An Effective Physics Simulation Methodology Based on a Data-Driven Learning Algorithm

Lin Jiang
Clarkson University
Potsdam, New York, USA
jiangl2@clarkson.edu

Martin Veresko
Clarkson University
Potsdam, New York, USA
vereskm@clarkson.edu

Yu Liu
Clarkson University
Potsdam, New York, USA
yuliu@clarkson.edu

Ming-C. Cheng
Clarkson University
Potsdam, New York, USA
mcheng@clarkson.edu

ABSTRACT

A methodology of multi-dimensional physics simulations is investigated based on a data-driven learning algorithm derived from proper orthogonal decomposition (POD). The approach utilizes numerical simulation tools to collect solution data for the problems of interest subjected to parametric variations that may include interior excitations and/or boundary conditions influenced by exterior environments. The POD is applied to process the data and to generate a finite set of basis functions. The problem is then projected from the physical domain onto a mathematical space constituted by its basis functions. The effectiveness of the POD methodology thus depends on the data quality, which relies on the numerical settings implemented in the data collection (or the training). The simulation methodology is developed and demonstrated in a dynamic heat transfer problem for an entire CPU and in a quantum eigenvalue problem for a quantum-dot structure. Encouraging findings are observed for the POD simulation methodology in this investigation, including its extreme efficiency, high accuracy and great adaptability. The models constructed by the POD basis functions are even capable of predicting the solution of the problem beyond the conditions implemented in the training with a good accuracy.

CCS CONCEPTS

• **Computing methodologies** → **Modeling and simulation**.

KEYWORDS

Model order reduction, proper orthogonal decomposition, thermal simulation of CPU, quantum-dot simulation

ACM Reference Format:
Lin Jiang, Martin Veresko, Yu Liu, and Ming-C. Cheng. 2022. An Effective Physics Simulation Methodology Based on a Data-Driven Learning Algorithm. In *Proceedings of the Platform for Advanced Scientific Computing (PASC22).* ACM, New York, NY, USA, 10 pages. https://doi.org/XXXXXXX.XXXXXXX

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
PASC22, June 27–29, 2022, Basel, Switzerland
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

A vast amount of mathematical and scientific problems rely on partial differential equations (PDEs) to offer numerical solutions and to understand the insight into their system behaviors. PDEs are also responsible for development and design of many modern technologies in the fields of automobiles, aeronautics, energy, power, materials, semiconductors, nanostructures, drugs, etc. Phenomena described by the PDEs needed in these industries are mainly relevant to the conservation laws or wave propagation. The former includes, for example, heat flow, fluid dynamics, particle transport, etc., and the latter includes sounds, electromagnetic waves, elastodynamic waves, quantum eigenvalue problems, etc.

Many commercial tools and open-source simulation platforms, offering multi-physics simulations using direct numerical simulations (DNSs) based on finite difference, finite element and finite volume methods, have been developed in the last several decades. These DNS tools generally offer accurate numerical solutions for the above-mentioned problems. However, because of a large number of degrees of freedom (DoF) needed in the DNS methods, these well-developed simulation tools are usually computationally intensive and demand extensive computational resources, especially when a fine resolution and/or dynamic information are needed.

To overcome the drawbacks inherited in the DNS methods and to effectively utilize the DNS resources developed for decades, in this study we investigate a data-driven learning technique for different simulation problems to significantly reduce the numerical DoF for the problems of interest with a desired accuracy. The data-driven learning algorithm is based on a reduced order modeling technique derived from the proper orthogonal decomposition (POD)[7, 27] that has been shown in many areas of research[2, 5, 8, 11, 14, 21, 23, 24, 26, 35] to be effective and accurate if the data collected in the training process is adequate for the problem of interest.

The POD projects the problem of interest onto a mathematical space described by a finite set of orthogonal basis functions (hereafter named POD modes). To generate an optimal set of POD modes in space via the decomposition, solution data from the PDEs for the problem are needed from DNSs of the structure governed by the PDEs. The DNSs need to cover a range of parametric variations to train the POD modes to adopt the effects of various excitations and

boundary conditions (BCs). This is different from many other orthogonal projection-based methods, such as Fourier basis, wavelets, Legendre polynomials, Bessel functions, etc., whose basis functions are selected (i.e., assumed) based on the solution characteristics influenced by excitations, BCs and/or geometries. The POD modes are generated from the training process (or data collection) and specifically tailored to the geometry and parametric variations of the problem. The POD is thus able to substantially minimize the DoF needed in the simulation of PDEs with a high accuracy. The approach does not require a priori assumptions on the physical phenomena, geometry, BCs, etc. Its accuracy is however dependent on the quality of the collected data (influenced by the training settings), the selected number of modes (i.e., DoF) used in the physics simulation and the computer precision.

In this work, we investigate the applications of the POD methodology to two distinct physics simulation problems, including dynamic heat transfer in a CPU and a quantum eigenvalue problem of a quantum dot (QD) structure. The first application concerns an accurate and efficient prediction of thermal gradients and hot spots that are desperately needed for thermal management of semiconductor chips, thermal-aware task scheduling for high performance computing CPUs and GPUs, thermal-aware design exploration of chip floorplans, etc.[6, 19, 25, 28, 29, 32, 34, 36, 39]. The second application demonstrates an effective methodology to solve the Schrödinger equation for electron wave functions (WFs) in quantum structures. Effective simulations for solutions of the quantum eigenvalue problems (or the Schrödinger equation) are also desired for a wide range of engineering and scientific areas involving physical dimensions as small as the electron wavelength or molecular/nuclear scales. These subjects cover, for example, materials, physics, electronics, photonics, biology, medicines, chemistry, etc.[1, 3, 4, 9, 10, 12, 16–18, 22, 30, 33, 37, 38, 44–46].

This work demonstrates the initial effort for developing an effective data-driven methodology for multi-physics simulations. The POD simulation methodology for dynamic heat transfer problems is first presented in the next section, followed by the POD quantum eigenvalue problem described by the a Hamiltonian in a POD space. The POD methodology is then applied to analyze the dynamic thermal distribution over an entire CPU and predict the WFs in various quantum states (QSs) in a quantum-dot structure. The former is excited by dynamic power dissipation initiated by CPU computing. The latter is influenced by external electric field. The investigation demonstrates the efficiency, accuracy and adaptability of the POD methodology for solution prediction of multi-dimensional problems within and beyond the training conditions.

2 BACKGROUND OF POD FOR PHYSICS SIMULATIONS

POD generates a set of basis functions from several sets of data samples collected from the PDE for the problem of interest over the entire simulation domain. This is done by seeking a POD mode $\eta(\vec{r})$ that maximizes its mean square inner product with the solution $Q(\vec{r},t)$[7, 27].

$$\left\langle \left(\int_\Omega Q(\vec{r},t)\eta(\vec{r})d\Omega \right)^2 \right\rangle \Big/ \int_\Omega \eta(\vec{r})^2 d\Omega, \tag{1}$$

where Ω is the physical domain and the angled brackets $\langle\rangle$ indicate the average of the solution data ensemble collected over many sets of numerical observations accounting for the parametric variations. Using the maximization process in (1) to find a POD mode, it ensures that the component projected onto the mode $\eta(\vec{r})$ contains the maximum least squares (LS) information of the system behavior in the data ensemble. In the space orthogonal to this mode, the maximization process is performed again to generate the second mode. Repetition of the maximization process in this fashion leads to an eigenvalue problem given by the Fredholm equation [7, 27],

$$\int_{\Omega'} \left\langle Q(\vec{r},t) \otimes Q(\vec{r}',t) \right\rangle \vec{\eta}(\vec{r}')d\Omega' = \lambda\vec{\eta}(\vec{r}), \tag{2}$$

where $\otimes$ is the tensor product and λ is the POD eigenvalue of the two-point correlation matrix and represents the mean squared Q variations captured by its POD mode. The value of λ reveals the importance of the corresponding mode. The data Q in (2) are usually collected from the DNSs of the PDE for the problem. For discrete data, (2) involves solving an $N_r \times N_r$ eigenvalue problem, where N_r is the number of grid points in the domain. In a multi-dimensional problem, this requires an intensive computational effort, especially for problems with a high resolution. The method of snapshots [20, 40–42] is thus applied to transform the $N_r \times N_r$ problem to an $N_s \times N_s$ eigenvalue problem with a considerably smaller dimension determined by the number of samples (i.e., snapshots N_s).

