Reduced Basis Method for Nanodevices Simulation

George S.H. Pau*
Lawrence Berkeley National Laboratory,
1 Cyclotron Road MS 50A-1148,
Berkeley, CA94720.
(Dated: May 23, 2008)

Ballistic transport simulation in nanodevices, which involves self-consistently solving a coupled Schrödinger-Poisson system of equations, is usually computationally intensive. Here, we propose coupling the reduced basis method with the subband decomposition method to improve the overall efficiency of the simulation. By exploiting *a posteriori* error estimation procedure and greedy sampling algorithm, we are able to design an algorithm where the computational cost is reduced significantly. In addition, the computational cost only grows marginally with the number of grid points in the confined direction.

PACS numbers: 02.70.-c, 75.40Mg, 73.23.Ad

I. INTRODUCTION

As size of electronic devices shrinks to nanometer scale, ballistic charge transport becomes increasingly important in describing the transport phenomena in these devices [1]. However, ballistic charge transport simulation is usually computationally intensive — it involves self-consistently solving a coupled Schrödinger – Poisson system of equations [2–4]. Described in greater details in Section II, the iterative procedure involves repetitively solving a Schrödinger equation with open boundary conditions [2] at many different energy states within each iteration. The large number of states required to accurately determine the distribution of the electron density and the number of self-consistent iterations needed to achieve convergence lead to the large computational cost usually associated with ballistic charge transport simulation. A more efficient method to solve the Schrödinger equation can thus greatly improve the overall efficiency of ballistic charge transport simulation. Note that another popular approach to ballistic transport simulation involves solving the non-equilibrium Green's function equations (NGEF) – Poisson system of equations [5, 6]. In this paper, we will concentrate on the approach based on the Schrödinger equation although the methodology we describe can potentially be applied to the approach based on the NGEF formalism as well.

The finite difference method and the finite element method are most widely used to approximate the Schrödinger equation due to their flexibility [3, 4, 7–9]. However, a direct application of these methods, especially in higher spatial dimensions, can lead to a large algebraic system of equations, of which the solution is computational expensive. The subband decomposition method [10, 11] or more commonly known as the coupled-mode approach [5, 12] attempts to reduce the computational cost by decomposing the Schrödinger equation into two smaller subproblems, resulting in a bounded Schrödinger equation in the confined directions and an open Schrödinger equation in the transport direction. In particular, by first solving the bounded Schrödinger equation at different locations along the transport direction, we are able to obtain a smaller algebraic system of equations for the open Schrödinger equation, which can then be solved more efficiently; the procedure is then effective in the limit where we need to solve the Schrödinger equation at large number of different energy levels. The efficiency can be further improved by a WKB approximation of the open Schrödinger equation [11]. Nevertheless, solving the bounded Schrödinger equation, which involves solving an eigenvalue problem at different locations along the transport direction, can still be potentially ex-

*Electronic address: gpau@lbl.gov

pensive, especially when strong confinement of the electron demands a finely discretized simulation domain. This paper proposes an efficient method based on the reduced-basis approach to reduce the computational cost of solving the bounded Schrödinger equation.

The reduced basis method is a model-order reduction technique which exploits dimension reduction afforded by the *smooth* and *low-dimensional* parametrically induced solution manifold. Instead of using general basis sets such as finite element, an approximation to a solution of an underlying parameterized partial differential equation is obtained by a projection onto a finite and low dimensional vector space spanned by a basis sets consisting of solutions at a number of judiciously selected parameter points. The reduced basis method was first introduced in the late 1970s in the context of nonlinear structural analysis [13, 14] and subsequently abstracted, analyzed, and extended to a much larger class of parameterized partial differential equations [15–19]. In the more recent past the reduced basis approach and in particular associated *a posteriori* error estimation procedures have been successfully developed for many different types of PDEs that are affine in the parameters [20–26], general nonaffine PDEs [27, 28], and linear eigenvalue problem [21, 29]. We will elaborate further the methodology in Section III. In particular, we extend the methodology described in [29] to eigenvalue problem that is nonaffine in the parameter, and describe how reduced basis methodology can be incorporated into the overall solution procedure for the Schrödinger – Poisson system of equations.

This paper is organized as follows. We first describe the problem that we would like to solve. To simplify the presentation of the methodology, we will use the double-gate MOSFET as a model problem. We then provide the weak formulation of the equations involved and briefly describe the subband decomposition method. This serves as a platform for us to describe the reduced basis method, and how it fits into the overall solution procedure. We conclude with some numerical results and comparison to the subband decomposition method.

II. PROBLEM STATEMENT

With the effective mass approximation [30], the electron is described by a wavefunction $\psi(E) \in H^1(\Omega) \subset \mathbb{C}^2$ which for a given E, satisfies the following schrödinger equation:

$$-\nabla \cdot \left(\frac{1}{2m^*} \nabla \psi(E) \right) + V_{\text{eff}}(\psi)\psi(E) - E\psi(E) = 0, \tag{1}$$

with appropriate open boundary conditions [2]. The potential $V_{\text{eff}} \in L^2(\Omega)$ is given by

$$V_{\text{eff}}(\psi) = -\phi(\psi) + V_{xc}(\psi) + V_b, \tag{2}$$

where $\phi \in H^1(\Omega)$, $V_{xc} \in H^1(\Omega)$ and $V_b \in L^2(\Omega)$. We ignore the exchange-correlation term V_{xc} for simplicity but the methodology described will easily accommodate the V_{xc} term; and V_b describes the potential gap between the insulator and the semiconductor. The potential ϕ in turn satisfies a Poisson equation given by

$$-\nabla \cdot [\epsilon \nabla \phi] = -n(\psi) + N_D, \tag{3}$$

with appropriate boundary conditions. Here, ϵ is the dielectric function of the materials, $n(\psi)$ is the density of free electrons, and N_D is the concentration of donor impurities; we ignore contribution of hole and acceptor impurities for simplicity. Equation (1) and (2) are thus coupled through the term $n(\psi)$, which can be defined as

$$n(\psi) = C \int_0^\infty f_{-1/2} \left(\frac{E_f - E}{T} \right) |\psi(E)|^2 dE, \tag{4}$$

where C is a model-dependent coefficient and $f_{-1/2}(\cdot)$ is the Fermi-Dirac integral of order $-1/2$. Here, E_f is the fermi level of the electrons entering the system.

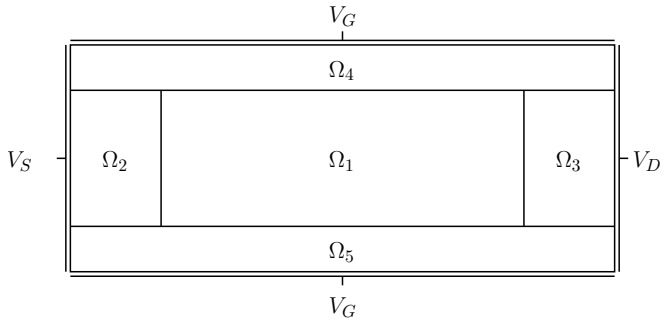

FIG. 1: A model problem based on the double-gate MOSFET.

To solve the above coupled system of equations, an iterative scheme is needed. We utilize a relaxed fixed-point method. Starting from an initial guess n^0, we construct the sequence n^k where

$$n^k = \alpha n^{k-1} + (1-\alpha) n^f,$$

α is an adjustable parameter to improve convergence, and n^f is determined from (4) with ψ^k computed from (1) with $V_{\text{eff}} = \phi^k + V_b$. We note that (4) must be evaluated numerically, and thus (1) must be evaluated many times. We then solve (3) for ϕ^{k+1} with the new value of n^k. The procedure is repeated until $\|\phi^k - \phi^{k-1}\| \leq \varepsilon_{\text{tol}}$, where ε_{tol} is our desired tolerance. Clearly, even if numerical approximation of (1) is not particularly expensive, the cumulative effects due to the iterative nature of the algorithm and numerical approximation of (4) can be computationally challenging.

For the purpose of this paper, we will consider a 2-dimensional nanodevice (a double-gate MOSFET) shown in Figure II. Given a source potential, V_S, a drain potential V_D and a gate potential V_G, we would like to determine the current flow I in the x_1-direction. The simulation domain $\Omega \equiv [0,a] \times [0,b] \subset \mathbb{R}^2$ can be further divided into 5 subdomains denoted by Ω_i, $i = 1, \ldots, 5$; (x_1, x_2) denotes a point in Ω. The material properties we will be using is that of Si in Ω_1, Ω_2 and Ω_3, and SiO_2 in Ω_4 and Ω_5. In addition, Ω_2 and Ω_3 are doped to provide free carriers for the charge transport. We assume the crystal structure of the device is oriented such that x_1 is in the $\langle 100 \rangle$ direction and x_2 is in the $\langle 001 \rangle$ direction. The axes are then aligned with the principal axes of the six equivalent ellipsoids of the conduction band. Based on the effective mass approximation, we then have three configurations for $m^* \equiv (m_1^*, m_2^*, m_3^*)$ and

$$\nabla \cdot \left(\frac{1}{2m^*} \nabla \psi \right) = \frac{\partial}{\partial x_1} \left(\frac{1}{2m_1^*} \frac{\partial}{\partial x_1} \psi \right) + \frac{\partial}{\partial x_2} \left(\frac{1}{2m_2^*} \frac{\partial}{\partial x_2} \psi \right) + \frac{\partial}{\partial x_3} \left(\frac{1}{2m_3^*} \frac{\partial}{\partial x_3} \psi \right).$$

The three configurations of m^* are given by (m_l, m_l, m_t), (m_t, m_l, m_l), (m_l, m_t, m_t); m_t and m_l are the transverse and longitudinal masses of the material. We assume m_t and m_l for Si and SiO_2 are the same. Finally, we assume we have a two dimensional electron gas with a parabolic dispersion relation in the x_3 direction.

