AFRL-AFOSR-VA-TR-2020-0157

Dimensional Reduction of Highly Nonlinear Multiscale Models Using Most Appropriate Local Reduced-Order Bases

Charbel Farhat
LELAND STANFORD JUNIOR UNIVERSITY

08/15/2020
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTA2
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
01-09-2020

2. REPORT TYPE
Final Performance

3. DATES COVERED *(From - To)*
15 Apr 2017 to 14 Apr 2020

4. TITLE AND SUBTITLE
Dimensional Reduction of Highly Nonlinear Multiscale Models Using Most Appropriate Local Reduced-Order Bases

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-17-1-0182

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Charbel Farhat

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
LELAND STANFORD JUNIOR UNIVERSITY
450 SERRA MALL
STANFORD, CA 94305-2004 US

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AF Office of Scientific Research
875 N. Randolph St. Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR RTA2

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-AFOSR-VA-TR-2020-0157

12. DISTRIBUTION/AVAILABILITY STATEMENT
A DISTRIBUTION UNLIMITED: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Higher-fidelity mathematical models, better approximation methods, and faster numerical algorithms have been developed for the solution of many computational problems. Linux clusters are now ubiquitous, GPUs (Graphics Processing Units) have shattered computing speed barriers, and exascale machines will increase computational power by at least two orders of magnitude. Most importantly, the potential of simulation based engineering science for providing a deeper understanding of complex engineering systems, improving design reliability, reducing design-cycle time, and enhancing their performance is well recognized today in many fields. Yet, in many applications such as turbulent flow computations at high Reynolds numbers, high-fidelity simulations remain so computationally intensive that they cannot be performed as often as needed. Consequently, their impact on engineering has been strong so far for the analysis and verification of carefully selected system configurations, system verification, and forensic applications. However, it has not been as strong for routine analysis, what-if scenarios, parametric studies, and time-critical applications such as design, design optimization, optimal control, and test support. Such applications demand a game-changing computational technology that leverages the power of high performance computing with the unique ability of low-dimensional computational models to perform in real-time. Nonlinear, Projection-based Model Order Reduction (PMOR) can provide this leverage. For this reason, this three-year basic research effort focuses on developing a rigorous, systematic approach for parametric, highly nonlinear, multiscale PMOR based on the recently developed mathematical concept of local reduced-order bases. Its main objective is to advance the state-of-the-art of nonlinear model reduction with respect to enhancing the feasibility of thi

15. SUBJECT TERMS
reduced order modeling, non-linear projection

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			FAHROO, FARIBA
Unclassified	Unclassified	Unclassified	UU		19b. TELEPHONE NUMBER *(Include area code)* 703-696-8429

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Table of Contents

1 Project Objectives and Expected Significance

1.1 Motivations and Identification of Research Problems

Higher-fidelity mathematical models, better approximation methods, and faster numerical algorithms have been developed for many solid and fluid applications. Linux clusters are now ubiquitous, GPUs (Graphics Processing Units) have shattered computing speed barriers, and exascale machines will increase computational power by at least two orders of magnitude. Most importantly, the potential of Computational Structural Dynamics (CSD)- and Computational Fluid Dynamics (CFD)-based simulations for providing a deeper understanding of complex engineering systems, improving their design reliability, reducing their design-cycle time, and enhancing their performance is well recognized today in many engineering fields. Yet, for many applications, high-fidelity CSD and CFD simulations remain so computationally intensive that they cannot be performed as often as needed, or are more often performed in special circumstances than routinely. This is the case, for example, for turbulent flow computations at high Reynolds numbers. Consequently, the impact of CSD and CFD on engineering had been strong until the beginning of this research effort for the analysis and verification of carefully selected system configurations, system verification, and forensic engineering. However, it has not been as strong for routine analysis, what-if scenarios, parametric studies, and time-critical applications such as design, design optimization, optimal control, and test support. Such applications demand a game-changing computational technology that leverages the power of high performance computing with the unique ability of low-dimensional computational models to perform in real-time. Projection-based Model Order Reduction (PMOR) – a process which aims to lower the computational complexity of a given computational model by reducing its dimensionality (or order) – can provide this leverage. For this reason, it is a serious contender for such a game-changing computational technology. Indeed, PMOR is becoming increasingly indispensable for parametric CSD and CFD studies, CSD- and CFD-based computational design, stochastic CSD and CFD computations for uncertainty quantification, and embedded CSD or CFD computing for model predictive control.

Linear PMOR, which consists of projecting a linear high-dimensional model onto a reduced subspace represented by a Reduced-Order Basis (ROB) in order to generate a corresponding Projection-based Reduced-Order Model (PROM), has deep roots in many engineering fields. These include structural dynamics [17], acoustics [27], and electronic circuit analysis [34]. For applications with a single or no system parameter, it has matured to the point where it has already been the subject of several textbooks [9, 11]. In this setting, it is embraced today by many practitioners. By comparison, *parametric* and *nonlinear* PMOR is still in its infancy. Nevertheless, major advances have been recently made in these two aspects of PMOR [4] which has consequently gained importance for those applications where parametric, nonlinear, high-dimensional simulations remain computationally intensive if not prohibitive. These include, among others, high-speed, multiscale, solid mechanics and structural dynamic computations, failure analysis and crack propagation, and turbulent CFD computations [13]. For such applications, the state-of-the-art of parametric and nonlinear PMOR has significantly advanced at several research institutions worldwide. These include, to name only a few, MIT, Rice University, and Stanford University in the US, Université Pierre-Marie Curie in France, École Polytechnique Féderale de Lauzanne in Switzerland, and the Max Planck Institute in Germany.

In the area of CFD, the Principal Investigator (PI) and his research team at Stanford University have developed a rigorous computational methodology for parametric nonlinear PMOR based on local ROBs [5] that has achieved a factor of 438 reduction in the CPU time required by a High-Dimensional CFD Model (HDM) for the solution of a benchmark turbulent flow problem, while maintaining 99%

of the level of accuracy of the HDM [13].

In the area of CSD, they have also pioneered a structure-preserving hyper reduction method [21] – that is, a PROM-related approximation method with provable stability properties that enables a low-dimensional computational model to perform in real-time. This method has reduced the CPU time required for the simulation of an underbody blast problem by a factor greater than 60,000, while maintaining 96% of the accuracy of the underlying high-dimensional multi-physics computational model [20].

In the area of computational aeroelasticity (or coupled CFD-CSD computations), the same team at Stanford University has exploited PMOR to enable real-time CFD-based flutter analysis on smart phones and tablets [8], and model predictive control [18]. Most importantly, the PI and his research team, in collaboration with Boeing and Volkswagen, have recently demonstrated for the first time the successful application of RANS (Reynolds-Averaged Navier-Stokes)-based CFD PROMs for the solution in real-time of two different, realistic, *what-if?* aerodynamic design problems. The first one focused on NASA's Common Research Model (CRM), which is representative of modern airline jets such as the Boeing 767. The second application centered on the 2009 Volkswagen Passat vehicle. The first case study is summarized below to illustrate the potential of CFD-based PROMs, and demonstrate the impact of the CFD-based PMOR capabilities readily available for the proposed basic research effort.

Specifically, the first aforementioned PMOR case study focused on the steady-state aerodynamics of a *parameterized* version of NASA's CRM (see Figure 1-left). The cruise condition was set to Mach 0.85, angle of attack 2.32°, and sideslip angle $0°$. For these fixed conditions and the chosen altitude, the Reynolds number is $R_e = 5 \times 10^6$. The starting point was a RANS CFD HDM based on the Spalart-Allmaras turbulence model and Reichardt's wall-law. The CRM geometry was parameterized in a four-dimensional parameter space consisting of (1) the wing span, (2) the stream-wise wing tip rake, (3) the vertical wing tip rake, and (4) the outboard twist (washout). The corresponding deformable mesh had 11,454,702 grid points and therefore 57,273,510 unknowns. A local, parametric, CFD PROM based on 3 local ROBs of dimensions 20, 23, and 21 only (reduction factor greater than 2 million) was constructed and hyper reduced on Excalibur (Cray XC40 at the Army Research Laboratory) in about 2 hours wall clock time. Using this CFD PROM, *what-if?* design scenarios where the four aforementioned shape parameters were varied to obtain 16 unsampled parameter configurations were simulated. Each CFD PROM computation was performed on a MacBook Air laptop and verified using a counterpart CFD HDM computation performed on 1,024 cores of Excalibur. In each case, the CFD PROM reproduced in less than 3 minutes on this laptop the numerical results computed in two hours wall-clock time on Excalibur using the CFD HDM, with less than 3% error (see Figure 1-right for the assessment of the accuracy of the pressure coefficient profiles predicted using the CFD PROM).

The advances in PMOR highlighted above have paved the way for the application of this potentially game-changing computational technology to the solution of many parametric, nonlinear, dynamical systems of practical importance to both of the scientific and engineering communities. However, numerous challenges remain to be addressed before nonlinear PMOR can achieve reliability and computational efficiency, particularly for highly nonlinear, multiscale dynamical systems of relevance to the Air Force. The basic research objectives stated next focus on addressing some of these challenges.

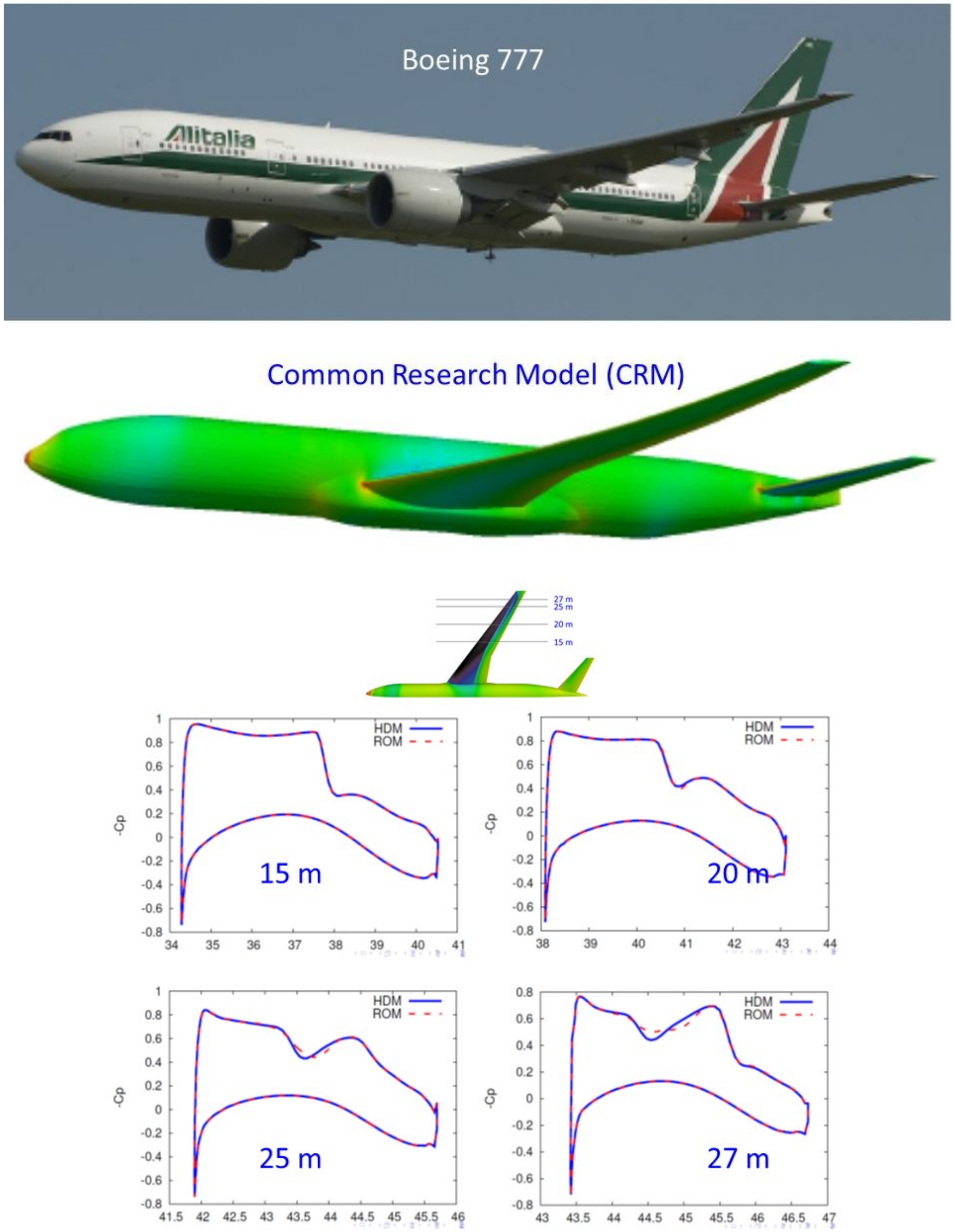


Figure 1: Aerodynamic analysis of a parameterized NASA Common Research Model (left) using a CFD HDM of dimension 57,273,510, and a parametric PROM counterpart of dimension 23 - Assessment of the accuracy of the pressure coefficient profiles predicted at various cross sections of the wing using the parametric nonlinear CFD PROM.

