WWC-6

- Disclaimer
- Chicken and Egg Problem Solved !
- Where are the Media Applications
- Runtime environments
- Ultimate Challenge for Computing

Disclaimer

- Projecting future anything is futile
- History is the only guide
- My views might not be those of Intel's

Chicken and Egg Problem Solved !

Future Needs
To
Future Workloads
To
Future Architectures

Order has not always been right !!!

Managed Runtimes: Key Trends

- **Changes driven by need for faster, more complex development and rapid deployment**
  - "Plumbing" moving from application software to platform software
    - Multithreading, Dynamic memory management
    - Security, Reliability, Bounds check, Type checking
- **Increasingly decoupled software components in the enterprise**
  - Web Services, Transactional message queues, and distributed transactions usage is growing
- **Object oriented design / web services result in data being wrapped in increasing layers of abstraction**

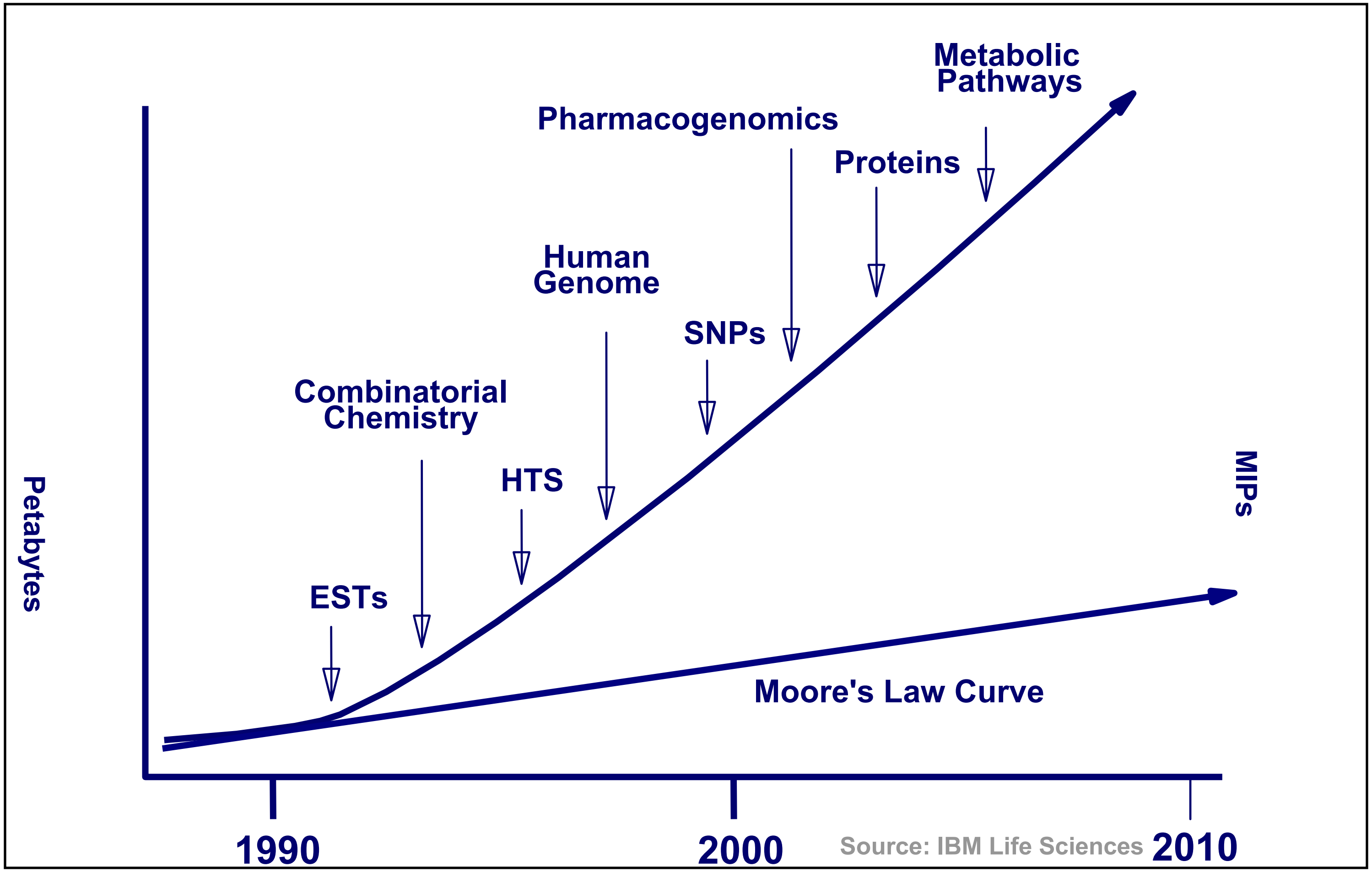
Media Applications

- **Voice recognition to Lip Reading**
- **Video Streams and Transcoding becomes common place**
  - Partial search to Full Search
  - From Batch Job to real time to better than Realtime

