

# Performance Opportunities and Obstacles for Earth System Prediction

James B White III (Trey)  
trey@ucar.edu

Earth System Prediction Capability Workshop  
NOAA Earth System Research Laboratory  
September 8, 2010



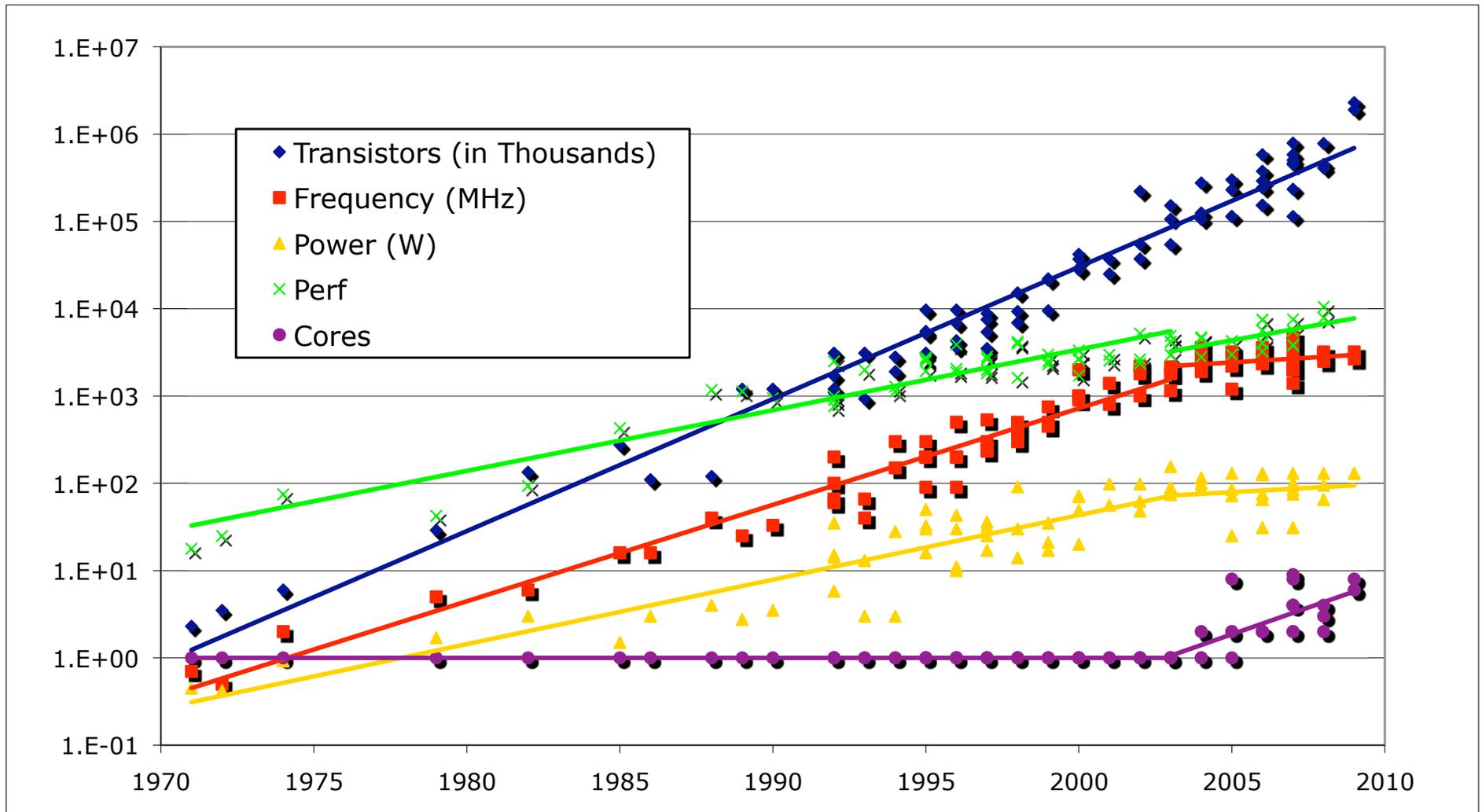
U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science

