A Novel Parallel Approach for Numerical Solution of the Schrödinger and Poisson Equations in Semiconductor Devices

YIMING LI, CHUAN-SHENG WANG, JAM-WEM LEE, TIEN-SHENG CHAO, and S. M. SZE
1 Department of Electronics Engineering and Institute of Electronics, National Chiao Tung University
2 Department of Mathematics, National Tsing Hua University; 3 National Nano Device Laboratories
P.O. Box 25-178, Hsinchu
TAIWAN

Abstract: - A new parallel implementation of quantum confinement effects simulations for semiconductor devices is presented. In this simulation, a set of self-consistent Schrödinger and Poisson (SP) equations is solved with parallel divide and conquer and monotone iterative algorithms on a Linux-cluster with message-passing interface (MPI) library. To solve the Schrödinger equation, instead of the conventional large-scale approach for eigenvalue problem, a novel parallel divide and conquer scheme is developed to find all the corresponding wave functions and energy levels. Moreover, the nonlinear Poisson equation is solved with monotone iterative method instead of the Newton’s iterative method. The parallel implementation shows that a well-designed simulation can reduce the execution time up to many orders of magnitude. Compared with the measured data and classical results, numerical simulations on a realistic metal-oxide-semiconductor (MOS) device are presented to show the accuracy and efficiency of the method.

Key-Words: - Schrödinger and Poisson Equations, Parallel Divide and Conquer, Monotone Iterative Technique, Semiconductor Device Simulation, Quantum Confinement Effects

1 Introduction

Numerical modeling and simulations of ultra-thin semiconductor devices [1, 2] have been proven to be an indispensable tool for the analysis and optimal design of various semiconductor devices. As the dimensions of devices continue to shrink [3], the development of sophisticated and efficient semiconductor device technology CAD software will provide the engineers significant leverage in conducting research into new integrated circuits technologies. For the most of MOS devices, quantum confinement effects in inversion layers have been become important technologically [3, 4], it has been well known that SP approach is the best one since it fully includes the quantum mechanics in a consistent fashion [5]. However, the numerical solution of SP model is a time consuming task and not widely used for engineering applications so far [4].

In this present work, a robust and efficient parallel simulation approach for the solution of SP model is proposed and successfully implemented on a Linux-cluster with MPI. The Schrödinger and Poisson equations are discretized with finite volume method (FVM) [2] and a self-consistent iterative algorithm is applied to compute the solution. The discretized Schrödinger equation leads to an eigenvalue problem, the corresponding eigenvalue (energy levels) and eigenvectors (wave functions) are computed with parallel divided and conquer (DC) algorithm. Furthermore, the discretized nonlinear Poisson equation is solved with monotone iterative (MI) method instead of Newton’s iterative method. Compared with Newton’s iterative method, the major features of the monotone iterative method are as follows: (i) it converges globally for various semiconductor devices under various bias conditions, (ii) it is inherently parallel and its implementation is much easier than Newton’s iterative method [1, 2].

The developed SP simulator is then applied to fast study the impacts of quantum confinement effects on the interpretation of capacitance-voltage (C-V) data for ultra-thin gate dielectrics in MOS structure. Compared with the measured and classical C-V results, numerical simulations of the SP model on a 2.2nm oxide thickness MOS device are presented to show the accuracy of the method. Benchmarks, such as speedup and efficiency are also given to show the computational performance.

This paper is organized as follows. In Sections 2, the Schrödinger and Poisson equations in semiconductor devices are described. In Section 3, the numerical methods and parallel approach on a Linux-cluster with MPI are presented, respectively. In Section 4, simulation results and measurement
data for a submicron MOS device are presented to demonstrate the accuracy and efficiency of the method. Benchmark results included the achieved speedup, efficiency, and related performances are also given in this section. Finally, summary and conclusion are given in Section 5.

2 The Schrödinger and Poisson Equations in Semiconductor Devices

The most obvious quantum mechanical effect, seen in a very thin oxide, is gate leakage via direct tunneling through the oxide [3, 4]. It also has been shown that the inversion layer charge density calculated using quantum mechanics approach is smaller than that calculated classically for a given applied gate voltage, thus affecting the shift of the subthreshold curves. The inversion-layer capacitance is a capacitance inherent to MOS structure and this capacitance has been thought to have a significant influence on the performance of scaled MOS devices with ultra-thin gate dielectrics.

