Optimum System Balance for Systems of Finite Price

John D. McCalpin, Ph.D.
IBM Corporation
Austin, TX

SuperComputing 2004
Pittsburgh, PA
November 10, 2004

Overview (continued)

- Q: What is a “balanced” system?
- My answer:
  “A balanced system is one for which the primary applications are limited in performance by the most expensive component(s) of the system.”

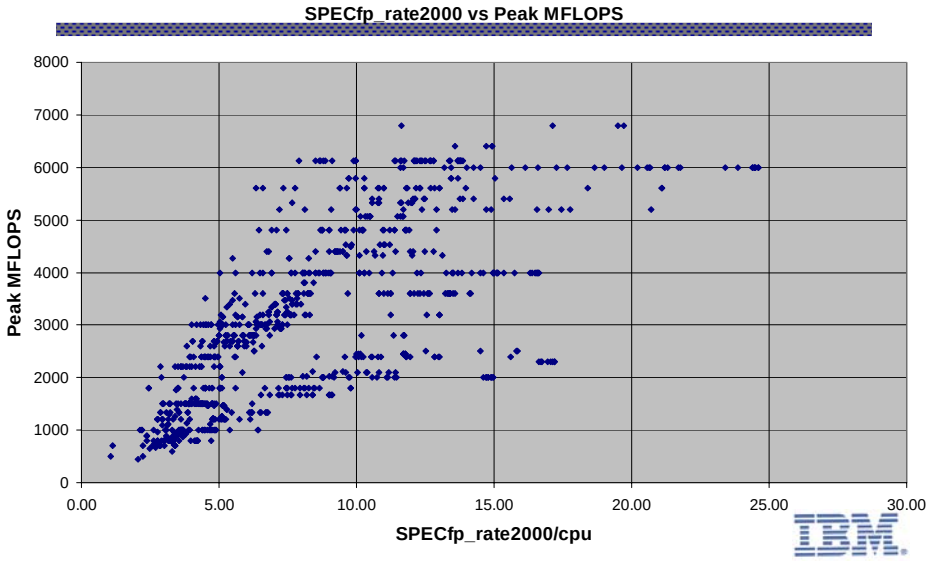
A Simple Composite Model

- Assume the time to solution is composed of a compute time proportional to peak GFLOPS plus a memory transfer time proportional to sustained memory bandwidth
- Assume “x Bytes/FLOP” to get:

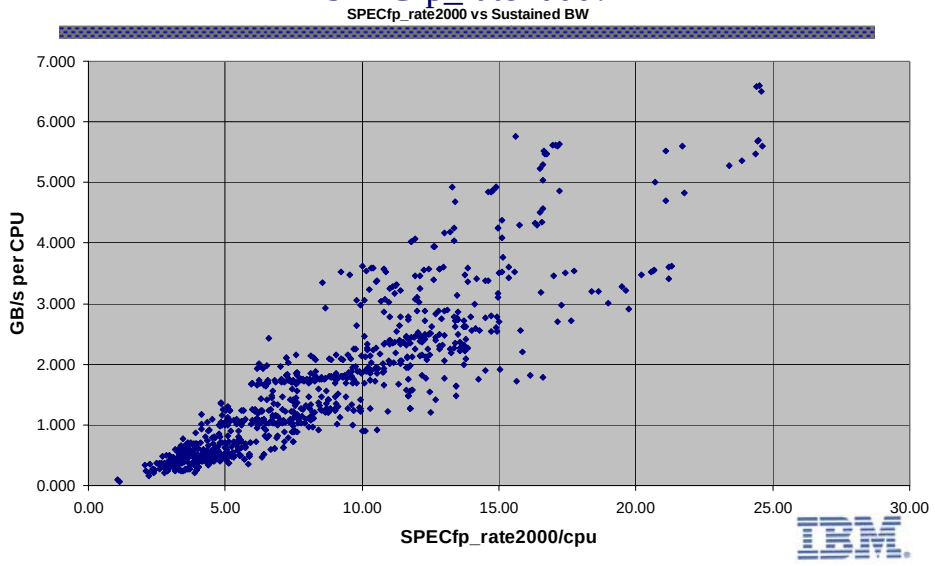
$$\text{"Balanced GFLOPS"} \equiv \frac{1\ \text{"Effective FP op"}}{\left(\frac{1\ \text{FP op}}{\text{Peak GFLOPS}}\right)+\left(\frac{x\ \text{Bytes}}{\text{Sustained GB/s}}\right)}$$

- Target SPECfp_rate2000 as the workload

Does Peak GFLOPS predict SPECfp_rate2000?



