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

Leigh Tesfatsion

Abstract
U.S. centrally-managed 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. 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 transacted product as grid-delivered energy (MWh) competitively priced at designated grid delivery locations during successive operating periods, supported by ancillary services. The study then discusses an alternative conceptually-consistent "linked swing-contract market design" that appears well-suited for the scalable support of increasingly decarbonized grid operations with more active participation by demand-side resources. This alternative design entails a fundamental switch to a dynamic insurance focus on advance reserve procurement permitting continual balancing of real-time net load. Reserve consists of the guaranteed availability of diverse power-path production capabilities for possible centralized dispatch during future operating periods, offered into centrally-managed linked forward reserve markets by two-part pricing swing contracts in firm or option form.

Key words: U.S. RTO/ISO-managed wholesale power markets, grid decarbonization, market design, fundamental conceptual issues, alternative linked swing-contract market design, physically-covered insurance, two-part pricing, swing contracts, power-path production capability sets, digital-twin representation

*ISU Digital Repository WP #22005: Initially posted: 9/13/2022; Latest revision: 4/10/2023. Author Contact Info: Leigh Tesfatsion, Research Professor of Economics, Courtesy Research Professor of Electrical & Computer Engineering, and Professor Emerita of Economics, Iowa State University, Ames, IA 50011-1054, Homepage: https://www2.econ.iastate.edu/tesfatsi/, Email: tesfatsi@iastate.edu; https://www2.econ.iastate.edu/tesfatsi/EconGridSupportedPowerMarkets.SlideSet.LTesfatsion.pdf.

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 diverse 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*: namely, the possible disruption or collapse of grid operations due to real-time imbalance between the injection of power *into* the grid and the withdrawal and/or inadvertent loss of power *from* the grid. Grid power withdrawals occur when the power usage of customers electrically connected to the grid exceeds their use of locally-generated power. Inadvertent grid power losses occur whenever power flows across transmission grid 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 renewable power facilities. The latter facilities include *intermittent power resources (IPRs)*,[2] such as grid-connected wind farms and photovoltaic solar arrays 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 recent initiatives such as FERC Order No. 2222 [14] encouraging more active participation in these markets by aggregations of distribution-level power resources and customers, has increased the uncertainty and volatility of grid *net load*, i.e., customer power withdrawals and inadvertent power losses minus net non-dispatched power injections. 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 types of system security issues [4].

In consequence, as reported in [15], RTOs/ISOs are finding it harder to procure the dependable advance availability of RTO/ISO-dispatchable power-path production capabilities with sufficiently diverse attributes to maintain reliable real-time balancing of net load.[3]

---

[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 [16].

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

The remaining sections of this study are organized as follows. Section 2 presents a careful summary description of the *DAM/RTM Two-Settlement System Design* at the core of current U.S. RTO/ISO-managed wholesale power market operations: namely, the daily RTO/ISO-coordinated operation and settlement of a *Day-Ahead Market (DAM)* and a *Real-Time Market (RTM)*. Section 3 defines and explains basic economic concepts essential for undertaking a fundamental reconsideration of this DAM/RTM Two-Settlement System Design.

The following four conceptually-problematic economic presumptions reflected in the DAM/RTM Two-Settlement System Design are then highlighted and analyzed in Section 4:

Problematic Presumption (P1):
The basic transacted product is grid-delivered energy (MWh), i.e., flows of power (MW) accumulated at designated grid locations during designated operating periods with duration measured in hours (h).

Problematic Presumption (P2):
For careful analysis of supplier revenue sufficiency, it suffices to partition total supplier cost into two components: namely, a "variable" component determined by supplier grid-delivered energy; and a "fixed" component independent of supplier grid-delivered energy.

Problematic Presumption (P3):
Grid-delivered energy (MWh) is a commodity whose energy units u = 1MWh – perfectly substitutable (economically equivalent) conditional on grid delivery location b and operating period T – should be transacted in a competitive spot market at a uniform per-unit locational marginal price LMP(b,T) ($/MWh) determined in accordance with the competitive (marginal cost = marginal benefit) spot-pricing rule.

Problematic Presumption (P4):
Total supplier revenue attained in these competitive commodity spot markets will suffice over time to cover total supplier cost.

Presumptions (P1)–(P4) reflect the misleading view that U.S. RTOs/ISOs are fiduciary managers for weakly-correlated collections of competitive commodity spot markets. The reality is far more daunting: U.S. RTOs/ISOs are fiduciary "conductors" tasked with orchestrating the availability and subsequent possible dispatch of increasingly-diverse dispatchable power resources to service just-in-time power demands of increasingly diverse grid-connected customers while meeting just-in-time power requirements for reliable grid operation.

Section 4 carefully develops and justifies the following four counter-claims (CC1)–(CC4) to the conceptually-problematic presumptions (P1)–(P4):

Counter-Claim (CC1):

Within the context of a grid-supported centrally-managed wholesale power market, grid-delivered energy (MWh) is *not* the basic transacted product. To the contrary, suppliers provide *two distinct types of products*:

Physically-Covered Insurance: *Guaranteed availability* of diverse power-path[4] production capabilities for *possible* central-manager dispatch during *future* operating periods to reduce volumetric grid risk;

Real-Time Power-Path Delivery: *Actual delivery* of power-paths 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):

Within the context of a grid-supported centrally-managed wholesale power market, the total costs of suppliers must be partitioned into *three* distinct components – Unavoidable Fixed Cost ("Sunk Cost"), Avoidable Fixed Cost, and Variable Cost – to permit a conceptually-sound analysis of their revenue sufficiency.

Counter-Claim (CC3):

Within the context of a grid-supported centrally-managed wholesale power market, *grid-delivered energy (MWh) is not a commodity*. Although grid-delivered energy has a standard unit of mesurement $u = 1$MWh, market participants do *not* consider these units u to be perfect substitutes (economically equivalent) conditional on grid delivery location and operating period.

Counter-Claim (CC4):

Grid-supported centrally-managed wholesale power markets are necessarily *forward* markets due to the speed of real-time operations. To ensure revenue sufficiency, each supplier i participating in such a market should be permitted to submit a *two-part pricing*[5] supply offer ensuring *full compensation* for:

(1) any *avoidable fixed cost* incurred for provision of physically-covered insurance for *future* operating periods, *whether or not* supplier i is subsequently dispatched to provide *actual* power-path deliveries during these periods;
(2) any *variable cost* incurred for *actual* power-path deliveries in response to received real-time dispatch instructions.

Section 5 carefully considers how the retention of the four legacy economic presumptions (P1)–(P4) in the core design of current U.S. RTO/ISO-managed whole-

[4] A *power-path* $\mathbf{p}_b(\mathrm{T}) = (\, p_b(t) \mid t \in \mathrm{T}\,)$ is a sequence of injections and/or withdrawals of power $p_b(t)$ (MW) at a *single* grid-location b during a designated time-interval T.

[5] 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 Sec. 4.5 below. 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.

sale power markets is hindering the ability of these markets to transition smoothly to decarbonized grid operations. Indeed, this retention appears to have led to a dangerous conundrum. Recalling the slow and 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), this conundrum can be characterized as follows:

Ptolemaic Epicycle Conundrum for Power Markets ("Onion Problem"):

> Current net-load balancing issues are addressed by instituting a new rule-layer ("epicyle") governing trade and settlement for an additional ancillary-service support product, which in turn gives rise to new net-load balancing issues.

Section 6 briefly reviews the key features of an alternative design for grid-supported RTO/ISO-managed wholesale power markets, consistent with Counter-Claims (CC1)–(CC4), that appears well-suited for the scalable support of increasingly decarbonized grid operations and more active participation by demand-side resources: namely, the *Linked Swing-Contract Market Design* [39]. The latter design calls for a fundamental change in the envisioned main role of the RTO/ISO: namely, a switch *from a static focus* on RTO/ISO successive scheduling of grid-delivered energy amounts for near-term operating periods, supported by proliferating ancillary services, *to a dynamic focus* on RTO/ISO advance procurement of physically-covered insurance guaranteeing the availability of diverse power-path production capabilities for *possible* RTO/ISO dispatch during *future* operating periods.

Concluding remarks are given in Section 7. Quick-reference glossaries and guides for key terms and concepts appearing in the main text of this study are provided in appendices.

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

The development of the legacy core *Two-Settlement System Design* supporting current U.S. RTO/ISO-managed wholesale power market operations can be traced in a series of reports released by the U.S. Federal Energy Regulatory Commission (FERC), culminating in a 2003 White Paper [8].

In [8], FERC envisions grid-delivered energy (power accumulations) at designated grid delivery (pricing) locations during designated operating periods to be the basic transacted product. These grid-delivered energy amounts are to be determined in accordance with a two-settlement system [32] consisting of a daily bid/offer-based RTO/ISO-managed *Day-Ahead Market (DAM)* operating in tandem with a daily bid/offer-based RTO/ISO-managed *Real-Time Market (RTM)*; see Fig. 1.

The overall goal of the DAM/RTM 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 DAM/RTM two-settlement system is designed to be in accordance with the determination of market-clearing prices

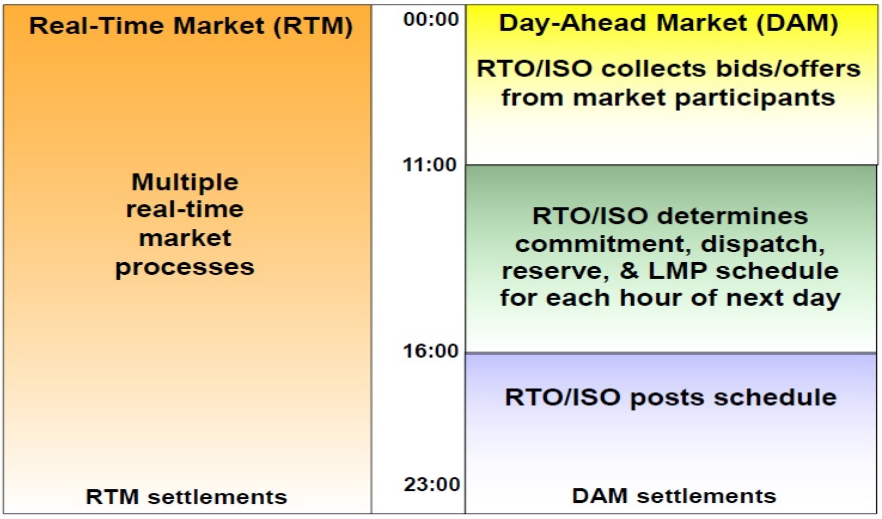


Fig. 1 Simplified depiction of daily operations for an RTO/ISO-managed two-settlement system.

and quantities in *competitive commodity spot markets* to an extent consistent with maintaining the reliable support of a transmission grid susceptible to transmission-line congestion.

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

Load-Serving Entities (LSEs), acting on behalf of managed customers, submit demands bids into the day-D DAM for the purchase of energy at grid delivery locations for each hour H of day D+1. Each such demand bid can take the form of a *fixed (non-dispatched must-service) energy demand*. It can also include or take the form of a *dispatchable price-sensitive energy demand schedule* if: (i) the LSE has installed real-time telemetry permitting the RTO/ISO to incrementally increase or decrease the LSE's energy demand by dispatch signals; and (ii) the LSE can implement any received RTO/ISO dispatch signals by suitable instructions communicated to its managed customers.

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 increase or decrease the unit's energy supply by dispatch signals.

The RTO/ISO conducts a bid/offer-based SCUC/SCED optimization[6] for the day-D DAM in combined or consecutive form. This optimization is conditional on current state conditions, submitted bids and offers, and forecasts for IPR injections and/or withdrawals of power at each grid delivery location during each hour H of day D+1. It is also subject to system constraints that include a net-load balancing requirement at each grid delivery location for each hour H of day D+1.

The optimization determines 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. It also determines anticipated dispatch schedules for price-sensitive energy demands and/or supplies at each grid delivery location for each hour H of day D+1.

Settlements for cleared bids and offers are determined by *locational marginal pricing* [35]; 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 previous SCUC-determined generation-unit commitments, is the dual variable solution for the net-load balancing constraint at b for hour H.[7]

An RTM is a daily collection of sub-markets for near-term future time-periods τ with relatively short durations (e.g., 5 minutes). These RTM sub-markets are cleared by RTO/ISO-managed SCED optimizations conditional on previously-determined unit commitments plus RTO/ISO forecasts for fixed (non-dispatched must-service) energy demands and supplies for τ.[8]

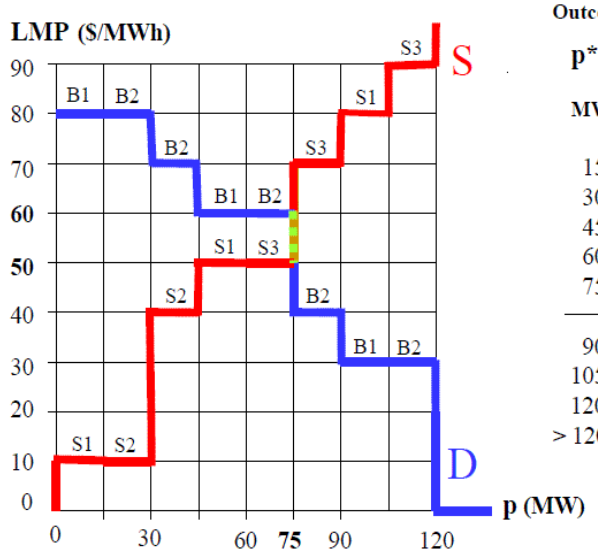
The purpose of these RTM sub-markets (plus supplemental unit-commitment processes) is to permit the successive updating of previously determined optimal SCUC/SCED solutions to take into account updated RTO/ISO forecasts as well as unanticipated changes in other relevant factors. Any adjustments needed in the scheduled energy deliveries determined in the day-D DAM for some hour H of day D+1, as indicated by the solutions for RTM sub-markets conducted after the close of the day-D DAM but prior to hour H, are settled using the LMPs determined in these RTM sub-markets.

