The Path to Extreme Supercomputing—LACSI Workshop

DARPA HPCS

David Koester, Ph.D.
DARPA HPCS Productivity Team

12 October 2004
Santa Fe, NM

This work is sponsored by the Department of Defense under Army Contract W15P7T-05-C-D001. Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the United States Government.

Outline

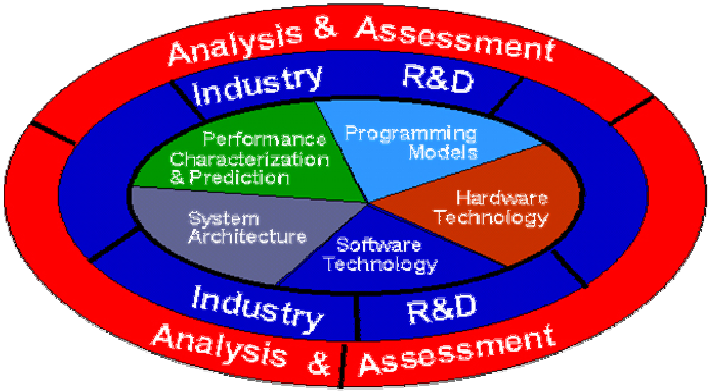
- Brief DARPA HPCS Overview
  - Impacts
  - Programmatics
  - HPCS Phase II Teams
  - Program Goals
  - Productivity Factors — Execution & Development Time
  - HPCS Productivity Team Benchmarking Working Group
- Panel Theme/Question
  - How much?
  - How fast?

High Productivity Computing Systems

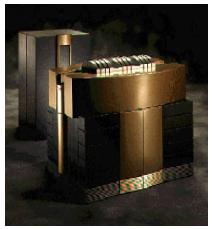
➢ **Create a new generation of economically viable computing systems (2010) and a procurement methodology (2007-2010) for the security/industrial community**

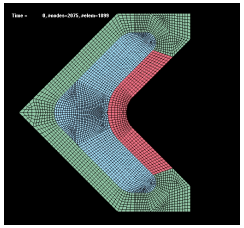
Impact:

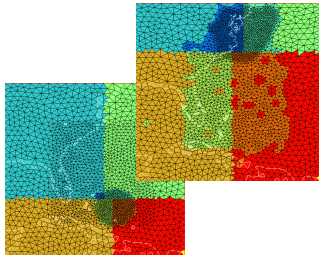
- **Performance** (time-to-solution): speedup critical national security applications by a factor of 10X to 40X
- **Programmability** (idea-to-first-solution): reduce cost and time of developing application solutions
- **Portability** (transparency): insulate research and operational application software from system
- **Robustness** (reliability): apply all known techniques to **protect against outside attacks**, hardware faults, & programming errors

