Experiences With IBM Server Performance

Michael Ignatowski

IBM Server Group
ignatow@us.ibm.com

In Days of Old (1980s).....

- Bipolar Mainframes
  - ~$1 million per processor
    - (with several MB of memory)
- Goals:
  - CPU is an expensive piece of property
    - Do not allow it to sit idle due to other system constraints
  - Common Measure:
    - Execution-Unit Utilization - strive for 100%
      - Understand what keeps this from happening

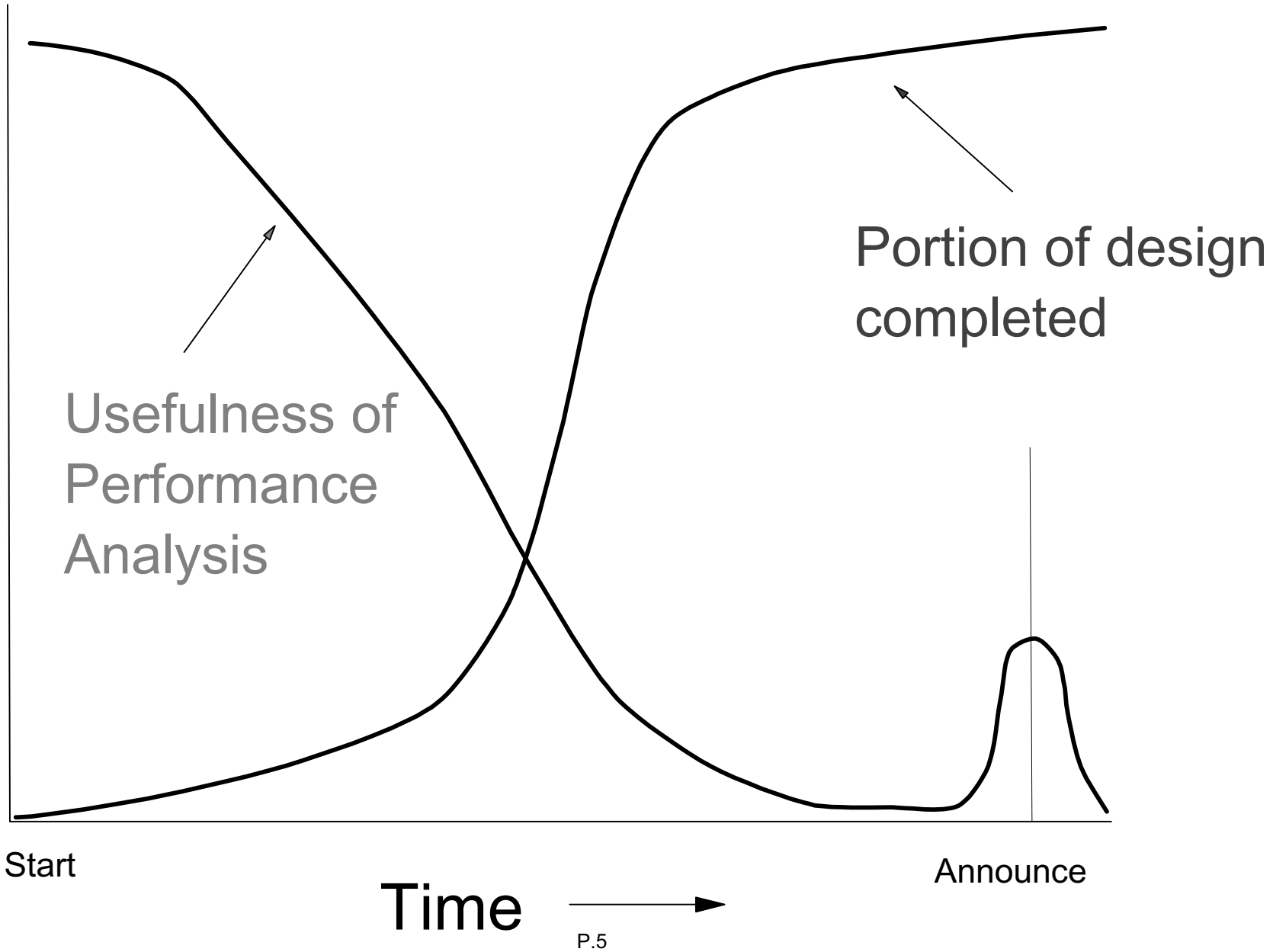
Historical Performance Process (S390)

1. Define Internal benchmarks
   - Batch, interactive, data base, technical....
2. Collect and select traces
   - Update regularly as OS and applications are updated
3. Run traces through various levels of models
   - Cycle-by-cycle timers
   - Cache models (miss rates)
   - Trace analyzers (opcodes, branches, etc...)

