Supercomputing and its Applications in Biomedical Engineering

Yuefan Deng
Yuefan.Deng@StonyBrook.edu
Applied Mathematics Department
Stony Brook University

August 19, 2011

STONY BROOK
STATE UNIVERSITY OF NEW YORK

World Market of 5 Computers


- Remark attributed to Thomas John Watson(Chairman of the Board of IBM), 1943
  - *“I think there is a world market for about five computers.”*
- IBM responded with two sentences, 2008
  - *“Well, Mr. Watson is wrong.”*
  - *“The world will need only one computer.”*

Moore’s Law: Speed = (T/1.5)^2

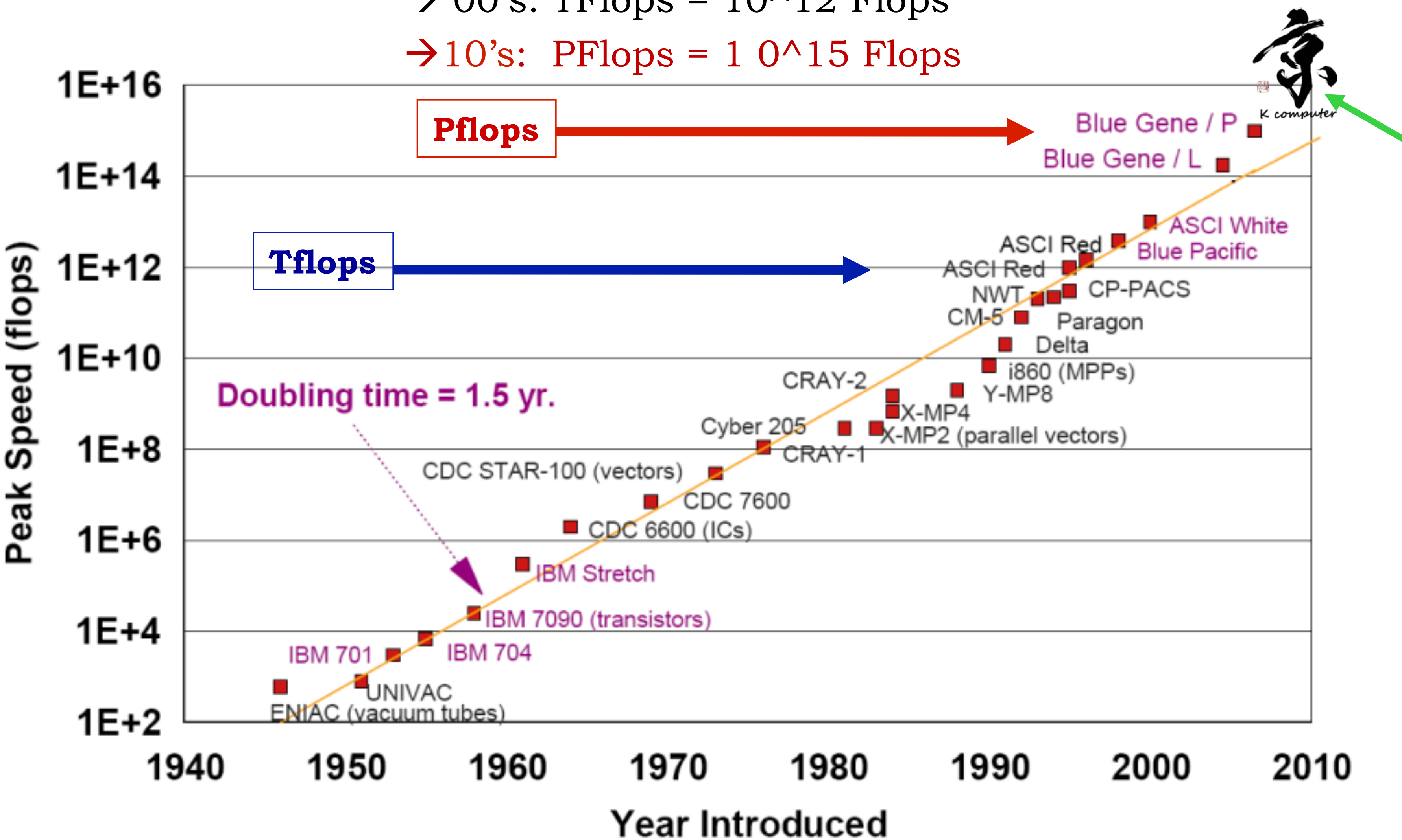
→ 40’s: 10^2 Flops

→ 70’s: MFlops = 10^6 Flops

→ 80’s: GFlops = 10^9 Flops

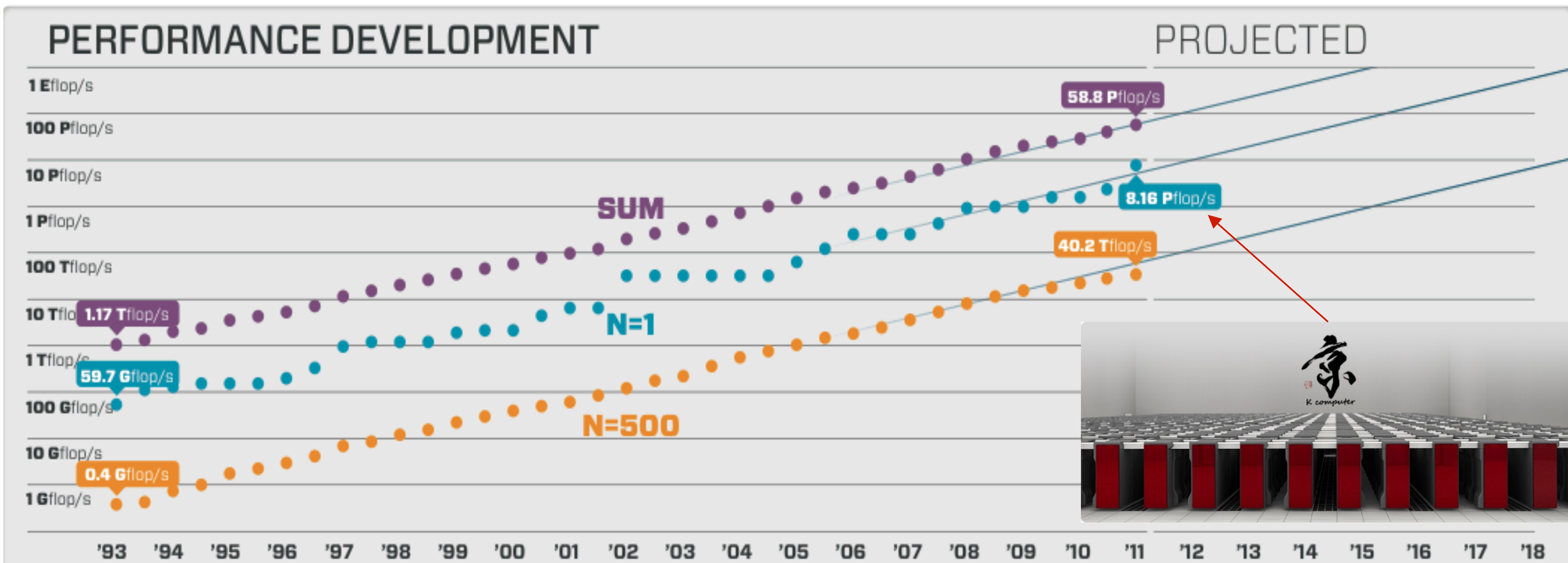
→ 00’s: TFlops = 10^12 Flops

→10’s: PFlops = 1 0^15 Flops



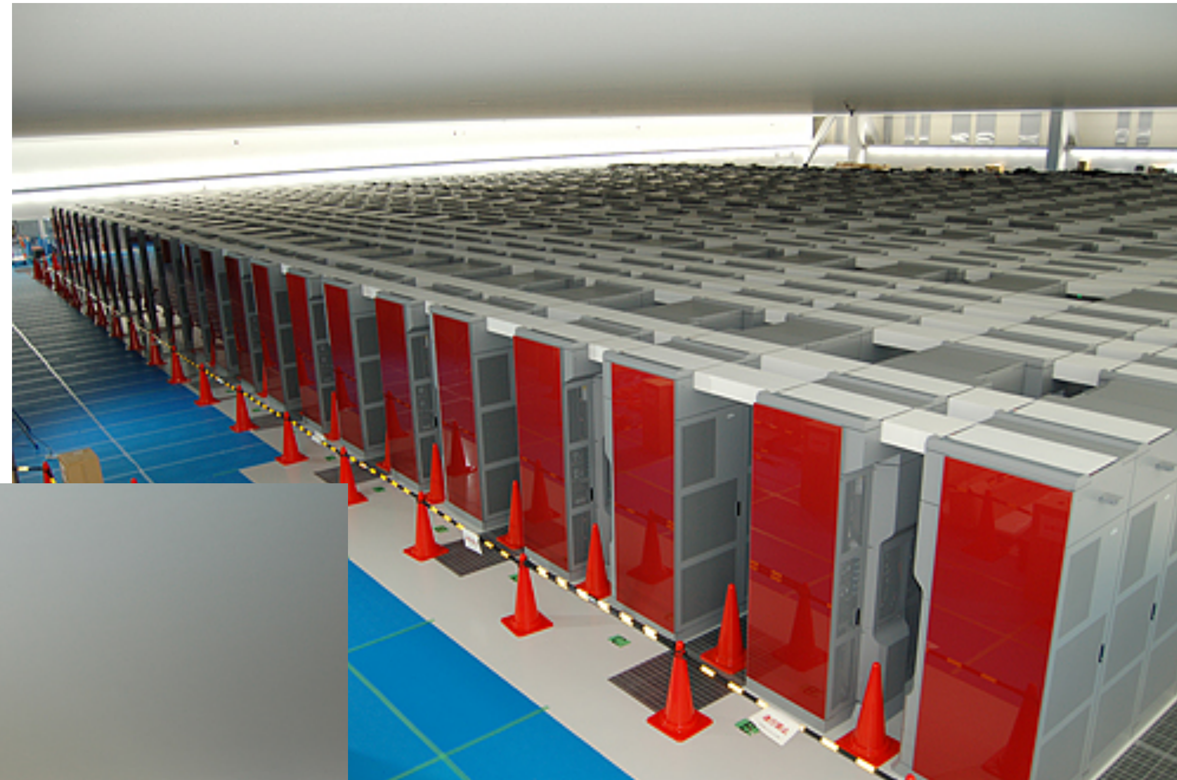
Top 5 Supercomputers as of June 2011

	NAME/MANUFACTURER/COMPUTER	LOCATION	COUNTRY	CORES	R_{max} Pflop/s
1	**K Computer** SPARC64 VIIIfx 2.0GHz, Tofu interconnect	RIKEN	Japan	548,352	8.16
2	**Tianhe-1A** 6-core Intel X5670 2.93 GHz + Nvidia M2050 GPU w/custom interconnect	NUDT/NSCC/Tianjin	China	186,368	2.56
3	**Jaguar** Cray XT-5 6-core AMD 2.6 GHz w/custom interconnect	DOE/SC/ORNL	USA	224,162	1.76
4	**Nebulae** Dawning TC3600 Blade Intel X5650 2.67 GHz, NVidia Tesla C2050 GPU w/ Iband	NSCS	China	120,640	1.27
5	**Tsubame 2.0** HP Proliant SL390s G7 nodes (Xeon X5670 2.93GHz) , NVIDIA Tesla M2050 GPU w/Iband	TiTech	Japan	73,278	1.19



Source: www.top500.org

Japan RIKEN AICS' K Computer



LINPACK:	**8.162x10^15**
Peak Perf:	**8.773x10^15**
#CPUs (2 GHz, 8-core):	**68,644**
#Cores:	**548,352**
Efficiency:	**93%**
Network (6D torus):	**Tofu**
Total Power:	**10 MW, $10M/year**

How Fast is the Fastest Supercomputer as of June 2011?

The fastest supercomputer (K Computer) can perform:

8,162 Quadrillion =8,162,000,000,000,000

Floating-Point Operations Per Second

How Fast is the Fastest Supercomputer as of June 2010?

The Total US Population (as of June 2011) **300,000,000**

How Fast is the Fastest Supercomputer as of June 2010?

Assuming each American aging 0-100 years can perform 1 FP/s

Operations taken supercomputers 1 second
would have taken all Americans

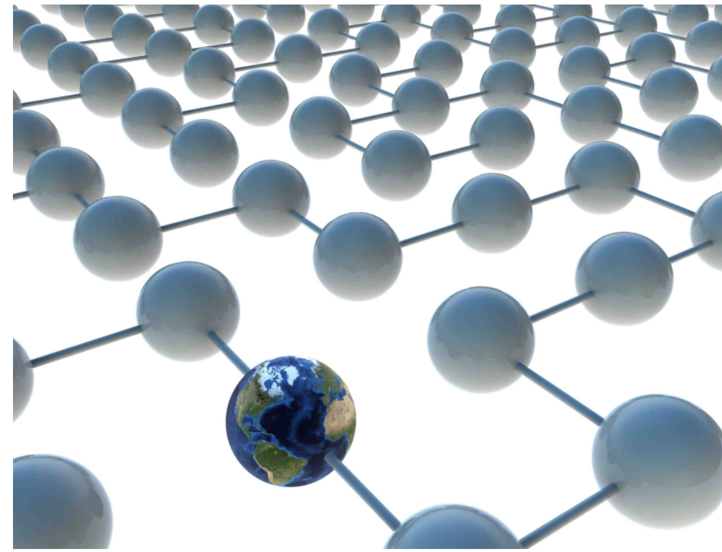
8,162,000,000,000,000/300,000,000=27.2 Million Seconds

OR

315 Days

Supercomputer's key components

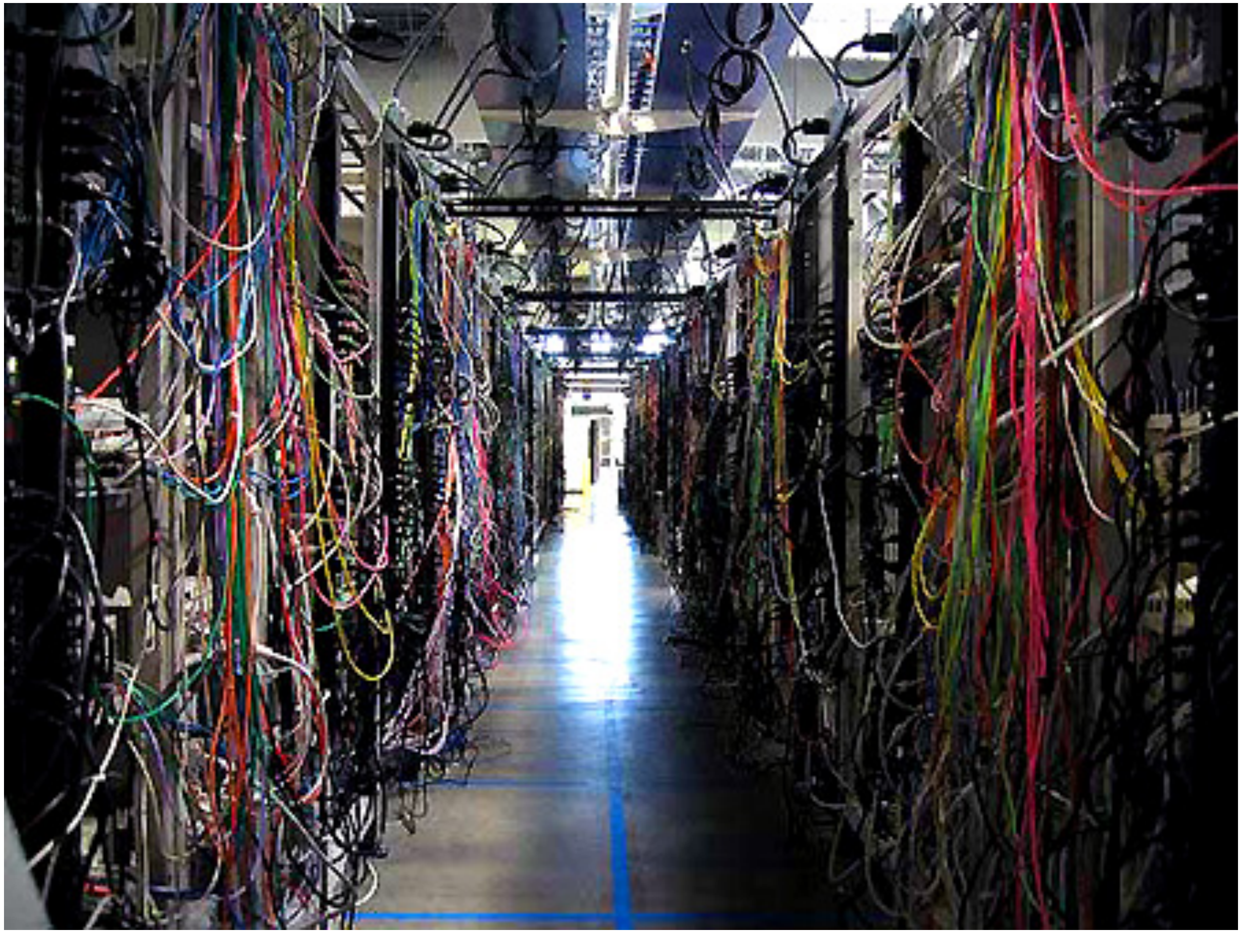
- Interconnection networks
- Node processors
- Storage subsystem



A supercomputers (outside)?

10

A supercomputer (inside)?



