Performance Engineering FUN3D at Scale with TAU Commander

John C. Linford*[1], Srinath Vadlamani[1], Sameer Shende[1], Allen D. Malony[1], William Jones[2], William Kyle Anderson[2], and Eric Nielsen[2]

[1]*ParaTools, Inc. Eugene, OR 97405, U.S.A.*
[2]*NASA Langley Research Center, Hampton, VA 23681, U.S.A.*

The FUN3D Solver

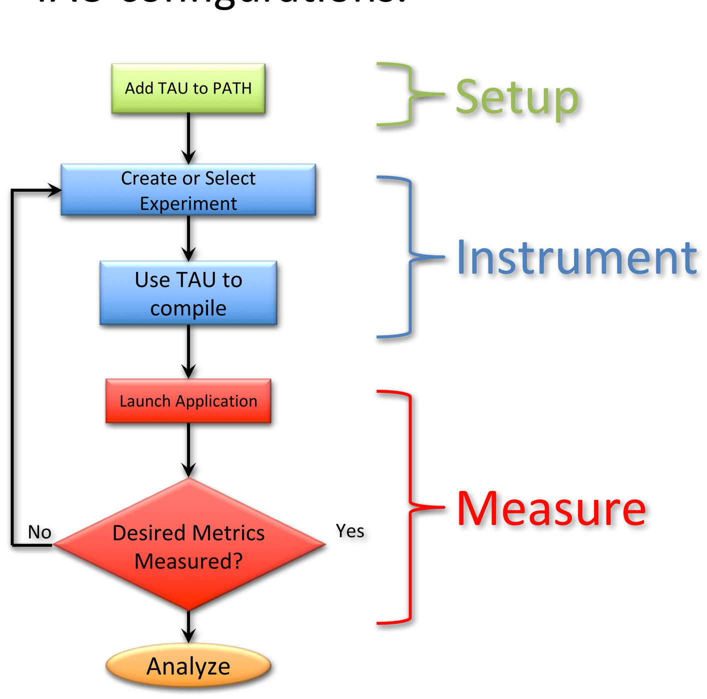
NASA Langley Research Center's FUN3D software is an unstructured-grid computational fluid dynamics suite used to tackle complex aerodynamics problems. The toolset enables multidisciplinary capabilities through coupling to variable fidelity models encompassing structural effects, multi-body dynamics, acoustics, radiation, optics, propulsion, and ablation. FUN3D provides the world's foremost adjoint-based design capability, enabling formal optimization of time-dependent moving-body simulations involving turbulent flows. The adjoint formulation is also used to perform mathematically-rigorous mesh adaptation and error estimation.

Credit: US Army

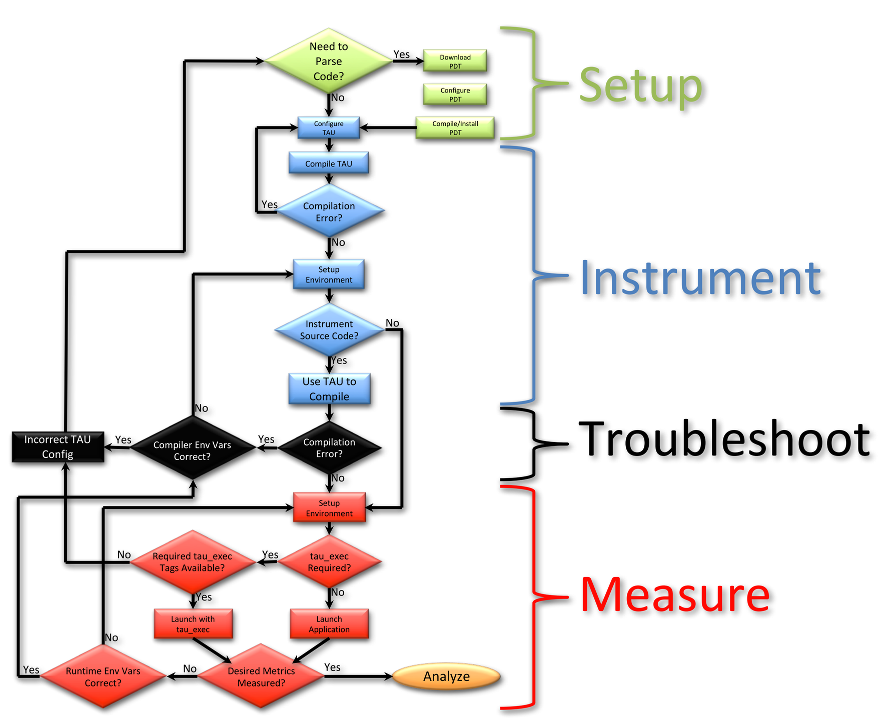
FUN3D is widely used to support major national research and engineering efforts at NASA and among groups across U.S. industry, other government agencies, and academia. A past collaboration with the Department of Energy received the prestigious Gordon Bell Prize, which recognizes outstanding achievements in high-performance computing.

