

Copyright
by
Cara Rose Touretzky
2016

The Dissertation Committee for Cara Rose Touretzky
certifies that this is the approved version of the following dissertation:

**Embedding Dynamics and Control
Considerations in Operational Optimization of
Process and Energy Systems**

Committee:

Michael Baldea, Supervisor

Thomas F. Edgar

Thomas M. Truskett

Roger T. Bonnecaze

Atila Novoselac

**Embedding Dynamics and Control
Considerations in Operational Optimization of
Process and Energy Systems**

by

Cara Rose Touretzky, B.Chem.E.

DISSERTATION

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2016

Dedicated to my grandmother Caroline Bors
and our cat Sheila.

Acknowledgments

My parents Pam and Ron, my sister Julie, and my grandmother Caroline, for making sure that Texas never felt too far from New Jersey.

Dr. Ogunnaike, for noticing and encouraging my passion for systems engineering, and for telling me I had to apply to graduate school.

All of the industry mentors and collaborators I have had the opportunity to work with through TWCCC and NGPV. Iiro Harjunoski at ABB Corporate Research for his expertise in scheduling. Rakesh Patil for the internship opportunity at NEC Labs.

Julie Cushen, for letting me live on her couch during my internship at NEC.

Dana McGuffin and Jena Ziesmer, for doing most of the work presented in Chapter 5, and for being my friends in addition to “my undergrads”.

The EPA STAR Graduate Fellowship Program, which provided funding for my last two years of study and conference travel, and a confidence boost after seven other fellowship applications.

My fellow PSE researchers Richard Pattison, Siyun Wang, Conan Park, Ray Wang, Matt Walters, Abby Ondeck, Ankur Kumar, Corey James, Forrest Estep, Akshay Sriprasad, and Vincent Heng. My time working in CPE would not have been the same without you. I hope my next office is as fun as the one I shared for

five years with Richard, Siyun, and Conan.

My adviser, Michael Baldea, who has always said that Richard, Siyun, Conan, and I were crazy for wanting to be his first students. In hindsight, you were right (as always), but it was worth it. My growth as an engineer, researcher, writer, and presenter is a direct result of your support. Thank you for your continuous reinforcement, patience, and technical guidance, despite my discontinuous ability to follow your advice.

Embedding Dynamics and Control Considerations in Operational Optimization of Process and Energy Systems

Cara Rose Touretzky, Ph.D.
The University of Texas at Austin, 2016

Supervisor: Michael Baldea

Embedding dynamics and control considerations within operational optimization decisions can result in improved performance of processes and energy systems. These efforts are motivated by modern sustainability initiatives, in particular demand response and demand management strategies for improving the efficiency of the electric grid. In these scenarios residential, commercial, and industrial electricity consumers are provided with a financial incentive to shift their demand such that the total load on the grid can be satisfied using efficient generation technologies and renewable energy sources. The financial incentive is typically a time-dependent price structure, where rates reflect the demand level and stress on the grid. Reacting to such fast-changing energy markets requires that process and energy systems be highly flexible, which is a significant departure from traditional steady state operation under fixed market conditions. In this context,

flexibility means the ability to make frequent changes to the system operation (e.g., production setpoints, constraint levels, etc.) while still maintaining stability and satisfying operating constraints at all times. This necessitates the development of advanced control and decision making strategies which are aware of system dynamics.

Accounting for dynamics by incorporating detailed, first-principles models of a system into optimization-based controllers or scheduling calculations would provide ample dynamic information. However, the resulting dynamic optimization formulations would be plagued by a large problem size, numerical difficulties associated with stiff equations and multiple time scales, and the presence of integer decisions. In this dissertation, we address these challenges through hierarchical controller designs and novel scheduling (and rescheduling) formulations including low-order models of relevant system dynamics, which are identified through an appropriate model reduction or system identification procedure. Case studies involving the built environment and chemical processes are used to demonstrate the proposed methods.

Table of Contents

Acknowledgments	v
Abstract	vii
List of Tables	xvi
List of Figures	xviii
Chapter 1. Overview	1
1.1 Motivation	1
1.2 Problem Formulations and Solution Approaches	3
1.3 Guide to Chapters	6
Part I Hierarchical Control for Processes and the Built Environment	11
Chapter 2. Nonlinear Model Predictive Control of Energy-Integrated Processes	12
2.1 Introduction	12
2.2 Process Systems with Energy Recovery	14
2.3 Dynamic Analysis and Model Reduction	19
2.4 Supervisory Nonlinear Model Predictive Control	24
2.5 Case Study: Control of a Reactor Feed-Effluent Heat Exchanger Process	33
2.5.1 Example 1	38
2.5.2 Example 2	39
2.6 Conclusions	42