Implications of Java / .Net on System Design

- **More mips consumed in general**
  - Instruction path length is longer due to abstraction
  - Deeper stacks due to smaller functions (OO)
  - More type casting
- **Scheduling is complicated**
  - More branches due to OO (Object Oriented)
  - Branch prediction complicated by virtual calls
  - Load and call intensive, long call chains
- **Higher cache pressure**
  - Garbage Collection (GC)
  - More / deeper pointer chasing
  - Larger working set size
- **More self modifying code (JIT,GC)**

The Life Sciences Computing Challenge



The Life Sciences Community must innovate faster than Moore's law

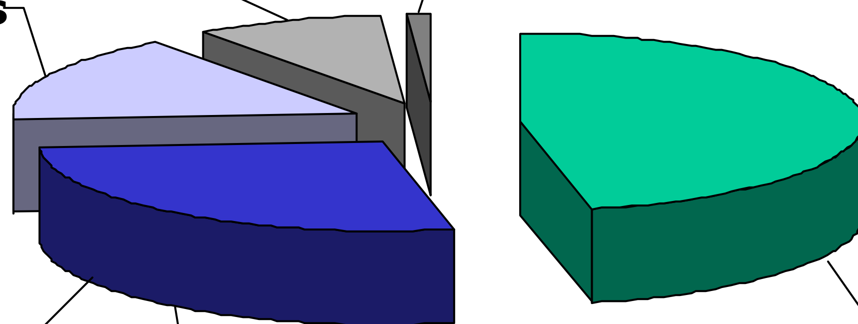
Massive Parallelism, new algorithms, and new approaches for I/O must be developed.

Life Sciences Drives HPC: Processor Forecast Units

	2001	2002	2003	2004	2005	CAGR
Life Sciences	**70,811**	**116,331**	**190,689**	**297,811**	**451,860**	**62%**
Manufacturing Total	**72,273**	**94,630**	**129,179**	**172,556**	**231,024**	**35.8%**
University	**88,263**	**112,650**	**148,809**	**193,536**	**250,875**	**32.3%**
EDA	**38,795**	**52,152**	**73,214**	**100,440**	**138,554**	**39.3%**
Technical Mgmt.	**25,619**	**35,890**	**52,499**	**74,815**	**107,275**	**45.2%**
Govt Total	**52,213**	**67,074**	**89,443**	**117,013**	**152,330**	**33.1%**
Petroleum Total	**27,025**	**37,463**	**54,146**	**76,846**	**109,336**	**43.9%**
Digital Media/Content	**9,271**	**12,845**	**18,612**	**26,418**	**37,671**	**43.8%**
Technical Other	**15,956**	**20,564**	**29,201**	**44,244**	**72,063**	**55.7%**
Chemistry	**22,617**	**26,133**	**30,556**	**33,445**	**34,043**	**14.3%**
Software Engineering	**8,493**	**10,744**	**14,127**	**18,047**	**23,075**	**30.3%**
Finance Total	**13,542**	**16,953**	**21,935**	**27,587**	**34,263**	**29.1%**
Total	**444,878**	**603,429**	**852,410**	**1,182,758**	**1,642,369**	**40.8%**

Source: IDC, 2001

Sources of Bio Applications



BLAST –FASTA – DOCK - HMMER
Phred/Phrap

BLAST – GCG - ClustalW – DOCK - HMMER
GeneSpring – LabBook –MassLynx Phred/Pharp
Discovery Center – Netgenics - Rosetta Resolver
Lion SRS

Source: IDC custom end user survey 2001

"Life Sci. companies" are software companies!

*the names BLAST, FASTA, DOCK, HMMR, GCG, ClustalW, GeneSpring, LabBook, MassLynx, Phred/Phrap, Netgenics, Rosetta resolver and Lion SRS are the property of their respective owners

Market segmentation in Life Sciences

- **Market segments (according to Burrill & Co.)**
  - Basic Science: Research to build basic knowledge in biological sciences. (NIH/academic/National-lab research community).
  - BioTechnology: Technology development to support life sciences.
  - Pharmaceuticals : Drug R&D, clinical trials, marketing.
  - AgBio/biochemical :Genetically Modified food, biochemicals from living systems.
  - Bio Defense : Detection of bioweapons in the field in real time.
  - Equipment/suppliers
- **The near-term opportunities for Intel are in the most compute intensive segments:**
  - Basic Science
  - BioTechnology
  - Pharmaceuticals

Life Sciences Solution Focus Areas

SFA's	**Massive Data Management**	**Bio-informatics**	**Bio-Simulation**	**Biomolecular Analysis Systems**
High Level Insights	• Effective management of massive amounts of data spread between one or more data bases. • Multi-terabyte today, petabytes by 2010. • Technology is all over the map: from Oracle and DB2 to flat file systems. • Data Grids will play a major role.	• Understanding and correlating patterns in data to biological behavior (e.g. target discovery). • Working with patterns in sequences (DNA and Proteins) and other data. • Its more than the sequences – its machine learning and knowledge engineering.	• Modeling biological systems on computers. • Physical modeling: spanning DNA to protein to metabolites to cell to organ to organism. • Long term goal: In-silico modeling of biological systems (e.g. drug design).	• Silicon manufacturing technology applied to devices/instruments for life science applications. • Microfluidics, single molecule detection – "lab on a chip". • Includes Biochips – both DNA and Protein.