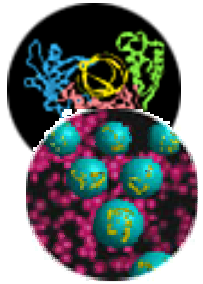


HPCS Program Focus Areas









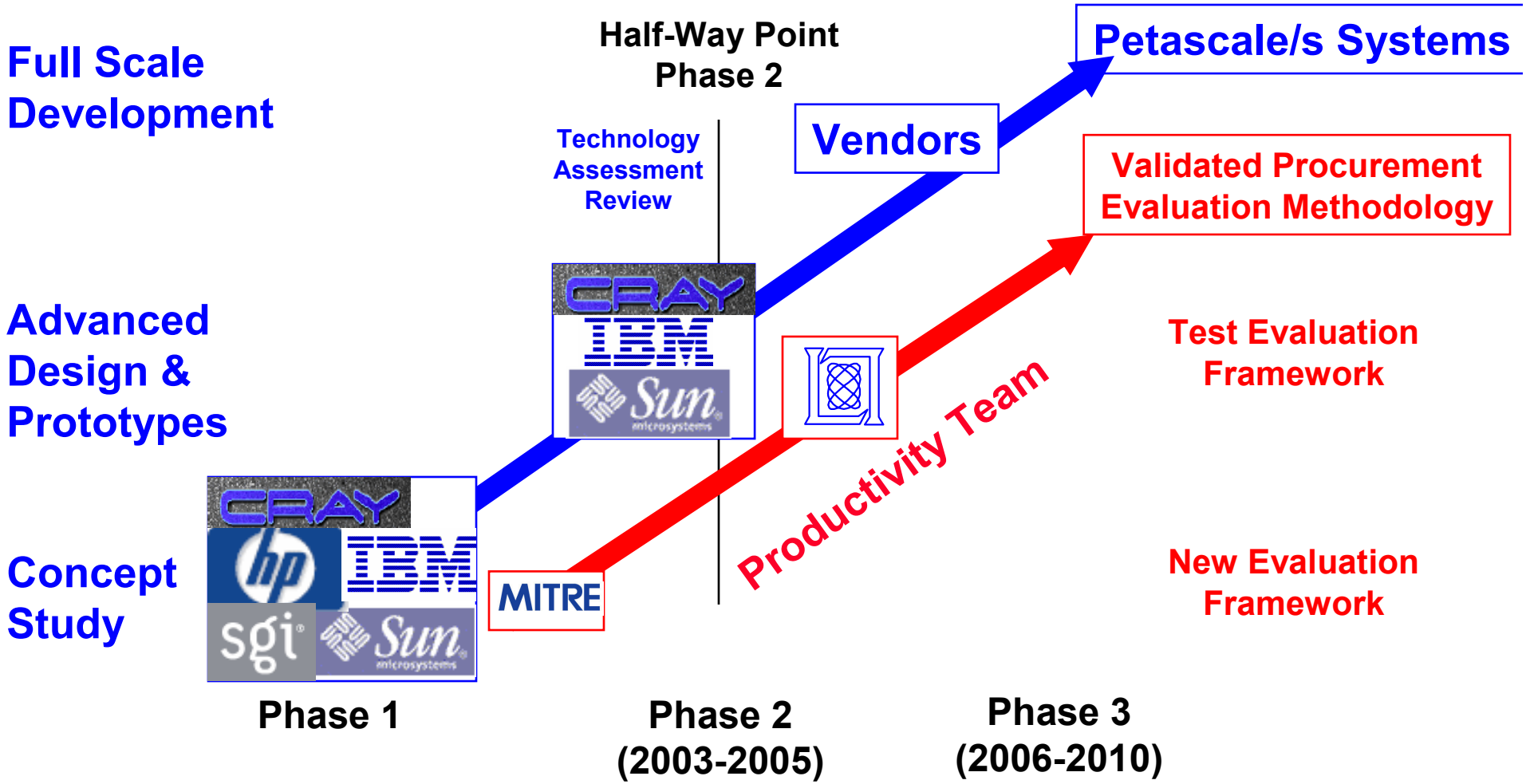
Applications:

- Intelligence/surveillance, reconnaissance, cryptanalysis, weapons analysis, airborne contaminant modeling and biotechnology

Fill the Critical Technology and Capability Gap
Today (late 80's HPC technology).....*to*.....Future (Quantum/Bio Computing)

High Productivity Computing Systems
-Program Overview-

- Create a new generation of economically viable computing systems (2010) and a procurement methodology (2007-2010) for the security/industrial community



HPCS Phase II Teams

Industry

PI: Elnozahy

PI: Mitchell

PI: Smith

Mission Partners

Office of Science
U.S. Department of Energy

National Nuclear Security Administration

Productivity Team (Lincoln Lead)

