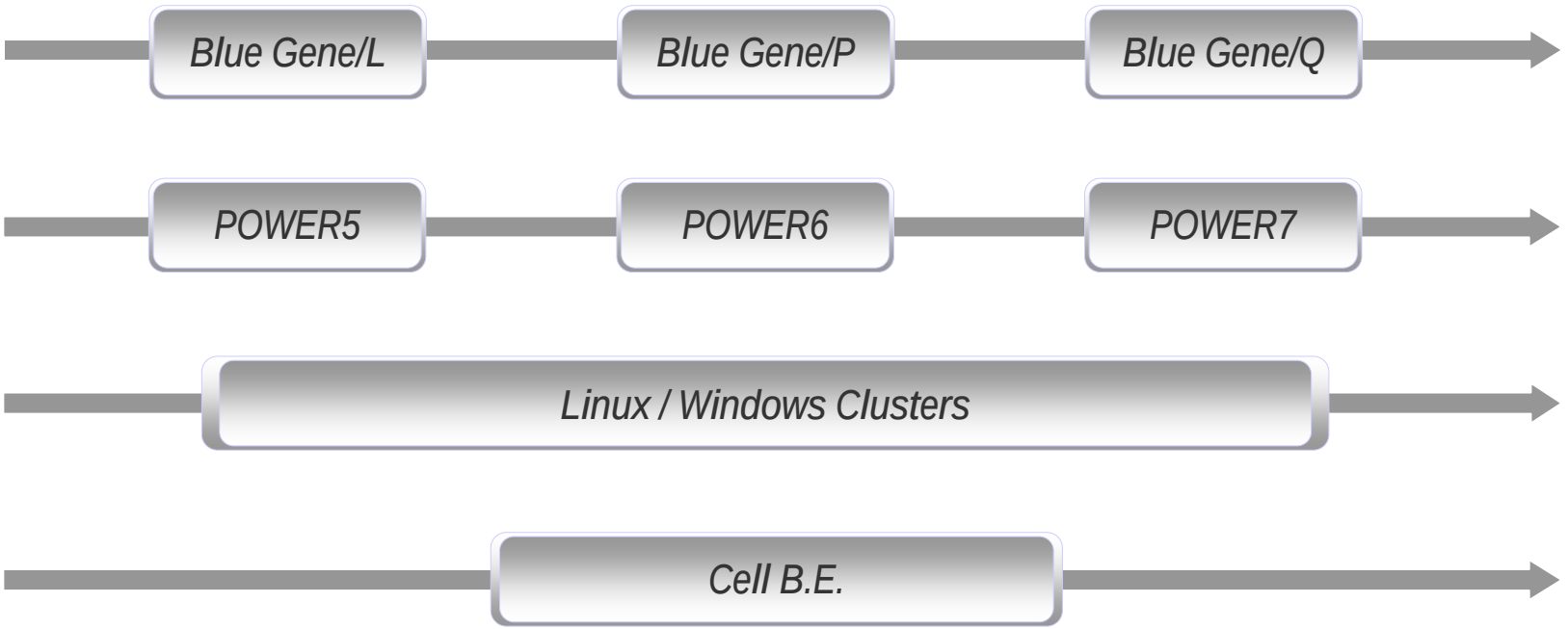
PetaFLOPS and Megawatts: The 'Green' Blue Gene/P

