electronics

MDPI

Article

Two-Stage Day-Ahead Scheduling for Coordinated Peak Shaving and Frequency Regulation in High-Renewable Low-Inertia Power Systems with Heterogeneous Energy Storage

Yuxin Jiang [1], Yufeng Guo [1,*], Junci Tang [2,3], Qun Yang [4], Yihang Ouyang [1], Lichaozheng Qin [1] and Lai Jiang [1]

[1] School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin 150001, China
[2] School of Electrical Engineering, Shenyang University of Technology, Shenyang 110178, China
[3] State Grid Liaoning Electric Power Supply Co., Ltd., Shenyang 110006, China
[4] State Grid Dalian Power Supply Co., Ltd., Dalian 116019, China
* Correspondence: guoyufeng@hit.edu.cn

Abstract

As power-electronic-interfaced renewable generation displaces synchronous machines, modern power systems face coupled day-ahead challenges: net-load variability demands peak shaving, while declining inertia necessitates explicit frequency-regulation scheduling. In sequential security-constrained unit commitment (SCUC) and Security-Constrained Economic Dispatch (SCED), the reserve procured in SCUC may lose deliverability after redispatch because the same storage bandwidth is reassigned to energy service. This paper proposes a two-stage day-ahead framework that addresses both challenges for low-inertia systems with high inverter-based resource (IBR) penetration. Stage I embeds Rate-of-Change of Frequency (RoCoF), frequency nadir, and quasi-steady-state (QSS) constraints in SCUC, with a piecewise-linear outer approximation for the non-convex nadir limit. Stage II strictly inherits the SCUC commitment and reserve reservation, and it applies bandwidth deduction to prevent peak-shaving redispatch from consuming committed frequency reserve. A technology-aware partition further assigns fast-response Lithium Iron Phosphate (LFP) batteries to sub-second frequency support and long-duration Vanadium Redox Flow Batteries (VRFBs) to energy shifting. Evaluated under the adopted reduced-order frequency-response framework and disturbance representation, tests on a modified IEEE 39-bus system under an extreme-wind scenario demonstrate that explicit frequency constraints eliminate all post-contingency violations, the inheritance mechanism closes a 23.85 MW reserve gap after redispatch, and heterogeneous storage partitioning preserves essentially the same disturbance sensitivity while increasing the peak-shaving ratio to 45.85%, lowering the day-ahead cost to CNY 10.483 × 10^6 and reducing the average system price to 209.33 CNY/MWh.

Keywords: high-penetration renewable energy; low-inertia power systems; frequency stability; coordinated peak shaving and frequency regulation; IBR–synchronous-machine interaction; reserve executability; security-constrained unit commitment; heterogeneous energy storage

Check for updates

Academic Editor: Ahmed Abu-Siada

Received: 8 April 2026
Revised: 21 April 2026
Accepted: 21 April 2026
Published: 23 April 2026

1. Introduction

The rapid integration of wind and photovoltaic (PV) generation is transforming both the energy balance and the dynamic behavior of modern power systems. Under high-penetration conditions, renewable output intensifies net-load fluctuations and therefore increases the need for day-ahead peak shaving. At the same time, the progressive displacement of synchronous generators by power-electronic-interfaced inverter-based resources (IBRs) erodes system inertia and weakens post-contingency frequency support capability. O'Sullivan et al. [1] quantified this effect for the Irish system by studying the maximum instantaneous non-synchronous generation that can be accommodated without compromising frequency stability, while Ela et al. [2] characterized the additional operating-reserve requirements introduced by variable generation at the system level. As a result, high-renewable operation creates a coupled scheduling problem: storage and other flexible resources must smooth the net-load profile and, at the same time, preserve sufficient headroom to satisfy Rate-of-Change of Frequency (RoCoF), frequency nadir, and quasi-steady-state (QSS) limits after a credible disturbance. In such systems, coordinated peak shaving and frequency regulation can no longer be deferred to real-time correction alone and must instead be addressed explicitly in day-ahead scheduling.

The literature on embedding frequency security in unit commitment has evolved through several methodological stages. Restrepo and Galiana [3] were among the first to introduce primary frequency regulation constraints into the unit commitment formulation, establishing a direct link between commitment decisions and post-contingency frequency performance. Ahmadi and Ghasemi [4] extended this approach by linearizing system frequency limit constraints within security-constrained unit commitment (SCUC), enabling tractable mixed-integer linear programming (MILP) embedding. Carrión and Arroyo [5] developed a computationally efficient MILP formulation for thermal unit commitment that has become a standard reference for commitment modeling. Teng et al. [6] proposed a stochastic scheduling framework with inertia-dependent fast frequency response requirements, linking reserve sizing to the time-varying inertia level determined by the commitment schedule. Chávez et al. [7] formulated a governor-rate-constrained optimal power flow for primary frequency control adequacy, showing that governor ramp limitations can bind before capacity limits under low-inertia conditions. Trovato et al. [8] investigated the role of flexible demand-side resources in supporting frequency stability in future low-carbon systems, demonstrating that thermostatically controlled loads can complement conventional generators in providing post-contingency response. Liu et al. [9] developed a framework for modeling frequency security constraints and quantifying frequency control reserve capacities within unit commitment, systematically linking RoCoF, nadir, and QSS limits to reserve sizing. Badesa et al. [10] proposed simultaneous scheduling of multiple frequency services in stochastic unit commitment, coordinating inertia, primary response, and enhanced frequency response within a unified formulation. Liang et al. [11] similarly addressed tractability for hybrid AC/DC microgrids by convexifying bi-directional converter models, reinforcing the broader imperative to preserve MILP tractability as renewable-side characteristics enter scheduling formulations. Yet no existing formulation simultaneously embeds explicit RoCoF, nadir, and QSS limits in SCUC, preserves MILP tractability, and accounts for storage that must co-serve both energy shifting and frequency regulation.

A second challenge arises in the sequential SCUC–SCED pipeline that governs day-ahead market clearing, regardless of how the Stage I reserve is sized. Chen et al. [12] provided a comprehensive survey of SCUC modeling, solution methods, and open challenges for electricity markets, documenting the prevailing sequential structure in which SCED follows SCUC. Chen et al. [13] focused on improving large-scale day-ahead SCUC performance through algorithmic enhancements, while Liu et al. [14] studied security-constrained economic dispatch with multi-stage rescheduling, examining how sequential redispatch adjustments affect system cost and constraint satisfaction. Within this sequential structure, the same storage power bandwidth is valuable to both peak-shaving energy service and frequency regulation. If the SCED does not strictly inherit the Stage I commitment and dynamic reserve reservation, the redispatch can absorb headroom that has already been reserved for frequency regulation. Moreira et al. [15] demonstrated that storage deliverability can be overstated in wholesale markets when service restrictions are not represented consistently across scheduling stages.

A third challenge concerns how coordinated peak-shaving and frequency-regulation duties should be allocated across heterogeneous storage technologies. Wen et al. [16] embedded frequency dynamics constraints in unit commitment with battery energy storage, demonstrating that storage can provide both energy and frequency services within a single scheduling framework. Shi et al. [17] showed that joint optimization of peak shaving and frequency regulation with battery storage yields superlinear economic gains compared with sequential service provision. Chen and Baldick [18] examined battery storage formulations and their impact on day-ahead SCUC, analyzing how different modeling choices affect commitment outcomes and storage utilization. Wang et al. [19] proposed a joint scheduling method for peak shaving and frequency regulation using hybrid energy storage with explicit degradation modeling. Hu et al. [20] developed day-ahead and hour-ahead optimal scheduling for battery storage at renewable energy stations participating in primary frequency regulation. Wen et al. [21] addressed the economic allocation of energy storage systems considering wind power distribution. However, treating storage as a uniform service pool obscures the fast-response/long-duration distinction, yielding schedules that are feasible but misaligned with constituent-technology characteristics.

Taken together, these unresolved issues constitute three interlinked research gaps in coordinated peak-shaving and frequency-regulation scheduling for power systems with high renewable penetration:

- G1. Tractable frequency-security representation in Stage I. Day-ahead SCUC for high-renewable systems requires explicit post-contingency RoCoF, nadir, and QSS constraints, yet existing formulations either omit these limits or sacrifice computational tractability.
- G2. Inter-stage reserve preservation during peak-shaving redispatch. Without strict inheritance of commitment and dynamic reserve reservation across scheduling stages, the same storage headroom may serve both peak-shaving and frequency-regulation functions, severing the link between scheduled and executable reserve.
- G3. Technology-aware service allocation for heterogeneous storage. Within a coordinated two-stage framework, heterogeneous storage technologies differ in response speed and energy duration; these differences should govern the allocation of peak-shaving and frequency-regulation duties rather than being masked by a uniform service pool.

1.1. Contributions

To address these gaps, this paper proposes a two-stage day-ahead scheduling framework for coordinated peak shaving and frequency regulation in low-inertia power systems with high renewable penetration. The principal contributions are as follows:

1. Explicit frequency-security modeling in Stage I (*G1*). RoCoF, nadir, and QSS stability constraints are formulated directly within the SCUC, capturing the post-contingency frequency dynamics that arise from the interaction between IBRs and the synchronous fleet. The non-convex nadir condition is resolved through a piecewise-linear outer approximation, preserving mixed-integer linear tractability.
2. SCED with strict inheritance of reserve and peak-shaving redispatch coordination (*G2*). A reservation-inheritance and bandwidth-deduction mechanism maintains frequency-reserve deliverability across the two scheduling stages. Reservation inheritance passes the Stage I commitment and frequency-service reservation to Stage II as

binding constraints, and bandwidth deduction prevents peak-shaving dispatch from consuming the power headroom already committed to frequency regulation.

3. Technology-aware heterogeneous storage partitioning (*G3*). LFP and VRFB units are assigned frequency-support and energy-shifting roles, respectively, that are consistent with their dynamic response speeds and energy durations. A four-case comparative study isolates and quantifies how this technology-aware partition improves service-role consistency and economic coordination while preserving the reserve executability established by the inter-stage inheritance mechanism.

1.2. Paper Organization

The remainder of this paper is organized as follows. Section 2 introduces the two-stage scheduling architecture, problem definition, and role differentiation of heterogeneous storage. Section 3 formulates Stage I, the frequency-secure SCUC with explicit RoCoF/QSS/nadir constraints and the PWL nadir approximation. Section 4 formulates Stage II, the peak-shaving-constrained SCED with reservation inheritance and bandwidth deduction. Section 5 describes the solution procedure and implementation, including a single-stage benchmark for computational comparison. Section 6 presents the case study following an evidence chain from frequency-security necessity through executability to heterogeneous-partition value. Section 7 concludes the paper.

2. System Model and Problem Statement

Day-ahead scheduling for power systems with high renewable penetration conventionally follows a sequential SCUC–SCED pipeline. Addressing the three gaps identified in Section 1 requires specifying (i) the scheduling architecture and inter-stage interface, (ii) the post-contingency frequency-security criteria that govern reserve sizing, and (iii) the technology-specific rationale for assigning distinct service roles to LFP and VRFB units. This section presents each of these elements in turn.

2.1. Two-Stage Scheduling Architecture

The proposed framework uses the sequential SCUC–SCED structure adopted in standard day-ahead operation [12], with an explicit inter-stage interface for reserve deliverability. The system comprises a set $\mathcal{G}$ of thermal generating units, a set $\mathcal{E} = \mathcal{E}_\ell \cup \mathcal{E}_v$ of energy storage units (LFP: $\mathcal{E}_\ell$; VRFB: $\mathcal{E}_v$), a set $\mathcal{W}$ of wind farms, a set $\mathcal{V}$ of photovoltaic (PV) plants, a transmission network with buses $\mathcal{N}$ and branches $\mathcal{L}$, and an aggregate load P_t^L (p.u.) served over $T = 96$ quarter-hour intervals ($\Delta t = 0.25$ h, 24 h horizon).

As shown in Figure 1, the scheduling pipeline proceeds in two stages. Stage I solves a frequency-secure SCUC (MILP) that determines unit commitment schedules, frequency reserve allocations, and storage service assignments over the 24 h horizon. Stage II solves a peak-shaving-constrained SCED (LP) that re-dispatches generation and storage subject to a hard peak-shaving constraint, extracts Locational Marginal Prices (LMPs) from the power-balance dual, and verifies that the energy dispatch does not use the reserve capacity awarded in Stage I.

The critical link between the two stages is the reservation inheritance mechanism: the commitment vector $\boldsymbol{u}^{\text{sc}}$, the storage frequency-service mode vector $\boldsymbol{z}^{\text{sc}}$, and the EFR reserve capacities $\boldsymbol{r}^{f,\text{sc}}$ determined in Stage I are passed to Stage II as fixed inherited parameters. A companion bandwidth deduction rule then prevents the SCED energy dispatch from consuming the power headroom already committed to frequency services, thereby closing the executability gap between reserve award and physical delivery.

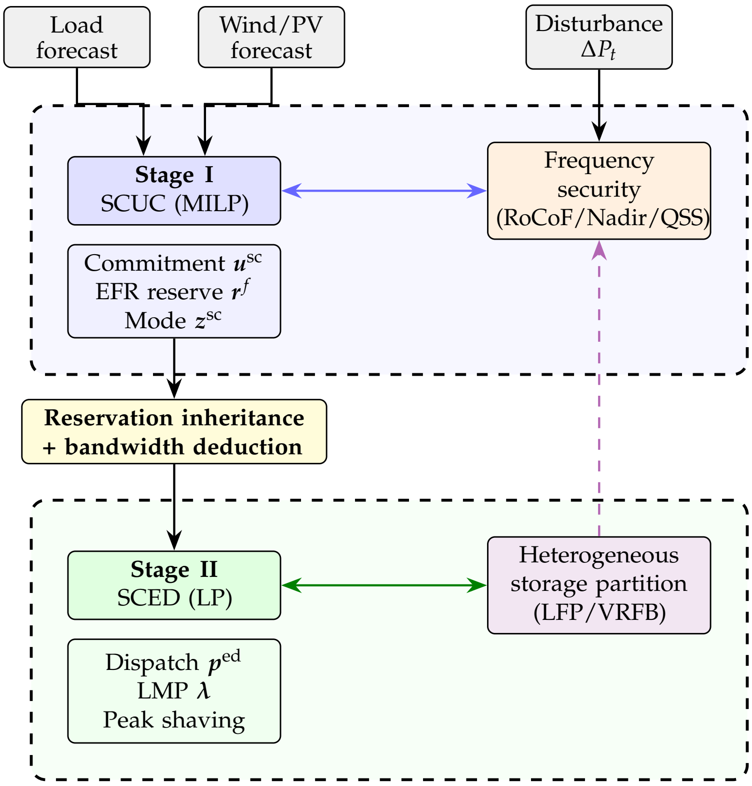

Figure 1. Two-stage scheduling architecture for coordinated peak shaving and frequency regulation under high renewable penetration.

