Local Flexibility Market

Peter Cramton[1]

22 September 2019

Abstract

A local flexibility market is presented that addresses intrazonal congestion in a zonal electricity market. The market lets system operators access flexibility at multiple voltage levels to satisfy transmission constraints. Market participants offer local flexibility intraday in a continuous trading process. This supply is matched with demand from system operators. The market-based redispatch is transparent and technology neutral. Inc-dec gaming is mitigated with features to detect and sanction the behavior. Cost-based redispatch from conventional generation serves as a backstop if additional flexibility is required close to real time. This further mitigates inc-dec gaming by disciplining behavior in the spot market. An advantage of the approach in Europe is that it represents a modest change from the current market and therefore can be implemented sooner and with less risk.

Keywords: local flexibility, congestion pricing, market design, electricity markets, electricity regulation

[1] University of Cologne and University of Maryland, pcramton@gmail.com. Since 1983, Peter Cramton has conducted research on market design; he has applied that research to design auction-based markets of radio spectrum, electricity, financial securities, and other products. For funding, I thank EPEX Spot and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2126/1–390838866.

Introduction

The zonal market design of European electricity markets faces a serious challenge. Consider Germany. Germany is a single zone and thus has a single price for energy at any time. Rapid entry of renewable energy resources, both wind and solar, has strained the transmission grid. The best renewable resources are often far from load and are intermittent. It is now common that transmission constraints bind within Germany. This intrazonal congestion is alleviated with redispatch. In 2017, the cost of redispatch, including the curtailment of renewables, was 1.4 billion Euros and it totaled 473 million Euros in the first quarter of 2019 (BNA 2019).

In the zonal approach, within-zone transmission constraints are initially ignored for purposes of pricing in the spot market, but of course they cannot be ignored in physical delivery. To satisfy constraints, the system operator does a redispatch. Some out-of-merit generation on the import side of the constraint (downstream) is asked to produce more, displacing some in-merit generation on the export side of the constraint (upstream). This approach works well if the transmission grid is so robust that constraints rarely bind. Otherwise, the approach is problematic.

The current approach in Germany is cost-based redispatch. Resources asked to decrease are paid their lost profits—zonal price minus regulated cost. Resources asked to increase are paid their regulated costs for the additional energy. Resources are remunerated in a way that makes them indifferent to the redispatch measure. This is problematic for two reasons. First, the regulator does not know the resource's true cost, especially for a limited-energy resource such as storage. The regulated cost is at best an approximation of true cost. In Germany, the regulated cost is based on a complex set of rules that are negotiated with industry. Second, the resource providing flexibility is not rewarded for providing the service, since the payment just covers the cost. Participants have no incentive to invest in resources that can provide flexibility. Flexibility is offered not as an opportunity to provide a valuable service, but as a regulatory obligation. This greatly limits the offering and development of much needed flexibility.

Recognizing these limitations of cost-based redispatch, the European Commission has proposed market-based redispatch. The key difference is that the renumeration to those providing flexibility is based on market prices, rather than regulated cost. This too can be problematic. Participants anticipating congestion have incentives to distort bids, even absent any market power. Some upstream resources, anticipating being asked to decrease, can improve profits by underbidding cost on the spot market. Some downstream resources, anticipating being asked to increase, can improve profits by overbidding cost on the spot market, anticipating a higher payment in redispatch. This inc-dec gaming results in inflated profits for resources. Worse it aggravates congestion. The redispatch opportunity introduces a ceiling on the upstream offers and a floor on the downstream offers. Thus, a larger quantity must be redispatched to satisfy the transmission constraint. The extra payment to upstream resources is especially problematic as it creates a perverse incentive to build additional capacity in the wrong place, further aggravating the constraint.

The bid distortions can be thought of as the bidders arbitraging between the two opportunities to sell—the spot market or the subsequent redispatch market. Normally, arbitrage improves efficiency as it causes prices between markets to converge to the efficient level, such as the arbitrage between forward and spot markets. In this case, however, the arbitrage is harmful as it artificially exacerbates the binding transmission constraint. The reason is that the zonal pricing, ignoring intrazonal transmission constraints, is fundamentally inconsistent with the pricing when constraints are respected.

To address these incentive problems, many markets, especially in the United States, have switched to nodal markets. With nodal markets, energy prices properly reflect transmission constraints. The price equals the marginal value of energy at each time and location. These locational marginal prices send the right price signal to support the real-time dispatch respecting transmission constraints. And in the long run, the locational prices also send the right signal for siting of new resources and transmission planning. In most times, transmission constraints do not bind and there is a single energy price, but as constraints bind, price differences emerge that support the efficient dispatch satisfying network constraints.