1.2 Statement of Objectives

Consequently, ww have proposed to develop a rigorous, *systematic* approach for parametric, highly nonlinear, multiscale PMOR based on the concept of local ROBs introduced in [5]. To this effect, we have formulated the following basic research objectives:

1. Enhance the feasibility of the concept of subspace approximation in a lower-dimensional subspace generated by most appropriate local basis vectors for computations that must be performed in relatively large time-intervals.
2. Refine this concept by considering both row and column clusterings so that features can be tracked in space, time, or space and time.
3. Guarantee the smoothness of an approximation in a lower-dimensional subspace generated by most appropriate local basis vectors.
4. Preserve the fine scales when approximating the solution of a multiscale problem in a low-dimensional subspace.

To achieve the above objectives, we have structured our research program as a 3-year basic research effort and organized it around 5 research tasks.

2 State-of-the Art of Nonlinear PMOR at the Beginning of this Research Effort

2.1 Parametric Projection-Based Model Order Reduction

Mathematically speaking, we are mainly interested in the fast numerical solution of μ-parametric dynamic problems using appropriate, PDE (Partial Differential Equation)-based computational models. (Here, and throughout the remainder of this proposal, μ denotes a vector of problem parameters such as, for example, *shape variables* or *operating points*). Such problems are particularly computationally intensive because they require multiple simulations. They are frequently encountered in the optimal design of engineering systems and their operating conditions (or operating points). Their discretization on an appropriate spatial grid using an equally appropriate time-step leads to a μ-parametric HDM (High-Dimensional Model) – that is, a computational model whose dimension N is significantly larger than what is referred to in this proposal as a PROM (Projection-Based Reduced-Order Model). Such an HDM is limited by the physics it does not incorporate. In the context of a multi-fidelity approach for the solution of a flow configuration of interest, it defines a certain level of fidelity. Parametric PMOR (Projection-Based Model Order Reduction) can be applied in principle to any such HDM to reduce its dimensionality and generate a corresponding parametric PROM. In this sense, PMOR is not about reducing the level of fidelity of the computational model to which it is applied, but mainly its order (or dimensionality). In particular, given an HDM of a certain fidelity level, the ultimate objective of PMOR is to produce a counterpart PROM that retains as much as possible its level of fidelity, but has a significantly lower dimension and therefore lower computational cost. For parametric problems, this may be mathematically possible, and practically feasible, for the reasons briefly explained below.

Typically, a μ-parametric, PDE-based computational model is an HDM. This is because its underlying spatial discretization is performed *a priori* – that is, before any significant knowledge about the response of the system it represents is developed – using local shape functions. Such a parametric model can be represented by a residual equation of the form $r(u, \mu) = 0$, where $r \in \mathbb{R}^N$ may be a linear, linearized, or nonlinear, semi-discrete or discrete vector function, and $u \in \mathbb{R}^N$ is the state vector or solution of interest. When the dimension (or number of degrees of freedom or unknowns) N is very large, this model becomes computationally intensive if not cost-prohibitive, and in any case, ill-suited for time-critical applications and real-time computing. On the other hand, performing an *affine* approximation of u in a reduced subspace represented by a *right* Reduced Order Basis (ROB) $V \in \mathbb{R}^{N\times n}$, where $n \ll N$ – that is, approximating u as $u \approx u_c + Vy$, where $u_c \in \mathbb{R}^N$ denotes a reference state and $y \in R^n$ is typically referred to as the vector of *generalized* or *reduced coordinates* – and projecting the μ-parametric HDM represented by the residual function $r \in \mathbb{R}^N$ onto this subspace using a *left* ROB $W \in \mathbb{R}^{N \times n}$, leads to the μ-parametric computational model $W^T r(Vy, \mu) = 0$ whose dimension is $n \ll N$. If $W \neq V$, the projection is known as a Petrov-Galerkin projection. Otherwise ($W = V$), it is known as a Galerkin projection. When V is carefully constructed *a posteriori* – that is, after some knowledge about the system response has been developed – the corresponding PROM can capture the dominant behavior of the underlying HDM and therefore retain most of its fidelity.

Knowledge about the system response can be obtained during a *training* procedure similar to a design of (numerical) experiments and performed *offline*. During this procedure, the model parameters represented by the vector $\mu \in \mathcal{D}_\mu \subset \mathbb{R}^{N_\mu}$, where $\mathcal{D}_\mu$ designates the parameter domain of interest and N_μ denotes its dimension, (which is the same as the number of parameters of interest) are sampled at a few points in $\mathcal{D}_\mu$ using a greedy but effective sampling strategy [12] equipped with a residual-based error indicator. At each sampled point in the parameter domain $\mathcal{D}_\mu$ – that is, sampled parameter vector – a set of problems related to the main problem of interest is solved, and the obtained numerical solutions are referred to as solution "snapshots". Next, these snapshots are compressed using, for example, the Singular Value Decomposition (SVD) to construct the right ROB V. The well-known Proper Orthogonal Decomposition (POD) method based on snapshots fits within this framework of data collection and compression.

To address the issue of robustness of a right ROB V and its corresponding parametric PROM with respect to changes in the parameters of interest, two strategies had been developed by the PMOR community by the time this research effort had started: a *point-wise* strategy based on a *database* of PROMs [2], and a *global* strategy based on the notion of a μ-global ROB. The *point-wise* strategy, which was pioneered by the PI and his research team, is particularly effective for linear or linearized problems [30]. For nonlinear problems, the global approach is more practical and effective. For linearized problems such as those which arise in stability analysis, the local strategy can be described as follows. A pair of individual right ROB V_i and left ROB W_i and their corresponding point-wise PROM $W_i^T r(V_i y, \mu_i)$ are computed offline at each sampled parameter point μ_i and stored in a database of linearized PROMs. Then, during the *online* stage where real-time computations are desired, a PROM $W_\star^T r(V_\star y, \mu_\star)$ is constructed at each queried but unsampled parameter point $\mu_\star$ by interpolating in real-time the matrix components of the stored point-wise linearized PROMs on appropriate matrix manifolds [3]. Subsequently, the interpolated PROM is used to perform in real-time the desired predictions for the queried but unsampled parameter configuration $\mu_\star$. For nonlinear problems however, it is more convenient to collect all solution snapshots computed at all sampled parameter points and compress them into a single global right ROB V, and compute a corresponding global left ROB W. In this case, the parameter sampling strategy is designed so that V (and therefore W) is (are) reliable in a large, if not the entire, region of the parameter domain $\mathcal{D}_\mu$. As stated earlier, greedy algorithms have been proposed in the literature for this purpose. However, the data sampling problem remains today an active topic of research, particularly when the dimension N_μ of the parameter domain $\mathcal{D}_\mu$ is large – say, $N_\mu \geq 10$. The main challenges include the development of error estimators or good error indicators for efficiently guiding the greedy algorithm, keeping the number of sampled points and therefore the amount of collected data as small as possible, and obtaining a right ROB of the lowest possible dimension.

Despite its low dimension n, a parametric PROM of the form $W^T r(Vy, \mu)$ does not necessarily guarantee, for nonlinear problems, computational feasibility or real-time performance. This is because the construction of a PROM scales not only with its small size n, but also with the large size of the underlying HDM, $N \gg n$. For linear problems, this construction is performed only once. More importantly, it is performed offline and therefore does not interfere with the desired online (real-time) computations. On the other hand for nonlinear problems, a PROM needs to be repeatedly reconstructed online to address, for example, time-dependency, whether the chosen computational strategy is an explicit or implicit one, or the dependence of the updated tangent operator on the previous solution, if the chosen computational strategy is an implicit one. This caveat is typically remedied by equipping a nonlinear PMOR method with a rigorous procedure for approximating $W^T r(Vy, \mu)$ whose computational complexity scales only with the small size n of the parametric nonlinear PROM. Such a procedure is often referred to in the literature as hyper reduction. It transforms the nonlinear PROM into a hyper reduced nonlinear PROM (or HPROM) that guarantees computational feasibility and enables real-time performance, while maintaining the desired level of accuracy. In the context of computational mechanics, two main approaches had been developed by the beginning of the this research effort for hyper reduction. The first one, originally proposed for continuous systems in [10], was later extended in [16] to semi-discrete and discrete systems. It is known as the Empirical Interpolation Method (EIM). Its discrete counterpart developed in [16] is commonly referred to as DEIM (where D stands for discrete). This approach is related to the Gappy POD method which was first introduced in [19] for the purpose of image reconstruction. However, EIM is characterized by a greater level of theoretical support. It can also be described as an approximate-then-project approach – that is, an approach where $r(Vy, \mu)$ is first approximated, then the product $W^T[r(Vy, \mu)]$ is evaluated. The GNAT (Gauss-Newton with Approximated Tensors) [14, 13] method developed by the PI and his collaborators belongs to the same category of hyper reduction methods. It enjoys however unique stability properties for CFD applications that distinguish it from competing approaches. It has also been developed around the elegant and practical concept of a reduced CFD mesh. Specifically, EIM, DEIM, GNAT and related methods seek an approximation of a high-dimensional nonlinear function such as $r(u, \mu)$ via expansion in an appropriate reduced basis and therefore are analogous to high-dimensional interpolation schemes. Then, they project the interpolated function onto the left subspace represented by W, whether $W = V$ or $W \neq V$. Such hyper reduction methods have been shown to be effective for nonlinear parabolic problems. However, they have also been shown to lead to numerically unstable HPROMs when applied to nonlinear second-order hyperbolic problems [20]. Alternatively, the second approach developed for hyper reduction approximates a nonlinear reduced-order operator such as $W^T r(Vy, \mu)$ directly. Hence, it can be described as a project-then-approximate approach. It has been developed mainly in the context of finite element semi-discretizations, but is equally applicable in the contexts of finite difference and finite volume methods. The Reduced Integration Domain (RID) [35], generalized cubature [6], and Energy Conserving Sampling and Weighting (ECSW) [20] methods are examples of this second approach for hyper reduction. They have been successfully applied to structural dynamics and solid mechanics problems, which are second-order hyperbolic problems. In particular, the ECSW method, which was developed by the PI and his co-workers, was shown in [21] to preserve the Lagrangian structure associated with Hamilton's principle for second-order dynamical systems, and thus to preserve the numerical stability properties of the applied time-integration scheme.

