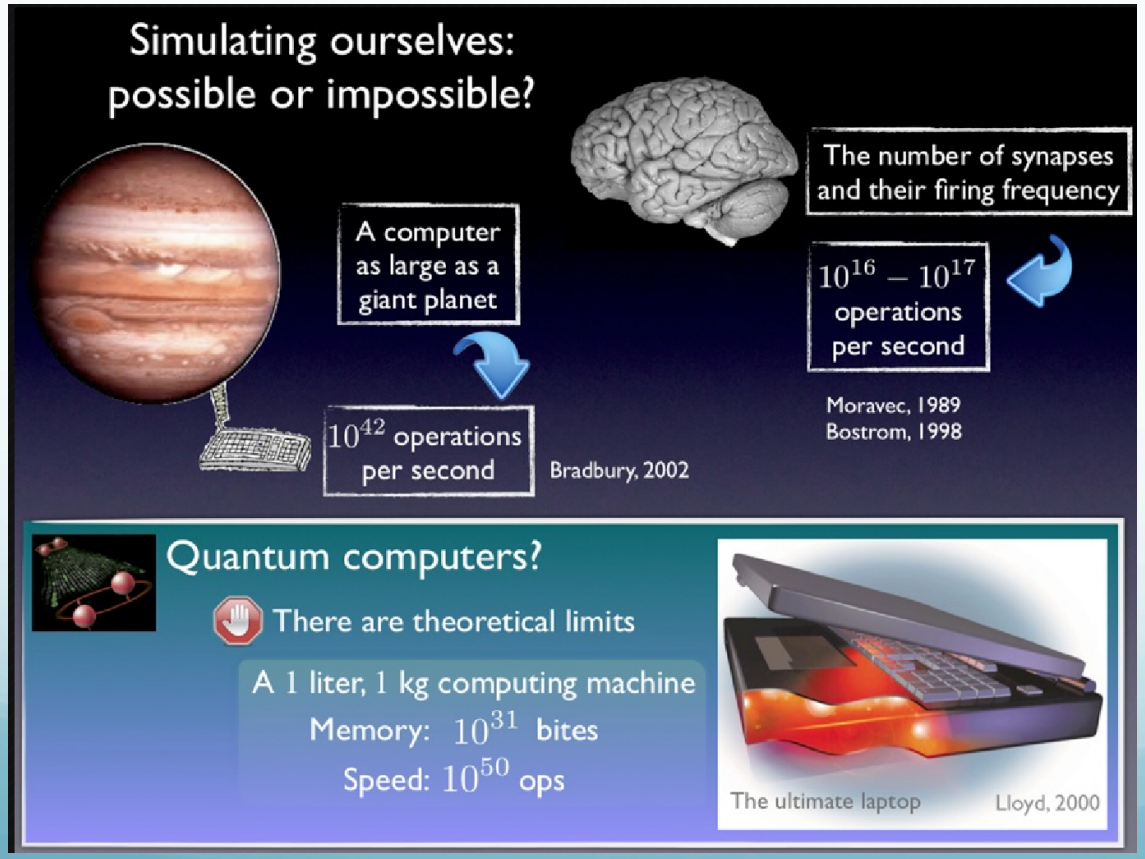
The answer depends on the structure of space at the smallest scale and/or the boundary conditions at the large scale ("end" of the Universe). A lattice Universe can be computable.

http://www.phys.washington.edu/users/savage/Simulation/Universe/

Related arguments:

- Our Universe is just a numerical simulation.

  (e.g. Nick Bostrom, Philosopher, Oxford: "At some point, humans will be able to simulate parts of its own history. Then, it is likely we are just being simulated. "13th floor" movie.)



A quantum version of Church-Turing conjecture.

Related arguments:

- The Brain function != computation.
  (e.g. Roger Penrose, Mathematical Physics, Oxford)

The argument is based on Goedel's incompleteness theorem.

Basically, it goes like this: "Mathematical understanding
is incomplete and contradictory,
hence can't be reduced to algorithmic rules, hence can't be created by a logical machine. Yet, we have created it".

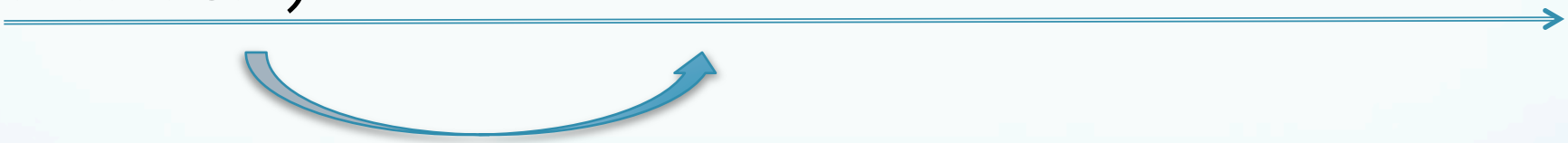
The place of computational thinking in the context of human creativity.

Tools (e.g. paper, screwdriver, calculator)

LANGUAGE

Brain

The UNIVRSE Itself



Degree of being external to creativity:

High Low