SAND2020-1548C

Sandia ATDM Program: Application, Productivity and Performance Portability

February 6th, 2020

Bill Rider for the Sandia ATDM Team

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525

SNL ATDM Team Members

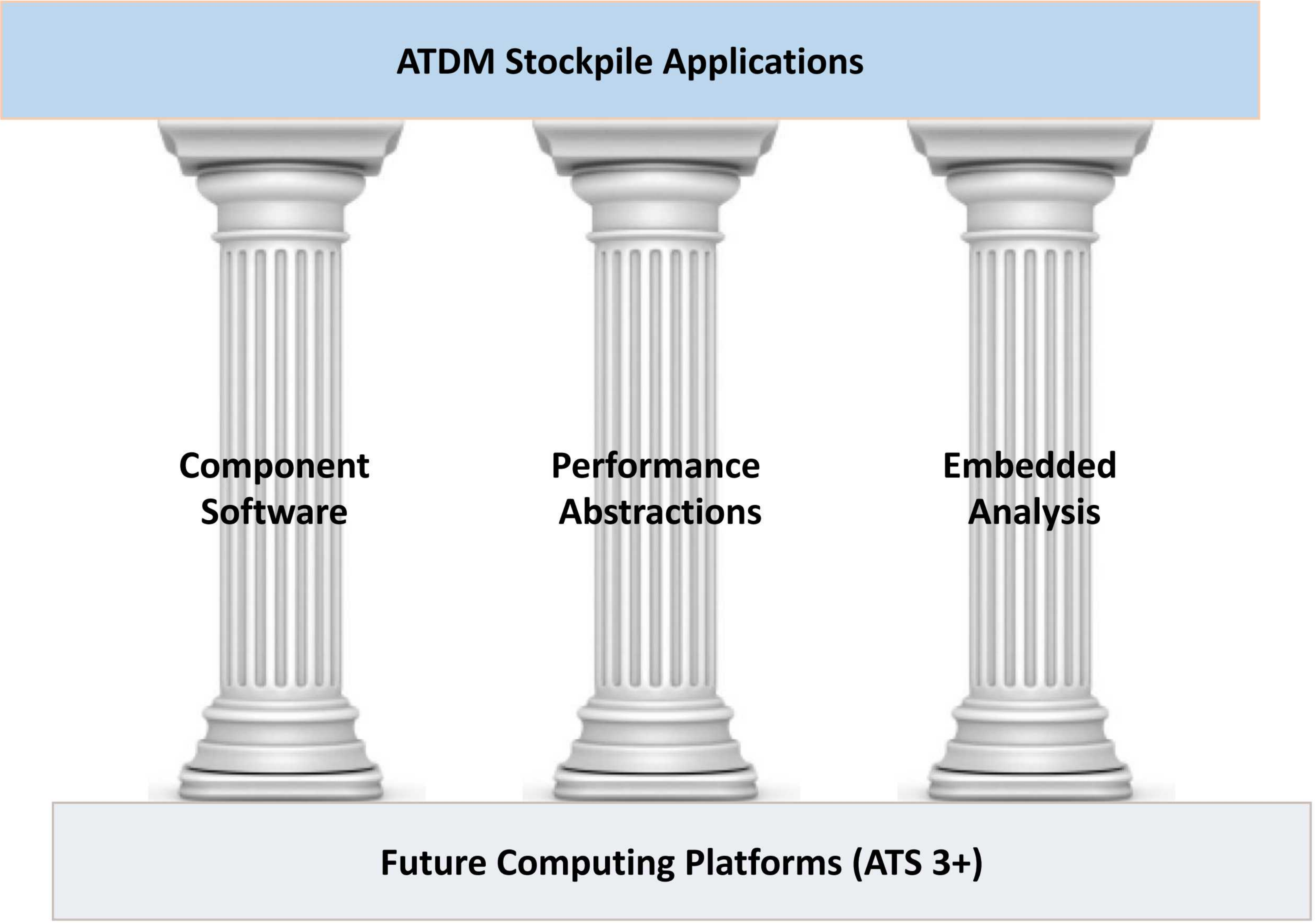
Program Management:
Jim Stewart, Mike Glass, Bill Rider, Paul Crozier, Richard Kramer, Ken Alvin

EMPIRE PI's:
Matt Bettencourt, Eric Cyr
EMPIRE PO:
Keith Cartwright
EMPIRE Staff:
Kris Beckwith, Adam Blake, Wade Burgess, Jason Gates, Daniel Jensen, Paul Lin, Ed Love, Billy McDoniel, Duncan McGregor, Brandon Medina, Logan Meredith, Sean Miller, Chris Moore, Stan Moore, Edward Philips, Tim Pointon, Troy Powell, Gregg Radtke, Nick Roberds, Sidney Shields, Scot Swan, Andrew Steyer, David Turner

SPARC PI:
Micah Howard
SPARC Staff:
Matt Barone, Matt Bopp, Sam Browne, Jeff Fike, Travis Fisher, Mike Hansen, Victor Brunini, Mark Hoemmen, Brad Maeng, James Overfelt, Julia Plews, Cory Stack, Ross Wagnild, Jerry Watkins
SPARC V&V Team:
Sarah Kieweg, Brian Carnes, Eric Mussoni, Brian Freno, Greg Weirs, Derek Dinzl, Jaideep Ray, Ryan Keedy, Micah Howard, Ross Wagnild, Justin Winokur

Components Project Managers:
R. Brightwell, R. Clay, R. Hoekstra, R. Oldfield, A. Salinger, D. Turner, M. Wolf,
Components Principal Investigators:
R. Bartlett, S. Hammond, B. Hanks, J. Hu, P. Lin, K. Moreland, S. Olivier, E. Phipps, S, Rajamanickam, G. Sjaardema, C. Trott
Components Staff:
S. Acer, A. Bradley, L. Berger-Vergiat, B. Carnes, J. Ciesko, S. Conde, K. Copps, V. Dang, M. Ebeida, H. Edwards (Nvidia), N. Ellingwood, J. Elliott, N. Evans, K. Ferreira, J. Forester, J. Frye, J. Gates C. Glusa, R. Grant, G. Hansen, E. Harvey, D. Hollman, D. Ibanez, S. Levy, B. Kelley, J. Mauldin, D. Mavriplis (U. Wyoming), D. McGregor, J. Miles, S. Miller, S. Moore, S. Mukherjee, C. Ober, T. Otahal, S. Owen, K. Pedtretti, M. Perego, E. Phillips, D. Poliakoff, N. Roberts, R. Quadros, W. Scott, C. Siefert, W. Spotz, P. Stallings, D. Sunderland, G. Templet, I. Tezaur, A. Toth, R. Tuminaro, C. Ulmer, L. Ward, J. Watkins, J. Wilke, A. Younge,

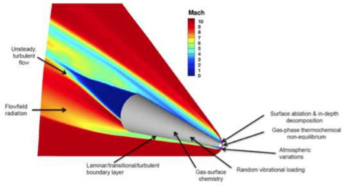
ATDM @ Sandia



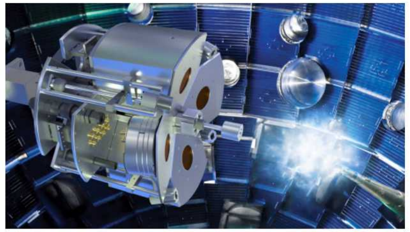
ATDM Stockpile Applications
Component Software
Performance Abstractions
Embedded Analysis
Future Computing Platforms (ATS 3+)

Agile Component Strategy

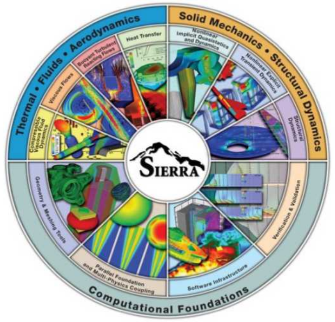
- Sandia has decades of software development experience using component strategy
- Start from current Agile Components (Trilinos)
- Design new components/APIs based on ATDM requirements
- Explore new technologies
- Deep integration of ATDM technologies
- Deep integration of ATDM application and component teams



SPARC



EMPIRE



Engineering Mechanics Code Suite

RAMSES

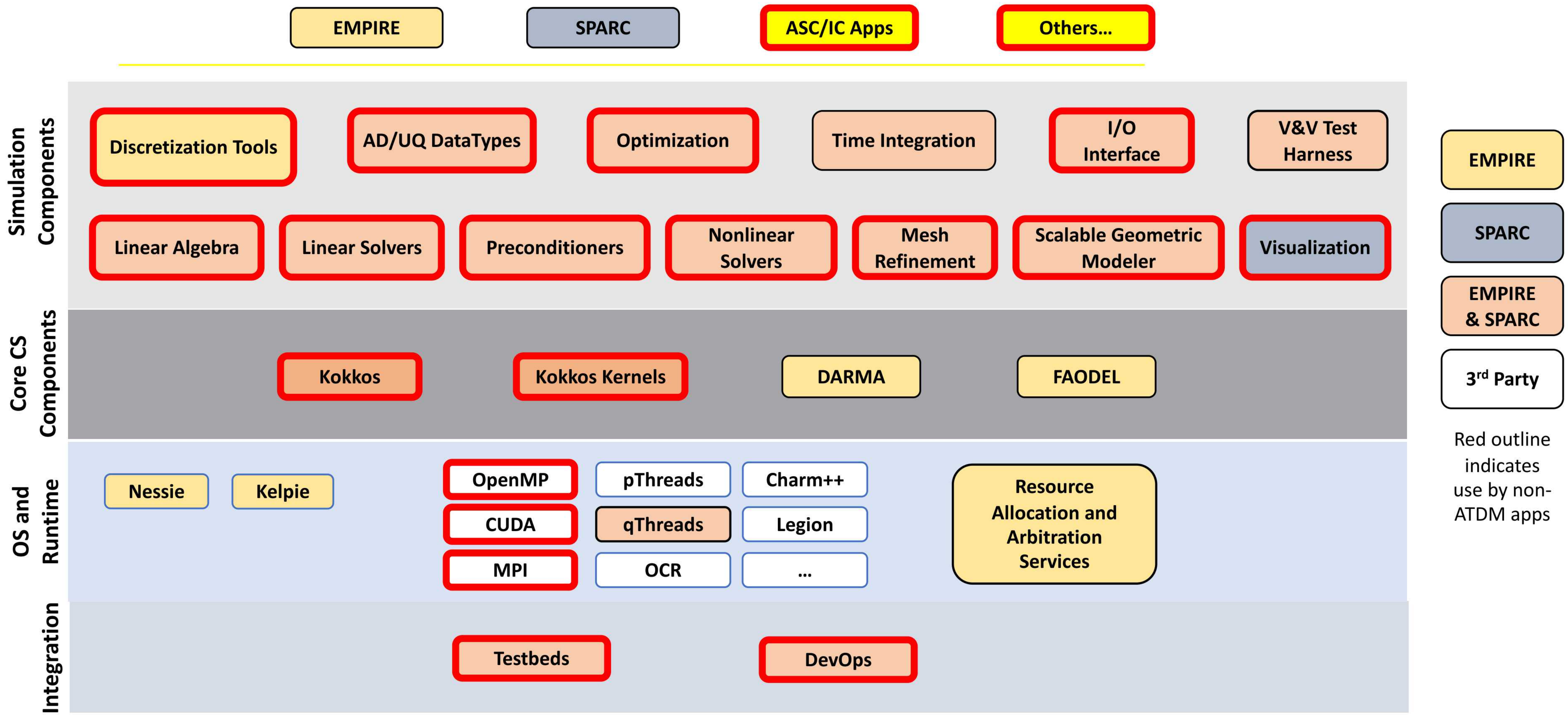
ITS SCEPTRE NuGET EIGER EMPHASIS Xyce Charon CHEETAH Gemma EMPIRE