Equation (1) is used to describe the behavior of the electron distribution along z direction. The potential energy, \( E(z) \), is related to the electron distribution via the Poisson equation (2) and it is a sum of potential energy and pseudopotential energy due to the band offset at the semiconductor insulator interface. Fig. 1 shows a typical MOS device structure with ultra-thin film gate oxide. The one-dimensional SP model is defined along z direction from metal gate to substrate. In (1), \( z \) is the coordinate along the perpendicular direction, \( \Psi_{i,j}(z) \) is the wave function for the jth subband in the ith valley, \( E \) is the potential energy, and the \( m_{ni}^* \) is electron effective mass in the ith valley. The \( E(z) \) in (1) is connected to the potential \( \phi(z) \) in (2), and its expression is given by (3). The wave function \( \Psi_{i,j}(z) \) in (1) and the electron density \( n(z) \) in (2) are related by (4), where \( g_i \) and \( m_{ni}^* \) are the ith valley degeneracy and the ith density of state effective mass. The basic SP model to describe hole quantization phenomena is similar to equations (1)-(4). For simplicity, the dielectric constants \( \varepsilon(z) \) used in this study are 11.9 for Si and 3.9 for SiO₂, respectively [3]. The solution of equations (1)-(4) provides the electron density, and hence the device capacitance with quantum confinement effects could be calculated from the simulated surface charge density [3]. The C-V characteristic of MOS structure with different gate dielectrics is fast investigated by means of the developed parallel quantum mechanical simulator. The quantum transport results are found to compare quite well with experimental data. Comparisons are also made with results obtained using classical simulations and this approach as well.

3 Numerical Methods and Parallel Algorithms

In this section, the self-consistent, parallel divided and conquer, and monotone iterative methods will be presented. Due to the dependence relationships among the potential \( \phi(z) \), wave functions \( \Psi_{i,j}(z) \), and energy levels \( E_{i,j} \), the solution of system (1)-(4) should consist a self-consistent loop to reach the final convergent solution. In addition, in order to solve the SP model (1)-(4) efficiently, a scalable and portable parallel divided and conquer method is proposed. It is designed for the fast calculating electron all wave functions and energy levels for a large sparse band
matrix that raising from the FVM discretization of Schrödinger equation. This method is stable and for a large class of matrices it is, asymptotically, faster by an order of magnitude than the conventional one. When the electron density is obtained, the FVM discretized Poisson equation is then solved with monotone iterative method.

Fig. 2 shows a self-consistent procedure for the solution of SP model. The iteration loops will be terminated until a specified stopping criterion is reached.

![Fig. 2. A flowchart diagram for solving a set of SP model self-consistently.](image)

Firstly, as shown in Fig. 2, the MOS device simulation domain and needed input parameters are initialized. Then the eigenvalue problem that comes from the discretized Schrödinger equation is solved with the proposed parallel DC method. Thirdly, if the convergence tests do not meet the specified stopping criteria, we update the newer computed wave functions \( \Psi_{i,j}(z) \) and energy levels \( E_{i,j} \), and solve the discretized Poisson equation with the MI method. When the final solutions are convergent, we perform a post-process to calculate the device C-V curves. It can be proved that the convergence property of this self-consistent iteration algorithm is guaranteed due to a contraction mapping theory.

The proposed DC algorithm to fast solve the corresponding eigenvalue problem is illustrated in Fig. 3. The DC algorithm for a tridiagonal eigenvalue problem has been applied to some practical problems [6, 7]. In this work, we applied the DC algorithm to solve the Schrödinger equation in semiconductor device. From the numerical experiment, it can be concluded that DC algorithm, when properly implemented, could be many times faster than traditional ones, such as bisection followed by inverse iteration or the QR algorithm, even on serial computers. The parallel implementation shows that a well-designed simulation can reduce the execution time up to many orders of magnitude. As shown in Fig. 3, the global matrix is split and scheduled equally by host server processor and then sends to each client processor. The eigensystems are computed with QR or Jacobi method independently. Only the eigenvalues will be estimated from these submatrices. Finally, computed results are conquered by a merge method, combined, and returned to the host processor. If all exact eigenvalues are obtained from each matrix, the solved eigenvalues will be merged directly and hence the corresponding wave functions will be constructed. If the engenvalues are partially obtained, the multiplicity of approximated eigenvalues will be checked by a proper estimation strategy, such as Householder transformation or solving a secular equation will be provided for these eigenvalues.