2.2 Reduced Basis Approximation Based on Locality in Solution Manifold

Highly nonlinear problems in fluid and solid mechanics can feature different physical regimes, moving features such as discontinuities and fronts, multiple scales, and physical instabilities. Consequently, capturing the dominant features of such problems using an approximate solution constructed in a fixed, low-dimensional subspace of global basis vectors respresented by a right ROB V is in many cases very challenging, if not simply impossible. To address this issue, the PI and his research group have proposed in [5] to approximate the solution of such highly nonlinear problems in a lower-dimensional subspace generated by most appropriate *local* basis vectors. In this approach, locality does not necessarily refer to space or time, but to the region of the manifold where the solution lies. This concept of a reduced basis approximation requires partitioning during the offline stage of PMOR the solution space into sub-regions, and constructing and assigning to each sub-region a local, low-dimensional, right ROB. It also requires identifying in real-time during the online incremental solution of the highly nonlinear reduced-order problem of the closest local region where the current high-dimensional solution lies, and updating this solution in the subspace generated by the identified local ROB. In this manner, different physical regimes and different scales should be capturable, at least *in principle*, using different low-dimensional ROBs which can be pre-computed and stored offline, and retrieved online only when needed.

The theoretical and algorithmic foundations of the local ROB approach outlined above have been laid out in [5]. They have also been successfully demonstrated in [5] for the solution of simple but parametric one-dimensional traveling shock wave problems, and in [36] for the solution of complex but non parametric three-dimensional unsteady aerodynamic problems.

2.3 Galerkin and Petrov-Galerkin Projections and Residual Minimization

If the nonlinear HDM is an implicit computational model, the residual $r(u_c + Vy)$ is typically linearized by a Newton or Newton-like method, which gives rise at each m-th Newton iteration to the overdetermined system of equations $J(u^{(m)}, \mu)V\Delta y = -r(u^{(m)}, \mu)$, where $J = \dfrac{\partial r}{\partial u}$ is the Jacobian of the residual with respect to the state vector, $u^{(m)}$ is the reconstructed high-dimensional solution at the m-th Newton iteration, and $\Delta y \in \mathbb{R}^n$ is the increment vector of generalized coordinates to be determined. In this case, an alternative approach to solving $\left(W^T J(u^{(m)}, \mu)V\right)\Delta y = -W^T r(u^{(m)}, \mu)$ in order to determine Δy, with $W = V \Leftrightarrow$ the HDM is associated with an elliptic problem, is to minimize some Θ-norm of the linearized residual $\|V\Delta y + J^{-1}(u^{(m)}, \mu) r(u^{(m)}, \mu)\|_\Theta$ – that is, to determine Δy as follows

$$\Delta y = \arg\min_{z\in\mathbb{R}^n} \|Vz + J^{-1}(u^{(m)}, \mu) r(u^{(m)}, \mu)\|_\Theta \tag{1}$$

In this case, it can be proved that choosing $W = V$ when appropriate, constructing a parametric, nonlinear, Galerkin PROM of the form $V^T r(Vy, \mu) = 0$, and determining Δy by solving the PROM equation $\left(V^T J(u^{(m)}, \mu)V\right)\Delta y = -V^T r(u^{(m)}, \mu)$ is equivalent to performing the minimization underlying (1) using $\Theta = J(u^{(m)}, \mu) = J^{(m)}$; hence, this choice is optimal in the sense of minimizing the discrepancy between the reduced-order and high-dimensional solutions only if the Jacobian $J^{(m)}$ is symmetric positive definite so that it can define a norm. While this is the case for most applications in solid mechanics and structural dynamics, it is not the case for most applications in fluid mechanics. However, it can also be proved that choosing $W = J^{(m)}V$ of the form $W^T r(Vy, \mu) = 0$ and determining Δy by solving the PROM is equivalent to performing the minimization underlying (1) using $\Theta = {J^{(m)}}^T J^{(m)}$ which is symmetric positive definite and therefore defines a norm [14]; hence, this choice is unconditionally optimal in the sense of minimizing the discrepancy between the reduced-order and high-dimensional solutions. These results, which were developed by the PI and his research team, were exploited in [14] and [13] to develop the first parametric, nonlinear, implicit PROMs in CFD and demonstrate their robustness, accuracy, and computational efficiency. Specifically, these nonlinear CFD PROMs were constructed using a Petrov-Galerkin projection approach based on $W = J^{(m)}V$, or equivalently, the minimum residual method (1) with $\Theta = {J^{(m)}}^T J^{(m)}$. Noting that $\|Vz + J^{-1}(u^{(m)}, \mu)r(u^{(m)}, \mu)\|_{J^{(m)T} J^{(m)}} = \|J(u^{(m)}, \mu)Vz + r(u^{(m)}, \mu)\|_2$, it follows that determining the increment vector of generalized coordinates Δy by solving the Petrov-Galerkin PROM based on $W = J^{(m)}V$ is also equivalent to solving the above nonlinear least-squares problem using the Gauss-Newton method, which is known to deliver an optimal solution.

3 Proposed Research

In summary, by the beginning of this research effort, nonlinear PMOR had been successfully applied to elliptic and parabolic problems. By comparison, it was still in its infancy for first-order hyperbolic dynamical systems such as those associated with high-speed flows, and second-order dynamical systems such as those associated with various forms of wave propagation, particularly in the presence of moving features such as discontinuities and fronts, multiple spatial and/or temporal scales, instabilities, and when the solution has statistical properties that must be preserved. Hence, we proposed to build on our latest efforts and results in this field to address these fundamental issues and advance the state-of-the-art of nonlinear PMOR. Specifically, we proposed a three-year basic research program focused on developing a rigorous and systematic approach for performing parametric, highly nonlinear, multiscale PMOR. We have based our approach on the concept of subspace approximation in a lower-dimensional subspace generated by most appropriate local basis vectors introduced in [5]. To this effect, we have proposed to organize our fundamental research effort around the following 5 complementary basic research tasks designed to achieve the objectives stated in Section 1.2:

- Task 1. Machine Learning Algorithms for Snapshot Clustering.
- Task 2. Online Reduced-Order Basis Selection Algorithms.
- Task 3. Scalable Data Compression Algorithms.
- Task 4. Smooth Approximations in a Lower-Dimensional Subspace Generated by Most Appropriate Local Basis Vectors.
- Task 5. Fine Scales Preserving Subspace Approximation of a Multiscale Solution.

Tasks 1, 2 and 3 address the first and second basic research objectives stated in Section 1.2. These are: to enhance the feasibility of the concept of subspace approximation in a lower-dimensional subspace generated by most appropriate local basis vectors for computations that must be performed in relatively large time-intervals; to refine this concept by considering both row and column clusterings so that features can be tracked in space, time, or space and time. Tasks 4 and 5 address the third and fourth basic research objectives stated in Section 1.2, respectively – that is, to guarantee smoothness and preserve the fine scales when approximating the solution of a multiscale problem in a low-dimensional subspace. Tasks 1–3 leverage advances in machine learning and linear algebra. Tasks 4 and 5 leverage recent advances in multiscale modeling and applied mathematics.

Throughout all tasks identified above, we have used examples from high-speed structural dynamics and/or turbulent flow applications to assess, illustrate, and demonstrate the developed mathematical results and numerical algorithms.

4 Accomplishments and State-of-the-Art at the End of this Research Effort

4.1 Machine Learning Algorithms for Snapshot Clustering (Task 1)

We have pursued two different types of clustering: *column* clustering, and *column-row* clustering. Let

$$U = [u^{1,\mu_1},\, u^{2,\mu_1},\, ,\cdots,\, u^{N_1,\mu_1}, u^{1,\mu_2},\, \cdots, u^{N_2,\mu_2},\, \cdots,\, u^{1,\mu_p},\, \cdots, u^{N_p,\mu_p}]$$

denote the matrix of N_s solution snapshots collected along the trajectories computed for the sampled parameter points $\mu_1, \mu_2, \cdots, \mu_p$, where $N_s = N_1 + N_2 + \cdots + N_p$. Column clustering consists in computing the partitioning $S_C = \{S_1,\, S_2,\, \cdots, S_k\}$ of U into $k \leq N_s$ sets S_j, each grouping together solution snapshots represented by columns of U that share a same important feature of the problem of interest such as, for example, a same physical regime or a same temporal scale of the nonlinear solution, and therefore define and characterize a certain region of its manifold. This is the type of clustering that was adopted in [5]. It is ideal for tracking and grouping features in time, in an ordered or unordered fashion – when μ is varied in $\mathcal{D}_\mu$. However, it is not necessarily optimal for tracking features in space. For this reason, we have also pursued here colum-row clustering strategies where after U has been column-clustered, a partitioning $S_R = \{S_{1i},\, S_{2i},\, \cdots, S_{li}\}$ of S_j into $l \leq N$ sets S_{ij}, each grouping together components of solution snapshots represented by rows of S_j that share a same important highly nonlinear feature of the problem of interest, and therefore define and characterize a local region of its solution manifold. The main idea behind this *refined* partitioning is that it should be preferrable for tracking and grouping features in both the temporal and spatial domains, in an ordered or unordered fashion. Speficically, we have shown that such a clustering is capable of tracking both spatial and temporal aspects of nonlinear solution on its manifold [7].

In either case, we have developed Machine Learning (ML) algorithms to compute, for a given collection of parametric solution snapshots represented by the matrix $U \in \mathbb{R}^{N\times N_s}$, the desired clusters. To this effect, we note that the celebrated k-means algorithm is perhaps the most popular clustering algorithm. As such, it was also adopted in [5]. This ML algorithm partitions the N_s snapshots of U into k clusters in which each solution snapshot belongs to the cluster with the nearest mean, serving as a prototype of the cluster. It also results in a partitioning of the data space into Voronoi cells. By construction, it fulfills the desire to group together data pieces that define and characterize a region of the solution manifold. Unfortunately, it does not fulfill *à priori* the implicit requirement that each cluster of solution snapshots S_j or S_{ij} be low-dimensional in order to enable the construction of a low-dimensional local ROB V_j, or its component V_{ij}, respectively.

Hence, under this basic research task, we have pursued the development of alternative ML algorithms for data clustering that partition a data set into data points that lie in a union of low-dimensional subspaces. As a starting point, we have construct a sparse subspace clustering algorithm based on the concept of a similarity matrix [15]. Given the snapshot matrix $U \in \mathbb{R}^{N\times N_s}$, this algorithm builds a similarity matrix $C \in \mathbb{R}^{N\times N_s}$ column-by-column by solving for each column the convex optimization problem

$$c_j = \min \|\xi\|_1 \, s.t. \, u_j = S\xi,\, \xi_j = 0$$

where u_j is the corresponding column of the snapshot matrix U. Then, it performs a spectral clustering on the symmetrized similarity matrix $|C| + |C|^T$ to obtain simultaneously the optimal number of clusters k and the partitioning S_C or S_R, depending on the scope of the snapshot matrix U.

Next, we have also designed soft clustering methods based on the fuzzy c-means algorithm that are robust at the boundaries of overlapping clusters and demonstrated their unique potential for multidisciplinary analysis and optimization problems using local PROMs.

4.2 Online Reduced-Order Basis Selection Algorithms (Task 2)

The proposed local ROB approach for building effective parametric nonlinear PROMs operates in two stages. In the first one, the solution space is sub-divided into sub-regions where the nonlinear solution exhibits significantly different features (for example, flow regime, spatial or temporal scale, $\cdots$), and a local right ROB is constructed and assigned to each sub-region. This stage can be performed offline. In the second stage, the sub-region where the current solution state lies must be first identified, its assigned ROB must be retrieved, and the corresponding parametric nonlinear PROM $W^T r(Vy, \mu) = 0$ must be constructed on the fly, after a method for constructing W is chosen (see Section 2.3). This stage must be performed online. Hence, its implementation requires two real-time computational technologies: one for indentifying the local ROB to use for constructing the reduced-order approximation where the current solution state lies, and one for performing the hyper reduction of $W^T r(Vy, \mu)$ or related operators. In this basic research task, we have first addressed the former requirement by developing a real-time algorithm for this purpose and demonstrating its performance in [23]. Then, we have addressed the second requirement by expanding the scope of ECSW (see Section 2.1) to Petrov-Galerkin PROMs [23].

