DEVCOM
ARMY RESEARCH LABORATORY

High Performance Computing Review
2020 - 2021

CONTENTS

INTRODUCTION & INFORMATION

FACILITIES

COMPUTATIONAL SYSTEMS

SERVICES

HPC SUCCESS STORIES

WORKFORCE OF THE FUTURE

MESSAGE FROM THE DIRECTOR

Ms. Cynthia Bedell, SES

Director,
DEVCOM Army Research Laboratory
Computational & Information Sciences Directorate

In the last HPC Review, I wrote of all of the changes that had affected our Department of Defense (DOD) Supercomputing Resource Center (DSRC) since the inaugural edition. No one could have anticipated how much more change would occur between that edition and this one. One thing that has not changed, however, is our mission to operationalize science for transformational overmatch, and our quest to achieve enduring excellence at providing transformational overmatch as part of persistent Army modernization. The Army's ability to operationalize science is an enduring strategic competitive advantage against any adversary. Computational and information sciences are key to that advantage.

The year 2021 brought a reorganization within ARL to align competencies, more rapidly and efficiently bringing our technical capabilities to bear for Army Futures Command (AFC) challenges. One result is that my Directorate is now the home of scientists and engineers from the former Vehicle Technology Directorate, merging computational and information sciences with vehicle expertise and future autonomous systems. We are also positioned to engage our DSRC and Army Artificial Intelligence Innovation Institute (A2I2) in support of Project Convergence, the Army's new campaign of learning designed to aggressively advance and integrate our Army's contributions to the Joint Force. For the Army, the core challenges of the future fight are speed and scale.

The dramatic changes of 2020 were of course brought about as a result of COVID-19 and continued into 2021. Here we are, two years later, furthering our mission, in some respects as if nothing happened. Your response to the pandemic, adopting telework in stride and maintaining your pace and productivity, makes me proud, though I am not surprised. ARL really set an example Department wide in the way we managed our summer intern programs. While many other organization cancelled their summer programs, ARL adapted curriculum and processes in the moment and executed virtual summer intern programs that are the model for how this should be accomplished in the future. I commend all who made this successful, organizers, mentors, teachers, and students. To celebrate this success, we have included a special student section in this HPC Review highlighting developments from our 2020 and 2021 HPC Internship Program interns.

MESSAGE FROM THE CENTER DIRECTOR

Mr. Matt Goss

Director,
DEVCOM Army Research Laboratory
Department of Defense
Supercomputing Resource Center

Welcome back to the ARL DSRC. The Greek philosopher, Heraclitus, who lived around 500 B.C., is quoted as saying "change is the only constant in life" and we have ample recent evidence. COVID-19 changed our landscape in so many ways, socially, professionally and in terms of health. Yet here we are two years later, COVID-19 still prominent, yet we have adapted and done more than survive, we have thrived. The High Performance Computing (HPC) Modernization Program (HPCMP) continues transforming from a program focused on HPC to one focused on High End Computing (HEC). HEC combines HPC with high performance data analytics, embracing machine learning and other artificial intelligence tools, statistical analysis, and graph analysis. Users will be able to expand their workflows beyond traditional compute intensive jobs to embrace massive amounts of disparate data to further their science.

The ARL DSRC bid good bye to our previous workhorse, Excalibur, making room for a new generation of computational resources. Through the HPCMP Technology Insertion process for FY20, we received Jean, a Liqid Computing reconfigurable system with 57,696 Intel XEON Cascade Lake Advanced Performance compute cores and 280 NVIDIA Ampere A100 General-Purpose Graphics Processing Units (GPGPUs). It employs a 200 Gigabit per second InfiniBand interconnect supported by 323 terabytes of memory, and 12.5 petabytes of usable non-volatile memory express (NVMe)-based solid state storage. We received a companion Liqid platform, Kay, with 48,480 Intel XEON Cascade Lake Advanced Performance compute cores and 76 NVIDIA Ampere A100 GPGPUs, with a 200 Gigabit per second InfiniBand interconnect and supported by 240 terabytes of memory, and 10 petabytes of usable NVMe-based solid state storage. Due to acceptance testing issues with the new platforms, Centennial and Hellfire remain online.

The new supercomputers celebrate the remarkable achievements and enduring legacy of Jean Jennings Bartik and Kathleen 'Kay' McNulty Mauchly, computing pioneers who were part of the original team of ENIAC programmers. The systems join the ARL DSRC 'Betty' system, named In honor of Frances Elizabeth "Betty" (Snyder) Holberton, another of the original ENIAC programmers.

DoD's push toward Digital Engineering, the DevSecOps based acquisition model, and delivering capabilities to the field faster than ever before requires us to integrate tightly with the cloud. We must align our user interfaces, middleware and back end operations to what a user expects from a cloud experience. ARL serves as lead for a new Army Enterprise Cloud Management Agency initiative to integrate ARL computational resources with Cloud Army (cArmy) over the DREN. The new capability, called cArmy-X, will mirror capabilities over DREN for the S&T community to reflect what exists on cArmy for the operational community.

OUR VISION, OUR MISSION

Disruptive Science And Technology

The mission of ARL is to operationalize science for transformational overmatch. ARL is the Army's sole corporate research laboratory strategically placed in the Army Futures Command and focused on cutting-edge disruptive foundational research. ARL is shaping and informing the future operating environment, and is the primary link to the world-wide scientific community. In the last several years, as the Army has established modernization priorities, stood up the Army Futures Command and Combat Capabilities Development Command, focused efforts through Cross Functional Teams, emphasized Soldier touchpoints, and linked new concepts with science, ARL has been a key part of shaping thought and adapting as an organization and the leader of the Army's foundational research ecosystem.

We have expanded from our major sites to across the nation and are established in four distributed regions. We are executing priority research and cross-competency programs guided by our new campaign plan framework, which enables us to support AFC and DEVCOM, nesting our efforts within those of our higher headquarters, while emphasizing our unique value proposition for the Army.

The AFC Team Ignite is a collaboration between the Futures and Concepts Center and DEVCOM. It has expanded from early events designed to overcome the language barrier between scientists and operational concept writers, to establishing four new offices to evolve Ignite from a Task Force to an enduring Team. Team Ignite is one part of AFC's persistent modernization of the Army to provide future warfighters with the concepts, capabilities, and organizational designs they need to dominate a future battlefield.

The Army Research Laboratory is also investing in the future workforce, for example, through the Army Educational Outreach Program (AEOP) and the HPCMP Internship Program (HIP). The AEOP offers our nation's youth and teachers a portfolio of opportunities that effectively engage future workforce generations in meaningful, real world Science, Technology, Engineering and Mathematics experiences, competitions and paid internships. The HIP provides experience to prospective DoD employees in defense related research and development and cultivates the next-generation workforce in an active and collaborative effort among the HIP Mentors/Supervisors, HIP Interns, and the HIP Program Manager. The demand for qualified scientists and engineers, and indeed the full array of skills required to execute the mission of the Army Research Laboratory, is greater than ever and the Army is actively pursuing innovative approaches for attracting and retaining the future workforce.

New Discoveries, Cost Cutting Innovation

The Army Research Laboratory's driving purpose is to create and exploit scientific knowledge for transformational overmatch. It is part of the U.S. Army Combat Capabilities Development Command (CCDC) team, which is postured to deliver scientific, technological, engineering, and analytical solutions across the near-, mid- and far-term time horizons as a part of the Army Futures Command and in support of the Army's Modernization Strategy and beyond. In an era of rapid change and global access to science and technology, talent, and resources marked by rising strategic threats, quite simply the nation that can best and most timely operationalize science for military competition and combat will have a game-changing competitive advantage.

ARL has itself been transitioning through a multi-year inflection point building the future Army while vectoring us boldly forward for the first half of the 21st century. ARL has developed a campaign plan framework, nesting our efforts within those of our higher headquarters, while emphasizing our unique value proposition for the Army.

This plan encompasses four main lines of efforts focused on our science, our people, our engagements, and our operations. The campaign plan guides us in achieving the ARL mission to operationalize science for transformational overmatch, as part of the CCDC mission to provide the research, development, lifecycle engineering, and analytical expertise to deliver capabilities that enable the Army to deter and, when necessary, decisively defeat any adversary now and in the future.

WHY ARMY RESEARCH?

FACILITIES

Army Research Laboratory DOD Supercomputing Resource Center

Army Research Laboratory DOD Supercomputing Resource Center The ARL DSRC is one of five HPC centers provided by the HPCMP and hosts unclassified and classified computing platforms and storage. Combined with HPC software, secure broadband networks, and subject matter expertise, it is a powerful tool for research, discovery, innovation, problem solving, and creation and sustainment of future weapon systems. The confluence of basic and applied research expertise and facilities is the key to unlocking basic physical phenomena and harnessing the potential for defeating future threats and protecting U.S. personnel and property.

The ARL DSRC also offers services beyond those provided by the HPCMP, or extends them. We host customer HPCs, providing our expertise in systems administration along with available space, power, cooling, and back up power. We provide expertise in machine learning and customize ML tools for customer-specific needs. We will help containerize applications and build applications in a Persistent Services Framework. We can help create user interfaces and dashboards to simplify user interaction with applications, especially for data analysis. With the HPCMP missions space expanded to include machine learning, digital engineering, and acquisition engineering, the demand for support has increased beyond what the Program itself can provide in a timely manner. The ARL DSRC is poised to meet its users' needs.

Defense Research and Engineering Network (DREN)

The Defense Research and Engineering Network (DREN), and its classified counterpart, the SECRET DREN (SDREN), are the HPCMP's broadband networks supporting the S&T, Test and Evaluation, and Acquisition Engineering communities. The networks are high speed, high availability, high capacity, low latency, and nationwide providing access to the DSRCs by DOD, industry and university partners.

Root Name Server

ARL hosts one of 13 global root name servers, a critical element of the internet infrastructure. The ARL H root server is one of only three operated by the Federal government, the other two hosted by the Defense Information Systems Agency and NASA Ames Research Center, respectively. All others are hosted by industry or universities. Root name servers are the first stop in translating (resolving) human readable host names into machine readable internet protocol (IP) addresses used in communications among hosts and devices. The ARL root name server legacy dates back to the original ARPANET.

In 2019 ARL added two nodes to its Domain Name System (DNS) root server architecture, reducing the average global DNS query response time for the server from 122ms to 68ms. Additional nodes were added in 2020 bringing the total to 8: Frankfurt, Germany, Hong Kong, China, Johannesburg, South Africa, Sao Paulo, Brazil, Dubai, UAE, and Sydney, Australia, in addition to the original two nodes located in Aberdeen Proving Ground, MD and San Diego, CA. The average global DNS query response time is currently 45.54ms.

COMPUTATIONAL SYSTEMS

BETTY

Cray CS500
102,400 2.9 GHz AMD (EPYC) Rome Cores
446 TB System Memory
350 TB NVMe Solid State Drive
15 PB RAID Storage
GPUs: 292 NVIDIA Volta V100
Interconnect: HDR InfiniBand
Operating System: Red Hat Enterprise Linux

JEAN

Liqid Computing
8.8 petaFLOPS Theoretical Peak Capacity
55,392 Intel XEON Cascade Lake Advanced Performance compute cores
308 TB of memory
12.5 PB NVMe Solid State Disk
GPUs: 294 NVIDIA A100 Ampere
Interconnect: 200 gigabit per second Mellanox InfiniBand
Operating System: CentOS 8

GPUs (Graphics Processing Units)
NVMe (Non-Volatile Memory express)

SCOUT

IBM Power9
22 Training Nodes
- *132 NVIDIA V100 GPUs*
- *11 TB Memory*
- *330 TB Solid State Storage*

128 Inference Nodes
- *512 NVIDIA T4 GPUs*
- *32 TB Memory*
- *512 TB Solid State Storage*

Interconnect: EDR InfiniBand
Operating System:
Red Hat Enterprise Linux

KAY

Liqid Computing
4.9 petaFLOPS Theoretical Peak Capacity
46,176 Intel XEON Cascade Lake Advanced Performance compute cores
215 TB of memory
10.3 PB NVMe Solid State Disk
GPUs: 86 NVIDIA A100 Ampere
Interconnect: 200 gigabit per second Mellanox InfiniBand
Operating System: CentOS 8

SERVICES

DOD High Performance Computing Modernization Program (HPCMP)

The HPCMP celebrated its 30th birthday on December 5, 2021. It was created when President George H. W. Bush signed into law the National Defense Authorization Act for Fiscal Years 1992 and 1993. It was established under the Office of the Director, Defense Research and Engineering and in 2011 was transferred to the U. S. Army, which assigned it to the Engineer Research and Development Center in Vicksburg, Mississippi.

After 30 years, the Program still executes its mission to modernize the supercomputer capability of Department of Defense laboratories and test centers with large-scale computing resources; high bandwidth, low latency networks; commercial and government owned software and open source tools; and resident subject matter expertise and training. The Program was born amid rapid technological change internationally, which we continue to witness today. It has adapted to changing requirements, responded to national emergencies, and enables current and future generations of DOD scientists and engineers.

Originally focused on support for science and technology requirements, the HPCMP expanded to support DOD test and evaluation, and more recently, the acquisition engineering community. The ever-growing set of capabilities and services include digital engineering, virtual prototyping, virtual testing, large-scale data analytics, artificial intelligence and machine learning, and use of off premise resources such as government and commercial clouds.

Persistent Services Framework

The Persistent Services Framework (PSF) at the ARL DSRC provides functionality to bridge gaps users experience as part of their HPC workflows. Traditional HPC jobs "tear down" services and connections established during execution. Persistent services allow continually accessible applications, such as user-owned databases, a container image library, and ML toolboxes.

The initial deployment of the PSF capability on Centennial supports a limited subset of current users. The capability is targeted for those interested in incorporating data science, for example, into their existing simulations, and new users interested in leveraging the compute power offered by HPC systems to perform large-scale analysis and machine learning on their existing data sets.

Both ARL DSRC TI-20 systems, Jean and Kay, will be equipped with dedicated PSF resources (20 compute nodes with 96 cores per node, and 4 storage nodes). PSF will be available as a limited-access capability, with access and resources allocated to HPC projects/applications having bona fide PSF requirements.

Data Science Team

As the HPCMP purview has expanded so has the range of needs presented by DOD scientists and engineers. We see the biggest growth in data science, which includes ingesting and working with very large, diverse data sets (classified and unclassified), applying statistical methods to the data; training, testing and applying ML tools; visualizing in conventional and new ways; developing tools for integrating multiple types and formats of data, and more. The ARL DSRC Data Science Team reaches out to existing customers and identifies new ones, presenting skills in science, engineering, computer science, and mathematics, and offering creative approaches for solving data driven problems. Our Army civilian and contractor team assists customers in transitioning to HPC platforms, brokers partnerships between organizations, develops workflows and applications, and integrates with PET.

WHAT IS HIGH-PERFORMANCE COMPUTING?

The following pages contain HPC Success Stories from across DOD Services and Agencies. They represent work done by ARL researchers at any of the HPCMP DSRCs, and work done by all users at the ARL DSRC.

High performance computing, or supercomputing, is dynamic to say the least. From early vector and later massively parallel computers employing custom processors and running proprietary operating systems, to clusters of computers containing commodity processors and running an open source operating system, Linux. The drivers were mainly physics-based models, requiring a large number of the fastest possible processors and as much memory as was possible. Those requirements still exist but new ones have arisen, such as machine learning with a demand for GPUs, at least for training algorithms, and data science, working with possibly millions of files and extensive data movement. HPC architectures for the traditional and the new look very different. Other changes have also arisen, such as HPC for mobile or edge computing, and access to computational resources at above secret classifications.

The HPCMP uses its technology insertion process to capture ever changing DOD requirements and implement DSRC capabilities to meet those requirements. The HPCMP continually expands DREN connectivity between users and the DSRCs, provides subject matter expertise through its User Productivity Enhancement and Training (PET) initiative, and maintains high capacity repositories for user data, backup storage, and disaster recovery.

Streaming Detection and Classification Performance on SCOUT

Wesley Brewer
DoD High Performance Computing Modernization Program
User Productivity Enhancement & Training (PET)
General Dynamics Information Technology

Project Description

Real-time object detection may be used to automatically detect and classify specific targets in motion imagery (MI). High-performance computing systems such as SCOUT provide computational support at the tactical edge for training object detection and classification systems and inferencing on streaming MI. One such example would be to provide analytics support for a fleet of drones streaming MI using the Video Processing Exploitation Framework (VPEF). Another application is providing support for training classification systems such as the Ordnance Threat Target Automated Recognition (OTTAR) system for classification of ordnances on the battlefield. In this work, we port, benchmark, and optimize VPEF and OTTAR to run on SCOUT, a GPU-dense edge supercomputer. We use Horovod, an MPI-based framework, to distribute the training across 48 GPUs at 91% parallel scaling efficiency. We also observe a 10x speedup by inferencing at reduced precision via TensorRT. We find that V100s work well for training and offline inferencing in batches, while T4 GPUs outperform V100 GPUs only in specific usage scenarios, such as streaming detection at reduced precision.

Relevance of Work to DOD

SCOUT is a mobile, containerized supercomputer that can be deployed anywhere in the world. With 128 inference nodes and 22 training nodes, SCOUT was designed to support both massively parallel inferencing and parallel training of deep neural networks. While deployment of production-level AI/ML systems to the cloud is more prevalent, deployment of the same to supercomputers, especially POWER9 architecture, has been less well understood. In this work, we provide information for how to optimally deploy streaming object detection and classification models to a mobile containerized supercomputer. Ultimately the impact of such a study is to provide better support to analysts on the battlefield making critical decisions in real time. The results of this study are provided in a technical report and a publication with specific recommendations on configuration and settings to achieve optimal performance on SCOUT.

Computational Approach

For training neural networks, we used the TensorFlow Model Garden object detection library to train Faster R-CNN using data from the Wide Area Surveillance Activity Based Intelligence (WASABI) dataset. To speed up training, we used Horovod over Nvidia's Collective Communications Library (NCCL). To optimize and reduce the precision of the models, we used TensorRT within TensorFlow. For streaming object detection, we created a GStreamer pipeline that uses VPEF components to stream images to the GPU for inference. The pipeline outputs timestamps just before and after the inference, which are post-processed to compute an average frames per second (FPS) metric.

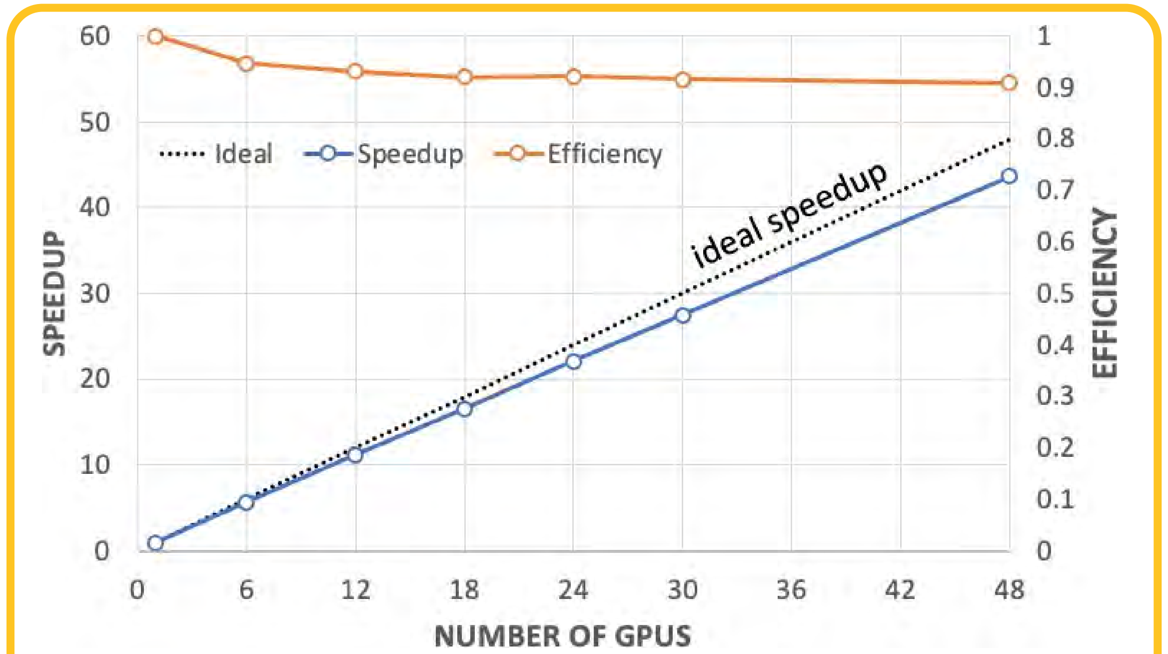
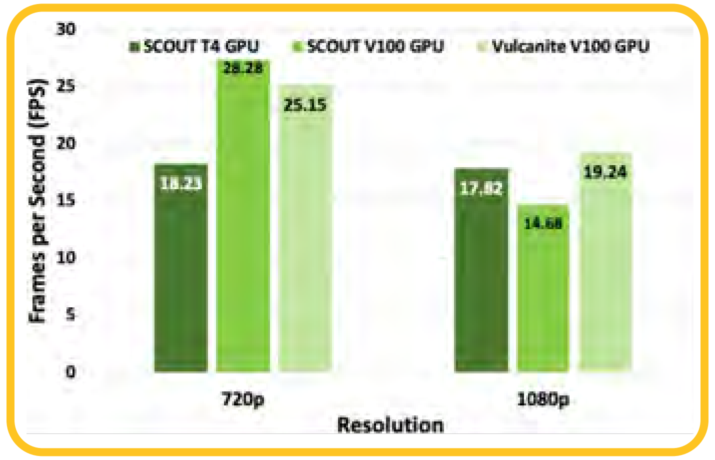
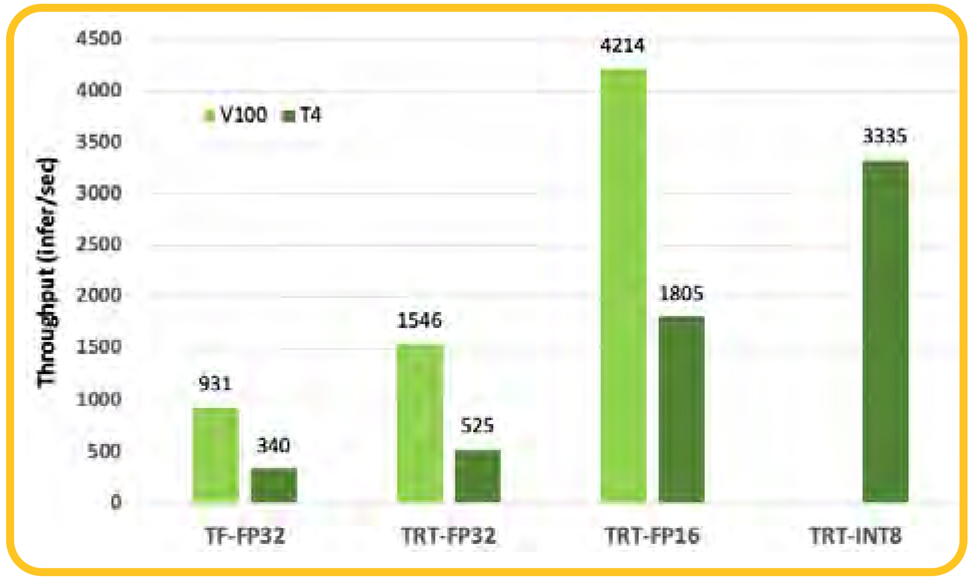


Figure 1:
Speedup and parallel scaling efficiency training ResNet-50 with Horovod on SCOUT.