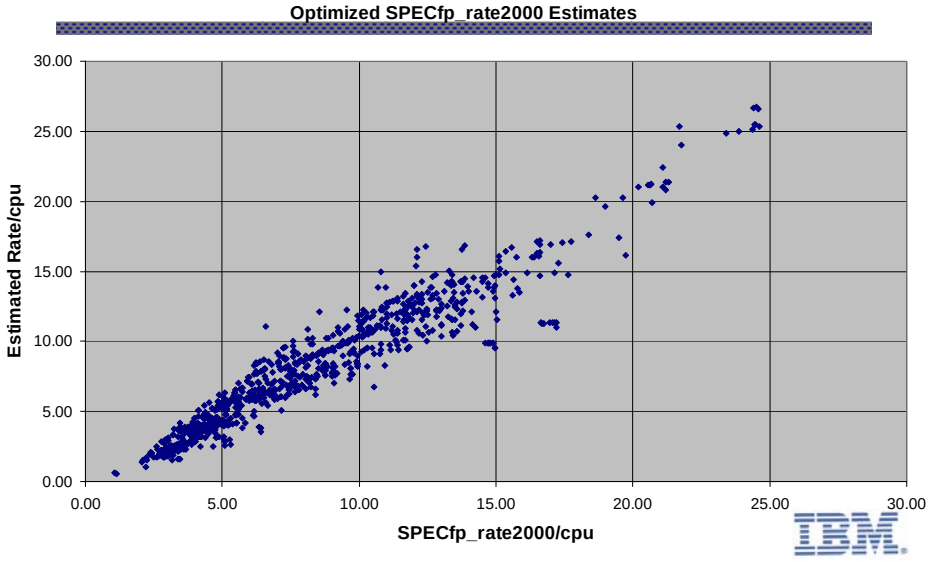
Does Sustained Memory Bandwidth predict SPECfp_rate2000?



Optimized Model Results

- Results rounded to nearby round values:
  - Bytes/FLOP for large caches === **0.16**
  - Bytes/FLOP for small caches === **0.80**
  - Size of asymptotically large cache === **~12 MB**
  - Coefficient of best fit === ~6.4
  - The units of the coefficient are
    SPECfp_rate2000 / Effective GFLOPS

Does this Revised Metric predict SPECfp_rate2000?



Cost Model

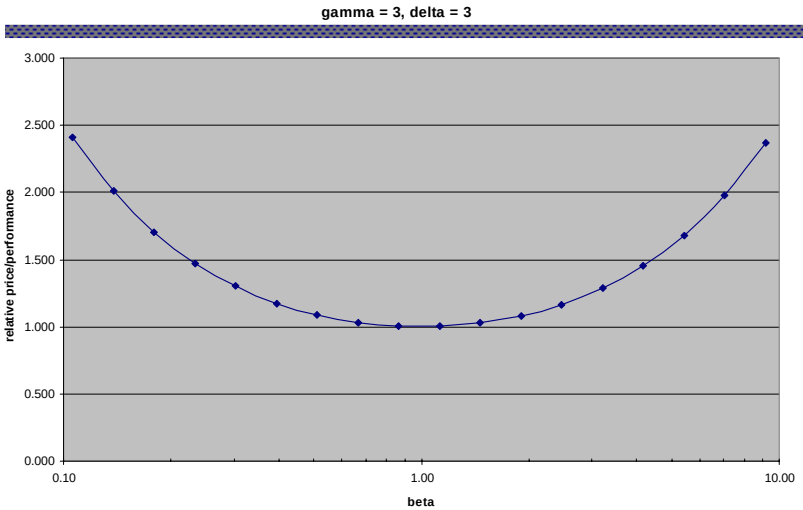
- Assume simple linear additive model
  - FLOPS cost some amount
  - Sustained BW costs a different amount
  - Define:

$$\text{beta} = R_{mem} / R_{cpu}$$
$$\text{gamma} = W_{mem} / W_{cpu}$$
$$\text{delta} = (\$/BW) / (\$/FLOPS)$$