TAU Commander

TAU Commander simplifies the TAU Performance System® by using a structured workflow approach that gives context to a TAU user's actions. This eliminates the troubleshooting step inherent in the traditional TAU workflow and avoids invalid TAU configurations.



TAU Commander Workflow

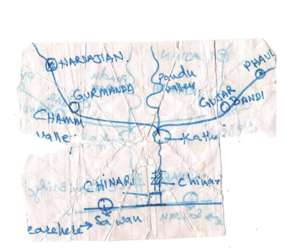


Traditional TAU Workflow

A study of 124 workflows demonstrated that using TAU Commander reduces the number of unique steps in the performance workflow by approximately 50% and reduces the number of commands a user must know from approximately eight to exactly one. TAU Commander is installed on several DoD DSRC systems.

Improve productivity by specifying the desired result, not step-by-step directions.

VS



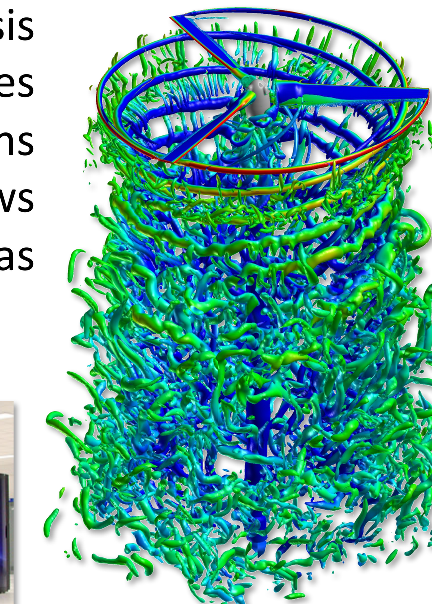
Acknowledgments

This work was developed in part by the User Productivity Enhancement, Technology Transfer and Training (PETTT) Project No. PETA-KY07-001 and by DoE SBIR Grant No. DE-SC0009593. The authors would like to acknowledge the computational resources and PETTT software support from the DoD High Performance Computing Modernization office under Contract No. GS04T09DBC0017.

* Corresponding author: jlinford@paratools.com

FUN3D at the Department of Defense

FUN3D is being applied to a broad spectrum of analysis and design problems across all the major service branches at the Department of Defense (DoD). These applications span the speed range from subsonic to hypersonic flows and include both fixed- and rotary-wing configurations as well as a diverse array of weapons systems.

