SciDAC-3 Institute for Sustained Performance, Energy, and Resilience

Bob Lucas
University of Southern California
Sept 23, 2011

SUPER

SciDAC

Scientific Discovery through Advanced Computation

DOE Office of Science's program for trans-petascale computational science

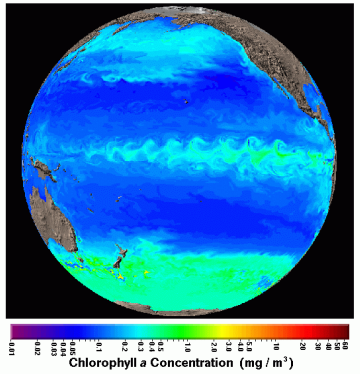
Maximizing performance is getting increasingly difficult:

- **Systems are more complicated**
  - **O(100K) multi-core CPUs**
  - **GPU accelerators**
- **Codes are more complicated**
  - **Multi-disciplinary**
  - **Multi-scale**

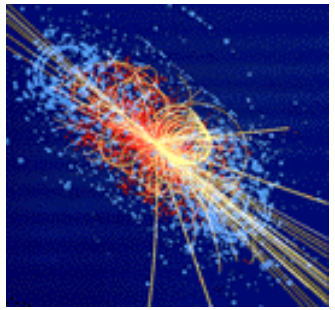
IBM BlueGene at LLNL

Cray XT5

ORNL NCCS



POP model of El Nino



BeamBeam3D
accelerator modeling

SciDAC
Scientific Discovery
through
Advanced Computing

SUPER

SciDAC Performance Efforts

SciDAC has always had an effort focused on performance

Performance Evaluation Research Center (PERC)

- **Benchmarking, modeling, and understanding**

Performance Engineering Research Institute (PERI)

- **Performance engineering, modeling, and engagement**
- **Three SciDAC-e projects**

Institute for Sustained Performance, Energy, and Resilience (SUPER)

- **Performance engineering**
- **Energy minimization**
- **Resilient applications**

SUPER

SUPER Team

ANL
Paul Hovland
Boyana Norris
Stefan Wild

LBNL
David Bailey
Lenny Oliker
Sam Williams

LLNL
Bronis de Supinski
Daniel Quinlan

Oregon
Allen Malony
Sameer Shende

ORNL
Gabriel Marin
Philip Roth
Patrick Worley

UCSD
Laura Carrington

UMD
Jeffrey Hollingsworth

UNC
Rob Fowler
Allan Porterfield

USC
Jacque Chame
Robert Lucas (PI)

UTK
Shirley Moore
Dan Terpstra

Utah
Mary Hall
Chun Chen

SciDAC
Scientific Discovery through Advanced Computing

SUPER

Broadly Based Effort

- **Twenty-seven people at the kick-off meeting**
  - **University of Oregon, Sept 20-21**
  - **Not everybody made it**
- **All PIs have independent research projects**
  - **SUPER money alone isn't enough to support any of its investigators**
  - **SUPER leverages other work and funding**
- **SUPER contribution is integration, results beyond any one group**
  - **Follows successful PERI model (tiger teams and autotuning)**
  - **Collaboration extends to others having similar research goals**
    - **Already talking to LANL and Rice (both were invited to our meeting at Oregon)**
    - **Other likely collaborators include PNNL, Portland State, and UT San Antonio**
    - **Perhaps Juelich and Barcelona too?**

SciDAC
Scientific Discovery through Advanced Computing

SUPER

Management Structure

- **Overall management**
  - **Bob Lucas and David Bailey**
- **Distributed leadership of research**
  - **Follows PERI model, adapts as needed**
- **Weekly project teleconferences**
  - **Wednesdays, noon Eastern**
  - **All hands every four weeks (management and planning)**
  - **Technically focused otherwise**
- **Regular face-to-face project meetings**
  - **Monday mornings at SC**
  - **All hands, twice per year, each institution takes a turn hosting**
    - **Allows students and staff to attend at least once**

SUPER Meeting Schedule

Oregon	**Sept. 21-22, 2011**
UNC RENCI	**March 29-30, 2012**
ANL	**Sept. 2012**
UTK ICL	**March 2013**
LBNL	**Sept. 2013**
Utah	**Feb. 2014**
Maryland	**Sept. 2014**
UCSD SDSC	**March 2015**
ORNL	**Sept. 2016**
USC ISI	**March 2017**