Both stages are executed at day-ahead gate closure (typically 6–12 h before delivery) using 15 min forecasts of load, wind, PV, and the worst-case disturbance. Real-time balancing and intra-day re-dispatch are outside the scope of this work. For each interval t, the worst-case credible infeed loss is parameterized as ΔP_t (p.u.), defined as the largest coincident wind and PV output that could disconnect at a single network bus; its computation is detailed in Section 5.1. The temporal coordination between gate closure, the two optimization stages, and the delivery horizon are all shown in Figure 2.

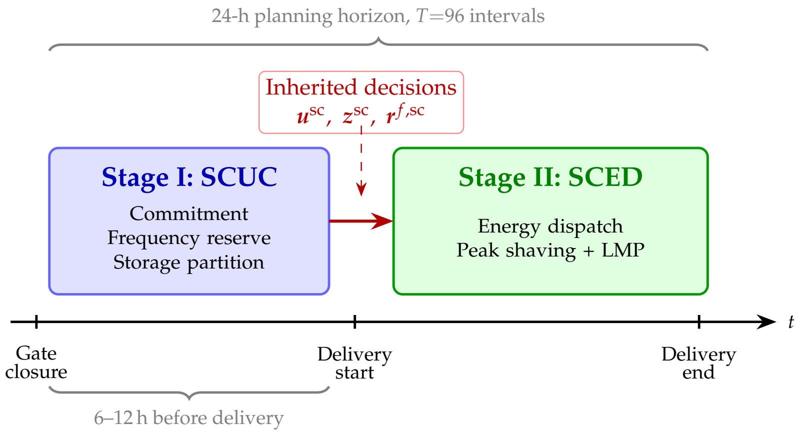

Figure 2. Day-ahead scheduling timeline and inter-stage data flow.

2.2. Frequency-Security Criteria

Instead of treating frequency support as a generic capacity product, the proposed framework ties the reserve requirement to three physically grounded post-contingency metrics—RoCoF, frequency nadir, and QSS deviation—all derived from a reduced-order system frequency response (SFR) model.

2.2.1. System Frequency Response Model

Following a sudden generation loss of magnitude ΔP_t (p.u.) at interval t, the per-unit rotor swing equation for the aggregated system is [22,23]

$$\frac{2H_t}{f_0}\frac{\mathrm{d}\Delta f}{\mathrm{d}t} = \Delta P_t - \Delta P_m(t) - D\,P_t^L\,\Delta f(t), \tag{1}$$

where $\Delta f(t)$ is the frequency deviation from nominal f_0, H_t (s) is the system equivalent inertia constant (load-weighted sum of committed synchronous machines), $\Delta P_m(t)$ is the combined mechanical-power increment from governor response and storage injection, and D (p.u./Hz) is the composite load-frequency damping coefficient.

In the simplified SFR model, the mechanical response is approximated by a governor ramp and a fast storage step [23]:

$$\Delta P_m(t) = R_t^G\,\frac{\min(\tau, t_R)}{t_R} + R_t^F\,\mathbf{1}(\tau \geq t_E), \tag{2}$$

where R_t^G is the total primary frequency response (PFR) capacity, R_t^F is the emergency frequency response (EFR) capacity, $t_R = 3.0$ s is the governor response time, $t_E = 0.5$ s is the EFR activation delay, and τ denotes elapsed time since the disturbance. Substituting (2) into (1) yields the governing frequency-trajectory equation:

$$\begin{aligned} 2H_t\frac{\mathrm{d}\Delta f}{\mathrm{d}t} = f_0\Big[\Delta P_t - R_t^G\frac{\min(\tau, t_R)}{t_R} \\ - R_t^F\,\mathbf{1}(\tau \geq t_E) - D\,P_t^L\,\Delta f\Big]. \end{aligned} \tag{3}$$

As illustrated in Figure 3, the resulting trajectory exhibits three characteristic phases: ① an inertial decline (slope set by H_t and ΔP_t) from $t = 0$ to EFR onset at t_E; ② a decelerating decline as governor and storage response build up to t_R; and ③ quasi-steady-state settling governed by the aggregate droop characteristic. One frequency-security metric is derived from each phase, as detailed below.

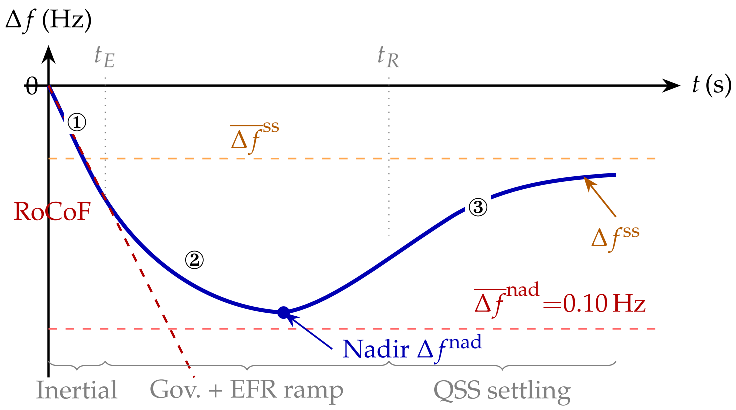

Figure 3. Post-disturbance frequency trajectory of the aggregated SFR model showing three characteristic phases.

2.2.2. RoCoF Metric

With the nominal frequency $f_0 = 50$ Hz and the composite load-frequency damping coefficient $D = 0.015$ p.u./Hz, the three security metrics are derived from (3) as follows.

Before any active response can act, the frequency decline is governed solely by system kinetic energy. The instantaneous RoCoF at $t = 0^+$ is

$$\left.\frac{\mathrm{d}f}{\mathrm{d}t}\right|_{t=0^+} = \frac{f_0\,\Delta P_t}{2H_t}, \tag{4}$$

and must satisfy $|\mathrm{d}f/\mathrm{d}t|_{t=0^+} \le \overline{\dot{f}} = 0.4$ Hz/s to prevent mal-operation of rate-of-change protection relays.

2.2.3. QSS Deviation Metric

After the full primary and EFR capacity has been deployed, the system settles at a quasi-steady-state deviation

$$\Delta f_t^{\mathrm{ss}} = \frac{\Delta P_t - R_t^G - R_t^F}{D\,P_t^L}, \tag{5}$$

which must satisfy $|\Delta f_t^{\mathrm{ss}}| \le \overline{\Delta f}^{\mathrm{ss}} = 0.01$ Hz.

2.2.4. Frequency Nadir Metric

The deepest instantaneous excursion occurs at $t^\star = (\Delta P_t - R_t^F)\,t_R/R_t^G$ and is approximated by

$$\Delta f_t^{\mathrm{nad}} = \frac{(\Delta P_t - R_t^F)^2\,t_R}{4\,H_t\,R_t^G/f_0}, \tag{6}$$

subject to $\Delta f_t^{\mathrm{nad}} \le \overline{\Delta f}^{\mathrm{nad}} = 0.1$ Hz to avoid under-frequency load shedding. The bilinear product $H_t \cdot R_t^G$ in the denominator renders direct MILP embedding non-trivial [4]; the piecewise-linear outer approximation developed in Section 3.6 resolves this non-convexity. Table 1 summarizes the three metrics.

Table 1. Frequency-security metrics and their SCUC embedding.

Metric	Physical Origin	Limit	Controlling Variable
RoCoF	Inertial response, (4)	0.4 Hz/s	System inertia H_t (via commitment $u_{i,t}$)
QSS	Primary + EFR, (5)	0.01 Hz	$R_t^G + R_t^F \ge \Delta P_t - DP^L \bar{\Delta f}$
Nadir	Governor + storage, (6)	0.10 Hz	$H_t \cdot R_t^G$ (bilinear; PWL approx.)

2.2.5. Reserve Deliverability Requirement

A committed reserve is physically deliverable only if two conditions hold: (i) the allocated capacity has not been simultaneously claimed for energy dispatch or peak shaving, and (ii) the storage SOC at the start of the contingency interval is sufficient to sustain the full injection over the required delivery window ($t_E = 0.5$ s for EFR; $t_R = 3.0$ s for PFR). The bandwidth deduction mechanism in Stage II enforces condition (i), while the SOC deliverability margin in the Stage I storage constraints enforces condition (ii) (Sections 4 and 3.7).

2.3. Heterogeneous Storage Roles and Partition Rationale

LFP batteries and vanadium redox flow batteries (VRFBs) are both grid-scale electrochemical storage technologies, yet their dynamic characteristics favor distinct service roles. LFP units achieve full power ramp within approximately 100 ms and exhibit high round-trip efficiency (∼95%), rendering them well suited to EFR delivery within the 0.5 s activation window and to frequent cycling. VRFB units require 1–3 s to reach rated output as electrolyte-pump pressure stabilizes, but they operate over a substantially wider SOC range (5–95%) and offer 4–12 h of energy-to-power duration. These properties make VRFB units more appropriate for sustained multi-hour discharge in peak shaving and energy shifting. Table 2 summarizes the representative technology parameters adopted in this study for scheduling-level role differentiation.

Table 2. Representative technology parameters for LFP and VRFBs adopted in this study.

Property	LFP Battery	VRFB
Energy-to-power ratio (h)	1–4	4–12
Response time	∼100 ms	∼1–3 s
Safe SOC range	[20%, 80%]	[5%, 95%]
Usable SOC range (p.u.)	0.60	0.90
Round-trip efficiency	∼95%	∼75%
Designated service role		
Fast frequency support	Primary	Secondary
Energy arbitrage	✓	✓
Peak shaving	Partial	Primary

If both technologies compete for fast reserve and long-duration energy services without explicit role differentiation, the optimizer may produce a low-cost schedule that does not respect their inherent response speed or energy duration characteristics. The proposed behavior partitioning policy addresses this by assigning each storage unit e a service-eligibility set

$$\mathcal{S}_e \subseteq \{\text{EFR}, \text{Energy}\}, \quad e \in \mathcal{E}, \tag{7}$$

with characteristic delivery windows

$$\tau_{\text{EFR}} = 0.5\ \text{s}, \qquad \tau_{\text{Energy}} = \text{intra-day}. \tag{8}$$

LFP units ($e \in \mathcal{E}_\ell$) are assigned to fast-response frequency support, with eligibility $\mathcal{S}_\ell = \{\text{EFR}, \text{Energy}\}$, whereas VRFB units ($e \in \mathcal{E}_v$) are limited to long-duration energy services. In Cases A–C, all storage units share the uniform eligibility $\mathcal{S}_e = \{\text{EFR}, \text{Energy}\}$; in Case D, the partitioned sets are enforced so that VRFB capacity is reserved for peak shaving rather than fast-response support.

3. Frequency-Security-Constrained Day-Ahead SCUC

This section presents the Stage I model. The explicit RoCoF, nadir, and QSS constraints provide the frequency-security basis for the executability mechanism introduced later in Section 4.

3.1. Objective Function

The Stage I SCUC co-optimizes thermal commitment, frequency reserve allocation, and storage service assignment over the 24 h horizon. The key sets and parameters are listed in Table 3, and the decision variables are summarized in Table 4. The objective minimizes a composite cost comprising thermal generation costs, storage degradation costs, renewable curtailment penalties, transmission violation penalties, and—depending on the case variant (defined in Section 6.2)—an LFP opportunity cost term and a nadir slack penalty:

$$\begin{aligned} \min \quad & \sum_{i\in\mathcal{G}}\sum_{t=1}^{T}\Big[c_i^0 u_{i,t} + \sum_{k=1}^{K_i} c_i^k\, p_{i,k,t} \\ & + c_i^{\text{su}}\,\delta_{i,t}^{+} + c_i^{\text{sd}}\,\delta_{i,t}^{-}\Big] \\ & + \sum_{e\in\mathcal{E}}\sum_{t=1}^{T}\left(c_e^d\, p_{e,t}^d + c_e^c\, p_{e,t}^c\right) \\ & + \lambda_w \sum_{w\in\mathcal{W}}\sum_{t}\left(P_{w,t}^{\text{av}} - p_{w,t}\right) \\ & + \lambda_v \sum_{v\in\mathcal{V}}\sum_{t}\left(P_{v,t}^{\text{av}} - p_{v,t}\right) \\ & + M_{\text{trs}} \sum_{\ell\in\mathcal{L}}\sum_{t}\left(s_{\ell,t}^{+} + s_{\ell,t}^{-}\right) \\ & + \underbrace{W_{\text{ps}}\textstyle\sum_{t=1}^{T}\Psi_t\, z_{e^\star,t}}_{\text{Case C only}} \\ & + \underbrace{M_{\text{nad}}\textstyle\sum_{t=1}^{T} s_t^{\text{nad}}}_{\text{Cases B/C}}, \end{aligned} \tag{9}$$

where c_i^0 (CNY) is the no-load cost of unit i, c_i^k (CNY/p.u.) is the incremental heat-rate coefficient for segment k of the piecewise-linear bid, c_i^{su} and c_i^{sd} (CNY) are the start-up and shut-down costs, c_e^d and c_e^c (CNY/p.u.) are the per-unit degradation cost rates for storage discharging and charging, λ_w and λ_v (CNY/p.u.) are renewable curtailment penalty prices that prioritize full utilization of available wind output $P_{w,t}^{\text{av}}$ and PV output $P_{v,t}^{\text{av}}$, $M_{\text{trs}} = 10^4$ CNY/p.u. is a large penalty that relaxes branch flow limits through slack variables $s_{\ell,t}^{+}, s_{\ell,t}^{-} \geq 0$, $e^\star \in \mathcal{E}_\ell$ denotes the LFP unit, $W_{\text{ps}} = 9100$ CNY/unit is the peak-shaving opportunity cost weight scaled by an urgency index Ψ_t, where Ψ_t denotes a normalized measure of the deviation of forecast net load from its daily mean (Section 3.7), $s_t^{\text{nad}} \geq 0$ is a soft slack for the nadir constraint, and $M_{\text{nad}} = 10^8$ CNY/p.u. is the associated penalty that renders the nadir limit effectively hard in practice.