A. Abstract Formulation

We now derive the weak formulation for (1) and (3) for the model problem described in Section II. For (1), the weak formulation is: given $E \in \mathbb{R}$, find $\psi \in Y \equiv H^1(\Omega)$ such that

$$\frac{1}{2m_1^*}\int_\Omega \frac{\partial\psi}{\partial x_1}\frac{\partial v^*}{\partial x_1}+\frac{1}{2m_2^*}\int_\Omega \frac{\partial\psi}{\partial x_2}\frac{\partial v^*}{\partial x_2} + \int_\Omega \psi V_{\text{eff}} v^* - E\int_\Omega \psi v^* = \frac{1}{2m_1^*}\int_{\Gamma_S\cup\Gamma_D}\frac{\partial\psi}{\partial x_1}v^*, \quad \forall v\in Y. \tag{5}$$

where Γ_S and Γ_D are respectively the boundaries in contact with source and drain electrodes. Based on the quantum transmitting boundary method [2], we expand the R.H.S of (5): for $g = S, D$,

$$\begin{aligned}\int_{\Gamma_g}\frac{\partial\psi}{\partial x_1}v^* &= \sum_{m=1}^{N^g} \mathrm{i}\ 2a_m^g k_m^g \int_{\Gamma_g}\chi_m^g v^* + \int_{\Gamma_g}\chi_m^g v^*\int_{\Gamma_g}\chi_m^g\psi \\ &\quad - \sum_{m=N^g+1}^{\infty} k_m^g\int_{\Gamma_g}\chi_m^g v^*\int_{\Gamma_g}\chi_m^g\psi,\end{aligned} \tag{6}$$

where (ξ_m^g, E_m^g), $1\le m\le\infty$ are the eigenstates along Γ_g; $k_m^g = \sqrt{2m^*(E-E_m^g)}$; N^g is the largest m for which $E > E_m^g$, b_m^g, $1\le m\le N^g$ are the coefficients of outgoing traveling-wave states, and b_m^g, $m > N^g$ are coefficients of the evanescent states. For a particular problem, a_m^g is a parameter that we can vary while b_m^g and N^g is determined as part of the solution.

To facilitate the variational formulation, We now define the following functional forms : $\forall\ w\in Y$, $v\in Y$, $V\in L^2$, $\chi^g\in H_0^1(\mathbb{R})$,

$$a_0(w,v;\alpha) = \int_\Omega \alpha\nabla w\ \nabla v^*, \tag{7}$$

$$a_1(w,v;m^*) = \frac{1}{2m_1^*}\int_\Omega\frac{\partial w}{\partial x_1}\frac{\partial v^*}{\partial x_1}+\frac{1}{2m_2^*}\int_\Omega\frac{\partial w}{\partial x_2}\frac{\partial v^*}{\partial x_2}, \tag{8}$$

$$a_2(w,v;V) = \int_\Omega w\ V\ v^*, \tag{9}$$

$$a_3(w,v) = \int_\Omega w\ v^*, \tag{10}$$

$$c(w,v;\chi^g) = \frac{1}{2m_1^*}\int_{\Gamma_g}\chi^g w\int_{\Gamma_g}\chi^g v^*, \tag{11}$$

$$b(v;\chi^g) = \frac{1}{2m_1^*}\int_{\Gamma_g}\chi^g v^*. \tag{12}$$

The abstract formulation is then: given $E\in\mathbb{R}$, find $\psi\in Y$ that satisfies

$$\begin{aligned}&a_1(\psi,v;m^*)+a_2(\psi,v;V_{\text{eff}})-Ea_3(\psi,v) \\ -\sum_{g=S,D}\sum_{m=1}^{N^g}\mathrm{ik}_m^g c(\psi,v;\chi_m^g) &+ \sum_{g=S,D}\sum_{m=N^k+1}^{\infty}\mathrm{k}_m^g c(\psi,v;\chi_m^g) \\ &= -\sum_{g=S,D}\sum_{m=1}^{N^g}\mathrm{i}\ 2a_m^g\mathrm{k}_m^g b(v;\chi_m^g), \quad \forall v\in Y.\end{aligned} \tag{13}$$

For (3), the weak formulation is: given $n(\psi)$, the solution $\phi\in H^1(\Omega)$ is given by

$$\int_{\Omega/\Gamma_0}\epsilon\nabla\phi\nabla v^* + \int_{\Gamma_0}\epsilon\nabla V_G\nabla v^* = \int_\Omega(-n(\psi)+N_D)v^*, \quad \forall v\in Y,$$

where Γ_0 is the boundary in contact with gate electrode. We have imposed the following boundary conditions:

$$\phi|_{\Gamma_0} = 0, \quad \text{and} \quad \frac{d\phi}{dx_1}\Big|_{\Gamma_S \cup \Gamma_D} = 0.$$

Let $f(v;V) = \int_\Omega V v^*$, and $h(v;V) = \int_{\Gamma_0} \epsilon \nabla V \nabla v^*$. Then, the abstract formulation is: given $n(\psi)$, the solution $\phi \in Y$ is given by

$$a_0(\phi, v; \epsilon) \;=\; f(v; -n(\psi) + N_D) - h(v; V_G), \quad \forall v \in Y. \tag{14}$$

For the current problem where we have assumed a 2-dimensional electron gas, the charge density n is given by [6]

$$n(\psi) = \frac{2}{\pi}\sqrt{\frac{2m_3^* T}{\pi}} \int_0^\infty f_{-1/2}\left(\frac{E_f - E}{T}\right) |\psi(E)|^2 dE, \tag{15}$$

and this is sum over the three different configuration of m^*. We assume $E_f = 0$ at zero bias. Finally, the current intensity I is given by

$$I = \int_0^b j_1(x_1, x_2; \psi) dx_2 \tag{16}$$

where j_1, the current density in the x_1 direction, is defined as

$$j_1(\psi) = \frac{1}{m_1^*}\frac{2}{\pi}\sqrt{\frac{2m_3^* T}{\pi}} \int_0^\infty \mathrm{Im}\left(\bar{\psi}(E)\frac{\partial \psi(E)}{\partial x_1}\right) f_{-1/2}\left(\frac{E_f - E}{T}\right) dE. \tag{17}$$

Numerical approximation of (13) – (14) based on, say, finite element method, can however be computationally very expensive since (13) must be solved many times in a single iteration in order to numerically determine the density n. In particular, suppose we substitute the unbounded upper limit in (15) by $E_{\max}$ and subdivide the interval $[0, E_{\max}]$ into n_E intervals. We then use Gauss quadrature formulation within each interval to arrive at the following approximation of n:

$$n(\psi) \approx \frac{2}{\pi}\sqrt{\frac{2m_3^* T}{\pi}} \sum_{g=S,D} \sum_{m=1}^{N^g} \sum_{i=1}^{n_E} \sum_{q=1}^{Q} f_{-1/2}\left(\frac{E_f - E_q^i}{T}\right) |\psi(E_q^i; a_m^g)|^2 w_q \tag{18}$$

where N^g is the number of modes considered at Γ_g; $a_{m'}^{g'} = 1$ if $m' = m$ and $g' = g$, and 0 otherwise; E_q^i are the quadrature points in interval i; w_q is the quadrature weight; and Q is the number of quadrature points used per interval. Then, in each iteration, the maximum number of times we must solve (13) is $(N^S + N^D) n_E Q$. This can be somewhat smaller by excluding E for which $f_{-1/2}((E_f - E)/T)$ is negligibly small.

B. Subband Decomposition Approach

The subband decomposition method is first described in [5, 10]. Assuming that the wavefunction is bounded in the x_2-direction, we can write Y as $X^1 \times X^2$ where $X^1 = H^1(\Omega^1 \equiv [0, a])$ and $X^2 = H_0^1(\Omega^2 \equiv [0, b])$. Then, we can express $\psi \in Y$ as

$$\psi(x; E) = \sum_{i=1}^{\infty} \varphi(x_1; E)\xi_i(x_2; x_1), \quad \varphi(x_1; E) \in X^1, \quad \xi_i(x_2; x_1) \in X^2. \tag{19}$$

Here, $\xi_i(\cdot;\mu\equiv x_1)\in X^2$, $i=1,\ldots,\infty$ are solutions to the following eigenvalue problem:

$$\begin{aligned}\tilde{a}_1(\xi_i(\mu),v;m_2^*)+\tilde{a}_2(\xi_i(\mu),v;V_{\text{eff}}(\mu)) &= \lambda_i(\mu)\tilde{a}_3(\xi_i(\mu),v),\\ &\qquad 1\le i\le\infty,\quad \forall v\in X^2, \qquad (20)\\ \tilde{a}_3(\xi_i(\mu),\xi_j(\mu)) &= \delta_{ij},\quad 1\le i,j\le\infty; \qquad (21)\end{aligned}$$

where $V_{\text{eff}}(\mu)=V_{\text{eff}}(x_2;\mu\equiv x_1)$; and

$$\tilde{a}_1(w,v;\alpha)=\int_{\Omega^2}\frac{1}{2\alpha}\nabla w\nabla v,\quad \tilde{a}_2(w,v;t)=\int_{\Omega^2}x\,t\,v,\quad \tilde{a}_3(w,v)=\int_{\Omega^2}w\,v, \qquad (22)$$

for $w\in X^2$, $v\in X^2$ and $t\in L^2(\Omega^2)$.

Substituting (19) into (13), we obtain a one dimensional problem for $\varphi_i(E)$:

$$\begin{aligned}&\sum_{i=1}^{\infty}\frac{1}{2m_1^*}\left\{\int_{\Omega^1}\frac{d\varphi_i(E)}{dx_1}\frac{dt}{dx_1}\tilde{a}_3(\xi_i(x_1),\xi_j(x_1))+\int_{\Omega^1}\frac{d\varphi_i(E)}{dx_1}t(x_1)\tilde{a}_3(\xi_i(x_1),\frac{\partial\xi_j}{\partial x_1}(x_1))\right.\\ &\quad+\int_{\Omega^1}\varphi_i(x_1;E)\frac{dt(x_1)}{dx_1}\tilde{a}_3(\frac{\partial\xi_i}{\partial x_1}(x_1),\xi_j(x_1))\\ &\quad\left.+\int_{\Omega^1}\varphi_i(x_1;E)t(x_1)\tilde{a}_3(\frac{\partial\xi_i}{\partial x_1}(x_1),\frac{\partial\xi_j}{\partial x_1}(x_1))\right\}\\ &\quad+\int_{\Omega^1}(\lambda_i(x_1)-E)\varphi_i(x_1;E)t(x_1)\delta_{ij}-\sum_{g=S,D}\sum_{m=1}^{N^g}\mathrm{i}\mathrm{k}_m^g\frac{\varphi_i(x_g)t(x_g)}{2m_1^*}\delta_{mi}\delta_{mj}\\ &\quad+\sum_{g=S,D}\sum_{m=1}^{N^g}\mathrm{k}_m^g\frac{\varphi_i(x_g)t(x_g)}{2m_1^*}\delta_{mi}\delta_{mj}\\ &\qquad=-\sum_{g=S,D}\sum_{m=1}^{N^g}\mathrm{i}2a_m^g\mathrm{k}_m^g\frac{t(x_g)}{2m_1^*}\delta_{mj},\quad\forall t\in X^1,\quad 1\le j\le\infty. \qquad (23)\end{aligned}$$

This is simply the weak form for the following one-dimensional Schrödinger equation [10]:

$$-\frac{d}{dx_1}\left(\frac{1}{2m_1^*}\frac{d}{dx_1}\varphi_i\right)-\sum_{j=1}^{\infty}\frac{a_{ij}(x_1)}{m_1^*}\frac{d}{dx_1}\varphi_j-\sum_{j=1}^{\infty}\left(\frac{b_{ij}(x_1)}{2m_1^*}-\lambda_i\delta_{ij}+E\right)\varphi_j=0, \qquad (24)$$

for $i=1,\ldots,\infty$ with the appropriate open boundary condition; $a_{ij}(x_1)=\int_{\Omega_2}\xi_i(x_1)\,\{\partial\xi_j(x_1)/\partial x_1\}$ and $a_{ij}(x_1)=\int_{\Omega_2}\xi_i(x_1)\{\partial^2\xi_j(x_1)/\partial x_1^2\}$. It is further found that only finite number of ξ_i is needed, which we denote as n_e. If these n_e $\xi_i(x_1)$ are known, this one-dimensional problem can be solved very efficiently.

