1

Introduction

The electrical power system has been revolutionizing over the decades into a highly interconnected, large, and complex renewable system. Populations and economic growth globally demand higher electricity. Transactions crossing large areas are encouraged to make more economic and environmental sense and result in large power flowing over a wide area. High-voltage transmission technologies boosted voltage levels to 1000 kV Ultra-High-Voltage Alternating-Current (UHVAC) and ± 800 kV Ultra-High-Voltage Direct Current (UHVDC) to transmit power over thousands of miles [1]. Advanced power electronic devices enable flexible alternating current transmission system (FACTS), for instance, static var compensator and voltage source converter-based STATCOM, being adapted to control power flow agilely and accurately in electric power grids [2, 3].

The accelerating decarbonization of energy systems promoted and promised worldwide requires a rising penetration level of renewable energy, distributed energy, and energy storage making the power system ever large and more complex. To make the large, complex, and dynamic power system secure and cost-effective, real-time monitoring, operating, and control are crucial. Accurate and fast calculation, combined with intelligent decisions on power systems is more vital than ever, shifting from the analytical Energy Management System (EMS) to the intelligent EMS [4].

While the power system has been evolving to be bigger, the power system is getting to be more intelligent. In the power industry, the smart and intelligent grid is developed based on the following technologies: big data, deep learning, and high-performance computing [5, 6]. Big data and deep learning usually involve intensive computing efforts, thus high-performance computing is the key to making smart decisions on time. Intelligent real-time analysis based on multi-source big data analysis, deep learning techniques, and high-performance computing, as the trend, makes the power system adaptive and predictive possible.

Compared with the traditional grid, the modern power system is operated under more uncertainty because of the intermittent renewable energy and power market transactions as described above. Intelligent real-time analysis and calculation need to adopt these uncertainties more quickly to anticipate extreme events to make better decisions timely. The natural reflection of human beings in complex operating environments requires more accurate and intuitive data models and powerful calculation methods. For example, to provide powerful technical support for intelligent real-time analysis and calculation, introducing the latest numerical calculation technology into the power system power flow, state estimation, contingency analysis, security-constrained automatic generation control, security-constrained unit commitment, and faster-than-real-time transient simulation require novel data structure and calculation approach. The current data

Graph Database and Graph Computing for Power System Analysis, First Edition. Renchang Dai and Guangyi Liu.

processing and calculation approaches of the existing power system applications face the following three challenges:

1) Data Management and Analysis: It is a requirement and challenge to develop data acquisition, processing, and storage technologies that can simultaneously meet the needs of grid online analysis and offline planning for converged multi-source big datasets.
2) Mathematical Methods and Computation: It is necessary and challenging to develop new mathematical tools and algorithms to achieve faster-than-real-time grid simulation.
3) Models and Simulations: The parallel dynamic simulation framework of power systems is not currently developed enough to support real-time, wide-area protection, and control.

The trends of the electric power system are challenging the existing EMS and Market Management System (MMS) in their computationally intensive applications, such as power flow, state estimation, contingency analysis, multi-time point network analysis, security-constrained automatic generation control, security-constrained unit commitment, security-constrained economic dispatch, and faster-than-real-time transient simulation. Next-generation EMS/MMS is required to be evolving to accommodate larger scale, higher complex, more constrained, and uncertain power systems with a faster than real-time manner or even look-ahead capability with future situational awareness [7].

To meet the above challenges, the requirements and goals of parallel analytical technologies and tools are urgently needed to support new data management tools and rapid computational analysis methods.

Due to their efficient data management and rapid computational analysis capabilities, graph database and graph computing technologies are gaining more and more attention. They are promising technologies to effectively solve the problem of big data rapid analysis and processing. In the field of e-commerce, graph computing technology based on a graph database plays a key role in real-time trading and real-time analysis such as anti-money laundering, bad transaction detection, intelligent navigation, and other fields. Many Internet companies have also developed their graph computing technologies and products, such as Pregel, a graph computing system developed and designed by Google. The Trinity project of Microsoft Research is about graph database and graph computing projects. Google's Pregel products have become one of the industry examples of successful graph database applications.

In the field of analysis of power grids, using graph data management and computation technology is new. This book summarizes the recent research and development achievements on this topic. Graph database architecture to power systems is first introduced, which is needed to support fine-grained parallel computing to improve power system computation efficiency. The traditional relational database is replaced by a graph database to model power systems and implement applications. Using the graph database, the program to solve large-scale algebraic equations, high-dimensional differential equations, and optimization problems is reconfigured and redesigned to accommodate graph parallel computing in this book. By using graph database and graph computing, the computational model can be integrated with the grid model, data storage, and numerical calculation, while making full use of big data technologies such as memory computing, distributed parallel computing, and decomposition aggregation. The technology has the significant advantages of large scale, high speed, and high efficiency of computing data, and provides a technical solution with great potential for data management and analysis and calculation of a giant power grid.

Without fast and accurate calculations, a timely response to real-time events is impossible and the system is running at risk. Analysis of the North American blackout of 14 August 2003 shows that delayed and missed responses are the main reasons for the wide-area blackout [8].