Figure 2 illustrates the determination of an optimal demand-equals-supply (D=S) solution for a given hour H on day D+1 by means of a bid/offer-based RTO/ISO-managed DAM SCED optimization conducted during day D. The depicted optimal

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

[7] See [28] for a detailed discussion of the mathematics of LMP determination in U.S. RTO/ISO-managed wholesale power markets.

[8] RTM SCED optimizations are similar in form to DAM SCED optimizations except that RTMs impose stricter restrictions on the submission of *price-sensitive demand bids*. For example, ERCOT restricts RTM submission of price-sensitive demand bids to Qualified Scheduling Entities (QSEs) managing QSE-controllable load sources; see [2, Sec. 4.3].



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.

solution for hour H of day D+1 consists of a *set* of points with a common optimal power level $p^* = 75$ (MW) and a range of optimal price levels LMP* ($/MWh). This optimal price indeterminacy 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 D and S. Also, the depicted optimal solution for hour H of day D+1 is not conditioned on grid delivery location because absence of grid congestion is assumed at this optimal solution.[9]

[9] An optimal 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. As explained in [24], *market efficiency* for a market M means that participating buyers and suppliers are extracting maximum possible total net surplus – i.e., total buyer benefit ($) minus total supplier variable cost ($) – from this participation. As shown in [24, Sec. II, Fig. 2], if LMP separation occurs at an optimal solution for hour H of day D+1, the RTO/ISO itself extracts a non-negative (typically positive) "RTO/ISO net surplus" ($) from market operations at this optimal solution. In this case, accurate determination of market efficiency would require accurate determination of the subsequent use made of this extracted RTO/ISO net surplus, and the effects of this use on the welfare of the market participants.

The depicted aggregate demand and supply schedules D and S are constructed[10] from the LSE demand bids and generation-unit supply offers submitted to the day-D DAM by two LSE buyers (B1,B2) and three generation-unit suppliers (S1,S2,S3).

Note the optimal market-clearing outcomes depicted in Fig. 2 are indeed outcomes for 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). Hence, choice of a power level p is equivalent to choice of a grid-delivered energy-block $p \cdot 1h$ (MWh).

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

- Each *price-sensitive* energy demand (supply) schedule that is bid (offered) by a buyer (supplier) k into an RTO/ISO-managed DAM held on day D for a particular 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 [31].
- 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 $b(k) = b$.

[10] 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). Conversely, 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 (market-clearing) solution points, by definition, are then given by the intersection points of these aggregate D and S schedules with all horizontal and vertical segments included. Compare, for example, the demand bid, supply offer, and market-clearing concepts used by the Midcontinent Independent System Operator (MISO) for its day-ahead and real-time energy markets [31]. For an extended discussion of these fundamental economic concepts, see [38] and [39, Ch. 12].

- **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-paths, without any way to express preferences regarding the dynamic attributes of these power-paths: e.g., capacity profile, ramp-rate profile, and exact timing within H.*

Participants in an RTO/ISO-managed DAM (or RTM) for a future operating period T are assured, by design, that any grid-delivered energy amounts the RTO/ISO announces have been scheduled for T are supported by scheduled transmission capacity. Traders who determine and settle physically-covered bulk energy trades for T through other venues, such as privately-negotiated bilateral trades, must secure transmission-capacity support for these physically-covered trades by self-scheduling them as fixed-form energy bids and offers in a DAM (or RTM) conducted for T. In addition, these traders might need to procure supporting contracts (e.g., Financial Transmission Rights and Contracts-for-Difference) to ensure that the settlement terms they agreed to in these other venues are not disrupted by obligatory DAM/RTM LMP price settlements; see [39, 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 Design*, even though ERCOT (lying entirely within the state of Texas) is not in fact subject to FERC jurisdiction.

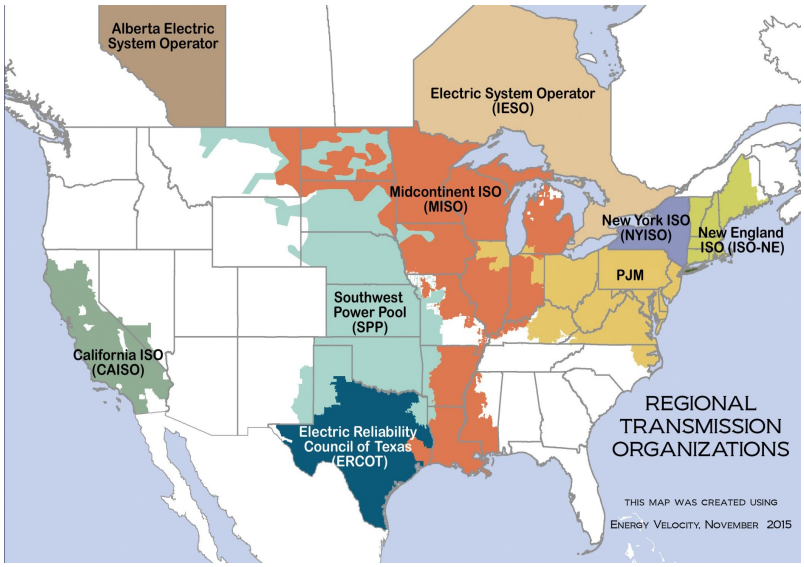


Fig. 3 North American RTO/ISO-managed wholesale power markets. (Public domain: [13])

As seen in Fig. 4, these seven RTOs/ISOs operate over a physical high-voltage AC transmission grid consisting of three separately-synchronized parts.

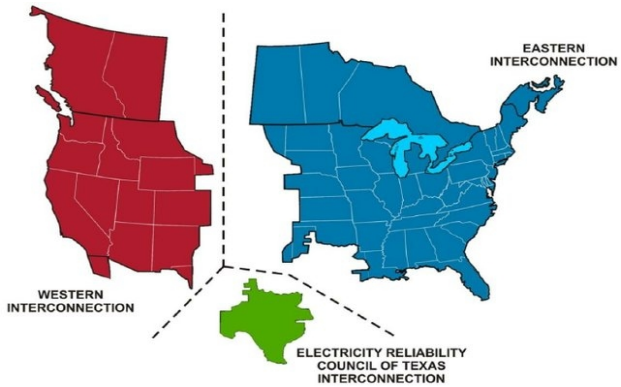


Fig. 4 North American RTOs/ISOs operate over a physical high-voltage AC transmission grid consisting of three separately-synchronized parts. (Public domain image by Jong Suk Kim)

Finally, FERC's proposed Two-Settlement System Design did not include guidance for provision and settlement of ancillary services.[11] Rather, FERC explicitly delegated ancillary service aspects of power system management to the individual states participating in each RTO/ISO [8, p. 11]. Consequently, as reported in [6, Tables 1-2] and [15, Table 1, p. 6], ancillary service procurement and settlement processes differ widely across the seven U.S. RTOs/ISOs.

3 Essential Economic Concepts

3.1 Overview

This section defines, explains, and illustrates economic concepts essential for the careful analysis of grid-supported centrally-managed wholesale power market operations. These concepts will be used throughout the remaining sections of this study.

3.2 Asset Definitions: Unit Measurement Distinctions

Definition D1: A *standard unit of measurement* is 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.

[11] *Ancillary services* are support services for grid reliability [15, Appendix]. Examples include: "black-start" services for restoration of power flow to a collapsed grid; reactive power support for voltage control; and net-load balancing services provided by on-line generation units with unencumbered capacity or by off-line relatively quick-start generation units.

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°F and boils at 212°F; *British Thermal Unit* (Btu), the quantity of heat required to raise by 1°F the temperature of one pound of liquid water currently at the temperature that water has its greatest density ($\approx$ 39°F); *Person-Hour*, a unit of human labor defined to be one hour of work by one person.

Definition D2: An *asset* is 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 value-relevant possibly-correlated attributes.

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

Human Asset Examples: *Health*; *Hand-Grip Strength*; *Intelligence*.

Social Asset Examples: *Beauty*; *Labor Capability*; *Language Fluency*.

Physical Asset Examples: *Hardness*; *Apple*; *Grid-Delivered Energy*; *Honey-Crisp Apple*; *DURACELL AA 1.5v Battery*.

Financial Asset Examples: *Personal Loan*; *Bank-Issued Home Mortgage*; *U.S. Treasury Bill*; *Common Stock Share*.

Definition D3: A *u-asset* is an asset *A* that has a standard unit of measurement.[12]

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.

[12] A *u*-asset is a new asset categorization, introduced in the current study. The set of all *u*-assets (Definition D3) is strictly nested between the set of all assets (Definition D2) and the set of all commodities (Definition D4). As will be carefully discussed in Section 3.5, grid-delivered energy is an example of a *u*-asset that is *not* a commodity. Hence, an explicit recognition of this new *u*-asset categorization appears to be essential for the careful economic analysis of grid-supported electric power market operations.

Physical u-Asset Examples: *Apples*, measured by the number of apples; *Grid-Delivered Energy*, measured by Watt-hour (Wh); *HoneyCrisp Apples*, measured by pound (lb); *DURACELL AA 1.5v Batteries*, measured by number of batteries.

Financial u-Asset Examples: *Common Stocks Listed on the New York Stock Exchange (NYSE)*, measured by ($/share); *Shares of Duke Energy Common Stock (NYSE:DUK)*, measured by number of shares; *U.S. 52-Week Treasury Bills*, measured by number of bills; ...

Examples of Assets that are *Not* u-Assets: *Health*; *Beauty*; *Hardness*.

Definition D4: A *commodity* is a physically-exchangeable *u*-asset Q such that, conditional on location and time, each individual Q-trader k (buyer and/or supplier) considers all Q-units available for trade to be *perfect substitutes* for each other; that is, to have the same economic value.[13]

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

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

Financial Commodity Examples: *Shares of Duke Energy Common Stock (NYSE:DUK)*, measured by number of shares; *U.S. 52-Week Treasury Bills issued 1 January 2003*, measured by number of bills.

Examples of u-Assets that are *Not* Commodities: *Hand-Grip Strength*; *Intelligence*; *Labor-Capability*; *English Language Verbal Fluency*; *Apples*; *Houses*; *Grid-Delivered Energy*; *Fire-Insurance Contracts*; *Mortgages*.

3.3 Unit and Per-Unit Calculations Can Mask Conceptual Error

Preliminaries:

Let R denote the set of real numbers.[14] The standard algebraic operators that act on elements of R include: addition (+); subtraction (-); multiplication ($\times$); division

[13] This "same economic value" assigned to all units u of a commodity Q available for trade at a given location and time can differ across the 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 π does not fall below the common economic value assigned to Q-units by supplier i or exceed the common economic value assigned to Q-units by buyer j.

[14] 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.

($\div$); and equality ($=$). The set R together with its standard algebraic operators are hereafter referred to as the *Real Number System*.

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 listed in Section 3.2 together with real-valued units derived from these seven real-valued SI Base Units by means of standard algebraic operators.

The seven SI Base Units $\{m, kg, s, A, K, mol, cd\}$ 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 of 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 the Planck constant, $\hbar$. The SI Base Unit for electric current is an Ampere (A), defined in terms of the latest internationally agreed-upon value for the electrical charge carried by an electron.

Unit and Per-Unit Calculations Must be Undertaken With Care:

Consider the status (True, False, Undecidable, Undefined, Ambiguous, ...) 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 both the Metric System and 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 about 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 about the Real World (what type of "equality" is "=" meant to signify?).

Regarding Real World ambiguity, consider the following possibilities. The energy (2MWh) on each side of the operator "=" could be stored energy located at a grid location *b at* a particular point in time; thus, 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 power 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 bus location *b during* operating day D as an accumulated flow of power.[15]

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 within 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 every Q trader.*

Status: Undefined statement within the Real Number System (what is u?) and the Metric System (what is a commodity?). True statement (standard economic definition of a commodity) for the Real World.

The above 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 in terms of 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); and 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-distinct 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, and *substantially different* economic effects when used as consumption goods or as inputs to a production process. //

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

[15] In Schweppe et al. [34, fn, p. 1153] and Schweppe et al. [35, Appendix F.1], the proposed *Frequency Adaptive Power Energy Rescheduler (FAPER)* is carefully restricted to *energy loads* ("energy-type usage devices") characterized by: (1) a need for a certain amount of energy over a period of time T in order to fulfill their functions (or purposes); and (2) 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 period of time T in order to full 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 [35] is not addressed by the authors.

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; or (b) equal energy amounts (measured in MWh) that have been grid-delivered at a designated grid location b during a designated operating-period T as the accumulation of two power-injection sequences with distinctly different physical attributes (e.g., different ramp-rate *profiles* during T, different capacity *profiles* during T, different *delivery timing* within T, ...).

Dividing the u-amounts on each side of "=" by a common "base u-value" (for example, "1 pound of apples" for the apple example, or "1MWh of energy" for the grid-delivered energy example), 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 problems highlighted in **Lemmas 3.1–3.3** regarding the representation of real-world quantity amounts as points along Cartesian coordinate axes

suggests the desirability of considerating alternative constructive mathematical approaches permitting "holistic" representations of real-world phenomena and their interactions.[16] Crucial ramifications of **Lemmas 3.1–3.3** specifically for the design and operation of grid-supported RTO/ISO-managed wholesale power markets will be identified and explored in Section 4.

3.4 Market Analysis: Key Concepts

3.4.1 Basic Market Definitions

Definition BM1: A market for an asset A is a ***spot market*** if 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: A market for an asset A is a ***forward market*** if 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: (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: (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: (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: (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 $[AFC_i^o + VC_i(\mathrm{a})]$ of supplier i's avoidable fixed cost (\$) and variable cost (\$) if supplier i selects $a > 0$.