(Left) Figure 2:
Profiling of real MI using AVAA's VPEF-DNN on SCOUT with SSD detector.



(Below) Figure 3:
Inference benchmark performance of ResNet-50 on V100 and T4 GPUs at multiple precisions on SCOUT
(Note: V100 does not support INT8 inferencing).

Results

In Fig. 1 we show the speedup obtained on SCOUT when training ResNet-50 with Horovod up to 48 GPUs. We achieved good quasilinear speedup with 91% parallel scaling efficiency at 48 GPUs. We also investigated streaming object detection performance using VPEF with SSD-MobiletNetV2 object detection model trained on COCO images to detect pedestrians, cars, and bicycles as shown in Fig. 2. The T4 GPU gives similar inference performance as the V100 GPU but has half the number of CUDA and Tensor cores and is four times more power efficient. Finally, we investigated inferencing at reduced precision via TensorRT. In Fig. 3 we show the results of a TensorFlow ResNet-50 model versus TensorRT converted models at three different precisions: FP32, FP16, and INT8. This study shows that by using reduced precision we are able to achieve a 10x speedup in inference performance from TF-FP32 to TRT-INT8.

Future

In the future, we would like to investigate reduced-precision streaming using Nvidia's DeepStream SDK framework and modify VPEF to support TensorRT models. DeepStream is not open source and Nvidia does not provide ppc64le binaries, making it impossible to test on SCOUT. However, we have performed initial promising tests using DeepStream on Vulcanite HPC, an x86_64 HPC with V100 GPUs, showing a speedup of about 6x at FP16.

Co-Investigators

Chris Geyer *(Parsons Corporation)*
Dardo Kleiner *(CIPS, Corporation)*
Connor Horne *(Naval Research Laboratory)*
William "Rich" Thissell *(Deftec Corporation)*

Publications

W. Brewer, C. Geyer, D. Kleiner, C. Horne, "Streaming detection and classification performance of a POWER9 edge supercomputer", IEEE High Performance Extreme Computing, 2021.

Reduced-Order Modeling of Aeroelastic Resonances in Extreme-Speed Turbochargers

Daniel J. Bodony & David W. Fellows
Department of Aerospace Engineering,
University of Illinois at Urbana Champaign

Project Description

Small, highly versatile aircraft are becoming increasingly important to provide increased situational awareness, reduced workloads, and reduced risk to warfighters. These aircraft are often powered by intermittent combustion engines that utilize turbochargers to achieve the pressure ratios needed to sufficiently power the aircraft; however, turbocharger manufacturers have not yet broadly embraced the aviation industry. Consequently, the engines utilize turbochargers designed for ground applications which operate at off-design conditions during flight. Over prolonged exposure to these conditions, the turbocharger's rotors are susceptible to fluid-structural interactions (FSI) that induce high-cycle fatigue of the blade tips. Identifying the specific operating conditions at which FSI phenomena arise is important to avoid the onset of FSI-induced high-cycle fatigue or optimize turbomachinery designs to eliminate the FSI altogether. This work seeks to develop a reduced-order model of the fluid-structural dynamics to aid in the quick identification of FSI onset conditions.

Relevance of Work to DOD

This work supports Technical Area 3.1 (Electried Lightweight Turbomachinery) of ARL's Versatile Tactical Power and Propulsion Essential Research Program (VICTOR ERP). The objectives of this technology area include, in part, identifying, understanding, and modeling contributors to the high-cycle fatigue of turbomachinery blades. These objectives are currently being achieved via experimental investigations, which are expected to provide novel insight of the aeroelastic behavior of turbomachinery in the extreme-speed regime but are ill-suited to rapidly assess a particular machine's aeroelastic stability. Numerical methods are a viable alternative, but the existing approaches require coupled fluid and structural domains, requiring significant computational power and execution time. The reduced-order model developed in this work aims to provide a general numerical methodology utilizing uncoupled simulation methods to substantially mitigate the required computational expense needed to obtain an aeroelastic stability prediction.

Computational Approach

The reduced-order model methodology is displayed in Figure 1. The work presented here derives a reduced-order model for a turbocharger's turbine wheel. ANSYS Mechanical Enterprise was used to solve the turbine wheel's free vibration problem utilizing a finite element method. The solutions consist of a series of

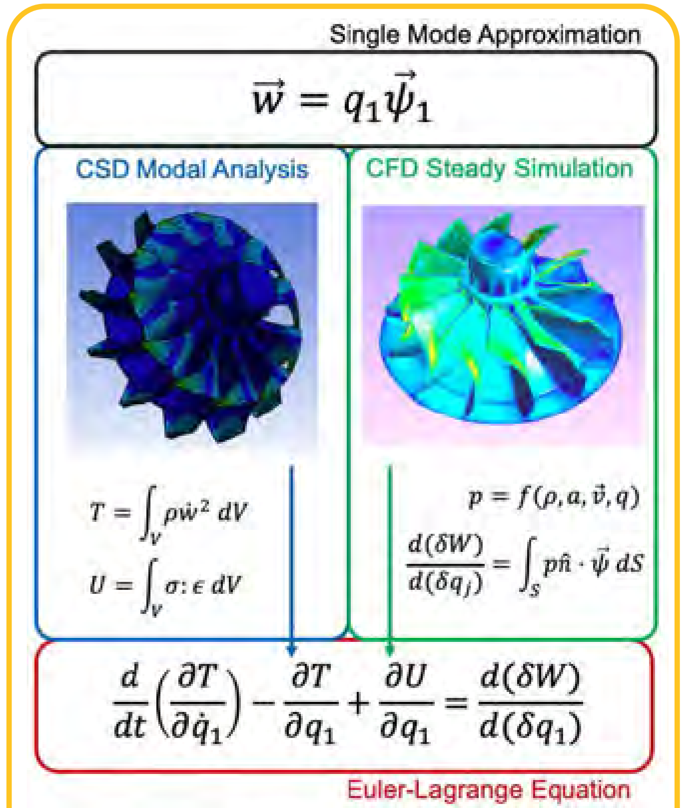


Figure 1:
Single mode approximation (top) prescribes structural response using a shape function and a generalized coordinate. The shape function is obtained using CSD techniques (left), and a governing equation for the generalized coordinate is derived from the Euler-Lagrange equation (bottom). The response of the fluid to the structural deformation is approximated using a CFD-informed simplied aerodynamic model (right) which computes the unsteady pressure loads that act on the structure.

fundamental vibratory mode shapes and corresponding excitation frequencies. A single mode shape is chosen to approximate the structural response, from which the deforming structure's kinetic and potential energy may be computed.

The response of the surrounding fluid to the deforming structure is modeled utilizing a CFD-informed local piston theory that uses the mean steady flow field to approximate the unsteady pressure loads acting on the turbine blades. CFD simulations are being performed on HPCMP machines, Mustang and Centennial, using an element-based finite volume formulation present in ANSYS CFX. The 3D Reynolds-Averaged Navier-Stokes (RANS) equations are solved within the turbine domain, with the exception of the turbine blade subdomain where the 3D RANS equations are solved with additional centripetal and Coriolis source terms to model the effect of the rotating turbine blades without physically moving the blades within the domain. The boundary conditions were informed from experimental measurements. All solid domain boundaries were modeled as free-slip surfaces to acquire a pseudo-Euler solution on the turbine blade surfaces.

Results

To identify the mode shape most likely being excited during an aeroelastic event, CSD modal analyses were performed over a parametric sweep of turbospool rotational speeds and environmental temperatures. The computed excitation frequencies were then plotted as a function of turbospool rotational speed and the associated turbospool engine orders were overlaid on the plot. A non-dimensional version of this plot, known as a Campbell diagram, is displayed in Figure 2. The turbospool rotational speed has been non-dimensionalized by a rotational speed encountered at flight conditions, and the modal frequencies have been non-dimensionalized by the associated blade passing frequency. The intersection of an engine order and a modal frequency imply a potential resonance condition associated with harmonics of the turbospool rotational speed.

Experiments conducted at ARL's Small Engine Altitude Research Facility have identied two operating conditions where excessive vibrations of the turbocharger are observed. The more extreme of these vibration events is being considered as the resonance event of interest. Comparisons of the predicted resonance conditions to the experimental observations imply an axisymmetric mode shape (depicted in Figure 1) could be responsible for the high-cycle fatigue of the turbine blade tips, however, this hypothesis is still being evaluated. Additional comparisons against recently acquired data from non-intrusive stress measurement probes installed within the turbocharger's turbine are being conducted to aid this effort. The axisymmetric mode shape, in conjunction with the local piston theory, has been used to compute the terms present in the Euler-Lagrange equation (see Figure 1) and derive the equation governing the fluid-structural dynamics. The resulting equation is a nonlinear, second-order ordinary dierential equation for which linear stability analysis techniques may be used to understand the behavior of a perturbation about the unreformed structural configuration.
The tools required to inform the linear stability analysis are currently being developed and validated against existing aeroelastic studies of panel
flutter on comparatively simpler fluid-structural configurations.

Future

A series of methods to validate the reduced-order modeling methodology have been proposed. A high-fidelity, coupled aero-thermo-elastic simulation will be performed on the HPCMP machines to observe the fluid-structural behavior of the turbine wheel at operating conditions associated with FSI behavior. Additionally, ARL will use the existing engine cells and commission a new hot gas bench, known as the Versatile Turbomachinery Experimental Facility, to measure blade deformation in a variety of turbomachines at altitudes up to 30,000 feet. The reduced-order modeling methodology will be applied to each turbomachine and operating condition to evaluate the accuracy of the model.

Co-Investigators

Dr. Ryan C. McGowan *(U.S. Army Combat Capabilities Development Command Army Research Laboratory)*

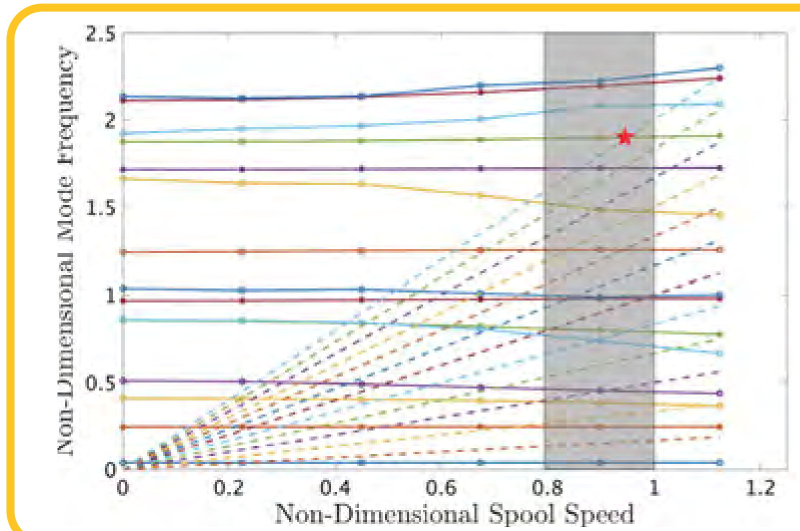


Figure 2:
Non-dimensional Campbell diagram of selected turbospool vibratory modes. Solid lines correspond to modal frequencies, while dashed lines correspond to turbospool engine orders. The gray box represents the range of turbospool rotational speeds that are encountered during flight. The red star corresponds to a potential resonance condition that is being investigated using the reduced-order model.

High Performance Computing Platforms for Rapid Optimization of Computationally Complex DNN Models for the Tactical Edge

Venkat Dasari
US Army Combat Capabilities Development Command Army Research Laboratory, Computational & Information Sciences Directorate

Project Description
Deep Neural Network (DNN) architectures are widely used for object detection, semantic scene labeling, and contextual learning by autonomous systems. Due to their ability to learn from data in solving complex problems, DNNs are finding a role at the tactical edge. However the DNNs are computationally complex, resource intensive, and often not suitable for deployment over resource-constrained edge computing platforms. For the DNNs to be deployable at the tactical edge, we optimize them to reduce their computational complexity and resource requirements for their execution on target platforms. The DNNs are analyzed, code execution is optimized, unnecessary features and weights are pruned, and the model precision is reduced from their default fp32 operation to fp18 or int8 in order to generate an optimized DNN model requiring fewer resources and yet no or minimal reduction in their prediction accuracy, Figure 1. A modified Taylor series approximation function is used to rank the features and weights, and low ranking features and weights are deleted from the model.

Relevance of Work to DOD
With the revolution of artificial intelligence and deep learning technologies, researchers have started implementing AI enabled intelligent mission applications at the tactical edge. AI stacks have been proposed for sensing, reasoning and computing situational awareness to provide human-in-the-loop actionable intelligence in mission execution. Many of these functions are highly reliant on high-performance computing and memory. However, Size, Weight and Power (SWaP) are critical to soldiers already burdened with large amounts of equipment. We have explored a wide variety of approaches to reduce SWaP so warfighters can focus more on the mission at hand and carry less equipment, spend less time on logistics, transportation and setup. Therefore, tactical edge computing platforms must employ small-form-factor modules for compute, storage, and networking functions that conform to strict SWaP constraints. In other words, tactical edge computing platforms have limited computation, communication, memory and battery life. While they can support mission applications, computationally complex DNN models may not be deployable without modifications at tactical edge platforms due to resource constraints. Thus, various optimization approaches are leveraged to reduce the computational complexity of a DNN model to significantly reduce the model size while preserving the prediction accuracy. This model could be directly deployed in multiple tactical edge computing platforms including NVidia Xavier, Movidius, and FPGA. The outcome of this project will be of tremendous value to the AIMM ERP in making their AI algorithms mission deployable

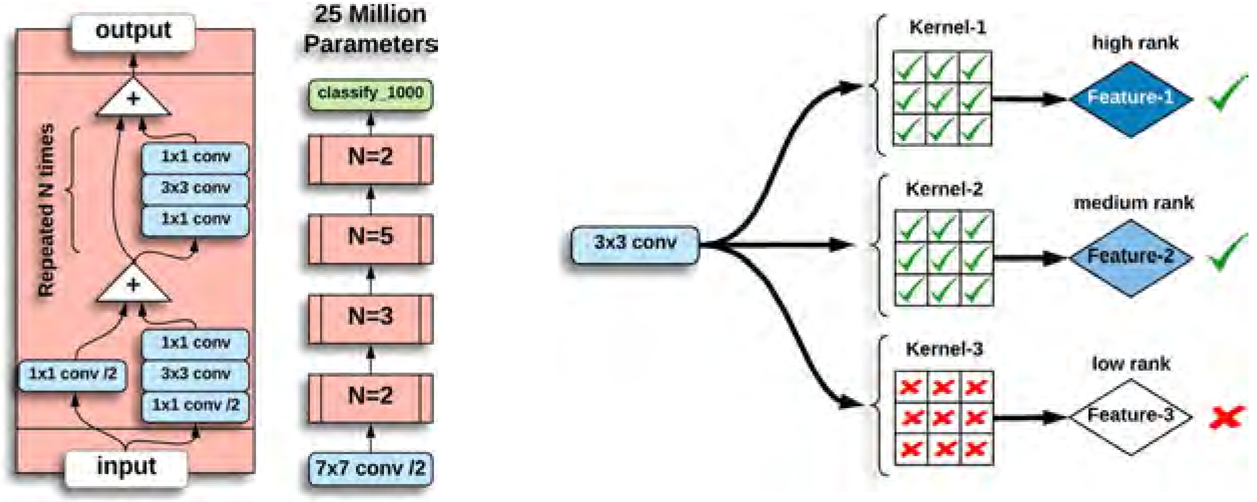


Figure 1:
The model on the left shows a typical DNN composed of feed-forward and residual branches, while figure on the right shows a convolution being ranked and pruned.

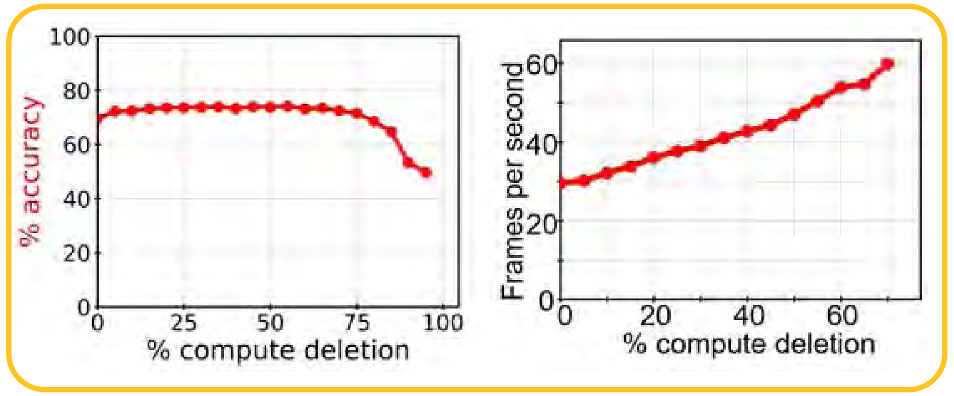


Figure 2: Experimental results from SCOUT.

Alg	Framework	GPU Mem	FPS
Base	PyTorch	1.56 GB	27
Fp32	TensorRT	2.05 GB	49
FP16	TensorRT	1.96 GB	118
Int8	TensorRT	1.75 GB	148

Table 1: ICNet quantization.

Computational Approach
Our project heavily depends upon GPU computing resources and also needs a reasonable number of CPUs to reprocess our training data. The DNN model optimization process is computationally intense and involves retraining of the model as we systematically delete low rank features. DSRC HPC computing platforms with GPU-dense nodes are ideal for optimization computations. Thus, we train and optimize the DNN models on SCOUT and deploy them on resource constrained edge computing platforms. The essential software for our research is Anaconda; Keras and PyTorch are open-source frameworks for deep learning. The training data set is ImageNet, a large dataset of over 1.4 million images that occupy 600GB of disk space. The state-of-the-art DL models, e.g., Resnet, Mobilenet, Squeezenet etc., may need to be tailored, optimized and retrained for resource-limited computing platforms.

Results
Figure 2 describes the relationship between the compute deletion and model accuracy and its performance in terms of frames per second. On the left hand side of the figure, we can delete up to 60% of computations in our target model using various pruning techniques and still maintained the pruned model accuracy close to its original model. Left hand side of the graph demonstrate a linear relationship between the compute deletion and increase in inference acceleration in terms of frame per second. Table 1 shows the inference acceleration of ICNet as the model precision is reduced from fp32 to fp16 and int8.

Future
We have converted ICNet from Pytorch to Tensor-RT and used the available tools to quantize the model from fp32 down to fp16 and int8. As can be seen in Table 1, quantizing the model showed a drastic increase in performance. We are currently analyzing the stability of the quantized models in comparison to the original model as well as using SCOUT to re-train the original model using quantization-aware loss functions to minimize the accuracy loss when quantizing down to int8.

Co-Investigators
Billy Geerhart *(Army Research Laboratory Computational & Informational Sciences Directorate and Parsons)*
David Alexander and Peng Wang *(Army Research Laboratory Computational & Informational Sciences Directorate)*

Publications
V.R. Dasari, B.E Geerhart, P. Wang, D.M. Alexander (2021) Deep neural network model optimizations for resource constrained tactical edge computing platforms. Disruptive Technologies in Information Sciences V. International Society for Optics and Photonics. Volume 11751 , Pages:117510E

Geerhart III, V.R. Dasari, P. Wang, and D. Alexander (2021) Efficient normalization techniques to optimize AI models for deployment in tactical edge. Disruptive Technologies in Information Sciences V. International Society for Optics and Photonics. Volume 11751 , Pages:117510F.

Army Vulnerability and Lethality Analysis Bolstered Through High-Performance Computing

Kaylan Hutchison
Booz Allen Hamilton (Contractor)
at the U.S. Army Combat Capabilities Development Command Analysis Center

Project Description

The ability to process massive amounts of data is crucial as organizations and their technological innovations evolve. It is through analysis of data that the U.S. Army can continue to develop engineering breakthroughs, modernize technology and make informed decisions. However, standard computing workstations are not always the most effective agent for the challenge of processing immense amounts of data and operations within a reasonable timeframe. That's why the U.S. Army Combat Capabilities Development Command Data Analysis Center (DAC) leverages High Performance Computing (HPC) to perform faster and more complex analysis on tri-service survivability, vulnerability and lethality modeling.

Relevance of Work to DOD

Computational capability plays a major role in analysis. HPC is a computing technology that harnesses the power of large-scale computer clusters to process complex calculations and perform data-intensive tasks at high speeds. A high performance computer processes quadrillions of calculations per second, compared to the few billion calculations with a standard laptop. As DAC strives to be the best analytical workforce out there, it must take advantage of the best computer resources; DAC is leveraging HPC for the DOD's accredited Joint-Service approved Advanced Joint Effectiveness Model (AJEM). AJEM is managed by DAC and used for analyses to support a wide variety of acquisition efforts. Results produced from AJEM inform system design, incident recreation and analysis of alternatives to ensure systems used by warfighters are safe, lethal, functional and effective. AJEM also provides target vulnerability data to support operational planning by the Joint Technical Coordinating Group for Munitions Effectiveness (JTCG/ME). The JTCG/ME serves as the DOD's focal point for credible, standardized tri-service munitions effectiveness data and methodologies for operations and acquisition. Chartered over 50 years ago, agencies across the Joint community collaborate to develop, implement and leverage the work. The immense amount of data and analysis DAC produced for JTCG/ME has created demand for a system like HPC.

Computational Approach

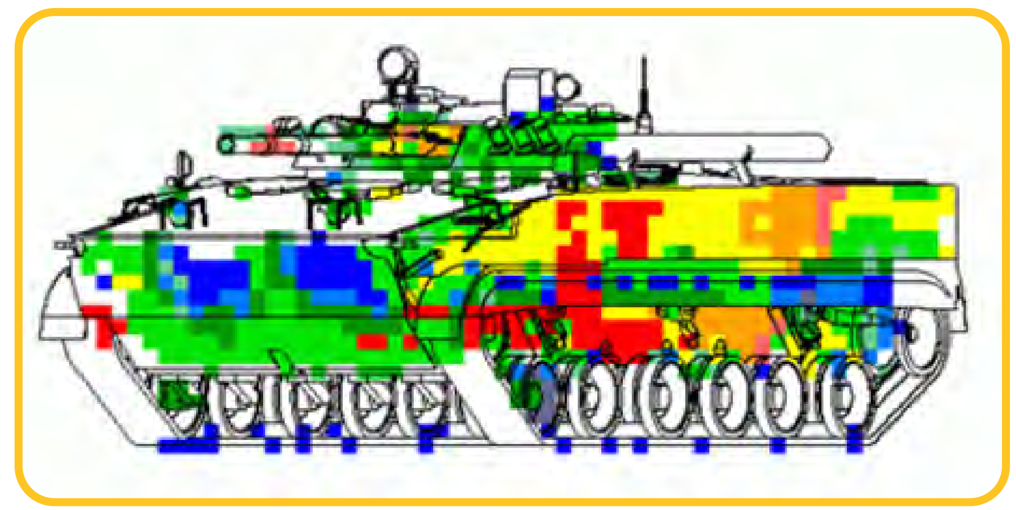
To guarantee a reliable and reputable model as part of the continuous software development process, AJEM undergoes quarterly methodology improvements and upgrades features that allow analysts to better process data. AJEM has kept pace with the changes in technology. As new ideas come through, or new methodology or materiel, AJEM adapts and accommodates to meet high-priority mission requirements. Those accommodations result in more detailed, complex AJEM models. Thus, to expedite run-time, DAC is partnering with the Army Research Laboratory to get more information to customers faster and more efficiently.

