The Eleventh International Conference on
Engineering Computational Technology
23–25 August, 2022 | Montpellier, France

Proceedings of the Eleventh International Conference on
Engineering Computational Technology
Edited by B.H.V. Topping and P. Iványi
Civil-Comp Conferences, Volume 2, Paper 10.2
Civil-Comp Press, Edinburgh, United Kingdom, 2022, doi: 10.4203/ccc.2.10.2

Effective Physics Simulations based on Model Reduction and Domain Decomposition

Lin Jiang, Yu Liu and Ming-C. Cheng

Department of Electrical & Computer Engineering
Clarkson University, Potsdam, NY, USA

Abstract

This investigation implements domain decomposition in proper orthogonal decomposition (POD) to construct an effective multi-block methodology for physics simulations of engineering and scientific problems. To develop such a methodology, the structure of interest is first partitioned into smaller blocks, and solution data of each block are collected from detailed numerical simulation (DNS) accounting for parametric variations in the block. The collected data that represent the block are used to generate (or *train*) a set of basis functions (or POD modes) that are therefore tailored to the characteristics of the block accounting for its parametric variations. With the well-trained modes, the approach significantly reduces the degree of freedom (DoF) needed to reach an accurate solution. To construct a model for a larger domain, the trained POD blocks are then glued together using the discontinuous Galerkin method to enforce thermal continuity at the block interfaces. The multi-block concept further minimizes the computational effort in the training process and allows the POD methodology to offer efficient simulation models for large-scale structure with a high resolution, which may be crucial for many engineering and scientific applications.

The multi-block POD methodology has been applied to physics simulations in two distinct areas, including a prediction of the dynamic thermal distribution in a 2-block 3D semiconductor structure and simulation of a 3-block 1D quantum-well structure whose electron wave functions are governed by the Schrödinger equation. It has been illustrated that the POD methodology in both applications is able to offer very good agreement with the DNS results using just 3 or 4 POD modes in the 3D and 1D problems.

Keywords: physics simulation, reduced-order model, domain decomposition, proper orthogonal decomposition, data driven, thermal simulation, Schrödinger equation.

1 Introduction

Solutions of complex physics problems usually rely on direct numerical simulations (DNSs), such as finite-difference (FD) and finite-element (FE) method. Many commercial and open-source codes have been developed in recent decades, covering a wide range of scientific and engineering applications with good accuracy. These DNS tools are however computationally intensive, especially when a high resolution is needed. To minimize the computational effort, numerical degrees of freedom (DoF) needs to be substantially reduced. The common approach is to project the problem onto a functional space described by a finite set of orthogonal basis functions, such as the Fourier basis, wavelets, Bessel functions, Legendre Polynomials, Wannier functions, etc. Each set of these basis functions for a problem is however assumed based on its solution form, physical geometries and excitations but may not be optimal for the problem of interest.

To significantly improve the projection method, the basis functions need to be optimized for the particular problem of interest. Proper orthogonal decomposition (POD) [1,2] is an effective projection-based method that minimizes the DoF needed to reach an accurate solution via decomposition of solution data collected from DNS accounting for parametric variations. However, such a decomposition (training) process is problematic on 2 issues. Firstly, to collect the massive amount of data from DNS in a large-scale multi-dimensional structure, particularly with a high resolution, it is computationally prohibitive. Secondly, to account for enough variations of physical parameters and boundary conditions (BCs) in the training, time-consuming DNS needs to be performed many times to cover a wide range of operating/BC conditions, which is impractical for engineering design of a structure.

This work investigates an approach implementing domain decomposition in POD, where the domain is partitioned into subdomains (named elements or blocks). The training is then performed for each smaller block locally instead of the entire domain globally. This approach offers more efficient/thorough training for the POD basis functions (or modes). It also provides a more practical, flexible way of analysing and designing complex engineering or scientific problems based on the concept of building blocks. Use of building blocks has been an important practice for modern technology development, including semiconductor chips, nanostructures, material synthesis, metamaterials, etc. Two test cases of physics simulations using the multi-block POD methodology are presented below, including chip-level dynamic thermal simulation of a 2-block semiconductor structure and a quantum eigenvalue problem of a 3-block quantum well (QW) structure.

