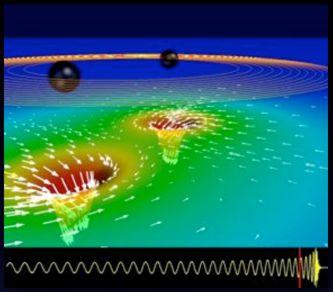
Engineering Computational Science & Engineering

Mark Stalzer
Center for Advanced Computing Research
California Institute of Technology
stalzer@caltech.edu

May 26, 2011

This is an informal talk about making science & engineering happen faster using algorithms & computers...

CACR Mission & Partners



Computational Infrastructure for Geodynamics

bnmc

cell center

SEISMOLOGICAL LABORATORY
California Institute of Technology

LIGO

US CMS

UltraLight

JPL
Jet Propulsion Laboratory
California Institute of Technology

CPES
Center for Plasma Edge Simulation

VAO

DANSE

GALCIT
MECHANICS: FLUIDS, SOLIDS, BIOSYSTEMS, AND SPACE

ASC™

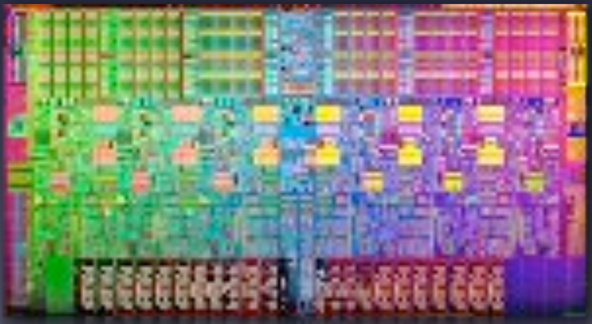
CALTECH
PSAAP

Accelerating scientific discovery & engineering through advanced computation, collaboration and research

CACR

Intel Touchstone Delta:
World's Fastest Computer in 1991
(30 Gflops)

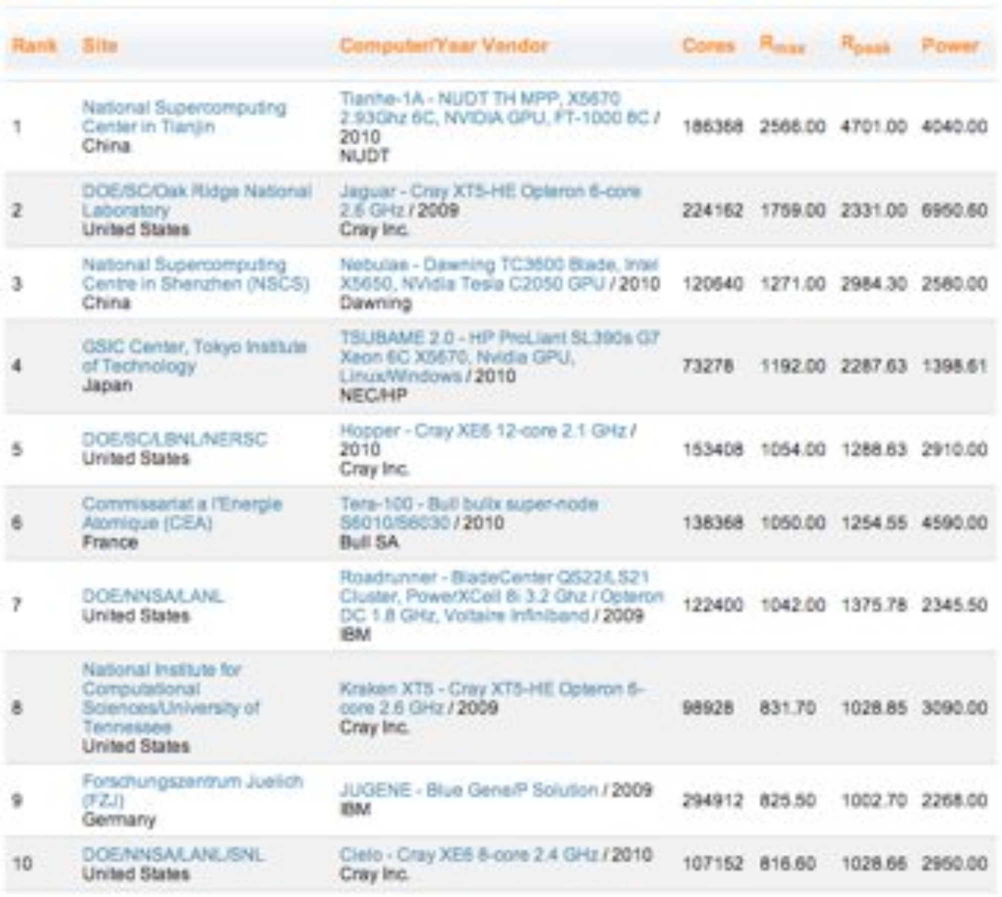
Westmere-EX
(60 Gflops)



Limitations of Hardware Performance Measures

Measure: How fast does the machine go [flops]?

- Important, but can be disconnected from science
- CSE problems do not *necessarily* need the fastest machines
- Dot product machines distort R&D - What if BLAST was the key app?



Rank	Site	Computer/Year Vendor	Cores	R_{max}	R_{peak}	Power
1	National Supercomputing Center in Tianjin China	Tianhe-1A - NUDT TH MPP, X5670 2.93Ghz 6C, NVIDIA GPU, FT-1000 8C / 2010 NUDT	186368	2566.00	4701.00	4040.00
2	DOE/SC/Oak Ridge National Laboratory United States	Jaguar - Cray XT5-HE Opteron 6-core 2.6 GHz / 2009 Cray Inc.	224162	1759.00	2331.00	6950.60
3	National Supercomputing Centre in Shenzhen (NSCS) China	Nebulae - Dawning TC3600 Blade, Intel X5650, NVidia Tesla C2050 GPU / 2010 Dawning	120640	1271.00	2984.30	2580.00
4	GSIC Center, Tokyo Institute of Technology Japan	TSUBAME 2.0 - HP ProLiant SL390s G7 Xeon 6C X5670, Nvidia GPU, Linux/Windows / 2010 NEC/HP	73278	1192.00	2287.63	1398.61
5	DOE/SC/LBNL/NERSC United States	Hopper - Cray XE6 12-core 2.1 GHz / 2010 Cray Inc.	153408	1054.00	1288.63	2910.00
6	Commissariat a l'Energie Atomique (CEA) France	Tera-100 - Bull bullx super-node S6010/S6030 / 2010 Bull SA	138368	1050.00	1254.55	4590.00
7	DOE/NNSA/LANL United States	Roadrunner - BladeCenter QS22/LS21 Cluster, PowerXCell 8i 3.2 Ghz / Opteron DC 1.8 GHz, Voltaire Infiniband / 2009 IBM	122400	1042.00	1375.78	2345.50
8	National Institute for Computational Sciences/University of Tennessee United States	Kraken XT5 - Cray XT5-HE Opteron 6-core 2.6 GHz / 2009 Cray Inc.	98928	831.70	1028.85	3090.00
9	Forschungszentrum Juelich (FZJ) Germany	JUGENE - Blue Gene/P Solution / 2009 IBM	294912	825.50	1002.70	2268.00
10	DOE/NNSA/LANL/SNL United States	Cielo - Cray XE6 8-core 2.4 GHz / 2010 Cray Inc.	107152	816.60	1028.66	2950.00

Need a larger context.

Source: top500.org 11/10

Outline

- Introduction
- Algorithmic Complexity
    - How to beat (almost) any Top500 machine with a Mac
- CSE as Systems Engineering
    - The questions change
- Structural Complexity in CSE
    - Examples from CACR
- Some Implications
    - Clear costs of computation lead to design trades

Randomized Interpolative Decomposition

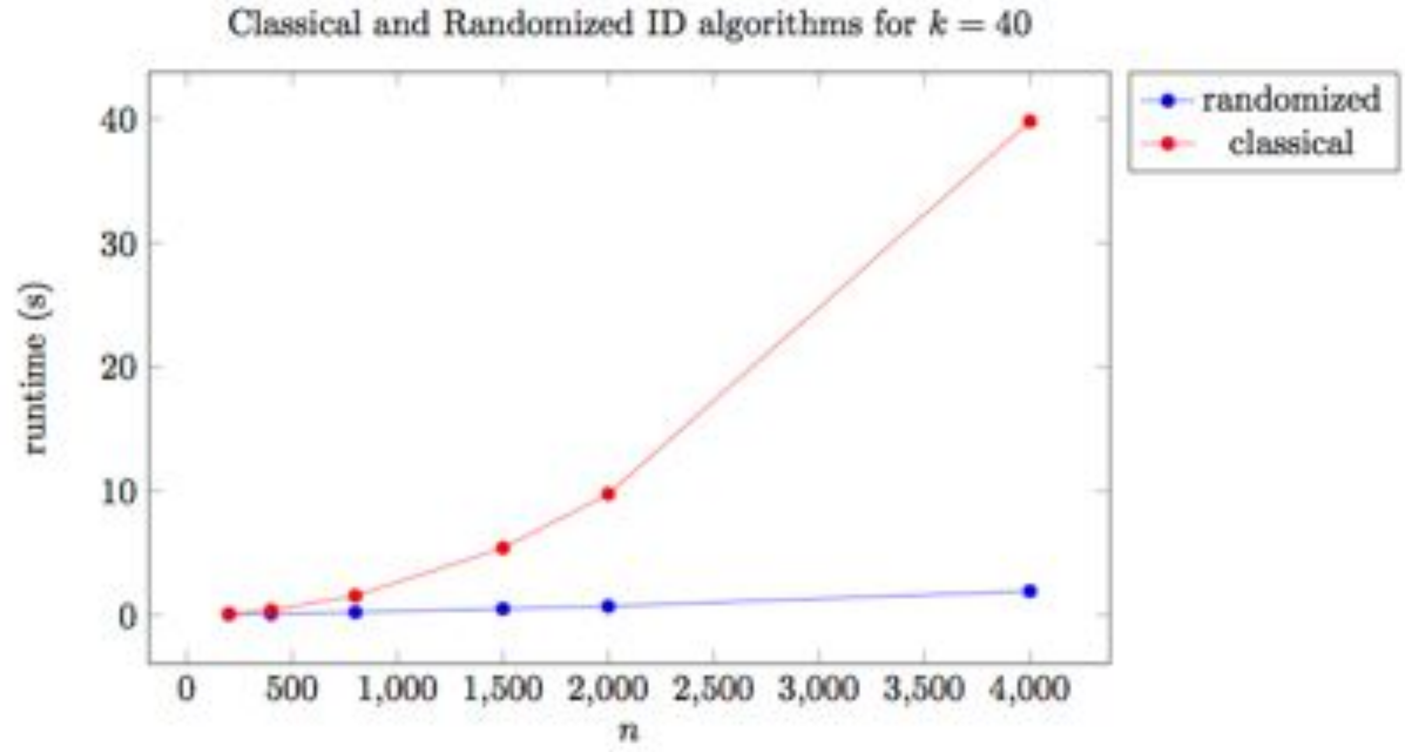


Figure 1: A comparison of the runtimes of the randomized and non-randomized versions of the algorithm. k refers to the rank of the approximation.

- Accelerates $O(mnk)$ to $O(mn \log(l) + kln)$ with $l-k>0$ & small

(See Liberty et al., PNAS 104(51) Dec 2007)

Source: Andy Lucas, Caltech SURF

F117-A “NightHawk”

LA Times 30 Mar 1999

Stealth Fighter’s Crash Reveals a Design’s Limits

… [T]o engineers familiar with stealth technology, one look at the triangle-shaped aircraft speaks volumes about how far science has come since the F117-A first took to the air. The reason its body is made up of flat surfaces, for example, is that ***the 1970’s computers used to design it couldn’t perform calculations to measure the radar resistance of three-dimensional objects.***

Source: John Ottusch, Hughes

Problem Formulation (scalar)

Field given by:

$$\psi(\mathbf{x}) = \int_S G(\mathbf{x}-\mathbf{x}')\sigma(\mathbf{x}')d\mathbf{x}', \text{ where } G(r) = e^{ikr}/r$$

Boundary condition:

$$\varphi(\mathbf{x}) + \psi(\mathbf{x}) = 0, \ \ \mathbf{x} \text{ on } S$$

Discretize and apply bc:

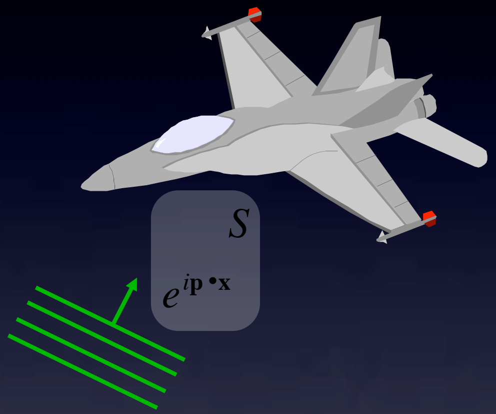
$$\psi(\mathbf{x}) = \sum_i w_i G(\mathbf{x}-\mathbf{x}_i)\sigma(\mathbf{x}_i)$$

$$e^{i\mathbf{p}\bullet\mathbf{x}} = -\sum_i w_i G(\mathbf{x}-\mathbf{x}_i)\sigma(\mathbf{x}_i), \ \ \mathbf{x} \text{ on } S$$