Definition BM7: (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.

[16] See, for example, Sec. 3 titled "Completely Agent-Based Modeling (c-ABM): A Mathematics for the Real World?" in Tesfatsion [41].

Definition BM8: 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: 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.

3.4.2 Commodity Market Definitions

Definition CM1: A ***commodity spot market*** is a spot market for a commodity.

Definition CM2: A ***futures market*** is a forward market for a commodity.

Definition CM3: 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: 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: 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: 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 $D_j(q) = \pi$ (measured in \$/u) that buyer j is willing to pay to procure a *next* Q-unit, given that buyer j has already procured q.[17]

[17] The following regularity conditions are sufficient to ensure an *inverse* demand schedule $D_j(q) = \pi$ for a buyer j, as defined in **CM6**, can be inverted to obtain a well-defined *ordinary* demand schedule $q = D_j^o(\pi)$ for buyer j as defined in **CM3**, and vice versa, where $D_j(q)$ coincides with buyer j's marginal benefit function $MB_j(q)$; see [39, 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 $q \geq 0$. Buyer j's *marginal benefit*, evaluated at any $q' \geq 0$, is then the non-negative derivative of buyer j's benefit function $B_j(q)$ with respect to q, evaluated at $q = q'$. This mapping $D_j(q')$ of q' into a non-negative *incremental* benefit evaluation $\partial B_j(q')/\partial q =: MB_j(q') =: \pi'$ (\$/u) is buyer j's *inverse demand schedule for* Q. Finally, if buyer j's marginal benefit function $MB_j(q)$ is a *strictly* decreasing function of q for $q \geq 0$, a common "diminishing marginal returns" assumption for commodity spot markets, it can be inverted over $q \geq 0$ to give a *strictly* decreasing *ordinary*

Definition CM7: 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: 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 a the total avoidable cost $C_i(q)$ (measured in \$) that supplier i would have to incur to supply q.

Definition CM9: 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: 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 $S_i(q) = \pi$ (measured in \$/u) that supplier i is willing to be paid for a *next* Q-unit, given that supplier i has already supplied q.[18]

---

demand schedule $q = D_j^o(\pi)$ for buyer j. In this case, by construction, the price π' that satisfies $q' = D_j^o(\pi')$ is the marginal benefit of buyer j evaluated at the Q-demand level q'. Economists studying competitive commodity spot markets typically work with ordinary demand schedules mapping prices into quantities because, as will be seen below in definition **CM11**, all buyer participants in such markets are assumed to be price-takers. However, in U.S. RTO/ISO-managed wholesale power markets, demand schedules ("demand bids") are typically expressed in inverse form, as mappings from quantities into prices.

[18] The following regularity conditions are sufficient to ensure an *inverse* supply schedule $S_i(q) = \pi$ for a supplier i, as defined in **CM10**, can be inverted to obtain a well-defined *ordinary* supply schedule $q = S_i^o(\pi)$ for supplier i as defined in **CM7**, and vice versa, where $S_i(q)$ coincides with supplier i's marginal cost function $MC_j(q)$; see [39, 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) defined as in **CM9** is then the derivative of supplier i's total avoidable cost function $C_i(q)$ with respect to q, evaluated at $q' \geq 0$. This mapping of $q' \geq 0$ into a non-negative *marginal* cost evaluation $\partial C_i(q')/\partial q =: MC_i(q') =: \pi'$ (\$/u) is supplier i's *inverse supply schedule for Q*. Finally, if supplier i's marginal cost function $MC_i(q)$ is a *strictly* increasing function of q for $q \geq 0$, a common "increasing marginal cost" assumption for commodity spot markets, it can be inverted over the range $q \geq 0$ to give an *ordinary* supply schedule for supplier i; that is, to give a strictly-increasing function $S_i^o(\pi)$ mapping each non-negative Q-unit price π' (measured in \$/u) into a non-negative Q-supply $q' = S_i^o(\pi')$ (measured in u). In this case, by construction, the Q-unit price π' that maps into q' is supplier i's marginal cost $MC_i(q')$ (measured in \$/u), evaluated at Q-supply level q'. Economists studying competitive commodity spot markets typically work with ordinary supply schedules mapping prices into quantities because, as will be seen below in definition **CM11**, all supplier participants in such markets are assumed to be price-takers. However, in U.S. RTO/ISO-managed wholesale power markets, supply schedules ("supply offers") are typically expressed in inverse form, as mappings from quantities into prices.

3.5 Competitive (MC=MB) Spot-Pricing Requires Priced Assets to be Commodities

CM11 (KEY DEFINITION): Let Q denote a commodity with standard unit of measurement u, and let CSM denote a commodity spot market for Q. Then CSM is a ***Competitive Commodity Spot Market (CCSM)*** if the following conditions hold:[19]

(CCSM1) The participants in CSM 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.[20]

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

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

(CCSM5) The equilibrium concept for CSM is *competitive equilibrium*, defined as follows. Let $q = D^o(\pi) =: \sum_j D_j^o(\pi)$ denote the *(ordinary) aggregate demand schedule* for Q, and let $q = S^o(\pi) =: \sum_i S_i^o(\pi)$ 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 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 condition:

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

$$q^* \;=\; D^o(\pi^*) \;=\; S^o(\pi^*) \;\; . \tag{1}$$

The following lemma, an immediate implication of definitions **CM3–CM10** in Section 3.4.2, establishes an important alternative "marginal pricing" form for the Competitive (D=S) Market Clearing Condition (1).

Lemma 3.4: *Suppose regularity conditions*[21] *hold such that:*

(a) *Each buyer j has an inverse demand schedule $D_j(q_j) = \pi$ that can be inverted to give a well-defined 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., $D_j(q_j) = MB_j(q_j)$.*

[19] See [39, Ch. 12] for a more detailed illustrated presentation of the standard economic definition of a CCSM, including key related concepts such as net surplus extraction and market efficiency.

[20] 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).

[21] Regularity conditions ensuring that properties (a) and (b) in Lemma 3.4 both hold are provided in Footnotes 17 and 18.

(b) *Each supplier i has an inverse supply schedule $S_i(q_i) = \pi$ that can be inverted to give a well-defined 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., $S_i(q_i) = MC_i(q_i)$.*

*Then the Competitive (D=S) Market Clearing Condition (1) in definition **CM11** is equivalent to the following marginal-pricing condition:*

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

$$\pi^* \;=\; \mathrm{MC}_i(q_i^*) \;=\; \mathrm{MB}_j(q_j^*) \tag{2}$$

Outline of Proof for Lemma 3.4: To see the claimed equivalence, given properties **(a)** and **(b)**, apply appropriate *inverse* demand and supply function operations to the terms in condition (1), and apply appropriate *ordinary* demand and supply function operations to the terms in condition (2). //

A CCSM for which the Competitive (D=S) Market Clearing Condition (1) is equivalent to the Competitive (MC=MB) Spot-Pricing Rule (2) will be called a ***Marginal-Pricing CCSM***, or ***MP-CCSM*** for short. MP-CCSMs have a variety of attractive efficiency and optimality properties. Several of these properties are stated below as lemmas for later reference.

Lemma 3.5: *An MP-CCSM is a **uniform-price market** in the following sense. At any competitive equilibrium $e^* = (\pi^*, q^*)$, the same Q-unit price π^* (\$/u) is: (a) **paid** by each inframarginal (cleared) buyer j for each Q-unit that buyer j purchases; and (b) **received** by each inframarginal (cleared) supplier i for each Q-unit that supplier i sells.*

Lemma 3.6: *All fixed cost for each supplier i participating in an MP-CCSM is sunk cost, i.e., non-avoidable fixed cost.*

Proof for Lemma 3.6: By definition, an MP-CCSM is a commodity 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 MP-CCSM and *instantly* cleared (or not cleared) to determine competitive equilibrium outcomes. Thus, no supplier participating in an MP-CCSM is a decision-maker able to avoid (or not avoid) some cost depending on a decision the supplier makes at this given location and time. //

Lemma 3.7: *Revenue sufficiency* holds for an MP-CCSM. That is, at any competitive equilibrium $e^* = (\pi^*, q^*)$ for an MP-CCSM, the total revenue earned by each supplier *i* is sufficient to cover supplier *i*'s total avoidable cost.

Outline of Proof for Lemma 3.7: The equivalent defining conditions (1) and (2) for an MP-CCSM competitive equilibrium $e^* = (\pi^*, q^*)$ imply that the total

(possibly-zero) revenue earned by each supplier i – whether cleared or not – is sufficient to cover the sum (or integral) of the possibly-zero marginal costs that supplier i incurs at e^*.[22] By the supplier cost definitions given in Section 3.4.1, the sum (or integral) of marginal costs for a supplier i at e^* constitutes the *total variable cost* incurred by supplier i at e^*. Moreover, by Lemma 3.6, all fixed cost for supplier i at e^* is sunk cost, i.e., *unavoidable* fixed cost. Thus, the total *avoidable* cost of a supplier i at e^* coincides with supplier i's total *variable* cost. It thus follows from definition **BM7** that supplier i is revenue sufficient. //

Lemma 3.8: *At any competitive equilibrium $e^* = (\pi^*, q^*)$ for an MP-CCSM, the* ***Total Net Surplus (TNS)***[23] *extracted by participating buyers and suppliers is the largest possible Total Net Surplus that buyers and suppliers can extract from the underlying commodity spot market CSM. Thus,* ***market efficiency*** *holds for CSM at e^*, in the following sense: At e^* there is no wastage of opportunity to extract additional net surplus from CSM.* //

Outline of Proof for Lemma 3.8: See Tesfatsion [39, Ch. 12]. //

Lemma 3.8 implies that defining conditions for an MP-CCSM are *sufficient* to ensure market efficiency holds for the underlying commodity spot market CSM at any competitive equilibrium $e^* = (\pi^*, q^*)$. However, these defining conditions are *not necessary* for the market efficiency of the underlying CSM.

For example, it can be shown that market efficiency holds for the underlying CSM if the Competitive (MC=MB) Spot-Pricing Rule (2) is replaced by *any* price-rule PR that satisfies the following three price-rule conditions:

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

PRC(b): All Q-units *failing to trade* in CSM 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.[24]

[22] See footnote-10 in Section 2 and Tesfatsion [39, Ch. 12].

[23] By definition, *Total Net Surplus* attained at any point $e = (\pi, q)$ is the sum of *Total Net Buyer Surplus* and *Total Net Supplier Surplus* attained at e. *Total Net Buyer Surplus* at e is the difference between the maximum amount that buyers would have been willing to pay for q and the amount that they actually pay for purchase of q at e. *Total Net Supplier Surplus* at e is the difference between the payment that suppliers actually receive for sale of q at e and the minimum payment that they would have been willing to receive for sale of q.

[24] Condition **PRC(c)** essentially holds for U.S. RTO/ISO-managed DAM/RTM SCED optimizations in the absence of transmission grid congestion because SCED-determined bus LMPs then collapse to a single uniform LMP across the grid. Conversely, when the grid is congested (i.e., when at least one transmission-line capacity constraint is active), at least some LMP separation occurs across the buses of the grid. 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)**.

An example of a price-rule PR for CSM that satisfies **PRC(a)–PRC(c)**, distinct from the Competitive (MC=MB) Spot-Pricing Rule (2), 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.[25] 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 (MC=MB) Spot-Pricing Rule (2) and *against* the use of a k-discriminatory-price rule for a CSM 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 *strictly* inframarginal 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 *strictly* inframarginal 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.

A second intuitive argument commonly given *in favor of* using the Competitive (MC=MB) Spot-Pricing Rule (2) and *against* the use of a k-discriminatory-price rule for a CSM 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 supply schedule for his strictly inframarginal 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. This incentive only disappears when the supplier has shifted up his reported supply schedule to a point that the discriminatory price assigned to each of his strictly inframarginal Q-units equals the discriminatory price assigned to the marginal (last) sold Q-unit.

[25] 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.

In summary, an MP-CCSM indeed has a number of attractive efficiency and optimality properties. However, the following essential caution – a direct implication of Lemma 3.3 in Section 3.3 – *must* be kept carefully in mind.

> **IMPORTANT CAUTION.** In order for **definitions CM3–CM11**, the **defining conditions for an MP-CCSM**, and the **MP-CCSM properties formalized in Lemmas 3.5–3.8** to be *conceptually meaningful*, the transacted asset *A must be a commodity Q* as defined by **D4** in Section 3.2. That is, the transacted asset A must have a standard unit of measurement u such that, at any given location and time, all A-traders consider all available units u of asset A to be perfect substitutes (economically equivalent). Otherwise, a real-line "quantity axis" *cannot* be constructed in a conceptually coherent manner for any of the functional forms appearing in these definitions, conditions, and properties.

All units u of a commodity Q available for trade at a given location and time *must by definition* be perfect substitutes for each Q-trader at this location and time. Important implications of this observation specifically for the existence of demand and supply functions, the core underpinnings of an MP-CCSM, are stated below in the form of two lemmas.

Lemma 3.9: *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.***

Lemma 3.10 *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 **Lemma 3.9** and **Lemma 3.10** 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**, 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 Honey-Crisp Apple to be a perfect substitute for a Dole Mandarin Orange; that is, suppose the specific apple-versus-orange attribute of a fruit-piece *matters* to buyer j? *In this case, the economic value that buyer j attains from any procured fruit-quantity q will depend on the specific apple-orange composition of q.*

Consequently, the maximum fruit-quantity q that buyer j is willing to purchase at each listed fruit-unit price 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-pieces in the *sealed* bag? Second, given this composition, what will be the likely apple-orange composition of the fruit-quantity q^* that is randomly drawn from the *unsealed* bag if buyer j reports that $q^* = D^o_j(\pi^*)$ is the maximum fruit-quantity he is willing to purchase at the announced price π^*?

