Economics of Grid-Supported Electric Power Markets: A Fundamental Reconsideration*

Leigh Tesfatsion

Preface

This study highlights four conceptually-problematic presumptions underpinning the current design of U.S. centrally-managed wholesale power markets that are greatly hindering their transition to decarbonized grid operations with increasingly diverse participants. These problematic presumptions are: (1) the basic transacted product is grid-delivered energy; (2) supplier revenue sufficiency can be adequately analyzed using a two-part partition of supplier cost into fixed and variable parts; (3) grid-delivered energy is a commodity that can (and should) be transacted in short-run markets at competitively determined per-unit prices; and (4) supplier revenue sufficiency is assured in such markets.

The current reality is far more daunting. The central managers for these markets are fiduciary conductors tasked with orchestrating the advance availability and possible future dispatch of increasingly diverse reserve (dispatchable power-production capabilities) to service the just-in-time power demands of increasingly diverse customers while continuing to meet all requirements for grid reliability. Thus, the essential need is long-to-short forward reserve markets able to operate as nimble physically-covered insurance mechanisms for protection against volumetric grid risk (power imbalance) during future operating periods.

One option is to muddle through, forcing central managers and market participants to focus primary attention on short-run benefits and costs for “next” energy units (MWh) in per-unit form ($/MWh), conditional on delivery location and time, without regard for the true benefits and costs of flexible just-in-time power delivery. Another option, advocated in this study, is to transition gradually to an insurance market design permitting grids to operate as efficient, reliable, and resilient flexibility-support mechanisms well-aligned with the local goals and constraints of market participants. To illustrate the latter option, a recently proposed Linked Swing-Contract Market Design is briefly reviewed. This proposed design permits dispatchable power resources to offer diverse types of reserve into a centrally-managed collection of linked forward bid/offer-based reserve markets via two-part pricing insurance contracts taking a flexible “swing” form.

*Latest revision: 7/14/2024. This study is a revised updated version of a supporting document [48] for comments e-filed to FERC for Docket AD21-10-0000 [18]. Author Information: Leigh Tesfatsion, Professor Emerita of Economics, Courtesy Research Professor of Electrical & Computer Engineering, Iowa State University, Ames, IA 50011-1054. Email: tesfatsi@iastate.edu, Homepage: https://www2.econ.iastate.edu/tesfatsi.
# Table of Contents

1 Introduction ................................................................. 3

2 Legacy Core Design of U.S. Wholesale Power Markets ............... 7

3 Essential Measurement Concepts ........................................... 14
   3.1 Overview .......................................................... 14
   3.2 Asset Definitions: Unit Measurement Distinctions ............... 14
   3.3 Unit and Per-Unit Calculations Can Mask Conceptual Error ...... 16

4 Essential Economic Concepts .............................................. 20
   4.1 Overview .......................................................... 20
   4.2 Basic Market Definitions ........................................ 20
   4.3 Commodity Market Definitions .................................. 21
   4.4 Marginal Pricing Requires Commodities ......................... 23

5 Legacy Core Design: Fundamental Conceptual Issues ................. 29
   5.1 Overview .......................................................... 29
   5.2 Reserve, Not Energy, is the Basic Product ...................... 30
   5.3 Supplier Cost Analysis Requires a 3-Part Partition ............. 35
   5.4 Grid-Delivered Energy is Not a Commodity ..................... 44
   5.5 Supplier Revenue Sufficiency Requires 2-Part Pricing ........ 47

6 Legacy Core Design: Roadblock for Grid Decarbonization ........... 50
   6.1 Overview .......................................................... 50
   6.2 Proliferation of Participation Models ........................... 50
   6.3 Proliferation of Flexibility Products ............................ 51
   6.4 Proliferation of Out-of-Market Make-Whole Payments .......... 52
   6.5 Growing Revenue Insufficiency Concerns ....................... 55
   6.6 Ptolemaic Epicycle Market-Design Conundrum ................... 55

7 What Else Can be Done: An Illustrative Alternative ................. 56
   7.1 Overview: The Linked Swing-Contract Market Design .......... 56
   7.2 Four Motivating Design Principles ................................ 57
   7.3 Innovative Design Features: Summary Form ..................... 58
   7.4 Expanded Discussion of Innovative Design Features .......... 59
   7.5 Legacy Core Design Comparisons ................................ 62

8 Conclusion: Grids as Flexibility-Support Mechanisms ............... 65

Appendices: ................................................................. 68
   A.1 Acronyms ......................................................... 68
   A.2 Glossary of Standard Transmission System Terms ................ 69
   A.3 Glossary of Standard Economic Terms ........................... 70
   A.4 Cost Types for Power Markets: Empirical Examples .......... 72
   A.5 Glossary of Swing-Contract Market Terms ..................... 73
   A.6 Invertibility of Demand and Supply Functions .................. 74

References ................................................................. 75
1 Introduction

The basic purpose of centrally-managed wholesale power markets operating over high-voltage AC transmission grids is to maintain efficient just-in-time production and transmission of bulk power to satisfy just-in-time customer power demands and grid reliability requirements.

To achieve this dynamic open-ended purpose, central managers must continually protect against volumetric grid risk. This physical risk is the possible disruption or collapse of grid operations due to real-time imbalance between withdrawal and/or inadvertent loss of power from the grid and the injection of power into the grid. Grid power withdrawals occur when the power usage of customers electrically connected to a grid exceeds their use of locally-generated power. Inadvertent power losses occur whenever power flows across a grid’s transmission lines.

In response to private economic incentives and public policy mandates encouraging grid decarbonization [22], U.S. RTO/ISO-managed wholesale power markets are transitioning from a traditionally heavy reliance on fossil-fuel based power generators to a greater reliance on Intermittent Power Resources (IPRs). These IPRs include wind farms and photovoltaic solar arrays whose weather-dependent power generation is not fully firm by storage.

The increasing participation of IPRs in U.S. RTO/ISO-managed wholesale power markets, together with initiatives such as FERC Order No. 2222 [15] encouraging more active participation by demand-side resources, has increased the uncertainty and volatility of grid net load. In consequence, as reported in [16], RTOs/ISOs are finding it harder to procure the dependable advance availability of RTO/ISO-dispatchable power-production capabilities with sufficiently diverse attributes to maintain reliable real-time balancing of net load.

Moreover, many IPRs connect to high-voltage AC transmission grids by means of power electronic inverters that convert DC to AC power, a connection technology that differs fundamentally from the traditional connection technology for fossil-fuel based power generators. At higher IPR penetration levels, this new connection technology can pose new security issues [4].

The recognition of these difficulties has led to increasingly urgent calls for action. For example, in 2021 the National Academies of Sciences (NAS) and the National Renewable Energy Laboratory (NREL) issued separate reports [36, 42] identifying

---

1 Current U.S. RTO/ISO-managed wholesale power markets consist of energy, ancillary service, and capacity markets whose operations over high-voltage AC transmission grids are managed by a Regional Transmission Organization (RTO) or Independent System Operator (ISO); see [17].

2 For the purposes of this study, an Intermittent Power Resource (IPR) is defined to be a grid-connected power resource whose power injections and/or withdrawals are not mediated through some form of aggregator and are not fully controllable by centrally-managed dispatch.

3 The net load of a grid at a given point in time consists of power withdrawals and inadvertent power losses (e.g., transmission line losses) net of non-dispatched power injections.

4 In practice, reliable real-time balancing of net load means maintaining net-load balance within acceptable tolerance levels over time.

Strongly encouraged by these calls for action, efforts are underway to improve the conceptual and operational design of U.S. electric power markets. As discussed in later sections of this study, these efforts are taking diverse forms. Nevertheless, they largely adhere to the following eight broadly-accepted goals:5

**Goal (G1): Incentive Alignment.** The market design should be well-aligned with the local objectives and constraints of market participants, including privacy concerns, thus ensuring their voluntary participation.

**Goal (G2): Resource Adequacy.** The market design should provide incentives for new resources to enter in sufficient quantity to accommodate retirements, de-ratings, and increases in power demand over time while maintaining adequate reserve to address uncertainty and volatility of net load.

**Goal (G3): Efficiency.** The market design should be efficient, i.e., it should not waste resources. To promote short-run efficiency, the design should permit the production, transmission, and distribution of power from existing resources to be based on accurate assessments of benefits and costs. To promote longer-run efficiency, the design should encourage the development and adoption of new technologies permitting increased benefit from power use and reduced cost for power production and transmission.

**Goal (G4): Reliability and Resiliency.** The market design should ensure continual net-load balancing during normal power system operations, despite weather events and other anticipated types of disturbances. The design should also support rapid recovery and return to net load balancing following sudden major disruptions, such as the loss of a line or a generation unit.

**Goal (G5): Fairness.** The market design should be fair, i.e., it should provide an even playing field for all actual and potential market participants. Thus, it should permit and encourage actual and potential market participants to compete for the provision of reserve and for the production, procurement, delivery, and use of electric power. It should also avoid the unintended creation of structural and strategic market advantages for some participants to the detriment of others.

**Goal (G6): Conceptual Coherency and Transparency.** The market design should be conceptually coherent, and market rules and operations under the design should be as transparent as possible.

---

5 The specific expressions (G1)–(G8) for these eight goals are based on Oren [38, Sec. II.A], Tesfatsion et al. [44, Sec. 2], and Tesfatsion [45, Sec. 2.2].
Goal (G7): Minimum Administrative Intervention. The market design should discourage ad-hoc rule-making and decision-making by administrators. To further this goal, market rules and operations should be based on service requirements rather than on irrelevant physical and operational attributes of resources, to an extent compatible with the attainment of other design goals. Wherever possible, mechanisms should be instituted to permit and encourage transition to a design with limited administrative control.

Goal (G8): Supportive of Previous Reform Efforts. The market design should be in accordance with FERC, RTO/ISO, and stakeholder efforts to promote increased market access, pay for verified performance, demand-side participation, and encouragement of private initiative.

Despite the general acceptance of goals (G1)–(G8), ongoing efforts to reform the core design of current U.S. RTO/ISO-managed wholesale power markets have been contentious. A key theme of this study is that much of this contention arises from four conceptually-problematic economic presumptions built into this core design. In brief preliminary form, these presumptions are as follows:

Problematic Presumption (P1):
The basic transacted product for grid-supported centrally-managed wholesale power markets is grid-delivered energy (MWh), i.e., accumulations of flows of power (MW) at designated grid locations b during designated operating periods T with duration measured in hours (h).

Problematic Presumption (P2):
For careful analysis of supplier revenue sufficiency in such markets, it suffices to partition total supplier cost into a “variable” component dependent on the quantity supplied and a “fixed” component independent of the quantity supplied.

Problematic Presumption (P3):
Grid-delivered energy conditional on delivery location b and delivery period T is a commodity, i.e., its units (MWh) are perfect substitutes. Thus, these units can (and should) be transacted in a spot market M(b,T) at a uniform per-unit locational marginal price LMP(b,T) ($/MWh) determined in accordance with the standard competitive (marginal benefit = marginal cost) spot-pricing rule.

Problematic Presumption (P4):
The total supplier revenue attained in the spot markets in (P3) will suffice to cover total supplier cost.

Presumptions (P1)–(P4) reflect the comforting static view that the primary role of U.S. RTOs/ISOs is to oversee the determination of unit prices ($/MWh) for grid-delivered energy (MWh) in collections of short-run competitive markets, weakly cross-correlated by needed real-time ancillary service adjustments.6

6 The need for ancillary service adjustments, e.g., the real-time dispatch of generation capacity unencumbered by market-determined dispatch obligations, arises from inevitable discrepancies between scheduled and delivered energy, and between delivered energy and the actual flow of customer power withdrawals. These discrepancies require continual real-time corrective actions across distinct grid locations to maintain continual power balance at each of these locations.
The current dynamic reality is far more daunting: U.S. RTOs/ISOs are fiduciary conductors tasked with orchestrating the availability and possible future dispatch of increasingly-diverse dispatchable power resources to service the just-in-time power demands of increasingly diverse customers while meeting just-in-time power requirements for reliable grid operation. This orchestration is severely constrained by the physical complexity of power flows across transmission grids: a power injection anywhere flows everywhere.

Recognition of this dynamic reality results in strong counterclaims to (P1)–(P4), expressed below in brief preliminary form:

**Counter-Claim (CC1):**
Suppliers participating within a grid-supported centrally-managed wholesale power market provide **two** basic types of product:

- **Physically-Covered Insurance:** Guaranteed availability of power-production capabilities for possible central-manager dispatch during future operating periods, to reduce volumetric grid risk;
- **Real-Time Power Delivery:** Actual delivery of power in response to central-manager dispatch signals received during an operating period to satisfy just-in-time customer power demands and grid reliability requirements.

**Counter-Claim (CC2):**
A conceptually-sound analysis of revenue sufficiency for a supplier participating within a grid-supported centrally-managed wholesale power market requires a partitioning of this supplier’s total cost into **three** components: (i) non-avoidable fixed cost (“sunk cost”); (ii) avoidable fixed cost; and (iii) variable cost.

**Counter-Claim (CC3):**
Within the context of a grid-supported centrally-managed wholesale power market, **grid-delivered energy is not a commodity**. Although grid-delivered energy has a standard unit of measurement – a megawatt-hour (MWh) – central managers and market participants do **not** consider these units to be perfect substitutes (economically equivalent) conditional on grid delivery location and time. Thus, “marginal benefit” and “marginal cost” are not well-defined concepts for grid-delivered energy.

**Counter-Claim (CC4):**
A grid-supported centrally-managed wholesale power market M(T) for an operating period T must necessarily be a **forward** market due to the speed of real-time operations. To ensure revenue sufficiency, a supplier i participating in M(T) should be permitted to submit supply offers in a **two-part pricing** form enabling full compensation for:

---

7 It has long been recognized by economists that two-part pricing can be used by monopolistic suppliers in spot-market settings as price-discrimination instruments permitting extraction of “net surplus” from buyers; see, for example, the discussion of this spot-market issue in Section 4.4. The recommended use of two-part pricing in (CC4) is for an altogether different context: namely, suppliers participating in forward markets might have to incur avoidable fixed costs to guarantee their ability to fulfill a range of possible real-time delivery obligations under contracts with swing
(1) *avoidable fixed cost* that supplier $i$ must incur to guarantee the *availability* of reserve (dispatchable power-production capabilities) for possible central dispatch during $T$, whether or not supplier $i$ is actually dispatched to provide power delivery during $T$;

(2) *variable cost* (if any) that supplier $i$ incurs for *actual* dispatched power delivery during $T$.

The remaining sections of this study are organized as follows. Section 2 presents a careful summary description of the *Two-Settlement System* constituting the core design feature for all seven U.S. RTO/ISO-managed wholesale power markets. Basic measurement and economic concepts essential for undertaking a fundamental reconsideration of this core design feature are reviewed in Sections 3 and 4.\(^8\)

Section 5 highlights the dependence of the Two-Settlement System on the four economic presumptions (P1)–(P4) and carefully presents and analyzes the counter-claims (CC1)–(CC4) to these four presumptions. Section 6 then considers how the retention of the Two-Settlement System – hence presumptions (P1)–(P4) – as a core design feature is hindering the ability of U.S. RTO/ISO-managed wholesale power markets to transition smoothly to decarbonized grid operations.

Section 7 considers what else can be done. Specifically, could the Two-Settlement System be advantageously replaced by a conceptually-consistent alternative? Or, as some have argued, would the only alternative be the inefficient adoption of zonal pricing, or a return to an inefficient reliance on top-down cost-based prices set by administrators?

As a counterpoint to the latter pessimistic view, Section 7 briefly reviews an alternative *Linked Swing-Contract Market Design* \([45]\) proposed for grid-supported centrally-managed wholesale power markets. It is argued that this alternative design is consistent with goals (G1)–(G8) and counterclaims (CC1)–(CC4), and is well-suited for the scalable support of increasingly decarbonized grid operations with more active participation by demand-side resources.

Concluding remarks are given in Section 8. Quick-reference guides for acronyms, terms, and key concepts used in this study are provided in Appendices A.1–A.5. Technical materials regarding the invertibility of demand and supply functions, used in support of counterclaims (CC1)–(CC4), are provided in Appendix A.6.

## 2 Legacy Core Design of U.S. RTO/ISO-Managed Markets

The development of the legacy core *Two-Settlement System* \([35]\) supporting current U.S. RTO/ISO-managed wholesale power market operations can be traced in a series of reports released by the U.S. Federal Energy Regulatory Commission (FERC), culminating in a 2003 White Paper \([8]\).
This White Paper envisions grid-delivered energy (power accumulations) at designated grid delivery (pricing) locations during designated operating periods to be the basic transacted product for U.S. RTO/ISO-managed wholesale power markets. As illustrated in Fig. 1, these grid-delivered energy amounts are to be determined in accordance with a daily bid/offer-based RTO/ISO-managed Day-Ahead Market (DAM) operating in tandem with a daily bid/offer-based RTO/ISO-managed Real-Time Market (RTM).

The overall goal of the Two-Settlement System is to permit energy transactions at designated grid delivery locations during designated operating periods to be efficiently determined by the demand bids and supply offers of energy buyers and suppliers. With this overall goal in mind, the Two-Settlement System is designed to be in accordance with the determination of market-clearing prices and quantities in competitive commodity spot markets to an extent consistent with maintaining the reliable support of a transmission grid susceptible to transmission-line congestion.

The purpose of the RTO/ISO-managed DAM held on each day D is to commit RTO/ISO-dispatchable generation units for day D+1 that permit the RTO/ISO to ensure efficient continual net-load balancing during day D+1.

Load-Serving Entities (LSEs), acting on behalf of managed customers, submit demands bids into the day-D DAM for the purchase of energy at grid delivery locations for each hour H of day D+1. Each such demand bid can take the form of a fixed (non-dispatched must-service) energy demand. It can also include or take the form of a price-sensitive energy demand schedule.

Generation units submit supply offers into the day-D DAM for the sale of energy at grid delivery locations for each hour H of day D+1. Each such supply offer can take the form of a fixed (non-dispatched must-service) energy supply. It can also
include or take the form of a dispatchable price-sensitive energy supply schedule if the generation unit has installed real-time telemetry permitting the RTO/ISO to incrementally adjust the unit’s energy supply by real-time dispatch signals.

The RTO/ISO conducts a bid/offer-based SCUC/SCED optimization for the day-D DAM in combined or consecutive form. This optimization is conditional on current state conditions, submitted bids and offers, and RTO/ISO forecasts for non-dispatched injections and/or withdrawals of power at each grid delivery location $b$ during each hour $H$ of day $D+1$. The optimization is subject to system constraints that include transmission line capacity limits, reserve requirements, and a power balance constraint at each grid delivery location $b$ for each hour $H$ of day $D+1$.

Fixed supply offers and fixed demand bids are directly entered into the power-balance constraints for the SCUC/SCED optimization. A SCUC optimization is then conducted to determine a binary (yes/no) commitment solution for each dispatchable generation unit for each hour $H$ of day $D+1$ indicating whether or not this generation unit is required to be available for possible RTO/ISO-dispatch during hour $H$ of day $D+1$. Conditional on this SCUC commitment solution, a SCED optimization is conducted to determine cleared (accepted) price-sensitive supply offers from committed dispatchable generation units and cleared (accepted) price-sensitive demand bids from LSEs at each grid delivery location $b$ for each hour $H$ of day $D+1$.

DAM settlements for these fixed/cleared offers and fixed/cleared bids are determined by locational marginal pricing [40]; that is, by the pricing of grid-delivered energy (MWh) conditional on grid delivery location and operating period, subject to system constraints. The Locational Marginal Price $LMP(b,H,D+1) (\$/MWh)$ determined in a day-D DAM SCED optimization for scheduled energy deliveries at a grid delivery location $b$ during some hour $H$ of day $D+1$, conditional on SCUC-determined generation-unit commitments, is the dual variable solution for the power balance constraint at location $b$ for hour $H$.10

An RTM is a daily collection of sub-markets for near-term future time-periods with relatively short durations (e.g., 5 minutes). An RTM sub-market for a near-term future time-period $\tau$ is cleared by an RTO/ISO-managed SCED optimization conditional on previously-determined unit commitments plus RTO/ISO forecasts for non-dispatched injections and/or withdrawals of power at each grid delivery location $b$ during $\tau$. RTM SCED optimizations are similar in form to DAM SCED optimizations except that RTMs generally impose stricter restrictions on the submission of price-sensitive demand bids.11

The purpose of these RTM sub-markets (plus supplemental unit-commitment processes) is to permit the successive updating of previously determined optimal SCUC/SCED solutions to take into account updated RTO/ISO forecasts as well

---

9 SCUC is an acronym for Security-Constrained Unit Commitment, and SCED is an acronym for Security-Constrained Economic Dispatch.

10 See [31] and [50] for studies focusing on the conceptualization and mathematical derivation of LMPs in U.S. RTO/ISO-managed wholesale power markets.

11 For example, ERCOT restricts RTM submission of price-sensitive demand bids to Qualified Scheduling Entities (QSEs) that manage QSE-controllable load sources; see [2, Sec. 4.3].
as unanticipated changes in other relevant factors. Any adjustments needed in the scheduled energy deliveries determined in the day-D DAM for some hour H of day D+1, as indicated by the solutions for RTM sub-markets conducted after the close of the day-D DAM but prior to hour H, are settled using the LMPs determined in these RTM sub-markets.

Figure 2 illustrates the determination of an optimal market-clearing demand-equals-supply (D=S) solution for a given hour H on a designated operating day D+1 by means of a bid/offer-based RTO/ISO-managed DAM SCED optimization conducted during day D, conditional on a SCUC-committed collection of generation suppliers. For ease of depiction, absence of grid congestion and line losses is assumed to hold at this optimal solution.

Note the market depicted in Figure 2 is indeed an energy market, despite the appearance of power levels (MW) along the quantity axis. These power levels represent possible choices for a maintained power level $p$ (MW) during operating hour H (1h). Consequently, the choice of a power level $p$ is equivalent to the choice of a grid-delivered energy-block $p \cdot 1h$ (MWh).
The aggregate demand and supply schedules \(D\) and \(S\) in Fig. 2 are constructed from the LSE demand bids and generation supply offers submitted to the day-D DAM by two LSE buyers (B1,B2) and three generation suppliers (S1,S2,S3). Specifically, the aggregate demand schedule \(D\) in Fig. 2 gives, from left to right, the highest purchase reservation value \(($/\text{MWh})\) – i.e., the highest maximum willingness to pay \(($/\text{MWh})\) – for each successive unit (MW) increase in the maintained power level \(p\) for \(H\), where this highest purchase reservation value is calculated across all buyers (here B1 and B2). The aggregate supply schedule \(S\) in Fig. 2 gives, from left to right, the lowest sale reservation value – i.e., the lowest minimum acceptable payment – for each successive unit (MW) increase in the maintained power level \(p\) for \(H\), where this lowest sale reservation value is calculated across all suppliers (here S1, S2, and S3).

The optimal solution for the market depicted in Fig. 2 consists of the intersection points of the aggregate demand and supply schedules, \(D\) and \(S\). This solution consists of a set of points in the \((p,\text{LMP})\) plane with a common optimal power level \(p^* = 75\) (MW) and a range of optimal price levels \(\text{LMP}^*\) ($/\text{MWh})\). The multiplicity of optimal price levels arises because the demand bids and supply offers submitted to this DAM take a required step-function form that results in flat vertical and horizontal segments for the aggregate demand and supply schedules. Note, also, that the depicted optimal solution is not conditioned on grid delivery location because absence of grid congestion and line losses is assumed at this optimal solution.\(^{12}\)

Market efficiency is said to hold for a market \(M\) if the participating buyers and suppliers are achieving maximum possible total net benefit from this participation – i.e., maximum possible total buyer benefit ($) net of total supplier variable cost ($). Market efficiency holds at the optimal market-clearing (\(D=S\)) solution depicted in Fig. 2 due to the assumed absence of grid congestion and line losses.\(^{13}\)

For later purposes, additional aspects of the U.S. RTO/ISO-managed DAM SCED optimization formulation depicted in Fig. 2 are highlighted below.

- Each price-sensitive energy demand (supply) schedule that is bid (offered) by a buyer (supplier) \(k\) into an RTO/ISO-managed DAM held on day \(D\) for a particular

---

\(^{12}\) An optimal market-clearing (\(D=S\)) solution for hour \(H\) of day \(D+1\), determined by a day-D DAM SCED optimization formulated as a DC optimal power flow problem for a loss-less grid, will determine a common optimal LMP level (or a common set of optimal LMP levels) at each grid delivery location for hour \(H\) if no grid congestion occurs at this optimal solution, i.e., if no transmission line capacity constraint is active at this optimal solution. Conversely, if any transmission line congestion occurs at this optimal solution, some separation of optimal LMP levels (or sets of optimal LMP levels) across grid delivery locations will usually (but not necessarily) occur for hour \(H\). See [31] for a more extensive discussion of these points.

