Distributed coordination for transmission-distribution systems with nonlinear flexibility aggregation

Xinliang Dai[1,2], Yanlin Jiang[2], Frederik Zahn[2], Yi Guo[3], Veit Hagenmeyer[2]

[1]Andlinger Center for Energy and the Environment, Princeton University, USA.
[2]Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, Germany.
[3]School of Automation, Beijing Institute of Technology, China.

arXiv:2608.06341v1 [eess.SY] 6 Aug 2026

Abstract

High shares of distributed energy resources (DERs) transform distribution systems into active participants in integrated transmission and distribution (ITD) operations. Linear models enable scalable distribution-level flexibility aggregation but can misclassify AC feasible operating points, whereas direct nonlinear aggregation becomes costly, especially in multiperiod ITD coordination. This paper reformulates transmission–distribution coordination within a hierarchical optimization framework and introduces a non-iterative predictor–corrector aggregation method. By leveraging path-following techniques from real-time optimal control, the approach achieves tractable computation with guaranteed error bounds. Across 24 radial distribution-network cases and seven meshed variants, including the real KIT Campus North grid, the proposed method yields substantially lower sampled false- and lost-flexibility rates than linear surrogates and a convex relaxation. On two 24-period ITD testcases, the formulation reduces end-to-end wall-clock time by factors of **6** relative to the corresponding centralized formulation, primarily through dimensionality reduction. An open-source implementation supports transparency and further research.

1 Introduction

The increasing penetration of distributed energy resources (DERs), including rooftop photovoltaics (PV), energy storage systems (ESS), and electric vehicles (EV) charging, is transforming distribution networks from passive loads into active participants in system operation, socalled active distribution network [1–3]. This transition introduces variable and distributed power injections driven by weather conditions and consumer behavior, and motivates closer transmission system operator (TSO)–distribution system operator (DSO) coordination, as also reflected in recent European regulations [4]. A central challenge is to coordinate power exchange at the TSO–DSO interface while accounting for controlling the DERs in the distribution system.

A straightforward approach to compute set points for the DERs is a fully centralized managed model, in which the TSO collects detailed information from all DSOs, and directly optimizes the dispatch of all controllable DERs across the entire integrated transmission-distribution (ITD) system [1]. In this way, it can effectively manage the entire ITD system as a single optimization problem. However, this approach is not feasible in practice. The joint optimization of multiple nonlinear distribution systems with potentially thousands of individual DERs quickly exceeds available computational capacities. Additionally, centralized management requires detailed consumer data, which creates privacy concerns and places substantial demands on communication infrastructure [5–7].

An alternative framework is the decomposition of the problem into two layers. This approach relies on flexibility aggregation to represent the set of feasible power exchanges at the coupling points between TSO and DSOs, which enables distributed coordination between transmission and distribution systems. In this framework, each DSO provides its flexibility set to the TSO. The DSO ensures that any operating point within this set can be securely accommodated in its network, while TSO schedules power exchanges without explicitly modeling distribution-level details. The aggregation-based interface can be computed locally by each DSO to communicate only compact flexibility sets to the TSO and, thereby, replaces detailed distribution-level models in the upper-level problem. It therefore enables

parallel preprocessing, reduces privacy and communication burdens, and lowers the dimension and complexity of the transmission-level NLP.

Despite these advantages, this non-iterative flexibility aggregation faces significant challenges. The nonlinear nature of distribution networks complicates the accurate representation of feasible regions. The presence of storage-based DERs further exacerbates this difficulty by introducing temporal coupling across multiple periods, making the problem both mathematically intricate and computationally demanding.

1.1 Related Work

Existing flexibility aggregation methods can be classified along two main axes: the power flow model used to represent distribution-network physics and the method used to project or approximate the feasible region in the TSO-DSO coupling space. Table 1 summarizes representative approaches according to these two axes.

On the modeling side, the most scalable approaches rely on linearized power flow models. These include DC power flow [8–10], enhanced DC power flow [11], LinDistFlow [12, 13], and linearized AC models [14, 15]. DC power flow is lossless and was originally developed for transmission systems, where branch reactances dominate resistances, differences between adjacent bus voltage angles are small, and voltage magnitudes are close to 1 p.u. [16–18]. LinDistFlow is a lossless simplification of the DistFlow model and is mainly suited to radial distribution networks [19–21]. Enhanced DC models improve the classical DC approximation by adding voltage magnitude or loss corrections [22], while linearized AC models use the Jacobian at an operating point to approximate the relation between injections and state variables. These linearized models enable tractable aggregation and have been extended to multi-period and three-phase settings [23–29], but their accuracy depends on the validity of the underlying assumptions.

On the aggregation side, several set representation methods have been proposed for the linearized models. Outer approximations provide supersets of the feasible region and may therefore require additional feasibility checks [8, 24]. Inner approximations provide guaranteed feasible but potentially conservative subsets [12, 25, 26]. Boundary detection methods sample or approximate the flexibility boundary [9–11, 14, 15, 23, 27, 28], but their cost can increase rapidly with the dimension of the coupling space. Projection-based methods, such as Fourier-Motzkin elimination, eliminate internal variables to obtain a representation in the coupling-variable space, but can become expensive for large multi-period models [29]. More recently, implicit-function-based methods have introduced temporal decomposition strategies for LinDistFlow-based aggregation [13].

However, using linearized models can compromise accuracy. The resulting aggregated sets may overestimate the feasible region, leading to infeasible dispatch commands, or underestimate it, leaving valuable flexibility unutilized [30]. Moreover, such modeling mismatches can accumulate over time in multi-period settings, as demonstrated in [31]. To improve accuracy, several works move beyond purely linear models. Direct nonlinear boundary detection has been proposed in [32], but it becomes computationally expensive for large networks. DistFlow-based methods can capture nonlinear effects and provide approximations of flexibility regions [31, 33], but their reliance on branch flow formulations restricts them mainly to radial networks [21]. Approximate-dynamic-programming methods approximate value functions of local distribution-level subproblems [34, 35], but face scalability limitations. Convex inner approximations of the AC power flow model have been developed [36]. Originally intended for uncertainty management, these methods often produce overly conservative flexibility sets, which constrain their practical value in aggregation.

Table 1: Summary of Aggregation Methods and Power Flow Models

Ref	System Model	Network Topology	Time Period	Temporal Decomp.	Aggregation Method	Nonlinear AC Physics
[8]	DC power flow	mesh	single	-	Outer approximation	-
[9, 10]	DC power flow	mesh	single	-	Boundary detection	-
[23]	Enhanced DC power flow (3-phase)	mesh	single	-	Boundary detection	-
[11]	Enhanced DC power flow	mesh	single	-	Boundary detection	Yes
[24]	Enhanced DC power flow	mesh	multiple	No	Outer approximation	-
[25, 26]	Enhanced DC power flow (3-phase)	mesh	multiple	No	Inner approximation	-
[27, 28]	Enhanced DC power flow (3-phase)	mesh	multiple	No	Boundary detection	-
[29]	Enhanced DC power flow (3-phase)	mesh	multiple	No	Fourier–Motzkin elimination	-
[14, 15]	Linearized AC power flow	mesh	multiple	No	Boundary detection	-
[12]	LinDistFlow	radial	multiple	No	Inner approximation	-
[13]	LinDistFlow	radial	multiple	Yes	Implicit function reformulation	-
[32]	DistFlow	radial	single	-	Boundary detection	Yes
[33]	DistFlow	radial	single	-	Analytical expressions	Yes
[31]	LinDistFlow	radial	single	-	Quadratic loss compensation	Yes
[34]	LinDistFlow with linearized losses	radial	single	-	Approximate dynamic programming	Yes
[35]	AC power flow	mesh	single	-	Approximate dynamic programming	Yes
[36]	AC power flow	mesh	single	-	Convex restriction	Yes
This paper	AC power flow	mesh	multiple	Yes	Predictor-corrector	Yes

*All the models using small-angle assumptions to linearly approximate the sin and cos functions are classified as DC power flow.
*Enhanced DC power flow includes voltage difference and/or line losses compared with the standard DC model.

1.2 Challenges and Contributions

Existing methods face a trade-off between scalability, AC accuracy, topology generality, and temporal decomposition. Linearized methods are scalable but may lose accuracy; nonlinear methods are more accurate but are often restricted to radial networks, single-period settings, or computationally expensive formulations. Motivated by this gap, we formulate the nonconvex ITD coordination problem within a graph-based hierarchical optimization framework that reflects the layered TSO-DSO structure.

Within this framework, we adapt predictor–corrector techniques—originally developed for real-time path-following control problems [37, 38]—to approximate distribution-level flexibility sets. The proposed approach is conceptually related to recent work in generic nonlinear programming (NLP) [39, 40], where implicit function reformulations are used to iteratively condense derivatives, thereby improving the reliability of interior-point methods [41, 42] and accelerating computation [43, 44]. In contrast to these iterative solver-level reductions, the present paper uses a non-iterative predictor-corrector construction to precompute local surrogates of distribution-level implicit feasible sets. The resulting aggregation directly accounts for nonlinear AC power flow constraints, is applicable to both radial and meshed distribution networks, and enables scalable pre-computation through spatial and temporal decomposition.

The main contributions of this paper are in the following:

1. We reformulate the ITD dispatch problem into a hierarchical optimization framework and introduce a predictor–corrector aggregation scheme for distribution networks' nonconvex implicit feasible sets. We establish error bounds and validate accuracy on tutorial examples and extensive MATPOWER benchmarks, covering both radial and meshed topologies.
2. We extend the approach to multi-period operation by constructing a coupled power–energy envelope that preserves temporal consistency while enabling parallel preprocessing across spatial and temporal scales. This reduces dimensionality and achieves $5-7$ times speedups compared to centralized state-of-the-art NLP solvers such as IPOPT [45].

After notations[1], Section 2 presents the problem formulation for ITD coordination, including the graph representation of the network, the AC power flow models, the TSO–DSO management models, and the corresponding hierarchical optimization framework. Section 3.1 introduces the proposed predictor-corrector method for nonlinear flexibility aggregation and presents the corresponding single-period benchmarks. Section 4 extends the method to multi-period scenarios. Finally, Section 5 concludes the paper.

2 Problem Formulation for ITD Coordination

This section introduces the modeling ingredients underlying the coordination of integrated transmission-distribution (ITD) systems. We first describe the graph representation of the network, then present the AC power flow models of the transmission and distribution subsystems, and discuss the corresponding TSO-DSO management models. Based on these ingredients, we formulate the ITD coordination problem as a graph-based hierarchical optimization problem and introduce the associated implicit feasible sets that motivate the aggregation method developed in the next section.

2.1 Graph Representation of Power Networks

A power network can be represented as an undirected graph $\mathcal{G} = (\mathcal{N}, \mathcal{L})$, where $\mathcal{N}$ denotes the set of buses and $\mathcal{L}$ denotes the set of branches.

To represent the subsystem structure of the ITD network, let $\mathcal{S}^{\mathrm{T}} = \{0\}$ and $\mathcal{S}^{\mathrm{D}}$ denote the index sets of the transmission and distribution subsystems, respectively, and define

$$\mathcal{S} := \mathcal{S}^{\mathrm{T}} \cup \mathcal{S}^{\mathrm{D}}.$$

The overall ITD network is represented by the graph

$$\mathcal{G} = (\mathcal{N}, \mathcal{L}),$$

[1] *Notations:* Let $\mathbb{R}^n$ denote the n-dimensional real Euclidean space, and $\mathbb{C}^n$ the n-dimensional complex Euclidean space. For two column vectors, $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$, their concatenation is denoted by $(x, y) := [x^\top, y^\top]^\top$. For a set $\mathcal{S}$ with cardinality $|\mathcal{S}|$, the concatenation of subvectors x_i across all $i \in \mathcal{S}$ is denoted by $\{x_i\}_{i\in\mathcal{S}} := (x_1, \cdots, x_{|\mathcal{S}|})$. Distance from a point x to a set $\mathcal{S} \subseteq \mathbb{R}^n$ is denoted by $\mathrm{dist}(x, \mathcal{S}) := \inf_{s\in\mathcal{S}} \|x - s\|$. Following [46], the Hadamard (element-wise) product of vectors $x, y \in \mathbb{R}^m$ is denoted by $x \circ y$, with elements given by $[x \circ y]_i := [x]_i \cdot [y]_i$. Element-wise squaring of a vector $x \in \mathbb{R}^m$ is denoted by $x^{\circ 2} := x \circ x$.