Electromagnetics, Radiation, and Electrical Code Suite

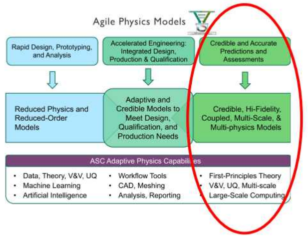
Agile Components

ATDM applications benefit as well as setting a broad foundation for other applications - *"Write once, use many"*

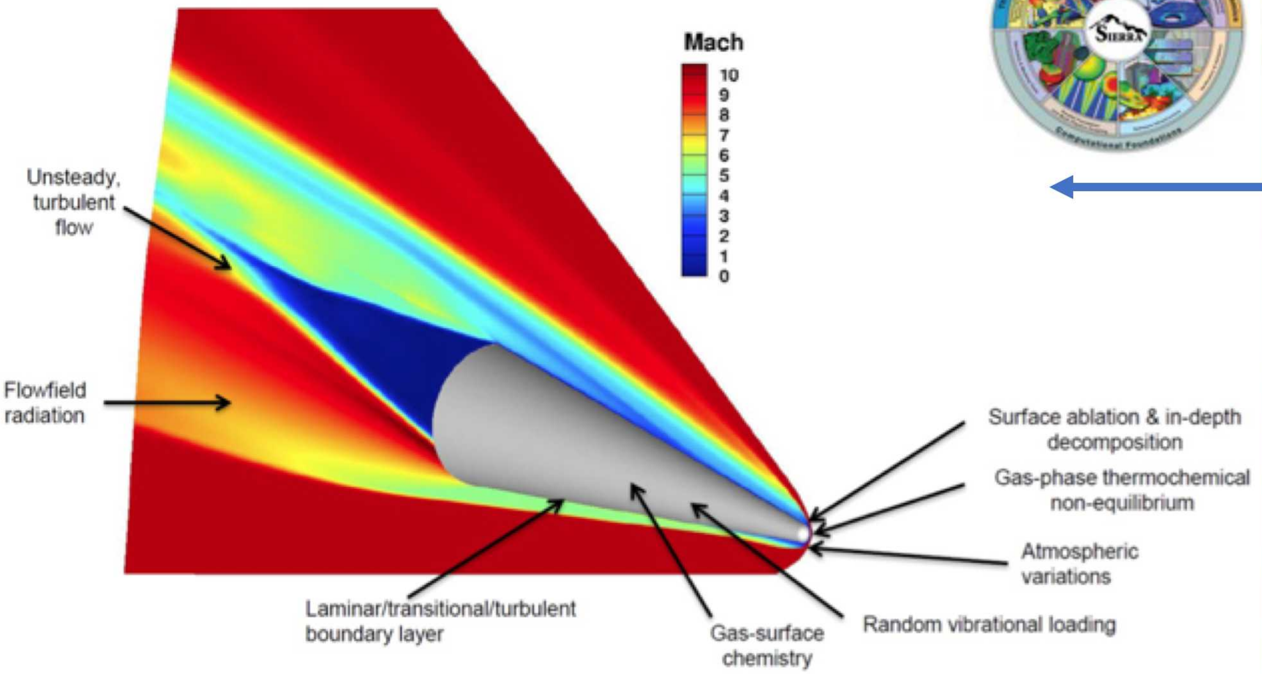
ATDM Components



Credible and Accurate Predictions and Assessments: (3-5-year goals)

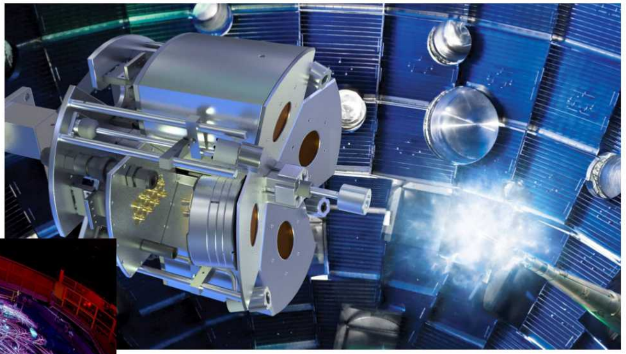


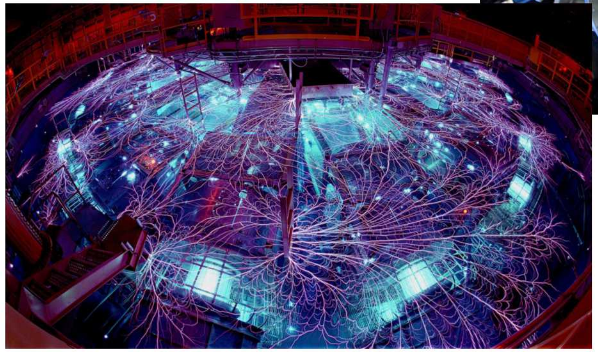
Re-entry Environments
SPARC



RAMSES
ITS SCEPTRE NuGET EIGER EMPHASIS Xyce Charon
CHEETAH Gemma EMPIRE

Plasma Simulation
EMPIRE





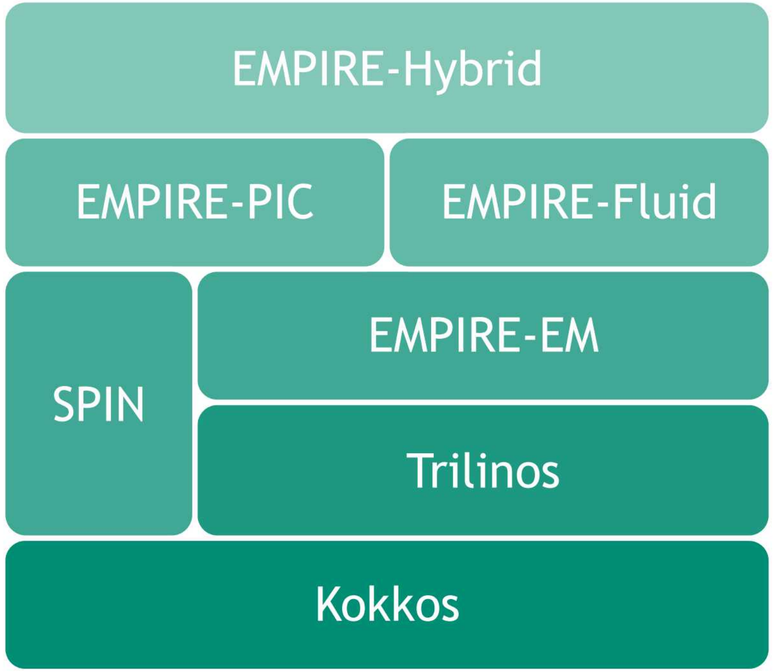
Gemma
Electromagnetic Radiation (EMR)

CHEETAH
Radiation Effects

- Productionize ATDM codes (transition to Integrated Codes)
- Develop key combined environments simulation capabilities (new physics + coupling)
- Follow-through on production code preparation for Next-Generation Platforms
- Leverage Strategic Partnership Programs (SPP) where possible

