ADVANCED QUANTUM TECHNOLOGIES
www.advquantumtech.com

Consensus-Based Distributed Quantum Decomposition Algorithm for Security-Constrained Unit Commitment Considering Optimal Transmission Switching

Dejian Huang, Wei Dai, Fang Gao, Mingyu Yang, Yu Pan, Jiangwei Ju, and Yin Ma*

To enhance controllability, flexibility, and economic performance, this approach develops a dynamic model for the joint optimization of unit commitment (UC) and optimal transmission switching (OTS), considering grid topology changes. This mixed-integer nonlinear problem sees computational complexity rise sharply with system size. A consensus-based distributed quantum decomposition algorithm (QDA) is proposed, splitting the problem into locally relaxed sub-problems (SPs) and consensus-constrained master problems (MPs). Multi-stage strategy is employed to iteratively refine solutions by incorporating cutting planes. For MPs, quadratic unconstrained binary optimization Hamiltonians are designed to be compatible with quantum computing, their complexity is simplified by inserting up/down constraint Hamiltonians that eliminate auxiliary binary variables. Leveraging quantum annealing and a photonic quantum computer, these Hamiltonians are efficiently solved to explore optimal UC and grid topology. Results on the 6-Bus System validate the centralized QDA, while the consensus-based distributed QDA outperforms Gurobi9 in operational costs and solution time for the IEEE RTS 24-Bus System. The joint UC-OTS optimization reduces costs compared to static networks' UC, demonstrating the mutual influence of unit and topology optimization. This approach won the Grand Prize of the 2024 "Challenge and Leadership Selection" Special Competition of the 19th National Challenge Cup Competition.

1. Introduction

The unit commitment (UC) problem aims to determine the day-ahead unit scheduling, with the objective of minimizing the overall operating costs of the power system while satisfying security constraints. Recent advancements have introduced quantum computing (QC) to enhance UC solutions, such as the divide-and-conquer UC approach, inserting hard constraints into hybrid Hamiltonians.[1] UC is widely applied in various domains, such as security-constrained unit commitment (SCUC) considering N-1 contingencies to mitigate network emergency overloads,[2] the scheduling of service restoration and load shedding events,[3] and UC considering the electricity consumption experience of customers.[4] The optimal transmission switching (OTS) seeks to establish the optimal topology of the power grid. When flow violations or other security-related issues occur within the grid, economical and swift switching of transmission lines can effectively modify the grid topology to alleviate flow congestion. As a flexible and reliable control method, OTS is extensively utilized in areas such as alleviating transmission congestion caused by the integration of renewable energy through the flexibility of transmission lines,[5,6] overvoltage issues,[7] wind power integration models,[8,9] cost reduction,[10,11] and enhancing the resilience and reliability of systems, particularly in consideration of disaster resilience and post-disaster recovery.[12] As a crucial optimization issue in power systems, both UC and OTS mathematical models contain a significant number of binary and continuous variables, classifying them as mixed-integer nonlinear programming (MINLP) problems. They inherently exhibit characteristics of discreteness, nonlinearity, and non-convexity, making them NP-hard problems. With the continuous growth in electricity demand and the development of power markets, the scale and complexity of these models are expected to increase significantly. As a result, NP-hard optimization problems in power systems are increasingly being addressed through the adoption of new computational paradigms to accelerate their solution processes.[13,14]

The traditional static grid structure's security verification strategies struggle to meet the demands for power flow adjustments, while the joint optimization of UC and OTS enables coordinated control of generator scheduling and grid topology. By constructing an optimal grid topology and corresponding

D. Huang, W. Dai, F. Gao, M. Yang
School of Electrical Engineering
Guangxi University
Nanning 530004, China
E-mail: fgao@gxu.edu.cn
Y. Pan
College of Control Science and Engineering
Zhejiang University
Hangzhou 310027, China
J. Ju, Y. Ma
Beijing QBoson Quantum Technology Co., Ltd.
Beijing 100015, China

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/qute.202500241

DOI: 10.1002/qute.202500241

generation plan without compromising the supply to nodal loads, it is possible to significantly reduce the operational costs of the system while also enhancing the grid's reliability in the face of unforeseen incidents. This approach aligns with the direction of smart grid development. Nonetheless, the joint optimization model for UC and OTS necessitates the introduction of a substantial number of binary variables to represent the on/off state of unit, as well as the state of transmission lines. This results in an "exponential explosion" issue as the model scales, posing significant computational challenges for classical computers. Despite these challenges, Oren et al. [15] proposed a joint optimization model for UC and OTS that ensures N-1 reliability. This model decomposes the joint optimization into two groups of subproblems (SPs), which are solved iteratively, demonstrating that optimizing network topology can enhance grid scheduling. However, the iterative method that alternately inputs binary and continuous variable solutions may lead to convergence at local optima, often requiring a substantial number of iterations. Li et al. [16] emphasized the importance of OTS for flexible transmission in UC but noted that it could also cause significant system disturbances, thereby increasing optimization difficulty. Consequently, OTS is regarded primarily as a corrective measure to mitigate or eliminate power flow violations in emergencies. To circumvent the non-convexity induced by binary variables, Neill et al. [17] fixed all integer variables to their optimal values and constructed a dual problem for the N-1 reliable joint optimization for UC and OTS that contained only continuous variables, conducting an economic analysis based on dual concepts. To align the joint optimization model of UC and OTS more closely with the practical construction of smart grids, Shafie-Khah et al. [18] introduced constraints on the number of transmission line switches and proposed a UC model that considered dynamic thermal ratings of transmission lines. They decomposed the original problem into two relaxed SPs that only considered partial constraints based on constraint simplification, iteratively solving to obtain a final solution. However, their simplified model assumes linear costs and does not account for intelligent transmission switching under emergency conditions. Additionally, the decomposition and iteration framework does not consider the interaction between UC and OTS. This results in a somewhat simplistic cutting plane construction for each iteration, making it difficult to guarantee algorithm convergence. Lin et al. [19] introduced a joint optimization model for UC and OTS that incorporated short-circuit current constraints. This model defined the impact of each transmission line's operational state on short-circuit current as the sensitivity of the short-circuit current for that system. They constructed linearized short-circuit current constraints in the form of Benders cutting planes and utilized a stepwise solving strategy to solve the model within a reasonable time. However, their simplified model only provides a fixed network topology with a single transmission switching state, failing to achieve hourly dynamic network topology optimization, thus somewhat avoiding the complexity associated with optimizing transmission line topology. Shahidehpour et al. [20] introduced OTS into UC to reduce transmission violations and lower operational costs. However, their model overlooked the OTS in post-emergency scenarios, i.e., the real-time topological changes in response to sudden incidents. Street et al. [21] built upon this model by incorporating an emergency switching model for transmission lines after incidents. They achieved the joint optimization through an improved exact nested column-and-constraint generation algorithm. Nonetheless, this precise decomposition technique remains centralized, failing to significantly reduce the optimization complexity of the SPs and the master problems (MPs), both of which are NP-Hard MINLPs. This leads to combinatorial explosion issues and considerable optimization difficulty, resulting in low computational efficiency. He et al. [22] proposed a method that simultaneously considered transmission topology and physical vulnerability indices. By utilizing graph theory-based flow algorithms, their approach identified potentially overloaded lines after transmission line faults, effectively prioritizing emergencies to achieve safer and more economical solutions.

The aforementioned works have significantly advanced the research on the joint optimization of UC and OTS, highlighting the interrelationship between grid topology and unit scheduling while continuously enriching the model's constraints to align with practical operational requirements. However, two main issues persist: 1) There has been insufficient improvement in the solving efficiency of the joint optimization model, along with challenges in ensuring algorithm convergence. Although much of the literature employs a strategy of decomposing the original problem into multiple SPs, this process often fails to leverage the interactions between UC and OTS, leading to low efficiency in decomposition. Furthermore, the resulting SPs may still be classified as MINLPs, posing substantial optimization difficulties; 2) The simplifications made to the joint optimization model, such as neglecting real-time topological changes due to sudden incidents or considering only static topology models to address exponential search space issues, hinder the ability to effectively manage joint optimization problems involving complex system constraints. These challenges ultimately stem from the difficulties classical computation faces in addressing the "exponential explosion" problem. As a new computational paradigm, QC may offer a potential solution. The rapid development of quantum algorithms and the availability of real quantum hardware have propelled QC into various fields. Utilizing QC to tackle various optimization problems has emerged as a compelling research topic.

QC is an emerging computational paradigm with significant development potential and prospects. It leverages quantum properties such as superposition and entanglement to perform computations, demonstrating quantum advantage in solving problems like random quantum circuit sampling,[23,24] Gaussian boson sampling,[25,26] and combinatorial optimization.[27,28] Quantum computers operate according to the laws of quantum mechanics, with the fundamental unit being the qubit. n qubits can store information analogous to 2^n classical bits, and they can also operate in parallel, reducing computational complexity and substantially enhancing computational power for specific problems. Currently, quantum computers are categorized into universal quantum computers and specialized quantum computers designed for specific computational tasks. Universal quantum computers operate on the quantum gate model,[29,30] implementing computations through a sequence of quantum gate operations on qubits. These quantum computers can execute a variety of standard quantum algorithms, including Shor's algorithm for integer factorization,[31] Grover's algorithm for unstructured database search,[32] and the HHL algorithm for solving linear equations.[33] However, these quantum computers re-

quire extremely low-temperature environments and, in the current noisy intermediate-scale quantum era,[34] the number of available qubits is insufficient to solve practical problems. Therefore, classical computing is often utilized to process information, reducing the demands on the number of qubits and the depth of quantum circuits. This has led to the development of many hybrid quantum-classical algorithms, such as the improved HHL algorithms HIPEA [35] and HMPEA,[36] the Variational Quantum Eigensolver,[37] and the Quantum Approximate Optimization Algorithm.[38] Typical specialized quantum computers include quantum annealer based on the quantum adiabatic theorem [39] and coherent Ising machine (CIM) based on optical systems,[40] These computational platforms can be applied to solve combinatorial optimization problems based on the Ising model [40,41] and have been successfully implemented on real quantum hardware. Quantum annealers leverage the tunneling effect of quantum mechanics to find global optima more rapidly and can, under certain conditions, converge to global optimal values more easily than classical annealers. Quantum gate-based QC is polynomial-time equivalent to adiabatic quantum computation.[42] CIMs utilize a dissipative architecture based on optical quanta to perform computations by constructing networks connected by optical pulses, demonstrating strong resistance to environmental noise and errors. Combinatorial optimization problems can be equivalently transformed into Quadratic Unconstrained Binary Optimization (QUBO) problems,[43] typically represented as min X^TQX, where $X \in \{0,1\}^n$, $Q \in R^{n\times n}$. Quantum annealers and CIMs exhibit high computational efficiency in solving QUBO problems and have already been applied in various practical scenarios, including chemical synthesis,[44–46] vehicle and job-shop scheduling,[47,48] and financial investments.[49–51] In particular, CIMs have shown remarkable scalability and efficiency in addressing optimization problems in real-world applications.[52]

To enhance the solving efficiency of the joint optimization model for UC and OTS under security constraints, this work proposes a consensus-based distributed quantum decomposition algorithm (QDA). This algorithm employs a multi-stage optimization strategy that synchronously and in parallel constructs cutting planes across different buses, effectively providing a day-ahead unit scheduling and the corresponding optimal grid topology for power flow management and system reliability. The specific contributions of this paper are as follows:

1) For the joint optimization problem of UC and OTS, we introduce a centralized QDA. This algorithm decomposes the joint optimization problem into an upper UC module that optimizes unit scheduling, a lower OTS module that addresses flow constraints and topology optimization, and a contingency verification module that considers unexpected scenarios. Iterative interactions between these modules are facilitated by adding cutting planes, ultimately achieving dynamic grid topology switching to avoid security constraint violations while reducing operational costs. Experimental results confirm the interdependence between generator scheduling and network topology, demonstrating that effective transmission line switching can significantly lower system operational costs.
2) In the upper UC module, we utilize dual information from the SP of UC to construct optimality and feasibility cutting planes based on a generalized Benders framework. In the lower OTS module, inspired by sensitivity analysis, we define the impact of transmission switching state on system operational costs or flow violations as sensitivity and construct linear optimality and feasibility cutting planes based on this analysis. We propose a method to convert these cutting planes into a QUBO model (cutting plane Hamiltonian) suitable for QC. The MP formed by these cutting plane Hamiltonians can be efficiently solved using quantum annealing algorithms on the D-Wave platform or our proprietary photonic quantum CIM, providing a viable QC paradigm for optimizing power system operations and verifying the feasibility and effectiveness of the quantum algorithm on real quantum machines for the joint optimization problem.
3) Building on the centralized algorithm, we further propose a consensus-based distributed QDA with higher solving efficiency. This algorithm uses the phase angles and interactive power of neighboring buses as consensus variables, allowing each bus to independently and in parallel solve local problems with consensus constraints. It constructs local cutting plane Hamiltonians, decoupling the MP into local MPs at the bus level. The local MPs are smaller in scale than the centralized MP, and the parallel-solving characteristic significantly enhances the solving efficiency of the joint optimization problem. The consensus-based distributed QDA demonstrates improved operational costs and solving speed in the IEEE RTS 24-Bus System compared to traditional commercial solvers like Gurobi9.

2. Joint Optimization Model of UC and OTS

The UC problem is responsible for generating a day-ahead unit scheduling that ensures both economic efficiency and system security, while the OTS aims to determine the optimal power system topology. When the grid topology is altered, the previously planned unit schedule might no longer be applicable, resulting in potential interdependencies between the grid topology and the unit dispatch plan. Additionally, the model must account for day-ahead contingency constraints to ensure the system's stability and reliability. This joint optimization model involves both continuous variables, such as unit output and line power flows, and discrete variables, like unit on/off state and transmission switching state, making it a MINLP problem, which can be formulated as follows:

$$\min_{P,U} \sum_{t=1}^{NT} \left[\sum_{i=1}^{NG} \mathcal{K}\left(P_{i,t}\right) + \mathcal{J}\left(U_{i,t}\right) + \sum_{l=1}^{NL} \mathcal{L}\left(Z_{l,t}\right) \right] \tag{1}$$

$$\mathcal{K}\left(P_{i,t}\right) = \alpha_i \cdot P_{i,t}^2 + \beta_i \cdot P_{i,t} \tag{2}$$

$$\mathcal{J}\left(U_{i,t}\right) = \gamma_i \cdot U_{i,t} \tag{3}$$

$$\mathcal{L}\left(Z_{l,t}\right) = \kappa_l \left(1 - Z_{l,t}\right) \tag{4}$$

$$s.t. \sum_{i\in NG(n)} P_{i,t} - \sum_{l\in NL_{from}(n)} \vec{P}_{l,t} + \sum_{l\in NL_{to}(n)} \vec{P}_{l,t} = D_{n,t}, \forall n, \forall t \tag{5}$$

$$P_i^{\min} \cdot U_{i,t} \cdot \tilde{U}_{i,t}^c \le P_{i,t} \le P_i^{\max} \cdot U_{i,t} \cdot \tilde{U}_{i,t}^c, \forall i, \forall t \tag{6}$$

$$\sum_{j=t-T_i^{on}}^{t-1} U_{i,j} \geq T_i^{on} U_{i,t-1} \left(1 - U_{i,t}\right), \quad \forall i, \forall t \geq T_i^{on} \tag{7}$$

$$\sum_{j=t-T_i^{off}}^{t-1} \left(1 - U_{i,j}\right) \geq T_i^{off} U_{i,t} \left(1 - U_{i,t-1}\right), \quad \forall i, \forall t \geq T_i^{off} \tag{8}$$

$$U_{i,t} \in \{0,1\}, P_{i,t} \geq 0, \quad \forall i, \forall t \tag{9}$$

$$\sum_{i=1}^{NG} \tilde{U}_i^c = NG - 1, \quad \forall c \geq 1 \tag{10}$$

$$-P_l^{\max} \cdot Z_{l,t} \cdot \tilde{Z}_{l,t}^c \leq \vec{P}_{l,t} \leq P_l^{\max} \cdot Z_{l,t} \cdot \tilde{Z}_{l,t}^c, \quad \forall l, \forall t \tag{11}$$

$$\left(\vec{\theta}_{n,t} - \vec{\theta}_{m,t}\right)/x_l - \vec{P}_{l,t} \leq M\left(2 - Z_{l,t} - \tilde{Z}_l^c\right), \quad \forall l\,(n \to m), \forall t \tag{12}$$

$$\left(\vec{\theta}_{n,t} - \vec{\theta}_{m,t}\right)/x_l - \vec{P}_{l,t} \geq -M\left(2 - Z_{l,t} - \tilde{Z}_{l,t}^c\right), \forall l\,(n \to m), \forall t \tag{13}$$

$$\sum_{l=1}^{NL} \left(1 - Z_{l,t}\right) \leq J, \quad \forall t \tag{14}$$

$$\sum_{l \in NL(n)} Z_{l,t} \geq 1, \quad \forall n, \forall t \tag{15}$$

$$Z_{l,t} \in \{0,1\}, \quad \forall l, \forall t \tag{16}$$

$$\theta_n^{\min} \leq \vec{\theta}_{n,t} \leq \theta_n^{\max}, \quad \forall n, \forall t \tag{17}$$

$$\sum_{l=1}^{NL} \tilde{Z}_l^c = NL - 1. \quad \forall c \geq 1 \tag{18}$$

Here, NG, NT, and NL represent the set of units, time periods, and transmission lines, respectively. $NG(n)$ denotes the set of units located at bus n; $NL_{from}(n)$, $NL_{to}(n)$ and $NL(n)$ represent the sets of outgoing, incoming, and total transmission lines at bus n, satisfying $NL(n) = NL_{from}(n) \cup NL_{to}(n)$; The continuous variable $P_{i,t}$ indicates the output power of unit i at time period t; $\vec{P}_{l,t}$ represents the power flow on transmission line l at time period t; $\vec{\theta}_{n,t}$ denotes the phase angle at bus n at time period t; The binary variable $U_{i,t}$ indicates the on/off state of unit i at time period t; $U_{i,t} = 1$ when the unit is online, and $U_{i,t} = 0$ when it is offline; The binary variable $Z_{l,t}$ represents the state of transmission line l at time period t; $Z_{l,t} = 1$ when the line is connected, and $Z_{l,t} = 0$ when it is disconnected; Parameters α_i, β_i, and γ_i correspond to the quadratic fuel cost coefficient, linear fuel cost coefficient, and startup cost coefficient of unit i, respectively; Parameter κ_l denotes the switching cost of transmission line l; Parameters $P_i^{\min}$ and $P_i^{\max}$ represent the minimum and maximum output power limits of unit I; The binary parameter $\tilde{U}_{i,t}^c$ indicates the state of unit i during contingency k at time period t, with $c = 0$ representing the normal operation; Parameters T_i^{on} and T_i^{off} specify the minimum up and down times for unit I; Parameter $P_l^{\max}$ represents the maximum power flow limit of transmission line l; The binary parameter $\tilde{Z}_{l,t}^c$ indicates the state of transmission line l during contingency k at time t, with $c = 0$, $\tilde{Z}_{l,t}^0 = 1$ representing the normal operation; x_l denotes the reactance of transmission line l; M is a sufficiently large constant introduced by the big-M method, ensuring $M \geq (\theta_{m,t}^{\max} - \theta_{n,t}^{\min})/x_l$.