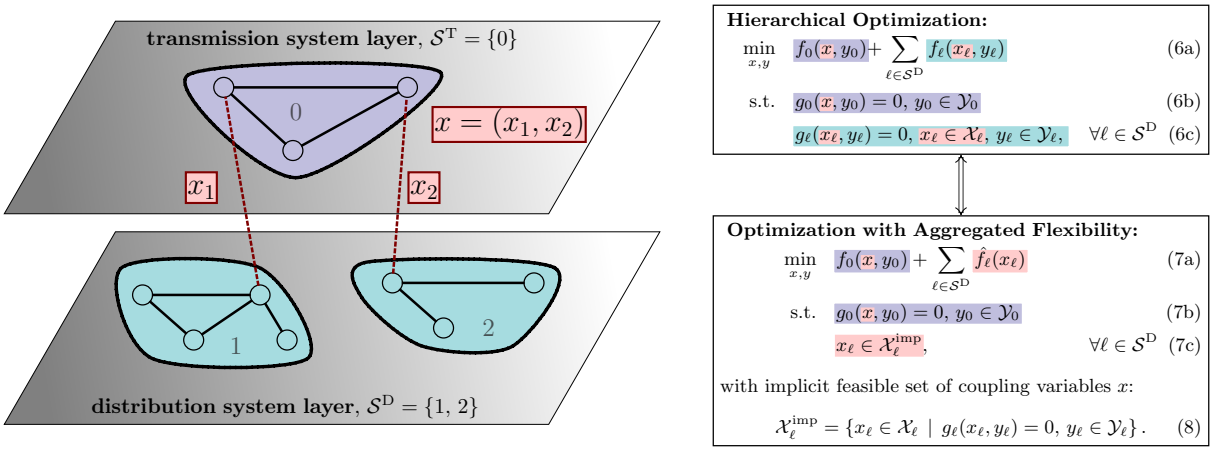

Figure 1: Hierarchical optimization framework for coordinating integrated transmission-distribution (ITD) systems. The transmission system (purple) and distribution systems (green) are modeled as layered subsystems, with coupling variables (red) linking the two layers. The right panel illustrates the corresponding mathematical reformulation: the original hierarchical optimization problem (6) is reduced to an aggregated optimization problem (7), where distribution-level feasibility is captured by the implicit feasible set (8).

together with the subsystem node sets $\{\mathcal{N}_\ell\}_{\ell \in \mathcal{S}}$, which satisfy

$$\mathcal{N} = \bigcup_{\ell \in \mathcal{S}} \mathcal{N}_\ell, \qquad \mathcal{N}_\ell \cap \mathcal{N}_m = \varnothing \quad \text{for all } \ell \neq m.$$

For each subsystem $\ell \in \mathcal{S}$, define

$$\mathcal{L}_\ell := \{(i,j) \in \mathcal{L} \,|\, i, j \in \mathcal{N}_\ell\}, \qquad \mathcal{G}_\ell = (\mathcal{N}_\ell, \mathcal{L}_\ell),$$

where $\mathcal{N}_\ell$ and $\mathcal{L}_\ell$ contain the buses and branches internal to subsystem ℓ. The coupling-edge set is

$$\mathcal{L}_s := \{(i,j) \in \mathcal{L} \,|\, \exists\, \ell, m \in \mathcal{S},\ \ell \neq m,\ i \in \mathcal{N}_\ell,\ j \in \mathcal{N}_m\}.$$

Hence,

$$\mathcal{L} = \left(\bigcup_{\ell \in \mathcal{S}} \mathcal{L}_\ell \right) \cup \mathcal{L}_s.$$

The edges in $\mathcal{L}_s$ represent coupling between subsystems. This partition is used to distinguish subsystem-internal variables and constraints from the PCC/interface quantities through which the subsystems are coordinated.

Remark 1 (Extension to multiple TSOs)**.** *The assumption $|\mathcal{S}^{\mathrm{T}}| = 1$ is adopted for notational simplicity and is not essential to the proposed aggregation method. In a more general setting, multiple interconnected transmission systems can be represented by $|\mathcal{S}^{\mathrm{T}}| \geq 1$, with additional tie-lines between transmission subsystems. The predictor-corrector aggregation remains local to each distribution subsystem and can be applied independently. The main additional challenge is the upper-layer coordination among TSOs after aggregation, which results in a distributed nonconvex nonlinear programming (*NLP*). Such transmission-level coordination can be handled by distributed optimization methods, e.g., a distributed optimization framework for large-scale AC OPF based on network decomposition and condensed coordination [7].*

Based on this partition, the AC power flow model of each subsystem $\ell \in \mathcal{S}$ contains the variables and constraints associated with its internal buses and branches $(\mathcal{N}_\ell, \mathcal{L}_\ell)$, while the quantities associated with the interconnections in $\mathcal{L}_s$ define the interface variables used in the coordination model below.

2.2 AC Power Flow Models

Let $Y_\ell \in \mathbb{C}^{|\mathcal{N}_\ell| \times |\mathcal{N}_\ell|}$ denote the complex bus admittance matrix of subgraph $\mathcal{G}_\ell$, with real and imaginary parts $G_\ell, B_\ell \in \mathbb{R}^{|\mathcal{N}_\ell| \times |\mathcal{N}_\ell|}$. For each bus $i \in \mathcal{N}_\ell$, the complex voltage $V_i \in \mathbb{C}$ is expressed in rectangular coordinates as $V_i = u_i + \mathbf{j} w_i$, with $u_i, w_i \in \mathbb{R}$. The active and reactive power generated at bus i are denoted by $p_i^g, q_i^g \in \mathbb{R}$, and the corresponding demand by $p_i^d, q_i^d \in \mathbb{R}$. For each distribution subsystem $\ell \in \mathcal{S}^{\mathrm{D}}$, let $p_\ell^{\mathrm{pcc}}, q_\ell^{\mathrm{pcc}} \in \mathbb{R}$ denote the active and reactive power exchanged with the transmission system through the PCC.

For each subsystem $\ell \in \mathcal{S}$, the nodal power injection from their neighboring buses can be expressed as

$$p_\ell^{\text{inj}}(u_\ell, w_\ell) = (G_\ell\, w_\ell + B_\ell\, u_\ell) \circ w_\ell + (G_\ell\, u_\ell - B_\ell\, w_\ell) \circ u_\ell \tag{1a}$$
$$q_\ell^{\text{inj}}(u_\ell, w_\ell) = (G_\ell\, u_\ell - B_\ell\, w_\ell) \circ w_\ell - (G_\ell\, w_\ell + B_\ell\, u_\ell) \circ u_\ell \tag{1b}$$

where nonlinear mappings p_ℓ^{inj}, $q_\ell^{\text{inj}} : \mathbb{R}^{|\mathcal{N}_\ell|} \times \mathbb{R}^{|\mathcal{N}_\ell|} \to \mathbb{R}^{|\mathcal{N}_\ell|}$ collect the nodal injections across all buses in $\mathcal{N}_\ell$, and u_ℓ, w_ℓ are the real and imaginary voltage components, i.e., $u_\ell = \{u_i\}_{i\in\mathcal{N}_\ell}$ and $w_\ell = \{w_i\}_{i\in\mathcal{N}_\ell}$.

For a distribution system $\ell \in \mathcal{S}^{\text{D}}$, the AC power flow model is given by

$$1 = e_1^\top u_\ell, \tag{2a}$$
$$0 = e_1^\top w_\ell, \tag{2b}$$
$$0 = p_\ell^{\text{inj}}(u_\ell, w_\ell) - C_\ell^g\, p_\ell^g + p_\ell^d - e_1\, p_\ell^{\text{pcc}}, \tag{2c}$$
$$0 = q_\ell^{\text{inj}}(u_\ell, w_\ell) - C_\ell^g\, q_\ell^g + q_\ell^d - e_1\, q_\ell^{\text{pcc}}, \tag{2d}$$
$$v_\ell = u_\ell^{\circ 2} + w_\ell^{\circ 2}, \tag{2e}$$
$$\underline{v}_\ell \le v_\ell \le \overline{v}_\ell, \quad \underline{p}_\ell^g \le p_\ell^g \le \overline{p}_\ell^g, \quad \underline{q}_\ell^g \le q_\ell^g \le \overline{q}_\ell^g \tag{2f}$$

with $p_\ell^d = \{p_i^d\}_{i\in\mathcal{N}_\ell}$ and $q_\ell^d = \{q_i^d\}_{i\in\mathcal{N}_\ell}$. Constraints (2a)-(2b) fix the voltage at the PCC, where $e_1^\top = [1, 0, 0, \ldots]$ [27, 47]. In the ITD model, the PCC represents the secondary side of the interface transformer. We assume that its voltage is regulated at a prescribed operating condition and, following the convention adopted in [34, 48], characterize the distribution-network flexibility in the P-Q domain under a fixed PCC voltage.

Constraints (2c)-(2d) denote the nodal power balance with the active and reactive power generation p_ℓ^g, $q_\ell^g \in \mathbb{R}^{n_\ell^g}$ for all n_ℓ^g controllable DERs, as well as the power injection from transmission system $p_\ell^{\text{pcc}}, q_\ell^{\text{pcc}} \in \mathbb{R}$. Here $C_\ell^g \in \mathbb{R}^{|\mathcal{N}_\ell| \times n_\ell^g}$ denote the connectivity matrix to these controllable DERs. Constraints (2e)-(2f) limit voltage magnitude, active power generation, and reactive generation, where overline and underline represent the corresponding upper and lower bounds respectively.

Analogously, the transmission system $\ell = 0$ is modeled as

$$1 = e_1^\top u_0, \tag{3a}$$
$$0 = e_1^\top w_0, \tag{3b}$$
$$0 = p_0^{\text{inj}}(u_0, w_0) - C_0^g\, p_0^g + p_0^d + C_0^{\text{pcc}} p^{\text{pcc}}, \tag{3c}$$
$$0 = q_0^{\text{inj}}(u_0, w_0) - C_0^g\, q_0^g + q_0^d + C_0^{\text{pcc}} q^{\text{pcc}}, \tag{3d}$$
$$v_0 = u_0^{\circ 2} + w_0^{\circ 2}, \tag{3e}$$
$$\underline{v}_0 \le v_0 \le \overline{v}_0, \quad \underline{p}_0^g \le p_0^g \le \overline{p}_0^g, \quad \underline{q}_0^g \le q_0^g \le \overline{q}_0^g. \tag{3f}$$

The voltage at the slack bus is fixed, following the convention in the AC power flow problem. The key distinction between (2) and (3) lies in the active and reactive power exchanges via PCC between them. While $p_\ell^{\text{pcc}}, q_\ell^{\text{pcc}}$ denote the power injection from the transmission system at the substation (the first bus in the subgraph $\mathcal{G}_\ell$), $p^{\text{pcc}}, q^{\text{pcc}} \in \mathbb{R}^{|\mathcal{S}^{\text{D}}|}$ stack the power transferred to distribution systems, i.e., $p^{\text{pcc}} = \{p_\ell^{\text{pcc}}\}_{\ell\in\mathcal{S}^{\text{D}}}$ and $q^{\text{pcc}} = \{q_\ell^{\text{pcc}}\}_{\ell\in\mathcal{S}^{\text{D}}}$. Moreover, the connectivity matrix $C_0^{\text{pcc}} \in \mathbb{R}^{|\mathcal{N}_0| \times |\mathcal{S}^{\text{D}}|}$ in the constraints (3c)-(3d) specifies the transmission buses to which distribution systems are attached.

These detailed subsystem models in (2) and (3) constitute the foundation of the hierarchical optimization framework (6). In particular, the transmission model (3) corresponds to the upper-layer constraints (6b), while the distribution models (2) represent the lower-layer constraints (6c).

2.3 TSO-DSO Management Models

How these subsystem models are embedded into the ITD coordination problem depends on the adopted TSO-DSO management model, which determines the structure of the coupling variables between transmission and distribution.

The choice of whether the TSO can directly dispatch DERs shapes the structure of the coupling variables linking transmission and distribution. In particular, the choice of operational paradigm, such as whether the TSO is allowed to issue dispatch commands to each distribution-level DER, determines the structure of the coupling variables x_ℓ that link the transmission system with each distribution subsystem ℓ.

These management models determine the definition of the coupling variables x_ℓ and local variables y_ℓ, and thereby specify how each subsystem enters the hierarchical optimization problem introduced next.

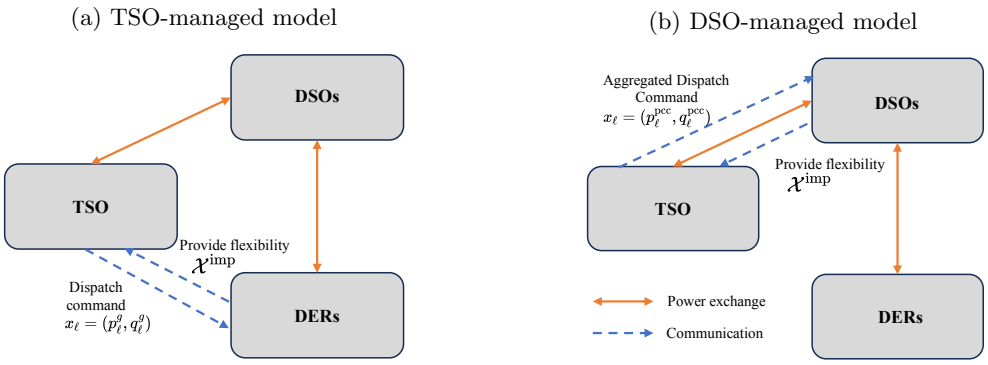

Figure 2: Both TSO–DSO management models keep detailed distribution-level data hidden from the TSO but differ mainly in communication and coordination. In the TSO-managed model, each distribution system ℓ provides $\mathcal{X}^{\text{imp}}$ with respect to DER outputs (p_ℓ^g, q_ℓ^g), enabling the TSO to dispatch DERs directly. In the DSO-managed model, each DSO instead supplies $\mathcal{X}^{\text{imp}}$ with respect to PCC exchanges $(p_\ell^{\text{pcc}}, q_\ell^{\text{pcc}})$ and disaggregates individual DERs based on the scheduled PCC exchanges.