Use same points for $\mathbf{x}$:

$$\mathbf{V} = \mathbf{ZI}$$

FMM accelerates computation of **ZI**



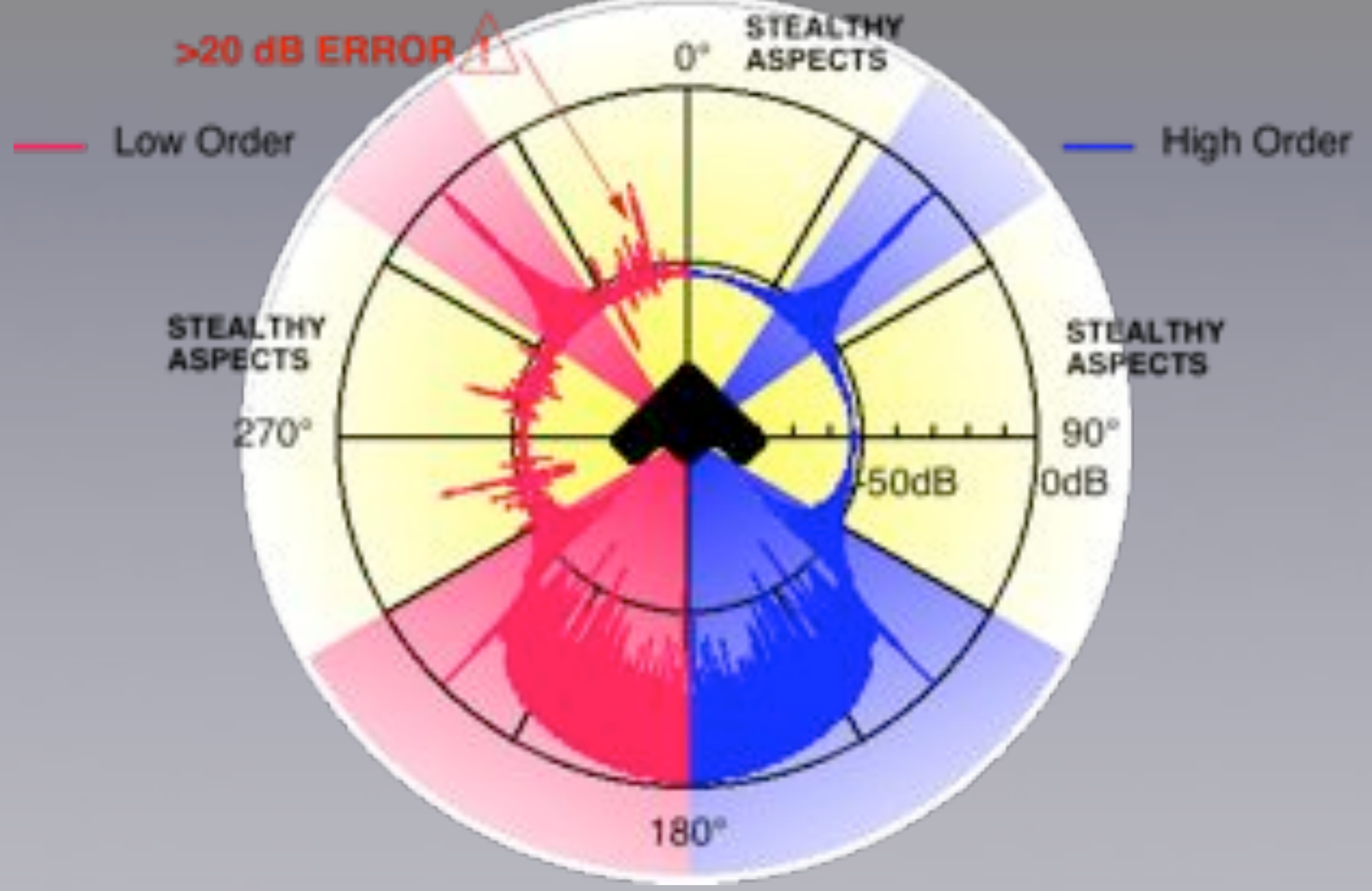
Rokhlin’s Dogma

Engineering radar cross section prediction:

1. Methods must be high order (Nystrom discretization)
2. Methods must be fast (Helmholtz FMM)
3. Implementations must be scalable

HIGH-ORDER DISCRETIZATION FOR STEALTHY TARGETS

"HRL BAT" [2d TARGET] L = 300λ 6000 UNKNOWNS



Source: Stephen Wandzura, Hughes

FastScat Results

Year	*1992*	*1999*
Code	Patch	FastScat
Computer	Touchstone Delta	SGI Origin 2000
Processors	512	64
Radius (wavelengths)	5.31	60
Area (sq. wavelengths)	354	45,239
Accuracy (db rms)	2 (est)	0.12
Unknowns	48,672	2,160,000
Memory (Gb)	38	45.5
Time (hours)	19.6	27.9
Cost/ Sq.Wavelength	$10,000 (approx)	$50

- 10x greater accuracy on much larger target
- Patch would require 2,197,000,000x time for same accuracy
- *Nine orders of magnitude improvement in seven years*

(See Ottusch, Stalzer, Visher, and Wandzura, Scalable electromagnetic scattering calculations on the SGI Origin 2000, *Proc. SC99,* Portland)

Are we done yet?
Verification & Validation

Verification

- Module tests
- Comparison to theory
- Solution convergence
- Time complexity checks

Validation

- Comparison to experiments for fixed parameters
- Uncertainty quantification

Bad science and policy can result from divergent un-validated codes

CSE as Systems Engineering

- Science is a closed loop process between theory and experiment
- Significant resources devoted to experiment
- Computers -
  - Enable prediction, operation and data collection
  - Disintermediate the process

Step back from computation or data centric views

Take a broader *systems* view...

Structural Complexity in CSE
Examples from CACR

- Weapons certification
- Biochemistry
- Real time astronomy
- Earth tomography
- High energy physics

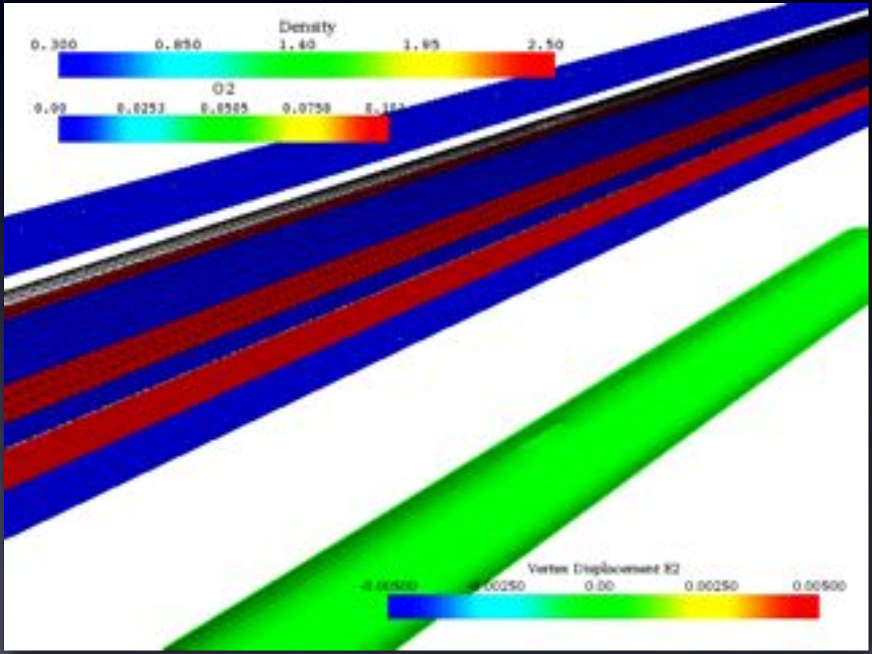
Sample CSE Systems Design Questions

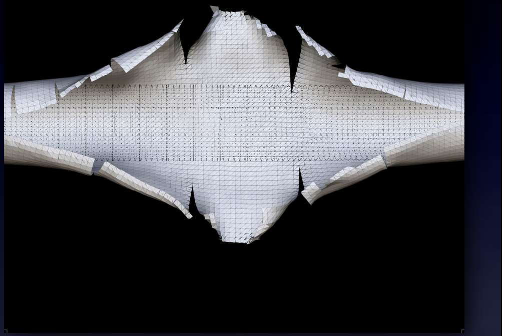
In a systems view, the questions affecting computer resources are different -

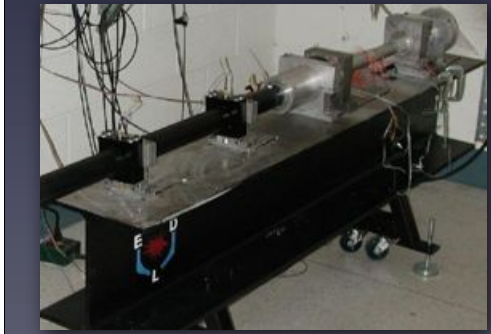
- How best to deploy resources (experiments, model development, computation) to certify a given design to 99.99%?
- What's the next best experiment?
- How do we (partially) automate discovery in astronomical data sets?
- Which earthquakes best resolve Earth's structure?
- How does experiment sensitivity change with computer resources?

Motivation for Systems View: Validated Simulations

Coupled fracture simulation with detailed chemistry



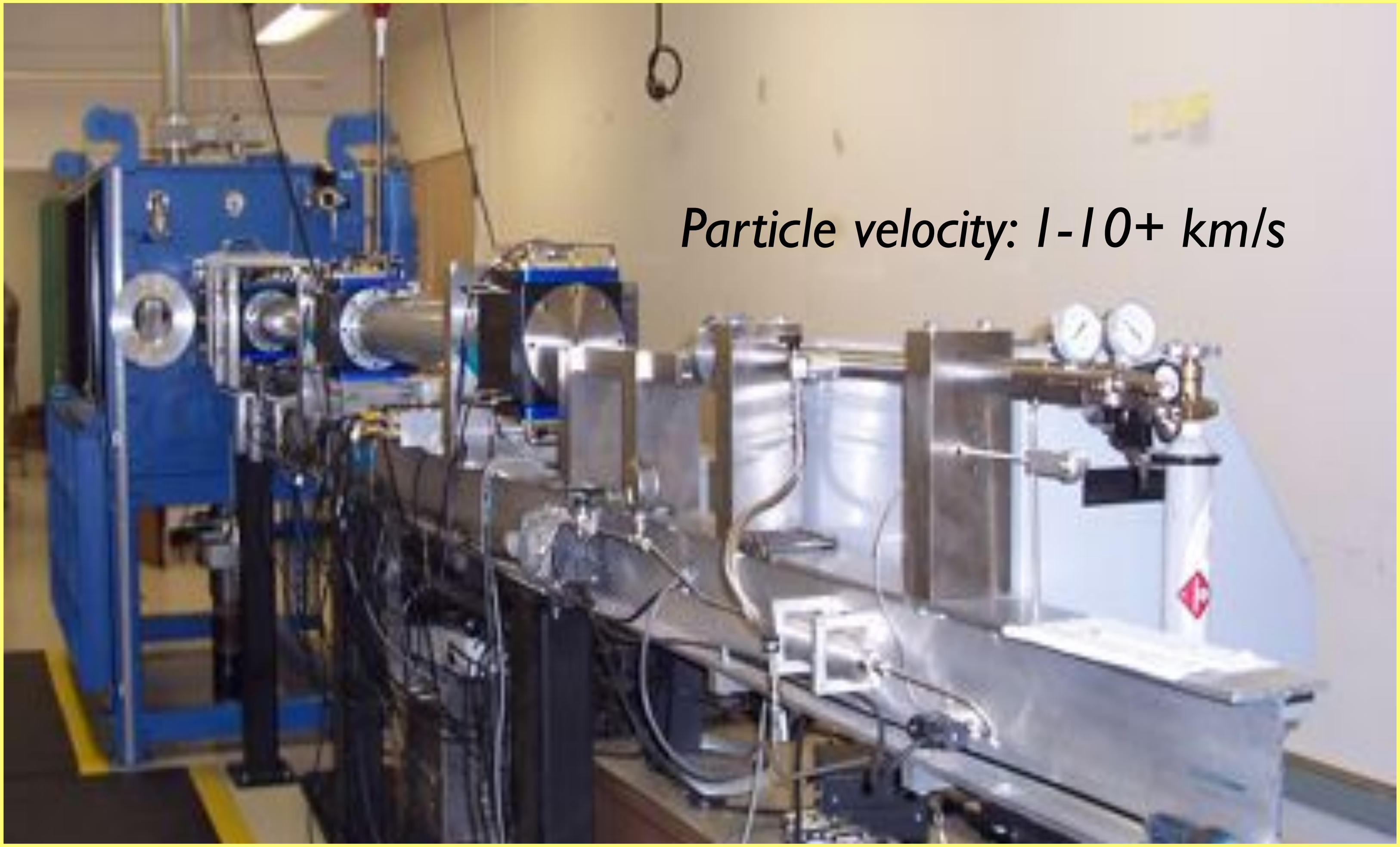




Essence of CSE: Fast feedback between models, simulation and experiment

Source: Dan Meiron and Joseph Shepherd, Caltech

Caltech-JPL Small Particle Hypervelocity Impact Range



Particle velocity: 1-10+ km/s

Source: Michael Ortiz and Ares Rosakis, Caltech

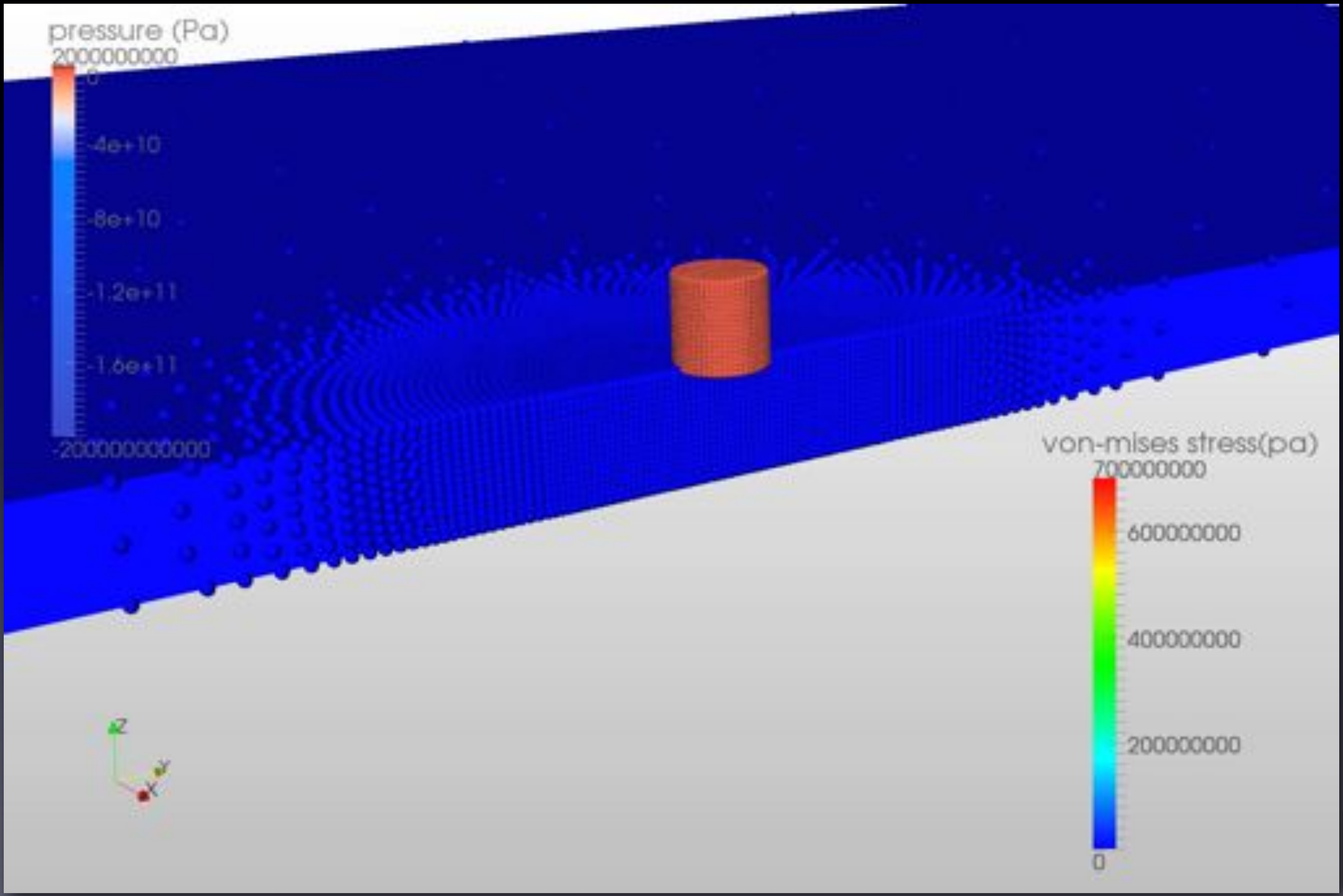
SPHR Firing

Photron FASTCAM SA1.1 mo...

