United States Energy Information Administration

INTERNATIONAL ELECTRIC MARKET MODEL

COMPONENT DESIGN REPORT

Dr. Dale M. Nesbitt
dale.nesbitt@arrowheadeconomics. com

ArrowHead Economics

27121 Adonna Ct.
Los Altos Hills, CA 94022
(650) 218-3069 mobile

www. arrowheadeconomics.com

January 27, 2016

CONTENTS

1 INTRODUCTION 12
1.1 Dimensions and Structure of Regional Wholesale Electricity 14
1.2 Model Temporality 18
1.3 Dimensions and Structure of Regional Retail/Point of Use Power Models 18
1.4 CDR Topics Beyond Original Requirements 21
2 THERMAL ELECTRIC GENERATION AS IMPACTED BY EMISSIONS REGULATION 24
2.1 Plant Sub-Models That Account for Economics and Emissions 24
2.2 Sidebar—Where Do We Get These Emissions Coefficients? 33
2.3 The Economic Meaning of an Unregulated, Uncontrolled Scenario 39
2.4 The Addition of Renewables to the Picture with the Possibility of Tradable Renewable Energy Credits (RECs) 50
2.5 From Where Do Emissions Prices Come? 52
2.5.1 Endogenous Pollutant Pricing Under Cap and Trade Regulation 53
2.5.2 Pollutant Taxes (As Contrasted with Cap and Trade) 66
2.5.3 Completely Uncontrolled Emissions 68
2.5.4 Command and Control Regulation (e.g., BACT, Point Source, CPP) 69
2.6 From Where Do REC Prices Come? 70
3 MODELING THE INTERACTION BETWEEN FUEL SWITCHING, PLANT RETROFITTING, AND NEW CAPACITY ENTRY 76
3.1 Coal-Waste-Wood Generation 76
3.1.1 Fuel Switching and Plant Retrofitting for Coal Plants 76
3.1.2 Fuel Substitution and Discrete "Exemplar" Coal Types 78
3.1.3 Interplay between Retrofit Options and Coal Use 79
3.1.4 Retrofit Nodes—Data and Structure 85
3.1.5 Summary of Concepts Thus Far 88
3.1.6 Fuel Substitution—A Second Coal Type 92
3.1.7 Capacity in Place and Capacity Expansion (for Coal) 94
3.2 Gas Fired Generation (Which Is Directly Extrapolatable to Oil) 99
3.2.1 Plant Retrofitting for Gas CC plants 99
3.2.2 Interplay between Retrofit Options and Gas Use 99
3.2.3 Retrofit Nodes—Data and Structure 102
3.2.4 Summary of Concepts for Gas Thus Far 104
3.2.5 Capacity in Place and Capacity Expansion (for Gas CC) 106
3.2.6 Oil Generation 108
3.2.6.1 Distillate Combustion Turbines 108
3.2.6.2 Residual Oil Based Steam Turbine Generation 108
3.3 Nuclear Generation 110
3.4 Total Emissions Allowances, Coal, Gas, Oil, and Renewables 111
3.5 RECs and Where They Come From 113
3.6 Concluding Comments 115

4 LOAD AND RESERVE MARGIN ... 116
4.1 Model and Methodology ... 117
4.1.1 A Note on Model Time Points (Time Discrimination) ... 118
4.1.2 Retaining the Hour by Hour Chronology (Coincidence) ... 120
4.1.3 Overriding Hour-by-Hour Chronology to Develop Load Duration Curves ... 120
4.1.4 Demand Elasticity ... 134
4.2 Chronology by Hour ... 139
5 SUPPLY-DEMAND BALANCE WITHIN THIS LOAD AND GENERATOR FRAMEWORK ... 140
5.1 A Market Clearing Price at Every Load Tranche ... 141
5.2 Reserve Margin ... 141
5.3 How Will IEMM Represent Such Reserves? ... 142
5.4 The "Dispatch" Mindset Does Not Work for IEMM ... 146
6 PLANT PROFITABILTIY AND CAPACITY ADDITION ... 149
7 RENEWABLES—WIND ... 167
7.1 Renewables Summary, Definition, and Scope ... 167
7.2 Fundamentals of Wind ... 172
8 RENEWABLES—SOLAR ... 181
9 ENFORCED ENTRY OF RENEWABLES TO A RENEWABLES PORTFOLIO STANDARD ... 185
10 TRANSMISSION ... 191
10.1 How Should IEMM Emulate Transmission? ... 191
10.2 Alternative Transmission Tariff and Pricing Structures ... 197
10.3 Transmission Congestion ... 200
11 RESERVOIR HYDRO, RUN OF RIVER HYDRO, AND PUMPED HYDRO FOR STORAGE AND LOAD-SHIFTING ... 204
11.1 Electricity Storage ... 204
11.2 Fundamentals of Electricity Storage at or Near Point of End-use ... 207
11.3 Storage and Load Leveling with Load Duration Curves ... 214
11.4 Fundamentals of Electricity Storage at or Near Irregular Renewables ... 218
11.5 Development of the IEMM Storage Model in the Pumped Hydro Context ... 221
11.6 Pumped Hydro Storage ... 222
11.6.1 Storage Reduces Temporal Swings ... 223
11.6.2 Storage and Generation Capacity ... 226
11.6.3 Transportation/Transmission Deliverability ... 226
11.6.4 Storage Reduces the Peak-Off-peak Price Differential ... 228
11.6.5 Storage of Water Rather than Storage of Electricity ... 228
11.7 Seasonal Hydro Is Important Almost Everywhere in the World (e.g., Ecuador) 230
11.8 The ArrowHead Storage Model ... 230
11.9 Energy Limited Supplies ... 236
11.10 Summary ... 239

12 EMISSIONS AND REC BANKING 241
13 CONSERVATION 244
14 GLOBAL SOCIAL WELFARE MAXIMIZATION (MPEC, COMPLEMENTARITY) CANNOT POSSIBLY WORK 253
15 APPENDIX: ASSEMBLING GENERATION DATA—RAW DATA IDENTIFICATION, GATHERING, AND ASSEMBLY 254

FIGURES

Figure 1: Structure of Regional Wholesale Power Systems 16
Figure 2: Dimensions of Regional Retail Power Model 19
Figure 3: Integrating Fuel Models (Oil and Gas) with North American IEMM Example 22
Figure 4: Each Generator Produces MWh Jointly with a Slate of Pollutants 24
Figure 5: Each Generator Purchases and Surrenders Emissions Allowances Consistent with Its Level of Generation 26
Figure 6: Parameters of a Representative Thermal Plant in the IEMM Model 26
Figure 7: Short-Run Supply Curve for the Plant (Marginal Cost Curve) 27
Figure 8: Heat Rate Curve (Input-Output Curve) for a Thermal Power Plant 28
Figure 9: A Quadratic Heat Rate Curve for a Thermal Power Plant 29
Figure 10: Short Run Marginal Cost Curve for a Plant with a Quadratic Heat Rate 31
Figure 11: Linear Programming Is Based on a Stairstep Supply Curve 32
Figure 12: Nonlinear Supply Curves Require "Zillions" of Tranches to Emulate Because Real World Short-term Supply Curves Are So Decidedly Nonlinear 32
Figure 13: Emissions Parameters for Uncontrolled Plants (lb/MMBtu Fuel Input)—Illustrative 34
Figure 14: SO2 and Coal Accounting--Illustrative 35
Figure 15: NOx and Coal Accounting--Illustrative 38
Figure 16: Mercury and Coal Accounting--Illustrative 38
Figure 17: CO2 and Coal Accounting--Illustrative 38
Figure 18: Environmental Emissions Coefficients and Heat Rates, and VOCs--Illustrative. 39
Figure 19: Fuel and Emissions Prices in an Unregulated, Unconstrained Power System 40
Figure 20: Calculating Environmental Cost 40
Figure 21: Economic-Only Variable Cost 40
Figure 22: Emissions Allowance and Fuel Prices (Illustrative) 41
Figure 23: Fuel and Emissions Prices in a Regulated Power System 41
Figure 24: Environmentally Burdened Variable Costs of Generation Plants without CO2 Sequestration 42
Figure 25: Environmentally Burdened Variable Costs of Generation Plants withCO2 Sequestration 44
Figure 26: Environmentally Burdened Short-Run Supply Curve (Marginal Cost Curve) for the Plant 45
Figure 27: The "Supply Stack" Based on Strictly Economic Costs 45
Figure 28: Environmentally Adjusted Supply Stack 47
Figure 29: The Economic Supply Stack and the Environmentally Adjusted Supply Stack and Their Implications for Market Prices 47
Figure 30: The North American Electric Generation Supply Stack on a Purely Economic Basis 48
Figure 31: The North American Electric Generation Supply Stack on an Environmentally Adjusted Basis 49
Figure 32: Each Generator Purchases and Surrenders Emissions Allowances and a REC Consistent with Its Level of Generation 51

Figure 33: The Power Business Before Environmental Regulation Is Local and Balkanized 53
Figure 34: All Plants Vie for a Fixed Supply of All Emissions Allowances 57
Figure 35: Every Plant Decides How Much to Operate at Its Individual, SO2-Modified Production Cost 58
Figure 36: IEMM Sums All the Emissions Allowances Required by All the Plants Given Their Operations 59
Figure 37: The Aggregate of the Plants Is a "Price IN-Quantity OUT" Model 60
Figure 38: The Plant Model Is a Demand Curve for Emissions Allowances 60
Figure 39: The Supply of Emissions Allowances 61
Figure 40: The IEMM Representation of the Supply of Emissions Allowances 61
Figure 41: The "Cap"—The Number of Emissions Allowances Created and Authorized 62
Figure 42: Various Levels of Proposed CO2 Caps (Pew Research Website) 63
Figure 43: The Economic Context of the Cap and Trade Regulation Model 64
Figure 44: Accumulate All Emission Allowances Cap and Trade Domain 65
Figure 45: Power Prices Depend on Emissions Prices AND Emissions Prices Depend on Power Prices 65
Figure 46: Uncapping at a Given Pollutant Tax Rate Imposes That Pollutant Tax Rate on the Model 67
Figure 47: An Uncontrolled System Has a Zero Pollutant Tax Rate and No Cap 68
Figure 48: Each Generator Must Have Capacity as a Factor If It Is to Generate 71
Figure 49: A Model of REC Trading to Achieve 20 Percent RPS 72
Figure 50: Zero RPS—No Mandated or Subsidized Renewables Entry 74
Figure 51: A Ramp-up or "Grace" Period to Full RPS Implementation 74
Figure 52: A "Green Tax" Levied on Non-Qualifying Generators 75
Figure 53: Combinations of Plant Equipment Installations 77
Figure 54: Six Potential Coal Bins 79
Figure 55: The Sixteen Retrofit-Non Retrofit Options for Coal Type k 79
Figure 56: The Sixteen Retrofit-Non Retrofit Options for Coal Type k with Capacity Limitation 80
Figure 57: The Non-Retrofit Electricity Generation Node (Upper Right in Figure 56) 81
Figure 58: The Non-Retrofit Electricity Generation Node (Upper Right in Figure 56) 81
Figure 59: The Non-Retrofit Electricity Generation Node (Upper Right in Figure 56) with Input Output Coefficients Carefully Enumerated 83
Figure 60: The Material Balance for the Completely Non-Retrofit Node—Bituminous Pulverized Coal Plant 87
Figure 61: The Material Balance for the SO2 Removal Retrofit Node—Bituminous Pulverized Coal Plant 87
Figure 62: Network and Markets for the SO2 Removal Retrofit Node—Bituminous Pulverized Coal Plant 88
Figure 63: Network and Markets for the NOx Removal and SO2 Removal Retrofit Node—Bituminous Pulverized Coal Plant 88
Figure 64: Network and Markets for the NOx removal/SO2 Removal/Mercury Removal/CO2 Sequestration Retrofit Node—Bituminous Pulverized Coal Plant 89
Figure 65: Generation and Retrofit Alternatives for Coal Type k as Represented in IEMM 91
Figure 66: Adding a Second Coal Type 93

Figure 67: Total Coal Generation Capacity—Spans All Coal Nodes in the Model............... 95
Figure 68: SO2 Removal Capacity Price—Spans All Coal Nodes in the Model to Which SO2 Removal Equipment Is Added... 96
Figure 69: NOx Removal Coal Capacity Price—Spans All Coal Nodes in the Model to Which NOx Removal Equipment Is Added.. 97
Figure 70: Mercury Removal Coal Capacity Price—Spans All Coal Nodes in the Model to Which Mercury Removal Equipment Is Added... 97
Figure 71: CO2 Sequestration Coal Capacity Price—Spans All Coal Nodes in the Model .. 98
Figure 72: Combinations of Plant Retrofits.. 99
Figure 73: The Four Retrofit-Non Retrofit Options for Gas CC Plants 100
Figure 74: Capacity in Place for Gas CC 100
Figure 75: The Non-Retrofit Electricity Generation Node.. 101
Figure 76: The Non-Retrofit Gas CC Generation Node (Upper Right in Figure 75)........... 101
Figure 77: The Non-Retrofit Gas CC Generation Node.. 102
Figure 78: The Material Balance for the NOx removal Retrofit Node—Gas CC plant....... 103
Figure 79: Network and Markets for the NOx removal Retrofit Node—Gas CC plant....... 104
Figure 80: Network and Markets for the NOx removal and Sequestration Retrofit Node—Gas CC plant 104
Figure 81: Generation and Retrofit Alternatives for Gas 105
Figure 82: Total Installed Capacity for Gas CC.. 106
Figure 83: NOx Removal Retrofit for Gas CC.. 107
Figure 84: CO2 Removal for Gas CC 108
Figure 85: Combinations of Distillate Combustion Turbine Retrofits................................ 109
Figure 86: Combinations of Plant Equipment Installations for Residual Steam Turbine Generation.. 109
Figure 87: Nuclear Generation Node 110
Figure 88: Total Installed Capacity for Nuclear.. 110
Figure 89: CO2 Emissions Allowances Produced and Consumed...................................... 111
Figure 90: Mercury Emissions Allowances Produced and Consumed 112
Figure 91: NOx Emissions Allowances Produced and Consumed 113
Figure 92: SO2 Emissions Allowances Produced and Consumed...................................... 113
Figure 93: RECs and Renewable MWh Produced and Consumed 114
Figure 94: Demand Variation Is Assembled by Time of Day, Week, Season, and Year..... 117
Figure 95: Time Period Structure of IEMM Allows Non-chronological Time Points......... 119
Figure 96: Time Period Structure of IEMM Allows Fully Chronological Time Points 119
Figure 97: Create Monthly Load Duration Curve by Sorting 121
Figure 98: A Historical Pattern of Load for a Region... 122
Figure 99: Each Region in the Model Has a Continuous Monthly Load Duration Curve ... 122
Figure 100: Stairstep Approximation to Monthly Load Shape 123
Figure 101: Disaggregation of Each Month into Load Tranches.. 126
Figure 102: Representing Future Load Growth.. 129
Figure 103: Forward Growth of Heights of the Load Curve Tranches 130
Figure 104: IEMM Should Represent Price Elasticity in Each Load Period 135
Figure 105: Exponential Lag Function.. 137
Figure 106: We Represent Price Elasticity in Each Load Period.. 138

Figure 107: Inelastic Supply-Demand Equilibria Are Calculated for Each of the Six Load Tranches 140
Figure 108: Inelastic Supply-Demand Equilibria Are Calculated for Each of the Load Tranches Augmented for Reserves at Time of Peak 142
Figure 109: The "Supply Stack" of All Plants in Ascending Order of Cost 146
Figure 110: The "Supply Stack" of All Plants in Ascending Order of Cost with Demand 147
Figure 111: Economic Reality—Price Has No Relationship Whatsoever to the Cost of ANY Plant 148
Figure 112: Plant Profitability Calculation Considers Power, Gas, Emissions, RECs, Fixed, and Variable Costs 150
Figure 113: Conceptual Representation of Asset Value 150
Figure 114: Asset Value Given Actual, Highly Time-Varying Power Prices 151
Figure 115: Levelized Capital Recovery Assumption, Which Is Incorrect 155
Figure 116: Levelized Margin Capture Assumption Fails under Backwardated Product Price 156
Figure 117: Levelized Margin Capture Assumption Fails under Contango Product Price 156
Figure 118: Base through Peak Market Clearing Prices and Quantities in the Absence of Capacity Addition 157
Figure 119: Adding a Small Amount of New Capacity 158
Figure 120: More Capacity K Increases Volume K but Decreases Prices 159
Figure 121: Still More Capacity Addition Would Kill the Peak Price 160
Figure 122: Margin Capture Is a "Laffer Curve" with Regard to Capacity Additions. 161
Figure 123: A Growing Market that Requires Capacity Additions 162
Figure 124: A Little Bit of Entry 163
Figure 125: A Moderate Amount of Capacity Addition 164
Figure 126: Capacity Addition for N-1 of the N Hours 164
Figure 127: Peaky Price Duration Curve 165
Figure 128: History of Renewable Energy in the United States 169
Figure 129: Renewable Portfolio Standards or Voluntary Targets (2015) 170
Figure 130: Generation Pattern for a Wind Turbine 173
Figure 131: Group Hourly Wind Generation for Each Month into the Peak-Intermediate-Base Hour Tranches 173
Figure 132: The Wind Generation Duration Curve for 1,000 MW of Wind Capacity 175
Figure 133: Stairstep Adjustment of Monthly Load Shape to Account for Stochastic Wind Contribution 176
Figure 134: Wind Energy Generated by Month in a Given Region 177
Figure 135: Typical Day-Night Summer Wind Pattern 177
Figure 136: Wind Generation Pattern (Annual by Hour Tranche) 178
Figure 137: Modification of the Load Pattern Resulting from the Introduction of Wind 179
Figure 138: Solar Insolation during the Day-Night Cycle 181
Figure 139: Group Hourly Solar Generation for Each Month into the Specific Hour Tranches 182
Figure 140: The Solar Duration Curve for 1,000 MW of Solar Capacity 182
Figure 141: Stairstep Adjustment of Monthly Load Shape to Account for Stochastic Solar Contribution 183

Figure 142: Solar Generation Pattern (Annual by Hour Tranche) 184
Figure 143: A Model of REC Acquisition to Achieve 20 Percent RPS 186
Figure 144: RECs and Renewable MWh Produced and Consumed 189
Figure 145: An Interconnected Generation-Transmission-Load System in IEMM 193
Figure 146: Each of the Busbars in Figure 145 193
Figure 147: IEMM Represents Interregional Postage Stamp-Distance-Zonal Ratemaking with Constraints 198
Figure 148: IEMM Represents Interregional Distance Dependent (Pancaked) Ratemaking with Constraints 199
Figure 149: Supply Curve for Transmission Service 201
Figure 150: No Congestion on the Transmission Link 201
Figure 151: Congestion on the Transmission Link 202
Figure 152: Storage Is a Dispatchable “Tank” with Injections and Withdrawals 205
Figure 153: Without Storage, Markets Equilibrate in Real-time 206
Figure 154: Storage Buys from and Sells to a Market 206
Figure 155: Daily Load Shapes Are Quite Predictable in Every Industrialized Electric System 208
Figure 156: Capacitor/Batter Storage at Point of Use 209
Figure 157: How Do You Dispatch Generation without Storage? 210
Figure 158: If You Had a PERFECT Power Storage Technique (Battery or Capacitor), What Would You Do in Summer? 211
Figure 159: How Would You Dispatch in - Winter without Storage? 212
Figure 160: If You Had a Perfect Power Storage Technique (Battery or Capacitor), What Would You Do in Winter? 213
Figure 161: The Monthly Load Duration Curve 214
Figure 162: Modification of the Load Pattern Resulting from the Introduction of Electricity Storage 215
Figure 163: Capacity Reduction Resulting from Electricity Storage 216
Figure 164: Monthly Load Duration Curves with Power Storage 217
Figure 165: Plant Operation with Power Storage 217
Figure 166: United States Production Is Flat While Consumption Is Strongly Seasonally Peaked 218
Figure 167: Capacitor/Battery at Wind Turbine Site 219
Figure 168: Wind Generation Pattern Averaged Over the Month 220
Figure 169: Power Consumption Is Hourly 223
Figure 170: The System Must Be Sized to the Peak 224
Figure 171: Hydro Consumption Is Temporal 224
Figure 172: Hydro Storage Reduces Generation Capacity in Place 225
Figure 173: Massively Oversized Hydro Completely Flattens Generation 226
Figure 174: Storage Reduces Generation (and Generation Additions) 227
Figure 175: Transmission Sizing Interacts with Storage 227
Figure 176: Storage Reduces the Peak-Off-peak Price Differential 228
Figure 177: Pumped/Reservoir Hydro Example—California Water Project 229
Figure 178: IEMM Representation of Pumped Hydroelectric 229
Figure 179: Renewables and Hydro Are Important Everywhere (e.g., Ecuador) 231

Figure 180: The IEMM Storage Model 232
Figure 181: Time Points of Equal Length for Storage Model Illustration 232
Figure 182: Injection and Withdrawal Balance at Time 1 233
Figure 183: Inelastic Supply-Demand Equilibria Are Calculated for Each of the Six Load Tranches 237
Figure 184: Network Structure for Energy Limited Resources 237
Figure 185: Disaggregation of Each Month into Six Load Tranches 238
Figure 186: Supply (Availability) of Energy Limited Resource 238
Figure 187: Price that Expends the Energy Limited Volume 239
Figure 188: Banking Deposits from and Withdraws to a Market 241
Figure 189: Purchasing Emissions Allowances for Disposal 243
Figure 190: Model of Residential Energy Consumption and Load Modification 245
Figure 191: The Air Conditioning Fleet is Carefully Vintaged; Vintage Rollover Is Very Significant 247
Figure 192: Stairstep Hourly Air Conditioning Demand by Tranche 248
Figure 193: Calculated Stairstep Approximation to Fuel Consumption 248

TABLES

Table 1: Number of Hours that Exist During Each Month and within 124
Table 2: The Temporal Structure for June in IEMM 131
Table 3: Total Monthly Energy and Assignment of Total Monthly Energy to Time Tranche ...
.... 132
Table 4: Energy and Capacity per Sub-Time Tranche during the Month of June 134
Table 5: Representative Reserve Margin Assumptions 145

1 INTRODUCTION

This document comprises our Component Design Report (CDR) to meet the United States Energy Information Administration's (EIA's) requirements for its International Electricity Market Model (IEMM). We believe and intend this CDR to be definitive and to be fully responsive to EIA's objectives for IEMM design and fully capable of allowing EIA to make prototyping and implementation decisions. We are focused on the methodology and calculations, but we give attention to data and data availability in an appendix because electricity is such an exceedingly data intensive business.

The linchpin of a successful, accurate, and workable EIA IEMM is a CDR articulating the scope, scale, structure, methodology, data, and applicability of the IEMM. IEMM will have at its heart both Balkanized, individual electricity systems as well as integrated, interconnected, multi-country systems. Both are ubiquitous throughout the world. Absent a clear, comprehensive, workable, methodologically respectable, and definitive CDR roadmap, it would be difficult or impossible to implement effective world and regional electricity models.

IEMM is unlike a world oil or a world gas model. Electricity is not and can never be interconnected to the same degree. For some regions [e.g., Organization for Economic Cooperation and Development (OECD) Europe, North America], IEMM will be composed of a fairly large, interconnected model of the electrical system. The decisions people have to make in such systems are extremely sophisticated, ranging from capacity addition, environmental remediation via plant retrofit, renewables entry, capacity market and value, storage and load-shifting, and other important phenomena. For other regions (e.g., Africa, Asia, Other Non-OECD) that are less developed, IEMM will be composed of country- or subregion-specific models that are not interconnected, i.e., a series of Balkanized subregional models, a series of regions that lack the capital to retrofit plants or maintain a world-class reserve margin. Such regions tend to be more primitive and more capital-constrained. Both types of models have to embed the thermal, hydro, renewable, technological, and other generation, load, as well as the reliability endogenous to the region. The CDR will articulate a method that can accurately represent all of these issues, and do so in the correct and relevant economic and process context of the region.

In late 2014, ArrowHead produced for EIA a definitive CDR for Transportation and Logistics in support of EIA's Global Hydrocarbon Supply Model (GHySMo). It was a detailed, comprehensive design and method, and many of the issues for IEMM modeling were de facto addressed there. We will not repeat the themes that were developed there but instead will refer to that CDR. This CDR will be a methodologically oriented approach, as the GHySMo CDR was. However, this IEMM CDR will specifically and systematically deal with the full range of electricity dimensions that exist in real world electrical systems and that are so critically crucial to the success of IEMM. With regard to the former CDR, one cannot blithely extrapolate a less temporal commodity such as oil or gas to a "just-in-time" commodity and industry such as electricity. This CDR has to be centered on "just-in-time" issues and the full range of the electric system that serves and even moderates those just-in-time needs (e.g., via storage).

As with the previous CDR, we offer a generic (non-platform-, non-model-specific) approach and put forth a microeconomic, agent-based methodology (independent of any platform, commercial application, or model). Our intention is to stress methodology, approach, scope, and comprehensiveness independent of ArrowHead's or any other modeling platform about which we know. Our proposal is not a thinly disguised "user guide" to any specific software platform, including our own. Rather, it is a stand-alone design of the best possible electric model, which can be referenced and used by anyone. Let us be clear. This CDR is not an articulation of the design or methodology of an existing electric modeling system (e.g., ArrowHead, MarketPoint, Aurora, ProSym, ProMod). There is no confidential or proprietary information regarding those systems (to the extent we know them) included herein. Rather, this is a methodological document that puts forth a comprehensive design on paper.

We stress that, just as with the former CDR, much of the work required in this IEMM CDR has not been previously integrated into a single document. The present CDR is intended as a greenfield, first of a kind document that puts forth a full methodological design for IEMM and its various components. Keep in mind, temporality is very much more complicated with power than with virtually any other economic commodity. Electricity is a "manufacturing" and real-time logistics business, not a depletable resource and logistics business. Electricity is a "just in time" business. We know that fundamentals are important to EIA, and this IEMM CDR strives to focus on and integrate fundamentals.

We also point out in this prefatory section that there is no way that a regional, country, or multiregional power model could ever be built based on global social welfare maximization or its close cousin aggregate cost minimization (such as with a linear program). MPEC and complementarity models, which are intrinsically, inescapably based on global social welfare maximization, cannot possibly work. As shown in the former CDR, you would have to "integrate" the agent equations articulated herein (and they would have to be integrable, which they are likely not). There is no need for them to be integrable for network economic equilibrium formulations, a huge advantage for the recommended IEMM model. Issues such as storage, load leveling, hydro, renewables, etc. do not lend themselves to "integration" or to a global social welfare maximization approach and, therefore, to an MPEC or complementarity approach. With linear programming (LP), "Been there. Done that." The previous ArrowHead CDR shows the futility of LP for problems of much smaller scope than this. It simply doesn't work. Unlike the previous ArrowHead CDR on logistics and transportation, we are not going to spend an inordinate amount of time herein on the equivalency of global social welfare maximization and Walrasian, agent-based network equilibrium in which the network spans the power generation, transmission, and consumption grid. Such equivalency or comparison would be a waste of time for models as large as electricity models; they are simply not amenable to the global social welfare maximization method. (If they were, wouldn't it already have been done by now?) Anyone who knows anything about optimization knows that if you integrate a function in a multivariate fashion, you get a function that you can maximize and in so doing generate the original function. We dare say that no one, and we mean NO ONE, will be able to integrate the agent-based, behavioral equations in the electric sector and get a problem that can be optimized. (If anyone can integrate the equations put forth herein, give us a call! Dinner will be on us.) We have discussed this during our briefing with EIA, and they agreed even in the simplest electricity model that there was no way to use global social welfare maximization where things like storage and load leveling might be involved.

Importantly, the electrical system model design articulated herein is more than mere conjecture. Dr. Nesbitt has built and run electricity models quite similar to the one articulated herein and used them in a number of applications, and he has done so almost continuously since 1978 with the Tennessee Valley Authority SAM model. We know that the proposed design works. We have more than “test calculated” it; we have used the approach for modeling and decision making for many years We have built and interpreted models of electrical systems, international as well as domestic, based on the design articulated herein. There is no question that the method and calculations work, and they represent electrical systems very well. One of the models Dr. Nesbitt built using the recommended methodology, the entirety of North America (as defined by NERC), was as large as 50 million equations and unknowns. Others we have built have been as small as several dozen equations. We doubt whether there has ever been a larger economic model than the multiregional North American electricity model we built, and we are positive through direct experience that the method and components recommended here will work for EIA for any Balkanized or interconnected electrical system anywhere in the world. Even though the CDR does not consider or propose implementation, it is nice to know that the proposed approach has been implemented and has worked and is not idle conjecture. If a model in our format were 50 million equations, the corresponding MPEC maximization problem would have 80-100 million equations, a result derived in our transportation and logistics CDR from 2014. There would be a Jacobean matrix of 100 million by 100 million within their solution method that would have to be successively inverted and solved, and there is no doubt it would be degenerate, nearly singular, and highly problematic. It just wouldn’t be tenable.

Many of the illustrations of model structure presented here come from the regional electricity model that exists within the ArrowHead modeling system, or more properly from concepts that are embedded therein. Many of the images are taken directly from that model. Also, many of the graphics and text are taken or derived from the Dale Nesbitt Economic Training Program. It is our intent that no information is taken from any other proprietary electricity modeling system or model.

1.1 Dimensions and Structure of Regional Wholesale Electricity

Several dimensions need to be included to correctly model an electricity market at the “wholesale” level. (We use the term “wholesale” to refer to bulk, regional supplies of electricity, “big” volumes, pre-distribution electric power, even if there is no liquid or traded market, as there is not in most countries and regions of the world.) This CDR will describe these critical, interconnected dimensions and recommend how EIA should build regional electricity models and interconnect them via a multi-region transmission system to fully meet your IEMM needs. These wholesale electricity dimensions, shown schematically and generically in Figure 1 and which fully span EIA categories that have been articulated (demand or load; generation and capacity; and end-use price), should **not** be misinterpreted as an “ArrowHead network diagram.” The figure is much too aggregated and conceptual for that. Rather, it is a comprehensiveness guide, making sure to enumerate all the parts and the whole.

1. **Fuels.** Whether by exogenous assumption or by connection to a single or multifuel model, IEMM must represent fuel prices and quantities available in every generation region of the world. This CDR will discuss how to do that, whether by connecting to a fuel model or connecting to local price or price-quantity projections of fuels, collectively or individually. We will discuss how to dynamically link to oil-gas-coal supply models (rather than fuel price projections or manual iteration). Fuels comprise an extremely large portion of electricity cost. Fuels comprise an extremely large fraction of the cost of electric power generation, and they must not be taken lightly; the fuel dimension will not be given short shrift in our IEMM design.

2. **Thermal generation**, including plant operation (sometimes called "dispatch")

   - Existing thermal generation
   - New thermal generation (both endogenously determined capacity additions and exogenously given capacity additions)
   - Retirements (both endogenously determined and exogenously given capacity retirements, idlements, or decommissionings)

   Thermal generation is a crucially important dimension, often misunderstood and represented. Some people assume they can default to a global social welfare maximization/cost minimization model for dispatch. Others think they have to represent bidding and operation. Others think the central issue is a purely economic decision on the part of individual plant owners (including the government) to decide what plant to run. Most realize, however, that the notion of "dispatch" from a centralized perspective is gone from real-world electrical systems and gone from valid models of real-world electrical systems. In fact, the modern electrical system has become so complex that dispatch is not even an operative notion anymore. Our CDR puts forth various methods, interrelates them, and recommends the appropriate decentralized, agent-based method. The recommended representation of thermal generation, so central to any electricity model, is profound, accurate, and comprehensive.

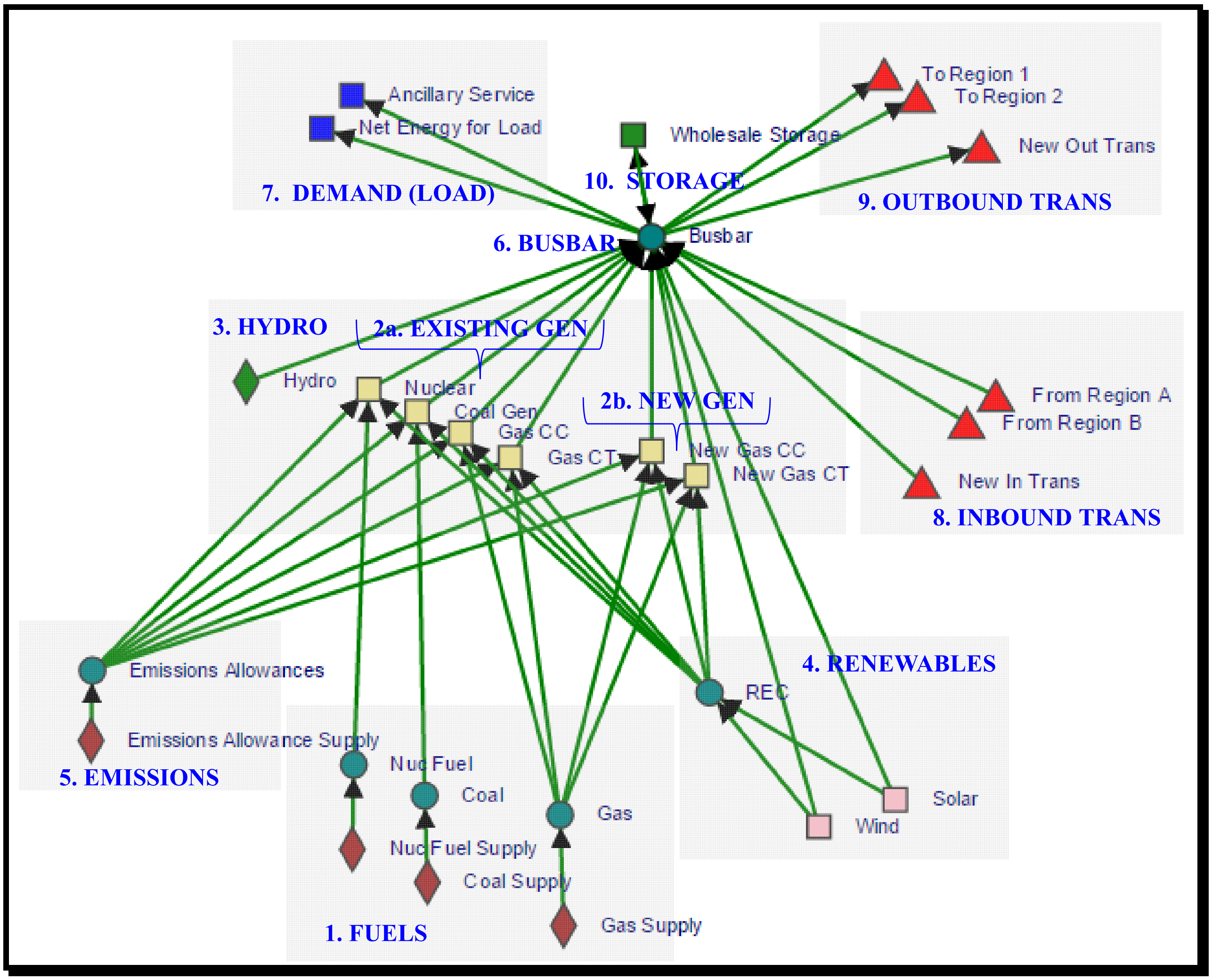
3. **Hydroelectric** (short run view and long-run view).

   - Reservoir hydro (fill and generation dynamics coordinated with price)
   - Run of river (generation dynamics coordinated with hydrology and price)
   - Pumped hydro (fill dynamics coordinated with hydrology, reservoir size, and short-term price variation)
   - Energy limited resources (resources that can generate for no more than say 3000 hours per year coordinated with price. There just isn't a big enough reservoir!

   Hydroelectric is crucial, centrally pivotal, in many power systems, particularly within developing countries (and some developed locales like western Canada). It is so easy to sweep hydroelectricity under the rug when designing a power system. Power models have traditionally been thermal generation and dispatch oriented. Alas, failure to represent hydroelectric properly leads to failure of one's power model and will lead to failure of IEMM. As far as we know, no one has figured out how to include hydroelectric in a global social welfare maximization model; it just has not been done to our knowledge. Perhaps models like

Aurora or Plexos would argue that point, but Aurora and Plexos are understood to be global social welfare maximization models. Our CDR will deal with hydro definitively. There is no developing country in the world, and certainly no industrialized country in the world, in which hydroelectricity can be glossed over, assumed away, or approximated subjectively.

Figure 1: Structure of Regional Wholesale Power Systems



4. **<u>Renewables and regulations</u>**.

- Mandates
- Renewable energy credits (RECs), which go by other names in other countries
- Direct subsidies
- Government or private installations
- Behind the fence installations
- Within-establishment installations

The comments about hydroelectric apply equally well to renewables and regulations/subsidies. The CDR will articulate how to represent renewables and renewable entry and exit. The contributions we put forth on renewables to our knowledge have not been put forth before in electric system modeling around the world. They break free from the intellectual and practical restrictions of cost minimization or global social welfare maximization that do not get them right.

5. **Emissions and Regulations**.

    - Retrofits to achieve compliance
    - Various emission regulations and how they impact generation, transmission, and load

    Emissions and regulations are yet another huge and growing topic in electricity that simply cannot be omitted from IEMM. With regulations ranging from pollutant taxes, to cap and trade restrictions, to "minimum available control technology/best available control technology" (MACT/BACT), to "new source performance standards/existing source performance standards" (NSPS/ESPS), to the command and control policy termed "Clean Power Plants" or CPP, to a veritable plethora of other command and control, centrally planned and legislated environmental and emissions regulations; this CDR shows how to consider and integrate them, which of course must and will be country or region contextual within IEMM under this design. IEMM will require a wide range of regulatory representations given the broad diversity of regulatory regimes across the world, and this CDR puts forth the necessary science and design.

6. **Price competition at the busbar**. The busbar is where notions of least-cost dispatch (implying marginal cost or "system lambda" pricing), average-cost pricing, price caps, and other pricing methods have entered the historical discussion. The CDR addresses how to produce a cogent price at the wholesale level. Transferring that cogent price to the point of consumption is discussed in the retail sections.

7. **Load and (locational) reserves**. Load representations can vary from chronological hourly to highly tranched monthly or annual. All loads in IEMM must be located, geographically specific loads. This CDR builds up a load model from first principles—hours or fractions thereof and electricity consumption during those periods. Load representation is a sophisticated issue, and this CDR will address it from hourly chronological to tranched annual with important gradations in the middle. Many developing and other countries have no notion of reserves or system reliability, which is an important additional dimension of load. (Many electric systems choose to have generators standing in reserve and ready to go, implicitly meeting a demand for reserves, i.e., a requirement for reserves. IEMM has to represent such reserves as well as net energy for load.) Developed countries do indeed have reserves and reserve requirements. The method we put forth spans both net energy for load and reserve requirements. Representation of load has proven a thorny issue in electricity models, and this CDR aspires to clarify. In fact, the load representation developed here has not to our knowledge been used in previous electric dispatch models because, as dispatch models, they assume inelastic load (demand). Assuredly the science and methodology put forth here will elucidate.

8. **<u>Inbound transmission</u>**. This CDR addresses costs, losses, tariffing (e.g., postage-stamp, distance-based, pancaked), areal extent of a transmission system, and, of course, the inbound transmission system to any area or market. This transmission discussion will be limited to "energy" and not "complex arithmetic" (meaning wave patterns) such as in load flow models. Anyone proposing "complex arithmetic" load flow methods will not be credible in an economic CDR such as this. The loop has never been closed between fully AC or cyclical power on the one hand and energy prices and volumes on the other hand. It is an unsolved problem as we see it.

9. **<u>Outbound transmission</u>**. Same as inbound transmission.

10. **<u>Wholesale electricity storage</u>**. As an example, California has implemented mandatory electricity storage using a "qualified" electricity storage method. Some electricity storage techniques occur at wholesale, and some at the point of use. Storage is a potentially game-changing issue, having a potentially profound effect on the entire electricity system—installed capacity, operation, retirements, and fuel choice. This CDR will elaborate and recommend an approach. ArrowHead has implemented the approach commercially within California to simulate the recent California law related to mandatory storage. Global social welfare maximization has profound difficulty representing storage or endogenous load-shifting. We have not seen a global social welfare maximization model that can do so, and there is no prospect of that occurring as we see it. For anyone who is skeptical, please go ahead and try to integrate the equations to develop the maximand, and please phone ArrowHead when you are done!

The ArrowHead CDR will address the preceding dimensions individually and then address how to connect them into a cogent IEMM model.

1.2 Model Temporality

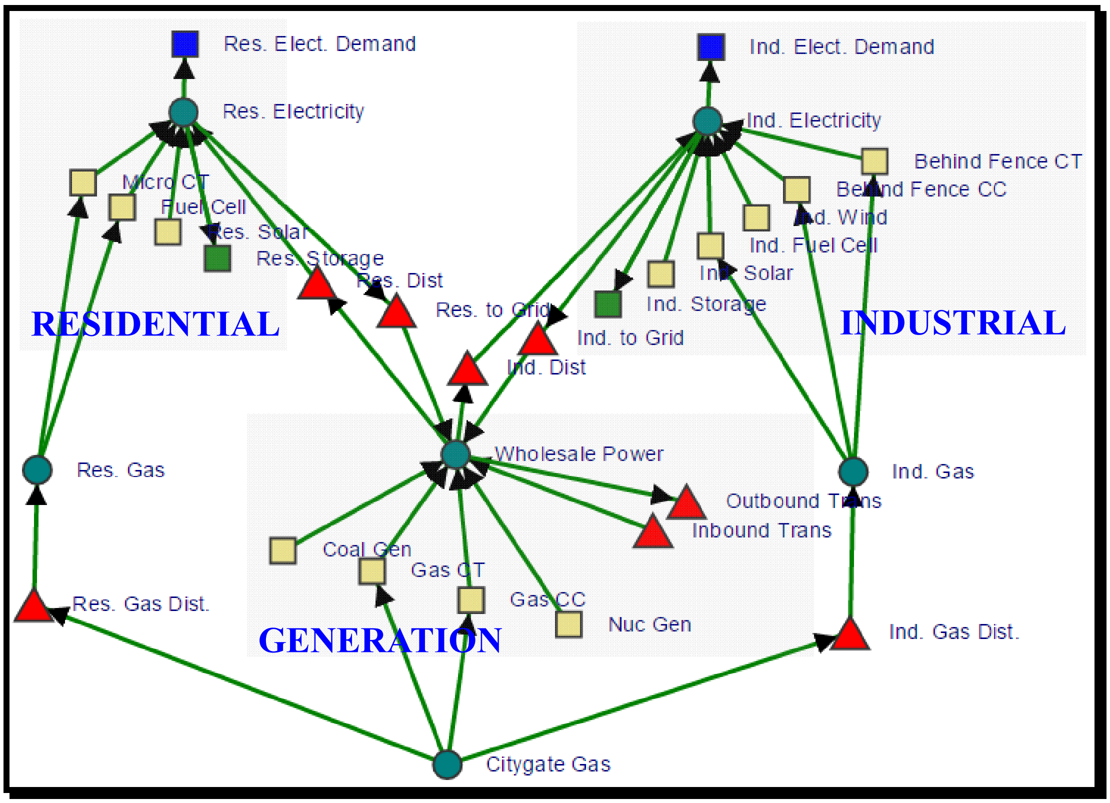
Model temporality is a central issue for IEMM, and there need to be several versions of IEMM at EIA for each country to properly model temporality. Some versions of IEMM have to be as short-run as hourly (power storage, pumped hydro, rapid load variation, ramping). Temporality that short cannot be used for capacity expansion decisions, comprised as they would be of 25 years times 8760 hours per year and 8784 hours for leap years. That is simply not viable for longer term projections or decisions. Longer term versions of IEMM have to tranche or group hours to be viable for capacity expansion, retirement, or pricing decisions. Shorter-term versions of IEMM have to represent hourly loads or lower. Our CDR will describe a method for IEMM, which has the same or similar structure with temporality that can be adapted for both uses.

1.3 Dimensions and Structure of Regional Retail/Point of Use Power Models

The load side of the wholesale model in Figure 1 above needs to be expandable in many countries to represent "retail" markets, a term we use to refer to "point of consumption," as depicted in Figure 2 below. Retail or point of consumption dimensions are important if IEMM is to get load right, both MWh and MW, and endogenize load and load-shifting. There can be no escape in

IEMM from these retail concepts and an endogenous representation of load and load-shifting. In the absence of that, load is completely arbitrary and difficult to justify. There is zero prospect of modeling any of this endogenous load and endogenous load-shifting using global social welfare maximization or cost minimization. The problem is simply too detailed, too agent specific, and too autonomous among agents in practice. Retail decisions are fundamentally agent-based. The notion that somehow decentralized retail decisions can march in simultaneous lockstep to a global social welfare function is specious at best. The retail and consumption dimensions included in the CDR are depicted in Figure 2:

Figure 2: Dimensions of Regional Retail Power Model



1. **<u>Sectorization</u>** (e.g., residential, industrial, commercial, transportation). The different scales, machinery, and use patterns in each sector require disaggregation. In developing countries, this may be all that is necessary. In developed countries, more is needed. We think of these sectors as individual classes or types of consumption. We conceive of representative or aggregate "establishments" in each of these sectors, and we think of these "establishments" as buying end-use fuels (gas and electricity) and consuming them using highly time sensitive technologies to satisfy highly time sensitive end-uses such as air conditioning, cooking, clothes drying, lighting, and so forth. We need an establishment model for each sector to characterize load and load temporality, which is crucial to IEMM. We will put forth one establishment model in this CDR as an example.

2. **<u>Distributed generation,</u>** which refers to customer site or micro-regional generation.

- Distributed thermal
  - Microturbines and similar small scale machinery
  - Distributed generation (DG)
  - Fuel cell
- Distributed solar, photovoltaic, or district heat
- Distributed wind
- Distributed geothermal

3. **Distributed storage**

   - Battery
   - Capacitor
   - Compressed air energy storage (CAES)
   - Thermal (ice, hot material)

4. **Load control** (e.g., direct load control or DLC, clocked operation, and prioritized service cuts). One part of load control includes curtailments or interruptions, which are quite prevalent in power systems outside North America and Europe. People just don't get power at certain times. Another mechanism might be priced reliability bands. The IEMM design allows these types of load control scenarios to be implemented and studied for various electrical systems around the world.

5. **Reversible distribution lines** (power by direct haul or backhaul from the point of consumption back to wholesale). We will discuss moving point of use energy back into the grid in real-time. Moving energy back into the grid from point of use has become a huge issue, and it interacts with the physics and economics of the distribution system.

6. **Conservation** (technology, building shells, lifestyle changes, and mode shifts). Conservation becomes a key issue, and we will elaborate substantially in the establishment section of the CDR.

7. **New loads (e.g., electric vehicles).** There is a rather humorous yet illustrative case study from Los Altos, CA on that, but one that is highly illustrative of what IEMM must model. That case study is potentially repeated in most real-world power systems. With the potential advent of electric or plug-in hybrid vehicles, there is possibly a colossally large impact on residential distribution and therefore wholesale electricity, both the absolute magnitude of load as well as the timing of load. IEMM will represent that.

8. **Distributed generation and reserves.** Distributed generation and other techniques provide distributed reserves. Distributed reserves can be "black start machines" that can provide reliability when they are not running or larger machines that must be kept "hot," i.e., "spinning." In systems where reserves are not provided by the local electric generation company, distributed reserves can be crucially important.

9. **Retail prices to each segment**. Prices can be (and are) discriminated among segments. Price discrimination by sector or by use needs to be addressed.

Regulations, constraints, market design, and similar issues may have to be superposed across the wholesale, retail, and integrated multiregional models within IEMM. We address the regulatory dimension generically and methodologically in our CDR. We do not herein research or delve into specific electricity regulations in every venue of the world; no one can do that until country by country implementation is put together with an eye toward modeling. We will emphasize how to implement the impact of such regulation in the IEMM model. The best approach to regulation for IEMM is handling it as a generic mechanism. By so doing, the methods herein can be extrapolated to specific situations around the world. EIA will be able to represent the specific regulatory regime in each electricity model which collectively comprise IEMM. There is no uniformity of regulation, and assuredly there is no single world model or method when it comes to regulation. That is one of the great fallacies of a common, abstract method such as global social welfare maximization.

1.4 CDR Topics Beyond Original Requirements

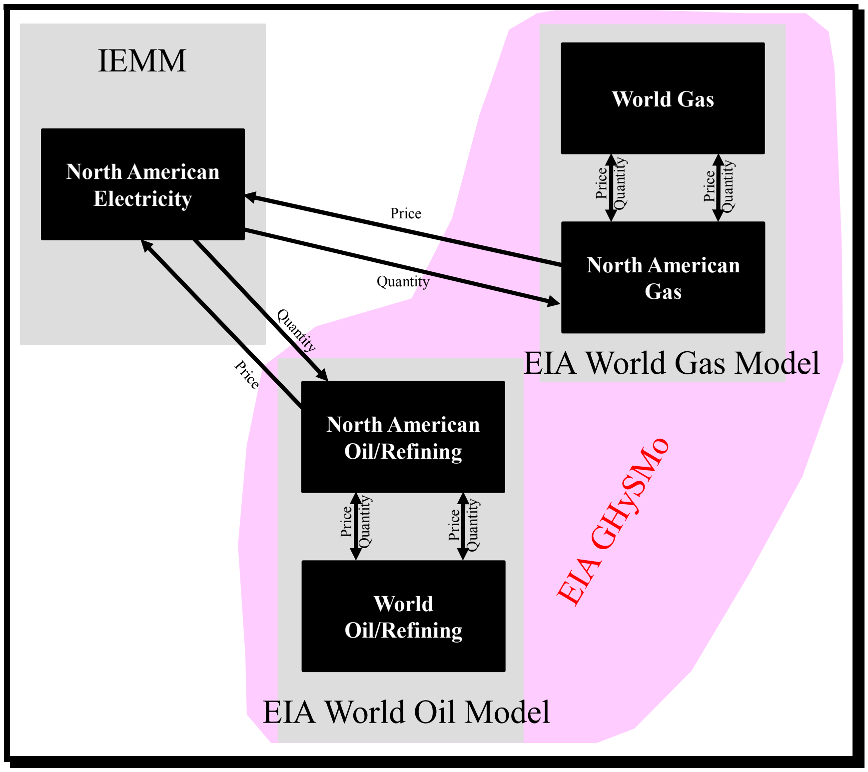
There are a few aspects of electricity that are critical to achieving a successful IEMM that may not have been sufficiently emphasized in the requirements. Several of those, which overlap with some of the preceding requirements, are the following:

1. **Hydropower.** Hydroelectric generation is a critical component in numerous electrical systems ranging from Scandinavia to Brazil to China to Canada to the United States. It is essential that hydro be represented accurately and in an integrated, interconnected way, and the design of IEMM in this CDR does so. Hydroelectric power has hourly, daily, weekly, and certainly seasonal impacts on the electric power system, and all must be considered in IEMM. The proposed design ensures that they will. Hydroelectric power is ubiquitous throughout the world, and it must be carefully represented. Hydroelectric power does not subscribe to the global social welfare function maximization paradigm. It comes when it wants, or it comes when an agent chooses to use it or chooses to exploit its value.

2. **Fuel model integration.** In electricity markets, fuels are major influences on market and electrical system behavior; fuel cost comprises a substantial fraction of total generation cost. To represent electricity markets and the electrical system accurately, natural gas, coal, oil, nuclear fuel, and other fuels ultimately need to be modeled in an integrated, interconnected way. IEMM needs to capture the inter-sectoral relationships dynamically. EIA is beginning to use the ArrowHead Global Gas Model and Global Oil Model to represent oil, products, and gas. For the IEMM project, those models can be interconnected and integrated. Fuel models foster uniformity of approach and assumptions on all the various fuels right along with electricity. Figure 3 illustrates schematically. With the design we put forth here, it is possible to integrate any fuel model except for a linear programming model and most probably an

MPEC or complementarity model (which likely has the same discontinuities) with the power model. [1]

Figure 3: Integrating Fuel Models (Oil and Gas) with North American IEMM Example



3. **<u>Emissions and Emission Credits</u>.** Emissions, emissions credits, and regulations are critical elements in electricity markets. To represent electrical system and market behavior accurately, emissions related commodities need to be modeled in an integrated way to dynamically capture the electricity-environment interrelationship. The IEMM design in this CDR does precisely that. For analyzing many issues related to power, emissions have become a centrally important aspect in understanding electricity, fuel, and other markets. Emissions need to be endogenized rather than just being included as exogenous inputs in which emissions prices are "guessed" or ignored altogether. Emissions, ranging from the older and more mundane pollutants such as SO2 to the more modern and concerning emissions such as CO2 promise to have increasing impact on the power system over time. Conversely, the power system promises to have increasing impact on these emissions over time. How many times do we read that the utility of the future will be extremely CO2 sensitive?

[1] We understand that EIA's own Dr. Harvey Greenberg has proven that linear programming yields discrete and countable prices, not continuous prices. Even if the attribution is not correct, the result certainly is. Linear programming has a countable, discrete number of prices; price is not continuous. That renders linear programming models incapable of integration with any model save for another linear programming model, and such integration would have to be direct and temporally consistent. Dr. William Hogan related the problem to us years ago. The same problem likely besets complementarity and mathematical programs with equilibrium constraints (MPEC). There is no way to disprove the discontinuity and countability effect in nonlinear programs.

4. **<u>Regulatory Actions</u>.** Models need to be able to easily represent an extremely wide variety of regulatory actions and any potential regulatory actions. This CDR is poised to represent many disparate regulations. It is not realistic to enumerate every existing and conceivable regulation in this CDR. It is realistic, and indeed important, to enumerate the various generic regulations and show how they can and will be incorporated.

5. **<u>Storage</u>**. In certain areas that have hydro resources, electricity storage (either direct storage or indirect storage) has always been an important market element. It is going to be even more important moving into the future, and every IEMM in every region must represent it. With the potential advent of many new power storage technologies such as the Tesla Powerwall and with certain states such as California legislating the adoption of storage withdrawal at time of peak, accurately representing power storage in models in an integrated way is essential. The design in this CDR does precisely that. IEMM simply cannot be without a comprehensive, sophisticated representation of storage and load-shifting, and it must be represented throughout the electrical system from generation to wholesale to retail to point of use. No portion of the power system is going to be without some influence from storage in North American and literally every other system in the world.

6. **<u>New technologies and future market developments</u>**. Our recommended approach for the IEMM can accommodate new technologies and new market evolution. Our generation models allow the full range of technological change. Assuredly, generation is not the only location in the electric system in which technological change must be considered. It also must be considered in transmission, end-use technology, renewables, environmental-removal technology, and perhaps others. As we proceed, we will see that this IEMM design accommodates the full range of new technology and equality importantly new operation based in whole or in part on that technology, and it accommodates at whatever level of the network it occurs—fuel, generation, transmission, end use, etc.

7. **<u>Endogenous Capacity Changes</u>.** Endogenous capacity change is a capability we have invented and championed over the years (it began in the SRI-Gulf model in 1973) that is critical to characterizing the impact of capacity addition and capacity retrofits on power. This CDR addresses endogenous capacity changes in great detail.

A final and very important point is the following. After understanding the scope and scale of electricity generation, transmission, load, etc., who with a straight face would propose a "reduced form" or "simple" model of electricity? No one with any sanity; such a model would be a travesty. Unequivocally, there would be no possibility of an aggregate or reduced form IEMM. Electricity is one of those commodities that requires detail and sophistication. JIT demands it. The sheer number and type of generation stations, all different, demands it. Electricity is not a commodity that wears well under aggregate or reduced form models.

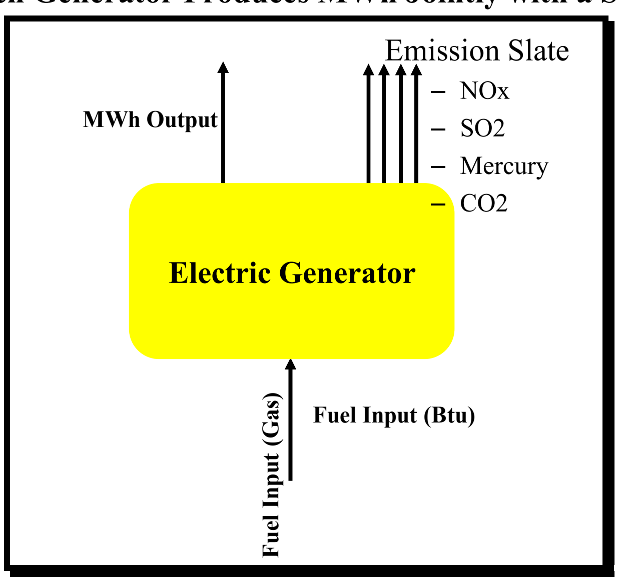
2 THERMAL ELECTRIC GENERATION AS IMPACTED BY EMISSIONS REGULATION

A linchpin of IEMM, as it is in other power models with lesser capability and aspirations, is the thermal electricity generation component. This section lays out the scope and design of the thermal electricity generation component, taking into account environmental issues (prices and emission levels), renewable constraints, and regulatory signals as well as renewables. This section covers elements 2a, 2b, and 5 in Figure 1.

2.1 Plant Sub-Models That Account for Economics and Emissions

IEMM needs to include every electricity-related thermal asset in the country or region of interest, whether individually or as part of an aggregate generation activity (meaning that generation nodes in IEMM might represent individual plants, or they might represent aggregates of plants into aggregated nodes). It needs to quantify how each such asset competes and complements, and how every asset combines to determine market prices of power, allowances, and fuels. The IEMM design begins by recognizing that each generator in each region being modeled produces MWh as well as ejecting its attendant slate of emissions as shown in Figure 4. The figure shows iconically that the plant takes in a fuel (indicated in the diagram), executes its thermal generation process, and produces MWh output and a slate of pollutants, which occur in part because of the chemicals in the fuel and in part because of oxygen or nitrogen fixation to those fuels as they are combusted at high-temperature in air (which is 80 percent nitrogen). Combustion chemistry is not arcane rocket science; it simply recognizes that generators jointly produce MWh and pollutants. This representation is not limited to thermal generators, but nonthermal generators generally eject fewer or no pollutants.

Figure 4: Each Generator Produces MWh Jointly with a Slate of Pollutants



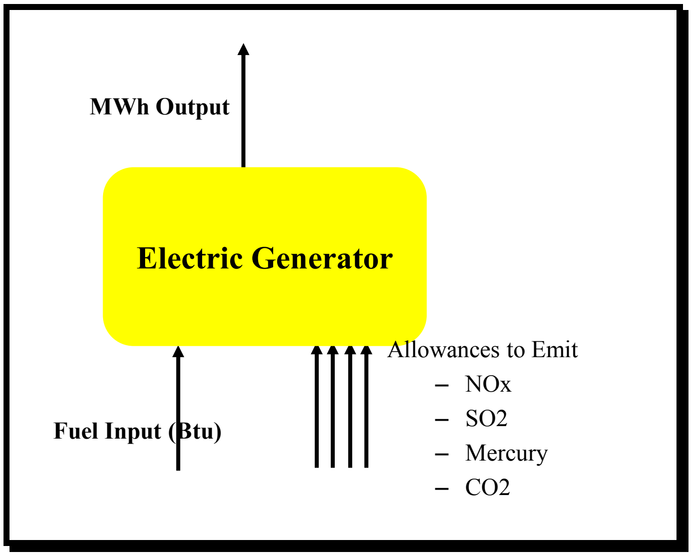
The specific slate of pollutants indicated in Figure 4 has been a central consideration in the United States, Europe, and other power systems throughout the world over the years. The method put forth herein allows a regional or integrated model to include any subset of these particular pollutants, and it allows consideration of other pollutants as well that are not indicated in the figure (e.g., particulates, water, and land use). IEMM must not be restricted to any specific, irrevocable, inflexible slate of pollutants, i.e., not restricted to the specific set in Figure 4. IEMM must be able to adapt instantly to any slate of pollutants or "externalities" that need to be represented so that it can characterize each power system around the world and whatever environmental restrictions that are imposed upon it. In the ArrowHead electricity model used here for illustration, we have included only the four pollutants in the figure—SO2, NOx, mercury, and CO2—as a matter of convenience. (In the past, our models of the type designed here have included pollutants and externalities such as particulates, cooling water, land use, etc. in addition to the four pollutants depicted in Figure 4. There must be no limitation in IEMM on the slate or types of externalities that can accompany each generator.) Our continued discussions will focus on the four pollutants in Figure 4, but keep in mind that the list can be expanded or contracted in every country or region IEMM might choose. There is nothing in the recommended method that is married to these specific four pollutants.

To be allowed to generate in a pollution-regulated regime (of which complete absence of regulation and complete freedom of action is an important and quite prevalent special case throughout the world), each generator must obtain and surrender emissions allowances that exactly offset the emissions it generates. In an unregulated situation, these emissions allowances are free—i.e., zero price. In fact, **the quintessence of non-regulation is a zero allowance price**. (In economic parlance, that is exactly what an externality is, right out of the economics textbook. An externality is a commodity that has a zero price. It is a commodity needed to generate but that has no market and, therefore, a zero price or cost. Varian's[2] classic microeconomic text gives examples of externalities and the lack of pricing.) This notion of obtaining allowances or entitlements that "match up with" and, therefore, offset pollutants is extremely important and quite equivalent to the joint output notion in Figure 4. In particular, we think of the notion of generating MWh and producing a slate of pollutants as equivalent to the notion of generating MWh and obtaining a required, exactly matching, exactly offsetting slate of pollution allowances. The generator obtains an allowance for every unit of pollution it creates as illustrated in Figure 5. Obtaining an allowance one for one is precisely equivalent to the accounting calculus in Figure 4, but it allows one to price, regulate, or even preclude through constraints or other signals the emission of certain pollutants. An important capability IEMM must embody is being able to perform the proper accounting and to perform the proper pricing or consideration of the specific magnitude of pollution that accompanies its own generation. Furthermore, the structure in Figure 5 is going to allow IEMM to price (or not) the various emissions or restrict or ban them through regulation, trading, or other mechanisms, all of which are crucial to the proper operation and productivity of IEMM. To interpret Figure 5 and its implications for the cost and profitability of generating, keep in mind that the fuel is priced (prevailing coal, gas, oil, or nuclear fuel prices) and the emissions allowances are also priced. Such pricing allows a simple yet comprehensive pro forma of plant profitability and, therefore, plant operation or dispatch to be implemented. This design of a generation node is crucial yet frequently overlooked or accomplished in a different fashion.

---

[2] Varian, Hal R., **Microeconomic Analysis**, Third Edition, Norton, 1992.

Figure 5: Each Generator Purchases and Surrenders Emissions Allowances Consistent with Its Level of Generation



We want to emphasize that this construct in Figure 5 does not mean that we assume that every IEMM generation model is a pollution tax or a cap and trade model or even a pollution-priced model. Clearly if the pollution tax is zero, the model in Figure 5 does all the pollution accounting, and does it correctly, but it does not price it and, therefore, does not elevate the variable cost of generation. The externality remains, and its magnitude is quantified. It does not internalize the externality. The notion of doing the pollution accounting correctly from the outset, whether or not the pollutant is priced, is central to IEMM. IEMM cannot really succeed without it. EIA will be asked for the pollution footprint of various electric entities around the world whether or not those pollutants are regulated.

Using the plant construct in Figure 5, we can begin to specify the basic economics and the basic pollution footprint of that plant. That is, for each plant input to IEMM, it must measure or judgmentally estimate the variable cost related parameters of the given plant, which are illustratively tabulated as in Figure 6 for a modern combined-cycle unit (the parameters are analogous but have different numerical values for all other individual plants and types of plants in the market).

Figure 6: Parameters of a Representative Thermal Plant in the IEMM Model

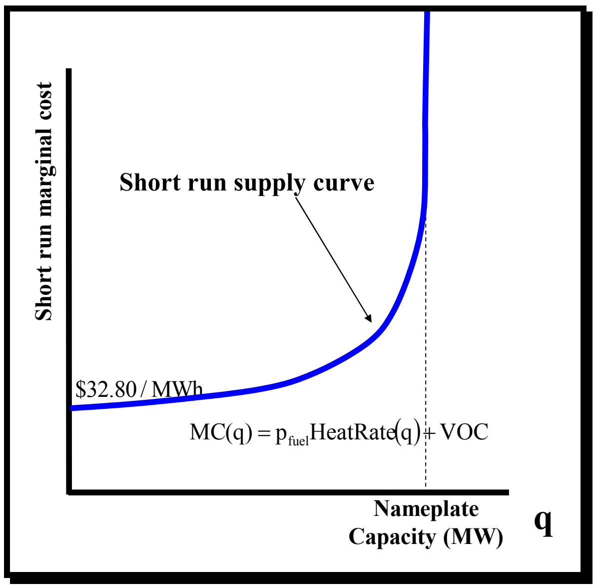
Combined Cycle Plant	
Capacity (MW)	500
Heat Rate (Btu/KWh)	7200.00
O&M ($/MWh)	$4.00
SO2 Coeff (lb/MWh)	0.00
NOx coeff (lb NOx/MWh)	0.20
Hg coeff (lb/MWh)	0.00
CO2 coeff (ton/MWh)	0.41

Given the plant parameters in Figure 6 and an assumed gas price of \$4.00/Mcf, for a natural gas combined-cycle plant, one can combine the natural gas price and the plant parameters to calculate the short run, economics-only[3] supply curve for the plant:

$$\text{Generation Cost} = p_{GAS} * \text{Heat Rate} + \text{Var O\&M}$$
$$= 4\frac{\$}{\text{MMBtu}} * 7.2\frac{\text{MMBtu}}{\text{MWh}} + \$4.00/\text{MWh} = \$32.80/\text{MWh}.$$

Generation cost means short run variable cost in the sense that it contains fuel cost and variable operating cost only (but no recurrent fixed cost, e.g., maintenance cost, capital cost, or pollutant cost. Short run variable cost means that this entire cost can be avoided if the plant is not operating but will be borne if the plant is operating. That is the quintessential definition of variable cost—it can be avoided in its entirety by not running the plant. Variable cost is truly "variable" with respect to output. All of the other categories of cost will be incorporated and superimposed as we proceed.) We use the preceding economic-only generation cost calculation to create the short run supply curve (short run marginal cost curve) for the plant as shown in Figure 7. This short run variable cost has been the mother's milk of dispatch models, but we shall see that it is sorely inadequate.[4]

Figure 7: Short-Run Supply Curve for the Plant (Marginal Cost Curve)



It is crucial to recognize that the short run marginal cost curve in Figure 7 is **highly nonlinear,** and it incorporates a production function.[5] A constant heat rate such as that implicitly considered herein thus far represents a Leontief, fixed input-output proportion production function. We will address the notion of nonlinear production functions immediately below. Toward that end, there

---

[3] By economics-only, we mean there have been no environmental or renewables costs incorporated yet.
[4] The specific functional equation for short run supply is not as important as the way in which it is used, so we defer that discussion.
[5] This notion of a production function is spelled out in substantial detail in the GHySMo CDR. That is such a crucial concept and crucial to proper electricity generation modeling that we will emphasize it shortly.

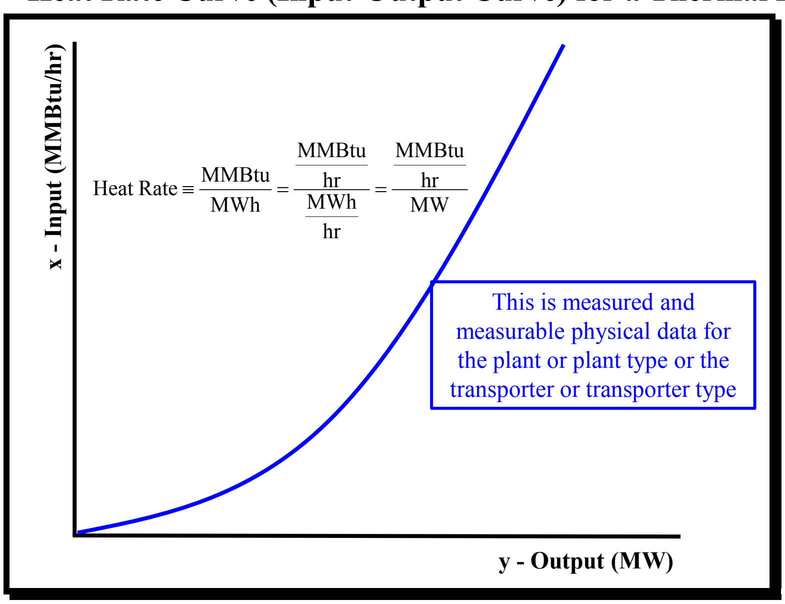
has been a goodly bit of study and statistical analysis that the underlying heat rate curve and, therefore, the production function may be well approximated by a quadratic form. In fact, a quadratic form is known to be akin to a second order Taylor approximation to any form, whether or not one can even see the actual form itself. Joel Kline of the California Energy Commission wrote a nice analysis of that some years ago. We will appeal to the salient parts of that herein as we proceed. For now, we simply assume that we have a production function that gives a cogent shape to the short run supply curve in Figure 7.

In previous publications, we (and others) have referred to the quadratic heat rate function. We summarize that briefly here using our generation node representation herein. A power plant "heat rate" is an input-output coefficient, reflecting the amount of input (fuel) that is required per unit of output. To wit, the heat rate is the number of units of **input** you need (generally expressed in terms of Btus) to produce one unit of **output** (generally expressed in terms of MWh). The quadratic heat rate equation, indicated in Figure 8, has the specific function form:

$$x = a_2 y^2 + a_1 y + a_0$$

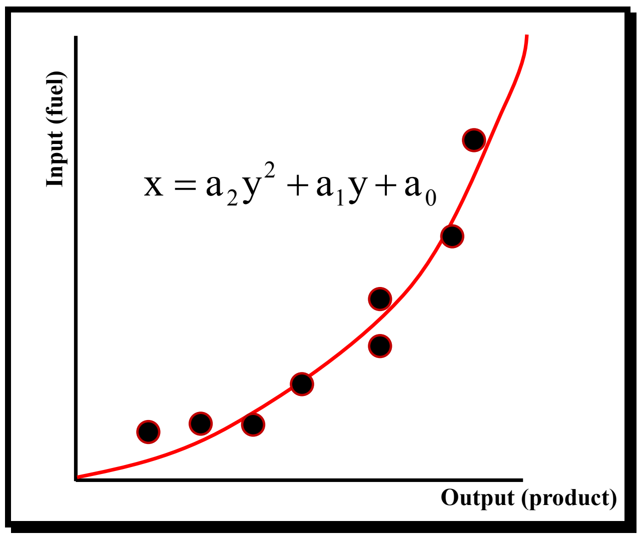
in which x is the level of input (fuel consumption) and y is the corresponding level of output (electric generation). One can fit any measured input-output curve to a quadratic form such as this. Such fitting is immensely convenient as well as accurate in representing the phenomenon that "the faster you drive the plant, the more fuel at the margin it consumes." Degradation of performance at high utilization is true for power plants, pipes, furnaces, cars, and a lot of technologies. The faster you drive your motorcycle, the more fuel it burns per vehicle mile traveled. Power plants are not particularly different.

Figure 8: Heat Rate Curve (Input-Output Curve) for a Thermal Power Plant



Lest someone assert "EIA couldn't do that; they couldn't get heat rates curves," keep in mind, people who model thermal generation are already doing it and have already done it. EIA has no choice but to do it. Modelers are measuring losses on every power plant in every system in the world, just as shown in Figure 9. Some of these estimates are generic. A particular thermal power plant type may not be built yet or isn't reported, but we know the basic thermodynamics. Plant thermodynamics is hardly a secret, hardly arcane, hardly unique. A power plant is a power plant is a power plant. Pulverized coal steam turbine power plants have common thermodynamics everywhere. Thermodynamics and combustion chemistry aren't rocket science nor highly unusual, and heat rates have typically been measured and reported rather accurately. It is not acceptable to say: "We don't have the data." We do have the data, and if we don't, we can get it, infer it, or model it. Extrapolation from a coal plant in one region that has been scrupulously measured to a coal plant in another region is perfectly acceptable; the prime mover and technology are identical. For thermal plants, it is believed that this quadratic heat rate curve is highly descriptive up to but below installed capacity. To be precise, we believe that any true and correct heat rate curve of higher order can be well approximated for modeling purposes by a quadratic in the range of interest.

Figure 9: A Quadratic Heat Rate Curve for a Thermal Power Plant



If we assume a quadratic heat rate for the power plant, the producer can be expected to produce electricity and consume fuel so as to solve the following profit maximization problem:

$$\begin{aligned}
&\underset{y,x}{\text{MAX}} \quad py - \phi y - wx \\
&\text{SUBJECT TO} \quad x = a_2 y^2 + a_1 y + a_0 \\
&x \geq 0 \\
&y \geq 0
\end{aligned} \tag{1}$$

in which

ϕ = nonfuel operating cost ($/MWh)
w = fuel price ($/MMBtu)

x = quantity of fuel burned (MMBtu).
p = power price.
y = quantity of power produced (MWh).

The mathematics of analytically solving this problem is derived in substantial detail in the previous GHySMo CDR, but we summarize the analytical answer herein. We emphasize that the solution to this problem is strictly analytical. There is no need (and no attractiveness) for any numerical solution to problem (1). No one need sprint to a nonlinear optimization program to solve it at great expense! On the contrary, one breaks out his or her calculus book to solve it, and quite simply so. The solution is strictly algebraic and analytical. Herein lies the huge advantage of the agent-based approach—there is no need for costly numerical optimization. So many operations research people express incredulity at this—we solve the profit maximization problem analytically, just like they showed us in our economics texts, and not numerically. There is simply no need. The profit maximizing production of electricity and consumption of fuel are given by the equations:

$$\begin{aligned} y^*(p,w) &= \frac{p-(\phi+wa_1)}{2wa_2} \quad \text{if } p \geq \phi + wa_1 \\ y^*(p,w) &= 0 \quad \text{if } p < \phi + wa_1 \end{aligned} \tag{2}$$

$$\begin{aligned} x^*(p,w) &= a_2\left[\frac{p-(\phi+wa_1)}{2wa_2}\right]^2 + a_1\left[\frac{p-(\phi+wa_1)}{2wa_2}\right] + a_0 \quad \text{if } p \geq \phi + wa_1 \\ x^*(p,w) &= 0 \quad \text{if } p < \phi + wa_1 \end{aligned} \tag{3}$$

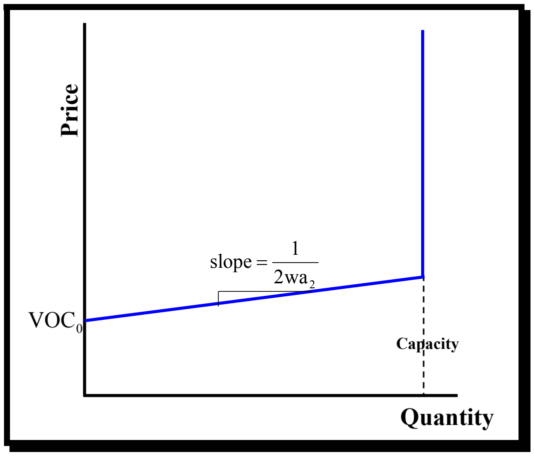
It is interesting and appropriate to regard the term $VOC_0 = \phi + wa_1$ as the variable cost for the very first increment of output from the plant, including both nonfuel cost and fuel cost.

$$\begin{aligned} y^*(p,w) &= \frac{p-VOC_0}{2wa_2} \quad \text{if } p \geq VOC_0 \\ y^*(p,w) &= 0 \quad \text{if } p < VOC_0 \end{aligned} \tag{4}$$

$$\begin{aligned} x^*(p,w) &= a_2\left[\frac{p-VOC_0}{2wa_2}\right]^2 + a_1\left[\frac{p-VOC_0}{2wa_2}\right] + a_0 \quad \text{if } p \geq VOC_0 \\ x^*(p,w) &= 0 \quad \text{if } p < VOC_0 \end{aligned} \tag{5}$$

The first equation (4) is the insightful one. It says that output grows linearly with price, i.e., the marginal cost function is linear. That means that the specific form of short-run marginal cost curve in Figure 7 is linear and upward tilting as shown in Figure 10. The upward tilt in the curve is related to the curvature of the quadratic heat rate. In our conceptual developments to follow, we will use the nonlinear form in Figure 7 rather than the linear form in Figure 10.

Figure 10: Short Run Marginal Cost Curve for a Plant with a Quadratic Heat Rate



A nonlinear short run marginal cost function such as that in Figure 7 or the specific quadratic form in Figure 10 cannot realistically be incorporated into a linear programming (or global social welfare maximization model) of a power system, a country, or a region. There would be so much discretization required as to "kill" the application. Quite the contrary, one requires an intrinsically nonlinear technique. The only option within a linear programming model would be a stairstep function or a sequence of stairstep functions arrayed horizontally so as to discretize the nonlinear function. Figure 11 illustrates the former and Figure 12 illustrates the latter. A single stairstep function as in Figure 11 is unsatisfactory, yielding nasty step response "bang-bang" behavior so characteristic of linear programming. (There is no controversy here—linear programming is known to display unseemly "bang-bang" behavior.[6]) Modelers revile the "bang-bang" nature of plant operation intrinsic within a linear program. This behavior is not at all rectified by a nonlinear global social welfare maximization approach such as MPEC or complementarity. There is no theorem for nonlinear programming guaranteeing that answers are continuous rather than stairsteppy. In fact, there are many nonlinear counterexamples with Kuhn-Tucker conditions that give decidedly bang-bang results even though one cannot discern such behavior at the outset from their specific nonlinear structure. With regard to the computational realities, discretization of intrinsically nonlinear curves obviates the ability to implement and compute sufficiently comprehensive models. There are simply too many power plants (or aggregates of power plants) to allow that. The methods have to be inherently nonlinear.

A sequence of stairstep functions sufficient to represent the true and correct nonlinear nature of each short-run supply curve as in Figure 12 is no more satisfactory. If we needed 20 tranches to represent the true nonlinearities, that would increase the number of rows in the linear program by a factor of 20. We know that the computation time in a linear program grows as $MN^2/2$ where N is the number of rows (hubs), and M is the number of columns (i.e., plants, which is always bigger than the number of rows). If the number of rows increases by a factor of 20, that means the computation time increases by a factor of 20^3=8,000: this is why linear programming models of

[6] Dr. Dale Nesbitt of ArrowHead studied linear programming at Stanford University under the late Dr. George Dantzig, the person who **invented** linear programming. Dr. Dantzig himself said that repeatedly—supply functions are stairsteps. There is no controversy about that.

electricity or world gas take many, many hours or even weeks of wasteful computation. Excessive computation time is why power modelers and IEMM abandon linear programming in favor of intrinsically, directly nonlinear methods such as the design put forth herein. There is just no value to endure the type of pain LP puts you through. The situation will not be markedly different with nonlinear programming, complementarity, or MPEC types of models. They are plagued with the same bang-bang problems and the same excessive size problems.

Figure 11: Linear Programming Is Based on a Stairstep Supply Curve

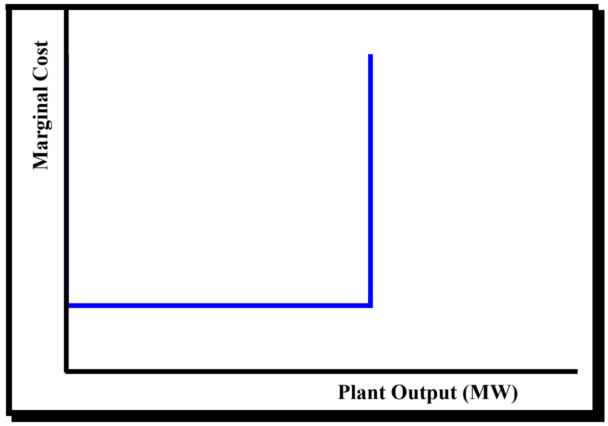
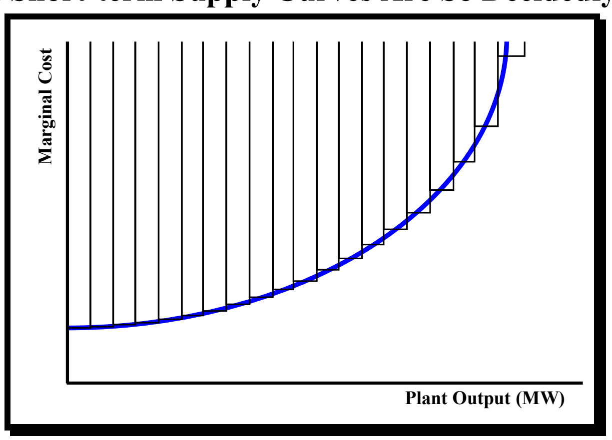


Figure 12: Nonlinear Supply Curves Require "Zillions" of Tranches to Emulate Because Real World Short-term Supply Curves Are So Decidedly Nonlinear



There is no question that IEMM will require a direct, nonlinear method. That is because the shape of the short-term supply curve is nonlinear. Inescapably, IEMM must input the nonlinear supply curve in Figure 7 directly, accurately, and without wasteful "discrete disaggregation" such as in Figure 12. It must derive whatever nonlinear supply curve is to be used from an inherently nonlinear production function, just as we showed in great detail in the GHySMo CDR, our economic training program, and just as we shall resummarize here. The design here not only allows that, it requires it.

We begin with this economic-only short-term supply curve in Figure 7 (rather than Figure 10, which applies equally well in the quadratic heat rate case). We have one such curve for every plant or aggregate of plants that compose a thermal generation region or country, and we need to overlay the environmental emissions costs atop the pure economic variable costs so that there is a total variable cost measure that includes both. As presently configured,

- the SO2 price (calculated endogenously in the model or input exogenously) times the number of SO2 allowances per MWh of generation (estimated from plant chemistry, thermodynamics, combustion, and stoichiometry) plus

- the NOx price (calculated endogenously in the model or input exogenously) times the number of NOx allowances (estimated from plant chemistry, thermodynamics, combustion, and stoichiometry) plus

- the mercury price (calculated endogenously in the model or input exogenously) times the number of mercury allowances (estimated from plant chemistry, thermodynamics, combustion, and stoichiometry) plus

- the CO2 price (calculated endogenously in the model or input exogenously) times the number of CO2 allowances (estimated from plant chemistry, thermodynamics, combustion, and stoichiometry)

can be computed and added vertically to the short run supply curve in Figure 7. When IEMM adds this plus the nonfuel variable cost, it computes the full, environmentally burdened production cost of the plant, taking account not only the economic costs (variable operating and maintenance cost and fuel cost) but also the fully embedded, internalized environmental costs for each of the pollutants, which are computed individually and summed. Keep in mind, these costs will be positive if there is any emissions regulation, and they will be zero if there is no emissions regulation for the particular country or regional IEMM in question. IEMM needs to be able to deal with both because there are regulated and unregulated systems around the world. (Keep in mind, zero regulation means unpriced, uninternalized externality.) The structure is fully general and fully applicable to the regulated as well as uncontrolled case. It is the emissions prices that differ. Such additions must be calculated identically for every plant in IEMM, as this design allows. It does it for every plant—every node—throughout the region or regions of focus for IEMM.

2.2 Sidebar—Where Do We Get These Emissions Coefficients?

The Environmental Protection Agency (EPA) and various commercial data vendors maintain databases for generation units in North America—a database of SO2, NOx, mercury, and CO2 emissions per MMBtu of fuel, heat rate, nonfuel operating and maintenance cost, and effective capacity of each plant. These coefficients are neither arcane nor mysterious for plants of similar characteristics around the world. They extrapolate to similar plants with similar fuels and similar prime movers around the world. A representative slice of such a plant database, taken from our Stanford lectures, appears in Figure 13 uncontrolled plant configurations with different coal types. The figure, approximately consistent with such a data set, expresses the emissions coefficients per pound of fuel. (That is the typical way they are measured). That is rather sensible because it is the fuel that has the "bad stuff" in it, and generation simply involves combustion of that fuel and that "bad stuff" and ejects it to the environment.

Figure 13: Emissions Parameters for Uncontrolled Plants (lb/MMBtu Fuel Input)—Illustrative

Plant	Emission Control	Fuel	CO2 Emission Coefficient (lb/MMBtu)	Hg Emission Coefficient (lb/MMBtu)	SO2 Emission Coefficient (lb/MMBtu)	NOx Emission Coefficient (lb/MMBtu)
Combined Cycle	None	Natural Gas	117.0	0.00000000	0.000	1.29
Pulverized Coal	None	Bituminous Coal	205.0	0.00001045	2.200	3.10
Supercritical Coal	None	Bituminous Coal	205.0	0.00001045	2.200	3.10
Pulverized Coal	None	Sub-bituminous Coal	210.0	0.00000670	1.400	3.10
Supercritical Coal	None	Sub-bituminous Coal	210.0	0.00000670	1.400	3.10
Pulverized Coal	None	Lignite	215.0	0.00000979	2.100	4.00
Supercritical Coal	None	Lignite	215.0	0.00000979	2.100	4.00

For every plant/fuel in the generation mix (or the region of interest), we assemble the environmental emissions data in Figure 13. (The numerical values in Figure 13 and subsequent figures in this section are illustrative. We have not attempted to render them accurate for the plants or fuels in question nor to disclose any proprietary or confidential data.) The data represents plants without SO2 or NOx control equipment (i.e., no limestone scrubbing to remove SO2 or selective catalytic reduction to remove NOx). These data in Figure 13 are not arcane, obscure, or controversial. Quite the contrary, they are well understood and in fact quite ubiquitous for power plants around North America and around the world. We understand that such data are available for every existing power plant in the United States, and they are in fact similar elsewhere in the world because the prime mover is the same, and the combustion chemistry and stoichiometry are the same. Notice that these data express the emission of each of the four pollutants considered herein—SO2, NOx, mercury, and CO2—per MMBtu of fuel. The reason is that it is the stoichiometry of combustion (and not plant configuration) that determines emissions to the atmosphere, and much of this stoichiometry depends intrinsically upon what is in the fuel and how much of it has to be burned per MWh generated. Any notion that EIA cannot get these type of data is just plain wrong. They can be garnered directly from countries or certainly by direct analogy with United States plants with similar prime mover and fuel.

Figure 13 reports the various emissions coefficients for uncontrolled plants with various fuel types and various plant configurations. It begins with a gas combined-cycle plant in the first row. It then reports, for two basic plant configurations (pulverized coal and supercritical coal), the emissions coefficients for the three basic types of coal—bituminous coal, sub-bituminous coal, and lignite. These are decreasing quality fuels. The table represents plant configurations with no environmental control equipment, uncontrolled plants. They have no SO2 control (e.g., no limestone scrubber), no NOx control (e.g., no SCR or low NOx burner), no mercury control (e.g., no activated carbon injection or baghouse filtration), and no CO2 control (i.e., no flue gas arrest and no capture and sequestration). There are a large number of such plants around the world, and these data are likely to pertain fairly directly for uncontrolled plants. These represent what uncontrolled combustion ejects to the atmosphere.

Thus far we have been addressing environmental emissions coefficients per MMBtu of fuel consumed by the power plant. These are to be regarded as the emitted chemicals embedded in the various fuels as they go through the various plant configurations and undergo combustion and gas ejection from those configurations. They are based on combustion chemistry and removal chemistry on the power plants, exactly as they should be. Power plants burn fuels, and whatever was resident within the fuel (sulfur, mercury, etc.) is chemically bonded to the combustion products via heat (NOx) will appear in the plant gas ejectant or the ash.

The values in Figure 13 are based on one MMBtu of fuel input; thus, they must be properly adjusted for power plants with different heat rates. If a plant has a heat rate of 16,000 Btu/KWh, it has to burn a lot more fuel, i.e., push a lot more molecules through the burner and onward into the atmosphere, than a plant with a 9,500 Btu/KWh heat rate. By adjusting these coefficients to plants with different heat rates, we are going to calculate estimates of emissions coefficients per MWh of power plant output. That will be an important objective of the IEMM generation modules. Assuredly, EIA realizes that IEMM must perform the emissions accounting for generation units correctly, flawlessly, and continuously throughout the model time horizon.

In this IEMM CDR, stoichiometric calculations are performed using the fuel data in Figure 13 for each of the pollutant types in the following manner. We begin with SO2 and then proceed to the other fuels in a precisely analogous fashion. To analyze SO2, we apply the heat rate of the plant and the sulfur content of the coals being burned to calculate the SO2 related information we need—the pounds of SO2 emitted per MWh generated. That calculation for SO2 is summarized in Figure 14, which is structured so as to calculate the final column shaded green—pounds of SO2 per MWh generated. It is worthwhile to walk through the calculation for a non-retrofit plant configuration in Figure 14. This walkthrough will ensure that the IEMM calculation is made correctly given the chemistry of the fuel, the combustion process, and any removal chemistry.

Figure 14: SO2 and Coal Accounting--Illustrative

Plant	Emission Control	Fuel	Heat Rate of Plant (MMBtu/MWh)	Heat Content of Coal (Btu/lb)	Heat Content of Coal (MMBtu/Short Ton)	Sulfur Content of Coal (Fraction)	lb of Sulfur per Short Ton of Coal	lb of SO2 per Short Ton of Coal	lb of SO2 per MMBtu of Coal	SO2 Removal Efficiency	lb SO2/MMBtu of Coal	lb SO2/MWh Generated
Combined Cycle	None	Natural Gas	7.200								0.0000	0.0000
Pulverized Coal	None	Bituminous Coal	10.000	12,000	24.0	0.0132	26.400	52.80	2.2000	0	2.2000	22.0000
Supercritical Coal	None	Bituminous Coal	9.500	12,000	24.0	0.0132	26.400	52.80	2.2000	0	2.2000	20.9000
Pulverized Coal	None	Sub-bituminous Coal	10.500	8,500	17.0	0.0060	12.000	24.00	1.4118	0	1.4118	14.8235
Supercritical Coal	None	Sub-bituminous Coal	10.000	8,500	17.0	0.0060	12.000	24.00	1.4118	0	1.4118	14.1176
Pulverized Coal	None	Lignite	11.000	6,800	13.6	0.0075	15.000	30.00	2.2059	0	2.2059	24.2647
Supercritical Coal	None	Lignite	10.500	6,800	13.6	0.0075	15.000	30.00	2.2059	0	2.2059	23.1618
Combined Cycle	SCR	Natural Gas	7.200								0.0000	0.0000
Pulverized Coal	SCR+Scrub	Bituminous Coal	10.500	12,000	24.0	0.0132	26.400	52.80	2.2000	0.98	0.0440	0.4620
Supercritical Coal	SCR+Scrub	Bituminous Coal	9.975	12,000	24.0	0.0132	26.400	52.80	2.2000	0.98	0.0440	0.4389
Pulverized Coal	SCR+Scrub	Sub-bituminous Coal	11.025	8,500	17.0	0.0060	12.000	24.00	1.4118	0.98	0.0282	0.3113
Supercritical Coal	SCR+Scrub	Sub-bituminous Coal	10.500	8,500	17.0	0.0060	12.000	24.00	1.4118	0.98	0.0282	0.2965
Pulverized Coal	SCR+Scrub	Lignite	11.550	6,800	13.6	0.0075	15.000	30.00	2.2059	0.98	0.0441	0.5096
Supercritical Coal	SCR+Scrub	Lignite	11.025	6,800	13.6	0.0075	15.000	30.00	2.2059	0.98	0.0441	0.4864

Consider row 2 in Figure 14, a non-retrofit pulverized coal plant burning bituminous coal. The heat rate for such a plant is reported in Column 4 of the table to be 10.000 MMBtu/MWh, which is equivalent to 10,000 Btu/KWh. Column 5 of the table gives the Btu content of the bituminous coal to be 12,000 Btu/lb, which in Column 6 translates to 24 MMBtu/ton through simple unit conversion. (For IEMM, estimates will have to be made of the coal chemistry and heat content for the coal or coals burned.) The sulfur content of this coal is reported in Column 7 to be 1.32 percent by weight. If 1.32 percent of a ton of coal by weight is sulfur, Column 8 calculates the actual weight of elemental sulfur in a ton of bituminous coal to be 26.4 lb/ton. Sulfur is atomic weight 16, and oxygen is atomic weight 8. That means that a molecule of SO2 weighs exactly twice as much as an atom of elemental sulfur. That means that the weight of SO2 that will be produced by the burning of this bituminous coal in the atmosphere will be 52.8 lb/ton as reported in Column 9. If we divide the 52.8 pounds of SO2 per ton of coal burned by the 24 MMBtu/ton heat content, we will calculate in Column 10 that there are 52.8/24 = 2.2 lb SO2 created from the burning of one MMBtu of bituminous coal. Column 10 tells us that every MMBtu of coal that is burned creates 2.2 lb of SO2 that has to be arrested and dealt with (or ejected to the atmosphere or the ash). Thus far, we have considered nothing except combustion stoichiometry, strictly chemistry.

Column 11 then presents the fraction of SO2 entering the flue gas following combustion in the boiler that will be captured and arrested by whatever environmental equipment that has been installed on the plant. We see for the non-retrofit plant that the capture efficiency is zero—100 percent of the SO2 that is created by combustion in the boiler exits to the atmosphere through the flue. That means, in Column 12, that the quantity of SO2 ejected to the environment given whatever equipment has been installed on the plant is 2.2 lb SO2 per MMBtu of fuel burned. The heat rate of the plant reported in Column 4 is 10.000 MMBtu/MWh. If we multiply column 12 by column 4, we will have our final answer—the number of pounds of SO2 that are ejected from the plant per MWh of power generated by the plant, taking account of the heat rate and therefore the thermal losses of the plant. Notice that the answer in row 2 is 22.0 lb SO2/MWh generated from the plant. For every MWh generated burning 1.3 percent elemental sulfur bituminous coal in a pulverized coal plant with a 10,000 Btu/KWh heat rate, there will be 22.0 lb SO2 ejected to the atmosphere. An uncontrolled, non-retrofit plant would eject this quantity. Ejection data such as this will be associated with the input links to uncontrolled, non-retrofit plants that exist throughout the world (and may not exist in the future).

If we compare the retrofit configuration of this plant in the second white row of the table below, we will see that all calculations through and including Column 10 are identical. This configuration would, as the non-retrofit configuration, eject 2.2 lb SO2 into the environment for every MMBtu of coal burned. However, the removal efficiency of the limestone scrubber is 0.98, meaning only 0.02 percent of the SO2 that enters the flue proceeds all the way through the calcium carbonate absorption matrix and exits to the environment. That means that (1-0.98) * 2.2 lb SO2/MMBtu coal, which is equivalent to 0.02 * 2.2 or 0.044 lb SO2/MMBtu coal will be ejected to the environment from this plant configuration, substantially below the uncontrolled configuration. One can see how different removal efficiencies, different coal sulfur contents, and different plant heat rates combine to create different environmental emissions coefficients for the plant. That is evident in the rest of the table.

This calculation needs to be made in IEMM for every power plant and every possible type of coal in the country, region, or system of interest. That is the only way to obtain the correct environmental emissions coefficients, which are going to be needed to characterize regulation, retrofits, modern plant construction, etc., as well as doing all the environmental emissions accounting in the region correctly. There is certainly an opportunity to aggregate a bit and choose a handful of coal compositions and a handful of plant heat rates for every configuration of plant, but the aggregate of fuels and whatever aggregates of plant configurations that are selected cannot omit an important generation option from an environmental perspective.

