



- (51) International Patent Classification:
G06F 17/50 (2006.01)
- (21) International Application Number:
PCT/US2014/057840
- (22) International Filing Date:
26 September 2014 (26.09.2014)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
61/883,158 26 September 2013 (26.09.2013) US
61/883,942 27 September 2013 (27.09.2013) US
61/889,355 10 October 2013 (10.10.2013) US
- (71) Applicant: SYNOPSYS, INC. [US/US]; 700 E. Middlefield Rd., Mountain View, CA 94043 (US).
- (72) Inventors: LIU, Jie; 1273 McKendrie Street, San Jose, CA 95126 (US). MOROZ, Victor; 17035 Bohlman Road, Saratoga, CA 95070 (US). SHAUGHNESSY-CULVER, Michael, C.; 1068 Wood Duck Avenue, Santa Clara, CA 95051 (US). SMITH, Stephen, Lee; 1662 Columbia Drive, Mountain View, CA 94040 (US). OH, Yong-Seog; 2667 Calle Moreilia, Pleasanton, CA 94566 (US). BALASINGAM, Pratheep; 6249 Plymouth Avenue, San Jose, CA 95129 (US). MA, Terry Sylvan, Kam-Chiu; 510 Preakness Court, Danville, CA 94506 (US).
- (74) Agent: SUZUE, Kenta; Haynes Beffel & Wolfeld LLP, P.O. Box 366, Half Moon Bay, CA 94019 (US).
- (81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: PARAMETER EXTRACTION OF DFT

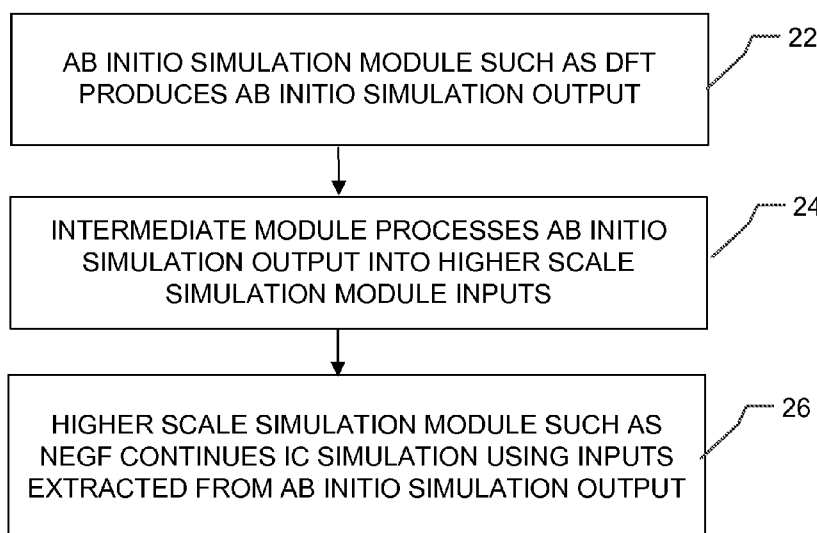


FIG. 2

(57) Abstract: Electronic design automation modules for simulate the behavior of structures and materials at multiple simulation scales with different simulation modules.



Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

— *with international search report (Art. 21(3))*

PARAMETER EXTRACTION OF DFTInventors:

Jie Liu
Victor Moroz
Michael C. Shaughnessy-Culver
Stephen Lee Smith
Yong-Seog Oh
Pratheep Balasingam
Terry Sylvan Kam-Chiu Ma

REFERENCE TO RELATED APPLICATIONS

- [1] The present application claims the benefit U.S. Application No. 61/883,158 filed September 26, 2013; and U.S. Application No. 61/883,942 filed September 27, 2013; and U.S. Application No. 61/889,355 filed October 10, 2013, incorporated by reference herein.
- [2] The following U.S. patent applications are incorporated by reference herein: U.S. Application titled "SIMULATION SCALING WITH DFT AND NON-DFT" with docket no. SYNP 2378-2 filed September 26, 2014; U.S. Application titled "ITERATIVE SIMULATON WITH DFT AND NON-DFT" with docket no. SYNP 2383-2 filed September 26, 2014.

BACKGROUNDField of the Invention

- [3] The present invention relates to electronic design automation, and modules for simulating the behavior of structures and materials at multiple simulation scales with different simulation modules.

SUMMARY

- [4] One aspect of the technology is an EDA tool comprising a data processor; and storage configured to provide computer program instructions to the processor.
- [5] The storage includes an intermediate module interacting with a plurality of simulation modules to perform an EDA simulation at a plurality of different simulation scales. The plurality of simulation modules includes a first set of one or more ab initio simulation modules at a first simulation scale; and a second set of one of more non-ab initio simulation modules at a second simulation scale larger than the first simulation scale.
- [6] The intermediate module processes outputs of the first set of one or more simulation tools to generate a set of semiconductor material quantities received and processed by the second set

of one or more simulation modules including any of: effective mass tensor, non-parabolicity, and N-band k.p model parameters where N is an integer.

[7] In one embodiment, the data processor is responsive to one or more of the input parameters to at least one of the second set of one or more non-ab initio simulation modules having an inaccuracy, such that the procedures automatically execute at least one of the first set of one or more simulation modules to provide results as said one or more of the input parameters to said at least one of the second set of one or more simulation modules, thereby lessening the inaccuracy.

[8] In one embodiment, the first set of one or more ab initio simulation modules includes a density functional theory module, and said second set of one or more non-ab initio simulation modules excludes the density functional theory module.

[9] In one embodiment, the set of semiconductor material quantities includes the effective mass tensor.

[10] In one embodiment, the set of semiconductor material quantities includes the non-parabolicity.

[11] In one embodiment, the set of semiconductor material quantities includes the N-band k.p model parameters.

[12] Another aspect of the technology is a computer-implemented method comprising:

[13] causing a plurality of simulation modules to perform an EDA simulation at a plurality of different simulation scales, the plurality of simulation modules including:

[14] a first set of one or more ab initio simulation modules at a first simulation scale; and

[15] a second set of one or more non-ab initio simulation modules at a second simulation scale larger than the first simulation scale; and

[16] processing outputs of the first set of one or more simulation tools to generate a set of semiconductor material quantities received and processed by the second set of one or more simulation modules including any of: effective mass tensor, non-parabolicity, and N-band k.p model parameters where N is an integer.

[17] Various embodiments are disclosed herein.

[18] Other aspects and advantages of the present technology can be seen on review of the drawings, the detailed description and the claims, which follow.

BRIEF DESCRIPTION OF THE DRAWINGS

[19] Figure 1 is a block diagram of various simulation modules at multiple simulation scales coupled together and using output of one module as input for another module, and automatically

executing another module at the same simulation scale or different simulation scale to improve inaccurate input (such as missing input).

[20] Figure 2 is an example block diagram of simulation modules at multiple simulation scales coupled together via an intermediate module using output of an ab initio simulation module as input for a higher scale simulation module

[21] Figure 3 is an example block diagram of simulation modules at multiple simulation scales coupled together via an intermediate module using output of a higher scale simulation module as input for an ab initio simulation module.

[22] Figure 4 is an example block diagram of simulation modules at multiple simulation scales coupled together via intermediate module processing, to iterate between an ab initio simulation module and a higher scale simulation module.

[23] Figure 5 is an example process flow of executing simulation tools automatically at multiple simulation scales, to improve inaccurate input (such as missing input).

[24] Figure 6 shows different simulation volumes of simulation modules at different simulation scales.

[25] Figures 7-8 shows a switch between simulation modules at different simulation scales, from a larger simulation volume to a smaller simulation volume.

[26] Figures 9-10 shows a switch between simulation module volumes, with horizontal and oblique asymptotes.

[27] Figures 11-12 show an error at the border between simulation modules at different simulation scales, from a smaller simulation volume to a larger simulation volume.

[28] Figures 13A, 13B, 13C and 13D illustrate various lattice configurations, as examples of arrangements in which a semiconductor property has a finite distortion range.

[29] Figure 14 is a simplified block diagram of a computer system configured to perform IC simulation at multiple scales.

[30] Figure 15 illustrates EDA tools and process flow for integrated circuit design and manufacturing.

DETAILED DESCRIPTION

[31] A detailed description of embodiments of the present invention is provided with reference to the Figures.

[32] As processing power increases sufficiently, scaling tools can simulate transistor behavior at increasingly granular levels, even at the atomic level. A trade-off exists among granularity of the simulation, and required computation resources. For example, fine granularity increases

required computation resources, and so tends to have reduced simulation volume. In another example, coarse granularity reduces computation resources, and so tends to have enlarged simulation volume.

[33] There are many levels of how detailed the physical model can be. Typically, simpler models can handle larger examples whereas more sophisticated models typically can only handle a smaller subset of such examples within comparable computation time. For example, so-called compact transistor models like BSIM represent a transistor as several empirical formulas that can be calculated within milliseconds. This enables such compact models to be used in characterizing large circuits that contain thousands of transistors. A more sophisticated Technology Computer-Aided Design (TCAD) model represents transistor as hundreds of thousands of interconnected points that are scattered throughout different parts of the transistor. Several key Partial Differential Equations (PDEs) are solved on these interconnected points and determine distribution of the charges and electrostatic potential throughout the transistor, as well as the charge transport between the terminals. The TCAD model typically handles a handful of transistors (say, from 1 to 5) in the same timeframe as the compact model handles thousands of transistors in a circuit. The upside for the TCAD model is that it can predict transistor behavior based on the known material properties, whereas the compact model typically needs to be calibrated either to TCAD model or to experiments. Yet another modeling level is Density Functional Theory (DFT), which typically cannot represent an entire transistor, but represents a small part of it, considering each atom separately and how the atoms are connected to each other. A collection of atoms of the order of 100 hundred, which is less than one percent of the atoms in a transistor, can be characterized by DFT in the same timeframe as TCAD handles a few transistors and the compact model handles a thousand transistors. The DFT model can predict material properties based on the atomic properties, and therefore can provide the input for the TCAD model to characterize the transistor. So, there is a hierarchy of modeling approaches, including DFT (sometimes referred to as “first principles” approach), TCAD, and compact models. The tools in this hierarchy are applied sequentially, going from detailed physics/small structure towards the simpler physics and larger size, and usually this works fine. However, there are cases where this hierarchical approach fails, and there are more such cases as the transistor scaling continues. There are some characteristics that typically can be obtained only with large enough structure size. For example, typically one can only get accurate electric field distribution and the current crowding/filamentation effects on transistor scale, i.e. a TCAD level model. Such characteristics can affect the inter-atomic bonds and the band structure, and therefore alter results

of DFT analysis that cannot calculate such characteristics on the scale it can handle. The present approach bridges this gap.

[34] **Figure 1** is a block diagram of various simulation modules at multiple simulation scales coupled together and using output of one module as input for another module, and automatically executing another module at the same simulation scale or different simulation scale to improve inaccurate input (such as missing input). Details of the blocks are discussed in U.S. Provisional Applications 61/883,158 filed September 26, 2013; and U.S. Application No. 61/883,942 filed September 27, 2013; and U.S. Application No. 61/889,355 filed October 10, 2013, incorporated herein by reference.

[35] This technique adaptively combines multiple simulation and design modules and methodologies, in order to offer automated, self-adaptive, and multi-scale simulation capabilities of various physical processes in the fabrication and operation of integrated circuit devices. The modules can include several hierarchical simulation modules, each of which focuses on the physical processes in one particular spatial scale. The modules can communicate in one direction, or back-and-forth with each other via parameter extraction in an automated and self-adaptive way.

[36] The multiple modules can have one or more of multiple benefits. Firstly, multiple modules offer systematic and comprehensive simulation capabilities of the physical processes in multiple spatial scales. While the existing simulation methodologies focus on the physical processes in one particular spatial scale, the modules include them as internal modules or at least modules communicating via an intermediate module, and offer systematic simulation functionalities in multiple spatial scales. Secondly, the modules help reducing the cost of experiment and testing by including ab initio simulations modules as the internal modules, or as external modules in communication via an intermediate module, which can offer valuable physical insights and replace empirical experiments and calibrations. Thirdly, the modules can significantly reduce computational time and consumption of computational resources, by replacing some computationally expensive calculations via parameter extraction. Fourthly, the modules are able to control the simulation automatically and self-adaptively, hence minimizing the needs of human control and intervention.

[37] Without the modules, one needs to be an expert of many sophisticated simulation modules and methodologies, to perform simulations of physical processes which span multiple spatial scales. The simulations performed in this fashion are functionality limited, physical expertise demanding, computationally expensive, human time intensive, and error prone. The modules are dedicated to solve these problems.

[38] Example modules shown in Figure 1 include: quantum Monte Carlo (QMC) 19; time-dependent density functional theory (TD-DFT) 15; density functional theory (DFT), DFT with on-site U (DFT+U) 14, and dynamic mean field theory (DMFT) 14; ab initio molecular dynamics (AIMD) 18; tight binding (TB) 13; classical molecular dynamics (MD) 17; Monte Carlo (MC) 16; non-equilibrium Green's function (NEGF) 12; Wigner and wave function formalism quantum transport 21; quantum and deterministic Boltzmann transport 10; and various continuum simulation modules and methodologies (continuum), etc.

[39] Example information flows between the different modules, are follows:

[40] The QMC module solves the many-body physics problems. The electronic structure and lattice structure properties calculated from QMC are the output of the modules. Since the mean field theory methodologies DFT and TD-DFT sometimes give imprecise results (e.g. band structure, band gap, etc.), the QMC module can verify the results. Similarly, the QMC module can benchmark the MD, AIMD, TB, MC, NEGF, and continuum modules, to simulate the atomic movement, transport properties, and various device performance metrics.

[41] The TD-DFT module solves the time dependent Schrodinger equation. The various optical response properties of materials and the atomic movement under external excitation can be obtained from the TD-DFT module. Since the DFT may not account for electronic excitation, the TD-DFT module can benchmark DFT results when the electronic excitation cannot be ignored. The inter-atomic force computed from TD-DFT is used in the AIMD and MD modules to obtain the atomic trajectory. The optical response related parameters used in the MC, NEGF, and continuum modules are computed from the TD-DFT module.

[42] The DFT module can include not only the DFT modules based on mean field theory and single particle approximation, but also strong correlation simulation algorithms like DMFT and DFT+U. The DFT module can generate the electronic structure. The inter-atomic forces computed in the DFT module is used to calculate the atomic movement in the AIMD module. The energy calculated in the DFT module is used to generate empirical force fields, which can be used in the MD module, by using methods like parameter fitting and/or optimization algorithms. The various phenomenological parameters in the TB, MC, and continuum modules are calibrated and optimized using batch calculations in the DFT module. The Hamiltonian and overlap matrices in the NEGF module can be obtained from the DFT module.