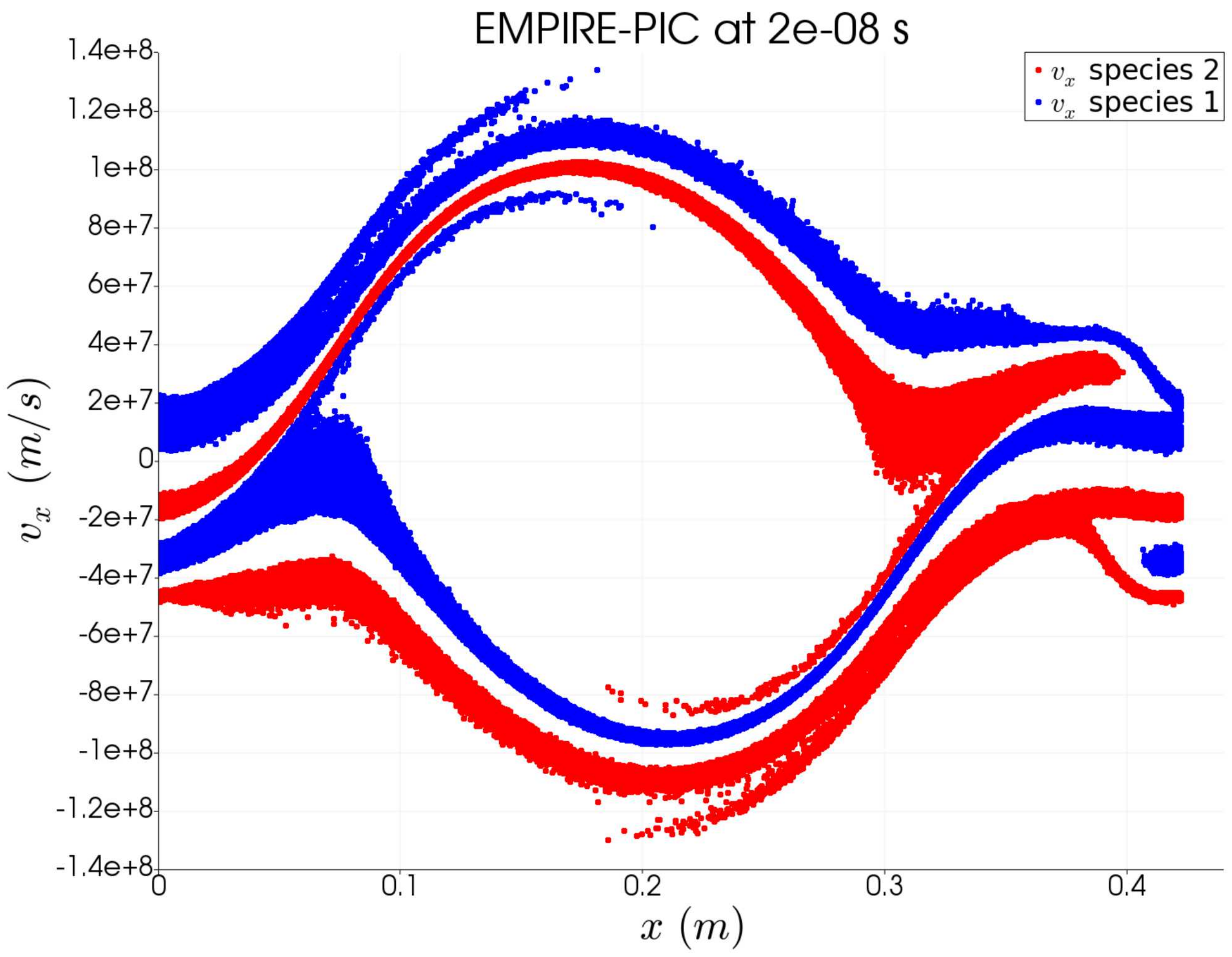
Advancing Plasma Physics Modeling

- EMPIRE leverages the opportunity from ATDM to advance plasma simulation capability on two fronts:
  - Component-based software design for portability across next-generation hardware architectures
  - New fluid and hybrid kinetic/fluid algorithms for validity and performance across a wider range of plasma density regimes
- EMPIRE is built upon Trilinos components:
  - Panzer: FEM discretization infrastructure
  - Tempus: General time integration package
  - Uses the modern Tpetra-based linear solver stack
  - Kokkos: Portable threading library
- EMPIRE will enable:
  - Higher fidelity modeling of critical plasma applications
  - Towards exascale simulation



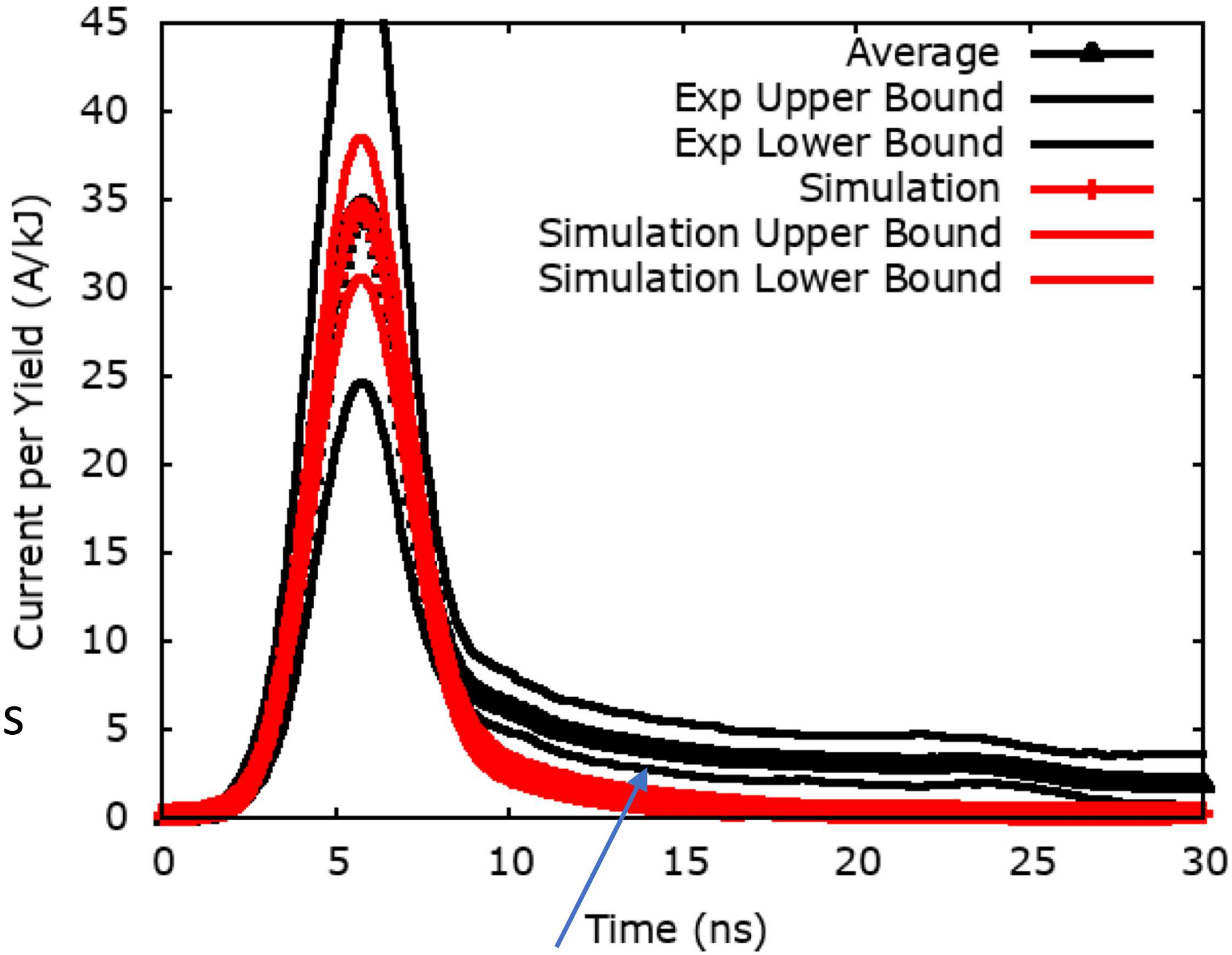
The Two-stream Instability Problem

- This test is looking forward to exercising the fluid and hybrid capabilities.
- The two-stream instability is one of the few phenomena that are present at number densities applicable to PIC, fluid, and hybrid and that has an analytic solution.
- Analytic theory (using an approximation) predicts the growth rate of the fastest growing mode until the instability becomes saturated to be $\frac{1}{\sqrt{8}}\omega_{pe} \approx 631\text{MHz}$.



Simulation of diode experiments on Z

- Diodes are driven by the photoelectric effect from X-rays created from Z machine shots
- Assume 4 current sensors from an experiment are independent measurements
- Over estimation of the error because the environment might not be the same
- Simulation error is the confidence interval assuming first order in dx, dt (CFL=6), and number or particles
    - dx=~height/8, ~height/16, and ~height/32