2 Methods

POD generates its modes $\eta(\mathbf{r})$ from solution data $Q(\mathbf{r},t)$ by maximizing the mean square inner product of the mode and the solution [1,2]. This leads to an eigenvalue problem,

$$\int_{\Omega'} \langle Q(\mathbf{r},t) \otimes Q(\mathbf{r}',t)\rangle \, \eta(\mathbf{r}')d\Omega' = \lambda \, \eta(\mathbf{r}), \tag{1}$$

where λ is the eigenvalue and the brackets indicate an average over many snapshots in time. In static problems, the "snapshots" are collected with variations of excitations and/or BCs. The heat transfer problem is discussed first below, followed by the Schrödinger equation.

With $Q(\mathbf{r},t)$ replaced by $T(\mathbf{r},t)$, temperature is expressed as a linear combination of the modes,

$$T(\mathbf{r},t) = \sum_{j=1}^{M} a_j(t) \, \eta_j(\mathbf{r}), \tag{2}$$

where *M* is the selected number of modes for the solution and $a_j(t)$ is the *temperature* in the POD space for each mode. To solve a_j, the heat transfer equation can be projected onto a POD space,

$$\int_{\Omega} \left(\eta \frac{\partial \rho C T}{\partial t} + \nabla \eta \cdot k \nabla T \right) d\Omega = \int_{\Omega} \eta P_d(\vec{r},t) d\Omega + \int_{\mathrm{S}} \eta \, k \nabla T \cdot d\mathbf{S}, \tag{3}$$

with *k* as thermal conductivity, P_d power density, ρ density, *C* specific heat and **S** the surface vector. Using Equation (2) in Equation (3), a set of coupled ordinary differential equations (ODEs) can be derived,

$$\sum_{j=1}^{M} c_{ij} \frac{da_j}{dt} + \sum_{j=1}^{M} g_{ij} \, a_j = P_{int,i} + P_{suf,i}, \quad \text{for } i = 1 \text{ to } M, \tag{4}$$

where the coefficients (c_{ij} and g_{ij}) and interior and boundary surface powers ($P_{int,i}$ and $P_{suf,i}$) are defined by the integrals in Equation (3) [3,4].

For a multi-block domain, using the discontinuous Galerkin method [5,6], Equation (3) becomes

$$\int_{\Omega} \left(\eta \frac{\partial \rho C T}{\partial t} + \nabla \eta \cdot k \nabla T \right) d\Omega =$$
$$\int_{\Omega} \eta P_d(\vec{r},t) d\Omega + \int_{\mathrm{S}} (\llbracket kT \rrbracket \langle \nabla \eta \rangle - \langle k \nabla T \rangle \llbracket \eta \rrbracket) \cdot d\mathbf{S} - \mu \int_{\mathrm{S}} \llbracket kT \rrbracket \llbracket \eta \rrbracket dS, \tag{5}$$

where the penalty factor $\mu = N_\mu/\Delta r$ with N_μ as the penalty number and Δr as the local mesh size, and $\langle \cdot \rangle$ and $\llbracket \cdot \rrbracket$ are the average and difference across the interface, respectively. Using Equation (2) for each block, an ODE matrix equation for an *N*-block domain can be derived. With the *n*-th block vector, $\boldsymbol{a}_n$ ($n = 1$ to N), determined, the *n*-th block temperature can be estimated in Equation (2).