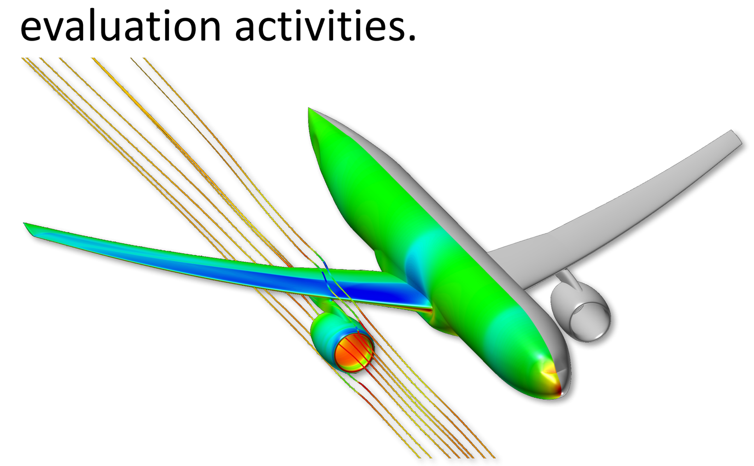


Coupled Helios-FUN3D Simulation Courtesy Rohit Jain, US Army Aviation Development Directorate - AFDD

FUN3D is in routine use across the various DoD Supercomputing Resource Centers (DSRCs). To accommodate an ever-increasing demand for larger and more complex simulations, the FUN3D development team is partnered with computational experts from ParaTools, Inc. through the DoD/Engility Productivity Enhancement, Technology Transfer and Training Project (PETTT). Through this collaboration, the team is effectively identifying and addressing computational barriers encountered at scale.

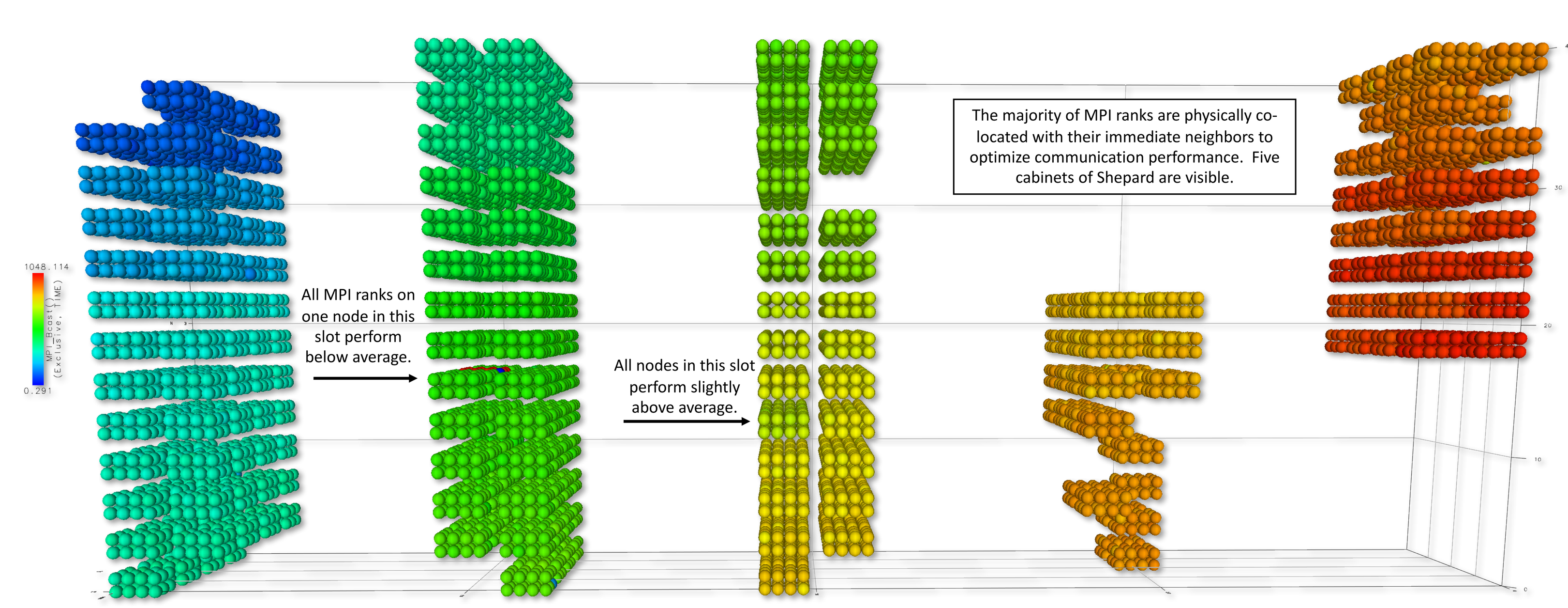
Project Goal

The team is using TAU commander and related tools on the DoD DSRC systems to study FUN3D computational performance and to guide optimization efforts. TAU Commander highlights source code regions that limit scalability through profiling, tracing, and aggregate summary statistics with respect to computational time, memory allocation, and memory access patterns. The analysis approach is being carefully documented to assist other DoD groups in similar performance evaluation activities.