72000 fps 1/72000 sec 384 x 176

Center frame : -31 -00:00:00.000431

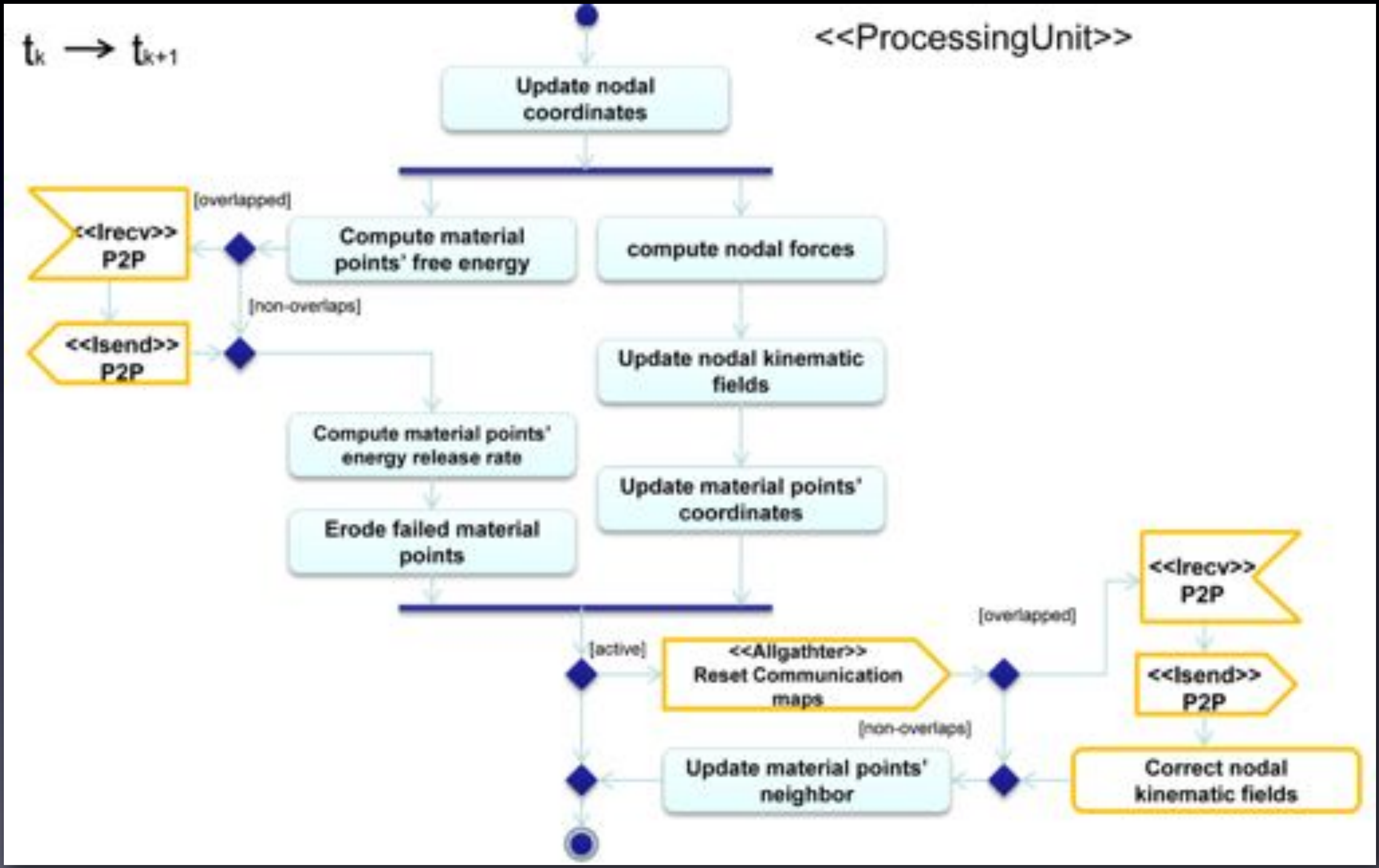
Parallel OTM Results



Source: Bo Li & Santiago Lombeyda, Caltech

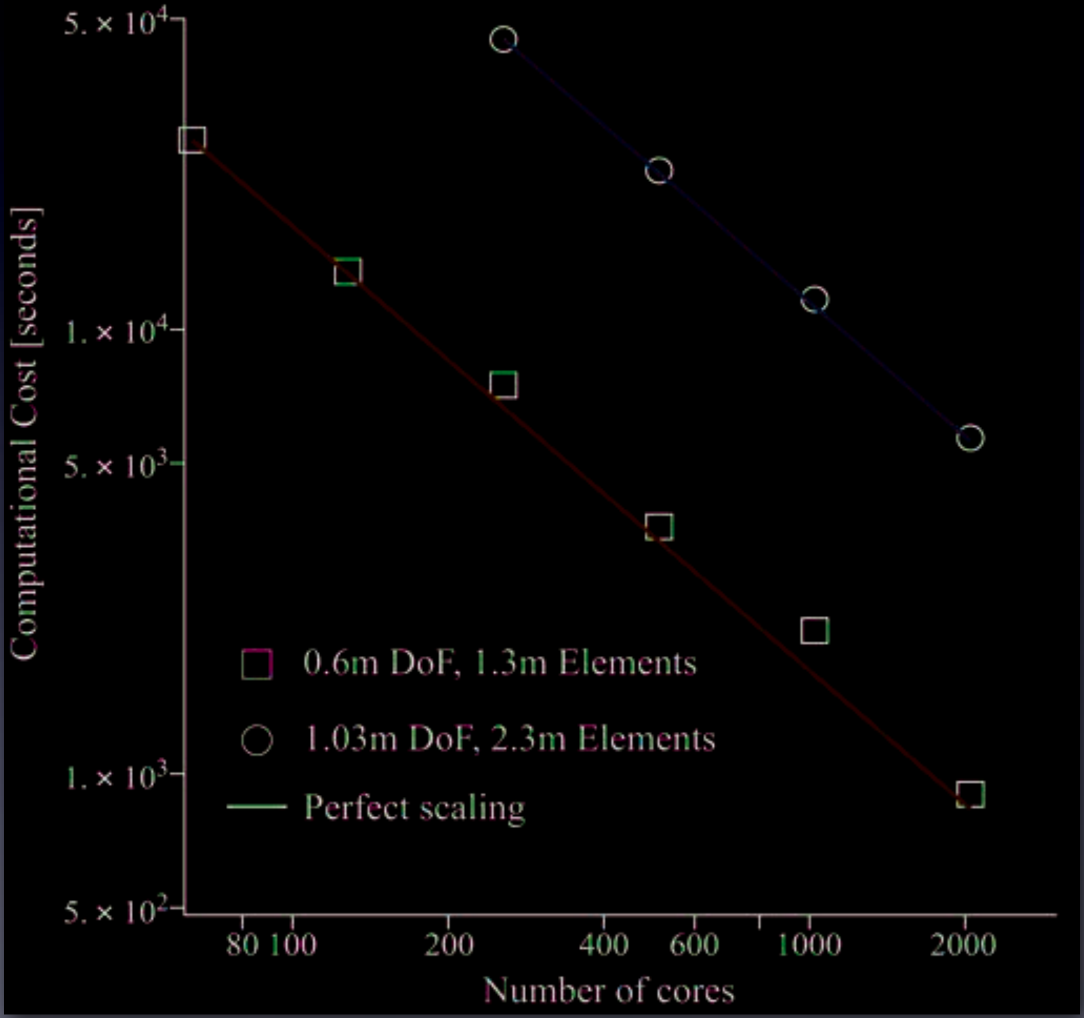
CACR

Parallel OTM Architecture



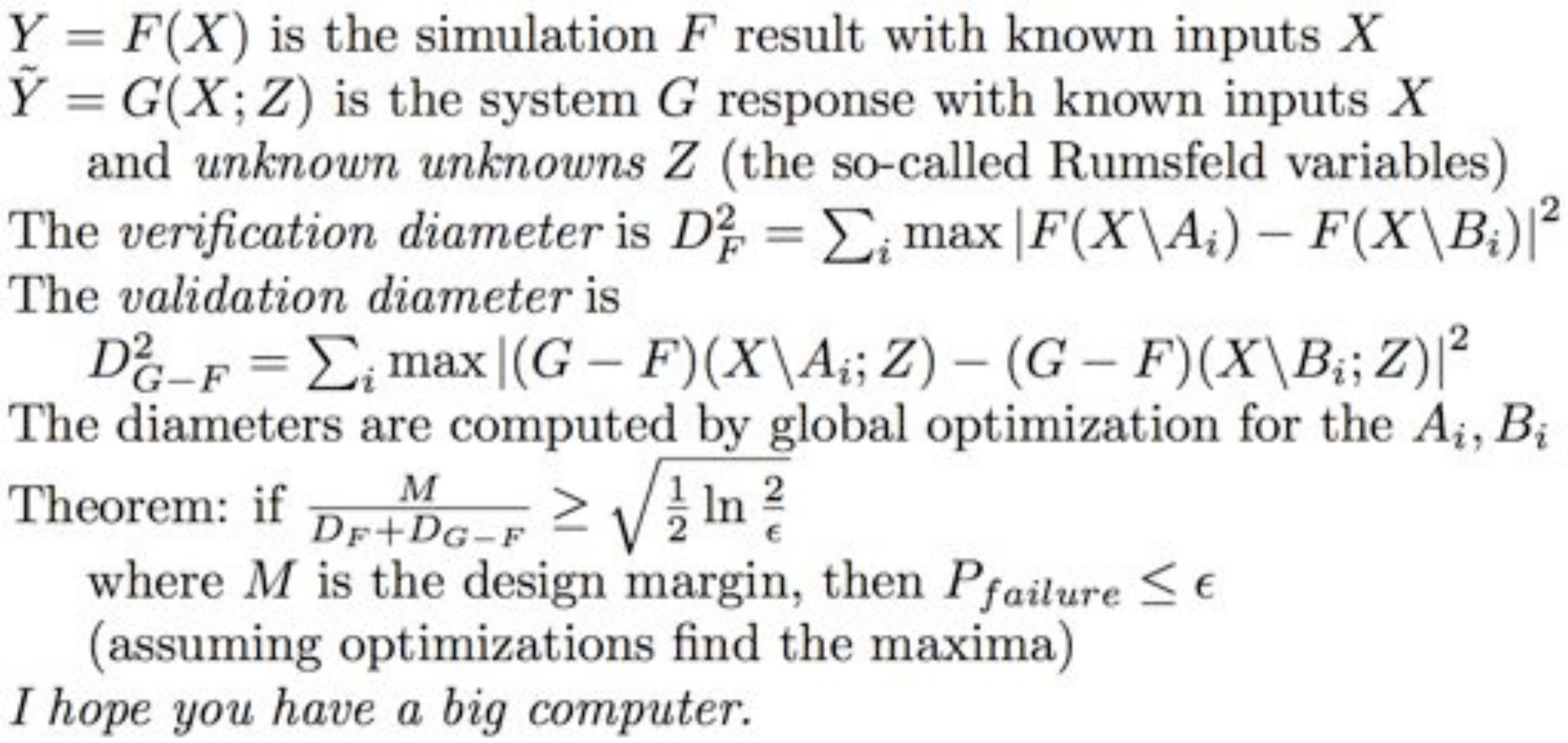
Source: Bo Li, Caltech

Parallel OTM Scaling



Source: Bo Li, Caltech

Uncertainty Quantification by Concentration of Measures



$Y = F(X)$ is the simulation F result with known inputs X

$\tilde{Y} = G(X;Z)$ is the system G response with known inputs X and *unknown unknowns* Z (the so-called Rumsfeld variables)

The *verification diameter* is $D_F^2 = \sum_i \max |F(X\backslash A_i) - F(X\backslash B_i)|^2$

The *validation diameter* is

$$D_{G-F}^2 = \sum_i \max |(G-F)(X\backslash A_i;Z) - (G-F)(X\backslash B_i;Z)|^2$$

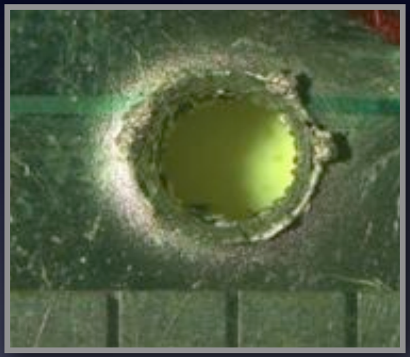
The diameters are computed by global optimization for the A_i, B_i

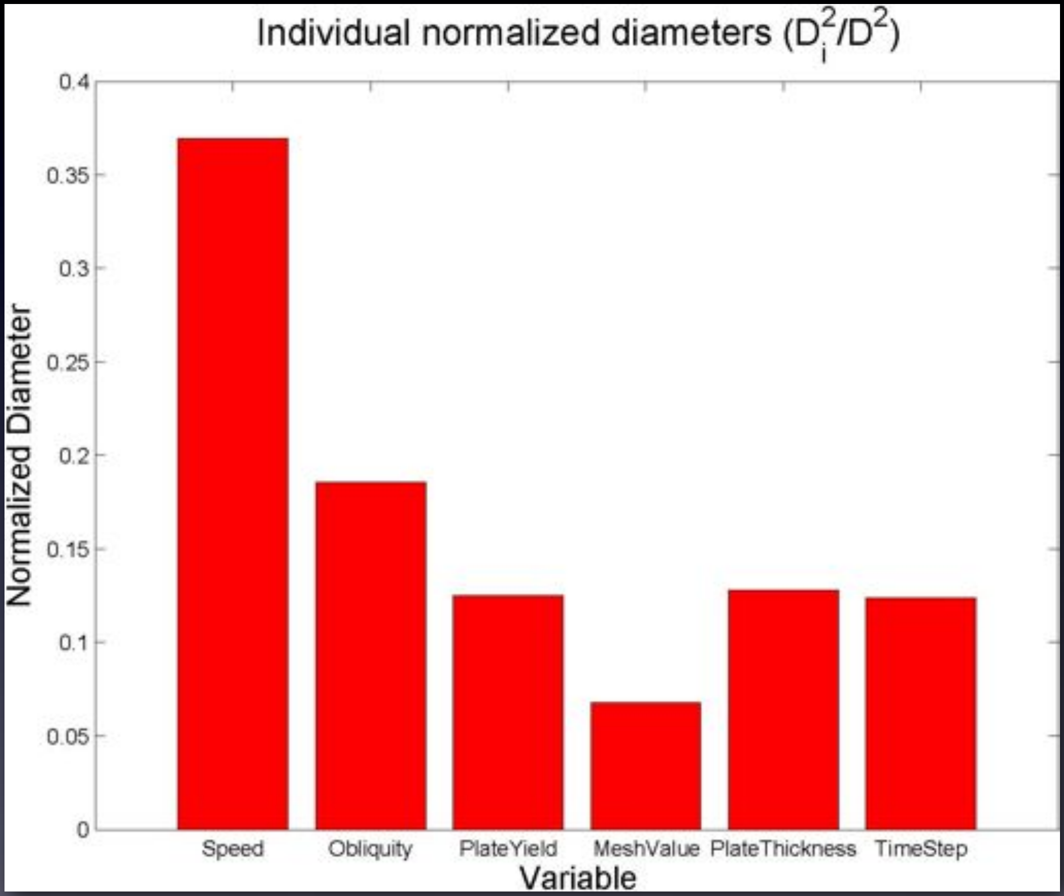
Theorem: if $\frac{M}{D_F + D_{G-F}} \geq \sqrt{\frac{1}{2}\ln\frac{2}{\epsilon}}$ where M is the design margin, then $P_{failure} \leq \epsilon$ (assuming optimizations find the maxima)

I hope you have a big computer.