Chapter 3. Nonlinear Model Reduction and Model Predictive Control of Residential Buildings with Energy Recovery	44
3.1 Introduction	44
3.2 System Description	49
3.2.1 Minimal Aggregate Model	52
3.2.2 Time-scale Analysis	53
3.2.3 Energy Management and Control System Design	59
3.2.4 Model reduction	61
3.3 Case Study	62
3.3.1 Detailed System Model	62
3.3.2 Granular Aggregate Model	68
3.3.3 Analysis and Model Reduction	70
3.3.4 Model Stiffness Analysis	73
3.3.5 Energy Management	74
3.3.6 Results and Discussion	76
3.4 Conclusions	80
Chapter 4. Hierarchical Scheduling and Control Strategies for Thermal Energy Storage Systems	84
4.1 Introduction	84
4.2 Background: Optimal Control Strategies for Buildings	87
4.3 Proposed Method: Optimization Framework for Energy Management	93
4.3.1 Centralized Optimization Formulation	94
4.3.2 Hierarchical Energy Management Strategy	97
4.3.2.1 Slow Layer: TES Scheduling System	100
4.3.2.2 Fast Layer: Air-Side Controller	103
4.3.2.3 Interaction and Information Exchange Between Control Tiers	106
4.4 Solution Methods	108
4.5 Overview of Case Studies	110
4.6 Case Study: Chilled Water TES Scheduling	111
4.6.1 TES Model	112
4.6.2 Building Model	113

4.6.3	Low Level Controllers	114
4.6.4	Hierarchical Control Implementation	116
4.6.5	Model Inputs	117
4.6.6	Results and Discussion	118
4.7	Case Study: PCM TES Scheduling with Detailed Dynamic Model	123
4.7.1	System Description	125
4.7.2	Equation-Oriented System Modeling	128
4.7.2.1	Coolant Loop Model	129
4.7.2.2	Modeling of PCM-based Active Thermal Energy Storage	131
4.7.2.3	The Coolant Domain	132
4.7.2.4	The PCM domain	132
4.7.2.5	Building Load and Cost Model	135
4.7.3	Controller Design	136
4.7.4	Results and discussion	138
4.7.4.1	Discussion: Optimal Scheduling vs. Heuristic Operation	139
4.7.4.2	Assumptions and Potential Design Limitations	141
4.7.4.3	Effect of Forecasting Error	143
4.8	Case Study: PCM TES Scheduling with a Low Order Model	146
4.8.1	System Description	148
4.8.1.1	Building	148
4.8.1.2	PCM Storage	151
4.8.1.3	Constraints	152
4.8.2	Analysis of Operation Scenarios	153
4.8.3	Optimal Operation Problem	157
4.8.4	Results	160
4.9	Conclusion	163
Chapter 5. The Effect of Distributed Electricity Generation Using Natural Gas on the Electric and Natural Gas Grid		166
5.1	Introduction	166
5.2	Background	167

5.2.1	Overview of the Electric and Natural Gas Grids	167
5.2.2	Distributed Generation Technology	170
5.2.3	Points of Connectivity	172
5.3	System Model	175
5.3.1	Model formulation	176
5.3.2	Optimization Problem Formulation	181
5.3.3	Model Parameters	187
5.4	Results	192
5.4.1	Reporting Approach	192
5.4.2	Unconstrained Case Results	195
5.4.3	Constrained Case Results	196
5.4.4	Decentralized Case Results	197
5.4.5	Cold Season Cases	198
5.4.6	Electric Heating Cases	201
5.5	Conclusion	202

Part II Scheduling with Dynamic Process Models 205

Chapter 6.	Scheduling with Data-Driven Process Models	206
6.1	Introduction	206
6.2	Problem Formulation	209
6.2.1	Scheduling	210
6.2.2	Interaction of Scheduling and Process Dynamics	212
6.3	Integrated Scheduling and Control of Continuous Processes	213
6.3.1	Scale-Bridging Models	214
6.4	Handling Discrete-Time Dynamic Process Models	216
6.4.1	Discrete-time Scale Bridging Models	216
6.4.2	Case Study	220
6.5	Conclusions	222

