Design by Competitive Benchmarking:
Tackling the Smart Grid Challenge with Innovative IS Artifacts

Markus Peters\textsuperscript{1**}, Wolfgang Ketter\textsuperscript{1}, and John Collins\textsuperscript{2}

\textsuperscript{1} Erasmus University Rotterdam, The Netherlands, peters@rsm.nl, wketter@rsm.nl
\textsuperscript{2} University of Minnesota, jcollins@cs.umn.edu

Abstract. Our digital economy relies on a power infrastructure that was not designed to support it. The fault lines are starting to show, and they are exacerbated by initiatives to liberalize electricity markets, to increase production from renewables, and to electrify road transport. The Smart Grid, the vision of a new distributed control paradigm for the electric grid, could come to the rescue but critical parts of this vision remain unclear. We make three contributions towards resolving the Smart Grid challenge: First, we propose that Brokers, an innovative class of IS artifacts, should play a pivotal role in a future Smart Grid. Second, we formalize Competitive Benchmarking, a general-purpose design methodology for IS artifacts in complex Smart Market environments. We instantiate this methodology for the case of electric grids in the Power Trading Agent Competition (Power TAC). And third, we analyze seven Broker designs from Power TAC participants in five different countries to derive a tentative Broker design theory. Our work contributes to the design of a power infrastructure for the twenty-first century.

Keywords: Brokers, Computational Economics, Design Theory, Smart Grid, Smart Markets, Trading Agents

\textsuperscript{**} The first author is a student at Erasmus University Rotterdam.
1 Introduction

Our society runs on electricity. Consider the America of 1940, for example, where 10% of energy was consumed in the form of electricity and contrast this number with the digital America of today, where more than 40% of energy is electrical. (US Department of Energy, 2003; US Energy Information Administration, 2013a,b). And the trend towards further electrification is intact: between one and six million electric vehicles are projected to be driving on American roads by 2020 (MIT, 2011). Each one of these vehicles will require roughly the same electricity per charge that a family household consumes in a day. Of course, such developments are in addition to the continuing digitization of our economy, on which global businesses will spend up to 4.5 trillion dollars annually over the next five years according to Gartner’s 2013 IT spending report.

But electricity’s remarkable success story comes not without downsides. Most American electricity today is generated in large, centralized power plants from fossil (68%) or nuclear (20%) fuels that have been heavily criticized for their environmental impact (US Energy Information Administration, 2013a). An average power plant converts only about one third of its primary energy into usable electricity, while 8% of its output are lost on the long way to the consumer (US Department of Energy, 2003). Generation from renewable or small-scale decentralized sources addresses these problems but their intermittent, decentralized nature makes them difficult to control when deployed at appreciable scale. A series of blackouts, including the 2003 Northeast Blackout that affected 50 million customers, has vividly demonstrated our unconditional reliance on a grid that is becoming vulnerable (Fox-Penner, 2005).

The reasons behind these challenges are numerous and intertwined in complex ways. But one important theme underlying many of them is a mismatch between growing demands for reliability, sustainability, affordability, volume, and novel services on the one hand, and a hierarchical control paradigm that is essentially unchanged from the grid’s early days on the other. Modernizing this control paradigm is an extremely challenging proposition. The Smart Grid of the future will have to (a) efficiently allocate electricity among hundreds of millions of stakeholders with unique preferences and behavioral idiosyncrasies; (b) accommodate renewable and decentralized power
sources, as well as novel storage capabilities like electric vehicle fleets; (c) respect the complicated side-constraints imposed by power flow physics and several layers of regulation; (d) provide fine-grained, decentralized control possibly down to the appliance level; and (e) uphold real-time control under uncertainty due to the infinitely perishable nature of electricity, all the while ensuring a smooth transition from the operational grid of today. By bringing together insights from the behavioral and engineering sciences, Information Systems (IS) scholars have substantial contributions to make to this grand interdisciplinary challenge.

Our own contributions in this article are threefold. First, we describe the Smart Grid challenge from an IS viewpoint, and argue for the importance of a new class of IS artifacts that we call **Brokers**. Brokers can play a pivotal role in modern control paradigms for the electric grid specifically, and other Smart Market environments more generally. Second, we formalize **Competitive Benchmarking**, a general-purpose design methodology for IS artifacts in complex Smart Market environments. We apply the Competitive Benchmarking methodology in the Power Trading Agent Competition (Power TAC, Ketter et al. 2013a) which challenges participants to create innovative Broker designs. And third, we present a tentative **Broker design theory** based on insights gleaned from Power TAC’s Competitive Benchmarking process to-date.