Once the POD modes are found, the physical quantity $Q(\vec{r},t)$ can be described by a linear combination of the POD modes,

$$Q(\vec{r},t) = \sum_{j=1}^{M} a_j(t)\eta_j(\vec{r}), \tag{3}$$

where M is the selected number of modes representing the solution ($1 \le M \le N_s$), a_j are weighting coefficients responding to the parametric variations, and the POD modes are normalized.

Depending on the problem of interest, a set of equations for a_j is derived by projecting its governing PDE onto a functional space constituted by its POD modes. This imposes a clear physics-based guideline on the POD modes to comply with the PDE and thus to offer an accurate and efficient prediction for the simulation problem. The POD models for the dynamic heat transfer and quantum eigenvalue problems are formulated below.

2.1 Heat Transfer Equation

For dynamic thermal problems, heat transfer is described by the heat transfer equation,

$$\frac{\partial \rho C T(\vec{r},t)}{\partial t} - \nabla \cdot k\nabla T(\vec{r},t)) = P_d(\vec{r},t), \tag{4}$$

where k is the thermal conductivity, P_d the power density, ρ the density and C the specific heat. To develop a POD model, the dynamic heat transfer equation can be projected onto a POD space along the generated POD modes using the Galerkin projection method. This gives rise to the weak form of the heat transfer equation,

$$\int_\Omega (\eta \frac{\partial \rho C T}{\partial t} + \nabla\eta \cdot k\nabla T)d\Omega = \int_\Omega \eta P_d d\Omega + \int_S \eta(k\nabla T)\cdot d\vec{S}, \tag{5}$$

where $\vec{S}$ the surface vector on the boundary of the domain Ω. After inserting (3), with $Q(\vec{r},t)$ replaced by $T(\vec{r},t)$, into (5), a set of M

ordinary differential equation can be derived from (5) for a_j,

$$\sum_{j=1}^{M} c_{i,j}\frac{da_j(t)}{dt} + \sum_{j=1}^{M} g_{i,j}a_j(t) = P_{pod,i}, \text{ for } i = 1 \text{ to } M, \quad (6)$$

where $c_{i,j}$ and $g_{i,j}$ are elements of the thermal capacitance and conductance matrices in the POD space and given by

$$c_{i,j} = \int_\Omega \rho C \eta_i \eta_j d\Omega, \text{ and } g_{i,j} = \int_\Omega k\nabla\eta_i \cdot \nabla\eta_j d\Omega, \quad (7)$$

and $P_{pod,i}$ represents the projected power density in Ω and heat flux on the surface along the ith POD mode,

$$P_{pod,i} = \int_\Omega \eta_i P_d(\vec{r},t)d\Omega - \int_S \eta_i(-k\nabla T)\cdot d\vec{S}. \quad (8)$$

With $a_j(t)$ determined by (6), $T(\vec{r},t)$ can be predicted from (3) as a linear combination of $\eta_j(\vec{r})$, and thus the resolution of the POD solution is as fine as the DNS. Since the eigenvalue represents the mean squared temperature variations captured by each POD mode, the theoretical LS error over the entire domain and all the snapshots for the M-mode POD model can be estimated from

$$Err_{LS} = \sqrt{\frac{\sum_{i=M+1}^{N_s}\lambda_i}{\sum_{i=1}^{N_s}\lambda_i}}. \quad (9)$$

This error prediction is however only reasonable if good quality of the collected data is guaranteed with enough computer precision. To ensure good data quality, it is twofold. Firstly, the solution data need to be accurate and consistent with the governing PDE. Secondly, the training needs to cover as much as possible a range of parametric variations which the simulation will encounter.

2.2 Schrödinger Equation

n a quantum structure, the electron WF $\Psi(\vec{r})$ is determined by the Schrödinger equation,

$$\nabla\cdot\left[-\frac{\hbar^2}{2m^*}\nabla\Psi(\vec{r})\right] + U(\vec{r})\Psi(\vec{r}) = E\Psi(\vec{r}), \quad (10)$$

where $\hbar$ is the reduced plank constant, m^* the effective mass, $U(\vec{r})$ the potential energy and E the total energy of the electron. Similar to the heat transfer problem, the Schrödinger equation is projected onto the POD modes using the Galerkin projection method. The projection leads to the weak form of the Schrödinger equation,

$$\int_\Omega \nabla\eta_i\cdot\frac{\hbar^2}{2m^*}\nabla\Psi d\Omega + \int_\Omega \eta_i U\Psi d\Omega - \int_S \eta_i\frac{\hbar^2}{2m^*}\nabla\Psi\cdot d\vec{S} = E\int_\Omega \eta_i\Psi d\Omega. \quad (11)$$

After substituting (3) with $Q(\vec{r},t)$ replaced by $\Psi(\vec{r})$ into (11), an $M\times M$ eigenvalue problem for $\vec{a}$ in the POD space is expressed as,

$$\boldsymbol{H}_\eta\vec{a} = E\vec{a}, \quad (12)$$

where $\boldsymbol{H}_\eta$ is the Hamiltonian in POD eigenspace given as

$$\boldsymbol{H}_\eta = \boldsymbol{T}_\eta + \boldsymbol{U}_\eta + \boldsymbol{B}_\eta, \quad (13)$$

with the interior kinetic energy matrix expressed as

$$\boldsymbol{T}_{\eta,i,j} = \int_\Omega \nabla\eta_i(\vec{r})\cdot\frac{\hbar^2}{2m^*}\nabla\eta_j(\vec{r})d\Omega, \quad (14)$$

the potential energy matrix expressed as

$$\boldsymbol{U}_{\eta,i,j} = \int_\Omega \eta_i(\vec{r})U(\vec{r})\Psi(\vec{r})d\Omega, \quad (15)$$

and finally, the boundary kinetic energy matrix expressed as

$$\boldsymbol{B}_{\eta,i,j} = \int_S \eta_i(\vec{r})\frac{-\hbar^2}{2m^*}\nabla\Psi(\vec{r})\cdot d\vec{S}. \quad (16)$$

For the homogeneous Neumann and Dirichlet BC, the boundary kinetic energy matrix vanishes.

3 DEMONSTRATION OF POD SIMULATION METHODOLOGY

3.1 Thermal Simulation of a CPU

The POD simulation methodology is first demonstrated in dynamic thermal analysis of a quad-core CPU, AMD ATHLON II X4 610e, whose floorplan is shown in Figure 1 [13]. Units of the quad-core processor include four 512KB L2 caches, a northbridge in the center and I/O and DDR3 placed in the periphery. Dynamic DNS in FEniCS [15], an open computational platform based on the finite-element method (FEM), is performed to generate thermal data to train the POD modes for the quad-core processor. The size of the simulation domain for this processor is 14 mm×12 mm×650 μm in the x, y and z directions. The dynamic power is assumed uniform in each unit and dissipated only on the top 150 μm layer of the chip. All the surfaces of the chip are assumed adiabatic except the bottom surface where the convection BC is implemented with a heat transfer coefficient calculated form the structure dimensions and material properties with an ambient temperature of 45℃. The total power applied to the CPU is approximately 50W and the dynamic power density in each unit in the DNS is averaged over 48,000 CPU cycles (~ 13.7 μs) at 3.5 GHz with the percentage of the power consumption in each unit similar to the power map in [13]. The power density over each average period in each unit is generated randomly. A simulation over 4.1 ms with a mesh of $128\times128\times13$ is performed to collect the training thermal data at a time step over 8,000 clock cycles.