The bottom line is that an *ordinary demand* schedule **CM3** is *not* well-defined for fruit for a fruit-buyer j at a given location and time *unless* buyer j considers all pieces of fruit available for purchase at this location and time to be perfect substitutes. Analogous arguments can be used to demonstrate that an *ordinary supply* schedule **CM7** is *not* well-defined for fruit for a fruit-supplier i at a given location and time *unless* supplier i considers all pieces of fruit available for supply at this location and time to be perfect substitutes.

What about *inverse* demand and supply schedules for fruit, defined in accordance with the standard economic definitions **CM6** and **CM10** in Section 3.4.2? Here the import of **Lemmas 3.9–3.10** 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 (i.e., his *purchase reservation value* **BM8**) for a "next" piece of fruit, given he has already procured a fruit-quantity q, without knowing: (i) which specific fruit piece, apple or orange, is to be his "next" procured fruit piece; and (ii) what is the specific apple-orange composition of his already-procured fruit-quantity q?

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

4 Legacy Core Market Design: Fundamental Conceptual Issues

4.1 Counter-Claims (CC1)–(CC4) to Design Presumptions

The DAM/RTM Two-Settlement System reviewed in Section 2, originally proposed by FERC [8] in 2003, constitutes the core design of all seven current U.S. RTO/ISO-managed wholesale power markets. As noted in Section 1, this core design reflects *four conceptually-problematic economic presumptions*, reproduced below in summarized forms:

Problematic Presumption (P1): Presumption that *the basic transacted product is grid-delivered energy (MWh),* i.e., flows of power (MW) accumulated *at* designated grid locations *during* designated time-periods (h).
Problematic Presumption (P2): Presumption that, for careful analysis of supplier revenue sufficiency, it *suffices* to partition total supplier cost into *only two* components: a "variable" component *determined by* supplier grid-delivered energy; and a "fixed" component *independent of* supplier grid-delivered energy.
Problematic Presumption (P3): Presumption that grid-delivered energy (MWh) is a *commodity* whose perfectly-substitutable (economically equivalent) units $u =$ 1MWh (conditional on location and time) should be transacted at a *uniform per-unit energy price ($/MWh) determined in a competitive commodity spot market*.
Problematic Presumption (P4): Presumption that *total supplier revenue* attained in these competitive commodity spot markets will suffice over time to cover *total supplier cost*.

Specific conceptually-problematic aspects of each legacy economic presumption (P1) through (P4) will next be taken up in turn.

4.2 (CC1): Reserve, Not Energy, is the Basic Transacted Product

4.2.1 Presumption (P1) Critique: Two Concerns

Presumption (P1) is consistent with the focus of current U.S. RTO/ISO-managed wholesale power markets on amounts of grid-delivered energy bought and sold at designated grid locations during designated operating periods. This section stresses two important conceptual concerns regarding this presumption.

- *Physical Reliability Concern:* Energy transactions in U.S. RTO/ISO-managed wholesale power markets must be supported by the physical operations of underlying transmission grids. The necessary requirements for the reliable operation of these grids over successive operating periods T cannot be expressed solely in terms of transacted and grid-delivered amounts of energy.
- *Benefit and Cost Valuation Concern:* How power (MW) is injected *at* designated grid locations *during* successive operating periods can matter greatly to RTOs/ISOs, power producers, and power customers. Grid-delivered energy amount is only one of many possible valued attributes of this flow of power.

4.2.2 Physical reliability Concern

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/or inadvertent loss of power from the grid at each point in time.

More carefully stated, continual net-load balance for a transmission grid means that Kirchhoff's Current Law must hold for this grid. Applied to any 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 current I is measured in Amperes (A).

U.S. RTO/ISO-managed wholesale power markets operate over high-voltage *alternating-current (AC)* transmission grids. Consider the analytical modeling developed in [39, Ch. 6 & Sec. 9.2] for an RTO/ISO-managed wholesale power market M(T) for a future operating period T, where M(T) operates over a high-voltage AC transmission grid with buses b in a bus-set $\mathbb{B}$. The participants in M(T) consist of the following entities. For each bus $b \in \mathbb{B}$,

- a collection $\mathbb{M}(b)$ of *dispatchable* generation-units m with unique electrical connection to the transmission grid at transmission bus b;
- a collection $\mathbb{LS}(b)$ of LSEs j, each of whom manages power-usage for a distinct collection $\mathbb{C}_j(b)$ of customers with unique electrical connection to the transmission grid at transmission bus b;
- a collection $\mathbb{NG}(b)$ of *non-dispatchable* generation-units n with unique electrical connection to the transmission grid at transmission bus b.

The relative timing of M(T) and T are depicted in Fig. 5.

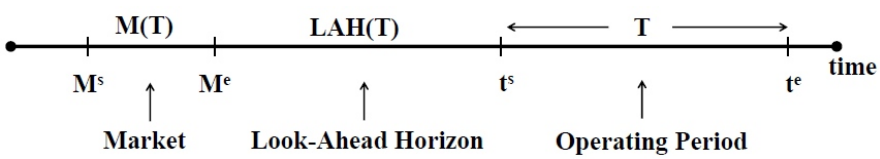


Fig. 5 Time-line for a grid-supported RTO/ISO-managed wholesale power market M(T) conducted for a future operating period T.

The AC power-flow operations of the transmission grid for M(T) are approximated in [39, Ch. 6 & Sec. 9.2] as *direct-current (DC)* power-flow operations. Given this DC power-flow approximation, losses are zero; and 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 $p = V_o \cdot I$, where V_o measured in Volts (V) denotes the constant voltage magnitude assumed for the DC power-flow approximation.[26]

[26] See [36, 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 DC-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 $t \in \mathrm{T}$, 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 $j \in \mathbb{LS}(b)$ plus the *total forecasted fixed customer load* at bus b for customers of the LSEs $j \in \mathbb{LS}(b)$ 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 form [39, Ch. 6 & Sec. 9.2]:

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

$$G_b^{\mathsf{dis}}(t) \;+\; \mathrm{NLPI}_b(t) \;=\; L_b^{\mathsf{dis}}(t) \;+\; \widehat{NL}_b^{\mathsf{f}}(t) \tag{3}$$

where:

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

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

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

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

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

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

The important take-away from the net-load balance constraints (3) 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*[27] for operating-period T.

Specifically, constraints (3) impose joint restrictions on the following four types of power-paths at each bus $b \in \mathbb{B}$ during T:

- the *dispatched* power-path $\mathbf{p}_m^{\mathsf{dis}}(\mathrm{T}) = (\, p_m^{\mathsf{dis}}(t) \mid t \in \mathrm{T}\,)$ at b during T for each *dispatchable* generation-unit $m \in \mathbb{M}(b)$;
- the *dispatched* power-path $\mathbf{p}_j^{\mathsf{dis}}(\mathrm{T}) = (\, p_j^{\mathsf{dis}}(t) \mid t \in \mathrm{T}\,)$ at b during T for the managed customers of each LSE $j \in \mathbb{LS}(b)$;

[27] Recall from Sec. 1 that a *power-path* $\mathbf{p}_b(\mathrm{T}) = (\, p_b(t) \mid t \in \mathrm{T}\,)$ is a sequence of power injections/withdrawals $p_b(t)$ (MW) at a *single* grid-location b *during* an operating-period T.

- the *forecasted fixed* power-path $\widehat{\mathbf{p}}^{\mathrm{f}}_j(\mathrm{T}) = (\,\widehat{p}^{\mathrm{f}}_j(t) \mid t \in \mathrm{T}\,)$ for *fixed power withdrawals* at b during T by the managed customers of each LSE $j \in \mathbb{LS}(b)$;
- the *forecasted fixed* power-path $\widehat{\mathbf{p}}^{\mathrm{f}}_n(\mathrm{T}) = (\,\widehat{p}^{\mathrm{f}}_n(t) \mid t \in \mathrm{T}\,)$ for *fixed power injections* at b during T by each *non-dispatchable* generator $n \in \mathbb{NG}(b)$.

As demonstrated in [39, Ch. 7], given mild regularity conditions and a finite-duration operating period $\mathrm{T} = [t^{\mathrm{s}}, t^{\mathrm{e}})$, it is possible to approximate each of these power-paths as closely as desired by a step-function[28] consisting of a discretized sequence of "energy blocks" E_k for a collection $\mathbb{K}(\mathrm{T}) = \{k_n \mid n = 1, \ldots \mathrm{N}(\mathrm{T})\}$ of half-open sub-periods $k_n = [k^{\mathrm{s}}_n, k^{\mathrm{e}}_n)$ constituting a suitably-refined partition of T. However, as demonstrated in [39, Chs. 7,16] and [26], it is then important to use market optimization formulations expressed in run-time variables to ensure system constraints are imposed with sufficient accuracy. For example, the constraints imposed at the start-time k^{s}_n of each successive sub-period k_n of $\mathbb{K}(T)$ should be expressed in terms of *run-time* min/max limits on power-capacities and ramp-rates for sub-periods $k_{n'} \geq k_n$, i.e., min/max limits that are conditional on the state of each market participant at the start-time k^{s}_n for sub-period k_n, given the specific solution trajectory assumed through sub-period k_{n-1}.

In addition, it could be advantageous, or even necessary, to use partitions of T with *different* sub-period durations Δk for *different* types of market participants in order to capture adequately the specific static and dynamic attributes of their power-paths. For a related discussion, see [20, Sec. 3.1.1].

4.2.3 Benefit and Cost Valuation Concern

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

Let M(T) denote an RTO/ISO-managed wholesale power market M(T) operating over a high-voltage AC transmission grid for a future operating period $\mathrm{T} = [t^{\mathrm{s}}, t^{\mathrm{e}})$. Let $\mathbf{p}_b(\mathrm{T}) = (\,p_b(t) \mid t \in \mathrm{T}\,)$ denote a power-path for T that consists of a sequence of power injections and/or withdrawals $p_b(t)$ (MW) at a particular grid location b during times $t \in \mathrm{T}$. Let $m(b)$ denote an RTO/ISO-dispatchable generation unit electrically connected to b during T, and let $C_j(b)$ denote a collection of customers $c_j(b)$ electrically connected to b during T who are serviced by an LSE $j \in \mathbb{LS}(b)$.

Suppose T has finite duration; and suppose the power-path $\mathbf{p}_b(\mathrm{T})$ has a continuous extension over $\bar{\mathrm{T}} = [t^{\mathrm{s}}, t^{\mathrm{e}}]$, the compact closure of T. It then follows from the discussion in Section 4.2.2 that $\mathbf{p}_b(\mathrm{T})$ can be approximated arbitrarily closely over T by a suitably-constructed step function. Plotted in a time-MW plane, this approximating step-function consists of a finite sequence of one or more "energy blocks" E_k

[28] 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.

for a collection $\mathbb{K}(\mathrm{T}) = \{k_n \mid n = 1, \ldots \mathrm{N}(\mathrm{T})\}$ of half-open sub-periods $k_n = [k_n^{\mathsf{s}}, k_n^{\mathsf{e}})$ that partition the operating period $\mathrm{T} = [t^{\mathsf{s}}, t^{\mathsf{e}})$.

Consequently, if power-path $\mathbf{p}_b(\mathrm{T})$ is dispatched at b during T, its total energy delivery at b during T can be calculated in close approximate form by adding up the energy-blocks for its approximating step-function. However, this in no way guarantees that the actual value assigned to $\mathbf{p}_b(\mathrm{T})$ by a producer $m(b)$, a customer $c_j(b)$, or the RTO/ISO can be expressed solely as a function of this energy delivery.

A generation unit $m(b)$ dispatched by the RTO/ISO to deliver power-path $\mathbf{p}_b(\mathrm{T})$ 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 their power withdrawals at b during T in a flexible *just-in-time* manner to run their personally-owned electrical devices for locally-determined purposes. The value they 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 the guaranteed availability of a *collection* of suitably diverse RTO/ISO-dispatchable power-paths enabling the RTO/ISO to balance uncertain net-power withdrawal at b and other grid locations during T by suitable *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 the above observations is that the 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 transacted energy amounts.

4.3 (CC2): Supplier Cost Analysis Requires a Three-Part Partition

4.3.1 Conceptual Concerns Regarding Supplier Cost Presumption (P2)

Presumption (P2) has two troublesome aspects. First, it reflects the traditional economic overly-simplistic partition of total supplier cost into only two parts: fixed and variable. Second, its focus on delivered energy amount as the sole determinant of supplier variable cost is based on the conceptually-problematic presumption (P1). These two aspects of (P2) are separately addressed in the next two sub-sections.