SciDAC
Scientific Discovery through Advanced Computing

SUPER

SUPER Objectives

Automatic performance tuning

Energy minimization

Resilient computing

Optimization of the above

Application engagement

Tool integration

Outreach and tutorials

SciDAC
Scientific Discovery through Advanced Computing

SUPER

Performance Engineering (1)

Measurement and monitoring

- **Adopting University of Oregon's TAU system**
- **Still plan to collaborate with Rice and its HPCToolkit**

Performance Database

- **Extending TAU's PerfDMF to enable online collection and analysis**

Performance modeling

- **PBound and Roofline models to bound performance expectations**
- **MIAMI to model impact of architectural variation**
- **PSINS to model communication**

SciDAC Scientific Discovery through Advanced Computing

SUPER

Performance Engineering (2)

- **Automatic tuning for performance portability**
  - **Led by Mary Hall, University of Utah**
- **Extend PERI autotuning system for future architectures**
  - **New TAU front-end for triage**
  - **CUDA-CHiLL to target GPUs**
  - **OpenMP-CHiLL for SMP multicores**
  - **Active Harmony provides search engine**
    - **Drive empirical autotuning experimentation**
    - **Balance threads and MPI ranks in hybrids of OpenMP and MPI**
    - **Extend to surface/area to volume, or halo size, experiments**
  - **Targeted Autotuning**
    - **Similar in spirit to Domain Specific Langauges**
    - **Users write simple code and leave the tuning to us**
  - **Whole program autotuning**
    - **Parameters, algorithm choice, libraries linked, etc.**

The SUPER Autotuning Framework

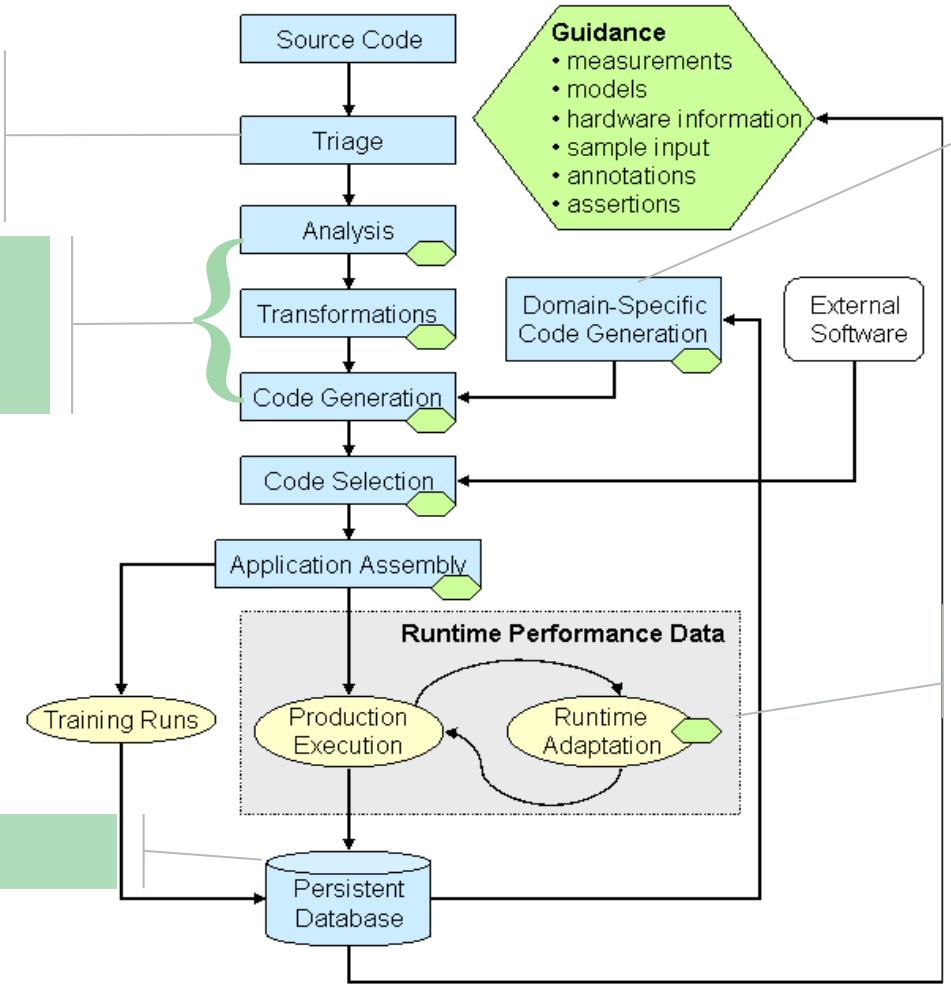
TAU (Oregon)
HPC Toolkit (Rice)
ROSE (LLNL)