Specifically, we have developed under this basic research task a real-time algorithm for the identification of the sub-region of the manifold where the current solution of the problem of interest lies. We have used as a starting point the real-time algorithm previously presented by the PI and co-workers in [5]. This algorithm identifies a sought-after sub-region by identifying the cluster centroid whose state is the closest to the current state. For this purpose, it adopts the same metric as that used for clustering the snapshots. In [5], this algorithm was proved to be optimal and real-time capable. However, it is attached to the notion of a mean state and therefore most appropriate for column clustering. Under this basic research task, we have explored its suitability for column-row clustering and developed an alternative for such snapshot partitionings that proved to be computationally more efficient [23].

We have also extended the ECSW hyperreduction method to Petrov-Galerkin PROMs where the underlying high-dimensional model can be associated with an arbitrary finite element, finite volume, or finite difference semi-discretization method. We have extended the scope of this method to cover local PROMs based on piecewise-affine approximation subspaces, such as those designed for mitigating the Kolmogorov barrier issue associated with convection-dominated flow problems. We have shown that the resulting hyperreduction method is robust and accurate [23]. In particular, we have shown that its offline phase is fast and parallelizable. We have demonstrated the potential of its online phase for large-scale applications of industrial relevance using turbulent flow problems with $O(10^7)$ and $O(10^8)$ degrees of freedom. We have also shown that for such problems, the online part of the extended ECSW method enables for challenging CFD problems wall-clock time and CPU time speedup factors of three and four orders of magnitude, respectively, while delivering an exceptional accuracy.

4.3 Scalable Data Compression Algorithms (Task 3)

Once the snapshot clusters have been generated as discussed in Section 4.1 above, it remains to compress the data in each column cluster S_j or column-row cluster S_{ij} to obtain the desired local right ROB V_j, or its component V_{ij}, respectively. Typically, the SVD algorithm is used for this purpose. However, when the high-fidelity simulations must be performed in large time-intervals because of the physics of interest, the anticipated number of snapshots N_s needed for performing the training procedure is such that the snapshot matrix $U \in \mathbb{R}^{N \times N_s}$ can be extremely large. In this case, the SVD decomposition of such a matrix can become overwhelming even when performed on a massively parallel computing system, and may be even prohibited by memory limitations. At the beginning of this research effort, we were concerned that clustering the snapshots as explained above may or may not alleviate this issue, depending on the characteristics of the problem of interest. For this reason, we have developed under this basic research task scalable alternatives to the standard thin SVD algorithm including: limited memory recursive SVD algorithms where we decompose the snapshots or snapshot clusters into blocks and compress them in a recursive manner; SVD with low-rank update algorithms where we perform a thin SVD on a block of snapshots and enrich the obtained result with low-rank updates of the remaining snapshots; and stochastic or probabilistic SVD algorithms which are known to be memory efficient. By having all such alternatives and enhancing the computational complexity of our PMOR software within our CFD and CSD analyzers, we have been able to apply PMOR to the showcase applications reported in Section 4.6.

4.4 Smooth Approximations in a Lower-Dimensional Subspace Generated by Most Appropriate Local Basis Vectors (Task 4)

As already discussed above, the approximation of a solution in a lower-dimensional subspace generated by most appropriate *local* basis vectors [5] is a promising method for the reduction of highly nonlinear computational models. It is supported by theoretical and algorithmic foundations that have been laid out in [5]. Its potential for the real-time solution of highly nonlinear problems in computational mechanics has been successfully demonstrated in [5, 36]. However, this nonlinear PMOR suffers from one caveat: for time-dependent problems, the solutions delivered by the nonlinear PROMs resulting from this local approximation method are vulnerable to spurious oscillations. These may be introduced in a solution computed online when switching from one local ROB to another. Hence, in this basic research task, we have addressed this issue and eliminated this caveat. Specifically, we have pursued two different approaches for this purpose:

- *Overlapping Clustering*. In this approach, we have considered overlapping clustering algorithms. Here, the basic idea is that if two neighboring snapshot clusters overlap, by construction, their associated local ROBs will share some directional components. In this case, switching between two ROBs does not generate a spurious "impulse" in the simulated system and therfore does not introduce spurious oscillations in the predicted numerical response. In particular, we have:
  - Developed an understanding of *which* and *how many* snapshots to overlap in order to achieve a certain *level* of continuity of the approximation when switching from one ROB to another.
  - Determined whether the online solution process may require switching at a certain time-instance between two non neighboring clusters, and how to define in this case an overlap between two such snapshot clusters.
- *Discontinuous Space-Time Solution*. Alternatively, instead of treating the method of approximation in a lower-dimensional subspace generated by most appropriate local basis vectors introduced in [5] as a discontinuous Galerkin (or Petrov-Galerkin) semi-discretization method that may introduce spurious oscillations in its associated discretization (or time-integration) scheme, we have reformulated this method as a genuine discontinuous space-time method. Then, we

have equipped it with (hybrid) variational formulations for enforcing various weak forms of the continuity of various quantities of interest (such as, for example, a kinetic energy) at the left and right sides of each time-instance.

4.5 Fine Scales Preserving Subspace Approximation of a Multiscale Solution (Task 5)

Nonlinear multiscale problems – defined here as those nonlinear problems that exhibit vastly different scale features that are significant to the macroscopic behavior – are ubiquitous in science and engineering. Numerical methods that attempt to resolve all such scales lead to massive discretized problems. Consequently, a number of multiscale methods have been introduced to model highly heterogeneous solid and structural systems in solid mechanics, and turbulent flows and combusion in fluid mechanics, without requiring monolithic discretizations. Despite these advances, the numerical solution of multiscale methods remains today computationally intensive, if not simply unaffordable. Hence, there is a significant interest in understanding if and how nonlinear PMOR can be applied to reduce the dimensionality of multiscale models and accelerate their processing by orders of magnitude. In particular, the main concern centers around the question of whether a nonlinear PMOR can preserve the fine scales of a high-dimensional, multiscale solution?

Consider again the nonlinear PMOR based on the approximation of a solution in a lower-dimensional subspace generated by most appropriate *local* basis vectors introduced in [5]. In its offline stage, any fine scale captured by the HDM and represented in a sampled snapshot will find its way to a column (time) or column-row (time-space) cluster. Then, the first question becomes whether this scale can be preserved during data compression within a cluster using SVD or an SVD-like algorithm?

In its online stage, the computation of the generalized coordinates vector Δy requires the solution of a minimization problem using a θ-norm (see Eq. (1) in Section 2.3). Hence, the second question becomes whether a fine scale that survived an SVD or SVD-like data compression can survive this minimization process in the presence of larger scales?

Rather than answering the above two questions *directly*, we have designed here a rigorous process that ensures that all scales of interest represented in a high-dimensional solution are preserved during a nonlinear PMOR. We have achieved this objective by separating the scales *à priori*, and performing nonlinear PMOR at all scales. To this effect, we have considered both fields of solid and fluid mechanics as the techniques for scale separation in these two fields are different.

For nonlinear multiscale problems in solid mechanics, we have considered the well-known concept of a locally attached microstructure (for example, see [22]) which leads to a general approach for separating scales *à priori*. It can be described as follows. At each macroscale material point, the stress-strain relationship is determined via a detailed computational analysis of a microstructure boundary value problem. The corresponding boundary conditions are determined from the localization of the pointwise deformation gradient at the macroscale. The pointwise macroscale stress tensor is determined from the distribution of this tensor over the microstructure and a homogenization method. An analytic constitutive law is prescribed at the microscale where the length scale is sufficiently small so that a homogeneous medium can be assumed, and existing material models can accurately capture the response. This concept is very computationally intensive because it requires performing a microstructure analysis at each Gauss point of the macroscale model. Consequently, several PMOR methods have been recently developed to reduce its computational cost (for example, see [37, 31, 32, 1, 25]). A number of other methods based on the so-called reduced basis method have also been designed for elliptic (static), multiscale, *scalar* problems with affine parameter dependence

[32, 33, 26]. However, the relevance of these methods to real-world, nonlinear, multiscale dynamic problems and their impact on the acceleration of their solution time has yet to be determined. Hence, we have designed under this basic research task a nonlinear PMOR framework based on an $N+1$-level locally attached microstructure approach and incorporate in it a hyperreduction process. In this computational framework, we have reduced the dimensionality of the governing equations at all $N+1$ scales, in sharp contrast with the aforementioned previous works on this topic where model reduction was performed only at the finest scale. Specifically, we have performed PMOR using a Galerkin projection onto a well-chosen subspace of approximation. We have used SVD to extract this subspace from relevant training data. We have used averaging-based localization and homogenization schemes to transmit strain information to finer scales and stress information to coarser scales, respectively, and the deformation gradient at the k-th scale to define uniform or periodic essential boundary conditions at the $k+1$-th scale. We have computed the pointwise first Piola-Kirchhoff stress tensor at the k-th scale as the volumetric average of the corresponding stress tensor over the locally attached microstructre, i.e., the k-th scale. To achieve a computational complexity that depends on the lowest dimensions only, we have tailored the ECSW method [20, 21] to the context of the locally attached microstructure approach by redesigning its training algorithm to ensure that the reaction forces associated with the essential boundary conditions at the miscroscales are are well-approximated, and used it for hyper-reducing the projected nonlinear internal forces. In [38], we have demonstrated the performance of this multilevel PMOR framework for nonlinear dynamic multiscale problems in structural and solid mechanics. Specifically, we have demonstrated for two benchmark multiscale problems in structural and solid mechanics that our developed multiscale PMOR framework is capable of delivering speedup factors as high as five orders of magnitude.

Next, we have pointed out that conventional offline training of ROBs in a predetermined region of a parameter space leads to parametric reduced-order models that are vulnerable to extrapolation. This vulnerability manifests itself whenever a queried parameter point lies in an unexplored region of the parameter space. In [24], we have addressed this issue by presenting an in-situ, adaptive framework for nonlinear model reduction where computations are performed by default online, and shifted offline as needed. The framework is based on the concept of a database of local ROBs, where locality is defined in the parameter space of interest. It achieves accuracy by updating on-the-fly a pre-computed ROB, and approximating the solution of a dynamical system along its trajectory using a sequence of most-appropriate local ROBs. It achieves efficiency by managing the dimension of a local ROB, and incorporating hyperreduction in the process. While this framework is sufficiently comprehensive, we have developed it in the context of dynamic multiscale computations in solid mechanics. Most importantly, we have shown that even in a nonparametric setting of the macroscale problem and when all offline, online, and adaptation overhead costs are accounted for, the computational framework we have developed can accelerate a single three-dimensional, nonlinear, multiscale computation by at least an order of magnitude, without compromising accuracy.

For nonlinear multiscale problems in fluid dynamics, we have considered Large Eddy Simulations (LES) where the large-scale features of a turbulent flow are resolved, and the subgrid-scale stresses representing the effect of the unresolved (fine) scales on the resolved (coarse) ones are modeled. To separate these scales *à priori*, we have adopt the Variational Multiscale (VMS) framework introduced in [28], where one does not filter the Navier-Stokes equations but uses instead a variational projection. Specifically, we have adopted the method based on cell agglomeration proposed in [29] by the PI and co-workers for separating *à priori* the coarse and fine scales. We have tailored the nonlinear, multiscale PMOR outlined above to reduce dimensionality at both resolved and modeled scales, in order to maintain the proposed nonlinear multiscale PMOR framework unified at the fundamental level. In particular, we have substituted the Galerkin projection with a Petrov-Galerkin projection to maintain numerical stability, and the GNAT procedure by the new ECSW method that we have developed under this research effort for the hyperreduction of Petrov-Galerkin PROMs. As in the case of solid mechanics problems, we have shown that we have achieved all objectives associated with this task.