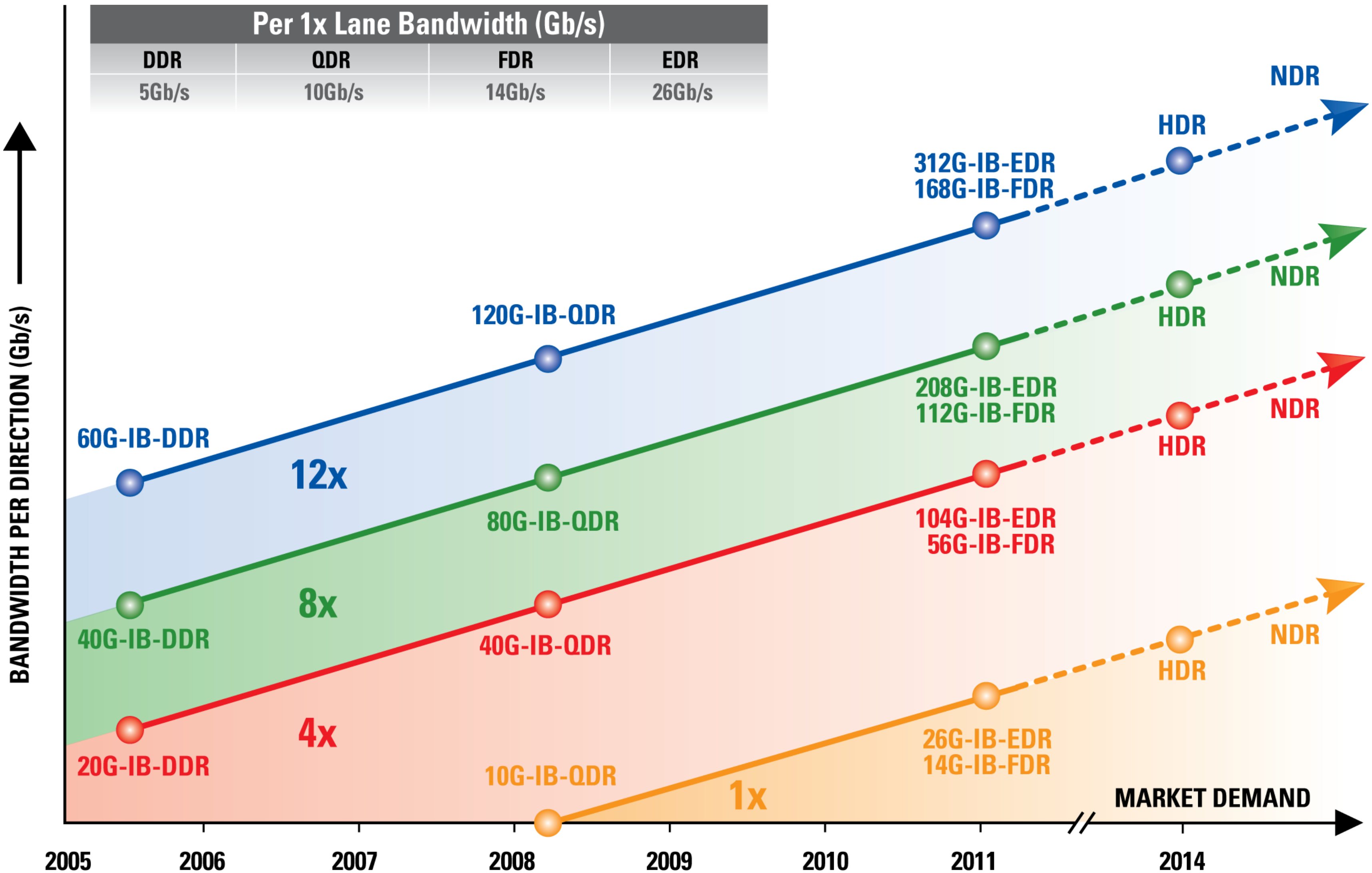
Key Technologies

- **System power** is a first class constraint on exascale system performance and effectiveness.
- **Memory** is an important component of meeting exascale power and applications goals.
- **Programming model**. Early investment in several efforts to decide in 2013 on exascale programming model, allowing exemplar applications effective access to 2015 system for both mission and science.
- **Investment in exascale processor design** to achieve an exascale-like system in 2015.
- **Operating System strategy for exascale** is critical for node performance at scale and for efficient support of new programming models and run time systems.
- **Reliability and resiliency are critical at this** scale and require applications neutral movement of the file system (for check pointing, in particular) closer to the running apps.
- ***HPC co-design strategy and implementation*** *requires a set of a hierarchical performance models and simulators as well as commitment from apps, software and architecture communities.*

Exascale Expectations

Systems	2010	2018	Difference Today & 2018
System peak	2 Pflop/s	1 Eflop/s	O(1000)
Power	6 MW	~20 MW (goal)	
System memory	0.3 PB	32 - 64 PB	O(100)
Node performance	125 GF	1.2 or 15TF	O(10) – O(100)
Node memory BW	25 GB/s	2 - 4TB/s	O(100)
Node concurrency	12	O(1k) or O(10k)	O(100) – O(1000)
Total Node Interconnect BW	3.5 GB/s	200-400GB/s (1:4 or 1:8 from memory BW)	O(100)
System size (nodes)	18,700	O(100,000) or O(1M)	O(10) – O(100)
Total concurrency	225,000	O(billion) + [O(10) to O(100) for latency hiding]	O(10,000)
Storage Capacity	15 PB	500-1000 PB (>10x system memory is min)	O(10) – O(100)
IO Rates	0.2 TB	60 TB/s	O(100)
MTTI	days	O(1 day)	- O(10)

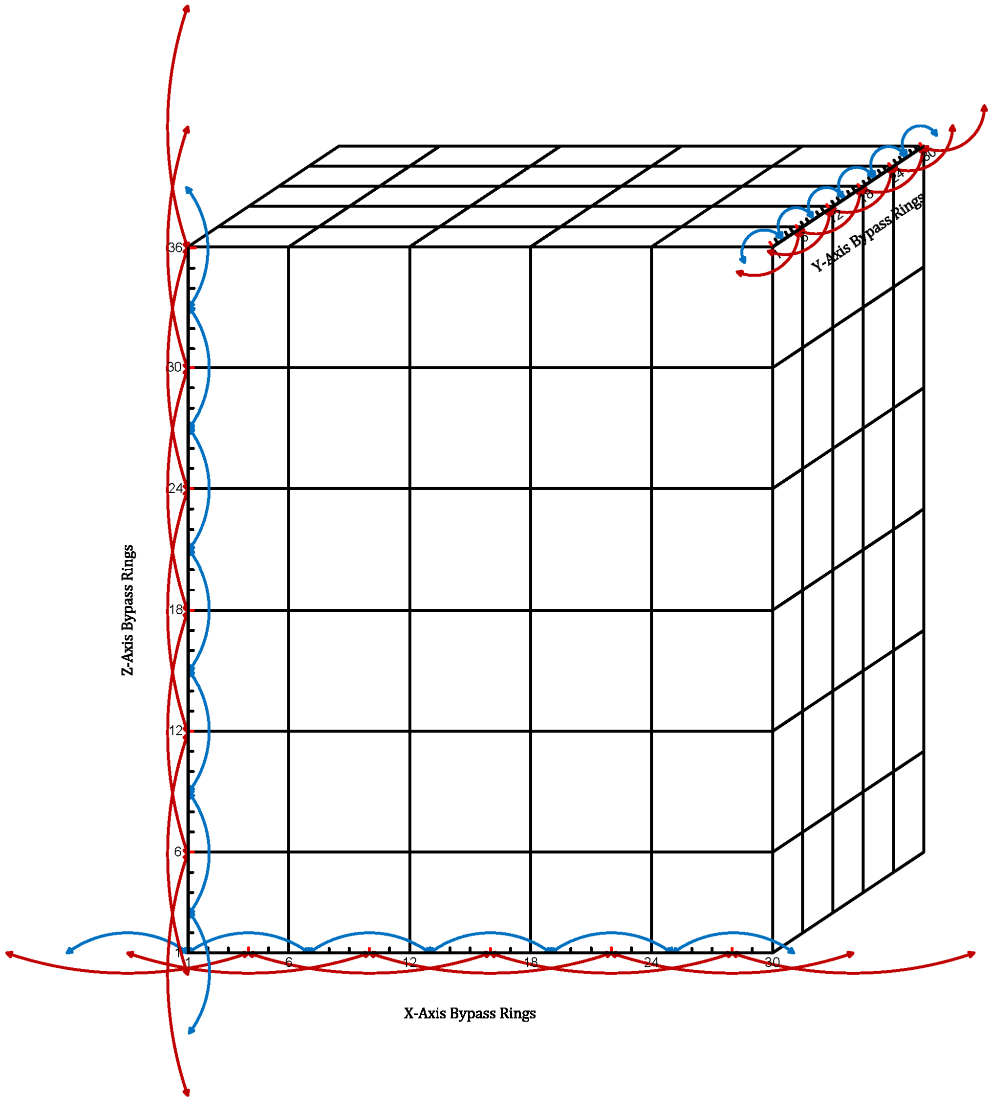
Network Road Map



Traffic (Before)

Traffic (after)

Interlace Bypass Torus



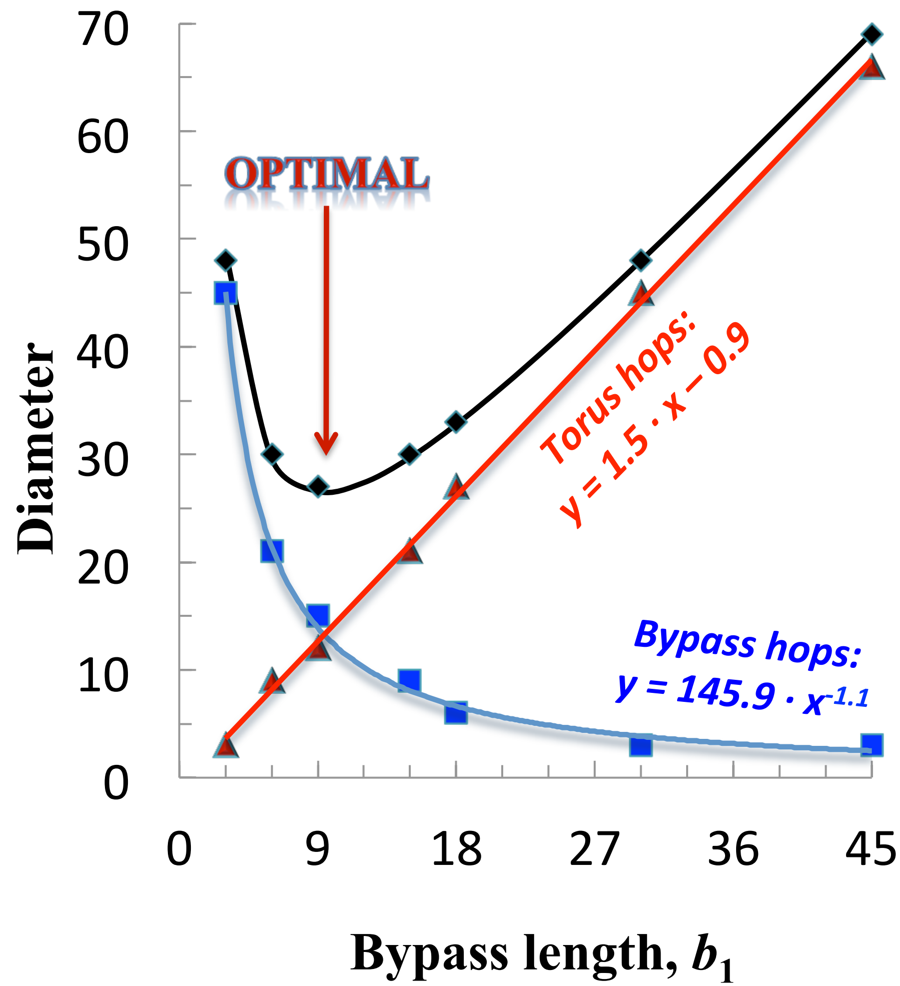
Source: P. Zhang, R. Powell, and Y. Deng, IEEE **Trans. Parallel and Distributed Systems Vol. 22** Issue 2 (2011) pp. 287-295

Impact of Bypass Schemes

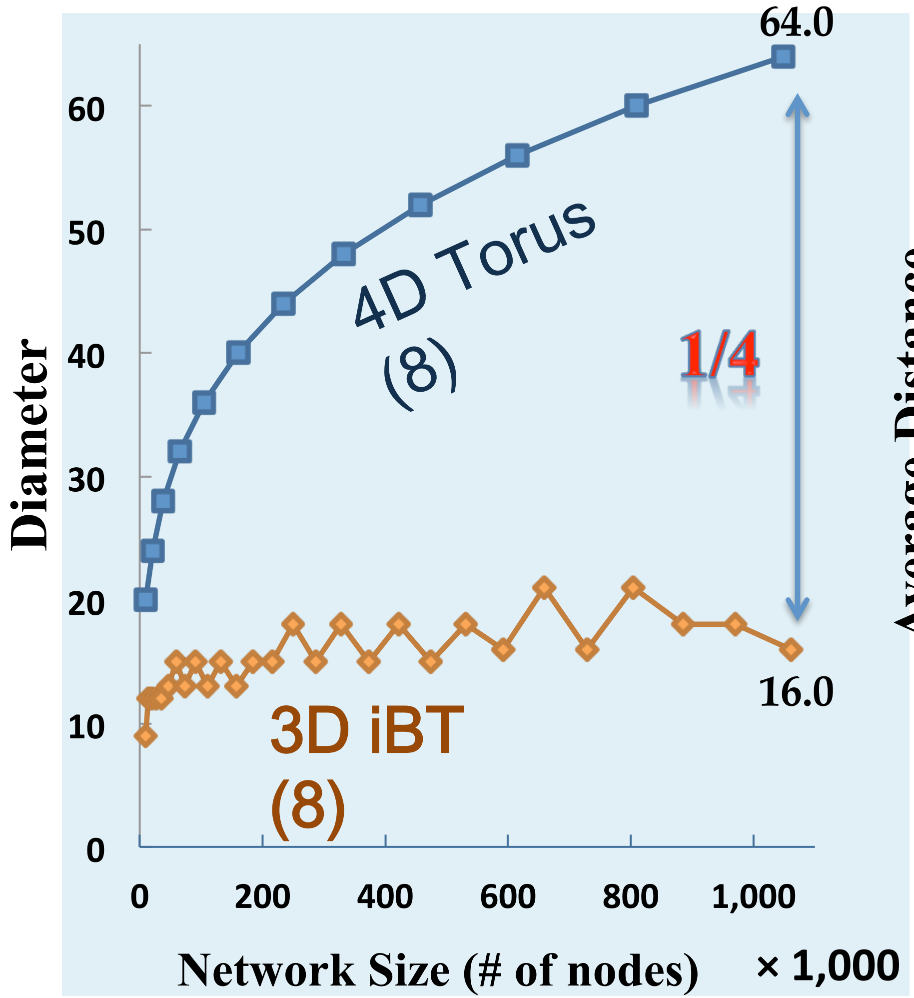
- Diameter:
  - Longest shortest path
- Path in iBT:
  - Torus hops
  - Bypass hops
- Network:
  - iBT(90^3; $\boldsymbol{b}$=<b_1>)

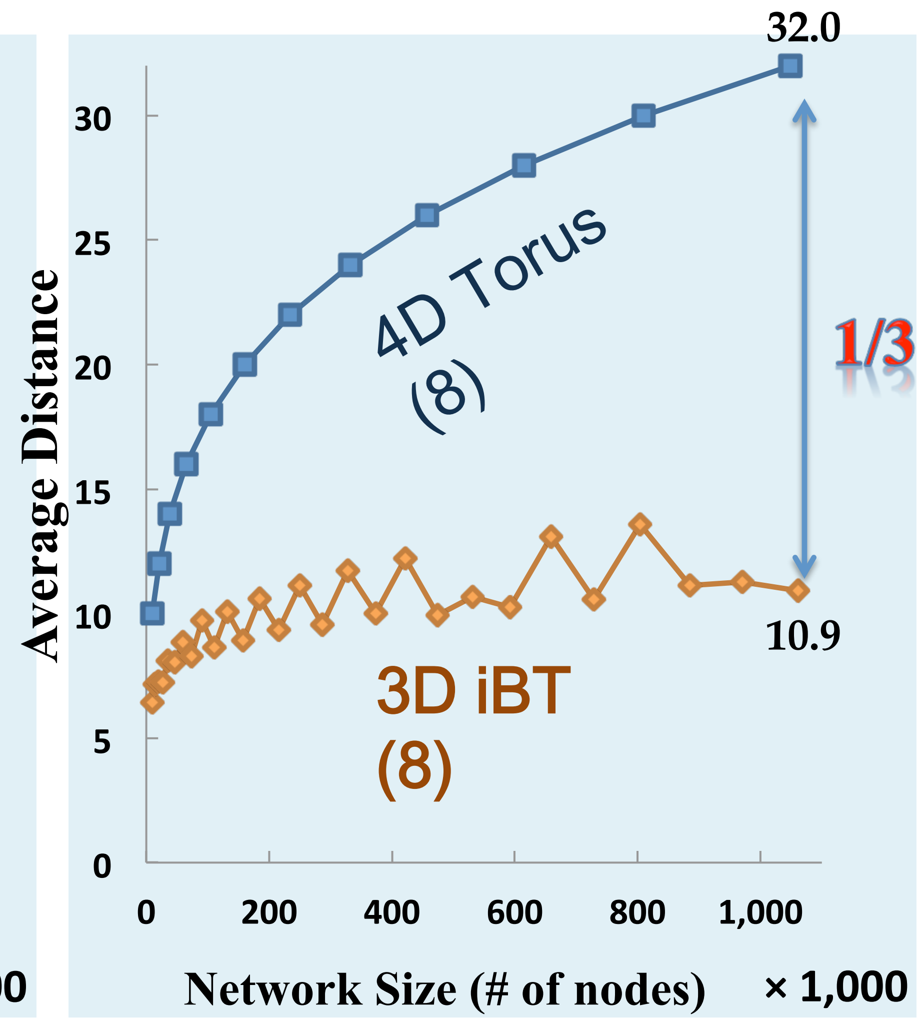
Diameter(y) = Function of b_1 (x)