Sequence-of-events records show that the system experienced two and a half minutes of disturbances from the initial event to the system collapse. In the initial critical nine seconds, the hundreds of generators that tripped offline were not fully captured by the EMS system, since the EMS update cycle (typically in minutes) is much longer than the critical events interval. Clearly, the opportunity to take timely measures to prevent the blackout was missed and the possibility of a more timely response would have been enhanced by an EMS with a faster cycle time.

The time has come when it is critical to improve the computation efficiency of EMS applications to accommodate modern power systems that are of ever-increasing size and operational complexity. The goal today is to provide a cycle time equivalent to the Supervisory Control and Data Acquisition (SCADA) cycle time or faster to provide look-ahead capability with future situational awareness using forecasted and scheduled information of load forecasting, unit commitment, and outage schedule. The analytical processing time needs to be reduced from tens of seconds to subseconds [7]. To meet this requirement, technologies for the next-generation of high-performance EMS are being studied [9–11]. However, the computation capability to complete the core EMS applications, such as state estimation, power flow, and contingency analysis at a SCADA sampling rate has not yet been achieved.

To achieve EMS computation cycle times that are faster than the SCADA sampling rate, a novel database architecture along with fast computational methods are presented in this book. Among the various computational tools available to improve computation efficiency, parallel computing is a promising technology, providing abundant storage along with multiple processing paths [12, 13]. In [12], parallel state estimation using a preconditioned conjugate gradient algorithm and an orthogonal decomposition-based algorithm is proposed. The proposed algorithm can solve state estimation problems faster using parallel computing, but it is infeasible to deal with a large condition number of a gain matrix. Alves and Monticelli in [13] proposed an approach to solving contingency analysis by parallel computer and distributed network. To utilize the linearity of the power system component, current balance equations subjected to Kirchhoff's current law are used to model power flow problems in [14]. The result shows a reduction in computational time by over 20% when using the current balance equations.

The state-of-the-art communication technologies and computation technologies in information systems are brilliantly showing power system engineers a technical solution to measure, monitor, and analyze electric power systems widely and quickly. However, when the exponentially growing data is acquired at the control center, the database and computation engine consumes a longer time to process which deteriorates the computation efficiency. To achieve analysis with high computation efficiency, novel system architecture, and fast computational algorithms are needed to assist operators to ensure a reliable, resilient, secure, and efficient electric power grid promptly. Among the various computational algorithms, parallel computing is a promising technology to improve computation efficiency taking advantage of modern computation technology, abundant storage space, and parallel capability of database and GPU. Multiple-core CPUs and GPUs are available nowadays as affordable hardware configurations to facilitate parallel computing. However, the state-of-art EMS/MMS does not effectively harness the multi-threaded parallelization capability in their applications [12, 13] for the reason of the traditional relational database and computation algorithms applied by EMS/MMS were not designed for parallel computing.

To accommodate parallel computing, both the database and calculation approaches for the EMS/MMS applications need to be redesigned to fit into a parallel database management system and parallel computing. Previous works investigated the feasibility of adopting graph computing on topology processing, state estimation, power flow analysis, "$N-1$" contingency analysis, and security-constrained economic dispatch [15–21]. Realizing that parallel processing of the power system applications needed by real-time operation and long-term planning can be enhanced by taking advantage of the embedded graph characteristics of a power system, this book has married a graph-based database with graph computing to achieve high computational effective power system analysis to accommodate the evolving power systems and power market.

To accommodate parallel computing, database and mathematical model for power system calculation need to be redesigned to fit into parallel database management, parallel analysis, and fast visualization.

In this book, the critical power system applications are revisited. The computational approaches involved in these applications are introduced in detail. These approaches are abstracted to be mathematical problems in solving large-scale algebraic equations, high-dimensional differential equations, and mixed integer linear optimization problems. Graph data structure and graph parallel computing are introduced to model the power system in graph and solve the problems in parallel.

1.1 Power System Analysis

1.1.1 Power Flow Calculation

Power flow calculation is a well-known application in power system analysis. The intention of power flow calculation is to obtain bus voltage magnitude and angle information. Once the voltage information is known, active power and reactive power flow on each branch can be analytically determined. In mathematics, the power flow calculation model is a set of high-dimensional nonlinear algebraic equations.

There are several different methods to solve nonlinear equations. The well-known Newton–Raphson method linearizes equations using a Taylor series with the linear term only. Industry-grade EMS also uses the Fast-decoupled power flow method to approximate active and reactive flow equations by decoupling voltage magnitude and angle calculations. Although decoupled power flow method takes a few more iterations than Newton–Raphson method to converge, each iteration takes much less time. For reactance-dominated transmission networks, decoupled power flow method outperforms the Newton–Raphson method on computation efficiency. The cost is the approximation on Jacobian matrices by decoupled power flow method deteriorates power flow convergence. Usually, in industry-grade EMS, decoupled power flow method is conducted first, then Newton–Raphson method second if decoupled power flow method diverges. This strategy practically provides supporting evidence of its effectiveness for contingency analysis for a large-scale system with thousands of contingencies.