\(^{13}\) As carefully explained and illustrated in [26], [43], and [45, Ch. 12], market efficiency holds at a DAM optimal market-clearing (\(D=S\)) solution point for a loss-less grid in the absence of any LMP separation. However, if LMP separation occurs at this solution point, the RTO/ISO itself extracts a non-negative (typically positive) ”congestion rent” ($) from market operations at this solution point. In this case, accurate determination of market efficiency would require accurate determination of the subsequent use made of this RTO/ISO-extracted congestion rent, and the effects of this use on the total net benefit attained by the market participants. Conversely, no RTO/ISO extraction of congestion rent occurs in the absence of LMP separation.
operating hour $H$ during day $D+1$ must include $k$’s grid-location $b(k)$ together with a finite number $N_k \leq N$ of (MW/price)-blocks $B_n(k)$, $n = 1, \ldots, N_k$, where $N$ is set by the RTO/ISO; e.g., $N = 10$ in ISO-NE [23] and $N = 9$ in MISO [34].

- Each $B_n(k)$ consists of a range $[p_{k,n-1}, p_{k,n}]$ of power levels along the horizontal power axis satisfying $0 \leq p_{k,n-1} < p_{k,n}$ and a non-negative per-unit energy price $\pi_{k,n} ($/MWh) along the vertical price axis.

- The interpretation of $B_n(k)$ for a buyer (supplier) $k$ is that $\pi_{k,n}$ is the maximum (minimum) per-unit energy price that $k$ is willing to pay (be paid) for procurement (supply) of a next (“marginal”) increment $E_{k,n} = (p_{k,n} - p_{k,n-1}) \cdot 1h$ of grid-delivered energy at $b(k)$ during $H$, given that $k$ has already agreed to procure (supply) grid-delivered energy in amount $p_{k,n-1} \cdot 1h$ at $b(k)$ during $H$.

- If $k$ is a buyer (supplier), the resulting price sequence $(\pi_{k,1}, \ldots, \pi_{k,N_k})$ is required to be non-increasing (non-decreasing).

- The RTO/ISO then constructs and plots an aggregate demand (supply) schedule in the $(p, \pi)$-plane for grid-delivered energy at each grid-location $b$ during $H$ by plotting – in descending (ascending) price order – all of the blocks $B_n(k)$ submitted for $H$ by all of the buyers (suppliers) $k$ at grid location $b(k) = b$.

**Bottom Line:** Each buyer (supplier) $k$ participating in this DAM is required to express their willingness to pay (be paid) for procurement (supply) of successive increments $E_{k,n}$ of energy (accumulated power), to be grid-delivered at $b(k)$ during $H$ by means of RTO/ISO-dispatched power flows, without any way to express preferences regarding the dynamic attributes of these power flows: e.g., capacity profile, ramp-rate profile, and exact timing within $H$.

Participants in an RTO/ISO-managed DAM (or RTM) for a future operating period $T$ are assured, by design, that any grid-delivered energy amounts the RTO/ISO announces have been scheduled for $T$ are supported by scheduled transmission capacity. Traders who determine and settle physically-covered bulk energy trades for $T$ through other venues, such as privately-negotiated bilateral trades, must secure transmission-capacity support for these physically-covered trades by self-scheduling them as fixed-form energy bids and offers in a DAM (or RTM) conducted for $T$. In addition, these traders might need to procure supporting contracts (e.g., Financial Transmission Rights and Contracts-for-Difference) to ensure that the settlement terms they agreed to in these other venues are not disrupted by obligatory DAM/RTM LMP price settlements; see [45, Sec. 13.9].

All seven U.S. RTO/ISO-managed wholesale power markets (CAISO, ERCOT, ISO-NE, MISO, NYISO, PJM, SPP) – depicted in Fig. 3 – are currently operating in accordance with FERC’s proposed Two-Settlement System, even though ERCOT (lying entirely within the state of Texas) is not in fact subject to FERC jurisdiction. The seven RTOs/ISOs depicted in Fig. 3 operate over a physical high-voltage AC transmission grid consisting of three separately-synchronized parts; see Fig. 4.
Fig. 3 North American RTO/ISO-managed wholesale power markets. (Public domain: [14])

Fig. 4 North American RTOs/ISOs operate over a physical high-voltage AC transmission grid consisting of three separately-synchronized parts. (Public domain image by Jong Suk Kim)
Finally, FERC’s proposed Two-Settlement System did not include guidance for provision and settlement of ancillary services. Rather, FERC explicitly delegated ancillary service aspects of power system management to the individual states participating in each RTO/ISO [8, p. 11]. Consequently, as reported in [6, Tables 1-2] and [16, Table 1, p. 6], ancillary service procurement and settlement processes differ widely across the seven U.S. RTOs/ISOs.

3 Essential Measurement Concepts

3.1 Overview

This section defines, explains, and illustrates measurement concepts essential for the careful analysis of Two-Settlement System operations in U.S. RTO/ISO-managed wholesale power markets. These concepts will be used throughout the remaining sections of this study.

3.2 Asset Definitions: Unit Measurement Distinctions

Definition D1: Standard Unit of Measurement. A specified positive amount $u$ of some phenomenon that is commonly used (by law or by convention) to measure the magnitude of general amounts of this phenomenon in a comparable manner.

Seven Standard International (SI) Base Units for Physical Phenomena: Length measured by meter (m); Mass measured by kilogram (kg); Time measured by second (s); Electric Current measured by Ampere (A); Thermodynamic Temperature measured by degree Kelvin (K); Amount of Substance measured by mole (mol); Luminous Intensity measured by candela (cd).

Examples of Units Defined as Functions of SI Base Units: Pound (lb) =: Unit for Weight: 1lb =: 0.45359237kg; Metric Ton (mt) =: Unit for Weight: 1mt =: 1000kg; Watt (W) =: Unit for Power: 1W =: [1kg][1m]$^2$[1s]$^{-3}$; 1kW =: 1000W; 1MW =: 1000kW; Volt (V) =: Unit of Electric Potential: 1V =: [1W][1A]$^{-1}$; Hour (h) =: Unit for Time: 1h =: 60s; Watt-hour (Wh) =: Unit for Energy: 1Wh =: [1W][1h]; 1kWh =: 1000Wh; 1MWh =: 1000kWh; Hertz (Hz) =: Unit for Frequency: 1Hz =: [1 cycle] · [1s]$^{-1}$.

Other Commonly Used Physical Measurement Units: Degree Fahrenheit (°F), a normalized temperature unit such that water freezes at 32°F and boils at 212°F; British Thermal Unit (Btu), a unit of heat defined to be the quantity of heat required to raise by 1°F the temperature of 1lb of liquid water currently at the temperature ($\approx$ 39°F) at which water has its greatest density; Person-Hour, a unit of human labor defined to be one hour of work by one person.

Ancillary services are support services for grid reliability [16, Appendix]. Examples include: “black-start” services for restoration of power flow to a collapsed grid; reactive power support for voltage control; and net-load balancing services provided by on-line generation units with unencumbered generation capacity or by off-line relatively quick-start generation units.
Definition D2: Asset. Anything in physical or financial form that can function as a store of value. In principle, an asset can be constructively characterized as a vector of multiple possibly-correlated valued attributes.

Asset Example: Apple =: (location; time; weight; shape; color; crispness; ... )

Human Asset Examples: Health; Stamina; Hand-grip strength; Intelligence.

Social Asset Examples: Beauty; Labor capability; English language verbal fluency.

Physical Asset Examples: Electric power quality; Mineral hardness; Fruit; Battery; House; Grid-delivered energy.

Financial Asset Examples: Personal loan; Home fire insurance contract; Home mortgage contract; U.S. Treasury bill; Common stock share.

The next definition D3 – for a \(u\)-asset – is a new asset categorization introduced in [48]. The set of all \(u\)-assets is strictly nested between the set of all assets, previously defined in D2, and the set of all commodities defined below in D4. As will be shown in Section 4.4, grid-delivered energy is an example of a \(u\)-asset that is not a commodity. Consequently, the explicit recognition of this new asset category could facilitate the careful economic analysis of grid-supported electric power markets.

Definition D3: \(u\)-Asset. An asset \(A\) that has a standard unit of measurement \(u\).

Human \(u\)-Asset Examples: Hand-grip strength measured by standardized test score; Intelligence measured by Intelligence Quotient (IQ).

Social \(u\)-Asset Examples: Labor-capability measured by person-hour; English language verbal fluency measured by standardized test score.

Physical \(u\)-Asset Examples: Mineral hardness measured by Mohs scale; Fruit measured by pounds (lbs); Battery measured by number of batteries; House measured by number of houses; Grid-delivered energy measured by Watt-hour (Wh).

Financial \(u\)-Asset Examples: Home fire insurance contract measured by coverage amount ($); 30-year fixed home mortgage contract at 7.12% interest measured by mortgage principle ($); 1-year U.S. Treasury bill measured by redemption value ($); Share of Duke Energy common stock (NYSE:DUK) measured by current market value ($/share).

Examples of Assets that are Not \(u\)-Assets: Health; Stamina; Beauty; Electric power quality.
Definition D4: Commodity. A physically-exchangeable \( u \)-asset \( Q \) such that, conditional on location and time, each \( Q \)-trader (supplier and/or buyer) considers all \( Q \)-units \( u \) available for trade to be perfect substitutes for each other; that is, to have the same economic value.\(^{15}\)

Commodities Defined by Legally-Enforceable Standards:

Agricultural Examples: HoneyCrisp Apples measured by pound (lb); No. 1 Hard Red Winter Wheat measured by metric ton (mt); Champagne measured by millilitres (ml).

Industrial Examples: DURACELL AA 1.5v Batteries measured by number of batteries; Henry Hub Natural Gas (Louisiana) measured by metric million Btu (mmBtu).

Financial Examples: 1-year U.S. Treasury bills issued on 1 January 2024 with $100 redemption value measured by number of bills; Shares of Duke Energy common stock (NYSE:DUK) measured by number of outstanding shares.

Examples of \( u \)-Assets that are Not Commodities:

Hand-grip strength; Intelligence; Labor capability; Verbal English language fluency; Mineral hardness; Fruit; House; Grid-delivered energy; Home fire insurance contract; Home mortgage contract.

3.3 Unit and Per-Unit Calculations Can Mask Conceptual Error

3.3.1 Technical Preliminaries

Let \( R \) denote the set of real numbers.\(^{16}\) The standard algebraic operators that act on elements of \( R \) include: addition (+); subtraction (−); multiplication (×); division (÷); and equality (=). The set \( R \) together with its standard algebraic operators is hereafter referred to as the Real Number System.

\(^{15}\) The “same economic value” assigned to all units \( u \) of a commodity \( Q \) available for trade at a given location and time can differ across different \( Q \)-traders. Nevertheless, \( Q \)-trading is facilitated as follows: Commodity \( Q \) can be sold by a \( Q \)-supplier \( i \) to a \( Q \)-buyer \( j \) in bulk (multi-unit) amount \( q' \) (measured in \( u \)) at a common per-unit price \( \pi' \) (measured in $/\( u \)) as long as: (i) \( \pi' \) is greater or equal to the common economic value assigned by supplier \( i \) to each possible “last” \( Q \)-unit that supplier \( i \) could provide at \( q' \); and (ii) \( \pi' \) is less than or equal to the common economic value assigned by buyer \( j \) to each possible “last” \( Q \)-unit that buyer \( j \) could procure at \( q' \).

\(^{16}\) In standard texts on real analysis, the set \( R \) is often defined axiomatically as a complete Archimedean ordered field. Alternatively, \( R \) is sometimes defined as the end-result of a process taking the following general form: Step-1: Assume the existence of various primitive set-theoretic concepts; Step-2: Use the Step-1 assumptions to develop the set \( N = \{1, 2, 3, \ldots\} \) of natural numbers; Step-3: Use the Step-1 assumptions and development of \( N \) to develop the set \( Z = \{0, 1 \ldots, 2, 3, \ldots \} \) of integers; Step-4: Use the Step-1 assumptions and development of \( Z \) to develop the set \( Q = \{m/n \mid m, n \in Z, \text{ and } n \neq 0\} \) of rational numbers; Step-5: Use the Step-1 assumptions and development of \( Q \) to develop the set \( R = \{r, \ldots\} \) of real numbers.
The International System of Units (SI) is commonly referred to as the Metric System. The Metric System consists of the seven real-valued SI Base Units \(\{m, \text{kg}, \text{s}, \text{A}, \text{K}, \text{mol}, \text{cd}\}\) identified in Section 3.2 together with real-valued units derived from these SI Base Units by means of standard algebraic operators.

The SI Base Units are each defined in terms of a latest internationally agreed-upon value for a physical constant pertaining to some physical aspect of the Real World, where these physical constants are assumed to be mutually independent of each other. For example, the SI Base Unit for length is a meter (m), defined in terms of the latest internationally agreed-upon value for \(c\), the speed of light in vacuum space. The SI Base Unit for mass is a kilogram (kg), defined in terms of the latest internationally agreed-upon value for \(\hbar\), the Planck constant. The SI Base Unit for electric current is an Ampere (A), defined in terms of the latest internationally agreed-upon value for \(e\), the electrical charge carried by an electron.

3.3.2 Unit and Per-Unit Calculations Must be Undertaken with Care

Consider the status (True, False, Undecidable, Ambiguous, Undefined, ...) assigned to each of the following five statements:

Statement S1: \(10 = 10\)

Status: True statement within the Real Number System. Ambiguous statement (10 of what?) within the Metric System and for the Real World.

Statement S2: \(10\) pounds of apples = \(10\) pounds of apples

Status: Undefined statement within the Real Number System (what is a pound? what is an apple?) and the Metric System (what is an apple?). Ambiguous statement for the Real World: no two separate apples are physically identical, and physical differences can affect production cost, eating preferences, and consumption benefits; thus, what type of “equality” is “=” meant to signify?

Statement S3: \(2\text{MWh} = 2\text{MWh}\)

Status: Undefined statement within the Real Number System (what is a MWh?). True statement within the Metric System. Ambiguous statement for the Real World (what type of “equality” is “=” meant to signify?).

Important Additional Remarks Regarding the Status of Statement S3:

Regarding the Real World ambiguity of Statement S3, consider the following possibilities. The energy (2MWh) on each side of the operator “=” could be identically-stored energy located at a grid location b at a particular point in time. In this case the operator “=” could represent physical equivalence.

Alternatively, the energy (2MWh) on each side of the operator “=” could represent energy that has been grid-delivered at b during the course of some operating day D, i.e., the accumulation of a flow of power (MW) injected at b during D. For example, these injections might have occurred: (i) throughout
all 24 hours of day \(D\) at a constant level 1MW/12; or (ii) only during the first 12 hours of day \(D\) at a constant level 1MW/6; or (iii) every other half hour during day \(D\) at a constant level 1MW/6. The operator “=” could thus signify customer indifference regarding the exact manner in which energy (2MWh) has been delivered at their location \(b\) during operating day \(D\) as an accumulated flow of power.

In Schweppe et al. [39, fn, p. 1153] and [40, App. F.1], a proposed \textit{Frequency Adaptive Power Energy Rescheduler (FAPER)} is carefully restricted to energy loads (”energy-type usage devices”) characterized by: (i) a need for a certain amount of energy over a period of time \(T\) in order to fulfill their functions (or purposes); and (ii) indifference as to the exact times within \(T\) during which the energy is furnished. \textit{Power loads} are characterized as the loads of devices requiring power at specific times during a period of time \(T\) in order to fulfill their functions (or purposes).

Surprisingly, however, the critical nature of the distinction between energy loads and power loads for the hourly nodal “spot pricing” approach proposed in the main chapters of [40] is not addressed by the authors.

\textbf{Statement S4:} 1 DURACELL AA 1.5v Battery = 1 DURACELL AA 1.5v Battery

\textbf{Status:} Undefined statement within both the Real Number System and the Metric System (what is a “DURACELL AA 1.5v Battery”?). Ambiguous statement for the Real World; even for a single brand and type of battery, no two distinct manufactured batteries are ever exactly the same in terms of their physical attributes. Thus, what type of “equality” is “=” meant to signify?

\textbf{Statement S5:} Let \(Q\) denote a commodity (Definition D4) with \(Q\)-amounts \(q\) measured in terms of a specific standard unit of measurement \(u\); and let the operator “=” signify “is a perfect substitute for”. Then, conditional on a given location and time, \(10u = 10u\) for each \(Q\) trader.

\textbf{Status:} Undefined statement within the Real Number System (what is \(u\)?) and the Metric System (what is a commodity?). True statement (by definition) for the Real World.

The above five statements and status assignments have three important implications, expressed below as Lemmas for later reference.

\textbf{Lemma 3.1:} The use of the operator “=” to equate amounts of two \(u\)-assets measured by the same standard unit of measurement \(u\) can result in conceptual error.

\textbf{Outline of Proof for Lemma 3.1:} By design, the standard unit of measurement \(u\) for a \(u\)-asset \(A\) typically measures the “amount” of \(A\) based on \textit{only one attribute} of \(A\), such as: weight measured in pounds (lb); energy measured in megawatt-hours (MWh); or economic value measured in U.S. dollars ($). No attempt is made to ensure that \(u\) characterizes the attributes of \(A\) in a physically or economically complete
manner. Thus, the “same” u-amounts for two physically and/or economically distinct u-assets with the same standard unit of measurement u can have substantially different physical effects when used within a physical application (e.g., as inputs to a production process) and substantially different economic effects when used within an economic application (e.g., as consumption goods for households). //

**Lemma 3.2:** The standard use of per-unit (p.u.) calculations in economics and power engineering can mask conceptual error.

**Outline of Proof for Lemma 3.2:** Conditional on a given location and time, suppose: (i) assets $A'$ and $A''$ are two u-assets that share a common standard unit of measurement $u$; (ii) $a'$ is an amount of $A'$ measured in $u$; (iii) $a''$ is an amount of $A''$ measured in $u$; (iv) $a' = a''$ measured in $u$; but (iv) the u-units for assets $A'$ and $A''$ are not equivalently exchangeable for a purpose at hand.

For example, the u-assets on each side of “=” could be: (a) equal apple amounts (measured in pounds) for two distinct apple varieties that are being offered for sale at a given location and time, as in **Statement S2**; or (b) equal energy amounts (measured in MWh) that have been grid-delivered at a designated grid location $b$ during a designated operating-period $T$ as the accumulation of two power-injection sequences with distinctly different physical attributes (e.g., different ramp-rate profiles during $T$, different capacity profiles during $T$, different delivery timing within $T$, ...), as in **Statement S3**.

Dividing the u-amounts on each side of “=” by a common “base u-value” (for example, “1 pound of apples” for the apple example in **Statement S2**, or “1MWh of energy” for the grid-delivered energy example in **Statement S3**), one is left with “per-unit” equations such as “10 = 10” for the apple example and “2 = 2” for the grid-delivered energy example that appear to be correct equations because they are true statements for the Real Number System. Any differences in the full collection of attributes characterizing the two underlying u-assets $A'$ and $A''$ that conceptually invalidate the unqualified use of an equality operator “=” in the original versions of these equations – that is, the use of “=” without the qualification “measured in $u$” – are now lost from sight. //

**Lemma 3.3:** A conceptually-meaningful real-line “quantity axis” cannot be constructed for an asset $A$ conditional on location and time unless asset $A$ is a u-asset whose u-units are equivalently exchangeable for the purpose at hand, conditional on this location and time.

**Outline of Proof for Lemma 3.3:** Suppose, first, that an asset $A$ is not a u-asset. Then there is no way to measure “amounts” of $A$ along a real-line “quantity axis” by measuring these amounts in terms of a real-valued unit-of-measurement $u$.

Suppose, next, that a u-asset $A$ is to be used as an input for a physical and/or economic process $Z$ to take place at a location $b$ at start-time $t$. However, suppose the u-units of $A$ are not equivalently exchangeable for process $Z$. Finally, suppose a process manager is tasked with the construction of a function mapping different amounts of input $A$ (measured in $u$) into corresponding physical and/or economic outcomes for process $Z$, taking as given a particular configuration of all other inputs.
As a first task-step, the manager sets about the construction of a “quantity axis” for \( A \) by identifying each real number \( r \geq 0 \) along the real-line with an amount of \( A \) of size \( r \) (measured in \( u \) units). Unfortunately for the manager, the precise selection of \( u \)-units comprising each given amount \( r \) of \( A \) can affect the resulting physical and/or economic outcomes of process \( Z \) because, by assumption, the \( u \)-units of \( A \) are not equivalently exchangeable for process \( Z \).

Thus, the physical and/or economic outcomes for process \( Z \) cannot be expressed as a conceptually well-defined function of the “amount” of input \( A \) represented as a non-negative \( r \)-value along the real line. //

The fundamental issues highlighted in Lemmas 3.1 – 3.3 regarding the representation of real-world quantity amounts as points along Cartesian coordinate axes suggest the desirability of considering alternative constructive mathematical modeling approaches permitting “holistic” representations of real-world phenomena and their interactions. See, for example, the extended discussions and illustrations of this point provided in Tesfatsion [46] and [49, Sec. 3].

Crucial ramifications of Lemmas 3.1 – 3.3 for the design and operation of grid-supported centrally-managed wholesale power markets will be explored below in Section 5.

4 Essential Economic Concepts

4.1 Overview

This section provides economic definitions that are essential for understanding the conceptually-problematic aspects of the four economic presumptions (P1)–(P4) underpinning the Two-Settlement System, the core design feature for all seven U.S. RTO/ISO-managed wholesale power markets.

For ease of exposition, several key cost definitions are expressed in a preliminary simplified form. More rigorous definitions and illustrations for these cost concepts will subsequently be provided in Section 5.3.2 as essential support for the proper conceptualization and practical assurance of supplier revenue sufficiency.

4.2 Basic Market Definitions

Definition BM1: Spot Market. A market for an asset \( A \) such that transacted amounts of \( A \), payments for these transacted amounts of \( A \), and deliveries of these transacted amounts of \( A \) all occur at the same location and time (“on the spot”).

Definition BM2: Forward Market. A market for an asset \( A \) such that transacted amounts of \( A \) and payment obligations for these transacted amounts of \( A \) are determined in advance of the delivery of these transacted amounts of \( A \).
Definition BM3: Non-Avoidable Fixed Cost (“Sunk Cost”) in Preliminary Simplified Form. Conditional on location and time, the non-avoidable fixed cost (“sunk cost”) ($) of a supplier \(i\) in the process of selecting a non-negative supply-level \(a\) (measured in \(u\)) for a \(u\)-asset \(A\) is the fixed cost \(SC_i^o\) ($) that supplier \(i\) has incurred to date that cannot be modified by any current or future decision that supplier \(i\) makes, including selection of \(a\).

Definition BM4: Avoidable Fixed Cost in Preliminary Simplified Form. Conditional on location and time, the avoidable fixed cost ($) of a supplier \(i\) in the process of selecting a non-negative supply-level \(a\) (measured in \(u\)) for a \(u\)-asset \(A\) is the fixed cost \(AFC_i^o\) ($) that supplier \(i\) incurs if and only if supplier \(i\) selects a positive supply-level \(a\).

Definition BM5: Variable Cost in Preliminary Simplified Form. Conditional on location and time, the variable cost ($) of a supplier \(i\) in the process of selecting a non-negative supply-level \(a\) (measured in \(u\)) for a \(u\)-asset \(A\) is the \(a\)-dependent cost \(VC_i(a)\) ($) that supplier \(i\) would have to incur for each selection of \(a\), where \(VC_i(0) = 0\).

Definition BM6: Total Avoidable Cost in Preliminary Simplified Form. Conditional on location and time, the total avoidable cost ($) of a supplier \(i\) in the process of selecting a non-negative supply-level \(a\) (measured in \(u\)) for a \(u\)-asset \(A\) equals \(0\) ($) if \(a = 0\) and equals the summation [\(AFC_i^o + VC_i(a)\)] of supplier \(i\)'s avoidable fixed cost ($) and variable cost ($) if supplier \(i\) selects \(a > 0\).

Definition BM7: Supplier Revenue Sufficiency in Preliminary Simplified Form. A supplier \(i\) participating in a market \(M\) for a \(u\)-asset \(A\) is revenue sufficient for \(M\) if the total revenue ($) that supplier \(i\) attains from participation in \(M\) suffices to cover the total avoidable cost ($) that supplier \(i\) incurs from participation in \(M\).

Definition BM8: Purchase Reservation Value. Conditional on location and time, a buyer \(j\)'s purchase reservation value ($) for an item \(z\) available for purchase from a supplier \(i\) is the maximum payment ($) that buyer \(j\) is willing to make to supplier \(i\) for item \(z\).