CHiLL (USC/ISI and Utah)
ROSE (LLNL)
Orio (Argonne)

PerfTrack (LBNL, SDSC, RENCI)

OSKI (LBNL)

Active Harmony (UMD)
GCO (UTK)



SciDAC
Scientific Discovery through Advanced Computing

SUPER

First PERI Autotuning Success SMG2000

SMG2000: Semicoarsening multigrid code, used for various applications, including modeling of groundwater diffusion.

PERI researchers integrated several tools, then developed a "smart" search technique to find an optimal tuning strategy among 581 million different choices.

Achieved 2.37X performance improvement on one key kernel.

Achieved 27% overall performance improvement.

SciDAC
Scientific Discovery through Advanced Computing

SUPER

Autotuning the central SMG2000 kernel

Outlined code (from ROSE outliner)

```
for (si = 0; si < stencil_size; si++)
   for (kk = 0; kk < hypre__mz; kk++)
      for (jj = 0; jj < hypre__my; jj++)
         for (ii = 0; ii < hypre__mx; ii++)
            rp[((ri+ii)+(jj*hypre__sy3))+(kk*hypre__sz3)] -=
               ((Ap_0[((ii+(jj*hypre__sy1))+ (kk*hypre__sz1))+
                (((A->data_indices)[i])[si])])*
                (xp_0[((ii+(jj*hypre__sy2))+(kk*hypre__sz2))+(( *dxp_s)[si])]));
```

CHiLL transformation recipe

permute([2,3,1,4])
tile(0,4,TI)
tile(0,3,TJ)
tile(0,3,TK)
unroll(0,6,US)
unroll(0,7,UI)

Constraints on search

$0 \le TI, TJ, TK \le 122$
$0 \le UI \le 16$
$0 \le US \le 10$
compilers $\in$ {gcc, icc}

Search space:

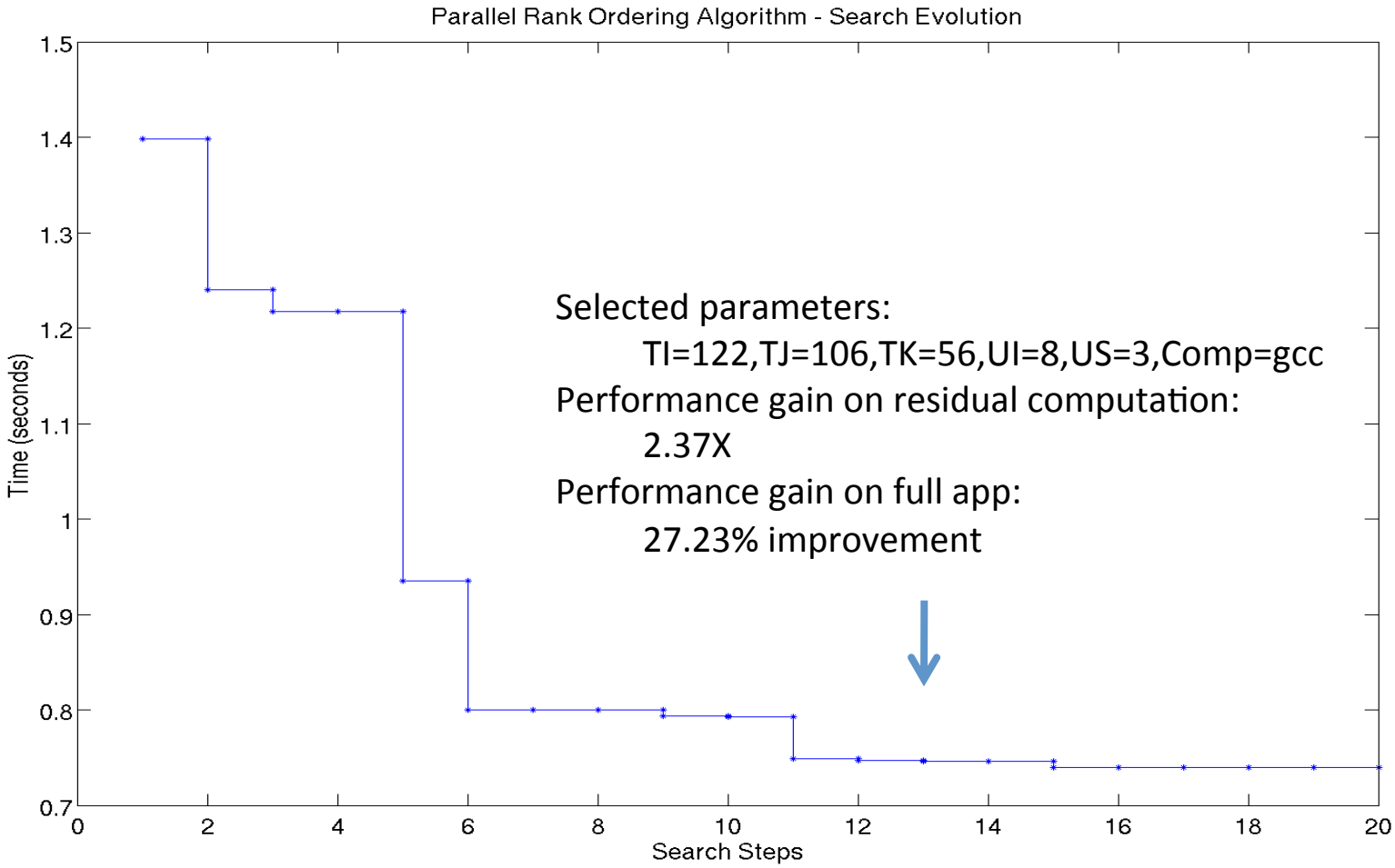
$122^3 \times 16 \times 10 \times 2 = 581{,}071{,}360$ points