The objective function (1) of the model aims to minimize the total cost of the power system, comprising the generation fuel cost $\mathcal{K}$ of the unit, the startup cost $\mathcal{J}$ of the unit, and the switching cost $\mathcal{L}$ of the transmission lines. The constraints of the model include the UC constraints Equations (5–10) and the OTS constraints Equations (11–18), which take into account day-ahead contingencies. Constraint (5) represents the power balance at each bus, ensuring that the total power generated by the unit at the bus plus the incoming and outgoing power flows equals the load at that bus. The unit capacity constraint (6) defines the maximum and minimum output power of the unit under contingency c. Constraints (7) and (8) enforce the minimum up and down times for the unit, ensuring the unit maintains on/off state for specified durations. Equation (9) defines the domain of the UC-related variables. Constraint (10) specifies the unit contingency constraint, which ensures that under an unusual event, at most one unit is offline. The transmission line power flow capacity constraint (11) sets the upper and lower bounds on power flow through the lines under contingencies. Power flow constraints (12) and (13) follow Kirchhoff's laws, using the DC power flow equations to model the power flows in the system. The upper bound constraint (14) limits the number of disconnected lines within the structure. The anti-islanding constraint (15) ensures that at least one transmission line connects to each bus when all corresponding lines are capable of switching. Equation (16) defines the domain of the transmission line state variables, while Equation (17) defines the domain of the bus phase angle variables. Finally, constraint (18) specifies the transmission line contingency condition, ensuring that in a contingency scenario, at most one transmission line is disconnected.

3. Consensus-Based Distributed QDA

The consensus-based distributed QDA is a distributed version of the centralized quantum decomposition approach. It introduces consensus among units or buses into the optimal SPs of UC, optimal SPs of OTS and relaxed SPs of OTS within the centralized algorithm. The primary distinction between the distributed and centralized approaches lies in the model construction and solution process of local optimization SPs, while their overall framework remains identical, as illustrated in **Figure 1** This process mainly decomposes the joint optimization problem into an upper UC module, a lower OTS module, and a contingency validation module, taking into account the interactions among these three components. The upper UC module generates the unit scheduling, which is then passed to the lower OTS module to determine the optimal grid topology that complies with safety checks while fitting the unit scheduling. If the lower OTS module consistently fails to meet safety validation criteria, a security cutting plane for the unit is constructed and returned to the upper module to adjust the unit scheduling. This iterative process gradually aligns unit scheduling and grid topology towards a near-global optimal solution. The contingency validation module acts as a reactive intelligent error correction mechanism to prevent security issues arising from N-1 contingencies. When the process returns to the lower OTS module after passing through the contingency valida-

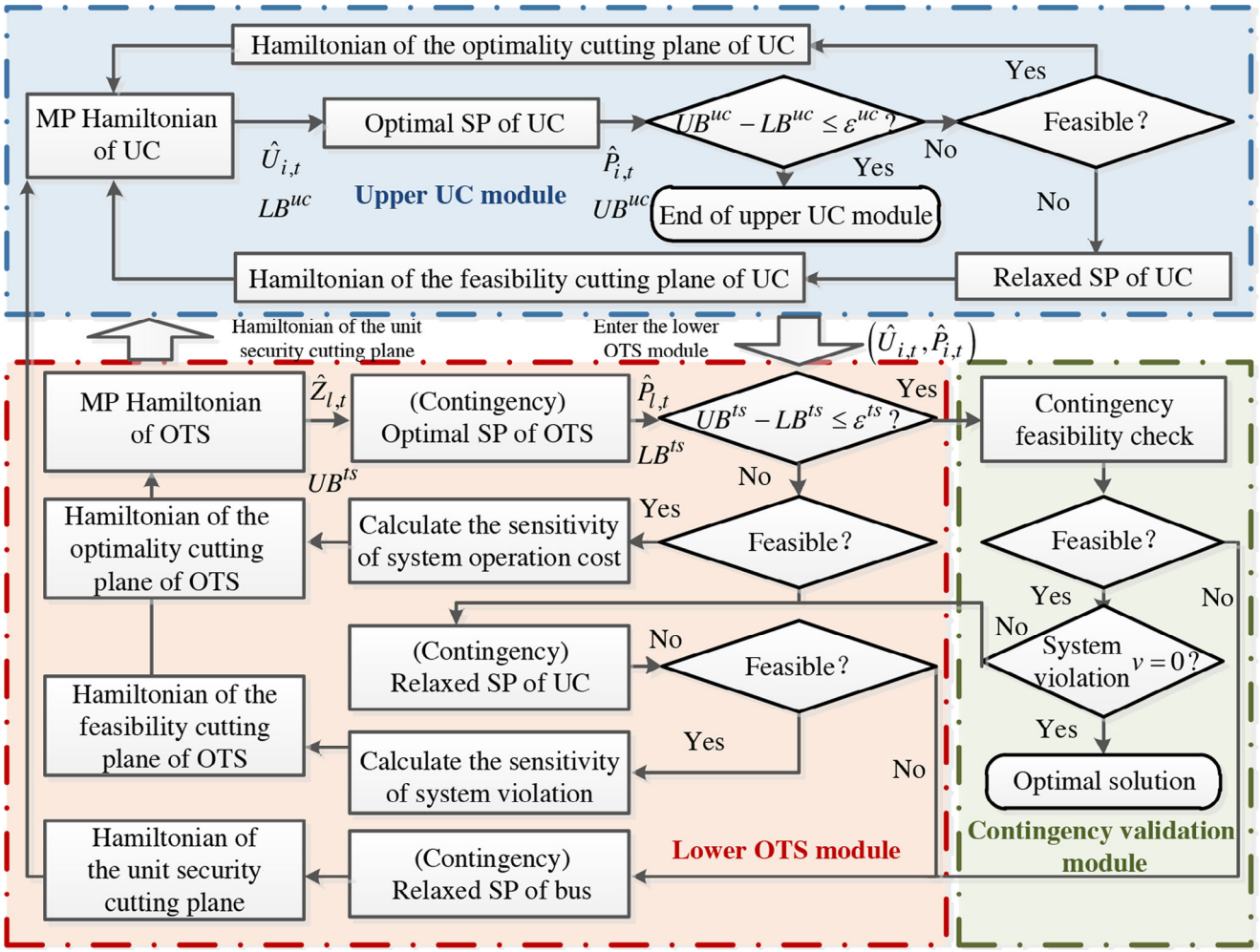

Figure 1. Flowchart of centralized or consensus-based distributed QDA.

tion module, the optimal SPs of OTS, relaxed SPs of OTS, and relaxed SPs of bus change the related constraints into a variant form considering a contingency situation, hence these SPs are prefixed with "contingency" to denote their altered formulations.

After completing the inner and outer iterative loops between the upper and lower modules, the optimal solution for unit scheduling and grid topology under normal conditions can be obtained. If N-1 contingencies are considered, this optimal solution is passed to the contingency validation module for feasibility checks. Should the solution fail the validation, the process returns to the lower-level module to adjust the grid topology through its internal iterations. At this stage, all SPs prefixed with "contingency" in the lower module will have their relevant constraints modified to account for the specific contingency scenario. If the contingency still cannot be resolved, the process returns to the upper UC module to further refine the unit scheduling. The MPs of UC and OTS modules are responsible for optimizing the binary decision variables, which are the unit on/off state and transmission line state. These are discrete optimization problems where the computational challenge arises from the combinatorial explosion of the feasible solution space as the number of discrete variables increases. The computational complexity of solving these combinatorial optimization problems grows exponentially with the number of discrete variables. Given the potential of QC to accelerate the solution of constrained combinatorial optimization problems, we reformulate the MPs of UC and OTS modules into a QUBO format Hamiltonian, making them amenable to QC techniques. This transformation allows us to leverage quantum annealers and CIMs, which possess quantum advantages, to efficiently solve these problems. To enhance the decomposition efficiency, we introduced a synchronous distributed solving strategy based on the centralized QDA mentioned above, leading to the development of the consensus-based distributed QDA. Unlike the centralized approach, this distributed algorithm employs consensus mechanisms across adjacent buses to decompose SPs into local SPs, thereby generating local cutting planes to formulate local MPs. Specifically, the consensus-based distributed QDA introduces consensus variables (inter-bus power exchanges and phase angles) into the optimal SPs of UC, optimal SPs of OTS, and relaxed SPs of OTS. This approach allows these SPs to be decoupled at the bus level. The addition of consensus constraints ensures that the consensus variables gradually converge through iterative consensus processes. The dual variables, continuous variables,

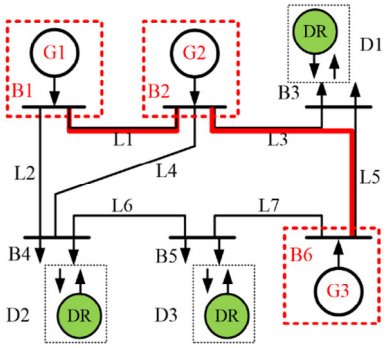

Figure 2. Simplified topology diagram of the power grid for buses with unit. The red solid lines represent the assumed adjacency relationships between the buses.

system costs, and constraint violations derived from the local SPs can be used to construct local cutting planes, further ensuring the decouplability of the MPs. As a result, the number of binary variables in the decoupled local MPs is significantly reduced, facilitating a more efficient solution process.

3.1. The Upper UC Module

The upper module is responsible for determining the day-ahead optimal unit scheduling, aiming to minimize the operational costs of the power system by optimally managing the on/off state $U_{i,t}$ and the output power $P_{i,t}$ of the unit. This module decomposes the problem based on the types of variables involved, resulting in the formulation of the local optimal SP of UC, the relaxed SP of UC, and the local MP of UC. The local optimal SP and relaxed SP focus on solving continuous variables $P_{i,t}$ and slack variables $S_{i,t}$, generating dual information that can be used to construct cutting planes. These cutting planes are then returned to the local MP, refining its feasible region to enhance the binary solution $U_{i,t}$.

1) Consensus-based local optimal SP of UC:

The local optimal SP of UC is responsible for solving continuous variables $P_{i,t}$ while simultaneously validating the feasibility of the unit on/off state $\hat{U}_{i,t}$. Therefore, this module only needs to consider buses equipped with units. Based on the potential transmission line topology, the connections between these buses are assumed, introducing the power exchange between neighboring buses as consensus variables. Taking the simplified power grid topology shown in **Figure 2** as an example, the dashed red boxes represent buses equipped with units (B1, B2, B6), while the solid red lines indicate the assumed adjacency between these buses. We focus solely on the adjacency relationships among buses that have units, decomposing the optimal SP of UC into a local optimal SP specific to each bus equipped with units. For instance, the local optimal SP at bus n is:

$$\min \sum_{i=1}^{NG(n)} \sum_{t=1}^{NT} \mathcal{K}\left(P_{i,t}\right) + \sum_{m \in \mathcal{N}_{ad}(n)} \mathcal{M}\left(\vec{P}^{n}_{n,m}, \vec{P}^{+}_{n,m}\right) \tag{19}$$

$$\mathcal{M} = \lambda^{n}_{n,m}\left(\vec{P}^{n}_{n,m} - \vec{P}^{+}_{n,m}\right) + \frac{\rho^{n}_{n,m}}{2}\left(\vec{P}^{n}_{n,m} - \vec{P}^{+}_{n,m}\right)^{2} \tag{20}$$

$$s.t. \sum_{i=1}^{NG(n)} P_{i,t} \pm \sum_{m \in \mathcal{N}_{ad}(n)} \vec{P}^{n}_{n,m} = D_{n,t}, \quad \forall t \tag{21}$$

$$P_{i}^{\min} \hat{U}_{i,t} \le P_{i,t}, \quad \forall i \in NG\left(n\right), \forall t \tag{22}$$

$$P_{i,t} \le P_{i}^{\max} \hat{U}_{i,t}. \quad \forall i \in NG\left(n\right), \forall t \tag{23}$$

Since the binary variables $U_{i,t} = \hat{U}_{i,t}$ are predetermined fixed inputs, and the topology among buses equipped with units is also assumed to be static, the local optimal SPs of UC are formulated as nonlinear programming problems. $NG(n)$ represent the set of units installed at bus n. $\lambda^{n}_{n,m}$ and $\rho^{n}_{n,m}$ denote the dual variables and penalty coefficients of the consensus constraints, respectively. The superscripts n and subscripts n, m together indicate the copies of variables associated with bus n and its neighboring bus m in the local SP of bus n. $\vec{P}^{+}_{n,m}$ is the global consensus variable, represents the power flow between buses n and m, while $\vec{P}^{n}_{n,m}$ is the local consensus variable, which is the copy of the power flow in the local SP at bus n.

The consensus constraint $\vec{P}^{n}_{n,m} - \vec{P}^{+}_{n,m} = 0$ is treated as a soft constraint in Equations (19 and 20), i.e., consensus deviation $\mathcal{M}$, and incorporated into the objective function. The objective function is then minimized to ensure that the local consensus variables $\vec{P}^{n}_{n,m}$ gradually converge to the global consensus variables $\vec{P}^{+}_{n,m}$. The power balance constraint in Equation (21) accounts for the consensus power flow, with the sign of the consensus variables $\vec{P}^{n}_{n,m}$ determined by the assumed bus connections. This sign reflects the direction of power flow between bus n and its neighboring bus m, $-\vec{P}^{n}_{n,m}$ indicating the power flow from the local bus n to its adjacent bus m, and vice versa for $+\vec{P}^{n}_{n,m}$. Since the direction of power flow is not restricted, the assumed connections between buses do not affect the solution process. Equations (22 and 23) define the unit output power capacity constraints of bus n. The update rules for the global consensus variables, the dual variables of the consensus constraints, and the penalty coefficients are as follows:

$$\vec{P}^{+}_{n,m}\left(k+1\right) = \frac{\vec{P}^{n}_{n,m}\left(k\right) + \vec{P}^{m}_{n,m}\left(k\right)}{2} \tag{24}$$

$$\lambda^{n}_{n,m}\left(k+1\right) = \lambda^{n}_{n,m}\left(k\right) + \rho^{n}_{n,m} \cdot \left(\vec{P}^{n}_{n,m}\left(k\right) - \vec{P}^{+}_{n,m}\left(k\right)\right) \tag{25}$$

$$r^{n}_{n,m}\left(k\right) = \left|\vec{P}^{n}_{n,m}\left(k\right) - \vec{P}^{+}_{n,m}\left(k\right)\right| \tag{26}$$

here, k denotes the number of consensus iterations. When the local consensus variables are sufficiently close to the global consensus variables (i.e., $r^{n}_{n,m}(k) = 0$), the neighboring buses equipped with units are considered to have reached consensus, indicating that they have sufficiently approached the global optimal solution of the optimal SP of UC. The sum of the objective functions of all local optimal SPs (Equation 19) effectively represents an upper bound on the optimal solution of the original UC problem.

If the optimal SP of UC is feasible and consensus is achieved among all neighboring buses, the optimal local dual variables $\hat{a}_{i,t}$ and $\hat{b}_{i,t}$ of constraints (22) and (23), as well as the optimal local continuous variables $\hat{P}_{i,t}$, can be obtained. These values are then

used to construct the local optimality cutting planes of UC:

$$h \geq \sum_{i=1}^{NG(n)} \sum_{t=1}^{NT} \mathcal{K}\left(\hat{P}_{i,t}\right) + \hat{a}_{i,t}\left(P_i^{\min} U_{i,t} - \hat{P}_{i,t}\right) + \hat{b}_{i,t}\left(\hat{P}_{i,t} - P_i^{\max} U_{i,t}\right)$$
$$= \sum_{i=1}^{NG(n)} \sum_{t=1}^{NT} C^{UC,op} + F_{i,t}^{UC,op} \cdot U_{i,t}, \quad \forall n \tag{27}$$

here, h represents the feasible region of the local optimality cutting plane of UC. $\hat{a}_{i,t}$ and $\hat{b}_{i,t}$ are the dual variables associated with the unit output power capacity constraints (Equations 22 and 23), while $\hat{P}_{i,t}^{op}$ denotes the solution of the continuous variables obtained from solving the local optimal SP of UC. The term $C^{UC,op}$ is a constant in the cutting plane, and $F_{e,i,t}^{UC,op}$ represents the coefficient of the unit on/off state variable $U_{i,t}$ in the cutting plane. This cutting plane reveals the optimality of the proposed solution with respect to the original joint optimal problem; therefore, we incorporate it into the local MP of UC as a local lower bound on the optimal solution of the UC problem.

2) Relaxed SP of UC:

If the optimal SP of UC is found to be infeasible, continuous slack variables ($S_{i,t}^2$, $S_{i,t}^1$) are introduced to quantify the extent of the violation for unit i concerning the power output constraint (Equation 6) during time period t. This leads to the construction of the original relaxed SP of UC. Similar to the local optimal SP of UC, the original relaxed SP of UC utilizes the average consensus method decomposition to construct the local relaxed SP of UC, which is solved independently by each bus. Taking bus n as an example, the local relaxed SP of UC is:

$$\min \mu' = \sum_{i=1}^{NG} \sum_{t=1}^{NT} \left(S_{i,t}^1 + S_{i,t}^2\right) + \sum_{m \in \mathcal{N}_{ad}(n)} \mathcal{M}\left(\vec{P}_{n,m}^n, \vec{P}_{n,m}^+\right) \tag{28}$$

$$s.t. \sum_{i=1}^{NG(n)} P_{i,t} \pm \sum_{m \in \mathcal{N}_{ad}(n)} \vec{P}_{n,m}^n = D_{n,t}, \quad \forall t \tag{29}$$

$$P_i^{\min} \hat{U}_{i,t} - P_{i,t} - S_{i,t}^1 \leq 0, \quad \forall i \in NG(n), \forall t \tag{30}$$

$$P_{i,t} - P_i^{\max} \hat{U}_{i,t} - S_{i,t}^2 \leq 0. \quad \forall i \in NG(n), \forall t \tag{31}$$

$$S_{i,t}^1, S_{i,t}^2 \geq 0, \ \forall i \in NG(n), \forall t \tag{32}$$

where $S_{i,t}^1$ are the non-negative slack variables for the unit minimum power constraint (22), and $S_{i,t}^2$ are the non-negative slack variables for the unit maximum power constraint (23). The consensus soft constraint $\mathcal{M}(\vec{P}_{n,m}^n, \vec{P}_{n,m}^+)$ in the objective function is similar to Equation (20). The update rules for the global consensus variable, dual variables, and penalty coefficient of the consensus constraint follow Equations (24–26).

To avoid the complex consensus iterations between multiple local relaxed SPs and achieve a faster optimal allocation of relaxed variables, we construct a relaxed variable objective function that reflects cost information. Additionally, the Karush-Kuhn-Tucker (KKT) conditions of the original relaxed SP are included as one of the constraints, and the constraint of relaxation variables is also considered, leading to the improved consensus-inspired relaxed SP for UC:)

$$\min \mu = \sum_{i=1}^{NG} \sum_{t=1}^{NT} \mathcal{K}\left(S_{i,t}^1\right) + \mathcal{K}\left(S_{i,t}^2\right) \tag{33}$$

$$s.t. \quad Karush - Kuhn - Tucker \ conditions, \tag{34}$$

$$S_{i,t}^2 \leq P_i^{\max}\left(1 - \hat{U}_{i,t}\right). \quad \forall i, \forall t \tag{35}$$

The objective function (33) incorporates the generation cost coefficients to achieve the minimum violation cost μ under the current unit on/off state, ensuring that the slack variables are preferentially allocated to units with lower generation costs. Since the original relaxed SP satisfies convex optimization conditions, its optimality is ensured by including the KKT conditions as constraints (34) in the consensus-inspired relaxed SP, as detailed in Appendix A. Equation (35) represents the relaxation limitation constraints, which ensure that deviations are only allocated to units in the off state, with the maximum deviation constrained by the maximum output limit.

In the UC problem, the original relaxed SP of UC often results in multiple feasible solutions for continuous variables that correspond to the same minimum violation. This occurs because the original relaxed SP only minimizes the total system violation, which means minimizing the sum of the slack variables. As a result, the corresponding single global feasibility cutting plane essentially reduces the Boolean polytope formed by the feasible region of the global binary variables. In contrast, the newly constructed consensus-inspired relaxed SP of UC utilizes the KKT conditions of the original relaxed SP as constraints to ensure the optimality of the system violation. Additionally, it incorporates an economic-oriented violation cost objective function and bounds on the slack variables in constraints, which guide the allocation of slack variables across different buses to align with the generation cost characteristics of the units. Therefore, the purpose of the entire consensus-inspired relaxed SP is to minimize the system violation while allocating violations preferentially to the power constraints of units with lower operating costs. The consensus-inspired relaxed SP leads to the construction of bus-level cutting planes, which allow for multiple reductions of the Boolean polytope, while also maintaining the decouplability of the MP of UC. Similarly, the consensus strategy in Equations (24–26) utilized in the original relaxed SP also facilitates multi-cutting and MP decoupling, so the individual local relaxed SP can be solved in parallel across multiple classical computers. However, the process of achieving consensus among the local relaxed SPs of different regions requires multiple iterations, which is both complex and time-consuming. In contrast, the improved consensus-inspired relaxed SP not only eliminates the need for multiple consensus iterations but also provides the possibility for distributed solving of the MP, due to its guiding the allocation of slack variables across different buses to align with the generation cost characteristics of the units.