4.6 State-of-the-Art of Nonlinear PMOR at the End of this Research Effort

By the end of this research effort, the PI and his research team at Stanford have truly re-set the state-of-the-art of nonlinear PMOR. The following slides describe this new state-of-the-art for challenging Detached Eddy Simulations, LES, and nonlinear, dynamic, multiscale modeling of solid mechanics problems.

5 Bibliography & References Cited

References

[1] Assyr Abdulle and Yun Bai. Adaptive reduced basis finite element heterogeneous multiscale method. *Computer Methods in Applied Mechanics and Engineering*, 257:203–220, 2013.

[2] D. Amsallem, J. Cortial, and C. Farhat. Toward real-time CFD-based aeroelastic computations using a database of reduced-order information. *AIAA Journal*, 48(9):2029–2037, 2010.

[3] D. Amsallem and C. Farhat. An online method for interpolating linear parametric reduced-order models. *SIAM Journal on Scientific and Statistical Computing*, 33:2169–2198 — second most downloaded paper of this journal in September and October 2011, 2011.

[4] D. Amsallem, C. Farhat, and B. Haasdonk. Special issue on model reduction. *International Journal for Numerical Methods in Engineering*, 102:931–932, 2015.

[5] D. Amsallem, M. Zahr, and C. Farhat. Nonlinear model order reduction based on local reduced-order bases. *International Journal for Numerical Methods in Engineering*, 92:891–916, 2012.

[6] Steven S An, Theodore Kim, and Doug L James. Optimizing cubature for efficient integration of subspace deformations. *ACM Transactions on Graphics*, 27(5):165, 2008.

[7] S Anderson, T White, P Avery, and C Farhat. Row-wise clustering for sparse and spatial local reduced-order bases. In *Thirteenth World Congress on Computational Mechanics (WCCM 2018)*, New York City, New York, July 22-27, 2018.

[8] Good Vibrations: Stanford Engineers Put a Damper on Aeroelastic Flutter Andrew Myers. Stanford report. http://news.stanford.edu/news/2011/march/airplane-aeroelastic-flutter-032411.html.

[9] A. C. Antoulas. *Approximation of Large-Scale Dynamical Systems (Advances in Design and Control)*. SIAM, 2010.

[10] Maxime Barrault, Yvon Maday, Ngoc Cuong Nguyen, and Anthony T Patera. An empirical interpolationmethod: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathematique*, 339(9):667–672, 2004.

[11] P. Benner, M. Hinze, and E. Jan W. ter Maten. *Model Reduction for Circuit Simulation, Lectures Notes in Electrical Engineering*, volume 74. Springer, 2011.

[12] Tan Bui-Thanh, Karen Willcox, and Omar Ghattas. Model reduction for large-scale systems with high-dimensional parametric input space. *SIAM Journal on Scientific Computing*, 30(6):3270–3288, 2008.

[13] K. Carlberg, C. Farhat, J. Cortial, and D. Amsallem. The gnat method for nonlinear model reduction: Effective implementation and application to computational fluid dynamics and turbulent flows. *Journal of Computational Physics*, 242:623–647 — most cited paper of this journal in 2013, 2013.

[14] Kevin Carlberg, Charbel Bou-Mosleh, and Charbel Farhat. Efficient non-linear model reduction via a least-squares petrov–galerkin projection and compressive tensor approximations. *International Journal for Numerical Methods in Engineering*, 86(2):155–181, 2011.

[15] T Chapman and C Farhat. Sparse subspace clustering for nonlinear projection-based model order reduction. In *Fourteenth US National Congress on Computational Mechanics*, Montreal, Canada, July 17-20, 2017.

[16] Saifon Chaturantabut and Danny C Sorensen. Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal on Scientific Computing*, 32(5):2737–2764, 2010.

[17] R. Craig and M. Bampton. Coupling of substructures for dynamic analysis. *AIAA Journal*, 6:1313–1319, 1968.

[18] F. Gurrola D. Amsallem, S. Deolalikar and C. Farhat. Model predictive control under coupled fluid-structure constraints using a database of reduced-order models on a tablet. In *AIAA-2013-2588, 31st AIAA Applied Aerodynamics Conference*, San Diego, CA, June 24-27, 2013.

[19] R. Everson and Lawrence Sirovich. Karhunen-Loeve procedure for gappy data. *Journal of the Optical Society of America A*, 12(8):1657–1664, 1995.

[20] C. Farhat, P. Avery, T. Chapman, and J. Cortial. Dimensional reduction of nonlinear finite element dynamic models with finite rotations and energy-conserving mesh sampling and weighting for computational efficiency. *International Journal for Numerical Methods in Engineering*, 98:625–662, 2014.

[21] C. Farhat, T. Chapman, and P. Avery. Structure-preserving, stability, and accuracy properties of the energy-conserving sampling and weighting (ecsw) method for the hyper reduction of nonlinear finite element dynamic models. *International Journal for Numerical Methods in Engineering*, 102:1077–1110, 2015.

[22] Frédéric Feyel and Jean-Louis Chaboche. FE^2 multiscale approach for modelling the elasto-viscoplastic behaviour of long fibre sic/ti composite materials. *Computer methods in applied mechanics and engineering*, 183(3):309–330, 2000.

[23] Sebastian Grimberg, Charbel Farhat, Radek Tezaur, and Charbel Bou-Mosleh. Mesh sampling and weighting for the hyperreduction of nonlinear petrov-galerkin reduced-order models with local reduced-order bases. *arXiv preprint arXiv:2008.02891*, 2020.

[24] W He, P Avery, and C Farhat. In-situ adaptive reduction of nonlinear multiscale structural dynamics models. *International Journal for Numerical Methods in Engineering*, (in press).

[25] JA Hernández, J Oliver, Alfredo Edmundo Huespe, MA Caicedo, and JC Cante. High-performance model reduction techniques in computational multiscale homogenization. *Computer Methods in Applied Mechanics and Engineering*, 276:149–189, 2014.

[26] Jan S Hesthaven, Shun Zhang, and Xueyu Zhu. Reduced basis multiscale finite element methods for elliptic problems. *Multiscale Modeling & Simulation*, 13(1):316–337, 2015.

[27] U. Hetmaniuk, R. Tezaur, and C. Farhat. Review and assessment of interpolatory model order reduction methods for frequency response structural dynamics and acoustics problems. *International Journal for Numerical Methods in Engineering*, 90:1636–1662, 2012.

[28] TJR Hughes, L Mazzei, and KE Jansen. Large eddy simulation and the variational multiscale method. *Comput. Vis. Sci.*, 3(47), 2000.

[29] B Koobus and C Farhat. A variational multiscale method for the large eddy simulation of compressible turbulent flows on unstructured meshes application to vortex shedding. *Computer Methods in Applied Mechanics and Engineering*, pages 1367–1384, 2004.

[30] T. Lieu, C. Farhat, and M. Lesoinne. Reduced-order fluid/structure modeling of a complete aircraft configuration. *Computer Methods in Applied Mechanics and Engineering*, 195(41-43):5730–5742, June 2009.

[31] E Monteiro, Julien Yvonnet, and Qi-Chang He. Computational homogenization for nonlinear conduction in heterogeneous materials using model reduction. *Computational Materials Science*, 42(4):704–712, 2008.

[32] Ngoc Cuong Nguyen. A multiscale reduced-basis method for parametrized elliptic partial differential equations with multiple scales. *Journal of Computational Physics*, 227(23):9807–9822, 2008.

[33] Mario Ohlberger and Felix Schindler. Error control for the localized reduced basis multiscale method with adaptive on-line enrichment. *SIAM Journal on Scientific Computing*, 37(6):A2865–A2895, 2015.

[34] L. T. Pillage and R. A. Rohrer. Asymptotic waveform evaluation for timing analysis. *IEEE Transactions Computer Aided Design*, 9:352–366, 1990.

[35] D Ryckelynck. A priori hyperreduction method: an adaptive approach. *Journal of computational physics*, 202(1):346–366, 2005.

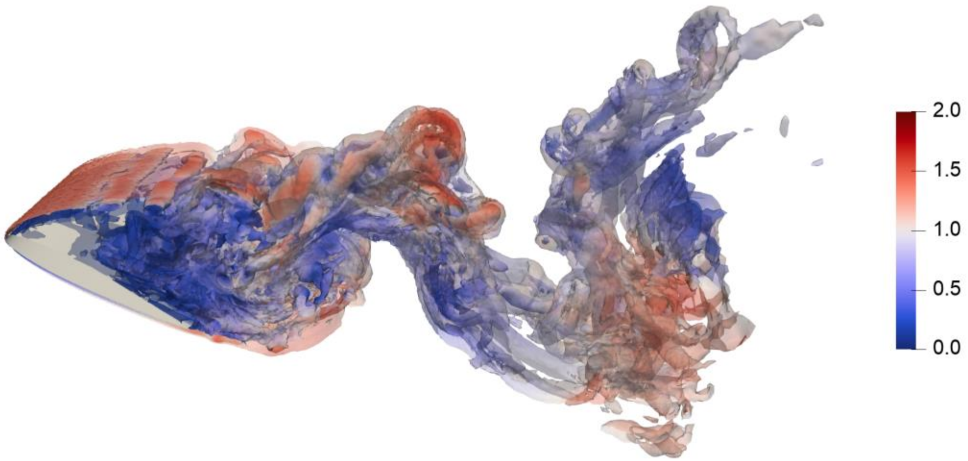
[36] K. Washabaugh, D. Amsallem, M. Zahr, and C. Farhat. Nonlinear model reduction for cfd problems using local reduced order bases. In *AIAA-2012-2686, 42nd AIAA Fluid Dynamics Conference and Exhibit*, New Orleans, Louisiana, June 25-28, 2012.

[37] Julien Yvonnet and Q-C He. The reduced model multiscale method (r3m) for the non-linear homogenization of hyperelastic media at finite strains. *Journal of Computational Physics*, 223(1):341–368, 2007.

[38] Matthew J Zahr, Philip Avery, and Charbel Farhat. A multilevel projection-based model order reduction framework for nonlinear dynamic multiscale problems in structural and solid mechanics. *International Journal for Numerical Methods in Engineering*, 112(8):855–881, 2017.