My view is that the benefits of a nodal market are great, and ultimately worth the substantial increase in complexity. Moreover, the benefits of a nodal market will increase with further renewable penetration. The future market will require greater locational flexibility from both supply and demand. Only a nodal market sends the right locational marginal prices that will induce the efficient response in real time and motivate efficient investment and innovation in the long run.

Nonetheless, the appetite for a nodal market in Europe appears weak, as nodal pricing comes with substantial changes in the system. The required adaptions from a technical and regulatory point of view are extensive and would lead not only to new roles for market actors, but also challenges in adapting the support schemes for renewables. Therefore, nodal pricing is unlikely to appear in Europe for many years.

One alternative is market splitting—adding zones in response to persistent structural congestion. For example, Germany could split into North and South zones. This would partially address the redispatch problem in Germany, since much of the congestion is caused by moving wind power from north to south. However, market splitting in Germany also seems unlikely. And it is at best a partial solution. Constraints change fluidly over time. Any static zonal structure will be far from perfect.

Thus, the remaining solution on the table, in addition to improved transmission, is to make redispatch work as best it can. This paper examines the design of a local flexibility market for redispatch including renewable curtailment. The market is intended to give system operators the local flexibility necessary to effectively address intrazonal congestion. I focus especially on how the local flexibility market can mitigate inc-dec gaming and thereby improve price signals and investment incentives.

Local flexibility markets operate at both the transmission and distribution levels. Today, local congestion in the distribution grid is primarily addressed with curtailment of renewables in a process that lacks market price signals. The local flexibility market can be especially powerful in rationalizing congestion management at the distribution level, which is becoming increasingly important with the expansion and integration of distributed energy resources. Alternatives to curtailment become visible and properly priced. Moreover, the scope for inc-dec gaming is much reduced at the distribution level.

Effective markets for local flexibility are essential as the market design adapts to new technologies. Demand for flexibility, both system-wide and local, will increase with the penetration of intermittent renewables. Many technologies, both old and new, provide the means to fill that demand. Good market design is vital to motivate the investments needed to provide reliable electricity at least cost to consumers. The future electricity market must support a high penetration of renewables with an efficient mix of flexible resources. This will only happen with good market design.

I first outline the design of a local flexibility market consistent with the European intraday trading model. Then I examine inc-dec gaming. Next I discuss an approach to mitigate inc-dec gaming.

Local flexibility

The local flexibility market I outline here is intended to be consistent with the intraday trading model operated in most of Europe. This is to enable more rapid implementation of the market. Indeed, many of the features are already implemented on a test basis in northwest Germany as part of the enera project.

The day-ahead auction operates as it does today. The optimization identifies a single zonal price, ignoring intrazonal transmission constraints. Local flexibility is introduced in the continuous intraday trading that follows.

The purpose of the market is to efficiently centralize local flexibility offers. This allows system operators, both TSOs and DSOs in a coordinated effort, to reliably and economically relieve physical congestion on the grid close to real time. It also provides a means for flexibility providers to offer and price their services.

The market has a high level of transparency. Prices and volumes by location are public and determined in a clear process. The market encourages coordination between system operators with clear communication protocols. A neutral power exchange operates the market. The exchange uses a continuous double auction with a displayed order book. Physical certification and verification are ensured by the system operators.

Throughout the intraday timeframe, flexibility providers offer flexibility. The offers are in separate order books that run in parallel with the zonal intraday market. The key new element of orders for flexibility is the location of the resource or need. The product is deviation from a baseline (flexibility), rather than energy. Providers include power plants, storage, renewables, aggregators, and virtual power plants. Special attention is given to expanding the set of providers both in terms of voltage (low, medium and high) and technology (power-to-x, combined heat and power (CHP), etc.). Flexibility demand comes from the TSO, mid-voltage DSO and low-voltage DSO. The exchange matches bids and asks continuously. This gives both sides of the market a great deal of control over the timing of trade to resolve congestion.

During periods of no congestion, the global market runs unconstrained. When there is congestion, then a local order book appears for market-based congestion management intraday. If congestion is still not resolved, then the system operators engage in controlled congestion management, as is done today, where a control signal is sent to resources. For example, an upstream wind resource may be instructed to curtail output. Thus, controlled congestion management sits as a backstop in case there is inadequate offers for local flexibility. The advantage of the flexibility market is it enables market-based congestion management. A broader set of resources can offer flexibility and compete in a transparent and technology neutral way to relieve congestion. The approach harnesses the advantages of market-based congestion management yet retains cost-based congestion management as a backstop both for resiliency and to improve behavior in the flexibility market. Offers in the flexibility market are disciplined by a system operator's alternative of cost-based redispatch.

