A Novel Machine-Learning Based Mode Space Method for Efficient Device Simulations

Yeongjun Lim and Mincheol Shin
School of Electrical Engineering
Korea Advanced Institute of Science and Technology
Daejeon, 34141, Republic of Korea
mshin@kaist.ac.kr

***Abstract*—In this work, we present a machine-learning based mode space method for efficient quantum transport simulations. By introducing a novel concept of projectability in training a machine learning model, our method reduces the size of the Hamiltonian more effectively compared to previous methods, while faithfully reproducing both the real and imaginary bands of interest. Through performing InAs nanowire FET simulations using the density functional theory (DFT) Hamiltonian and non-equilibrium Green's function (NEGF) method, we demonstrate that our method enables highly efficient simulations without losing accuracy.**

***Index Terms*—density functional theory, non-equilibrium Green's function, mode space method, field-effect transistors**

I. Introduction

As the technical node of electronic devices shrinks to the nanoscale regime, it becomes increasingly important to use atomistic simulations that accurately account for quantum effects. One powerful method for studying nanodevices is to utilize Non-equilibrium Green's function (NEGF) method with density functional theory (DFT) Hamiltonian. However, the computational challenge of employing large-sized DFT Hamiltonians for realistic device simulations remains a significant obstacle to overcome.

To tackle this issue, the mode space (MS) method has emerged as a successful technique for reducing the size of the DFT Hamiltonian while the errors in the transport characteristics are kept within a few percent or less [1], [2]. To further advance the methodology, we have developed in this work a novel MS method based on machine learning scheme that accurately reproduces the electrical transport properties of the device.

To validate the effectiveness of our proposed method, we have applied it to the DFT Hamiltonian and conducted quantum transport simulations of InAs nanowire field-effect transistors (FETs) using the NEGF method. Our method enables highly efficient simulations that reliably capture the complex characteristics of nanoscale devices.

II. Simulation Approach

The fundamental concept of the MS method is to reduce the Hamiltonian size by unitarily transforming the full Hamiltonian using a smaller basis set which mainly contributes to the carrier transport. The transformation is done by

$$\begin{aligned} h &= U^{+}HU, \\ s &= U^{+}SU \end{aligned} \tag{1}$$

where H and S are the full Hamiltonian and overlap matrix of size $N \times N$, respectively, U is the transformation matrix of size $N \times m$ $(m \ll N)$, so the reduced Hamiltonian h and overlap matrix s has a size of $m \times m$.

Finding the appropriate basis set for such a transformation is a complex problem that does not have a unique solution. In the previous MS method applied to tight-binding or $ab\text{-}initio$ Hamiltonians, a subset of the Bloch states is selected to build the initial transformation matrix [1], [3]. Consequently, the removal of unphysical states becomes necessary, a process

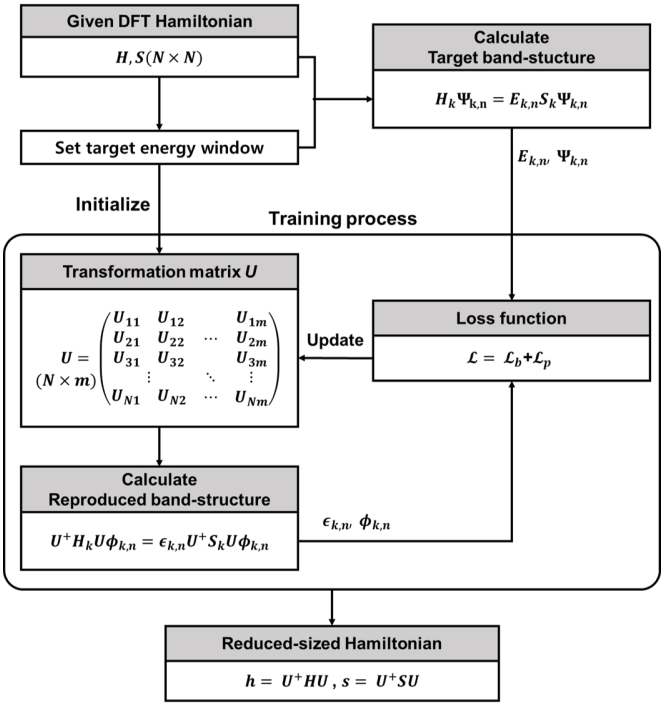

Fig. 1. Schematic of the training workflow to find the transformation matrix U that reduces the Hamiltonian size.