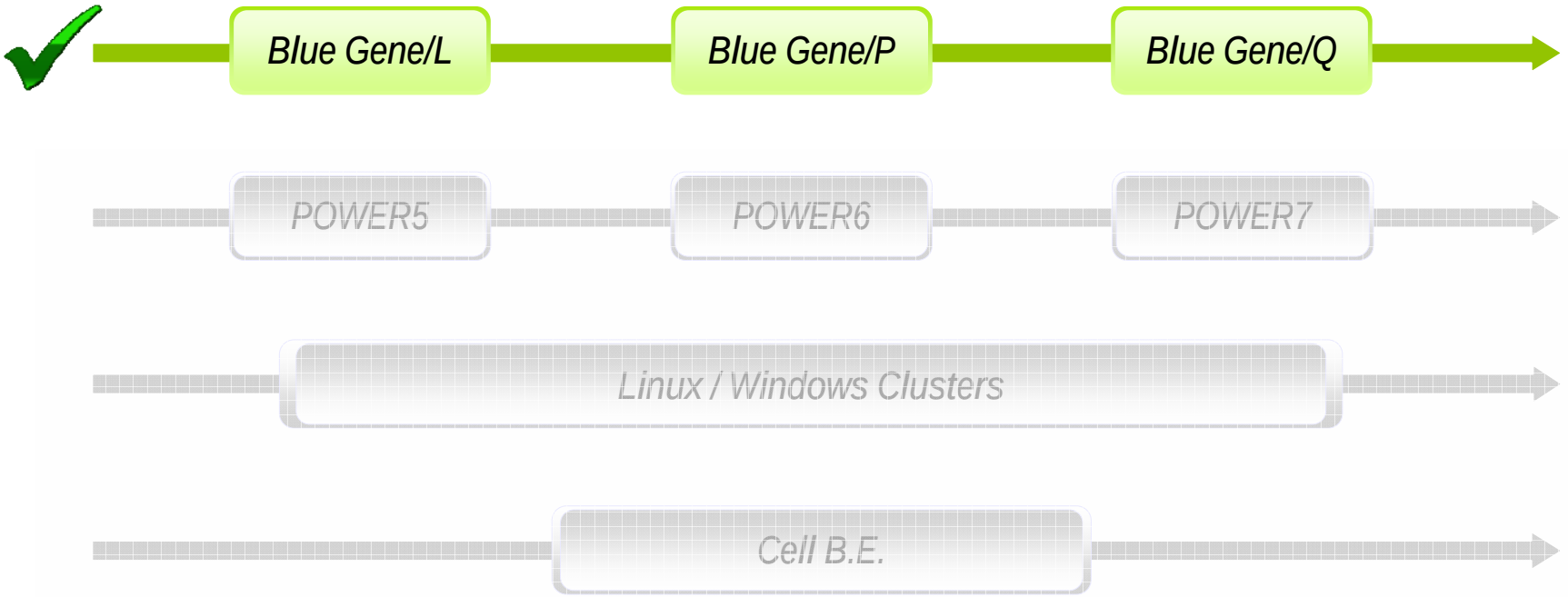
Patrick Carey
Blue Gene, IBM Worldwide

IBM Green Supercomputing Roadmap



IBM Green Supercomputing Roadmap

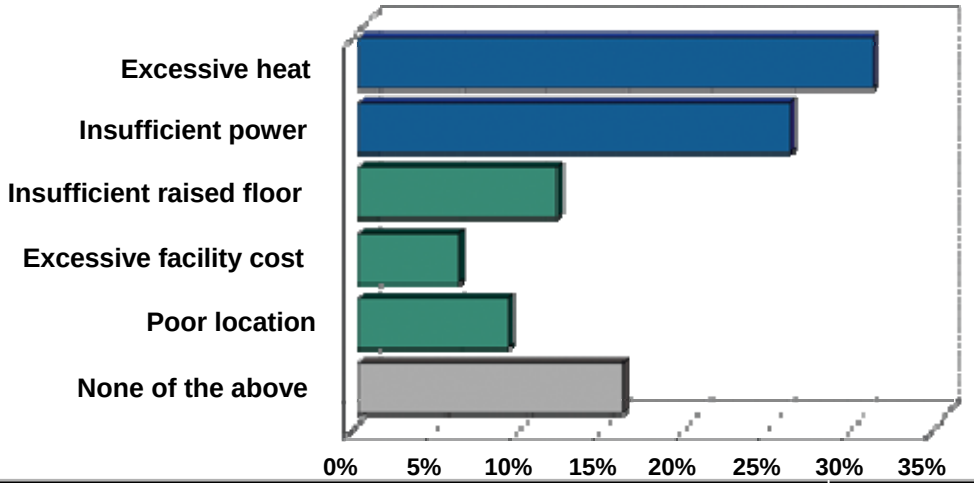
What we cover here…



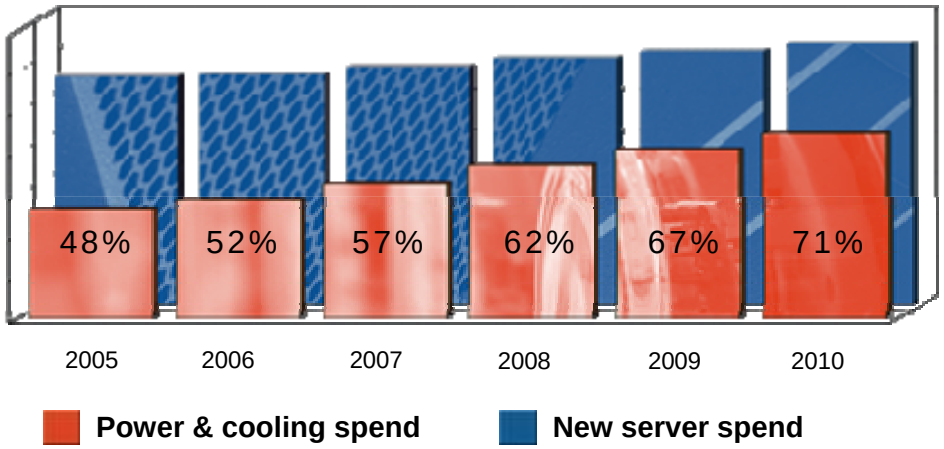
Blue Gene is the Leader in Green Supercomputing

- Low Power Chips
- Highest FLOPs/per Watt
- Highest Scalability of any platform
- Solving the world's largest Research Challenges

What is the greatest facilities problem with your primary data center? Gartner 2006



Expense to power and cool installed base ($US)