To achieve self-consistency, each fixed point iteration now consists of three parts: (i) determination of the subbands $\xi_i(x_2;x_1)$, $1\le i\le n_e$ for finite points in Ω^1, (ii) determination of $n(\psi)$ by solving (23) for $(N^S+N^D)n_EQ$ different combination of E and a_m^g, and (iii) determination of $\phi(\psi)$ by solving (14) given $n(\psi)$. In [10], finite element method is used to approximate the solutions at all stages of the algorithm. It is hope that the computational overhead incurred in part (i) will significantly reduce the computational cost of solving the open Schrödinger equation needed to determine the electron density. However, part (i) can be computationally expensive, especially if very fine mesh is needed to resolve the strong confinement of the electrons in the x_2-direction or when (20) must be solve at large number of points if finer mesh is needed in the x_1-direction. Our goal is to speed-up the determination of ξ_i for any given x_1 through the reduced basis method.

Before we proceed, we describe how we can determine $\partial\xi_i/\partial\mu$. Let $\partial\xi_i/\partial\mu\in X^2$. Then, by taking

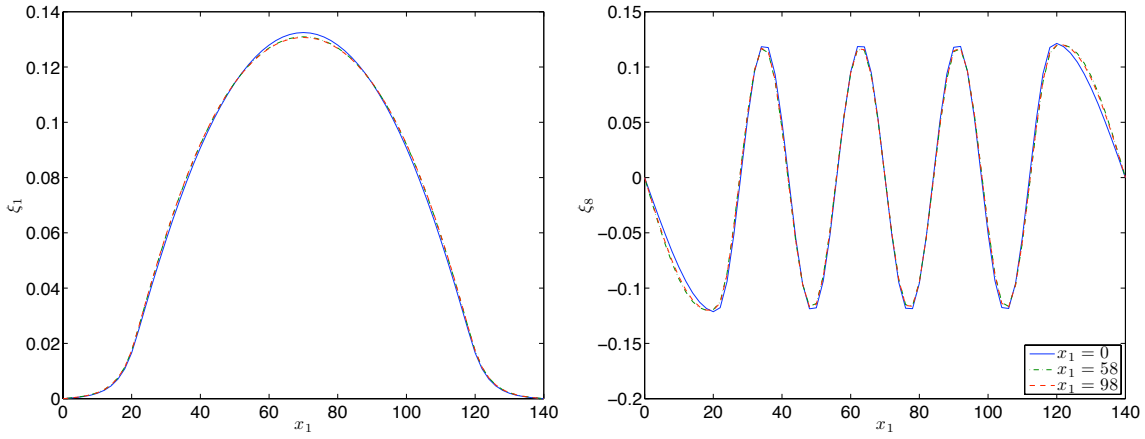

FIG. 2: ξ_1 and ξ_8 for three different values of x_1 based on FE approximation for $V_D = 0.015$.

the derivative of (20) with respect to μ, we obtain

$$\begin{aligned}\tilde{a}_1(\frac{\partial \xi_i}{\partial \mu}(\cdot;\mu), v; m_2^*) + \tilde{a}_2(\frac{\partial \xi_i}{\partial \mu}(\cdot;\mu), v; V_{\text{eff}}(\cdot;\mu)) - \lambda_i(\mu)\tilde{a}_3(\frac{\partial \xi_i}{\partial \mu}(\cdot;\mu), v) \\ = -\tilde{a}_2(\xi_i(\cdot;\mu), v; \frac{\partial V_{\text{eff}}(\cdot;\mu)}{\partial \mu}) + \frac{d\lambda_i(\mu)}{d\mu}\tilde{a}_3(\xi_i(\cdot;\mu), v), \\ 1 \le i \le \infty, \quad \forall v \in X^2. \end{aligned} \tag{25}$$

In addition, by letting $v = \xi_i$, and invoking (20), we obtain

$$\frac{d\lambda_i(\mu)}{d\mu} = \tilde{a}_2(\xi_i(\cdot;\mu), \xi_i(\cdot;\mu); \frac{\partial V_{\text{eff}}(\cdot;\mu)}{\partial \mu}). \tag{26}$$

Finally, by substituting (26) into (25), we can solve for $\partial \xi_i / \partial \mu$. At present $\partial V_{\text{eff}}/\partial \mu$ is computed using a difference formula. In Appendix A, we describe a formulation that is more consistent with the finite element approximation space of ϕ; it however leads to a higher computational cost. We also note that since V_b does not depend on x_1, $\partial V_{\text{eff}}/\partial \mu = -\partial \phi / \partial \mu$.

III. REDUCED BASIS METHOD

We now present the reduced basis formulation for (20). Figure 2 shows the variation of ξ_1 and ξ_8 for several different values of x_1. The variation is small but nontrivial, and we can discern a smooth variation of ξ_i with respect to x_1. This suggests that reduced basis method can very efficiently approximate ξ_i. This section describes how ξ_i and $\partial \xi_i / \partial x_1$ can be approximated by the reduced basis method. Let $\mu \equiv x_1$, and $\mathcal{D} \equiv \Omega^1$. In additional, for notational convenience, we have $\xi_i(\mu) = \xi_i(x_2;\mu)$, $\phi(\mu) = \phi(x_2;\mu)$, $d\xi_i(\mu) = \partial \xi_i(x_2;\mu)/\partial \mu$ and $d\phi(\mu) = \partial \phi(x_2;\mu)/\partial \mu$.

A. Approximation Spaces

We first introduce nested sample sets $S_N = (\mu_1, \ldots, \mu_{N_s})$, $1 \le N_s \le N_{s,\max}$ and define the associated nested reduced-basis spaces as

$$\begin{aligned} W_N &= \text{span } \{\xi_i(\mu_j), 1 \le i \le n_e, 1 \le j \le N_s\}, \quad 1 \le N_s \le N_{s,\max}, \\ &= \text{span } \{\zeta_n, 1 \le n \le N \equiv N_s n_e\}, \quad 1 \le N_s \le N_{s,\max}; \end{aligned} \tag{27}$$

where $\xi_1(\mu_j), \ldots, \xi_{n_e}(\mu_j)$ are the solutions of (20) at $\mu = \mu_j$; and ζ_n are basis functions obtained after $\xi_i(\mu_j)$, $1 \le i \le n_e$, $1 \le j \le N_s$ are orthonormalized.

We also construct collateral approximation spaces for $\phi(\mu)$ and $d\phi(\mu)$ based on the empirical interpolation procedure [27, 28, 31]. For $p = \phi(\mu)$ and $d\phi(\mu)$, we construct nested sample sets $S_M^p \equiv \{\mu_1^p, \ldots, \mu_M^p\}$, $1 \le M \le M_{\max}^p$, nested approximation spaces $W_M^p \equiv \text{span } \{q_1^p, \ldots, q_M^p\}$, $1 \le M \le M_{\max}^p$, and nested interpolation points $T_M^p \equiv \{t_1^p, \ldots, t_M^p\}$, $1 \le M \le M_{\max}^p$.

In (27), we have assumed $\xi_i(\mu_j)$ are known exactly. In practice however, $\xi_i(\mu_j)$ must be determined through some form of "truth" approximation — here, we use the finite element method with $\mathbb{P}_1$ elements. We build our reduced basis approximation on, and measure the error in the reduced basis approximation relative to this "truth" approximation. Note that since reduced basis approximation is build upon this "truth" approximation, it cannot perform better than this "truth" approximation. Thus, the number of elements used to obtain our "truth" approximation, $\mathcal{N}$, must usually be large. Similarly, the W_M^ϕ and $W_M^{d\phi}$ are constructed from a "truth" approximation of ϕ and $d\phi$, here based on finite element method utilizing $\mathbb{Q}_2$ elements.

B. The Approximation

Our reduced basis approximation to (20) is then given by: find $(\xi_{i,N,M}(\mu), \lambda_{i,N,M}(\mu)) \in \mathcal{Y}_N \equiv (W_N \times \mathbb{R})$, $1 \le i \le n_e$ such that

$$\begin{aligned} \tilde{a}_1(\xi_{i,N,M}(\mu), v; m_2^*) + \tilde{a}_2(\xi_{i,N,M}(\mu), v; V_{\text{eff},M}(\mu)) &= \lambda_{i,N,M}(\mu)\tilde{a}_3(\xi_{i,N,M}(\mu), v), \\ &\quad 1 \le i \le n_e, \ \forall v \in W_N, \end{aligned} \tag{28}$$

$$\tilde{a}_3(\xi_{i,N,M}(\mu), \xi_{j,N,M}(\mu)) = \delta_{ij}, \quad 1 \le i, j \le n_e, \tag{29}$$

where $V_{\text{eff},M} = V_b + \phi_M$.