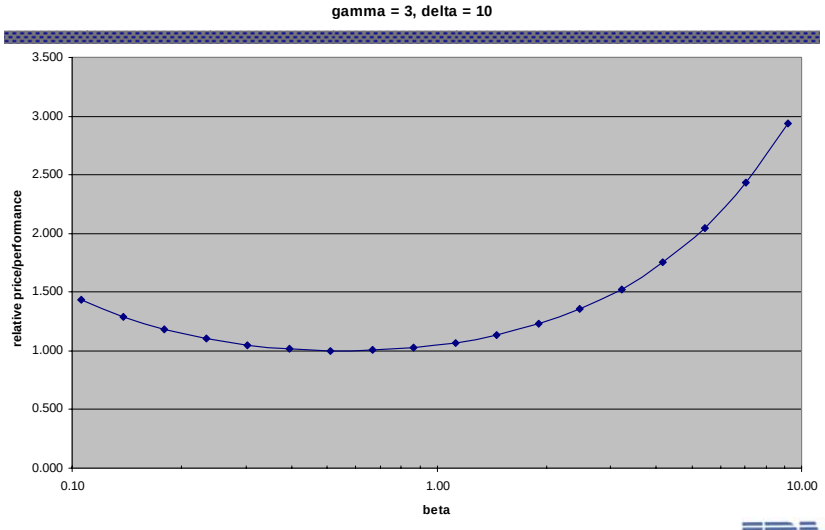
The Correct Optimization

- This is actually an easy optimization problem
- Minimize cost/performance
  - Same as minimizing cost * time
- Optimum cost/performance occurs at
  - beta = sqrt(gamma/delta)

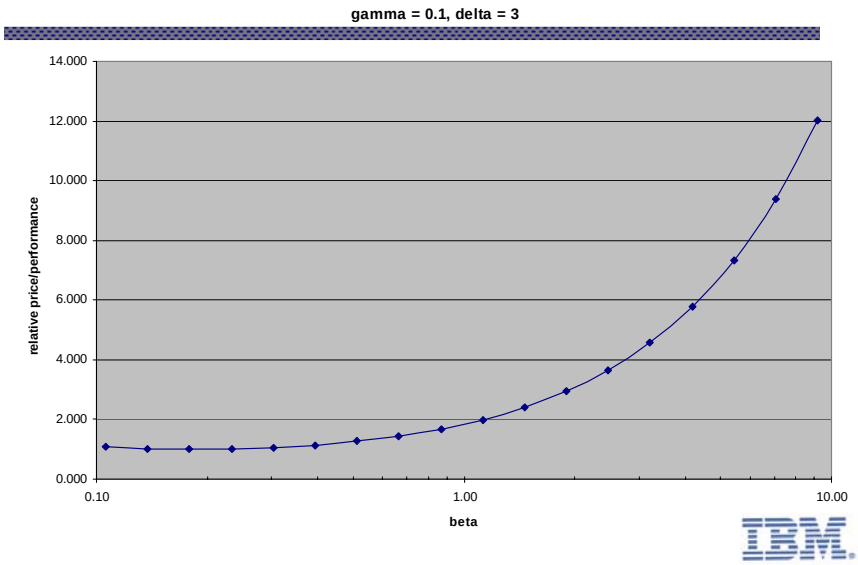
Example: High BW, expensive BW



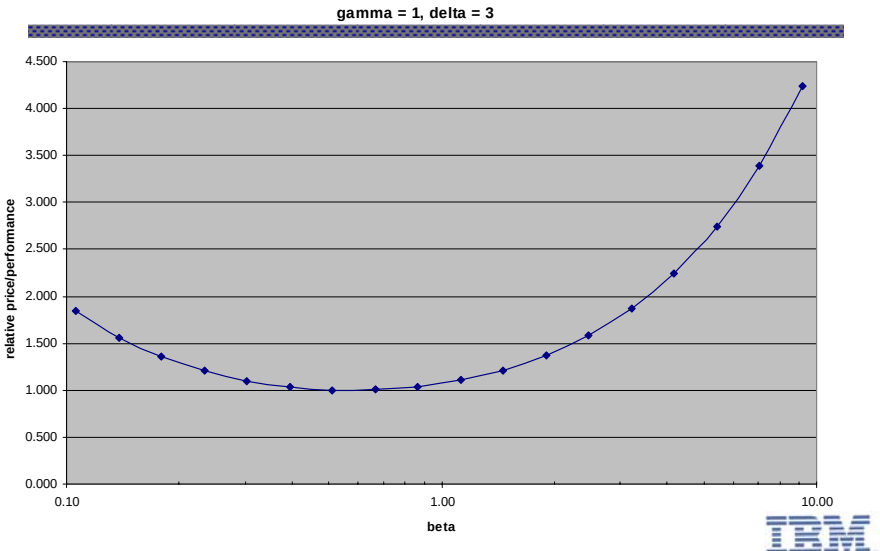
High BW, very expensive memory



Low-BW, expensive BW



Medium BW, expensive BW



Summary

- Balance is important to cost/performance
- You must understand performance
- You must understand cost
- Optimum cost-performance is not intuitive!