Inc-dec gaming

A well-recognized criticism of market-based redispatch is its vulnerability to inc-dec gaming in which downstream resources, asked to increase, overbid and those upstream from a constraint, asked to decrease, underbid. The intent of both parties is to capture congestion rents and thereby increase profits. The incentive to do so derives from inconsistent pricing between the spot and redispatch markets. In the spot market intrazonal transmission constraints are ignored, but in redispatch they are not. This inconsistency creates an arbitrage opportunity to shift trade to the more favorable market. This is accomplished in the spot market by underbidding upstream—indicating a greater willingness to supply—and overbidding downstream—indicating a reduced willingness to supply. The impact is to aggravate congestion and increase the cost of redispatch.

When pricing is consistent, as in a nodal system where constraints are recognized both day ahead and in real time, arbitrage between the day-ahead and real-time markets is a good thing. It leads to the desirable convergence of the day-ahead prices to the expected real-time prices. The arbitrage improves pricing and leads to a day-ahead plan that is more consistent with real time.

Unfortunately, the redispatch arbitrage is not of this efficiency-enhancing form. The inc-dec behavior is motivated by providers exploiting the model inconsistency between spot and redispatch. The underbidding upstream and overbidding downstream magnify the congestion on the constraint, increasing redispatch costs. Worse yet these inflated congestion costs are going to the wrong parties. Especially problematic is the underbidding of upstream providers, who are being paid more for producing less. The compensation of upstream providers is much higher than their value to the system, leading to excessive investment in resources upstream. On the positive side, the inflated congestion costs are presumably motivating transmission investment to relieve the constraint, but the long-run transmission investment is excessive. Only an overbuilt grid can eliminate congestion nearly all the time.

Inc-dec gaming is a general phenomenon of inconsistent pricing between the spot and redispatch markets. Its implications are most easily seen in the context of a simple example, like the one presented in Hirth and Schlecht (2018), which I summarize below.

There is a single bottleneck separating the lower-cost upstream supply from the downstream market. To further simplify, assume the market is competitive, both upstream and downstream, and there is no uncertainty. No uncertainty means that the providers can correctly anticipate the redispatch outcome when deciding on spot behavior. To make the example concrete, we can think of a highly stylized model of Germany. The upstream market is in the north with ample wind resources and the downstream market is in the south.

It is instructive to compare the market outcomes under three scenarios: nodal pricing, cost-based redispatch, and market-based redispatch. In nodal pricing, the price inconsistency is removed. The bottleneck is correctly modeled in the spot market and no redispatch is needed. In cost-based redispatch, the redispatch is performed using regulated costs, which are assumed to be true costs. In market-based redispatch, providers make offers to maximize profits, correctly anticipating the redispatch opportunity.

Nodal pricing. As a result of the ample wind in the north, there is a low nodal price in the north and a high nodal price in the south. Both prices reflect the value of the marginal resource in each market. The price difference is the difference between the cost of the marginal generating resource in the south and the north. The bottleneck prevents additional lower-cost supply in the north from moving south.

Cost-based redispatch. The spot market ignores the bottleneck and so determines a single price that is between the two nodal prices. The price is the cost of the marginal resource needed to serve aggregate demand. However, the implied spot quantities result in more flow through the bottleneck than is feasible. Redispatch is needed to decrease the spot quantities of the most expensive suppliers in the north and increase the quantities in the south. Redispatched generators are compensated for costs. Those decreased are paid lost profits and those increased are paid costs incurred. Since there is no opportunity to earn profits in redispatch, the spot market is not distorted. However, there are no incentives to offer the flexibility or to invest in flexibility. Flexibility is provided as an obligation of generators.

Market-based redispatch. Without uncertainty, providers can anticipate the redispatch market. This alters bidding incentives in the spot market. Nonetheless, without market power, the final dispatch is efficient. But there are important distortions in the spot market. Anticipating a high redispatch price in the south, south providers overbid until the spot price reaches the redispatch price of the south. (If all the demand in the south can be satisfied by the north, then the spot price equals the nodal price from the north.) The higher spot price means that more generation from the north will be dispatched in the spot, requiring greater redispatch. Northern providers anticipate the redispatch price and quantity. All those northern providers that are decreased in redispatch underbid to the nodal price of the north.

