

# Performance Analysis of a Hybrid Parallel Linear Algebra Kernel

Sue Goudy, Lorie Liebrock, Steve Schaffer

New Mexico Institute of Mining and Technology

May 18, 2004

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company,  
for the United States Department of Energy's National Nuclear Security Administration  
under contract DE-AC04-94AL85000.

## That was then ...

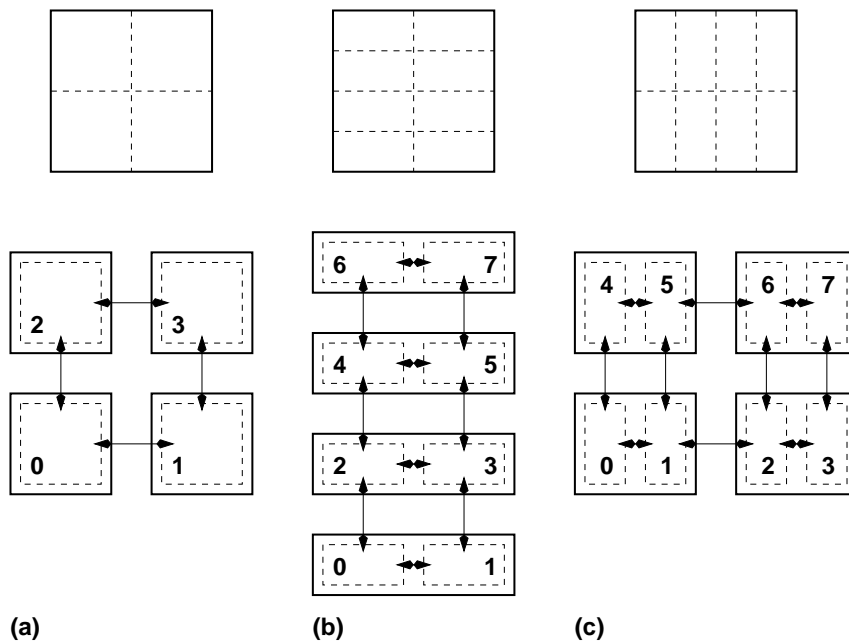
- Pick a good candidate for hybrid parallelism and show speedup over pure MPI.
- What characteristics should this application have?
  - Both coarse- and fine-grained parallelism
  - Large number of messages when compared with computation
- Decrease the number of messages compared with useful work.

# Semicoarsening Multigrid (SMG)

- Iterative method for solution of linear algebraic systems arising from the discretization of elliptic partial differential equations
- 3-dimensional version has at least three levels of parallelism
  - Matrix manipulations of data in the simulation volume
  - Plane solves (done with 2D SMG)
  - Line solves (aka tridiagonal system solves)

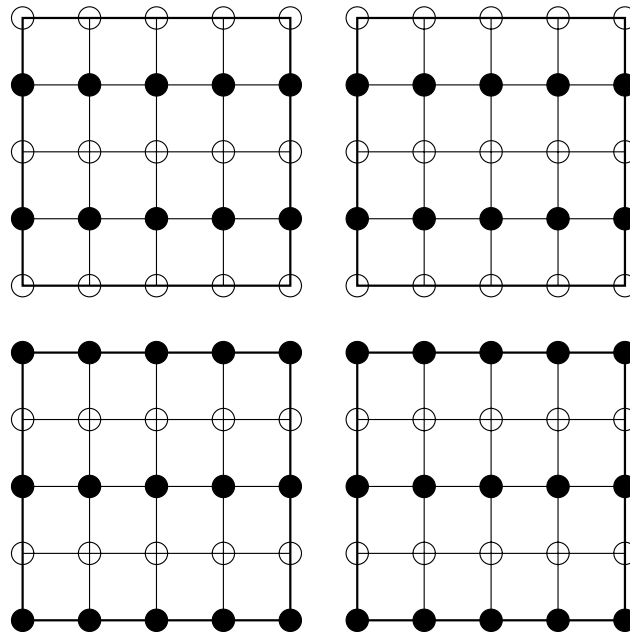
NMT:Goudy - Linux Clusters, The HPC Revolution – p.3