(See Lucas, Owhadi, and Ortiz, Rigorous verification, validation, uncertainty quantification and certification through Concentration-of-Measure inequalities, *J. Comp. Methods in Applied Mechanics and Engineering,* 197:4591-4609, 2008)

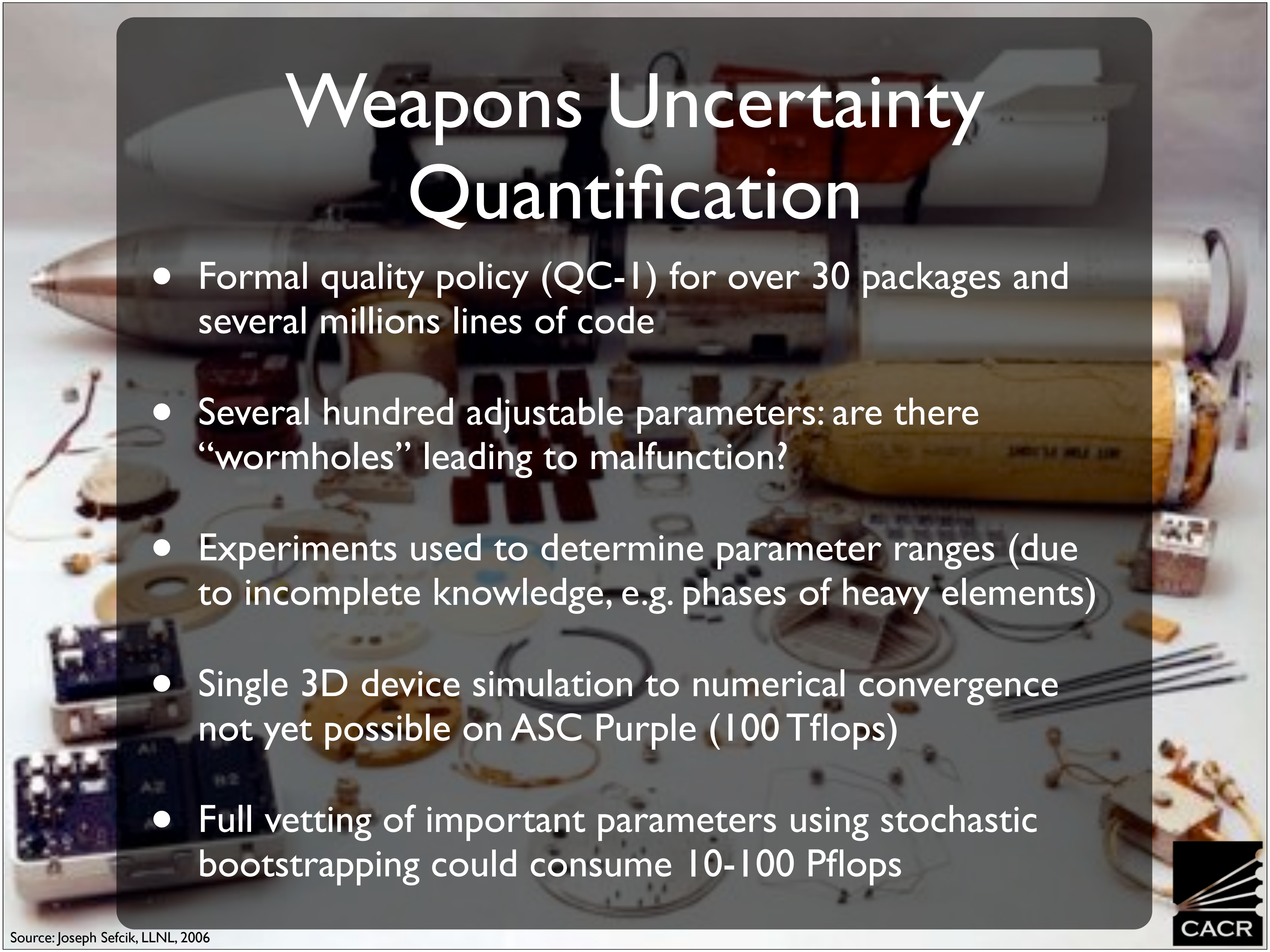
Verification Sub-diameters





Sensitivity to both experimental and modeling parameters.
It's an experimental-theoretical system.

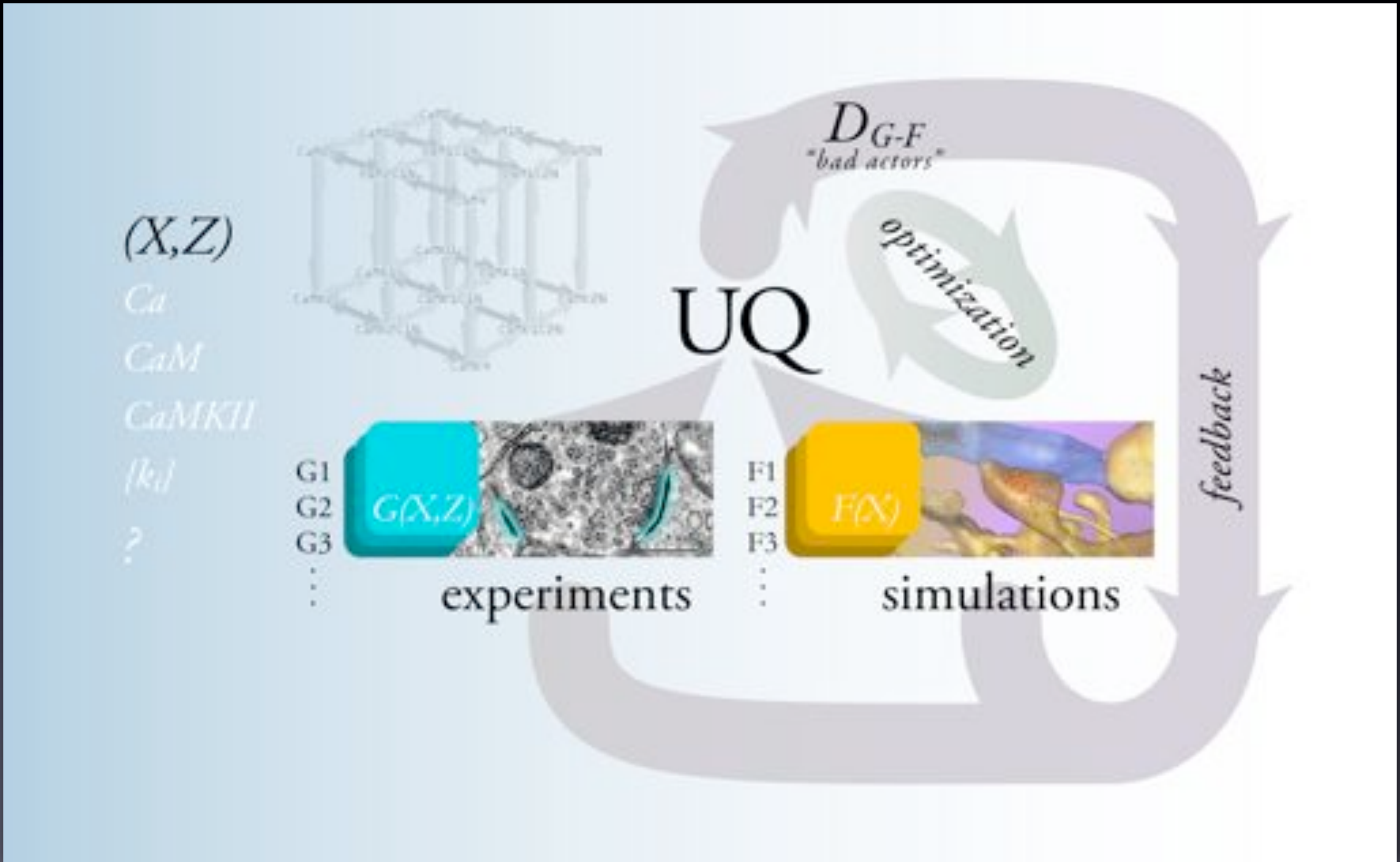
Weapons Uncertainty Quantification



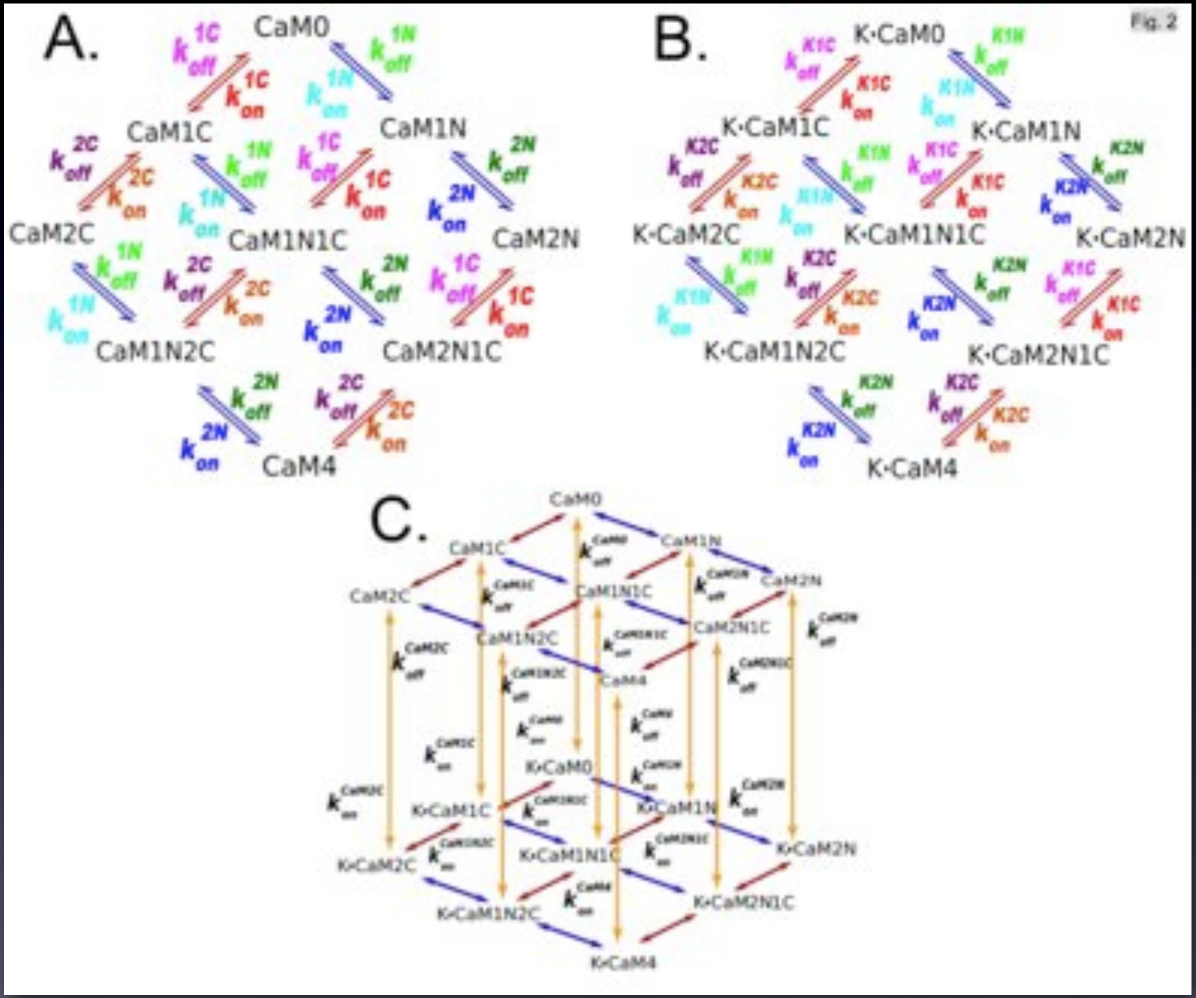
- Formal quality policy (QC-1) for over 30 packages and several millions lines of code
- Several hundred adjustable parameters: are there “wormholes” leading to malfunction?
- Experiments used to determine parameter ranges (due to incomplete knowledge, e.g. phases of heavy elements)
- Single 3D device simulation to numerical convergence not yet possible on ASC Purple (100 Tflops)
- Full vetting of important parameters using stochastic bootstrapping could consume 10-100 Pflops

Source: Joseph Sefcik, LLNL, 2006

Synaptic Uncertainty Quantification



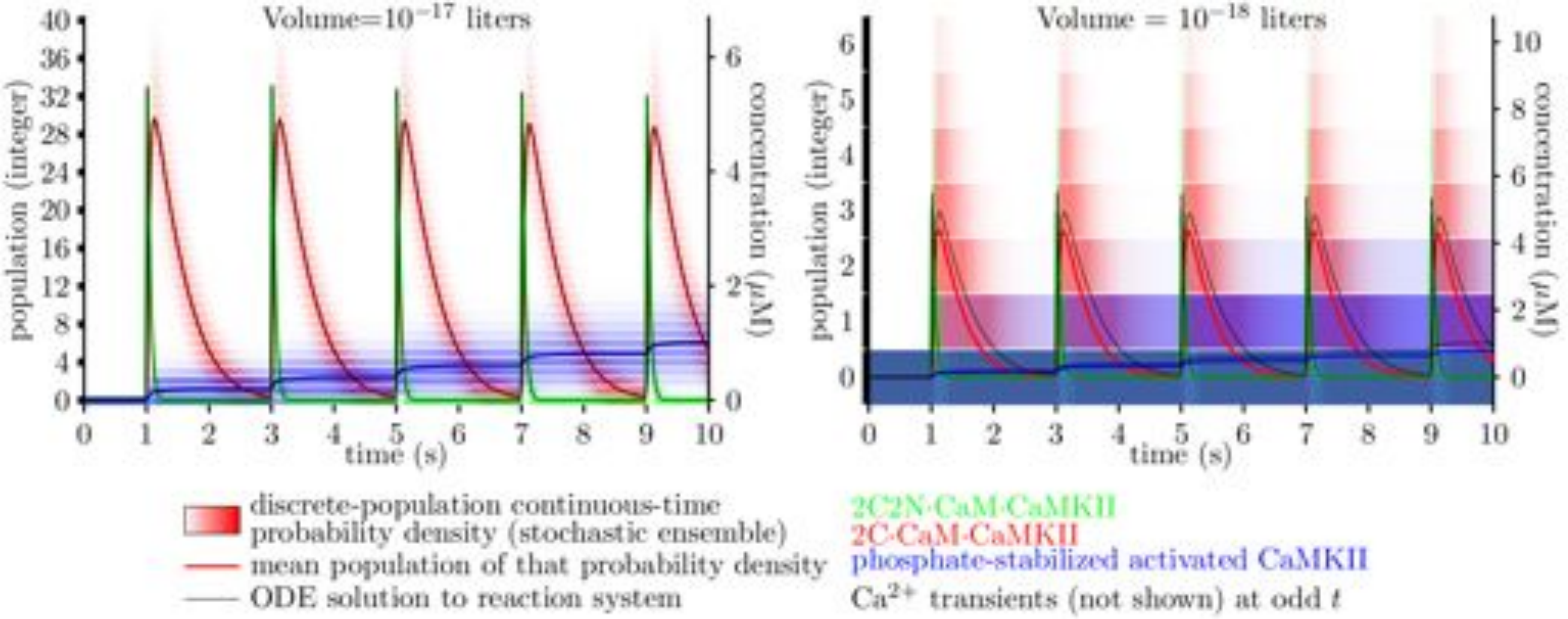
CaMKII & Subunits Model



Large SSA model: 146 species & 453 equations: Matches ODE for large concentrations.