Table 3. Key sets and parameters.

Symbol	Description	Value
$\mathcal{G}$	Set of thermal generators	$\lvert\mathcal{G}\rvert = 10$
$\mathcal{E}$	Set of storage units ($\mathcal{E}_\ell \cup \mathcal{E}_v$)	$\lvert\mathcal{E}\rvert = 2$
$\mathcal{W}$	Set of wind farms	–
$\mathcal{V}$	Set of PV plants	–
$\mathcal{T}$	Set of time periods	$T = 96$
f_0	Nominal frequency	50 Hz
S_b	System power base	50 MW
Δt	Dispatch interval	0.25 h
$\overline{\dot{f}}$	RoCoF limit	0.4 Hz/s
$\overline{\Delta f}^{\text{ss}}$	QSS limit	0.01 Hz
$\overline{\Delta f}^{\text{nad}}$	Nadir limit	0.1 Hz
D	Load damping coefficient	0.015 p.u./Hz
t_R	Governor response time	3.0 s
t_E	EFR response time	0.5 s
η_{ps}	Peak-shaving target	0.25
W_{ps}	Opportunity cost weight	9100 CNY

Table 4. Decision Variables (SCUC).

Symbol	Description	Type/Dim.
$u_{i,t}$	Thermal commitment state	Binary, $\mathcal{G} \times \mathcal{T}$
$p_{i,t}$	Thermal dispatch	Cont., p.u.
$p_{i,k,t}$	Segment k output	Cont., p.u.
$r^g_{i,t}$	Thermal PFR capacity	Cont., p.u.
$\delta^+_{i,t}, \delta^-_{i,t}$	Start/stop binary	Binary
$p^d_{e,t}, p^c_{e,t}$	Storage discharge/charge	Cont., p.u.
$s_{e,t}$	Storage SOC	Cont., p.u.·h
$r^f_{e,t}$	Storage EFR capacity	Cont., p.u.
$z_{e,t}$	Freq.-service mode	Binary, $\mathcal{E}_\ell \times \mathcal{T}$
$y_{e,h}$	Hourly mode lock	Binary
R^G_t, R^F_t	System PFR/EFR capacity	Cont., p.u.
H_t	System inertia constant	Cont., s

From a modeling perspective, c^d_e and c^c_e are interpreted as technology-specific equivalent throughput-cost parameters introduced to preserve tractability and provide simplified economic differentiation between storage technologies rather than detailed cycle-depth-dependent degradation processes. Scaled by Ψ_t, W_{ps} acts as an auxiliary coordination coefficient that quantifies the opportunity cost of assigning the fast-response LFP unit to peak-shaving duty during high-urgency intervals. The coefficients M_{trs} and M_{nad} are large security-enforcement penalties set well above ordinary dispatch-cost coefficients so that the associated branch-flow and nadir slack variables are activated only as last-resort feasibility measures, with M_{nad} maintained at a higher level to keep the nadir limit effectively hard.

3.2. Thermal Generator Constraints

The active power output of unit i is modeled by a piecewise-linear bid curve with K_i segments:

$$p_{i,t} = P_i^{\min} u_{i,t} + \sum_{k=1}^{K_i} p_{i,k,t}, \tag{10}$$

where $P_i^{\min}$ (p.u.) is the minimum stable output, $u_{i,t} \in \{0,1\}$ is the commitment status, and $p_{i,k,t}$ (p.u.) is the segment-k output bounded by $0 \le p_{i,k,t} \le \overline{P}_{i,k} u_{i,t}$. The total dispatch satisfies

$$P_i^{\min} u_{i,t} \le p_{i,t} \le P_i^{\max} u_{i,t}, \quad \forall i, t. \tag{11}$$

Inter-temporal ramp limits are

$$-R_i^d \le p_{i,t} - p_{i,t-1} \le R_i^u, \quad \forall i, t \ge 2, \tag{12}$$

where R_i^u and R_i^d (p.u./interval) are the up- and down-ramp rates. Minimum up/down time constraints require that, once a unit starts ($\delta^+_{i,t} = 1$) or stops ($\delta^-_{i,t} = 1$), it remains in the new state for at least T_i^u or T_i^d intervals:

$$\sum_{\tau=t}^{\min(t+T_i^u-1,\,T)} u_{i,\tau} \ge T_i^u \delta^+_{i,t}, \quad \forall i, t, \tag{13}$$

$$\sum_{\tau=t}^{\min(t+T_i^d-1,\,T)} (1-u_{i,\tau}) \ge T_i^d \delta^-_{i,t}, \quad \forall i, t, \tag{14}$$

with $\delta^+_{i,t} - \delta^-_{i,t} = u_{i,t} - u_{i,t-1}$.

3.3. Renewable and Storage Constraints

Wind dispatch is bounded by the available forecast and must satisfy a daily accommodation target of $\phi_w = 0.95$:

$$0 \le p_{w,t} \le P^{\text{av}}_{w,t}, \quad \sum_t p_{w,t} \ge \phi_w \sum_t P^{\text{av}}_{w,t}. \tag{15}$$

PV plants have analogous constraints. The storage SOC evolves as

$$s_{e,t} = s_{e,t-1} + \left(\eta^c_e\, p^c_{e,t} - p^d_{e,t}/\eta^d_e\right)\Delta t, \quad \forall e,\, t, \tag{16}$$

where η^c_e, η^d_e are the charging and discharging efficiencies, subject to the technology-specific SOC bounds (Table 2):

$$\underline{S}_e \le s_{e,t} \le \overline{S}_e, \quad \forall e,\, t. \tag{17}$$

Simultaneous charging and discharging is precluded by binary mode variables with $u^d_{e,t} + u^c_{e,t} \le 1$.

3.4. Power Balance and Network Constraints

System-wide power balance requires

$$\sum_{i\in\mathcal{G}} p_{i,t} + \sum_{e\in\mathcal{E}} (p^d_{e,t} - p^c_{e,t}) + \sum_w p_{w,t} + \sum_v p_{v,t} = P^L_t, \quad \forall t. \tag{18}$$

Branch flows follow the DC power flow model $f_{\ell,t} = b_\ell(\theta_{n,t} - \theta_{m,t})$, with transmission limits relaxed as $|f_{\ell,t}| \le F^{\max}_\ell + s^+_{\ell,t} + s^-_{\ell,t}$ via slack variables penalized at M_{trs} in (9).

3.5. Frequency-Security Constraints

The three frequency-security metrics derived in Section 2.2 are embedded as hard constraints in the SCUC formulation. These constraints are activated in Cases B–D; Case A omits them entirely to serve as a peak-shaving-only baseline.

The equivalent system inertia constant at interval t is determined by the committed thermal portfolio:

$$H_t = \frac{1}{S_b} \sum_{i\in\mathcal{G}} H_i\, P^{\max}_i\, u_{i,t}, \quad \forall t, \tag{19}$$

where H_i (s) is the inertia constant of unit i, $P^{\max}_i$ (p.u.) is its rated capacity, and S_b (MW) is the system base.

RoCoF constraint. Substituting (19) into the RoCoF metric (4) and requiring $|\mathrm{d}f/\mathrm{d}t|_{t=0^+} \le \overline{\dot{f}}$ yields a linear lower bound on the system inertia:

$$H_t \ge \frac{f_0\, \Delta P_t}{2\, \overline{\dot{f}}}, \quad \forall t, \tag{20}$$

where ΔP_t (p.u.) is the worst-case credible infeed loss at period t.

QSS constraint. Requiring $|\Delta f^{\text{ss}}_t| \le \overline{\Delta f}^{\text{ss}}$ in (5) produces a linear inequality on the aggregate primary and fast-response reserve:

$$R^G_t + R^F_t \ge \Delta P_t - D\, P^L_t\, \overline{\Delta f}^{\text{ss}}, \quad \forall t, \tag{21}$$

where R^G_t (p.u.) is the total PFR capacity and R^F_t (p.u.) is the total EFR capacity.

Nadir constraint. Unlike the RoCoF and QSS limits, which are enforced as hard linear inequalities, the nadir constraint is implemented with a soft slack s^{nad}_t penalized at $M_{\text{nad}} = 10^8$ CNY/p.u. in (9), rendering it effectively hard in practice while preserving solver feasibility. Bounding the frequency nadir (6) by $\overline{\Delta f}^{\text{nad}}$ gives:

$$\frac{(\Delta P_t - R_t^F)^2\, t_R}{4\, H_t\, R_t^G / f_0} \le \overline{\Delta f}^{\text{nad}} + s_t^{\text{nad}}, \quad \forall t. \tag{22}$$

Constraint (22) is non-convex due to the bilinear product $H_t \cdot R_t^G$ in the denominator and the quadratic numerator in R_t^F. The piecewise-linear outer approximation that resolves this non-convexity is developed in Section 3.6.

The system-level PFR and EFR capacities are aggregated from unit-level allocations:

$$R_t^G = \sum_{i \in \mathcal{G}} r_{i,t}^g, \quad R_t^F = \sum_{e \in \mathcal{E}} r_{e,t}^f, \quad \forall t, \tag{23}$$

where $r_{i,t}^g$ (p.u.) is the PFR capacity provided by thermal unit i and $r_{e,t}^f$ (p.u.) is the EFR capacity provided by storage unit e.

In addition to the three frequency-security constraints, an additional spinning-reserve adequacy requirement is enforced to cover the conventional synchronous-generator N-1 contingency, mandating that the total online headroom exceeds ΔP_t^{sg} (p.u.):

$$\sum_{i \in \mathcal{G}} (P_i^{\max} - p_{i,t})\, u_{i,t} + R_t^G + R_t^F \ge \Delta P_t^{\text{sg}}, \quad \forall t. \tag{24}$$

Here, ΔP_t^{sg} denotes the synchronous-generator N-1 contingency used in the reserve-adequacy check. Together with (20)–(22), constraint (24) jointly determines the minimum reserve allocation for the conventional synchronous-generator N-1 condition and the renewable-disturbance-driven frequency constraints.

3.6. PWL Outer Approximation for the Nadir Constraint

Rearranging (22), the nadir constraint can be expressed as a bilinear inequality in the equivalent variables H_t and R_t^G:

$$H_t\, R_t^G \ge \frac{f_0\, t_R\, (\Delta P_t - R_t^F)^2}{4\, \overline{\Delta f}^{\text{nad}}}, \quad \forall t. \tag{25}$$

The left-hand side is bilinear in the two continuous decision variables, and the right-hand side is convex quadratic in R_t^F. Direct embedding into an MILP is therefore not possible, and a three-step linearization procedure is adopted.

Step 1: Binary expansion of R_t^G. The PFR capacity R_t^G is discretized into P_b binary bits with a resolution step Δ_R (p.u.):

$$R_t^G = \sum_{p=1}^{P_b} w_p\, z_{p,t}, \quad w_p = \Delta_R\, 2^{p-1}, \quad z_{p,t} \in \{0, 1\}, \tag{26}$$

where $P_b = \lceil \log_2(R_{\max}^G / \Delta_R + 1) \rceil$ and $R_{\max}^G$ (p.u.) is the upper bound on the total PFR capacity.

Step 2: Big-M linearization of bilinear products. Auxiliary variables $m_{p,t} = H_t\, z_{p,t}$ and $\gamma_{p,t} = R_t^F\, z_{p,t}$ are introduced and linearized by the standard McCormick envelope tightened with big-M bounds:

$$0 \le m_{p,t} \le \overline{H}\, z_{p,t}, \tag{27}$$

$$H_t - \overline{H}(1 - z_{p,t}) \le m_{p,t} \le H_t, \tag{28}$$

$$0 \le \gamma_{p,t} \le R^F_{\max}\, z_{p,t}, \tag{29}$$

$$R^F_t - R^F_{\max}(1 - z_{p,t}) \le \gamma_{p,t} \le R^F_t, \tag{30}$$

where $\overline{H}$ (s) and $R^F_{\max}$ (p.u.) are the upper bounds on the system inertia and total EFR capacity, respectively.

Step 3: PWL cutting-plane bundle. The convex quadratic right-hand side of (25) is outer-approximated by $K = 10$ tangent hyperplanes evaluated at evenly spaced breakpoints $\rho_k = (k-1)R^F_{\max}/(K-1), k = 1, \ldots, K$. Defining $c_h = f_0 t_R / (4\overline{\Delta f}^{\text{nad}})$, the linearized nadir constraint at each breakpoint k reads

$$c_h \sum_p w_p\, m_{p,t} \ge (\Delta P_t - \rho_k)^2 + 2(\rho_k - \Delta P_t)(R^F_t - \rho_k) - s^{\text{nad}}_t, \tag{31}$$

for all $t \in \mathcal{T}$ and $k = 1, \ldots, K$. Together, (27)–(31) replace the non-convex constraint (22) with a set of linear inequalities that form an outer approximation of the feasible region: any solution satisfying all K cuts also satisfies the original nadir constraint, and the conservatism is controlled by the number of tangent planes. The binary expansion adds $P_b T = 576$ variables ($P_b = 6$, $T = 96$), a modest overhead.