## Sample decompositions for SMG data sets



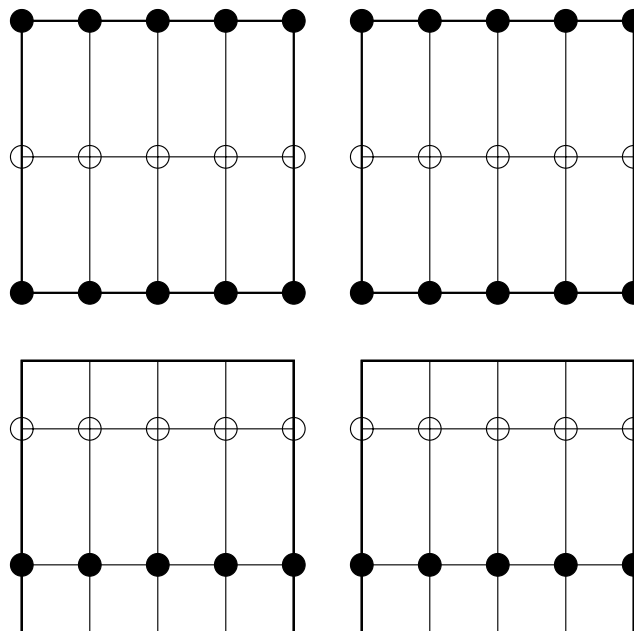
NMT:Goudy - Linux Clusters, The HPC Revolution – p.4

# Sample SMG fine grid across processors



NMT:Goudy - Linux Clusters, The HPC Revolution - p.5

# Sample SMG coarse grid across processors



NMT:Goudy - Linux Clusters, The HPC Revolution - p.6

## First catch your hare ...

- Lifted serial SMG to MPI, using work decomposition based on discretization of the simulation domain
- Added OpenMP directives at lowest loop level (easy to do but not enough payoff)
- Created a block version of our line solver and raised the OpenMP directives up one level
- Ran the code with different data distributions w.r.t. MPI and OpenMP work
- Wondered why the performance was not what we expected

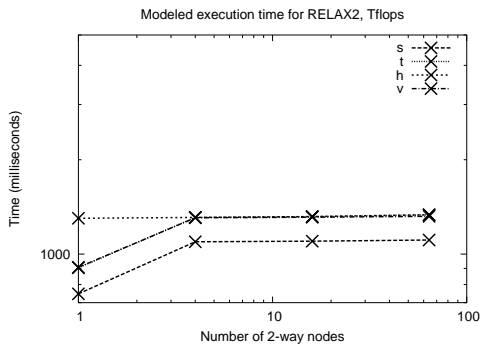
NMT:Goudy - Linux Clusters, The HPC Revolution – p.7

## 2D SMG work estimate

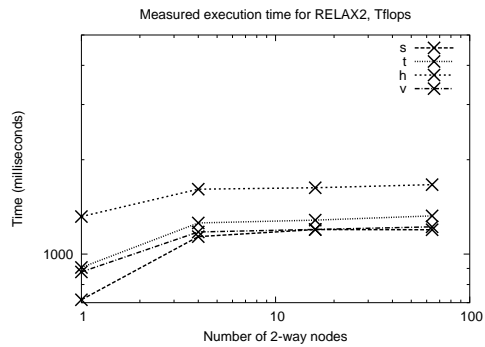
- Computational work on fine grid is  $O(N^2)$  .  
Work on all grids is roughly 2 times that of fine grid alone.
- Communication on fine grid is  $O(N)$  .  
Number of messages on all grids is roughly  $\log_2(N)$  times that of fine grid alone.