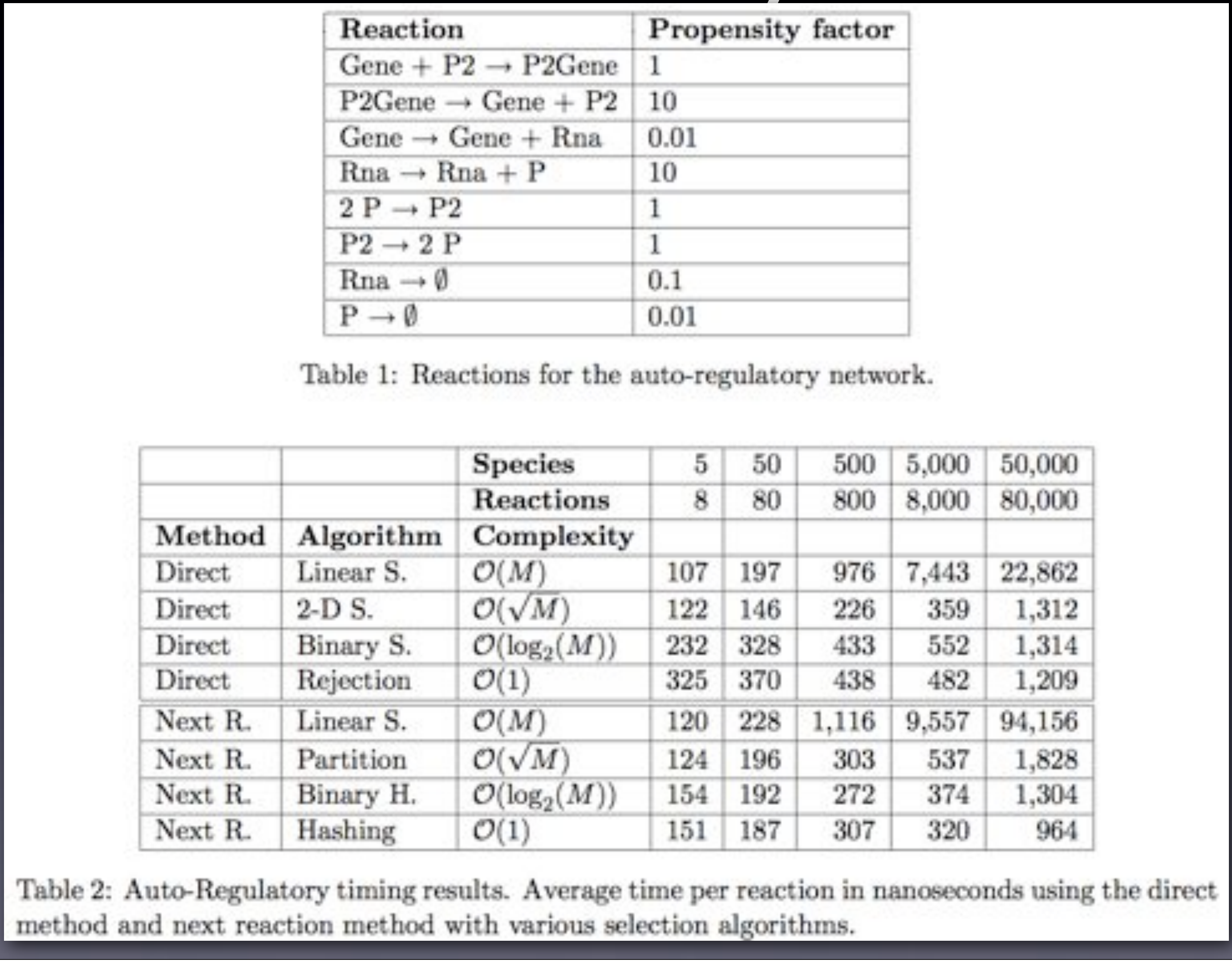
(Thanks to J. McCorquodale, S. Pepke, T. Kinzer-Ursem, S. Mihalas and M. Kennedy)

CaMKII Phosporylation: Effects of Small Concentrations



Statistical pressure from illegality of negative populations, and concentration spikes.

Stochastic Simulation for Biochemical Systems



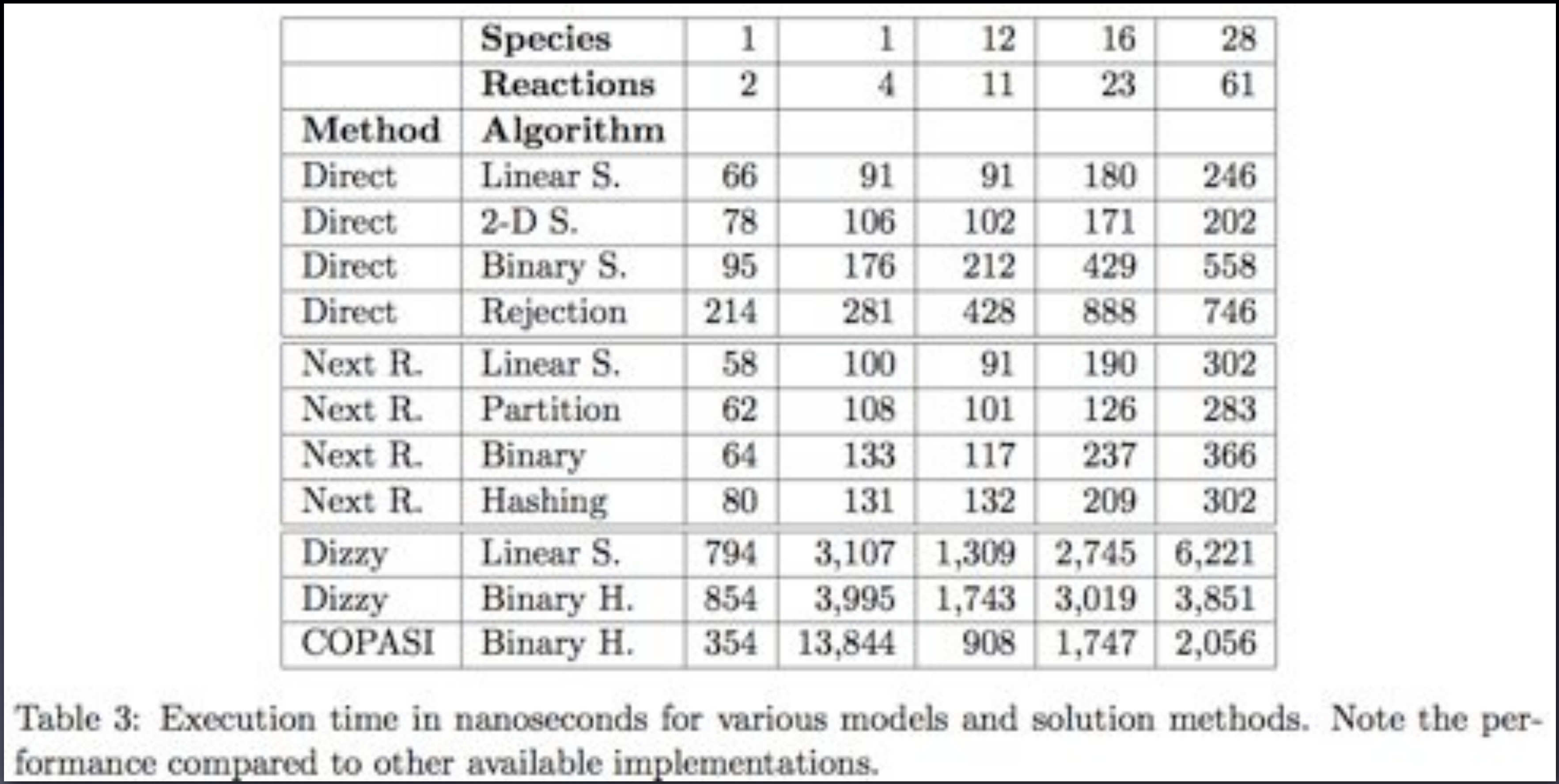
Reaction	Propensity factor
Gene + P2 → P2Gene	1
P2Gene → Gene + P2	10
Gene → Gene + Rna	0.01
Rna → Rna + P	10
2 P → P2	1
P2 → 2 P	1
Rna → ∅	0.1
P → ∅	0.01

Table 1: Reactions for the auto-regulatory network.

		Species	5	50	500	5,000	50,000
		Reactions	8	80	800	8,000	80,000
Method	**Algorithm**	**Complexity**					
Direct	Linear S.	$\mathcal{O}(M)$	107	197	976	7,443	22,862
Direct	2-D S.	$\mathcal{O}(\sqrt{M})$	122	146	226	359	1,312
Direct	Binary S.	$\mathcal{O}(\log_2(M))$	232	328	433	552	1,314
Direct	Rejection	$\mathcal{O}(1)$	325	370	438	482	1,209
Next R.	Linear S.	$\mathcal{O}(M)$	120	228	1,116	9,557	94,156
Next R.	Partition	$\mathcal{O}(\sqrt{M})$	124	196	303	537	1,828
Next R.	Binary H.	$\mathcal{O}(\log_2(M))$	154	192	272	374	1,304
Next R.	Hashing	$\mathcal{O}(1)$	151	187	307	320	964

Table 2: Auto-Regulatory timing results. Average time per reaction in nanoseconds using the direct method and next reaction method with various selection algorithms.

Use of Algorithms



	Species	1	1	12	16	28
	Reactions	2	4	11	23	61
Method	**Algorithm**					
Direct	Linear S.	66	91	91	180	246
Direct	2-D S.	78	106	102	171	202
Direct	Binary S.	95	176	212	429	558
Direct	Rejection	214	281	428	888	746
Next R.	Linear S.	58	100	91	190	302
Next R.	Partition	62	108	101	126	283
Next R.	Binary	64	133	117	237	366
Next R.	Hashing	80	131	132	209	302
Dizzy	Linear S.	794	3,107	1,309	2,745	6,221
Dizzy	Binary H.	854	3,995	1,743	3,019	3,851
COPASI	Binary H.	354	13,844	908	1,747	2,056

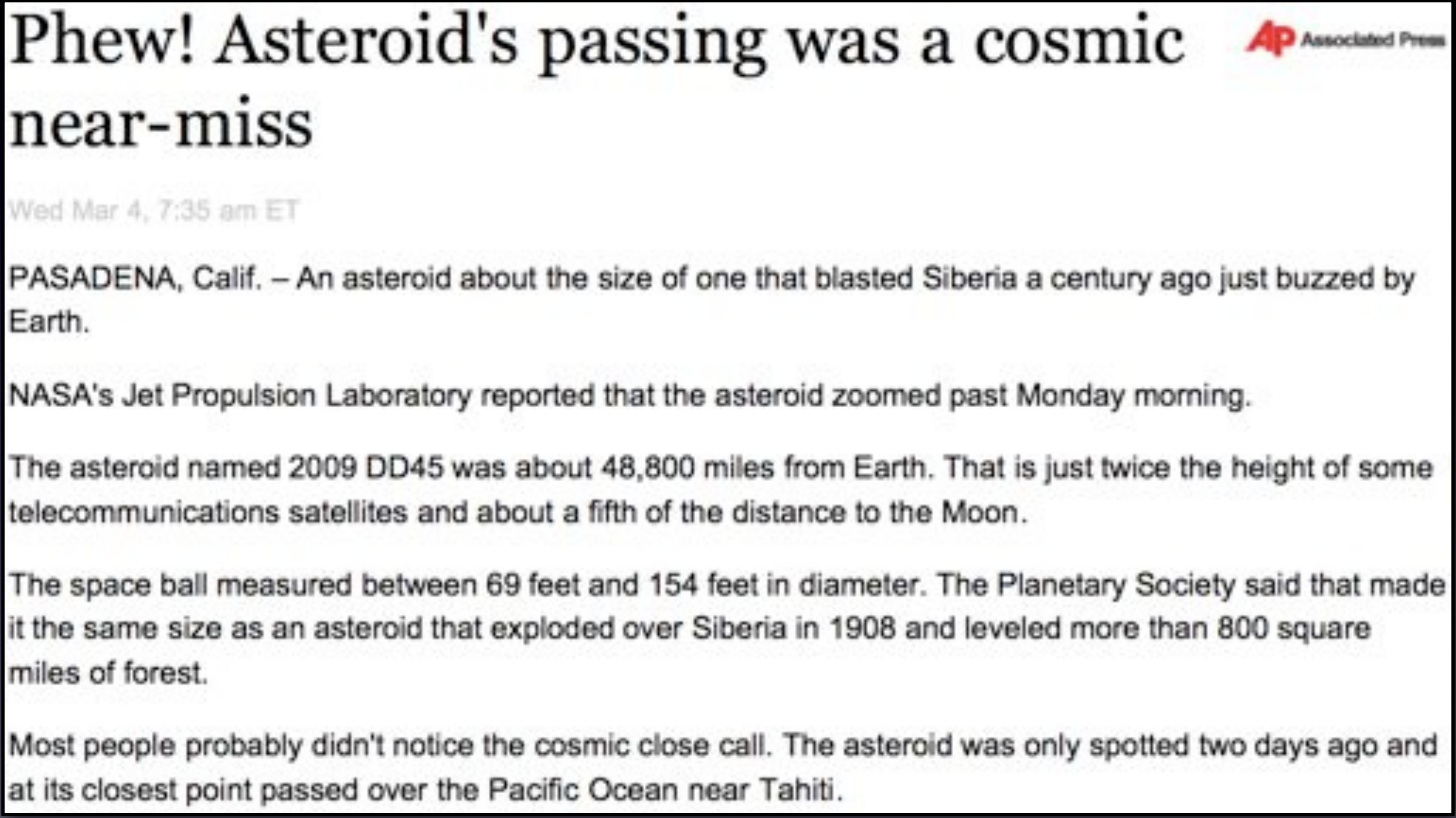
Table 3: Execution time in nanoseconds for various models and solution methods. Note the performance compared to other available implementations.

