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

Leigh Tesfatsion

Abstract:

Centrally-managed U.S. wholesale power markets operating over high-voltage AC transmission grids are transitioning from heavy reliance on fossil-fuel based power to greater reliance on renewable power with increasingly diverse suppliers and customers. This study highlights four conceptually-problematic economic presumptions reflected in the legacy core design of these markets that are hindering this transition. The key problematic presumption is the static conceptualization of the basic product as grid-delivered energy (MWh) transacted in short-run (day-ahead and intra-day) markets at competitively determined unit prices ($/MWh), conditional on delivery location and time. This study argues, to the contrary, that the basic product in need of efficient reliable transaction in these markets is reserve (physically-covered insurance) for protection against power imbalance (volumetric grid risk). This reserve is the guaranteed availability of dispatchable nodal power-production capabilities for possible central dispatch during designated future operating periods at designated grid delivery locations to satisfy just-in-time customer power demands and grid reliability requirements. For illustration, a recently proposed Linked Swing-Contract Market Design is briefly reviewed. The latter 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. The swing in these contracts permits efficient planning for real-time reliability, and the two-part pricing form of these contracts permits cleared suppliers to assure their revenue sufficiency. A principled cost allocation rule supports the independence of the fiducial central manager by assuring break-even revenue adequacy for system operations as a whole.

Keywords:

Market design, wholesale electric power markets, renewable power integration, volumetric grid risk, linked forward reserve markets, physically-covered insurance, flexible dispatch, nodal multi-interval pricing, revenue sufficiency, digital twinning

---

*1 October 2024. Forthcoming in *Foundations and Trends in Electric Energy Systems (FnT-EES) Series*, NOW Publishers, Delft, the Netherlands. This study is a revised version of Working Paper #22005 (Iowa State University Digital Repository) submitted as a supporting document [45] for comments e-filed to FERC for Docket AD21-10-000 [16]. Author Contact Information: Leigh Tesfatsion, Professor Emerita of Economics, Heady Hall 260, Iowa State University, Ames, IA 50011-1054. Email: `tesfatsi@iastate.edu`; Homepage: `https://www2.econ.iastate.edu/tesfatsi` **NOTE:** This compiled version includes revised contact information (4/21/2026) plus corrections to the exposition of paragraph 2 on p. 42 (2/25/2026).

Table of Contents

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 behind-the-meter 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 [21], U.S. RTO/ISO-managed wholesale power markets[1] are transitioning from a traditionally heavy reliance on fossil-fuel based power generators to a greater reliance on *Intermittent Power Resources (IPRs)*.[2] These IPRs include wind farms, photovoltaic solar arrays, and hydropower facilities whose weather-dependent power generation is not fully firmed 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 [13] encouraging more active participation by demand-side resources, has increased the uncertainty and volatility of grid *net load*.[3] In consequence, as reported in [14], 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.[4]

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 [35, 41] 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 [15].

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*.

key challenges facing current U.S. RTO/ISO-managed wholesale power markets. In 2022 the U.S. Federal Energy Regulatory Commission (FERC) issued an order [16] requesting a fundamental reconsideration of the design and operation of these markets. In 2024 a group of researchers at Resources for the Future (RFF) released a report [31] titled "Time for a Market Upgrade?" that examines current U.S. wholesale power market operations in relation to critical future needs.

Strongly encouraged by these calls for action, efforts are underway to improve the conceptual and operational design of U.S. wholesale power markets. As discussed in later sections of this study, these efforts are taking diverse forms. Nevertheless, they largely adhere to the following nine 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, deratings, 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 potental 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 the first eight goals are based on Oren [37, Sec. II.A], Tesfatsion et al. [49, Sec. 2], and Tesfatsion [43, 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.

Goal (G9): *Internalization of Externalities.* The market design should permit the net-benefit (i.e., benefit minus cost) objective functions used in centrally-managed market-clearing processes to internalize *social* benefits and costs reflecting the environmental impacts of electric power production, transmission, and distribution.

Despite the general acceptance of goals (G1)–(G9), 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 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]

The current dynamic reality is far more daunting: U.S. RTOs/ISOs are fiducial 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: *Availability* of nodal 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)

[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.

should be permitted to submit supply offers in a *two-part pricing*[7] form enabling full compensation for:

(1) *avoidable fixed cost* that supplier *i* must incur to guarantee the *availability* of reserve (dispatchable nodal 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 counterclaims (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* [43] proposed for grid-supported centrally-managed wholesale power markets. It is argued that this alternative design is consistent with goals (G1)–(G9) and counterclaims (CC1)–(CC4), and is well-suited for the scalable support of increasingly decarbonized grid operations with more active participation by diverse suppliers and customers.

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.

---

[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 (flexibility) in their delivery terms, as well as *variable costs* for actual real-time deliveries, and both types of costs must be fully covered in order for these suppliers to stay in business.

[8] Shortened versions of the essential background materials in Sections 2–4 appear in Tesfatsion [46, Secs. III-IV], a companion study focused more narrowly on locational marginal pricing.

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

The development of the legacy core *Two-Settlement System* [34] 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 [7].

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 *Real-Time Market (RTM)* for current-day planning that operates in parallel with a daily bid/offer-based RTO/ISO-managed *Day-Ahead Market (DAM)* for next-day planning.

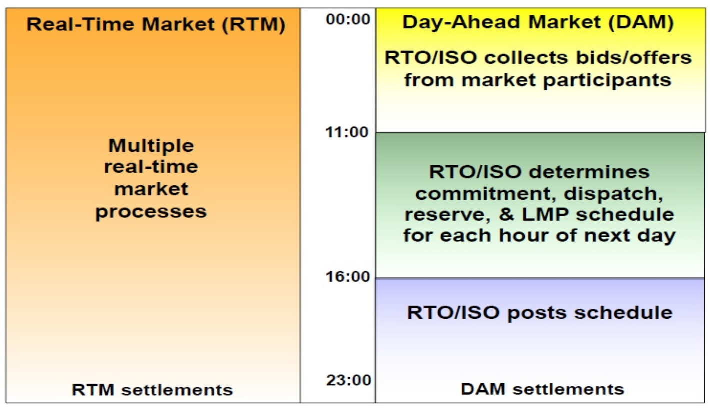


Fig. 1 Simplified depiction of Two-Settlement System operations on a typical Day D: Parallel real-time and day-ahead market processes are conducted and settled for days D and D+1.

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 customers in geographically disjoint regions, submit demands bids into the day-D DAM for the purchase of energy at designated 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[9] 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* [39]; 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 τ 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 τ. 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]

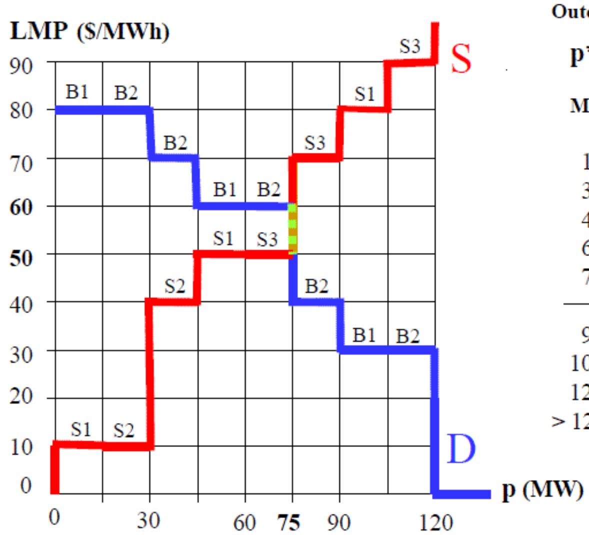
[9] SCUC is an acronym for *Security-Constrained Unit Commitment*, and SCED is an acronym for *Security-Constrained Economic Dispatch*.

[10] See [30] and [46] 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].

The purpose of these RTM sub-markets (supported by 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 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.

Real-time operations for U.S. RTO/ISO-managed wholesale power markets depend on real-time telemetry to dispatch an ultimately-determined bid/offer-based SCED solution for an operating period T. This ultimate SCED solution is conditioned on available dispatchable resources, up-to-date forecasts for net fixed loads, and up-to-date assessments of physical operating conditions.



Market-Clearing Demand = Supply Outcomes (p*, LMP*) for hour H

p*=75, 50 ≤ LMP* ≤ 60

MW	***max* B Price**	***min* S Price**
15	$80	$10
30	$80	$10
45	$70	$40
60	$60	$50
75	$60	$50
90	$40	$70
105	$30	$80
120	$30	$90
> 120	$0	$∞

Fig. 2 Illustrative depiction of the optimal market-clearing (demand = supply) solution for the *maintained* power-withdrawal levels of cleared buyers and the *maintained* power-injection levels of cleared suppliers for a given hour H during operating day D+1, as determined by a bid/offer-based RTO/ISO-managed DAM SCED optimization conducted on day D.

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 day-D 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) on the following day D+1. 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* ($/MWh) – i.e., the highest *maximum willingness to pay* ($/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,LMP) plane with a common optimal power level p^* = 75 (MW) and a *range* of optimal price levels LMP* ($/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 surplus* from this participation – i.e., maximum possible total buyer benefit ($) net of total supplier production cost ($). 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]

---

[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 [30] for a more extensive discussion of these points.

[13] As carefully explained and illustrated in [25], [42], and [43, 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

For later purposes, additional aspects of the U.S. RTO/ISO-managed DAM SCED optimization formulation depicted in Fig. 2 are highlighted below, where the symbol ":=" denotes "*is equal by definition to*".

- 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 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 $B_n(k)$, $n = 1, ..., N_k$, where N is set by the RTO/ISO: e.g., $N = 10$ in ISO-NE [22] and $N = 9$ in MISO [33].
- 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, π)-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 delivery 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

---

use on the total net surplus attained by the market participants. Conversely, no RTO/ISO extraction of congestion rent occurs in the absence of LMP separation.

settlement terms they agreed to in these other venues are not disrupted by obligatory DAM/RTM LMP price settlements; see [43, 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.