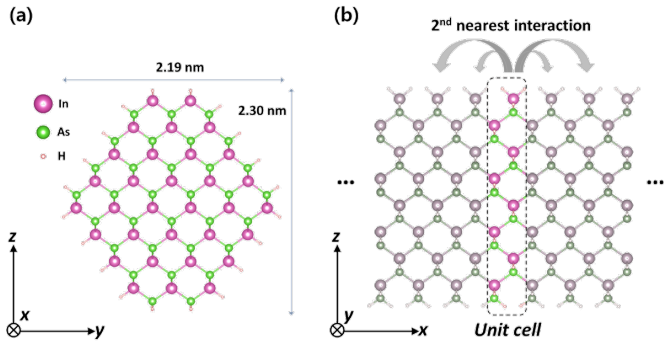

Fig. 2. Atomic structure of the simulated InAs nanowire unit cell which has 2^{nd} nearest interaction range: (a) cross-sectional view and (b) side view.

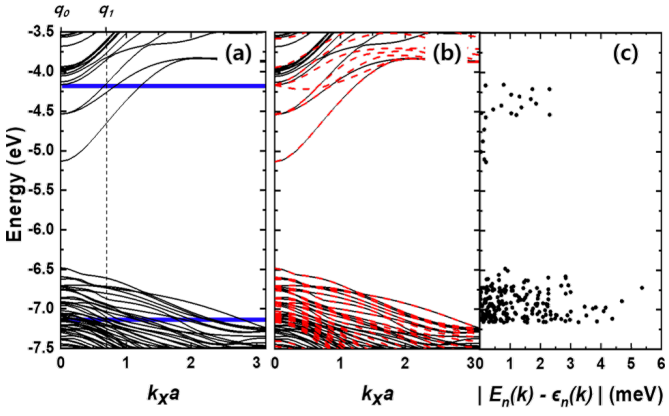

Fig. 3. (a) The band structure of the InAs nanowire unit cell calculated by using the full Hamiltonian and (b) comparison with the use of the reduced Hamiltonian (red dashed lines). (c) Error between the target and reproduced band structure. The target energy window for the MS method is highlighted with the blue line.

that is often challenging and relies primarily on the initial choice of the Bloch states [4].

In contrast to the conventional MS method constructs U by sampling Bloch states, we utilized machine learning algorithms to directly obtain an optimized U. We trained a network customized to a given structure, as opposed to a typical neural network that is trained to predict the properties of other structures. The network is trained to ensure that the reduced-sized Hamiltonian accurately reproduces the properties contributing to carrier transport within a given energy window.

To this end, we have introduced a novel approach that takes into account not just the band structure, but also the projectability factor. The latter factor is crucial in assessing how effectively the reduced-sized Hamiltonian represents the original wavefunctions. To address this, we have defined the loss function $\mathcal{L}$ as

$$\mathcal{L} = \mathcal{L}_b + \mathcal{L}_p \tag{2}$$

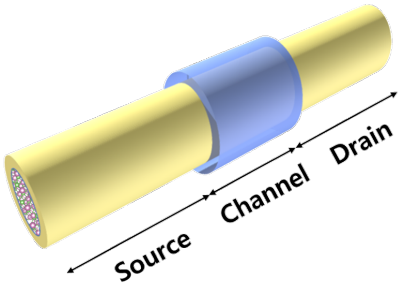

Fig. 4. Schematic of gate-all-around InAs nanowire FET constructed by using the unit cell of Fig. 2.

$$\mathcal{L}_b = w_b \sum_k \sum_n \{E_n(k) - \epsilon_n(k)\}^2 \tag{3}$$

$$\mathcal{L}_p = \sum_q \sum_n w_p (1 - p_{qn})^2 \tag{4}$$

where w_b and w_q are weighting factors, $E_n(k)$ is target band structure within an energy window of interest obtained from the full Hamiltonian, $\epsilon_n(k)$ is reproduced band structure, and $p_{qn} (0 \leq p_{qn} \leq 1)$ is the projectability between the wavefunctions from the full Hamiltonian and those from the reduced Hamiltonian at selected wave vectors q.

By minimizing $\mathcal{L}$, the optimized U is obtained, leading to a reduced-size effective Hamiltonian that can be used in place of the original Hamiltonian for transport simulations. The effective Hamiltonian greatly reduces computational burden while maintaining all the characteristics needed for accurate simulations. The overall flow of the method is shown in Fig. 1.