$$y = 145.9 \cdot x^{-1.1} + 1.5 \cdot x - 0.9$$



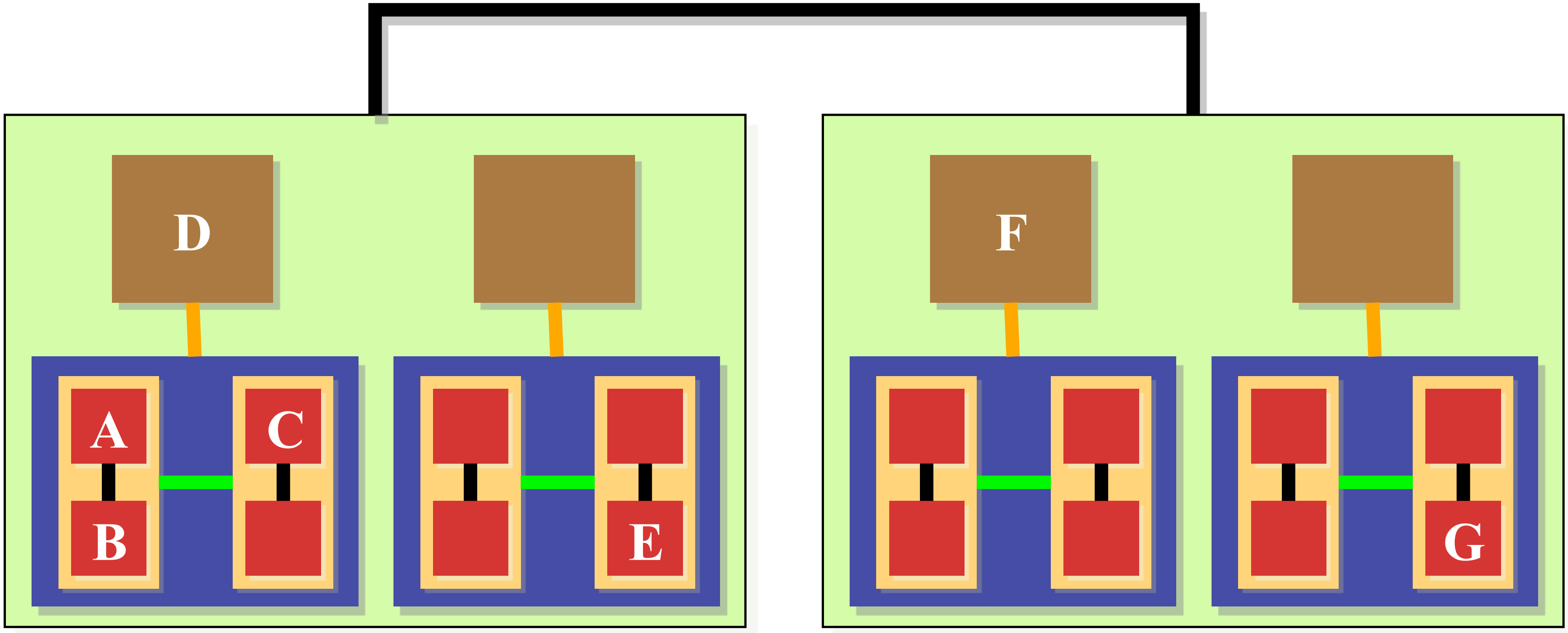
3D iBT vs. 4D Torus





Algorithm:
Task Mapping

Data Access Rates for RoadRunner



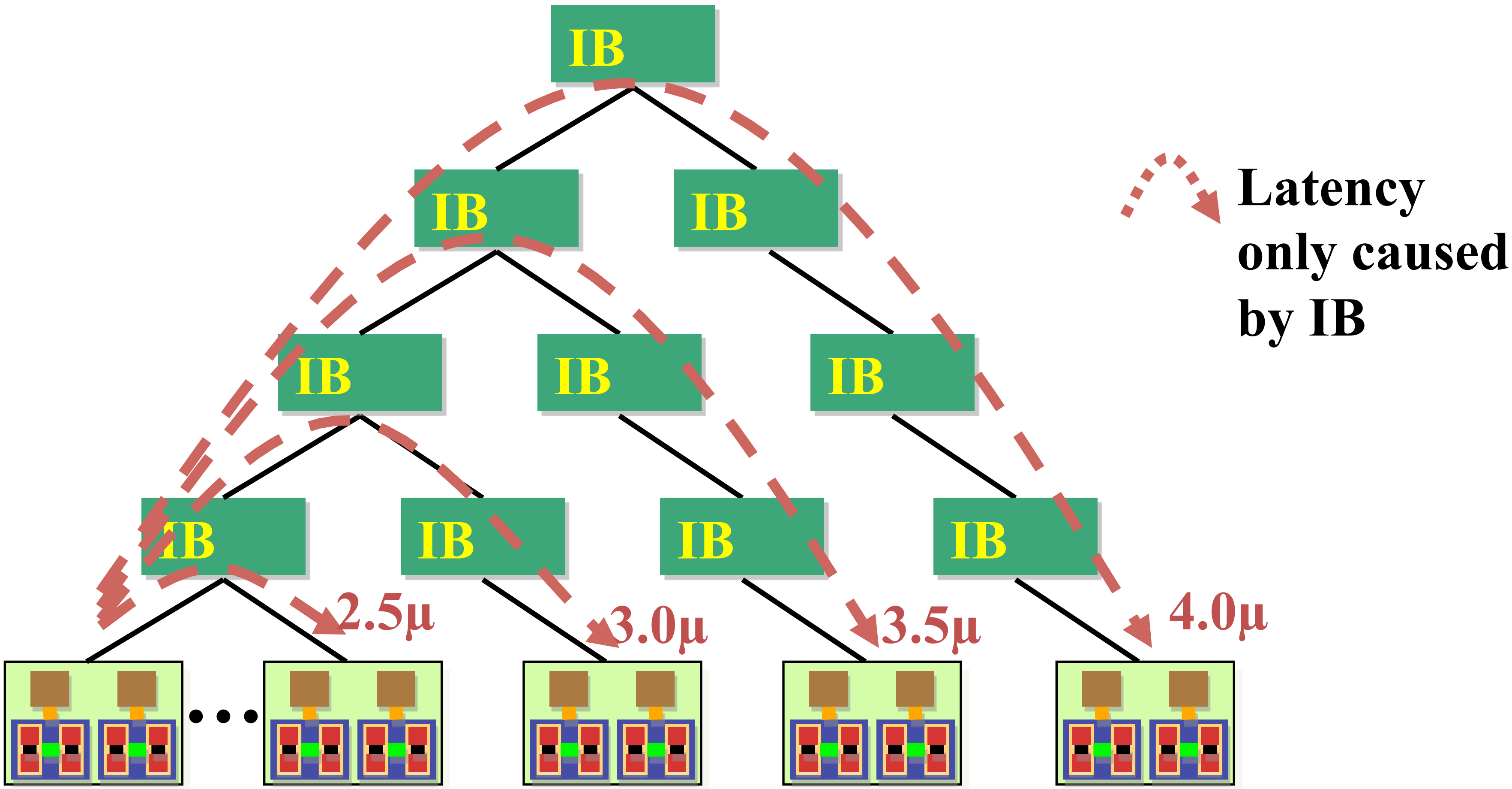
A One Cell Core

D AMD Opteron

A—B	0.3μs	A—F	5.47μs
A—C	0.8μs	A—E	6.4μs
D—F	2.16μs	A—G	8.78μs
A—D	3.19μs		

Reference: K. Barker, K. Davis, A. Hoisie, D. Kerbyson, M. Lang, S. Pakin, J.C. Sancho, Entering the Petaflop Era: The Architecture and Performance of Roadrunner, SC2008

Long-Ranged Messaging in RR



Reference: K. Barker, K. Davis, A. Hoisie, D. Kerbyson, M. Lang, S. Pakin, J.C. Sancho, Entering the Petaflop Era: The Architecture and Performance of Roadrunner, SC2008

Task Mapping Graph Theory Model

Application Demand Matrix

$$D = \left[d_{ij}\right] \in R^{n\times n}$$

Number of bytes from task *i* to *j*

Network Supply Matrix

$$S = \left[s_{ij}\right] \in R^{n\times n}$$

Number of hops from node *i* to *j*

Mapping Operator

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ i_1 & i_2 & \cdots & i_n \end{pmatrix} \begin{matrix} \leftarrow task\ id \\ \leftarrow node\ id \end{matrix}$$

Node *i* executes task *σ*(*i*)

# of bytes from task i to j

# of hops from task i to j

$$f_{S,D}(\sigma) = \sum_{i,j=1}^{n} d_{ij} \cdot s_{\sigma(i)\sigma(j)}$$

Hop-Bytes Metric

Mapping Model – Eigenanalysis

Application Demand Matrix

$$D = [d_{ij}] \in R^{n \times n}$$

$$B = (D + D^T)/2$$

β_i and $\boldsymbol{p}_i$ are eigenpairs

Network Supply Matrix

$$S = [s_{ij}] \in R^{n \times n}$$

α_i and $\boldsymbol{q}_i$ are eigenpairs

Mapping Operator

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ i_1 & i_2 & \cdots & i_n \end{pmatrix} \begin{matrix} \leftarrow task\ id \\ \leftarrow node\ id \end{matrix}$$

Mapping Matrix

$$M = [m_{ij}] \in R^{n \times n}$$

$$m_{ij} = \begin{cases} 1, & \text{task } j \text{ on node } i, \\ 0, & \text{otherwise.} \end{cases}$$

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$$

$$C \triangleq \left[c_{ij} = (\boldsymbol{q}_i^T M \boldsymbol{p}_j)^2\right]$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}$$

$$f_{S,D}(\sigma) = \boldsymbol{\alpha}^T \cdot C \cdot \boldsymbol{\beta}$$

Hop-Bytes Metric

Basic Model

$$Min \quad \left\{ \sum_{t=1}^{n} \sum_{p=1}^{m} L(t,p) \cdot x_{tp} + \sum_{t=1}^{n} \sum_{t'=1}^{n} D(t,t') \cdot \left(\sum_{p=1}^{m} \sum_{p'=1}^{m} S(p,p') \cdot y_{tt'pp'} \right) \right\}$$

subject to:

$$\begin{cases} \sum_{p=1}^{m} x_{tp} = 1 \quad (t = 1,\ldots,n) \\ \sum_{t=1}^{n} x_{tp} \geq 1 \quad (p = 1,\ldots,m) \\ \sum_{t=1}^{n} x_{tp} \leq \left\lceil \frac{A_p}{A_T} \times n \right\rceil \quad (p = 1,\ldots,m) \\ x_{tp} + x_{t'p'} \leq 1 + y_{tt'pp'} \quad (t,t' = 1,\ldots n, t < t'; p, p = 1,\ldots m) \end{cases}$$

where

$$x_{tp} = \begin{cases} 1, & \text{if task } t \text{ is assigned to processor } p \\ 0, & \text{otherwise} \end{cases}$$

$$y_{tt'pp'} = \begin{cases} 1, & \text{if task } t \text{ and } t' \text{ are assigned to processor } p \text{ and } p' \text{ respectively} \\ 0, & \text{otherwise} \end{cases}$$

Source: Y. Chen, Y. Deng. Task Mapping on Supercomputers with Cellular Networks. *Computer Physics Communications*, 2008. doi:10.1016/j.cpc.2008.04.011

Enhanced Model

$$Min \quad \left\{ \sum_{t=1}^{n} \sum_{p=1}^{m} L(t,p) \cdot x_{tp} + \sum_{i=1}^{k} (D_i S)_{\max} \right\}$$

subject to:

$$\begin{cases} \sum_{p=1}^{m} x_{tp} = 1 & (t = 1, \dots, n) \\ \sum_{t=1}^{n} x_{tp} \geq 1 & (p = 1, \dots, m) \\ \sum_{t=1}^{n} x_{tp} \leq \left\lceil \frac{A_p}{A_T} \times n \right\rceil & (p = 1, \dots, m) \end{cases}$$

where $x_{tp} = \begin{cases} 1, & \textit{if task } t \textit{ is assigned to processor } p \\ 0, & \textit{otherwise} \end{cases}$

$(D_i S)_{\max}$ *Max in the i batch communication*

Mapping PDEs to 6D Topology

Analysis of Task Mapping for Parallel Supercomputers

A Dissertation Presented

by

Janet Laura Braunstein

to

The Graduate School

in Partial Fulfillment

of the Requirements

for the Degree of

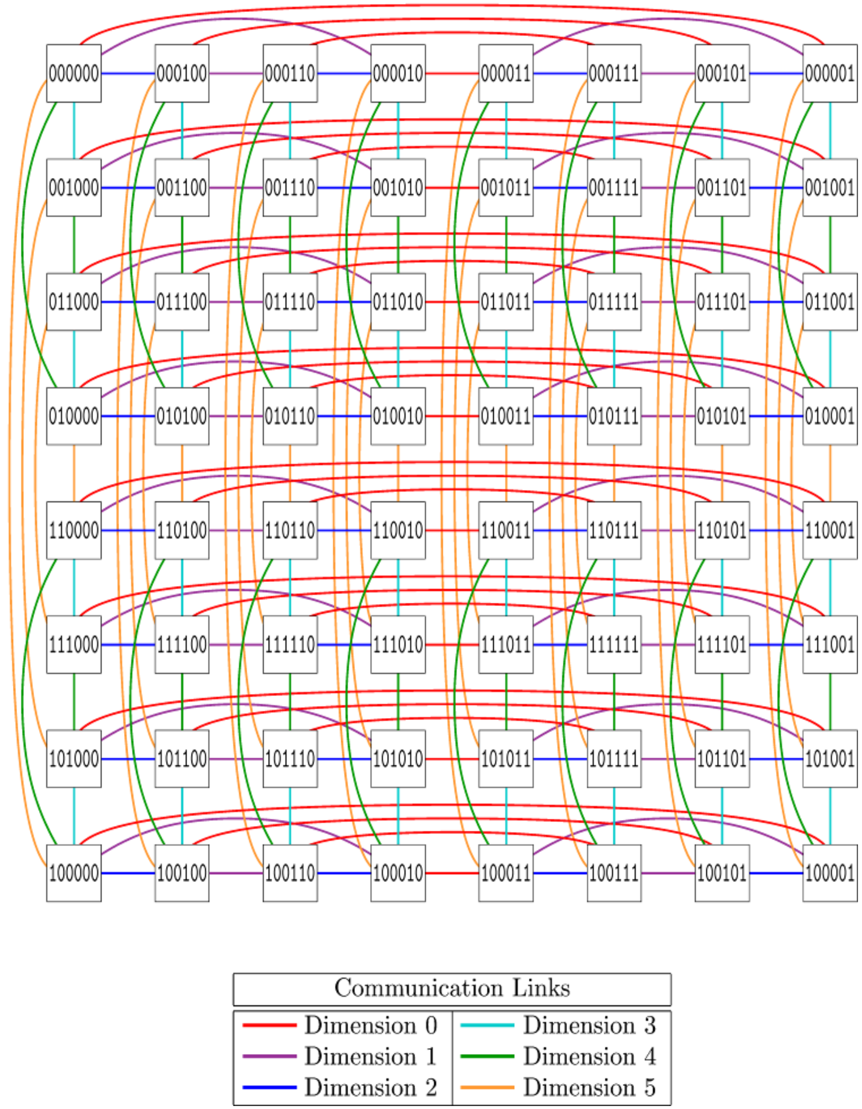
Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