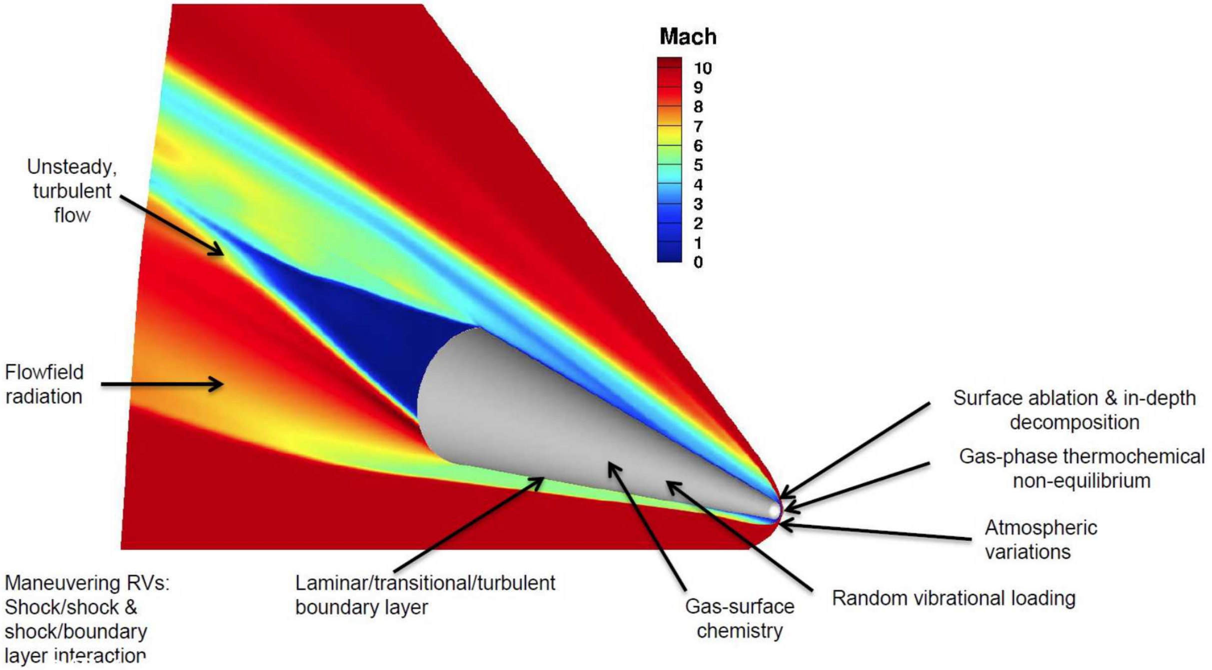


Late time current is larger in experiment, could be because of outgassing

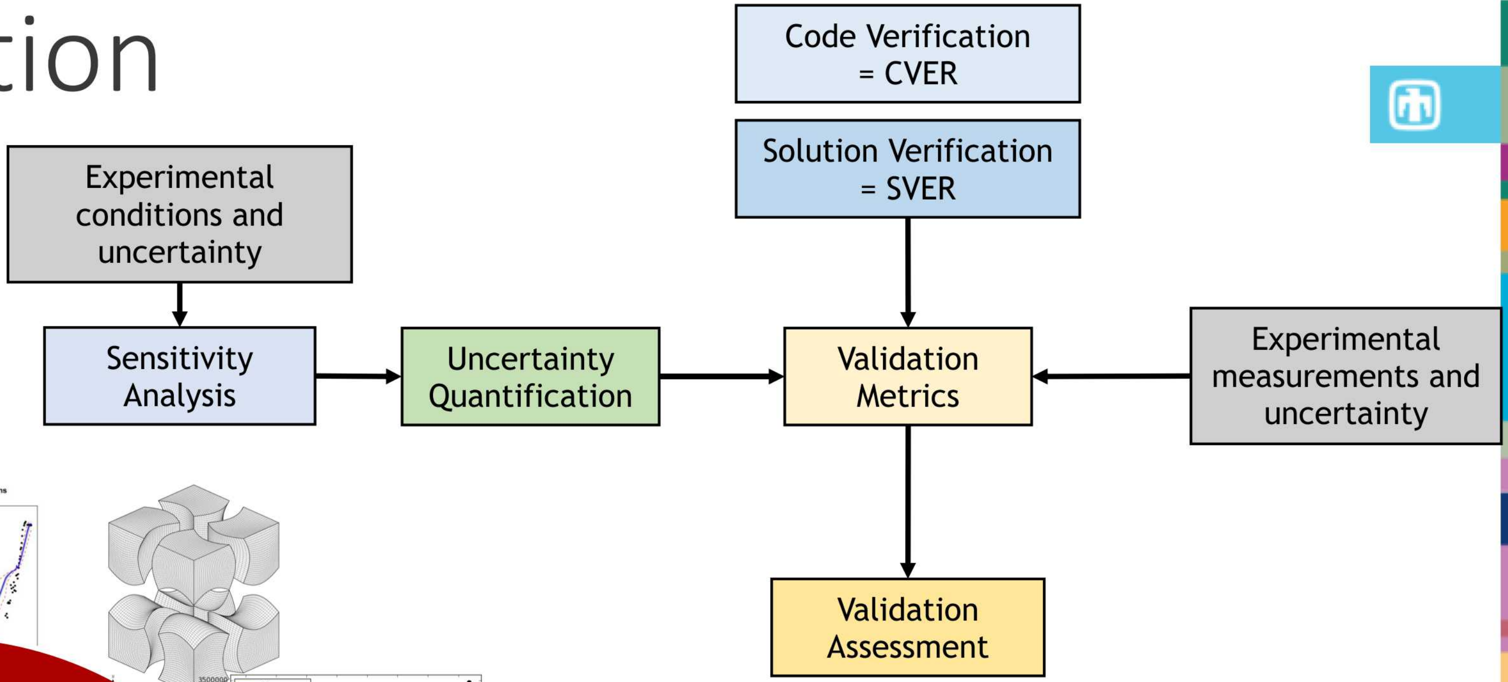
SPARC Basics

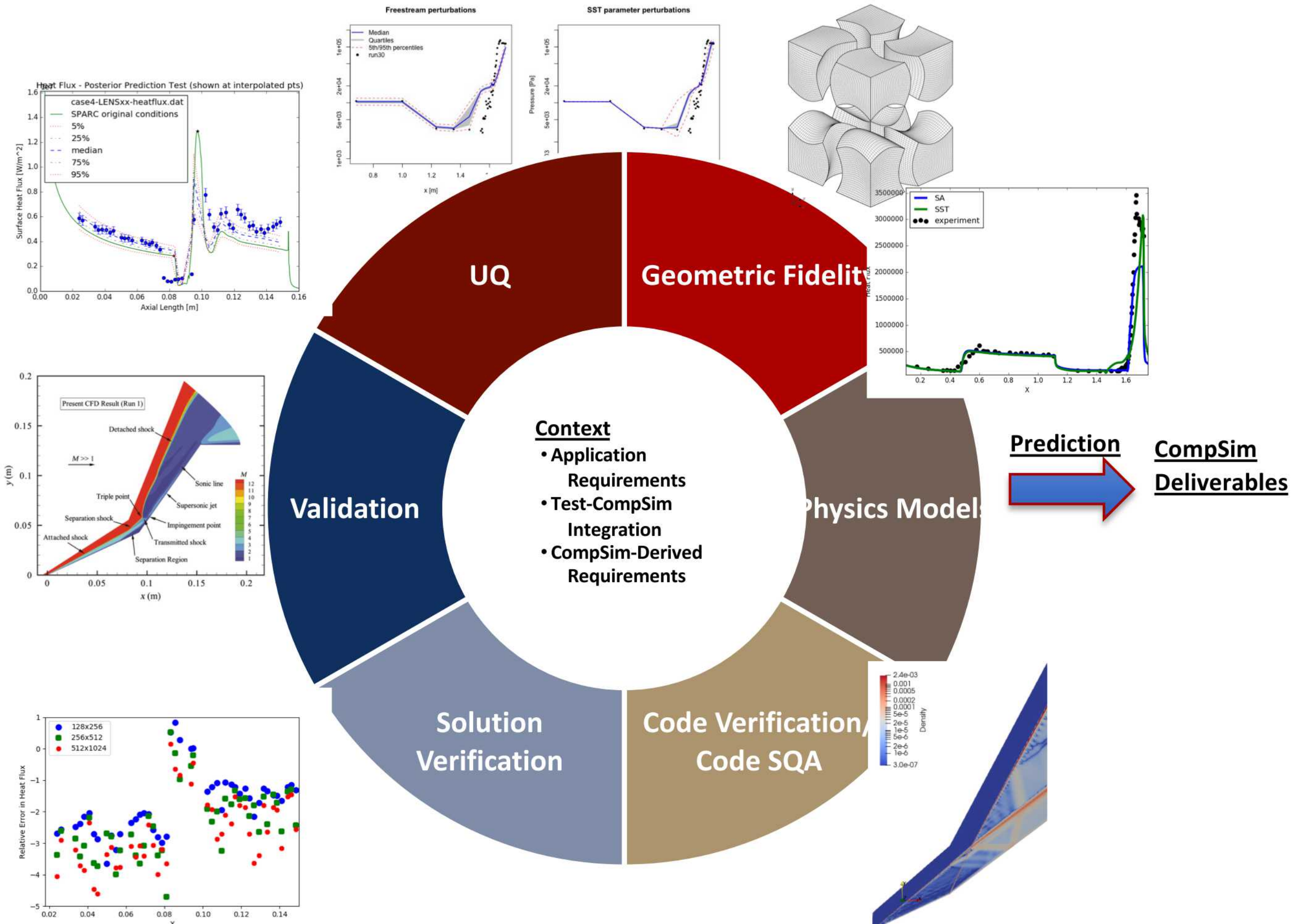
- State-of-the-art reentry simulation on next-gen platforms
  - Continuum compressible CFD (Navier-Stokes), hypersonic gas dynamics
  - Hybrid structured-unstructured finite volume methods. R&D: high order unstructured disc. collocation element methods
  - Perfect and thermo-chemical non-equilibrium gas models
  - RANS and hybrid RANS-LES turbulence models; R&D: Direct Numerical Simulation
- Enabling technologies/components
  - Performance portability through Kokkos
  - Scalable solvers
  - Embedded geometry & meshing
  - Embedded UQ and model calibration

- Credibility
  - Validation against wind tunnel and flight test data
  - Visibility and peer review by external hypersonics community
- Software quality
  - Rigorous regression
  - V&V
  - Performance testing

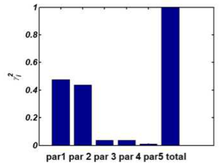


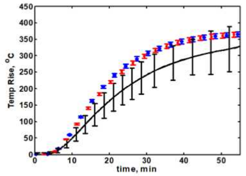
Approach: Verification & Validation / UQ Workflow



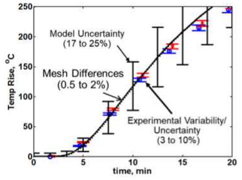


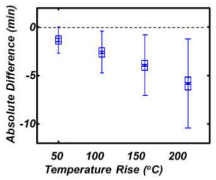
CompSim V&V/UQ Elements



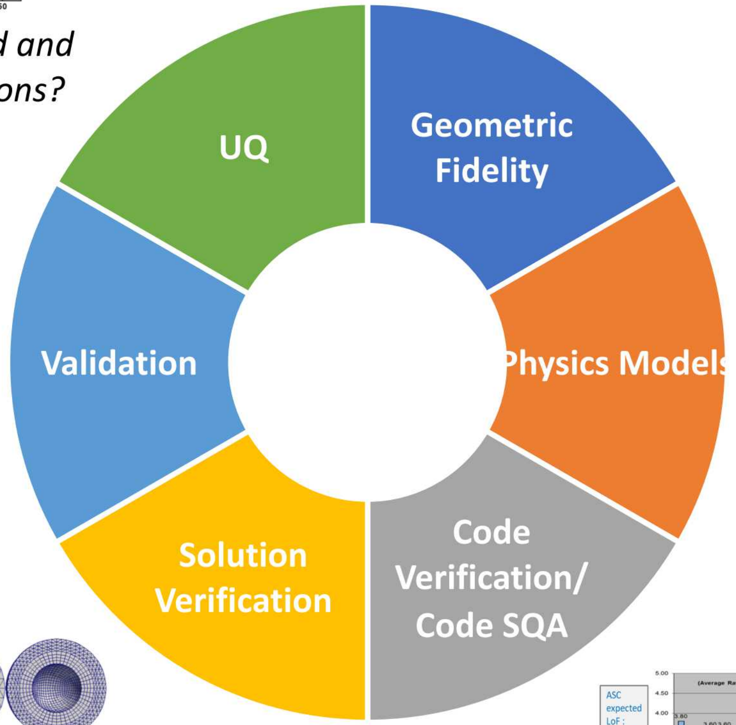


How are uncertainties assessed and reflected in simulation predictions?





What is the discrepancy between simulation and experiments?



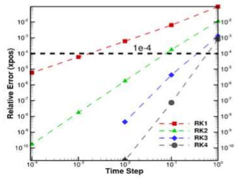
As-Modeled As-Designed

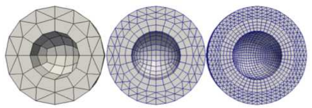
How are geometric feature simplifications influencing simulation results?

