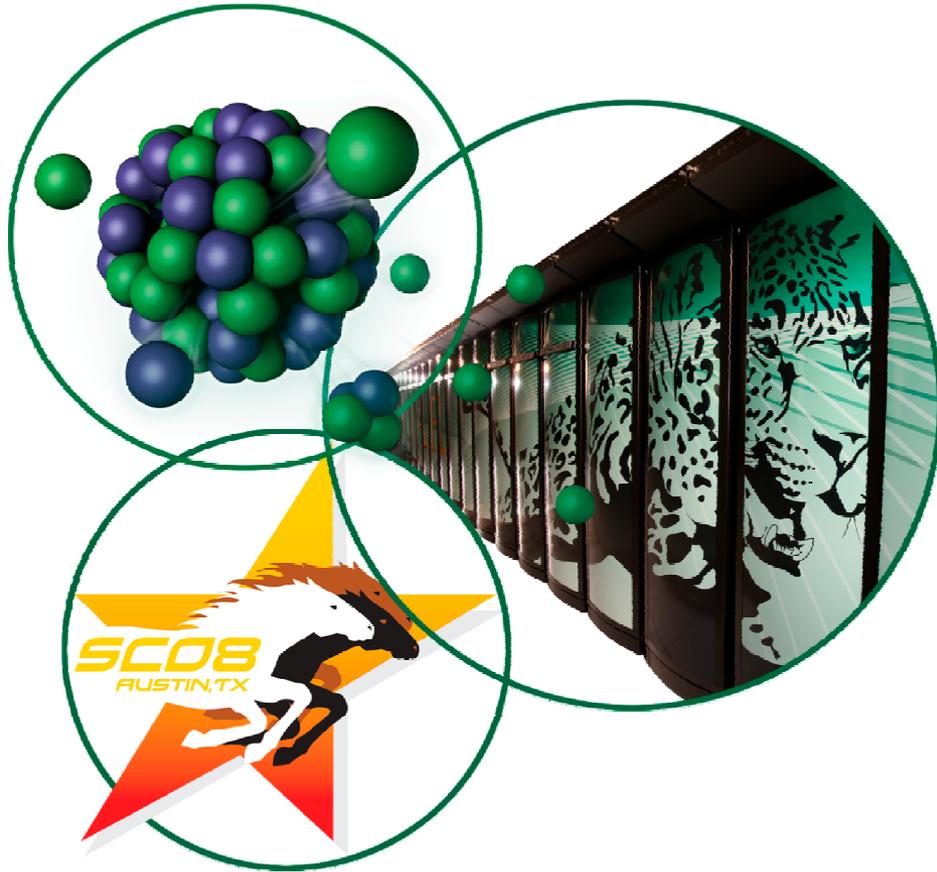


# Performance and Productivity of Emerging Architectures

Presented by

**Jeremy Meredith**  
**Sadaf Alam**  
**Jeffrey Vetter**

Future Technologies



# Overview

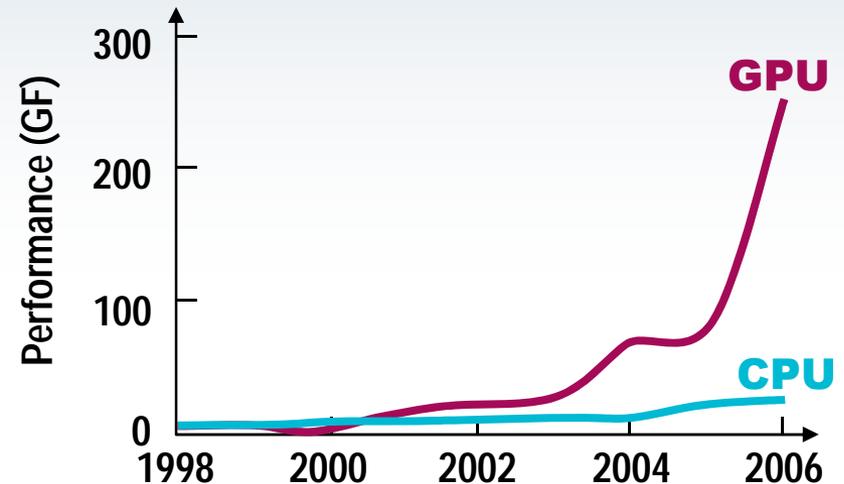
## Motivation

- Technologies on the horizon several orders of magnitude more powerful and power-efficient than microprocessors
- Grand challenge applications requiring several orders of magnitude more performance than now available