4.3.2 Three-Part Partitioning of Supplier Cost

For reasons carefully articulated in seminal work by Baumol et al. [3], the traditional economic partition of total cost into two components, “fixed” and “variable,” is conceptually incomplete and empirically problematic for many U.S. industries. Rather, total cost should be partitioned into *three* economically-distinct types of costs:

$$\textit{Total Cost} \;=:\; \textit{Sunk Cost} \;+\; \textit{Avoidable Fixed Cost} \;+\; \textit{Variable Cost} \tag{4}$$

As will be stressed in subsequent sections of this study, the need for the three-part partition (4) 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 $t+\Delta t$. The DM's total cost at time t can be partitioned into three components – sunk cost, avoidable fixed cost, and variable cost – as follows:

$$\begin{aligned} \textit{Sunk Cost} =:\; & \textit{Non-Avoidable Fixed Cost} \\ =:\; & \text{Cost } SC^o \text{ that:} \\ & \text{(i) the DM incurs } \textbf{whether or not} \text{ the DM commits at time } t \\ & \quad \text{to undertaking a type-}A \text{ action at time } t+\Delta t; \\ & \text{(ii) } \textbf{does not} \text{ depend on the specific form of type-}A \text{ action} \\ & \quad \text{the DM undertakes, should the DM choose to commit.} \end{aligned} \tag{5}$$

$$\begin{aligned} \textit{Avoidable Fixed Cost} \;=:\; & \text{Cost } AFC^o \text{ that:} \\ & \text{(i) the DM incurs } \textbf{if and only if} \text{ the DM commits at} \\ & \quad \text{time } t \text{ to undertaking a type-}A \text{ action at time } t+\Delta t; \\ & \text{(ii) } \textbf{does not} \text{ depend on specific form of type-}A \text{ action.} \end{aligned} \tag{6}$$

$$\begin{aligned} \textit{Variable Cost} \;=:\; & \text{Cost } VC(a) \text{ that:} \\ & \text{(i) the DM incurs } \textbf{if and only if} \text{ the DM commits at} \\ & \quad \text{time } t \text{ to undertaking a type-}A \text{ action at time } t+\Delta t; \\ & \text{(ii) } \textbf{does} \text{ depend on specific form } a \text{ of type-}A \text{ action.} \end{aligned} \tag{7}$$

$$\textit{Fixed Cost} \;=:\; [\textit{Sunk Cost} \;+\; \textit{Avoidable Fixed Cost}] \tag{8}$$

$$\textit{Avoidable Cost} \;=:\; [\textit{Avoidable Fixed Cost} \;+\; \textit{Variable Cost}] \tag{9}$$

For illustration, consider a *currently off-line* dispatchable thermal generator m at the start of a day-D U.S. RTO/ISO-managed DAM that the RTO/ISO is conducting to prepare for day-D+1 grid operations. An example of a *sunk cost* for m is an amount of money that m previously spent to purchase a piece of generation equipment that now has no resale value. An example of an *avoidable fixed cost* for m 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[29] 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.[30] An example of a *variable cost* for m is the

[29] 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.

[30] Kirschen and Strbac [23] refer to start-up cost as a "quasi-fixed cost".

fuel cost $VC_m(p,\text{H},\text{D+1})$ that m would have to incur in order to maintain a specific positive power-injection level p during some 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 (4) for total cost.

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

"Sunk Cost is Sunk" Dictum:
A decision-maker at time t who is required to make a decision at time t should make this decision independently of his sunk cost at time t.

Define the DM to be *risk-averse*[31] if the DM is *not* willing to participate:

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

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

$$\text{net benefit} \;=:\; \text{benefit} \;-\; \text{avoidable cost} \tag{10}$$

Note a strictly positive net benefit (10) would permit the DM to pay down at least part of his time-t *sunk* cost, assuming this sunk cost is positive.

Conversely, a risk-averse DM should agree to the commitment at time t *only if* the DM's *expected* net benefit from this commitment is *non-negative*, where net benefit is again defined as in (10). Otherwise, commitment at time t would be an agreement to participate in an undertaking with a strictly negative expected payoff.

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

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

- *if* the DM believes this commitment would result *for sure* in a net benefit (10) for himself that is *strictly positive*;
- *only if* the DM believes his *expected* net benefit from this commitment is *non-negative*, where net benefit is again defined as in (10).

[31] In economic theory, survival is often guaranteed *a priori*, e.g., by postulating zero subsistence needs. The risk aversion of a decision-maker DM is then characterized as a preference attribute of DM: namely, the degree to which DM's utility function expressing preference orderings over possible payoffs exhibits concave curvature properties. In reality, strictly negative payoffs can pose grave *survival* risks for *any* person or commercial entity without deep financial pockets that simply wishes to avoid starvation or insolvency, inducing them to behave in a "risk averse" manner.

Finally, the following rigorous version of definition M10 in Section 3.4.1 will be needed for use in subsequent sections of this study:

Definition R-M10: (Rigorous Form of Definition M10). *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 market participation is sufficient to cover the total avoidable cost that supplier i incurs from this market participation, where avoidable cost is defined as in (9).

4.3.3 Variable Cost Compensation in U.S. RTO/ISO-managed DAM/RTMs

Multiple types of variable costs incurred by power resources participating in U.S. RTO/ISO-managed wholesale power markets are listed in Appendix A.4. In addition to the commonly-considered fuel-cost category, the list includes: 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.

The key concern raised in this sub-section is whether the standardized supply-offer forms required in current U.S. RTO/ISO-managed wholesale power markets permit suppliers to receive appropriate conceptually-coherent compensation for their incurred variable costs. As will next be shown, these supply-offer forms force suppliers to express their variable costs as functions of 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 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[32] 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.

More precisely, apart from ISO New England,[33] these scheduled dispatch set-points determine co-optimized scheduled *maintained* power levels or *maintained* power-slope levels (hence scheduled energy deliveries) *and* scheduled *operating reserve* (*unencumbered generation capacity levels*) for each committed dispatchable generation unit for each hour H of day D+1. These determinations are subject to sys-

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

[33] ISO New England conducts Forward Reserve Market (FRM) auctions for 10-minute contingency reserve and 30-minute supplemental reserve *separately* from DAM energy scheduling; see, for example, [6, Tables 1-2] and [15, Table 1].

tem constraints that include nodal and/or zonal reserve requirements for operating reserve with different availability characteristics.[34]

The 24 hourly supply offers that a dispatchable power resource m submits to a day-D RTO/ISO-managed DAM SCED optimization for the 24 hours of an operating day D+1 are intended to convey variable cost information about m to the RTO/ISO. As detailed in [38], each of these per-hour supply offers is generally[35] required to take a simple step-function form: namely, a possibly-zero fixed maintained power-level demand $\bar{p}$ for hour H (with no associated price information) plus a small upper-limited number N of successive (MW/price)-blocks $n = 1, \ldots, N$ in the MW-\$/MWh plane indicating m's requested per-unit compensation \$/MWh over successively higher non-overlapping intervals of possible maintained power levels p (MW) for H.

Starting at the fixed demand $\bar{p}$ for maintained power, and continuing through some maintained power level $p' > \bar{p}$, the summation of the finite number of (MW/price)-block \$/h-compensations required by m at p', multiplied by 1h, is then considered to be an approximation of m's non-decreasing *variable cost* $VC_m(p')$ (\$) for H, evaluated at maintained power level p' for H. For example, given $N = 3$ and $0 \leq \bar{p} < p_1 < p_2 < p' \leq p_3$, there are three (MW/price)-blocks to consider for the calculation of $VC_m(p')$ (\$): namely, the three (MW/price)-blocks corresponding to the three successive intervals $(\bar{p}, p_1]$, $(p_1, p_2]$, and $(p_2, p']$ of possible maintained-power levels for H. Let π_n (\$/MWh) for $n = 1, 2, 3$ denote m's requested per-unit compensation for grid-delivered energy for each of these three successive intervals. Then, m's variable cost $VC_m(p')$ (\$) evaluated at the maintained power level p' for H is approximated as:

$$VC_m(p') \;=:\; \Big[\pi_1 \cdot [p_1 - \bar{p}] \;+\; \pi_2 \cdot [p_2 - p_1] \;+\; \pi_3 \cdot [p' - p_2]\Big] \times [1h]\,. \qquad (11)$$

The crucial point illustrated by (11) is that the RTO/ISO-managed DAM/RTM two-settlement system forces suppliers to report their variable costs to RTOs/ISOs *as functions solely of grid-delivered energy amounts (MWh).* In actuality, as indicated by the empirical variable-cost examples given in Appendix A.4 (e.g., depreciation of owned machinery, assessed charges for transmission services, ...), many types of variable costs arise for suppliers in part or in whole from their need to provide real-time RTO/ISO-dispatched deliveries of *power-paths* during successive

[34] The types of operating reserve procured on a co-optimized basis with energy in U.S. RTOs/ISOs (apart from ISO New England) include Regulation, Spinning Reserve, and Supplemental Reserve. See [15, Sec. II.A & Appendix].

[35] 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 unit-commitment 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: *Startup Offer* (\$/start); *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) whose power levels p commence at the LSL level.

operating periods to satisfy just-in-time customer power demands and grid reliability requirements.

4.4 (CC3): Grid-Delivered Energy is Not a Commodity

4.4.1 Overview

Presumption (P3) implies that, within the context of a grid-supported U.S. RTO/ISO-managed wholesale power market, grid-delivered energy (MWh) conditional on grid delivery location and operating period is a *commodity* whose perfectly substitutable units (MWh) should be bought and sold in a *spot market* at a *competitively-determined uniform market price* ($/MWh). The fundamental concern regarding presumption (P3) is its serious conceptual inconsistency with the four Counter-Claims (CC1)–(CC4) in Section 1 that are supported with care in [39] and throughout the present study.

The economic background materials presented in Section 3 are used in the following subsections to carefully explain and critique this conceptual inconsistency.

4.4.2 Grid-Delivered Energy is a u-Asset that is Not a Commodity

This sub-section demonstrates that, in contradiction to presumption (P3), energy (MWh) does *not* function as a commodity within the context of current U.S. RTO/ISO-managed wholesale power markets. More precisely, a participant in such a market typically does *not* view a MWh of energy to be a perfect substitute for any other MWh of energy, conditional on delivery location and operating period.

To the contrary, power producers and power customers typically care about the *dynamic* attributes of the power-paths they use to inject/withdraw power at their grid locations during successive operating periods. For example, power producers dispatched to inject power at a grid location *b* during an operating period T might reasonably care about equipment wear-and-tear cost incurred due to the fast ramping required to follow received dispatch set-points. And power customers electrically connected to *b* might reasonably care about the degree of flexibility they have to meet their diverse power requirements during T by just-in-time determined power withdrawals at *b*.

Finally, RTOs/ISOs also typically care about power-paths, not energy deliveries. Given any future operating period T, an RTO/ISO needs availability of dispatchable power-paths with diverse dynamic attributes for possible just-in-time RTO/ISO dispatch during T to ensure continual net-load balancing.

Thus, if a producer, customer, or RTO/ISO were asked to assign a monetary value to a specific amount of grid-delivered energy E^* (MWh) at a specific grid location *b* for a specific future operating period T, typically they would not be able to do so without knowing the power-path to be used for this delivery. For example, what value would each of these entities assign to each of the following four power-path

options $\mathbf{p}_b(\mathrm{T}) = (\, p_b(t) \mid t \in \mathrm{T}\,)$ for the injection of power (MW) at a specific grid location b during a specific future 24-hour operating period T, where each power-path option accumulates to the *same* amount E* = 12 (MWh) of grid-delivered energy:

Power-Path (a): Power (MW) is injected at b at level $p = 24$ during the *first half-hour* of T and at level $p = 0$ during the remainder of T.
Power-Path (b): Power (MW) is injected at b at level $p = 0.5$ throughout T.
Power-Path (c): Power (MW) is injected at b at level $p = 1$ during *every other* hour of T, with $p = 0$ during the remaining hours of T.
Power-Path (d): Power (MW) is injected at b during T in a flexible manner that is entirely up to the entity, apart from the requirement that the resulting total energy delivery must equal E* = 12 (MWh).

Note that 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/withdrawal at b during T subject only to the delivered-energy requirement E* = 12 (MWh) since option (d) encompasses all such options.

The clear implication of these examples is that the product "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. Hence, recalling the important caution **CCSM4** carefully justified in Section 3.5, the following constructions essential for locational marginal pricing are conceptually ill-defined for this product: demand schedule (ordinary or inverse); supply schedule (ordinary or inverse); marginal cost function; and marginal benefit function. Attempts to justify LMP settlements for the DAM/RTM two-settlement system by pointing to the efficiency and optimality properties of competitive commodity spot markets satisfying the competitive (MC=MB) spot-pricing rule (2) are therefore not conceptually supportable.

Many products that are not commodities are successfully transacted in real-world markets. Indeed, product innovation commonly occurs through continual striving to produce *variations* of existing products that permit at least temporary market advantages. Grid-supported centrally-managed wholesale power markets are necessarily *forward* markets due to the speed of real-time operations. The key to the conceptually-consistent design of such markets is to consider with care how real-world *forward* markets transact products whose units are *not* viewed as perfect substitutes by market participants. As will be seen in Section 6, the short answer is *appropriate contract design*.

4.4.3 Conceptually Unsupportable Use of Spot-Market Pricing Results in Time-Inconsistent Settlements

A second fundamental conceptual concern regarding presumption (P3) is that a market process M(T) conducted within a grid-supported RTO/ISO-managed wholesale power market for an operating period T *cannot be a spot market* due to the speed of real-time grid operations. That is, *M(T) cannot coincide with T*.

Indeed, 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); see Fig. 5. 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[36] 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[37] 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.

4.5 (CC4): Supplier Revenue Sufficiency Requires 2-Part Pricing

Presumption (P4) implies that the revenues earned by suppliers from participation in a *Competitive Commodity Spot Market (CCSM)* will suffice over time to cover their total cost, i.e., the sum of their total fixed cost and total variable cost. However, presumption (P4) is false for a CCSM.

The final defining assumption **(A.7)** given in Section 3.5 for a CCSM implies that all suppliers participating in a CCSM have *zero* avoidable fixed cost; that is, all of their fixed cost is sunk cost, not avoidable cost. However, there is nothing in the seven defining assumptions **(A.1)-(A.7)** for a CCSM that guarantees suppliers will be able to cover any or all of their *sunk* costs.

For example, by derived property CCSM3 in Section 3.5, note that *market efficiency* holds at any competitive equilibrium point $e^* = (\pi^*, q^*)$ for a CCSM in the sense that total net surplus at e^* is as large as possible for the underlying commodity spot market CSM. However, referring to the precise definition of market efficiency given in Footnote 23, it is seen that market efficiency does *not* guarantee that the revenues received by participating suppliers cover *any* of their *sunk* costs.

This is a special case of a broader economic fact: There is *no economic efficiency justification* for instituting market rules that ensure suppliers are reimbursed for sunk

[36] These settlements include *out-of-market (OOM)* make-whole payments for partial reimbursement of avoidable fixed costs (e.g., start-up 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.