3.7. Heterogeneous Storage Partition Constraints (Case D)

LFP mutual exclusion of frequency-service and energy modes. The binary indicator $z_{e,t} \in \{0,1\}$ selects between frequency-service mode ($z_{e,t} = 1$, all capacity reserved for EFR) and energy mode ($z_{e,t} = 0$, bulk dispatch enabled):

$$p^d_{e,t} + p^c_{e,t} \le P^{\text{cap}}_e\,(1 - z_{e,t}), \quad e \in \mathcal{E}_\ell,\ \forall t, \tag{32}$$

$$r^f_{e,t} \le P^{\text{cap}}_e\, z_{e,t}, \quad e \in \mathcal{E}_\ell,\ \forall t, \tag{33}$$

where P^{cap}_e (p.u.) is the rated power capacity of storage unit e. Constraint (32) precludes energy dispatch in frequency-service mode, while (33) restricts EFR allocation to intervals in which the unit is committed to frequency service.

Hourly mode continuity lock. To prevent impractical intra-hour toggling, the horizon is divided into $N_H = 24$ hourly windows of $W = 4$ intervals each, and the mode indicator is locked to a one-hourly variable $y_{e,h}$:

$$z_{e,t} = y_{e,h}, \quad t \in \{W(h-1)+1, \ldots, Wh\}, \quad h = 1, \ldots, N_H, \tag{34}$$

for $e \in \mathcal{E}_\ell$.

SOC deliverability margin. The SOC must retain enough energy to sustain the committed EFR capacity over one dispatch interval Δt:

$$s_{e,t} \ge \underline{S}^\ell_e + \Delta t\, r^f_{e,t} / \eta^d_e, \quad e \in \mathcal{E}_\ell,\ \forall t, \tag{35}$$

where $\underline{S}^\ell_e$ (p.u.·h) is the lower SOC bound for LFP. This margin ensures that even if the EFR service is fully activated throughout the period, the storage unit will not breach its energy floor.

VRFB exclusion from frequency-service mode. VRFB units are restricted to energy mode throughout the horizon because their slower response speed precludes effective EFR delivery:

$$r^f_{e,t} = 0, \quad z_{e,t} = 0, \quad e \in \mathcal{E}_v,\ \forall t. \tag{36}$$

Peak-shaving urgency index. The opportunity cost term in (9) uses a normalized urgency index

$$\Psi_t = \frac{|P_t^{NL} - \bar{P}^{NL}|}{\max_\tau |P_\tau^{NL} - \bar{P}^{NL}|}, \tag{37}$$

where $P_t^{NL} = P_t^L - \sum_w P_{w,t}^{\text{av}} - \sum_v P_{v,t}^{\text{av}}$ (p.u.) is the forecast net load and $\bar{P}^{NL}$ its daily mean. Because $\Psi_t \to 1$ at peak/valley periods, the penalty $W_{\text{ps}}\Psi_t z_{e^\star,t}$ discourages frequency-service commitment precisely when energy-mode dispatch is most valuable for peak shaving.

4. Peak-Shaving-Constrained Day-Ahead SCED with Reserve Preservation

The Stage I SCUC yields the commitment schedules $\boldsymbol{u}^{\text{sc}}$, the storage frequency-service mode assignments $\boldsymbol{z}^{\text{sc}}$, and the EFR reservations $\boldsymbol{r}^{f,\text{sc}}$. These decisions lock the unit on/off status and the frequency-service obligations for each storage unit. Stage II coordinates peak-shaving redispatch with the frequency-regulation commitment obtained in SCUC. Without an explicit inter-stage mechanism, the SCED—formulated as a pure LP—can freely reallocate any idle capacity, including headroom already committed to frequency-regulation services, to peak-shaving energy dispatch. The proposed Stage II therefore imposes two coupled requirements that are absent from a conventional economic dispatch: (i) the frequency-service reservations committed in Stage I serve as hard power-bandwidth constraints through reservation inheritance and bandwidth deduction, and (ii) a peak-shaving target constrains the residual net-load range. Together with Stage I, these requirements complete a two-stage day-ahead model for coordinated peak shaving and frequency regulation. This section formulates the physical dispatch, the bandwidth deduction mechanism, the peak-shaving constraint, and the pricing run that extracts locational marginal prices (LMPs).

4.1. Physical Dispatch Formulation

With commitment variables fixed at $u_{i,t} = u_{i,t}^{\text{sc}}$, the SCED reduces to a linear program (LP) that minimizes total dispatch cost:

$$\begin{aligned} \min \quad & \sum_{i\in\mathcal{G}} \sum_{t=1}^{T} \Big[c_i^0 u_{i,t}^{\text{sc}} + \sum_{k=1}^{K_i} c_i^k\, p_{i,k,t} \Big] \\ & + \sum_{e\in\mathcal{E}} \sum_{t=1}^{T} (c_e^d\, p_{e,t}^d + c_e^c\, p_{e,t}^c) \\ & + M_B \sum_{t=1}^{T} |P_t^{\text{mis}}| + M_T \sum_{\ell\in\mathcal{L}} \sum_{t=1}^{T} (s_{\ell,t}^+ + s_{\ell,t}^-), \end{aligned} \tag{38}$$

where the first line represents thermal no-load and piecewise-linear energy costs inherited from the SCUC cost structure, the second line accounts for storage charging and discharging costs, and P_t^{mis} (bounded by ± 0.1 p.u.) is the power-balance slack penalized at $M_B = 10^8$ to ensure near-exact balance. Transmission violations are penalized at $M_T = 10^6$. All thermal operating constraints (10)–(12) and storage SOC dynamics ((16) and (17)) are retained from Stage I with the integer variables fixed.

4.2. Reservation Inheritance and Bandwidth Deduction

To prevent the energy dispatch from consuming frequency headroom, the proposed framework enforces reservation inheritance: the EFR capacity allocated in Stage I is locked into Stage II as a fixed parameter rather than a decision variable. Specifically, for each storage unit e and period t, the EFR reservation $r_{e,t}^{f,\text{sc}}$ is passed from the SCUC solution to the SCED as a hard input:

$$r_{e,t}^f = r_{e,t}^{f,\text{sc}}, \quad e \in \mathcal{E},\ \forall t. \tag{39}$$

Building on this inheritance, the bandwidth deduction mechanism implements a priority-based shared-bandwidth scheme in which frequency-security service takes priority and energy dispatch can use only the remaining headroom. Rather than imposing complete separation between the two services, the available dispatch headroom on both the discharge and charge sides is reduced by the inherited EFR reservation

$$p^d_{e,t} \le P^{\text{cap}}_e - r^{f,\text{sc}}_{e,t}, \tag{40a}$$

$$p^c_{e,t} \le P^{\text{cap}}_e - r^{f,\text{sc}}_{e,t}, \tag{40b}$$

for all $e \in \mathcal{E}$ and $\forall t$. When the SCUC allocates a large EFR reservation ($r^{f,\text{sc}}_{e,t} \to P^{\text{cap}}_e$), the available dispatch headroom shrinks toward zero; conversely, when the EFR reservation is small, the storage unit retains substantial room for peak-shaving dispatch. This formulation allows frequency-security service and energy dispatch to coexist within the same rated capacity envelope while ensuring that the frequency-side obligation always has priority. In Cases A and B, where the executability closure is inactive, bandwidth deduction is disabled and the full rated power remains available for peak shaving ($p^d_{e,t} \le P^{\text{cap}}_e$), so that the performance gap relative to Cases C/D can be quantified.

In addition to the power-bandwidth constraint, the SOC must retain sufficient energy to sustain the inherited EFR commitment over the delivery window t_E:

$$s_{e,t} \ge \underline{S}_e + \frac{t_E}{\eta^d_e} r^{f,\text{sc}}_{e,t}, \quad e \in \mathcal{E}_\ell,\ \forall t, \tag{41}$$

where $\underline{S}_e$ is the lower SOC bound, and η^d_e is the discharge efficiency. This constraint guarantees that even if the EFR service is fully activated during any period, the storage unit will not violate its SOC floor.

4.3. Hard Peak-Shaving Constraint

The primary objective of the SCED storage dispatch is to flatten the residual net-load profile seen by thermal generators. The reference net load before storage dispatch is defined as

$$\text{NL}^{\text{ref}}_t = P^L_t - \sum_{w \in \mathcal{W}} P^{\text{av}}_{w,t} - \sum_{v \in \mathcal{V}} P^{\text{av}}_{v,t}, \tag{42}$$

and the residual net load after storage and renewable dispatch as

$$\text{NL}^{\text{res}}_t = P^L_t - p^w_t - p^v_t - \sum_{e \in \mathcal{E}} (p^d_{e,t} - p^c_{e,t}). \tag{43}$$

To impose the peak-shaving target as a hard constraint, auxiliary variables $\overline{nl}$ and $\underline{nl}$ are introduced to envelope the residual net load

$$\text{NL}^{\text{res}}_t \le \overline{nl}, \quad \text{NL}^{\text{res}}_t \ge \underline{nl}, \quad \forall t, \tag{44a}$$

$$\overline{nl} - \underline{nl} \le (1 - \eta_{\text{ps}})\, R^{\text{ref}}, \tag{44b}$$

where $R^{\text{ref}} = \max_t \text{NL}^{\text{ref}}_t - \min_t \text{NL}^{\text{ref}}_t$ is the pre-dispatch peak-valley range, and η_{ps} is the peak-shaving target (set to 25% for Cases A, C, and D). Constraint (44b) couples all $T = 96$ time periods through the range variables $\overline{nl}$ and $\underline{nl}$ and is the principal driver of the storage dispatch trajectory.

When bandwidth deduction is active (Cases C and D), the storage headroom available for peak shaving is reduced by the inherited frequency reservations. If this reduction is too severe, the peak-shaving constraint may become infeasible, signaling that the SCUC has over-committed frequency services at the expense of energy dispatch flexibility. This feedback mechanism ensures that the two stages reach a mutually consistent operating point.

4.4. Pricing Run and LMP Extraction

As a supportive system-level interpretive analysis, a separate pricing run is performed after the physical dispatch LP. Once the executable storage schedule $\boldsymbol{p}^{d\star}, \boldsymbol{p}^{c\star}$ is obtained, the pricing run re-solves the system dispatch with storage outputs fixed in order to extract nodal prices that reflect the marginal cost of serving load [24]. The pricing-run objective retains only thermal and renewable generation costs:

$$\begin{aligned} \min \quad & \sum_{i\in\mathcal{G}}\sum_{t=1}^{T}\Big[c_i^0 u_{i,t}^{\mathrm{sc}} + \sum_{k=1}^{K_i} c_i^k\, p_{i,k,t}\Big] \\ & + \sum_{w\in\mathcal{W}}\sum_t c_w^w\, p_{w,t} + \sum_{v\in\mathcal{V}}\sum_t c_v^v\, p_{v,t}, \end{aligned} \tag{45}$$

subject to thermal operating constraints (10)–(12), network power flow, and the power-balance equality with fixed storage injection:

$$\sum_{i\in\mathcal{G}} p_{i,t} + \sum_{w\in\mathcal{W}} p_{w,t} + \sum_{v\in\mathcal{V}} p_{v,t} = P_t^L - \sum_{e\in\mathcal{E}}(p_{e,t}^{d\star} - p_{e,t}^{c\star}), \quad \forall t. \tag{46}$$

Generation outputs are constrained to a narrow band around the physical-run solution ($\pm\delta$, with $\delta = 0.05$ p.u.) to preserve dispatch consistency while allowing marginal perturbation. The LMP at period t is then extracted from the dual variable $\lambda_t^\star$ of the power-balance equality (46)

$$\pi_t = \frac{\lambda_t^\star}{S_b\,\Delta t} \quad [\text{yuan/MWh}], \tag{47}$$

where S_b is the power base and Δt the time-step duration. Decoupling storage from the pricing run prevents price-taker storage from distorting LMPs and ensures that nodal prices reflect only the marginal cost of dispatchable generation. Because the storage schedule has already been determined by the physical run, the extracted LMPs represent the system's marginal cost of serving load given the executable storage dispatch—providing an economically meaningful price signal even though storage itself does not set the marginal price. The pricing analysis is supportive evidence for cross-case comparison rather than a core contribution of the framework. These extracted LMPs are interpreted here primarily as system-level signals of locational scarcity under the secured and executable dispatch, indicating where network, supply–demand, or security-related constraints become marginally binding.

5. Solution Procedure and Implementation

Sections 3 and 4 formulate the frequency-secure SCUC and the peak-shaving-constrained SCED as two coupled optimization stages. This section completes the implementation description by specifying (i) the disturbance model that parameterizes the frequency constraints, (ii) the two-stage clearing workflow, (iii) the approximation quality of the PWL nadir linearization, and (iv) the resulting problem scale and solver configuration. All models are implemented in MATLAB R2024a with YALMIP as the algebraic modeling layer and IBM CPLEX 12.10 as the back-end solver, executed on a workstation equipped with an Intel i7 processor and 32 GB RAM.

5.1. Disturbance Characterization

The frequency-security constraints (20)–(22) require a per-period disturbance magnitude ΔP_t (p.u.). In this study, ΔP_t is used as a time-varying renewable disturbance term under high wind/PV operating conditions so that the frequency-security requirements can track the renewable-side disturbance level over time. The conventional synchronous-generator N-1 condition is treated separately through the spinning-reserve adequacy constraint (24).

At each period t, the worst-case credible infeed loss is defined as the largest coincident wind-plus-PV output at any single network bus:

$$\Delta P_t = \frac{1}{S_b} \max_{b \in \mathcal{N}} \left[\sum_{w \in \mathcal{W}_b} P^{\mathrm{av}}_{w,t} + \sum_{v \in \mathcal{V}_b} P^{\mathrm{av}}_{v,t} \right], \tag{48}$$

where $\mathcal{W}_b$ and $\mathcal{V}_b$ are the sets of wind and PV plants connected at bus b, and S_b (MW) is the system base. Here, P^{av} denotes confidence-adjusted available wind/PV output obtained from an offline calibrated probabilistic forecasting module, rather than a deterministic point forecast. Accordingly, ΔP_t is defined as the largest credible coincident renewable infeed loss at a single bus for interval t. Because the confidence-adjusted output P^{av} is an upper bound on the actual dispatch (which may be curtailed), (48) provides a conservative but operationally meaningful disturbance estimate for all three frequency constraints. For the test scenario used in this study, ΔP_t ranges from 0.24 p.u. during the load valley to 1.504 p.u. (≈75 MW) during the peak wind period; the time-varying profile is computed offline from the renewable forecast and passed to the SCUC as a parameter array.