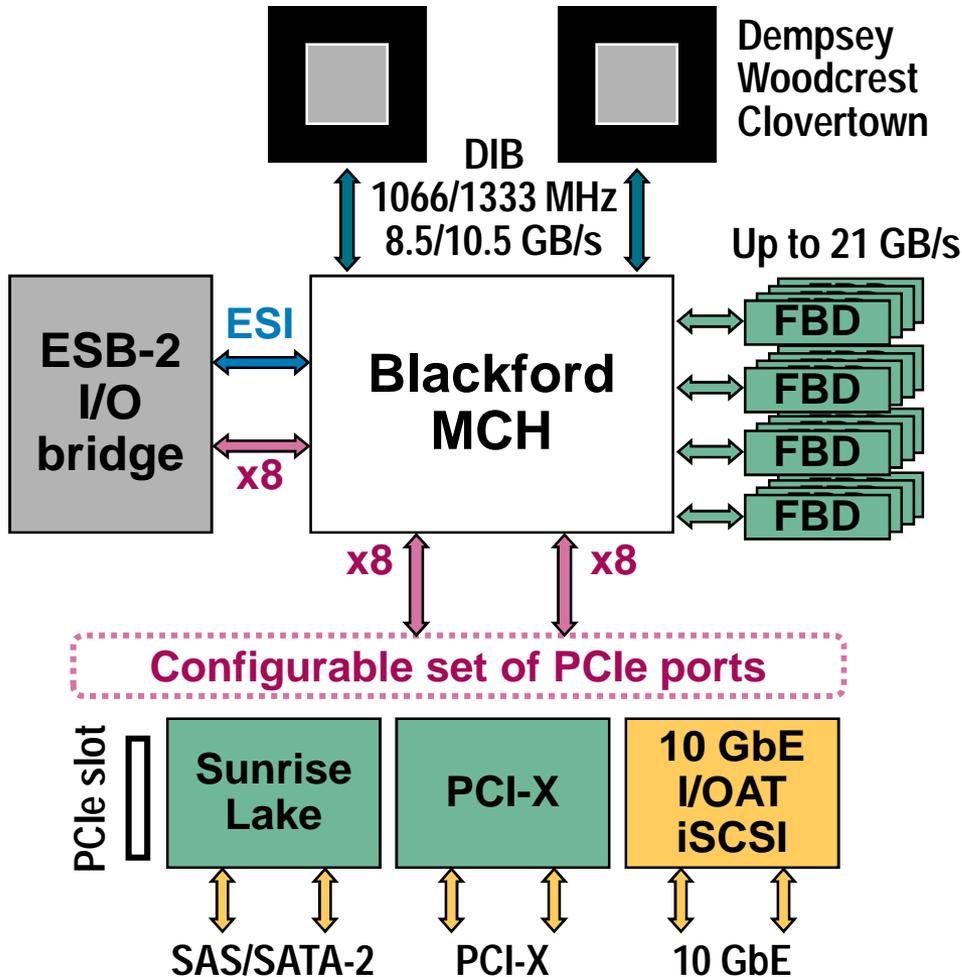


## Goals

- Gauge performance of emerging devices using high-level languages
- Estimate productivity with respect to contemporary microprocessing devices