Brokers are the core of a series of radically new IS-driven, customer-centric business models. They afford customer participation where it is currently unfeasible, incentivize behavioral change, and leverage distributed social intelligence for the achievement of societal goals. Our work on Broker design contributes to a solution for the “grand challenge” of providing affordable, reliable, and sustainable energy for the twenty-first century (Amin and Wollenberg, 2005).

## 2 The Smart Grid as an Information Systems Challenge

Today’s electric grid has evolved out of a fragmented system of regional electric utilities. In North America, these vertically integrated monopolies dominated the generation, transport, and distribution of electricity until the late 1990s when the Federal Energy Regulatory Commission (FERC) mandated that every generator was to be given equal access to the grid.
A model of the physical and control structures that resulted from this initiative is depicted in Figure 1. The model is separated into a **transmission system** where large-scale generators like coal and nuclear power plants feed high-voltage electricity into a transmission network, and a **distribution system** responsible for regional electricity provisioning to commercial and residential end-customers. Some large industrial loads capable of operating at high voltage levels may be connected directly to the transmission system, but the bulk of loads in terms of volume and number are served through distribution systems. In the following, we contend that IS research should focus on distribution systems where it will have greater impact than at the transmission level.

The key differences between the two types of system are summarized in Table 1. **Transmission systems** are far evolved in terms of sensory capabilities and ICT deployment. The Independent System Operators (ISO) responsible for operating these systems and other participants have significant incentives for investing in sensing, optimization, and decision-support capabilities due to the tremendous costs and benefits of their capital-intensive businesses. A key difference between
<table>
<thead>
<tr>
<th>Attribute</th>
<th>Transmission System</th>
<th>Distribution System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of nodes and stakeholders</td>
<td><strong>Low</strong> - Approximately</td>
<td><strong>High</strong> - Approximately</td>
</tr>
<tr>
<td></td>
<td>– $10^2$ large-scale generators</td>
<td>– $5 \times 10^4$ distribution feeders</td>
</tr>
<tr>
<td></td>
<td>– $10^3$ transmission lines</td>
<td>– $10^7$ customer meters</td>
</tr>
<tr>
<td></td>
<td>– $10^4$ substations</td>
<td>– $5 \times 10^8$ appliances</td>
</tr>
<tr>
<td></td>
<td>per system</td>
<td>per system</td>
</tr>
<tr>
<td>Stakeholder sophistication</td>
<td><strong>High</strong> - Commercial generators use sophisticated forecasting and optimization</td>
<td><strong>Low-to-Medium</strong> - Distribution Utilities (DU) use relatively simple forecasting</td>
</tr>
<tr>
<td></td>
<td>routines to compute optimal bids in wholesale markets;</td>
<td>schemes; Residential customers make ad-hoc consumption decisions and potentially</td>
</tr>
<tr>
<td></td>
<td>Independent System Operators (ISO) use advanced power flow analysis tools for</td>
<td>sporadic, manual tariff decisions; Commercial customers</td>
</tr>
<tr>
<td></td>
<td>operations and contingency planning</td>
<td>may exhibit some sophistication in tariff negotiations and use of controllable loads</td>
</tr>
<tr>
<td>Sensing ability</td>
<td><strong>Medium-to-High</strong> - Mature commercial sensing devices are available and widely</td>
<td><strong>Low-to-Medium</strong> - Commercial customer on-site sensing devices available but questions</td>
</tr>
<tr>
<td></td>
<td>deployed; increasing deployment of Phasor Measurement Units (PMU) for wide-area</td>
<td>remain w.r.t. standards, optimal level of functionality, etc.; DUs are gradually</td>
</tr>
<tr>
<td></td>
<td>monitoring</td>
<td>installing feeder sensors</td>
</tr>
<tr>
<td>ICT deployment</td>
<td><strong>High</strong> - ISOs use sophisticated Energy Management Systems (EMS), Supervisory</td>
<td><strong>Low-to-Medium</strong> - Customer-site deployment efforts are still tentative; uncertainty</td>
</tr>
<tr>
<td></td>
<td>Control and Data Acquisition (SCADA) systems, and Phasor Data Concentrators (PDCs)</td>
<td>w.r.t. standards, regulatory framework, required functionality, and business cases;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DUs use Distribution Management Systems (DMS) but proprietary technology makes them</td>
</tr>
<tr>
<td></td>
<td></td>
<td>difficult to extend to new applications</td>
</tr>
<tr>
<td>Control paradigm</td>
<td><strong>Market-based / Direct</strong> - Long-term matching of supply and demand, as well as</td>
<td><strong>Direct</strong> - Predominantly passive management</td>
</tr>
<tr>
<td></td>
<td>ancillary services such as regulation, load-following, and various types of reserves</td>
<td>to match supply to demand; some use of direct control for load shedding</td>
</tr>
<tr>
<td></td>
<td>Real-time operation starts from previous market transaction, applies direct control</td>
<td></td>
</tr>
<tr>
<td></td>
<td>to handle real-time deviations and contingencies</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Summary of current transmission and distribution system characteristics. Numeric examples for a typical U.S. system are taken from (Widergren et al., 2004)