[43] DFT is an example of a class of approaches known as "*ab initio*" or "first principles" approaches. Such approaches require minimal empirical input to generate accurate ground state total energies for arbitrary configurations of atoms. These approaches make use of fundamental quantum mechanical equations, and require very little in the way of externally-supplied materials

parameters. This capability makes these methods well-suited to investigating new materials and to providing highly detailed physical insight into material properties and processes, but also renders them extremely computation intensive. As such, a task control system can control a multi-scale simulation project in which any of the *ab initio* approaches is used in combination with less computation intensive approaches such as 2D Schrödinger and TCAD. As used herein, an "*ab initio*" or "first principles" analysis approach or module is an approach or module that develops its results at least in part by solving Schrödinger's equation based on positions and types of atoms. Other example first principles approaches that can be used herein include EPM (Empirical Pseudo-potential Method) modules, and ETB (Empirical Tight Binding) modules, and combinations of approaches.

[44] The TB module is based on phenomenological expressions of various physical quantities. It can be used to compute electronic and transport properties. The inter-atomic forces calculated in the TB module can be used to generate the force field in the MD module. The TB description of the target system (e.g. Hamiltonian) is used in the MC, NEGF, and continuum models to obtain the physical properties pertaining to electron transport and device process.

[45] The NEGF module computes the transport properties, which are the output of the modules. The NEGF module can simulate not only transport with scattering, but also ballistic transport. So it can be used to benchmark the continuum module, especially when the transport is largely ballistic. Also, the physical properties (local density of states, transmission, mean free path, etc.) obtained from the NEGF module can offer valuable insights in the ultra-scaled devices. NEGF iterates between (i) the Poisson equation to get a 3D potential profile U , and (ii) the NEGF transport equation to get a density matrix ρ . NEGF also generates the following distributions: electron charge density, hole charge density, electron velocity, hole velocity, electrostatic potential, current (product of respective charge density and charge velocity).

[46] The MC module includes, but is not limited to, device MC, kinetic MC, and lattice kinetic MC. The device operation physics and fabrication process are the output of the MC module. The MC module can take the output from other modules as input. For example, the energy dumped by the DFT module can be used to estimate the activation energy in the MC module to evaluate the atomic migration probability. The MC module is capable of simulating many physical processes, e.g. dopant diffusion during process, electron transport, atom migration, etc.

[47] The quantum and deterministic Boltzmann module relies on the Boltzmann equation for calculation of transport properties.

[48] The Wigner and wave function formalism module calculates quantum transport properties.

[49] The continuum module accepts output of the other modules and use them as input, to calculate various device performance metrics, like charge distribution, current-voltage curve, etc. The continuum module includes various numerical algorithms like finite element method, finite volume method, finite difference time domain method, boundary element method, etc. The continuum module can be considered an intermediate module that communicates between different modules.

[50] The coordination module coordinates the other modules. If a particular module has missing input, the coordination module executes another module to provide the missing input. If a particular simulation scale is not specified in the execution instruction, the coordination module determines the appropriate scales to run. The proposed modules receive target simulation quantities via a user interface, and instruct the coordination module to start the simulation. After receiving the simulation instructions, the coordination module decides which functional block(s) should be called to perform these simulation tasks. To accomplish the entire user request, the functional blocks of the proposed module may be called many times adaptively and iteratively by the coordination module. The communication of information (like parameter request and extraction) between/among the functional blocks is controlled by the coordination module. The coordination module can be considered an intermediate module that coordinates different modules.

[51] A particular simulation module can be integrated into a suite of multiple simulation modules. Alternatively, intermediate modules can process output from discrete simulation modules for use as input by each other.

[52] In one example division of different simulation scales, a complete semiconductor device scale includes SPICE and other continuum modules. The ab initio simulation scale can include AIMD, DFT, DMFT, DFT+U, TD-DFT, and QMC. The intermediate simulation scale can include MC, NEGF, MD, and TB. Drift-diffusion can be a large scale. Different simulations scales can be viewed also as a continuum of scales, with finer gradations of scale, such that any pair of tools can have overlapping or non-overlapping simulation volumes.

[53] **Figure 2** is an example block diagram of simulation modules at multiple simulation scales coupled together via an intermediate module using output of an ab initio simulation module as input for a higher scale simulation tool. To take advantage of the accuracy of more granular simulation, and the lowered computation intensity of less granular simulation, parameters are passed between multiple levels of simulation.

[54] One such combination uses DFT to develop inputs for NEGF, and uses NEGF to simulate larger volumes than would be possible with DFT alone.

[55] At 22, results are produced from an ab initio simulation module such as DFT. At 24, results are processed into higher simulation scale input. Processing can be as minimal as passing a parameter, or performing multiple operations to extract data. At 26, the simulation is continued using the results from the ab initio simulation module as input of an intermediate scale simulation module such as NEGF.

[56] DFT receives input such as types and coordinates of all atoms that comprise the structure, coordinates of additional electrons and holes, and boundary conditions of electrostatic potential. DFT generates as output, band structure, bandgap E_g , Young modulus E , Poisson ratio ν , hole and electron effective mass m^* , permittivity, into larger scale simulations. For example, NEGF can directly use this output as the Hamiltonian. TCAD can adjust the mobility with this output. SPICE parameters can be adjusted with this output.

[57] One DFT embodiment performs the simulation with only internal fields or potential barriers. Another DFT embodiment performs the simulation with an external field output from NEGF or other higher scale modules.

[58] The scheme can offer one or more of multiple benefits. (1) It enables the calculation of large systems which are too computationally expensive for the ab initio simulation modules. (2) It enables simulation of the physical process whose duration is too long for the ab initio simulation modules. (3) It enables batch simulations of a large number of systems which are too computational resource consuming for the ab initio simulation modules. (4) It improves the calculation precision of the phenomenological simulation modules and methodologies by adaptively extracting parameters from the comparatively more accurate ab initio simulation modules. (5) It offers automated and self-adaptive calculation scheme, which minimizes needs of human intervention and maximizes productivity of simulation and design.

[59] Without the scheme, one typically needs to perform the parameter extraction process based on prior simulation experience and intuition, which are not only human time intensive but also error prone. This scheme solves this problem, by automating the parameter extraction from the ab initio simulation modules with an intermediate module, and using the extracted parameters in the phenomenological simulation modules.

[60] The ab initio scale simulation modules and methodologies mentioned above can include, but are not limited to, the following modules: quantum Monte Carlo (QMC), time dependent density functional theory (TD-DFT), density functional theory (DFT), and ab initio molecular dynamics (AIMD), etc.

[61] The intermediate scale phenomenological simulation modules and methodologies” mentioned above include, but not limited to, the following modules: classical molecular dynamics (MD), Monte Carlo (MC), tight-binding (TB), non-equilibrium Green’s function (NEGF), and continuum simulation methods. Different simulations scales can be viewed also as a continuum of scales, with finer gradations of scale, such that any pair of tools can have overlapping or non-overlapping simulation volumes.

[62] The intermediate module extracts parameters from ab initio density functional theory (DFT) simulations. The extracted parameters can be used as the inputs of other modules (e.g. Monte Carlo), to facilitate multi-scale simulations. Example applications of the DFT intermediate tool output are: capacitance calculator, circuit simulator for simulation operation of a circuit of devices, device simulator for computing current-voltage in one device or combination of devices, device simulator based on the solution of partial differential equations, device simulator based on the solution of the Boltzmann transport equation, device simulator based on quantum transport, and device simulator for the simulation of magnetic based devices.

[63] Multi-scale simulation assists downscaling integrated circuit semiconductor devices and exploration of novel materials. Downscaling makes the quantum mechanical effects more important. The new modules help understand new materials’ properties and help improve the device design. Users can simulate the devices by combining the ab initio quantum mechanical simulations from first-principles.

[64] One of the most mature and widely-applied ab initio algorithms is the density functional theory (DFT). Simulation modules, such as DFT modules, can be internal to the EDA tool, or an external module by a third party. Examples of DFT tools are VASP, SIESTA, Quantum Espresso, OpenMX, etc.

[65] The physical semiconductor quantities extracted from DFT discussed below in more detail include: bandgap (direct and indirect), effective mass scalar, effective mass tensor, non-parabolicity, and N-band k.p model parameters (such as Luttinger parameters). Although the example below discusses 6-band k.p model Luttinger parameters, other embodiments are directed to various numbers of bands (e.g., 2, 6, 8, 20, etc.). Such preceding parameters can be calculated in different directions such as x, y, and z due to anisotropic behavior. These quantities can be used as the input parameters of other tools.

[66] To extract the bandgap, the intermediate module relies on input of multiple k points, such as k_1 (k_{1x} , k_{1y} , k_{1z}) and k_2 (k_{2x} , k_{2y} , k_{2z}). Here, k_1 specifies the conduction band maximum (CBM) location, and k_2 specifies the valence band minimum (VBM) location.

[67] The intermediate module parses the DFT output data, to locate these two k points in the DFT data. Using the Fermi level, intermediate module will automatically determine from which

band the eigenvalues $E(\mathbf{k}1)$ and $E(\mathbf{k}2)$ are extracted. Finally, the result of $E(\mathbf{k}2) - E(\mathbf{k}1)$ is used as the bandgap value.

[68] The intermediate module can perform data interpolation. Alternatively, it parses DFT data and directly locates $E(\mathbf{k}1)$ and $E(\mathbf{k}2)$. The two $\mathbf{k}$ points can be specified when performing the DFT calculations, such that $E(\mathbf{k}1)$ and $E(\mathbf{k}2)$ are explicitly contained in the DFT data.

[69] To extract the effective mass tensor, the intermediate module uses three quantities: the band label (which band to compute effective mass tensor), the band structure valley location $\mathbf{k}_0(k_{ox}, k_{oy}, k_{oz})$ in the Brillouin zone, and the cutoff value k_{cut} .

[70] Then, the intermediate module performs the extraction calculations. The intermediate module parses the DFT output data and chooses all $\mathbf{k}$ points $\mathbf{k}(k_x, k_y, k_z)$ which can satisfy

$$[71] \quad \sqrt{(k_x - k_{ox})^2 + (k_y - k_{oy})^2 + (k_z - k_{oz})^2} < k_{cut}$$

[72] Then, using the band label specified, the eigenvalues $E(\mathbf{k})$, including $E(\mathbf{k}_0)$, corresponding to the chosen $\mathbf{k}$ points in the target band are read out from DFT data.

[73] Then, the effective mass tensor m_{ij} ($i, j = x, y, z$) extraction is performed based on the Taylor expansion of the $E(\mathbf{k})$ relation near $\mathbf{k}_0$

$$[74] \quad E(\mathbf{k}) = E(\mathbf{k}_0) + \frac{\hbar^2}{2} m_{ij}^{-1} (k_i - k_{oi}) (k_j - k_{oj}) + O(|\Delta \mathbf{k}|^2)$$

[75] To obtain the numerical values of the effective mass tensor, which has multiple (e.g., six) independent components due to symmetry $m_{ij} = m_{ji}$, the linear algebraic equations $\mathbf{A} \mathbf{X} = \mathbf{B}$ is formulated and solved, where $\mathbf{A}$ is an $n \times 6$ matrix; $\mathbf{B}$ and $\mathbf{X}$ are 6×1 matrices; and n is the number of chosen $\mathbf{k}$ points. The solution of the $\mathbf{A} \mathbf{X} = \mathbf{B}$ is based on least square solution algorithm. The effective mass tensor matrix inversion is calculated.

[76] Therefore, appropriate DFT data such as the k_{cut} value are provided to intermediate module, such that there are sufficient (e.g., 6) linearly independent equations in $\mathbf{A} \mathbf{X} = \mathbf{B}$. Otherwise, the extraction can crash or give unphysical results.

[77] Also, k_{cut} can be set as an appropriate value. If small enough, all chosen (k_x, k_y, k_z) points are in the parabolic region near (k_{ox}, k_{oy}, k_{oz}) . Secondly, if k_{cut} is large enough, the difference $E(\mathbf{k}) - E(\mathbf{k}_0)$ is large enough to ensure extraction accuracy. Although an example $k_{cut} = 0.01$ can be used, k_{cut} is a material-specific parameter to be customized.

[78] To extract the effective mass scalar, the intermediate module uses three inputs: the band label (which band to compute effective mass scalar), the band structure valley location $\mathbf{k}_0(k_{ox}, k_{oy}, k_{oz})$ in the Brillouin zone, and another $\mathbf{k}$ point $\mathbf{k}p(k_{px}, k_{py}, k_{pz})$ near $\mathbf{k}_0$.

[79] The intermediate module parses the DFT data and find out the eigenvalues $E(\mathbf{k}p)$ and $E(\mathbf{k}_0)$ in the specified band. Then, using the Taylor expansion of the $E(\mathbf{k})$ relation at $\mathbf{k}_0$

$$[80] \quad E(k_p) = E(k_o) + \frac{\hbar^2}{2m} |k_p - k_o|^2 + O(|\Delta k|^2)$$

[81] the effective mass scalar m along the direction $k_p - k_o$ can be computed.

[82] Similar to the effective mass tensor extraction, the $|k_p - k_o|$ can be small enough such k_p that is still in the parabolic region of the valley. And it can be large enough such that the difference $E(k_p) - E(k_o)$ is large enough to ensure extraction accuracy.

[83] To extract the non-parabolicity parameter α , the intermediate module relies on multiple inputs: the band label (which band to compute effective mass scalar), the band structure valley location $k_o(k_{ox}, k_{oy}, k_{oz})$ in the Brillouin zone, a k point $k_p(k_{px}, k_{py}, k_{pz})$ in the parabolic region near k_o , and a k point $k_n(k_{nx}, k_{ny}, k_{nz})$ in the non-parabolic region far away from k_o .

[84] To extract non-parabolicity, in the first step, the effective mass scalar m along the direction $k_p - k_o$ is computed as described. Using the value of m ,

$$[85] \quad E_p(k_n) = E(k_o) + \frac{\hbar^2}{2m} |k_n - k_o|^2$$

is computed to obtain the eigenvalue at k_n if the band structure were parabolic. Actually the band structure is non-parabolic at k_n , so in the second step, using

$$[86] \quad E(k_n) = \frac{1}{2\alpha} [-1 + \sqrt{1 + 4[E_p(k_n) - E(k_o)]\alpha}] + E(k_o)$$

[87] the non-parabolicity parameter α is computed.

[88] The choice of k_p follows the principles introduced in effective mass scalar extraction. The choice k_n can be pre-specified or up to the user, provided that k_n is in the non-parabolic region.