1.1.2 State Estimation

The power system state estimation (SE) is based on real-time telemetry from SCADA. The network topology connection of the power system is determined in real-time, along with the real-time operating state of the power grid which forms the basis of the online analysis software. It serves to monitor the state of the grid and enables EMSs to perform various important control and planning tasks such as establishing near real-time network models for the grid, optimizing power flows, and bad data detection and analysis.

There are at least three major aspects of the future power grid that will directly impact SE research. First, more advanced measurement technologies like phasor measurement units have offered hope for near real-time monitoring of the power grid.

Second, new regulations and market pricing competition may require utility companies to share more information and monitor the grid over large geographical areas. This calls for distributed control, and hence, distributed SE to facilitate interconnection-wide coordinated monitoring.

Lastly, to facilitate smart grid features such as demand response and two-way power flow, utility companies will need to have more timely and accurate models for their distribution systems. This calls for SE at the distribution level, which places more stringent requirements on SE algorithms. So far, utility companies have done little in implementing SE in distribution systems, even though SE has been deployed extensively in transmission systems for decades. However, as the electric power grid becomes smarter, more distribution automation will be needed and SE at the distribution level will become more important. The control mechanism in the distribution system will most likely be distributed and active in nature, and so will be the corresponding SE functions. This necessitates the development of new distributed SE algorithms that avail themselves of the substantially increased number of real-time measurements.

1.1.3 Contingency Analysis

It is a challenge and a goal to operate a large-scale, complex, and dynamic power grid with safety and cost-effectiveness. Contingency analysis is one of the applications to secure power systems operating with no violation. Contingency analysis uses base case power flow driven from SE to assess the security of power systems under the contingency of a single equipment outage and their combinations. Contingency analysis is usually running periodically every one to two minutes.

The contingency analysis is usually based on an online power grid analysis to figure out the weak point and security risks of the power grid and issue an alarm when the system is running at risk. It facilitates dispatching operators to deal with potential operation issues in time to prevent cascading events and blackouts.

Contingency analysis is time-consuming as it involves a large number of computations of AC load flow. To reduce the computational time, an automatic contingency screening approach is being adopted which identifies and ranks only those outages which cause the limit violation on power flow in the lines or voltages on the buses. Practically, only selected contingencies will lead to severe conditions in the power system. Therefore, the process of identifying these severe contingencies is referred to as contingency selection and this can be done by calculating severity indices for each contingency. This is important to target the vulnerable point in a large-scale power system network with a minimum time requirement.

The potential of artificial neural networks for nonlinear adaptive filtering and control, their ability to predict solutions from past trends, their enormous data processing capability, and their ability to provide fast responses in mapping data make them a promising tool for their application to power systems.

Looking forward, real-time and intelligent technologies need to be developed for contingency analysis to promote a look ahead and predictive security awareness. The development of big data and high-performance computing technology is the key to making this goal possible.

1.1.4 Security-Constrained Automatic Generation Control

The automatic generation control (AGC) is used to balance active power and regulate tie-line power flow while minimizing the power generation cost. In the present state of the art, the AGC base point is determined by the economic dispatch (ED) and AGC regulates the area control error to be zero and controls the tie-line power flow to the desired command.

ED optimizes generation under network security constraints. There are two commonly used methods for active ED in power systems: (i) offline ED and (ii) online ED.

Offline ED calculates unit commitment and dispatches unit active power output for the next day or the next few days in a time interval of hours based on the generation capacity, grid network constraints, as well as the forecasted load.

Since offline ED is based on load forecasting, the generation dispatch may not accurately meet the actual load. The operating conditions of the power system are changing, and the active power output of the generators may deviate from the scheduled power generation set point. Therefore, online ED adjusts the generation output set by offline ED continuously to satisfy the power system's actual operating point in a short time interval (5–15 minutes).

In practice, for a middle-scale power system, ED calculation with network security constraints takes minutes. Due to the short cycle of AGC control conflicting with the extensive computation efforts of the security-constrained optimization, network security constraints are not modeled in AGC in real-time. When power system operation changes significantly in the ED cycle, AGC commands cannot guarantee that the network security constraints are satisfied.

The present state of the art assumes the power system operation point does not change significantly enough between two ED executions to push the AGC base point determined by ED to violate network security constraints. This assumption is not always true in power system operations.

In the case of intermittent renewable energy penetrated power systems and fast-response power electronics-based generation integrated transmission and distribution power network, the system power flow has a high probability of shifting away from the base point which is optimized in the cycle of ED. AGC does not optimize power flow within network security constraints. With the present AGC command, the power flow may result in violations. Thus, the AGC without network security constraints presents risks in power system operation.

When SE estimates violated power flow, the present state of the art heuristically changes generation limits of generators for AGC regulation to alleviate the risk of overflow under AGC command. This approach has drawbacks in that it is heuristic and the SE execution cycle in minute-scale cannot fit into AGC execution in seconds.