Importance of an *Independent* **Performance Team**
(i.e. not part of any development team)

- Accumulation of experience, knowledge, tools.
  - Able to apply knowledge and tools from one project to other projects
  - Development of a performance community to share knowledge and insight
- Unbiased
  - Reputation depends on accuracy and completeness.
  - Less motivation to make a project "look good"
- Less chance of performance team being plundered by development when schedules get tight.

Performance Analysis and the Design Cycle



Components of Performance

Component	Uncertainty in Projections	Comments
CPI (Cycles per Instr)	+/- 5%	Detailed Tools influenced by server organization
Frequency	+/- 15%	Less detailed tools controlled by the technology organization
Pathlength	+/- 40% (SMP Scaling)	Little or no tools

- Projecting absolute TPC-C values beyond 6 months is questionable
- Excessive emphasis on CPI studies in practice

Some Current Issues:

- SMP Software Scaling
  - Behavior of current software less interesting than extrapolation of future trends
  - Understand sources of SMP software degradation
    - Assume no serious lock contention (?)
  - How to deal with multithreading?

Some Current Issues (cont.):

- Latency Hiding
  - Two major components
    - Useful instruction processing under a miss
    - Generate multiple outstanding misses
  - Important aspects of processor design
  - Effects of interlocks and synchronization
  - Speculative execution into critical locked segments

Some Current Issues (cont.):

- Future Software Trends
  - Possibly the biggest issue in future server design
  - What are the specific trends in:
    - MP scalability, cross-interrogate rates
    - Caching characteristics
    - Memory size and bandwidth requirements
    - I/O bandwidth requirements

Some Current Issues (cont.):

- Specific Workloads
  - OLTP: problems with TPC-C (vers. 3)
    - Need longer more complex transactions
    - More realistic cluster version (less partitioned)
  - BI (Business Intelligence): problems with TPC-H
    - Hard to use as a design metric, too many components
    - System highly tuned to benchmark is not representative
  - e-Commerce Where is TPC-W ?
  - Other Generic Trends:
    - Significantly increased media content
    - Convergence of communication and computation
  - What is the relative importance of these?

Some Aspects of Future Server Usage

- SMP software limits no longer a factor
  - Improvements in software scaling
  - Pervasive use of Partitioning
- Dynamic Workload Variations
- Unpredictable Growth
- System Availability may be the key metric
  - i.e. the ability to provide required level of service at all times under all conditions

INFINIBAND™ Technologies (IBT)

- The I/O bussing technology formerly known as "SIO"
  - Convergence of NGIO and FIO efforts
- The future commodity I/O bus (?)
- Some big questions:
  - How / when to introduce into commercial servers
  - Network management
  - Support for partitioning
  - Support for IPC message passing, standard APIs ?

Use of Large MP Systems

- Metrics for NUMA and Cluster systems
  - Multiple instances of database application (?)
    - Important issue for NUMA configurations
  - Assume non-partitioned data
    - as opposed to extremely partitioned data in TPC-C
    - Most users do not have time to partition data, the e-business world is changing too fast.
    - If effort was made to partition data, it would become "out of tune" within a few short weeks.
  - Do not assume perfectly balanced workload

Final Topic...

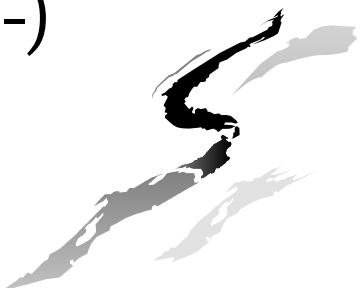
A slightly "tongue-in-cheek" view of the future of computer architecture.

...from a Storage Bigot

Trends into the Future

- Processor cost: order of ~$1000 and declining
- The memory system is the expensive real-estate
  - It determines:
    - Number of users supported
    - Size of problems computed
    - Ultimate bandwidth limits to processors and I/O
    - System Cost and physical size

- Don't let it sit around idle !
  - strive for 100% memory utilization :-)



Some Revised Definitions:

- Computer System:
  - A memory system with assorted peripherals (disks, processors, power supplies, etc....)
- Computer Program:
  - An algorithm for defining a sequence of storage references (~ cache misses)
- Processors:
  - A device that analyses a computer program and generates the correct series of cache misses
- System Performance:
  - Rate at which the memory is servicing misses

Goals

Design goal of processors:

- Generate misses as rapidly as possible
- Preferably multiple outstanding misses at a time

"Never let an inexpensive component become a limiting factor in system performance"

- Processors are relatively inexpensive...
- Try to avoid more than 30%-40% average processor utilization when the system is running at full capacity.
- Execution units will cost a couple of dollars each. Don't worry about their utilization.

In the future, this leads us to.....

The "Cup Holder" Analogy

- System designers should view processors in the same way as mini-van designers view cup holders.
- Put in as many as any customer could conceivably use.
- Then throw in a few more for good measure.