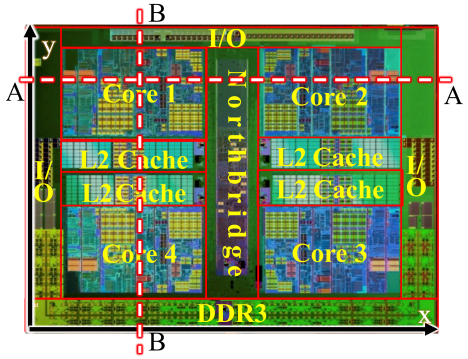

Figure 1: Floorplan of the selected CPU. The dashed lines indicate the plotting paths for the temperature profiles in the demonstration. The intercept of 2 lines is at (3.8mm, 9.8mm).

Applying the method of snapshots [20, 40–42] to (2), the eigenvalues and POD modes are generated using the training data obtained

from the 3D dynamic DNS of the CPU chip. With the POD modes, POD model parameters or coefficients in (6) can then be determined. The eigenvalue spectrum is shown in Figure 2, where the eigenvalue drops nearly four orders of magnitude from the first to the third mode and more than 4 orders to the fourth and fifth modes. This indicates that the first few POD modes are able to capture the mean squared temperature variations effectively and thus only a small number of POD modes is needed to predict the temperature solution with a good accuracy, as indicated by (9).

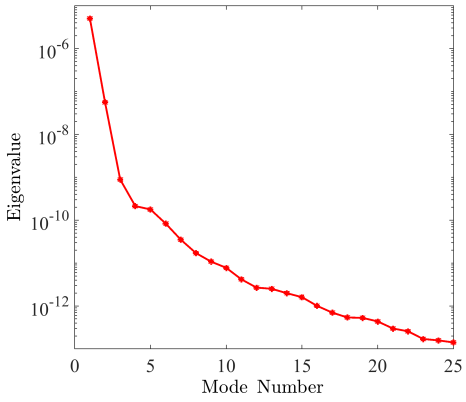

Figure 2: Eigenvalue spectrum for the thermal data.

To verify the validity of the training process, the training error is estimated by performing the POD simulation of the CPU using the same dynamic power map in the training. Figure 3 shows that the LS training error is near 1.32%, 1.2% or 1.1% with 3, 5 or 7 modes, respectively, in the heating layer. However, the training error in the entire chip is larger and becomes 3.2%, 3.15%, 3.03% or 2.9% with 3, 5, 7 or 8 modes, respectively. This is because the power is dissipated in the thin heating layer (23% of the chip volume in our study), and the temperature in most of regions in the thick substrate below the heating layer is close to the ambient, where the percentage error tends to be larger. Although the idea error estimated from (9) continues declining as more modes are included, the training error becomes nearly invariant beyond 4 modes due to the computer precision and numerical fluctuation in the higher modes.

To demonstrate the developed POD thermal model for the CPU, a dynamic power density in each unit is initiated by a different random sequence from that used in the thermal data collection. However, the power density percentage distribution over all the units remains similar. To test the robustness of the model, LS test errors are estimated over 3 different simulation times including 4.1 *ms*, 6.2 *ms* and 10.3 *ms* even though the training was only carried out for a period of 4.1 *ms*. Figure 3 shows that, as the number of modes increases, the LS test error for the 4.1 *ms* case is slightly greater than the LS training error and stays near 3.2% in the entire chip and 1.25%-1.27% in the heating layer with 5 or more modes.

When testing the POD model for a period of 6.2 *ms*, beyond the 4.1 *ms* training time, the LS error is slightly greater than the 4.1 *ms*-case LS error below 5 modes in the entire CPU or the heating layer. With 5 or more modes, the 6.2 *ms* LS error becomes nearly identical to the 4.1 *ms*-case error. This is evidently different from the general understanding of the deep/machine learning methods that usually work well in the interpolation case but fail in the case of extrapolation. This stems from the fact that, as mentioned previously, the POD simulation methodology offers a clear guideline for its solution to follow, based on the physical principle enforced by the projected governing equation. Thus, the POD model still works accurately in the untrained range when using more modes. When it goes far beyond the training time, such as the 10.3 *ms* case shown in Figure 3, the LS test error with 5 modes increases from 3.2% (for the 4.1 *ms* case) to 5% in the entire chip and 1.25% to 2% in the heating layer. Its LS error continues declining toward the 4.1 *ms*-case error as the number of modes increases, and it is as small as 1.63% in the heating layer and 3.1% in the entire chop with 8 modes. The results suggest that the POD methodology performs reasonably well in a extrapolation situation with more modes included unless the situation is very far beyond the training conditions.

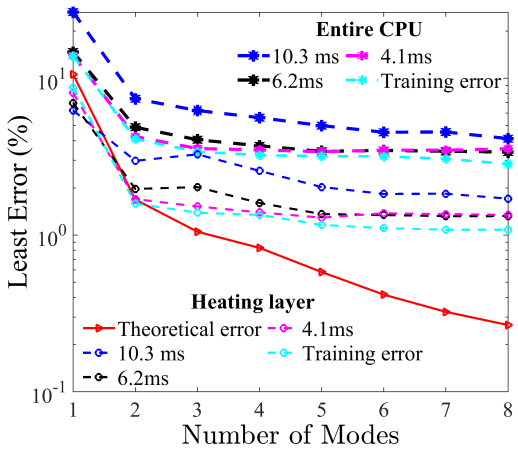

Figure 3: LS percentage errors with respect to the ambient.

A comparison of dynamic temperatures at the location of (3.8 *mm*, 9.8 *mm*) in the CPU (see Figure 1) derived from the POD thermal model and FEniCS-FEM is illustrated in Figure 4 for simulation over a period of 6.2 *ms*. Temperature distributions along Paths A and B at 6.2 *ms* are also shown in Figures 5 and 6, respectively. In general, more POD modes used in the simulation lead to a better agreement with FEniCS-FEM. However, it is interesting to observe that the predicted temperature by 5 modes in Core 1 near the I/O unit is actually slightly more accurate than that by 7 modes, as shown in Figure 5. In addition, it is shown in Figure 6 that in Core 4 the 3-mode POD model offers better agreement with FEniCS-FEM than the 5 modes model. Overall, the maximization process given in (1) minimizes the LS error over the entire simulation domain and time, as shown in Figure 3, instead of the local error.

To illustrate the variation of the test error induced by different random sequences for each unit in the dynamic power map, a close-up view of the LS error with error bars for the case of 6.2 *ms* is displayed in Figure 7. The variation is larger in the entire CPU than in the heating layer. The variation of the error evaluation shrinks when more modes are included and becomes as small as 0.055% and 0.037% for the entire CPU and the heating layer, respectively, with 5 or more modes for a confidence level of 95%.

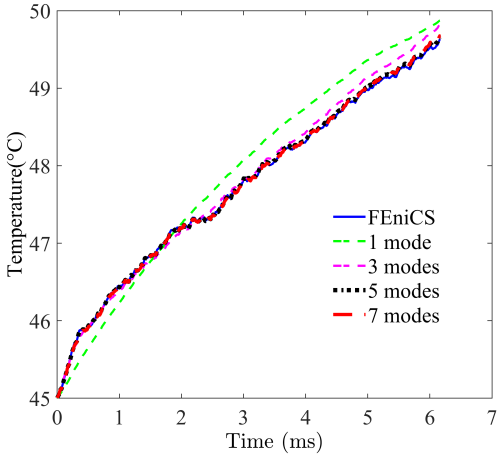

Figure 4: Dynamic temperature at the intercept of Paths A and B indicated in Figure 1.