electricity provisioning and other businesses is the presence of two separate layers of control that participants engage in: **Economic Control** at the transmission level is provided by wholesale markets where generators sell energy to aggregated distribution-level loads or large industrial users of electricity. In an ideal world, forward markets would provide allocation ahead of time and spot markets would resolve residual imbalances between supply and demand. But because electricity is currently not stored at scale, and because generators are subject to ramp-up times and other physical constraints, most generators can only sell their capacity forward, i.e., there exists no conventional, competitive spot market for electricity. The upshot of all wholesale market transactions
is a tentative schedule of production and consumption. Based on this preliminary schedule, the ISO operates the grid using **Direct Control** in real-time. This includes control actions in the seconds-to-milliseconds range (for which markets are too slow) and all system-level considerations (for which markets do not account). An important part of this process is invoking reserve generators to balance supply and demand whenever the tentative schedule is flawed. This balancing service is coordinated through an ISO-managed **balancing market** where the originators of imbalances are charged rates that are typically higher than those for wholesale purchases ahead of time. The managed balancing market replaces the competitive spot market that exists in other commodities.

**Distribution systems**, in contrast, are far less evolved. The difference can be attributed to their smaller scope in terms of energy supplied over wider areas, and to their larger number of unsophisticated participants. Most distribution networks use a traditional, hierarchical control paradigm where customers subscribe to fixed or simple time-of-use tariffs, make consumption decisions independently of the availability of electricity, and receive monthly bills based on sporadic meter readings. The Distribution Utility (DU) that operates the system forecasts the aggregate consumption of all participants and procures offsetting generation commitments in the wholesale market. The wholesale market, in turn, invokes the right number of generators.

This simple distribution system control paradigm has worked well historically, but has lately been criticized for at least four reasons: First, participants have currently no incentives for shifting their non-time-critical loads (e.g., washing machines) to times when electricity is available. This is especially problematic in light of increasing deployment of renewables that cannot be dispatched to match given demand. Second, grid capacity planning is based on peak demand. Between 10% and 18% of American power systems capacity is utilized less than 1% of the time and this spare capacity is both environmentally harmful (e.g., excess transmission corridors) and expensive for rate-payers (MIT, 2011). Third, more than 90% of outage minutes are due to distribution-level failures. These failures affect digital economies that rely on high-quality power.3 And fourth,

---

3 Some studies have estimated that the “cost of power disturbances to the U.S. economy ranges between $119 and $188 billion per year” (Electric Power Research Institute, 2011), while others contend that “power outages and power quality disturbances cost the economy from $25 to $180 billion annually.” (US Department of Energy, 2003)
unresponsive customer demand combined with competitive wholesale markets leads to market power on the side of highly reactive marginal generators, a key reason for market failures like the California energy crisis (Borenstein, 2002).

Several practical and conceptual issues hamper the modernization of distribution systems. Visibility in distribution systems is mostly limited to sporadic meter readings that are too coarse for real-time optimization and control. Much less than real-time visibility do distribution systems provide real-time controllability of loads. Selected commercial customers may allow the DU to control certain loads remotely (curtailment), but this level of control remains elusive for the bulk of electricity customers. And finally, even if more granular data was available would a centralized control paradigm not scale to tens of millions of customers or hundreds of millions of appliances.