DAC can speed up analysis quite considerably with the performance power of HPC— not only with the architecture behind it but with the software and resources the HPC Modernization Program has. HPC clusters and nodes operate in a way that allows users to run their analyses through parallel processes in a fraction of the time. This enables multitasking without a particular process getting "stuck" and holding up subsequent processes. DAC is currently working on merging parallel processing into the current state of AJEM.

Results

At DAC, some analysts have a metaphor of a 50,000 acre farm to describe HPC's impact. A single small tractor may eventually get the harvest done, but a "mega-harvester" with automation and GPS designed for large jobs will do it faster and more effectively. DAC has been able to accomplish with HPC what would not have been possible using AJEM on a standard desktop workstation. The cluster-nodes infrastructure allows for hundreds of thousands of AJEM results in a week, versus a few hundred with a standard PC system.

Aside from the time-saving benefits for analysts and DOD decision-makers, HPC opens the scope of computationally-intensive AJEM analyses. However, it is not just the technology, but the people behind the technology that make this successful. DAC has an innovative workforce: the people who reach out for an answer, who see an opportunity and take advantage, who adapt to new challenges and see different perspectives. The benefit of leveraging HPC for AJEM will be far-reaching for JTCG/ME and the acquisition community.



Advanced Joint Effectiveness Model (AJEM) visualization of a system's probability of kill given a hit (Pk/h), the statistical probability that a target will be degraded or lose functionality when a threat hits the target.

Future

While DAC tests and transitions parallelized code to software users, DAC will also continue to redesign the software to further develop the capability and take advantage of HPC hardware. Additionally, DAC will leverage other initiatives sponsored by the HPC Modernization Program to train users and modernize codes.

According to the AJEM model manager, there has been a fast-growing number of AJEM requests from internal and external agencies, many of which leverage HPC to use DAC software. As DAC develops the link with AJEM and HPC, the number of stakeholders will grow. This not only brings high potential for a greater impact, but increases potential for most, if not all, AJEM simulations to leverage HPC computing resources as the new standard.

Co-Investigators

L. Ashley Bomboy, Douglas B. Howle, Jr., Roberto Gonzalez, Jr., and Timothy D. Mallory.
(U.S. Army Combat Capabilities Development Command Data Analysis Center)

Publications

J. Michael Barton, Douglas B. Howle, Jr., Craig S. Barker and Andrew W. Drysdale, "Ground Vehicle and Occupant Vulnerability Assessment Using High Performance Computing," The ITEA Journal, 40 (4), December 2019, 260-270

Integrating Analysis and Visualization for Army Test Data

Vincent Perry,
US Army Combat Capabilities Development Command
Army Research Laboratory,
Computational & Information Sciences Directorate
DOD Supercomputing Resource Center

Project Description

In the past few years, the ARL DSRC began to shift focus from traditional scientific visualization and simulation to data-focused test and evaluation analysis support. Visual analytics has become an increasingly important capability for users to get a quick first interpretation and interaction ability with their data. The goal of this project is to develop a visual analytics workbench using open-source tools that enables advanced data analysis capability for the T&E community, leveraging HPC resources for data processing and modeling, with a visual front-end for analysts to interact and interface directly with their data.

Relevance of Work to DOD

The DOD is investigating the use of autonomous vehicles for resupply and reconnaissance efforts. With the large amount of test data being collected, this manual, labor-intensive process of meaningful discovery succeeds only with precise queries, suffering in timeliness, broadness, and lack of scalability. The current work directly supports Army T&E analysts to expedite evaluation workflows, as well as the Army's Next Generation Combat Vehicle modernization priority. Our visual analytics workbench provides a scalable system and interface for timely harvesting of important information from vast and often disparate data sets.

Computational Approach

We utilized two sources of data: structured instrumentation data stored in a PostgreSQL database and unstructured written reports. We created a Singularity container for PostgreSQL to run the database in the HPC batch processing environment. We then scripted an analysis workflow that dispatches the database to compute nodes while also launching and connecting RStudio or Jupyter notebook to the database. The data from the written reports were processed into csv files; they are time-stamped and can be synchronized with the instrumentation data. Next we developed targeted analysis capabilities that can be automated to run on HPC for the end-users. We applied a ranking algorithm retrieval engine that ranks the written reports by similarity score to a user-input query. We performed natural language processing of the written reports which led to experimenting with and training clustering and classification models to further expedite an analyst's evaluation of the reports. We then explored anomaly detection and feature importance models for the instrumentation data.

Our next step was to build the visual analytics workbench. We limited the scope to a single day of data collection, specifically the day that had the most written reports available with which to coordinate. We developed a front-end dashboard application using Python's Dash that allows users to visually query and interact with large amounts of data. The application was organized with two tabs – one for a data overview and another for micro analysis. Checklists, radio items, drop downs, and buttons were all used as inputs inside the application, while the graphs were the output. The final step was scripting ways to automate launching the different components. This includes the back-end databases, the behind-the-scenes analysis models, and the front-end UI. To facilitate the usability of interacting with this application, we developed an iLauncher plugin that automates job launching.

Results

The data analysis workflow allowed the analysts to process almost 20 TB of test data spanning 134 days of testing in only 10 days on Centennial. This not only reduced processing times, it enabled analysts to generate enhanced performance statistics, correlate disparate heterogeneous data sets and ultimately determine causality of previously unexplained system failures.

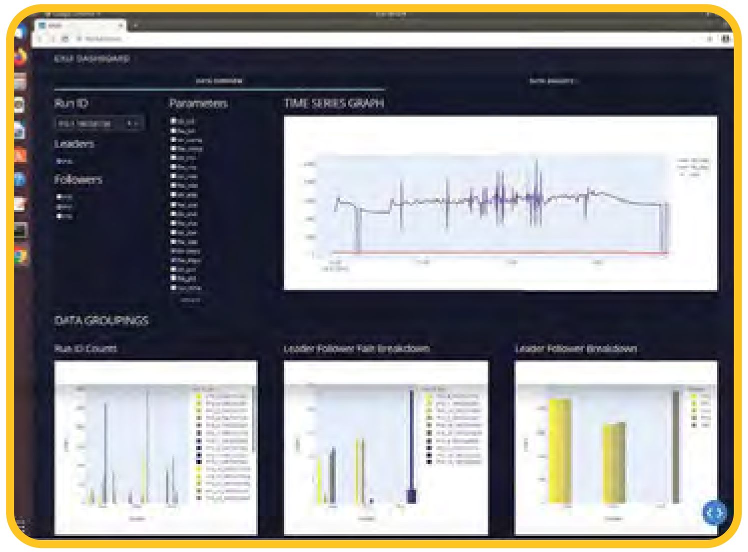


Figure 1:
The "Data Overview" Tab of the Dashboard Upon Launch

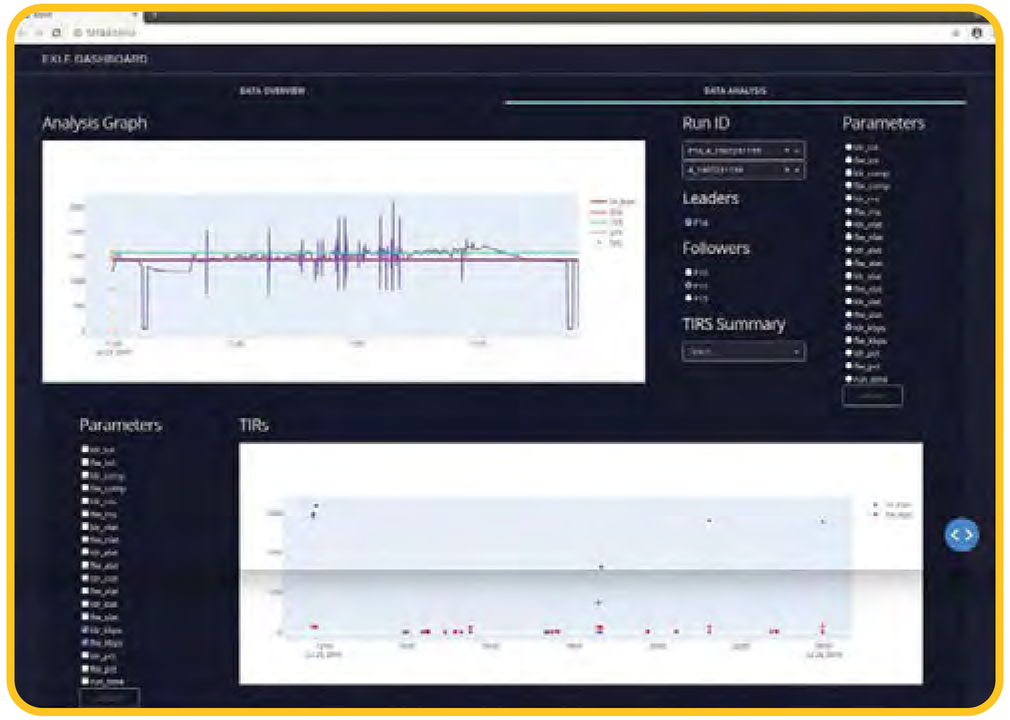


Figure 2:
The "Data Analysis" Tab of the Dashboard

The final visual analytic capability result includes a user-facing visualization dashboard that interfaces with two separate back-end data sources and integrates R analysis code into the Python application and data management code. The dashboard includes two tabs: a "Data Overview" tab and a "Data Analysis" tab, Figure 1. While data is pre-loaded to provide an overview, the user may begin interacting with the application by selecting the particular test trial from the drop-down in the top left. The top visualization plots each of the selected features over time, allowing the user to compare the data for several features at once. The three bottom visualizations provide summary statistics about the selected data identifiers.

In the second "Data Analysis" tab, Figure 2, users can select trials, identifiers, and parameters of the data set. On top, parameter values are plotted over time along with summary statistics depicting the mean, first quartile, median, and third quartile as horizontal lines. The written reports recorded during this time are also marked on this visualization. The second plot shows a timeline with markers depicting where the written reports exist in time. In all, the user can detect anomalies, recognize patterns, correlate data across visualization types, and further inspect the data without performing command line queries.

Future

In the future, we would like to expand the capability to include a Spark or Dask processing layer to distribute the data processing state before delivering content to the front-facing application. To perform statistical analysis across large subsets of the data, we would like to incorporate packages such as Parallel or Rmpi to distribute the R analysis jobs across multiple compute cores and nodes on the HPC systems. We would also like to incorporate more machine learning models into the analysis section of the pipeline. Using the instrumentation data, we plan to apply logistic regression to develop baseline models of the system performance over time to further identify trends and anomalies in the data. We plan to leverage SCOUT for training our machine learning models, experiment with running this pipeline within the Persistent Services Framework, and containerize all the different services that run together for this analytics workflow.

Co-Investigators

Dale Shires, *(U.S. Army Combat Capabilities Development Command Army Research Laboratory)*
Mariya Occorso, *(Parsons)*
Aiden Kenny, *(Columbia University)*
Jacob LeBlanc, *(Stanford University)*
Dan Owens, *(Army Test & Evaluation Command Army Evaluation Center)*

Publications

Vincent Perry, Mariya Occorso, Aiden Kenny, Dale Shires. Integrating Analysis and Visualization for Heterogeneous T&E Data. The ITEA Journal of Test and Evaluation 2021; 42: 85-92.

Aiden Kenny, Mariya Occorso, Vincent Perry. Visual Analytic Application Toolset for Large-Scale Data. Workforce of the Future: Student Article. The ITEA Journal 2021; 42: 72-76.

Realtime Key-Point Detections For Missiles in Flight Using A Deep Learning Framework

Gedion Teklemariam
Computational Scientist
DoD High Performance Computing Modernization Program
User Productivity Enhancement & Training (PET)
General Dynamics Information Technology
Naval Research Laboratory - DC

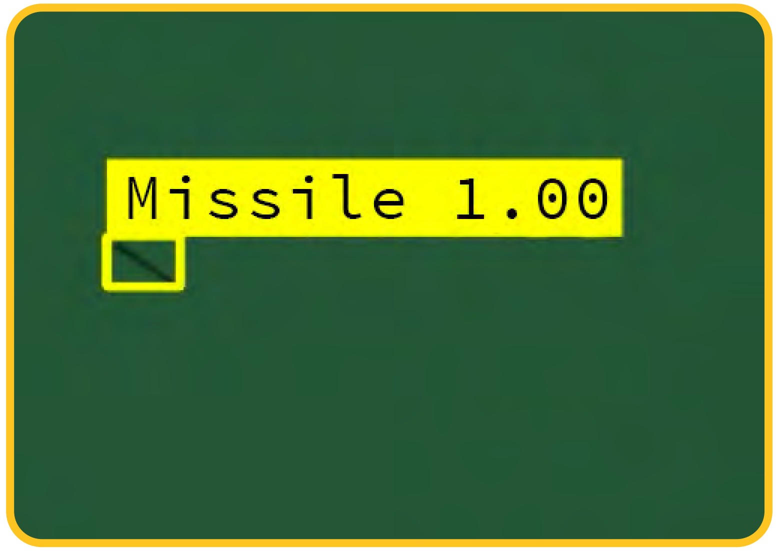


Figure 1: YOLOv3 Model detecting missile captured from distance of about 2 miles.

Project Description

The objective of the project is to produce a key-point detection model for missiles in-flight. The detection model will have the capability to automatically determine location of the flying missiles within the camera frame, as well as producing the tip and tail locations for the detected missiles. The White Sands Missile Range (WSMR) team, ARL, and the USAF AFMC AEDC/TST team have intentions to utilize the output of this project to improve efficiency on their current processes for detection of missiles in-flight and the ability to dynamically perform analysis in a live test setting, as well as post-test analysis.

At WSMR, the current process for determining the actual tip and tail of the missiles is a manual process, which is very time consuming and costly. There can be multiple cameras capturing an engagement. For the WSMR team, in the future, engagements that include multiple targets and multiple missiles are likely, which would increase the complexity of terminal engagements. Automating the detection and labeling of missile tip and tails, as well as estimation of the object's orientation in the image frames, would enable the WSMR team to be free of the need to manually or semi-manually label objects.

Relevance of Work to DOD

The immediate impact for the DOD will be to reduce the high cost of labor for manual-labeling of the tip and tails of missiles. Some of the cameras used at testing sites can produce up to 1000 frames per second and having a manual process to go through each image frame costs the DOD tremendously in time and money.

Furthermore, the AEDC/TST team has a national defense-related mission need in the area of Missile Warning. Specifically, detection and defeat of surface to air missiles. For the AEDC/TST team, a key measurement currently unavailable to missile warning and countermeasure system designers is the attitude of missiles in flight. The model output from this project can be utilized to improve efficiency at performing visual based detections for missile warnings and provide key inputs to perform dynamic missile attitude analysis in a live setting.

Computational Approach

As a primary option, we use the Detectron2 deep learning framework, which is developed by Facebook Research. Detectron2 is based on MaskRCNN-benchmark and built in PyTorch. With Detectron2, many levels of image detections are possible including key-point detection. Detectron2 is flexible and allows for customizations. Depending on identifiable needs in the future, the model can be extended to do several levels of segmentations including panoptic segmentation (includes background segmentation). In addition to the Detetron2 framework, for general detection steps, the YOLOv3 and Mask-RCNN frameworks were also considered.

From the WSMR and the AEDC/TST teams, to prepare model inputs, we acquired video files as well as synthetic missile images. However, the resolutions of the inputs are quite limited, and to adapt to resolution related issues, the development of dynamic image enhancement solutions will be considered. For data ingestion steps and post-processing steps, which accommodate analysis related needs, pipelines are developed to automate the transformational steps. As a prerequisite for the key-point detection model, the development of a custom pose estimator for missiles is required, which will need to be adapted into the Detectron2 framework to accurately detect tip and tail of the missiles in-flight. The HPC systems being used for trainings are Vulcanite, SCOUT, and Gaffney.

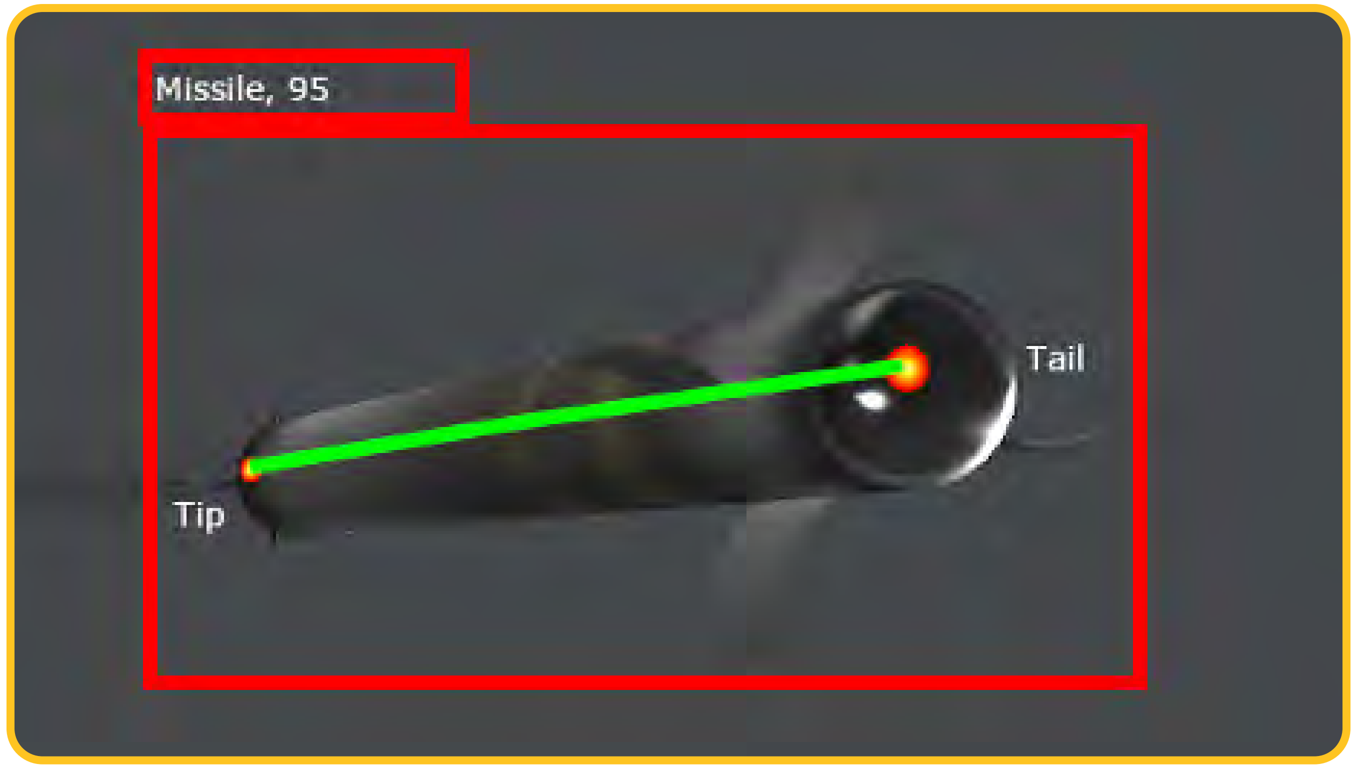


Figure 2: Conceptual drawing on synthetic missile image highlights what the key-point detection model will produce.

Results

A box-level detection model was trained with utilization of the YOLOv3 deep learning framework. The YOLOv3 model can successfully detect a distant missile object in-flight and generate the x-y coordinate values for the missile's location within the frame. For lower quality images, if resolution enhancement is required, the YOLOv3 model can be utilized as a layered pre-preprocessing step to identify the x-y coordinate values; allowing for dynamic zoom operations.

For operational efforts as well as future development efforts, pipelines for pre-processing and post-processing steps are created. The pipelines are designed to handle several image formats. Also, to produce the training and test datasets, image composing, and auto-labeling algorithms have been produced. With minimal manual preparations, the image composer solution utilizes a small collection of actual images to generate large quantities of training datasets with labeling files. Many of the pre-processing and post-processing steps have been completed. However, the custom missile pose estimator and the key-point detection model are still in development.

Future

This project can be scaled to perform multi-class foreground and background detections. Although the scope of the current project is limited to the detection of a single object-class, the model can be scaled to detect for multi-class objects simultaneously; as well as perform full environment analysis by including panoptic segmentation, which includes background detections. Detectron2 is a modular framework with benchmarking samples for adapting additional capabilities. In addition to understanding the attitude of missiles in-flight, a scaled model can be trained to identify background objects and spaces with which environmental setups can be defined dynamically.

Co-Investigators

Larry W. Alejo *(Army Test & Evaluation Command White Sands Missile Range)*
Janice Denny *(USAF Arnold Engineering Development Complex / Test Technology Analysis and Evaluation Branch)*

WORKFORCE OF THE FUTURE

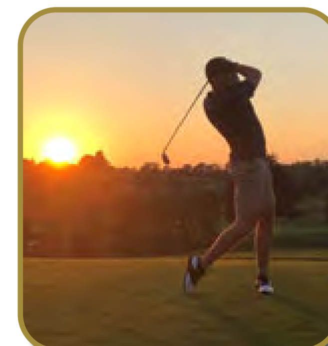
The High Performance Computing Modernization Program sponsors the annual HPCMP Internship Program (HIP) as part of workforce development, supporting undergraduate and graduate students. It is a mentor driven process where mentors submit proposals to the HIP team and successful mentors then select interns. Interns are expected to write a paper and prepare and present a brief at the end of the internship. The DOD and its contractors have succeeded in hiring greater than 20% of the interns. ARL submits proposals every year and has been quite successful, supporting 64 students over the last 4 summers.

Candidate HIP interns are US citizens exploring science, technology, engineering, and mathematics (STEM)-related career fields. However, the pool of STEM students who are US citizens and are interested in pursuing federal employment continues to dwindle due to competing opportunities offered by industry. Additionally, the country continues to see a relative decline in the number of US citizens enrolled in science and engineering undergraduate and graduate schools. It is essential to develop interest and computational skills earlier in the student pipeline, in high school, before students decide on an academic career path. In 2020, the ARL DSRC submitted a proposal to the HIP team to extend the STEM workforce pipeline. In addition to working with local schools to recruit high school interns, the project included engaging and training teachers to work with computer scientists who will impart their new knowledge and experience in their own classrooms. Although this proposal was not a typical HIP proposal, which encompasses computational technologies and high performance computing, Dr. Kevin Newmeyer, Deputy Director of the HPCMP, funded the proposal. Three teachers and four high school students from the surrounding counties were supported by this non-traditional internship.