5.2. Two-Stage Scheduling Workflow

Algorithm 1 summarizes the complete clearing procedure for one case mode. For each of the four experimental cases (A–D) defined in Section 6.2, the same two-stage pipeline is executed with different constraint configurations.

Algorithm 1 Two-stage day-ahead scheduling workflow

Require: Forecast: P^L, P^{av}_w, P^{av}_v, ΔP_t; case mode $\in \{A, B, C, D\}$
Ensure: Commitment $\boldsymbol{u}^{\mathrm{sc}}$, dispatch $\boldsymbol{p}$, LMP $\boldsymbol{\pi}$

// Stage 1: Frequency-secure SCUC (MILP)
1: Compute ΔP_t from renewable forecast via (48)
2: Construct PWL cutting-plane bundle ((26)–(31))
3: Assemble MILP (9):
  base constraints (10)–(17), (18)
  + freq. security (20)–(23), PWL cuts (31) (B/C/D)
  + partition (32)–(37) (D)
4: Solve with CPLEX (gap $\leq 2 \times 10^{-4}$)
5: Extract $\boldsymbol{u}^{\mathrm{sc}}$, $\boldsymbol{z}^{\mathrm{sc}}$, $\boldsymbol{r}^{f,\mathrm{sc}}$
// Stage 2a: SCED physical run (LP)
6: Compute dispatch headroom via (40) (C/D)
7: Formulate SCED LP (38) s.t. (39), (44)
8: Solve; extract $\boldsymbol{p}^{d\star}$, $\boldsymbol{p}^{c\star}$, $\boldsymbol{s}^\star$
// Stage 2b: SCED pricing run (LP)
9: Fix storage at $\boldsymbol{p}^{d\star}$, $\boldsymbol{p}^{c\star}$; formulate (45)–(46)
10: Solve (dual simplex); extract LMP via (47)

In Stage 1, the frequency-secure SCUC is formulated as an MILP that jointly determines thermal commitment $\boldsymbol{u}^{\mathrm{sc}}$, EFR reservations $\boldsymbol{r}^{f,\mathrm{sc}}$, and—for Case D—the storage frequency-service mode vector $\boldsymbol{z}^{\mathrm{sc}}$. The PWL cutting-plane bundle (26)–(31) is constructed before the solve call, and the constraint set is assembled incrementally: Case A includes only (10)–(17) (commitment and storage constraints); Cases B/C add the frequency-security constraints (20)–(23) and the PWL nadir cuts; and Case D further adds the heterogeneous partition constraints (32)–(37). The MILP is solved by CPLEX with an MIP gap tolerance of 2×10^{-4}.

In Stage 2, the SCED is executed in two sequential linear programs. The physical run inherits the SCUC decisions via (39), applies bandwidth deduction (40) in Cases C/D, and minimizes the dispatch cost (38) subject to the peak-shaving constraint (44). The resulting optimal storage schedule $\boldsymbol{p}^{d\star}$, $\boldsymbol{p}^{c\star}$ is then fixed, and the pricing run (45)–(46) re-solves the thermal and renewable dispatch using the dual simplex method to extract stable LMP dual variables.

5.3. PWL Approximation Quality

The K-segment outer approximation of the nadir constraint (Section 3.6) introduces conservatism because the linearized feasible region is a strict subset of the true quadratic feasible set (cf. the hatched band in Figure 4). The maximum approximation error at the midpoint between two adjacent breakpoints is bounded by

$$\varepsilon_K \leq \frac{(\Delta P_{\max})^2\, t_R}{4\overline{\Delta f}^{\text{nad}}} \cdot \frac{1}{K^2}, \tag{49}$$

which decreases quadratically with the number of segments K.

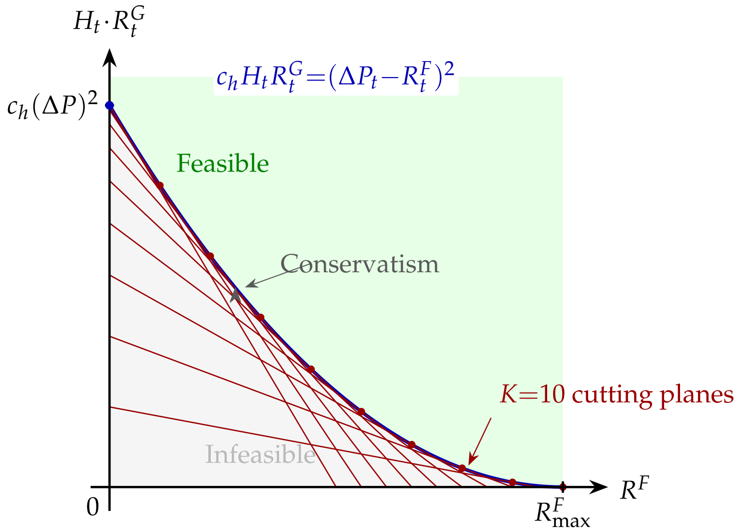

Figure 4. PWL outer approximation of the nadir feasibility boundary ($K = 10$ cutting planes).

To quantify the practical impact of the PWL approximation, Table 5 reports a Case D sensitivity sweep over $K \in \{5, 10, 20\}$ and binary discretization resolutions $P_b \in \{6, 10\}$. The solver times reported in Table 5 refer only to the Stage I SCUC MILP. Across all six settings, the day-ahead cost varies only from CNY 10.48272 to 10.48397 million (within 0.012%), while the time-accumulated EFR reservation remains within 554.71–613.32 MWh. Using the original nonlinear nadir expression for ex-post evaluation, the minimum nadir stays between 49.94205 and 49.94301 Hz, and the maximum violation of the 49.9 Hz limit remains zero in all cases. These results indicate that the reported economic and frequency-security outcomes are insensitive to moderate changes in both the number of PWL segments and the binary discretization resolution. In particular, at the practical discretization $P_b = 6$, changing K from 20 to 10 changes the day-ahead cost by only CNY 148 and leaves the ex-post nadir practically unchanged (49.94206 Hz versus 49.94219 Hz). A finer discretization $P_b = 10$ does not improve the ex-post nadir check, but it increases the solver burden substantially (e.g., 21.42 s at $K = 20$, $P_b = 6$ versus 92.80 s at $K = 20$, $P_b = 10$). Accordingly, $K = 10$ with $P_b = 6$ is adopted as the final accuracy–complexity trade-off in the revised study.

Table 5. Case D sensitivity of the nadir PWL approximation to K and P_b with Stage I SCUC solver times.

K	P_b	Cost (10^6 CNY)	Accum. EFR (MWh)	SCUC Solver (s)	Min Exact Nadir (Hz)	Max Violation (Hz)
5	6	10.48272	613.32	20.78	49.94301	0.00000
5	10	10.48358	590.41	134.55	49.94224	0.00000
10	6	10.48304	594.66	19.21	49.94219	0.00000
10	10	10.48314	560.13	88.48	49.94227	0.00000
20	6	10.48319	587.94	21.42	49.94206	0.00000
20	10	10.48397	554.71	92.80	49.94205	0.00000

5.4. Problem Scale and Solver Configuration

Table 6 summarizes the SCUC problem dimensions for each case variant. The columns reflect the three distinct constraint configurations: Case A (energy-only baseline), Cases B/C (frequency-secure with a uniform storage pool), and Case D (frequency-secure with heterogeneous partition).

Table 6. SCUC problem scale and solver settings.

Parameter	Case A	Cases B/C	Case D
Variable counts			
Binary (commitment)	960	960	960
Binary (mode z)	0	0	96
Binary (PWL)	0	576	576
Continuous	∼8 k	∼10 k	∼10 k
Constraint counts			
Thermal	∼15 k	∼15 k	∼15 k
Storage (SOC)	∼2 k	∼2 k	∼2 k
Network (DC OPF)	∼8 k	∼8 k	∼8 k
Frequency security	0	∼300	∼300
PWL cuts (KT)	0	960	960
Partition	0	0	∼500
Solver settings (CPLEX 12.10)			
MIP gap tolerance		2×10^{-4}	
Branching priority		commitment > mode > PWL	
Time limit		3600 s	

Case A contains only $|\mathcal{G}|T = 960$ commitment binaries. Cases B and C activate the frequency-security constraints and the PWL binary expansion, adding $P_bT = 576$ auxiliary binaries ($P_b = 6$, $T = 96$); this raises the explicit binary count from 960 to 1536 relative to Case A, but the PWL variables admit a tight LP relaxation (each $z_{p,t} \in [0, 1]$ in the McCormick envelope) that limits branch-and-bound overhead. Case D further introduces $|\mathcal{E}_\ell|T = 96$ frequency-service mode binaries and approximately 500 additional partition constraints ((32)–(36), (34) and (35)), a modest increment over Cases B/C.

The Stage II SCED consists of two sequential LPs (physical run and pricing run), each containing only continuous variables. The physical-run LP is solved by the barrier method, while the pricing-run LP is solved by the dual simplex method to ensure numerically stable dual variables for LMP extraction (47). In all cases, the SCED wall-clock time is under 60 s, and the total two-stage clearing time is dominated by the SCUC MILP. Typical overall SCUC–SCED computational times are reported in Section 6.8.

For computational reference only, a single-stage integrated benchmark is also constructed by embedding the peak-shaving constraint (44) directly into the SCUC MILP alongside the frequency-security constraints. This single large formulation co-optimizes commitment, frequency reserve, and energy dispatch in one pass. The benchmark serves to contextualize problem scale and solve time; it is not a main comparison case because it collapses the SCUC–SCED interface and therefore cannot test reserve executability. A comparison of the benchmark dimensions and solve statistics against the two-stage framework is presented in Section 6.8.

6. Case Study and Results

6.1. Test System Description

The proposed framework was validated on a modified IEEE 39-bus (New England) system derived from the MATPOWER case archive [25], with all parameters scaled to a 50 MW base. The test system comprises:

- 10 thermal generating units with a total installed capacity of 3600 MW;
- 1 LFP storage unit: 109 MW rated power, 199 MWh capacity (1.8 h duration);
- 1 VRFB storage unit: 125 MW rated power, 500 MWh capacity (4.0 h duration);
- Multiple wind farms with a peak available output of 1342 MW;
- PV plants with a peak available output of 353 MW;
- System load ranging from a valley of 2264 MW to a peak of 2939 MW.

The load and renewable profiles are drawn from an extreme-wind day recorded in northeast China in 2024. This particular day was selected because sustained high wind availability induces a pronounced low-inertia operating condition with prolonged low-net-load intervals, thereby providing a stringent stress test for evaluating coordinated peak shaving and frequency regulation under high renewable penetration. On the selected day, the aggregate renewable output (wind + PV) supplied approximately 46.8% of the daily energy demand, comprising 27,513 MWh of wind generation and 1506 MWh of solar generation. This penetration level places the system squarely in the regime where the IBR–synchronous-machine interaction governs post-contingency frequency dynamics, rendering the scenario directly relevant to the operational challenges addressed by high-penetration renewable energy research.

6.2. Experimental Design

Four cases were constructed to isolate the incremental contribution of each framework component, as summarized in Table 7. The experimental design followed a progressive activation pattern that traces a three-link evidence chain: Case A $\rightarrow$ Case B establishes the necessity of explicit frequency-security constraints; Case B $\rightarrow$ Case C establishes the necessity of executability closure via reservation inheritance and bandwidth deduction; and Case C $\rightarrow$ Case D quantifies the marginal value of heterogeneous storage partitioning.

Table 7. Definition of experimental cases.

Mechanism	Case			
	A	B	C	D
Peak-shaving only baseline	✓	×	×	×
Explicit frequency-security constraints	×	✓	✓	✓
Uniform storage pool	✓	✓	✓	×
Reservation inheritance (39)	×	×	✓	✓
Bandwidth deduction	×	×	✓	✓
Heterogeneous partition (LFP/VRFB)	×	×	×	✓

6.3. Performance Analysis of the Proposed Framework Under Case D

Prior to the comparative analysis, this subsection characterizes the operational profile of the proposed framework under the target scenario. Table 8 and Figure 5 summarize the deterministic solution of Case D, in which all four mechanisms are active: explicit frequency-security constraints, reservation inheritance, bandwidth deduction, and heterogeneous storage partitioning.

Table 8. Baseline performance of the proposed framework (Case D).

Metric	Value
RoCoF max. (Hz/s)	0.217
QSS deviation max. (Hz)	0.010
Nadir deviation max. (Hz)	0.058
Lowest frequency point (Hz)	49.942
Reserve deviation max. (MW)	≈0
VRFB/LFP frequency-service share (%)	1.04/100.00
VRFB/LFP energy-service share (%)	100.00/68.75
Peak-shaving ratio (%)	45.85
Ramp/curvature reduction (%)	53.91/55.84
Wind utilization (%)	95.00
Day-ahead/redispatch/total cost (CNY 10^6)	10.483/3.021/13.504
Average system price (CNY/MWh)	209.33

The results of Case D confirm that the full framework remains feasible and executable under the extreme-wind scenario. The maximum reserve deviation is effectively eliminated, verifying that the EFR capacity awarded in Stage I remains physically deliverable after Stage II redispatch. The resulting operating point is constrained yet secure: the RoCoF peak reaches 0.217 Hz/s (54% of the 0.400 Hz/s limit), the lowest post-contingency frequency remains at 49.942 Hz (above the 49.900 Hz threshold), and the QSS deviation attains the 0.010 Hz boundary without exceeding it. Both storage units return to approximately 50% state of charge by the end of the scheduling horizon, confirming that the schedule does not borrow energy from the subsequent operating day.

Stage II substantially reduces both the amplitude and the curvature of the residual load profile. Relative to the pre-dispatch profile, the adjusted net load attains a 45.85% peak-shaving ratio, a 53.91% reduction in aggregate ramp magnitude, and a 55.84% reduction in curvature, while maintaining 95% wind utilization. The dispatch absorbs wind-rich intervals, attenuates the steepest ramps, and preserves the headroom required for the inherited frequency reserve, demonstrating that the energy reshaping extends well beyond simple load-valley filling.