PIRT

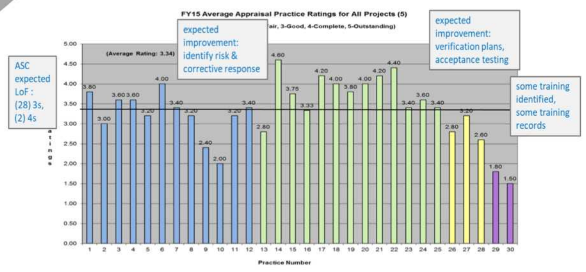
Phenomena	Importance	Adequacy for Intended Use			
		Math Model	Code	Validation	Model Parameter
Phenomena 1	H	H	M	L	L
Phenomena 2	M	H	M	L	L
Phenomena 3	L	H	M	L	L

Are important physics models adequate? Key gaps mitigated?



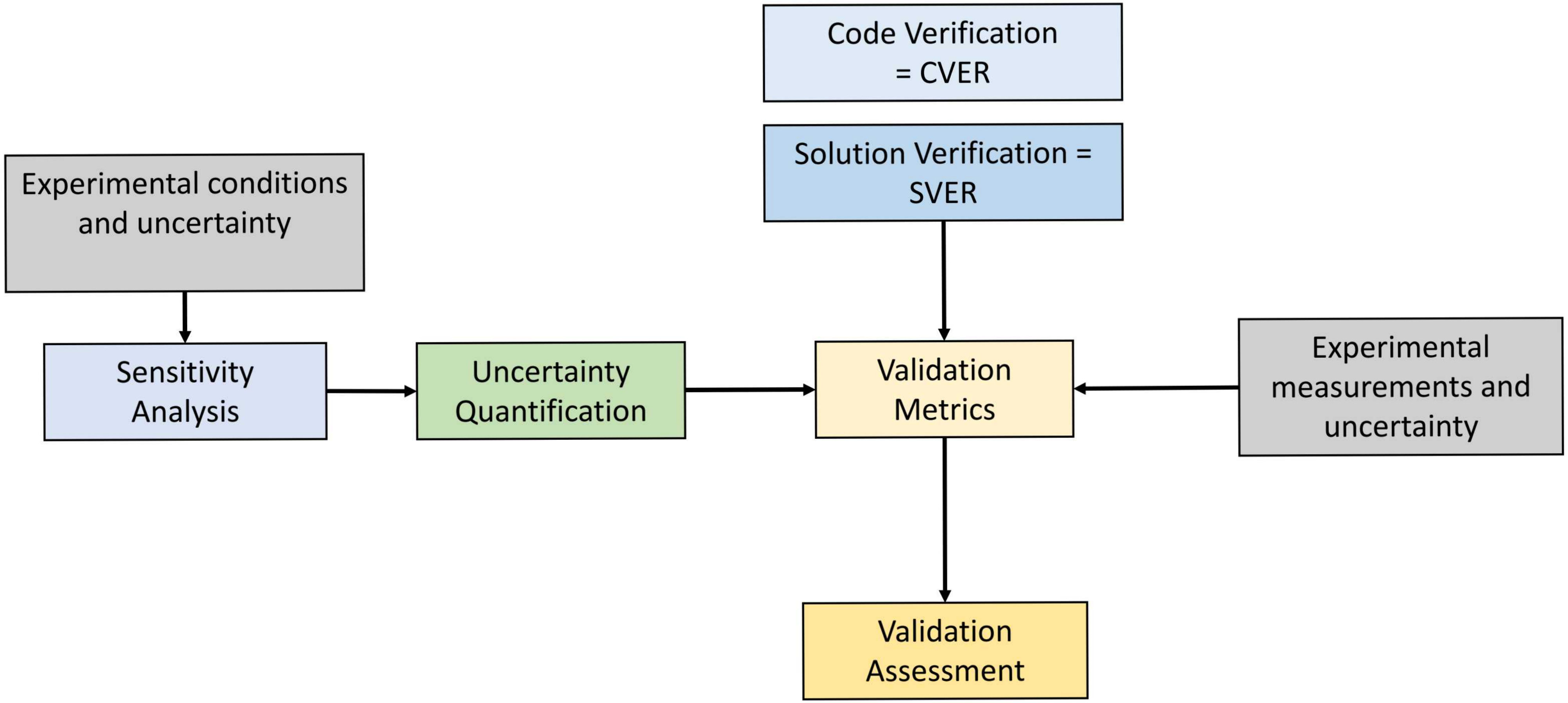


How do numerical solution or human errors affect simulation results?



What is the evidence for code credibility?

Approach: Verification & Validation / UQ Workflow



SPARC Code Verification Study

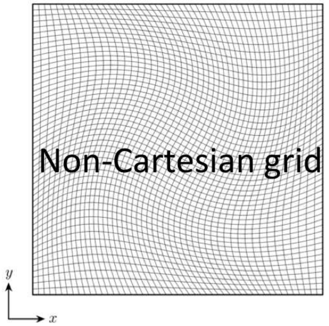
2D Inviscid Flow using a Trigonometric Manufactured Solution

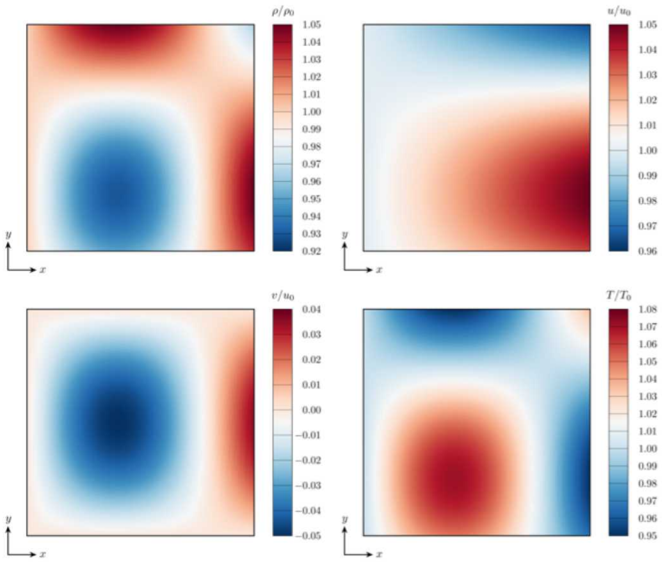
$$\rho(x,y) = \rho_0 \left[1 - \epsilon \sin\left(\tfrac{5}{4}\pi x\right)\left(\sin(\pi y) + \cos(\pi y)\right)\right],$$

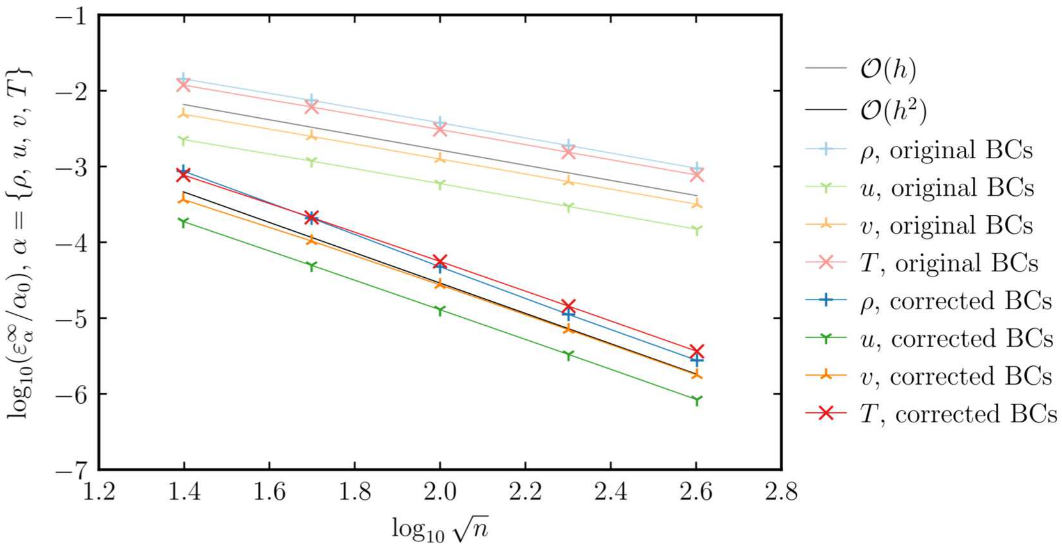
$$u(x,y) = u_0 \left[1 + \epsilon \sin\left(\tfrac{1}{4}\pi x\right)\left(\sin(\pi y) + \cos(\pi y)\right)\right],$$

$$v(x,y) = -\epsilon u_0 \sin\left(\tfrac{5}{4}\pi x\right)\sin(\pi y),$$

$$T(x,y) = T_0 \left[1 + \epsilon \sin\left(\tfrac{5}{4}\pi x\right)\left(\sin(\pi y) + \cos(\pi y)\right)\right],$$







Corrected boundary conditions lead to design convergence rates

Case #1: HIFiRE-1turbulent flow simulations

Relevance: RANS turbulence models in 3D flows

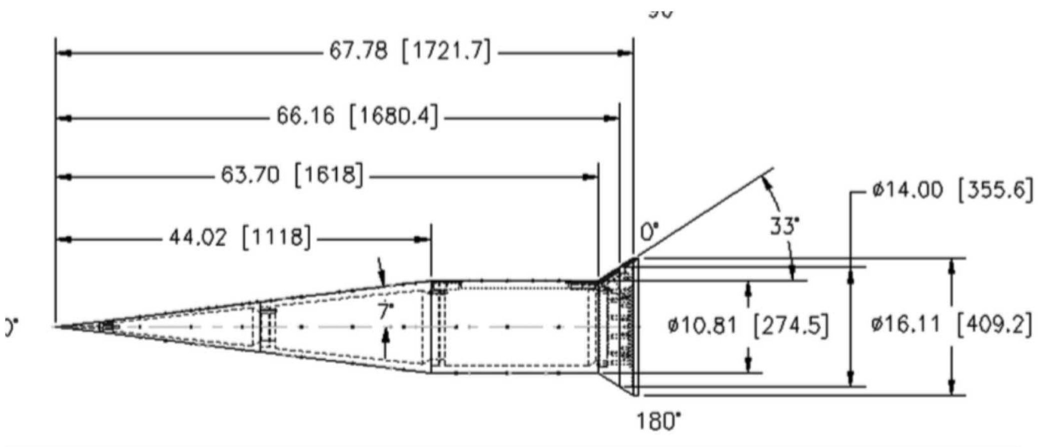
Tasks and Deliverables (ASC V&V):

- Q2: Completed SVER, UQ, and validation assessment.

