Increasingly, regulators, customers and third parties are pushing for more innovation within the electric utility industry. Utilities have been slow to move much beyond small pilots, and these piloted programs tend to focus on one or two aspects of the opportunity (e.g., DR, voltage, behavioral-based customer engagement). None of these pilots jointly addresses the cost savings accruing across all of the avoided cost categories that exist in the varied and diverse silos of utility value (e.g., supply, ancillary services, bank deferral, line loss mitigation, KVAR support, voltage impacts, customer-specific avoided costs).

The realization of the need for this joint consideration of both grid and supply, and the need for distribution level clearing prices to spark efficient innovation, has been apparent for some time (e.g., see Caramanis, March 2012, IEEE Smart Grid, or more recent posts on “Transactive Energy”). But a comprehensive modeling and valuation process is still missing. Nothing exists, quite yet. And so, the questions continue:

“How do we get more innovation within this space?” “Where is that efficient price signal”?

“How do I know when the utility death spiral begins?” “Is it real, and how will I know?”

“Will California’s recent push for more granular effectiveness/ targeting spread to all of U.S.?”

“How can we integrate grid-side costs and supply-side costs into a single framework?”

We address these questions directly here. Our work over the past few years has specifically focused on uncovering, and quantifying, the “hidden costs” within the utility. The purpose is to arrive at a specific price signal of the type shown below by customer and location (e.g., voltage, power factor, circuit capacity deferral, long-run LMP or Locational Marginal Price, and others). This work has required a much more granular level of customer forecasting and optimizations. But the detailed granularity is exactly what is required to jointly value both grid-side and supply-side opportunities, simultaneously.
In essence, these DMPs are simply the “shadow prices” (natural product of formal optimization modeling, analogous to derivations of LMPs) from our IDROP software. These shadow prices ARE that efficient price signal, just like LMPs are shadow prices for the transmission network. Here, we call them the “Distributed Marginal Prices” or “DMPs”. In the above map, we show DMPs for a 4pm example, for one circuit in the West. Red areas denote higher DMPs and blue areas are lower priced locations. We review the details of how this is done in the following pages, but first we discuss a bit about the concept of the DMP and some of the industry-changing implications that are inevitable, if DMPs are widely adopted.

The true strength of this approach is that it simultaneously and jointly values both supply-side (KW) avoided costs and grid-side costs (KVAR, voltage, power factor) at the same time. It forces collaboration within the utility across silos. It reveals a single price signal per house, per customer, for third parties to see with certainty, which will spark considerable innovation, and do so at exactly the right PLACE, the right TIME, and right AMOUNT. We are not only including the short term, more operational value and benefits, but we also incorporate the longer term benefits (capacity deferral for T&D, commodity cost to serve, future LMPs) such that resources with higher fixed costs can participate. This stands in contrast to the current status quo which tends to favor either the grid or the supply, to the exclusion of the other.

**What We Cover**

We start with an overview of eight, or so, key strategic implications of how DMPs can advance the industry, and bring regulators, utilities and third party innovators jointly to the table. The goal: an efficient grid at least cost. Second, we lay out our process and methods which can be used to obtain the required granular, locational avoided costs (in contrast to current averaged avoided costs). Third, a series of critiques to DMP-type approaches are provided which we expect will arise from various parties. This helps frame how DMPs might be used, or abused. Last, we offer some brief thoughts on regulatory policy. We try to remain agnostic on policy, focusing primarily on the math and the methods. But, given some risks we see with rate basing and decoupling, we recommend that shared savings type earnings mechanisms be promoted by regulators to encourage utility participation.

This white paper provides significant detail and insights into the work we have done over the past years toward the measurement of more granular avoided costs. Even if DMPs never become an actual market-traded price, the methods and models described here show regulators and utilities how to more intelligently calculate the avoided costs that lie at the heart of our work. All we have done here is “sharpen our pencils,” with respect to the measurement and calculation of avoided costs. The States have a mandate to provide reliable service at the least cost. To know if one has the least cost, you must measure costs. In the past, average avoided costs were fine. But with the advance of PV, storage and other distributed resources, we are now required to apply locational avoided costs at a much more granular level. Our DMP methodology is nothing more than that. It identifies the marginal contribution of a KW, and now a KVAR, to each customer. That is the Distributed Marginal Cost (DMC). We call it a DMC, even though it is the same as the DMC, simply to remain consistent with the industry’s conceptual understanding of LMPs. You don’t need an ISO or DSO, or a market traded DMP, to achieve grid efficiency or spark third party innovation. Simply use the methods described here, and call it a DMC. The calculations are the same. The only difference lies with how you choose to use it.
IMPLICATIONS

There are several key implications of how the use of DMPs will change the industry. Some of these will be almost immediate, and others will take more time.

First, and most importantly, DMPs provide a transparent and efficient pricing signal that reflects the true costs to the utility. Currently, most of these costs are hidden from third parties. Indeed, many of the costs are actually hidden from the utility themselves, at least at the marginal level. Only average tariff-based pricing has mattered to utilities to date, so there has never really been any motivation to uncover what these marginal costs were. The changing mix of resources below the bus, however, mandates that this type of local marginal cost analysis be done. And in creating the DMP, we have a common metric on which to more accurately value which micro grid resources matter, and where.

Second, the DMP opens up a transparent price signal to third party innovation. Google, EnerNOC and other’s apparent frustration with utilities likely stems, in part, from not having this type of clear transparency of costs. Once established, DMPs are likely to spark significant innovation in exactly the right places. Vendors will know what the payoffs will be for their KVAR injections, voltage support, DR, solar and many other resources. They will be able to take financial risks with their higher fixed cost assets and programs, and plan accordingly. Whether or not utilities participate is unknown, nor is it known if their activity will be regulated or non-regulated. But in all cases, significant innovation is inevitable, sparked by a simple DMP. Expect Samsung, Apple, Microsoft, Honeywell, Comcast, and a host of others to now begin integrating their existing services with utility services. The DMP opens up access to just about anyone to participate, yet still ensures a DSO focus on reliability.

Third, we jointly value both grid-side and commodity-side costs in a mathematically integrated fashion, not too dissimilar from current LMP derivations. As such, DMPs should offer a commonly understood metric for utilities, regulators and third parties. Contrast this to the current state where we find many pilot activities across the country, each myopically focused on just one or two of the several cost silos or value buckets. Voltage control pilots. DR pilots. Behavioral programs. Audits. Distribution automation. The list goes on. Few, if any, consider the impacts of their efforts on the other utility silos. A single DMP resolves this myopia in a fairly elegant and accurate way. In this sense, DMPs are the killer app.

Fourth, DMPs may also be the start of the utility’s death spiral, or eventually be called the “utility killer”. If all innovation now comes from third parties, and those third parties then begin to hedge their supply through power contracts, or physical iron in the ground, then the utilities risk being simply a “wires” company. With perhaps 70% of utility margin accruing to supply side activity currently, the financial impacts to utilities could be very significant. Of course, utilities can opt to participate via non-regulated affiliates. Or perhaps regulators will mandate that some percentage of the resources be housed within the regulated utility. This tact may be necessary where regulators observe “reverse gaming” by third parties. Here, just as Enron and others withheld supply years ago, to artificially spike prices, so too can demand side third parties artificially pre-cool homes, heat water and run pumps, to spike demand and loads, leading to high prices. Then, the third parties will “double dip” during the afternoons, by taking DR price credits when prices are high. The home is already pre-cooled, so the customer won’t notice much, but the financial impact will be significant. We have calculated that a large enough third party can game this system with as little as 10% of the load under its control. The only utility hedge is to have load under its direct control, to mitigate the gaming. Finally, we may see regulators adopt a new perspective on utility
investment in these seemingly non-regulated type activities. In reality, the DMP framework shows us that these micro grid resources and DG (distributed generation), which tout the KW benefits, can have positive benefits to the grid, voltage, KVAR and “wires” type impacts. As such, these micro grid investments could easily fall under a regulator’s purview, motivating utility investment in micro grid assets directly, instead of just third parties. And given the potential for third party gaming, this type of blended approach to regulating the new paradigm has some merit.

Fifth, DMPs will lead to additional job security for regulators and policy analysts. Many and varied policy discussions will be had, State by State, regarding how many and how much should be included within the DMP price signal. Is it only short term focused? How many of the longer term cost buckets should be considered? How does the DMP impact specific customer segments? All good questions.

Sixth, the DMPs will clearly motivate development of micro level resources. This makes the job of the distribution planners much more difficult. The newly created DSOs will have their hands full managing the process, protecting the grid, and still desiring to create efficiency across their circuits. Hosting fees for solar customers that self-generate 99% of the time, but still want that option to get grid service in a pinch, and other examples will change the way that utility prices and tariffs get structured. We are likely to see more and more KW centric tariffs vs. KWH, as more and more of these innovations push customers toward self-generation. We will also see regulatory discussions around individual settlement shapes and individual customer tariff pricing options (unique $/KW or $/KWH vs. TOU pricing). We are likely to see third parties create sophisticated hedging services, even flat bill options where the third party accepts the risks and customer’s bill volatility is zero. We expect increased attention to de-regulation within states that have resisted this path in the past. And we can probably expect a rapid adoption of new resources that, in the past, were considered “out of the money.” Once unlocked, the DMP makes transparent the hidden utility costs and, almost by definition, suggests more and varied micro resources lie in our future. The most valuable job in the utility may soon be the Distribution Planner.

Seventh, DMPs will force changes in how regulators and utilities approach IRP planning and cost effectiveness valuations, especially for micro grid resources like solar, EE, DR, etc. In the past, these evaluations rarely took KVAR, voltage or power factor considerations into account. The focus was almost solely on KW/KWH reductions. Now, we have a DMP valuation platform on which we can assess both the KW/KWH contributions simultaneously with the KVAR/voltage/power factor benefits. Moreover, the reduced latencies of the cost effectiveness and IRP analyses will necessarily migrate from hourly level to sub-5 minute valuations. And as the time latency decreases, the covariance between loads and prices tends to rapidly increase. This requires a much more robust and comprehensive set of valuation tools than are used today. We have found that our 5 minute level valuations for resources such as thermal ceramic heating bricks (for wind firming), physical battery dispatching and water heater frequency following yield 2X to 5X higher valuations when analyzed at this 5 minute latency. In practice, our current IRP models will not go away. Rather, they will be supplemented by a series of DIRPs, or Distribution IRPs per substation. We already have optimization systems that value power flows within the networked set of buses substation to substation. What is missing is the granular valuation below the bus (DMPs). Both matter. And DMPs do not replace the need for system wide IRP planning. But the operation of DSOs, and the increased efficiency and innovation sparked by DMPs will almost certainly inform the system-wide IRP plan, for the better.
Eighth, the DMP signals can also be provided for “next day” or even “next years” as estimates or forecasts that can be used for investors to judge the cost effectiveness of their resources. Such DMP forecasts can be used to value the payback threshold for, say, EV charging during the day or thermal ice storage during high solar gain periods, to mitigate the risk of “duck curve” system loads. The financial analysis is made transparent and easily observed, such that investors’ financial risk is substantially mitigated to the point where investment decisions can be made to resolve coming changes in system or circuit loadshapes that simply shift peaks to new hours (early evening vs. afternoon). Without some type of transparent pricing signal, these innovations will be slow in coming. As well, wind intermittency, solar intermittency and other issues resolved by local ancillary service type resources benefit from knowing the value of their resource at a 1 to 5 minute latency level. And as these resources lower the load volatility, the updated DMPs will reflect the diminishing returns, as more and more of these resources are added. So, there is a natural valuation process which serves to limit over-spending on any circuit or area, too.

Finally, it is likely that we will see some utilities embrace the notion of DMPs and others will resist it. But what is knowable is that, for a utility to know whether to embrace or resist, they will first need to beef up their analytics, by customer, and conduct the necessary financial analysis to know what to do. This can only be identified by using some, if not all, of the types of very granular forecasting, optimization and valuation approaches described later in this paper. DMPs are a single metric of value, but within DMP calculations lie all of the “hidden costs” and hidden value. Once utilities embark on the more detailed financial valuations, they will see unique opportunities for margin, on both the regulated and the de-regulated side. And the key to utility survival is in knowing where future margin exists. In fact, the utilities are best positioned to know these future margins, before regulators or third parties receive the transparent DMP signals. Utilities are the source of the data. So, one can surmise that a reasonable leading indicator of a utility’s future stock price may lie within investors’ assessments as to whether or not the utility is actively using the DMP-type granular methods which inform the DMP. If they are not, it is extremely unlikely that they will know what future margin strategies to pursue and which to resist.

Of course, there are more implications that we have yet to uncover, and we welcome your opinion and insights in this regard. Our purpose here is simply to lay out a future vision that is not only conceptually grounded, but technically and practically implementable, today. Yes, the solution described here does rely heavily on our own software. But this software was designed, in the first place, to create the DMP (they are the IDROP shadow prices) and we knew we had to unlock the many and varied, “hidden” costs within the utility silos for IDROP’s optimization framework to really shine. And over the past few years, we have proven out this capability, the feasibility and the need for more granular cost analysis and optimization.

We see the increased attention to these issues in NY and CA, particularly. Other States will follow suit. Europe and Australia appear to be moving in this direction, as well. And with Google and others working fervently in this space, solutions need to occur quickly. Yes, our agenda is to sell software, but we also care about grid reliability and fear Enron-style demand gaming. So, we feel the need to openly share methods and thoughts, to speed progress toward a reliable grid at least cost. We do hold some patents in this space, but are happy to consider limited free use of these to spark innovation, since we all will benefit. We do favor a cost-based approach within our optimizations as our patent approvals lie with the use of these direct utility avoided costs, versus simply responding to the familiar, zonal or local LMP. It
is exactly these utility avoided costs that currently remain “hidden,” yet they form the backbone of regulatory cost-to-serve mandates, and which must be included in any type of distributed avoided cost approach. These are the key to unlocking the innovation we seek.

The Process

We are all familiar with LMPs. Current ISOs provide a zone-level LMP for an area that is based on the weighted average of the local bus-level marginal costs. One can call this a “macro” LMP applied to all customers served by that substation or transformer bank. It is a straightforward step to simply pull the local bus-level LMP, and apply its value to all customers (per KW basis) within than local substation area. No one would argue this is infeasible, though some would balk (briefly) at the potential for discriminatory pricing policies. Nevertheless, this local LMP is the cost to serve that area, on average, and supports accurate cost-to-serve regulatory policies.

We separate these policy considerations from the current discussion, favoring the advancement of the technical/conceptual discussion. But no question, at the end of this discussion, several important policy issues become obvious. We simply ask that you suspend these objections for the time being, and see what is possible, and even practical, first.

Our IDROP software performs similar types of forecasting and optimization analytics as those used by the ISOs in their LMP calculation process, but the big difference is that IDROP does this at a much more granular level (by customer, even by end use) for all customers served by substation, and all circuit sections below the bus. These “micro” LMPs can be more descriptively called a DMP, or Distributed Marginal Price.

DMP price signals value, simultaneously, both grid (KVAR) and supply (KW). Here, we directly value KVA and power factor, and put “voltage/kvar” and “KWH/KW” on a level playing field, valued by a single, local DMP price, by customer site. In some cases, KVAR improvement may constitute the bulk of the DMP value. In other locations, it might be more KW centric. In most cases, it is a blend of both, and any KW reduction will also have voltage and power factor benefits. The DMP price can be split into its KW vs. KVAR components for use in targeting either grid-side or supply-side resources, but the key point is the DMP puts both on the same playing field to compete for resources and valuation attention. Moreover, the DMP can simply be added to the LMP per bus, and this TOTAL price signal then revealed to third parties to drive investment planning and allocation of resources. This creates the very type of efficient price signaling that third parties seek (e.g., Google, ComCast, Samsung, EnerNOC, solar/wind providers, many others). Moreover, it focuses the right type of attention (KVAR vs. KW) at the right location, which benefits the DSO (Distribution System Operator). The concept of a DSO establishes a new role for the distribution function of utilities. While the associated policy concerns are outside the scope of this paper, discussion can now certainly begin. Someone will need to take responsibility for balancing reliability needs with the desire for innovation and advancement in customer-focused Smart Grid solutions.
The Method

Let’s step back a minute and identify the “value buckets” that comprise the DMP. Below is a descriptive categorization of the value opportunities across both supply-side and grid-side sectors. Our DMP jointly integrates both, placing both on equal footing (in terms of dollars saved). The DMP is the price for the next increment of KW/KVAR at the house (or customer). We can calculate the DMP price down to the end use, as well, but let’s assume a customer-centric view to start. Some examples include the following:

Both KW and KVAR are forecasted, given the weather and, in the short term, time-series/econometric based trending. Usually, utilities that are supply-side focused only forecast KW, while distribution-centric innovations favor power factor. Both matter, and both must be modeled simultaneously on a level playing field. Utility IT Departments have generally been averse to spending money on the collection and archiving of KVAR data, favoring only KW for use in billing. But the birth of the DMP requires that we track both. With both, we can now have KVA forecasted directly, and can assess impacts from power factor changes, voltage drops, and line loss impacts, the key grid-side value buckets. These forecasts can be estimated, or updated daily, with the use of a nightly AMI system download. Of course, 15 minute KW and KVAR data reads are preferred, in near real time, but remain at the mercy of the IT and telecom cost to collect it, at that latency. Two-way communication apart from the AMI system, either by house, or across sections of circuits, in more real-time, would help, too. But for the time being, we will assume that these KVAR and KW forecasts are reasonably accurate, irrespective of its source or its latency.

Now, in practice, we can calculate the impacts from a 5 minute change in KVAR/KW by using batch processing of the utilities distribution power flow tool, and produce estimates of the resulting changes in voltage, losses and other grid-centric effects. These can be converted into their dollar equivalents and input into Integral Analytics IDROP optimization engines to determine the shadow price (DMP) contribution or value for placing one increment of KW or KVAR, or both, at specific locations. We can
perform these analytics simply using the daily forecasts, without batch run power flow tools, to test and run use-case scenario analysis. But the more robust solution must directly integrate the distribution power flows in near real time, in the long run. We have been working with several of the major tools which appear to be sufficient to the task. Given the prevalence of existing 5 minute LMPs, we assume that this 5 minute latency, or even hourly perspectives suffice for the time being. The loss in assuming a 5 minute view lies with the under-estimation of the value accruing from ancillary services that are responding to a 4 second, or shorter latency, signal. In addition, various arbitraging and dynamic dispatching projects we have completed will require this (e.g., see Integral Analytics WindStore, GridStore, IDROP arbitraging, wind firming examples on our website, or ask for more details).

The discussion to this point has focused on short-term valuations, generally within the hour, or within a day or month. We extend the methodology to also include pricing components for longer term capacity value and deferral. These are often more important than the short-term cost savings. Indeed, much of the regulator’s current attention is squarely focused on longer term value attributable to solar, EE, DR, wind and other innovative supply-side and grid-side resources. To do this, we leverage our LoadSEER software which is a comprehensive econometric/geospatial forecasting tool. There is substantial detail and sophistication embedded within LoadSEER. It can be used for several purposes beyond DMP calculations, but the key aspect here is that we can forecast very granular load growth, or reduction, at the customer level, or acre level. The reason we designed LoadSEER was to optimally target DR, EE, solar, wind, etc. to the right locations such that we maximize T&D avoided costs, identifying which circuits would benefit the most from DSM, first, from implementation of programs to defer a capacity need for the grid. Since LoadSEER forecasts 10 years, or more, of very granular local growth, we are able to now forecast 10 year LMPs instead of just next day LMPs. This greatly advances the ISOs ability to plan and locate future capacity, and to identify future areas of congestion (and hence, higher LMPs). And most importantly, it shows both third parties and utilities where to best locate micro grid resources (or new supply).

We always remain consistent with the overall Corporate Forecast of load growth, reconciling forecast error and differences in certain areas. But it is made easier and more accurate each year as the distribution planner becomes a more integral part of the forecasting process. The planner actually locates new load within the tool, in real time, month to month (e.g., Walmart store, Honda plant, residential building) and the forecasts are automatically updated to zero in on “the truth.”

As an interesting side note, consider this. Distribution planners historically were never consulted, nor a part of, the utility’s IRP (Integrated Resource Plan) process. It was typically the Corporate Forecaster working with the Supply Stack staff (the IRP) who then worked with the Rate Department to create average tariff prices for customer classes. No distribution planner was needed. Today, with increasing solar, advancing DR, intermittency, electric vehicle load growth, etc. the distribution planner arguably plays the key role in the DSO, managing hundreds of DIRPs in some cases (e.g., one DIRP per substation/bank). In addition, with advance of new supply technologies, rate designs will be forced to become localized or even individualized to reflect the local cost to serve as well as competitive supply costs. Rate Departments will be relying on the distribution planner to give them detailed views on these costs in the future.
An example circuit level forecast is shown below, using LoadSEER. Here, the red polygons identify circuits at risk of exceeding capacity (targets for DSM, KVAR) and green areas depict areas of zero risk (alternatively, ideal zones for EV charging stations, new economic development).

![Image of circuit level forecast](image)

Engineers also enter the cost of capacity additions so that the Corporation can see not only where T&D avoided costs are highest, but when and how much. Moreover, because we use multiple forecasting methods within LoadSEER, including a geo-spatial methodology based on NASA satellite histories of 30 years, we are able to quantify regression-based functions of how load grows, how it clusters together, depends on proximities to roads, economic centers, hospitals, airports, entertainment, etc. This enables a more accurate forecasting of new events (e.g., new highway, commuter rail, Honda plant, EV charging locations) all of which have no load history on which traditional econometric regression modeling relies. With no data history, one cannot forecast anything with regression. Hence, the need for geo-spatial methods. The other benefit from using this approach is that LoadSEER performs the forecast using three independent methodologies, including the geo-spatial method, an econometric/time-series method applied to KWH, and a regression on past peak KW circuit loads (using both weather and 100 economic factors). This approach enables the distribution planner to triangulate these forecasts. If all three forecasts essentially predict the same future loads, the planner has confidence in the prediction. Even if 2 of the 3 are similar, some confidence exists. Today, planners typically use only one method, a simple regression, and it usually only includes temperature. In cases where mild weather coincides with a down economy (as occurred from 2008 to 2011 in many areas), these forecasts are quite biased and inaccurate. If the economy returns, and it is an extreme weather year, the utility is blind to a considerable risk.
The graph below depicts a typical forecast where the predictions converge (Western State, summer peaking).

The methodology also directly separates risks from weather vs. economic factors. Some micro resources like DR or HVAC efficiency tend to impact the weather sensitive portion of the forecast (the dotted line forecast), whereas other technologies like solar tend to reshape more hours (impacting more of the economic, weather normal forecast). In addition, LoadSEER tracks all impacts at the hourly level such that the correct circuit coincident KW reduction is captured (e.g., solar may only reduce a 5pm peaking circuit at 30% to 40% of its nominal rating, the rating expected at noon, or peak solar gain). In the example below, one can see the larger spread between the dotted line forecast (1 in 10 year weather risk) vs. the weather normal load (economic risk, 1 in 2 year risk).

As a result of the acre-level forecasts, we can produce a very localized and targeted set of resource plans capturing both grid asset additions, as well as value coming from micro grid resources, such as EE, solar and DR. This creates a focused plan by circuit, at the right time, leading to significantly improved “bang
for the same buck.” This stands in stark contrast to the current mass marketing type efforts currently employed by utilities, and differs substantially from the current use of an average avoided T&D cost applied equally to all customers, for evaluating demand side resource cost effectiveness.

Where the typical avoided average T&D cost is, say, $50/KWyr, in reality this average value is a weighted average of quite a few $0 values, where no capacity is needed, and perhaps 10% to 20% much higher value on constrained circuits. So, right off the bat, a more nuanced, granular and targeted costing of the grid asset deferral savings has value. And this value can be incorporated into the “long term” portion of the DMP price overall.

Again, there are several policy questions to be asked (e.g., refer to map below, do we lower prices in the green areas?), but we remain agnostic regarding the eventual incorporation of all, or some, of this localized capacity deferral value. The important point is that it is technically feasible to do so, and this can be used to optimize investment and resource allocation.

The same long-term valuation philosophy can be applied to the commodity. We have discussed how LoadSEER can be used to forecast 10 year LMPs, by bus, for incorporation into the DMP, either in whole or in part. Additionally, we can accurately calculate the customer’s cost to serve (commodity) using commonly accepted mark-to-market valuation methods (albeit at a much more granular level, requiring significantly more processing and analytics). As we all know, tariffs are average prices reflecting an average cost to serve for a customer class. The reality of this process is that significant cross-subsidization occurs not only between classes, but WITHIN classes.
In the example below, we have calculated the actual cost to serve for each of several thousand commercial customers.

To obtain these valuations, we run our DSMore software for each customer, to calculate the exact (long-term full requirements price) cost to serve, for energy and capacity, for each customer. We use a series of econometric regression functions to forecast customer load (with AMI data, the forecasts are very accurate, but we can also use monthly data with “borrowed” hourly shapes from customers with similarly situated characteristics, demographics, firmo-graphics, etc.). We assess impacts from temperature, humidity, wind speed, cloud cover, economic conditions, “bend over” effects, splines (slope changes), and other factors, by hour, by month, by day-type, for each customer. We are able to estimate the customer loads through 30+ years of actual hourly weather for the local micro climate, and couple these hour by hour load forecasts with 20 GARCH-based forward pricing curves, keyed to the local ISO hub, or utility’s system lambdas. The result is approximately 700 mark-to-market valuations.

To the extent that our statistical forecasts are accurate, say 95% goodness of fits, this is analogous to having an AMI meter installed on the customer’s site for 30+ years recording usage with 95% accuracy. By customer. It took us a few years to automate this process, and get the point where we could analyze hundreds of thousands of loads, but it was worth it. The result is, literally, the true cost to serve for each house (at least within the margin of statistical error, the 95%). In the graphic above, where the average tariff price is 6 cents for the commodity, the range of true costs goes from 3 cents to 10 cents. This is the weather normal 8,760 average price which one would charge this customer, if they were a non-reg provider of full-requirements power.

Again, we can hear the regulatory policy analysts cringe at the technical feasibility of individual customer tariffs, and of the potential for individualized settlement shapes, and the likely adverse harm unfairly placed on some low income customers or small business (or large). Agreed. But note that simply because we are technically capable to price out the true cost to serve does not mean we are forced to apply
it. Rather, policy analysts will debate the issue and likely arrive at some type of compromise that rewards the higher cost customers a touch more than the low-cost ones. And this policy decision can be applied as a judged amount, applied to the DMP, vs. using all of the true cost to serve. The inclusion of some portion of the cost into the DMP does spark innovation at exactly the right place. As such, we include it in the DMP. The only question is how much is enough to include, or how much is reasonable. Or, is there only upside financial gain to customers and no down-side loss or risk (i.e., apply some DMP portion adder for all, which is an incentive to customers, and is higher for higher cost customers)?

Finally, some people ask us, “Is all that granular level processing of loads and prices really necessary?” Answer: Yes. We tried short cuts, such as applying Black-Scholes type derivative methods, and averaging loads and prices over longer time periods, but the results always underestimated the avoided costs, especially at peak times (95th %ile to the 99th %ile). So, we inevitably returned to a more comprehensive full enumeration of avoided costs, and leaned on causal-based methods instead of the easier, but insufficient Monte Carlo and Black-Scholes type approaches. When we used any form of load or price averaging up front, we lost the important hourly covariance information between prices and loads. We always saw a clear underestimation with pre-averaging. With the more robust method, we not only get quite reasonable and accurate covariance forecasts for 30+ years, keyed to the customer, we also get a complete distribution of loads and costs. This full distribution is necessary to identify the costs at the extreme tails, where prices spike and weather is extreme. This is the core source of value for many Smart Grid resources. If we under-estimate, or if we use assumed Black-Scholes distributions (or 95% VAR), we miss the critically important 99% load, or the 99.5% load where significant cost savings and attention is focused.

So, until the industry has enough storage under its belt, at least enough to dampen price volatility, we prefer the more accurate, albeit data intensive, approach. Besides, it represents the true cost to serve. No shortcuts. Below is a typical, full 3D distribution of avoided 8760 hourly costs for a customer, across 30+ weather years and 20 forward pricing/cost curves. Note the obvious skew, as prices climb and weather gets hot. This is the covariance distribution that we want, and which we create for each customer. Any averaging necessarily leads to an average cost to serve estimate that might be as much as 2X to 3X lower than actual during these times.
A final clarification is needed with respect to the commodity cost to serve component of the DMP. We refer to this cost as if it were all energy. In reality, we are using the energy usage as a means to get at the capacity cost for that customer. The true cost of energy for customers within that substation is the bus level LMP for that hour. But this has no capacity value in it at all. To get the right capacity cost per customer, we must analyze that customer’s load across many actual weather patterns, and over all possible forward curves (e.g., during extreme weather, forced outages, high natural gas prices, etc.). As the forward price markets “boom” and “bust”, the covariances between prices and loads change. The only accurate way to value this is via the full enumeration process we lay out here. Fully modeled, we can observe and calculate how much capacity each customer contributes to the total, and then add just this piece to the LMP. We rely on the hourly LMP for the energy component, and this is already embedded within the DMP directly. The issue is similar for the DMP adder associated with long-term capacity deferral of banks/circuits. We do not add into the DMP value a constant value or a per KW value. We assign the DMP component value based on the coincident hour’s contribution of that customer’s load with the circuit’s peak. In this regard, street lighting is not likely to get any value for circuit capacity deferral represented in its DMP. Similarly, the DMP adder for the commodity cost to serve (aka capacity) will also be zero.

Moving on to the grid assets, another component of the DMP lies with grid asset protection, such as primary or service transformer overloading predictions. Traditionally, utilities would approach their Load Research Department for typical customer shapes, to be used in the assignment of service transformers to locations. See a typical example below which compares the (over) averaged load research shape vs. the actual shape.

Generally, the field crews, or planners, are risk-averse, desiring transformer loads to be around 30% of the nominal transformer rating (most of the excess to accommodate cold-load pick up). However, over the years, some homes build additions, take up a pottery hobby (20 KW electric kiln), or clandestinely install grow lights. Instead of using heuristics, or averaged tables, it is a very simple matter of calculating the exact loss of life expected per transformer (ANSI tables) whether winter or summer, pole mount or pad mount, and across various durations of above-nominal rating hours. This more granular approach is much more accurate than heuristic algorithms, and moreover, allows us to simulate added electric vehicle loads on top of existing loads. Since utilities will never know exactly which house adopts what EV, this process enables the identification of overloading transformers well in advance of their adoption. Adding DMP to the valuation helps optimize investment, and improves asset protection and reliability. Moreover, field crews can change out these overloaded assets with comparable rated under-loaded ones, possibly even on the same street.
Our results from this analytical process (see example below) depict that all but a handful of service transformers are just fine, but 2 or 3 of them should be replaced now (or homes targeted with audits, DR, EE). Too much solar, and reverse power flow, might exacerbate the problem, but at least we now have a firm grasp of the problem and have quantified the risk. Additionally, adding EV loads reveals several more transformers potentially at risk, and these can be addressed during non-peak times for the field crew.

The same process can also be applied to primary transformers, sections of circuit, specific wiring issues, and other grid assets, as they generally are aggregations of the micro-level load forecasting we already have in place.

One additional value bucket component of the DMP also lies within our IDROP optimization and arbitraging, or choreography, of loads. IDROP can be used to optimally coordinate, or choreograph, the end uses along the circuit as a function of the total cost to serve for the circuit. We find that about twice (to 3X) the cost savings can be achieved by optimizing over the circuit, versus having each house optimize their own bills. This can also be done in a manner that additionally pays attention to the nominal ratings of the service transformers. When DR signals are optimally coordinated, we can dramatically lower the number of ratings violations and their durations. The load that the transformers experience is much flatter, thereby preserving asset life. Note that the choreographed loads are not flat for two reasons: 1) we are maximizing the total costs savings, not focusing just on the transformer, and 2) there are important engineering constraints placed on us by the end-use equipment creating little “bumps” in the flattened load (e.g., HVAC needs to run a minimum of 7 minutes, EVs must charge 35 minutes minimum to get the temperature raised enough).

These engineering constraints lower our total cost savings, but importantly preserve the appliance life, and hence, customer satisfaction. The DMP contribution in this case is also an improvement in the power factor, which has a distribution dollar value that can be incorporated into the DMP (either estimated, or calculated from the local power flow model). Moreover, this type of optimal choreography can be targeted to areas of weaker voltage support (e.g., the DMP algorithms pick up the voltage drops and IDROP then identifies this area as deserving of a higher shadow price, or DMP contribution). This is a natural outcome of the DMP optimization calculations, and does not need to be estimated externally. An example is shown below taken from one of our IDROP pilot projects, for about 30 customers on 5 or so...
transformers. Note that this flattening of the load is achieved with 30% of customers participating in the program. Essentially, the optimization algorithms “dance” around the non-participants to create this effect. Of course, this requires some type of one way signaling from non-participants regarding the near real time total load.

If we extend this example to the circuit, we observe the type of results below. This is not an actual outcome, it is simulated. All of our pilots to date consist of 50 to 1,000 customers and none are collocated all on the same circuit. But applying the known results to a peak day for an example circuit, we can see below that the circuit load can be “flattened” using only a couple appliances and 30% participation.

One can also add in EV loads, all arriving at 4pm from 10% of the customers, heroically assuming that they all agree to nighttime charging. But we see a few important insights from this simulation.

First, we only needed 30% participation. We will likely never get 100%, but even if we could, it would be inefficient to pay them to participate (diminishing returns). Second, we have created virtual storage on
the circuit in near real time, flattening the load (sans the one-minute volatility shown). Third, we can “stair step” loads in the morning (pre-cooling a couple degrees, heating water heaters a bit) to actually match the preferred operating sweet spots for our plants as they come on line.

So now, we can begin to talk about plant following, instead of plants following load. We begin to talk about “supply response,” not just demand response. We now get to turn the tables on the normal, unenviable position in which utilities are placed, that of an industry without an inventory or storage warehouse buffer. We have created it from the thermal inertia in a few of the end-uses already in place. The best part is that in all our pilot projects, no customer has noticed any discomfort or had an issue with this process. The days of 5 hour DR and customer suffering are likely nearing an end.

Fourth, we can also incorporate wind following and cloud following, by choreographing the end use in conjunction with the wind or cloud forecasts and the ISO 5 min LMP price (or now, in addition, the DMP price). For details on the financial results of wind following and wind firming, see our WindStore software overviews. For details on how this process makes physical storage batteries more cost effective, see our GridStore results. In both cases, we are simply applying the IDROP optimization engine to batteries or (in the case of wind or water heaters) thermal ceramic bricks and 2 and 3 element water heaters. And with increasing prevalence of “duck curves” (i.e., excess solar generation produced during sunny, summer afternoons), ice-making or afternoon ice-storage is likely to emerge to take advantage of low afternoon-LMPs. And EV charging incentives, ironically, may appear for summer afternoon hours, albeit targeted locationally at employment centers instead of residential homes. The DMP signal will serve to motivate these innovations, and in the right places. The principle is the same, and the same optimizations used here are the same ones used in the derivation of the DMPs (think shadow prices).

The final consideration within the DMP calculation lies with voltage and power factor. These impacts are naturally valued within the optimization results, based on KW and KVAR forecasts. And as we discussed, a 5-minute DMP seems a reasonable latency to use, given the use of LMP at 5 minutes. But note that DMPs can be computed for almost any timeframe. Necessarily, there will be unavoidable operating consequences on the system within the 5-minute period, just as they exist today. And the DSO will likely manage the circuits similarly to what is done today, to ensure reliability. Our projected load and Kvar for the period is based on factors computed a priori for each location, from our KW/KVAR forecasts and the associated power flow results (which may or may not require 5-min updates). Lower latency DMPs are limited simply by processing time and telecommunication lags. In the interim, the utility can do nothing but operate the system as well as possible, as they do today. Even so, we expect that lower latency benefits will accrue to the circuits via sub-minute DMP signal following, as is possible today with 4 sec frequency following, or AGC-type signals.

With respect to voltages, we currently constrain our DMP calculations and optimizations to not permit Low Voltage. However, it is possible to relax this constraint and assign a cost based on 1) a predicted actual customer cost for the low-voltage performance on end use equipment, by customer class (a linearized marginal cost to compensate for reduced performance of lighting, cooling, etc.) and/or 2) an addition of an expected cost that is added as a voltage violation penalty. At some point, low voltage becomes an operating security issue for the utility, dealt with by the utility’s DMS system or automated protection equipment as a means of averting a system collapse/local blackout of the feeder (e.g., overcurrent relays would step in to open circuits at some point). The utility and rule-makers (or the DSO)
locally may decide it is best to avoid such load levels by simply issuing increasingly higher price signals as the point of collapse is approached, as a means of discouraging reaching it. The quantification of this price adder derives from a) value of customer’s lost load and b) system security. As such, this portion of the DMP price necessarily requires local DSO specification. And for our purposes here, we simply constrain our DMP optimizations to not allow for low voltage.

The voltage violation portion of the DMP signal carries two costs that are computed in what is basically a distribution version of existing methods that are currently used to compute marginal costs at the transmission level. First, consider a poor power factor. This is analogous to the cost to the utility of the VAR penalty that must be paid at the transmission level, and can be determined from transmission postings for this time period. To this, we can add the cost of any consequences of low-VAR operating status on the distribution equipment between this location and the transmission bus, computed from load flows. Second, losses are reflected as the cost of the power that must be purchased to replace the losses plus the cost of moving it through the distribution system, computed from load flow analysis, or less accurately estimated from wire type, impedance and distance. In addition, users may choose to compute an additional cost, in some cases, for the Loss of Life at high equipment loadings that we presented previously. A portion of the DMP can be ascribed to the expected loss of life in equipment operating under extreme loadings.

And so, we have reviewed several “value bucket” components of the DMP to this point. All of this is technically and practically feasible using KW and KVAR forecasts and optimizations (IDROP), as well as valuations for system LMPs (LoadSEER coupled with a network power flow), grid asset deferral (LoadSEER), commodity-based cost to serve (DSMore), grid asset protection (TLM), and the utilities’ distribution power flow tools (CYME, SynerGEE, Milsoft, OpenDSS, Nexant, etc.). The DMPs can be derived without the direct use of these distribution power flow tools, but with considerably less accuracy.

The true strength of this approach is that it simultaneously and jointly values both supply-side (KW) avoided costs and grid-side costs (KVAR, voltage, power factor) at the same time. It forces collaboration within the utility across silos. It reveals a single price signal per house, per customer, for third parties to see with certainty, which will spark considerable innovation, and do so at exactly the right PLACE, the right TIME, and right AMOUNT. We are not only including the short term, more operational value and benefits, but we also incorporate the longer term benefits (capacity deferral for T&D, commodity cost to serve, future LMPs) such that innovations with higher fixed costs can participate. This stands in contrast to the current status quo which tends to favor either the grid or the supply, to the exclusion of the other.

The following “heat maps” depict what typical DMPs per hour would be for a peak day for a circuit. These prices are simply the shadow prices that are naturally generated from the IDROP forecasting and optimization modeling process, analogous to the larger scale ISO creation of LMPs. We calculate these DMPs using avoided cost results we have gleaned from 4 years of IDROP pilots, DSMore analysis across 30+ States, and recent LoadSEER work. Intentionally, we are not citing a specific utility here. However, much of the data is adapted from Western loads and prices. So, results are simulated, not from one utility.

Note that IDROP’s DMP price signal is being added to the bus level LMP to reveal the total value or price per location/customer. As LMPs change, and as loads vary, the DMPs change, as shown on the heat maps shown below. Note that the substation in this case lies to the East, and hence the majority of the eastern customers exhibit lower DMPs, due to stronger voltage and to a lesser extent, fewer losses. In this
example, we intentionally exclude longer term DMP value adders in an effort to simplify the interpretation of the results (e.g., capacity/grid deferral, 10 year LMP or commodity cost to serve). When we do include the longer term DMP value adders, the mapped output is much more “spotted” with red-colored higher DMPs interspersed among blue, low cost areas, as one would expect. But it is more difficult to view the trade-offs between KW and KVAR value. So, here we are more focused on just the short term DMP value components, for ease of interpretation and understanding. The Western portion needs much more attention and resources, and hence there are clear pockets of high DMPs. But there are still some pockets of higher DMPs nearer the substation. Red colors depict high DMPs, and blue zones are low DMPs (using $/MWH). Voltage and KVAR essentially are converted into KWH equivalents for $ valuation/reporting. We tend to prefer this approach, for simplicity. But we can parse out the relative contributions to DMP which are KW centric vs. KVAR related, but note that both are usually impacted from many of the micro grid resources. In the Base Case below, we do see some pockets of higher DMP closer to the substation, but the majority lie to the East. Remember that this DMP specification does not include any of the longer term value buckets, so the preponderance of higher DMPs to the West is not surprising. Adding in longer term value buckets of avoided costs would create a more patched mosaic of DMPs, particularly within the Eastern region.

Now, let’s add in DR for the top 100 hours, for 25% load reductions across 20% of the customer base. This is simulated, as we don’t have 20% participation in DR in this region, but it is instructive. One can see that the DMPs are lowered, as DR helps lower the DMP value for subsequent innovation, as we expect, and want.
The map below shows this, and eventually there will be diminishing returns to additional DR to the point where only a few pockets of focused DR attention is worth the marketing effort, or incentives. The DMPs in the West, particularly, are reduced, but don’t quite reach the lower DMP levels of their neighbors.

Next, let’s allow for the installation of equipment that enables locational KVAR injections. Here, we assume that third parties see the transparent DMPs and respond optimally to the location of these assets, just to see what is technically possible. We see that the remaining pockets of higher DMPs on the Western front are lowered, as we would want, and much progress has been made on the Eastern side to lower DMPs as well. At this point, with the optimal allocation of both DR and KVAR, we have driven the efficiency of this circuit toward a much more optimal state.
Finally, let’s review what happens to the voltage, power factor, losses and KVAR from the more efficient circuit state. Below, we see the average hourly results for each. Overall power factor is significantly improved throughout the day, with modest KVAR injection (albeit optimally allocated). In this case, line losses were estimated, and likely are artificially high due to the estimation of service-drop distances. And the majority of the line loss mitigation likely accrues to the secondary line loss where the distance estimation was required. Nevertheless, the key point here is that DMPs can drive innovation at the right location, and circuit performance can be improved at the same time that KW reducing programs are pursued. Importantly, we are valuing these on the same playing field and jointly.

In conclusion, the strength of this approach is that it simultaneously and jointly values both supply-side (KW) avoided costs and grid-side costs (KVAR, voltage, power factor) at the same time. It forces collaboration within the utility across silos. It reveals a single price signal per house, per customer, for third parties to see with certainty. And we include both short-term and long-term sources of avoided costs. Yes, there are several important regulatory and policy-based decision in practice, but the technical feasibility exists and these decisions can be made State by State. The implications of DMP implementation are not totally known, yet we can assume that DMPs will spark significant innovation and that they will transform the electric utility industry in important ways.
The Critics

As is true with any new paradigm, criticism is inevitable. Pioneers take the arrows, and settlers get the land. But criticism and debate is necessary, and healthy, as it leads to resolution, understanding, and growth. Toward that end, here are some of the barbs that expect will arise, and which serve to spark an increased understanding of what DMPS and are not.

Critic 1: This is nothing new. Europe has had DNOs for a decade (Distribution Network Org.).

Yes, but their focus has been limited to largely toward settlements and de-regulation rather than a broader view on distribution level avoided costs and the gains from integration. What is new here with DMPs is the direct specification of distributed avoided costs and pricing. This enables more efficient DG integration, more accurate individual customer settlement shapes, calculation of hosting fees for solar (higher KW charge or allocation), and optimal choreography of distribution level resources, among others, which are beyond the current role of the DNOs.

Critic 2: More focus should be placed on transmission level assets. The amount of energy traversing the circuit is 1% of the total. So, an independent operator for a DSO is too expensive, per MWH.

All energy use travels across both transmission and distribution, so its 100% on both, in reality (exception, large transmission service customers). And given the advance of PV and the need to coordinate end uses via virtual power plant opportunities on the circuits, a more granular focus below the substation does improve the performance of the transmission network system. Reductions in load growth provide reliability benefits as well as opportunities for capital deferral. So, there are joint benefits to both sides. Further, the implementation of DMPs does not need to be an independent function. Utilities can manage it, directly, and simply reveal a DMP signal for use by third parties. They do this now, on average, in their specification of avoided costs for energy, capacity and avoided T&D in DSM filings and EE/DR cost effectiveness analyses. Here, the DMP framework could simply be used to assign these same average costs at a more granular locational level. Little additional work would be required of utilities, but third party investors would see more accurate avoided cost values, even on a forward basis. At a high level, this is exactly what is done today, albeit averaged. Regulators provide EE/DR earnings to utilities today based on forward average estimates of avoided costs, for the projected life of the measure. All DMPs do is to calculate these per location, instead of on average. But there are both short term (spot) and long term (capacity) valuations, just like what exist today in the current avoided cost framework. The main point here is that one cannot easily argue that DMPs are too complicated or too costly to implement, or that a DSO must be independent. DMPs can simply be an extension of the current avoided cost methods, simply done at a more granular level without a significant increase in costs. Whether regulators or utilities desire to establish daily or forward trading markets on DMPs, like LMPs, is a separate question.

Critic 3: A DSO infrastructure would be much too complicated to implement.

Perhaps, but this argument does not prevent the implementation of DMPs as a more refined basis for avoided cost measurement, as described above. And this type of argument is a bit empty in its support. ISOs are complicated, yet we continue to improve their operation and control year by year. A car engine is complicated, but I still drive one every day. What matters is not the level of complication, but rather
reliability and validity. Ironically, the implementation of DMP-centric price signals will actually improve the ability of the ISOs to control plant ramp rates, scheduling, operations and other issues which are part of the ISO complexity. DMP price signals enable more optimal choreographing of pre-heating, pre-cooling, arbitraging water pumping, and other end uses on the demand side, to match desired plant operations (we call it plant following, wind following, cloud following). Because we now have control over some demands within the Smart Grid (via 2 way signaling), we can create virtual storage within the system which serves as the desired buffer between supply and demand, and this reduces the complexity in plant operations. We can ideally adjust loads to bring on plants, and hold plants, at their sweet spots of operating conditions, even in near real time.

**Critic 4: It is better to have a centralized, hierarchical control approach.**

This is perhaps part of the basis for Critic #3 saying, “It’s too complicated.” No doubt, many people that will engage in the discussions will want to maintain some type of centralized control. It is a natural part of how utilities, regulators and ISO operators think. They must. Reliability is priority one. But in doing so, they also limit their views to top down perspectives, instead of bottoms up. Neither perspective alone is sufficient, but a blended approach is what we argue is needed. First and foremost, it is not feasible, not to mention not practical, to attempt optimizations of electric systems across all end uses. It simply cannot be done today, with existing computing power and algorithms. As such, some level of decentralization is necessary. Besides, why would you want to replace the network-centric operations which already exist and work to optimally coordinate plants with substations. We suggest that the distribution DMPs be performed locally, for each substation, separately. We have successfully simulated optimal coordination of up to 100,000 customers in near real time. So, a substation focus for DMP price specification is tractable and achievable for at least at a 5 minute level, if not less than 1 minute. Further, our DMP framework rests squarely on the bus level LMP within the substation, and it captures all the relevant information (short term) within the network system. Our LoadSEER methodology is then applied regionally to calculate the long run LMP forward prices. Then, we can estimate both spot and forward LMP information within the DMP in a very similar manner to that proposed by Amory Lovins (Rocky Mountain Institute) and others. But importantly, we don’t need to optimize the whole electric system, or replace existing systems. We stand on the shoulders of the ISO giants, leveraging their sophistication, and actually helping them out a bit with the provision of forward 10 year LMPs (vs. their current state ability to only forecast next day LMPs). Are these forward 10 year LMPs perfect? Of course not. But, their accuracy increases with increasing forecast accuracy of loads, plants and transmission lines, to be sure. And surely this accuracy will improve over time. But this process is no more, or less, problematic than the current resource and load forecast uncertainties embedded within existing State IRP processes, performed for average loads and average supply needs. In fact, we have been quite careful in our LoadSEER and DSMore calculations and methods to precisely remain consistent with these existing IRP methods, the system wide corporate forecasts for the city, and planned DG and larger resource additions. So, there is no loss in resolution from our approach. Only more informed locational value and opportunity.

**Critic 5: The DSO must be the interface to the ISO for all retail customers, prosumers (solar) and DR. The ISO will then dispatch the DSO virtual power plants in its day-head and real-time dispatch based.**
This suggestion continues with the theme that DSOs must adopt centralized type operations. We have described above why this is not necessary, at least at the start. But it does highlight the fact that there needs to be some coordination between the ISO and the application of DMPs. We suggest at a minimum that this coordination begins to occur, and arguably sufficiently so, simply via the use of the LMP as the base for the DMP. Even with this limited approach to the coordination, third party actors will respond to near real time DMPs (with some type of regulator earnings incentive permitted based on DMPs vs. current averaged avoided costs), and the existing ISO and utility DMS systems will simply respond, as they do today, to the system it observes. Ideally, if the DMPs are accurate, and the third party investors or DR operators are efficient, the ISOs and the DMS systems will observe a more efficient flow of power. They don’t need to be tied together, operationally, to achieve this. The DMP forecasts shown within this document are performed independent of DMS systems. We simply used static power flow models (e.g., CYME, SynerGEE, Milsoft, Nexant, OpenDSS). Sure, there is some loss of accuracy in using static specifications of the radial power flows, but we have also side-stepped 2 to 3 years of IT integration work. We do this because it is fast, and cheap, and are willing to forego some accuracy in the short term to prove out the methodologies. Of course, short term operations and control at the DMS system will be greatly improved from direct information feeds from the DSM system, but this information flow could be a one way flow from DMS to DMP price updates. The reverse integration may not necessarily be worth the cost. Alternatively, our current calculations of DMPs simply use more “static” power flow analysis results from CYME, SynerGEE and other existing connectivity models in the calculation of KW and KVAR impacts on voltage, power factor and losses. Here, we forecast near real time estimates of the most appropriate DMPs, given forecasted KW and KVAR. So, direct DMS integration is not necessary at this level, and provides a reasonable DMP forecast on which to begin pilot applications. Again, this is simply a set of avoided cost forecasts, exactly analogous to what is applied today for EE/DR earnings mechanisms. It’s simply made marginal, and locational now, instead of averaged. But there is no operational IT system integration cost, or implementation delay, that is required. We are all aware of the slow progress of Smart Grid innovation, caused in part by some IT-centric (centralized) planning focus and requirements. Full integration of utility IT systems takes years, and tends to slow innovation and learning. So, perhaps conducting DMP analytics outside of these requirements, in the short term, is preferred. Regulators could couch this approach as simply an improvement in their current avoided cost methods, monitor progress and adapt over time and third party actions are observed. Adjustments can be made fairly quickly, learnings gleaned, and joint understanding achieved, prior to any IT system integration or discussions of ISO/DSO independence, control or centralization. At that point, those discussions would certainly be much more informed by the observed actions of the players. And certainly this tact poses fewer reliability risks.

Critic 6: *The DSO, and perhaps the utility, is now a wires only entity.*

Although this is possible, it may not be optimal. As we have seen in the DMP creation, both KW and KVAR are required to optimize the loads and costs below the bus. DG is a local resource, as is CHP and DR and others. And utilities provide KVAR, not KW, in reality. Almost every resource and end use impacts both. As such, utilities may argue that they deserve rate based returns (or other earnings mechanisms) on their own investments in these types of resources, as they may be in the best position to actually see it, or to understand the impacts of DMPs. This utility investment may not only enhance the rate of innovation, overall, but importantly limit the potential for DMP gaming by unscrupulous third party providers that gain a high market penetration, relative to other providers. Analogous to Enron-like
supply side gaming, a third party vendor with sufficient demand under its control can manipulate both DMP and LMP prices. They might pre-cool, pre-heat in the early afternoon, driving up prices, then collect DR incentives in the afternoon. Unless the utility has sufficient and equal resources under its own control, the mitigation of gaming may be difficult. The smarter third parties will likely even optimize their gaming such that they are not obvious in their tactics. So, some type of utility involvement in the supply side resources below the substation does make sense, and perhaps leads the overall best combination of regulated and non-regulated resources.

Critic 7: There is a Transactive Energy Group (Cazalet et al) which suggests 5 points, a TE Plan, which is a good review, and consistent with DMPs, and we recommend readers review their discussions. Our only departures from their thinking lie with their “top down” focus, and some concerns of complexity.

First, they suggest that regulators should reform tariffs based on forward subscriptions and spot transactions (see http://lnkd.in/wZ_pHm). This will provide stable revenues to retailers, distribution owners and generator owners and greater flexibility and efficiency though more responsive end use devices, distributed generation, and distributed storage. Rocky Mountain Institute holds this same view, and so do we, with the DMP methodology using both short term and long term avoided costs. The methodology we use within our spatial forecasting and our econometric circuit modeling (down to the acre and customer level) are precisely the type of methods required to identify future pockets of load growth, or decrease, such that a PowerWorld or other transmission level power flow can identify future areas of congestion, or higher substation LMPs. Without localized forecasting, this is not possible. We use standard production cost modeling to forecast energy costs, and these combined methods enable exactly the type of forward price calculations required to enable this open market transaction. Today, we only have next day LMPs, and no insight into forward congestion, for the Transport tender described by these proponents.

Second, FERC and the State regulators should raise wholesale price caps and lower price floors (negative LMPs), and post 5-minute locational tenders (see Transactive Energy Roadmap at http://bit.ly/X1lw6x). We agree here as well, and this squares well with a DMP framework. Further, we have shown that this can be done without direct system integrations or top down ISO type control. This would limit the complexity which they cite as potentially problematic, in the short term, and perhaps long term. However, unrestricted floors and caps does pave the way for potential gaming by non-regulated parties of the sort we observed under non-ISO, bi-lateral energy markets during the Yr 2000 timeframe in California. This time, though, it will come from third parties that manipulate demands, instead of supply, by pre-cooling homes, heating water, and other actions which drive loads and prices higher in the morning, then these same parties are likely to collect DR incentives in the afternoon. This “double dipping” or market gaming requires some minimum level of MW of demand control. But note also the recent focus of Google, and others, at the “grid’s edge” where supply and delivery is more vulnerable and easier to potentially game (speculation, but theoretically the right place for gaming to test the waters).

Third, regulators should require that prices be fully locational reflecting transmission and distribution operating criteria, congestion and marginal losses. Locational fairness can be maintained by side payments. This is almost a direct call for DMPs, in our opinion, though the Transactive Energy proponents believe that only the Transport portion of the tender be calculated by utilities/regulators. The
energy portion, they feel, should be left to unregulated bi-lateral markets. In either case, DMPs can be used as a relatively low risk transition toward a bi-lateral market structure, to mitigate gaming risk in the short term.

Fourth, regulators should resist approving utility generation procurements to minimize potential stranded investments and high rates until this full plan is in operation. The current surplus, new customer distributed generation and storage investments, and flexibility from the response to fully dynamic transactive tariffs should be given the opportunity to balance the investment and operation of the grid without the introduction of a centralized capacity market. Yes, this makes sense, too, and again suggests the use of DMPs in our proposed forecasted form versus fully operational form. And it seems a bit risky to release the State’s IRP process, the focus of which is just as much on future supply reliability as it is on lowest cost to serve. So, full scale reliance of future supply on free markets, at least in the short term, may incur reliability risks that cause undesirable economic or reliability consequences.

Fifth, the regulators should accelerate competitive access for all customers, limit the concentration of generator ownership or control, and aggressively monitor participant wholesale and retail tenders, transactions, positions, and committed capital and enforce such restrictions as may be necessary. Here, we agree that regulators should monitor the transition to more open competitive access, and even provide new earnings mechanisms to utilities to participate. Our view is that both utilities and non-regulated third parties should be motivated to both participate in the development of resources below the substation. If for no other reason, it will limit the potential gaming of the DMP prices. Further, the use of DMPs provides the regulator with the long term desired equilibrium traded tenders. If the bi-lateral market prices and supply stray considerably from the optimized DMP prices, we can infer that either irrational markets are in play (which is fine, short term) or that some type of gaming is afoot (not fine).

So, we really only differ in our opinions with respect to 1) potential complexity is not a reason to not try DMPs, and 2) full scale competition below the bus may cause unforeseen consequences to reliability. Utilities and third party innovators should be given equal opportunity. Full scale competition may not give utilities enough time to respond. Gaming may arise from Google, or others. And these market players are not responsible for reliability. So, we caution regulators to perhaps slow the march toward full scale competition below the bus, using DMPs as a transition, either within a DSO/ISO managed context or as a guide for regulatory caps or pilot demonstration.

**Critic 8: Implicit within the arguments of Transative Energy proponents is the assumption that overall system efficiency is maximized by having individual buyers and sellers optimize their own needs. This implies that an ISO or DSO overseer is not required, as well, except for the Transport delivery fees.**

We have performed several analyses regarding what is optimal for the system overall, vs. what is the value if all actors act independently. We have found that there are 2X to 3X more cost savings that can be achieved if market actors (buyers) act in concert with one another vs. separately. This is not really surprising, when you think about it. All customers are tied to the same grid, and increased or decreased load in one pocket of the grid causes unforeseen changes elsewhere on the grid. So, if a house is merely optimizing its own needs, it may well install numerous solar panels, sell back to the grid, causing reverse flows that are not anticipated or for which the grid was not designed. Conversely, low local prices may spark more EV charging stations, with subsequent voltage drops and service problems for neighbors. One
example analysis, shown below, quantifies the cost saving differences between applying DMPs for each home, conditioned on the needs of the whole circuit, versus having each home optimize its own needs. Here, we see that centralized dispatch (DMP methods) lead to overall lower cost to serve results than the decentralized method (each home limits their own demand, given their tariff rate). The difference is about 3X, or $76 vs. $26, per home, on average.

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Centralized</th>
<th>Decentralized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer Bill</td>
<td>$209</td>
<td>$200</td>
<td>$175</td>
</tr>
<tr>
<td>Total Cost to Serve</td>
<td>$285</td>
<td>$255</td>
<td>$194</td>
</tr>
<tr>
<td>Net Revenue</td>
<td>($77)</td>
<td>($55)</td>
<td>($19)</td>
</tr>
<tr>
<td>Utility Cost Savings</td>
<td>$22</td>
<td>$57</td>
<td>$76</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This implies that the system loses $50 per month (peak month) in potential cost savings, if taking the decentralized (non-DMP) approach. This equates to millions of dollars of lost cost savings.

**Critic 9:** T&D circuits are too dynamic to estimate avoided costs. We perform switching and transfers to accommodate capacity issues, and this is essentially free. Or the more entrenched version of this stance holds that “there are no avoided costs on the circuit.” It is what it is.

This argument generally is held among traditionalist distribution planners and operators. Yes, it is true that switching and transfers are the first, best step in mitigating capacity issues at the circuit and bank level, and yes this is very low cost. But eventually, the loads increase to a point beyond which this type of switching solution is no longer sufficient. Further, these switching activities are generally performed prior to the peak season, in anticipation of higher peak season loads, however radial network flows typically remain static through this peak season. Sure, sometimes adjustments are made within the season, too, but these changes can be made known and included. We account for these types of switching transfers within our LoadSEER software platform so that distribution planners can better forecast the hourly net load that is being transferred and account for this in the circuit forecast overall. Even if changes are made within the peak season, whatever DMP price signals are forecasted for the season will still send reasonable avoided cost estimates to the locations that will eventually need the added support. This is true even if we use a single, static CYME-type connectivity model for the whole season. Yes, it is not exactly accurate in the event of added switching, but more often than not the distribution operators re-switch the circuit back to its original configuration at season’s end. If we were to attempt to chase the increase in accuracy too much, it is possible that we might be sending artificially low DMP price signals to those locations that need it the most, in the long run. These types of nuances can only be identified, and addressed, empirically. And resolution will depend on the extent to which policy makers desire a focus on short term accuracy over long term investment motivations. In all cases, any attempt is necessarily no worse than our current regulatory approach of using average avoided costs, both across the whole system and even at the substation level.
Regulatory Policy

We have remained fairly agnostic with respect to regulatory policy, so far. We have focused primarily on methods for measuring avoided costs, to this point. This is intentional. Each State is likely to evolve in unique ways as they move toward distributed platforms for encouraging grid resiliency, EE/DR/PV promotion or proposed earnings mechanisms for utilities. Further, States are likely to transition carefully from their current averaged avoided cost methods to the more granular and marginal avoided cost methods described here. It is unlikely that any one State would embark on a full scale implementation of distributed DMPs immediately. Pilots will test the feasibility and third parties will offer their own nuances to DMP-type implementation. However, we do believe that any regulatory policy should address some key aspects.

First, utilities will need earnings incentives to move toward more granular avoided cost platforms. And whatever earnings mechanisms are put in place by the regulatory policy should address the natural disincentives embedded within current tariff rate structures (which are average costs). We have shown that it is feasible to provide a unique cost to serve price to each customer using accurate marginal avoided cost methods (see DSMore discussion and results). We don’t recommend applying these cost-to-serve prices to customers, immediately. The price shock for some customers would be difficult to manage. Rather, we argue that regulators use the DMP avoided cost methods to provide an incentive to higher cost customers to become more efficient. This is no different than what occurs today with traditional EE and DR incentives. We simply are using locational information to provide higher incentives in some places. The customer-specific cost-to-serve measurements can be used in settlement shapes, however, or grouped into similar load factor blocks to achieve more efficient grid and supply outcomes than is currently observed.

Second, we recommend that regulators begin their focus on the Long Term avoided costs first, leaving implementation of the Short Term factors for later (refer to the 2 by 2 table of “avoided cost value buckets” described early on in this paper. Short Term factors are necessarily more operational and more difficult to implement, also carrying more reliability risk if not implemented properly. Pilot test the Short Term factors carefully, to work out operational issues prior to wider application. Most of the utilities we work with are currently focused on the Long Term factors and are able to see how to implement these factors simply as refinements of their current averaging approaches.

Third, don’t think of DMPs as tradable market prices, at first. Consider them simply as more granular calculations of avoided costs. In this vein, reasonable estimates can be made of the avoided costs per house, annually, without any impact to operational reliability. All you are doing here is refining your cost-effectiveness modeling of the benefits and costs, just as is done today, but at a more granular level. In some States, DMPs are likely to never reach the level of a market-traded price signal, and this is just fine. But even here, we will see increased efficiency in the grid and supply as the correct avoided costs are being transparently revealed to both utilities and third parties. This directs innovation and investment toward those areas where it provides the greatest returns. And even if the avoided costs are calculated once a year, and are not dynamic, the investments will eventually reveal increased grid efficiency. In fact, some States may find that the added costs and complexities of creating a fully functioning ISO or DSO-type platform based on real time DMPs is not worth the effort. We think it will be, given the importance and value of wind/cloud following, ancillary services (frequency, in particular), voltage
support, power factor benefits, and optimally locating storage, PV and DR. But this level of sophistication does not need to occur immediately, nor does any perceived complexity in doing so necessarily block progress toward this end state (simply start with the Long Term, non-operational avoided costs).

Fourth, we recommend that regulators wait, or be slow to adopt, policies that create independent DSOs immediately. Again, operational reliability is perhaps the most important function provided by utilities today, and jeopardizing existing reliability brings potentially disastrous consequences. Utilities should continue to operate the grid. Use DMPs as avoided cost pricing incentives to motivate more intelligent investments. Whatever improvements that occur on the grid will be obvious to the distribution operators, and they will respond appropriately, as they do today. Removing grid operations from the utilities to be managed by an independent entity is a risky tactic, at least at first.

Utilities also need an appropriate earnings mechanism to participate. We have shown how current policies risk continued inefficiency (via averaged tariffs). We assert, with reasonable confidence, that Google and others have the potential to game the grid with Enron-style demand manipulation (artificially increasing prices via pre-cooling, pre-heating, pumping). Unless regulators are willing, and able, to limit this potential market power, the best tact is to allow utilities to participate in distribution-side investments just as non-regulated third parties do. This dramatically limits the potential for gaming, and provides a check on non-regulated activities. There are several options, but we argue that a shared savings mechanism is preferred.

First, traditional rate basing approaches face the familiar risk of potential over-investment by utilities. One could cap the total amount of investment, but this approach still does not insure that investments are made in the right locations. We need the DMP-type avoided costs values to insure locational appropriateness. And if we have to measure the marginal costs anyway, regulators have a ready-made source for shared savings measurements. Many States offer this today, albeit averaged and non-locational. DMP calculations provide an accurate source for demonstrated shared savings (and say 10% to 15% are offered to utilities as their earnings incentive).

Second, decoupling could be used. The problem with decoupling is that it tends to apply policies to the whole system, and may not focus attention where it is most needed, locationally. Unintended consequences may arise from decoupling where utilities focus on cost saving efforts that have nothing to do with improving grid efficiency. Moreover, it does not force the application of more granular avoided costs, which is the only way to insure least cost planning based on actual cost-to-serve factors (the main goal of regulatory policy).

Third, the earnings mechanism obviously needs to be equal to, or greater than, the alternative returns from central plant and grid investment (currently rate-based). This comparison, which is a simple empirical analysis, will guide the specification of whether the shared savings mechanism should be 10%, 15%, or whatever level. Regulators might want to start at 25%, and gradually decrease the percentage as grid efficiency is achieved, perhaps. But only a shared savings type approach insures 1) that more granular avoided costs get estimated in the first place, and 2) that utilities focus their investments in the right locations, without getting sidetracked by decoupling nuances or overinvestment from rate basing.
Finally, regulators should consider extending the application of the earnings mechanisms to all resources that improve grid efficiency. This extends beyond PV installations and storage. It should include earnings returns for DR, voltage support, KVAR injection, HVAC and WH leasing, and other resources which carry operational or service-type components. Since utilities will continue to operate the grid, at least in the short run, they are well positioned to locate, operate and integrate these “softer” resources into the grid. Of course, regulatory oversight will be required to separate grid operations from distributed resource investment arms of the utilities. But this is no different than what we see today for bulk and retail supply, where utilities maintain regulated services and non-regulated arms.

We close out this section with an invitation. We encourage open and rigorous dialog and critiques of the DMP approach. Our view holds that focusing discussions on the analytics, on what is testable, and on what can be feasibly implemented all are more productive efforts than is a general discussion of how DSOs, DSPPs, or DNOs should be structured, organized and controlled. It is natural to want to go directly to a process-centric solution, but we believe that too much of this process focus might be “getting the cart before the horse.” Process should follow form, content, and substance, and not vice versa. The mandate from the States is reliable power at the least cost. Hence, we designed DMPs to focus squarely on avoided cost measurements, as required within a world of more granular and distributed resources. Even if one balks at the concept of a DMP, this paper lays out a comprehensive approach to the measurement of granular avoided costs. And this is the key need to motivate a more efficient grid.

We welcome questions, critiques and input. Contact information is provided below.

Tom Osterhus, PhD (CEO, Integral Analytics). Tom.Osterhus@IntegralAnalytics.com

Michael Ozog, PhD (VP, Integral Analytics). Michael.Ozog@IntegralAnalytics.com

KEYWORDS: Distributed Marginal Price, DMP, Transactive Price Signal, transactive energy, plant following, cloud following, supply response, virtual power plant, wind firming, dynamic dispatching, least cost planning, grid optimization, solar integration, integrating renewables, future of electric utility, LMP, IRP, distributed automation, smart grid, electric storage, spatial forecasting, mark-to-market valuation, DSMore, LoadSEER, IDROP, Integral Analytics.