[89] To extract the k.p model parameters, the user or intermediate module provides two inputs to the intermediate module: the location of the band structure valley $k_o(k_{ox}, k_{oy}, k_{oz})$ in the Brillouin zone, and the cutoff value k_{cut} .

[90] The intermediate module extracts the k.p model parameters in multiple steps. In the first step, it parses the DFT data and select the k points $k(k_x, k_y, k_z)$ whose distances to k_o are smaller than k_{cut} . Also, the corresponding eigenvalues $E_{DFT}(k)$ in the highest six valence bands are read from DFT data.

[91] In the second step, the initial guess of the model parameters ($\gamma_1, \gamma_2, \gamma_3$, and Δ_{so}) is used to generate the 6 x 6 Hamiltonian matrix

$$[92] \quad H = \begin{bmatrix} P+Q & -S & R & 0 & -S/\sqrt{2} & \sqrt{2}R \\ & P-Q & 0 & R & -\sqrt{2}Q & \sqrt{3/2}S \\ & & P-Q & S & \sqrt{3/2}S^* & \sqrt{2}Q \\ & & & P+Q & -\sqrt{2}R^* & -S^*/\sqrt{2} \\ & & & & P+\Delta_{so} & 0 \\ & & & & & P+\Delta_{so} \end{bmatrix}$$

[93] where

$$[94] \quad \begin{cases} P = \frac{\hbar^2}{2m_e} \gamma_1 (k_x^2 + k_y^2 + k_z^2) \\ Q = \frac{\hbar^2}{2m_e} \gamma_2 (k_x^2 + k_y^2 + k_z^2) \\ R = \frac{\hbar^2}{2m_e} \sqrt{3} [-\gamma_2 (k_x^2 - k_y^2) + 2i\gamma_3 k_x k_y] \\ S = \frac{\hbar^2}{2m_e} 2\sqrt{3} (k_x - ik_y) k_z \end{cases}$$

[95] according to the k.p theory. Here, m_e is the electron mass; $\gamma_1, \gamma_2, \gamma_3$ and are the Luttinger parameters; and Δ_{so} is the spin-orbit split-off energy. Then, the eigenvalue problem

$$[96] \quad H(k)X(k) = EBKP(k)X(k)$$

[97] is solved, using the ZGEEV subroutine as implemented in LAPACK package, to obtain the eigenvalues (e.g., 6) for each selected k point.

[98] In the third step, the difference between the EDFT(k) and the EBKP(k) is evaluated using the cost function, which is defined as

$$[99] \quad c = \sqrt{\frac{1}{N_k} \sum_k |E_{BKP}(k) - E_{DFT}(k)|^2}$$

[100] where N_k is the number of selected k points in the extraction. The cost function describes how much the guessed model parameters ($\gamma_1, \gamma_2, \gamma_3$ and Δ_{so}) deviate from the DFT data.

[101] In the fourth step, the cost function is iteratively reduced, by using the simplex optimization algorithm to optimize the model parameters ($\gamma_1, \gamma_2, \gamma_3$ and Δ_{so}). In the iterative optimization of the model parameters, the second and third steps are repeated and the model parameters are updated until convergence.

[102] Various embodiments are directed to a different number of bands in the k.p model.

[103] In the following, the bulk silicon six-band k.p model parameter extraction is used as an example, to show how the simplex optimization algorithm works in the extraction.

[104] In this example, the calculation is done in the Nd -dimensional ($Nd = 4$) parameter space, since there are four parameters to optimize. To start with, (0, 0, 0, 0) is used as the initial guess of the model parameters ($\gamma_1, \gamma_2, \gamma_3$ and Δ_{so}). A simplex with $Nd + 1$ vertices $X_i(x_{i1}^{(0)}, x_{i2}^{(0)}, x_{i3}^{(0)}, x_{i4}^{(0)})$ is formed surrounding the initial guess, where the superscript means the iteration step number and $i = 1, 2, \dots, Nd + 1$ is the vertex label. Then, the second and third steps are executed to obtain the cost function values (ci) for all vertices. After sorting, the vertex (Xm) with the largest cost function value (cm) will be updated.

[105] To update Xm , Xm is reflected with respect to the geometric average of all other vertices, to obtain the reflected image Xr . Then, the cost function value (cr) of Xr is evaluated using the

second and third steps. According to the value of c_r , the vertex is updated differently in the following three different situations.

[106] If $\min_{i \neq m} \{c_i\} \leq c_r \leq \max_{i \neq m} \{c_i\}$,

[107] Then the vertex Xm is updated as Xr .

[108] If $c_r \leq \min_{i \neq m} \{c_i\}$,

[109] Then it means that the trial Xr is on the correct update direction in the parameter space to reduce cost function but the reflection length may be too short to be optimal. Then the reflection is extended by twice as $X'r$ whose cost function value is $c'r$. If $c'r$ is smaller than c_r , the vertex Xm is updated as $X'r$. If $c'r$ is larger than c_r , the vertex Xm is updated as Xr .

[110] If

[111] $c_r > \max_{i \neq m} \{c_i\}$

[112] it means that the trial Xr is on the wrong update direction in the parameter space. Then the search direction is reversed as $Xm - Xr$, The reversed search is repeated until $c_r < \max_{i \neq m} \{c_i\}$ is satisfied.

[113] The above calculations continues iteratively until either the difference between the maximum cost function values of the two adjacent iteration steps is smaller than a floor, e.g. 1 meV, or the maximum distance between any two vertices in the simplex is smaller than a floor, e.g. 1e-6.

[114] Results can converge in a stepwise way to the experimentally measured values, with a mismatch one the other of, for example, several percent in the final converged results, due to the fact that DFT results are approximations of physical reality.

[115] **Figure 3** is an example block diagram of simulation tools at multiple simulation scales coupled together via an extraction tool using output of a higher scale simulation tool as input for an ab initio simulation tool. To take advantage of the accuracy of more granular simulation, and the lowered computation intensity of less granular simulation, parameters are passed between multiple levels of simulation. Such embodiments provide data from simpler/large scale tools to the more sophisticated/small scale tools to accurately evaluate effects on the microscopic scale.

[116] Although the example discusses NEGF as the larger scale simulation tool, other embodiments are directed to other larger scale simulation tools other than NEGF. Although NEGF and DFT are discussed, other modules can be substituted.

[117] At 36, results are produced from an intermediate scale IC simulation tool such as NEGF. At 34, results are processed into lower simulation scale input. Processing can be as minimal as passing a parameter, or performing multiple operations to extract data. At 32, the simulation is

continued using the results from the intermediate scale tool as input of an ab initio simulation tool such as DFT.

[118] The non-equilibrium Green's function (NEGF) is an important algorithm in the TCAD tools and methodologies we propose here. The NEGF module can be used to simulate the transport properties of the devices, e.g. current, potential distribution, transmission coefficient, etc. Therefore, it is a vital part of the proposed TCAD tools.

[119] The NEGF simulator is based on iterative self-consistent solution of the Green's functions, the self-energies, and the potential profile. In the NEGF simulator, the calculation starts from the Hamiltonian matrix H , which can be obtained from either ab initio density functional theory (DFT) simulations, or tight binding parameterization, or effective mass approximation, or other techniques.

[120] The steps are explained in more details below:

[121] 1. The contact self-energies are used to represent the electrodes/contacts (like source and drain) that are linked to the transport channel. It is evaluated iteratively at energy points of interests, by using exponentially converging contact self-energy Green's function algorithms.

[122] 2. To start with, the potential profile is used as an initial guess of the final converged potential profile. The typical choice is a linear drop from drain to source. Of course, other initial guess shapes are allowed. The principle of choosing the initial potential profile is that it should be as close to the final converged potential profile as possible, in order to reduce the iteration number and computational cost to achieve convergence.

[123] 3. The retarded Green's function G^r , the election Green's function G^n , and the hole Green's function G^p are evaluated at all energy points of interests, by using numerical algorithms like (but not limited to) matrix factorization, recursive Green's function algorithm, and the Fast inverse using nested dissection, etc. The proposed approach uses efficient algorithms to accelerate the calculation by taking advantage of the special matrix structure of the Hamiltonian.

[124] 4. The in-scattering and out-scattering self-energies are computed by using the Green's functions, which are computed in the previous step. These scattering self-energies are used to represent the various scattering mechanism during transport.

[125] 5. The convergence check is performed to determine whether or not the calculation of the scattering has converged. If no, more self-consistent calculation loops will be performed. If yes, the calculation continues to the next step.

[126] 6. Poisson's equation is calculated using the algorithms like, but not limited to, direct matrix computation algorithms, iterative Krylov subspace matrix computation algorithms,

fast Fourier transform, and fast multipole method, domain decomposition method, etc. The purpose is to obtain the updated potential profile from the known charge distribution.

[127] 7. The convergence check is performed to determine whether or not the potential profile calculation has converged. If no, more iteration loops will be performed. If yes, the calculation continues to the post-process module.

[128] 8. The post-process module calculates the physical quantities of interest, which includes (but not limited to), current density, current-voltage curve, transmission coefficient, density of states, potential profile, charge distribution, etc.

[129] **Figure 4** is an example block diagram of simulation tools at multiple simulation scales coupled together via intermediate tools to iterate between an ab initio simulation tool and a higher scale simulation tool. The block diagram of 42, 44, 46, 48, 50 and 52 essentially combines Figures 2 and 3. In addition to one-way hierarchical modeling flow from more sophisticated/small scale tools down to the simpler/large scale tools, one or several iterations can be performed by going in both directions along that hierarchy. Such feedback and iterations can be applied across two or more levels of the model, potentially spanning the range from circuit/system level to the first principles level. To take advantage of the accuracy of more granular simulation, and the lowered computation intensity of less granular simulation, parameters are passed between multiple levels of simulation.

[130] An iterative and self-adaptive scheme (mentioned as the “scheme” hereafter) of the parameterization process of the phenomenological simulation modules and methodologies, is used to simulate the fabrication and operation of integrated circuit devices. The scheme combines the strengths of the ab initio simulation modules and methodologies, which are comparatively more accurate but computationally expensive, and the merits of the phenomenological simulation modules and methodologies, which are relatively less accurate but computationally inexpensive, to offer a set of iterative, self-adaptive, and automated simulation and design tools and methodologies.

[131] The “iterative parameterization” mentioned above covers multiple different categories of iterative processes. Iterative process (i) refers to the iterative parameter extraction between/among the modules mentioned in the above two points. Iterative process (ii) refers to the iterative parameter extraction between the “top-down approach”, which means that the physical quantities to be calculated are expressed as phenomenological equations with the phenomenological parameters extracted from the ab initio simulations, and the “bottom-up approach”, which means that the calculation starting point is the ab initio equations and the

computational expenses are alleviated by approximating some computationally expensive components in the equations.

[132] Although NEGF and DFT are discussed, other modules can be substituted.

[133] In an iterative looped feedback, the NEGF module and the DFT module are combined together to provide enhanced simulation capabilities, in two directions. The NEGF module outputs non-equilibrium transport properties and electronic structure information, which are used in the DFT module to account for the non-equilibrium conditions. The DFT module outputs equilibrium electronic structure information, which is used in the NEGF module to simulate transport properties. These two directions iterate as a bi-directional feedback loop, to offer unprecedented TCAD simulation capabilities. For example, to simulate the source-channel-drain device structure, the DFT module is first run to obtain the Hamiltonian and overlap matrices. Then these matrices are input into the NEGF module to simulate the transport properties. This generates the non-equilibrium electron transport properties like potential distribution, which in turn is fed back into the DFT module to calculate physical properties like forces on atoms and the internal stress under non-equilibrium conditions, etc.

[134] Different types of an iterative process are as follows: (1) simulation functionality feedback; (2) calculation efficiency and precision feedback; and (3) the iterative looped feedback. Although NEGF and DFT are discussed, other modules can be substituted. They are introduced in the three following sections below:

[135] In the simulation functionality feedback, The NEGF module decides what should be calculated and sends instructions to the lower-level DFT module to get the computations done. For example, in the ab initio electron transport simulation, the NEGF module uses ab initio Hamiltonian matrix and overlap matrix as input. So the NEGF module sends instructions to the DFT module, requesting related calculations. In the effective mass Hamiltonian electron transport calculations, the NEGF module can request the DFT module to perform band structure calculations and extract the effective mass.

[136] In the efficiency and precision feedback, the NEGF module runs electron transport simulations and compares the results against benchmark results. The comparison is used to feedback into the DFT module, to strike a good balance between the computational efficiency and precision. For instance, the calculation is largely determined by the size of the matrix that represents the system of interests. The larger the matrix size, the more precise the results are. But the calculation will be less efficient. In contrast, if the system is represented by using smaller matrices, the calculation will be less precise, but the calculation will be more efficient. By comparing against benchmark results and/or experimental calibrations of example systems, the

NEGF module can work jointly with the lower-level DFT module, to determine the optimum matrix size to represent the target system. By using this optimum matrix size, the simulation can achieve both good calculation precision and high simulation efficiency.

[137] The three processes introduced above are designed to be automated and self-adaptive, in order to minimize the human intervention and to maximize productivity.

[138] **Figure 5** is an example process flow of executing simulation tools automatically at multiple simulation scales, to improve inaccurate input (such as missing input).

[139] Multiple simulation scales of various simulation modules, such as in Figure 1, use output of one module as input for another module, and automatically executing another module at the same simulation scale or different simulation scale to improve inaccurate input (such as missing input).

[140] In the example of Figure 1, if a particular simulation module has missing input, the coordination module executes another module to provide the missing input. If a particular simulation scale is not specified in the execution instruction, the coordination module determines the appropriate scales to run. The proposed modules receive target simulation quantities via a user interface, and instruct the coordination module to start the simulation.

[141] At 60, a user (or automated) instruction is received to begin simulation without specifying at least one simulation scale (or at least without specifying one simulation tool). There are multiple options. In option 62, simulation is initiated automatically at a particular simulation scale (or a particular simulation tool out of many) depending on simulation input that does not specify the particular simulation scale (or depending on simulation input that does not specify the particular simulation tool). In option 64, simulation is initiated at a default simulation scale without regard to other simulation input (or initiated with a default simulation module without regard to other simulation input), or with regard to an explicitly selected simulation scale (or with regard to an explicitly selected simulation module). At 66, inaccurate parameters (such as missing parameters) are identified. At 68, simulation scales (or simulation tools) are automatically switched to acquire inaccurate parameters such as missing parameters.

[142] Figure 6 shows different simulation volumes of simulation modules at different simulation scales. A 3D simulation volume is indicated. In other embodiments, the simulation volume may be 2D or 1D. Because simulator A has more granularity than simulator B, simulation volume A 70 of simulator A is smaller than simulation volume B 72 of simulator B. Because simulation volume A 70 is smaller, the border between simulation volumes A and B is also a border 70 of simulation volume A 70.