# Performance Opportunities and Obstacles for Earth System Prediction

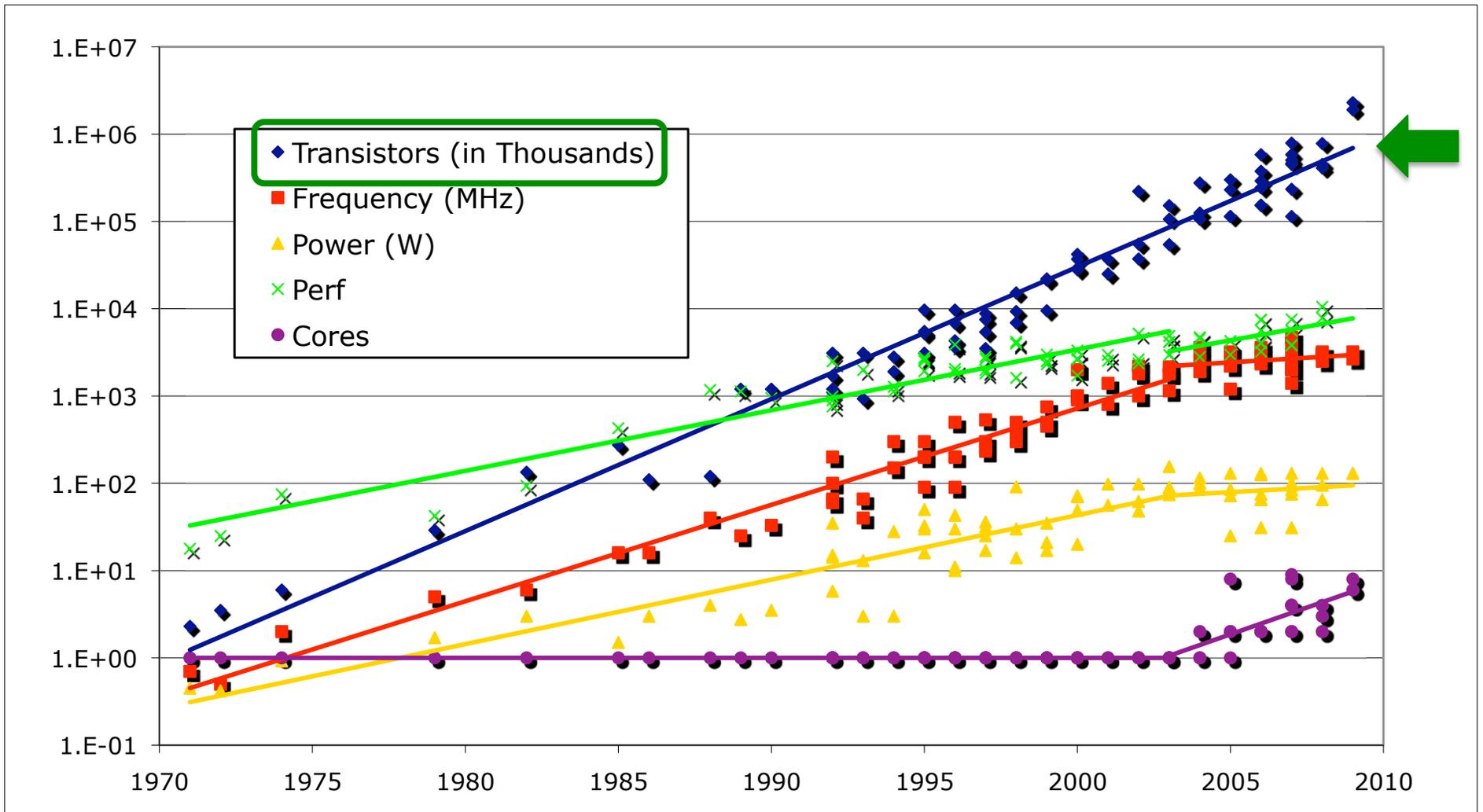
- Computer trends
- Portable-performance engineering
- Workflows
- Scalability vs. throughput
- Opportunities and obstacles

# Kathy Yelick's Processor Trends



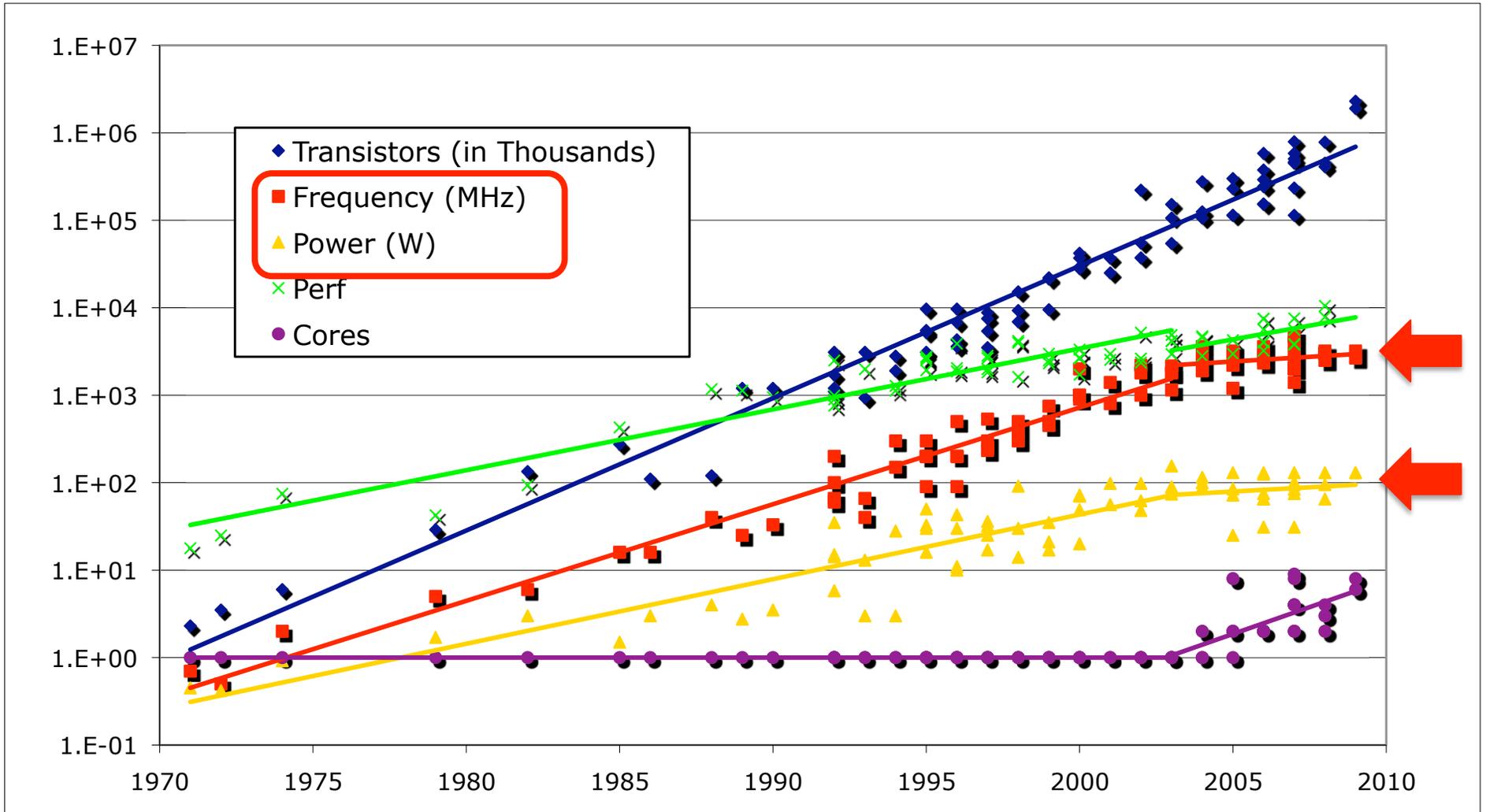
*From Kathy Yelick's "Ten Ways to Waste a Parallel Computer", using data from Kunle Olukotun, Lance Hammond, Herb Sutter, Burton Smith, Chris Batten, and Krste Asanovic*