We are pleased to feature the work of twelve of the interns from the summers of 2020 and 2021; three of these are our high school students. This took place during a time when COVID-19 halted many programs but ARL's flourished thanks to innovative and energetic scientist and engineer mentors and program coordinators, creative thinking, hours of planning, and adaptable students. The students are co-authors on ten articles that appeared in the June, September, and December 2021 issues of the ITEA Journal (or will in the March 2022 issue) as detailed below. Three of the articles were presented at the 38th International Test and Evaluation Symposium in September 2021. The paper by Sage Leone et al. won the Student Best Paper Award at the Symposium.

STUDENT BIOS

KATHERINE BRECZINSKI
is an undergraduate computer science and government double major at Franklin & Marshall College where she is now a senior. Katherine has taken many computer science classes, including artificial intelligence, where she first gained interest in the topic and learned about sentiment analysis, game theory, and heuristic and adversarial search. She took part in the High-Performance Computing Internship to expand her knowledge of artificial intelligence topics and high-performance computing and to gain research experience. She values the experience she gained working on a project with a team and improving her collaborative and communication skills. After her experience in HIP, she decided to keep learning more about artificial intelligence topics, and she started doing research involving language models with her computer science advisor. Katherine is currently the head tutor of the computer science tutor team, and she is also a part of the F&M orchestra, where she is the current concertmaster.
(HIP-21-010 Tools and Processes for Health Analysis Related to Blast Over-pressure)



CHASE CARBAUGH
graduated from the University of Maryland, Baltimore County in May 2021 with a Bachelor of Science in Computer Science. His HIP internship at the ARL DSRC was from June to December 2020. He loves to play golf with his friends and is passionate about football and gardening. Chase has since begun work as a Software Engineer for Booz Allen Hamilton and is immensely grateful for ARL and the opportunity the HIP programs offered him as the experience gained during this internship was pivotal in obtaining his current role.
(HIP-20-018: HPC Workload Characterization: Optimizing Data Ingestion and Initial Analysis of SCOUT Workloads)

TONY CRUZ
is an undergraduate student at the University of Maryland, Baltimore County. He is on track to graduate in the winter of 2021 with a Bachelor of Science in Computer Science. Since his internship at the ARL DSRC in the summer of 2020 he has accepted a fulltime position as a data science software engineer for MITRE and will begin in January of 2022. The HPC Workload Characterization project gave him an effective understanding of the elastic stack and how to properly characterize log data. Furthermore, Tony's internship at the ARL DSRC gave him a foothold within his computer science career that has helped him establish himself a voice within the professional IT community.
(HIP-20-018: HPC Workload Characterization: Optimizing Data Ingestion and Initial Analysis of SCOUT Workloads)

SELENA HAMILTON
is a sophomore at Harford Community College, transferring to University of Maryland, Baltimore County to major in Computer Science. While in high school, she participated in FRC as a Safety Captain, and later on the Programming Lead. She followed the Project Lead the Way (PLTW) track in high school which exposed her to many STEM classes and projects, and participated with the Science National Honor Society. Currently, she still participates in FRC as a programming mentor. She enjoys reading, sewing, longboarding, taking hikes and traveling. During her internship with HIP, she learned a lot about the development process for games and software, as well as how to work with and visualize large amounts of data. Her internship has inspired her to begin several passion projects (a website, and video game). As a result, she started to work towards certification for Web programming and Game development. She is working towards this by taking online classes using the website edX.
(HIP-20-023: Immersive Environments for Visual Analytics)

JAMES HUGHES

is a college sophomore studying at University of Maryland College Park in the Clark School of Engineering. Looking to be a Mechanical Engineer, he aspires to work in the Aerospace field working in a corporate laboratory. James is also a UMD Public Leadership Scholar and seeks to better lead and prepare his generation of peers for technical and social problems of the future. As a 10-year plus Scouting veteran and Eagle Scout, he enjoys taking long backpacking trips in ludicrously short amounts of time and teaching others about the Great Outdoors.
(HIP-20-023: Magical Music Machinations: A Virtual Reality Outreach Tool)

NICHOLAS KENDALL

is a sophomore at University of Wyoming studying Mechanical Engineering. He has a passion for additive manufacturing, and he put his skills to the test with developing a new methodology to visualize the results of HPC simulations of high velocity events that occur during the formation of a shaped charge jet. Upon graduation, he aspires to work in a DoD, federal or industry laboratory utilizing his engineering and mechanical skills. Nicholas is an Eagle Scout, an avid skier and mountain biker and a mountain bike coach. He is also owner and head mechanic of Cycle On LLC, a bike repair service.
(HIP-20-023: 3D Printing of High Performance Computing Simulation Results)

ALEXANDRA LEMOINE

is a Ph.D. student in the Department of Aerospace Engineering at the University of Maryland. During her time at the University of Maryland, she has acted as a teaching assistant and substitute instructor for undergraduate aerospace engineering classes. In addition, she has collaborated with the Women in the Engineering Department to design and instruct a 2-week outreach program focused on introducing incoming engineering students to computational methods for solving engineering problems. Alexandra holds a M.S. in Mechanical Engineering from the University of Wisconsin, Madison, and a dual B.S. in Aerospace Engineering and Mathematical Sciences from the Florida Institute of Technology. Prior to attending the University of Maryland, she worked as a research engineer focusing on computational fluid dynamics of automotive engines for 2 years at Convergent Science Inc. Her academic interests are in Computational Fluid Dynamics, Scientific Computing, and Fluid Mechanics.
(HIP-20-17 Deposition Models for CMAS Attack in Gas Turbine Engines)

JACOB LEBLANC

is a senior at Stanford University studying computer science in the human-computer interaction track. Although he goes to school in California, he lives in the small town of Lafitte, Louisiana. Outside of schoolwork, he can be found working for the Ideas Out Loud (a club that organizes Stanford's TEDx) tech and innovations team. After graduation,he plans to pursue software development.
(HIP-21-011: Distributed Processing and Analysis for Large-Scale Heterogeneous Data: Text Classification and Analysis through Clustering)

SEAN LEONARD

is a master's student in the Computer Science Department at Old Dominion University. He was an intern at the Army Research Lab in Aberdeen, Maryland, building an implementation for ParaView Catalyst into a CFD simulation framework used at ARL. This objective started when he was an intern and then spilled over into a collaborative effort where Sean Leonard, as an undergraduate, led a team of undergraduates to finish building the in-situ visualization implementation. Mr. Leonard has worked at Newport News Shipbuilding on a natural language processing project. Mr. Leonard has also worked on a flow visualization tool and porting basic features into a VR environment. His research interests are in machine learning, bioinformatics, and scientific visualization.
(HIP-20-017 Application of CinemaScience Methods to Visualize and Analyze Complex Multiphase Flow Processes Supporting Next Generation Future Vertical Lift (FVL) Concepts)

SAGE LEONE

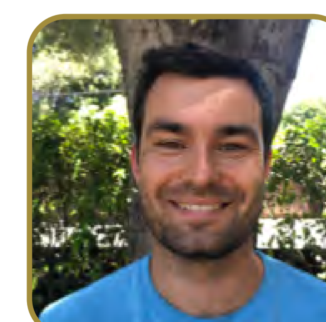
is a junior at the University of Maryland, College Park majoring in computer science and minoring in mathematics. During her internship with ARL and the HPCMP, she developed skills in data science and high-performance computing that she has continued to build in the classroom and further internships. Currently, she is involved in undergraduate computer science research at the University of Maryland developing a platform to facilitate conversation among communities of social chatbots. She is also an organizer for a campus hackathon for underrepresented genders in tech and is a part of the University of Maryland club sailing team. Outside of school, she enjoys spending time outside, including hiking, skiing, and running. After graduation, she plans to enter the workforce as a software developer.
(HIP-20-046: Topic Modeling and Visualization for Test Incident Report Data)

PATRICK MORTIMER

is a PhD student at the University of Texas at Austin in Aerospace Engineering. He earned his MS from the same. His research focus is on experimental flow field and performance measurements for coaxial, counter-rotating rotors. A specific emphasis is placed on investigating the nonlinear effects of dynamically changing rotor operating conditions. Dynamic changes involve executing a step or sine input to vary operating conditions with respect to time. The two main operating conditions under investigation are the blade pitch (collective and cyclic) or rotor rotational speed. Particle Image Velocimetry (PIV) measurements are used to collect and analyze the flow field data, whereas performance data is collected through various force transducers. He participated in the High-Performance Computing Internship Program (HIP) with the goals to gain knowledge in Computational Fluid Dynamics (CFD) research, as well as his high-performance computing (HPC) knowledge.
(HIP-21-012: Nonlinear Response of a Coaxial Rotor to Blade Pitch Change)

MARIYA OCCORSO

is an undergraduate student at Arcadia University where she is studying for her Bachelors of Science in Computer Science. Originally a student at the University of Maryland Baltimore County, Mariya transferred in the Fall of her sophomore year to Arcadia University. She is now a rising senior having taken numerous computer science and mathematics courses during her time at Arcadia University and UMBC. Mariya has earned a place on the Dean's Honor List or Dean's Distinguished List since she has been at Arcadia. Before the COVID-19 Pandemic sent Arcadia University online, Mariya worked as a student assistant at both the Digital Fabrication Lab and Virtual Reality Lab on Arcadia's campus. She is currently a member of the varsity Women's Lacrosse team at Arcadia, having just completed her second season. Upon graduation, Mariya will return to the ARL DSRC
(HIP-20-019: Visual Analytic Application Toolset for Large-Scale Heterogeneous Data)
(HIP-21-011: Distributed Processing and Analysis for Large-Scale Heterogeneous Data: Text Classification and Analysis through Clustering)



NIKOS P. TREMBOIS

earned a Master of Science degree in Aerospace Engineering at University of California, Davis where he is currently pursuing a PhD. He entered with an interest in computational fluid dynamics and aeroacoustics and has primarily focused on computational aeroacoustics of rotorcraft. Along the way, he also worked on grid generation. He found grid generation to be a long process that took away from the research which he really refers to spend most time on, so he automated the grid creation process. Through automating the pre- and post-processing steps of CFD, he gained more time for researching aeroacoustics and enabled himself to carry out a wider variety of simulations. He took part in the High-Performance Computing Modernization Program Internship Program internship with the expectation of researching rotorcraft aeroacoustic
(HIP-21-012: Computational Aeroacoustics of Stacked Rotors)

Tools and Processes for Health Analysis Related to Blast Over-pressure

Katherine M. Breczinski
Franklin & Marshall College

Project Description

The goal of this project was to better understand data analyses to improve operational medicine for the military in terms of blast over-pressure studies (BOS). Also of interest was how this data should be cataloged and made available to researchers pursuing this topic. The majority of data related to this field consists of concentrations of blood biomarkers present, Defense Automated Neurobehavioral Assessment (DANA) metrics, and self-reported symptoms by study participants. Python tools were used to reproduce the analyses from the literature, including Spearman's rank correlation, analysis of variance, and interquartile range assessments. Machine learning methods were also explored to see if useful correlations could be found. This work shows the advantages of having a structured approach for data storage and access, as well as highlighting the ability for common open-source tools and statistical and machine learning packages to produce meaningful results.

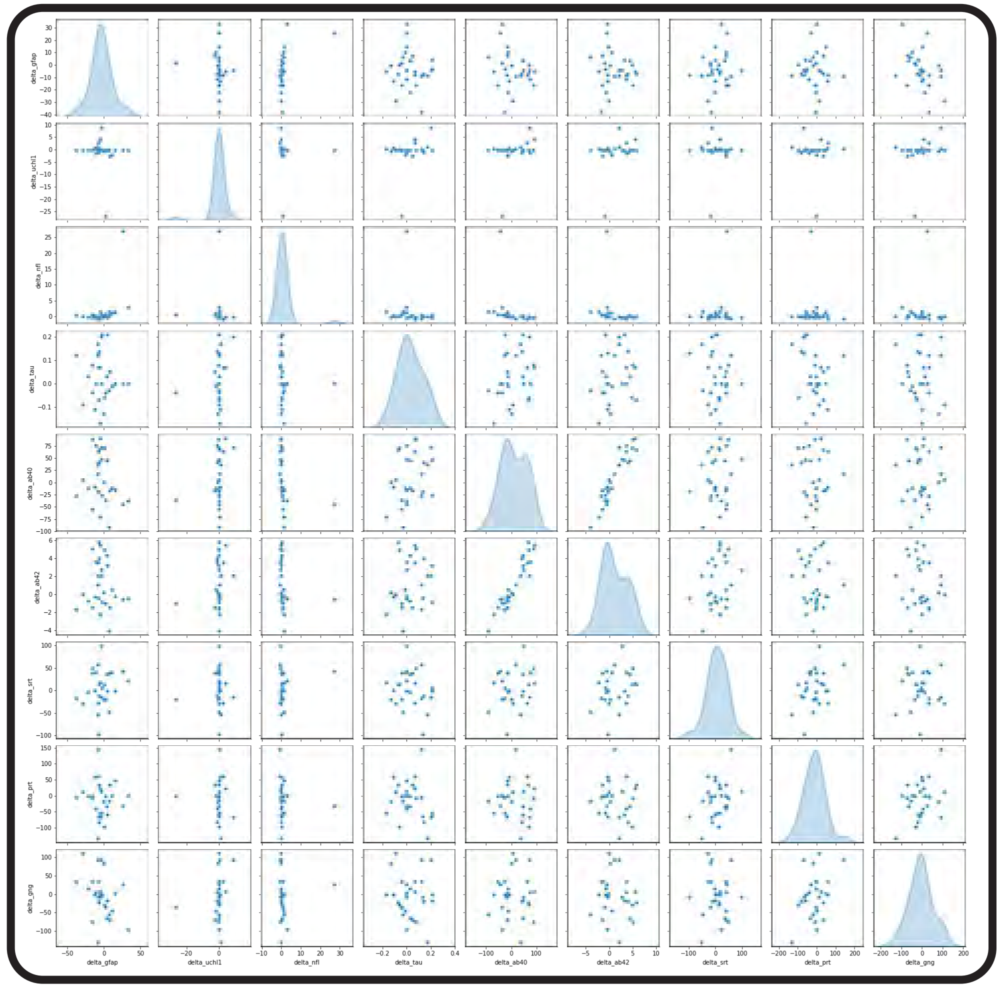


Figure 1: Pairplot of blood biomarkers and DANA metric

Relevance of Work to DOD

Over-pressure events generate pressure waves exceeding atmospheric pressure. Exposure to blast over-pressure can have severe negative effects on people and their cognitive health. For example, soldiers involved in breaching events, such as those that use explosives to punch holes into structures like walls or fences, are commonly exposed to over-pressure. Soldiers proximate to weapon systems can also experience over-pressure. Exposure to over-pressure events can lead to symptoms similar to those experienced with concussion or mild Traumatic Brain Injury (mTBI). The BOS effort is attempting to understand and mitigate the effects of blast over-pressure events for the military, and in particular address issues related to TBI

Computational Approach

Jupyter Notebooks, Python, and open-source Python libraries were imported and used to assist in the extract, transform, and load process of the data from a comma-separated value (CSV) file. Pandas DataFrames were used to store the columns of data for the analyses. The Seaborn library was used for visualization of the blood biomarkers and DANA metrics. Figure 1 shows the pairplot of the blood biomarkers and DANA metrics. The SciPy.stats library was used to perform one-tailed Spearman's rank correlation. The Seaborn library was also used to create heatmaps using blood biomarker and DANA metrics data.

Results

Some of our analyses produced results divergent from those reported in the open literature. These remain under investigation and are currently unresolved. To see if outliers were causing discrepancies between the reported results and the reproduced results, the correlation of blood biomarkers GFAP and UCH-L1 were recalculated after taking away entries 6 and 18 as potential outliers. However, the recalculated result was similar to the results with the outliers included, and did not account for the discrepancies with the paper.

Differences were found in the median and interquartile range calculations as well. The k-means clustering did not add any significant information. The data points were not tightly clustered around the centroids, so it was difficult to see whether there was any pattern to how people were grouped based on changes in the DANA metrics. There were three DANA tests considered in this data: Simple Reaction Time (SRT), Procedural Reaction Time (PRT), and Go/No-Go (GNG). SRT assesses basic reaction times to visual cues, PRT measures attention and processing speed, and GNG assesses speed, accuracy, and possible omissions.

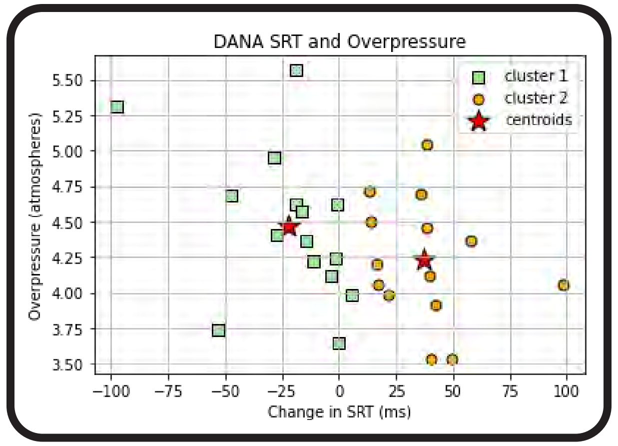
An analysis using the elbow-curve method showed that there were two, possibly three, distinct groups. Two groups were explored under the assumption that there was one group that did better from pre- to post-blast exposure, and one that did worse. However, since the data is so scattered on the change in performance for each of the k-means graphs, it is hard to draw any conclusions. Figures 2, 3, and 4 show the k-means clustering of the DANA metrics against SRT, PRT, and GNG respectively. The scattered nature of the graphs was most likely due to the fact that n=29 in the data set being used. A larger data set might show tighter clusters around the centroids. The discrepancies found between this analysis and that reported in the literature shows that having a dedicated and streamlined platform to perform calculations and house data would be beneficial

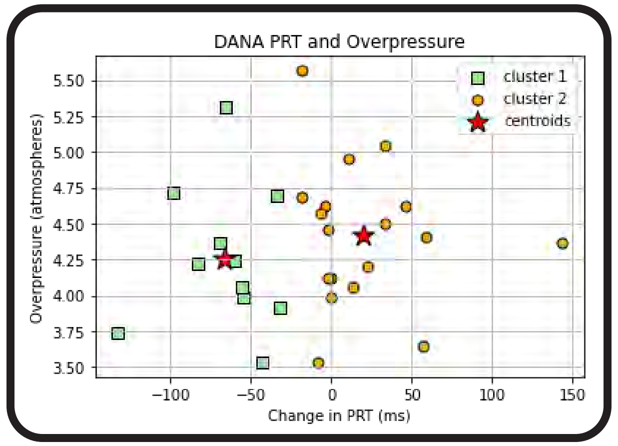
Future

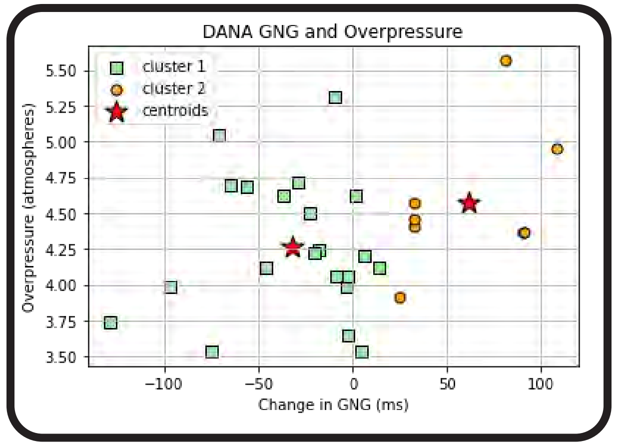
We want to understand discrepancies found between this analysis and that reported in the literature. The small size of the data set limited the potential impact of machine learning approaches; we are obtaining data from a study of 200 – 300 subjects. Random forests are being explored for Likert-scale data as a possible analysis to help show and predict the relationships between the blood biomarkers, DANA metrics, and self-reported symptoms.

HIP Mentors

Dale Shires *(DEVCOM ARL DOD Supercomputing Resource Center)*
Michael An *(DEVCOM ARL DOD Supercomputing Resource Center & Parsons)*







(top) Figure 2: SRT & overpressure
(middle) Figure 3: SRT & overpressure
(bottom) Figure 4: GNG & overpressure

Publications

Katherine M. Breczinski, Dale Shires, and Michael An, "Tools and Processes for Health Analysis Related to Blast Over-pressure," The ITEA Journal, 42 (4), December 2021, pp. 256-261

HPC Workload Characterization: Optimizing Data Ingestion and Initial Analysis of SCOUT Workloads

Tony Cruz,
University of Maryland, Baltimore County

Chase Carbaugh,
University of Maryland, Baltimore County

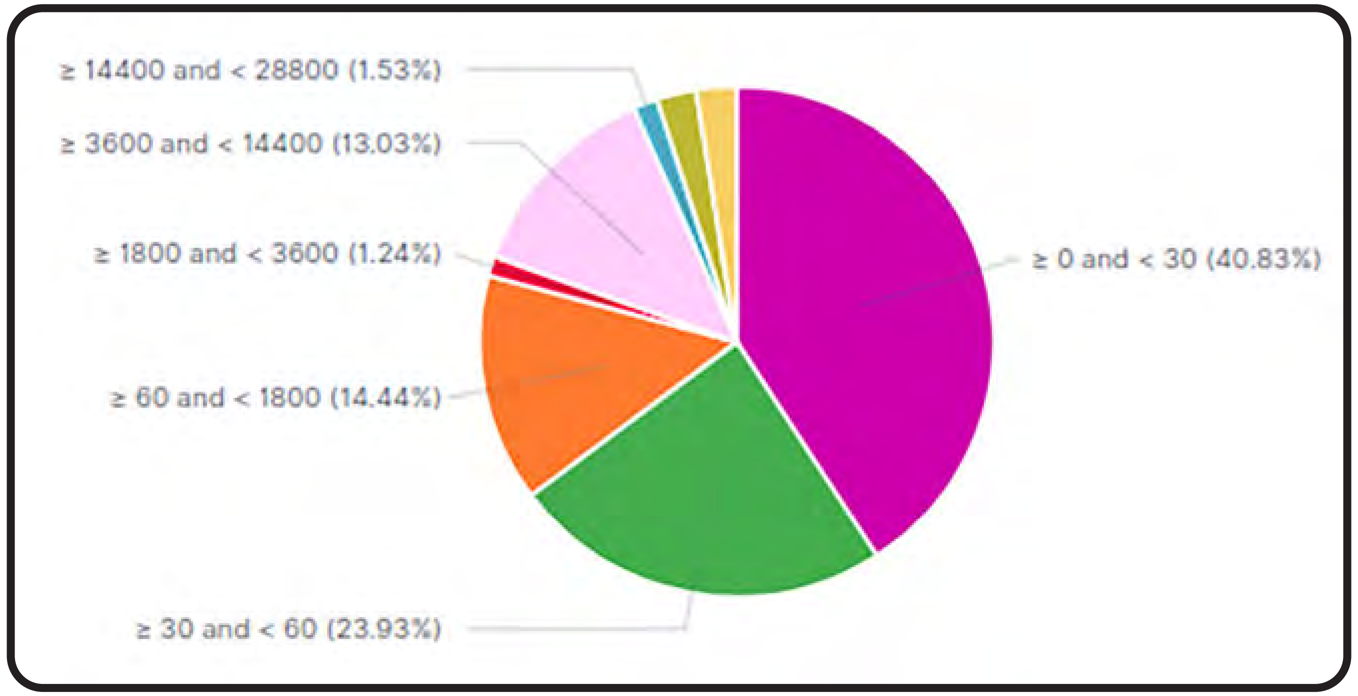


Figure 1: Job runtime organized into discrete bins

Project Description

The size and power of a computer positively correlates with the complexity of its management, making High Performance Computing (HPC) systems management difficult. We describe an application and optimization of methods to characterize the workload of the new artificial intelligence (AI) and machine learning (ML) platform at ARL, SCOUT (SuperComputing OUTpost). In the current project we built upon a prior implementation of the ELK stack to ingest workload data from SCOUT's load sharing facility. With an ingestion pipeline in place we are able to collect, transform, visualize, and characterize job level data from ad-hoc, managerial, and exploratory standpoints.