The green column at the right of the diagram is the sought after answer. The green column represents the pounds of SO2 that are emitted by the various plant configurations per MWh of generation from those plant configurations given the type of coal they burn, the heat rate they have, and the elemental sulfur content of their feedstock. These green numbers are the emissions numbers that are needed to make the calculation implicit in Figure 14.

The real world situation is not quite so simple as the illustration in Figure 14. In fact, installation of one type of remediation equipment (e.g., selective catalytic reduction) will, in fact, remove not only its target (e.g., NOx), but it will also remove to different degrees many of the other reactive chemicals in the flue gas (e.g., SO2, mercury). It is not acceptable to assume that an SCR removes ONLY NOx, its primary target. There are what have come to be termed “cobenefits” of plant retrofit additions. The discussion in these tables has not fully considered the “cobenefit” or synergy issue that occur because the pollution emissions rates will depend on the individual and combination of equipment installed. The proposed IEMM model structure will allow complete consideration of these “cobenefits,” which are synergies among the removal percentages as a function of the combination of remediation equipment that is installed, but we will not be addressing that in detail. The IEMM model illustrations later in this document can consider these cobenefits.

It is necessary to make the analogous stoichiometry and recovery efficiency calculation for NOx for all of the plant configurations and all the coal types. The NOx calculations are summarized in Figure 15. Keep in mind, NOx occurs because of “nitrogen fixation” during combustion. As such, NOx output probably should not be strictly associated with the composition of fuel input. However, this has become a convention because flame temperature can be a function of type and volume of fuel throughput. The green column at the right is the sought after answer—the pounds of NOx emitted to the atmosphere for every MWh generated. These are the figures needed to make the environmental cost calculations and to embed them in the operation costs. Notice that the installation of an SCR knocks NOx emissions down by 90 percent. However, the heat rate is 5 percent worse because of the addition of capture equipment, meaning that there isn’t a strict 90 percent reduction from the uncontrolled case.

Mercury calculations are summarized in Figure 16. The mercury contents of each of the incoming coals are enumerated, the elemental mercury emitted to the environment in the absence of controls is estimated, the control efficiency is estimated, and the pounds of mercury emitted into the atmosphere per MWh of generation is estimated. The green column at the far right is the sought after answer—the pounds of mercury emitted to the environment per MWh generated. Recall that the installation of ACI knocks mercury emissions down by 90 percent.

The CO2 calculations for all of the plant types are summarized in Figure 17. As with the previous calculations, the sought after CO2 calculations per MWh of production are enumerated in the green column at the right.

Figure 15: NOx and Coal Accounting--Illustrative

Plant	Emission Control	Fuel	Heat Rate of Plant (MMBtu/MWh)	Heat Content of Coal (Btu/lb)	Heat Content of Coal (MMBtu/Short Ton)	NOx Removal Efficiency	lb NOx/MMBtu of Coal	lb NOx/MWh Generated
ombined Cycle	None	Natural Gas	7.200			0	1.30	9.3600
ulverized Coal	None	Bituminous Coal	10.000	12,000	24.0	0	3.10	31.0000
upercritical Coal	None	Bituminous Coal	9.500	12,000	24.0	0	3.10	29.4500
ulverized Coal	None	Sub-bituminous Coal	10.500	8,500	17.0	0	3.10	32.5500
upercritical Coal	None	Sub-bituminous Coal	10.000	8,500	17.0	0	3.10	31.0000
ulverized Coal	None	Lignite	11.000	6,800	13.6	0	4.00	44.0000
upercritical Coal	None	Lignite	10.500	6,800	13.6	0	4.00	42.0000
ombined Cycle	SCR	Natural Gas	7.200			0.9	0.13	0.9360
ulverized Coal	SCR+Scrub	Bituminous Coal	10.500	12,000	24.0	0.9	0.31	3.2550
upercritical Coal	SCR+Scrub	Bituminous Coal	9.975	12,000	24.0	0.9	0.31	3.0923
ulverized Coal	SCR+Scrub	Sub-bituminous Coal	11.025	8,500	17.0	0.9	0.31	3.4178
upercritical Coal	SCR+Scrub	Sub-bituminous Coal	10.500	8,500	17.0	0.9	0.31	3.2550
ulverized Coal	SCR+Scrub	Lignite	11.550	6,800	13.6	0.9	0.40	4.6200
upercritical Coal	SCR+Scrub	Lignite	11.025	6,800	13.6	0.9	0.40	4.4100

Figure 16: Mercury and Coal Accounting--Illustrative

Plant	Emission Control	Fuel	Heat Rate of Plant (MMBtu/MWh)	Heat Content of Coal (Btu/lb)	Heat Content of Coal (MMBtu/Short Ton)	Removal Efficiency	lb Hg/MMBtu of Coal	lb Hg/MWh Generated
Combined Cycle	None	Natural Gas	7.200			0	0.0000E+00	0.0000E+00
Pulverized Coal	None	Bituminous Coal	10.000	12,000	24.0	0	1.0450E-05	1.0450E-04
Supercritical Coal	None	Bituminous Coal	9.500	12,000	24.0	0	1.0450E-05	9.9275E-05
Pulverized Coal	None	Sub-bituminous Coal	10.500	8,500	17.0	0	6.7000E-06	7.0350E-05
Supercritical Coal	None	Sub-bituminous Coal	10.000	8,500	17.0	0	6.7000E-06	6.7000E-05
Pulverized Coal	None	Lignite	11.000	6,800	13.6	0	9.7900E-06	1.0769E-04
Supercritical Coal	None	Lignite	10.500	6,800	13.6	0	9.7900E-06	1.0280E-04
Combined Cycle	SCR	Natural Gas	7.200			0.9	0.0000E+00	0.0000E+00
Pulverized Coal	SCR+Scrub	Bituminous Coal	10.500	12,000	24.0	0.9	1.0450E-06	1.0973E-05
Supercritical Coal	SCR+Scrub	Bituminous Coal	9.975	12,000	24.0	0.9	1.0450E-06	1.0424E-05
Pulverized Coal	SCR+Scrub	Sub-bituminous Coal	11.025	8,500	17.0	0.9	6.7000E-07	7.3868E-06
Supercritical Coal	SCR+Scrub	Sub-bituminous Coal	10.500	8,500	17.0	0.9	6.7000E-07	7.0350E-06
Pulverized Coal	SCR+Scrub	Lignite	11.550	6,800	13.6	0.9	9.7900E-07	1.1307E-05
Supercritical Coal	SCR+Scrub	Lignite	11.025	6,800	13.6	0.9	9.7900E-07	1.0793E-05

Figure 17: CO2 and Coal Accounting--Illustrative

Plant	Emission Control	Fuel	Heat Rate of Plant (MMBtu/MWh)	Heat Content of Coal (Btu/lb)	Heat Content of Coal (MMBtu/Short Ton)	Removal Efficiency	lb CO2/MMBtu of Coal	lb CO2/MWh Generated
Combined Cycle	None	Natural Gas	7.200			0	117.0	842.4000
Pulverized Coal	None	Bituminous Coal	10.000	12,000	24.0	0	205.3	2053.0000
Supercritical Coal	None	Bituminous Coal	9.500	12,000	24.0	0	205.3	1950.3500
Pulverized Coal	None	Sub-bituminous Coal	10.500	8,500	17.0	0	212.7	2233.3500
Supercritical Coal	None	Sub-bituminous Coal	10.000	8,500	17.0	0	212.7	2127.0000
Pulverized Coal	None	Lignite	11.000	6,800	13.6	0	215.4	2369.4000
Supercritical Coal	None	Lignite	10.500	6,800	13.6	0	215.4	2261.7000
Combined Cycle	SCR	Natural Gas	7.200			0.9	11.7	84.2400
Pulverized Coal	SCR+Scrub	Bituminous Coal	10.500	12,000	24.0	0.9	20.5	215.5650
Supercritical Coal	SCR+Scrub	Bituminous Coal	9.975	12,000	24.0	0.9	20.5	204.7868
Pulverized Coal	SCR+Scrub	Sub-bituminous Coal	11.025	8,500	17.0	0.9	21.3	234.5018
Supercritical Coal	SCR+Scrub	Sub-bituminous Coal	10.500	8,500	17.0	0.9	21.3	223.3350
Pulverized Coal	SCR+Scrub	Lignite	11.550	6,800	13.6	0.9	21.5	248.7870
Supercritical Coal	SCR+Scrub	Lignite	11.025	6,800	13.6	0.9	21.5	237.4785

All the emissions coefficients calculated for SO2, NOx, mercury, and CO2 in Figure 14, Figure 15, Figure 16, and Figure 17 and the plant heat rates are tabulated in a single table in Figure 18. That is, all the green columns from Figure 14, Figure 15, Figure 16, and Figure 17 and the plant heat rates are tabulated in Figure 18 expressed on a lb/MWh output basis in the green columns in Figure 18. The data in Figure 18 give us what EIA needs to calculate environmentally burdened generation costs **and** to do all the requisite environmental accounting in any regional electrical system. This will be an essential part of IEMM.

Figure 18: Environmental Emissions Coefficients and Heat Rates, and VOCs--Illustrative

Plant	Emission Control	Fuel	lb SO2/MWh Generated	NOx/MWh Generated	lb Hg/MWh Generated	CO2/MWh Generated	Heat Rate (MMBtu/MWh)	Variable Cost ($/MWh)
Combined Cycle	None	Natural Gas	0.0000	9.3600	0.0000E+00	842.4	7.200	$4.00
Pulverized Coal	None	Bituminous Coal	22.0000	31.0000	1.0450E-04	2053.0	10.000	$7.00
Supercritical Coal	None	Bituminous Coal	20.9000	29.4500	9.9275E-05	1950.4	9.500	$7.00
Pulverized Coal	None	Sub-bituminous Coal	14.8235	32.5500	7.0350E-05	2233.4	10.500	$7.00
Supercritical Coal	None	Sub-bituminous Coal	14.1176	31.0000	6.7000E-05	2127.0	10.000	$7.00
Pulverized Coal	None	Lignite	24.2647	44.0000	1.0769E-04	2369.4	11.000	$7.00
Supercritical Coal	None	Lignite	23.1618	42.0000	1.0280E-04	2261.7	10.500	$7.00
Combined Cycle	SCR	Natural Gas	0.0000	0.9360	0.0000E+00	84.2	7.200	$4.20
Pulverized Coal	SCR+Scrub	Bituminous Coal	0.4620	3.2550	1.0973E-05	215.6	10.500	$7.35
Supercritical Coal	SCR+Scrub	Bituminous Coal	0.4389	3.0923	1.0424E-05	204.8	9.975	$7.35
Pulverized Coal	SCR+Scrub	Sub-bituminous Coal	0.3113	3.4178	7.3868E-06	234.5	11.025	$7.35
Supercritical Coal	SCR+Scrub	Sub-bituminous Coal	0.2965	3.2550	7.0350E-06	223.3	10.500	$7.35
Pulverized Coal	SCR+Scrub	Lignite	0.5096	4.6200	1.1307E-05	248.8	11.550	$7.35
Supercritical Coal	SCR+Scrub	Lignite	0.4864	4.4100	1.0793E-05	237.5	11.025	$7.35

Following the assembly of plant environmental coefficients, each node in IEMM must obtain from the environmental supply sector (where the emissions allowances are "supplied" in an economic sense) the prices of SO2, NOx, mercury, CO2, and fuels. Keep in mind, even though the illustrative calculation herein uses representative prices, the IEMM model designed here will, under cap and trade regimes, be iterating on these prices, and each iteration of the calculation outlined here will be made using the then-current estimates of market-clearing prices. (IEMM may not need to iterate for regimes other than cap and trade.) When the iteration is complete, those prices will be the **right** estimates of market clearing prices for both emissions and electricity. That is crucially important, and an important capability of IEMM.

2.3 The Economic Meaning of an Unregulated, Uncontrolled Scenario

What does it mean for pollutants to be unregulated? The economics books are quick to tell us it means that allowances to produce those pollutants are **free**, zero price, no economic value and no economic cost. In such a situation, the prices we would see in the IEMM model for the region of interest are those in Figure 19.

If we made the economic calculations only, multiplying the various prices from Figure 19 by the input-output coefficients from Figure 18 and adding the variable operating costs from Figure 18, we could calculate the economic-only operation costs for every plant configuration. The environmental cost calculation would be calculated as in Figure 20. We can see the mnemonic we use—yellow for SO2, gray for mercury, brown (as in smog) for NOx, and green for CO2. We try to continue that mnemonic here.

Figure 19: Fuel and Emissions Prices in an Unregulated, Unconstrained Power System

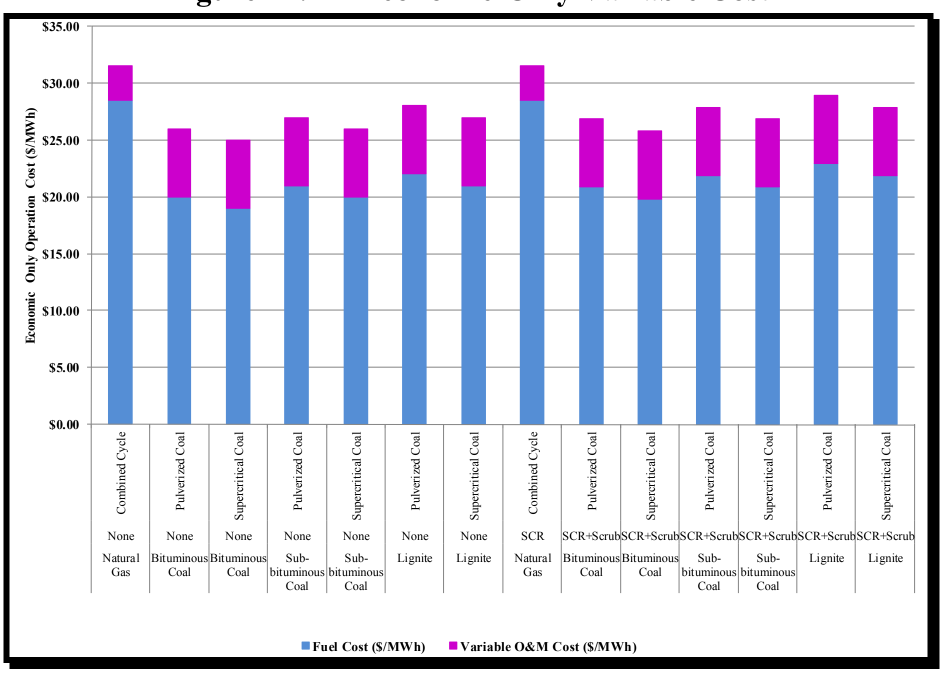
SO2 Price ($/ton)	NOx Price ($/ton)	Hg Price ($/ton)	CO2 Price ($/ton)	Gas Price ($/MMBtu)	Coal Price ($/MMBtu)
$0	$0	$0	$0	$4	$2

Figure 20: Calculating Environmental Cost

$$
\begin{aligned}
\text{EnvCost} = 0 = {} & p_{SO_2}\frac{\$}{\text{lb } SO_2}\times IO_{SO_2}\frac{\text{lb } SO_2}{\text{MWh}} + p_{NO_x}\frac{\$}{\text{lb } NO_x}\times IO_{NO_x}\frac{\text{lb } NO_x}{\text{MWh}} \\
& + p_{Hg}\frac{\$}{\text{lb Hg}}\times IO_{Hg}\frac{\text{lb Hg}}{\text{MWh}} + p_{CO_2}\frac{\$}{\text{lb } CO_2}\times IO_{CO_2}\frac{\text{lb } CO_2}{\text{MWh}}
\end{aligned}
$$

The array of those calculations appears in Figure 21 for every plant in Figure 18. These calculations are made in IEMM directly for every plant in any region or country of the world. These calculations have zero pollutant prices and, therefore, zero internalization of the environmental externalities.

Figure 21: Economic-Only Variable Cost



Notice in Figure 21 that the gas plants variable costs are high as compared with coal plants. The economic-only variable costs of the gas plants—retrofit or non-retrofit—are substantially higher than the economic-only variable costs of the coal plants. Higher costs for gas plants has occurred for many years in North America and throughout the world (but not in the recent two years in which gas price has sunk below coal price). Lower variable cost is why there are so many coal plants in North America and other places in the world and why the coal plants run in baseload service, and the gas plants run in cycling and peak load service. However, these cost estimates in Figure 21 do not incorporate or "internalize" any of the environmental costs as developed in this section thus far. They need to internalize those costs, and we now proceed to calculate how they do so. These are the costs that dispatch models tend to use for the various power plants configurations, which we term non-environmentally burdened costs.

To address the notion of environmentally burdened or adjusted costs, suppose we had used the prices in Figure 22, which are representative of what might occur in a regional IEMM model. If we saw this in a region, we would know that some regulatory authority priced all pollutants, perhaps directly with a pollution tax, indirectly with cap and trade, or in a more sophisticated fashion using CPP or similar regulation.

Figure 22: Emissions Allowance and Fuel Prices (Illustrative)

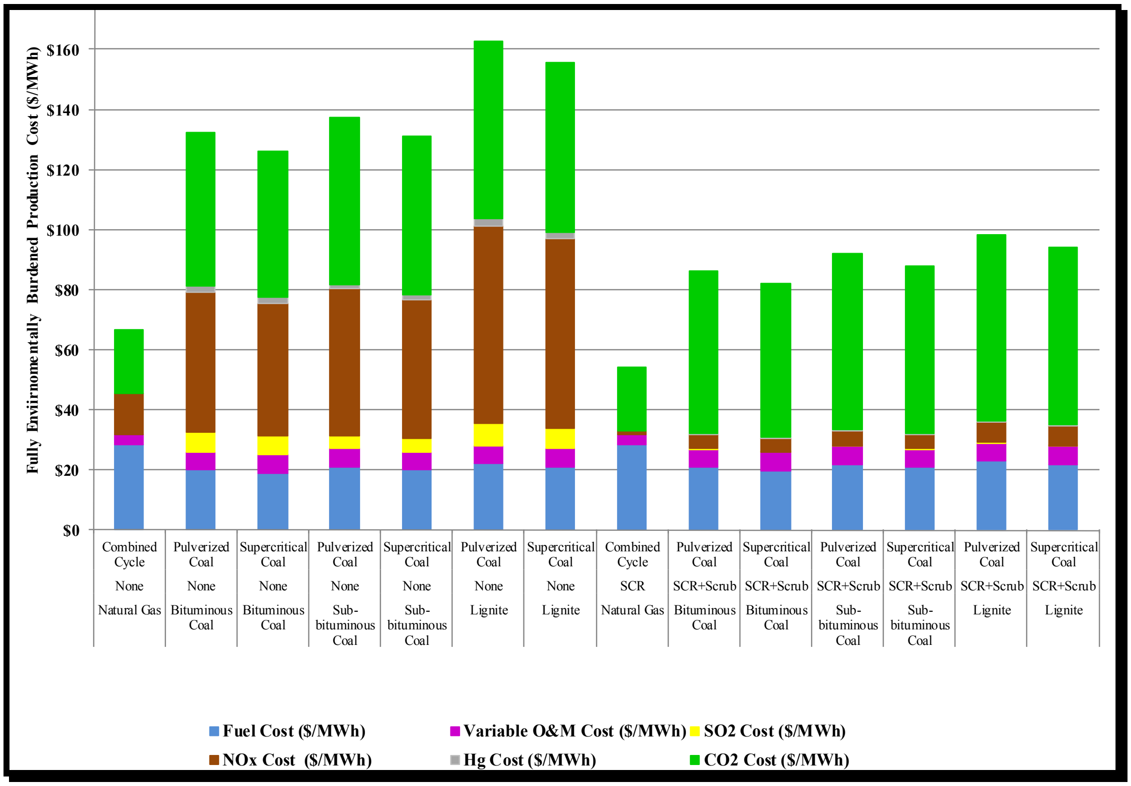
SO2 Price ($/ton)	Nox Price ($/ton)	Hg Price ($/ton)	CO2 Price ($/ton)	Gas Price	Coal Price
$600	$3,900	$40,000,000	$50	$4.00	$2.00

IEMM multiplies the emissions allowance prices in Figure 22 by the emissions coefficients in Figure 18 before and after retrofits respectively for every one of the plants within an IEMM market model (or any other region or system of interest). Doing so makes the calculation in Figure 23 to perform the calculations summarized in Figure 24, to which we now turn in some detail. We should emphasize that we have excluded CO2 removal retrofits for the moment in Figure 24. We will layer them on after we discuss the other pollutants.

Figure 23: Fuel and Emissions Prices in a Regulated Power System

$$0 \le EnvCost - p_{SO_2}\frac{\$}{\text{lb }SO_2}\times IO_{SO_2}\frac{\text{lb }SO_2}{MWh} + p_{NO_x}\frac{\$}{\text{lb }NO_x}\times IO_{NO_x}\frac{\text{lb }NO_x}{MWh} + p_{Hg}\frac{\$}{\text{lb Hg}}\times IO_{Hg}\frac{\text{lb Hg}}{MWh} + p_{CO_2}\frac{\$}{\text{lb }CO_2}\times IO_{CO_2}\frac{\text{lb }CO_2}{MWh}$$

Figure 24: Environmentally Burdened Variable Costs of Generation Plants without CO2 Sequestration



In this equation, input-output designates the input-output coefficient for the emission, i.e., the pounds of emission emitted to the environment per MWh of generation. These input-output terms are the calculated input-output coefficients from Figure 18.

The economic cost of the plants is the sum of the bottom two bars, the same pink and blue bars that were shown in Figure 21. The environmental emissions allowance costs are represented by the sum of the yellow (SO2) plus brown (NOx) plus gray (mercury) plus green (CO2) bars. The plants on the left in the diagram are non-retrofit or uncontrolled plants, and the plants on the right are retrofit, environmentally controlled plants. This calculation is using the emissions coefficients developed in Figure 18. There is no bypassing these emissions coefficients for thermal systems around the world. And yet they are quite well known and quite peculiar to electricity generation. They are readily extrapolated from regions such as North America, in which they have been measured, to various developed and developing country locales around the world in which the plant configurations and prime movers are similar and in fact quite well known. Thermodynamics is thermodynamics. Chemistry is chemistry. Physics is physics. Combustion is combustion. A steam turbine is a steam turbine is a steam turbine. When you put a hydrocarbon in the front door, you are fairly certain what is going to come out the back door. Plants are not specific or peculiar to one region. A gas turbine in one region is not intrinsically different from a gas turbine in another region. EIA will rely on technological commonalities as it assembles environmental input-output coefficient data for worldwide generator sets.

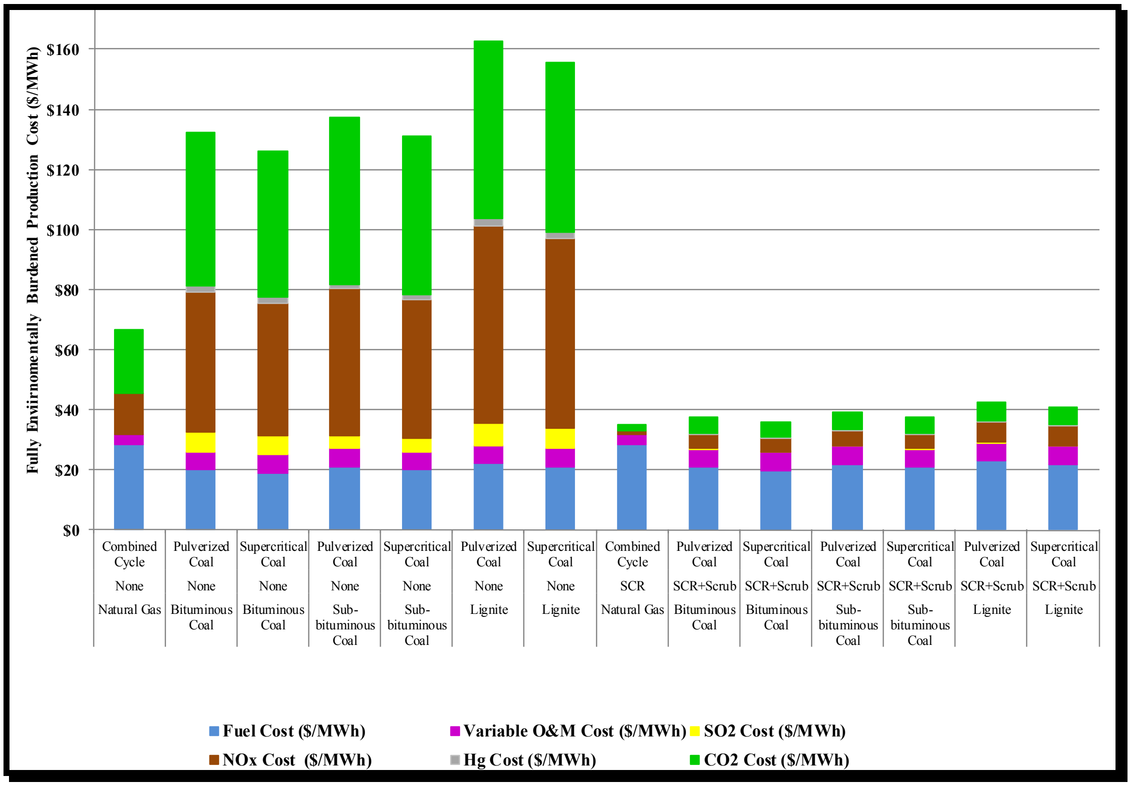
Notice that even if the price of the emissions allowances is zero, the emissions accounting (i.e., how much of every type of pollutant in the slate) is released into the environment remains operational and intact within the IEMM design herein. That is precisely why even for uncontrolled systems, IEMM needs to use this allowance or entitlement model of emissions (inputs to power plants rather than outputs from power plants or some more arcane type of environmental model.) We return to this theme repeatedly.

Note in Figure 24 that the environmental costs can be rather large relative to the economic costs, which are represented in the bottom two bars, particularly for non-retrofit plants (and not shown in the diagram and not analyzed herein yet) for high heat rate plants. The less efficient the plant, the more fuel you have to put through it, the more "atoms" you have to put through it, and the more atoms of effluent it produces. It is precisely the point of cap and trade, pollutant tax, MACT/BACT, CPP, and NSPS/ESPS environmental regulation to render the bars for the higher polluting plants so large that those plants are relegated to reduce operations, thereby reduce output, and thereby reduce pollution either through shutting down, retrofitting, or reducing their level of output. This type of phenomenon is not isolated to the Unites States. It happens today in many power systems around the world (how did the SO2 that was harming the Black Forest get abated? It was evidently a combination of fuel switching and SO2 remediation), and it is likely to increase under scenarios of increased environmental concern in countries and systems that have not formerly had such concern. Environmental remediation is "in the wind" around the world, and EIA's IEMM must correctly perform the environmental accounting and the environmental incentives. CO2 is in the wind in many countries around the world, as we witnessed recently during the Paris conference. This IEMM design will allow EIA to address such issues.

It is elucidating to represent what would happen if CO2 removal is also implemented in the retrofit cases on the right in Figure 24. The result appears in Figure 25. Notice that the CO2 cost even at $50/ton is substantially arrested. The externality has largely been internalized.

The decisions plant owners in IEMM make regarding plant operation fully embeds the preceding environmental costs. Under certain types of regulation, such environmental costs can be passed through to parties other than ratepayers without elevating rates (e.g., tax subsidized costs). Under other types of regulation, such environmental costs are passed through in the form of increased electricity rates or taxes on particular fuels. The proposed IEMM design will allow both as we shall see shortly. In either case, they decrement from the profitability of the plant, **and** they markedly impact market prices of electricity. In most real world thermal generation systems, one cannot operate plants based on the purely economic considerations implicit in the two bars of Figure 21, which are the bottom two bars of Figure 24, i.e., the fuel and operating cost only. The days of doing that are long since behind us in many systems, certainly in all the industrialized countries of the world and increasingly in developing countries. One must operate plants based on the economic considerations in the two bars at the bottom of Figure 24 **plus** the environmental emissions allowance costs as represented in the top bars of Figure 24. It is the total height that drives the decision whether to operate or shut down, not just the bottom two. IEMM has to be able to undertake fully environmentally burdened operation (as well as fully unburdened or pretty much any hybrid in between), for that is precisely the way that emissions taxes, cap and trade, and much of the command and control style of emissions regulation such as CPP works. Only if IEMM does this correctly can one calculate market prices of power and fuels (gas, coal, nuclear fuel, heavy fuel oil) correctly. The IEMM design herein allows that.

Figure 25: Environmentally Burdened Variable Costs of Generation Plants withCO2 Sequestration



IEMM must make the calculation in Figure 24 for every one of the generating units in the United States (individually or in aggregates) or in any power system being modeled, each with their own individual emissions and fuel and operating characteristics. It then must create the short run supply curve for that plant using the fully environmentally burdened operation cost and replace the economics-only cost that formerly appeared in Figure 7. The economics only curve from Figure 7 is reproduced in Figure 26. The environmental emissions costs are incorporated in the figure as calculated just above, and those environmental emissions costs elevate the economics curve straight upward as indicated in Figure 26. Because each plant has different environmental input-output coefficients, the degree of elevation of the short-term marginal cost curve is plant-specific and potentially quite different across plants.

This point that environment and economics are intertwined and enmeshed in IEMM cannot be overemphasized. If we think of the "supply stack" of plants in a given region based on purely economic costs, we think of a plot conceptually like that in Figure 27. The nuclear, renewable and hydro plants are at the left. Next come the coal plants. The gas combined-cycle plants are in the middle of the supply stack, followed by the old gas steam turbine units more to the right and finally further to the right the gas and oil combustion turbine and internal combustion peakers. The lineup of plants would be strongly associated with fuel price and with plant age if nothing but economics were taken into account in variable cost. Alas, however, economics is not sufficient in the world in which environmental costs are being fully internalized in the power generation system via cap and trade or taxation or other methods, and that must be accounted for in IEMM.

Figure 26: Environmentally Burdened Short-Run Supply Curve (Marginal Cost Curve) for the Plant

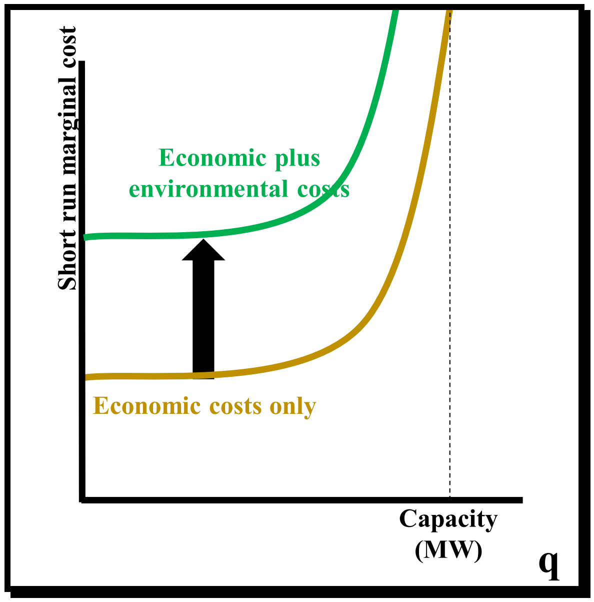
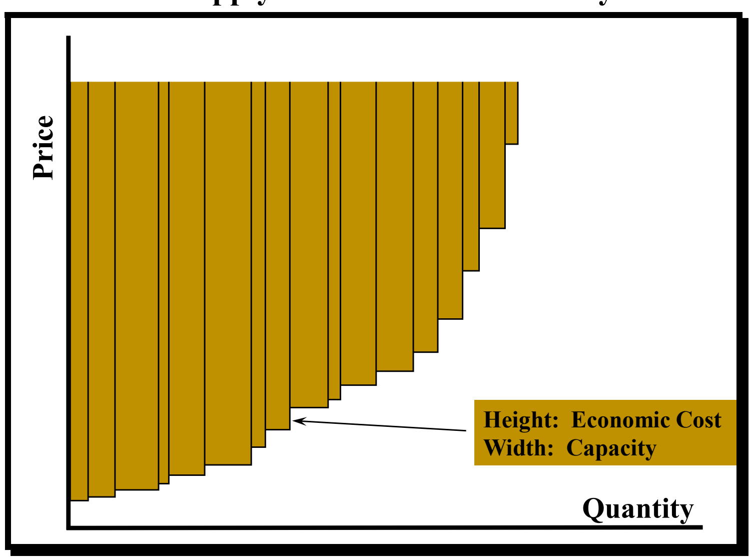


Figure 27: The "Supply Stack" Based on Strictly Economic Costs



In precisely the sense of Figure 7 being elevated to the environmentally burdened curve in Figure 26, each individual plant in the supply stack in Figure 7 is individually impacted by environmental costs based on its particular technology and configuration, its particular set of environmental input-output coefficients, and the prices of pollutants that it faces. The environmental input-output coefficients can be and often are dramatically different for different plants, and such variation is critically important for determining new, environmentally modified supply stacks. For example, if a plant has been retrofit for "selective catalytic reduction" or SCR, its NOx input-output coefficient is reduced by some 90 percent, as characterized in the numerical illustrations in Figure 15. If a coal plant has been retrofit with a scrubber, its SO2 output is reduced by 98 percent or so, as characterized in the previous numerical illustrations in Figure 14. If a plant has been retrofit with activated carbon injection, its mercury output is reduced by some 90 percent as illustrated in Figure 16. Older, non-retrofit plants that have not been retrofit, even though they might burn the same fuel, are penalized to a much higher degree because they must buy or acquire many more emissions allowances to offset the much higher quantity of pollutants they produce because of their generally higher heat rate. A plant with a 16,000 Btu/KWh heat rate will buy and burn a lot more fuel than a plant with a 10,000 Btu/KWh heat rate and will, therefore, spray a lot more atoms into the air. The notion of paying money to retrofit plants to achieve better input-output coefficient is going to come squarely into play in IEMM. Other methods completely miss this—environmental costs affect power prices and rates to consumers.

When the environmental penalties are applied unit by unit to the supply stack in Figure 27, one obtains the recalculated supply situation in Figure 28, i.e., the environmentally adjusted supply stack. Relative to Figure 27, coal plants are penalized to an extreme degree in terms of higher production cost. They must buy SO2, NOx, mercury, and CO2 emissions allowances. Nuclear and renewables plants are not affected because they do not need to buy any emissions allowances. (This is why wind and solar and nuclear have so much environmental benefit; they don't have to buy any emissions allowances because they do not emit and, therefore, have no emissions to offset.) The more efficient gas plants have to buy relatively smaller quantities of NOx and CO2 allowances. Less efficient gas plants (including peakers) are hit harder because they produce more NOx and CO2 that have to be offset with allowances in higher quantities. Oil plants have to buy allowances commensurate with the heavy fuel oil or diesel that they buy and burn, and they can emit SO2 and mercury. Plants that would burn ultra-high-sulfur residual oil (such as in Mexico burning combustolio) would have to spend a lot of money on SO2 allowances. All such calculations are highly individualistic on a plant by plant basis. Without the plant by plant calculation, we cannot know the true operating cost of every plant and, therefore, cannot know the environmentally adjusted supply stack for plants in the given region. This design allows careful plant by plant accounting and costing, which IEMM most definitely needs.

After the plant by plant by plant cost adjustments from Figure 27 to Figure 28 are made to account for environmental costs, the brown supply stack is transformed into the green supply stack. After reordering of the environmentally burdened supply stack in ascending order of environmentally burdened production cost, whose production costs reflect both economic costs plus environmental emissions costs, we obtain an adjustment from the brown curve in Figure 29 to the green curve in Figure 29. If one does not make the calculation from the brown curve to the green curve carefully, accurately, and individually for each and every plant, one does not get the supply stack right. If one does not get the supply stack right, one does not get plant operation nor the price of power right (which is seen in Figure 29 to be substantially higher with the environmental penalties than without), one does not get the operation of plants right, and one does not get plant profitability right. That will not be acceptable for IEMM. We will show later that one does not get the rate of capacity additions right either. IEMM needs to get all of those right so that all its prices are calculated correctly and, therefore, can serve EIA's electrical system analysis and projection requirements.

Figure 28: Environmentally Adjusted Supply Stack

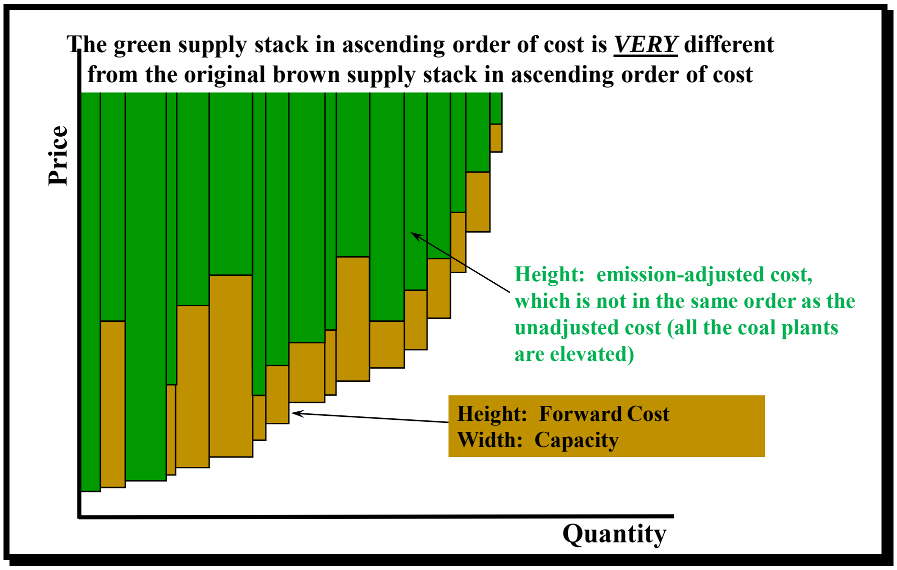
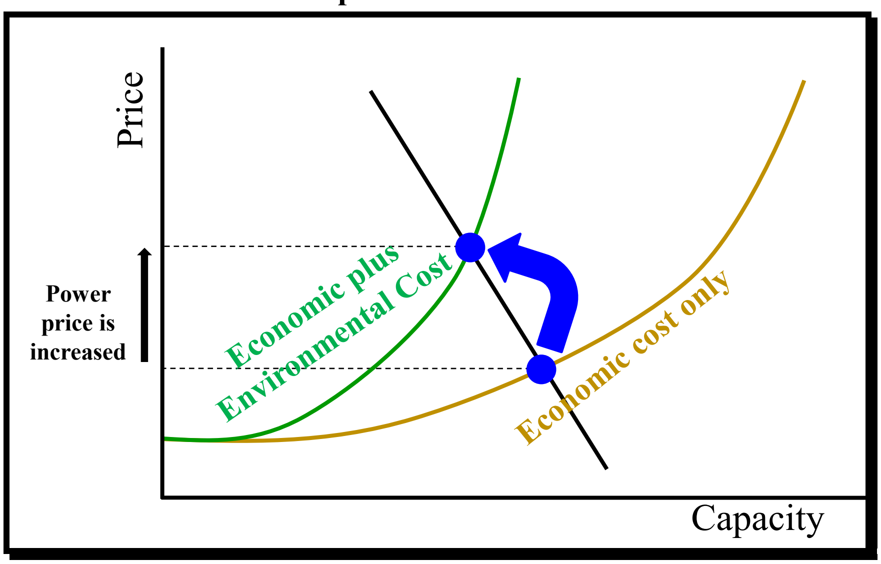
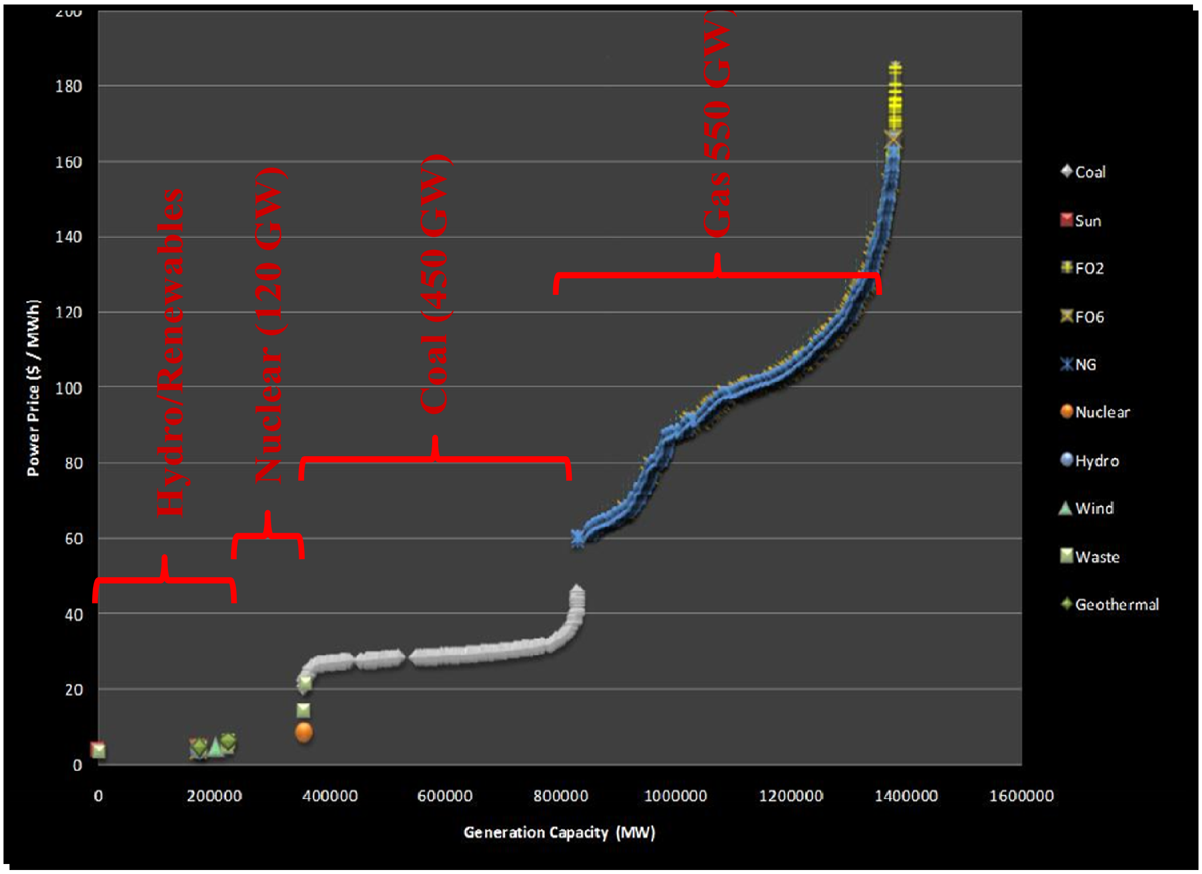


Figure 29: The Economic Supply Stack and the Environmentally Adjusted Supply Stack and Their Implications for Market Prices



It is instructive to plot the brown curve for the entirety of North America (at a point in time about fifteen years ago, accumulated and plotted by Michael Blaha as part of our Economic Training Program.). It appears in Figure 30. The curve requires a detailed calculation, but precisely the one summarized previously in this section. The curve included every generation station in North America and included their full slate of cost variables and environmental coefficients. The hydro and renewables capacity number at the left requires additional interpretation because those resources are energy limited and do not typically operate at nameplate capacity. The thermal plants of each type are indicated and color coded across the page, and they are pretty much in the standard order—hydro and renewables, nuclear, coal, gas, oil. The diagram is very intuitive. This diagram is the brown curve in Figure 29, the non-environmentally adjusted supply stack for North America.

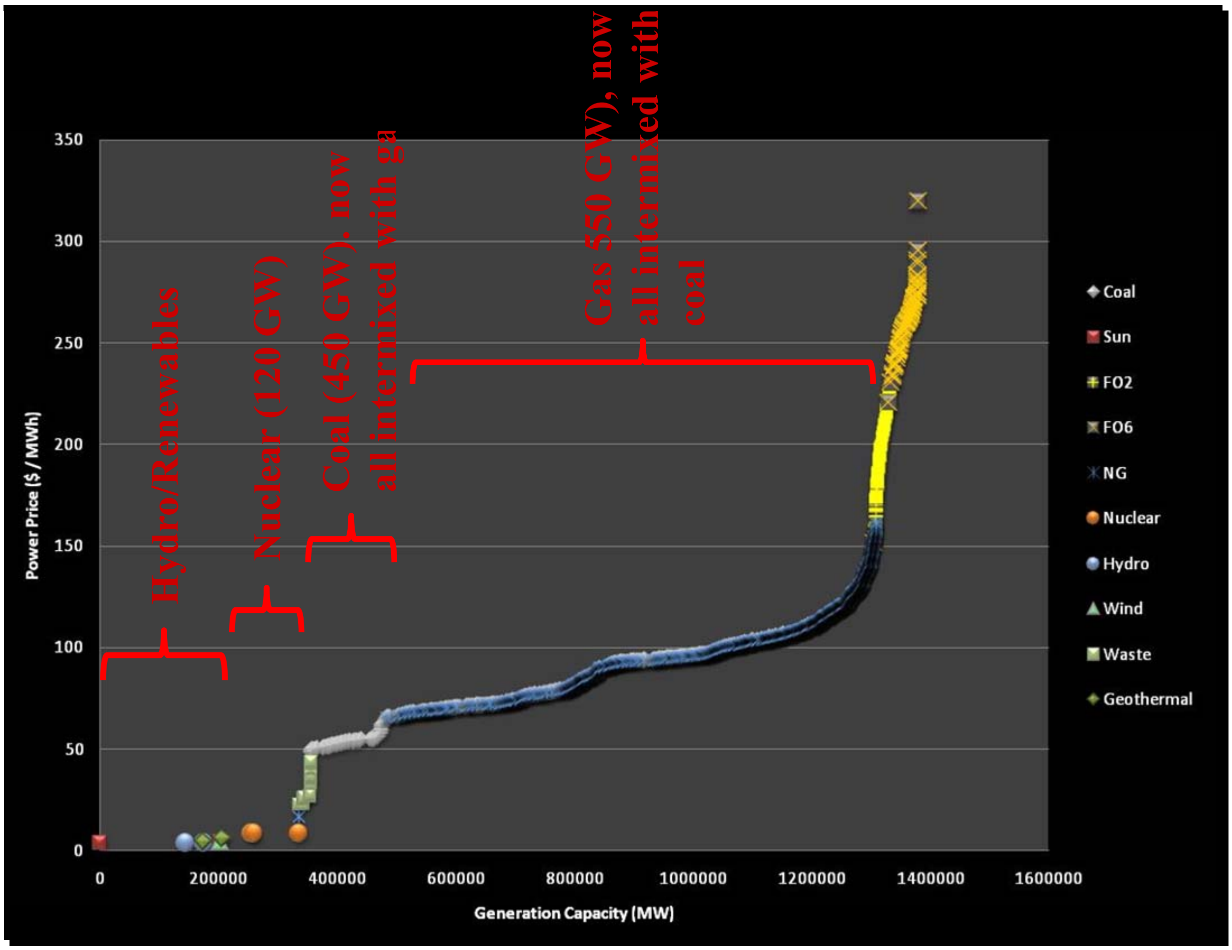
Figure 30: The North American Electric Generation Supply Stack on a Purely Economic Basis



How does this supply stack look after the imposition of environmental emissions allowances, including a $20/ton CO2 tax? The answer is very insightful for plant operation and plant profitability and modeling electricity generation, transmission, and consumption. It is crucial to the success of IEMM. It is crucial to getting good market and price information. The environmentally adjusted North American supply stack, plotted by Michael Blaha some fifteen years ago, appears in Figure 31.

The height of the entire stack has been shifted upward to a significant degree because of the internalized CO2 environmental cost. Be clear that the scale of the vertical axis has substantially changed between Figure 30 and Figure 31. The scales on the two curves are different because of the upward shift in cost when environmental costs are included. At first blush, it may appear that we have lost most of the coal plants. In reality, those plants have not been lost. In reality, those coal plants have become directly intermixed and interleaved with the natural gas plants in blue, which are not penalized by CO2 emissions allowance cost and other environmental emissions allowance costs to anywhere nearly as great an extent as coal. Coal has moved away from its former cost-preferred status in Figure 30. It has moved to the margin (and partially out of the energy mix) even at a $20/ton carbon tax, intermixing all along the way with the blue gas plants in the diagram. The impact of such an environmental tax is profound and shows why a full plant by plant modeling approach is required within IEMM. One cannot possibly model power prices without quantifying and understanding this supply stack (in reality on a highly regionalized basis, not this national total supply stack).

Figure 31: The North American Electric Generation Supply Stack on an Environmentally Adjusted Basis



Calculating environmentally adjusted regional supply stacks such as these is an important part of what the IEMM technology needs to do—it needs to modify the supply stack and calculate its logical implications for market prices of power on an interconnected but multiregional basis. It is important to reiterate the scope of this supply stack and how challenging it is to get it and calculate it. It requires knowledge of all the cost and all the environmental emissions coefficients for every one of the generators that comprised the system being modeled, all assigned to their appropriate regions.

The modification of the plant production cost for every plant in the stack as a function of the then current (perhaps model-iterated) emissions prices or any emissions regulatory signal is the central point of this section. This section has shown what environmental emissions allowance prices do to the supply stack. The next section overlays the additional dimension of renewables, renewables credits, and RECs. We thereafter turn to the question of how the prices of these allowances are formed and from whence emanates the supply of emissions allowances. That will take us to the conclusion for thermal generation.

2.4 The Addition of Renewables to the Picture with the Possibility of Tradable Renewable Energy Credits (RECs)

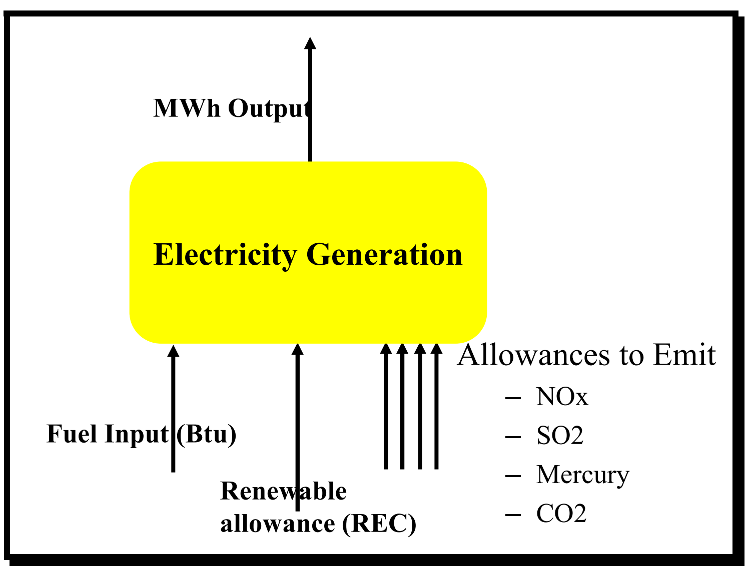
Renewable Energy Credits (RECs) represent a right to generate MWh by virtue of a contractual and economic connection to a "green" or an environmentally friendly generating facility. Renewable Energy Credits have been referred to as green tags, Renewable Energy Certificates, Tradable Renewable Certificates (TRCs), and perhaps other monikers around the world. This section summarizes briefly how they trade and how they should be incorporated into IEMM. The technique is quite akin to the way that environmental emissions allowances are incorporated and potentially traded, but there are subtle differences. Of course, there are a number of subtleties and different ways that renewables and tradable or non-tradable (mandated) renewables can enter the market, but if we can model them in the most sophisticated context of tradable RECs, then we can model them in simpler contexts such as command and control regulation, taxes, subsidies, or mandates. We think of a REC as an allowance to generate based on the demonstrated existence in the market of some accompanying quantity of a "green" source of energy.

RECs are considered a commodity separate from power produced—producers of "green" power generate and can sell RECs as well as electrical energy per se. What is "green" power? That is strictly a question of regulation and definition. Some government body declares what sources of power qualify as "green" power. Almost every jurisdiction of which we are aware would designate wind and solar as "green," and most jurisdictions would designate geothermal as "green." Some would designate incineration of waste as "green," although some would not. Some jurisdictions might consider landfill biomass as "green" or ethanol from waste or crops to be "green." Some jurisdictions would consider run of river, reservoir, or pumped hydroelectric to be "green." However, others would not, wanting to dismantle hydro dams and create free running rivers again. Regardless, one must craft a list of sources that are considered "green" from a jurisdictional perspective. These sources are often called "qualifying" sources, implying that they must legitimately qualify as meeting the definition of "green" under the jurisdiction in which they operate. We term the electrical energy generated by the qualifying "green" source as "qualifying electrical energy."

RECs enter the picture in the following way. For every MWh of electrical energy that a "green" source delivers to the market, i.e., to the busbar, the "green" source prints a piece of "scrip," a paper certificate in effect, that bears the denomination 1 MWh. That piece of scrip that bears the denomination 1 MWh is a REC. The REC is the piece of paper, the scrip that demonstrates that 1 MWh of qualifying "green" energy has been delivered to the market. To emphasize, for each MWh of qualifying "green" or renewable energy that a qualifying facility generates to the market, that renewable facility owner would generate one REC of 1 MWh, i.e., he or she would get the right to print a 1 MWh REC "scrip." MWh of scrip are generated one for one as MWh of green energy are generated. There are exactly as many MWh of scrip printed and circulated as there are MWh of qualifying "green" energy MWh generated and delivered to the market. It is as simple as metering the amount of qualifying "green" energy delivered.

As those REC scrip certificates are generated, the local (or national) regulator would simply require that every "non-green," non-qualifying thermal generator in a region must obtain, present, and surrender say 0.25 RECs for every 1 MWh of non-qualifying thermal electricity he or she chooses to generate. Every non-green MWh would have to surrender 0.25 "green" MWh of RECs as a condition of generation. The renewable energy generator would sell his or her 1 MWh of REC on the open market, and that 1 MWh of REC would allow 4 MWh of thermal energy to be generated by non-qualifying thermal facilities that are vying to generate. When RECs are traded and ultimately held by thermal generators, the entity purchasing the REC gains the right to generate and thereby consumes those RECs (or hoards and holds them and thereby claims environmental benefits by eliminating even more thermal electric generation). This situation of non-qualifying generators having to surrender a REC is depicted in the expanded plant model in Figure 32. As indicated in the figure, the non-qualifying thermal generator in IEMM will require fuel, the full slate of emissions allowances, one for each regulated emission, and a REC. The purpose of the REC is to ensure that renewables participate in the market to a preordained fraction as a condition of non-qualifying thermal generators to generate. This preordained fraction is the notion of a Renewable Portfolio Standard or RPS. We are going to see that this is an extremely general model that allows IEMM to characterize the full range of economics, environment, and renewables in its generation nodes and thereby allow a sophisticated and accurate representation of power generation. This plant structure is going to allow a richness of treatment of renewables.

Figure 32: Each Generator Purchases and Surrenders Emissions Allowances and a REC Consistent with Its Level of Generation



2.5 From Where Do Emissions Prices Come?

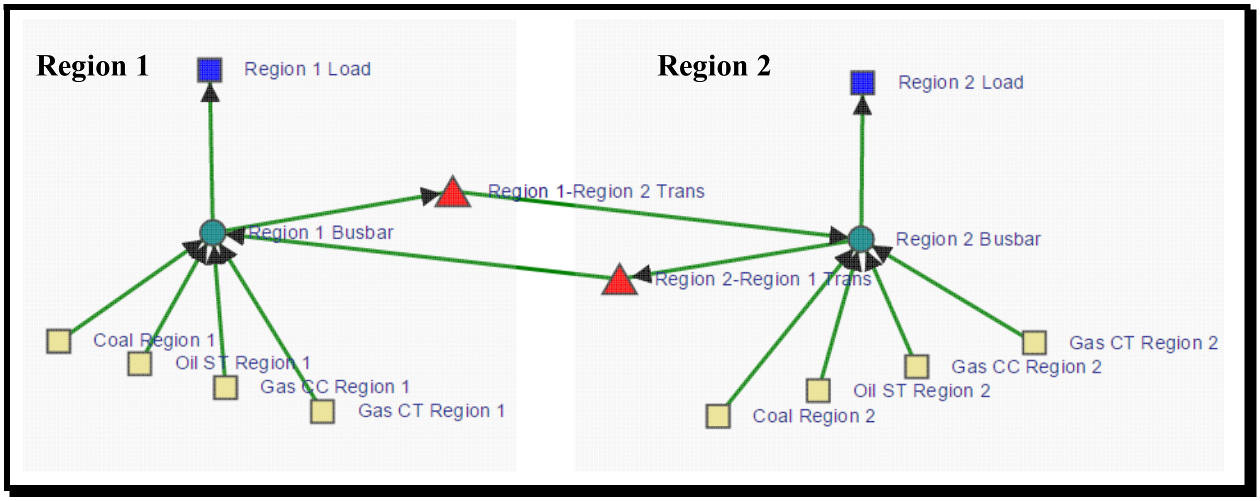
In many, in fact most, jurisdictions around the world, emissions prices are zero at present. In many electricity systems in the world, there is no control exercised over emissions from generation on an ongoing basis, and there is no mandate for "green" entry. (Such is true for CO2 in the United States today except in California under AB 32.) In such situations, the emissions prices for those uncontrolled pollutants or renewables are zero by construction. Zero prices is exactly what economists mean when they say something in the environment is uncontrolled. You can pollute with impunity and pay nothing. The beauty of the generation node design we have put forth is that it can easily give IEMM an accurate accounting of the magnitude of emissions by plant and thus total emissions throughout the region or country even if the assumed emissions or REC prices are zero. Zero price is one case, but not the only required case for IEMM. That is crucially important for this design.

Nonzero emissions prices can emanate from pollutant taxes, from cap and trade regulation that may be occurring within the region or country, from MACT/BACT (which, to reiterate, stands for minimum available control technology/best available control technology) mandates that may be in place, from NSPS/ESPS regulations (which stand for new source performance standards/existing source performance standards) that may be in place, or from CPP style regulations and have emerged recently in the United States. The latter is rather a draconian command and control regulation, and it sends quite a strong signal on the pollutant input links (and as we shall see in the capacity addition and retrofit logic that is attached to these various generation nodes in discussions below). The pollutant input links carry these regulatory signals to the generation nodes. Bans and mandates to control or eliminate a particular pollutant are also carried in as costs of allowances to emit a given pollutant. This construct of carrying a price signal to a generator, i.e., to the generation node, related to pollutants and renewables is a powerful way to get renewables and the environment into the generation operation and capacity addition decisions. That is precisely the point of environmental regulation—to change the way that plants are built and operated.

We begin our discussion of how to model environmental regulation by using signals sent to generators with cap and trade and after that move to signals sent under exogenously imposed pollutant tax and command and control regulations. Cap and trade is an economically efficient style of regulation that occurs in some jurisdictions around the world and must be modeled within IEMM. Once it is modeled in IEMM, the full range of other environmental regulations becomes very easy and straightforward as a modeling special case of the logic developed for cap and trade.

The situation before the imposition of environmental regulation is depicted in Figure 33. The only thing that previously crossed between and among generation regions in those pre-regulation days was transmission. Disparate regions were managed and operated individually by utilities, merchants, and regulatory bodies within those regions. Regions were completely disconnected except by transmission and power pooling and interconnection agreements, which have been established for reliability and for trading reasons. Operation was a strictly local issue. Profitability was a strictly local issue. There were no overlapping considerations that crossed between regions.

Figure 33: The Power Business Before Environmental Regulation Is Local and Balkanized



2.5.1 Endogenous Pollutant Pricing Under Cap and Trade Regulation

Cap and trade regulation (whether based on allocation/assignment of allowances or original auction of allowances) changes the situation to that depicted in Figure 33.[7] We are going to simplify, exaggerate, and caricature how cap and trade regulation works and what its economic rationale is so as to foster careful understanding not only at the conceptual level but also at the modeling level. Following the caricature, we then get specific and detailed about the types of programs currently in place or have been put in place since the mid-1990s in the United States and Europe and perhaps now and in the future elsewhere and how they affect the power businesses in those venues.

One of the motivations for cap and trade regulation is that it does not require government mandates other than the cap. It is designed to be economically efficient as it reduces aggregate pollution to fall below a regulator-specified cap. Implementing cap and trade regulation within IEMM will be the most difficult regulatory framework to implement. All other forms of environmental regulation can be represented as easy special cases. That is the motivation for articulating in detail cap and trade first and then specializing to the other important, prevalent forms. No one should think that the proposed IEMM design will be intrinsically married to and only consider cap and trade regulation. It is not. Cap and trade is the most difficult to represent, and other forms that occur around the world can be achieved as simplified special cases.

Following is a caricature of how cap and trade regulation has worked for SO2 in the United States, quite an accurate and intuitive caricature in fact. It is the way CO2 cap and trade regulation works in California under AB 32, the assembly bill that originated it.

[7] Allocation/assignment and original auction are mechanisms for placing emissions allowances into circulation. We will address those later in this discussion. For now, we simply assume that the allowances exist in the market and area tradable and traded.

1. A government policy body passes a law or regulation that enables cap and trade regulation. That law empowers and instructs the regulatory body (in the United States, it has been the EPA, EPA, or the California regulatory body) to implement, administer, and enforce the law. Alternatively, a court mandates or upholds or extends a regulation and mandates the regulatory body such as the environmental regulator to administer it. In the United States, Congress passed the Clean Air Act, the President signed it, and the environmental regulator was empowered to administer it. Such administration carries with it a tremendous degree of definition as well.

2. Using SO2 as an example, the environmental regulator decides how many tons of SO2 they will allow to be produced in a given year such as 2016, say 7 million tons in that given year. An upper bound is the "Cap" part of cap and trade. The environmental regulator establishes the cap in physical units (tons of SO2, tons of NOx, tons of mercury, tons of CO2) that shall not be exceeded by the aggregate of all the generators in the market, whether qualifying or non-qualifying generators.

3. In the case of SO2, the environmental regulator then prints up 7 million 1-ton emissions allowance certificates, i.e., 7 million pieces of "scrip." Each scrip certificate says: "The bearer of this certificate is hereby entitled to emit one 1 ton of SO2 during the year 2016." The emissions allowance scrip is required as a precondition for operation of any generation unit that produces SO2 to the environment.

4. The environmental regulator decides if there will be any additional allowances that can be generated by planting banana trees in the Philippines or other approved offsetting acts. The regulator decides if allowances from another jurisdiction can be used to offset pollution in this jurisdiction. These are methods whereby adjustments in the regulator-established cap are allowed or not. Offsets have the effect of increasing the number of allowances in circulation above and beyond the cap and de facto simply making a new cap.

5. The environmental regulator then places those 7 million certificates into circulation. They can do so using what is termed an "allocation or assignment method," which entails merely handing allowances out for free to plant owners without cost approximately pro rata based on how much they emitted last year, according to whatever rule or practice they might want. The key observation is all allowances are handed out for free to agents in the market. The specific scheme for deciding who gets what is immaterial to the operation of the IEMM model. However, it has huge impacts on distribution of benefits and costs under cap and trade regulation. IEMM will keep track of who gets the initial endowment of allowances.

   Alternatively, regulators can use what is termed an "auction method," in which the environmental regulator prints up official, certified, legal "scrip" and auctions such scrip certificates from a central repository to the highest bidder. Original auction is how AB32 works in California. The state government sells the scrip certificates, all proceeds going to the state.

   To date in the United States, environmental scrip certificates have generally been "allocated" among existing generators using some type of pro rata type of allocation scheme. In pending and anticipated environmental regulation, there are proposals not to use the allocation or assignment method (hand them out for free) but rather to force every generator to bid for them from a common auction pot, the auction money going to the coffers of the regulating government. The auction method is what California AB 32 uses. There is economic impetus to implement the auction method (which some have termed the "original auction") of putting emissions allowances into circulation because it ensures that emissions allowance prices are internalized into power prices from upstream all the way to downstream and consumption. For those who truly want to internalize the externality, the original auction method does precisely that. However, it forces end-users to pay the entire externality cost, and many end-users have resisted that politically.

6. The environmental regulator allows those emissions allowances to be "burned up" by the plant owner who ultimately acquires them, each allowance compensating and offsetting 1 ton of SO_2 for the year 2016. The environmental regulator also allows them to be sold, traded, bartered, or otherwise transferred in a secondary market. Operating in this secondary market is the "Trade" part of cap and trade. People can buy and sell emissions allowances to match the emissions they are going to create during power generation. Emissions allowances are not "black marketed." They are "white marketed." With full encouragement and compliance by regulators and everyone else, these emissions allowances are actively traded on a legal and transparent market. Generators buy and sell the right to pollute, but EPA establishes the aggregate right to pollute according to the cap, which is the number of pieces of scrip that have been put into the market plus the number of endogenously generated and allowed offsets. Caps are generally set to decrease in magnitude over time so as to mitigate the pollutant being regulated. Such is the point of cap and trade regulation—to continuously reduce the cap and thereby continuously decrease the right to pollute to the environment. We will show some examples of CO_2 caps that were proposed a decade or so ago in the United States as examples.

7. As an emission allowance is "burned up" by a generator, there is a cost of doing so. An emission allowance is a perishable commodity just like a fuel, and when a plant owner "burns it up" in his own plant, he acquires a generation cost from so doing, precisely as derived thus far. It is just as if the generator had bought and burned an extra lump of coal or an extra hour of labor. He converts two perishable goods each having a nonzero economic value (coal cost plus an SO_2 emission allowance cost) into another perishable good with a nonzero value (electricity). Consumption of the former two perishable goods—fuel and allowances—imposes costs in precisely the same way as shown earlier in this section. Any emissions allowance cost that is "burned up" enters a plant's production cost and affects that plant's operation decisions as shown earlier in this section.

8. Traded emissions allowances are, via trading, "marked to market." Accounting rules, e.g., Financial Accounting Standards Board (FASB), require companies to value them that way, just as they would value a lump of coal. Under an allocated/assigned emissions allowance scheme, the government mails out emissions allowances that have positive market value, and revenues or receipts must be booked at the level of that positive market value. It is just as if someone gave them silver bullion; it is revenue. They then burn those allowances, and offsetting emissions costs occur and are imposed on generator operating cost. Both the receipt of allowance transaction and the burning of allowance transaction, one a revenue and the other a cost, are valued on a mark-to-market basis. Under an allocation scheme, there is revenue and

there is offsetting cost. Under an auction scheme in which the generator has to buy each and every emissions allowance from a central repository, there would be no revenue but only a cost. That cost would have to be marked to market as it entered his generation cost. The different accounting treatments are a profound profitability difference. In the former case, the regulator gives you scrip for free to meet most of your obligation under the regulation. In the latter case, you have to buy every piece of scrip to meet your obligation under the regulation. Big difference, which IEMM as designed in this CDR, must capture.

9. Nobody tells anybody what to do. There is no centralization or "command and control" aspect to cap and trade environmental regulation. Generators act unilaterally based on allowance prices and their desired and required generation patterns. The price of the emissions allowance is the market signal that affects behavior. It is an intrinsically decentralized, market-based system. The market selects the worst polluter(s) and drives it from the market. The worst polluters are those who cannot afford to buy enough allowances to offset their pollution and still remain profitable. They become the high-cost providers by virtue of the high imposed cost of their pollution. Their environmentally burdened production cost drives them out of the money. Thus, they are individually selected (in a Darwinian sense) and are driven from the market altogether via retirement or are forced to retrofit their equipment and clean themselves up.

Under a cap and trade regulatory regime, the situation is markedly different from the independent situation in Figure 33. Under cap and trade, all plants in the United States or the IEMM jurisdiction being modeled have to vie for a fixed supply of NOx, SO2, mercury, and perhaps CO2 emissions allowances. Figure 34 illustrates by representing the same two interconnected electric power regions at the top as were in Figure 33.[8] Each of those regions contains all the fundamental dimensions of the power business: fuel substitution; indigenous generation including investment, operation, and retirement; mark-to-market electricity competition at the local busbar; load at the local busbar including net energy for load and local ancillary service load; inbound transmission; and outbound transmission. The regions at the top are power markets, which before cap and trade regulation were completely localized save for whatever transmission capacity might exist between and among them. However, with the imposition of cap and trade, they will, as a result, have been interconnected coast to coast, border to border, sea to shining sea, by emissions allowance markets. Law has mandated that the correct number of emission allowances must be held in advance by every power generator in every region as a precondition for him or her generating. By design, the requisite allowances are issued and can be purchased across many regions, not just within a single region. Voila! We now see the value of the plant model with the multiple inputs! IEMM can operate no other way.

This structure, as we shall see, is not peculiar to cap and trade regulation. It is going to be easily generalized to consider all types of environmental regulation. The way one might implement a straight pollutant tax is obvious. It is easiest to articulate and understand in the context of cap and trade regulation. The reason is that the signals that enter each generation node along the pollutant path can be outright bans, forced retrofits, etc., as well as simple pollutant pricing. (If one were to

[8] Obviously, the real world and IEMM have many more than two regions, but two regions illustrates. The generalization to more regions is straightforward.

build a constrained linear programming model, the same set of signals would enter the generation nodes.) Readers need to bear with the discussion and internalize it. IEMM regional models must have this structure so that IEMM can accurately represent the pollutant regulation in force, if any. The word “if any” means that many jurisdictions, particularly in developing countries, simply do not have any environmental regulation. Those are easy to represent with this structure—just impose a zero price on all pollutants. There is more to come regarding this point.

Figure 34: All Plants Vie for a Fixed Supply of All Emissions Allowances

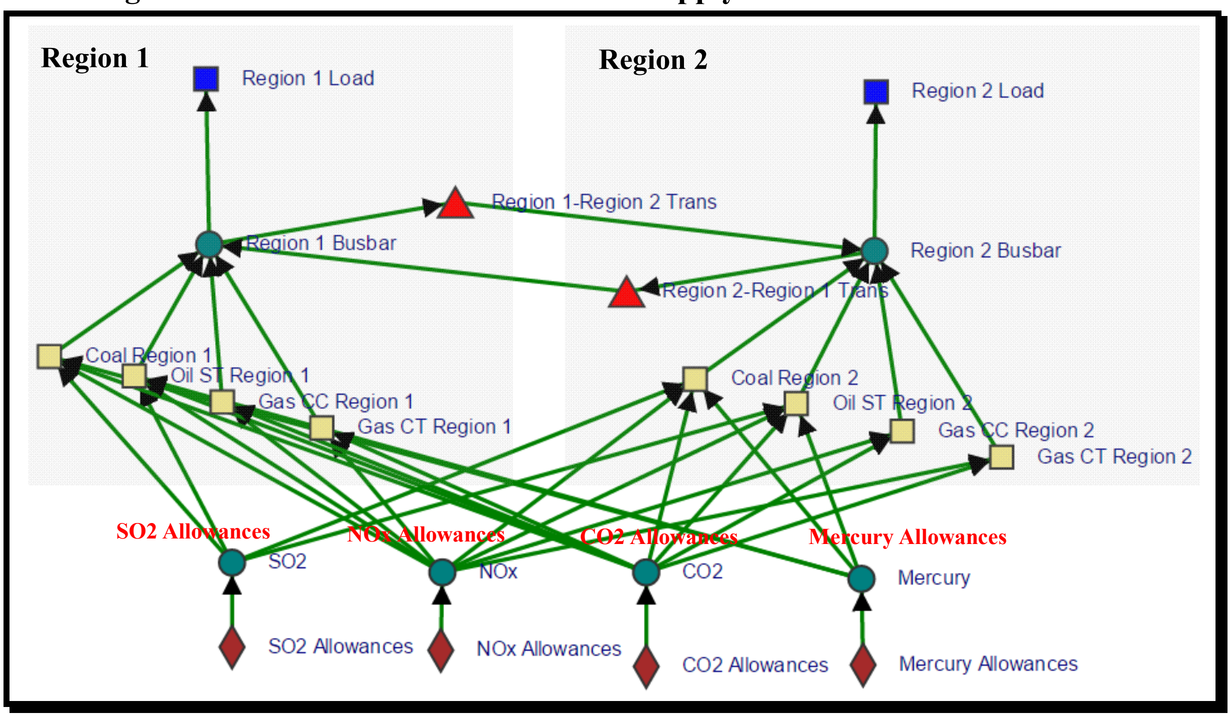
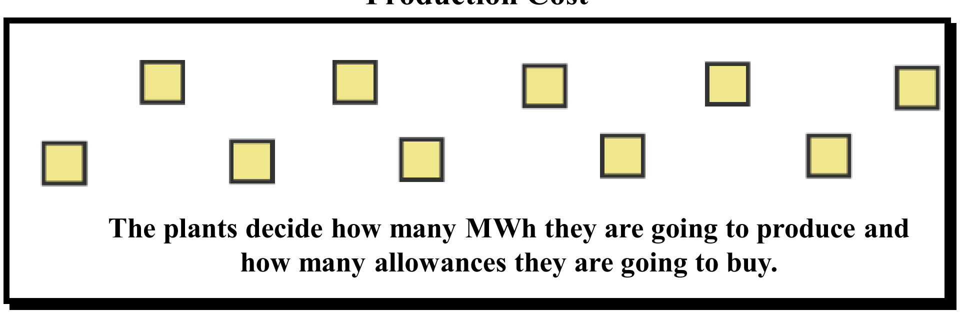


Figure 32 illustrates how the situation regarding emissions allowances appears at each individual plant in the power system under consideration. In return for the right to produce each MWh of output, each plant owner must own and surrender a slate of emissions allowances that offsets the emissions he is going to produce as a result of such generation. He must purchase those emissions allowances from the market or otherwise obtain them as a precondition of generation. He could obtain them via an allocation or assignment scheme, but he must, through those mechanisms or outright trading, obtain the requisite emissions allowances. The specific quantities of emissions allowances he must hold are dictated by the specific engineering-chemistry-physics properties of his plant given its technology and fuel configuration, which we have discussed at length previously. Higher emitters per MWh must hold or purchase more allowances than lower emitters per MWh to be able to produce each MWh. The idea behind cap and trade regulation is obvious—it forces higher emitters to buy more allowances than lower emitters and thereby forces higher emitters to incur more variable costs than lower emitters. It is designed to render higher emitters costlier to operate and thus push them closer and closer to the margin and out of the supply stack so that their pollution no longer is emitted. It is designed to force them to retrofit if they want to stay within the margin. The President himself has spoken about this in the context of the coal fleet.

Cap and trade emissions regulation is Darwinian in concept, driving out the highest emitters in preference to the lowest emitters. That is the whole point of cap and trade regulation. IEMM must implement the plant model in Figure 32 for each and every plant in each and every region in the regional IEMM model, and thereby it will be able to compute exactly how many emissions allowances are required per MWh output for each plant in each region.[9] Using the plant model in Figure 32, one can, for each assumed level of plant operation for each plant, calculate the emissions allowances consumed by each plant throughout the entire electric system. After that, one can accumulate all the allowances consumed by every plant and calculate how many emissions allowances would be required to sustain a particular generation level from every plant in the system.

To see how the simple plant model in Figure 32 is used in the context of an entire fleet of generators and to understand its economic context, suppose we specified an emission allowance price p_S for SO2 to the North American electric generation fleet. That SO2 price would be delivered to all the plant nodes throughout North America. Those plant nodes would incorporate that price together within their own individual emissions rates to calculate the emissions cost they would have to bear at that specified emission price p_S for SO2.Calculating their individual emissions costs is precisely what the calculations associated with Figure 32 instruct us to do, and the calculation has been summarized in detail previously in this section. As shown in Figure 35, every plant would take the price p_S for SO2, calculate its individual variable operating cost, fully burdened by the cost of the SO2 allowance, and would operate if it were profitable to do so under the variable cost thus calculated and not operate if it were not profitable. De facto, it would maximize profits in the face of its economic costs plus its emissions costs. In a truly economic sense, it would operate if it were profitable and would not operate if it were not profitable. (Internalizing emissions costs lies at the heart of the notion of internalizing otherwise external environmental costs into plant operation decisions.) Internalization would occur given a price p_S for SO2 and given the plant's own unique requirements to buy SO2 allowances, which adds to its production cost an amount dictated by its own SO2 emission characteristics. The procedure is as simple as what we saw previously—creating an environment-adjusted supply stack and using that environment-adjusted supply stack to operate plants.

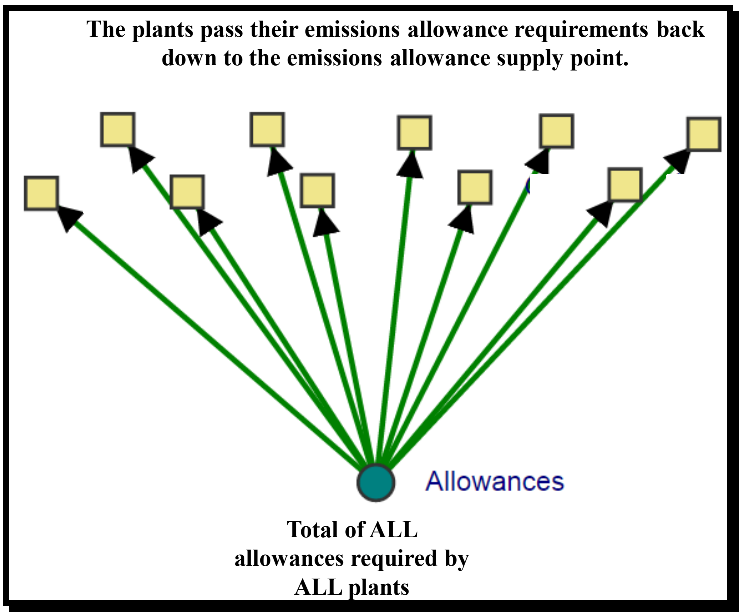
Figure 35: Every Plant Decides How Much to Operate at Its Individual, SO2-Modified Production Cost



[9] IEMM requires the requisite emissions coefficients for SO2, NOx, mercury, and CO2 for every plant in the United States. They are available in the public domain in EPA databases.

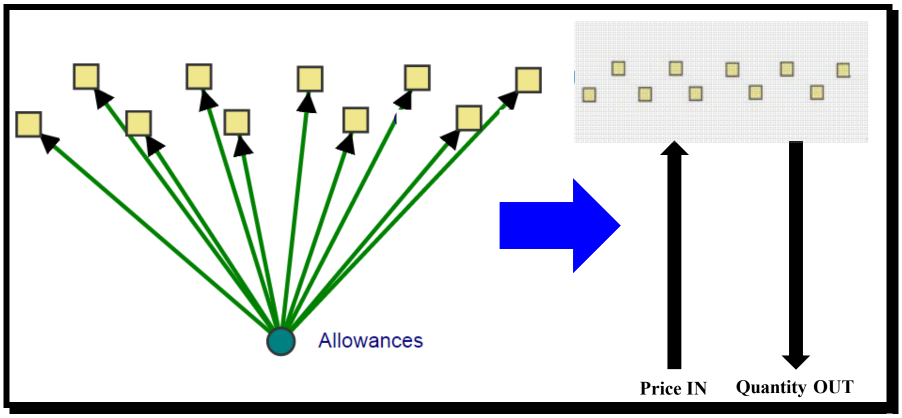
Every plant in Figure 35 would generate its individual level of MWh output and would thereby generate its individual level of SO2 emissions allowances required. As this occurs, IEMM then extends a link back from each plant in Figure 35 to an aggregation point as shown in Figure 36 to sum up the individual SO2 emissions allowances for every individual plant so as to compute a grand total summed across all the plants. This aggregate amount is what any cap in place will limit by law. The network structure in Figure 36 is no more than delivering the levels of plant emissions back along the links from the individual generators to a common point of aggregation where they can be added up.

Figure 36: IEMM Sums All the Emissions Allowances Required by All the Plants Given Their Operations



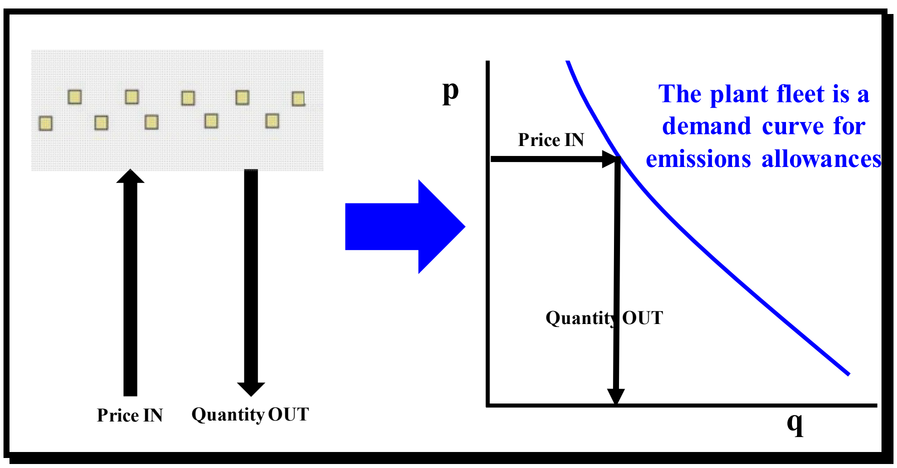
Notice what we have done from an economic perspective. We have delivered an emissions price ps for SO2 into the full fleet of plants in Figure 36, we have made adjustments in those plants' operating costs to account for emissions volumes and prices, we have operated them in the way that would be indicated by their environmentally adjusted production costs, we have thereby inferred how many MWh they will individually produce, and we have thereby inferred how many pounds of SO2 they will produce and, therefore, need to offset by buying emissions allowances. That tells us for each plant and thereafter by summation in aggregate how many emissions allowances we will need by summing the pounds of SO2 over all plants in the market. To summarize, we have put a price ps for SO2 into the plant mix, and we have taken out the quantity q of SO2 emissions entitlements that will be consumed by that plant mix. The fleet of plants is a "price IN-quantity OUT" model for SO2 emissions allowances as emphasized in Figure 37. One puts the emission allowance price into the model, and out comes the quantity of emissions allowances demanded by the mix of generators in the market. It is the individual plant model in Figure 32 that allows this. That is why it is such a linchpin of IEMM.

Figure 37: The Aggregate of the Plants Is a "Price IN-Quantity OUT" Model



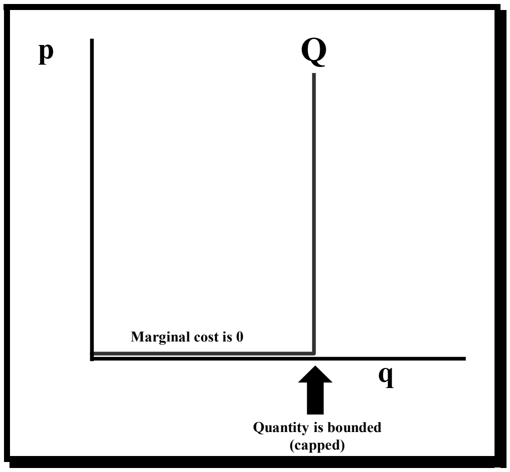
From an economic perspective, what does a "price IN-quantity OUT" model for SO2 emissions allowances as represented by the plant mix really mean? Simply, it is a demand curve for SO2 emissions allowances. You input a price of SO2 emissions allowances, and out comes a corresponding quantity of SO2 emissions allowances. You put in a higher price of SO2 emissions allowances, and out comes a lower quantity of SO2 emissions allowances. Voila! The plant capacity addition, operation, and retirement model is, in reality, a demand curve for emissions allowances in the classic economic sense. Figure 38 emphasizes.

Figure 38: The Plant Model Is a Demand Curve for Emissions Allowances



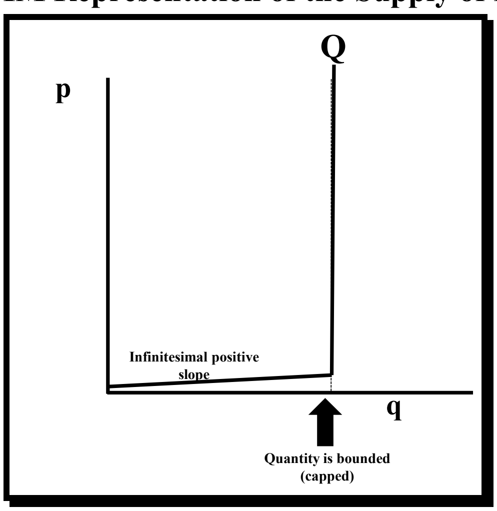
Voila, by this logic, IEMM will contain a robust aggregate demand curve for emissions allowances, and it is an explicit function of the price of those emissions allowances. And what do economic models look for when they have a demand curve? They look for a supply curve to lay on top of it (or an exogenously given price)! And how can IEMM think about the supply of emissions allowances? We think about the supply very simply as shown in Figure 39. There is a maximum quantity Q of emissions allowances, a cap, that is issued into circulation by the EPA or similar regulatory body either by allocation/assignment or by auction. The marginal cost in an economic sense of those emissions allowances is zero. (The cost of the printing of the emissions allowance scrip and placing it into circulation is negligibly small, too small to be considered as an important part of cost.) Thus, the regulatory mechanism de facto places a supply curve for emissions allowances into the market that is horizontal at marginal cost 0 until the prescribed cap amount Q is reached. At that point, the marginal cost of the next emissions allowance becomes infinitely high; there is no "next" emissions allowance available once the cap is hit. No price is too high; Q is a hard cap. You cannot produce one additional emissions allowance above and beyond the quantity Q at any cost; it is illegal under the controlling law and regulation. (This is not unlike the situation we shall see later where the price of power does not necessarily have to be equal to the cost of the last or marginal plant.)

Figure 39: The Supply of Emissions Allowances



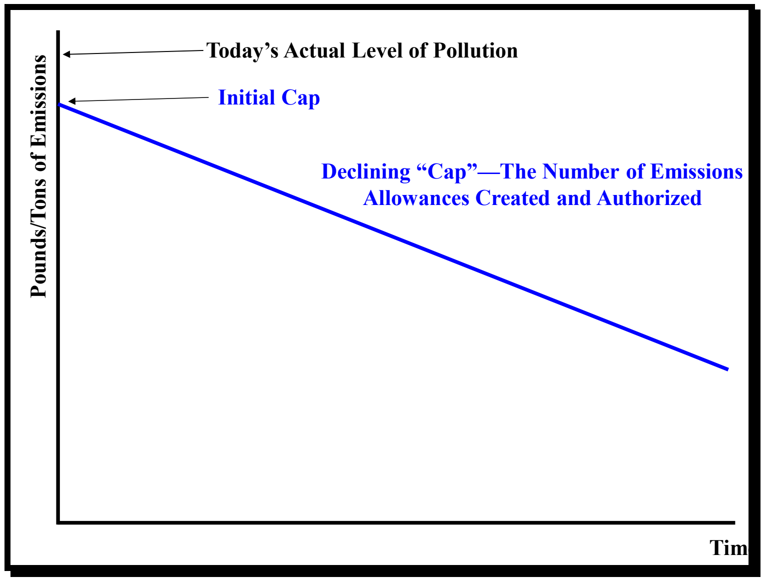
IEMM does not have to use strictly horizontal lines in such supply curves. All otherwise horizontal lines can have strictly positive slopes. Thus within IEMM, we represent the actual cap in Figure 39 using an almost horizontal segment as shown in Figure 40. This requirement is no particular problem. We can use a slope of 10^{-8} and be so close to 0 that nothing material is affected. (We can do this simply so that we can use the quadratic production function model to emulate the emissions allowance supply curve.)

Figure 40: The IEMM Representation of the Supply of Emissions Allowances



A simple supply activity in IEMM sets up precisely the emissions allowance supply curve in Figure 40 for each of the local or continental pollutants considered. Figure 40 is a perfectly well-designed supply curve in the economic sense. The IEMM curve allows the maximum quantity, the cap Q, to vary over time and, therefore, to represent downward sloping or more stringent emissions allowances through time.

Figure 41: The "Cap"—The Number of Emissions Allowances Created and Authorized

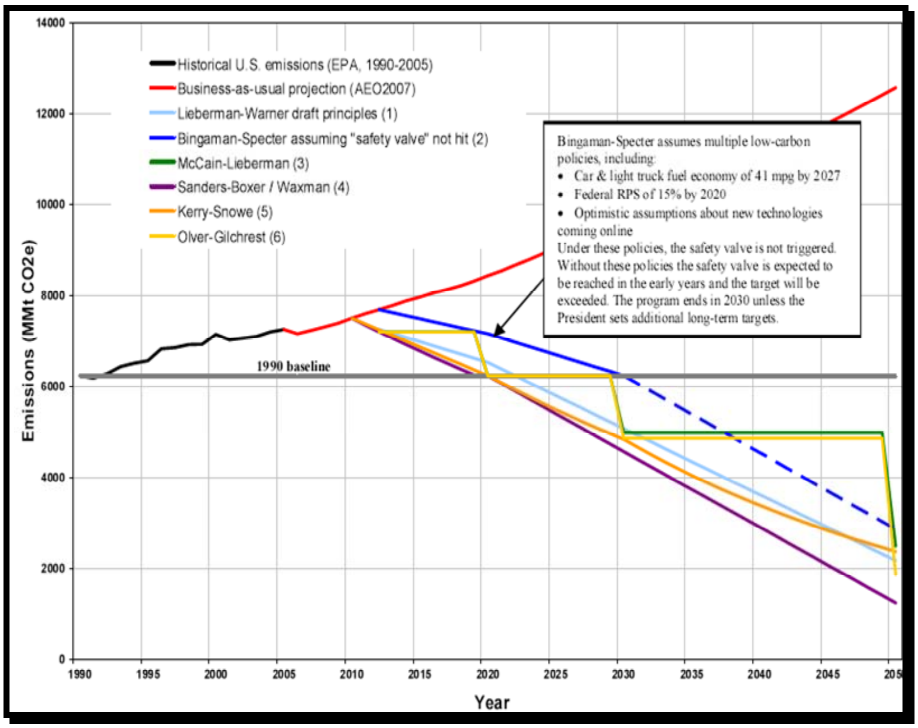


The way IEMM would use Figure 40 is to posit that the cap Q is a time-varying function Q(t) forward in time within IEMM, the cap representing how the regulators are anticipated to vary the cap. For example, we can posit Q(t) to represent say a 7.5-million-ton cap through 2017, a 4-million-ton cap from 2017-2023, and a 2.75-million-ton cap after that. Figure 41 illustrates. We can represent one of the linearly varying CO2 caps by simply setting Q to decline linearly from a present high level to a lower level. Doing so will create a supply curve for emissions allowances in each forward year that represents the assumed CO2 regulation and the consequent supply of emissions in that year. (We discuss banking of environmental allowances and temporal tradeoffs later in this CDR.) In Figure 41, the cap, which represents the number of emission allowance "scrip" certificates that are printed and placed into circulation by the regulatory body, in practice typically declines over time. The idea of the declining cap is to "ease" the offending pollutant out of the air and water gradually and measurably over time and to have a decentralized market pick the highest polluting, least economic plant(s) to move to the margin and exit the generation mix altogether or to retrofit to curb its polluting.[10] Furthermore, the initial cap is typically below current aggregate production, and it decreases from there. A declining cap ensures that any price of emissions allowances that comes out of a cap and trade system will begin at a positive level and will continue at a positive level. The emission externality will always be priced and thereby internalized via cap and trade into the economic system. The entire point of a cap is to positively price the externality and have that positive price move plants into and out of the system.

[10] We will see shortly how to override the assumption that the market sets the pollutant price in favor of a centralized command and control regulatory authority.

The Pew Research Center website a few years ago enumerated many or all of the CO2 capping activities proposed at that time in the United States related to greenhouse gasses. (There have been others since then. These are merely representative.) It was clear at the time that carbon dioxide regulation portended to occur in some form of cap and trade scheme with or without a “safety valve” maximum carbon dioxide price. The various caps that were proposed are summarized in Figure 42. It is not important what the actual caps are or were; it is important to recognize that they are in fact caps and that there is strong possibility that CO2 regulation will manifest in the form of cap and trade regulation. As we shall see, pollutant taxation and completely uncontrolled scenarios are trivial special cases of this rather sophisticated cap and trade logic.

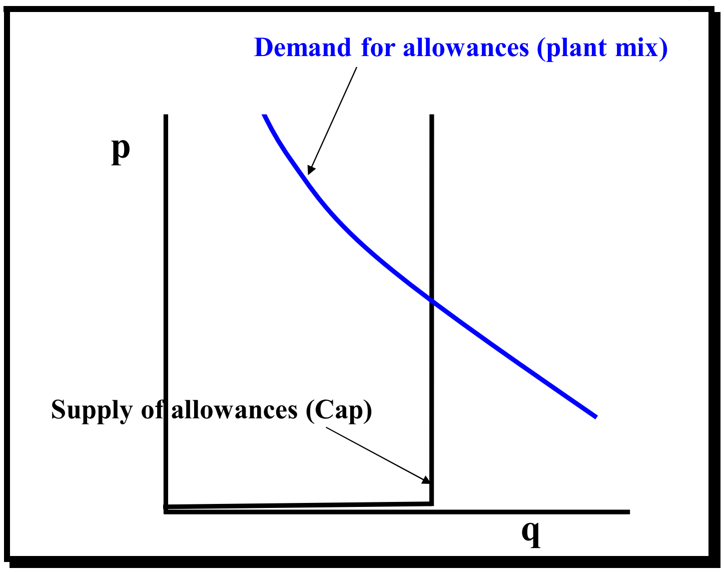
Figure 42: Various Levels of Proposed CO2 Caps (Pew Research Website)



With this market-determined pollutant price method, the economic loop is closed—there is both a supply curve and a demand curve. We have economic context for the model of emission allowance capping and trading. The proper economic context is that in Figure 43. The pollutant regulation is represented in IEMM by a capped supply curve for emissions allowances. The entire plant fleet model is in an important sense the demand curve, telling us how many of those emissions allowances will be consumed as a function of their price; this is simple, straightforward, and potentially applicable to a lot of countries and regions. California is already doing this for CO2. IEMM needs this capability in its quiver, and this design delivers that.

The accumulation of emissions allowances across all the energy generators is represented in Figure 44. Notice at the top left, all the emissions allowances from every plant are accumulated into a single circle or market hub for emissions allowances. The model sums all the emissions allowance requirements as plants make their corresponding investment, operation, and retirement decisions in the face of emissions costs as well as economic, REC, and all other costs. As shown in the structure at the left, the sum of all the plants creates a very large demand equation for each emission allowance. Once the emissions allowances are accumulated in the circle at the left, we note that government policy regarding those emissions allowances prescribes an overall emissions allowance supply function, precisely as described previously. It is the intersection of the prescribed emissions allowance supply function and the demand function that sets the price of the emissions allowances. In our CDR, that price is an endogenous function of every plant in the system today and every plant that could possibly enter or exit the system in the future. Emissions allowance prices are fully endogenous and fully a function of fuel prices; investment, operation, and retirement of generators; net energy for load; ancillary service; mark-to-market power-on-power competition at the busbar; inbound transmission; and outbound transmission. It is supply-demand balancing in the emissions allowance market that sets the price of the emissions allowance. One cannot analyze emissions allowances on an isolated scenario basis; they must be endogenous to the system.

Figure 43: The Economic Context of the Cap and Trade Regulation Model



The cap and trade system in the United States, California, and Europe melds generation, transmission, and load considerations in the energy regions at the top of Figure 45 with supply-demand logic for emissions entitlements at the bottom, thereby forming fully integrated energy-emissions prices not only in each individual, local energy market but also fully coupled across all regional energy markets of the United States throughout the country or throughout any other venue around the world. Using the emissions allowance aggregation and pricing logic at the bottom and sending calculated allowance prices up to each and every individual generator, cap and trade (and the IEMM model) sends precisely the correct signals from the allowance market to each generator. The system in Figure 45 comes to a fully interdependent and integrated equilibrium that spans emissions allowances and electric energy throughout the applicable region simultaneously and completely. The price of each is intrinsically related to the price of the other—neither is independent of the other. Power prices depend on emissions prices, and emissions prices depend on power prices.