NMT:Goudy - Linux Clusters, The HPC Revolution – p.8

# First attempt at model



(a)



(b)

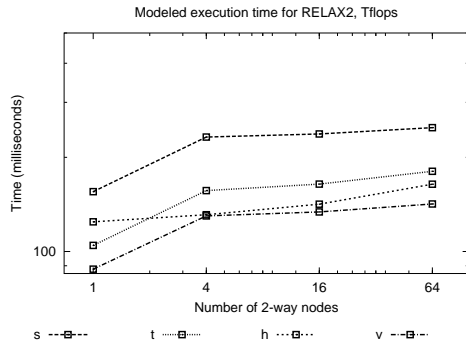
NMT:Goudy - Linux Clusters, The HPC Revolution - p.9

## Model methodology changes

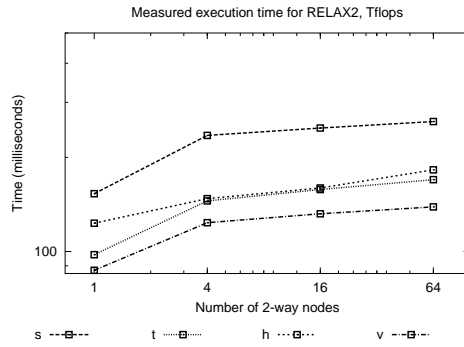
- FP performance - moved from simple benchmark constants to measurements that incorporate memory hierarchy effects
- Communication - moved from simple pingpong benchmark to halo exchange benchmark
- Currently examining effects of MPI task distribution

NMT:Goudy - Linux Clusters, The HPC Revolution - p.10

# Better model

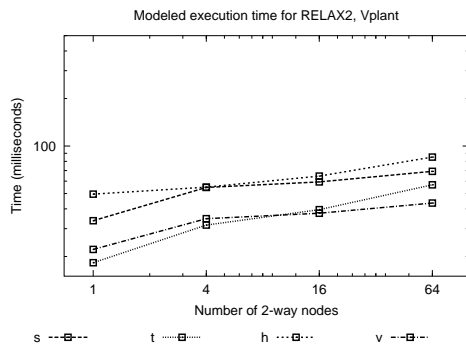


(c)

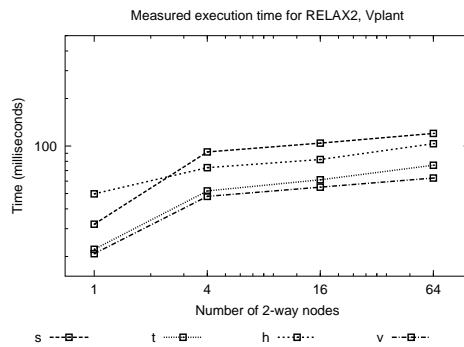


(d)

# Not quite there

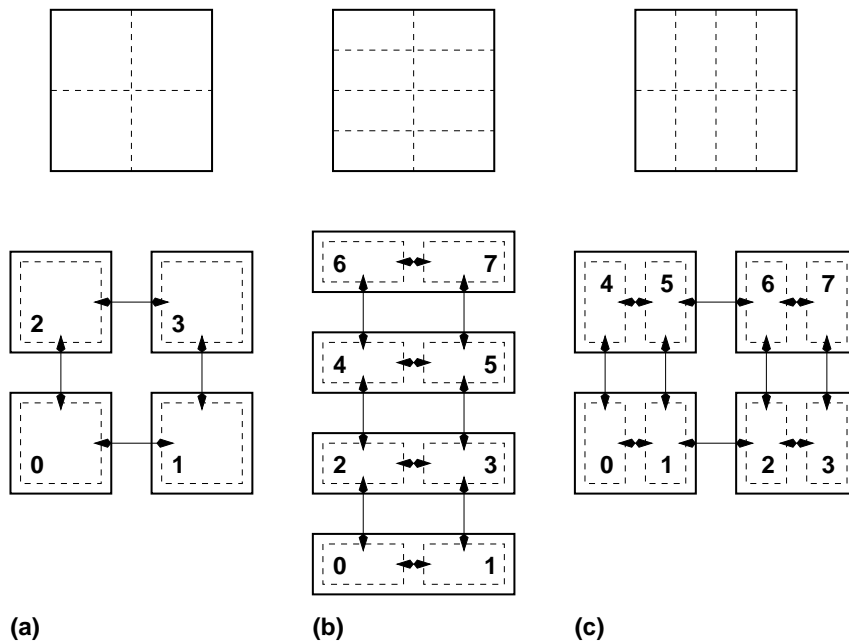


(e)