III. Results

A. InAs nanowire

We applied the MS method to the DFT Hamiltonian of the InAs nanowire unit cell shown in Fig. 2, which has a 2^{nd} nearest neighbor interaction range. We used the SIESTA package to calculate electronic structures and relax the atomic structure [5]. The generalized gradient approximation exchange-correlation functional by Perdew, Burke, and Ernzerhof (GGA-PBE) was used [6]. The atomic structure is relaxed until the maximum force became less than 0.05 eV/Å. The unit cell contains 88 atoms including pseudo-hydrogen atoms passivating the surfaces, and the Hamiltonian size is 1272×1272.

In this case, two wave vectors $q(q_0, q_1)$ were selected to calculate $\mathcal{L}_p$ as shown in Fig. 3 (a). Through the process of training the network, the size of the Hamiltonian is reduced to 54×54 which is only 4.2 % of the full Hamiltonian. Although the size of the Hamiltonian is greatly reduced, the band structures are well reproduced with minimal error

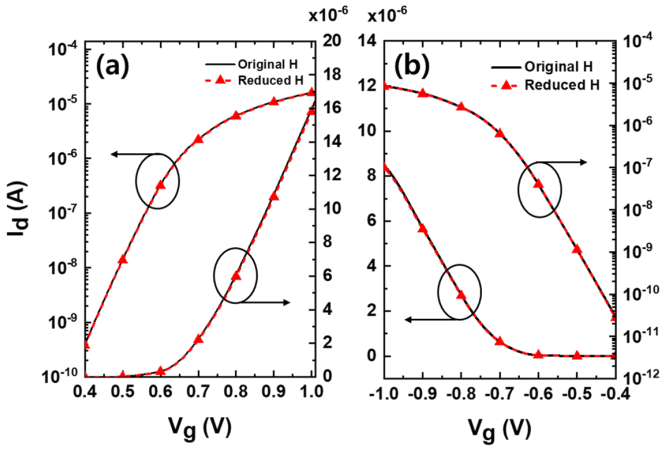

Fig. 5. Comparison of the transfer curves calculated by using the full Hamiltonian (black lines) and reduced Hamiltonian (red triangles on dashed lines): (a) n-type and (b) p-type InAs nanowire FETs. The length of the source, channel, and drain is 10 nm each. The applied drain bias V_{d} is 0.5 V.

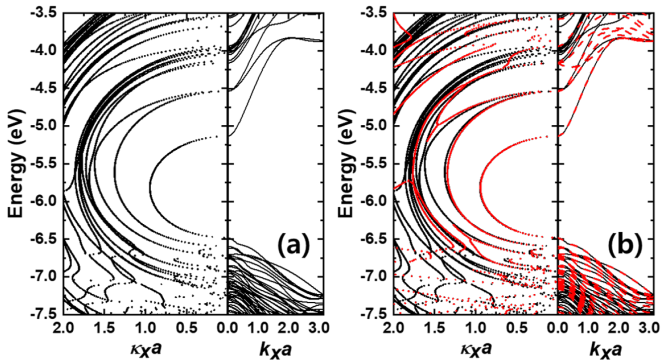

Fig. 6. (a) The complex band structure of the InAs nanowire unit cell calculated by using the full Hamiltonian and (b) comparison with the use of the reduced Hamiltonian (red dotted lines).

within the target energy window of [-7.16, -4.15] eV as shown in Fig. 3 (b) and (c).

B. n / p-type FETs

The effectiveness of the MS method on device simulation is evaluated by employing the reduced-sized Hamiltonian in our in-house quantum transport simulator that solves NEGF and Poisson's equation self-consistently [1]. We first considered typical n-type and p-type FETs. The schematic of the simulated gate-all-around InAs nanowire FET is shown in Fig. 4. We calculated the transfer characteristics of the FETs using both original and reduced Hamiltonian as shown in Fig. 5. Our results indicate that the use of the reduced Hamiltonian with the MS method leads to high accuracy for both types of FETs.