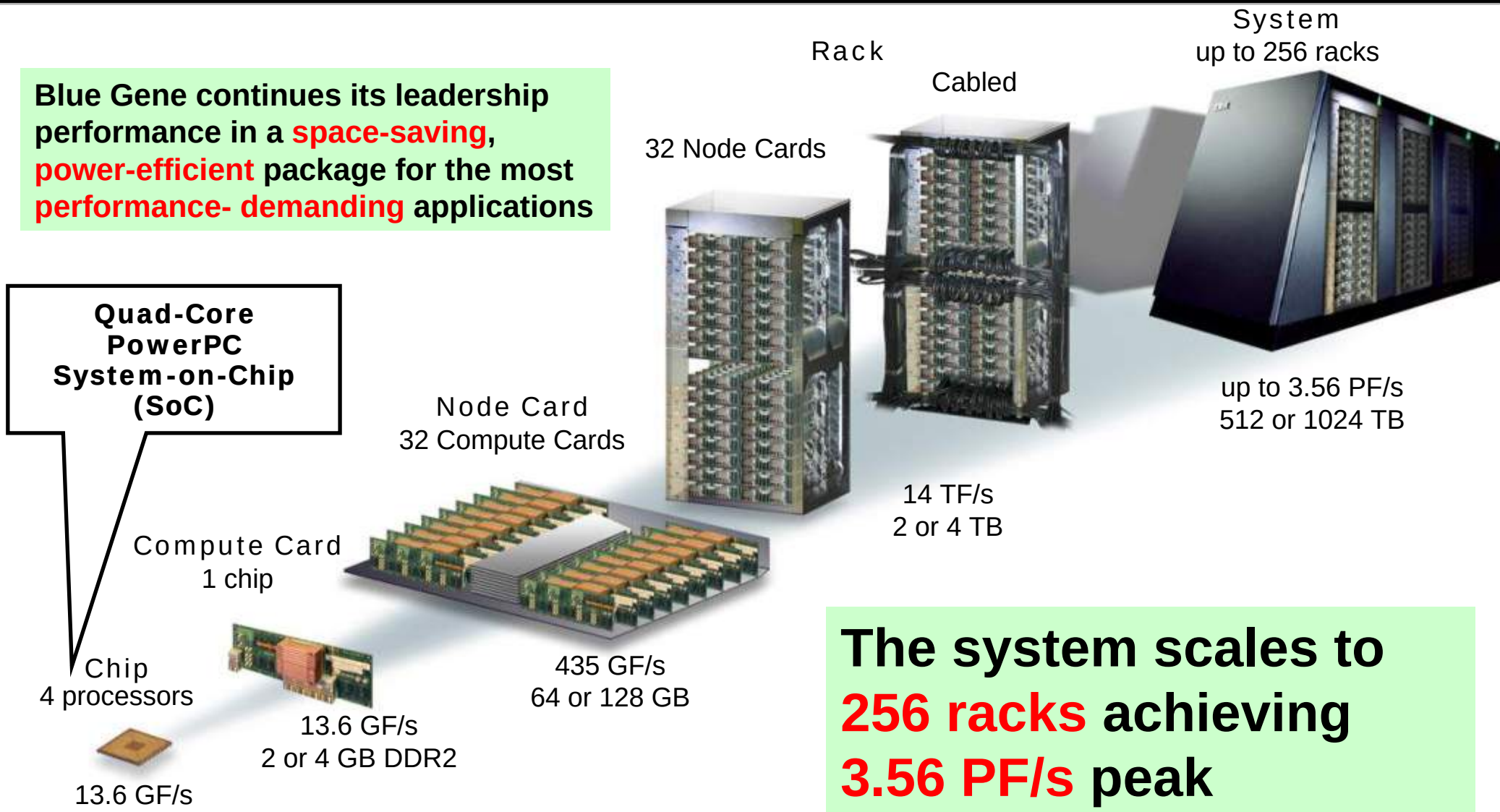


The Blue Gene Approach – A Path Finder to Petascale Supercomputing

- **Design Driven by Integration, Power, and Technology Considerations:**
  - Multiple modest cores running at reduced voltage and frequency ($P = CV^2F$) on a single chip rather than one high-performance processor.
    - Single system-on-chip nodes reduce system complexity
    - System-on-chip reduces power as interconnects are shortened and chip crossings removed.
    - Embedded DRAM shortens latency, reduces application stalling, simplifies design.
    - Enables high density, further reducing interconnect distances and saving facility cost.
- **Stress Simplicity**
  - Choose the right areas to innovate (networks, memory system, clocking, power regulation and distribution)
  - Note: There is NO exotic or bleeding edge technology in Blue Gene
- **Enable familiar, standard programming model and mature compiler support.**
  - Linux environment, MPI (standard messaging interface), and OpenMP (standard shared memory parallelism)
- **Close attention to RAS (reliability, availability, and serviceability) at all system levels.**
  - One of the biggest challenges
  - Infused throughout Blue Gene design