Redispatch creates an arbitrage opportunity that distorts bids. For the north, the opportunity is to be redispatched down at the north nodal price. This puts a ceiling of the north nodal price on spot offers. For the south, the opportunity is to be redispatched up at the south nodal price. This puts a floor of the south nodal price on spot offers from the south.

In this simplified example without uncertainty or market power, the final dispatch is the same in all three scenarios. There are, however, substantial differences in terms of prices and payments.

Under nodal pricing, resources are paid locational marginal prices that properly reflect transmission constraints. The system operator collects congestion rents, which then can be returned to transmission owners and reduce transmission fees. Redispatch is unnecessary.

Under cost-based redispatch, the spot price is the same as in a world without transmission constraints. The system operator pays redispatch costs to generators so that additional profits are not earned from redispatch.

Under market-based redispatch, the spot price is one of the two nodal prices, and therefore may be higher or lower than the cost-based redispatch. Congestion in the bottleneck is aggravated by inc-dec gaming. As a result, the system operator pays higher redispatch costs to generators. Bidding becomes more difficult as providers must consider redispatch opportunities when placing bids. Spot prices can become more volatile and difficult to interpret. The higher payments received in the north motivate inefficient generation investment in the north. For the system operator, the system becomes more difficult to manage as more redispatch must occur close to real time.

Despite its limitations, market-based redispatch would work well if 1) transmission constraints rarely bind, and 2) there is much uncertainty about which constraints are apt to bind, making successful inc-dec gaming difficult. The first requirement seems unlikely to be satisfied without excessive investment in transmission. This is especially true with a high penetration of intermittent renewables located far from load. The second requirement also seems unlikely. Binding constraints are often predictable in the spot market based on near-term supply and demand forecasts. Thus, it seems important in practice to mitigate inc-dec gaming.

Experience with market-based redispatch in the United Sates also supports this conclusion. The U.S. restructured markets in California and Texas began as zonal markets with some form of market-based redispatch. In both cases, the zonal approach failed and was replaced with nodal pricing (Cramton 2017). Inc-dec gaming and high redispatch costs were key factors in moving to nodal pricing. However, the failure of the California electricity market in 2000 was caused by utilities' inability to hedge retail commitments with forward contracting. The PJM and New England markets had an approach more like cost-based redispatch. These markets also found the zonal approach had serious problems and switched to nodal pricing. Efforts to mitigate inc-dec gaming in all these early markets were weak. None of the markets benefited from a market monitor in the early years. The market monitor innovation was introduced after the California market failed.

Mitigating inc-dec gaming

I now explore approaches to mitigate inc-dec gaming. If successful, mitigation would limit redispatch and its associated costs. System operators would access the local flexibility market and redispatch a smaller volume at lower cost and higher transparency. Resources would be motivated to offer flexibility because of the profits earned. And profits would be appropriately constrained by competition among flexibility providers.

A prerequisite for successful mitigation is to promote desirable attributes in the market design. The market should be transparent, competitive, simple, and robust.

- *Transparent*. The local flexibility market should have a high level of transparency. Trade volume and price by location should be public. This is a feature of the proposed intraday trading model.
- *Simple*. The market should be as simple as possible. An important advantage of retaining the current intraday trading model is then the local flexibility market is a consistent extension that appears when transmission constraints bind.
- *Competitive*. The local flexibility market should encourage participation. Transparency and simplicity support participation. In addition, the market should be broadened to include all potential flexibility providers, regardless of voltage or technology. Participation is further motivated by the profits earned from offering flexibility at market prices, rather than cost.

Competition reduces the rents of inc-dec gaming by creating more options for the system operators to choose among and more discipline on offers of the flexibility providers.

- *Robust*. The approach must work well even when there is insufficient liquidity in the continuous market. In the proposed market, the system operators can access additional flexibility as a backstop from the large set of resources that are under cost-based redispatch today. This backstop assures that the system operators have enough flexibility to manage congestion. The backstop also disciplines offers in the local flexibility market.

Mitigation is especially important for high-cost upstream generation asked to decrease. These resources uneconomically schedule production in the spot market, which is then undone in redispatch. The resources are paid for not producing, which they are happy to do—indeed their risk is that they are not redispatched and are forced to generate at a price below cost. Scheduling production that is uneconomic at the spot price is a chief indicator of inc-dec gaming.