SHOWCASE APPLICATION #1: LES OF TURBULENT FLOW PAST A NACA AIRFOIL

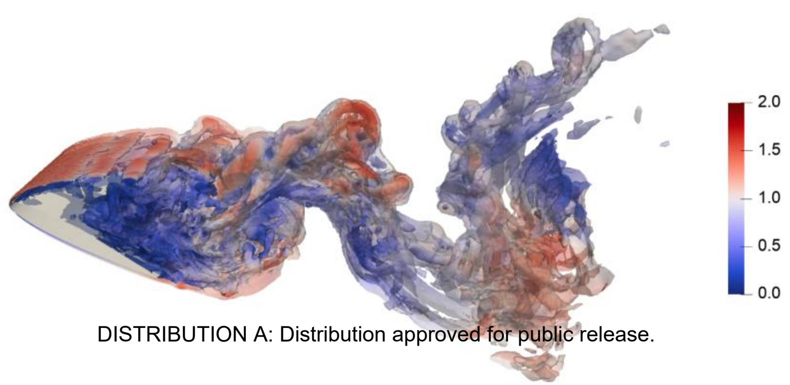
- **Large eddy simulation (LES) of turbulent compressible flow past a NACA 0012 airfoil at $\alpha = 30°$**
  - $Re = 10{,}000$, $M_\infty = 0.2$
  - solution computed in nondimensional simulation time-interval [0, 30], corresponding to $\sim$10 periods of vortex shedding



SHOWCASE APPLICATION #1:
LES OF TURBULENT FLOW PAST A NACA AIRFOIL

- **Spatial and temporal discretization**
  - 5th-order low diffusion FV scheme for convective term, 2nd-order linear Galerkin FE for diffusive terms
  - Vreman subgrid-scale model
  - time-integration performed using third-order DIRK scheme with nondimensional $\Delta t = 0.02$

- **Computational domain**
  - one chord length extrusion in spanwise direction, periodic BC applied on spanwise faces, no-slip at airfoil surface
  - 2,079,546 vertices and 11,909,406 tetrahedral elements yields $N =$ **10,397,730**



SHOWCASE APPLICATION #1: LES OF TURBULENT FLOW PAST A NACA AIRFOIL

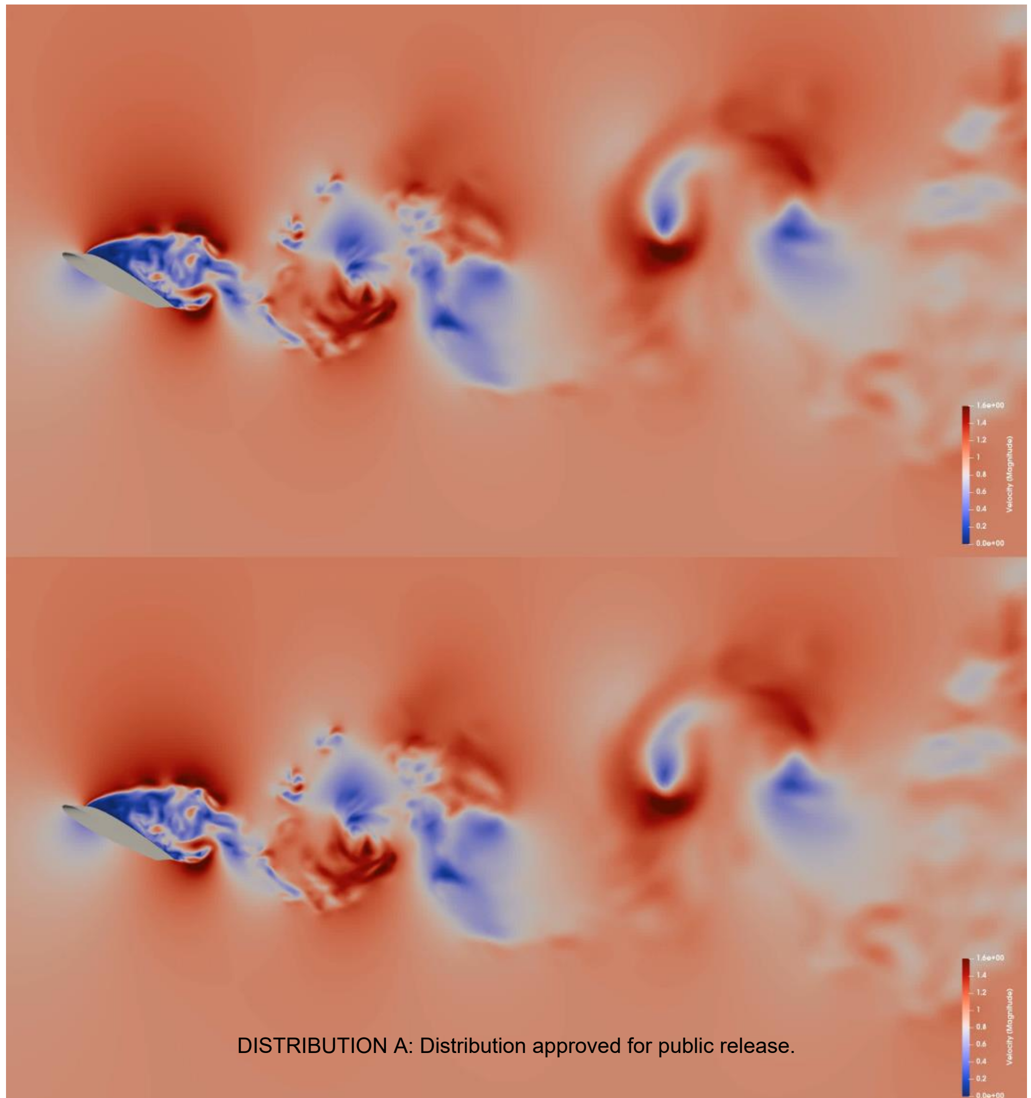
- **LSPG-based PROM**
  - 501 solution snapshots collected using sampling frequency $\Delta t = 0.06$
  - construct of ROB of dimension $\boldsymbol{n} = \mathbf{83}$ via POD, corresponding to SV energy threshold of 95%

- **HPROM training**
  - FV-adapted ECSW with training tolerance of $\varepsilon = 0.01$ and 51 training snapshots
  - mesh sampling yields $|\widetilde{\mathcal{E}}| = 2{,}390$ sampled cells ***(0.12% of HDM mesh cells)***

DISTRIBUTION A: Distribution approved for public release.

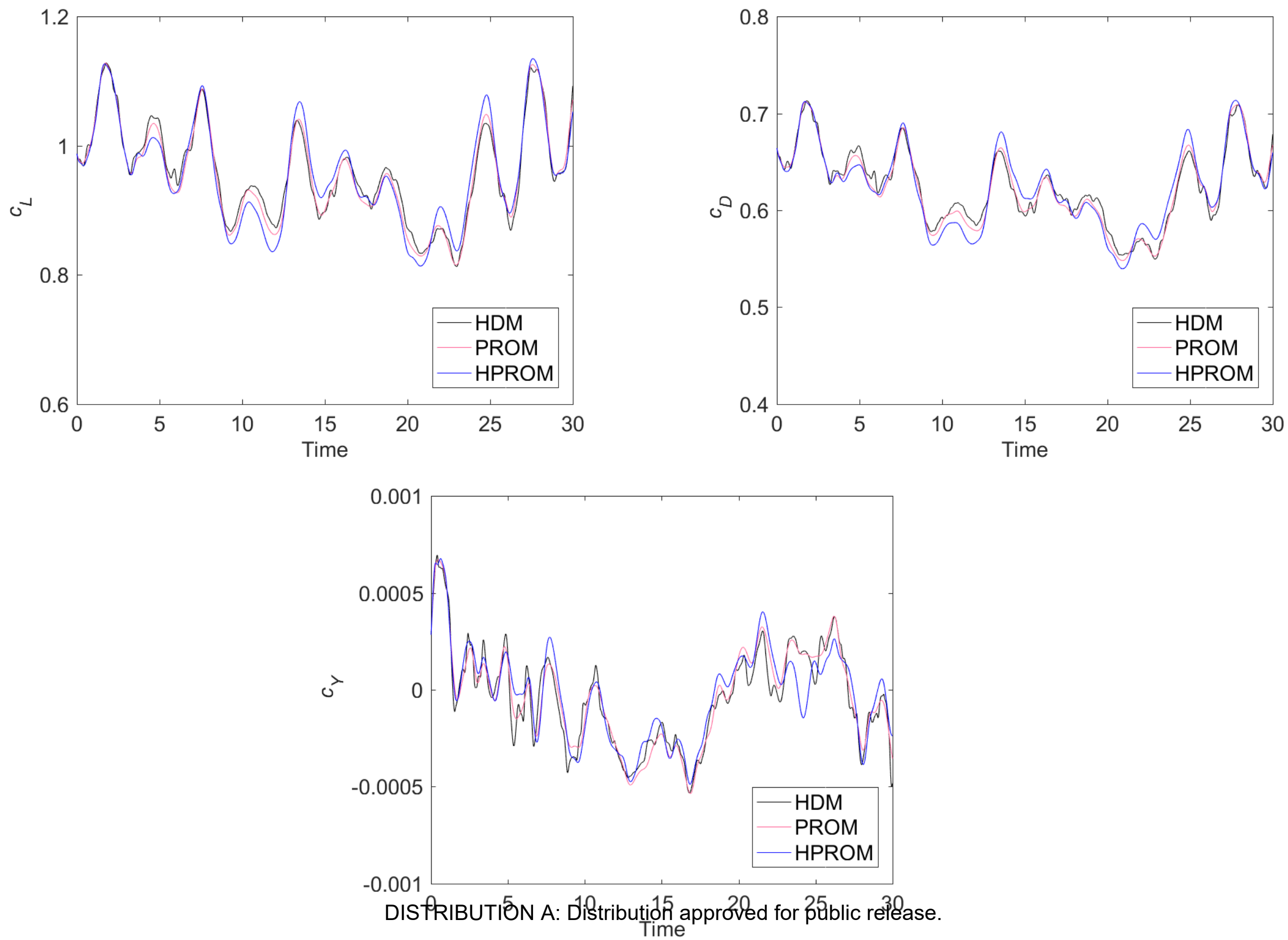
SHOWCASE APPLICATION #1:
LES OF TURBULENT FLOW PAST A NACA AIRFOIL

- **Visualization of HDM and reconstructed HPROM solutions by velocity magnitude**

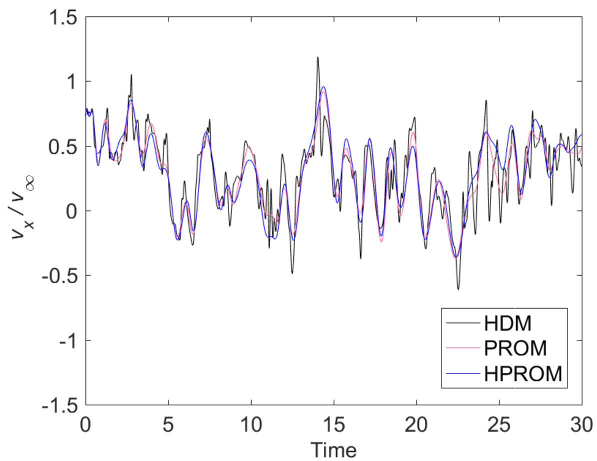


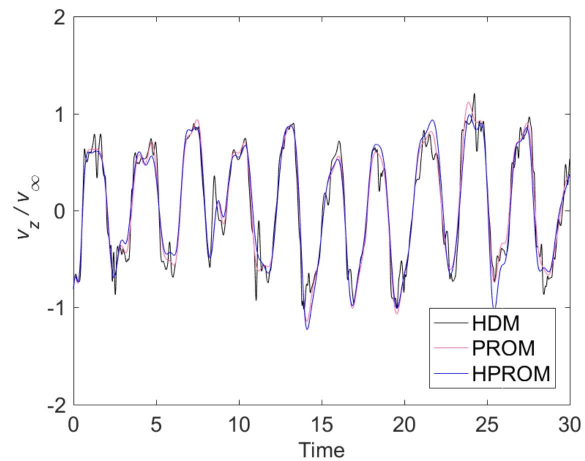
SHOWCASE APPLICATION #1: LES OF TURBULENT FLOW PAST A NACA AIRFOIL

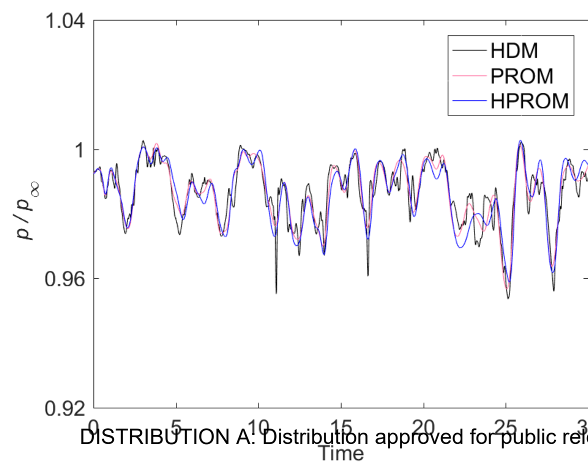
- **Comparison of lift, drag, and side force coefficient time-histories computed using HDM, PROM, and ECSW-based HPROM**