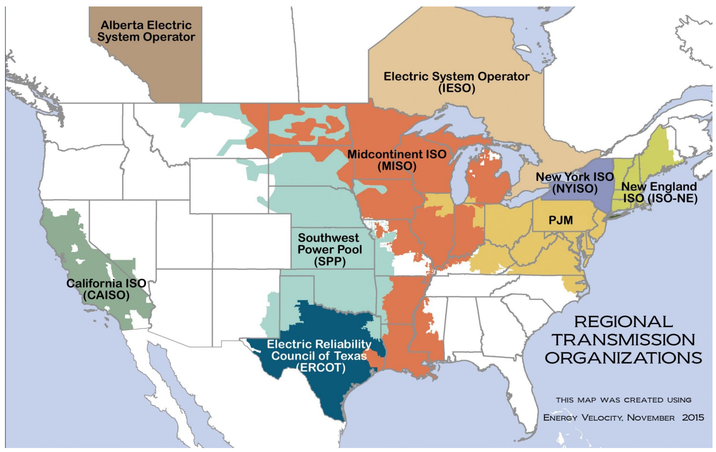


Fig. 3 North American RTO/ISO-managed wholesale power markets. (Public domain: [12])

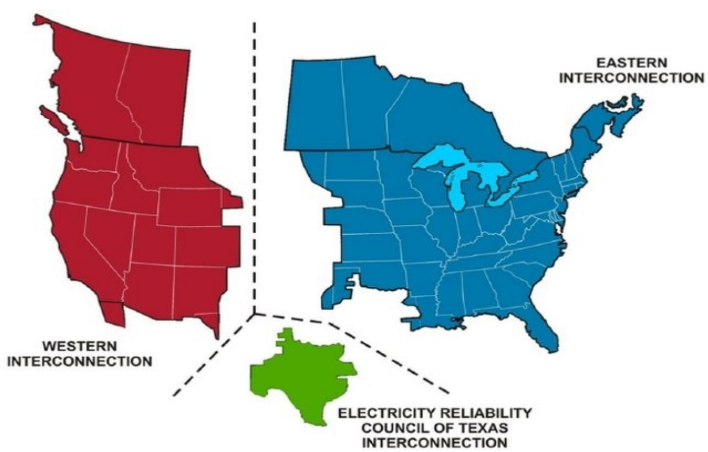


Fig. 4 North American RTOs/ISOs operate over a physical high-voltage AC transmission grid consisting of three separately-synchronized parts. (Public domain image by Jong Suk Kim)

Finally, as originally envisioned by FERC [7, p. 11], the Two-Settlement System did not include specific rules for the provision and settlement of ancillary services.[14] Rather, resource adequacy assurance was specifically left up to the states participating in each energy region managed by an RTO or ISO. As reported in [5, Tables 1-2] and [14, Table 1, p. 6], ancillary service procurement and settlement processes currently 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 := $[1\text{kg}][1\text{m}]^2[1s]^{-3}$; 1kW := 1000W; 1MW := 1000kW; *Volt* (V) := Unit of Electric Potential: 1V := $[1\text{W}][1\text{A}]^{-1}$; *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 := $[1\text{ cycle}] \cdot [1\text{s}]^{-1}$.

Other Commonly Used Physical Measurement Units: *Degree Fahrenheit* (^{o}F), a normalized temperature unit such that water freezes at 32^{o}F and boils at 212^{o}F; *British Thermal Unit* (Btu), a unit of heat defined to be the quantity of heat required to raise by 1^{o}F the temperature of 1lb of liquid water currently at the temperature ($\approx$ 39^{o}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.

[14] *Ancillary services* are grid-reliability support services [14, Appendix]. Examples include: "black-start" power-flow restoration for a collapsed grid; reactive power for voltage control; operating reserve (net-load balancing services) provided by on-line generation units with unencumbered capacity; and net-load balancing services provided by off-line 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 attributes.

Asset Example: *Apple* := *(Location; Time; Weight; Shape; Color; ...)*

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 [45]. 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 $R := \{r, \ldots\}$ denote the set of real numbers.[16] The standard algebraic operators that act on elements r of R include: addition $(+)$; subtraction $(-)$; multiplication $(\times)$; division $(\div)$; 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 π' (measured in \$/u) as long as: (i) π' 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) π' 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 $Z := \{0, 1, -1, 2, -2, 3, -3, \ldots\}$ 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 $\{m, kg, s, A, K, mol, cd\}$ 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: *2MWh = 2MWh*
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. [38, fn, p. 1153] and [39, App. F.1], a proposed *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 time-period 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. *Power loads* are characterized as the loads of devices requiring power at specific times during a time-period 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 [39] is not addressed by the authors.

Statement S4: *1 DURACELL AA 1.5v Battery = 1 DURACELL AA 1.5v Battery*
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?

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.*
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.

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.*

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 *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 scheduled for grid delivery at a designated grid location b during a future 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 power-delivery *timing* during 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 a real-line quantity axis suggest the desirability of considering alternative constructive mathematical modeling approaches, such as agent-based modeling, that permit "holistic" representations of real-world phenomena and their interactions. See, for example, the extended discussions and illustrations of this point provided in Tesfatsion [47] and [44, Sec. 3].

Crucial ramifications of **Lemmas 3.1 – 3.3** for the design and operation of grid-supported centrally-managed wholesale power markets are explored 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, all economic concepts are based on revealed preference theory constructions (i.e., measured willingness to pay or be paid) rather than on the postulated existence of preference orders with ordinal utility-function representations; see Tesfatsion [43, Sec. 9.3.4] for expanded discussion of these alternative approaches. Also, 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 consistent 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 $[AFC_i^o + VC_i(\text{a})]$ 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 π (measured in \$/u) into the maximum Q-amount $q := D_j^o(\pi)$ (measured in u) that buyer j is willing to procure at price π.

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 π (measured in \$/u) into the maximum Q-amount $q := S_i^o(\pi)$ (measured in u) that supplier i is willing to supply at price π.

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:[17]

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

(CCSM2) Each buyer j and supplier i is a price-taker.[18]

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

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

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

$$q \;:=\; \sum_j q_j \;:=\; \sum_j D^o_j(\pi) \;:=\; D^o(\pi) \tag{1}$$

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

$$q \;:=\; \sum_i q_i \;:=\; \sum_i S^o_i(\pi) \;:=\; S^o(\pi) \tag{2}$$

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 (π, 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:*

$$q^* \;=\; D^o(\pi^*) \;=\; S^o(\pi^*) \;\; . \tag{3}$$

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 [43, 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 π (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 ensuring the following conditions (a) and (b) hold for CCSM(Q):*

(a) *Each buyer j participating in CCSM(Q) has a strictly-decreasing inverse demand schedule $\pi := D_j(q_j)$ that can be inverted to give a strictly-decreasing ordinary demand schedule $q_j := D_j^o(\pi)$ 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 strictly-increasing inverse supply schedule $\pi := S_i(q_i)$ that can be inverted to give a strictly-increasing ordinary supply schedule $q_i := S_i^o(\pi)$ 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 $q_i^* > 0$,

$$\pi^* \;=\; \mathrm{MB}_j(q_j^*) \;=\; \mathrm{MC}_i(q_i^*) \tag{4}$$

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 ordinary demand and supply expressions $q_j^* := D_j^o(\pi^*)$ and $q_i^* := S_i^o(\pi^*)$ appearing in (3) by appropriate *inverse* demand and supply operations. //

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"). //

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 benefit and cost functions for CCSM(Q) participants.

price π^* (\$/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 partication.*

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 variable 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 supplier i at e^*. If $q_i^* = 0$, supplier i incurs zero variable 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 π^* at e^*. Moreover, from condition (b) in Lemma 4.1, supplier i's marginal cost function coincides with supplier i's strictly-increasing 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 variable 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) at any competitive equilibrium* $e^* := (\pi^*, q^*)$ *for MP-CCSM(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 [43, 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 [43, Ch. 12] for careful discussion and illustrations of these and related market concepts.

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 π^b and the supplier's sale reservation value π^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 π^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 π^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* π^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* π^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) 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 π 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 π 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 π. At the end of the experiment, one of the listed fruit-unit prices, π^*, 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 HoneyCrisp Apple to be a perfect substitute for a Dole Mandarin Orange; that is, suppose the specific apple-versus-orange attribute of a fruit-unit *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 π 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-units 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 π^*?

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 units 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 conceptually well-defined for fruit for a fruit-supplier i at a given location and time *unless* supplier i considers all units 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" unit of fruit, given he has already procured a fruit-quantity q, without knowing: (i) which specific fruit unit, apple or orange, is to be his "next" procured fruit unit; 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" unit of fruit, given he has already supplied a fruit-quantity q, without knowing: (i) which specific fruit unit, apple or orange, is to be his "next" supplied fruit unit; 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 (4).

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 U.S. practice; the emphasis of U.S. RTO/ISO-managed wholesale power 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 grid-supported centrally-managed wholesale power 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 nodal power-production capabilities for possible central dispatch during future operating periods to protect against volumetric grid risk.

Counterclaim (CC1) to presumption (P1) is the primary message stressed throughout this study. Section 5.2 commences support for (CC1) by carefully considering fundamental physical and valuation concerns regarding the validity of presumption (P1), as follows:

- *Physical Reliability Concerns:* Energy transactions in grid-supported centrally-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 nodal flows of power (MW) at single designated grid locations b during designated operating periods T.
- *Benefit and Cost Valuation Concerns:* How power is injected or withdrawn *at* grid locations b *during* successive operating periods T can matter greatly to power producers, power customers, and central managers. The total amount of energy E(b,T) (MWh) resulting from a grid-delivered flow of power (MW) at a grid location b during an operating period T (h) is only one of many possible valued attributes this flow of power could possess.

Additional support for counterclaim (CC1) is then provided in Section 5.3 through Section 5.5 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* [50, pp. 35-38]. Expressed for 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).

Consider, for example, the analytical modeling developed in [43, Ch. 7 & Sec. 9.2] for a grid-supported RTO/ISO-managed wholesale power market M(T) for a future operating period T $:= [t^{\mathrm{s}}, t^{\mathrm{e}})$ with finite duration. The relative timing of M(T) and T are depicted in Fig. 5.

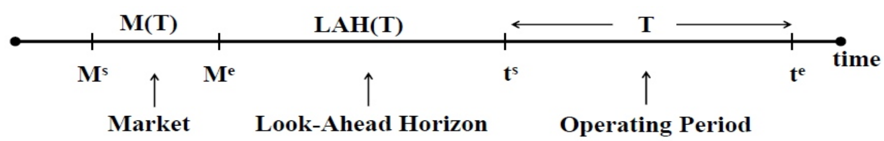


Fig. 5 Time-line for a grid-supported RTO/ISO-managed wholesale power market M(T) conducted for a future operating period T.

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

- a set $\mathbb{M}(b)$ of *dispatchable* generation units *m*, each with a unique electrical point of connection to the transmission grid at bus b;
- a set $\mathbb{LSE}(b)$ of LSEs j that function as market intermediaries for disjoint geographically-located sets $\mathbb{C}_j(b)$ of power customers with unique electrical points-of-connection to the transmission grid at bus b;
- a set $\mathbb{NG}(b)$ of *non-dispatchable* generators *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 [43, 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_o. 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 $p := V_o \cdot I$.

[24] See [40, 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 $\in \mathbb{B}$ and time $t \in$ T, the *total dispatched power injection* at bus b by the dispatchable generation units $m \in \mathbb{M}(b)$, plus the *total net line-power inflow* at bus b from buses b$'$ in $\mathbb{B}$ 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 $j \in \mathbb{LSE}$(b) plus the *total forecasted fixed customer load* at bus b for customers of the LSEs $j \in \mathbb{LSE}$(b) minus the *total forecasted fixed power injection* at bus b by the non-dispatchable generators $n \in \mathbb{NG}$(b).