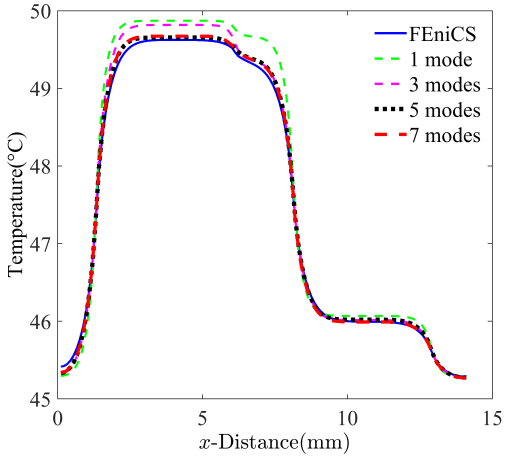

Figure 5: Temperature distribution along Path A at 6.2 *ms*.

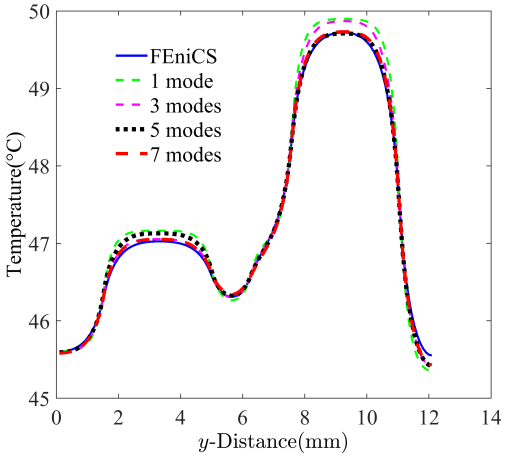

Figure 6: Temperature distribution along Path B at 6.2 *ms*.

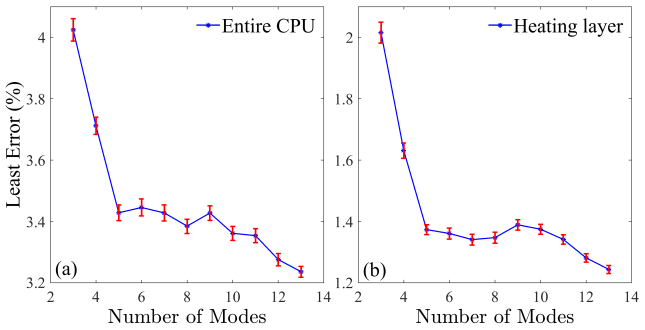

Figure 7: The LS test error with error bars for the 6.2 *ms* **case.**

Our investigation reveals that use of a finer resolution in DNS to improve the quality of the collected data for POD mode generation leads to more accurate numerical calculations for the gradients of modes needed in (7) for $g_{i,j}$. This thus improves the POD accuracy and leads to a smaller LS error but increases the training time needed to collect data, to generate POD modes and to evaluate the model coefficients in (6). We also found that the eigenvalue spectrum is not influenced by the spatial resolution (once the resolution is reasonably fine) except for the very high mode number if the number of snapshots is large enough. This implies that, to reach a similar accuracy, the DoF needed in the simulation using the POD model derived from finer-resolution modes is equal to or even smaller than the model derived from coarser-resolution modes.

3.2 Simulation of a Quantum-Dot Structure

The second demonstration of the POD simulation methodology involves a 3×3 grid of GaAs/InAs QDs displayed in Figure 8, where effective mass in GaAs $m^*_{GaAs} = 0.067m_o$ and InAs $m^*_{InAs} = 0.023m_o$ and the band off $\Delta E = 0.544eV$ with m_o as the free electron mass. To generate the POD modes for the QD structure, WF data are collected from DNS of the Schrödinger equation using a finite difference method, subjected to different applied electric fields. The training data are collected in response to a total of 8 single electric fields separately from each of the two orthogonal directions in x and y with each component spaced evenly between $-40kV/cm$ and $+40kV/cm$. Namely, there are 8 snapshots of WFs in each direction. With an additional sample at zero field, a total of 17 snapshots of WFs are collected. To train POD modes, only the WFs in the first 6 QSs are collected at each electric field and there are 17×6 sets (snapshots) of WFs used in the training. Because of the degenerate QSs (see Table 1), a very small grid size of 0.1 nm is used along x and y in simulations for the data collection to offer accurate degenerate states in DNSs and thus in the POD simulation.

With the 102 sets of WF training data, the method of snapshots [20, 40–42] is applied to generate the POD modes and eigenvalues. The eigenvalue spectrum of the WF data is illustrated in Figure 9, which is significantly different from the spectrum for the thermal problem shown in Figure 2 because of the distinct characteristics embedded in the collected data between these 2 problems. Unlike the thermal problem where the eigenvalue decreases rapidly in the first few modes, the POD modes in the QD structure contain behaviors of WFs in all the 6 selected QSs. To maximize the mean

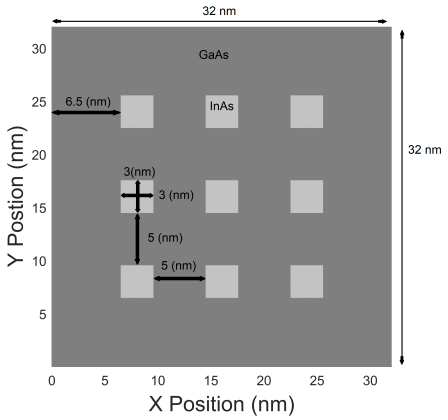

Figure 8: Simulation domain of 2D QDs.

Table 1: Quantum Eigenenergy and POD Percentage Error with respect to the Lowest Band Energy

Quantum State	POD QS Energy(eV)	DNS QS Energy(eV)	POD Error(%)
1	0.419948	0.419004	0.225081
2	0.433175	0.432143	0.238516
3	0.439549	0.438534	0.231015
4	0.447263	0.446110	0.258153
5	0.452561	0.451459	0.243761
6	0.459061	0.457970	0.238099
7	0.466329	0.465109	0.261987
8	0.471819	0.470642	0.249776
9	0.485228	0.483937	0.266492

square variations of WFs of all the selected states guided by (1), the eigenvalue remains nearly unchanged for the first six modes.

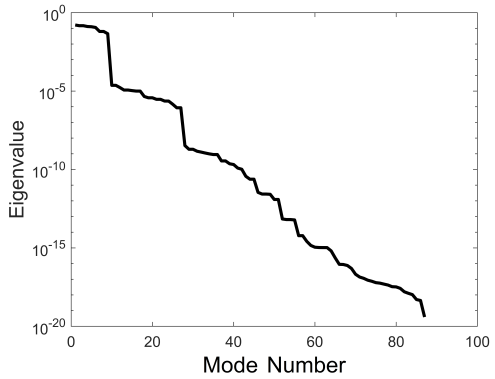

Figure 9: Eigenvalue spectrum.

The training error of the WFs in the trained 6 QSs is presented in Figure 10 for each field direction at 20 kV/cm. LS training errors below 1% in all 6 QSs can be achieved with 10 or 11 modes. In QS 4, a training error below 0.7% can be reached with 9 modes, and an error below 0.5% can be obtained the QS 2 with 15 or more modes at either field direction.