Relevance of Work to DOD

Collectively, the Department of Defense Supercomputing Resource Centers (DSRCs) provide scientists and engineers with in excess of 7 billion processor hours of computing power per year. The ARL DSRC runs roughly 5 million jobs on its HPC systems each year. Without access to job and node level metrics users are unaware of the resources that their jobs consume. Metrics offer a multi-level view of the interactions between a computer and its users, providing a better understanding of machine operation and the relationships between the user, the user workloads, and the system. Sub-optimal use of HPC system resources leads to downsides for the HPCMP, such as longer wait times for the user and higher operating costs per facility. With workload data indexed we can perform predictive analysis and forecast anomalies.

Computational Approach

SCOUT is a non-traditional HPC architecture designed specifically for AI/ML workloads. It is the first production IBM POWER9 architecture and LSF workload manager in the HPCMP. It is also the first production system with multiple GPUs on every node: 6 Nvidia Volta V100 GPUs on 22 training nodes and 4 Nvidia Tesla T4 GPUs on 128 inference nodes. Because the system is so different and is designed for a unique workload, we expected users to interact with the system in a different way.

We implemented the open-source ELK stack to construct an ingestion pipeline and provide front end visualization. We then optimized the transformation process and characterized SCOUT's initial workloads. Workload characterization is the process of leveraging log data ingested from the workload manager, individual compute nodes, and applications. The ELK Stack includes Filebeat for collection, Logstash for transformation, Elasticsearch as a search engine, and Kibana for front end visualization.

Elasticsearch allowed us to quickly pull data for ad hoc and exploratory analyses within Kibana's front end visualizations. We implemented Ruby with splits and scans to match key field data. Workload characterization is a data centric process, so we first identified the properties of our data set. The data contained information for 10,467 jobs and included jobs run between April 5th and July 29, 2020. We used graphical visualizations to characterize the distribution of workloads, focusing on the distribution of jobs in relation to their runtimes, how they exited, the number of GPUs they requested, and their node types.

Results

Figure 1 sorts jobs by runtime bins, with numbers representing seconds. The majority of jobs had runtimes from 0-60 seconds, totaling over 65% of job runtimes. Including jobs with runtimes between 1 and 30 minutes, runtimes below 30 minutes account for over 80% of the initial jobs run on SCOUT.

Shifting focus to how jobs exited (not pictured), we learned that nearly 80% of the initial workloads on SCOUT completed successfully. Coupling this with the fact that over 65% of the jobs had runtimes below 1 minute reinforced our idea that users were warming up to the system by running small jobs to test system capabilities. We next analyzed the distribution of jobs based on the number of GPUs requested, discovering that nearly 90% contained jobs that ran with at least 1 GPU, with almost 80% of jobs running with either 1 or 4 GPUs.

We visualized the distribution of GPUs per job by runtime buckets, see the dashboard in Figure 2. Combining this with information about the distribution of jobs by GPU models and runtimes by GPU models, we learn that 1/4 of the jobs used training nodes. In fact, the total runtime of these jobs also accounted for 1/4 of the total runtime. We conclude that there are nearly 4 Tesla T4 GPUs for every 1 Tesla V100 GPU, confirming that GPU model counts are proportioned well on SCOUT, with its model ratio of 4:1.

Future

Future work will include more detailed node level data from which we will be able to characterize how each node performed throughout the duration of a workload. Other work includes setting up the ELK stack on SCOUT to ingest logs in real time and applying Ruby methods to the complete ingestion pipelines. Real-time analytics are useful for cases such as detecting anomalous behavior as it occurs or simply monitoring job metrics during runtime.

HIP Mentors

Matthew Dwyer *(DEVCOM ARL Computational & Informational Sciences Directorate)*

Publications

Tony Cruz, Chase Carbaugh, and Matthew Dwyer, "HPC Workload Characterization: Optimizing Data Ingestion and Initial Analysis of SCOUT Workloads," The ITEA Journal, 42 (3), September 2021, pp. 147-152

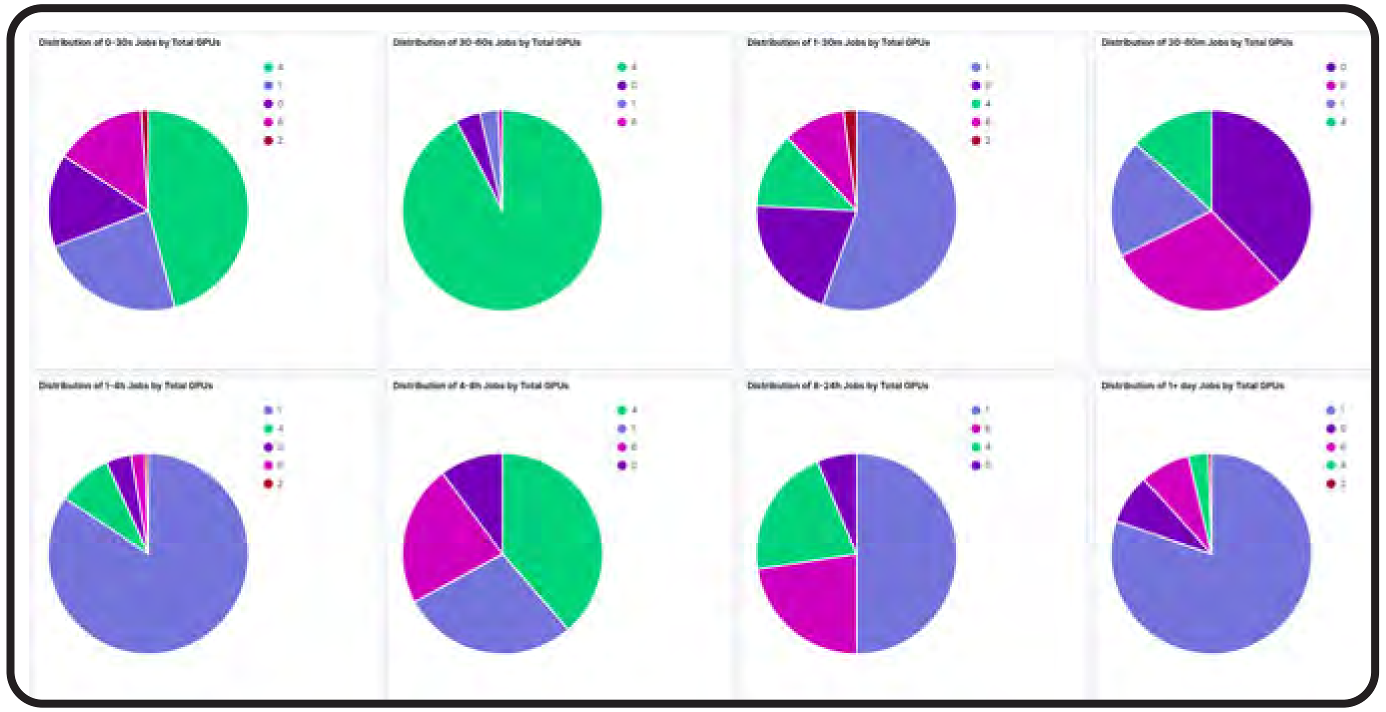


Figure 2: Distribution of GPUs by runtime bucket

Magical Music Machinations: A Virtual Reality Outreach Tool

Selena Hamilton,
University of Maryland, Harford Community College

James Hughes,
University of Maryland, College Park

Figure 1:
Flappy Trombone Room

Project Description
We describe a virtual reality experience that could be used at outreach events to introduce a wide range of students to the capabilities within computer science and help stimulate their interest in Science, Technology, Engineering and Mathematics (STEM) related fields. We created a virtual escape room game that offers an entertaining and interactive experience for the user, as well as a showcases for a few of the capabilities within Computer Science and Mechanical Engineering.

Relevance of Work to DOD
The objective of this project is to inspire students to pursue STEM fields by showing them how computer science and engineering can be fun. Additionally, we wanted to show them that this virtual reality project was created by high school students and that extra-curricular STEM programs can be fun and engaging.

Computational Approach
We identified all the requirements of the project as well as things we wanted to incorporate. We chose to create our own controller for the game since the design and construction of the controller would demonstrate mechanical engineering skills. We decided on an escape room style game in which the player must play four different instruments in four separate rooms. Each room offers its own challenge which the player must overcome to exit that room.

Once all challenges are completed in all four rooms, the player can escape through the main door. In addition to the main room and the four challenge rooms, we created a free playroom that allows the user to practice playing all instruments, providing them more experience to complete the room challenges. The practice playroom also provided us with an area to test our code. We named this application Magical Music Machinations.

We created visual representations of the game flow as well as floor plans of the individual rooms. This allowed us to develop a list of required assets before building the game itself. We then scoured the Unity asset store and internet for premade assets. After finding as many premade and usable assets as possible, we divided the remaining assets among ourselves for creation, and created the rooms.

Results
We created the following rooms.

Main Hub. This is where the player spawns in and where they access doors to the games as well as the Free Playroom and the door leading to escape.

Flappy Trombone Room (Figure 1). Similar to the Flappy Birds Game, this room was made to look like an arcade game in which the player guides the flappy trombone through obstacles using the trombone instrument, holding notes for a specific amount of time. Once the player keeps the flappy trombone safe for a specified amount of time, they receive a key instead of tickets usually gifted in arcades games.

Rhythm Room. The idea is to keep a steady beat using a drum for a stream of soldiers marching in front of the player. If the player manages to hold the tempo through a series of changes then a key will fall out of a soldier's pocket in front of the player, allowing them to leave.

Simon Says Room (Figure 2). Each light corresponds to a piano key. The lights play in a random order and get increasingly difficult. If the player follows the lights correctly for five rounds, the game key will fall from a compartment on the piano.

Snake Charmer Room. Using a flute, the player, who is on a balcony in this room, must go through the process of elimination and memorization to charm the snake into rising high enough to allow the player to grab the key tied around its neck/head. Using the key, the player can escape the room.

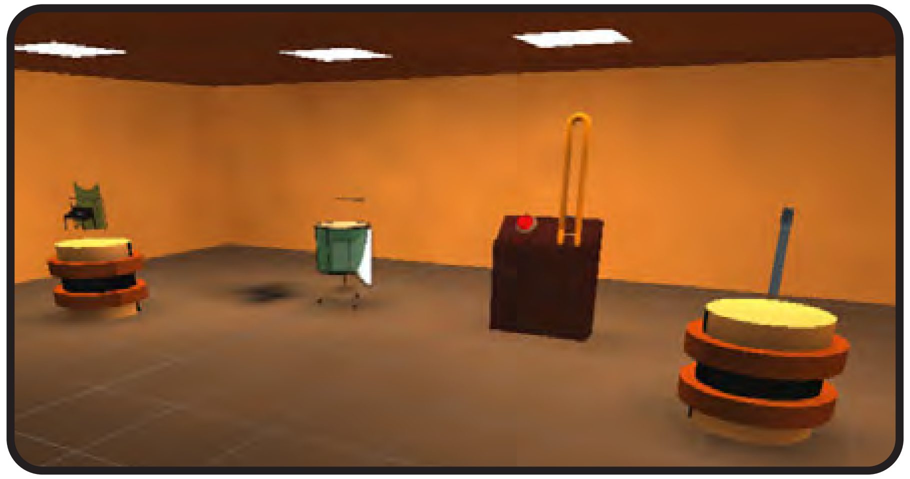
Free Playroom (Figure 3). This room has access to all the instruments in the game and allows players to interact with and practice all of the instruments in case certain challenges prove too difficult. This room was also used to test our code and instruments to ensure they worked correctly.

Once the rooms were created, we split up the remaining tasks, one person coding the interactions within the game and ensuring everything worked properly, and the other designing and creating the controller. We built a prototype controller using Play-Doh to fashion buttons anchored on cardboard. Using a circuit board, alligator clips, and a USB cable, the Makey Makey uses closed-loop signals to send the computer keyboard or mouse click signals mapped to our game. With assistance from another intern, the button tray was 3-D printed with Polylactic Acid filament (PLA) and the buttons themselves were printed with conductive PLA.

At the completion of our internship, the rooms are complete, a player can visit and interact minimally with the assets in each room, and the controller is integrated into the game. All instruments are capable of being played and making noise as well as being controlled by the controller except for the trombone.

Future
Magical Music Machinations is suitable to show to students as an outreach tool; however, many additional improvements can be made. We want to add better sound quality for the instruments as well as give the user more control over the notes and how long the notes sustain. We want to complete the trombone and integrate all the rooms so that the user can complete all the challenges, win the keys, and escape from the game.

HIP Mentor
Virginia To *(DEVCOM ARL DOD Supercomputing Resource Center & Parsons)*



(Above) Figure 2:
Simon Says Room

(Below) Figure 3:
Free Playroom

3D Printing of High-Performance Computing Simulation Results

Nicholas Kendall,
University of Wyoming

Project Description
This work demonstrates the feasibility and methods used to 3D print high-performance computing (HPC) simulation results of the formation of shaped-charge jets and penetration into armor. The 3D simulations were conducted using computing resources from the Department of Defense HPC Modernization Program and the fused deposition modeling printer used was the Ultimaker S3. The conversion of 3D HPC simulation results for use with additive manufacturing is presented and shown as a new tool for researchers to visualize, analyze, and compare HPC simulations of shaped charge jet formation and penetration to experimental results.

Relevance of Work to DOD
This capability provides researchers with a new tool for analyzing and visualizing complex phenomena through 3D physical representations of those simulations. We generate 3D physical representations of complex ballistics interactions that are impossible to obtain through experimentation. It also provides a tool for physical comparison of final experimental results to simulation results.

Computational Approach
A shaped charge warhead consists of a liner made from a ductile material, typically copper, that is recessed into an explosive forming a hollow cavity. These liners come in different shapes, materials, and sizes, however, the ones used in this study were all simulated as conical copper liners. Copper produces a long, thin, high-velocity, stretching penetrator called a jet to form from the detonating explosive, which can efficiently penetrate armor. Figure 1 shows the setup for a shaped charge simulation. The pink and yellow objects on the left make up the shaped charge warhead, which consists of the explosive (pink) and a copper liner (yellow). The alternating shades of grey blocks are rolled homogeneous armor (RHA) steel target blocks. When the explosive detonates, a shockwave propagates through the copper liner causing the copper liner to invert from the cone apex forward to the base within the hollow cavity, forming a high velocity jet that penetrates the RHA plates.

We employ solid mechanics codes in computational simulations to analyze these multi-material, large deformation, strong shock wave events. We exploit additive manufacturing with the computational simulations for a three-dimensional (3D) representation of the high velocity events. The software tool used to visualize the simulation was Paraview version 5.8.0. The initial model of the warhead was taken before it was detonated. The liner and explosive were isolated and a filter was applied to the data to extract the surface. This then allowed for a triangulation filter to be applied to the extracted surface. These two steps were essential because the fused deposition modeling (FDM) printer requires the surface geometry of a 3D object to be made out of facets. Since the two models were encoded with triangular surface features, the models could be exported out of Paraview as a stereo lithography (STL) file. These STL files were imported into the Ultimaker Cura slicer program for the Ultimaker S3 3D printer to produce a geometry code (g-code). Once the g-code was produced the models were 3D printed.

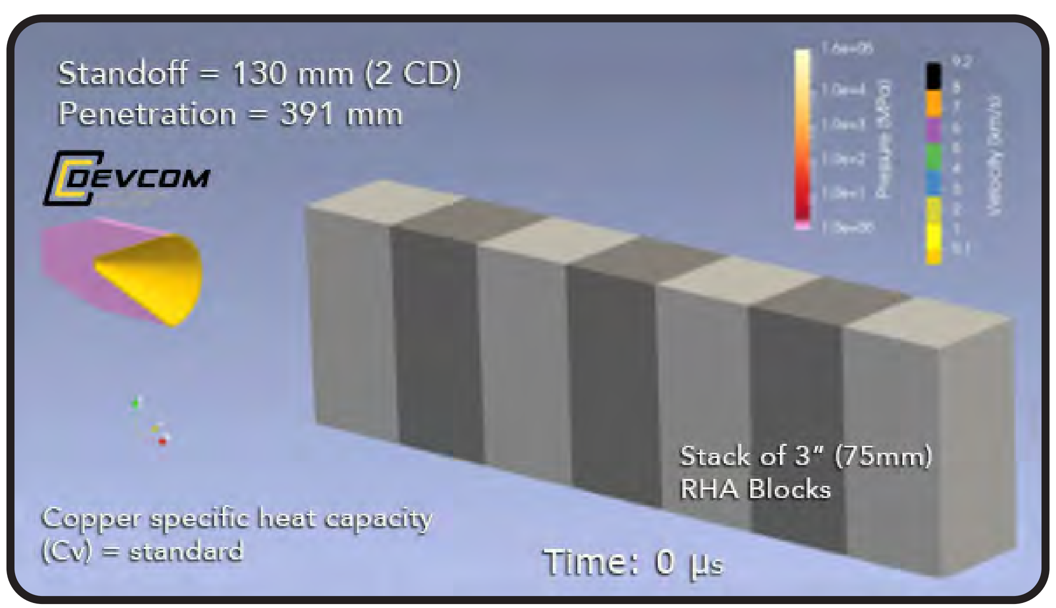


Figure 1:
Setup for a shaped charged simulation

Figure 2:

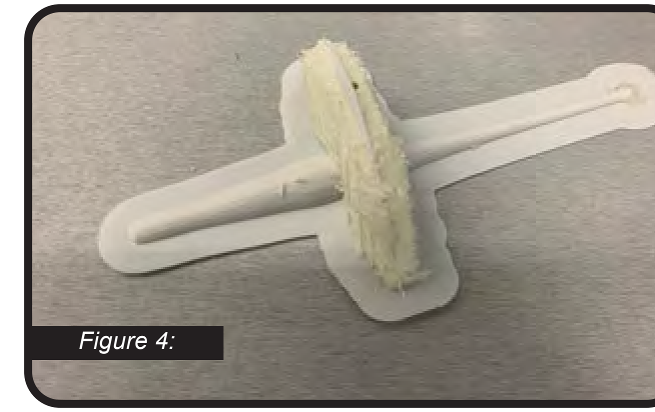


Figure 4:

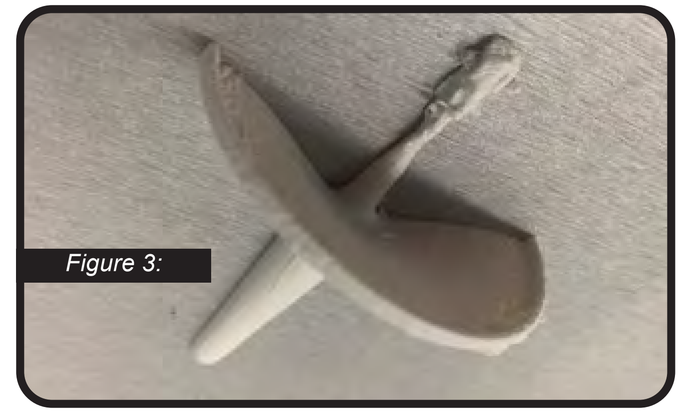


Figure 3:

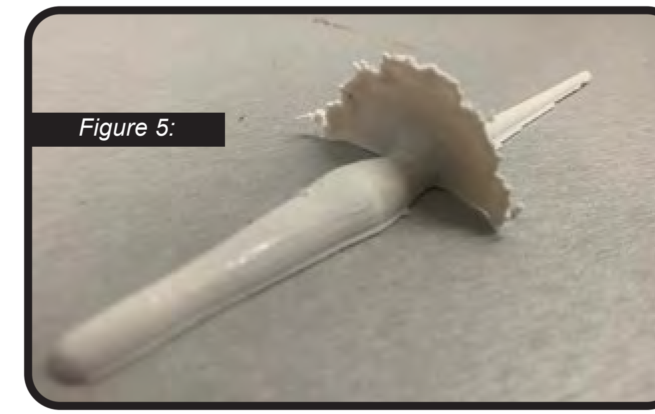


Figure 5:

Figures 2 & 4: 3D printed shaped charge jet in half symmetry with PVA supports.
Figures 3 & 5: The same shaped charge jets with PVA supports removed.

Results
Large deformations in the liner occur during jet formation resulting in difficulties printing a 3D model, creating overhangs, small surface features, and floating entities. Since FDM printers use a thermoplastic filament that is extruded layer by layer, they cannot replicate overhangs or suspended particles without support structures. The Ultimaker S3 3D Printer used for this work used two different materials. The first is Polylactic Acid (PLA) plastic, the most common plastic used in desktop 3D printing. It is made from renewable resources such as sugar cane or corn starch. The second material is Polyvinyl Alcohol (PVA). PVA enables more complicated geometries and overhangs to be created, which can be removed later by dissolving the PVA in water leaving behind the solidified PVA model.

Figures 2 and 4 show the 3D printed shaped charge jet in half symmetry along with the PVA supports. Since the jet formation at a later timestamp was more delicate, addition of a PLA raft was required to increase the surface area of the print to ensure the print adhered to the print surface. Figures 3 and 5 show the same shaped charge jets with the PVA supports removed. The PVA worked reliably as a support system for these 3D prints and a method for printing shaped charge jet formation from 3D HPC simulations was demonstrated.

Future
The exporting process and complexity of generating the required g-code varied between the shaped charge jet and the target blocks. The shaped charge jet exported cleanly to the slicer; the target blocks did not. The target blocks had lots of manifold errors, meaning that the model was either invalid or undefined at certain locations, causing certain import/export errors during the process. Some of these errors go on to cause printing errors that we do not fully understand. We need to conduct further research and experimentation to investigate and fix the root causes.