For the purposes of this section, however, it is important to express these net-load balance constraints in their explicit mathematical forms [43, Ch. 6 & Sec. 9.2]:

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

$$G_b^{\mathsf{dis}}(t) \ + \ \mathrm{NLPI}_b(t) \ = \ L_b^{\mathsf{dis}}(t) \ + \ \widehat{NL}_b^{\mathsf{f}}(t) \tag{5}$$

where:

$$G_b^{\mathsf{dis}}(t) \ := \sum_{m\in\mathbb{M}(b)} p_m^{\mathsf{dis}}(t) \qquad (\textit{total dispatched power injection});$$

$$\mathrm{NLPI}_b(t) \ := \ \Big[\sum_{\ell\in\mathbb{L}_{E(b)}} w_\ell(t) - \sum_{\ell\in\mathbb{L}_{O(b)}} w_\ell(t) \ \Big] \qquad (\textit{total net line-power inflow});$$

$$L_b^{\mathsf{dis}}(t) \ := \sum_{j\in\mathbb{LSE}(b)} p_j^{\mathsf{dis}}(t) \qquad (\textit{ total dispatched customer load });$$

$$\widehat{NL}_b^{\mathsf{f}}(t) \ := \ [\widehat{L}_b^{f}(t) \ - \ \widehat{G}_b^{\mathsf{f}}(t)] \qquad (\textit{total forecasted net fixed load});$$

$$\widehat{L}_b^{\mathsf{f}}(t) \ := \sum_{j\in\mathbb{LSE}(b)} \widehat{p}_j^{\mathsf{f}}(t) \qquad (\textit{total forecasted fixed customer load});$$

$$\widehat{G}_b^{\mathsf{f}}(t) \ := \sum_{n\in\mathbb{NG}(b)} \widehat{p}_n^{\mathsf{f}}(t) \qquad (\textit{total forecasted fixed power injection}).$$

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 $\in \mathbb{B}$ for operating period T. Rather, they are complicated joint restrictions on *power-paths*, defined as follows:

> **Definition:** A *power-path* $\mathbf{p}_b$(T) = $(\, p_b(t) \mid t \in \mathrm{T}\,)$ 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 $\in \mathbb{B}$ during T:

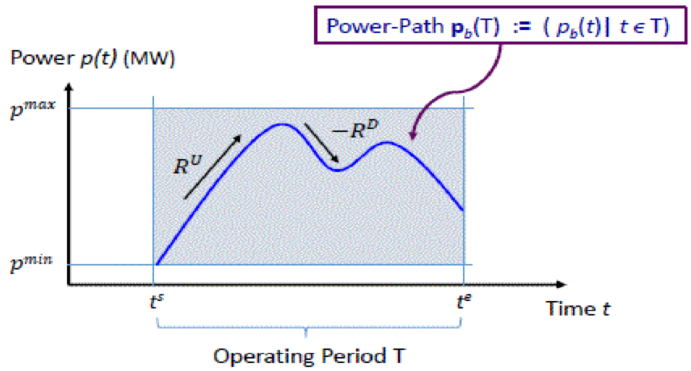


Fig. 6 Illustration of a power-path at a single grid location b during an operating period T. Note that a power-path is inherently a nodal conception.

- the *dispatched* power-path $\mathbf{p}^{\text{dis}}_m(\mathrm{T}) := (\, p^{\text{dis}}_m(t) \mid t \in \mathrm{T}\,)$ at b during T for each *dispatchable* generation unit $m \in \mathbb{M}(b)$;
- the *dispatched* power-path $\mathbf{p}^{\text{dis}}_j(\mathrm{T}) := (\, p^{\text{dis}}_j(t) \mid t \in \mathrm{T}\,)$ at b during T for the customers serviced by each LSE $j \in \mathbb{LSE}(b)$;
- the *forecasted fixed* power-path $\widehat{\mathbf{p}}^{\text{f}}_j(\mathrm{T}) := (\, \widehat{p}^{\text{f}}_j(t) \mid t \in \mathrm{T}\,)$ for *fixed power withdrawals* at b during T by the customers serviced by each LSE $j \in \mathbb{LSE}(b)$;
- the *forecasted fixed* power-path $\widehat{\mathbf{p}}^{\text{f}}_n(\mathrm{T}) := (\, \widehat{p}^{\text{f}}_n(t) \mid t \in \mathrm{T}\,)$ for *fixed power injections* at b during T by each *non-dispatchable* generator $n \in \mathbb{NG}(b)$.

As demonstrated in [43, Ch. 7], given mild regularity conditions and a finite-duration operating period $\mathrm{T} := [t^{\text{s}}, t^{\text{e}})$, it is possible to approximate a power-path $\mathbf{p}_b(\mathrm{T})$ as closely as desired by a step-function.[25] For example, this *step-function approximation* for $\mathbf{p}_b(\mathrm{T})$ could consist of a discretized sequence

$$\mathbf{p}_b(\mathrm{K}(\mathrm{T})) \;:=\; (p_b(k) \;\mid\; k \in \mathrm{K}(\mathrm{T})) \tag{6}$$

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

$$k_n \;:=\; [k^{\text{s}}_n, k^{\text{e}}_n), \;\; n = 1, \ldots, \mathrm{N}(\mathrm{T}) \tag{7}$$

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

$$p_b(k_n) \;:=\; p_b(k^{\text{s}}_n) \;\; (\mathrm{MW}) \tag{8}$$

[25] Step functions are universal approximators for the class of all continuous real-valued functions $f{:}[a,b] \to R$ 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 $f_\varepsilon{:}[a,b] \to R$ 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 ε. This assertion follows immediately from the Heine-Cantor Theorem, which establishes that any continuous function $f{:}X \to Y$ between metric spaces X and Y, with X compact, is *uniformly* continuous.

denotes the power-level of power-path $\mathbf{p}_b$(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 [43, Chs. 7,16] and [27], 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 Δk_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 [20, 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 nameplate capacity limits.

See [43, 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 [43, Ch. 7 & Sec. 9.2] for an RTO/ISO-managed wholesale power market M(T) that was used in Section 5.2.2 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 $:= [t^{\mathrm{s}}, t^{\mathrm{e}})$ with finite duration.

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

Let $\mathbf{p}_b$(T) $:= (\, p_b(t) \mid t \in \mathrm{T}\,)$ 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 $\mathbf{p}_b$(T) has a continuous extension over $\bar{\mathrm{T}} := [t^{\mathrm{s}}, t^{\mathrm{e}}]$, the compact closure of T. Then $\mathbf{p}_b(\mathrm{T})$ can be approximated arbitrarily closely over T by a suitably-constructed step function $\mathbf{p}_b(\mathrm{K(T)})$ taking form (6). Plotted in a time-MW plane, this approximating step-function $\mathbf{p}_b(\mathrm{K(T)})$ consists of a finite sequence of *energy-blocks* (MWh):

$$E_{b,n} \ := \ p_b(k_n^s) \times [k_n^e - k_n^s] \quad \text{for} \ \ n = 1, \ldots, \mathrm{N(T)}. \tag{9}$$

Consequently, if the power-path $\mathbf{p}_b(\mathrm{T})$ 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 $\mathbf{p}_b(\mathrm{T})$ in no way guarantees that the *actual* cost and/or benefit value assigned to $\mathbf{p}_b(\mathrm{T})$ 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 $\mathbf{p}_b(\mathrm{T})$ 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 $\mathbf{p}_b(\mathrm{T})$ 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 $\mathbf{p}_b(\mathrm{T})$ 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-simplistic 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:

$$\textit{Total Cost} \;:=\; \textit{Sunk Cost} \;+\; \textit{Avoidable Fixed Cost} \;+\; \textit{Variable Cost} \tag{10}$$

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:

$$\begin{aligned}\textit{Sunk Cost} &:= \textit{Non-Avoidable Fixed Cost} \\ &:= \text{Cost } \; SC^o \; \text{ that:} \\ &\quad \text{(i) DM incurs } \textbf{whether or not} \text{ DM commits at time } t \\ &\qquad \text{to undertaking a type-}A\text{ action at time } t+\Delta t; \\ &\quad \text{(ii) } \textbf{does not} \text{ depend on the specific type-}A\text{ action} \\ &\qquad \text{DM undertakes, should DM choose to commit.}\end{aligned} \tag{11}$$

$$\begin{aligned}\textit{Avoidable Fixed Cost} \;:=\; & \text{Cost } \; AFC^o \; \text{ that:} \\ & \text{(i) DM incurs } \textbf{if and only if} \text{ DM commits at time } t \\ & \quad \text{to undertaking a type-}A\text{ action at time } t+\Delta t; \\ & \text{(ii) } \textbf{does not} \text{ depend on the specific type-}A\text{ action} \\ & \quad \text{DM undertakes, should DM choose to commit.}\end{aligned} \tag{12}$$

$$\begin{aligned}\textit{Variable Cost} \;:=\; & \text{Cost } \; VC(a) \; \text{ that:} \\ & \text{(i) DM incurs } \textbf{if and only if} \text{ DM commits at time } t \\ & \quad \text{to undertaking a type-}A\text{ action at time } t+\Delta t; \\ & \text{(ii) } \textbf{does} \text{ depend on specific form } a \text{ of type-}A\text{ action} \\ & \quad \text{DM undertakes, should DM choose to commit.}\end{aligned} \tag{13}$$

$$\textit{Fixed Cost} \;:=\; [\, \textit{Sunk Cost} \;+\; \textit{Avoidable Fixed Cost} \,] \tag{14}$$

$$\textit{Avoidable Cost} \;:=\; [\, \textit{Avoidable Fixed Cost} \;+\; \textit{Variable Cost} \,] \tag{15}$$

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 $VC_m(p,\mathrm{H},\mathrm{D+1})$ 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 DM in terms of the three-part partition (10) for total cost.

By definition, DM's sunk cost at time t is non-avoidable, hence incurred whether or not DM agrees to the commitment at time t. DM's sunk cost at time t should therefore play no role in 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.*

Let DM's "net benefit" be defined as follows:

$$\big[\text{Net Benefit}\big] \;:=\; \big[\text{Benefit}\big] - \big[\text{Avoidable Cost}\big] \tag{16}$$

Define DM to be *risk-averse*[28] if DM is *not* willing to participate:

- in any *risky* undertaking with *zero* expected net benefit, where the qualifier "risky" means there is a positive probability of experiencing a strictly negative net benefit outcome;
- in *any* undertaking with *strictly negative* expected net benefit.

Thus, if risk-averse, DM will agree to the commitment at time t *only if* DM's *expected* net benefit from this commitment is *non-negative*.

---

[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] Kirschen and Strbac [24] 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 *utility functions* representing preference orderings over lotteries – i.e., gambles offering possible payoffs ($) with associated probabilities – exhibit concave curvature properties. In reality, negative payoffs ($) 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 utility as expected lottery payoffs approach extreme negative levels. However, the routine explicit inclusion of survival risks within economic models would permit a more profound understanding and appreciation of the role of institutional scaffolding in real-world economies as ameliorators of these survival risks.

On the other hand, even if risk-averse, DM *should* agree to the commitment at time t *if* DM believes this would result *for sure* in a *strictly positive* net benefit. Note a strictly positive net benefit (16) with non-negative avoidable cost would permit DM to compensate for at least part of any *sunk* cost he has at time t.

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

Commitment Principle for Decison-Maker DM: If risk averse, DM should commit at time t to undertaking an action of type A at the future time $t+\Delta t$:

- *if* DM believes this commitment would result *for sure* in a net benefit (16) that is *strictly positive*;
- *only if* DM believes the *expected* net benefit (16) from agreeing to this commitment is *non-negative*.

Finally, making use of the 3-part total cost partitioning (10) developed earlier 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 actual participants in current 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 design of current 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 required in current 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[29] 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.

Roughly characterized, apart from ISO New England,[30] these DAM-scheduled dispatch set-points simultaneously determine 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.[31]

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 [42], 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[32] required to take the following two-part form: (i) a possibly-zero power-supply level $\bar{p}$ (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 Δp_n accompanied by associated incremental energy prices π_n ($/MWh). Each power interval Δp_n consists of a range $(p_n^{\mathsf{s}}, p_n^{\mathsf{e}}]$ 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 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, [5, Tables 1-2] and [14, Table 1].

[31] The types of operating reserve procured on a co-optimized or coupled basis with energy in U.S. RTO/ISO-managed wholesale power markets include Regulation, Spinning Reserve, and Supplemental Reserve. See [5] and [14, Sec. II.A & 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, π) that commence at the LSL point. Start-up cost is an important form of avoidable fixed cost for QSEs considering the submission of supply offers to ERCOT DAMs.