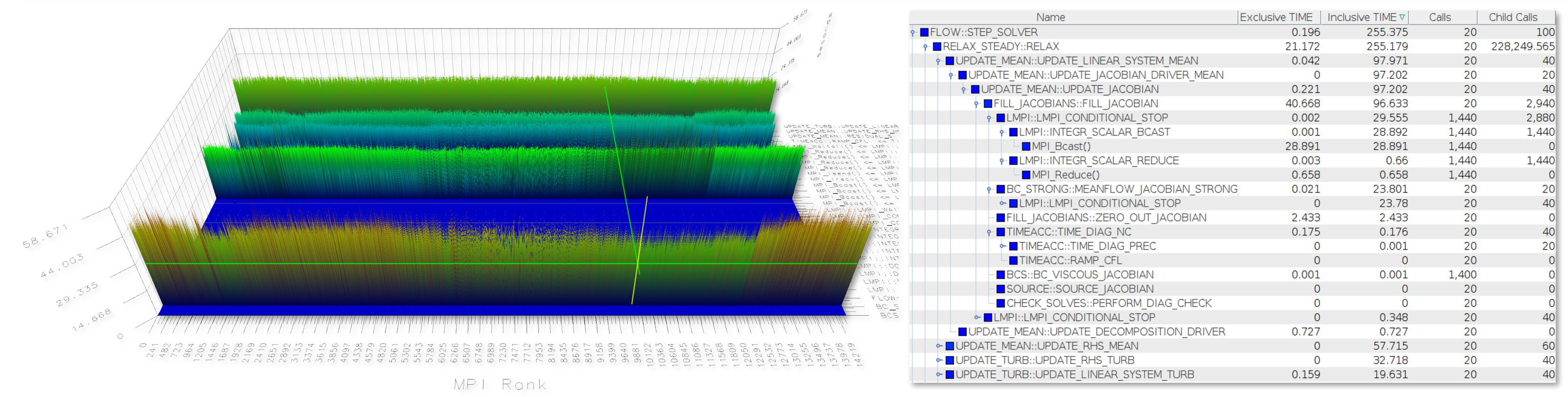
- **1,651,089,924 grid points**
- **5,902,801,476 tetrahedral elements**
- **1,310,290,264 prismatic elements**
- **14,400 Ivy Bridge cores**

To establish a baseline database of FUN3D performance metrics, the team used TAU Commander to profile a high Reynolds number simulation of the flow over a wing-body-pylon-nacelle geometry. The computation was performed using 600 nodes (14,400 Intel® Xeon® Ivy Bridge cores) on *Shepard*, a Cray XC30 system located at the Navy DSRC.

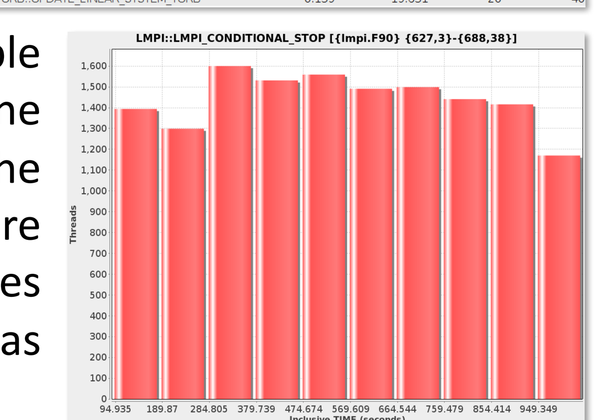


Topology plot showing physical locations of all MPI ranks.