MIT Lincoln Laboratory

PI: Kepner

PI: Lucas

PI: Basili

PI: Benson & Snavely

PI: Dongarra

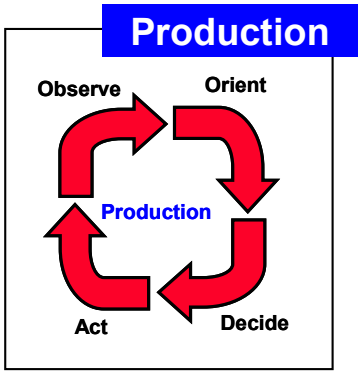
PI: Koester

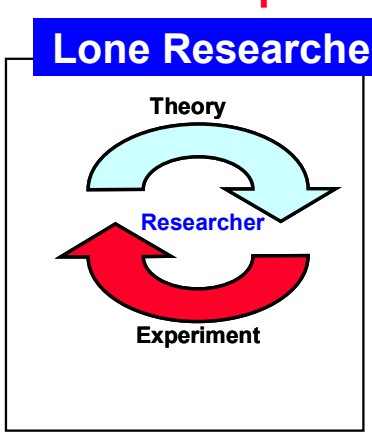
PIs: Vetter, Lusk, Post, Bailey

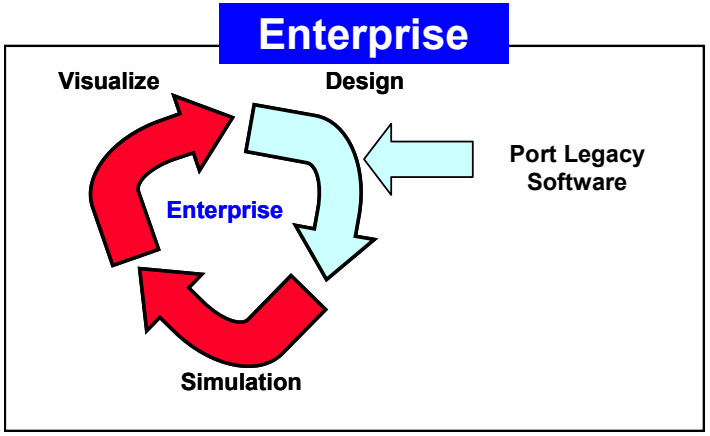
PIs: Gilbert, Edelman, Ahalt, Mitchell

HPCS Program Goals
Productivity Goals

- **HPCS overall productivity goals:**
  - **Execution (sustained performance)**
    - **1 Petaflop/s (scalable to greater than 4 Petaflop/s)**
    - **Reference: Production workflow**
  - **Development**
    - **10X over today's systems**
    - **Reference: Lone researcher and Enterprise workflows**