Definition BM9: Sale Reservation Value. Conditional on location and time, a supplier \(i\)'s sale reservation value ($) for an item \(z\) that supplier \(i\) is offering for sale to a buyer \(j\) is the minimum payment ($) that supplier \(i\) is willing to accept from buyer \(j\) for item \(z\).

### 4.3 Commodity Market Definitions

Definition CM1: Commodity Spot Market. A spot market for a commodity.

Definition CM2: Futures Market. A forward market for a commodity.
Definition CM3: Ordinary Demand Schedule. Conditional on location and time, a buyer j’s ordinary demand schedule for a commodity Q with standard unit of measurement u is a function that maps each non-negative Q-unit price \( \pi \) (measured in $/u) into the maximum Q-amount \( q = D_j(\pi) \) (measured in u) that buyer j is willing to procure at price \( \pi \).

Definition CM4: Benefit Function. Conditional on location and time, a buyer j’s benefit function for a commodity Q with a standard unit of measurement u is a function that maps each non-negative Q-amount \( q \) (measured in u) into the benefit \( B_j(q) \) (measured in $) that buyer j would obtain from procurement of \( q \).

Definition CM5: Marginal Benefit Function. Conditional on location and time, a buyer j’s marginal benefit function for a commodity Q with standard unit of measurement u is a function that maps each non-negative Q-amount \( q \) (measured in u) into the incremental benefit \( MB_j(q) \) (measured in $/u) that buyer j would obtain from procurement of a next Q-unit, given that buyer j has already procured \( q \).

Definition CM6: Inverse Demand Schedule. Conditional on location and time, a buyer j’s inverse demand schedule for a commodity Q with standard unit of measurement u is a function that maps each non-negative Q-amount \( q \) (measured in u) into the maximum Q-unit price \( \pi = D_j(q) \) (measured in $/u) that buyer j is willing to pay to procure a next Q-unit, given that buyer j has already procured \( q \).

Definition CM7: Ordinary Supply Schedule. Conditional on location and time, a supplier i’s ordinary supply schedule for a commodity Q with standard unit of measurement u is a function that maps each non-negative Q-unit price \( \pi \) (measured in $/u) into the maximum Q-amount \( q = S_i(\pi) \) (measured in u) that supplier i is willing to supply at price \( \pi \).

Definition CM8: Total Avoidable Cost Function. Conditional on location and time, a supplier i’s total avoidable cost function for a commodity Q with a standard unit of measurement u is a function that maps each of supplier i’s feasible non-negative Q-supply levels \( q \) (measured in u) into the total avoidable cost \( C_i(q) \) (measured in $) that supplier i would have to incur to supply \( q \).

Definition CM9: Marginal Cost Function. Conditional on location and time, a supplier i’s marginal cost function for a commodity Q with standard unit of measurement u is a function that maps each non-negative Q-amount \( q \) (measured in u) into the incremental cost \( MC_i(q) \) (measured in $/u) that supplier i would have to incur to supply a next Q-unit, given that supplier i is currently supplying \( q \).

Definition CM10: Inverse Supply Schedule. Conditional on location and time, a supplier i’s inverse supply schedule for a commodity Q with standard unit of measurement u is a function \( S_i(q) \) that maps each non-negative Q-amount \( q \) (measured in u) into the minimum non-negative Q-unit price \( \pi = S_i(q) \) (measured in $/u) that supplier i is willing to be paid for a next Q-unit, given that supplier i has already supplied \( q \).
4.4 Marginal Pricing Requires Commodities

Definition CM11: Competitive Commodity Spot Market (CCSM). Let $Q$ denote a commodity with standard unit of measurement $u$, and let $CSM(Q)$ denote a commodity spot market for $Q$. Then $CSM(Q)$ is a CCSM for $Q$ if the following five conditions hold:

1. (CCSM1) The participants in $CSM(Q)$ consist of a fixed set of $Q$-buyers $j$ and a fixed set of $Q$-suppliers $i$.

2. (CCSM2) Each buyer $j$ and supplier $i$ is a price-taker.

3. (CCSM3) Each buyer $j$ has a non-increasing ordinary demand schedule $D^o_j(\pi)$ that maps each non-negative $Q$-unit price $\pi$ (measured in $$/u$) into a non-negative $Q$-demand $q^*_j =: D^o_j(\pi)$ (measured in $u$).

4. (CCSM4) Each supplier $i$ has a non-decreasing ordinary supply schedule $S^o_i(\pi)$ that maps each non-negative $Q$-unit price $\pi$ (measured in $$/u$) into a non-negative $Q$-supply $q^*_i =: S^o_i(\pi)$ (measured in $u$).

5. (CCSM5) The equilibrium concept for $CSM(Q)$ is competitive equilibrium, defined as follows. Let

$$q =: \sum_j q_j =: \sum_j D^o_j(\pi) =: D^o(\pi)$$

(1)

denote the (ordinary) aggregate demand schedule for $Q$, and let

$$q =: \sum_i q_i =: \sum_i S^o_i(\pi) =: S^o(\pi)$$

(2)

denote the (ordinary) aggregate supply schedule for $Q$. Then a price-quantity pair $e^* = (\pi^*, q^*)$ with $q^* > 0$ is a competitive equilibrium for $CSM(Q)$ if $e^*$ is an intersection point of the aggregate demand and supply schedules $q =: D^o(\pi)$ and $q =: S^o(\pi)$ plotted in the $(\pi, q)$ plane; that is, if $e^*$ satisfies the following market-clearing rule:

**Competitive (D=S) Market Clearing Rule at $e^* = (\pi^*, q^*)$ with $q^* > 0$,**

$q^*_j =: D^o_j(\pi^*)$ for each buyer $j$, and $q^*_i =: S^o_i(\pi^*)$ for each supplier $i$:

$$q^* = D^o(\pi^*) = S^o(\pi^*)$$

(3)

The following lemma establishes an important “marginal pricing” implication of the Competitive (D=S) Market Clearing Rule (3) in definition CM11, given suitable additional regularity conditions.

---

17 See [45, Ch. 12] for a detailed presentation and illustration of the standard economic definition for a CCSM, including key related concepts such as net surplus extraction and market efficiency.

18 A participant in a spot market for a commodity $Q$ with a standard unit of measurement $u$ is said to be a price-taker if the participant behaves as if his own market transactions have no effect on the market-determined $Q$-unit price $\pi$ (measured in $$/u$).
Lemma 4.1: Let CCSM(Q) denote a competitive commodity spot market for a commodity Q. Suppose CCSM(Q) satisfies the sufficient conditions\(^{19}\) in Appendix A.6 that are shown to ensure the following conditions (a) and (b) hold for CCSM(Q):

(a) Each buyer j participating in CCSM(Q) has a non-increasing inverse demand schedule \(\pi = D_j(q_j)\) that can be inverted to give a non-increasing ordinary demand schedule \(q_j^* = D_j^*(\pi)\) for buyer j, and vice versa, where \(D_j(q_j)\) coincides with buyer j’s marginal benefit function, i.e., \(D_j(q_j) = MB_j(q_j)\).

(b) Each supplier i participating in CCSM(Q) has a non-decreasing inverse supply schedule \(\pi = S_i(q_i)\) that can be inverted to give a non-decreasing ordinary supply schedule \(q_i^* = S_i^*(\pi)\) for supplier i, and vice versa, where \(S_i(q_i)\) coincides with supplier i’s marginal cost function, i.e., \(S_i(q_i) = MC_i(q_i)\).

Let \(e^* = (\pi^*, q^*)\) with \(q^* > 0\) denote a competitive equilibrium for CCSM(Q). Then the following marginal-pricing rule holds for \(e^*\):

**Competitive (MB=MC) Spot-Pricing Rule** at \(e^* = (\pi^*, q^*)\) with \(q^* > 0\): For each buyer j and supplier i such that \(q_j^* > 0\) and \(q_i^* > 0\),

\[
\pi^* = MB_j(q_j^*) = MC_i(q_i^*)
\]

**Proof for Lemma 4.1:** To establish (4) holds at \(e^* = (\pi^*, q^*)\) with \(q^* > 0\) if \(e^*\) is a competitive equilibrium for CCSM(Q) and conditions (a) and (b) are satisfied, invert the expressions \(q_j^* = D_j^*(\pi^*)\) and \(q_i^* = S_i^*(\pi^*)\) appearing in the Competitive (D=S) Market Clearing Rule (3) by appropriate inverse demand and supply operations.\(^{19}\)

A CCSM(Q) for some commodity Q that satisfies the conditions in Appendix A.6 ensuring conditions (a) and (b) in Lemma 4.1 both hold, hence the marginal pricing rule (4) holds at any competitive equilibrium \(e^* = (\pi^*, q^*)\) for CCSM(Q), will be called a Marginal-Pricing CCSM, or MP-CCSM for short. CCSMs and MP-CCSMs have a variety of attractive efficiency and optimality properties. Several of these properties are stated below as lemmas for later reference.

**Lemma 4.2:** All fixed cost for each supplier i participating in a CCSM is sunk cost.

**Proof for Lemma 4.2:** By definition, a CCSM is a spot market that takes place at a given location and time for a given set of participants whose demand and supply schedules are automatically submitted to the CCSM and instantly cleared (or not cleared) by a market process to determine equilibrium outcomes. Thus, any fixed cost for a participating supplier at the time of this participation is non-avoidable fixed cost ("sunk cost").\(^{19}\)

**Lemma 4.3:** An MP-CCSM is a uniform-price market.

**Proof for Lemma 4.3:** Let MP-CCSM(Q) denote an MP-CCSM for some commodity Q. By definition, MP-CCSM(Q) is a uniform-price market for Q in the following sense. At any given competitive equilibrium \(e^* = (\pi^*, q^*)\), the same Q-unit

\(^{19}\) These sufficient conditions impose non-negativity, monotonicity, differentiability, and curvature (concavity and convexity) conditions on the demand and supply schedules for CCSM(Q).
price \( \pi^* \) ($/u) is: (a) paid by each buyer \( j \) for each \( Q \)-unit that buyer \( j \) purchases at \( e^* \); and (b) received by each supplier \( i \) for each \( Q \)-unit that supplier \( i \) sells at \( e^* \).

**Lemma 4.4:** Supplier revenue sufficiency holds for an MP-CCSM. That is, the total revenue attained by a supplier \( i \) from participation in this MP-CCSM is sufficient to cover the total avoidable cost that supplier \( i \) incurs from this participation.

**Proof for Lemma 4.4:** By Lemma 4.2, all fixed cost for supplier \( i \) is sunk cost, i.e., non-avoidable fixed cost. Thus, by definitions BM3–BM6 in Section 4.2, the total avoidable cost incurred by supplier \( i \) from participation in MP-CCSM(\( Q \)) coincides with the total marginal cost incurred by supplier \( i \) from this participation.

Let \( e^* = (\pi^*, q^*) \) denote a competitive equilibrium for MP-CCSM(\( Q \)), and let \( q^*_i \) denote the amount of \( Q \) provided by \( i \) at \( e^* \). If \( q^*_i = 0 \), supplier \( i \) incurs zero marginal cost at \( e^* \) and attains zero revenue; hence, supplier \( i \) is revenue sufficient. Suppose \( q^*_i > 0 \). Then, by definition of an MP-CCSM, it follows from (4) that supplier \( i \)'s marginal cost \( MC(q^*_i) \) at \( q^*_i \) coincides with the competitive equilibrium market price \( \pi^* \) at \( e^* \). Moreover, from condition (b) in Lemma 4.1, supplier \( i \)'s marginal cost function coincides with supplier \( i \)'s non-decreasing inverse supply schedule from 0 to \( q^*_i \). It follows that the total revenue \( \pi^* \times q^* \) ($\) attained by supplier \( i \) at \( e^* \) suffices to cover supplier \( i \)'s total incurred marginal cost ($\) at \( e^* \), calculated as the total (integrated) area under supplier \( i \)'s marginal cost function from 0 to \( q^*_i \). Thus, supplier \( i \) is revenue sufficient.

**Lemma 4.5:** Let MP-CCSM(\( Q \)) denote an MP-CCSM for a commodity \( Q \), and let CSM(\( Q \)) denote the underlying CSM for MP-CCSM(\( Q \)). Then market efficiency holds for CSM(\( Q \)) in the following sense: At \( e^* \) the Total Net Surplus\(^{20} \) extracted from CSM(\( Q \)) is maximized; i.e., there is no wastage of opportunity to extract additional net surplus from CSM(\( Q \)).

**Proof for Lemma 4.5:** See Tesfatsion [45, Ch. 12].

Lemma 4.5 implies that the defining conditions for an MP-CCSM are sufficient to ensure market efficiency holds for the underlying commodity spot market CSM at any competitive equilibrium \( e^* \). However, these defining conditions are not necessary for the market efficiency of this underlying CSM.

For example, it can be shown that market efficiency holds for the underlying CSM(\( Q \)) for an MP-CCSM(\( Q \)) for a commodity \( Q \) if the Competitive (MB=MC) Spot-Pricing Rule (4) is replaced by any price-rule PR that satisfies each of the following three conditions:

\(^{20}\) In economics, the Total Net Surplus attained at any competitive equilibrium \( e^* = (\pi^*, q^*) \) for MP-CCSM(\( Q \)) is defined to be the sum of the Total Net Buyer Surplus and Total Net Supplier Surplus attained at \( e^* \). The Total Net Buyer Surplus attained at \( e^* \) is defined to be the difference between the maximum amount that buyers would have been willing to pay for procurement of \( q^* \) and the amount that buyers actually pay for the procurement of \( q^* \) at \( e^* \). The Total Net Supplier Surplus attained at \( e^* \) is defined to be the difference between the payment that suppliers actually receive for the sale of \( q^* \) at \( e^* \) and the minimum payment that suppliers would have been willing to receive for the sale of \( q^* \). See Tesfatsion [45, Ch. 12] for careful discussion and illustrations of these and related market concepts.
Leigh Tesfatsion

**PRC(a):** All $Q$-units traded in CSM($Q$) under price-rule PR also trade at some competitive equilibrium $e' = (\pi', q')$ for CSM($Q$);

**PRC(b):** All $Q$-units failing to trade in CSM($Q$) under price-rule PR also fail to trade at the competitive equilibrium $e' = (\pi', q')$ in **PRC(a)**;

**PRC(c):** Under price-rule PR, the price paid by a buyer $j$ to purchase a unit of $Q$ is the same as the price received by the supplier $i$ who supplies this unit of $Q$.\(^{21}\)

An example of a price-rule PR for CSM($Q$) that satisfies **PRC(a)**–**PRC(c)**, distinct from the Competitive (MB=MC) Spot-Pricing Rule (4), is the $k$-discriminatory-price rule ($k \in [0, 1]$) defined as follows: For any matched buyer-supplier pair for which the buyer’s purchase reservation value $\pi^b$ and the supplier’s sale reservation value $\pi^s$ satisfy $\pi^b \geq \pi^s$, set the strike price for this pair at the weighted-average level $\pi^k = k\pi^b + [1-k]\pi^s$ lying between their reservation values.\(^{22}\) Thus, the division between buyer and supplier of the net surplus increment $[\pi^b - \pi^s]$ resulting from their trade is determined by $k$; however, the total amount of this net surplus increment is not affected by $k$.

One intuitive argument commonly given in favor of using the Competitive (MB=MC) Spot-Pricing Rule (4) and against the use of a $k$-discriminatory-price rule for a CSM($Q$) is that competitive spot-pricing provides more incentive to suppliers (buyers) of a commodity $Q$ to use the most efficient available technology for extraction of net supplier (buyer) surplus. For example, suppose a supplier switches to a new technology that strictly lowers his marginal cost of production (hence his sale reservation value $\pi^s$) for each unit of $Q$ he sells. Competitive spot-pricing permits this supplier to keep all of his resulting increased net supplier surplus; the $k$-discriminatory-price rule with $k < 1$ does not. An analogous argument holds for a buyer able to switch to a new technology that permits him to increase his marginal benefit (hence his purchase reservation value $\pi^b$) for each unit of $Q$ he buys.

However, under either price-rule, a welfare-maximizing supplier or buyer will not switch to a more efficient technology unless the cost of this switch is less than the expected net surplus gain from future market transactions. Yet low-cost or costless technology switching could strongly deter engagement in the costly research and development (R&D) efforts needed to develop more efficient technologies. Clearly, a dynamic joint analysis of market and R&D processes is needed to examine with care the long-run efficiency implications of alternative market-price rules.

---

\(^{21}\) Condition **PRC(c)** holds for DAM/RTM SCED optimizations in the absence of LMP separation. However, since power injected or withdrawn at any one grid location rapidly affects power flow on all directly or indirectly connected transmission lines, power injections and withdrawals do not in fact constitute a collection of bilateral buyer-supplier trades as presumed in the statement of condition **PRC(c)**.

\(^{22}\) The two extremes of the $k$-discriminatory-price rule are of special interest: The 0-discriminatory-price rule awards all generated net surplus to buyers because the price received by suppliers is their minimum acceptable sale price $\pi^s$. Conversely, the 1-discriminatory-price rule awards all generated net surplus to suppliers because the price charged to buyers is their maximum acceptable purchase price $\pi^b$. 
A second intuitive argument commonly given in favor of using the Competitive (MB=MC) Spot-Pricing Rule (4) and against the use of a k-discriminatory-price rule for a CSM($Q$) concerns incentives for truthful revelation. Under a k-discriminatory-price rule with $k < 1$, a self-interested supplier $i$ would have a strategic incentive to report a higher-than-true sale reservation value for his strictly inframarginal$^{23}$ $Q$-units in order to receive a higher discriminatory price for these units, thus increasing his true net revenue (i.e., his revenue minus his true variable cost) from the sale of these units.

In summary, as seen in the presentation and discussion of Lemmas 4.1–4.5, competitive commodity spot markets (CCSMs) and MP-CCSMs satisfying the Competitive (MB=MC) Spot-Pricing Rule (4) have attractive efficiency and optimality properties. However, the following caution must be kept carefully in mind:

**Key Caution:** The defining properties of a CCSM, as well as all derived properties established for CCSMs and MP-CCSMs in Lemmas 4.1–4.5, require the transacted asset $A$ to be a commodity.

Specifically, in accordance with definition D4 in Section 3.2, the transacted asset $A$ must have a standard unit of measurement $u$ such that, conditional on location and time, all $A$-traders consider all available units $u$ of asset $A$ to be perfect substitutes (economically equivalent). Otherwise, as established in Lemma 3.3 in Section 3.3, a real-line “quantity axis” cannot be constructed in a conceptually coherent manner for the ordinary demand and supply functions appearing in the defining properties for a CCSM.

Important implications of this key caution specifically for the existence of demand and supply functions, the essential underpinning for the conception of a competitive equilibrium, are stated below in lemma form for later reference.

**Lemma 4.6:** Conditional on location and time, any buyer $j$ of a commodity $Q$ that receives one additional $Q$-unit $u$ is completely indifferent with regard to which precise $Q$-unit he receives because, by definition of a commodity, the incremental economic benefit that buyer $j$ gains from the receipt of an incremental “next” $Q$-unit is the same for all available $Q$-units. This indifference is a necessary condition for buyer $j$ to have a conceptually well-defined demand schedule for $Q$ at this given location and time, either inverse or ordinary.

---

$^{23}$ The units $u$ of a commodity $Q$ that are purchased by buyers (equivalently, sold by suppliers) in a competitive equilibrium $e^* = (\pi^*, q^*)$ for a CSM($Q$) are said to be **inframarginal units**. An inframarginal unit $u$ of a commodity $Q$ is said to be a **marginal unit** if this unit is the last $Q$-unit purchased and sold at $e^*$, assuming all $Q$-units purchased at $e^*$ are purchased in the descending order of their purchase reservation values (i.e., from highest to lowest) calculated across all participating buyers conditional on all of their past purchases, and in the ascending order of their sale reservation values (i.e., from lowest to highest) calculated across all participating sellers conditional on all of their past sales. All other $Q$-units purchased and sold at $e^*$ are said to be **strictly inframarginal units**. Finally, all $Q$-units demanded for purchase (or offered for sale) at $e^*$ that are not in fact purchased (or sold) at $e^*$ are called **extramarginal units**. See Fig. 2.
Lemma 4.7  Conditional on location and time, any supplier $i$ of a commodity $Q$ that supplies one additional $Q$-unit is completely indifferent with regard to which precise $Q$-unit he supplies because, by definition of a commodity, the incremental economic cost that supplier $i$ incurs from the supply of an incremental “next” $Q$-unit is the same for all available $Q$-units. This indifference is a necessary condition for a supplier $i$ to have a conceptually well-defined supply schedule for $Q$ at this given location and time, either inverse or ordinary.

To understand the critical implications of Lemma 4.6 and Lemma 4.7 in more concrete terms, consider the following situation. At a given location and time, an experimental economist plans to use a sealed bag containing a mixture of HoneyCrisp Apples and Dole Mandarin Oranges to construct an ordinary demand schedule for fruit for a human subject called “buyer $j$. The standard unit of measurement $u$ for fruit is taken to be a piece of fruit; hence, fruit-quantities $q$ are measured by the number of included fruit pieces $u$, and fruit-unit prices $\pi$ are measured by dollars per fruit-piece ($/u$).

In accordance with condition CCSM3, a defining condition appearing in the definition CM11 for an MP-CCSM given in Section 4.4, the experimenter hands buyer $j$ an ordered list of successively higher fruit-unit prices $\pi$ and asks buyer $j$ to report the maximum fruit-quantity $q := D^o(\pi)$ that he would be willing to buy at each listed fruit-unit price $\pi$. At the end of the experiment, one of the listed fruit-unit prices, $\pi^*$, will be randomly announced, the bag of fruit will be unsealed, and buyer $j$ will be required to pay $\pi^* \times q^*$ ($\$) for a fruit-quantity $q^* := D^o(\pi^*)$ that the experimenter draws randomly from the unsealed bag.

Unfortunately for the experimenter, suppose buyer $j$ does not consider a Honey-Crisp Apple to be a perfect substitute for a Dole Mandarin Orange; that is, suppose the specific apple-versus-orange attribute of a fruit-piece matters to buyer $j$? In this case, the economic value that buyer $j$ attains from any procured fruit-quantity $q$ will depend on the specific apple-orange composition of $q$.

Consequently, the maximum fruit-quantity $q$ that buyer $j$ is willing to purchase at each listed fruit-unit price $\pi$ will depend on how buyer $j$ resolves his uncertainty regarding two related aspects of the experiment. First, what is the apple-orange composition of fruit-pieces in the sealed bag? Second, given this composition, what will be the likely apple-orange composition of the fruit-quantity $q^*$ that is randomly drawn from the unsealed bag if buyer $j$ reports that $q^* := D^o_j(\pi^*)$ is the maximum fruit-quantity he is willing to purchase at the announced price $\pi^*$?

The bottom line is that an ordinary demand schedule $q := D^o(\pi)$, as defined in CM3 in Section 4.3, is not conceptually well-defined for fruit for a fruit-buyer $j$ at a given location and time unless buyer $j$ considers all pieces of fruit available for purchase at this location and time to be perfect substitutes. Similar arguments can be used to demonstrate that an ordinary supply schedule $q := S^o(\pi)$, as defined in CM7 in Section 4.3, is not well-defined for fruit for a fruit-supplier $i$ at a given location and time unless supplier $i$ considers all pieces of fruit available for supply at this location and time to be perfect substitutes.
What about inverse demand and supply schedules for fruit, defined in accordance with the standard economic definitions CM6 and CM10 in Section 4.3? Here the import of Lemma 4.6 and Lemma 4.7 is even clearer.

Suppose apples and oranges are not perfect substitutes for a fruit-buyer \( j \). How can buyer \( j \) express his maximum acceptable purchase price \( \pi = D_j(q) \) (i.e., his purchase reservation value BM8) for a “next” piece of fruit, given he has already procured a fruit-quantity \( q \), without knowing: (i) which specific fruit piece, apple or orange, is to be his “next” procured fruit piece; and (ii) what is the specific apple-orange composition of his already-procured fruit-quantity \( q \)?

Suppose apples and oranges are not perfect substitutes for a fruit-supplier \( i \). How can supplier \( i \) express his minimum acceptable sale price \( \pi = S_i(q) \) (i.e., his sale reservation value BM9) for a “next” piece of fruit, given he has already supplied a fruit-quantity \( q \), without knowing: (i) which specific fruit piece, apple or orange, is to be his “next” supplied fruit piece; and (ii) what is the specific apple-orange composition of his already-supplied fruit-quantity \( q \)?