Taking advantage of the high performance of graph computing, the network security-constrained AGC is potentially achievable.

1.1.5 Security-Constrained ED

Optimization theory is an important tool in decision science and the analysis of power system operation. In security-constrained ED, objectives are set as a quantitative measure of the performance of the system under study, which could be the power system generation cost, renewable energy, or a combination of quantities. The main objective is to find values of the variables that optimize the objective. In addition, the variables are restricted or constrained to meet physical laws and security requirements. The process of identifying objectives, variables, and constraints for a given problem is known as modeling. Once the model has been formulated, an optimization algorithm can be used to find its solution.

Unconstrained optimization approaches are the basis of constrained optimization algorithms. Particularly, most of the constrained optimization problems in power system operation can be converted into unconstrained optimization problems. The major unconstrained optimization approaches that are used in power system operation are the gradient method, line search, Lagrange multiplier method, Newton–Raphson optimization, trust-region optimization, quasi-Newton method, double dogleg optimization, conjugate gradient optimization, and so on.

A general formulation of constrained optimization approaches can be modeled as:

$$\min_{x \in \Omega} f(x) \tag{1.1}$$

where $\Omega = \{x \mid c_i(x) = 0,\ \ i \in \mathcal{E}; c_j(x) \geq 0,\ \ j \in \mathrm{I}\}$.

Linear programs have a linear objective function and linear constraints, which may include both equalities and inequalities. The feasible set is a polytope, a convex, connected set with flat, polygonal faces. The contours of the objective function are planar.

Power system operation problems are nonlinear. Thus, nonlinear programming (NLP)-based techniques can handle power system operation problems such as the optimal power flow problem and security-constrained ED with nonlinear objective and constraint functions.

The linear programming models discussed above have been continuous, in the sense that decision variables are allowed to be fractional. However, fractional solutions are not realistic in power system unit commitment. This problem is called the integer-programming problem. It is said to be a mixed integer program when some, but not all, variables are restricted to be integers.

1.1.6 Electromechanical Transient Simulation

Transient stability analysis assesses the state of the power system after a severe disturbance using transient simulations. The mathematical model for describing power system dynamic behavior is a nonlinear dynamic system that includes high-dimensional nonlinear differential equations and larger scale nonlinear algebraic equations (DAEs). When the time domain method is applied to power system transient simulation, differential-algebraic equations are solved by using time-consuming numerical integration methods. To improve the computation efficiency, the choice of a proper step size and parallel computing is essential.

If the step size is too large, the result will become inaccurate or even completely wrong when the large step size is not within the range of numerical stability. If the step size is too small, the transient simulation will take longer than necessary to keep the accuracy. The adaptive time step is a solution that uses the smallest possible time step to obtain an accurate result, thereby increasing the calculation speed while ensuring the calculation accuracy.

In each time step, we need to solve high order high-dimensional differential-algebraic equations. To further improve the computation efficiency, a parallel computing algorithm has been investigated in power system transient simulation. In this book, a graph-based parallel computing method is demonstrated with the adaptive time-step numerical integration method.

Using the sequential method, in each iteration, the network equations are solved to update the network bus voltage including the generator terminal voltage. Differentiate equations use generator terminal bus voltage as a boundary condition to solve the dynamic states of the generator, exciter, governor, Power System Stabilizer (PSS), and current injections from generators to networks. The updated current injections are applied to solve network equations in the next iteration until the converged solution is achieved.

Graph computing demonstrated outperformance on power system steady-state applications where the technology will be used to solve algebraic equations in the sequential method-based transient simulation. In the sequential method, since the differential equations for each generation system are independent once the terminal voltages are solved by algebraic equations, the differential equation sets can be solved by graph parallel computing naturally which will be addressed in this book in detail.

1.1.7 Photovoltaic Power Generation Forecast

In recent years, the rapid exhaustion of fossil fuel sources, environmental pollution concerns, and the aging of developed power plants are considered crucial global concerns. As a consequence, renewable energy resources including wind and solar have been rapidly integrated into the existing power grids. The reliability of power systems depends on the capability of handling expected and unexpected changes and disturbances in production and consumption while maintaining quality and continuity of service. The variability and stochastic behavior of photovoltaic (PV) power are caused by including voltage fluctuations, as well as local power quality and stability issues [22]. Hence, accurate photovoltaic power generation forecasting is required for the effective operation of power grids [23].

The studies in solar irradiance and photovoltaic power forecasting are mainly categorized into three major classes:

1) The persistence models sever as a baseline that assumes the irradiance values at future time steps are equal to the same values at the forecasting time [22].
2) Physical models employ physical processes to estimate future solar radiation values using astronomical relationships [24], meteorological parameters, and numerical weather predictions (NWPs) [25].
3) Statistical and artificial intelligence techniques estimate or regress solar irradiance and photovoltaic power generation [26–32].