[37] A multi-stage optimization problem that jointly determines an optimal solution $(s_0^*, s_1^*, \ldots, s_N^*)$ for successive time-periods $(s_0, s_1, \ldots, s_N)$ is said to be *time-inconsistent* if re-optimization undertaken at the beginning of some later time-period s_n with $0 < n \leq N$ results in an optimal solution for $(s_n, \ldots, s_N)$ that deviates from $(s_n^* \ldots, s_N^*)$. See [39, Sec. 10.2].

costs. As stressed by the "Sunk Cost is Sunk" dictum in Section 4.3.2, sunk costs are already incurred costs (using up of resources) that suppliers cannot avoid by current or future decisions; hence, sunk costs should have no effect on these decisions.

Indeed, the only way a supplier *i* can ensure coverage of some or all of his sunk cost through a market participation is if he has some type of *structural or strategic market advantage*[38] relative to other actual or potential suppliers that reduces or eliminates the ability of these other suppliers to compete for supplier *i*'s customers if supplier *i* attempts to charge these customers for sunk costs. Examples of situations giving rise to supplier market advantage include:

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

Conversely, buyers participating in a "perfectly contestable" market have *no* incentive to compensate participant suppliers for sunk costs. Roughly defined, a market is said to be *perfectly contestable* if any participating supplier charging a product price that results in revenues strictly exceeding the product's avoidable cost of production can be successfully challenged and replaced by an existing or newly entering rival supplier able to charge a lower price for the same product. See Baumol et al. [3].

What about supplier revenue sufficiency as defined in Section 4.3.2; that is, the ability of suppliers to earn sufficient revenues over time to cover their *avoidable fixed cost* as well as their variable cost? As seen in Section 3.5, suppliers participating in spot markets have no avoidable fixed cost. In contrast, avoidable fixed cost essentially always arises for suppliers participating in forward market settings because avoidable fixed cost includes opportunity cost, i.e., earnings foregone by not committing assets to an alternative next-best use.

As discussed in Section 4.4.3, market processes conducted within the context of RTO/ISO-managed wholesale power markets operating over high-voltage AC transmission grids are necessarily *forward* markets, not spot markets, due to the speed of

[38] As discussed more carefully in [33], *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.

real-time grid operations. Various types of avoidable fixed cost that arise for suppliers participating in U.S. RTO/ISO-managed wholesale power markets are listed in Appendix A.4. This list includes: capital investment cost; transaction cost; opportunity cost, and unit commitment cost. Specific examples are given for each type of avoidable fixed cost. It is difficult to conceive how these avoidable fixed costs could be expressed in an empirically credible manner as functions solely of grid-delivered energy amounts. Yet, this is what would be needed in order for suppliers participating in RTO/ISO-managed DAMs/RTMs to be assured coverage of their avoidable fixed costs solely through some form of extended LMP pricing mechanism.

Fortunately, this is not necessary. Forward markets instituted in other industries routinely rely on *two-part pricing contracts* to ensure supplier revenue sufficiency, i.e., supplier revenues sufficient to cover supplier avoidable fixed costs as well as supplier variable costs. Section 6 of this study reviews key features of an alternative *Linked Swing-Contract Market Design* [39] for RTO/ISO-managed wholesale power markets that demonstrates how two-part pricing contracts could also advantageously be introduced in these markets to ensure supplier revenue sufficiency.

5 Legacy Core Design: Roadblock for Grid Decarbonization

5.1 Overview

U.S. RTO/ISO-managed wholesale power markets are large complex organizations. From an external vantage point, the continued reliance of these markets on the legacy core 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 briefly discusses several external indicators in support of this concern; a more detailed discussion of these indicators is given in [39].

5.2 Proliferation of Participation Models

The continued focus of U.S. RTO/ISO-managed Day-Ahead Markets (DAMs) and Real-Time Markets (RTMs) on energy as the key transacted product, as reflected in legacy economic presumption (P1), appears to be resulting in a proliferation of *participation models* functioning as artificial market entry barriers.[39]

More precisely, to participate in these DAMs/RTMs, a power resource must be classified in accordance with a designated taxonomy of participation models, each with its own eligibility rules and performance requirements. At the top of this taxonomy are two categorizations: "Energy (MWh)" and "Operating Reserve (MW)." The latter category consists of various forms of unencumbered generation capacity (MW) distinguished by availability characteristics; see Section 4.3.3.

However, this entire DAM/RTM taxonomy is conceptually problematic because, as discussed with care in [39, Sec. 3.2.1], "Energy" and "Operating Reserve" are

[39] For example, see the complicated eligibility requirements that power resources must satisfy to participate in MISO DAM/RTM processes as variously defined types of "qualified resources," covered in MISO's 278-page Energy and Operating Reserve Markets manual [30].

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.

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.

5.3 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.[40]

Three types of operating reserve procured on a *co-optimized* basis with energy in U.S. RTO/ISO-managed DAMs (apart from ISO New England) are Regulation, Spinning Reserve, and Supplemental Reserve [15, 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 participa-

[40] In economics, two or more products are said to be *joint products* if their production results jointly from the use of *common* inputs in a production process. Operating reserve and scheduled generation dispatch are *rival* joint products, meaning that – for any given level of inputs (here generation capacity) – an increase in one of the joint products requires a decrease in the other. The use of a single fuel source to co-generate electricity and steam is an example of a *non-rival* joint-production process well known to engineers.

tion. Indeed, net loads for CAISO and other U.S. RTOs/ISOs are already exhibiting more frequent and dramatic down/up ramping swings. Consequently, the current heavy reliance on unencumbered generation capacity (MW) for operating reserve is becoming increasingly risky.

Industry and academic researchers are thus exploring the possible introduction of new types of "flexibility products" to facilitate the balancing of more uncertain and volatile net loads. For example, FERC has approved proposed ramp flexibility products for CAISO, MISO, and SPP; see [15, p. 12].

A major concern regarding these developments is that the newly approved RAMP (MW/min) products, together with already instituted CAP (MW) and ENERGY (MWh) products, are *not* independently produced products that can be *separately* transacted at *separately* determined prices in a conceptually consistent manner. To the contrary, they are the *correlated* attributes of *individual* power-paths, hence *joint* products. See Footnote 40.

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 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:

$$\phi_m(\mathbf{p}_m(\mathrm{T})) \;\approx\; \mathrm{C}^{\mathsf{CAP}}(\mathbf{p}_m(\mathrm{T})) \,+\, \mathrm{C}^{\mathsf{RAMP}}(\mathbf{p}_m(\mathrm{T})) \,+\, \mathrm{C}^{\mathsf{ENERGY}}(\mathbf{p}_m(\mathrm{T}))\,. \qquad (12)$$

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 (12) are in fact *highly-correlated* attributes of a *single* power-path $\mathbf{p}_m(\mathrm{T})$.

5.4 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 [12, Sec. I.2] on "'uplift transparency," released in 2018, directs RTOs/ISOs to provide a more transparent monthly reporting of "uplift payments" characterized as follows:

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

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

In contrast, RTOs/ISOs have instituted various types of *OOM make-whole payments* for suppliers participating in DAMs/RTMs whose market-determined revenues fail to provide coverage for certain types of avoidable fixed cost. Examples include ERCOT's payments for accumulated power usage (energy) required for start-up, and for the maintenance of an on-line state at a minimum possible maintained power-injection level; see Footnote 35.

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 the maintenance of an on-line state by no means provide *full* coverage for *all* supplier avoidable fixed costs; see the avoidable fixed cost examples listed in Appendix A.4.

On the other hand, various OOM make-whole payment methods proposed to ensure better coverage of supplier costs have tended to blur the operationally-critical distinctions among sunk cost, avoidable fixed cost, and variable cost discussed in Section 4.3. 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. As reviewed by Hartman [18, pp. 6-7], the original NOPR release encouraged commentators to suggest that unit-commitment costs for other types of generation should be incorporated into these prices as well. Moreover, industry and academic researchers are continuing to explore extended-LMP methods for broader-based incorporation of avoidable fixed costs into DAM/RTM energy prices, i.e., into LMPs ($/MWh).

A key conceptual argument against the incorporation of supplier avoidable fixed costs into DAM/RTM *energy* prices is as follows. These avoidable fixed costs are insurance costs, not production costs. That is, they are the costs that suppliers incur to be able to fulfill their contractual commitments to provide *availability* of power-paths for *possible* RTO/ISO-dispatch during *future* operating periods. This *assured availability*, in and of itself, provides a critically important service: namely, reduction of volumetric grid risk for these future operating periods. A supplier should be compensated for providing risk-reduction service (dispatchable power-path availability) for a future operating period T *whether or not* the RTO/ISO subsequently chooses to dispatch this supplier for *actual* power-path delivery during T.

However, incorporation of supplier avoidable fixed costs solely into DAM/RTM *energy* prices for a future operating period T would prevent these suppliers from receiving compensation for period-T risk-reduction services per se. Rather, suppliers would only receive compensation for period-T services if they were dispatched for *energy deliveries* during T.

Consider the following analogous situation. Suppose a fire-insurance company FIC is interested in providing risk-reduction products to households for some future period T. These risk-reduction products are fire-insurance contracts sold to households *in advance of T* that promise to provide make-whole house repairs in case of a period-T house-fire. Define company FIC's *insurance pool* to be the subset of households that purchase a fire-insurance contract from FIC in advance of T. Suppose company FIC is *not* permitted to require each household in its insurance pool to pay a common "premium payment" ($) *whether or not* the household experiences house-fire damage during T.[41]

To stay in business, company 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. Thus, why limit house repairs to *burned* houses? The would-be "fire insurance" company FIC is thus incentivized to function as an ordinary "home repair" company with no provision of risk-reduction services.

[41] Suppose: (i) the number N of households in FIC's insurance pool is large; (ii) each household in this insurance pool has the same small independent probability $\beta \in (0,1)$ of experiencing a house-fire during T; and (iii) each household in this insurance pool would have the same house-repair cost HRC ($) in case of a house-fire. Then, by the Law of Large Numbers, the FIC can "almost surely" guarantee full coverage of its actual total house-fire repair cost *during* T by: (a) requiring each household in its insurance pool to pay a relatively small premium $\beta \times$ HRC *in advance of* T; and (b) offering each household in its insurance pool *free* make-whole house-fire repair *during* T.

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:

$$\pi'_A(T) \times a(T) \;=\; c^o \;+\; \pi_A(T) \times a(T)\,, \tag{13}$$

hence:

$$\Delta\pi_A(T) \times a(T) \;=\; c^o\,. \tag{14}$$

However, as detailed in Section 4.3.2, an avoidable fixed cost is defined to be 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, as stressed in Section 4.4.2, grid-delivered energy and operating reserve do *not* function as commodities within the context of a co-optimized DAM/RTM. Hence, the uniform "unit prices" derived in these market processes for grid-delivered energy and operating reserve, conditional on a given delivery location and operating period, are conceptually-problematic constructs with no efficiency or optimality justification. The incorporation of an avoidable fixed cost into such a presumed uniform market price would thus simply result in a double-layered price distortion.

5.5 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 market participation is sufficient to cover the total avoidable cost that supplier i incurs from this market participation. See the careful definition of this concept provided in Section 4.3.2.

The proliferation of OOM make-whole payments reported in Section 5.4 indicates that U.S. RTO/ISO-managed wholesale power markets are not currently ensuring revenue sufficiency for their participant suppliers. As discussed in previous sections of this study, and addressed in greater detail in [39], this revenue insufficiency appears to be arising as the result of fundamental conceptually-problematic economic presumptions embedded in their legacy core DAM/RTM two-settlement system designs: namely, presumptions (P1)–(P4).

5.6 Ptolemaic Epicycle Market-Design Conundrum

Participation models, flexibility products, OOM make-whole payments, and revenue insufficiency concerns all appear to be proliferating in current U.S. RTO/ISO-managed wholesale power markets. This proliferation suggests these markets could be caught up in a dangerous market-design conundrum, 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 in Section 4.3.2 – correction of the fundamental conceptual inconsistencies in the core design principles is persistently deemed to be too costly to correct.

6 An Alternative Linked Swing-Contract Market Design

6.1 Overview

Previous sections of this study support the contention that the four conceptually-problematic economic presumptions (P1)–(P4) reflected in the legacy DAM/RTM two-settlement system design at the core of current U.S. RTO/ISO-managed wholesale power market operations are hindering the smooth transition of these markets to lower-carbon grid operations with increasingly diverse participants.

This section briefly reviews the innovative aspects and key features of an alternative design for grid-supported centrally-managed wholesale power markets, referred to as the *Linked Swing-Contract Market Design* [39], that appears better suited for the scalable support of these objectives.[42]

[42] Details regarding the development and testing of the Linked Swing-Contract Market Design can be found in the following studies:[1, 19, 25, 26, 27, 29, 37, 39, 40].

The Linked Swing-Contract Market Design is based on the following four conceptually-consistent economic principles [EP1]-[EP4] that were presented in preliminary form as Counter-Claims (CC1)–(CC4) in Section 1:

[EP1]: Within the context of a grid-supported centrally-managed wholesale power market, grid-delivered energy (MWh) is *not* a commodity because the cost-benefit valuations that participants assign to any given energy delivery E^* (MWh) at a grid location *b* during an operating period T typically depend on the *dynamic* attributes of the power-path that *delivers* E^* at *b during* T.

[EP2]: Any market process carried out within a grid-supported centrally-managed wholesale power market must necessarily be a *forward* market process M(T) for a *future* operating period T due to the speed of real-time grid operations.