5 Legacy Core Design: Fundamental Conceptual Issues

5.1 Overview

The Two-Settlement System reviewed in Section 2 constitutes the core design feature for all seven current U.S. RTO/ISO-managed wholesale power markets. As noted in Section 1, this core design feature reflects four conceptually-problematic economic presumptions (P1)–(P4). These presumptions, re-stated below, can now be given more precise meaning using the essential measurement and economic concepts reviewed in Sections 3 and 4.

Problematic Presumption (P1):

The basic transacted product for grid-supported centrally-managed wholesale power markets is grid-delivered energy (MWh), i.e., accumulations of flows of power (MW) at designated grid locations \( b \) during designated operating periods \( T \) with duration measured in hours (h).

Problematic Presumption (P2):

For careful analysis of supplier revenue sufficiency in such markets, it suffices to partition total supplier cost into a “variable” component dependent on the quantity supplied and a “fixed” component independent of the quantity supplied.

Problematic Presumption (P3):

Grid-delivered energy conditional on delivery location \( b \) and delivery period \( T \) is a commodity, i.e., its units (MWh) are perfect substitutes. Thus, these units can (and should) be transacted in a spot market \( M(b, T) \) at a uniform per-unit locational marginal price \( LMP(b, T) \) ($/MWh) determined in accordance with the standard competitive (marginal benefit = marginal cost) spot-pricing rule.

Problematic Presumption (P4):

The total supplier revenue attained in the spot markets in (P3) will suffice to cover total supplier cost.
The counterclaims (CC1)–(CC4) to presumptions (P1)–(P4) that were stated in brief preliminary form in Section 1 are carefully highlighted and analyzed in the remainder of this section.

### 5.2 Reserve, Not Energy, is the Basic Product

#### 5.2.1 Critique of Presumption (P1): Overview

Presumption (P1) asserts that grid-delivered energy (MWh) is the basic product transacted in grid-supported centrally managed wholesale power markets. This presumption is consistent with current practice in U.S. RTO/ISO-managed wholesale power markets; the emphasis of these markets is on grid-delivered energy bought and sold in short-run (day-ahead and intra-day) markets at designated grid locations.

Counterclaim (CC1) asserts to the contrary that suppliers in such markets provide two basic types of product: namely, physically-covered insurance (reserve), and real-time power delivery. Here “reserve” is conceived to be the guaranteed availability of dispatchable power-path production capabilities for possible central dispatch during future operating periods to protect against volumetric grid risk.

Counterclaim (CC1) is the primary message stressed throughout this study. This section commences support for (CC1) by carefully considering the following fundamental physical and valuation concerns regarding the validity of presumption (P1):

- **Physical Reliability Concerns:** Energy transactions in U.S. RTO/ISO-managed wholesale power markets must be supported by the physical operations of underlying transmission grids. The necessary requirements for the reliable operation of these grids cannot be expressed solely in terms of grid-delivered energy (MWh). Rather, these reliability requirements constitute joint restrictions on “power-paths,” i.e., on flows of power (MW) at single designated grid locations during designated operating periods.

- **Benefit and Cost Valuation Concerns:** How power (MW) is injected at grid locations during successive operating periods can matter greatly to power producers, power customers, and RTOs/ISOs. The total amount of energy resulting from a grid-delivered flow of power at a grid location during an operating period is only one of many possible valued attributes this flow of power could possess.

Additional support for counterclaim (CC1) is then provided in subsequent subsections as each of the remaining three counterclaims (CC2), (CC3), and (CC4) is supported in turn.
5.2.2 Physical Reliability Concerns

In order for a transmission grid to operate reliably over time, the grid must be in continual net-load balance. Roughly stated, this means that the injection of power into the grid must balance the withdrawal and inadvertent loss of power from the grid at each instant of time.

More carefully stated, continual net-load balance is a reliability requirement for a transmission grid because any transmission grid in stable operation must satisfy Kirchhoff’s Current Law. Applied to a general electrical network at a given point in time, this law asserts the following: The algebraic sum of all currents entering a network node \( n \) must equal the algebraic sum of all currents exiting node \( n \), where currents are measured in Amperes (A).

U.S. RTO/ISO-managed wholesale power markets operate over high-voltage Alternating Current (AC) transmission grids. For illustration, consider the analytical modeling developed in [45, Ch. 7 & Sec. 9.2] for an RTO/ISO-managed wholesale power market \( M(T) \) for a future operating period \( T = [t^*, t^e] \) with finite duration. The relative timing of \( M(T) \) and \( T \) are depicted in Fig. 5.

![Fig. 5 Time-line for a grid-supported RTO/ISO-managed wholesale power market \( M(T) \) conducted for a future operating period \( T \).](image)

The market \( M(T) \) operates over a high-voltage AC transmission grid with buses \( b \) in a bus-set \( B \). The participants in \( M(T) \) are as follows: For each bus \( b \in B \),

- a collection \( M(b) \) of dispatchable generation-units \( m \), each with a unique electrical point of connection to the transmission grid at bus \( b \);
- a collection \( LSE(b) \) of LSEs \( j \), each of whom manages power-usage for a distinct collection \( C_j(b) \) of customers with unique electrical points-of-connection to the transmission grid at bus \( b \);
- a collection \( NG(b) \) of non-dispatchable generation-units \( n \), each with a unique electrical point-of-connection to the transmission grid at bus \( b \).

The AC power-flow operations of the transmission grid for \( M(T) \) are analytically modeled in [45, Ch. 7 & Sec. 9.2] using a standard Direct-Current (DC) power-flow approximation.\(^{24}\) This approximation assumes a loss-less grid with a constant voltage magnitude \( V_0 \). In this case Kirchhoff’s Current Law, expressed in terms of current \( I \) measured in Amperes (A), can equivalently be expressed in terms of (active) power \( p \) measured in megawatts (MW) using \( p = V_0 \cdot I \).

\(^{24}\) See [41, Sec. 3.1] for a careful discussion of the standard assumptions used to derive a DC power-flow approximation for the AC power-flow of a high-voltage AC transmission grid.
The net-load balance constraints for this approximated transmission grid for a given operating period $T$ can then be summarized as follows, where the qualifier fixed is used as a short-hand expression for non-dispatched must-service. For each bus $b \in B$ and time $t \in T$, the total dispatched power injection at bus $b$ by the dispatchable generation-units $m \in M(b)$, plus the total net line-power inflow at bus $b$ from buses $b'$ in $B$ with $b' \neq b$, must equal the total forecasted net load at bus $b$, calculated as the total dispatched customer load at bus $b$ for customers of the LSEs $j \in LSE(b)$ plus the total forecasted fixed customer load at bus $b$ for customers of the LSEs $j \in LSE(b)$ minus the total forecasted fixed power injection at bus $b$ by the non-dispatchable generators $n \in NG(b)$.

For the purposes of this section, however, it is important to express these net-load balance constraints in their explicit mathematical forms [45, Ch. 6 & Sec. 9.2]:

**Net-load balance constraints:** For each bus $b \in B$ and time $t \in T$,

\[ G_b^{\text{dis}}(t) + \text{NLPI}_b(t) = L_b^{\text{dis}}(t) + \hat{NL}_b(t) \]

where:

- $G_b^{\text{dis}}(t) = \sum_{m \in M(b)} p_m^{\text{dis}}(t)$ (total dispatched power injection);
- $\text{NLPI}_b(t) = \left[ \sum_{\ell \in LE(b)} w_{\ell}(t) - \sum_{\ell \in LO(b)} w_{\ell}(t) \right]$ (total net line-power inflow);
- $L_b^{\text{dis}}(t) = \sum_{j \in LSE(b)} p_j^{\text{dis}}(t)$ (total dispatched customer load);
- $\hat{NL}_b(t) = \left[ \hat{L}_b(t) - \hat{G}_b(t) \right]$ (total forecasted net fixed load);
- $\hat{L}_b(t) = \sum_{j \in LSE(b)} \hat{p}_j(t)$ (total forecasted fixed customer load);
- $\hat{G}_b(t) = \sum_{n \in NG(b)} \hat{p}_n(t)$ (total forecasted fixed power injection).

The important take-away from the net-load balance constraints (5) is that these are not static demand=supply restrictions on transacted amounts of energy (MWh) to be delivered at each bus $b \in B$ for operating period $T$. Rather, they are complicated joint restrictions on power-paths, defined as follows:

**Definition:** A power-path $p_b(T) = (p_b(t) \mid t \in T)$ is a sequence of power injections and/or withdrawals $p_b(t)$ (MW) that take place at a single designated grid-location $b$ during a designated operating period $T$. See Fig. 6 for an illustrative example.

Specifically, constraints (5) impose joint restrictions on the following four types of power-paths at each bus $b \in B$ during $T$:
• the dispatched power-path \( p_{m}^{\text{dis}}(T) = (p_{m}^{\text{dis}}(t) \mid t \in T) \) at b during T for each dispatchable generation-unit \( m \in M(b) \);

• the dispatched power-path \( p_{j}^{\text{dis}}(T) = (p_{j}^{\text{dis}}(t) \mid t \in T) \) at b during T for the managed customers of each LSE \( j \in \text{LSE}(b) \);

• the forecasted fixed power-path \( \hat{p}_{j}^{f}(T) = (\hat{p}_{j}^{f}(t) \mid t \in T) \) for fixed power withdrawals at b during T by the managed customers of each LSE \( j \in \text{LSE}(b) \);

• the forecasted fixed power-path \( \hat{p}_{n}^{f}(T) = (\hat{p}_{n}^{f}(t) \mid t \in T) \) for fixed power injections at b during T by each non-dispatchable generator \( n \in \text{NG}(b) \).

As demonstrated in [45, Ch. 7], given mild regularity conditions and a finite-duration operating period \( T = [t^s, t^e] \), it is possible to approximate a power-path \( p_b(T) \) as closely as desired by a step-function.\(^{25}\) For example, this step-function approximation for \( p_b(T) \) could consist of a discretized sequence

\[
p_b(K(T)) =: (p_b(k) \mid k \in K(T))
\]

where \( K(T) \) consists of a collection of successive half-open sub-periods

\[
k_n =: [k^s_n, k^e_n), \quad n = 1, ..., N(T)
\]

constituting a suitably-refined partition of operating period T, and

\[
p_b(k_n) =: p_b(k^s_n) \quad (\text{MW})
\]

\(^{25}\) Step functions are universal approximators for the class of all continuous real-valued functions \( f:[a, b] \rightarrow \mathbb{R} \) defined over compact intervals \([a, b]\) of the real line. For example, given any such function \( f \), and any \( \varepsilon > 0 \), it is straightforward to establish the existence of a step function \( f_{\varepsilon}:[a, b] \rightarrow \mathbb{R} \) with finitely many time-steps for which the maximum absolute approximation error \( |f(x) - f_{\varepsilon}(x)| \) over \( x \in [a, b] \) is less than \( \varepsilon \). This assertion follows immediately from the Heine-Cantor Theorem, which establishes that any continuous function \( f:X \rightarrow Y \) between metric spaces \( X \) and \( Y \), with \( X \) compact, is uniformly continuous.
Leigh Tesfatsion

denotes the power-level of power-path \( p_b(T) \) evaluated at the start-time \( k_n^s \) of the sub-period \( k_n \in K(T) \).

However, as carefully discussed and demonstrated in [45, Chs. 7,16] and [29], the attributes of a power-path are highly correlated (jointly produced) outputs. For example, the ramp-rate (MW/min) profile of a power-path for an operating period \( T \) is highly correlated with the capacity (MW) profile of this power path for \( T \). Consequently, step-function approximations for power-paths need to be carefully tailored for the purpose at hand to ensure correlations among valued attributes are properly taken into account. In addition, it could be advantageous, or even necessary, to use \( T \)-partitions (7) with different sub-period durations \( \Delta k_n \) for different types of market participants in order to capture adequately the specific static and dynamic attributes of their power-paths. For a discussion related to this point, see [21, Sec. 3.1.1].

Moreover, market optimization formulations making use of step-function approximations (6) for power-paths should be expressed in run-time variables to ensure system constraints are imposed with sufficient accuracy. For example, the ramping and capacity constraints imposed at the start-time \( k_n^s \) of each successive sub-period \( k_n \) in \( K(T) \) should be expressed in terms of run-time min/max limits on ramp-rates and power-capacities for sub-periods \( k_{n'} \) with \( n' \geq n \) to adequately capture the run-time endogenously-determined dependence of ramping capability on current operating state as well as on exogenously-rated capacity limits.

See [45, Ch. 7] for a complete analytical modeling of an RTO/ISO-managed SCED optimization in Mixed-Integer Linear Programming (MILP) form for a market \( M(T) \) that makes use of step-function power-path approximations (6) with system constraints expressed entirely in run-time variables.

5.2.3 Benefit and Cost Valuation Concerns

A more serious conceptual concern regarding presumption (P1) is that it prevents a comprehensive high-fidelity valuation of benefits and costs.

Consider once again the analytical model developed in [45, Ch. 7 & Sec. 9.2] for an RTO/ISO-managed wholesale power market \( M(T) \) that was previously used to illustrate physical reliability concerns regarding (P1). Recall that \( M(T) \) operates over a high-voltage AC transmission grid for a future operating period \( T = [t_s, t_e) \) with finite duration.

The participants in \( M(T) \) include a collection \( M \) of RTO/ISO-dispatchable generation units \( m \), a collection \( \text{LSE} \) of load-serving entities \( j \), and a collection \( \text{NG} \) of non-dispatchable generators \( n \). For each bus \( b \) in the grid bus-set \( \mathbb{B} \), \( M(b) \) denotes the collection of all generation units \( m(b) \) in \( M \) that are electrically connected to bus \( b \), and \( \text{LSE}(b) \) denotes the collection of all load-serving entities \( j \) in \( \text{LSE} \) that service customers electrically connected to bus \( b \). Finally, the collection of all customers \( c_j(b) \) managed by a load-serving entity \( j \in \text{LSE}(b) \) is denoted by \( C_j(b) \).

Let \( p_b(T) = (p_b(t) \mid t \in T) \) denote a power-path for operating period \( T \) that consists of a sequence of power injections and/or withdrawals \( p_b(t) \) (MW) at bus \( b \) during times \( t \in T \). Suppose this power-path has a continuous extension over \( \tilde{T} = \)
$[t^*, t^*]$, the compact closure of $T$. Then $p_b(T)$ can be approximated arbitrarily closely over $T$ by a suitably-constructed step function $p_b(K(T))$ taking form (6). Plotted in a time-MW plane, this approximating step-function $p_b(K(T))$ consists of a finite sequence of *energy-blocks* (MWh):

$$E_{b,n} =: p_b(k^e_n) \times [k^e_n - k^v_n] \text{ for } n = 1, \ldots, N(T). \quad (9)$$

Consequently, if the power-path $p_b(T)$ is scheduled for dispatch at $b$ during $T$, its anticipated total energy delivery $E(b,T)$ (MWh) at $b$ during $T$ can be approximated by adding up the energy-blocks (9) for its step-function approximation. However, the ability to approximate $E(b,T)$ for an RTO/ISO-dispatched power-path $p_b(T)$ in no way guarantees that the *actual* cost and/or benefit value assigned to $p_b(T)$ by a generation unit $m(b)$, a customer $c_j(b)$, or the RTO/ISO can be expressed solely as a function of $E(b,T)$.

A generation unit $m(b)$ dispatched by the RTO/ISO to deliver $p_b(T)$ at $b$ during $T$ would presumably care about the *dynamic* attributes of this power-path as well as its static attributes. For example, $m(b)$ might be concerned about equipment depreciation cost incurred *during* $T$ from *ramping* wear and tear, and the fuel costs incurred *during* $T$ for power production.

Moreover, what each customer $c_j(b)$ would presumably value in advance of $T$ is a guaranteed ability to determine power withdrawals at $b$ during $T$ in a flexible *just-in-time* manner to operate personally-owned electrical devices for locally-determined purposes. The value that $c_j(b)$ would attach in advance of $T$ to *any one* pre-specified power-path $p_b(T)$ would presumably be low, simply because of its inflexibility.

Finally, what the RTO/ISO would presumably value in advance of $T$ is having the reliable availability of a *suitably diverse collection* of dispatchable power-paths for $T$ enabling the RTO/ISO to balance net-power withdrawals at $b$ and other grid locations during $T$ by *just-in-time* dispatched net-power injections. The value that the RTO/ISO would attach in advance of $T$ to the availability of *any one* specified power-path $p_b(T)$ would presumably be low, simply due to its inflexibility.

The key implication of these observations is that benefits and costs of producers, customers, and the RTO/ISO itself in RTO/ISO-managed wholesale power markets cannot properly be assessed solely in terms of grid-delivered energy amounts.

### 5.3 Supplier Cost Analysis Requires a 3-Part Partition

#### 5.3.1 Critique of Presumption (P2): Overview

Presumption (P2) regarding the analysis of total supplier cost has two problematic aspects. First, it reflects an overly-simplest 2-part partition of total supplier cost into fixed and variable components. Second, its focus on grid-delivered energy amounts as the sole determinant of supplier variable cost is based on the conceptually-problematic presumption (P1). These two problematic aspects of presumption (P2) will next be separately addressed.
5.3.2 Presumption (P2) Reflects an Overly Simplistic Cost Partitioning

For reasons carefully articulated in seminal work by Baumol et al. [3], the traditional economic partitioning of total cost into two components, “fixed” and “variable,” is conceptually incomplete and empirically problematic. Total cost at a given time \( t \) should instead be partitioned into three economically-distinct components:

\[
\text{Total Cost} = \text{Sunk Cost} + \text{Avoidable Fixed Cost} + \text{Variable Cost} \tag{10}
\]

As will be stressed in subsequent sections of this study, the need for the three-part partition (10) is particularly critical for the conceptually-coherent design and operation of grid-supported centrally-managed wholesale power markets.

Consider a Decision-Maker DM at a current time \( t \) who must decide now whether or not to commit to undertaking an action of type A at a future time \( t + \Delta t \). DM’s total cost at time \( t \) can be partitioned into three components – sunk cost, avoidable fixed cost, and variable cost – as follows:

\[
\text{Sunk Cost} =: \text{Non-Avoidable Fixed Cost} \tag{11}
\]

\[
=: \text{Cost } SC^0 \text{ that:}
\]

(i) DM incurs \textbf{whether or not} DM commits at time \( t \) to undertaking a type-A action at time \( t + \Delta t \);

(ii) \textbf{does not} depend on the specific type-A action DM undertakes, should DM choose to commit.

\[
\text{Avoidable Fixed Cost} =: \text{Cost } AFC^0 \text{ that:} \tag{12}
\]

(i) DM incurs \textbf{if and only if} DM commits at time \( t \) to undertaking a type-A action at time \( t + \Delta t \);

(ii) \textbf{does not} depend on the specific type-A action DM undertakes, should DM choose to commit.

\[
\text{Variable Cost} =: \text{Cost } VC(a) \text{ that:} \tag{13}
\]

(i) DM incurs \textbf{if and only if} DM commits at time \( t \) to undertaking a type-A action at time \( t + \Delta t \);

(ii) \textbf{does} depend on specific form \( a \) of type-A action DM undertakes, should DM choose to commit.

\[
\text{Fixed Cost} =: [ \text{Sunk Cost} + \text{Avoidable Fixed Cost} ] \tag{14}
\]

\[
\text{Avoidable Cost} =: [ \text{Avoidable Fixed Cost} + \text{Variable Cost} ] \tag{15}
\]

For illustration, consider a currently off-line dispatchable thermal generator \( m \) at the start-time \( t \) for a day-D U.S. RTO/ISO-managed DAM that the RTO/ISO is conducting to prepare for day-D+1 system operations.
An example of a *sunk cost* for \( m \) at time \( t \) is an amount of money that \( m \) previously spent to purchase a piece of generation equipment that now has no resale or scrap value. An example of an *avoidable fixed cost* for \( m \) at time \( t \) is the start-up cost that \( m \) would have to incur in order to transition from its currently off-line state to a synchronized state\(^{26}\) by the start of day \( D+1 \) if \( m \) submits an offer into the day-D DAM to provide positive power injection during Hour 1 of day \( D+1 \) and the RTO/ISO clears this offer.\(^{27}\) Finally, an example of a *variable cost* for \( m \) at time \( t \) is the fuel cost \( V C_m(p,H,D+1) \) that \( m \) would have to incur to maintain a specific positive power-injection level \( p \) during some designated hour \( H \) of day \( D+1 \).

Consider, once again, a decision-maker DM at a current time \( t \) who must decide *now* whether or not to commit to undertaking an action of type \( A \) at a *future* time \( t + \Delta t \). We conclude this sub-section by expressing several decision principles for this DM in terms of the three-part partition (10) for total cost.

By definition, the DM’s sunk cost at time \( t \) is non-avoidable, hence incurred whether or not the DM agrees to the commitment at time \( t \). The DM’s sunk cost at time \( t \) should therefore play no role in the DM’s time-\( t \) commitment decision.

**“Sunk Cost is Sunk” Dictum:** A decision-maker at time \( t \) who is required to make a decision at time \( t \) should make this decision independently of his sunk cost at time \( t \).

Define the DM to be *risk-averse\(^{28}\)* if the DM is *not* willing to participate:

- in any *risky* undertaking with *zero expected payoff*, where the qualifier “risky” means there is some dispersion of positive-probability payoffs *around* the zero expected payoff, hence a positive probability of a negative payoff;
- in *any* undertaking (risky or certain) that has a *strictly negative expected payoff*.

If the DM is risk-averse, the DM should agree to the commitment at time \( t \) *if* the DM believes this would result *for sure* in a *strictly positive* net benefit, where:

\[
\text{Net Benefit} = \text{Benefit} - \text{Avoidable Cost} \tag{16}
\]

Note a strictly positive net benefit (16) would permit the DM to pay down at least part of his time-\( t \) sunk cost, assuming this sunk cost is positive.

---

\(^{26}\) A thermal generator is said to be in a *synchronized state* if it is an operating state that permits it to inject power into the grid, even if no such power injection is currently being undertaken.\(^{27}\) Kirsch and Strbac [25] refer to start-up cost as a “quasi-fixed cost”.

\(^{28}\) In economic theory, survival is often guaranteed *a priori* by postulating zero subsistence needs for consumers and assured solvency for firms. The risk aversion of decision-makers is then characterized as a preference attribute: namely, the degree to which their net benefit functions expressing preference orderings over possible payoffs exhibit concave curvature properties. In reality, strictly negative payoffs can pose grave *survival risks* for *any* person or commercial entity without deep financial pockets, inducing them to behave in a “risk averse” manner. One could choose to interpret this behavior as arising from a preference for “survival,” modeled by dramatic declines in net benefit as “death” or “insolvency” is approached. However, the routine explicit inclusion of subsistence needs and insolvency risks within economic models would permit a more profound understanding and appreciation of the role of institutional scaffolding in real-world economies.
Conversely, a risk-averse DM should agree to the commitment at time $t$ only if the DM’s expected net benefit from this commitment is non-negative, where net benefit is again defined as in (16). Otherwise, commitment at time $t$ would be an agreement to participate in an undertaking with a strictly negative expected payoff.

These observations are summarized in the form of a commitment principle, as follows:

**Commitment Principle:** A risk-averse DM should agree to commit at time $t$

- if the DM believes this commitment would result for sure in a net benefit (16) for himself that is strictly positive;
- only if the DM believes his expected net benefit from this commitment is non-negative, where net benefit is again defined as in (16).

Finally, making use of the 3-part partitioning (10) of total cost developed in this section, the preliminary definition BM7 given in Section 4.2 for supplier revenue sufficiency is stated below in a revised more-rigorous form for use in subsequent sections of this study.

**Definition RBM7: Rigorous Definition for Supplier Revenue Sufficiency.**
A supplier $i$ participating in a market $M$ for a $u$-asset $A$ is revenue sufficient for $M$ if the total revenue earned by supplier $i$ from this participation is sufficient to cover the total avoidable cost that supplier $i$ incurs from this participation, where avoidable cost is defined as in (15).

### 5.3.3 Presumption (P2) Results in Inadequate Avoidable Cost Remuneration

Multiple types of avoidable cost incurred by participants in U.S. RTO/ISO-managed wholesale power markets are listed and described in Appendix A.4. The list for avoidable fixed cost includes: capital investment cost; transaction cost; opportunity cost; and unit commitment cost. The list for variable cost includes: fuel cost; labor cost; intermediate good (supply chain) cost; equipment/software rental cost; equipment depreciation cost; transmission service charges; variable-cost offsets for sales of valuable bi-products; and disposal costs for waste bi-products.

A key concern is whether the current design of U.S. RTO/ISO-managed wholesale power markets permits suppliers to receive appropriate conceptually-coherent compensation for their incurred avoidable costs, i.e., their avoidable fixed costs and variable costs.