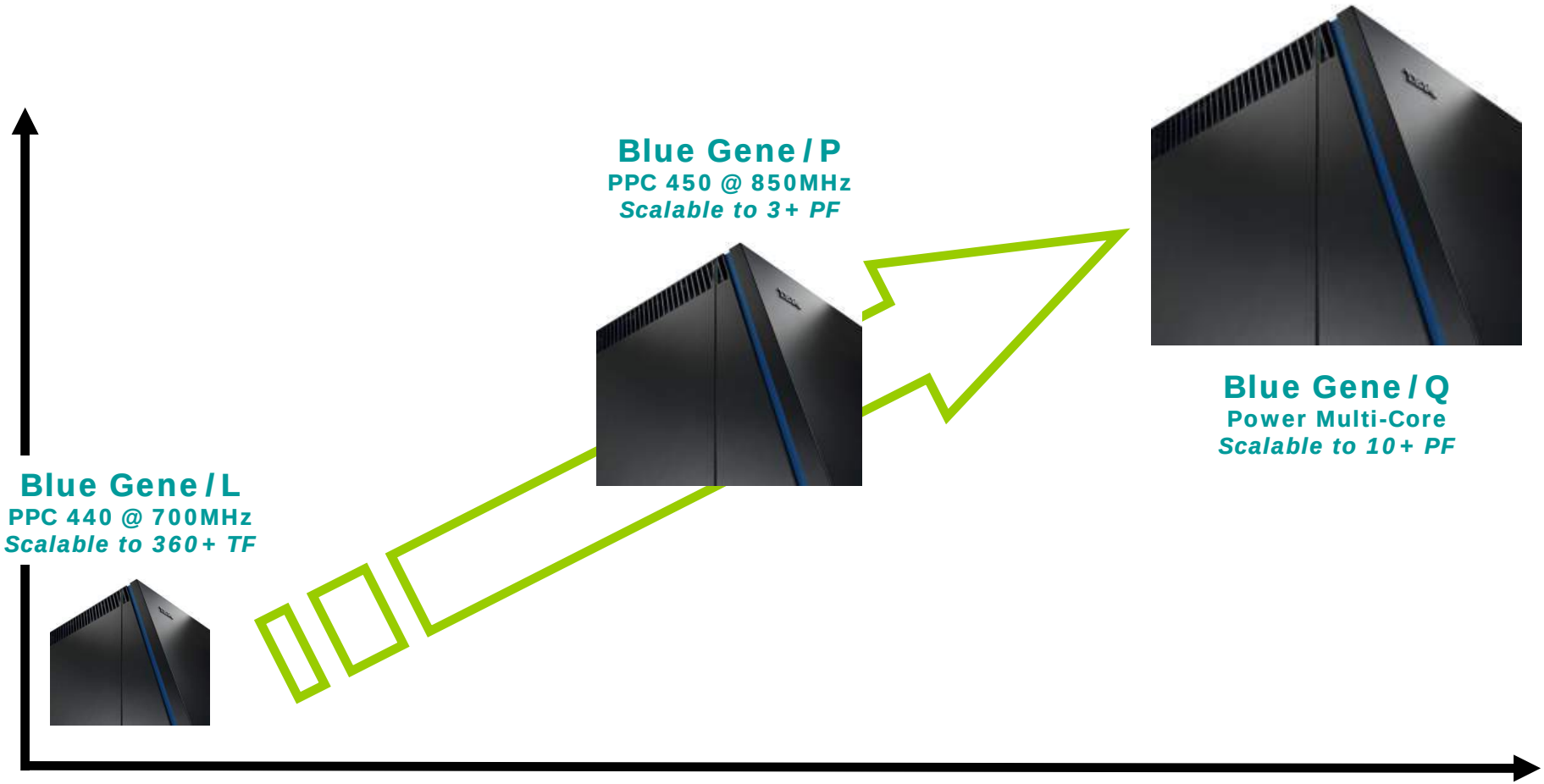
What is Blue Gene/P?

Blue Gene continues its leadership performance in a space-saving, power-efficient package for the most performance- demanding applications



The system scales to 256 racks achieving 3.56 PF/s peak

Blue Gene technology roadmap:



Blue Gene is #1 in the 30th TOP500 list (12 November, 2007)

4 BG entries in the TOP 10
15 Entries in the 20 Teraflop Club

Source: www.top500.org

Top500	Installation	Processor	Rmax TF/s
1	DOE BlueGene/L LLNL	700 MHz PPC 440 (104 racks)	478.2
2	FZJ Juelich Blue Gene/P	850 MHz PPC 450 (16 racks)	167.3
8	BlueGene at Watson	700 MHz PPC 440 (20 racks)	91.29
10	Stony Brook/BNL BG/L	700 MHz PPC 440 (18 racks)	82.16
11	ASC Purple LLNL	1.9 GHz POWER5 p5 575	75.76
12	RPI Blue Gene/L	700 MHz PPC 440 (16 racks)	73.03
13	Barcelona SC	2.3 GHz PPC970 JS21	63.83
24	IBM Rochester Blue Gene/P	850 MHz PPC 450 (4 racks)	43.16
28	FZ J- Juelich Blue Gene/L	700 MHz PPC 440 (8 racks)	37.33
37	Univ Groningen (ASTRON)	700 MHz PPC 440 (6 racks)	27.45
40	Max Planck/IPP Blue Gene/P	850 MHz PPC 450 (2 racks)	21.91
41	Oak Ridge NL Blue Gene/P	850 MHz PPC 450 (2 racks)	21.91
42	Indiana University	2.5 GHz PPC 970 JS21	21.79
45	NAVO	p5 575 1.9 GHz	20.07
46	IBM Adv Client Tech Center	p5 575 1.9 GHz	20.07