These issues will likely to become critical when combined with several impending technological and regulatory changes as illustrated in Figure 2. Some of these changes exacerbate the need for modernization, whereas others enable IS innovations that can bring this modernization about.
New **Sensing and Control Capabilities** like digitally connected meters, smart appliances, and remotely controllable loads are among the key enablers. **Renewables and Decentralized Generation** promise to significantly reduce the environmental impact of electricity production, but they pose a formidable challenge when deployed at appreciable scale. **New Load Types** like electric vehicles and electric space heatings will increase overall load levels and change load profiles significantly. **Electricity Storage** can increase the utilization of generation, transmission, and distribution assets when properly controlled. And **Retail Competition** is the institutional foundation for a scalable, distributed approach to managing distribution systems.

Retail competition creates opportunities for intermediaries that offer digitally connected customers personalized, sustainable, and affordable electricity services. While intermediation can to some extent be performed by traditional retailers, we propose that the data-intensity and complexity of future retail electricity markets will breed a sophisticated class of IS artifacts that we call **Brokers**. Brokers will fulfill the intermediary’s role first in support of a human decision-maker but increasingly also in an autonomous fashion (Davenport and Harris, 2005). We consider Brokers a key innovation on the path to future sustainable power systems. Our proposition goes well with the observation that sustainability “can be much more than a cost, a constraint, or a charitable deed – it can be a source of opportunity, innovation, and competitive advantage.” (Porter and Kramer, 2006) Brokers provide customers with information and incentives to adjust their consumption based on availability; they hold the potential to improve allocative efficiency, asset utilization, and economic welfare; and they are the core of a new competitive, distributed control paradigm. A design theory (Walls et al., 1992) for Brokers would address many of the challenges electric grids face today.

What is unclear is how a Broker design theory can be established. It seems daunting to design an artifact that concurrently trades in multiple interrelated markets with strategizing competitors; that has to construct and manage portfolios of long- and short-term obligations under uncertainty; that faces suboptimal decisions and idiosyncratic behavioral influences on the customer side; and that is exposed to uncertainty from weather and multiple layers of regulation. We should note that these challenges are not limited to Smart Grids: Strategic interactions arise in a broad class of
modern Smart Market environments, such as multiechelon supply chains (Collins et al., 2010), or any strategic marketing effort for that matter (Bucklin et al., 1998). We content that a design theory for Brokers will generalize to these other settings as well. What makes the Broker design problem for the Smart Grid particularly interesting is its tremendous societal relevance, paired with the high degree of complexity stemming from electricity’s peculiarities.

The most effective way we know for addressing such high-complexity design problems is a combination of empirical science and open competition. By building new artifacts, empirically evaluating their performance relative to competing designs, learning and improving, and occasional breakthrough innovations, a virtuous cycle is created where designers are constantly challenged to improve upon their existing designs. The competitive design approach combines two of the most important modes of learning: imitation and experimentation. On the downside, it does not automatically account for the risks of failure (regulatory bodies like the Federal Aviation Administration partly manage these risks in high-stakes domains) and it only applies where effective benchmarking can take place.

Both factors are significant in the electric grid. First, the livelihood of many depends on a properly functioning electricity supply. It is therefore impossible to evaluate Broker quality on actual distribution grids. Pilot studies are being performed (Faruqui et al., 2009) but their realism is bounded by a homogenous, small-scale setup where one consortium controls the entire pilot grid. And second, outside of pilot studies, the Smart Grid exists primarily as a vision. Even if customers were willing to accept the risks that come with improperly designed Brokers, many of the infrastructural prerequisites simply do not exist at this moment.

3 Design by Competitive Benchmarking

We propose to circumvent these limitations through Competitive Benchmarking (CB), a general-purpose design methodology for IS artifacts in complex Smart Market environments. CB unleashes competitive innovation on high-complexity design problems in an environment that promotes em-
Empirical science, accelerates feedback and learning, and that allows high-risk experimentation. It comprises two core elements that support and reinforce each other, see Figure 3.

**Fig. 3:** The Competitive Benchmarking design methodology with its two core elements: the **Competitive Benchmarking Process** is a semi-structured, iterative design process; the **Competitive Benchmarking Platform** is a laboratory environment that supports both design and benchmarking activities.

**The Competitive Benchmarking process** is a semi-structured, iterative process in which multiple independent research groups design novel IS artifacts for a given real-world problem using their unique expertise. They benchmark the performance of their artifacts, jointly analyze and learn from the outcomes, and improve their designs for the next iteration.