Similarly, our reduced basis approximation to (25) is given by: find $d\xi_{i,N,M}(\mu) \in W_N$, $1 \le i \le n_e$ such that

$$\begin{aligned} \tilde{a}_1(d\xi_{i,N,M}(\mu), v; m_2^*) + \tilde{a}_2(d\xi_{i,N,M}(\mu), v; V_{\text{eff},M}(\cdot;\mu)) - \lambda_{i,N,M}(\mu)\tilde{a}_3(d\xi_{i,N,M}(\mu), v) \\ = \tilde{a}_2(\xi_{i,N,M}(\mu), v; d\phi_{N,M}(\mu)) + \frac{d\lambda_{i,N,M}(\mu)}{d\mu}\tilde{a}_3(\xi_{i,N,M}(\mu), v), \\ 1 \le i \le n_e, \quad \forall v \in W_N, \end{aligned} \tag{30}$$

where

$$\frac{d\lambda_{i,N,M}(\mu)}{d\mu} = \tilde{a}_2(\xi_{i,N,M}(\mu), \xi_{i,N,M}(\mu); \phi_M(\mu)). \tag{31}$$

It is not immediately clear that $d\xi_{i,N,M}(\mu)$ can be sufficiently approximated in W_N. In Section IV B, we will examine if it is necessary to replace W_N by an enlarged space W_N^d given by

$$\begin{aligned} W_N^d &= \text{span } \{\xi_i(\mu_j), \ldots, \xi_{n_e}(\mu_j), d\xi_i(\mu_j), \ldots, d\xi_{n_e}(\mu_j), 1 \le j \le N_s\}, \\ &= \text{span } \{\zeta_n, 1 \le n \le N \equiv 2N_s n_e\}. \end{aligned} \tag{32}$$

C. Offline-online Decomposition

We first expand our reduced basis approximation as

$$\xi_{n,N,M}(\mu) = \sum_{j=1}^{N} \xi_{n,N,M\,j}(\mu)\zeta_j, \quad 1 \le n \le n_e, \tag{33}$$

where $\zeta_j \in W_N$, and $\xi_{n,N,M\,j}(\mu) \in \mathbb{R}$.

We then expand our empirical interpolation approximation for $\phi(\cdot;\mu)$ as

$$\phi_M(\cdot;\mu) = \sum_{m=1}^{M} \beta_{M,m}(\mu) q_m^\phi(\cdot) \ , \tag{34}$$

where $\beta_M(\mu) \in \mathbb{R}^M$ is given by

$$\sum_{k=1}^{M} B_{m,k}^{M,\phi} \beta_{M\,k}(\mu) \ = \ \phi(t_m^\phi;\mu), \quad 1 \le m \le M; \tag{35}$$

and $B^{M,\phi} \in \mathbb{R}^M \times \mathbb{R}^M$ is given by $B_{m,k}^{M,\phi} = q_m^\phi(t_k^\phi)$, $1 \le m,k \le M$. We note that $\{q_m^\phi, 1 \le m \le M\}$ is pre-constructed offline based on the empirical interpolation method, and we compute $\beta_M(\mu)$ given any $\mu \in \mathcal{D}$. Inserting the above representations (33) and (34) into (28), we obtain the following discrete equations

$$\sum_{j=1}^{N} \left\{ A_{i,j}^N + \left(\sum_{m=1}^{M} C^{N,\phi,m} \beta_{M\,m}(\mu) \right) \right\} \xi_{n,N,M\,j}(\mu) \ =$$

$$\lambda_{n,N,M}(\mu) \sum_{j=1}^{N} M_{i,j}^N \xi_{n,N,M\,j}(\mu), \quad 1 \le i \le N, \quad 1 \le n \le n_e \ ; \tag{36}$$

$$\sum_{i=1}^{N}\sum_{j=1}^{N} \xi_{n,N,M\,i}(\mu) M_{i,j}^N \xi_{m,N,M\,j}(\mu) \ = \ \delta_{nm}, \quad 1 \le n,m \le n_e \ ; \tag{37}$$

where $A^N \in \mathbb{R}^{N\times N}$, $M^N \in \mathbb{R}^{N\times N}$, $C^{N,\phi,m} \in \mathbb{R}^{N\times N}$, $1 \le m \le M$ are given by $A_{i,j}^N = \tilde{a}_1(\zeta_j,\zeta_i;m_2^*) + \tilde{a}_2(\zeta_j,\zeta_i;V_b)$, $1 \le i,j \le N$, $M_{i,j}^N = \tilde{a}_3(\zeta_j,\zeta_i)$, $1 \le i,j \le N$, and $C_{i,j}^{N,\phi,m} = \tilde{a}_2(\zeta_j,\zeta_i;q_m^\phi)$, $1 \le i,j \le N$, $1, \le m \le M$ respectively.

Similarly, for (30), we expand

$$d\xi_{n,N,M}(\mu) = \sum_{j=1}^{N} d\xi_{n,N,M\,j}(\mu)\zeta_j, \tag{38}$$

where $\zeta_j \in W_N$, and $d\xi_{n,N,M\,j}(\mu) \in \mathbb{R}$. We also expand our empirical interpolation approximation for $d\phi(\mu)$ as

$$d\phi_M(\cdot;\mu) = \sum_{m=1}^{M^{d\phi}} \gamma_{M,m}(\mu) q_m^{d\phi}(\cdot) \ , \tag{39}$$

where $\gamma_M(\mu) \in \mathbb{R}^M$ is given by

$$\sum_{k=1}^{M} B^{M,d\phi}_{m,k} \gamma_{M\,k}(\mu) \;=\; d\phi(t^{d\phi}_m;\mu), \quad 1 \le m \le M; \tag{40}$$

and $B^{M,d\phi} \in \mathbb{R}^{M^{d\phi}} \times \mathbb{R}^{M^{d\phi}}$ is gven by $B^{M,d\phi}_{m,k} = q^{d\phi}_m(t^{d\phi}_k)$, $1 \le m,k \le M$. Inserting the above representations (33), (38) and (39) into (30), we obtain the following discrete equations

$$\begin{aligned}
&\sum_{j=1}^{N} \left\{ A^N_{i,j} + \left(\sum_{m=1}^{M^\phi} C^{N,\phi,m} \beta_{M\,m}(\mu) \right) - \lambda_{n,N,M}(\mu) M^N_{i,j} \right\} d\xi_{n,N,M\,j}(\mu) \;= \\
&\sum_{j=1}^{N} \left\{ \frac{d\lambda_{i,N,M}(\mu)}{d\mu} M^N_{i,j} - \left(\sum_{m=1}^{M^{d\phi}} C^{N,d\phi,m} \gamma_{M\,m}(\mu) \right) \right\} \xi_{n,N,M\,j}(\mu), \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad 1 \le i \le N\ ;
\end{aligned} \tag{41}$$

where $C^{N,d\phi,m} \in \mathbb{R}^{N\times N}$, $1 \le m \le M^{d\phi}$ is given by $C^{N,d\phi,m}_{i,j} = \tilde{a}_2(\zeta_j,\zeta_i;q^{d\phi}_m)$, $1 \le i,j \le N$, $1, \le m \le M^{d\phi}$, respectively.

Finally, the linear functional $\tilde{a}_3$ is simply approximated by

$$\begin{aligned}
\tilde{a}_3(w_n(\mu), v_m(\mu)) \;&\approx\; \tilde{a}_3(w_{n,N,M}(\mu), v_{m,N,M}(\mu)) \\
&=\; \sum_{i=1}^{N}\sum_{j=1}^{N} M^N_{i,j} w_{n,N,M\,i}(\mu) v_{m,N,M\,j}(\mu)
\end{aligned} \tag{42}$$

when we want to compute the $\tilde{a}_3$ terms in (23).

The computational decomposition is then clear. At the beginning of each inner iteration, we generate nested reduced-basis spaces W_N, $1 \le N \le N_{\max}$, nested approximation spaces W^ϕ_M, $1 \le M \le M^\phi_{\max}$ and $W^{d\phi}_M$, $1 \le M \le M^{d\phi}_{\max}$, and the associated nested sets of interpolation points T^ϕ_M and $T^{d\phi}_M$. For determining $\xi_{i,N,M}$, $1 \le i \le n_e$, we form and store $A^N, M^N, B^{M,\phi}$, $C^{N,\phi,m}$, $1 \le m \le M^\phi_{\max}$ and $C^{N,d\phi,m}$, $1 \le m \le M^{d\phi}_{\max}$. This is equivalent to the offline stage in a more typical reduced-basis formulation. The computational cost is (to leading order) $O(N\mathcal{N}^\bullet + n_e N\mathcal{N}^\dagger + M^2N^2\mathcal{N})$, where $\bullet$ and $\dagger$ depend on the complexity of the eigenvalue solver and linear solver used, $M = \max(M^\phi, M^{d\phi})$, and $\mathcal{N}$ is the dimension of our "truth" approximation.

In the online stage — during construction of discrete matrices for (23) — we solve (36) – (37) for $\xi_{n,N,M\,j}(\mu)$, $1 \le j \le N$, $1 \le n \le n_e$, and $d\xi_{n,N,M\,j}(\mu), 1 \le j \le N^{d\xi_n}$, $1 \le n \le n_e$, and evaluate (42). The computational costs for each μ is then $O((n_eN)^3 + n_eN^3 + MN^2)$, which is then independent of $\mathcal{N}$.

D. *A Posteriori* Error Estimation

The derivation of the *a posteriori* error estimator follows [29]. We note that the eigenvalues λ_i are of multiplicity *one* but $\tilde{a}(v,v;V_{\text{eff}}(\mu)) = \tilde{a}_1(v,v;m^*_2) + \tilde{a}_2(v,v;V_{\text{eff}}(\mu))$ is not strictly positive for all $\mu \in \mathcal{D}$.

For $i = 1,\ldots,n_b$, we define the residual as

$$R_i(v;\mu) = \tilde{a}(\xi_{i,N,M}(\mu), v; V_{\text{eff}}(\mu)) - \lambda_{i,N,M}(\mu)\tilde{a}_3(\xi_{i,N,M}(\mu), v), \tag{43}$$

for $\forall v \in Y$ where $\tilde{a}(w,v;V_{\text{eff}}(\mu)) = \tilde{a}_1(w,v) + \tilde{a}_2(w,v;V_{\text{eff}}(\mu))$. We also define a reconstructed error

$\hat{e}_i$ in Y, such that

$$\hat{a}(\hat{e}_i, v) = R_i(v;\mu), \quad \forall v \in Y, \tag{44}$$

where

$$\hat{a}(w,v) = \tilde{a}_1(w,v;m_2^*) + \tilde{a}_2(w,v;V_b) + (\gamma + \max_{\mu\in\mathcal{D},x_2\in\Omega^2} \phi(x_2;\mu))\tilde{a}_3(w,v); \tag{45}$$

$$\gamma = \left| \min_{\mu\in\mathcal{D},x_2\in\Omega^2} \phi(x_2;\mu) \right|; \tag{46}$$

$$\|R_i(\,\cdot\,;\mu)\| \equiv \sup_{v\in Y} \frac{R_i(v;\mu)}{\hat{a}(v,v)^{1/2}} = \hat{a}(\hat{e}_i,\hat{e}_i)^{1/2}; \tag{47}$$

and $\|\,\cdot\,\| = \hat{a}(\,\cdot\,,\,\cdot\,)^{1/2}$.