HIP-Mentor
David Kleponis *(DEVCOM ARL Weapons & Materials Research Directorate)*

Co-Investigators
Ethan Eldridge, and Stephen Schraml *(DEVCOM ARL Weapons & Materials Research Directorate)*

Publications
Nicholas Kendall, Ethan Eldridge, David Kleponis, and Stephen Schraml, "3D Printing High Performance Computing Simulation Results for Visualization and Analysis," The ITEA Journal, 43 (1), March 2022, accepted for publication

Deposition Models for CMAS Attack in Gas Turbine Engines

Alexandra Le Moine,
University of Maryland

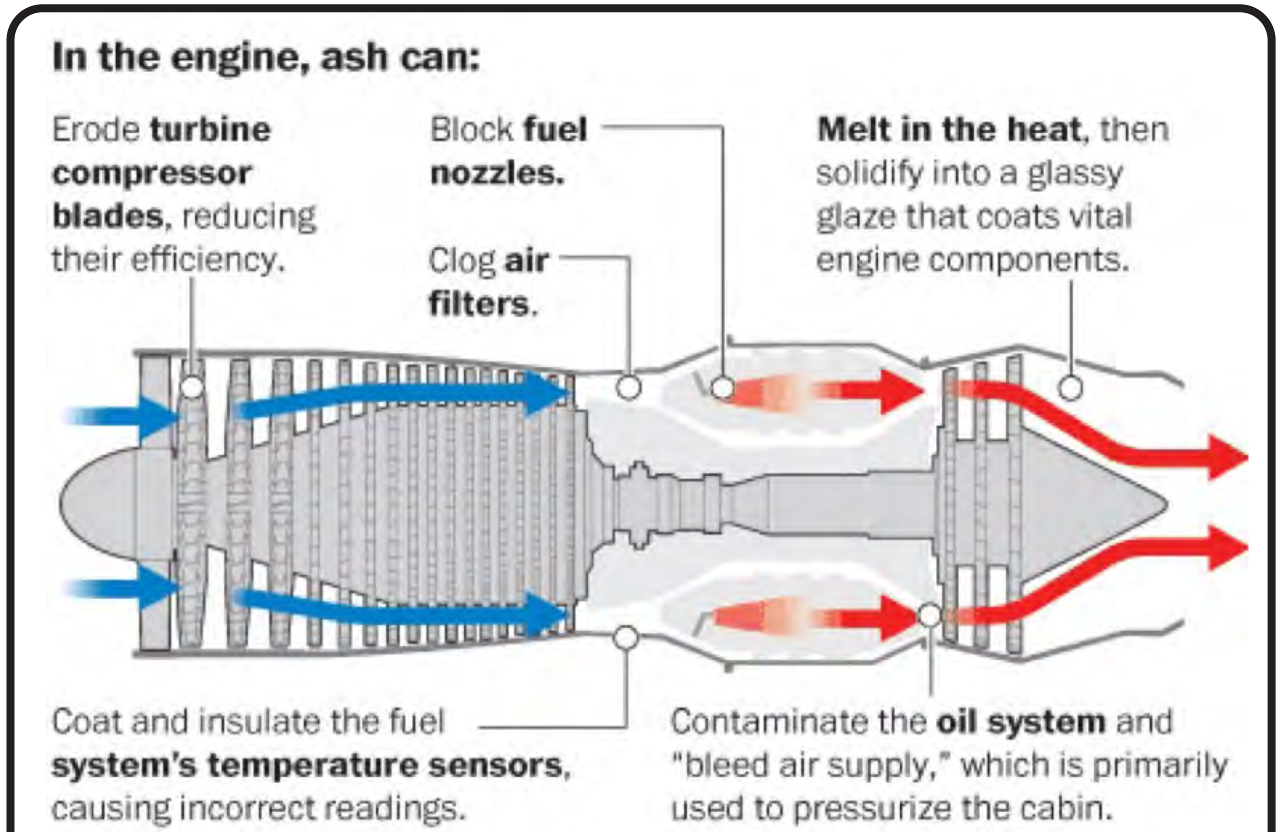


Figure 1: Volcanic ash effects on engine components

Project Description

We lay the groundwork for investigating the effects of particle deposition on thermal and environmental barrier coatings in gas turbine engines subject to environmental particle ingestion (e.g. sand/ash/salt). Research indicates that high-fidelity point-particle computational simulations of the flow regime of interest will be useful in studying the underlying flow physics surrounding particle deposition on gas turbine engine components, see Figure 1. As such, several deposition models have been incorporated into high-fidelity computational fluid dynamics software, Athena-RFX. The deposition models have been rigorously tested for functionality and preliminary tests of particle size on deposition have been performed.

Relevance of Work to DOD

The ingestion of solid particulates, such as sand or volcanic ash, in fixed-wing and rotary-wing engines can lead to severe engine damage. Similarly, there have been many reported rotorcraft incidents caused by brownout, a phenomenon in which the rotor downwash causes a cloud of sand to engulf the helicopter, such as the fatal V-22 Osprey crash landing in May 2015. In many of these incidents, sand or volcanic ash ingestion in the gas turbine engine led to power-loss or complete engine failure. The current work is aimed at understanding the effects of erosion, clogging, and deposition within a gas turbine engine.

Computational Approach

The primary objective of this study was to use an existing high-fidelity computational framework and implement and test a particle deposition model for calcium-magnesium-alumino-silicate (CMAS) particles under gas turbine engine operating conditions. We employ the Athena-RFX computational fluid dynamics solver to model the continuum fluid with Lagrangian particle tracking for massive particles. Athena-RFX incudes 4 depositions models: critical velocity model, critical viscosity model, Singh and Tafti model, and Bons OSU model. The direct numerical simulation calculations were perform on the ERDC DSRC Cray XC40/50 platform, Onyx.

Results

Results are presented for the critical viscosity model, which considers particle deposition to occur based solely on the particles temperature. In these test cases, the soften temperature was set to the softening temperature of CMAS. A simple rectangular geometry was used with the lengths in the x, y, and z directions equal to 1.0 cm, 1.0 cm, and 0.5 cm, respectively. A circular injector boundary condition was set in the x-y plane at z = 0 cm and a wall boundary condition was set in the x-y plane at z = 0.5 cm. The remaining boundary conditions were set to an outflow. In the next tests we forced the particle temperature to be much greater than the softening temperature and much less than the softening temperature. Heat transfer in the particle was turned off to remove particle heating or cooling. Figure 2a shows the results when the particle temperature is greater than the CMAS softening temperature.

As expected, all particles deposit onto the wall. Figure 2b shows the results when the particle temperature is less than the CMAS softening temperature. As expected, no particles deposit onto the wall and instead experience full rebound. The results from this preliminary study indicate that the particle diameter is an important parameter regarding the characterization of CMAS deposition in gas turbine engines. In addition, the study suggests that the high thermal and turbulent fluctuations in the turbine stage can impact the dynamics of the particles

Future

It is known that CMAS particles have higher surface tension, density and viscosity than fluids that are traditionally used to validate deposition models. As such, there is a need for a more realistic, physics-based deposition model for molten CMAS particles that includes surface tension, density, and viscosity variations. The Athena-RFX simulations were performed using very coarse grids. In principle, Athena-RFX, a Direct Numerical Simulation code, allows for all scales of turbulence to be resolved. Further work will be conducted refining the grid size to approach the order of the Kolmogorov length scale.

HIP-Mentors

Dr. Simon Su *(National Institute of Standards & Technology formerly of the DEVCOM ARL DOD Supercomputing Resource Center)*
Dr. Luis Bravo, *(DEVCOM ARL DOD Supercomputing Resource Center)*

Co-Investigators

Dr. Anindya Ghosha & Dr. Muthuvel Muruga *(DEVCOM ARL Computational & Informational Sciences Directorate)*
Dr. Alison Flatau *(University of Maryland)*

Publications

Alexandra Le Moine, Alison Flatau, Simon Su, Luis Bravo, Anindya Ghoshal, and Muthuvel Murugan, "Deposition Models for CMAS Attack in Gas Turbine Engines," The ITEA Journal, 42 (3), September 2021, pp. 164-172

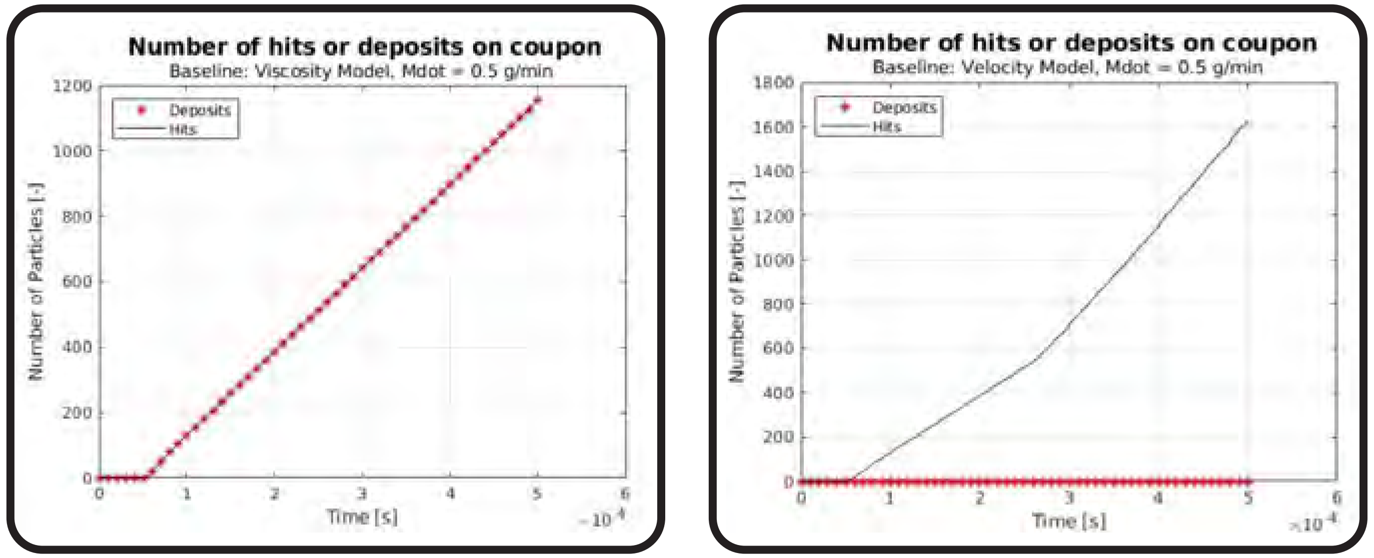


Figure 2A and 2B:
Particle temperature greater than softening temperature

Application of CinemaScience Methods to Visualize and Analyze Complex Multiphase Flow Processes

Sean Leonard,
Old Dominion University

Project Description

The volume of data generated by Athena and other simulation software necessitates the use of in-situ visualization due to the very large amount of data written to disk for post-processing analysis. Typically, the visualization step is handled post-simulation but for in-situ data analysis, visualization occurs intra-simulation eliminating the need to write the intermediate results to disk. Thus, the development of a robust, portable in-situ visualization implementation is a necessity. In addition, an image-based visualization tool, CinemaScience, must be implemented for one of the two modes of in-situ visualization, batch mode. Live mode is highly interactive and allows the use of the graphical user interface while batch mode is the opposite. Batch mode merely feeds a list of steps into a visualization tool (Paraview, in this case) to be carried out at each time step. As a result, successful implementation of in-situ visualization through the use of Paraview Catalyst and the add on of CinemaScience culminates in a viable big data solution for large scale simulations.

Relevance of Work to DOD

Sand ingestion is a substantial problem for gas turbine engines operating in austere environments. If an engine is subjected to high doses of calcium-magnesium-alumino-silicate (CMAS), the constituents of sand, the result can be degradation of the thermal barrier coating, reducing component life and requiring costly maintenance. Our proposal allows analysts and designers to more rapidly analyze the massive data sets produced by physics-based computational simulations.

Computational Approach

Data has moved from residing on dusty pages or archival CDs in a storage room to being feverishly produced and devoured online with a complementary hunger. This shift is attributable to the proliferation of digital devices and associated technical improvements enabling improved scalability, such as High-Performance Computing (HPC). The resultant volume, velocity and variety of data is an emerging field termed Big Data. Data visualization is one approach for addressing Big Data, exploiting the human eye as a powerful device that can immediately tune into inconsistencies and patterns. We examined two visualization tools, VTK with its companion Paraview, and VisIt; they offer similar functionality. Both provide a module for in-situ visualization, Catalyst for Paraview and Libsim for VisIt, and have options for batch and live mode. Batch mode is non-interactive and script-driven while live mode is highly interactive and user-driven. In-situ visualization provides access to information generated by the simulation on the fly. Scientists have access to results from simulations almost immediately. In-situ visualization is becoming a vital means of responsibly applying limited HPC resources and improving end-user productivity.

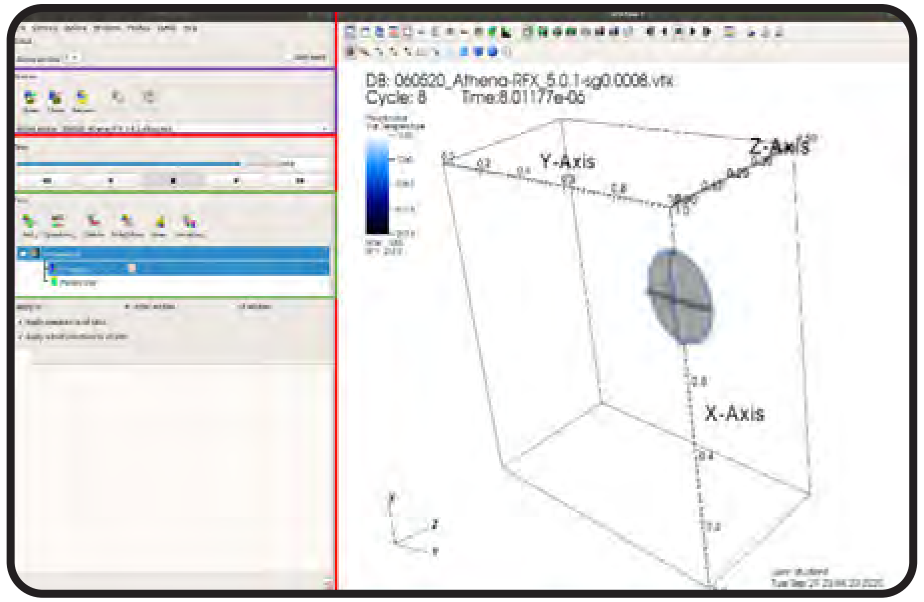


Figure 1:
VisIt GUI on startup with color boxes.
Red (Navigation),
Green (Coloring),
Blue (Filters),
Purple (Data Handling)

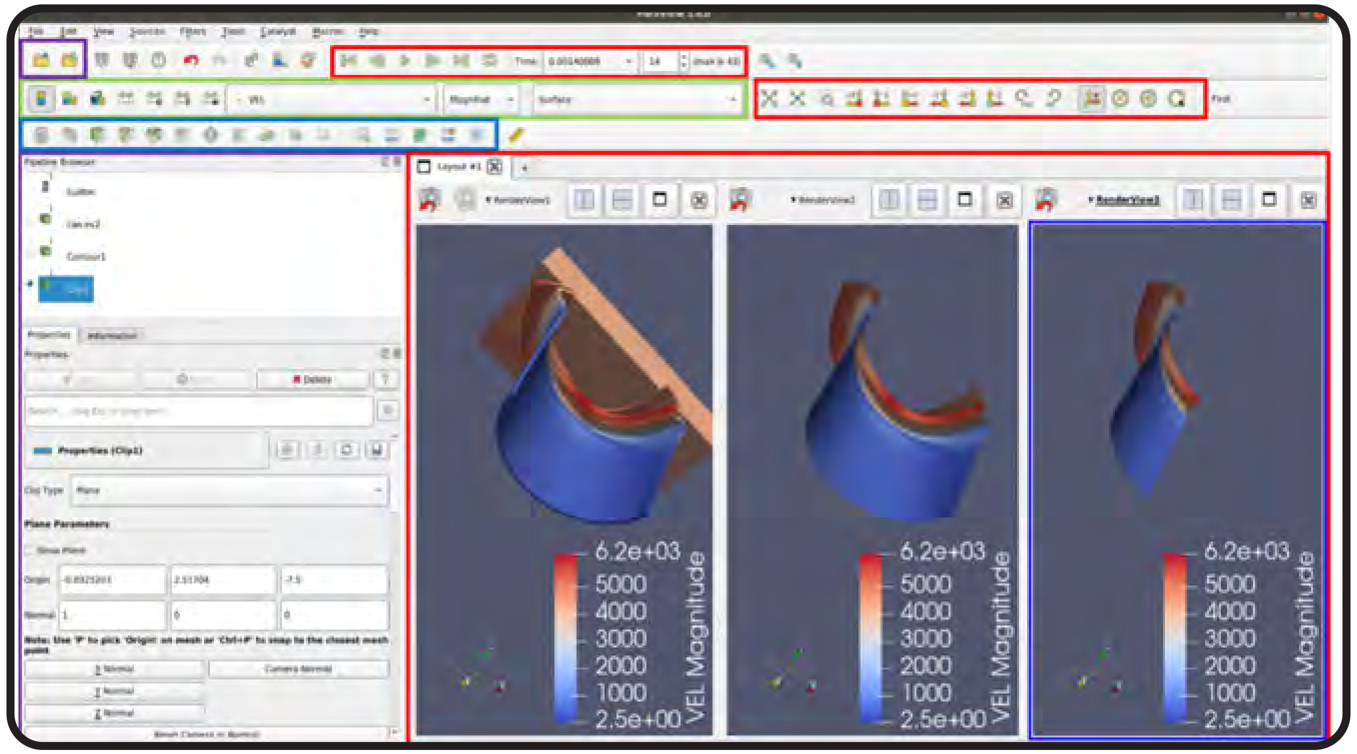


Figure 2:
Paraview GUI on startup with color boxes. Red (Navigation), Green (Coloring), Blue (Filters), Purple (Data Handling)

Simulations can produce a wide variety of data, such as statistics, 3-D field variables, images, polygonal data, and time series data. Working with images strains the capabilities of both VisIt and Paraview, for example, there may be thousands or tens of thousands of images to process. CinemaScience was developed to process a database of images using new image sampling methods. CinemaScience's interchangeable modules allow various visualization techniques to be used, enabling a user to tailor CinemaScience to their needs.

Results

VisIt and Paraview's functionality is composed of four main classes (Navigation, Coloring, Filters, Data Handling). These classes and their locations in their interfaces are displayed in Figure 1 for VisIt and Figure 2 for Paraview. CinemaScience organizes the images in an intuitive way and then assembles them. It can then slice and dice the assembled image in any way that the user requires, and allows the image to be viewed from any angle. However, the real power of CinemaScience is the in-situ batch mode, where traditional visualization tools are crippled by the large volume of data required for creating a complete "big picture" representation of a simulation.

Future

Libsim has been implemented in Athena by our co-investigators at Texas A&M University. Next steps include implementing Paraview in Athena and connecting both with CinemaScience to reduce the load on users and eliminate the need for users to become expert in all 3 visualization tools.

HIP Mentors

Dr. Simon Su *(National Institute of Standards & Technology formerly of the DEVCOM ARL DOD Supercomputing Resource Center)*
Dr. Luis Bravo *(DEVCOM ARL Weapons & Materials Research Directorate)*

Co-Investigators

Dr. Anindya Ghoshal & Dr. Muthuvel Murugan *(DEVCOM ARL DOD Supercomputing Resource Center)*
Dr. Alison Flatau *(University of Maryland)*
Dr. Zhanping Liu *(Old Dominion University)*

Publications

Sean Leonard, Zhanping Liu, Simon Su, Luis Bravo, Anindya Ghoshal, Muthuvel Murugan, and Alison Flatau, "Application of CinemaScience Methods to Visualize and Analyze Complex Multiphase Flow Processes Supporting Next Generation Future Vertical Lift (FVL) Concepts," The ITEA Journal, 42 (3), September 2021, pp. 153-163

Nonlinear Response of a Coaxial Rotor to Blade Pitch Change

Patrick Mortimer,
University of Texas at Austin

Project Description

There were two objectives of the project: investigate the nonlinear response of rotor inflow when subjected to time-varying changes in blade pitch, and serve as a learning experience in developing and executing computational fluid dynamics (CFD) simulations. Using the CREATE CFD software, Helios, the flow field around a coaxial, counter-rotating rotor was simulated.

Two test cases were created. A baseline case where the collective blade pitch for the upper and lower rotor was set and held constant throughout the simulation, and a dynamic test case where the upper rotor was subjected to a change in collective of $\Delta\theta_0 = 3°$ over one rotor revolution. A Rotor Comprehensive Analysis System (RCAS) structural model was developed and coupled to the CFD model providing blade deflections for accurate rotor loads for both the baseline and dynamic test cases.

Relevance of Work to DOD

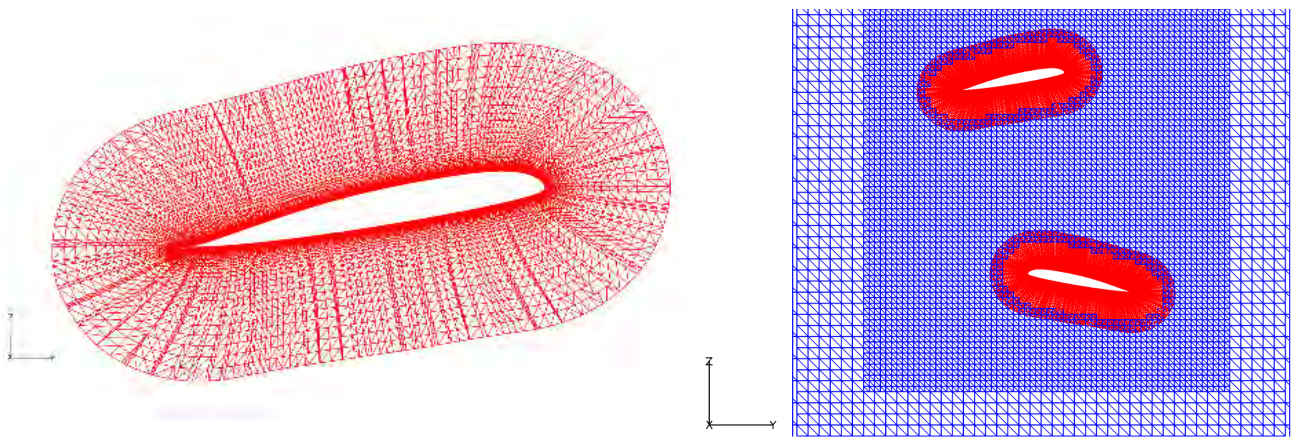
The accurate computation of the rotor inflow in hover is still a challenge for modern computational methods. The already high physical complexity of the aerodynamic environment in hover is further complicated when the transient response due to control inputs needs to be considered. A dynamic pitch input results in a sudden change in sectional angle of attack, which leads to a sudden change in sectional forces.

Development of accurate numerical models (i.e. design tools) affects a wide area of applications such as flight dynamics and control, and flight simulations for pilot training. Better inflow models lead to more maneuverable vehicle designs, which can help soldiers on the battlefield, especially in urban environments.