SciDAC Scientific Discovery through Advanced Computing

SUPER

Early Result from PERI

Parallel search evaluates 490 points and converges in 20 steps

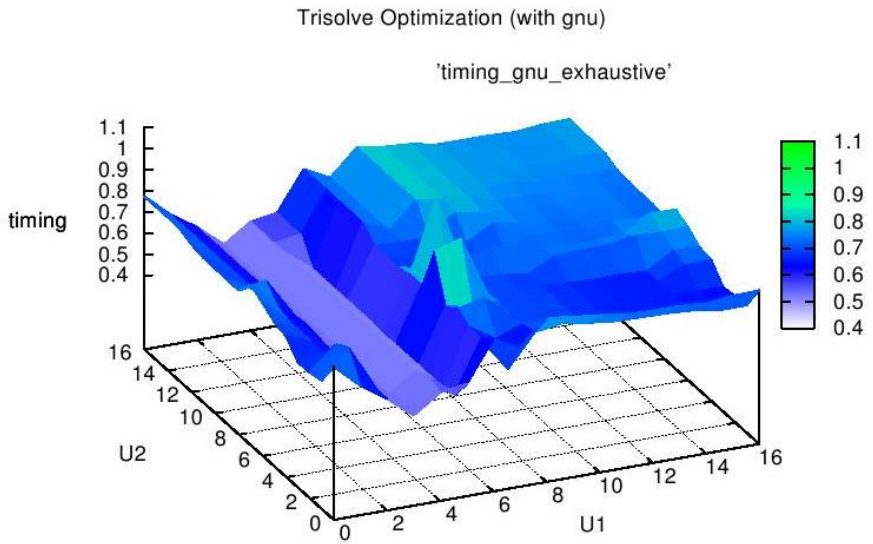


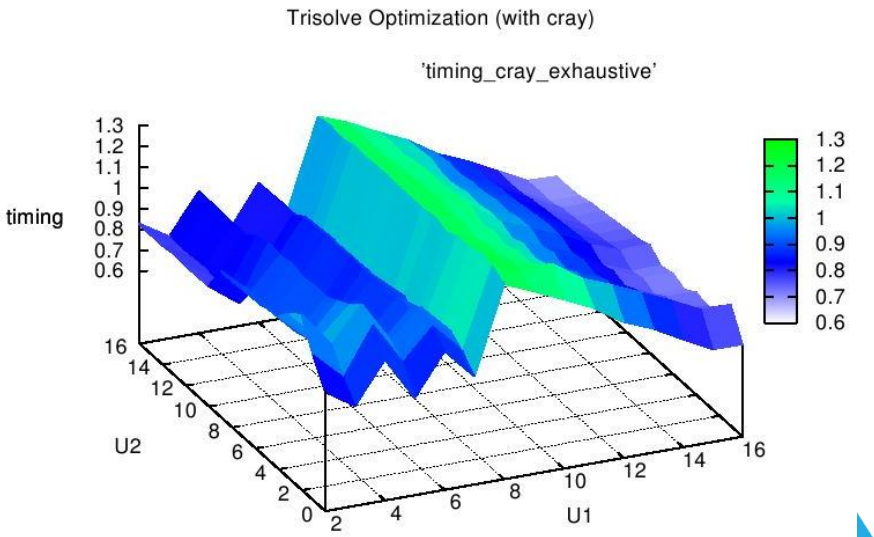
Scientific Discovery through Advanced Computing

SUPER

Autotuning the triangular solve kernel of the Nek5000 turbulence code

Compiler	Original	Active Harmony			Exhaustive		
	Time	Time	(u1,u2)	Speedup	Time	(u1,u2)	Speedup
pathscale	0.58	0.32	(3,11)	1.81	0.30	(3,15)	1.93
gnu	0.71	0.47	(5,13)	1.51	0.46	(5,7)	1.54
pgi	0.90	0.53	(5,3)	1.70	0.53	(5,3)	1.70
cray	1.13	0.70	(15,5)	1.61	0.69	(15,15)	1.63





Scientific Discovery through Advanced Computing

SUPER

Energy Minimization

Led by Laura Carrington, University of California at San Diego

Develop new energy aware APIs for users

- **I know the processor on the critical path in my multifrontal code**

Obtain more precise data regarding energy consumption

- **Extend PAPI to sample hardware monitors**
- **Build new generation of PowerMon devices**
- **Extend performance models**

Transform codes to minimize energy consumption

Inform systems to allow them to exploit DVFS

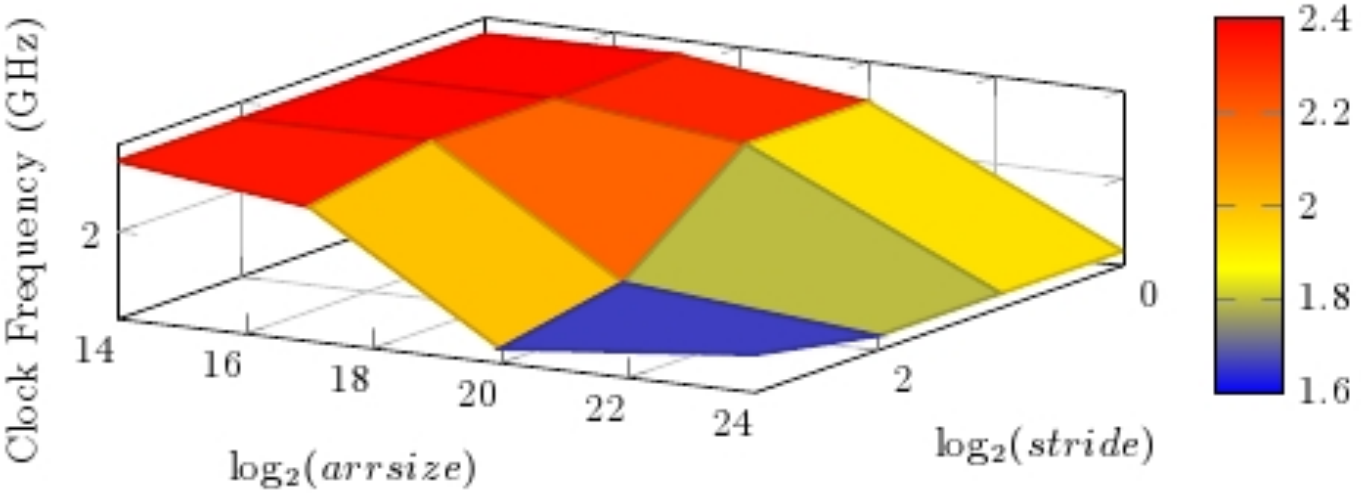
SciDAC
Scientific Discovery through Advanced Computing

