

FASTMath: Frameworks, Algorithms and Scalable Technologies for Mathematics

Esmond G. Ng, Lawrence Berkeley National Laboratory, Director Karen Devine, Sandia National Laboratories, Deputy Director



SAND2019-8338 PE

FASTMath brings leading edge computational mathematics to the SciDAC Program



Develop advanced, robust numerical techniques for DOE science applications

- Eight focused topical areas based on application needs
- High level synergistic techniques

Deploy high-performance software on DOE supercomputers

- Algorithmic and implementation scalability
- Performance portability





FASTMath mission: Reduce the barriers facing computational scientists





Demonstrate basic research technologies from applied mathematics

- Build from existing connections with basic research
- Focus on results that are most likely to meet application needs

Engage and support the computational science community

- Publications and presentations in highly visible venues
- Team tutorials
- Workforce pipeline and training



Algorithmic and performance-oriented R&D on next-generation architectures is key component of FASTMath

Algorithms for multicore & GPU

- GPU-enabled mesh adaptivity
- Semi-structured multigrid methods
- Parallel-in-time time integration

Reducing data movement

- Low-communication Poisson solvers
- Asynchronous smoothers and preconditioners
- Data reordering in tri-diagonal solvers
- MPI task placement for reduced communication

Exploiting performance-portable paradigms

- Kokkos-based distance-2 coloring for multigrid aggregation in KokkosKernels
- GDSW preconditioners for domain decomposition methods using Kokkos
- Asynchronous multitasking (AMT) for UQ ensembles using Legion

Partitioning effectively and efficiently

- GPU and KNL enabled graph partitioning for unstructured meshes
- Spectral graph partitioning and sparse matrix ordering

FASTMath is Partnership-aware: actively engaged with 23 SciDAC-4 partnerships



BER (5/8)

- Structured AMR
- Unstructured AMR
- Time integration
- Linear/Nonlinear solvers, Preconditioners
- Optimization
- Verification / UQ

FES (9/9)

- Unstructured
 meshes AMR
- Unstructured mesh PIC methods
- Discretization technologies
- Iterative solvers
- Time integration
- UQ

HEP (3/5)

- Direct solvers
- Structured Grid AMR
- Optimization
- Sensitivity Analysis
- Inference and machine learning

NP (2/3)

- Structured grid AMR
- Eigenvalue problems
- Inference and Machine Learning

BES (3/4)

- Nonlinear and tensor eigenvalue problems
- Linear solvers and Preconditioners

NE (1/1)

- UQ
- Time integration

FASTMath is focused on eight core technology areas





FASTMath brings together an exceptional team of researchers and software library capabilities



Our software has 100s of Our team comprises over 50 researchers from 5 national person-years of development laboratories and 5 universities behind it SuperLU PARPACK Argonne **PETSc** mfem ZOLTAN Nationa svmPACK **AMReX** Rensselaer SMU Trilinos 🕻 kokkos kernels

FASTMath team



M. Adams, LBNL A. Almgren, LBNL R. Archibald, ORNL T. Casey, SNL P. Colella, LBNL B. Debusschere, SNL A. Dener, ANL K. Devine, SNL V. Dobrev, LLNL M. Eldred. SNL R. Falgout, LLNL D. Gardner, LLNL G. Geraci, SNL R. Ghanem, USC P. Ghysels, LBNL J. Hu, SNL X. Huang, ANL D. Ibanez, SNL M. Jacquelin, LBNL J. Jakeman, SNL

K. Jansen, Colorado H. Johansen, LBNL M. Knepley, Buffalo T. Kolev, LLNL S. Leyffer, ANL R. Li, LLNL X. Li, LBNL Y. Liu, LBNL J. Loffeld, LLNL O. Marques, LBNL D. Martin, LBNL Y. Marzouk, MIT P. McCorquodale, LBNL L. C. McInnes, ANL M. Minion, LBNL J. Müller, LBNL T. Munson, ANL H. Najm, SNL E. Ng, LBNL M. Perego, SNL