We now define $\tilde{a}^+(w,v;V_{\text{eff}}(\mu)) = \tilde{a}(w,v;V_{\text{eff}}(\mu)) + \gamma\tilde{a}_3(w,v)$ and introduce the following eigenvalue problem: for $\mu \in \mathcal{D}$, find $(\xi_i^+(\mu), \lambda_i^+(\mu)) \in Y \times \mathbb{R}$, $1 \le i \le n_e$ such that

$$\begin{aligned} \tilde{a}^+(\xi_i^+(\mu), v; V_{\text{eff}}(\mu)) &= \lambda_i^+(\mu)\tilde{a}_3(\xi_i^+(\mu), v), \quad \forall v \in Y, \quad 1 \le i \le n_e, & (48)\\ \tilde{a}_3(\xi_i^+(\mu), \xi_j^+(\mu)) &= \delta_{ij}, \quad 1 \le i \le j \le n_b. & (49) \end{aligned}$$

It is clear that $\xi_i^+(\mu) = \xi_i(\mu)$ and $\lambda_i^+ = \lambda_i + \gamma$.

Proposition 1. *Given $\hat{a}(w,v)$ as defined in (45), we have*

$$\hat{a}(v,v) \ge \tilde{a}^+(v,v;V_{\text{eff}}(\mu)) \ge \tilde{a}_3(v,v) \ge 0, \tag{50}$$

for all $\mu \in \mathcal{D}$.

Proof. We first prove left inequality. Let $f(\cdot) = \max_{\mu\in\mathcal{D},x_2\in\Omega^2} \phi(x_2;\mu)$. By expanding $\tilde{a}^+$, we obtain

$$\begin{aligned} \tilde{a}^+(v,v;V_{\text{eff}}(\mu)) &= \tilde{a}_1(v,v;m_2^*) + \tilde{a}_2(v,v;V_b) + \tilde{a}_2(v,v;\phi(\mu)) + \gamma\tilde{a}_3(v,v) \\ &\ge \tilde{a}_1(v,v;m_2^*) + \tilde{a}_2(v,v;V_b) + \tilde{a}_2(v,v;f) + \gamma\tilde{a}_3(v,v), \end{aligned} \tag{51}$$

since $\tilde{a}_1(v,v;m_2^*) \ge 0$; $\tilde{a}_2(v,v;V_b) \ge 0$ as $V_b \ge 0$, $\tilde{a}_3(v,v) \ge 0$, and $\gamma \ge 0$. Since the R.H.S of (51) is equivalent to $\hat{a}(v,v)$, left inequality is proven.

To prove the right inequality, we first note that

$$\tilde{a}(v,v;V_{\text{eff}}(\mu)) \ge \lambda_1(\mu)\tilde{a}_3(v,v),$$

and $\lambda_1(\mu) \ge \min_{x_2\in\Omega^2}\{V_b(x_2) + \phi(x_2;\mu)\}$. Then,

$$\begin{aligned} \tilde{a}^+(v,v;V_{\text{eff}}(\mu)) &= \tilde{a}(v,v;V_{\text{eff}}(\mu)) + \min_{\mu\in\mathcal{D},x_2\in\Omega_2} \phi(x_2;\mu)\tilde{a}_3(v,v) \\ &\ge (\min_{\mu\in D} \lambda_1(\mu) + \min_{\mu\in\mathcal{D},x_2\in\Omega_2} \phi(x_2;\mu))\tilde{a}_3(v,v) \\ &\ge \tilde{a}_3(v,v), \end{aligned} \tag{52}$$

since $\min_{x_2\in\Omega^2} V_b(x_2) = 0$. This concludes the proof for Proposition 1. □

Hypothesis 1. *Assuming our reduced-basis approximation is convergent in the sense that*

$$\lambda_{i,N,M}(\mu) \to \lambda_i(\mu), \quad 1 \le i \le n_e, \quad \textit{as} \quad N \to \infty. \tag{53}$$

Then, for sufficiently large N,

$$i = \arg \min_{1 \le j \le \mathcal{N}} \left| \frac{\lambda_{i,N,M}(\mu) - \lambda_j(\mu)}{\lambda_j^+(\mu)} \right|. \tag{54}$$

Proposition 2. *Assume our reduced-basis approximation is convergent in the sense given by (53). Then, for large N and $i = 1, \ldots, n_e$,*

$$\left| \frac{\lambda_{i,N,M}(\mu) - \lambda_i(\mu)}{\lambda_i(\mu) + \gamma} \right| \le \frac{\|R_i(\,\cdot\,;\mu)\|}{(\lambda_{i,N,M}(\mu) + \gamma)^{1/2}} \tag{55}$$

In addition, for $\lambda_{i,N,M}(\mu)$ of multiplicity one and associated $u_{N,i}(\mu)$, we have

$$\|u_{i,N,M}(\mu) - u_i(\mu)\| \le \frac{\|R_i(\,\cdot\,;\mu)\|}{d_i}, \tag{56}$$

and

$$|\lambda_{i,N,M}(\mu) - \lambda_i(\mu)| \le \frac{\|R_i(\,\cdot\,;\mu)\|^2}{d_i^2}, \tag{57}$$

where $d_i = \min_{j \ne i} \left| \frac{\lambda_{j,N,M}(\mu) - \lambda_{i,N,M}(\mu)}{\lambda_{j,N,M}(\mu) + \gamma} \right|$.

Proof. The proof utilizes Proposition 1 and details of the proof can be found in [29]. □

We note that (57) will in general be a better bound due to the $\|R_i\|^2$ term. Numerical experiments also indicate this is so. We thus define our error estimators based on (56) and (57) :

$$\Delta_{N,M}^{\lambda}(\mu) = \max_{1 \le i \le n_e} \frac{1}{d_i^2} \frac{\|R_i(\,\cdot\,;\mu)\|^2}{|\lambda_{i,N,M}(\mu)|}, \tag{58}$$

$$\Delta_{N,M}^{\xi}(\mu) = \max_{1 \le i \le n_e} \frac{1}{d_i} \frac{\|R_i(\,\cdot\,;\mu)\|}{\|\xi_{i,N,M}(\mu)\|}. \tag{59}$$

We can construct efficient offline-online computational strategies for the evaluation of our error estimators (58) – (59). From (45) and our reduced basis approximation, we have

$$\begin{aligned} \hat{a}(\hat{e}_i, v) &= \tilde{a}_1(\xi_{i,N,M}(\mu), v; m_2^*) + \tilde{a}_2(\xi_{i,N,M}(\mu), v; V_b) \\ &+ \sum_{m=1}^{M^\phi} \beta_m(\mu) \tilde{a}_2(\xi_{i,N,M}(\mu), v; q_m^\phi) + \bar{\varepsilon}_{M+1} \tilde{a}_2(\xi_{i,N,M}, v; q_{M^\phi+1}^\phi) \\ &- \lambda_{i,N,M}(\mu) \tilde{a}_3(\xi_{i,N,M}(\mu), v), \quad v \in Y, \quad 1 \le i \le n_e. \end{aligned} \tag{60}$$

where $\bar{\varepsilon}_M = \max_{\mu \in \mathcal{D}} \hat{\varepsilon}_M(\mu)$ and $\hat{\varepsilon}_M(\mu) = |\phi(t_{M^\phi+1}^\phi; \mu) - \phi_M(t_{M^\phi+1}^\phi; \mu)|$. It then follows from linear superposition that

$$\begin{aligned} \hat{e}_i(\mu) &= \sum_{n=1}^{N} \xi_{i,N,M\,n}(\mu) \left\{ p_n^1 + p_n^2 + \sum_{m=1}^{M^\phi} \beta_m p_n^{2+m} + \bar{\varepsilon}_M p_n^{M+3} \right\} \\ &- \lambda_{i,N,M}(\mu) \sum_{n=1}^{N} \xi_{i,N,M\,n}(\mu) p_n^0, \end{aligned} \tag{61}$$

where

$$\begin{aligned}
\hat{a}(p_n^1, v) &= a_1(\zeta_n, v; m_2^*), \quad v \in Y, \quad 1 \le n \le N, \\
\hat{a}(p_n^2, v) &= a_2(\zeta_n, v; V_b), \quad v \in Y, \quad 1 \le n \le N, \\
\hat{a}(p_n^{2+m}, v) &= a_2(\zeta_n, v; q_m^\phi), \quad v \in Y, \quad 1 \le n \le N, \quad 1 \le m \le M^\phi + 1 \\
\hat{a}(p_n^0, v) &= a_3(\zeta_n, v), \quad v \in Y, \quad 1 \le n \le N.
\end{aligned}$$

Then, $\|R_i(\,\cdot\,;\mu)\|$ is given by

$$\begin{aligned}
\|R_i(\,\cdot\,;\mu)\|^2 &= \hat{a}(\hat{e}_i, \hat{e}_i) \\
&= \sum_{k=1}^{3+M^\phi} \sum_{k'=0}^{3+M^\phi} \sum_{n=1}^{N} \sum_{n'=1}^{N} \Theta_k(\mu)\Theta_{k'}(\mu)\xi_{i,N,M\,n}(\mu)\xi_{i,N,M\,n'}(\mu)\hat{A}_{n,n'}^{k,k'} \\
&\quad + \sum_{n=1}^{N} \sum_{n'=1}^{N} \lambda_{i,N,M}^2(\mu)\xi_{i,N,M\,n}(\mu)u_{i,N,M\,n'}(\mu)\hat{A}_{n,n'}^{0,0} \\
&\quad + \sum_{n=1}^{N} \sum_{n'=1}^{N} \sum_{k=1}^{3+M^\phi} u_{N,i\,n}(\mu)\lambda_{N,i}(\mu)\Theta_k(\mu)\hat{A}_{n,n'}^{q,0}; \qquad (62)
\end{aligned}$$

where $\hat{A}^{k,k'} \in \mathbb{R}^{N\times N}$, $0 \le q, q' \le Q$ are given by $\hat{A}_{n,n'}^{k,k'} = \hat{a}(p_n^k, p_{n'}^{k'})$, $0 \le p, p' \le M^\phi + 3$, $1 \le n, n' \le N$, $\Theta_1 = 0$, $\Theta_2 = 0$, $\Theta_{2+m} = \beta_m, 1 \le m \le M^\phi$, and $\Theta_{M^\phi+3} = \bar{\varepsilon}_M$. We now see that the dual norm of the residual is the sum of products of parameter-dependent functions and parameter-independent functionals. The offline-online decomposition is now clear.