$[p - p^{\mathsf{s}}_n]$ (MW) that m is offering to maintain during hour H (1h) of day D+1 in return for receiving an incremental compensation ($) in amount $\pi_n \times [p - p^{\mathsf{s}}_n] \times$ 1h.

Given S_m(H,D+1), the RTO/ISO can approximate m's *variable* cost function for hour H of day D+1 as follows. Let p' denote any maintained power-supply level offered by m that strictly exceeds m's fixed power-supply level $\bar{p} \geq 0$. Then the summation of the incremental compensations required by m at p' provides an approximation of the variable cost $VC_m(p', \text{H,D+1})$ ($) that m would incur for hour H of day D+1 if m's supply offer S_m(H,D+1) is cleared at power level p'.

For example, suppose $N = 3$ and $0 \leq \bar{p} < p_1 < p_2 < p' \leq p_3$. Suppose, also, that m's supply offer S_m(H,D+1) for hour H of day D+1 consists of: (i) a fixed maintained power-supply level $\bar{p}$ for hour H of day D+1; and (ii) three ordered power intervals $(\bar{p}, p_1]$, $(p_1, p_2]$, and $(p_2, p']$ with associated incremental compensation levels π_1, π_2, and π_3. Then m's variable cost $VC_m(p', \text{H,D+1})$ ($) for hour H of day D+1, evaluated at p', can be approximated as follows:

$$VC_m(p', \text{H,D+1}) \ := \ \pi_1 E_1 \ + \ \pi_2 E_2 \ + \ \pi_3 E_3(p') \tag{17}$$

where the three energy-blocks E_i (MWh) in (17) are given by

$$E_1 \ := \ [p_1 - \bar{p}][1h]; \ \ E_2 \ := \ [p_2 - p_1][1h]; \ \text{ and } \ E_3(p') \ := \ [p' - p_2][1h] \tag{18}$$

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 the subsequently dispatched power-paths for these resources.

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 := $[t^{\mathrm{s}}, t^{\mathrm{e}})$. This supply offer consists of a single energy-block for possible RTO/ISO dispatch during T. For concreteness, it is assumed: m is in an off-line state (zero power generation) at time 0; m needs to return to an off-line state as soon as possible after T concludes; and m has a constant ramp-up rate $R^U > 0$ and a constant ramp-down rate $-R^D < 0$.

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.

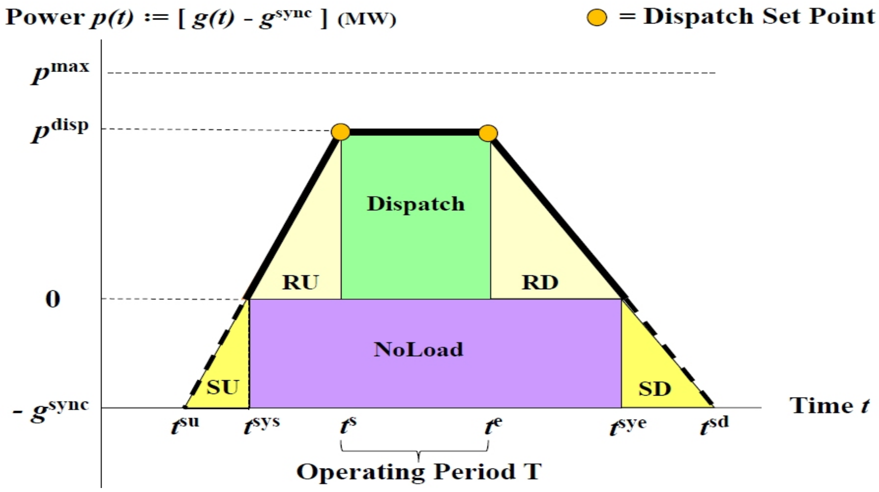


Fig. 7 Illustration of the energy production (MWh) needed before, during, and after an operating period T to support the dispatch of a single energy block ("Dispatch") during T: namely, start-up (SU); ramp-up (RU); no-load (NoLoad); ramp-down (RD), and shut-down (SD).

As depicted in Fig. 7, the level $p(t) := [g(t) - g^{\text{sync}}]$ (MW) of m's power *injection* into the grid at grid-location b(m) at a time point t is the difference between m's *total* power generation $g(t)$ at t and the power generation g^{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^{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^{disp} at the start-time $t^{\text{s}} > t^{\text{sys}}$ for operating period T. At time t^{s}, m's total total power generation is $g(t^{\text{s}})$ = $p^{\text{disp}} + g^{\text{sync}}$. However, only the injected portion p^{disp} is visible to the RTO/ISO.

Generator m then maintains the power injection level p^{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^{e} to time t^{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^{sync} at time t^{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^{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 grid-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 *DispatchCost*.

The accumulation of the behind-the-meter power that m generates during the interval $[t^{\mathsf{su}}, t^{\mathsf{sys}})$ in which m is transitioning from an off-line state to a synchronized state is labeled "SU" in Fig. 7. The accumulation of the behind-the-meter power that m generates to maintain itself in a synchronized state during the next time interval $[t^{\mathsf{sys}}, t^{\mathsf{sye}})$ is labeled "NoLoad" in Fig. 7. Finally, the accumulation of the behind-the-meter power that m generates during the time interval $[t^{\mathsf{sye}}, t^{\mathsf{sd}})$ in which m is transitioning from a synchronized state at time t^{sye} 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 *SUCost*, *NoLoadCost*, and *SDCost*.

Note Fig. 7 also depicts two additional power-injection regions "RU" and "RD". The ramp-up region RU reflects m's need to ramp up its power *injection* level to p^{disp} at the start-time t^{s} 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 m's need 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 *RUCost* and *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). Thus, *focusing solely on energy production costs,* revenue sufficiency for m requires that m receive compensation for the following costs:

$$\textit{Avoidable Fixed Cost:} \quad \text{SUCost, NoLoadCost, SDCost;} \tag{19}$$

$$\textit{Variable Cost:} \quad \text{RUCost, DispatchCost, RDCost.} \tag{20}$$

The cost components SUCost, NoLoadCost, and 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 RUCost, DispatchCost, and RDCost are *variable costs* for m because: (i) they depend on m's commitment to produce an energy-block for operating period T; and (ii) they *also* depend on the *specific form* of this energy block, namely, the power level p^{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 a dispatchable generator m at a grid-location $b(m)$ has a minimum maintainable power-injection level $p^{\min} > 0$. Suppose, also, that the supply offer S_m(D+1) submitted by m on day D to an RTO/ISO-managed day-ahead market M(D+1) for an operating day D+1 offers the guaranteed availability of a *set* $\mathbb{PP}_m$ of power-paths $\mathbf{p}_{b(m)}$ 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 $\mathbb{PP}_m$. Finally, for ease of depiction, suppose all of the power-paths in $\mathbb{PP}_m$ have the same initial power level, $p^{\min} > 0$, at the start-time t^{s} for day D+1; and that m has a constant ramp-up rate $R^U > 0$ and a constant ramp-down rate $-R^D < 0$.

Figure 8 depicts one of the power-paths in generator m's offered power-path set $\mathbb{PP}_m$ for day D+1. This power-path consists of *two* successive energy blocks, E1 and E2, to be grid-delivered at generator m's grid-location b(m) during day D+1, together with two associated ramping regions R1 and R2. Ramping region R1 reflects m's need to ramp up to the maintained power level for E1, starting from $p^{\min}$ at the start-time t^{s} for day D+1. 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.

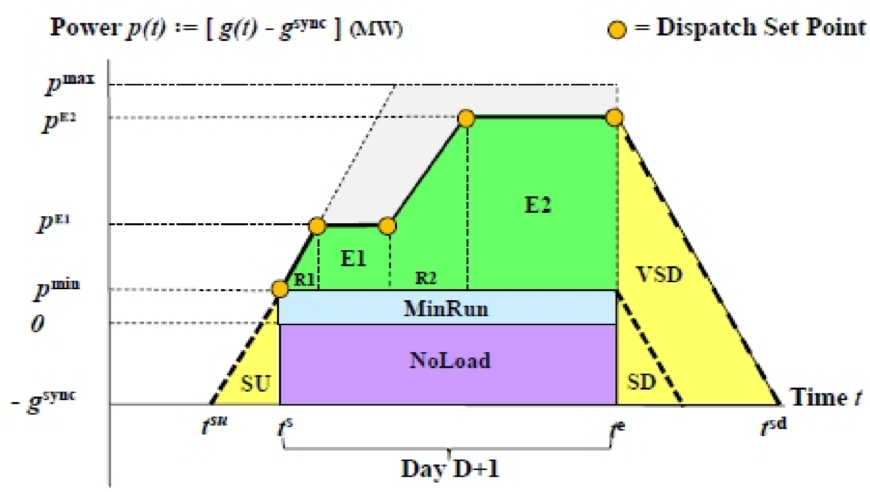


Fig. 8 Illustration of the energy production (MWh) needed before, during, and after an operating day D+1 to support the successive dispatched delivery of two energy blocks E1 and E2 during day D+1: namely, start-up (SU); no-load (NoLoad); minimum run-time (MinRun); transitional ramping (R1 and R2); and shut-down (SD and VSD).

If the RTO/ISO clears generator 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 $\mathbb{PP}_m$ 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 performance cost.

Thus, as depicted in Fig. 8, generator m's avoidable fixed cost includes the energy production cost for the start-up region SU, the no-load region NoLoad, and the shut-down region SD. However, it also includes the region labeled "MinRun" that arises because m has a positive minimum maintainable power-injection level $p^{\min} > 0$.

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 ramp regions R1 and R2, the energy-blocks E1 and E2, and the shut-down region VSD. Note these energy production costs depend on the specific attributes of the dispatched power-path for day D+1, including the dispatched power level at the end-time t^{e} for day D+1 that can vary across the power-paths in $\mathbb{PP}_m$.

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, as illustrated in Figure 7 and Figure 8, 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 during successive operating periods.

For example, dispatchable power producers offering availability of power-path production capabilities for possible dispatch at a grid location b during a future operating period T will typically care about the wear-and-tear cost they could incur for equipment damage if fast ramping is required to match received dispatch set-points. Power customers electrically connected to b will typically care about the degree of flexibility they will have to make idiosyncratic just-in-time power withdrawals at b during future operating periods T to meet their diverse local power requirements.

Moreover, a primary fiducial 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 $T := [t^s, t^e)$, 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 $\mathbf{p}_b(T) := (\, p_b(t) \mid t \in T\,)$ 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* := 12MWh:

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

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

Power-Path (c): Power p is injected (or withdrawn) at b at level p = 1MW during *every other* hour of T, with p = 0MW 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* = 12MWh 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 centrally-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 the start of time-period s_0 is said to be *time-inconsistent* if re-optimization undertaken at the start of some later time-period s_n with $0 < n \leq N$ results in an optimal solution for $(s_n, \ldots, s_N)$ that deviates from $(s_n^* \ldots, s_N^*)$. See [43, 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 LMP(b,T) ($/MWh), will suffice to cover the total fixed and variable cost that suppliers incur in this market.