# Moore's Law Continues



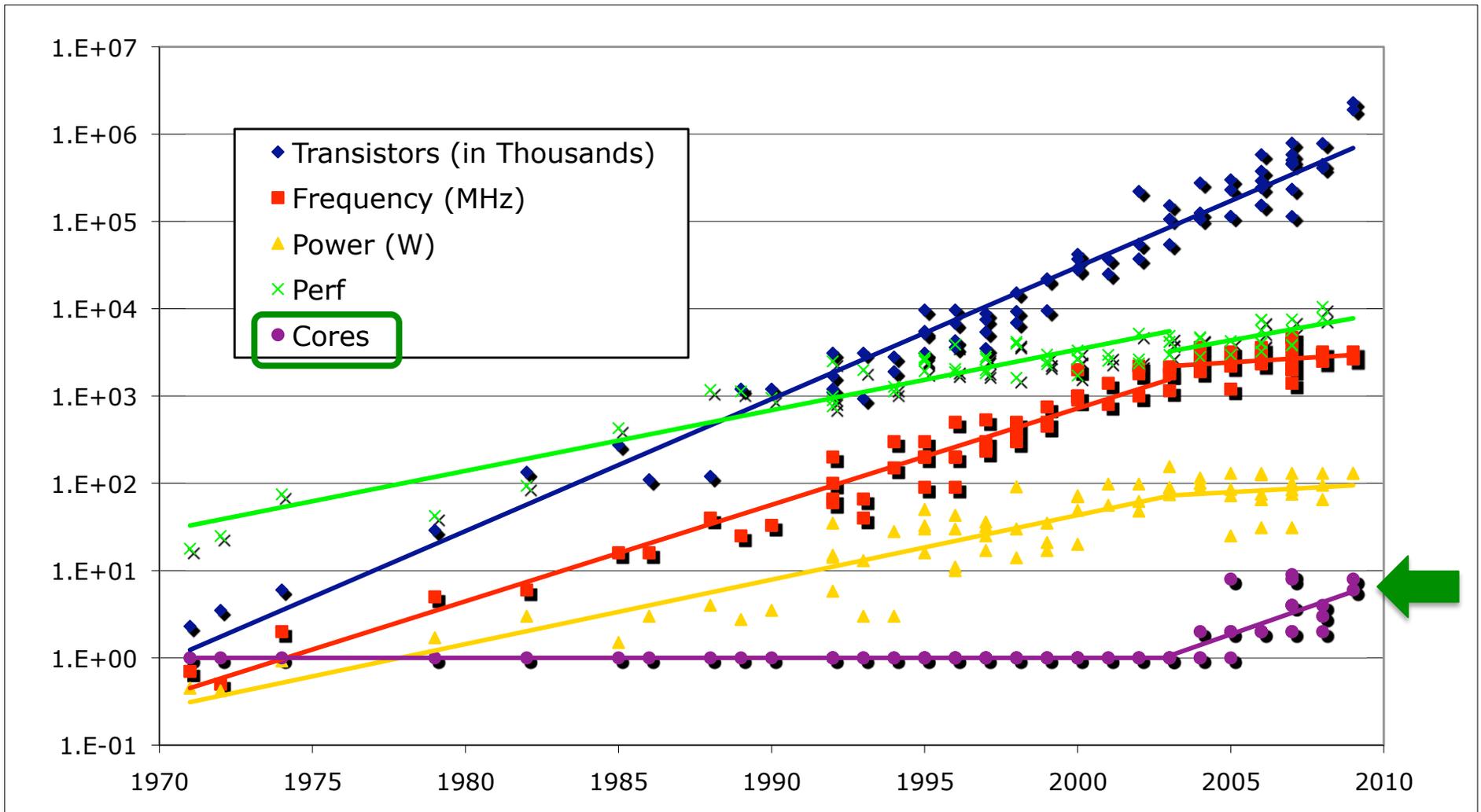
*From Kathy Yelick's "Ten Ways to Waste a Parallel Computer", using data from Kunle Olukotun, Lance Hammond, Herb Sutter, Burton Smith, Chris Batten, and Krste Asanovic*