Consider, first, the remuneration of variable cost. As will next be shown, the supply-offer formulations currently required in U.S. RTO/ISO-managed wholesale power markets force suppliers to express their variable costs as functions of grid-delivered energy (MWh), with no consideration of dynamic power-path implementations. It is difficult to understand how any of the variable-cost categories listed in Appendix A.4 – including fuel cost – can be accurately measured and reported
solely as a function of grid-delivered energy, with no consideration of power-path implementation.

As reviewed in Section 2, the core design element for all seven U.S. RTO/ISO-managed wholesale power markets is a bid/offer-based DAM/RTM two-settlement system. In all seven RTOs/ISOs, the DAM SCED optimization conducted during the morning of each day D (conditional on given commitments for generation units) determines scheduled power dispatch set-points (MW) for each committed generation unit at the start of each hour H during the following day D+1.

Apart from ISO New England, these scheduled dispatch set-points determine co-optimized scheduled maintained power levels or maintained power-slope levels (hence scheduled energy deliveries) and scheduled operating reserve (unencumbered generation capacity levels) for each committed dispatchable generation unit for each hour H of day D+1. These determinations are subject to system constraints that include nodal and/or zonal reserve requirements for operating reserve with different availability characteristics.

The supply offers that dispatchable power resources submit to a day-D RTO/ISO-managed DAM for each hour H of day D+1 are intended to convey information to the RTO/ISO about their supply costs for day D+1.

More precisely, as detailed in [43], the supply offer $S_m(H, D+1)$ that a dispatchable power resource $m$ submits to a day-D RTO/ISO-managed DAM for some hour H during day D+1 is generally required to take the following two-part form: (i) a possibly-zero power-supply level $\bar{p}_m$ (MW) that $m$ is planning to maintain during hour H as a fixed (non-dispatched must-service) power supply, with no accompanying compensation request; and (ii) a finite (possibly zero) number $N$ of ordered non-overlapping power intervals $\Delta p_n$ accompanied by associated incremental energy prices $\pi_n$ ($/MWh). Each power interval $\Delta p_n$ consists of a range $[p_{sn}, p_{en}]$ of successively increasing power levels $p$ determining incremental power increases

---

29 SCED optimizations for U.S. RTO/ISO-managed DAMs and RTM sub-markets are similar, apart from operating-period duration and restrictions on LSE submission of fixed demand bids. For simplicity of exposition, this sub-section focuses solely on DAM SCED optimizations.

30 ISO New England conducts Forward Reserve Market (FRM) auctions for 10-minute contingency reserve and 30-minute supplemental reserve separately from DAM energy scheduling; see, for example, [6, Tables 1-2] and [16, Table 1].

31 The types of operating reserve procured on a co-optimized basis with energy in U.S. RTOs/ISOs (apart from ISO New England) include Regulation, Spinning Reserve, and Supplemental Reserve. See [16, Sec. IIA & Appendix].

32 As detailed in [2], ERCOT permits Qualified Scheduling Entities (QSEs) to submit hourly supply offers in a three-part form that allows inclusion of some avoidable fixed cost information in addition to variable cost information. For example, a supply offer submitted to a day-D ERCOT DAM by a currently off-line QSE for some hour H during day D+1 typically consists of three parts: a Startup Offer ($/start); a Minimum-Energy Offer consisting of an energy price ($/MWh) and a Low-Sustained Limit (LSL) power level (MW); and a non-decreasing piecewise linear Energy Offer Curve in the MW$/MWh plane consisting of a finite collection of linearly-connected power-price points $(p, \pi)$ whose power levels $p$ commence at the LSL level. Start-up cost is an important form of avoidable fixed cost for QSEs considering the submission of supply offers to ERCOT DAMs.
The crucial point illustrated by (17) is that the supply offer formulations currently used in U.S. RTO/ISO-managed DAMs force suppliers to report their variable costs as functions solely of grid-delivered energy blocks $E_i$ (MWh). However, as reported in Appendix A.4, many types of variable costs actually incurred by dispatchable power resources in U.S. RTO/ISO-managed wholesale power markets cannot be adequately expressed in this manner because these costs depend crucially on the dynamic properties of their subsequently dispatched power-paths.

Moreover, even if a dispatched power-path for an operating day D+1 in fact consists of one or more successive energy-blocks, and the only costs taken into account are energy production costs, these costs can take a much more complicated form than suppliers participating in U.S. RTO/ISO-managed wholesale power markets are currently permitted to report in their supply offers.

For example, consider a dispatchable generator $m$ at a grid location $b(m)$ that submits a supply offer at time 0 to an RTO/ISO-managed wholesale power market $M(T)$ for a future operating period $T = [t^s, t^e)$. This supply offer consists of a single energy-block for possible RTO/ISO dispatch during $T$. For concreteness, it is assumed that $m$ is in an off-line state (zero power generation) at time 0, and that $m$ needs to return to an off-line state as soon as possible after $T$ concludes.

Suppose generator $m$’s supply offer is cleared by the RTO/ISO at time 0. Generator $m$ must then take steps to ensure it is able to commence delivery of its offered energy-block at the start of operating period $T$.

As depicted in Fig. 7, the level $p(t) = g(t) - g^{synch}$ (MW) of $m$’s power injection into the grid at $b(m)$ at a time point $t$ is the difference between $m$’s total power
Fig. 7 Illustration of the energy production (MWh) needed before, during, and after an operating period T to support the dispatch of a single energy block (“Dispatch”) during T: namely, Start-Up (SU); Ramp-Up (RU); No-Load; Ramp-Down (RD), and Shut-Down (SD).

Generation $g(t)$ at $t$ and the power generation $g^{\text{sync}}$ that $m$ needs to be producing to maintain its synchronization\(^{33}\) to the grid. In order to ensure it can deliver its offered period-T energy-block, labeled “Dispatch” in Fig. 7, $m$ starts up its generating facilities at time $t^{\text{SU}} > 0$; and $m$ continues to ramp up its total power generation $g(t)$ until time $t^{\text{SYS}} > t^{\text{SU}}$, the time that $m$ first attains a synchronized state to the grid. At time $t^{\text{SYS}}$, generator $m$ is spinning at synchronous speed, ready to inject power into the grid; however, $m$ is not yet injecting any power into the grid.

Generator $m$ then continues to ramp up its total power generation $g(t)$ until its power injection level at $b(m)$ reaches its cleared dispatch level $p^{\text{disp}}$ at the start-time $t^s > t^{\text{SYS}}$ for operating period T. At time $t^s$, $m$’s total power generation is $g(t^s) = p^{\text{disp}} + g^{\text{sync}}$. However, only the injected portion $p^{\text{disp}}$ is visible to the RTO/ISO.

Generator $m$ then maintains the power injection level $p^{\text{disp}}$ at $b(m)$ during operating period T. At the end-time $t^e$ for operating period T, generator $m$ initiates a ramp-down process. During the initial phase of this ramp-down process, from time $t^s$ to time $t^{\text{SYE}}$, $m$ is still injecting power into the transmission grid at $b(m)$. However, once $m$’s total power generation level $g(t)$ falls to its synchronized total power generation level $g^{\text{sync}}$ at time $t^{\text{SYE}}$, $m$ is no longer injecting power into the grid. Generator $m$ then continues to ramp down its total power generation $g(t)$ until it reaches a shut-down state $g(t^{\text{SD}}) = 0$ at time $t^{\text{SD}}$.

\(^{33}\) A dispatchable power resource $m$ with a single electrical point-of-connection $b$ to a grid is said to be in a synchronized state with respect to this grid if it is in an operating state that permits it to inject power into the grid at $b$, even if $m$ is not currently injecting any power into the grid at $b$. 
The accumulation of the power that $m$ injects into the grid at location $b(m)$ during operating period $T$ to meet its supply-offer obligations to the RTO/ISO results in the delivery of the energy-block labeled “Dispatch” in Fig. 7. The energy production cost that $m$ incurs during period $T$ for this energy-block delivery will hereafter be referred to as $\text{DispatchCost}$.

The accumulation of the power that $m$ injects into the grid at location $b(m)$ during the interval $[t_{su}, t_{sys})$ in which $m$ is transitioning from an off-line state to a synchronized state is labeled “SU” in Fig. 7. The behind-the-meter energy produced by $m$ to maintain itself in a synchronized state during the next time interval $[t_{sys}, t_{sde})$ is labeled “NoLoad” in Fig. 7. Finally, the accumulation of the power that $m$ injects into the grid at location $b(m)$ during the time interval $[t_{sde}, t_{sd})$ in which $m$ is transitioning from a synchronized state at time $t_{sde}$ to a shut-down state at time $t_{sd}$ is labeled “SD” in Fig. 7. The costs incurred by $m$ for the three produced energy amounts SU, NoLoad, and SD will hereafter be referred to as $\text{SU}Cost$, $\text{NoLoadCost}$, and $\text{SDCost}$.

Note that Fig. 7 also depicts two additional power-injection regions labeled “RU” and “RD”. The ramp-up region RU reflects the need for $m$ to ramp up its power injection level to $p_{\text{disp}}$ at the start-time for operating period $T$, commencing from a zero power injection level at time $t_{sys}$, in order to meet its energy-block dispatch obligations during $T$. The ramp-down region RD reflects the need for $m$ to ramp down its power injection level to 0 after these period-T dispatch obligations are met, as part of $m$’s required progression towards an off-line state. The costs incurred by $m$ for the produced energy amounts RU and RD will hereafter be referred to as $\text{RUCost}$ and $\text{RDCost}$.

Revenue sufficiency for generator $m$ requires that $m$ receive full compensation for all avoidable fixed cost and variable cost that $m$ incurs due to its participation in $M(T)$. Focusing solely on energy production costs, revenue sufficiency for $m$ requires that $m$ receive compensation for the following costs:

\begin{align*}
\text{Avoidable Fixed Cost:} & \quad \text{SU}Cost, \, \text{NoLoadCost}, \, \text{SDCost}; \quad (19) \\
\text{Variable Cost:} & \quad \text{DispatchCost}, \, \text{RUCost}, \, \text{RDCost}. \quad (20)
\end{align*}

The cost components $\text{SU}Cost$, $\text{NoLoadCost}$, and $\text{SDCost}$ are avoidable fixed costs for $m$ because they depend on $m$’s commitment to produce an energy-block for operating period $T$ but they do not depend on the specific form of this energy-block. The cost components $\text{DispatchCost}$, $\text{RUCost}$, and $\text{RDCost}$ are variable costs for $m$ because they depend on $m$’s commitment to produce an energy-block for operating period $T$ and they also depend on the specific form of this energy block – namely, the power level $p_{\text{disp}}$ determining the height of this energy-block.

More complicated forms of dispatch entail more complicated forms of required energy production whose costs must be compensated if suppliers are to achieve revenue sufficiency.

For example, suppose the supply offer $S_{m}(D+1)$ submitted on day $D$ by a dispatchable generator $m$ to an RTO/ISO-managed day-ahead market $M(D+1)$ for an operating day $D+1$ offers the guaranteed availability of a set $\mathcal{P}_m$ of power-paths $p_{b(m)}$.
for possible RTO/ISO dispatch at \( m \)'s grid location \( b(m) \) during day \( D+1 \). More precisely, if the RTO/ISO clears \( S_{m}(D+1) \) on day \( D \), then on day \( D+1 \) the RTO/ISO is obligated to select and dispatch exactly one power-path from \( \mathbb{P}_m \).

Figure 8 depicts one of the power-paths in generator \( m \)'s offered power-path set \( \mathbb{P}_m \). This power-path consists of two successive energy blocks to be grid-delivered at generator \( m \)'s grid location \( b(m) \) during day \( D+1 \), labeled \( E1 \) and \( E2 \), together with two associated ramping regions \( R1 \) and \( R2 \). The first ramping region \( R1 \) reflects \( m \)'s need to ramp up to \( E1 \) starting from \( m \)'s minimum maintainable power injection level \( p_{\text{min}} > 0 \). The second ramping region \( R2 \) reflects \( m \)'s need to ramp up from the maintained power level for \( E1 \) to the higher maintained power level for \( E2 \).

If the RTO/ISO clears \( m \)'s supply offer \( S_{m}(D+1) \), the avoidable fixed cost incurred by \( m \) is the same no matter which particular power-path in \( \mathbb{P}_m \) the RTO/ISO selects and dispatches on day \( D+1 \). This follows because, by definition, avoidable fixed cost is an availability cost, not a dispatch cost.

Thus, as seen in Fig. 8, \( m \)'s avoidable fixed cost includes the energy production costs for the power injection regions labeled SU, NoLoad, and SD. However, in contrast to the example depicted in Fig. 7, this avoidable fixed cost also includes the energy production cost for the power injection region labeled “Min-Run” in Fig. 8 that arises because \( m \) is now assumed to have a positive minimum maintainable power injection level \( p_{\text{min}} > 0 \).
Moreover, suppose the RTO/ISO chooses to dispatch the particular power-path depicted in Fig. 8 during day D+1. Then the variable cost incurred by generator \( m \) includes the energy production costs for the ramping regions R1 and R2 as well as for the two successive energy-blocks E1 and E2. Note these energy production costs are strongly dependent on the specific dynamic attributes of the dispatched power-path for day D+1.

The key implications of Section 5.3.3 are highlighted below:

As detailed in Appendix A.4, suppliers participating in U.S. RTO/ISO-managed wholesale power markets incur a wide variety of avoidable fixed costs and variable costs. Many of these costs are not-well approximated as functions of grid-delivered energy. Moreover, even costs that are expressible as functions of grid-delivered energy can take a complicated form. Thus, the supply offer formulations for current U.S. RTO/ISO-managed wholesale power markets that force suppliers to express their costs as simple functions of grid-delivered energy do not permit these suppliers to ensure their revenue sufficiency.

### 5.4 Grid-Delivered Energy is Not a Commodity

#### 5.4.1 Critique of Presumption (P3): Overview

Presumption (P3) asserts that grid-delivered energy (MWh) conditional on delivery location and time is a commodity whose perfectly substitutable units (MWh) can (and should) be bought and sold in a spot market at a competitively-determined uniform market price ($/MWh).

The measurement and economic materials previously presented in Sections 3 and 4 are used in this section to demonstrate the conceptual incoherency of (P3) as a design principle for U.S. RTO/ISO-managed wholesale power markets. It is first argued that grid-delivered energy units (MWh) are not perfect substitutes (economically equivalent), conditional on delivery location and time; hence, grid-delivered energy fails to satisfy the core defining condition for a commodity. It is then argued that insistence on treating grid-delivered energy as a spot-market commodity necessitates a reliance on time-inconsistent market settlements.

#### 5.4.2 Grid-Delivered Energy Units (MWh) Are Not Perfect Substitutes Conditional on Delivery Location and Time

Participants in U.S. RTO/ISO-managed wholesale power markets typically do not consider a MWh of grid-delivered energy to be a perfect substitute for any other MWh of grid-delivered energy, conditional on delivery location and time. To the contrary, as previously noted in Section 5.2.3, power producers, power customers,
and RTOs/ISOs care about the *dynamic* attributes of the power-paths they inject, withdraw, or dispatch for injection/withdrawal during successive operating periods.

For example, power producers dispatched to inject power at a grid location \( b \) during an operating period \( T \) typically care about the wear-and-tear cost they incur for physical equipment damage when fast ramping is required in order to match received dispatch set-points. Power customers electrically connected to \( b \) during an operating period \( T \) typically care about the degree of flexibility they have to make idiosyncratic just-in-time power withdrawals at \( b \) during \( T \) to meet their diverse local power requirements.

Moreover, a primary fiduciary obligation of an RTO/ISO is protection against volumetric grid risk. This protection requires the continual advance availability of diverse dispatchable power-paths for possible RTO/ISO dispatch across the grid during successive operating periods \( T \) to ensure the continual balancing of net load.

Thus, if a producer, customer, or RTO/ISO were asked to assign a monetary value to a specific energy amount \( E^* \) (MWh) resulting from a power-path consisting of a sequence of injections (or withdrawals) of power \( p_b(t) \) at a designated grid location \( b \) during a designated operating period \( T = [t^*, t^*] \), typically they would not be able to do so without knowing the dynamic attributes of this power-path.

For example, consider the valuation that each of these entities would assign to each of the following four power-path options \( p_b(T) = (p_b(t) \mid t \in T) \) at a particular grid location \( b \) during a particular 24-hour operating period \( T \), where each option results in the *same* injected (or withdrawn) energy amount \( E^* = 12 \text{MWh} \):

**Power-Path (a):** Power \( p \) is injected (or withdrawn) at \( b \) at level \( p = 24\text{MW} \) during the first half-hour of \( T \) and at level \( p = 0\text{MW} \) during the remainder of \( T \).

**Power-Path (b):** Power \( p \) is injected (or withdrawn) at \( b \) at level \( p = 0.5\text{MW} \) throughout \( T \).

**Power-Path (c):** Power \( p \) is injected (or withdrawn) at \( b \) at level \( p = 1\text{MW} \) during every other hour of \( T \), with \( p = 0\text{MW} \) during the remaining hours of \( T \).

**Power-Path (d):** Power \( p \) is injected (or withdrawn) at \( b \) during \( T \) in a flexible manner that is entirely up to the entity, apart from the requirement that the total energy resulting from these injections (or withdrawals) must equal 12MWh.

The flexible power-path option (d) would presumably be assigned a higher value than the rigid options (a) through (c). Indeed, option (d) would presumably be assigned at least as high a value as any other power-path option for power injection (or withdrawal) at \( b \) during \( T \) that results in an energy amount \( E^* = 12 \text{MWh} \) since option (d) encompasses all such options.

As this example clearly illustrates, grid-delivered energy (MWh) conditional on delivery location and time does *not* function as a commodity within the context
of U.S. RTO/ISO-managed wholesale power markets. It thus follows from Section 4.4 that the following constructions – essential underpinnings for the Two Settlement System – are conceptually ill-defined: demand schedule (ordinary or inverse); supply schedule (ordinary or inverse); marginal cost function; marginal benefit function; the Competitive (D=S) Market Clearing Rule (3); and the Competitive (MB=MC) Spot-Pricing Rule (4).

The key implication of Section 5.4.2 is highlighted below:

Attempts to justify the current reliance of U.S. RTO/ISO-managed wholesale power markets on the Two Settlement System by pointing to the efficiency and optimality properties of competitive commodity spot markets satisfying the competitive (MB=MC) spot-pricing rule (4) are not conceptually supportable.

5.4.3 The Two-Settlement System Results in Time-Inconsistent Settlements

A second fundamental conceptual concern regarding presumption (P3) is that a grid-supported RTO/ISO-managed wholesale power market $M(T)$ conducted for an operating period $T$ cannot be a spot market due to the speed of real-time grid operations. That is, referring back to the market time-line depicted in Fig. 5, the market $M(T)$ cannot coincide with its operating period $T$.

Indeed, as carefully reviewed in Section 2, the DAM/RTM two-settlement system at the core of each U.S. RTO/ISO-managed wholesale power market is a collection of grid-supported forward markets $M(T)$ with positive-duration look-ahead horizons LAH(T), i.e., with positive-duration time intervals between the close of $M(T)$ and the start of $T$. The scheduled generation-unit commitments and dispatch set-points for hour $H$ of day $D+1$, determined in the day-D DAM, are subject to change in supplementary unit-commitment processes and in RTM sub-markets held between the close of the day-D DAM and the start of hour $H$ on day $D+1$.

Nevertheless, the settlements\(^{34}\) for these scheduled next-day unit commitments, generation levels, and operating reserve levels are determined at the end of day $D$ as if they were actual spot-market transactions carried out on day $D$; see Fig. 1. This pay-for-performance in advance of actual performance typically results in time-inconsistent\(^{35}\) settlements, i.e., settlements determined and assigned to resources on day $D$ for unit commitments, energy levels, and operating reserve levels scheduled for day $D+1$ that are subsequently adjusted by OOM and RTM LMP payments due to discrepancies that arise between scheduled and actual outcomes on day $D+1$.

---

\(^{34}\) These settlements include out-of-market (OOM) make-whole payments for partial reimbursement of avoidable fixed costs (e.g., opportunity costs) for committed generation units plus DAM SCED-determined nodal (dual-variable based) price payments intended to cover the variable costs for scheduled energy and reserve provision.

\(^{35}\) A multi-stage optimization problem that jointly determines an optimal solution $(s_0, s_1, \ldots, s_N)$ for successive time-periods $(s_0, s_1, \ldots, s_N)$ at some time $0$ is said to be time-inconsistent if re-optimization undertaken at the beginning of some later time-period $s_n$ with $0 < n \leq N$ results in an optimal solution for $(s_0, \ldots, s_N)$ that deviates from $(s_0, \ldots, s_N)$. See [45, Sec. 10.2].
5.5 Supplier Revenue Sufficiency Requires 2-Part Pricing

5.5.1 Overview

Considered together, the four presumptions (P1)–(P4) make the following strong claim regarding supplier revenue sufficiency:

*(P(1)–(P4) Claim:* The total revenue attained by suppliers participating in a market \( M(b,T) \) for grid-delivered energy (MWh) that is conditioned on grid delivery location \( b \) and operating period \( T \), and that is settled by means of Locational Marginal Prices \( \text{LMP}(b,T) \) ($/MWh), will suffice to cover the total fixed and variable cost that suppliers incur in this market.

In the following two sub-sections it will be shown that this strong revenue sufficiency claim is conceptually problematic on two grounds:

- Spot and forward markets do not assure coverage of sunk costs;
- Forward markets do not assure coverage of avoidable fixed costs.

5.5.2 Spot and Forward Markets Do Not Assure Coverage of Sunk Costs

Let \( M \) denote a marginal-pricing competitive commodity spot market \( \text{MP-CCSM}(Q) \) for a commodity \( Q \), as defined in Section 4.4. Also, let supplier \( i \) denote a participant in \( M \) at a given time \( t \).

As established in *Lemma 4.2*, any fixed cost incurred by supplier \( i \) as a result of its prior decision to participate in \( M \) at time \( t \) is sunk cost for supplier \( i \) at time \( t \); i.e., non-avoidable fixed cost. Moreover, nothing in the defining conditions for an MP-CCSM\((Q)\) ensures that the revenues attained by supplier \( i \) from participation in \( M \) at time \( t \) will suffice to cover any or all of supplier \( i \)’s sunk cost at time \( t \).

More precisely, as established in *Lemma 4.5*, market efficiency holds at any competitive equilibrium point \( e^* = (\pi^*, q^*) \) for \( M \) in the sense that the total net surplus ($) extracted by suppliers and buyers at \( e^* \) from the underlying commodity spot market \( \text{CSM}(Q) \) for \( M \) is as large as possible. However, market efficiency does not guarantee that the net surplus ($) extracted by suppliers at \( e^* \) will cover any or all of their sunk costs.

This is a special case of a basic economic principle: *There is no economic justification for instituting market rules that ensure market suppliers are reimbursed for sunk costs, whether these markets are constituted as spot markets or forward markets.* As stressed by the well-known “Sunk Cost is Sunk” dictum in Section 5.3.2, sunk costs at a given point in time are already-incurred costs that cannot be changed by any current decisions; hence, sunk costs should have no effect on these decisions.
Indeed, the only way a supplier receives assured coverage for some or all of his sunk costs through a market participation is if he has some type of structural or strategic market advantage\footnote{As discussed more carefully in [37], structural market advantage refers to an instituted feature of a market that systematically favors some market participants over others. In contrast, strategic market advantage is an opportunity available to a market participant to influence market outcomes in their favor in an officially unintended manner through some behavioral means. The standard economic term for these types of market advantages is “market power.” However, the use of “market power” in studies of electric power markets could cause confusion.} relative to other actual or potential suppliers that reduces their ability to compete for his customers should he attempt to charge his customers for sunk costs. Examples of situations giving rise to supplier market advantage include:

- **Regulatory Protection (Entry Barrier):** A supplier might have patent protection for his product that prevents other suppliers from producing this same product.
- **Product Differentiation:** A supplier’s product might have a special attribute (e.g., sale location, flavor based on secret recipe), highly valued by buyers, that differentiates it from all other products currently being supplied in the market and that is hard (or impossible) for other suppliers to copy.
- **Supply-Capacity Constraints:** A supplier might be a monopolist (sole supplier) with respect to “residual demand” customers that other capacity-constrained suppliers are unable to service, enabling him to include “extra” charges in the product price he sets for these customers for coverage of his sunk cost.

Conversely, buyers participating in a “perfectly contestable” market have no incentive to compensate supplier sunk costs. Roughly defined, a market for a product is perfectly contestable if any participating supplier who is charging a price for this product that results in revenues strictly exceeding the avoidable cost that must be incurred to produce this product can be successfully challenged and replaced by existing or newly entering rival suppliers able to charge a lower price for this product while still ensuring coverage of this avoidable cost. See Baumol et al. [3].