August 2007



BG/L Network Supply Matrix

Task Mapping on Supercomputer with Cellular Network

A Dissertation Presented

by

Yongzhi Chen

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

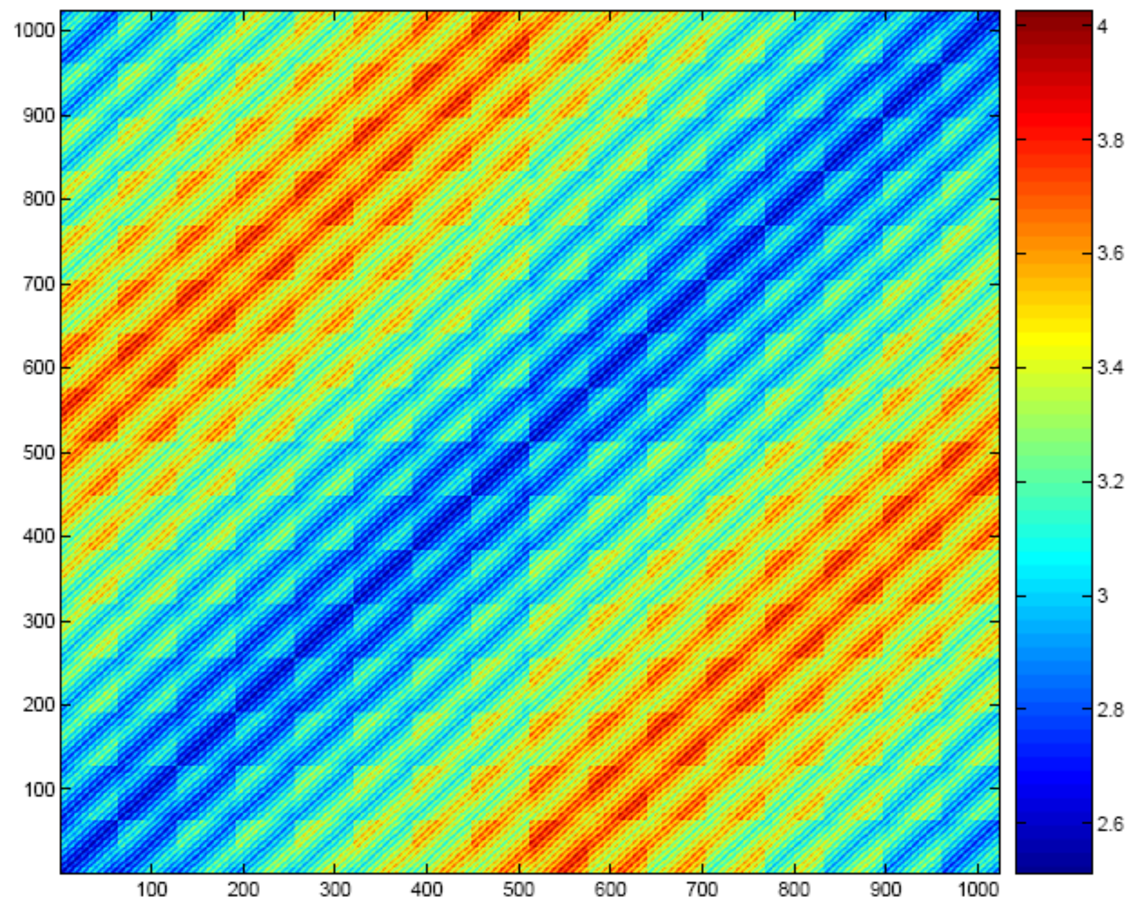
Doctor of Philosophy

in

Applied Mathematics and Statistics

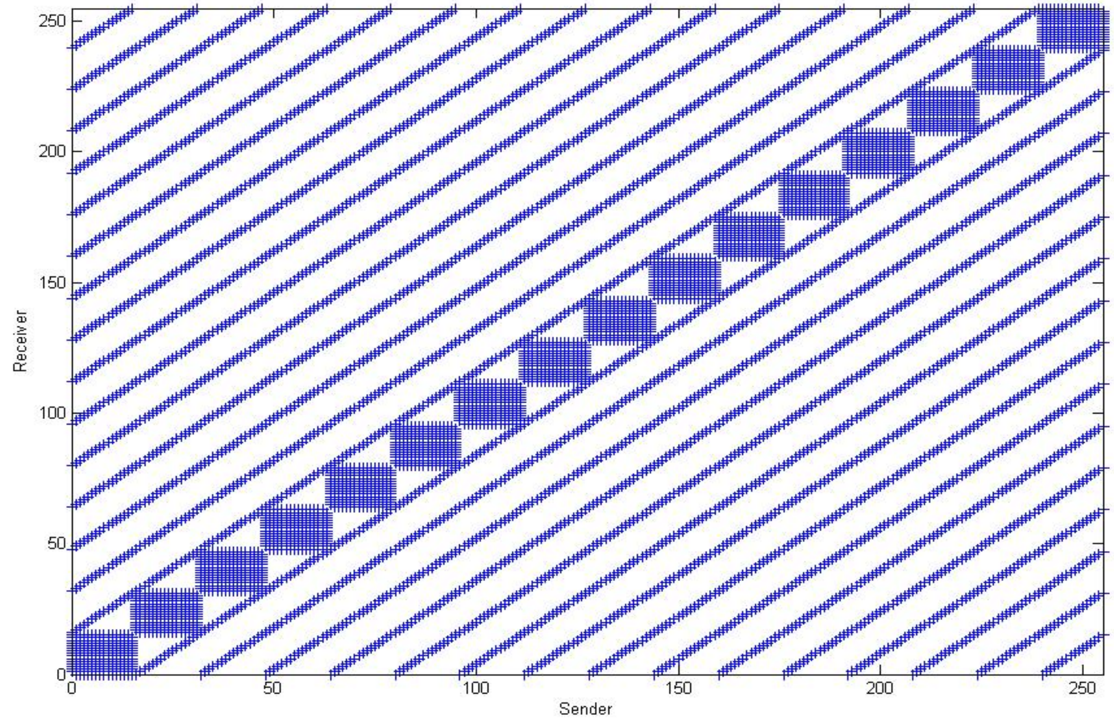
Stony Brook University

May 2008

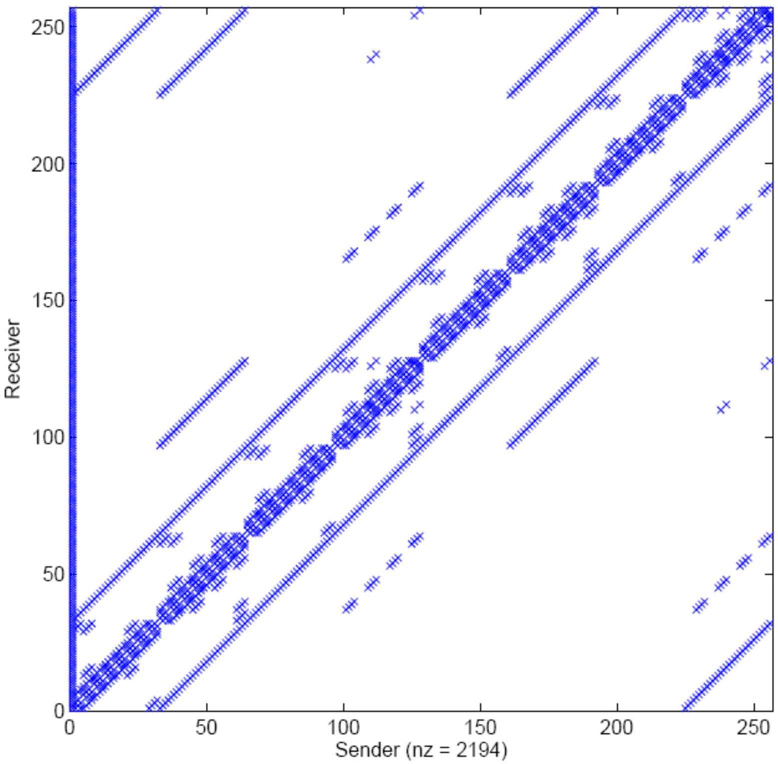


Latency in micro-second on 8x8x16 BG/L computer

Demand Matrices



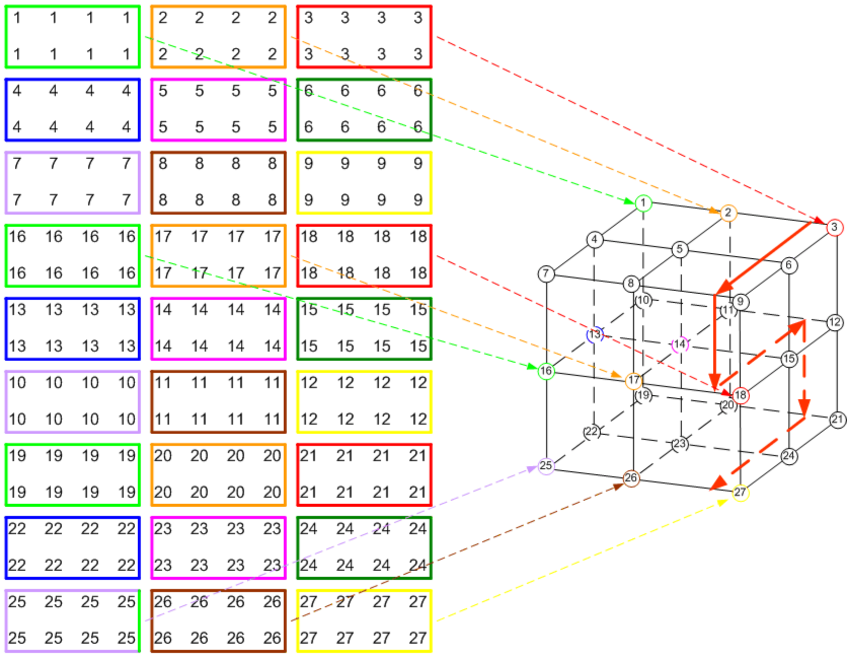
2D FFT Communication Demand Matrix



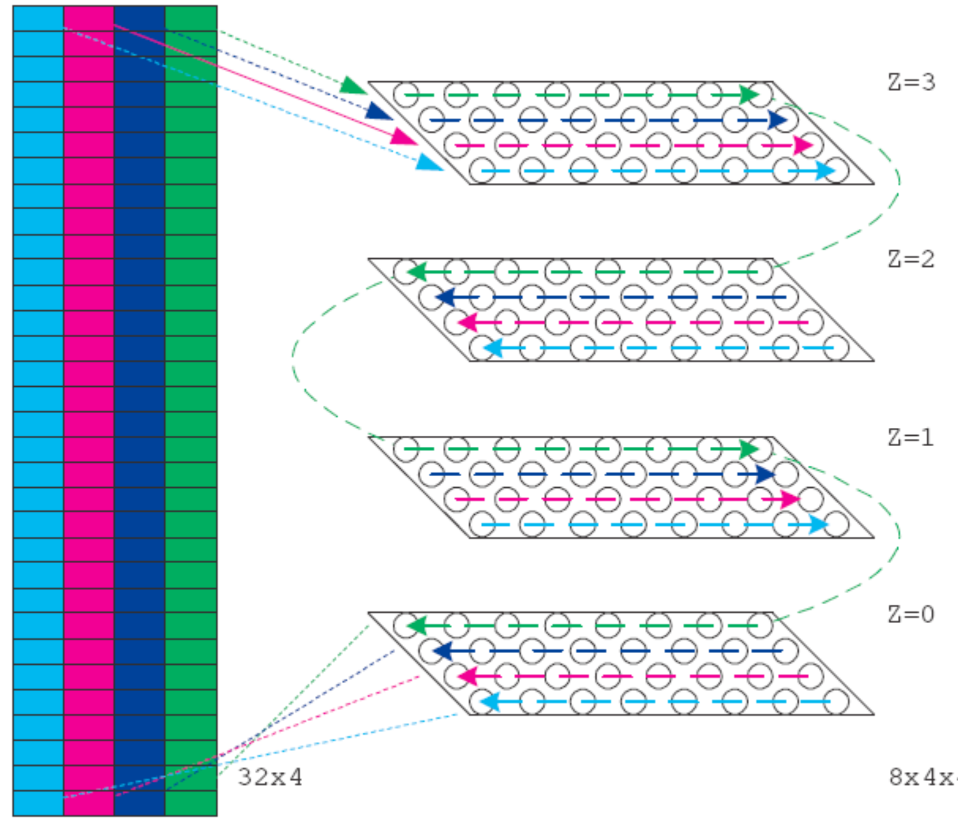
Multi-Grid Communication with 256 Nodes

Source: Y. Chen and Y. Deng, *Task mapping on supercomputers with cellular networks*, Comp. Phys. Comm., 179 (2008), pp. 479-485

Mapping 2D Physics to 3D Network



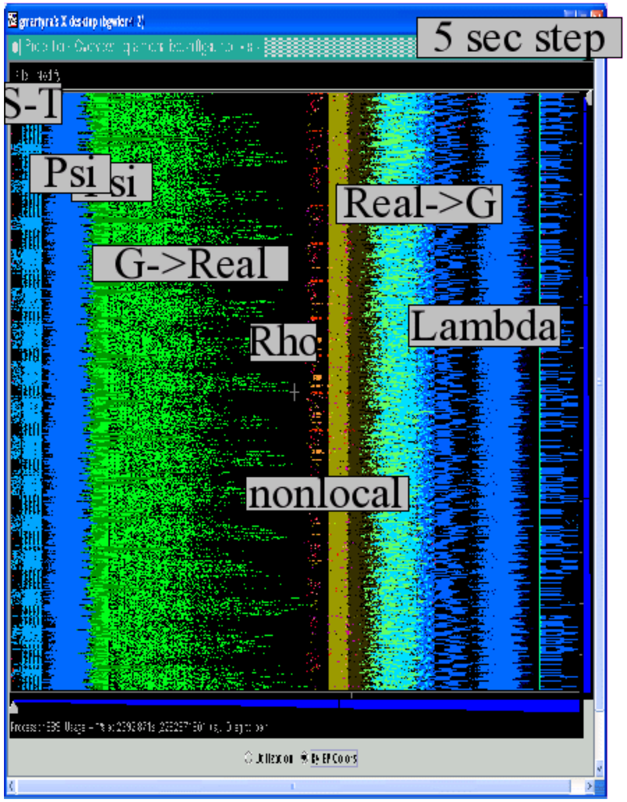
2D Hyperbolic PDEs (12x18)
Mapped to 3D (3x3x3) Network

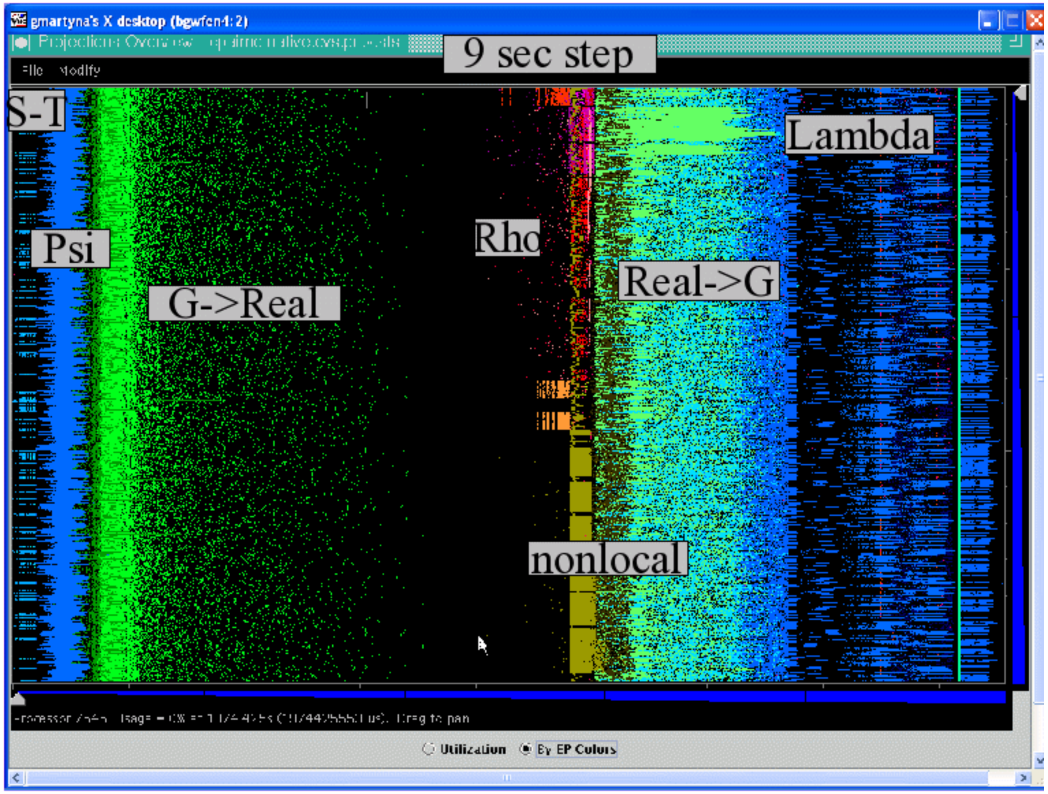


2D Hyperbolic PDEs (32x4)
Mapped to 3D (8x4x4) Network

Source: Y. Chen and Y. Deng, *A Detailed analysis of communication load balance on BlueGene supercomputer*, Comp. Phys. Comm., 180 (2009), pp 1251-1258

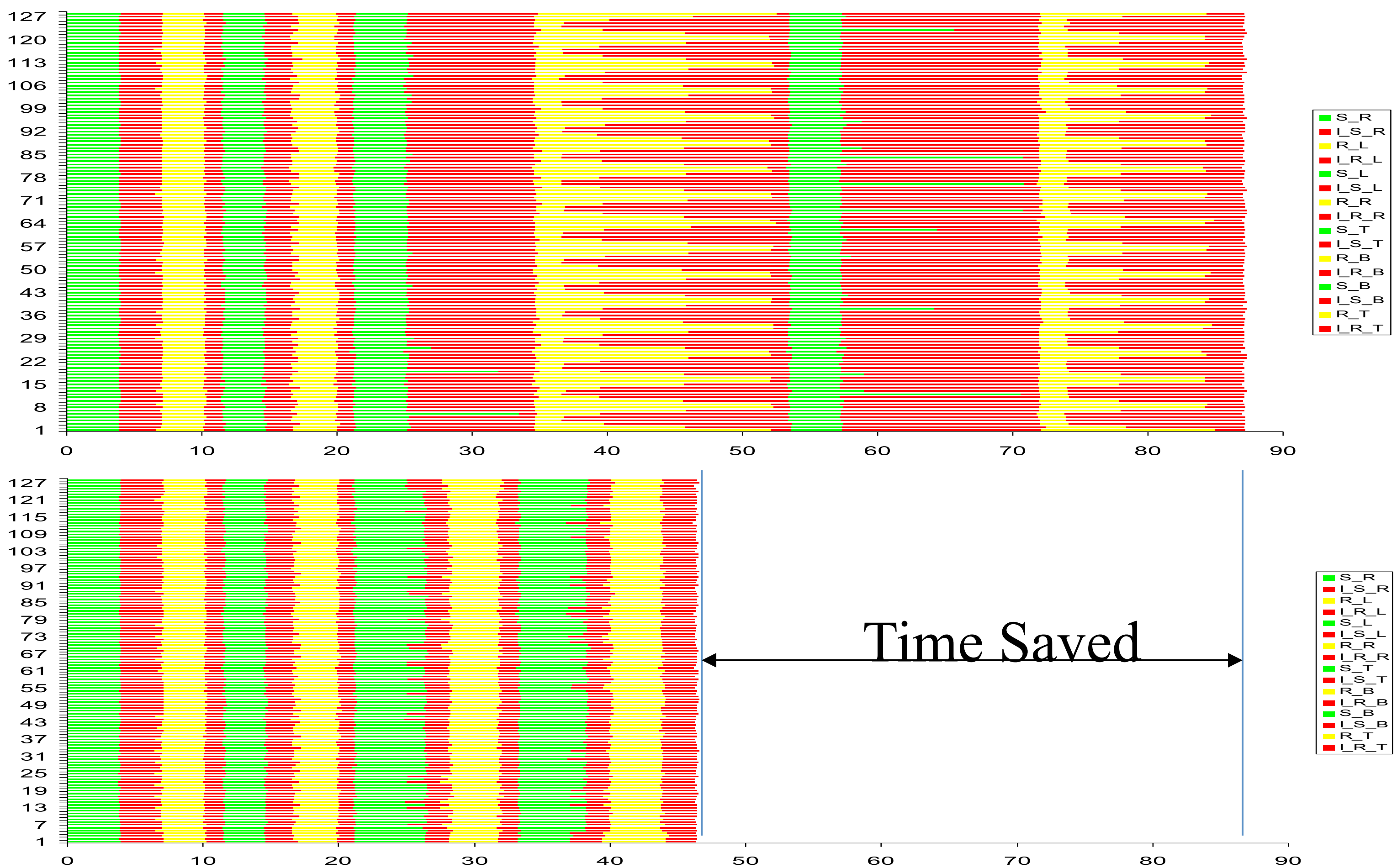
Mapping Improves Performance





Density (R) reduction and multicast to State (R) improved.
State (G) communication to/from orthogonality chares improved.

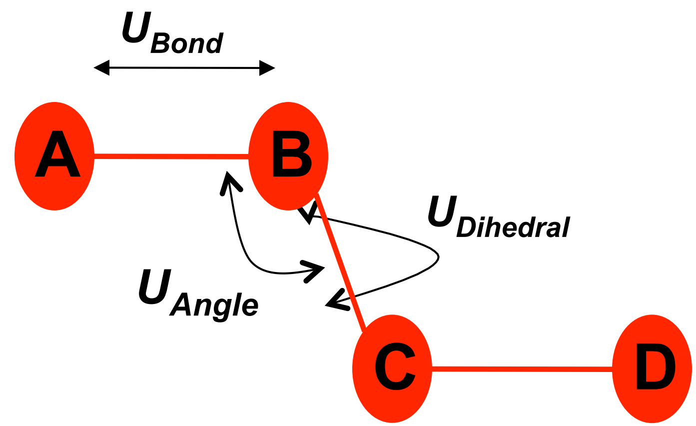
Link Load Balance for 2D Wave Eq. on 128 CPUs BG



Source: Y. Chen and Y. Deng, *A Detailed analysis of communication load balance on BlueGene supercomputer*, Comp. Phys. Comm., 180 (2009), pp 1251-1258

Algorithm:
Multiple Time Stepping

Classical Force Field



$$
\begin{aligned}
U(\vec{R}) = & \sum_{\text{Bonds}} k_i^{\text{bond}} (r_i - r_0)^2 \\
& + \sum_{\text{Angles}} k_i^{\text{angle}} (\theta - \theta_0)^2 \\
& + \sum_{\text{Dihed}} k_i^{\text{dihed}} \left(1 + \cos(n_i \phi_i - \delta_i)\right) \\
& + \sum_{i} \sum_{\substack{j \neq i \\ i,j \text{ not excl}}} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + U_{\text{Coulomb}}(\vec{R})
\end{aligned}
$$

Coulomb Potential

$$U_{\text{Coulomb}}(\vec{R}) = \frac{1}{4\pi\varepsilon_0}\sum_{n}\sum_{i=1}^{N^*}\sum_{j=i+1}^{N^*} q_i q_j \frac{erfc(\alpha\left|r_{ij}+n\right|)}{\left|r_{ij}+n\right|}$$

$$+\frac{1}{\varepsilon_0 V}\sum_{k>0}\frac{1}{k^2}e^{-\frac{k^2}{4\alpha^2}}\left\{\left|\sum_{i=1}^{N} q_i\cos(k\cdot r_i)\right|^2+\left|\sum_{i=1}^{N} q_i\sin(k\cdot r_i)\right|^2\right\}$$

$$-\frac{1}{4\pi\varepsilon_0}\sum_{i=1}^{N}\sum_{i\le j}^{M^*} q_i q_j\left\{\delta_{ij}\frac{\alpha}{\sqrt{\pi}}+\frac{erf(\alpha r_{ij})}{r_{ij}^{\,1-\delta_{ij}}}\right\}$$

Definitions of Symbols

n = lattice vector of periodic MD cell

k = reciporical lattice vector

N = total number of atoms

N^* = N discounting any exluded itramolecular charges

M^* = Number of excluded atoms

α = Ewald parameter

δ_{ij} = Kronecker delta symbol

V = Volume of MD cell

MTS: Cut-off vs. Jump Steps

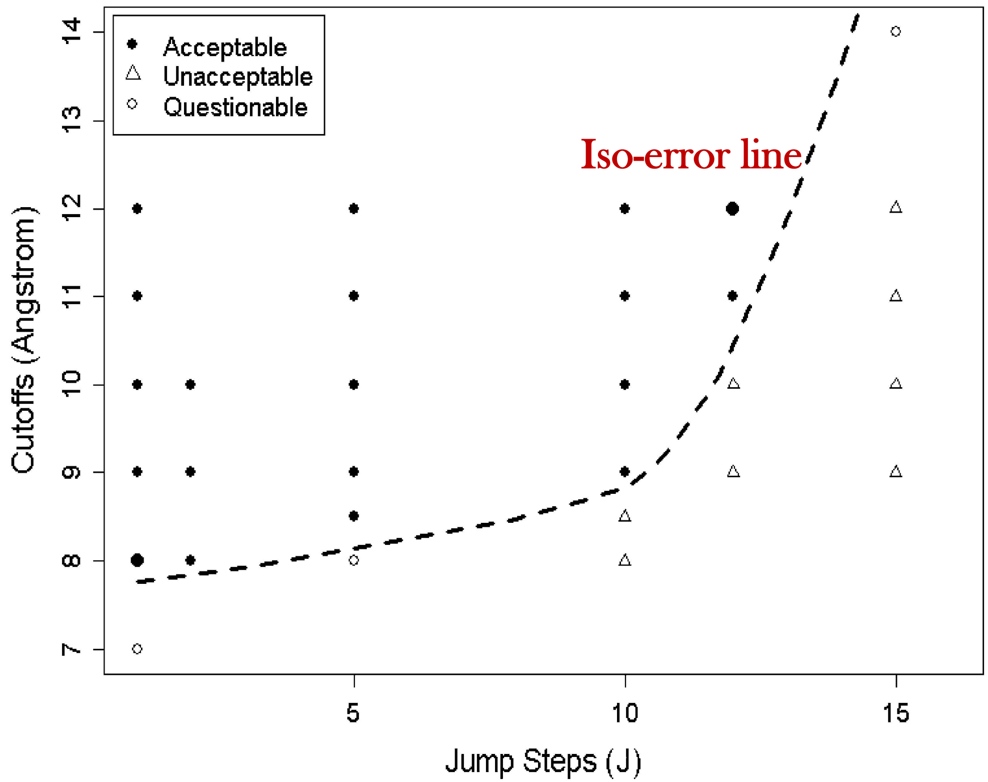


Fig. 9. The cutoff and jump step phase diagram. This diagram reveals the boundary in the cutoff-jump plane between convergent (the upper left region) and divergent (the lower right region) simulations. The value $J = 12$, the maximum possible jump step predicted by the 1d model, also serves as a cutoff for the jump steps in the many-body system.