[EP3]: *Dispatchable Power Resources (DPRs)* participating in a forward market process M(T) for a future operating period T within a grid-supported centrally-managed wholesale power market provide *two distinct types of products*:

Physically-Covered Insurance: *Guaranteed availability* of diverse power-path production capabilities for *possible* central-manager dispatch during the *future* operating period T to protect against volumetric grid risk;

Real-Time Power-Path Delivery: *Actual delivery* of power-paths in response to central-manager dispatch signals received *during* operating period T to meet just-in-time customer power demands and grid reliability requirements.

[EP4]: Each DPR in **[EP3]** can use a *two-part pricing contract* to submit to M(T) a coupled offer of physically-covered insurance for T and real-time power-path delivery during T that ensures revenue sufficiency for the DPR, as follows:

- The two-part pricing contract can include a DPR-specified *offer price* to compensate the DPR *prior to T* for any *avoidable fixed cost* the DPR must incur to guarantee its offered physically-covered insurance for T;
- The two-part pricing contract can include a DPR-specified *performance payment method* to compensate the DPR *after T* for any *variable cost* the DPR incurs for power-path delivery during T carried out in response to received central-manager dispatch signals.

Consequently, under this alternative design there is no need for separate or co-optimized *energy* markets. Moreover, there is also no need to rely on *out-of-market (OOM) make-whole payments* in order to ensure supplier revenue sufficiency.

6.2 Innovative Aspects and Basic Features

One innovative aspect of the Linked Swing-Contract Market Design is the conceptualization of a "power-path" for a grid-supported centrally-managed wholesale power market. This conceptualization permits the design of market rules and contractual bid/offer forms to be considered from the distributed vantage points of market participants as well as the centralized vantage point of the system manager.

Key Definition: A *power-path* $\mathbf{p}_b(\mathrm{T}) = (\, p_b(t) \mid t \in \mathrm{T}\,)$ for a grid-supported centrally-managed wholesale power market is a sequence of injections and/or withdrawals of power $p_b(t)$ (MW) at a *single* grid-location b during a designated time-interval T.

For example, Fig. 6 depicts the dispatchable power-path production capabilities of a dispatchable power resource m participating at grid-location b_m in a grid-supported RTO/ISO-managed wholesale power market. The figure highlights one of multiple possible power-paths that m could deliver to the grid at b_m during operating period T in response to dispatch signals received from the RTO/ISO.

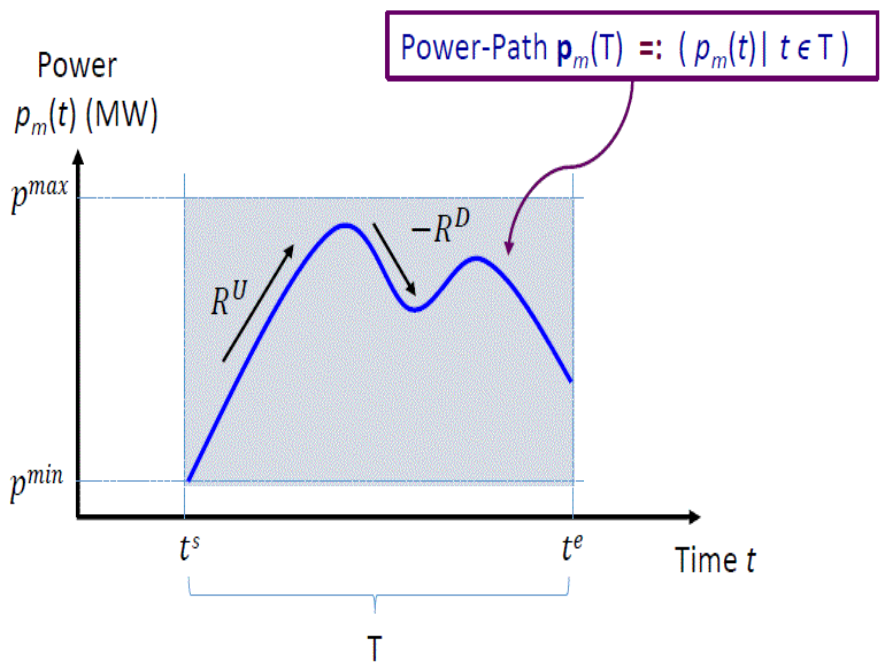


Fig. 6 One of multiple possible power-paths $\mathbf{p}_m(\mathrm{T}) = (\, p_m(t) \mid t \in \mathrm{T}\,)$ consisting of successive injections/withdrawals of power $p_m(t)$ (MW) that a dispatchable power resource m with swing (flexibility) in ramp-rate (MW/min) and power capacity (MW) could be RTO/ISO-dispatched to deliver at its grid location b_m during a future operating period T.

A second innovative aspect of the Linked Swing-Contract Market Design is a fundamental change in the envisioned role of the RTO/ISO.

In current U.S. RTO/ISO-managed wholesale power markets, the primary concern of the RTO/ISO is the daily DAM/RTM scheduling of grid-delivered energy for near-term operating periods, supported by ancillary services (e.g., operating reserve, reactive power support for voltage control, ...) and supplemental procurement processes (e.g., residual unit commitment processes, capacity markets, ...).

In the Linked Swing-Contract Market Design, the envisioned primary concern of the RTO/ISO is the advance clearing of reserve offers from dependable dispatchable power resources for *possible* RTO/ISO dispatch during *future* operating periods T to protect against volumetric grid risk.[43] A *reserve offer for T* is a contractually expressed offer to provide guaranteed availability of power-path production capabilities for possible RTO/ISO dispatch during period T.

These two design innovations require changes in product definitions, contract forms, settlement rules, and RTO/ISO management practices, but *not* in real-time operations. Consequently, as demonstrated in [39, Ch. 16], these design innovations can be introduced gradually into current U.S. RTO/ISO-managed wholesale power markets without disruption of their real-time operations.

The basic features of the Linked Swing-Contract Market Design are as follows. The design consists of a linked collection of RTO/ISO-managed forward reserve markets M(T) for *future* operating periods T. As noted above, *reserve for T* consists of the guaranteed availability of diverse power-path production capabilities in advance of T for *possible* RTO/ISO dispatch *during* T, as protection against volumetric grid risk.

More precisely, as carefully explained and illustrated in [39, Chs. 10-11], the Linked Swing-Contract Market Design posits a collection $\mathscr{M} =: \{\mathscr{M}(\mathrm{T}) \mid \mathrm{T} \in \mathscr{T}\}$ of sets $\mathscr{M}(\mathrm{T}) = \{\mathrm{M}'(\mathrm{T}), \mathrm{M}''(\mathrm{T}), \ldots\}$ of *RTO/ISO-managed forward reserve markets* $\mathrm{M}'(\mathrm{T})$, $\mathrm{M}''(\mathrm{T})$, ... for each operating period T in a designated operating-period set $\mathscr{T}$. The forward reserve markets $\mathrm{M}'(\mathrm{T})$, $\mathrm{M}''(\mathrm{T})$, ... in $\mathscr{M}(\mathrm{T})$ for a given future operating period T are differentiated by the durations of their *look-ahead horizons* $\mathrm{LAH}'(\mathrm{T})$, $\mathrm{LAH}''(\mathrm{T})$, ... , which can range from years to minutes; cf. Fig. 5.

The participants in each market M(T) consist of *Load-Serving Entities (LSEs)*, *Intermittent Power Resources (IPRs)*, and *Dispatchable Power Resources (DPRs)*.

A *reserve bid* submitted to M(T) by an LSE j on behalf of managed customers electrically connected to the grid at a common grid-location b_j is a demand for power-path delivery at b_j during T. This demand can take a dispatchable price-sensitive form and/or a *fixed (non-dispatchable must-service)* form.

The power injections of an IPR n unfirmed by storage that take place at n's grid point-of-connection b_n during T are considered to be a fixed power-path delivery that must be forecasted by the RTO/ISO. This forecast is directly entered into the system constraints for the contract-clearing optimization that the RTO/ISO conducts for M(T): specifically, it is entered into the bus-b_n power balance constraint for operating period T.

As previously noted, and illustrated in Fig. 6, a *reserve offer* submitted to M(T) by a DPR m is a contractually expressed offer to provide guaranteed *availability* of power-path production *capabilities* for *possible* RTO/ISO dispatch at m's grid-location b_m during T.

[43] *Volumetric grid risk* is systemic risk of grid collapse due to net load imbalance. A *systemic risk* is a *system-wide* risk, i.e., a correlated risk arising for system operations as a whole.

Specifically, a reserve offer submitted to M(T) by a DPR m is a two-part pricing[44] swing-contract SC_m expressible in the following standardized form:

$$\mathrm{SC}_m(\mathrm{T}) = \Big(\alpha_m(\mathrm{T}),\ \mathbb{T}_m^{\mathsf{ex}}(\mathrm{T}),\ \mathbb{PP}_m(\mathrm{T}),\ \phi_m(\mathrm{T})\Big) \tag{15}$$

The first component $\alpha_m(\mathrm{T})$ in (15) is the *offer price* that m specifies for $\mathrm{SC}_m(\mathrm{T})$. A non-negative[45] offer price designates the amount that must be paid to m – either directly or in amortized payment-schedule form – if the RTO/ISO clears $\mathrm{SC}_m(\mathrm{T})$ for T. The offer price $\alpha_m(\mathrm{T})$ permits m to ensure recovery *ex ante* (i.e., in advance of T) for all avoidable fixed cost[46] that m must incur to guarantee its offered availability of power-path production capabilities for possible RTO/ISO dispatch during T, should the RTO/ISO choose to clear $\mathrm{SC}_m(\mathrm{T})$ for T.

The second component $\mathbb{T}_m^{\mathsf{ex}}(\mathrm{T})$ in (15) is the set of possible *exercise times* t_m^{ex} that m designates for $\mathrm{SC}_m(\mathrm{T})$ between the close of M(T) and the start-time for the future operating period T. The number and positioning of these exercise times determine whether $\mathrm{SC}_m(\mathrm{T})$ is a firm contract or some type of option contract (e.g., European, American, Bermudan, ...).

The third component $\mathbb{PP}_m(\mathrm{T})$ in (15) is the set of power-paths that m is offering for possible RTO/ISO-dispatched delivery at m's grid location b_m during T. Ideally, $\mathbb{PP}_m(\mathrm{T})$ should be a "digital twin" expressing the full physical power-path production capabilities of m during T. In practice, $\mathbb{PP}_m(\mathrm{T})$ will typically be an approximation for this digital twin that conveys the key physical attributes of m's offered power-paths.

These key physical attributes must include m's grid delivery location b_m. Additional attributes can include *static* features, such as total amount of grid-delivered energy (MWh) and down-time/up-time parameters, as well as *dynamic* features such as: power capacity (MW) *profile* during T; ramp-rate (MW/min) *profile* during T; and power-factor (kW/kVA) *profile* during T. The forms and ranges of these power-path attributes determine the degree of *swing (flexibility)* in m's offered reserve.

The final component $\phi_m(\mathrm{T})$ in (15) is a performance payment method that maps each power-path $\mathbf{p}$ in the power-path set $\mathbb{PP}_m(\mathrm{T})$ into a real-valued dollar payment

[44] As detailed in [39, Ch. 10.2], this two-part pricing enables *separate time-consistent DPR cost settlements* (i) for reduction of volumetric grid-risk provided for T *in advance* of T, and (ii) for any RTO/ISO-dispatched power-path delivery occurring *during* T subject to RTO/ISO verification *subsequent* to T.

[45] Swing-contract offer prices are not restricted in sign. For example, a DPR m with a positive minimum sustainable power-injection level could specify an offer price $\alpha_m < 0$ in an attempt to ensure the RTO/ISO clears his swing contract SC_m. This negative offer price commits the DPR to *pay* the amount $-\alpha_m$ should SC_m be cleared.

[46] See Appendix A.4 for a listing of different types of avoidable fixed cost that real-world DPRs can incur. These types include: capital investment cost; transaction cost; opportunity cost; and unit-commitment cost.

$\phi_m(\mathrm{T})(\mathbf{p})$ (\$) for incurred variable cost.[47] More precisely, $\phi_m(\mathrm{T})(\mathbf{p})$ designates the *ex-post* (i.e., after T) compensation for variable cost (\$) that m would require for delivery of $\mathbf{p}$ during T in accordance with received RTO/ISO dispatch instructions. This variable-cost compensation is subject to post-T verification by the RTO/ISO of m's performance (power-path delivery) during T, which is an incentive for m to follow the RTO/ISO's dispatch instructions as closely as possible.

The performance payment method $\phi_m(\mathrm{T})$ can take a wide variety of forms. Ideally, however, it should be expressible in standardized metrics that permit both DPR m and the RTO/ISO: (i) to agree *ex ante* on the precise nature of DPR m's offered risk-reduction service (dispatchable power-path availability) for T; and (ii) to verify *ex post* the extent to which any contractually-admissible dispatch instructions conveyed by the RTO/ISO to DPR m for period-T performance (power-path delivery) have accurately been followed.

The RTO/ISO clears a subset of the reserve bids and offers submitted to M(T) to maximize the expected *net benefit (benefit minus avoidable cost)* of M(T) participants. This contract-clearing optimization problem is conditioned on current state conditions (including market linkages)[48] and is subject to standard types of system constraints for T such as nodal net-load balancing constraints, nodal/zonal reserve requirements, and line capacity limits. The conceptual and practical advantages of this contract-clearing optimization formulation relative to the SCUC/SCED optimization formulations implemented in current RTO/ISO-managed wholesale power markets are demonstrated by findings from analytical and computational test-cases reported in [39].

Finally, to preserve its status as an independent fiduciary, the RTO/ISO allocates all net reserve procurement costs and transmission service costs incurred for M(T) operations back to M(T) participants based on their relative contributions to these costs. The specific cost-allocation rules used by the RTO/ISO for these purposes are carefully presented and motivated in [39, Sec. 6.7].