To remove the strong smoothness assumption, increase the generalization capability, and improve the computation efficiency, in this book, the problem of spatio-temporal probabilistic solar radiation forecasting is presented as a graph distribution learning problem. In the approach, a set of solar measurement sites in a wide area is modeled as an undirected graph, where each node represents a site and each edge reflects the correlation between historical solar data of its corresponding nodes/sites to model the solar radiation spatio-temporal characteristics.

1.2 Mathematical Model

In general, power system analysis could be transformed to solve a linear system $Ax = b$, differential equations $\frac{dx}{dt} = f(x, t)$, and/or optimization problems $\min_{x \in \Omega} f(x)$. As a fundamental function, graph parallel computing approaches to solve the three typical mathematical problems involved in power system analysis, optimization, and simulations are key components in this book.

1.2.1 Direct Methods of Solving Large-Scale Linear Equations

Direct methods are widely used to solve linear equations by a finite sequence of operations. In the absence of rounding errors, direct methods would deliver an exact solution. Besides their high efficiency to solve moderate-size linear systems, direct methods are also popular to solve large sparse linear systems, like power flow, SE, and other power system problems. Sparse direct methods are a tightly coupled combination of techniques from numerical linear algebra, graph theory, graph algorithms, permutations, and other topics in discrete mathematics [33]. And such problem has been extensively studied.

This book focuses on direct methods for sparse linear systems, such as lower–upper (LU), Cholesky, and other factorization, and the implementation by graph parallel computing. It first introduces basic concepts, such as the definition and data structure of the sparse matrix, used in direct methods. As lots of graph concepts and algorithms are used in direct methods, the relationship between matrix and graph is included in the book of basic concepts.

1.2.2 Iterative Methods of Solving Large-Scale Linear Equations

The iterative method includes a series of techniques that use successive approximations to obtain more accurate solutions to a linear system at each step. Iterative methods can be expressed as $x^k = B \cdot x^{(k-1)} + c$, where B and c are constant. Direct method is widely used in solving power system problems. Iterative methods are applied in power system analysis as well including the Jacobi method, Gauss–Seidel method, successive over-relaxation method, symmetric successive over-relaxation method, conjugate gradient method, generalized minimal residual method, and bi-conjugate gradient.

1.2.3 High-Dimensional Differential Equations

The ordinary differential equations (ODEs) with an initial value that appeared in transient simulations are of the form

$$\frac{dx}{dt} = f(x, t), \quad x(t_0) = x_0. \tag{1.2}$$

Here the solution $x = x(t)$ needs to be solved for any time $t > t_0$. The variable x, the right-hand-side function $f(x, t)$, and the initial value x_0 can be either a scalar value or a vector. In addition, $f(x, t)$ can be either linear or nonlinear.

There are generally two classes of numerical methods for solving ODEs: (i) one-step methods and (ii) linear multistep methods.

One-step methods make use of the previously computed x_n to produce a value of x_{n+1}. They do make use of one or more evaluations of the function at intermediate points between t_n and t_{n+1} in order to improve accuracy. However, these function evaluations are then discarded and are not reused in making future steps. In contrast, multistep methods make use of the previously found values x_n, x_{n-1}, ... in order to produce a value of x_{n+1}.

The simplest class of one-step methods is the Euler's methods and their relatives – forward Euler, backward Euler, and Trapezoidal rule. One disadvantage of backward Euler and the Trapezoidal rule is that they require solving implicit equations at each time t_{n+1}. In addition to achieving higher orders of accuracy, methods in the Runge–Kutta family self-adapt the time step sizes. These methods always include two methods, one with higher order which is used for the error estimation. The other lower-order method is for the approximated solution. These two methods are designed to have the same intermediate steps, which ensure that the extra computation effort for error estimation is negligible.

1.2.4 Mixed Integer-Programming Problems

Consider an optimization problem modeled as a Mixed-Integer Linear Program (MILP) has the following structure:

$$\min c^T x \tag{1.3}$$

$$s.t.\ \ Ax + By + Cv \leq a \tag{1.4}$$

$$A'x + B'y + C'v \leq a' \tag{1.5}$$

$$x \in \{0,1\}^n, y \in \{0,1\}^m, v \in R^p \tag{1.6}$$

This is an optimization problem that has both continuous (v) and binary (x and y) sets of variables, and only some of the binary variables (x) have nonzero objective function coefficients. The constraint set can be divided into two subsets. The first set of constraints (1.4) models aspects of the problem that can be represented efficiently in the MILP framework (e.g. assignment constraints) and has a significant impact on the Linear Program (LP) relaxation. The second set of constraints (1.5), on the other hand, is assumed not to significantly affect the LP relaxation and is usually large in number because of the limited expressive power of MILP methods.

1.3 Graph Computing

When performing power system computing, from topology process, admittance matrix formation, matrix factorization, forward and backward substitution, optimal search, and numerical integration, to state visualization, a large number of database operations are called repeatedly on data reading, writing, searching, and concurrent accessing. A relational database uses join-intensive queries for the whole database for many database operations inviting more computation time for large datasets. On the contrary, the graph database outperformed a relational database in these database operations [34]. The database operation mimeograph dataset is proportional to the number of subgraphs other than the entire graph leveraging the graph database's nodal parallel and hierarchical parallel capabilities.