Based on the solutions of the continuous variables and dual variables within the bus, we can construct local feasibility cutting

planes for the UC at each bus:

$$0 \geq \sum_{i=1}^{NG(n)} \sum_{t=1}^{NT} k_{i,t}^{1} \left(P_i^{\min} U_{i,t} - \hat{P}_{i,t}^{fea} \right) + k_{i,t}^{2} \left(\hat{P}_{i,t}^{fea} - P_i^{\max} U_{i,t} \right),$$

$$= \sum_{i=1}^{NG(n)} \sum_{t=1}^{NT} C^{UC,fea} + F_{i,t}^{UC,fea} \cdot U_{i,t}, \quad \forall n \tag{36}$$

here, $k_{i,t}^{1}$ and $k_{i,t}^{2}$ are the dual variables associated with the relaxed unit output power capacity constraints (30) and (31). $\hat{P}_{i,t}^{fea}$ denotes the solution of the continuous variables obtained by solving the improved relaxed SP of UC. $C^{UC,fea}$ is the constant term in the cutting planes, while $F_{e,i,t}^{UC,fea}$ represents the coefficient of the unit on/off state variable $U_{i,t}$ in the cutting planes. These cutting planes are incorporated into the MP of UC in the form of binary variable constraints, effectively reducing the feasible region of the binary variables and refining the unit on/off states.

3) Local MP of UC

Since both the consensus-based optimal SP of UC and the improved relaxed SP of UC return local cutting planes on the bus level, it is possible to construct a bus-level local MP of UC:

$$\min \quad \underline{Z} = h \tag{37}$$

$$s.t.\ h \geq \sum_{i=1}^{NG(n)} \sum_{t=1}^{NT} C^{UC,op} + F_{i,t}^{UC,op} \cdot U_{i,t}, \forall n, \forall t \tag{38}$$

$$0 \geq \sum_{i=1}^{NG(n)} \sum_{t=1}^{NT} C^{UC,fea} + F_{i,t}^{UC,fea} \cdot U_{i,t}, \forall n, \forall t \tag{39}$$

$$\sum_{j=t-T_i^{on}}^{t-1} U_{i,j} \geq T_i^{on} U_{i,t-1} \left(1 - U_{i,t}\right), \forall i \in NG(n), \forall t \geq T_i^{on} \tag{40}$$

$$\sum_{j=t-T_i^{off}}^{t-1} \left(1 - U_{i,j}\right) \geq T_i^{off} U_{i,t} \left(1 - U_{i,t-1}\right). \forall i \in NG(n), \forall t \geq T_i^{off} \tag{41}$$

This MP considers the constraints related to binary variables within each bus, specifically the minimum up/down constraints for the unit (Equations 40,41). Constraints (38) and (39) are concise representations of the local optimality cutting plane (27) and the feasibility cutting plane (36), respectively. As the SPs involve constraints, continuous variables, and deviations that are independent across different time periods, the cutting planes (38) and (39) naturally decouple over time periods. The local MP of UC focuses on optimizing the binary variables $U_{i,t}, i \in NG(n)$ representing the on/off state of units within each bus, providing a local lower bound $\underline{Z}$ for the objective function of the original UC problem. To address this constrained combinatorial optimization problem, we reformulate the MP into an Ising model (as discussed in Section 3.4) and employ the quantum annealer (as detailed in Section 3.5) and CIM (as detailed in Section 3.6) to accelerate the solution.

After completing the internal iterations between the local MP and SPs of UC for all buses, a unit scheduling $(\hat{U}_{i,t}, \hat{P}_{i,t})$ is obtained. This scheduling is then transmitted to the lower OTS module. If the unit scheduling does not satisfy the power flow security constraints of the transmission lines, the lower-level internal iteration in the lower OTS module generates the corresponding optimal grid topology.

3.2. The Lower OTS Module

In both centralized and distributed frameworks, the lower OTS module first evaluates whether all security constraints are met through the optimal SP of OTS. If the security constraints are satisfied, an optimality cutting plane is constructed and returned to the MP of OTS. This optimality cutting plane reveals the optimal conditions for the OTS, providing insights into minimizing the objective function of the original joint optimization problem and accelerating the convergence of the algorithm. If the security constraints are not satisfied, the relaxed SP of OTS is solved, followed by the construction of a feasibility cutting plane, which is then returned to the MP of OTS. The goal of this step is to adjust the transmission switching state to eliminate violations of the security constraints. The MP of OTS is constrained by the binary variables constraints, as well as by the optimality and feasibility cutting planes. It focuses on optimizing these binary variables $Z_{l,t}$ and providing a lower bound for the objective function of the original joint optimization problem. If the SPs yield no feasible solutions, it indicates that no suitable grid topology can be found for the given unit scheduling $\hat{U}_{i,t}$. In this case, the lower module will solve the relaxed SP of the bus and construct a unit security cutting plane to return to the upper module for correcting the unit scheduling.

Based on the above description, we will introduce the consensus-based lower OTS module suitable for distributed scenarios. Unlike the consensus-based upper UC module, the SPs in the lower module will consider consensus variables associated with neighboring buses under the actual grid topology $\hat{Z}_{l,t}$.

1) Local optimal SP of OTS

Based on the unit on/off state $\hat{U}_{i,t}$ determined by the upper module, we need to perform a security check on the current transmission switching schedule $\hat{Z}_{l,t}$. The optimal SP of OTS takes the binary variables $\hat{U}_{i,t}$ and $\hat{Z}_{l,t}$ as fixed inputs and is responsible for optimizing the continuous decision variables $p_{i,t}$ and $p_{l,t}$ in the joint optimization problem. Its mathematical formulation is as follows:

$$\min_{P} \ w_n = \sum_{t=1}^{NT} \left[\sum_{i \in NG(n)} \mathcal{J}\left(\hat{U}_{i,t}\right) + \mathcal{K}\left(P_{i,t}\right) + \sum_{l \in NL(n)} \mathcal{L}\left(Z_{l,t}\right) \right]$$
$$+ \sum_{l \in NL(n)} \mathcal{I}\left(\vec{P}_l^n, \vec{P}_l^+\right) + \sum_{m \in N_{adj}(n)} \mathcal{O}\left(\vec{\theta}_m^n, \theta_m^+\right) \tag{42}$$

$$\mathcal{I} = \upsilon_l^n \left(\vec{P}_l^n - \vec{P}_l^+\right) + \frac{\varsigma_l^n}{2} \left(\vec{P}_l^n - \vec{P}_l^+\right)^2 \tag{43}$$

$$\mathcal{O} = \gamma_m^n \left(\vec{\theta}_m^n - \theta_m^+\right) + \frac{\varphi_m^n}{2} \left(\vec{\theta}_m^n - \theta_m^+\right)^2 \tag{44}$$

$$s.t. \quad Z_{l,t} = \hat{Z}_{l,t}, \quad \forall l \in NL(n), \forall t \tag{45}$$

$$\sum_{i \in NG(n)} P_{i,t} \pm \sum_{l \in NL(n)} \vec{P}_l^n = D_{n,t}, \quad \forall t \tag{46}$$

$$P_i^{\min}\hat{U}_{i,t} \le P_{i,t}, \quad \forall i \in NG(n), \forall t \tag{47}$$

$$P_{i,t} \le P_i^{\max}\hat{U}_{i,t}. \quad \forall i \in NG(n), \forall t \tag{48}$$

$$-P_l^{\max}\cdot Z_{l,t} \le \vec{P}_l^n, \quad \forall l \in NL(n), \forall t \tag{49}$$

$$\vec{P}_l^n \le P_l^{\max}\cdot Z_{l,t}, \quad \forall l \in NL(n), \forall t \tag{50}$$

$$\left(\vec{\theta}_m^n - \vec{\theta}_n^n\right)/x_l - \vec{P}_l^n \le M\left(1 - Z_{l,t}\right), \forall l \in NL(n), \forall t \tag{51}$$

$$\left(\vec{\theta}_m^n - \vec{\theta}_n^n\right)/x_l - \vec{P}_l^n \ge -M\left(1 - Z_{l,t}\right), \forall l \in NL(n), \forall t \tag{52}$$

here, $NL(n)$ denotes the set of transmission lines associated with bus n; $N_{ad}(n)$ represents the set of neighboring buses connected to bus n; $NG(n)$ indicates the set of units located at bus n. To construct a bus-level local optimal SP of OTS, it is necessary to consider not only the power flow of transmission lines between the local bus and its neighboring buses but also the relationship between the power flow and the phase angles of the neighboring buses. Therefore, the optimal SP of OTS introduces two types of consensus variables, $\vec{P}_l^+$ and θ_m^+ are the global consensus variables, representing the power flow of transmission line l connected to bus n and phase angle of the neighboring bus m, respectively, while $\vec{P}_l^n$ and $\vec{\theta}_m^n$ are the local consensus variables, which are the copies of the power flow and the phase angle of the neighboring bus in the local optimal SP at bus n, respectively. The subscripts and superscripts together indicate that these physical quantities of the transmission line l or bus m are local copies at bus n. The power flow consensus constraint $\vec{P}_{n,m}^n - \vec{P}_{n,m}^+ = 0$ and phase angle consensus constraint $\vec{\theta}_m^n - \vec{\theta}_m^+ = 0$ are expressed as soft constraints in the form of consensus deviations, $\mathcal{I}$ and $\mathcal{O}$, and are incorporated into the objective function(42). The dual variables and penalty coefficients associated with the power flow consensus constraint are denoted as v_l^n and ς_l^n, respectively, while those related to the phase angle consensus constraint are denoted as γ_m^n and φ_m^n, respectively. The optimal SPs of OTS consider the power balance constraint (46) for the topology of bus n, unit output power capacity constraints (47) and (48), transmission line power flow capacity constraints (49) and (50), and the power flow Equations (51) and (52). The update rules for the global consensus variables, dual variables of the consensus constraints, and penalty coefficients are similar to those described in Equations (24–26). Once the local consensus variables of each bus converge, the local optimal SP of OTS approaches the global optimal solution.

If the SP yields an optimal solution for the continuous variables, the sensitivity of the system operating cost of bus n to each transmission line state is calculated. Based on this sensitivity, a local optimality cutting plane of OTS is constructed:

$$h_n \ge \hat{w}_n + \sum_{t=1}^{NT}\left[\sum_{l\in NL(n):\hat{Z}_{l,t}=1}\eta_{l,t}^{op}\left(1 - Z_{l,t}\right) + \sum_{l\in NL(n):\hat{Z}_{l,t}=0}\eta_{l,t}^{op}Z_{l,t}\right],$$
$$= \sum_{l=1}^{NL(n)}\sum_{t=1}^{NT} C^{TS,op} + F_{l,t}^{TS,op}\cdot Z_{l,t}, \forall n \tag{53}$$

here, $\hat{w}_n$ represents the operating cost of bus n under the given unit scheduling and the current grid topology (42); The sum of the operating costs of all buses provides the upper bound for the objective function of the joint optimization problem, denoted as $UB^{ts} = \sum_{n=1}^{NB}\hat{w}_n$; $\eta_{l,t}^{op}$ denotes the change in operating cost due to a state change in transmission line l during time period t, which is defined as the sensitivity of the system operating cost to the transmission line state; $C^{TS,op}$ is the constant term in the cutting plane; $F_{l,t}^{TS,op}$ is the coefficient of the transmission line state variable $Z_{l,t}$ in the cutting plane.

In the traditional Benders decomposition framework, cutting planes are typically constructed using the dual variables of the continuous variable constraints in the SP. These dual variables indicate the influence of the constraints on the objective function. However, the discrete nature of the transmission line state makes it difficult to calculate the dual variables for these SPs. Inspired by the construction of cutting planes in the Benders framework and sensitivity analysis methods, we have developed an optimality cutting plane of OTS that resembles the Benders cutting plane formulation, based on the sensitivity of the system operating cost to the transmission line state. To illustrate how to calculate $\eta_{l,t}^{op}$, we first determine the operating cost $\hat{w}_n$ of bus n under the given unit scheduling and the current grid topology. Then, we temporarily modify the state of transmission line l one by one under the current topology and recalculate the operating cost $\hat{w}_n'$ of bus n after the topology change. The resulting change in system operating cost, $\hat{w}_n' - \hat{w}_n$, due to the state change of transmission line l, is defined as the sensitivity $\eta_{l,t}^{op} = \hat{w}_n' - \hat{w}_n$ of system operating cost. If the temporary change in the transmission line state leads to an infeasible solution or forms an electrical island, the sensitivity of system operating cost to that transmission line is set to zero.

If the SP has no feasible solution, it indicates that under the determined unit scheduling, the current grid topology fails to pass the security check. In such a case, slack variables need to be introduced to construct the relaxed SP of OTS.

2) Local relaxed SP of OTS

Based on the local optimal SP of OTS, the unit power capacity constraints and the transmission line power flow capacity constraints are relaxed to formulate a local relaxed SP of OTS, which potentially yields a feasible solution. Similar to the local optimal SP of OTS, this relaxed SP uses the binary variables $\hat{U}_{i,t}$ and $\hat{Z}_{l,t}$ as fixed inputs and introduces two types of consensus variables, $\vec{P}_l^n$ and $\vec{\theta}_m^n$, along with their corresponding soft consensus constraints. Under the specified unit and transmission line state, this SP is responsible for minimizing the local system violation. Its mathematical formulation is expressed as follows:

$$\min v_n = \sum_{t=1}^{NT}\left[\sum_{i\in NG(n)} S_{i,t}^- + S_{i,t}^+ + \sum_{l\in NL(n)} F_{l,t}^- + F_{l,t}^+\right]$$
$$+ \sum_{l\in NL(n)}\mathcal{I}\left(\vec{P}_l^n, \vec{P}_l^+\right) + \sum_{m\in\mathcal{N}_{ad}(n)}\mathcal{O}\left(\vec{\theta}_m^n, \theta_m^+\right) \tag{54}$$

$$\mathcal{I} = v_l^n\left(\vec{P}_l^n - \vec{P}_l^+\right) + \frac{\varsigma_l^n}{2}\left(\vec{P}_l^n - \vec{P}_l^+\right)^2 \tag{55}$$

$$\mathcal{O} = \gamma_m^n\left(\vec{\theta}_m^n - \theta_m^+\right) + \frac{\varphi_m^n}{2}\left(\vec{\theta}_m^n - \theta_m^+\right)^2 \tag{56}$$

$$s.t. \quad Z_{l,t} = \hat{Z}_{l,t}, \quad \forall l \in NL(n), \forall t \tag{57}$$

$$\sum_{i \in NG(n)} P_{i,t} \pm \sum_{l \in NL(n)} \vec{P}_l^n = D_{n,t}, \quad \forall t \tag{58}$$

$$P_i^{\min} \hat{U}_{i,t} \le P_{i,t} + S_{i,t}^-, \quad \forall i \in NG(n), \forall t \tag{59}$$

$$P_{i,t} \le P_i^{\max} \hat{U}_{i,t} + S_{i,t}^+, \quad \forall i \in NG(n), \forall t \tag{60}$$

$$-P_l^{\max} \cdot Z_{l,t} \le \vec{P}_l^n + F_{l,t}^-, \quad \forall l \in NL(n), \forall t \tag{61}$$

$$\vec{P}_l^n \le P_l^{\max} \cdot Z_{l,t} + F_{l,t}^+, \quad \forall l \in NL(n), \forall t \tag{62}$$

$$\left(\vec{\theta}_m^n - \vec{\theta}_n^n\right)/x_l - \vec{P}_l^n \le M\left(1 - Z_{l,t}\right), \forall l \in NL(n), \forall t \tag{63}$$

$$\left(\vec{\theta}_m^n - \vec{\theta}_n^n\right)/x_l - \vec{P}_l^n \ge -M\left(1 - Z_{l,t}\right), \forall l \in NL(n), \forall t \tag{64}$$

$$0 \le S_{i,t}^-, S_{i,t}^+ \quad , \forall i \in NG(n), \forall t \tag{65}$$

$$0 \le F_{l,t}^-, F_{l,t}^+ \quad , \forall l \in NL(n), \forall t \tag{66}$$

here the unit capacity slack variables, $S_{i,t}^-$ and $S_{i,t}^+$, represent the downward and upward relaxed power of unit i at time period t, respectively; The transmission line capacity slack variables, $F_{l,t}^-$ and $F_{l,t}^+$, represent the downward and upward relaxed power flow of transmission line l at time period t, respectively; The sum of all slack variables represents the local system violation v_n at bus n; Equation (58) represents the local power balance constraint for bus n; Equations (59) and (60) are the relaxed unit power capacity constraints for bus n, while Equations (61) and (62) are the relaxed transmission line power flow capacity constraints; Equations (63) and (64) define the power flow constraints for bus n; Finally, Equations (65) and (66) limit the range of the transmission line relaxation variables for bus n. The update rules for global consensus variables, dual variables of consensus constraints, and penalty coefficients are similar to those described in Equations (24–26).

If the relaxed SP of OTS has a solution, it indicates that a suitable grid topology can be constructed for the unit scheduling through transmission switching. Similar to the approach used in the local optimal SP of OTS, the sensitivity of system violation of bus n to each transmission line state can be computed as $\eta_{l,t}^{fea} = \hat{v}_n' - \hat{v}_n$. This sensitivity can then be used to construct the local feasibility cutting plane of OTS:

$$0 \ge \hat{v}_n + \sum_{t=1}^{NT} \left[\sum_{l \in NL(n): \hat{Z}_{l,t}=1} \eta_{l,t}^{fea}\left(1 - Z_{l,t}\right) + \sum_{l \in NL(n): \hat{Z}_{l,t}=0} \eta_{l,t}^{fea} Z_{l,t} \right],$$
$$= \sum_{l=1}^{NL(n)} \sum_{t=1}^{NT} C^{TS,fea} + F_{i,t}^{TS,fea} \cdot Z_{l,t}, \forall n \tag{67}$$

here, $\eta_{l,t}^{fea}$ represents the sensitivity of system violation to the transmission line l at time period t. $C^{TS,fea}$ is the constant term in the cutting plane; $F_{i,t}^{TS,fea}$ is the coefficient associated with the transmission line state variables $Z_{l,t}$ in the cutting plane.

If the relaxed SP of OTS still has no solution, it indicates that there is no feasible grid topology for the unit scheduling provided by the upper UC module. In such cases, it is necessary to return to the MP of UC in the upper UC module to adjust the unit scheduling.