G. Han, Y. Deng, J. Glimm, G. Martyna, *Error and timing analysis of multiple Time Stepping for MD*, Comp. Phys. Comm. 176 (2007) 271-291

Speedup Due to MTS

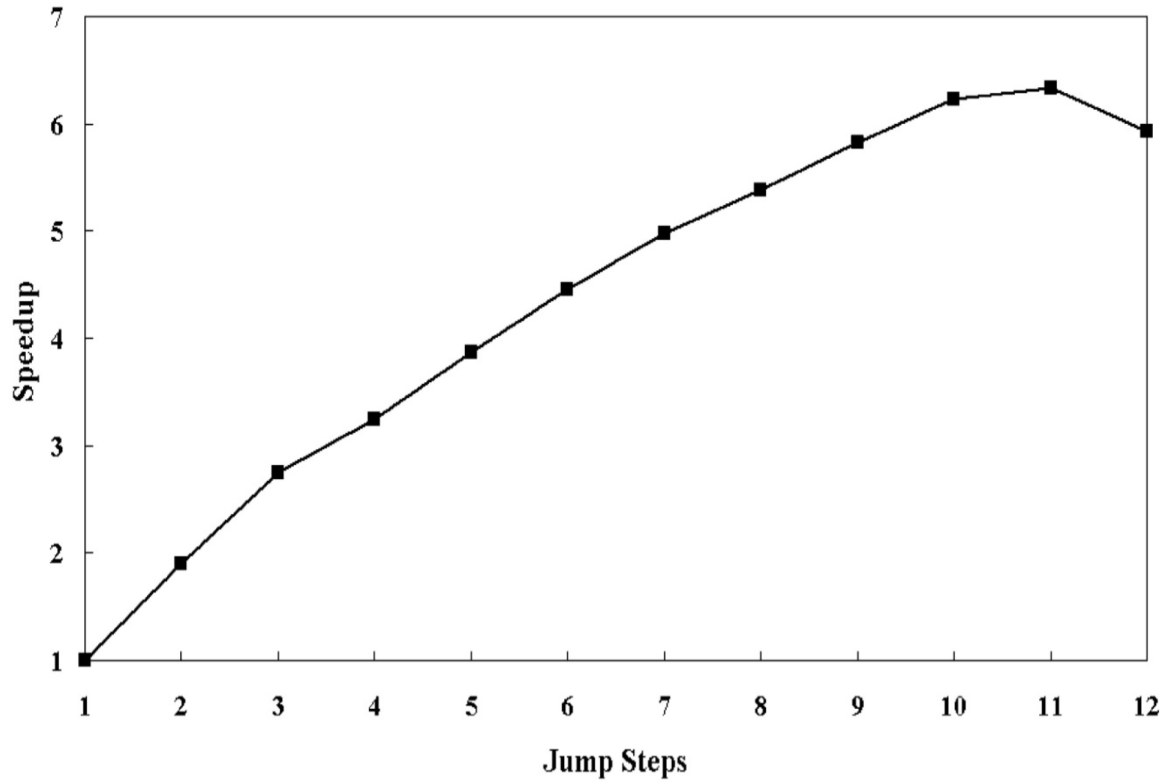
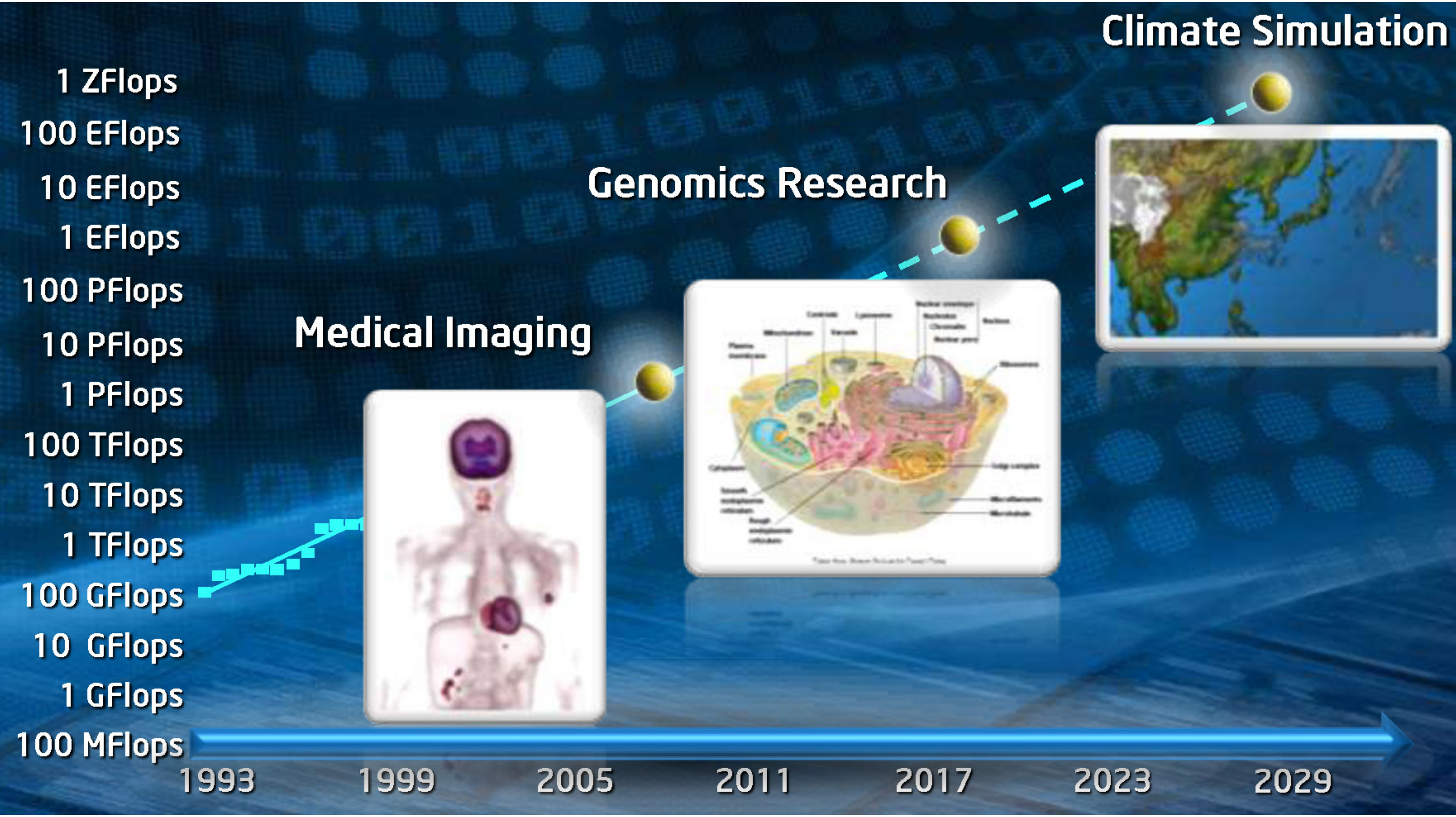


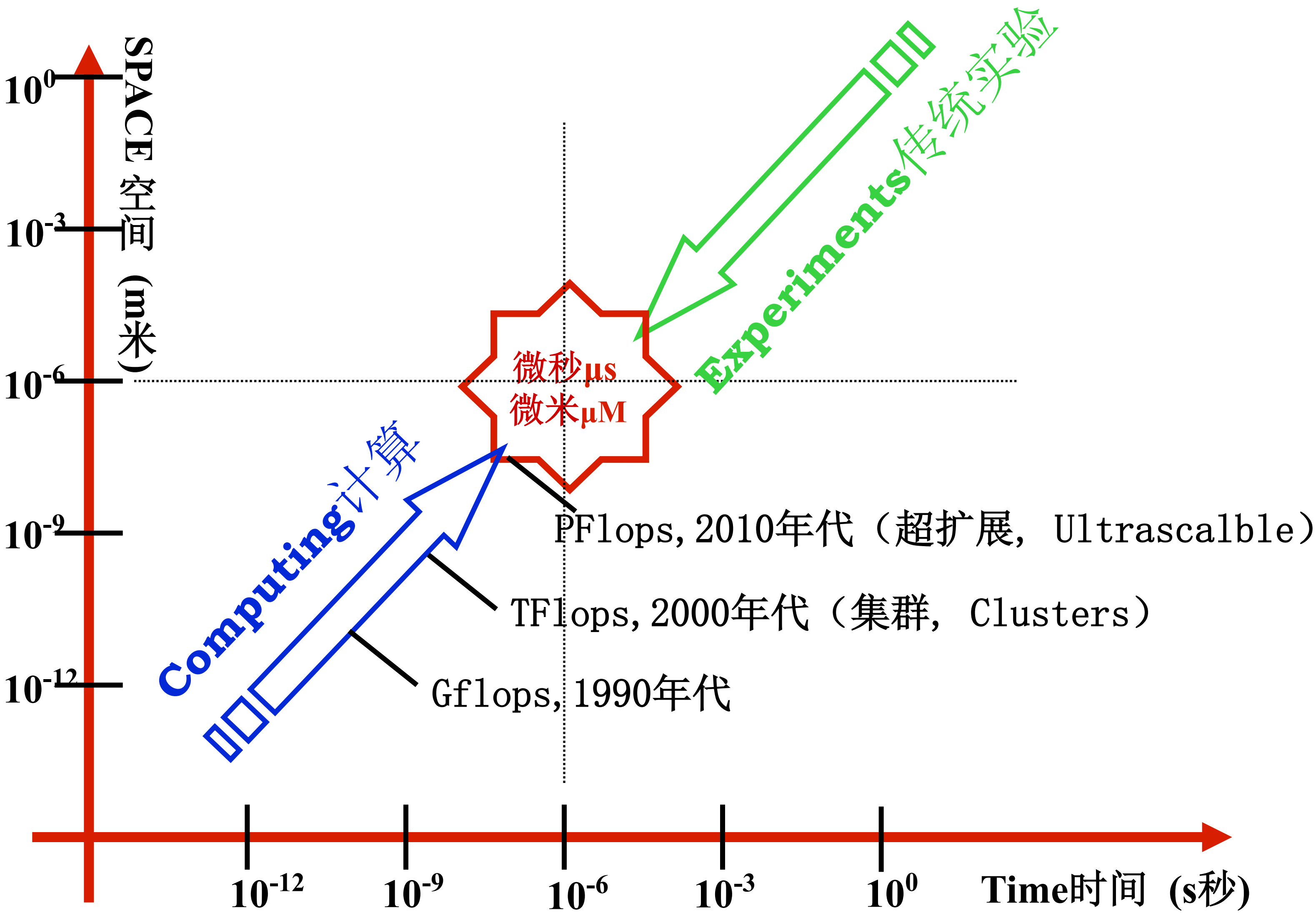
Fig. 11. MTS speedup for MD computations performed with fixed real space Ewald cutoff $R = 15$ Å and short-range cutoff, R_S, selected along the iso-error boundary in Fig. 8. Results were generated using our program MDoC.

G. Han, Y. Deng, J. Glimm, G. Martyna, *Error and timing analysis of multiple Time Stepping for MD*, Comp. Phys. Comm. 176 (2007) 271-291

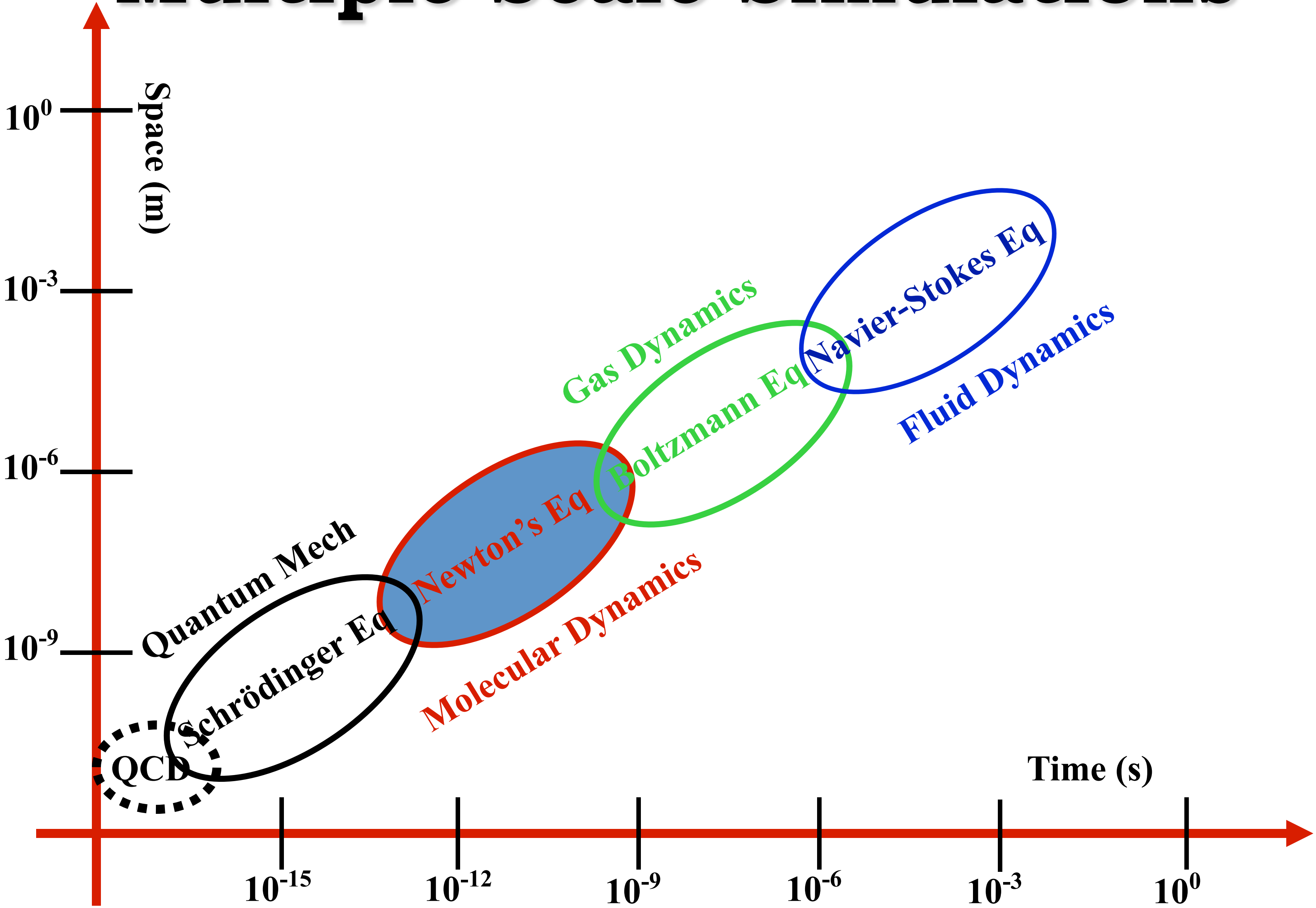
Key Application Drivers



Computing Vs. Experiments

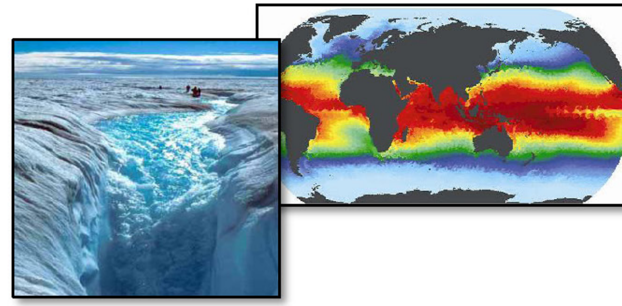


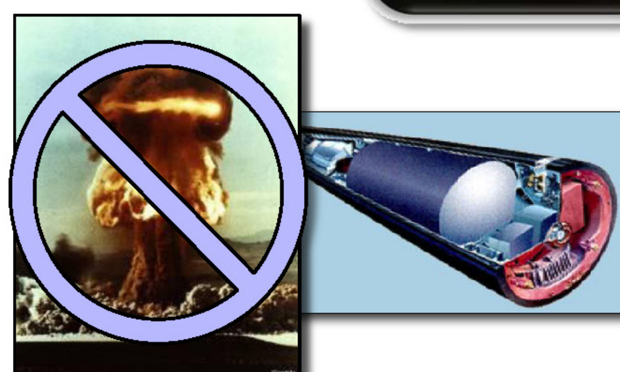
Multiple Scale Simulations



Key Application Drivers

- *Climate Change*: Understanding, mitigating and adapting to the effects of global warming
  - Sea level rise
  - Severe weather
  - Regional climate change
  - Geologic carbon sequestration
- *Energy*: Reducing U.S. reliance on foreign energy sources and reducing the carbon footprint of energy production
  - Reducing time and cost of reactor design and deployment
  - Improving the efficiency of combustion energy sources
- *National Nuclear Security*: Maintaining a safe, secure and reliable nuclear stockpile
  - Stockpile certification
  - Predictive scientific challenges
  - Real-time evaluation of urban nuclear detonation



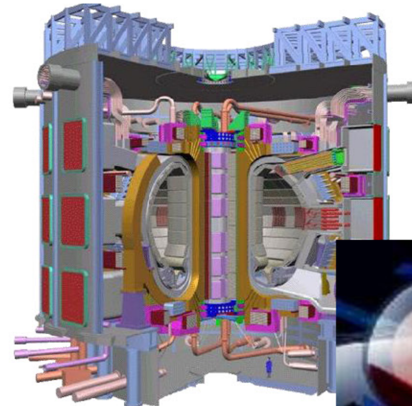


Accomplishing these missions requires exascale resources.