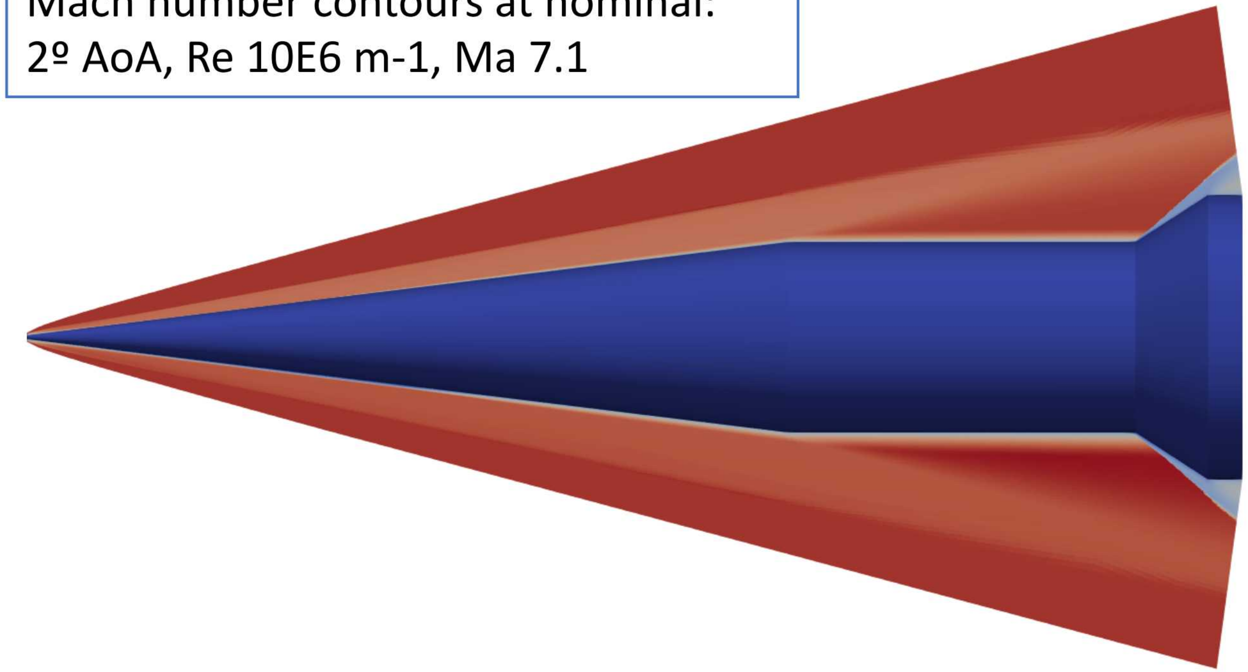
- **Aim:** Validate the Spalart-Allmaras (SA) and Shear Stress Transport (SST) turbulence models in SPARC
  - Low enthalpy flow (H_0 ~ 2.6MJ/kg), perfect gas
  - Transition to turbulence @ x = 0.45 m
  - Measurements: surface pressure and heat flux

Extend last year's HIFiRE-1 work to 3D:

- Generate regularly refined sequence of 3D meshes;
- Element counts: **2M/16M/128M**/1B/8B
- Full UQ, SVER for run 34, 2 deg. AoA (following slides)
- 2D/3D comparisons for axisymmetric runs



Mach number contours at nominal:
2º AoA, Re 10E6 m-1, Ma 7.1



HIFiRE-1uncertainty quantification / validation

Four input parameters (run 34)

- AoA, density, temperature, velocity
- Specify uniform random variables for each (+/-10%)

Propagation of uncertainty

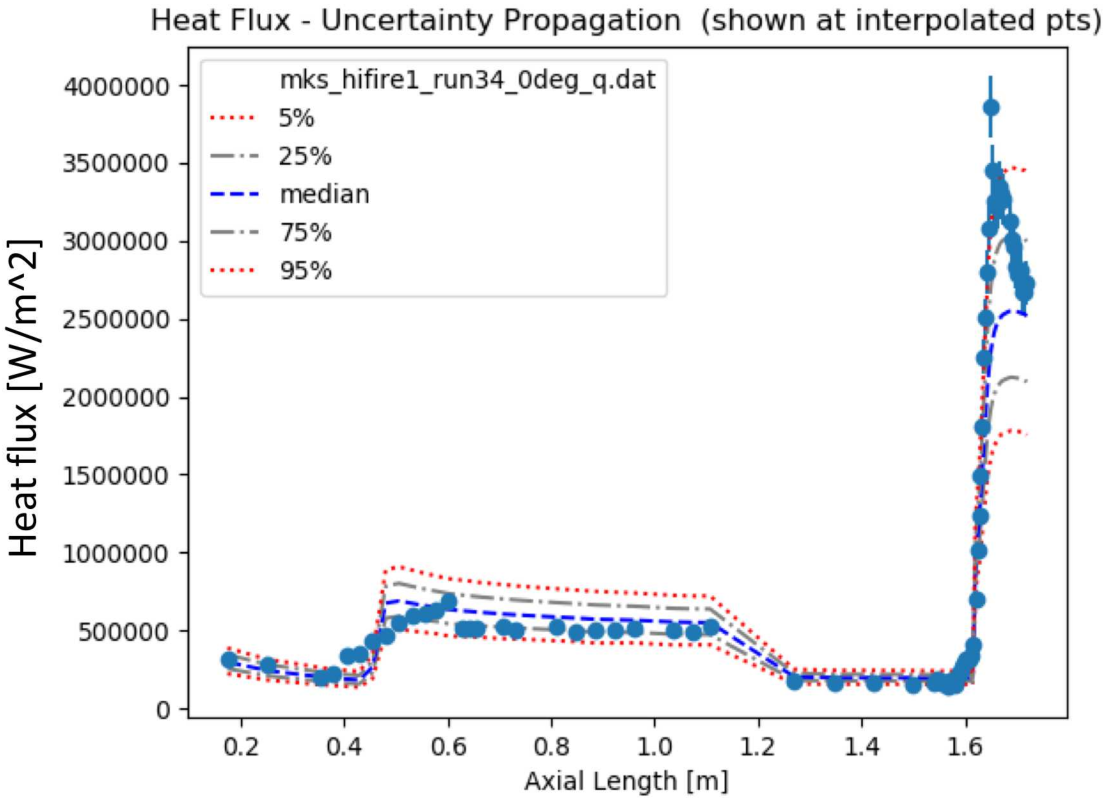
- PCE surrogate using sparse grid quadrature (level 2 = evals)
- Sampling of surrogate for statistics (10K samples)

Validation with uncertainty

- Compute probability levels (5/10/90/95%) at exp locations
- Compare SA/SST models

Conclusions

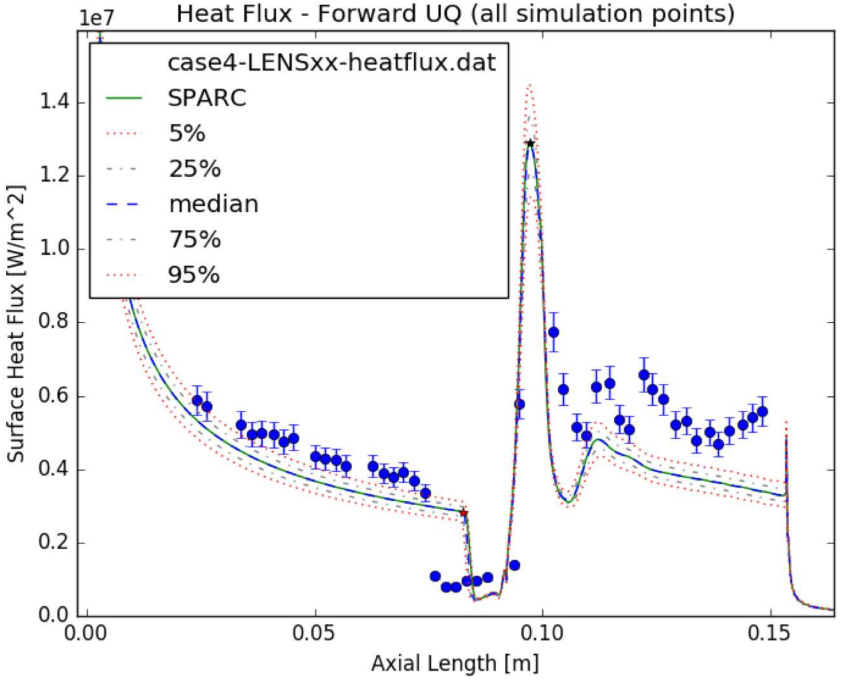
- As expected, RANS models not reliable predictors of transition or separation, but surface heat flux and pressure are otherwise well predicted.

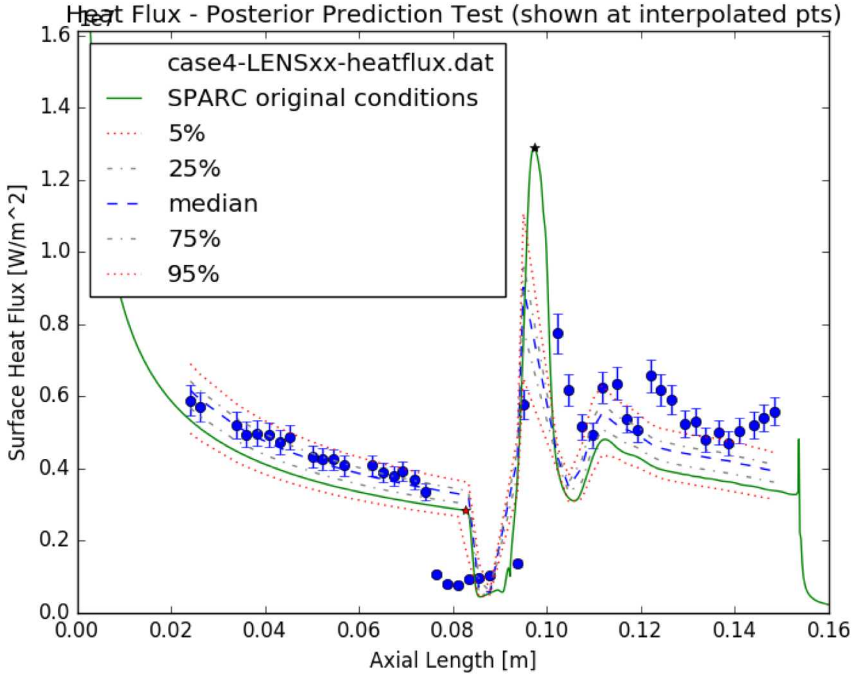


Heat flux for fully 3D study flow using SPARC's SA turbulence model. Quantile curves represent uncertainty. Propagation. Experiment error bars +/- 5%, as provided. Flow at 2º AoA, Re~10E6 m-1, Ma~7.1