Figure 5 and Table 8 illustrate the heterogeneous partition in practice. The LFP unit carries all frequency-service participation periods (100% share), whereas the VRFB enters the frequency layer only as a marginal 1.04% fallback. The mode-occupancy panel further reveals that the VRFB remains in energy-only operation for 98.96% of the intervals, while the LFP spends 31.25% of the horizon in frequency-only mode and the remaining 68.75% in dual-service operation. The partition, therefore, anchors frequency support on the LFP while permitting it to contribute to peak shaving whenever additional load reshaping is economically beneficial. The accompanying nodal-price heat map provides supplementary economic information, with an average system price of 209.33 CNY/MWh and clustered high-price periods coinciding with peak load. Collectively, these baseline results demonstrate more than basic feasibility: the closed-loop framework coordinates peak shaving and frequency regulation while remaining secure, executable, and technology-consistent under a high-wind stress scenario.

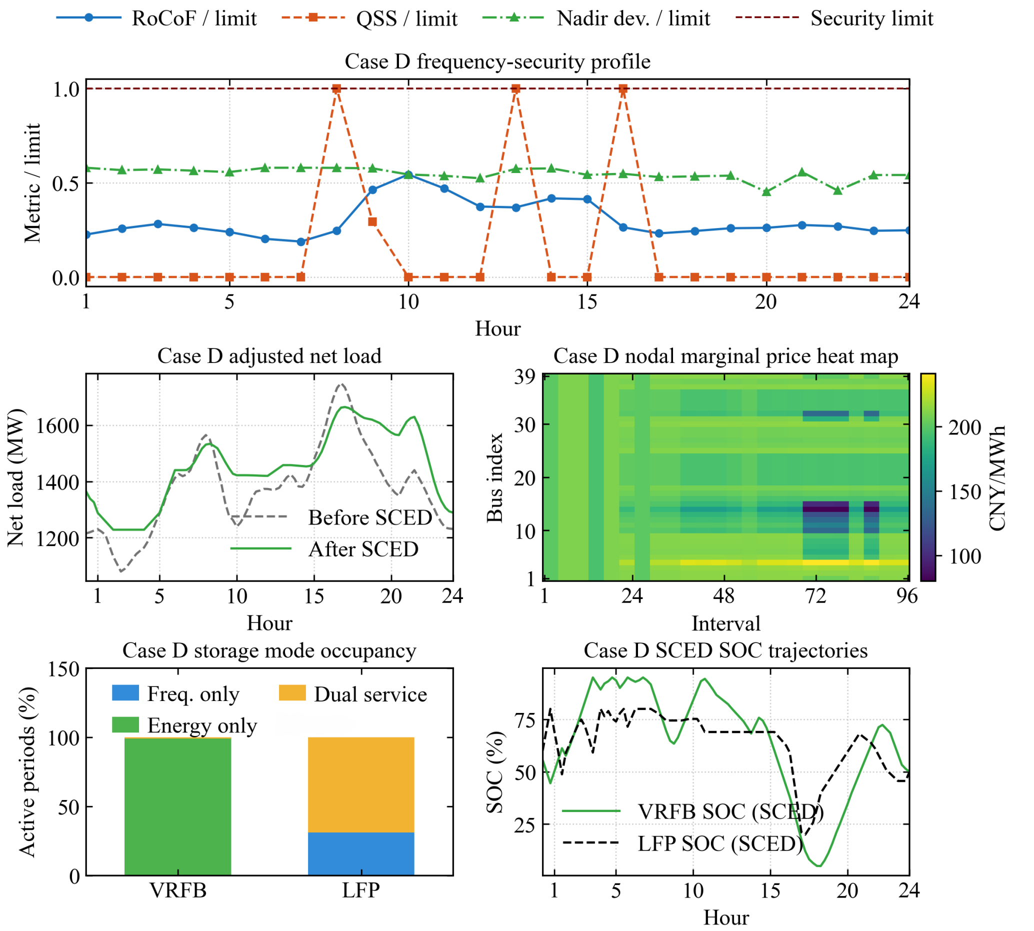

Figure 5. Baseline Benchmark results of the proposed framework for Case D.

6.4. Necessity of Explicit Frequency-Security Constraints

The first link in the evidence chain examines whether a peak-shaving schedule can be regarded as frequency-secure in the absence of explicit post-contingency frequency constraints. Table 9 reports the three frequency-security indicators for all four cases, while Figure 6 presents the two time-series signals most directly relevant to operational assessment: the RoCoF trajectory and the lowest post-contingency frequency.

Table 9. Frequency-security comparison.

Metric	Limit	Case A	Case B	Case C	Case D
RoCoF					
Max. value (Hz/s)	0.400	0.217	0.217	0.217	0.217
Violations/96	0	0	0	0	0
QSS Frequency Deviation					
Max. value (Hz)	0.010	1.391	0.000	0.000	0.010
Violations/96	0	96	0	0	0
Lowest Frequency Point					
Min. value (Hz)	49.900	49.348	49.946	49.946	49.942
Violations/96	0	96	0	0	0
Nadir Deviation					
Max. value (Hz)	0.100	0.652	0.054	0.054	0.058
Violations/96	0	96	0	0	0

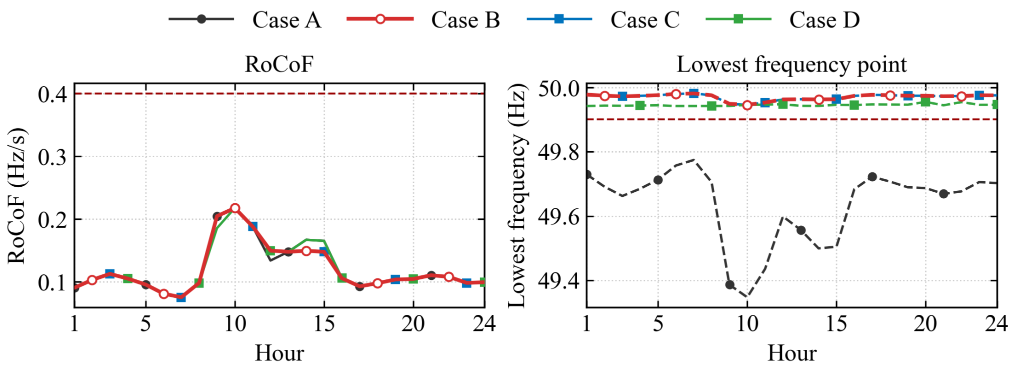

Figure 6. Frequency-security comparison across Cases A–D. The dotted horizontal lines indicate the corresponding frequency-security.

For the extreme-wind benchmark adopted in the main case study comparisons, the comparison between Cases A and B quantifies the impact of embedding explicit frequency-stability constraints in the day-ahead schedule under high renewable penetration. Although Case A successfully reduced net-load variability, it violated all three post-contingency stability metrics: the QSS deviation reached 1.391 Hz (139× the 0.010 Hz limit) across all 96 intervals, the nadir deviation peaked at 0.652 Hz (6.5× the 0.100 Hz limit), and the lowest post-contingency frequency deteriorated to 49.348 Hz—well below the 49.900 Hz threshold—in every interval. These results unequivocally confirm that energy-side reshaping alone is insufficient to guarantee post-contingency frequency stability when IBRs constitute a substantial share of the generation mix.

Upon embedding the RoCoF, QSS, and nadir constraints in SCUC (Case B), both the QSS and nadir violation counts fell from 96 to 0, and the lowest frequency recovered to 49.946 Hz. Notably, the RoCoF metric remained at 0.217 Hz/s across all four cases and was never binding; the dominant factor distinguishing Cases A and B is therefore the explicit reservation of PFR and EFR capacity rather than the system inertia level.

Cases C and D retained zero violation counts after the subsequent executability and partitioning mechanisms were activated, confirming that the Stage I security basis is preserved throughout the full two-stage pipeline. Relative to Cases B and C, Case D operates marginally closer to the admissible QSS and nadir boundaries: its QSS deviation reaches 0.010 Hz, its nadir deviation reaches 0.058 Hz, and its lowest post-contingency frequency decreases to 49.942 Hz, while the RoCoF peak remains at 0.217 Hz/s. This modest tightening is consistent with the more aggressive peak-shaving-oriented storage reallocation in Case D, yet all three frequency-security criteria remain fully satisfied.

Figure 7 further extends the assessment of Case D to three representative operating scenarios. The extreme-wind benchmark retains the baseline peak RoCoF of 0.217 Hz/s and the hourly minimum post-contingency frequency of 49.942 Hz. Under the maximum peak-valley and minimum-RE scenarios, the corresponding peak RoCoF values are reduced to 0.156 Hz/s and 0.156 Hz/s, while the hourly minimum frequency remains above 49.942 Hz in both cases. Hence, within the adopted representative disturbance envelope, the proposed framework preserves adequate frequency-security margins across materially different renewable-output and net-load conditions.

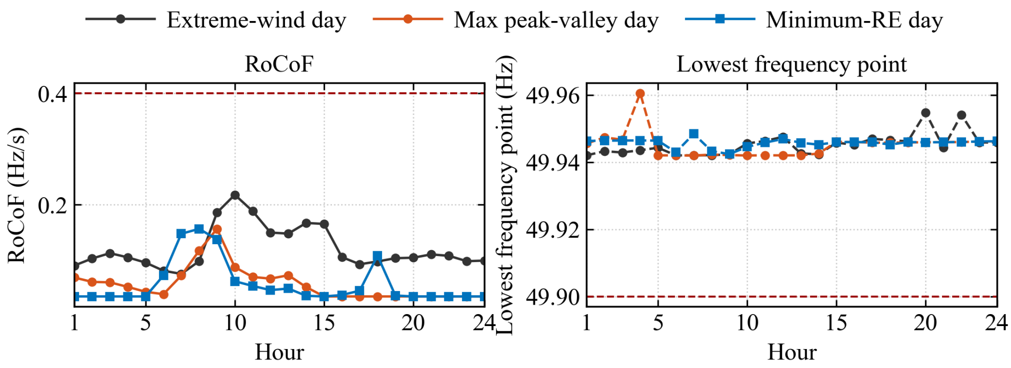

Figure 7. Case D frequency-security comparison across representative scenarios. The dotted horizontal lines indicate the corresponding frequency-security limits.

6.5. Reserve Executability Under Peak-Shaving Redispatch

The second link in the evidence chain examines whether the frequency-secure reserve awards obtained in SCUC remain physically deliverable after the SCED peak-shaving redispatch. The critical interface between the two stages can be characterized by comparing two quantities at each interval: the reserve obligation inherited from Stage I and the residual storage power bandwidth that remains after Stage II energy redispatch. An executability gap arises whenever the post-dispatch bandwidth falls below the inherited reserve requirement, regardless of the gross headroom available in other intervals.

Figure 8 visualizes this interface from the power-bandwidth perspective of the primary reserve-carrying storage unit. In each panel, the bars denote the realized SCED net power output, while the dashed envelopes delineate the time-varying executable power band that remains after subtracting the frequency reservation from the rated capacity. In Case B, this envelope is evaluated under the same conservative derating, margin, and activation-loss assumptions used in the ex-post deliverability audit; the red-shaded intervals mark where the realized dispatch exceeds the executable range. Although Case B satisfies all Stage I frequency constraints, it incurs a maximum reserve deviation of 23.85 MW (mean 2.79 MW) after SCED redispatch. The deviation does not originate from incorrect reserve sizing in SCUC, but rather from the fact that Stage II consumes the same physical headroom for peak-shaving energy dispatch without an explicit mechanism to safeguard the inherited reservation.

Table 10 provides a representative before/after comparison at the critical interval (17.0 h) by focusing on the reserve-carrying storage unit itself. In Case B, the unit inherits 23.85 MW of reserve from Stage I, but its post-SCED available headroom drops to 0.00 MW, resulting in a reserve shortfall of the same magnitude. After reservation inheritance and bandwidth deduction are activated, the mismatch disappears: in Case C, the inherited reserve and the remaining headroom are both 23.85 MW, while in Case D, the inherited reserve is 17.20 MW, and the remaining headroom is 18.75 MW, again with zero shortfall.

Table 10. Representative before/after executability comparison for the reserve-carrying storage unit at 17.0 h.

Case	Interval (h)	Inherited Reserve (MW)	Post-SCED Headroom (MW)	Shortfall (MW)
Case B	17.0	23.85	0.00	23.85
Case C	17.0	23.85	23.85	0.00
Case D	17.0	17.20	18.75	0.00

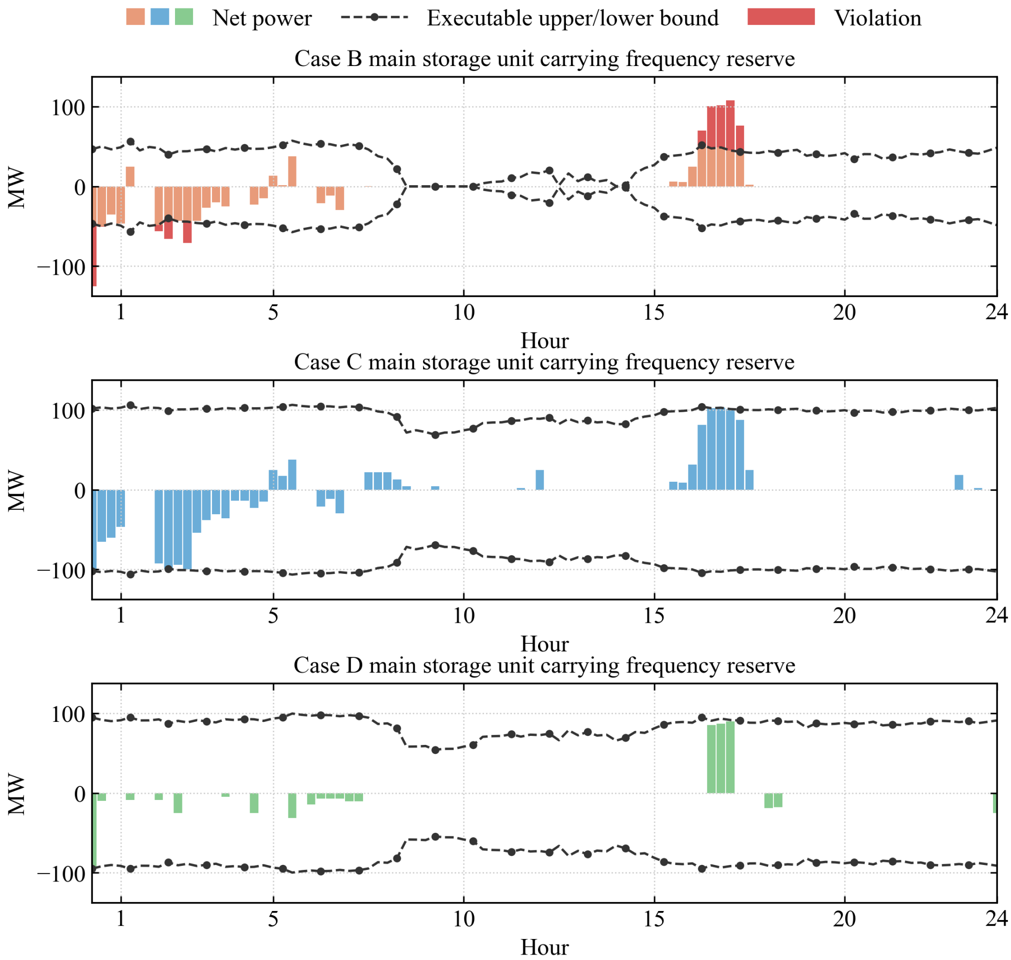

Figure 8. Reservation executability comparison.