In a quantum block surrounded by others, $Q(\mathbf{r},t)$ in Equation (1) is replaced by the electron wave function (WF) $\psi(\mathbf{r})$. Similar to Equation (5), the Schrödinger equation can be projected onto a set of POD modes for each block as

$$\int_\Omega \nabla\eta \cdot \frac{\hbar^2}{2m^*}\nabla\psi d\Omega + \int_\Omega \eta U\psi d\Omega - \oint_S \left(\left[\!\left[\frac{\hbar^2}{2m^*}\psi \right]\!\right] \langle \nabla\eta \rangle + \left\langle \frac{\hbar^2}{2m^*}\nabla\psi \right\rangle [\![\eta]\!] \right) \cdot d\mathbf{S}$$
$$-\mu \oint_S \left[\!\left[\frac{\hbar^2}{2m^*}\psi \right]\!\right] [\![\eta]\!] dS = E \int_\Omega \eta\, \psi\, d\Omega\,, \tag{6}$$

where m^* is electron effective mass, U potential energy and $\hbar$ the reduced Plank's constant. ψ in each block is expressed as a linear combination of its modes,

$$\psi(\mathbf{r}) = \sum_{j=1}^{M} a_j\, \eta_j(\mathbf{r}). \tag{7}$$

where a_j representing the time-independent Schrödinger equation in the POD space. Using Equation (7) in Equation (6), a multi-block quantum eigenvalue problem can be formulated for an (N×M)×(N×M) Hamiltonian in the multi-block POD space for an N-block domain each with M modes.

3 Results

For demonstration of POD on thermal analysis of a 2-block 3D domain consisting of Core-1 and L2-Cache in a CPU shown in Figure 1, uniform power is applied to each block with additional higher power to the red square area. Dynamic power is averaged over 120k cycles at 3.2GHz and assigned randomly. Heating thickness is 0.0558mm on the top. Adiabatic BCs are assumed except the bottom with constant heat flux.

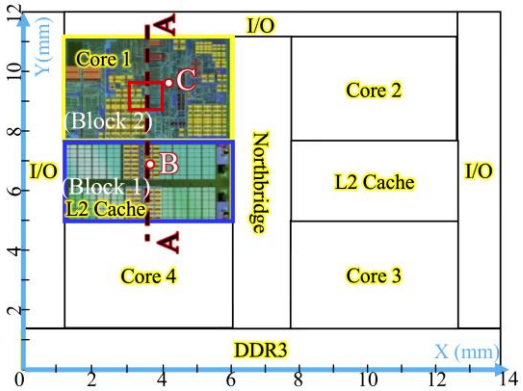

Figure 1: Floorplan of AMD-ATHLON-II X4-610e CPU with the 2-block domain, including L2-Cache (Block-1) and Core-1 (Block-2). Path A and Points B and C are the locations for the demonstrations below.

With POD modes generated for each block via data collected from an FE method, 2-block POD thermal model is validated against FE simulation. Figure 2 shows that

least square errors (LSEs) with 4 modes in the entire domain and heat layer are near 2.48% and 1.78%, respectively, at $N_\mu \approx$ 40-80. Dynamic temperatures at Points B and C are illustrated in Figures. 3(a) and 3(b), and the profile along Path A in Figure 3(c), where results with 4 modes from these 2 approaches nearly overlap except near the temperature peak.

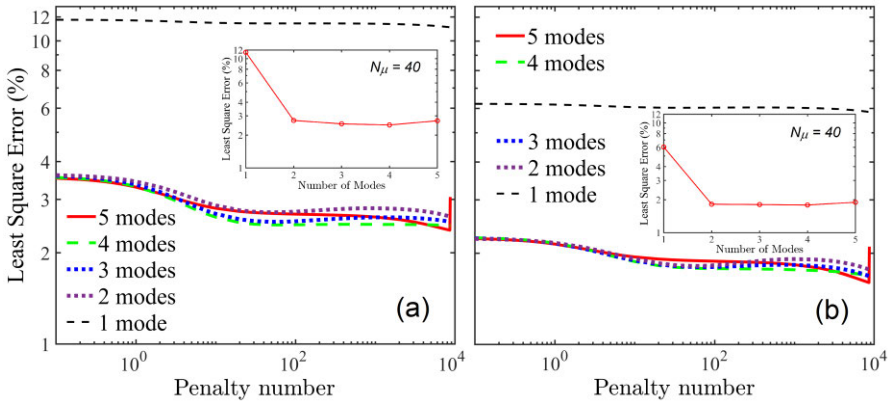

Figure 2. LSEs in the (a) entire domain and (b) heating layer.