2.3.1 TSO-managed model

In a TSO-managed model (Fig. 2a), the transmission operator directly controls the DERs within each distribution subsystem. The coupling variables are given by

$$x_\ell = (p_\ell^g, q_\ell^g) \in \mathcal{X}_\ell \subseteq \mathbb{R}^{2n_\ell^g},$$

i.e., the active and reactive power outputs of all DERs. The remaining state variables

$$y_\ell = (u_\ell, w_\ell, v_\ell, p_\ell^{\text{pcc}}, q_\ell^{\text{pcc}}) \in \mathcal{Y}_\ell \subseteq \mathbb{R}^{3|\mathcal{N}_\ell|+2}$$

are determined internally through the distribution model (2), which can be expressed as

$$g_\ell(x_\ell, y_\ell) = 0,\ x_\ell \in \mathcal{X}_\ell,\ y_\ell \in \mathcal{Y}_\ell, \tag{4}$$

where nonlinear function $g_\ell : \mathbb{R}^{2n_\ell^g} \times \mathbb{R}^{3|\mathcal{N}_\ell|+2} \to \mathbb{R}^{3|\mathcal{N}_\ell|+2}$ summarizes the equality constraints (2a)-(2e). In this management model, the PCC exchanges $(p_\ell^{\text{pcc}}, q_\ell^{\text{pcc}})$ are dependent variables: once x_ℓ is fixed, the power exchanged with the transmission system is uniquely determined. The proposed aggregation method thus directly yields an approximate implicit feasible set (14).

2.3.2 DSO-managed model

By contrast, in a DSO-managed model (Fig. 2b), the TSO interacts with each distribution subsystem only through the PCC, without direct control of internal DERs. The coupling variables reduce to the aggregated exchanges,

$$x_\ell = (p_\ell^{\text{pcc}}, q_\ell^{\text{pcc}}) \in \mathcal{X} \subseteq \mathbb{R}^2.$$

A difficulty arises when multiple DERs are present, since the system then has more internal degrees of freedom.

To address this, we follow the strategy of adjusting system imbalance as presented in [49]. In the DSO-managed model, we introduce the *distributed slack* where each DER adjusts its power output to account for the system-wide power flexibility. For DER i in subsystem ℓ,

$$p_{\ell,i}^g = p_{\ell,i,\text{ref}}^g + \alpha_i^p \Delta p_\ell^g, \tag{5a}$$
$$q_{\ell,i}^g = q_{\ell,i,\text{ref}}^g + \alpha_i^q \Delta q_\ell^g \tag{5b}$$

with $\sum_i \alpha_i^p = 1$ and $\sum_i \alpha_i^q = 1$. Here, $\Delta p_\ell^g, \Delta q_\ell^g$ represent system-wide adjustments, and $p_{\ell,i,\text{ref}}^g$, $q_{\ell,i,\text{ref}}^g$ are the nominal setpoint for the DER output. The participation factors, determined by the DSO, e.g., for minimizing system losses, remain constant during optimization, while the system-wide power adjustment can be implicitly

governed by the coupling variables x_ℓ. As a result, the distribution model (2) can also be expressed as (4) with $g_\ell : \mathbb{R}^2 \times \mathbb{R}^{3|\mathcal{N}_\ell|+2} \to \mathbb{R}^{3|\mathcal{N}_\ell|+2}$ and

$$y_\ell = (u_\ell, w_\ell, v_\ell, \Delta p_\ell^g, \Delta q_\ell^g) \in \mathcal{Y} \subseteq \mathbb{R}^{3|\mathcal{N}_\ell|+2}.$$

An alternative to participation factors is using the pseudo-inverse of $\frac{\partial g_\ell}{\partial y_\ell}$ to resolve invertibility challenges.

Remark 2 (DER–PCC mapping)**.** *The fixed participation policy removes internal dispatch degrees of freedom by parameterizing individual DER outputs through aggregate adjustments. Under regularity assumptions, this parameterization yields a locally single-valued mapping from PCC exchanges to the corresponding policy-restricted internal state.*

The two-dimensional P–Q flexibility set considered in this paper is conditional on a prescribed PCC voltage magnitude. When the dependence of distribution-level flexibility on the interface voltage must be represented explicitly, the coupling variables may be extended to include the PCC active-power exchange, reactive-power exchange, and voltage magnitude:

$$x_\ell = (p_\ell^{\text{pcc}}, q_\ell^{\text{pcc}}, v_\ell^{\text{pcc}}),$$

where yields a *P-Q-V* flexibility aggregation representation and allows the transmission-level problem to account explicitly for the dependence of distribution-network feasibility on the PCC voltage magnitude. The predictor-corrector aggregation remains applicable to the resulting smooth implicit formulation under the regularity conditions. The PCC voltage magnitude may be treated as a continuous coupling variable without explicitly modeling transformer tap positions. If tap positions are optimized directly, additional discrete variables are required. One possible alternative is to construct a separate continuous flexibility representation for each admissible tap position.

With the subsystem models and coupling variable definitions now specified, we can formulate the overall ITD coordination problem as a graph-based hierarchical optimization problem.

2.4 Graph–Based Hierarchical Optimization

The coordination of ITD systems can be expressed as a hierarchical optimization problem:

$$\min_{x,y} \quad f_0(x, y_0) + \sum_{\ell \in \mathcal{S}^{\text{D}}} f_\ell(x_\ell, y_\ell) \tag{6a}$$

$$\text{s.t.} \quad g_0(x, y_0) = 0,\ y_0 \in \mathcal{Y}_0, \tag{6b}$$

$$g_\ell(x_\ell, y_\ell) = 0,\ x_\ell \in \mathcal{X}_\ell,\ y_\ell \in \mathcal{Y}_\ell, \qquad \forall \ell \in \mathcal{S}^{\text{D}}, \tag{6c}$$

with $x = \{x_\ell\}_{\ell \in \mathcal{S}^{\text{D}}}$, where f_ℓ and g_ℓ denote the objective functions and constraints of the subsystem $\ell \in \mathcal{S} = \mathcal{S}^{\text{T}} \bigcup \mathcal{S}^{\text{D}}$. The vector y_ℓ collects the variables that are totally local to the subsystem $\ell \in \mathcal{S}$, while the vector x_ℓ represents the coupling variables that link the transmission system 0 with distribution subsystems $\ell \in \mathcal{S}^{\text{D}}$, with feasible sets $\mathcal{X}_\ell$ and $\mathcal{Y}_\ell$, respectively. Variable partitioning depends on the adopted control strategy discussed in Section 2.3.

As illustrated in Fig. 1, constraints (6b) and (6c) describe the transmission and distribution system models, respectively. The essence of flexibility aggregation is to replace detailed distribution models (6c) and also the local variables y_ℓ with implicit feasible sets in the coupling space. Then, the optimization problem (6) can be rewritten as

$$\min_{x,y_0} \quad f_0(x, y_0) + \sum_{\ell \in \mathcal{S}^{\text{D}}} \hat{f}_\ell(x_\ell) \tag{7a}$$

$$\text{s.t.} \quad g_0(x, y_0) = 0,\ y_0 \in \mathcal{Y}_0, \tag{7b}$$

$$x_\ell \in \mathcal{X}_\ell^{\text{imp}}, \qquad \forall \ell \in \mathcal{S}^{\text{D}} \tag{7c}$$

with implicit feasible sets for all $\ell \in \mathcal{S}^{\text{D}}$

$$\mathcal{X}_\ell^{\text{imp}} = \{x_\ell \in \mathcal{X}_\ell \mid \exists\, y_\ell \in \mathcal{Y}_\ell \text{ such that } g_\ell(x_\ell, y_\ell) = 0\} \tag{8}$$

where $\hat{f}_\ell$ approximates the local objective f_ℓ.

The nonlinear implicit feasible sets $\mathcal{X}_\ell^{\text{imp}}$ in (8) are generally nonconvex and computationally intractable. The predictor-corrector approximation introduced in the next section is not a convexification of these sets and does not provide a global optimality certificate for Problem (6). Instead, it constructs a tractable local surrogate of the implicit distribution-level feasible set around a regular operating point. The theoretical guarantees derived below are local approximation error bounds for the reconstructed state (Theorem 2) and for the induced feasible-set approximation (Corollary 1).

3 Nonlinear Flexibility Aggregation

This section develops the proposed nonlinear flexibility aggregation method for approximating the implicit feasible sets introduced in Section 2.4. We first present the predictor-corrector approximation for flexibility aggregation, then derive local error bounds, and finally illustrate its performance through numerical examples and benchmark studies.

3.1 Predictor–Corrector Approximation

We propose a predictor–corrector aggregation method to construct tractable surrogates of the implicit sets, thereby enabling coordination without requiring explicit distribution-grid models in the upper-level coordination problem. Starting from a known feasible point, the proposed method first computes a tangential predictor based on local Jacobian information, and then applies a corrector step to compensate for the residual of the nonlinear equality constraints. The resulting approximation $\hat{y}(x)$ provides a local surrogate of the exact implicit mapping while avoiding iterative nonlinear solves.

In the following, we illustrate how the proposed method approximates the implicit feasible set (8). For notational simplicity, the subsystem subscript index ℓ is omitted in the following subsections.

Consider the nonlinear system associated with a single distribution grid:

$$g(x, y) = 0,\ x \in \mathcal{X} \subseteq \mathbb{R}^m,\ y \in \mathcal{Y} \subseteq \mathbb{R}^n, \tag{9}$$

where $g : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^l$ represents the system equality constraints in (6c).

The exact projected feasible set of the coupling variable x is given by

$$\mathcal{X}^{\text{imp}} = \{x \in \mathcal{X} \mid \exists\, y \in \mathcal{Y} \text{ such that } g(x, y) = 0\}.$$

Hence, for every $x \in \mathcal{X}^{\text{imp}}$, there exists at least one $y \in \mathcal{Y}$ satisfying $g(x, y) = 0$. Under the regularity conditions stated below, such a feasible y can be represented locally by an implicit mapping $y^\star(\cdot)$ around a feasible point.

Since computing the exact implicit set $\mathcal{X}^{\text{imp}}$ is intractable, we propose a predictor-corrector method to approximate it. We first impose the regularity assumptions in the following:

Assumption 1. *The function g is continuously differentiable on $\Pi := \mathcal{X} \times \mathcal{Y}$.*

Assumption 2. *Let $(x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$ be a feasible base point satisfying $g(x_0, y_0) = 0$. The dimensions satisfy $n = l$, and the base-point Jacobian $\partial_y g(x_0, y_0) \in \mathbb{R}^{n\times n}$ is nonsingular.*

Assumption 2 is the standard regularity condition ensuring local existence and local uniqueness of the implicit mapping around a feasible operating point. Its numerical verification and practical implementation are provided in Remark 4. Based on the assumpption, we introduce the implicit function theorem in the following:

Theorem 1 (Implicit Function Theorem)**.** *Let $(x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$ satisfy*

$$g(x_0, y_0) = 0.$$

Suppose Assumptions 1 and 2 hold in a neighborhood of (x_0, y_0). Then there exist neighborhoods $\mathcal{U} \subseteq \mathcal{X}$ of x_0 and $\mathcal{V} \subseteq \mathcal{Y}$ of y_0, and a locally unique continuously differentiable mapping $y^\star : \mathcal{U} \to \mathcal{V}$ such that

$$g(x, y^\star(x)) = 0, \qquad \forall x \in \mathcal{U}. \tag{10}$$

Moreover,

$$\partial_x y^\star(x) = -\left[\partial_y g(x, y^\star(x))\right]^{-1} \partial_x g(x, y^\star(x)),\ \forall x \in \mathcal{U} \tag{11}$$

Note that (11) follows from differentiating (10) with respect to x, which yields

$$\partial_y g(x, y^\star(x))\, \partial_x y^\star(x) + \partial_x g(x, y^\star(x)) = 0.$$

For a given x in the neighborhoods of x_0, the tangential predictor of y at the base point (x_0, y_0) is

$$\bar{y}(x) := y_0 - \left[\partial_y g(x_0, y_0)\right]^{-1} \partial_x g(x_0, y_0)\,(x - x_0). \tag{12}$$

Since $\bar{y}(x)$ is only a first-order approximation, it generally does not satisfy the nonlinear constraints $g(x, \bar{y}(x)) = 0$. We therefore introduce the corrector

$$\Delta y^{\text{cor}}(x) := -\left[\partial_y g(x_0, y_0)\right]^{-1} g(x, \bar{y}(x))$$

which yields the predictor-corrector approximation:

$$\begin{aligned}\hat{y}(x) &:= \bar{y}(x) + \Delta y^{\mathrm{cor}}(x) \\ &= y_0 - \left[\partial_y g(x_0, y_0)\right]^{-1} \partial_x g(x_0, y_0)\,(x - x_0) \\ &\quad - \left[\partial_y g(x_0, y_0)\right]^{-1} g(x, \bar{y}(x))\end{aligned} \tag{13}$$

The role of (13) is to approximate the equality-defined implicit relation $g(x, y) = 0$. The original set constraint $y \in \mathcal{Y}$ is not enforced by the predictor-corrector step itself. Instead, it is incorporated through the reduced feasible set in the x-space induced by the implicit mapping. Accordingly, the approximated implicit feasible region is defined as

$$\hat{\mathcal{X}}^{\mathrm{imp}} = \{x \in \mathcal{X} \mid \hat{y}(x) \in \mathcal{Y}\}. \tag{14}$$

Thus, the predictor-corrector construction approximates the equality-constrained manifold, while the inequality and bound constraints on y are enforced through the membership condition $\hat{y}(x) \in \mathcal{Y}$. If the reconstructed state $\hat{y}(x)$ satisfies the operational limits, i.e., $\hat{y}(x) \in \mathcal{Y}$, the corresponding coupling point x is retained in $\hat{\mathcal{X}}^{\mathrm{imp}}$. Otherwise, if $\hat{y}(x) \notin \mathcal{Y}$, the coupling point is excluded from the predictor-corrector approximation of the implicit feasible region.