![Fig. 3 A divide and conquer algorithm for solving the eigenvalue system \( Ax = \lambda x \). The \( A \) is the matrix rising from discretized Schrödinger equation and boundary conditions, \( x \) and \( \lambda \) are the corresponding eigenvectors and eigenvalues of the matrix \( A \).](image)

To solve the FVM discretized Poisson equation globally, a monotone iterative algorithm is applied. This MI method was proposed and applied by us earlier to the semiconductor device simulation successfully [1, 2]. The MI formula is of the following form
\[ y^{(m+1)} = \left( L + U \right)^{(m)} - F(y^{(m)}) + \lambda y^{(m)} \]

\[ (D + \lambda I) \], \quad \text{(6)} \]

where \( Y \) is the unknown vector, \( F \) is the nonlinear vector form, and \( D, L, U, \) and \( I \) are diagonal, lower triangular, upper triangular, and identity matrices, respectively. The parameter \( \lambda \) depends on the device structure, bias condition, and nonlinear property of the Poisson equation. The Poisson equation in SP model is a highly nonlinear partial differential equation and is a tightly coupled with Schrödinger equation so that the convergence behavior does not always satisfied for all biases [4]. The MI method applied here for SP semiconductor device simulation [1, 2] does not involve any Jacobian matrix and simulation results show a very excellent global convergent property. However, the Newton’s iterative method inherently requires a Jacobian matrix and a sufficiently accurate initial guess to begin with the solutions. Note that the system (6) is of Jacobi type and hence is highly parallel. It does not require any assembling process for a global matrix. Consequently, the MI method is very cost effective in terms of both computational time and data storage memory.

Fig. 4 A parallel architecture for the DC method.

Fig. 4 shows the parallel computing architecture for the DC method. The server processor \#0 divides and schedules the jobs firstly, and all jobs will be send to client processors \#1-\#7 via network connection. All processors calculate the eigenpairs independently. All solved results are merged and feedback to the host processor for the next process.

Fig. 5 demonstrates the constructed and used Linux-cluster system and the network configuration in this work. Each cluster contains 16 PCs in this study; files access and share are through network file system (NFS) and network information system (NIS). The user datagram protocol (UDP) that controlled by MPI is applied to the short distance communication.

**4 Simulation Results and Discussion**

The SP simulation results are provided in this section, the experimental C-V curves and classical results [3] are compared with the SP results. It demonstrates that the SP results have a good consistent with the measured data for a thin-oxide thickness MOS device. It also shows that the inversion layer charge density calculated using quantum mechanics SP approach is indeed smaller than that calculated classically for a given applied gate voltage. Furthermore, the achieved parallel performance, for example the parallel timing table, speedup, and efficiency are presented here.

Fig. 6. The electron concentration distribution \( n(z) \) resulting from SP and classical simulation.

The studied MOS device structure is shown in Fig. 1. In this study, a MOS device with 2.2nm oxide film is prepared to do a comprehensive study for quantum confinement effects. First of all, as shown in Fig. 6, this example presents the SP simulation is necessary.
for thin-oxide MOS devices. Different from the SP approach, the most widely used classical result is obtained by solving the Poisson equation (2) with Boltzmann statistics [3]. The electron density shows that the classical approach becomes overestimation when the device oxide thickness down to 2.2nm region.

As shown in Fig. 6, the effective inversion layer thickness in Si surface is increased by several angstroms owing to quantization effects. The fundamental difference between them lies in the fact that a maximum of the electron concentration distribution is localized at the semiconductor surface according to the classical model whereas it is shifted from the surface according to the quantum mechanical theory. Moreover, it also has been verified theoretically that the average distance of electrons from the semiconductor surface resulting from the conventional classical method is half of that for the SP simulation. This effect is reason that classical method results in over estimated device capacitance. Figures 7 and 8 show the result clearly.