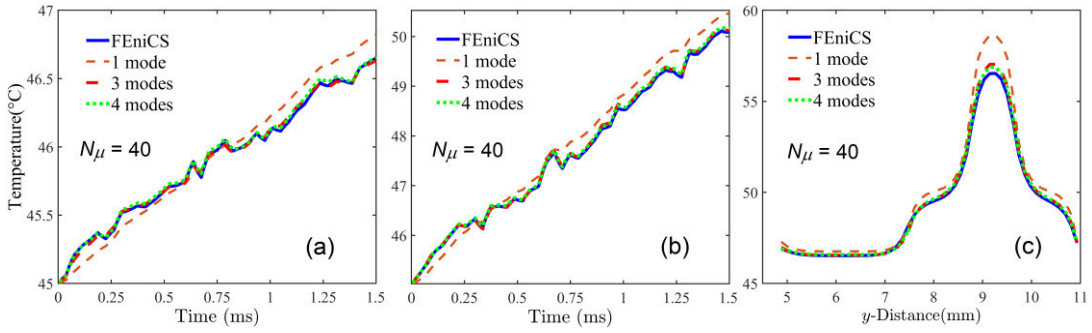

Figure 3: Dynamic temperatures at (a) Point B and (b) Point C and temperature distribution along (c) Path A.

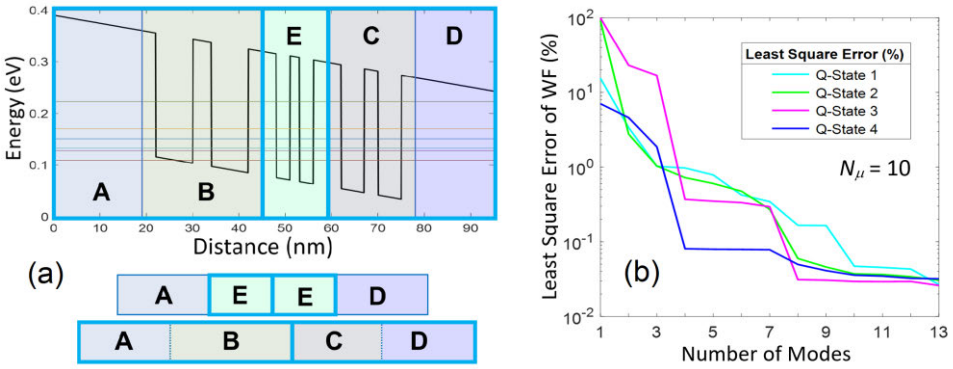

Figure 4: (a) Top: QW structure for the demonstration. Bottom: 2 QW structures for the training. (b) LSEs of WFs in QSs 1-4 over the entire domain.

The quantum POD simulation is demonstrated in a 6-QW structure, shown in Figure 4(a), partitioned into 3 blocks labelled as AB-E-CD. FD Schrödinger simulations of the A-E-E-D and A-B-C-D domains, shown in Bottom of Figure 4(a), are performed separately to collect WF data to generate POD modes for Blocks AB, E and CD. The demonstration in Figure 4(b) shows that an LSE of the 3-block POD model below 1%, compared to the FD simulation, is achieved with just 3 modes for the first 2 quantum states (QSs), and for QSs 3 and 4, 4 modes are needed. WFs from the POD model with 3 or 4 modes shown in Figures 5(a) and 5(b) for QSs 1 and 3 are in excellent agreement with the FD simulation.