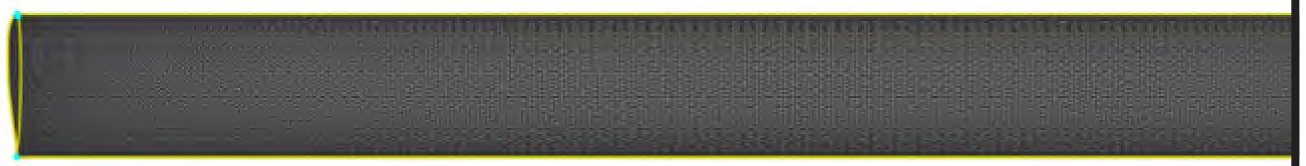
Computational Approach

Helios version 11.03 was employed to understand the complex aerodynamic interactions of a coaxial, counter-rotating rotor in hover subject to time-varying collective pitch inputs. An RCAS structural model was coupled to the CFD model providing blade deflections in order to achieve accurate rotor loads. Currently, a baseline test case at a blade loading (thrust condition) of $C_T/\sigma = 0.10$ has been simulated. An RCAS model of the dynamic pitch test case has been created, and coupled with the CFD model. As a first test case for dynamic inflow, the upper rotor will be subject to a collective pitch ramp resulting in a pitch increase of $\Delta\theta_0 = 3°$ over one rotor revolution.

Figure 1: Grids for CFD simulation



(a) Cross-section of near-body grid (red) and off-body grid (blue)



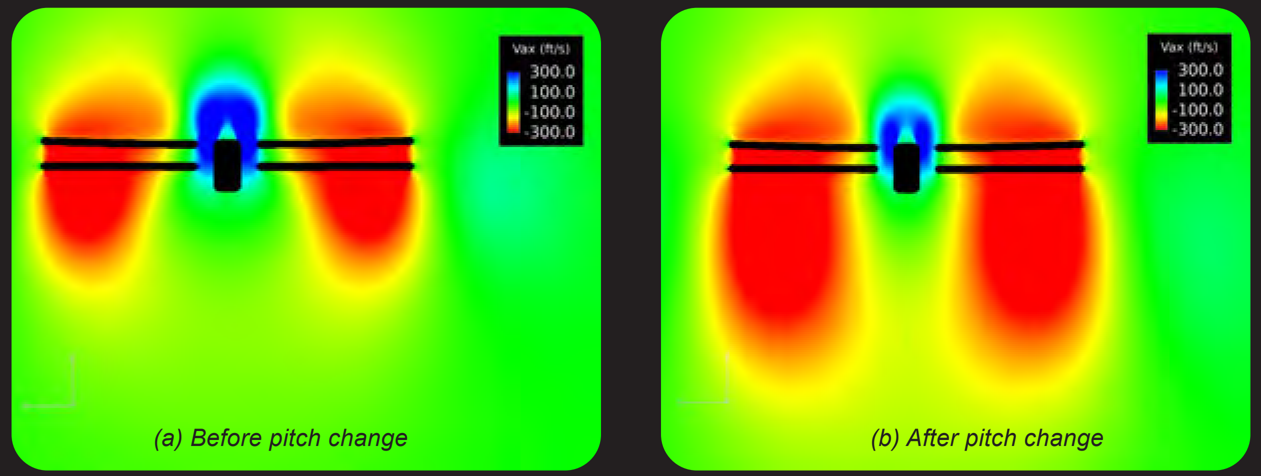
(b) Blade surface grid

Grids were created for the blades as well as a cylindrical body representing the hub and mounting equipment. CREATE Capstone was used to generate the surface mesh for both the blades and cylindrical body. We used 12.2 million total near-body nodes and approximately 200 million off-body nodes., see Figure 1. Adaptive mesh refinement was used, which refined the off-body grid according to the scaled Q-criteria scheme proposed by Kamkar et. al. The flow field around the rotor was solved using the Reynolds

Averaged Navier-Stokes equations with the Spalart-Allmaras turbulence model near the blade boundary and detached eddy simulation away from it. A time step size of 4.1653e-05 seconds corresponding to 0.225° was chosen to provide sufficient temporal resolution in order to capture the transient effects due to the blade pitch change. All simulations were run on Centennial at the ARL DSRC.

Results

The RCAS model for the dynamic pitch test case was investigated by looking at the commanded pitch change and the corresponding change in upper rotor thrust. The pitch ramp occurred over one rotor revolution and resulted in a change in rotor thrust of ~40 N. Blade deflections for the lower rotor did not appear to be accurate as there was little to no change in blade motion observed. Positive blade pitch inputs in the RCAS model resulted in a decrease in collective. The exact cause is currently being investigated.



(a) Before pitch change (b) After pitch change

Figure 2:
Instantaneous flow field contours of axial velocity (along rotor rotational axis)

Data from the CFD simulations were extracted at an azimuthal interval of ≈1° (four time-steps). Figure 2 shows instantaneous flow fields with contours of axial velocity (velocity component along the rotational axis of the rotor). Figure 2a shows the flow field before the executed pitch change; Figure 2b shows the flow field after the pitch change. An increase in the axial velocity can be clearly seen due to the increased thrust achieved from a larger collective pitch angle.

Future

The grid resolution using the adaptive mesh refinement was too low when the rotor blades were not located in the set slice for data extraction. This resulted in the tip vortices not being resolved. This has been remedied by using static mesh refinement in the blade region. Simulations need to be completed with the new mesh strategy. The second remaining issues pertains to the RCAS model. The lower rotor deflections are not accurate. Positive blade pitch inputs in the RCAS model result in a decrease in collective. The exact cause is currently being investigated, but is likely the cause of the problems pertaining to the lower rotor.

HIP-Mentor

George Jacobellis, PhD *(DEVCOM ARL Weapons & Materials Research Directorate)*

Publications

Patrick Mortimer and George Jacobellis, "Nonlinear Response of a Coaxial Rotor to Blade Pitch Change," The ITEA Journal, 43 (1), March 2022, accepted for publication

Distributed Processing and Analysis for Large-Scale Heterogeneous Data: Text Classification and Analysis through Clustering

Mariya Occorso,
Arcadia University

Jacob LeBlanc,
Stanford University

Project Description

The Army Test and Evaluation (T&E) Command generates massive amounts of heterogeneous systems data during test events, requiring analysts to process, evaluate, and report on that data. For the past decade the T&E community has leveraged HPC resources for their data reduction and analysis efforts; however, as data volumes increase, more useful analysis tools are needed to evaluate the data. With test programs producing thousands of test incident reports (TIRs) per system under test, evaluators spend substantial time reviewing and scoring the incidents to determine system readiness.

Relevance of Work to DOD

Every program of record undergoes test and evaluation. For every item tested, one ore more test incident reports are generated to document everything done to and with the system, from routine oil changes and replacing wiper blades on a truck to fires, explosions and rollover events. For large acquisition programs that last for decades, thousands to tens of thousands of TIRs can be generated and an evaluator must read through all of them to score them for relevance to a specific evaluation need. Our goal is to expedite the evaluation phase by integrating machine learning into the data analysis and scoring process for evaluators.

Computational Approach

The TIR text data contained 85,906 unique words, each of which represents a distinct piece of information. We preprocessed the data using stemmers and lemmatizers and tokenization to simplify the task, creating fewer unique words and less complex data for our algorithms. We then converted the documents to number vectors before employing machine learning algorithms. To reduce the dimension of this resulting vector space, we used Principal Component Analysis (PCA) to determine which features were most important as well as which features could be combined.

We applied density based clustering algorithms, a form of unsupervised learning, to the TIR data. The two clustering algorithms that we chose from this category were Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Ordering Points To Identify Cluster Structure (OPTICS). We also examined centroid based clustering algorithms, K-Means and Mini Batch K-Means, which determine clusters by a central vector. These algorithms require the user to decide the K value, which is the number of clusters. To find the optimal K value, we employed the elbow method. Figure 1 is the output of DBSCAN on a 2 dimensional scatter plot showing over 100 clusters, along with the outliers. Figure 2 is a three dimensional plot of clusters obtained using K-Means.

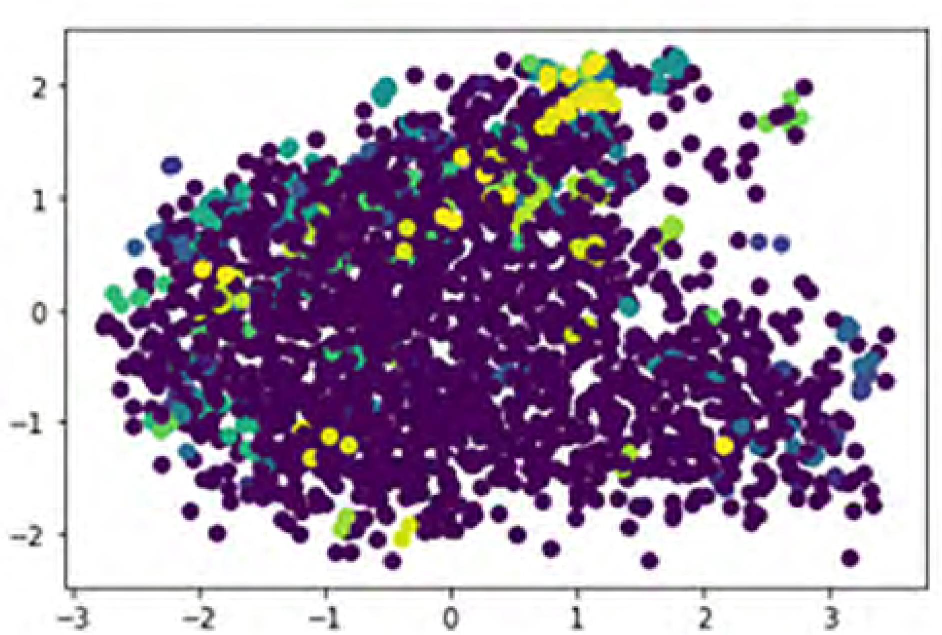


Figure 1:
Two-dimensional clustering of data using DBSCAN

Results

We chose silhouette scores and overall usability as metrics to determine clustering algorithms performance. After running several iterations of the density based clustering algorithms, we found that they scored well for silhouette scores and poorly for usability.

In particular DBSCAN resulted in an average silhouette score of .713, but had over 100 clusters and 1,800 outliers. With the number of outliers ranging from 100 to 2,000 it was clear that DBSCAN did not capture all of the data. The centroid based clustering algorithms both tended to have relatively lower silhouette scores compared to the density based algorithms. K-Means had an average silhouette score of .27 and Mini Batch K-Means had an average silhouette score of .25.

We found the centroid based clustering algorithms outperformed both density based algorithms in terms of usability. After applying the elbow method with our data points, we found that the optimal K value was thirteen. Refining our tokenizing process did not significantly improve silhouette scores, it did improve the most representative terms per cluster list. Overall, we found that centroid based clustering algorithms performed best for our project.

We tried several classification algorithms in creating a model for the TIRs. We tested each algorithm with a training/test model using k-folds. Each of these folds produced precision, recall and f-1 scores as a measure of performance. We focused on recall because our data had already been classified by experts. Preliminary results led us to dismiss the random forest classifier and Gaussian Naive Bayes. Neither performed well due to the overlap between classifications. On the other hand, we continued pursuing both the SVC and MLPClassifier algorithms. Overall precision and recall were best when using all of the data with c values greater than or equal to 5 and dimensions greater than or equal to 150. Figure 3 show the classification report and confusion matrix for the results at c = 5 and the data scaled to 50 dimension.

Future

We want to add more data to our models. Machine learning is very reliant on the amount and quality of data; it can only benefit with more data. The end goal of this project is to create a front-end web application where a user can access these models to analyze new data.

HIP-Mentors

Jamie Infantolino *(DEVCOM ARL Computational & Informational Sciences Directorate)*
Vincent Perry *(DEVCOM ARL DOD Supercomputing Resource Center)*

Publications

Mariya Occorso, Jacob LeBlanc, Jamie Infantolino, and Vincent Perry, “Distributed Processing and Analysis for Large-Scale Heterogeneous Data: Text Classification and Analysis through Clustering,” The ITEA Journal, 42 (4), December 2021, pp. 262-270

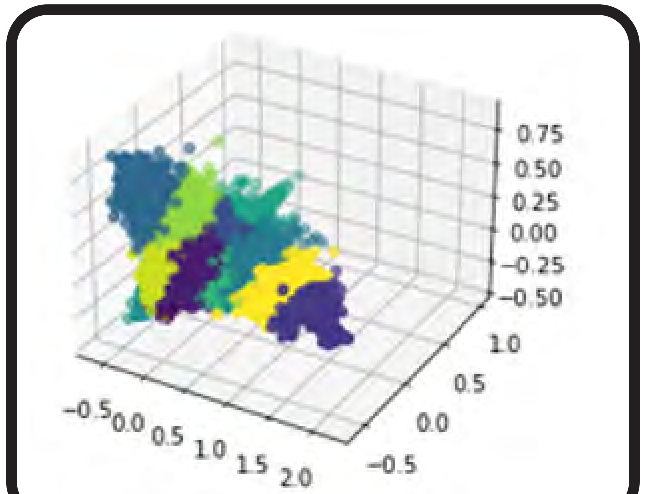


Figure 2:
Three dimensional graph of K-Means clustering with 13 clusters.

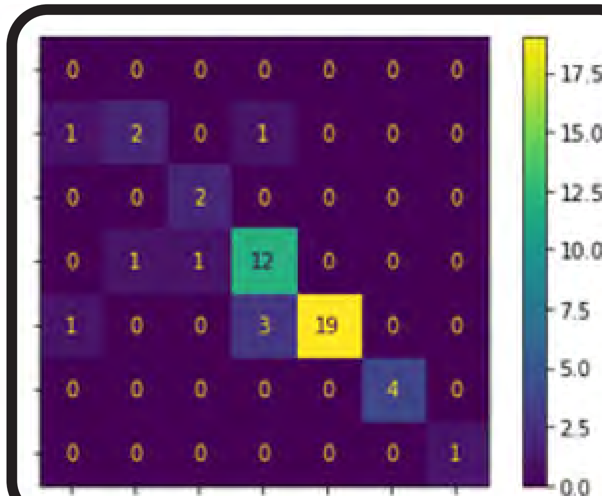


Figure 3:
The confusion matrix for SVC with c = 5 and data scaled to 50 dimensions. Data on the diagonal from the top left to the bottom right are considered to be classified correctly and other points are not.

Broadband Noise Prediction from CFD Wall Spectrum

Nikos P. Trembois,
University of California, Davis

Project Description

The objective is to assess the ability to predict self-noise of rotorcraft using wall pressure spectrum models. This is a new need that has emerged from the downscaling of rotors on rotorcraft, partly due to new battery technology that has expanded the missions rotorcraft can carry out. We used high-fidelity computational fluid dynamics (CFD) simulations to evaluate the effect of 3D rotorcraft flow on the blade boundary layer, thus self-noise production. A higher resolution grid is required to accurately capture the boundary layer to the degree required for acoustic assessment, leading to a grid study to find the appropriate balance between acoustic accuracy and computational time. It was found that the boundary layer parameters can be successfully extracted from the CFD simulation and used with a wall pressure spectrum model, namely the Lee model, to predict the self-noise of a rotor blade.

Relevance of Work to DOD

Despite rotorcraft's unique ability to hover efficiently, as well as take off and land vertically, rotorcraft noise needs to be curbed to fully realize its capabilities. New rotorcraft designs reduce rotor radii, tip speed, and disk loading by distributing loads throughout multiple rotors or stacking rotors. Additionally, smaller scale vehicles are in development to fill new roles like resupply and reconnaissance. Modern technology, including light weight and efficient battery storage, enable these new designs. However, the acoustics of these configurations still need to be understood to increase their viability.

Computational Approach

Rotors with reduced disk loading powered by electric motors have a distinctively different sound signature when compared to a conventional helicopter. In particular, self-noise is no longer masked by other sound sources and becomes audible. Unfortunately, airfoil self-noise research has largely been overlooked due to the dominance of tonal noise in helicopters and other common rotorcraft designs. However, a large influx of research in rotorcraft noise reflects the emerging requirements of rotorcraft aeroacoustics. One need that has not been met is predicting self-noise with high fidelity simulations.

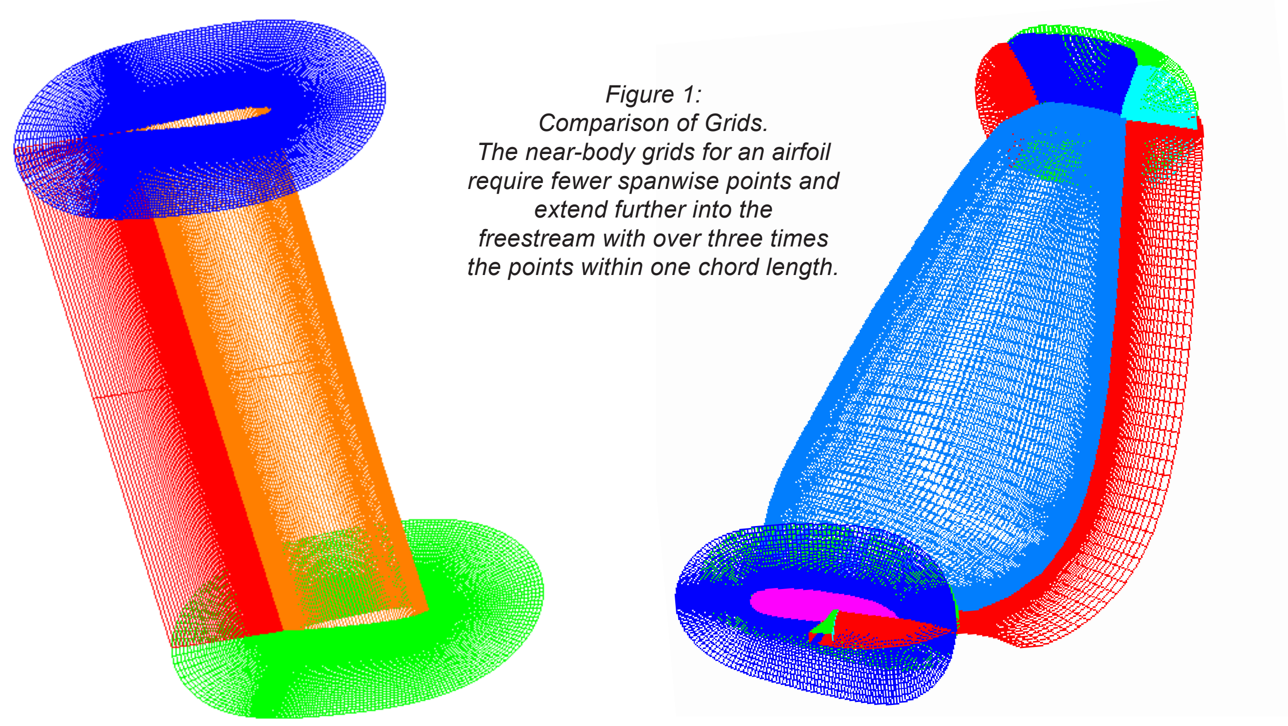


*Figure 1:
Comparison of Grids.
The near-body grids for an airfoil require fewer spanwise points and extend further into the freestream with over three times the points within one chord length.*

A fine near-body grid is required to resolve the boundary layer for the acoustic parameters and the resolution of the off-body grids needs to be fine enough to capture 3D flow characteristics like tip vortices and wakes that interact with and affect the boundary layer, see Figure 1.

We utilize a full 3D Reynolds Averaged Navier Stokes rotorcraft simulation to predict self-noise of rotorcraft blades. A fine near-body grid is required to resolve the boundary layer for the acoustic parameters and the resolution of the off-body grids needs to be fine enough to capture 3D flow characteristics like tip vortices and wakes that interact with and affect the boundary layer. The result is an acoustic prediction that achieves higher accuracy than 2D or semi-empirical flow models and includes the effects of 3D flow structures.

The simulation was carried out using Helios. The near-body grid was solved with OVERFLOW and the off-body grid solver was SAMCART. The boundary layer parameters at many radial positions were used as inputs for Brooks, Pope, and Marcolini acoustic models to calculate the self-noise generated at each radial section, then the total self-noise for the blade. The acoustic model uses the wall pressure spectrum to calculate the acoustic frequencies. The range of frequencies calculated are from 100 to 10,000 Hz as these are the frequencies humans are most sensitive to. The boundary layer properties of interest are freestream velocity, turbulence time scale, boundary layer thickness, displacement thickness, momentum thickness, maximum friction coefficient, and pressure gradient in the chordwise direction.

Results

The necessary grid statistics were determined to accurately capture boundary layer parameters. Then, with new grids, the acoustics were computed. The grid study indicated that 168 volume grids were necessary for accurate simulation of aeroacoustics. This is a 158% increase in points when compared to a rotorcraft grid and a 20% decrease when compared to airfoil grids. However, since this reflects changes in near-body grid points, while the off-body grids are unchanged, this only results in a 12% increase in computational time per solver step. Once the sections were extracted with their boundary layer parameters, the acoustics were calculated using the Lee model, which is a wall pressure spectrum model based on the Brooks, Pope and Marcolini model. The acoustics of a radial section is shown in Figure 2. So far, only sectional acoustic noise has been calculated and total noise will be calculated in future work.

Future

Acoustic results are just the beginning and further study is required to fully validate this method. While the sectional acoustics were calculated, the acoustics of the whole blade need to be calculated and validated against experimental results. The total noise will be calculated with UCD-Quietfly. Once the total noise is calculated the different turbulence models and simulations parameters will be varied to determine the ideal setup for small-scale rotor self-noise prediction.