3) Relaxed SP of the bus

The relaxed SP of the bus ensures the feasibility by relaxing the bus power balance constraints. It does so by fixing all transmission lines to be in a connected state while assigning the unit on/off state $\hat{U}_{i,t}$ provided by the upper UC module to the variables $U_{i,t}$ as equality constraints and calculating the corresponding Lagrange multipliers $\hat{\lambda}_{i,t}$. The relaxation SP of bus computes the minimum power balance overload under the fully connected state and the upper-level unit dispatch scheme, as expressed below:

$$\min \quad \varpi = \sum_{t=1}^{NT} \sum_{n=1}^{NB} R_{n,t}^- + R_{n,t}^+ \tag{68}$$

$$s.t. \quad U_{i,t} = \hat{U}_{i,t}, \quad \forall i, \forall t \tag{69}$$

$$\sum_{i \in NG(n)} P_{i,t} - \sum_{l \in NL_{from}(n)} P_{l,t} + \sum_{l \in NL_{to}(n)} P_{l,t} - R_{n,t}^- + R_{n,t}^+ = D_{n,t}, \quad \forall n \tag{70}$$

$$P_i^{\min} U_{i,t} \le P_{i,t} \le P_i^{\max} U_{i,t}, \quad \forall i, \forall t \tag{71}$$

$$-P_l^{\max} \le \vec{P}_l \le P_l^{\max}, \quad \forall l, \forall t \tag{72}$$

$$\left(\vec{\theta}_m - \vec{\theta}_n\right)/x_l - \vec{P}_l = 0, \quad \forall l, \forall t \tag{73}$$

$$R_{n,t}^-, R_{n,t}^+ \ge 0, \quad \forall n, \forall t \tag{74}$$

here, ϖ represents the minimum system violation of the bus power balance constraint; The bus power slack variables $R_{n,t}^-$ and $R_{n,t}^+$ represent the surplus and deficit relaxed power of bus n at time period t, respectively; Constraint (69) assigns the unit on/off continuous variables to the unit scheduling for this iteration; Constraint (70) denotes the bus power balance constraint after incorporating the relaxed power; Constraint (71) represent the unit power capacity constraints, while constraint (72) correspond to the transmission line capacity constraints in a fully connected state; Constraint (73) ensures the power flow consistency under full connectivity; Constraint (74) imposes the non-negativity of these two types of slack variables, ensuring, together with the objective function (68), that at any given bus n and time period t, at least one of the slack variables $R_{n,t}^-$ or $R_{n,t}^+$ equals zero, that is, that only one state—either power surplus or deficit—can occur.

Solving the above expression yields the minimum violation of the power balance constraint $\hat{\varpi}$ and the Lagrange multiplier $\hat{\lambda}_{i,t}$ corresponding to the unit dispatch constraint (69). This allows us to construct the unit security cutting plane, which is then returned through the external loop and added to the MP of UC in the upper module to refine the unit scheduling:

$$0 \ge \hat{\varpi} + \sum_{i=1}^{NG} \sum_{t=1}^{NT} \hat{\lambda}_{i,t}\left(U_{i,t} - \hat{U}_{i,t}\right) \tag{75}$$

4) Local MP of OTS

The local MP of OTS is responsible for determining the binary state variables $Z_{l,t}, l \in NL(n)$ of the transmission lines connected to bus n. It provides the optimal partial grid topology that matches the upper unit scheduling and gives a local lower bound of the joint optimization problem's objective function. This MP can be expressed as follows:

$$\min \quad h_n + \sum_{l=1}^{NL(n)} \sum_{t=1}^{NT} \mathcal{L}\left(Z_{l,t}\right) \tag{76}$$

$$s.t. \quad \sum_{l \in NL(n)} Z_{l,t} \geq 1, \quad \forall t \tag{77}$$

$$h_n \geq \sum_{l=1}^{NL(n)} \sum_{t=1}^{NT} C^{TS,op} + F_{l,t}^{TS,op} \cdot Z_{l,t}\ , \tag{78}$$

$$0 \geq \sum_{l=1}^{NL(n)} \sum_{t=1}^{NT} C^{TS,fea} + F_{i,t}^{TS,fea} \cdot Z_{l,t} \tag{79}$$

here the anti-islanding constraints (77) are related to the binary variables; Constraints (78) and (79) are simplified representations of the local optimality cutting planes (53) and feasibility cutting planes (67) of OTS, respectively; The decision variable is $Z_{l,t}$, while the local operating cost $\hat{w}_n$, the system violation $\hat{v}_n$, and the sensitivities of the transmission lines to operating cost $\eta_{l,t}^{op}$ and system violation $\eta_{l,t}^{fea}$ are hyperparameters input to the MP. Since in the relaxed SP of OTS, the objective function and constraints are independent across time periods, the feasibility of cutting planes generated from this SP can naturally decouple over these time periods.

3.3. Contingency Validation Module

The contingency validation module is responsible for incorporating parameters $\tilde{U}_{i,t}^c$, $\tilde{Z}_{l,t}^c$ that represent N-1 contingencies into the relevant constraints. As depicted in the flowchart in Figure 1, the process begins with a contingency feasibility check. If no feasible solution is identified during this check, it indicates that it is impossible to prevent violations of security constraints under the given contingency by merely transmission switching. In such cases, it becomes necessary to construct a unit security cutting plane for the contingency relaxed SP of the bus and return it to the upper UC module to adjust the unit scheduling corresponding to the contingency. If a feasible solution does exist, the next step is to assess whether the system violation under the contingency scenario is zero. If zero violations are detected, it signifies that the current unit scheduling and grid topology remain valid under the contingency, and the optimal solution is then output. If violations are present, the process proceeds to the contingency relaxed SP of OTS in the lower OTS module, where the grid topology is recalibrated through the internal iteration of that module.

The contingency feasibility check takes the optimal UC solution ($\hat{U}_{i,t}$, $\hat{P}_{i,t}$) and corresponding grid topology ($\hat{Z}_{l,t}$, $\hat{P}_{l,t}$) under normal conditions as fixed inputs while accounting for the constraints imposed by the unexpected contingency. This process is mathematically formulated as follows:

$$\min \quad v_c = \sum_{t=1}^{NT} \left[\sum_{i \in NG} S_{i,t}^- + S_{i,t}^+ + \sum_{l \in NL} F_{l,t}^- + F_{l,t}^+ \right] \tag{80}$$

$$s.t. \quad \sum_{i \in NG(n)} P_{i,t} \pm \sum_{l \in NL(n)} \vec{P}_l^n = D_{n,t}, \quad \forall n, \forall t \tag{81}$$

$$P_i^{\min} \cdot \hat{U}_{i,t} \cdot \tilde{U}_{i,t}^c \leq P_{i,t} + S_{i,t}^-, \quad \forall i, \forall t \tag{82}$$

$$P_{i,t} \leq P_i^{\max} \cdot \hat{U}_{i,t} \cdot \tilde{U}_{i,t}^c + S_{i,t}^+, \quad \forall i, \forall t \tag{83}$$

$$-P_l^{\max} \cdot \hat{Z}_{l,t} \cdot \tilde{Z}_{l,t}^c \leq P_{l,t} + F_{l,t}^-, \quad \forall l, \forall t \tag{84}$$

$$P_{l,t} \leq P_l^{\max} \cdot \hat{Z}_{l,t} \cdot \tilde{Z}_{l,t}^c + F_{l,t}^+, \quad \forall l, \forall t \tag{85}$$

$$\left(\vec{\theta}_m^n - \vec{\theta}_n^n\right)/x_l - \vec{P}_l^n \leq M\left(2 - \hat{Z}_{l,t} - \tilde{Z}_{l,t}^c\right), \forall l, \forall t \tag{86}$$

$$\left(\vec{\theta}_m^n - \vec{\theta}_n^n\right)/x_l - \vec{P}_l^n \geq M\left(\hat{Z}_{l,t} + \tilde{Z}_{l,t}^c - 2\right), \forall l, \forall t \tag{87}$$

$$0 \leq S_{i,t}^-, S_{i,t}^+ \quad , \forall i, \forall t \tag{88}$$

$$0 \leq F_{l,t}^-, F_{l,t}^+ \quad , \forall l, \forall t \tag{89}$$

Here Equation (81) represents the power balance constraints for each bus; Equations (82) and (83) define the relaxed unit power capacity constraints under contingencies; Equations (84) and (85) describe the relaxed transmission line power flow capacity constraints during a contingency; Equations (86) and (87) correspond to the power flow constraints under contingencies; Finally, Equations (88) and (89) restrict the value ranges for the slack variables of unit and transmission lines. The contingency feasibility check computes the relaxed unit power, denoted as $S_{i,t}^-$, $S_{i,t}^+$, and the relaxed transmission line power flow, denoted as $F_{l,t}^-$, $F_{l,t}^+$, under contingencies c. Summing these slack variables provides the system violation v_c under the contingency scenario c.

As illustrated in Figure 1, there are three potential cases of the contingency feasibility check: (1) If a feasible solution exists and the objective function (80) representing the system violation v_c under the contingency equals zero, it indicates that the contingency feasibility check has been passed, and the optimal unit scheduling and grid topology derived under normal conditions remain applicable to the contingency scenario; (2) If a feasible solution exists but the system violation is non-zero, the process moves to the consensus-based local relaxation SP of OTS in the lower OTS module, where the constraints are adjusted to reflect the contingency conditions. Specifically, the unit power capacity constraints (59) and (60) are replaced by (82) and (83), and the transmission line power flow capacity constraints (61) and (62) are modified to (84) and (85), and the power flow constraints (63) and (64) are adjusted to (86) and (87). The lower OTS module's internal iteration process is then re-executed to recalibrate the grid topology under the contingency; (3) If no feasible solution is found, a unit security cutting plane is constructed, which is then passed back to the upper UC module to adjust the unit scheduling under the contingency. This iterative process between the upper and lower modules, involving both internal and external loops (identical to the procedures under normal conditions but

with the relevant constraints modified for the contingency scenario), continues until an optimal solution satisfying the contingency feasibility check is obtained. In summary, the contingency validation module aims to intelligently correct the optimal solution derived from normal conditions into the optimal response for post-contingency scenarios.

3.4. Optimization Method of MP Hamiltonian Based on QC

The fundamental unit of a quantum computer is the qubit, which can exist in a superposition of states $|0\rangle$ and $|1\rangle$. A quantum system composed of multiple qubits can also be prepared in an entangled state. For instance, a two-qubit system may exist in a superposition of four entangled states, namely $|0\rangle|0\rangle$, $|0\rangle|1\rangle$, $|1\rangle|0\rangle$, and $|1\rangle|1\rangle$. Therefore, a quantum system consisting of n qubits can potentially be in a superposition of quantum states. Unlike classical computing, QC leverages the properties of superposition and entanglement, allowing computational tasks to be processed in a parallel manner. To harness these quantum features for accelerating the solution of specific problems, this section systematically reformulates the cutting planes in the MP of UC or OTS into a QUBO form of the Hamiltonian. Additionally, this section proposes a novel method for transforming unit minimum up/down constraints into a QUBO form Hamiltonian to serve as a penalty term. This approach algebraically handles the unit minimum up/down constraints, avoiding the introduction of auxiliary variables and thereby significantly reducing the size of the QUBO matrix for the MP.

3.4.1. Construction Method of Cutting Plane QUBO Hamiltonians

The MPs of UC and OTS consist of different cutting planes. To construct the QUBO Hamiltonian for the MP, it is necessary to systematically reformulate each component. Specifically, this involves the optimality cutting plane of UC (38), the feasibility cutting plane of UC (39), and the unit security cutting plane of UC (75) from the MP of UC, as well as the optimality cutting plane of OTS (78) and the feasibility cutting plane of OTS (79) from the MP of OTS, into the QUBO model.

The mathematical model of the QUBO Hamiltonian for the optimality cutting plane of UC is formulated as follows:

$$H_n^{UC,op} = \sum_{i=1}^{NG(n)} \sum_{t=1}^{NT} \left[\sum_{e=1}^{NE^{UC}} F_{e,i,t}^{UC,op} + \xi_n^{UC,op} \left(C_e^{UC,op} - UB_e^{UC} \right) F_{e,i,t}^{UC,op} \right] \cdot U_{i,t} + \sum_{i=1}^{NG(n)} \sum_{t=1}^{NT} \sum_{i'=1}^{NG(n)} \sum_{t'=1}^{NT} \left(\sum_{e=1}^{NE^{UC}} F_{e,i,t}^{UC,op} F_{e,i',t'}^{UC,op} \right) U_{i,t} U_{i',t'} \quad (90)$$

Here, $H_n^{UC,op}$ denotes the Hamiltonian of the optimality cutting plane of UC for bus n; NE^{UC} represents the internal iteration count of the upper UC module; $C_e^{UC,op}$ is the constant term of the optimality cutting plane of UC during the e-th iteration; $F_{e,i,t}^{UC,op}$ refers to the coefficient of the binary variable $U_{i,t}$ in the optimality cutting plane for the e-th iteration; UB_e^{UC} signifies the upper bound of the objective function in the upper UC module for the e-th iteration; $\xi_n^{UC,op}$ is the penalty term coefficient introduced during the conversion of the optimality cutting plane of CU into the QUBO form.

The mathematical model of the QUBO Hamiltonian of the feasibility cutting plane of UC is expressed as follows:

$$H_n^{UC,fea} = \xi_n^{UC,fea} \sum_{i=1}^{NG(n)} \sum_{t=1}^{NT} \times \left[\sum_{e=1}^{NE^{UC}} \left(C_e^{UC,fea} + F_{e,i,t}^{UC,fea} \cdot U_{i,t} + \sum_{\varphi=1}^{N\varphi^{UC}} 2^{\varphi} S_{\varphi}^{UC,fea} \right) \right]^2 \quad (91)$$

Here, $H_n^{UC,op}$ represents the Hamiltonian of the feasibility cutting plane of UC for bus n; $N\varphi^{UC}$ denotes the number of auxiliary binary variables required for the feasibility cutting plane; $S_{\varphi}^{UC,fea}$ refers to the auxiliary binary variables introduced for the feasibility cutting plane of UC; $C_e^{UC,fea}$ is the constant term of the feasibility cutting plane of UC during the e-th iteration; $F_{e,i,t}^{UC,fea}$ represents the coefficient of variable $U_{i,t}$ in the feasibility cutting plane during the e-th iteration.

The mathematical model of the QUBO Hamiltonian of the optimality cutting plane of OTS is expressed as follows:

$$H_n^{TS,op} = \sum_{l=1}^{NL(n)} \sum_{t=1}^{NT} \left[\sum_{e=1}^{NE^{TS}} F_{e,l,t}^{TS,op} + \xi_n^{TS,op} \left(C_e^{TS,op} - UB_e^{TS} \right) F_{e,l,t}^{TS,op} \right] \cdot Z_{l,t} + \sum_{l=1}^{NL(n)} \sum_{t=1}^{NT} \sum_{l'=1}^{NL(n)} \sum_{t'=1}^{NT} \left(\sum_{e=1}^{NE^{TS}} F_{e,l,t}^{TS,op} F_{e,l',t'}^{TS,op} \right) Z_{l,t} Z_{l',t'} \quad (92)$$

Here, $H_n^{TS,op}$ denotes the Hamiltonian of the optimality cutting plane of OTS at bus n; NE^{TS} represents the internal iteration count of the lower OTS module; $C_e^{TS,op}$ is the constant term of the optimality cutting plane of OTS during the e-th iteration; $F_{e,l,t}^{TS,op}$ refers to the coefficient of binary variable $Z_{l,t}$ in the optimality cutting plane during the e-th iteration; UB_e^{TS} signifies the upper bound of the objective function in the lower OTS module for the e-th iteration; $\xi_n^{TS,op}$ is the penalty term coefficient introduced during the conversion of the optimality cutting plane of OTS into the QUBO form.

The mathematical model of the QUBO Hamiltonian of the feasibility cutting plane of OTS is expressed as follows:

$$H_n^{TS,fea} = \xi_n^{TS,fea} \sum_{L=1}^{NL(n)} \sum_{t=1}^{NT} \times \left[\sum_{e=1}^{NE^{TS}} \left(C_e^{TS,fea} + F_{e,l,t}^{TS,fea} \cdot Z_{l,t} + \sum_{\varphi=1}^{N\varphi^{TS}} 2^{\varphi} S_{\varphi}^{TS,fea} \right) \right]^2 \quad (93)$$

here, $H_n^{UC,op}$ represents the Hamiltonian of the feasibility cutting plane of OTS at bus n; $N\varphi^{TS}$ denotes the number of auxiliary binary variables required for the feasibility cutting plane; $S_{\varphi}^{TS,fea}$ refers to the auxiliary binary variables introduced for the feasibility cutting plane of OTS; $C_e^{TS,fea}$ is the constant term of the feasibility cutting plane during the e-th iteration; $F_{e,l,t}^{TS,fea}$ represents

the coefficient of the binary variable $Z_{l,t}$ in the feasibility cutting plane for the e-th iteration; $\xi_n^{TS,fea}$ is the penalty term coefficient introduced during the conversion of the feasibility cutting plane of OTS into the QUBO form.

The mathematical model of the QUBO Hamiltonian of the unit security cutting plane is expressed as follows:

$$H^{SE} = \xi^{SE}\sum_{i=1}^{NG}\sum_{t=1}^{NT}\left[\left(\hat{\varpi} - \lambda_{i,t}\hat{U}_{i,t}\right) + \lambda_{i,t}U_{i,t} + \sum_{\varphi=1}^{N\varphi^{SE}} 2^{\varphi} S_{\varphi}^{SE}\right]^2 \quad (94)$$

Here, H^{SE} denotes the Hamiltonian of the security cutting plane of UC; $N\varphi^{SE}$ represents the number of auxiliary binary variables required for the security cutting plane; S_{φ}^{SE} refers to the auxiliary binary variables introduced for the security cutting plane; ξ^{SE} is the penalty term coefficient introduced during the conversion of the security cutting plane into the QUBO form.

3.4.2. The Improved QUBO Hamiltonian of the Unit Minimum Up/Down Constraints

The unit minimum up/down constraints (Equations 7 and 8) are complex binary variable constraints that couple multiple time periods and are responsible for preventing abrupt changes in the unit's operational state. When considering these constraints, the start-up and shut-down decisions for a unit are influenced not only by the current period's demand but also by the state of preceding and succeeding periods. This temporal dependency introduces strong inter-period coupling, significantly increasing the complexity of decision-making. Fundamentally, the minimum up/down constraints are discrete, nonlinear inequality constraints. Such constraints complicate model optimization, especially in large-scale models with many variables, where the demands for computational time and resources increase dramatically.

In the joint optimization problem of UC and OTS over all time periods T, $2 * T$ unit minimum up/down constraints are introduced. In traditional penalty construction methods, a set of auxiliary binary variables is needed to equivalently transform these constraints into nonlinear equality constraints. These equality constraints are then squared to serve as penalty terms, but this approach introduces a large number of auxiliary variables. Thus, there is an urgent need for a penalty construction method for unit minimum up/down constraints that does not rely on auxiliary variables.

In constraints (7) and (8), the number of binary variables involved in each term is $T_i^{on} + 2$ and $T_i^{off} + 2$, respectively. To reduce this number, the complex unit minimum up/down constraints are decomposed into simpler constraints for each time period. The mathematical model is as follows:

$$U_{i,t}\left(1 - U_{i,t-1}\right) \le U_{i,\tau}, \tau \in \left[t+1, \min\left\{t + T_i^{on}, NT\right\}\right] \quad (95)$$

$$U_{i,t-1}\left(1 - U_{i,t}\right) \le 1 - U_{i,\tau}, \tau \in \left[t+1, \min\left\{t + T_i^{off}, NT\right\}\right] \quad (96)$$

Based on the unit minimum up/down constraints, the above formulation decomposes each original constraint into T_i^{on} and T_i^{off} constraints, respectively. This approach increases the number of minimum up/down constraints while reducing the number of binary variables in each constraint to three, thereby decreasing the complexity of each constraint. Constraint (95) indicates that if unit i starts up at period t, subsequent constraints are applied to ensure that unit i remains in the "on" state during the following $t + T_i^{on}$ periods. Similarly, constraint (96) enforces the minimum downtime requirement by applying $t + T_i^{off}$ constraints, ensuring that the unit remains "off" for the specified minimum number of periods.

Table 1. Truth table of the QUBO Hamiltonian of unit minimum up/down constraints.