- **Comparison of streamwise and vertical velocity components and pressure computed using HDM, PROM, and ECSW-based HPROM**
  - probe located 1.5*c* downstream of airfoil half-chord







SHOWCASE APPLICATION #1:
LES OF TURBULENT FLOW PAST A NACA AIRFOIL

- **Comparison of wall-clock and CPU-times for HDM- and HPROM-based simulations**
  - HDM solution: 39.1 hours on 240 cores
  - HPROM solution: 32.5 minutes on 12 cores
  - offline training (on 240 cores): 5.84 minutes for subspace construction via POD + 28.4 minutes for mesh sampling

Wall-clock time speedup factor = 72.1
CPU-time speedup factor = 1,420

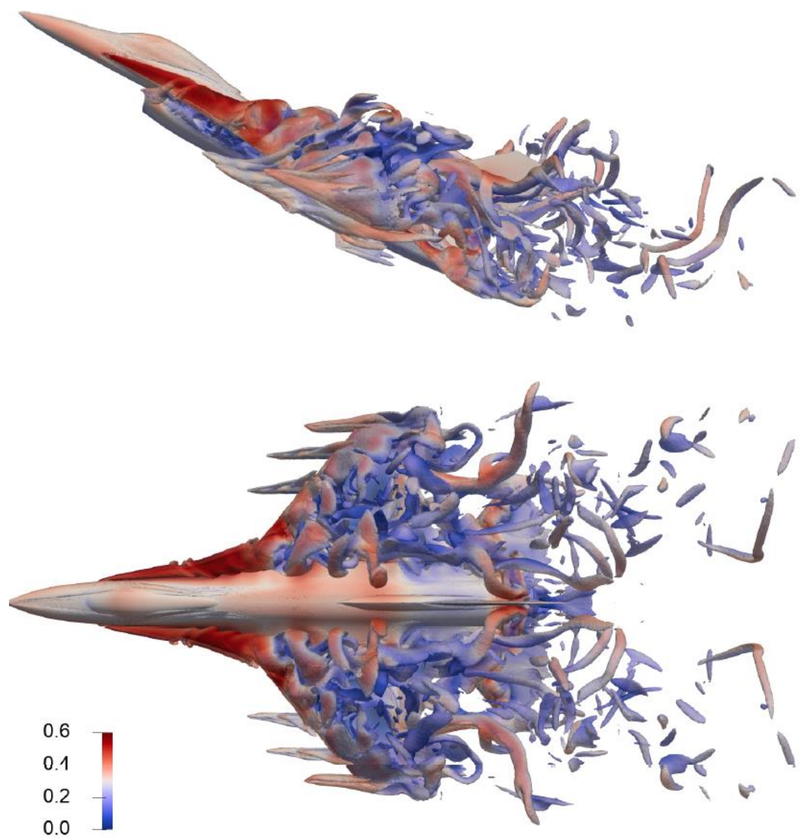
SHOWCASE APPLICATION #2: F-16 C/D BLOCK 40 AT HIGH ANGLE OF ATTACK

- **F-16 C/D Block 40 aircraft model at 30° angle of attack and 10,000 ft flight conditions, Re = 18,200,000, M_∞ = 0.3**
- High-dimensional model (HDM)
  - - ***DES*** turbulence model
  - - CFD mesh with 26,919,879 vertices and ***158,954,429 tetrahedral elements***
  - - dimension ***N = 161,519,274***
  - - time-discretization using second-order DIRK in time-interval $t \in [0, 1.29]$ s

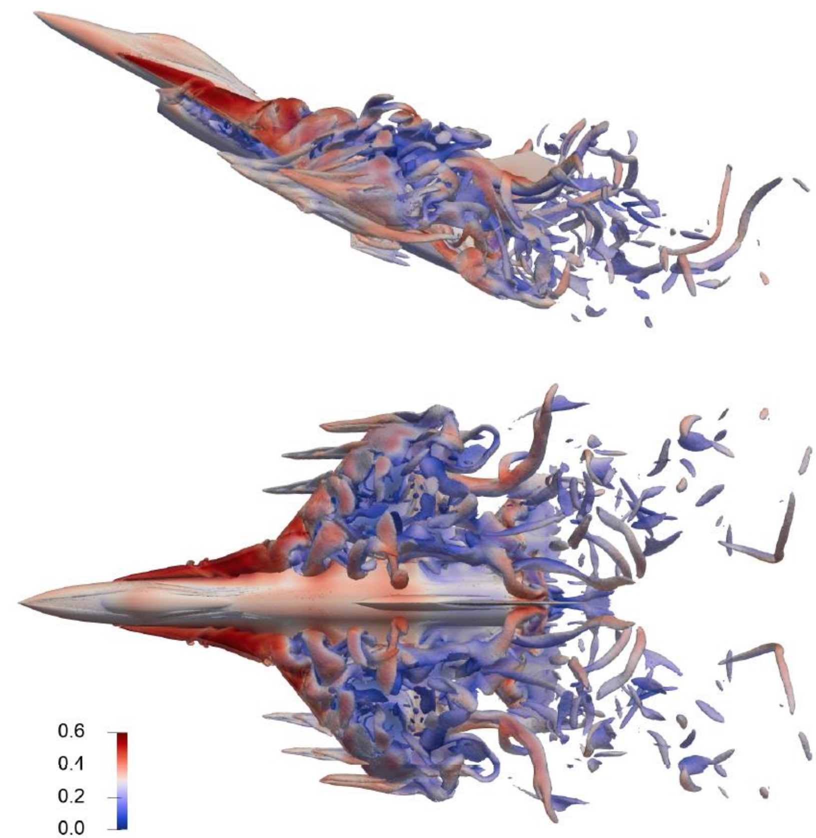
- PMOR and hyperreduction to obtain a hyperreduced PROM (HPROM)
  - - 5,001 solution snapshots collected
  - - local subspace approximation with N_c = ***50*** clusters
    - o $\boldsymbol{min_k\, n_k}$ **= 26**; $\boldsymbol{mean_k\, n_k}$ **= 54**; $\boldsymbol{max_k n_k}$ **= 115**
  - - data compression in each cluster using SVD and construction of local ROBs using $\varepsilon_{SVD} = 10^{-5}$
  - - hyperreduction using ***ECSW***, 21 training residual snapshots (every 250th computed time-step), $\varepsilon_{ECSW} = 10^{-2}$ ➡ reduced mesh with $\widetilde{N_e}$ = ***605*** elements only (0.0022%)!

SHOWCASE APPLICATION #2: F-16 C/D BLOCK 40 AT HIGH ANGLE OF ATTACK

- **F-16 C/D Block 40 aircraft model at 30° angle of attack and 10,000 ft flight conditions, Re = 18,200,000, M_∞ = 0.3**



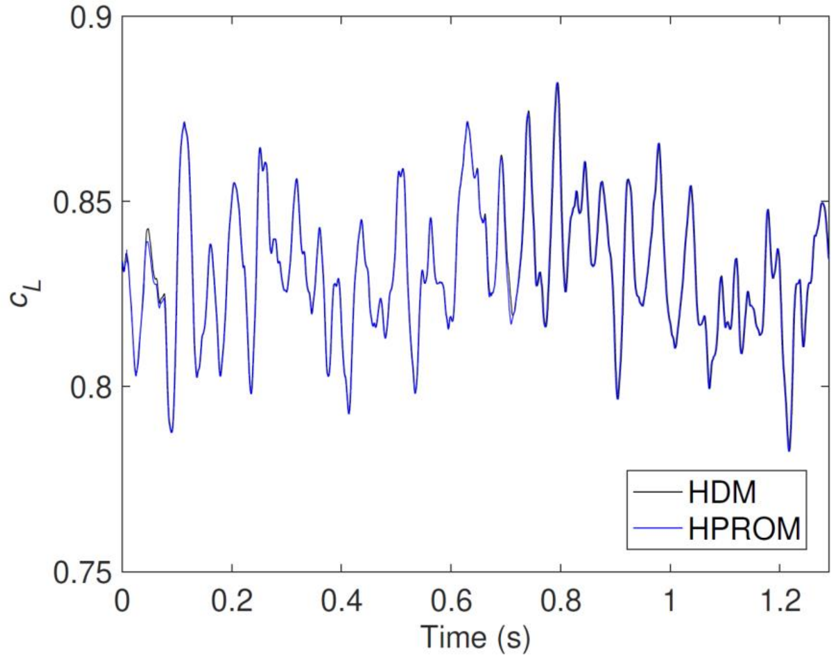
HDM

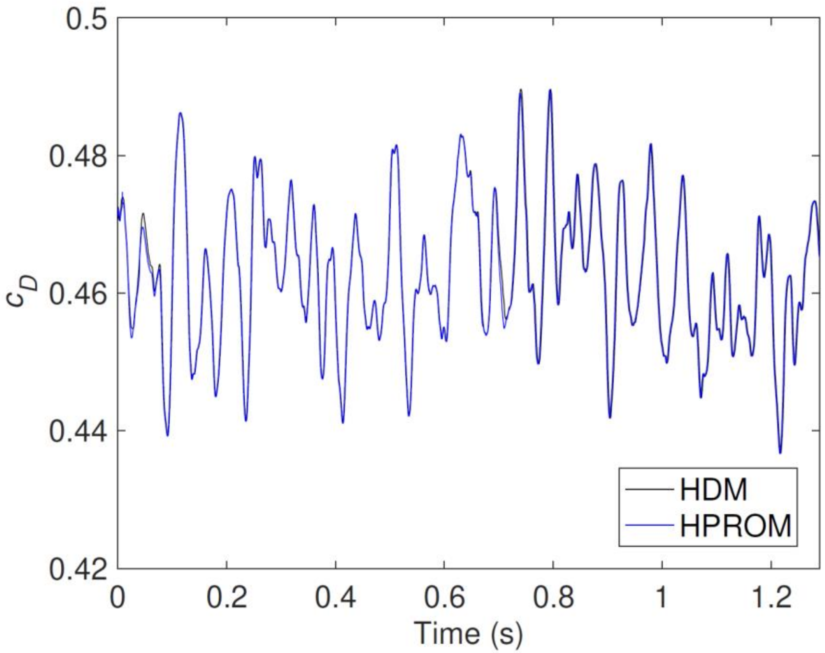


HPROM

SHOWCASE APPLICATION #2: F-16 C/D BLOCK 40 AT HIGH ANGLE OF ATTACK

- **F-16 C/D Block 40 aircraft model at 30° angle of attack and 10,000 ft flight conditions, Re = 18,200,000, M_∞= 0.3**