- S. Rajamanickam, SNL
- D. Reynolds, SMU
- C. Safta, SNL
- O. Sahni, RPI
- K. Sargsyan, SNL
- R. Saye, LBNL
- S. Seol, RPI
- M. Shephard, RPI
- B. Smith, ANL
- C. Smith, RPI
- G. Slota, RPI
- H. Tran, ORNL
- R. Van Beeumen, LBNL
- R. Vogt, NCSU
- S. Wild, ANL
- C. Woodward, LLNL
- C. Yang, LBNL
- U. M. Yang, LLNL
- H. Zhang, ANL



Structured Grid Spatial Discretization

- Goal: Provide software support for efficient parallel solution of a large variety of problems in science and engineering using blockstructured adaptive mesh approaches
- Technologies: adaptive mesh refinement, high-order discretizations, mapped multi-block domains, dynamic load balancing, particle dynamics
- **Software**: AMReX, Chombo, Algoim
- SciDAC-4 Partnerships:
 - NP: TEAMS
 - BER: ProSPect
 - FES: AToM
 - HEP: ComPASS
- Area Lead: Ann Almgren <u>ASAlmgren@lbl.gov</u>



Astrophysics (Castro)



Accelerators (WarpX)



Cosmology (Nyx)



Combustion (LMC)

In structured AMR, embedded boundary methods address dynamically changing geometries



Embedded Boundary/Multifluid approach for tracking ice boundaries in ProSPect BISICLES

- Location of transition from grounded to floating ice is dynamically important
- Space-time formulation: Development case implemented in 1D, with extension to 2D in progress





Representing a grounding line using stair-step (left) and embeddedboundary (right) discretizations



Grounding lines can move arbitrarily fast depending on bathymetry and ice thickness \rightarrow contact point problem



High-order mapped multiblock methods enable first-of-kind tokamak computations for COGENT



First-ever 4D continuum gyrokinetic simulation capability for problems in the edge plasma region of tokamak reactors that spans both sides of the magnetic separatrix

 FASTMath's mapped-multiblock software discretizes PDE in a nearly field-aligned coordinate system





- Fourth-order finite volume methods
- High-order ghost value interpolation
- Flux-matching at block boundaries



Unstructured Mesh Spatial Discretization

- Goal: Deliver scalable, high-performing unstructured mesh tools to support applications with complex geometries
- Technologies: Unstructured mesh adaptation, parallel unstructured mesh infrastructure, high-order discretizations, error estimation, load balancing, task mapping, graph algorithms, PIC methods
- Software: Albany/LGR, EnGPar, MFEM, MeshAdapt, Omega_h, PHASTA, PUMI, PUMIpic, PuLP/XtraPuLP, Zoltan/Zoltan2
- SciDAC-4 Partnerships and Other Interactions:
 - FES: PSI2, HPBS, RF, CTTS, TDS, SCREAM
 - BER: DEMSI, ProSPect
 - BER: E3SM and CMDV
 - BLAST, LGR, ATDM, ASC
- Area Lead: Mark Shephard (shephard@rpi.edu)







FASTMath provides unstructured mesh technology to Six Fusion SciDAC partnerships



Unstructured mesh capabilities provided:

- New physics discretization technologies
- CAD model simplification
- Mesh generation with specific mesh controls
- High-order adaptation for curved geometries
- Adaptive simulation workflows
- Dynamic load balancing
- Mesh infrastructure for finite element assembly
- New unstructured mesh PIC method



MFEM EM analysis of vacuum region on ~1M element mesh of RF CAD, wall geometry and flux surfaces



PUMIPic tools for Particle-in-Cell methods

- GPU- and multicore-optimized mesh and particle data structures
- Distributed mesh with overlapping domains for on-processor particle push

Data analysis and graph algorithms contribute to success of unstructured mesh applications



Unstructured mesh infrastructure simplifies data analysis, adaptive UQ control, and extraction of fundamental insights from simulations

- Challenging due to complex geometries, discretization size, number of uncertain input parameters
- New tools for computational steering, multi-fidelity modeling and in situ visualization







Graph biconnectivity algorithms enable ice sheet simulations for ProSPect MALI

- Parallel algorithms detect degenerate mesh features; enable solver convergence without expensive pre-processing
- Best paper award at 2019 Int. Conf. Parallel Processing