In the offline stage, we compute p_n^k, $0 \le k \le M^\phi + 3$, $1 \le n \le N$, based on (60) at the cost of $O((4+M)N\mathcal{N}^\bullet)$, where the $\bullet$ denotes computational complexity of the linear solver used to obtain p_n^k. We then evaluate $\hat{A}^{k,k'}$ at the cost of $O(4+M)N^2\mathcal{N}^2)$. We store the matrices $\hat{A}^{k,k'}$ at a total cost of $(4+M)N^2$.

In the online stage, we simply evaluate the sum (61) for a given $\xi_{i,N,M}(\mu)$ and $\lambda_{i,N,M}(\mu)$, $1 \le i \le n_e$. The operation count is only $O(n_e M^2 N^2)$. The online complexity is thus independent of $\mathcal{N}$. Unless M is large, the online cost to compute the error estimator is then a fraction of the cost required to obtain $\xi_{i,N,M}(\mu)$ and $\lambda_{i,N,M}(\mu)$.

E. Solution Method

In each fixed point iteration, part (i) of the solution method described in Section II B will now consist of (a) constructing the reduced basis machinery required to approximate $\xi(x_2; x_1)$ and $\lambda(x_1)$, and their derivatives to a required level of accuracy, and (b) approximating $\xi_i(x_2; x_1)$ and $\lambda_i(x_1)$ for finite points on Ω_1 by $\xi_{i,N,M}(x_2; x_1)$ and $\lambda_{i,N,M}(x_1)$. In part (a), we must construct the reduced basis approximation space W_N and the relevant reduced basis matrices described in earlier sections; we construct W_N based on an adaptive greedy algorithm [23, 26] that utilize the error estimator to very efficiently choose a good set of S_N. Note that we do not need to reconstruct the reduced basis machinery at each fixed point iteration. Armed with the *a posteriori* error estimators, we only reconstruct the reduced basis machinery when the estimated errors of the solutions based on W_N of the previous iteration are above the tolerance we desired. This significantly reduces the cost of reduced basis method by limiting the number of times we need to perform the expensive "offline" computation. The procedure is summarized in Figure 3.

Several variations to the above procedure. For example, a more frequent reconstruction may lead to smaller N, thus reducing the cost of "online" calculation. Thus, one could impose compulsory reconstruction of W_N at fixed intervals; at present we do not impose this as N required is generally small. In addition, we do not expect N to change drastically since ϕ only changes slightly for each

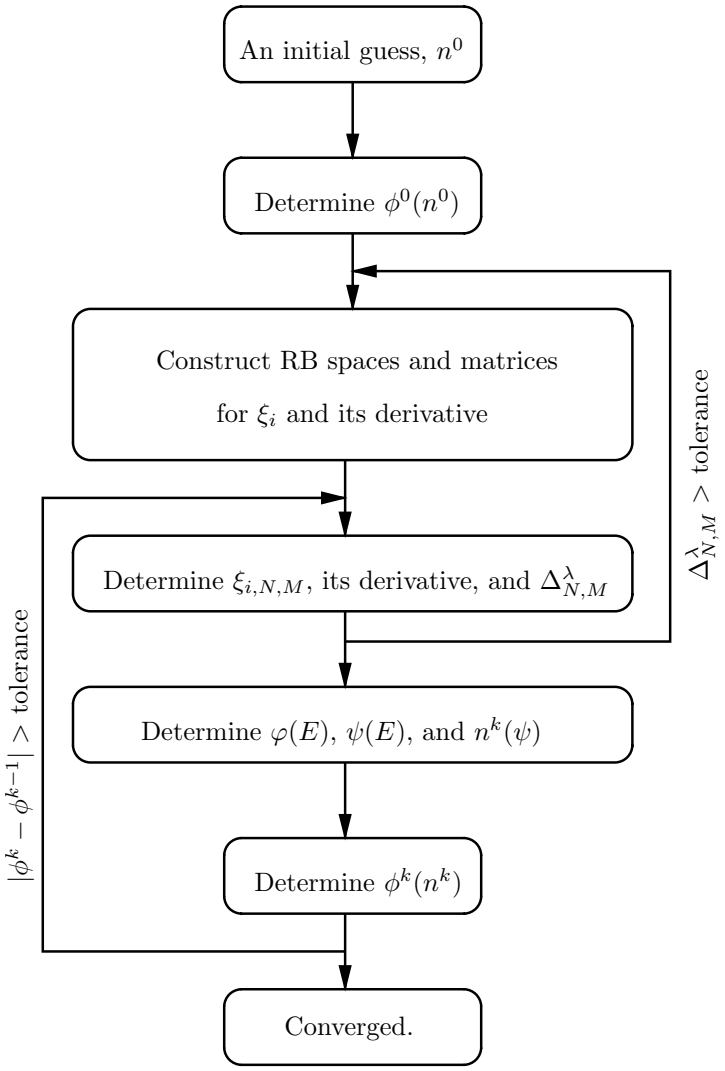

FIG. 3: Subband decomposition procedure with reduced basis approximation in part (i).

iteration. We could also reduce the offline computational cost by reconstructing the W_N based on existing S_N. While this removes the cost associated with greedy sampling procedure, we are less certain that the approximation space will be optimal and the solutions within the tolerance we desired.

IV. NUMERICAL RESULTS

We consider a domain $\Omega = [0, 340] \times [0, 120]$, which can be divided into subdomains detailed in Table I, which also gives the relative dielectric constant, ϵ_r, and donor concentration, N_D in each subdomain. The source and gate voltages applied are $V_S = 0$ and $V_G = 0.015$ while drain voltage, V_D is allowed to vary. We consider a temperature of 10^{-4}, and $P^S_{\max} = P^D_{\max} = 8$. To evaluate (18), we use $E_{\max} = 20T$ since $f_{1/2}(-E_{\max}/T) < 10^{-8}$. In addition, $n_{\mathrm{E}} = 50$ and $Q = 3$.

In Section IV A and IV B, we will first look at the accuracy of reduced basis method in approx-

	extent	$\epsilon_r = \epsilon/\epsilon_0$	N_D
Ω_1	$[80, 260] \times [20, 100]$	11.7	0
Ω_2	$[0, 80] \times [20, 100]$	11.7	2.96×10^{-5}
Ω_3	$[26, 340] \times [20, 100]$	11.7	2.96×10^{-5}
Ω_4	$[0, 340] \times [100, 120]$	3.9	0
Ω_5	$[0, 340] \times [0, 20]$	3.9	0

TABLE I: Definition of Ω_1 – Ω_5, and ϵ_r and N_D used in the model problem; $\epsilon_0 = 1/4\pi$.

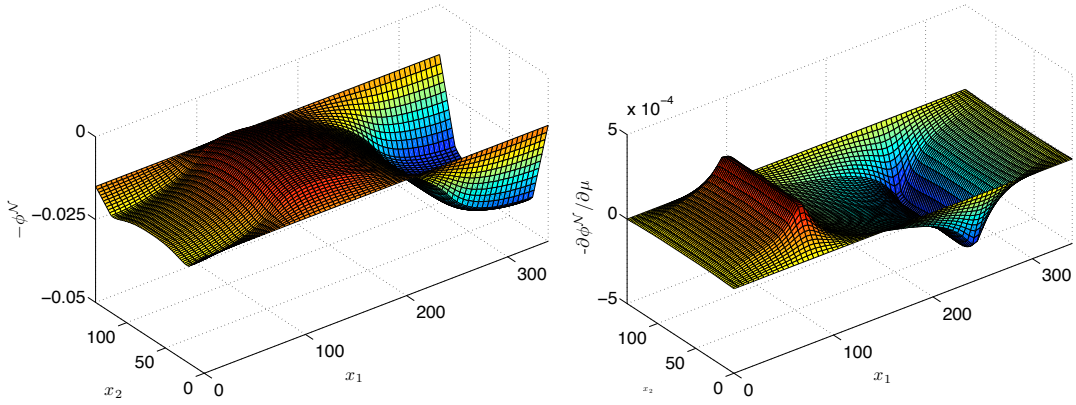

FIG. 4: $\phi^{\mathcal{N}}$(Left) and $\partial\phi^{\mathcal{N}}/\partial\mu$ (Right) for $V_D = 0.015$. The superscript $\mathcal{N}$ indicates that it is a finite element approximation of ϕ.

imating $\xi_i(x_1)$, $1 \leq i \leq n_e$. We denote solutions obtained through finite element method as the "truth" approximation — we build our reduced basis approximation on, and measure the error in the reduced basis approximation relative to this "truth" approximation. Note that since reduced basis approximation is build upon this "truth" approximation, it cannot perform better than this "truth" approximation. In Section IV C, we compare effects of using reduced basis method in part (i) on accuracy and efficiency of subband decomposition method. In our fixed point iterative scheme, $\alpha = 0.1$ and the convergence criteria is given by $|\phi^k - \phi^{k-1}|_{L_\infty} < 10^{-5}$, following the criteria used in [10].

A. Empirical interpolation approximation of ϕ and $\partial\phi/\partial\mu$

We first examine the approximation of ϕ and $\partial\phi/\partial\mu$ based on the empirical interpolation method. Figure 4 shows the solutions of ϕ and $\partial\phi/\partial\mu$ at convergence for the case $V_D = 0.015$. We note that the variation of $\phi(x_2; \mu)$ with respect to μ is nontrivial. The empirical interpolation errors of ϕ_M and $\partial\phi_M/\partial\mu$, denoted by $\bar{\varepsilon}_M^{\phi}$ and $\bar{\varepsilon}_M^{d\phi}$ respectively, are shown in Figure 5. The figure shows that we have a rapidly converging approximation — with $M^{\phi} = 20$ and $M^{d\phi} = 21$, the error $\bar{\varepsilon}_M^{\phi}$ and $\bar{\varepsilon}_M^{d\phi}$ are less than 10^{-8}.

B. Convergence of the reduced basis approximation

For our convergence analysis, the test sample Ξ_μ is given by the number of grid points in the x_1 direction — for the test problem, the size of Ξ_μ is 68. We will also define the following error

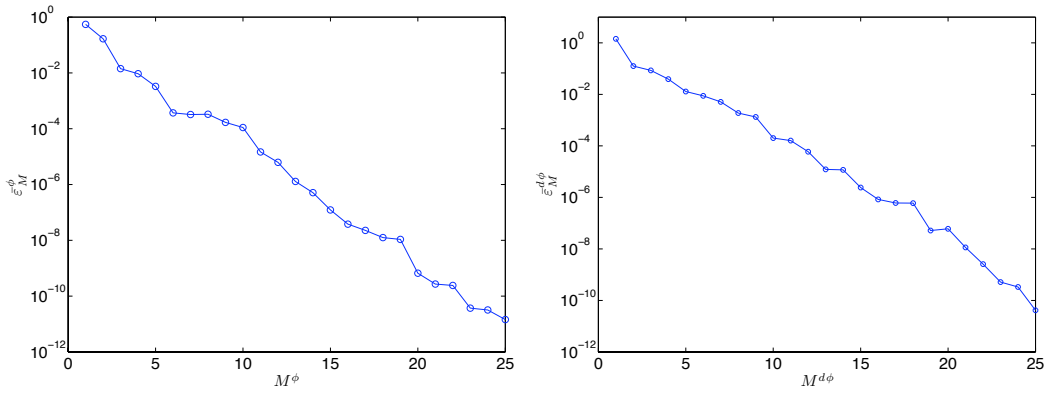

FIG. 5: ε_M versus M for ϕ_M (Left) and $\frac{\partial \phi_M}{\partial \mu}$ (Right).