6.3 Current U.S. DAMs vs. Swing-Contract DAMs

This section provides high-level comparisons of the basic design features and market-clearing optimization formulations for current U.S. DAMs and SC DAMs, assuming all customer demand takes a fixed-load form to simplify the comparison.

---

[47] As shown in Appendix A.4, these variable costs include payments for inputs (fuel, labor, ...) needed for power-path production, charges for transmission services, and wear-and-tear depreciation of physical equipment.

[48] *Market linkages* refers to the specific linkages established among the collection of markets $\mathscr{M} =: \{\mathscr{M}(\mathrm{T}) \mid \mathrm{T} \in \mathscr{T}\}$ as a result of the RTO/ISO's contract-clearing and dispatch decisions in successive operating periods T. The RTO/ISO keeps tracks of these linkages by carrying forward on its books an adaptively updated record of its cleared reserve bids/offers and its dispatch decisions.

		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. 7 Basic Market Design features: Current U.S. RTO/ISO-managed DAMs vs. SC DAMs.

Figure 7 provides high-level comparisons of the *basic market design features* for SC DAMs and current U.S. DAMs. The *main similarities* are:

– Both DAM designs are RTO/ISO managed;

– Both DAM designs have the same types of market participants;

– Both DAM designs are subject to standard SCED-types of system constraints, such as nodal power balance constraints, nodal/zonal reserve requirements, and transmission-line capacity limits.

The *main differences*, listed below, involve product definition, contract forms, settlement rules, and RTO/ISO managedment practices, *not* real-time operations. Thus, as demonstrated in [39, Ch. 16], these differences could be introduced *gradually* into current U.S. RTO/ISO-managed DAMs.

– SC DAMs are forward markets for *reserve*, i.e., for offered availability of diverse power-path production capabilities for possible next-day RTO/ISO dispatch;

– SC DAM reserve offers 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 submitted reserve offer SC_m is cleared for day D+1 receives an *offer-price payment* α_m (in lump-sum or amortized form) as compensation for the reduction of volumetric grid-risk this cleared reserve offer provides for day D+1 operations.[49]

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

– A DPR m participating in an SC DAM can ensure its revenue sufficiency (i.e., market revenue $\geq$ avoidable cost) by appropriate specification of the offer-price α_m and performance payment method ϕ_m that m includes in its submitted reserve offer SC_m; no resort to OOM make-whole payments is needed.

		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. 8 Market-clearing optimization formulations: Current U.S. RTO/ISO-managed DAM SCUC/SCED optimization vs. SC DAM optimization.

Figure 8 provides high-level comparisons of the *market-clearing optimization formulations* for SC DAMS and current U.S. DAMs. The *main similarities* are:

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

[49] Note the *amount of this risk-reduction* depends on the inherent swing (flexibility) of the power-path set $\mathbb{PP}_m$ that m includes in its submitted reserve offer SC_m. However, the amount of this risk-reduction does *not* depend on which power-path the RTO/ISO actually dispatches from this power-path set during day D+1.

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

The *main differences* are as follows. For an SC DAM contract-clearing optimization conducted on day D for an operating-day D+1:

– the objective function fully incorporates the availability cost (i.e., avoidable fixed cost) and performance cost (i.e., variable cost) of each participating dispatchable power resource m, as conveyed to the RTO/ISO by the offer price α_m and the performance payment method ϕ_m that m includes in its submitted swing-contract reserve offer SC_m;

– the only payment obligation determined by the day-D SC DAM that is settled in advance of operating-day D+1 is coverage of the availability costs incurred by dispatchable power resources with cleared reserve offers;

– no performance payments are made to cleared dispatchable power resources in advance of actual performance (dispatched power-path delivery) on day D+1 that has been verified by the RTO/ISO;

– "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 dispatchable power resource m are conveyed to the RTO/ISO by the power-path production-capability set $\mathbb{PP}_m$ that m includes in its submitted swing-contract reserve offer SC_m.

7 Conclusion: Grids as Flexibility-Support Mechanisms

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

This study concludes that fundamental design changes are indeed needed. Multiple conceptually-problematic aspects of the legacy DAM/RTM two-settlement system design at the core of all seven current U.S. RTO/ISO-managed wholesale power markets are identified, analyzed, and illustrated in Sections 2–4. Briefly summarized, these aspects are as follows:

Key Product-Definition Issues

(1) Conceptually-problematic focus on grid-delivered energy (MWh) as the basic transacted product.

(2) *Grid-delivered* energy strongly fails to satisfy a unit homogeneity requirement essential for the conceptual coherency of Locational Marginal Pricing (LMP).

Key Settlement-Rule Issue

(3) Sequential provisional forward-market determination of LMP settlements in advance of final ex-post LMP settlements for actual real-time dispatched performance results in time-inconsistent settlements, hence in unnecessarily complex and confusing settlement rules.

Key Supply-Offer Formulation Issues

(4) Suppliers are *forced* to express their supply costs as functions of grid-delivered energy amounts (MWh).

(5) Suppliers are *not required* to distinguish carefully between their avoidable and non-avoidable costs.

(6) Suppliers are *unable* to specify their supply offers in a manner that ensures their *revenue sufficiency*, i.e., in a manner that ensures their *market*-attained earnings are sufficient to cover their *market*-incurred avoidable costs.

Section 5 highlights and discusses four indications that the conceptual design issues **(1)**–**(6)** are causing increasing problems for current U.S. RTO/ISO-managed wholesale power market operations. These four troublesome indications are:

- Proliferation of participation models functioning as artificial entry barriers;
- Proliferation of conceptually ill-defined flexibility products;
- Proliferation of out-of-market (OOM) make-whole payments;
- Growing supplier revenue insufficiency concerns.

To illustrate how an alternative conceptually-consistent market design could be implemented for a grid-supported centrally-managed wholesale power market – a design well-suited for the scalable support of increasingly decarbonized grid operations and increasingly diverse participants – Section 6 briefly reviews the key features of the *Linked Swing-Contract Market Design* developed and tested in studies [1, 19, 26, 27, 29, 37, 39, 40].

Finally, throughout this study, emphasis has been placed on grid-supported centrally-managed wholesale power markets as *flexibility-support mechanisms*: that is, as mechanisms enabling just-in-time production and transmission of bulk power to satisfy just-in-time customer power demands and grid reliability requirements. As seen, ensuring the efficiency and reliability of such mechanisms is a complex multi-faceted problem.

A physically-feasible alternative might be to transit to a fully storage-supported world. Flexible producer determination of power production levels and flexible customer determination of power usage levels could be supported entirely by local producer and customer storage devices, linked via a supply chain consisting of retail storage-device stores and/or a network of charge/discharge stations.

However, a key economic concern regarding this fully storage-supported world is the potential for substantial inefficiency (wasted resources). Suppose the *average*

state-of charge (SOC) that producers and customers maintain for their local storage devices over time is uniformly bounded above zero. Then, from a global vantage point, it would appear as if a positive, possibly-large, and possibly growing inventory of (potential) energy were being carried forward through time, forever unused, instead of contributing to the creation of net benefit.

Consequently, at least at present, striving to redesign current grid-supported centrally-managed wholesale power markets to enable them to operate as reliable efficient flexibility-support mechanisms over lower-carbon grids would seem to be the better option. Moreover, this redesign, itself, should 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: Quick-Reference Glossaries and Guides

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	Customer load & inadvertent power loss minus net non-dispatched power injection
Net fixed load	Customer fixed load minus net 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	Unavoidable fixed cost
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 a specific 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 = $(\alpha, \mathbb{T}^{\mathsf{ex}}, \mathbb{PP}, \phi)$
$\mathrm{SC}_m(\mathrm{T})$	SC submitted by a DPR m to a swing-contract market M(T) for T
t^{ex}	Exercise time in an exercise set $\mathbb{T}^{\mathsf{ex}}$
$t_m^{\mathsf{ex}}(\mathrm{T})$	Exercise time in an exercise set $\mathbb{T}_m^{\mathsf{ex}}(\mathrm{T})$
$\mathrm{T} = [t^{\mathsf{s}}, t^{\mathsf{e}})$	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 $j \in \mathbb{LS}(b)$
$\mathbb{L} \subseteq \mathbb{B} \times \mathbb{B}$	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
$\mathbb{LS}$	Index set for the load-serving entities j participating in a swing-contract market
$\mathbb{LS}(b) \subseteq \mathbb{LS}$	Subset of load-serving entities in $\mathbb{LS}$ 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 non-dispatchable generation units n participating in a swing-contract market
$\mathbb{NG}(b) \subseteq \mathbb{NG}$	Subset of 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 m
$\mathbb{RR}_m$	Set of feasible ramp-rates r (MW/min) for m
$\mathbb{T}^{\mathsf{ex}}$	Set of possible exercise times t^{ex} for a swing-contract SC
$\mathbb{T}_m^{\mathsf{ex}}(\mathrm{T})$	Set of possible exercise times $t_m^{\mathsf{ex}}(\mathrm{T})$ for a swing contract $\mathrm{SC}_m(\mathrm{T})$

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. DOI: 10.1016/j.apenergy.2020.115182
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. DOI: https://doi.org/10.26889/9781784671600
5. Birge, J., Hortaçsu, A., Mercadal, I., Pavlin, M. (2017). "Limits to arbitrage in electricity markets: A case study of MISO," CEEPR WP 2017-003, MIT Center for Energy and Environmental Policy Research (CEEPR) Working Paper Series, Jan.
6. 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.
7. Farid, A.M. (2022). "A profit-maximizing security-constrained IV-AC optimal power flow model and global solution," *IEEE Access*, vol. 10, 2842-2859. doi: 10.1109/ACCESS.2021.3138972 `https://ieeexplore.ieee.org/document/9663377`
8. FERC (2003). *Notice of White Paper*, U.S. Fed. Energy Regulatory Comm. (FERC), April 28.
9. FERC (2016). "Fast-Start Pricing Operated 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), Dec. 15. `https://www.ferc.gov/whats-new/comm-meet/2016/121516/E-2.pdf`
10. FERC (2017). "Fast-Start Pricing in Markets Operated by by Regional Transmission Organizations and Independent System Operators: Withdrawal of Notice of Proposed Rulemaking and Termination of Rulemaking Proceeding," *Docket No. RM17-3-000*, U.S. Federal Energy Regulatory Commission (FERC), Dec. 21.
11. FERC (2018a). "Essential Reliability Services and the Evolving Bulk-Power System: Primary Frequency Response," *Final Rule, Order No. 842*, U.S. Federal Energy Regulatory Commission (FERC), February 15th.
12. FERC (2018b). "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), April 25. `https://www.federalregister.gov/documents/2018/04/25/2018-08609/`
13. FERC (2019). "Map of Regional Transmission Organizations and Independent System Operators," U.S. Federal Energy Regulatory Commission (FERC). `https://www.ferc.gov/industries/electric/indus-act/rto.asp`
14. 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), Sept. 17th.
15. 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, September.
16. FERC (2022a). "Shaping the Grid of the Future: RTOs and ISOs," U.S. Federal Energy Regulatory Commission (FERC). Available online: `https://www.ferc.gov/power-sales-and-markets/rtos-and-isos`
17. FERC (2022b). "Modernizing Wholesale Electricity Market Design," *Order Directing Reports: Docket No. AD21-10-000*, U.S. Federal Energy Regulatory Commission, , April 21. `https://www.ferc.gov/media/ad21-10-000-0`
18. Hartman, D. (2017). "Refreshing price formation policy in wholesale power markets," Study No. 106, *R Street Policy*, Aug.

19. Heo, D.Y., Tesfatsion, L.S. (2015). "Facilitating appropriate compensation of electric energy and reserve through standardized contracts with swing," *Journal of Energy Markets* 8(4), 93–121. https://doi.org/10.21314/JEM.2015.135
20. Hsieh, E., Anderson, R. (2017). "Grid flexibility: The quiet revolution," *Electricity J.* 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. Kirschen, D.S., Strbac, G. (2018). *Fundamentals of Power System Economics*, Second Edition, John Wiley & Sons, New York, NY.
24. 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.
25. 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.
26. 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.
27. Li, W. (2023). Market Design Innovations for Decarbonized Grid Operations, Ph.D. Thesis, Dept. of Electrical & Computer Engineering, Iowa State University, Ames, IA.
28. 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.
29. 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. Preprint: https://www2.econ.iastate.edu/tesfatsi/SwingContractsWithDynamicReserves.MaEtAl.pdf
30. MISO (2020). "Energy and Operating Reserve Markets," *Business Practice Manual*, BPM-002-r21, Effective Date October 15th.
31. MISO (2021). "Energy and Operating Reserve Markets," *Business Practice Manual*, BPM-002-xxxx, Midcontinent Independent System Operator (MISO), https://www.misoenergy.org/business-practice-manuals/
32. NAS (2016). *Analytic Research Foundations for the Next-Generation Electric Grid*, National Academies of Science, The National Academies Press, Washington, D.C.
33. 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.
34. Schweppe, F.C., Tabors, R.D., Kirtley Jr., J.L., Outhred, H.R., Pickel, F.H., Cox, A.J. (1980). "Homeostatic utility control," *IEEE Trans. on Pow. Apparatus and Syst.* PAS-99(3), May/June.
35. 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.
36. Sun, J., Tesfatsion, L. (2010). "DC Optimal Power Flow Formulation and Testing Using Quad-ProgJ," Working Paper No. 06014, Dept. of Econ., Iowa State U. Last Revised: March 2010. https://www2.econ.iastate.edu/tesfatsi/DC-OPF.JSLT.pdf
37. 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
38. 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.
39. Tesfatsion, L. (2021). *A New Swing-Contract Design for Wholesale Power Markets*, 20 Chapters, 288pp., John Wiley & Sons, Inc. (IEEE Press Series on Power Engineering), Hoboken, New Jersey, USA.

40. 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
41. 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, 325pp. https://www2.econ.iastate.edu/tesfatsi/ace.htm