SPARC Validation Result

Improved Predictions After Calibration of Experimental Input Flow Conditions to Measurements

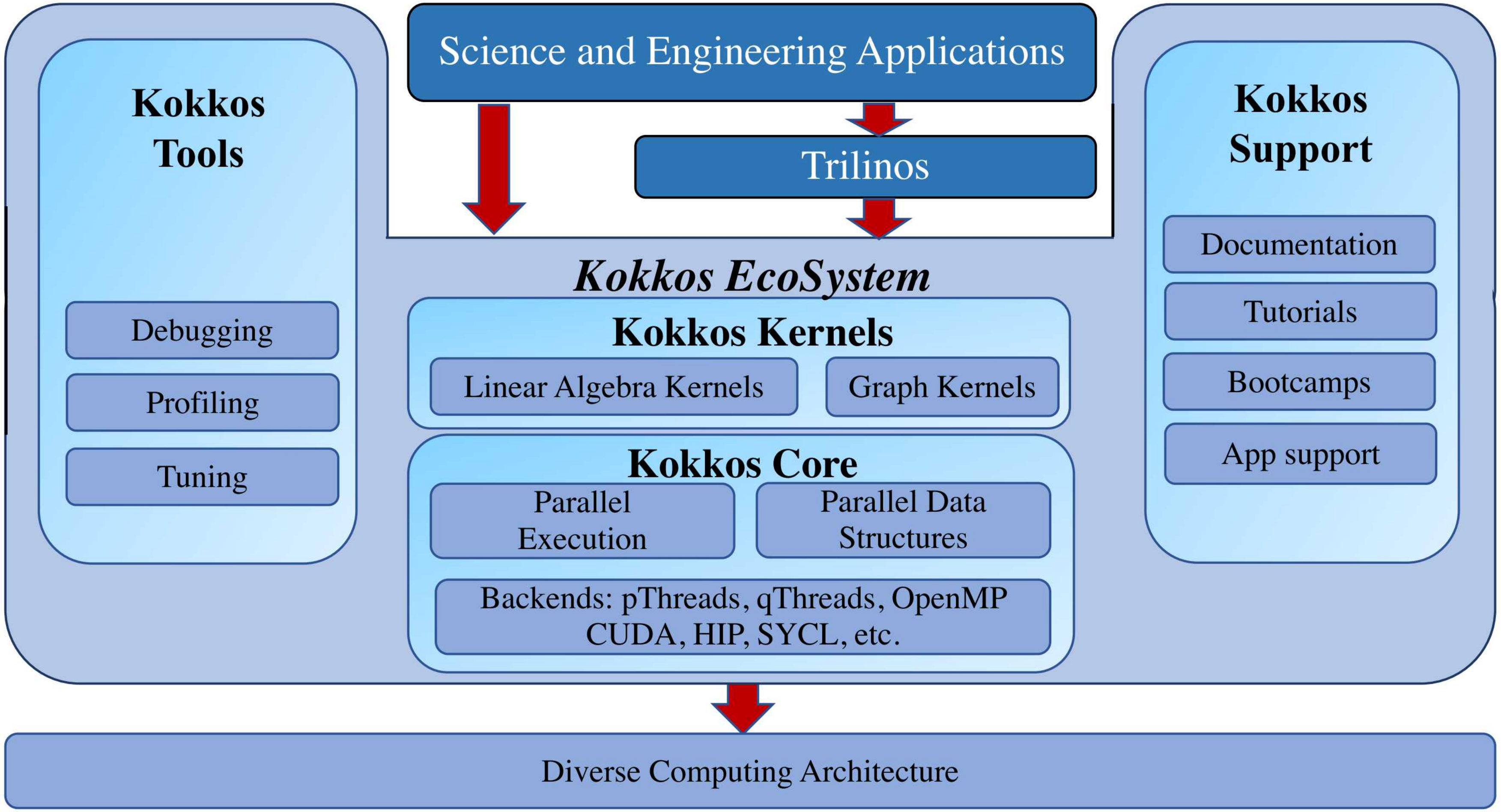




Before calibration → DAKOTA Explore and predict with confidence. → After calibration

Challenge: Initial validation result indicated poor agreement with experimental measurements *(left plot)*
Hypothesis: Poor agreement is due to **mis-specified input (freestream) flow conditions** in experiment
Approach: Apply Dakota's Bayesian capability to calibrate freestream density and velocity to measurements
Result: Significant improvement of computed SPARC heat flux when compared with experiment *(right plot)*

The Kokkos EcoSystem



Kokkos Core: parallel patterns and data structures; supports several execution and memory spaces

Kokkos Kernels: performance portable BLAS; sparse, dense and graph algorithms

Kokkos Tools: debugging and profiling support

Kokkos enables performance portability and the complexity of supporting numerous architectures that are central to DOE HPC enterprise

Kokkos Impact and Growth

- Expanding solution for common needs of modern science/engineering codes
- It is Open Source - maintained and developed at https://github.com/kokkos
- It has many users at wide range of institutions
- Now funded about 50/50 by ATDM and Office of Science ECP
  - Kokkos ECP project extended and refocused to include developers at Argonne, Oak Ridge, and Lawrence Berkeley - staffing is in place
  - HIP backend for AMD: main development at ORNL
  - SYCL for Intel: main development at ANL
  - OpenMP target for AMD, Intel and NVIDIA: lead at Sandia
- The Kokkos Team is a primary HPC contributor to the C++ standard
  - About half of all HPC representatives at C++ committee are associated with Kokkos
- Goal: make Kokkos a sliding window of advanced capabilities for HPC Performance Portability
  - Develop and prove new techniques, concepts and abstractions then introduce into the C++ standard

Sandia National Laboratories

Los Alamos NATIONAL LABORATORY EST. 1943

Argonne NATIONAL LABORATORY

OAK RIDGE National Laboratory

BERKELEY LAB

CSCS

Kokkos Ecosystem will be supported on all of the DOE leadership class platforms up through El Capitan

A Word on Performance and Portability

Relative node performance (**measured against CTS systems**)

		CTS1	Trinity		Sierra		Astra
		Broadwell	Haswell	KNL	POWER9	V100 GPU	ThunderX2
LINPACK FLOP Rates (per Node)	Perf	1.09 TF/s	~0.86 TF/s	~2.06 TF/s	~1 TF/s	~21.91 TF/s	~0.71 TF/s
	Rel	1.00X	0.79X	1.89X	0.91X	20.01X	0.65X
Memory Bandwidth (STREAM) (per Node)	Perf	~136 GB/s	~120 GB/s	~90 GB/s / ~350 GB/s	~270GB/s	~850 GB/s x 4 = ~3.4 TB/s	~250 GB/s
	Rel	1.00X	0.88X	0.66X	1.99X	25.00X	1.84X
Power (TDP, per Node)	Perf	120W x 2 = 240W	135W x 2 = 270W	~250W	190W x 2 = 380W	~300W x 4 = 1.2kW	~180W x 2 = 360W
	Rel	1.00X	1.13X	1.04X	1.58X	5.00X	1.50X

“Next Generation Platforms” and performance portability

- Next Generation Platform (NGP): a high-performance computer that has a new generation computing architecture that requires very different programming model to fully utilize the hardware

ATS-1 (Trinity): ~30 Pflops aggregate,
~3 Tflop 68-core Xeon Phi processors,
Top500: #6 on HPL, #4 on HPCG
production in 2017

For reference:
Top DoD HPC (Onyx): ~6 Pflops
Top500: #48 on HPL

Vanguard-1 (Astra): ~2 Pflops aggregate,
~1 Tflop 56-core Thunder X2 ARM processors,
Top500: #204 on HPL, #36 on HPCG
production in late 2019

ATS-2 (Sierra): ~125 Pflops aggregate,
4x ~8 Tflop Volta100 GPUs,
Top500: #2 on HPL, #2 on HPCG
production in mid 2019

ATS-3 (Crossroads): ~500 Pflop aggregate,
??? Architecture, delivery in late 2021

ATS-4 (El Capitan): 1 Eflop aggregate,
??? Architecture, delivery in late 2023

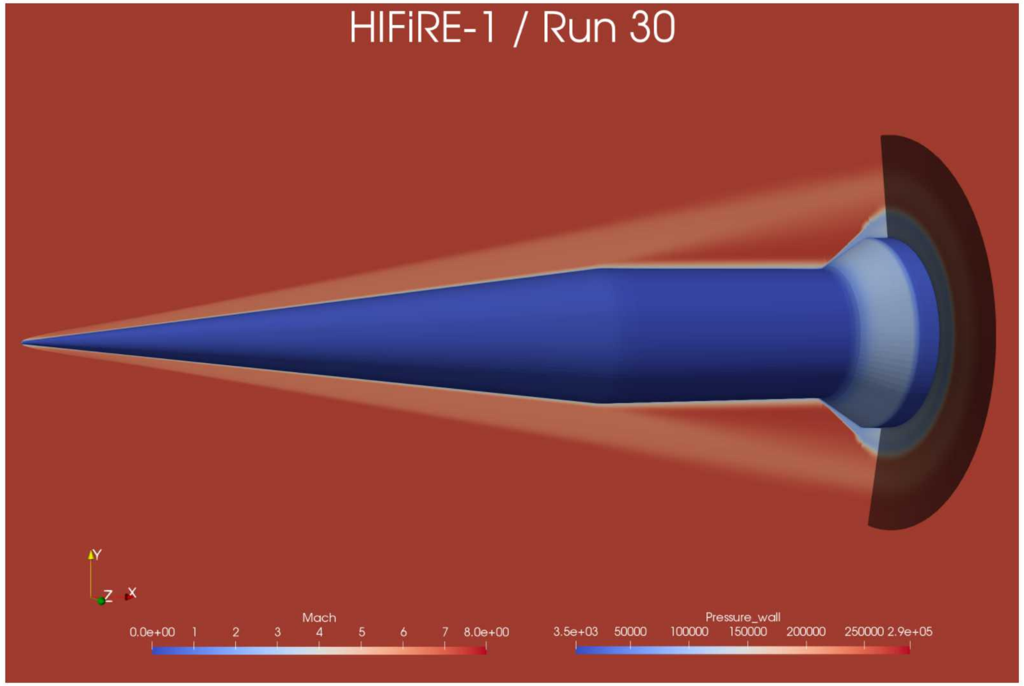
We estimate it takes 10 years to develop, validate and productionize a mod/sim code for weapons qualification