(See Mauch and Stalzer, Efficient formulations of exact stochastic simulation of chemical systems, *IEEE/ACM Trans. on Comp. Biology and Bioinformatics,* April 2009)

Much to Do

- Current simulator just an incomplete inner loop
- Spatially explicit stochastics?
- Multi-scale in time (e.g. slow scale SSA)?
- Potential impact of high performance computing (cloud)?
- Discretization of biological structures
- Uncertainty quantification

Real-Time Astronomy



Phew! Asteroid's passing was a cosmic near-miss

AP Associated Press

Wed Mar 4, 7:35 am ET

PASADENA, Calif. – An asteroid about the size of one that blasted Siberia a century ago just buzzed by Earth.

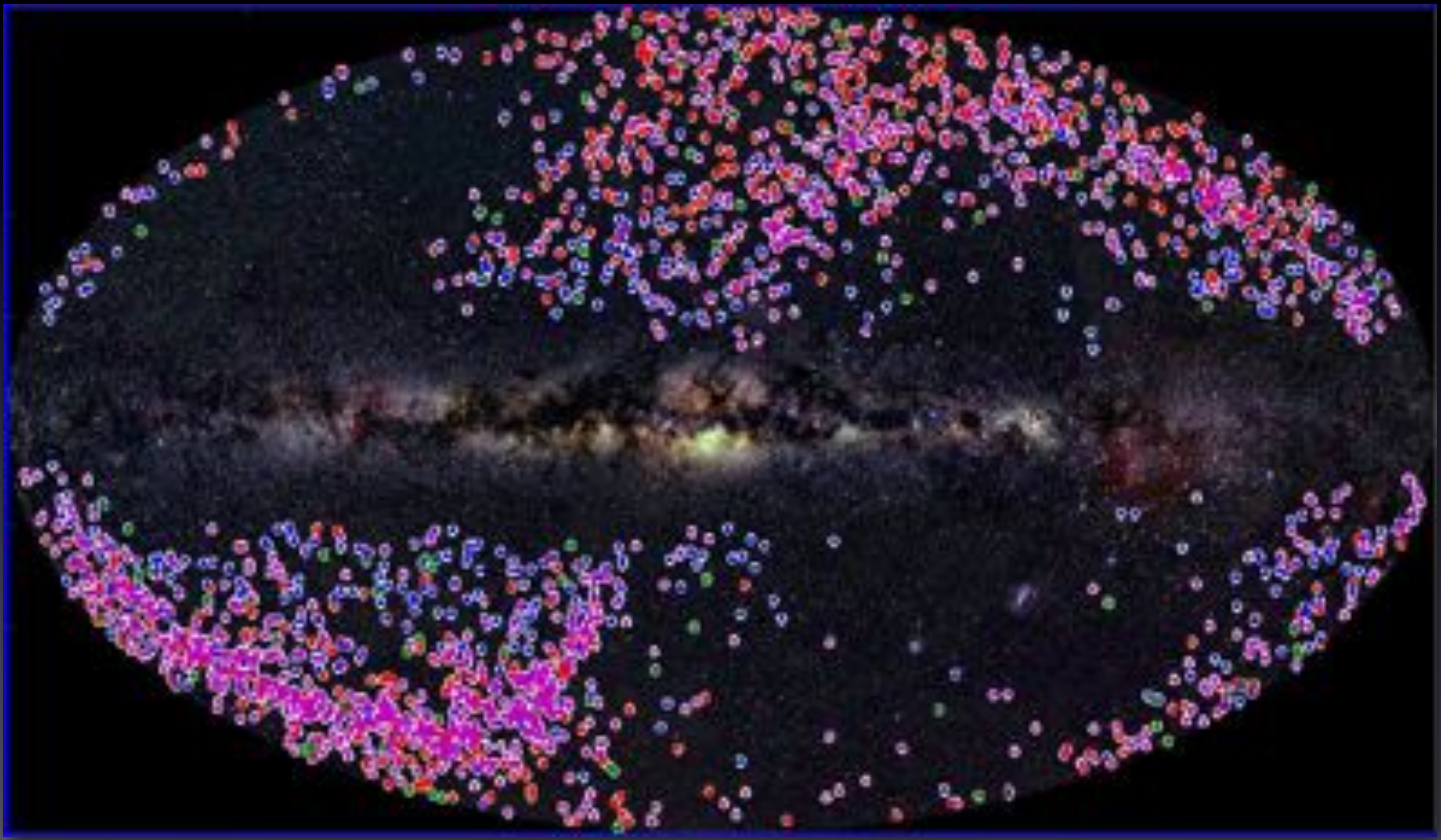
NASA's Jet Propulsion Laboratory reported that the asteroid zoomed past Monday morning.

The asteroid named 2009 DD45 was about 48,800 miles from Earth. That is just twice the height of some telecommunications satellites and about a fifth of the distance to the Moon.

The space ball measured between 69 feet and 154 feet in diameter. The Planetary Society said that made it the same size as an asteroid that exploded over Siberia in 1908 and leveled more than 800 square miles of forest.

Most people probably didn't notice the cosmic close call. The asteroid was only spotted two days ago and at its closest point passed over the Pacific Ocean near Tahiti.

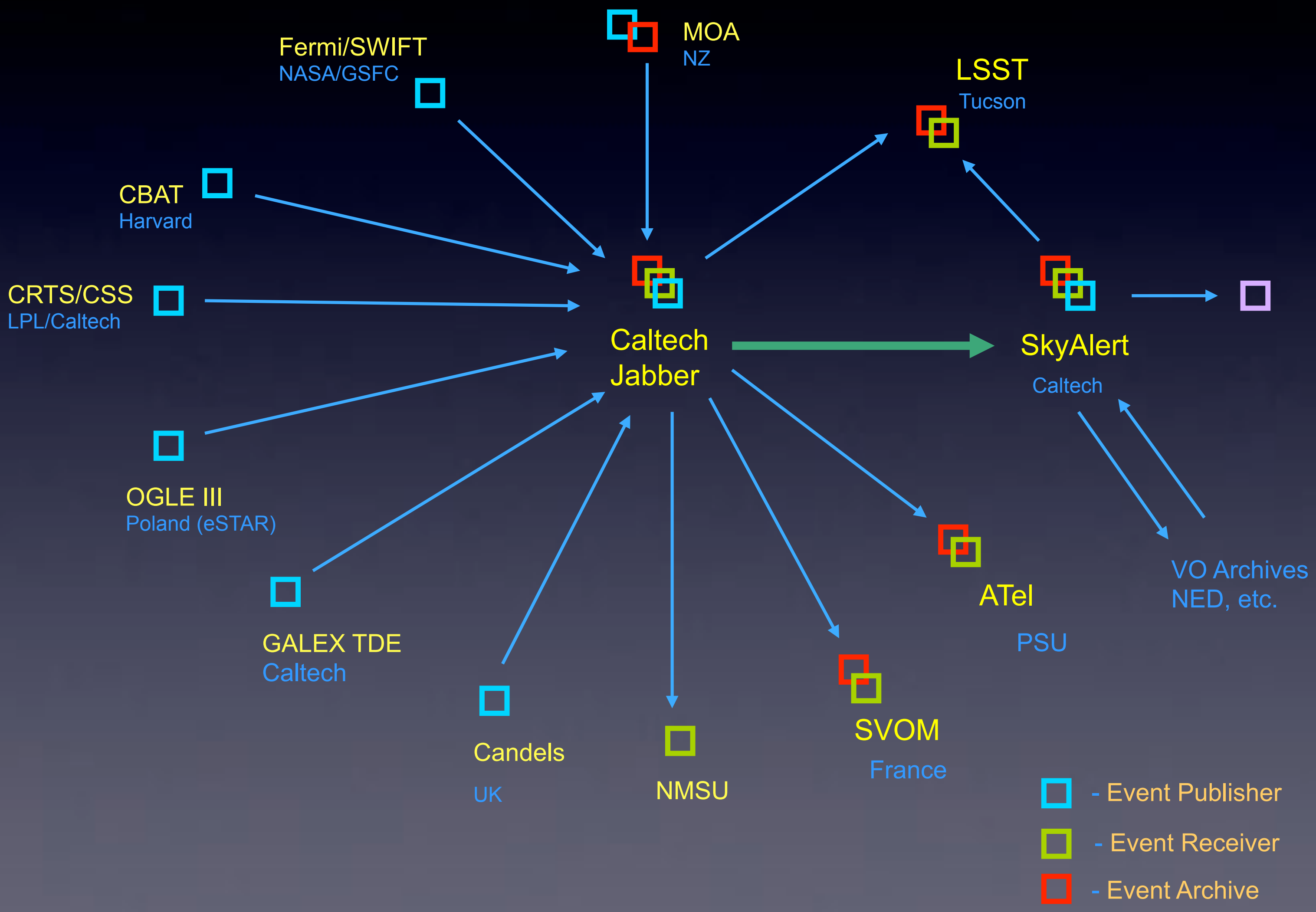
Catalina Real-Time Transient Survey: Discovery Locations



CRTS Catalina Real-Time Transient Survey

Source: Andrew Drake, Caltech

VOEvent Architecture



MOA
NZ

Fermi/SWIFT
NASA/GSFC

LSST
Tucson

CBAT
Harvard

CRTS/CSS
LPL/Caltech

Caltech
Jabber

SkyAlert
Caltech

OGLE III
Poland (eSTAR)

VO Archives
NED, etc.

ATel
PSU

GALEX TDE
Caltech

Candels
UK

NMSU

SVOM
France

🟦 - Event Publisher
🟩 - Event Receiver
🟥 - Event Archive

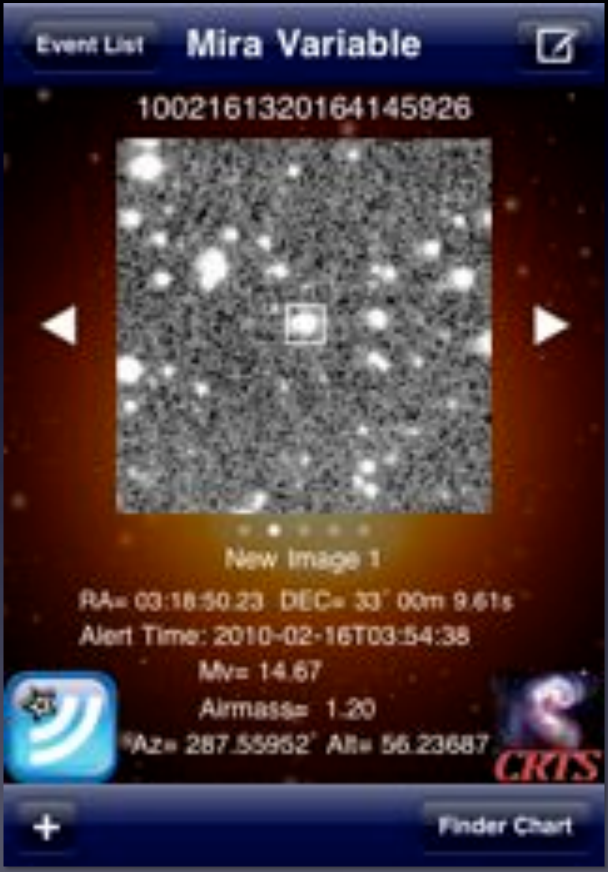
Event Distribution

1978 P 1

Smith, J. C.; Christy, J. W.; Graham, J. A.

IAU Circ., 3241, 1 (1978). Edited by Marsden, B. G.

Capt. J. C. Smith, U.S. Navy, reports: "Elongation of the photographic image of Pluto has been detected by J. W. Christy on plates taken with the U.S. Naval Observatory's 155-cm astrometric reflector on 1978 Apr. 13, 20 and May 12; 1970 June 13, 15, 16, 17 and 19; and 1965 Apr. 29 and May 1. The maximum elongation is ~0"9 in p.a. 170o-350o. Observed position angles are consistent with a uniform revolution/rotation period equal to the lightcurve period of 6.3867 days. The data suggest that there is a satellite, 2-3 magntiudes fainter than Pluto, revolving around Pluto in this period at a mean distance of ~20 000km; the implied sun-Pluto mass ratio is 140 000 000:1. ...



10021613201641459 26

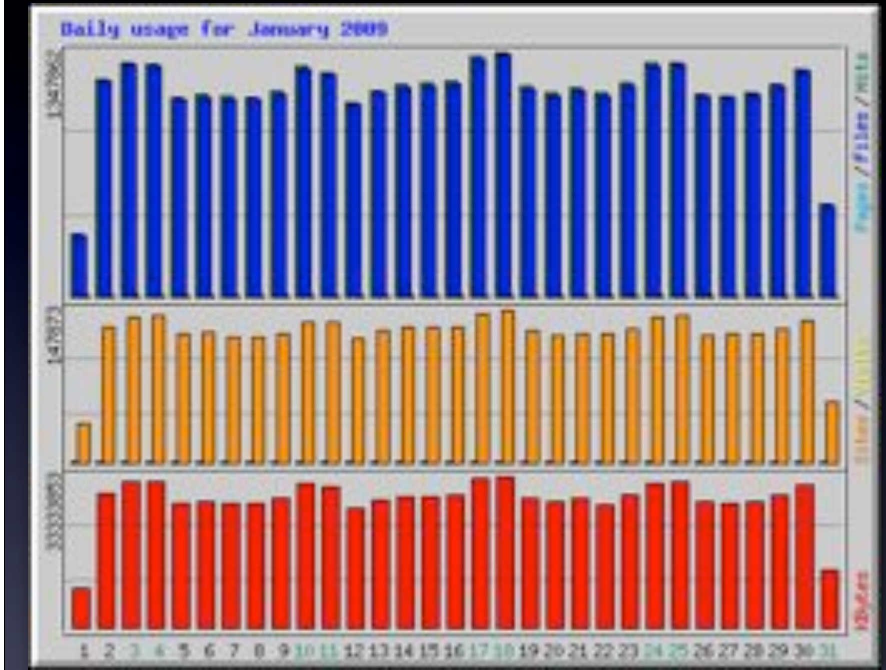
RA= 03:18:50.23 DEC= 33˚ 00m 9.61s
Alert Time: 2010-02-16T03:54:38
Mv= 14.67
Airmass= 1.20
Az= 287.55952˚ Alt= 56.23687