Remark 3. *The proposed predictor-corrector aggregation provides a non-iterative local surrogate of the nonlinear implicit feasible set $\mathcal{X}^{\mathrm{imp}}$. Compared with the existing aggregation methods summarized in Table 1, it directly accounts for nonlinear AC power flow equations, is applicable to both radial and meshed distribution networks under the stated regularity assumptions, and enables scalable precomputation through spatio-temporal decomposition.*

3.2 Approximation Error Analysis

We next analyze the local approximation error of the proposed predictor-corrector construction under the additional assumption of twice continuous differentiability.

Assumption 3. *The function g is twice continuously differentiable on $\Pi := \mathcal{X} \times \mathcal{Y}$.*

For convenience, we simplify notation by writing

$$y^\star := y^\star(x), \qquad \bar{y} := \bar{y}(x), \qquad \hat{y} := \hat{y}(x),$$

and define the sensitivity operator along the feasible manifold as

$$S^\star(x) := \partial_x y^\star(x) = -\left[\partial_y g(x, y^\star(x))\right]^{-1} \partial_x g(x, y^\star(x)).$$

The following regularity result summarizes the local smoothness properties required in the subsequent analysis.

Lemma 1. *Under the assumptions of Theorem 1, and additionally Assumption 3, there exist constants $L_1, L_2, L_3 > 0$ such that, for all $x, x_1, x_2 \in \mathcal{U}$, and all $y_1, y_2 \in \mathcal{V}$,*

$$\|S^\star(x_1) - S^\star(x_2)\| \le L_1 \|x_1 - x_2\|, \tag{15}$$

$$\|\partial_y g(x_1, y^\star(x_1)) - \partial_y g(x_2, y^\star(x_2))\| \le L_2 \|x_1 - x_2\|, \tag{16}$$

$$\|\partial_y g(x, y_1) - \partial_y g(x, y_2)\| \le L_3 \|y_1 - y_2\|. \tag{17}$$

We first establish the local error bound for the tangential predictor (12).

Lemma 2. *Under the assumptions of Lemma 1, for all $x \in \mathcal{U}$, the tangential predictor $\bar{y}(x)$ defined in (12) satisfies*

$$\bar{y}(x) - y^\star(x) = \mathcal{O}(\|x - x_0\|^2).$$

Proof. For a fixed $x \in \mathcal{U}$, we simplify notation by writing

$$y^\star := y^\star(x), \qquad \bar{y} := \bar{y}(x).$$

From (12) we have

$$\bar{y} - y^\star = y_0 - y^\star + S^\star(x_0)(x - x_0).$$

Since $y_0 = y^\star(x_0)$, we apply the integral form of the mean value theorem to $y^\star(\cdot)$ and obtain

$$y_0 - y^\star = -\int_0^1 S^\star(x_0 + \tau(x - x_0))(x - x_0) d\tau.$$

Hence,

$$\bar{y} - y^\star = -\int_0^1 \left[S^\star(x_0 + \tau(x - x_0)) - S^\star(x_0)\right](x - x_0)d\tau$$

Taking norms yields

$$\begin{aligned}
&\|\bar{y} - y^\star\| \\
\leq &\|x - x_0\| \int_0^1 \|S^\star(x_0 + \tau(x - x_0) - S^\star(x_0)\| \, d\tau \\
&\Downarrow \text{Lipschitz continuity (15)} \\
\leq &\|x - x_0\|^2 \int_0^1 L_1 \tau d\tau \\
\leq &\frac{1}{2} L_1 \|x - x_0\|^2 , \qquad (18)
\end{aligned}$$

which concludes the proof. □

To state the error bound with an explicit constant for the predictor-corrector approximation (13), let $M_0 := \partial_y g(x_0, y_0)$ denote the base-point Jacobian, which is nonsingular by Assumption 2, set $\omega := \left\|M_0^{-1}\right\|$, and fix $\bar{r} > 0$ such that

$$\mathcal{B}(x_0, \bar{r}) := \{x \in \mathcal{U} \subseteq \mathcal{X} \mid \|x - x_0\| \leq \bar{r}\}.$$

Theorem 2. *Suppose the assumptions of Lemma 1 hold. Then, for all $x \in \mathcal{B}(x_0, \bar{r})$, the predictor-corrector approximation $\hat{y}(x)$ defined in (13) satisfies*

$$\|\hat{y}(x) - y^\star(x)\| \;\leq\; C \, \|x - x_0\|^3 ,$$

with

$$C := \tfrac{\omega}{2} \left(L_1 L_2 + \tfrac{1}{4} L_1^2 L_3 \, \bar{r}\right).$$

In particular, $\hat{y}(x) - y^\star(x) = \mathcal{O}(\|x - x_0\|^3)$.

Proof. From (12) and (13), we obtain

$$\begin{aligned}
\hat{y} - \bar{y} &= -M_0^{-1} g(x, \bar{y}) \\
\hat{y} - y^\star &= \bar{y} - y^\star - M_0^{-1} \left(g(x, \bar{y}) - g(x, y^\star)\right)
\end{aligned}$$

Applying the integral form of the mean value theorem to $g(x, \cdot)$, we obtain

$$\begin{aligned}
&g(x, \bar{y}) - g(x, y^\star) \\
&= \int_0^1 \partial_y g\big(x, y^\star + \tau(\bar{y} - y^\star)\big) \, d\tau \, (\bar{y} - y^\star).
\end{aligned}$$

Hence,

$$\hat{y} - y^\star = M_0^{-1} \Big\{ M_0 - \int_0^1 \partial_y g(x, y^\star + \tau \, (\bar{y} - y^\star)) d\tau \Big\} \, (\bar{y} - y^\star)$$

Taking the norm of both sides, we have

$$\begin{aligned}
&\|\hat{y} - y^\star\| \\
\leq &\left\|M_0^{-1}\right\| \left\{ \|M_0 - \partial_y g(x, y^\star)\| + \int_0^1 \|\partial_y g(x, y^\star + \tau \, (\bar{y} - y^\star)) - \partial_y g(x, y^\star)\| \, d\tau \right\} \|\bar{y} - y^\star\| \\
&\Downarrow \text{Lipschitz continuous (17)} \\
\leq &\left\|M_0^{-1}\right\| \left\{ \|M_0 - \partial_y g(x, y^\star)\| + L_3 \, \|\bar{y} - y^\star\| \int_0^1 \tau d\tau \right\} \|\bar{y} - y^\star\| \\
&\Downarrow \text{Lipschitz continuous (16)}
\end{aligned}$$

$$
\begin{aligned}
&\leq \left\|M_0^{-1}\right\| \left\{ L_2 \left\|x - x_0\right\| + \frac{L_3}{2} \left\|\bar{y} - y^\star\right\| \right\} \left\|\bar{y} - y^\star\right\| \\
&\Downarrow \text{Remainder of } \bar{y} \text{ (18)} \\
&\leq \frac{1}{2}\omega \left\{ L_2 L_1 \left\|x - x_0\right\|^3 + \frac{1}{4} L_3 L_1^2 \left\|x - x_0\right\|^4 \right\}, \\
&\leq \frac{\omega}{2} \left(L_1 L_2 + \tfrac{1}{4} L_1^2 L_3 \, \bar{r} \right) \left\|x - x_0\right\|^3 = C \left\|x - x_0\right\|^3
\end{aligned} \tag{19}
$$

which concludes the proof. □

Theorem 2 also yields a guarantee for the induced feasible set approximation (14).

Corollary 1 (Feasible-set approximation error)**.** *Suppose the assumptions of Theorem 2 hold. Then, for all $x \in \mathcal{B}(x_0, \bar{r})$, the following conditons are satisfied:*

$$
\begin{aligned}
\operatorname{dist}\left(y^\star(x),\, \mathcal{Y}\right) &\leq C \left\|x - x_0\right\|^3, \forall x \in \hat{\mathcal{X}}^{\text{imp}} \setminus \mathcal{X}^{\text{imp}}, \\
\operatorname{dist}\left(y^\star(x),\, \mathbb{R}^n \setminus \mathcal{Y}\right) &\leq C \left\|x - x_0\right\|^3, \forall x \in \mathcal{X}^{\text{imp}} \setminus \hat{\mathcal{X}}^{\text{imp}}.
\end{aligned}
$$

Proof. By Theorem 2, $\left\|\hat{y}(x) - y^\star(x)\right\| \leq C \left\|x - x_0\right\|^3$ on $\mathcal{B}(x_0, \bar{r})$. In the first case, $\hat{y}(x) \in \mathcal{Y}$ while $y^\star(x) \notin \mathcal{Y}$, hence $\operatorname{dist}(y^\star(x), \mathcal{Y}) \leq \left\|y^\star(x) - \hat{y}(x)\right\| \leq C \left\|x - x_0\right\|^3$. In the second case, $\hat{y}(x) \in \mathbb{R}^n \setminus \mathcal{Y}$ while $y^\star(x) \in \mathcal{Y}$, hence $\operatorname{dist}(y^\star(x), \mathbb{R}^n \setminus \mathcal{Y}) \leq \left\|y^\star(x) - \hat{y}(x)\right\| \leq C \left\|x - x_0\right\|^3$. □

Hence misclassification can occur only where the exact state lies within $C \left\|x - x_0\right\|^3$ of the boundary of the operational limits. Coupling points whose exact state is bounded away from this boundary are correctly classified once $\left\|x - x_0\right\|$ is sufficiently small.

The next proposition provides an interpretation of the predictor-corrector approximation in the special case where g is quadratic.

Proposition 1. *Suppose the assumptions of Theorem 1 hold, and additionally assume that $g : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$ is quadratic. Then the predictor-corrector approximation $\hat{y}(x)$ defined in (13) coincides with the second-order Taylor approximation of the implicit mapping $y^\star(x)$ around x_0.*

Proof. For a quadratic function g, the Taylor expansion of $g(x, y)$ about the feasible point (x_0, y_0) is exact up to second order. By construction, the tangential predictor $\bar{y}$ cancels the constant and first-order terms in this expansion. Hence, the residual $g(x, \bar{y})$ contains only second-order terms in $\Delta x := x - x_0$.

The corrector step in (13) is obtained by premultiplying this residual by $-[\partial_y g(x_0, y_0)]^{-1}$, thereby producing exactly the second-order correction to the tangential predictor. Therefore, $\hat{y}(x)$ coincides with the second-order Taylor approximation of the implicit mapping $y^\star(x)$ around x_0. □

Remark 4 (Regularity of the base point)**.** *Assumption 2 is a local regularity condition imposed only at the selected feasible base point (x_0, y_0). In implementation, the corrector is computed by solving linear systems with $M_0 = \partial_y g(x_0, y_0)$, whose numerical regularity can be assessed using its smallest singular value, condition number, or factorization pivots. Solvability [50, Thm. 2] or voltage stability indicators [51, Thm. 2] may also be used as auxiliary screening tools. If M_0 is singular or severely ill-conditioned, the base point can be rejected and the surrogate reconstructed around a better-conditioned feasible operating point or over a smaller neighborhood. A pseudo-inverse correction may also be used as a numerical fallback, although such a regularized step lies outside the theoretical predictor-corrector construction analyzed here.*

3.3 Illustrative Example: IEEE 33-Bus System

We first illustrate the effectiveness of the proposed method using the IEEE 33-bus distribution network [52]. The system is evaluated under both TSO- and DSO-managed coordination models, where the predictor–corrector aggregation method is compared against two widely used linear surrogates.

Conventional studies typically rely on linear models, such as the LinDistFlow approximation or first-order tangential predictors of the AC equations. As shown in Figs. 3a and 3b, these models capture the general orientation of the reference feasible region $\mathcal{X}^{\text{imp}}$ but do not accurately reproduce its shape. Figs. 3c and 3d show that the convex relaxation of the DistFlow model [21, 53] provides an outer approximation that contains the exact AC feasible set but also introduces an additional false-feasible region.