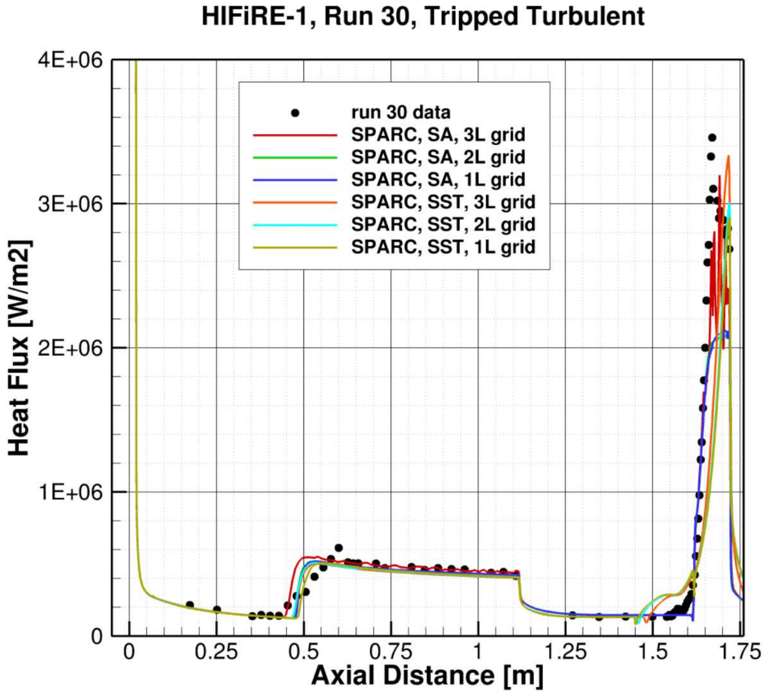
We’re looking at 5 different architectures that may require 5 different programming models in an 8 year span

Kokkos is a foundational element of SPARC in achieving performance portability

HIFiRE-1 Performance Analysis and Optimization

- Use-case: compute the steady-state, RANS aero solution for the HIFiRE-1 geometry

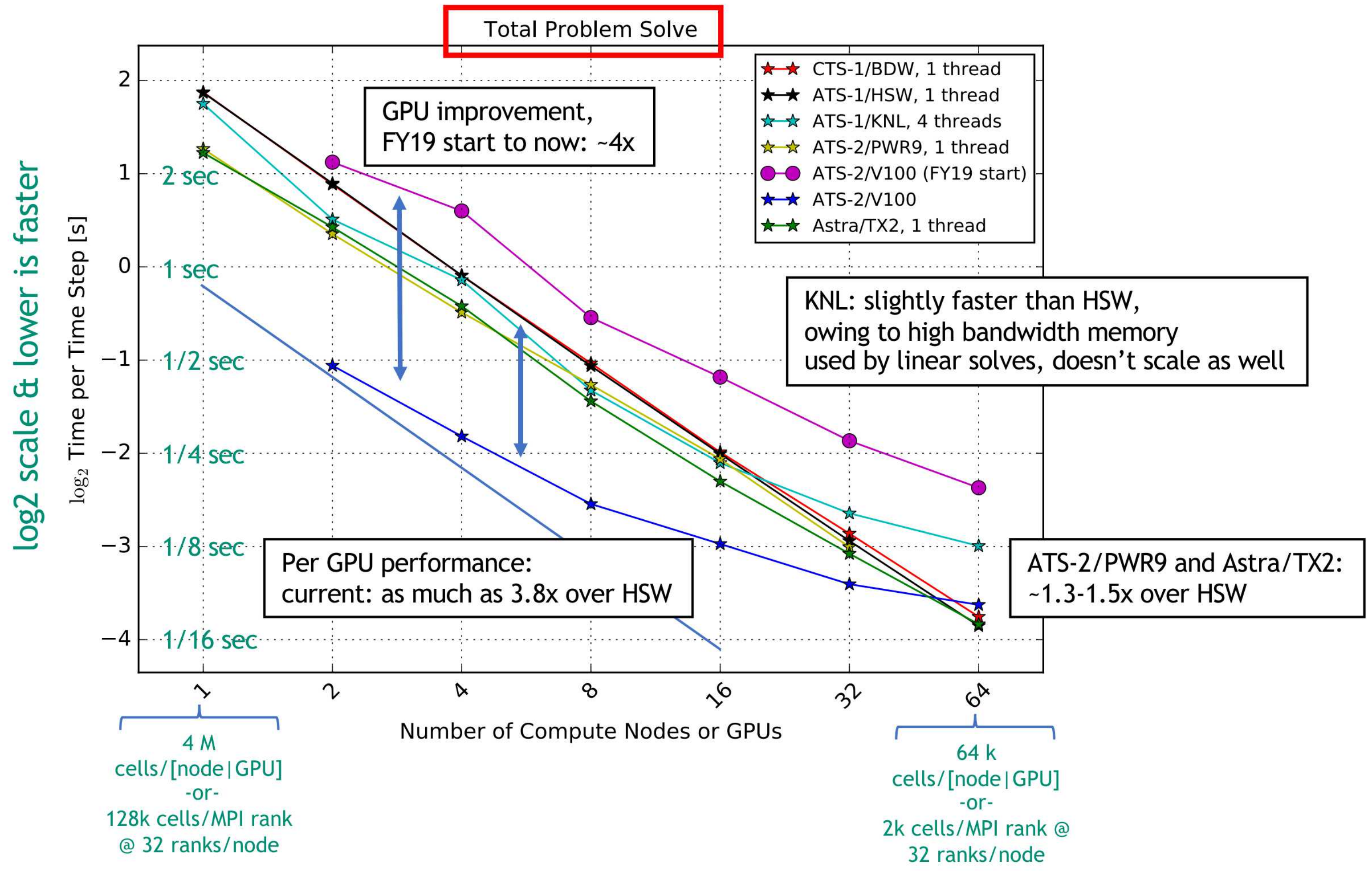




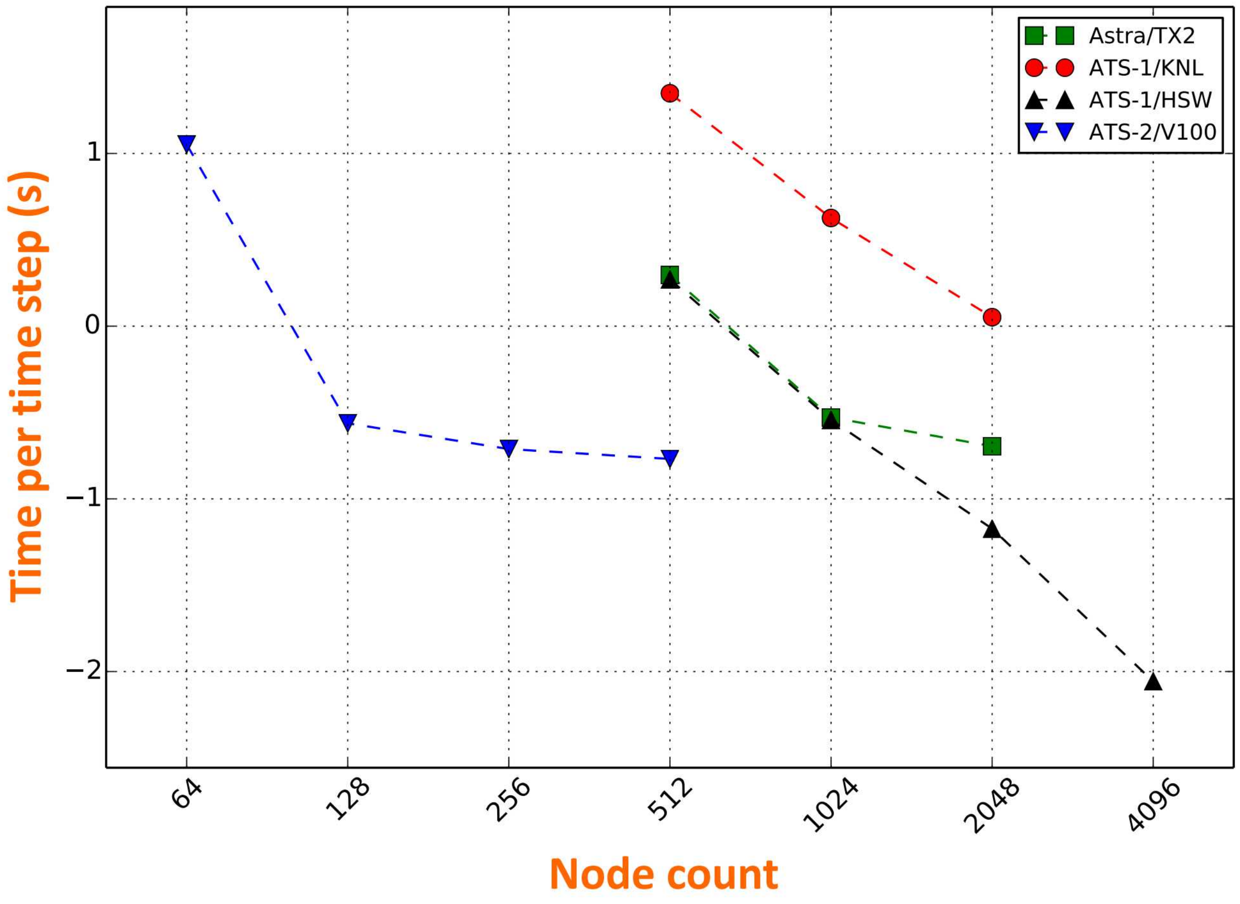
Using the following systems:
- ATS-1's (Trinity) Xeon (Haswell) nodes
- ATS-1's (Trinity) Xeon Phi (Knight's Landing) nodes
- ATS-2's (Sierra) V100 GPU nodes

Note: this is the same case being considered by the validation team

HIFiRE-1 Strong Scaling



Strong scaling performance of EMPIRE



Comparison EMPIRE performance on Astra (ARM), Trinity/KNL, Trinity/HSW, and Sierra for a problem with 166 million elements, 195 million DOFs in the linear solve, and 8.2 billion particles. Linear solve struggles to scale on GPU architecture.