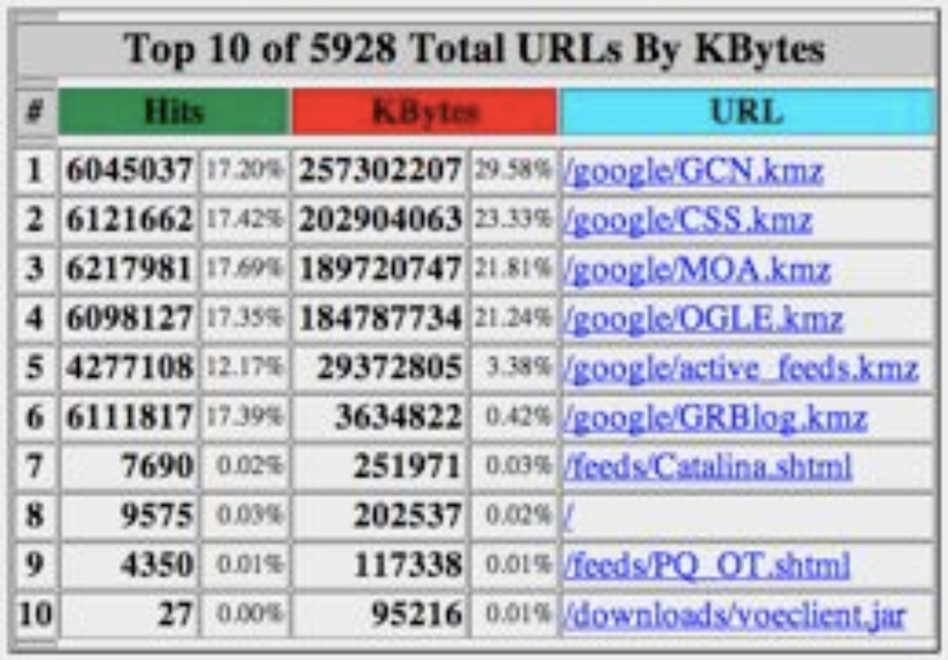
Three layers of event distribution:

- Real-time discovery (minutes)
  - VOEvents
- Human classified (mins to hours)
  - HTML, iPhone, Circulars (> 1200) twitter.com/skyalert
- Events of interest (mins to days)
  - ATel (> 40), CBAT (> 90), VSNET (> 100)

The glue is the architecture, APIs and social organization

Google Sky Usage



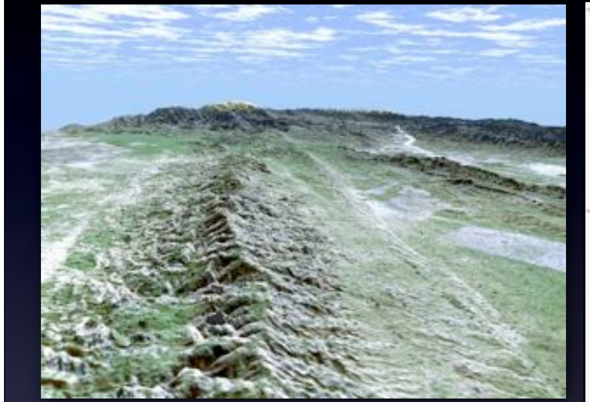


Top 10 of 5928 Total URLs By KBytes

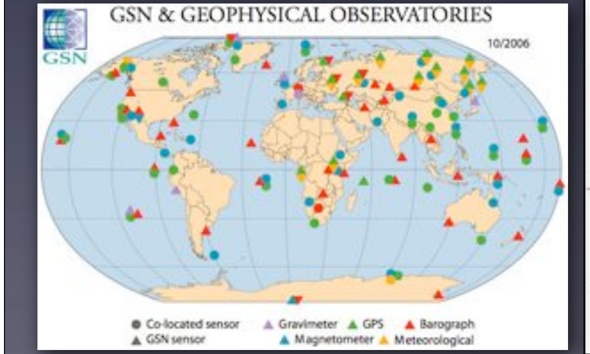
#	Hits		KBytes		URL
1	6045037	17.20%	257302207	29.58%	/google/GCN.kmz
2	6121662	17.42%	202904063	23.33%	/google/CSS.kmz
3	6217981	17.69%	189720747	21.81%	/google/MOA.kmz
4	6098127	17.35%	184787734	21.24%	/google/OGLE.kmz
5	4277108	12.17%	29372805	3.38%	/google/active_feeds.kmz
6	6111817	17.39%	3634822	0.42%	/google/GRBlog.kmz
7	7690	0.02%	251971	0.03%	/feeds/Catalina.shtml
8	9575	0.03%	202537	0.02%	/
9	4350	0.01%	117338	0.01%	/feeds/PQ_OT.shtml
10	27	0.00%	95216	0.01%	/downloads/voeclient.jar

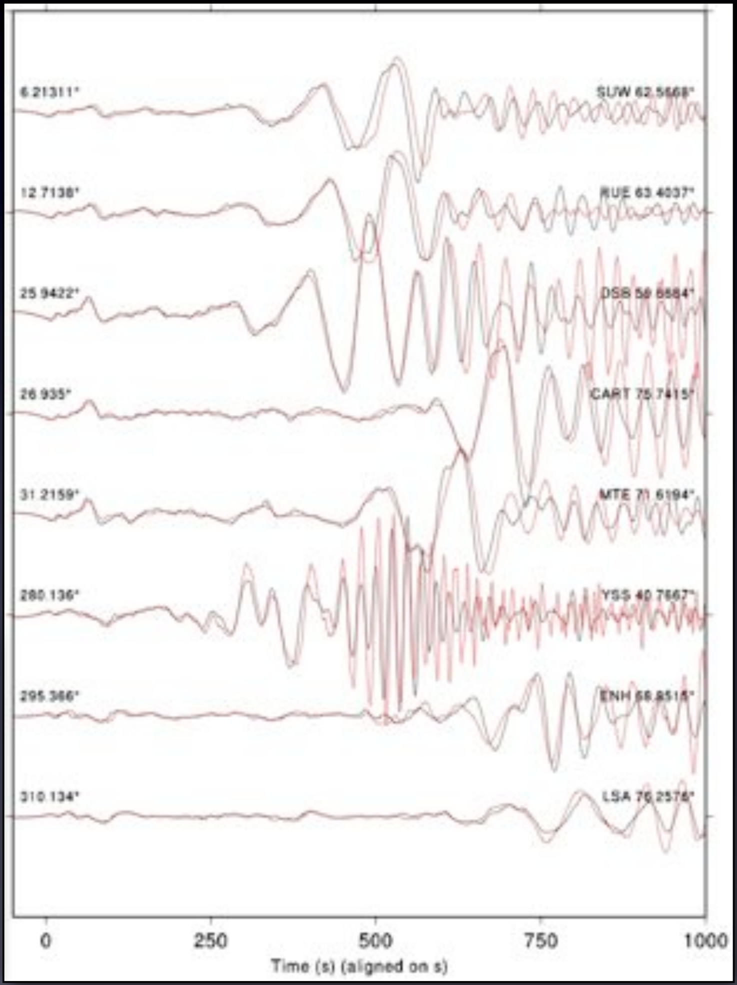
- Nearly 1 TB/month (827 GB) to support user interaction
- Data feeds (in this case) are insignificant
- We found out by over-run system logs!

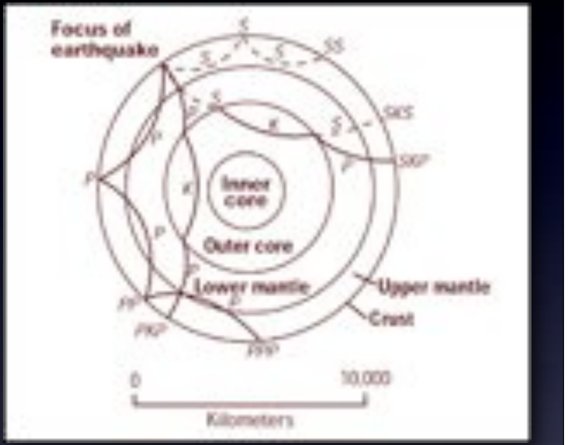
Earth Tomography



Earthquake
recorded globally
and stored.

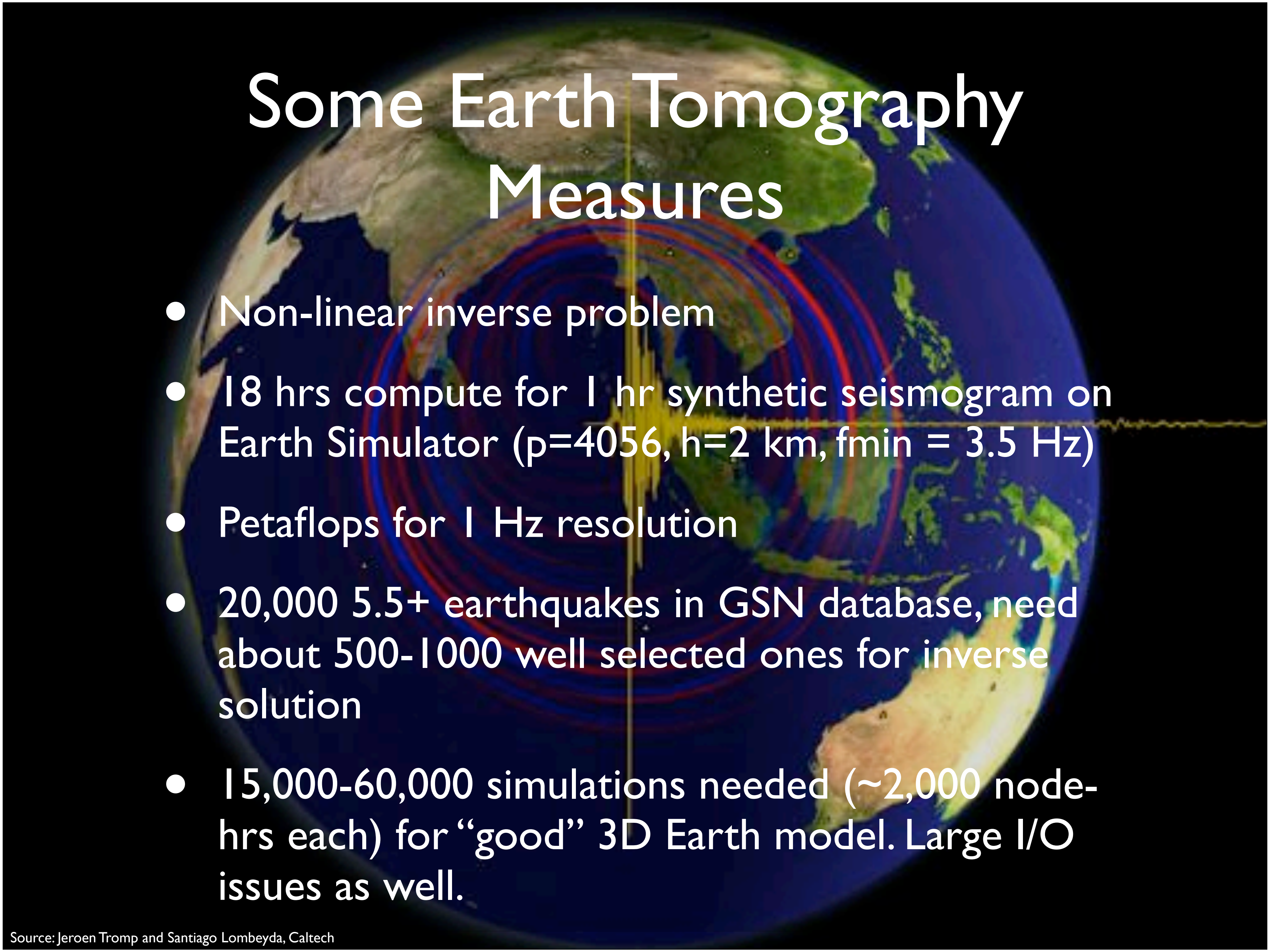






3D Earth model and
sources used to compute
synthetic seismograms.
These are compared with
recorded data and used to
refine the Earth model.

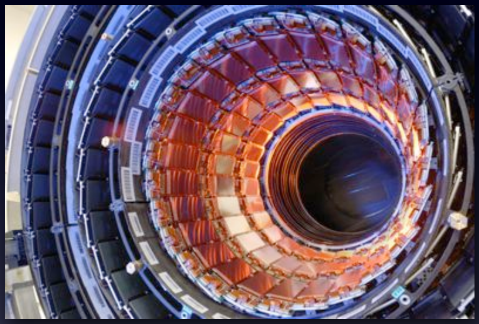
Some Earth Tomography Measures



- Non-linear inverse problem
- 18 hrs compute for 1 hr synthetic seismogram on Earth Simulator (p=4056, h=2 km, fmin = 3.5 Hz)
- Petaflops for 1 Hz resolution
- 20,000 5.5+ earthquakes in GSN database, need about 500-1000 well selected ones for inverse solution
- 15,000-60,000 simulations needed (~2,000 node-hrs each) for “good” 3D Earth model. Large I/O issues as well.

Source: Jeroen Tromp and Santiago Lombeyda, Caltech

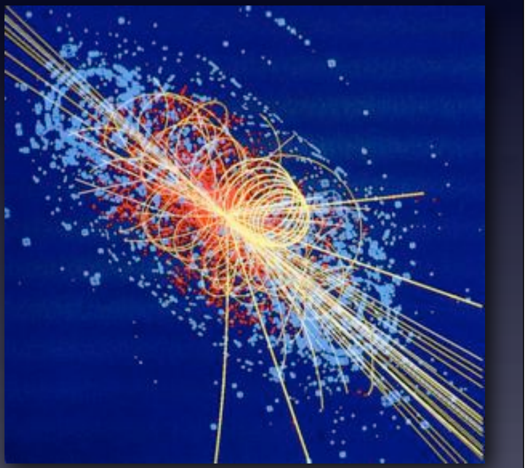
LHC/CMS Data Analysis

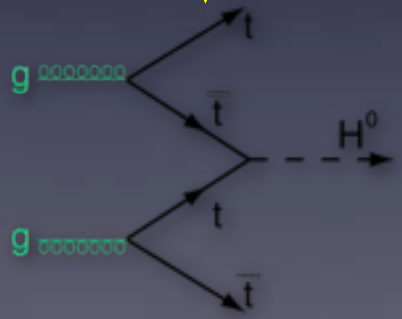


Tier0

Tier1_1
Tier1_2
.
.
.

Tier2_1
Tier2_2
Tier2_3
Tier2_4
.
.
.