# Power & Frequency Stagnating



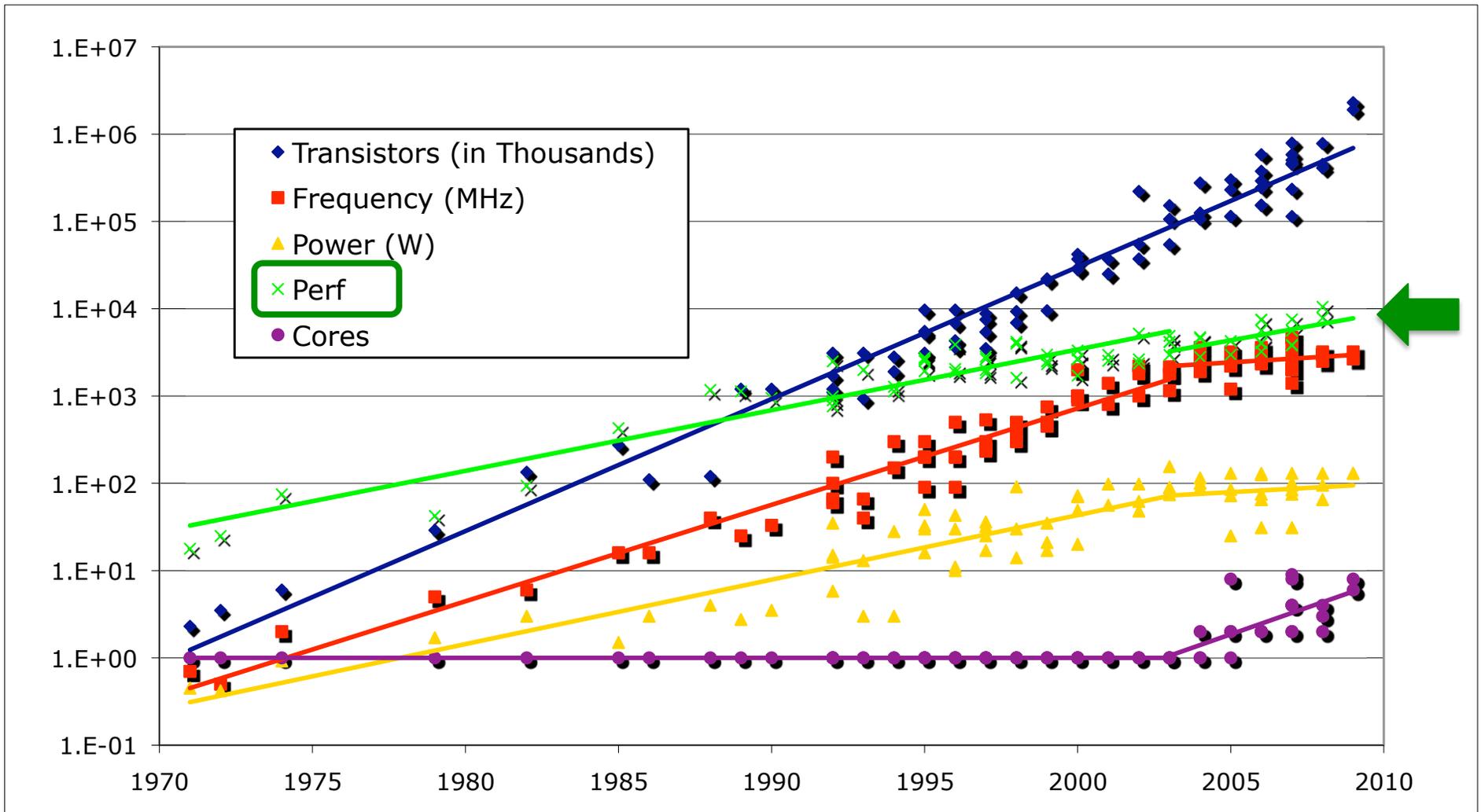
*From Kathy Yelick's "Ten Ways to Waste a Parallel Computer", using data from Kunle Olukotun, Lance Hammond, Herb Sutter, Burton Smith, Chris Batten, and Krste Asanovic*

# Cores Per Socket Increasing



*From Kathy Yelick's "Ten Ways to Waste a Parallel Computer", using data from Kunle Olukotun, Lance Hammond, Herb Sutter, Burton Smith, Chris Batten, and Krste Asanovic*

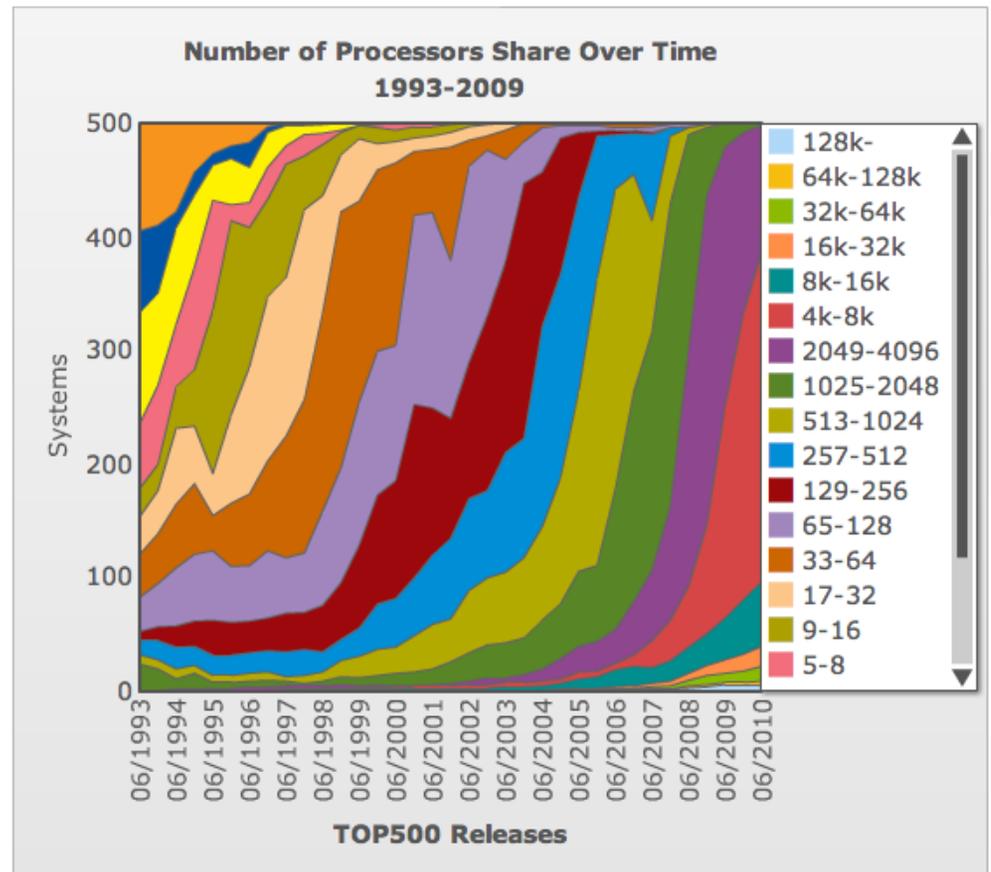
# Performance Increasing (Some)