Figure 44: Accumulate All Emission Allowances Cap and Trade Domain

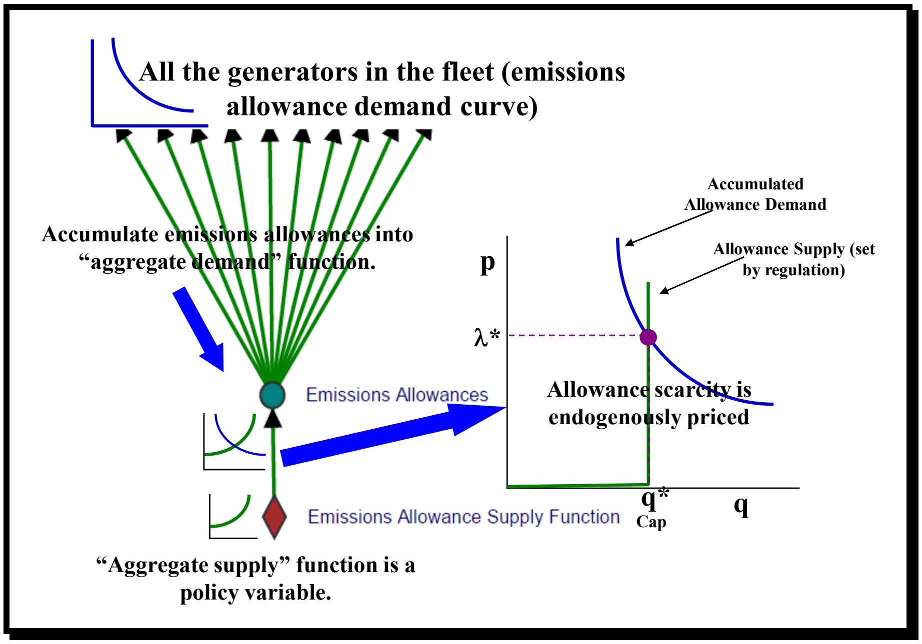
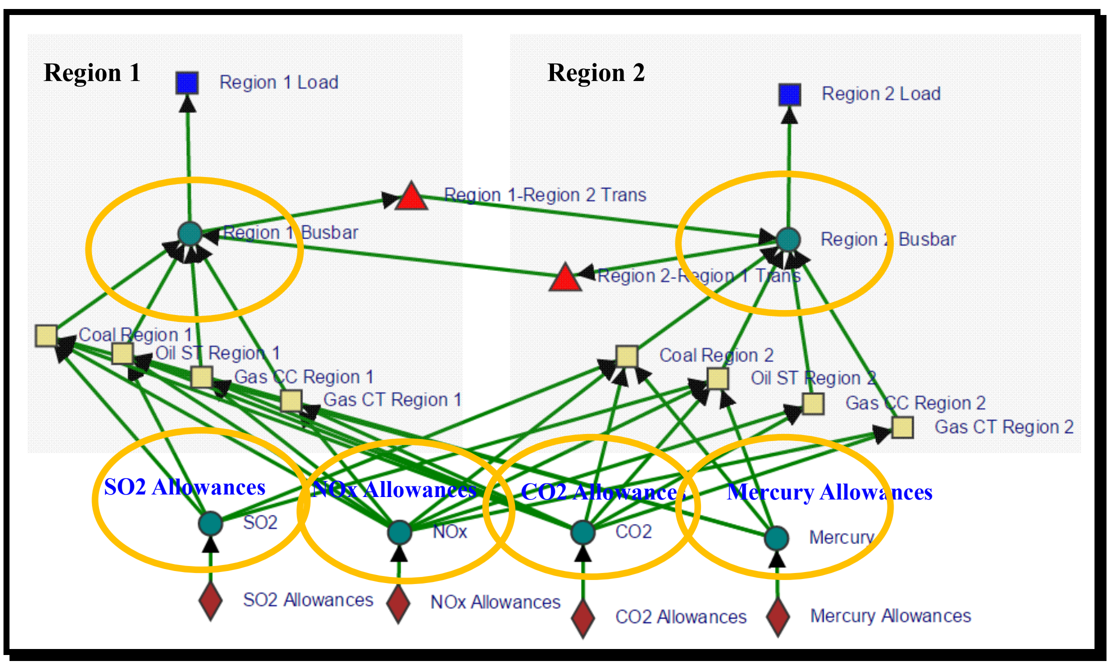


Figure 45: Power Prices Depend on Emissions Prices AND Emissions Prices Depend on Power Prices



It is important given the diversity of regions and plants that the underlying power model be granular and detailed. The electric generation portion of the IEMM model as represented in Figure 45 usually needs to be highly granular and, therefore, to allow the preceding considerations to be represented. In fact, a previous economic model we have built using this method has been highly granular, representing every one of the generation stations in North America either individually or as part of regional aggregates. It spanned the continent. It was comprised of some 50 million equations and unknowns, a monstrous and colossally detailed model.

It is now time to move beyond cap and trade and to think about what this structure means for environmental control methods other than cap and trade, pollutant taxes, or no controls at all. The answer is simple. If a particular plant configuration is eliminated by regulation (as can easily occur with ESPS or NSPS or CPP), we just put an infinitely high factor price on that link for that factor, and that will eliminate that plant from the mix. There is no power price at which it can be profitable, run, or survive if that happens, and it will be driven from the mix immediately. That is the right way to implement ESPS for example. The factor price is truly a "signal" to the generating node. If, in the real world, that generation node (e.g., an unretrofit pulverized coal plant) were not allowed to operate because of an emission regulation, how might we model that? We would simply put an infinitely high price on a factor to that plant, and it would cease to operate. As economists, we realize that putting an infinitely high price on some activity is the same thing as imposing a zero volume on that activity. By understanding that a ban or constraint is equivalent to an infinite price, we can use this construct in Figure 45 for the complete and total range of environmental regulations or constraints. There is no permutation of environmental regulation that we know of that cannot be represented using the construct in this CDR as represented in Figure 45. (Perhaps someone could posit an outlier, but we have not encountered one yet.) Think about how powerful that is.

The IEMM model needs to be a nodal, price-quantity-capacity addition model of each region with full and complete factor inputs. IEMM needs to include a highly detailed, regionally disaggregated network representation of electric generation assets (not to mention transmission corridors and demand patterns articulated later in this CDR.) The IEMM model design is based on economic principles that govern agents in commodity markets.[11]

2.5.2 Pollutant Taxes (As Contrasted with Cap and Trade)

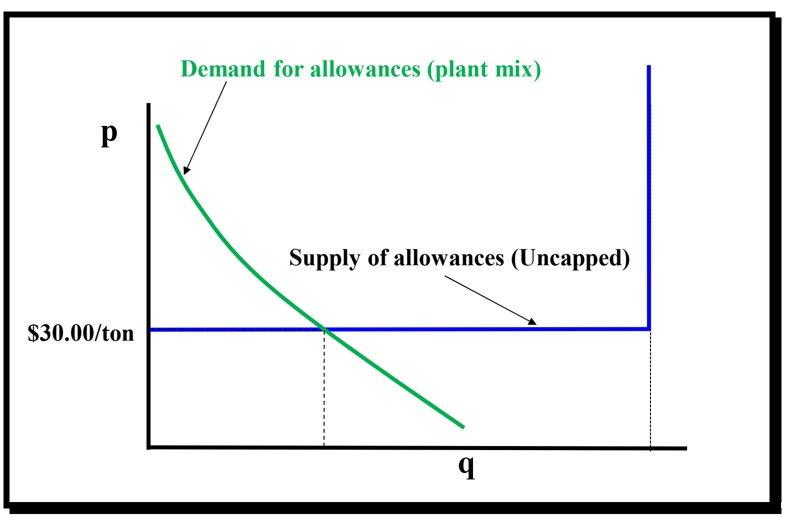
Why did we go to all this trouble to articulate pollutant cap and trade? After all, isn't cap and trade "on the outs" with some jurisdictions and some courts? That may be true in certain venues in the United States, but it is definitely the chosen technique in other jurisdictions and courts that will be the subject of IEMM modeling. It is far from dead. It exists in Europe and California, and it could well emerge elsewhere. The reason we have spent the time and detail developing it is that it is the hardest technique to implement from a modeling perspective, and most of the other techniques of regulation can be achieved as special cases of it. That is so, so important. If we get cap and trade right, we can specialize to zero controls, pollutant taxes, and a plethora of other regulations IEMM might want to implement straightforwardly. As an example, government imposed pollutant taxes are trivial to implement using the cap and trade representation. They are a trivial special case in which the supply of emissions allowances is infinite at an exogenously specified price, which is

[11] See Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green (Harvard Economics Professors), Microeconomic Theory, Oxford University Press, 1995, 981 pages or Hal R. Varian (University of Michigan economics professor), Microeconomic Analysis, Third Edition, 1992, W. W. Norton and Company, New York for a careful and comprehensive articulation of producer and consumer behavior and market equilibrium.

the assessed pollutant tax. Under a pollution tax regime, regulators choose to posit and mandate a tax rate on particular emissions, perhaps broadly or perhaps regionally, thereby attempting to penalize those emissions by internalizing the tax into the price passed on to consumers and plant operators. However, and this is crucially important, taxes do not by construction cap or limit emissions to any particular level. They cannot. There is no preordained relationship between a tax and the magnitude of pollutant reduction. The exact impact of a tax needs to be figured out by the IEMM model in each jurisdiction EIA might implement. The impact of a tax is an extremely important distinction—with a pollutant tax, there is no realistic attempt and no legal or regulatory infrastructure to achieve a particular cap. A tax is a tax, and a cap is a cap. Ne'er do the twain meet. And yet IEMM under this design will correctly model either. Importantly, it allows EIA to flip between the two and clearly understand both types of regulation and how the generation system, pollution, renewables, price, etc. could be expected to respond. Such adaptability is invaluable when IEMM moves around to the world to represent all the disparate jurisdictions around the world.

If we refer to the diagram in Figure 46, consider what happens when we set the height of the supply curve to be the desired CO2 tax, e.g., $30/ton. That means as long as the demand curve cuts through to the left of the vertical point, the price of CO2 will be $30/ton. All we have to do to implement a tax of this type in IEMM is to expand the magnitude of the cap to infinity, i.e., to move the cap, the vertical part of the supply curve, to the right all the way out to infinity, well past any reasonable range of operation of the model. At the same time, we must set the height of the horizontal part of the supply curve to be constant at the desired CO2 tax, e.g., $30/ton. Creating an unrealistically large upper bound ensures that there will never be enough CO2 produced to hit that cap, and, therefore, the cap will be inactive from the standpoint of the model. However, the CO2 will always be priced at $30/ton regardless of the level of CO2 output. As that occurs, Voila! We have a CO2 tax at $30/ton. And we will have the full aggregate CO2 accounting. The figure is in fact how IEMM implements a tax regime—just lift the cap to infinity and lift the height of the CO2 supply curve to the level of the desired tax.

Figure 46: Uncapping at a Given Pollutant Tax Rate Imposes That Pollutant Tax Rate on the Model

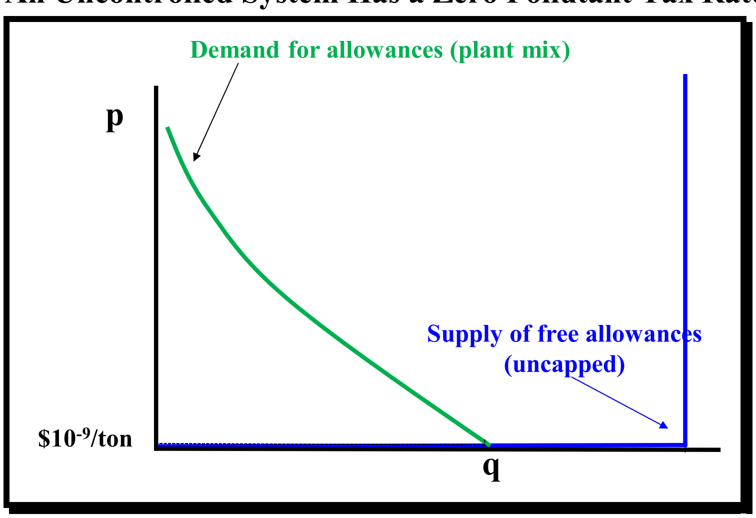


The magnitude of the tax (the height of the blue allowance supply curve) must be allowed to vary over time. It can firm (growing from \$30 in this illustration) or erode (declining from \$30) in various scenarios IEMM will want to run. The beauty of this method is that IEMM still does all the individual plant accounting, and all the multiplant accumulation of allowances articulated previously in this section. The model will easily report as a function of the tax level exactly how much the entire power system will pollute. This monitoring and measuring aspect of the proposed structure is crucially important for IEMM. We see how easy a straight pollutant tax is to implement with this IEMM design.

2.5.3 Completely Uncontrolled Emissions

Now that we have cap and trade and pollutant taxes straight, what is the nature of a totally uncontrolled system (so common across the world)? With the IEMM structure articulated here, the answer becomes trivial. It is merely the situation in Figure 47. IEMM just enters zero or infinitesimally low pollutant price with an effectively infinite supply of allowances, meaning exactly what economic theory tells us that the externality is not priced and is, therefore, a true externality in the classic sense. And the power system will run without any control whatsoever of the given pollutant because there is no cost or price. That is going to be an important case or set of cases to be run for power systems outside the industrialized world, power systems that have never or have been very loath to control pollutant output, to substitute cleaner fuels, or to install retrofit cleanup equipment on power plants, particularly coal or fuel oil plants. This uncontrolled situation too, because of its inherent structure of accumulating the emissions from each and every plant in the system and properly accumulating them into an aggregate, gives IEMM full knowledge of how much the system will pollute if left uncapped and uncontrolled. How much the system will pollute is not only an important policy consideration, but it is also an important baseline against which to measure other actions the power system might take and how much environmental cleanup that would engender even if the environment were technically unregulated and uncontrolled. A large part of the world power system remains unregulated and uncontrolled, yet prospects for environmental regulation grow every day throughout every venue of the world.

Figure 47: An Uncontrolled System Has a Zero Pollutant Tax Rate and No Cap



The beauty of the proposed approach for IEMM is that it does full and complete accounting of individual plant and total system emissions even if it prices them at zero. It allows IEMM to quantify the magnitude of emissions automatically under every model scenario it runs for places like China or the Middle East or Indochina or South America. Quantifying emissions will give a realistic understanding of developing and developed countries that elect not to control pollution. IEMM can do all the accounting, and IEMM can consider what would happen should they elect to control. That is going to be invaluable to EIA.

2.5.4 Command and Control Regulation (e.g., BACT, Point Source, CPP)

Command and control regulation affects the ability to build a plant in the first place or to operate a plant at all because of some particular pollutant. It is a generalization of the three types of environmental regulation presented previously, delivering a different signal to the power plants other than a simple price. De facto, the plant "capacity" is not available for a certain type of service under CPP because that plant emits a certain pollutant at a rate that is deemed by regulators to be unacceptable and unallowable. IEMM will represent that type of command and control regulation via the "capacity" portion of the model, which will be articulated in detail in the next section. To presage, it will deny use of capacity to certain uses if they pollute too much under the CPP regulation. (IEMM would put an infinite price and that capacity link, and it couldn't be used for that regulation-busting activity.) For now, it is important to conceive that there will be another link incoming to the generation node in Figure 32, and that link will be a "capacity" or "enablement of capacity" link. For a generation node to run in a certain configuration, it must have "capacity." If that capacity is used for some other mode of operation or denied because of regulation, that generation node cannot and will not run and the pollutants it would otherwise emit will not be emitted. In the case of an environmental regulation that does not allow a certain plant or a certain type of capacity to run, that capacity will simply be unavailable and shut off by the emission or the capacity link having its price set to infinity. That is a very clever and efficacious operational way to ensure that command and control regulations that disable or change capacity can be represented. There are several command and control regulations we will summarize here and will explore in some detail later in this CDR. Command and control regulation has burst into the United States system, and IEMM has to represent it. It has not burst into other systems, which continue to prefer decentralized solutions such as cap and trade or pollutant taxes. IEMM needs to represent both.

As a vox populi example, consider NSPS or new source performance standard regulation. NSPS means that a certain type of entering capacity with a certain set of input-output coefficients is simply not available for construction or operation. It is banned by regulators because it produces too much CO2 per MWh of generation in the judgment of regulators and/or the law. We represent this in the capacity addition section, ruling out capacity additions of certain types by assuming an infinite capital cost of entry for those types. NSPS might take the form for example that a new plant cannot legally emit more than 1,000 pounds of CO2 per MWh generated, thereby precluding coal without sequestration by fiat. The capacity addition section of IEMM would not allow any plant that violated that standard to be built or operated if built. Quite simple and straightforward to implement using this structure.

ESPS or "existing source performance standard" means that a certain type of pre-existing, embedded, already operational operating capacity with a certain set of input-output coefficients is forced to derate its capacity to 0 because of the regulation. Under this type of regulation, the regulators basically "kill" any plant that violates the standard, forcing it to retire or be decommissioned according to some period of time. Retirement opens up the rest of the system to endogenous capacity addition, which must occur using alternative capital stocks that meet the NSPS and the ESPS requirement. The ESPS type of regulation is easy to represent in the IEMM design by disabling generation capacity above the standard according to a schedule. The structure easily, for example, disables unretrofit generation. Such disablement will force the model to add only retrofit generation or to add only retrofits. The generation nodes in Figure 32 are readily amenable to that style of regulation.

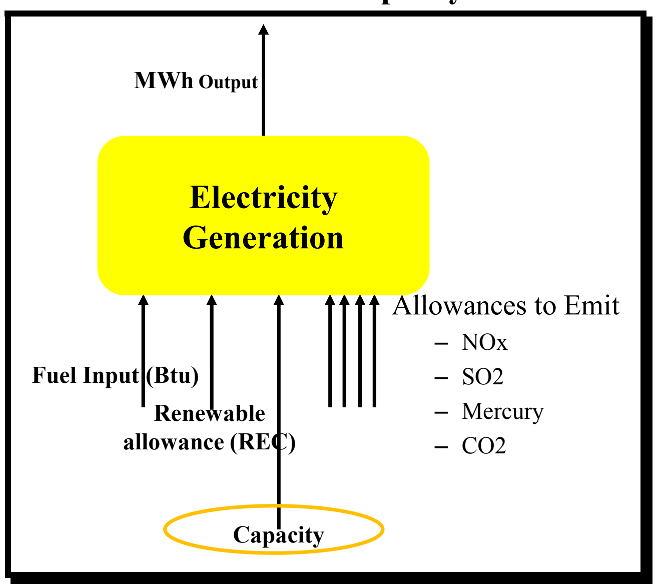
Other forms of command and control regulation have been "best available control technology." Best available control technology is akin to remodeling one's house and being allowed to receive permits only if you install on a retrofit basis certain HVAC technology that the regulators have deemed appropriate or allowable. You have to "bring your house up to code" as a condition to maintain it. By analogy, when you maintain your power plant or engage some other activity, the permits to do so require you to install certain remediation equipment. IEMM represents this type of regulation by requiring certain types of capacity to be implemented at certain maintenance intervals or other time-schedules.

All of these forms of regulation will be dealt with in the capacity addition and retrofit addition portion of the discussion that is forthcoming so that we will see how IEMM can incorporate them. There is no other way this full range of potential regulation can be implemented as far as we know. To presage that discussion, we postulate that "capacity" is also a factor of production by a generator. To generate, a generator has to have "capacity" dedicated to such generation. That capacity must be available one for one, i.e., if you want 1 MWh of energy in every hour in a given year, you need 1 MW of capacity supplied to that node. If because of some regulation such as ESPS or NSPS, you cannot generate in a certain configuration, you merely turn off the capacity to that node for that type of operation by inserting an infinitely high price on that inbound capital link. That, in IEMM, will do it. (Operations research people generally do not default to this type of representation. They want to "constrain" something out by using a quantity relationship rather than "price" something out by using a high implicit cost. Economists want to set infinitely high costs or prices. The effect is the same, and yet the economic approach allows IEMM to do this with signals to plant nodes. Very insightful, and very efficient. Signals to plant nodes; that is the key.) You either price the capacity to that node at an almost infinitely high level, or you set the quantity of capacity as a factor to that node to zero if it violates the NSPS or ESPS constraint. Simple. The structure of the generation nodes is going to have the structure in Figure 48. The capacity discussion later in this paper will pick up on the crucial idea represented by Figure 48.

2.6 From Where Do REC Prices Come?

This section articulates how and why a Renewable Portfolio Standard or RPS model based on tradable RECs must be represented in IEMM. (A Renewables Portfolio Standard is a target level of renewables, usually expressed as a fraction of total MWh consumed, that will or should be met by qualifying renewables generators.) A goodly bit of our renewables/REC discussion will parallel the pollutant discussion. In particular, we begin with a sophisticated RPS/REC design centered on endogenously generated and fully tradable RECs that will meet a certain RPS. We then make various specialized assumptions to show how the entire structure can represent no RPS/no renewables, a "green tax" that can catalyze renewables, or various command and control mechanisms that can catalyze renewables or direct subsidies of renewables by taxpayers (or any party outside the ratepayer pool). All are needed by IEMM to get renewables right because there will be so many combinations and permutations across the various venues around the world in which IEMM will have to operate. We will talk about the temporality of renewables generation later and how IEMM will have to represent it. This whole area of renewables and renewables generation has become white hot around the world. China is being pressured to build wind and solar rather than coal. Germany, amazingly to many, has literally coated its residential roofs with solar photovoltaic. The Mojave Desert has a lot of central station solar projects and proposals that are out of the money but still being actively promoted. However, politicians and environmentalists are questioning entry there. There are a tremendous number of potential sites for renewables generation around the world. That is why IEMM needs to embed the broad-based renewables design in this CDR.

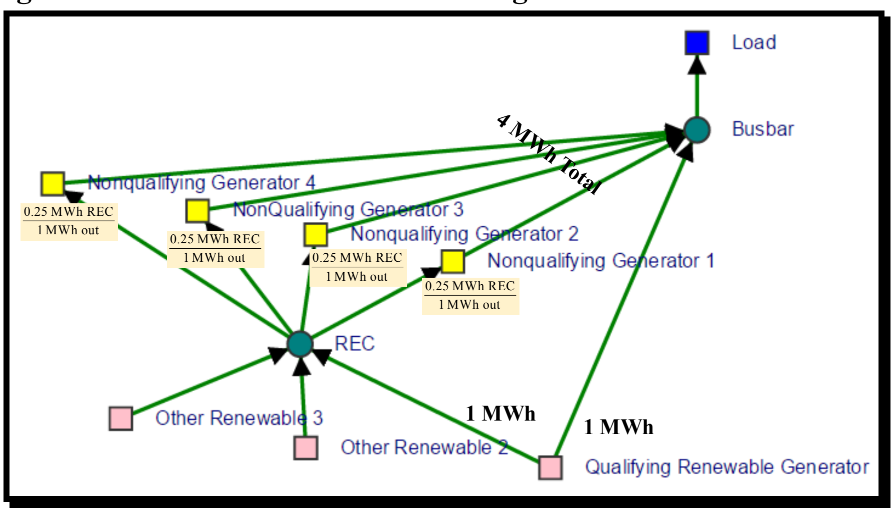
Figure 48: Each Generator Must Have Capacity as a Factor If It Is to Generate



Think about a mandated RPS supported by a traded REC system and how IEMM needs to be structured in order to represent it. The first thing that has to happen is that the regulatory authority has to specify what a "qualifying renewable" is and what a "qualifying renewable" is not. Generally, technologies such as wind or solar are deemed to be "qualifying renewables," but thermal generators and often hydroelectric are not. We have discussed that previously in this CDR.

Suppose we have a qualifying renewables technology that generates 1 MWh of qualifying renewable energy. Such a technology appears at the lower right of Figure 49. Notice that the qualifying renewable energy technology produces 1 MWh of electricity, and it simultaneously produces or "prints" 1 MWh of REC, i.e., it "prints" 1 MWh of REC scrip. That 1 MWh of REC scrip after that must be purchased in whole or in part by the fleet of thermal power plants in a ratio of ¼ to 1 in order for those thermal power plants to be able to operate. In the example in Figure 49, each thermal plant must surrender 1 MWh of REC scrip for every 4 MWh it generates. It must purchase those RECs from the open REC market, or it must otherwise acquire them (perhaps by building its own portfolio of renewable or green generation) or by putting solar collectors on the roofs of its thermal generators. A large number of renewables technologies may be putting RECs into the market, but the requirement that each non-qualifying, non-renewable generator must surrender 1 REC for every 4 MWh generated is what ensures that there will be 4 total MWh of thermal generated for every 1 MWh of renewable energy generated. This specific requirement would, therefore, meet a renewable portfolio standard of 20 percent. There would be 1 MWh of renewables for every 4 MWh of thermal.

Figure 49: A Model of REC Trading to Achieve 20 Percent RPS



It is the surrender ratio imposed on the non-qualifying thermal plants (the yellow plants in the diagram) that sets the 20 percent RPS. If that surrender ratio were to change, the renewable portfolio standard would adjust to a different number. If that surrender ratio were 3 to 1 rather than the 4 to 1 in the example, i.e., if every thermal generator had to surrender 1 MWh of REC scrip for every 3 MWh it generated, the RPS would thereby increase to 25 percent. IEMM will need this calculus in Figure 49, adjustable to any RPS if it is to represent renewables entry and cost correctly. Furthermore, as we will describe below, IEMM is going to have to represent the specific temporal generation pattern of the qualifying renewable accurately because the MWh that are generated and delivered to the market are not necessarily generated at time of base or time of peak. They enter the market at somewhat random but statistically predictable times. IEMM has to carefully quantify the generation duration curve for each renewables generator and insert the renewables MWh into the system at the correct time, correct in the sense that it must represent the true and correct timing that those renewables processes enter the market.

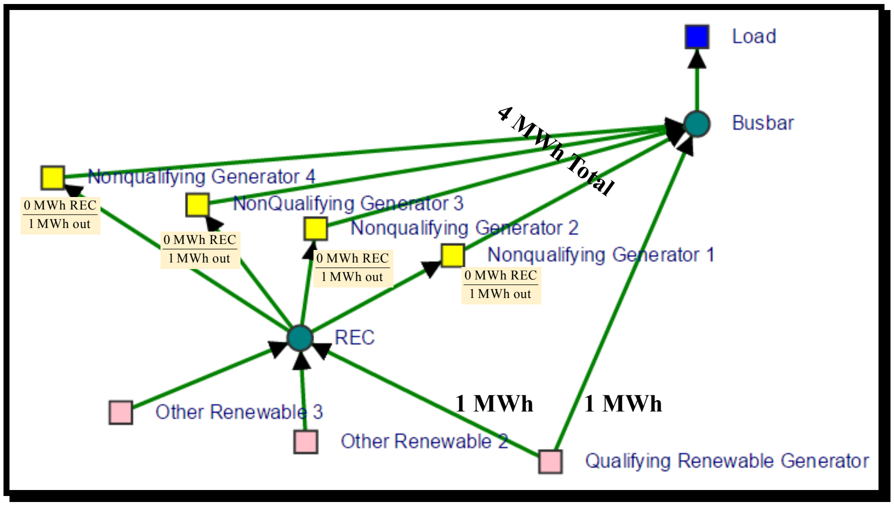
To reiterate, the network structure in Figure 49 requires that renewables must be built at a rate of 20 percent of the rate of thermal plants. If renewables were not built and operating at 20 percent of the total, there would not be sufficient generation from the nonqualifying generators at a ratio of 4 to 1 to meet the other 80 percent of the demand. That is, there would not be enough renewable plus thermal generation in place to meet the demand for power. As that occurs, the price of power will escalate to a very high level because there will not be sufficient thermal plus renewable generation to meet the load. The absence of renewables would create 4 for 1 an absence of nonqualifying generation, and that joint absence of nonqualifying generation together with the renewable shortfall would cause very, very much higher power prices. Sooner or later, the price of power would rise to the point at which a qualifying facility will profitably enter. Until that qualifying facility enters, a nonqualifying facility cannot enter because the system will not meet the specified RPS if that occurs. The only thing that can enter is a renewable. Renewable entry will bring the system to a situation in which aggregate supply meets aggregate demand, and 20 percent of the aggregate supply will come from renewables, even if there has to be substantial entry of renewables. Sufficient entry of renewables is precisely what an RPS means. A flat 20 percent of total MWh must come from qualifying renewables. Period. That is precisely what this structure ensures when placed into IEMM.

We can now see how the RPS fraction can be maintained through the emanation, trading, and consumption of offsetting RECs that are properly priced by the structure in Figure 49 and, therefore, install precisely the correct quantity of renewables to meet the specified RPS in the region. This trading mechanism is straightforward to build into the IEMM model with this proposed design (but assuredly impossible with global social welfare maximization). If IEMM did so, we would put the capital and operating cost of the renewable plant into the model, and the model would establish a price for RECs and a price for energy that would induce the renewable producer to enter and produce at maximum possible output. It would calculate the de facto cost of the subsidy or economic dislocation required to achieve the assumed RPS. Knowing the cost of a subsidy is going to be crucial to the functioning of IEMM, which otherwise would not site a sufficient magnitude of renewables, and no jurisdiction would meet any particular renewable energy fraction or RPS. Far from being free, renewables can be quite expensive. The construct put forth herein allows the increased cost of those renewables to find their way into electricity prices and ensures that the market embraces the full extent of the mandated RPS.

Of course, if RECs were unpriced (zero price), they might as well not be mandated at all. That is the situation without any renewables consideration, subsidy, or mandate. If RECs are priced at all, renewables will be entering the mix at some overall rate. That is the situation in which some regulator or government body has decided to subsidize or mandate the entry of renewables into the generation mix. That is modeled easily using the structure in Figure 49. If we wanted to represent no RPS or REC at all, we would simply postulate that non-qualifying generators needed 0 MWh of RECs for every MWh they generated. The situation would be that in Figure 50. No non-qualifying generator would have to surrender any RECs to allow any level of generation. In such a situation, there would be no market for the RECs and their price would fall to zero. Zero price RECs would be a completely unsubsidized, unmandated renewables situation, and IEMM would properly simulate that for various countries or regions in which there is no operational RPS or REC. That does not mean there would be no renewables; it would mean that there would be no subsidy or mandate for renewables. They could enter on their own if they were economic.

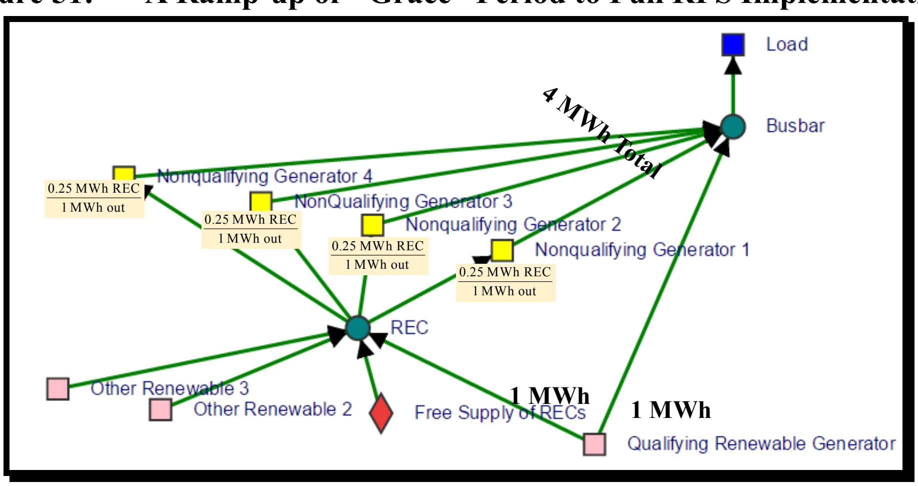
What if there were a deep subsidy of renewables, i.e., a deep subsidy of the pink dual output node at the lower right that represents qualifying renewables? That would be represented by putting in a subsidized capital cost into the pink qualifying renewables box at the lower right. That would drive down the price of the tradable REC and, therefore, the necessary incentive it would require to get additional renewables to enter the market.

Figure 50: Zero RPS—No Mandated or Subsidized Renewables Entry



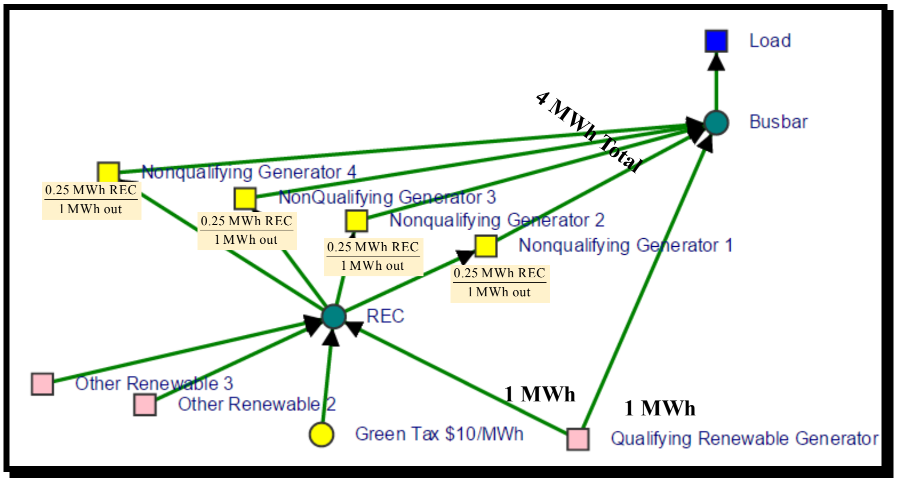
What if there were a "grace period" to get up to some RPS standard? One could simply put in a declining number of "free" RECs that would simulate the shortfall-during-ramp-up period. The situation in Figure 51 illustrates. In the figure, there is seen to be a (presumably declining) supply of "free" RECs introduced. These free RECs will reduce the number of qualifying renewables at the lower right that must enter if the electric market is to have sufficient aggregate capacity. If that schedule of free RECs declines over time, we will see the magnitude of renewables MWh enter the system be short of the RPS initially but ramp up to the RPS over time as the shortfall is erased. If we had an infinite or very large supply of free RECs in this construct that would be equivalent to the situation of no mandate or subsidy for renewables.

Figure 51: A Ramp-up or "Grace" Period to Full RPS Implementation



A final green scenario that people might consider is a "green tax" against non-qualifying generators. Rather than letting the price of RECs float as they have in these discussions thus far, suppose the government just wanted to impose a regulator-established green tax against non-qualifying generators. (Governments do this all the time as a way to raise revenue.) To simulate that, the structure becomes that in Figure 52. Notice that the "green tax" does not necessarily achieve the RPS, or it might, in fact, exceed the RPS. There is no known connection between the magnitude of a green tax, exogenously imposed, and the impact it will have on the generation fleet and the market. The proposed IEMM structure will deliver that.

Figure 52: A "Green Tax" Levied on Non-Qualifying Generators



Notice that the environmental implications of RECs and renewables entry are perfectly handled because every plant in the system—non-qualifying thermal as well as qualifying renewables alike—are represented using their environmental input-output coefficients. Furthermore, the cost of renewables enters the wholesale and retail price of electricity properly in this proposed IEMM construct. There is no question that the proposed structure handles renewables perfectly and completely. The renewables that enter do so because the agents that own them are made whole and profitable and add precisely the correct RPS quantity of renewables voluntarily. Furthermore, the distortion of the market in terms of the price of thermal energy in the region is correctly and properly calculated.

3 MODELING THE INTERACTION BETWEEN FUEL SWITCHING, PLANT RETROFITTING, AND NEW CAPACITY ENTRY

This section is the most sophisticated in this CDR, representing the interplay between fuel switching, retrofitting of plant and equipment to accomplish environmental remediation, and new greenfield generation entry. We begin with the most difficult type of thermal plant to represent—coal. We then move to the easier gas, oil, and nuclear plants. There is no model design we have seen that addresses the problems in this section in an integrated, interconnected fashion, as we do in this section and recommend for the design of IEMM.[12]

3.1 Coal-Waste-Wood Generation

3.1.1 Fuel Switching and Plant Retrofitting for Coal Plants

Fuel switching (particularly switching among coal types or from coal to fuel oil or gas or from fuel oil to gas) is one of the more important dimensions of the IEMM model and of many or most electrical systems throughout the world. IEMM has to carefully represent fuel switching/coal switching (which even the Chinese are considering right now!) as well as rehabilitating/retrofitting a fleet to control pollutants, which industrialized and industrializing electric systems are doing. Fuel switching interacts in a very complex fashion with the fleet of plant and equipment in place and the fleet of plant and equipment that could or will be built.

In the United States (and various other venues such as Europe), there are three or more pollutants for which retrofit equipment is chemically, physically, and economically straightforward—SO2, NOx, and mercury. CO2 is a bit more complicated; it requires sequestration of the plant effluent or some more aggressive chemistry, and it involves almost inconceivably high volumes of materials. CO2 is not chemically reactive like the other pollutants.[13] SO2, NOx, mercury, and CO2 retrofits are de facto a "greenfield capacity addition" problem because they are greenfielded onto an existing plant configuration when they were not there before. Adding SO2 removal (e.g., flue gas desulfurization or FGD) to arrest and remove SO2 is de facto greenfielding a scrubber onto an existing power plant. Some might argue that it is brownfielded because the plant site already exists. However, we believe it is greenfielded (built anew and new construction costs) contiguous to an existing plant. Assuredly the pre-retrofit plant and site already exist. CO2 mitigation via sequestration requires the construction of a new, greenfield plant component by assumption. We recognize that the plant is truly in fact a brownfield site. However, these retrofit technologies are to be regarded as greenfield, one-of-a-kind add-ons to an existing plant. We address elsewhere the notion that CO2 mitigation can, of course, be mitigated by switching from coal to gas or nuclear or renewables, and this possibility is thoroughly represented in the model structure presented here.

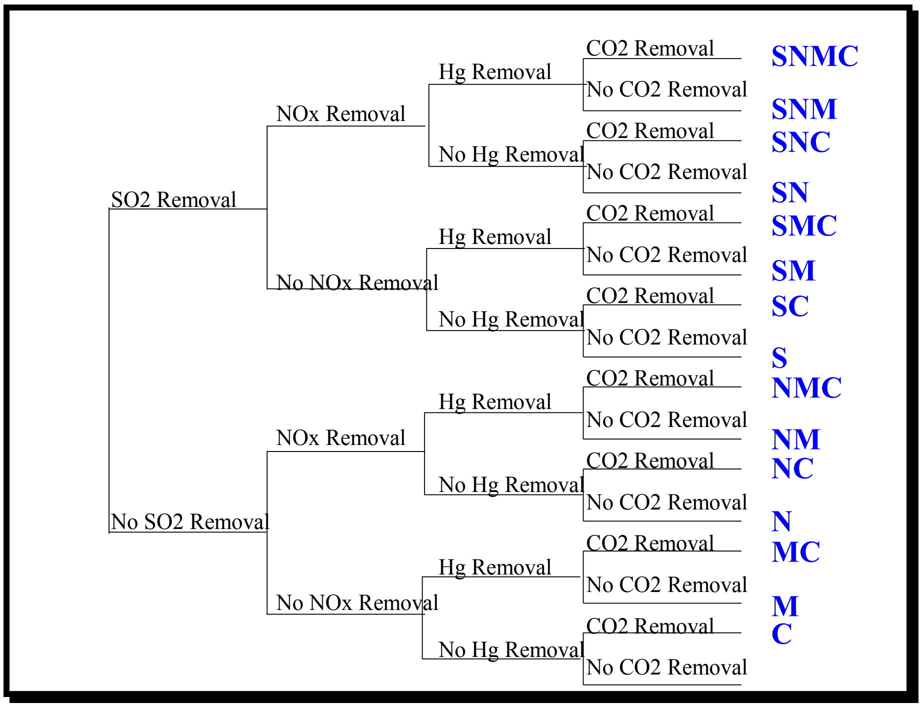
---

[12] Illustrations and diagrams in this section come from ArrowHead's regional electricity model.

[13] Keep in mind that the coal pile one sees is almost entirely carbon with an atomic weight of 6. When that coal pile is burned to CO2, the carbon "sticks" to two oxygens each with an atomic weight of 8 and thus produces CO2 with an atomic weight of 22. The weight of the effluent is 22/6 or 3.7 times the weight of the coal pile. When one sees a large coal stack in a power plant yard, keep in mind that the CO2 effluent from that plant will be almost 4 times larger. That puts into perspective the magnitude of CO2 removal that people are considering. And CO2 is literally as inert and nonreactive as a water molecule or a noble gas! It just doesn't easily chemically combine. That is understandable; so much energy was "squeezed" out of it during combustion.

The IEMM model is going to have to enumerate individual and combinations of SO2, NOx, mercury, and CO2 removal retrofits as shown in Figure 53 for many generation systems in the world.[14] (We use the notation S to designate SO2 removal, N to designate NOx removal, M to designate mercury removal, and C to designate CO2 removal.) If one of the four removal technologies happened not to be necessary in a particular regional power venue, it is very easy to delete it from consideration of it by simply redacting certain removal technologies from Figure 53. Notice that the paths through the diagram enumerate exhaustive combinations of NOx removal, SO2, removal, mercury removal, and CO2 removal/sequestration hardware that may be grafted onto a plant. (It is straightforward to consider more retrofit types, but these four are sufficient to represent to get a good answer for IEMM for American and other power systems.) There may be additional pollutants required, but the structure remains the same.

Figure 53: Combinations of Plant Equipment Installations



The equipment installations represented in Figure 53 interact in several important dimensions. First, the percentage removals of the various pollutant streams are not merely additive, multiplicative, or linear. They are nonlinearly synergistic, and the enumerative structure represents that combinatorially. Because removal technologies tend to be chemically reactive and chemically interactive, the plant pollutant coefficients are neither simply additive nor multiplicative through the combinatorial paths in Figure 53. We alluded to that in the illustrative removal calculations above, but this combinatorial structure cements that consideration even more firmly into the proper input-output coefficients for pollutants. Second, the operating cost changes induced by retrofitting are neither simply additive nor multiplicative. They are nonlinear and synergistic. Finally, the heat

[14] The tree could have more or fewer options than this, but the logic is fairly clear.

rate degradations are neither simply additive nor multiplicative but are themselves nonlinear and synergistic. IEMM must be able to represent all three types of synergies (sometimes euphemistically referred to as "cobenefits" in the electric and environment industry), and this design will do precisely that. As a fourth dimension of synergy, combinations of equipment installations can have nonlinearly related capital costs. There are economies of scale here. One gets two pollutants impacted together for less the sum of the individual costs of each pollutant in isolation. IEMM will need to estimate and insert emissions coefficients, the heat rates, and the capital costs for all the combinations of installations, not merely individual incremental installations. Later sections explain.

Equipment installations are not the only way to change plant operation or change plant emissions. Fuel switching does the same thing. We will continue with the discussion of hardware for the time being, and then we will introduce fuel switching against that backdrop.

3.1.2 Fuel Substitution and Discrete "Exemplar" Coal Types

IEMM might be tempted to consider tens or hundreds of steam coal types, one for every distinction in coal chemistry or impurity content. That level of detail is not operationally useful for IEMM. Instead, from a practical perspective, IEMM needs to aggregate coal types, defining a modest, manageable, representative set of "exemplar" coals. (We use the word "exemplar" to mean an aggregate with properties representative of each of the coals that is aggregated into the category.) Ultimately, one is compelled to aggregate or "bin" coals into exemplar coal types. The reason is simple based on the particular stoichiometry presented earlier. Managing the specific number of coal types or bins is important. IEMM must trade off the number of coals (and thus the number of "bins") against the incremental benefit (with regard to predictability) from adding additional coal "bins." It is important to understand this "binning" technique so that one can understand the extension of the technique to add as many more bins as are needed in a particular IEMM venue. We suspect there might be more bins in a Chinese generation system, for example, than in a Middle East generation system.

As an illustration, coals might be "binned" as shown in Figure 54. In categorizing coals, the heat content of the coal (the dimension down the rows) can be important. The sulfur content of the coal can be important. The mercury content of the coal may be important as well; there might or might not be a separate mercury dimension. There may be other pollutants. All coals in the model have to be categorized into one of the selected bins. There is no operational upper bound on the number of coal bins that can be considered, but each additional bin increases the dimensionality and execution time. There is no country power system model that has any degree of coal generation within it that can be correct without such considerations.

Binning of coals is precisely the analog to what has been done with crude oils of different assays going into refining models. Precisely as is the case with the many different coals, there are many different crude assays—high gravity crudes, low gravity crudes, high-sulfur crudes, high aromatics crudes, high paraffin crudes, etc., and each has substantially different implications for crude oil refining and product mix. By analogy, there are a large number of different types of coals, and they produce electricity and generate pollutants differently in a given power plant configuration.

We have categorized the various coals into bins so as to keep the dimensionality of the IEMM model under control. Figure 54 is put forth as an example, and probably a pretty good one, of what IEMM will have to ponder.

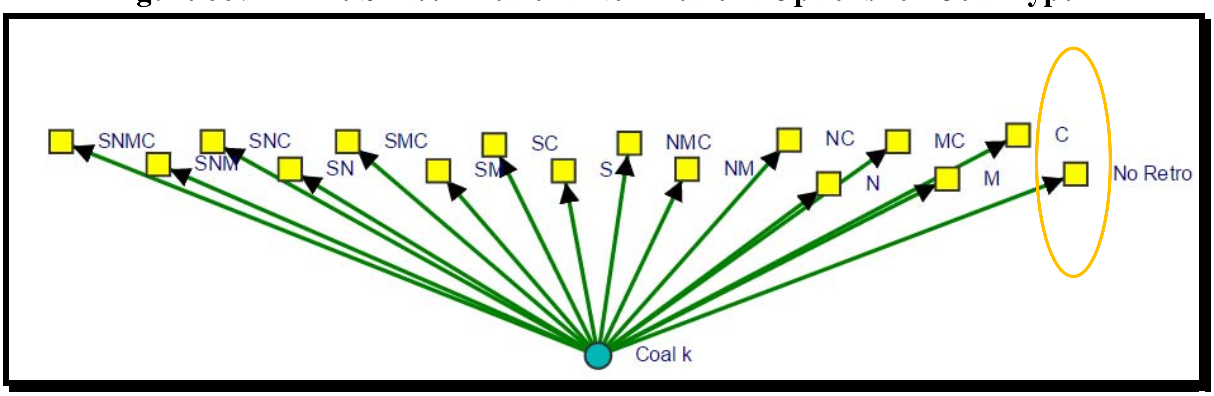
Figure 54: Six Potential Coal Bins

	Low sulfur	High sulfur
Bituminous		
Sub-bituminous		
Lignite		

3.1.3 Interplay between Retrofit Options and Coal Use

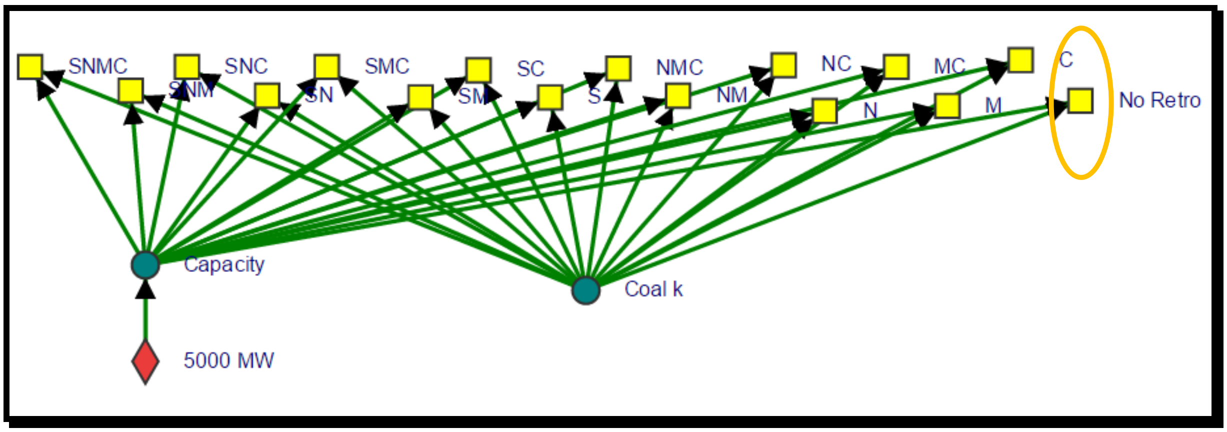
This section characterizes, using the preceding retrofit and coal type distinctions, how IEMM should be built and how it represents all the interacting configurations enumerated in the hardware retrofit enumerations in Figure 53. We use the ArrowHead network representation and conventions to show graphically how IEMM should represent those options. We draw the diagram of the exhaustive set of retrofit and non-retrofit options using a network format as shown in Figure 55. Each of the generation options is seen to draw from a particular coal type, so we must repeat this diagram for every one of the assumed exemplar coal types, i.e., to replicate the diagram in Figure 55 for each coal type. Every node in Figure 55 represents one full combinatorial path through the tree in Figure 53, and the diagram, in fact, exhausts and enumerates all the paths. In fact, Figure 55 is a network-centric enumeration of installed equipment combinations in Figure 53.

Figure 55: The Sixteen Retrofit-Non Retrofit Options for Coal Type k



If there existed say 5,000 MW of existing capacity embedded in a completely uncontrolled, non-retrofit plant in the region an original plant, i.e., the upper right node in Figure 55 that is circled in orange, each of these sixteen plant configurations (represented as nodes in the diagram) must “compete” to use this 5,000 MW of installed generation capacity. That is, this 5,000 MW must be expended in some combination on the sixteen possible plant configurations. To represent the retrofit and utilization of this capacity, we build the representation in Figure 56, augmenting the representation from Figure 55 to make sure that generation capacity in place never exceeds the installed 5,000 MW and therefore that retrofit additions of every type do not increase the installed 5,000 MW. Aggregate generation capacity before and after retrofitting will be 5,000 MW, no more and no less. The model will operate the retrofit plus non-retrofit capacity so that the sum total of retrofit plus non-retrofit capacity does not exceed 5,000 MW. That capacity limitation will also limit any coal substitution that might occur, as we shall see shortly.

Figure 56: The Sixteen Retrofit-Non Retrofit Options for Coal Type k with Capacity Limitation



The inputs to each of the yellow retrofit generation nodes and the yellow non-retrofit nodes are extremely important. We begin with the “No Retro” node at the upper right to illustrate. The inputs to that generation node are indicated conceptually in the diagram in Figure 57, precisely and identically as they were in Figure 32. (We have included the four environmental inputs and the REC input, which are not yet specifically drawn in Figure 56.) We have drawn in Figure 57 this “No Retro” node complete with all the input markets to which it is hooked and the output markets to which it is hooked and all the capacity sources to which it is hooked. We have drawn it using the ArrowHead network drawing interface in Figure 58. We see the direct analogy, in fact, the direct parallelism, between Figure 57 and Figure 58. The labels in Figure 58 are slightly altered, but that is the sole difference.

We see in Figure 57 and Figure 58 the four pollutants entering the generator, reflecting the fact that this uncontrolled, non-retrofit generator will have to purchase sufficient SO2, NOx, mercury, and CO2 allowances to generate at whatever MWh level it chooses to generate. Indeed, the SO2, NOx, mercury, and CO2 input-output coefficients quantify how much SO2, NOx, mercury, and CO2 will be produced with each MWh of electrical energy that is produced and therefore how many SO2, NOx, mercury, and CO2 allowances the generator will have to obtain (through purchase on the open market or through allocations by a regulator) to generate. We see the need for a REC coming into this node to enable its generation. It is not sufficient for the generator to merely offer four environmental allowances. It must also offer the REC or green energy allowance. We see the coal type k coming into this node, meaning that its input-output coefficients and non-derated capacity are specific to this coal type. If there were a different coal type, these input-output coefficients will all be different. There is the capacity link coming into this node, meaning that the aggregate generation by this node cannot exceed the MW capacity in place. Indeed, the capacity input will carry the signal that there is 5,000 MW of capacity available to be used for power generation in this node, and no more.

Figure 57: The Non-Retrofit Electricity Generation Node (Upper Right in Figure 56)

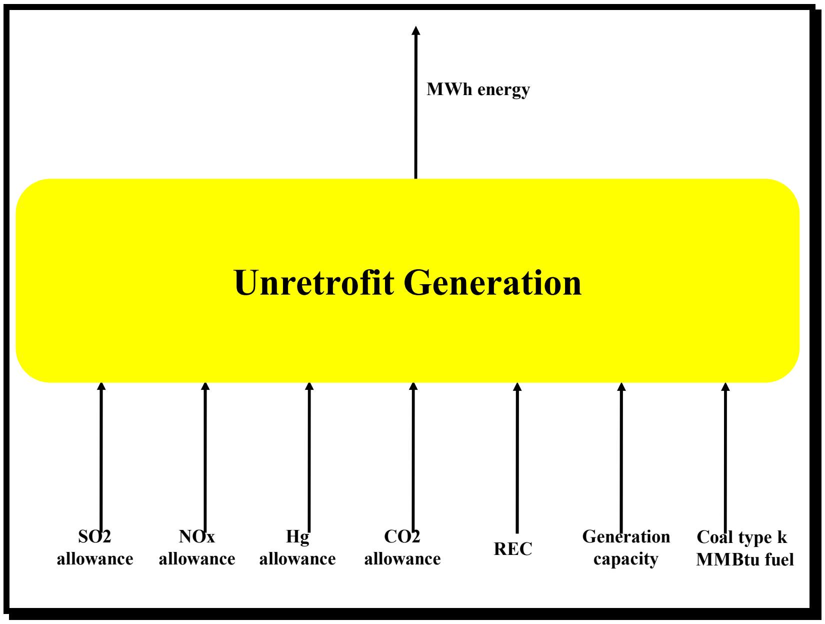
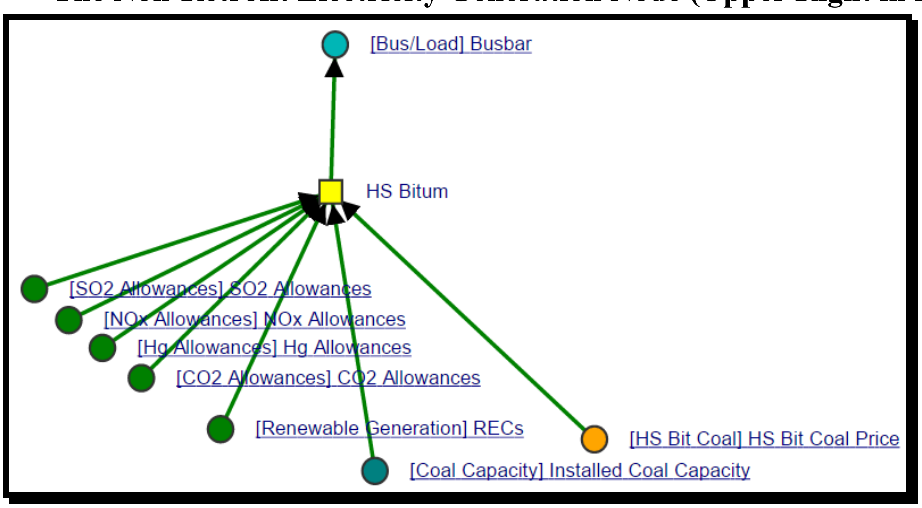


Figure 58: The Non-Retrofit Electricity Generation Node (Upper Right in Figure 56)



For each of the seven inputs to the non-retrofit electricity generation node, we relate the quantity of input to the quantity of output (MWh). Economists term such relationships "input-output coefficients," and we adopt that terminology. IEMM should focus strictly on input-output relationships, the direct analogy of heat rates, throughout its IEMM design. IEMM could appeal to efficiencies as they do in Europe, but we find that to be a bit confusing. Input-output relationships were discussed in extensive detail previously, but in IEMM these concepts need to be put to work. These input-output coefficients are specific of course to the coal type used. (This is why we need a number of discretely binned coal types.) The input-output coefficients required to quantify the non-retrofit electricity generation node in Figure 57 appear in Figure 59.

It is worthwhile to walk through the input-output coefficient calculations to emphasize how and why they work and are crucial to the IEMM design. The seven input-output coefficients, depicted carefully for clarity in Figure 59, are the following. The first input, the capacity input, is the following:

$$\text{Capacity} = 1 \frac{\text{MW input}}{\text{MW output}}$$

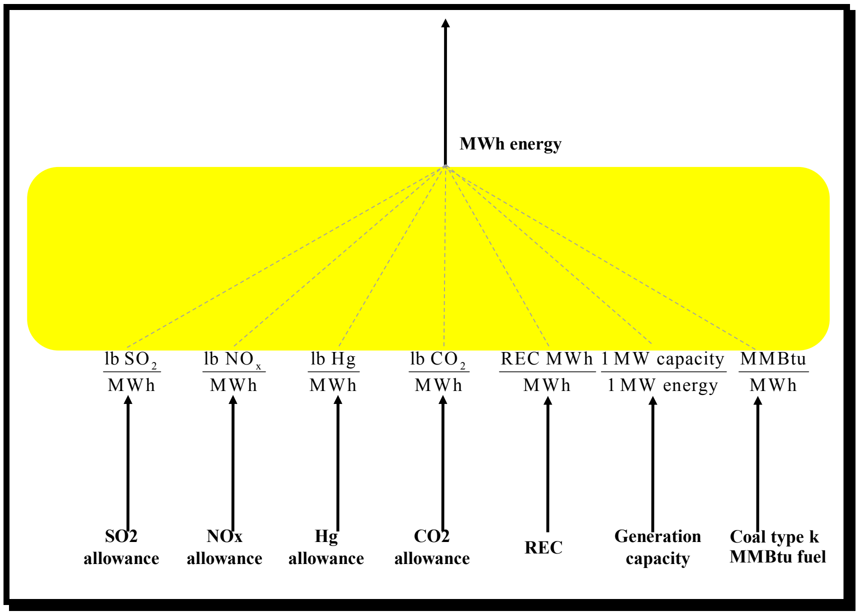
This generation node needs 1 MW of the total 5,000 MW of capacity if it wants to produce 1 MW of output (or equivalently 1 MWh in each hour or some fraction of the hours) using this non-retrofit, uncontrolled generation technique. By assigning every MW in the region (every one of the 5,000 MW) to one and only one of the plant configurations, we are able to constrain the total capacity that is either non-retrofit or retrofit in any configuration to be lower than or equal to the 5,000 MW that is actually in place. The region can choose to operate any number of the 5,000 MW for generating electricity in this uncontrolled, non-retrofit, non-sequestering plant configuration, but it can never deploy more than 5,000 MW. Similarly, if we have other configurations (with particular, installed retrofit combinations), we will be able to use those to limit the amount of utilization of retrofit equipment to be less than or equal to actual installed retrofit equipment as well. This notion of capacity and aggregate capacity utilization is fundamental to the IEMM design. You need 1 unit of capacity to make 1 unit of product in each hour, and you must use each MW in one and only one generation configuration. If you don't have that 1 MW of capacity assigned to your plant configuration, you don't have a way to make product using that plant configuration. This notion is not particularly different in concept from the fact that you need some number of units of fuel to make 1 unit of product. The representation here is that you must combine 1 unit of capacity plus some number of units of fuel to make 1 MWh of product. By representing capacity in this fashion, we are going to be able to consider expansion of capacity in an absolute sense and also in a retrofit sense. It all becomes endogenous, both addition of new capacity and addition of retrofit hardware to ameliorate emissions. IEMM will need this for sure.

Heat rates h are invariably expressed in terms of Btu/KWh, and they are converted as follows:

$$\text{HeatRate} = \text{h} \frac{\text{Btu input}}{\text{KWh output}} * \frac{1\text{MMBtu}}{10^6\,\text{Btu}} * \frac{10^3\,\text{KWh}}{1\text{MWh}} = \frac{\text{h}}{10^3} \frac{\text{MMBtu}}{\text{MWh}}$$

Thus, the second input, the heat rate input, represents the quantity of fuel needed to produce 1 MWh of output. Heat rate is a very standard and obvious input for each plant configuration. It is related in detail to the discussion in previous sections of input-output coefficients and quadratic (or other) relationships. This discussion assumes a heat rate constant with output, although IEMM will not.

Figure 59: The Non-Retrofit Electricity Generation Node (Upper Right in Figure 56) with Input Output Coefficients Carefully Enumerated



If we express coal (or other fuel) prices, p_c, in terms of \$/MMBtu, IEMM can make the dimensionally correct calculation in terms of \$/MWh for the coal cost portion of total operation cost:

$$\text{CoalCost}(\$/\text{MWh}) = p\frac{h}{10^3}\frac{\$}{\text{MWh}}$$

This calculation ensures that the coal cost properly enters the operating cost expressed in terms of \$/MWh. IEMM has to make sure that all the other pollutants enter the calculation in precisely the same fashion, and this design highlights that.

We now move to the SO2 input-output coefficient. A typical input-output coefficient would be expressed in terms of pounds of SO2 per MWh of generation output, precisely as we have discussed previously. The SO2 coefficient is “input” (emissions of SO2 ejected to the environment) divided by “output” (MWh of power produced). The units of output are MWh, and the units of input are pounds of SO2. The input/output coefficient is analogous to a heat rate, namely pounds of SO2 produced per MWh of generation.

We re-summarize how we obtain the SO2 input-output coefficient, even though we went through it is some detail in Section 2. We summarize the fundamental coal chemistry and stoichiometry of the plant as follows:

$$1.1\,\%\text{ sulfur coal at }20\text{MMBtu / ton}$$

$$0.011*2000\text{lb}=22\frac{\text{lb S}}{\text{ton coal}}=44\frac{\text{lb SO2}}{\text{ton coal}}$$

We briefly summarize the stoichiometric calculation made previously in Section 2 for this non-retrofit, uncontrolled plant configuration. The reason the weight doubles is because sulfur has an atomic weight of 16 and oxygen has an atomic weight of 8. That is, when the two oxygens bond with the one sulfur, the weight of the sulfur doubles. If we assume that the coal has a heat content of 20 MMBtu/ton or 10,000 Btu/lb, the SO2 content per MMBtu of coal would be:

$$IO_S=\frac{44\dfrac{\text{lb SO2}}{\text{ton coal}}}{20\dfrac{\text{MMBtu}}{\text{ton coal}}}=2.2\frac{\text{lb SO2}}{\text{MMBtu coal}}$$

This equation is basic stoichiometry. If this coal were to be burned in the 11,000 Btu/KWh plant considered here, we could calculate the pounds of SO2 that were released per MWh of power generation as follows:

$$2.2\frac{\text{lb SO2}}{\text{MMBtu coal}}*11\frac{\text{MMBtu coal}}{\text{MWh}}=24.2\frac{\text{lb SO2}}{\text{MWh}}$$

This relationship provides the number that goes into the IEMM generation node, pounds of SO2 released per megawatt hour generated. It ensures in IEMM that the output from the power generation node is properly expressed in terms of MWh, and the input to the power generation process is properly expressed in terms of pounds of SO2. It will further ensure that the price of the SO2 emission allowances expressed in terms of \$/lb, denoted p_S, will properly enter the operation cost calculation as follows:

$$SO_2\text{ Cost}\left(\$/\text{MWh}\right)=p_S 24.2\frac{\$}{\text{MWh}}$$

We have to make this calculation for every pollutant for this non-retrofit plant.

Consider next the renewable requirement for generation. The REC input-output coefficient specifies the number of REC MWh that are required to enable generation of 1 MWh of non-REC-qualifying MWh. If we want to consider a 20 percent RPS, we saw previously that we should set the REC input-output coefficient at 0.25 MWh REC/MWh. Thus, the REC cost to generate would be calculated:

$$\text{REC Cost}\left(\$/\text{MWh}\right)=p_R\frac{\$}{\text{MWh REC}}0.25\frac{\text{MWh REC}}{\text{MWh}}$$

3.1.4 Retrofit Nodes—Data and Structure

Retrofit nodes are exactly analogous to non-retrofit nodes as analyzed previously with one exception—they need a second type of installed capacity, namely the requisite retrofit hardware. As an example, a coal plant that wants to remove SO2 would have to have an SO2 removal technology (e.g., a wet limestone scrubber or something similar) installed. A coal plant that wants to remove NOx would have to have a selective catalytic converter (SCR) or something similar installed. Plants would have to install activated carbon injection (ACI) if they are to remove mercury, and CO2 removal if they are to remove or arrest CO2. We begin to understand that the generation node needs a second kind of capacity, namely the installed hardware capacity enumerated in the tree in Figure 53, for each type of plant retrofit capacity that has been installed on the uncontrolled plant.

Alas, such retrofits do not occur for free. There are significant capital costs involved. Plant owners have to spend capital dollars to reduce their environmental input-output coefficients. Retrofits cause adjustments in performance in the plant, and those adjustments come at an operating cost premium. Retrofit hardware requires energy (often called parasitics) to run and thus increases plant heat rate (i.e., reduces plant efficiency) as it saps energy for the pollutant removal.

In the forthcoming discussion, we assume a wet limestone scrubber as being representative of SO2 removal, selective catalytic reduction (SCR) as being representative of NOx removal, activated carbon injection (ACI) in conjunction with a fabric filter as being representative of mercury removal, and some sort of unspecified sequestration process for CO2 removal as indicated in Figure 53. The input-output coefficient data for these four pieces of hardware are potentially different for every node and coal type in the model, but we will try to be representative in the forthcoming discussion. We are using data here that is representative of a particular regional IEMM model to illustrate the calculations. No proprietary or confidential data are put forth here. Actual regional IEMM data will be gathered by EIA. Actual data may be somewhat country specific, particularly things such as installed retrofit capacity today.

It is illustrative to walk through some specific numbers for the non-retrofit plant configuration in Figure 59, which were calculated by applying the various SCR, scrubber, and mercury removal technologies to obtain the performance of the retrofit plant. As stated previously, generally an installed scrubber removes approximately 49/50 of the SO2, an installed SCR removes approximately 9/10 of the NOx, and an installed ACI with a fabric filter removes approximately 9/10 of the mercury. Such removal levels are typical of modern state of the art scrubbers, SCR, and mercury removal/fabric filters that could be retrofit onto plants today. These technologies indeed modify the input-output coefficients (emissions coefficients) for the various plant types from their values in the top sections of Figure 18 to the values in the bottom of Figure 18.

Notice for the supercritical, bituminous type coal in line 3:

- SO2 emissions input-output coefficients drop from 22.0 lb SO2/MWh to 0.462 lb SO2/MWh

- NOx emissions input-output coefficients drop from 29.45 lb NOx/MWh to 3.09 lb NOx/MWh
- mercury emissions input-output coefficients drop from 9.9275E-4 lb mercury/MWh to 1.0424E-5 lb mercury/MWh.

if we retrofit those units with control technology. (The decreases are not straight percentage decreases because retrofit equipment increases the plant heat rate (i.e., reduces plant operating efficiency) to run the additional equipment and endure the additional backpressure and the like.

If we were to retrofit the unit with an expensive CO2 capture and sequestration technology:

- CO2 emissions input-output coefficients drop from 2053 lb CO2/MWh to 215.6 lb CO2/MWh.

(Notice that CO2 emissions control would get a coal plant well below CPP and other potential output limits.) Plant retrofits, however, cost money. There is a $/MW charge to install an SCR and a $/MW charge to install a scrubber. There is a $/MW charge to install both an SCR and a scrubber. Thus, the efficiency equivalents in the bottom of Figure 18 are accompanied by a $/MW capital charge. There is a capital cost required to achieve them. There is also an increase in operating cost when we install extra equipment on a plant, and we have calculated the operation cost in the non-retrofit case in the top of Figure 18 and for the retrofit case in the bottom of Figure 18. There is no free lunch where capital equipment, thermodynamics, and operating chemistry are concerned.

Referring to the structure in Figure 60, the completely non-retrofit plant, there is one more type of capacity IEMM would need to be able to generate with SO2 removal. The node would have to have SO2 removal capacity in place. One does not get to operate using the input-output parameters of a generator that has an installed SO2 removal until and unless the facility has paid the cost to have SO2 removal capacity installed. You would have to have an input from aggregate capacity in the region as shown in Figure 60, but you would have to have an input from installed SO2 removal capacity in the region. The configuration of a unit with SO2 removal installed would, therefore, take the augmented form in Figure 61 and its equivalent in Figure 62. Notice that a lot of things are identical to the uncontrolled case in Figure 60 except that there is an additional link to account for capital installation of SO2 removal and except that most of the input-output coefficients are different because of the reduced plant efficiency to run the SO2 removal. This node cannot run until and unless there is some capacity incoming to the node and unless there is some SO2 removal capacity incoming to the node. We are going to see that we will be able to endogenously add both types of capacity, so SO2 removal addition and SO2 amelioration by retrofitting become endogenous.

What about a retrofit that has both SO2 removal **and** NOx removal installed? That network would appear as in Figure 63, containing an input link from SO2 removal capacity **and** an input link from NOx removal, both of which are required if one is to operate a generator that has the parameters of such a plant.

What about a retrofit that has all four--NOx removal, SO2 removal, mercury removal, and CO2 removal (sequestration) retrofits installed? See Figure 64 for the representation.

Figure 60: The Material Balance for the Completely Non-Retrofit Node—Bituminous Pulverized Coal Plant

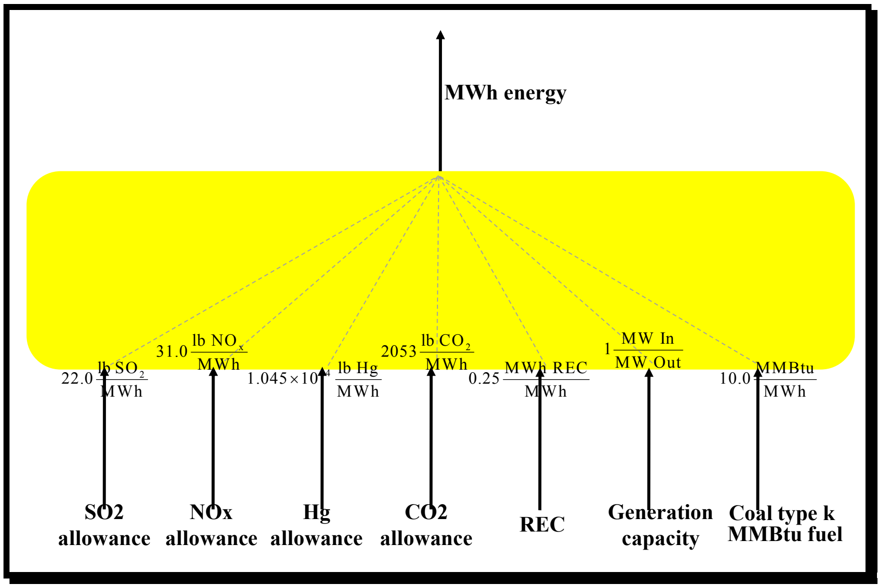
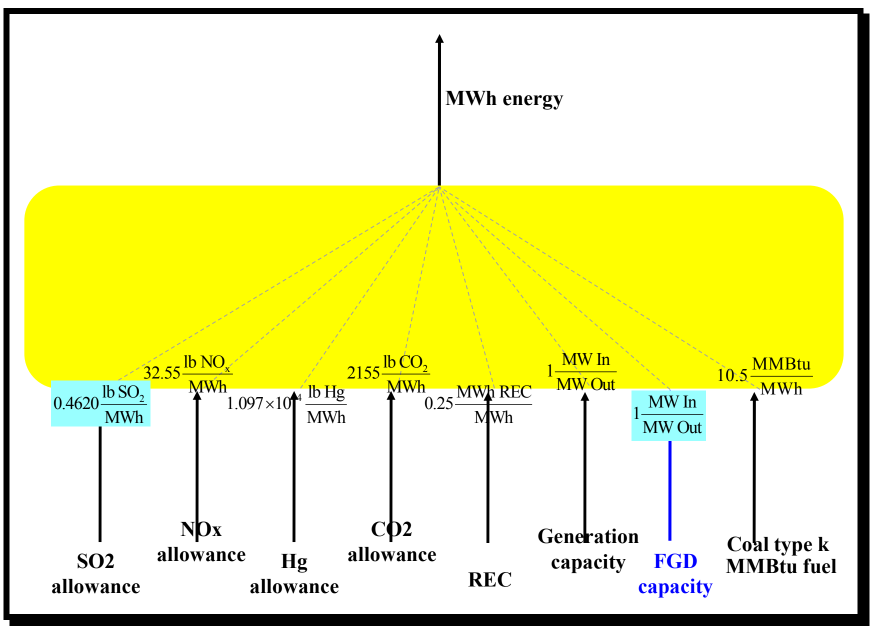


Figure 61: The Material Balance for the SO2 Removal Retrofit Node—Bituminous Pulverized Coal Plant



By considering the capacity that a plant needs to deploy depending on the installed equipment, the simple generation node configuration we have here can represent the full range of environmental emissions, retrofit or remediation options, and capacity expansion. There is no substitute for this when modeling North American, European, and other energy systems using IEMM. IEMM is going to have to represent the complete range of plant capacities and capacity additions and equipment retrofits. How many times have we heard that we want China to consider carbon capture and sequestration and/or substitution of gas for coal? How many times have people complained that China is releasing some huge fraction of world mercury into the air, and it is blowing into the United States? IEMM under this design will model that. EIA has to be able to model retrofit and modification of the Chinese system and every other system in the world. This IEMM design allows you to do so. Typical dispatch or global social welfare maximization approaches have no chance of doing so.

Figure 62: Network and Markets for the SO2 Removal Retrofit Node—Bituminous Pulverized Coal Plant

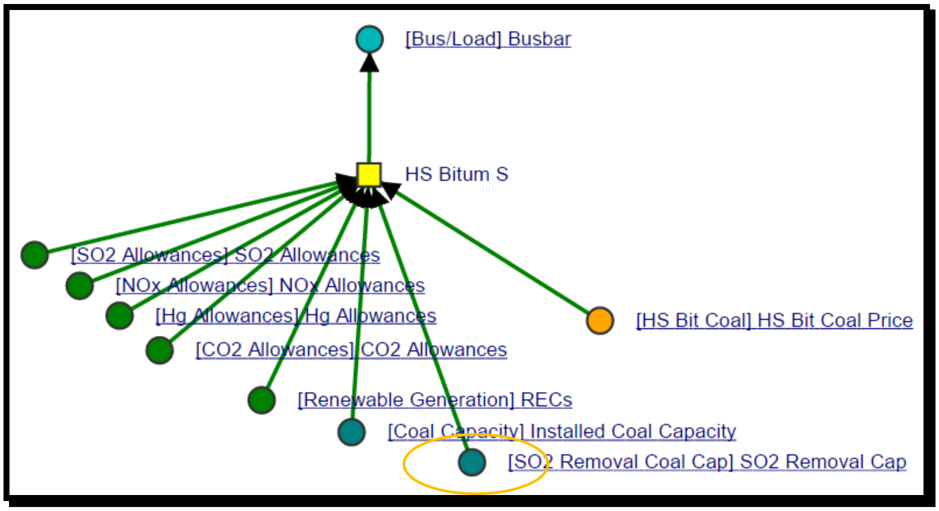
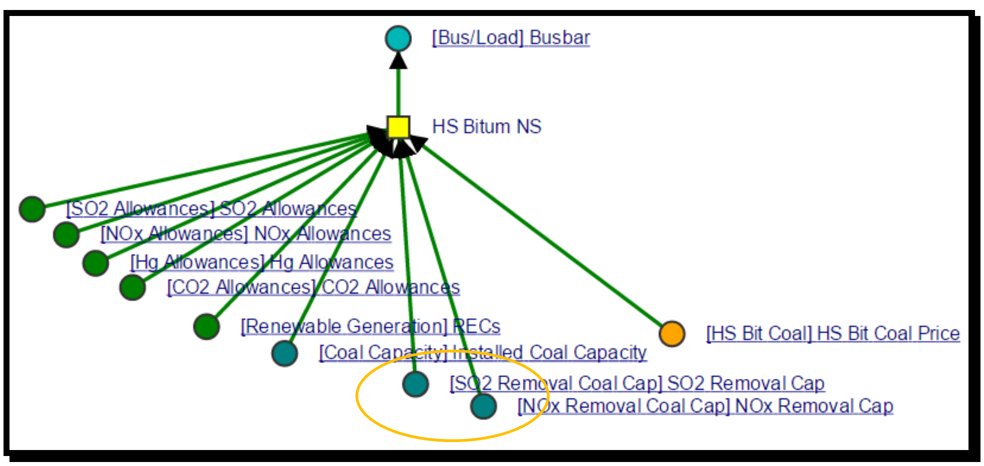


Figure 63: Network and Markets for the NOx Removal and SO2 Removal Retrofit Node—Bituminous Pulverized Coal Plant

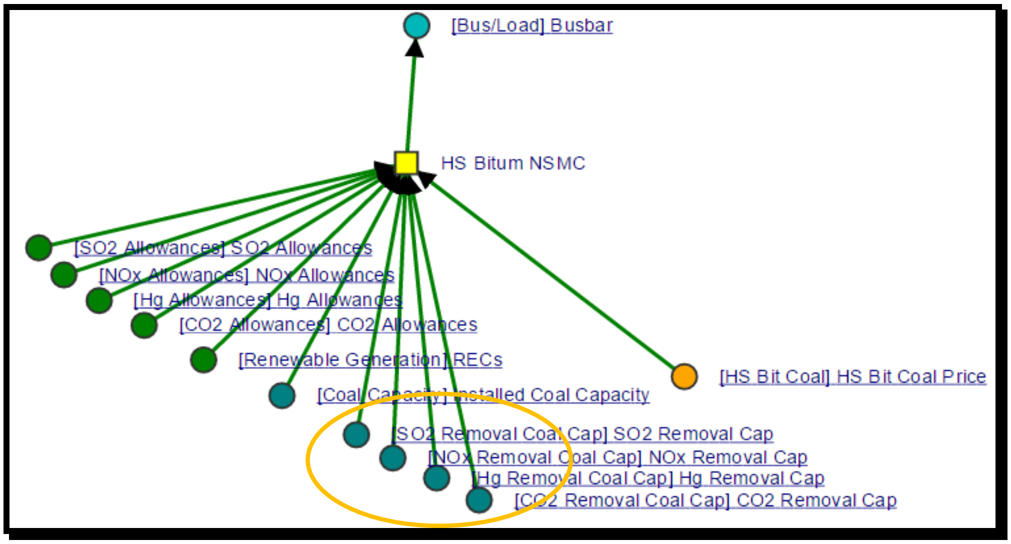


3.1.5 Summary of Concepts Thus Far

Consider coal type k. For this specific coal type, Figure 65 is the configuration of the operation sector and the capacity addition sector. Notice the following:

- There are 16 operating configurations, one for each possible combination of retrofit/environmental-removal configurations. This was clear in the equipment installation tree in Figure 53. If we change the retrofit options that are available, the number of configurations will change. If additional retrofit technologies were to be considered, they would have to be added to this collection of sixteen configurations. We will see fewer configurations for gas. Oil will be more like coal.

Figure 64: Network and Markets for the NOx removal/SO2 Removal/Mercury Removal/CO2 Sequestration Retrofit Node—Bituminous Pulverized Coal Plant



- For each of the 16 operating configurations, there must be an input for the aggregate capacity in the region. That ensures that the sum of all 16 operating configurations does not exceed the 5,000 MW of coal capacity that exists in the region. This aggregate capacity input link ensures that the installed capacity in the region is not violated and that in fact installed capacity in the region can be reduced over time.

- For each of the 16 operating configurations, the environmental emissions coefficient must be input for that specific plant configuration:

  - SO2, expressed as pounds of SO2 emitted per MWh generated.
  - NOx, expressed as pounds of NOx emitted per MWh generated.
  - Mercury, expressed as pounds of mercury emitted per MWh generated.
  - CO2, expressed as pounds of CO2 emitted per MWh generated.

If there are any synergies or "cobenefits" of various combinations of retrofit technologies, they will be incorporated in the slate of SO2, NOx, mercury, and CO2 emission coefficients for each combination. It is the slate of emission coefficients that matters. If an operating configuration in the model indicates:

- S, indicating SO2 removal (e.g., scrubbing), that configuration must accept as input a link from the SO2 removal capacity node. If a configuration has SO2 removal on it designated by an S, it has to have a capacity link from the SO2 removal capacity hub. Its capacity input-output coefficient for that capacity link will be 1. That is, if a plant has SO2 removal on it, it has to buy SO2 removal capacity from the SO2 removal node, and it has to buy capacity from the total capacity node, both with input-output coefficients of 1.

- N, indicating NOx removal (e.g., SCR), that configuration must accept as input a link from the NOx removal capacity node. If a plant has NOx removal on it, it has to have a capacity link from the NOx removal capacity hub. Its capacity input-output coefficient for that capacity link will be 1. That is, if a plant has NOx removal on it, it has to buy NOx removal capacity from the NOx removal node, and it has to buy capacity from the total capacity node, both with input-output coefficients of 1.

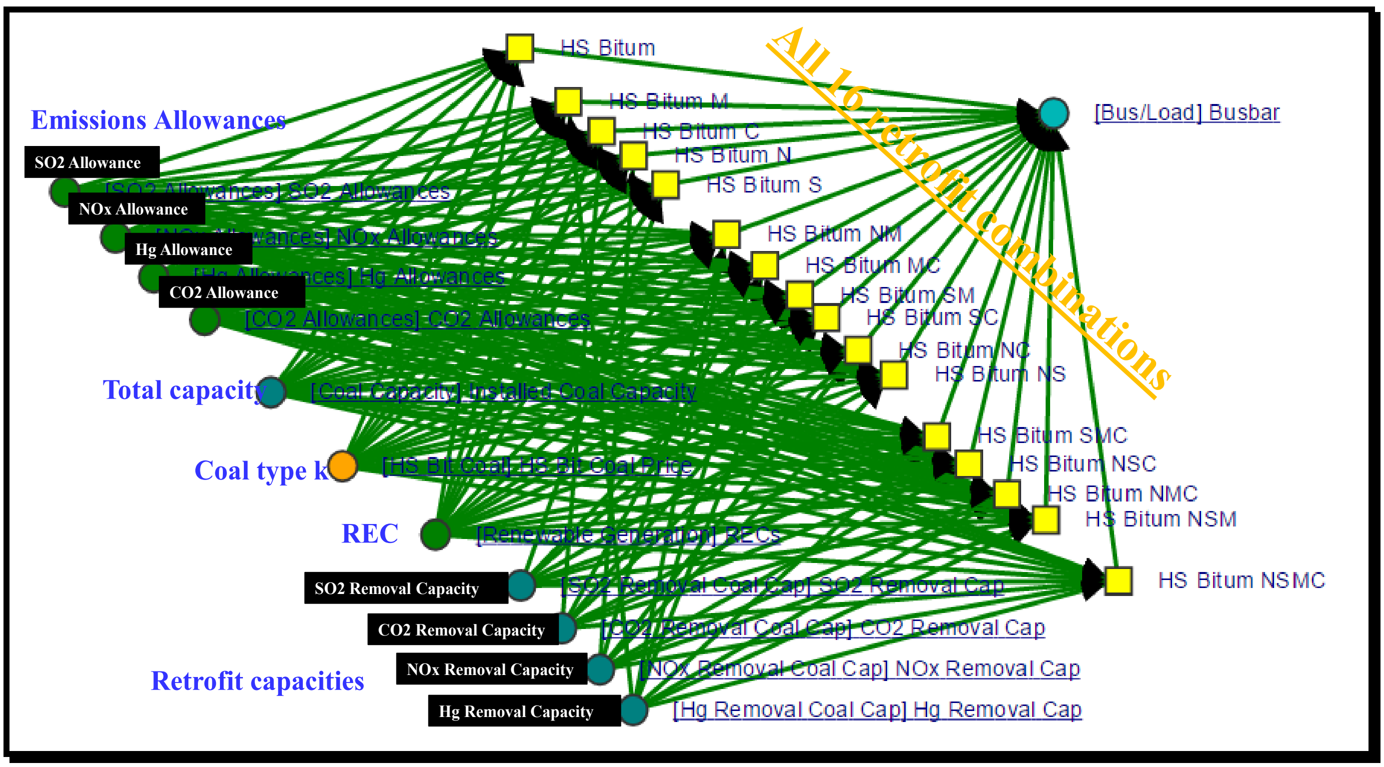
- M, indicating mercury removal (e.g., ACI), that configuration must accept as input a link from the mercury capacity node. If a plant has mercury removal on it, it has to have a capacity link from the mercury removal capacity hub. Its capacity input-output coefficient for that capacity link will be 1. That is, if a plant has mercury removal on it, it has to buy mercury removal capacity from the mercury removal capacity node, and it has to buy capacity from the total capacity node, both with input-output coefficients of 1.

- C, indicating CO2 removal (e.g., sequestration), that configuration must accept as input a link from the CO2 removal capacity node. If a plant has CO2 removal on it, it has to have a capacity link from the CO2 removal capacity hub. Its capacity input-output coefficient for that capacity link will be 1. That is, if a plant has CO2 removal on it, it has to buy CO2 removal capacity from the CO2 removal node, and it has to buy capacity from the total capacity node, both with input-output coefficients of 1. It is these capacity links that "enable" that corresponding piece of equipment to operate. That is, the CO2 removal link "signals" that particular operating facility with the CO2 removal configuration installed that it can operate.

A configuration such as this allows IEMM to fully consider everything about coal type k that is relevant. It allows IEMM to fully represent generation, both in an economic setting and in an environmental setting. This generation model allows consideration of all the environmental pollutants (whatever the list EIA has specified for each particular country or region), the full range of environmental retrofits and equipment additions, the complete range of coal types and, therefore, coal substitution, and renewables considerations via RECs.

The notion of plant operations is clear in Figure 65. Plants buy capacity from the various capacity nodes commensurate with their retrofit configuration. The amount of capacity they buy from the five capacity nodes—total capacity in the region node, SO2 removal capacity node, NOx removal capacity node, mercury removal capacity node, and CO2 removal capacity node. The total removal capacity in the region node is fixed at the removal capacity that exists in the region. However, at an assumed capital cost of say \$425,000/MW for SO2 removal capacity, \$325,000/MW for NOx removal capacity, \$105,000/MW for mercury removal capacity, and \$2,600,000/MW for CO2 removal capacity, the model can expand the SO2 removal, NOx removal, mercury removal, and CO2 removal capacity nodes respectively. Then that expanded node sends out price information to all the operating plant configurations, and the ones that bid the most for that capacity are the ones that operate. That allows IEMM to quantify the environmental remediation equipment, if any, that will be installed in the given region being modeled and fully price it into the system. It has endogenous capacity addition of all the retrofit techniques. There is no substitute for this in IEMM, which is chartered to understand environmental issues associated with generation as well as generation around the world.

Figure 65: Generation and Retrofit Alternatives for Coal Type k as Represented in IEMM



To amplify, the capital costs of the various retrofit technologies are located in the retrofit technology capacity addition nodes at the bottom of Figure 65. The output of those nodes represents capacity available for use by all the plant configurations. Those capacity nodes are going to be allowed to build endogenously within IEMM, the specific methodology to be described in a later section. None of the plant operation nodes build endogenously; they have unlimited capacity; there is no limitation on the plant operation nodes. It is the interaction of all five of the capacity nodes that governs which plant configurations exist and operate. To wit, the operational and environmental limitations are taken care of in the capacity nodes.

If one examines the configuration in Figure 65, there are only **five** nodes that contain any notion of capacity in place. We will see, therefore, there are only five nodes therefore that take account of endogenous capacity addition. Those five include total capacity in place, SO2 removal capacity, NOx removal capacity, mercury removal capacity, and CO2 removal capacity.

3.1.6 Fuel Substitution—A Second Coal Type

We are not aware that anyone has represented the coal substitution-equipment modification aspect of anyone's power system besides ourselves and MarketPoint. We do so as described herein. Suppose we want to add a second coal type. The structure becomes that in Figure 66, which is a simple and very direct expansion of the network in Figure 65. Notice that the five types of capacity—total capacity, SO2 removal capacity, NOx removal capacity, mercury removal capacity, and CO2 removal capacity—span both of the coal types. That means that the capacity in place must not exceed the capacity in the region for any of the five types of capacity. This structure trades off coal substitution against plant equipment modification via retrofit. Think how powerful this will be for IEMM, not only in an industrialized electrical system in which CO2 and pollutant removal are already current, hot regulatory and economic issues but also for developing country electrical systems where it is not (yet). IEMM also needs to be positioned for developing countries and electrical systems in which it becoming an issue. There is no electrical system in the world in which CO2 is not an issue, whether the host country likes it or not. The IEMM structure here allows EIA to consider the issue of plant emissions, retrofits to reduce emissions, and fuel switching. This analysis is for the most complicated fuel of all, namely coal. Oil and natural gas are much easier, as we shall see. And yet because coal (and its equivalent wood and waste) are so prevalent and important worldwide, there is no choice but to deploy this type of structure.

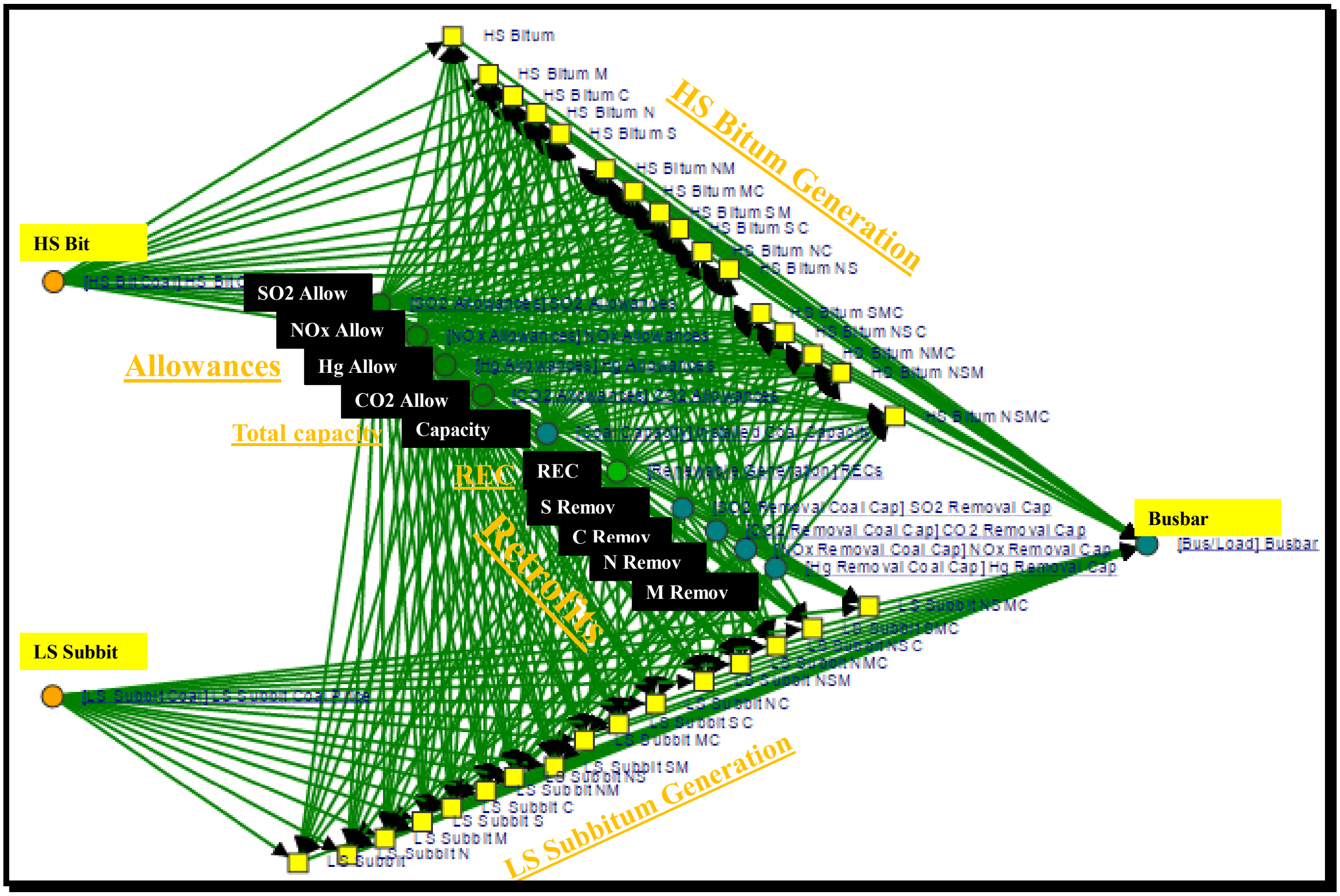
Note the following

- The sixteen operating configurations are repeated for each of the two coals, which we have chosen to be LS Sub-bituminous coal and HS Bituminous coal. Distinguishing these two coals is rather typical. (We could distinguish as many more as we want.)

- The sixteen emissions coefficients are specific to each coal type for each specific plant configuration. That is, the sixteen emissions coefficients in the HS Bituminous collection of sixteen operating configurations are specialized to HS Bituminous coal depending on the particular equipment installed on the plant. The sixteen emissions coefficients in the LS Subbituminous Coal operating configurations are specialized to LS Sub-bituminous coal depending on the particular equipment installed on the plant. For HS Bituminous versus LS Sub-bituminous coal, the sixteen emissions coefficients will potentially be very different. IEMM might decide to substitute coal if that is the best way to get rid of SO2 or mercury, and EIA definitely wants that capability within IEMM.

- The heat rates and plant operating parameters for the sixteen operating configurations for HS Bituminous coal will be different from the sixteen operating configurations for LS Sub-bituminous coal because of the differences in coal chemistry and combustion.

Figure 66: Adding a Second Coal Type



- The total generation capacity in the region is common to all plants. All plants in the region will have a certain set of environmental technologies installed, and they will freely substitute between LS Sub-bituminous and HS Bituminous coal within this configuration.

- The total capacity of NOx removal is common to the two coal operating sectors. All plant configurations that have NOx removal in both coal generation groups—LS Sub-bituminous and HS Bituminous—draw from a common NOx removal capacity node. That means that coal types can be freely substituted for each unit, and whether it has NOx removal or not will be properly represented.

- The total capacity of Hg removal is common to the two coal operating sectors. All plant configurations that have Hg removal in both coal generation groups—LS Sub-bituminous and HS Bituminous—draw from a common Hg removal capacity node. That means that coal types can be freely substituted for each unit, and whether it has Hg removal or not will be properly represented.

- The total capacity of CO2 removal is common to the two coal operating sectors. All plant configurations that have CO2 removal in both coal generation groups—LS Sub-bituminous and HS Bituminous—draw from a common CO2 removal capacity node. That means that coal types can be freely substituted for each unit, and whether it has CO2 removal or not will be properly represented.

- The total capacity of SO2 removal is common to the two coal operating sectors. All plant configurations that have SO2 removal in both coal generation groups—LS Sub-bituminous and HS Bituminous—draw from a common scrubber capacity node. That means that coal types can be freely substituted for each unit, and whether it has SO2 removal or not will be properly represented.

- The coal supply nodes become specific to each of the two coal operation sectors. HS Bituminous coal is provided only to the plant configurations that burn HS Bituminous coal. The situation is analogous for LS Sub-bituminous coal.

The potential for accurate modeling of a thermal power system and its emissions consequences and possible remediation through equipment modification and fuel switching becomes immediately obvious, and IEMM needs all those dimensions if it is going to properly represent developed or developing country electrical systems, much less North America and Europe in which there remains a tremendous quantity of coal and coal generation.