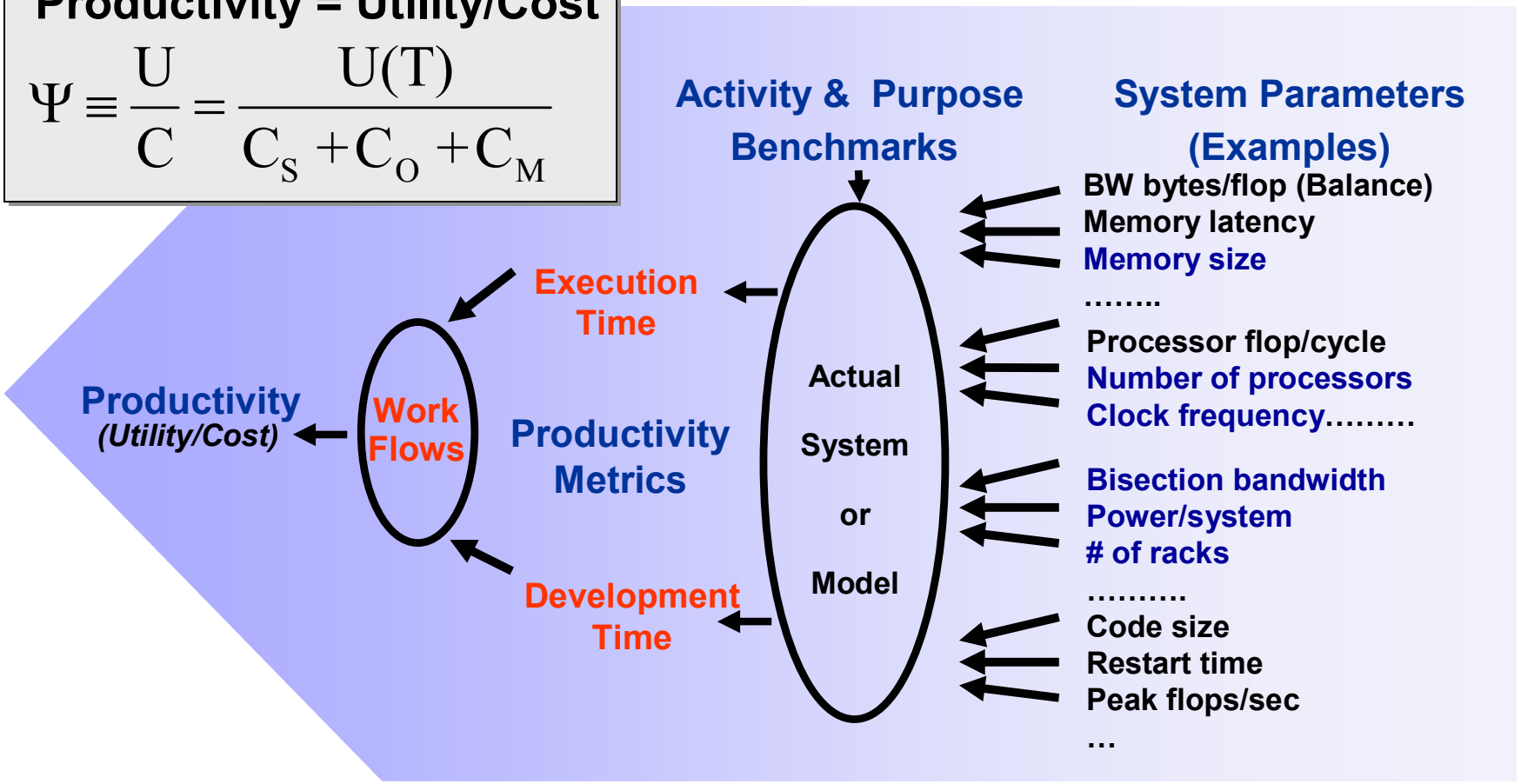


10x improvement in time to first solution!

HPCS Program Goals Productivity Framework

Productivity = Utility/Cost

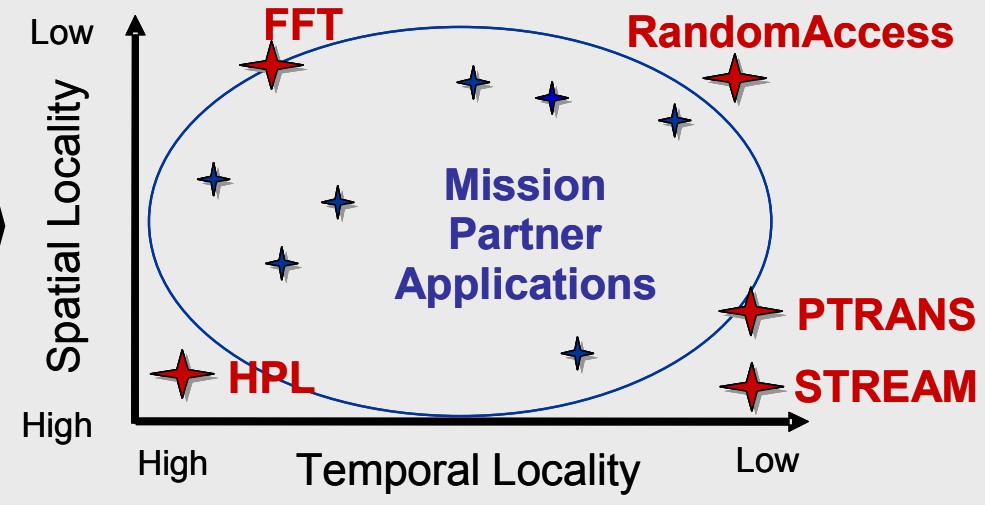
$$\Psi \equiv \frac{U}{C} = \frac{U(T)}{C_S + C_O + C_M}$$



HPCS Program Goals Hardware Challenges

- General purpose architecture capable of:
  - Subsystem Performance Indicators
  1) 2+ PF/s LINPACK
  2) 6.5 PB/sec data STREAM bandwidth
  3) 3.2 PB/sec bisection bandwidth
  4) 64,000 GUPS