However, it is important to understand that a supplier’s coverage (or not) of his sunk costs through a market participation does not violate the “Sunk-Cost is Sunk” Dictum in Section 5.3.3. By definition, such coverage does not affect the amount of these sunk costs. Thus, a supplier at a given time $t$ seeking to maximize his expected net benefit should choose a course of action compatible with this objective. The greater the expected net benefit, the greater the likelihood that sunk costs are covered in part or in whole. However, the converse is false. Taking an action to maximize sunk-cost coverage does not ensure maximization of expected net benefit.

### 5.5.3 Forward Markets do Not Assure Coverage of Avoidable Fixed Costs

By definition BM2 in Section 4.2, a forward market for an asset $A$ is a market for which transacted amounts of $A$ and payments for these transacted amounts of $A$ are
determined in advance of the delivery of these transacted amounts of $A$. Avoidable
fixed cost essentially always arises for suppliers participating in forward asset mar-
kets because avoidable fixed cost includes opportunity cost, i.e., earnings foregone by not committing these assets to an alternative next-best use.

Forward markets for assets take highly diverse forms, particularly if the trans-
acted asset is not a commodity. For example, consider the varied types of forward markets instituted for salaried employment (person-hours), auto accident coverage (damage repair), and health-care provision (e.g., annual check-ups). The suppliers participating in these forward markets receive varied degrees of reimbursement for the avoidable fixed costs they incur to guarantee future asset availability and for the variable costs they incur for real-time asset deliveries.

As discussed in Section 5.4.3, U.S. RTO/ISO-managed Day-Ahead Markets (DAMs) and Real-Time Markets (RTMs) for grid-delivered energy (MWh) operate over high-voltage AC transmission grids. These markets are necessarily forward markets, not spot markets, due to the speed of real-time grid operations. That is, these markets have look-ahead horizons with positive time-durations; see Fig. 5.

Various types of avoidable fixed cost and variable cost that arise for suppliers participating in U.S. RTO/ISO-managed DAMs and RTMs are listed and described in Appendix A.4. The listed types of avoidable fixed cost include: capital investment cost; transaction cost; opportunity cost, and unit commitment cost. The listed types of variable cost include: fuel cost; labor cost; intermediate-good (supply chain) cost; equipment/software rental cost; depreciation cost; transmission service charges; cost offsets from sales of valuable bi-products; and waste bi-product disposal costs.

For many of these costs it is difficult to envision how they could be expressed in a conceptually sound and empirically-credible manner as functions solely of grid-delivered energy (MWh). Yet, this is what would be needed in order for suppliers participating in U.S. RTO/ISO-managed DAMs and RTMs to be assured full coverage of these costs through some form of LMP ($/MWh) settlement system.

Fortunately, reliance on an LMP settlement system is not necessary. Forward markets instituted in other industries routinely rely on two-part pricing contracts to ensure supplier revenue sufficiency, i.e., to ensure supplier revenues are sufficient to cover supplier avoidable fixed costs as well as supplier variable costs. Section 7 of this study reviews key features of an alternative Linked Swing-Contract Market Design [45] for RTO/ISO-managed wholesale power markets that demonstrates how two-part pricing contracts could be advantageously introduced in these markets to ensure supplier revenue sufficiency.
6 Legacy Core Design: Roadblock for Grid Decarbonization

6.1 Overview

U.S. RTO/ISO-managed wholesale power markets are large complex organizations. From an external vantage point, the continued reliance of these markets on the legacy two-settlement system design reviewed in Section 2 appears to be greatly hindering these markets from transitioning smoothly to decarbonized grid operations. This section discusses several external indicators in support of this concern.

6.2 Proliferation of Participation Models

The continued emphasis of U.S. RTO/ISO-managed Day-Ahead Markets (DAMs) and Real-Time Markets (RTMs) on grid-delivered energy (MWh) as the key transacted product, in keeping with presumptions (P1)–(P4), has resulted in a proliferation of participation models functioning as artificial market entry barriers.

More precisely, to participate in these DAMs/RTMs, a power resource must be classified in accordance with a designated taxonomy of participation models, each with its own eligibility rules and performance requirements. At the top of this taxonomy are two categorizations: “Energy (MWh)” and “Operating Reserve (MW).” The latter category consists of various forms of unencumbered generation capacity (MW) distinguished by availability characteristics.

However, this entire DAM/RTM taxonomy is conceptually problematic because, as discussed with care in [45, Sec. 3.2.1], “Energy” and “Operating Reserve” are not conceptually well-defined as independent participation categories for such markets. For example, consider an RTO/ISO-managed DAM SCUC/SCED optimization conducted on day D in order to co-optimize scheduled energy deliveries and scheduled operating reserve for day D+1. As usual, suppose the forecasted net fixed load-profile at a grid bus b (or for a grid zone z) for day D+1 functions as a centroid (i.e., mid-point anchor) for a nodal (or zonal) uncertainty set expressing RTO/ISO-specified operating reserve requirements for day D+1 at b (or for z). Then:

- **Energy** levels determined in the day-D DAM SCUC/SCED optimization for scheduled dispatched delivery at bus b (or within zone z) for day D+1 consist of a collection of RTO/ISO-dispatchable power-paths deemed capable of covering the centroid of the uncertainty set for bus b (or zone z) during day D+1.
- **Operating reserve** levels determined in the day-D DAM SCUC/SCED optimization for scheduled availability at bus b (or zone z) during day D+1 consist of a collection of RTO/ISO-dispatchable power-paths deemed capable of covering the remainder of the uncertainty set for bus b (or zone z) during day D+1.

---

37 For example, see [34] for a description of the complicated eligibility requirements that power resources must satisfy in order to participate in DAM/RTM processes for the Midcontinent Independent System Operator (MISO) as one of MISO’s permitted types of “qualified resources.”
Thus, there is no fundamental conceptual distinction between optimal scheduled energy dispatch levels and optimal scheduled operating reserve levels for day \( D+1 \). Both are collections of RTO/ISO-dispatchable power-paths for day \( D+1 \) whose purpose is to ensure the balancing of uncertain net fixed load during day \( D+1 \).

### 6.3 Proliferation of Flexibility Products

A related concern is the proliferation of flexibility products as supplemental support for real-time net-load balancing.

U.S. RTO/ISO-managed DAMs/RTMs have traditionally relied on operating reserve products taking the form of unencumbered generation capacity distinguished by speed of availability. At any given time, unencumbered generation capacity is generation capacity that is currently without scheduled dispatch obligations. Thus, operating reserve and scheduled generation dispatch are joint products.\(^{38}\)

Three types of operating reserve procured on a co-optimized basis with energy in U.S. RTO/ISO-managed DAMs (apart from ISO New England) are Regulation, Spinning Reserve, and Supplemental Reserve [16, Sec. II.A & Appendix]. Next-day hourly prices for each operating reserve type, derived as dual variable solutions for hourly reserve requirement constraints incorporated into the system constraints for SCED optimizations, take the form of foregone energy-price (LMP) payments.

As noted in Section 1, net loads for U.S. RTO/ISO-managed wholesale power markets are expected to become increasingly uncertain and volatile as these markets transit to increased reliance on IPRs and more active demand-side participation. Indeed, net loads for CAISO and other U.S. RTOs/ISOs are already exhibiting more frequent and dramatic down/up ramping swings. Consequently, the current heavy reliance on unencumbered generation capacity (MW) for operating reserve is becoming increasingly risky.

Industry and academic researchers are thus exploring the possible introduction of new types of “flexibility products” to facilitate the balancing of more uncertain and volatile net loads. For example, FERC has approved proposed ramp flexibility products for CAISO, MISO, and SPP; see [16, p. 12].

A major concern regarding these developments is that the newly approved RAMP (MW/min) products, together with already instituted CAP (MW) and ENERGY (MWh) products, are not independently produced products that can be separately transacted at separately determined prices in a conceptually consistent manner. To the contrary, they are the correlated attributes of individual power-paths, hence they constitute joint products.

The conceptually-problematic treatment of RAMP, CAP, and ENERGY as independently produced and priced products presumes the value of a power-path can

---

\(^{38}\) In economics, two or more products are said to be joint products if their production results jointly from the use of common inputs in a production process. Operating reserve and scheduled generation dispatch are rival joint products, meaning that – for any given level of inputs (here generation capacity) – an increase in one of the joint products requires a decrease in the other. The use of a single fuel source to co-generate electricity and steam is an example of a non-rival joint-production process well known to engineers.
be appropriately measured by means of separate prices assigned to its attributes, treated as independent products. In actuality, the attributes of a power-path $p_b(T) = (p_b(t) | t \in T)$ for an operating period $T$ – such as power-delivery start-time $t^o \in T$, power capacity (MW) profile, ramp-rate (MW/min) profile, power-factor (kW/kVA) profile, power-delivery duration $\Delta t$, and total grid-delivered energy (MWh) – are correlated jointly-produced attributes. A change in any one attribute of a power-path can necessitate changes in its other attributes.

In economics, hedonic pricing is the pricing of a product on the basis of prices separately assigned to its intrinsic physical attributes as well as to its external circumstances. In some situations it might be desirable to use a hedonic-price approximation for the variable cost $\phi_m(p_m(T))$ ($) that a dispatchable power resource $m$ would incur for RTO/ISO-dispatched delivery during $T$ of each of its offered power-paths $p_m(T)$.

For example, the variable cost $\phi_m(p_m(T))$ might be approximated as a linear combination of contractually-agreed metric functions that separately assign costs ($) for post-T verified power capacity (MW) profile, ramp-rate (MW/min) profile, and total grid-delivered energy (MWh), as follows:

$$\phi_m(p_m(T)) \approx C^{\text{CAP}}(p_m(T)) + C^{\text{RAMP}}(p_m(T)) + C^{\text{ENERGY}}(p_m(T)). \quad (21)$$

However, it would be conceptually incorrect and highly problematic in practice to ignore that the CAP, RAMP, and ENERGY “products” whose costs are evaluated in (21) are in fact highly-correlated attributes of a single power-path $p_m(T)$.

### 6.4 Proliferation of Out-of-Market Make-Whole Payments

U.S. RTO/ISO business practice manuals provide detailed descriptions of business operations for stakeholders and other interested parties. The manuals that focus specifically on the trade and settlement of energy and operating reserve have become extremely complex over the years. Much of this growing complexity has arisen from the need to explain various types of out-of-market (OOM) payments that RTOs/ISOs have instituted for their supplier participants as supplements to their market-determined revenues in an attempt to ensure coverage of their incurred costs.

The names and definitions of these OOM payments are not standardized across the seven U.S. RTOs/ISOs, and their relationship to the OOM payments discussed in FERC Orders is not entirely clear. An essential aspect in need of clarification is the distinction between “uplift” OOM payments and “make-whole” OOM payments.

For example, FERC Order No. 844 [13, Sec. I.2] on “uplift transparency,” released in 2018, directs RTOs/ISOs to provide a more transparent monthly reporting of “uplift payments” characterized as follows:

“RTO/ISO markets can be affected by a number of operational challenges such as unplanned transmission and generation outages and the need to maintain adequate voltage throughout the system. Limitations in the ability of the
market software to incorporate all reliability considerations can at times result in prices that fail to reflect some of these challenges. In such situations, certain resources needed to reliably serve load may not economically clear the market and RTOs/ISOs must take out-of-market actions (i.e., operator-initiated commitments) to ensure system needs are met. These actions give rise to uplift costs. ... Uplift payments reflect the portion of the cost of reliably serving load that is not included in market prices.” [pp. 4-7]

Thus, FERC Order 844 characterizes uplift payments as OOM reimbursements to power resources for undertaking RTO/ISO-requested OOM actions deemed necessary to maintain grid reliability.

In contrast, RTOs/ISOs have instituted various types of OOM make-whole payments for suppliers participating in DAMs/RTMs whose market-determined revenues fail to provide coverage for certain types of avoidable fixed cost. Examples include ERCOT’s payments for accumulated power usage (energy) required for start-up, and for the maintenance of an on-line state at a minimum possible maintained power-injection level; see Fig. 8 in Section 5.3.3.

A key concern regarding existing OOM make-whole payments is that they do not ensure supplier revenue sufficiency. For example, the OOM make-whole payments that ERCOT awards to suppliers for start-up and for maintenance of an on-line state by no means provide full coverage for all of the avoidable fixed costs that these suppliers incur; see Appendix A.4.

However, OOM make-whole payment methods proposed to address this concern have tended to blur the operationally-critical distinctions among sunk cost, avoidable fixed cost, and variable cost discussed in Section 5.3.2. For example, the Notice Of Proposed Rule-making (NOPR) released by FERC in 2016 [10], and subsequently withdrawn by FERC in 2017 [11], would have required unit-commitment costs for fast-start resources to be incorporated into the energy and operating reserve prices determined in co-optimized DAMs/RTMs for variable-cost remuneration. Unit commitment costs are an important form of avoidable fixed cost; see Appendix A.4. As reviewed by Hartman [19, pp. 6-7], the original NOPR release encouraged commentators to recommend that unit-commitment costs for other types of generation should also be incorporated into these energy and operating reserve prices.

Moreover, industry and academic researchers are continuing to explore extended-LMP methods for the broad-based incorporation of avoidable fixed costs into DAM/RTM-determined LMPs ($/MWh). A key conceptual argument against these extended-LMP initiatives is they do not permit proper remuneration for supplier risk-reduction services.

More precisely, avoidable fixed costs are insurance costs, not production costs. That is, they are the costs that suppliers with cleared supply offers for a future operating period T must now incur in order to guarantee the availability of the reserve (dispatchable power-production capabilities) that they have offered for T. This guaranteed availability provides a critically important service: namely, reduction of volumetric grid risk for T. Thus, these suppliers should be appropriately compensated
for the cost of providing this service whether or not the RTO/ISO subsequently dispatches them for actual power-path delivery during T.

The incorporation of supplier avoidable fixed costs into DAM/RTM energy prices via an extended-LMP method would thus prevent suppliers from receiving appropriate compensation for risk-reduction services. Under an extended-LMP method, a supplier would only receive compensation for the provision of risk-reduction services for a future operating period T if the RTO/ISO subsequently chooses to dispatch this supplier for actual power-path delivery during T.

Consider the following analogous situation. Suppose a Fire-Insurance Company (FIC) is interested in providing risk-reduction products to households for some future period T. These risk-reduction products are fire-insurance contracts sold to households in advance of T that promise to provide make-whole house repairs in case of a period-T house-fire. The FIC’s insurance pool consists of the subset of households that purchase one of these fire-insurance contracts in advance of T.

Suppose the FIC is not permitted to require each household in its insurance pool to pay a common “premium payment” ($) in advance of T whether or not the household subsequently sustains house-fire damage during T. In order to ensure its solvency, the FIC would then be forced to require each household in its insurance pool to pay the full avoidable cost of any house-fire repair that the FIC provides to this household during period T. Thus, why limit house repairs to burned houses? The would-be “fire insurance” company FIC is thus incentivized to function as an ordinary “home repair” company with no provision of risk-reduction services.

Finally, various conceptual and practical arguments can be given against the proposed incorporation of supplier avoidable fixed cost into any energy or operating reserve price determined for a co-optimized DAM/RTM. The incorporation of a positive avoidable fixed cost $c^o$ incurred in advance of an operating period T into the period-T price $\pi_A(T)$ determined for a product A that would be procured in total period-T amount $a(T)$ at price $\pi_A(T)$ requires changing the recorded book-price $\pi_A(T)$ to a price level $\pi'_A(T) := \pi_A(T) + \Delta \pi_A(T)$ such that:

$$\pi'_A(T) \times a(T) = c^o + \pi_A(T) \times a(T), \quad (22)$$

hence:

$$\Delta \pi_A(T) \times a(T) = c^o. \quad (23)$$

However, as carefully discussed in Section 5.3.2, an avoidable fixed cost is a cost that: (i) arises from a decision to commit now to undertaking some designated type of action at a future time; and (ii) does not depend on the specific form of

---

39 Suppose: (i) the number N of households in FIC’s insurance pool is large; (ii) each household in this insurance pool has the same small independent probability $\beta \in (0, 1)$ of experiencing a house-fire during T; and (iii) each household in this insurance pool would have the same house-repair cost HRC ($) in case of a house-fire. Then, by the Law of Large Numbers, the FIC can “almost surely” guarantee full coverage of its actual total house-fire repair cost during T by: (a) requiring each household in its insurance pool to pay a relatively small premium $\beta \times HRC$ in advance of T; and (b) offering each household in its insurance pool free make-whole house-fire repair during T.
the undertaken action. Thus, there is no guarantee that an avoidable fixed cost $c^o$ incurred in advance of some future operating period $T$ due to a commitment to be available for possible RTO/ISO-instructed delivery of a currently uncertain amount of product $A$ during $T$ can be expressed as a stable function of the actual delivery $a(T)$ of product $A$ during $T$. Consequently, the recorded book-prices $\pi'_A(T)$ for each operating period $T$ could be highly unstable, implying there is no practical advantage gained relative to making simple direct OOM make-whole payments $c^o$.

Second, if the book-price $\pi'_A(T)$ were instead implemented as the period-$T$ price of product $A$ in place of $\pi_A(T)$, the transacted $A$-amount would presumably change from $a(T)$ to $a'(T) = a(T) + \Delta a(T)$ for some non-zero increment $\Delta a(T)$. That is, any attempt to endogenize the needed “price distortion” by incorporating it into a co-optimized DAM/RTM SCUC/SCED formulation in advance of the determination of an optimal solution would typically result in an inefficient distortion of the resulting quantity solution.

Third, grid-delivered energy and operating reserve do not function as independently produced products, let alone as independently produced commodities, within the context of a co-optimized DAM/RTM. Hence, the uniform “unit prices” currently derived in these markets for grid-delivered energy and operating reserve, conditional on a given delivery location and operating period, are conceptually-problematic constructs with questionable efficiency and optimality justifications. The incorporation of an avoidable fixed cost into such a presumed uniform market price would thus simply result in a double-layered price distortion.

6.5 Growing Revenue Insufficiency Concerns

Revenue sufficiency is said to hold for a supplier $i$ participating in a market $M$ if the total revenue earned by supplier $i$ from this participation is sufficient to cover the total avoidable cost that supplier $i$ incurs from this participation.\footnote{See the rigorous definition RBM7 for supplier revenue sufficiency provided in Section 5.3.2.}

The proliferation of OOM make-whole payments reported in Section 6.4 indicates that U.S. RTO/ISO-managed wholesale power markets are not currently ensuring revenue sufficiency for their participant suppliers. As discussed in previous sections of this study, and addressed in greater detail in [45], this revenue insufficiency appears to be arising as the result of fundamental conceptually-problematic economic presumptions embedded in their legacy core DAM/RTM two-settlement system designs: namely, presumptions (P1)–(P4).

6.6 Ptolemaic Epicycle Market-Design Conundrum

As reported in previous sections, the retention of the DAM/RTM Two-Settlement System – hence presumptions (P1)–(P4) – as the core underpinning for current U.S. RTO/ISO-managed wholesale power market operations is causing increasing difficulties: namely, a perceived increasing need to rely on conceptually-problematic
participation models, flexibility products, and OOM make-whole payments to counter growing supplier revenue insufficiency problems.

Recall the difficult transition from the earth-centric circular-orbit solar-system model developed by Claudius Ptolemy (circa 100-170 AD), supported by a proliferation of postulated “epicycles,” to the sun-centric elliptical-orbit solar-system model due to Nicolaus Copernicus (1473-1543) and Johannes Kepler (1571-1630). In analogy to this earlier episode, the conundrum facing current U.S. RTOs/ISOs can be characterized as follows:

**Ptolemaic Epicycle Conundrum for Market Design (“Onion Problem”):**

- Fundamental conceptual inconsistencies in the core design principles instituted for a market result in operational difficulties.
- These operational difficulties are addressed by introducing a new layer of rules (an “epicycle”) around the initial core design principles, which results in further operational difficulties.
- Rule-layer (epicycle) accretion then continues to occur because – ignoring the “Sunk Cost is Sunk” Dictum – correction of the fundamental conceptual inconsistencies in the core design principles is persistently deemed to be too costly to undertake.

7 What Else Can Be Done: An Illustrative Alternative

7.1 Overview: The Linked Swing-Contract Market Design

Previous sections of this study provide support for two contentions. First, four conceptually-problematic presumptions (P1)–(P4) are built into the DAM/RTM Two-Settlement System design at the core of current U.S. RTO/ISO-managed wholesale power market operations. Second, these four presumptions are hindering the smooth transition of these markets to decarbonized grid operations with increasingly diverse participants.

One common response to this contention is that, while imperfect, the DAM/RTM Two-Settlement System should be retained because it is the best available option. Specifically, the only viable alternatives to this system would be the adoption of an inefficient zonal pricing system or a return to an inefficient centrally-administrated cost-based pricing system.

This study disputes this pessimistic assessment. As noted in previous sections, grid-supported power markets must necessarily be *forward* markets due to the speed of real-time grid operations. The essential product transacted in these forward markets is *reserve* (physically-covered insurance) for protection against volumetric grid risk during future operating periods. Thus, the key to the appropriate redesign of current U.S. RTO/ISO-managed wholesale power markets is to recognize that grids must be permitted to operate as nimble flexibility-support insurance mechanisms.
Physically-covered insurance in other industries is routinely offered by means of legally-enforceable two-part pricing contracts. These contracts ensure coverage of avoidable fixed costs for the insuring parties through contractually-specified insurance premiums (in lump-sum or amortized form) and coverage of variable costs (if any) for the insuring parties through contractually-specified terms such as co-payments, deductibles, and negligence penalties.

It would thus seem reasonable for reserve offers in U.S. RTO/ISO-managed wholesale power markets to take the form of two-part pricing insurance contracts that permit reserve suppliers to be separately compensated for two distinct types of cost: (i) *avoidable fixed cost* incurred to guarantee reserve availability, hence reduction of volumetric grid risk, for future operating periods; and (ii) *variable cost* (if any) incurred for RTO/ISO-dispatched delivery of reserve, hence for the actual delivery of power-paths during real-time operations to meet just-in-time customer power demands and grid reliability requirements.

A proposed use of two-part pricing insurance contracts to reduce volumetric grid risk for future operating periods is a key innovative feature of the *Linked Swing-Contract Market Design* [45]. This design, proposed for grid-supported centrally-managed wholesale power markets, has been developed and tested in a series of studies\textsuperscript{41} at design readiness levels\textsuperscript{42} ranging from conceptual formulation through computational performance testing for 118-bus test systems.

The remainder of this section briefly discusses motivating principles, key innovative features, and legacy-design comparisons for the proposed Linked Swing-Contract Market Design. As will be seen, this proposed design appears well-suited for the scalable support of decarbonized grid operations with increasingly diverse participants.

### 7.2 Four Motivating Design Principles

The development of the Linked Swing-Contract Market Design [45] has been motivated by four fundamental *Power-Market Design (PMD)* principles for grid-supported centrally-managed wholesale power markets, as follows:

**[PMD1]:** Grid-delivered energy (MWh) is not a commodity because the benefit and cost valuations that participants assign to an energy amount $E^*$ delivered at a grid location $b$ during an operating period $T$ typically depend strongly on the dynamic attributes of the sequence of power injections and/or withdrawals (MW) that implement this delivery of $E^*$ at $b$ during $T$.

**[PMD2]:** A grid-supported wholesale power market $M(T)$ for an operating period $T$ is necessarily a *forward* market for $T$ due to the rapidity of real-time grid operations.

\textsuperscript{41} See [1, 20, 28, 29, 27, 33, 44, 45, 47].

\textsuperscript{42} In analogy to DOE Technology Readiness Levels (TRLs) for physical technology, nine Design Readiness Levels (DRLs) for institutional design ranging from conceptual formulation DRL-1 to real-world implementation DRL-9 are formulated in [45, Ch.18.3].
Leigh Tesfatsion

[PMD3]: A Dispatchable Power Resource (DPR) participating in a grid-supported centrally-managed wholesale power market M(T) for a future operating period T provides two distinct types of product:

**Physically-Covered Insurance (Reserve Availability):** Guaranteed availability of dispatchable power-path production capabilities for possible central-manager dispatch during future operating periods to protect against volumetric grid risk;

**Real-Time Performance (Reserve Delivery):** Real-time centrally dispatched delivery of power-paths during successive operating periods to meet just-in-time customer power demands and grid reliability requirements.

[PMD4]: A Dispatchable Power Resource (DPR) participating in a grid-supported centrally-managed wholesale power market M(T) for a future operating period T should be permitted to submit its supply offer for M(T) in a two-part pricing form to ensure its revenue sufficiency, as follows:

- The supply offer should permit the DPR to include an offer price sufficient to compensate the DPR prior to T for any avoidable fixed cost the DPR must incur to guarantee the availability of its offered reserve for T;
- The supply offer should permit the DPR to include a performance payment method that assures the DPR receives appropriate compensation after T for any variable cost the DPR incurs for actual verified power-path delivery undertaken during T in accordance with period-T dispatch instructions.