This suggests a mitigation strategy. First, upstream resources participating in the flexibility market are required to demonstrate their flexibility. This limits resources from falsely inflating a baseline. Second, where possible, system operators identify uneconomic upstream generators from regulated cost. The settlement then is modified so that the uneconomic upstream generation cannot profit from redispatch. For example, redispatch payments are withheld from upstream generators with regulated cost above the spot price. This is consistent with what would happen under cost-based redispatch for these resources. There is no opportunity cost from not producing, so no payment is required. This is a form of automatic mitigation. It is fast and effective in instances where the regulated cost is a good proxy for true cost.

Explicit rules against inc-dec gaming would provide additional mitigation. Participants in the local flexibility market would agree to no inc-dec gaming. Sanctions would result from violations. The challenge here is that detecting violations may be difficult, especially under continuous trading, where offers reflect market prices rather than costs. Still, explicit rules against the behavior would increase the risks of inc-dec gaming. Eliminating the obvious abuses can go a long way in limiting the costs from inc-dec gaming.

For certain resources with clear baselines, such as CHP and Power-to-X plants, the baseline can be used to identify gaming. Since the baseline should on average be the same regardless of whether there is congestion, a baseline that varies with congestion suggests inc-dec gaming. Other resources, such as wind and solar resources, have a clear baseline that varies with congestion, but this is not inc-dec gaming. Their energy production absent curtailment is observable, so inc-dec gaming is not an issue.

A related mitigation approach comes from the examination of production and consumption schedules. Each participant in the local flexibility market submits unit-specific schedules every day to the market operator. Inc-dec behavior would be identified from persistent differences in schedules on days with and without congestion. For example, an upstream generator engaged in inc-dec gaming would inflate its schedule in congested periods to increase its redispatch payment. Redispatch payments would be withheld from those engaging in inc-dec gaming. One would need to be cautious not to withhold payments from those engaging in perfectly legitimate behavior. Wrongly withholding payments would discourage participation in the market.

The mitigation approach should be narrowly tailored to apply where participants have both the capability and the incentive to engage in inc-dec gaming. This will vary based on the type of resource and the form and extent of congestion. Mitigation is most important in instances of significant structural congestion that is easy to predict. Then mitigation should be applied to those resources that are able to take advantage of the arbitrage between the spot price and the redispatch price that inc-dec gaming exploits.

The mitigation strategy is a hybrid of market-based and cost-based. The cost-based backstop assures that system operators will have enough flexibility to manage the system at reasonable cost. The market-based first stage expands the set of participants offering local flexibility, providing more options for the system operators to manage constraints as they appear.

These mitigation approaches are imperfect. To improve them over time, a role of the market monitor is to review redispatch (including curtailment) and the performance of the local flexibility market periodically. When problems are identified, the market monitor makes recommendations for improving the market. These would include better methods for addressing inc-dec gaming.

Despite imperfect mitigation approaches, market-based redispatch is the second-best solution at hand if nodal pricing is not possible. The integration of demand-side flexibility is a necessity and cannot be achieved in a cost-based regime.

Conclusion

Europe's zonal energy market would work well absent intrazonal congestion. Unfortunately, the rapid expansion of intermittent renewables introduces transmission constraints, which must be addressed at least in the medium term with redispatch. As constraints appear, the system operators redispatch to relieve constraints. The issue is especially pronounced in Germany, where substantial wind resources in the north are needed to serve load in the south. The growth in transmission has been unable to keep up with the growth in renewables. Substantial and costly redispatch is needed to relieve congestion.

The pricing inconsistency between the spot and redispatch markets creates an undesirable arbitrage opportunity. Inc-dec gaming exploits this opportunity. Resources both upstream and downstream of a transmission constraint anticipate prices in the spot and redispatch markets and distort offers to maximize profits. The behavior aggravates congestion and inflates redispatch costs.

I describe approaches to mitigate inc-dec gaming. Provided structural congestion is not a pervasive long-term problem, the mitigation may be effective at addressing intrazonal congestion without excessive costs and inefficiency. However, if intrazonal congestion is too great, it may be preferable to adopt nodal pricing, especially in the long term.

References

BNA (2019), "Quartalsbericht zu Netz- und Systemsicherheitsmassnahmen: Erstes Quartal 2019," Bundesnetzagentur.

Cramton, Peter (2017), "Electricity Market Design," *Oxford Review of Economic Policy*, 33:4, 589-612.

Hirth, Lion and Ingmar Schlecht (2018), "Market-based Redispatch in Zonal Electricity Markets," Working Paper, Neon Neue Energieokonomik.