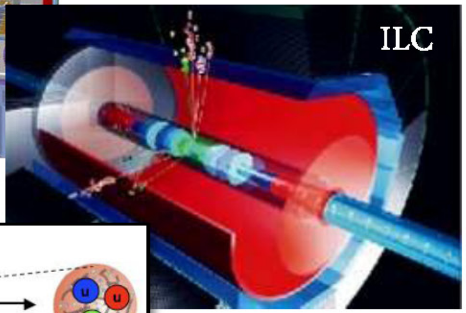
Key Application Drivers

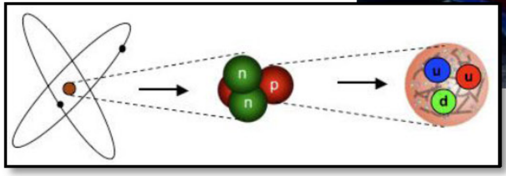
- **Nuclear Physics**
  - Quark-gluon plasma & nucleon structure
  - Fundamentals of fission and fusion reactions
- **Facility and experimental design**
  - Effective design of accelerators
  - Probes of dark energy and dark matter
  - ITER shot planning and device control
- **Materials / Chemistry**
  - Predictive multi-scale materials modeling: observation to control
  - Effective, commercial, renewable energy technologies, catalysts and batteries
- **Life Sciences**
  - Better biofuels
  - Sequence *to* structure *to* function

These breakthrough scientific discoveries and facilities require exascale applications and resources.

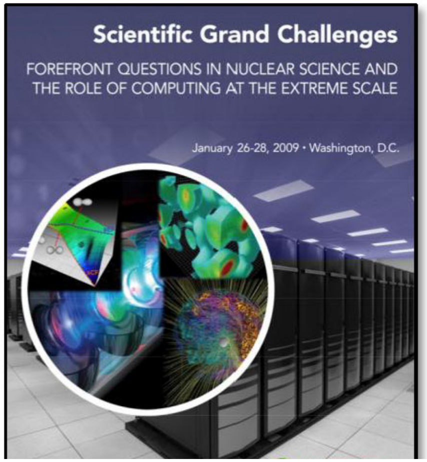


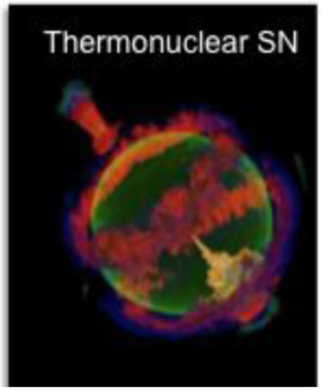
ITER





Structure of nucleons





Supercomputing Super-challenges

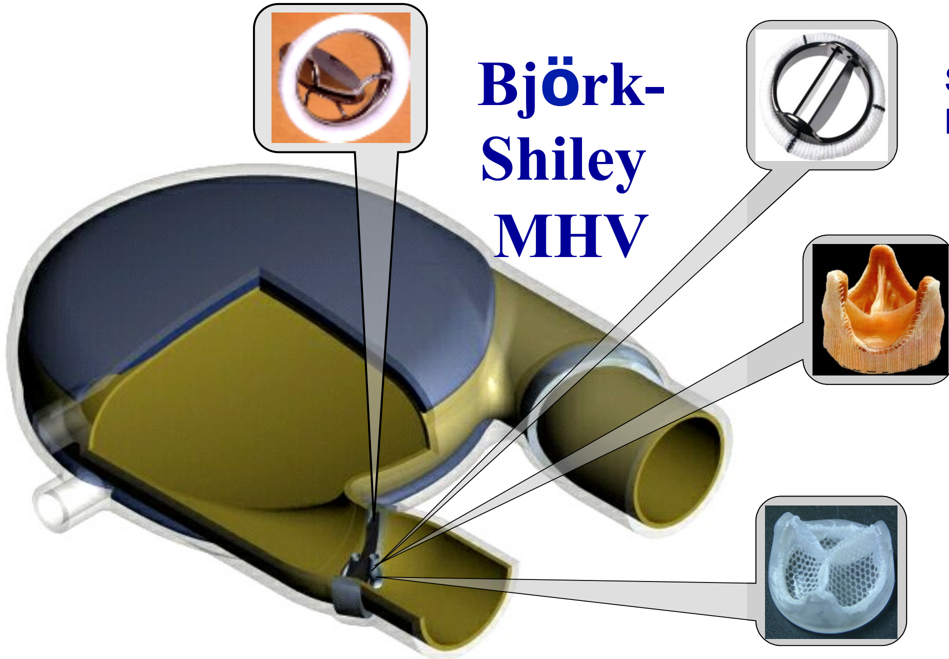
- **Architectures**: Complex memory accesses
- **Applications**: Massive and dynamical loads
- **Performance**: Sensitive to algorithms

- **Spatial resolution:** stresses memory
  - More mesh points and more precisions
- **Temporal resolution:** stresses processor speeds & processor-memory bw
- **Both spatial & temporal:** stress inter-processor latency and bandwidth

Application:
Platelet Activation

Platelet Activity Measurements in LVAD

LVAD – Prof. Affeld Humboldt University, Berlin - implantable part of a pneumatic heart-assist system



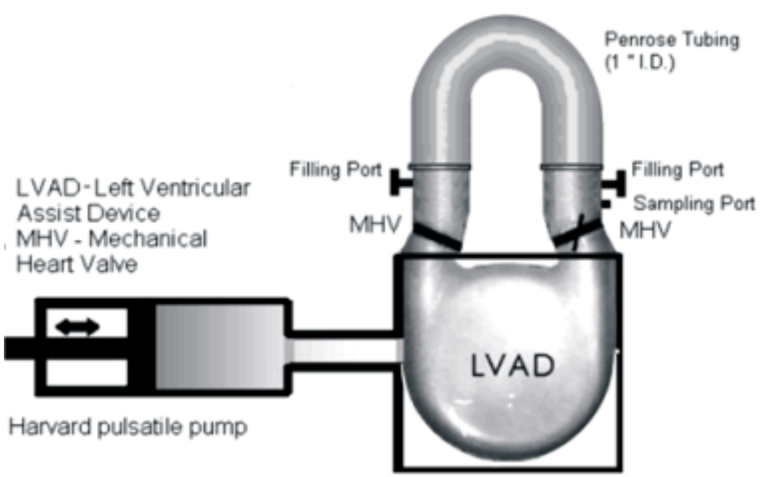
Björk-Shiley MHV

St. Jude bileaflet MHV

St. Jude tissue valve

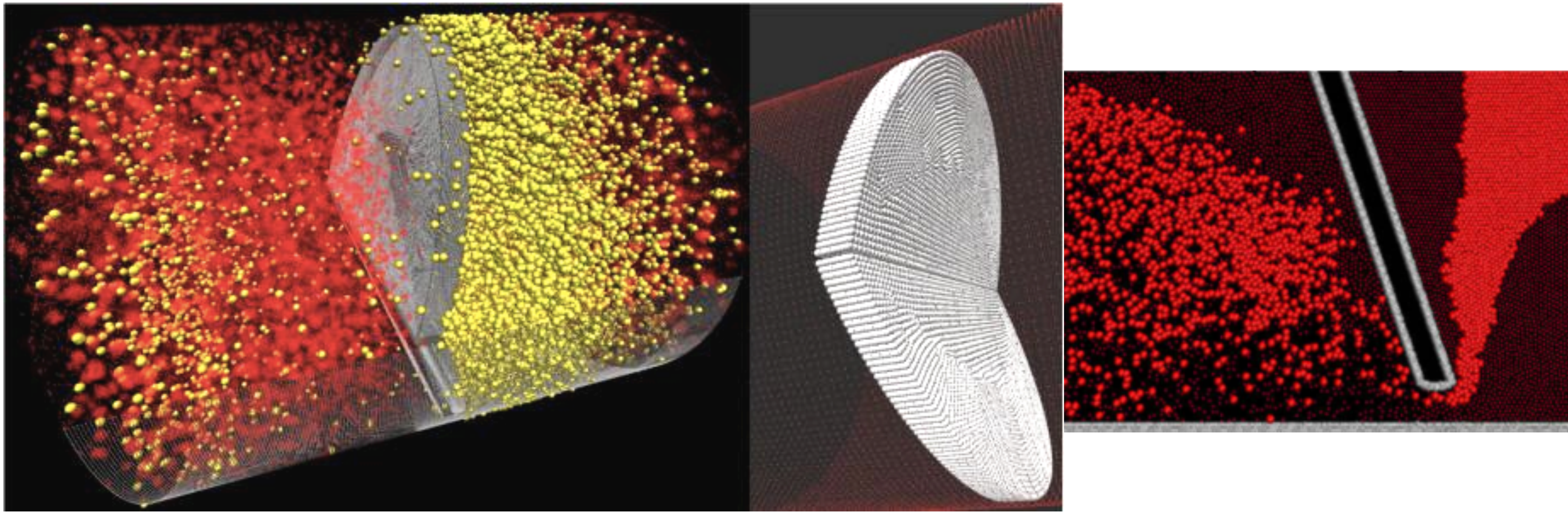
Tri-leaflet polymeric valve

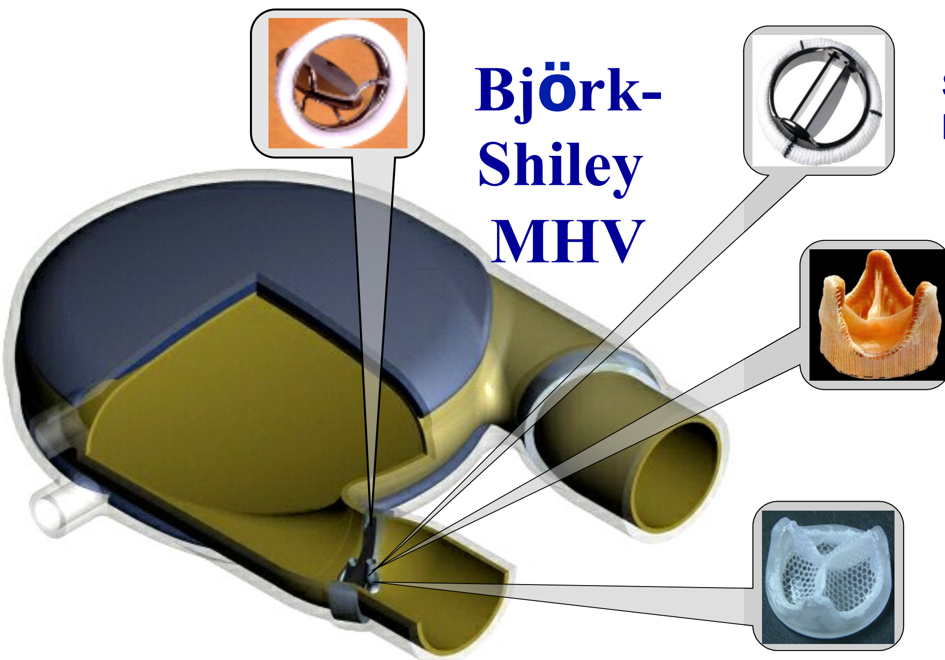
Harvard pulsatile blood pump: quasi-physiologic flow waveforms SV = 65 ml; 70 BPM



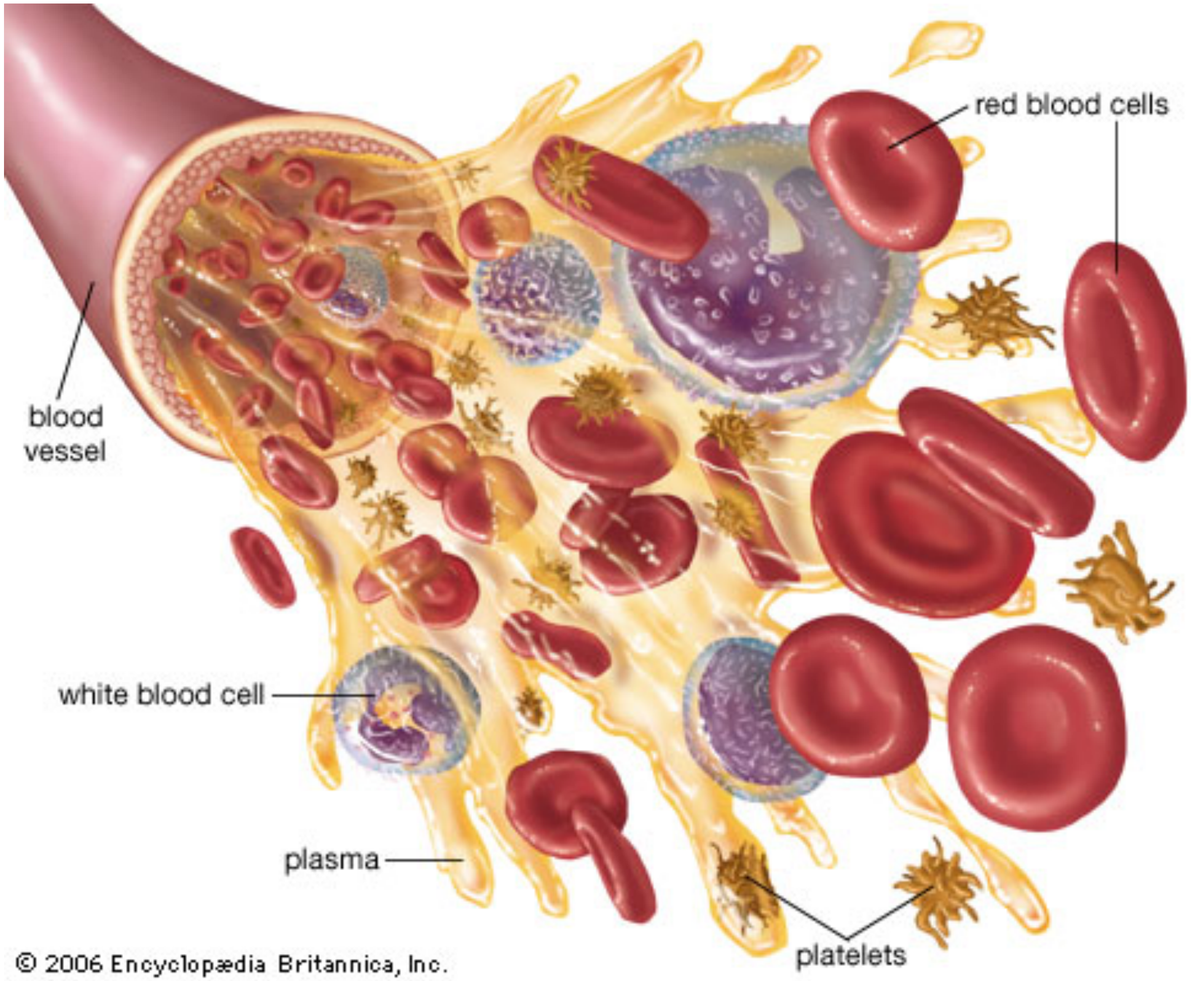
- PRP bags: platelets separated from plasma by gel filtration
- Recirculated in LVAD past MHV (incubator at 37°C)
- Timed aliquots removed
- Factor Xa, acetylated prothrombin, Ca^{++} ⇒ prothrombinase complex
- Samples assayed for thrombin generation rates (chromogenic thrombin substrate, using a microplate reader)

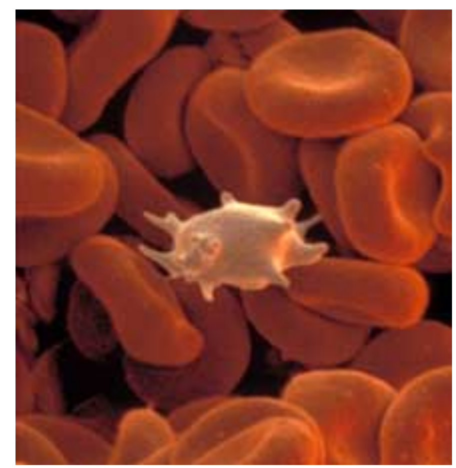
Artificial Heart Simulation Challenges





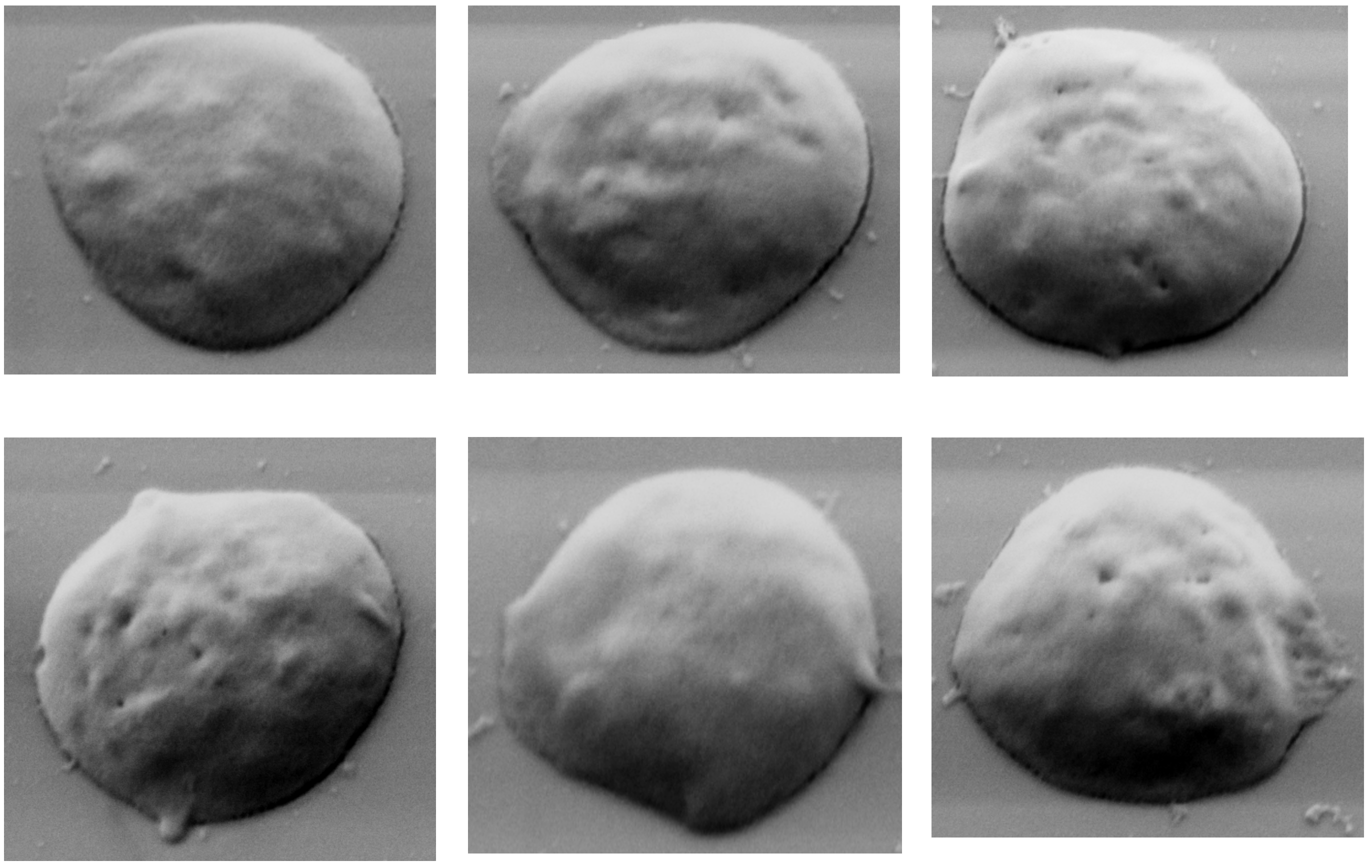
Blood Components



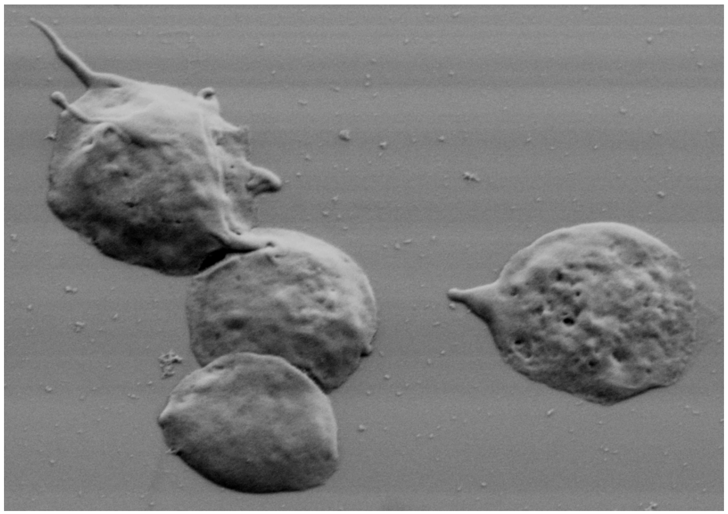


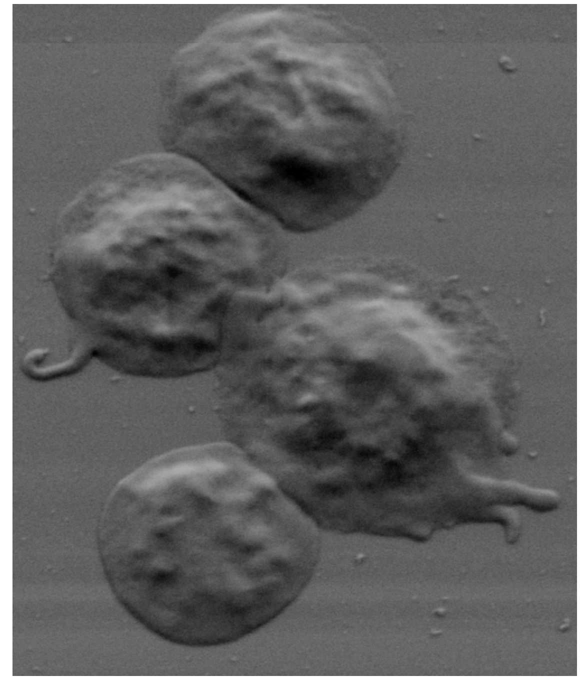
Blood is made up of multiple components, including red blood cells, white blood cells, platelets, and plasma. Source: Encyclopedia Britannica, Inc.