### 7.3 Innovative Design Features: Summary Form

In accordance with [PMD1]–[PMD4], the proposed Linked Swing-Contract Market Design [45] includes four innovative features intended to facilitate the efficient and reliable operation of U.S. RTO/ISO-managed wholesale power markets as they transition to decarbonized grid operations with increasingly diverse participants.

The first innovative feature is a fundamental change in the current focus of U.S. RTO/ISO-managed wholesale power markets on grid-delivered energy as the basic transacted product. Under the proposed design, the basic transacted product is instead conceived to be a general form of reserve: namely, the offered availability of dispatchable power-production capabilities for possible dispatch at designated grid locations during designated future operating periods.

The second innovative feature is a fundamental change in the current primary focus of RTO/ISOs on short-run (day-ahead and intra-day) markets for grid-delivered energy, supported by ancillary services and supplemental capacity commitment processes. Under the proposed design, the primary role of an RTO/ISO is instead envisioned to be the continual procurement of physically-covered insurance (reserve) for protection against volumetric grid risk. To enable this procurement, the RTO/ISO manages a linked collection of forward bid/offer-based reserve markets M(T) for future operating periods T.
The third innovative feature is a fundamental change in current supply-offer formulations. Under the proposed design, each Dispatchable Power Resource (DPR) participating in a reserve market M(T) as a potential reserve supplier for a future operating period T is permitted to submit its supply offer in a 2-part pricing swing-contract form. The “swing” (flexibility) in these DPR-submitted supply offers facilitates the RTO/ISO’s ability to dispatch power during T to meet just-in-time customer power demands and grid reliability requirements. The two-part pricing form of these DPR-submitted supply offers permits each DPR to ensure its revenue sufficiency.

The fourth innovative feature is a fundamental change in settlement processes. Under the proposed design, these processes are based on fundamental pricing, risk-sharing and cost-allocation principles. For example, in accordance with FERC Order 755 [9], no performance payments are paid to market participants in advance of verified real-time performance.

The incorporation of these four features into current RTO/ISO-managed wholesale power markets would require changes in product definitions, management practices, supply-offer formulations, and settlement processes, but not in real-time operations. Consequently, as illustrated in [45, Ch. 16], these changes could be introduced gradually, without disruption of real-time operations.

### 7.4 Expanded Discussion of Innovative Design Features

The previous section briefly summarizes four innovative features characterizing the Linked Swing-Contract Market Design [45] involving product definitions, management practices, supply-offer formulations, and settlement processes. This section provides a more careful discussion of these features.\(^{43}\)

The Linked Swing-Contract Market Design posits an RTO/ISO-managed collection of linked forward bid/offer-based reserve markets M(T) for future operating periods T. The look-ahead horizon LAH(T) between the close of M(T) and the start of T can range in duration from years to seconds, as can the duration of T itself.

Reserve for T consists of dispatchable power-paths offered for possible RTO/ISO dispatch during T. As defined in Section 5.2.2, a power-path for T is a sequence \( p_b = (p_b(t) | t \in T) \) of injections and/or withdrawals of power \( p_b(t) \) (MW) that take place at a single designated grid location b during T. See, for example, the power-paths depicted in Figs. 6–8.

This nodal power-path conceptualization focuses attention on actions and events that take place over time at specific grid localities. It thus permits market operations and contractual bid/offer forms to reflect the distributed vantage points of market participants as well as the centralized vantage point of the RTO/ISO, which facilitates incentive alignment.

\(^{43}\) A detailed analytical formulation for the Linked Swing-Contract Market Design can be found in [45, Chs. 4-10].
Core types of participants in a reserve market M(T) for a future operating period T include Load Serving Entities (LSEs) managing power usage for collections of power consumers and Intermittent Power Resources (IPRs) whose generation is not firmed by storage. LSEs submit reserve bids for T; a reserve bid submitted into M(T) by an LSE can include both a fixed (non-dispatched must-service) amount of power withdrawal for T, without accompanying information regarding willingness to pay, and a price-sensitive demand schedule for power-path delivery during T. IPRs submit forecasts for their fixed (non-dispatched must-service) amount of power injection at specific grid locations during T without accompanying information regarding required compensation.

In addition, the core types of participants in a reserve market M(T) for a future operating period T include Dispatchable Power Resources (DPRs). Examples include: dispatchable generators; dispatchable loads providing power absorption as an ancillary service; dispatchable storage facilities providing down/up power as an ancillary service; and dispatchable renewable power facilities (e.g., hydropower plants, wind farms, and PV solar arrays whose power is firmed by storage).

A reserve offer SC\(m(T)\) submitted by a DPR \(m\) into a reserve market M(T) for a future operating period T is a two-part pricing swing contract that offers a collection of dispatchable power-paths for possible RTO/ISO dispatch during T in return for appropriate compensation. More precisely, SC\(m(T)\) consists of four \(m\)-specified components:

- a offer price \(\alpha_m(T)\) in lump-sum or amortized form permitting \(m\) to receive compensation ex ante (before T) for any avoidable fixed cost that \(m\) must incur to ensure the availability of \(m\)’s offered reserve for possible RTO/ISO-dispatched delivery during T;
- a set \(\mathbb{T}_m(T)\) of contract exercise times available to the RTO/ISO between the close of M(T) and the start of T;
- a production-possibility set \(\mathbb{P}_m(T)\) that conveys the degree of flexibility (“swing”) in the physical attributes of the reserve (dispatchable power-paths) that \(m\) is offering to make available for possible RTO/ISO dispatch during T;
- a performance payment method \(\phi_m(T)\) permitting \(m\) to receive compensation ex post (after T) for any variable cost that \(m\) incurs for verified period-T delivery of a power-path in \(\mathbb{P}_m(T)\) in accordance with RTO/ISO dispatch instructions.

Thus, the reserve offer SC\(m(T)\) permits DPR \(m\) to ensure its revenue sufficiency. That is, \(m\) can specify the offer price \(\alpha_m(T)\) and performance payment method \(\phi_m(T)\) appearing in SC\(m(T)\) to ensure that the revenue it attains from participation in M(T) suffices to cover all of the avoidable cost (i.e., avoidable fixed cost plus variable cost) that \(m\) incurs due to its decision to participate in M(T).

The RTO/ISO conducts a contract-clearing optimization for each reserve market M(T). The goal of this optimization is to maximize expected total net benefit (i.e., total benefit minus total avoidable cost) for M(T) participants. This optimization is

\[\text{maximize } E\{\text{total net benefit}\} = \text{total benefit} - \text{total avoidable cost}\]

44 As in current RTO/ISO-managed wholesale power markets, these LSE fixed demands are automatically incorporated as nodal must-serve loads in the nodal power-balance constraints for T.
conditioned on current state conditions and subject to power balance, line capacity limits, and other standard types of system constraints for operating period T.

To maintain the independence of the RTO/ISO, all costs incurred for reserve allocation are allocated back to market participants. More precisely, the cost of procuring reserve for an operating period T, net of LSE payments for cleared price-sensitive demand bids for T, is allocated across M(T) participants ex post (after T) based on their contribution to the anticipated volatility and realized size of these costs. This cost allocation determines principled functional relationships among:

- the prices paid by LSEs for power usage during T in accordance with their cleared price-sensitive power-usage demands for T;
- the costs charged to LSEs for power usage during T that result from their fixed power-usage demands for T;
- the offer-price payments and performance payments received by DPRs with cleared supply offers for T;
- IPR payments (or charges) resulting from their fixed power injections during T.

Regarding the last-listed cost-allocation category, IPRs receive payments (negative charges) to the extent their fixed power injections during T reduce the RTO/ISO’s need to dispatch cleared DPR reserve offers during T, hence reduce the RTO/ISO’s need to compensate dispatched DPRs for incurred variable costs. On the other hand, IPRs whose forecasted fixed power injections for T increase the anticipated volatility of period-T net fixed load – hence period-T volumetric grid risk – are charged a corresponding portion of the RTO/ISO’s reserve acquisition costs for T.

IPRs thus face revenue insufficiency risk; the costs they are charged for period-T reserve procurement can exceed the payments they receive for period-T reductions in DPR variable cost compensation. This risk-exposure is deliberate; it gives IPRs an incentive to firm their power injections with storage, thus enabling their participation in reserve markets M(T) as DPRs able to ensure their revenue sufficiency through proper formulation of the offer prices and performance payment methods they include in their submitted supply offers.

The durations of the look-ahead horizons LAH(T), LAH*(T), LAH**(T),... for the RTO/ISO-managed forward reserve markets M(T), M*(T), M**(T),... that pertain to a single designated future operating period T can range from multiple years to minutes. Linkages among these markets for T arise because the bids and offers cleared for T in each successive market affect the perceived desirability or need to clear additional bids and offers for T in the next. The RTO/ISO carries forward

---

45 A system operator for a grid-supported wholesale power market is said to be independent if the operator has no financial or ownership stake in the operations of this market, either in the market outcomes determined for market participants or in the physical operations of the grid per se.

46 More precisely, cost allocations under the Linked Swing-Contract Market Design are determined in accordance with the RTO/ISO Cost Allocation Rule defined in [45, Sec. 6.7] whose form reflects three principles: (i) participants ought to be paid or charged based on their positive or negative contribution to total net surplus (Shapley value principle for multi-player games); (ii) costs incurred to protect against risk should be charged to those giving rise to this risk; and (iii) no performance payments should be made in advance of actual verified performance (FERC Order 755 [9]).
through time a continually updated portfolio of the cleared bids and offers for T in preparation for real-time period-T operations.

In order to encourage appropriate entry of new DPRs, the forward reserve markets pertaining to distant future operating periods T can include Long-Term Forward Markets LTFM(T) whose look-ahead horizons have multiple-year durations. Bids and offers cleared in LTFM(T) can be adjusted in subsequent years-ahead, months-ahead, days-ahead, and/or intra-day reserve markets for T, and in supplementary unit-commitment processes that are held prior to T.

However, the credible specification of system constraints for LTFM(T) could be extremely difficult due to deep uncertainty arising from its long look-ahead horizon. Thus, the RTO/ISO tasked with managing LTFM(T) might need to supplement any LSE-submitted demand bids for T with demand bids based on its own long-term assessment of power-usage needs for T in order to provide appropriate signals to potential reserve suppliers regarding the types and amounts of reserve to be sought for T. This could help to reduce a potential reserve supplier’s perceived risk of not being cleared for LTFM(T) due to uncertain reserve needs, regardless of any supply-cost advantages it might possess.

On the other hand, the two-part pricing form of the reserve offers that DPRs are permitted to submit into LTFM(T) allows these resources to assure their revenue sufficiency through appropriate specification of the offer prices and performance payment methods they include in their supply offers for LTFM(T), independently of the length of its look-ahead horizon. Also, since all performance payments are paid ex post (after T) contingent on verified period-T performance, a DPR can make its required performance payments contingent on actual period-T market conditions (e.g., fuel prices) tailored to its own unique operational attributes.

In summary, under the proposed Linked Swing-Contract Market Design [45] there is no need for separate or co-optimized energy markets to ensure net load balancing. There is also no need to rely on conceptually-problematic participation categories and flexibility products to ensure reserve adequacy. Finally, there is no need to rely on out-of-market (OOM) make-whole payments, or on conceptually-problematic settlement methods such as extended-LMP, in order to ensure supplier revenue sufficiency.

7.5 Legacy Core Design Comparisons

Detailed comparisons between the Linked Swing-Contract Market Design and the current design of U.S. RTO/ISO-managed wholesale power markets are provided in [45]. This section illustrates key similarities and differences in these designs by focusing on one specific form of market: namely, a Day-Ahead Market (DAM).

More precisely, this section provides high-level comparisons of the basic design features and market-clearing optimization formulations for current U.S. DAMs and the proposed form of a DAM under the Linked Swing-Contract Market Design, referred to below as an SC DAM. For simplicity, all customer demand in these DAMs is assumed to take a fixed (non-dispatched must-service) form.
Figure 9 provides high-level comparisons of the basic market design features for SC DAMs and current U.S. DAMs. The main similarities are:

- Both DAM designs are RTO/ISO-managed markets based on bids and offers;
- Both DAM designs have the same types of market participants;
- Both DAM designs are subject to standard types of system constraints, such as nodal power balance constraints, nodal and/or zonal reserve requirements, and transmission-line capacity limits.

The main differences, listed below, involve product definition, supply-offer forms, and settlement rules, not real-time operations. Thus, as illustrated in [45, Ch. 16], these differences could be introduced gradually into current U.S. DAMs.

- SC DAMs are forward markets for reserve, i.e., for offered availability of diverse power-path production capabilities for possible next-day RTO/ISO dispatch;
- SC DAM supply offers for reserve are two-part pricing swing contracts in either firm or option form;
- A DPR $m$ participating in an SC DAM held on day D whose supply offer

$$SC_m(D+1) = (\alpha_m(D+1), T_{m}^D(D+1), PP_m(D+1), \phi_m(D+1))$$

(24)
Leigh Tesfatsion

is cleared for day D+1 receives as payment its offer price $\alpha_m(D+1)$ (in lump-sum or amortized form) as compensation for the reduction of volumetric grid-risk this cleared supply offer provides for day D+1 operations.\(^{47}\)

– A DPR $m$ participating in an SC DAM for day D+1 can ensure its revenue sufficiency (i.e., revenue $\geq$ avoidable cost) by appropriate specification of the offer-price $\alpha_m(D+1)$ and performance payment method $\phi_m(D+1)$ that $m$ includes in its supply offer (24); no resort to OOM make-whole payments is needed.

– For an SC DAM, no performance payment occurs in advance of actual verified performance (dispatched power-path delivery) during next-day operations;

Figure 10 provides high-level comparisons of the market-clearing optimization formulations for current U.S. DAMs and SC DAMs.

<table>
<thead>
<tr>
<th>Differences</th>
<th>Current DAM SCUC</th>
<th>Current DAM SCED</th>
<th>SC DAM Optimization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objective</td>
<td>Min [startup/shut down costs + spoiled costs + dispatch costs + reserve costs + constraint penalties]</td>
<td>Min [Dispatch costs + reserve costs + constraint penalties]</td>
<td>Min [Availability cost + performance cost + constraint penalties]</td>
</tr>
<tr>
<td>Unit commitment constraints</td>
<td>Yes</td>
<td>No</td>
<td>Each DPR includes its unit commitment constraints in its submitted swing contract</td>
</tr>
<tr>
<td>Key RTO/ISO decision variables</td>
<td>Unit commitments</td>
<td>Energy dispatch &amp; reserve levels</td>
<td>Which swing contracts are cleared</td>
</tr>
<tr>
<td>Settlement</td>
<td>No</td>
<td>LMPs calculated as SCED dual variables</td>
<td>Each cleared DPR receives the offer price it has included in its submitted swing contract</td>
</tr>
</tbody>
</table>

Fig. 10 Optimization comparison: U.S. DAM SCUC/SCED vs. SC DAM Contract Clearing.

The main similarities in these optimization formulations are as follows:

– Each DAM optimization can be formulated as a Mixed Integer Linear Programming (MILP) problem;

– The objective function of each DAM optimization includes penalties for system constraint violations.

\(^{47}\) The amount of this risk reduction for day D+1 depends on the swing (flexibility) inherent in the specific collection of dispatchable power-paths that $m$ includes in its day-D offered production possibility set $\mathbb{PP}_m(D+1)$ in (24). However, the amount of this risk reduction does not depend on which specific power-path the RTO/ISO chooses to dispatch from $\mathbb{PP}_m(D+1)$ during day D+1.
The main differences in these optimization formulations are as follows. For an SC DAM conducted on day D for an operating-day D+1:

– The objective function for the SC DAM contract-clearing optimization fully incorporates the availability cost and performance cost of each participating DPR $m$, as conveyed to the RTO/ISO by the offer price $a_m(D+1)$ and the performance payment method $\phi_m(D+1)$ that $m$ includes in its submitted swing-contract supply offer $SC_m(D+1)$ depicted in (24);

– Each participating DPR $m$ can ensure full coverage of its avoidable fixed cost by appropriate specification of the offer price $a_m(D+1)$ it includes in (24);

– Each participating DPR $m$ can ensure coverage of any variable cost it might incur for dispatched power-path delivery during D+1 by appropriate specification of the performance payment method $\phi_m(D+1)$ it includes in (24);

– the only payment obligation determined by the day-D SC DAM that is settled in advance of operating-day D+1 is payment of offer prices, hence coverage of avoidable fixed costs incurred by DPRs with day-D cleared supply offers to ensure the availability of their offered reserve for possible RTO/ISO dispatch during day D+1;

– no performance payments are made to cleared DPRs in advance of verified actual performance (power-path delivery) on day D+1 that is undertaken in accordance with RTO/ISO dispatch instructions;

– “unit commitments” are replaced by “cleared contracts,” and the only binary-valued RTO/ISO decision variables are yes/no contract-clearing indicators;

– the “unit-commitment constraints” of each participating DPR $m$ are conveyed to the RTO/ISO through the particular attributes of the power-path production possibility set $PP_m(D+1)$ that $m$ includes in its submitted supply offer (24).

8 Conclusion: Grids as Flexibility-Support Mechanisms

The purpose of this study is to contribute to the development of a conceptually-consistent market design for U.S. RTO/ISO-managed wholesale power markets that is capable of supporting the smooth transition of these markets to greater reliance on renewable power and more active participation by demand-side resources. An essential consideration is whether achievement of this goal requires fundamental changes in the existing design of these markets.

This study concludes that fundamental design changes are indeed needed. Multiple conceptually-problematic aspects of the legacy DAM/RTM Two-Settlement System at the core of all seven current U.S. RTO/ISO-managed wholesale power markets are carefully highlighted, analyzed, and illustrated in Sections 2–6. These aspects did not prevent feasible work-arounds in earlier years, when net loads were relatively stable and predictable. They are causing increasingly serious issues now as these markets attempt to transition to decarbonized grid operations with greater dependence on renewable power and more diverse participants.
Section 7 thus asks what else can be done. Given the increasing volatility and uncertainty of net loads, it is argued that grid-supported U.S. RTO/ISO-managed wholesale power markets must be reorganized to permit them to operate, efficiently and reliably, as flexibility-support insurance mechanisms well-aligned with the local goals and constraints of market participants.

The needed insurance protection is protection against volumetric grid-risk, the systemic risk of grid collapse due to physical imbalance between net load and centrally-dispatched power injection at one or more grid locations. The needed flexibility-support is support for the continual just-in-time power needs of increasingly diverse and numerous market participants distributed across the grid.

Section 7 reviews an alternative market design proposed for these markets that has specifically been tailored to meet these needs: the Linked Swing-Contract Market Design [45]. This proposed design has been developed and tested at design readiness levels DRL-1 (conceptual formulation) through DRL-4 (performance testing using moderate-scale computational platforms) in a series of studies. This design, along with other proposed market designs at early design-readiness levels, is reviewed in a recent Resources for the Future (RFF) report [32].

What would be needed to increase the design readiness level of the Linked Swing-Contract Market Design from DRL-4 all the way to DRL-9 (real-world implementation)?

Perhaps the greatest perceived practical drawback of the design is its requirement that dispatchable power resources submit reserve supply offers (24) into reserve markets $M(T)$ for future operating periods $T$ that include production-possibility sets $PP(T)$ providing physical descriptions of their offered dispatchable power-path production capabilities for $T$. Here is what I said about this challenging goal in my 2021 book [45, Sec. 5.1, p. 35]:

“How (this goal) is met in practice is a function of current capabilities. One can imagine, for example, that future capabilities might permit a power-path $p$ to be represented as a linear combination of basis functions, as a calculus-of-variation perturbation, in pictorial or bit form rather than analytical expression, or in ways that today cannot even be conceived.”

Today there are indeed exciting possibilities for meeting this goal that were not conceived back in 2021: namely, the digital twinning of these production-possibility sets. As formally defined in [7, p. 1042], a digital twin (DT) is a mathematical model with an updating mechanism that generates data that are indistinguishable from its physical counterpart. Less formally, here is how a DT was characterized in a 2024 virtual symposium sponsored by the National Academies of Sciences:

“A digital twin couples a real-world facility (and the data available from physical sensors and other building systems) with a computational model of the facility that is dynamically updated as conditions change.” (NAS Symposium, Digital Twins in the Built Environment, 13 June 2024)

---

48 Specifically, see [1], [20], [28], [29], [27], [33], [44], [45], and [47].
However, the representation of power-path production possibility sets also raises fundamental conceptual considerations: namely, what is the true physical nature of a “power-path” in relation to current procedures for the measured determination of power levels? More broadly, how must the conception of a power-path be generalized to encompass the interconnected physical reality of active and reactive power as described by Schwepppe et al. [40, Appendices] and von Meier [51]?

In a 2022 study, Kirkham et al. [24] attribute critical operational issues arising for grid-supported electric power systems to fundamental many-to-one measurement problems for active power, reactive power, apparent power, and power factor. Specifically, the definitions for these concepts – currently expressed in static Root-Mean-Square (RMS) time-averaged forms for a conventionally designated operating period T – can in fact correspond to multiple underlying dynamic realities with important effects on the resulting period-T physical operation of these systems.

The present study focuses on economic fundamentals for grid-supported U.S. RTO/ISO-managed wholesale power markets. The main finding is that the core DAM/RTM Two-Settlement System underpinning the operation of each of these markets is conceptually problematic due to a serious many-to-one measurement problem for benefit/cost valuations. A companion study [50] focusing more narrowly on locational marginal pricing (LMP) reports a similar troublesome finding for the reliance of these markets on LMP settlements.

A key concern is that these economic many-to-one measurement problems are conjoined with the physical many-to-one measurement problems identified by [24] in the optimal power flow (OPF) formulations commonly used for DAM/RTM SCED optimizations. Hence, they are also conjoined in the implementations of the DAM/RTM Two-Settlement Systems underpinning current U.S. RTO/ISO-managed wholesale power market operations. These serious economic and physical many-to-one measurement problems must be carefully addressed by power system researchers as they strive to improve the design and operation of these markets.

In conclusion, many critical conceptual and operational issues confront power system researchers attempting to redesign U.S. RTO/ISO-managed wholesale power markets to enable them to operate reliably and efficiently as they transition to lower-carbon grid operations with increasingly diverse participants. This is no time to be intellectually or emotionally wedded to what has worked in the past; the current needs for redesign are too pressing.