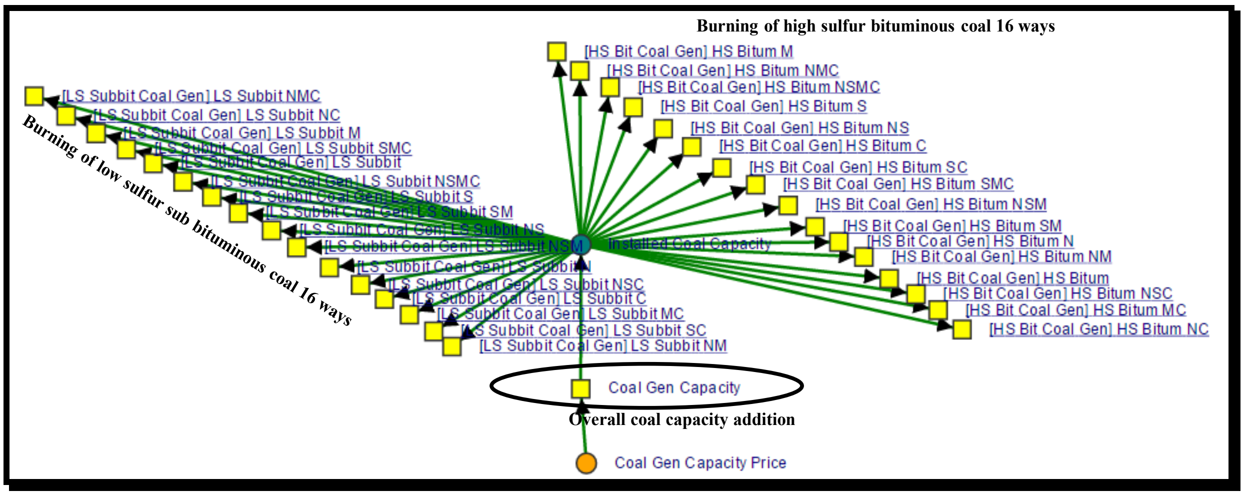
3.1.7 Capacity in Place and Capacity Expansion (for Coal)

We reiterate the five capacity addition possibilities, and IEMM can and must consider independent but interdependent capacity additions of all five types, which are generally characterized as follows:

1. Total Generation Capacity
2. SO2 removal, e.g., wet limestone scrubber
3. NOx removal, e.g., selective catalytic reduction or SCR
4. Mercury Removal, e.g., activated carbon injection
5. CO2 removal, e.g., unspecified CO2 removal

The remainder of this section considers coal generation with four possible retrofits and two coal types. Extension of this to more coal types if IEMM needs to do so is straightforward; it merely expands the preceding models laterally. Four retrofit techniques and two coal types are illustrative of the directness and the sophistication of the economics and environmental implications of thermal generation. It is also a template for correct modeling that can immediately be "copy and pasted" to consider more coal types within IEMM. Notice that the foregoing capacity nodes provide existing capacity and/or new capacity to each of the five capacity addition nodes. Let us enumerate them briefly in network format. The structure in Figure 67 shows the aggregate magnitude of coal capacity available in the region, notwithstanding whether it is used to burn coal type 1 on the left (low-sulfur sub-bituminous) or coal type 2 on the right (high-sulfur bituminous). This structure, which appears at first blush to be difficult or sophisticated or overly detailed, is in fact extremely simple. It is merely enumerative. It ensures that the coal generation capacity in place for the region is not exceeded no matter what type of coal is burned or what type of retrofit installations it is burned within in the region. Capacity expansion logic for aggregate coal capacity in the region occurs at the yellow node second from the bottom in Figure 67. The yellow node is where a number such as the $2,750,000/MW capacity cost of installation of new pulverized coal plants would reside. Alas, this is not the only capacity that might be needed. There might be SO2 removal, NOx removal, mercury removal, and even CO2 removal that would be economic to add or that would be forced by regulators under some sort of NSPS, ESPS, or CPP regulation. However, this structure fully reflects and allows endogenous capacity addition.

Figure 67: Total Coal Generation Capacity—Spans All Coal Nodes in the Model

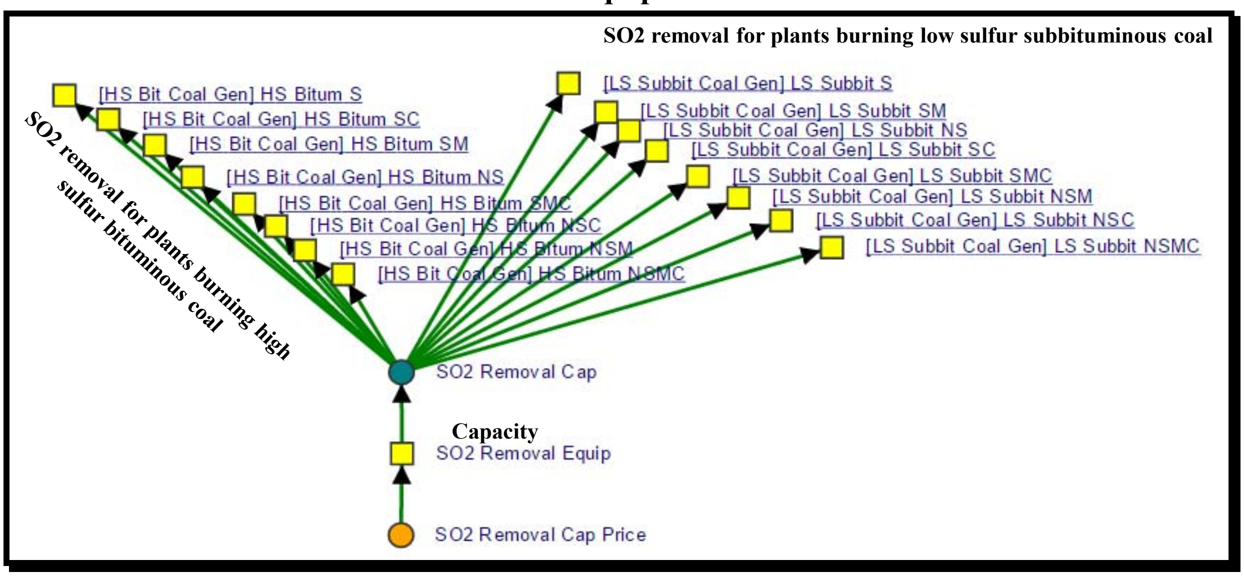


One of the most important results of this structure is the following. What is the "price" of coal capacity? What is the fair market value of coal capacity in Figure 67? The answer is disarmingly simple, and so, so, so valuable to electricity modeling. The price on the output of the Installed Coal Capacity hub is the **price of capacity**. That capacity can be used in a number of different configurations with a number of different environmental inputs, but there still is one single value of capacity. However, the price that is computed as the supply-demand balance between capacity supplied and all the various uses of capacity in Figure 67 is the oft sought after "**capacity price**." The capacity price is the price for example that PJM seeks to establish at auction in PJM. It is price a customer would have to pay for the right to use 1 MW of coal-fired generation capacity. It is the price of the "keys to the plant." Electricity people say it is the price to "buy the button" to the plant. Assuredly if new capacity is being built, the price of capacity will become equal to long-run marginal capital cost (amortized in a very special, very market sensitive way. It is not acceptable to levelize or amortize it. This IEMM design needs to be able to calculate it on a fully dynamic basis. We discuss how to do that below.) People who are in or near the electricity business covet this capacity price. It is related to the minimum amount of money they should hold out for to secure a power purchase agreement or PPA. This capacity price is immensely valuable, and it is a byproduct of our proposed IEMM design.

Using the structure in Figure 67, IEMM is able to represent retirements or idlements by simply ramping back the installed capacity in Figure 67 as old coal plants exit the fleet. Retirements and idlements are some of the more difficult phenomena to model, but the place they are modeled is in the capacity of gas combined-cycle capacity available in the given region in Figure 67. This IEMM design will nail them.

The structure in Figure 68 shows the aggregate magnitude of SO2 removal capacity that has been installed in the region, notwithstanding whether it is used to burn low-sulfur sub-bituminous (coal type 1) or high-sulfur bituminous (coal type 2). This structure ensures that the SO2 removal capacity in place for the region is not exceeded no matter what type of coal is burned. It also gets the input-output coefficient magnitudes correct for plants that have installed SO2 removal. The yellow generation configurations into which the SO2 removal capacity points flow are only those generation configurations that have SO2 removal installed and have SO2 removal input-output coefficients within those nodes. Notice that there are fewer coal plant configurations that are pointed to by the SO2 removal capacity node than the total number of generation configurations in place. That is because it only points to configurations that have SO2 removal installed, not to configurations that do not.

Figure 68: SO2 Removal Capacity Price—Spans All Coal Nodes in the Model to Which SO2 Removal Equipment Is Added



The structure in Figure 69 shows the aggregate magnitude of NOx removal capacity that has been installed in the region, notwithstanding whether it is used to burn low-sulfur sub-bituminous (coal type 1) or high-sulfur bituminous (coal type 2). This structure ensures that the NOx removal capacity in place for the region is not exceeded no matter what type of coal is burned. It also gets the input-output coefficient math correct for plants that have installed NOx removals. Notice that there are a limited number of coal plant configurations that are pointed to by the NOx removal capacity node. It points to all configurations that have NOx removal installed, not to nodes that do not.

The structure in Figure 70 shows the aggregate magnitude of mercury removal capacity that has been installed in the region, notwithstanding whether it is used to burn low-sulfur sub-bituminous (coal type 1) or high-sulfur bituminous (coal type 2). This structure ensures that the mercury removal capacity in place for the region is not exceeded no matter what type of coal is burned. It also gets the input-output-coefficient math correct for plants that have installed mercury removals. Notice that there it only points to configurations that have mercury removal installed, not to nodes that do not.

Figure 69: NOx Removal Coal Capacity Price—Spans All Coal Nodes in the Model to Which NOx Removal Equipment Is Added

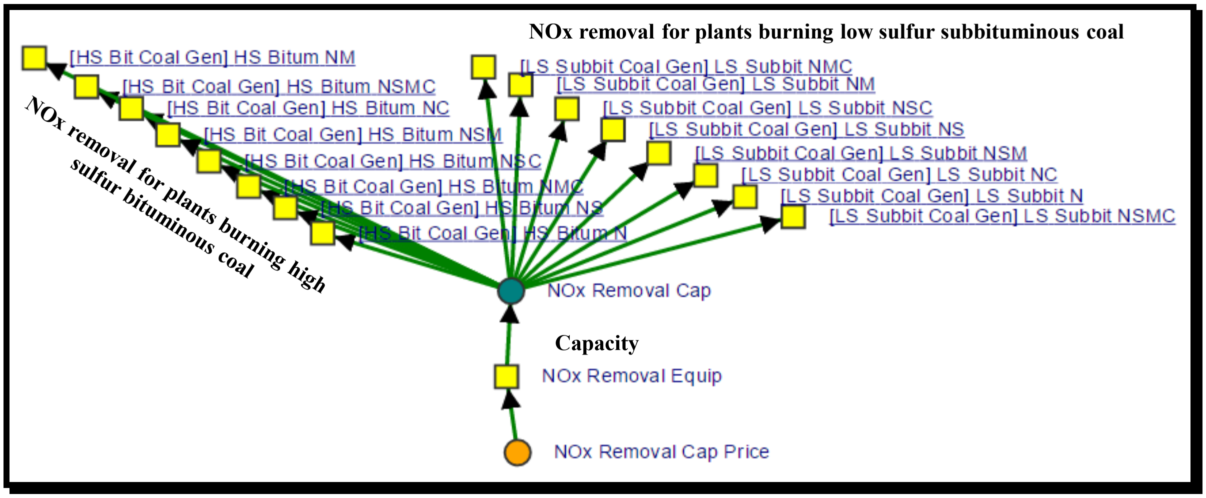
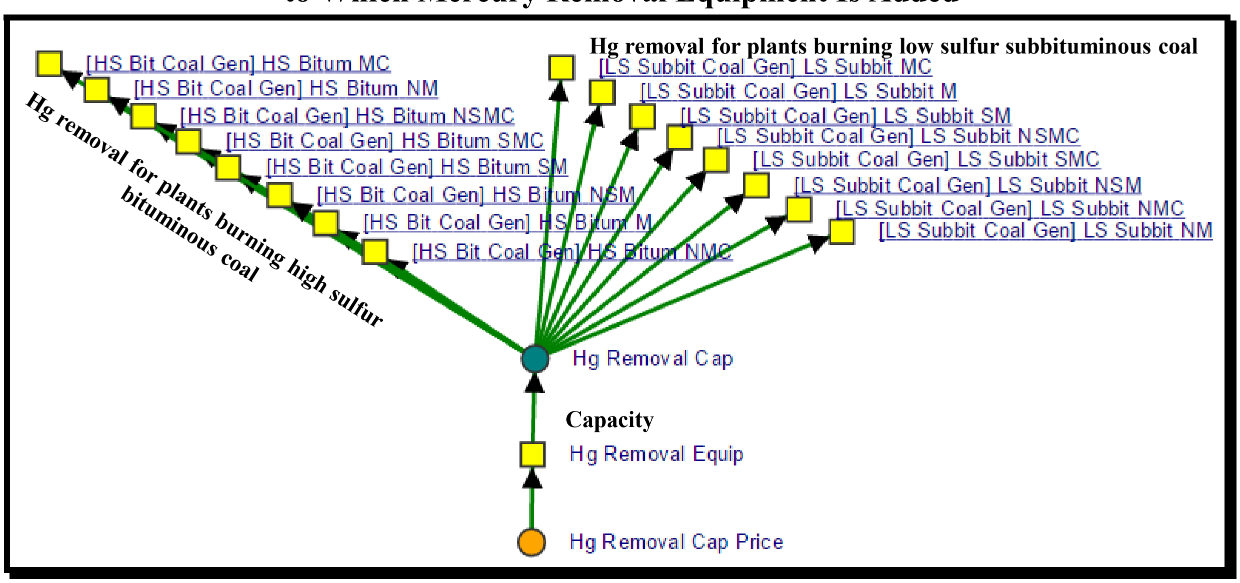
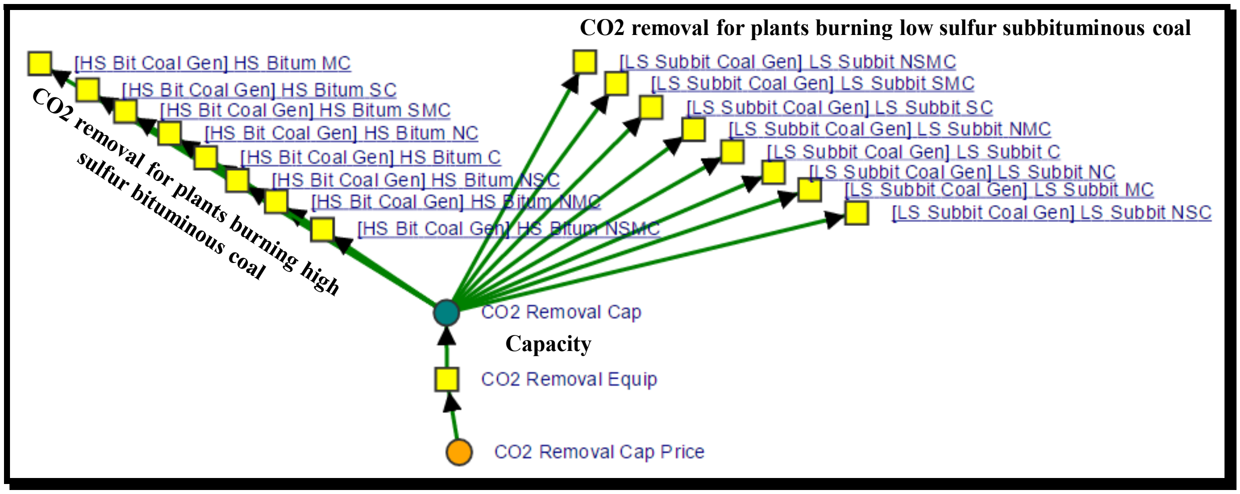


Figure 70: Mercury Removal Coal Capacity Price—Spans All Coal Nodes in the Model to Which Mercury Removal Equipment Is Added



The structure in Figure 71 shows the aggregate magnitude of CO2 removal capacity that has been installed in the region, notwithstanding whether it is used to burn low-sulfur sub-bituminous (coal type 1) or high-sulfur bituminous (coal type 2). This structure ensures that the CO2 removal capacity in place for the region is not exceeded no matter what type of coal is burned. It also gets the input-output-coefficient math correct for plants that have installed CO2 removals. Notice that there are fewer coal plant configurations that are pointed to by the CO2 removal capacity node. It only points to configurations that have CO2 removal installed, not to nodes that do not.

Figure 71: CO2 Sequestration Coal Capacity Price—Spans All Coal Nodes in the Model



We reiterate that five nodes govern total installed coal capacity, installed SO2 removal capacity, installed NOx removal capacity, installed mercury removal capacity, and installed CO2 removal capacity. These five nodes allow the model to expand or retire coal capacity and to keep proper track of what type of retrofit capacity has been installed. This simple structure allows complete representation of capacity addition and retirement, retrofit capacity addition and retirement, and through that mechanism pretty much any range of capacity regulation related to the environment such as NSPS/ESPS or BACT/MACT or CPP. Under this design, IEMM will become a profoundly important and accurate model of thermal generation, taking account not only the capacity that was added but also any retrofits or cleanup equipment that might have been or will be installed.

Why would one go to this type of detail? The answer is simple and very relevant to all electrical systems in the world that IEMM must characterize. If regulation were command and control, certain types of capacity (e.g., NOx removal) would be mandated. If regulation were market-based, there would be endogenously determined partial plant retrofits over time to achieve the mandated market-based regulation. Both are crucial. Also, if one is to represent a system like China with its mammoth coal fleet, a lot of it unretrofitted, one must carefully consider issues of plant retrofit (e.g., CO2 removal) versus fuel switching (because there are so many very high-cost, very low-quality coals in China that could easily be substituted) versus capacity additions using a new fuel altogether such as gas or wind. Failure to do so is failure to represent the Chinese electrical system. We happen to know through direct work in China that they are considering importing Australian coal, which is much cleaner and cheaper than some of their own indigenous coal. This IEMM design will allow a more sophisticated, correct, and accurate representation of the Chinese electrical system than perhaps the Chinese themselves have. As the Chinese are pressured to reduce or modify coal generation or substitute renewables or gas, EIA will be able to anticipate and represent the impact on their electrical system and predict how it is likely to evolve.

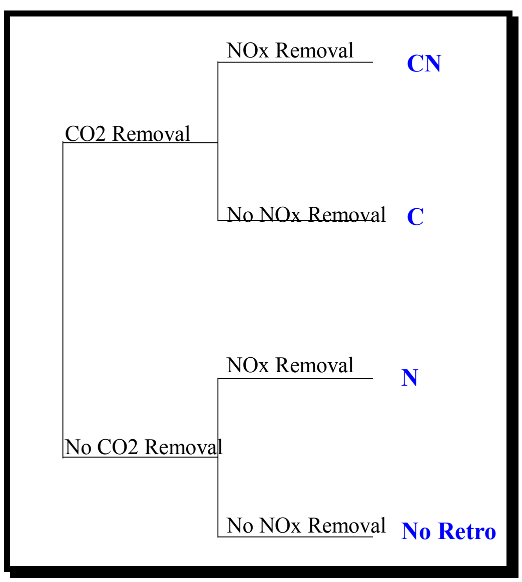
3.2 Gas Fired Generation (Which Is Directly Extrapolatable to Oil)

The discussion of coal generation in the preceding section was complex because coal chemistry, combustion, and pollution are complex. There are more pollutants that come out of coal-fired power generation and many coal types that have significant impacts on plant ratings and on certain emissions. By contrast, with gas, there are fewer pollutants. There is no possibility for fuel substitution. Gas is gas. The primary complexity is that there are three prime movers for gas generation that all have to be considered individually–simple cycle combustion turbine, combined-cycle, and conventional steam turbine. We are going to be illustrating using only one of these three gas generation types—combined-cycle, steam turbine, and combustion turbine. However, the discussion easily extrapolates to the other two.

3.2.1 Plant Retrofitting for Gas CC plants

Natural gas CC plants release no SO2 and no mercury. Neither exists in the fuel. Thus, the retrofit combinations for gas CC plants are substantially simpler than for coal as indicated in Figure 72. All one can do is install NOx removal to mitigate NOx, and one can install sequestration to arrest or mitigate CO2. Both are generally approximately 90 percent efficient. Notions of CO2 removal work just as well (in fact better) for gas-fired plants as for coal fired plants. The reason is that at the lower heat rates of gas CC plants, a much lower volume of flue gas with a much lower volume of CO2 must be addressed. Low flue gas volume can be a big advantage for sequestration. Why would one sequester a coal plant when one could just as easily sequester a gas CC plant or two and free up a lot more capacity for generation? IEMM will tell.

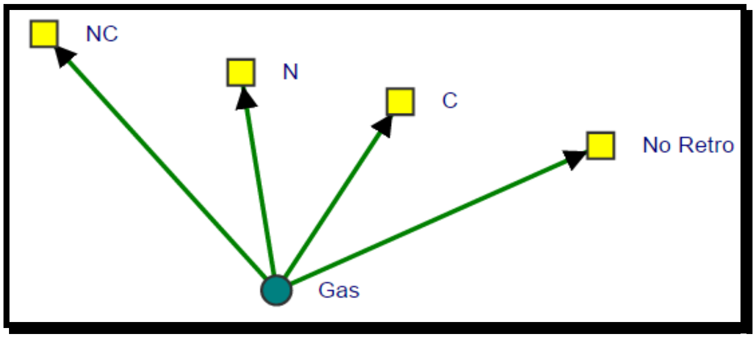
Figure 72: Combinations of Plant Retrofits



3.2.2 Interplay between Retrofit Options and Gas Use

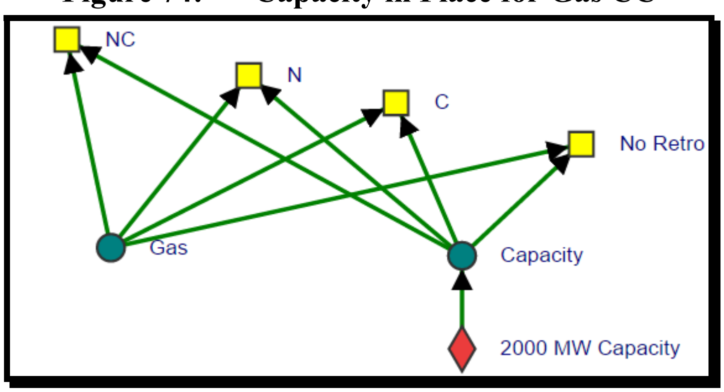
This section characterizes, using the preceding retrofits, how the model is built, and how it represents all the interacting configurations enumerated in Figure 72. We use the IEMM network representation to explain how IEMM will represent those options. We draw the diagram of the exhaustive set of retrofit and non-retrofit options using a network format as shown in Figure 73. Every node in Figure 73 represents one full path through the tree in Figure 72. In fact, Figure 73 is nothing more than the enumeration of equipment combinations in Figure 72.

Figure 73: The Four Retrofit-Non Retrofit Options for Gas CC Plants



If there existed 2,000 MW of existing gas CC capacity embedded in a completely uncontrolled, non-retrofit plant in the region, each of the preceding four plant configurations (represented as nodes in the diagram) must compete to use this 2,000 MW of installed generation capacity, i.e., for this plant site. To represent the retrofit and utilization of this capacity, we build the representation in Figure 74, augmenting the representation from Figure 73 to make sure that retrofit additions do not increase the aggregate capacity of the generation capacity in place inappropriately, i.e., never exceed the installed 2,000 MW. To wit, this IEMM design will ensure that the aggregate generation capacity before and after retrofitting will be 2,000 MW, no more and no less. The model will operate the retrofit and non-retrofit capacity so that the sum total of retrofit plus non-retrofit capacity will not exceed 2,000 MW.

Figure 74: Capacity in Place for Gas CC



The inputs to each of the yellow retrofit generation nodes and the yellow non-retrofit nodes are extremely important. We begin with the No Retro node at the upper right to illustrate. The inputs to that generation node are indicated conceptually in the diagram in Figure 75, precisely and identically as they were in Figure 32. (We have included the two environmental inputs and the REC input, which are not specifically drawn in Figure 74.) We have drawn in Figure 75 this "Non-Retrofit/Non-Sequestration" node complete with all the input markets to which it is hooked and the output markets to which it is hooked. We have drawn it using the ArrowHead network drawing interface in Figure 76. (We will be using our network drawing interface for our graphics for simplicity and clarity later in this section.) We see the direct analogy, in fact, the direct parallelism, between Figure 75 and Figure 76. The labels in Figure 76 are slightly altered, but that is the sole difference.

Figure 75: The Non-Retrofit Electricity Generation Node

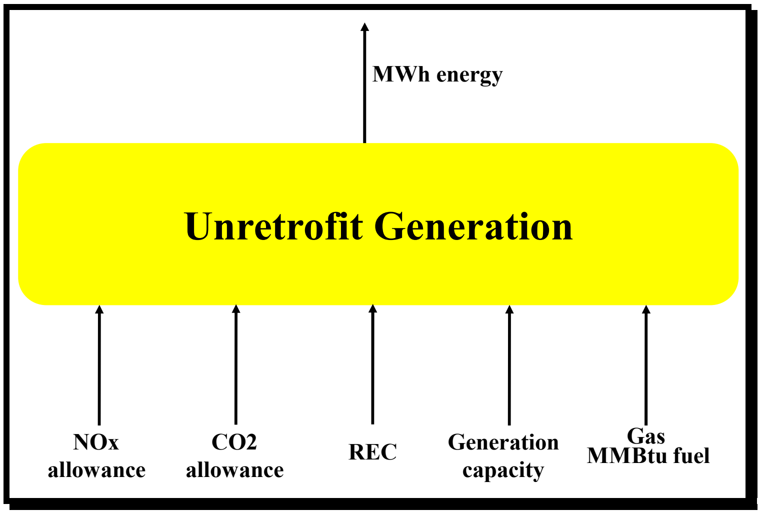
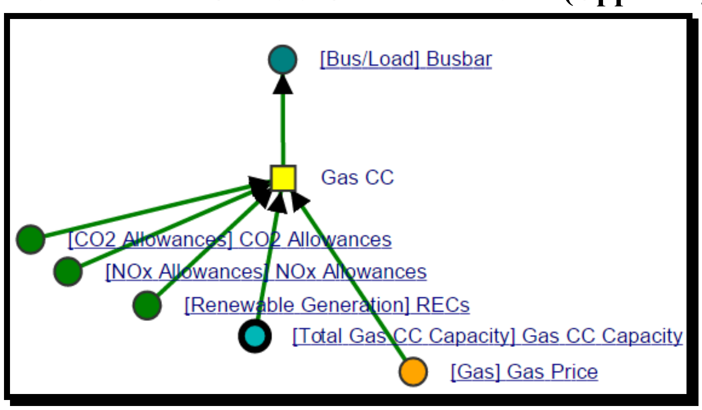


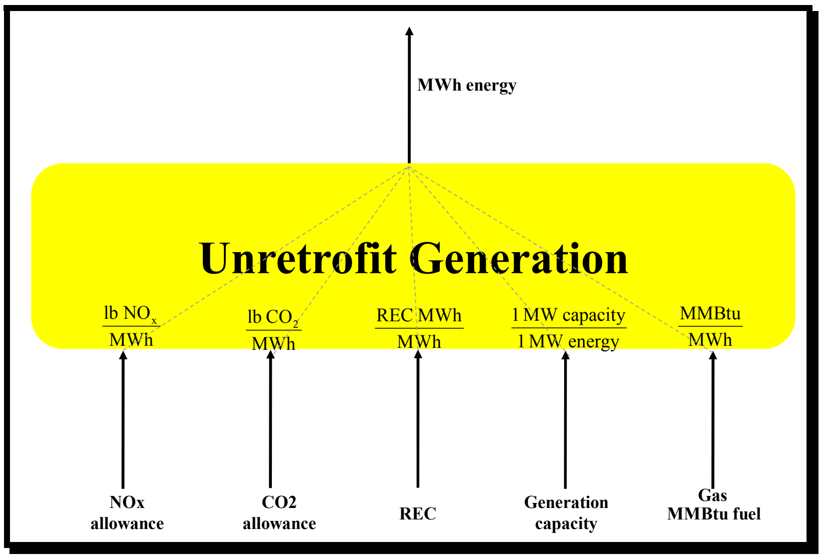
Figure 76: The Non-Retrofit Gas CC Generation Node (Upper Right in Figure 75)



We see in Figure 75 and Figure 76 the two pollutants entering the generator, reflecting the fact that this uncontrolled, non-retrofit generator will have to surrender sufficient NOx and CO2 allowances to generate at whatever MWh level it chooses to generate. Indeed, the NOx and CO2 input-output coefficients quantify how much NOx and CO2 will be produced with each MWh of electrical energy that is produced and therefore how many NOx and CO2 allowances the generator will have to have (via purchase on the open market or through allocations by a regulator) to generate. If those allowances are not priced, we have a completely unregulated system. We also see in the figure requirement for a REC coming into this node to enable its generation. It is not sufficient for the generator to merely acquire two environmental allowances. It must also acquire the REC or green energy allowance. We see gas coming into this node, meaning that its input-output coefficients and non-derated capacity are specific to gas. There is the capacity link coming into this node, meaning that the aggregate generation by this node cannot exceed the aforementioned 2,000 MW. Indeed, the capacity input will carry the signal that there is 2,000 MW of capacity available to be used for power generation in this node, and no more.

For each of the five inputs to the non-retrofit electricity generation node in Figure 76, we relate the quantity of input to the quantity of output (MWh). People term such relationships "input-output coefficients," and we adopt that terminology. Input-output relationships were discussed in extensive detail for coal previously, but now these concepts are similarly put to work for gas. The input-output coefficients are specific to gas. The input-output coefficients required to quantify the non-retrofit electricity generation node in Figure 75 appear in Figure 77.

Figure 77: The Non-Retrofit Gas CC Generation Node



We will dispense with calculating numerical input-output coefficients as we did for coal, but that preceding discussion about coal is directly relevant.

3.2.3 Retrofit Nodes—Data and Structure

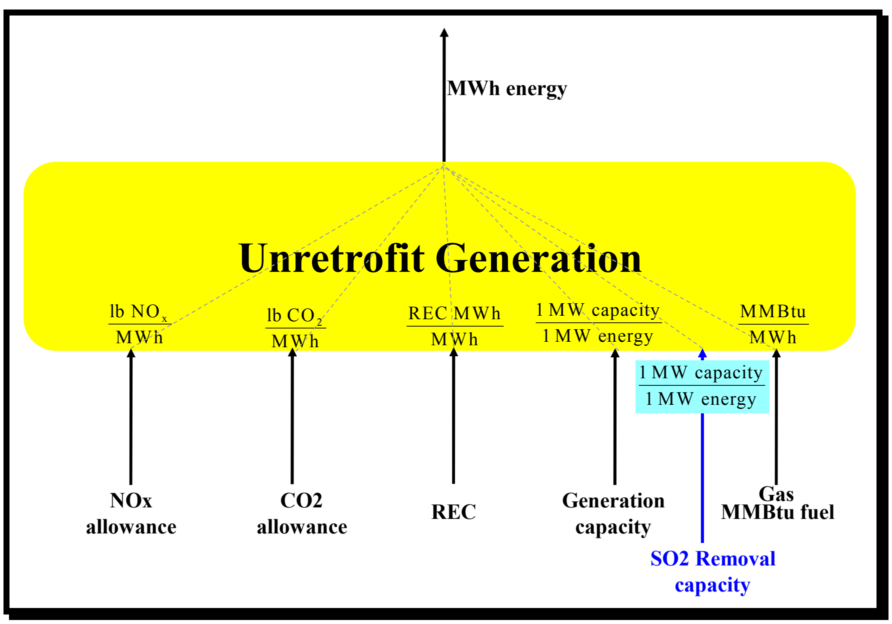
Retrofit nodes are exactly analogous to non-retrofit nodes as analyzed previously with one exception—they need a second type of installed capacity, namely NOx removal and CO2 removal. You need a second kind of capacity, namely the capacity enumerated in the tree in Figure 73, for each type of plant retrofit you might use for the uncontrolled plant. And, alas, such retrofits do not occur for free. Retrofits cause adjustments in performance in the plant, and those adjustments come at a capital cost. We will consider a combination of NOx removal and sequestration for CO2 removal as indicated in Figure 73.

It is illustrative to walk through some specific numbers for the non-retrofit plant configuration in Figure 77, which were calculated by applying the various NOx removal and sequestration technologies. An installed NOx removal removes approximately 9/10 of the NOx. These technologies indeed modify the input-output coefficients (emissions coefficients) for the various plant types from their values in the top sections of Figure 18 to the values in the bottom of Figure 18.

Plant retrofits, however, cost money. There is a $/MW charge to install NOx removal and a $/MW charge to install sequestration. There is also an increase in operating cost when we install extra equipment on a plant, and we have calculated the operation cost in the non-retrofit case in the top of Figure 18 and for the retrofit case in the bottom of Figure 18.

Referring to the structure in Figure 77, there is one more type of capacity you would need to be able to generate with NOx removal. You would have to have NOx removal capacity in place. You do not get to operate using the input-output parameters of generator that has an installed NOx removal until and unless you have NOx removal capacity. You would have to have an input from aggregate capacity in the region as shown in Figure 77, but you would have to have an input from installed NOx removal capacity in the region. The configuration of a unit with NOx removal installed would, therefore, take the augmented form in Figure 78 and its equivalent in Figure 79. Notice everything is identical to the uncontrolled case in Figure 78 except that there is an additional link to account for installation of NOx removal. This node cannot run until and unless there is some capacity incoming to the node and unless there is some NOx removal capacity incoming to the node. We are going to see that we will be able to endogenous add both types of capacity, so NOx removal addition and NOx amelioration by retrofitting become endogenous.

Figure 78: The Material Balance for the NOx removal Retrofit Node—Gas CC plant



What about a retrofit that has both NOx removal and CO2 removal retrofits installed? See Figure 80 for the representation in the ArrowHead graphical system.

Figure 79: Network and Markets for the NOx removal Retrofit Node—Gas CC plant

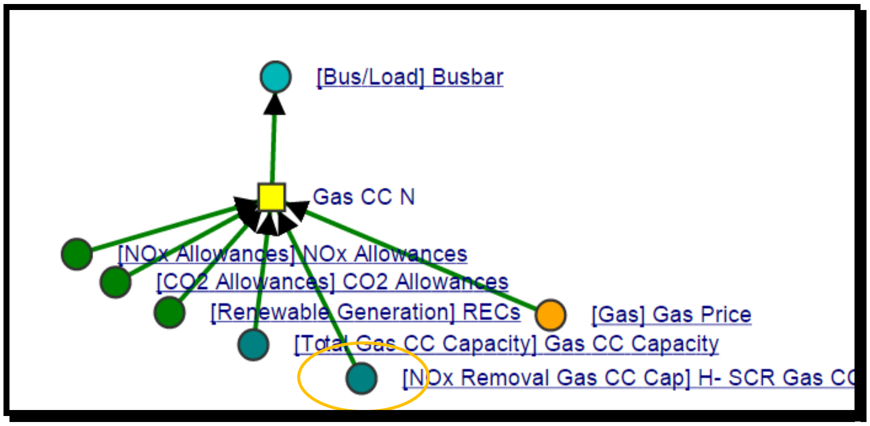
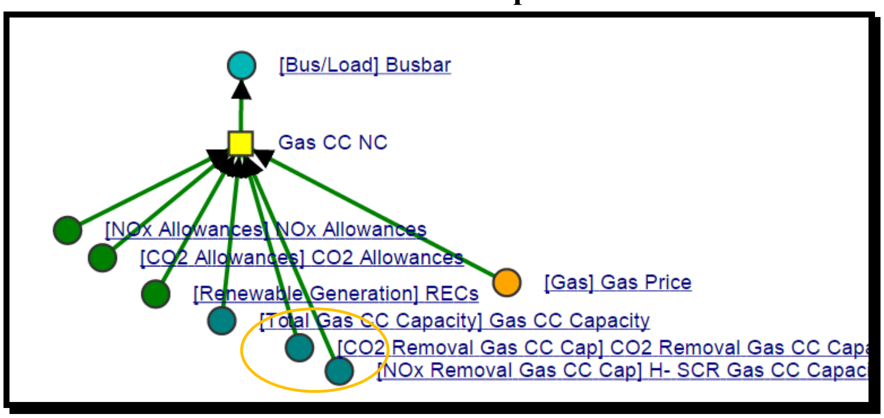


Figure 80: Network and Markets for the NOx removal and Sequestration Retrofit Node—Gas CC plant



By considering the capacity that a plant needs to deploy depending on the installed equipment, the simple generation node configuration we have here can represent the full range of environmental emissions, retrofit or remediation options, and capacity expansion. It endogenizes the retrofit and environmental arrest decision. That ensures that the cost becomes embedded in the market price of electricity.

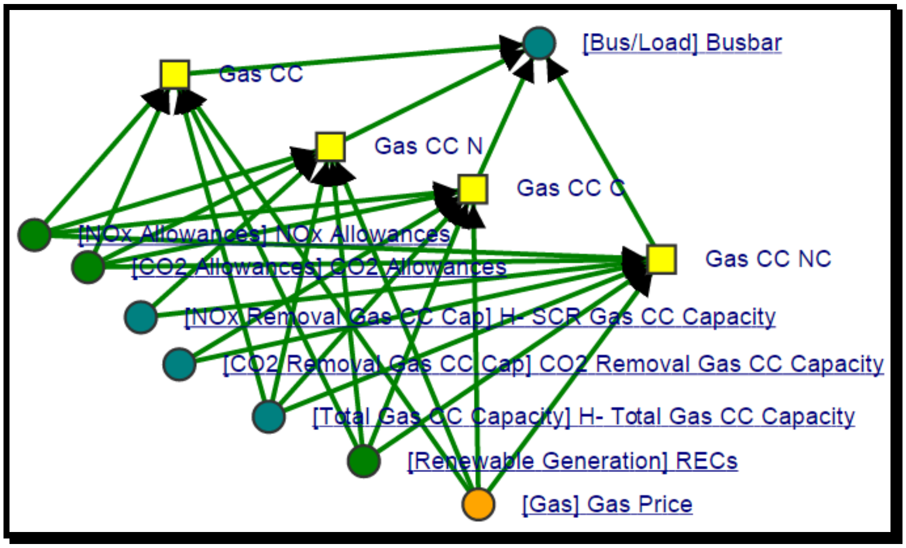
3.2.4 Summary of Concepts for Gas Thus Far

Figure 81 is the configuration of the operation sector and the capacity addition sector. Notice the following:

- There are four operating configurations, one for each possible combination of retrofit/environmental removal configurations. If additional retrofit technologies were to be considered, they would have to be added to this collection of sixteen configurations.
- For each of the four operating configurations, there must be an input for the aggregate capacity in the region. That ensures that the sum of all four operating configurations does not exceed the 2,000 MW of gas combined-cycle capacity that exists in the region. This aggregate capacity

input link ensures that the installed capacity in the region is not violated and that in fact installed capacity in the region can be reduced over time.

Figure 81: Generation and Retrofit Alternatives for Gas



- For each of the four operating configurations, the environmental emissions coefficient must be input for that specific plant configuration
  - NOx, expressed as pounds of NOx emitted per MWh generated.
  - CO2, expressed as pounds of CO2 emitted per MWh generated.

If there are any synergies or "cobenefits" of various combinations of retrofit technologies, they will be incorporated in the slate of NOx and CO2 emission coefficients. It is the slate of emission coefficients that matters.

- If an operating configuration has:
  - N, designating NOx removal (e.g., SCR), that configuration must accept as input a link from the NOx removal capacity node. If a plant has NOx removal on it, it has to have a capacity link from the NOx removal capacity hub. Its capacity input-output coefficient for that capacity link will be 1. That is, if a plant has NOx removal on it, it has to buy NOx removal capacity from the NOx removal node, and it has to buy capacity from the total capacity node, both with input-output coefficients of 1.
  - C, designating CO2 removal (e.g., sequestration), that configuration must accept as input a link from the CO2 capacity node. If a plant has CO2 removal on it, it has to have a capacity link from the CO2 capacity hub. Its capacity input-output coefficient for that capacity link will be 1. That is, if a plant has CO2 removal on it, it has to buy CO2 removal capacity from the scrubber node, and it has to buy capacity from the total

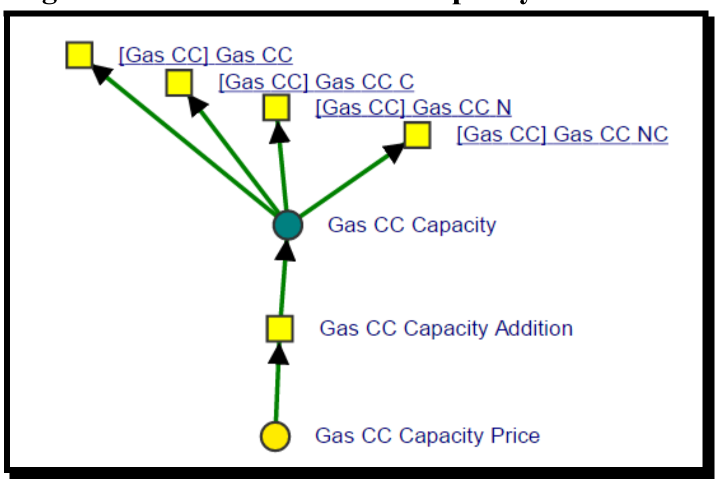
capacity node, both with input-output coefficients of 1. It is these capacity links that "enable" that corresponding piece of equipment to operate. That is, the scrubber link "signals" that particular operating facility with the scrubber configuration installed that it can operate.

All three types of gas generation are represented in this fashion, and in parallel—simple cycle combustion turbine, combined-cycle, and gas steam turbine. The preceding example has considered only combined-cycle, but three parallel gas systems must be implemented. Extrapolation to the other two types of gas generation is straightforward. We will see that parallelism in the final model configuration defined below.

3.2.5 Capacity in Place and Capacity Expansion (for Gas CC)

The situation for total installed capacity for any one of the three gas technologies appears in Figure 82. Notice that there all four plant configurations are included.

Figure 82: Total Installed Capacity for Gas CC

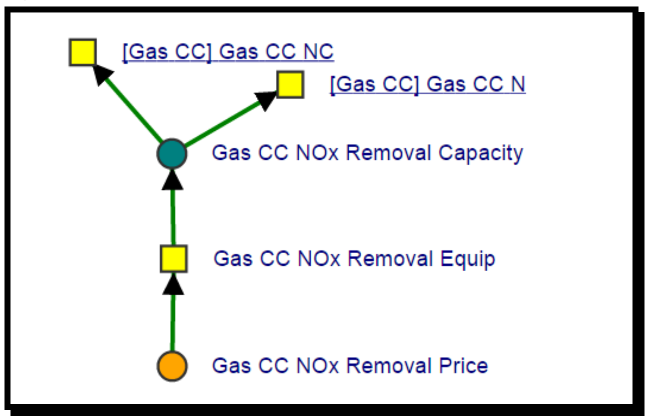


Endogenous capacity additions for gas combined-cycle appears in the yellow node second from the bottom in Figure 82. That is the node in which we have a number such as $1,000,000/MW for the addition of new gas combined-cycle generation and where we allow gas combined-cycle capacity to expand endogenously.

To continue the notion of the price of capacity, one of the profound results of this IEMM structure is the following, just as true for gas combined-cycle as it was above for coal. What is the "price" of gas combined-cycle capacity? What is the fair market value of gas combined-cycle capacity in Figure 82? The answer is disarmingly simple, and so valuable to electricity modeling. The price on the output of the Gas CC Capacity Addition node is the price of capacity. That capacity can be used in a number of different configurations with a number of different capacity factors and a number of different environmental outputs. However, the price that is computed as the supply-demand balance between capacity supplied and all the various uses of capacity in Figure 82 is the ubiquitously sought after "capacity price." Capacity price is the price for example that PJM seeks to establish at auction in PJM. It is price a customer would have to pay for the right to use 1 MW of gas combined-cycle generation capacity. It is the price of the “keys to the plant.” Electricity people say it is the price to “buy the button” to the plant. Assuredly if new capacity is being built, the price of capacity will become equal to long-run marginal capital cost (amortized in a very special, very market sensitive way. It is not acceptable to levelize or amortize it. This IEMM design will calculate it on a fully dynamic basis.) People who are in or near the electricity business very much want to compute and know this capacity price. It is related to the minimum amount of money they should hold out expect to secure a power purchase agreement or PPA. Capacity value is immensely valuable information, a byproduct of the proposed IEMM design.

The situation for NOx removal capacity for any one of the three gas technologies appears in Figure 83. Notice that there are only two plant configurations that have NOx removal. Endogenous capacity additions for NOx removal attached to gas combined-cycle appears in the yellow node second from the bottom in Figure 83. That is the node in which we insert $325,000/MW for NOx removal retrofit of an existing plant.

Figure 83: NOx Removal Retrofit for Gas CC

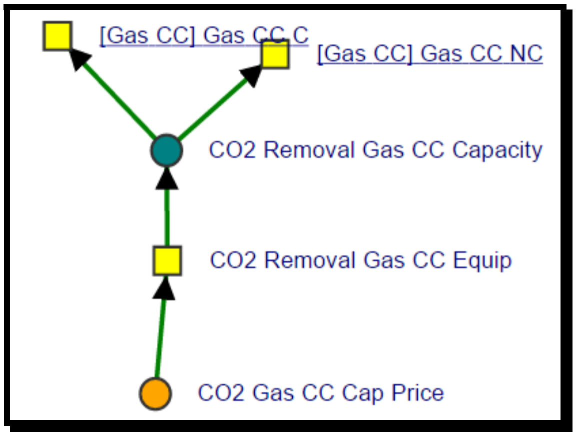


The situation for CO2 removal capacity for any one of the three gas technologies appears in Figure 84. Notice that there are only two plant configurations that have CO2 retrofit equipment. Endogenous capacity additions for CO2 removal appears in the second node from the bottom in Figure 84, and that is the node in which we insert $1,500,000/MW for CO2 removal in a gas combined-cycle power plant.

Collectively in Figure 82, Figure 83, and Figure 84, we represent the endogenous addition of gas combined-cycle capacity without any controls and after that the addition of the complete, exhaustive, combinatorial set of capacity and environmental controls that can be implemented in a new or existing gas combined-cycle plant.

We are also able to represent gas CC retirements or idlements by simply ramping back the installed capacity in Figure 82 as old plants exit the fleet. Retirements and idlements are some of the more difficult phenomena to model, but the place they are modeled is in the capacity of gas combined-cycle capacity available in the given region in Figure 82.

Figure 84: CO2 Removal for Gas CC



3.2.6 Oil Generation

There are oil based generators in many countries in the world, particularly in those locales that lack gas local distribution companies (LDCs). This section puts forth in summary form some basics for oil. There are two types of oil we have to consider—distillate quality liquid fuels in a combustion turbine and resid quality liquid fuels in a steam turbine. These types of plants have become obsolete in developed electricity systems because of the availability of gas and become of the large disparity between gas and oil price. However, in some economies in which remoteness and lack of transmission are an issue, IEMM will have to consider oil based generation.

3.2.6.1 Distillate Combustion Turbines

We can dispense with distillate combustion turbines quickly. The situation looks precisely like gas combustion turbines. There is no sulfur (to speak of) in the incoming distillate and no mercury in the incoming distillate, and therefore there is no SO2 or Hg in the retrofit or operational tree. Thus, distillate combustion turbine generation retrofits appear as in Figure 85. The logic and method for incorporating distillate based combustion turbines into the generation mix is exactly analogous to the discussion of gas combined cycle and combustion turbines. IEMM will clearly, under this design, be able to represent distillate based power generation. (If needed, this discussion pertains just as well to oil-based combined cycle, not merely combustion cycle, generation equipment.

3.2.6.2 Residual Oil Based Steam Turbine Generation

Residual oil is a bit of a different situation. Residual oil generates NOx and CO2, very much like coal or gas. That is a byproduct of hydrocarbon fuel, no matter how clean the fuel. Residual oil can be very dirty and very sulfur-laden in many venues around the world. Local refineries in many venues around the world are not high technology refineries, and they do not desulfurize. They merely eject the sulfur in the incoming crude into their residual oil stream. (That has happened historically in venues such as Mexico, which eject "combustolio" with sulfur content as high as 6 percent and give it to boilers to burn.) For residual oil, IEMM will need a retrofit combinatorial tree with three levels. Figure 86 illustrates.

Figure 85: Combinations of Distillate Combustion Turbine Retrofits

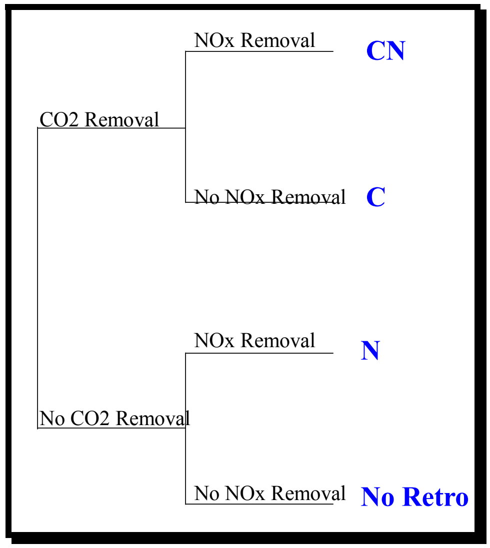
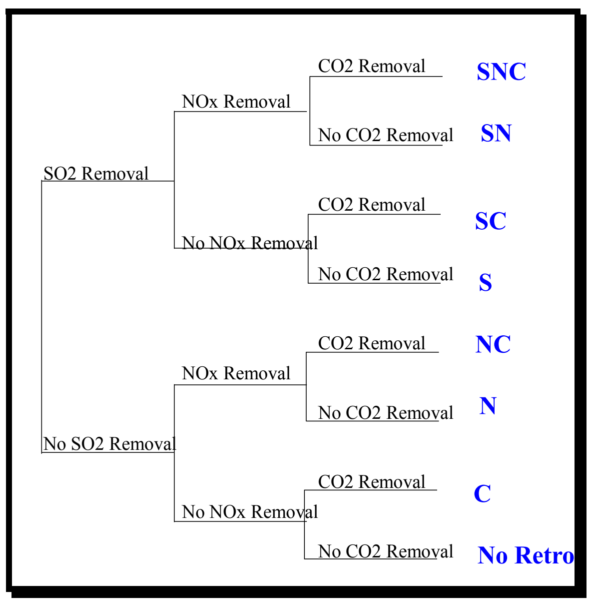


Figure 86: Combinations of Plant Equipment Installations for Residual Steam Turbine Generation

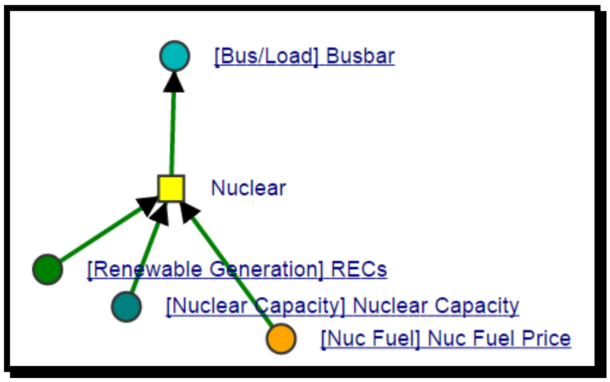


Once we have this residual oil steam generation tree in Figure 86, we can easily see how to proceed. We develop eight plant configurations for redid steam turbine generation representing the eight combinations in Figure 86 and proceed precisely as we did with coal (which had sixteen) and gas (which had four). The logic and ultimate structure of the oil steam turbine generation sector is obvious. We are not aware that international models have given sufficient attention to oil fired generation. It is quite prevalent in various venues around the world because of their remoteness and the lack of natural gas infrastructure. It cannot be ignored in many venues, even though it probably can in North America and Europe in light of the widespread availability of gas and coal.

3.3 Nuclear Generation

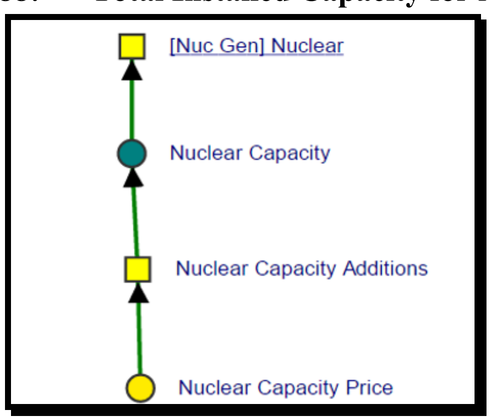
There are no emissions from nuclear, so there are no configurations with and without a retrofit related to nuclear. This nuclear generation situation appears as in Figure 87. Notice that there is only one plant configuration about which IEMM need worry. There are no emissions to worry about because nuclear generation does not emit SO2, NOx, mercury, or CO2. Nuclear requires a REC because it is not a qualifying renewable in any regulatory jurisdiction we have seen. Nuclear generation requires nuclear fuel.

Figure 87: Nuclear Generation Node



There is only one type of nuclear generation capacity, namely "regular," plain vanilla nuclear capacity. Figure 88 illustrates. If we wanted to consider breeder reactors, fusion reactor, or other nuclear configuration, we could. Those would fall into the technological change category.

Figure 88: Total Installed Capacity for Nuclear



Endogenous capacity additions for nuclear appear in the yellow node second from the bottom in Figure 88. That is the node in which we have a number such as $5,000,00 /MW for the addition of new nuclear generation and where we allow nuclear capacity to expand endogenously.

To continue again the notion of the price of capacity, one of the most beneficial results of this structure is the following, just as true for nuclear. What is the "price" of nuclear capacity? What is the fair market value of nuclear capacity in Figure 88? The price of nuclear capacity is the price that a customer would have to pay for the right to use 1 MW of nuclear generation capacity. It is the price of the “keys to the plant.”

How would we represent plant decommissioning? Simple. We simply decrement the magnitude of capacity in the Nuclear Capacity Additions node. Such decrement reduces the installed nuclear capacity in place. Capacity reduction would, in turn, cut nuclear generation in the system and force the addition of next best capacity. That is precisely what EIA will want to do in considering Fukushima decommissioning scenarios that might occur around the world as well as in Japan itself. Interestingly, as one cuts the amount of nuclear capacity in place, the price of nuclear capacity for the remaining plants might well grow. That is a result we might expect from IEMM.

3.4 Total Emissions Allowances, Coal, Gas, Oil, and Renewables

If we consider the supply of emissions allowances and the collection of plants eligible in the market to consume them, EIA gets an accurate look at both the pricing and the aggregate accounting for pollution in IEMM. Figure 89, a screen shot from ArrowHead’s electricity model, indicates where all the CO2 allowances would have to be surrendered. If we begin at the left, we see that they will have to be surrendered in gas steam turbines, gas combustion turbines, gas combined-cycle, coal plants burning high-sulfur bituminous coal, and coal plants burning bituminous coal. They will **not** have to be consumed in nuclear or renewables plants. No CO2 is produced from such plants. Figure 89 illustrates the supply of all the CO2 allowances and the various types of plants that compete to consume these CO2 allowances. Figure 89 is the incarnation of the diagram in Figure 44 for CO2 allowances. The accumulation of all the allowances from all the plant configurations into the aggregate supply node is the environmental accounting we have been addressing, and it occurs regardless of the price or supply of those allowances.

Figure 89: CO2 Emissions Allowances Produced and Consumed

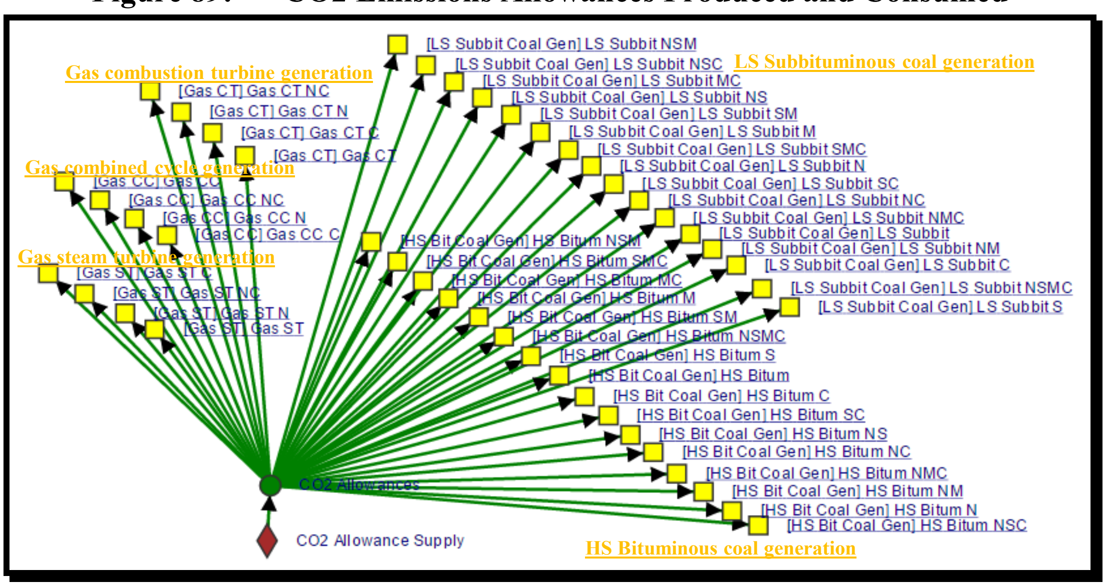


Figure 90 indicates where all the mercury allowances would have to be surrendered. If we begin at the left, we see that they will have to be surrendered in coal plants burning high-sulfur bituminous coal and coal plants burning bituminous coal. They will **not** have to be consumed in gas, nuclear, or renewables plants. No mercury is emitted from such plants. Figure 90 illustrates the supply of all the mercury allowances and the various types of plants that compete to consume these mercury allowances. Figure 90 is the incarnation of the diagram in Figure 44 for mercury allowances. The accumulation of all the allowances from all the plant configurations into the aggregate supply node is the environmental accounting we have been addressing, and it occurs regardless of the price or supply of those allowances.

Figure 90: Mercury Emissions Allowances Produced and Consumed

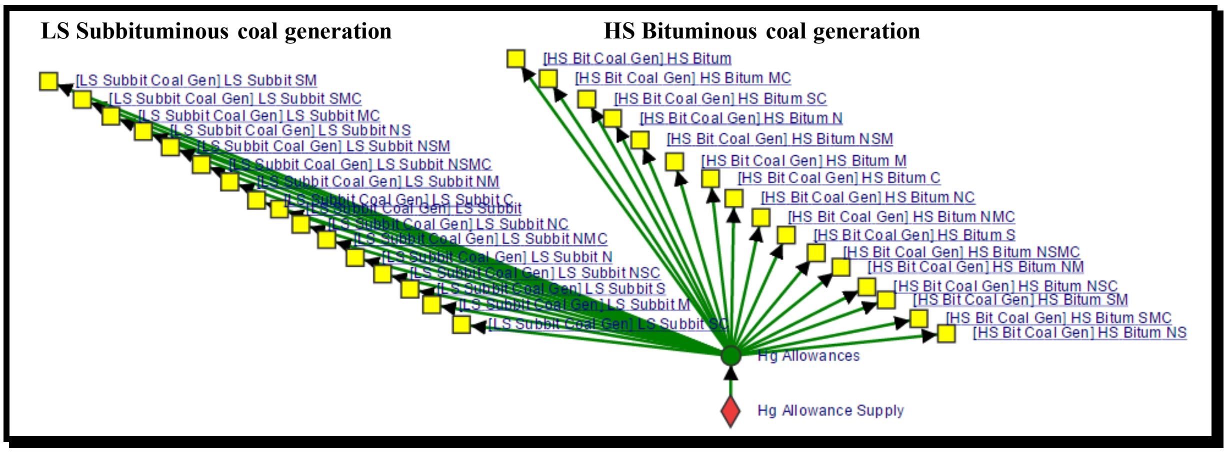


Figure 91 indicates where all the NOx allowances would have to be surrendered. If we begin at the left, we see that they will have to be surrendered in coal plants burning low-sulfur sub-bituminous coal, coal plants burning high-sulfur bituminous coal, gas steam turbines, gas combustion turbines, and gas combined-cycle plants. They will **not** have to be consumed in nuclear or renewables plants. No NOx is produced from such plants. Figure 91 illustrates the supply of all the NOx allowances and the various types of plants that compete to consume these NOx allowances. Figure 91 is the incarnation of the diagram in Figure 44 for NOx allowances. The accumulation of all the allowances from all the plant configurations into the aggregate supply node is the environmental accounting we have been addressing, and it occurs regardless of the price or supply of those allowances.

Finally, Figure 92 indicates where all the SO2 allowances would have to be surrendered. If we begin at the left, we see that they will have to be surrendered in coal plants burning low-sulfur sub-bituminous coal and, coal plants burning high-sulfur bituminous coal. They will **not** have to be consumed in gas, nuclear, or renewables plants. No SO2 is produced from such plants. Figure 92 illustrates the supply of all the SO2 allowances and the various types of plants that compete to consume these SO2 allowances. Figure 92 is the incarnation of the diagram in Figure 44 for SO2 allowances. The accumulation of all the allowances from all the plant configurations into the aggregate supply node is the environmental accounting we have been addressing, and it occurs regardless of the price or supply of those allowances.

Figure 91: NOx Emissions Allowances Produced and Consumed

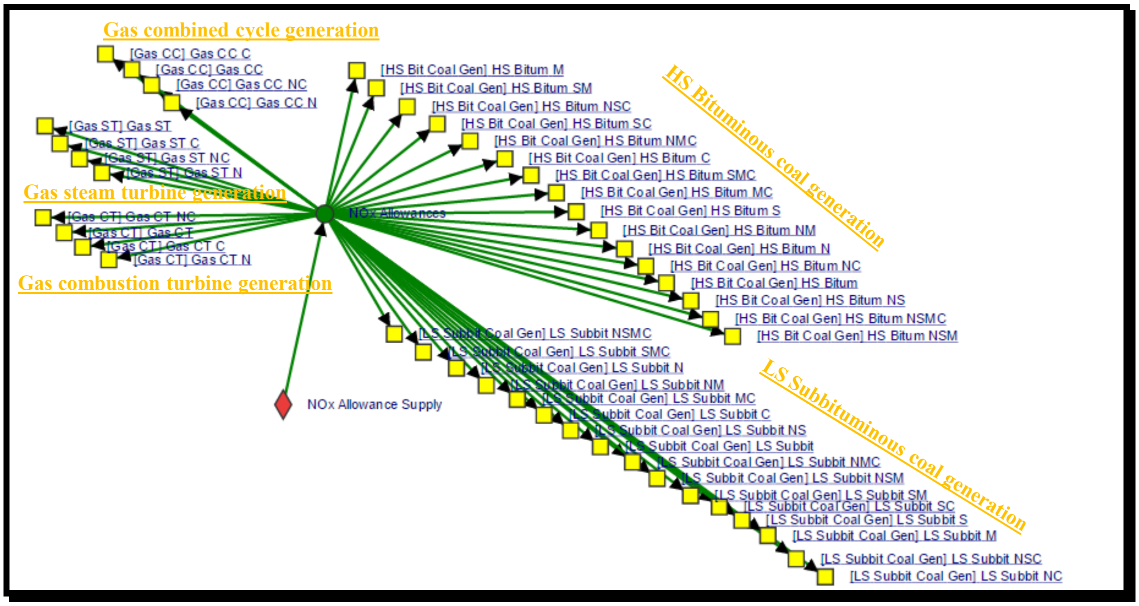
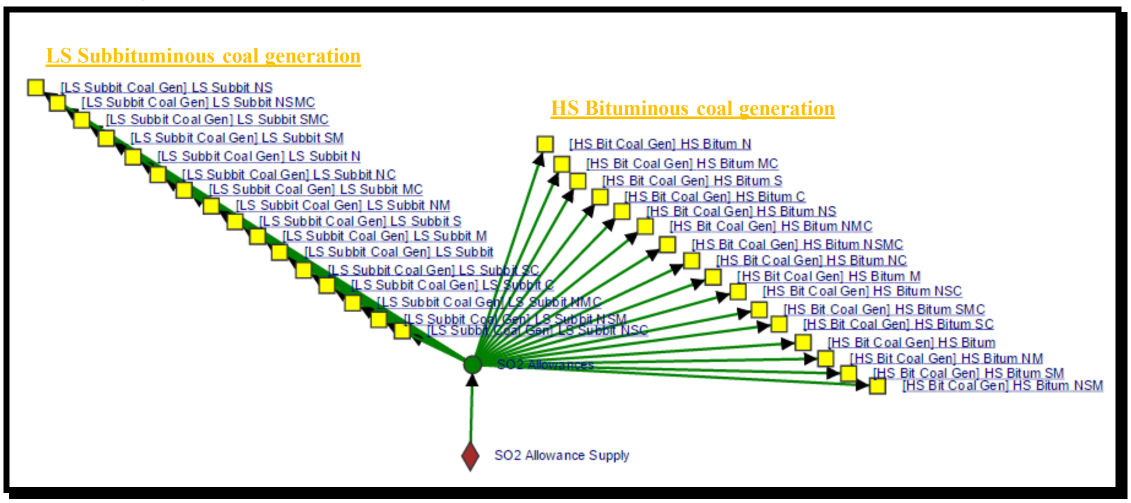


Figure 92: SO2 Emissions Allowances Produced and Consumed

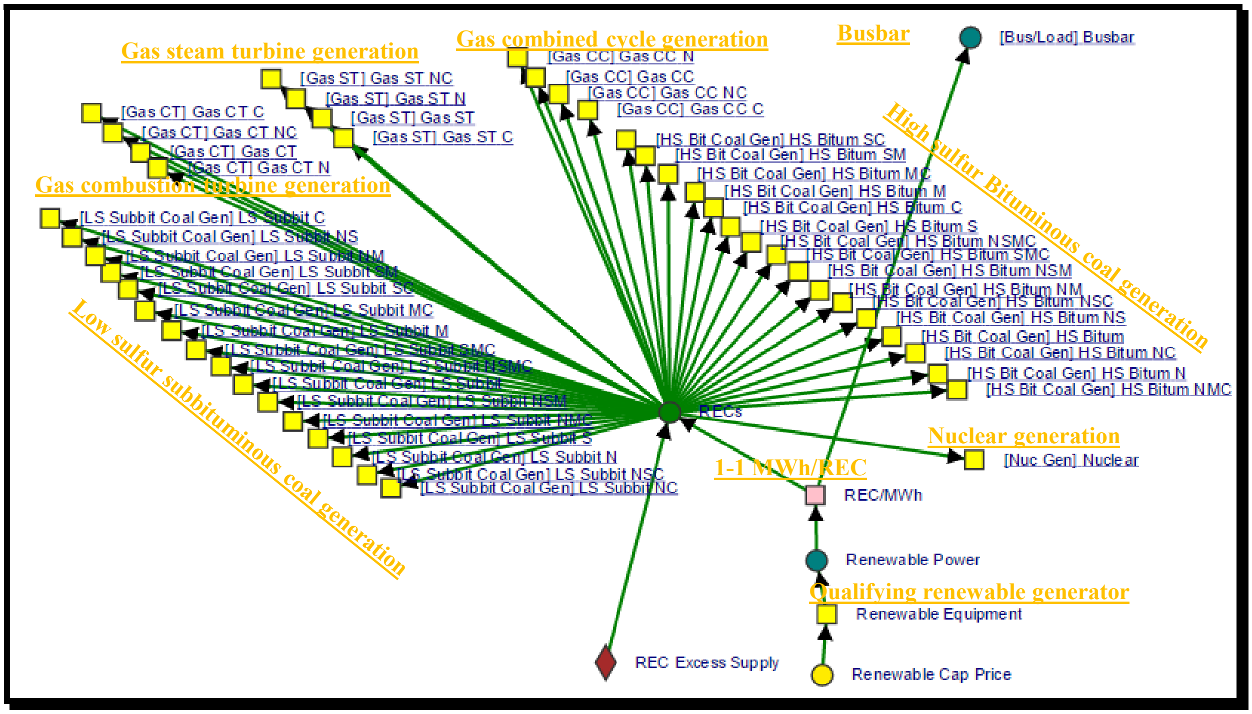


3.5 RECs and Where They Come From

Figure 49 showed conceptually the origin and disposition of RECs. Notice at the lower right in the pink node, we see that the entry of renewables creates both MWh flowing into the market (the busbar) and the exact same number of RECs flowing into a tradable REC market. Figure 93 is the incarnation of the diagram in Figure 49, containing the production and disposition of RECs.

In Figure 93, the disposition of RECs is obvious. Every one of the non-qualifying plants in the region must surrender a certain number of RECs per MWh it generates. We see the RECs going to coal plants burning high-sulfur bituminous coal, low-sulfur sub-bituminous coal, gas combined-cycle, gas simple cycle, gas steam turbine, and nuclear. All the RECs generated are consumed in total by all these generators. The non-qualifying generators (all the yellow generators in the diagram) are required to surrender a certain number of RECs for every MWh they generate to the busbar. Those non-qualifying generators, therefore, consume RECs at a given rate. These non-qualifying generators represent our demand curve for RECs in an analogous way that the fleet of plants represents a demand curve for environmental emissions allowances.

Figure 93: RECs and Renewable MWh Produced and Consumed



On the REC supply side, we see two sources of RECs in the diagram. In the yellow figure at the lower right titled Renewable Equipment (such as a windmill), we see the operation of a qualifying renewable, in this example a wind asset. That yellow figure delivers 1 MWh to the Renewable Power hub. That hub, in turn, delivers 1 MWh to the REC/MWh Splitter. The function of the splitter is to issue 1 MWh of wind power to the busbar (i.e., to the grid) at the upper right and simultaneously to issue 1 MWh of REC to the tradable REC market to its left. The pink node delivers the MWh to the market and delivers the MWh of REC to the REC market. We also see in the diagram the supply of "shortfall" RECs in the brown REC Excess Supply diamond at the bottom. This represents the fact that quite often there is a "grace period" or an impossibility of RECs to enter the market at such a rate that they achieve the full RPS level. This is represented by a supply of free RECs to the market from the brown diamond at some ostensibly decreasing level that allows the non-qualifying generators to generate at a level above the level commensurate with REC MWh being delivered to the market. In this way, IEMM can model lagged compliance with an RPS. Virtually all REC/RPS systems around the world lag regulatory requirement in terms of full and complete compliance, and this allows us a way to represent alternative full compliance schedules.

3.6 Concluding Comments

This section shows the scope of the plant operation (dispatch) and variable cost calculation. It is extremely sophisticated and takes full account of all capacity in place.

There is no element of thermal plant generation omitted in this section:

- Capacity in place, completely endogenous with regard to capacity addition, operation, and retirement/decommissioning.
- Capacity addition and retirement
- Fuel substitution.
- Equipment retrofits (SO_2, NO_x, mercury, and CO_2 and any other pollutant EIA may need to consider), completely endogenously
- RPS and RECs
- Emissions entitlements and accounting (SO_2, NO_x, mercury, and CO_2), whether or not emissions are priced

We will talk about load in detail in the next section. Putting this type of detail and sophistication into a global social welfare maximization approach defies logic and experience. It has never been done, and that is unlikely to be an accident or oversight. It is not going to happen. The size alone would doom it, not to mention the algorithm or the poorness of formulation. The sheer size of the Jacobean that is embedded in complementarity solution algorithms would doom it.

People might quip: "Oh, but this proposed design is complicated." Electricity, the quintessential just-in-time commodity is complicated. One cannot simply "wish complexity away" and appeal to a simplistic dispatch model, historically based on global social welfare maximization or its cousin cost minimization. The idea of a simplified or "reduced form" or aggregated generation model for electricity is absolutely ludicrous on its face. IEMM cannot risk that. To miss the dimensions developed thus far is to render the IEMM essentially valueless. This IEMM design will be central to successful IEMM modeling around the world. When we import load, capacity expansion, and other dimensions onto this model, EIA will see the benefit of this design. The design will only become more apparent as we consider load, hydro, and all the other requisite dimensions later in this report, and the ability to represent disparate power systems around the world will become more apparent.

4 LOAD AND RESERVE MARGIN[15]

Electricity demand (which people have termed "load") varies by hour of day, day of week, week of year, and year. Because electricity has not been storable (let us not debate that assertion quite yet; there are certainly storage and load-shifting alternatives, but they are modest in the grand scheme of electricity production and consumption[16], which is substantially real-time), electricity is the ultimate in "just-in-time" (JIT) manufacturing product. With today's electric power system technology, one cannot store power in bulk quantities for future use. That is not necessarily going to be true in the future, however. There is a substantial amount of work ongoing to advance and perfect electricity storage. At present, effectively 100 percent of all electric demand must be met by available, operational generating capacity—either native assets that exist within proximity (via transmission and distribution) to the load or assets accessible through transmission systems from distant generation points. The rapid temporal variability and magnitude of load are critically important elements of an electric power system and an electric power market. The magnitude and variability of load have to be represented carefully and accurately in IEMM if one is to represent and quantify energy and capacity markets, prices, and volumes in electricity markets. They need to be represented carefully within IEMM as described in this section. The lack of viable storage is more chronic in lesser developed country electric systems than in Europe or the United States.

In addition to net energy for load, reserve margins over and above demand requirements must be maintained to provide both longer-term and short-term demand response and system balancing. The reason is simple. In the just-in-time manufacturing environment of electric power, one must have just-in-time resources ready at a moment's notice to "click on" in the event one of the scheduled generators should "click off" because of an outage or some other reason. In a just-in-time environment, one must have production capacity held in reserve to meet the actual time-varying load as it occurs in the market. It is not just the problem that load varies, and, in fact, varies through a very wide range. Such variation is structural, i.e., deterministic. In addition, the ability to supply load is probabilistic because of plant or other asset outages. That problem is generally solved in an electric power system by assuming a required magnitude of reserves that can be implemented within an electric system and can meet a given reliability criterion. Quantifying and representing those reserves is an important element of electricity system market and operation modeling.

This section describes the proposed IEMM methodologies for:

- Defining forward projected demand curves (called "load" curves), including levels of regional and temporal granularity.

- Operating an energy generation system to meet those forward, time-varying demand or load schedules.

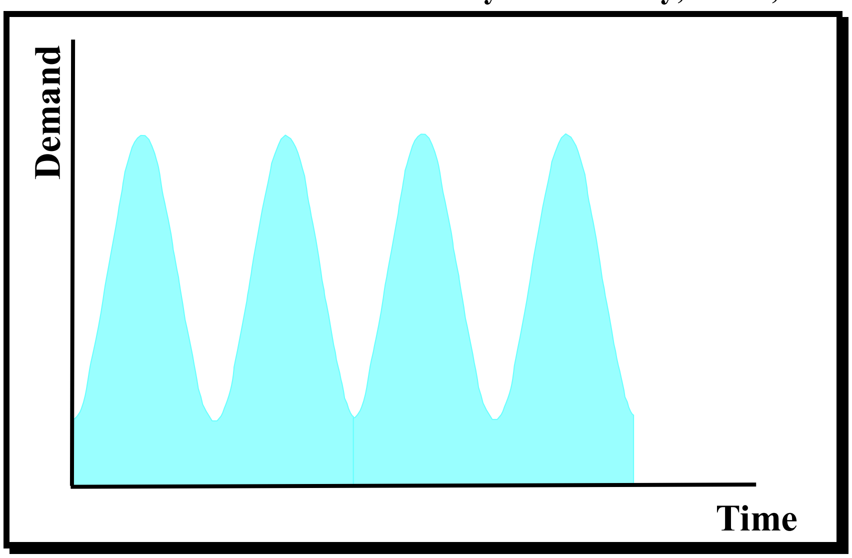
[15] The concepts and methods of characterizing load both chronologically and with nonchronological load duration curves are documented. This section is designed as a summary.
[16] We will be giving a good deal of attention to electricity storage and how to model it later in this CDR.

- Estimating and assuming reserve margins, including ancillary services and transmission-line-loss issues that impact nodal and system reliability, and after that calculating their individual and combined effect on electric system prices and quantities and capacity additions. We will be discussing what these ancillary services are below.

4.1 Model and Methodology

Projecting electric energy demand is similar to predicting demand for the beam of light from a flashlight. When it is on, it is on. When it is off, it is off. There is no storage of the light beam. Electricity is the same; it is real-time. We need to project the hour by hour, day by day, week by week, and year by year demand for the beam of light, i.e., the aggregate stream of electric energy required by the customers in a given region. Figure 94 summarizes for a given day how hourly demand for electric energy must be projected. Such projection must occur subregion by subregion, location by location, throughout the IEMM regions that comprise the North American, European, or another electricity market.

Figure 94: Demand Variation Is Assembled by Time of Day, Week, Season, and Year



How do we determine what the hourly load variation is, and how can we project it ten or more years into the future? To do so for the United States, one can access the hourly demand reports by every utility and other reporting entity in the country, which have been reported to the FERC in the form of FERC 714 load reports. When data such as the FERC 714 load data exists, data that displays hour by hour load (consumption) over a several year historical period, it is straightforward to download and process it into load or demand shapes as characterized here. (The FERC 714 reports are comprehensive, generating histories of demand by hour for every reporting entity in the United States over the past several reporting years. Because generation equals load, they are thought to characterize the precise time pattern of load, i.e., of power consumption.) By so doing, we are able in effect to develop a historical load histogram for every hour of the year. This highly detailed hourly information can either be used directly in its raw form or aggregated to build comprehensive models of demand at different levels of detail over different time horizons for our model.

The raw, reported FERC 714 data is intrinsically a **historical** demand schedule, which varies by time of day, week, season, and year. (Figure 94 is a conceptual illustration of the time-varying historical data that can be assembled for every nodal subregion in North America and for many other regions of the world.) Indeed, there is a time-varying schedule of demand of the form in Figure 94 stretching over the past several years at every point in North America and elsewhere. Once we have downloaded the historical hourly load data for each individual region, the next step is to aggregate the raw information hour-by-hour, region-by-region to create specific regional aggregates of hour-by-hour electric energy consumption. They have the form in Figure 94, but they pertain to larger regional aggregates. To wit, we have taken the curves of the form in Figure 94 and aggregated them hour by hour according to the geographic subdivisions in the model. After such aggregation, we have an overall curve of the form in Figure 94 for each of the subregions that compose our model.

Once IEMM has assembled this hour-by-hour load pattern over the historical period for each of the subregions of our model, we after that organize the load data according to the temporality we wish to use. We could if we wanted maintain the hour-by-hour chronological form of the data (and we will return to this hourly chronological theme below), but we generally use it to calculate daily, weekly, seasonal, and/or annual representations of load and load duration.

4.1.1 A Note on Model Time Points (Time Discrimination)

The inherent structure of IEMM must be highly temporal and, in fact, chronological. The model can use more strategic, longer-term forms such as non-chronological temporal forms (e.g., load duration curves). In fact, the model can use hybrids of chronological and non-chronological forms such as chronological by month but non-chronological within month. There is great power in this, and we describe it herein. Notwithstanding, because electricity begins as a non-storable "just-in-time" or JIT commodity with some degree of moderation caused by storage and load management, temporal detail cannot be avoided. We have to take special care to model real-time delivery and load-shifting through (imperfect) electricity storage or load-shifting. This is the reason why a reduced form or simply electricity model would be a travesty.

IEMM should be based on what we call "time points," which are instants in time at which IEMM calculates a full general equilibrium market solution, and the solution is fully interconnected and simultaneous over whatever time points are inserted into the model. The model time points need to be separated in time by any interval. For example, the model should allow three 1 year intervals followed by three 2 year intervals followed by three 3 year intervals followed by three 5 year intervals. It also must allow a year to be represented by 8760 time intervals, each 1 hour in length, perhaps aggregated but never missing any hours. Conversely, it should allow any other equally spaced or unequally spaced time intervals. The time points are interpreted as instants in time at which samples are taken from a continuous forward price curve, a continuous forward generation and power consumption curve, and a continuous forward investment schedule.

IEMM allows any number of forward time points and any number of forward “sub-time” points (load tranches for each time point) to be represented within the model’s forecast horizon. Figure 95 illustrates how IEMM uses this temporal structure to represent any time horizon and any number of **non-chronological** time points one might desire. Figure 96 illustrates how IEMM uses this temporal structure to represent any time horizon and any number of **chronological** sub-time points one might desire. At each of the model time points, the model assumes a set of “sub-time” points. Sub-time points are defined to be a set of times needed to fully represent each time point. If the time points are months, then the sub-time points must represent days in the month for example.

Figure 95: Time Period Structure of IEMM Allows Non-chronological Time Points

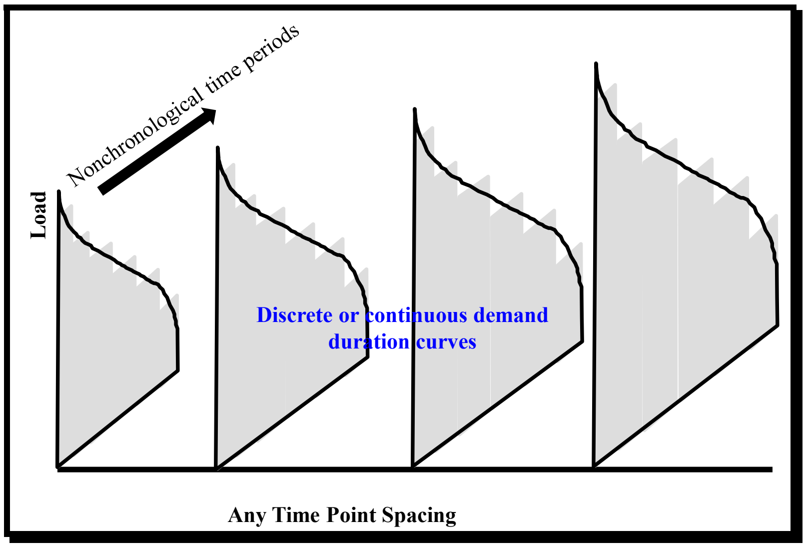
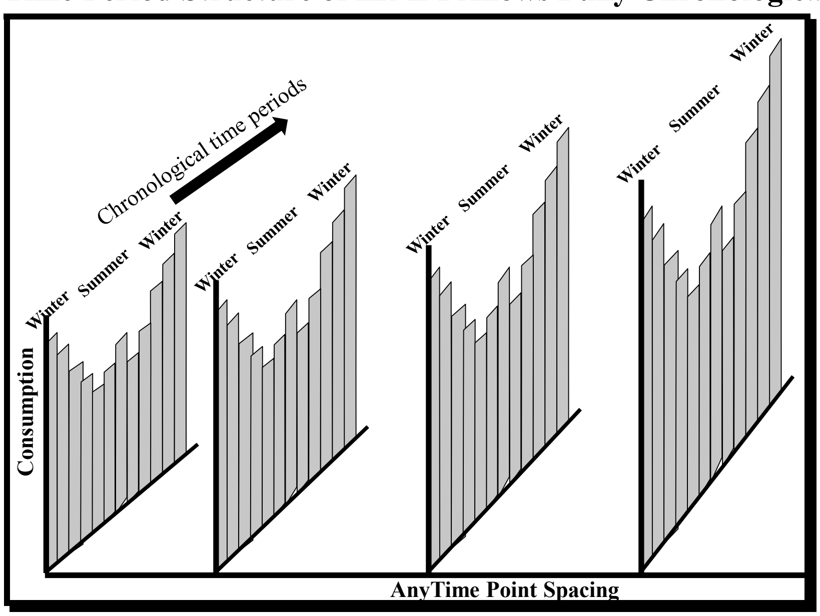


Figure 96: Time Period Structure of IEMM Allows Fully Chronological Time Points



If one reorders the sub-time points from highest to lowest load, he or she has systematically overridden the chronological assumption inherent in the data and created a non-chronological model. If one organizes the sub-time points chronologically and synchronously across regions and nodes, he or she has created a fully chronological market model. This IEMM design allows the complete range of time points, and one selects different combinations of times and sub-times for specific applications.

This time period structure in Figure 95 or Figure 96 allows a complete, detailed representation of the price temporal detail for every plant in the model.

4.1.2 Retaining the Hour by Hour Chronology (Coincidence)

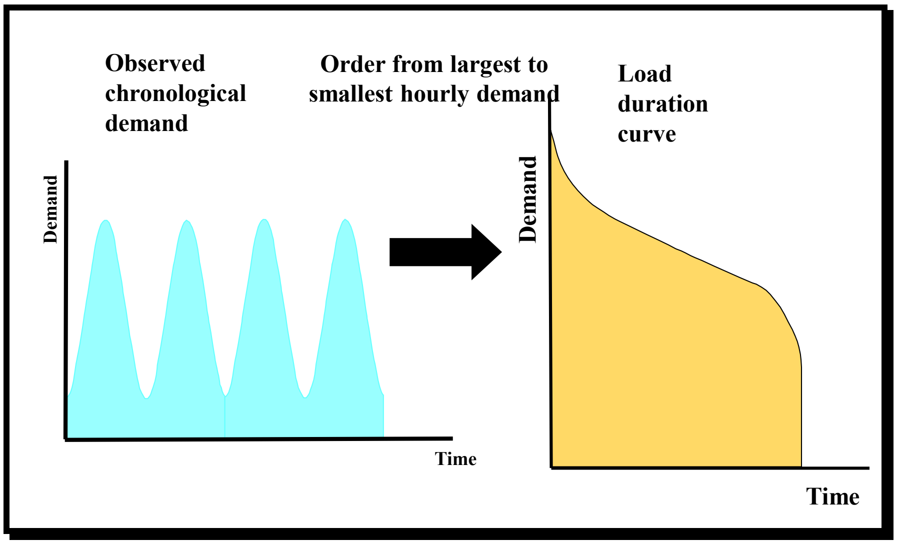
For short-term electricity modeling and representation, it is important to shorten the time intervals and to carefully and scrupulously maintain the chronology. Shortening time intervals is needed for example to represent electrical energy storage, pumped storage, load management, real-time pricing, etc. Maintaining the hour by hour chronology is the simplest but most detailed of all the timing assumptions for electricity. We will defer that discussion to the electricity storage section later herein. We will give a detailed hour-by-hour time point model of short-term electricity storage in that section. Maintaining the chronology is impossible for strategic level models designed to predict capacity addition and capital stock addition and rollover.

ArrowHead has recently built a one week, 168-hour forward model of electricity generation, consumption, and storage. The venue was Southern California. Such a model has to be totally and scrupulously chronological. The Southern California model was the simplest possible temporal assumption, but it imposes the most size and detail on the underlying model. It would never work for hour to hour storage and at the same time for endogenous capacity addition.

4.1.3 Overriding Hour-by-Hour Chronology to Develop Load Duration Curves

For long-run or strategic modeling, we recommend IEMM develop non-chronological load duration curves within the months but maintain the chronology by month. We begin with the hour-by-hour demand schedules for each of the subregions within each month as depicted in Figure 94 and reorder them in sequence from highest demand to lowest demand in each given month. Such a demand reordering process creates a series of monthly regional load duration curves as depicted in Figure 97. Such load duration curves represent total demand that occurs during a year, and they represent the total demand that occurs in each month of the year, but they do not represent total demand that occurs by hour within each month of the year. They implicitly do consider each and every hour, but they aggregate hours into portions of a monthly load duration curve. To wit, no hour is or can be omitted. But the hours are “grouped” or aggregated according to common load during that hour. By reordering demands from highest hour to lowest hour, we lose the chronology of hourly demand by day and by week within the month. However, we preserve the month to month distinctions. After reordering within each month, what we have is a highest-to-lowest snapshot of monthly demand as distributed throughout the hours of the month. Whether this reordered series of demands is sufficient for a given need depends on whether we are studying hourly and daily load following or unit commitment. If so, this would not be sufficient temporal detail. If not, the unordered representation of loads within the month will be sufficient. To wit, if we need hourly/daily cycle information, we need to retain the chronology. ArrowHead and IEMM allow retention of chronological information, and they allow the type of aggregation indicated in Figure 97. We should point out that IEMM should allow switching between short-term chronological loads and longer term grouped loads. There is on "one size fits all" with regard to temporality in IEMM.

Figure 97: Create Monthly Load Duration Curve by Sorting



Let us become concrete. Using a historical year as an example, Figure 98 illustrates three different load duration curve analyses of some demand data downloaded several years ago for a particular region. The figure depicts the average daily chronological demand pattern for each of the twelve months of the year, which is termed "Monthly Average Hourly Demand" in the figure and is represented in red curves. The average daily load shapes are the red solid curves that embody the characteristic double peak in the winter and the characteristic single peak in the summer. They are derived from the historical average hour by hour demand by month. They represent the average daily usage pattern within the indicated month. Such daily load shape information would be used to simulate daily operation in any short-term models we might build. This type of load shape information would be suitable for short-term market analysis and unit commitment analysis.

Returning to Figure 98, we construct 12 monthly load duration curves (depicted in gray with blue lines). The series of 12 monthly load duration curves is used to populate a strategic version of IEMM. That is, IEMM can contain the blue curves from Figure 98 to represent load by month, which we represent conceptually using the continuous curve in Figure 99. The 12 monthly load duration curves in Figure 98 are truly a hybrid chronological-non-chronological representation of load, and each one of them has the form in Figure 99. They maintain the chronology month to month, yet they erase the chronology within each month. They are conceptually ideally suited to the needs of long-term electricity modeling, as we will see shortly.

Figure 98: A Historical Pattern of Load for a Region

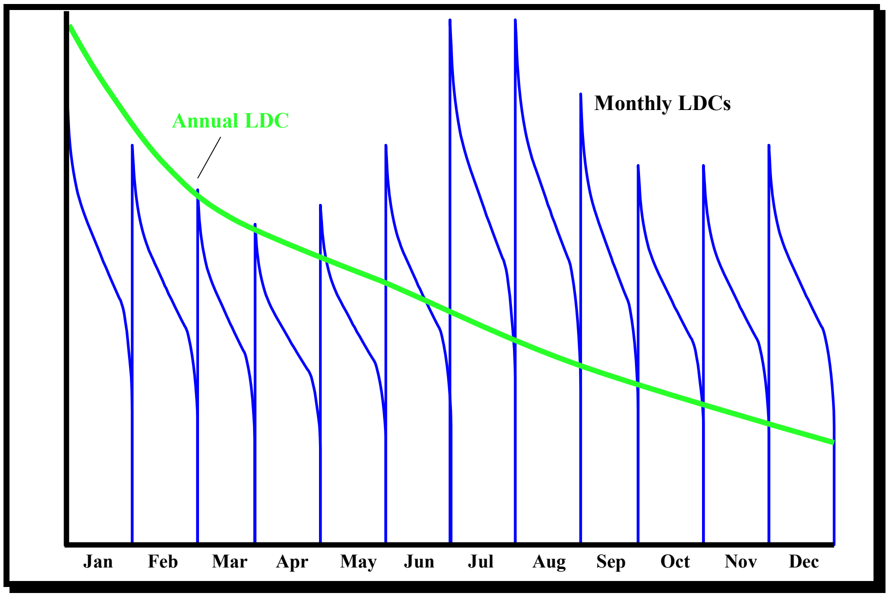
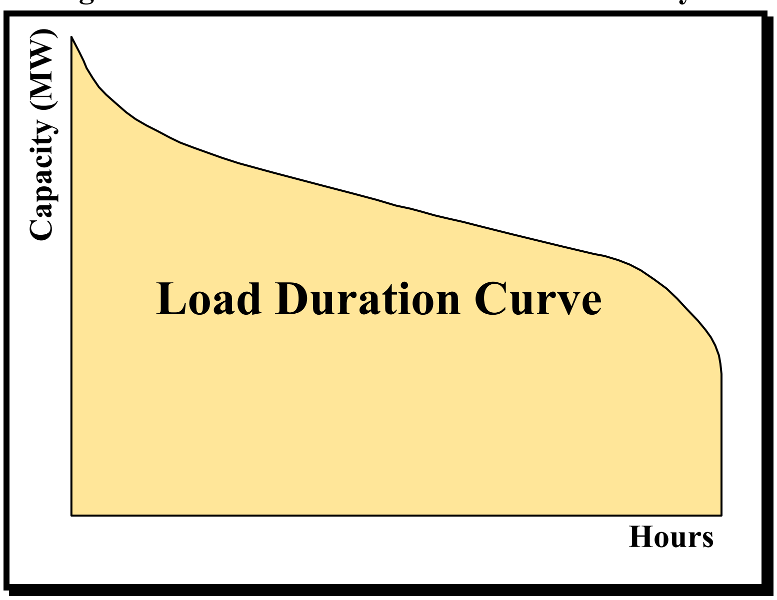


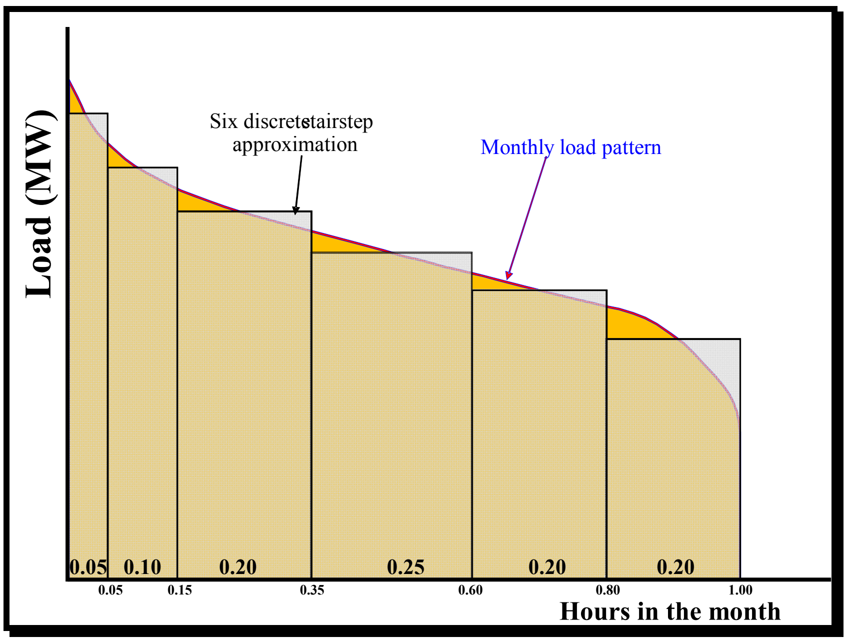
Figure 99: Each Region in the Model Has a Continuous Monthly Load Duration Curve



It is useful to review how we use the fundamental demand information in the form of Figure 98 that is derived from the fundamental data sources such as the FERC 714 reports. (EIA will assemble equivalent hourly chronological load information in other countries or regions.) The monthly load duration curves that come from the FERC 714 information have the continuous form in Figure 99 for each month of the year. Each of the months has approximately 730 hours, so each monthly load duration curve in Figure 99 has approximately 730 hours that compose that month. The 730-odd hours of the month ordered from highest load hour to lowest load hour forms a relatively continuous curve as shown in Figure 99.

For purposes of calculating prices at different levels of load, one takes the monthly load shape at the right in Figure 99 and makes a discrete stairstep approximation, dividing the hours of the month into discrete increments. We have selected six increments in the forthcoming examples (IEMM can use any number of discrete increments it wishes), thereby creating the stairstep approximation in Figure 99 for each month. Such an analysis renders load chronological between months and non-chronological within month. As we can see in Figure 100, this is a very standard mathematical and engineering problem: How do we approximate the continuous curve by a discrete stairstep curve? How many discrete stairsteps do we use? There is no obviously correct answer to this question, but there is an "energy preserving" answer. That is the answer in which the areas under the discrete load duration curve above the continuous curve are exactly equal to the areas below the discrete load duration curve below the continuous curve. Such an answer is pragmatic; it renders total energy identical for the discrete curve as for the continuous curve. If we use more and more and more discrete stairsteps, we get a better approximation to the continuous curve. However, as we shall see, more and more discrete stairsteps requires more and more computational detail on the load side of the IEMM. There is always the design tradeoff between more and more detail (more accuracy) and pragmatism of calculation.

Figure 100: Stairstep Approximation to Monthly Load Shape



Keep in mind as one thinks about how many stairsteps and thus how "fine" an approximation to the continuous curve in the figure one might want, the horizontal axis in Figure 100 is divided into the increments of hours in Table 1, showing from left to right, time of highest peak to time of baseload operation, respectively. Knowing that there are only approximately 720 hours in a 30-day month and knowing that the historical loads across each of those hours (and the future loads) are expected to be somewhat uncertain, we need to be extremely careful about creating too many sub-tranches of hours. There is motivation to disaggregate the hours at time of extreme peak and at time of extreme base, but there is much less motivation to do so at times of intermediate load. There just is not the benefit of discriminating middle intermediate load to the same degree.

Table 1: Number of Hours that Exist During Each Month and within Each Tranche within Month

	Hrs/Mo	Peak (0-5 % of hours)	Next (5-15 % of hours)	Next (15-35 % of hours)	Next 35-60 % of hours)	Next (60-80 % of hours)	Base (80-100 % of hours)
Jan	744	37.2	74. 4	148.8	186.0	148.8	148.8
Feb	672	33. 6	67.2	134. 4	168.0	134. 4	134. 4
Mar	744	37.2	74. 4	148.8	186.0	148.8	148.8
Apr	720	36.0	72.0	144. 0	180.0	144. 0	144. 0
May	744	37.2	74. 4	148.8	186.0	148.8	148.8
Jun	720	36.0	72.0	144. 0	180.0	144. 0	144. 0
Jul	744	37.2	74. 4	148.8	186.0	148.8	148.8
Aug	744	37.2	74. 4	148.8	186.0	148.8	148.8
Sep	720	36.0	72.0	144. 0	180.0	144. 0	144. 0
Oct	744	37.2	74. 4	148.8	186.0	148.8	148.8
Nov	720	36.0	72.0	144. 0	180.0	144. 0	144. 0
Dec	744	37.2	74. 4	148.8	186.0	148.8	148.8
Totals	8760						

The reason we discretize the monthly load duration curves is so that we can calculate a general equilibrium (integrated price-quantity-capacity addition) solution for each of the discrete load tranches. Each of the discrete tranches is going to be regarded as a distinct "market." There is going to be a complete region-wide price-quantity-capacity addition equilibrium for each of the time tranches in the figure, and the market equilibria across the tranches are going to be strongly interconnected. Load tranches are not independent, not in the slightest. (Market equilibrium solutions are not compatible with continuous load or demand curves; no economist knows how to do that to our knowledge. The best that can be done is temporal discretization. There may be a Nobel Prize and a bit of prize money waiting for someone who can deal with continuously varying load.) To do so, we must take care in selecting the number of discrete increments into which we want to disaggregate the continuous curve in Figure 99. Wanting to use a "stairstep" approximation to Figure 100, we ask how many stairs would be required to get a good enough approximation to the curve so that our price-quantity equilibrium solution is both robust in an economic sense as well as representing a sufficiently close approximation to the curve in Figure 99. As an example, if we elect to use six stairs to approximate the curve, we would be able to draw the 6-stair discrete approximation to the continuous curve in Figure 99 as shown in Figure 100. In a recent ArrowHead application in PJM in the year 2014, we have used 218 stairsteps to represent a year, 12 months in the year times 18 load tranches per month (not 6 tranches per month). That is much more realistic than the simple example used here of six discrete tranches. The process of moving from the continuous curve in Figure 99 to the discrete approximation in Figure 100 can be done using extremely fancy, sophisticated, statistical fitting methods, or it can be done approximately. What we see in Figure 100 is the following:

- Average load during top 5 percent of hours (designated 5%). The average load across the top 5 percent of the hours is lower than the true, maximum hourly superpeak, and this is a problem for electricity modeling. IEMM wants to define the superpeak hour fairly carefully.

- The average load including the top 15 percent of the hours but excluding the top 5 percent of the hours (designated 15%).

- The average load including the top 35 percent of the hours but excluding the top 15 percent of the hours (designated 35%).

- The average load including the top 60 percent of the hours but excluding the top 35 percent of the hours (designated 60%).

- The average load including the top 80 percent of the hours but excluding the top 65 percent of the hours (designated 80%).

- The average load including the bottom 20 percent of the hours (designated 100%).

Notice that the discretized monthly curve in Figure 100 and again in Figure 101 distinguishes six different, discrete demand levels because it has six different, discrete horizontal tranches in this illustrative discussion. Beginning at the left, we see the highest tranche, and we see a series of declining tranches until we get to the lowest horizontal tranche at the lower right. Each of these six horizontal tranches corresponds to a specific, discrete level or tranche of demand, each tranche of which we express in MW. But the discrete curve also tells us how many hours occur within each of the six discrete levels of demand in the curve. For example, the highest and leftmost level of demand persists in the diagram for 5 percent of the hours in the month. Assuming that there are 730 hours in the month, this means that the highest and leftmost level of demand persists in the diagram for 36.5 hours in the month. The second to highest and second to leftmost level of demand persists for 15-5=10 percent of the hours in the month, i.e., 73 hours in the month. Continuing this logic across the diagram, we see that it is in effect a histogram for the occurrence for six different levels of demand.

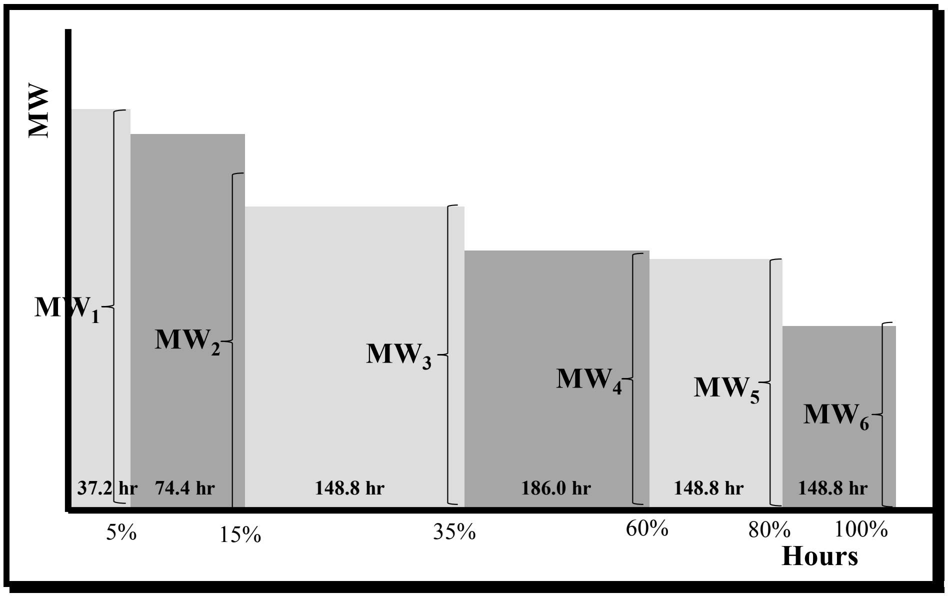
If we consider the curve in Figure 101, we have the peak load in the rectangle at the extreme left. It tells us what the average load is for the highest load collection of hours in the month, and it is the indicated level of megawatts MW_1.It is crucially important to note that the stairstep approximation assumes that the load MW_1 persists across the entire first interval, which represents 5 percent of the hours in the month or 37.2 hours in the month. This discrete stairstep approximation is extremely important. The first rectangle tells us that the load of MW_1 persists for exactly 37.2 hr. That means there are

$$MW_1 \times 37.2\text{hr} = 37.2MW_1 \text{ MWh}$$

of energy transacted under the leftmost load tranche. We think of that leftmost load tranche as a peak load "market" for electrical energy. The total size of that market is

$$37.2MW_1 \text{ MWh}.$$

Figure 101: Disaggregation of Each Month into Load Tranches



The number of hours that market persists is 37.2 hr., and the rate of energy production within that market is MW_1 megawatts, assumed to be constant and flat.

By contrast, the most off-peak, the deepest base load, is the rightmost rectangle. The diagram tells us what the average load is for the lowest load collection of hours in the month, and it is the indicated level of megawatts MW_6.It is crucially important to note that the stairstep approximation assumes that the load MW_6 persists across the entire sixth interval, which represents 20 percent of the hours in the month or 148.8 hours in the month. This discrete stairstep approximation is extremely important. The sixth rectangle tells us that the load of MW_6 persists for exactly 148.8 hr. That means there are

$$MW_6 \times 148.8\,\text{hr} = 148.8 MW_6 \text{ MWh}$$

of energy transacted under the rightmost load tranche. We think of that rightmost load tranche as a baseload "market" for electrical energy. The total size of that market is

$$148.8 MW_6 \text{ MWh}.$$

The number of hours that deep baseload market persists is 148.8 hr., and the rate of energy production within that market is MW_6 megawatts, assumed to be constant and flat.

Precisely the same logic persists for every block the discrete diagram. Keep in mind, this logic defines six discrete, average, representative levels of load that compose the entire month, one for each tranche of hours, and it specifies how many hours each representative level of load persists during the month. The stairstep curve approximates the continuous curve. We point out the following critically important dimensions of the diagram in Figure 101:

- The **heights** of the six rectangles are the megawatt loads (expressed in MW, a rate of energy delivery) that occur in each of the six-hour tranches. The height of the leftmost or peak load is the number we see so often reported by agencies such as the NERC. They only tell you the peak megawatts in the market, i.e., the height of the leftmost red curve in the diagram. They tell you nothing about energy; they tell you about the capacity that must be in place because this is a just-in-time industry.

- The **widths** of the six rectangles are the number of hours in the month that correspond to each of the loads.

- The **areas** of the six rectangles are the energies delivered from the generation system to the loads in each of the six hourly tranches expressed in megawatt hours. The areas represent total energy delivered during each of the six tranches of hours.

- The **sum of the areas of the six rectangles** represents the total energy delivered during **all** tranches of hours during the month. It is the total energy delivered during the month. Total energy is the number we see so often reported by agencies such as the NERC. It contains zero information regarding the shape of the load. It is just the total quantity of energy delivered to and consumed by the market. In this example, total electrical energy is

$$\text{Energy} = 37.2\text{MW}_1 + 74.4\text{MW}_2 + 148.8\text{MW}_3 + 186.0\text{MW}_4 + 148.8\text{MW}_5 + 148.8\text{MW}_6 .$$

It is important to note that once the load duration curve has been discretized as in Figure 101, there are only six (6) levels of generation in the whole analysis

$$\begin{gathered} \text{MW}_1 \\ \text{MW}_2 \\ \text{MW}_3 \\ \text{MW}_4 \\ \text{MW}_5 \\ \text{MW}_6 \end{gathered}$$

That is, during every hour of the month, demand will occur precisely at one of these six levels. Demand (and aggregate generation) will never occur at any level other than the above six levels. These levels of demand will occur for different numbers of hours, but there are only six discrete levels. A small number such as six is convenient for this illustration, but applications of IEMM will probably choose to use more.

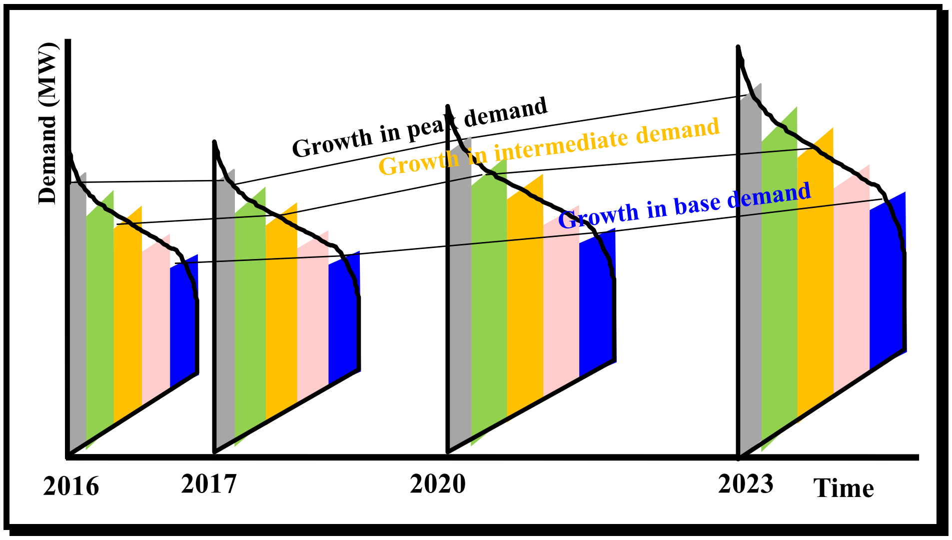
Organizations such as NERC and foreign governments present total annual or total monthly electrical energy generation, i.e., the preceding sum. Annual or monthly totals are not a particularly useful number in and of themselves. We need to understand the shape, the variation of load, and not just total energy of load. They also present MW_1, the peak MW load for the year. They often report the peak hour, not the peak tranche of hours. To wit, organizations such as NERC report: (1) the height of the leftmost (peak) rectangle and (2) the total area under all the rectangles. As can be seen, that is insufficient to properly characterize the load that the generation organization faces. One requires the height, width, and area of **every** rectangle. Calibration of loads to such organizations as NERC is often difficult or futile. One requires the actual distribution of loads during each and every hour of the month and year to obtain the proper characterization of load as in Figure 101. That is why IEMM will require the equivalent of FERC 714 data (which are hourly) rather than more aggregate or summarized information such as NERC. For many countries, this is available by analogy from contiguous countries or from electric utility reports. It should be available from in-country measured electricity deliveries. EIA can see that the typical "energy output" reports for indigenous electric utilities is inadequate. IEMM needs load shape information.

The diagram in Figure 99, as used in the model, can be regarded as specifying the load and hourly duration of load that occurred during a base year, say the year 2016. The original continuous load duration curve for that year, as represented by Figure 99, is derived directly from the FERC 714 or equivalent load data set, which tells us the exact load in every hour of every month that occurred during that year. To the extent that the FERC 714 is correct, there is no uncertainty about the process or the curve in Figure 99 for the historical year. The next question that arises in the economic context is how this curve grows or changes forward in time. IEMM has to specify the demand/load side of the model, and that means specifying the growth pattern of the curve in Figure 99 or equivalently the growth pattern of its discrete equivalent in Figure 101.

The process of future load growth and change is might be represented as in Figure 102. We have chosen only five tranches to create Figure 102 to keep the illustrative diagrams from becoming too busy, but the logic extends to as many tranches as one might need to use in IEMM. (There are likely to be more tranches in industrialized countries or regions and fewer in developing countries or regions. There will not be a common IEMM standard across every country or region in the world.). Notice the gray rectangles evolving forward in time in Figure 102. The height of the gray rectangle represents the load in MW that occurs during the peak tranche of hours. There is a height in the year 2016, a height in the year 2017, a height in 2020, and a height in 2023 in the diagram, and those heights are connected by the peak load line in the diagram. That line in the diagram represents the projected growth in peak MW. It represents the growth in that "market." That line is the projected growth in peak electricity demand that is assumed for the given market represented by in Figure 102 into the future. Notice in the diagram, the peak line is increasing through time, meaning that in this example, the peak load is projected to increase through time. Increasing peak is characteristic of a location or region that is experiencing growth in peak load demand, i.e., growing population and growing GDP. IEMM will have to characterize this carefully.

Turning to the orange intermediate line, which touches the heights of the middle intermediate curves forward through time, that is the growth of the intermediate load projected forward through time. Similarly, the base load line, which touches the heights of the baseload rectangles forward through time is the growth of the base load projected forward through time. We can see that projecting load and load growth is impossible with summary information such as the NERC. That information might help use with the red line (peak MW growth), and it might help us with the sum of all the lines (total MWh energy growth, which itself is rather meaningless when forecasting the nature of load growth.) IEMM needs to be more sophisticated, as this design is.

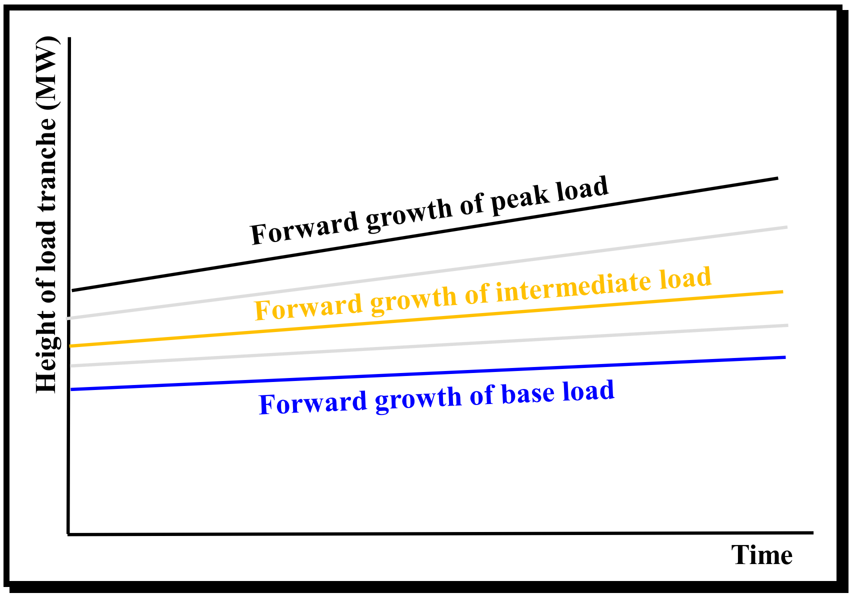
Figure 102: Representing Future Load Growth



We can plot the five forward load growth curves in this illustration, emphasizing peak, intermediate, and base, on a common plot as in Figure 103. The concept becomes very simple. IEMM postulates the rate of change forward in time of peak load, next to peak load, next to next to peak load,…, and so forth all the way down to base load. The widths of the load duration curves in Figure 103 remain constant, so this procedure grows the rate at which energy is being consumed (the heights of the rectangles forward in time expressed in MW) as well as the total energy being consumed in each tranche of hours (the areas of the rectangles forward in time expressed in MWh). This is a convenient forward projection of capacity, hours, and energies, and it is general as one can see. It allows us to project the **full forward load shape** and not be satisfied with the far less complete and more summary forward energy or peak load projections by organizations such as NERC. It is clear that we must project not only the full forward energy or the full forward peak load, but rather we must project the full forward load **shape** as well as peak, intermediate, and base load energies and capacities. The procedure is straightforward, and the accuracy is straightforward using the methodology and data that we do. We calibrate each of the loads to a historical time or period of times and then project forward in each tranche of load using growth or change logic for each tranche of load.

There is one key point that merits serious mention, and that is the issue of peak growth relative to energy growth. The curve in Figure 103 shows peak growing faster than intermediate and base. The load duration curve in Figure 101 projected forward through time would become more "peaky" and less baseloaded. The peak rectangle would grow relative to the other two rectangles. Increasing peakiness is what has happened over the past 20 years as developed countries have moved from industrialized economies to service economies. Increasing peakiness is what many assume in the load growths in most models of industrialized electricity systems, and it is likely to be what IEMM observes. If we grew the three curves at the same rate, there would be no marked change in the peakiness or base loadedness of the forward load shape. If we grew the peak more slowly than the intermediate or base, we would become flatter in our load shape, and that is what the politically correct forecasters eager to show the benefits of load management or load control generally want. Ostensibly, that has been the goal of peak load management, but it has not been particularly successful during recent years in North America. This type of load growth has occurred in Europe and may be occurring in places such as China. It needs to be carefully represented in IEMM.

Figure 103: Forward Growth of Heights of the Load Curve Tranches



It is very easy to see how to implement rapid growth of peak versus energy, modest growth of peak versus energy, or load management policies that trim the peak faster than they trim the off-peak. The proposed IEMM load representation allows examination of all those eventualities and their consequent impact on electricity market prices (capacity and energy) as well as quantities, capacity additions, and capacity retirements, and extremely critical requirement. As we shall see in subsequent sections of this CDR, it also has major implications for hydro, storage, and peaking techniques. The IEMM design in those sections will make that clear.

Pure, lockstep coincidence among times of load is a bit difficult and requires chronological treatment in the IEMM model. As an example, the absolute peak tends to sweep across the United States and Canada from East to West with the sun. The peak time in New York, say 200 pm Eastern, will not be the peak time in Chicago (200 pm Central, or 300 pm Eastern). There is a slight amount of excess capacity that rolls from east to west on the peak day, and this excess can only be captured with a fully synchronous model. That would require a lot more time-tranching of loads or even movement to an hourly model. It is not going to occur with most IEMM models, but this design would accommodate it.

Let us examine exactly how load and load shape data should enter the IEMM model. This is an important yet subtle mathematics that has to be done carefully to characterize the load shape for the given month. (We cannot count the number of times this simple mathematics has been done incorrectly!) Using the month of June as an example, the input data enters in the form in Table 2. Notice in the top portion of the table, we have 720 hours per month. In the middle portion of the table, we have the fraction of hours of the month that have been assigned to each load tranche. These are precisely the widths of the hour tranches in Table 1. They are the same hour fractions as in Table 2. Finally, in the bottom section in Table 2, we have calculated the absolute number of hours that exist in each of the hour tranches for the month of June. They are the product of the hour fractions in the middle section times the total number of hours in the month in the top section. The bottom of the table represents hours in the classic sense of hours—one hour of time elapsed.

Table 2: The Temporal Structure for June in IEMM

Hours per month (June)

2016	720
2017	720
2018	720
2019	720
2020	720

Fraction of Hours per Time Tranche

	1	2	3	4	5	6
2016	0.05	0.1	0.2	0.25	0.2	0.2
2017	0.05	0.1	0.2	0.25	0.2	0.2
2018	0.05	0.1	0.2	0.25	0.2	0.2
2019	0.05	0.1	0.2	0.25	0.2	0.2
2020	0.05	0.1	0.2	0.25	0.2	0.2

Hours per Time Tranche

	1	2	3	4	5	6
2016	36	72	144	180	144	144
2017	36	72	144	180	144	144
2018	36	72	144	180	144	144
2019	36	72	144	180	144	144
2020	36	72	144	180	144	144

We now turn to the issue of energy and capacity in each of the six tranches of load during June. If we begin with a source such as NERC, they report the number of megawatt hours that occur in the entire month of June. The total energy for the month of June expressed in terms of MWh consumed in the month is the sum of the areas of the blocks that compose Figure 101, i.e.,

$$\text{Total Energy} = 37.2MW_1 + 74.4MW_2 + 148.8MW_3 + 186.0MW_4 + 148.8MW_5 + 148.8MW_6$$

in the top section of Table 3. We can think of the fraction of energy consumption that occurs in each of the time tranches.

$$EF_1 = \frac{37.2MW_1}{37.2MW_1 + 74.4MW_2 + 148.8MW_3 + 186.0MW_4 + 148.8MW_5 + 148.8MW_6}$$

$$EF_2 = \frac{74.4MW_2}{37.2MW_1 + 74.4MW_2 + 148.8MW_3 + 186.0MW_4 + 148.8MW_5 + 148.8MW_6}$$

$$EF_3 = \frac{148.8MW_3}{37.2MW_1 + 74.4MW_2 + 148.8MW_3 + 186.0MW_4 + 148.8MW_5 + 148.8MW_6}$$

$$EF_4 = \frac{186.0MW_4}{37.2MW_1 + 74.4MW_2 + 148.8MW_3 + 186.0MW_4 + 148.8MW_5 + 148.8MW_6}$$

$$EF_5 = \frac{148.8MW_5}{37.2MW_1 + 74.4MW_2 + 148.8MW_3 + 186.0MW_4 + 148.8MW_5 + 148.8MW_6}$$

$$EF_6 = \frac{148.8MW_6}{37.2MW_1 + 74.4MW_2 + 148.8MW_3 + 186.0MW_4 + 148.8MW_5 + 148.8MW_6}$$

We term these the "energy fractions." It is crucial to note that these energy fractions cannot be calculated without the detailed monthly loads sorted into a continuous monthly load duration curve as in Figure 97 and after that discretizing that load duration curve as in Figure 100. One really needs the full spectrum of loads to accomplish this.

Table 3: Total Monthly Energy and Assignment of Total Monthly Energy to Time Tranche

Energy Consumption in June (MWh)

2016	5,004,000
2017	5,724,000
2018	6,444,000
2019	7,164,000
2020	7,884,000

Fraction of Energy per Time Tranche

	1	2	3	4	5	6
2016	0.071942	0.129496	0.230216	0.251799	0.172662	0.143885
2017	0.069182	0.125786	0.226415	0.251572	0.176101	0.150943
2018	0.067039	0.122905	0.223464	0.251397	0.178771	0.156425
2019	0.065327	0.120603	0.221106	0.251256	0.180905	0.160804
2020	0.063927	0.118721	0.219178	0.251142	0.182648	0.164384

Energy per Time Tranche (MWh)

	1	2	3	4	5	6
2016	360,000	648,000	1,152,000	1,260,000	864,000	720,000
2017	396,000	720,000	1,296,000	1,440,000	1,008,000	864,000
2018	432,000	792,000	1,440,000	1,620,000	1,152,000	1,008,000
2019	468,000	864,000	1,584,000	1,800,000	1,296,000	1,152,000
2020	504,000	936,000	1,728,000	1,980,000	1,440,000	1,296,000

The energy fractions represent the area of each block divided by the areas of all the blocks. Total energy for June and the fraction of energy per time tranche in June is accumulated as in Table 3.

We see in Table 3 the assumption of linear growth in total MWh (total electrical energy) in June in the coming five years. In the second section of Table 3, we have calculated the total energy consumed during the month in each of the six (in this case) hour tranches. In the context of Figure 100, the product of the top section of Table 3 times the bottom section of Table 3 gives the **area** of each of the rectangles in Figure 100, i.e., the energy consumed during each of the load tranches. That is, the area of tranche 3 in June 2016 (the total MWh consumed in tranche 3 during 2016) is the product of the total energy in June 2016 (5,004,000 MWh) times the fraction 0.230216 assigned to tranche 3. This product calculates the total MWh of energy consumed during the 144 hours of tranche 3, which is 1,152,000 MWh. Multiplication of the total energies in the top portion of Table 3 times the bottom section of Table 3 is tabulated in the bottom table in Table 3. (We are going to see that these energy fractions can actually be back-calculated from the MW we want to have in each of these discrete load tranches.)

At the bottom of Table 3, we have the total energies in each of the blocks in Figure 101. We have the areas of each of the blocks. Since the height of each block is expressed in terms of MW and the width of each block is expressed in terms of hours, these areas are expressed in terms of MWh. At the bottom of Table 2, we have the number of hours that represents the horizontal width of each energy block in Figure 101. We have MWh, and we have hours. If we divide MWh by hours, we obtain (dimensionally) MW.

Notice in Table 4 that we have arrayed the quantity of energy that is consumed during each of the six load tranches for June of the indicated year in the top portion of the table. These are the absolute numbers of MWh that are consumed in each of those tranches of time. These are de facto the "sizes" of the energy markets that IEMM will be modeling. The top section in Table 4 contains MWh of energy in each tranche. The middle portion of Table 4 contains the absolute number of hours in each load tranche. Clearly megawatt-hours divided by hours yields megawatts. That is, if we divide the energies in the top portion of Table 4 by the middle portion of Table 4, we obtain the MW of capacity being consumed in each of the tranches of load. (This is the measure of load that many are used to seeing, and here it is embedded in the IEMM load design.) That calculation appears in the bottom section of Table 4. This is what we normally think of as load—the frequency distribution of MW of load during June. This frequency distribution needs to be calculated in a very specific fashion and input to IEMM as indicated in this brief discussion. We could if we desire reverse the process, specifying the number of MW of capacity by tranche and after that calculating the energy consumed in each tranche of time. That would be equivalent to the discussion of the calculations in these tables. The point here is that MW, MWh, and hours must be lockstep consistent. Notwithstanding, the markets represented in IEMM are the six tranches of hours, and the IEMM is designed to meet the demands in those markets and price energy into those markets one by one.

It is the set of MW loads at the bottom of Table 4 that people are used to seeing. Yet, that begs the question of hours and the question of energy (MWh). That is just the rate of demand in each tranche of hours. The IEMM design makes that crystal clear.

Table 4: Energy and Capacity per Sub-Time Tranche during the Month of June

Energy per Time Tranche (MWh)						
	1	2	3	4	5	6
2,016	360,000	648,000	1,152,000	1,260,000	864,000	720,000
2,017	396,000	720,000	1,296,000	1,440,000	1,008,000	864,000
2,018	432,000	792,000	1,440,000	1,620,000	1,152,000	1,008,000
2,019	468,000	864,000	1,584,000	1,800,000	1,296,000	1,152,000
2,020	504,000	936,000	1,728,000	1,980,000	1,440,000	1,296,000

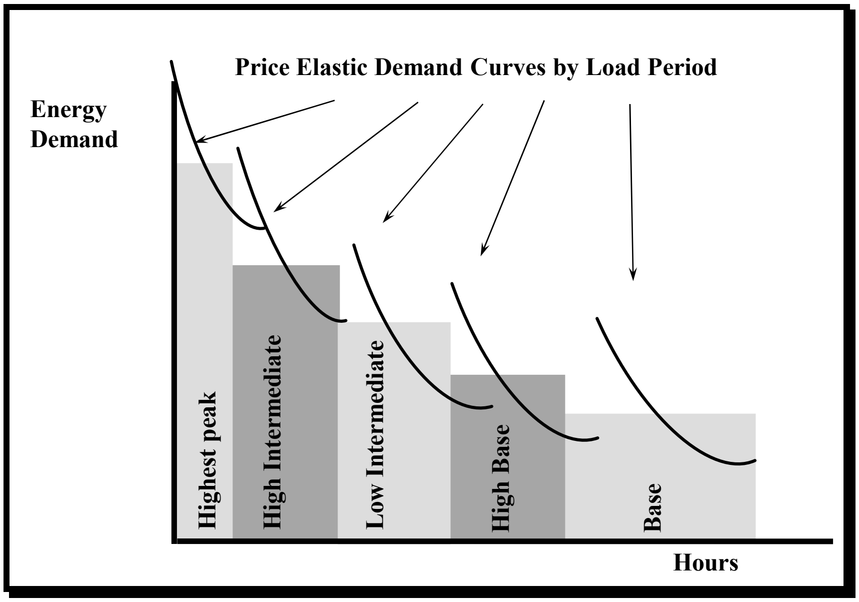
Hours per Time Tranche (MWh)						
	1	2	3	4	5	6
2,016	36	72	144	180	144	144
2,017	36	72	144	180	144	144
2,018	36	72	144	180	144	144
2,019	36	72	144	180	144	144
2,020	36	72	144	180	144	144

MW by Tranche						
	1	2	3	4	5	6
2,016	10,000	9,000	8,000	7,000	6,000	5,000
2,017	11,000	10,000	9,000	8,000	7,000	6,000
2,018	12,000	11,000	10,000	9,000	8,000	7,000
2,019	13,000	12,000	11,000	10,000	9,000	8,000
2,020	14,000	13,000	12,000	11,000	10,000	9,000

4.1.4 Demand Elasticity

The discussion thus far has specifically assumed inelastic, price-insensitive demand. (Alas, that is exactly what "dispatch" models invariably assume.) IEMM must have the capability on a load tranche by load tranche basis to specify a price elasticity of demand so that the actual projected demand for electricity by the model is price-sensitive. There is growing prospect that electricity pricing will become more real time than today, and that requires the notion of price elasticity. The IEMM price elasticity capability is represented schematically in Figure 104. (Again, demand elasticity can be assumed in an hourly model, but the number of hourly elasticities would be prohibitively large. This is one of the reasons IEMM will need to aggregate load temporally.) Notice in the figure that we specify price-sensitive demand curves for each of the discrete increments of load. There is a peak load demand curve complete with price sensitivity, an intermediate load demand curve complete with price sensitivity, and a base load demand curve complete with price sensitivity. Many believe this price sensitivity to be potentially important if the true marginal cost of on-peak power becomes increasingly exposed to electricity customers for the first time in one or more (but not necessarily all) regions that IEMM has to model. In the past, customers have been insulated from the true marginal cost (i.e., the true price) of peak load power because of regulatory cross-subsidies of peak load prices by base load prices and of residential customers (who cause the peak to a significant degree) by industrial customers. Our price sensitivity feature promises to be important today and in the future. Without it, one simply cannot project power prices at time of peak. IEMM needs to have the capability to quantify these peak for off-peak subsidies and calculate the modification the generation fleet that logically persists in the face of that.

Figure 104: IEMM Should Represent Price Elasticity in Each Load Period



Where would one get these elasticities? One could get them subjectively, and one could see how to estimate them statistically. To represent demand side elasticity, we have adapted the general Koyck lag model.[17] In this discussion, we assume time periods of equal length, say one month. (We can generalize to variable length time periods.) In the case of monthly time points, the Koyck lag model postulates the following functional form for the demand function, assuming all the parameters are expressed on a monthly time basis and that the constants represent monthly changes:

$$\frac{q(t)}{Q(t)} = \left[\frac{q(t-1)}{Q(t-1)}\right]^{\alpha} \left[\frac{p(t)}{P(t)}\right]^{\varepsilon}.$$

This formulation calculates the ratio between the demand q(t) and a reference demand Q(t) as a function of the ratio between the demand in the previous month q(t-1) relative to the reference demand in the previous month Q(t-1). This formulation requires as inputs:

α = lag coefficient.

ε = short-term price elasticity coefficient.

[17] The Koyck lag is what is termed a one period autoregressive model in the logarithms of prices and quantities, well known and studied in the public and professional literature. IEMM needs to have variable length time periods of arbitrary length. This discussion will assume equal length time periods for simplicity. Variable length time period mathematics have been published in Dr. Jill Scotcher's dissertation.

Q(t) = reference demand quantity through time.

P(t) = reference price through time.

To formulate the model, we begin by ignoring the lag term in the demand equation altogether. If there were no lag term, the market would adjust immediately and completely to the level dictated by the reference price. By so doing, we would compute the level of immediate and full response or "target" value of demand, which we denote with an asterisk:

$$q^*(t) = Q(t)\left[\frac{p(t)}{P(t)}\right]^{\varepsilon}$$

We can think of q*(t) as the instantaneous target value of demand, i.e., the demand that would be achieved if there were zero inertial of lag effects of any type in the market. If there were no lags, the market would adjust immediately to this target value q*(t).

Parenthetically, this formulation works with values of 1.0 for the reference quantities, prices, incomes, and degree days. If we set all the reference parameters to 1.0, all parameters q*(t) and p(t) would be expressed as absolute values rather than relative values. If we were to use 1.0 in place of the reference values, it is clear that the elasticity parameters α and ε, will have different magnitudes than if the reference quantities are used with nontrivial values.

Having conceived an instant response target value of demand, what we now recommend is an autoregressively lagged response function that uses the function q*(t) as a "target" value. That is, we want a lagged response function q(t) that adjusts itself continuously over time so as to approach q*(t) in a lagged fashion. Continuous adjustment is what we need to have a general lagged demand structure. The specific lag function we will use, the standard lag function in the Koyck lag formulation, is the generalized, continuous time, geometric/exponential lag form. In particular, the functional form of the lag function is

$$q(t) = a + be^{-\beta t}$$

where a and β are constants of the lag function. We want this lag function to satisfy two boundary conditions. That is how we calibrate the parameters α and β. In particular, we want as boundary conditions the following two equations:

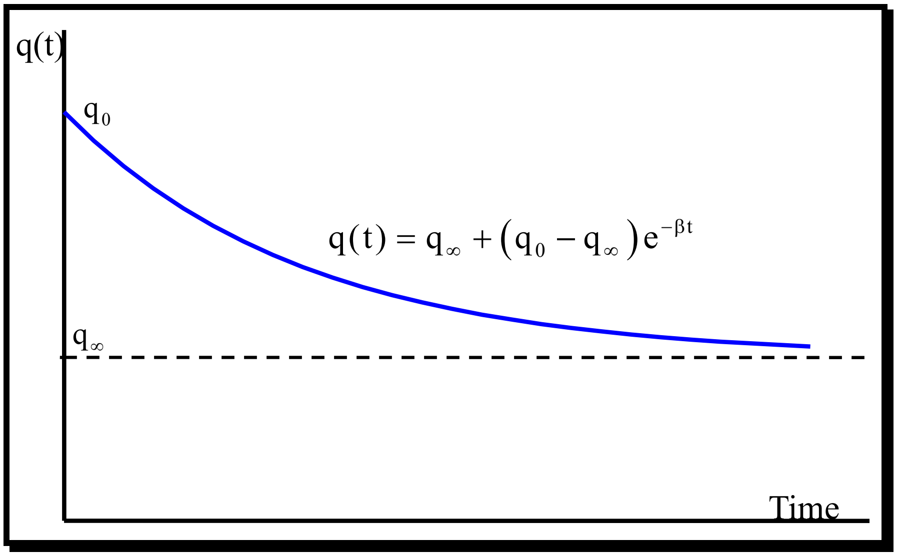
$$q(0) = q_0$$
$$q(\infty) = q_\infty$$

That is, we want to specify the value exogenously at time 0 and time infinity. This function along with the boundary conditions can be written in the following general asymptotic form:

$$q(t) = q_\infty + \left(q_0 - q_\infty\right)e^{-\beta t}.$$

which is an asymptotic exponential lag function as shown in Figure 105. Notice that the larger the value of β, the faster the adjustment and the shorter the lag. We will be equating the asymptotic value with the "target value" from the demand formulation above, and we will be lagging the adjustment toward that value in the demand model.

Figure 105: Exponential Lag Function



The general formulation of a Koyck lag model (a first order autocorrelated lag model in the literature) is based on the dynamic target value equation. If we precompute a target value q*(t) over the model horizon and a first-time point demand level q(0), we can lag the rate of adjustment of the actual demand q(t) toward the target demand q*(t) using the variable length time points. Using this concept, the following lagged demand response equation is

$$q(t) = q(t-1)e^{-\beta} + q^*(t-1)\left(\frac{1-e^{-\beta}}{\beta} - e^{-\beta}\right) + q^*(t)\left(1 - \frac{1-e^{-\beta}}{\beta}\right).$$

It is clear from the structure of this equation that the demand curve data includes the following parameters:

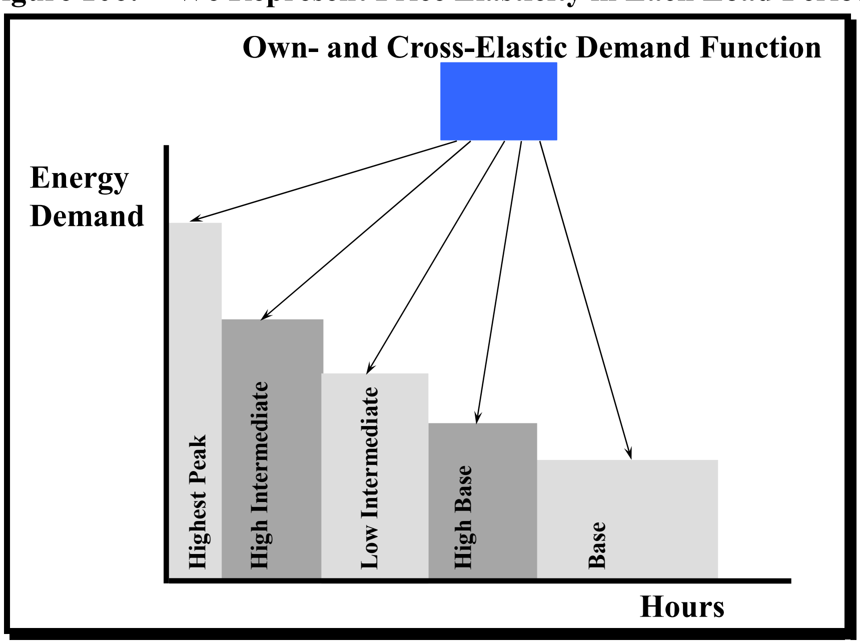
β = lag coefficient, expressed as a fraction per year
ε= short-term price elasticity coefficient (negative in sign).
Q(t) = reference demand quantity through time.
P(t) = reference price through time.

That is the entirety of the IEMM load data—a lag coefficient, a short-term elasticity coefficient, a reference price and a reference quantity. (IEMM can include other terms in this demand representation if EIA so chooses, but we defer that discussion.) This data is inserted into each demand curve in the model for each of the tranches of load. When this data is inserted, the demand curve in the model becomes sophisticated. If you insert into the demand curve a complete forward time track of prices p(1), p(2), p(3), …, the demand model returns a complete forward time track of quantities q(1), q(2), p(3). These quantities are price and price elasticity dependent, and they contain a lag term. That means that the model is completely general with respect to short-term and long-term price elasticity of demand for each of the tranches of load.

The reality is that no one really knows the short-term price elasticity of demand in most electricity venues around the world. IEMM can assume inelastic demand very easily, and we can postulate price elasticities subjectively by extrapolation from other commodities. To our knowledge, no one has ever successfully measured or observed the price elasticity of demand tranche by tranche. We know how to do that (and it is included in this IEMM CDR) based on FERC 714 and EIA data, but it is quite a large job. It also might not be totally appropriate at time of peak, which for issues such as real-time pricing is the essential piece of the load puzzle. IEMM has the richness of structure and methodology to represent the full range of price elastic effects. To date, such representation has been reasonable but subjective. That is very useful because we believe that subjective estimates of demand elasticities have a lot of value. Everything need not be "statistical."

Load management will require knowledge of cross elasticity between load periods, i.e., load-shifting. Methodologically, this is fairly easy. From a data perspective, this is difficult. The situation would appear conceptually as in Figure 106. We would need to understand the rate of substitution of peak for off-peak as a function of the absolute and relative prices of peak and off-peak. Installing the function would be easy. Fitting it to historical data or estimating it would be somewhat difficult. Cross elastic demand functions have been well studied in the literature.

Figure 106: We Represent Price Elasticity in Each Load Period



For the remainder of this section, we assume inelastic demand for simplicity. However, we can impose price elastic demand very simply and very immediately.

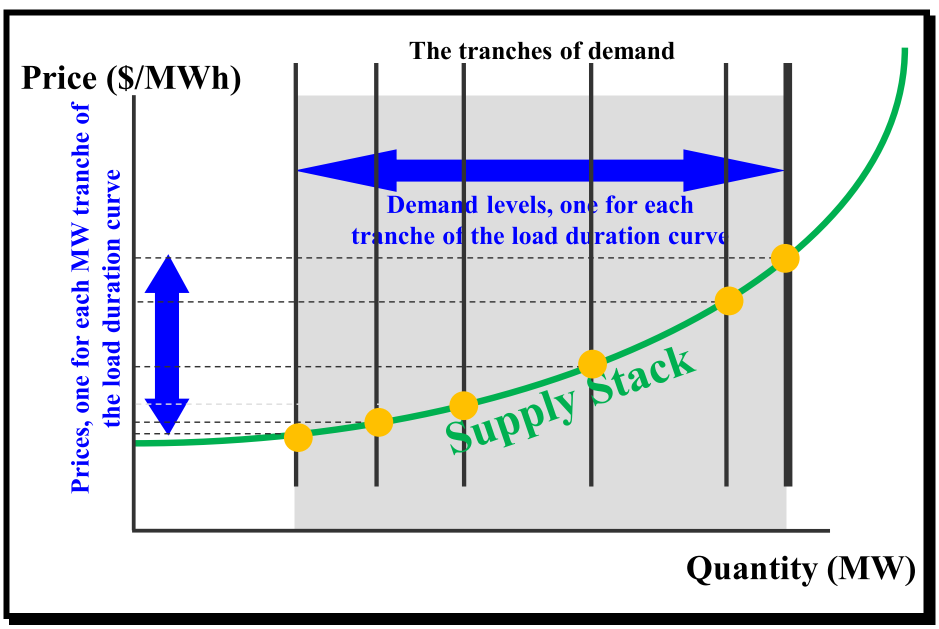
4.2 Chronology by Hour

No one need ask about chronology by hour. It works the same way as the LDC, but it has hourly time points. There will be a market in every hour, and IEMM will calculate it.

5 SUPPLY-DEMAND BALANCE WITHIN THIS LOAD AND GENERATOR FRAMEWORK

If we were to perform the conceptually simple exercise of calculating the supply-demand equilibrium for each of the different levels of demand by month developed in the previous section, we would be calculating a discrete histogram of market clearing prices, one for each tranche. The logic would be simple. Each tranche would represent a fixed, inelastic level of demand (recall that the discrete load duration curve de facto defines discrete markets), and the supply stack would cut across the six tranches at six different points. Figure 107 illustrates conceptually. There would thus be six different, distinct market clearing prices each with a corresponding frequency of occurrence (i.e., each persisting for a specific number of hours). Although the long-term IEMM model will do this in a more sophisticated fashion, the simple mnemonic in Figure 107 is a good way to express it. (Had we retained 8760 hours, we would have 8760 vertical curves, each persisting for an hour and therefore 8760 chronological prices. The logic is identical. However, we would have had to have load following and ramping in our marginal cost curves, not a particularly easy task.) In the six-tranche disaggregation of the monthly load duration curve, six demand points (each with a frequency of occurrence) are specified. Those six demand points correspond to six vertical lines in the diagram in Figure 107. Six supply-demand crossing points are calculated, one for each of those vertical capacity tranche curves, giving six market clearing prices distributed as indicated on the vertical axis in Figure 107. These six market clearing prices occur with the same frequency of occurrence as the six demand tranches that generated them. Therefore, the six prices compose in effect a histogram over prices during the month, a so-called price duration curve. (Precisely the same logic would occur with 8760 hours.) Keep in mind, the price duration curves compose a histogram of prices for a single, deterministic case.

Figure 107: Inelastic Supply-Demand Equilibria Are Calculated for Each of the Six Load Tranches



The reason that the simple representation in Figure 107 is merely that—a representation—is that the supply stack shown in the figure for the given region cannot, in fact, be drawn at all. It is, in point of fact a complex function of fuel substitution; investment, operation, and retirement of generation; inbound transmission from contiguous regions; outbound transmission to contiguous regions; environmental emissions costs, renewables, and other market phenomena, all the dimensions of the previous two sections. In spite of that complexity, however, the plant operation logic is conceptually as simple as that in Figure 107. The power of IEMM is to represent the sophistication and complexity of local supply within every region.

5.1 A Market Clearing Price at Every Load Tranche

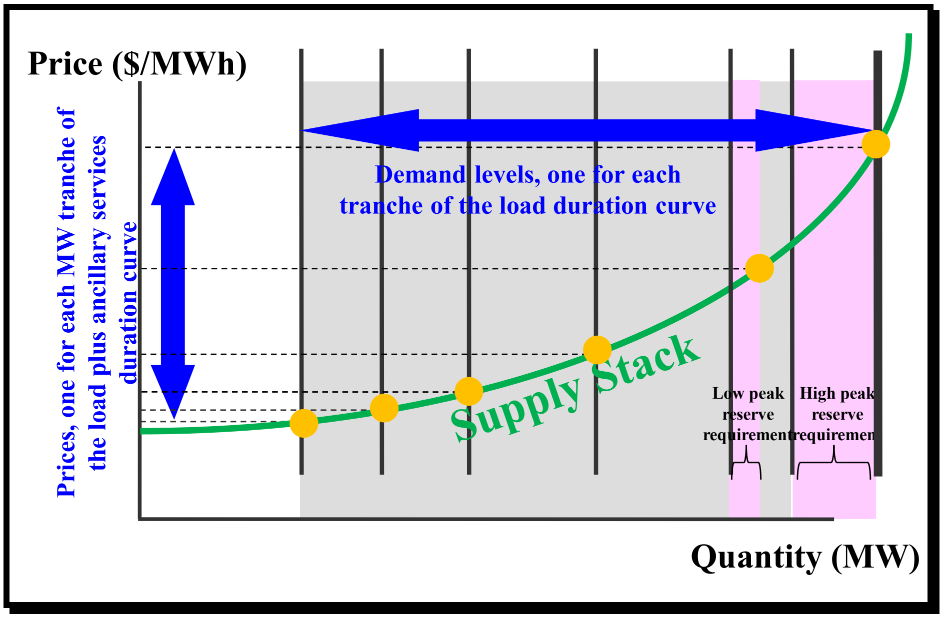
One of the key observations in this section is that there is a market clearing price of electricity at every load tranche. Every load tranche is, in fact, a discrete market, and the size of that market is the number of MWh, the quantity of energy, that transacts in that market. There is a price at time of peak. There is a price at time of intermediate. There is a price at time of base. Peak, intermediate, and base are markets. There is de facto a price of electricity for every hour of the month and year. The IEMM is determining the price of electricity at every time period and every hour (e.g., every tranche of hours) in the model. The fundamental unit of market clearing is the tranche of hours within the month or year. The answer the IEMM calculates is the price expressed in $/MWh for every MWh that is transacted at time t in hour tranche j. The energy transacting in each tranche of hours is deemed a market.

5.2 Reserve Margin

The notion of reserve margin (which should be defined locally throughout the IEMM model, just as it is and should be in the real world) is that there must be capacity online above and beyond what is needed to satisfy net energy for load. Electricity has been a real-time business, and it will continue to be to some degree. If one of the just-in-time assets is down for maintenance or experiences an unexpected outage, there needs to be another real-time alternative that can take over essentially instantaneously. That is what system reliability means. If we take the six vertical demand schedules in Figure 107, particularly the rightmost two; there has to be **more** generating capacity in place and operating than the level of actual load itself. If a unit fails, there must be another unit online to immediately initiate generation so as to continue serving power load. That is what redundancy means.

With regard to reserves or backup capacity, there are three basic categories to be considered—spinning reserves, standby reserves, and regulation. Together they add up to say 15 percent or so extra at time of peak. Reserves are appended to the right-hand side of the rightmost demand lines in Figure 107. Peak hour reserves can be as high as 15 percent of peak hour demand. There is additional demand appended to the right of the peak line in Figure 107 and shown in Figure 108 in lavender has to be added in the sense that this many plants have to be up and running to meet the peak hour net energy for load. These reserve plants have to be up and running and ability to generate very quickly so that the system does not go down in the event of a forced outage. Unless the market price falls at the point where the lavender reserves curve plus the black peak demand curve, these plants will not be "paid" to be up and running and thus will not be up and running.

Figure 108: Inelastic Supply-Demand Equilibria Are Calculated for Each of the Load Tranches Augmented for Reserves at Time of Peak



The data IEMM needs for reserve margin is simply the percent "overage" the system needs in each of the demand tranches. The reserve margin is higher at time of peak than at time of base because there is excess capacity at time of base.

For many countries in the world, there are low or no reserves. This reserve discussion would not be relevant to such countries unless they develop to a degree, and certainly many of them will do so during the EIA forward forecasting period. Their capacity would not be augmented to account for excess generation capacity online to allow for spinning or standby reserves or regulation capacity/energy until and unless they develop to the point of managing system reliability.

IEMM will have to worry about the spatial distribution of reserves. In some jurisdictions, power system people do not think of inbound power lines as providing any reserve value whatsoever. They reason that generation plants that come via transmission are geographically so far away and can easily be channeled to their own reliability needs that they cannot be counted as reserves in the destination region for the transmission link. By contrast, however, some believe that in fact generators upstream from transmission are de facto a "portfolio" asset and, in fact, can be counted as reserves in the destination region. The geographic/spatial nature of reserves, which we shall address in the transmission design below, has to be taken into careful account in most systems including the United States and various foreign systems.

5.3 How Will IEMM Represent Such Reserves?

It is known that the delivered value of electricity is complicated because of quality and reliability issues that must be dealt with by generators and transmission system operators. Real-world electric products and services can be aggregated into four primary categories:

- **Energy.** Energy can be regarded as megawatt-hours of electricity at a particular market location, i.e., a competitive hub, according to a particular time pattern. The discussion thus far has focused on energy demand. This is often called "net energy for load," implying that it is electrical energy that people actually use to do work or secure services.

- **Regulation Must Run (RMR).** These are plants that are technically situated in the transmission system that they must generate at roughly full capacity or else the transmission system becomes unstable, shuts down, or substantially reduces system capacity/throughput. These are units that cannot be interrupted, and if they are substituted for, they must be substituted for by units that run all the time that are located in essentially the same place on the transmission system. RMR is a technical requirement, and it has to be modeled as a different market.

- **Term Ancillary Services.** Term ancillary services include planning reserves, operating, reserves, spinning reserves, stand-by service, and backup service. For such services, there is typically a market hub pricing dimension and a basis differential dimension, where the basis differential is measured at some point downstream from the relevant market hub.

- **Real-Time Ancillary Services.** Real-time ancillary services are required to provide continuous-time system reliability and system stability. They include voltage support, frequency control, reactive power, load following services, and scheduling/operation services. Such services are in one sense highly localized and in another sense network-wide. It is not always clear how or where such services are monetized, but they must be monetized or they will not be provided. Because of the "network reliability externality" prevalent in many systems, they might well be monetized over a much broader area than they are produced and introduced.

- **Capacity or Reservation.** Capacity or reservation (sometimes called a demand charge or an access charge) is best regarded as a derivative product that derives its price from the value of energy and of ancillary services. IEMM will calculate capacity value explicitly as discussed previously, but it is the fully internalized value of capacity. Unless reserve capacity can augment the value of energy or ancillary services, it has no intrinsic value on its own. It goes without saying that the value of capacity depends on the regulatory environment. Regulation can distort or modify the value of capacity substantially.

The IEMM model as proposed herein represents the capacity or reservation dimension, the energy dimension, and the term ancillary service dimension. It only approximately represents real-time ancillary service—it has a "set aside" for real-time ancillary service. Models that ignore the various categories of ancillary service understate the need for and value of capacity. Most analysis to date has not attempted to quantify the fact that in an actual power system, one needs a number of localized ancillary services to operate the system and that each nodal market and in fact the aggregate market as a whole is willing to pay substantially for these services. Such services are intrinsically localized or Balkanized, and they intrinsically reside contiguous to major load centers. They are usually represented as "must-run" units (sometimes referred to as "regulatory must-run" or RMR units) exogenously inserted into the model solution. Ancillary service is an important element of IEMM and is calculated endogenously on a region-specific basis.

The nature of ancillary services is such that the supply stack that is normally thought to characterize a nodal market is not available in its entirety to satisfy energy demand as reflected in the load projections. Some of that supply stack is dedicated elsewhere. Indeed, a goodly bit of the capability in the supply stacks for each node (which are intended to be all-inclusive with regard to generation capacity) are in point of fact incapable of supplying energy to the demand side because they are committed to providing ancillary services such as spinning or standby reserves.

- **Regulating reserves (5 percent).** Peak power is defined in terms of some short time interval (e.g., 60 minutes continuous). Regulating reserves have to meet variations within that specified time interval on a continuous or almost continuous basis. They are not available to produce energy into the bulk market because they are committed to provide important localized reserve or backup services as their primary objective. Regulating reserves might eliminate 5 percent of the available capacity from the supply stack and thereby derate the supply as a result.

- **Spinning reserves (5 percent).** Spinning reserves are thought of as immediate outage backstops. If something fails, they must within a very short time interval such as 10 minutes be fully operational and provide energy to the grid to replace whatever unit or transmission link that has failed. Because they must be dedicated to providing reserve capacity, they necessarily cannot be dedicated to providing bulk energy in the supply stack to the market. They are therefore deleted from the overall supply stack for energy. Most people estimate spinning reserves at time of peak to be approximately 5 percent (slightly larger than the largest unit or two online). An alternative view would be that an amount of capacity equivalent to the largest plant or two in each region should be held in reserve.

- **First, Second, and Subsequent Contingency Operating Reserves (5 percent).** Typically, there is another echelon of reserves that must respond during some longer period (say an hour or so) of an outage. Operating reserves are dedicated to providing outage backstop and, therefore, cannot be dedicated to another use such as providing bulk energy for load. This block of reserves, therefore, must not be included in the supply stack that serves bulk energy load.

These various categories of ancillary services have very high value in certain localized regions but not necessarily in other localized regions. It is important to identify and quantify them nodally/locationally within the market because they affect costs and capacity in place. IEMM will explicitly and systematically do so, specifying specific reserve quantities for regulation, spinning, and operating reserves that persist during the tranche of super-peak load (tranche number 1) downward to the tranche of deepest base load (tranche number 6). As an example, we augment the net energy for load monthly curve in Figure 101 by the numerical values in Table 5. Such assumptions can be easily changed and must represent the region being represented within IEMM.

By representing reserves in this way, IEMM gets a better answer with regard to capacity in a market setting, and it gets a better answer to capacity value versus energy value in a market setting. We should also point out that by specifying ancillary service requirements as demand curves, they too are price elastic. That means that the quantity of ancillary services demanded may well be a function of their price.

Table 5: Representative Reserve Margin Assumptions

Load Tranche	Reserve Percentage
1 (ultrapeak)	14%
2	10.5%
3	4.5%
4	0%
5	0%
6 (base)	0%

Where might one assemble the appropriate reserve margin assumptions? Historically, people have used the rule of thumb that one must have spinning reserves equal in magnitude to the two largest plants on their system. That has often yielded numbers in the 10 percent range. People believe that 5 percent is needed for regulation of the system, yielding a total ancillary service number in the range of 15 percent. Some regulatory bodies have in the past postulated numbers like 18 percent reserves at time of peak, but people worry that such level of reliability might be excessively costly. 15 percent has become an informal standard in the United States as we perceive it, although that number might have to increase if renewables enter the system en masse. That 15 percent number is in our opinion an approximate standard for a thermal plant system. There is judgment regarding the appropriate reserve margin to use at time of peak and near peak, but 15 percent **local** or nodal reserve margin (which is what IEMM might often assume for places like the United States, Canada, or Europe) seems reasonable, representative of industry practice, and not overly costly or "gold plated" in an economic sense. Developing countries generally cannot afford reserves, so the reserve margin is zero or very close to it for a significant number of electrical systems around the world. IEMM needs to consider the full range, and this proposed IEMM design does.

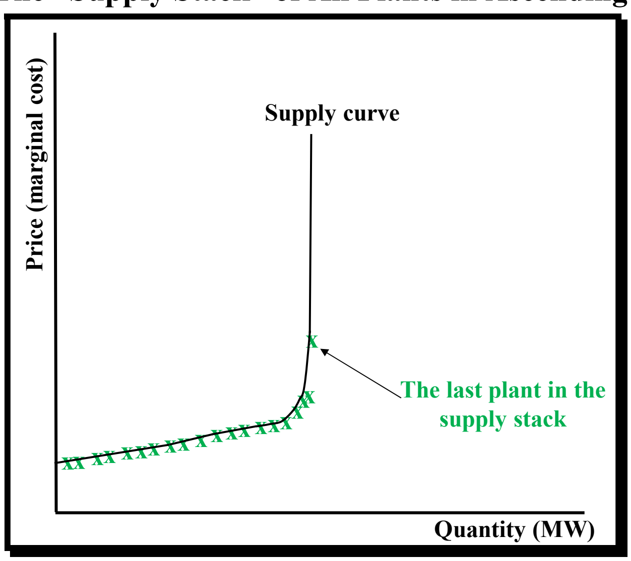
To secure the reserve margin say of 15 percent, there must be payments by the market to capacity that is not generating but rather is spinning or regulating. This IEMM design allows EIA to consider that the capacity is "oversized" relative to the pure net energy for load, and it requires that the market "pay for it." That is a correct and critical assumption—the market must pay for capacity that is used for reserves; otherwise, generators will not provide it, and it will not be available, and system reliability will degrade. IEMM, as designed here, does this by calculating a price of energy and capacity that at the margin just compensates the marginal plant that is providing reserve capacity. The market prices calculated by the model are correct in the sense that they compensate the reserve requirement at time of peak and near peak that is required to meet the reserve requirement, yet at the same time, they do not pay for reserves that are not specified as needed. The capacity component of the price at each of the located nodes is calculated correctly if we oversize the demand to account not only for net energy for load but also for reserves above and beyond net energy for load. This calculation is altered by the presence of environmental emissions allowances, which alter the capacity value for gas versus coal versus oil versus renewables and peakers versus non-peakers.

The model described in this CDR "idles" plants that are outside the margin (plus reserves). That is, plants that are not economically competitive to provide net energy for load **or** reserves are simply idled. They are paid nothing by the market; the market prices are not sufficient to pay any margin for them to operate or to maintain themselves or to decommission. This idea is likely the correct and proper model for retirements for all plants save nuclear. Plants when they are no longer economic (which happens because their fixed plus variable cost structure escalates above the capital plus variable plus fixed operating and maintenance cost of a new plant, which is then built) simply fade away and are no longer operated. The market pays them nothing. They are extra-marginal plants and are never operated again. Their cost structure continues to escalate upward because of age, and they are effectively displaced and replaced by new vintages of plants, that themselves get on their own aging and cost escalation track. The preceding is a very simple yet accurate model of plant retirements—they just cease to be part of the operating mix and the reserve mix as their costs become uneconomic due to aging and due to cost preference of new plant types.

5.4 The "Dispatch" Mindset Does Not Work for IEMM

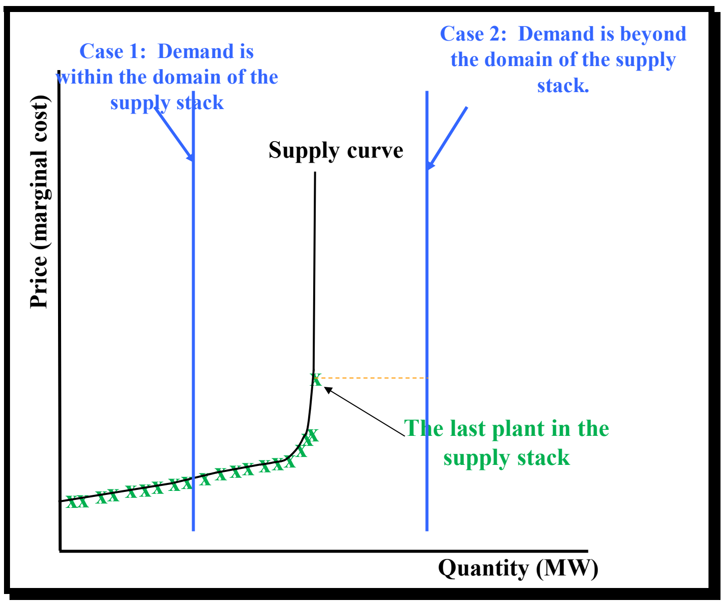
One of the problems with the "dispatch" mindset is the notion that the price of electricity (called system lambda by dispatch modelers, whose operations research background becomes painfully obvious by the very moniker) should be equal to the cost of the marginal generating plant. The marginal generating plant is the one with the highest generation cost that is still operating in the generation mix. The dispatch mentality leads people to conclude, erroneously, that the price of electricity must always be set by one of the plants in the market. IEMM will not succumb to the erroneous dispatch mindset but will calculate the power price correctly. To see why, consider in Figure 109 what the sequence of generation plants in a region lined up in ascending order of generation cost might appear.

Figure 109: The "Supply Stack" of All Plants in Ascending Order of Cost



If we were dispatch modelers, we would have one of two situations. In the first situation, the demand for power would cut through the supply stack at a point to the left of the vertical line. We term that Case 1, and we highlight it at the left of Figure 110. That is the normal case typical dispatch modelers anticipate. They anticipate that the reserve margin is sufficiently high that the vertical part of the supply stack always falls to the right of the inelastic demand line. In that case, the electricity price is seen to be equal to the cost of the marginal plant. The marginal plant sets the price, which is low and reasonable and reflects the operation of the marginal plant. This price has been called the system lambda, evolving from the age-old notion that traditional dispatch models minimize dispatch cost and this price is the Lagrange multiplier on the cost minimization problem. The other case, however, is the case in which the demand for power would be to the right of the vertical line. This case, termed Case 2, is also depicted at the right in Figure 110. Case 2 occurs when there is not enough capacity in place using each and every plant in the system to meet the demand in a particular hour. This case frequently occurs in the real world at time of peak, notwithstanding the caterwauls of dispatch modelers, particularly if there are outages that affect the supply stack at time of peak. What is the price in that situation? The answer is that there is no price. The dispatch construct with a fixed load is just plain wrong. IEMM does not want to be based on that type of assumption or construct. We are going to see the proper construct in the next diagram.

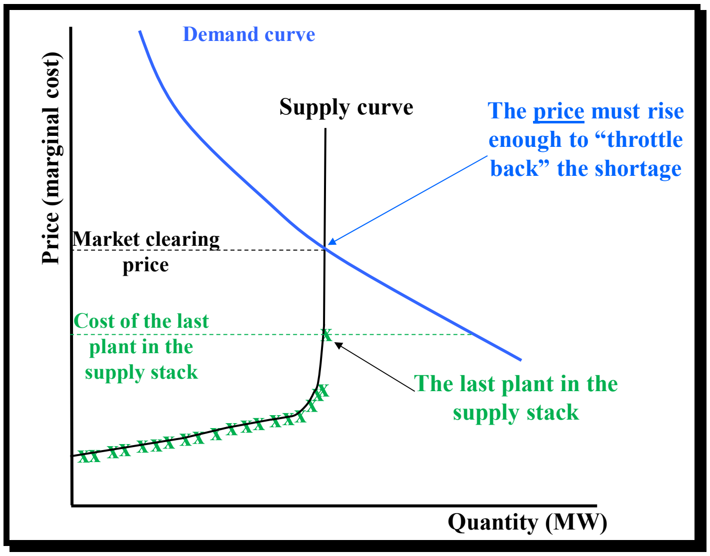
Figure 110: The "Supply Stack" of All Plants in Ascending Order of Cost with Demand



If we think a little bit about it, we realize that there is no such thing as an inelastic electricity (or any other) demand curve. The very notion, which lies at the heart of dispatch models, is incorrect. There are only elastic demand curves. When we consider that the only respectable supply curves in any economic setting are elastic supply curves, we get the insightful solution in Figure 111, which is the solution embedded in this design of IEMM.

There is no way we know of to get this type of plant operation and dispatch, complete with reserve requirement, out of a complementarity, MPEC, or linear programming approach. To implement a dispatch model based on minimization of variable cost or minimization of any cost would be mortal for IEMM. The situation in Figure 111 occurs all the time in real world electrical systems, and it is one of the things that triggers endogenous capacity addition. Dispatch based on cost minimization do not describe the real world and are problematic with regard to EIA modeling how power systems and prices truly work.

Figure 111: Economic Reality—Price Has No Relationship Whatsoever to the Cost of ANY Plant



6 PLANT PROFITABILTIY AND CAPACITY ADDITION[18]

The aggregate capacity in place and the price of energy are a strong function of capacity addition into the market. There is no question about it, the original EIA solicitation recognizes it, and standard dispatch approaches have ignored or guessed it offline and ported the offline guess into their model. The aggregate quantity of pollution, and thus the aggregate demand for emissions allowances, is a strong function of new plant entry as well as of retrofitting of old plants with pollution control equipment, retirement of old plants, and fuel management in old and new plants. Capital formation in the generation sector is one of the chief agents not only of power costs but also environmental cleanup and renewables entry. For example, new greenfield generation capacity enters the market; embeds new, low emission capacity in the existing fleet; operates; and thereby alters the individual and aggregate demand for emissions allowances (i.e., supply of emissions) throughout the fleet, and erases the profit incentives for certain old plants to continue operation. (This is what is being talked about, for example, for China—capital stock rollover.) New plants almost always have lower emissions coefficients, and new plants run many more hours than old plants. This combined effect—lower emissions coefficients and more hours of operation—reduces the aggregate production of emissions and thus the aggregate demand for emissions allowances (i.e., output of emissions), all else equal. (We have seen precisely this phenomenon in the rollover of the automobile fleet since the 1960s. New cars pollute much less than old cars, and the capital stock rollover of the automobile fleet embeds more and more low emission new cars over time. IEMM will represent that.) This IEMM design allows EIA to represent all of that, and throughout the entire electricity systems that IEMM must model.

IEMM uses the preceding characterization of the plant or a plant component such as a retrofit in a profitability calculation (a calculation precisely akin to how plant profitabilities or project profitabilities are calculated in practice) as summarized in Figure 112 (a diagram from the original Gulf-SRI model in 1973-4 and nothing more than a graphical representation of producer profitability as originally put forth by Cazalet in 1977 and expanded by Nesbitt in 1984.)[19] As indicated in the figure, the IEMM profitability takes account of power price, fuel price, all the emissions prices for every emission that must be offset by the plant through surrender of emissions allowances under the law, and the renewable REC price under the applicable law. If one has prices for the various emissions allowances (and those prices can be calculated endogenously in IEMM as discussed previously) and the REC price, the process of calculating economic costs and calculating emissions costs and REC costs and layering them together is straightforward. It can distinguish and carefully consider environmental as well as economic costs, and it has the virtue that it internalizes environmental costs in the operational decisions the plant makes. To support this calculation in the IEMM model, one needs to assemble the data for every power plant in North America individually and in the aggregate to make the calculations outlined herein.

It is insightful to depict the profitability calculation in Figure 112 conceptually in the diagram in Figure 113, which emphasizes the central importance of market prices to any profitability

[18] The earliest published antecedent of this section was by Edward G. Cazalet, **Generalized Equilibrium Modeling--The Methodology of the SRI-Gulf Model**, Stanford Research Institute, 1977.

[19] Because of the temporality of electric power, we need to generalize the diagram to represent the temporality. Nonetheless, it is very useful to put forth the salient concepts even without the temporal generalization.

calculation that occurs anywhere at any time in an economy. What IEMM does is calculate the green area of the diagram in Figure 113 both annually and in a present value sense. The green area of the diagram is the margin captured—price minus variable cost—in each and every future hour, day, month, and year of future plant operation. Obviously, the key drivers of the green area are the forward power price and the forward gas price for a gas plant and the same forward power price but the forward coal price in place of the forward gas price for a coal plant. This green area drives plant investment, operation, and retirement, for it is the intrinsic profitability of the plant. We have a detailed explication of the proper profitability measure to use, but this section enhances that discussion in the context of electric generation and transmission.

Figure 112: Plant Profitability Calculation Considers Power, Gas, Emissions, RECs, Fixed, and Variable Costs

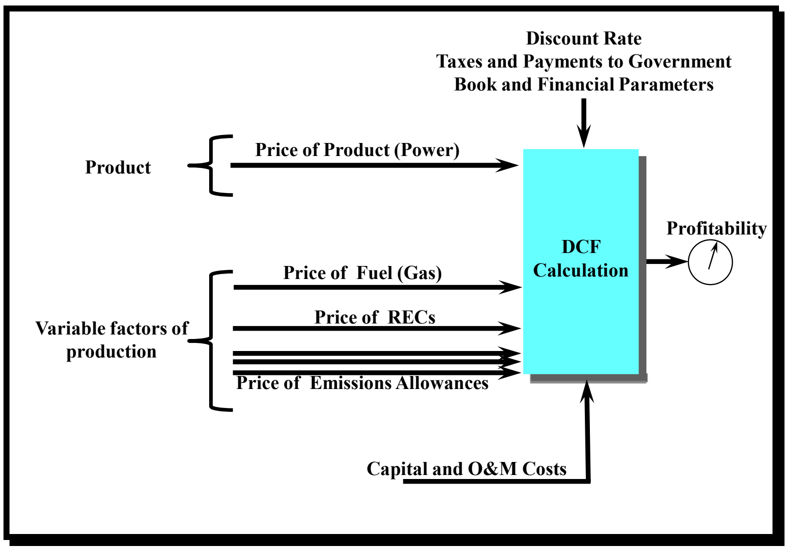
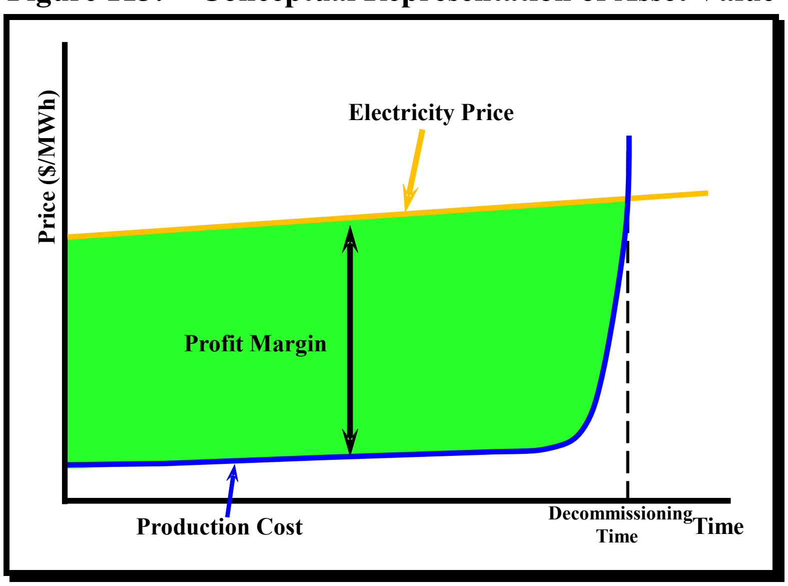


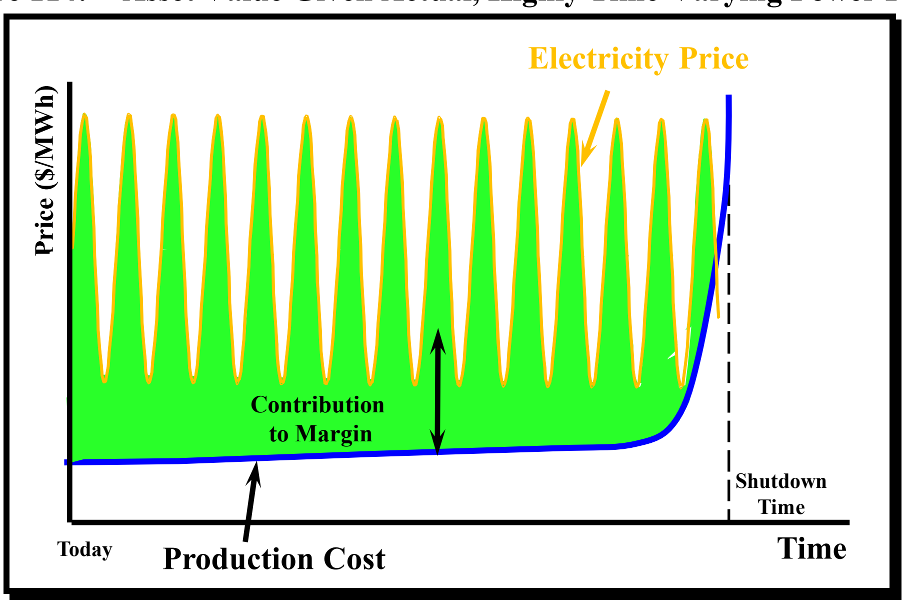
Figure 113: Conceptual Representation of Asset Value



This green area is so fundamental to the real world and to the IEMM model that it merits additional emphasis here. For every existing and possible node (every existing and possible plant in North America), the IEMM model sets up the diagram in Figure 113 and makes the indicated calculation of the net present value of profitability (including capital cost, not shown). If that net present value minus the capital cost is zero or less, no new plant will enter the market. If that net present value minus capital cost is positive and exceeds the capital cost of a new plant that could enter the market, that new plant can be expected to enter the market because the returns it will get at its rate of discount will exceed its capital cost of entry.

The reality is not exactly that in Figure 113, which is conceptual and not properly dynamic. The reality is that power price varies by minute, hour, day, week, month, and year. We have in fact a price for every tranche of load developed in the demand section above. The IEMM model power generation nodes view the conceptual diagram in Figure 113 as the more dynamic and volatile diagram in Figure 114 with regard to power (and fuel, not shown). In addition, not shown in the diagram, natural gas and other fuel prices are also volatile, and their volatility is not necessarily temporally coincident with power volatility. It is the profitability represented by the diagram in Figure 114 that IEMM uses to represent prospects for new plant entry, prospects for existing plant operation, and in fact by adjusting the blue curve in the diagram to internalize the impact of environmental regulation on plant owner decision makers. The IEMM model as defined here must calculate the diagram in Figure 114 for every plant (or plant aggregate) within IEMM—every existing and possible power plant in North America or whatever system is being modeled. It calculates the profitability of every existing and possible plant in IEMM using the basic concept in Figure 114 but augmented for emissions and renewables as discussed in previous sections. Figure 114 is a dynamic, graphical representation of the present value of profitability calculation so widely published.

Figure 114: Asset Value Given Actual, Highly Time-Varying Power Prices



Equally importantly as additions of new, greenfield capacity, there exists a wide range of possible retrofitting techniques for existing plants with a wide possible range of pollution control equipment. All such possible plant retrofits entail capital cost, operating cost, and heat rate impacts on the plant, but those are quantifiable. There is a similar philosophy to that in Figure 113, namely is the retrofit profitable in light of the costs or bans on various plant configurations or operations.

This section summarizes how IEMM considers entry of new capacity, taking full and explicit account of capital as well as operating costs (including fuel costs and heat rates as well as environmental emissions allowance costs and REC costs). Consider a plant that can possibly be added at a given point in time (denoted t_0), which was conceived as in Figure 114. It is easiest to think of the plant in Figure 114 as a natural gas combined-cycle plant, but it would be just as appropriate to think of a coal plant, nuclear plant, transmission line, gas pipeline, refinery, or coal gasification plant. The plant will incur a capital cost of construction (not shown in the diagram), and it will operate and sell its product into a market at a market-determined price indicated by the electricity curve at the top of the diagram. The plant will, if built, acquire its input (gas) at a market-determined price for that commodity indicated by the blue curve at the bottom of the diagram. It will also buy all the environmental emissions allowances (if any) and any RECs required. For simplicity of exposition, we ignore in the diagram the fact that the product price—i.e., the power price—varies hour to hour and therefore is very jagged in shape. The plant will last for some period into the future as indicated in the figure. Specifically, as the plant ages and ultimately obsoletes, its operating cost will escalate and ultimately skyrocket. That is the time at which either an extensive rehabilitation, perhaps approximately equal in magnitude to the entire capital cost of entry, must occur or else the plant will be retired because it is no longer price-competitive. During the life of the plant, it will generate a schedule of margin capture equal in magnitude to the time-varying green shaded area in the model, **which certainly is not levelized but rather is specifically market-determined.** This diagram is quite analogous to the one in Figure 114.

We want to associate some simple algebra with Figure 114. We adopt the following notation:

$p_k(t)$ = market price of the product at time t in the hour tranche k (load period k).

$\phi_k(t_0,t)$ = variable operating cost at time t in the hour tranche k (load period k) of the plant built at time t_0, including fuel cost, all environmental emissions costs, and REC cost.

$q_k(t_0,t)$ = production from the plant at time t in the hour tranche k (load period k) given that the plant built at time t_0.

r = discount rate.

Tx = income tax rate.

CapCost(t_0) = capital cost of the plant that opens at time t_0, appropriately adjusted for all taxation and other effects. CapCost is not merely the raw cost expressed in $/MW, but it contains all the tax and fiscal adjustments.

The present value of margin capture after income taxes during the forward life of the facility, i.e., the present value of the irregular green area in the diagram, is

$$\left(1-Tx\right)\sum_{t=t_0}^{\infty}\left(\frac{1}{1+r}\right)^{t-t_0}\sum_{k=1}^{S}\left[p_k(t)-\phi_k(t_0,t)\right]q_k(t_0,t).$$

This equation calculates the number of after-tax profit dollars the plant will generate over its full future life cycle. It is the present value of the green area in Figure 114, the present value of margin capture by the plant owner. If this number of after-tax dollars is larger than the after-tax capital cost of the facility, someone in the market would be expected to make the investment because the investment will pay a rate of return at least as high as the specified market rate of return as embedded in the assumed discount rate. If, on the other hand, the number of after-tax dollars is lower than the capital cost of the facility, no plant owner would make the investment because it would pay a rate of return strictly less than the specified discount rate. In a competitive, atomistic market, entry will occur up to the point at which the present value of margin capture in the preceding formula is precisely and exactly equal to the present value of the capital-related costs of the plant. Therefore, the condition will be that market entry will occur up to the point at which further entry is not profitable, i.e., up to the point at which

$$\text{CapCost}\left(t\right)=\left(1-Tx\right)\sum_{t=t_0}^{\infty}\left(\frac{1}{1+r}\right)^{t-t_0}\sum_{k=1}^{S}\left[p_k(t)-\phi_k(t_0,t)\right]q_k(t_0,t).$$

Note the structure of the preceding equation. It is a present value difference equation forward in time beginning with the opening time point t_0 and proceeding forward from that point onward. If we write the margin requirement at the opening time t_0 as a function of future margin requirements, the preceding equation becomes

$$\sum_{k=1}^{S}\left[p_k(t_0)-\phi_k(t_0,t_0)\right]q_k(t_0,t_0)=\frac{\text{CapCost}\left(t\right)}{\left(1-Tx\right)}-\sum_{t=t_0+1}^{\infty}\left(\frac{1}{1+r}\right)^{t-t_0}\sum_{k=1}^{S}\left[p_k(t)-\phi_k(t_0,t)\right]q_k(t_0,t).$$

This equation relates the price in the opening year t_0 with the time-schedule of prices and operating costs in all subsequent future years t_0+1, t_0+2, t_0+3, and so on. The equation is a direct application of the Cazalet 1997 and Nesbitt 1984 formulas.

The equation has an important interpretation. The margin capture in the opening year t_0 is equal to the total capital cost minus the margin capture in the second and subsequent years. People have improperly interpreted this as the “capital charge factor” or “capital recovery factor” and have levelized or annuitized it even though no real world market ever makes the slightest attempt to levelize anything. IEMM must specifically use the complex, intertemporal difference equation without approximation or levelizing. IEMM should use the fully dynamic specification of this difference equation, thereby de facto using a market determined rate of “depreciation.” Depreciation is critical to projecting forward prices in markets that allow entry, for the forward price is an explicit and systematic function of market-determined capital asset depreciation. At the same time, market-determined capital asset depreciation is an explicitly and systematic function of market prices. Price and depreciation are coupled; they are simultaneously determined. One cannot be quantified without the other, and the governing equation is the preceding difference equation. The specific mantra that governs capacity additions is this:

Today's capacity additions are a function of tomorrow's price,

and

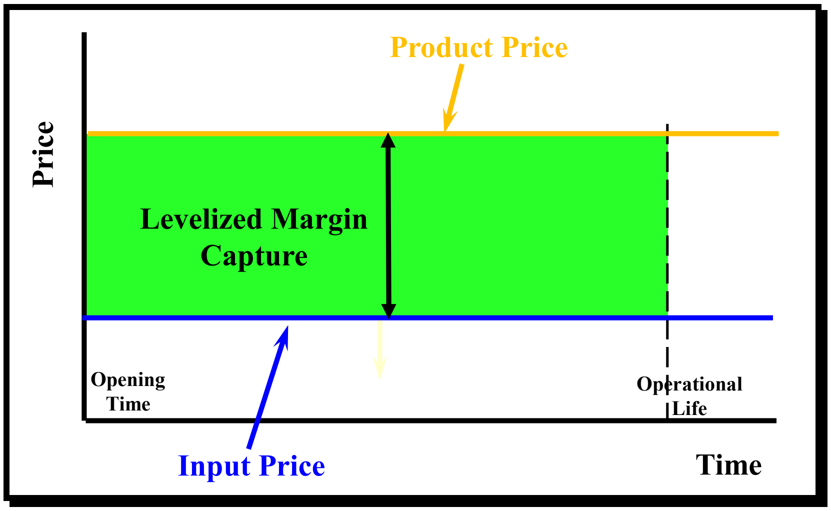
Tomorrow's price is a function of today's capacity additions.

The time loop is fully closed within IEMM, just as it is in the real world. Capital entry and temporality of price are fully consistent and fully closed and fully rational in the words of the Nobel Laureate Richard Lucas. (Lucas conceived the approach known as "rational expectations," which we propose that IEMM adopt. The approach articulated and recommended here is a rational expectations approach. There is no other approach even vaguely as acceptable.)

Levelizing or annuitization of capital recovery, i.e., levelizing of the green area in Figure 113 and Figure 114 is a mortal mistake. Most electricity models do it never the less. IEMM as recommended here will not. Models that purport to represent market entry sometimes mistakenly interpret the green area in Figure 113 and Figure 114 and the preceding price equation by arguing that "price equals O&M cost plus capital charge factor," and they assume that the capital charge factor is annuitized or levelized across the life of the plant in some prespecified way. It is as if a deity emerged from the firmament and deemed that all capital shall be recovered in level, equal, temporally predictable increments or annual installments. Would that a modeler's life or the behavior of a market were so easy! There is not a market in existence that has ever been levelized in any fashion whatsoever except perhaps serendipitously or ephemerally. To make a levelizing assumption in a purported market model is specious at best. It will ruin the ability of IEMM to project capacity additions, which is crucial to the success of IEMM.

It is important to emphasize this point again and again and again—typical models other than IEMM (and a lot of literature) assume that the green area in Figure 113 and Figure 114 is constant or "level" over the life of the plant, and, therefore, by assumption, that capital asset depreciation is constant over the assumed lifetime of the asset. In doing so, they assume that the analytically correct dynamic diagram in Figure 113 and Figure 114 can be approximated by the levelized capital recovery diagram in Figure 115. It is easy to see why this is incorrect, and why this IEMM CDR does not do so and cautions EIA against doing so. The latter diagram indicates what it means to assume a levelized capital recovery schedule, i.e., a constant capital asset depreciation schedule. Levelized capital recovery is tantamount to assuming a levelized difference between the price of the product (power) and the price of the input factor (gas). It is also tantamount to another simple but patently incorrect idea embedded in other models: *"If the market prices would pay an annuitized capital charge, there will be entry. Otherwise, there will be no entry."* Under such an assumption, which is dead wrong and bordering on laughable in real world markets, entry would be intrinsically tied to a **levelized** capital charge, which is ridiculous. No market levelizes or guarantees anything. There is strong empirical evidence to the contrary—every time there is a price change in fact.

Figure 115: Levelized Capital Recovery Assumption, Which Is Incorrect



What is levelized capital recovery truly assuming? Clearly, it is assuming that the **difference** between power price and fuel price must and will be flat and level in real terms to induce entry. The entering entity expects and receives a flat difference between electricity price and fuel price. It biases the calculation of price, and it biases the prediction of capital entry.

Let us further explore the veracity of this type of modeling approach. Could a levelizing approach such as that in Figure 115 realistically be expected give the right answer in a price-backwardated world? Not a chance, as indicated in Figure 116. Would a levelizing approach such as that in Figure 115 realistically be expected to give the right answer in a contango price world? Not a chance, as indicated in Figure 117. Could levelizing be correct if fuel prices were backwardated or contango? Not a ghost of a chance. Would fuel prices interact with capital recovery? Not at all under this levelizing assumption, but they would in IEMM, which does not make this unrealistic levelizing assumption. Do real-world market agents such as Intel, Duke Energy, Microsoft, Dow Chemical, American Airlines, Applied Materials, or USX ever in their wildest dreams receive "levelized" capital recoveries and do they make decisions on that basis? No. Are forward fuel prices or forward power prices ever level? Virtually never. It is well to consider that models that use levelized capital recoveries (if they allow entry at all) are largely worthless. They report capacity entry schedules, but those schedules are guesses. They report prices, but those prices are therefore guesses. IEMM calculates the full forward price schedule and never a levelized approximation to that schedule.

This recommended IEMM design assumes fully dynamic capital recovery, i.e., market-determined depreciation, as is implicit in Figure 113 and Figure 114. In IEMM, depreciation cannot be simplistically annuitized or subjected to any exogenously given time schedule; it must be market derived and must be very sensitive to contango, backwardation, and all other price dynamics both on the fuel side and on the power side. This view is critically important in predicting forward price as a function of capital asset entry and simultaneously capital asset entry as a function of forward price.

Figure 116: Levelized Margin Capture Assumption Fails under Backwardated Product Price

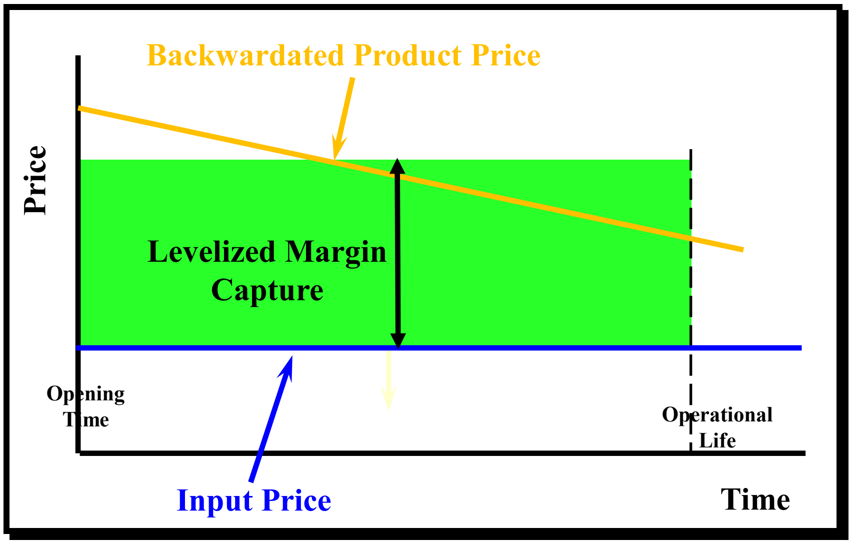
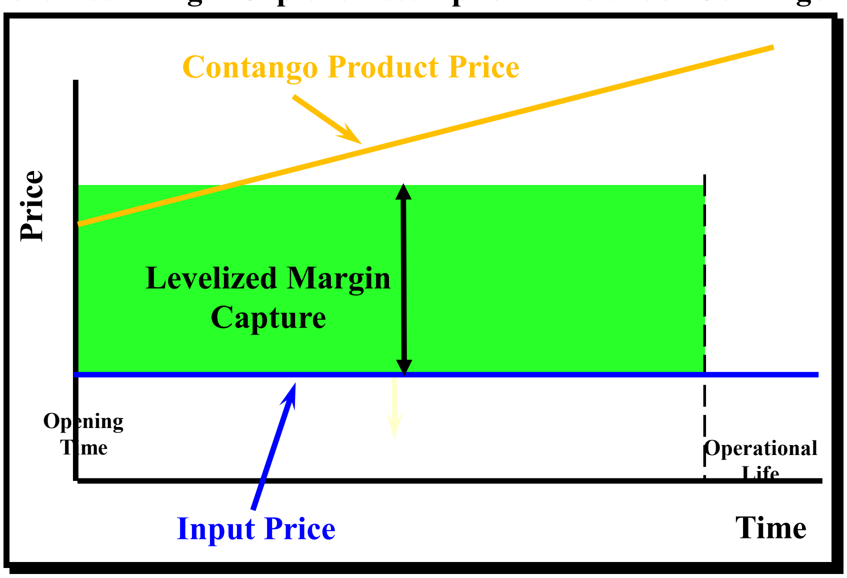
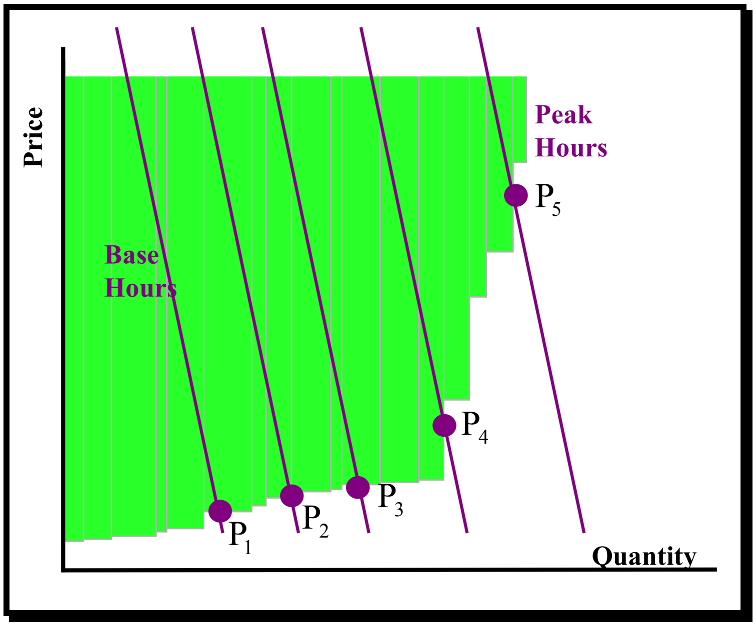


Figure 117: Levelized Margin Capture Assumption Fails under Contango Product Price



How does IEMM (as defined in this CDR) know how much new capacity to add to a given supply stack in a given nodal/local region as previously discussed? Magnitude of capacity entry is one of the toughest modeling problems, yet IEMM (as recommended here) addresses it. We summarize how by returning to the supply stack of existing plants that is in place in a given year. Each element of the supply stack, because it already exists, will run or not run based on whether its variable cost is above or below the market price. The existing fleet of plants is represented using the green supply stack in Figure 118. The green supply stack is the analogous environmentally adjusted (green) supply stack summarized in Figure 28 and Figure 29.

Figure 118: Base through Peak Market Clearing Prices and Quantities in the Absence of Capacity Addition



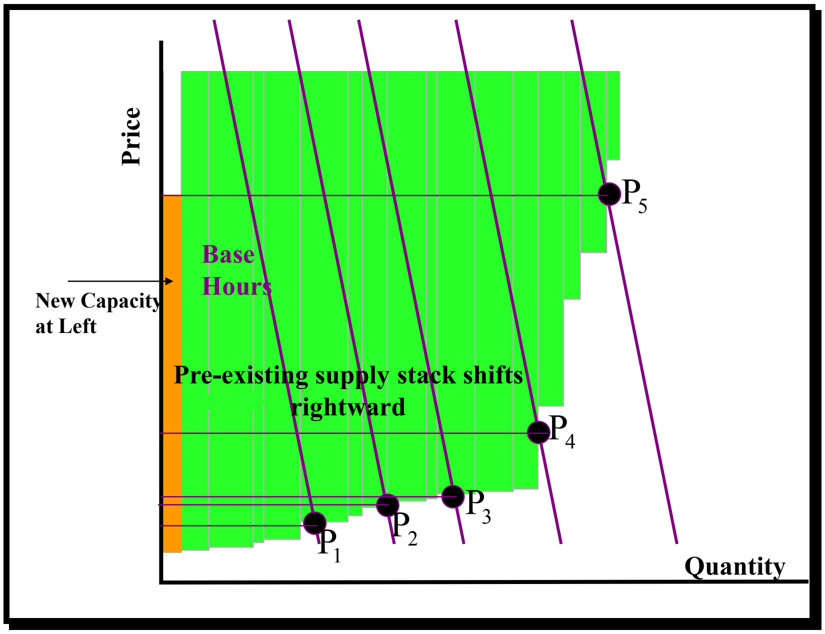
To quantify the time variation in demand that affects electric power generation, we postulate for simplicity five intervals of load ranging from base through intermediate to peak.[20] Figure 118 indicates the five market clearing prices and the five corresponding quantities that would occur at those five market clearing prices. The prices are arrayed from the peak price at the top right of the diagram to the base price at the bottom of the diagram. We have designated the prices that would occur in the market in Figure 118 as P_1, P_2, P_3, P_4, P_5 ranging from deepest baseload at the left all the way to highest peak load at the right.

Consider what would happen if someone were to install a very tiny increment of new capacity to this supply stack, as shown in Figure 119. In general, this new capacity would create a new, short-term supply curve pressed against the leftmost axis in the diagram, the lowest cost portion of the supply stack. (In the real world, new entrants typically have lower variable cost than existing plants, but this does not have to be the case either in IEMM or in the real world or the supply stack here. It is just more convenient to graph and understand if we assume that.) Figure 119 adds a small amount of capacity to the left-hand side of the original supply stack in Figure 118. The new capacity is pressed against the left axis because newer plants tend to be more efficient and have lower variable operating costs than the older plants in the mix. (This need not be the case, but it is a lot easier to draw and illustrate the important concepts by assuming that the new capacity enters against the left axis.) The profitability of this new capacity C would be given by the following equation:

[20] The specific nature of load is not important here as long as we distinguish several tranches of demand. The argument generalizes no matter how detailed a representation of demand we elect to carry. We could just as well have considered each and every one of the 8760 hours of load, but that would complicate the graph and the discussion unnecessarily.

$$C\left[(P_1-\phi_1)h_1+(P_2-\phi_2)h_2+(P_3-\phi_3)h_3+(P_4-\phi_4)h_4+(P_5-\phi_5)h_5\right].$$

Figure 119: Adding a Small Amount of New Capacity



The notation h_i represents the number of hours that exist in the i^{th} load tranche. The capacity would be represented by the width of the leftmost stairstep in the supply stack, very narrow in the diagram in Figure 119 because we are considering a minuscule amount of new capacity C. The margin captured by this new plant supply increment would be the vertical height of the orange rectangle at time of peak (tranche 5), i.e., the difference between the market price and the operating cost of the new plant increment. It would be a bit lower during tranches 1,2,3, and 4 because the prices would be lower than the peak price. Furthermore, the margin capture would be weighted by the number of hours that the new plant ran in base, intermediate, and peak time tranches. The profitability of the newly entering plant is the **area** of the orange rectangles at the left, i.e., the product of the magnitude of capacity added times the margin capture per unit of capacity. Because the magnitude of capacity C is so infinitesimally small in Figure 119, the prices have not dropped at all from their original settings in Figure 118.

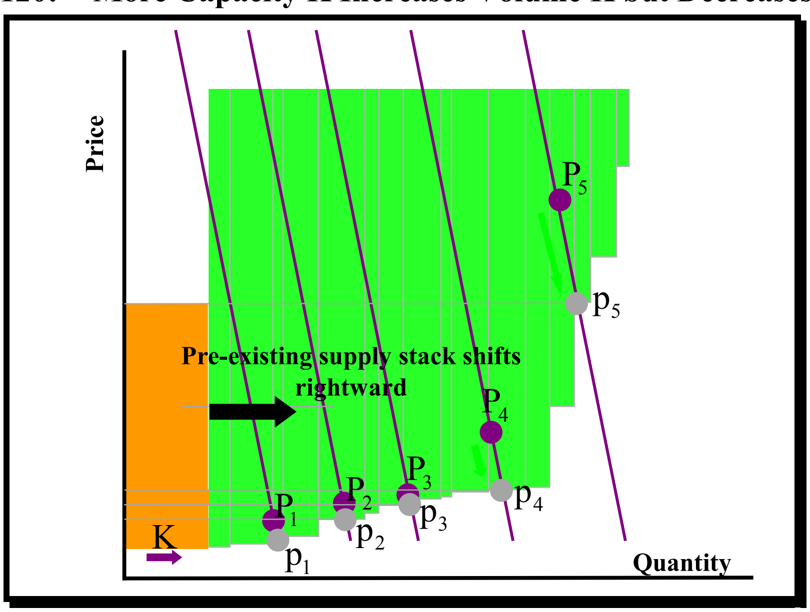
As the quantity of new capacity is increased from zero 0 as shown in Figure 119, the entire pre-existing, green supply stack is displaced outward and to the right by exactly the magnitude of new capacity that is added. The entry of new capacity shifts the previously existing supply stack directly horizontally to the right. The new capacity that is added generates margin capture for its owner that is indicated by the orange shaded areas at the left, each area being determined by the number of hours the new capacity runs in each time tranche. During time of peak, it generates margin equal to the difference between the peak price P_5 and the variable operating cost. During time of intermediate load, it generates margin equal to the difference between the intermediate price and the variable operating cost. During time of base, it generates margin equal to the difference between the base price and the variable operating cost. Each tranche of margin capture must be weighted by the number of hours that the particular tranche of demand occurs.

Now suppose we add a significant magnitude of new capacity K as shown in Figure 120 a magnitude of new capacity that can depress market prices because of its size, precisely as shown in the diagram. The figure indicates that the peak price is depressed from the purple level P_5 to the gray level p_5 as the supply stack is displaced outward and to the right by the entry of new capacity. All the other prices are depressed as the market equilibrium points simultaneously slide down the various demand curves from the purple points to the gray points in the diagram. The magnitude of price softening is, of course, highest at time of peak, but it is not necessarily insignificant or zero at time of intermediate or base, as the diagram indicates. Notice that the margin capture by the new capacity K that has entered has widened because the quantity of new capacity entry has grown, but the margins are not as high because of the price depression that has occurred because of the new entry. The new entry K increases the quantity term K but decreases the margin capture in each of the five-time tranches. The aggregate level of margin capture is equal to the hourly weighted average of the various colored areas in the leftmost supply curve in the figure. They are clearly larger in magnitude than they were in Figure 119. The new margin capture equation is the following:

$$K\left[(p_1-\phi_1)h_1+(p_2-\phi_2)h_2+(p_3-\phi_3)h_3+(p_4-\phi_4)h_4+(p_5-\phi_5)h_5\right].$$

The K term is much larger than it was before in the context of Figure 119, and the margin capture term is a bit smaller than it was before in the context of Figure 119. The product of the K term and the margin capture term is clearly larger than it was before in the context of in the context of Figure 119.

Figure 120: More Capacity K Increases Volume K but Decreases Prices

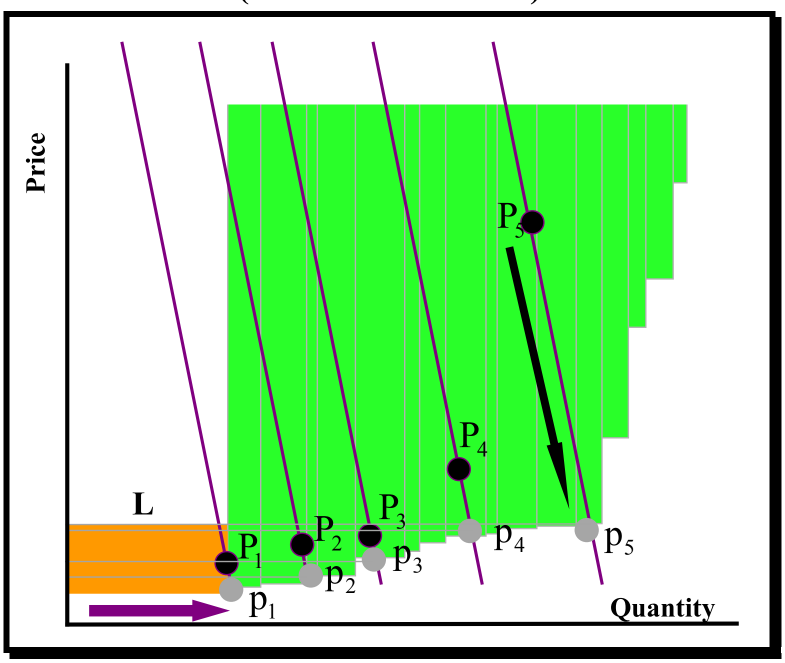


What would happen if one added even more capacity, a huge amount of capacity L in fact? The situation would appear as in Figure 121. The width L of the capacity addition tranche at the left would increase, but the prices would be “crashed” down all the way toward variable operating cost. The margin per MWh would become infinitesimal, but the quantity L of MWh would become very large. The hourly weighted sum of the shaded margin capture areas would become negligibly small as presaged in the diagram. The market would be overbuilt, and margins would crash precipitously. The relevant margin equation would be:

$$L\left[(p_1-\phi_1)h_1+(p_2-\phi_2)h_2+(p_3-\phi_3)h_3+(p_4-\phi_4)h_4+(p_5-\phi_5)h_5\right].$$

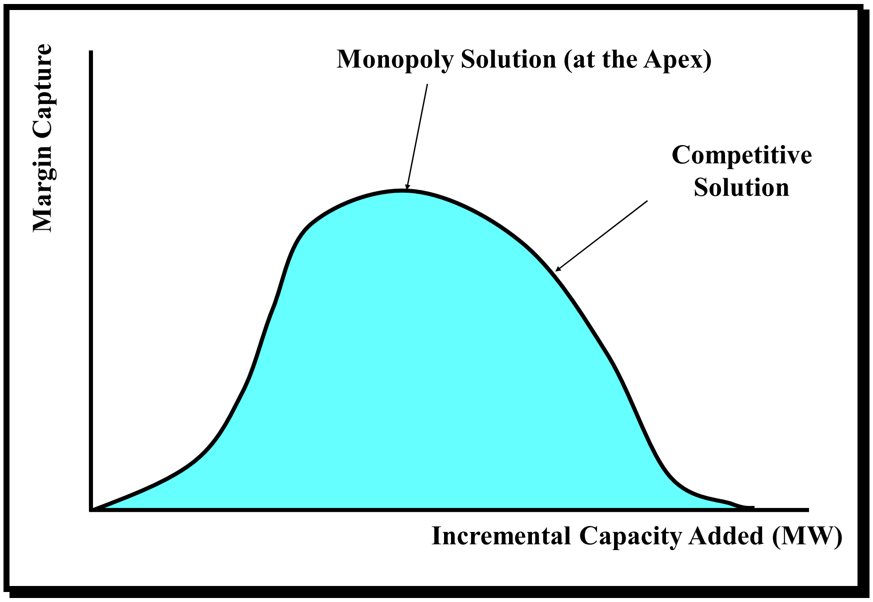
The term L would be very large, but there would be no difference between the price and the variable operating cost in any time tranche so the margin term on the right in braces would trend toward zero. There would be a lot of volume L, but there would be no margin in any time tranche. Clearly margin capture would be trending toward zero because there is no margin in the market.

Figure 121: Still More Capacity Addition Would Kill the Peak Price (and All Other Prices)



It is clear from the preceding diagrams that as the magnitude of capacity additions increases, the margin capture from the new capacity initially increases from its initial value of zero, passes through a maximum, and ultimately falls to a final value of zero for infinite capacity addition (which completely kills the market by driving all prices down to variable cost). The profit that results from new capacity entry takes the form of a "Laffer Curve" as in Figure 122 that relates margin capture to the magnitude of capacity additions. To see why there is a "Laffer Curve," notice that margin capture is zero if zero capacity is added because there is no quantity (the C case previously), and margin capture is zero if infinite capacity is added because there is no margin (prices have fallen all the way down to variable operating cost of the entering unit, the L case previously). Margin capture is clearly positive for middle levels of capacity additions and passes through at least one maximum between the two limits. IEMM must calculate this Laffer curve and predict market entry based on margin capture along the Laffer curve. In particular, the market adds capacity along this curve up to but not beyond the point at which margin capture is equal to or exceeds capital cost of entry on a discounted present value basis.

Figure 122: Margin Capture Is a "Laffer Curve" with Regard to Capacity Additions.



In a pure monopoly, the industry would add capacity equal in magnitude to the peak margin capture, taking full account of all price erosion they could cause though entry or alternative all price firming they could cause by withholding entry or operation if they had sufficient market concentration. Make no mistake; it would take a monopoly for the node to operate at the maximum point on the curve. It would take massive market concentration for this to be able to occur at all, and the requisite degree of market concentration never occurs. This is precisely what monopoly theory predicts. In a competitive market, the industry is not capable of achieving the absolute maximum. The industry would add capacity beyond the maximum because there will be positive aggregate profitability from so doing. The electric power generation market is not thought to be a monopoly or oligopoly market. Most markets are not monopolistic or oligopolistic.[21] Entry is atomistic and ubiquitously available; anyone can enter with literally the same technology. Anyone can build a power plant. There is no credible empirical evidence that electric generation has or exercises market power in industrialized countries.[22] The IEMM methodology we have here is agnostic whether there might be market power and works perfectly well under such an assumption. However, it also works perfectly well under competitive entry. There is no demonstrable "delta p over delta q" term that would allow entrants to operate strictly at the absolute peak of the Laffer curve. Individual generators cannot unilaterally drive up prices by withholding capacity. Keep in mind, exercise of market power is illegal in most places in the world. FERC conducts periodic reporting exercises to guard against it. It is not illegal to **have** market power; it is illegal to **exercise** it. There have been few if any successful prosecutions in the energy business regarding the exercise

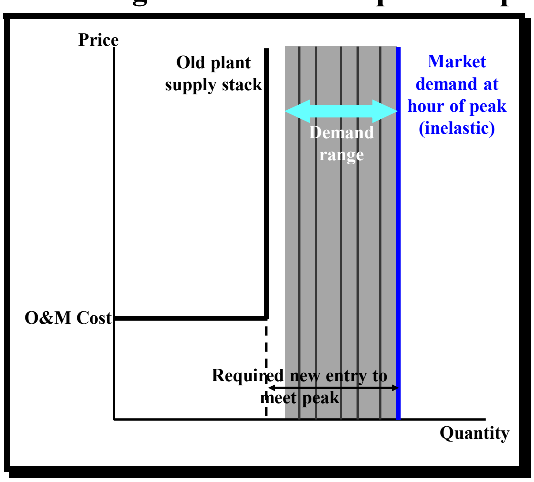
[21] If there are "granted" franchise monopolies, they are regulated, meaning that there is no prospect of entrants exploiting market power.
[22] There is no shortage of political argument or demagoguery. However, as we now discuss, such assertions are difficult to justify.

of market power. The idea that players have and exercise market power is pretty specious. There is little or no evidence that market power has ever been exercised in the electricity business. IEMM (as defined here) can represent both market power and competitive industries.

Greenfield entry is important in the sense that the rate and magnitude of capital stock rollover and new equipment entry has characteristically been underestimated. IEMM (as defined here) can identify specific greenfield entry opportunities and place new plants or transmission links into those places where profitable and where individual agents would be likely in the real world to build them. When that occurs, the prices that occur in the market are **strongly** a function of the capital cost of entry. It is not appropriate to assume that entry occurs because of price. That is only half the story. Price occurs because of entry. The loop between price and entry is closed, not open. IEMM can be equipped with estimates of the capital cost of entry by various plant types in all of the nodal regions that compose IEMM.

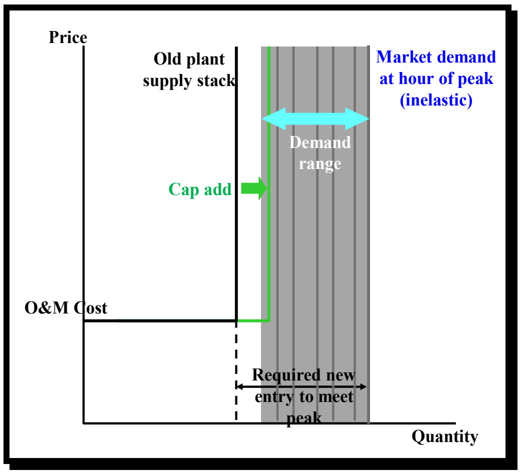
We conclude this section with a careful understanding of capacity entry and peak price in a market setting. Suppose that we are in a growing market (growing in the sense that peak load minus existing capacity is positive, which occurs in markets where load is not growing but retirements are occurring as well as demand growing). The situation in Figure 123 illustrates. We craft the supply stack from old (existing) plants in the supply stack in the black curve at the left. We see the gray area, the left side of which is the baseload energy and the right side which is the peak energy. The left side of the gray area is the deepest baseload, and the right side depicted by the solid blue line is the highest peak load. There is not enough capacity in place to satisfy the base or the peak or levels between the extremes. Clearly capacity is going to have to be added. The width of the gray area is the variation from deepest base to highest peak. We see at the bottom how much capacity would have to be added if the system is to meet the peak load and all loads beneath the peak.

Figure 123: A Growing Market that Requires Capacity Additions



Suppose we had a very small magnitude of entry, enough to get total installed capacity higher than the lowest baseload hours. That situation appears as in Figure 124. We see that the original black curve has moved rightward to become the green curve. Enough capacity has been added so that now there is sufficient capacity to meet baseload energy demand. For every hour in which the demand is less than the green curve, the price is the cost of the black supply curve—variable cost of the capacity added. There is no profitability during any of those hours. For every point to the left of the green curve, there is excess capacity in place. For every point to the right of the green curve, there is insufficient capacity in place, and the price can be infinitely high. The key is that the price in this market will be the height of the black supply/green curve out to the amount of capacity that is added. Beyond that, the price goes to infinity. Thus, there is very strong incentive for the market to keep adding. The price that new capacity will return is tremendously high.

Figure 124: A Little Bit of Entry



Suppose we considered even more entry, enough to get total installed capacity into the middle intermediate hours. That situation appears in Figure 125. We see that the original black supply curve has moved rightward into the middle intermediate to become the green curve. Enough capacity has been added so that now there is sufficient capacity to meet intermediate energy demand. For every hour in which the demand is below the green curve, the price is the cost of the black supply curve—variable cost of the capacity added. There is no profitability for any hour below the capacity that was added. For every point to the left of the green curve, there is excess capacity in place. There are a lot more hours here for which there is excess capacity in place. For every point to the right of the green curve, there is insufficient capacity in place, and the price is infinitely high. There is a narrowing amount of energy in that tranche to the right of the green curve. Capacity is being added so that the area to the right is shrinking. The key is that the price in this market will be the height of the black supply/green curve out to the amount of capacity that is added. Beyond that, the price goes to infinity. Thus, there is strong incentive for the market to keep adding. The price that new capacity will return is tremendously high.

Suppose we had a lot of entry, right up to but not including the last hour of peak. We add enough capacity to meet demand for the first N-1 hours, but not the Nth and final hour. That situation appears as in Figure 126. We see that the original black supply curve has moved rightward almost out to the end of the gray range, i.e., almost to the blue curve. Enough capacity has been added so that now there is sufficient capacity to meet every hour of demand save the single peak hour. For every hour in which the demand is below the green curve, the price is the cost of the black supply curve—variable cost of the capacity added. There is no profitability for any hour below the capacity that was added. There is excess supply at every hour save the peak hour. For every point to the left of the green curve, there is excess capacity in place. There are clearly a lot more hours here for which there is excess capacity in place. For every point to the right of the green curve, and there is only one hour left, namely the peak hour, there is insufficient capacity in place. Would the price be infinitely high? No, and we will see why.

Figure 125: A Moderate Amount of Capacity Addition

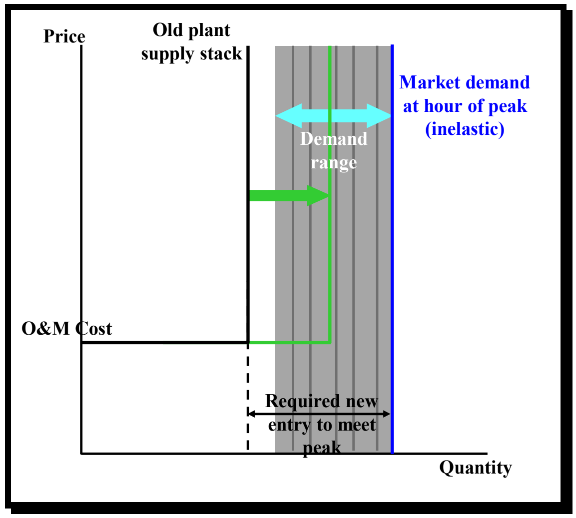
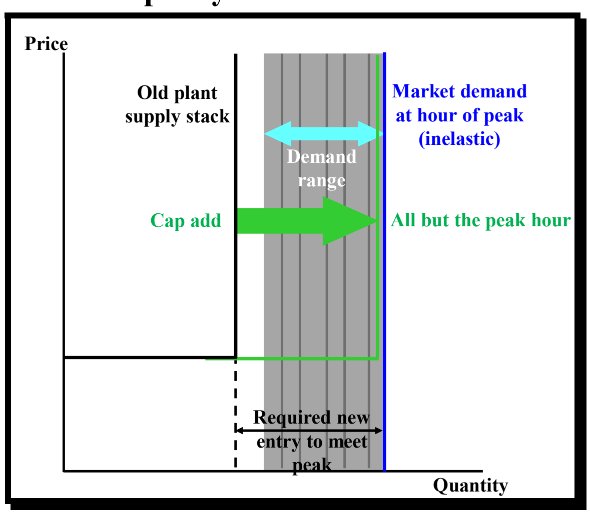


Figure 126: Capacity Addition for N-1 of the N Hours

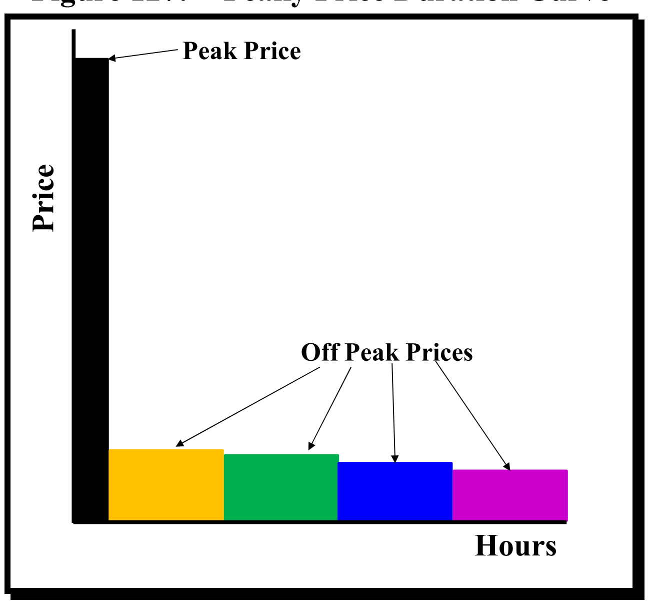


How high would the price in the peak hour have to be to attract the last MW of capacity so that you could meet peak capacity? It would have to be the full capital cost of the plant. And what would be the price of all N-1 hours below the peak. It would be the height of the black supply/green curve. There would be zero profit and zero return to capital during every one of the N-1 nonpeak hours. Zero profit in every hour other than the peak hour because there is excess capacity in place in every hour save for the peak hour. There would be only one profitable hour, and it would be the peak hour. To entice entry, the price in the peak hour would have to pay the full capital cost of entry. The price in all N-1 hours below the peak would be equal to variable cost, and there would be no profit and no return to capital investment. Very high price at time of peak. Very low prices in all other hours. That is what a proper electricity model must imply. IEMM will imply that.

This section has put forth a well know result in economics that most or all of the return to capital must occur during the peak time. It is because there is systematic overcapacity, by construction, at time of off-peak. Association of return with the causal event is the standard result in economics that the price must be high enough to precipitate the causal event, and the causal event is the addition of capacity to hit fully the peak load. The causal event is the peak hour. It is the peak hour that causes the investment to have to be made all the way out to the peak. Therefore, the peak hour must bear the price of getting the market to add out to the time of peak. The entire profit the plant earns must occur during the time of peak. There is excess capacity at time of off-peak, and the plant earns no margin.

Thus, we should see a price calculation that has prices down near marginal short run cost for all times save the peak. At time of peak, we should see a significant price spike. Peak price spikes are exactly what we see in real world markets. When we see a “smearing” of prices across off-peak and peak, we know the model is wrong and is giving arbitrary answers without any consideration of endogenous capacity expansion. The price duration curve in an expanding market (or a market that has to add capacity to offset retirements and decommissionings) should look somewhat like Figure 127.

Figure 127: Peaky Price Duration Curve



This section is somewhat of a caricature. The true situation is more complicated. However, it is true that the market MUST see a very strong price at time of peak relative to all the prices at times of off-peak. We see this in the real world. Alas, we do not see it in dispatch models, which are generally wrong as a result.

This IEMM design of new capacity entry has no peer. There is no hope to represent it using MPEC, complementarity, or linear programming with the accuracy and insight in this design. The dimensionality is simply too large to appeal to global social welfare maximization or its cousin aggregate cost minimization.

7 RENEWABLES—WIND

7.1 Renewables Summary, Definition, and Scope

The notion of what qualifying renewables is rather simple in practice but confounded by myriad political and other agendas. The following list contains what people mean by "renewables," or more properly qualifying renewables.

- **Wind:** According to renewables advocates, wind offers the greatest potential for growth in the United States (and elsewhere) renewable power generation. The best resources primarily lie in the center of the country (the Midwest), stretching from Texas to North Dakota. The most optimistic projections argue for perhaps a fivefold increase in onshore wind capacity, from 63 gigawatts (GW) in 2014 to perhaps 314 GW by 2030.The most optimistic projections also envisage an additional 40 GW of capacity in offshore wind. To make this happen will require a substantial RPS probably assisted by RECs or heavy subsidies to attract the money to the market that can be profitably deployed. We will be talking in this section in great detail about wind. Most people believe, because of the intrinsic uncertainty in wind patterns, that a wind machine is an "energy machine," not a "capacity machine." When we come to the issue of storage in subsequent sections, particularly endogenous storage coupled with decentralized wind generation, we will have quite a different story, one that is central for IEMM.

- **Solar photovoltaics (PV) and concentrated solar power (CSP):** Recent years have seen decreases in the cost of solar PV technologies as well as the launch of several CSP plants. Solar insolation in the United States varies between regions, but across the entire lower 48 and Hawaii, solar insolation is higher than in Germany, the current world leader in solar PV capacity. (German solar PV has been deeply subsidized.) At the upper edge of reasonableness, by 2030 total installed capacity of solar PV might reach 135 GW, as compared to 7 GW in 2012.This raises the prospect of dramatic differences in the pattern and magnitude of distributed generation, with over one-third of solar PV capacity installed on rooftops. In such a situation, many users will become producers during some hours and importers during others, perhaps requiring re-utilization of the grid system.

- **Biomass and Biogas:** The United States (and other countries) have vast arable land resources, high potential in residues from the agriculture sector, forest and mills, as well as unutilized waste and methane from landfills. There is a substantial amount of arable land (albeit not always accompanied by the requisite fresh water) that could allow for crop-based biomass. There is significant potential for biomass to be used for heating, particularly in the manufacturing industry, where its use could triple between 2010 and 2030.Biomass offers the potential for an additional 46 GW of power generation capacity, taking the total to 84 GW by 2030.The big problem with biomass in power generation is that it might well have **substantially** higher value in the liquid fuels sector (what we call the RIN markets, which stands for "renewable index number"). Any notion that biomass and biogas are necessarily destined for the electricity sector could be just plain wrong. We must have a careful understanding of the refining and liquid fuels sector before we could make such an assertion. Fuels with a high RIN number tend to go to liquid fuel (bioethanol, biodiesel) markets. Even landfill biomass has a high enough RIN number that it might gain more value in liquid fuels markets than in electricity markets. IEMM (as defined here) will take that into account.

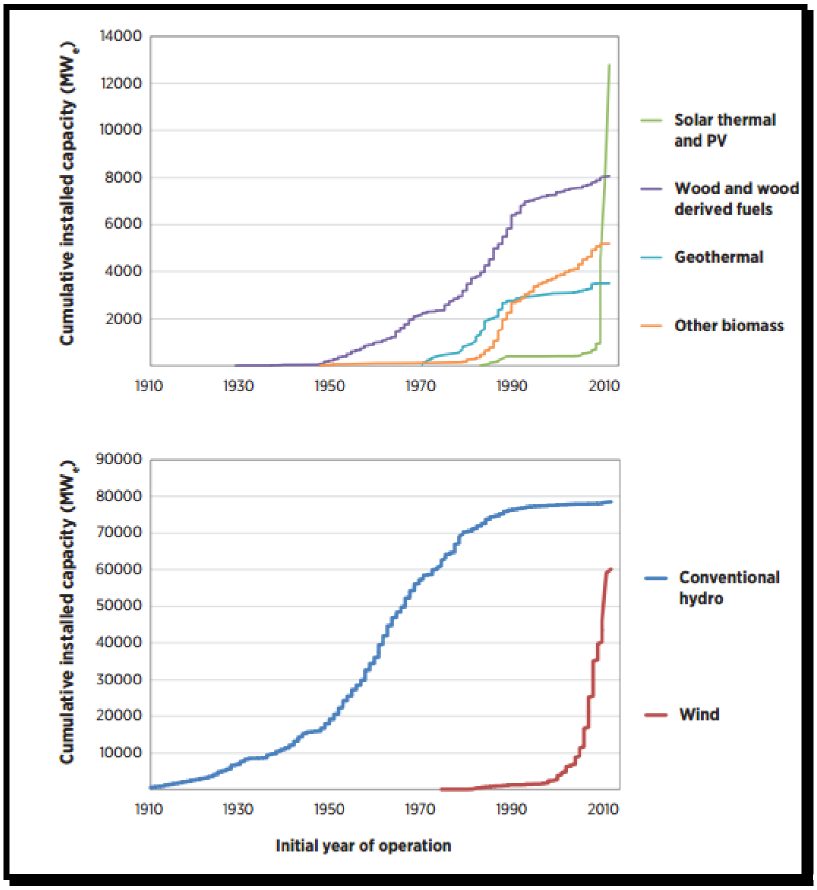
- **Geothermal:** The United States has a number of geothermal resources, significantly in the west. The United States is thought to be currently using perhaps 10 percent of its potential. Many other countries have large, untapped geothermal potential. It is not out of the realm of possibility to see an additional 18 GW in power generation from geothermal, adding to the 6 GW currently planned and likely over the coming 20 years. The rest of the world is no different.

- **Hydro:** Hydropower is currently the largest source of renewable power generation in the United States, although a number of jurisdictions have systematically defined hydropower as **not** a renewable. (It seems they want to dismantle dams and pumped hydro facilities for environmental and land use reasons, and they endeavor to justify it by not defining hydropower as a renewable.) In the United States and Europe, there is extremely limited potential for new large-scale developments. Additional potential can come from retrofitting and upgrading turbines at existing dams, adding power generation facilities at non-powered dams, and some new run-of-river hydro projects. In fact, run of river hydro can be some of the easiest capacity to add (politically). In Canada, there is more potential, but there is environmental opposition. Around the world, there are a large number of run of river hydro facilities, some reservoir facilities, and a few pumped hydro facilities. We introduce an extensive and accurate representation of hydro, which is critical for IEMM, in subsequent sections. IEMM cannot be without that.

The future of the power industry's generation portfolio seems destined to include a strong mix of renewable technologies, not just in North America but in every electricity system in the world. Renewables are advocated by virtually every government, and often subsidized, and their specific impact on the electrical system must be scrupulously modeled. Regulators and government agencies are bound and determined to ensure strong entry and presence of renewables. The recent trends in new renewable generation facilities have shown tremendous growth, especially in the wind sector, as depicted in Figure 128. The current total installed wind generation capacity in the United States and Europe have seen significant increases. The outlook for wind generation calls for further increases in capacity. Figure 128 certainly evidences the rapid growth in wind.

This rapid growth has occurred for several reasons, including strong political support for renewable energy including tax credits, availability of green pricing programs for consumers, improved manufacturing and performance capabilities, national and international pressure for renewables, and perhaps most importantly **renewable portfolio standards** that require a certain fraction of the mix be renewable (the definition of which is set by regulators and government agencies). In particular, arguably the key driver of the entry of renewable are the so-called Renewable Portfolio Standards (RPS), which require a certain percent of energy sales or of installed capacity to come from renewable resources. According to http://www. ncsl. org/research/energy/renewable-portfolio-standards. aspx, a number of states have mandated RPS. We have reproduced Figure 129 from that website to emphasize the extent of RPS in the United States. It is alive and well in Canada, Europe, and elsewhere around the globe. There is a lot of pressure for electrical systems around the world (e.g., China) to embed large and increasing quantities of renewables. This IEMM design will allow EIA to represent that very accurately.

Figure 128: History of Renewable Energy in the United States

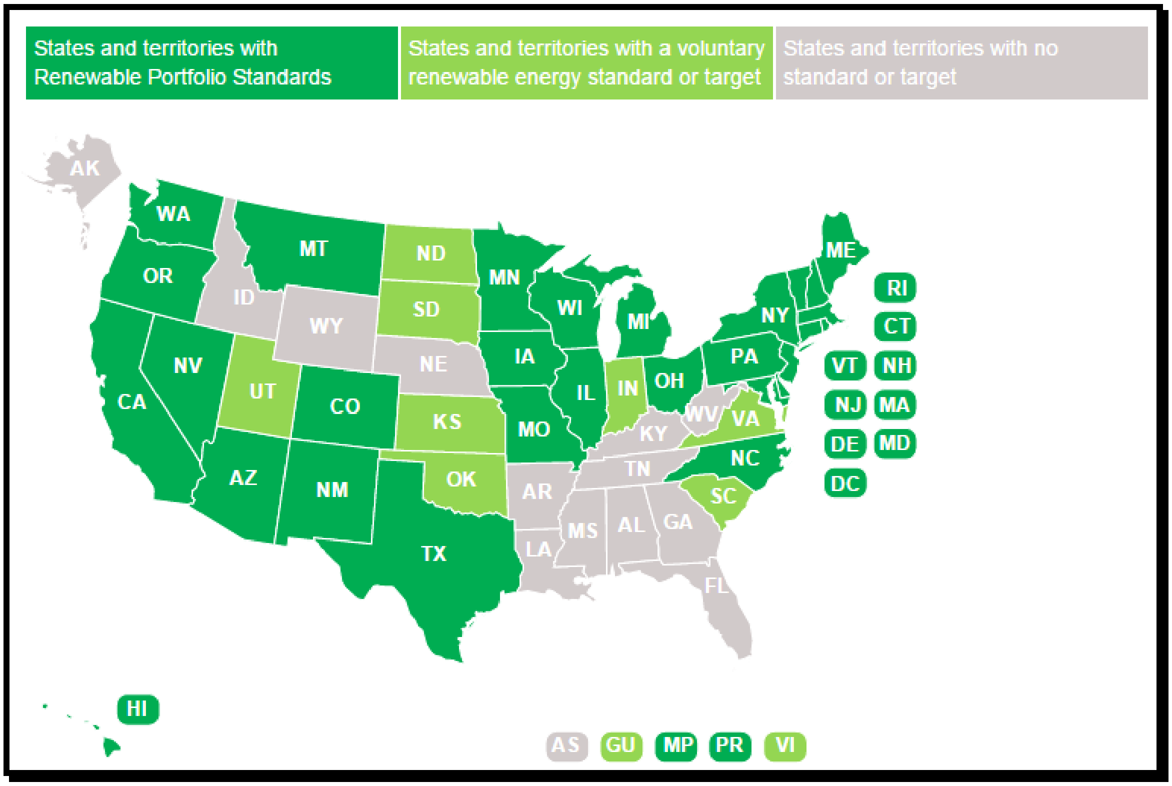


The idea behind RPS is to force renewables (both MW and MWh) into the generation mix, plain and simple. The first item of business is, of course, for government agencies to specify what specific technologies qualify as renewables. Generally, qualifying renewable technologies include:

- Solar Water Heat
- Solar Thermal Electric
- Photovoltaics
- Landfill Gas
- Wind
- Biomass combustion
- Hydroelectric
- Geothermal Electric
- Geothermal Heat Pumps
- Tidal Energy
- Wave Energy
- Ocean Thermal

- Municipal Solid Waste
- Anaerobic Digestion
- Biodiesel
- Fuel Cells using Renewable Fuels

Figure 129: Renewable Portfolio Standards or Voluntary Targets (2015)



Most states and many international governments have mandated that renewables should or must compose a certain fraction of total MWh in a region. The politics are crystal clear on renewables. There is consideration not only of mandating a renewables fraction but also trading entitlements generated from designated renewables plants (which are called renewable energy certificates or RECs) as a method of enforcement. IEMM must be able to represent RECs and tradable (or externally priceable) renewables allowances to create and examine alternative renewables scenarios and entries into electricity systems around the world. It is not OK to just guess; entry has to be endogenous, and it has to meet whatever RPS or equivalent might be in place. It is important if a trading regime for RECs is established to quantify the price of those RECs. Most anticipated RPS programs have contemplated REC origination and trading based on energy (MWh). Some anticipate programs based on MW of installed capacity rather than MWh of renewable energy, but MW of installed capacity do not have the same "teeth." The ones based on MWh of renewable energy will be more difficult to achieve because they have more "teeth." IEMM must model that implicitly and accurately, and the design in this section shows exactly how to do so.

There are a number of dimensions with RPS and RECs that merit consideration in the IEMM model:

- **<u>RPS timing and size</u>.** The IEMM design herein allows a broad range of RPS programs to be represented. Given the diversity and uncertainty in these programs, it is important to have the flexibility to posit programs and evaluate them. The issue of "how much" and "when" is a leading issue with RPS, and IEMM (as defined here) allows broad consideration.

- **<u>Local and national variation</u>.** There are literally myriads of RPS rules and proposals across many state and other venues. It is important to consider a reasonable base case and alternative cases. It is also important to consider their RPS programs in a geographically diverse fashion. France will not have the same RPS as Germany, and differences matter.

- **<u>Negative renewables balances</u>.** How fungibly traded can RECs be? What if states mandate that they will not allow a net negative or positive balance within the state, meaning that they intend to force local generators to comply with local renewables percentages locally and not trade responsibility to contiguous states? What might that do to REC prices? Renewables entry? Location of renewables entry? Impact on distant systems versus local systems?

- **<u>Renewables credits by building a sola collector in the Sahara</u>.** Will it be acceptable to generate RECs by building wind power in a distant, completely disconnected system such as a solar collector in the Sahara for example? Would a generator in New Jersey be allowed to take a REC from a wind turbine it built in Burkina Faso? If so, RECs will materialize in the New Jersey electrical system "for free." If not, RECs will have to be generated within the New Jersey system indigenously. That difference is very material to the New Jersey electrical system, and this design will model that.

- **<u>Who suffers (i.e., exits the market) under alternative REC regimes</u>**. Is it old coal plants? Old gas plants? New coal plants? New gas plants?

- **<u>Market driven or mandated</u>.** This refers to two different approaches to the RPS/REC problem. Mandated programs are those that posit renewables to be a certain fraction of total MWh or MW and do not allow net new addition of nonrenewable plants until the criteria are met. Market driven methods would involve trading of RECs.

- **<u>Policy shortfalls</u>**. If a government mandates an RPS but then allows shortfalls and "workups" and ramp-up times to the RPS, how does that affect the system? Do new thermal generators get embedded in the system early and affect long-run entry of delayed renewables?

- **<u>Alternative RECs and REC trading regimes</u>.** IEMM (as defined here), with customization, allows fairly broad representation of REC trading and REC entry. There will be customization required to represent the full range of REC rules because so many are under consideration among the various regions of the country.

- **<u>Renewables technology</u>.** IEMM (as defined here) allows a wide range of technological options for renewable energy including enhanced wind turbine efficiency, lower cost, and so forth. New entrants can be cheaper and more efficient than old, embedded entrants.

In the IEMM design (as defined here), nonrenewable and renewables generation, which include gas combined-cycle and combustion turbine (with anticipated learning forward through time and with or without CO2 removal), pulverized coal, supercritical coal, IGCC without CO2 removal, IGCC with CO2 removal, renewables, nuclear, fuel cell, and other exotic forms of thermal generation, as well as the preceding range of renewables options, all interact and compete. IEMM (as defined here) allows a wide range of technological options for plant retrofits, those that exist today and possibly more cost effective ones that might emerge in the future. The model allows the full range of timings of availability or entry of new technology.

Because wind, solar, biomass, and geothermal are the most important (especially wind), we will give special attention in particular to wind and solar, wind in this section and solar in the next. In particular, we are going to discuss the proper representation of the entry and operation of renewables—wind, solar, and/or biomass. We need to discuss wind separately from solar separately from biomass.

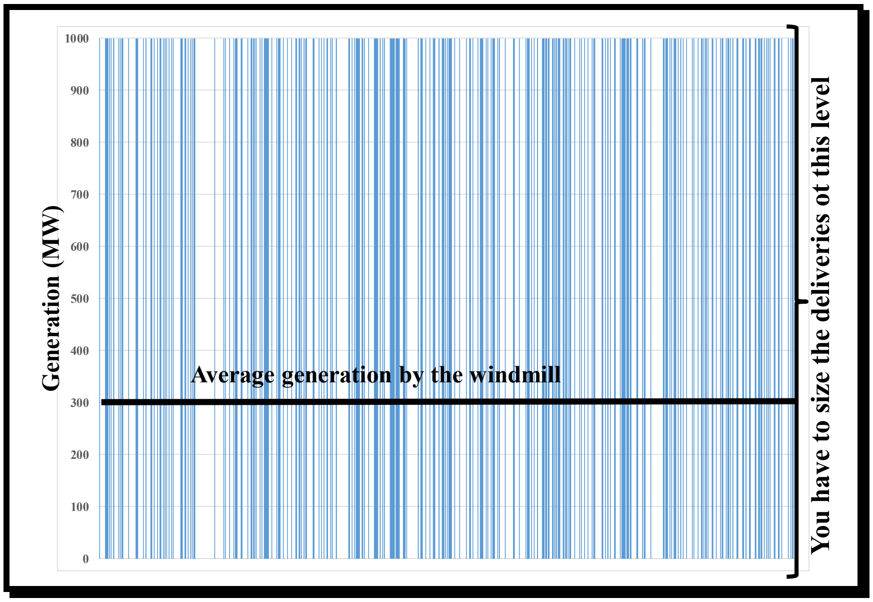
7.2 Fundamentals of Wind

The best wind turbines generate about 30 percent of the time because that is all the time that the wind blows. Wind turbines do not create wind (alas), and they only create energy when the wind blows. They do nothing the other 70 percent of the time. This is an important consideration when it comes to wind. These are machines with perhaps a 30 percent load factor, and the load factor can happen any time of the day, night, week, month, year, etc.

The wind pattern is very site specific, so in an aggregate model, we must consider the wind turbine generation pattern (i.e., the wind pattern) for the average or representative wind turbine in the region if we are not sufficiently disaggregated so that we can discern specific wind patterns in specific potential and existing locations. Figure 130 illustrates what the generation pattern for a wind turbine is. It is either "off" or it is "on." It is either running at 100 percent capacity (i.e., the wind is blowing) or at 0 percent capacity (i.e., the wind is not blowing). The wind is blowing 30 percent of the time and not blowing 70 percent of the time, meaning that for each hour, the colored line is maxed at 100 percent for 30 percent of the hours and is zeroed for 70 percent of the hours. The diagram in Figure 130 has 30 percent of the hourly lines at the maximum and 70 percent of the hourly lines at zero. It is very deceiving to observe the diagram in Figure 130, which looks "pink," and very much more deceiving to view it as an average or mean type of diagram.

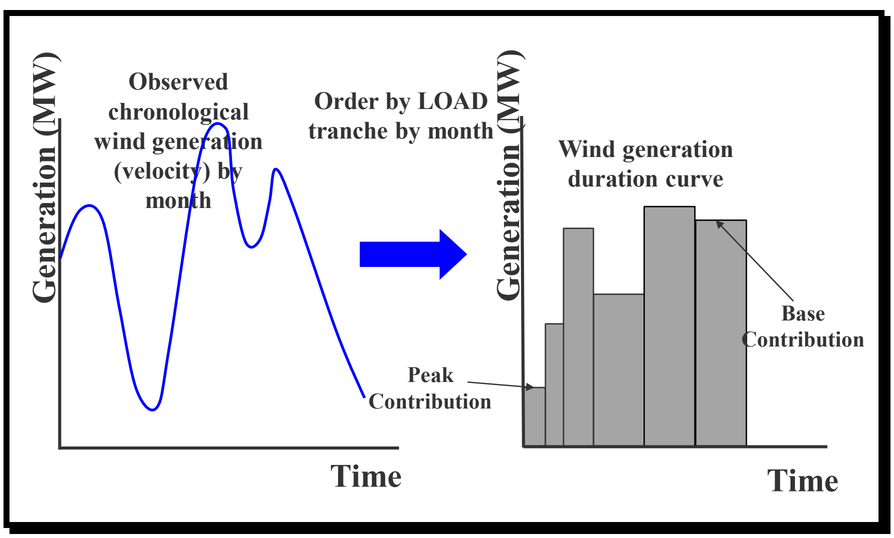
It would be tempting to assert that on average over the year that the wind is blowing 30 percent of the time and, therefore, to "derate" a 1,000 MW wind farm and assumed that it generated 300 MW in every hour of the year on average, i.e., to assume the black line in Figure 130. This is a typical practice for thermal power plants—derating capacity so that the average available capacity is "one minus the outage rate" times the nameplate capacity. This cannot be done for wind turbines. The wind duration and timing statistics of the wind have to be taken into account of at a finer level of detail than for a thermal plant, which can run pretty much whenever it is called (subject to ramp-up considerations).

Figure 130: Generation Pattern for a Wind Turbine



What needs to be done, therefore, is to develop the magnitude of wind generation (i.e., wind velocity capable of wind generation) by hour, by day, by month throughout a given year. That is, we must generate a statistical average magnitude of MWh of wind generation inserted into the electrical grid from the wind turbine for every hour of the year. Such a curve appears on the left of Figure 131. It is the observed wind turbine generation pattern by hour (or the statistical average thereof over a large number of observations) for the given month under consideration.

Figure 131: Group Hourly Wind Generation for Each Month into the Peak-Intermediate-Base Hour Tranches



Once we know what particular hours of the various months that the wind blows and generates in each month, we can assign those hours to the tranches of load in the load duration curve as structured through sorting in Section 4 above. There must be a mapping from every one of the "wind hours" into every time tranche in the load duration curve developed in Section 4. It is important to emphasize—we assign those hours precisely to the tranches of load in that table and NOT create a wind generation duration curve or sort the hours into some other criteria. The hour classes in which the wind turbine generates are the selfsame hour tranches in which the load duration curve is defined. One must retain the mapping of specific hours that load persists to the specific tranches of load in the load duration curve. One must then use that selfsame hour-to-tranche mapping to calculate the wind generation duration curve as in Figure 131. Chronology must be scrupulously maintained. This is important—the wind hours must be assigned systematically to the tranches that define the load, not ordered independently from highest to lowest wind generation hours or wind velocities at the site. This is important; independent ordering of wind hours would be wrong. Analyses of the order of wind hours from highest generation to lowest generation are sadly misleading and totally worthless. The wind has to be synchronized with the load and not analyzed independently.

If there is say 1,000 MW of wind generation capacity in a region (not energy, but capacity, i.e., maximum ability to generate), we develop using the preceding procedures a wind duration generation curve of the form at the right of Figure 131. As illustrated by the figure, there might, for example, be only 120 MW being generated at time of peak, and that is occurring during the peak 0-5 percent of the hours in the month. There are 0.05*730 or 36.5 hours in that tranche of peak hours. That means that there are 120*36.5 or 4380 MWh generated in the peak tranche. That is precisely equal to the area of the leftmost rectangle in Figure 131. The leftmost rectangle represents the quantity of energy from the wind turbine that is contributed at time of peak. Moving over to the second to the leftmost rectangle, the height of that rectangle might be say 200 MW, the number of hours in that tranche (10 percent of the hours in the month) is 73, and the energy contributed is 200 MW*73 hours or 14,600 MWh. By repeating this contribution of the 1,000 wind to each of the six tranches of load, all perfectly temporally correlated because of our knowledge of the timing of the wind generation pattern, we create the wind generation duration curve in Figure 132. We can interpret the diagram as follows:

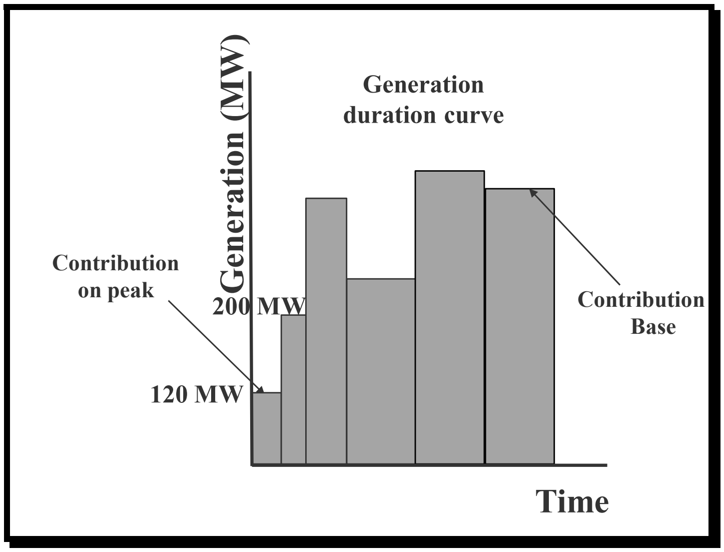
- The area of each rectangle is the amount of wind energy that is contributed to the grid during each tranche of hours.
- The height of each rectangle is the average rate of contribution expressed in MW contributed to the grid during each tranche of hours.
- The width of each rectangle is the number of hours in the tranche, i.e., the tranche driven by the load and not the wind.

The total sum of the areas within the six rectangles is the total wind energy contributed during the month. The heights of the curve represent the MW contributed during the various periods within the month, which reflect the exact timing during the month when these wind patterns occur. The areas represent the MWh contributed during each period within the month. This shows how we obtain the monthly MW and MWh contributed. Once we have the monthly wind contribution pattern, we then do several things:

1. We accumulate the capital cost and the operating and maintenance cost of the windmill (1,000 MW of generation capacity in this case) and run it to an account, or perhaps two accounts, one for fixed and one for variable cost.

2. We assign that cost to retail customers in the form of increased distribution tariffs, reflecting pass-throughs of fixed and variable cost of wind generation to ratepayers. By so doing, we correctly capture the phenomenon that renewables may **reduce** the wholesale price of power, but renewables do not reduce rates. They increase rates. The money from this renewables account we allude to here will exceed the cost of thermal energy. It must be included in retail (but not wholesale) rates.

Figure 132: The Wind Generation Duration Curve for 1,000 MW of Wind Capacity

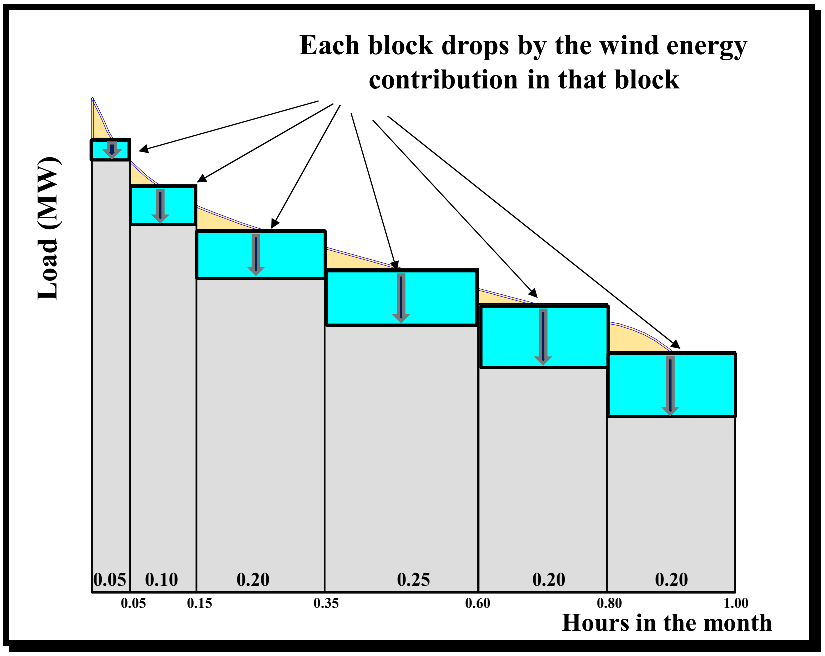


3. Alternatively, it might not be appropriate to instead assign the renewable energy generation capital and operating cost to retail customers in the form of increased distribution tariffs. Rather, it might be taxpayer subsidized, and that means the account must be assigned to taxpayers or third parties who provide incentives in the form of subsidies. In that situation, both wholesale and retail rates can decrease because taxpayers rather than ratepayers or market agents are picking up the bill. IEMM must (and will) consider alternative arrangements regarding who pays the entry and operating cost of wind.

4. Most importantly, we set up a simple supply curve with nearly zero marginal cost for wind (because the fixed and variable costs are sunk and assigned elsewhere in the energy system). This allows us de facto to operate the wind first to the local load (in the case that the local load cannot be **fully** met with wind, or to force wind into the local load at a level above and beyond local demand and force the model to export any excess energy and capacity to contiguous regions. At the price of wind, such export will be very economically justified. When we have forced wind into the local busbar at its statistically correct generation pattern, the original local demand duration curve from Figure 100 becomes adjusted as in Figure 133. Specifically, the heights of the wind energy contribution curve from Figure 132 are subtracted from the heights

of the load duration curve in Figure 100 to create the adjusted or residual demand duration curve in Figure 133. It is the gray curve at the bottom, ex-post adjustment for temporally correct contribution by wind energy. That is the demand duration curve that power plants in the region now see after the entry and operation of wind. If we use this gray curve instead of the original, pre-adjustment curve, then we get the power price and quantity correct. The price is set by the marginal, unsubsidized thermal unit.

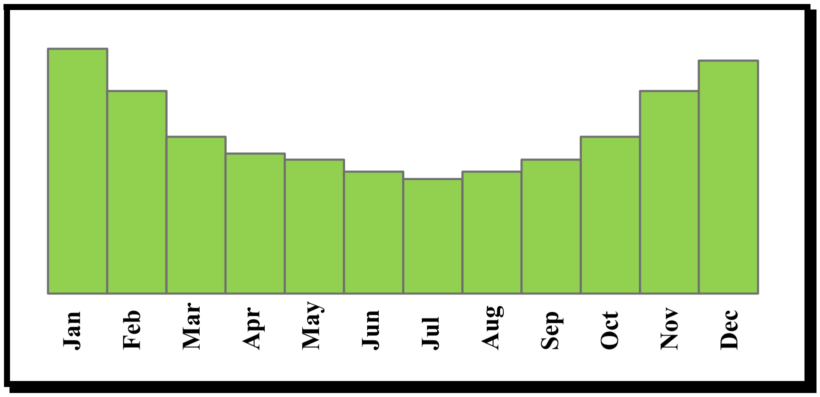
Figure 133: Stairstep Adjustment of Monthly Load Shape to Account for Stochastic Wind Contribution



Returning to the wind generation duration curve in Figure 132, there are two phenomena that we must and do represent that impact that curve within IEMM. There are two dimensions of non-randomness that must be accounted for—seasonal (temperature) variation and daily wind variation. It is well known that the wind blows less during the hot time of the year and much less during the hottest time of the day in the summer. Wind tends not to blow during the hottest days of the year because there is generally temperature inversion during those days. (Air during a temperature inversion moves "up and down," not "side to side." That means that during a temperature inversion, wind turbines are often becalmed.) In addition, wind tends not to blow as much during the middle of the day as during the middle of the night (contrary to what golfers think!). If we were to calculate the total wind energy generated during each month in the region (i.e., the total gray area under the wind generation duration curve), it might appear such as that in Figure 134. The height of each monthly curve is the average MW generation during that month, and the width of each monthly curve is the number of hours in that month. The green areas in Figure 134 are equivalent to the sum of the total blue rectangular areas in the wind duration generation curve. Hence, the area of each rectangle in the green curve is the MWh of wind energy generated during each month. We take account of its intra-monthly temporality with the wind generation curve, but we take account of its total monthly energy generation as in the green curve.

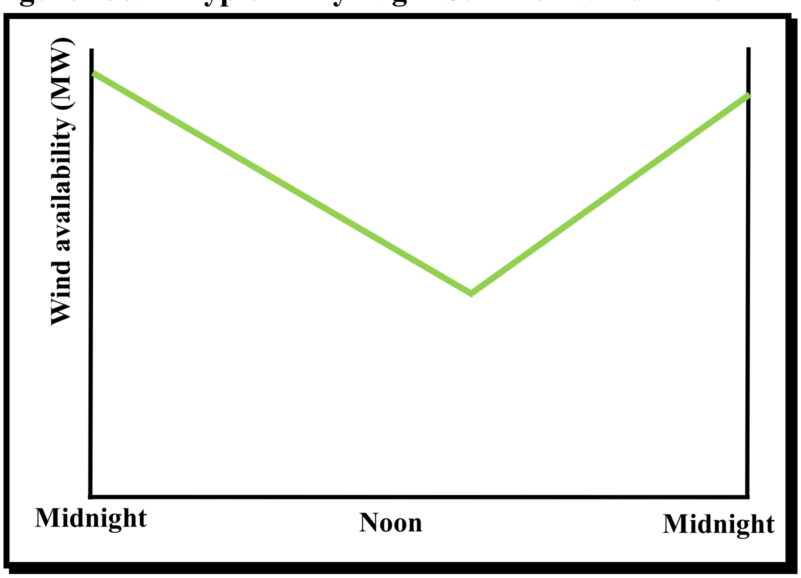
Thus, there are twelve total monthly wind generation duration curves of the temporal form in Figure 132 but sized so that their total area is that in Figure 134.

Figure 134: Wind Energy Generated by Month in a Given Region



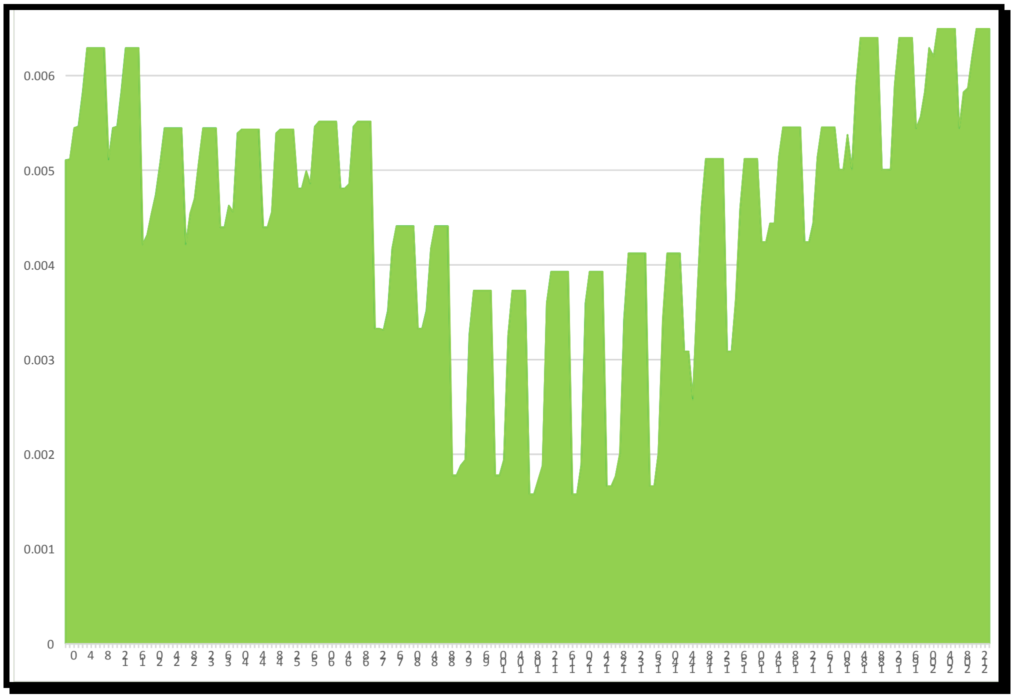
Within each month, there are of course individual days, and we know that the wind pattern across a day varies according to a statistical day-night pattern. A typical summer day pattern might appear as in Figure 135. Generally, on hot days, the wind velocity is notably lower during the heat of the day at 2:00 pm than in the evening or night. This is a typical wind pattern for a summer day. For a winter day, the pattern is higher on average and more randomly distributed across the day. The daily and hourly wind generation patterns observed historically are explicitly considered in the procedure articulated above.

Figure 135: Typical Day-Night Summer Wind Pattern



ArrowHead has been working in the PJM service territory in recent years. We have accumulated the wind generation duration curve for the PJM West region (around the Pittsburgh area). We have divided the year into 12 months, and we have divided the months into 18 tranches of load each. This client has wanted a lot of intra-monthly detail so that they could understand the precise shape of the peak and the precise shape of the base. We have plotted the annual wind generation duration curve for that region (normalized to 1 MWh of annual energy), and in so doing we obtain the shape in Figure 136. We can discern the seasonal variation in wind velocity from Figure 134, and we can certainly visualize the day/night pattern from Figure 135 in the monthly curves. This is a pretty "lumpy," non-regular contribution of MW and MWh to energy generation in the region, and that is precisely the proper characterization of wind (and, in fact, other renewables). Any notion that wind is a "capacity machine," easily available for energy on demand, is clearly not borne out by profiles of production that have the form in Figure 136.

Figure 136: Wind Generation Pattern (Annual by Hour Tranche)

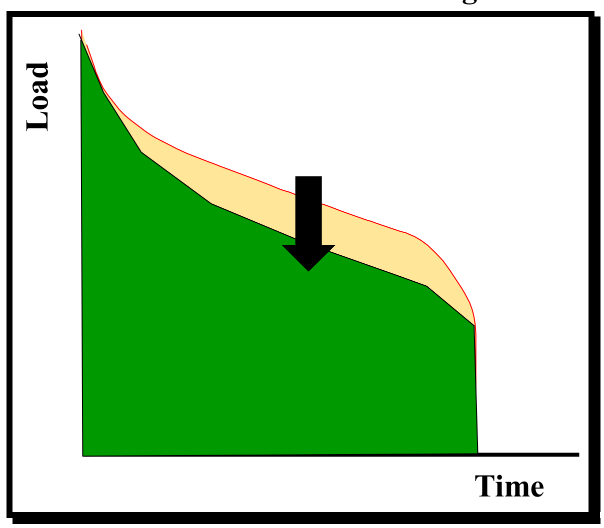


We are going to see in the forthcoming renewable supply section (following the analogous discussion for solar in the next section) that in fact this generation duration curve, normalized for the magnitude of capacity in the region, is injected at its variable operating cost (which is very low) into the regional busbar in which the wind exists. This is the correct model of wind. Other models are simply incorrect because they completely omit the synchronization of wind energy with load and with other thermal energy entering the grid and being delivered to users.

Assuming that the variable cost of the wind (i.e., the dispatch cost) is so low that it is beneath that of any other source, this pattern of wind energy is going to dispatch into the local busbar unconditionally. No one can beat it! That is the way IEMM (as defined here) will work—injecting the time pattern in Figure 136 for every MW of wind that is installed in the region and installing it at the variable operating cost of the wind turbine (which is virtually zero). For purposes of explication, and explication only, we can assume that this wind will always dispatch against local demand (not the way IEMM will work, but very illustrative of how wind affects an IEMM system), i.e., for didactic purposes, we will redact local wind from local load to anticipate what the true impact of wind is likely to be in any IEMM system.

By redacting the contribution of local wind from the local load tranches and doing so in parallel tranche by tranche, we effectively adjust the demand for electricity, taking explicit account of the energy generated from wind and of its specific temporality. When we do so, we will have derived precisely what the residual thermal load looks like, i.e., the load that remains to be served by the thermal plants in the region after the wind energy is absorbed at its "lower than anyone else" cost. This is precisely the adjusted green curve at the bottom of Figure 137. The green curve is the load that is left over for the thermal generators in the region following the contribution of wind, which is stochastic at its heart but somewhat statistically regular. To reiterate, the green curve is what the load that remains for thermal plants looks like (with no prospect of electric storage quite yet).

Figure 137: Modification of the Load Pattern Resulting from the Introduction of Wind



We note several phenomena here that are very important with regard to wind:

- Wind pushes down the off-peak more than it pushes down the peak. To wit, it cannot be argued that wind pushes down the peak with probability 1.0. Wind is to a significant degree an on-off phenomenon, and system reliability requires that one has thermal plant backup or reserves during those random, stochastic times during which the wind might not be blowing at time of peak. It will not be appropriate for the wind duration generation curve to be 100 percent attributed and attributable at time of peak. We must discount its contribution at time of peak. The change that one expects to see and model is represented in Figure 137. This is not because wind is "bad." It is because modern electrical systems must be close to 100 percent reliable. Developing country and poorer power systems may not be able to absorb wind for this reason, and IEMM will want to model that carefully. This design allows precisely that. This is going to be far from a trivial issue in other electrical systems, and, in particular, developing country electrical systems up to and including places like China.

- Thus, the introduction of wind renders the resulting load shape **worse** for the thermal plants because its energy contributions are most valuable at time of off-peak when there are plenty of thermal plants de facto providing reserves because of their very existence. Wind does not reduce the peak, but it reduces the off-peak.

- The deterioration of the load shape is likely a **theorem**; it literally always has to happen with a duration shape such as that in Figure 136. It is destined to drag down the load factor remaining to thermal generators for almost any load shape. The fly in the ointment here would be the introduction of electrical storage and applying it at the wind generation site. Subsequent discussions address that because technological change is increasingly raising that possibility. In California, regulation encouraging electricity storage has proceeded ahead before there is sufficiently efficacious electricity storage technology, reasoning that mandate will lead to subsidized development of intermittent renewables. IEMM has to be able to model whether that is the case, and this design will.

- An electric system needs complete redundancy at time of peak (gas). Otherwise, system reliability can fall to "developing country" levels. That is not going to be acceptable in the United States, Europe, or other electricity grids, and our demand and renewables generation patterns must reflect that. They do with the recommended IEMM method herein.

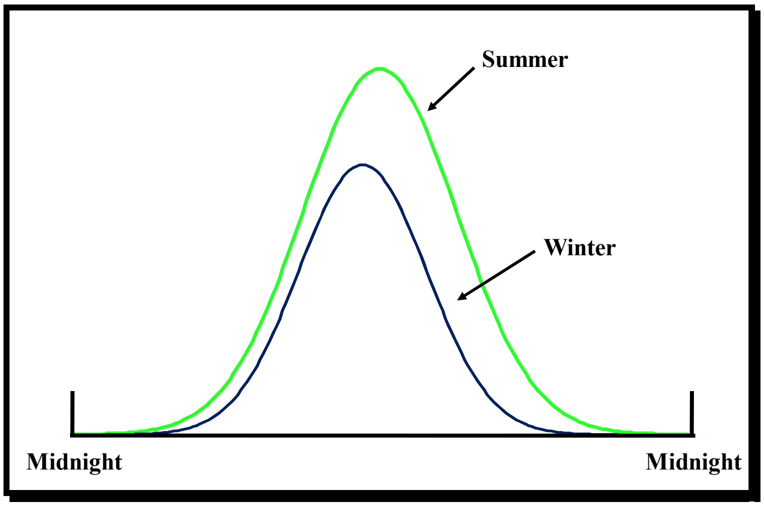
- As the off-peak is reduced, it can fall below critical levels at which base load plants would have to cycle and thereby induce operating strategies that are difficult. As the green curve drops on the right (base load), operations of coal plants can be affected. When that happens, either the coal plant must cycle to meet load, or one must drop (exclude) the wind energy by "feathering" turbines and that type of thing. (People in California have commented that they have to "feather" wind turbines in months like April when load drops to its minimum and the hydroelectric run is in full swing.) IEMM, configured as it is in the fashion outlined here, will correctly quantify that impact of wind on baseload generation.

- This phenomenon of dumping wind at time of off-peak has already occurred in California and Texas in recent years—both have dumped wind energy at time of off-peak. In both states, there has been literally zero wind energy at time of peak, so their curves do in fact look very much like those in Figure 137—no change at time of peak but large depression at time of off-peak. IEMM (as defined here) will capture that and thereby reliably and correctly capture the impact of wind at various RPS or indigenous levels in an electric system. This is crucial to model and understand the entry of renewables. The actual situation in IEMM (as defined here) is not quite this simple, but this is quite didactic as to what we would expect to happen. This phenomenon must be covered in every electrical system in the world.

8 RENEWABLES—SOLAR

Much written about wind in Section 7 previously pertains directly to solar with a few critical modifications. First, the day-night pattern of solar looks like a "normal curve" with solar insolation maximized at a time right after high noon and going to zero morning and night. (The presence of solar collectors, no matter how efficient, does not get the sun to shine at night.) There is a longer and stronger solar insolation pattern during the summer than during the winter. Thus, for solar, the daily solar insolation pattern and therefore solar generation pattern has the characteristic form in Figure 138.

Figure 138: Solar Insolation during the Day-Night Cycle



This pattern for daily solar must be assembled every day and a statistical average for each day of the month must be assembled. This procedure is precisely the process articulated in the previous section for wind. Precisely as with wind, we must find the solar insolation hour to hour curve as shown on the left of Figure 139 and map it into the precise load hours (as defined by the load duration curve tranches) on the right and thereby create the solar generation duration curve. The hour-to-hour synchronization of solar generation with the loads that have been assembled to represent final use is critical here, as it was for wind.

After that, one can assign solar generation to the various hours of the month to obtain a solar electric generation duration curve. Just as with wind, these assignments have to be made precisely based on the original chronology of the hour tranches used to characterize load. It cannot be accomplished independently of the consumption load tranching. Such a curve would look quite different from wind, for solar is correlated with time of peak in the summer but with time of intermediate in the winter. It would have a look such as the right-hand curve in Figure 139. We would subtract the right-hand curve in Figure 139, which is reproduced in Figure 140, from the curve in Figure 100 for each month just as we did in Figure 133 for wind.

Figure 139: Group Hourly Solar Generation for Each Month into the Specific Hour Tranches

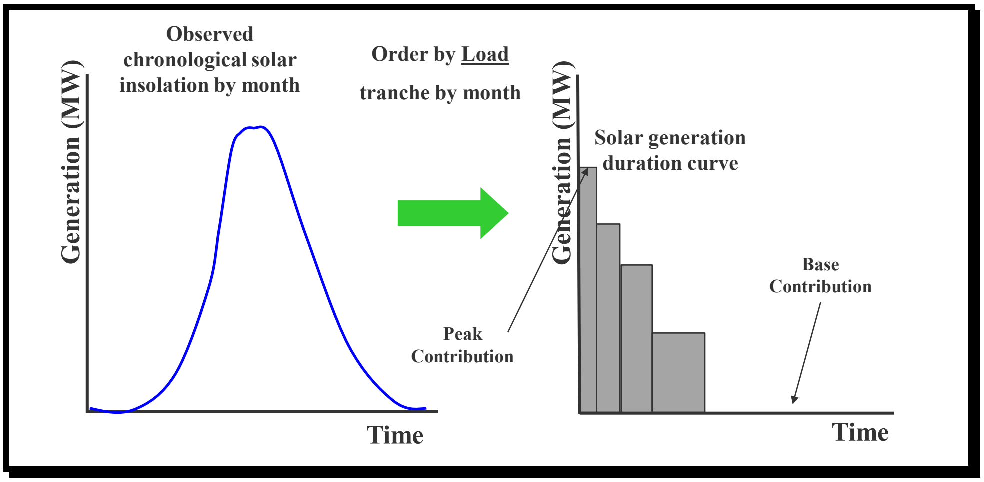
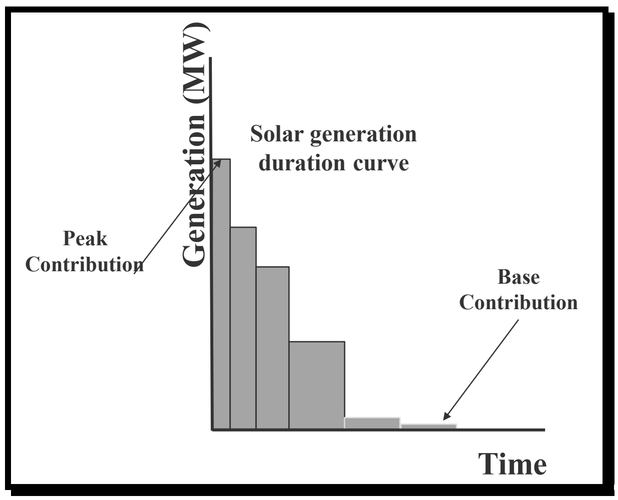


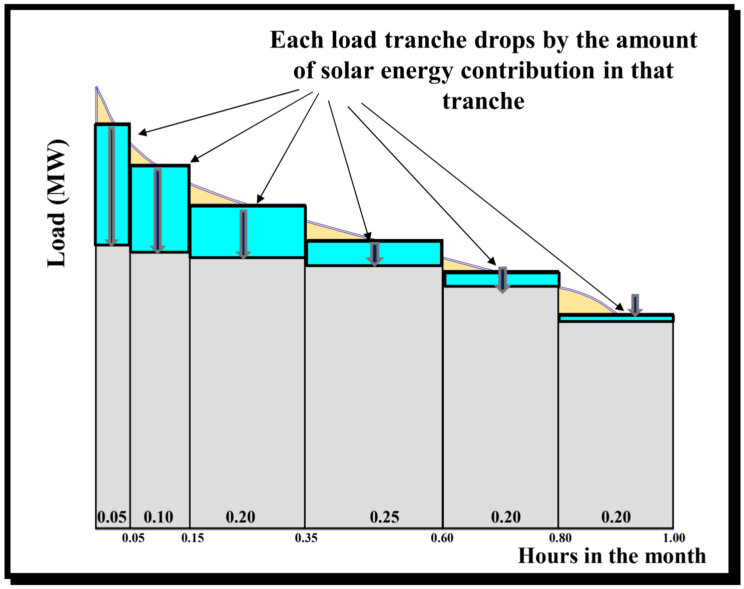
Figure 140: The Solar Duration Curve for 1,000 MW of Solar Capacity



Notice in Figure 141 that the addition of solar with its generation duration curve will tend to flatten the residual load duration curve and thereby render it less "peaky." This allows more of a base-loaded thermal operation profile. Notice how much flatter (more base-loaded) the gray area (post solar) of Figure 141 is than the original load curve (the red curve), which is peaky.

Technically, we would subtract the solar energy contribution in Figure 140 from the gray curve in Figure 100, which has already been adjusted for the wind energy that is contributed. With solar energy, because of the stochastic lack of reliability, we cannot assume that the sun will not go behind a cloud at time of peak. We must for reliability purposes possibly discount the solar energy MW at time of peak. We can do so by modifying the contributions of solar contributed in Figure 140, i.e., we can delete some of the peak contribution to take account of system reliability or weather randomness at time of peak.

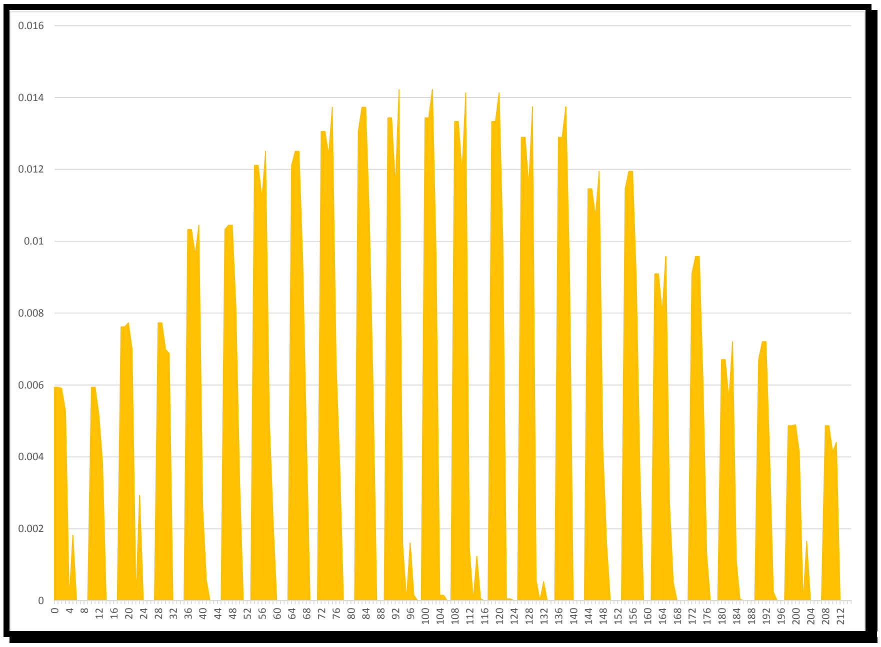
Figure 141: Stairstep Adjustment of Monthly Load Shape to Account for Stochastic Solar Contribution



We have seen in this section how we can take explicit account of the mandated entry of solar (and in the previous section wind). Mandated entry of other forms such as biomass or tide or other diverse sources is entirely analogous. By representing them in the fashion articulated herein, IEMM can calculate the correct power and fuel prices and burns and thereby the correct prices in the system. Equally importantly, we capture the specific temporality (at the statistical average level, inserting that statistical average as a deterministic input). Biomass and geothermal are much easier, for those sources are not stochastic like wind and solar. They enter the generation mix just like coal or nuclear. They are base-loaded, and they contribute their energy at the left side of the supply stack (after subsidies).

As discussed previously, we have been working in the PJM service territory. We have accumulated the solar generation duration curve (solar insolation curve) for the PJM West region (around the Pittsburgh area). As before, we have divided the year into 12 months, and we have divided the months into 18 tranches of load each. That client has wanted a lot of intra-monthly detail so that they could understand the precise shape of the peak and the precise shape of the base. We have plotted the annual solar generation duration curve for that region (normalized to 1 MWh of annual energy), and we have obtained the shape in Figure 142. We can see the seasonal variation in solar insolation from Figure 140, and we can certainly visualize the day/night pattern from Figure 138 in the monthly curves, which is a pretty "lumpy," non-regular contribution of MW and MWh to energy generation in the region, but it is substantially more regular in nature than its wind counterpart. With the exception of "clouding over," the temporality of solar energy is more regular than that of wind. The notion that wind is a "capacity machine," easily available for energy on demand, is clearly not borne out by profiles of production that have the form in Figure 142.

Figure 142: Solar Generation Pattern (Annual by Hour Tranche)



9 ENFORCED ENTRY OF RENEWABLES TO A RENEWABLES PORTFOLIO STANDARD

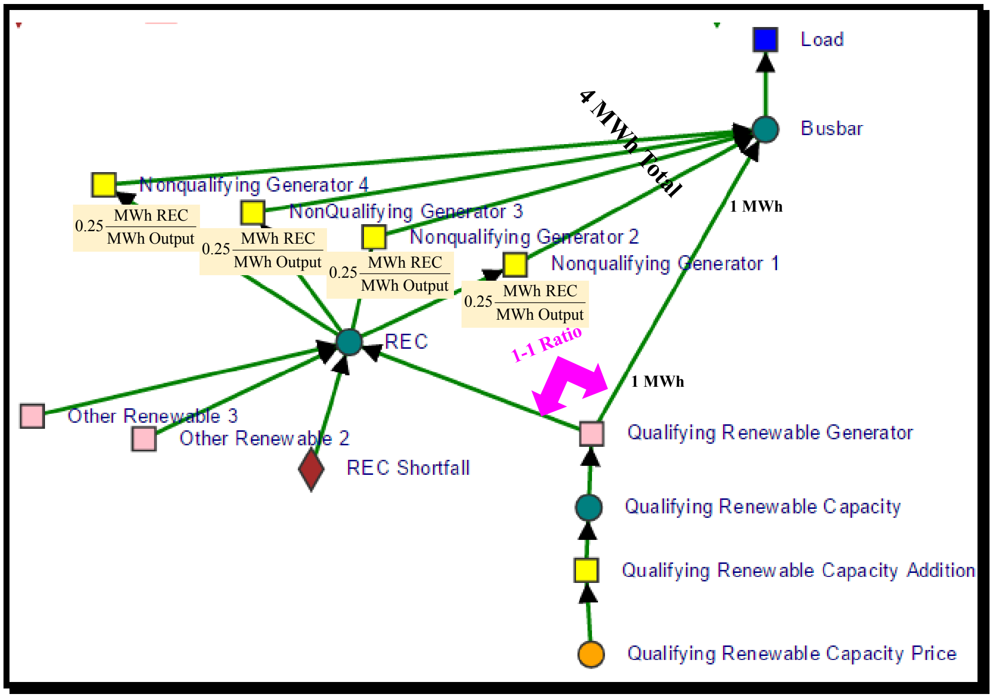
Now that we have discussed the nuances of wind and solar and the generation duration curves that they imply, we are now ready to represent the mandated entry of wind and solar en masse as would be commensurate with an RPS. That is the purpose of this section—for IEMM to model the actual entry and operation of a system according to an RPS.

One of the ways to ensure the entry of renewables into an electricity system is to establish a Renewable Energy Certificate (REC) system to force their entry. This is both a property of the real world and a modeling convenience that is very important in understanding and quantifying renewables entry and operation at different levels. RECs represent a right to do certain things because of a contractual and economic connection to a "green" or an environmentally friendly generating facility. Renewable Energy Certificates have been referred to as green tags, Renewable Energy Credits, Tradable Renewable Certificates (TRCs), and environmental attributes around the world. This section summarizes briefly how they trade and how they can be incorporated into IEMM and how thereby IEMM can include them in a sophisticated fashion. There are so many alternatives for doing so that one cannot hope to have them all in the model today, but as they are shaped, they are easy to add as discussed in this section.

In IEMM, RECs would be considered a commodity separate from power produced—producers of "green" power generate and can sell RECs as well as power itself. RECs are allowances that are given value via a mechanism outlined previously and below that a renewable plant owner can sell right alongside his or her power. For each MWh of green energy, a green facility generates, that green facility owner also generates a 1 MWh REC. He has the right to print (i.e., to obtain for free) a 1 MWh REC "scrip." The sale of MWh to the grid gets a $/MWh market price. He also has the right to sell the 1 MWh REC "scrip" to another generator and keep the money from such sale. As REC scrip is generated, the local (or national) regulator would simply require that every non-green, thermal generator in a region must surrender 0.25 of REC scrip for every 1 MWh of thermal electricity generated. The renewable energy generator would sell his 1 MWh of REC on the open market, and that 1 MWh of REC would allow 4 MWh of thermal energy to be generated. When RECs are traded and ultimately held by thermal generators, the entity purchasing the REC would gain the right to generate and thereby consume those RECs (or hold them and thereby and claim environmental benefits by eliminating thermal electric generation). Non-qualifying generators have to buy REC scrip; otherwise, they are not allowed to generate. REC scrip is an entitlement to generate for non-qualifying facilities.

The actual model structure necessary to get renewables into the generation mix at the level of 20 percent of the market (expressed in terms of MWh) is that in Figure 143. We will spend a bit of time walking through its constituents because it is such an integral part of the renewables part of IEMM. Beginning at the top, notice that there is a demand at the busbar in the blue box, and there is a busbar hub just entering that blue box. That busbar hub is where MWh that are consumed in the region are accumulated, and as shown in the diagram they are accumulated from qualifying renewables generators at the lower right and from non-qualifying thermal generators at the middle left. That is, busbar electricity comes from non-renewables and renewables alike.

Figure 143: A Model of REC Acquisition to Achieve 20 Percent RPS



At the lower right, suppose that 1 MWh of physical energy is delivered to the busbar market. That delivery of 1 MWh of physical energy to the busbar market is represented by the upper right output link from the pink node delivering 1 MWh along the link that points into the busbar market. It is the link that leaves the pink renewable operations node at its upper right. We want that 1 MWh of renewable energy leaving the pink link to enable 4 MWh (but no more than 4 MWh) to be delivered to the busbar market from non-qualifying thermal generators in yellow at the left. If we impose on every non-qualifying thermal generator in the middle left that they must purchase and surrender 0.25 MWh of REC (a REC being a contractual representation that indeed 1 MWh of qualifying renewable energy has been generated by the renewable operator and delivered to the busbar market), then the sum total of generation by the non-qualifying thermal generators in aggregate cannot exceed 4 MWh. That means that the busbar will receive a total of 5 MWh, and 1 MWh of that 5 MWh has to be qualifying renewable energy. That ensures that renewables will make up 20 percent of the total market in the region and that in fact a 20 percent RPS would occur. Twenty percent of all energy into the busbar would come from qualifying renewables, and 80 percent would come from non-qualifying plants.

The price of the REC will rise to the point at which the revenue dollars coming to the renewable generator by virtue of his sale of 1 MWh to the market at the price that persists in the busbar market **plus** the price he can get for the sale of 1 MWh of REC he obtains to the REC exchange will fully pay his capital plus operating cost plus a market rate of return and will, in fact, incentivize him into the market. This means that the price of the REC will have to equilibrate to the point at which it is equal to the "long-run marginal entry cost" of the next unit of renewables at the lower right minus the variable cost (dispatch cost) of the best (i.e., marginal) thermal generator that is not allowed to generate, i.e., is shut out, because of the existence of the RPS. The REC will equilibrate to the number of dollars it takes to erase the marginal thermal generator (at pure variable cost) and substitute a brand spanking new renewable generator at full life cycle cost. The REC will equilibrate to full life cycle cost of renewable minus variable cost of excluded thermal. That is the amount of money that will be required to induce entry of the requisite number of renewables. Without this type of model, IEMM would never figure this out. It could never figure out what the marginal thermal generator is, and it could never figure out what the marginal renewable process is. It could never keep track of the generation duration curve of the marginal (and infra-marginal) renewables processes entering at the lower right.

There are several noteworthy areas regarding the qualifying renewables generator.

1. Notice that there must be a strict 1-1 relationship between the number of physical MWh that the qualifying renewables generator delivers to the busbar market and the number of paper RECs that the qualifying renewables generator delivers to the REC market. They must print exactly one unit of REC for every one unit of renewable energy they deliver to the busbar. That is why the pink node has two outputs, and they are exactly equal in magnitude.

2. The renewables generator can and must add capacity. The capital cost, operating cost, thermal efficiency, book, tax, and financial parameters attendant with such capacity additions reside within the yellow box at the lower right of the diagram. Renewables owners will have to shell out capital dollars, operating dollars, etc. to build additional renewables capacity that will reside within the pink node at the lower right. This allows the IEMM model to input various different capital and operating cost estimates, perhaps varying over time. This is crucially important for renewables modeling. The capital cost of such renewables has to be considered and be endogenized in the price.

3. The generation duration curve (energy shape) of the qualifying renewables generator will have to satisfy the temporal pattern in Figure 136 for wind or the temporal pattern in Figure 142 for solar. Specifically, if we had 1 MW of wind capacity installed, it would create MWh according to the temporal pattern in Figure 136. The total area under that temporal curve would have to be calibrated to each 1 MW of installed capacity. It would represent how many MWh would be generated from a 1 MW wind turbine in every hour of the year at that site. Roughly 30 percent of the hours would generate 1 MWh for that hour, and roughly 70 percent of the hours would generate 0 MWh for that hour because the wind is not blowing. Using the generation duration pattern in Figure 136, one would simply input the output from the wind turbine per MW of installed capacity. As the model added wind capacity in the yellow node at the lower right, it would push this generation duration curve upward pro rata according to the amount of capacity the model would have to add.

   This is a highly nontrivial exercise for IEMM, yet it cannot be omitted or shortcutted. IEMM has to inject the specific time pattern of MWh into the physical MWh market that the renewable technology physically delivers. This is important, and for reasons made clear in the earlier sections on wind and solar. With wind, there are more MWh injected into the market at time of base than at time of peak. That tends to erode the load duration curve remaining to be served by the non-qualifying generators in the market. With solar, there are MWh injected into the market on or near time of peak but very little at time of base. That tends to improve the load duration curve remaining to be served by the non-qualifying generators in the market. It is crucially important to get both of these phenomena correct. Otherwise, the model will not properly reflect either the thermal non-qualifying fleet or the qualifying renewables fleet and, therefore, will miss the boat on renewables and their effect on prices, quantities, capacity additions and capacity in place, and plant decommissionings.

4. There are those who assert that electricity storage would change all this and would catalyze the entry of renewables, whose energy contributions could be leveled and firmed so that they begin to emulate what a thermal plant can do. We address that issue squarely below when we examine and recommend energy and hydro storage elements for the IEMM. To presage that discussion, tautological assumptions regarding the impacts of storage can be specious.

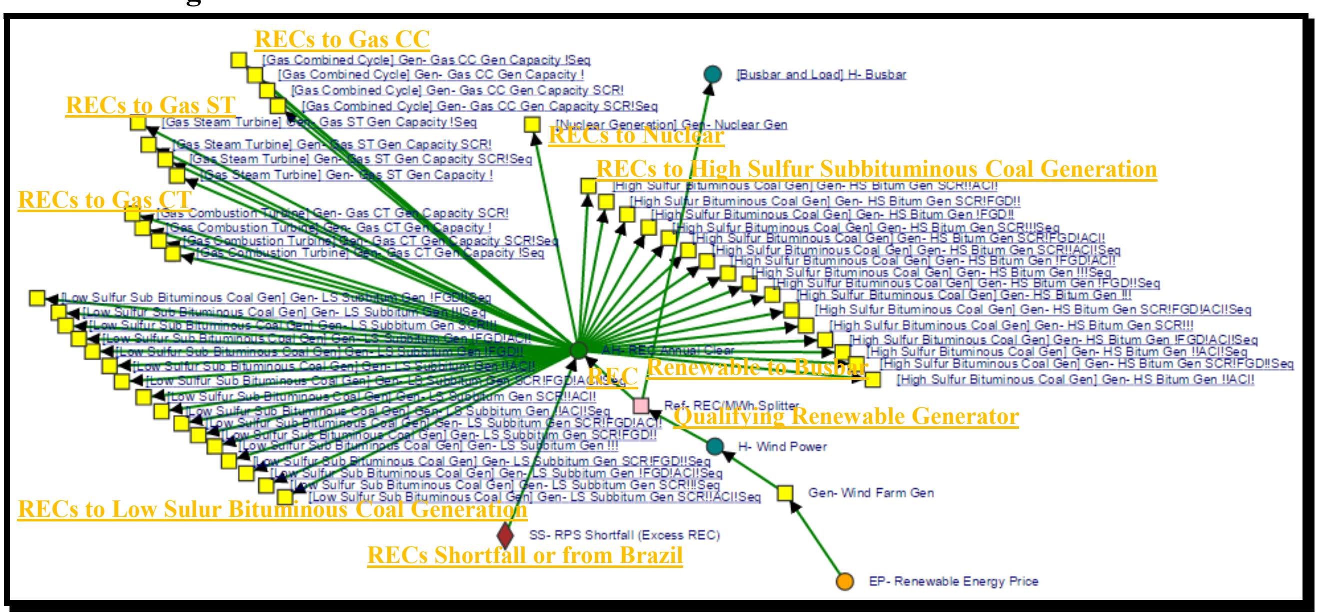
5. It is the surrender ratio in the non-qualifying thermal plants that sets the 20 percent (or whatever other) RPS EIA wants to implement. If that surrender ratio were to change say from 0.25 to 0.33, the renewable portfolio standard would rise to a higher number (25 percent). The magnitude of the RPS will be related to the redemption number for RECs by non-qualifying generators. In fact, if α is the required REC surrender rate in the nonqualifying plants, then the RPS achieved will be $\alpha/(1+\alpha)$. We can see that the structure in Figure 143, coupled with this nonqualifying plant surrender ratio, can simulate any RPS one might want.

6. If there are direct subsidies of renewables, the capital cost of the entering renewable is reduced. In that situation, the difference between the full life cycle cost of the renewable and the marginal thermal unit that is exiting drops, and the value of a REC will drop. In a word, subsidized-renewables entry drops the price of a REC. Any other answer is wrong. IEMM needs to represent that and with this design, it will. The reason is that renewables subsidies are ubiquitous throughout industrialized electrical systems, and they could materialize in any or all power systems throughout the world.

If we look at what the REC representation looks like for the electrical system with two coal types, coal generation, gas combined-cycle generation, gas combustion turbine generation, and nuclear generation, the diagram that is the equivalent of Figure 143 is that in Figure 144. Notice in the diagram that a single qualifying renewable occurs at the lower right in the pink rectangle. It delivers a certain number of MWh to the busbar and an equal number of RECs to the REC trading market. It delivers the MWh to the busbar according to the highly irregular, native pattern consistent with the native properties of the renewable source. It merely counts up the MWh delivered to the busbar for purposes of providing RECs to the traded market. The REC trading market offers RECs to every type of non-qualifying generator—coal generation burning high-sulfur bituminous coal, coal generation burning low-sulfur sub-bituminous coal, gas combined-cycle, gas simple cycle, gas steam turbine, and nuclear, all of which must surrender RECs for the right to generate. There is also a supply of "extra" RECs to simulate government policy that would allow the system to fall short of the RPS for some period of time and then catch up. (Those extra RECs could come from Brazil or some faraway, distant venue.)

The particular network structure in Figure 143 requires that renewables must be built at a fraction R percent of the total of thermal plants, and this will occur if the number of REC certificates that must be surrendered by each generator per MWh of generation is α=R/(1-R). We have been working with R=0.20 in the preceding examples. Lacking this number of REC certificates in the market, the price of the RECs would escalate to very high levels, and yet there could be no thermal generation because there are no RECs to de facto authorize it. The price of the REC at the margin would get so high that no thermal generation in the region at the margin is possible. If renewables were not built fast enough, thermal plants could not generate fast enough, and the price of a REC would skyrocket. People would pay a significant amount of money to get a new renewable into the market because to do so would be so immensely profitable and would leverage 4 MWh from nonqualifying generators into the market for every 1 MWh of qualifying renewable. This would accelerate the development of local renewables so as to obtain sufficient RECs to authorize operation of the thermal fleet. That would induce the renewable producer to enter the market. There would be a large profit motive for the renewable generator to enter the market. Not only does he generate profits from energy sales, but he also generates profit from sale of RECs that thermal generators need to be able to generate at all. This would indeed induce renewable entry so that the RPS is achieved, after which the value of the REC will drop to the point at which the marginal renewable entrant would be just indifferent between entering and not entering. We have never seen a model (other than this proposed design) do this.

Figure 144: RECs and Renewable MWh Produced and Consumed



We can see how the RPS fraction can be maintained through the emanation, trading, and consumption of offsetting RECs that are properly priced by the structure in Figure 143. This trading mechanism is easily built into IEMM model under this design, and it will be invaluable. When we do so, we are putting the capital and operating cost of the renewable plant into the model, and the model will establish a price for RECs and a price for energy that will induce the renewable producer to enter and produce at maximum possible output. There is no substitute for this as EIA considers world and North American systems that are policy-directed to meet a particular level of renewables. The direct benefit in the North American system is immediately obvious.

10 TRANSMISSION

Thus far, we have discussed generation and load. This section puts forth a recommended model of transmission for IEMM, the crucially important portion of the electric system that lies between generation and load. Transmission renders electricity a spatially distributed economic system in the classic sense.[23] Transmission is an economic agent, moving power from Point A to Point B, an agent no different from any other. We will discuss the significant aspects of transmission and spatiality in this section with regard to electricity. It is an obviously important issue in large and more advanced systems and countries. However, emerging systems and partially Balkanized systems such as China that have prospects of integrating increasingly have to represent transmission because they generally have more remote generation connected to load centers via transmission. Transmission must be carefully considered and represented, and this design shows how. The previous GHySMo CDR focused to a significant degree on transportation and logistics in a generic sense.

10.1 How Should IEMM Emulate Transmission?

The diagram in Figure 1 is fundamentally local in nature. However, the diagram is careful to include all the inbound transmission into the particular busbar/region and all the outbound transmission from the particular busbar/region to a contiguous region. Production simulation people (i.e., power dispatch people) would be quick to assume that transmission should be represented as a "resource" in the particular busbar/region, not particularly different from a generation plant. That is simply not correct, as we shall examine here. In fact, it is mortally incorrect; transmission is not a resource in the same sense that local generation is a resource, not in the slightest. That is one of the difficulties of typical production simulation and dispatch models. Transmission cannot be treated as a resource, a plant equivalent, in IEMM. Transmission is fundamentally different from generation, just as it was fundamentally different from supply. IEMM needs to represent operation, "dispatch," and adequacy of the transmission system carefully and accurately as a proactive player as articulated here. Such modeling gives deep insight into how the transmission system will and should be operated, where any congestion points might exist, and how and whether to overcome them. It will also tell us the distributed prices and the tariffs all around the transmission system.

In Figure 1, a simple relationship at the particular busbar/region will hold: Local supply plus inbound transmission plus withdrawals from electrical storage equals local demand plus injections to electrical storage plus outbound transmission. In the diagram, as well as in every particular transmission system in the world, some of the composition of the supply stack is represented by **inbound transmission**, and some of the composition of the demand tranches is represented by **outbound transmission**. That is pointedly clear in Figure 1. Transmission is fully embedded within and endogenous to the IEMM model. The model directs transmission **plus** generation **plus** withdrawals from electrical storage into a market, and it directs net energy for load **plus** ancillary services (reserves) **plus** injections to electrical storage **plus** outbound transmission out of a market. The material balances (losses) and the economics (tariffs and costs of losses) are preserved on the

[23] The GHySMo ArrowHead CDR talked at length about spatial equilibrium. The concepts put forth there apply directly, with the extension here related to temporality and just-in-time, real-time delivery issues.

interconnecting transmission system. This is unique to the proposed IEMM model design, which is calculating a so-called "Samuelsonian spatial equilibrium" across the transmission system, that answer referring to Nobel Laureate Dr. Paul Samuelson and his famous proper characterization of spatial equilibrium where transmission systems are concerned. His famous 1956 paper should be required reading for everyone in the power and energy modeling business. It tells precisely how a transmission system should be modeled in an economic sense. A careful reading of that paper shows how the transmission system should be represented in a spatial sense. A careful reading of that paper shows why the transmission system should probably not be represented using a linear program because of the limiting assumptions one has to make.

ArrowHead and our predecessors have modeled complex transmission grids in literally every industry in the world:

- Electricity
- Natural gas pipelines
- World LNG shipping
- World crude shipping
- World gasoil shipping
- World petroleum product shipping
- Airline origin and destination modeling (United Air Lines, Continental Air Lines, Air France, SAS)
- Truck origin and destination modeling (Yellow Freight)

One key characteristic of electricity that differentiates it from all other types of transmission is that with electricity there are not well defined "froms" and "tos." There are a large number of competing, parallel, and nonparallel paths. There are a lot of froms. There are a lot of tos. There are a lot of upload (energization) points. There are a lot of download (de-energization) points. One of our electric clients once quipped: "Electric transmission is like a spatial clothesline hanging up there, the wires being the conductors. There are places where you energize the clothesline by uploading from a power plant or withdrawing from storage. There are places where you de-energize the clothesline by downloading to a load or an ancillary service or inject to storage. All you have to worry about is not trying to take too much electricity from one energizing point to one de-energizing point. You'll melt 'er down. We'll put a breaker on there to make sure you don't, and that breaker'll be the maximum throughput we'll allow." That is a very good physical characterization, and it leads to an effective economic characterization for IEMM as well.

The relevant diagram for how we represent the interplay between generation and transmission is that in Figure 145. The diagram in Figure 1 is implemented at each of the nodes in the diagram in Figure 145. These are the so-called "inputs" to each of the local nodes and "outputs" from each of the local nodes. We have depicted the nodes in shorthand form as shown in Figure 146. That is, every node in Figure 145 has the orientation in Figure 146, and each and every node in Figure 145 has the representation in Figure 146. There is local generation, which contributes to the supply stack. There is inbound transmission, which contributes to the supply stack. There is local net energy for load and ancillary services/reserves, which contributes to the demand tranches. There is withdrawal from storage at certain times. On the other end of the spectrum, there is outbound transmission, which contributes to the demand tranches. There are injections to storage at certain times. This is disarmingly simple but obviously correct as a nodal representation of the power system from an economic as well as physical and electrical perspective.

Figure 145: An Interconnected Generation-Transmission-Load System in IEMM

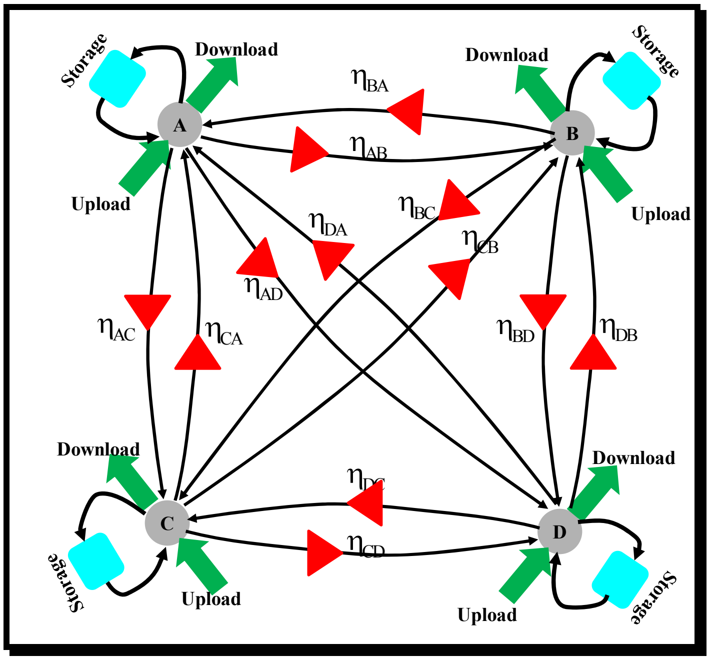
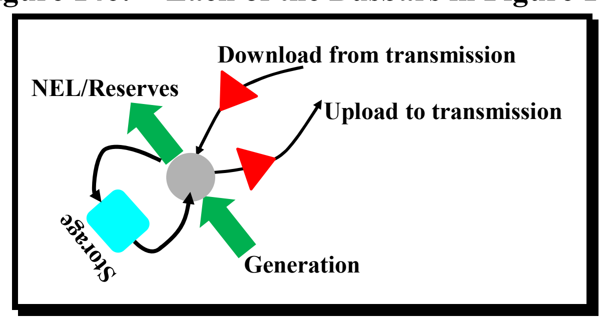


Figure 146: Each of the Busbars in Figure 145



This charactrization of IEMM transmission s unique in its ability to represent each local or nodal generation business by representing the logic in Figure 1 economically as in Figure 145 and at the same time forcing complete lockstep physical (MW and MWh) consistency and economic consistency (costs of losses and tariffs along the transmission system) as shown in Figure 146. In Figure 145, we have used the notation η to indicate that **every** transmission linkage in Figure 145 must be characterized by its own unique bundle of losses and economic costs. The use of the common notation η is not intended to communicate that every transmission link has the same losses (attributed or actual). It communicates that a cost and loss factor is present in the IEMM on each and every transmission link in the model. This diagram shows how the IEMM solution maintains the economic logic in Figure 1 at every node in the region that IEMM will be modeling while at the same time respecting all the loss (energy), capacity, and tariff relationships across each and every link of the transmission system and thereby across the transmission system as a whole.

The structure in Figure 145 and Figure 146 is precisely the structure of large, multi-country, multiregional transmission systems or small transmission systems for individual countries or regions, both of which must be built within IEMM. Balkanized systems will have their own inherent transmission structure, but they will not be connected to other transmission structures.

As far as we know, IEMM thus designed is the only model in existence that will calculate a true zero arbitrage solution over time and space, a true Samuelsonian spatial equilibrium, across the transmission system. Located pricing (i.e., pricing at the hubs in Figure 145 and Figure 146) and understanding intrinsically the relationships of those prices across the bidirectional conductor systems that compose the transmission grid is crucial to electric system transmission modeling, and IEMM has it in spades. Our recommended method, in the purest sense, **is** Samuelsonian spatial equilibrium (named after the first non-Norwegian economics Nobel laureate Dr. Paul Samuelson). No other transmission model makes such a claim as far as we know; they simply do not make the Samuelsonian calculation. Any approach other than Samuelsonian will be inaccurate, likely fatal, for understanding how to operate a transmission system for economics and reliability. Non-Samuelson approaches do not represent the economics or physics of transmission at all. They treat it completely arbitrarily. They end up with the wrong capacity in place, the wrong costs, and the wrong prices and basis differences.

This section articulates some of the concepts and misconceptions within transmission modeling as it has been practiced in the past and illustrates the key features of transmission modeling within IEMM.

- **Transport (Contract) Flow?** Contracted transport flow is explicitly modeled in the IEMM system. We described precisely how contracted flows were modeled in the GHySMo CDR. Contracted flows are, by agreement, the most infra-marginal flows in the system. Contracted flows dominate all other flows and come first. The IEMM construction has contracted flows nailed.

  Contracts are included in the general equilibrium solution as constraints or as parallel transmission systems, one of the two parallel systems being pre-contracted and forced to operate and the other being non-contracted and therefore completely discretionary and "spot" oriented. This is an important contribution of the Samuelsonian spatial equilibrium concept. Pre-contracted transmission entitlements and flows distort market clearing prices at both ends of the transmission system because they can force uneconomic carriage. IEMM, as designed here, will calculate the precise magnitude of such distortion and predict the existence, stability, and persistence of such distortion. We mark those transmission contracts to market at the distorted prices that result in the market in which those contracts are forced to be honored.

  This is critically important—contracts distort markets in particular ways, and such distortions are important to quantify and understand. This will give IEMM a singularly accurate estimate of the spatial and temporal economic impacts of pre-contracted transmission. There is another point that is worthy of mention. If the price dislocation caused by the contract distortion gets large enough, there is a lot of pressure brought to bear on the contract. In the vernacular of the oil industry, "If you have an out of market contract, you can look right down the pipe and not get your face wet." Out of the money contracts tend to get renegotiated, sometimes with regulatory assistance and sometimes not. By calculating the magnitude of price and market distortion, our users can understand when contract terms are likely to become imperiled. The IEMM design (which includes contracts as discussed in detail in the GHySMo CDR) applies directly here. IEMM has contracts nailed. We do not believe that is easily possible in an MPEC or complementarity global social welfare maximization context.

- **DC versus AC Emulation.** The IEMM is what is called a DC or energy flow construct. It is not a full AC sinusoidal wave-flow construct. That is never what is needed for most power systems in the world. Banish the thought! DC lines and converters are specifically distinguished and included as transmission links in the spatial IEMM model. Every DC line and its contribution to the spatial equilibrium calculated for the entire system is included. The second question is more subtle. The IEMM model completely represents and balances the real component of power flow, i.e., what electricity models term DC flow or energy-only. As such, it honors the first and second laws of thermodynamics and their electrical equivalent Kirchoff's law and extended Ohm's law. In the vernacular, losses and constraints in energy delivery are properly represented. Reactive power or frequency support is not represented.

  It is interesting as an example to consider what "reactive power" is and why it is not and likely will never be included in an economic model of electricity[24]. Alternating current (AC) is supplied in a 60Hz waveform. Reactive power is produced when the current waveform is out of phase with the voltage waveform due to inductive or capacitive loads. Current lags voltage with an inductive load, and leads voltage with a capacitive load. Only the component of current in phase with voltage produces real or active power that does real work. Current is in phase with voltage for a resistive load, like a resistance heater. Reactive power is necessary for producing the electric and magnetic fields in capacitors and inductors.

  The additional current flow associated with reactive power can cause increased losses and excessive voltage sags. Transmission system operators have to ensure that reactive reserves are available to handle system contingencies such as the loss of a generator or transmission line because increased current flows after the occurrence of contingencies can produce greatly increased reactive power absorption in transmission lines.

  When different sources have different phases, real world power systems have to put them back together with capacitors and inductors and the like. If they do not, they can have voltage excursions on their system, and that can shut down or even damage part of the system. These types of considerations require highly detailed sinusoidal and differential equation analysis of the physics of a transmission, generation, and load system. Neither our IEMM design nor any other economic model will have things like reactive power in their solution. The best we can do is constrain throughput on the transmission system to account for reactive power and other

[24] http://web. ornl. gov/sci/decc/RP%20Definitions/What%20is%20Reactive%20Power%20_ORNL_. pdf

constraints, and we can represent any additional energization that might be required. Such is the state of economic modeling of power systems.

- **<u>Right Sizing of the Transmission System</u>.** With regard to transmission, the IEMM model will build new transmission lines (if you enable the "new build" switch that will be in IEMM under this design for transmission), and it will retain the capacities and routes on the old lines only (if you turn off the "new build" switch on transmission). If you want IEMM to build and properly size the transmission system (and you will) and the generation system simultaneously, you have to estimate the capital cost, operating cost, and losses of entry not only for generation units but also for incremental new transmission entrants. That is, you have to estimate the capital cost, operating cost, losses, and financial and book parameters of every possibly entering power line link. There are estimates of many of those links, but EIA will merit from review these links in IEMM as you use the model to make better decisions.

The IEMM model is not a "physics of power flow" model such as GE MAPS or PowerWorld, which were originally designed to represent power system physics and sinusoidal waveforms. GE MAPS, PowerWorld, and PTI are Kirchoff's-Law/Ohm's-Law-Impedance-Inductance-physical wires and switches and buses or complex math models. Those types of models are not economic models, and they are not reliability models, and they have little hope of being either on a go forward basis. While they are indispensable for understanding bottlenecks and congestion points in transmission and serious analysis of interruptions that might occur, power flow models are prohibitively complex and massively detailed and are **<u>not</u>** economic models and contain **<u>no</u>** agent-based economic component. They are of little help in running a transmission system. The Optimal Power Flow or OPF elements of those models to not represent the true economics of operation of those systems and thus will not be

While IEMM is not a pure physics model of the power transmission system, it accepts the output of such models in simulating real world constraints and bottlenecks in transmission and generation and load. Furthermore, entry of new generators occurs rapidly, and retirement of old generators occurs just as rapidly, perhaps more so. As this occurs, new entrants tend to situate systematically in certain preferred locations. This is going to be exacerbated in the future as gas becomes more of a challenge in various regions of North America and the world.

All the production function considerations and all the capacity addition considerations pertain to the elements of the transmission system. This is uniquely invaluable, allowing transmission costs to be set by variables costs of old capacity as contrasted with full fixed plus variable cost of new capacity.

Getting back to power transmission as an energy flow problem, assuming that reactive power and other difficult physical issues have been resolved and input as constraints or retardations in the economic model, we recognize that the imposition of an LP, MPEC, or complementarity model on a physical model is economically wrong, for it assumes the existence of the proverbial dictator pulling the strings on each and every investment, operation, and retirement decision on every piece of equipment and every customer in the system. The sufficiency conditions, examined in the GHySMo CDR whereby there might be consistency, are literally impossible to hold in practice.

To say (or hope) that they will is wishful thinking in the extreme. To reference that CDR and argue that global social welfare maximization and network economic equilibrium is specious and incorrect. The behavioral equations for electricity are simply not integrable in the required sense of Samuelson.

10.2 Alternative Transmission Tariff and Pricing Structures

One of the strongest features of the IEMM transmission model is that it can represent a broad combination of postage-stamp, distance-based, zonal, or hybrid electric transmission pricing methods. (Although not addressed nearly as frequently as the constraints or bottlenecks of transmission, the tariffing and regulation of transmission can be critically important for IEMM.) We can insert into IEMM an accurate and important model of the various regional markets that allows explicit representation of postage stamp and other pricing of transmission.

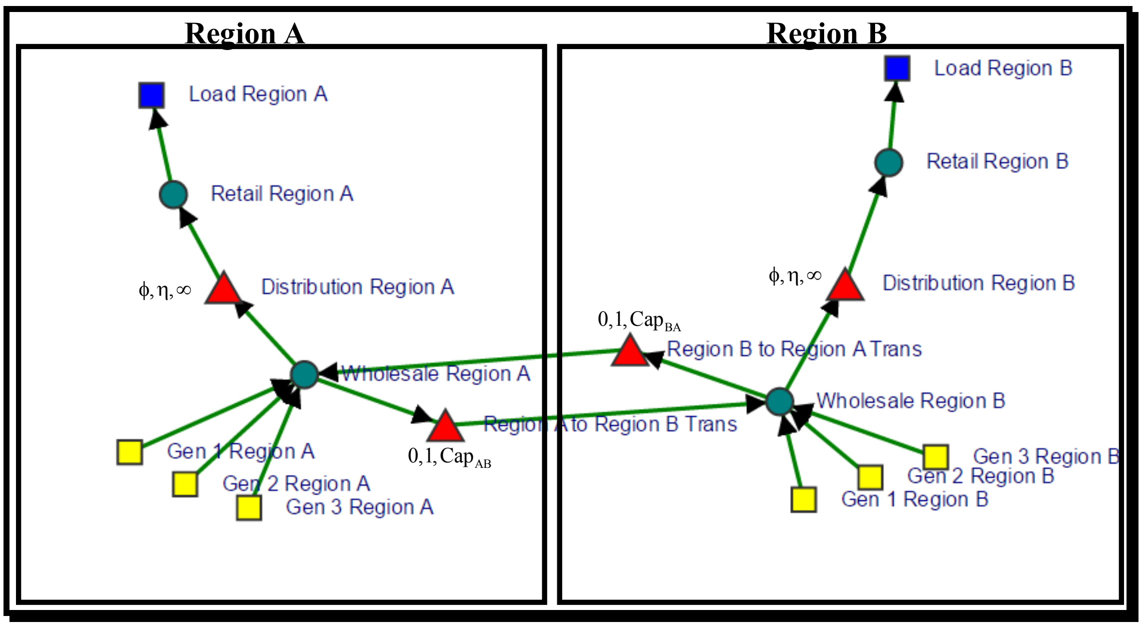
Transmission link tariffing is a crucially important issue, and there are a number of variations and distinctions IEMM has to consider. One of the most frequent forms of transmission tariffing is what is called "postage stamp" tariffing. Under postage stamp tariffing, the cost of transmission is the same from any generator anywhere to any user anywhere. Just like at the Post Office, you buy a "stamp" and you get to move your power from any generation point to any load point, whether they are 1,000 yards apart or 1,400 miles apart. It is quite akin to a first class letter; the same stamp can take your letter from here to Miami or to take your letter to a destination 2 miles away. How might we represent postage-stamp tariffing of the transmission system? Using the notation that a triangle is a transmission link, consider the IEMM structure in Figure 147. The three numbers associated with each transmission triangle represent the following

- The variable (volumetric) cost of the transmission link, excluding losses
- The thermal efficiency of the transmission link (i.e., 1 minus the percentage loss along the link)
- The capacity along the transmission link

Thus, the triplet (0, 0.975, and 1,750) would represent a transmission link with 0 volumetric cost, 97.5 percent throughput (2.5 percent losses), and 1,750 MW of capacity.

Using this notation, consider the orientation of the transportation links in Figure 147. Notice that the generators in Region A at the left in Figure 147 are clustered near the market hub that feeds net energy for load into Region A. They pay no cost to get onto the transportation system locally. They energize the transmission system in Region A at no cost. If their power is uploaded from the Region A busbar to the Region A load, the triple that comes into play is (ϕ, η, and infinity). That means that there is a nonfuel volumetric charge of ϕ, and there is an assigned loss of η, but there is no upper bound on capacity that can be utilized for uploading. That means that a generator in Region A will face the cost of ϕ and an assigned loss of η to move power from generators in Region A to load in Region A. Consider power that is produced in Region B. That power is uploaded for free to the busbar in Region B. That power moves to the busbar in Region A according to the triple on the B to A transmission link (0, 1, and Cap_{BA}). That means there is no nonfuel cost and no assigned losses, and there may be an upper bound on capacity of Cap_{BA} that represents the maximum physical flow through the B to A transmission link. When that power gets to the busbar in Region A at the downstream terminus of that transmission link, it is uploaded to Region A load in precisely the same way as before, according to the triple that comes into play (ϕ, η, and infinity). The customer in Region A will have faced the cost ϕ and the assigned loss of η to move power from generators in Region B to load in Region A. The customer in Region B will have paid the same transmission cost whether he bought power from a generator in Region B or Region A. This is postage-stamp transmission pricing according to its intrinsic definition—a single cost on and off the transmission system no matter where the on point and where the off point are physically located. He pays the same transmission cost regardless of where his power got onto the transmission system. Figure 147 is a perfect representation of postage-stamp ratemaking. The upload-transmission-download cost is precisely the same no matter where the uplift or downlift.

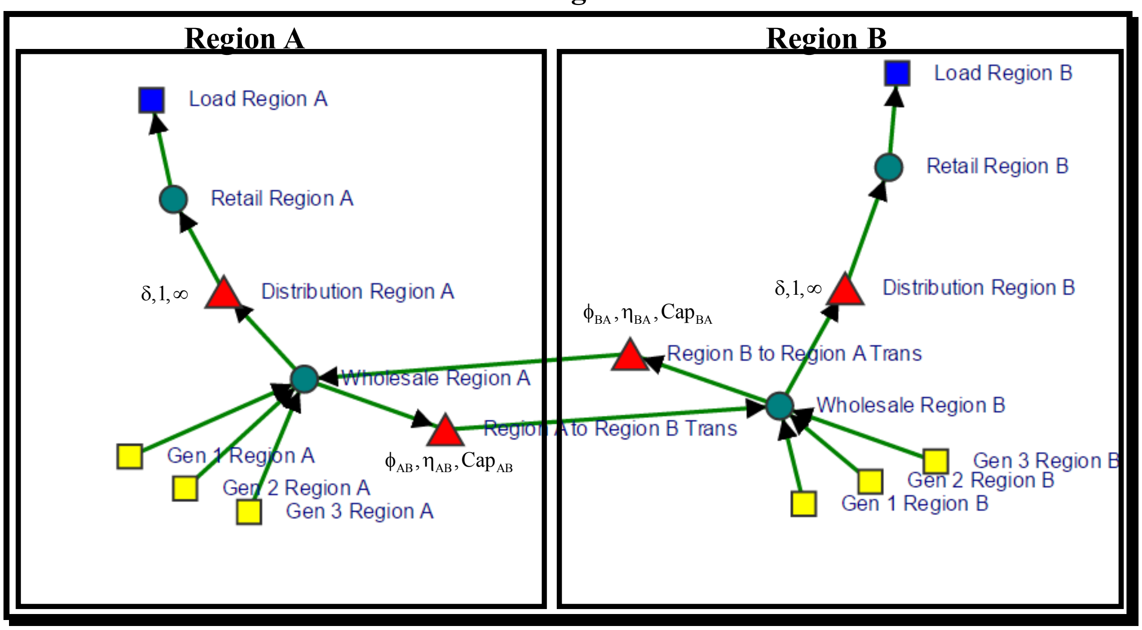
Figure 147: IEMM Represents Interregional Postage Stamp-Distance-Zonal Ratemaking with Constraints



Let's look at a slightly different situation, a "pancaking" situation with distance dependent electricity transmission tariffs. Consider the orientation of the transportation links in Figure 148. Notice that the generators in Region A at the left in Figure 148 are clustered near the market hub that feeds net energy for load into Region A. They pay no cost to get onto the transportation system locally. They energize the transmission system in Region A at no cost. If their power is uploaded from the Region A busbar to the Region A load, the triple that comes into play is (δ, 1, and infinity). That means that there is a nonfuel volumetric charge of δ (thought of as a distribution charge in Region A), and there is an assigned loss of 1, but there is no upper bound on capacity that can be uploaded. That means that a generator in Region A will face the local distribution cost of δ and an assigned loss of 1 to move power from generators in Region A to load in Region A. Ostensibly, it will be inexpensive for local generators to serve local customers in Region A because they do not have to get onto the inter-regional transmission system.

Consider power that is produced in Region B. That power is uploaded for free to the busbar in Region B. That power moves to the busbar in Region A according to the triple on the B to A transmission link (ϕ_{BA}, η_{BA}, and Cap_{BA}). That means there is an interregional transmission tariff of ϕ_{BA} and assigned losses of η_{BA} necessary to move power from B to A. There is an upper bound on capacity from B to A of Cap_{BA} (if applicable). That ϕ_{BA} and loss η_{BA} have to be paid by anyone moving power from B to A. Once the power gets to Region A, the local distribution charge of δ will pertain. Notice in this construct, customers who get their power from generators in Region A pay no transmission charge and only the distribution charge δ. Customers who get their power from generators in Region B pay the transmission charge ϕ_{BA} plus the distribution charge δ. They pay more by the amount of the inter-regional transmission tariff in the form of a pancaked cost ϕ_{BA}. The term pancaking means that if one grafts two or more of these types of regional structures together in series, the tariffs will add, sum, or "pancake" one on top of the other.

Figure 148: IEMM Represents Interregional Distance Dependent (Pancaked) Ratemaking with Constraints



It should be obvious from the structure of Figure 148 that it is straightforward to represent zonal or distance-based rate pancaking of the transportation system. If we link similar structures together for more than one region that span say the WECC in the United States or Europe, we obtain a sequenced pancaking system. This is precisely the way multi-region transmission works (and not just for power transmission. It works precisely the same for gas pipeline transmission.) Doing so would cause the IEMM model to properly represent interregional tariff pancaking no matter what country or set of countries EIA was attempting to model. Obviously, the structure in Figure 147 and Figure 148 allows a rich array of existing and possible future transmission pricing. It allows IEMM to consider alternative representations of transmission systems, alternatives that can impose significant value differences on various plants in the region.

10.3 Transmission Congestion

People have amplified the preceding analyses to inquire about constraints or “congestion.” They assert that for studies of transmission EIA might be working on, you need to be able to capture the costs and constraints of congestion. When a located price is calculated, is it the cost of the energy separable from the induced value of the congestion? Can the model decouple cost from congestion rent? Are the located prices equal to the cost of energy plus the cost of congestion plus losses? Is there scarcity rent or congestion rent coming into play here?

There has been a lot of misinformation about located or nodal prices, much of it put forth by LP or dispatch models. IEMM calculates the congestion value in terms of the market basis differential across each and every transmission linkage. This is the quintessential market definition of congestion: the market price differential across each and every asset. The 2009 Nesbitt and Scotcher paper makes this point very well. From an economic perspective, congestion means that the asset is running at full capacity for a given hour (or another period of time), and the price differential across the asset sustained by the market exceeds the cost of the asset for its last hour of operation. That is, if there were one more smidgeon of capacity during that hour, you could sell it for more than its marginal short run cost (but not necessarily at a price that would offset the capital cost of new entry). The market wants more production than you have, but it may or may not be willing to pay the capital cost for more production. The difference between the price differential across the asset and the cost of deploying the asset is the value of congestion rent, and IEMM (as defined here) calculates it explicitly and endogenously. We are not aware that anyone calculates it dynamically in the way that we have done here. One would think that systematic transmission rents would never exceed the cost of capacity expansion in the long-term, and that would be right unless transmission capacity expansion were precluded by policy, land use, right of way, or other noneconomic reason.

Within IEMM, and within our economic training seminar, we view each transmission link as having a "supply curve for transmission service." Figure 149 illustrates. (This is totally consistent with profit maximization subject to a production function constraint, as we articulated in detail in the GHySMo CDR and will not repeat herein.) The price of electricity at the upstream end of the transmission link is the green curve at the bottom. The transmission cost is the vertical height of the black curve minus the green curve. We see the notion of a bounded quantity of pipeline capacity drawing from a supply of generation at the upstream end of the transmission link.

When the demand curve cuts through the transmission supply curve on the flat portion as in Figure 150, this is a situation of plenty of transmission capacity. We see that the price becomes the cost of electricity plus the cost of the transmission link, exactly as one would expect. Supply is elastic because the transmission link is not fully deployed. Price obeys all the rules people have become used to seeing. However, this asset is not a resource in the classical production simulation/dispatch sense. If the throughput goes up, the price of power in the region upstream from the transmission link goes up. That is not a traditional dispatch resource.

Figure 149: Supply Curve for Transmission Service

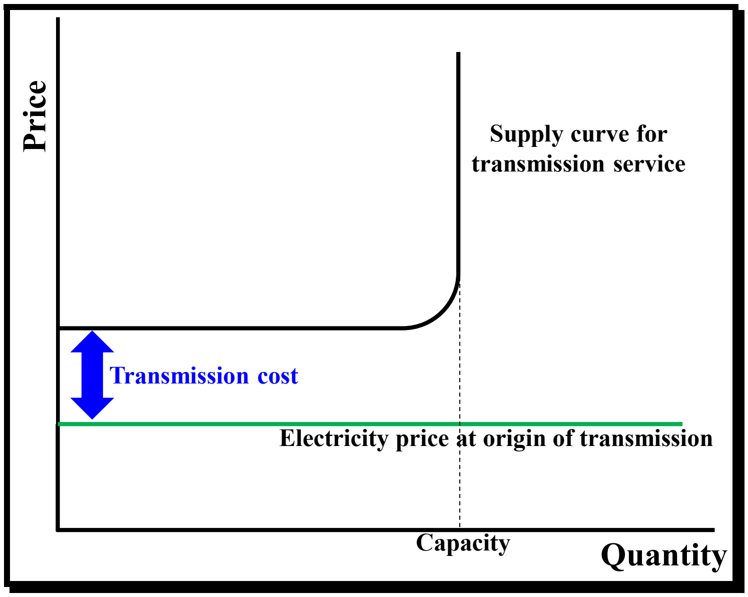
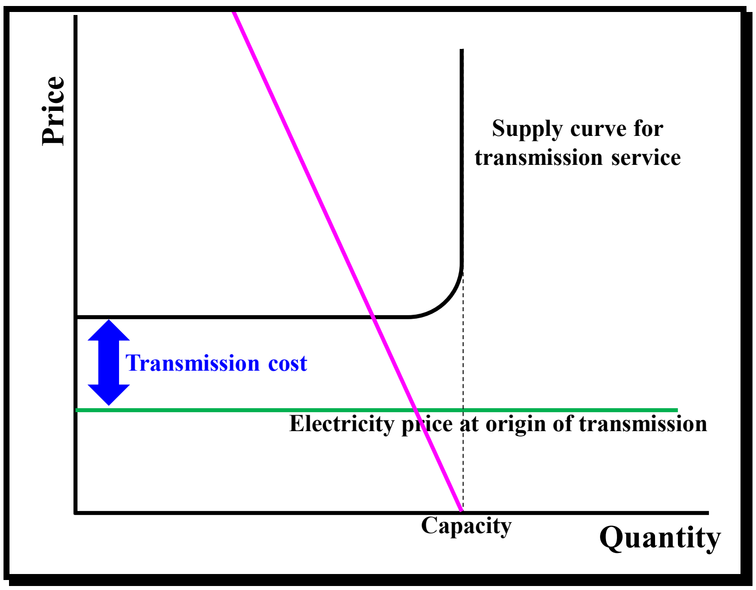
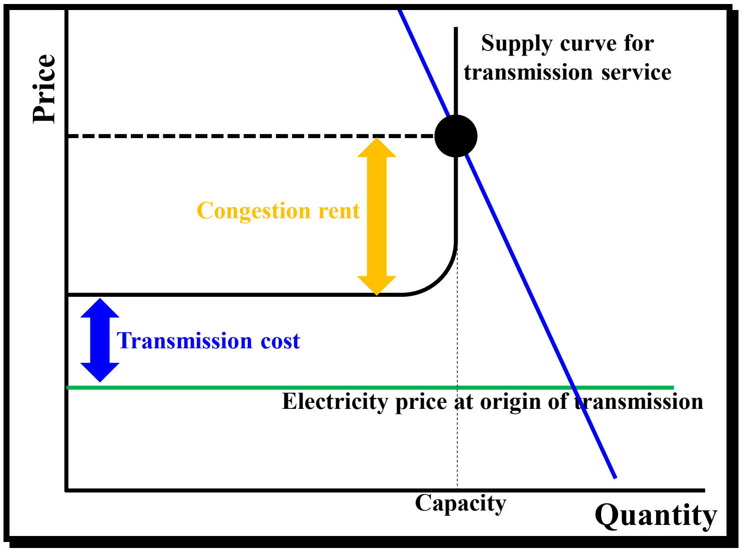


Figure 150: No Congestion on the Transmission Link



When demand walks upward toward the extreme right-hand end of that supply curve for transmission service as in Figure 151, IEMM calculates where the demand curve crosses the transmission service supply curve and thereby endogenously calculates the congestion rent. The price in the market is above the cost of electricity plus the cost of transmission, substantially so when there is congestion as in Figure 151. In such a situation, IEMM prices that transmission link on a competitive basis (as does the real world) and that price contains some congestion rent. Congestion rent is not necessarily the horrid thing that regulators so frequently criticize. It is simply a manifestation that the market is not "overbuilt" to the point at which there is a super-abundance of transmission capacity relative to demand for that transmission capacity. This is the normal situation. Why in the world should transmission be overbuilt? The IEMM design put forth here calculates the congestion rent if there is any or the normal, abundant supply transmission cost if there is none and thereby gets the right answer for transmission.

Figure 151: Congestion on the Transmission Link



As an example, we have calculated calculate congestion rents from the PJM West to PJM East transmission links, taking account of the transfer capability of those lines, possible generation and transmission entry and retirement in the eastern PJM regions, and possible entry and retirement in the western PJM regions. We have also calculated the nodal value of every power plant in the region, both existing plants and possible new entrants. We have disaggregated into a number of subregions with fairly complete representation of generation at each of the regions, transmission capability between and among the regions, possible entry and exit in each of the regions, and the value of transmission entitlements on every link within the region. The model is well positioned to value transmission entitlements, potential new build sites, potential plant purchase decisions, and in a nutshell how to negotiate from a position of informational strength with other companies. The IEMM design combines not only the physics of transmission but also the congestion (if any) and tariffing.

In sharp contrast to the omnipotent dictator view embedded in global social welfare maximization approaches as they (invariably) depart from the Samuelsonian sufficiency conditions, IEMM (as defined here) takes the textbook economic view. Each and every asset owner takes the locational, nodal prices for his or her **product** (e.g., delivered electricity) as given, takes the locational, nodal prices for his or her **feedstock** (e.g., purchased electricity in the generation region) as given, and sets his or her personal investment, operation, and retirement decisions so as to maximize his or her personal profits. Each and every agent in IEMM and in the particular market pursues his or her own profit bottom line. IEMM is an agent-based approach that assumes that all agents seek individually and independently to maximize profits. The market prices that come from the competitive "war game" as that results are the true market prices and bases that will occur in markets. If there are 1,000 agents in the IEMM representation of a market, there will be 1,000 simultaneous profit maximization problems going on within IEMM (and assuredly in the real world). IEMM gives EIA the agent-based, general equilibrium, "war game" solution, which is the sole correct solution in the world of located prices. To encourage individual behavior consistent with an exposed, transparent pricing system is allegedly a goal of RTOs/ISOs around the country. They do so in part by allowing locational bilateralism to expose current and forward prices locationally. In closing, it is well to consider that this profit-seeking approach is just as germane for transmission as it is for generation or end-use or storage. The IEMM profit-seeking approach stands in sharp contrast to global social welfare maximization, which in practice has difficulty with transmission.

11 RESERVOIR HYDRO, RUN OF RIVER HYDRO, AND PUMPED HYDRO FOR STORAGE AND LOAD-SHIFTING

Hydroelectricity simply cannot be ignored in IEMM. It has to be represented accurately and correctly, sometimes front and center. There is no choice. This is particularly true because so many developing and some developed countries have so much hydro. Hydro is one way to store electrical energy. It is stored as potential energy, but it is stored in a hydro facility, make no mistake. Hydro is fairly cheap in some countries, and that is why it has germinated over the years.

11.1 Electricity Storage

The notion of storage, electric or otherwise, is that of a temporary holding tank, a container that holds a valuable commodity. Economic agents buy a valuable commodity from the market and place it into that holding tank at certain points of time when the commodity is in excess supply and the price is low, and economic agents withdraw from that holding tank at other points of time when the commodity is in short supply and the price is high and sell it into the market at that point in time. Storage operators are de facto "buying low and selling high." They are arbitraging temporal prices in the market. They are shifting the temporal relationship between supply and load, physically and economically. This can cause rather dramatic changes in prices, especially for a commodity like electricity.

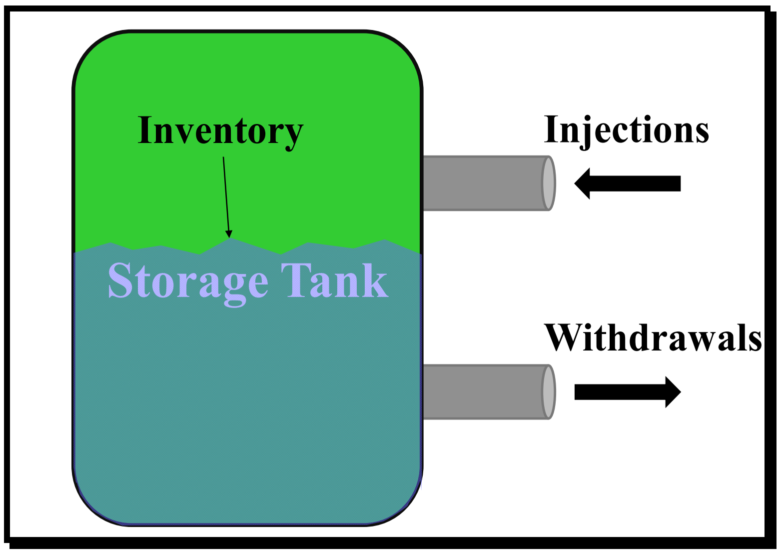
To wit, storage is an activity in which at some period of time, a person systematically buys a valuable commodity from the market and puts it into a temporary holding tank (e.g., capacitor, hydro reservoir, gas storage) or in the case of hydroelectric an activity in which "mother nature" puts a valuable commodity into a temporary holding tank (e.g., hydro reservoir high in the mountains in which snowmelt and runoff charge the reservoir). There is a notion of addition to storage, sometimes called "injections." A valuable commodity is bought from a market and injected into the temporary storage tank. The point of storage is to "store" things for later use.

It merits specific mention that with storage there is always the notion of a "tank," generally a tank with limited capacity. The reservoir has a certain maximum size for storing the commodity; no more can be held than the designed size of the tank. An operationally binding bound on the size of the tank is not always necessarily true, however. (We will see below certain examples where the capacity of the storage tank does not come into play.) There does not have to be an effective limit on the tank size relative to the aggregate injections and withdrawals that will be desirable in the market. The size of the tank might well be larger than the market could ever be able to use. This is not to say that there are "infinite" tanks in existence. It is to say that the size of the tank might well be much, much larger than the market can absorb via injection and dispatch. In that situation, it will be the injection and withdrawal rates that will take precedence over the tank size.

There is also the notion with storage of withdrawals from the market when supply is short and prices are high. Ostensibly the function of storage is to augment supply when supply is short and prices are high so as to have additional commodity available to market at the time when prices are high. You put the commodity away when markets are abundant so when that "rainy day" comes, you will have more abundant supply on hand from the market and from the storage field to meet the higher demand on that rainy day. The notion of storage intrinsically contains the notion of "withdrawal" at certain times when the commodity is needed or more valuable.

Figure 152 illustrates the tank at the left and injections to the tank from the right. Thus far, the notion of storage, tank contents, injections, and withdrawals is not particularly difficult. You store commodity you have bought from the market in a tank at certain periods of time and you take commodity out of the tank at other periods of time and sell it to the market. Your decisions will be constrained by the maximum rate of injection, the size of the tank, and the maximum rate of withdrawal. This is simple in concept, yet very sophisticated and difficult in practice, sufficiently so that it is not amenable to back of the envelope guesswork.

Figure 152: Storage Is a Dispatchable "Tank" with Injections and Withdrawals



Consider what happens in markets in which there is no storage, the situation indicated in Figure 153. In such markets, there are a number of sources of commodity coming into a local market hub. There are a number of destinations or uses or outflows of the commodity from the market. The market is an "exchange," a place where all the supply sources meet all the demands, and a price occurs in real-time so that the sum total of supply is exactly equal to the sum total of demand in that market. The various destinations are indicated as outputs (destinations) from the local market hub. In the absence of storage, the situation would be very simple. The sum of the inputs must be exactly equal to the sum of the outputs—the market must balance or equilibrate in real-time, every instant of time in fact—and supply would have to be exactly equal to demand at every instant in time. There would no way to temporally change or buffer the time pattern of supply to meet a different time pattern of demand. The situation in Figure 153 represents a just-in-time, real-time supply and demand balancing situation with no buffering through storage present in the market.

The way storage works in an electrical system and in a market (e.g., gas storage, gasoline storage to meet summer seasonal peaks in gasoline demand) augments the simple situation in Figure 153 as shown in Figure 154. The key augmentations are several.

1. We still notice the notion of a number of sources of commodity coming into the local market hub. If you didn't have sources of commodity into the market, there would be nothing to store! Storage does not create commodity. Storage merely intercepts commodity from the market, puts it in a tank for some period, and inserts that commodity back into the market at a different time. (With hydro, water materializes from seasonal rainfall and snowmelt so that it has zero price. Notwithstanding, it is still a valuable commodity that injects to storage. To wit, injection is either random or systematic, but withdrawal is always systematic.)

Figure 153: Without Storage, Markets Equilibrate in Real-time

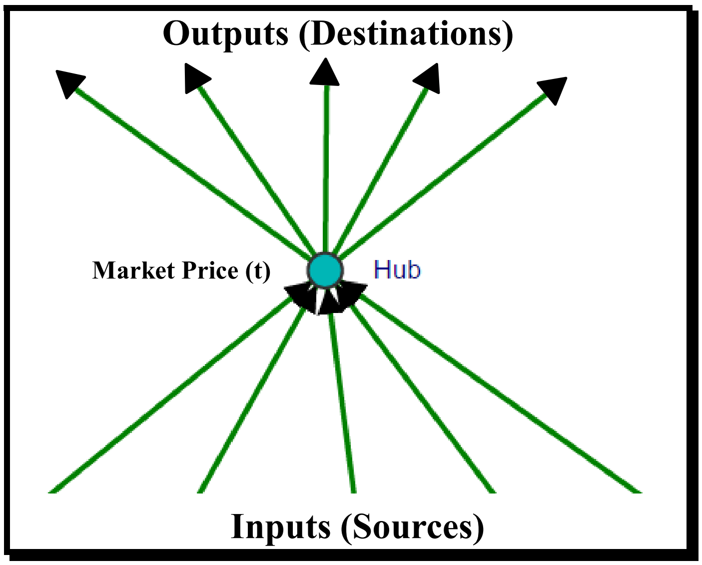
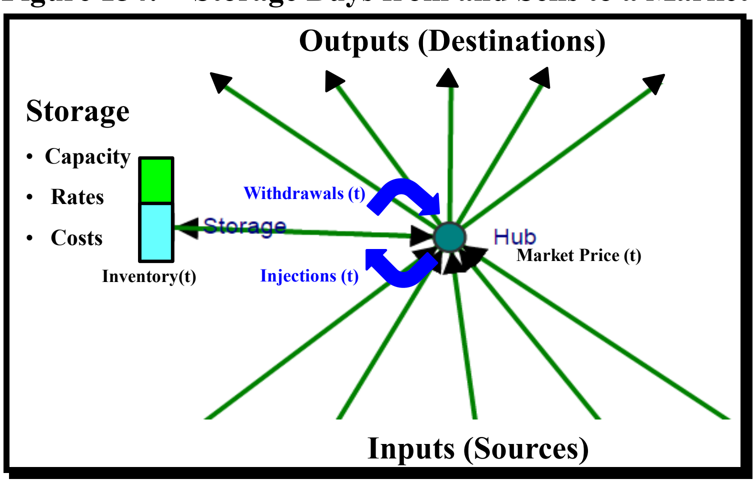


Figure 154: Storage Buys from and Sells to a Market



2. Notice that withdrawals from storage are one of the sources into the market. When commodity is withdrawn from storage, it is sold into and, therefore, enters the market.

3. There are a number of destinations or uses of the commodity in the market. Those destinations are indicated as outputs (destinations) from the local market hub.

4. Notice that injections to storage are one of the destinations from the market. When commodity is injected to storage, it is bought from and, therefore, leaves the market.

5. There is a time-varying price at the local market hub, designated Market Price (t). That market price tends to be high at time of peak and low at time of off-peak. At time of peak, demand is high, and the price is pulled up. At time of off-peak, demand is low or nonexistent, and the price is pulled down. Storage is not a passive actor in the market. Quite the contrary, storage is an active player in the market and affects the price over time in the market materially.

In the absence of storage, the situation was extremely simple, as in Figure 153. The real-time sum of the inputs must be exactly equal to the real-time sum of the outputs—the market must balance or equilibrate—supply would have to be exactly equal to demand at every instant in time. Without storage, there is no way to temporally buffer the time pattern of supply to meet a time pattern of demand that is intrinsically different from the time pattern of supply. With storage, however, the situation becomes markedly different, as in Figure 154. With storage, someone has built a tank at the market hub, and he or she has thereby built the ability to buy commodity from the market and inject it into the inventory in his or her tank on any injection and withdrawal schedule he or she wishes. He or she has also built the ability to extract commodity from the inventory in the tank and sell it back to the market. The owner of the storage facility has installed the ability to buy commodity from the market, put it into the tank, and extract and sell it to the market literally whenever he wants. He has bought the ability to temporally arbitrage prices in the market. He can buy low/sell high at his complete and unilateral discretion. (In the case of reservoir hydro, he doesn't have the discretion to buy low. Water falls from the sky and shows up in his reservoir for free. However, he very much has the discretion to "sell high," maximizing the revenues and therefore the profits he gets from operating his storage asset. We address the issue of "selling high" below when we consider energy limited resources.

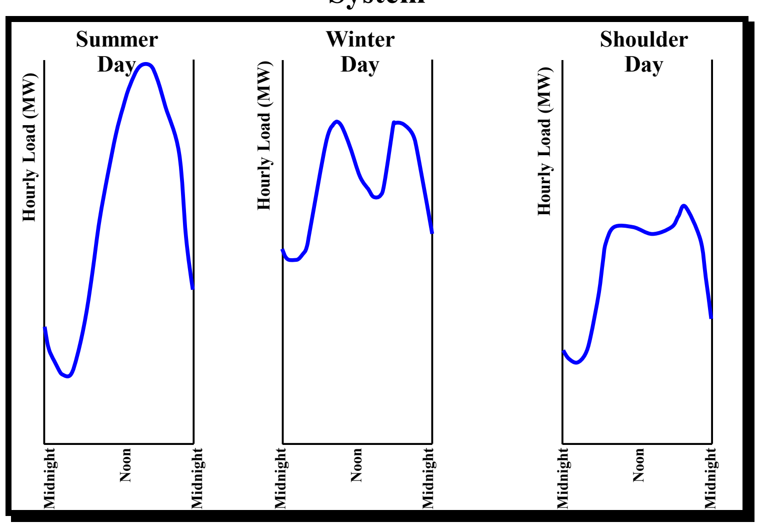
The notion that storage buys commodity from a market, saves it, and sells to a market is central to modeling storage correctly in IEMM. Storage buys from a power market at 1:00 am when the price is very low because there is no demand, stores it, and sells it into the market the very next day at 2:00 pm when the price is very high because there is so much more demand. Storage buys electricity to pump water uphill at 1:00 am so that it can put the water through the penstock and generate the next afternoon at 2:00 pm. Storage is buying from the market, saving, and selling to the market later. The Strategic Petroleum Reserve is no different from this—buying crude oil, saving it, and ostensibly selling it when times are tight and crude prices are high. Electricity storage is much faster and much closer to real-time than oil or gas storage, but it is nonetheless very similar.

11.2 Fundamentals of Electricity Storage at or Near Point of End-use

This section examines how electricity storage, if it existed technically and economically, would work in IEMM (as defined here) and in the real world at the point of consumption. Batteries, capacitors, small scale compressed air, Tesla Powerwall home lithium battery packs, thermal storage (e.g., ice making and melting), and those types of direct and indirect electrical storage capabilities are represented in IEMM and have the impacts we will discuss in this section. Some of these technologies are presently under development and others exist. There is a literal frenzy of activity in these arenas because the impact and benefit of these technologies are beginning to be more widely understood. When we see the impact these techniques can have and the economic modification of the generation business, we see exactly why these techniques are so important and why IEMM cannot omit them. When we envision the fundamentally decentralized nature of these technologies and the way they can modify/distort the generation system, we realize how futile and incorrect it would be to represent them using global social welfare maximization methods. It just is not going to happen. Storage is far too microeconomic in nature. (Is there anything more agent-oriented than a Tesla Powerwall? Probably not.)

People have long measured the daily-weekly-monthly load variation in homes and retail systems around the world. Measurements of load variation in domestic and commercial use have been observed to behave as depicted in Figure 155. In the middle, we invariably observe that the 24-hour load on a winter day begins a midnight at a very low level (everyone is asleep, residential load is low; businesses are idle so commercial load is low, only 24/7 industry is operating). People get up in the morning and business and industry start for the day, and we see the load ramping up quickly. It is still cold in the morning, so the ramp-up is rapid and high. During the midday, the ambient temperature rises to the point at which the magnitude of space heating and other use moderates a bit. Solar insolation is your "friend" on a winter day, warming ambient temperatures a bit and reducing load in the middle of the day. As evening approaches, solar insolation declines, temperature declines, and the evening temperature drops. Load increases to accommodate evening activities as the ambient temperature drops. Finally, people go to bed and commerce shuts down for the night, and the load falls. We can see at the left that the sun is your "friend" in the winter at the left. The graph at the left is ubiquitously observed during winter days virtually everywhere in the world.

Figure 155: Daily Load Shapes Are Quite Predictable in Every Industrialized Electric System

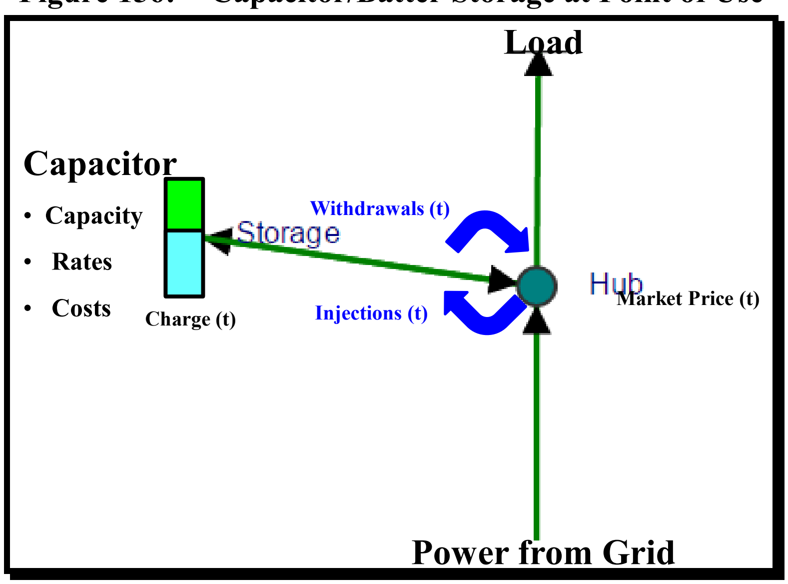


Moving to the far left, a summer day, the situation is different. The load drops from midnight at the left until about 4:00 am when it is minimum. That is when the daily temperature is at its lowest, and, therefore, air conditioning is at its lowest, commerce is shut down for the night, and industry is at a minimum. We see a morning ramp-up for two reasons. The ambient temperature begins to rise at sunrise, and people initiate domestic, commercial, and industrial work for the say. We see a rapid ramp-up in load from 6:00 am until it peaks at about 2:00 pm. The sun sears across the sky from about 10:00 am until about 4:00 pm and the midday load rises to very high levels. The sun is your "enemy" during the midday, accelerating the need for lifesaving and comfort- and productivity-enhancing air conditioning. In the evening, as the sun descends in the sky, solar insolation drops, and temperature drops, load ramps down heading into the midnight hour. The graph at the right is ubiquitously observed during summer days virtually everywhere. It has a massive midday peak at about 2:00 pm. It has a massive off-peak at about 4:00 am.

The shoulder months (spring and fall) are low and flat at night and moderate and flat during the day. There isn't much air conditioning or temperature driven heating or cooling demand during shoulder months. Temperatures are moderate. Load is lost when people go to sleep for the night. The load shape is fairly flat during the day.

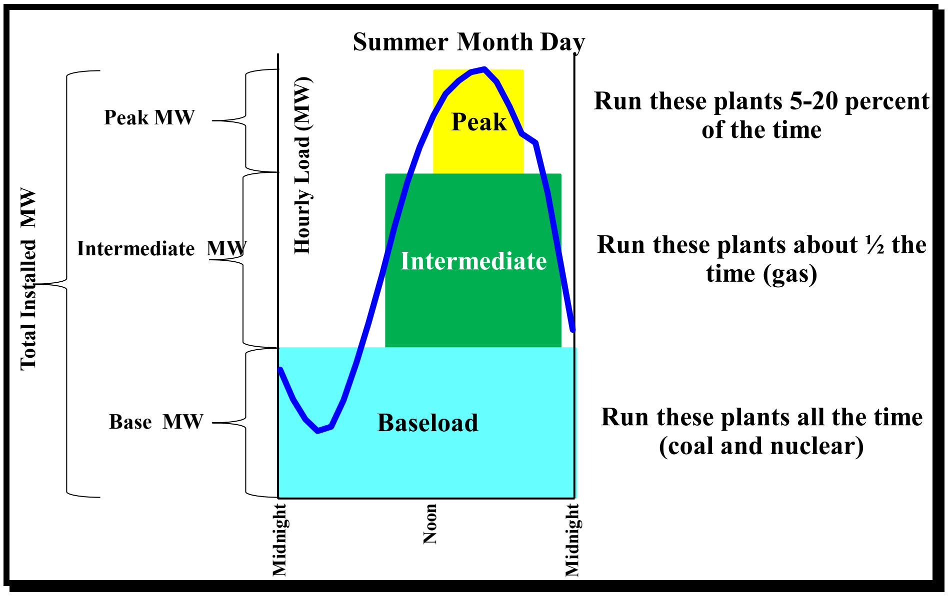
Suppose you had a perfect capacitor/battery in the market, a perfect electricity storage device. By perfect, we mean that you put electricity into it whenever you want, it stores that electricity as an electric charge, and it discharges that electricity back into the market whenever you want. This is akin to the Tesla Powerwall—taking in power whenever you want and withdrawing power whenever you want. We like to term this a "capacitor," but that does not imply any specific technology. Figure 156 illustrates.

Figure 156: Capacitor/Batter Storage at Point of Use



If we have the electricity storage device in Figure 156 in the market, consider how we might have to service the load for the summer day at the right-hand side of Figure 155 if there were no storage. As indicated in Figure 157, you would have to have some plants at the bottom of the stack designated "Baseload." Those are plants that would have to run all 24 hours of the day, never shutting off and always delivering to market.

Figure 157: How Do You Dispatch Generation without Storage?



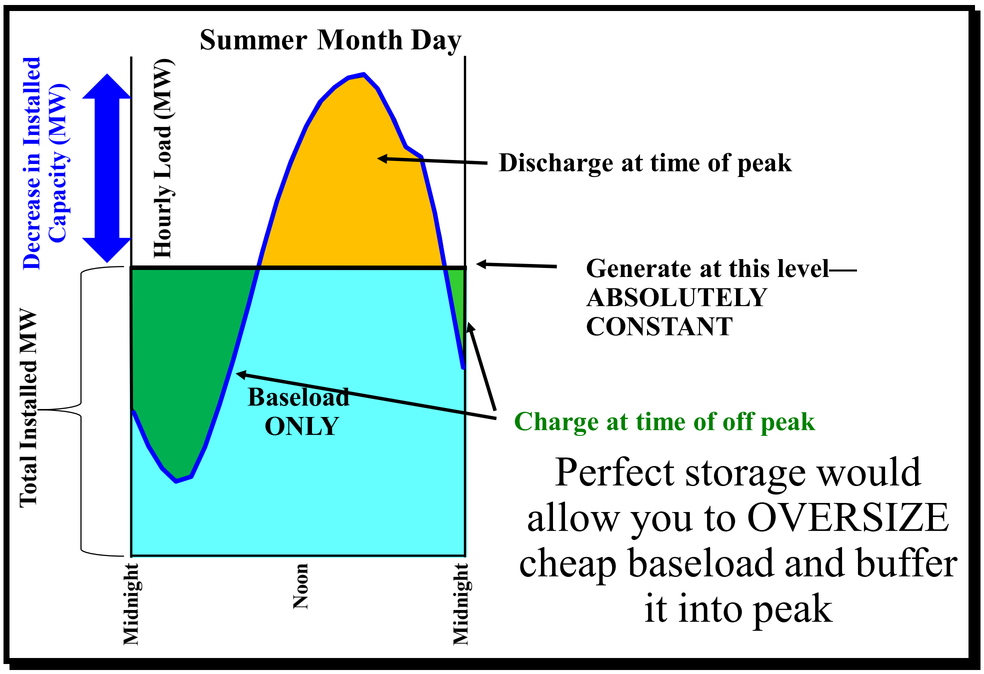
There is a collection of plants designated in green as "Intermediate" that would come on at approximately 10:00 am, run at full capacity, and shut down at 10:00 pm. Those plants would not be needed after 10:00 pm and before 10:00 am because there isn't enough aggregate load to absorb their output. These plants would run about half the time, starting up in the morning and shutting down in the evening. They are your standard, well known Intermediate power plants.

Finally, there is a collection of plants designated in yellow as peak plants. These plants start up at 11:00 am or so and shut down at 3:00 pm or so. The load is high enough to absorb the output from those plants for the hottest four hours of the day, and no other.

If we had a storage technique in this market as indicated in Figure 156, the situation would be markedly different. Suppose you were to generate at a constant level as indicated in Figure 158. Suppose further during the late night and early morning, you took electricity that was generated, used part of it to service the then-current load (which is low at time of off-peak), and inject the difference between average generation and actual load into your capacitor. In Figure 158, you would be injecting the energy indicated by the green area into your capacitor. Notice that a lot of that green area occurs between midnight and say 8:00 am, and some of it occurs from say 9:00 pm to midnight.

Your capacitor would contain its maximum stored inventory by about 10:00 am. You would then withdraw from your capacitor and place the electricity into the market beginning at 10:00 am when the actual load in the market exceeds the average generation level into the market. Specifically, the average generation would be provided to the market, and storage would provide the difference between the load (which exceeds average generation) and average generation to the market. This withdrawal from your capacitor is represented by the orange area in Figure 158. If you could, you would set daily generation at the constant level in Figure 158 so that the green area (injection) was exactly equal to the orange area (withdrawals). You would use storage to completely buffer the variation in load around a constant, average, daily generation level. The actual capacity delivered to end-use (the size of the local wires) is dramatically reduced with end-use storage. This is the genesis of the stranded cost arguments that have been seen throughout the United States electric distribution company business. Our design models it correctly.

Figure 158: If You Had a PERFECT Power Storage Technique (Battery or Capacitor), What Would You Do in Summer?



We also note in Figure 158 that the magnitude of installed and operational capacity in place during the particular day would be the difference between the peak MW (the top of the load curve) and the average monthly generation. The existence of the storage capacitor allows installed capacity to drop by the magnitude of the vertical red double pointed arrow at the upper left. This represents a very large reduction in necessary generation capacity brought about by the capacitor completely buffering the entire load variation in the region. The proposed storage model allows IEMM to correctly and properly quantify these capacity changes. There is no other reliable way to do so, particularly not global social welfare maximization. The methodology simply does not accommodate or represent storage. No one has ever done it. This design allows IEMM to properly quantify and understand the potentially profound consequences of storage and dispatch of storage.

These diagrams and analysis imply the following storage fundamentals:

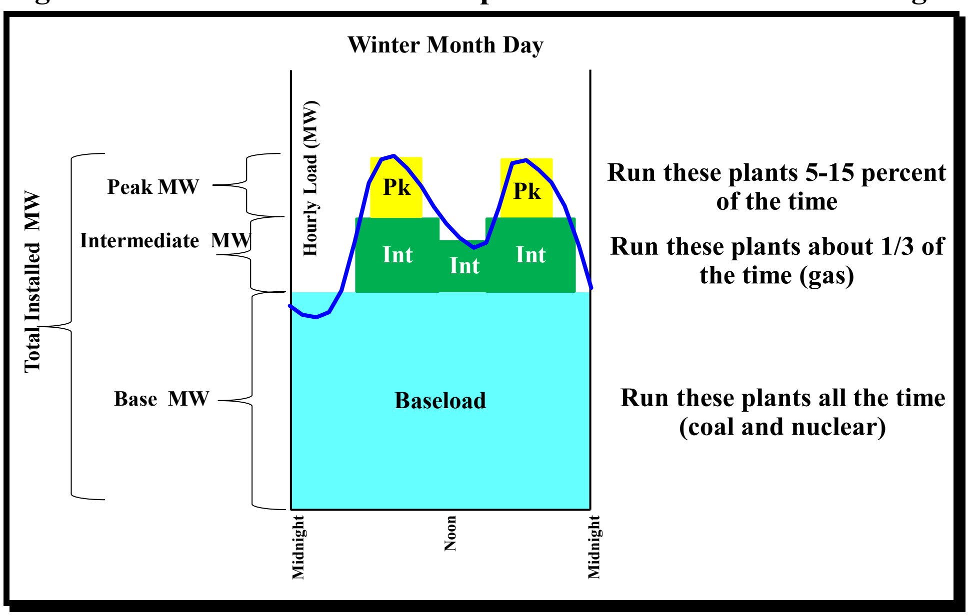
- Storage and load leveling must be considered at or near aggregate end-use.
- Storage would allow the system to generate with **nothing** but cheap, efficient baseload plants, plants that never shut off and never cycle.

- Storage would obviate the need for day-to-day or hour-to-hour peaking technology. That is the whole point of storage—to eliminate the need for time-varying supply even though demand is wildly time-varying.

- Depending on how long you could hold a charge, storage could cross days, weeks, or months.

If we have the electricity storage device in Figure 156 in the market, consider how we might have to service the load for the winter day at the left-hand side of Figure 155 if there were no storage. As indicated in Figure 159, you would have to have some plants at the bottom of the stack designated “Baseload.” Those are plants that would have to run all 24 hours of the day, never shutting off and always delivering to market. You would have intermediate plants that come on at 6:00 am when people are getting up and work is ramping up and run until 10:00 at night when industry and commerce shuts down during the night. Some of these intermediate plants are seen to ramp down during peak solar insolation from 10:00 am to about 3:00 pm. Finally, we see the peakers coming on stream in the morning hours and the afternoon/evening hours. The generation pattern necessary to meet the winter day is clear from Figure 159 in the absence of storage. There is a little bit more base-loading (relatively speaking) and a little bit less peaking (relatively speaking) than the summer day. Winter loads are slightly less “peaky” than searing summer loads.

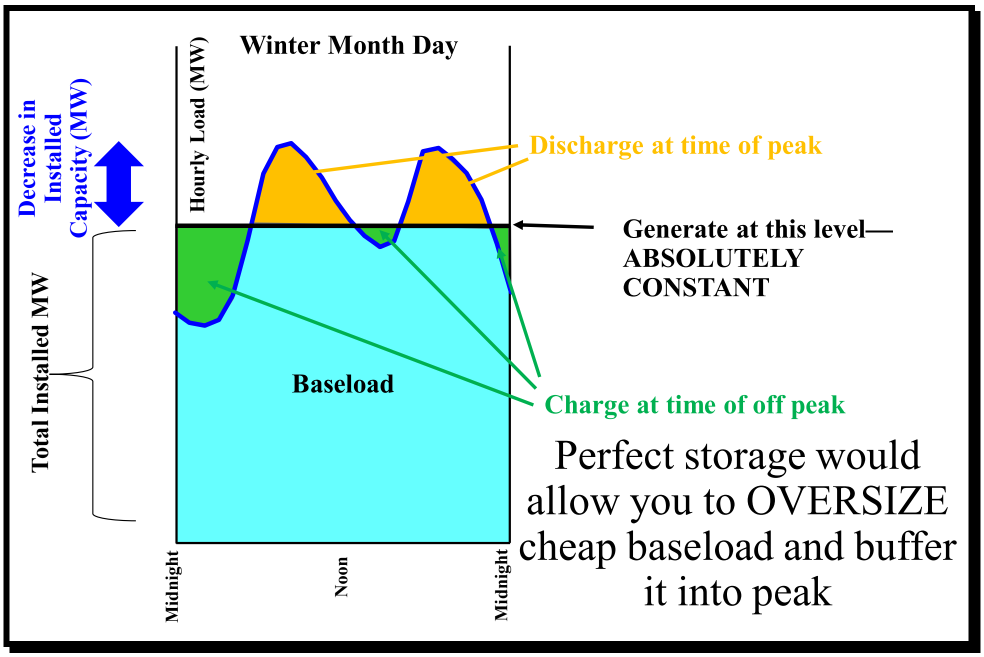
Figure 159: How Would You Dispatch in - Winter without Storage?



If we had a storage technique in this market as indicated in Figure 156, the situation would be very different. Suppose you were to generate at a constant level as indicated in Figure 160. Suppose further during the late night and early morning as well as the midday, you took electricity that was generated, used part of it to service the then-current load (which is low at these times of off-peak), and injected the difference between average generation and actual load into your capacitor. In Figure 160, you would be injecting the energy indicated by the green areas into your capacitor, one during the middle of the night and one during the midday. Notice that a lot of that green area occurs between 10:00 pm and say 5:00 am, and some of it occurs from say 10:00 am to 3:00 pm when midday solar insolation and consequent higher temperatures are helping rather than hurting you.

You would take energy out of your capacitor during the morning peak (the leftmost orange area in Figure 160) and during the evening peak (the rightmost orange area in Figure 160). The total orange area would be the magnitude of discharge from the capacitor. You would set the constant average generation level so that the sum of the injection energy (the green area) would be equal to the sum of the withdrawal energy (the orange area). Your generation would be constant, and it would be established at a level so that the green off-peak energy would be totally buffered into the orange on-peak area.

Figure 160: If You Had a Perfect Power Storage Technique (Battery or Capacitor), What Would You Do in Winter?



Just as with the summer day, notice that the total capacity in place and operating at the flat level is substantially lower than the peak demand. The advent of the storage technique reduces the capacity in place from the peak capacity to the average capacity during the winter day. The new capacity in place is the height of the horizontal line that bounds the blue area in Figure 160, i.e., the average daily generation. The amount of capacity saved is the difference between the 24-hour peak and the average generation line. This is one of the leading terms and, in fact, the raison d'etre for storage—reducing the amount of generation capacity in place and reducing the amount of electricity that has to be delivered to consumers.

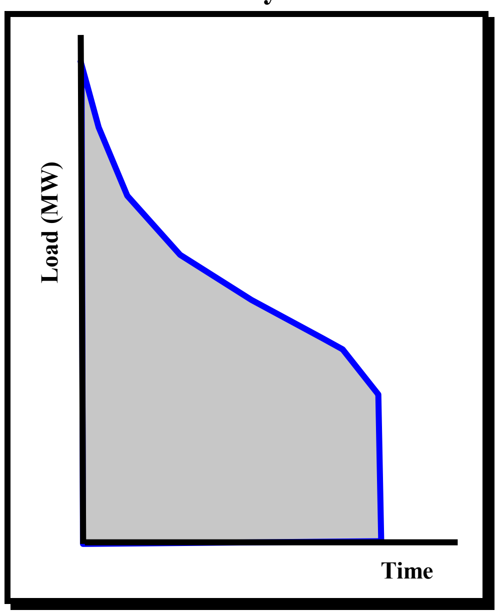
This analysis is extremely elucidating, and most certainly IEMM must have it in the mix at end-use. If you had a perfect power storage technique, if you sited it at end-use, you have the potential to charge/discharge it so as to **completely** buffer a constant supply into a time-varying load with literally a 100 percent capacity factor. You could have **very** low cost, VERY high load factor generation. You wouldn't have any expensive cycling and peaking capacity. Your system would gravitate toward rock bottom cost baseload sources and away from more expensive peakers, and you would eliminate a lot of capacity you would otherwise need in the fleet.

IEMM needs to model storage in this fashion, perfect or imperfect. This will not occur with anything but a decentralized, agent-based model. How in the world would anyone figure this out with a global social welfare maximization model? No one knows.

11.3 Storage and Load Leveling with Load Duration Curves

The preceding models of electricity storage have to be applied and applicable in a load duration curve context as well as the preceding chronological context. This section shows how that will be done in IEMM. Consider the load duration curve for a period of time, say a month, as indicated in Figure 161. The peak hours in the months are on the left and the off-peak hours in that month are on the right. The height of the curve represents how many MWh of energy you need to deliver in each hour of the month. Keep in mind, MWh of energy divided by 1 hr. of duration gives MW of load, the height of the curve.

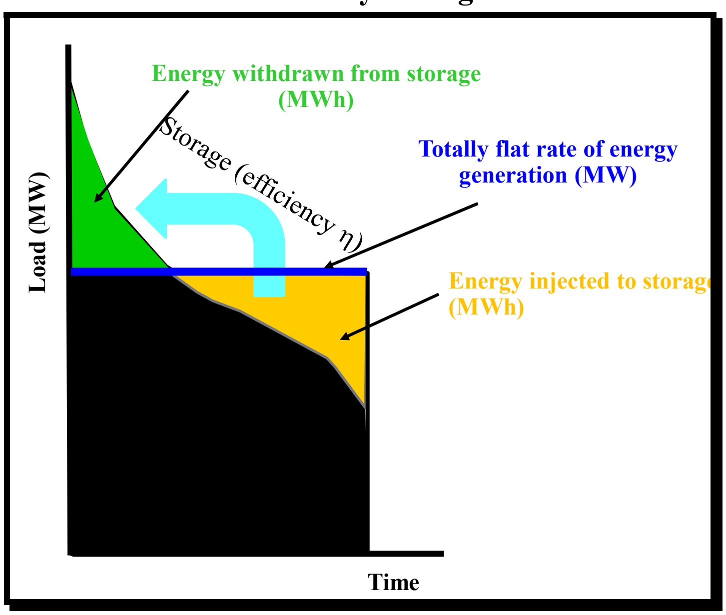
Figure 161: The Monthly Load Duration Curve



Suppose you had a perfect electricity storage system that you could insert into this load system. In particular, suppose that you could generate 1 MWh in hour j of the month but you could deliver that 1 MWh to the load in hour k of the month. This is the ideal storage system, is it not? It is a system that allows you to generate whenever you want, save the energy, and deliver it whenever you want. Think of it as a large-sized capacitor, one that accepts electricity injection whenever that electricity is generated and produces electricity withdrawal whenever the load occurs—perfect and complete buffering of the load.

The crucial observation is this. What time pattern of generation would you need in the region to meet the load? If all were perfect, you would need a perfectly flat, constant operation generation fleet at the **average** load for the month. The horizontal blue segment in Figure 162 represents the flat, non-varying, constant generation level that one might consider for the month. It is completely base-loaded, operating in each and every single hour of the month, and 100 percent capacity factor production. For every hour that load is less than the average load, you would inject the difference between the monthly average and that hour of generation into storage as shown in the yellow area in Figure 162. For every hour that load is larger than the average load, you would withdraw the difference between that hour of generation and the average generation for the month from storage and serve the load, as shown in the yellow area in Figure 162. The orange area in the diagram shows the total energy that would be injected into your capacitor. The green area in the diagram shows the total energy that would be withdrawn to market from your capacitor. The orange area would be bigger than the green area by precisely the amount of loss that you would bear as a result of injection and withdrawal. (There will be losses to inject and withdraw, the second law of thermodynamics guarantees it!)

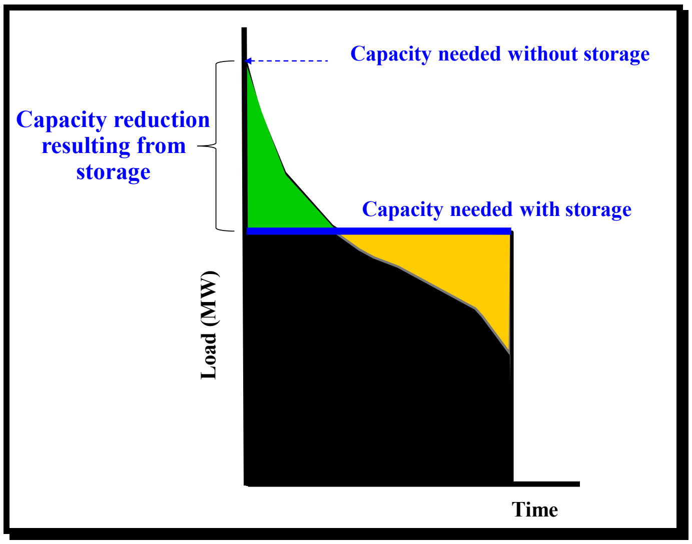
Figure 162: Modification of the Load Pattern Resulting from the Introduction of Electricity Storage



The perfect capacitor (with an efficiency of η), would allow you to service the load as indicated in Figure 162. You would run a strictly base-loaded 24/7 system at rock bottom cost at the MW range indicated in the blue horizontal segment. You would inject the difference between MW generation and load when load is below MW generation (the orange area). You would withdraw the difference between load and MW generation when load is above MW generation (the green area). Inventory in the capacity would go up at time of off-peak, and inventory in the capacitor would go down at time of peak. The inventory would never go below zero, and it would never go above the orange or green areas.

And why would this be the type of insight IEMM should or would give? The answer is seen by redrawing Figure 162 as shown in Figure 163, which emphasizes the capacity reduction that the entry of storage can induce. Without storage, as emphasized in Figure 163, there must be a fleet of capacity in place necessary to meet the peak hour of demand. That is, the generation system must be sized to the absolute peak hour in the system. If it is undersized, there will be a shortage at time of peak and load will not be met. In the figure, that amount of capacity is designated "capacity needed without storage." The fact that all the storage can buffer the constant amount of capacity into the time-varying load means that the magnitude of generation capacity in place is substantially lower, decreased to the level designated "capacity needed with storage."

Figure 163: Capacity Reduction Resulting from Electricity Storage



Let us consider a real world power system. The monthly load duration curves in a historical, real world power system in a previous year are assembled in Figure 164. The blue curves in the diagram are the monthly load duration curves for the twelve chronological months in the given region. We notice that the annual peak load occurs in July, and it was approximately 40 GW. That is the highest height of the blue monthly load duration curves in any month of the year, indicated by the blue arrow in the diagram.

We also see in the diagram in Figure 164 the notion that the area below the monthly average (indicated in yellow) can be shifted up to the area above the monthly average (indicated in green) by the insertion of grid level storage (or pumped hydro or perhaps other techniques). That means that the monthly generation that is required in each month is indicated by the heavy, solid, black line. This solid, black line lies substantially below the real-time load in the blue curve for many hours of every month. That presages the reduction in capacity that is required to meet each monthly load because of the existence of the storage and load-shifting.

By representing storage as implicit in Figure 164 (the way IEMM under this design will do it), we can accurately represent the capacity in place, capacity decommissioning, and capacity expansion that is required in the model. Importantly, we begin to see the generation pattern that is required in each month and for the entire year as indicated in Figure 165. Notice that the generation in place has declined markedly from the peak load period in July. The green arrow in Figure 165 shows the reduction in installed capacity that is achieved by the storage and load leveling. Notice in the diagram in Figure 165 that there is a big tranche of baseload generation that runs literally around the clock every day of the year. That is the blue area in the diagram. Those plants will run all the time. Some plants turn on in May and turn off in October and they run during the smaller winter peak, and they run virtually all the time for that 6 or so month span. Those are the orange plants in the diagram. They are intermediate and peak plants, but they do not cycle nearly as much. They let the capacitor do all the short-term cycling. They simply run at full capacity during the six-month season that they run. The notion of plant cycling is all but gone. All the cycling is done by injection and withdrawal from the capacitor.

Figure 164: Monthly Load Duration Curves with Power Storage

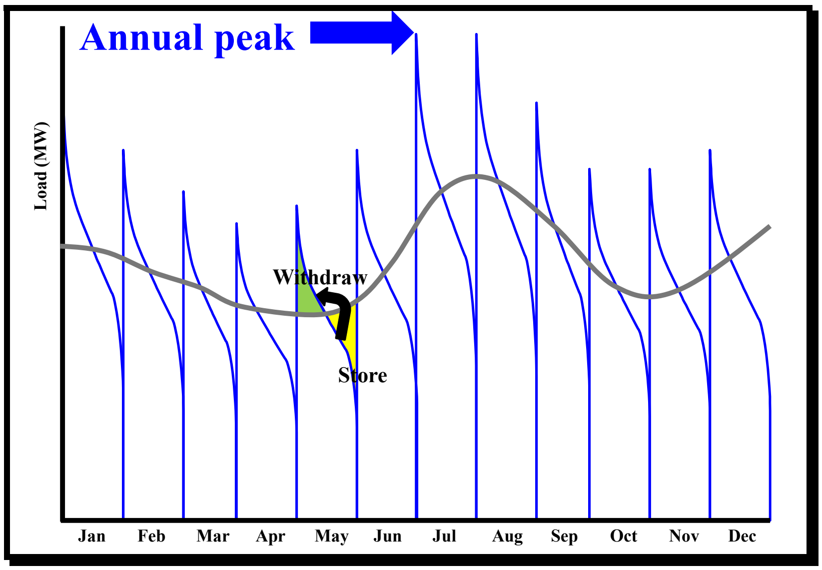
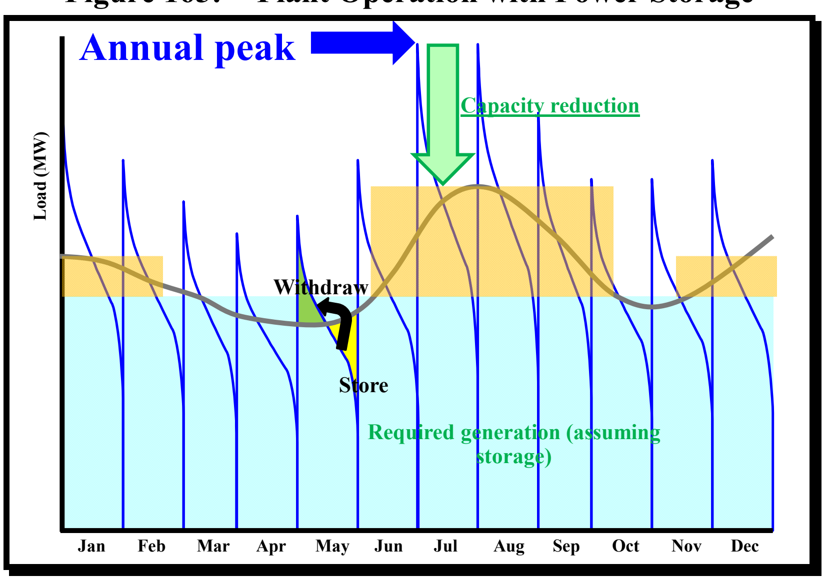
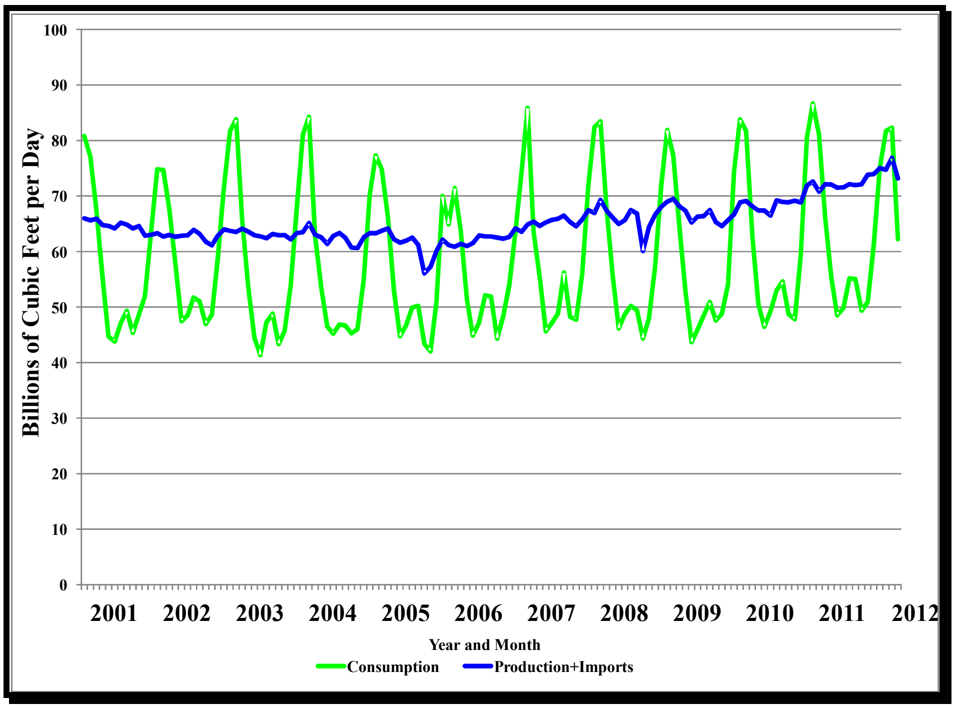


Figure 165: Plant Operation with Power Storage



This ain't rocket science! This is precisely what we see in natural gas in the United States. The relevant diagram is that in Figure 166, which shows that wellhead production is flat while demand is highly time-varying. This is what storage DOES! Clearly there is a large natural gas storage system buffering constant production into a time-varying load. The proposed storage model design for IEMM here will allow EIA to analyze whether the same phenomenon as gas occurs with electricity storage—batteries, thermal, CAES, capacitance, etc. That is necessary and invaluable. IEMM cannot make naïve assumptions about storage and load leveling from the outset. Storage and load leveling have to be model OUTPUTS, not model INPUTS.

Figure 166: United States Production Is Flat While Consumption Is Strongly Seasonally Peaked

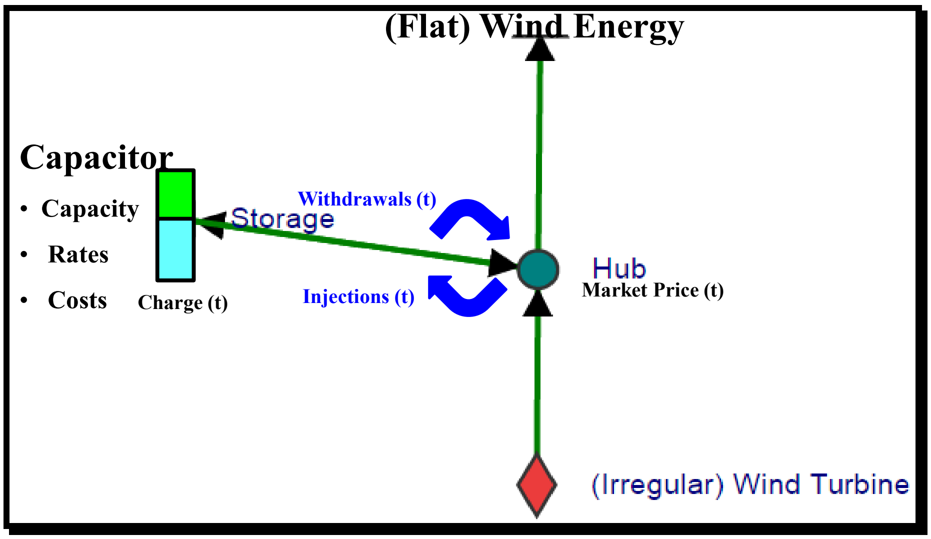


11.4 Fundamentals of Electricity Storage at or Near Irregular Renewables

If we were to site the same capacitor analyzed previously at the site of renewables generation, we could direct the irregular schedule renewables generation into a flat (or flatter) energy source. In particular, although we will not draw the buffering of load diagram, the wind turbine situated with storage can inject when the wind blows and withdraw at a constant rate, thereby turning the wind turbine into more of a capacity machine than if it were run directly. This has become a crucially important issue in renewables policy. Policy makers have become very interested in associating storage with intermittent renewables as a way to "firm" those renewables. That cannot be done without a decentralized agent-based IEMM representation. Firming and leveling of renewables generation is a microeconomic phenomenon.

Storage at the point of wind turbine generation is a way of flattening the highly variable wind generation duration curve and more importantly ostensibly turning it into a flat, reliable, 100 percent probability energy-from-capacity stream. It is a way of firming the temporal production from wind into a flat energy profile, flat at the average duration factor of the wind at the wind turbine. This would be a highly valuable product, for it would be akin to a chunk of capacity. It would be a way to change the irregular generation pattern in Figure 130 to its 300 MW constant equivalent in that diagram. It is not without cost, but with the IEMM model, the analysis is doable. Consider the structure in Figure 167, which sites a wind turbine and a storage facility (e.g., a capacitor or battery) at the same market location (hub.)

Figure 167: Capacitor/Battery at Wind Turbine Site

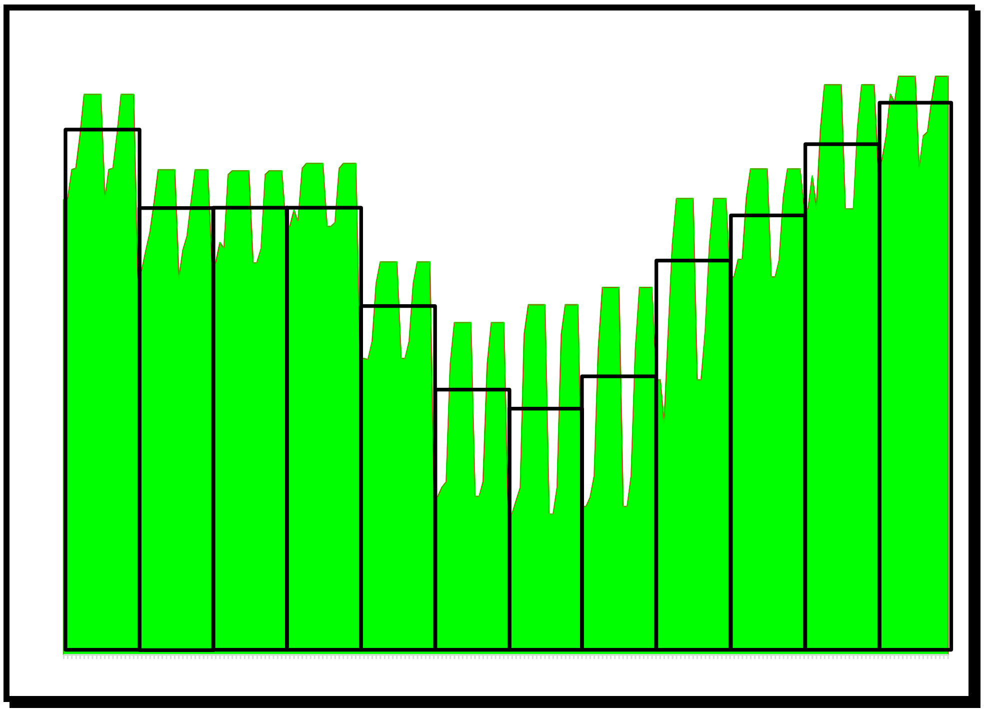


Under this structure in Figure 167, the monthly wind diagram from Figure 136 would in the presence of sufficient storage take on the form in Figure 168, which is precisely equal to the average monthly wind energies originally introduced in Figure 134 (subject of course to losses in injection, storage, and withdrawal from storage, if any.) This is a far more attractive generation pattern than the spiky, irregular generation pattern by the wind turbine in place. Furthermore, it looks very much like a baseload capacity generation pattern. The advent of reliable storage associated with a wind turbine has the potential to turn it from an “energy machine” to a “capacity machine.” This is potentially huge and could have a major impact on the generation fleet in place, both renewable and thermal, as well as the price of wholesale electricity. This consideration is not limited to the United States. It is in play everywhere in the world, and IEMM must model it accurately as this design allows. That price will be much more driven by wind than it has in the past, and much less driven by thermal generation. There is no way IEMM can fail to consider ideal or even imperfect storage at the renewable generation site. This technology is likely to materialize, whether only in part, and such materialization dramatically alters the economics and reliability of renewables generation as compared with traditional thermal generation. With the proposed IEMM design herein, which lets storage projects be associated with renewables generation projects be represented as indicated in Figure 167 and thereby allows IEMM to relate renewables and storage at the same site in an endeavor to firm their capacity.

To summarize, there are many possible locations at which storage can occur in the economic system, all the way from extreme upstream to extreme downstream. In the case of natural gas storage, which we need to understand as an analogy, there are primarily two locations with two fundamentally different characteristics:

- Market area storage
- Production area storage

Figure 168: Wind Generation Pattern Averaged Over the Month



These locations are quite fundamentally different. Production area storage (or generation area storage as in the case of renewables) is likely to be largely "merchant," owned and operated by private producers or generators. Market area storage might be "utility" owned and operated, although equipment on private premises might well be privately owned and operated. Examples might be so-called grid-level liquid metal batteries such as the Ambri Liquid Metal Grid Scale Battery, which is still very much in the R&D phase. IEMM will keep this distinction in our mind as we think about electricity storage, which is even more diverse in terms of possible location.

We have examined in the preceding electricity development the equivalent of market area storage when we analyzed storage at end-use. We have examined the equivalent of production area storage when we analyzed storage at the site of renewables generation. IEMM can and must analyze storage at both areas in the power system (as well as others)

- **End-use (retail).** The Tesla Powerwall (lithium batteries hanging on the wall) are an end-use, merchant, privately owned storage technique. So are ice in the basement, small scale compressed air, small scale batteries (like the sodium-sulfur or tri-liquid batteries we hear about) be end-use, merchant, retail storage devices. They would be injecting electrons at time of base and pulling them out at time of peak. Supercapacitors would fit into this segment.

- **Neighborhood (maxi-retail).** We use the term maxi-retail because it aggregates more than one customer both in ownership and operation. Large-scale sodium sulfur and other liquid metal batteries such as Ambri implemented within a neighborhood, cooperative, or collective would fit into this category.

- **Wholesale.** Large-scale run of river hydro, reservoir hydro, and pumped storage facilities operate at the wholesale level of an electric power system. Large-scale compressed air (CAES) projects would be wholesale.

- **Generator.** The reservoir hydro projects occur at the level of generation.

Electricity storage is a complicated problem with complicated locations of devices. IEMM under the proposed design will be able to handle that. Because of the broad diversity of possible storage siting, as is carefully presaged in Figure 1 and Figure 2, inclusion of imperfect storage at virtually every location, every busbar, in the system is a must.

11.5 Development of the IEMM Storage Model in the Pumped Hydro Context

Storage of hydro, refined products, gas, or electricity in batteries/capacitors is an intrinsic player in the marketplace. (In this section, we use hydro storage as our discussion example.) We know that storage adds to high head hydro supply during time of peak (more hydro available to deliver to time of peak demand) and adds to hydro demand at time of off-peak (buys hydro at time of off-peak to service the forthcoming peak need). As such, storage puts:

- Upward pressure on price during time of off-peak. By increasing hydro demand at time of off-peak, storage elevates off-peak price.

- Downward pressure on price during time of peak. By increasing hydro supply at time of peak, storage depresses on-peak price.

Keep these pricing ideas in mind, for they are central to understanding storage.

There are a number of applications of storage in energy markets, and they impact markets and markets impact them:

- Reservoir hydro storage for seasonal variation

- Pumped hydro storage for short-term variation

- Batteries and capacitors for very short-term variation, a growing area

- Petroleum product storage for seasonal variation

- Petroleum storage for strategic reasons

Modeling systematic commodity storage (e.g., hydro and liquid storage) is more difficult than modeling phenomena such as reservoir hydro storage because, in the hydro and liquids case, one has to make two decisions—injection and withdrawal. In the case of reservoir hydro, there is only one decision to make—when to generate. If the storage model can represent the hydro and liquids storage problem, the hydro reservoir problem (pumped as well as seasonal) is a simple special case in which injection is eliminated and water materializes via hydrology.

We are going to explain the hydro and liquids storage module EIA needs here. Specialization to pumped or reservoir hydro and other storage mechanisms is straightforward.

11.6 Pumped Hydro Storage

Storage of hydro, refined products, gas, or electricity in batteries/capacitors is an intrinsic player in the marketplace. (In this section, we use hydro storage as our discussion example.) We know that storage adds to high head hydro supply during time of peak (more hydro available to deliver to time of peak demand) and adds to hydro demand at time of off-peak (buys hydro at time of off-peak to service the forthcoming peak need). As such, storage puts:

- Upward pressure on price during time of off-peak. By increasing hydro demand at time of off-peak, storage elevates off-peak price.

- Downward pressure on price during time of peak. By increasing hydro supply at time of peak, storage depresses on-peak price.

Keep these pricing ideas in mind, for they are central to understanding storage. And hydroelectric storage is instrumental to understanding capacity in place and price in IEMM.

There are a number of applications of storage in energy markets, and they impact markets and markets impact them:

- Reservoir hydro storage for seasonal variation

- Pumped hydro storage for short-term variation

- Batteries and capacitors for very short-term variation, a growing area

- Petroleum product storage for seasonal variation

- Petroleum storage for strategic reasons

Modeling systematic commodity storage (e.g., hydro and liquid storage) is more difficult than modeling phenomena such as reservoir hydro storage because, in the hydro and liquids case, one has to make two decisions—injection and withdrawal. In the case of reservoir hydro, there is only one decision to make—when to generate. If the storage model can represent the hydro and liquids

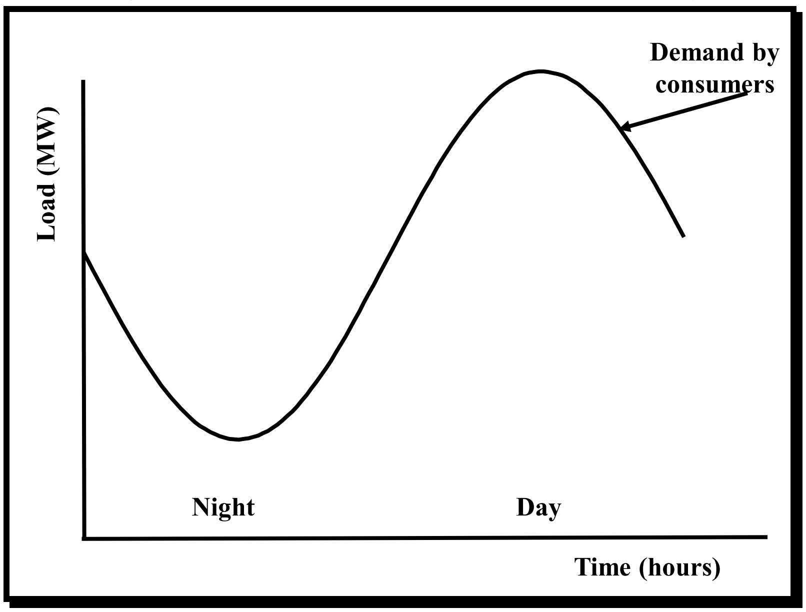
storage problem, the hydro reservoir problem (pumped as well as seasonal) is a simple special case in which injection is eliminated and water materializes via hydrology.

We are going to explain the hydro storage module EIA needs here. Specialization to pumped or reservoir hydro and other storage mechanisms is straightforward.

11.6.1 Storage Reduces Temporal Swings

The load (consumption) of a commodity in many cases follows a day-night pattern. Hydro is an example. Power load through time occurs as shown in Figure 169. We note that consumption at night is low. Similarly, consumption during the day is high because the weather is hot and many homes must be air conditioned for safety and convenience. For purposes of this section, we are not considering the stochastic variation in load, which is important for pumped hydro storage, but only the systematic, predictable load.

Figure 169: Power Consumption Is Hourly



If a generator had to meet this real-time pattern of load, several things would have to occur. First, he would have to size his total generation capacity to the absolute peak, which occurs during the day. Secondly, he would have to shut down much or all of his generation capacity during the night (off-peak). There would be no one to accept and consume his product. He would have to run at higher and higher capacity during the day peak period, running at full capacity only on the peak day. Figure 170 illustrates the situation without pumped hydro storage.

What happens when pumped hydro storage enters the picture? During the night, people are able to draw from the power market to pump and inject hydro into a pumped hydro storage facility **plus** serve the low level of consumption that occurs during the night. In effect, the market buys electricity to inject into pumped hydro storage plus serve the seasonally low night demand. During the day, people can withdraw hydro from the pumped hydro storage facility to generate real-time deliveries so as to serve the peak day market. Withdrawal from pumped hydro storage plus real-time deliveries are sufficient to satisfy the peak demand. The situation becomes that in Figure 171.

Figure 170: The System Must Be Sized to the Peak

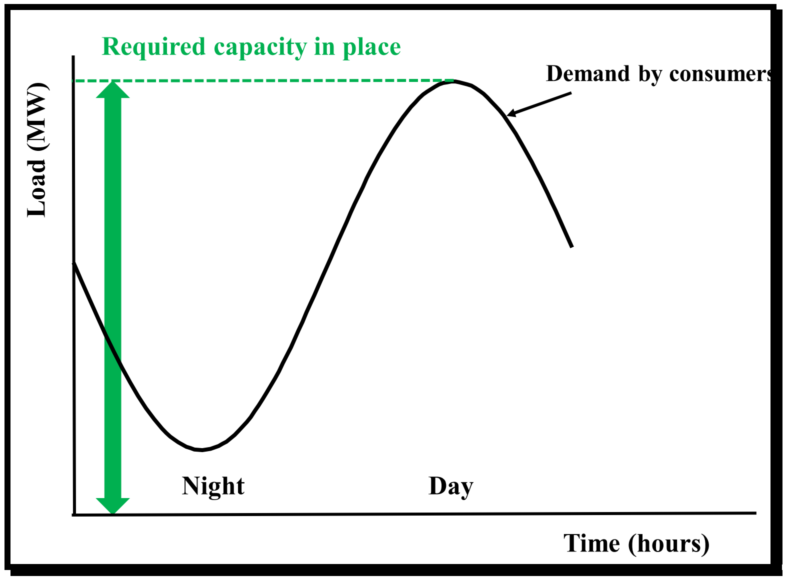
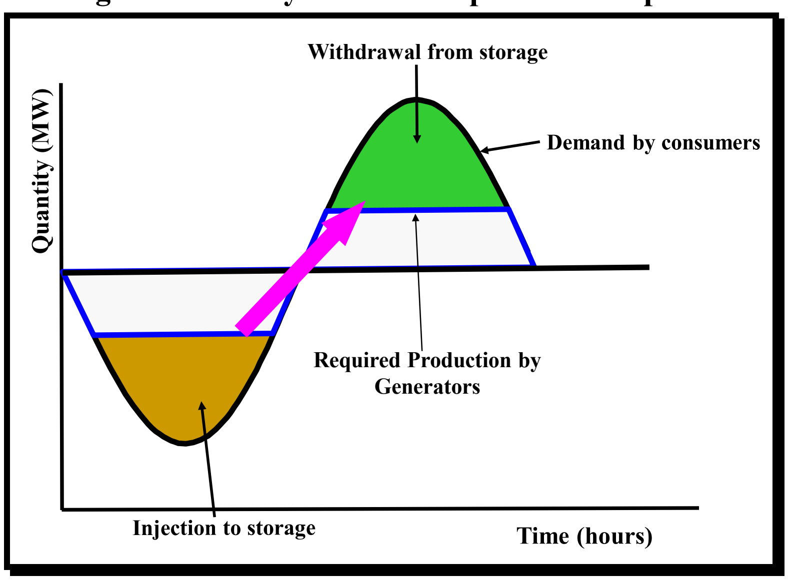


Figure 171: Hydro Consumption Is Temporal

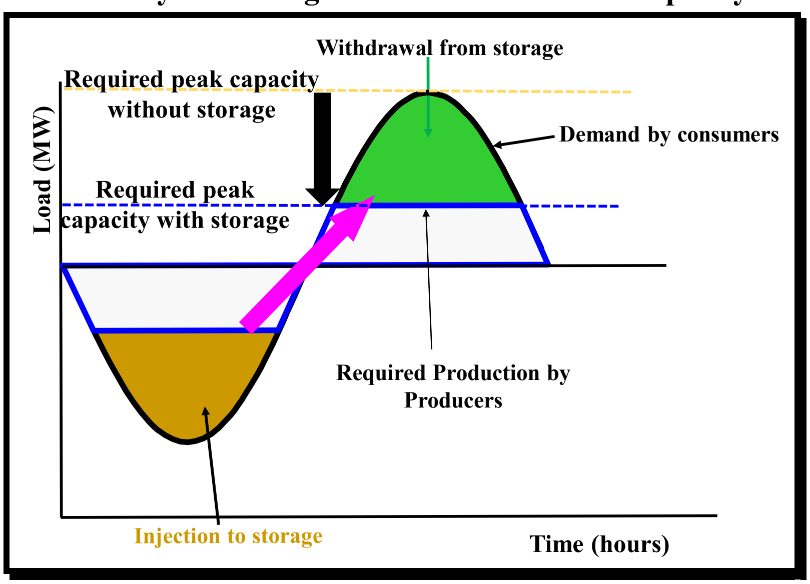


We note in Figure 171 the highly varying load by consumers, the outside black curve in the diagram. The load by consumers is the same as it was in Figure 169. We also see that during the night at the left, there are deliveries to customers **plus** injections to storage. The quantity of electricity that is provided by generators to the market during the night at the left is the solid blue line. As shown, it is composed of the actual consumption by consumers during the night plus the injections to storage during the night, which is represented by the brown area at the left during the night. Night consumption is **higher** than the very low consumption by customers, the difference during the off-peak period being delivered to the storage facility in the form of pumping.

We note in Figure 171 during the day at the right that the consumption is the very high black line at the top. However, the delivery by generators is the solid blue line above the annual average upward trending line but well below the black line. It is, in fact, the blue line at the right. That real-time delivery by generators indicated by the blue line is augmented by withdrawals from storage represented by the green area during the day. Deliveries through the distribution system (the blue line) are augmented by deliveries from storage inventory (the green area) so that the sum total satisfies demand by consumers. The notion here is that storage buys hydro at time of off-peak and sells hydro at time of peak, decreasing the need for real-time deliveries at time of peak but increasing the need for real-time deliveries at time of off-peak. Storage "levels the load" that generators have to meet. The leveled load that occurs as a result of storage, the load that generators have to meet, is the blue line in Figure 171. The effect of storage is that generators meet a much less variable, must less volatile load than they would without it.

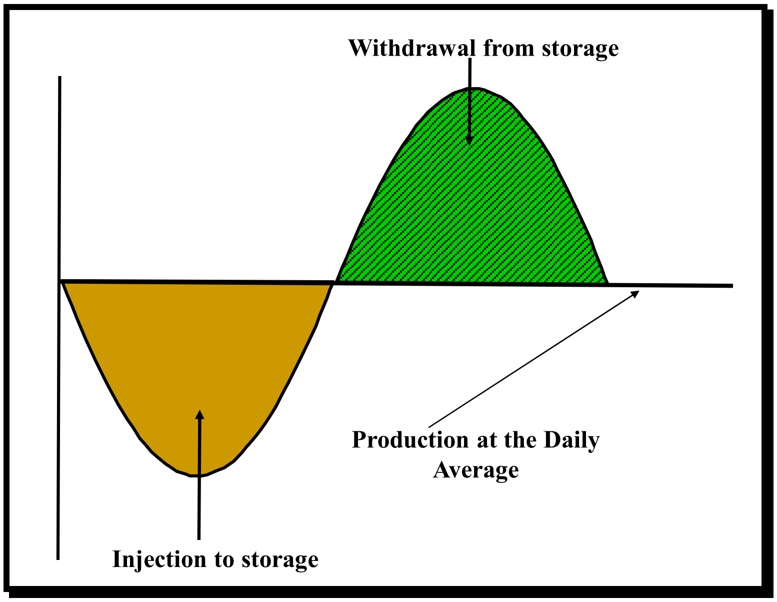
The impact of storage on the magnitude of generator capacity in place is indicated in Figure 172. The peak that generators must maintain is far lower in the "with storage" case than the "without storage" case. This can dramatically affect real-time prices as well as annual average prices and costs. It dramatically affects the capacity factor on the various supply and transportation facilities. This is why storage is potentially so important.

Figure 172: Hydro Storage Reduces Generation Capacity in Place



It is well to consider that if there were an oversized storage system, one could completely buffer the demand variation. One could in the night store all the hydro one needed in the day such that the quantity delivered by generators in the night would be equal to the quantity delivered by generators in the day. In such a situation, represented by Figure 173, the night-day price differential would collapse to pure variable cost plus holding cost, the entirety of peak deliverability having been buffered by storage during the night. This is meant to illustrate that the size of the storage reservoir and the rate at which hydro can be injected to and withdrawn from the storage reservoir is very important. We shall see that in the IEMM storage model.

Figure 173: Massively Oversized Hydro Completely Flattens Generation



11.6.2 Storage and Generation Capacity

What does this mean for crucial issues such as generation capacity in place? Without storage in place, one would have to add sufficient generation capacity to meet the peak load. Generation capacity added is represented by the top line in Figure 174 for the all-time peak. By contrast, if there were sufficient storage to inject (pump) significant quantities during the night and generate with them during the day, the production requirements met by the generator would be the black, smooth, generation line in the middle. The plants would run at full capacity, and the market would only need enough plants to hit the black lines in the second period. You would need far less peak generation capacity. You would run your plants close to 100 percent capacity factor. The key observation is that the market would need far fewer producing plants. Generation costs would be lower. Figure 174 illustrates the effect your storage model must have. It must carefully quantify and measure the reduction in peak generation capacity that would occur by virtue of storage being present in the market. The model outlined here does precisely that and thereby gives more accurate predictions of resource needs and depletion.

11.6.3 Transportation/Transmission Deliverability

The situation is very much the same for transportation as it is for generation. It is obvious from the discussion surrounding Figure 174 that market area storage contiguous to customers will reduce the quantity of inbound transmission capacity that is required. The recommended IEMM storage model represents that phenomenon. However, transmission sizing also interacts with short-term demand variation. Figure 175 illustrates how short-term buffering of short-term demand swings allows inbound transmission to be undersized relative to the time of peak demand, the short-term swings buffering in and out of market area storage. We see in this diagram the longer cycle demand variation and superposed on that the shorter cycle demand variation. Our storage model represents this phenomenon as well, buffering inbound transmission capacity over short-term demand variations and thereby downsizing inbound transmission capacity requirements. The recommend IEMM storage model represents this phenomenon as well.

Figure 174: Storage Reduces Generation (and Generation Additions)

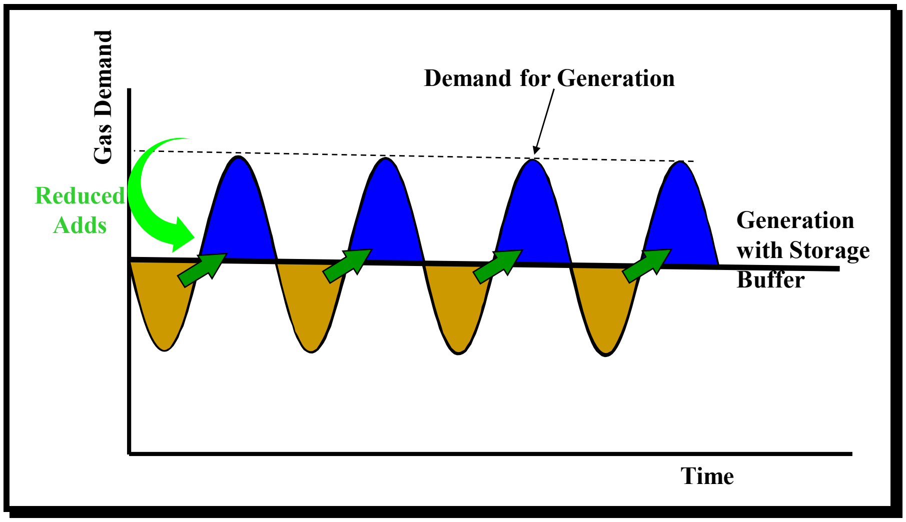
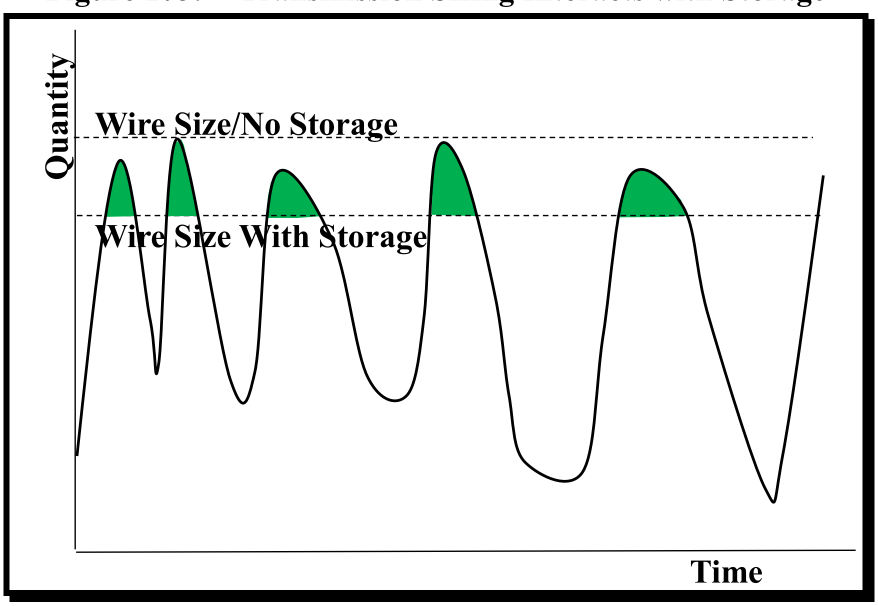


Figure 175: Transmission Sizing Interacts with Storage

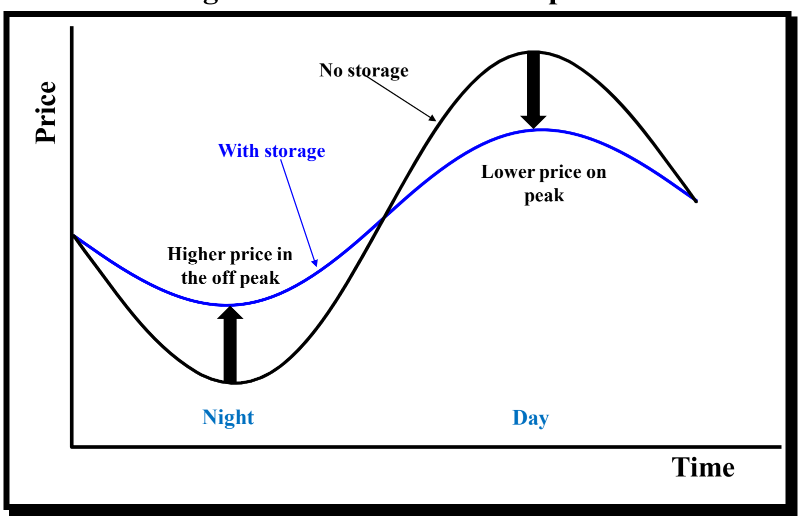


Storage clearly allows the market to undersize inbound transmission (relative to what it would be without storage). This saves **<u>lots</u>** of money and dramatically downsizes the necessary system, and it alters market prices and quantities. It also, alas, can lead to stranded cost and capacity on the electricity transmission and distribution system. IEMM needs to model and predict that.

11.6.4 Storage Reduces the Peak-Off-peak Price Differential

We have discussed, but it is worth strongly reiterating, that storage reduces the differential between the price that occurs at time of peak and the price that occurs at time of off-peak. Figure 176 illustrates. The IEMM storage model must lead systematically to this type of result, and the one we recommend here for IEMM does. Furthermore, if there is a surfeit of storage, the peak-off-peak price differential will fall toward the variable cost of storage, including inventory holding cost. The model we articulate here will do that. Lesser amounts of storage in place (the typical rule around the world) will impact but not collapse the day-night price differential and will retard the quantity of peak generation and transmission system capacity. This is a difficult and sophisticated calculation, and embedding it within a market framework has never been done before to our knowledge. This IEMM design does so.

Figure 176: Storage Reduces the Peak-Off-peak Price Differential



11.6.5 Storage of Water Rather than Storage of Electricity

This section gives a brief sidenote. With pumped hydro, one is not actually storing electricity. One is storing water. You convert electricity to high head water via pumping. You convert high head water to electricity by hydro generation. As an example, consider the California Water Project and its massive pumped hydro project. Notice at the left that water travels southward on the California aqueduct and pumped the top of the Tehachapi Mountains and deposited in Pyramid Lake on the Los Angeles side at the top of the mountains. This is high head water (meaning high elevation water). Water flows from Pyramid Lake through the generator (called a penstock) to Elderberry Estuary, which is part of Pyramid Lake. This is the low-head hydro. The project can pump water from Elderberry Estuary back up to Pyramid Lake, or it can send it to Los Angeles. The schematic for this project appears in Figure 177. The actual IEMM representation of this system would appear as in Figure 178. This representation, which is fairly obvious, represents the WATER as well as the ELECTRICITY. IEMM will have to represent the high head and low head water as well as the losses in pumping and hydroelectric generation for pumped hydro and reservoir hydro processes around the world. To miss this is to miss the essence of electricity modeling worldwide. When we introduce the notion of run of river hydro in the next section, it will become even more obvious why hydro representation will be central to IEMM.

Figure 177: Pumped/Reservoir Hydro Example—California Water Project

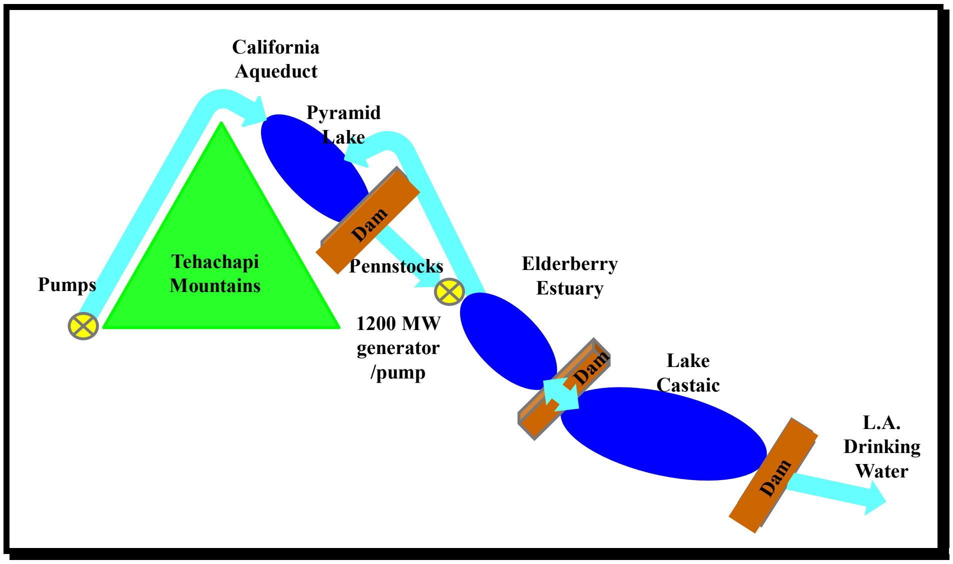
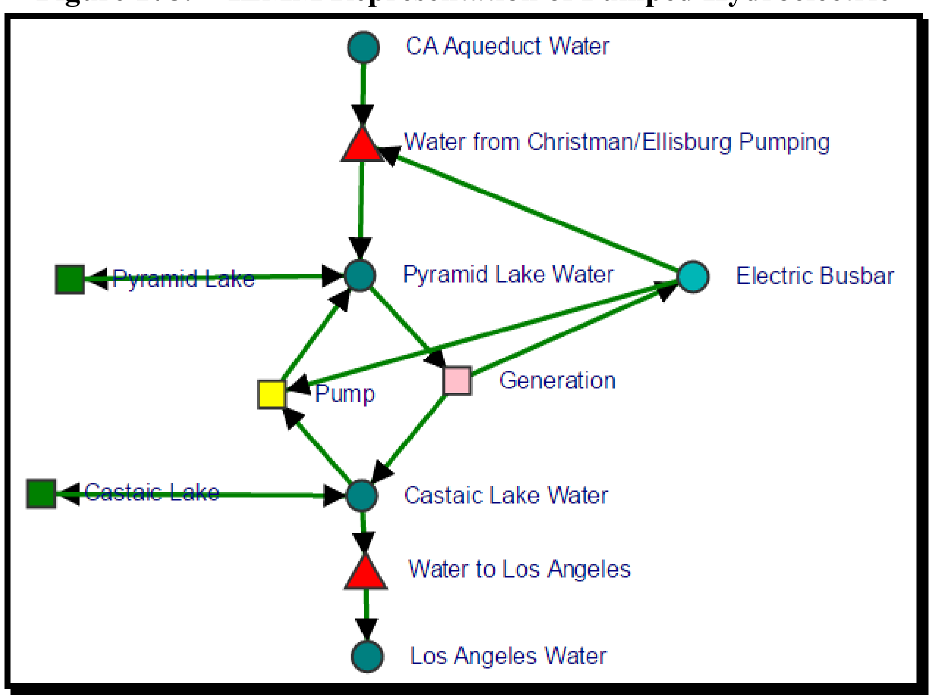


Figure 178: IEMM Representation of Pumped Hydroelectric



We see in Figure 178 a rather complex relationship between electricity and water. We see that the pumping activity that delivers water over the Tehachapis and into the high head hydro (Pyramid Lake) draws from the electric busbar in the region and pays the local price of electricity, which is highly time dependent. The pumping activity that moves water back up the hill from Castaic Lake to Pyramid Lake also draws from the same local price of electricity, which is highly time dependent. The generation activity that occurs when water is let out of high head Pyramid Lake and flows through the penstock (generator) to low head Castaic Lake delivers electricity to the local busbar. Both Pyramid Lake and Castaic Lake are characterized by storage nodes because they have a fixed maximum water reservoir and fixed upper bounds on inlet and outflow rates.

While the California water project might be me more complex than most (because it begins by pumping water over the Tehachapis into Pyramid Lake, the pumped hydro portion of the project involving Pyramid Lake and Castaic Lake is extremely standard. This IEMM design allows EIA to carefully represent hydro storage. As you know, this is not trivial around the world. Virtually every pumped hydro project in the nation and the world has common elements with the illustrated water project. IEMM must be able to represent this carefully and correctly for many systems in the world and the North American continent. IEMM cannot omit this.

11.7 Seasonal Hydro Is Important Almost Everywhere in the World (e.g., Ecuador)

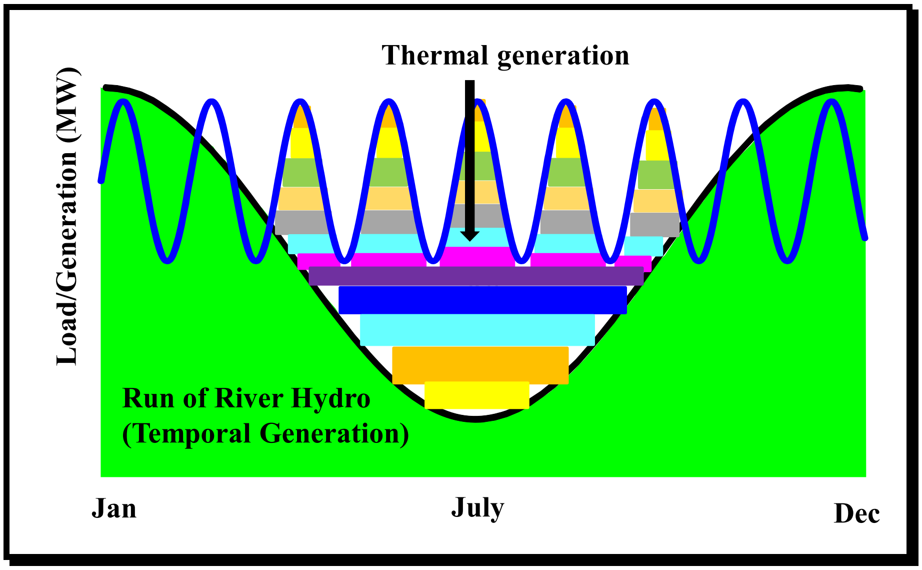
We use the Ecuadorian power system, which we modeled extensively for a client who was thinking of buying Ecuadorian power generation and transmission. In Ecuador, during their “wet” season, the rainfall in the high Andes and the flow down the streams is very high. Ecuador has installed run of river hydro in the mountains so that the rainfall can turn the paddlewheels on the generators but reservoirs and storage can be avoided. We are not sure why there are no reservoirs, but we suspected it was lack of capital and environmental sensitivity. De facto, those run of river hydro facilities generate when the rain falls, and the rivers run, but they do not generate during the “dry” season when there is no rainfall and the rivers do not run. The magnitude of generation during the normal and abnormally wet season was sufficient to completely meet all load in the country. However, the magnitude of generation during the dry season and during drought years was insufficient to meet Ecuadorian load. They were forced to have thermal generators installed to make up the difference between hydro run and load. Figure 179 illustrates. The green area is the hydro run, which is high during what the Northern Hemisphere considers the “winter months.” Run of river hydro meets the entire load in the country during those months, and there is no need for thermal generation. By contrast, during what the Northern Hemisphere considers the “summer months,” rainfall is low, and therefore run of river hydro is low. A whole sequence of installed thermal generators must run during the dry season. De facto, their thermal generators run about half the year and are thus classical intermediate and peaking plants.

During the dry season, the price of power is very high because the marginal generator is an oil or gas thermal generator. During the wet season, the price of power is very low because the marginal generator is a run of river hydro generator. It is crystal clear from Figure 179 that hydro needs to be represented seasonally and chronologically. This IEMM design allows EIA to do so.

11.8 The ArrowHead Storage Model

This section summarizes the correct and proper way to represent a storage asset in an economic context. The fundamental concept underlying the proper model of storage is that every storage asset resides at a hub. The storage asset buys electricity (or equivalently water) from the hub at the then-prevailing market price, and it sells electricity (or equivalently water) to the hub at the then-prevailing market price. In the network microeconomic context, there is a storage asset at the market hub contiguous to that storage asset, and it buys and injects hydro from the hub, places it into inventory, and sells hydro from that inventory back to the market at certain other times. The storage asset owner will strive to buy and sell hydro so as to maximize his profit by so doing.

Figure 179: Renewables and Hydro Are Important Everywhere (e.g., Ecuador)



There are storage assets located in a distributed fashion scattered throughout the entire electricity market, each one situated as in Figure 180.[25] Some storage lies in the upstream, some lies along power lines, and some lies in consumption areas. All are situated with respect to local market hubs as shown in Figure 180.

To illustrate how the IEMM storage model must work, consider a highly simplified situation in which there are N time intervals of equal length. For the sake of discussion, let us assume that they are hourly. Figure 181 illustrates the time structure for this summary discussion of the storage model. This might represent a 168-hour week (as in the example forthcoming).

Even though the time intervals are hours in this discussion, they could just as well be weeks or months for hydro or gas. We will require a bit of mathematical notation to summarize the storage model:

N=number of time points under consideration.
$I(t_0)$=inventory in the storage tank at the beginning of the model horizon.
$I(t)$=inventory in the storage tank at time t, $t=t_1, t_2, \ldots, t_{N-1}$.

---

[25] A definitive textbook on storage in an economic system has been written by Prof. Jeffrey Williams (Stanford and UCD), the preeminent storage economist in the world as we understand it. This storage model is quite consistent with Prof. Williams' models.

η_S=efficiency of injection (1 minus losses during injection).
s(t)=quantity purchased from the local hub and injected to storage at time t=0,1,2,…,N-1.
η_W=efficiency of withdrawal (1 minus losses during withdrawal).
w(t)=quantity withdrawn from storage and delivered to the local hub at time t=0,1,2,…,N-1.

Figure 180: The IEMM Storage Model

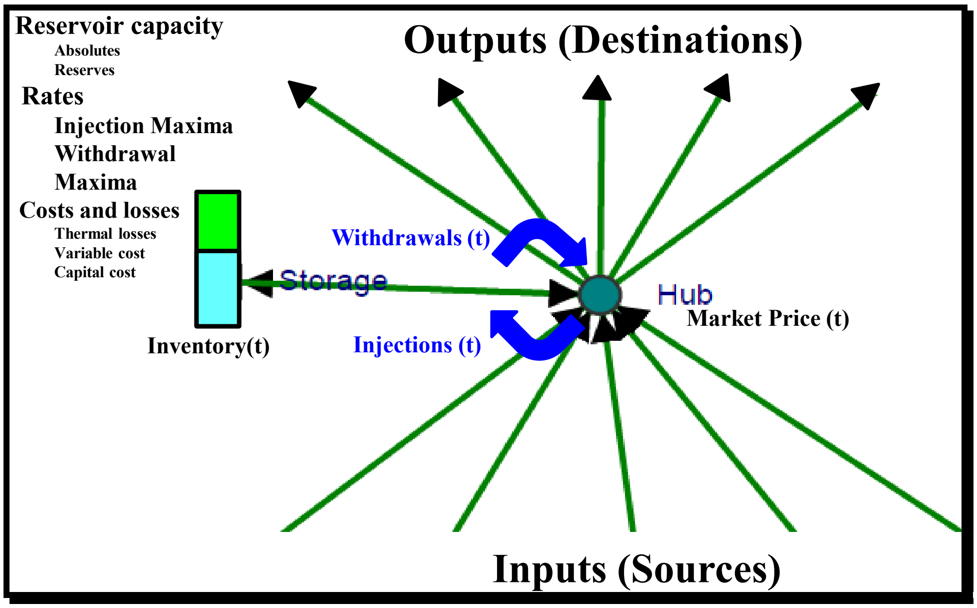
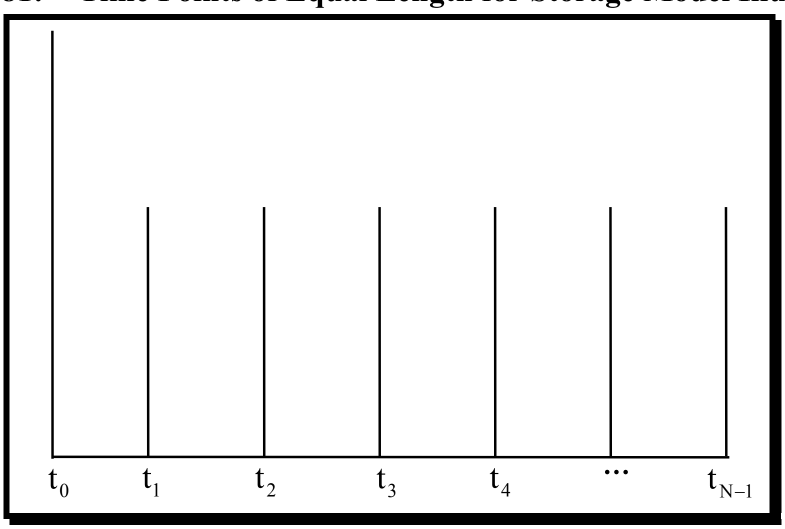


Figure 181: Time Points of Equal Length for Storage Model Illustration



Considering the first time interval, which goes from 0 to 1, notice that if we take the quantity s_0 from the market at time t_0, the quantity delivered to the storage tank will be $s_0(1-\eta_I)$. This accounts for losses upon injection to storage. Alternatively, if we want to place the quantity w_0 into the market, we have to take the quantity $w_0(1+\eta_W)$ out of the tank. This accounts for losses upon withdrawal to the market and is summarized in Figure 182.

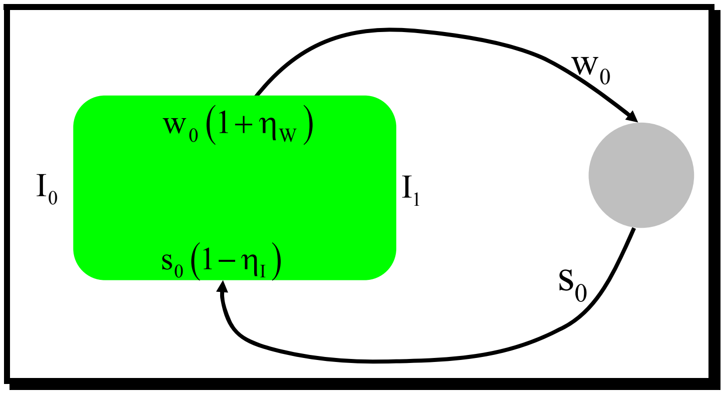
This means that the inventory in the tank at time 1 is

$$I_1 = I_0 + s_0\left(1-\eta_I\right) - w_0\left(1+\eta_W\right)$$

If we follow the same logic, the inventory in the tank at time 2 is

$$I_2 = I_1 + s_1\left(1-\eta_I\right) - w_1\left(1+\eta_W\right)$$

Figure 182: Injection and Withdrawal Balance at Time 1



Substitution of the first equation into this second equation gives the expression

$$I_2 = I_0 + \sum_{k=0}^{1} s_k\left(1-\eta_I\right) - \sum_{k=0}^{1} w_k\left(1+\eta_W\right)$$

The pattern of what the inventory is becomes clear

$$I_n = I_0 + \sum_{k=0}^{n-1} s_k\left(1-\eta_I\right) - \sum_{k=0}^{n-1} w_k\left(1+\eta_W\right) \quad n = 1, 2, ..., N-1 \qquad \textbf{(6)}$$

The inventory in the tank at any time point t_n is the inventory present in the storage reservoir at time 0 plus cumulative hourly injections (adjusted for losses) less cumulative hourly withdrawals (adjusted for losses) up until time t_{n-1}.

There are two concepts related to the size and capacity of the storage assets needed for our model:

R(t) = maximum size of the reservoir at time t.

ρ(t)=lower bound on working inventory at time t (reserve requirement mandated by the regulator or reserves carried by the operator as a matter of practice). Regulators or operators may not allow inventories to go to zero; operators generally keep some level of reserves in their tank so that they can meet peak deliverability requirements.

The inventory in the tank at time t_i must be larger than the reserve requirement but smaller than the maximum size of the reservoir:

$$\rho(t) \le I_0 + \sum_{k=0}^{n-1} s_k (1-\eta_I) - \sum_{k=0}^{n-1} w_k (1+\eta_W) \le R(t) \quad n = 1,2,\ldots,N-1. \qquad \mathbf{(7)}$$

This will <u>be the first constraint equation</u> (de facto the production function) that will govern profit maximizing behavior on the part of the storage facility owner.

In addition to reservoir constraints, there is a maximum rate of injection (governed by phenomena such as amperage on the inbound wire and a maximum rate of withdrawal (governed by phenomena such as amperage on the outbound wire). The storage model needs to take account of these injection and withdrawal maxima. The simplest possible way to do so is to posit the following types of constraints

$\sigma(t)$=upper bound on injection rate at time t (simplified for illustration).
$\eta(t)$=upper bound on withdrawal rate at time t (simplified for illustration).

This implies the following constraints that govern profit maximization:

$$\begin{aligned} 0 &\le s(t_i) \le \sigma(t_i) \\ 0 &\le w(t_i) \le \mu(t_i) \end{aligned} \qquad (8)$$

This will be the second constraint equation (de facto the production function) that will govern profit maximizing behavior on the part of the storage facility owner.

The preceding two phenomena govern the inventory level and how rapidly the storage asset can augment it or draw it down. We now must turn to the economic and behavioral part of the storage model. To do so, we need a few more mathematical terms:

ϕ_S = variable cost (handling cost) injection excluding any inventory storage charges.

ϕ_W = variable cost (handling cost) of withdrawal excluding any inventory storage charges.

D_t=discount factor from time t back to time 0.

We are now prepared to write the equation for the profitability that the owner of the storage facility can achieve by selling and buying electricity to and from the hub to which it is attached. The storage field operator takes the price of electricity at the contiguous hub to which he is connected as given and maximizes profit by buying electricity from the hub and selling electricity to the hub. In particular, the storage field operator defines his profit according to the equation:

$$\text{Profit} = \sum_{k=0}^{N-1}(p_k - \phi_W)\, w_k D_k - \sum_{k=0}^{N-1}(p_k + \phi_S)\, s_k D_k + \text{Ending Value} \tag{9}$$

The first summation is the discounted present value of profit the storage facility operator receives from the sale of electricity back to the local hub. Notice that it contains the price of electricity **minus** the variable cost of handling. Losses upon injection and withdrawal are going to be taken account in the constraint equations as we maximize this measure of profit. The second summation is the discounted present value of cost the storage facility operator bears from having to purchase electricity from the local hub. Notice it contains the price of the electricity **plus** the variable cost of handling. The third term is the value of the inventory of hydro in the tank at the end of the model horizon. We are not going to dwell upon that except to say that it is needed so that the storage facility operator does not "game" the end of the model horizon.

The storage facility owner will then set his injection and withdrawal strategy over time so as to maximize profit subject to the two constraints developed above. That is, the storage field operator will set his injection and withdrawal schedule so as to solve the following profit maximization problem:

$$\begin{aligned}
&\text{MAX} \quad \sum_{k=0}^{N-1}(p_k - \phi_W)\, w_k D_k - \sum_{k=0}^{N-1}(p_k + \phi_S)\, s_k D_k + \text{Ending Value} \\
&\text{SUBJECT TO} \\
&\rho(n) \le I_0 + \sum_{k=0}^{n-1} s_k (1-\eta_I) - \sum_{k=0}^{n-1} w_k (1+\eta_W) \le R(n) \quad n = 1,2,\ldots,N-1 \\
&0 \le s(t_i) \le \sigma(t_i) \\
&0 \le w(t_i) \le \mu(t_i)
\end{aligned} \tag{10}$$

(We are not going to discuss the ending value model.) This is the fundamental behavioral relationship governing storage operation no matter where or what time frame—profit maximization subject to injection and withdrawal maxima and subject to reservoir maxima and minimum reserve requirement. The losses, reservoir size, initial model inventory, and all other phenomena are explicitly considered and endogenized herein. One can see all those variables contained in the profit maximization problem. This model of storage, precisely the one that must be in IEMM, allows storage to be sited in the model in disparate locations throughout the world—generation areas, renewables areas, along the transmission system, or near end-use.

It doesn't take a rocket scientist to observe that there is precious little possibility of building or running a model of storage using a global social welfare maximization approach. No one has ever done it to our knowledge, and prospects are minuscule to zero. Cost minimization or global social welfare maximization simply do not respect the behavior of goal seeking storage agents in the electric power system. Any search for the goddess of global social welfare maximization is likely to be futile, and it will miss the issue of energy storage and load-shifting, which we have shown to be central to current and possible electricity systems.

In summary, to understand storage and load-shifting, IEMM must insert a storage node everywhere in the power system that a storage asset exists, or EIA thinks could exist and link it up to the local hub. It will operate as an economic agent would, and it will impact volumes and prices of logistical assets as well as production and consumption assets in precisely the way summarized in the first part of this section. We do not see how IEMM could model storage in any other way. Any other way (such as arbitrary storage injection-withdrawal schedules "matched" to history) is likely to be dead wrong and misleading.

With this construct, all the myriad storage operators everywhere in the world are explicitly assumed to be "players," i.e., agents. They all interact, and they all individually modify supply and demand patterns as indicated above, and in precisely the right way. And they are located all over heck and gone, de facto competing with and complementing each other. Storage is intrinsically an "optimization problem within an economic problem," and EIA is advised to represent it this way.

The recommended IEMM storage model represents every storage asset in the world:

- Individually
- Independently

By so doing, every storage node in the world thereby competes with and/or complements every other storage node and pipeline and supply source in the world. This is the quintessence of logistical modeling. Every storage node competes against every other node in the model to make money and re-temporalize flows. Every storage node is a proactive, price taking, profit maximizing competitor. This is what renders network microeconomic modeling unique and accurate.

11.9 Energy Limited Supplies

Suppose we are in a region with the supply stack and six demand tranches as shown in Figure 183. We see six prices, one for each of the six tranches of load. We see that

- the price p_6 persists for the number of hours H_6 at the megawatts MW_6,
- the price p_5 persists for the number of hours H_5 at the megawatts MW_5,
- the price p_4 persists for the number of hours H_4 at the megawatts MW_4,
- the price p_3 persists for the number of hours H_3 at the megawatts $MW_{3,}$
- the price p_2 persists for the number of hours H_2 at the megawatts $MW_{2,}$ and
- the price p_1 persists for the number of hours H_1 at the megawatts $MW_{1.}$

We now ask what will happen if there is an energy-limited resource that can be provided to this market at a cost lower than the price $p_{1.}$ In that situation, there are a number of MWh that can be provided to this market at a low cost. The situation would appear as in Figure 184. The supply stack in Figure 183, i.e., the green curve in Figure 183, is represented in aggregate in the rightmost node in Figure 184. The load duration curve in Figure 185 is contained in the Load node in in Figure 184. The energy limited resource is represented as in Figure 186. The important thing to notice about Figure 186 and its position in Figure 184 is that delivery of energy from the energy limited resource **<u>is not constrained to any particular subtime</u>**. The energy can come in any hour you want. There is complete ability to release the energy limited resource at any time one might wish. That is the key; the structure on the left of Figure 184 is not constrained to occur in real-time, coordinated precisely with any particular subtime of load. The energy limited resource can be husbanded and released whenever one might wish.

Figure 183: Inelastic Supply-Demand Equilibria Are Calculated for Each of the Six Load Tranches

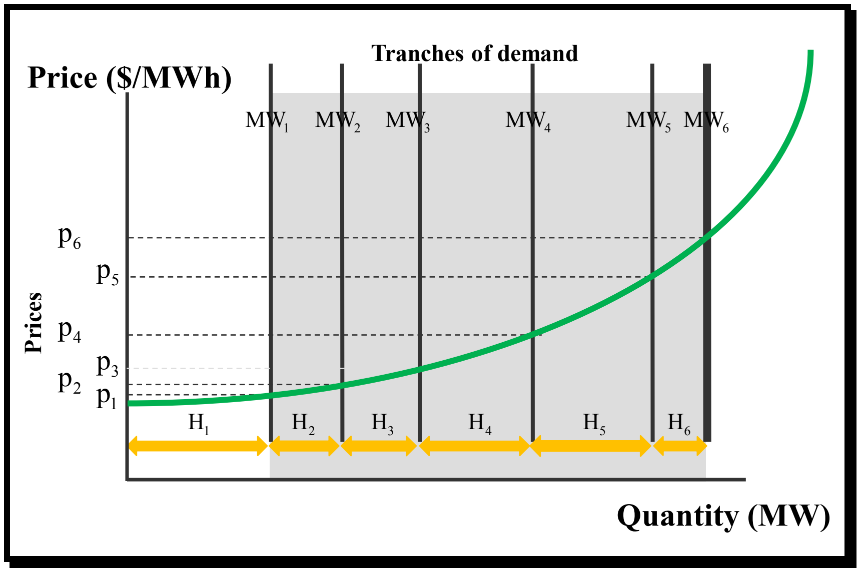
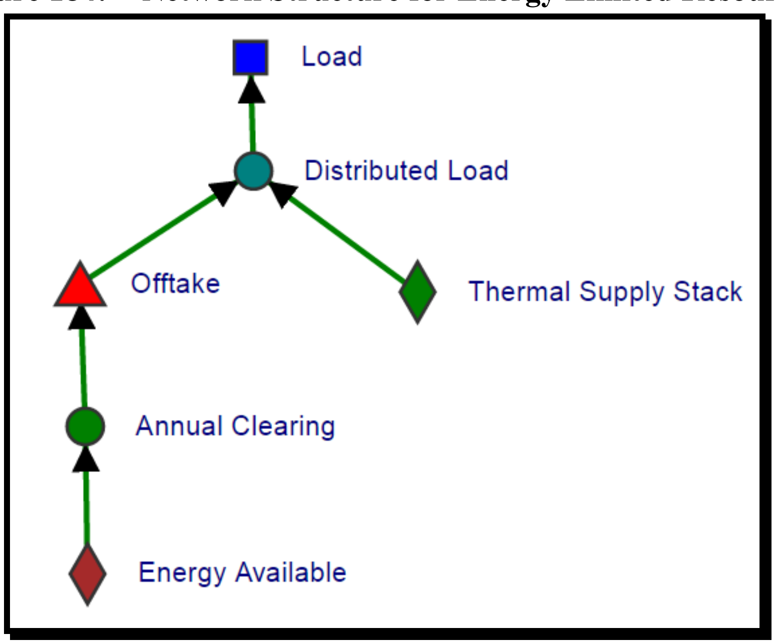


Figure 184: Network Structure for Energy Limited Resources



As indicated in Figure 186, the energy limited resource can provide energy at a constant marginal cost over all subtimes until the energy available is expended. And this occurs in a very specific and measurable way, one that IEMM can carefully model using the proposed design. And the method flies in the face of traditional, old school dispatch models. The method (and the real world) must find the specific price level that starts with the peak, moves to the next, moves to the next, etc. until all the energy available is expended. The situation appears as in Figure 187. The price starts with the peak price and inches its way down until precisely such time as 100 percent of the limited volume of energy is gone. Then it stops inching down. At that price, all the energy limited load is placed into the market and expended, and profit is at its absolute maximum level. This is precisely what a profit-seeking owner of an energy limited resource would do. And that is what the IEMM model will predict that he does. The resulting prices, very difficult to calculate in practice, are calculated precisely according to Figure 187 herein. The energy limited resources starts with peak and drives every price downward until the limited amount of energy is expended.

Figure 185: Disaggregation of Each Month into Six Load Tranches

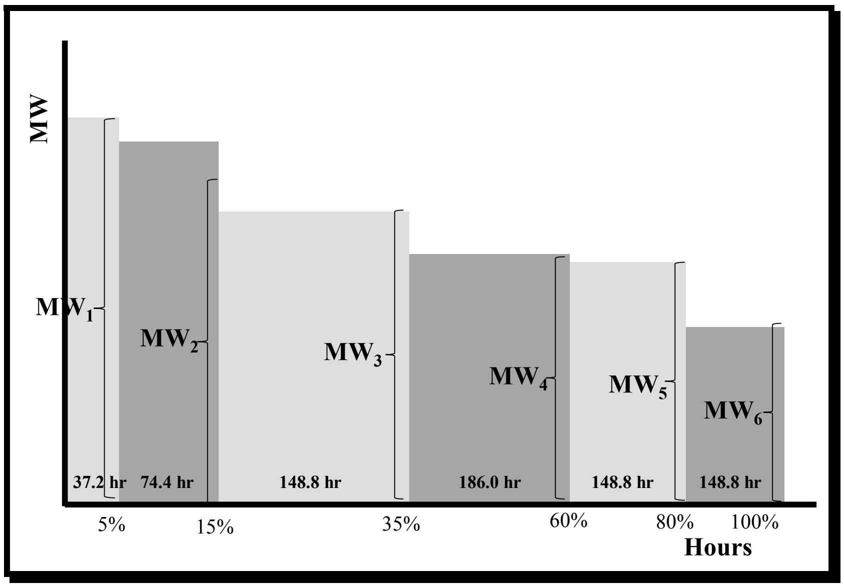
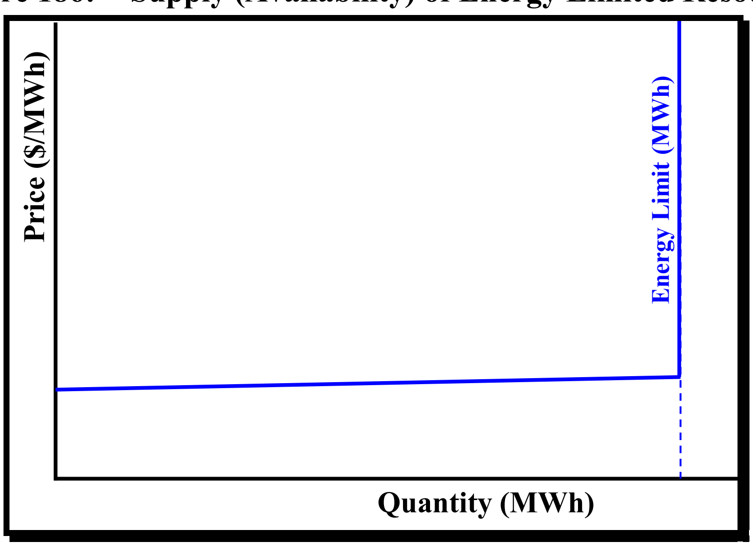
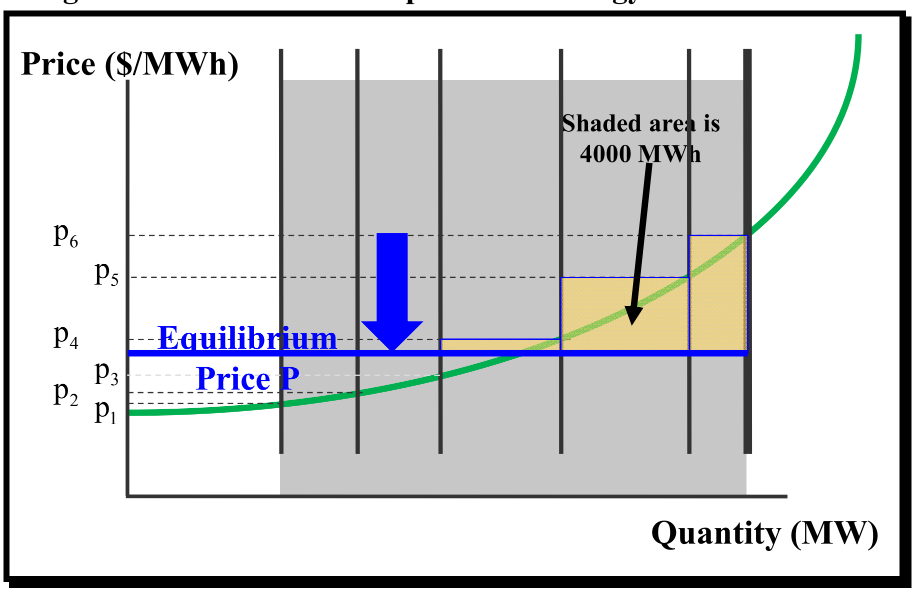


Figure 186: Supply (Availability) of Energy Limited Resource



The situation is slightly more difficult, but we can handle it using the structure in Figure 184. We simply put a capacity bound into the red triangle. The process is similar to that in Figure 187 except that when there is a capacity constraint, IEMM cannot fill all the blocks at the same rate. The ones that cannot be filled rapidly enough command a higher price. The marginal source in those blocks comes from the supply stack. The ones that can be filled rapidly enough command the lower price P. The energy limited supply is the marginal source during those times.

Figure 187: Price that Expends the Energy Limited Volume



11.10 Summary

The ArrowHead endogenous storage model will buy and sell from the hub to maximize profit. Such behavior will serve to decrease price at time of peak and increase price at time of off-peak. That is the quintessence of storage. Furthermore, the proper endogenous operation of storage naturally decreases the incentive to build further storage. The model can find the right amount of storage, the amount that is profitable in a given region but no more.

One important observation: with this storage model, the storage injection and withdrawal schedule is not exogenous, it is endogenous. We have seen all too many times linear programming and global optimization practitioners making the blithe assumption that storage asset operators will dispatch storage in the future the same way they have dispatched in the past. ("Dumb in the past; dumb in the future! Simpleton automatons, those storage guys! They do it the same every year!) This is laughably wrong, storage is real time and adaptive to economic signals, and the recommended IEMM storage model will be completely adaptive. (We have seen such models turn peak/off-peak prices upside down in practice, the peak price being lower than the off-peak price, because of some historical assumption.) The recommended model asserts that the storage asset owner will coldly and rationally react to prices and will endogenously operate his facility so as to maximize profit. He will not have total flexibility to do so. He will be restricted by the size of his reservoir and the reserve requirements he places on his reservoir. He will be restricted by maximum rates of injection and maximum rates of withdrawal.

The recommended storage model has the advantage that:

- Hydro/storage/electricity supply and demand are balanced at every hub at every time point. For example, the monthly hydro injection and withdrawal and electricity markets will equilibrate everywhere. That is powerful.

- Storage affects market price, and at the same time market price affects storage. The recommended model closes the loop, representing the fact that storage operation strategy depends on price and price depends on storage operation strategy. This is profound, and we are not aware that any other storage model embedded within an economic model has ever been able to accomplish it.

- It calculates inter-temporal opportunity cost of storing versus selling at every time point.

Some have wondered whether the recommended model is too cold hearted and rational a model of the storage asset owner. Evidence points to the contrary. We have seen empirical evidence that owners of storage entitlements (meaning either storage field owners per se or third parties who have contracted for access to storage assets) do indeed maximize profit in exactly the fashion articulated in this section. If one wants to assume irrationality or inertia as a driver of storage field operation, one can modify the assumptions or operation of the recommended model. However, with lesser models, EIA can never truly understand what effective operation really means, if even one of many scenarios.

Certainly local distribution companies and public utility commissions (or their equivalents) around the world have traditionally viewed storage as insurance against high peak demands. Their reward and penalty structure forces them to behave this way. The recommended storage model represents this phenomenon very nicely. Using the reserve requirement, we represent the fact that there is a "nondiscretionary" portion of storage, a regulated portion of storage, which simulates the insurance aspect of storage that say a regulatory body in Europe would strive to enforce at time of peak. However, equally importantly, there is a residual piece, which is dispatched economically in pursuit of "buy low/sell high" profits. It is the combination of the two that impacts market prices and the value of the storage asset, and the recommended storage module represents that beautifully.

There are a large number of subtle details that must be present in the storage model EIA uses, and we have omitted those here. We have put forth the essence. We do not see how this type of model could be incorporated into a linear programming, MPEC, or complementarity formulation. Those ilks of models are going to omit endogenous representations of storage. The dimensionality alone of a respectable storage model would dramatically exacerbate the dimensionality of those approaches, which was discussed in detail in our former CDR.

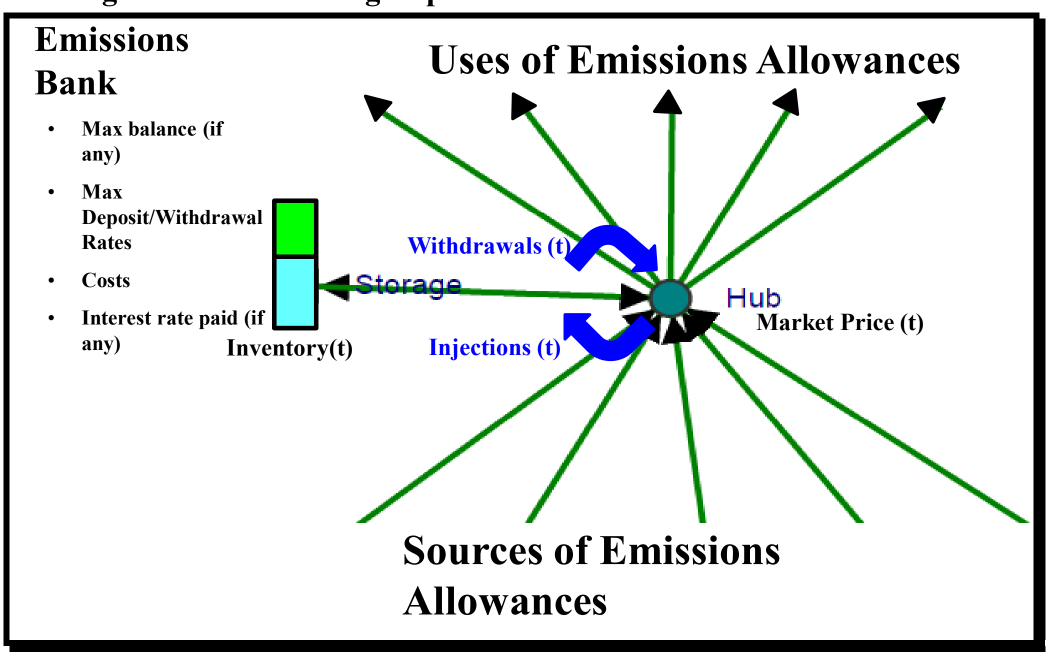
12 EMISSIONS AND REC BANKING

How many times have we heard people say that they are going to buy up CO2 allowances today (or generate CO2 allowances today by planting rubber trees in the Philippines) so that they can "bank" them for later use when they really need them? Some regulators have apparently bought into this type of thinking. If people were to buy up CO2 allowances today, they would assuredly bid up the price today. More demand for emissions allowances today means higher price today. If people were to consume or sell CO2 allowances in the future, they would assuredly drive down the price tomorrow. More allowances put onto the market in the future means lower prices in the future. Clearly the decision to buy or acquire CO2 allowances today and sell them tomorrow is a decision to drive up the price today and drive down the price tomorrow. Intertemporal speculation about emissions allowances is not at all benign about the price of those emissions allowances.

How would we represent emissions banking, which is possibly important in certain venues around the world and may become even more important as time goes forth?

The answer is clear, as indicated in Figure 188. Notice at the bottom, there are a number of disparate sources of emissions credits. Governments can auction them. Governments can assign them. People can engage in green activities and generate them. People can bring them in from other jurisdictions (if they are honored in this jurisdiction). This creates a supply of emissions allowances to a traded emissions market as indicated in Figure 188. At the same time that emissions allowances are being created in upstream markets, emissions allowances can be withdrawn from the emissions bank as shown in the bottom right arrow in the diagram in Figure 188.

Figure 188: Banking Deposits from and Withdraws to a Market



At any given time, a number of emissions allowances will have to be used in the market to offset emissions that are occurring. Those deliveries to the market come out the top of the hub in Figure 188. Out the top of the hub in Figure 188 also comes deposits to the emissions bank taken from the market.

Not surprisingly, emissions banking is nothing but a good, old-fashioned, run-of-the-mill storage process. People deposit emissions allowances into the bank when prices are low, and they withdraw emission allowances from the bank when prices are high. They are not marching to the beat of some mythical global social welfare function. The very thought is ludicrous on its face. They march to the beat of their own pocketbook.

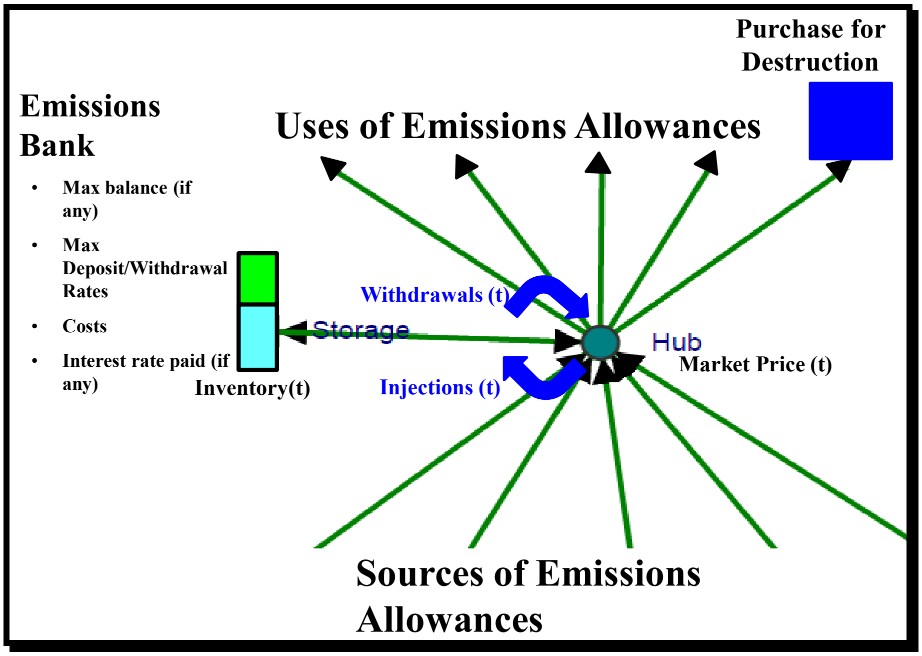
All we have to do is put the emissions bank in the form of Figure 188 at every point in the system at which emissions allowances are generated and disposed, i.e., the point of exchange.

We can do precisely the same thing with RECs, banking them today for potential future use tomorrow. People are speculating them.

In the past, there have been a number of parties who have said that they would buy up a number of emissions allowances or they will buy up a number of RECs and simply sit on them or destroy them so they cannot be traded or redeemed. They will systematically destroy them or take them out of service. They will do so because they want emissions regulations to be even more binding than government agencies may want. They want even more renewables than the RPS. There is nothing the matter with that and nothing to preclude that in most regulatory or power settings. We can represent that quite easily with the structure in Figure 189. At the upper left is an inelastic demand curve that buys and destroys emissions allowances from the tradable emissions allowance market. This can measure the impact of players who buy and destroy emissions allowances or RECs as a strategy to increase the pace of emissions regulation or renewables entry.

Emissions banking and emissions acquisition by means other than direct generation by government are potentially important in every electric system in the world. EIA cannot allow naïve consideration or no consideration of emissions banking. This requires the sophisticated model of emissions allowance storage put forth herein. These types of emissions or REC banking models result in emissions or REC price tracks that borrow from and sell to the emissions bank at a discount rate. Those prices equilibrate with discount rates, as they must. If anyone asserts that a government or agent will be buying and banking emissions, it is not acceptable to have a naïve or incomplete view of that. Quite the contrary, EIA will need the emissions banking model articulated in this section, which IEMM provides under this design.

Figure 189: Purchasing Emissions Allowances for Disposal



13 CONSERVATION

The ArrowHead modeling system and its predecessors have been unique in their ability to model conservation and end-use. Even though conservation and end-use are not totally synonymous, end-use is a sufficiently important contributor to what we know as conservation that it needs to be modeled on an agent basis within IEMM as characterized in this section. This section is a continuation of work that ArrowHead and its predecessors have done for the Electric Power Research Institute (EPRI) and the Gas Research Institute (GRI) and a number of utility companies' integrated resource plans (IRP) over the years but not disclosing any confidence from such work. The work actually dates back to the late 1970s for the TVA with their Strategic Analysis Model (SAM).

If we are really going to understand conservation and load shape, we have to build what we term an "establishment" model, i.e., a model of sectoral consumption. We build this model of sectoral consumption, and then we append it as a sector to the generation model. The establishment we consider in Figure 190 is a residence, a single family dwelling. (IEMM would have to model multifamily dwellings, commercial establishments, industrial establishments, transportation, etc.) We see at the top of Figure 190 the seven basic categories of residential establishment demand

- **Space Cooling (A/C).** The magnitude of space cooling/air conditioning demand varies by hour, day, week, month, and year. There is a space cooling demand duration curve, by months, and it has to be estimated. This can work by analogy from United States air conditioning loads, weighted of course by "penetration" or "saturation" in other world systems. Ultimately, A/C demand is hour by hour. Space cooling demand is akin to a load duration curve, but it is not a load duration curve. It is an energy service duration curve—when do you need the energy service we call air conditioning.

- **Space Heating.** The magnitude of space heating demand varies by hour, day, week, month, and year. There is a space heating demand duration curve, by month, and it has to be estimated. This can work by analogy from United States space heating loads, weighted of course by "Penetration" or "saturation" in other world systems. Ultimately, space heating demand is hour by hour. Space heating demand, like air conditioning, is akin to a load duration curve, but it is not a load duration curve. It is an energy service duration curve—when do you need the energy service we call space heating.

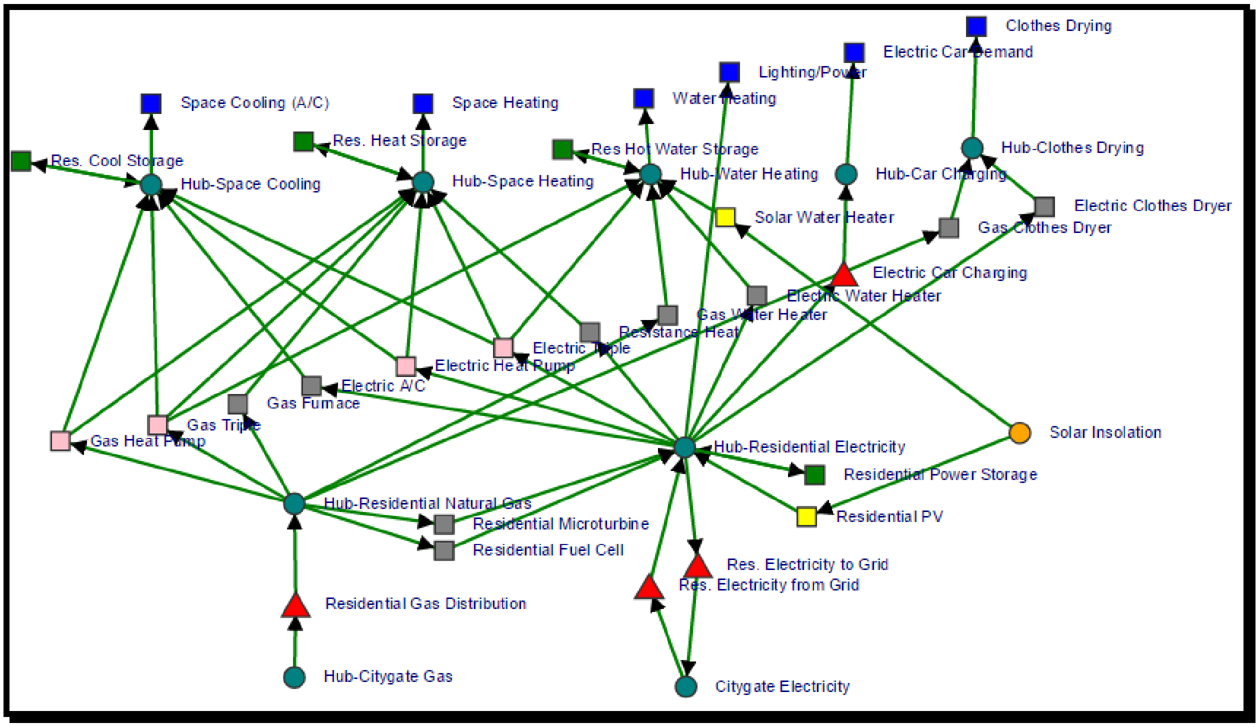
- **Water Heating.** Water can be heated as a direct product or a thermodynamic byproduct.

- **Lighting/Power.** This is a large and important component of residential demand—direct consumption of electricity. This would include dishwashing, parasitic losses, computers, clocks, electric stoves and ovens, etc.

- **Electric Car Demand.** This is potentially a large and highly times varying demand. In principle, it can occur every night of the year during time of off-peak, and it can disappear during the day. Conversely, if power prices are not correctly set, it can occur during the day (because of low prices) and can exacerbate daily load. IEMM has to consider this emerging component of direct electric demand. If these cars can generate electricity using gasoline during any hour, we would have to deliver that to the grid.

- **<u>Clothes Drying.</u>** This is a traditional gas versus electric device substitution matter.

Figure 190: Model of Residential Energy Consumption and Load Modification



Beginning at the upper left of the diagram, refer to the Hub-Space Cooling. Notice what comes into it and competes to meet the fundamental end-use of air conditioning, which means cool rooms in the summer. It doesn't mean Btus; it means cool rooms in the summer. The outputs of the air conditioning technologies are "cool rooms in the summer," i.e., Btu equivalents of cold air pushed through the register:

- Gas Heat Pump
- Gas Triple Integrated Appliance
- Electric Air Conditioning
- Electric Heat Pump
- Electric Triple Integrated Appliance

Each of these techniques feeds into the Hub-Space Cooling, which represents a fundamental end-use energy service. The demand for the energy service has a highly temporal, hour to hour pattern. Not only that, each of these techniques to provide it has a highly time dependent, highly temperature dependent "coefficient of performance," which is the reciprocal of what we represent as an input-output coefficient or a heat rate equivalent. Substitution (on a vintage basis) of the preceding air conditioning technologies means substitution of an hour by hour input-output coefficient, and that in turn leads to hour by hour reduction in fuel consumption because of the hour by hour change in heat rate.[26] As an example, if an electric heat pump enters, it will have to follow the hour by hour air conditioning load during the summer, and it will have to follow the hour by hour space heating load during the winter. At its elevated coefficient of performance (reciprocal heat rate), it will consume less electricity than some other cooling technique. This is the only conceivable way to model conservation, and from a decentralized decision-making perspective. That is precisely the way the recommended IEMM will work and will allow EIA to operate.

There are two other points worthy of mention—thermal storage of cool air and vintaging of capital stocks. We know in the summer we could freeze water at night and blow air across it to melt it during the day. We use the ice example only to illustrate that there are current and may be future technologies that produce "cold" at time of off-peak and deliver it to market at time of peak. That technology is represented by the Res. Cool Storage node at the upper left. This technology has the potential to create cooling during the off-peak and deliver it during the on-peak. This technology would level the air conditioning load (and thereby the gas or electricity load depending on the type of technology that is producing it). IEMM must have a storage node as described previously, and it must be hooked up to residential air conditioning load. Such storage of "cold" is likely to be easier to implement in new establishments

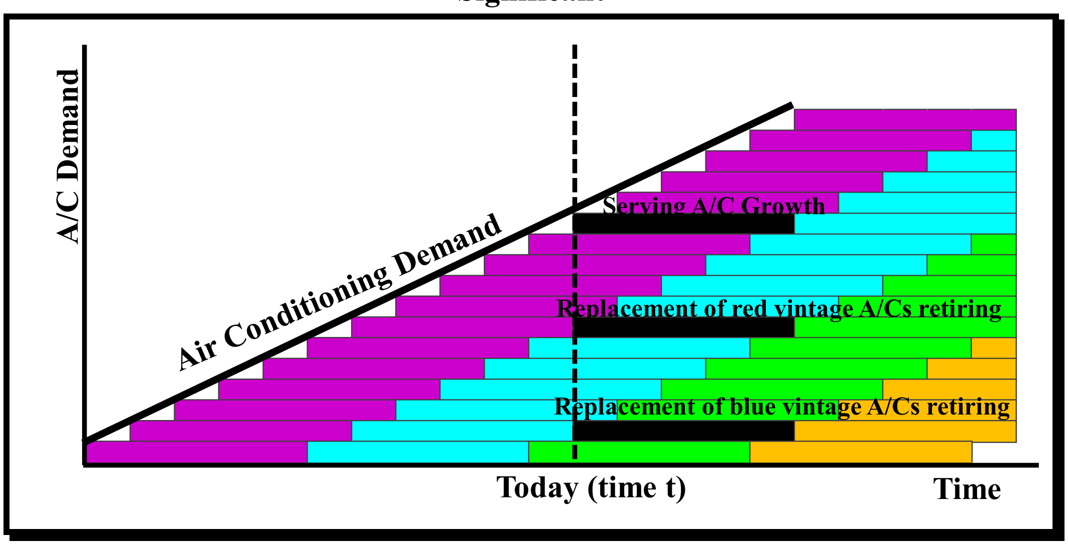
One of the other, and critically important, issues in conservation is capital stock rollover. HVAC equipment lasts 15-20 years, and people don't usually tear it out and replace it prematurely. We have the diagram in Figure 191, which is needs to be automatically calculated in the end-use device node calculations in IEMM. In particular, the end-use device node calculations we have carefully vintage the capacity in every end use node in the system. That means we get the vintage portion of capital stock rollover for every end-use technology in Figure 190 lockstep correctly and accurately. Lacking this vintaging (which as far as we know is only possible in network economic equilibrium models and not global social welfare maximization models), we have no hope of getting the air conditioning and other establishment conservation issues modeled. And we have no hope of modeling their impact on load.

The other phenomenon IEMM must see is the change in fuel consumption that results from the penetration of say an electric heat pump as constrained by the vintage schedule in Figure 191. To the extent this fuel is electricity (and it IS), this has a profound load-shifting and decreasing change. We will have the situation in Figure 192 and Figure 193. For a given summer month, Figure 192 represents the hour by hour air conditioning load. This is the demand for "cold air through the register" in each and every hour of the month. This is the output of the air conditioner that will have to happen during that month. If IEMM must have an electric air conditioner operating during that month, IEMM must will divide the air conditioning load by the tranche-by-tranche (load specific) input-output coefficient for the air conditioner and calculate the electric load and calculate the top duration curve in Figure 193. If on the other hand IEMM must have an electric heat pump operating during that month, IEMM will divide the air conditioning load by the tranche-by-tranche (load specific) input-output coefficient for the electric heat pump and calculate the bottom duration

[26] We should emphasize that we can put the environmental inputs into each of these end-use conversion techniques—SO2, NOx, mercury, CO2. That would allow us to

curve in Figure 193. In this fashion, IEMM calculates not only the pure "conservation" effect of the penetration of a higher efficiency device like and electric heat pumps, but IEMM must also calculate the detailed and specific implication for load variation. We thus calculate not only the aggregate MWh saving from penetration of a conservation technology but IEMM must also calculate the hour by hour, tranche by tranche change in electricity load that such conservation would imply. There is no substitute for the accuracy of this. We can see from the network structure in Figure 190 this is in IEMM. Just drag and drop the vintage residential establishment equipment model in, fill it with data, and voila, you have the requisite model of the residential establishment and its implications for electricity (and gas) temporal load. You also know the propensity of individual agents for adopting conservation technology. In the real world (as contrasted with the policy world), conservation technologies enter when and only when agents buy them. When you stack these graphs end to end, month by month, year by year, you get the true and correct impact of the entry of more efficient end-use technology, both in terms of total MWh but more importantly in terms of MW required to meet that load.

Figure 191: The Air Conditioning Fleet is Carefully Vintaged; Vintage Rollover Is Very Significant



Moving to the right in Figure 190, IEMM must encounter the residential space heating node. As with the air conditioning node (but for different months of the year of course), IEMM must have the prospect of making hot air during time of space heating off-peak and injecting it into the residential establishment at time of on-peak. Thus, IEMM must require the Res. Heat Storage node hooked to the space heat node as shown at the top second from left in Figure 190. This would level the power load substantially for those establishments that are heated by resistance heaters or electric heat pumps.

We note in Figure 190 that the space heating technologies include

- Gas heat pump
- Gas triple integrated
- Gas furnace

- Electric heat pump
- Electric triple
- Resistance heater/furnace

Figure 192: Stairstep Hourly Air Conditioning Demand by Tranche

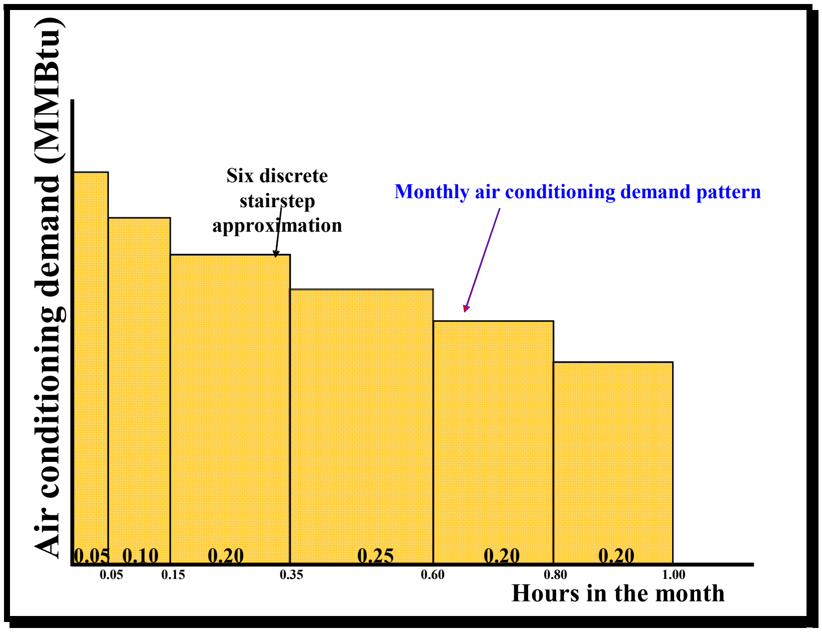
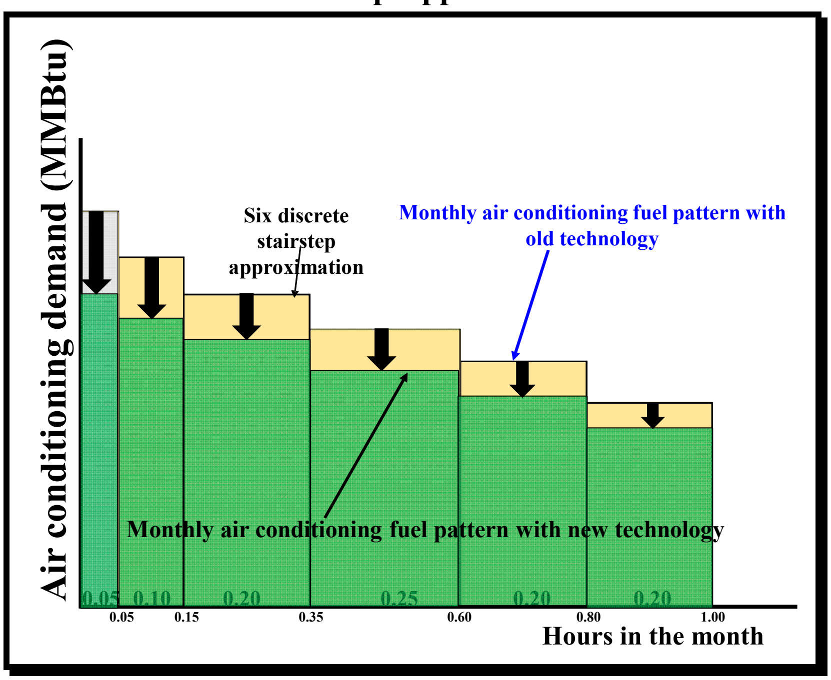


Figure 193: Calculated Stairstep Approximation to Fuel Consumption



Each of these technologies has a link that enters the residential space heating node. Each of these technologies is subject to the selfsame vintaging considerations as we saw with air conditioning, the same vintaging as we saw in Figure 191 for air conditioning. The IEMM design we have put forth here fully vintages air conditioning, space heating, and, in fact, all the other nodes in the model. That is an absolute prerequisite for understanding conservation, load modification, and load-shifting in IEMM. It is the only way IEMM will be able to get true and correct electric load and load shape as derived from the residential establishments in a region.

The tranche-by-tranche analysis of space heating demand and how it implies hour by hour fuel demand associated with Figure 192 and Figure 193 is exactly analogous. We begin with the analogous diagram as Figure 192 but for space heating. We then use the particular space heating technique from Figure 190 and the particular coefficient of performance for heating (which varies by load period potentially because of temperature and other differences) to calculate the time-schedule of fuel use, hour by hour, in space heating.

We move one more node to the right to consider water heating. The sources of water heating are seen in Figure 190 to include

- Gas Triple Integrated Appliance
- Electricity Triple Integrated Appliance
- Gas Water Heater
- Electric Water Heater
- Solar Water Heater

We also see in the diagram a representation of storage of hot water, which can be produced at some time other than time of use and stored until time of use.

Lighting and direct power demand comprise the many direct electricity uses in the residential establishment. Such loads include lighting of course, parasitic loads in space heating and air conditioning, electric cooking, computers, telephones, radios, electromechanical drive, televisions, and the full range of residential establishment electricity uses. These uses are obviously extremely time dependent (depending on time of particular end-use), and they have grown in magnitude in industrialized countries in recent years as homes have enjoyed more and more electric devices.

We have had to distinguish electric automobiles. The reasons are several. First, there is a lot of energy involved in automobile travel and therefore potentially a lot of electrical energy that will transact the local distribution company and into the car. In principle, this use for electric car charging could occur at night and off-peak. However, if electricity pricing were to stay irrational, regulated, subsidized, or cross-subsidized, people could begin to charge during intermediate or peak hours. This can dramatically change the capacity and energy flowing through the distribution system and into the electric vehicle and dramatically change cost structure and price. This structure in Figure 190 allows full and complete consideration of electric vehicles and their impact on retail power loads and prices and load patterns.

Finally, we see at the right the demand for clothes drying. This demand is met by

- Electric Clothes Dryer

- Gas Clothes Dryer

These are fairly straightforward techniques. For electric clothes dryers, there is the load management notion that these devices might or might not be clocked. If they are clocked, they would only be functional say from 900 pm to 900 am and not during the peak times of day. This would represent a mechanical shifting of load by committing dishwashing to the off-peak times. If they are not clocked, they can be operated whenever one wants, and they add to electric or gas demand in a substantially different fashion.

We can see a rather robust set of residential establishment demands. Each of the blue demand curves has a very time dependent, hourly, demand duration curve shape. Keep in mind, these are not "load duration curves" in the classic sense because they do not represent electricity. Quite the contrary, they are energy service demand duration curves, representing the fundamental commodity being demanded and consumed—cool rooms in the summer, warm rooms in the winter, etc. Those demand durations are highly dependent within the month, highly time dependent across the months, and highly time-dependently over the years. They are no different in concept from load duration curves except that they characterize fundamental commodities.

Given this demand structure for fundamental end-use services, the demand for fuels is what economists term an implied demand." Implied demand comes out of a model; it does not go into a model. This is the issue with market power published previously by ArrowHead, and it is the issue here. One simply CANNOT infer demand curves for gas or electricity in the residential sector. One has to model them, as we have put forth here. Once certainly cannot infer demand shapes for electricity in the residential sector. They have to be modeled as implied here.

We see all the new and existing technologies competing to satisfy the fundamental end-uses, all on a highly temporal basis in Figure 190. We see the inputs to those technologies coming from the two fuel hubs—Residential Natural Gas at the left center and Residential Electricity at the right center. These are the highly temporal implied demands for gas and electricity, taking full account of demand show, conservation, and load-shifting at end-use but not yet taking account of similar phenomena for distributed fuels.

Referring first to the Residential Natural Gas hub at the left of the first, notice that it receives gas from the city gate through the distribution system at the bottom left. Such receipts are highly temporal. Notice also that distributed generation equipment sized to the residential establishment, e.g., residential microturbines and residential fuel cells, draw gas from the residential market. These are technologies that exist today and may decrease in cost markedly. They are also technologies that are difficult to insert into existing establishments, but not in new establishments such as planned communities or multifamily developments. Thus they must be inserted into the retail market for a number of reasons. First, they are "capacity machines" rather than "energy machines" and can stop and start on a dime. (They would be termed "black start" machines.) As distributed generation sources, they have the ability to inject electricity into the residential establishment on a moment's notice. In the event that electricity distribution charges were to become high because of subsidies for renewables or other programs, these technologies can come quickly into the mix and revolutionize establishment consumption. If that were to happen, they would reduce the load on the electricity distribution process at the lower right. They would certainly change not only the aggregate electricity consumption but also its fundamental shape. Black start units change load shapes dramatically.

Moving over to the Residential Electricity hub, let's ponder the inputs

- Residential Microturbine
- Residential Fuel Cell
- Residential Electricity from the Grid
- Residential Photovoltaic
- Withdrawals from Residential Power Storage

You can get power from the grid, or you can get it from distributed generation, or you can get it from residential solar photovoltaic. And where might this power go. The outputs from the Residential Electricity node are equally important. Power can go to

- Electric A/C
- Electric Heat Pump
- Electric Triple Integrated Appliance
- Resistance Heating
- Electric Water Heater
- Electric Lighting and Power
- Electric Car Charging
- Electric Clothes Dryer
- Injections to Residential Electricity Storage
- Upload to the Electric Grid

The latter two are crucially important. The new Tesla Powerwall system is designed to inject and withdraw from the residential electric market, and that is exactly what it will do. That, and technologies like it, are represented in the yellow residential storage node, as they must be. Substantial penetration of that technology stands to level the load at the residential electricity hub and reduce the capacity required to distribute energy to residential establishments. The implications for the granted (or natural) monopoly represented by the distribution system, stranded costs, and even the very ability to load the cost of social programs onto the power distribution system are immense. Higher and higher electric distribution costs just encourage distributed generation/solar PV and load-shifting, all of which will be captured by this IEMM model structure.

The sophistication and accuracy of this model are evident, and IEMM must represent residential end-use in this fashion. Otherwise, it will miss load, tariff, and other considerations. The electricity company of the future is not a "dispatch" company. The regional electricity model of the future is not a "dispatch" model. Because it is not a "dispatch" model, it can never be a global social welfare maximization model. What is there to maximize when we are considering a disparate model that runs from fuels at the upstream, through environment and renewables, through generation, through distribution, and all the way downstream through end-use conversion to fundamental energy use.

“Dispatch” is the absolute height of naiveté, and IEMM cannot afford to be naïve. The IEMM model of the future is an integrated model of fuels, generation, end-use, end-use conversion, etc.as designed here.

One needs to build similar “establishment” models for the commercial sector, the industrial sector (perhaps broken down into subsectors), the transportation sector, and electricity only/electromechanical drive sectors if required. We will show EIA precisely how to do so and precisely how to gather the requisite data. Once EIA does so, you will have profound, endogenous models of electricity load and energy on a go forward basis.

14 GLOBAL SOCIAL WELFARE MAXIMIZATION (MPEC, COMPLEMENTARITY) CANNOT POSSIBLY WORK

First order necessary conditions, which are supposed to be isomorphic to decentralized profit maximization, could never work with electricity. We have $1000 tax free for the person who can integrate the equations herein and create a global social welfare function that he or she can differentiate and get our equations back. In the sense of our GHySMo CDR, you would have to INTEGRATE all the equations and relationships put forth in this design so that you would have a function whose derivative you could take to put into Kuhn-Tucker (complementarity) conditions. That is simply not possible. Period. There is no global social welfare function that we know of that can yield the design herein. If someone can demonstrate it, dinner is on us.

Even if there were, the size of it would be prohibitive, and the embedded Jacobean would be so large as to be untenable.

There is no important dimension of an electric system that the proposed IEMM design misses as far as we know. And it includes the full and complete range of design considerations.

15 APPENDIX: ASSEMBLING GENERATION DATA—RAW DATA IDENTIFICATION, GATHERING, AND ASSEMBLY

When building a model of a country, region, or continental electricity system, the type and origin of data is an important consideration. This section briefly addresses how and where to gather and obtain such data. We will use data sets available in the United States as an analogy to help builders and users of IEMM become efficient and confident.

At the heart of an IEMM model, regardless of what country or region one is considering, is the representation of and the data characterizing the thermal generation system. To wit, one must accumulate requisite generation data for each and every electric generation unit within the geographic scope of IEMM. In the case of North America (as defined by NERC), the IEMM generation representation includes all electric generation units in NERC. As far as we know, there exists no single available data source that contains the full vector of electric generating unit data for every generator in North America/NERC, so IEMM will have to appeal to multiple data sources. Several sources are required to assemble a database containing all the electric generation units and containing the vector of data that characterize each generator. What we term the vector of plant data much include generator capacity, location (country, state/province, city, and county level), generation technology/prime mover, operating characteristics, fixed and variable costs, vintage, and fuels.

To build the North American model, the following data sources can be used for both thermal and renewable generation

1. **Generation Capacity in Place**. EIA Form EIA-860, "Annual Electric Generator Report." This source is available from the EIA website at web page

   http://www. eia. doe. gov/cneaf/electricity/page/eia860.html.

   This data source provides all the electric generating units for the lower 48 states. This data source is updated annually, but it lags behind the present by some 1-3 years. That means that one has to appeal to different sources to delete and insert retirements/decommissionings and capacity additions between the time of publication and the time of model initiation. The EIA data set does not contain the complete vector of parameters for every plant, and those must be assembled from sources elsewhere.

   Some countries in the world have similarly detailed data, and some do not (yet). Visits to in-country electric companies in the native language is likely to secure plant information that is needed. And we understand Ventyx delivers generation data for a number of countries.

   Electric generation nodes represent categories of electric generation supply and must be characterized by the prime mover (or generation technology, e.g., hydro turbine, steam turbine, combustion turbine) and the primary fuel (e.g., hydro, coal, natural gas). It is useful to develop and use a lexicon, a set of abbreviations, for the various prime mover and fuel combinations in the plant fleet.

For the various electric generation nodes, the vector of plant data includes the following central components

a. **Installed Capacity (Sometimes Termed "Plant Rating")**: IEMM must assemble the installed capacity of each thermal generator by type expressed in MW. In IEMM, we have to derate each installed thermal plant during high-temperature summer operation or in high altitude operation, and we have to increase the rating of the plant in low-temperature winter operation or low altitude operation. The design rating and the seasonal rating variation are needed. We should emphasize that the design rating is generally easily available from various private vendors and from government agencies who oversee the indigenous electric company. People who go to the government in the native language should expect to get (or infer) installed MW. Seasonal temperature ratings are easy to get by analogy from United States power plants. Altitude ratings are easy to get by analogy to United States power plants. Temperature and altitude variation and based on science and thermodynamics and are easily estimatable

b. **Nonfuel Operative and Maintenance (O&M):** IEMM must assemble nonfuel operating and maintenance cost of installed generation capacity expressed in $/MWh. O&M can vary by load period (base versus peak), i.e., different O&M cost at time of base versus time of peak. For the O&M cost, the Federal Energy Regulatory Commission's Form 1 report can be used, which can be accessed at http://www. ferc. gov/docs-filing/forms/form-1/viewer-instruct. asp. This report provides a detailed categorization of all costs incurred on a yearly basis on a plant by plant, unit by unit basis. These costs should extrapolate worldwide.

c. **Heat Rate (Btu fuel/KWh output):** Heat rate measures the "thermal efficiency" of the plant in Btus of fuel per kilowatt-hour of generation (Btu/KWh). The heat rate is the reciprocal of the classical thermodynamic thermal efficiency of the unit. Heat rate in IEMM is allowed to vary by load period allowing for the possibility of a different heat rate for an electric generation node at time of base versus time of peak, when it is running at max design output. If there were duct firing at time of peak versus not at time of base, we would expect to see a heat rate decrement at time of peak.

   The Energy Information Administration (EIA) Form EIA-923 in http://www. eia. gov/electricity/data/eia923/ allows us to infer plant heat rates. It provides on a monthly basis the amount of fuel consumed and the amount of energy produced. Each type of fuel consumed is converted into a Btu equivalent using standard conversion factors. The amount of electrical energy produced requires no conversion to another unit, merely a scaling factor to convert from MWh to KWh. Heat rates for various fuels and prime movers in the United States are an accurate guide to same or similar plants worldwide.

d. **Availability (Scheduled and Unscheduled Outages):** Availability is the percent of time the generation category is available (percent). Availability varies by time point and load period allowing for a different percent available during each time point, load period pair for an electric generation node. Availability is used as a multiplier on the electric generation node's capacity to determine an effective capacity for each time point, load period pair. Since the availability data is gathered at the prime mover and primary fuel level of detail, the availability is simply an assignment for the data source to the node. For generator availability, the NERC's Generator Availability Data Set (GADS) is used to get estimates by Prime Mover and Primary Fuel.

2. **New Plants Added and Old Plants Retired Since the Prior Reporting Period.** Because of reporting lags in capacity enumerations, it is necessary to adjust for retirements that have occurred in the intervening years and for additions that have occurred in the intervening years. In both cases, one needs the vector of plant parameters for every plant added.

3. **Canadian and Mexican Plants (within NERC).** NERC's Annual Electricity Supply and Demand Report. This data source is available from NERC for a fee. This data source provides all the electric generating units for Canada (NERC portion) and Mexico plants. This complements the EIA data set.

4. **Environmental Retrofit Cost and Efficacy**. The data elements needed to represent plant retrofit includes

   a. Capital Cost for each retrofit technique ($/MW).
   b. Additional variable cost for retrofit installation (expressed in $/MWh) when a retrofit is installed.
   c. Heat rate penalty for retrofit installation (expressed as a percentage multiplier).
   d. Capacity penalty/derating for retrofit installation (expressed as a percentage multiplier).
   e. Retrofit efficiency (percentage reduction).

   The United States EPA publishes retrofit data, which can be extrapolated worldwide.

5. **Energy Produced (MW and MWh) by Hydro Units**: Hydro is modeled differently from other generation units. For hydro generation in the United States, the United States Department of Energy's Energy Information Administration (EIA) in http://www.eia.gov/electricity/monthly/pdf/epm.pdf provides detailed monthly hydro data. For hydro generation in Canada, the commercial source Statistics Canada gives hydroelectric energy produced on a monthly basis. Historical temporal analysis can give a better idea of hydroelectric supply functions. Hydro storage has to be modeled explicitly.

   Hydroelectric generation is treated differently from the other electric generation units in that hydro does not convert a fuel into energy. Hydro simply supplies energy (MWh) into each market. Hydroelectric supply nodes require two data elements:

   a. **Supply Quantity:** for hydroelectric generation, supply quantity is the maximum amount of energy (MWh) that can be provided into the energy market. Supply quantity varies by time point and load period allowing for a different amount of energy to be provided to the market during each time point, load period pair for a supply node representing hydroelectric generation. The data source (identified above) gives monthly generation. The monthly demand load shape is used to determine the energy available for operation into a particular load period, thereby simulating the manner in which hydroelectric generation is operated to maximize profit for the owners of hydroelectric generation.

b. **Forward Cost:** the cost expressed in terms of $/MWh of supplying one MWh. Forward cost varies by time period and load period allowing for a different energy cost to be incurred during each time period-load period pair for a supply node representing hydroelectric generation. Hydro is invariably quite inframarginal, meaning one could use a low cost number such as $1/MWh and see it dispatch. It is never the marginal source of supply (except in Ecuador!), so such an assumption will not distort model results.

6. **Energy Limited Resource Bounds.** If there are energy limited resources (e.g., biomass that has only enough carbonaceous fuel to run 5000 hr/yr or hydroelectric reservoirs that are good for a maximum of 4000 hours of generation out of the reservoir), they will have to be estimated.

7. **Environmental Input-Output Coefficients (lb/MWh output or lb/MMBtu fuel).** The fundamentals of combustion and output are constant across plants everywhere in the world. We know for example what the environmental input-output coefficients for an 11,800 Btu/KWh heat rate unit without any retrofit equipment installed on it that is burning 1.1 percent sulfur bituminous coal is. For SO2, NOx, mercury, and CO2, all the requisite numbers are inferred from similar equipment in the United States or Europe or China where they are reported or known.

8. **Wind Generation Patterns.** Wind generation patterns are going to be extremely local. EIA will have to measure or infer wind generation patterns or get them out of indigenous utility data sets in the local language. If they do not exist, they will have to be inferred by analogy.

9. **Solar Insolation Patterns.** Solar insolation and solar generation patterns are going to be extremely local. EIA will have to measure or infer wind generation patterns or get them out of indigenous utility data sets in the local language. If they do not exist, they will have to be inferred by analogy.

It is going to be difficult, but far from impossible, to gather the regional IEMM data that are required. Once IEMM does, however, the quality of model and forecasts will be uncanny.