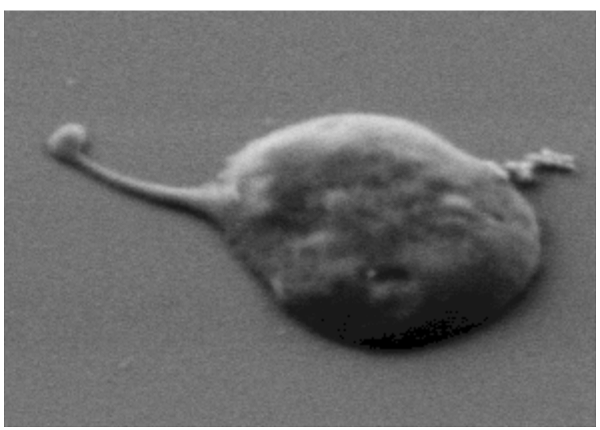
Unactivated Resting Discoid Platelets



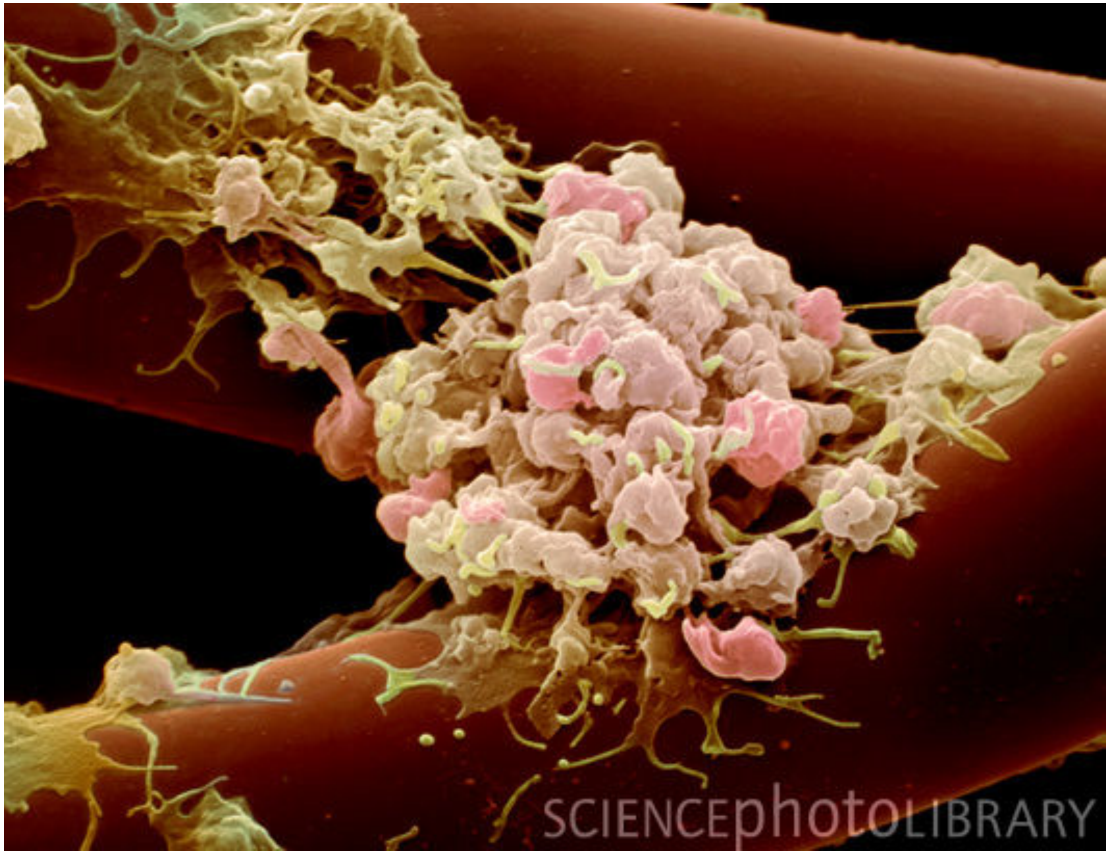
Partially Shear Activated Platelets

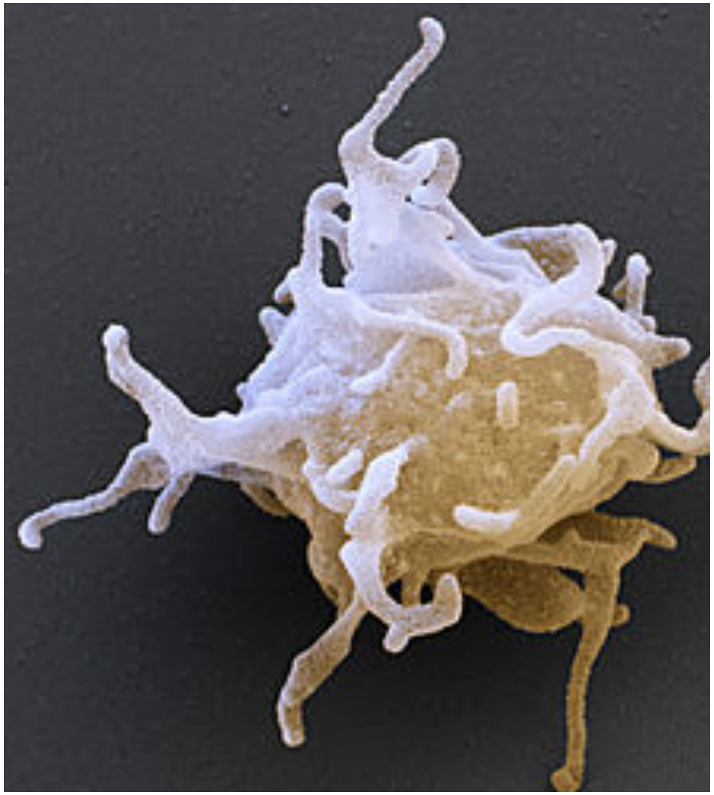






Fully Activated Platelets

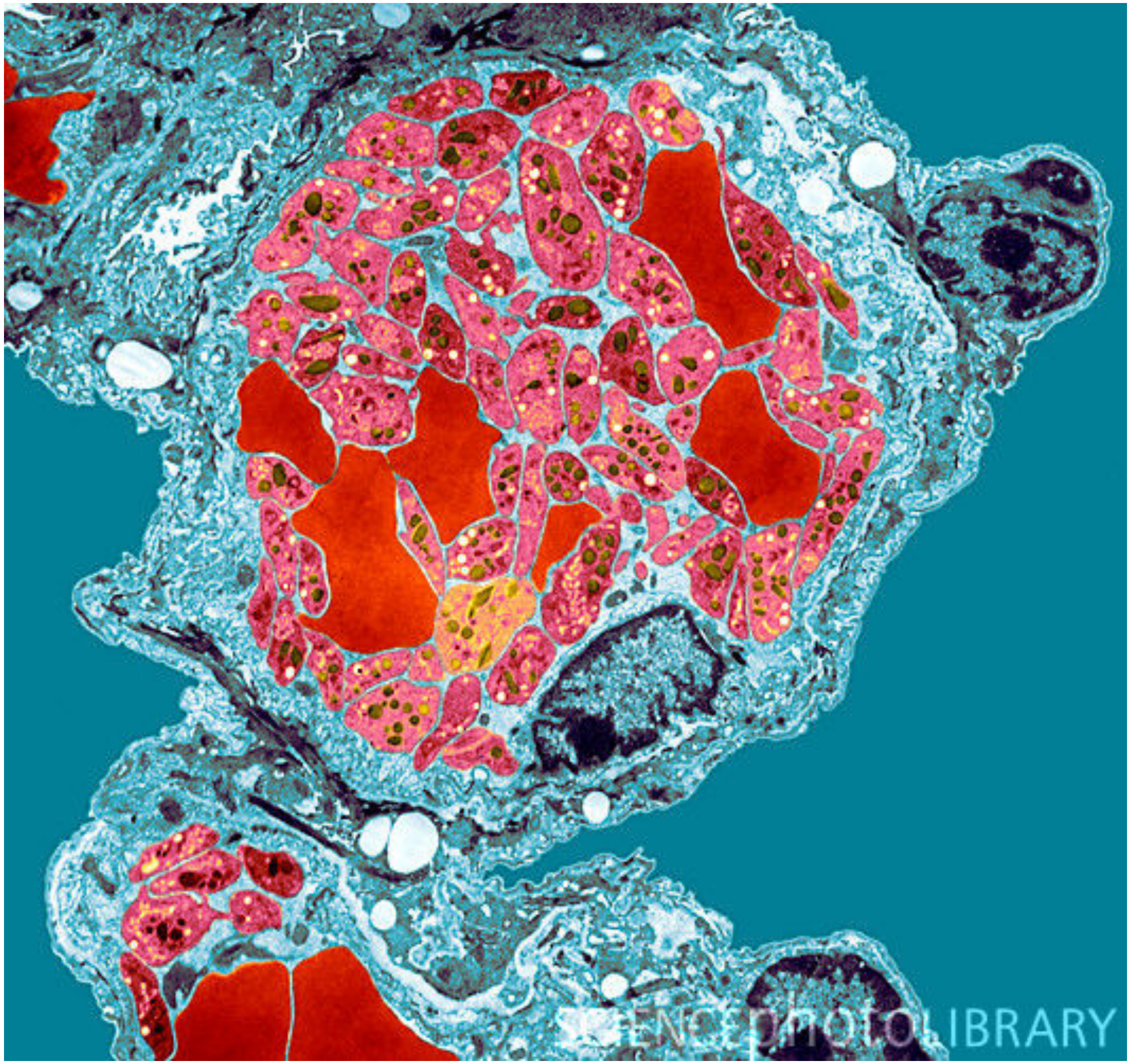




SEM photo. Platelet cells (thrombocytes) are formed in the bone marrow and circulate in the blood in large numbers. In an unactivated state they are round/oval, whereas this activated platelet has developed extensions from the cell wall known as pseudopodia. Platelets function in two ways: they plug defects in the walls of blood vessels, and are involved in clotting. They also release serotonin which constricts blood vessels.

Source: Science Photo Library

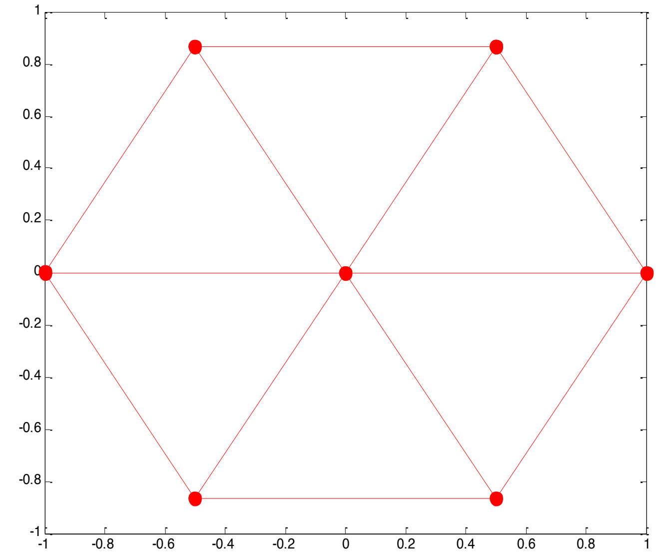
A Cross Section of Activated Platelet

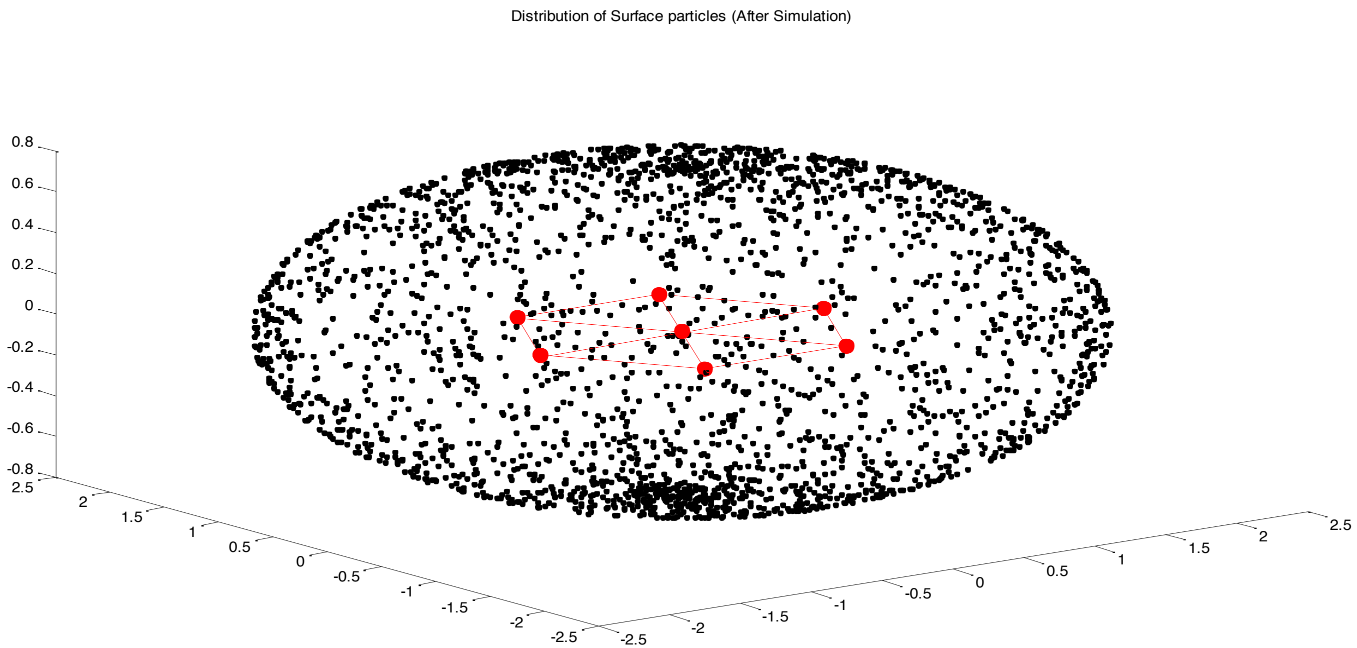


TEM of a section through an activated blood platelet. Platelets plug defects in the walls of blood vessels, and are involved in clotting. They also release serotonin which constricts blood vessels. Platelet cells (thrombocytes) are formed in the bone marrow, and circulate in blood in large numbers. In an unactivated state they are round/oval, whereas activated platelets develop pseudopodia or extensions from the cell wall. Platelets contain cytoplasmic organelles such as mitochondria, endoplasmic reticulum and granules.