In the next two subsections 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 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 ($) extracted by buyers and suppliers at e^* from the underlying commodity spot market CSM(Q) for M is as large as possible. However, market efficiency does *not* guarantee that the net surplus ($) 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*[36] 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, unique flavor based on a 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 i might be a monopolist (sole supplier) with respect to "residual demand" customers that other capacity-constrained suppliers are unable to service, enabling supplier i 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 5.3.2. By definition, such coverage does not affect the amount of these sunk costs. Thus, a supplier i at a given time t seeking to maximize his expected net benefit should choose an action compatible with this objective. The greater the expected net benefit, the greater the likelihood that supplier i's sunk costs will be covered in part or in whole. On the other hand, limiting attention to action choices that ensure sunk cost coverage can prevent supplier i from attaining maximum 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

[36] As discussed more carefully in [36], *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.

determined in advance of the delivery of these transacted amounts of A. Avoidable fixed cost essentially always arises for suppliers participating in forward asset markets 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 transacted 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 cost that actually arise for suppliers participating in current 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.

It is difficult to envision how many of these avoidable costs could possibly 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 current 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 supported by cost allocation rules* to ensure supplier revenue sufficiency, i.e., to ensure supplier revenues suffice to cover supplier avoidable fixed costs as well as supplier variable costs; see [17].

Section 7 of this study reviews key features of an alternative *Linked Swing-Contract Market Design* [43] for RTO/ISO-managed wholesale power markets that demonstrates how two-part pricing contracts supported by a principled cost-allocation rule 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 complex organizations. From an external vantage point, the continued reliance of these markets on the DAM/RTM 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.[37] 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 [43, 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 [33] 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*[39] basis with energy in U.S. RTO/ISO-managed DAMs (apart from ISO New England) are Regulation, Spinning Reserve, and Supplemental Reserve [14, 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 operating reserve (MW) for net-load balancing services 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 [14, p. 12].

A major concern regarding these developments is that the RAMP (MW/min) products, together with already instituted CAP (MW) and ENERGY (MWh) prod-

---

[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 power engineers.

[39] Here “co-optimized" means that the conducted DAM SCUC/SCED optimization jointly determines energy dispatch set-points and operating reserve (unencumbered generation capacity) levels for next-day operations. Energy and operating reserve are then both settled by means of prices derived as dual variables for system constraints – power-balance constraints for energy, and reserve requirement constraints for operating reserve. The coupled nature of these constraints results in LMPs ($/MWh) being used to determine prices for operating reserve as well as for energy. See [5] for a more detailed discussion of this point.

ucts, 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 be appropriately measured by means of separate prices assigned to its attributes, treated as independent products. In actuality, the attributes of a power-path $\mathbf{p}_b(\mathrm{T}) := (\, p_b(t) \mid t \in \mathrm{T}\,)$ for an operating period T – such as power-delivery start-time $t^{\circ} \in \mathrm{T}$, power capacity (MW) *profile*, ramp-rate (MW/min) *profile*, power-factor (kW/kVA) *profile*, power-delivery *duration* Δt, and total *grid-delivered* energy (MWh) – are correlated jointly-produced attributes. A change in any one attribute of a power-path scheduled for delivery at a grid location b during a future operating period T 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 $\phi_m(\mathbf{p}_m(\mathrm{T}))$ ($) that a dispatchable power resource m would incur for RTO/ISO-dispatched delivery during T of each of its offered power-paths $\mathbf{p}_m(\mathrm{T})$.

For example, the variable cost $\phi_m(\mathbf{p}_m(\mathrm{T}))$ 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:

$$\phi_m(\mathbf{p}_m(\mathrm{T})) \;\approx\; \mathrm{C}^{\mathsf{CAP}}(\mathbf{p}_m(\mathrm{T})) \,+\, \mathrm{C}^{\mathsf{RAMP}}(\mathbf{p}_m(\mathrm{T})) \,+\, \mathrm{C}^{\mathsf{ENERGY}}(\mathbf{p}_m(\mathrm{T}))\,. \qquad (21)$$

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 $\mathbf{p}_m(\mathrm{T})$.

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 [11, 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 OOM *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 [9], and subsequently withdrawn by FERC in 2017 [10], 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 reported by Hartman [18, pp. 6-7], the original NOPR release encouraged commentators to recommend that unit-commitment costs for other types of generation be incorporated into these energy and operating reserve prices as well.

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 nodal 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 cleared 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 cleared suppliers from receiving appropriate compensation for risk-reduction services. Under an extended-LMP method, a cleared 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 cleared 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 for some future period T. These products are fire-insurance contracts offered to households *in advance* of T at an offer price α (\$). Each contract promises to provide make-whole house repairs below period-T market cost[40] to a household that experiences a fire *during* T. FIC's *insurance pool* consists of all households that purchase these contracts.

However, suppose FIC is *not* permitted to charge an offer price to each household in its insurance pool in advance of T. In order to ensure its solvency, 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 FIC provides to this household during period T. The risk-sharing rationale for setting up an insurance pool thus disappears. The would-be "fire insurance" company FIC is instead incentivized to function as an ordinary *ex post* "fire repair" company with no *ex ante* 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:

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

hence:

[40] Suppose: (i) the number of households in FIC's insurance pool is very large; (ii) each of these households has the same small independent probability $\beta \in (0,1)$ of experiencing a house-fire during T; and (iii) each of these households would have the same house-repair cost HRC (\$) in case of a house-fire. Then, by the Strong Law of Large Numbers, 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 offer price $\alpha := [\beta \times \text{HRC}]$ *in advance of* T; and (b) offering each household in its insurance pool *free* make-whole house-fire repair *during* T.

$$\Delta\pi_A(T) \times a(T) \;=\; c^o \,. \tag{23}$$

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 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.[41]

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 carefully discussed in previous sections, this revenue insufficiency appears to be arising as the result of fundamental conceptually-problematic economic presumptions embedded in their core DAM/RTM Two-Settlement System designs: namely, presumptions (P1)–(P4).

[41] See the rigorous definition **RBM7** for supplier revenue sufficiency provided in Section 5.3.2.

6.6 Ptolemaic Epicycle Market-Design Conundrum

In summary, 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 from heavy reliance on fossil-fuel based power to increasing reliance on renewable power with increasingly diverse types of participants. Specifically, they are hindering the ability of these markets to ensure continual power balance in the face of increasingly volatile and uncertain net loads.

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* [43]. This design, proposed for grid-supported centrally-managed wholesale power markets, has been developed and tested in a series of studies[42] at design readiness levels[43] 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 [43] has been motivated by four fundamental *Power-Market Design (PMD)* principles for grid-supported centrally-managed wholesale power markets, as follows:

---

[42] See [1, 19, 26, 27, 28, 29, 32, 43, 48, 49].

[43] 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 Tesfatsion [43, Sec. 18.3].

[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.

[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 nodal 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 [43] 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 nodal power-production capabilities for possible central dispatch at designated grid locations during designated future operating periods.

The second innovative feature is a fundamental change in the current primary focus of RTOs/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 flexibility (swing) 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 [8], 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 carefully illustrated[44] in [43, 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 [43] involving product definitions, management practices, supply-offer formulations, and settlement processes. This section provides a more careful discussion of these features.[45]

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.

[44] The careful illustration alluded to here is a complete analytical modeling of a "Transitional DAM" SCUC/SCED optimization in computationally tractable Mixed Integer Linear Programming (MILP) form that permits dispatchable power resources to submit reserve supply offers either in an ERCOT 3-part form or in a swing-contract form.

[45] Detailed *analytical* and *computational* formulations for the Linked Swing-Contract Market Design are provided in Tesfatsion [43, Chs. 4-10].

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 $\mathbf{p}_b := (p_b(t) \mid t \in \mathrm{T})$ 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.

Core types of participants in a reserve market M(T) for a future operating period T include *Load Serving Entities (LSEs)* functioning as market intermediaries for power consumers, and *Intermittent Power Resources (IPRs)* whose power 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,[46] and a price-sensitive demand schedule for power-path delivery during T. Non-firmed 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 such as wind farms, PV solar arrays, and hydropower facilities whose power is firmed by storage.

A reserve offer $\mathrm{SC}_m(\mathrm{T})$ 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, $\mathrm{SC}_m(\mathrm{T})$ consists of four m-specified components as follows:

- an offer price $\alpha_m(\mathrm{T})$ 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 $\mathbb{T}^{\mathsf{ex}}_m(\mathrm{T})$ of contract exercise times available to the RTO/ISO between the close of M(T) and the start of T;
- a production possibility set $\mathbb{PP}_m(\mathrm{T})$ 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;

[46] 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.

- a performance payment method $\phi_m(\mathrm{T})$ permitting m to receive compensation *ex pos*t (after T) for any variable cost that m incurs for verified period-T delivery of a power-path in $\mathbb{PP}_m(\mathrm{T})$ in accordance with RTO/ISO dispatch instructions.

Thus, the reserve offer $\mathrm{SC}_m(\mathrm{T})$ permits DPR m to ensure its revenue sufficiency. That is, m can specify the offer price $\alpha_m(\mathrm{T})$ and performance payment method $\phi_m(\mathrm{T})$ appearing in $\mathrm{SC}_m(\mathrm{T})$ 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 bid/offer-based optimization for each reserve market M(T) for a future operating period T. The goal of this optimization is to maximize the expected total net benefit (i.e., total benefit minus total avoidable cost) of M(T) participants by appropriate selection of bids and offers to clear for T. The optimization is conditional on current state conditions, and on forecasts for non-dispatched injections and/or withdrawals of power at each grid delivery location during T. The optimization is subject to nodal power-balance constraints, line capacity limits, and other standard types of system constraints for operating period T.

To maintain the *independence*[47] of the RTO/ISO, all costs incurred for reserve acquisition 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[48] 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;
- the payments (or charges) allocated to non-firmed IPRs based on their *fixed* power injections during T.

Regarding the last-listed cost-allocation category, non-firmed 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

---

[47] 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.

[48] More precisely, cost allocations under the Linked Swing-Contract Market Design are determined in accordance with the *RTO/ISO Cost Allocation Rule* defined in [43, 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 [8]).

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'),... for a sequence of RTO/ISO-managed forward reserve markets M(T''), M(T'),... whose operating periods $T'' \supseteq T'$ ... $\supseteq T$ encompass a *common* designated future operating period T can range from multiple years to minutes, as can the duration of T itself. Linkages among these markets arises because the bids and offers cleared in each successive market affect the perceived desirability or need to clear additional bids and offers in the next. The RTO/ISO carries forward through time a continually updated portfolio of the cleared bids and offers covering a designated future operating period 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 [43] 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 throughout Tesfatsion [43]. These comparisons include test-case performance studies for wholesale power markets operating over 5-bus and 30-bus transmission grids.

The performance findings reported in [43] for relatively small-scale transmission grids are buttressed by favorable 118-bus test-case findings reported in a study by Li and Wang [29]. The key contributions of the latter study are as follows:

- Conceptual development and analytic modeling of a Linked Swing-Contract (SC) Market consisting of a linked RTO/ISO-managed SC Day-Ahead Market (DAM), SC Hour-Ahead Market (HAM), and SC Real-Time-Market (RTM);
- Conceptual development, analytic modeling, and systematic performance testing of an SC offer for privately owned and offered battery storage;
- Use of 118-bus test cases to undertake systematic performance testing for Linked SC DAM-HAM-RTM operations, given high renewable power penetration firmed by privately owned and offered battery storage;
- Use of 118-bus test cases to undertake systematic comparative performance testing for Linked SC DAM-HAM-RTM operations and current U.S. DAM-HAM-RTM operations under high renewable power penetration.