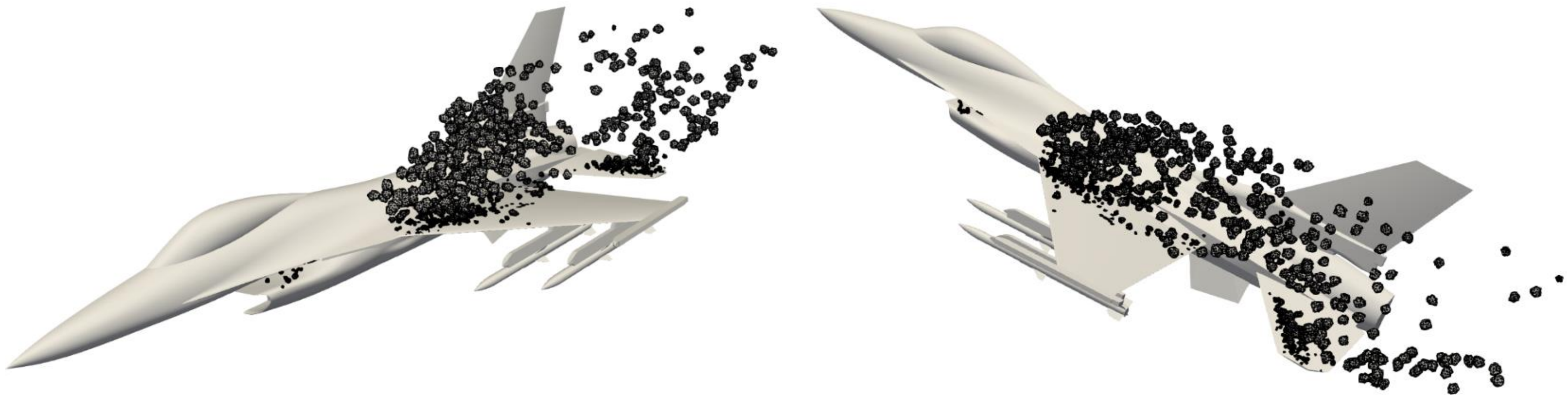


- stability, accuracy, and smoothness are achieved

SHOWCASE APPLICATION #2: F-16 C/D BLOCK 40 AT HIGH ANGLE OF ATTACK

- **F-16 C/D Block 40 aircraft model at 30° angle of attack and 10,000 ft flight conditions, Re = 18,200,000, M_∞ = 0.3**
  - *Frontera* supercomputer at the University of Texas at Austin

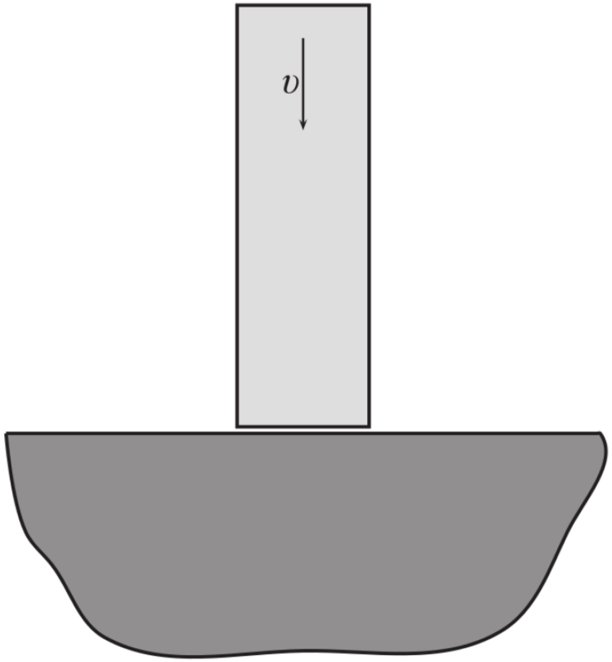
  - o HDM: ***100.3 hrs wall-clock time on 3,584 cores***
  - o ECSW (construction of reduced mesh): ***30.6 min wall-clock on 32 cores***



  - o HPROM: ***5.8 min wall-clock on 32 cores***
  - o ***speedup factor*** (wall-clock time) = ***1.04×10^3***
  - o ***speedup factor*** (CPU time) = ***1.17×10^5***

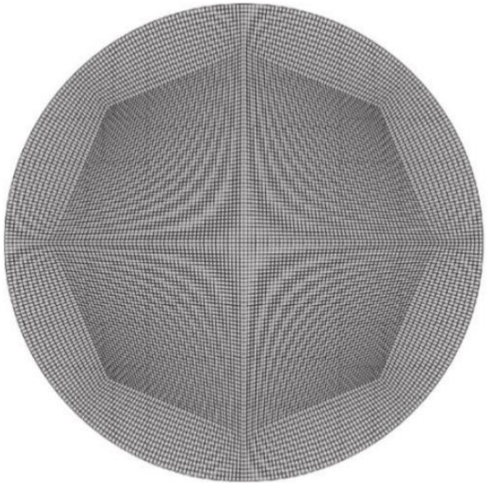
SHOWCASE APPLICATION #3: MULTISCALE MODELING OF THE TAYLOR IMPACT TEST PROBLEM

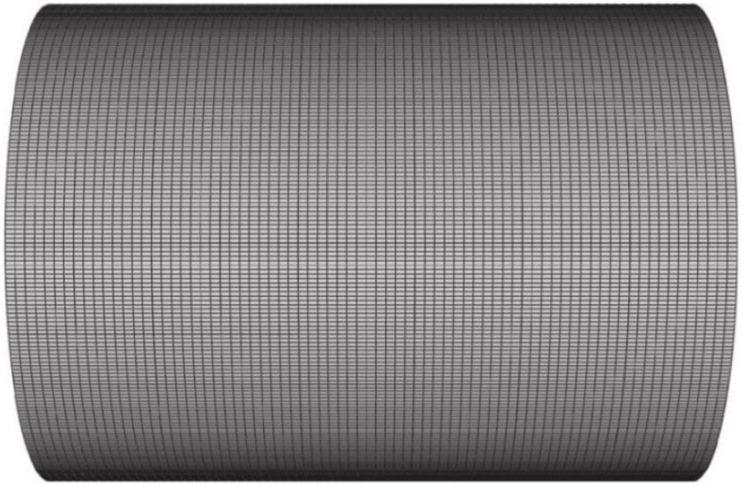
- **Approximate solution using most-appropriate LROB**
  - deformable cylinder travelling at a constant velocity and impacting a rigid anvil



SHOWCASE APPLICATION #3: MULTISCALE MODELING OF THE TAYLOR IMPACT TEST PROBLEM

- **Macroscale FE-based HDM**
  - cylinder diameter $= 9.52mm$, length $= 12.7mm$
  - initial velocity $v = 150$ m/s
  - 432,000 8-node hexahedral elements; 1,000 time-steps using explicit CD method with $\Delta t = 1 \times 10^{-9}s$
  - 3.456 billion constitutive relation evaluations





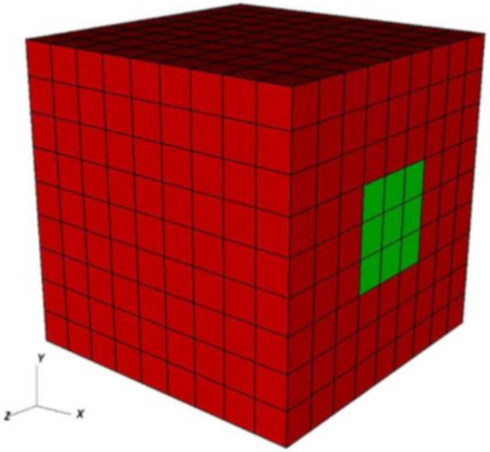
SHOWCASE APPLICATION #3: MULTISCALE MODELING OF THE TAYLOR IMPACT TEST PROBLEM

Microscale FE-based HDM

- fiber-reinforced composite
- both fiber and matrix modeled using a compressible neo-Hookean constitutive law
- 729 8-node hexahedral elements
- periodic boundary conditions (12 single-point constraints and 804 multi-point constraints)

Microscale PROM

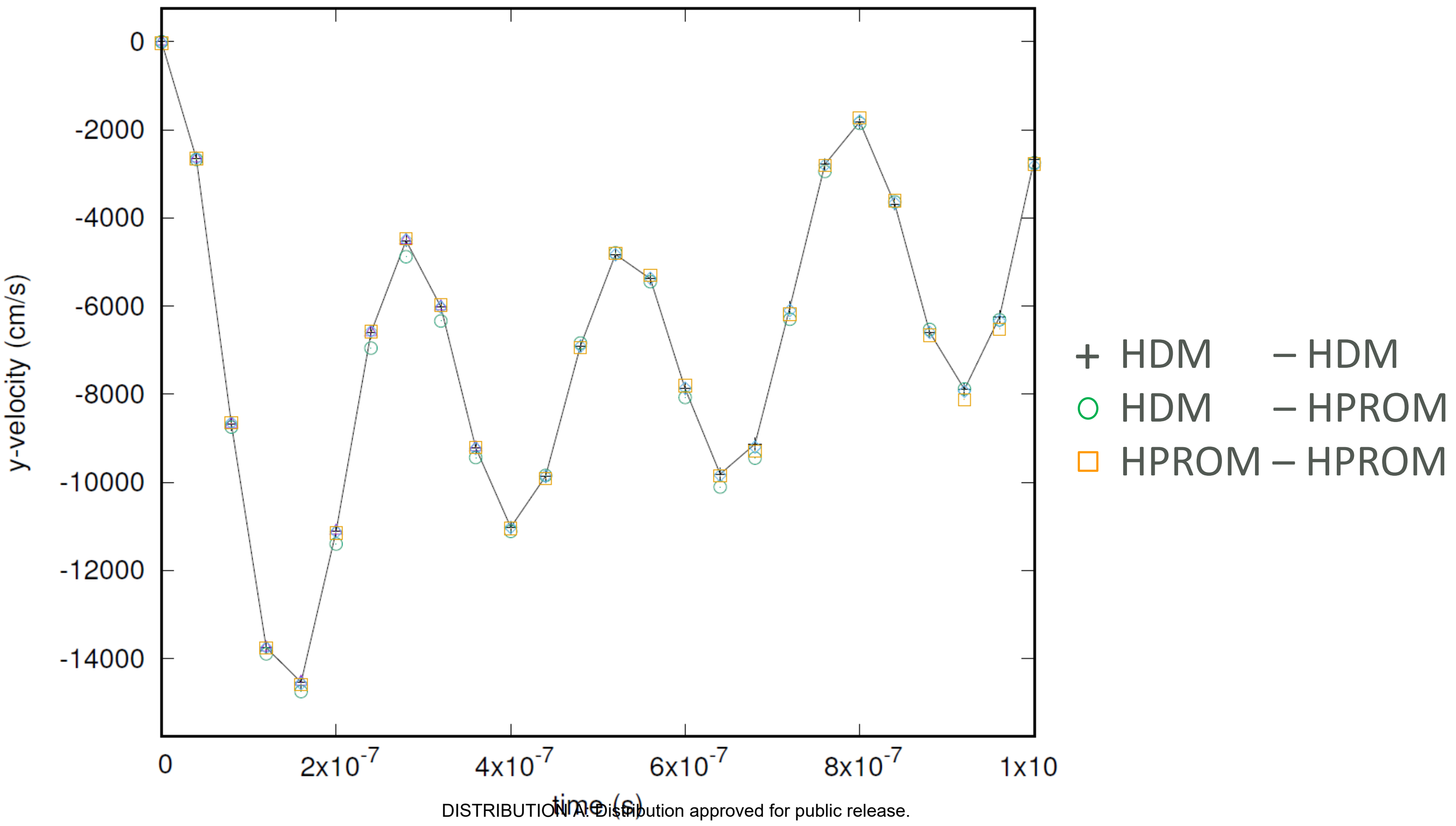
- ROB of dimension 9
- dual ROB of dimension 9
- reduced mesh with 7 sampled elements (ECSW)



FE mesh of the microscopic model: matrix (red) and fiber (green)

SHOWCASE APPLICATION #3: MULTISCALE MODELING OF THE TAYLOR IMPACT TEST PROBLEM

- **Velocity at a node (harder than displacement)**



SHOWCASE APPLICATION #3: MULTISCALE MODELING OF THE TAYLOR IMPACT TEST PROBLEM

- **Cray XC40 (Excalibur, ARL)**

HDM – HDM : 1,103 hrs wall-clock on 1,920 cores

HPROM – HPROM : ***7 hrs*** wall-clock on ***60 cores including training***

→ ***speedup factor = 144***