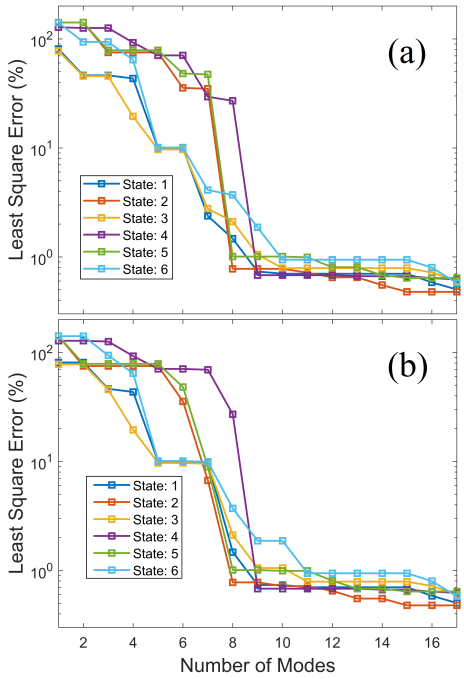

Figure 10: LS training errors in the first 6 QSs with applied electric field of 20 kV/cm in the (a) x and (b) y directions.

While each set of the collected data for the training was obtained from the DNS at a field in either x or y direction, in the demonstration an electric field combining the 2 orthogonal components, $\vec{E} = (20\hat{x} - 10\hat{y})\ kV/cm$, is applied. The LS test errors of the WFs predicted by the POD approach shown in Figure 11 decrease gradually as more modes are added. In general, the test error is slightly greater than the training error. The test errors for all the QSs suddenly drop significantly from 8 to 9 modes. The test errors for QSs 1-6 stay near 0.7%-1.2% with 10 modes and become as small as 0.7%-0.9% when 16 modes are included. It is interesting to observe that, even though QSs 7-9 were not included in the training, the quantum POD model are still able to predict the WFs in QSs 7-9 with a good accuracy. As shown in Figure 11, the LS test errors in QSs 7-9 are near 1.5%-2.23% with 9-11 modes and continue decreasing with more modes. For QSs 7 and 8, the error as small as 1%-1.4% is observed with 15 or more modes.

To observe the error statistics induced by field variation, the LS test error with error bars in each of the 6 trained QSs is shown in Figure 12 for fields varying within the training range. In general for each QS, the error variation is small for the first few modes and rises when using 7 or 8 modes. The variation however drops suddenly when the number of modes increases from 8 to 10 modes and remains near 0.1% to 0.27% beyond 10 modes.

The training in the quantum POD approach involves WF data from a few selected QSs, and the quantum POD model thus predicts the WF solutions in all the selected states subjected to a applied

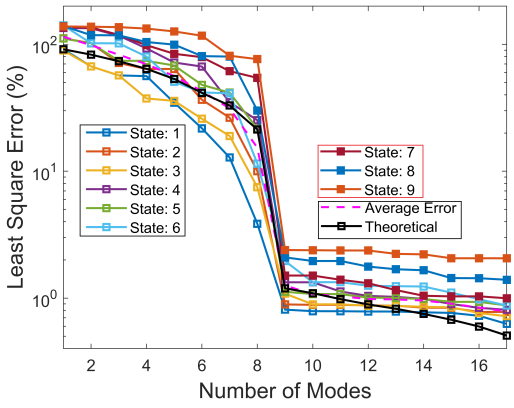

Figure 11: LS test error in each QS compared to the theoretical LS error estimated by (9). The average error shown in the dashed line denotes the LS average of the selected (six) QSs.

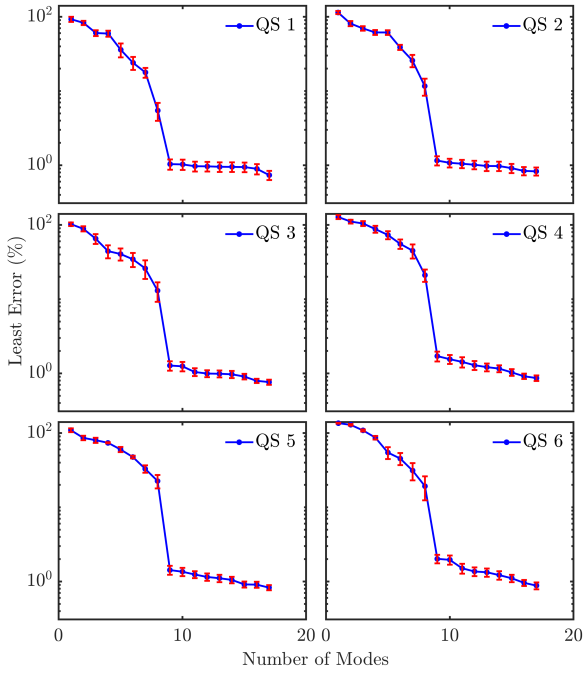

Figure 12: The LS test error with error bars for the first 6 QSs.

electric field. This is different from the thermal POD methodology that offers one thermal solution in the CPU structure subjected to a dynamic power map. The expression of the LS error given in (9) can therefore be used to approximately estimate the POD numerical error for the heat transfer problem provided that the data quality and computer precision are reasonably good. (9) is however not able to predict the POD numerical error of the WF in each individual state because the eigenvalue represents the WFs in all selected QSs. Instead, as shown in 11, (9) reasonably predicts the LS average of the LS errors over all the trained QSs.

The comparison of the quantum eigenenergies derived from the POD methodology and the DNS is included in Table 1, together with the percentage error of the POD prediction with respect to the lowest band energy. Apparently, QSs 2-3 and QSs 5-6 are nearly degenerate with energy difference of 6.39 meV and 6.5 meV, respectively. Using a grid size as small as 0.1 nm in the POD mode training, the quantum POD methodology accurately predicts these degenerate states. Moreover, a very accurate prediction of the QS energy is also achieved even for the untrained 7th - 9th QSs.

The profiles of $|\Psi|^2$ along x and y directions in the first 3 QSs solved from the quantum POD methodology are shown Figure 13, compared to the DNS results. The contours of $|\Psi|^2$ in QSs 1-3 are given in Figure 14, where the plotting paths along the x and y directions for $|\Psi|^2$ in Figure 13 are indicated. The first POD mode predicts the mean of the WF data sets used in the training and thus offers the unbiased and thus symmetric WF in each state, as shown in Figure 13. The inclusion of more POD modes gradually improves the accuracy of the POD WF in each state. With 9 or 10 modes, the POD WF in each of QSs 1-3 nearly overlaps that from the DNS of the Schrödinger equation for the QD structure. Similarly, the profiles and contours of $|\Psi|^2$ in QSs 4-6 derived from the quantum POD methodology are also illustrated in Figures 15 and 16, respectively, compared to those obtained from the DNS.