Upon activating the reservation-inheritance and bandwidth-deduction mechanism in Case C, the maximum reserve deviation is reduced to numerical zero across all 96 intervals, as evidenced by the complete absence of red-shaded violations in the corresponding panel of Figure 8. This result confirms that the executability gap is closed entirely by the proposed inter-stage inheritance mechanism rather than by any modification of the Stage I reserve allocation itself.

Case D further corroborates that heterogeneous storage partitioning does not compromise this closure. Despite concentrating the fast-response burden on the LFP unit, the reserve deviation remains negligible, and the post-redispatch deliverable headroom reaches 217.20 MW, as reflected by the wide executable band in the rightmost panel. The result demonstrates that the proposed technology-aware role assignment preserves full reserve executability while enabling the service-differentiation benefits discussed in Section 6.6.

6.6. Value of Heterogeneous Storage Partitioning

The third link in the evidence chain investigates whether explicit heterogeneous partitioning alters the allocation of frequency and energy services across storage technologies.

As shown in Table 11 and Figure 9, Case C is executable, yet its storage roles are misaligned with the intended technology characteristics: the VRFB carries all frequency-service participation periods (100%), while the faster-responding LFP provides none (0%). Upon imposing the heterogeneous partition in Case D, the frequency-service burden shifts entirely to the fast-responding LFP (100% share), while the long-duration VRFB is relegated to the energy layer and participates in frequency service only as a minimal fallback (1.04% share). In the energy layer, the VRFB accounts for all energy-service participation periods

(100%), whereas the LFP contributes to energy service in 68.75% of the intervals. The mode-occupancy panel confirms that this reallocation does not dilute the LFP's frequency-led role: the VRFB operates almost exclusively in energy-only mode (98.96%), while the LFP alternates between frequency-only operation (31.25%) and dual-service operation (68.75%). This constitutes the principal operational value of the proposed partitioning mechanism: it renders the service allocation transparent and technology-consistent without compromising executability.

Table 11. Storage service-sharing statistics.

Metric	Unit	Case C	Case D
VRFB frequency-service share	%	100.00	1.04
LFP frequency-service share	%	0.00	100.00
VRFB energy-service share	%	42.71	100.00
LFP energy-service share	%	52.08	68.75

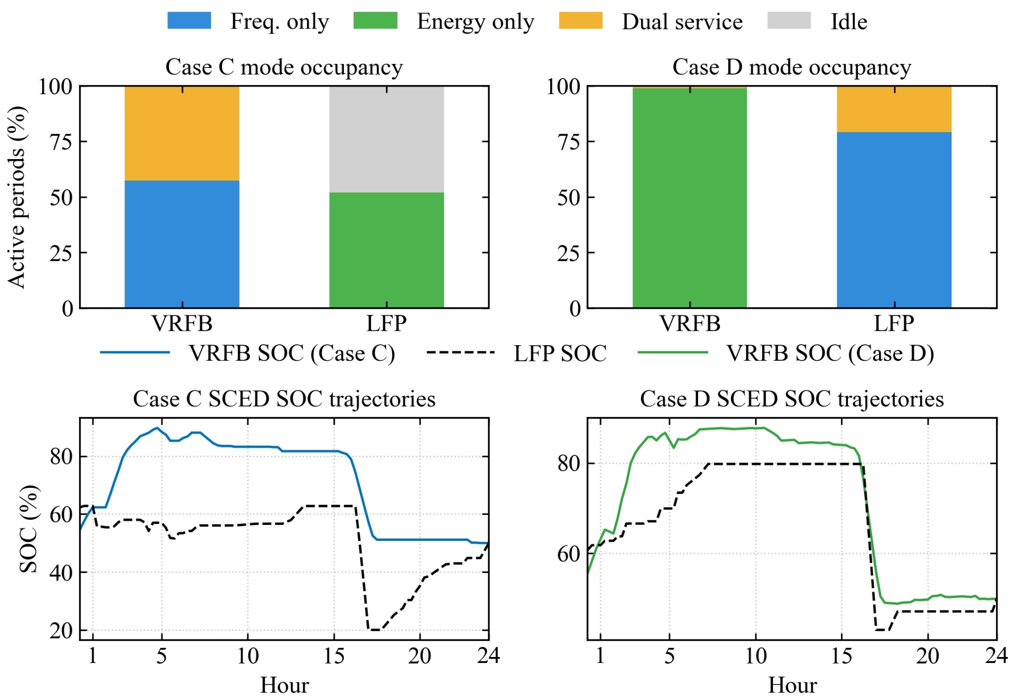

Figure 9. Service partition across storage technologies.

It is important to note that the transition from Case C to Case D should not be interpreted as a change in reserve quantity or as a precondition for executability—Case C already achieved zero reserve deviation. The maximum reserved EFR remained at 56.019 MW in both cases; only its composition changed: the optimizer concentrated fast-response duty on the LFP and released the VRFB capacity to the energy layer, consistent with the two technologies' response-speed and energy-duration characteristics. The value of heterogeneous partitioning, therefore, resides in rendering the dispatch more technology-consistent and economically coordinated while preserving the executability established by the inheritance mechanism.

The residual VRFB fallback also serves a structural purpose. Its frequency share in Case D is small but nonzero (1.04%, i.e., one interval out of 96), and its maximum reserved EFR contribution is only 1.64 MW, indicating that the partitioning mechanism does not impose a rigid hard separation. Instead, it establishes an asymmetric role structure: the LFP serves as the primary fast-response resource, while the VRFB functions as a secondary safeguard that can provide limited backstop support without dominating the frequency layer. In other words, when the optimization encounters binding constraints on the fast-response side, it can activate a bounded VRFB fallback to participate in primary frequency response. This role specialization is precisely what enables Case D to deepen the peak-shaving ratio without forfeiting the LFP-centered frequency mainline.

6.7. Economic Trade-Off

Table 12 and Figure 10 summarize the system-level economic indicators. To facilitate fair cross-case comparison, the total cost is decomposed into the day-ahead commitment cost, the Stage II redispatch cost, and an ex-post compliance penalty. The compliance penalty is computed as $C^{\text{pen}} = \kappa_f \sum_t v_t^f + \kappa_r \sum_t v_t^r$, where v_t^f (Hz) and v_t^r (MW) denote the per-period frequency-security and reserve-executability violations, respectively, and κ_f, κ_r are uniform penalty rates calibrated so that a fully compliant case incurs zero penalty. This metric is applied uniformly across all cases as an evaluation tool; it does not constitute a market payment or an in-model cost term.

Table 12. Economic performance comparison.

Indicator	Case A	Case B	Case C	Case D
Day-ahead cost (CNY 10^6)	10.491	10.491	10.491	10.483
Redispatch cost (CNY 10^6)	2.984	2.990	3.017	3.021
Compliance penalty (CNY 10^6)	0.220	0.119	0.000	0.000
Total cost (CNY 10^6)	13.696	13.601	13.508	13.504
Peak-shaving ratio (%)	34.46	36.78	39.37	45.85
Wind utilization (%)	95.00	95.00	95.00	95.00
Avg. system price (CNY/MWh)	210.53	210.84	210.42	209.33

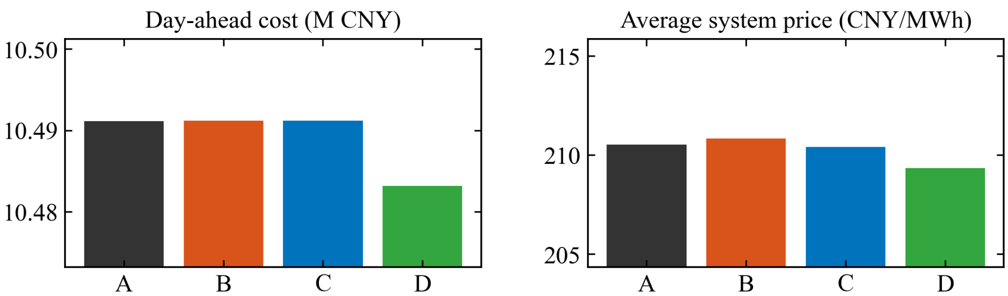

Figure 10. Economic comparison across Cases A–D.

Under the tested scenario, the penalty-related costs decrease monotonically as the security and executability mechanisms are progressively activated, whereas the final transition to heterogeneous partitioning further lowers the day-ahead cost and the average system price while deepening peak shaving, albeit with a shift in the composition of the ex-post total cost. Case A incurred the highest total cost (CNY 13.696 $\times$ 10^6) because the absence of frequency-security enforcement triggered the largest compliance penalty (CNY 0.220 $\times$ 10^6). Case B reduced that penalty to CNY 0.119 $\times$ 10^6 but still incurred a non-negligible executability penalty attributable to the 23.85 MW reserve deviation.

Once reservation inheritance and bandwidth deduction were activated in Case C, the compliance penalty vanished, and the schedule reached the lowest-cost feasible point among the uniform-pool cases. Relative to Case B, the executable peak-shaving ratio improved from 36.78% to 39.37%, and the total cost declined from CNY 13.601 $\times$ 10^6 to 13.508 $\times$ 10^6. This outcome underscores that the true cost of frequency-security coordination can only be assessed once the reserve award has been rendered executable.

Compared with Case C, Case D lowers the day-ahead cost from CNY 10.491×10^6 to 10.483×10^6 and reduces the average system price from 210.42 to 209.33 CNY/MWh, while increasing the peak-shaving ratio from 39.37% to 45.85%. Together with the near-identical sensitivity behavior reported in Section 6.9, this result demonstrates the operational value of heterogeneous partitioning: by assigning fast regulation to the LFP and releasing the VRFB to the energy layer, the framework deepens load reshaping and reduces both commitment cost and the average price level without degrading the disturbance-response level of the schedule. Although the stronger Stage II energy movement raises the redispatch cost slightly from CNY 3.017×10^6 to 3.021×10^6, this increase is more than offset by the reduction in day-ahead commitment cost, so the total evaluated cost further declines from CNY 13.508×10^6 to 13.504×10^6. The contribution of heterogeneous partitioning should therefore be interpreted as improving both the operational quality and the overall cost effectiveness of the schedule through a more suitable allocation of storage duties.

6.8. Computational Performance and Benchmark Discussion

Table 13 and Figure 11 report the overall SCUC–SCED computational times for the four main cases alongside a single-stage integrated benchmark.

All four cases converged successfully on the test platform, with solver times ranging from 11.24 s (Case A) to 46.70 s (Case D). Cases A and C are the fastest because they omit either explicit frequency-security coupling (Case A) or heterogeneous partitioning (Case C), whereas Cases B and D require additional computation owing to the supplementary security and service-allocation structure. Nevertheless, even the most demanding two-stage case remains well below the single-stage integrated benchmark.

Table 13. Computational performance summary of the SCUC–SCED framework.

Case	Solver Time (s)
Case A	11.24
Case B	35.83
Case C	11.95
Case D	46.70
Single-stage integrated benchmark	100.15

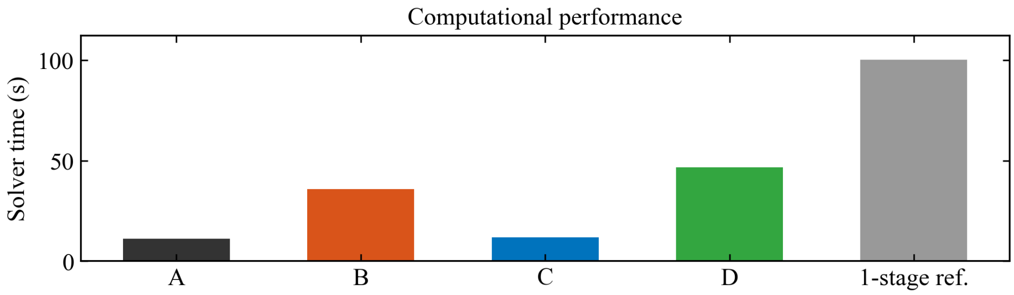

Figure 11. Computational performance of the SCUC–SCED framework and single-stage benchmark.

A single-stage benchmark is included solely as a computational reference. Its model dimensions are comparable to those of the proposed Case D formulation (1920 versus 1920 binary variables; 16,802 versus 16,800 continuous variables; 34,072 versus 32,439 constraints), yet it collapses peak-shaving and frequency-security couplings into a single monolithic MILP. The benchmark required 100.15 s, compared with 46.70 s for the proposed Case D. This comparison should not be construed as a claim that the two-stage framework is universally faster; the more significant distinction is architectural. The proposed framework preserves the SCUC–SCED interface explicitly, rendering reserve executability observable and enforceable, whereas the single-stage benchmark cannot isolate the source of infeasibility or over-commitment across scheduling stages.

