HPGMG: Benchmarking Computers Using Multigrid

This talk: http://59A2.org/files/20150324-HPGMG.pdf

**Jed Brown** jed@jedbrown.org (ANL and CU Boulder)
HPGMG Collaborators: Mark Adams, Sam Williams, John Shalf, Erich Strohmeier (LBNL)

Copper Mountain Multigrid, 2015-03-24
What is performance?

- Accuracy
- Model complexity
- Cost
- Compute Time
- Human Time

- Terms relevant to scientist/engineer
- Compute meaningful quantities of interest, not one iteration/time step
- No flop/s, number of elements/time steps
Work-precision diagram: *de rigueur* in ODE community

- Tests discretization, adaptivity, algebraic solvers, implementation
- No reference to number of time steps, flops, etc.

[Hairer and Wanner (1999)]
Exascale Science & Engineering Demands

- Model fidelity: resolution, multi-scale, coupling
  - Transient simulation is not weak scaling: $\Delta t \sim \Delta x$
- Analysis using a sequence of forward simulations
  - Inversion, data assimilation, optimization
  - Quantify uncertainty, risk-aware decisions
- Increasing relevance $\Rightarrow$ external requirements on time
  - Policy: 5 SYPD to inform IPCC
  - Weather, manufacturing, field studies, disaster response
- “weak scaling” [...] will increasingly give way to “strong scaling”
  [The International Exascale Software Project Roadmap, 2011]
- ACME @ 15 km scaling saturates at < 10% of Titan (CPU) or Mira
  - Cannot decrease $\Delta x$: SYPD would be too slow to calibrate
  - “results” would be meaningless for 50-100y predictions, a “stunt run”
- ACME v1 goal of 5 SYPD is pure strong scaling.
  - Many non-climate applications in same position.
High Performance LINPACK

- Solve $n \times n$ dense linear system: $O(N^{3/2})$ flops on $N = n^2$ data
- Top500 list created in 1993 by Hans Meuer, Jack Dongarra, Erich Strohmeier, Horst Simon
It's all about the memory

Memory motion dominates floating point cost
About half of die devoted to caches
Network moving on-die, maybe throughput cores
Algorithms keep pace with hardware (sometimes)

16 million speedup from each

Opportunities now: uncertainty quantification, design
Incentive to find optimal algorithms for more applications

Algorithmic and architectural advances work together!

[c/o David Keyes]
What does “representative” mean?

- **Diverse applications**
  - explicit PDE solvers (seismic wave propagation, turbulence)
  - implicit PDE solvers and multigrid methods (geodynamics, structural mechanics)
  - irregular graph algorithms (network analysis, genomics, game trees)
  - dense linear algebra and tensors (quantum chemistry)
  - fast methods for N-body problems (molecular dynamics, cosmology)
  - cross-cutting: data assimilation, uncertainty quantification

- **Diverse external requirements**
  - Real-time, policy,
Necessary and sufficient

Goodhart’s Law
When a measure becomes a target, it ceases to be a good measure.

- Features stressed by benchmark necessary for some apps
- Performance on benchmark sufficient for most apps
HPGMG: a new benchmarking proposal

- https://hpgmg.org, hpgmg-forum@hpgmg.org mailing list
- Mark Adams, Sam Williams (finite-volume), Jed (finite-element), John Shalf, Brian Van Straalen, Erich Strohmeier, Rich Vuduc
- Gathering momentum, SC14 BoF

- Implementations
  - Finite Volume memory bandwidth intensive, simple data dependencies
  - Finite Element compute- and cache-intensive, vectorizes, overlapping writes
- Full multigrid, well-defined, scale-free problem
- Matrix-free operators, Chebyshev smoothers
Full Multigrid (FMG): Prototypical Fast Algorithm

- start with coarse grid
- truncation error within one cycle
- about five work units for many problems
- no “fat” left to trim
- distributed memory – restrict active process set using Z-order
Multigrid design decisions

- $Q_2$ finite elements
  - Partition of work not partition of data – sharing/overlapping writes
  - $Q_2$ is a middle-ground between lowest order and high order
  - Matrix-free pays off, tensor-product element evaluation

- Linear elliptic equation with manufactured solution

- Mapped coordinates
  - More memory streams, increase working set, longer critical path

- No reductions
  - Coarse grid is strictly more difficult than reduction
  - Not needed because FMG is a direct method

- Chebyshev/Jacobi smoothers, $V(3, 1)$ cycle
  - Multiplicative smoothers hard to verify in parallel
  - Avoid intermediate scales (like Block Jacobi/Gauss-Seidel)

- Full Approximation Scheme
SuperMUC (FDR 10, E5-2680)

HPGMG-FE Performance

- supermuc np=140608
- supermuc np=32768

Variability

First solve after changing problem size
HPGMG-FE on Edison, SuperMUC, Titan

- Edison: np = 131072
- SuperMUC: np = 140608
- Titan: np = 262144

Graph showing HPGMG-FE Performance with:
- Mean time: 1.6B
- Variability: 12.9B, 155B, 309B
- Titan > 200ms

DOF/s vs Solve time (s)
HPGMG-FV distinguishes networks at 2M DOFs/node
MIC communication bottlenecks on Stampede
Outlook

- What is the cost of performance variability?
  - Measure best performance, average, median, 10th percentile?
- Should dynamic range enter into a ranking metric?
  - Why is NERSC installing DRAM in Cori?
  - Versatility is an essential part of Performance.
- Finite element or finite volume?
  - overlapping writes, cache reuse
  - FE: > 20% Intel, 6% Blue Gene/Q; vs 10% for FV
- Linear or nonlinear?
- Irregularity and adaptivity?
- Tensor-valued coefficients?
- Elasticity?