**The Competitive Benchmarking platform** is a laboratory environment that supports these design and benchmarking activities. It provides an evolving common language, a common set of abstractions that capture relevant facets of the problem domain, extensive scientific tools, and it reduces the need for costly duplication of infrastructure development. CB is primarily targeted at design problems that are inaccessible to analytic modeling, and CB platforms are therefore often Agent-based simulations (Tefsatsion, 2006) or Virtual Worlds (Chaturvedi et al., 2011).
While neither benchmarking nor scientific experimentation platforms are new, we know of no other work that combines both into a coherent design methodology for IS artifacts. We present the principles of CB in Section 3.1, discuss related concepts in Section 3.2, and instantiate the methodology for electric grids in Section 3.3.

### 3.1 Principles of Competitive Benchmarking

Any effective design methodology is a structured approach to exploring and learning about problem and solution spaces. Good designers create new designs, evaluate their usefulness, and learn from experiences to iteratively improve their designs. This structured form of organizational learning requires skills in

- systematic problem solving, experimentation with new approaches, learning from ... own experience and past history, learning from the experiences and best practices of others, and transferring knowledge quickly and efficiently. [Its best practitioners rely] on the scientific method, rather than guesswork, for diagnosing problems [and insist] on data, rather than assumptions, as background for decision making. (Garvin, 1993)

Competitive Benchmarking embraces these insights and distills them into a CB Process. This process does not aim to replace the existing process of scientific knowledge discovery. Rather, it removes several common obstacles from the process, and it adds a structured approach to benchmarking which we find insufficiently represented in current IS design practices.

Heterogeneity is crucial in designing artifacts for high-complexity environments. It is a source of novelty that leads to a broader set of solutions than any single research group could hope to find (Collins et al., 2009). However, some aspects of the design process are best viewed as a natural monopoly: comprehensive, high-quality CB platforms should be developed in a community effort. Beside the obvious benefit of not duplicating development hours, the approach also has several more subtle benefits: CB platforms encode a set of abstractions and a common vocabulary shared among participants; they are thoroughly validated by dozens of users, an important antecedent of useful artifacts and trustable policy guidance; they provide interfaces and execution contexts for
content from citizen developers (Chaturvedi et al., 2011); and they channel research questions and results between academia, industry, and policy.

3.2 Related Concepts

**Benchmarking** has long been used to improve products and organizational performance; in IS, it became popular through the Netflix competition (Bell and Koren, 2007). Imitative learning is among the method’s key benefits (Vorhies and Morgan, 2005), but it does not end with merely copying existing ideas. Researchers have found that successful benchmarking fosters creativity instead of reactivity, improves learning and other cognitive capabilities, and facilitates innovation based on deep introspection (Garvin, 1993; Shetty, 1993; Drew, 1997). But benchmarking does not by itself facilitate high-risk experimentation or provide accelerated feedback.

The latter two are hallmarks of **Trading Agent Competitions** (TAC, Greenwald and Stone, 2001; Wellman et al., 2007; Ketter and Symeonidis, 2012). A key distinction between TAC and our definition of CB is our emphasis on the **usefulness** of designed artifacts for solving real-world problems. This evolution is rooted in the design science tradition of IS (Walls et al., 1992; Hevner et al., 2004) and it shows in two ways. (1) **Real-World Alignment:** To satisfy the usefulness criteria of design science, an artifact must address an important and relevant business problem, and its utility, quality, and efficacy must be demonstrated (Hevner et al., 2004). Previous Trading Agent Competitions have been inspired by business problems (Collins et al., 2010) but their focus has been on stylized decision situations. In contrast, CB platforms generally model an interesting business problem to a level of detail sufficient for a demonstration of artifacts’ usefulness. (2) **Real-World Guidance:** An important secondary objective of CB is to inform policy makers of robust and efficient regulations for complex Smart Market environments. Here as above, the proper alignment of the simulation and the realism of its abstractions are important prerequisites for usefulness.
3.3 Competitive Benchmarking for Smart Grids

We instantiate the Competitive Benchmarking methodology for the case of electric grids in the Power Trading Agent Competition (Power TAC), a CB platform coupled with a CB process consisting of annual competitions (Ketter et al., 2012, 2013a,b). It models a distribution system under retail competition in a medium-sized city, in which consumers and small-scale producers may choose among a set of alternative electricity retailers represented by competing Brokers. These Brokers are autonomous agents, built by individual research groups. The remainder of the scenario is modeled by the CB platform. A typical simulation runs approximately 60 days of simulated time, although longer simulations are possible.