*From Kathy Yelick's "Ten Ways to Waste a Parallel Computer", using data from Kunle Olukotun, Lance Hammond, Herb Sutter, Burton Smith, Chris Batten, and Krste Asanovic*

# Scaling Out

- More cores
- More memory
- More performance
- More space
- More power



*Top500.org*

# Scaling Out → Scaling In

- More transistors
- More computational need
- Limited space
- Limited power

# Scaling In

- More cores per chip
- More threads per core
- Longer vector registers (2→4+ doubles)
- Block multithreading (GPUs)
- Heterogeneity on a chip
  - IBM Cell
  - CPU & GPU
- Power efficiency
- **More architectural uncertainty**

# Performance Opportunities and Obstacles for Earth System Prediction

- Computer trends
- **Portable-performance engineering**
- Workflows
- Scalability vs. throughput
- Opportunities and obstacles

# Portable-Performance Engineering of Community Earth System Model

- Block-oriented computation
- Hybrid parallelism
- Modular parallel communication
- Flexible task parallelism
- Modular built-in timers

# Block-Oriented Computation

- Pass blocks as procedure arguments
- Operate on a block at a time
- More than one element
  - Vectorization (SSE, AltiVec, double hummer, *etc.*)
  - Pipelining
  - Loop unrolling
- Less than whole domain
  - Cache blocking
  - Load balancing
- Tunable block size

# Hybrid Parallelism

- MPI and OpenMP
- OpenMP can target different parallelism
  - 3<sup>rd</sup> dimension in 2D decomposition
  - Too tightly coupled for distributed memory
  - Allows use of more cores
- Or same parallelism
  - Aggregate MPI messages
  - Fewer, larger messages can be more efficient
- Tunable number of threads/task

# Modular Parallel Communication

- Isolate parallel communication
- Allow different programming models
  - MPI, Co-Array Fortran, SHMEM
- Allow different algorithms
  - Performance tuning
  - Workarounds for system limitations