SUPER

Early Result from PERI

Use PERI automated performance modeling tools to automate DVFS in HPC applications to reduce energy consumption

Working to combine this DVFS work with existing PERI auto-tuning framework and Active Harmony, to search application space for optimal energy delay product



SciDAC
Scientific Discovery through Advanced Computing

SUPER

Resilient Computing

Led by Bronis de Supinski, Livermore National Laboratory

Novel idea for automating vulnerability assessment

- **Modeled on success of PERI autotuning**
- **Conduct fault injection experiments**
- **Determine which code regions or data structures fail catastrophically**
- **Determine what transformations enable them to survive**
- **Extend ROSE compiler to implement the transformations**

Investigate directive-based API for users

- **Augments empirically derived vulnerability assessment.**

SciDAC
Scientific Discovery through Advanced Computing

SUPER

Optimization

Led by Paul Hovland, Argonne National Laboratory

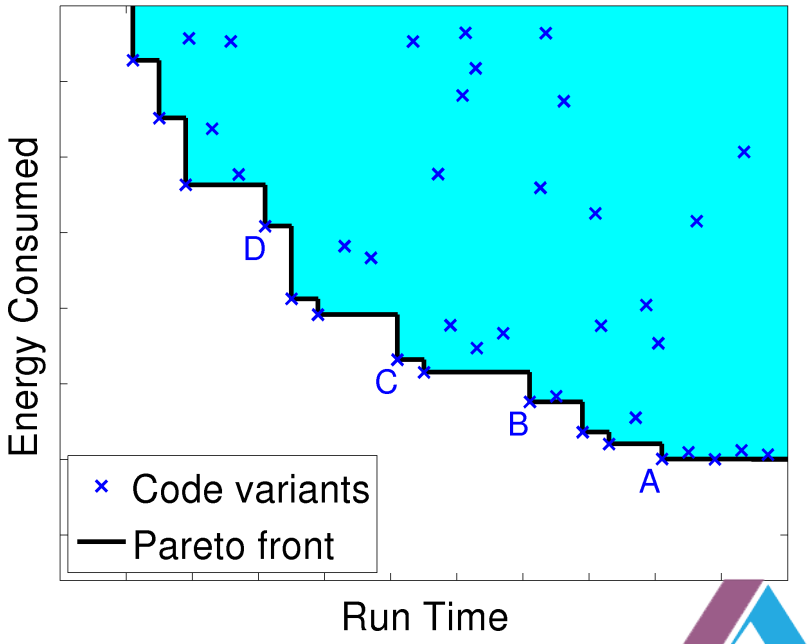
Performance, energy, and resilience are implicitly related and require *simultaneous* optimization

- **E.g., Processor pairing covers soft errors, but halves throughput**

Results in a stochastic, mixed integer, nonlinear, multi-objective, global optimization problem

Only sample small portion of search space:

- **Requires efficient derivative-free numerical optimization algorithms**
- **Need to adapt algorithms from continuous to discrete autotuning domain**



SciDAC
Scientific Discovery through Advanced Computing

SUPER

Application Engagement

- **Led by Pat Worley, Oak Ridge National Laboratory**
- **PERI strategy was proactively identify application collaborators**
  - **Based on comprehensive survey at beginning of SciDAC-2**
  - **Exploited proximity and long-term relationships**
- **SUPER strategy is to broaden our reach**
  - **Key is partnering with staff at ALCF, OLCF, and NERSC**
  - **Augment PerfDMF with IPM and other data from centers**
  - **Have already begun initial outreach to NERSC**
- **Collaborate with other SciDAC-3 institutes**
- **Focused engagement as requested by DOE**