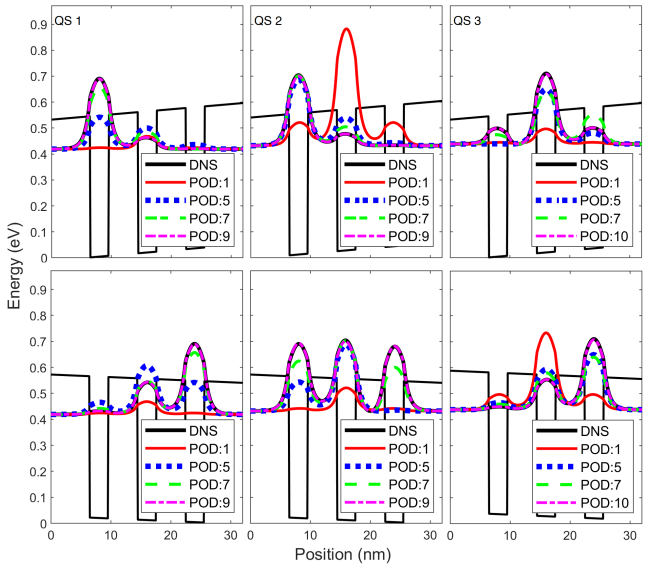

Figure 13: $|\Psi|^2$ in the first 3 QSs of the 2D QDs predicted by the POD method compared with the DNS of the Schrödinger equation along x (top row) and y (bottom row) directions. The paths along x and y are indicated in Figure 14

To further test the robustness of the quantum POD model beyond the training range of electric fields, Figure 17 illustrates the LS test error of WFs subjected to an applied field of $\vec{E} = (25\hat{x} + 50\hat{y})$, where WF data in QSs 7-9 were not included in the training. In general the LS error is only slightly greater than those shown in Figures 11 when more modes are included. For example, in the trained QSs 1-6, errors near 2.1%-3.5% are observed with 9 modes in Fig. 17. To

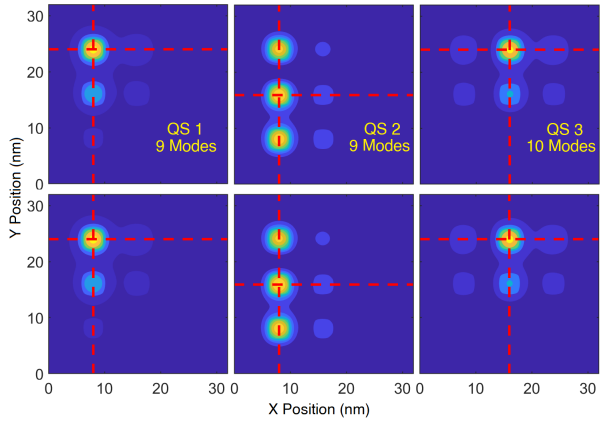

Figure 14: Contours of $|\Psi|^2$ in QSs 1-3 predicted by the POD method (top row) and DNS (bottom row). The red dashed lines indicate the plotting paths for Figures 13

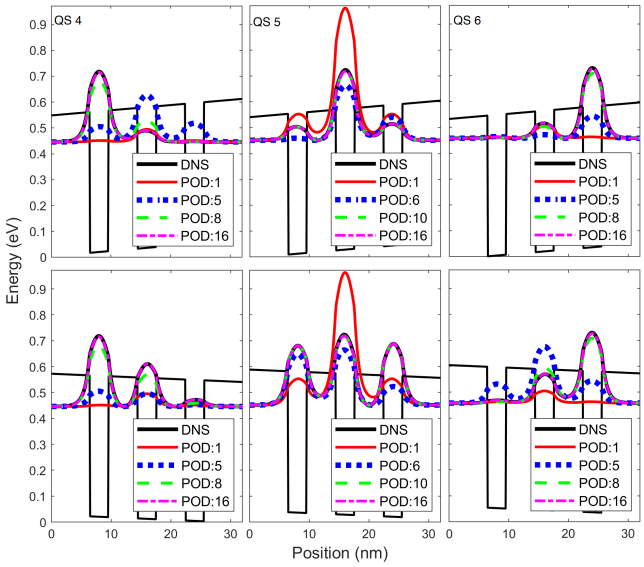

Figure 15: $|\Psi|^2$ in QSs 4-6 of the 2D QDs predicted by the POD method compared with the DNS of the Schrödinger equation along x (top row) and y (bottom row) directions. The paths along x and y are indicated in Figure 16.

reduce the error near or below 1.5% for all 6 selected QSs at such a high field, 15 or 16 modes are needed. It should also be noted that, for untrained QSs 7-8, use of 14 modes is able to keep the LS error near or below 2%; however, for the untrained 9th QS, it needs 15 or more modes to achieve 3.2%. Although the error is greater at applied field beyond the training range, the accuracy is still reasonably good even for the untrained QSs 7 - 9, as shown in figure 17, and the errors of QSs 1 - 8 still decrease as more modes are included. This indicates that, in the case of the extrapolation at electric field beyond the training range, the POD prediction is still reasonably accurate for the trained and untrained QSs, similar to the

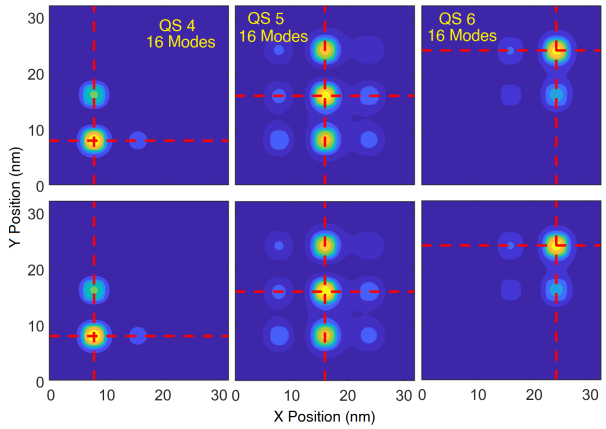

Figure 16: Contours of $|\Psi|^2$ in QSs 4-6 predicted by the POD method (top row) and DNS (bottom row). The red dashed lines indicate the plotting paths for Figures 15.

observation in the thermal prediction for the case of extrapolation by extending the simulation time beyond the training time.

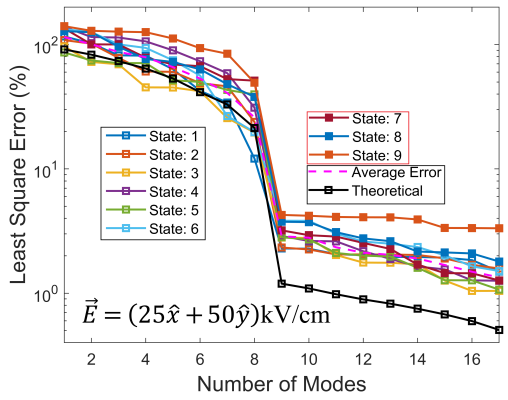

Figure 17: LS test error of WFs at an electric field whose y component is beyond the training field. The average LS error only include the selected (six) QSs.

4 DISCUSSIONS

Many learning algorithms only offer the training to repeat the system behavior instead of predicting the responses induced by unpredictable conditions, excitations or disturbances imposed on the system. These learning methods can be applied only to slowly varying behaviors with perhaps continuous learning to adapt the changes in the system behavior induced by the above-mentioned parametric variations. For scientific or engineering problems governed by PDEs, alternative learning methods are desired to adapt the changes effectively induced by the unpredictable parametric variations without continuous learning. Encouraging findings have been observed in the demonstrations of the POD simulation methodology in the dynamic thermal analysis of a CPU chip and in the prediction of electron WFs in a QD structure. The methodology derived from

Table 2: Computational Time for Thermal Simulations of CPU with Different Numbers of Modes

Simulation Time (ms)	Computational Time (s)						
	FEniCS FEM	No. of POD modes					
		1		3		5	
		ODEs	Post1/ Post2	ODEs	Post1/ Post2	ODEs	Post1/ Post2
4.1	6.78e3	0.11	0.07/ 1.10	0.11	0.15/ 2.08	0.11	0.27/ 3.16
6.2	1.02e4	0.15	0.11/ 1.28	0.16	0.23/ 2.68	0.16	0.40/ 4.78
10.3	1.69e4	0.26	0.18/ 2.14	0.28	0.38/ 4.51	0.30	0.69/ 8.05

the POD learning process for physics simulations has shown to be extremely efficient, very accurate and highly adaptable in response to the parametric variations near the training conditions.