Fig. 4: Main elements of the Power TAC scenario; Brokers are self-interested, autonomous software agents, built by individual research groups; the remainder of the scenario is modeled by the CB platform
Customers and Retail Market  Brokers interact through a retail tariff market with customer models that represent households and businesses. Some customers are equipped with solar panels and windmills, producing as well as consuming power. All customers are assumed to be equipped with digital meters from which consumption and production is reported every hour. Many customer models also include controllable capacities such as remotely controllable heat pumps or water heaters. Customer models exhibit sensitivity to price changes, weather conditions, and calendar factors such as day of week and hour of day, and they have a range of preferences over tariff terms. For example, some are willing to subscribe to variable-rate tariffs if they have the opportunity to save by adjusting their power usage, while others are willing to pay higher prices for the simplicity of fixed-rate or very simple time-of-use tariffs. Many customer models are built by citizen developers based on current research into tariff choice behavior, e.g., (Gottwald et al., 2011; Reddy and Veloso, 2012).

Tariff contracts may include usage-based and per-day charges, fixed and varying prices for both consumption and production of energy, rates that apply only above a specified usage threshold, signup bonuses, and early-withdrawal penalties. Separate contracts may be offered for charging electric vehicles, which could limit charging during high-demand periods, or even offer to pay customers for feeding energy back into the grid at certain times. Variable prices may follow a fixed schedule, or they may be fully dynamic with specified advance notice of price changes.

Wholesale Market and Generating Companies  Brokers may buy and sell energy from retail customers, and they may buy and sell energy for future delivery in the wholesale market. On the supply side, Power TAC includes models of utility-scale generators who sell their output through the wholesale market. These generators represent different price points and lead-time requirements, e.g, fossil and nuclear power plants, gas turbines, and wind parks.

The Distribution Utility  models the regulated monopoly that owns and operates the physical distribution facilities (lines, transformers, etc.) and is responsible for real-time balancing of supply and demand in the distribution system. It does this primarily by operating in the balancing market, and by exercising capacity controls provided by Brokers. The associated costs are allocated
to imbalanced Brokers. In the real world, balancing responsibility is typically handled at the transmission level. The simulation implements a generalization of proposals to move some balancing responsibility to the distribution level (Strbac, 2008).

**Brokers** develop customer portfolios by offering tariffs to anonymous residential and business customers, and by negotiating individual contracts with larger customers. Because controllable capacities can reduce costs significantly, Brokers can offer special tariffs for them, and then make offers to the DU for the right to exercise them to reduce imbalances. Given a portfolio of customers, Brokers compete with each other in the wholesale market to minimize the cost of power they deliver to their consuming customers, and to maximize the value of power delivered to them by their producing customers.

Official Power TAC competitions are usually held at major Artificial Intelligence conferences, and many Brokers have by now seen several iterations of improvement. The varied design approaches taken by Power TAC participants, together with a shared base of models, tools, and data formats, have accumulated into a rich archive of design ideas and artifact performance data that we analyzed. The result of our analysis is a tentative design theory for Brokers which we discuss next.

## 4 Towards a Broker Design Theory

We hosted Power TAC pilot competitions at international conferences in Barcelona (2011), Valencia (2012), and Nuremberg (2012). Towards a Broker design theory we examined data from 51 simulations in various Broker constellations in the most recent Nuremberg pilot. The purpose of our theory is to describe properties of high-performance Broker designs, so that future Broker developers can use and improve on them. Table 2 summarizes our findings.

To determine each Broker’s performance, we first calculated profit shares, the percentage of each Broker’s profit in all profits made throughout a simulation. Figure 5 shows pronounced differences between high- and low-performing Brokers in terms of profit share magnitude, certainty,