[143] Figures 7-8 shows a switch between simulation modules at different simulation scales, from a larger simulation volume to a smaller simulation volume. In Figure 7, nonzero simulated property 76 occupies a relatively small fraction of simulation volume B 72. In Figure 8, nonzero simulated property 76 occupies a relatively large fraction of simulation volume A 70. Accordingly, the simulation modules/scales switch automatically from simulation volume B 72 of simulator B to simulation volume A 70 of simulator A.

[144] Figures 9-10 shows a switch between simulation module volumes from a smaller simulation volume to a larger simulation volume. In Figure 9, nonzero simulated property 76 is not fully captured by simulation volume A 70. In Figure 10, nonzero simulated property 76 occupies fits in simulation volume B 72. Accordingly, the simulation modules/scales switch automatically from simulation volume A 70 of simulator A to simulation volume B 72 of simulator B.

[145] Examples of a simulated property include: electrostatic potential, electric field, non-equilibrium charge distribution, stress distribution and stress gradient, strain distribution and strain gradient, electron and hole generation, electron and hole recombination, atomic migration (such as hydrogen, lattice atoms, and impurities), point defect formation, extended defect formation, void formation, ionic migration, filament formation, phase change (such as between crystal and amorphous states), non-equilibrium spin distribution, electron and hole trapping at the defects, Joule heat generation, structure changes due to atomic and molecular chemical reactions, and optically generated electrons and holes. Such properties can be converted to DFT inputs.

[146] Figures 11-12 show an error at the border between simulation modules at different simulation scales, with horizontal and oblique asymptotes.

[147] In the example of Figures 9-10, a nonzero simulated property was not fully captured by simulation volume A of simulator A, so the simulation modules/scales switch automatically to simulation volume B 72 of simulator B. However, an alternative is to not switch and tolerate the nonzero error, remaining with simulation volume A of simulator A so long as the nonzero error does not exceed an error threshold.

[148] Figure 11 shows that at the border 74 between simulation volumes A and B, an error 82 exist in the nonzero simulated property 78. The nonzero simulated property 78 has failed to reach the horizontally asymptotic value 80 of the nonzero simulated property 78. In Figure 12, the nonzero simulated property 78 has failed to reach the obliquely asymptotic value 80 of the nonzero simulated property 78. So long as the error 82 does not does not exceed an error

threshold, the nonzero error may be tolerated by not switching from simulation volume A of simulator A to simulation volume B 72 of simulator B.

[149] Figures 13A, 13B, 13C and 13D illustrate various lattice configurations, as examples of arrangements in which a semiconductor property has a finite distortion range to be simulated as in Figures 6-12. The finite distortion range can depend on the direction. Distortions can be electrical, mechanical, or band structure. Examples include point defects such as vacancy defects, interstitial defects and substitutional defects, using a GaAs alloy as an example. A conventional 8-atom cell (e.g. 400), and a supercell composed of 4 conventional cells are shown in each of the figures. As used herein, a host is the material into which native defects and dopants diffuse, and a host atom is an atom in the host material without any diffusion into the host material. The host material can be an alloy.

[150] Fig. 13A illustrates a supercell 410 including host atoms Ga (Gallium) and As (Arsenic), and lattice vacancy defects (e.g. 411, 414) surrounded by adjacent host atoms Ga and As. A lattice vacancy defect refers to an atom site in a crystal lattice where a single host atom is missing. Neighbors of a point defect can be important. As shown in the example of Fig. 13A, a vacancy defect has first order neighbors and second order neighbors. For instance, vacancy defect 411 at an As site has 4 first-order neighbors that are Ga atoms (e.g. 412), and more second-order neighbors (e.g. 413). Vacancy defect 414 at a Ga site has 4 first-order neighbors that are As atoms (e.g. 415), and more second-order neighbors (e.g. 416).

[151] In a different alloy there can be more options for first-order neighbors. For instance in a SiGe alloy (not shown), a point defect such as a vacancy defect can have different first-order neighbors including 4 Si atoms, 3 Si atoms and 1 Ge atom, 2 Si atoms and 2 Ge atoms, 1 Si and 3 Ge atoms, or 4 Ge atoms.

[152] Fig. 13B illustrates an example of a dopant atom surrounded by various different numbers of host atoms adjacent thereto. Impurities or dopants can be doped into the GaAs host material. Group IV elements such as Si (silicon) can act as either donors on Ga sites or acceptors on As sites. In the example of Fig. 13B, Si is used as the dopant (e.g. 425).

[153] Fig. 13C illustrates an example of an interstitial defect atom surrounded by various different numbers of host atoms adjacent thereto, where an interstitial defect atom is present in the interstitial spaces between the host crystal lattice sites. In an alloy (e.g. GaAs), each alloying species (e.g. Ga, As) can be at a defect site between lattice sites or an interstitial location off the crystal lattice. In the example of Fig. 13C, a Ga interstitial is at a defect site (e.g. 435), while an As interstitial is at another defect site (e.g. 437). In a tetrahedrally coordinated GaAs configuration (not shown), a Ga interstitial or a As interstitial can have four equidistant first-

order neighbors. In a hexagonally coordinated GaAs configuration (not shown), a Ga interstitial or a As interstitial can have six equidistant first-order neighbors, including 3 Ga atoms and 3 As atoms.

[154] Fig. 13D illustrates an example of a substitutional defect atom surrounded by various different numbers of host atoms adjacent thereto, where a host atom is replaced by an atom of a different type than the host atoms. A substitutional defect atom can be smaller or larger than a host atom (covalent radius of Ga = 126pm, covalent radius of As = 119pm). In the example of Fig. 13D, Be (Beryllium, covalent radius = 90pm) is shown as a smaller substitutional defect atom at a defect site (e.g. 445), while Te (Tellurium, covalent radius = 135pm) is shown as a larger substitutional defect atom at another defect site (e.g. 447).

[155] For a more complex material, there can be more types of point defects. For instance, indium gallium arsenide (InGaAs) is a ternary alloy of indium, gallium and arsenic. Indium and gallium are both from the boron group (group III) of elements, and thus have similar roles in chemical bonding. InGaAs is regarded as an alloy of gallium arsenide and indium arsenide with properties intermediate between the two depending on the proportion of gallium to indium. For instance, compounds $\text{In}_{0.75}\text{Ga}_{0.25}\text{As}$, $\text{In}_{0.5}\text{Ga}_{0.5}\text{As}$, and $\text{In}_{0.25}\text{Ga}_{0.75}\text{As}$ include different proportions of gallium to indium, while InAs does not include Ga and GaAs does include In.

[156] In InGaAs, point defects can include As vacancy (i.e. missing lattice atom where As is supposed to be), In/Ga vacancy (i.e. missing lattice atom where In or Ga is supposed to be), As interstitial (i.e. extra As atom between lattice sites), In interstitial (i.e. extra In atom between lattice sites), Ga interstitial (i.e. extra Ga atom between lattice sites), As atom in the In/Ga lattice site, In atom in the As lattice site, and Ga atom in the As lattice site. For each of these point defects, there are different combinations of first-order neighbors, similar to the first-order neighbors described for the SiGe alloy.

[157] **Figure 14** is a simplified block diagram of a computer system configured to perform IC simulation to implement any of the methods and processes herein.

[158] Computer system 110 typically includes a processor subsystem 114 which communicates with a number of peripheral devices via bus subsystem 112. These peripheral devices may include a storage subsystem 124, comprising a memory subsystem 126 and a file storage subsystem 128, user interface input devices 122, user interface output devices 120, and a network interface subsystem 116. The storage subsystem 128 includes non-transitory memory storing computer programs, databases and other resources to configure the data processing systems as tool for controlling multiple simulation tools at multiple simulation scales as

described herein. The tool can include an API configured to use input parameter sets, and to perform the procedures of the tool using the input parameter sets.

[159] The input and output devices allow user interaction with computer system 110. Network interface subsystem 116 provides an interface to outside networks, including an interface to communication network 118, and is coupled via communication network 118 to corresponding interface devices in other computer systems. Communication network 118 may comprise many interconnected computer systems and communication links. These communication links may be wireline links, optical links, wireless links, or any other mechanisms for communication of information, but typically it is an IP-based communication network. While in one embodiment, communication network 118 is the Internet, in other embodiments, communication network 118 may be any suitable computer network.

[160] The physical hardware component of network interfaces are sometimes referred to as network interface cards (NICs), although they need not be in the form of cards: for instance they could be in the form of integrated circuits (ICs) and connectors fitted directly onto a motherboard, or in the form of macrocells fabricated on a single integrated circuit chip with other components of the computer system.

[161] User interface input devices 122 may include a keyboard, pointing devices such as a mouse, trackball, touchpad, or graphics tablet, a scanner, a touch screen incorporated into the display, audio input devices such as voice recognition systems, microphones, and other types of input devices. In general, use of the term "input device" is intended to include all possible types of devices and ways to input information into computer system 110 or onto computer network 118.

[162] User interface output devices 120 may include a display subsystem, a printer, a fax machine, or non visual displays such as audio output devices. The display subsystem may include a cathode ray tube (CRT), a flat panel device such as a liquid crystal display (LCD), a projection device, or some other mechanism for creating a visible image. The display subsystem may also provide non visual display such as via audio output devices. In general, use of the term "output device" is intended to include all possible types of devices and ways to output information from computer system 110 to the user or to another machine or computer system.

[163] Storage subsystem 124 stores the basic programming and data constructs that provide the functionality of certain embodiments of the present invention. For example, the various modules implementing the functionality of certain embodiments of the invention may be stored in storage subsystem 124. These software modules are generally executed by processor subsystem 114.

[164] Memory subsystem 126 typically includes a number of memories including a main random access memory (RAM) 130 for storage of instructions and data during program execution and a read only memory (ROM) 132 in which fixed instructions are stored. File storage subsystem 128 provides persistent storage for program and data files, and may include a hard disk drive, a floppy disk drive along with associated removable media, a CD ROM drive, an optical drive, or removable media cartridges. The databases and modules implementing the functionality of certain embodiments of the invention may have been provided on a computer readable medium such as one or more CD-ROMs, and may be stored by file storage subsystem 128. The host memory 126 contains, among other things, computer instructions which, when executed by the processor subsystem 114, cause the computer system to operate or perform functions as described herein. As used herein, processes and software that are said to run in or on "the host" or "the computer", execute on the processor subsystem 114 in response to computer instructions and data in the host memory subsystem 126 including any other local or remote storage for such instructions and data.

[165] Bus subsystem 112 provides a mechanism for letting the various components and subsystems of computer system 110 communicate with each other as intended. Although bus subsystem 112 is shown schematically as a single bus, alternative embodiments of the bus subsystem may use multiple busses.

[166] Computer system 110 itself can be of varying types including a personal computer, a portable computer, a workstation, a computer terminal, a network computer, a television, a mainframe, a server farm, or any other data processing system or user device. Due to the ever changing nature of computers and networks, the description of computer system 110 is intended only as a specific example for purposes of illustrating the preferred embodiments of the present invention. Many other configurations of computer system 110 are possible having more or less components than the computer system.

[167] In addition, while the present invention has been described in the context of a fully functioning data processing system, those of ordinary skill in the art will appreciate that the processes herein are capable of being distributed in the form of a computer readable medium of instructions and data and that the invention applies equally regardless of the particular type of signal bearing media actually used to carry out the distribution. As used herein, a computer readable medium is one on which information can be stored and read by a computer system. Examples include a floppy disk, a hard disk drive, a RAM, a CD, a DVD, flash memory, a USB drive, and so on. The computer readable medium may store information in coded formats that are decoded for actual use in a particular data processing system. A single computer readable

medium, as the term is used herein, may also include more than one physical item, such as a plurality of CD ROMs or a plurality of segments of RAM, or a combination of several different kinds of media. As used herein, the term does not include mere time varying signals in which the information is encoded in the way the signal varies over time.

[168] **Figure 15** illustrates EDA tools and process flow for integrated circuit design and manufacturing.

[169] Aspects of the invention can be used to support an integrated circuit design flow. At a high level, the process starts with the product idea (step 200) and is realized in an EDA (Electronic Design Automation) software design process (step 210). When the design is finalized, it can be taped-out (step 227). At some point after tape out, the fabrication process (step 250) and packaging and assembly processes (step 260) occur resulting, ultimately, in finished integrated circuit chips (result 270).

[170] The EDA software design process (step 210) is itself composed of a number of steps 212-230, shown in linear fashion for simplicity. In an actual integrated circuit design process, the particular design might have to go back through steps until certain tests are passed. Similarly, in any actual design process, these steps may occur in different orders and combinations. This description is therefore provided by way of context and general explanation rather than as a specific, or recommended, design flow for a particular integrated circuit.

[171] A brief description of the component steps of the EDA software design process (step 210) will now be provided.

[172] System design (step 212): The designers describe the functionality that they want to implement, they can perform what-if planning to refine functionality, check costs, etc. Hardware-software architecture partitioning can occur at this stage. Example EDA software products from Synopsys, Inc. that can be used at this step include Model Architect, Saber, System Studio, and DesignWare® products.

[173] Logic design and functional verification (step 214): At this stage, the VHDL or Verilog code for modules in the system is written and the design is checked for functional accuracy. More specifically, the design is checked to ensure that it produces correct outputs in response to particular input stimuli. Example EDA software products from Synopsys, Inc. that can be used at this step include VCS, VERA, DesignWare®, Magellan, Formality, ESP and LEDA products.

[174] Synthesis and design for test (step 216): Here, the VHDL/Verilog is translated to a netlist. The netlist can be optimized for the target technology. Additionally, the design and implementation of tests to permit checking of the finished chip occurs. Example EDA software products from Synopsys, Inc. that can be used at this step include Design Compiler®, Physical

Compiler, DFT Compiler, Power Compiler, FPGA Compiler, TetraMA1, and DesignWare® products.

[175] Netlist verification (step 218): At this step, the netlist is checked for compliance with timing constraints and for correspondence with the VHDL/Verilog source code. Example EDA software products from Synopsys, Inc. that can be used at this step include Formality, PrimeTime, and VCS products.

[176] Design planning (step 220): Here, an overall floor plan for the chip is constructed and analyzed for timing and top-level routing. Example EDA software products from Synopsys, Inc. that can be used at this step include Astro and Custom Designer products.

[177] Physical implementation (step 222): The placement (positioning of circuit elements) and routing (connection of the same) occurs at this step, as can selection of library cells to perform specified logic functions. Example EDA software products from Synopsys, Inc. that can be used at this step include the Astro, IC Compiler, and Custom Designer products.