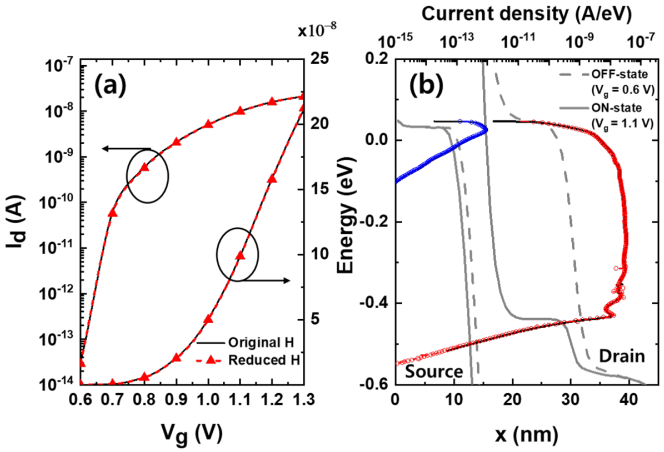

Fig. 7. (a) Transfer characteristics of InAs nanowire TFET calculated by using the full Hamiltonian (black lines) and the reduced Hamiltonian (red triangles on dotted lines). (b) Band profiles and current spectrum at ON and OFF-state calculated by using the original Hamiltonian (black) and reduced Hamiltonian (red for ON-state and blue for OFF-state). The length of the p-type source, intrinsic channel, and n-type drain is 15 nm each. The applied drain bias V_{d} is 0.5 V.

C. Tunnel FET

Accurately capturing tunneling behavior in nanoscale devices is of utmost importance. In particular, devices that exploit tunneling effects, such as tunnel FETs (TFETs), have gained significant attention as potential candidates for next-generation logic devices [7]. In this light, we conducted simulations on an InAs TFET using the effective Hamiltonian to test our method.

For a tunneling device, the complex band structure that represents the evanescent states plays an important role along with the real band structure [8], [9]. Fig. 6 demonstrates that even without explicitly including information on the complex band structure in the loss function, the 1$^{\mathrm{st}}$ evanescent mode that mainly contributes to the tunneling current is accurately replicated by minimizing the projectability loss $\mathcal{L}_p$.

The calculated transfer characteristics and current spectrums at OFF and ON-state of the InAs TFET are shown in Fig. 7. We set V_{OFF} as the gate voltage (V_{g}) for the onset of band-to-band tunneling (BTBT) where the potentials of the source and channel overlap, and V_{ON} = V_{OFF} + V_{d}. Despite considering the projectability of wavefunctions for only two wave vectors $q(q_0, q_1)$, the BTBT current contributed by evanescent modes is excellently reproduced.

IV. Conclusion

The practical impact of this work lies in two key contributions. Firstly, the new mode space (MS) method developed in this study achieves more reduction in Hamiltonian size than the previous MS method, which can lead to orders of magnitude improvement in computational speed. Secondly, by introducing the novel concept of projectability, the direct evaluation of the imaginary bands, which are often

cumbersome and computationally burdensome, is avoided. This facilitates easier implementtation of the MS method and improves its overall efficiency in device simulations. Although we used the DFT Hamiltonian in this work, our method is versatile and can be applied to other Hamiltonians obtained through various methods, such as effective-mass, k.p, and tight-binding.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Science and ICT (RS-2023-00243425).

REFERENCES

[1] M. Shin, W. J. Jeong, and J. Lee, "Density functional theory based simulations of silicon nanowire field effect transistors," J. Appl. Phys., vol. 119, no. 15, 154505, 2016.
[2] M. Shin, "Hetero-structure mode space method for efficient device simulations," J. Appl. Phys., vol. 130, no. 10, 104303, 2021.
[3] G. Mil'nikov, N. Mori, and Y. Kamakura, "Equivalent transport models in atomistic quantum wires," Phys. Rev. B, vol. 85, 035317, 2012.
[4] M. Shin, S. Jeon, and K. Joo, "Efficient atomistic simulations of lateral heterostructure devices with metal contacts," Solid-State Electron., vol. 198, 108456, 2022.
[5] J. M. Soler, E. Artacho, J. D. Gale, A. Garcia, J. Junquera, P. Ordejon, and D. Sanchez-Portal, "The SIESTA method for ab initio order-N materials simulation," J. Phys. Condens. Matter, vol. 14, no. 11, pp. 2745–2779, 2002.
[6] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," Phys. Rev. Lett., vol. 77, no. 18, pp. 3865–3868, 1996.
[7] A. M. Ionescu and H. Riel, "Tunnel field-effect transistors as energy efficient electronic switches," Nature, vol. 479, no. 7373, pp. 329–337, 2011.
[8] E. O. Kane, "Theory of Tunneling," J. Appl. Phys., vol. 32, no. 1, pp. 83-91, 1961.
[9] M. Luisier and G. Klimeck, "Simulation of nanowire tunneling transistors: From the Wentzel-Kramers-Brillouin approximation to full-band phonon-assisted tunneling," J. Appl. Phys., vol. 107, 084507, 2010.