However, any redesign must in turn be flexible and open to further adaptation. The ultimate goal, surely, must be robust wholesale power market design for transacting in a deeply-uncertain continually-evolving world.
### A.1 Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>Alternating Current</td>
</tr>
<tr>
<td>CAISO</td>
<td>California Independent System Operator</td>
</tr>
<tr>
<td>CFD</td>
<td>Contract-For-Difference</td>
</tr>
<tr>
<td>CCSM</td>
<td>Competitive Commodity Spot Market</td>
</tr>
<tr>
<td>CSM</td>
<td>Commodity Spot Market</td>
</tr>
<tr>
<td>D</td>
<td>Commonly used acronym for a day</td>
</tr>
<tr>
<td>DAM</td>
<td>Day-Ahead Market</td>
</tr>
<tr>
<td>DC</td>
<td>Direct Current</td>
</tr>
<tr>
<td>DM</td>
<td>Decision-Maker</td>
</tr>
<tr>
<td>DPR</td>
<td>Dispatchable Power Resource</td>
</tr>
<tr>
<td>ERCOT</td>
<td>Electric Reliability Council of Texas</td>
</tr>
<tr>
<td>FERC</td>
<td>Federal Energy Regulatory Commission</td>
</tr>
<tr>
<td>FTR</td>
<td>Financial Transmission Right</td>
</tr>
<tr>
<td>h</td>
<td>SI metric symbol for an hour (60s)</td>
</tr>
<tr>
<td>H</td>
<td>Commonly used acronym for an hour</td>
</tr>
<tr>
<td>IPR</td>
<td>Intermittent Power Resource</td>
</tr>
<tr>
<td>ISO</td>
<td>Independent System Operator</td>
</tr>
<tr>
<td>ISO-NE</td>
<td>Independent System Operator for New England</td>
</tr>
<tr>
<td>kW</td>
<td>SI metric symbol for a kilowatt (1000 W)</td>
</tr>
<tr>
<td>kWh</td>
<td>SI metric symbol for a kilowatt-hour (1000 Wh)</td>
</tr>
<tr>
<td>kVA</td>
<td>SI metric symbol for kilovolt-Amperes (1000 Volt-Amperes)</td>
</tr>
<tr>
<td>LMP</td>
<td>Locational Marginal Price (or Locational Marginal Pricing)</td>
</tr>
<tr>
<td>LSE</td>
<td>Load-Serving Entity</td>
</tr>
<tr>
<td>MILP</td>
<td>Mixed Integer Linear Programming</td>
</tr>
<tr>
<td>MISO</td>
<td>Midcontinent Independent System Operator</td>
</tr>
<tr>
<td>MW</td>
<td>SI metric symbol for a megawatt (1000 kW)</td>
</tr>
<tr>
<td>MWh</td>
<td>SI metric symbol for a megawatt-hour (1000 kWh)</td>
</tr>
<tr>
<td>NOPR</td>
<td>Notice Of Proposed Rule-making (FERC)</td>
</tr>
<tr>
<td>NYISO</td>
<td>Independent System Operator for New York</td>
</tr>
<tr>
<td>OOM</td>
<td>Out-of-Market</td>
</tr>
<tr>
<td>OPF</td>
<td>Optimal Power Flow</td>
</tr>
<tr>
<td>PJM</td>
<td>PJM Interconnection</td>
</tr>
<tr>
<td>QSE</td>
<td>Qualified Scheduling Entity (ERCOT)</td>
</tr>
<tr>
<td>RTM</td>
<td>Real-Time Market</td>
</tr>
<tr>
<td>RTO</td>
<td>Regional Transmission Organization</td>
</tr>
<tr>
<td>SCED</td>
<td>Security-Constrained Economic Dispatch</td>
</tr>
<tr>
<td>SCUC</td>
<td>Security-Constrained Unit Commitment</td>
</tr>
<tr>
<td>SI</td>
<td>Standard International (metric system)</td>
</tr>
<tr>
<td>SPP</td>
<td>Southwest Power Pool</td>
</tr>
<tr>
<td>TNS</td>
<td>Total Net Surplus</td>
</tr>
<tr>
<td>W</td>
<td>SI metric symbol for a watt</td>
</tr>
<tr>
<td>Wh</td>
<td>SI metric symbol for a watt-hour</td>
</tr>
</tbody>
</table>
# A.2 Standard Transmission System Terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ancillary service</td>
<td>Service that supports system reliability</td>
</tr>
<tr>
<td>Commitment</td>
<td>Scheduling of a dispatchable power resource for possible future central dispatch</td>
</tr>
<tr>
<td>Dispatch</td>
<td>Signaling a grid-connected power resource to inject/withdraw power</td>
</tr>
<tr>
<td>Energy</td>
<td>Abbreviation for electric energy (MWh)</td>
</tr>
<tr>
<td>Energy loads</td>
<td>Devices needing a certain amount of energy over an operating-period T, but indifferent with regard to exact timing of this energy provision during T.</td>
</tr>
<tr>
<td>Fixed power injection</td>
<td>Non-dispatched must-service power injection into a grid</td>
</tr>
<tr>
<td>Fixed load</td>
<td>Non-dispatched must-service power withdrawal from a grid</td>
</tr>
<tr>
<td>Generation</td>
<td>Production of power either for local behind-the-meter use or for grid injection</td>
</tr>
<tr>
<td>Grid-delivered energy</td>
<td>Energy (MWh) delivered at a location via accumulation of a power-path</td>
</tr>
<tr>
<td>Intermittent power</td>
<td>Power injections/withdrawals not fully under central-dispatchable control</td>
</tr>
<tr>
<td>Intermittent power resource</td>
<td>Grid-connected non-mediated source of intermittent power</td>
</tr>
<tr>
<td>Load</td>
<td>Commonly-used synonym for power withdrawal from a grid; technically, a grid device or grid component to which power is delivered</td>
</tr>
<tr>
<td>Locational marginal price</td>
<td>Energy price conditional on delivery location and operating period</td>
</tr>
<tr>
<td>Make-whole payment</td>
<td>OOM compensation for a market-incurred cost</td>
</tr>
<tr>
<td>Merit-order dispatch</td>
<td>Dispatch in accordance with net benefit contribution</td>
</tr>
<tr>
<td>Must-service</td>
<td>Power withdrawal (injection) that must be balanced by power injection (withdrawal) under normal grid operating conditions</td>
</tr>
<tr>
<td>Net load</td>
<td>Load &amp; inadvertent power loss minus non-dispatched power injection</td>
</tr>
<tr>
<td>Net fixed load</td>
<td>Fixed load minus fixed power injection</td>
</tr>
<tr>
<td>Net reserve cost</td>
<td>Reserve procurement cost minus reserve revenue receipts</td>
</tr>
<tr>
<td>Non-dispatchable power</td>
<td>Power not under RTO/ISO-dispatchable control</td>
</tr>
<tr>
<td>Operating reserve</td>
<td>Generation capacity (MW) unencumbered by energy delivery obligations</td>
</tr>
<tr>
<td>Performance</td>
<td>Delivery of a good or service in response to RTO/ISO-communicated instructions</td>
</tr>
<tr>
<td>Performance cost</td>
<td>Variable cost incurred for providing delivery of a good or service in response to RTO/ISO-communicated instructions</td>
</tr>
<tr>
<td>Power</td>
<td>Abbreviation for electric power (MW)</td>
</tr>
<tr>
<td>Power absorption</td>
<td>Incremental down/up changes in power withdrawal offered into a power system as an ancillary service</td>
</tr>
<tr>
<td>Power imbalance</td>
<td>Discrepancy between grid power injection &amp; grid power withdrawal/loss</td>
</tr>
<tr>
<td>Power injection</td>
<td>Insertion of power into a grid at an electrical point-of-connection</td>
</tr>
<tr>
<td>Power loads</td>
<td>Devices needing power at specific times to fulfill their functions or purposes</td>
</tr>
<tr>
<td>Power-path</td>
<td>Sequence of injections and/or withdrawals of power (MW) at a single grid location during a designated time-interval</td>
</tr>
<tr>
<td>Power-path delivery</td>
<td>Power-path implemented at a designated grid location during a designated time-interval in accordance with central-dispatch instructions</td>
</tr>
<tr>
<td>Power usage</td>
<td>Use of power as an intermediate good to further some end</td>
</tr>
<tr>
<td>Power withdrawal</td>
<td>Extraction of power from a grid at an electrical point-of-connection</td>
</tr>
<tr>
<td>Reserve</td>
<td>Service or product-provision capability that could be used to support grid reliability</td>
</tr>
<tr>
<td>Reserve bid</td>
<td>Contract requesting reserve availability</td>
</tr>
<tr>
<td>Reserve offer</td>
<td>Contract offering reserve availability</td>
</tr>
<tr>
<td>Transmission service cost</td>
<td>Variable cost incurred for grid operation and maintenance</td>
</tr>
<tr>
<td>Uplift payment</td>
<td>OOM compensation for required OOM action to maintain grid reliability</td>
</tr>
</tbody>
</table>
### A.3 Standard Economic Terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asset</td>
<td>Anything of durable value, whether physical or financial</td>
</tr>
<tr>
<td>Avoidable cost</td>
<td>Cost that can be avoided by not committing to undertake a specified type of action</td>
</tr>
<tr>
<td>Avoidable fixed cost</td>
<td>Avoidable cost not dependent on exact form of action as long as it has specified type</td>
</tr>
<tr>
<td>Benefit (or utility) function</td>
<td>Function measuring the increase in own-welfare attained by a customer from the consumption and/or use of goods and/or services</td>
</tr>
<tr>
<td>Commodity</td>
<td>Asset $Q$ with a standard unit of measurement $u$ such that, at any given location and time, $Q$-traders consider all available $Q$-units $u$ to be perfect substitutes</td>
</tr>
<tr>
<td>Competitive market</td>
<td>Commodity market whose buyers and suppliers are price-takers</td>
</tr>
<tr>
<td>Competitive equilibrium</td>
<td>Competitive market price-quantity outcome s.t. aggregate demand=aggregate supply</td>
</tr>
<tr>
<td>Consumer</td>
<td>Purchaser of goods/services for direct own consumption/use (no resale)</td>
</tr>
<tr>
<td>Contract in firm form</td>
<td>Non-contingent contract whose terms are binding on all parties</td>
</tr>
<tr>
<td>Contract in option form</td>
<td>Holder has the right, but not the obligation, to exercise the contract</td>
</tr>
<tr>
<td>Demand schedule (inverse)</td>
<td>Schedule expressing the maximum $Q$-unit price a buyer is willing to pay for each additionally demanded unit of a commodity $Q$</td>
</tr>
<tr>
<td>Demand schedule (ordinary)</td>
<td>Schedule expressing the maximum amount of a commodity $Q$ that a buyer is willing to purchase at each successively-higher $Q$-unit price</td>
</tr>
<tr>
<td>Efficiency</td>
<td>No wastage of resources</td>
</tr>
<tr>
<td>Fixed cost</td>
<td>Cost not dependent on a specific form of action undertaken</td>
</tr>
<tr>
<td>Forward market</td>
<td>Transacted amounts and payment obligations for these transacted amounts occur in advance of the delivery of these transacted amounts</td>
</tr>
<tr>
<td>Futures market</td>
<td>Forward market for a commodity</td>
</tr>
<tr>
<td>Good</td>
<td>Exchangeable physical item whose acquisition provides benefit to the procurer</td>
</tr>
<tr>
<td>Hedonic pricing</td>
<td>Pricing of a product by means of prices separately assigned to its intrinsic physical attributes and/or its external circumstances</td>
</tr>
<tr>
<td>Joint products</td>
<td>Products jointly produced from a given set of inputs</td>
</tr>
<tr>
<td>Law of One Price</td>
<td>In the absence of trade frictions (e.g., differences in trade locations, trade times, and/or trader product information), trader exploitation of arbitrage opportunities will ensure that every unit of a commodity available for purchase (sale) has the same purchase (sale) price,</td>
</tr>
<tr>
<td>Marked efficiency</td>
<td>No wastage of opportunity to increase total net surplus for buyer and supplier participants</td>
</tr>
<tr>
<td>Net benefit</td>
<td>Benefit minus avoidable cost</td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Net buyer surplus</td>
<td>Difference between the maximum payment a buyer is willing to make to purchase an item $z$ and the actual payment the buyer makes to purchase $z$</td>
</tr>
<tr>
<td>Net supplier surplus</td>
<td>Difference between the actual payment a supplier receives for the sale of an item $z$ and the minimum payment the supplier is willing to receive for the sale of $z$</td>
</tr>
<tr>
<td>Opportunity cost</td>
<td>Earnings foregone by not committing assets to an alternative next-best use</td>
</tr>
<tr>
<td>Pareto efficiency</td>
<td>No wastage of opportunity to increase benefit for some at no cost to others by means of a feasible reallocation of resources</td>
</tr>
<tr>
<td>Perfect substitutes</td>
<td>Two items are perfect substitutes (or economically equivalent) for a trader at a given location and time if substitution of either item for the other item does not affect the trader’s economic valuation of this item.</td>
</tr>
<tr>
<td>Price-taker</td>
<td>Trader participating in a market for a good or service who behaves as if his own market transactions cannot affect the market price of this good or service.</td>
</tr>
<tr>
<td>Product</td>
<td>Outcome of a production process</td>
</tr>
<tr>
<td>Production process</td>
<td>Process that transforms inputs into one or more outputs</td>
</tr>
<tr>
<td>Productive efficiency</td>
<td>No physical wastage of production inputs and/or production outputs</td>
</tr>
<tr>
<td>Purchase reservation value</td>
<td>Maximum payment a buyer is willing to make to procure a designated item</td>
</tr>
<tr>
<td>Revenue sufficiency</td>
<td>Supplier revenue is sufficient to cover supplier avoidable cost</td>
</tr>
<tr>
<td>Risk</td>
<td>Possibility of an adverse deviation from an expected outcome</td>
</tr>
<tr>
<td>Risk aversion</td>
<td>Unwillingness to participate in a risky undertaking with zero expected payoff</td>
</tr>
<tr>
<td>Sale reservation value</td>
<td>Minimum payment a supplier is willing to accept to supply a designated item</td>
</tr>
<tr>
<td>Service</td>
<td>Action taken by an entity that provides benefit to another entity</td>
</tr>
<tr>
<td>Spot market</td>
<td>Transacted amounts, payments for these transacted amounts, and deliveries of these transacted amounts all occur at the same location and time (“on the spot”).</td>
</tr>
<tr>
<td>Strategic market advantage</td>
<td>Unintended opportunity for a participant to exploit market rules to gain advantage.</td>
</tr>
<tr>
<td>Structural market advantage</td>
<td>Instituted market feature that systematically favors some participants over others</td>
</tr>
<tr>
<td>Sunk cost</td>
<td>Non-avoidable fixed cost</td>
</tr>
<tr>
<td>Supply offer</td>
<td>Offer to provide an item or service</td>
</tr>
<tr>
<td>Supply schedule (inverse)</td>
<td>Schedule expressing the minimum $Q$-unit price a supplier is willing to accept in payment for each additionally supplied unit of a commodity $Q$</td>
</tr>
<tr>
<td>Supply schedule (ordinary)</td>
<td>Schedule expressing the maximum amount of a commodity $Q$ that a supplier is willing to sell at each successively-higher $Q$-unit price</td>
</tr>
<tr>
<td>Systemic risk</td>
<td>System-wide risk, i.e., correlated risk arising for system operations as a whole</td>
</tr>
<tr>
<td>Transaction cost</td>
<td>Avoidable fixed cost incurred to organize a production process</td>
</tr>
<tr>
<td>Two-part pricing</td>
<td>Separately-requested compensation for avoidable fixed cost and variable cost</td>
</tr>
<tr>
<td>$u$-asset</td>
<td>An asset with a standard unit of measurement $u$</td>
</tr>
<tr>
<td>Variable cost</td>
<td>Avoidable cost dependent on specific form of an undertaken action (e.g., production level)</td>
</tr>
<tr>
<td>Volumetric grid risk</td>
<td>Systemic risk arising for a grid due to possible net load imbalance</td>
</tr>
</tbody>
</table>
A.4 Cost Types for Grid-Supported RTO/ISO-Managed Wholesale Power Markets: Empirical Examples

Types of Avoidable Fixed Cost:

1. **Capital Investment Cost.** Land acquisition, building construction; equipment purchases. Financed by *internal financing* (i.e., funds on hand), or by *external financing* taking two possible forms:
   - **Direct Financing:** Sell *newly issued* securities in primary security markets to lenders willing to invest in risky assets (i.e., assets with chance of loss) that also offer a sufficiently high chance of gain;
   - **Indirect Financing:** Obtain loans from financial intermediaries, typically secured by some form of collateral, that then result in amortized streams of payment obligations.

2. **Transaction Cost.** Insurance, building code compliance, licensing fees, employee search. Transaction costs are typically financed by internal financing.

3. **Opportunity Cost.** Expected net earnings from a best possible alternative use of assets, e.g., use of generation units directly (behind the meter) for local purposes.

4. **Unit Commitment Cost.** Start-up, no-load, minimum-run, and/or shut-down cost that are incurred for ensuring the availability of power-paths for possible RTO/ISO dispatched delivery during a future operating period but are not dependent on the specific form (if any) of this delivered power-path.

Types of Variable Cost:

1. **Fuel Cost.** Charges for pulverized coal, natural gas, nuclear, petroleum, and/or refuse-derived fuels as inputs to power production.

2. **Labor Cost.** Salaries/wages for: legal/tax advice; advertisement; planning; supervision; trading-desk operations; maintenance; and repair.

3. **Intermediate Good (Supply-Chain) Cost.** Rail/barge/pipeline/truck transport charges for fuel deliveries; replenishment of used-up supplies.

4. **Equipment/Software Rental Cost.** Rental charges for office equipment, cars, and software licenses.

5. **Depreciation of Owned Machinery.** Generation unit wear-and-tear due to start-up, normal, and/or shut-down ramping required to follow RTO/ISO-signal dispatch set-points during successive operating periods.

6. **Assessed Charges for Transmission Services.** Transmission grid operation and maintenance (O&M) costs allocated across market participants.

7. **Variable-Cost Offsets from Sales of Valuable Bi-Products.** Revenue offset to variable cost of a product due to joint production, e.g., co-generation of valuable heating services along with power by Combined Heat and Power (CHP) units.

8. **Disposal Cost for Waste Bi-Products.** Cost incurred by power plants (e.g., nuclear) to dispose of solid-waste output resulting from plant operations.
## A.5 Swing-Contract Market Terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Acronyms &amp; Generics:</strong></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>Generic symbol for a day</td>
</tr>
<tr>
<td>DPR</td>
<td>Dispatchable Power Resource</td>
</tr>
<tr>
<td>H</td>
<td>Generic symbol for an hour</td>
</tr>
<tr>
<td>IPR</td>
<td>Intermittent Power Resource</td>
</tr>
<tr>
<td>LAH(T)</td>
<td>Look-ahead horizon between close of M(T) and start of T</td>
</tr>
<tr>
<td>LSE</td>
<td>Load-Serving Entity</td>
</tr>
<tr>
<td>M(T)</td>
<td>Swing-contract market for a future operating period T</td>
</tr>
<tr>
<td>m</td>
<td>Generic symbol for a DPR</td>
</tr>
<tr>
<td>n</td>
<td>Generic symbol for an IPR</td>
</tr>
<tr>
<td>p</td>
<td>Generic symbol for a power level (MW)</td>
</tr>
<tr>
<td>p</td>
<td>Generic symbol for a power-path</td>
</tr>
<tr>
<td>p b(T)</td>
<td>Generic symbol for a power-path (p b(t)</td>
</tr>
<tr>
<td>r</td>
<td>Generic symbol for a ramp-rate (MW/min)</td>
</tr>
<tr>
<td>SC</td>
<td>Swing contract taking the general form SC = (α, T^e, PP, φ)</td>
</tr>
<tr>
<td>SC_m(T)</td>
<td>SC submitted by a DPR m to a swing-contract market M(T) for operating period T</td>
</tr>
<tr>
<td>t ex</td>
<td>Exercise time in an exercise set</td>
</tr>
<tr>
<td>T^e_m(T)</td>
<td>Exercise time in an exercise set T^e_m(T)</td>
</tr>
<tr>
<td>T = [t s, t e)</td>
<td>Operating period with start-time t s and end-time t e</td>
</tr>
<tr>
<td>α</td>
<td>Offer price ($) for a swing-contract SC</td>
</tr>
<tr>
<td>α_m(T)</td>
<td>Offer price ($) for a swing-contract SC_m(T)</td>
</tr>
<tr>
<td>∆T</td>
<td>Duration of operating period T, measured in real hourly units (e.g., 0.6h)</td>
</tr>
<tr>
<td>φ</td>
<td>Performance payment method for a swing contract SC that maps PP into payments</td>
</tr>
<tr>
<td>φ_m(T)</td>
<td>Performance payment method for a swing contract SC_m(T) that maps each power-path p_m(T) ∈ PP_m(T) into a dollar payment ($)</td>
</tr>
<tr>
<td><strong>Sets &amp; Subsets:</strong></td>
<td></td>
</tr>
<tr>
<td>B = {1, ..., NB}</td>
<td>Index set for the buses b of a transmission grid</td>
</tr>
<tr>
<td>C_j(b)</td>
<td>Collection of customers serviced by load-serving entity j ∈ LSE(b)</td>
</tr>
<tr>
<td>L ⊆ B × B</td>
<td>Index set for the distinct bus-to-bus line segments ℓ of a transmission grid</td>
</tr>
<tr>
<td>L_o(b) ⊆ L</td>
<td>Subset of transmission-grid line segments originating at bus b</td>
</tr>
<tr>
<td>L_e(b) ⊆ L</td>
<td>Subset of transmission-grid line segments ending at bus b</td>
</tr>
<tr>
<td>LSE</td>
<td>Index set for the load-serving entities j participating in a swing-contract market</td>
</tr>
<tr>
<td>LSE_j(b) ⊆ LSE</td>
<td>Subset of load-serving entities j in LSE that service power customers at bus b</td>
</tr>
<tr>
<td>M</td>
<td>Index set for DPRs m participating in a swing-contract market</td>
</tr>
<tr>
<td>M(b) ⊆ M</td>
<td>Subset of DPRs m in M that are electrically connected at bus b</td>
</tr>
<tr>
<td>NG</td>
<td>Index set for IPRs n participating in a swing-contract market</td>
</tr>
<tr>
<td>NG(b) ⊆ NG</td>
<td>Subset of IPRs n in NG that are electrically connected at bus b</td>
</tr>
<tr>
<td>PP</td>
<td>Set of dispatchable power-paths p offered by a swing contract SC</td>
</tr>
<tr>
<td>PP_m(T)</td>
<td>Set of dispatchable power-paths p_m(T) offered by a swing contract SC_m(T)</td>
</tr>
<tr>
<td>P_m</td>
<td>Set of feasible sustainable power levels p (MW) for DPR m</td>
</tr>
<tr>
<td>R_R_m</td>
<td>Set of feasible ramp-rates r (MW/min) for DPR m</td>
</tr>
<tr>
<td>T^e</td>
<td>Set of possible exercise times t^e for a swing-contract SC</td>
</tr>
<tr>
<td>T^e_m(T)</td>
<td>Set of possible exercise times t^e_m(T) for a swing contract SC_m(T)</td>
</tr>
</tbody>
</table>
A.6 Invertibility of Demand and Supply Functions

The following conditions suffice to ensure an inverse demand schedule \( D_j(q) = \pi \) for a buyer \( j \), as defined in \text{CM6}, can be inverted to obtain a well-defined ordinary demand schedule \( q = D_j'(\pi) \) for buyer \( j \) as defined in \text{CM3}, and vice versa, where \( D_j(q) \) coincides with buyer \( j \)'s marginal benefit function \( MB_j(q) \); see [45, Sec. 9.3.4] for extended discussion.

Suppose buyer \( j \) has a benefit function \( B_j(q) \), defined as in \text{CM4}, that is non-decreasing, differentiable, and concave over \( q \geq 0 \). Buyer \( j \)'s marginal benefit, evaluated at any \( q' \geq 0 \), is then the non-negative derivative of buyer \( j \)'s benefit function \( B_j(q) \) with respect to \( q \), evaluated at \( q = q' \). This mapping \( D_j(q') \) of \( q' \) into a non-negative incremental benefit evaluation \( \partial B_j(q')/\partial q =: MB_j(q') =: \pi' (\text{$/u}) \) is buyer \( j \)'s inverse demand schedule for \( Q \). Finally, if buyer \( j \)'s marginal benefit function \( MB_j(q) \) is a strictly decreasing function of \( q \) for \( q \geq 0 \), a common “diminishing marginal returns” assumption for commodity spot markets, it can be inverted over \( q \geq 0 \) to give a strictly decreasing ordinary demand schedule \( q = D_j'(\pi) \) for buyer \( j \). In this case, by construction, the price \( \pi' \) that satisfies \( q' = D_j'(\pi') \) is the marginal benefit of buyer \( j \) evaluated at the \( Q \)-demand level \( q' \).

The following conditions suffice to ensure an inverse supply schedule \( S_i(q) = \pi \) for a supplier \( i \), as defined in \text{CM10}, can be inverted to obtain a well-defined ordinary supply schedule \( q = S_i'(\pi) \) for supplier \( i \) as defined in \text{CM7}, and vice versa, where \( S_i(q) \) coincides with supplier \( i \)'s marginal cost function \( MC_i(q) \); see [45, Sec. 8.2] for extended discussion.

Suppose supplier \( i \) has a total avoidable cost function \( C_i(q) \) as defined in \text{CM8} that is non-decreasing, differentiable, and convex over \( q \geq 0 \). Evaluated at any \( Q \)-supply level \( q' \geq 0 \), supplier \( i \)'s marginal cost \( MC_i(q') \) (measured in $/u) defined as in \text{CM9} is then the derivative of supplier \( i \)'s total avoidable cost function \( C_i(q) \) with respect to \( q \), evaluated at \( q' \geq 0 \). This mapping \( C_i(q') \geq 0 \) into a non-negative marginal cost evaluation \( \partial C_i(q')/\partial q =: MC_i(q') =: \pi' (\text{$/u}) \) is supplier \( i \)'s inverse supply schedule for \( Q \). Finally, if supplier \( i \)'s marginal cost function \( MC_i(q) \) is a strictly increasing function of \( q \) for \( q \geq 0 \), a common “increasing marginal cost” assumption for commodity spot markets, it can be inverted over \( q \geq 0 \) to give an ordinary supply schedule for supplier \( i \); i.e., to give a strictly-increasing function \( S_i'(\pi) \) mapping each non-negative \( Q \)-unit price \( \pi' \) (measured in $/u) into a non-negative \( Q \)-supply \( q' = S_i'(\pi') \) (measured in \( u \)). In this case, by construction, the \( Q \)-unit price \( \pi' \) that maps into \( q' \) is supplier \( i \)'s marginal cost \( MC_i(q') \) (measured in $/u), evaluated at the \( Q \)-supply level \( q' \).
References