IBM Blue Gene

Other IBM platforms

Blue Gene's power efficiency is first rate

BG requires 75-80% less power and space than COTS clusters

Blue Gene took 26 of the top 26 spots

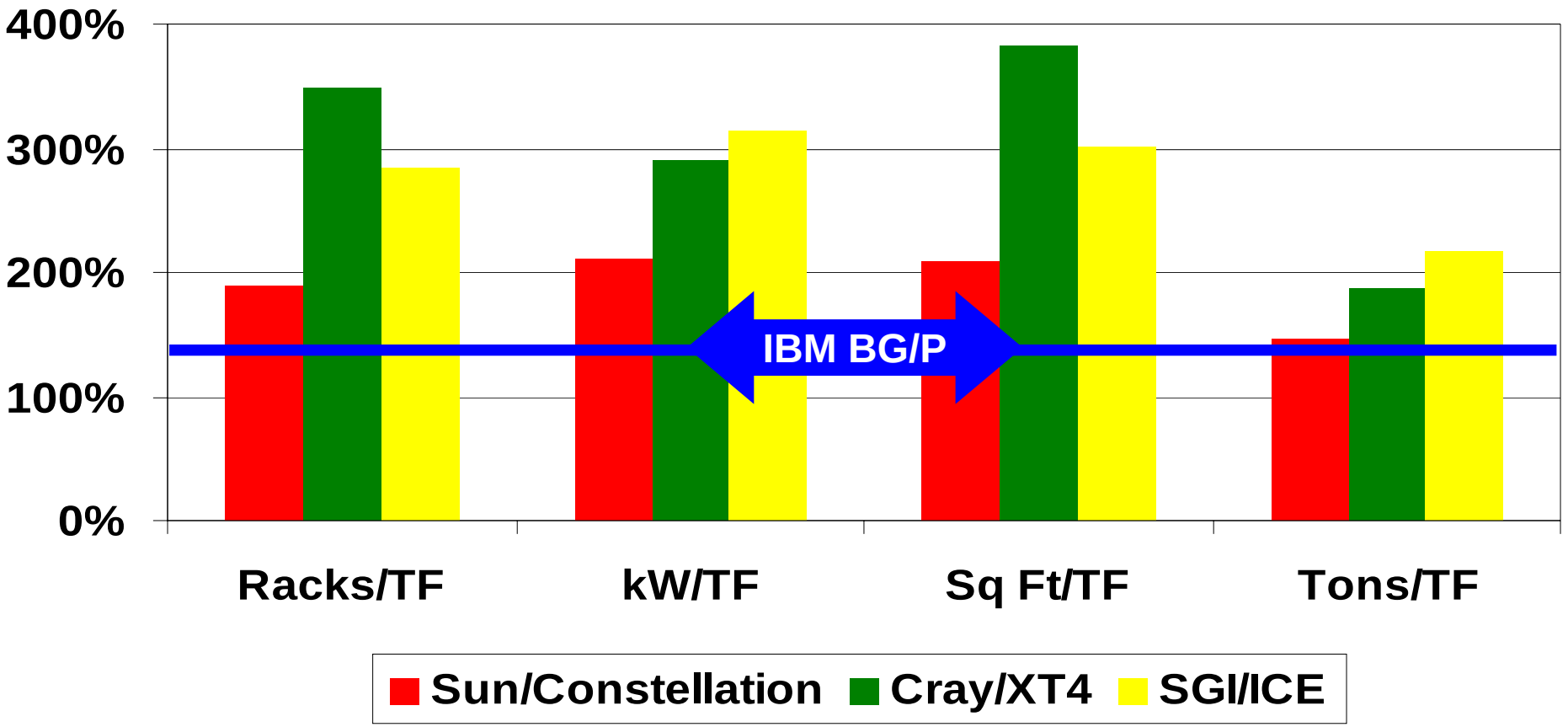
in the February, 2008 **Green500** list

Green500 www.green500.org	ENERGYGUIDE Compare the Energy Use of this Computer with Others Before You Buy.
1	Science and Technology Facilities Council
2	Max-Planck-Gesellschaft MPI/IPP
3	IBM - Rochester
4	Forschungszentrum Juelich (FZJ)
5	Oak Ridge National Laboratory

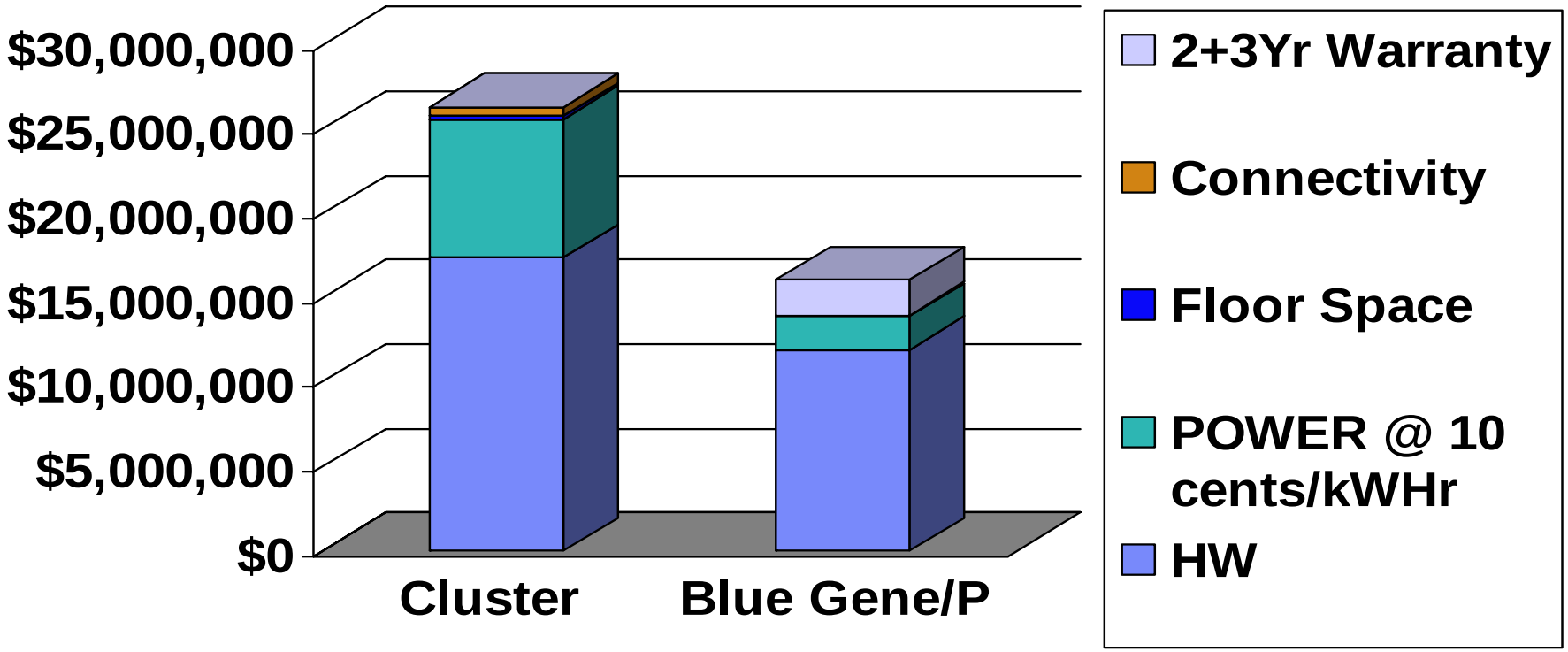
Computational efficiency is moving from sustained-to-peak (aka MPH/HP) to performance-per-watt (aka MPG)