In contrast, as shown in Fig. 4, the proposed predictor–corrector method substantially improves the approximation. Fig. 4a and Fig. 4b depict the feasible regions obtained from the proposed set $\hat{\mathcal{X}}^{\text{imp}}$, showing close alignment with the exact AC feasible region. The improvement becomes more evident in Fig. 5a and Fig. 5b, which display the error of the voltage magnitudes for the proposed approximation. In the flexibility sets, the voltage error is mostly

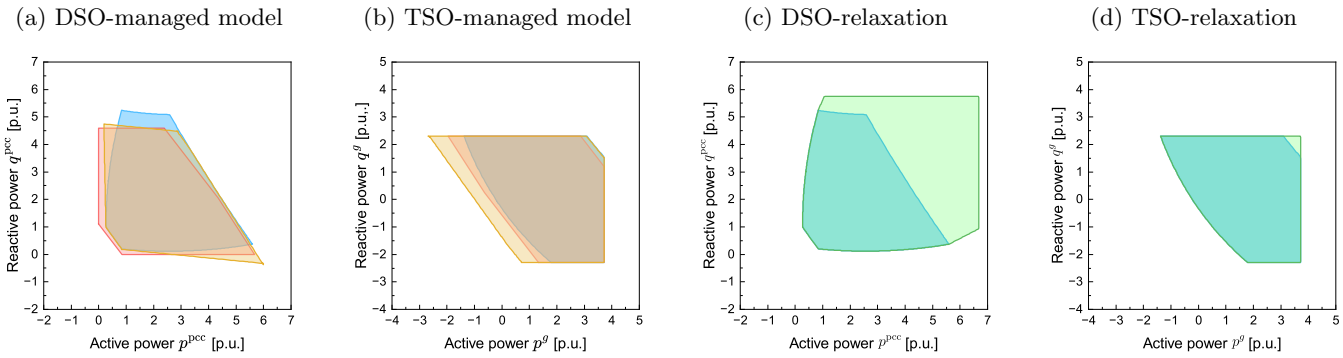

Figure 3: Comparisons of flexibility. The red and orange regions correspond to linear surrogate models, the green region to the convex relaxation, and the exact region is blue.

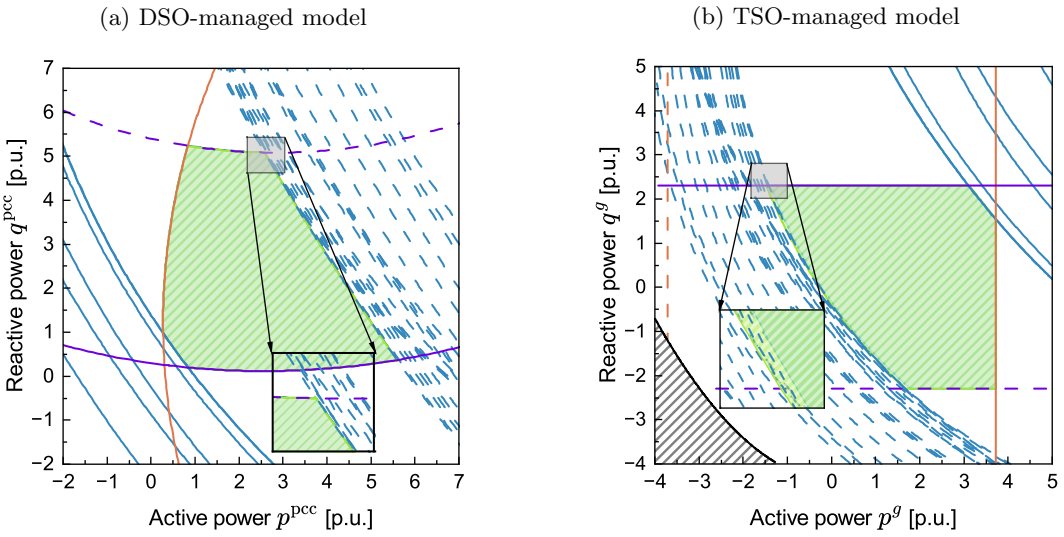

Figure 4: The proposed approximated set (14) vs. the exact feasible regions (8). The solid and dashed curves give upper and lower limits on squared voltage magnitudes (blue), active power (red), and reactive power (purple) from constraint (2f).

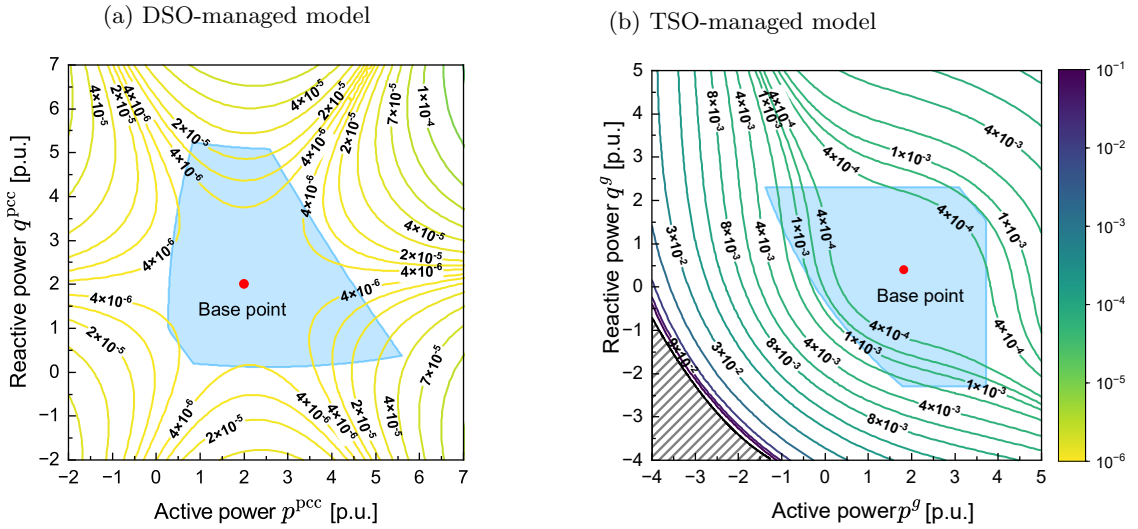

Figure 5: Error in voltage magnitude between the proposed approach (13) and exact AC power flow solutions.

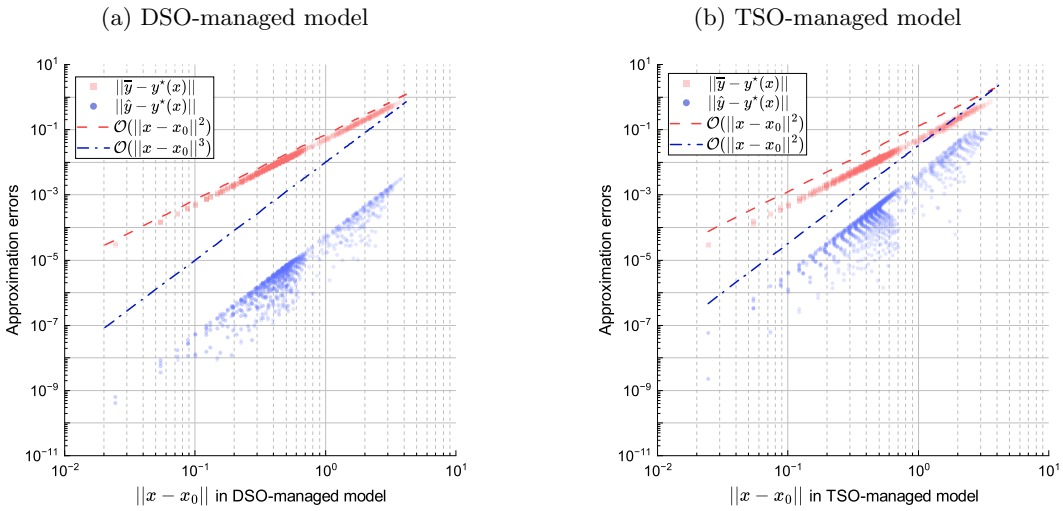

Figure 6: Error behavior of tangential predictor (red) and predictor–corrector (blue) approaches, with corresponding error upper bound (dotted lines).

below 4×10^{-6} for the DSO-managed model in Fig. 5a, and below 1×10^{-3} almost everywhere for the TSO-managed model in Fig. 5b. Moreover, Fig. 6 quantifies the approximation errors of both the tangential predictor and the predictor–corrector method, together with the theoretical error bounds derived in (18)–(19), with Lipschitz constants computed by [54, Proposition C.29].

Since the AC power flow equations are quadratic, the predictor–corrector aggregation method effectively provides a second-order approximation of the implicit feasible mapping (Proposition 1). This enables the method to capture boundary curvature and preserve feasibility with greater accuracy than purely linear surrogates, making it a powerful tool for representing distribution-level flexibility in hierarchical ITD coordination.

3.4 Benchmarking Approximation Accuracy

While the IEEE 33-bus system provides an illustrative comparison between methods, we construct a benchmark suite comprising 24 representative radial distribution system test cases with 7 meshed variants to systematically evaluate the accuracy and robustness of the proposed predictor-corrector method. These include all distribution system models from MATPOWER and a real-world distribution network, the KIT Campus Nord grid (`KIT Campus North`) [55], along with their meshed variants. ylAll optimization problems are implemented in MATLAB and solved with IPOPT [45] on a MacBook Pro with an M3 Pro chip.

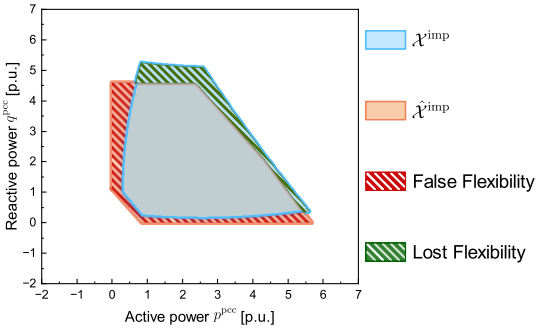

Figure 7: Illustration of the benchmark metrics.

Consider the nonlinear system associated with a single distribution grid (9). Accuracy is quantified by comparing the approximate feasible set (14),

$$\hat{\mathcal{X}}^{\text{imp}} = \{x \in \mathcal{X} \,|\, \hat{y}(x) \in \mathcal{Y}\},$$

with the exact implicit feasible set (8),

$$\mathcal{X}^{\text{imp}} = \{x \in \mathcal{X} \,|\, \exists\, y \in \mathcal{Y} \text{ such that } g(x, y) = 0\}.$$

As shown in Fig. 7, two complementary performance metrics are considered:

- **False Flexibility** [%]: the number of uniformly sampled points in $\hat{\mathcal{X}}^{\text{imp}} \setminus \mathcal{X}^{\text{imp}}$, normalized by the number of sampled points in $\mathcal{X}^{\text{imp}}$;
- **Lost Flexibility** [%]: the number of uniformly sampled points in $\mathcal{X}^{\text{imp}} \setminus \hat{\mathcal{X}}^{\text{imp}}$, normalized by the number of sampled points in $\mathcal{X}^{\text{imp}}$.

The same set of coupling points, sampled uniformly from a common bounding box, is used for all methods. For each sampled coupling point x, membership in $\mathcal{X}^{\text{imp}}$ is evaluated by fixing x, solving the full nonlinear AC feasibility problem for y, and checking whether $y \in \mathcal{Y}$. A sampled point is classified as feasible if the equality constraints (2a)–(2e) are satisfied up to the prescribed numerical tolerance and all operational constraints, including (2f), are satisfied. Note that the false flexibility may exceed 100% when the false-feasible region is larger than the exact feasible region.

By Corollary 1, both metrics count sampled points in the symmetric difference of $\hat{\mathcal{X}}^{\text{imp}}$ and $\mathcal{X}^{\text{imp}}$, which the theory predicts to concentrate in $C\,\|x - x_0\|^3$-thin band around the exact boundary of $\mathcal{Y}$. The benchmarks below quantify this misclassification empirically.

Table 2 compares five methods: the LinDistFlow approximation (LDS), the enhanced DC approximation (EDC), the convex relaxation of the DistFlow model (CR), the tangential predictor (TP), and the proposed predictor-corrector method (PC). The feasible region obtained using convex relaxation is an outer approximation of the exact implicit feasible set and therefore overestimates the available flexibility. This behavior is evident in Table 2, where CR has zero lost flexibility in all applicable radial cases but large false flexibility due to its false-feasible outer region. By contrast, the proposed method consistently yields the lowest false flexibility and lost flexibility values across the benchmark networks, indicating the closest agreement with the nonlinear AC feasible region.

Having demonstrated the accuracy of the predictor-corrector method in the static benchmarks, the next section extends the analysis to a multiperiod setting, where temporal coupling and time-varying demand further complicate flexibility aggregation and coordination.

4 Multiperiod Coordination Problems

Building on the spatial decomposition achieved by the predictor–corrector aggregation method, we now extend the framework to multiperiod settings. In this case, the storage DERs introduce intertemporal coupling, linking decisions across successive time steps. We demonstrate how implicit feasible sets can be constructed in parallel for each period while temporal consistency is enforced through coupling constraints, enabling efficient spatio-temporal decomposition.