N_s	N	$\varepsilon^\lambda_{N,M}$	$\varepsilon^\xi_{N,M}$	$\max_{\mu\in\Xi_\mu} \Delta^\lambda_{N,M}$	$\max_{\mu\in\Xi_\mu} \Delta^\xi_{N,M}$	$\max_{\mu\in\Xi_\mu} \eta^\lambda_{N,M}$	$\max_{\mu\in\Xi_\mu} \eta^\xi_{N,M}$
1	8	4.48 E−3	1.02 E−1	1.74 E−2	8, 85 E−2	4.38 E+1	5.87 E+0
2	16	1.42 E−5	1.66 E−3	1.79 E−4	8.10 E−3	4.99 E+1	6.69 E+0
3	24	2.02 E−9	3.53 E−5	9.10 E−8	2.58 E−4	1.48 E+1	7.31 E+0
4	32	5.69 E−11	6.09 E−6	2.09 E−9	3.49 E−5	6.31 E+2	4.27 E+1

TABLE II: Convergence of the reduced basis approximation for $V_D = 0.015$.

measures:

$$\varepsilon^\lambda_{N,M} = \max_{\mu\in\Xi_\mu} \epsilon^\lambda_{N,M}(\mu), \quad \varepsilon^\xi_{N,M} = \max_{\mu\in\Xi_\mu} \epsilon^\xi_{N,M}(\mu), \tag{63}$$

where

$$\epsilon^\lambda_{N,M}(\mu) = \max_{1\le i\le n_e} \frac{|\lambda_{i,N,M}(\mu) - \lambda_i(\mu)|}{|\lambda_i(\mu)|}, \tag{64}$$

$$\epsilon^\xi_N(\mu) = \max_{1\le i\le n_e} \frac{\|\xi_{i,N,M}(\mu) - \xi_i(\mu)\|}{\|\xi_i(\mu)\|}. \tag{65}$$

We also define the effectivity measures as

$$\eta^\lambda_{N,M}(\mu) = \frac{\Delta^\lambda_{N,M}(\mu)}{\epsilon^\lambda_{N,M}(\mu)}, \quad \eta^\xi_{N,M}(\mu) = \frac{\Delta^\xi_{N,M}(\mu)}{\epsilon^\xi_{N,M}(\mu)}. \tag{66}$$

Table II show that our reduced basis approximation is rapidly convergent. We require only 24 basis functions to reduce the relative errors $\varepsilon^\lambda_{N,M}$ to below 10^{-8} and $\varepsilon^\xi_{N,M}$ to below 10^{-4} for the case where $V_D = 0.015$ and $M^\phi = M^{d\phi} = 25$. In addition, the effectivity measures are small, indicating that our error estimators are good surrogate to the actual errors. Although $\eta^\lambda_{N,M}$ and $\eta^\xi_{N,M}$ increase with N, $\Delta^\lambda_{N,M}$ and $\Delta^\xi_{N,M}$ also decrease — thus the absolute difference between the actual errors and the error estimators is small.

We now look at the reduced basis error in $d\xi_{i,N,M}(\cdot)$, $d\lambda_{i,N,M}(\cdot)$, $1 \le i \le n_e$ and

N_s	N	$\varepsilon^{d\lambda}_{N,M}$	$\varepsilon^{d\xi}_{N,M}$	$\varepsilon^{\tilde{a}_3}_{N,M}$
1	8	3.94 E − 1	9.99 E − 1	2.10 E + 0
2	16	3.77 E − 4	1.15 E − 2	1.11 E − 2
3	24	8.83 E − 7	7.03 E − 6	6.59 E − 6
4	32	3.12 E − 9	7.98 E − 8	5.35 E − 8

TABLE III: Convergence of the reduced basis approximation of $d\lambda_{i,N,M}$ and $d\xi_{i,N,M}$, $1 \le i \le n_e$ for $V_D = 0.015$.

N_s	N	$\varepsilon^{d\lambda}_{N,M}$	$\varepsilon^{d\xi}_{N,M}$	$\varepsilon^{\tilde{a}_3}_{N,M}$
1	16	1.32 E − 3	1.54 E − 2	5.58 E − 2
2	32	5.40 E − 10	5.29 E − 9	4.14 E − 8

TABLE IV: Convergence of the reduced basis approximation of $d\lambda_{i,N,M}$ and $d\xi_{i,N,M}$, $1 \le i \le n_e$ for $V_D = 0.015$ with W^d_N.

$\tilde{a}_3(d\xi_{i,N,M}, \xi_{j,N,M}), 1 \le i, j, \le n_e$. We define

$$\varepsilon^{d\lambda}_{N,M} = \max_{\mu\in\Xi_\mu} \max_{1\le i\le n_e} \frac{|d\lambda_{i,N,M}(\mu) - d\lambda_i(\mu)|}{|d\lambda_i(\mu)|}, \tag{67}$$

$$\varepsilon^{d\xi}_{N,M} = \max_{\mu\in\Xi_\mu} \max_{1\le i\le n_e} \frac{\|d\xi_{i,N,M}(\mu) - d\xi_i(\mu)\|_Y}{\|d\xi_i(\mu)\|_Y}, \tag{68}$$

$$\varepsilon^{\tilde{a}_3}_{N,M} = \max_{\mu\in\Xi_\mu} \max_{1\le i,j\le n_e} \frac{|\tilde{a}_3(d\xi_{i,N,M}(\mu), \xi_{j,N,M}(\mu)) - \tilde{a}_3(d\xi_i(\mu), \xi_j(\mu))|}{|\tilde{a}_3(d\xi_i(\mu), \xi_j(\mu))|} \tag{69}$$

From Table III, we again see the rapid convergence in the errors defined by (67) – (69). In particular, the error in $\tilde{a}_3(\cdot,\cdot)$, which determines the effects of reduced basis approximation on the subband decomposition method, decreases rapidly with N. For a relative error of 10^{-5}, $N = 24$. Since the magnitude of $\tilde{a}_3(d\xi_i(\mu), \xi_j(\mu))$ is of order 10^{-4}, the absolute error in the approximation is actually very small.

As indicated in Section III B, we now examine the approximation of $d\xi_i(\mu)$ in W^d_N given by (32). We note that the solutions $(\xi_{i,N,M}, \lambda_{i,N,M})$ must also be determined in $W^d_N \times \mathbb{R}$. From Table IV, we indeed see a faster convergence in the errors with respect to N_s. However, the total number of basis, N, also increases with N_s at a rate double that of W_N. As such, for higher accuracy, W^d_N can indeed be a better approximation space although for the current purpose, W_N appears to be sufficient and leads to a smaller N.

C. Effects of reduced basis approach on efficiency of subband decomposition method

We denote the methods where we approximate part (i) of the subband decomposition method by finite element method and reduced basis method as SDM/FEM and SDM/RBM respectively; part (ii) and (iii) are approximated by finite element method for both approaches. The finite element approximation of part (i) is implemented using $\mathbb{P}_1$ elements with $\mathcal{N} = 71$ while the reduced basis approximation uses the accuracy criteria given by $\Delta^{\lambda}_{N,M} < 10^{-8}$.

In Table V, we compare the computational cost of the two methods for $V_D = 0$ and 0.015. For $V_D = 0$, the total computational time is reduced by a factor of 2, *without* any appreciable loss of accuracy — the L_2 norm error of $n_{N,M}$ obtained through SDM/RBM is of order 10^{-11}. If we consider only the computational time spent in part (i), the computational savings is significant higher — the computational time is reduced by a factor of 5. In addition, the reduced basis approximation spaces are only reconstructed 3 times, out of the 22 fixed point iterations needed for convergence. We see similar results for $V_D = 0.015$. Note that the computational cost of part (i) includes the cost of

Case	Subband		Reduced Basis				
	Time, s		Time, s				
	Total	Part (i)	Total	Part (i)	N_{offline}	$k_{\max}$	$\|n^{\mathcal{N}} - n_{N,M}\|_{L_2}$
$V_D = 0$	730	546	305	112	3	22	$4.03\,\text{E}-11$
$V_D = 0.015$	926	717	378	165	5	28	$2.28\,\text{E}-11$

TABLE V: Comparison of the computational cost for the subband decomposition method and the reduced basis method. Here N_{offline} is the number of times W_N is reconstructed, and $k_{\max}$ is the maximum number of fixed-point iteration.

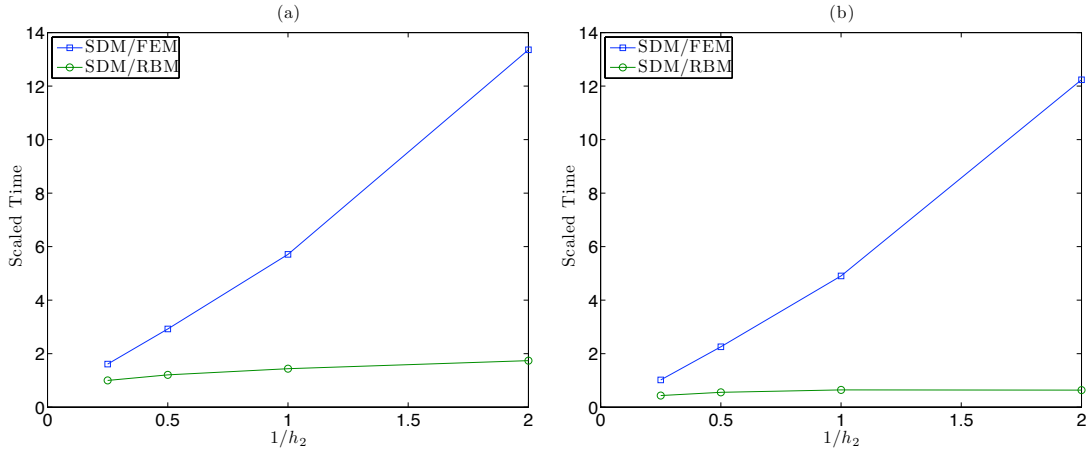

FIG. 6: Comparison of (a) the total computational time and (b) the computational time for part time, for the subband decomposition method and the reduced basis method with increasing mesh size. The time is scaled with respect to total time for the SDM/RBM method at $h_2 = 4$.

constructing the relevant matrices for use in part (ii).