[178] Analysis and extraction (step 224): At this step, the circuit function is verified at a transistor level, this in turn permits what-if refinement. Example EDA software products from Synopsys, Inc. that can be used at this step include AstroRail, PrimeRail, PrimeTime, and Star-RC1T products.

[179] Physical verification (step 226): At this step various checking functions are performed to ensure correctness for: manufacturing, electrical issues, lithographic issues, and circuitry. Example EDA software products from Synopsys, Inc. that can be used at this step include the Hercules product.

[180] Tape-out (step 227): This step provides the “tape out” data to be used (after lithographic enhancements are applied if appropriate) for production of masks for lithographic use to produce finished chips. Example EDA software products from Synopsys, Inc. that can be used at this step include the IC Compiler and Custom Designer families of products.

[181] Resolution enhancement (step 228): This step involves geometric manipulations of the layout to improve manufacturability of the design. Example EDA software products from Synopsys, Inc. that can be used at this step include Proteus, ProteusAF, and PSMGen products.

[182] Mask data preparation (step 230): This step provides mask-making-ready “tape-out” data for production of masks for lithographic use to produce finished chips. Example EDA software products from Synopsys, Inc. that can be used at this step include the CATS(R) family of products.

[183] Parallel flow. The integrated circuit manufacturing flow includes a parallel flow, as follows:

[184] (1) Develop individual process steps for manufacturing the integrated circuit. This can be modeled with EDA tools such as the Synopsys tools “Sentaurus Process”, “Sentaurus Topography”, and “Sentaurus Lithography”. The input information here is the process conditions like temperature, reactor ambient, implant energy, etc. The output information is the change in geometry or doping profiles or stress distribution.

[185] (2) Integrate the individual process steps into the complete process flow. This can be modeled with EDA tools such as the Synopsys tool “Sentaurus Process”. The input information here is the collection of the process steps in the appropriate sequence. The output is the geometry, the doping profiles, and the stress distribution for the transistors and the space in between the transistors.

[186] (3) Analyze performance of the transistor manufactured with this process flow. This can be done with EDA tools such as the Synopsys tool “Sentaurus Device”. The input information here is the output of step (3) and the biases applied to transistor terminals. The output information is the currents and capacitances for each bias combination. For silicon based processes or structures, much of the information about the materials needed for simulation of electrical behavior using these tools is well known. For other materials, it may be necessary to generate or provide parameters like lattice structure, diffusivity and concentration of defects, and the like in order to support device and process scale simulations. An EDA tool for generating parameters like this is described herein.

[187] (4) If necessary, modify the process steps and the process flow to achieve the desired transistor performance. This can be done iteratively by using tools such as the Synopsys tools mentioned above.

[188] Once the process flow is ready, it can be used for manufacturing multiple circuit designs coming from different fabless companies. The EDA flow 212-230 will be used by such fabless companies. The parallel flow described here is used at a foundry to develop a process flow that can be used to manufacture designs coming from their fabless customers. A combination of the process flow and the masks 230 are used to manufacture any particular circuit. If the integrated circuit is manufactured at an IDM (integrated device manufacturer) company instead of the combination of a fables company and a foundry, then both parallel flows described above are done at the same IDM company.

[189] There is also a bridge between these tools and the 212-230 EDA tools. The bridge can be an EDA tool a Synopsys tool “Seismos” that applies compact proximity models for particular circuit design and layout to obtain netlist with instance parameters for each individual transistor in the circuit as a function of its neighborhood and stress, including material conversion stress.

[190] Incorporated by reference herein are the following documents, which provide additional information about terms and components referenced herein:

[191] Mathieu Luisier, “Quantum Transport for Engineers Lecture 4: Wave Function (WF) formalism and electrostatics”, Integrated Systems Laboratory, ETH Zurich (2012).

[192] Kyoung-Youm Kim and Byoung-ho Lee, “Quantum transport modeling in anisotropic semiconductors using Wigner function formulation”, Proceedings Conference on Optoelectronic and Microelectronic Materials and Devices, COMMAD 2000. (2000)

[193] Daniel Arovas, “Lecture Notes on Condensed Matter Physics, Chapter 1 Boltzmann Transport”, Department of Physics, University of California, San Diego (2010).

[194] R. Grau-Crespo, “Electronic structure and magnetic coupling in FeSbO₄: A DFT study using hybrid functionals and GGA+U methods”, PHYSICAL REVIEW B 73, (2006).

[195] Michel Côté, “Introduction to DFT+U”, International Summer School on Numerical Methods for Correlated Systems in Condensed Matter, Université de Montréal.

[196] A. Muramatsu, “Quantum Monte Carlo for lattice fermions”, in: M.P. Nightingale, C.J. Umriga (Eds.), Proceedings of the NATO Advanced Study Institute on Quantum Monte Carlo Methods in Physics and Chemistry, Kluwer Academic Publishers, 1999.

[197] E.K.U. Gross and N.T. Maitra, “Introduction to TDDFT”, Chapter in Fundamentals of Time-Dependent Density Functional Theory, Springer-Verlag 2012.

[198] M.A.L. Marques, E.K.U. Gross, Time-dependent density functional theory, in: C. Fiolhais, F. Nogueira, M.A.L. Marques (Eds.), A Primer in Density Functional Theory, Springer Lecture Notes in Physics, vol. 620, Springer 2003, pp. 144–184.

[199] D.A. Ryndyk, “Tight-binding model”, Lectures 2006-2007, Dresden University of Technology.

[200] R.E. Bank, “Numerical Methods for Semiconductor Device Simulation”, IEEE Transactions on Electron Devices, Vol. ED-30, No. 9, 1983.

[201] J.F. Lee, “Time-Domain Finite-Element Methods”, IEEE Transactions on Antenna and Propagation, Vol. 45, NO. 3, 1997.

[202] R. Eymard, “Finite Volume Methods”, course at the University of Wroclaw, 2008.

[203] X.L. Chen, “An advanced 3D boundary element method for characterizations of composite materials”, Engineering Analysis with Boundary Elements 29, 513–523 (2005).

[204] D. Marx, “Ab initio molecular dynamics: Theory and Implementation”, Modern Methods and Algorithms of Quantum Chemistry, J. Grotendorst (Ed.), John von Neumann Institute for Computing, Julich, “NIC Series, Vol. 1, 2000.

[205] BURKE, K., et al., "The ABC of DFT" (2007), which among other things describes DFT.

[206] KRESSE, G., et. al., VASP the Guide (September 9, 2013).

CLAIMS

1 1. An EDA tool comprising:
2 a data processor;
3 storage configured to provide computer program instructions to the processor, including:
4 an intermediate module interacting with a plurality of simulation modules to
5 perform an EDA simulation at a plurality of different simulation scales, the plurality of
6 simulation modules including:
7 a first set of one or more ab initio simulation modules at a first simulation
8 scale; and
9 a second set of one or more non-ab initio simulation modules at a second
10 simulation scale larger than the first simulation scale, wherein
11 the intermediate module processing outputs of the first set of one or more simulation
12 tools to generate a set of semiconductor material quantities received and processed by the second
13 set of one or more simulation modules including any of: effective mass tensor, non-parabolicity,
14 and N-band k.p model parameters where N is an integer.

1 2. The EDA tool of claim 1, wherein, the data processor is responsive to one or more of the
2 input parameters to at least one of the second set of one or more non-ab initio simulation
3 modules having an inaccuracy, such that the procedures automatically execute at least one of the
4 first set of one or more simulation modules to provide results as said one or more of the input
5 parameters to said at least one of the second set of one or more simulation modules, thereby
6 lessening the inaccuracy.

1 3. The EDA tool of claim 1, wherein the first set of one or more ab initio simulation
2 modules includes a density functional theory module, and said second set of one or more non-ab
3 initio simulation modules excludes the density functional theory module.

1 4. The EDA tool of claim 1, wherein, the set of semiconductor material quantities includes
2 the effective mass tensor.

1 5. The EDA tool of claim 1, wherein, the set of semiconductor material quantities includes
2 the non-parabolicity.

1 6. The EDA tool of claim 1, wherein, the set of semiconductor material quantities includes
2 the N-band k.p model parameters.

1 7. A computer-implemented method comprising:
2 causing a plurality of simulation modules to perform an EDA simulation at a plurality of
3 different simulation scales, the plurality of simulation modules including:
4 a first set of one or more ab initio simulation modules at a first simulation scale;
5 and
6 a second set of one of more non-ab initio simulation modules at a second
7 simulation scale larger than the first simulation scale; and
8 processing outputs of the first set of one or more simulation tools to generate a set of
9 semiconductor material quantities received and processed by the second set of one or more
10 simulation modules including any of: effective mass tensor, non-parabolicity, and N-band k.p
11 model parameters where N is an integer.

1 8. The computer-implemented method of claim 7, wherein, the data processor is responsive
2 to one or more of the input parameters to at least one of the second set of one of more non-ab
3 initio simulation modules having an inaccuracy, such that the procedures automatically execute
4 at least one of the first set of one or more simulation modules to provide results as said one or
5 more of the input parameters to said at least one of the second set of one of more simulation
6 modules, thereby lessening the inaccuracy.

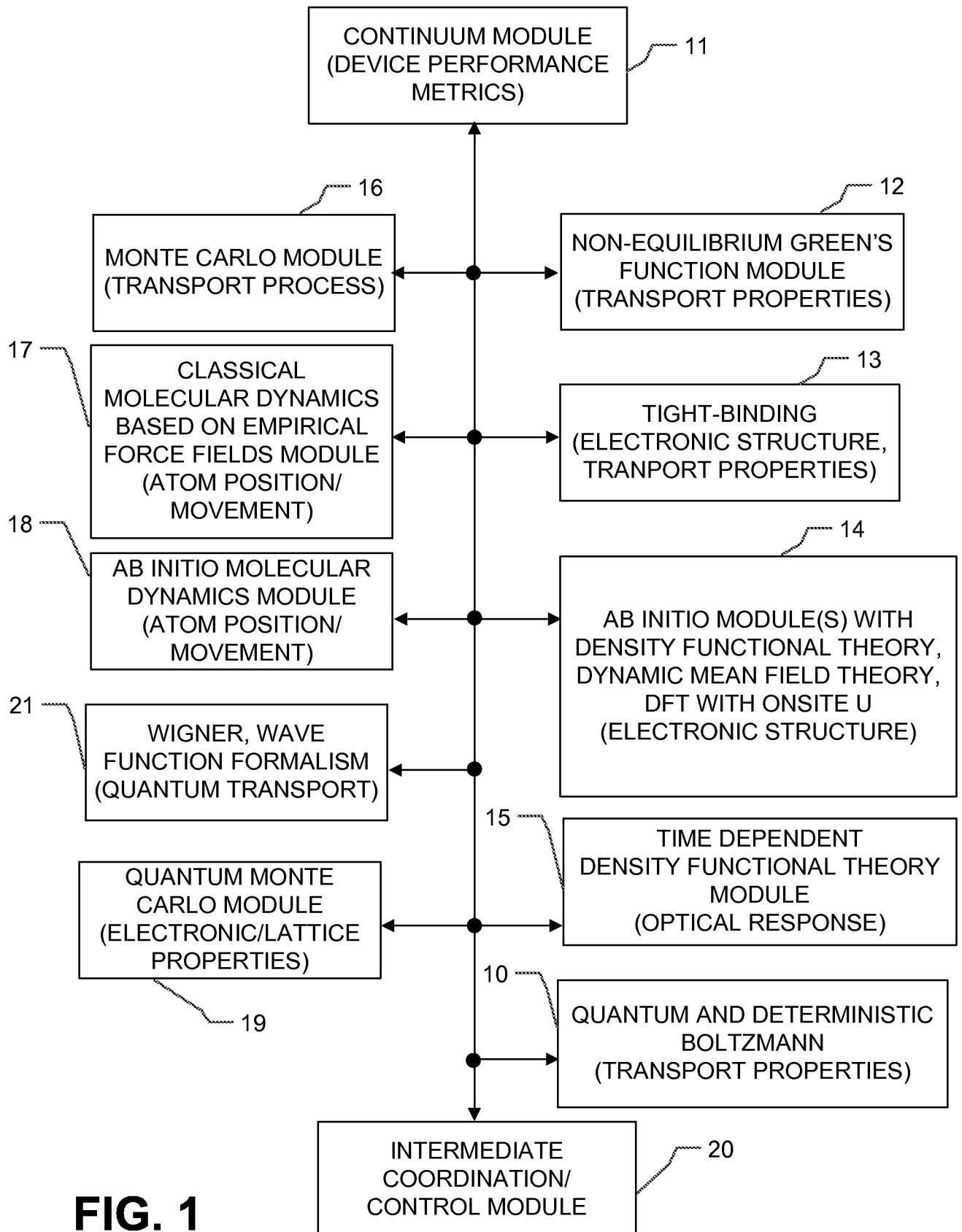
1 9. The computer-implemented method of claim 7, wherein the first set of one or more ab
2 initio simulation modules includes a density functional theory module, and said second set of one
3 or more non-ab initio simulation modules excludes the density functional theory module.

1 10. The computer-implemented method of claim 7, wherein, the set of semiconductor
2 material quantities includes the effective mass tensor.