Unstructured mesh technologies are advancing for efficiency on next generation systems



Accelerator-enabled unstructured mesh procedures

- Novel data structures and algorithms for irregular memory access patterns in graphs, meshes, and PIC
- Built on performance-portable paradigms (e.g., Kokkos, Raja) for "future proofing"





Zoltan2 task placement reduces communication time in the E3SM HOMME atmospheric modeling up to 31% on Mira

GPU-accelerated mesh adaptivity enables use of single GPU to evaluate designs that previously required more time on large CPU clusters

Architecture-aware task placement and partitioning

 Reduces application communication time by placing interdependent tasks in "nearby" cores

- Goal: Providing efficient direct and iterative linear solvers and preconditioners for large scale parallel computers
- Technology: linear solvers and preconditioners (direct, iterative, dense, sparse); multigrid methods
- Software: KokkosKernels, hypre, MueLu, PETSc, Trilinos, ShyLU, STRUMPACK, SuperLU, symPACK, ButterflyPACK
- SciDAC-4 Partnerships and Other Interactions
 - BER: ProSPect
 - HEP: ComPASS
 - FES: HBPS, SCREAM, ISEP, CTTS
 - BES: Computational Chemical Sciences Program
 - BER: E3SM
 - ATDM, ASC
- Area lead: Ulrike Yang (yang11@llnl.gov)









Focus on architecture-aware solvers prepares FASTMath solver libraries for next generation systems



KokkosKernels performance-portable sparse/dense linear algebra and graph kernels for CPUs, GPUs, KNLs
Batch BLAS, coloring, SpMM, SpMV, etc.





Domain decomposition preconditioners in ShyLu built on Trilinos' Tpetra software stack
Performance portability from Kokkos

Semi-structured multigrid methods in hypre

 Suitable for high levels of parallelism as in GPU



Focus on architecture-aware solvers prepares FASTMath solver libraries for next generation systems





Poisson solver for structured AMR with1/10th the communication costNew representation of nonlocal coupling

Localizing nonzeros in sparse matrix factorizations

 Higher computational efficiency in sparse triangular solves in symPACK



A sketch of the column reordering heuristic: columns adjacent to descendants are brought together, forming larger contiguous blocks



Spectral sparse matrix ordering built on communication avoiding eigensolversHigh quality benefits direct solvers



Time Integration

- Goal: Efficient and robust time integration methods and software for stiff implicit, explicit, and multirate systems in DOE applications
- Technologies: explicit/implicit/IMEX time integration, multirate methods
- Software: SUNDIALS, PETSc adjoint integration, Spectral Deferred Correction in AMReX



Baroclinic instability atmospheric climate simulation using SUNDIALS with the Tempest nonhydrostatic simulation code



t = 5000

Gray-Scott solution showing pattern formation on a 2-level AMR grid using SUNDIALS' multirate integration

Engine fuel simulation using SUNDIALS for fuel efficiency studies



lies Combustion

- SciDAC-4 Partnerships and Other Interactions
 - BER: Physics Convergence
 - NE: NUCLEI
 - FES: MGK
 - BER: E3SM, ParFlow watershed simulation
 - FES: BOUT++
 - OE: GridDyn
 - EERE: Zero-RK, PELE
 - ATDM, ASC
 - Area Lead: Carol Woodward (woodward6@llnl.gov)

New multirate time integration addresses needs of multiphysics applications



Multiphysics/multiscale simulations are limited in accuracy due to commonly used first order splittings

E.g., climate, combustion, fusion

High-order multirate integration advances different parts of systems with different time step sizes

- Reduce computation and communication: fewer function evaluations for slow terms
- Delivered in SUNDIALS ARKode
- Demonstrated in AMReX structured AMR framework

 $y' = f_{\text{fast}}(y) + f_{slow}(y)$



Two-rate integration method showing the coupling between the fast and slow methods in an MIS algorithm for a thirdorder explicit slow method

Same solution quality with 85% fewer advection evaluations for Brusselator advection-reaction system