4.1 Hierarchical Formulation

Consider a time horizon $\mathcal{K} = \{1, \ldots, N-1\}$. At each time step $k \in \mathcal{K}$, we collect the coupling variables and local variables into

$$x_k = \{x_{\ell|k}\}_{\ell \in \mathcal{S}^{\text{D}}} \text{ and } y_k = \{y_{\ell|k}\}_{\ell \in \mathcal{S}}.$$

Stacking them across the horizon yields

$$x = \{x_k\}_{k \in \mathcal{K}} \text{ and } y = \{y_k\}_{k \in \mathcal{K}}.$$

In the DSO-managed model, the TSO has no access to individual DER decisions within distribution systems. Temporal coupling therefore involves only local variables, arising for instance from State of Charge (SoC) dynamics or ramping-rate limits of DERs over the horizon. As a result, the hierarchical problem (6) can be reformulated as

$$\min_{x,y} \sum_{k \in \mathcal{K}} \Big\{ f_{0|k}(x_k, y_{0|k}) + \sum_{\ell \in \mathcal{S}^{\text{D}}} f_{\ell|k}(x_{\ell|k}, y_{\ell|k}) \Big\} \tag{20a}$$

subject to

$$g_{0|k}(x_k, y_{0|k}) = 0,\ y_{0|k} \in \mathcal{Y}_{0|k}, \qquad \forall k \in \mathcal{K} \tag{20b}$$

$$g_{\ell|k}(x_{\ell|k}, y_{\ell|k}) = 0, \qquad \forall \ell \in \mathcal{S}^{\text{D}}, \forall k \in \mathcal{K} \tag{20c}$$

$$(x_{\ell|k}, y_{\ell|k}) \in \mathcal{X}_{\ell|k} \times \mathcal{Y}_{\ell|k}, \qquad \forall \ell \in \mathcal{S}^{\text{D}}, \forall k \in \mathcal{K} \tag{20d}$$

$$M_{\ell|k+1}\, y_{\ell|k+1} = A_{\ell|k}\, y_{\ell|k}, \qquad \forall \ell \in \mathcal{S}^{\text{D}}, \forall k \in \mathcal{K} \tag{20e}$$

where the objective (20a) sums all transmission- and distribution-level costs over the entire time horizon $\mathcal{K}$. Constraints (20b) enforce the transmission power flow feasibility at each period, while (20c) (20d) impose the local

Table 2: Approximation errors: infeasibility & flexibility loss.

Cases	Topology	False Flexibility [%]					Lost Flexibility [%]				
		LDS	EDC	CR	TP	PC	LDS	EDC	CR	TP	PC
case10ba	radial	5.14	4.58	161.02	4.22	0.04	5.63	2.87	0.00	2.12	0.04
case12da	radial	2.18	2.90	67.44	3.25	0.00	6.45	3.81	0.00	2.12	0.03
case15da	radial	3.78	4.13	65.22	4.29	0.01	10.77	6.48	0.00	3.46	0.01
case15nbr	radial	2.41	2.81	70.31	2.99	0.00	9.12	6.22	0.00	2.99	0.00
case17me	radial	3.81	4.23	164.99	3.48	0.00	6.10	2.22	0.00	1.90	0.01
case18nbr	radial	2.39	2.96	68.89	3.36	0.00	10.02	6.35	0.00	3.06	0.01
case22	radial	0.99	1.31	55.30	1.41	0.01	5.25	3.39	0.00	1.32	0.01
case28da	radial	7.74	7.96	146.87	7.81	0.01	11.30	6.57	0.00	4.08	0.04
case33bw	radial	5.22	5.23	171.12	4.77	0.03	6.99	4.14	0.00	2.42	0.00
case33bw	mesh	–	3.56	–	3.96	0.00	–	6.63	–	3.43	0.01
case33mg	radial	5.51	5.41	204.76	4.84	0.00	6.56	3.91	0.00	2.35	0.01
case33mg	mesh	–	3.87	–	4.23	0.01	–	6.88	–	3.68	0.01
case34sa	radial	4.02	3.93	28.69	3.91	0.00	11.56	7.82	0.00	3.94	0.01
case38si	radial	8.18	8.01	198.12	7.07	0.04	10.24	7.29	0.00	4.06	0.11
case51ga	radial	6.83	6.25	171.91	4.82	0.04	5.47	2.70	0.00	1.69	0.03
case51he	radial	2.62	3.08	65.58	3.09	0.00	8.88	6.19	0.00	3.18	0.01
case69	radial	10.49	10.74	124.58	4.35	0.07	2.54	1.49	0.00	0.73	0.03
case74ds	radial	2.98	3.51	79.07	3.73	0.01	6.09	4.46	0.00	2.83	0.01
case85	radial	14.28	13.41	277.33	11.69	0.00	12.01	7.07	0.00	4.41	0.00
case94pi	radial	14.35	13.22	353.82	11.39	0.06	11.19	7.07	0.00	4.03	0.08
case118zh	radial	17.22	15.99	228.94	15.46	0.82	14.51	10.25	0.00	6.55	0.05
case118zh	mesh	–	6.99	–	7.07	0.05	–	6.26	–	3.90	0.03
case136ma	radial	10.27	8.77	83.39	7.08	0.22	11.10	7.84	0.00	4.96	0.27
case136ma	mesh	–	4.67	–	3.63	0.00	–	6.88	–	3.58	0.18
case141	radial	4.98	5.00	111.33	4.74	0.02	8.06	4.39	0.00	2.75	0.02
case533mt_hi	radial	4.47	2.75	419.33	2.11	0.18	5.69	4.35	0.00	1.64	0.02
case533mt_hi	mesh	–	2.30	–	2.31	0.20	–	3.67	–	1.78	0.00
case533mt_lo	radial	5.26	3.90	465.99	2.08	0.37	2.81	2.40	0.00	1.64	0.00
case533mt_lo	mesh	–	3.46	–	2.15	0.25	–	2.36	–	1.94	0.01
KIT-CN	radial	3.61	2.35	48.34	0.93	0.04	3.86	2.66	0.00	1.01	0.01
KIT-CN	mesh	–	1.75	–	1.11	0.48	–	1.82	–	0.83	0.00

* LDS, EDC, CR, TP, and PC denote, respectively, LinDistFlow, enhanced DC, convex DistFlow relaxation, tangential predictor, and the proposed predictor-corrector method.

distribution constraints and variable bounds. These first three groups of constraints are fully decoupled both in space (across subsystems) and in time, while (20e) introduces a local temporal coupling over horizon $\mathcal{K}$.

Remark 5. *The temporal coupling* (20e) *is a standard state-space relation. By partitioning* $y_{\ell|k}$ *into state variables* $y^{\text{state}}_{\ell|k}$ *and input variables* $y^{\text{input}}_{\ell|k}$*, we can rewrite* (20e) *into a standard state-space form:*

$$y^{\text{state}}_{\ell|k+1} = A^{\text{state}}_{\ell|k} y^{\text{state}}_{\ell|k} + B^{\text{input}}_{\ell|k} y^{\text{input}}_{\ell|k},$$

where $A^{\text{state}}_{\ell|k}$ *and* $B^{\text{input}}_{\ell|k}$ *are the corresponding transition and input matrices. See [56] for a discussion of multiperiod AC OPF in this framework.*

In contrast, under the TSO-managed model, the TSO has full access to individual DER dispatch and storage states (Section 2.3). The temporal link then shifts to the coupling variables themselves, replacing (20e) with,

$$M_{\ell|k+1}\, x_{\ell|k+1} = A_{\ell|k}\, x_{\ell|k}, \qquad \forall \ell \in \mathcal{S}^{\text{D}},\ \forall k \in \mathcal{K}, \tag{21}$$

so that state transitions are directly embedded in the coupling variables.

Remark 6. *In both cases,* $M_{\ell|k}$ *may vary over time to capture changing system dynamics or plug-and-play DER behavior.*

4.2 Aggregation under Local Temporal Coupling

When temporal consistency is imposed solely on the coupling variables $x_{\ell|k}$, i.e., TSO-managed model, the flexibility aggregation procedure introduced in Section 3.1 extends naturally to the multiperiod setting. For each distribution

system $\ell \in \mathcal{S}^{\mathrm{D}}$, the implicit feasible set at time step $k \in \mathcal{K}$ is given by

$$\mathcal{X}_{\ell|k}^{\mathrm{imp}} = \left\{ x_{\ell|k} \in \mathcal{X}_{\ell|k} \mid g_{\ell|k}(x_{\ell|k}, y_{\ell|k}) = 0,\ y_{\ell|k} \in \mathcal{Y}_{\ell|k} \right\},$$

These sets can be approximated independently at each time step, while the temporal coupling (21) enforces consistency across the horizon.

Under a DSO-managed model, however, the temporal coupling applies instead to the local variables $y_{\ell|k}$. In this case, the implicit feasible set $\mathcal{X}_{\ell|k}^{\mathrm{imp}}$ at time k depends explicitly on the local state $y_{\ell|k-1}$, preventing a fully parallel aggregation.

To address this, we adopt a temporal decomposition that preserves local state transitions while retaining period-wise parallelism. The key idea is to interpret each distribution grid as a *virtual storage* device. At each time step $k \in \mathcal{K}$, the coupling variables for distribution system $\ell \in \mathcal{S}^{\mathrm{D}}$ are defined as

$$x_{\ell|k} = (p_{\ell|k}^{\mathrm{pcc}},\ e_{\ell|k}^{\mathrm{pcc}}) \in \mathbb{R}^2,$$

where $e_{\ell|k}^{\mathrm{pcc}}$ denotes the virtual energy level of distribution system ℓ at time period k.

By assigning time-varying participation factors, analogous to the single-period case in Section 2.3.2, we establish a single-valued allocation between this coupling pair $(p_{\ell|k}^{\mathrm{pcc}}, e_{\ell|k}^{\mathrm{pcc}})$ and the internal DER power outputs and SoC values. This abstraction yields a virtual storage balance:

$$e_{\ell|k+1}^{\mathrm{pcc}} = e_{\ell|k}^{\mathrm{pcc}} - \Delta t\, p_{\ell|k}^{g} \qquad \forall \ell \in \mathcal{S}^{\mathrm{D}},\ k \in \mathcal{K}, \tag{22}$$

Remark 7 (Coupling for visualization)**.** *For clarity, we restrict the coupling vector to* $(p_{\ell|k}^{pcc}, e_{\ell|k}^{pcc})$*. Reactive exchanges* $q_{\ell|k}^{pcc}$ *can be included without altering the structure of the method, but are omitted here to simplify visualization.*

Remark 8 (Virtual storage and heterogeneity)**.** *The virtual-storage representation does not assume freely redistributable energy among heterogeneous units. The participation factors are imposed during aggregation, and a PCC-level point is retained only if the reconstructed unit-level powers and SoC values satisfy the device and network constraints in* $\mathcal{Y}_{\ell|k}$*. Hence, the exact feasible set under the prescribed participation policy is conservative relative to unrestricted internal redispatch, while the predictor-corrector approximation may introduce small boundary-classification errors.*

With this abstraction, multiperiod problems with local temporal coupling can be reformulated as

$$\min_{x,y} \sum_{k\in\mathcal{K}} \left\{ f_{0|k}(x_k, y_{0|k}) + \sum_{\ell\in\mathcal{S}^{\mathrm{D}}} f_{\ell|k}(x_{\ell|k}, y_{\ell|k}) \right\} \tag{23a}$$

subject to

$$\begin{aligned}
g_{0|k}(x_k, y_{0|k}) &= 0,\ y_{0|k} \in \mathcal{Y}_{0|k}, && \forall k\in\mathcal{K} && (23b)\\
g_{\ell|k}(x_{\ell|k}, y_{\ell|k}) &= 0, && \forall \ell\in\mathcal{S}^{\mathrm{D}}, \forall k\in\mathcal{K} && (23c)\\
(x_{\ell|k}, y_{\ell|k}) &\in \mathcal{X}_{\ell|k} \times \mathcal{Y}_{\ell|k}, && \forall \ell\in\mathcal{S}^{\mathrm{D}}, \forall k\in\mathcal{K} && (23d)\\
M_{\ell|k+1} x_{\ell|k+1} &= A_{\ell|k} x_{\ell|k} + B_{\ell|k} y_{\ell|k}, && \forall \ell\in\mathcal{S}^{\mathrm{D}}, \forall k\in\mathcal{K} && (23e)
\end{aligned}$$

Applying the predictor-corrector aggregation method (Section 3.1) to approximate the spatio-temporal decomposed constraints (23c)–(23d) yields the following upper-level problem:

$$\min_{x,y} \sum_{k\in\mathcal{K}} \left\{ f_{0|k}(x_k, y_{0|k}) + \sum_{\ell\in\mathcal{S}^{\mathrm{D}}} \hat{f}_{\ell|k}(x_{\ell|k}) \right\} \tag{24a}$$

subject to

$$\begin{aligned}
g_{0|k}(x_k, y_{0|k}) &= 0,\ y_{0|k} \in \mathcal{Y}_{0|k}, && \forall k\in\mathcal{K} && (24b)\\
x_{\ell|k} &\in \hat{\mathcal{X}}_{\ell|k}^{\mathrm{imp}}, && \forall \ell\in\mathcal{S}^{\mathrm{D}}, \forall k\in\mathcal{K} && (24c)\\
M_{\ell|k+1} x_{\ell|k+1} &= A_{\ell|k} x_{\ell|k} + B_{\ell|k} \hat{y}_{\ell|k}, && \forall \ell\in\mathcal{S}^{\mathrm{D}}, \forall k\in\mathcal{K} && (24d)
\end{aligned}$$

with the approximated implicit feasible sets

$$\hat{\mathcal{X}}^{\mathrm{imp}}_{\ell|k} = \left\{ x_{\ell|k} \in \mathcal{X}_{\ell|k} \;\middle|\; \hat{y}_{\ell|k}(x_{\ell|k}) \in \mathcal{Y}_{\ell|k} \right\}, \tag{24e}$$

defined for every distribution system $\ell \in \mathcal{S}^{\mathrm{D}}$ and each time step $k \in \mathcal{K}$.