H^{QT}	0	0	$2\xi^{QT}$	0	0	$2\xi^{QT}$	0	0
$U_{i,t-1}$	0	0	0	0	1	1	1	1
$U_{i,t}$	0	0	1	1	0	0	1	1
$U_{i,\tau}$	0	1	0	1	0	1	0	1

After simplifying the unit minimum up/down constraints, we further couple constraints (95) and (96) into a unified formulation, using $-U_{i,t} + U_{i,t-1}$ to represent the state transition of the unit. Specifically, if unit i shuts down at period t, then $-U_{i,t} + U_{i,t-1} = 1$; if unit i starts up at period t, then $-U_{i,t} + U_{i,t-1} = 0$; and if the state of the unit remains unchanged, then $-U_{i,t} + U_{i,t-1} = 0$. Given that the expected values of $U_{i,\tau}$ are either 0 or 1, we can construct a penalty term for period τ such that: when the unit is shutting down in the current period, make $U_{i,\tau} = 0$; when the unit is starting up in the current period, make $U_{i,\tau} = 1$; and when the unit's state does not change, no specific penalty is applied. This approach leads to the development of a QUBO Hamiltonian that satisfies the unit minimum up/down constraints without the need for auxiliary variables:

$$H^{QT} = \xi^{QT}\left[U_{i,\tau} + \left(-U_{i,t} + U_{i,t-1}\right) - 0.5\right]^2 \quad (97)$$

Here, H^{QT} represents the Hamiltonian of the unit minimum up/down constraints; ξ^{QT} denotes the penalty term coefficient introduced during the conversion of the minimum up/down constraints into the QUBO form. To simplify the above expression, we consider the properties of the binary variables $U_{i,t-1}$, $U_{i,t}$, and $U_{i,\tau}$, where $(U_{i,t-1})^2 = U_{i,t-1}$, $(U_{i,\tau})^2 = U_{i,\tau}$ and $(U_{i,t})^2 = U_{i,t}$. By neglecting the constant terms, we derive the simplified Hamiltonian of the unit minimum up/down constraints as follows:

$$H^{QT} = 2\xi^{QT}\left(U_{i,t} - U_{i,t-1}U_{i,t} + U_{i,t-1}U_{i,\tau} - U_{i,t}U_{i,\tau}\right).\tau$$
$$\in \left[t+1, \min\left\{t + T_i^{on}, t + T_i^{off}, NT\right\}\right] \quad (98)$$

If unit i performs a start-up or shut-down operation at period t, the QUBO Hamiltonian for the unit minimum up/down constraints is applied to each subsequent period up to $t + T_i^{on}(T_i^{off})$. This penalizes states that violate the unit minimum up/down constraints, with the Hamiltonian's truth values illustrated in **Table 1**. For any $U_{i,t-1}$, $U_{i,t}$ and $U_{i,\tau}$, $\tau \in [t+1, \min\{t + T_i^{on}, t + T_i^{off}, NT\}]$, there are a total of eight possible combinations. If $U_{i,t-1}$ and $U_{i,t}$ differ, it indicates a change in state at period t. In such cases, a penalty term should be added to

ensure that the state at subsequent period τ remains consistent with the changed state, preserving the minimum up/down constraints requirement. In Table 1, the third and sixth states violate the minimum up/down constraints. After the state changes, the unit fails to maintain its current state at time τ, resulting in the Hamiltonian H^{QT} being penalized, that is, the system is in an excited state with higher energy $2\xi^{QT}$. For the other unit state combinations that comply with the minimum up/down constraints, the corresponding Hamiltonian H^{QT} is not penalized, and the system remains in its ground state.

3.4.3. Optimization Method of Hamiltonian of MP

By summing the Hamiltonian of the cutting planes and the improved Hamiltonian of the unit minimum up/down constraints, we obtain a QUBO form Hamiltonian that is equivalent to the MP of UC (Equations 37–41) and OTS (Equations 76–79):

$$H^{UC} = H^{UC,op} + H^{UC,fea} + H^{SE} + H^{QT} \tag{99}$$

$$H^{TS} = \sum_{l=1}^{NL(n)} \sum_{t=1}^{NT} \mathcal{L}\left(Z_{l,t}\right) + H^{TS,op} + H^{TS,fea} \tag{100}$$

This QUBO form MP Hamiltonian can be easily transformed into an Ising model suitable for quantum computation. The Ising model can be efficiently mapped onto the hardware of quantum annealers and CIMs, enabling fast and efficient problem-solving. The objective of solving this model is to find the ground state of the Ising Hamiltonian, which corresponds to the optimal state of the power system with minimal operational cost. Moreover, the expected value of each qubit $1/2(I - \hat{\sigma}_z^{(i)})$ provides the solution to the unit on/off states as well as the states of the transmission lines.

The QUBO form MP of UC and OTS both adhere to a standard structure:

$$H = \sum_{(i,j)} \rho_{i,j} X_i X_j + \sum_i v_i\, X_i + a \tag{101}$$

the Ising model can be defined as an undirected graph $G(L, K)$, where L represents the set of vertices (qubits), and K represents the set of edges that indicate possible interactions between the qubits. The Hamiltonian of the Ising model is expressed as:

$$H_p = \sum_{(i,j)} k_{i,j} \hat{\sigma}_z^{(i)} \hat{\sigma}_z^{(j)} + \sum_i h_i\, \hat{\sigma}_z^{(i)} \tag{102}$$

where $\hat{\sigma}_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ is the Pauli-Z operator acting on the i-th qubit, the first term in the Hamiltonian H_p represents the interaction between qubits, while the second term accounts for the interaction between the qubits and the external magnetic field.

Based on the mapping relationship between the QUBO variables and the Pauli-Z operator:

$$Z_i \Leftrightarrow \frac{1}{2}(I - \hat{\sigma}_z^{(i)}) \tag{103}$$

the QUBO form MPs of UC and OTS can be transformed into the Ising Hamiltonian H_p. When the spin of the qubit i collapses to $|0\rangle$ (i.e., $|i\rangle = |0\rangle = \binom{1}{0}$) upon measurement, we have $\langle i|\hat{\sigma}_z^{(i)}|i\rangle = 1$. This implies that the expected value of the operator on the right-hand side of Equation (103) is $\langle i|\frac{1}{2}(I - \hat{\sigma}_z^{(i)})|i\rangle = 0$, indicating that the corresponding classical variable takes the value $Z_i = 0$ (i.e., representing the unit being "off" or the transmission line being "disconnected"). Conversely, when the spin of the qubit i collapses to $|1\rangle$ (i.e., $|i\rangle = |1\rangle = \binom{0}{1}$), the expected value of the operator on the right-hand side becomes $\langle i|\frac{1}{2}(I - \hat{\sigma}_z^{(i)})|i\rangle = 1$, indicating that the classical variable takes the value $Z_i = 1$ (i.e., representing the unit being "on" or the transmission line being "connected").

3.5. Model Calculation Based on Quantum Annealing

Quantum annealing algorithms search for the expected value of a user-defined problem Hamiltonian by adiabatically evolving the initial Hamiltonian of the system, ensuring that the system remains close to the ground state of the quantum system throughout the process. At the end of the process, each qubit collapses from a superposition state to either $|0\rangle$ or $|1\rangle$ upon measurement, yielding the classical solution to the problem represented by the Hamiltonian.

The process of controlling quantum annealing can be implemented using the D-Wave quantum annealers. The Hamiltonian of the D-Wave quantum annealers can be expressed as an Ising model:

$$H_{Ising} = -\frac{A(s)}{2}\left(\sum_i \hat{\sigma}_x^{(i)}\right) + \frac{B(s)}{2}\left(\sum_i h_i \hat{\sigma}_z^{(i)} + \sum_{i>j} k_{i,j} \hat{\sigma}_z^{(i)} \otimes \hat{\sigma}_z^{(j)}\right) \tag{104}$$

where $s \in [0, 1]$ is the normalized annealing parameter, $\hat{\sigma}_x^{(i)}$ represents the Pauli-X operator acting on qubit i, while $\hat{\sigma}_z^{(i)}$ and $\hat{\sigma}_z^{(j)}$ denote the Pauli-Z operators acting on qubits i and j, respectively. The coefficients h_i and $k_{i,j}$ represent the qubit's bias and coupling strength, respectively, corresponding to the linear and quadratic terms in the QUBO form of the MP. The first term in Equation (104) describes the initial Hamiltonian H_b, while the second term describes the problem Hamiltonian H_p, whose lowest energy eigenstate represents the optimal solution to the QUBO MP. The Hamiltonian H_p consists of two components: the first part describes the influence of the external magnetic field on the spin of each qubit, and the second part represents the interaction between qubits.

To achieve the adiabatic evolution of the quantum system, one must first identify the problem Hamiltonian H_p corresponding to the QUBO form MP and then prepare the initial Hamiltonian H_b of the quantum system before the annealing process begins. Given that the quantum system's Hamiltonian evolves with minimal external energy interference, it can transition from the initial Hamiltonian to the problem Hamiltonian sufficiently slowly. This evolution process can be mathematically represented as follows:

$$H(t) = \left(1 - \frac{t}{T}\right) H_b + \frac{t}{T} H_p, \quad \frac{t}{T} \in [0, 1] \tag{105}$$

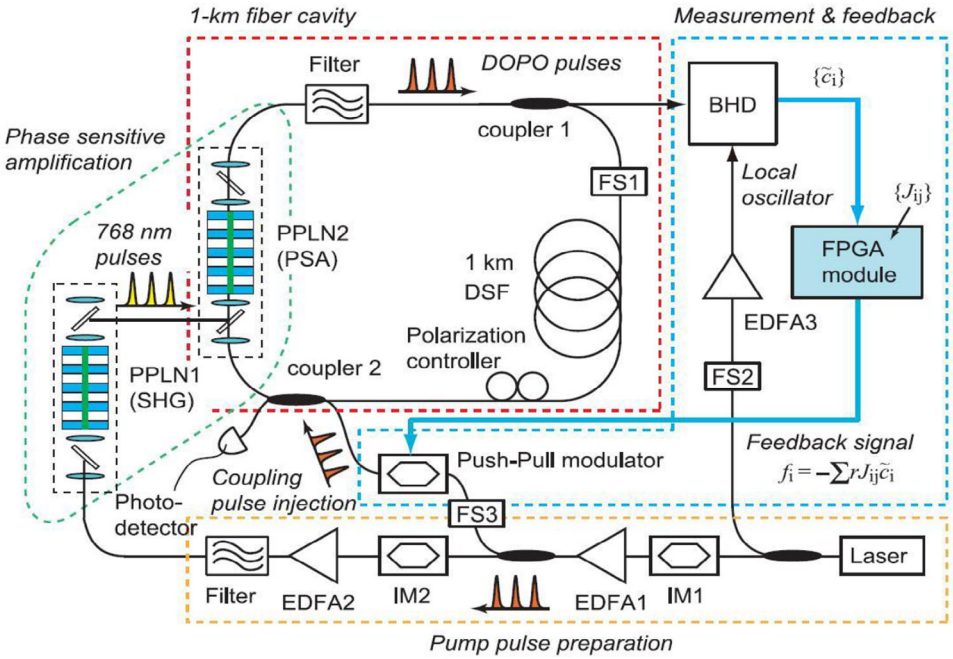

Figure 3. Fundamental principles of the CIM.

where t denotes the time parameter, and T is the total annealing time.

The solution principles of the quantum annealing algorithm can be described as follows: (1) Initial State of the Quantum System ($t = 0$): Before the quantum annealing process begins, at time $\frac{t}{T} = 0$, the quantum system is in a stationary state, specifically in the ground state of the initial Hamiltonian H_b. During this phase, all qubits exist in a superposition of state $|0\rangle$ and $|1\rangle$ (i.e., $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$); (2) Quantum Annealing Process Begins ($0 < t < T$): As the quantum annealing process proceeds, a slow perturbation is applied to the quantum system. This introduces the coupling and bias terms from the problem Hamiltonian into the qubits. As these interactions are gradually imposed, the qubits begin to entangle with each other. If this perturbation is applied over a sufficiently long period and with a sufficiently slow rate of change (representing an ideal quantum annealing process), the quantum system will consistently remain in its ground state without transitioning to an excited state; (3) Final State of the Quantum System ($t = T$): At the end of the annealing process, that is, $\frac{t}{T} = 1$, the quantum system reaches its final state. Given that no level crossing occurs during the entire annealing process, the system is now in the ground state of the problem Hamiltonian. This ground state can then be measured to yield the solution to the QUBO MP.

3.6. Model Calculation Vased on Optical Quantum CIM

Wang et al. introduced a novel approach for constructing a coherent quantum computer using a network of Degenerate Optical Parametric Oscillators (DOPOs) to solve NP-hard problems by finding the ground state of the Ising model in 2013.[53] In 2021, a coherent quantum computer with 100512 DOPO pulses functioning as Ising spins demonstrated a significant computational advantage over standard simulated annealing techniques when solving a 100 000-node graph.[54] This achievement marked the coherent quantum computer as the only QC platform in the academic field capable of handling problems at a scale of 100 000 bits, establishing a new benchmark in QC capabilities.

Degenerate Optical Parametric Oscillators (DOPOs) driven by femtosecond laser pulses are widely utilized in both classical and quantum optics. This process involves a nonlinear optical phenomenon where, through second-order nonlinear interactions, the input laser (pump light) is converted into two output beams (signal and idler waves) of the same frequency. Building on this DOPO process, Wen et al. within QBoson Quantum Technology platform developed a specialized optical quantum computer, known as the CIM. This quantum computer exhibits notable advantages in room-temperature optical quantum encoding, control, and full connectivity.

The fundamental principles of the CIM are illustrated in **Figure 3**, depicting a hybrid QC system comprising optical and electrical subsystems. The optical subsystem is responsible for the preparation and storage of qubits, while the electrical subsystem manages the control and computation of these qubits. The entire QC process involves four key modules: the pump pulse preparation module, the phase sensitive amplification module, the DOPO module, and the measurement & feedback module. These components work in tandem to execute the operations necessary for quantum computation. In the pump pulse preparation, femtosecond fiber lasers generate laser pulses, which are then modulated and amplified using an intensity modulator (IM) and an erbium-doped fiber amplifier (EDFA). In the phase sensitive amplification module, nonlinear optical effects are achieved using a periodically poled lithium niobate (PPLN1) crystal, generating a 768 nm second-harmonic laser. This second-harmonic laser

then serves as the pump laser for conversion into a 1560 nm signal laser via a second crystal (PPLN2), producing two squeezed laser states with frequencies that are half that of the pump laser and with matching polarization directions. These squeezed laser pulses are then routed into a fiber cavity to form a DOPO, generating optical pulses with specific phase and amplitude properties, which serve as the optical qubits. All optical qubits are maintained and stored within the fiber cavity for subsequent quantum computations. As the pump power increases and exceeds the oscillation threshold, the system transitions into a coherent state, where the optical phase bifurcates into two distinct states, 0 and π, which are analogous to the spin-up and spin-down states, respectively, for solving Ising optimization problems. In the measurement & feedback module, feedback signals are computed using a field programmable gate array (FPGA) based on the user-defined Ising problem matrix. This feedback guides the evolution of the optical qubits within the fiber cavity towards the lowest energy state of the Ising Hamiltonian, thereby facilitating the solution of the Ising optimization problem.

Specifically, the Ising matrix is loaded onto the FPGA, which then processes the phase and amplitude information of the optical pulses circulating in the fiber loop, obtained through balanced homodyne detection (BHD). Based on this information, the FPGA computes the feedback signal and modulates the feedback optical pulses using a push-pull modulator. These feedback pulses interact with the optical pulses within the fiber loop, directing the evolution of the optical qubits toward the lowest energy state of the Hamiltonian of the Ising problem. When the evolution is complete, the measured phase information of the optical qubits represents the final solution to the Ising problem. This mechanism endows the optical quantum computer with a potential advantage in handling high-dimensional, nonlinear optimization problems, showing significant promise in terms of parallel computing capabilities and processing speed.

4. Numerical Results

4.1. Experimental Preparation

This research presents a scalable distributed solution method (Distributed QDA) for the joint optimization problem of UC and OTS based on multi-decomposition and consensus strategies. Depending on the power system scenario, after the multi-decomposition in the QDA framework, the optimization subproblems within each module may not necessarily adopt the consensus strategy and can instead be solved using a global modeling approach (Centralized QDA). In the former case, due to the involvement of cross-regional power flow as a consensus variable, the optimization problems can be decoupled and solved independently at the regional level, making it suitable for large-scale, cross-regional power systems. The latter is more appropriate for smaller scale, centralized, unified scheduling scenarios. To verify the efficiency of the proposed method, we performed multiple scenario analyses using a 6-Bus System and the IEEE RTS 24-Bus System. Given the relatively simple structure of the 6-Bus System, with fewer continuous and binary variables, it allowed for a more straightforward analysis of the interactions between OTS and UC in the scenario experiments. Therefore, we first employed the Centralized QDA on the 6-Bus System to verify the correctness and feasibility of the algorithmic process. Building on the confirmed validity of the centralized approach, we extended our experiments to the more extensive and complex IEEE RTS 24-Bus System to demonstrate the superiority of the consensus-based Distributed QDA. In scenarios 6 and 7, the computational results and solution times achieved by this distributed approach showed clear advantages over the commercial solver Gurobi9. In our QDA, all SPs were solved using the classical solver Gurobi9, while the QUBO form MPs of UC and OTS were implemented on the D-Wave *Leap*^{TM} platform and the optical quantum CIM of Beijing QBoson Quantum Technology Co., Ltd.

The time required for executing a quantum instruction on a D-Wave QPU includes the initialization step's QPU programming time, as well as the QPU sampling time, which consists of annealing time, readout time, and delay time. In contrast to annealing machines, the optical quantum CIM does not include an initialization step during a single computation process, nor does it require multiple sampling processes. Instead, it relies on a unified synchronous evolution, where the optimal solution is read out when the compressed state of light undergoes spontaneous symmetry breaking. Therefore, the QPU time for the optical quantum CIM refers to the duration of a single sampling to output the current solution.

4.2. 6-Bus System

Figure 4a presents the schematic diagram of the 6-Bus System, while the load variations of the 6-Bus System over time period are illustrated in Figure 4b. **Table** 2 provides detailed transmission line data for the 6-Bus System, including the direction of power flow, line reactance, and power flow capacity limits. We applied the QDA to analyze multiple scenarios of the 6-Bus System, aiming to determine the optimal system operating cost under hourly load conditions, as well as to provide the unit scheduling schemes and transmission line switching strategies. The load was distributed among buses B3, B4, and B5, with allocations of 20%, 40%, and 40%, respectively, while the three units are located at buses B1, B2, and B6.

In our experiments with the 6-Bus System, we considered four distinct scenarios to evaluate the effectiveness of OTS in reducing system operating costs and to investigate the intelligent switching strategies in the case of N-1 transmission line contingencies:

Scenario 1: UC without considering transmission constraints or power flow equations. The UC model in this case includes only the system's overall power balance constraint, unit power capacity constraints, and unit minimum up/down constraints. The solution to the UC problem under these conditions yields the unit state and output power variations over the 24 h, as shown in **Figure** 5. In this scenario, unit G1 remains online throughout all time periods, G2 is operational from time period 8 to 24, while the costly G3 stays inactive. The total system operating cost for this scenario amounts to 104725.60.

In Scenario 1, which focuses purely on the UC problem, we test and discuss two different consensus mechanisms for the relaxed SP of UC: the local relaxed SP of UC with average consensus (Equations (28–32) and the consensus-inspired relaxed SP of UC (Equations 33–35). As shown in **Figure** 6a, the 6-Bus System is divided into three regions (denoted by red dashed boxes), with

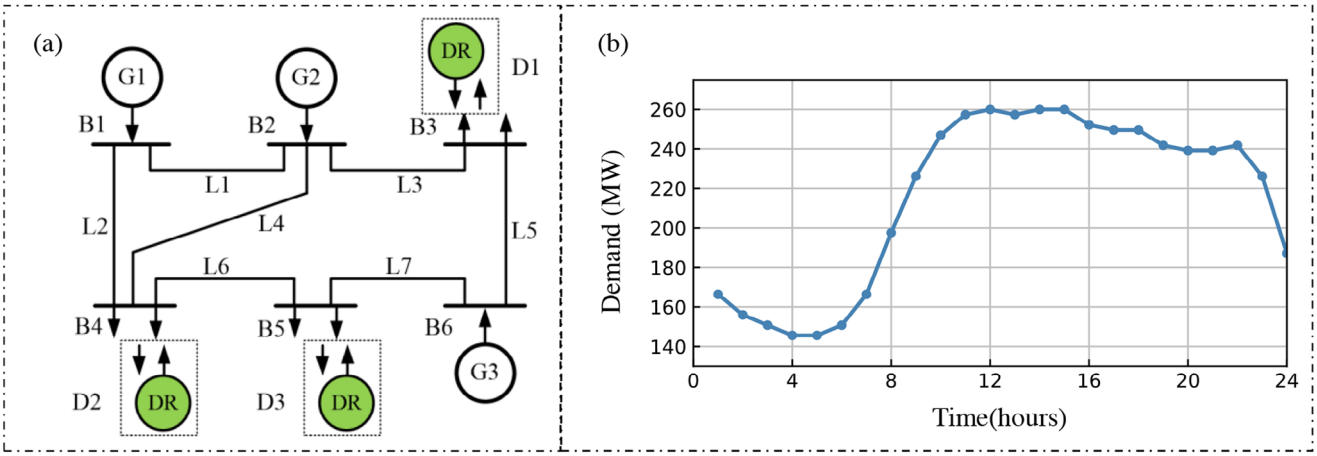

Figure 4. a) Schematic diagram, and b) Load variation of the 6-Bus System.