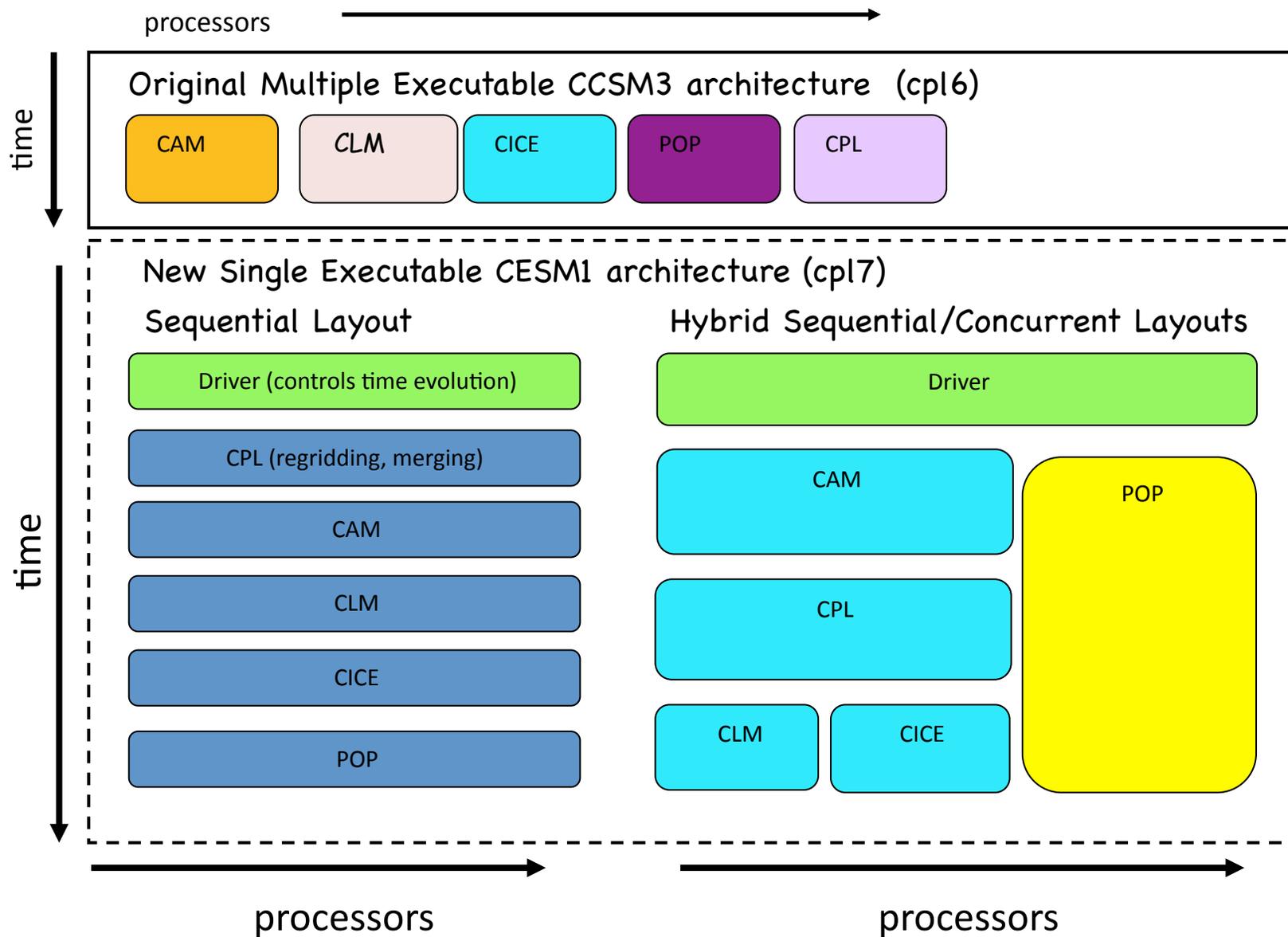
# Different Algorithms

- Sends before receives or vice versa
- Reproducible dot products
- Load balancing options:  
on task, on node, nearby nodes, global
- Flow control (critical on largest computers)

# Modular Parallel I/O

- Tunable number and location of I/O tasks
- Choices for underlying implementation:  
NetCDF3, NetCDF4, pNetCDF, binary
- Potential for asynchronous I/O
- Potential for in-memory checkpoint/restart

# Flexible Task Parallelism



# Modular Built-In Timers

- Portable performance tuning
  - Don't depend on vendor tools
  - Use newest computers
- Configurable level of detail
  - Load balance components
  - Choose tuning parameters for a given component
  - Find tuning “opportunities” and performance bugs
  - Optionally report hardware counters (PAPI)
- Opportunity for automatic tuning

# Performance Opportunities and Obstacles for Earth System Prediction

- Computer trends
- Portable-performance engineering
- **Workflows**
- Scalability vs. throughput
- Opportunities and obstacles

# IPCC AR5/CMIP5 Workflow

- Simulations output history files
  - Each file has many fields at a given time
- Post-processing scripts generate time series
  - One field over a long time
- Diagnostics use time series
- Earth System Grid hosts history and time series files
- Scientists worldwide download files and perform analysis

# Workflow Obstacles

- **Much** more data
  - Multiple PB for CMIP5
  - Higher resolution
  - More physical, chemical, and biological processes
  - More kinds of simulations
- Petascale simulation, gigascale analysis tools
- Unforeseen types of analysis
  - When in doubt, output

# Workflow Opportunities

- Parallel I/O
- Asynchronous I/O
- Output time series directly from simulation
- Parallel analysis tools
- Remote analysis tools

# Operational Workflow Opportunities

- Optimize number & frequency of output fields
- Perform analysis “near” simulation
- Move analysis into simulation

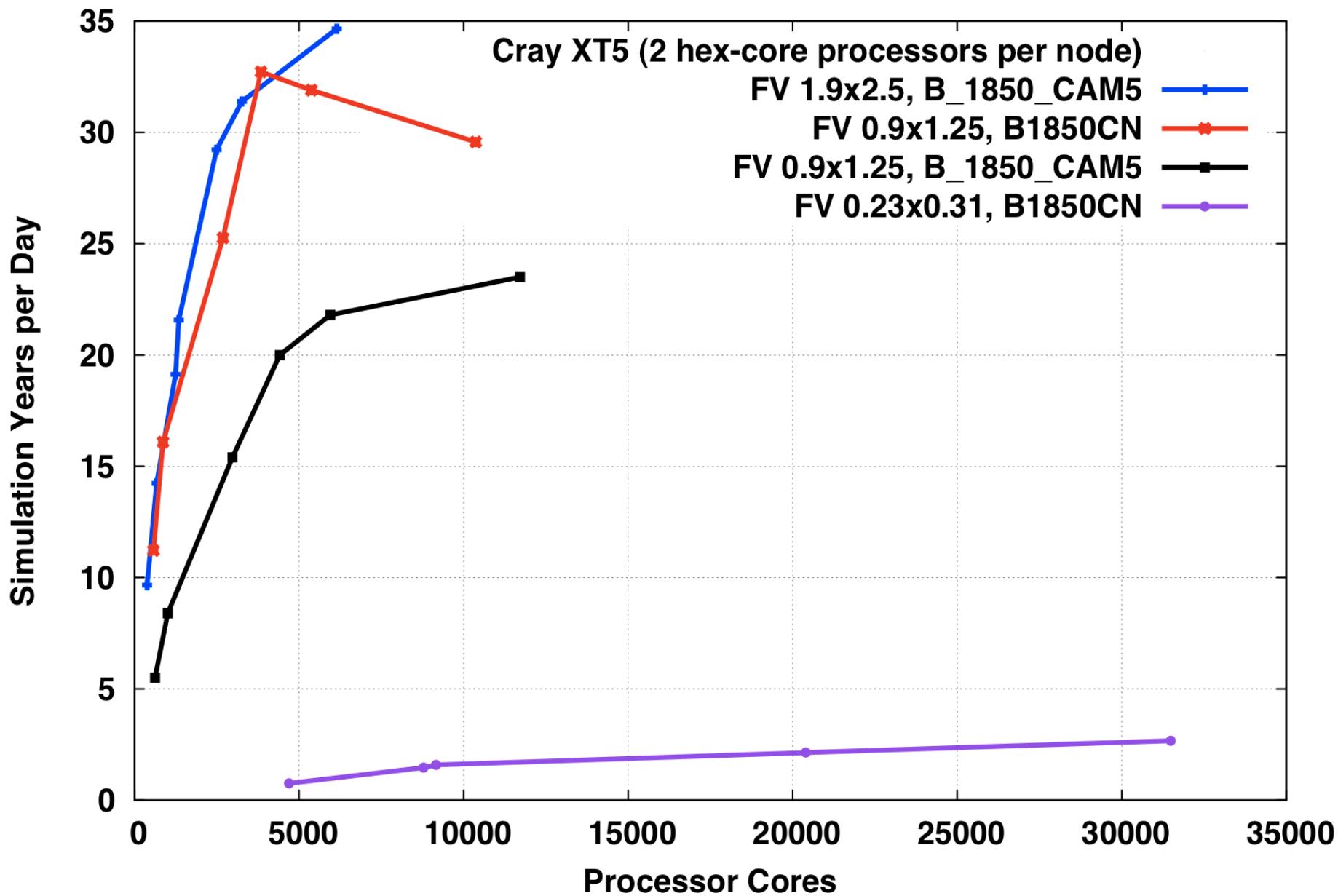
# Performance Opportunities and Obstacles for Earth System Prediction