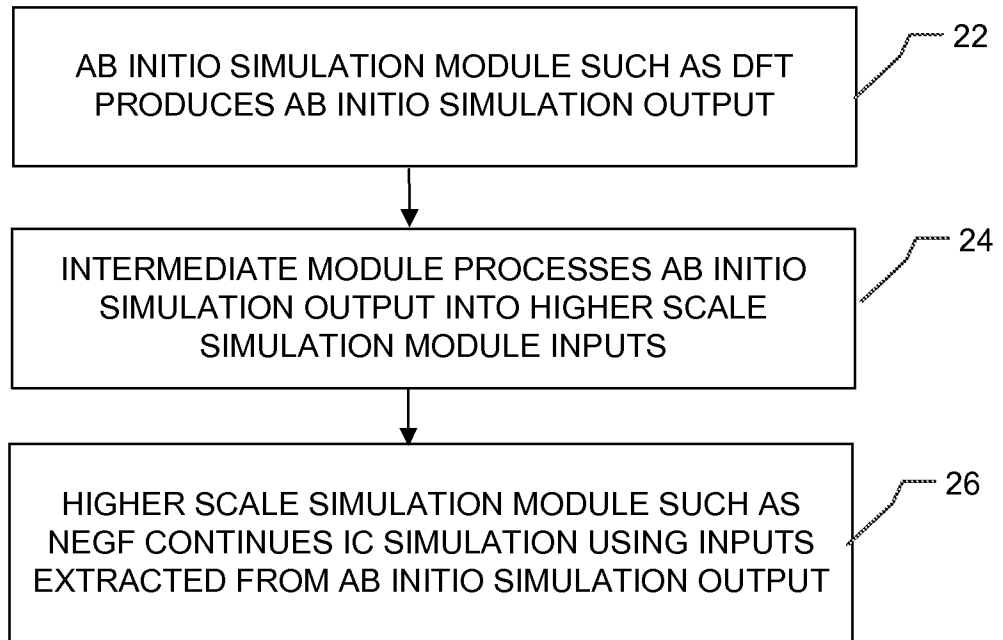
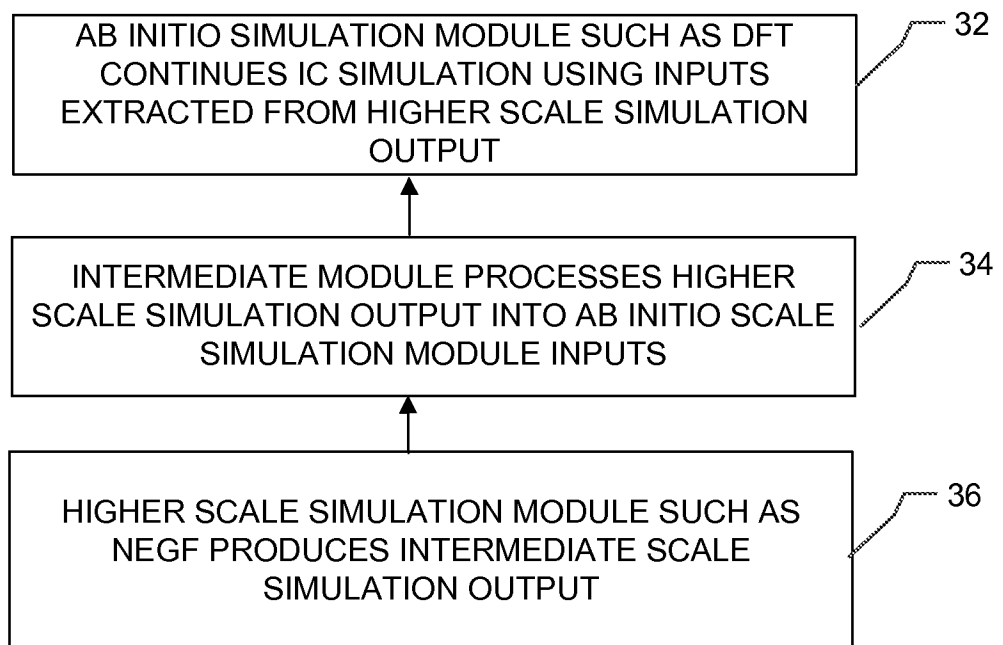
1 11. The computer-implemented method of claim 7, wherein, the set of semiconductor
2 material quantities includes the non-parabolicity.

1 12. The computer-implemented method of claim 7, wherein, the set of semiconductor
2 material quantities includes the N-band k.p model parameters.

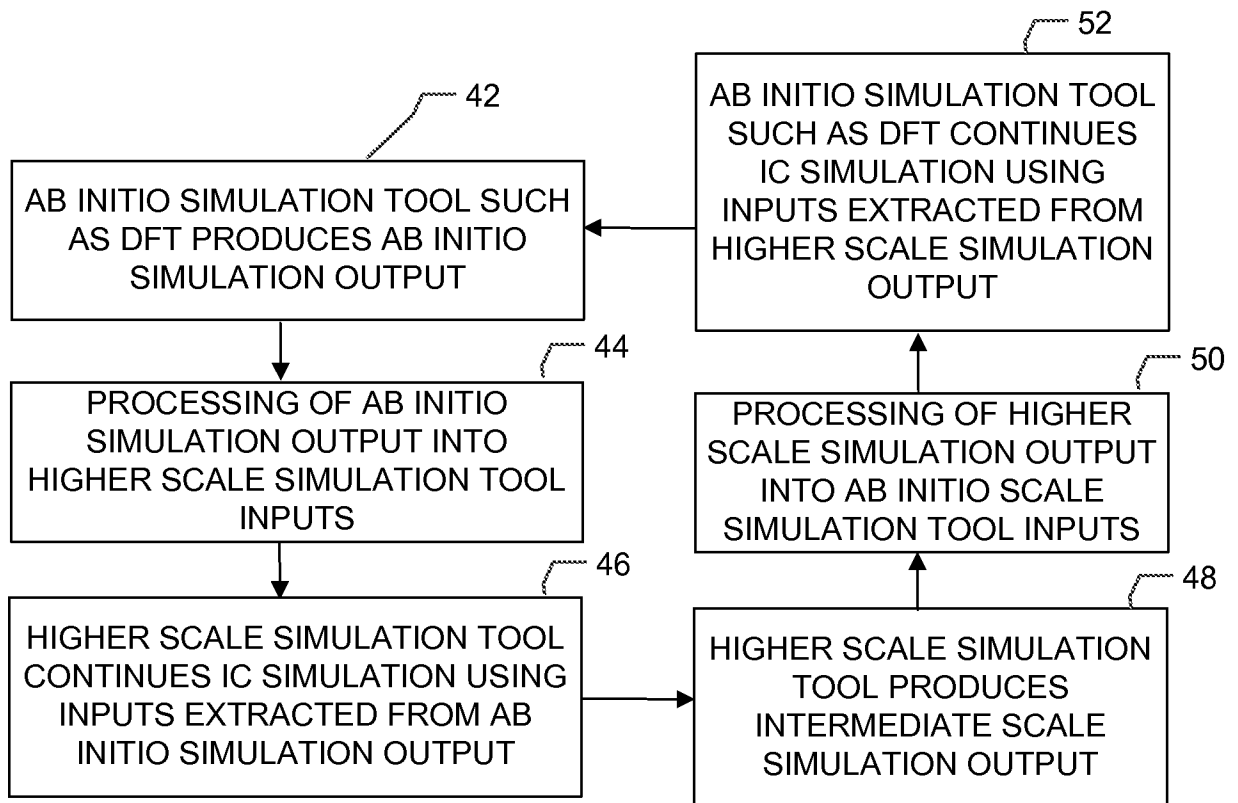
1 / 11



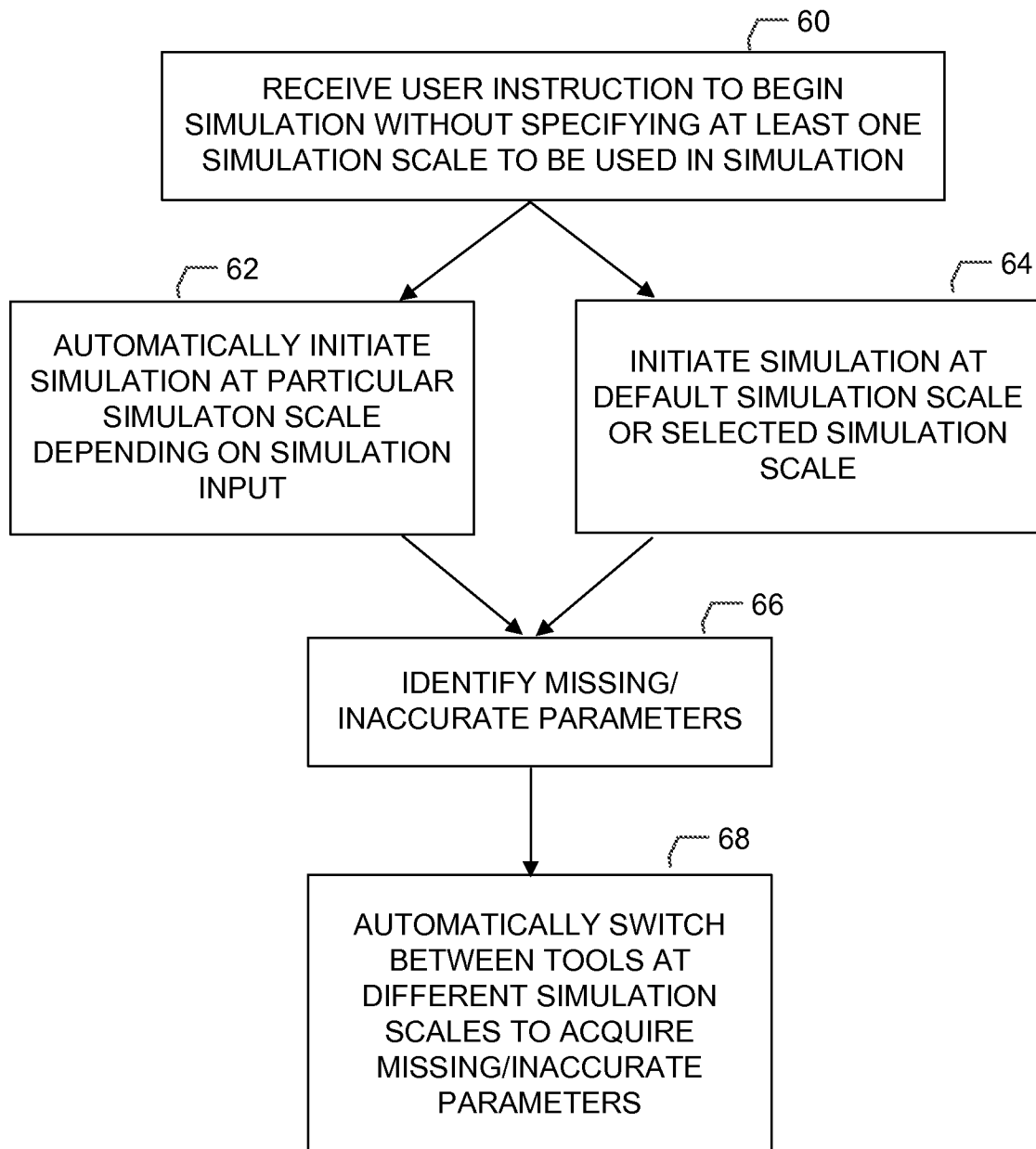
2 / 11

**FIG. 2****FIG. 3**

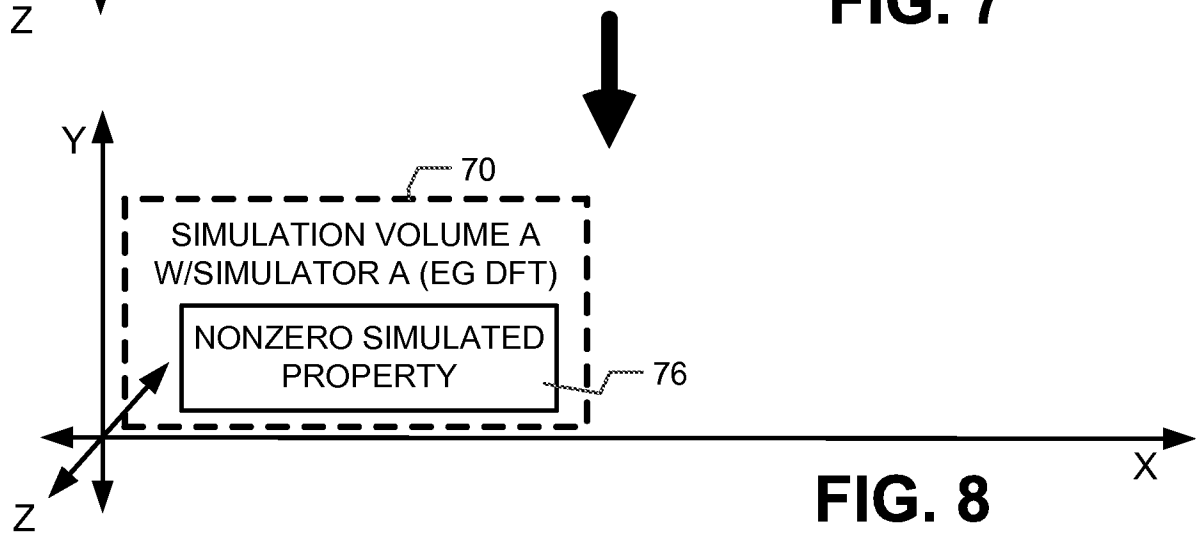
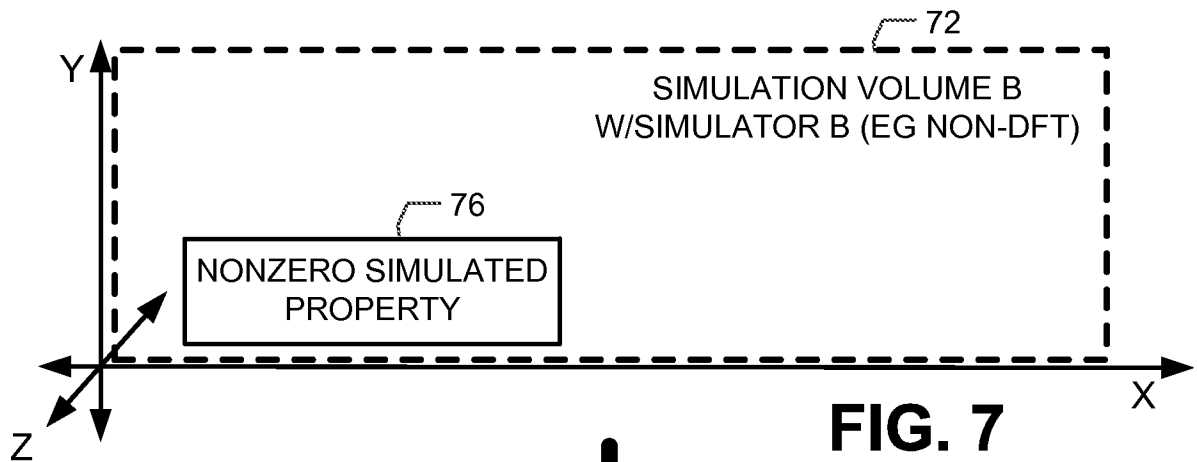
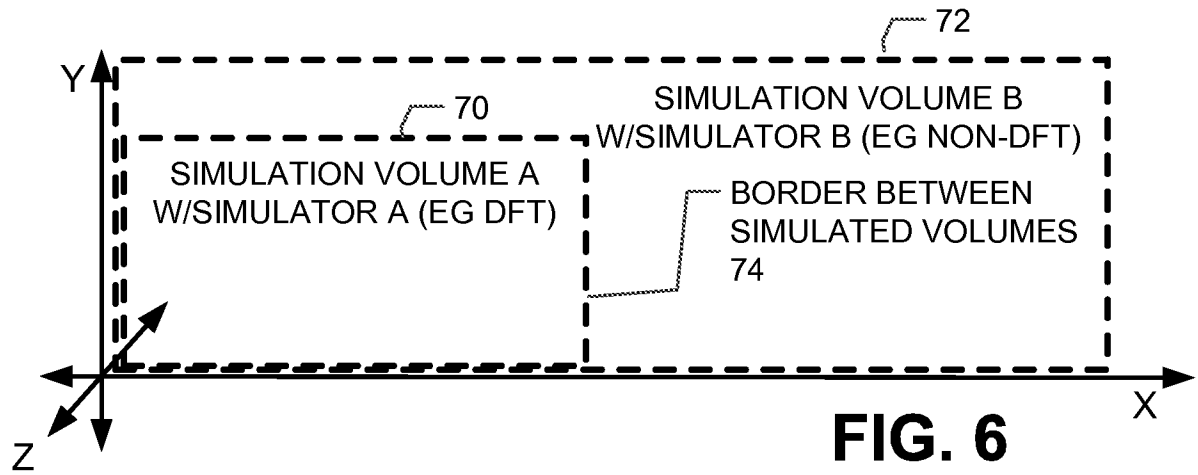
3 / 11