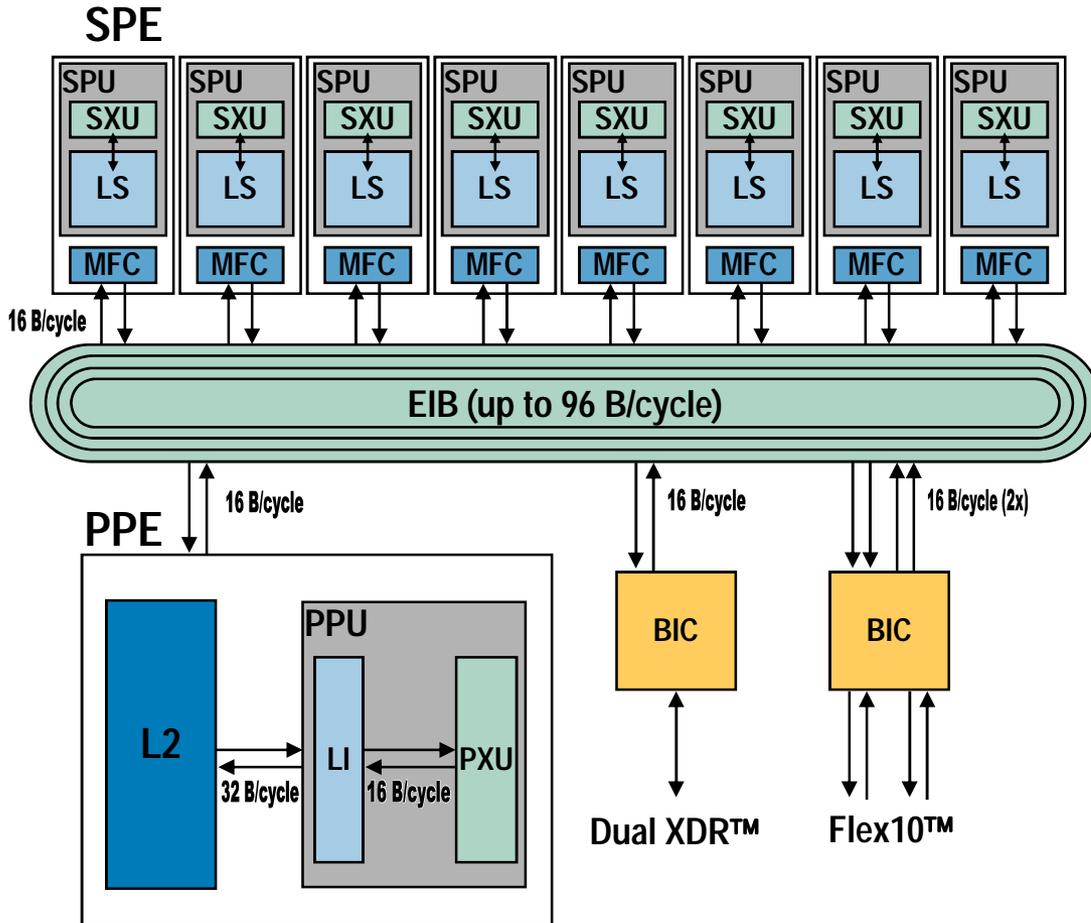


# Intel Xeon multicore CPU



- Dual-core (Woodcrest)
- Quad-core (Clovertown)
- Programming models
  - MPI
  - OpenMP
  - Pthreads
- Resource contention
  - Memory bandwidth
  - I/O bandwidth