6.9. Sensitivity Analysis

To assess the robustness of the foregoing conclusions, the frozen day-ahead schedules were stress-tested under a scaled disturbance sequence $\Delta P_t^{\text{scale}} = \alpha \, \Delta P_t$ with $\alpha \in [0.5, 2.0]$. The SCUC decisions are held fixed for each α; only the post-contingency frequency metrics are re-evaluated. The sensitivity test, therefore, measures the headroom embedded in the scheduled operating point rather than the system's ability to re-optimize. Because the RoCoF limit remains non-binding over the tested operating range while the nadir limit is the first frequency-security criterion to become active under disturbance scaling, the nadir-violation rate is adopted as the principal stress indicator in this subsection.

Figure 12 plots the nadir-violation rate as a function of the disturbance scaling factor. Case A remains highly vulnerable across the entire range, consistent with the absence of explicit frequency reserve. Cases B and C yield identical sensitivity curves because the executability closure modifies the SCED interface rather than the Stage I reserve sizing; their nadir-violation rate remains zero up to $\alpha = 1.0$ and rises sharply to 47.92% at $\alpha = 1.25$. Case D exhibits the same zero-violation region up to $\alpha = 1.0$ and subsequently follows the same nadir-violation trajectory as Cases B and C: the violation rate is likewise 47.92% at $\alpha = 1.25$ and reaches 100% for $\alpha \geq 1.5$. Simultaneously, the nadir deviations of Case D are consistently marginally smaller than those of Cases B and C (e.g., 0.2067 Hz versus 0.2171 Hz at $\alpha = 1.25$ and 0.6723 Hz versus 0.7063 Hz at $\alpha = 2.0$). The heterogeneous partition thus preserves essentially the same disturbance sensitivity as the executable uniform-pool cases, rather than materially weakening frequency robustness.

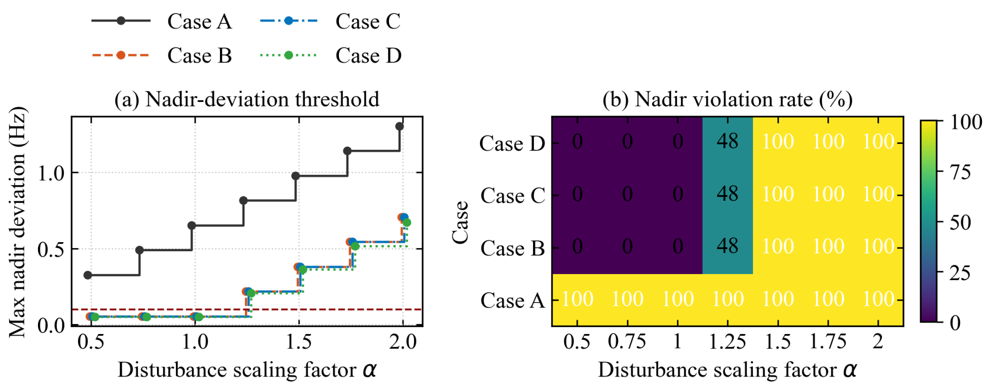

Figure 12. Sensitivity analysis under scaled disturbance stress. The red dotted horizontal line denotes the nadir-deviation security limit (0.10 Hz).

These results corroborate the interpretation drawn from the main cases. The executability closure does not, by itself, enlarge the intrinsic frequency-stability margin of the Stage I schedule; rather, it ensures that the scheduled reserve remains available after redispatch. Heterogeneous partitioning subsequently refines how that executable reserve is allocated across technologies. In the present setting, it maintains the disturbance sensitivity of Case D in close alignment with Cases B and C while simultaneously delivering a higher peak-shaving ratio, a lower day-ahead cost, and a lower average system price. The sensitivity analysis, therefore, complements the main-case findings: storage partitioning is valuable not because it alters the Stage I robustness class, but because it preserves that robustness while reallocating service duties toward a more effective and technology-consistent dispatch.

7. Conclusions

This paper has proposed a two-stage day-ahead scheduling framework for coordinated peak shaving and frequency regulation in low-inertia power systems with high renewable penetration. By embedding explicit post-contingency stability constraints in SCUC and preserving reserve deliverability throughout the SCED peak-shaving redispatch, the framework bridges the gap between frequency-security planning and executable operation. A systematic four-case study on a modified IEEE 39-bus system under an extreme-wind scenario yields the following principal findings:

1. Peak shaving alone is frequency-insecure. In the absence of explicit post-contingency constraints, the energy-only schedule (Case A) violated the QSS limit by a factor of 139 and drove the lowest frequency to 49.348 Hz in every interval. Embedding RoCoF, nadir, and QSS constraints (Case B) eliminated all 96 violations, confirming that frequency-security assurance in IBR-dominated systems cannot be inferred from energy-side reshaping alone.
2. Reserve awards require explicit inter-stage protection. Even after the Stage I reserve was correctly sized (Case B), the SCED redispatch consumed up to 23.85 MW of headroom already committed to frequency regulation. Activating reservation inheritance and bandwidth deduction (Case C) closed this executability gap entirely, reducing the maximum reserve deviation to a negligible level across all 96 intervals.
3. Technology-aware partitioning enhances service allocation without compromising executability. Under the tested scenario, the uniform-pool schedule (Case C) assigned all fast-response duty to the slower VRFB. Imposing heterogeneous partitioning (Case D) redirected frequency-service responsibility to the LFP (100% share), released VRFB capacity for energy shifting, increased the peak-shaving ratio to 45.85%, lowered the day-ahead cost to CNY 10.483×10^6, reduced the total evaluated cost to CNY 13.504×10^6, reduced the average system price to 209.33 CNY/MWh, and maintained zero reserve deviation with essentially the same disturbance sensitivity as the executable uniform-pool cases.

Collectively, within the adopted reduced-order frequency-response framework and disturbance representation considered in this study, these findings demonstrate that coordinated peak shaving and frequency regulation within sequential day-ahead scheduling for low-inertia systems hinge on three layered mechanisms: explicit frequency-security constraints furnish the reserve-sizing foundation; reservation inheritance and bandwidth deduction preserve that reserve throughout peak-shaving redispatch; and technology-aware heterogeneous partitioning aligns the dispatch with the physical characteristics of each storage technology. The proposed framework thereby provides a practical and transparent operational strategy for coordinating the interaction among IBRs, synchronous generators, and heterogeneous storage under high-renewable operating conditions.

Several directions warrant further investigation: (i) incorporating stochastic or robust optimization to account for renewable forecast uncertainty and its impact on frequency-stability margins; (ii) extending the contingency set to explicit $N - 1$ generator-trip scenarios in place of the current analytical disturbance proxy; (iii) validating the framework at higher penetration levels where the RoCoF constraint is expected to become the binding stability limit; (iv) examining the applicability of the proposed framework to hybrid AC/DC systems where converter-level dynamics introduce additional stability considerations; and (v) integrating data-driven approaches for predictive frequency-stability assessment that could inform adaptive reserve sizing within the proposed two-stage structure.

Author Contributions: Conceptualization, Y.J. and Y.G.; methodology, Y.J., Y.G. and J.T.; software, Y.J.; validation, Y.J., J.T., Q.Y., Y.O., L.Q. and L.J.; formal analysis, Y.J. and Y.G.; investigation, J.T. and Q.Y.; writing—original draft preparation, Y.J.; writing—review and editing, Y.G., J.T., Q.Y., Y.O., L.Q. and L.J.; supervision, Y.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Science and Technology Project of State Grid Corporation of China (2025YF-08).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data and code that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: Author Junci Tang was employed by State Grid Liaoning Electric Power Supply Co., Ltd., Shenyang, China; author Qun Yang was employed by State Grid Dalian Power Supply Co., Ltd., Dalian, China. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The authors declare that this study received funding from State Grid Corporation of China. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

Abbreviations

SCUC	Security-Constrained Unit Commitment
SCED	Security-Constrained Economic Dispatch
RoCoF	Rate of Change of Frequency
QSS	Quasi-Steady-State
EFR	Emergency Frequency Response
PFR	Primary Frequency Response
LFP	Lithium Iron Phosphate
VRFB	Vanadium Redox Flow Battery
PWL	Piecewise Linear

References

1. O'Sullivan, J.; Rogers, A.; Flynn, D.; Smith, P.; Mullane, A.; O'Malley, M. Studying the maximum instantaneous non-synchronous generation in an island system: Frequency stability challenges in Ireland. *IEEE Trans. Power Syst.* **2014**, *29*, 2943–2951. [CrossRef]
2. Ela, E.; Milligan, M.; Kirby, B. *Operating Reserves and Variable Generation*; Technical Report TP-5500-51978; National Renewable Energy Laboratory: Golden, CO, USA, 2011.
3. Restrepo, C.M.; Galiana, F.D. Unit commitment with primary frequency regulation constraints. *IEEE Trans. Power Syst.* **2005**, *20*, 1836–1842. [CrossRef]
4. Ahmadi, H.; Ghasemi, H. Security-constrained unit commitment with linearized system frequency limit constraints. *IEEE Trans. Power Syst.* **2014**, *29*, 1536–1545. [CrossRef]
5. Carrion, M.; Arroyo, J.M. A computationally efficient mixed-integer linear formulation for the thermal unit commitment problem. *IEEE Trans. Power Syst.* **2006**, *21*, 1371–1378. [CrossRef]
6. Teng, F.; Trovato, V.; Strbac, G. Stochastic scheduling with inertia-dependent fast frequency response requirements. *IEEE Trans. Power Syst.* **2016**, *31*, 1557–1566. [CrossRef]
7. Chavez, H.; Baldick, R.; Sharma, S. Governor rate-constrained OPF for primary frequency control adequacy. *IEEE Trans. Power Syst.* **2014**, *29*, 1473–1480. [CrossRef]
8. Trovato, V.; Teng, F.; Strbac, G. Role and benefits of flexible thermostatically controlled loads in future low-carbon systems. *IEEE Trans. Smart Grid* **2018**, *9*, 5067–5079. [CrossRef]
9. Liu, L.; Hu, Z.; Wen, Y.; Ma, Y. Modeling of frequency security constraints and quantification of frequency control reserve capacities for unit commitment. *IEEE Trans. Power Syst.* **2024**, *39*, 2080–2092. [CrossRef]
10. Badesa, L.; Teng, F.; Strbac, G. Simultaneous scheduling of multiple frequency services in stochastic unit commitment. *IEEE Trans. Power Syst.* **2019**, *34*, 3858–3868. [CrossRef]

11. Liang, Z.; Chung, C.Y.; Zhang, W.; Wang, Q.; Lin, W.; Wang, C. Enabling High-Efficiency Economic Dispatch of Hybrid AC/DC Networked Microgrids: Steady-State Convex Bi-Directional Converter Models. *IEEE Trans. Smart Grid* **2025**, *16*, 45–61. [CrossRef]
12. Chen, Y.; Pan, F.; Qiu, F.; Xavier, A.S.; Zheng, T.; Marwali, M.; Knueven, B.; Guan, Y.; Luh, P.B.; Wu, L.; et al. Security-constrained unit commitment for electricity market: modeling, solution methods, and future challenges. *IEEE Trans. Power Syst.* **2023**, *38*, 4668–4681. [CrossRef]
13. Chen, Y.; Casto, A.; Wang, F.; Wang, Q.; Wang, X.; Wan, J. Improving large scale day-ahead security constrained unit commitment performance. *IEEE Trans. Power Syst.* **2016**, *31*, 4732–4743 [CrossRef]
14. Liu, Z.; Ferris, M.C.; Zhao, J. Computational study of security constrained economic dispatch with multi-stage rescheduling. *IEEE Trans. Power Syst.* **2015**, *30*, 920–929. [CrossRef]
15. Moreira, A.; Singhal, N.G.; Kemp, J.M.; Ela, E.; Heleno, M. *Navigating Modeling Frontiers for Electric Storage Resources in Wholesale Electricity Markets*; Lawrence Berkeley National Laboratory: Berkeley, CA, USA, 2024.
16. Wen, Y.; Li, W.; Huang, G.; Liu, X. Frequency dynamics constrained unit commitment with battery energy storage. *IEEE Trans. Power Syst.* **2016**, *31*, 5115–5125. [CrossRef]
17. Shi, Y.; Xu, B.; Wang, D.; Zhang, B. Using battery storage for peak shaving and frequency regulation: joint optimization for superlinear gains. *IEEE Trans. Power Syst.* **2018**, *33*, 2882–2894. [CrossRef]
18. Chen, Y.; Baldick, R. Battery storage formulation and impact on day-ahead security constrained unit commitment. *IEEE Trans. Power Syst.* **2022**, *37*, 3995–4005. [CrossRef]
19. Wang, Y.; Zhao, X.; Huang, Y. Joint scheduling method of peak shaving and frequency regulation using hybrid energy storage considering degeneration characteristic. *IET Renew. Power Gener.* **2024**, *18*, 3765–3775. [CrossRef]
20. Hu, H.; Ma, Y.; Zhang, X.; Han, C.; Hao, Y. Day-ahead and hour-ahead optimal scheduling for battery storage of renewable energy power stations participating in primary frequency regulation. *Front. Energy Res.* **2024**, *12*, 1424389. [CrossRef]
21. Wen, S.; Lan, H.; Fu, Q.; Yu, D.C.; Zhang, L. Economic allocation for energy storage system considering wind power distribution. *IEEE Trans. Power Syst.* **2015**, *30*, 644–652. [CrossRef]
22. Kundur, P. *Power System Stability and Control*; McGraw-Hill: New York, NY, USA, 1994.
23. Anderson, P.M.; Mirheydar, M. A low-order system frequency response model. *IEEE Trans. Power Syst.* **1990**, *5*, 720–729. [CrossRef]
24. Kirschen, D.; Strbac, G. *Fundamentals of Power System Economics*; Wiley: Chichester, UK, 2004.
25. Zimmerman, R.D.; Murillo-Sanchez, C.E.; Thomas, R.J. MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Trans. Power Syst.* **2011**, *26*, 12–19. [CrossRef]