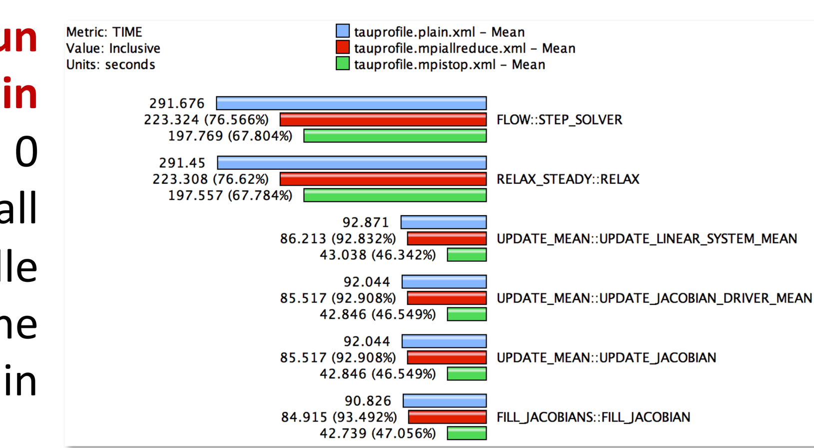
Performance Analysis



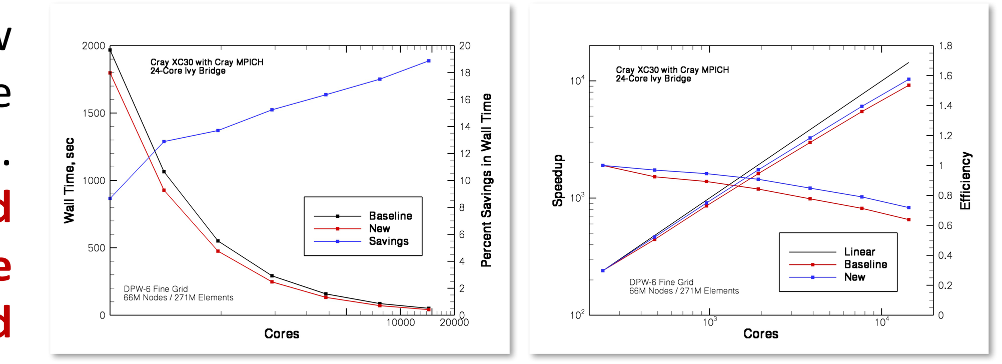
Load imbalance in `FILL_JACOBIAN` is easily visible by comparing against the lowest time in this routine and viewing the function histogram. On average, the routine took 96.6 seconds, of which 28.8 seconds are spent in `MPI_Bcast`. The broadcast operation takes place in `LMPI_CONDITIONAL_STOP`, which has dramatically different times across all MPI ranks.

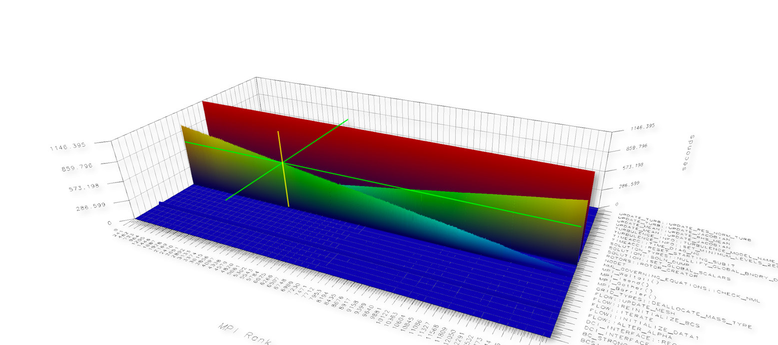


Analyzing data from the 14.4k core run revealed a communication bottleneck in `LMPI_CONDITIONAL_STOP`. Rank 0 performs a data reduction on behalf of all other ranks, so all other ranks remain idle until Rank 0 finishes the calculation. The severity of the problem grew linearly in the number of MPI ranks.