Table 2. Transmission line data of the 6-Bus System.

Line No.	From Bus	To Bus	X[pu.]	Flow Limit [MW]
1	1	2	0.170	150
2	1	4	0.258	150
3	2	3	0.037	150
4	2	4	0.197	150
5	3	6	0.018	150
6	4	5	0.037	37
7	5	6	0.140	150

each region containing one bus equipped with one unit, and adjacent regions connected by transmission lines. We first adopt the average consensus to solve the local relaxed SP of UC in a distributed manner. Achieving consensus requires multiple iterations of consensus variables related to the neighboring buses. There are three buses involved, with a total of four consensus variables (power exchange between neighboring buses) undergoing a convergence process. Figure 6b,c display the consensus iteration process of power exchange between buses B1 and B2 ($\vec{P}^{B1}_{B1,B2}$, $\vec{P}^{B2}_{B1,B2}$), and between buses B2 and B6 ($\vec{P}^{B2}_{B2,B6}$, $\vec{P}^{B6}_{B2,B6}$), respectively. From Figure 6b, it can be seen that for any given time period t, after several iterations, the consensus variables $\vec{P}^{B1}_{B1,B2}$ and $\vec{P}^{B2}_{B1,B2}$ converge finally. Similarly, in Figure 6c, the consensus variables $\vec{P}^{B2}_{B2,B6}$ and $\vec{P}^{B6}_{B2,B6}$ also converge. Once each consensus variable has converged, the relaxed variable distribution for each bus within the region can be determined.

We then adopt the consensus-inspired strategy by solving the consensus-inspired relaxed SP of UC, and the optimal relaxed variable distribution for each region can be obtained in a single, non-iterative step. **Table 3** presents a comparison of the data for the two consensus mechanisms. The data shows that the average consensus, due to its reliance on multiple iterations to achieve convergence, has a significantly higher time-consuming compared to the single-step solution of the consensus-inspired relaxed SP of UC. Specifically, the number of iterations for the average consensus fluctuates between 74 and 186 times in all time periods, and the average iteration time for each region's local relaxed SP of UC fluctuates between 0.05 and 0.21 s, resulting in an average total time of 2.36 s. In contrast, the consensus-inspired relaxed SP of UC in the Distributed QDA has a stable solution time between 1.00 and 2.74 ms, with a total time of 0.042 s, which is only 1.78% of the total time of the consensus mechanism, demonstrating a significant speed advantage.

Scenario 2: UC considering transmission line constraints and power flow equations. Compared to Scenario 1, the model here is

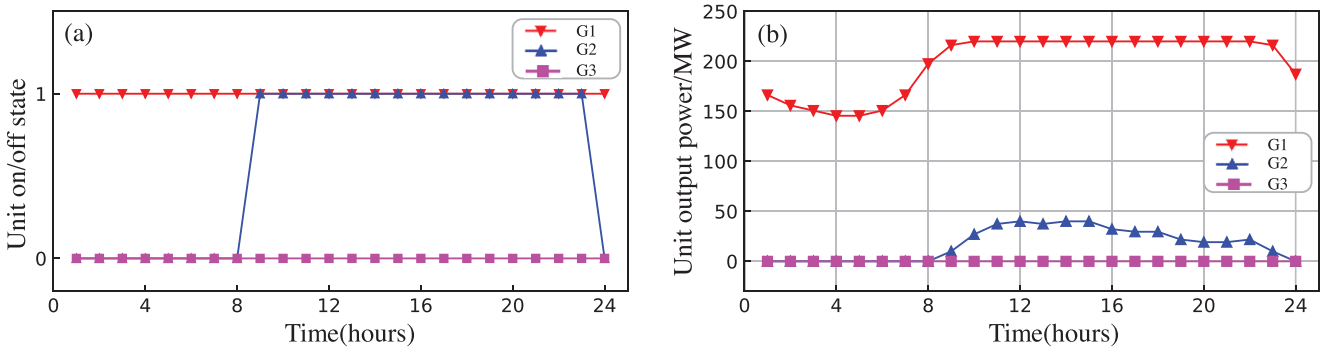

Figure 5. a) Unit state, and b) Unit output power in Scenario 1.

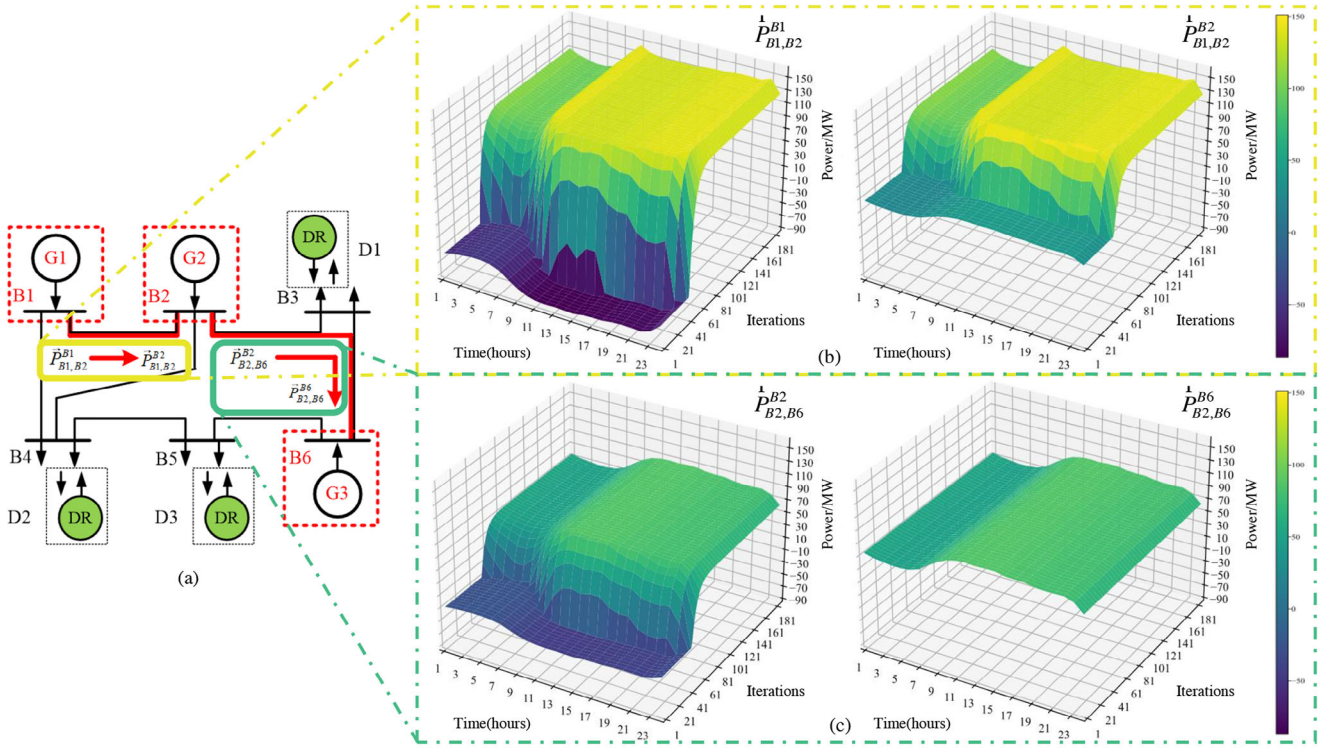

Figure 6. a) Consensus regions, b) Iterative convergence process of the consensus variable $\vec{P}^{B1}_{B1,B2}$ and $\vec{P}^{B2}_{B1,B2}$, c) Iterative convergence process of the consensus variable $\vec{P}^{B2}_{B2,B6}$ and $\vec{P}^{B6}_{B2,B6}$ in Scenario 1.

Table 3. Solution data of average consensus mechanism and the Distributed QDA based on consensus-inspired relaxed SP of UC in Scenario 1.

Time Period	1	2	3	4	5	6	7	8	9	10	11	12
Consensus Iterations	179	182	184	186	186	184	179	172	91	78	75	74
Average ConsensusTime-consuming (s)	0.14	0.14	0.13	0.13	0.21	0.13	0.13	0.21	0.06	0.10	0.05	0.06
Distributed QDA's Consuming (ms)	1.36	1.04	2.70	1.01	2.09	1.00	2.02	1.00	2.53	2.01	2.15	1.00
Time Period	13	14	15	16	17	18	19	20	21	22	23	24
Consensus Iterations	75	74	74	76	77	77	80	81	81	80	91	174
Average ConsensusTime-consuming (s)	0.08	0.05	0.05	0.05	0.07	0.06	0.06	0.13	0.06	0.06	0.07	0.13
Distributed QDA's Consuming (ms)	1.16	1.00	1.51	0.10	2.74	2.45	2.19	1.67	3.32	1.91	1.99	1.99

extended by adding a fixed transmission network topology with full connection, accounting for transmission line power flow capacity constraints and power flow equations. Additionally, Scenario 2 is divided into two sub-scenarios: Scenario 2-1 uses the default upper limits for transmission line power capacities; Scenario 2-2 reduces the upper limits for transmission line power capacities.

Figure 7 illustrates the unit on/off state, power output, and transmission power flow across time periods for both Scenario 2-1 and Scenario 2-2. In Scenario 2-1, the unit scheduling is adjusted compared to Scenario 1 due to the inclusion of power flow constraints and transmission line capacity constraints. G1 remains online throughout the entire time period, while the costly G3 remains offline. However, G2 is shut down during periods 2–6 and operates in the remaining periods. The unit output power is also adjusted accordingly to meet the power flow constraints. Compared to Scenario 1, G1's output power is reduced during periods of power flow congestion (time periods 1 and 7–24), leading to an increase in the output power of the relatively more expensive G2 to satisfy the system load at each bus. Consequently, this results in a higher system operating cost of 110103.45. To further emphasize the impact of power flow constraints on unit scheduling, in Scenario 2-2, we reduce the transmission line power limits, setting the maximum power capacities of lines L1, L2, L3, L4, L5, and L7 from the original 150 MW to 100 MW. The reduced maximum power limits lead to a more pronounced congestion issue on the transmission lines. As a result, to satisfy the load demand at each bus, the costly G3 is required to participate in power generation during time periods 9–22, further increasing the total system operating cost to 111735.23.

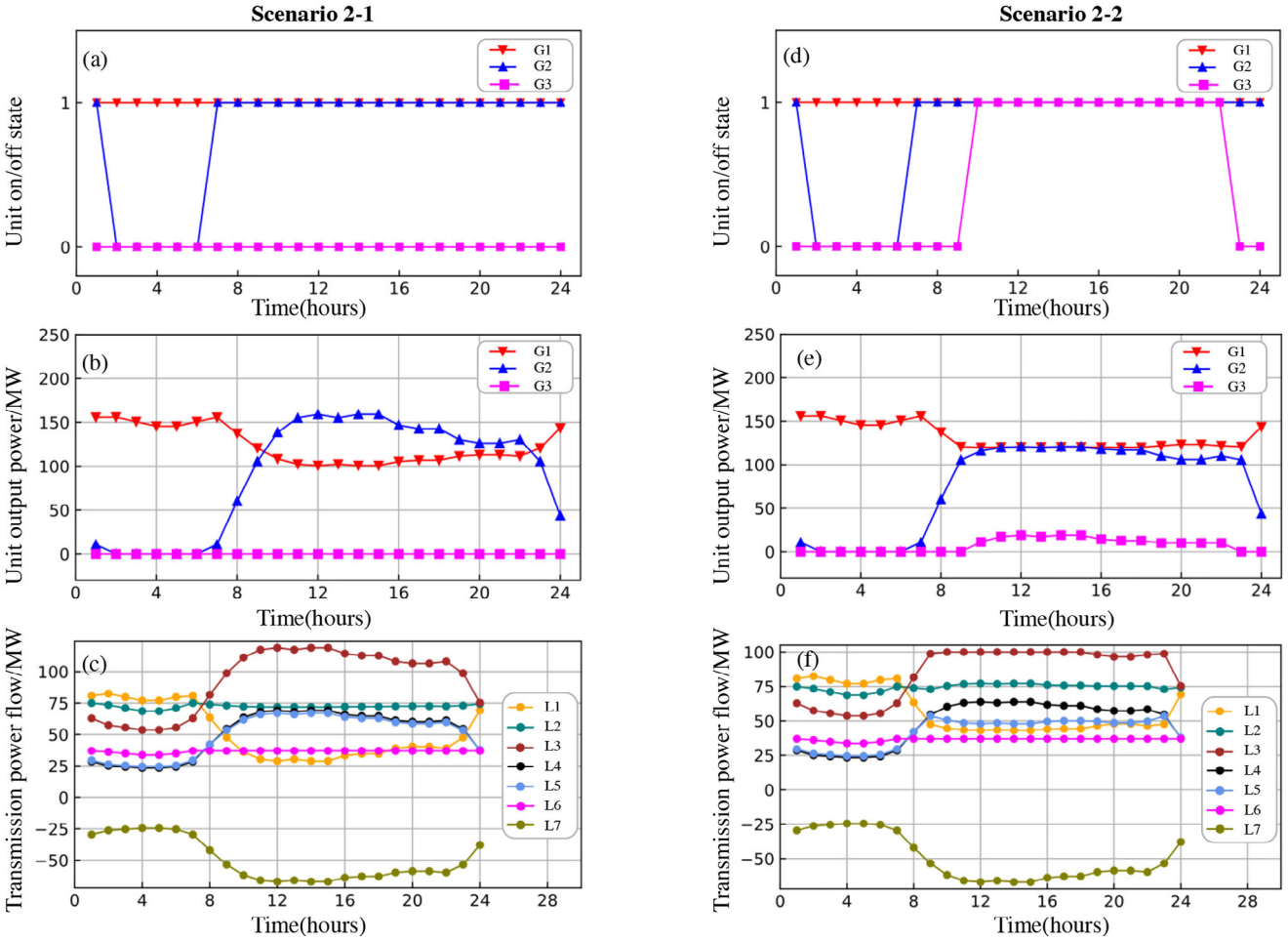

Figure 7. a)/d) Unit state, b)/e) Unit output power, and c)/f) Transmission power flow in Scenario 2-1/ Scenario 2-2.

Scenario 3: The joint optimization model for UC and OTS. This scenario builds on the UC model from Scenario 2 by introducing binary variables to represent the state of transmission lines. In the joint optimization model, both unit scheduling and transmission switching are optimized simultaneously to minimize system operating costs. Similar to Scenario 2, two subscenarios are considered: Scenario 3-1, with default transmission line power flow limits, and Scenario 3-2, with reduced transmission line power flow limits.

The Centralized QDA is used to solve the joint optimization problem for both sub-scenarios. The unit state, power output, transmission line state, and power flow results are presented in **Figure 8**. In comparison to Scenario 2-1, Scenario 3-1 achieves a significant reduction in system operating costs by optimizing transmission switching, reducing the total cost to 104725.60, which is the same as that from Scenario 1 and represents a 4.88% decrease compared to Scenario 2-1. The final unit scheduling also aligns with that of Scenario 1. Transmission line L4 is disconnected during time periods 11–16, and L6 is disconnected during time periods 1, 7–10, and 17–24. Similarly, in Scenario 3-2, the maximum power flow limits of L1, L2, L3, L4, L5, and L7 are reduced to 100 MW. The total system operating cost is 111409.07, reflecting a 6.38% increase compared to Scenario 3-1. However, compared to Scenario 2-2, which also reduced the transmission line power limits, the system cost in Scenario 3-2 is reduced by 0.29%. This demonstrates that incorporating transmission switching into the UC model effectively alleviates power flow congestion while reducing overall system operating costs.

Scenario 4: The joint optimization model for UC and OTS under N-1 contingency conditions. This scenario extends the joint optimization model from Scenario 3 by considering N-1 contingency conditions, specifically addressing the situation where unexpected line failures occur. The goal is to adapt to such contingencies through emergency adjustments in both unit scheduling and transmission line switching, thereby ensuring that bus loads are met without significant curtailments.

Figure 9a illustrates the scenario in which L1 of the 6-Bus System experiences a failure, taking it offline. In this diagram, the blue solid line indicates the offline state of L1, while the red solid lines highlight transmission lines with power flow violations. Under these conditions, the UC model from Scenario 2 would lead to severe power flow violations, potentially causing short-term or long-term instability in the power system, and in extreme cases, a complete system collapse. As shown in Figure 9b, the power flow constraint violations in Scenario 2 become apparent when L1 fails. The red solid lines represent the power flow limits

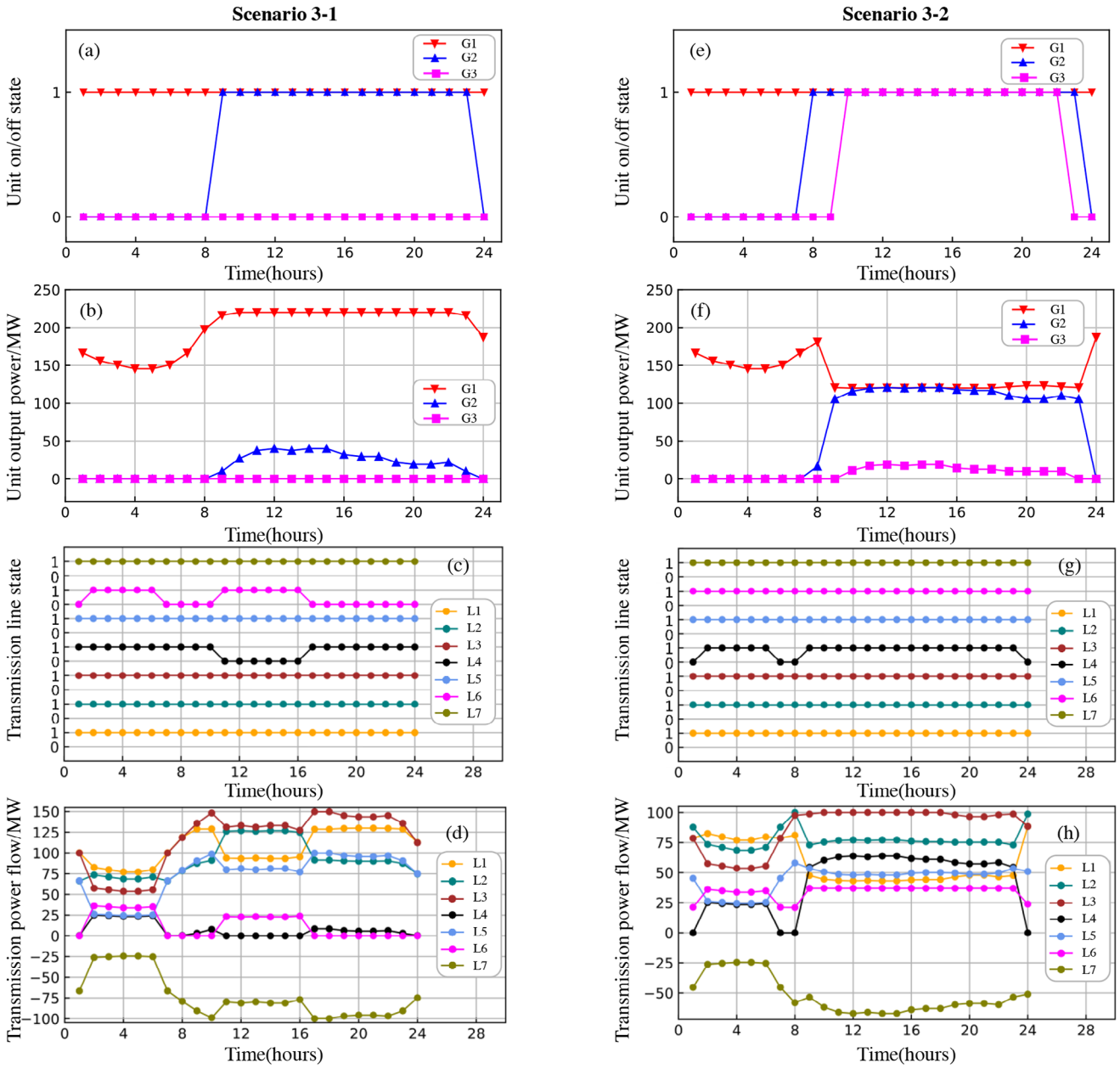

Figure 8. a)/e) Unit state, b)/f) Unit output power, c)/g) Transmission lines state, and d)/h) Transmission power flow in Scenario 3-1/Scenario 3-2.

of L1, L2, L3, L4, L5, and L7, while the red dashed line indicates the power flow limit of L6. In this scenario, L2 experiences power flow violations during time periods 1–3 and 6–24, and L6 exceeds its power flow limit for the entire 24-hour period.