In addition, most learning methods work well in the interpolation cases but fail to offer reasonable solution in the cases of extrapolation. Differently from other learning methods, the projection of the governing equation onto the POD modes imposes a clear physics-based guideline for the POD model to follow. The POD is thus able to offer a good prediction in situations far beyond the training condition if more modes are included. For the heat transfer problem, this was observed when the POD model trained for 4.1 ms was tested for a simulation of 10.3 ms, which leads to an LS error near 3% in the heating layer with 3 modes and 2% with 5 modes. For the QD problem, the POD modes trained for first 6 QS WFs were able to accurately predict the WFs in the untrained 7th and 8th QSs at an applied field beyond the training range, where an LS error of 2% can be reached with 14 modes. For the untrained 9th QS, when using 15 or more modes, a reasonable error near 3.2% can be achieved. Moreover, in the QD case, the training was carried out separately in 2 orthogonal directions, and yet the quantum POD model is able to accurately predict the WFs and eigenenergies at an electric field that combines these 2 orthogonal fields.

The demonstration of thermal analysis shows that the LS error with 5 POD modes is as small as 1.25% or 3.2% in the heating layer or the entire CPU, respectively. This leads to a 42,600-time reduction in the DoF and a significant speedup in computing. The DNS is performed in FEniCS-FEM using an iteration method based on the generalized minimal residual method [31]. The computing time for the POD simulations, compared to FEniCS-FEM, is given in Table 2, where Post1 and Post2 present the post-processes in (3) needed to calculate temperature for the heating layer and the entire chip, respectively. The ODEs provide the solutions for a_j in (6). If 5 modes are used to evaluate dynamic temperature in the heating layer of 4 cores, a reduction of more than 17,000 times in computing time is achieved (including the time needed for the ODE solution and post processing), compared to FEM simulation, and 2,000 times in the entire CPU. In realistic applications, temperature is needed only near the higher temperature areas (i.e., the 4 cores), which offers an improvement in the computing time over 4 orders of magnitude, compared to FEniCS-FEM. These are estimated using only one core in both approaches on Dell T7920 Dual Xeon 5122 CPUs.

The QD simulations for both DNS and POD approaches are performed in Matlab on an i7-8650U CPU. To reduce simulation time and memory space in DNS of the Schrödinger equation, the Krylov–Schur method [43] for sparse eigenvalue problems is applied. In the POD simulation, the LS error with 10 POD modes for the first 3 QSs is as small as 0.8%-0.9% and near 1%-1.2% for QSs 4-6. Because the fine mesh is needed to estimate the accurate degenerate state energies in the DNSs, the 2D POD QD simulation with 10 modes offers a reduction in DoF by 4 orders of magnitude. Compared to the dynamic heat transfer DNS, the sparse eigenvalue problem in DNS is significantly more efficient. The computing time for the POD QD simulation with 6 to 14 modes for the first 6 QSs is included in Table 3. The 10-mode POD simulation for all 6 QS WFs and eigenenergies plus the post-processing calculation offers a reduction in computing time by nearly 130 times. Realistically, when using WFs for further calculations, e.g., electron density, current density or kinetic energy, such a fine resolution is usually not needed, and coarser-resolution WFs can then be calculated in (3) to minimize the computational time and memory space. For example, if one uses only 1/3 or 1/2 of the spatial resolution in each direction, the speedup would be 3 orders of magnitude over the DNS.

Table 3: Computational Time for Simulations of QD Structure

Approach		No. of POD modes				
		6	8	10	12	14
POD Time (ms)	Eigenvalue Solver	0.0543	0.0586	0.0609	0.0729	0.0748
	Post Processing	10.247	13.977	16.679	18.803	23.385
DNS Eigenvalue Solver time(ms)	2154.4969					

5 CONCLUSIONS

This investigation proposes to develop compact but accurate models by using solution data derived from multi-physics DNS tools for computationally intensive scientific and engineering problems governed by PDEs. The data are processed via the POD to generate an optimal set of basis functions (or POD modes), and the PDE for the problem is then projected onto a mathematical space represented by the generated POD modes. This POD simulation methodology have been applied to 2 distinct physics problems, including 3D dynamic heat transfer in a CPU and a quantum eigenvalue problem in a 2D QD structure. The former analyzes the dynamic thermal profile over the entire CPU chip influenced by random heat excitations and BCs. The latter predicts the electron waves in a QD structure subjected to an applied electric field.

It has been demonstrated that the POD methodology offers very efficient predictions for these two physics simulation problems with a high accuracy within and beyond the training conditions. Our investigation also reveals that, in situations far beyond the training conditions, the POD methodology is still able to offer a reasonably accurate solution if more POD modes are added. In this work, the POD methodology offers a reduction in DoF by nearly 4 orders of magnitude in both problems and leads to a tremendous saving in computational time. To reach an accurate thermal prediction near the training condition, a speedup of 2,000 can be achieved for the entire CPU, compared to the DNS. In practical situations where only thermal information in the high temperature areas is needed, a

speedup over 17,000 times can be reached. For the 2D QD problem, a reduction in computing time over 130 times can be realized. The POD predictions offer spatial resolutions as fine as the DNSs.

It is worthwhile to point out that a finer resolution can be used in DNSs to improve the accuracy of the collected data and thus the quality of the training. Although this increases the computational effort in the training, use of the POD model with finer-resolution modes does not increase the DoF in simulation to reach a similar accuracy, compared to a coarser-resolution model. A practice for reaching a highly accurate POD model with improved simulation efficiency can therefore be carried out as follows. After performing POD simulation with a fine resolution POD modes, the spatial solution in (3) does not need to be calculated at every grid point. The prediction of the solution with a coarser spatial resolution or in some selected local regions may be enough in many applications.

ACKNOWLEDGMENTS

This work is supported by National Science Foundation under Grant Nos. OAC-1852102, ECCS-2003307 and OAC-2118079.

REFERENCES

[1] A. Abramo, A. Cardin, L. Selmi, and E. Sangiorgi. 2000. Two-dimensional quantum mechanical simulation of charge distribution in silicon MOSFETs. *IEEE Tran. Electron Devices* 47, 10 (2000), 1858–1863.

[2] D. Ahlman, F. Söderlund, J. Jackson, A. Kurdila, and W. Shyy. 2002. Proper orthogonal decomposition for time-dependent lid-driven cavity flows. *Numer. Heat Transfer, Part B.* 42, 4 (2002), 285–306.

[3] A. A Andronov, E. P Dodin, D. I Zinchenko, and Yu. N. Nozdrin. 2009. Transport in GaAs/Al x Ga1- x As superlattices with narrow minibands: Effects of interminiband tunneling. *Semiconductors* 43, 2 (2009), 228–235.

[4] M. H. N. Assadi and D. A. H. Hanaor. 2013. Theoretical study on copper's energetics and magnetism in TiO2 polymorphs. *J. Applied Phys.* 113, 23 (2013), 233913.

[5] J. A. Atwell. 2000. *Proper orthogonal decomposition for reduced order control of partial differential equations.* Ph. D. Dissertation. Virginia Polytechnic Institute and State University.

[6] O. Benedikt, M. Sojka, P. Zaykov, D. Hornof, M. Kafka, P. Šůcha, and Z. Hanzálek. 2021. Thermal-Aware Scheduling for MPSoC in the Avionics Domain: Tooling and Initial Results. In *2021 IEEE 27th International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA).* 159–168.

[7] G. Berkooz, P. Holmes, and J. L. Lumley. 1993. The proper orthogonal decomposition in the analysis of turbulent flows. *Annu. Rev. Fluid Mech.* 25, 1 (1993), 539–575.

[8] D. Binion and X. Chen. 2017. A Krylov enhanced proper orthogonal decomposition method for frequency domain model reduction. *Eng. Computations* (2017).

[9] O. Brea, H. Daver, J. Rebek, and F. Himo. 2019. Mechanism (s) of thermal decomposition of N-Nitrosoamides: A density functional theory study. *Tetrahedron* 75, 8 (2019), 929–935.