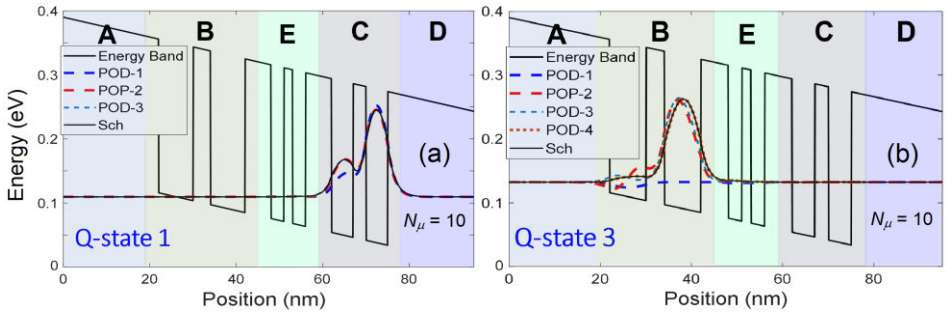

Figure 5: Energy band diagram and WFs in the (a) first and (b) third quantum states.

4 Conclusions and Contributions

POD has been found effective for physics simulations in different applications [3,4,7-10]. However, it has been limited to small/simplified structures due to computational effort needed for the training, including data collection, mode generation and calculations of model parameters. For a large-scale structure, this becomes prohibitive. For example, for thermal-aware task scheduling for CPUs/GPUs [11-14], the scheduling relies on information on hot spots. Currently, the thermal-aware CPU/GPU scheduling is almost all based on efficient RC thermal circuits with large RC elements, which very likely misestimate the peak values and locations of small-diameter hot spots. Although the POD approach offers an effective method with fine-enough resolution to capture all the hot spots, it is computationally intensive to train the large-scale CPU/GPU structure to develop a POD model.

Implementing domain decomposition in POD, the multi-block POD methodology offers an efficient alternative for training smaller blocks individually. With the multi-block approach, a more thorough training can be performed to generate a more accurate and robust set of POD modes for each block. In addition, a greater flexibility can be achieved if building blocks for the system are trained. A set of POD building blocks for a specific technology can then be stored in a database for design/simulation of a larger structure. The building-block concept is one of the major reasons for efficient engineering design and fast growth of modern technologies. Fast advance in multi-core CPUs/GPUs is a typical example of utilizing the building blocks, including cores, caches, memory units, streaming microprocessors and many other functional units. The concept of the POD building blocks can be implemented, for example, in

thermal-aware design exploration of GPU/CPU floorplans [15-18] that is a crucial step to minimize the heating issues and improve their reliability and performance. Another example is the use of building blocks for design/synthesis of nanostructures and materials, which requires efficient and accurate solution of the Schrödinger or Schrödinger-like equations to facilitate calculations of electronic structure. The quantum multi-block POD simulation methodology, if successfully implemented in multi-dimensional domain, will be useful for such applications.

This work has validated the great efficiency and the high accuracy of the multi-block POD simulation methodology in 2 different areas of applications. With the improved efficiency for the POD training provided by the multi-block approach, effective multi-physics multi-dimensional simulations enabled by the multi-block POD methodology will become possible in the near future.

Acknowledgements

This work is supported by the National Science Foundation under Grant No. ECCS-2003307.

References

[1] J. L. Lumley, "The structure of inhomogeneous turbulent flows", in Atmospheric Turbulence and Radio Wave Propagation, 166-177, 1967.

[2] G. Berkooz, P. Holmes, and J. L. Lumley, "The proper orthogonal decomposition in the analysis of turbulent flows", Annu. Rev. Fluid Mech., 25, 539–575, 1993.

[3] W. Jia, B.T. Helenbrook, Ming-C. Cheng, "Fast Thermal Simulation of FinFET Circuits Based on a Multi-Block Reduced-Order Model", IEEE Trans. CAD ICs & Systems, 35, 1114-1124, 2016.