Source: Science Photo Library

A Particle-based Model



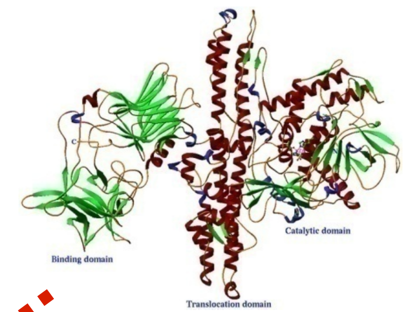


Modeling Botulinum

X. Chen and Y. Deng, *Botulinum structures at various temperatures and pH values*, J. Mol. Modeling, 13 (5) (2007) 559-572

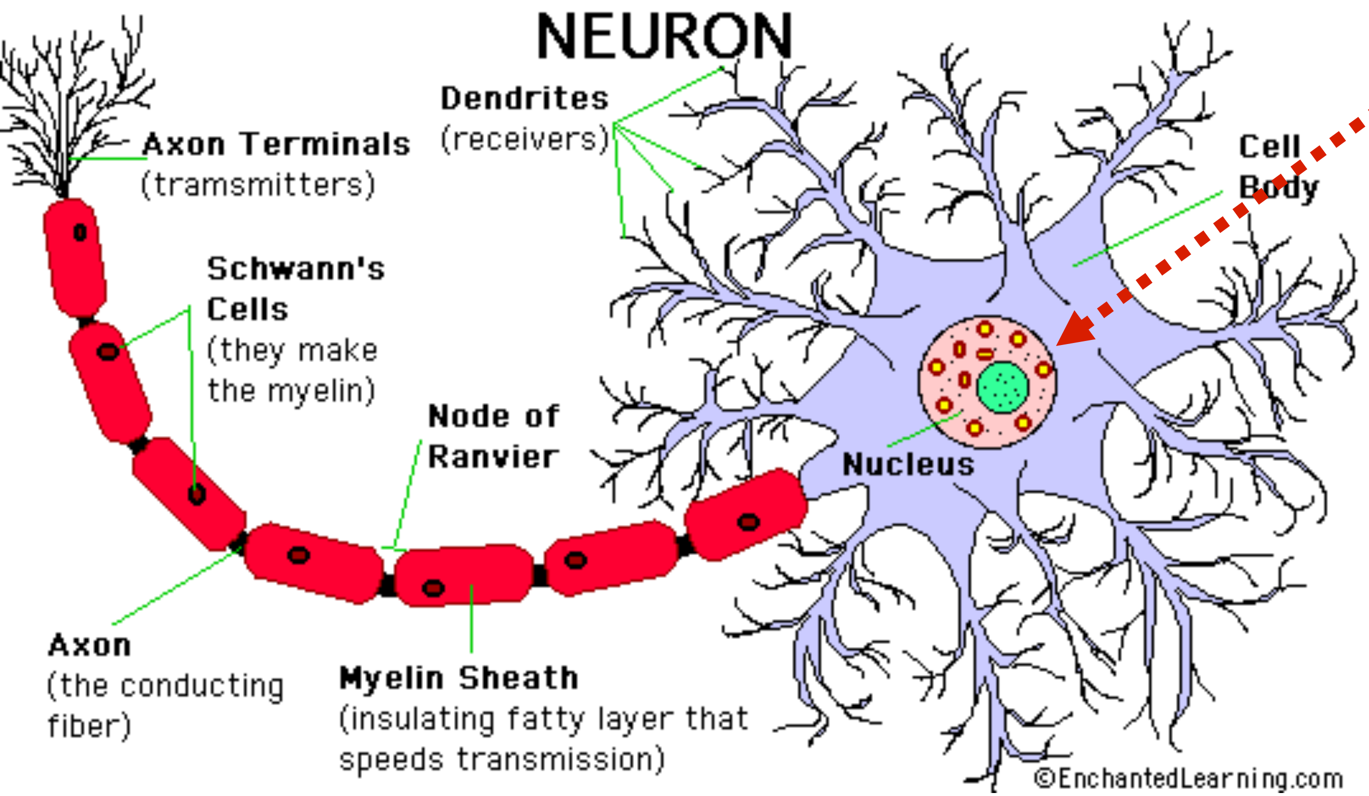
Y. Chen, X. Chen, and Y. Deng, *Simulating Botulinum Neurotoxin with Constant pH Molecular Dynamics in Generalized Born Implicit Solvent,* Comp. Phys. Comm. 177 (2007) 210-213

Botulinum Toxicity "Theory"

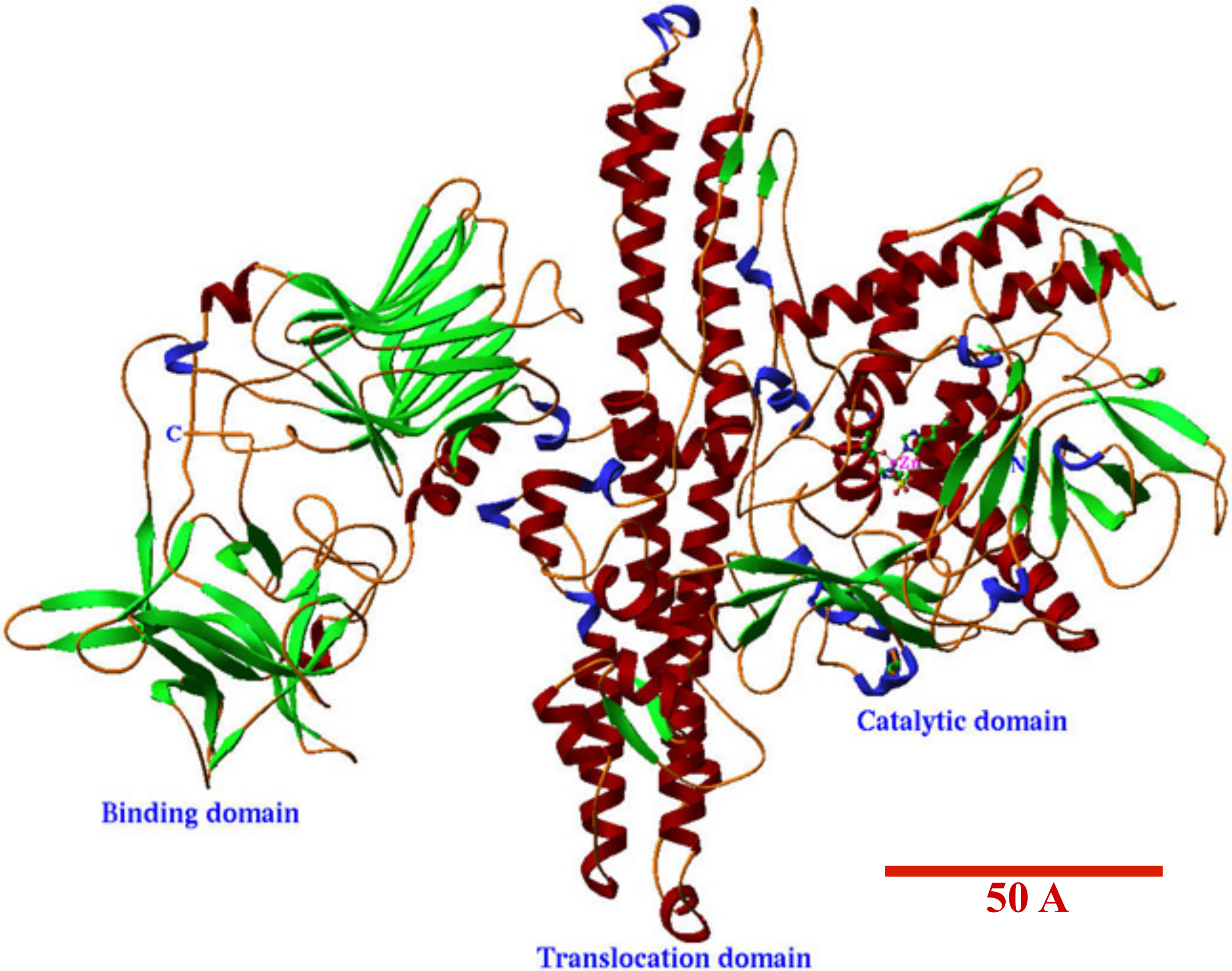


1 µM

1/50 µM



Botulinum



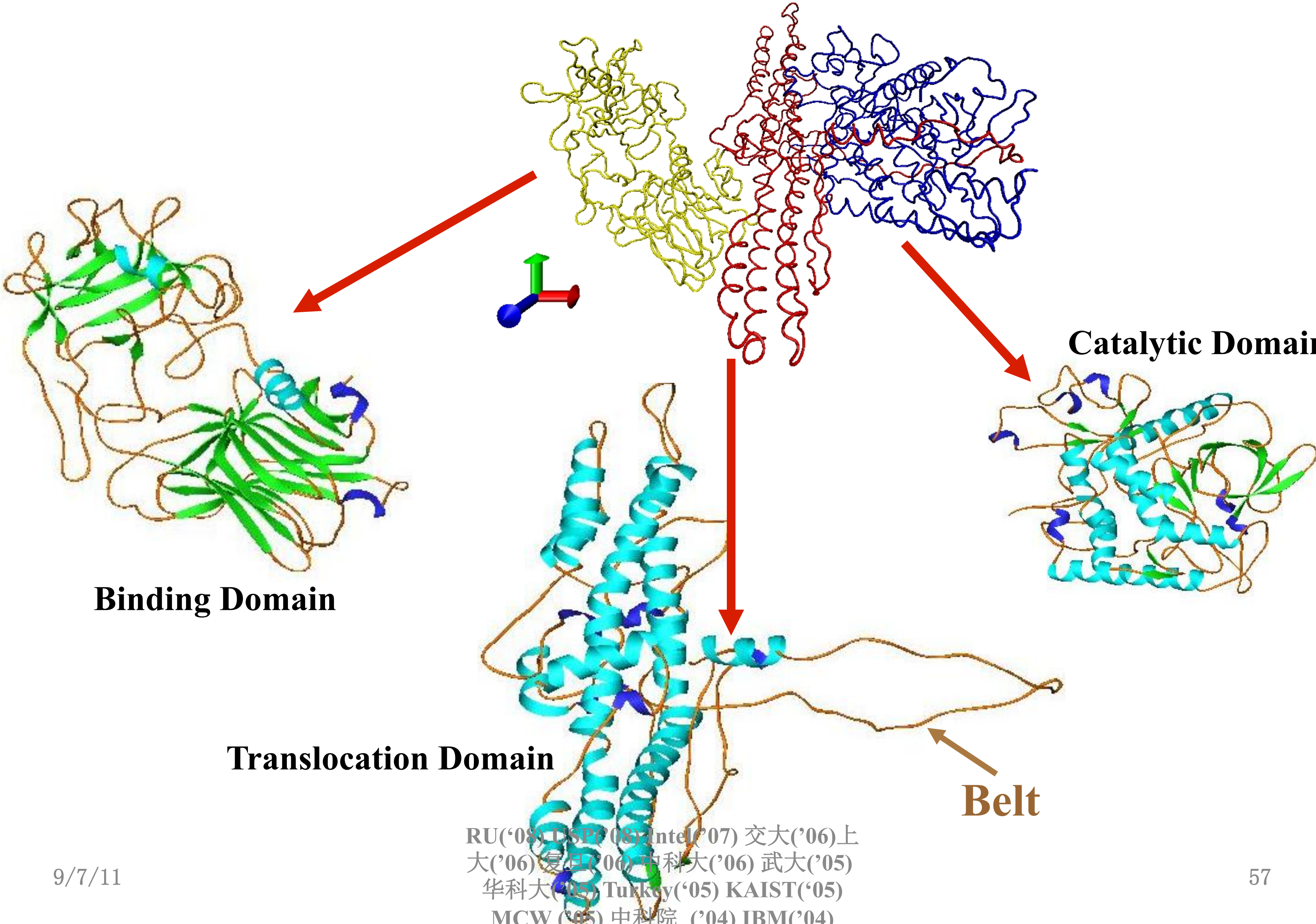
ce: S. Swaminathan

A Ribbons representation of Clostridium botulinum neurotoxin B (BoNT-B). Helices in blue represent 3-10 helices. The three functional domains are labeled. Zinc and the coordinating residues in the catalytic domain are shown in ball and stick representation.

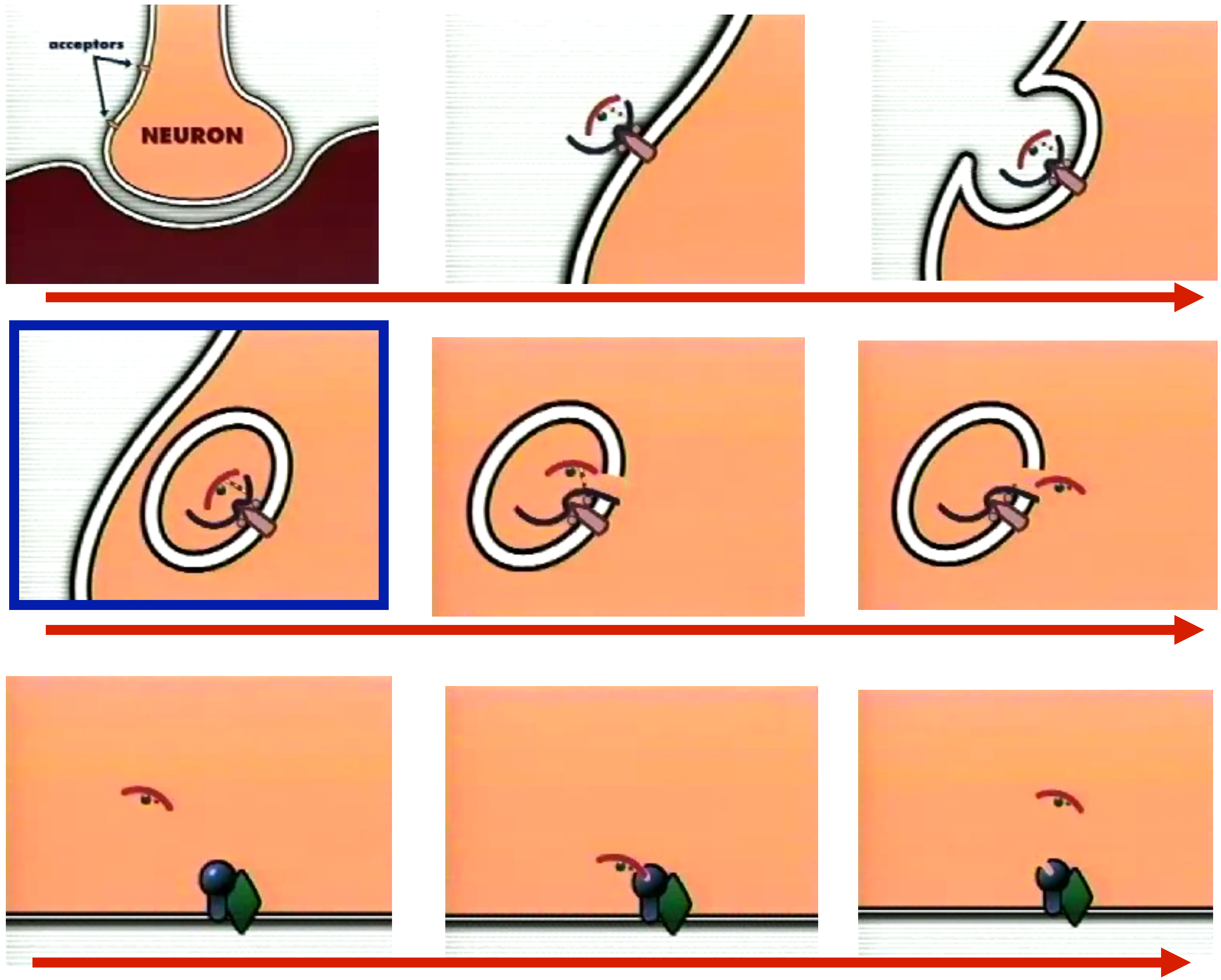
A few more facts:

(1) Three domains: Catalytic, Transloc., and Binding
(2) Two Chains: Light (50 kDa) and Heavy (100 kDa) connected by disulfide bond
(3) 1277 Residues
(4) 20,000+ atoms

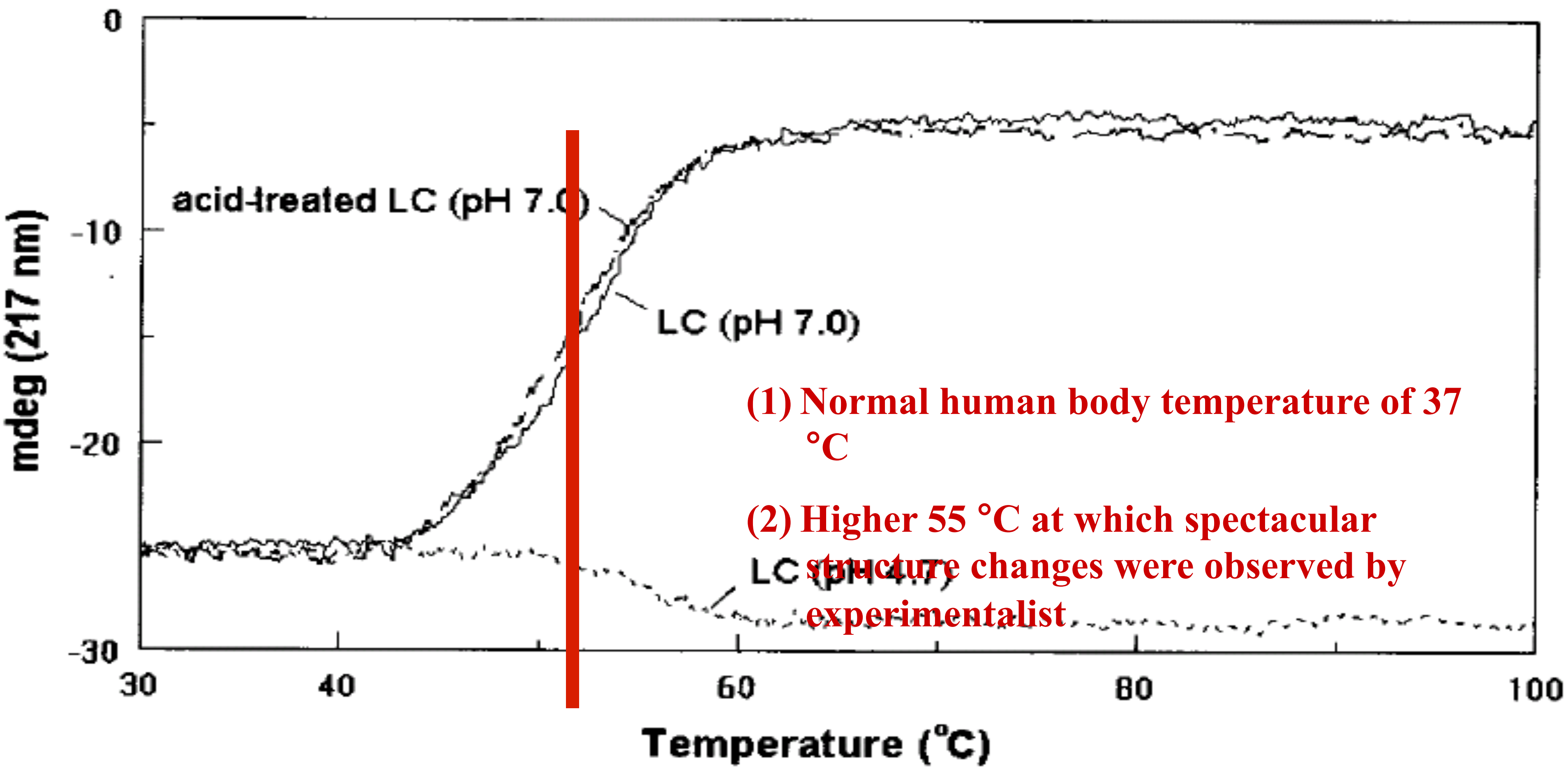
Three Domains of Botulinum



Botulinum Reaction Pathway



Simulating Botulinum in Different pH



Source: L. Li and B. Sigh: Biochem. V. 39 (6466-74)

Botulinum Modeling Parameters

Table 3 The difference between the systems used for the simulations of BoNT/A at high and low pH values

	Neutral pH (pH 7)	Acid pH (pH 4.7)
Histidine side chains	Not protonated	Protonated
Overall protein charge	−9	+3
Zinc ion	Present	Present
Bounding box	106×120×158	106×120×158
Volume ($Å^3$)	2,023,840	2,023,840
Total number of atoms	173,549	173,561
Protein atoms	20,698+Zn^{2+}	20,710+Zn^{2+}
Water molecules	50,950	50,950
Total mass	1,065,388 Da	1,065,400 Da
Total density	0.874 g/cc	0.874 g/cc

J Mol Model (2007) 13:559–572

Table 1 Summary of parameters for all numerical experiments

Temperature	pH values	
	pH 4.7	pH 7.0
37 °C	Whole protein (~64 ns) LC only (200 ns)	Whole protein (~63 ns) LC only (200 ns)
55 °C	Whole protein (~57 ns) LC only (200 ns)	Whole protein (~67 ns) LC only (200 ns)

Table 4 Simulation time needed

Name	No. of residues	Simulation time
Villin headpiece (alpha helical protein)	36	~10 ms 35/100 has conformational change in 1 ms simulation
Trp-cage	20	~100 ns at 315 K
Beta hairpin	54	38 ms at 300 K
BoNT/A	1,277	Unknown
BoNT/A LC only	431	Unknown

Table 5 Differences between the systems used for the simulations of LC for pH 4.7 at 37 °C

	Cut-off run	Whole protein run
Total atoms	59,672	173,561
Protein atoms	6,958+Zn^{2+}	20,710+Zn^{2+}
Water molecules	17,571	50,950

Thank You for Your Attention