Avi Kumar

Collaborations today

SFA's

Massive Data management

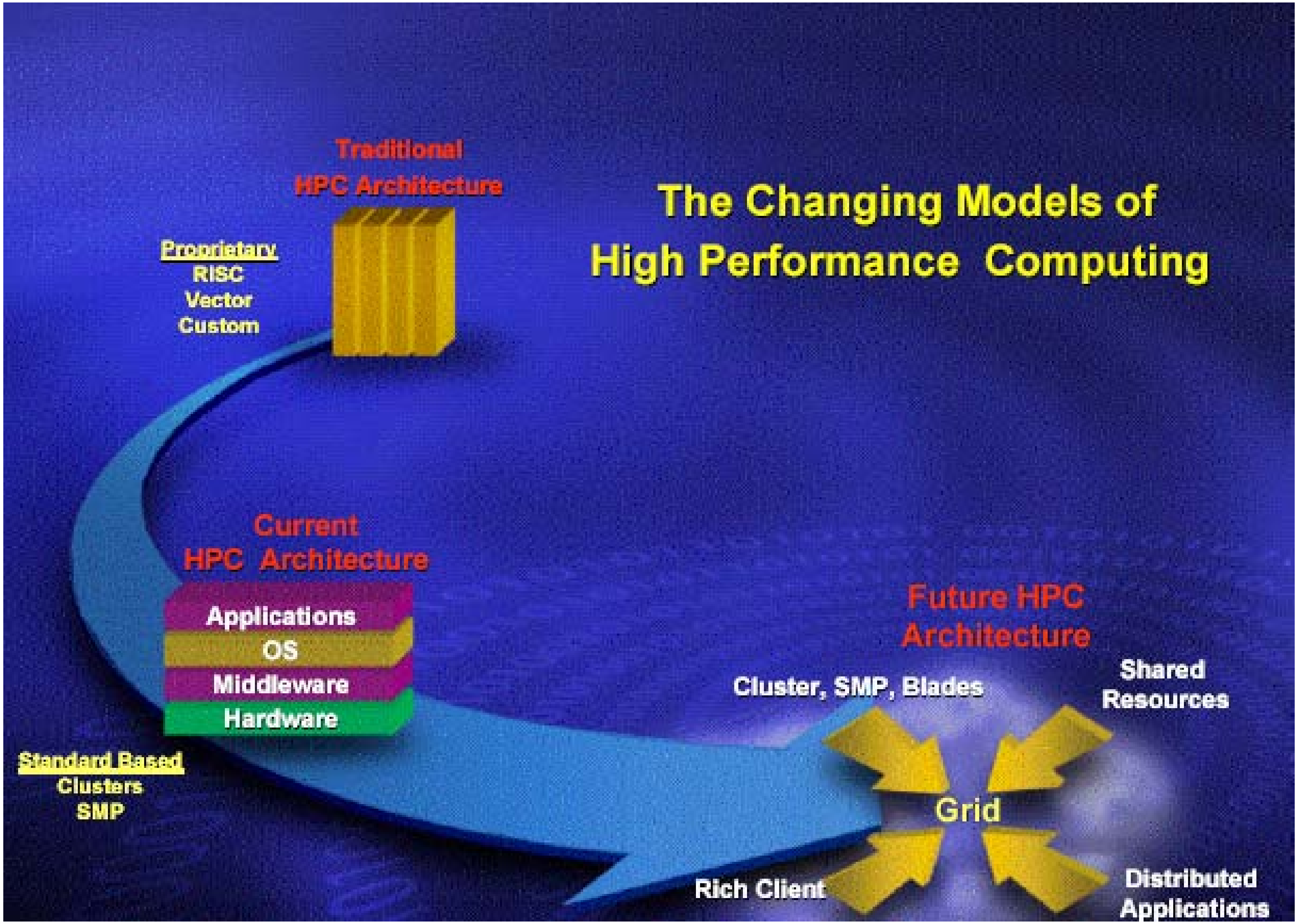
BioSimulation

BioInformatics

- HPC clusters and scalable file systems for Genomics.
  - **We've built an R&D cluster that will be used to explore different cluster file systems and different software stacks. Also studying large memory nodes to index and consolidate data.**
- Protein modeling
  - **Hypothesis: Itanium's floating point architecture maps well onto MD force calculations. This collaboration will test this hypothesis.**
- Bioinformatics software framework
  - **We will help build a common software framework to support complex bioinformatics collaborations across several institutions.**

Our lawyers won't let us name our collaborators yet. That will come soon.

Is Future HPC Computing Grid Computing ?



Timelines of different Computing Paradigms