The remainder of this section highlights key similarities and differences for the design of current U.S. RTO/ISO-managed wholesale power markets and the Linked Swing-Contract Market Design 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. To simplify the exposition, 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 as follows:

– Both DAM designs are RTO/ISO-managed markets based on bids and offers;

		Current DAM	SC DAM
Similarities		• Conducted day-ahead to plan for next-day operations • RTO/ISO-managed • Market participants include LSEs, DPRs, & IPRs • Same types of system constraints: Nodal power balance, zonal reserve requirements, line capacity limits, ...	
Differences	Optimization form	SCUC & SCED	Optimal contract clearing
	Settlement	Locational marginal prices	Swing contracts are two-part pricing contracts
	Market payments	Payment for next-day energy before actual energy delivery	Payment for resource availability now & resource performance ex post
	OOM payments	Make-whole payments	No make-whole payments
	Info released to participants	Unit commitments, LMPs, & next-day dispatch schedule	Which swing contracts have been cleared

Fig. 9 High-level comparison of basic market design features: Current U.S. DAMs vs. SC DAMs.

– 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 [43, 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 nodal 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

$$\mathrm{SC}_m(\mathrm{D+1}) \;:=\; \big(\,\alpha_m(\mathrm{D+1}),\; \mathbb{T}_m^{\mathrm{ex}}(\mathrm{D+1}),\; \mathbb{PP}_m(\mathrm{D+1}),\; \phi_m(\mathrm{D+1})\,\big) \qquad (24)$$

is cleared for day D+1 receives as payment its *offer price* α_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.[49]

[49] The *amount* of this risk reduction for day D+1 depends on the flexibility (swing) 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.

– 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 α_m(D+1) and performance payment method ϕ_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;

		Current DAM SCUC	Current DAM SCED	SC DAM Optimization
Similarities		• Both SCUC and swing-contract (SC) market clearing are solved as Mixed Integer Linear Programming (MILP) optimization problems subject to system constraints		
Differences	Objective	Min [Start-up/shut-down costs + no-load costs + dispatch costs + reserve costs + constraint penalties]	Min [Dispatch costs + reserve costs + constraint penalties]	Min [Availability cost + performance cost + constraint penalties]
	Unit commitment constraints	Yes	No	Each DPR includes its unit commitment constraints in its submitted swing contract
	Key RTO/ISO decision variables	Unit commitments	Energy dispatch & reserve levels	Which swing-contracts are cleared
	Settlement	No	LMPs calculated as SCED dual variables	Each cleared DPR receives the offer price it has included in its submitted swing contract

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

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

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 α_m(D+1) and the performance payment method ϕ_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 α_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 ϕ_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 response to 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 $\mathbb{PP}_m$(D+1) that m includes in its submitted supply offer (24).

8 Conclusion: Grids as Flexibility-Support Mechanisms

The basic goal of this study is to contribute to the development of a conceptually-consistent design for U.S. RTO/ISO-managed wholesale power markets enabling a smooth transition of these markets to greater reliance on renewable power with increasingly diverse participants. An essential consideration is whether this development will require fundamental changes in the existing design of these markets.

The conclusion of this study is that fundamental design changes will indeed be needed. As carefully analyzed in Sections 2–6, four conceptually-problematic presumptions are reflected in the DAM/RTM Two-Settlement System design that currently underpins operations for all seven U.S. RTO/ISO-managed wholesale power markets. Briefly summarized, these four problematic presumptions are as follows:

(P1) The basic transacted product is grid-delivered energy;
(P2) Supplier revenue sufficiency can be adequately analyzed using a two-part partition of supplier cost into fixed and variable parts;
(P3) Grid-delivered energy is a commodity that can (and should) be transacted in short-run commodity spot-markets at competitively determined per-unit prices;
(P4) Supplier revenue sufficiency is assured in such markets.

Carefully argued and illustrated counterclaims to each of the four presumptions (P1)-(P4) are provided in Sections 3–5. A high-level summary review of these counterclaims is provided below, with a stress on fundamental market design issues.

The four presumptions (P1)–(P4) convey a static view of U.S. RTO/ISO-managed wholesale power market operations as an orderly successive clearing of a collection of competitive commodity spot-markets distinguished by delivery location and delivery time-period. The reality is profoundly different.

Grid-delivered energy E is measured in megawatt-*hours* (MWh). The delivery of a designated amount E(b,T) of grid-delivered energy at a grid-delivery location b during an operating period T must necessarily be implemented as the accumulated power resulting from the delivery of a *power-path* at b during T, i.e., a sequence of injections and/or withdrawals of power p_b(T) (MW) at b during T (h). As illustrated in Section 5.4, multiple distinct power-paths could potentially be used to carry out the delivery of E(b,T) at b during T. The distinct dynamic attributes of these power-paths could have distinct serious economic and physical consequences for power producers, power distributors, power users, and fiducial power-system operators tasked with ensuring grid reliability.

Stated more precisely, making use of the measurement and economic concepts carefully reviewed in Sections 3–4, grid-delivered energy is a *u*-asset with a standard unit of measurement u = 1MWh for quantity size. However, it is *not* an asset whose attributes can be adequately represented and evaluated as a function of this one size attribute, conditional on delivery location and time. That is, *grid-delivered energy is not a commodity*.

To the contrary, as highlighted and illustrated in Section 5, buyers participating in grid-supported RTO/ISO-managed wholesale power markets typically care deeply about their ability to determine their future power usage in a flexible idiosyncratic just-in-time manner. Moreover, suppliers participating in these markets typically pay careful attention to the manner in which the costs entailed by their supply offers are strongly dependent on the dynamic attributes of the power-path deliveries these supply offers could subsequently entail.

Consequently, calculation of the true benefits and costs associated with power-usage and power-production decisions in these markets is not simply a matter of calculating potential gains and losses in terms of grid-delivered energy amounts E(b,T) (MWh) settled at per-unit prices LMP(b,T) ($/MWh) conditional on delivery location b and operating period T, even though current market rules require buyers and suppliers to express their demand bids and supply offers as if this were the case. Rather, the underlying dynamic reality forces market participants to undertake a more complex dynamic approach to benefit and cost calculation.

Consider, for example, the representation of supply costs for a profit-seeking Dispatchable Power Resource (DPR) who must decide now whether or not to submit a supply offer into a market M(T) that, if cleared, will commit DPR to be available for possible dispatch at a designated grid location b during a future operating period T. Economists are taught to express supplier cost curves as time-less constructions in a quantity-value plane, under an implicit assumption that the quantity axis refers to amounts of a commodity whose units are perfect substitutes. However, this simple approach is not available to DPR.

Rather, as explained and illustrated in Section 5.3, DPR will need to partition his costs into three components: non-avoidable fixed cost ("sunk cost"); avoidable fixed cost; and variable cost. Moreover, the accurate partitioning of these costs into three components will require DPR to consider with care the dynamic realization pattern for these costs before, during, and after period T.

As illustrated in Figs. 7–8 in Section 5.3, this dynamic cost analysis would still generally be necessary even if DPR were to focus attention solely on energy production costs, ignoring the many other types of avoidable fixed cost and variable cost incurred by suppliers in actual RTO/ISO-managed wholesale power markets; see Appendix A.4. This follows because DPR might have to incur energy production costs for purposes such as start-up, ramp-up, no-load, min-run, and ramp-down in order to ensure his availability for dispatch during period T, even if the RTO/ISO subsequently chooses not to dispatch DPR at a non-zero power-level during T.

In past years, with more predictable slower-changing aggregate loads, the mismatch between the static DAM/RTM Two-Settlement System design and dynamic reality was not a threat to reliable grid operations. However, in current times with a rapidly diversifying mix of market participants, this is no longer the case. As detailed in Section 6, the growing mismatch between design and reality is posing an increasing threat to the efficiency, reliability, and resiliency of grid operations.

To achieve a better match of design to reality, U.S. RTO/ISO-managed wholesale power markets will need to be redesigned to permit them to operate as flexibility-support insurance mechanisms well-aligned with the local goals and constraints of market participants. The needed flexibility-support is the ability to service the just-in-time power demands of increasingly diverse customers. The needed insurance is protection against volumetric grid-risk, the systemic risk of grid instability due to physical imbalance between increasingly volatile and uncertain net loads and centrally-dispatched power injections.

To illustrate what might be done, Section 7 reviews a Linked Swing-Contract Market Design [43] proposed for grid-supported centrally-managed wholesale power markets that has specifically been tailored to meet these flexibility and insurance needs. This design has been tested at design readiness levels DRL-1 (conceptual formulation) through DRL-4 (performance testing using moderate-scale computational platforms) in a series of previous studies.[50] A comparative review of this design and ten other proposed market designs at early design-readiness levels can be found in a report [31] prepared by Resources for the Future (RFF) researchers.

What additional design development would be needed to move the Linked Swing-Contract Market Design from DRL-4 to DRL-9 (real-world implementation)? Perhaps the greatest perceived practical drawback of the current design is its requirement that dispatchable power resources m submit swing-contract reserve offers $SC_m(T)$ into reserve markets M(T) for future operating periods T that include

[50] Specifically, see [1, 19, 26, 27, 28, 29, 32, 43, 48, 49].

production possibility sets $\mathbb{PP}_m$(T) providing physical descriptions of their offered dispatchable power-path production capabilities for T. Here is how this challenging requirement was characterized in 2021 in [43, Sec. 5.1,p. 35]:

> "How (this requirement) 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 requirement that were not conceived in 2021: namely, the digital twinning of the production possibility sets $\mathbb{PP}_m$(T). As formally defined in [6, 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 by the organizers of 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)

However, the representation of power-path production possibility sets also raises a fundamental conceptual consideration: 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 Schweppe et al. [39, Appendices] and von Meier [50, Sec. 3.4]?

In a 2022 study, Kirkham et al. [23] attribute critical operational issues arising in electric power systems to a fundamental many-to-one *physical* measurement problem arising for various power concepts. Roughly summarized, the current standardized definitions for active power, reactive power, apparent power, and power factor – expressed in operational measurement form as static Root-Mean-Square (RMS) time-averages for conventionally designated operating periods – can in fact correspond to multiple underlying dynamic realities with distinct consequential effects on the resulting physical operation of these systems.

An analogous finding of the current study and a companion study [46] is that the current reliance of U.S. RTO/ISO-managed wholesale power markets on the DAM/RTM Two-Settlement System results in a fundamental many-to-one *economic* measurement problem for grid-delivered energy.

The DAM/RTM Two-Settlement System requires an amount E(b,T) (MWh) of grid-delivered energy conditional on grid delivery location b and operating period T to be transacted at a single market-determined locational marginal price LMP(b,T) ($/MWh). However, as reviewed at the beginning of this section, grid-delivered energy E(b,T) is not a commodity because: (i) multiple power-paths could be used to deliver E(b,T) at b during T; and (ii) the true valuation that a market participant or system operator assigns to E(b,T) will typically depend strongly on the dynamic attributes of the power-path used to deliver E(b,T) at b during T. Thus, a single realization E(b,T) for grid-delivered energy can correspond to multiple possible valuation-patterns across market participants with multiple distinct effects on subsequent market operations.

A critical unresolved market design issue is therefore as follows: the *physical* many-to-one measurement problem for power concepts highlighted in Kirkham et al. [23] and the *economic* many-to-one measurement problem for grid-delivered energy highlighted in this study are *conjoined* in the DC Optimal Power Flow (DC-OPF) formulations routinely used to implement DAM/RTM SCED optimizations, a core facet of DAM/RTM Two-Settlement System operations.