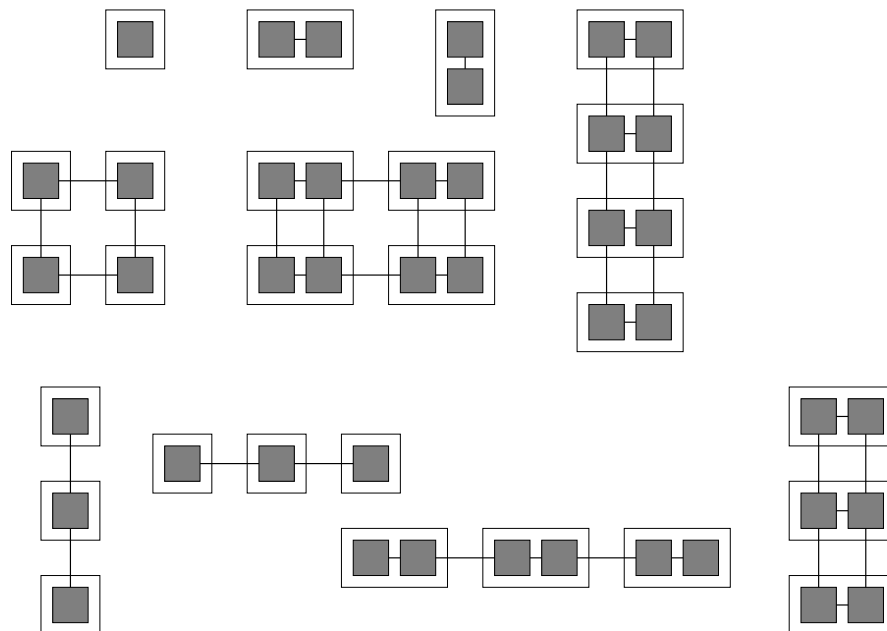
(f)

# Sample decompositions for SMG data sets



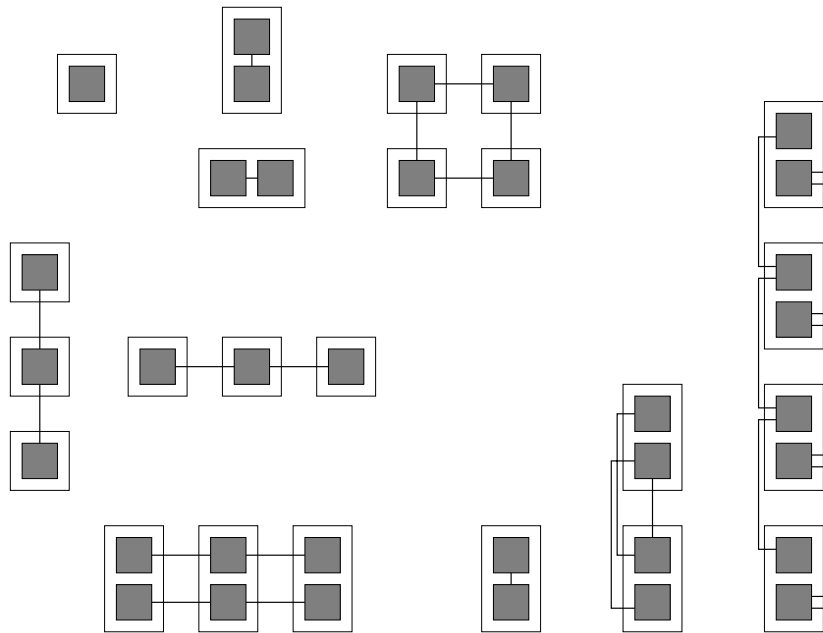
NMT:Goudy - Linux Clusters, The HPC Revolution - p.13

# Block-distributed MPI tasks



NMT:Goudy - Linux Clusters, The HPC Revolution - p.14

# Cyclic-distributed MPI tasks



NMT:Goudy - Linux Clusters, The HPC Revolution – p.15

## This is now ...

- Semicoarsening multigrid seemed like a good candidate for hybrid parallelism.
- This has not been true on any system we've tried.
- Could we have predicted this outcome with a good model?
- What are the components of such a model?

**spgoudy@nmt.edu**

NMT:Goudy - Linux Clusters, The HPC Revolution – p.16