4.3 Results from Real-World Scenarios

While Table 2 benchmarks diverse networks in single-period settings, practical transmission–distribution coordination also involves intertemporal coupling from storage DERs and time-varying demand. We therefore evaluate the proposed method in a dynamic ITD system comprising the IEEE 118-bus transmission grid and 28 heterogeneous distribution networks, for a total of 3,564 buses. The distribution systems span different sizes and characteristics, including `case10ba`, `case533mt`, `case118zh`, and the real-world KIT Campus North feeder (`KIT-CN`), which provides measured load and solar-generation profiles [55].

For evaluation, we first aggregate the flexibility of each distribution system independently at each hourly period. These flexibility sets are then incorporated into a transmission-level problem with temporal coupling constraints (24), enabling a direct comparison with the original centralized multiperiod formulation (20). We assessed the aggregation approaches using two performance metrics:

- **optimality gap**, defined as the deviation in objective value between the aggregated formulation (24) and the centralized optimization problem (20); and
- **constraints violations**, obtained by testing whether the aggregated dispatch decisions satisfy the original nonlinear distribution constraints (20c)-(20d).

(a) KIT North Campus (b) case118zh (c) case533mt

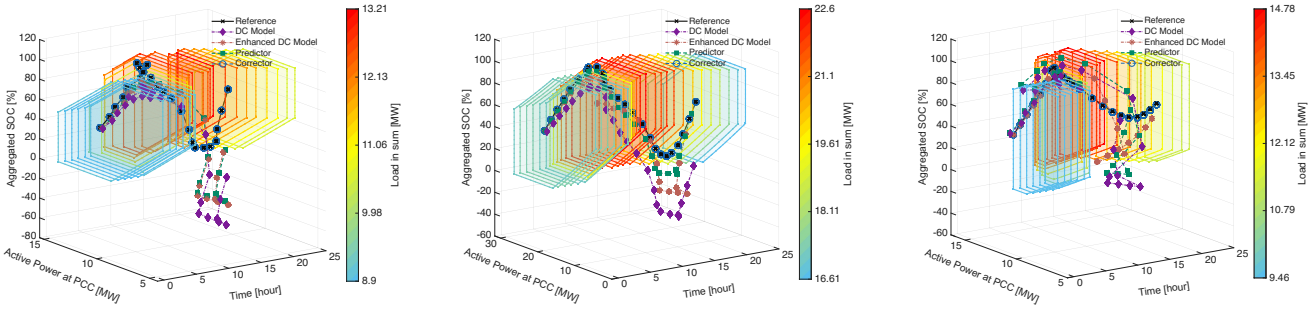

Figure 8: Comparison of optimal dispatch trajectories obtained from different aggregation methods: DC model (diamonds), enhanced DC model (stars), tangential predictor (squares), and the proposed predictor–corrector method (circles). The centralized solution with the exact AC power flow model is shown for reference (crosses). The background regions depict the time-decoupled implicit feasible sets generated by the predictor–corrector method, where red areas correspond to high-demand periods around noon and blue areas correspond to low-demand periods around midnight.

Fig. 8 illustrates the optimal dispatch trajectories for three representative distribution networks: `KIT-CN`, `case118zh`, and `case533mt`. The trajectories correspond to the proposed predictor–corrector aggregation (circles) and three linear surrogates, while the centralized AC solution to (20) serves as the reference (crosses). The background regions depict the time-decoupled implicit feasible sets generated by the predictor–corrector method, with color intensity indicating the level of total system demand. Across all three cases, the predictor–corrector trajectories align almost exactly with the reference solutions.

A complementary comparison in the power–energy domain is also shown in Fig. 9. While the linear surrogates can reasonably approximate the optimal trajectories in the power domain, their trajectories in the energy domain exhibit increasing deviations over time. This behavior, consistent with observations in [31, 57], reflects the accumulation of approximation errors inherent to linear models.

Detailed hourly evaluations of optimality gaps and constraint violations are reported in Fig. 10. The predictor–corrector aggregation method consistently tracks the reference solution with negligible cost deviation. Over the entire day, only some minor voltage violations occur in the `KIT-CN` system, each less than 10^{-5} p.u.

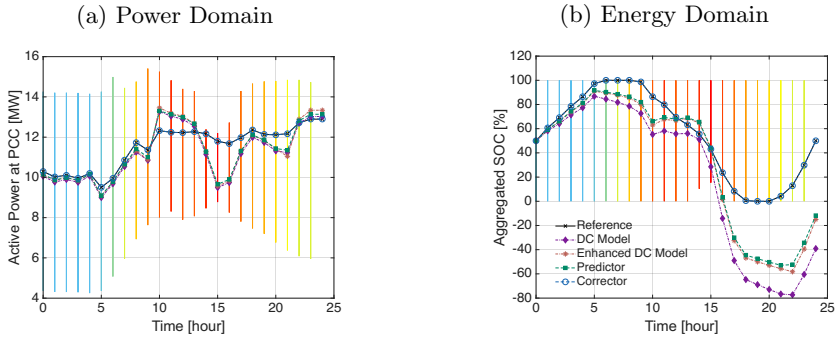

Figure 9: Comparison of optimal dispatch trajectories in the power–energy domain.

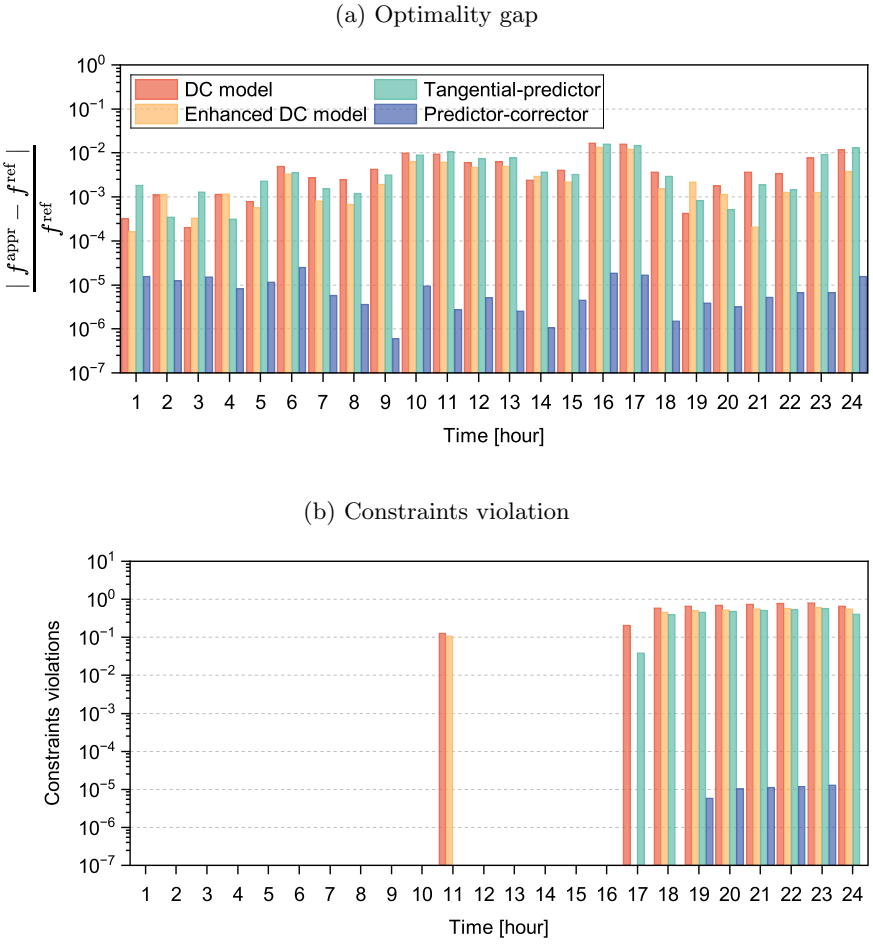

Figure 10: Performance matrices

Overall, these results confirm that the predictor–corrector aggregation provides an accurate and reliable representation of nonlinear system behavior, enabling robust flexibility characterization and secure TSO–DSO coordination.

4.4 Computation Time: Centralized vs. Aggregated

In this section, we identify and quantify the sources of the computational speedup achieved by the proposed aggregated approach. In addition to the IEEE 118-bus transmission system considered in Section 4.3, we consider a smaller ITD instance based on the IEEE 14-bus transmission system. As the centralized benchmark, the original problem (20) is solved directly using IPOPT [45]. In the aggregated approach, the predictor-corrector method first

constructs approximations of the spatio-temporally decomposed implicit feasible sets associated with (23c)-(23d), after which the resulting upper-level problem (24) is solved using the same solver. Thus, the comparisons within each test case reflect the computational effect of the aggregation reformulation rather than differences in solver choice.

Table 3: Performance of centralized and aggregated approaches

Case	Approach	AAggregation [s]	Initialization [s]	Solve [s]	Total [s]
`case14`	Centralized	–	2.12	4.94	7.06
	Aggregated	0.15	0.27	0.73	1.15
	Speedup	–	7.85×	6.77×	6.14×
`case118`	Centralized	–	37.33	20.01	57.34
	Aggregated	0.15	4.11	4.16	8.42
	Speedup	–	9.08×	4.81×	6.81×

Table 3 separates the wall-clock time into predictor-corrector aggregation, CasADi [58] initialization, including model construction and automatic differentiation, and the subsequent IPOPT solve. Including all three stages, the aggregated formulation reduces the total time from 7.06 to 1.15 seconds for `case14` and from 57.34 to 8.42 seconds for `case118`, corresponding to end-to-end speedups of 6.14 and 6.81, respectively. The initialization speedups of 7.85 and 9.08 demonstrate that dimensionality reduction lowers not only the cost of the iterative IPOPT solve but also the cost of model construction and automatic differentiation in CasADi. The sequential aggregation overhead is only 0.15 seconds in both cases and can be reduced further by executing the independent subsystem-period aggregation tasks in parallel.

Table 4: Solving time of centralized and aggregated problems

Test case	Metric	Centralized	Aggregated	Speedup
`case14`	Iterations	31	19	1.63×
	Linear Solver [s]	3.65	0.56	6.52×
	Function Eval. [s]	0.75	0.02	37.5×
	Total [s]	4.94	0.73	6.77×
`case118`	Iterations	31	22	1.41×
	Linear Solver [s]	14.96	3.14	4.76×
	Function Eval. [s]	3.07	0.16	19.2×
	Total [s]	20.01	4.16	4.81×

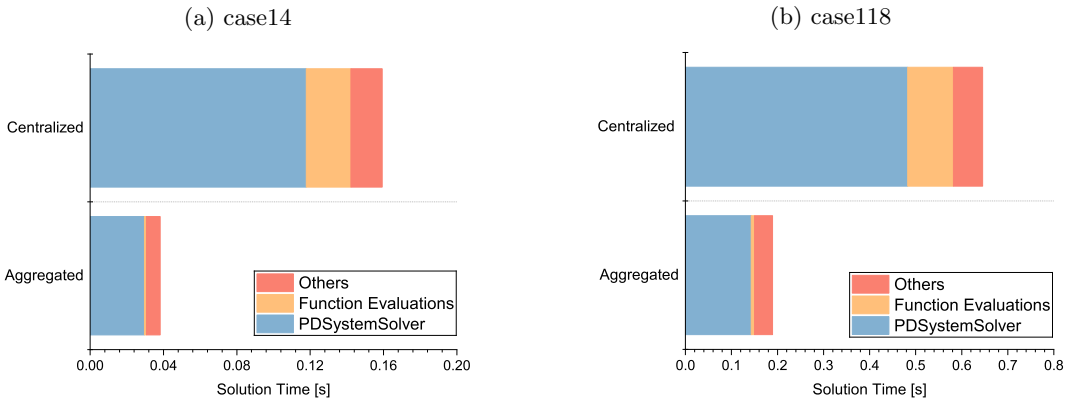

Figure 11: Average IPOPT time per iteration

To isolate the upper-level nonlinear-solver stage, Table 4 reports the IPOPT timing breakdown for the centralized and aggregated formulations. The aggregated formulation achieves IPOPT-stage speedups of 4.81-6.77, while the relative objective-value difference with respect to the centralized solution and the maximum constraint violation remain on the order of 10^{-5}, cf. Fig. 10. It also requires fewer IPOPT iterations in both test cases. Thus, the solver-stage improvement results from both a lower computational cost per iteration and a smaller number of iterations.

To explain the lower per-iteration cost, Fig. 11 presents the average IPOPT timing breakdown and identifies the primal–dual KKT solve, reported under `PDSystemSolver`, as the dominant computational component. As discussed

Table 5: Problem dimensions and KKT structure

Test case	Structural metric	Centralized	Aggregated	Reduction
`case14`	No. of variables	1.67E4	1.68E3	89.91%
	No. of constraints	1.65E4	9.08E3	44.88%
	KKT order	3.31E4	1.85E4	44.05%
	NNZ of KKT	2.83E5	5.48E4	80.62%
	KKT density	2.58E−4	1.60E−4	38.09%
`case118`	No. of variables	1.77E5	1.09E4	93.82%
	No. of constraints	1.75E5	9.18E4	47.39%
	KKT order	3.52E5	1.87E5	46.86%
	NNZ of KKT	3.20E6	5.56E5	82.62%
	KKT density	2.59E−5	1.59E−5	38.46%

in [59, Sec. 5], IPOPT reformulates general inequality constraints using bounded slack variables. After this reformulation, Table 5 shows that aggregation reduces the KKT-system order by 44.05%–46.86% and its structural number of nonzeros by 80.62%–82.62%. The KKT density also decreases by approximately 38%. These structural reductions result in 4.76–6.52 reductions in linear-system solution time.