[10] Y. F. Chang. 2015. Some solutions of extensive quantum equations in biology, formation of DNA and neurobiological entanglement. *NeuroQuantology* 13, 3 (2015), 304–310.

[11] A. Chatterjee. 2000. An introduction to the proper orthogonal decomposition. *Current Sci.* (2000), 808–817.

[12] A. J. Cohen, P. Mori-Sánchez, and W. Yang. 2008. Insights into current limitations of density functional theory. *Science* 321, 5890 (2008), 792–794.

[13] K. Dev, A. N. Nowroz, and S. Reda. 2013. Power mapping and modeling of multi-core processors. In *IEEE Int. Symp. Low-Power Electron. Design.* IEEE, 39–44.

[14] L. Feng, Y. Yue, N. Banagaaya, P. Meuris, W. Schoenmaker, and P. Benner. 2016. Parametric modeling and model order reduction for (electro-) thermal analysis of nanoelectronic structures. *J. Math. Industry* 6, 1 (2016), 1–16.

[15] FEniCSx. 2021. The FEniCSx computing platform. https://fenicsproject.org/

[16] A. Goyal, P. Gorai, H. Peng, S. Lany, and V. Stevanović. 2017. A computational framework for automation of point defect calculations. *Comp. Materials Science* 130 (2017), 1–9.

[17] S. Hammes-Schiffer. 2017. A conundrum for density functional theory. *Science* 355, 6320 (2017), 28–29.

[18] P. Harrison. 2009. *Quantum wells, wires and dots: theoretical and computational physics of semiconductor nanostructures.* John Wiley & Sons.

[19] I. Hendry and A. G. Sumpter. 2013. Thermal management of graphics processing units. US Patent 8,525,840.

[20] W. Jia and Ming-C. Cheng. 2022. A methodology for thermal simulation of interconnects enabled by model reduction with material property variation. *J. Comput. Sci.* 61 (2022), 101665.

[21] W. Jia, B. T. Helenbrook, and M.-C Cheng. 2016. Fast thermal simulation of FinFET circuits based on a multiblock reduced-order model. *IEEE Trans. CAD ICs and Syst.* 35, 7 (2016), 1114–1124.

[22] W. Kohn and L. J. Sham. 1965. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* 140, 4A (1965), A1133.

[23] P. LeGresley and J. Alonso. 2001. Investigation of non-linear projection for pod based reduced order models for aerodynamics. In *39th Aerosp. Ind. Assoc. Amer. Aerosp. Sci. Meeting Exhibit. (AIAA).* 926.

[24] Y. C. Liang, W. Z. Lin, H. P. Lee, S. P. Lim, K. H. Lee, and H. Sun. 2002. Proper orthogonal decomposition and its applications–part II: model reduction for MEMS dynamical analysis. *J. Sound Vibrat.* 256, 3 (2002), 515–532.

[25] C. Y. Liu, Y. J. Chen, and M. Hariyama. 2020. Thermal-aware memory system synthesis for MPSoCs with 3D-stacked hybrid memories. In *Proceedings of the 35th Annual ACM Symposium on Applied Computing.* 546–553.

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

[27] J. L. Lumley. 1967. The structure of inhomogeneous turbulent flows. *Atmospheric turbulence and radio wave propagation* (1967).

[28] S. Maity, A. Ghose, S. Dey, and S. Biswas. 2021. Thermal-aware adaptive platform management for heterogeneous embedded systems. *ACM Transactions on Embedded Computing Systems (TECS)* 20, 5s (2021), 1–28.

[29] K. Manna and J. Mathew. 2020. Thermal-aware design strategies for the 3d noc-based multi-core systems. In *Design and Test Strategies for 2D/3D Integration for NoC-based Multicore Architectures.* Springer, 111–123.

[30] J. X. Mu, Y.-X. Shi, M.-Y. Yang, Z. H. Sun, X. H. Liu, B.-J. Li, and N.-B. Sun. 2016. Design, synthesis, DFT study and antifungal activity of pyrazolecarboxamide derivatives. *Molecules* 21, 1 (2016), 68.

[31] A. Murli, V. Boccia, L. Carracciuolo, L. D'Amore, G. Laccetti, and M. Lapegna. 2007. Monitoring and Migration of a PETSc-based Parallel Application for Medical Imaging in a Grid computing PSE. In *Grid-Based Problem Solving Environments.* Springer, 421–432.

[32] R. Nath, R. Ayoub, and T. S. Rosing. 2013. Temperature aware thread block scheduling in GPGPUs. In *50th ACM/EDAC/IEEE Design Automation Conf. (DAC).* 1–6.

[33] M. F. Ng, M. B. Sullivan, S. W. Tong, and P. Wu. 2011. First-principles study of silicon nanowire approaching the bulk limit. *Nano letters* 11, 11 (2011), 4794–4799.

[34] M. Niknafs, I. Ukhov, P. Eles, and Z. Peng. 2019. Runtime resource management with workload prediction. In *Proceedings of the 56th Annual Design Automation Conference 2019.* 1–6.

[35] A. Nokhosteen, M. Soltani, and B. Barabadi. 2019. Reduced order modeling of transient heat transfer in microchip interconnects. *ASME J. Electron. Packag.* 141, 1 (2019), 011002.

[36] H. Pourmeidani, A. Sharma, K. Choo, M. Hassan, M. Choi, K. Kim, and B. Jang. 2018. Dynamic temperature aware scheduling for CPU-GPU 3D multicore processor with regression predictor. *J. Semicond. Tech. Sci.* 18, 1 (2018), 115–124.

[37] S. Rastegar, N. L. Hadipour, H. Soleymanabadi, et al. 2014. Theoretical investigation on the selective detection of SO2 molecule by AlN nanosheets. *J. Molecular Modeling.* 20, 9 (2014), 1–6.

[38] M. Shen and W. Cao. 2003. Electronic band-structure engineering of GaAs/AlxGa1- xAs quantum well superlattices with substructures. *Materials Science and Engineering: B* 103, 2 (2003), 122–127.

[39] G. Singla, G. Kaur, A. K. Unver, and U. Y. Ogras. 2015. Predictive dynamic thermal and power management for heterogeneous mobile platforms. In *2015 Design, Automation & Test in Europe Conference & Exhibition (DATE).* IEEE, 960–965.

[40] L. Sirovich. 1987. Turbulence and the dynamics of coherent structures. I. Coherent structures. *Quart. Appl. Math.* 45, 3 (1987), 561–571.

[41] L. Sirovich. 1987. Turbulence and the dynamics of coherent structures. II. Symmetries and transformations. *Quart. Appl. Math.* 45, 3 (1987), 573–582.

[42] L. Sirovich. 1987. Turbulence and the dynamics of coherent structures. III. Dynamics and scaling. *Quart. Appl. Math.* 45, 3 (1987), 583–590.

[43] Gilbert W Stewart. 2002. A Krylov–Schur algorithm for large eigenproblems. *SIAM J. Matrix Anal. A.* 23, 3 (2002), 601–614.

[44] S. A. Tatulian. 2018. From the Wave Equation to Biomolecular Structure and Dynamics. *Trends in biochemical sciences* 43, 10 (2018), 749–751.

[45] T. Vo-Dinh et al. 2013. Plasmonic nanoprobes: from chemical sensing to medical diagnostics and therapy. *Nanoscale* 5, 21 (2013), 10127–10140.

[46] H.-S. Zhang, L. Shi, X.-B. Yang, Y.-J. Zhao, K. Xu, and L. W. Wang. 2017. First-Principles Calculations of Quantum Efficiency for Point Defects in Semiconductors: The Example of Yellow Luminance by GaN: CN+ ON and GaN: CN. *Advanced Optical Materials* 5, 21 (2017), 1700404.