[4] W. Jia, B.T. Helenbrook, Ming-C. Cheng, "Thermal Modeling of Multi-Fin Field Effect Transistor Structure Using Proper Orthogonal Decomposition", IEEE Trans. Electron Devices, 61, 2752-2759, 2014.

[5] D.N. Arnold, F. Brezzi, B. Cockburn, D. Marini, “Discontinuous Galerkin methods for elliptic problems”, Lecture Notes in Comp. Sci. Eng., 89-101, 2000.

[6] D. N. Arnold, F. Brezzi, B. Cockburn, L. D. Marini, “Unified analysis of discontinuous Galerkin methods for elliptic problems”, SIAM J. Numer. Anal., 39, 1749-1779, 2002.

[7] D. Ahlman, F. Soderlund, J. Jackson, A. Kurdila and W. Shyy, “Proper orthogonal decomposition for time-dependent lid-driven cavity flows”, *Numer. Heat Transfer, Part B.* 42, 285–306, 2010.

[8] P. A. LeGresley and J. J. Alonso. 2001. Investigation of non-linear projection for POD based reduced order models for aerodynamics. *Proc. 39th Aerosp. Ind. Assoc. Amer. Aerosp. Sci. Meeting Exhibit. (AIAA)*, 1–15, Reno, NV, Jan. 2001.

[9] K. Lu, Y. Jin, Y. Chen, Y. Yang, L. Hou, Z. Zhang, Z. Li and C. Fu, “Review for order reduction based on proper orthogonal decomposition and outlooks of applications in mechanical systems”, *Mech. Sys. Signal Proce.*, 123, 264–297, 2019.

[10] D. Binion and X. Chen, “A Krylov enhanced proper orthogonal decomposition method for frequency domain model reduction”, *Eng. Computations*, 34, 285–306, 2017.
[11] H. Pourmeidani, A. Sharma, K. Choo, M. Hassan, M. Choi, K. Kim, B. Jang, “Dynamic Temperature Aware Scheduling for CPU-GPU 3D Multicore Processor with Regression Predictor”, *J. Semicond. Tech. & Sci.*, 18, 115-124, 2018
[12] R. Nath, R. Ayoub, T. S. Rosing, “Temperature aware thread block scheduling in GPGPUs”, *50th ACM/EDAC/IEEE Design Automation Conf. (DAC)*, Austin, TX, 1-6, 2013.
[13] M. Niknafs, I. Ukhov, P. Eles and Z. Peng, “Runtime Resource Management with Workload Prediction”, Proc. 56th Annual Design Automation Conf. (DAC '19). Article No. 169, 2019.
[14] O. Benedikt *et al*., “Thermal-Aware Scheduling for MPSoC in the Avionics Domain: Tooling and Initial Results”, *2021 IEEE 27th Int. Conf. Embedded and Real-Time Computing Systems and Applications (RTCSA)*, 159-168, 2021.
[15] Y. W. Wu, C. L. Yang, P. H. Yuh and Y. W. Chang, "Joint exploration of architectural and physical design spaces with thermal consideration", ISLPED '05. Proc. Int. Symp. Low Power Electronics & Design, 123-126, San Diego, CA, 2005.
[16] P. Chaparro, J. Gonzáles, G. Magklis, Q. Cai and A. González, "Understanding the Thermal Implications of Multi-Core Architectures", in IEEE Tran. Parallel & Distributed Sys., 18, 1055-1065, 2007.
[17] I. Arnaldo, A. Vicenzi, J. L. Ayala et al., "Fast and scalable temperature-driven floorplan design in 3D MPSoCs", 13th Latin American Test Workshop (LATW), Quito, Ecuador, 1-6, 2012.
[18] K. Manna, M. Jimson, “Thermal-aware Design Strategies for the 3D NoC-based Multi-Core Systems”, *Design and Test Strategies for 2D/3D Integration for NoC-based Multicore Architectures*. Springer, Cham, 111-123, 2020.