Enabling Architectures for Large-Scale Applications

Marc Snir and Bill Gropp

University of Illinois

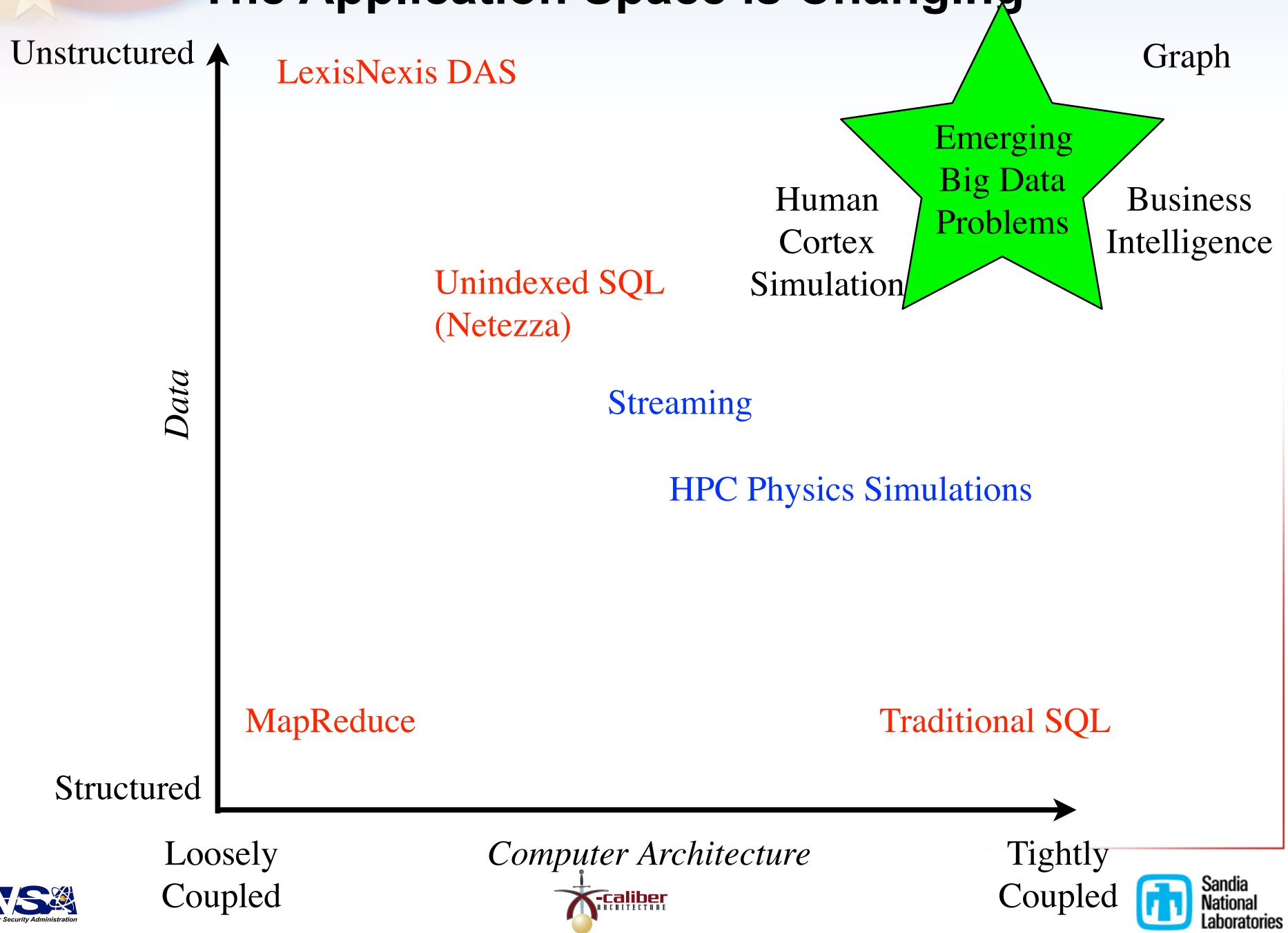**Mike Heroux and Richard Murphy**

Sandia National Laboratories

February 22, 2010

Key Points

- UHPC Scales: tera-embedded, peta-rack, exa-acre
- Two likely compute engines: Multithreaded or Accelerated
- Who cares? The data movement dominates power
- Application space changing (next slide)
- Programming Models are Changing
  - We cannot solely rely on BSP, but it may work at the low end
  - Must be asynchronous, message-driven, hierarchical, and possibly heterogeneous
- Weak Scaling is still key
- Academic, Industry, National Lab Partnerships are Required
  - to support meaningful off-roadmap prototyping
  - to provide impact to hard applications (and go beyond SPEC)
  - to stay competitive economically and for national security

The Application Space is Changing

Unstructured

Data

Structured

LexisNexis DAS

Graph

Emerging
Big Data
Problems

Human
Cortex
Simulation

Business
Intelligence

Unindexed SQL
(Netezza)

Streaming

HPC Physics Simulations

MapReduce

Traditional SQL

Loosely
Coupled

Computer Architecture

Tightly
Coupled

What about architecture research today?