---

### Market Design Principles

<table>
<thead>
<tr>
<th>Market</th>
<th>Design Principles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retail</td>
<td><strong>R1 – Decide on volume or margin</strong> High-volume and high-margin retail strategies can both be successful if properly synchronized with a wholesale market strategy; high-volume strategies appear to be easier to get right, likely because forecasting gets easier</td>
</tr>
<tr>
<td>Wholesale</td>
<td><strong>W1 – Use limit orders</strong> to control wholesale market exposure; consider the balancing market as strategic alternative only in cases of extreme wholesale market pricing</td>
</tr>
<tr>
<td>Balancing</td>
<td><strong>W2 – Use different lead times</strong> to lock in favorable prices for basic loads and adjust positions incrementally as new information arrives</td>
</tr>
<tr>
<td>Balancing</td>
<td><strong>W3 – Learn to forecast before optimizing bids</strong> Forecasting errors and resulting balancing charges dominate potential savings from better wholesale prices. While the best Brokers trade at below-average wholesale prices, it is more important to be in balance than to attain the absolutely lowest wholesale prices.</td>
</tr>
<tr>
<td>Balancing</td>
<td><strong>B1 – Exploit balancing last</strong> Optimize the Broker’s retail and wholesale functions before taking strategic positions in the balancing market</td>
</tr>
<tr>
<td>Balancing</td>
<td><strong>B2 – Avoid systematic balancing errors</strong> The best Brokers make small balancing errors and are not systematically biased towards either under- or oversupply</td>
</tr>
</tbody>
</table>

Table 2: Summary of Broker design principles

and resilience to competition. High-performing Brokers attain profit shares that are large, consistent, and resilient to competition. AstonTAC, for example, does well on all three counts with large average profits that are stable over different games and relatively unaffected by the additional competition in five-broker simulations. Mertacor and LARGEpower, on the other hand, perform well in some cases, but are not as consistent. To understand the reasons for these differences we analyzed Brokers’ actions in Power TAC’s three principal markets: retail, wholesale, and balancing.
Brokers’ strategic positions in the retail market for electricity consumption\(^5\) are shown in Figure 6. Mertacor, Aston TAC, and UTest are high volume players, offering highly competitive rates in exchange for significant retail market shares. Mertacor, in particular, consistently captures the majority of the market at extremely low prices which cut into its own margins (see Figure 5). AstonTAC and SotonPower are interesting in that their success hinges on fundamentally different causes: Aston TAC aims to attract large volumes in the retail market, whereas SotonPower’s suc-

\(^5\) Only AstonTAC and UTest also purchased power in the retail market during the games that we analyzed; volumes were small in relation to wholesale market activities.
cess is based on a more moderated combination of larger markups and other factors we discuss below.

This suggests retail market principle (R1) **Decide on a Volume or Margin strategy**. Both can be successful if synchronized with proper wholesale market trading. Volume strategies appear to be easier to get right, possibly due to the benefits of averaging effects in demand forecasts.

![Diagram](image)

**Fig. 6:** Retail market position of participating Brokers. Lighter coloring indicates higher market share uncertainty.

Brokers’ strategic **wholesale market** positions are illustrated in Figure 7. SotonPower’s low-volume strategy, likely combined with high forecasting accuracy, allows it to procure electricity cheaply. The relatively weak LARGEpower, Mertacor, and UTest are paying prices up to 1.6 times above market average. Note, that this is *not* solely a consequence of large volumes and the corresponding need to invoke marginal, high-cost generators. AstonTAC, for example, acquires more electricity than LARGEpower at lower prices on average, suggesting better timing of wholesale purchases or smarter use of bid prices. Mertacor pays the most for wholesale power; this together with its low-cost tariffs explains its relatively weak performance.
Different wholesale market performances can be explained by looking at Brokers’ bidding behavior. Wholesale bidding in Power TAC is highly complex; in addition to deciding the optimal bid price and volume, a Broker also needs to factor the optimal order leadtime into its decision. The impact of leadtime on prices for two typical games is shown in Figure 8. Both panels show the general downward-sloping price trend for increasing leadtimes that economic theory suggests, but there are irregularities in these graphs that underline the importance of reasoning about leadtimes and order limits. First, the graphs explain the high prices that Mertacor and UTest are paying in the wholesale market. In the left panel, Mertacor covered all its electricity demand in the two-hour-ahead auction at prices that are close to double the prices of longer leadtimes. In the process, it drives up prices for AstonTAC and LARGEpower, while SotonPower controls its exposure to this effect through limit orders. An analogous observation holds true for UTest in the right panel. LARGEpower (left panel) placed all its orders as market orders. This leads to relatively good balancing performance, as only forecasting error and no uncertainty in order filling affect its results, but it is also subject to unpredictable fluctuations in wholesale prices.
Taken together, these observations suggest three wholesale market principles: (W1) Use limit orders to control wholesale market exposure; (W2) Use different lead times to lock in favorable prices for basic loads long in advance and adjust positions as more information arrives; and (W3) Learn to forecast before optimizing bids. The latter principle is based on the observation that AstonTAC performs well in spite of not getting the overall lowest wholesale prices. Large imbalances, on the other hand, always led to inferior overall performance, as we explain next.