In conclusion, many serious conceptual and operational issues confront power system researchers who are attempting to redesign U.S. RTO/ISO-managed wholesale power markets. 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.

Appendices A.1–A.6: Quick-Reference & Technical Materials

A.1 Acronyms

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

A.2 Standard Transmission System Terms

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

A.3 Standard Economic Terms

Term	Description
Asset	Anything of durable value, whether physical or financial
Avoidable cost	Cost that can be avoided by *not* committing to undertake a specified type of action
Avoidable fixed cost	Avoidable cost not dependent on exact form of action as long as it has specified type
Benefit (or utility) function	Function measuring the increase in own-welfare attained by a customer from the consumption and/or use of goods and/or services
Commodity	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
Competitive market	Commodity market whose buyers and suppliers are price-takers
Competitive equilibrium	Competitive market price-quantity outcome s.t. aggregate demand=aggregate supply
Consumer	Purchaser of goods/services for direct own consumption/use (no resale)
Contract in firm form	Non-contingent contract whose terms are binding on all parties
Contract in option form	Holder has the right, but not the obligation, to exercise the contract
Customer	Purchaser of goods/services either for direct own consumption/use or for resale
Demand schedule (inverse)	Schedule expressing the maximum Q-unit price a buyer is willing to pay for each additionally demanded unit of a commodity Q
Demand schedule (ordinary)	Schedule expressing the maximum amount of a commodity Q that a buyer is willing to purchase at each successively higher Q-unit price
Efficiency	No wastage of resources
Fixed cost	Cost not dependent on a specific form of action undertaken
Forward market	Transacted amounts and payment obligations for these transacted amounts *occur in advance* of the delivery of these transacted amounts
Futures market	Forward market for a commodity
Good	Exchangeable physical item whose acquisition provides benefit to the procurer
Hedonic pricing	Pricing of a product by means of prices separately assigned to its intrinsic physical attributes and/or its external circumstances
Joint products	Products jointly produced from a given set of inputs
Law of One Price	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.
Marked efficiency	No wastage of opportunity to increase total net surplus for buyer and supplier participants
Net benefit	Benefit minus avoidable cost

Term	Description
Net buyer surplus	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
Net supplier surplus	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
Opportunity cost	Earnings foregone by not committing assets to an alternative next-best use
Pareto efficiency	No wastage of opportunity to increase benefit for some at no cost to others by means of a feasible reallocation of resources
Perfect substitutes	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.
Price-taker	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.
Product	Outcome of a production process
Production process	Process that transforms inputs into one or more outputs
Productive efficiency	No physical wastage of production inputs and/or production outputs
Purchase reservation value	Maximum payment a buyer is willing to make to procure a designated item
Revenue sufficiency	Supplier revenue is sufficient to cover supplier avoidable cost
Risk	Possibility of an adverse deviation from an expected outcome
Risk aversion	Unwillingness to participate in a risky undertaking with zero expected payoff
Sale reservation value	Minimum payment a supplier is willing to accept to supply a designated item
Service	Action taken by an entity that provides benefit to another entity
Spot market	Transacted amounts, payments for these transacted amounts, and deliveries of these transacted amounts *all occur at the same location and time* ("on the spot").
Strategic market advantage	Unintended opportunity for a participant to exploit market rules to gain advantage.
Structural market advantage	Instituted market feature that systematically favors some participants over others
Sunk cost	Non-avoidable fixed cost
Supply offer	Offer to provide an item or service
Supply schedule (inverse)	Schedule expressing the minimum Q-unit price a supplier is willing to accept in payment for each additionally supplied unit of a commodity Q
Supply schedule (ordinary)	Schedule expressing the maximum amount of a commodity Q that a supplier is willing to sell at each successively-higher Q-unit price
Systemic risk	System-wide risk, i.e., correlated risk arising for system operations as a whole
Transaction cost	Avoidable fixed cost incurred to organize a production process
Two-part pricing	Separately-requested compensation for avoidable fixed cost and variable cost
***u*-asset**	An asset with a standard unit of measurement u
Variable cost	Avoidable cost dependent on specific form of an undertaken action (e.g., production *level*)
Volumetric grid risk	Systemic risk arising for a grid due to possible net load imbalance

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-signaled 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

Term	Description
***Acronyms & Generics*:**	
D	Generic symbol for a day
DPR	Dispatchable Power Resource
H	Generic symbol for an hour
IPR	Intermittent Power Resource
LAH(T)	Look-ahead horizon between close of M(T) and start of T
LSE	Load-Serving Entity
M(T)	Swing-contract market for a future operating period T
m	Generic symbol for a DPR
n	Generic symbol for an IPR
p	Generic symbol for a power level (MW)
$\mathbf{p}$	Generic symbol for a power-path
$\mathbf{p}_b(\mathrm{T})$	Generic symbol for a power-path $(p_b(t) \mid t \in \mathrm{T})$
r	Generic symbol for a ramp-rate (MW/min)
SC	Swing contract taking the general form SC $:= (\alpha, \mathbb{T}^{\mathrm{ex}}, \mathbb{PP}, \phi)$
$\mathrm{SC}_m(\mathrm{T})$	SC submitted by a DPR m to a swing-contract market M(T) for operating period T
t^{ex}	Exercise time in an exercise set $\mathbb{T}^{\mathrm{ex}}$
$t_m^{\mathrm{ex}}(\mathrm{T})$	Exercise time in an exercise set $\mathbb{T}_m^{\mathrm{ex}}(\mathrm{T})$
$\mathrm{T} := [t^{\mathrm{s}}, t^{\mathrm{e}})$	Operating period with start-time t^{s} and end-time t^{e}
α	Offer price ($) for a swing-contract SC
$\alpha_m(\mathrm{T})$	Offer price ($) for a swing-contract $\mathrm{SC}_m(\mathrm{T})$
$\Delta\mathrm{T}$	Duration of operating period T, measured in real hourly units (e.g., 0.6h)
ϕ	Performance payment method for a swing contract SC that maps $\mathbb{PP}$ into payments
$\phi_m(\mathrm{T})$	Performance payment method for a swing contract $\mathrm{SC}_m(\mathrm{T})$ that maps each power-path $\mathbf{p}_m(\mathrm{T}) \in \mathbb{PP}_m(\mathrm{T})$ into a dollar payment ($)
***Sets & Subsets*:**	
$\mathbb{B}:=\{1,\ldots,NB\}$	Index set for the buses b of a transmission grid
$\mathbb{C}_j(b)$	Collection of customers serviced by load-serving entity $j \in \mathbb{LSE}(b)$
$\mathbb{L} \subseteq \mathbb{B} \times \mathbb{B}$	Index set for the distinct bus-to-bus line segments ℓ of a transmission grid
$\mathbb{L}_{O(b)} \subseteq \mathbb{L}$	Subset of transmission-grid line segments originating at bus b
$\mathbb{L}_{E(b)} \subseteq \mathbb{L}$	Subset of transmission-grid line segments ending at bus b
$\mathbb{LSE}$	Index set for the load-serving entities j participating in a swing-contract market
$\mathbb{LSE}(b) \subseteq \mathbb{LSE}$	Subset of load-serving entities j in $\mathbb{LSE}$ that service power customers at bus b
$\mathbb{M}$	Index set for DPRs m participating in a swing-contract market
$\mathbb{M}(b) \subseteq \mathbb{M}$	Subset of DPRs m in $\mathbb{M}$ that are electrically connected at bus b
$\mathbb{NG}$	Index set for IPRs n participating in a swing-contract market
$\mathbb{NG}(b) \subseteq \mathbb{NG}$	Subset of IPRs n in $\mathbb{NG}$ that are electrically connected at bus b
$\mathbb{PP}$	Set of dispatchable power-paths $\mathbf{p}$ offered by a swing contract SC
$\mathbb{PP}_m(\mathrm{T})$	Set of dispatchable power-paths $\mathbf{p}_m(\mathrm{T})$ offered by a swing contract $\mathrm{SC}_m(\mathrm{T})$
$\mathbb{P}_m$	Set of feasible sustainable power levels p (MW) for DPR m
$\mathbb{RR}_m$	Set of feasible ramp-rates r (MW/min) for DPR m
$\mathbb{T}^{\mathrm{ex}}$	Set of possible exercise times t^{ex} for a swing-contract SC
$\mathbb{T}_m^{\mathrm{ex}}(\mathrm{T})$	Set of possible exercise times $t_m^{\mathrm{ex}}(\mathrm{T})$ for a swing contract $\mathrm{SC}_m(\mathrm{T})$

A.6 Invertibility of Demand and Supply Functions

The following conditions suffice to ensure an *inverse* demand schedule $\pi := D_j(q)$ for a buyer j, defined as in **CM6**, can be inverted to obtain a well-defined *ordinary* demand schedule $q := D^o_j(\pi)$ for buyer j as defined in **CM3**, and vice versa, where $D_j(q)$ coincides with buyer j's marginal benefit function $MB_j(q)$ as defined in **CM5**. See Tesfatsion [43, Sec. 9.3.4] for extended discussion.

Suppose buyer j has a *benefit function* $B_j(q)$, defined as in **CM4**, that is non-decreasing, differentiable, and *concave* over $q \geq 0$. Evaluated at any Q-demand level $q' \geq 0$, buyer j's marginal benefit $MB_j(q')$ (measured in \$/u) as defined in **CM5** is then the non-negative derivative of buyer j's benefit function $B_j(q)$ with respect to q, evaluated at $q = q'$. The mapping $D_j(q')$ of q' into the non-negative *marginal* benefit evaluation π' (\$/u) $:= MB_j(q') := \partial B_j(q')/\partial q$ 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 $q := D^o_j(\pi)$ for buyer j. In this case, by construction, the Q-unit price π' that satisfies $q' = D^o_j(\pi')$ is the marginal benefit $MB_j(q')$ of buyer j evaluated at the Q-demand level q'.

The following conditions suffice to ensure an *inverse* supply schedule $\pi := S_i(q)$ for a supplier i, defined as in **CM10**, can be inverted to obtain a well-defined *ordinary* supply schedule $q := S^o_i(\pi)$ for supplier i as defined in **CM7**, and vice versa, where $S_i(q)$ coincides with supplier i's marginal cost function $MC_j(q)$ as defined in **CM9**. See Tesfatsion [43, Sec. 8.2] for extended discussion.

Suppose supplier i has a *total avoidable cost function* $C_i(q)$, defined as in **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) as defined in **CM9** is then the non-negative derivative of supplier i's total avoidable cost function $C_i(q)$ with respect to q, evaluated at $q = q'$. The mapping $S_i(q')$ of q' into the non-negative *marginal* cost evaluation π' (\$/u) $:= MC_i(q') := \partial C_i(q')/\partial q$ 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 a *strictly* increasing *ordinary* supply schedule $q := S^o_i(\pi)$ for supplier i. In this case, by construction, the Q-unit price π' that satisfies $q' = S^o_i(\pi')$ is the marginal cost $MC_i(q')$ of supplier i evaluated at the Q-supply level q'.