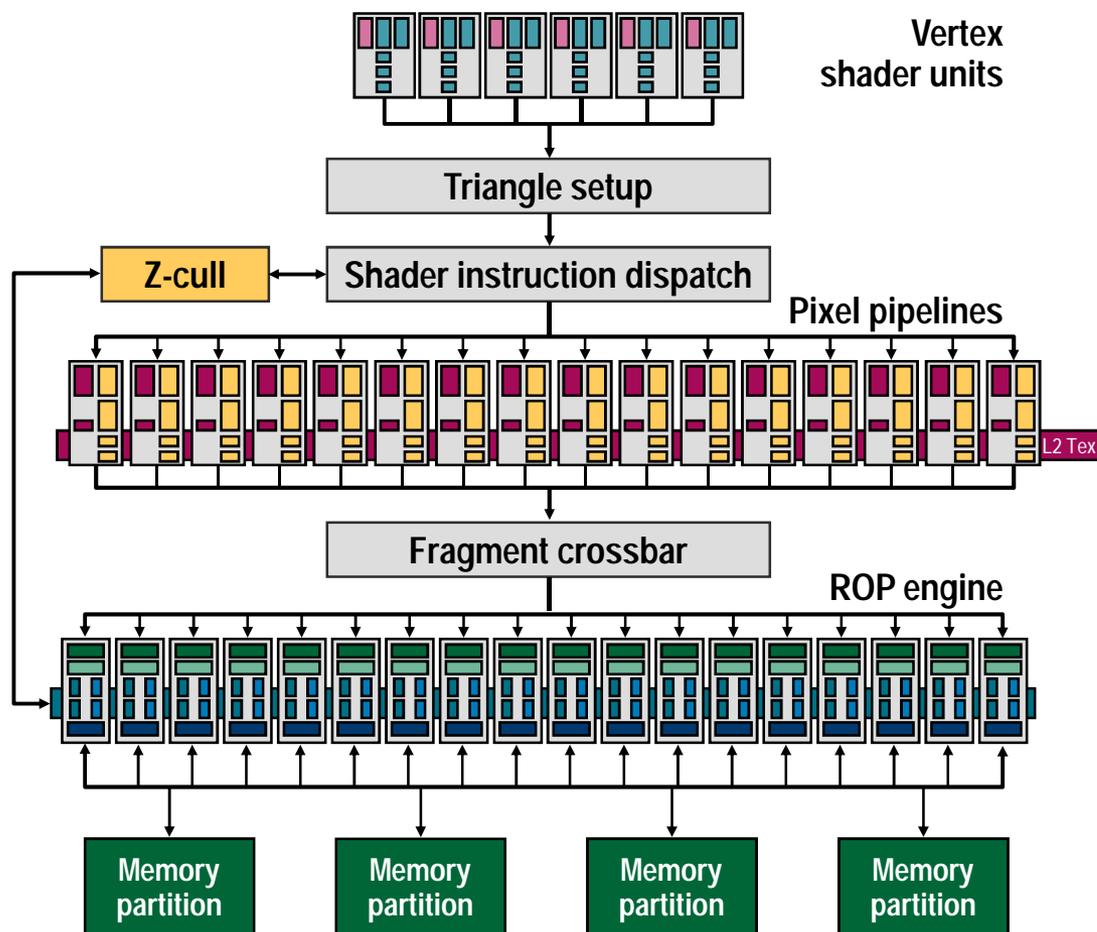
# IBM Cell broadband engine



- Cell heterogeneous multicore processor
  - 8 synergistic processing element (SPE) cores
  - 200 GFLOPS at 3.2 GHz
  - 200-GB/s memory bandwidth
- Programming models
  - Multisource compilation
  - Threadlike launch semantics for SPEs

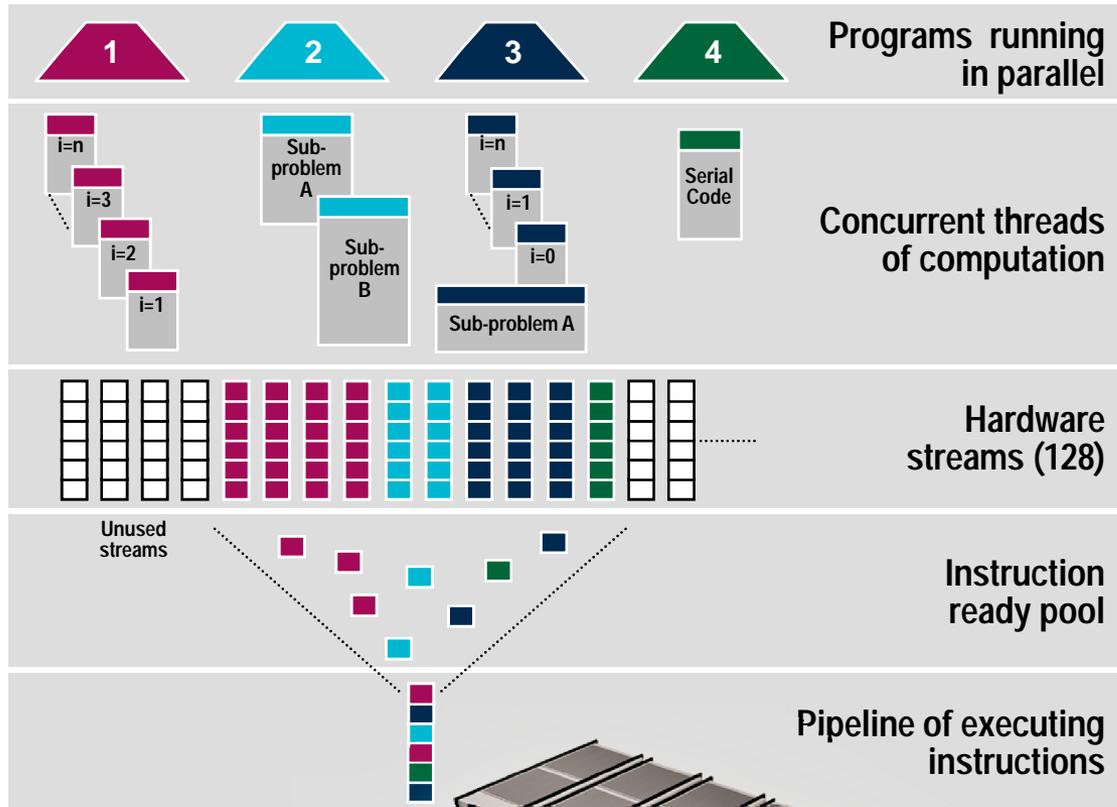
Source: M. Gschwind et al., Hot Chips-17, August 2005