PERI Engagement Impact

LBHMD (lattice Boltzmann) and GTC (plasma toroidal):

- **LBMHD: Up to 3X speedup via autotuning.**
- **GTC: Up to 1.77X speedup via autotuning.**

S3D (combustion):

- **12.7% overall performance improvement.**
- **762,000 CPU-hours are potentially saved each year.**

PFLOTRAN (subsurface reactive flows):

- **2X speedup on two key PETSc routines via autotuning.**
- **40X speedup in initialization; 4X improvement in I/O stage; overall 5X.**

Nek5000 (turbulence):

- **Up to 1.93X speedup.**

LS3DF (electronic structure):

- **Increased scalability from 1000-2000 to over 160,000 cores.**
- **Achieved 442 Tflop/s on Jaguar.**

SciDAC Scientific Discovery through Advanced Computing

SUPER

New Application: LS3DF

LS3DF: "linearly scaling 3-dimensional fragment" code for electronic structure calculation.

Developed at LBNL by Lin-Wang Wang and several collaborators.

Numerous applications in materials science and nanoscience.

Employs a novel divide-and-conquer scheme including a new approach for patching the fragments together.

Achieves nearly linear scaling in *computational cost versus size of problem*, compared with n^3 scaling in many other comparable codes.

Potential for nearly linear scaling in *performance versus number of cores*.

Challenge:

Initial implementation of LS3DF had disappointingly low performance and parallel scalability.

SciDAC
Scientific Discovery through Advanced Computing

SUPER

Performance Analysis of LS3DF

LBNL researchers (funded through PERI) applied performance monitoring tools to analyze run-time performance of LS3DF. Key issues uncovered:

Limited concurrency in a key step, resulting in a significant load imbalance between processors.

- **Solution: Modify code for two-dimensional parallelism.**

Costly file I/O operations were used for data communication between processors.

- **Solution: Replace all file I/O operations with MPI send-receive operations.**

SciDAC Scientific Discovery through Advanced Computing

SUPER

Resulting performance of LS3DF

135 Tflops/s on 36,864 cores of the Cray XT4 Franklin system at LBNL.

40% efficiency on 36,864 cores.

224 Tflops/s on 163,840 processors of the BlueGene/P Intrepid system at Argonne Natl. Lab.

40% efficiency on 163,840 cores.