Overall, dimensionality reduction is the primary source of the computational improvement, as it reduces both CasADi initialization costs and the cost of each IPOPT iteration. The spatio-temporal decomposition provides a complementary benefit by enabling parallel preprocessing, although the measured sequential aggregation overhead is already small.

5 Conclusion

This paper proposes a hierarchical optimization framework for integrated transmission-distribution (ITD) operation based on a predictor–corrector aggregation method. By reinterpreting path-following techniques, the method provides tractable surrogate models for distribution subsystems. The resulting aggregation directly accounts for nonlinear AC power flow constraints, is applicable to both radial and meshed distribution networks, and enables scalable pre-computation through spatio-temporal decomposition. Theoretical error bounds are established, and extensive benchmark studies demonstrate that the approach achieves speedups of 5–7 times compared to centralized nonlinear solvers. Analysis of different network configurations further highlights how structural factors shape aggregated flexibility, offering actionable insights for system operation. An open-source toolbox is released to support practical adoption and further study. Future work will extend the proposed framework by incorporating PCC voltage magnitude as an additional coupling variable, modeling transformer voltage control through discrete tap-position decisions, and evaluating its performance on larger-scale ITD networks. Another interesting direction is to investigate global optimality certification for special cases, for example through convex relaxations with verified zero-gap exactness.

Author Statements

X.D. and Y.J. contributed equally to this work. X.D. developed the methodology and algorithms. Y.J. implemented the computational framework and performed the numerical experiments. F.Z., Y.G., and V.H. contributed to the conceptual development, interpretation of the results, and revision of the manuscript. Corresponding: Y.G.

References

[1] Givisiez, A. G., Petrou, K. & Ochoa, L. F. A review on TSO-DSO coordination models and solution techniques. *Electric Power Systems Research* **189**, 106659 (2020).

[2] Karthikeyan, N., Pillai, J. R., Bak-Jensen, B. & Simpson-Porco, J. W. Predictive control of flexible resources for demand response in active distribution networks. *IEEE Transactions on Power Systems* **34**, 2957–2969 (2019).

[3] Dutta, R., Chakrabarti, S. & Sharma, A. Topology tracking for active distribution networks. *IEEE Transactions on Power Systems* **36**, 2855–2865 (2020).

[4] Kerscher, S. & Arboleya, P. The key role of aggregators in the energy transition under the latest european regulatory framework. *International Journal of Electrical Power & Energy Systems* **134**, 107361 (2022).

[5] Molzahn, D. K. *et al.* A survey of distributed optimization and control algorithms for electric power systems. *IEEE Transactions on Smart Grid* **8**, 2941–2962 (2017).

[6] Patari, N. *et al.* Distributed optimization in distribution systems: Use cases, limitations, and research needs. *IEEE Transactions on Power Systems* **37**, 3469–3481 (2021).

[7] Dai, X. *et al.* Distributed AC optimal power flow: A scalable solution for large-scale problems. *arXiv preprint arXiv:2503.24086* (2025).

[8] Wei, W., Liu, F. & Mei, S. Real-Time Dispatchability of Bulk Power Systems With Volatile Renewable Generations. *IEEE Transactions on Sustainable Energy* **6**, 738–747 (2015).

[9] Tan, Z., Zhong, H., Wang, J., Xia, Q. & Kang, C. Enforcing Intra-Regional Constraints in Tie-Line Scheduling: A Projection-Based Framework. *IEEE Transactions on Power Systems* **34**, 4751–4761 (2019).

[10] Tan, Z., Yan, Z., Zhong, H. & Xia, Q. Non-Iterative Solution for Coordinated Optimal Dispatch via Equivalent Projection—Part II: Method and Applications. *IEEE Transactions on Power Systems* **39**, 899–908 (2024).

[11] Tan, Z. *et al.* Estimating the Robust P-Q Capability of a Technical Virtual Power Plant Under Uncertainties. *IEEE Transactions on Power Systems* **35**, 4285–4296 (2020).

[12] Wen, Y., Hu, Z., He, J. & Guo, Y. Improved inner approximation for aggregating power flexibility in active distribution networks and its applications. *IEEE Transactions on Smart Grid* **early access** (2023).

[13] Dai, X. *et al.* Real-time coordination of integrated transmission and distribution systems: Flexibility modeling and distributed NMPC scheduling. *Electric Power Systems Research* **234**, 110627 (2024).

[14] Contreras, D. A. & Rudion, K. Time-Based Aggregation of Flexibility at the TSO-DSO Interconnection Point 1–5 (2019).

[15] Contreras, D. A., Müller, S. & Rudion, K. Congestion Management Using Aggregated Flexibility at the TSO-DSO Interface 1–6 (2021).

[16] Rau, N. S. Issues in the path toward an rto and standard markets. *IEEE Transactions on Power Systems* **18**, 435–443 (2003).

[17] Zhu, J. *Optimization of power system operation* (Wiley Online Library, 2009).

[18] Frank, S. & Rebennack, S. An introduction to optimal power flow: Theory, formulation, and examples. *IIE Trans.* **48**, 1172–1197 (2016).

[19] Baran, M. E. & Wu, F. F. Optimal capacitor placement on radial distribution systems. *IEEE Transactions on Power Delivery* **4**, 725–734 (1989).

[20] Baran, M. E. & Wu, F. F. Optimal sizing of capacitors placed on a radial distribution system. *IEEE Transactions on Power Delivery* **4**, 735–743 (1989).

[21] Farivar, M. & Low, S. H. Branch flow model: Relaxations and convexification (Parts I, II). *IEEE Transactions on Power Systems* **28**, 2554–2564 (2013).

[22] Yang, Z. *et al.* A linearized OPF model with reactive power and voltage magnitude: A pathway to improve the mw-only DC opf. *IEEE Transactions on Power Systems* **33**, 1734–1745 (2017).

[23] Kalantar-Neyestanaki, M., Sossan, F., Bozorg, M. & Cherkaoui, R. Characterizing the Reserve Provision Capability Area of Active Distribution Networks: A Linear Robust Optimization Method. *IEEE Transactions on Smart Grid* **11**, 2464–2475 (2020).

[24] Zhang, T. *et al.* On the Coordination of Transmission-Distribution Grids: A Dynamic Feasible Region Method. *IEEE Transactions on Power Systems* **38**, 1857–1868 (2023).

[25] Wang, S. & Wu, W. Aggregate Flexibility of Virtual Power Plants With Temporal Coupling Constraints. *IEEE Transactions on Smart Grid* **12**, 5043–5051 (2021).

[26] Wang, S., Feng, C. & You, F. Non-iterative coordination of interconnected power grids via dimension-decomposition-based flexibility aggregation. *IEEE Transactions on Power Systems* (2025).

[27] Chen, X., Dall'Anese, E., Zhao, C. & Li, N. Aggregate Power Flexibility in Unbalanced Distribution Systems. *IEEE Transactions on Smart Grid* **11**, 258–269 (2020).

[28] Chen, X. & Li, N. Leveraging Two-Stage Adaptive Robust Optimization for Power Flexibility Aggregation. *IEEE Transactions on Smart Grid* **12**, 3954–3965 (2021).

[29] Wen, Y., Hu, Z. & Liu, L. Aggregate temporally coupled power flexibility of DERs considering distribution system security constraints. *IEEE Transactions on Power Systems* **38**, 3884–3896 (2023).

[30] Lopez, L., Gonzalez-Castellanos, A., Pozo, D., Roozbehani, M. & Dahleh, M. QuickFlex: A Fast Algorithm for Flexible Region Construction for the TSO-DSO Coordination 1–6 (2021).

[31] Jiang, Y., Dai, X., Zahn, F. & Hagenmeyer, V. Enhanced flexibility aggregation using lindistflow model with loss compensation 1–6 (2025).

[32] Churkin, A., Kong, W., Gutierrez, J. N. M., Ceseña, E. A. M. & Mancarella, P. Tracing, Ranking and Valuation of Aggregated DER Flexibility in Active Distribution Networks. *IEEE Transactions on Smart Grid* 1–1 (2023).

[33] Jiang, X., Zhou, Y., Ming, W. & Wu, J. Feasible operation region of an electricity distribution network. *Applied Energy* **331**, 120419 (2023).

[34] Bandeira, M. B., Faulwasser, T. & Engelmann, A. An ADP framework for flexibility and cost aggregation: Guarantees and open problems. *Electric Power Systems Research* **234**, 110818 (2024).

[35] Engelmann, A., Bandeira, M. B. & Faulwasser, T. Approximate dynamic programming with feasibility guarantees. *IEEE Transactions on Control of Network Systems* (2025).

[36] Lee, D., Nguyen, H. D., Dvijotham, K. & Turitsyn, K. Convex Restriction of Power Flow Feasibility Sets. *IEEE Transactions on Control of Network Systems* **6**, 1235–1245 (2019).

[37] Diehl, M. *et al.* Real-time optimization and nonlinear model predictive control of processes governed by differential-algebraic equations. *Journal of Process Control* **12**, 577–585 (2002).

[38] Zavala, V. M. & Anitescu, M. Real-time nonlinear optimization as a generalized equation. *SIAM Journal on Control and Optimization* **48**, 5444–5467 (2010).

[39] Pacaud, F. Sensitivity analysis for parametric nonlinear programming: A tutorial. *arXiv preprint arXiv:2504.15851* (2025).

[40] Naik, S., Biegler, L., Bent, R. & Parker, R. Variable aggregation for nonlinear optimization problems. *arXiv preprint arXiv:2502.13869* (2025).

[41] Parker, R., Nicholson, B., Siirola, J., Laird, C. & Biegler, L. An implicit function formulation for optimization of discretized index-1 differential algebraic systems. *Computers & Chemical Engineering* **168**, 108042 (2022).

[42] Bugosen, S. I., Laird, C. D. & Parker, R. B. Process flowsheet optimization with surrogate and implicit formulations of a gibbs reactor. *arXiv preprint arXiv:2310.09307* (2023).

[43] Pacaud, F., Maldonado, D. A., Shin, S., Schanen, M. & Anitescu, M. A feasible reduced space method for real-time optimal power flow. *Electric Power Systems Research* **212**, 108268 (2022).

[44] Pacaud, F., Shin, S., Schanen, M., Maldonado, D. A. & Anitescu, M. Accelerating condensed interior-point methods on SIMD/GPU architectures. *Journal of optimization theory and applications* **202**, 184–203 (2024).

[45] Wächter, A. & Biegler, L. T. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming* **106**, 25–57 (2006).

[46] Johnson, C. R. *Matrix theory and applications* Vol. 40 (American Mathematical Soc., 1990).

[47] Brändle, J., Rousseau, J., Nahata, P. & Hug, G. On the Flexibility Potential of a Swiss Distribution Grid: Opportunities and Limitations (2026). arXiv:2510.13449.

[48] Früh, H. *et al.* Coordinated Vertical Provision of Flexibility From Distribution Systems. *IEEE Transactions on Power Systems* **38**, 1834–1844 (2023).

[49] Lee, D., Turitsyn, K., Molzahn, D. K. & Roald, L. A. Robust AC optimal power flow with robust convex restriction. *IEEE Transactions on Power Systems* **36**, 4953–4966 (2021).

[50] Wang, Z., Cui, B. & Wang, J. A necessary condition for power flow insolvability in power distribution systems with distributed generators. *IEEE Transactions on Power Systems* **32**, 1440–1450 (2016).

[51] Aolaritei, L., Bolognani, S. & Dörfler, F. Hierarchical and distributed monitoring of voltage stability in distribution networks. *IEEE Transactions on Power Systems* **33**, 6705–6714 (2018).

[52] Baran, M. E. & Wu, F. F. Network reconfiguration in distribution systems for loss reduction and load balancing. *IEEE Transactions on Power Delivery* **4**, 1401–1407 (2002).

[53] Gan, L., Li, N., Topcu, U. & Low, S. H. Exact Convex Relaxation of Optimal Power Flow in Radial Networks. *IEEE Transactions on Automatic Control* **60**, 72–87 (2015).

[54] Lee, J. M. *Introduction to Smooth Manifolds* (Springer, 2012).

[55] González-Ordiano, J. Á. *et al.* Probabilistic forecasts of the distribution grid state using data-driven forecasts and probabilistic power flow. *Applied Energy* **302**, 117498 (2021).

[56] Faulwasser, T. & Engelmann, A. Toward economic NMPC for multistage AC optimal power flow. *Optimal Control Applications and Methods* **41**, 107–127 (2020).

[57] Jiang, Y., Dai, X., Zahn, F., Guo, Y. & Hagenmeyer, V. Error accumulation using linearized models for aggregating flexibility in distribution systems 1–5 (2026).

[58] Andersson, J. A., Gillis, J., Horn, G., Rawlings, J. B. & Diehl, M. Casadi: a software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation* **11**, 1–36 (2019).

[59] Wächter, A. Short tutorial: Getting started with ipopt in 90 minutes (2009).