In the area of power system analysis, optimization, and simulation, using graph data management and computation technology, the computational model can be smoothly integrated with data storage and numerical calculation with in-memory computing, distributed parallel computing, and decomposition aggregation. The technology has great potential for large-scale power grid analysis and calculation.

The traditional Relational Database Management System (RDBMS) uses tables to store data. The RDBMS stores the structured records and their attributes in equal-length tables, and maintains the database using Structured Query Language. Ideally, structured tabular data of arbitrary complexity can be represented by relational databases. However, RDBMS has no flexibility to define unstructured datasets. Relational databases are not very accommodating of data interconnections in a dataset that is graph-based, rather than attribute-based. To represent the relational interconnections, the linking attributes of records in the database are stored in different tables for creating the relations, and the relationship between the different records is established by querying the same attribute of the corresponding records in different tables. Therefore, adding or deleting records in the RDBMS requires updating all tables with the associated shared attributes. When compared with the performance of a Graph Database Management System (GDBMS), an RDBMS takes much longer time to support attribute searching, optimal ordering, depth-first (or breadth-first) search, which limits the efficiency of topology analysis, parallel computing, and results visualization for the traditional power system applications.

In graph computing, the relationship between nodes and edges is self-defined by the graph. Unlike relational databases, graph databases model the power system using graph-oriented data structures for semantic queries with a set of nodes and edges [35, 36]. Unstructured attributes are then stored in data structures defined by the graph's nodes and edges. For power system applications, the GDBMS is more in line with the requirements of power system computing for complex data modeling, querying, sorting, and traversal.

Commonly used data structures for power system calculations include arrays, linked lists, trees, and graphs. For example, an array is often used for matrix operations; a linked list is used to represent the path set of generators and loads when taking advantage of fast forward and fast backward substitution (sparse vector methods) while solving matrix equations [37, 38]; an elimination tree is used for identifying parallel processes when parallelizing factorization [39, 40]; and a graph is used for topological processing [41].

The graph database is concise when modeling different data structures, from simple array data to the more complex structures which store graphs and trees. Hence, the graph database is very accommodating of data interconnections within a dataset, which fits the characteristics of power systems since the power system is naturally modeled as a graph consisting of nodes and branches. Nodes are physically connected through branches as edges. The unconstructed parameters of the bus, generator, load, and branch, such as active and reactive power of generator and load, resistance and reactance of lines and transformers, and so on are stored in the node or edge. The graph structure itself naturally represents the topology of the electric power grid.

The remarkable performance of the GDBMS results from its built-in parallel computing capability. The GDBMS allows us to take advantage of two types of parallelism: nodal parallelism and hierarchical parallelism.

In graph database operation parlance, nodal parallel computing refers to the computation of quantities associated with each node, where each computation is independent; in contrast, hierarchical parallelism partitions nodes into different levels according to their computing dependency and then performs the computation in parallel on nodes at the same level.

For example, nodal parallelism can be used when searching for a specific attribute value since no dependencies exist: GDBMS compares the desired value against the attribute value of all nodes and edges in parallel. In contrast, when dependencies exist and are characterized by precedence relationship depth level, the GDBMS processes all nodes using a depth-first search, exploiting hierarchical parallelism. By leveraging the graph database's nodal parallelism and hierarchical parallelism capabilities, the database operation time on graph datasets is proportional to the size of subgraphs rather than the size of the entire graph.

The parallel solution of power system application using the graph-based approach illustrated in this book will demonstrate that it can be used not only for SE, power flow, and contingency analysis but also for other applications, such as dynamic security assessment and transient stability simulation.

1.3.1 Graph Modeling Basics

A graph can be simply expressed as

$$G = (V, E) \tag{1.7}$$

where V denotes the set of nodes and E denotes the set of edges. When we design the structure of graph data in a power system, attributes of nodes and edges are used to store static/dynamic variables, such as voltage ratings, impedance (admittance), power capacity rating, real-time voltage, and real-time power. For each v_i, the description of its state is represented by a set of independent attributes $p_{vi} \in P$, and for the edge e_{ij} connecting with node v_i and node v_j, the description of its state is also represented by a set of independent attributes $p_{eij} \in P$. According to the above definition, a graph can be constituted by the V, E, and P sets. In numerical calculation, an element $a_{ij} \neq 0$ $(i \neq j)$ in admittance matrix A is equivalent to nodes v_i and v_j are connected, on the contrary, $a_{ij} = 0$ equivalent to the edge e_{ij} does not exist, which means nodes v_i and v_j are not connected.

1.3.2 Graph Parallel Computing

Graph parallel computing exploits both nodal and hierarchical parallelism. In power system applications, the tasks of matrix formation, right-hand-side correction vector calculation, and branch power flow calculation can be nodally parallelized. The tasks of matrix factorization and forward/backward substitution can be hierarchically parallelized.

When the power system is modeled using a graph structure, calculations such as admittance matrix formation can be performed using nodal parallel computing once the sparsity-based symbolic part of the data structure has been completed.