References

1. Battula, S. (2021). *Transactive Energy System Design for Integrated Transmission and Distribution Systems*, Ph.D. Dissertation, Department of Electrical and Computer Engineering, Iowa State University, Ames, Iowa, February.
2. Battula, S., Tesfatsion, L., McDermott, T.E. (2020). “An ERCOT test system for market design studies,” *Applied Energy* 275, October, 115182.
3. Baumol, W.J., Panzar, J.C., Willig, R.D. (1982). *Contestable Markets and the Theory of Industry Structure*. Harcourt Brace Jovanovich, Inc. New York, NY.
4. Billimoria, F., Mancarella, P., Poudineh, R. (2020). “Market design for system security in low-carbon electricity grids: From the physics to the economics,” OIES Paper: EL 41, The Oxford Institute for Energy Studies, June.
5. Ellison, J.F., Tesfatsion, L.S., Loose, V.W., Byrne, R.H. (2012). “A Survey of Operating Reserve Markets in U.S. ISO/RTO-Managed Electric Energy Regions,” Sandia National Laboratories Report (SAND2012-1000), Sept.
6. Emmert-Streib, F. (2023). “Defining a Digital Twin: A Data Science-Based Unification,” *Machine Learning Knowledge Extraction* 5(3), 1036-1054.
7. FERC (2003). “Notice of White Paper,” *Docket No. RM01-12-000*, U.S. Federal Energy Regulatory Commission (FERC), Washington, D.C., April 28.
8. FERC (2011). “Frequency Regulation Compensation in Organized Wholesale Power Markets,” *Final Rule, Order No. 755*, U.S. Federal Energy Regulatory Commission (FERC), Washington, D.C., October 20.
9. FERC (2016). “Fast-Start Pricing in Markets Operated by Regional Transmission Organizations and Independent System Operators: Notice of Proposed Rule-Making,” *Docket No. RM17-3-000*, U.S. Federal Energy Regulatory Commission (FERC), Wash. D.C., Dec. 15.
10. FERC (2017). “Fast-Start Pricing in Markets Operated by by Regional Transmission Organizations and Independent System Operators: Withdrawal of Notice of Proposed Rule-Making and Termination of Rule-Making Proceeding,” *Docket No. RM17-3-000*, U.S. Federal Energy Regulatory Commission (FERC), Washington, D.C., December 21.
11. FERC (2018). “Uplift Cost Allocation and Transparency in Markets Operated by Regional Transmission Organizations and Independent System Operators: Final Rule,” *Order No. 844*, U.S. Federal Energy Regulatory Commission (FERC), Washington, D.C., April 25.
12. FERC (2019). “Map of Regional Transmission Organizations and Independent System Operators,” U.S. Federal Energy Regulatory Commission (FERC), Washington, D.C. `https://www.ferc.gov/industries/electric/indus-act/rto.asp`
13. FERC (2020). “Participation of Distributed Energy Resource Aggregations in Markets Operated by Regional Transmission Organizations and Independent System Operators: Final Rule,” *Order No. 2222*, U.S. Federal Energy Regulatory Commission (FERC), Washington, D.C., September 17.
14. FERC (2021). “Energy and Ancillary Services Market Reforms to Address Changing System Needs,” *Docket No. AD21-10-000-0*, Staff Paper: U.S. Federal Energy Regulatory Commission, Washington, D.C., September.
15. FERC (2022a). “Shaping the Grid of the Future: RTOs and ISOs,” U.S. Federal Energy Regulatory Commission (FERC), Washington, D.C. Available online: `https://www.ferc.gov/power-sales-and-markets/rtos-and-isos`
16. FERC (2022b). “Modernizing Wholesale Electricity Market Design: Order Directing Reports,” *Docket No. AD21-10-000*, U.S. Federal Energy Regulatory Commission (FERC), Washington, D.C., April 21.
17. Frischmann, B.M., Hogendorn, C. (2015). “Retrospectives: The marginal cost controversy,” *Journal of Economic Perspectives* 29(1), 193–206.
18. Hartman, D. (2017). “Refreshing price formation policy in wholesale power markets,” Study No. 106, *R Street Policy*, August.
19. Heo, D.Y., Tesfatsion, L.S. (2015). “Facilitating appropriate compensation of electric energy and reserve through standardized contracts with swing,” *J. of Energy Markets* 8(4), 93–121.

20. Hsieh, E., Anderson, R. (2017). “Grid flexibility: The quiet revolution,” *Electricity Journal* Vol. 30, 1–8.
21. IEA (2021). *Net Zero by 2050: A Roadmap for the Global Energy Sector*, International Energy Agency (IEA), 4th Revision, Oct. Available online: `https://www.iea.org/`
22. ISONE (2020). *New England Manual for Market Operations*, Manual M-11, Revision 60, ISO New England Inc., Effective Date: October 1.
23. Kirkham, H., Strickland, D., Berrisford, A., Riepnieks, A., Voisine, J., Britton, J. (2022). “Overview of IEEE Standard 1459 Revision,” *IEEE Power & Energy Society General Meeting (PESGM) 2022*, 978-1-6654-0823-3/22. DOI: 10.1109/PESGM48719.2022.9916852
24. Kirschen, D.S., Strbac, G. (2018). *Fundamentals of Power System Economics*, Second Edition, John Wiley & Sons, New York, NY.
25. Li, H., Tesfatsion, L. (2011). “ISO net surplus collection and allocation in wholesale power markets under locational marginal pricing,” *IEEE Trans. on Power Systems* 26(2), 627–641.
26. Li, W. (2023). Market Design Innovations for Decarbonized Grid Operations, Ph.D. Thesis, Department of Electrical & Computer Engineering, Iowa State University, Ames, IA.
27. Li, W, Tesfatsion, L (2018) A swing-contract market design for flexible service provision in electric power systems, Chapter 5 (pp. 105–127) in: Meyn, S., Samad, T., Hiskens, I., Stoustrup, J. (Eds.), Energy Markets and Responsive Grids: Modelling, Control, and Optimization, Vol. 162, IMA Volumes in Mathematics and its Applications, Springer.
28. Li, W, Tesfatsion, L (2015) Market provision of flexible energy/reserve contracts: Optimization formulation, Economics Working Paper No. 15019, Iowa State University Digital Repository, Initial Posting: Nov. 12.
29. Li, W., Wang, Q. (2024). “A linked swing contract market design with high renewable penetration and battery firming,” *IEEE Transactions on Power Systems* 39(2), 2723-2734.
30. Liu, H., Tesfatsion, L., Chowdhury, A.A. (2009). “Derivation of locational marginal prices for restructured wholesale power markets,” *Journal of Energy Markets* 2(1), 3-27.
31. Lo Prete, C., Palmer, K., Robertson, M. (2024). *Time for a Market Upgrade? A Review of Wholesale Electricity Market Designs for the Future*, Report 24-09, Resources for the Future (RFF), Washington, D.C.
32. Ma, S., Wang, Z., Tesfatsion, L. (2019). “Swing contracts with dynamic reserves for flexible service management,” *IEEE Transactions on Power Systems* 34(5), 4024–4037.
33. MISO (2023). “BPM-002: Energy and Operating Reserve Markets,” *Business Practice Manual*, Midcontinent Independent System Operator (MISO), Effective Date: September 30th. `https://www.misoenergy.org/business-practice-manuals/`
34. NAS (2016). *Analytic Research Foundations for the Next-Generation Electric Grid*, National Academies of Sciences, The National Academies Press, Washington, D.C.
35. NAS (2021). The Future of Electric Power in the United States. Washington, DC: The National Academies Press. `https://doi.org/10.17226/25968`.
36. Nicolaisen, J., Petrov, V., Tesfatsion, L. (2001). “Market power and efficiency in a computational electricity market with discriminatory double-auction pricing,” *IEEE Transactions on Evolutionary Computation* 5(5), Oct., 504–523.
37. Oren, S. (2005) Generation adequacy via call options obligations: Safe passage to the promised land. *The Electricity Journal* 18(9), 8-42.
38. Schweppe, F.C., Tabors, R.D., Kirtley Jr., J.L., Outhred, H.R., Pickel, F.H., Cox, A.J. (1980). “Homeostatic utility control,” *IEEE Transactions on Power Apparatus and Systems*, PAS-99, No. 3, May/June, 1151-1163.
39. Schweppe, F.C., Caramanis, M.C., Tabors, R.D., Bohn, R.E. (1988). *Spot Pricing of Electricity*, Fourth Printing (2000). Kluwer Academic Publishers, Boston, MA.
40. Sun, J., Tesfatsion, L. (2010). “DC Optimal Power Flow Formulation and Testing Using QuadProgJ,” Working Paper No. 06014, Dept. of Econ., Iowa State U. Last Revised: March 2010. `https://www2.econ.iastate.edu/tesfatsi/DC-OPF.JSLT.pdf`
41. Sun, Y., Levin, T., Kwon, J., et al. (2021). “Research Priorities and Opportunities in United States Competitive Wholesale Power Markets,” NREL/TP-6A20-77521, National Renewable Energy Laboratory, Golden, CO, May.

42. Tesfatsion, L. (2009). "Auction basics for wholesale power markets: Objectives and pricing rules," *IEEE Power & Energy Society General Meeting*, July 26-30, Calgary, AB, Canada. DOI: 10.1109/PES.2009.5275970.
43. Tesfatsion, L. (2021). *A New Swing-Contract Design for Wholesale Power Markets*, John Wiley & Sons, Inc. (IEEE Press Series on Power Engineering), Hoboken, New Jersey, USA.
44. Tesfatsion, L. (2023). "Agent-based computational economics: Overview and brief history," Chapter 4 (pp. 41-53) in: Ragupathy Venkatachalam (Ed.), *Artificial Intelligence, Learning and Computation in Economics and Finance*, Book Series: Understanding Complex Systems, Springer Cham.
45. Tesfatsion, L. (2023), *Economics of grid-Supported electric power markets: A fundamental reconsideration*, Supporting Document (Attachment A) for Comments (Accession No. 20230117-5051) e-filed to U.S. Federal Energy Regulatory Commission (FERC) for Docket AD21-10-000 ("Modernizing Wholesale Electric Power Market Design"), January.
46. Tesfatsion, L. (2024), "Locational marginal pricing: A fundamental reconsideration," *IEEE Open-Access Journal of Power and Energy* 11, 104-116. DOI: 10.1109/OAJPE.2024.3361751. https://ieeexplore.ieee.org/document/10419119
47. Tesfatsion, L. (2021). "Agent-based platforms for electric power systems: Teaching,training, and research," Panel Presentation, *IEEE Power and Energy Society General Meeting (virtual),* July 26-29, 2021. https://www2.econ.iastate.edu/tesfatsi/ABMPowerSystemsIntro.LTesfatsion.pdf
48. Tesfatsion, L. (2022). "Transitioning to linked swing-contract wholesale power markets for Net-Zero 2050," *FERC Technical Conference 2022 (Virtual): Increasing Real-Time and Day-Ahead Market Efficiency Through Improved Software*, U.S. Federal Energy Regulatory Commission, Washington, D.C., June 21-23. https://www2.econ.iastate.edu/tesfatsi/FERCTechConf2022.LTesfatsionSlideSet.pdf
49. Tesfatsion, L., Silva-Monroy, C.S., Loose, V.W., Ellison, J.F., Elliott, R.T., Byrne, R.H., Guttromson, R.T. (2013). *New Wholesale Power Market Design Using Linked Forward Markets*. Sandia National Laboratories Report (SAND2013-2789). https://www2.econ.iastate.edu/tesfatsi/MarketDesignSAND2013-2789.LTEtAl.pdf
50. von Meier, A. (2024). *Electric Power Systems: A Conceptual Introduction*, 2nd Edition, John Wiley & Sons, Inc. (IEEE Press Series on Power Engineering), Hoboken, New Jersey, USA.