**FIG. 4**

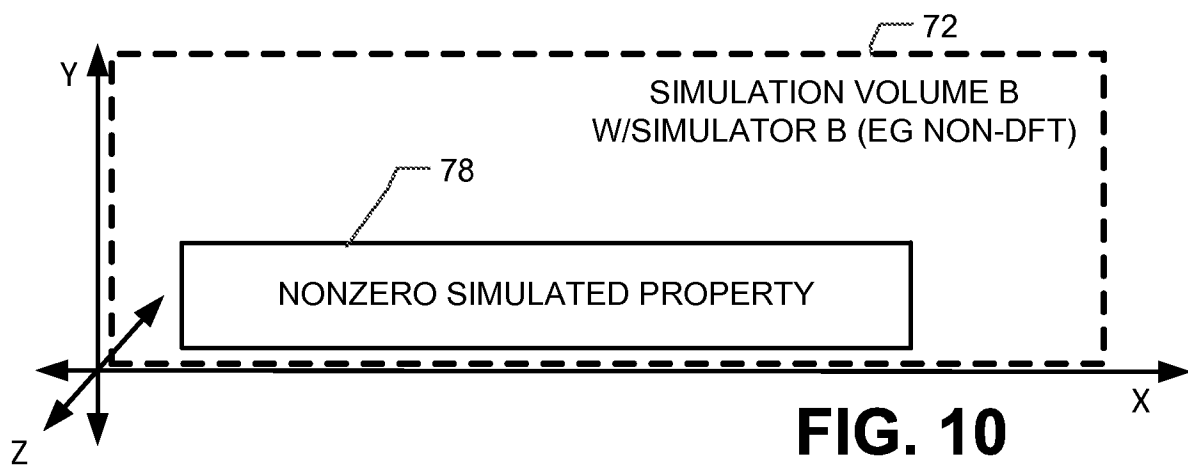
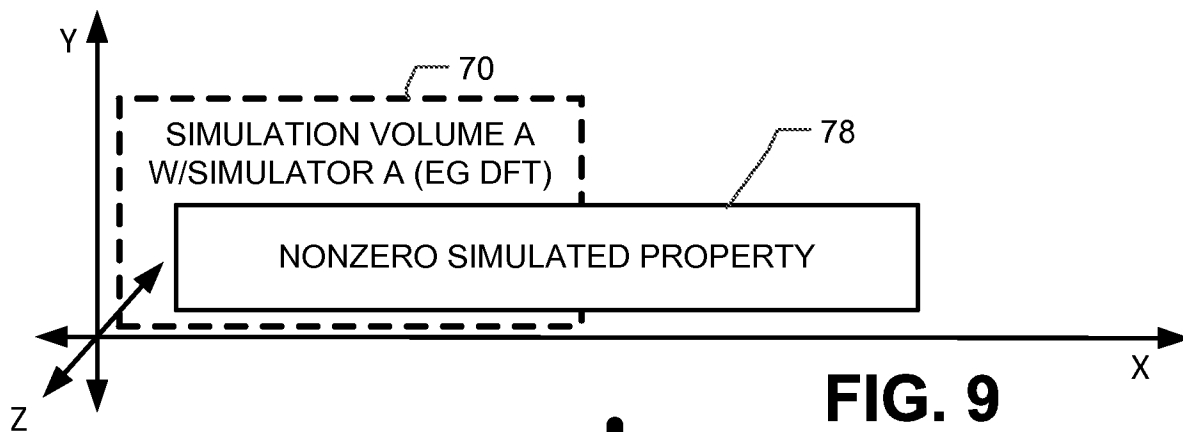
4 / 11

**FIG. 5**

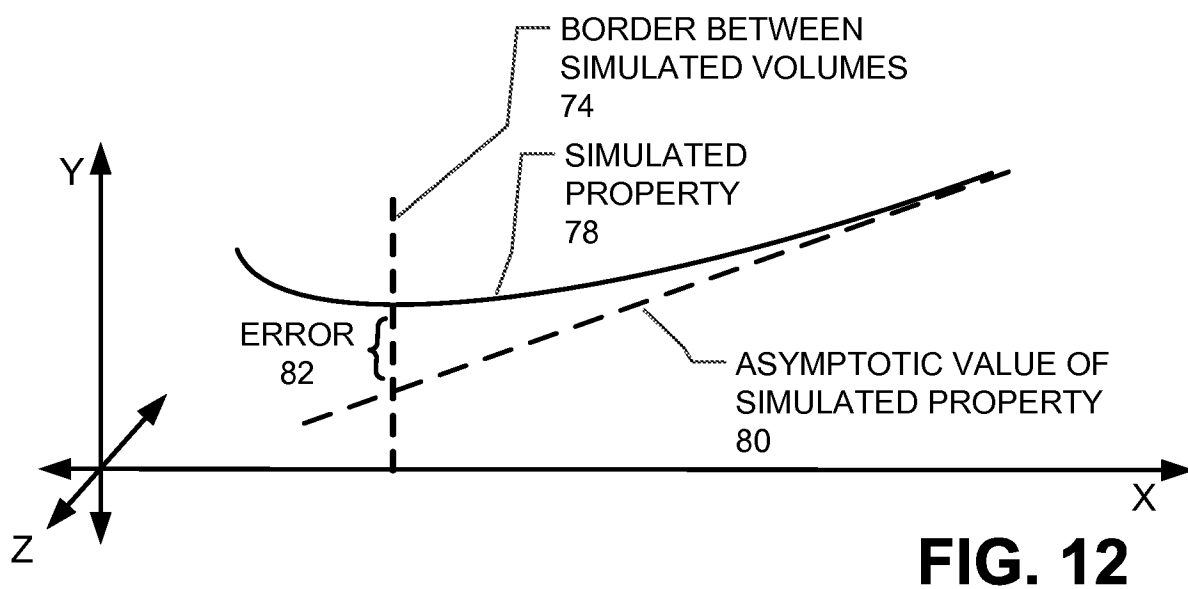
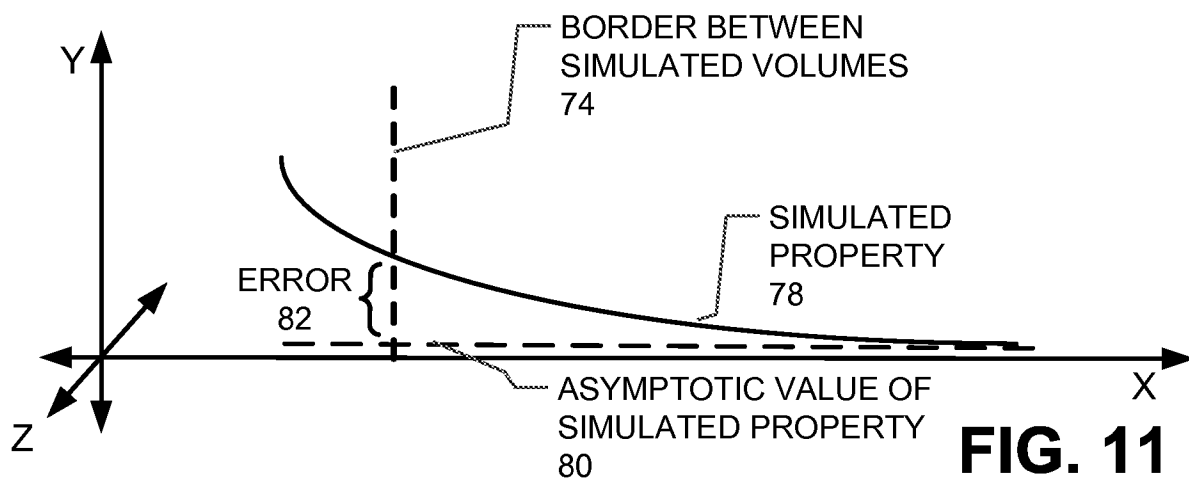
5 / 11



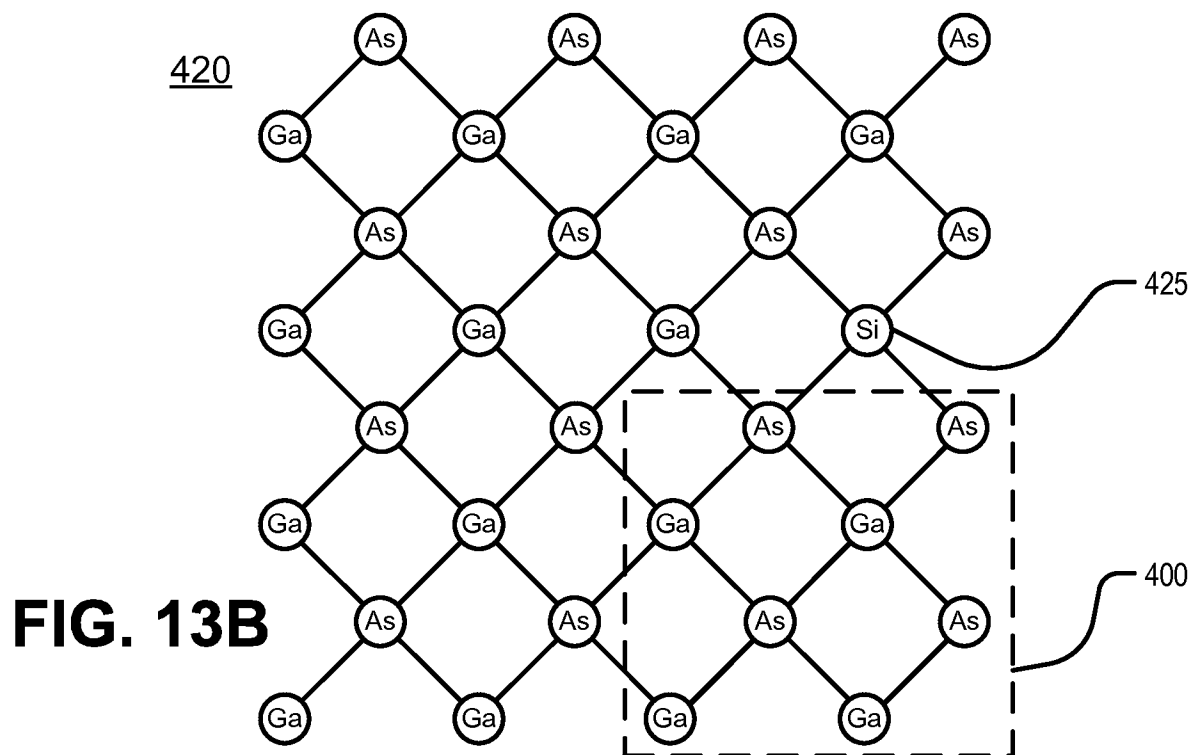
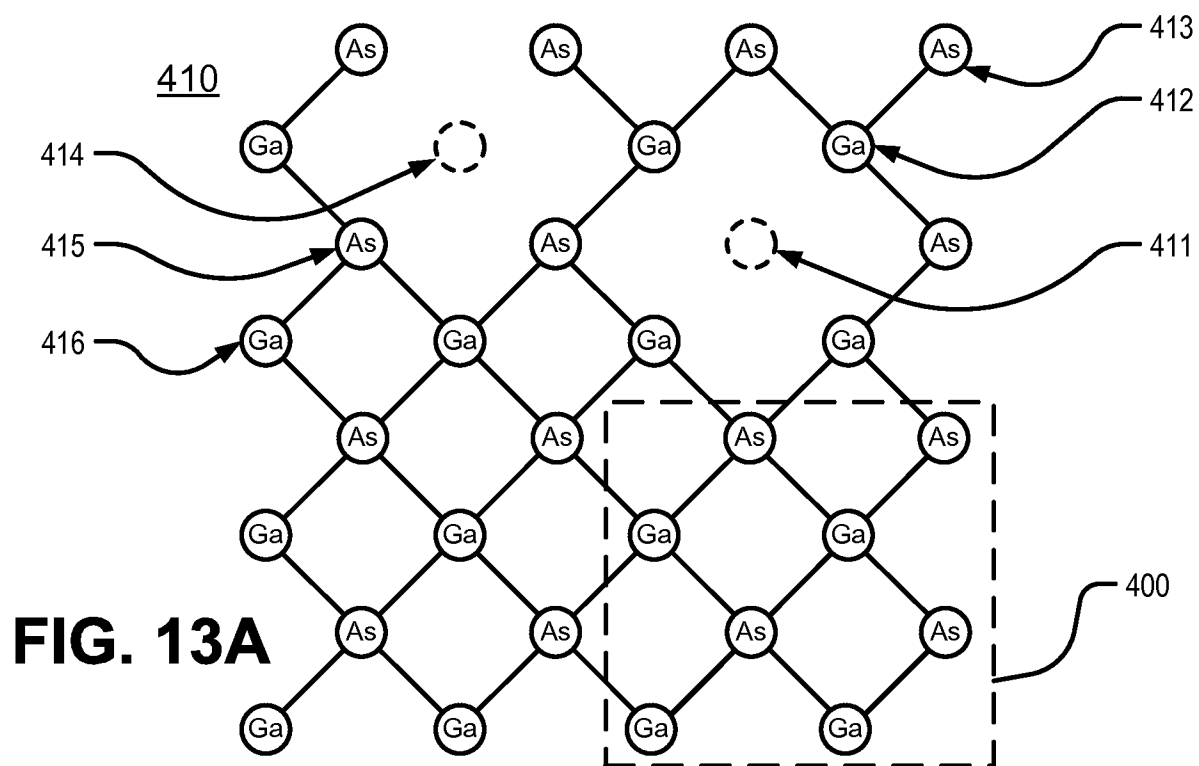
6 / 11