We next examine how the computational time scales with respect to mesh size in the x_2-direction. Figure 6(a) compares the computational time of SDM/FEM and SDM/RBM for increasingly fine mesh; the reported time has been scaled with respect to the total computational time of SDM/RBM at $h_2 = 4$, where h_2 is the mesh spacing in the x_2-direction. we see that the total computational time of SDM/FEM scales exponentially with the mesh size. In particular, this exponential growth originates from the computational time of part (i) of the algorithm, as shown in Figure 6(b), while the computational time in part (ii) and (iii) only contribute slightly to the increase of the total computational time. On the other hand, the increase in the total computational time for SDM/RBM is slow. In particular, computational time of part (i) only increases very slightly, due to marginal increase in the computational cost of the offline stage; there should be little or no increase in the computational cost of the online stage. This observation strongly suggests that the reduced basis approach is particularly suited for situations where computational cost of part (i) dominates the total computational cost. For example, fine resolution may be needed in the x_2-direction due to strong confinement of the electrons. In nanowires and nanotubes where we have a 2-dimensional confinement, the higher dimension will also lead to larger mesh size, thus increasing the computational cost of part (i).

Finally, we look at a quantity of interest, the drain current, I_{SD}. Figure 7 shows that we have a typical current-voltage relation for a MOSFET, where the rate of increase in I_{SD} decreases as the applied voltage V_{SD} increases. We further note that SDM/RBM method gives comparable result to SDM/FEM method.

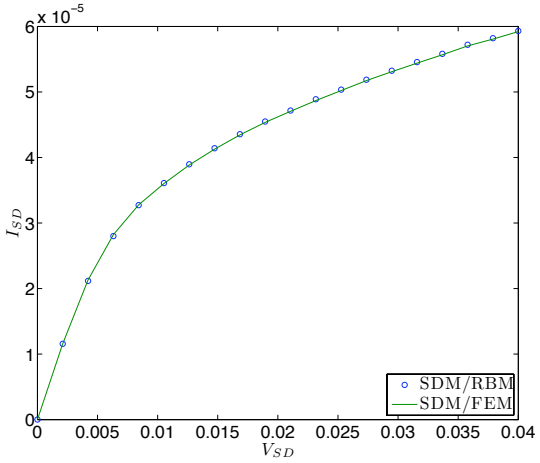

FIG. 7: Comparison of the computed drain current for SDM/FEM and SDM/RBM.

V. CONCLUSION

We have described how reduced basis method can improve the efficiency of the subband decomposition approach to ballistic transport simulation in nanodevices. In particular, the novel use of *a posteriori* error estimator and adaptive sampling procedure leads to a very efficient solution procedure. Numerical results based on a double-gate MOSFET show that the computational cost is reduced by 50% – 60% for a reasonably-sized problem and depends very weakly on the mesh size in the confined direction. We expect the computational savings to increase in cases of 2D confinement, such as those encountered in nanowires.

APPENDIX A: DERIVATIVE OF ϕ

To solve (23), we must evaluate $\partial\xi_n/\partial x_1$; in [10], ξ_n, $\partial\xi_n/\partial x_1$ and $\tilde{a}_3(\cdot,\cdot)$ are evaluated at the nodes (i,j) of the rectangular mesh, and interpolated to the quadrature points when evaluating the functionals in (23). In addition, $\partial\xi_n/\partial x_1$ are evaluated by difference formula. In our approach, $\partial\xi_n/\partial x_1$ are determined from (25), and this involves determining $\partial\phi/\partial x_1$ at the nodes (i,j). However, as we have used $\mathbb{Q}_2$ elements to solve for ϕ, its derivative is discontinuous, and thus not defined at the nodes. So, we compute the $\partial\phi/\partial x_1$ based on a difference formula. We then compute ξ_n, $\partial\xi_n/\partial x_1$ and $\tilde{a}_3(\cdot,\cdot)$ at the nodes (i,j) of the rectangular mesh, and interpolate to the quadrature points when evaluating the functionals in (23).

To avoid evaluating $\partial\phi/\partial x_1$ at the nodes, we can choose to compute $\partial\xi_n/\partial x_1$ directly at the quadrature points used to evaluate the functionals in (23). The reduced basis approximation procedure is as follows:

1. Compute $\partial\phi/\partial x_1$ at $(i+1/2,j)$, where $i+1/2$ is the midpoint between i and $i+1$.

2. Construct a magic point approximation for $\partial\phi/\partial x_1$, and the reduced basis machinery for $\partial\xi_n/\partial x_1$.

3. Evaluate the terms $\xi_{n,N,M}$, $\partial\xi_{n,N,M}/\partial x_1$ and $\tilde{a}_3$ at the quadrature points. To evaluate $\xi_{n,N,M}$ and $\partial\xi_{n,N,M}/\partial x_1$, values of ϕ_M and $\partial\phi_M/\partial x_1$ at the magic points for a given quadrature point must first be determined. For ϕ_M, these are obtained by the interpolation of the $\mathbb{Q}_2$ elements. For $\partial\phi_M/\partial x_1$, since the gradient between node $(i,t_M^{d\phi})$ and $(i+1,t_M^{d\phi})$ is a constant, the values

at the magic points for quadrature point falling between $(i, t_M^{d\phi})$ and $(i+1, t_M^{d\phi})$ is given by the value at node $(i+1/2, t_M^{d\phi})$; $t_M^{d\phi}$ are the magic points for $\partial\phi_M/\partial x_1$.

The above formulation should then be consistent with the $\mathbb{Q}_2$ elements we use. It is however more expensive: the computational cost of part (i) is increased by 66%. Determining the accuracy of the two approaches is also tricky. A comparison to, say, a full finite element approximation may be necessary although approximation error of subband decomposition method may dominate. In addition, the convergence criteria used in the fixed point iteration is not stringent, and any difference between the two approaches may not be discernible.

ACKNOWLEDGMENTS

I would like to thank A.T. Patera, Y. Maday and C. Le Bris for their guidance, and J.B. Bell for useful discussions. This work was supported by the Director, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

---

[1] K. Natori, Journal of Applied Physics **76**, 4879 (1994).
[2] C. S. Lent and D. J. Kirkner, Journal of Applied Physics **67**, 6353 (1990).
[3] E. Polizzi and N. Ben Abdallah, Phys. Rev. B **66**, 245301 (2002).
[4] S. E. Laux, A. Kumar, and M. V. Fischetti, Journal of Applied Physics **95**, 5545 (2004).
[5] R. Venugopal, Z. Ren, S. Datta, M. S. Lundstrom, and D. Jovanovic, Journal of Applied Physics **92**, 3730 (2002).
[6] Z. Ren, R. Venugopal, S. Goasguen, S. Datta, and M. Lundstrom, Electron Devices, IEEE Transactions on **50**, 1914 (Sept. 2003).
[7] A. Trellakis, A. T. Galick, A. Pacelli, and U. Ravaioli, Journal of Applied Physics **81**, 7880 (1997).
[8] L. R. Ram-Mohan, K. H. Yoo, and J. Moussa, Journal of Applied Physics **95**, 3081 (2004).
[9] L. R. Ram-Mohan, *Finite Element and Boundary Element Applications to Quantum Mechanics* (Oxford University, Oxford, 2002).
[10] E. Polizzi and N. B. Abdallah, J. Comput. Phys. **202**, 150 (2005).
[11] N. B. Abdallah, M. Mouis, and C. Negulescu, J. Comput. Phys. **225**, 74 (2007).
[12] J. Wang, E. Polizzi, and M. Lundstrom, Journal of Applied Physics **96**, 2192 (2004).
[13] B. O. Almroth, P. Stern, and F. A. Brogan, AIAA Journal **16**, 525 (1978).
[14] A. K. Noor and J. M. Peters, AIAA Journal **18**, 455 (1980).
[15] J. P. Fink and W. C. Rheinboldt, Z. Angew. Math. Mech. **63**, 21 (1983).
[16] M. D. Gunzburger, *Finite Element Methods for Viscous Incompressible Flows* (Academic Press, 1989).
[17] K. Ito and S. S. Ravindran, Journal of Computational Physics **143**, 403 (1998).
[18] J. S. Peterson, SIAM J. Sci. Stat. Comput. **10**, 777 (1989).
[19] T. A. Porsching, Mathematics of Computation **45**, 487 (1985).
[20] M. A. Grepl, N. C. Nguyen, K. Veroy, A. T. Patera, and G. R. Liu, in *Proceedings of the 2nd Sandia Workshop of PDE-Constrained Optimization: Towards Real-Time PDE-Constrained Optimization* (2007), SIAM Computational Science and Engineering Book Series, pp. 197 – 215.
[21] L. Machiels, Y. Maday, I. B. Oliveira, A. Patera, and D. Rovas, C. R. Acad. Sci. Paris, Série I **331**, 153 (2000).
[22] Y. Maday, A. T. Patera, and G. Turinici, C. R. Acad. Sci. Paris, Série I **335**, 289 (2002).
[23] C. Prud'homme, D. Rovas, K. Veroy, Y. Maday, A. Patera, and G. Turinici, Journal of Fluids Engineering **124**, 70 (2002).
[24] N. C. Nguyen, K. Veroy, and A. T. Patera, in *Handbook of Materials Modeling*, edited by S. Yip (Springer, 2005), pp. 1523–1558.
[25] K. Veroy and A. T. Patera, International Journal for Numerical Methods in Fluids **47**, 773 (2005).
[26] K. Veroy, C. Prud'homme, D. V. Rovas, and A. T. Patera, in *Proceedings of the 16th AIAA Computational Fluid Dynamics Conference* (2003), paper 2003-3847.
[27] M. Barrault, N. C. Nguyen, Y. Maday, and A. T. Patera, C. R. Acad. Sci. Paris, Série I. **339**, 667 (2004).

[28] M. A. Grepl, Y. Maday, N. C. Nguyen, and A. T. Patera, M2AN (Math. Model. Numer. Anal.) **41**, 575 (2007).
[29] G. S. H. Pau, Physical Review E **76**, 046704 (2007).
[30] G. Bastard, *Wave Mechanics Applied to Semiconductor Heterostructures* (Wiley-Interscience, New York, 1991).
[31] Y. Maday, N. C. Nguyen, A. T. Patera, and G. S. Pau, Constructive Approximation (2007), submitted.