ISCA 1973

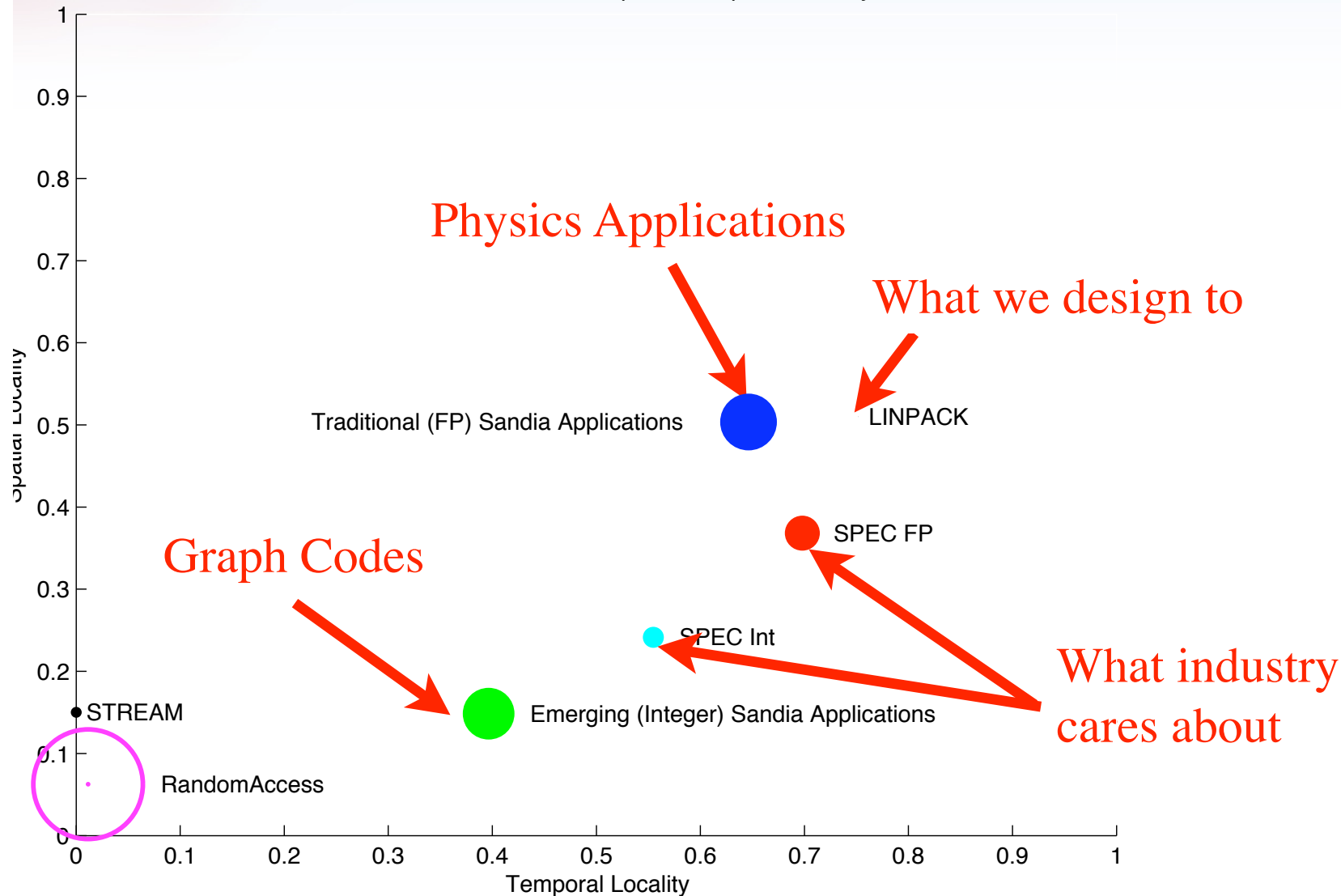Memory dominates 1/3 of cluster

ISCA 2008

Network, Power, Prediction
dominate > 50% of clusters

Bonus Material

Motivation: Emerging Applications are Different

Benchmark Suite Mean Temporal vs. Spatial Locality

From: Murphy and Kogge, *On The Memory Access Patterns of Supercomputer Applications: Benchmark Selection and Its Implications*, IEEE T. on Computers, July 2007

Example Graph Problem