In the graph structure, each numbered vertex represents a bus, and the edge between two vertices is a branch. In the graph structure, the counterparts of the connections between nodes are nonzero off-diagonal elements in the coefficient matrix A. Zero (absent) elements in the matrix schematic representation indicate that no direct connections between the nodes exist in the graph.

To form a row of the admittance matrix for any vertex, the off-diagonal nonzero elements of each row correspond only to the adjacent nodes. The graph-based approach searches for the neighboring vertex (or vertices) and the edges between them to form the admittance matrix numerical entries for all vertices simultaneously.

Other nodal parallelizable calculations in the power system analysis applications include the right-hand-side vector calculation, the convergence check, bad data detection, the branch power calculation, and the voltage and power flow violation check. These calculations, performed on node attributes, are independent from each other and can therefore be performed simultaneously.

In graph hierarchical parallel computing, computation is performed in parallel on nodes at the same depth level of the elimination tree. The nodes are partitioned into different levels in the elimination tree according to their calculation dependency. The calculation of quantities associated with the higher-indexed-level nodes depends on the calculation of quantities associated with the lower-level nodes. The calculations associated with nodes at the same level are independent and are performed in parallel.

In power system analysis, the hierarchically parallelizable tasks include matrix factorization and forward/backward substitution.

References

1 Liu, Z. (2014). *Ultra-High Voltage AC/DC Grids*. Waltham, MA: Academic Press.

2 Jayant Baliga, B. (2011). *Advanced High Voltage Power Device Concepts*. New York: Springer.

3 Tyll, H.K. (2004). *2004 IEEE/PES Transmission and Distribution Conference and Exposition: Latin America* (IEEE Cat. No. 04EX956), Sao Paulo, Brazil, pp. 976–980. https://doi.org/10.1109/TDC.2004.1432515.

4 Erkeng, Y., Liu, G., and Zhou, J. (1998). *Energy Management System*. Science Press.

5 Najafabadi, M.M., Villanustre, F., Khoshgoftaar, T.M. et al. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data* 2: 1.

6 Chen, Y., Huang, Z., Jin, S., and Li, A. (2022). Computing for power system operation and planning: then, now, and the future. *iEnergy* 1(3): 315–324.

7 Mansoor, A. and Gellings, C. (2021). *Needed: A Grid Operating System to Facilitate Grid Transformation*. Palo Alto, CA: Electric Power Research Institute. 94304-1338.

8 North American Electric Reliability Council (2004). Technical analysis of the August 14, 2003, blackout: what happened, why, and what did we learn? *Report to the NERC Board of Trustees by the NERC Steering Group*, 13 July 2004.

9 Myrda, P.T. and Grijalva, S. (2012). The need for next generation grid energy management system. *2012 CIGRE US National Committee Grid of the Future Symposium*, Kansas City, MO (July 2021).

10 Wix, S.D. and Plunkett, P.V. (2009). *Conference: Proposed for presentation at the Clean Technology Conference & Expo 2009*, Houston, TX (3–7 May 2009).

11 The U.S. Department of Energy (2015). An assessment of energy technologies and research opportunities. In: *Chapter 3. Enabling Modernization of the Electric Power System*. Washington, DC. https://www.energy.gov/sites/prod/files/2017/03/f34/qtr-2015-chapter3.pdf.

12 Chen, Y., Jin, S., Rice, M., and Huang, Z. (2013). Parallel state estimation assessment with practical data. *2013 IEEE Power & Energy Society General Meeting*, Vancouver, BC, pp. 1–5. https://doi.org/10.1109/PESMG.2013.6672742.

13 Alves, A.C.B. and Monticelli, A. (1995). Parallel and distributed solutions for contingency analysis in energy management systems. *38th Midwest Symposium on Circuits and Systems. Proceedings*, vol. 1. Rio de Janeiro, Brazil, pp. 449–452. https://doi.org/10.1109/MWSCAS.1995.504473.

14 da Costa, V.M., Martins, N., and Pereira, J.L.R. (1999). Developments in the Newton Raphson power flow formulation based on current injections. *IEEE Transactions on Power Systems* 14 (4): 1320–1326. https://doi.org/10.1109/59.801891.

15 Shi, J., Liu, G., Dai, R. et al. (2018). Graph based power flow calculation for energy management system. *2018 IEEE Power & Energy Society General Meeting (PESGM)*, Portland, OR, pp. 1–5. https://doi.org/10.1109/PESGM.2018.8586233.

16 Yuan, C., Zhou, Y., Zhang, G. et al. (2018). Exploration of graph computing in power system state estimation. *2018 IEEE Power & Energy Society General Meeting (PESGM),* Portland, OR, pp. 1–5. https://doi.org/10.1109/PESGM.2018.8586535.

17 Chen, T., Yuan, C., Liu, G., and Dai, R. (2018). Graph based platform for electricity market study, education and training. *2018 IEEE Power & Energy Society General Meeting (PESGM)*, Portland, OR, pp. 1–5.