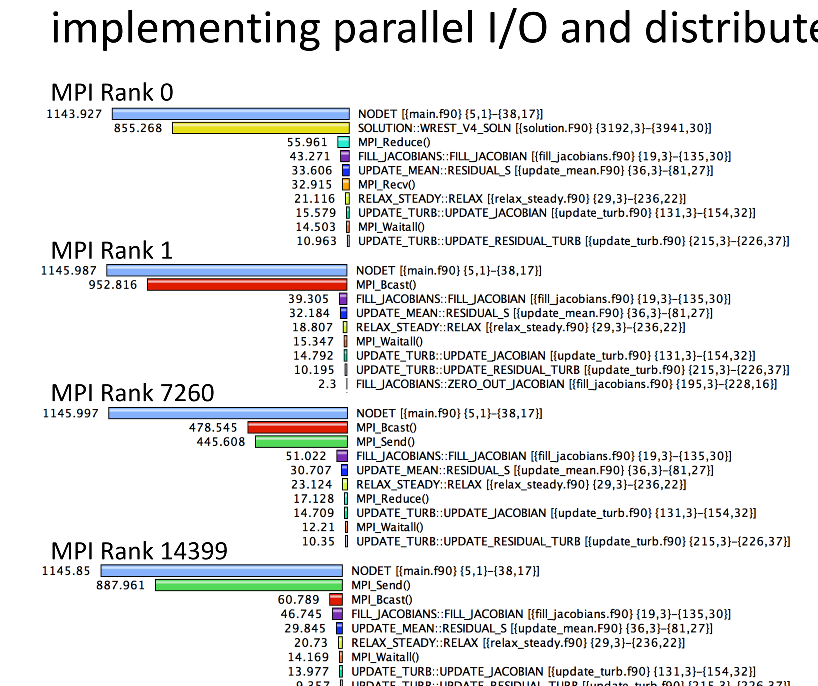
The team implemented a new approach that did not require the data reduction operation. **This improvement reduced runtime by up to 33% in core solver routines and is included in the FUN3D 13.0 release.**

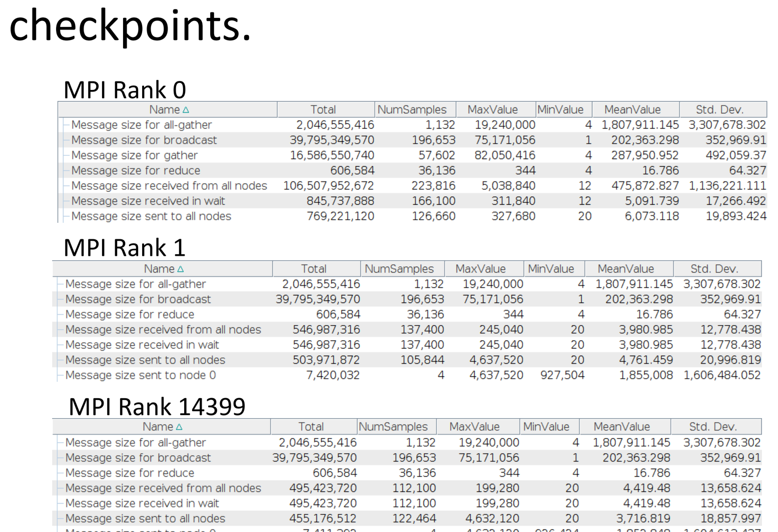




When writing the solution file to disk, there is a strong correlation between MPI rank and time spent in `MPI_Bcast` and `MPI_Send`. This operation occurs only once every few hours in production. **Further optimization should focus on load balance in the PDE solution.**

`MPI_Bcast` is called the same number of times with the same message size for all ranks, yet time spent in `MPI_Bcast` varies dramatically by rank. Rank 0 receives data from all others and writes it to disk, hence N-1 ranks spend the majority of their time in this operation waiting for I/O on Rank 0. This suggests implementing parallel I/O and distributed checkpoints.





Additional information, visualizations, and slides.

DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.