Chapter 7. Optimal Process Operations in Fast-Changing Electricity Markets: Framework for Scheduling with Low-Order Dynamic Models and an Air Separation Application	223
7.1 Introduction	223
7.2 Motivating Example: Air Separation Units	226
7.3 Scheduling under Dynamic Constraints	230
7.3.1 Challenges of Scheduling under Dynamic Constraints	240
7.3.1.1 Problem Size	240
7.3.1.2 Choice of PCs and QCs	243
7.4 Scheduling with rPCs and rQCs	245
7.4.1 Step 1: Selection of Variables for Scheduling-Oriented Low-Order Dynamic Models	247
7.4.2 Step 2: Identification of Scheduling-Oriented Low-Order Dynamic Models	251
7.4.3 Step 3: Optimal Scheduling Under Low-Order Dynamic Constraints	252
7.5 Case Study	256
7.5.1 ASU Model	256
7.5.2 Scheduling-Oriented Low-Order Dynamic Modeling of ASU Dynamics	266
7.5.3 Building ASU Low-Order Dynamic Models	272
7.5.3.1 Data Collection	273
7.5.3.2 Low-Order Dynamic Model Identification	276
7.5.4 ASU Scheduling Under Low-Order Dynamic Constraints	279
7.6 Results and Discussion	281
7.6.1 Optimal Sequence of Production Rate and Inventory Level Setpoints for Problems P1, P2, and P3	284
7.6.2 Implementation on Full-Order Model	287
7.6.3 Comparison of P3 and P4	289
7.7 Conclusion	292

Chapter 8. Closing the Scheduling Loop: Rescheduling Formulations Incorporating Dynamic Process Models, Fault Detection-Based Triggers, and Moving Horizon Concepts	294
8.1 Introduction	294
8.2 Rescheduling in Response to Dynamic Disturbances	298
8.2.1 Motivation	298
8.2.2 Framework	301
8.2.2.1 Process Monitoring Provides Rescheduling Triggers	303
8.2.2.2 Formulation	305
8.2.2.3 Opportunities for Model Based Process Monitoring	306
8.2.3 Summary	307
8.2.4 Case Study	309
8.2.4.1 Scheduling Problem Formulation	310
8.2.4.2 System Model	315
8.2.4.3 Results and Discussion	319
8.3 Moving Horizon Scheduling with Low-Order Dynamic Models	325
8.3.1 Comparison to Economic MPC	326
8.3.2 Case Study: Moving Horizon ASU Scheduling	328
8.3.2.1 Results and Discussion	330
8.4 Conclusion	333
Chapter 9. Conclusion and Future Directions	334
9.1 Summary of Contributions	334
9.2 Future Research Directions	337
Dissemination of Research	340
Journal Publications	340
Conference Publications	341
Patents	342
Presentations	342

Bibliography	345
Vita	373

List of Tables

2.1	Process parameters (adapted from [18])	36
3.1	Model naming conventions	63
3.2	HVAC Properties	65
3.3	Wall Dimensions	65
3.4	Wall Material Properties	66
3.5	Window Properties	66
3.6	Estimated parameters used in the granular model	66
3.7	Nomenclature	67
4.1	TES Case Study Summary	110
4.2	Operating Modes	112
4.3	Parameter values	119
4.4	Objective function value at $t = 24$ hours (as $J/10^7$)	119
4.5	Operating Modes	127
4.6	Flow conditions. F_2 is calculated using the option indicated. The charge only mode (m_2) is not used in the current operation of the coolant loop.	131
4.7	Total cost (scaled by $1e7$) and the reduction of the baseline cost through the use of optimal control and heuristics	141
4.8	Case Study Parameters	146
4.9	Variable Nomenclature	147
4.10	Variables	151
4.11	Objective Values and Solution Times	163
4.12	Parameters	164
5.1	Model Parameters and Variables	177
5.2	Scenario Definitions	193
5.3	Summary of Case Results	194

7.1	Nomenclature	234
7.2	Variables described by the scheduling-oriented low-order dynamic models	273
7.3	Identified model details	279
7.4	Summary of problem formulations.	283
7.5	Optimal solutions of problems P1-P4. The cost values are for the 3 day horizon.	289
8.1	Rescheduling Strategies	297
8.2	General Scheduling with Dynamic Model Problem Statement	302
8.3	Nomenclature - Indices and Sets	310
8.4	Nomenclature - Parameters and Variables	311