- Computer trends
- Portable-performance engineering
- Workflows
- **Scalability vs. throughput**
- Opportunities and obstacles

# Scaling Out CESM

- Flexible, hybrid parallelism
- Parallel I/O
- More-scalable atmosphere grids and dynamics
- Higher resolution
- More physical processes per grid point
  - Higher computational intensity

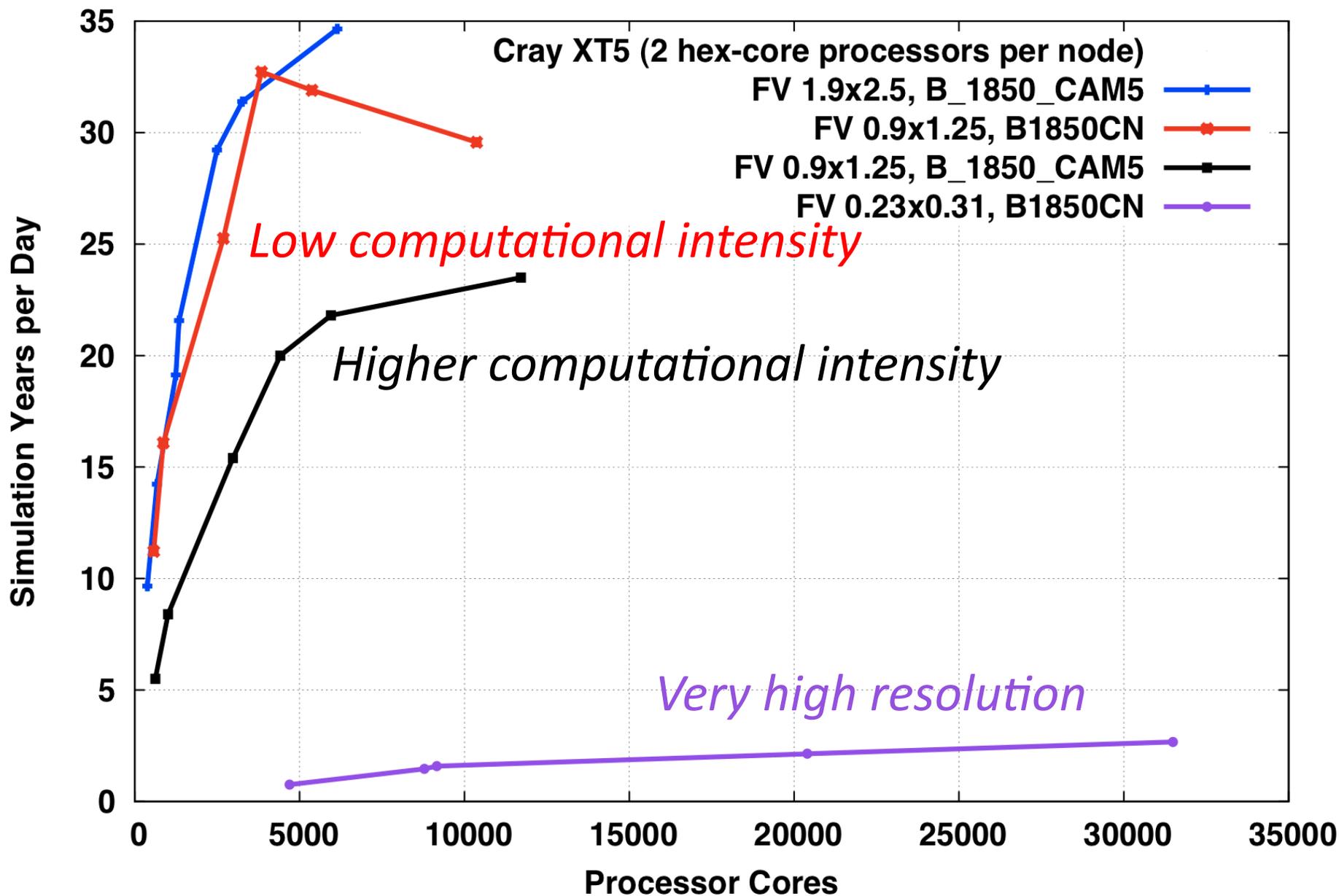
# CESM Performance



*Courtesy of Pat Worley, ORNL*

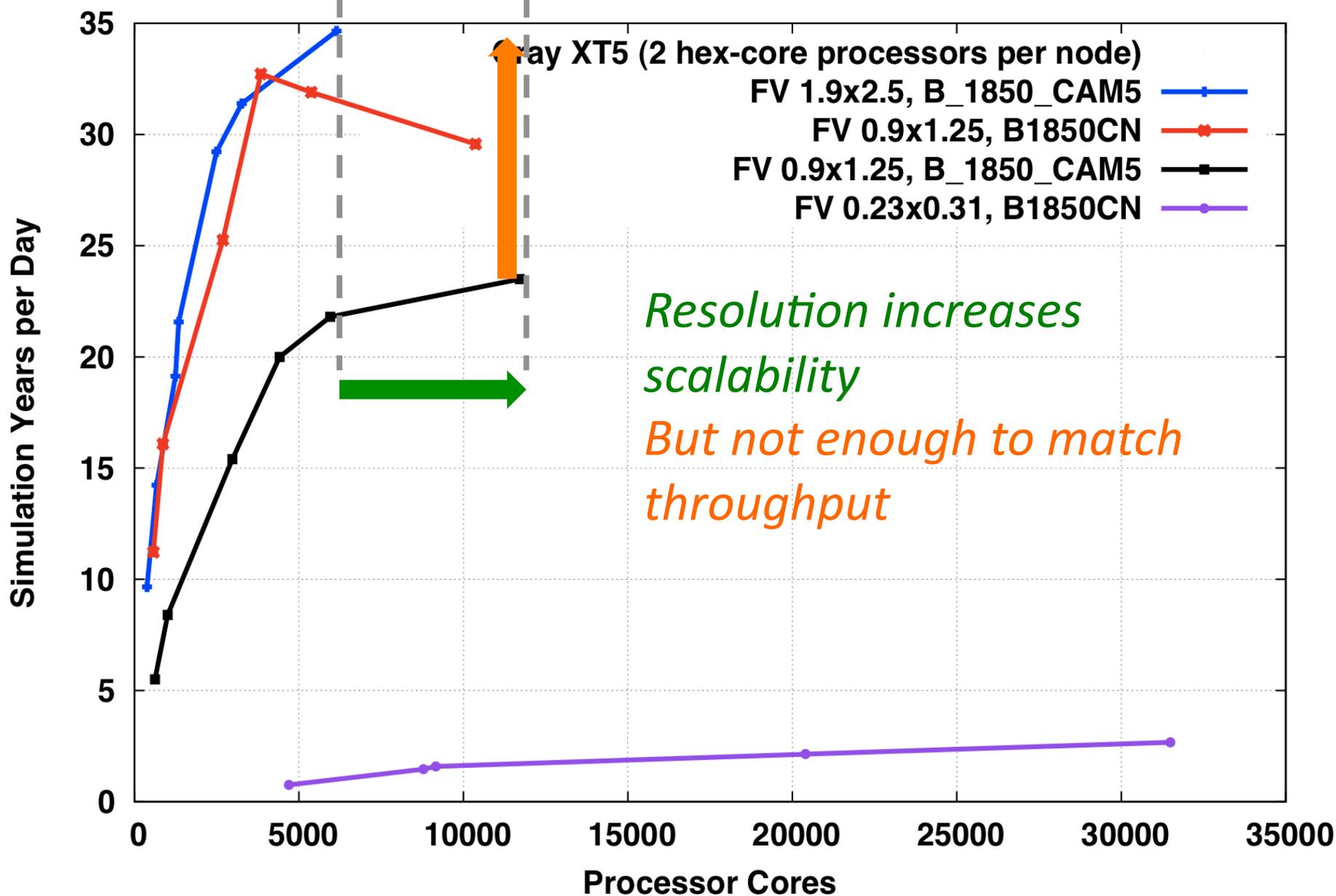
*Low resolution*

## CESM Performance



Courtesy of Pat Worley, ORNL

# CESM Performance



Courtesy of Pat Worley, ORNL

# Time-Integration Barrier

- Explicit methods
  - Scalable and cheap per time step
  - Resolution goes up, time step must go down
  - Single-thread performance no longer improving
- Implicit methods
  - Stable for large time steps
  - Expensive: linear and often nonlinear solvers
  - Less scalable: global reductions, latency bound

# Performance Opportunities and Obstacles for Earth System Prediction

- Computer trends
- Portable-performance engineering
- Workflows
- Scalability vs. throughput
- **Opportunities and obstacles**

# Performance Obstacles

- Stagnant single-thread performance
- Stagnant memory and communication latency
- Computer architectural uncertainty
- Software complexity
- Time-integration barrier
- I/O and metadata volume

# Performance Opportunities

- Increasing aggregate computing power
- Portable-performance engineering
  - Block-oriented computation
  - Hybrid, flexible parallelism
  - Modular parallel communication and timers
- Targeted I/O and/or embedded analysis
- Algorithms

## Questions?

# Performance Opportunities

- Increasing aggregate computing power
- Portable-performance engineering
  - Block-oriented computation
  - Hybrid, flexible parallelism
  - Modular parallel communication and timers
- Targeted I/O and/or embedded analysis
- Algorithms

*I thank the US Department of Energy Office of Biological and Environment Research for financial support, the workshop organizers for inviting me, and Mariana Vertenstein, Pat Worley, Kathy Yelick, and the Top500 maintainers for their content. Finally, I thank Warren Washington for his support and encouragement.*