# Graphics processing unit (GPU)



- **NVIDIA 7900GTX**
  - 24 SIMD pixel pipelines
  - 200 GFLOPS at 650 MHz
  - 50-GB/s memory bandwidth
- **Programming models**
  - Multiple-source compilation
  - Gather semantics define parallelism
  - Host CPU drives program setup and data transfer

# Extreme multithreading with Cray MTA-II

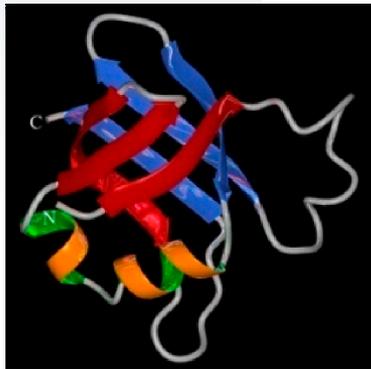


- Cray XMT system
  - MTA-I and MTA-II processor architecture
  - XT3/4 scalable infrastructure
  - AMD Torrenza technology
- Fine-grain multithreading (128 concurrent threads)
- Uniform memory hierarchy in MTA-I and MTA-II

# Application kernels

## Molecular dynamics (MD) calculations

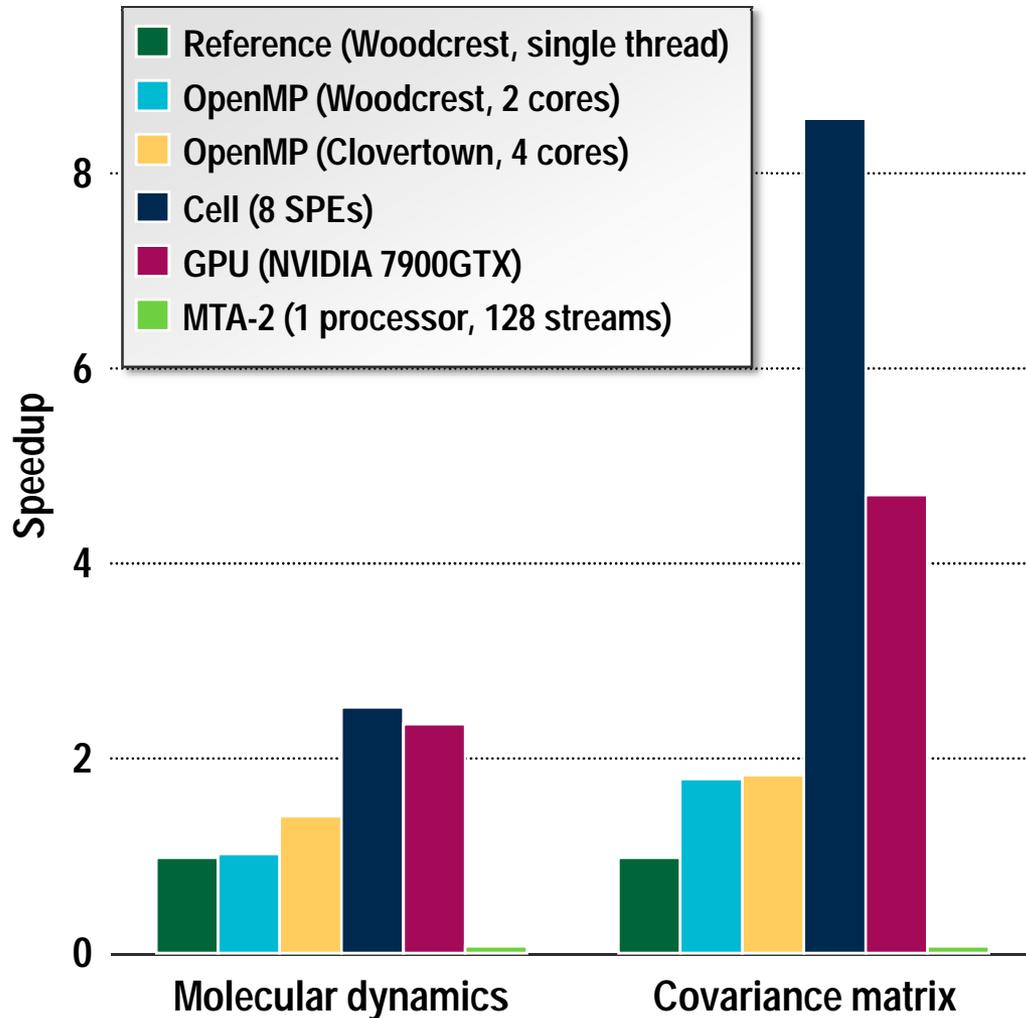
- Applications in biology, chemistry, and materials
- Newton's second law of motion
- Bonded calculations
- Nonbonded electrostatic and van der Waals forces
- Workload characteristics
  - Floating-point intensive
  - Irregular memory access patterns