To address this issue, we employ the Centralized QDA to solve the joint optimization model under N-1 contingency conditions. The results are presented in **Figure 10**, which demonstrates that the joint optimization of UC and OTS successfully adapts to the offline condition of transmission line L1, thereby ensuring bus loads are met while avoiding power flow violations. To specifically mitigate the violations on L2 and L6, the solution involves disconnecting L4 during time periods 11–16 and L6 during time periods 1–10 and 17–24. Correspondingly, G1 remains online throughout the entire 24-hour period, while G2 is scheduled to be online during time periods 1–3 and 6–24. The total system operating cost under these conditions is 108473.20, which represents a 3.58% increase compared to Scenario 3-1, where no contingency occurred for L1.

4.3. IEEE RTS 24-Bus System

To evaluate the performance of the consensus-based Distributed QDA in solving the joint optimization model for UC and OTS in larger-scale systems, we conducted computational analysis on the IEEE RTS 24-Bus System. **Figure 11a** presents the schematic di-

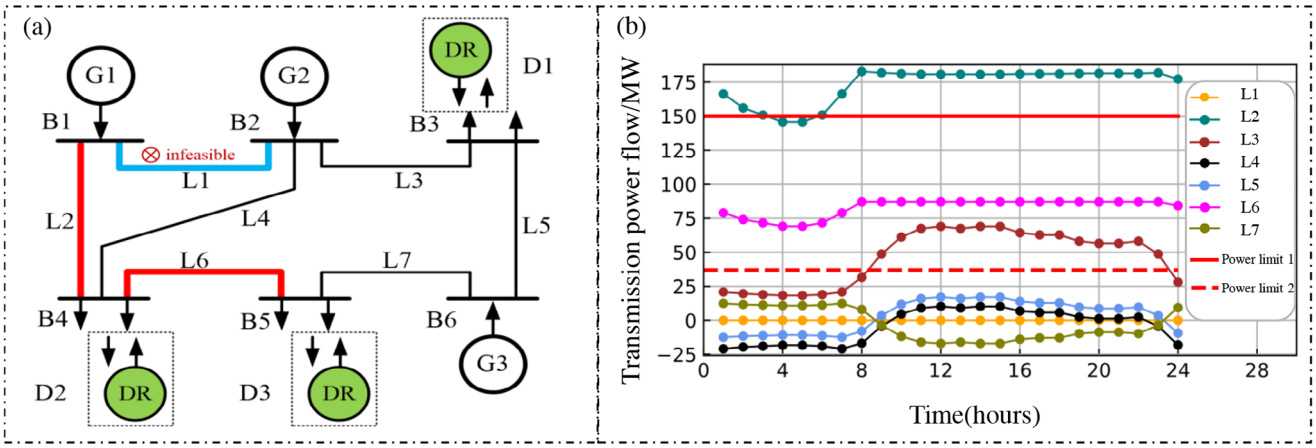

Figure 9. a) The 6-Bus System under a contingency on transmission line L1, and b) Flow limits exceeded under contingency.

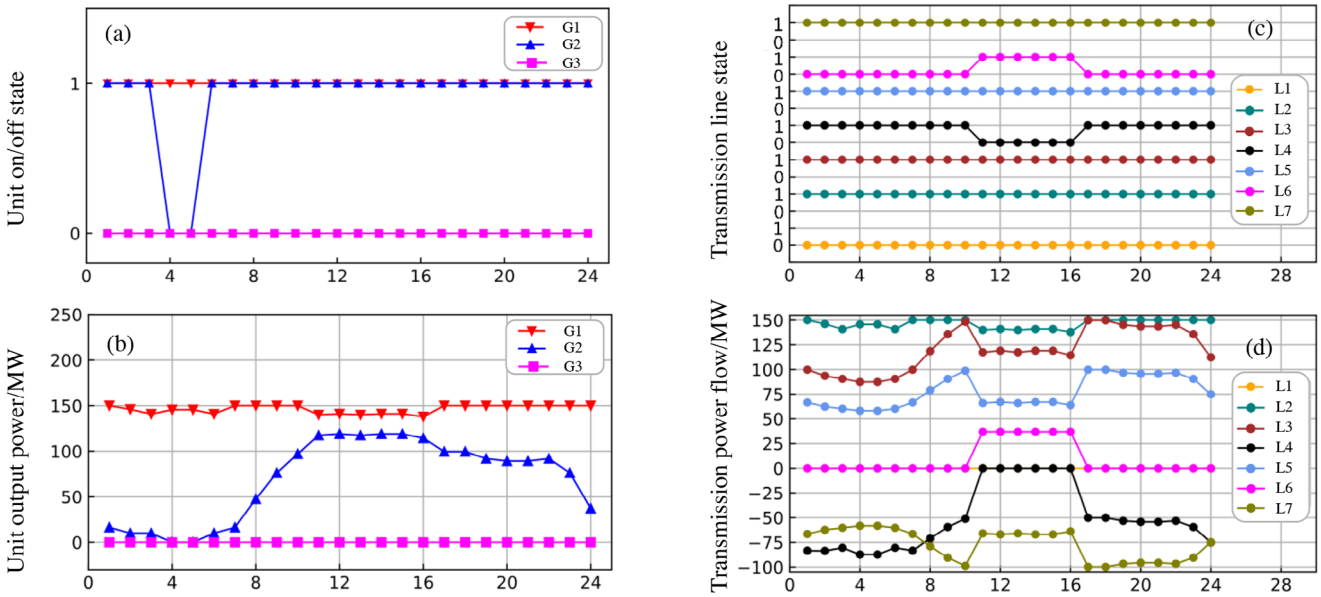

Figure 10. a) Unit state, b) Unit output power, c) Transmission lines state, and d) Transmission power flow in Scenario 4.

agram of the IEEE RTS 24-Bus System, while the load variations of the IEEE RTS 24-Bus System over time periods are illustrated in Figure 11b. This model consists of 24 buses, 34 transmission lines, and 26 units, which are distributed across buses B1, B2, B7, B13, B15, B16, B18, B21, and B23. The detailed bus load distribution and the parameters for unit and transmission lines are consistent with the specifications provided in Ref. [55]. Using the Distributed QDA, we analyzed three different scenarios for the IEEE RTS 24-Bus System. In each scenario, we provided the hourly unit scheduling and transmission switching strategies, as well as the corresponding system operating costs, while ensuring that the bus load requirements were met. This analysis aims to demonstrate the algorithm's capability to handle complex power system operations in a larger-scale system efficiently.

Scenario 5: The joint optimization of UC and OTS. In this scenario, we focused on the joint optimization problem under the same constraints as those applied in Scenario 3. The final resulting transmission line state is a full connection. **Figure 12** shows the bus output power (the total output power from all units connected to the bus) at different time periods, representing the bus's contribution to system generation across various time periods. The total system operating cost achieved was 738159.68, matching the result obtained using the commercial solver Gurobi9.

Scenario 6: The joint optimization with reduced transmission line power flow limits. To further examine the impact of transmission switching on reducing operational costs in the IEEE RTS 24-Bus System, we extended the analysis from Scenario 5 by lowering the power flow limits of transmission lines (L1-L34) by 80 MW from their original values. While this reduction led to an infeasible solution under the fully connected line configuration, the joint optimization model of UC and OTS identified a feasible solution. **Figure 13** presents the results of the Distributed QDA for this scenario: L1 and L6 were disconnected during time

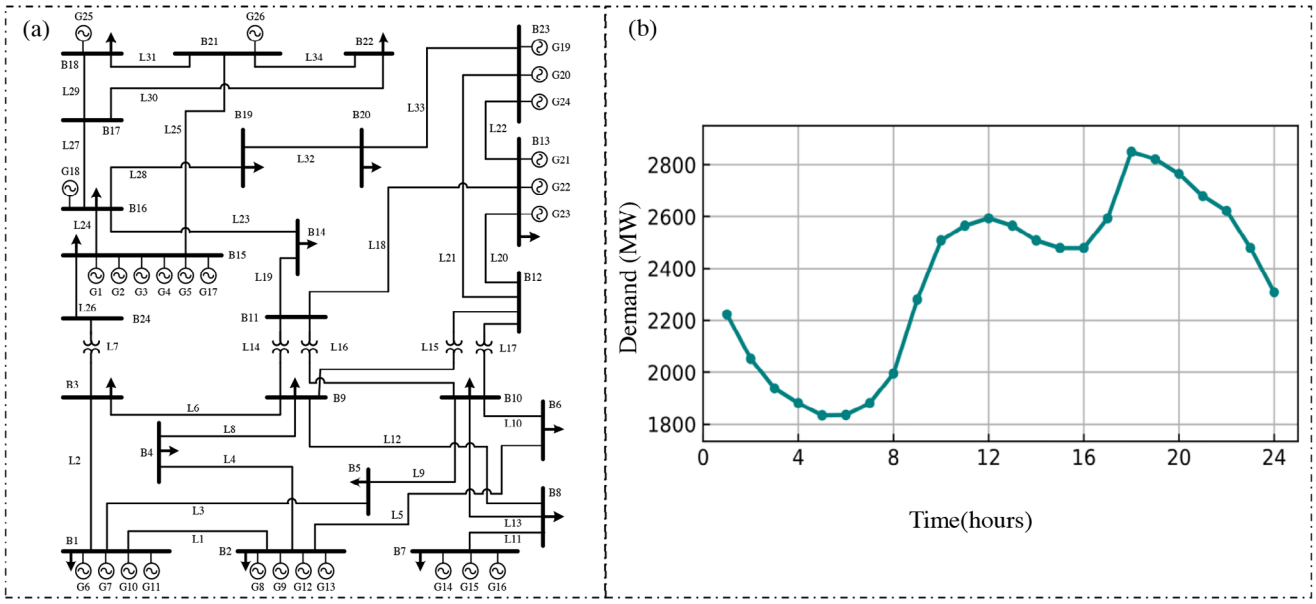

Figure 11. a) Schematic diagram, and b) Load variation of the IEEE RTS 24-Bus System.

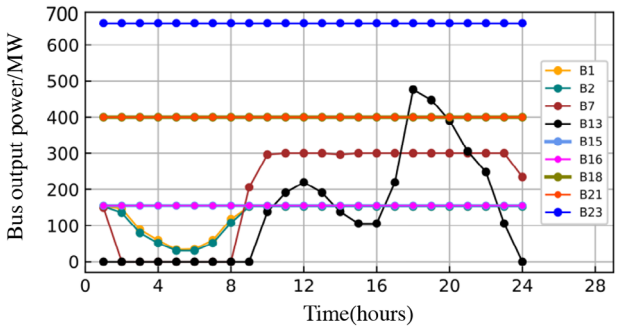

Figure 12. Bus output power in Scenario 5.

periods 3 and 5–6, L7 was disconnected during time periods 2 and 8, L16 was disconnected during time period 18, and L23 was disconnected during time period 19, while the remaining lines stayed connected throughout all time periods. The resulting system operating cost was 740095.81, representing a 0.003% reduction compared to Gurobi9's calculation of 740118.16.

Scenario 7: The Joint optimization under N-1 transmission line contingency. To further analyze the impact of an N-1 transmission line contingency on the IEEE RTS 24-Bus System, Scenario 7 builds upon the setup of Scenario 6 by assuming that transmission line L10 is offline due to a failure. In this scenario, the traditional UC model with fixed transmission topology (similar to Scenario 2) remained infeasible. However, the joint optimization model of UC and OTS identified a feasible solution. The results, obtained using the Distributed QDA, are illustrated in **Figure** 14: L4 was disconnected during time period 18, and L6, L22, and L29 were all disconnected during time period 19, while the remaining lines stayed connected throughout all time periods. The total system operating cost was reduced to 759120.52, representing a 0.34% decrease compared to Gurobi9's solution of 761688.25.

4.4. Summary of Model Examples

Tables 4 and 5 present detailed computational results for the 6-Bus System and IEEE RTS 24-Bus System models across Scenarios 1 to 7. The tables include results such as the objective function values, outer iteration counts between the upper UC module and lower OTS module, the number of inner interactions in these two modules, the total computation time required, and the utilization times of both the Quantum Processing Unit (QPU) and the Classical Processing Unit (CPU). For the 6-Bus System, given its relatively small scale and simple structure, there was no need to apply the Distributed QDA. Instead, the Centralized QDA was sufficient to validate the workflow's effectiveness. In contrast, for the larger and more complex IEEE RTS 24-Bus System, the centralized method struggled to converge to a feasible solution within a reasonable timeframe. Consequently, the consensus-based Distributed QDA was employed to demonstrate its computational efficiency. The convergence criteria of the upper and lower bounds of Gurobi9 and QDA are both set to error tolerance 10^{-4}. The OTS is only involved starting from Scenario 3 and beyond, so the CIM is employed to solve the MP Hamiltonian beginning with Scenario 3. During the iterative process, the number of cutting planes in the MP increases linearly with the number of iterations, theoretically leading to a progressively increasing optimization complexity and, consequently, slower solution speed. Due to the high cost of using CIMs to solve QUBO problems, for optimization tasks involving the CIM, we only solve the QUBO form MP in the final iteration, considering its solution time as an upper bound for the time required to solve the MP in each iteration. This approach allows us to estimate the upper limits of the Quantum Processing Unit (QPU) usage time and the total computation time. This strategy is both cost-effective and sufficient to validate the algorithm's efficacy while also highlighting the quantum advantage of the CIM. In the results of the 6-Bus System shown in Table 4, the objective function values obtained

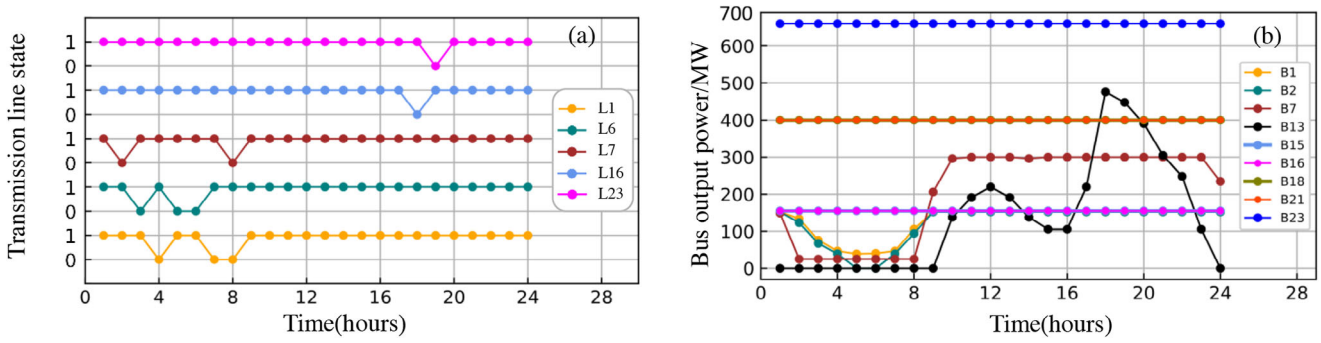

Figure 13. a) Transmission line state, and b) Bus output power in Scenario 6.

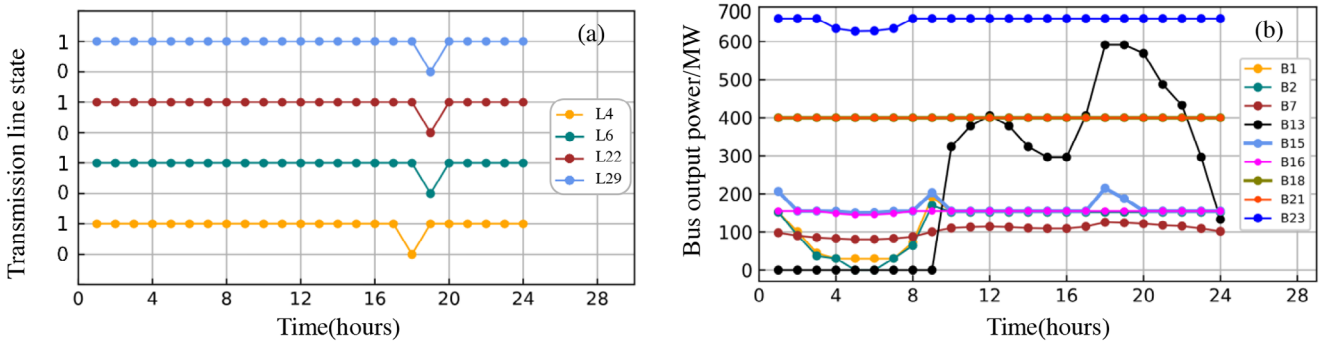

Figure 14. a) Transmission line state, and b) Bus output power in Scenario 7.

using the Centralized QDA are identical to those from Gurobi9 (with a superior result in Scenario 4). However, both the D-Wave and CIM versions of the QDA require more computation time to converge compared to Gurobi9. On the other hand, the results for the IEEE RTS 24-Bus System in Table 5 show that the consensus-based Distributed QDA consistently achieves objective function values that are equal to or better than those from Gurobi9 (with superior results in Scenarios 6 and 7). Moreover, the convergence times for all scenarios are significantly shorter than those of Gurobi9, with the CIM implementation of the Distributed QDA outperforming the D-Wave version in terms of solution speed. These findings further underscore the potential of QC, particularly the QC based on the photonic quantum computer, in the field of optimization. The results also demonstrate the effectiveness and superiority of the Distributed QDA.

5. Conclusion

To address the joint optimization of UC and OTS with consideration of day-ahead contingency constraints, this study proposes both Centralized and consensus-based Distributed QDAs, leveraging the advantages of QC for solving QUBO problems. The consensus distributed version employs a multi-stage optimization strategy to construct cutting planes in parallel at the bus level, providing a day-ahead scheduling solution for units along with the corresponding optimal grid topology. This approach achieves a coordinated optimization of both the grid topology and unit output power, offering a flexible and reliable solution for enhancing the economic efficiency and security of power generation. In the consensus-based Distributed QDA, the upper UC module introduces the interactive power of neighboring buses as consensus variables, while the lower OTS module incorporates the real interactive power flows and phase angles of neighboring buses as consensus variables. This consensus-based construction of local relaxed SPs and consensus-inspired relaxed SP enables both the upper and lower modules to perform parallel computation and optimization, and consensus-inspired relaxed SP not only effectively allocates the relaxation variables and decouples the cutting planes, but also reduces the computation time to just 1.78% of the time required by the average consensus mechanism. The cutting planes that constitute the MP are sequentially transformed into the QUBO form Hamiltonian, enabling the discrete variable optimization of the MP to leverage the quantum annealer or CIM based on optical systems for accelerated computation. This significantly enhances the optimization speed of the algorithm. Additionally, we introduce a novel method for constructing the QUBO Hamiltonian for unit minimum up/down constraints. Unlike traditional discrete methods, this approach eliminates the need for auxiliary binary variables, substantially reducing the size of the QUBO matrix for the MP and thereby minimizing the quantum resource requirements. Finally, our developed method effectively mitigates and avoids power flow violations by optimizing the grid topology through transmission line switching, resulting in a reduction in operational costs. Comparative experiments using the 6-Bus System across multiple scenarios against Gurobi9 demonstrate the effectiveness of the Centralized QDA in solving the joint optimization problem of UC and OTS. Furthermore, tests on the IEEE RTS 24-Bus System indicate

Table 4. Results for the Centralized QDA in various scenarios for 6-Bus System.