John Shalf, NERSC/LBNL, "The Landscape of Computer Architecture" presented at ISC07, Dresden, June 2007

IBM Blue Gene/P is the most power, space, and cooling efficient supercomputer



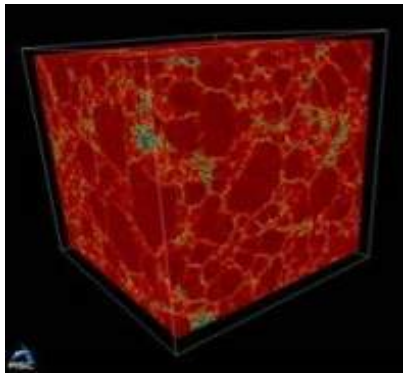
3 Year Comparison "Typical" Cluster vs. Blue Gene/P TCO for 100 Linpack TeraFLOPs



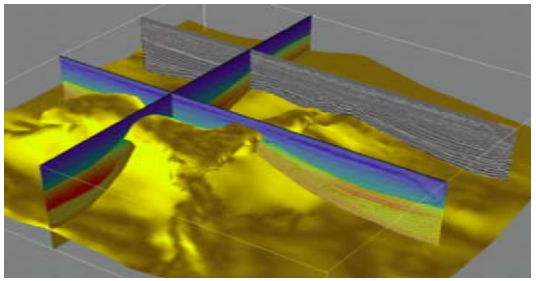
Where in the world is Blue Gene?

Why and for What is Blue Gene used?

- Improve understanding **– significantly larger scale, more complex and higher resolution models; new science applications**
- Multiscale and multiphysics **– From atoms to mega-structures; coupled applications**
- Shorter time to solution **– Answers from months to minutes**



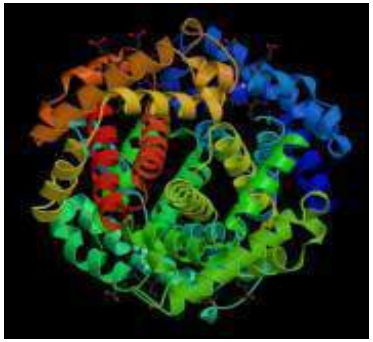
Physics – Materials Science
Molecular Dynamics



Geophysical Data Processing
Upstream Petroleum

Biological
Modeling – Brain Science

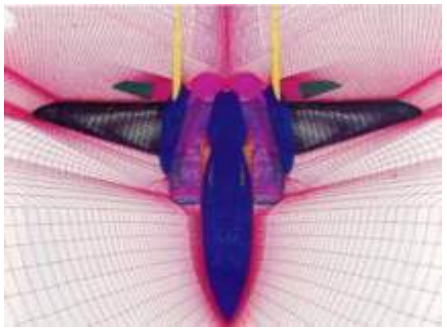
Life Sciences: In-Silico
Trials, Drug Discovery

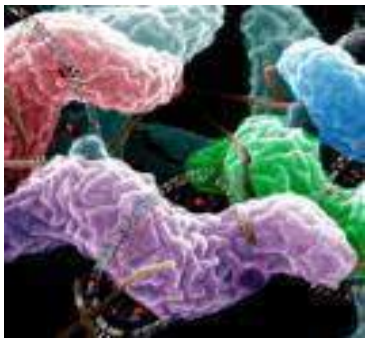


Financial Modeling
Streaming Data Analysis

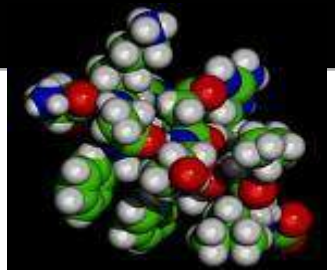
Environment and Climate Modeling

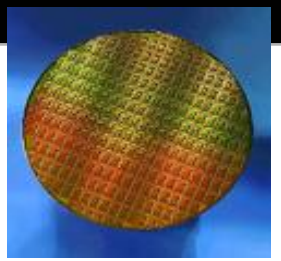
Computational Fluid Dynamics

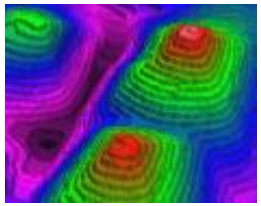




Life Sciences: Sequencing

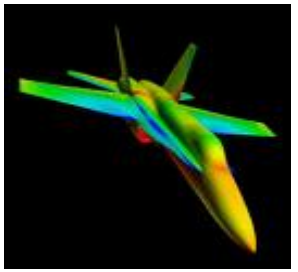







Blue Gene makes an impact in:

- **Government and Fundamental Science**
  - ***Material Science, Physics, Astrophysics***
  - ***Classified/Defense applications***
  - ***Climate, Environmental Modeling***
- **Upstream Petroleum**
- **Computer Aided Engineering: Computational Fluid Dynamics, EDA**
  - ***Automotive, Aerospace, Manufacturing***
  - ***Nuclear Engineering***
- **Life Sciences**
- **Large-Scale Data Streaming and Visualization**

What is driving the need for Petascale computing in Environment and Geosciences?