![Leadtime effects in the wholesale market](image)

Fig. 8: Leadtime effects in the wholesale market; left panel: AstonTAC vs. default broker vs. LARGEpower vs. Mertacor vs. SotonPower; right panel: default broker vs. MinerTA vs. UTest

We finally looked at the balancing market. This market is implicit in that residual imbalance between consumption and production is automatically offset by the DU in real-time. The penalties (or rewards) that a Broker incurs in the process depend on the overall imbalance on the grid at that instant, as well as on their own imbalances. As a result, it can be worthwhile for a Broker to risk an imbalance if it has the opposite sign from the overall imbalance. This relationship makes the regulating market an interesting strategic element for Brokers. Figure 9 reveals no simple connection between the level of imbalance and the corresponding reward or penalty to the Broker. Some general trends, however, can be observed: undersupply (lower half of the figure) generally
leads to higher penalties than the rewards afforded by oversupply (upper half of the graph). This is reflective of the asymmetric balancing mechanism employed in Power TAC (de Weerdt et al., 2011), and it illustrates the difficulties in explicitly using the balancing market as part of a strategy. AstonTAC, Mertacor, and to a lesser extent LARGEpower, are mostly well-balanced. They only sometimes incur sizable penalties as a result of small errors on large volumes. MinerTA and Soton-Power generally make larger mistakes on smaller volumes, and these mistakes are more favorably distributed between over- and undersupply, leading to smaller overall balancing penalties or even to rewards in the balancing market. UTest makes large, systematic errors on large volumes, leading to hefty penalties.

![Balancing market positions](image.png)

Fig. 9: Balancing market positions. Each point represents one broker/game combination; larger symbols indicate higher overall volumes. Notice the log-log scaling.

Together, these findings suggest two design principles: (B1) **Exploit balancing last**, only after committing to a retail strategy and optimizing wholesale trading with respect to it. And (B2) **Avoid systematic errors** in balancing. While some form of predictive error is unavoidable, the best Brokers showed no significant biases in their predictions of supply and demand.
5 Discussion and Conclusions

A power infrastructure for the twenty-first century will require a new distributed control paradigm, but critical parts of this “Smart Grid” vision remain unclear. We argued that IS scholars have much to contribute in this situation, and proposed that the research efforts of the community should focus on distribution networks where they will likely have the highest impact. Brokers, an innovative class of IS artifacts, are an important step towards a distributed control paradigm. Like other IS artifacts in complex economic environments, Brokers are exceedingly difficult to design. Part of this difficulty lies in the prohibitive costs of experimenting with Brokers in real-life power systems.

We proposed CB as a general purpose design-methodology that alleviates this difficulty. We instantiated the methodology in Power TAC, a rich CB platform coupled with a series of annual competitions that challenge participants to design Brokers. We analyzed data from a pilot competition hosted on the Power TAC platform and derived a tentative design theory for Broker artifacts. The preliminary analysis we presented is evidence of significant performance differences between different approaches to electricity trading. Once participating Brokers are fully developed, tools like empirical game theory (Jordan et al., 2007) can be leveraged to generate compelling, actionable insights into novel technologies and public policies for future sustainable energy systems. This work contributes to a solution for the “grand challenge” of providing affordable, reliable, and sustainable energy for the twenty-first century.

Power TAC is an open-source project, designed and documented to be accessible to advanced students. Access to the software and documentation, along with a Broker repository, will be maintained through the powertac.org website. We look forward to many years of vigorous competition and high-impact research results.

Acknowledgements – Thanks to the Power TAC developers at Minnesota, Aristotle in Thessaloniki, Zagreb, CMU, and Erasmus in Rotterdam, and to all the teams who saw the promise of Power TAC and have been willing to invest in Broker development for a rapidly evolving platform.
Bibliography


Peter Fox-Penner. Rethinking the grid: avoiding more blackouts and modernizing the power grid will be harder than you think. *The Electricity Journal*, 18(2):28–42, 2005.