18 Zhao, Y., Yuan, C., Liu, G., and Grinberg, I. (2018). Graph-based preconditioning conjugate gradient algorithm for "N−1" contingency analysis. *2018 IEEE Power & Energy Society General Meeting (PESGM)*, Portland, OR, pp. 1–5. https://doi.org/10.1109/PESGM.2018.8586214.

19 Qiu, H., Zhou, A., Hu, B. et al. (2018). TAnalyzer: "a graph database based topology analysis tool for power grid". *Proceedings of the 2nd International Conference on Computer Science and Application Engineering*, Hohhot, China, October 2018, pp. 1–5.

20 Kan, B., Zhu, W., Liu, G. et al. (2017). Topology modeling and analysis of a power grid network using a graph database. *International Journal of Computational Intelligence Systems* 10: 1355–1363.

21 Lv, X., Chen, S., Zheng, S. et al. (2018). Understanding the graph databases and power grid systems. *International Conference on Advanced Electronic Materials, Computers and Materials Engineering*, vol. 439, Issue 3, Singapore, September 2018.

22 Wan, C., Zhao, J., Song, Y. et al. (2015). Photovoltaic and solar power forecasting for smart grid energy management. *CSEE Journal of Power and Energy Systems* 1 (4): 38–46.

23 Jiang, Y., Long, H., Zhang, Z., and Song, Z. (2017). Day-ahead prediction of bihourly solar radiance with a Markov switch approach. *IEEE Transactions on Sustainable Energy* 8 (4): 1536–1547.

24 Hottel, H.C. (1976). A simple model for estimating the transmittance of direct solar radiation through clear atmospheres. *Solar Energy* 18: 129–134.

25 Pfenninger, S. and Staffell, I. (2016). Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data. *Energy* 114: 1251–1265.

26 Bae, K., Jang, H., and Sung, D. (2016). Hourly solar irradiance prediction based on support vector machine and its error analysis. *IEEE Transactions on Power Systems* 32 (2): 935–945.

27 Voyant, C., Notton, G., Kalogirou, S. et al. (2017). Machine learning methods for solar radiation forecasting: a review. *Renewable Energy* 105: 569–582.

28 Lauret, P., Voyant, C., Soubdhan, T. et al. (2015). A benchmarking of machine learning techniques for solar radiation forecasting in an insular context. *Solar Energy* 112: 446–457.

29 Wan, C., Xu, Z., Pierre, P. et al. (2014). Probabilistic forecasting of wind power generation using extreme learning machine. *IEEE Transactions on Power Systems* 29 (3): 1033–1044.

30 Golestaneh, F., Pinson, P., and Gooi, H. (2016). Very short-term nonparametric probabilistic forecasting of renewable energy generation – with application to solar energy. *IEEE Transactions on Power Systems* 31 (5): 3850–3863.

31 Zhang, Y. and Wang, J. (2015). GEFCom2014 probabilistic solar power forecasting based on k-nearest neighbor and kernel density estimator. *2015 IEEE Power & Energy Society General Meeting*, Denver, CO, USA, pp. 1–5. https://doi.org/10.1109/PESGM.2015.7285696.

32 Lauret, P., David, M., and Pedro, H. (2017). Probabilistic solar forecasting using quantile regression models. *Energies* 10 (10): 1591. https://doi.org/10.3390/en10101591.

33 Davis, T.A., Rajamanickam, S., and Sid-Lakhdar, W.M. (2016). A survey of direct methods for sparse linear systems. *Acta Numerica* 25: 383–566.

34 Vicknair, C., Macias, M., Zhao, Z. et al. (2010). A comparison of a graph database and a relational database. *Proceedings of the 48th Annual Southeast Regional Conference* (April 2010), Oxford, MS.

35 Malewicz, G., Austern, M.H., Bik, A.J. et al. (2010). Pregel: a system for large-scale graph processing. *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*, Indianapolis, June 2010, pp. 13–146.

36 Lu, Y., Cheng, J., Yan, D., and Wu, H. (2014). Large-scale distributed graph computing systems: an experimental evaluation. *Proceedings of the VLDB Endowment* 8 (3): 281–292.

37 Tinney, W.F., Brandwajn, V., and Chan, S.M. (1985). Sparse vector methods. *IEEE Transactions on Power Apparatus and Systems* PAS-104 (2): 295–301.

38 Zhang, B. and Chen, S. (1998). *Advanced Power Grid Analysis*. Science Press.

39 Schreiber, R. (1982). A new implementation of sparse Gaussian elimination. *ACM Transactions on Mathematical Software* 8 (3): 256–276.

40 Kumar, P.S., Kumar, M.K., and Basu, A. (1992). A parallel algorithm for elimination tree computation and symbolic factorization. *Parallel Computing* 18 (8): 849–856.

41 Pradeep, Y., Seshuraju, P., Khaparde, S.A., and Joshi, R.K. (2011). CIM-based connectivity model for bus-branch topology extraction and exchange. *IEEE Transactions on Smart Grid* 2 (2): 244–253.