Climate: Broader Scale

- 5 years analysis/day at 10km resolution, global scale (mesoscale structure)

Weather: More resolution

- 1 day/hour at 5km resolution, global scale
- 1 day/hour at 1km resolution, continental scale (thunderstorms)
- 2 hours/day at 10m resolution, local scale (tornadoes)

Oceanography: Long term Analysis

- 40 years/month at 10km resolution, global scale (eddy scale with biology)

Hydrology: Long term Analysis

- 10 years/week at 1km resolution, large basin scale

Space Weather: Massive Data

- Coronal Mass Ejections (CME) faster than real-time (coronal B-field fine structure)

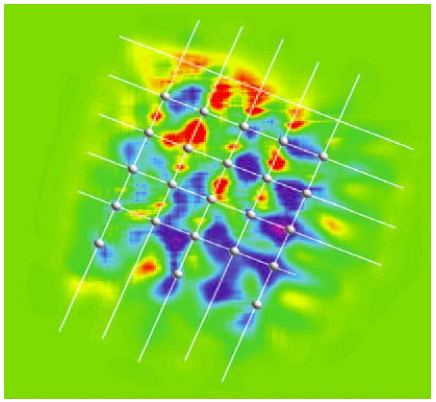
Earth System Science

- Model coupling across multiple scales and domains

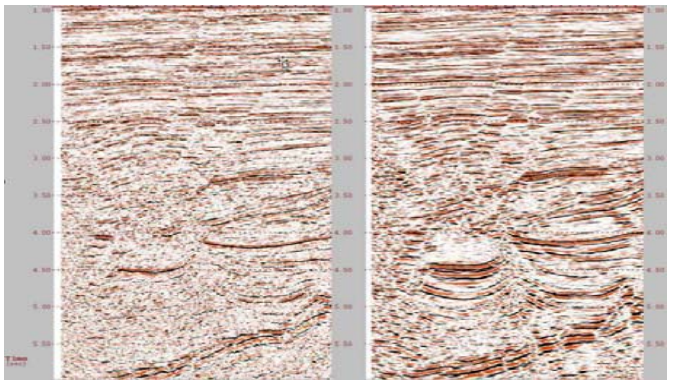
Petroleum Industry Applications on the Blue Gene

- **ExxonMobil Upstream Research**
  - Electrical Method for deep sub-salt imaging.
  - Node performance near parity with 2.8 GHz Xeon processors
  - 16 rack run for 27 hours (32,768 processors)
  - Run is roughly equivalent to a 28,000 processor Xeon cluster (which doesn't exist, and if it did would consume roughly 5x floor space and power)
  - Equivalent run on a standard Intel 1024 node (dual-processor nodes) Linux Cluster would take 16 days (assuming all nodes could stay operational for that long)
  - See Publication – Commer, et.al. (2008)*

ExxonMobil
Taking on the world's toughest energy challenges.™



- **Tsunami Development (ISV)**
  - Kirchoff-based PSTM and PSDM seismic imaging apps
  - Port complete, production tests to 4096 processors
  - High Resolution Seismic Imaging
  - Process 300 SqKm/day/rack PSTM, 41Million traces.
  - Tsunami states that one Blue Gene /L rack equals approximately a 1500 processor Opteron cluster



supercomputer, IBM Journal of Research and Development, Vol. 52 1/2 , 2008

Computational Fluid Dynamics on Blue Gene

Significant market/industry/segment drivers

- **Shorten the time to solution – Answers from days/weeks to minutes**
- **Improve understanding – Larger (3D and full body) models, with higher resolution**
- **Multiscale and multiphysics – Coupled and/or staggered thermal-structural-combustion-exhaust modeling**

CFD application areas for the BG:

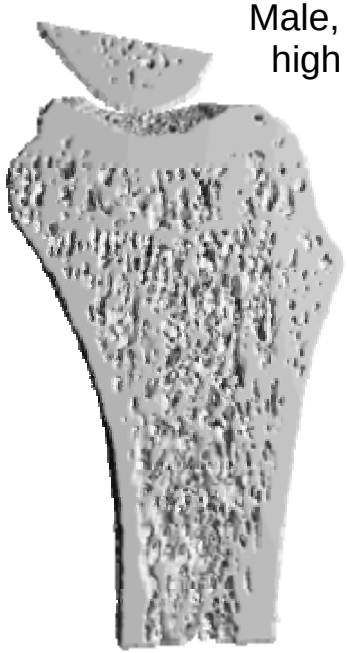
- **Aircraft and Automotive aerodynamics: structural design and turbulence modeling**
  - **commercial: Overflow, NSU3D, StarCD, CFD++, Powerflow**
  - **many proprietary codes: Stanford Univ/CDP (tested with GM), Georgia Tech, EPFL, Barcelona-BSC, GE-Tacoma**
- **Combustion**
  - **commercial: AVBP**
  - **many proprietary codes**
- **Environment – exhaust modeling**
  - **many proprietary codes**
- **Nuclear engineering**
  - **commercial: StarCD**
  - **many proprietary codes: EDF**

Examples of CFD codes that scale well on BG/L and BG/P:

- **CDP (Stanford U), PETSc-FUN3D, PPanel, AVBP (CERFACS), Code_Saturne (EDF), NEKTON, DNS, SAGE, Miranda, Raptor, Overflow-2, FUN3D, Ludwig, NEK5000, NSU3D, StarCD, AVUS, CTH**

Large Scale Strength Analysis of a Human Bone using Finite Elements

Male, 78 years
high density



Male, 72 years
low density

ZRL C. Bekas, A. Curioni

ETHZ P. Arbenz, H. v. Lenthe, A. Wirth, R. Müller

- Osteoporosis is a disease that causes significant loss of bone mass and deterioration of internal bone structure.
- According to the WHO, the risk of an osteoporotic fracture rises to 40% for women and 13% for men, causing health costs that are second only to cancer.