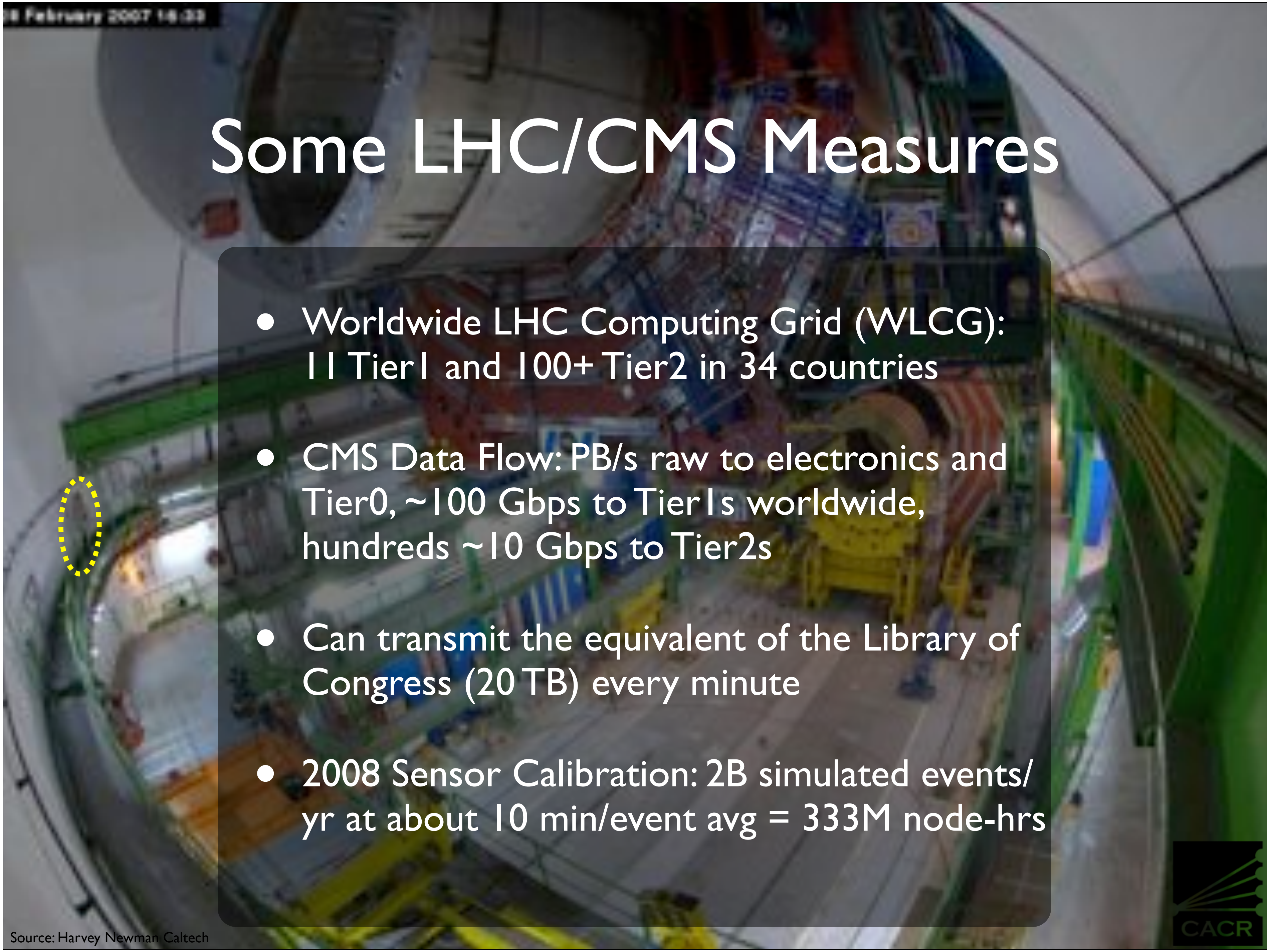




Virtual Detector
- -Design
- -Calibration
- -Transport
- -Classification

New physics?

Could UQ CoM help find Rumsfeld variables Z in the Standard Model?



Some LHC/CMS Measures

- Worldwide LHC Computing Grid (WLCG): 11 Tier1 and 100+ Tier2 in 34 countries
- CMS Data Flow: PB/s raw to electronics and Tier0, ~100 Gbps to Tier1s worldwide, hundreds ~10 Gbps to Tier2s
- Can transmit the equivalent of the Library of Congress (20 TB) every minute
- 2008 Sensor Calibration: 2B simulated events/yr at about 10 min/event avg = 333M node-hrs

Source: Harvey Newman Caltech

"Well, anything is possible. It's possible that an alien spaceship will materialize in your studio there and take over the microphone - but I'd rate the possibility of an LHC black hole engulfing the Earth as slightly less probable than that."

- Dr. Julian Bunn, CACR Principal Computational Scientist

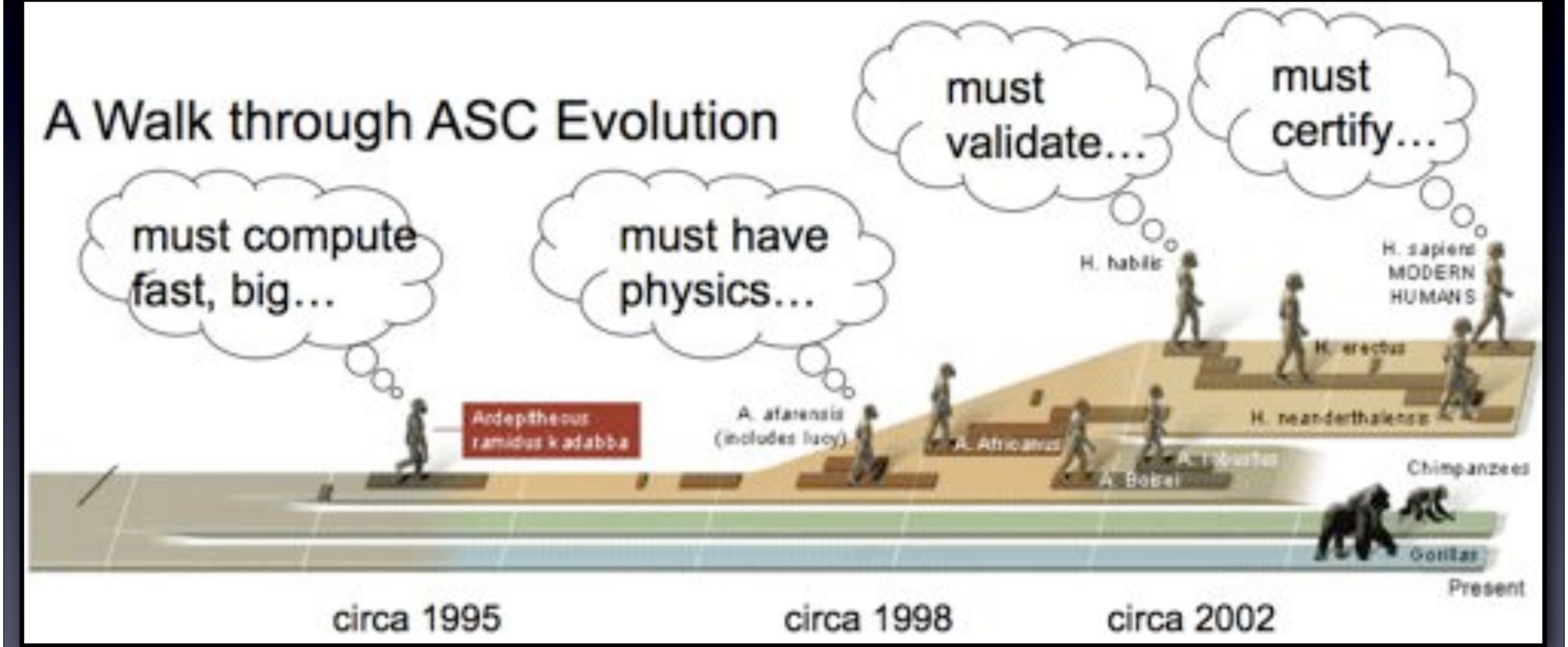
Currently running at 7 TeV with lead ions studying quark-gluon plasma.
We're still here...

Source: KPCC Public Radio Sept. 10 2008; CERN PR20.10 Nov. 4, 2010

Some Implications

- This was all planned...
- Commonalities
- Making computing costs explicit
- Partial question answers
- Final thoughts
- CSE systems engineering philosophy

This was all planned...



Source: Michael Ortiz, Caltech

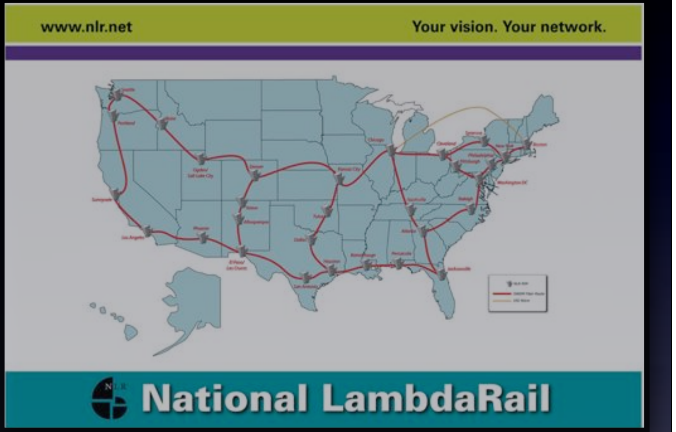
Commonalities

- Science driven; engineering enabled
- Complex social (virtual) organizations
- ***Integrated*** experiments & computation - near real-time
- Large simulations as "inner loops" & sensor calibrators
    - Importance of algorithms...
- Massive data sets due to sensors that are large and complex, or cheap and numerous, that require transportation and storage

It used to be that people built the best instruments and then figured out the results; now people build the best sensor architectures and then hope that computers figure it out.

Making Costs Explicit (Amazon EC2)

+



Instances

United States Europe

Standard Instances	**Linux/UNIX**	**Windows**
Small (Default)	$0.10 per hour	$0.125 per hour
Large	$0.40 per hour	$0.50 per hour
Extra Large	$0.80 per hour	$1.00 per hour
High CPU Instances	**Linux/UNIX**	**Windows**
Medium	$0.20 per hour	$0.30 per hour
Extra Large	$0.80 per hour	$1.20 per hour

Now $0.085

Sources: New York Times & National LambdaRail & Amazon

CACR

Applications Costs for EC2

FastScat sphere: $200 (assuming adequate MPI performance)

Certification: $5.4B (3 months at 100Pflops; weapons labs ~$6B/yr)

Single reaction: $2.7x10^{-11}$ (1 ms/r; how much to simulate a human?)

New Earth model: $7.5M

LHC/CMS calibration for 2008: $33M (LHC & detectors ~$5B)

Not having to run a supercomputing center: *priceless*

Explicit costs allow design trades

(See Walker, The real cost of a CPU hour, *Computer*, 35-41, April 2009. Also, Strand, KEYWORD: EVIL: Google's addiction to cheap electricity, *Harpers Magazine*, March 2008)

Partial Question Answers

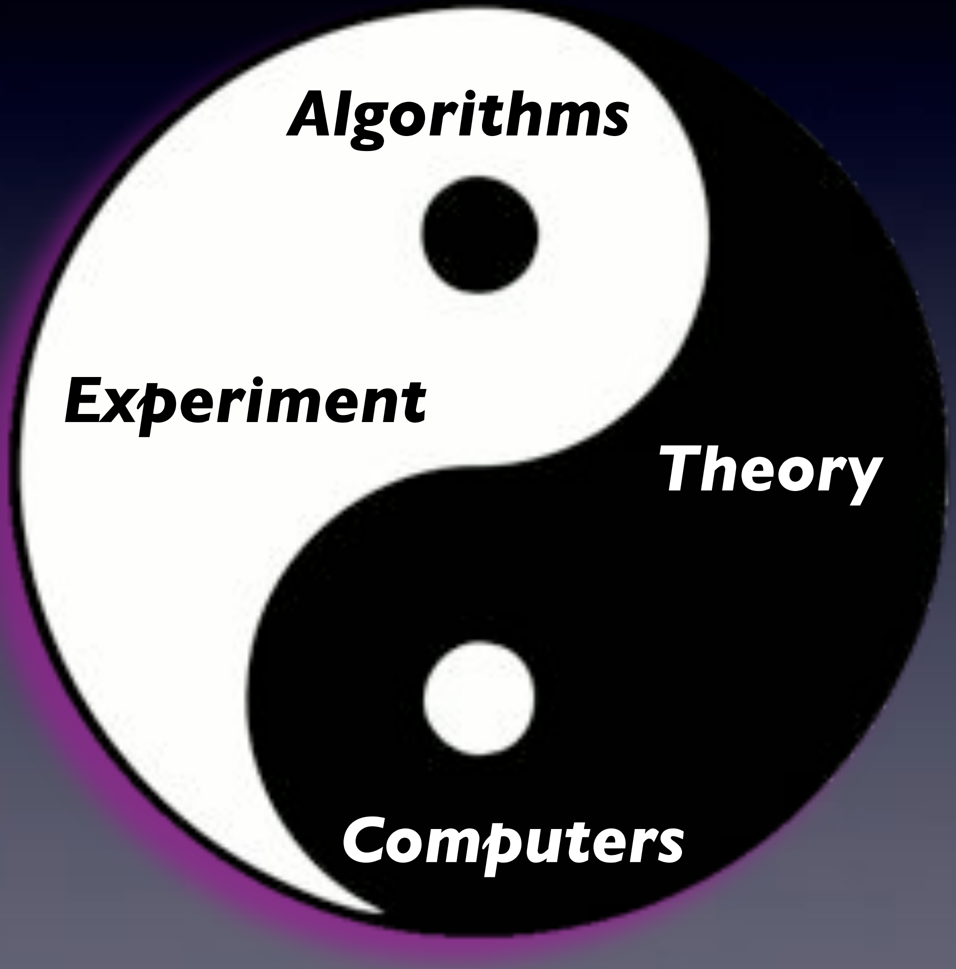
- How best to deploy resources to certify a given design to 99.99%?
  - Have theorem, need to compute $D_{G\text{-}F}$ & perform experiments to focus on “bad” sub-diameters (but not allowed to do full testing). Substantial legacy data. Need to trade resources devoted to F and G.
- What’s the next best experiment?
  - Same approach (biologists are very clever experimentalists; observability).
- How do we (partially) automate discovery in astronomical data sets?
  - Services architecture and buy-in from international community.
- Which earthquakes best resolve Earth’s structure?
  - Inverse problem, enabled by global sensor network & efficient forward code (SPECFEM3D). Lots of structure to exploit.
- How does experiment sensitivity change with computer resources?
  - If standard model is correct, no worse than 1/sqrt(N). What if it’s incomplete?

Final Thoughts

- CSE as Systems Engineering
  - Holistic and interdisciplinary (sociology challenging)
  - Design trades by potential science & engineering outcomes
  - Convergence as bounded unpredictability (complexity)
- Other applications?
  - Genome informatics, climate modeling, ...
  - What can we learn from commercial systems (e.g. search)?
- Technology pull by 100x to discover new phenomena -
  - Curse of dimensionality: *10-100 Pflops* (UQ), Distributed systems *10-100 Tbps* (CMS), Inverse problems (ET)
  - Large simulations as "inner loops" (buy more algorithms)

CACR

CSE Systems Engineering Philosophy



Web Links

CACR - www.cacr.caltech.edu

PSAAP - www.psaap.caltech.edu; OUQ arXiv:1009.0679v1

Stochastic simulation - cain.sourceforge.net

VAO & Catalina - www.usvao.org; crts.caltech.edu

Earth tomography - www.seismo.caltech.edu

LHC/CMS - www.uscms.org

Amazon EC2 - aws.amazon.com/ec2

This talk - www.cacr.caltech.edu/director