System	Scenarios	Data	Gurobi9	Centralized QDA (Gurobi + D-Wave)	Centralized QDA (Gurobi + CIM)
6-Bus System	Scenario 1	Objective function value	102547.60	102547.60	–
		Number of iterations within upper UC module	–	3	–
		Number of iterations within lower OTS module	–	–	–
		Number of iterations between upper and lower modules	–	–	–
		Solution time (s)	0.0450	0.3270	–
		Solution time of QPU (s)	–	0.3180	–
		Solution time of CPU (s)	0.0450	0.0090	–
	Scenario 2-1	Objective function value	110103.45	110103.45	–
		Number of iterations within upper UC module	–	4	–
		Number of iterations within lower OTS module	–	48	–
		Number of iterations between upper and lower modules	–	2	–
		Solution time (s)	0.0708	0.3850	–
		Solution time of QPU (s)	–	0.3810	–
		Solution time of CPU (s)	0.0708	0.0040	–
	Scenario 2-2	Objective function value	111735.23	111735.23	–
		Number of iterations within upper UC module	–	4	–
		Number of iterations within lower OTS module	–	48	–
		Number of iterations between upper and lower modules	–	2	–
		Solution time (s)	0.0718	0.5049	–
		Solution time of QPU (s)	–	0.4770	–
		Solution time of CPU (s)	0.0718	0.0279	–
	Scenario 3-1	Objective function value	102547.60	102547.60	102547.60
		Number of iterations within upper UC module	–	3	3
		Number of iterations within lower OTS module	–	53	53
		Number of iterations between upper and lower modules	–	1	1
		Solution time (s)	0.1725	0.7886	< 0.5767
		Solution time of QPU (s)	–	0.2570	< 0.0593
		Solution time of CPU (s)	0.1725	0.5316	0.5316
	Scenario 3-2	Objective function value	111409.07	111409.07	111409.07
		Number of iterations within upper UC module	–	4	4
		Number of iterations within lower OTS module	–	55	55
		Number of iterations between upper and lower modules	–	2	2
		Solution time (s)	0.36901	1.0684	< 0.6117
		Solution time of QPU (s)	–	0.4770	< 0.0203
		Solution time of CPU (s)	0.36901	0.5914	0.5914
	Scenario 4	Objective function value	108504.36	108473.20	108473.20
		Number of iterations within upper UC module	–	4	4
		Number of iterations within lower OTS module	–	94	94
		Number of iterations between upper and lower modules	–	2	2
		Solution time (s)	0.0908	1.4003	< 0.9983
		Solution time of QPU (s)	–	0.4450	< 0.0430
		Solution time of CPU (s)	0.0908	0.9553	0.9553

Table 5. Results for the consensus-based Distributed QDA in various scenarios for IEEE RTS 24-Bus System.

System	Scenarios	Data	Gurobi9	Distributed QDA (Gurobi + D-Wave)	Distributed QDA (Gurobi + CIM)
IEEE RTS 24-Bus System	Scenario 5	Objective function value	738159.68	738159.68	738159.68
		Number of iterations within upper UC module	–	5	5
		Number of iterations within lower OTS module	–	24	24
		Number of iterations between upper and lower modules	–	1	1
		Solution time (s)	10.3473	8.2175	< 0.9060
		Solution time of QPU (s)	–	7.3383	< 0.0268
		Solution time of CPU (s)	10.3473	0.8792	0.8792
	Scenario 6	Objective function value	740118.16	740095.81	740095.81
		Number of iterations within upper UC module	–	6	6
		Number of iterations within lower OTS module	–	102	102
		Number of iterations between upper and lower modules	–	2	2
		Solution time (s)	48.5053	22.6671	< 3.5188
		Solution time of QPU (s)	–	19.3862	< 0.2379
		Solution time of CPU (s)	48.5053	3.2809	3.2809
	Scenario 7	Objective function value	761688.25	759120.52	759120.52
		Number of iterations within upper UC module	–	6	6
		Number of iterations within lower OTS module	–	50	50
		Number of iterations between upper and lower modules	–	2	2
		Solution time (s)	58.2102	11.3709	< 2.2148
		Solution time of QPU (s)	–	9.3026	< 0.1465
		Solution time of CPU (s)	58.2102	2.0683	2.0683

that the results and computation time of the QDA are superior to those of the classical commercial solver Gurobi9, validating the advantages of the consensus Distributed QDA in tackling large-scale joint optimization problems of UC and OTS.

Appendix A

In addressing the UC problem, the Benders decomposition algorithm introduces non-negative slack variables into the unit capacity constraints, thereby formulating original relaxed SP of UC:

$$\min \quad \eta = \sum_{i=1}^{NG}\sum_{t=1}^{NT} S_{i,t}^1 + S_{i,t}^2 \tag{A1}$$

$$s.t. \quad \sum_{i=1}^{NG} P_{i,t} = D_t, \rightarrow k_t^0 \quad \forall t \tag{A2}$$

$$P_i^{\min}\hat{U}_{i,t} \le P_{i,t} + S_{i,t}^1, \rightarrow k_{i,t}^1 \quad \forall i, \forall t \tag{A3}$$

$$P_{i,t} - S_{i,t}^2 \le P_i^{\max}\hat{U}_{i,t}, \rightarrow k_{i,t}^2 \quad \forall i, \forall t \tag{A4}$$

$$S_{i,t}^1, S_{i,t}^2 \ge 0. \rightarrow k_{i,t}^3, k_{i,t}^4 \quad \forall i, \forall t \tag{A5}$$

The objective function (A1) aims to minimize the sum of all slack variables, representing the deviation from the unit capacity constraints. Equation (A2) ensures the overall power balance constraint for the system. Equations (A3) and (A4) denote the relaxed unit capacity constraints, while Equation (A5) imposes the non-negativity condition on the slack variables. Solving this original relaxed SP of UC allows for the calculation of the minimum total deviation from the unit capacity constraints for all units in the system. Since there may exist multiple optimal solutions corresponding to the same optimal deviation value, our goal is to develop a consensus-inspired relaxed SP of UC, distributing the slack variables in a manner that aligns with the economic characteristics of units. This facilitates the construction of feasible cutting planes for local UC at each bus based on the solutions of the continuous and dual variables within the bus.

The above formulation constitutes a linear programming problem, exhibiting convex optimization properties. Consequently, the optimality of the solution can be represented using the Karush-Kuhn-Tucker (KKT) conditions, ensuring the attainment of the optimal deviation η. The Lagrangian function for this formulation is expressed as follows:

$$\begin{aligned}
&\mathcal{L}\left(P_{i,t}, S_{i,t}^1, S_{i,t}^2, k_t^0, k_{i,t}^1, k_{i,t}^2, k_{i,t}^3, k_{i,t}^4\right) \\
&= \sum_{i=1}^{NG}\sum_{t=1}^{NT} S_{i,t}^1 + S_{i,t}^2 + k_t^0\left(\sum_{i=1}^{NG} P_{i,t} - D_t\right) \\
&\quad + k_{i,t}^1\left(P_i^{\min}\hat{U}_{i,t} - P_{i,t} - S_{i,t}^1\right) - k_{i,t}^3\left(S_{i,t}^1\right) \\
&\quad + k_{i,t}^2\left(P_{i,t} - S_{i,t}^2 - P_i^{\max}\hat{U}_{i,t}\right) - k_{i,t}^4\left(S_{i,t}^2\right)
\end{aligned} \tag{A6}$$

The KKT conditions can be derived from the Lagrangian function of the following optimization problem:

$$\frac{\partial \mathcal{L}\left(P_{i,t}, S^1_{i,t}, S^2_{i,t}, k^0_t, k^1_{i,t}, k^2_{i,t}, k^3_{i,t}, k^4_{i,t}\right)}{\partial P_{i,t}} = 0, \quad \forall i, \forall t \tag{A7}$$

$$\frac{\partial \mathcal{L}\left(P_{i,t}, S^1_{i,t}, S^2_{i,t}, k^0_t, k^1_{i,t}, k^2_{i,t}, k^3_{i,t}, k^4_{i,t}\right)}{\partial S^1_{i,t}} = 0, \quad \forall i, \forall t \tag{A8}$$

$$\frac{\partial \mathcal{L}\left(P_{i,t}, S^1_{i,t}, S^2_{i,t}, k^0_t, k^1_{i,t}, k^2_{i,t}, k^3_{i,t}, k^4_{i,t}\right)}{\partial S^2_{i,t}} = 0, \quad \forall i, \forall t \tag{A9}$$

$$k^0_t\left(\sum_{i=1}^{NG} P_{i,t} - D_t\right) = 0, \quad \forall t \tag{A10}$$

$$k^1_{i,t}\left(P^{\min}_i \hat{U}_{i,t} - P_{i,t} - S^1_{i,t}\right) = 0, \quad \forall i, \forall t \tag{A11}$$

$$k^2_{i,t}\left(P_{i,t} - S^2_{i,t} - P^{\max}_i \hat{U}_{i,t}\right) = 0, \quad \forall i, \forall t \tag{A12}$$

$$k^3_{i,t}\left(S^1_{i,t}\right) = 0, \quad \forall i, \forall t \tag{A13}$$

$$k^4_{i,t}\left(S^2_{i,t}\right) = 0, \quad \forall i, \forall t \tag{A14}$$

$$P^{\min}_i \hat{U}_{i,t} - P_{i,t} - S^1_{i,t} \le 0, \quad \forall i, \forall t \tag{A15}$$

$$P_{i,t} - S^2_{i,t} - P^{\max}_i \hat{U}_{i,t} \le 0, \quad \forall i, \forall t \tag{A16}$$

$$S^1_{i,t}, S^2_{i,t}, k^1_{i,t}, k^2_{i,t}, k^3_{i,t}, k^4_{i,t} \ge 0 \tag{A17}$$

These conditions represent the KKT conditions of the original relaxed SP of UC within the Benders decomposition algorithm, ensuring the optimality of the given optimization problem.

Acknowledgements

This work was supported by the National Key Research and Development Program of China under Grant. 2022YFB3304700, Chinese Society of Optimization, Overall Planning and Economic Mathematics – Qboson Fund, and the Open Research Project of the State Key Laboratory of Industrial Control Technology under Grant. ICT2024B21.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

consensus, distributed, optimal transmission switching, quantum computing, unit commitment

Received: March 27, 2025
Revised: May 16, 2025
Published online:

[1] M. Yang, F. Gao, W. Dai, D. Huang, Q. Gao, F. Shuang, *Adv. Quantum Technol.* **2025**, *8*, 2400286.
[2] J. Huang, H. Qin, K. Shen, Y. Yang, B. Jia, *Energy* **2024**, *305*, 132229.
[3] A. Fusco, D. Gioffrè, A. Francesco Castelli, C. Bovo, E. Martelli, *Appl. Energy* **2023**, *336*, 120739.
[4] H. Wang, H. Yue, Y. Zheng, *IOP Conf. Ser.: Earth Environ. Sci.* **2021**, *675*, 012141.
[5] E. J. C. Fatule, Y. Sang, J. F. Espiritu, *Energy* **2025**, *315*, 134275.
[6] H. Gharibpour, F. Aminifar, M. H. Bashi, *IET Gener. Transm. Distrib.* **2017**, *12*, 217.
[7] M. Khanabadi, H. Ghasemi, M. Doostizadeh, *IEEE Trans. Power Syst.* **2012**, *28*, 542.
[8] E. Nasrolahpour, H. Ghasemi, *Electr. Power Syst. Res.* **2015**, *121*, 341.
[9] A. Nikoobakht, M. Mardaneh, J. Aghaei, V. Guerrero-Mestre, J. Contreras, *IET Gener. Transm. Distrib.* **2017**, *11*, 142.
[10] G. Granelli, M. Montagna, F. Zanellini, P. Bresesti, R. Vailati, M. Innorta, *Electr. Power Syst. Res.* **2006**, *76*, 549.
[11] R. P. O'Neill, R. Baldick, U. Helman, M. H. Rothkopf, W. Stewart Jr., *IEEE Trans. Power Syst.* **2005**, *20*, 171.
[12] W. Fu, H. Xie, H. Zhu, H. Wang, L. Jiang, C. Chen, Z. Bie, *Energy* **2023**, *284*, 129314.
[13] F. Barani, M. Mirhosseini, H. Nezamabadi-pour, M. M. Farsangi, *Appl. Soft Comput.* **2017**, *60*, 180.
[14] Z. Bi, X. Yang, B. Wang, W. Zhang, Z. Dong, D. Zhang, *Appl. Soft Comput.* **2023**, *149*, 110973.
[15] K. W. Hedman, M. C. Ferris, R. P. O'Neill, E. B. Fisher, S. S. Oren, *IEEE Trans. Power Syst.* **2010**, *25*, 1052.
[16] A. V. Ramesh, X. Li, 2019 North American Power Symposium (NAPS), IEEE, Wichita, KS, USA **2019**, pp. 1–6.
[17] R. P. O'Neill, K. W. Hedman, E. A. Krall, A. Papavasiliou, S. S. Oren, *Energy Syst* **2010**, *1*, 165.
[18] M. Sheikh, J. Aghaei, A. Letafat, M. Rajabdorri, T. Niknam, M. Shafie-Khah, J. P. S. Catalao, *IEEE Syst. J.* **2019**, *13*, 3933.
[19] J. Lin, Y. Hou, G. Zhu, S. Luo, P. Li, L. Qin, L. Wang, *Int. J. Electr. Power Energy Syst.* **2019**, *110*, 309.
[20] A. Khodaei, M. Shahidehpour, *IEEE Trans. Power Syst.* **2010**, *25*, 1937.
[21] R. Saavedra, A. Street, J. M. Arroyo, *IEEE Trans. Power Syst.* **2020**, *35*, 4408.
[22] L. Nan, Y. Liu, L. Wu, T. Liu, C. He, *J. Mod. Power Syst. Clean Energy* **2020**, *9*, 1458.
[23] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, et al., *Nature* **2019**, *574*, 505.
[24] M. Gong, S. Wang, C. Zha, M.-C. Chen, H.-L. Huang, Y. Wu, Q. Zhu, Y. Zhao, S. Li, S. Guo, H. Qian, Y. Ye, F. Chen, C. Ying, J. Yu, D. Fan, D. Wu, H. Su, H. Deng, H. Rong, K. Zhang, S. Cao, J. Lin, Y. Xu, L. Sun, C. Guo, N. Li, F. Liang, V. M. Bastidas, K. Nemoto, et al., *Science* **2021**, *372*, 948.
[25] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, P. Hu, X.-Y. Yang, W.-J. Zhang, H. Li, Y.

Li, X. Jiang, L. Gan, G. Yang, L. You, Z. Wang, L. Li, N.-L. Liu, C.-Y. Lu, J.-W. Pan, *Science* **2020**, *370*, 1460.
[26] L. S. Madsen, F. Laudenbach, M. F. Askarani, F. Rortais, T. Vincent, J. F. F. Bulmer, F. M. Miatto, L. Neuhaus, L. G. Helt, M. J. Collins, A. E. Lita, T. Gerrits, S. W. Nam, V. D. Vaidya, M. Menotti, I. Dhand, Z. Vernon, N. Quesada, J. Lavoie, *Nature* **2022**, *606*, 75.
[27] B. Wu, H. Yu, F. Wilczek, *Phys. Rev. A* **2020**, *101*, 012318.
[28] X. F. Yin, X. C. Yao, B. Wu, Y.-Y. Fei, Y. Mao, R. Zhang, L.-Z. Liu, Z. Wang, L. Li, N.-L. Liu, F. Wilczek, Y.-A. Chen, J.-W. Pan, *Proc. Natl. Acad. Sci. USA* **2023**, *120*, 2212323120.
[29] E. Gibney, *Nature* **2017**, *541*, 447.
[30] Y. Wang, Y. Li, Z.-Q. Yin, B. Zeng, *npj Quantum Inf.* **2018**, *4*, 46.
[31] P. W. Shor, Proc. 35th Annu. Symp. Found. Comput. Sci, IEEE, Santa Fe, NM, USA **1994**, pp. 124–134.
[32] L. K. Grover, *Phys. Rev. Lett.* **1997**, *79*, 325.
[33] A. W. Harrow, A. Hassidim, S. Lloyd, *Phys. Rev. Lett.* **2009**, *103*, 150502.
[34] J. Preskill, *Quantum* **2018**, *2*, 79.
[35] F. Gao, G. Wu, M. Yang, W. Cui, F. Shuang, *Quantum Inf. Process.* **2022**, *21*, 111.
[36] F. Gao, G. Wu, S. Guo, W. Dai, F. Shuang, *Appl. Soft Comput.* **2023**, *137*, 110147.
[37] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, J. L. O'Brien, *Nat. Commun.* **2014**, *5*, 4213.
[38] E. Farhi, J. Goldstone, S. Gutmann, *arXiv:1411.4028* **2014**.
[39] E. Farhi, J. Goldstone, S. Gutmann, M. Sipser, *arXiv:quant-ph/0001106* **2000**.
[40] S. Utsunomiya, K. Takata, Y. Yamamoto, *Opt. Express* **2011**, *19*, 18091.
[41] C. C. McGeoch, C. Wang, Proceedings of the ACM International Conference on Computing Frontiers, **2013**, pp. 1–11.
[42] D. Aharonov, W. van Dam, J. Kempe, Z. Landau, S. Lloyd, O. Regev, *SIAM Rev.* **2008**, *50*, 755.
[43] G. Kochenberger, J.-K. Hao, F. Glover, M. Lewis, Z. Lü, H. Wang, Y. Wang, *J. Comb. Optim.* **2014**, *28*, 58.
[44] Y. Wang, X. Zhou, *Quantum Sci. Technol.* **2024**, *10*, 015056.
[45] V. Marx, *Nat. Methods* **2021**, *18*, 715.
[46] Y. Mizuno, T. Komatsuzaki, *Phys. Rev. Res.* **2024**, *6*, 013115.
[47] I. D. Leonidas, A. Dukakis, B. Tan, D. G. Angelakis, *Adv. Quantum Technol.* **2024**, *7*, 2300309.
[48] M. Schmid, S. Braun, R. Sollacher, M. J. Hartman, *Quantum Sci. Technol.* **2024**, *10*, 015051.
[49] D. Qu, E. Matwiejew, K. Wang, J. Wang, P. Xue, *Quantum Sci. Technol.* **2024**, *9*, 025014.
[50] D. Cividino, R. Westphal, D. Sornette, *Phys. Rev. Res.* **2023**, *5*, 013009.
[51] K. Tatsumura, R. Hidaka, J. Nakayama, T. Kashimata, M. Yamasaki, *IEEE Access* **2023**, *11*, 120023.
[52] W. Li, C. Wang, H. Wei, S. Hou, C. Cao, C. Pan, Y. Ma, K. Wen, *Quantum Sci. Technol.* **2025**, 025059.
[53] Z. Wang, A. Marandi, K. Wen, R. L. Byer, Y. Yamamoto, *Phys. Rev. A* **2013**, *88*, 063853.
[54] T. Honjo, T. Sonobe, K. Inaba, T. Inagaki, T. Ikuta, Y. Yamada, T. Kazama, K. Enbutsu, T. Umeki, R. Kasahara, K.-I. Kawarabayashi, H. Takesue, *Sci. Adv.* **2021**, *7*, abh0952.
[55] S. J. Wang, S. M. Shahidehpour, D. S. Kirschen, S. Mokhtari, G. D. Irisarri, *IEEE Trans. Power Syst.* **1995**, *10*, 1294.