Multirata

		Multilate
Steps	6,805	1,429 / 8,813
Advection evals (slow)	29,379	4,288
Reaction evals (fast)	98,917	116,949

New adjoint time-stepping gives rise to new PDE constrained optimization tool in PETSc / TAO



New tool for PDE-constrained optimization

- Applications: Global data assimilation, power system planning and operation
- Adjoint time-stepping schemes give consistent gradients for time-dependent PDEs, leading to faster optimization convergence and better accuracy
- Implementation allows multiple objective functions and time intervals



Variational data assimilation for a global shallow water model





Eigenvalue Calculations

- Goal: Develop algorithms and scalable software for linear and nonlinear eigenvalue problems
- Technologies: Efficient eigensolver algorithms; scalable multicore and GPU implementations
- **Software**: FASTEig eigensolver collection
- SciDAC-4 Partnerships and Other Interactions:
 - BES: CompCat-SciDAC
 - NP: NUCLEI
 - BES: Quantum Coherence EFRC
 - BES: Computational Materials Science Center
 - BES: Computational Chemical Science Center
- Area Lead: Chao Yang (CYang@lbl.gov)



Eigensolvers enable new discoveries in material science



Solving nonlinear eigenproblems from DFT based electronic structure calculations

 Spectrum-slicing method provides greater scalability above 5K cores than ELPA, ScaLAPACK





On-node performance of tensor eigensolver on Cori-KNL. For large spin chains, we can achieve 70-fold speedup (68 cores x 4 hyperthreads/core) New matrix-free iterative tensor eigensolver is capable of computing interior eigenvalues of disordered Heisenberg spin chains with more than 28 spins

- Previous record: 26 spins
- Enables study of many-body localization and thermalization properties of quantum materials
- Near perfect on-node scaling to 68 KNL cores for > 20 spins

Model Order Reduction speeds electron excitation simulation by orders of magnitude



Modeling linear response of molecular system to light (photons)

- Full-system approach requires solving thousands of linear systems from samples of spectral function on fine spectral grid
- Model order reduction approach needs fewer than 100 linear solves using rational Krylov subspace of dimension less than 100
- Can reduce execution time several orders of magnitude





Data Analytics

- Goal: Sparse functional representation of data to enable faster IO and analysis of big datasets
- Technology: Data compression, data reconstruction, machine learning workflows
- Software tools: Tasmanian, PUMI, TAO, Phasta, Catalyst
- SciDAC Applications and Other Interactions:
 - RAPIDS Institute
 - FES
- Area Lead: Rick Archibald (<u>ArchibaldRK@ornl.gov</u>)





Developing workflows and ML methods to accelerate discovery and validation in simulation and experimental data



Super Resolution methods for experimental data



Accurate Synthetic aperture radar reconstruction

in situ Visualization unlocks unsteady dynamics at extreme scale



In situ visualization allows comparison of instantaneous vortical structures with phase-averaged quantities from experiments and simulations

- Isosurface construction at each time step adds 3% overhead to simulation – far less than writing full data
- Collaboration with RAPIDS Institute





Comparison to experimental results.

Reconstruction methods benefit and exploit multiple FASTMath technical areas



MGARD-Multigrid Adaptive Reduction of Data

 Reconstruction of simulation data preserves *physical dynamics or quantities of interest* to a predescribed tolerance level





Original and reconstructed data from online dictionary learning

Compression and Reconstruction of Streaming Data via dictionary learning

Matrix factorization approach



Numerical Optimization

- Goal: Develop methods for numerical optimization problems with constraints and for sensitivity analysis using adjoint capabilities
- Technology: PDE-constrained optimization, sensitivity analysis, adjoints
- **Software:** MATSuMoTo, MINOTAUR, ORBIT, ROL, TAO
- SciDAC-4 Partnerships:
 - NP: NUCLEI
 - HEP: ComPASS, Accelerating HEP Science, Data Analytics on HPC
 - BER: ProSPect
- Area Lead: Todd Munson (tmunson@mcs.anl.gov)

Preferential parameter selection speeds optimization of large-scale black-box problems



We consider large-scale, black-box, multimodal, time consuming optimization



- Reduce the problem dimension by sensitivity analysis on a surrogate model
- In each iteration, optimize only over the most sensitive (important) parameters