## Covariance matrix calculation

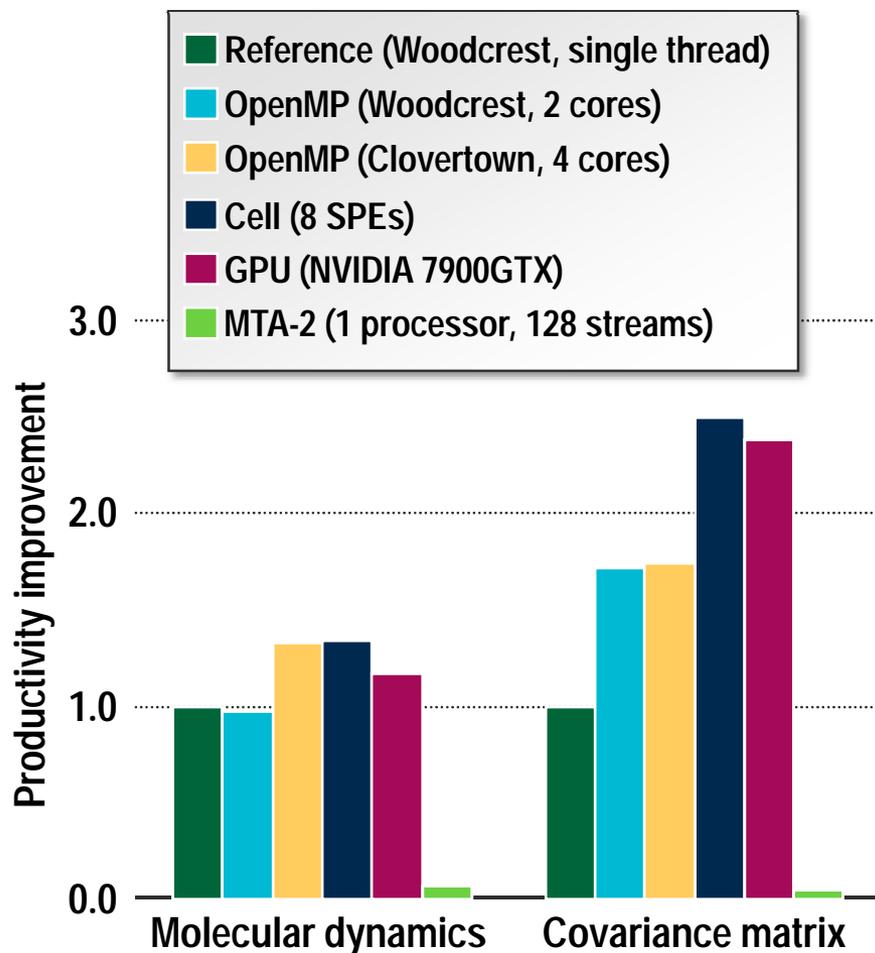
- Applications in hyperspectral imaging and machine learning
- Determines covariance among samples in data
- Workload characteristics
  - Memory-bandwidth and floating-point intensive
  - Regular memory access patterns