HPCS Program Goals & The HPCchallenge Benchmarks

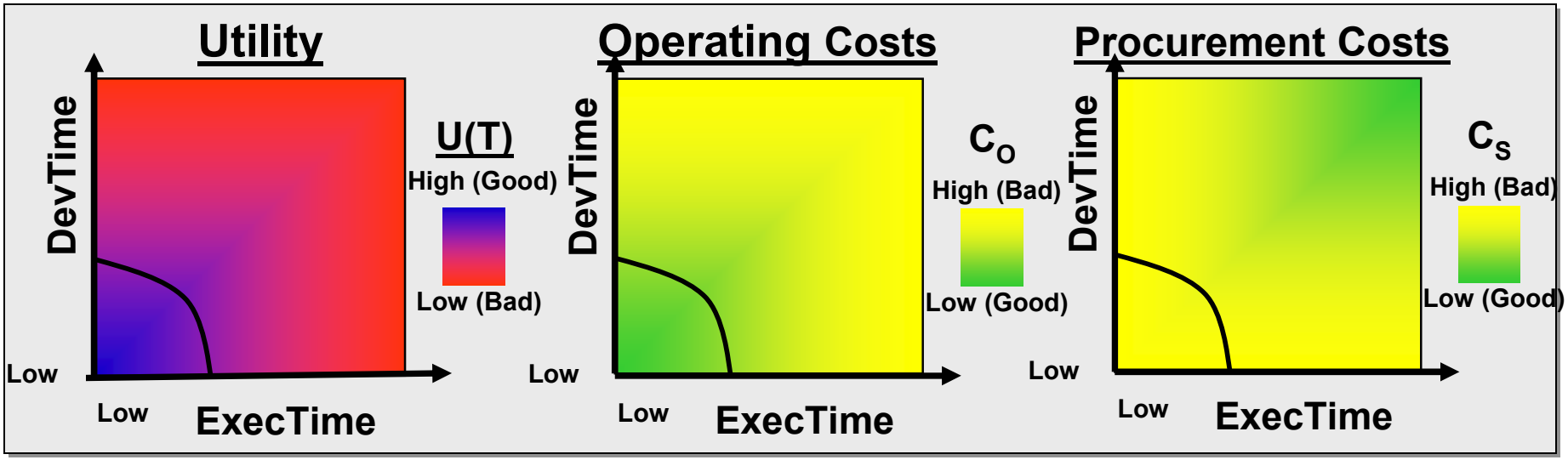


Productivity Factors
Execution Time & Development Time

Productivity = Utility/Cost

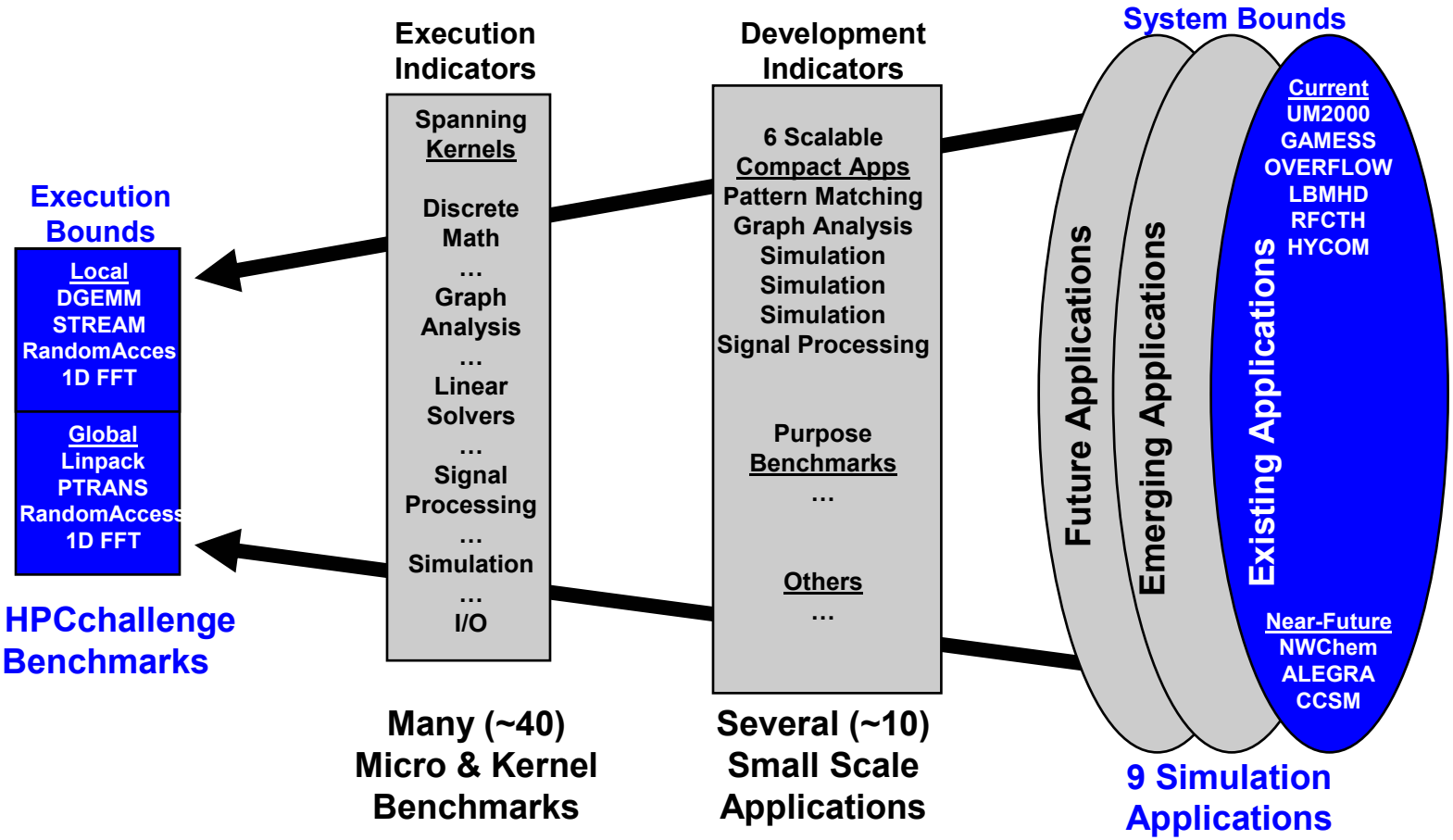
$$\Psi \equiv \frac{U}{C} = \frac{U(T)}{C_S + C_O + C_M}$$

- Utility and some Costs are relative to
  - Workflow (WkFlow)
  - Execution Time (ExecTime)
  - Development Time (DevTime)



- Reductions in both Execution Time and Development Time contribute to positive increases in Utility
  - Utility generally is inversely related to time
  - Quicker is better
- Reductions in both Execution Time and Development Time contribute to positive decreases in operating costs
  - Reduction in programmer costs
  - More work performed over a period
- However, systems that will provide increased utility and decreased operating costs may have a higher initial procurement cost
  - Need productivity metrics to justify the higher initial cost

HPCS Benchmark Spectrum



- Spectrum of benchmarks provide different views of system
  - HPCchallenge pushes spatial and temporal boundaries; sets performance bounds
  - Applications drive system issues; set legacy code performance bounds
- Kernels and Compact Apps for deeper analysis of execution and development time

Panel Theme/Question

- "How much should we change supercomputing to enable the applications that are important to us, and how fast?"

- How much? — HPCS is intended to "Fill the Critical Technology and Capability Gap between Today's (late 80's HPC technology)…..*to*…..Future (Quantum/Bio Computing)
- How fast?
  - Meaning when — HPCS SN-001 in 2010
  - Meaning performance — Petascale/s sustained

- I'm here to listen to you — the HPCS Mission Partners
  - Scope out emerging and future applications for 2010+ delivery
    (What applications will be important to you?)
  - Collect data for the HPCS Vendors on future
    - Applications
    - Kernels
    - Application characterizations and models

Statements

- Moore's Law cannot go on forever

  Proof: $2^x \rightarrow_{x \rightarrow \infty} \infty$

  So what?

- Moore's Law doesn't matter as long as we need to invest the increase in transistors into machine state — i.e., overhead — instead of real use