![Fig. 7. The C-V characteristics on the test n-MOS. The SP simulated (dot), measured (solid), and classical calculated (dash) full C-V curves of an Al-gate n-MOS device are presented. The oxide thickness is 2.2nm.](image)

The results, as shown in Figures 7 and 8, are presented to demonstrate the accuracy for the developed parallel quantum mechanical simulation with SP model. Fig. 7 shows the SP simulated and measured C-V curves. The SP model indicates both an increase in the magnitude of the threshold voltage and a degradation of the oxide capacitance in the inversion layer. Comparison among the classical, SP, and measurement data suggests the developed SP simulator provides a good accuracy for the test n-MOS device with $T_{\text{ox}}=2.2\text{nm}$. A similar test is also presented to show the robustness of the SP simulation in the Fig. 8. It demonstrates the simulated and measured C-V curves for the test n-MOS device with $T_{\text{ox}}=5.0\text{nm}$. The classical result in this case is close to the measured data, but it still have a 1.2fF/µm² difference at $V_{\text{GS}}=-2\text{V}$. However, the SP model has proved itself able to fit a full C-V curve of the test examples, as shown in Figures 7 and 8.

![Fig. 8. The C-V characteristics on the test n-MOS. The SP simulated (dot), measured (solid), and classical calculated (dash) full C-V curves of an Al-gate n-MOS device are presented. The oxide thickness is 5.0nm.](image)

Furthermore, parallel speedup and efficiency are given to show the computational performance for the novel approach. The following results contain the parallel simulation timing for various problem sizes. The sequential and parallel wall clock execution time is illustrated. The parallel efficiency of the method is measured by the parallel speedup [8]. Speedup is defined as the ratio of the code execution time on one processor to that on multiple processors; efficiency is the speedup divided by the number of processors. Tab. 1 shows the parallel time as a function of number of processors for the implemented parallel method.

<table>
<thead>
<tr>
<th>Number of Processor</th>
<th>Parallel timing (Sec.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP model with various matrix size</td>
<td></td>
</tr>
<tr>
<td>500²</td>
<td>1K²</td>
</tr>
<tr>
<td>2</td>
<td>41</td>
</tr>
<tr>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td>8</td>
<td>17</td>
</tr>
<tr>
<td>12</td>
<td>17</td>
</tr>
<tr>
<td>16</td>
<td>17</td>
</tr>
</tbody>
</table>

The measured execution time includes all self-consistent iteration loops where each eigenvalue
problem is solved with Jacobi method. For $V_G = 3V$, the self-consistent iteration loop takes only about 8-10 times to reach a specified maximum error $1E-6$ on the potential. For a single CPU, the execution time, for various matrix sizes $500^2$, $1k^2$, $2k^2$, $3k^2$, and $4k^2$ are 270, 2422, 29643, 110236, and 299196 seconds, respectively. In this case, the parallel DC method shows that it provides a new alternative to fast solve the SP model efficiently and can obtain a reasonable parallel speedup on a Linux-cluster with MPI library. Moreover, compared with the sequential execution time, the test also demonstrates that there is a good parallel time reduction when the matrix size is up to $3k^2 (3,000^2)$ and $4k^2$.

Fig. 9 shows a typical achieved parallel timing plot for matrix size is $3k^2$; for this case, the parallel time is reduced and tends to a fixed value significantly at 12 CPU. Fig. 10 and Table 2 give the achieved parallel speedup and efficiency on a 16-CPU Linux-cluster with MPI.

<table>
<thead>
<tr>
<th>Number of Processor</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SP model with various matrix size</td>
</tr>
<tr>
<td></td>
<td>$500^2$</td>
</tr>
<tr>
<td>1</td>
<td>100%</td>
</tr>
<tr>
<td>2</td>
<td>329%</td>
</tr>
<tr>
<td>4</td>
<td>375%</td>
</tr>
<tr>
<td>6</td>
<td>375%</td>
</tr>
<tr>
<td>8</td>
<td>196%</td>
</tr>
<tr>
<td>12</td>
<td>126%</td>
</tr>
<tr>
<td>16</td>
<td>91%</td>
</tr>
</tbody>
</table>

5 Conclusion

A computational efficient parallel simulation approach for SP model has been proposed. The proposed parallel DC with MI method for the first time to solve a self-consistent SP model has been successfully implemented on a Linux-cluster with MPI. This new simulation technique provides a fast, efficient, and accurate approach to extract device parameters in which quantum confinement effects are significant. The quantum transport results are also found to compare quite well with experimental data. Benchmarks, such as speedup and efficiency are also presented to show the computational performance.

References: