

ECONOMIC PRINCIPLES OF DEMAND RESPONSE IN ELECTRICITY

PREPARED BY
Larry E. Ruff, PhD

PREPARED FOR
Edison Electric Institute

OCTOBER 2002

LARRY E. RUFF

Dr. Larry E. Ruff is an internationally recognized expert on the restructuring of electric and gas utilities to create competition and on the operations of the resulting competitive markets. He has advised governments and utilities in the United States, Canada, Europe, the former Soviet Union, the Indian subcontinent, the Asia-Pacific region, and Latin America on competitive restructuring. He has been an expert witness in numerous proceedings at the Federal Energy Regulatory Commission (FERC), state regulatory commissions, and foreign jurisdictions. He writes and lectures extensively on competition in electricity and gas markets.

Dr. Ruff has been awarded a PhD in economics by Stanford University (1968) and a BS in physics (with honors) by the California Institute of Technology (1963). He has held senior positions in government, industry and non-profit organizations, and has been a Senior Vice President with National Economic Research Associates (NERA) and a Managing Director of Putnam, Hayes & Bartlett Inc. (PHB). He is currently a Senior Advisor with Charles River Associates.

Edison Electric Institute (EEI) is the association of United States shareholder-owned electric companies, international affiliates and industry associates worldwide. In 2001, our U.S. members served more than 90 percent of the ultimate customers in the shareholder-owned segment of the industry, and nearly 70 percent of all electric utility ultimate customers in the nation. They generated almost 70 percent of the electricity generated by U.S. electric utilities.

Organized in 1933, EEI works closely with its members, representing their interests and advocating equitable policies in legislative and regulatory arenas. In its leadership role, the Institute provides authoritative analysis and critical industry data to its members, Congress, government agencies, the financial community and other influential audiences. EEI provides forums for member company representatives to discuss issues and strategies to advance the industry and to ensure a competitive position in a changing marketplace.

EEI's mission is to ensure members' success in a new competitive environment by:

- Advocating Public Policy
- Expanding Market Opportunities
- Providing Strategic Business Information

For more information on EEI programs and activities, products and services, or membership, visit our web site at www.eei.org.

ECONOMIC PRINCIPLES OF DEMAND RESPONSE IN ELECTRICITY

Prepared by
Larry E. Ruff, PhD
8017 Oak Way
Windsor, CA 95492
Voice: 707-836-0437
Fax: 206-202-4647
leruff@attglobal.net

Prepared for
Edison Electric Institute
701 Pennsylvania Avenue, N.W.
Washington, D.C. 20004-2696
202-508-5000
www.eei.org

OCTOBER 2002

Edison Electric Institute has developed this publication as a service to the electric utility industry and to the public. Use of this publication is voluntary and should be undertaken after an independent review of the applicable facts and circumstances of the particular projects, although Edison Electric Institute has made all reasonable efforts to present comprehensive and accurate information. NO GUARANTEES OR WARRANTIES INCLUDING ANY EXPRESS OR IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS ARE MADE WITH RESPECT TO THIS PUBLICATION BY EEI, ITS OFFICERS, DIRECTORS, OR EMPLOYEES, OR AGENTS, WHO ALSO ASSUME NO LEGAL RESPONSIBILITY FOR THE ACCURACY OF THE PRESENTATIONS, COMMENTS OR OTHER INFORMATION IN THIS PUBLICATION. IN ADDITION, NO LIABILITY IS ASSUMED AND ALL LIABILITY IS EXPRESSLY DISCLAIMED FOR NEGLIGENCE OR DAMAGES OF ANY KIND, ANY DECISIONS, CONTRACTS, COMMITMENTS, OBLIGATIONS OR ANY OTHER ACTIONS UNDERTAKEN OR MADE ON THE BASIS OF THE INFORMATION CONTAINED IN THIS PUBLICATION.

Copyright © 2002 by Edison Electric Institute (EEI). All rights reserved. Printed in the United States of America. No part of this publication may be reproduced or distributed in any form or by any means, or stored in a data retrieval system, without prior written permission of the publisher, Edison Electric Institute.

Published by:
Edison Electric Institute
701 Pennsylvania Avenue, NW
Washington, D.C. 20004-2696
Phone: 202-508-5000
Web site: www.eei.org

Table of Contents

- 1. INTRODUCTION AND SUMMARY1
 - 1.1 The Context1
 - 1.2 A History of Demand Response in Electricity2
 - 1.3 Summary Conclusions and Recommendations3

- 2. THE BENEFITS OF IMPROVED DEMAND RESPONSE6
 - 2.1 Improving versus Simply Increasing Demand Response6
 - 2.2 Cost Reductions versus Rent Transfers.....7
 - 2.2.1 The Economic Benefit of Increased DR Is Lower
Supply-Side-Plus-Demand-Side Costs7
 - 2.2.2 Small Increases in DR Cannot Cause Large Changes
in Costs and Hence Cannot Have Large Benefits9
 - 2.2.3 Bill Reductions Due to Lower Prices Are Rent Transfers,
Not Cost Reductions or Social Benefits10
 - 2.2.4 Rent Transfers from an Increase in DR Are Transient
and Depend on Surprising Suppliers.....11
 - 2.2.5 Even If DR Reduces Extraordinary Peak-Period Market Power,
the Rent Transfers Will Probably Be Small13
 - 2.2.6 Consumers Can Gain from Implementing Cost-Effective DR,
But Not from Subsidizing Non-Cost-Effective DR.....15
 - 2.2.7 This Is All True Despite the “California Crisis,” Which
Has Little Relevance for DR in ITP Markets15
 - 2.3 Price Volatility and System Reliability Effects16
 - 2.3.1 Price Volatility Can and Should Be Managed by Contracting,
Not by Subsidizing DR17
 - 2.3.2 DR Should Be Paid for Reliability Services in Ancillary
Service Markets, Not by Subsidizing DR in Energy Markets.....17

- 3. THE THEORY OF DEMAND RESPONSE IN ELECTRICITY 18
 - 3.1 Demand and Demand Response in Normal Markets18
 - 3.1.1 The Outcome of a Competitive Market Process Is Efficient,
Even When Demand and Supply “Participate” Very Differently18
 - 3.1.2 A Consumer Can Sell What It Does Not Consume—
But Only If It Owns It19
 - 3.1.3 Paying a Consumer for DR “Resources” It Would Have
Bought But Didn’t Is Paying Twice for the Same Thing20
 - 3.1.4 Fixed Retail Prices Complicate Things But Do Not Change
the Logic That a Consumer Must Own What It Sells21
 - 3.2 Demand and Demand Response in Electricity Markets22

3.2.1	Minimizing Supply-Side Costs to Meet a “Fixed” But Market—Determined Demand Treats DR Fairly and Efficiently	23
3.2.2	Treating DR As an Energy “Resource” That Adds to Supply Often Leads to Confusion, Errors, and Inefficiencies	24
3.2.3	If the Concept of DR As an Energy “Resource” Is Defined and Used Correctly, It Becomes Largely Pointless	27
3.2.4	It Is More Natural and at Least As Efficient to Treat DR As Price-Responsive Demand	28
4.	INCORPORATING DEMAND RESPONSE IN ITP MARKETS	30
4.1	The Division of Responsibilities in ITP-Based Markets.....	30
4.1.1	The ITP’s Job Is to Operate a Wholesale Dispatch/Pricing Process That Fully and Efficiently Incorporates DR	31
4.1.2	It Is the Job of LSEs and ESPs—Not the ITP—to Manage Retail Prices, Price Risks, and Demand Response Services	32
4.1.3	It Is the Job of Regulators—Not the ITP—to Assure That Regulations Provide the Right DR Incentives	33
4.2	An Economically Efficient ITP DR Program	33
4.2.1	The ITP Should Accept and Use Demand Bids in Clearing Its Day-Ahead (and Any Other) Forward Markets	34
4.2.2	The ITP Should Use Demand Bids in Real-Time Dispatch and Pricing—as Demand Reductions, Not “Resources”	35
4.2.3	The ITP Should Not Interpret, Administer, or Override Retail Tariffs or Contracts Among Market Participants	36
4.2.4	If Price Caps Result in “Out-of-Market” Payments, DR Should Be Treated Just Like Supply	37
4.2.5	DR Should Be a Resource in Ancillary Service and ICAP Markets Just As Supply Resources Are	39
4.3	Operating in an Efficient ITP Market.....	40
4.3.1	DR Should Be Based As Far As Practical on Real-Time, Not Day-Ahead, Conditions and Prices	40
4.3.2	Purchases in the Day-Ahead Market Should Reflect Expected Real-Time Consumption Including the Expected Effects of Any DR	41
5.	CONCLUSIONS.....	43
	APPENDIX: A NUMERICAL EXAMPLE	45

1. INTRODUCTION AND SUMMARY

1.1 THE CONTEXT

Improving the ability of electricity demand to respond to wholesale spot prices will reduce the total costs of meeting demand reliably and can reduce the level and volatility of spot prices during critical periods. Recognizing this, the Federal Energy Regulatory Commission has said that its Standard Market Design (SMD) should and will allow “demand resources ... to participate fully in energy, ancillary services and capacity markets,” and the “demand side ... to participate in the real-time market.”¹

Despite the growing interest in short-run demand response (DR), recent analyses and proposals contain remarkably little discussion of the basic economic principles involved. This could be interpreted to mean that these economic principles, which are laid out in a substantial literature going back at least to the days of least-cost planning (LCP) and demand-side management (DSM),² are so well understood and widely accepted that there is no need to restate them—except that fundamental and important economic errors continue to be common where DR is concerned. In particular, the benefits of increased DR are often described and quantified in ways that lead to unrealistic expectations and inappropriate policies; and proposed or actual procedures for integrating DR into spot markets, particularly those based on treating DR as a “resource” that adds to supply rather than as a reduction in demand, are often incorrect in concept and inefficient in practice.

This report reviews the basic economic principles of DR and discusses some of the implications for electricity markets. The focus is on decreasing demand in response to price spikes during critical periods—called “peak” periods here, even if the cause is a drop in supply more than an increase in demand—because that is the most pressing issue for the SMD that FERC is defining for Independent Transmission Providers (ITPs).³ Reducing demand in response to longer-term price “bulges” such as that in the western United States in 2000-2001, or even in response to permanent price increases, involves the same economic principles but different institutional and implementation issues and is not specifically discussed here.

1 FERC “Working Paper on Standardized Transmission Service and Wholesale Electric Market Design,” March 15, 2002 (FERC SMD Working Paper), pp. 13, 6, and 13, respectively.

2 For examples, see: Larry Ruff, “Least-Cost Planning and Demand-Side Management: Six Common Fallacies and One Simple Truth,” *Public Utilities Fortnightly*, April 1988; William Hogan and Charles Cichetti, “Including Unbundled Demand-side Options in Electric Utility Bidding Programs,” *Public Utilities Fortnightly*, June 8, 1989; and Larry Ruff, “Equity vs. Efficiency: Getting DSM Pricing Right,” *The Electricity Journal*, November 1992.

3 The FERC SMD Notice of Proposed Rule Making of July 31, 2002 (the “SMD NOPR”) introduces the concept of an Independent Transmission Provider (ITP) to encompass both Regional Transmission Organizations (RTOs) and other jurisdictional entities that operate under the SMD. The term ITP is used in this paper to refer to any system operator that uses spot markets to manage and price real-time operations.

1.2 A HISTORY OF DEMAND RESPONSE IN ELECTRICITY

Until the 1970s, price was usually ignored in forecasting electricity demand even in the long run. When price-independent demand forecasts were used during the oil crises of the 1970s to justify a rapid expansion of generating capacity, technological critics argued that it would be cheaper to reduce demand than to increase supply, at least to some large extent.⁴ Then some economists pointed out that regulated prices below the costs of incremental supplies give consumers too little incentive to conserve and make it unprofitable to expand supply. Under these conditions, a utility with an obligation to supply can improve consumers' incentives and reduce its own full-cost-recovery prices by paying for demand reductions. As long as such payments do not exceed the difference between marginal costs and retail prices they are "price corrections" rather than subsidies.⁵ The LCP movement was born—but soon forgot the critical qualification about the relationship between prices and incremental costs.

Electricity demand continued growing in the 1970s despite increases in electricity prices, because oil prices were increasing even faster and price-regulated natural gas was in short supply. Utilities rapidly added generating capacity even as the Public Utilities Regulatory Policy Act (PURPA) required them to buy power from non-utility generators (NUGs). Then, in the early 1980s, world oil prices collapsed and recently deregulated natural gas made a comeback as electricity prices continued increasing to cover the costs of the new capacity. Electricity demand growth slowed below expectations, creating excess capacity that led to even more rate increases and, in some cases, cost disallowances that created financial problems for utilities.

In the late 1980s, utility DSM programs based on the LCP concepts from the 1970s finally caught on with state regulators. By now, however, electricity rates were generally above incremental costs, so utilities could pay for DSM activities only by raising rates further; utility payments for DSM were now real subsidies, not price corrections. Nonetheless, in the late 1980s and early 1990s many state regulators approved further rate increases to pay for subsidized DSM programs and the incentives that made them palatable to utilities. To the extent these programs reduced demand, they exacerbated the excess capacity problem and pushed up regulated rates even more.

4 The "Technical Fix Scenario" developed in the Ford Foundation's Energy Policy Project (*A Time to Choose: America's Energy Future* [Ballinger Publishing Co., Cambridge, MA, 1974]) was perhaps the first attempt to compare the costs of supply expansion to the costs of demand reduction on a national basis. Amory Lovins then followed with many works, including *World Energy Strategies: Facts, Issues and Options* (Friends of the Earth, 1975), and "Soft Energy Paths," *Foreign Affairs*, 1976.

5 If the incremental costs (C) of expanding supply are more than the cost-recovery price (P) a utility is allowed to charge, a utility loses money (C-P) if it expands supply at the current price, and hence must increase prices in order to recover its full costs. But if the utility can buy demand reductions at prices less than the amount (C-P) it would lose expanding supply, it can meet growing demand with lower incremental losses and hence smaller increases in full-cost-recovery prices. Roger Sant (*The Least-Cost Energy Strategy* [Carnegie-Mellon University Press, Pittsburgh, 1979]) was the first economist to popularize this idea. But the arithmetic magic works only when $P < C$; when $P > C$, consumers already have too much incentive to conserve, a utility can lower its average cost and hence its full-cost-recovery price by expanding supply, and a utility that pays anything to reduce demand must raise its average price to cover both the payment itself and its lost sales margin (P-C).

In the early and mid-1990s, the increasing divergence between high regulated retail utility rates and low incremental wholesale costs created pressure for competition in electricity, particularly from large consumers wanting to buy at low wholesale prices rather than high retail rates and from traders wanting to arbitrage such differences. This pressure was particularly strong in states that had had the most active and “successful” NUG and DSM programs and now had the largest gap between retail rates and incremental wholesale costs. The development of competition then significantly weakened the utility monopoly power necessary to support subsidized NUG and DSM programs and these began fading away.

The events outlined above demonstrated conclusively that electricity demand is strongly affected by prices and that ignoring this reality will lead to costly mistakes; today, nobody would think of forecasting electricity demand without considering prices as critical explanatory variables—and everybody is more modest about their ability to forecast at all. More importantly for the purposes here, these events also demonstrate the importance of correctly understanding and applying basic economic principles in the design of demand reduction policies and programs. The LCP concepts of the 1970s, the DSM subsidies of the late 1980s and early 1990s, and the DR programs being developed for ITPs today all involve the same economic and commercial principles. If the new DR programs are to avoid the mistakes and excesses of the past, and are to become permanent, stabilizing features of electricity markets rather than transient disruptions, they must be based on a correct understanding and application of economic and market principles.

1.3 SUMMARY CONCLUSIONS AND RECOMMENDATIONS

This report makes the following principal points:

- There is a critical distinction between improving DR by providing better price signals, technology, and information and then letting market participants respond to these, and simply increasing DR by forcing or subsidizing demand reductions that cost more than they save. Improving DR may require some socialized investment in infrastructure, technology, and transition costs, but will both increase DR and—if the socialized investments are well targeted and limited—reduce total costs. The brute-force approach of increasing DR without making it cheaper will increase total costs for society and ultimately for consumers.
- The net economic benefit of decreasing demand is the resulting reduction in supply-side costs less the increase in demand-side costs. This net benefit can be large if the demand reduction is large and costs significantly less than the resulting reduction in supply-side costs. But reducing demand by a few percent in a few peak hours cannot realistically reduce even peak-

period costs more than a few percent and hence cannot have a large net benefit. And if the demand reductions are costly, the net benefit can be negative.

- Much of the current enthusiasm for increasing DR is based on claims that small decreases in demand can cause large decreases in the prices and bills paid by consumers even it does not have a large effect on costs. But any reduction in consumers' bills in excess of the reduction in suppliers' costs is a transfer of economic rents, not a real benefit, and is not even all or even primarily from "greedy" producers to deserving consumers once the effects of contracts, taxes and widespread corporate shareholdings are taken into account.
- Even if reducing prices to transfer rents from suppliers to consumers is regarded as a desirable objective, small increases in DR will not produce large transfers once all anticipatory, consequential and longer-run effects are taken into account. A sudden decrease in peak demand can cause peak prices and bills to fall dramatically in the short run, given the inelasticity of short-run supply during peak periods, but has no effect on the need for or costs of non-peak capacity. If the non-peaker generation market is workably competitive, non-peak prices must soon increase to make up for any loss of peak-period rents to non-peaking capacity. And even inelastic short-run supply curves shift in the long run, so peak prices must eventually recover enough to cover the full costs of meeting incremental peak demand. In the long run, average prices to consumers will be slightly lower because the load factor has improved, and total bills will fall by the actual decrease in supply-side costs plus any decrease in the (probably small) excess profits the few peakers had been collecting—less any "uplift" used to subsidize the new DR. But if the reduction in peak demand is small, any rent transfer from suppliers to consumers will be transient and/or small, particularly if suppliers see it coming.
- Normal markets allow consumers to sell what they do not consume as long as they own it, but no rational market pays consumers for not consuming what they do not own, even if they can prove that they would have bought it but didn't. Paying somebody because they might have bought more but didn't is as illogical, unfair, and inefficient as buying the Brooklyn Bridge from somebody who thought about buying it but decided to sell it instead.
- A retail tariff or full-requirements contract (FRC) is sometimes interpreted as giving consumers the right to sell at market prices what they do not consume relative to some baseline consumption level (BCL). But even if this interpretation is accepted, a consumer cannot sell back something unless it first (or simultaneously) buys and pays for it, which implies that the net payment for selling back some of any BCL should be no more than the market price less the retail tariff/FRC price. In principle, LSEs and others selling at fixed tariff/FRC prices have the incentive and should have the financial responsibility to make such payments, but in

practice defining, allocating, and buying back BCLs is so difficult, contentious, and distorting that it is often impractical. Giving this difficult job to a third party such as the ITP does nothing to make it any easier to do correctly, but does make it more likely that mistakes will be made, hidden, and subsidized.

- The pricing and dispatch processes used by ITPs under the SMD do not inherently discriminate against DR, even if they include a step in which they select supply options to minimize the cost of meeting “fixed” but ultimately market-determined demands. Treating DR as a “resource” equivalent to supply resources in such processes is confusing and tends to produce illogical processes and inefficient outcomes. DR can be treated as a “resource” correctly, but doing so makes it obvious that it is more natural and less confusing simply to treat DR as what it is: a reduction in demand in response to price.
- The ITP’s principal responsibility is to operate the central dispatch/pricing processes needed in a competitive electricity market, including assuring that DR is fully and efficiently incorporated into those processes. It is the job of load serving entities (LSEs), energy service providers (ESPs) and others, not the ITP, to design and administer retail tariffs and contracts, manage risks and provide DR services. It is the job of regulators, primarily at the state level, to assure that retail tariffs and LSE regulation provide the right DR incentives. There are good reasons for such a division of responsibility, and it should be maintained in the design and implementation of DR programs.
- The ITP’s ancillary service (including operating reserve) markets and any capacity/capability (e.g., ICAP) markets should treat demand facilities as resources just like generation resources and should pay both types of resources the same amounts for equivalent services. Paying for ancillary services and ICAP is not the same as paying for not-consuming-energy and hence does not create the same opportunities for logical errors and practical inefficiencies.
- The ITP’s day-ahead market (DAM) and any other forward markets should accept and fully integrate demand bids—as price-responsive demand, not as negative supply “resources.” LSEs and others should submit day-ahead demand bids reflecting expected real-time demand, including the effects of any DR. There is no need or justification for the ITP paying anybody not to buy in the DAM. Any consumer or ESP wanting to be paid a known price for DR at the day-ahead stage can and should sell day-ahead energy contracts reflecting the energy it expects to save by DR in real time.
- The ITP’s real-time dispatch/pricing process should accept and fully integrate demand bids—as price-responsive demand, not as negative supply “resources.” In fact, DR may be more valuable in the RTM than in the DAM for reducing price spikes and improving reliability. In

a well-designed dispatch/pricing process, accepted demand bids will always influence price by displacing higher-cost supply bids, whether it is a supply or a demand bid that “sets” the price. Once the market-clearing dispatch is determined, the ITP simply tells each successful demand bidder to reduce its takes as it said it would and then charges the bidder the market price for whatever it takes; there should be no payment for reducing demand except that resulting from the normal settlement of day-ahead contracts against real-time operations.

2. THE BENEFITS OF IMPROVED DEMAND RESPONSE

Policies and programs that improve DR by providing more efficient incentives for and lowering the costs of DR will produce real, lasting benefits for society and for consumers. But mischaracterizing or overstating the benefits of DR can lead to poor policies and unrealistic expectations that result in inefficient and unsustainable DR programs.

2.1 IMPROVING VERSUS SIMPLY INCREASING DEMAND RESPONSE

There are many ways to increase peak DR without making it cheaper, such as subsidizing demand reductions and recovering the subsidy costs through a tax or “uplift” on consumption. Such a tax/subsidy combination may, if it is expected to continue long enough, stimulate consumers to invest in devices and processes that reduce the short-run costs of responding to price spikes and hence shift the short run demand curve. But such a tax/subsidy combination does nothing to reduce the overall costs of reducing demand, so any reduction in demand beyond what individual consumers would find worthwhile without the tax/subsidy arrangement costs more than it is worth and hence is inefficient and wasteful.

There are also many ways to improve peak DR. For example, the ITP can facilitate demand bidding in its markets; LSEs and state regulators can improve price signals/incentives for tariff customers; LSEs can design energy contracts with better DR incentives; and many entities can improve the technological options and information available to consumers and ESPs. Such actions improve incentives and options for responding to high energy prices, inducing consumers to make the private investments and incur the private operating costs necessary to increase short-run DR.

Some of the actions that can improve DR in the above sense involve public goods—i.e., services and systems that, once provided to some, can be used by all—and hence it can be appropriate to subsidize or socialize their costs. Some private goods and services may be cost-effective for individual consumers only if many of them buy the same service from the same supplier, in which case it may be appropriate for a monopoly to provide the service at regulated, cost-based

prices. It may even be appropriate in some cases for regulators to require that all consumers in some class buy certain devices such as time-of-use (TOU) meters. But once the cost-effective public-good investments have been made and the necessary monopoly services have been provided/required, there is no need or reason to subsidize and socialize demand reductions per se. As long as the socialized and monopoly costs of improving DR in this way are not too large, total costs decline for consumers and for the economy as a whole.

In this paper, a policy or program is said to increase DR if it results in a greater reduction in demand at high prices, no matter how or at what cost it does so, while a policy or program is said to improve DR if it lowers DR costs and hence reduces market-determined demand at higher prices. These definitions imply that, if everything else stays the same, improving DR will also increase DR and in the process will lower the total cost of meeting demand, but not all increases in DR are an improvement in DR. Indeed, any increase in DR resulting from programs that do not improve DR in the sense used here will increase demand-side costs more than it decreases supply-side costs and hence will create net costs, not benefits, for the economy and ultimately for consumers.

2.2 COST REDUCTIONS VERSUS RENT TRANSFERS

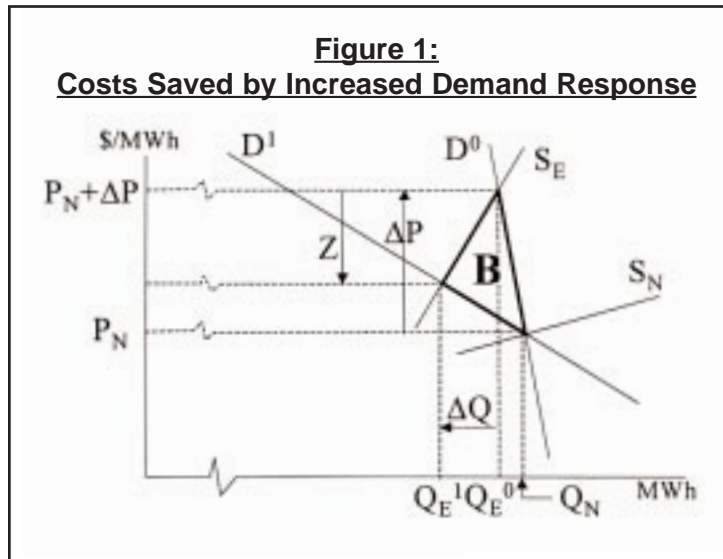
Discussions of DR often confuse prices with costs, and rent transfers with real cost reductions, and as a result make incorrect and unrealistic claims about the potential benefits of increasing DR. This section outlines the critical differences between prices and costs, and between rent transfers and cost reductions, and describes some of the implications for what increasing DR—and subsidizing DR—can and cannot really do.

2.2.1 THE ECONOMIC BENEFIT OF INCREASED DR IS LOWER SUPPLY-SIDE-PLUS-DEMAND-SIDE COSTS

The net economic benefit of an increase in DR is the reduction in the total supply-side plus demand-side costs of meeting consumers' demand.⁶ For example, if peak DR is improved by making it easier and cheaper to run back-up generation and turn off electricity-using industrial processes when prices are high, demand and hence the need for costly supply during high-priced peak periods will be reduced. The gross benefit of the improvement in peak DR will be the reduction in supply-side costs because fewer peakers are running in the short run and needed in the long run. But these gross DR benefits are obtained by running and perhaps buying more back-up generators and by changing energy-using industrial processes, which are costly. The net benefit of the increase in DR is the saving in peaker costs less the increase in the costs of back-up generation and industrial operations.

6 "Meeting consumers' demand" is shorthand here for something like "providing a mix of goods and services of equal value to consumers." For a good discussion of the conceptual and practical problems involved in defining the costs and benefits of DR, see Steven Stoft, "The Economics of Conserved-Energy 'Supply' Curves," April 1995 (available at <http://www.stoft.com/lib/papers>).

The concept of reducing total costs by increasing DR, and the magnitude of the resulting cost savings compared to some other possible measures of benefits, can be illustrated using the simple model in Figure 1. In this model, the supply curve is S_N “normally” but shifts to S_E during “events” that reduce supply and cause price spikes.⁷ The market price is P_N under normal conditions but increases during price spike events by an amount that depends on the demand response to price increases. Initially, demand is highly inelastic as illustrated by the demand curve D^0 ,



so during price spike events demand decreases only from Q_N to Q_E^0 and the market price increases to $P_N + \Delta P$. But then something is done to make demand more elastic as illustrated by demand curve D^1 , so the reduction in demand during price spikes increases by $\Delta Q = Q_E^0 - Q_E^1$ and price spikes are reduced by the amount Z .

Under standard economic assumptions, the reduction in total costs resulting from the increase in DR is approximately the area B in Figure 1,⁸ which is a triangle if the demand and supply curves are straight lines. If the initial demand curve D^0 is perfectly inelastic or vertical, so that $Q_N = Q_E^0$, the area of that triangle is:

$$\text{Area } B = \text{Net Benefit from the Increase in DR} = \Delta P \times \Delta Q / 2.$$

This expression says that the economic benefit of increasing DR during price spikes depends on the size of price spikes before the increase in DR (ΔP) and on the resulting increase in demand reduction during price spike events (ΔQ), but does not depend on how much price spikes are reduced (Z); the above formula for area B does not contain Z at all.⁹ The striking difference between this conclusion—that benefits do not depend on how much or even whether price

7 Shifts in demand could be used to make the same points made here and would be more consistent with the use of the term “peaks” to describe price-spike events, but are more difficult to represent in a simple diagram that also includes a shift in demand due to improved DR.

8 At each MWh level on the horizontal axis, the height of a supply curve represents the cost of producing the incremental MWh and the height of a demand curve represents the cost to consumers of “producing” the incremental MWh by not consuming it. Thus, the area under the supply and demand curves forming the top of area B represents the cost of meeting the MWh of demand between Q_E^1 and Q_N during price spike events before the increase in DR using a combination of supply increases and demand reductions, and the area under the demand curve forming the bottom of area B represents the cost of meeting this demand by reducing consumption after the increase in DR. The difference between these two areas is area B —the net cost saving or benefit due to the increase in DR.

9 If D^0 has some small elasticity the expression for area B does depend on Z , but not much and only because the factors that increase net benefits—e.g., higher demand elasticity after the increase in DR—also increase the Z associated with any ΔQ .

spikes are reduced—and the common view that the main point of DR programs is to reduce price spikes is discussed and explained below.

2.2.2 SMALL INCREASES IN DR CANNOT CAUSE LARGE CHANGES IN COSTS AND HENCE CANNOT HAVE LARGE BENEFITS

The above expression for the net benefit of an increase in DR has another implication that runs counter to conventional thinking about DR: a small fractional reduction in demand will produce benefits that are only a small fraction of consumers' total bills, or of changes in total bills if these are large relative to total bills. For example, in Figure 1 the total increase in consumers' bills during each price spike event—call it the “bill spike”—before the increase in DR was $\Delta P \times Q_N$ (assuming D^0 is totally inelastic or vertical). As a fraction of this amount, the net benefit resulting from the increase in DR is:

$$\begin{aligned} \text{Net Benefit as a Fraction of the “Bill Spike”} &= \text{Area B} / \Delta P \times Q_N \\ &= (\Delta P \times \Delta Q / 2) / (\Delta P \times Q_N) \\ &= \Delta Q / 2Q_N \end{aligned}$$

If the increase in DR reduces demand during price spikes by a fraction X , then $\Delta Q = X \times Q_N$ and this reduces to:

$$\text{Net Benefits as a Fraction of the “Bill Spike”} = X \times Q_N / 2Q_N = X/2.$$

Thus, an increase in DR that reduces demand during price spikes by some small percentage X will have net benefits that are about half as large as a percentage of the total bill spike, even if the increase in DR dramatically lowers or even totally eliminates the spikes in prices and bills. For example, if price spikes are producing bill spikes that cost consumers \$100 million and a 5 percent reduction in peak demand cuts this in half (or eliminates it altogether), the reduction in bills will be \$50 million (or \$100 million) but the actual benefits will be only about \$100 million $\times 0.05/2 = \$2.5$ million.

Again, this result is very different from the widespread view that reducing peak demand by a few percent can have benefits that are large percentages of total bills or at least of bill spikes. The basic reason for this difference is that what are commonly called benefits are, as shown below, actually transient rent transfers. The reality is that a small percentage reduction in demand can cause only a similarly small percentage reduction in supply and hence the reduction in supply-side costs can be at most a somewhat larger (if the reduction in demand eliminates particularly costly supply) but still small percent of total supply-side costs. And the reduction in demand is not free, so the net benefits will be less than that. Unfortunately, the real world tends to be continuous and niggardly, so it is wishful thinking to expect a small reduction in demand to yield very large benefits.

2.2.3 BILL REDUCTIONS DUE TO LOWER PRICES ARE RENT TRANSFERS, NOT COST REDUCTIONS OR SOCIAL BENEFITS

The real attention-grabbing effects of increasing DR are not the relatively small net reductions in costs discussed above, but the much more dramatic reductions in prices and consumers' bills that can result from small reductions in demand during peak periods when short-run supply is highly inelastic. But reductions in consumers' bills due to price declines are not benefits for society as a whole; they are transfers of economic rents¹⁰ from one set of pockets to another, often in the same pair of pants.

Reducing prices and transferring rents from suppliers to consumers may be regarded as desirable, particularly when suppliers are perceived to be taking advantage of or even creating price spikes that serve no useful purpose. But price spikes are normal and even essential in any commodity market, because they stimulate short-run increases in supply and decreases in demand, encourage long-run investments that moderate future price spikes, and make significant contributions to the capital costs of all producers. Reducing or eliminating excessive or artificially created price spikes is a worthy goal, although it is as difficult to define as it is to achieve. But regarding all or even most of the rents accruing to producers during price spikes as found-money that really belongs to consumers is incorrect in concept and dangerous in practice.

Furthermore, it is not always clear who ultimately gains and loses how much because of rent transfers resulting from price changes. The ultimate impact will depend on contracts, regulatory arrangements, tax incidence, and asset ownership. Changes in producers' profits change the value of mutual funds and retirement accounts held by the same people who buy electricity; for example, the California Public Employees Retirement Fund (CALPERS) was one of the biggest shareholders in Enron. It is never clear even after the fact who actually gains and loses how much from a change in prices once all effects are considered, but the final impacts of any price drop due to an increase in DR will certainly be more diffuse and mutually-offsetting than is implied by the us-versus-them, consumers-versus-producers rhetoric often heard in discussions of DR.

Most estimates of the benefits of increasing DR focus on the dramatic bill reductions that are said to be achievable with modest reductions in demand. For example, McKinsey and Company estimates that exposing essentially all U.S. consumers to real-time prices could produce total annual "benefits" of \$10–15 billion per year; but at least 80 percent of this amount is a pure rent transfer due to the projected price decreases, not a reduction in real costs. Many estimates of benefits from increased DR do not even try to estimate real cost reductions, focusing entirely on reductions in

¹⁰ An economic rent is any payment to a supplier that exceeds the actual cost (including competitive profit) of supply in the relevant time frame, e.g., short-run or long-run. In a capital-intensive industry short-run rents must be large to cover suppliers' fixed and capital costs.

consumers' bills, and hence misrepresent and significantly overstate the "benefits" from an increase in DR.¹¹ And virtually no estimates of the benefits of increased DR include or even acknowledge the demand-side costs necessary to achieve the supply-side benefits.

2.2.4 RENT TRANSFERS FROM AN INCREASE IN DR ARE TRANSIENT AND DEPEND ON SURPRISING SUPPLIERS

Even if rent transfers from suppliers to consumers are regarded as desirable, at least most of any such transfers resulting from an increase in peak DR will be transient at best and cannot be expected to occur at all unless the increase in DR surprises suppliers. Presumably consumers would be happy to accept even a transient windfall created by surprising suppliers, but policies designed on the assumption that consumers can gain from the mistakes made by surprised suppliers are unlikely to be successful, much less efficient.

Given the short-run inelasticity of both supply and demand during peak periods, a sudden increase in peak DR can cause large immediate declines in peak prices and consumers' bills. But any decline in bills in excess of the small reduction in supply-side costs resulting from mothballing or not using a few peakers must be reducing the short-run rents of the large amount of capacity that is still needed. Assuming that all needed generation capacity will, in an expected value sense over time, recover its total costs plus a competitive profit and not much more or less—which is, after all, a basic assumption behind the move to competitive electricity markets—time-weighted average prices must sooner or later, and probably sooner, get back up to about where they would have been if peak DR had not increased. In the long run, even the level and volatility of peak prices might not be reduced much or at all. Demand-weighted average prices paid by consumers will be slightly lower than they would otherwise have been because demand is relatively lower during the high-price peak periods, but this load-factor effect will be much smaller than the immediate, transient drops in peak and average prices that are often cited as reasons to encourage and even to subsidize DR.

These effects can be illustrated with a simple example. Assume an electricity system with a constant non-peak demand of 100 MW in 90 percent of the hours, a constant peak demand—initially 120 MW—in 10 percent of the hours, and only two kinds of generators, non-peakers and peakers. Initially the system is in an equilibrium with (ignoring outages and the need for reserves) 100 MW of non-peakers that run all the time and 20 MW of peakers that run in peak hours, and with prices of \$50/MWh and \$200/MWh in non-peak and peak periods, respectively. The assumed demands and prices imply that the demand-weighted average price—the average price paid by the average consumer—is \$67.65/MWh and the average hourly consumer bill is \$6,900/hour.¹² If there are no barriers to entry or exit for either non-peakers or peakers, the

11 See "Demand Response: Principles for Regulatory Guidance," February, 2002, Peak Load Management Alliance (PLMA "Principles") for summaries of the McKinsey study and others in which dramatic bill reductions are a major, the major or even the only "benefit" cited for increased DR.

12 The details of all calculations for this example are in the Appendix.

equilibrium prices must be just covering the total long-run marginal costs (LRMC) of both types of plants.

The transient and longer-term effects of an increase in peak DR can be illustrated by assuming that peak demand drops suddenly to 110 MW—a 9 percent drop in peak demand and a 1 percent drop in total MWh consumption. This creates an excess of peaker capacity that immediately drives peak prices down to the (assumed) \$100/MWh running cost of peakers, which decreases the peak price, the demand-weighted average price, and the average consumers' total bill by 50, 18, and 19 percent, respectively. Such large declines in prices and bills from such a small drop in demand are not out of line with some of the more enthusiastic claims about the expected “benefits” of increases in peak DR,¹³ and are not necessarily implausible as a short-run effect given the excess peaking capacity. But large price declines due to excess capacity are short-run transients that cannot be sustained if generation is to remain a commercially viable business.

Because there is now excess peaking capacity, peak prices can remain below peaker LRMC for some time. But the increase in peak DR has done nothing to reduce the need for or costs of non-peakers, so as soon as demand growth or plant retirements require investment in these the average price received by non-peakers—i.e., the time-weighted average of the peak and non-peak prices, given the assumption that non-peakers run all the time—must increase to whatever it would have been without the increase in DR. If the peak price stays at \$100/MWh and LRMCs do not change over this “medium run,” non-peak prices must increase enough to increase average consumer prices and bills until they are only 3.4 and 4.4 percent below their initial levels, compared to the immediate but transient declines of 18 percent and 19 percent, respectively.

In the long run, new peakers will be needed as well, so the peak price must eventually increase to cover the LRMC of peakers. The higher peak prices will allow non-peakers to cover their costs with lower non-peak prices, but in the long run, peak and non-peak prices—and hence price volatility¹⁴ - will be what they would have been without the increase in peak DR. If LRMCs do not change, long run average consumer prices will be down less than 2 percent because of the improved load factor, and the gross benefit to consumers—the decline in total bills—will be less than 3 percent, which is exactly the same as the decline in suppliers' costs. The net benefit of the increase in DR, both to consumers and to society as a whole, will be the 3 percent decline in bills/costs less the costs consumers incur to reduce their peak demand by 10 percent. On a long run, on-going basis, consumers will gain nothing at the expense of producers and hence will be ahead of the game only if the increase in DR is really cost-effective for society as a whole.

13 For example, an EPRI study concluded that “... a 2.5% reduction in electricity demand statewide could reduce wholesale spot prices in California by as much as 24%; a 10% reduction in demand might slash wholesale price spikes by half.” Taylor Moore, “Energizing Customer Demand Response in California,” *EPRI Journal*, Summer 2001, p. 8.

14 This example demonstrates that increasing DR may not decrease even price volatility in the long run. In fact, some patterns of increased DR could increase volatility. For example, if increasing DR reduced the peak but changed its shape so that demand was less price-elastic during, say, 5 percent of the hours, the peak price during these 5 percent of hours could increase, increasing price volatility. An example is shown in Section 5 of the Appendix.

Even the transient gains for consumers at the expense of producers in this example are obtained only because suppliers are surprised by the increase in DR and hence are caught with excess peaker capacity. But surely, if suppliers realize in advance that peak demand will suddenly fall on some “DR Day” they will—and should—stop investing in and maintaining peakers (and perhaps non-peakers as well) some time before DR Day. Although the details of the transition will depend on many things not specified in the simple example, supplier anticipation of DR Day will cause peak prices to be higher before DR Day, fall less on DR Day, and increase sooner after DR Day. If suppliers could perfectly anticipate and adjust everything, the total reduction in consumers’ bills over the period would be about the same as the reduction in suppliers’ costs over that period and there would be no rent transfer for the period as a whole.

In reality, DR does not increase suddenly, suppliers cannot perfectly anticipate and adjust to everything, there are many different kinds of generators, etc., so the effects of an increase in peak DR will be much more complex and unpredictable than this simple example suggests. But the basic principles illustrated by this simple example apply much more broadly. If suppliers underestimate future peak DR, they may provide too much peaking capacity that will drive down peak prices as DR takes effect, temporarily reducing the level and volatility of peak prices and creating transient windfalls for consumers and losses for suppliers. If suppliers believe unrealistically bullish projections for future DR, they may provide too little peaking capacity, causing peak prices to be higher and more volatile and giving suppliers a temporary windfall at consumers’ expense when the promised DR fails to materialize. Anything can happen in a complex and uncertain world, but the best bet is that consumers will not gain at the expense of suppliers in any long-run, expected-value sense.

2.2.5 EVEN IF DR REDUCES EXTRAORDINARY PEAK-PERIOD MARKET POWER, THE RENT TRANSFERS WILL PROBABLY BE SMALL

Even if price spikes are exacerbated by the extraordinary market power that generators can have during peak periods, any rent transfer from producers to consumers due to a small increase in DR will be much smaller than estimates based on the immediate price effects assuming that suppliers are surprised and no other prices change. The fact that peak-period gaming and market power may be pushing up peak prices and rents does not change the arithmetic conclusion that most of the “too high” peak rents are going to non-peaking generators. If the generation market is workably competitive outside peak hours, the non-peakers must earn about the same over all hours both before and not-too-long-after the increase in peak DR, so time-average prices—and perhaps even the level and volatility of peak prices—cannot change much in the medium or long runs.

If the generation market is reasonably competitive outside the peak hours when both demand and supply are highly inelastic, the amount that consumers can permanently gain at the expense of producers by reducing peak-period market power cannot be more than the excess or monopoly profits being earned by generators who make most of their rents during peak periods, i.e., peakers. Given that peakers are not exceptionally profitable and that there are not many of them, the amount of excess peaker profits potentially available to consumers will be less, and probably much less, than the immediate drops in consumers' bills estimated by looking at highly inelastic supply curves and assuming that these will not move in response to a decline in peak demand and prices.

These conclusions can be illustrated by modifying the example above to assume that the LRMC of peakers is only \$150/MWh but that the peculiarities of the peak period would allow generators to keep the peak price at \$200/MWh indefinitely unless peak DR increases. If the unexpected increase in peak DR totally eliminates peak-period market power, the peak price will drop 50 percent to the assumed \$100/MWh running costs of peakers, implying that average consumer prices and bills will immediately fall 18 and 19 percent, respectively, as in the previous example. But also as before, if the non-peaker market is workably competitive, time-average prices must soon increase back to the LRMC of non-peakers, so the immediate large price drops will be short-lived. In the long run (and assuming that LRMCs do not change), the peak price will increase to the peaker LRMC of \$150/MWh, which is 25 percent less than it was initially; but the non-peak price will have to be \$55.55/MWh, or 11 percent higher than it was initially, so that non-peakers can cover their total costs. (See the Appendix for all calculations.) The total consumer bill will fall by the small actual reduction in peak supply costs plus any decrease in the excess profits that were being earned by peakers, which is also likely to be small. In the example, the excess profits being earned by peakers before the increase in peak DR was only 1.5 percent of the initial consumers' bill, so even total elimination of peak-period market power would not do much for consumers. And if peaker owners see the increase in DR coming, they will reduce peaker capacity in advance and reduce even the transient gain to consumers—and perhaps even maintain or quickly reestablish their ability to keep prices above LRMC indefinitely.

The suggestion that an increase in peak DR may not greatly reduce peak-period market power in the long run may appear heretical or (much worse) illogical. But an increase in peak DR will not reduce the barriers to entry implied by the proposition that peakers can make monopoly profits indefinitely unless peak DR is increased, particularly if the increase in DR threatens to reduce peak prices. Nor will an increase in peak DR necessarily change the peak-period market game much in the long run, given that even highly inelastic short-run supply curves shift in the long run. For example, if most of the new DR takes effect at relatively low prices, the peak demand curve will shift but will still be nearly vertical at higher prices, so once a few peakers

have been shut down or not built the old peaker cartel will be reestablished—but now with fewer players who may have more effective pricing power.

2.2.6 CONSUMERS CAN GAIN FROM IMPLEMENTING COST-EFFECTIVE DR, BUT NOT FROM SUBSIDIZING NON-COST-EFFECTIVE DR

The analysis above demonstrates that consumers should expect an increase in peak DR to reduce their bills about as much as it reduces suppliers' costs. Because prices in a reasonably efficient market approximate suppliers' incremental costs, it follows that if an individual consumer finds it cheaper to implement peak DR than to pay high market prices for peak electricity, implementing that DR must increase that consumer's—and hence total consumers'—demand-side costs less than it decreases suppliers' costs, and hence implementing that DR is good for the individual consumer, for consumers as a group, and for society as a whole. Conversely, if a consumer finds that it is cheaper to keep paying high peak prices than to implement some DR, that DR would cost that consumer—and hence consumers as a whole—more than it saves suppliers and hence more than the likely reduction in the total consumers' bill, so implementing that DR would not be good from any standpoint. What is good (or not) for individual consumers at competitive market prices is good (or not) for consumers as a group and for the economy as a whole.

This result demonstrates the fallacy in the notion that increasing peak DR creates consumer “externalities” in the form of lower prices, and hence consumers should agreed to subsidize DR and tax themselves with an uplift to pay for the subsidies. Such a tax-subsidy arrangement will end up subsidizing DR that costs consumers more than the resulting reduction in suppliers' costs and hence ultimately in consumers' bills. DR that must be subsidized is unlikely to reduce wholesale prices significantly in any predictable, permanent sense, but even if it does the uplift necessary to pay for the subsidies will be larger than any fall in wholesale prices, increasing the average prices paid by consumers. If consumers as a whole try to drive down prices by taxing themselves to subsidize some of them to reduce demand more than they would find individually beneficial—in effect, creating a buyers' cartel—consumers as a whole will lose.

2.2.7 THIS IS ALL TRUE DESPITE THE “CALIFORNIA CRISIS,” WHICH HAS LITTLE RELEVANCE FOR DR IN ITP MARKETS

The economic arguments above may be—in fact, have been—criticized for seeming to ignore the huge rent transfers and allegedly monopoly profits created by dramatic events such as the crisis in western power markets in 2000–2001. But this experience—call it the “California crisis” for short—has few implications for the arguments above and even fewer for the integration of DR into short-run ITP markets.

It is no doubt true that a small percentage reduction in demand (or increase in rainfall) during the California crisis could have reduced prices and transferred proportionally larger rents (back) to California consumers¹⁵ from suppliers. But the fact that a hypothetical sudden increase in DR might hypothetically have saved consumers a lot of money at suppliers' expense two years ago does not mean that realistic increases in DR can realistically be expected to do so in the future. If suppliers expect increasing DR to reduce the probability and level of high prices in the future, they will adjust their plans and operations in ways that will reduce or eliminate any rent transfers in the future; some generation projects intended for the California market have already been delayed or cancelled. California should increase DR because, and to the extent that, doing so is expected to decrease supply-side costs more than it increases demand-side costs, not in the illogical and unrealistic hope that increasing DR will benefit consumers at suppliers' expense.

The California crisis was caused by a regional scarcity—whether real or contrived makes little difference for the issues here—over many months, compounded by a lack of contracting, a badly flawed central market design, and the almost total insulation of consumers from wholesale market prices; it had little to do with the fact that the California ISO was not accepting or using demand bids in its real-time dispatch and pricing decisions. With consumers paying a fixed price of about \$50/MWh in every hour when the monthly average wholesale (PX) price was more like \$200/MWh for many months in a row, nobody needed sophisticated TOU meters to know that it was cheaper to reduce demand than to increase supply in virtually any week-day hour, and if demand had responded to market prices the ISO's demand forecasts would have captured this with no need for real-time demand bidding. A large fraction of demand already had time-of-use meters, and regulated utility LSEs were already reporting on the monthly bill of each consumer the average PX price and both this year's and last year's consumption for the billing month. If utility LSEs had been ordered or allowed to charge consumers the monthly average PX price for monthly consumption above (say) 90 percent of last year's, total demand, prices, and the financial distress of utilities would have been significantly reduced. California needed a better short-run political response in the capital a lot more than it needed more short-run demand response programs in the ISO.

2.3 PRICE VOLATILITY AND SYSTEM RELIABILITY EFFECTS

Improving peak DR in the energy market, in the sense of making it more cost-effective for market participants to reduce demand when wholesale spot prices are high, will make it easier and cheaper to meet demand reliably and will probably reduce price volatility. But these effects are not benefits over and above the benefits discussed above—i.e., the reduced total costs of meeting demand—and in particular do not justify subsidies to increase DR above market-determined levels.

15 California electricity consumers were almost fully insulated from high prices during the crisis, but will end up paying most of the bill in the form of higher prices and higher taxes—and lower value of utility shares in their retirement portfolios—in the long run.

2.3.1 PRICE VOLATILITY CAN AND SHOULD BE MANAGED BY CONTRACTING, NOT BY SUBSIDIZING DR

The level and volatility of prices in ITP markets should be no higher—but should also be no lower—than necessary to reflect the level and volatility of the underlying costs. Once the ITP has incorporated DR fully into its markets and others have done what they should to improve DR, price spikes and volatility will be reduced as much as is cost-effective. Market participants should then use contracts and other commercial arrangements to manage the remaining volatility. The ITP should not try to drive prices or volatility in its markets to artificially low levels, because this would distort the markets and increase total costs.

The concern about the volatility of spot prices is based largely on misconceptions about the role of spot prices in a commodity market. Most transactions in most commodity markets, including essentially all electricity markets (except the California market in 1998–2001), are covered by bilateral contracts (or vertical integration, which is a form of contract) with varying lengths and prices. A spot market is not a substitute for contracts, but is the most efficient way to manage and price short-run system effects and to facilitate contracting by automatically pricing and trading contract imbalances and determining contract reference prices. If spot prices are regarded as too volatile and unpredictable for most commercial transactions—as they should be—the solution is not to dampen volatility artificially by imposing price caps or subsidizing DR or peakers, but to use contracts to manage the risks. Incremental changes in DR will not change volatility enough—if at all¹⁶—to have much effect on the optimal degree and cost of contracting or on the residual risks due to volatility.

An ITP may use price caps and/or an available/installed capacity/capability (ICAP) requirement in efforts to reduce volatility in its energy market. If it does so, DR should be incorporated into these processes on an equivalent basis with supply resources. Price caps and ICAP are discussed in sections 4.2.4 and 4.2.5 respectively.

2.3.2 DR SHOULD BE PAID FOR RELIABILITY SERVICES IN ANCILLARY SERVICE MARKETS, NOT BY SUBSIDIZING DR IN ENERGY MARKETS

In principle, reliability in a market-based electricity system is not something separate from price volatility but is implicit in it. Market prices should go as high as necessary as often as necessary to assure that there is always an adequate supply (including reserves, and including DR as reserves when DR can provide equivalent services) to meet demand at those prices. If the market always clears in this sense, an improvement in reliability is the same thing as a reduction in the cost of meeting demand reliably.

¹⁶ As discussed in section 2.2.4, there is no guarantee that increasing peak DR will reduce volatility in the long run, although under many “reasonable” assumptions it probably will.

In practice, ITP energy markets do not always clear in the sense outlined above, because energy prices sometimes cannot as a technical matter or are not allowed as a political matter to get to market clearing levels. In these cases the ITP must take actions outside the energy market to deal with the problems, and when it does so by buying non-energy-market services from generators it should also buy the same or equivalent services from DR resources to the extent they can provide them.

For example, energy prices determined every hour (or even every five minutes) cannot capture all the costs associated with short-run events such as the sudden loss of a large generator or transmission line, so the ITP pays for ancillary services to do what the energy market cannot. As discussed in section 4.2.5, DR resources should be allowed to provide such reliability services just as supply resources are. But there is no reason to subsidize DR in the energy market in an effort to improve reliability.

3. THE THEORY OF DEMAND RESPONSE IN ELECTRICITY

Asking whether it is demand or supply that is more important in determining market prices and quantities is like asking whether it is the hammer or the anvil that is more important in shaping the metal: it is the interaction between the two, not one or the other, and not even one more than the other, that gets the job done. But demand is not supply, just as a hammer is not an anvil. To get the job done well, the system must be designed to reflect the different characteristics and functions of each of its component parts, not to force a superficial equality of form or process. This section discusses how demand and supply are—and are not—treated in normal markets, and how DR should—and should not—be treated in ITP markets.

3.1 DEMAND AND DEMAND RESPONSE IN NORMAL MARKETS

In a normal market, both supply and demand respond to prices—in the long run if not in the short run—and these responses affect the market outcomes. But normal markets incorporate price-responsive demand without any concept of DR as a “resource” and without paying anybody not to buy the product. ITP energy markets should do the same.

3.1.1 THE OUTCOME OF A COMPETITIVE MARKET PROCESS IS EFFICIENT, EVEN WHEN DEMAND AND SUPPLY “PARTICIPATE” VERY DIFFERENTLY

In a normal competitive market, each seller sells the amount that maximizes its profits at the market price, implying that the marginal cost of increasing that seller’s supply equals the market price. Each buyer in such a market buys the amount that best meets its needs given the

market price, implying that the incremental cost in money, time, and convenience to that consumer of reducing its demand equals the market price. Because all producers are paid and all consumers are paying the same market price, the incremental cost each producer is incurring to increase supply equals the incremental cost each consumer is incurring to reduce demand. Any significant change in or reallocation of the quantities produced and consumed would cause the incremental costs of individual buyers and sellers to diverge from one another, creating more costs than benefits and increasing total costs.

The efficiency of a market-clearing process does not require that both demand and supply participate in the market in the same way. For example, in many markets—e.g., for tomatoes and real estate—sellers post prices, observe whether and how much consumers choose to buy at those prices, and change the posted prices until supply equals demand. The fact that suppliers rather than demanders initially post and then change the prices while demanders “only” decide how much to buy at those prices does not mean that demand does not participate fully or appropriately in the market. There is no presumption that supply and demand must participate in “the same” way in order for a market to reach an efficient solution.

In many efficient real-world markets either demand or supply is totally passive. For example, in the Tokyo fish market the sushi-on-the-hoof brought in last night is laid out on the floor each morning, wholesale distributors bid in an auction to buy that fixed quantity, and the fishermen wait on the sidelines to see what price they get. If the fishermen do not like today’s prices they may not go fishing tomorrow, which will tend to increase tomorrow’s price; but in today’s market suppliers are passive bystanders. A market in which either producers or consumers cannot or choose not to respond to normal changes in short-run prices can reach an efficient, cost-minimizing equilibrium if the other side of the market is adequately competitive—although what it takes for one side of a market to be “adequately competitive” may depend on what the other side does.

3.1.2 A CONSUMER CAN SELL WHAT IT DOES NOT CONSUME—BUT ONLY IF IT OWNS IT

A consumer who normally buys in a market can become a supplier in that market if it brings to or buys in the market more than it consumes. But a consumer must own everything it consumes itself plus everything it sells, and the only way it can get ownership of something is to produce it itself or buy it, ultimately from somebody who does produce it. Such simple but fundamental economic and commercial realities are often forgotten where DR is concerned.

As a matter of logic, not to mention common sense, if a consumer neither produces anything nor buys anything from somebody who does, it has nothing to sell in the market and there is no reason to pay it for anything. Paying a consumer for something it does not own is like buy-

ing the Brooklyn Bridge from somebody who does not own it: illogical on its face and patently unfair to those who get tricked into paying for nothing. Most importantly for the analysis here, paying consumers for something they would have bought but didn't is inefficient, because it encourages consumers to incur unproductive costs doing whatever it is they must do to increase the amount they are—illogically and inefficiently—allowed to sell.

3.1.3 PAYING A CONSUMER FOR DR "RESOURCES" IT WOULD HAVE BOUGHT BUT DIDN'T IS PAYING TWICE FOR THE SAME THING

Even if a consumer can prove conclusively that it would have consumed some specific quantity at some price, paying the consumer to reduce its consumption below that level without requiring that it either produce or buy what it sells is essentially paying twice for the same thing. To see this, suppose a consumer usually buys and consumes some "baseline consumption level" (BCL) x when the market price is near P_x , but when the market price P is higher than P_x would prefer to buy and consume some smaller quantity q . The reduction in consumption at the higher price P , $dr = x - q$, can be called a DR "resource"—although this is a misleading term that creates more confusion than insight.

If the difference between what a consumer would normally consume, x , and what it does consume, q , is called a DR "resource," it sounds reasonable to say that the consumer should be paid for its DR "resource" just as suppliers are paid for their supply resources. And it also sounds reasonable to say that no consumer should have to pay for something it does not consume, so it should pay the market price only for the q it actually consumes. Together, these two reasonable-sounding statements imply that, if the consumer's actual consumption is $q < x$, it should pay the market price P for the q it actually consumes and should be paid P for the amount $dr = x - q$ of DR "resources" it provides. In this case, the net payment or credit to the consumer would be:

$$\text{Net Payment TO Consumer} = P \times dr - P \times q$$

But, because $q = x - dr$, this can be rewritten as:

$$\begin{aligned} \text{Net Payment TO Consumer} &= P \times dr - P \times (x - dr) \\ &= 2P \times dr - P \times x \end{aligned}$$

This expression says that a consumer who consumes less than its fixed BCL of x should buy x at the market price and then be paid a price of $2P$ for selling back DR "resources." But a market that pays $2P$ to buy back DR "resources" when consumers are buying and real suppliers are

providing the same stuff for a price of P would clearly be illogical, inefficient, unfair—and short-lived. How did two reasonable-sounding propositions lead to such an unreasonable conclusion?

The problem here is that one of the two reasonable-sounding propositions above—nobody should have to pay for something they do not consume—is in fact logically and practically untenable. Of course a consumer should pay for something it does not consume if it is going to sell that something to someone else (and did not produce it itself or buy it from someone who did). Letting a consumer sell something just because it would have bought it but didn't is logically equivalent to letting the con artist sell the Brooklyn Bridge if he can prove that he thought about buying it before deciding to sell it instead.

The obvious solution to the double-payment problem here is simply to say that nobody can sell anything they do not own. If the consumer does not produce or buy its BCL, it has nothing to sell. The consumer in the example who actually consumes q could be allowed to buy its BCL x in the market at the market price P and then sell DR “resources” $dr = x - q$ at the same price, so that the net payment to the consumer would be:

$$\begin{aligned} \text{Net Payment TO Consumer} &= P \times dr - P \times x \\ &= P \times (x - q) - P \times x \\ &= - P \times q \end{aligned}$$

But a negative net payment TO the consumer is a payment BY the consumer, so this is just another way of saying that the consumer should pay the spot price for what it consumes - period. A consumer who brings nothing to the market has nothing to sell and hence should not be paid anything for what it might have bought but didn't.

3.1.4 FIXED RETAIL PRICES COMPLICATE THINGS BUT DO NOT CHANGE THE LOGIC THAT A CONSUMER MUST OWN WHAT IT SELLS

In electricity markets, many consumers buy from LSEs under tariffs or full-requirements contracts (FRCs) that allow them to take as much or as little as they please at a fixed price. Such a tariff or FRC is sometimes interpreted to give the consumer some right to the BCL x that it would normally buy at the tariff/FRC price P_x , including the right to sell at the market price P any of the amount x that it does not consume. It is not clear that a tariff or FRC should be interpreted in this way, but even if it is, any consumer who exercises its right to take its full BCL x so that it can sell some of it must also fulfill its obligation to pay the price P_x for its full BCL. With this interpretation of a tariff/FRC, a consumer with a BCL of x who consumes some amount $q < x$ can sell at the market price P the DR “resource” amount $dr = x - q$ that it does not consume, but only if it pays the tariff/FRC price P_x for its full BCL of x , i.e.:

$$\text{Net Payment TO Consumer} = P \times dr - P_x \times x$$

Because $x = q + dr$, this can be rewritten as:

$$\text{Net Payment TO Consumer} = (P - P_x) \times dr - P_x \times q$$

This expression says that a consumer who pays the tariff/FRC price P_x for the amount q that it actually consumes has no right to be paid for its DR “resources” $dr = x - q$ any more than the difference $P - P_x$ between the market price and the tariff/FRC price. Paying the full market price for DR “resources” the consumer does not buy or otherwise own is, in effect, subsidizing demand reductions.

Of course, even paying the difference $P - P_x$ for demand reductions makes sense only if a tariff/FRC really does give a consumer the right to buy and resell what it does not consume, which is far from obvious. A normal commercial contract that gives the buyer the right to buy and resell some defined quantity will usually require the consumer to buy that quantity at the contract price even when the consumer does not want it (perhaps because the market price is less), and to pay the full market price for anything it consumes in excess of the contract quantity—either that, or the buyer will pay a high option fee for the valuable rights-without-obligations deemed to exist in a tariff or FRC. In the absence of such reciprocal obligations or option fees it is not at all clear that buyers under a tariff or FRC should be deemed to have rights to resell at a higher price what they would “normally” buy if they do not consume it.

In practice, not only are rights under tariffs and FRCs unclear, but trying to define BCLs for individual consumers inevitably creates inequities and distortions. Furthermore, if the LSE has contracted with a FRC supplier to serve the LSE’s tariff customers, it is the FRC supplier who stands to gain from any decline in consumption at high market prices and hence should be ultimately responsible for any DR payments made to induce consumers to consume less. It is hard enough for the parties directly involved to figure out who has what rights and obligations with respect to whom in a complex web of regulatory and commercial arrangements, but letting some third party—e.g., a regulator or ITP—decide for them would be even worse. The only way to avoid such problems is to convert tariffs and full-requirements contracts into something more akin to real commercial contracts in which not only the prices, but also the contract quantities and the reciprocal rights and obligations, are well defined.

3.2 DEMAND AND DEMAND RESPONSE IN ELECTRICITY MARKETS

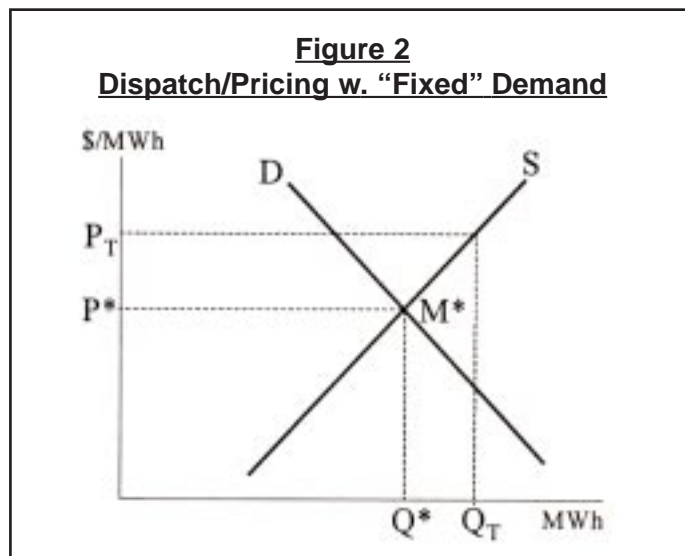
The new focus on DR in ITP markets raises the question of whether and how the dispatch and pricing processes used by ITPs should—or should not—be modified to take account of DR. This section discusses the planning and dispatch process in traditional utilities and in the new ISOs/ITPs in light of the economic principles discussed above, with emphasis on the fundamental errors that often occur when DR is treated as a “resource” that adds to supply rather than as a reduction in demand.

3.2.1 MINIMIZING SUPPLY-SIDE COSTS TO MEET A “FIXED” BUT MARKET-DETERMINED DEMAND TREATS DR FAIRLY AND EFFICIENTLY

Traditional vertically-integrated utilities did/do their long-run planning with the objective of minimizing the discounted long-run supply costs or utility revenue requirements of meeting forecast demand; they also did/do their short-run dispatching with the objective of minimizing the short-run supply costs of meeting actual demand. ISOs/ITPs now find market-clearing prices and quantities by minimizing the short-run supply costs of meeting forecast or actual demand; this is essentially the traditional utility dispatch objective, but with generator bids rather than engineering data used to estimate short-run supply costs and with much more emphasis on finding market-clearing energy prices that are consistent with the least-cost dispatch.¹⁷

For the purposes here, the essential feature of these processes is that their explicit cost-minimization step usually focuses exclusively on supply-side options and supply-side costs, and minimizes the supply-side costs of meeting an apparently “fixed” demand. But any competent ITP dispatch/pricing process (now) recognizes that demand is not independent of price. Such processes focus on supply when they explicitly minimize costs because the detailed information and control needed to evaluate and implement DR options efficiently resides with thousands or millions of consumers and hundreds or thousands of ESPs, not with the ITP. So the ITP deals in its explicit cost-minimization process only with the few hundred supply options it can realistically know and control in some detail, but uses—or at least should use—anything it knows about market demand to adjust the level of “fixed” demand that it meets at least supply-side cost.

The logic of an iterative dispatch/pricing process is illustrated in Figure 2. The ITP chooses some initial “fixed” target demands Q_T and then uses the supply costs implied by bids/offers from suppliers to determine a dispatch of supply resources that minimizes the total bid-cost¹⁸ of meeting the “fixed” demands Q_T ; this process also determines the prices P_T implied by the least-cost dispatch. The ITP then checks to see if market demands at the prices P_T are consistent with the “fixed” demands met, and if they are not adjusts things and tries again. As drawn in



17 The best ISO/ITP dispatch/pricing processes determine locational marginal prices (LMPs) for each of hundreds or thousands of nodes on the system as often as every five minutes. The discussion here abstracts from these complexities, speaking of a single market-clearing price each day or each hour.

18 Total bid-cost is defined as the total costs implied by the offers, as contrasted with total payments to suppliers at the uniform price defined by the highest bid accepted. If the ITP minimized the latter measure of cost, it would reduce purchases below the efficient level in order to drive down the price and its total payments at that price, in effect acting as a monopolist.

Figure 2, market demand at price P_T would be less than Q_T , so the ITP picks a lower “fixed” target demand, repeats the cost-minimization step, determines the prices associated with the new least-cost supply dispatch, estimates the market demand at those prices, etc. Ultimately, the ITP finds a “fixed” demand Q^* for which the least-cost dispatch of supply resources implies prices at which market demand is Q^* . This is a market-clearing equilibrium and hence it minimizes the sum of the supply-side and demand-side costs of meeting demand—even though each time costs were explicitly minimized in the process, only supply-side options and supply-side costs were explicitly considered and demand was “fixed.”

In practice, institutional and technological factors may prevent most demand from actually seeing and/or responding to wholesale prices, and ITPs may not always apply economic principles properly. But in concept the dispatch/pricing process outlined above does not ignore or discriminate against demand or DR options; it simply uses market demand as the best information available on how real consumers see and evaluate their real options. If the ITP does more to encourage market participants to submit demand bids indicating how much they will consume at different prices, and/or if the ITP, LSEs and ESPs provide consumers with better information, communication systems, meters, prices, technology, and services, the ITP will have a more price-responsive market demand curve to use in the above process and the result will be more DR and lower total costs. But at any time, the willingness of consumers to pay good money for electricity rather than consume less of it, as reflected in market demand—including any demand-side bids available at the time—is the best evidence there is of how real consumers perceive, value, and implement the real DR options available to them at that time.

It is critical to recognize that demand-side costs can be ignored in the explicit cost-minimization step of the above process precisely because DR is not being treated as a “resource” in this step. By definition, if DR “resources” are added to the options to be explicitly evaluated and compared in the cost-minimization step, the costs-to-be-minimized in that step must include all the incremental costs consumers incur when they reduce their demand in order to supply DR “resources.” Demand-side costs can be estimated using market data and bids much as supply-side costs are, but this must be done correctly, as discussed below.

3.2.2 TREATING DR AS AN ENERGY “RESOURCE” THAT ADDS TO SUPPLY OFTEN LEADS TO CONFUSION, ERRORS, AND INEFFICIENCIES

Utility DSM programs of the late 1980s and early 1990s typically maintained the traditional planning objective of minimizing total utility supply-side costs without regard to demand-side costs, even though they were now considering utility-funded DSM programs as “resources” in the utility planning process. Demand-side costs were ignored partly because they were too hard

to estimate and partly because—according to DSM advocates—they were insignificant anyway; DSM was said to be too cheap to meter. But the theoretical justification for ignoring demand-side costs was that utility planning had always sought to minimize utility costs or revenue-requirements without explicit consideration of demand-side costs, so changing this objective now just because DSM “resources” were in the planning mix would be unfair and undue discrimination against DSM.

The practice of treating demand reductions as a “resource” similar to supply resources and then minimizing only the “resource” costs of meeting demand has carried over to some proposed or actual DR programs in ISOs/RTOs/ITPs. The explicit or implicit argument for this approach is that ITP dispatch/pricing processes minimize the resource costs of meeting demand when they consider only supply resources, so to do anything different just because DR is now a “resource” would be discriminating against DR. What this argument ignores in both the utility DSM and the ITP DR contexts is that the fundamental objective of these processes is to minimize total costs, not just utility or ITP resource costs, and that it is logical to minimize only utility/ITP resource costs in the explicit cost-minimization step if, but only if, demand and hence demand-side costs are fixed for that step. When DSM/DR options are considered as “resources” in the planning/dispatch mix, this essential condition is not met.

To illustrate the problems created by trying to treat DR as an energy “resource,” consider a model ITP DR program that, while hypothetical, has features in common with most proposed and/or functioning ITP DR programs. In this model program, a qualifying ESP—who may be affiliated with a LSE,¹⁹ and is sometimes called a “demand resource response provider” (DRRP) or something similar—can act as a demand manager for a consumer or group of consumers. Each such ESP is given a baseline consumption level (BCL) based on what its consumer-clients would “normally” consume, and then each ESP submits to the ITP DR supply curves indicating how much actual consumption will be reduced below its BCL in response to DR payments from the ITP. The ITP uses some version of the dispatch/pricing process outlined in the preceding section to find a combination of supply resources and DR “resources” that minimizes the total bid-cost of meeting demand, pays the resulting “market-clearing” price to all supply resources and DR “resources,” and recovers its costs from LSEs.²⁰

The ITP’s first challenge in this type of DR program is to determine the BCL for each ESP. In some DR programs, each ESP’s BCL for one period is based on average or maximum consumption in recent days; for example, in New York the BCL is the average of the highest five daily demands over the previous ten days. Other programs base an ESP’s BCL on actual consump-

19 A large consumer who buys directly from the ITP markets is a “LSE” for the purposes here.

20 For example, the report “Recommended FERC Actions to Facilitate Demand Response Resource Programs Within Regional Transmission Organizations and Independent System Operators,” RETX, February 14, 2002, says: “The [real-time] Objective Function will be to minimize overall costs of Generation and Demand Response Resources” (p. 19); and DR “resources” should “...keep the tariff savings of unused energy as well as the market-clearing price ...” (p. 18).

tion immediately prior to the hour in which the ITP calls and begins paying the ESP for a demand reduction. Given that any consumer's consumption can vary widely from hour-to-hour or day-to-day for many reasons, any such rules are arbitrary and invite ESPs to manage their demand in order to increase their BCLs. As long as BCLs have value and are given away free by the ITP (or anybody else), there is no way to solve these problems by finding a more clever give-away process.

It is worth noting that nothing similar to the assignment of BCLs is required where supply resources are concerned. The ITP does not define a "baseline generation level" for each generator and pay/charge it for generating more/less; generators are simply paid for the energy that flows through their meters. The ITP does not give a contract energy seller a "baseline sales level" of free energy that it can sell today based on its contract sales yesterday; anybody selling a MWh today must generate that MWh today, buy it from somebody who does, or pay today's spot price for it. There must be something fundamentally wrong with any DR program that requires the ITP to give away valuable assets.

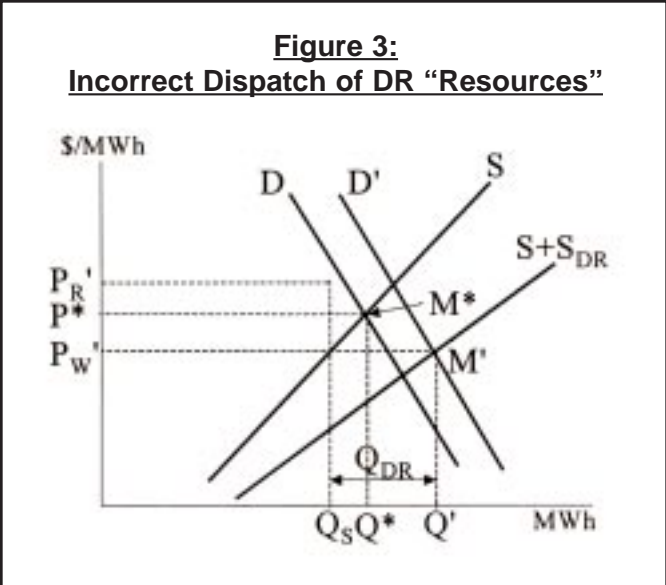
Once the ITP somehow defines the BCLs, each ESP submits to the ITP a DR supply curve indicating how much it will reduce its demand below its BCL at different DR "resource" prices P_{DR} , just as each generator tells the ITP how much energy it will supply at different wholesale market prices P_W . But the total incentive for demand reductions in this program is not just or even primarily the DR "resource" price P_{DR} , because consumers also pay some incremental retail price P_R for the MWh they continue to consume; a consumer who pays P_R for what it consumes and is paid P_{DR} for what it does not consume is effectively being paid $P_R + P_{DR}$ for not-consuming on the margin. If the DR supply curves submitted to the ITP indicate that large amounts of DR "resources" are available at each price P_{DR} , it is largely because consumers are actually being paid $P_R + P_{DR}$ for DR "resources." The DR payment from the ITP will significantly understate the incremental costs ESPs are incurring to reduce demand and supply DR "resources."

Once the ITP has obtained the DR supply curve S_{DR} and the "real" supply curve S , it adds the amount of DR "resources" offered at each level of P_{DR} to the amount of supply resources offered at the same level of P_W to obtain the total supply of "resources" at each price; this results in the supply curve $S + S_{DR}$ in Figure 3. The ITP may or may not also add some amount to the expected demand at each price to reflect the fact that some price-responsive demand has been incorporated into the supply curve; the adjusted demand curve is D' in Figure 3.²¹ The ITP then uses some version of the dispatch/pricing process outlined in section 3.2.1 above to find the prices that equate the total supply of "resources" to total demand. This solution is represented by point M' in Figure 3, where Q_S MWh of real supply resources and Q_{DR} MWh of DR "resources" are being dispatched at a wholesale price of P_W' . The ITP pays the price P_W' to the total ($Q_S + Q_{DR}$)

21 Clearly, this is a critical step. If the ITP adds to demand at each price the full amount of DR "resources" offered at that price, it will find the efficient solution M^* , as discussed in the next section. But the type of program discussed in this section does not usually add enough, if anything, to demand.

MWh of dispatched resources, but has only Q_S MWh of real resources to sell, so the ITP must charge buyers an average “uplifted” retail price of $P_R' = P_W' \times Q'/Q_S$ in order to balance its settlement accounts.

The solution M' obtained in this process is inefficient and does not clear any markets. Compared to the efficient, market-clearing solution M^* , M' has too much DR, not enough real supply, and a too-high uplifted retail price P_R' paid by LSEs. LSEs are paying one price P_R' , energy producers are being paid a lower price P_W' , and each DR “resource” provider is being paid its own



effective incremental price equal to P_W' plus its own incremental retail price. Such an inefficient solution would be unsustainable in a normal market, because consumers—including those who are being paid by the ITP to reduce their demand—would arbitrage the difference between the price P_W' paid to sellers and the price P_R' paid by buyers, making it impossible to maintain the DR subsidy. But the monopoly inherent in the need for centralized control of an integrated electricity system gives the ITP the power to collect the uplift tax from consumers and to subsidize DR by paying twice for it, so such an inefficient solution can be maintained indefinitely.

Clearly, this way of treating DR as a “resource”—which seems reasonable on the surface and is advocated by some serious people—does not reach an efficient, market-clearing solution. This approach results in more than the efficient amount of DR and creates a demand for some otherwise unneeded services, so it will have its supporters. But it does not accomplish the objective of integrating DR into ITP markets in a way that treats demand and supply equivalently to achieve an efficient mix of supply resources and DR.

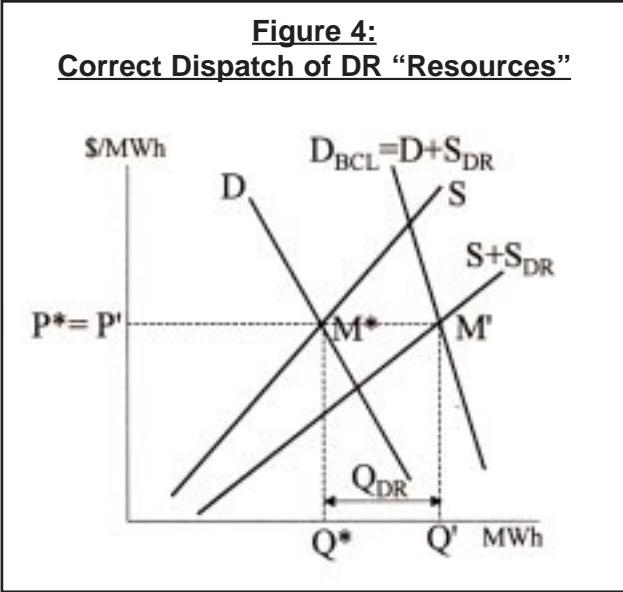
3.2.3 IF THE CONCEPT OF DR AS AN ENERGY “RESOURCE” IS DEFINED AND USED CORRECTLY, IT BECOMES LARGELY POINTLESS

Treating DR as a “resource” as just described is illogical and inefficient because it pays twice for DR “resources” and recovers the costs of this subsidy through a DR uplift. The ITP can cure the illogic and inefficiency simply by requiring that each ESP own the MWh in its BCL if it consumes or sells them. Ignoring the possibility that ESPs have their own production or bilateral contracts,²² the only way an ESP can get ownership of the MWh in its BCL is to buy those MWh

²² There is no reason not to allow an ESP to provide its BCL with own-production or bilateral contract purchases, and doing so changes none of the conclusions here. Allowing for this possibility would simply complicate the exposition.

in the spot market. Thus, to treat DR correctly as a “resource,” the ITP’s settlement system should charge each ESP the spot price for its BCL. If this is done, there is no need for the ITP to assign and give away BCLs; each ESP can be allowed to choose any BCL it wants, because it will pay for it what it is worth.

When BCLs are defined in this way, an offer from an ESP to sell any amount Q_{DR} of DR “resource” at any market price implies an offer to buy at that price a BCL equal to its actual consumption plus Q_{DR} , i.e., the amount Q_{DR} must be added horizontally to both the supply curve and the demand curve at that price. When both the supply curve and the demand curve are shifted in this way, a new market-clearing solution is obtained, illustrated by point M' in Figure 4. Compared to M^* , M' has the same price $P'=P^*$, the same dispatch of supply resources to provide the same amount of real energy Q^* , and some essentially arbitrary MWh of DR “resources” Q_{DR} . It is all logically correct—but clearly rather pointless. As discussed next, everything worth doing with DR can be done

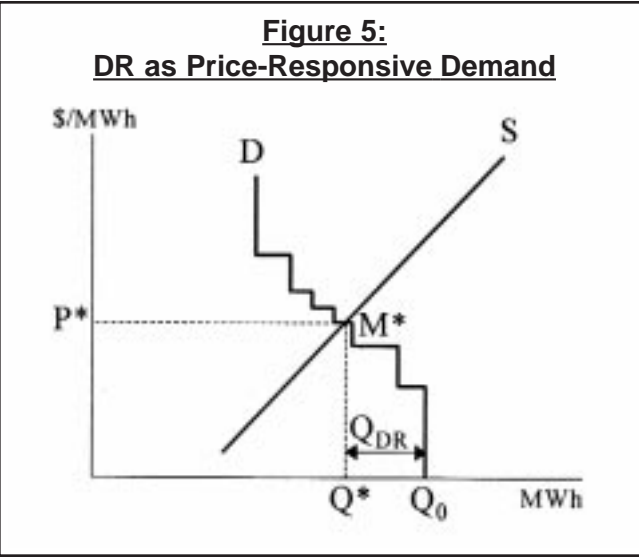


in a more straightforward and less confusing way simply by treating DR as a reduction in demand in response to price.

3.2.4 IT IS MORE NATURAL AND AT LEAST AS EFFICIENT TO TREAT DR AS PRICE-RESPONSIVE DEMAND

Figure 5 illustrates the pricing/dispatch process for an ITP that uses the same basic processes, systems and information as the model DR program of the preceding sections, but treats DR as a reduction in demand in response to price, not as a “resource” that adds to supply. LSEs and ESPs submit to the ITP demand bids indicating how much they expect their consumer-clients to consume at various energy prices —not in response to explicit DR payments from the ITP in addition to the market energy price, but purely in response to the market energy price itself. These demand bids can be expressed in terms of demand reductions relative to some BCLs representing expected consumption at “normal” prices (summing to the quantity Q_0 in Figure 5), as long as everybody understands that this is just a matter of form and that the ITP will not pay anybody for not-consuming-energy.

The ITP uses the demand bids from ESPs to determine the market demand curve D , which is represented as a series of discrete steps in Figure 5 to emphasize that individual demand bids can and do influence the price. The ITP then uses some version of the dispatch/pricing process outlined in section 3.2.1 to find the market-clearing equilibrium M^* at which the supply-side dispatch costs of meeting “fixed” demand Q^* are minimized and the market demand at the resulting price P^* is the same Q^* . The equilibrium M^* is efficient, in the sense that it minimizes the sum of supply-side and demand-side costs given the current state of



technology, information systems, retail pricing arrangements, etc. The decrease in demand below what it would be at low prices can be called the quantity of demand response, Q_{DR} , but this quantity plays no role in anything, and in particular nobody is paid anything for “supplying” Q_{DR} . If somebody can make DR more cost-effective by improving technology, information, retail pricing, etc., they should do so and if and when they do the equilibrium will change and Q_{DR} will increase.

As drawn in Figure 5, the price P^* is the price in one of the demand bids, so the market price is being “set” by that demand bid. But even if the supply curve S intersects the demand curve D on a vertical section, implying that P^* is the price in a supply offer rather than in a demand bid, the accepted demand bids would still have strongly influenced the price by displacing higher-price supply offers. That is how demand affects price in any market.

If Figure 5 represents a forward market, such as a day-ahead or hour-ahead market that results in binding financial commitments, the ITP records the quantities of buy and sell offers that clear in the market at the price P^* , collects the price P^* from the successful buyers and pays the price P^* to the successful sellers (or makes the corresponding debits and credits in the settlement accounts for net-settlement later). No payments from the ITP to anybody for “not-buying” in that market are needed or would be appropriate.

If Figure 5 represents the near-real-time operational market process used for dispatch, the ITP tells the suppliers whose supply offers cleared at the price P^* to run their plants as they said they would at that price and tells the consumers whose demand bids cleared at that price to

reduce their demands because their demand-bids indicated that they want to buy less at this price. If Figure 5 represents the ex post process used for final settlement and pricing, suppliers are paid the price P^* for what they actually delivered in real time, and consumers—including those who reduced their demand in response to the high price—pay the price P^* for what they actually took in real time.²³ The efficient balance between demand and supply is reached and the ITP's settlement books balance with no need for an uplift.

If some ESPs are managing the demand of consumers who buy their energy from separate LSEs, either the LSEs or the ESPs can submit the demand bids and the ITP simply bills the LSE for what the consumers actually take; if a LSE and an ESP agree that they want the ITP to make payments directly to the ESPs, they can enter into a bilateral energy contract that has the desired effect. There is no reason for the ITP to get in the middle of the commercial arrangements between the LSEs and the ESPs and, as discussed below, many good reasons why it should not do so.

This essentially traditional approach to ITP dispatch, pricing and settlements is logical and workable. It uses the same information, processes, and systems necessary to treat DR as a “resource,” and gets the same result if the DR “resource” approach is implemented correctly. The only difference is that this approach is straightforward and easily avoids the mistakes of paying somebody for something it might have bought but didn't, or paying twice for DR, or giving away valuable BCLs. There is no good reason not to do the straightforward and efficient thing: treat DR as a reduction in demand in response to price, not as a “resource” that adds to supply.

4. INCORPORATING DEMAND RESPONSE IN ITP MARKETS

Competitive electricity markets will work much more smoothly and efficiently if DR is appropriately and effectively incorporated into ITP dispatch/pricing processes. This section begins by reviewing some principles of ITP-based markets that are most relevant for DR programs, then outlines an ITP DR program designed on the principles discussed in this paper, and finally discusses how market participants should respond to such a DR program.

4.1 THE DIVISION OF RESPONSIBILITIES IN ITP-BASED MARKETS

Logic and experience demonstrate, and FERC's SMD for ITPs confirms, that there are some things the ITP must do, some things it may do, and some things it should not do. The appropriate division of responsibilities should be maintained in designing and implementing DR policies and programs.

²³ This discussion ignores the many complexities involved in determining forecast dispatches and prices and then determining ex post settlement quantities and prices reflecting what actually happened, given that forecasts are always wrong and dispatch instructions are never followed precisely. Such complexities do not change the basic principles or processes discussed here.

4.1.1 THE ITP'S JOB IS TO OPERATE A WHOLESALE DISPATCH/PRICING PROCESS THAT FULLY AND EFFICIENTLY INCORPORATES DR

A successful competitive electricity market requires at its core a centralized dispatch/market operator—here called an ITP—whose principal responsibility is to maintain reliable and efficient physical operations using a market-based real-time dispatch/pricing process. The ITP may and in the SMD will also operate day-ahead (and perhaps hour-ahead) markets, and should use market arrangements for assuring adequate operating reserves and ancillary services.

A corollary of the ITP's primary responsibility to operate an integrated dispatch/pricing process is that the ITP must provide the procedures, systems, and information needed to incorporate DR appropriately into that process at all stages. The ITP can also provide services that market participants may value (and pay for) such as forecasting and technical advice. But any such services must not interfere with the ITP's primary responsibilities or disrupt normal commercial arrangements among market participants.

Just as important as what the ITP should and might do are the things it should not do. It is not the ITP's job to tell market participants how they should manage their own operations in response to market prices, or whether or how they should manage their risks. The ITP should not be a party to contracts intended to manage energy market risks for market participants.²⁴ The ITP should not try to interpret, enforce, or override state retail regulations. And the ITP should not distort the markets it operates by trying to make market prices lower or less volatile than the actual costs these prices must reflect.

Most importantly for the purposes here, it is not the ITP's job to get involved in complex commercial matters such as defining the rights and obligations that various market participants may or may not have vis-à-vis one another under the many and diverse tariffs and commercial contracts. Getting the ITP involved in these matters will divert the ITP from its main role, create conflict and uncertainty, and erode normal commercial relationships among market participants. It will also inevitably lead to the socialization of large costs, because the easiest way for the ITP to settle the inevitable disputes with individual market participants is to buy them off with consumers' uplift money.

²⁴ The ITP must be a party to the congestion revenue rights (CRRs) called for in the FERC's SMD—also called financial/fixed transmission rights (FTRs), transmission congestion contracts (TCCs), and other things. Also, the ITP may buy ancillary services as part of its system operations role and may use contracts for this purpose.

4.1.2 IT IS THE JOB OF LSES AND ESPS—NOT THE ITP—TO MANAGE RETAIL PRICES, PRICE RISKS, AND DEMAND RESPONSE SERVICES

The primary job of LSEs is to sell electricity to consumers at market prices plus a reasonable—ideally, competitively determined—mark-up reflecting the costs of the services they provide, and to offer price and service options that encourage consumers to manage their demand in response to spot prices or let ESPs do so for them. Both LSEs and ESPs should design and market services that make DR more cost-effective for consumers, but LSEs should be the primary interface between consumers and the wholesale markets.

There are many ways a LSE can and should encourage DR by its customers. For example:

- The LSE can offer tariffs and contracts under which it sells defined quantities of energy at fixed prices and then buys and sells incremental energy at the wholesale price, as closely as this can be determined given the consumer's metering; such an arrangement protects consumers against most of the financial effect of spot price fluctuations but gives them incentives to respond to market prices on the margin.
- An LSE that must sell at a fixed tariff price can offer to share with consumers or ESPs the difference between the spot and tariff prices on demand reductions (somehow defined and verified), and can negotiate similar sharing arrangements with FRC suppliers who will gain if demand is reduced when spot prices are high.
- If consumers want a fixed-price, no-hassle supply but it is cost-effective (e.g.) to shut off or cycle water heaters or other equipment when the spot price is high, the LSE can offer a lower fixed price in exchange for the right to install and use a remotely controlled switch.
- If an ESP unaffiliated with the LSE can do any of this more cost-effectively than the LSE can, the LSE can contract with the ESP or encourage consumers to do so.

None of the activities described above is natural or appropriate for the ITP. The ITP should not sell energy to anybody on any terms other than spot, and hence should never—except perhaps when spot prices are capped, as discussed in section 4.2.4—have any commercial or other reason to pay anybody not to consume energy. The ITP is a natural monopoly that should stick to its critical but limited job of operating the central dispatch/spot market process. It is up to LSEs and ESPs to make DR more cost-effective for each of thousands or millions of consumers.

The interactions between the ITP, LSEs, and ESPs may need to change in order to make DR more cost-effective and to integrate DR into the ITP's dispatch/pricing processes. The ITP's systems will have to accept and use more information from, and will have to provide more real-time information to, LSEs, ESPs, and consumers directly. It may be appropriate for the ITP to deal directly with ESPs who deal directly with consumers who buy energy from separate LSEs. But such changes in ITP systems and processes should be designed and implemented so that they do not interfere with the ITP's principal responsibility and do not alter the basic division of responsibilities among the ITP and market participants.

4.1.3 IT IS THE JOB OF REGULATORS—NOT THE ITP—TO ASSURE THAT REGULATIONS PROVIDE THE RIGHT DR INCENTIVES

Many of the impediments to improving DR are the result of inefficient regulation of retail tariffs and LSEs, particularly at the state level. The most obvious example is the widespread use of retail tariffs that allow consumers to buy as much or as little as they please at fixed prices that may bear little or no relationship to current market prices. The problem here is often attributed to the fact that most small consumers do not have the TOU meters necessary to charge hourly spot prices for incremental hourly consumption, but the more fundamental and difficult-to-solve problem is the political opposition to exposing small—or sometimes even large—consumers to spot prices even on incremental consumption.

Retail tariffs that insulate consumers from wholesale prices even on the margin give consumers poor DR incentives, but should give the LSEs the incentive to encourage DR when the wholesale price is higher than the retail tariff prices. But some forms of regulation can eliminate the LSE's DR incentive. For example, if LSEs are allowed to charge a tariff price this month (or other billing period) that is expected to recover the costs of last month's consumption at last month's spot prices, even the LSEs have little incentive to encourage DR by tariff consumers. The LSE incentives under such an arrangement can be greatly improved by allowing LSEs to recover each month the cost of last month's (or, much better, this month's) consumption plus somehow-verified DR valued at last (this) month's spot market prices. But correcting the incentive problems caused by regulation is the responsibility of regulators, not the ITP. The ITP and its markets will drown in controversy and bureaucracy if the ITP tries to use ad hoc procedures in the wholesale market to correct the many inefficiencies of retail regulation.

4.2 AN ECONOMICALLY EFFICIENT ITP DR PROGRAM

ITPs can and should incorporate DR into their wholesale market processes using normal economic, commercial, and market concepts and processes, as outlined in this section.

4.2.1 THE ITP SHOULD ACCEPT AND USE DEMAND BIDS IN CLEARING ITS DAY-AHEAD (AND ANY OTHER) FORWARD MARKETS

The ITP should allow and encourage demand to bid to buy in the ITP's DAM and any other forward markets²⁵ just as suppliers offer to sell in that market, and should use all of these bids and offers to determine day-ahead market-clearing prices and quantities. As discussed at length above, the most logical and straightforward way for the ITP to treat DR in the DAM (and other markets) is to view a reduction in demand as a reduction in demand, not as a "resource," and certainly not to pay for DR "resources" in energy markets.

There is no fundamental reason that the ITP needs to operate a DAM²⁶ or that net purchases in an ITP's DAM should reflect expected actual real-time demand. Market participants could manage their risks with long-run contracts that specify delivery/settlement in real time, and if they did so any day-ahead trading would be for incremental or speculative quantities only. But under FERC's SMD, an ITP will operate a DAM, will define its CRRs to clear against day-ahead prices (DAPs), and may allocate some ancillary service costs based on the difference between real-time purchases and day-ahead purchases. Because such design features create strong incentives for net sales and purchases in the DAM to reflect expected real-time operations, the discussion here assumes this will be the case. Most energy will be traded under longer-term contracts, but these will be "rolled over" in the DAM, i.e., any longer-term contracts will be settled against DAPs and then expected real-time quantities will be bought and sold in the DAM.²⁷

There is no good reason for the ITP to pay anybody for not buying anything in the DAM. Efficient wholesale market prices and reasonable regulation of retail prices and LSEs will give all market participants the right incentives to encourage or implement DR and to bid in the DAM in ways that reflect their expected DR. If the demand bids do not indicate as much price-responsiveness as some analysts think there should be, the analysts are underestimating either the inherent difficulties and costs of or the regulatory disincentives for short-term DR. Either way, the ITP should not try to overcome these "obstacles" to DR by subsidizing DR with uplift money or by trying to make difficult judgments that are too hard for those closer to the problem. Doing so would do nothing to make DR more cost-effective, but would simply stimulate non-cost-effective DR actions that increase costs and prices for consumers.

If an ESP needs a guaranteed price at the day-ahead stage before it will implement DR, it can simply sell its expected DR in the DAM at the DAP. For example, if 1 MWh of DR costs

25 The implications of, say, an hour-ahead ITP market are not discussed here but should be clear from the discussion here.

26 The electricity markets in New Zealand, Australia, Argentina, and elsewhere operate successfully without an ISO-operated DAM. The main reason for an ITP-operated DAM is to give system operators more day-ahead information about and control over the system.

27 For example, if a LSE has a long-term contract with a generator for 100 MWh, at the day-ahead stage the generator will pay the LSE the DAM price for 100 MWh and the LSE will buy at the same price 100 MWh in the DAM for delivery in real time. The net effect is 100 MWh delivered in real time at the long-term contract price.

\$250/MWh, the ESP can offer to sell 1 MWh in the DAM at any price above \$250/MWh. Then, whenever the DAP is higher than \$250/MWh the ESP will collect that price in the DAM, implement its DR and “deliver” on its day-ahead contract by paying the real-time price (RTP), which is presumably about what it is being paid for reducing real-time demand.

The market-clearing prices in the DAM are just the prices at which the quantities demanded equal the quantities supplied, determined in the simplest case as illustrated in Figure 5.²⁸ The ITP records the concluded contracts for use in real-time settlements, collects (debits) the market-clearing price from the successful buyers and pays (credits) the price to the successful day-ahead sellers. The outcome is efficient and stable with no need for the ITP to do anything except operate the market.

4.2.2 THE ITP SHOULD USE DEMAND BIDS IN REAL-TIME DISPATCH AND PRICING—AS DEMAND REDUCTIONS, NOT “RESOURCES”

Any consumer who wants to “lock in” a price in the DAM and then ignore RTPs should and presumably will be free to do so. But some demand can respond to expected prices or ITP dispatch signals in near-real-time as well as some supply can and should be encouraged to do so.²⁹ In fact, demand bids in the RTM are arguably more valuable than demand bids in the DAM for reducing price spikes and improving reliability.

The ITP should provide market rules, processes, and communication systems that make very-short-run DR more cost-effective and integrate it fully into the ITP’s real-time dispatch/pricing processes. This process should accept demand bids indicating how much a LSE’s customers will take at different RTPs just as it accepts supply offers indicating how much a generator will deliver at different RTPs. Such bids and offers can take the form of offers to deviate from some baseline or expected levels at various RTPs, but this is purely a matter of form; in the end, generators should be paid/credited for the full amount they deliver and consumers should pay/be debited for the full amount they take, net of any day-ahead contracts. As discussed above, the ITP can find the efficient market-clearing dispatch and prices using the traditional concept of minimizing the supply cost of meeting the demand expected at the prices implied by the least-cost dispatch.³⁰

28 In more complex cases the DAM will include constraints that require a more complex market-clearing process. A day-ahead LMP market will include essentially the same complex system security constraints used for real-time dispatch and hence may have hundreds or thousands of potentially different LMPs. CRRs/FTRs/TCCs should be available to hedge the locational differences.

29 Ancillary services are usually bought, paid and scheduled at the day-ahead stage (or sooner), so any DR providing ancillary services must usually be sold and committed then.

30 In some cases the ITP may have to “trick” old software by treating demand bids as “negative supply” and adding them to supply rather than subtracting them from demand. As discussed in section 3.2.3, this can be done correctly if the same demand bids are added to projected or metered real demands and if all actual consumption pays the market price.

There is no reason for the ITP to pay anybody for not buying in the RTM (unless price caps are binding, as discussed in the next section), beyond the amounts automatically provided by the logical settlement of day-ahead contracts. Anybody who consumes in real time less than they bought in the DAM will be automatically credited for the RTP multiplied by the amount sold back. Anybody who did not buy in the DAM but in real time is persuaded by a high RTP to reduce its demand will save the high RTP on what it does not consume; this is all the “payment” anybody deserves or would get in a more normal, decentralized market.

The ITP should treat failure of a consumer to reduce demand as promised in its demand bid essentially the same way it treats failure of a generator to produce as promised in its supply offer. In either case, if failure to respond to ITP instructions causes no operational problem, there is no need for any penalty other than that implied by the RTP; a generator who fails to deliver simply loses the RTP and a consumer who fails to reduce demand simply pays the RTP. Large consumers, like large generators, may be required to inform the ITP if they are unable to perform as promised and may be subject to penalties if failure to perform causes system problems or costs not reflected in RTPs. Fundamentally, however, both consumers and generators are responding to RTPs that reflect their responses, and participate actively in the ITP’s dispatch/pricing process only because this improves two-way communications and the efficiency of the market-clearing process.

4.2.3 THE ITP SHOULD NOT INTERPRET, ADMINISTER, OR OVERRIDE RETAIL TARIFFS OR CONTRACTS AMONG MARKET PARTICIPANTS

In some proposed or operating ITP DR programs, the ITP bypasses LSEs and gives DR incentives directly to final consumers or to ESPs who deal directly with consumers. To the extent that there is an economic rationale for such programs, it is that regulatory and commercial realities give LSEs and/or their FRC suppliers too little (or negative) incentives to help their customers or ESPs implement cost-effective DR, so the ITP should cut the LSE/FRC middlemen out of the loop. The ITP should interpret tariffs and contracts to decide who has what rights and obligations with respect to whom at the retail level and should then provide the implied incentives and payments at the wholesale level.

At the most abstract level—i.e., if the ITP could know and control everything “perfectly”—this could work. For example, in some DR proposals/programs the ITP would set a BCL for each consumer, pay ESPs the difference between the wholesale and incremental retail prices for each consumer, and then recover its costs from the LSEs or FRC suppliers responsible for supplying those consumers at the tariff/FRC prices. In principle, if the ITP does all this “correctly,” everybody will have the right incentives to encourage and implement cost-effective DR. In practice, however, the ITP cannot possibly know or do everything it would need to know and do to do this correctly, so the results are anything but efficient or fair.

In practice, the ITP cannot know who has what contractual rights and obligations in each of hundreds of tariffs and contracts or what LSEs, ESPs, and consumers themselves are doing on their own to improve DR, so the setting of BCLs and buy-back prices is even more arbitrary, controversial, and distorting when done by the ITP than when done by LSEs/FRC suppliers themselves. In practice, ESPs and consumers push for high BCLs and payment of the full spot price for DR, while LSEs/FRC suppliers resist paying anything for the BCLs assigned to their customers. In practice, many players have incentives to exploit the confusion about DR “resources” and subsidies to advance their interests. So in practice, the ITP and/or its regulators often take the path of least resistance by setting high BCLs, paying the full wholesale spot price for DR—even to consumers who are paying spot prices for incremental consumption and hence should get no additional DR payment—and burying all the costs in an uplift that is ultimately paid by consumers.³¹

In principle, it may be possible to identify some specific flaw in retail tariffs, LSE regulation, or commercial contracts that could be fixed with carefully designed and focused ITP programs. But in practice, starting down this slippery slope diverts the ITP from its primary wholesale market functions, reduces the efficiency and predictability of commercial arrangements, socializes costs, and seriously distorts the entire market. Any ITP involvement in retail regulatory and commercial matters should be strongly resisted and, if irresistible, should be allowed only for truly compelling reasons and within very strict limits.

4.2.4 IF PRICE CAPS RESULT IN “OUT-OF-MARKET” PAYMENTS, DR SHOULD BE TREATED JUST LIKE SUPPLY

Price caps are never the best solution to any real or perceived problem, but they may be the only available “second best” solution to some problems and in any case appear to be a fact of political life. When price caps are binding in the ITP spot markets, demand at the capped price exceeds the supply available at that price, leaving the ITP with only two basic options for closing the gap: (1) use arbitrary, non-price methods to curtail demand and (try³²) to compel additional supply; or (2) make out-of-market (OOM) payments to induce some producers to produce more and/or some consumers to consume less. The second of these two basic options is more market-oriented and hence is preferred, but is easy to do incorrectly and even if done correctly cannot solve the serious problems caused by price caps.

31 For example, designers of PJM’s pilot economic load reduction program started with the correct concept that any payment for DR should be limited to the difference between the wholesale LMP and the retail price (with the complication that some fixed costs are recovered through \$/MWh payments). But the proposal to subtract the retail price from LMP was compromised and may be (or may already have been) abandoned in response to political pressure to pay the full LMP.

32 Compelling supply is logically equivalent to curtailing demand but in practice is much harder, at least in real time. The ITP can always pull a switch to cut off some real-time demand, but has no equivalent way to guarantee that “compelled” supply will actually be produced.

If demand may exceed supply in an ITP market at the capped price P_{CAP} , the supply offers and demand bids submitted to that market should indicate how much will be supplied and demanded at market prices up to and including P_{CAP} , and should also indicate how much more will be supplied and how much less will be demanded relative to the P_{CAP} levels in response to various levels of the OOM price (P_{OOM}). When the price cap is binding, the ITP dispatches all in-market supply offered and all demand bid at the price P_{CAP} and then selects a mix of OOM supply and OOM DR offers that closes the remaining gap between demand and supply at least OOM-bid-cost. All actual demand (supply) pays (is paid) the “market” price P_{CAP} and all accepted OOM supply and DR offers are paid the price OOM price P_{OOM} that “clears the OOM market.”³³

In concept—i.e., ignoring the effects of the distorted incentives discussed below—this OOM solution and payment process can produce the efficient outcome, but may be challenged because it appears that an incremental MWh of supply is paid the sum $P_{CAP}+P_{OOM}$ but an incremental MWh of DR is paid only P_{OOM} . As explained at length above, this appearance is incorrect. Both generators and LSE/ESPs are effectively paid the same price $P_{CAP}+P_{OOM}$ for OOM MWh, but because the LSE/ESP does not actually produce anything it must buy its OOM MWh at the price P_{CAP} before it has anything to sell, and hence is left with an apparent net payment of only P_{OOM} . But confusion caused by calling DR a “resource” and the misguided belief that consumers can benefit by subsidizing DR can produce strong political pressure in ITP, stakeholder, and regulatory processes to pay the full market price for DR without requiring that DR providers buy a MWh before they can sell it back. The result of such pressure may be illogical and inefficient OOM programs that subsidize DR.

Even if an OOM process is designed and implemented logically as outlined above, it will create seriously distorted incentives in practice. When the market price cap is likely to be reached, market participants will have incentives to understate the supply available and overstate the demand expected at the price cap in order to shift supplies and DR into the higher-priced OOM market, which will increase the frequency and level of “crises” and OOM payments. OOM payments may be the best way in principle for the ITP to deal with price caps, but in practice an OOM process may be badly designed and is easily manipulated. There is almost always a better solution to any problem than price caps.³⁴

33 For example, suppose P_{CAP} is \$1,000/MWh and that at this price market supply is 1,200 MWh and market demand is 1,400 MWh, leaving a demand-supply gap of 200 MWh. Suppose further that 150 MWh of OOM supply and 50 MWh of OOM DR are offered at OOM prices up to and including \$300/MWh. In this case, the ITP would accept the 1,200 MWh of market supply and the 150 MWh of OOM supply to meet the 1,400 MWh of market demand minus the 50 MWh of OOM DR. The ITP would then pay the 1,350 MWh of total physical supply and charge the 1,350 MWh of total physical demand the \$1,000/MWh capped price, and in addition would pay the 150 MWh of OOM supply and 50 MWh of OOM DR the \$300/MWh OOM price, recovering its OOM costs with an uplift on the price P_{CAP} .

34 For example, if generator market power is a problem, bid caps or bid mitigation may create fewer problems than price caps.

4.2.5 DR SHOULD BE A RESOURCE IN ANCILLARY SERVICE AND ICAP MARKETS JUST AS SUPPLY RESOURCES ARE

The analysis in this paper deals almost entirely with DR in the energy market, with only parenthetical references to the role of DR in ancillary service (including reserve) and installed/available capacity/capability (ICAP) markets. This is because the most serious confusion and errors surrounding DR come from treating DR as a “resource” in the energy market, and in particular from the incorrect idea that the energy market should pay consumers for not-buying-energy. When DR is providing ancillary services it is providing “reliability services and benefits to the grid as a whole,”³⁵ not demand reductions that affect the dispatch and pricing of energy, and paying somebody to provide specific reliability services is very different from paying somebody to not-buy energy. Furthermore, ancillary services are sold to the ITP or to LSEs who are required by the ITP to buy them, not to market participants who are free to buy them or not, and their costs are usually socialized on the grounds that they benefit all system users. These differences make it appropriate to regard DR as a resource that can provide and should be paid for providing ancillary services and ICAP just as supply (or, in some cases, transmission) resources are.

Whatever the ITP pays generators for non-energy services, it should also pay to consumers who provide equivalent services by managing their demands. For example, the ITP pays some “spinning reserve” generators to be ready to increase their output suddenly if a large generator or transmission facility fails, and then pays them for any energy they produce when they respond to ITP instructions. Large consumers who can reduce their demand quickly—sometimes much faster than any generator can ramp up—should be (and often already are) paid to be ready to do so, and when they reduce their demand will not pay for the energy they do not consume; the DR resources are paid for being ready to modify the system supply-demand balance quickly, not for “not buying energy.” Such treatment of DR as an ancillary service resource is a matter of basic fairness and efficiency.

In ITP markets with ICAP requirements, the most natural and efficient way to incorporate DR in the ICAP process would be for an LSE to use DR to reduce the peak demand that determines the LSE’s ICAP requirement. But to get credit for ICAP in this way, the DR resource would have to be available and used when needed, which is more than ICAP generation resources are required to do in most ICAP markets.³⁶ Both supply and DR resources should be paid the same for the same ICAP service, however that is defined and measured.

A DR resource paid for ICAP should get no payment for “not consuming,” but should simply not pay for the energy it does not take; this is analogous to paying generators for providing ICAP and then paying them the market price for any energy they generate. However, care must be

35 Eric Hirst and Richard Cowart, “Demand Side Resources and Reliability,” Framing Paper #2 of the New England Demand Response Initiative, March 20, 2002, p. 3.

36 In many/most ICAP markets, a generation unit can sell ICAP if it passed a performance test some time in the past, even if it is unavailable when more capacity is really needed and valuable.

taken to assure that a LSE does not use the same DR twice for ICAP, once to reduce the peak demand that determines its ICAP requirement and then again as a source of ICAP for meeting that requirement.

4.3 OPERATING IN AN EFFICIENT ITP MARKET

Once DR is integrated fully and efficiently into the ITP's dispatch/pricing process as outlined above, market participants should and will respond to the improved DR incentives and opportunities by reducing demand when and to the extent that it is cost-effective for them to do so. This section outlines some of the desirable and likely responses.

4.3.1 DR SHOULD BE BASED AS FAR AS PRACTICAL ON REAL-TIME, NOT DAY-AHEAD, CONDITIONS AND PRICES

It is often said that “demand can best respond by participating in the day-ahead market,”³⁷ and this is no doubt true in the sense that most consumers—just like most generators—would prefer to know the actual value of their DR far in advance. But the actual cost of meeting demand, and hence the actual value of reducing demand, depends on what actually happens in real time, which cannot be truly known by anybody in advance. If DR (or generation) is guaranteed a price at which it can buy whatever it chooses in real time, that guaranteed price will often turn out to be different from the actual RTP, and decisions made on the basis of the advance price will often be wrong. Furthermore, many price spikes and reliability problems in ITP markets arise when real-time conditions differ from what was anticipated at the day-ahead point. Thus, while any consumer or ESP should be free to lock in a price in the DAM and then ignore later developments if that is the way it must or chooses to operate, there is real value in DR that can be turned off or on in response to changes in expected RTPs after the DAM has closed.

If the DR decision must be made at or about the day-ahead point it should be based on the DAP if that is the best estimate of the expected RTP available at that time. But even an optimal decision made at the day-ahead point will turn out to be “wrong” in retrospect much of the time, in the sense that a different DR decision would have been made if the actual RTP could have and had been known when the decision was made.³⁸ Consumers and any LSEs/ESPs who manage demand for them should wait as long as practical to make their DR decisions and should then make their decision based on the expected value of the RTP at that time, no matter what their day-ahead or longer-term contract position.³⁹

37 FERC SMD Working Paper, p. 13.

38 For example, if DR that costs \$250/MWh is implemented because the DAP is \$300/MWh and then the RTP turns out to be \$200/MWh, somebody has wasted \$50/MWh by implementing too-costly DR. Conversely, if the \$250/MWh DR is not implemented because the DAP is only \$200/MWh and then the RTP turns out to be \$300/MWh, somebody has wasted \$50/MWh by consuming energy that was not worth as much as it cost to produce.

39 This is a simplified statement of a complex decision process. The timing of DR decisions should take into account the fact that delaying a DR decision increases the costs of implementing DR but may provide more information about how much (if any) DR is worth implementing.

Anything that gives consumers better information in advance about expected RTPs or reduces the costs and risks of waiting longer to make DR decisions will improve the cost-effectiveness and value of DR for consumers and the system. Conversely, anything that encourages consumers to make DR decisions earlier will reduce the ability of DR to help with very-short-run problems. For example, if a DAM encourages consumers to make their decisions based on the DAP rather than wait for later information, the DAM may result in more day-ahead DR but less very-short-run DR.

The possibility that a DAM may reduce very-short-run DR is not fanciful. Consumers often say that they want predictable prices and contracts so that they do not have to watch the short-run markets so closely. Some consumers who could wait until, say, four hours in advance to make their DR decisions may respond to introduction of a DAM by making their decisions at the day-ahead stage, buying what they need in the DAM, and then ignoring later price developments even if they could respond to them. This effect of a DAM is not necessarily inefficient—staying awake all night to watch prices is a real cost just like others—but it does not increase the very-short-run DR that may be most valuable.

Conversely, the existence of a DAM may induce more day-ahead DR from risk-averse consumers who would otherwise make their DR decisions even earlier based on longer-term contract prices. For example, consider an industrial energy manager who can reduce real-time demand by 1 MWh at a cost of \$250 and must make its decision at the day-ahead stage. In the absence of a DAM, there may be no clear estimate of the expected RTP at the day-ahead point and no way to hedge a DR decision at that point, so the energy manager may simply contract for long-term energy at a price below \$250/MWh and ignore the potential DR. But if there is a DAM, the energy manager can offer to sell 1 MWh in the DAM at any price above \$250/MWh, and when this offer clears in the DAM can implement its DR and book the difference between \$250 and the long-term contract price as a sure profit (as long as the DR actually happens in real time). In this case, the DAM would increase DR at the day-ahead stage.

Consumers who can cost-effectively change their demand on very short notice—e.g., an hour or less—should do so, either by actively participating in the ITP's RTM (and hour-ahead market if there is one) or simply by getting the best information about expected RTPs as real time approaches and then making their consumption decisions as late as they can.⁴⁰

4.3.2 PURCHASES IN THE DAY-AHEAD MARKET SHOULD REFLECT EXPECTED REAL-TIME CONSUMPTION INCLUDING THE EXPECTED EFFECTS OF ANY DR

Under the conditions assumed in this discussion—i.e., demand-bids accepted and used in the

⁴⁰ In concept the only difference between actively participating in the RTM and simply responding to expected RTPs as late as possible is that active participation gives both the consumer and the ITP better last-minute-time information. In practice, a consumer who actively participates may have some rights and obligations that a passive market forecaster/responder may not have, such as the right to compensation for following ITP dispatch instructions that turn out to be unprofitable once actual RTPs are known.

DAM and RTM, all long-term contracts rolled-over in the DAM, and risk-averse market participants who view the DAP as an unbiased estimate of the expected RTP—total net purchases in the DAM should reflect expected real-time consumption including the expected effect of DR. How these total net purchases should be allocated among various market participants depends on the details of tariffs, contracts and DR programs, which can be very complex. In simple terms, however, each market participant should buy in the DAM what it expects to resell or consume at any price that does not depend on RTPs. The principal categories of participants in the DAM, and what they should and can be expected to do, are summarized below.

- **Load-Serving Entities (LSEs):** LSEs should buy in the DAM the quantity they expect to sell in real time at tariff or contract prices that are not based on actual RTPs, taking into account the effects of DR on expected sales at such tariff/contract prices. For example, if the LSE sells to some consumers at the DAM price, the LSE should buy in the DAM what it expects those consumers to take given the DAP, taking into account any DR that the LSE or ESPs may implement if the expected RTP increases above the DAP in time to implement DR before real time. The LSE should not buy (or sell) in the DAM any quantities that it expects to sell (or buy) at RTPs, e.g., any total or incremental quantities that consumers will buy from the LSE at RTPs.
- **Large Consumers Serving as Their Own LSEs.** A large consumer serving as its own LSE will roll over any longer-term contracts in the DAM and hence should buy in the DAM what it actually expects to consume in real time. Its expected consumption should include the expected effects of any DR that it knows at the day-ahead stage it will implement plus any DR that it may decide to implement later if conditions change. For example, if the consumer has a base consumption level of 10 MWh but will decide at the hour-ahead stage to implement 2 MWh of DR if the expected RTP then exceeds \$250/MWh, and the probability of that is 0.5 given the DAP, the consumer should buy 9 MWh ($10 \text{ MWh} - 0.5 \times 2 \text{ MWh}$) in the DAM.⁴¹
- **Consumers Who Buy/Sell Incremental Quantities at RTPs:** If a consumer is buying some fixed amount at a contract price and will buy (sell, if negative) at RTPs any difference between its real-time consumption and its contract quantity, it should sell (buy) in the DAM the incremental amount it expects to buy (sell) in real time, taking into account the expected value of DR it knows it will or might implement.
- **ESPs Who Are Paid RTPs for DR:** If an ESP has a contract under which it is paid the RTP for energy it saves by implementing DR, it should sell in the DAM the quantity of DR it expects to implement. This would be the quantity expected from any DR committed at the day-ahead stage plus the expected value of DR that might or might not be committed later, depending on developments after the DAM closes.

41 If the probability that expected RTP exceeds \$250/MWh at the hour-ahead stage is higher at higher DAPs, as it probably is, the consumer should bid in the DAM to buy less at higher DAPs.

- **Generators:** Generators should sell in the DAM what they expect, in a probabilistic sense, to generate in real time, including the effects of dispatch decisions and forced outages. Any longer-term contracts a generator has will be rolled over in the DAM, so the generator will effectively collect (or pay) the DAP only on its expected real-time generation net of its longer-term contract sales.
- **Marketers and Speculators:** Marketers will clear all their contract positions in the DAM. If, contrary to the assumptions here, a predictable difference between the day-ahead and RTPs develops, speculators (including LSEs, marketers and others acting as speculators) will buy or sell in the day-ahead the market whatever they choose in order to take advantage of—and, in the process, reduce—the difference.

The most important point to note about these brief summaries of how different market participants should operate in the DAM is that there is no need for any concept of DR “resources” and no reason for the ITP ever to pay anybody for not buying in the DAM. Normal commercial considerations and interactions provide the right incentives and compensation for DR, viewed simply as a reduction in demand in response to price.

5. CONCLUSIONS

Improving short-run DR can yield significant benefits by reducing the total supply-side and demand-side costs of meeting demand during critical periods. It may also reduce the price spikes in ITP markets that are needed to maintain reliable operations and that are often increased by the exceptional market power that generators can have during such periods, although this cannot be guaranteed as a long-run effect given that short-run supply curves will shift as DR takes effect. But only reductions in the actual costs of meeting demand are real, sustainable, and predictable benefits. DR should be implemented because it can and to the extent that it does lower the total supply-side and demand-side costs of meeting demand, not in an ultimately futile attempt to drive down prices more than supply-side costs and benefit consumers at the expense of producers.

ITPs should assure that demand bids can be submitted to and will be fully incorporated into day-ahead energy markets and real-time energy dispatch/pricing processes, and that DR can provide and will be paid for ancillary services and ICAP in ways that are equivalent to the ways supply resources are used and paid. The most logical and natural way to do this where energy is concerned is to treat a reduction in demand in response to price as exactly what it is: a reduction in demand in response to price. ITP bidding, dispatch, pricing, and settlement processes can handle price-responsive demand quite easily in principle and with relatively small changes in practice.

When reductions in demand are treated as reductions in demand, a consumer who consumes less when the price is high buys less at that high price and gets to keep the money it would otherwise have paid—just as in any other market. If the consumer produced or purchased under contract more than it consumes at a high price, it can sell the difference at the high price. If a consumer buying under a tariff or full-requirements contract is deemed to have a right to buy and resell some baseline quantity at a low tariff price, the LSE or its full-requirements supplier with the obligation to sell at that price can buy back some of its obligation if it is cost-effective to do so, either directly or through an ESP. But there is no need or justification for the ITP to pay anybody for not consuming something—not even if the consumer might have consumed more but didn't.

It is logically possible to treat DR as a “resource.” After all, as Humpty Dumpty said, “When I use a word, it means just what *I* choose it to mean—neither more nor less,”⁴² so if everybody agreed what is meant by a DR “resource” such rhetorical flourishes might be harmless. But references to DR as a “resource” are seldom accompanied by the clear definitions and distinctions needed for clear thinking about such complex matters, and in practice calling DR a “resource” creates confusion that leads to illogical and inefficient DR programs.⁴³ And treating DR as a “resource” in the energy market correctly makes it clear that the whole concept is largely pointless.

DR programs that require the ITP to define the rights and obligations that consumers, LSEs, contract suppliers, and ESPs have vis-à-vis one another under tariffs and contracts and then to collect and make the implied payments among consumers and market participants are inconsistent with the fundamental division of responsibility in ITP-based electricity markets - and with even more basic principles of reasonableness and practicality. The impossibility of doing all this with any degree of fairness or efficiency inevitably results in the ITP and its regulators taking the path of least resistance and simply using uplift money to subsidize as much DR activity as the politics of the situation require or allow. This is not the way to integrate DR fully and efficiently into competitive electricity markets—or, indeed, to encourage the further development of competitive electricity markets at all.

42 Lewis Carroll, *Through the Looking Glass*, Chapter 6.

43 That is a lot to attribute to mere words, but imprecise use of words can create serious problems. Humpty Dumpty went on to say: “When I make a word do a lot of work like that, I always pay it extra.” *Ibid.* By Humpty’s standard, the term “DR resources” deserves a large extra payment—even if DR “resources” themselves do not.

ECONOMIC PRINCIPLES OF DEMAND RESPONSE IN ELECTRICITY

Appendix: A Numerical Example

Consider a simple electricity system that has a constant non-peak demand in 90 percent of the hours and a constant peak demand in 10 percent of the hours, and has two kinds of generation, non-peakers and peakers. The variables used to describe the situation in scenario S are:

Non-Peak Demand	=	NQ_S
Peak Demand	=	PQ_S
Non-Peak Price	=	NP_S
Peak Price	=	PP_S
Demand-Weighted Average Price	=	$DWAP_S$
Time-Weighted Average Price	=	$TWAP_S$
Average Hourly Consumer Bill	=	$AHCB_S$

Scenario 0: The Initial Competitive Situation

In this initial scenario, the system is in a long-run competitive equilibrium with demand and prices as follows:

NQ_0	=	100 MW	PQ_0	=	120 MW
NP_0	=	\$50/MWh	PP_0	=	\$200/MWh

These non-peak and peak demands and prices imply that the following variables have the indicated values:

$DWAP_0$	=	$(0.9 \times 100 \times \$50/\text{MWh} + 0.1 \times 120 \times \$200/\text{MWh}) / (0.9 \times 100 + 0.1 \times 120)$
	=	\$67.647/MWh
$TWAP_0$	=	$(0.9 \times \$50/\text{MWh} + 0.1 \times \$200/\text{MWh}) / (0.9 + 0.1)$
	=	\$65/MWh
$AHCB_0$	=	$0.9 \times 100 \times \$50/\text{MWh} + 0.1 \times 120 \times \$200/\text{MWh} / (0.9 + 0.1)$
	=	\$6,900/hour.

The system has:

- 100 MW of non-peakers that run in all hours
- 20 MW of peakers that run in the 10 percent of hours that are peak hours

Because both peakers and non-peakers must be just covering their long-run marginal costs (LRMC_P and LRMC_N, respectively) at the assumed market prices, these LRMCs are:

$$\begin{aligned} \text{LRMC}_P &= PP_0 &= \$200/\text{MWh} \\ \text{LRMC}_N &= \text{TWAP}_0 &= \$65/\text{MWh} \end{aligned}$$

Assume that peakers have a running cost of \$100/MWh. Then, they earn short-run rents of \$100/MWh when they run and sell their energy for \$200/MWh, and this \$100/MWh in rents must cover their total annualized fixed costs even though they run in only ten percent of the hours. Thus, their total “hourized” fixed costs must be \$10/MW/hour, i.e., they would just cover their annual fixed costs if they earned rents of \$10/MW in every hour. (This implies an annualized fixed cost of \$87.6/kilowatt/year for a year with 8,760 hours. The break-down of non-peakers’ LRMC into fixed and variable components is not needed in this example.)

Scenario 1: The Short-Run Competitive Situation

Starting with the system in the competitive equilibrium described above, a sudden increase in peak DR reduces peak demand to 110 MW (PQ₁), creating an excess of peaker capacity and causing the peak price to fall to the incremental cost of peakers, \$100/MW. The new price variables are:

$$\begin{aligned} PP_1 &= \$100/\text{MWh} \\ \text{DWAP}_1 &= (0.9 \times 100 \times \$50/\text{MWh} + 0.1 \times 110 \times \$100/\text{MWh}) / (0.9 \times 100 + 0.1 \times 110) \\ &= \$55.45/\text{MWh}. \\ \text{TWAP}_1 &= (0.9 \times \$50/\text{MWh} + 0.1 \times \$100/\text{MWh}) / (0.9 + 0.1) \\ &= \$55/\text{MWh}. \\ \text{AHCB}_1 &= (0.9 \times 100 \times \$50/\text{MWh} + 0.1 \times 110 \times \$100/\text{MWh}) / (0.9 + 0.1) \\ &= \$5,600/\text{hour}. \end{aligned}$$

The ratios between the short-run and the initial DWAPs and AHCBs are:

$$\text{DWAP}_1/\text{DWAP}_0 = 55.45/67.65 = 0.820 \quad \text{AHCB}_1/\text{AHCB}_0 = 5,600/6,900 = 0.811$$

Thus, DWAP and AHCB have fallen by 18 and 19 percent, respectively.

Scenario 2: The Medium-Run Competitive Situation

In the medium run, peak prices stay at \$100/MWh, but the non-peak price must increase enough to get the TWAP back to the LRMC of non-peakers, LRMC_N = \$65/MWh. This requires

a non-peak price NP_2 of \$61.11.../MWh, i.e.:

$$\begin{aligned} TWAP_2 &= LRMC_N = (0.9 \times \$61.11.../MWh + 0.1 \times \$100/MWh)/(0.9+0.1) \\ &= \$65/MWh \end{aligned}$$

With $PP_2 = \$100/MWh$ and $NP_2 = \$61.11.../MWh$, the demand-weighted average price and the average hourly consumer bill become:

$$\begin{aligned} DWAP_2 &= (0.9 \times 100 \times \$61.11/MWh + 0.1 \times 110 \times \$100/MWh)/(0.9 \times 100 + 0.1 \times 110) \\ &= \$65.346/MWh. \end{aligned}$$

$$\begin{aligned} AHCB_2 &= (0.9 \times 100 \times \$61.11/MWh + 0.1 \times 110 \times \$100/MWh)/(0.9 + 0.1) \\ &= \$6,600/hour. \end{aligned}$$

The ratios between the medium-run and the initial DWAPs and AHCBs are:

$$DWAP_2/DWAP_0 = 65.346/67.65 = 0.966 \quad AHCB_2/AHCB_0 = 5,600/6,900 = 0.9565$$

Thus, DWAP and AHCB have now increased so that their net declines from the initial situation are only 3.4 and 4.4 percent, respectively.

Scenario 3: The Long-Run Competitive Situation

In the long run competitive situation, peak prices go back to \$200/MWh (because the remaining peakers need to recover their fixed costs), and non-peak prices go back to \$50/MWh, but peak demand is now lower, so the price/bill variables take on the following values:

$$\begin{aligned} TWAP_3 &= (0.9 \times \$50/MWh + 0.1 \times \$200/MWh)/(0.9+0.1) \\ &= \$65/MWh \end{aligned}$$

$$\begin{aligned} DWAP_3 &= (0.9 \times 100 \times \$50/MWh + 0.1 \times 110 \times \$200/MWh)/(0.9 \times 100 + 0.1 \times 110) \\ &= \$66.337/MWh. \end{aligned}$$

$$\begin{aligned} AHCB_3 &= (0.9 \times 100 \times \$50/MWh + 0.1 \times 110 \times \$200/MWh)/(0.9 + 0.1) \\ &= \$6,700/hour. \end{aligned}$$

The ratios between the long-run and the initial DWAPs and AHCBs are:

$$DWAP_3/DWAP_0 = 66.337/67.65 = 0.981 \quad AHCB_3/AHCB_0 = 6,700/6,900 = 0.971$$

Thus, DWAP and AHCB have declined 1.9 and 2.9 percent, respectively, compared to the initial situation.

Scenario 4: Long-Run Situation with Peak-Period Market Power

Suppose the LRMC of peakers is only \$150/MWh, but peaker owners have enough peak-period market power to keep peak prices to \$200/MWh indefinitely unless DR increases. If the increase in peak DR totally eliminates peak-period market power, the peak price will fall to \$100/MWh as before and all price variables will remain the same in the short- and medium-run. The only difference is that in the long run the peak price will increase only to \$150/MWh. The new variables in the long-run are:

$$\begin{aligned} NP_4 &= \$55.55\dots/\text{MWh} & PP_4 &= \$150/\text{MWh} \\ TWAP_4 &= (0.9 \times \$55.55\dots/\text{MWh} + 0.1 \times \$150/\text{MWh}) \\ &= \$65/\text{MWh} \\ DWAP_4 &= (0.9 \times 100 \times \$55.55\dots/\text{MWh} + 0.1 \times 110 \times \$150/\text{MWh}) / (0.9 \times 100 + 0.1 \times 110) \\ &= \$65.84/\text{MWh}. \\ AHCB_4 &= (0.9 \times 100 \times \$55.55\dots/\text{MWh} + 0.1 \times 110 \times \$150/\text{MWh}) \\ &= \$6,650/\text{hour}. \end{aligned}$$

The ratios between the long-run and the initial DWAPs and AHCBs are:

$$DWAP_4/DWAP_0 = 65.84/67.65 = 0.973 \quad AHCB_4/AHCB_0 = 6,650/6,900 = 0.964$$

Thus, DWAP and AHCB have fallen 2.7 and 3.6 percent, respectively, relative to the initial situation.

In this case, the \$250/hour or 3.6 percent decline in total consumers' bills is the sum of an actual cost decrease because fewer peakers are needed and a decrease in producers' rents due to the elimination of peak-period market power. The actual cost savings is just the LRMCs of the 10 MW of peakers that are no longer needed, i.e.:

$$\text{Decline in Costs} = 10 \text{ MW} \times \$150/\text{MWh} \times 0.1 = \$150/\text{hour on average}$$

Initially, 20 MW of peakers had been getting \$200/MWh when their actual costs were only \$150/MWh, so the decline in rents when their market power is eliminated is:

$$\text{Decline in Rents} = 20 \text{ MW} \times \$50/\text{MWh} \times 0.1 = \$100/\text{hour on average}$$

That is, of the \$250/hour (3.62%) decline in bills, \$150/hour (2.17%) is a real cost decline and

\$100/hour (1.55%) is a rent transfer.

Scenario 5: The Effects of Broadening or Narrowing the Peak

Initially, peakers were recovering their fixed costs earning \$100 MWh rent (the difference between the initial peak price of \$200/MWh and their assumed running cost of \$100/MWh) in the 10 percent of the hours they ran. If the peak is broadened to 20 percent of the hours (whatever its level), peakers need only half as much rent per peak hour to cover their fixed costs, so:

$$\begin{aligned}\text{Peak Price with constant peak in 20\% of hours} &= \$100/\text{MWh} + 0.5 \times \$100/\text{MWh} \\ &= \$150/\text{MWh}\end{aligned}$$

If the peak is reshaped so that demand is 110 MW in 5 percent of the total hours and 105 MW in the other 5 percent of the peak hours, there will be 5 MW of peakers that run in 10 percent of the hours—call these the “shoulder” hours—and 5 MW that run in the 5 percent of the hours that are now the peak. The prices in the shoulder and peak hours must produce enough rents to allow both sets of (identical, in this example) peakers to cover their total costs. The new peak price, call it PP_5 , must be high enough to recover all the fixed costs of peakers that run only 5 percent of the time, which means the rents per MWh during this period must be twice the \$100/MWh they were during the entire peak initially. Thus,

$$\begin{aligned}PP_5 &= \$100/\text{MWh} + 2 \times \$100/\text{MWh} \\ &= \$300/\text{MWh}.\end{aligned}$$

With this PP_p , the peakers that run in both peak and shoulder hours will cover their fixed costs in the peak hours so the shoulder price, call it SP_5 , will be driven to the \$100/MWh running cost of peakers, i.e.:

$$SP_5 = \$100/\text{MWh}.$$

At these prices, the total revenue during the peak and shoulder hours accruing to non-peakers is the same as it was when the price was \$200/MWh for the whole period, so no change in the non-peak price is necessary to cover the LRMC of non-peakers.



**EDISON ELECTRIC
INSTITUTE**

701 Pennsylvania Avenue, N.W.
Washington, D.C. 20004-2696
202-508-5000
www.eei.org