# Performance



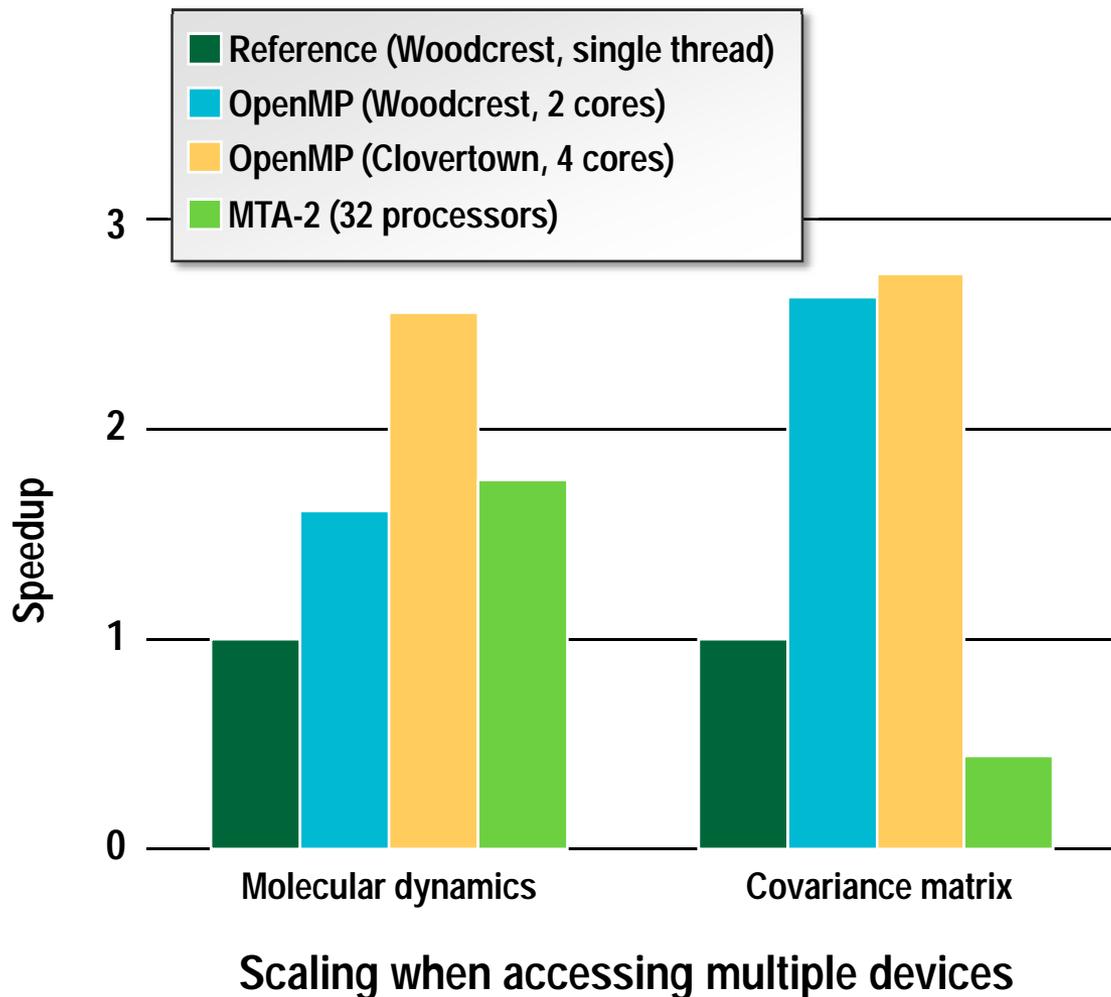
- **OpenMP improves performance moderately on commodity CPUs**
- **High parallelism of Cell and GPU can improve performance, but more so when memory access is regular**

# Productivity



- **Despite small performance increases through OpenMP, the ease of using it means that productivity can still be increased**
- **Conversely, the high speed of Cell and GPU means that even substantial effort results in higher productivity**

# Productivity scaling



**Increased parallelism from all architectures with little increased effort results in higher productivity**

# Contacts

## **Jeremy Meredith**

**Future Technologies Group**

**(865) 241-5842**

**jsmeredith@ornl.gov**

## **Sadaf Alam**

**Future Technologies Group**

**(865) 241-1533**

**alamrs@ornl.gov**

## **Jeffrey Vetter**

**Future Technologies Group**

**(865) 576-7115**

**vetter@ornl.gov**