Figure: 100-dim test example. Preferentially optimizing over the most important parameters leads to finding *improved solutions faster* than optimizing over all parameters

Future applications include the optimization of high energy physics simulations



A Mixed-Integer PDE-Constrained Optimization (MIPDECO) Method for the Design of an Electro-Magnetic Scatterer



New mixed-integer PDE-constrained optimization (MIPDECO) methods optimize design of an electromagnetic scatterer to cloak a region

- Set-based steepest-descent trustregion method for MIPDECO
- Promising numerical results and theoretical foundations

(cloaking device)

$$\min_{\substack{u,w \ u,w}} \| u + u_0 \|_{2,\Omega_0}^2$$
s.t. $-\Delta u - k_0^2 (1 + qw) u$
 $= k_0^2 qw u_0$
 $w \in \{0,1\}$

MIPDECO Formulation

Example: Design of electro-magnetic scatterer •Objective: Cloak the top-right corner •PDE: 2D Helmholtz equation •Discrete variables: 0-1 design of scatterer





Scatterer

Difference of Waves

Uncertainty Quantification

- Goal: Provide robust and efficient capabilities for uncertainty quantification in large-scale computations of physical systems.
- Technologies: Forward and inverse UQ, Bayesian learning, model error, multifidelity methods, optimization under uncertainty
- **Software**: DAKOTA, UQTk
- SciDAC-4 Partnerships and Other Interactions:
 - FES: Plasma surface interactions
 - NE: Fission gas in uranium oxide nuclear fuel
 - BER: OSCM
 - BER: ProSPect
 - EERE: redox potentials, A2e
 - BES: experimental design, E3SM
 - NSF, NNSA, DARPA
- Area Lead: Habib N. Najm (hnnajm@sandia.gov)







FASTMath UQ research is addressing challenges of high-dimensionality in applications



Efficiently building surrogates of high-dimensional models is difficult

Seeking low-rank functional tensor-train representations that reveal coupling in high-dimensional models

Applications:

- Optimizing sensor networks for improving climate model predictions (E3SM)
- Estimate uncertainty in sea-level rise predictions due to ice-sheet evolution (ProSPect)

E3SM Simulation



Tensor-train Schematic



Total Order Sensitivities



Advances in Bayesian Methods increase accuracy and robustness of UQ tools



UQ framework incorporating model error enables predictive uncertainty attribution due to BOTH data and model errors

Implemented in UQTk; used in DOE fusion science and land model

MCMC with local surrogates when global surrogates are infeasible

 Guaranteed convergence for heavier-tailed distributions broadens robustness and applicability

Likelihood distance metrics for parameter estimation in chaotic dynamical systems

 Diffusion-map distance between densities defined in manifold coordinates

> Density of points (in 2D diffusion-map coordinates) from Lorentz 63 system



Prototyping next generation AMT management of UQ ensembles with Dakota + Legion



Prototype ensemble management based on Asynchronous Many-Task (AMT) systems using Dakota + Legion.

- Exploit large-scale hybrid architectures
- Dependency-driven task management eliminates artificial synchronization and streamlines workflows



Modular architecture for Dakota-MPI, Dakota-X,

Joint Capability Roadmap for AMT ensemble management



Joint work with Stanford PSAAP2

FASTMath is focused on eight core technology areas





FASTMath has expertise needed for Machine Learning research and development



- Machine learning is built on core technologies in FASTMath: UQ, Linear Algebra, Optimization, Statistics, Data Analytics
 - Linear regression, regularization, SVD, PCA, probabilistic methods, stochastic steepest descent, risk minimization, Bayesian statistics, network inference,
- FASTMath advances in algorithms, models, computation, performance are directly applicable
 - Ex: Neural networks often viewed/used as surrogate representations of model output as function of set of inputs
 - Surrogates broadly used in optimization and UQ
 - Neural network training equivalent to surrogate fitting/regression
 - Ex: GPU-enabled linear algebra kernels accelerate deep neural network computations

FASTMath's research and technology support machine learning



FASTMath researchers' efforts (through FASTMath or other projects):