442 Tflops/s on 147,456 processors of the Cray XT5 Jaguar system at Oak Ridge Natl. Lab.

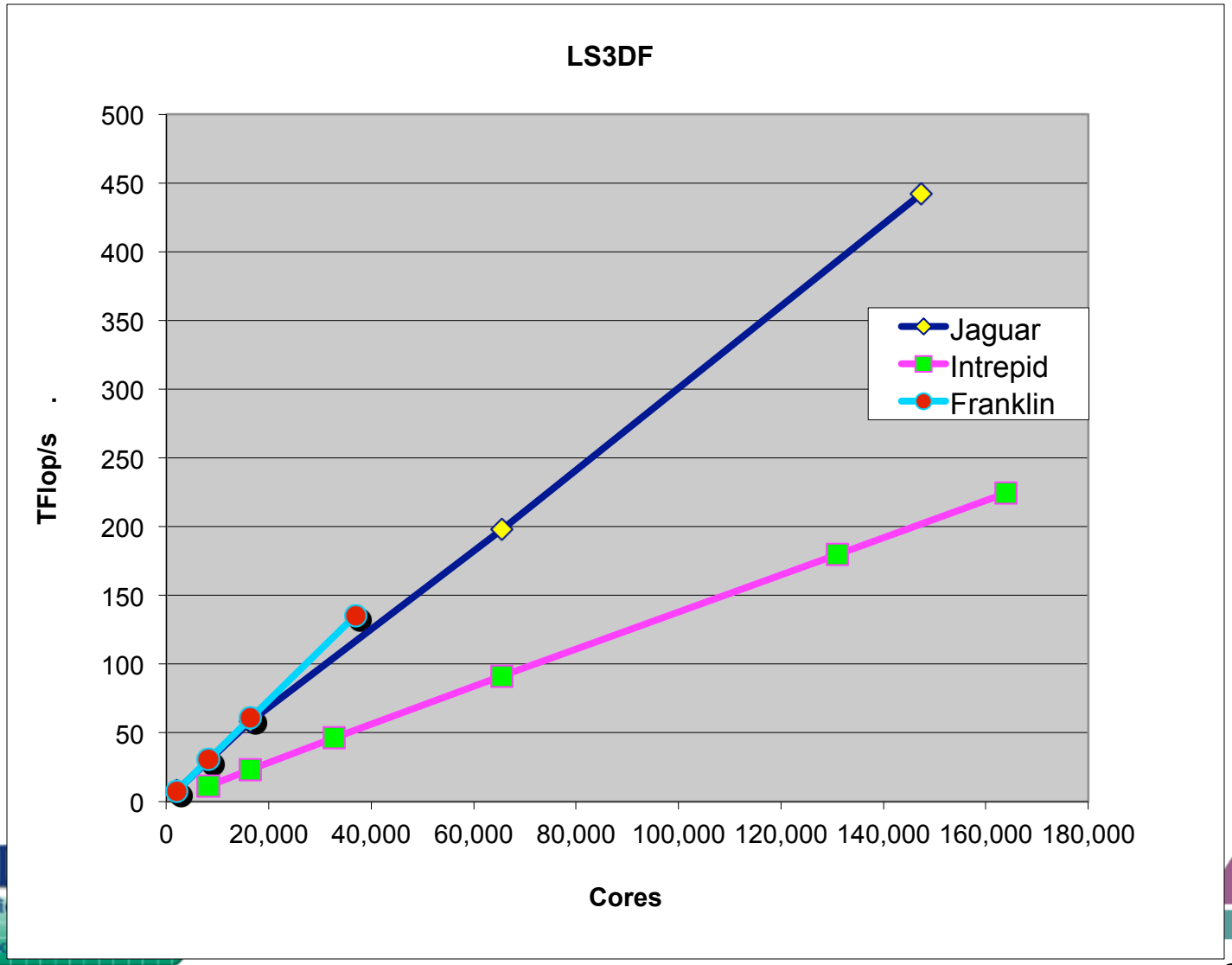
33% efficiency on 147,456 cores.

2008 ACM Gordon Bell Prize in a special category for "algorithm innovation."

Scientific Discovery through Advanced Computing

SUPER

Near-Linear Scaling to 163,840 Cores



Tool Integration

Led by Al Maloney, University of Oregon

TAU replaces HPCToolkit as primary triage tool

PerfDMF replaces PERI performance database

New tools to enable performance portability

- **CUDA-CHiLL and OpenMP-CHiLL**
- **PAPI GPU**

Integration of autotuning framework and TAUmon

- **Enable online autotuning**
- **Already using online binary patch for empirical tuning experiments**

SciDAC
Scientific Discovery through Advanced Computing

SUPER

Outreach and Tutorials

Led by David Bailey, Lawrence Berkeley National Laboratory

We will not provide training workshops as did SciDAC-2 CScADS

We will offering training to ALCF, OLCF, and NERSC staff

- **Enables limited deployment of our research artifacts**

We will organize tutorials for end users of our tools

- **Offer them at widely attended forums such as SC11**

UTK is standing up a SUPER Web site

SciDAC
Scientific Discovery through Advanced Computing

SUPER

Outreach to SciDAC-3 Institutes

PERI was directed not to work with math and CS institutes

- **Instead, focused on JOULE or applications of importance to DOE SC**
- **Even though math libraries have very broad impact**

PERI nevertheless found itself tuning math libraries

- **E.g., PETSc kernels are computational bottlenecks in PFLOTRAN**

SUPER needs new code to focus on beyond SciDAC-e

- **SciDAC-3 applications won't be known for many months**

Initial effort with NNSA and ParaDiS

What else can we do together?

SciDAC
Scientific Discovery through Advanced Computing

SUPER

Summary

Research worthy of DOE SC ASCR

- **Automatic performance tuning**
  - **New focus on portability**
- **Addressing the "known unknowns"**
  - **Energy minimization**
  - **Resilient computing**
- **Optimization of the above**

Near-term impact on DOE computational science applications

- **Application engagement coordinated with ALCF, NLCF, and NERSC**
- **Tool integration, making research artifacts more approachable**
- **Outreach and tutorials**

SUPER