Osteoporotic fractures deficient healing process has lead to the development of novel osteoinductive biomaterials

AIST – Japan

- **Researchers are using Blue Gene for genome decoding, protein engineering, and drug design to shed light on how drugs interact with their targets in the body.**

Blue Brain - EPFL

- **The first objective of the "Blue Brain" EPFL-IBM joint research project is to create a cellular level, software replica of the Neocortical Column to study its function and dysfunction and to lay the foundation for large scale modeling of the mammalian brain.**

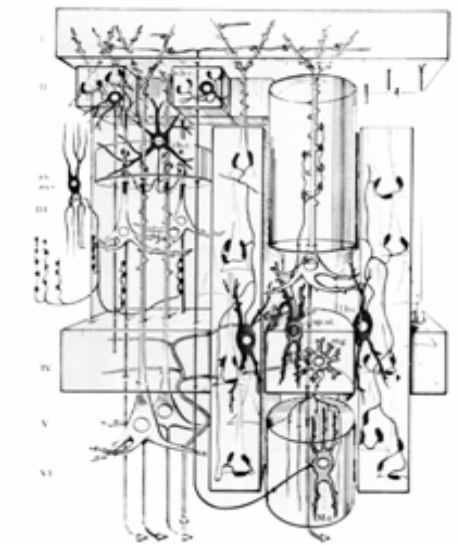
SBI (Stockholm Brain Institute)

- **SBI and IBM have embarked on a partnership that gives Swedish brain researchers access to BG/L – the first of its kind in the Nordic region.**
  - **"The combination of such enormous computer capacity and a high-resolution PET camera is unique in the world," says Hans Forssberg, Vice President of Karolinska Institutet and representative of the SBI. "Add to this the proximity to patients and clinical practice and we get entirely new opportunities for brain research from both a Swedish and international perspective.**
  - **Recent plan is to accelerate the porting of Siemens HRRT (high resolution research tomograph) code to Blue Gene. Demonstrate significant performance improvement such that it provides value to SBI and can be used as proof of value-add in selling Blue Gene to the other known 17 users of HRRT.**

Brain Mind Institute at Ecole Polytechnique Fédérale de Lausanne (EPFL)

"IBM's Blue Gene supercomputer allows a quantum leap in the level of detail at which the brain can be modeled. I argue that the time is right to begin assimilating the wealth of data that has been accumulated over the past century and start building biologically accurate models of the brain from first principles to aid our understanding of brain function and dysfunction."

Henry Markram



Blue Brain will search for new insights into how human beings think and remember, and how specific defects in our circuitry may contribute to **autism**, **schizophrenia** and **Parkinson's disease** With Blue Brain, research inquiries that used to require **several years of laboratory work can now be done in a matter of days, or even minutes**.

ASTRON: Low Frequency Radio Telescope

- Blue Gene enables LOFAR to provide higher resolution and sensitivity than any other low-frequency radio telescope
- Digital techniques provide extreme agility in frequency and pointing
- Multi-beaming capability allows simultaneous, full-sensitivity observations in widely separated directions
- High-bandwidth, fiber-optic network handles terabits/second
- Data buffers provide powerful, multi-steradian look-back capability

LOFAR

(LOw Frequency ARray) digitizes 10-240 MHz signals from an array of simple omni-directional antennas and processes the data on a central computer system to emulate a conventional dish antenna

STELLA

(Supercomputing Technology for Linked Lofar Applications) uses 6144 dual-CPU compute nodes of eServer Blue Gene® providing 27.45 Tf/sec

Blue Gene Consortium http://www.bgconsortium.org/

- **Purpose**

  A community of Blue Gene users who share experiences, help understand an d develop the application space, assist in application porting and tuning and p rovide feedback on functional requirements for next generation

- **Membership Criteria**

  Membership is open to may be from any geography and any industry or organization, committed to provide a skill or undertaking such as performance measurements, application porting, OS/compiler testing

- **Consortium Facilitator**

  Argonne National Lab (ANL)
  http://www-fp.mcs.anl.gov/bgconsortium/default.htm

Information Sources

- ✓ **Official Website**
  - www.ibm.com/servers/deepcomputing/bluegene.html
- ✓ **Research Site**
  - www.research.ibm.com/bluegene/index.html
- ✓ **Blue Gene Redbook**
  - There are several Blue Gene Redbooks.
- ✓ **ANL BGL Wiki**
  - www.bgl.mcs.anl.gov/wiki/index.php/ANLBGLWiki_Home
- ✓ **Blue Gene Consortium**
  - http://www.bgconsortium.org/
- ✓ **LLNL Blue Gene Site**
  - www.llnl.gov/asci/platforms/bluegenel/
- ✓ **TOP500 List**
  - www.top500.org
- ✓ **IBM Journal of Research and Development**
  - www.research.ibm.com/journal/rd49-23.html
- ✓ **IBM Journal of Research and Development**
  - http://www.research.ibm.com/journal/rd52-12.html