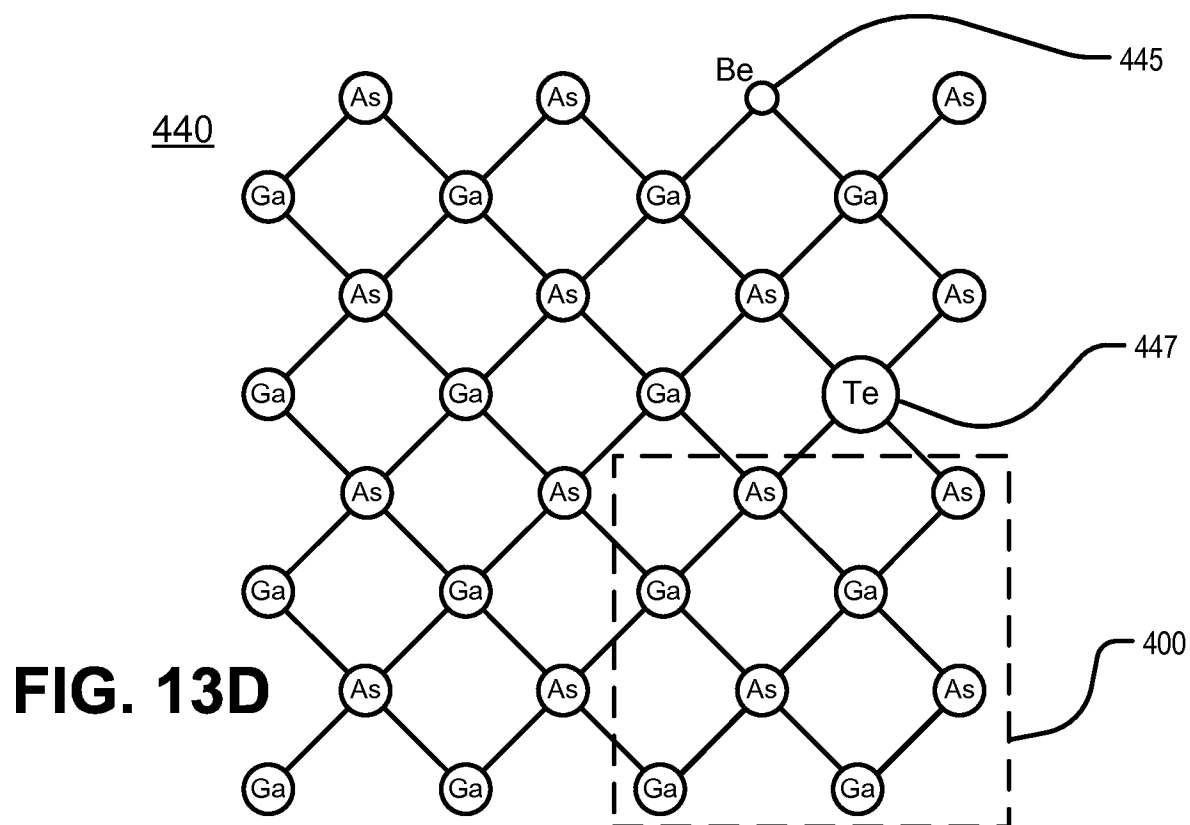
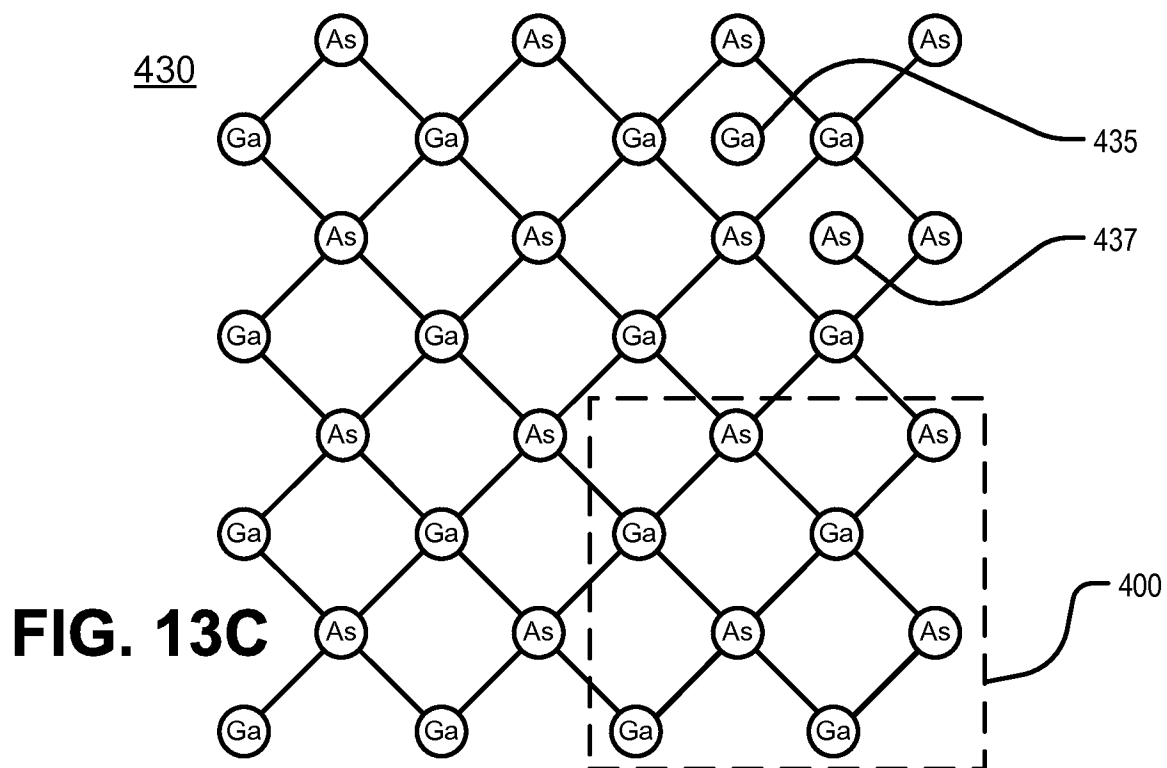
7 / 11



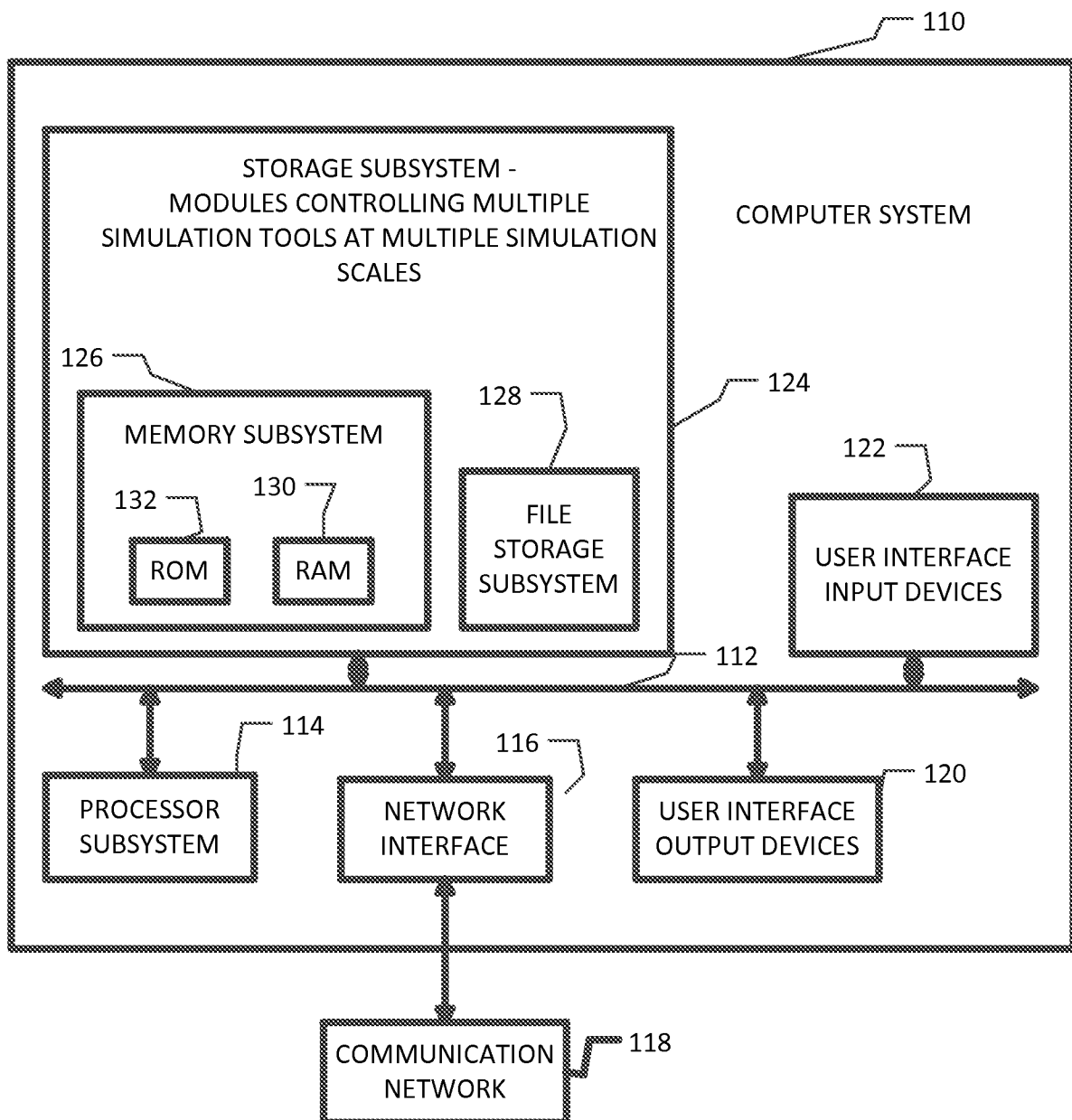
8 / 11



9 / 11



10 / 11

**FIG. 14**

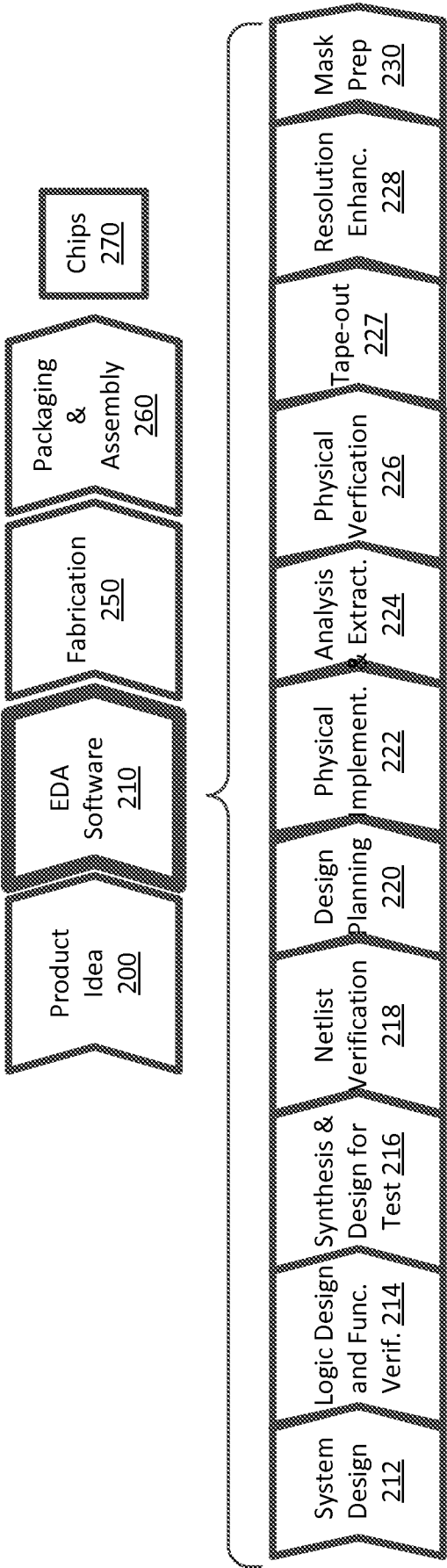


FIG. 15

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2014/057840**A. CLASSIFICATION OF SUBJECT MATTER****G06F 17/50(2006.01)i**

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F 17/50; G06F 19/00; G06F 17/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS(KIPO internal) & Keywords: EDA, design, automation, simulation, parameter, intermediate

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 00-72185 A2 (GET2CHIP) 30 November 2000 See abstract; claims 1-10; and figures 1-3.	1-12
A	WO 01-08028 A2 (KONINKLIJKE PHILIPS ELECTRONICS NV) 01 February 2001 See abstract; claims 1, 3, 11; and figures 1, 2.	1-12
A	US 2013-0139121 A1 (ZE-MING WU et al.) 30 May 2013 See abstract; claims 1, 9, 16; and figures 2-5.	1-12
A	US 2003-0217343 A1 (ROCHIT RAJSUMAN et al.) 20 November 2003 See abstract; claims 1-9; and figures 2, 4, 5.	1-12



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

28 November 2014 (28.11.2014)

Date of mailing of the international search report

28 November 2014 (28.11.2014)

Name and mailing address of the ISA/KR

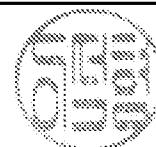
International Application Division
Korean Intellectual Property Office
189 Cheongsu-ro, Seo-gu, Daejeon Metropolitan City, 302-701,
Republic of Korea

Facsimile No. +82-42-472-7140

Authorized officer

LEE, Seok Hyung

Telephone No. +82-42-481-5983



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2014/057840

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 00-72185 A2	30/11/2000	AU 2000-51671 A1 CN 1408092 A0 EP 1248989 A2 JP 04495865 B2 JP 2003-500745 A JP 2003-500745 T US 6470486 B1 US 6516453 B1 US 6782511 B1 WO 00-72185 A3	12/12/2000 02/04/2003 16/10/2002 07/07/2010 07/01/2003 07/01/2003 22/10/2002 04/02/2003 24/08/2004 15/11/2001
WO 01-08028 A2	01/02/2001	EP 1196861 A2 WO 01-08028 A2 WO 01-08028 A3	17/04/2002 01/02/2001 14/02/2002
US 2013-0139121 A1	30/05/2013	US 2014-0082578 A1 US 8607179 B2	20/03/2014 10/12/2013
US 2003-0217343 A1	20/11/2003	CN 100341110 C CN 1647248 A DE 10392497 T5 JP 04508657 B2 JP 2005-522859 A JP 2005-522859 T KR 10-0936855 B1 KR 10-2004-0101421 A US 7178115 B2 WO 03-085706 A1	03/10/2007 27/07/2005 17/02/2005 21/07/2010 28/07/2005 28/07/2005 14/01/2010 02/12/2004 13/02/2007 16/10/2003