HIP Mentors

George Jacobellis, PhD *(DEVCOM ARL Weapons & Materials Research Directorate)*

Publications

Nikos P. Trembois George Jacobellis, "Broadband Noise Prediction from CFD Wall Spectrum," The ITEA Journal, 43 (1), March 2022, accepted for publication

*Figure 2:
Lee Model Self-Noise Prediction.*

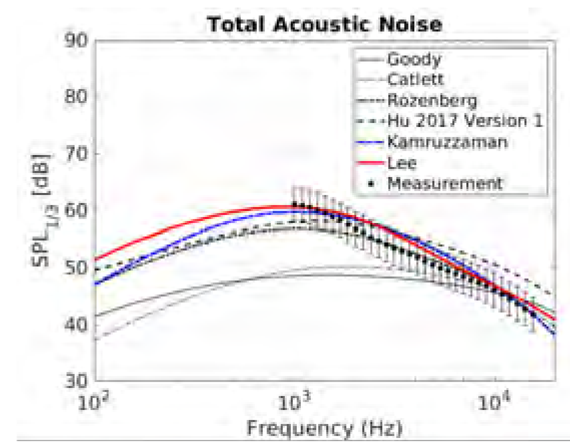
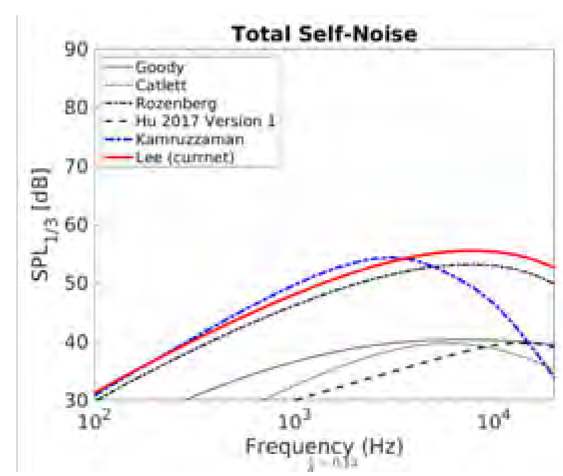


*Figure 2a (above)
compares Lee Model with experimentation and other wall pressure spectrum models for an airfoil with an angle of attack.*

*Figure 2b (below)
shows the self-noise of a Hitec SUI endurance rotor blade section.*



Visual Analytic Application Toolset for Large-Scale Data

Mariya Occorso,
Arcadia University

Aiden Kenny,
Columbia University

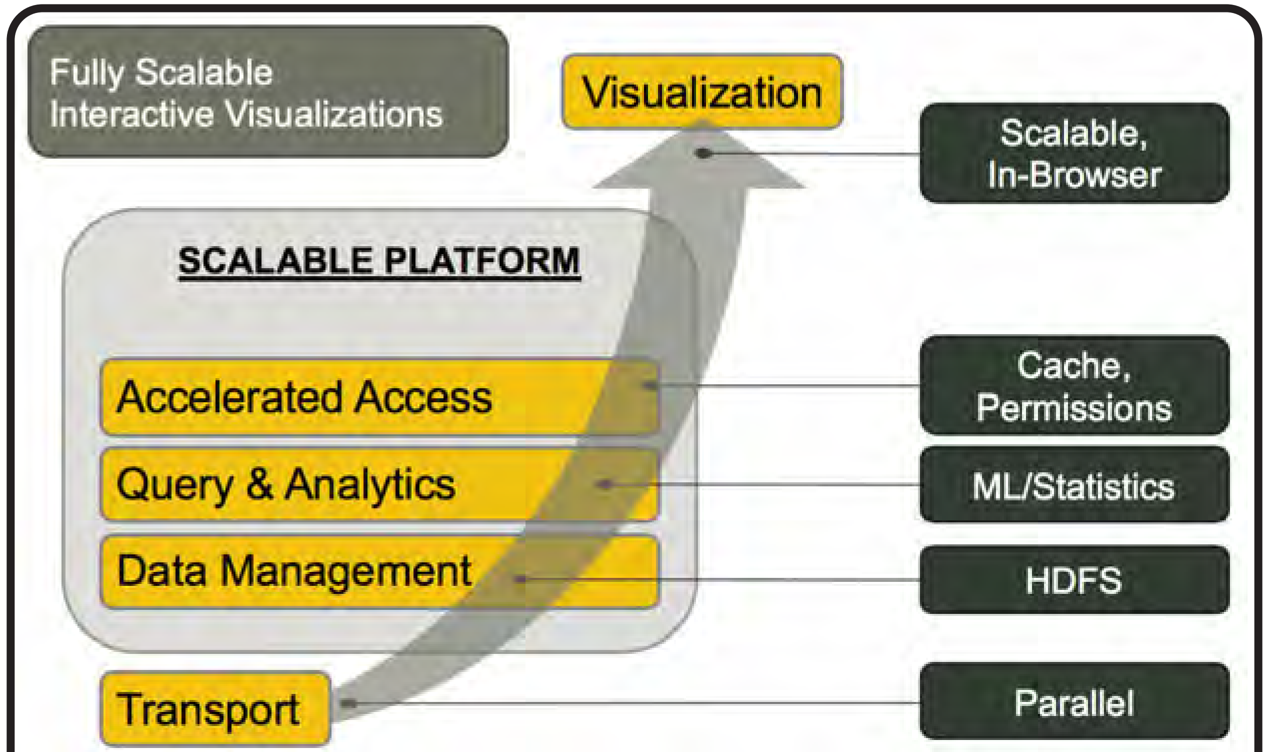


Figure 1:
Overview of the dashboard application and all of its components

Project Description

We built a framework for an interactive user dashboard to help analyze large-scale data sets, with specific emphasis on creating interactive visuals. We focused our dashboard on data collected from the Army's Expedient Leader Follower program, which focuses on the development of a convoy of autonomous vehicles. The leader vehicle in the convoy is driven by a soldier, while the rest of the convoy is autonomous. We wrote code in R that automatically accesses the database, cleans the data, and returns useful summary statistics. We also wrote code in Python that builds the actual dashboard itself, accepts the cleaned data and produces interactive visualizations.

Relevance of Work to DOD

Our goal was to create an interactive user dashboard to help data analysts easily and quickly explore test data. We sought to create a system to automatically connect to a PostgreSQL database, perform statistical analysis and machine learning on demand, and present the desired results, be it tables or interactive visualizations, on the dashboard. The impact will be a reduction in the time and labor to perform the analyses, and reduction in possible human error.

Computational Approach

A conceptual design of the dashboard application is displayed in Figure 1 showing all of the components to be utilized. We focused on the query & analytics portion of the dashboard using R, and on the visualization portion, building the dashboard in Python.

There were two main sources of data for the project. The first was the instrumental data collected during the various ExLF trials throughout 2019 and stored in a PostgreSQL database on Centennial. This database contains trial results from 134 separate days of testing, amounting to over 20 terabytes of data. The second source of data was from test incident reports (TIRs). Here, any failures during the trials observed by the testers were manually recorded and stored as a collection of PDF files.

We limited the scope to a single day of data, specifically to all of the tests from 23 July 2019. We chose this day specifically because it had the most TIR data available. Within this database there were seven different schemas of data, 547 tables of data, and 5,355 variables. We decided to further limit our scope to a single schema that contained 29 tables, and from those we selected a single table, labeled "comms analysis".

We set up a container to run the latest version of R via a Jupyter notebook and to connect to the PostgreSQL database; the Jupyter notebook contained the query and analytics code. The first script connects to the PostgreSQL database and reads in the data. It then calculates summary statistics for each of the numeric variables in the table for each vehicle in a given trial. A second script allows the user to specify which variables and trials they want to examine. The output of these two scripts becomes input for the dashboard and is used to make an interactive visualization for the specified variable for a given trial. Following the work with the R-Scripts, the dashboard application was created with user interactivity and adaptability in mind.

Results

The final result is a working prototype of the user dashboard having two tabs, a data overview tab and a data analysis tab. Screenshots of the data overview tab are displayed in Figure 2. The top left of the screen allows the user to select which trial they wish to investigate, and then further select which vehicles they want to analyze from that specific trial. Finally, they choose which variables are of interest to them, and several interactive plots are generated.
The first plots each of the selected variables over time; this allows the user to compare the data for several variables at once. The three plots toward the bottom provide more summary statistics about the selected vehicles. The data display tab is arranged analogously. The user selects variables to analyze and a plot of the time and magnitude of the TIRs for each variable is created. This shows where and how frequently an error occurred for each variable. Fails are plotted along with each of the variables to show where the errors occurred over time.

We have laid the groundwork for an effective user dashboard for visualizing data from the Army's ExLF program trials. As of now, the user is able to specify which trial they are interested in analyzing, and the app will generate interactive visualizations that will help with spotting anomalies or interesting patterns in the data. The user could then follow up and further examine the specified trial, specifically the TIRs, to see if a significant fault was observed during the trial.

Future

The next step is to combine the instrumented data with the TIRs. This will allow us to perform supervised learning and make inferences about which variables are significant in causing a trial to fail. We would also like to determine if unobserved errors occurred. We would like to containerize the entire project ecosystem so that all of the PostgreSQL, R, and Python environments can be run together using a single point of execution.

HIP-Mentors

Vincent Perry *(ARL DEVCOM DOD Supercomputing Resource Center)*

Publications

Aiden Kenny, Mariya Occorso, and Vincent Perry, "Visual Analytic Application Toolset for Large-Scale Data," The ITEA Journal, 42 (2), June 2021, pp. 72-76

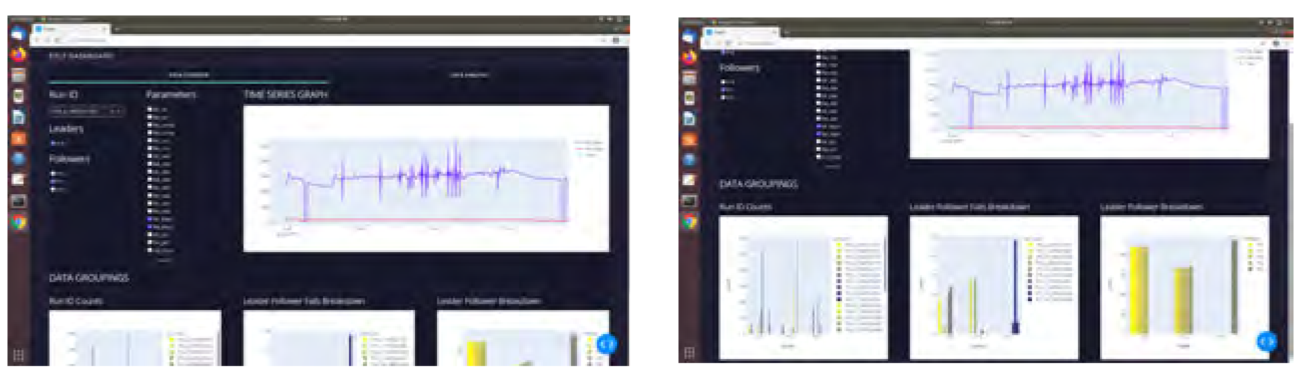


Figure 2:
The data overview tab of the dashboard

Topic Modeling and Visualization for Test Incident Report Data

Sage Leone,
University of Maryland, College Park

Project Description
Testing at the U.S. Army Aberdeen Testing Center (ATC) is producing data at an increasing volume, velocity, and variety. In order to handle the data, less labor intensive approaches to analysis need to be implemented. This research leverages the big data resources available at the Army Research Laboratory (ARL) DSRC to create an automated classification system for data records. Using Apache Spark's MLlib and SQL libraries, a topic model using Latent Dirichlet Allocation was developed to quickly and correctly categorize Test Incident Reports from vehicle testing. These results were compiled into clear visualizations, allowing trends to be quickly observed and used to directly improve the testing process. Building on previous ARL research, this topic model further reduces the time and labor spent on data analysis.

*Figure 1:
An Army vehicle during survivability/lethality tests at Aberdeen Test Center*

Relevance of Work to DOD
As a lead testing facility for the Department of Defense, the US Army Aberdeen Test Center collects massive amounts of data that must be compiled, reduced and analyzed. In the past, analysis was done manually, which is a costly and time-intensive process. As the volume and velocity of data continue to increase, automated analysis methods are being implemented to improve the process. Previous research at ARL utilized the power of High Performance Computing (HPC) resources to automate data visualization. Combining the processing power of HPC with big data technologies provides more tools for automated analysis, further reduces the need for human processing, and provides data to decision makes sooner, for example, from live fire test events, Figure 1.

Computational Approach
The purpose of this research is to demonstrate the validity of using Spark MLib and SQL libraries on a big data stack to produce a topic model for classification of army data. Topic modeling is an unsupervised classification approach to automatically derive topics from a corpus of text documents. The Latent Dirichlet Allocation (LDA) was chosen. It assumes each document can be described by a probability distribution of topics. Similarly, it describes each topic as a probability distribution of words. The goal was to reproduce a chart of test results that previously required weeks of work by analysts to produce, and to classify and graph TIRs by topic.

Apache Spark uses a master-worker architecture to distribute work over multiple nodes on the HPC. Spark has several associated libraries to provide further data analysis capabilities, including Spark MLlib and Spark SQL. Spark MLlib is a scalable machine learning library providing machine learning algorithms able to run at speed on large volumes of data. Spark SQL is a module for working with structured data, compatible with structured query language. It also provides access to Hive data warehouses and tables.

TIRs contain information specific to each test, descriptions of occurrences during testing, and other general information. Topic modeling on these texts was used to categorize the reports based on similar problems and methods of resolution. This work utilized a five-month period of vehicle TIR data running from June 2019 to October 2019 and containing around 2400 records. First, the TIR data was read from the csv files stored on an HPC into a Spark DataFrame. Text cleaning functions were applied to the narratives column of the DataFrame. Punctuation, special characters, and stopwords were removed and abbreviations expanded.

The text was also lemmatized, reducing each word to its base form. The text was tokenized, ngrams were formed, and term frequency vectors were created. Finally, the vectors were rescaled based on term importance—measured by frequency of occurrence in the corpus—to create the term frequency inverse document frequency matrix for the topic model algorithm. Spark provides an implementation of LDA topic modeling in the pyspark.ml python library.

Results
Topics were extracted from the document using the LDA algorithm and defined in probabilities of words. Each TIR was described by a combination of topics and related percentages, then matched with the topic with the highest percentage. From this data we generated a display of the total count of documents per topic, Figure 2. This displays the most common cause of problems to be determined at a glance, allowing feedback on the testing to be received and integrated with little delay. A related figure provides the same information as a function of time, allowing focus on specific periods and changes that occurred throughout the testing process. These plots are interactive, including display of more data and flexibility in selection of data traces.

The volume of data produced by testing at ATC requires a more time- and resource-efficient way to provide evaluation and classification of data. While previous research has been conducted on the use of HPC systems to handle processing of large quantities of data, the machine learning and SQL capabilities present with HDFS and a big data stack have been largely unemployed. Using LLNL's Magpie allowed a big data stack and accompanying projects to be run on any general parallel file system, the typical file system of an HPC machine. This project demonstrated the validity of leveraging Apache Spark to develop a topic model for quick classification of test data, reducing the time between testing and return of analysis.

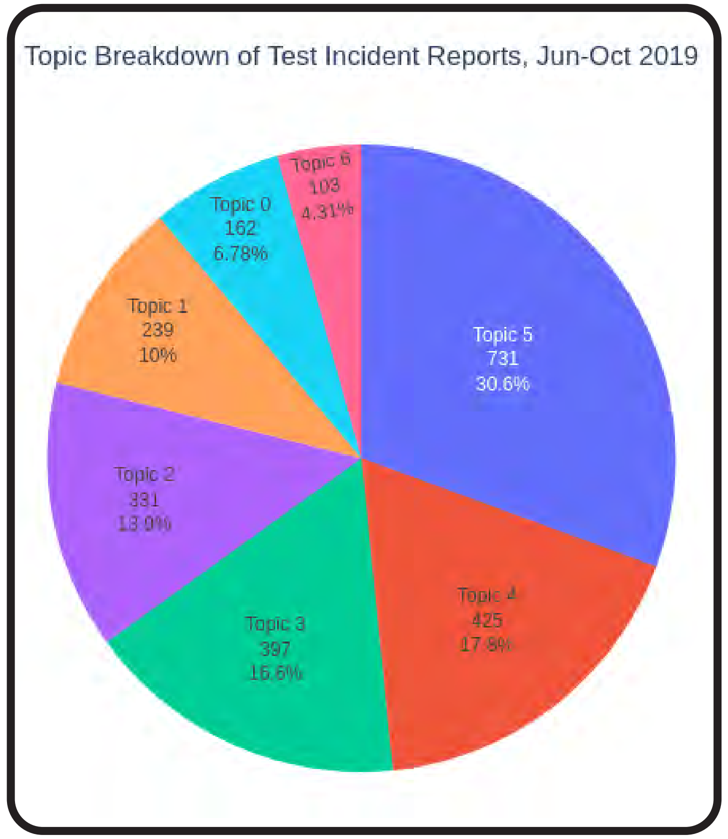


*Figure 2:
The percentage of documents in each topic of the entire dataset.*

Future
Generalization of this process for a variety of datasets would improve analysis and testing in other areas. An expansion of the text cleaning process to deal with a greater variety and inconstancy in data would permit the spread of the usefulness of this development.

HIP Mentor
Jamie Infantolino *(DEVCOM ARL Computational & Informational Sciences Directorate)*

Co-Investigators
Adam Childs & Cleon Anderson *(DEVCOM ARL Computational & Informational Sciences Directorate & Parsons)*

Publications
Sage Leone, Jamie Infantolino, Adam Childs, and Cleon Anderson, "Topic Modeling and Visualization for Test Incident Report Data," The ITEA Journal, 42 (2), June 2021, pp. 77-82

Immersive Environments for Visual Analytics

James Hughes,
University of Maryland, College Park

Selena Hamilton,
Harford Community College

Project Description

The US Army Aberdeen Test Center has been in the business of collecting test data on Aberdeen Proving Ground for more than 100 years and possesses a world recognized expertise. Managing and analyzing the magnitude of collected data today presents new challenges. The Army Research Laboratory DoD Supercomputing Resource Center in collaboration with the Aberdeen Test Center and its higher headquarters, the Army Test and Evaluation Command, is applying high performance computing and data analytics to test and evaluation. This project successfully demonstrated the use of immersive environments to render the movement of Expedient Leader Follower (ExLF) program vehicles under test.

Relevance of Work to DOD

During resupply convoys and reconnaissance missions, conditions such as long sustainment, poor weather or night operations adversely affect the safety and the situational awareness of the Soldier driver. The U.S. Army's ExLF program removes soldiers from the vehicle while operating in highly contested areas. However, the Army still wants the lead vehicle in the convoy to have a human driver, at least at initially. The unmanned Follower trucks need to stay on course without the benefit of street signs, lane markings, pavement, or GPS. The Follower may not have a clear line of sight to the Leader vehicle; it might have turned a corner or be hidden by a tree or building. In addition to avoiding other vehicles, animals, and structures, Follower vehicles must also avoid rocks, holes and other rubble. The instrumentation and programming of these autonomous vehicles is challenging; therefore, testing the vehicles requires gathering data on all vehicle instrumentation and sensors as well as any external sensors to monitor the overall performance of the vehicles and ensure safety for people and property.

Computational Approach

We began with data for the terrain map and an extremely large comma separated variable (CSV) file containing ExLF test data. The CSV file contains several test parameters including test number, Universal Transverse Mercator (UTM) data, GPS data, and vehicle identifiers. We use the cross-platform game engine, Unity, to create three-dimensional, virtual reality environments and provide the user the ability to quickly create 3D experiences.
We first set up the viewing area of the 3D environment in Unity. This required the track terrain piece and correct track path, as well as creating the truck asset, Figures 1 and 2.

A script file reads the CSV files to get position for the trucks in simulation. The script also pulls several public variables, the starting Current Point Index (CPI), the CPI Increase, and the Time Multiplier. The starting CPI is used to determine the start point in the data for each truck. The CPI Increase determines how many data points are jumped for each frame of the simulation by adding the number to the CPI at the end of every frame. Lastly, the Time Multiplier multiplies Unity's Engine Clock output by the given number, forcing the program to move through data points faster. This gives higher end machines the option to run the simulation faster than real time with the same amount of detail.

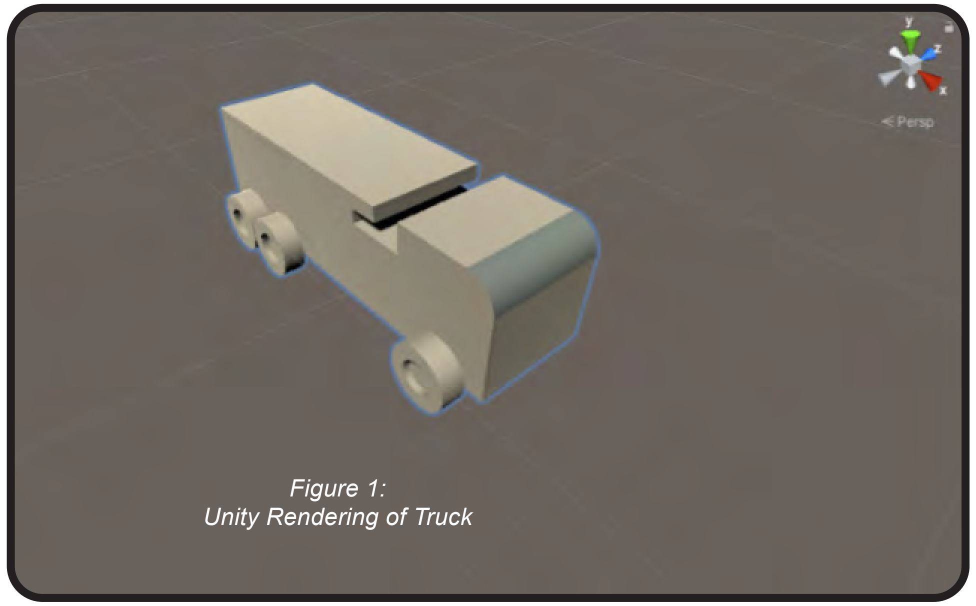


Figure 1:
Unity Rendering of Truck

Figure 2:
Unity View of Terrain Map & Test Track

Results

The simulation has successfully run for two trucks operating simultaneously. Each truck has its own respective set of settings. More trucks can be added to the simulation as necessary, although each truck will need to be moved to its proper starting point according to its data set. Each truck will also need it's starting CPI calculated so that truck start time matches up with runtime. Overall, this project achieved its goal of creating a simulated environment for the trucks to perform in using the given test track assets. Using this simulation as a basis, there is now a reliable way to visualize the ExLF test data on the ATC test track. Further Unity simulation engine work will be required to input different tracks, however the scripting device employed can be transferred to those other tests.

Future

Future work will involve automating the calculation of the starting CPI for the simulation, which currently needs to be calculated by the user in order to ensure the simultaneous start of the trucks in simulation. This could be resolved through a separate script that pulls data then assigns the correct starting CPI to each of the trucks based on the latest runtime number. Another improvement would be more changeable and moveable camera options in simulation. While the top-down camera view locked to the lead truck is adequate, having multiple camera angles to use and have the ability to turn would greatly increase the ability to spot flaws in the truck's movement.

HIP Mentors

Virginia To *(DEVCOM ARL DOD Supercomputing Resource Center & Parsons)*

Publications

James Hughes, Selena Hamilton, and Virginia To, "Immersive Environments for Visual Analytics," The ITEA Journal, 42 (2), June 2021, pp. 68-71

How to Request an Account

https://centers.hpc.mil/users/index.html#accounts

Who Can Use DoD HPC?

- DOD researchers and engineers
- DOD contractors
- University staff working on a DOD Research Grant. (Non-US Citizens may not have access to resources)
- Accounts are maintained through the Portal to the Information Environment (pIE) and managed by the Service/ Agency Approval Authorities (S/AAAs)

To Get Started:

(You will need to know the following items to complete your pIE account application)

- Citizenship Status
- Preferred Kerberos Realm
- Organizational ID
- Name, Title and Position
- Mailing address (no PO Box), Company Name, Phone, Fax and Email address
- Preferred User name
- Government Employee Status
- Who is your S/AAA?

Contact Information

More detailed information is available at

High Performance Computing Modernization Program (HPCMP)

Website: https://centers.hpc.mil/users/
HPCMP Help Desk Email: help@helpdesk.hpc.mil
HPCMP Help Desk Phone: 1.877.222.2039
Accounts Email: accounts@helpdesk.hpc.mil

DEVCOM Army Research Laboratory (DEVCOM ARL), DEVCOM Army Research Laboratory DOD Supercomputing Resource Center (ARL DSRC)

Website: https://www.arl.hpc.mil/
Email: outreach@arl.hpc.mil
User Services Phone: 1.800.275.1552