- Co-organize annual Machine Learning and Uncertainty Quantification workshops at USC (UQ)
- Neural network surrogates for the E3SM land model: 2X more accurate vs sparse regression polynomials (UQ)
- Deep learning for compression artifact removal (Data analytics)
- ML for Fusion research workflows (Data analytics)
- Modeling and solvers for goal-oriented ML-based regression (Numerical Optimization)
- KokkosKernels used in 2019 DARPA Sparse Deep Neural Network Challenge submission (Linear solvers)
- Performance-portable batch operations in KokkosKernels (Linear Solvers)
- Graph expertise relevant to ML in biology and material science (Unstructured Mesh, Linear Solvers)
- Expanding support for deep neural networks in Dakota's surrogate libraries (UQ)
- Distributed-memory sparse tensor decomposition

Several ML-related collaborations underway by FASTMath researchers



- HEP: Accelerating HEP Science: Inference and Machine Learning at Extreme Scales
- HEP: Data Analytics on HPC
- NP: NUCLEI (Deep learning for ab initio nuclear theory)
- RAPIDS Institute
- DARPA: Probablistic Machine Learning for physics discovery
- ECP: ExaLearn and ExaGraph
- IBM: Bayesian neural networks for yield-aware decision making

Collaboration with RAPIDS: Deep-learning based compression artifact removal provides fast enhancement compared to state-ofthe-art compressed sensing (CS)



Enhanced deep

super-resolution Network

TV Compressed Sensing

Residual Dense

Network



Visit the FASTMath posters

Tuesday

Time Integration Activities Eigensolvers Activities Unstructured Meshing Technologies UQ Algorithms Kokkos Kernels and Linear Solvers Numerical Optimization Activities Fast and Parallel Direct Linear Solvers Numerical Optimization Activities FASTMath Overview Carol Woodward Chao Yang Dan Ibanez; Mark Shephard Habib Najm Siva Rajamanickam Todd Munson Yang Liu Todd Munson Esmond Ng; Karen Devine

Thursday

Optimizing Computationally Expensive Large-scale Black-box Problems **Juliane Mueller** Unstructured Mesh Technologies for Fusion Simulation Codes Mark Shephard MFEM - Scalable Finite Element Methods Mark Stowell Mike Eldred **UQ** Software **Data Analytics Activities Rick Archibald** Multilevel Linear Solvers Ruipeng Li Structured Grid Activities Hans Johansen E. Ng; K Devine **FASTMath Overview**





RAPIDS + FASTMath Collaboration

Application Performance

- 3-9x faster triangular solves in CTTS tokamak simulations on KNL
- MPI communication and on-node parallelism optimizations for DGDFT density functional theory computation
- New low-memory PETSc solver interface for Xolotl cluster dynamics code

Data Analysis

- Deep convolutional neural networks for fast artifact reduction in lossy compressed images
- Machine learning as a coupler between simulation and experiments
- Parallelization approaches for deep learning training on leadership-class systems

Numerical Library Performance

- New structured adaptive mesh refinement support in DIY data analysis toolkit
- Participation in NERSC Roofline
 training workshop
- Explore GPU acceleration for ODEs in AMReX

Visualization

- In situ visualization demonstrated in fluid dynamics simulations at > 1M cores
- Computational steering to allow realtime analyst control of simulations
- Fast, low-memory in situ visualization of high-order fields in unstructured meshes

See posters on these collaborations both days!

FASTMath leadership council https://fastmath-scidac.org





Esmond Ng Director EGNg@lbl.gov

Karen Devine Deputy Director kddevin@sandia.gov



Ulrike Yang Linear Solvers yang11@llnl.gov

Habib Najm UQ hnnajm@sandia.gov





Ann Almgren Structured Mesh asalmgren@lbl.gov

Mark Shephard Unstructured Mesh shephard@rpi.edu





Rick Archibald Data Analytics archibaldrk@ornl.gov

Todd Munson Optimization tmunson@mcs.anl.gov





Carol Woodward Time Integration woodward6@llnl.gov



Chao Yang Eigensolvers cyang@lbl.gov