List of Figures

2.1	Prototype energy-integrated system [127, 18]	14
2.2	Reactor-FEHE process with fired furnace and exchanger bypass stream.	34
2.3	Evolution of the reactor temperature and furnace heat duty using a tracking objective	39
2.4	Evolution of the reactor temperature and furnace heat duty using an economic objective	40
2.5	Open-loop response of the reactor temperature for an increase in both feed flow rate and inlet temperature	40
2.6	Evolution of the reactor temperature and furnace heat duty using an economic objective	41
2.7	Evolution of the reactor temperature and furnace heat duty using a tracking objective	42
2.8	Evolution of the reactor temperature and furnace heat duty using an economic objective	42
3.1	Prototype single-zone building HVAC model and control system . .	49
3.2	University of Texas Thermal Façade Laboratory (photo courtesy of the UT School of Architecture).	50
3.3	Interactions between the building zone and environment. The temperature of the air in the zone is influenced by conductive heat transfer through the window and walls (only one wall is shown) and convective heat transfer at these surfaces. The structural element temperatures are influenced by solar radiation (I), long wave radiation, and the outdoor air temperature.	51
3.4	Minimal aggregate model. Dashed lines represent mass flow while solid lines represent heat transfer. Each block labeled T_i corresponds to a state in the model, and the heat streams of enthalpy H_i show the connections between the lumped building elements. Q_{HVAC} represents the heat removed from the zone feed air stream by the AHU.	52
3.5	Granular aggregate model. Dashed lines represent mass flow while solid lines represent heat transfer. The enthalpy labels are not shown but are identical to those in the minimal aggregate model.	69

3.6	Model eigenvalues for 404-dimensional full-order model, the 15-dimensional granular aggregate model, and the 12-dimensional reduced-order aggregate model	75
3.7	Weather Profile for August 1	78
3.8	Energy Cost	78
3.9	Occupancy Disturbances	79
3.10	Zone Temperature Profiles	82
3.11	AHU Power for Occupant Profile 1	83
3.12	Plant-Model Mismatch Using the Economic Objective for Occupant Profile 1	83
4.1	Energy Reliability Council of Texas (ERCOT) demand and day ahead settlement point prices for June 25, 2012 (Source: www.ercot.com)	85
4.2	Centralized EMPC	94
4.3	Hierarchical energy management scheme	100
4.4	Cyclical mode transition behavior.	102
4.5	The slow layer of the energy management scheme using approximates of the load instead of a detailed building model.	107
4.6	Coolant loop in the TFL HVAC system. Also shown are two proportional controllers that manipulate F_2 and the chiller cooling load.	111
4.7	Model of the building zone	114
4.8	Disturbance profiles. The ‘actual’ values are provided to the system being controlled (i.e., the plant), while the ‘forecast’ profiles are used by the model in the optimization calculations.	118
4.9	Cost profile	120
4.10	From top to bottom: Zone temperature and setpoint in the optimization model and plant, operating mode, TES charge state, and chiller load for the surplus storage case.	122
4.11	From top to bottom: Zone temperature and setpoint in the optimization model and plant, operating mode, TES charge state, and chiller load for the insufficient storage case.	123
4.12	From top to bottom: Zone temperature and setpoint in the optimization model and plant, operating mode, TES charge state, and chiller load for the insufficient storage case when using the ‘charge and cool’ mode instead of ‘charge only’.	124
4.13	Flexible configuration of the coolant circuit in the University of Texas Thermal Facade Lab HVAC system and additional PI controllers (control loops are indicated by red dotted lines). Small dashed lines show heat transfer in/out of the system	126

4.14	Configuration of the PCM TES tank.	127
4.15	Building thermal load profile.	136
4.16	Real-time energy price profiles for a 24-hour time period. In the first case the energy cost rises from the baseline by a factor of ten from hours 14 to 16. The second case has a broader, four-hour peak price schedule, where the cost increases during the hours 12 to 16	137
4.17	Comparison of optimal results and heuristic behavior for the two load profiles subject to the 2 and 4 hour peak price schedule (peak from hours 14-16 or 12-16). The chiller power, TES charge level, and active mode are shown for each optimal and heuristic controller. In the mode subplot, blue represents charging (m_3) and red discharging (m_4). All other periods are the normal operating mode (m_1).	142
4.18	The effect of error in the Q_{build} forecast	145
4.19	Coolant loops in the building system including the active PCM TES. The charging and discharging configurations are shown.	149
4.20	Types of load profiles, shown for the hours where $Load$ is nonzero.	155
4.21	Comparison of electricity price profiles (cost in EU cents/kWh)	156
4.22	Results of the MINLP for the Case 3 load using the flat cost schedule. The dashed line shows the P_{hp}^{min} limit.	161
4.23	Results of the MINLP for the Case 3 load using the peak cost schedule. The dashed line shows the P_{hp}^{min} limit.	162
4.24	Results of the NLP for the Case 3 load using the peak cost schedule. The dashed line shows the P_{hp}^{min} limit, which has been relaxed.	163
5.1	Points of connectivity between the electric and natural gas grids	173
5.2	Structure of residential neighborhood grids with distributed generation units	176
5.3	Varying Weighting Factors	187
5.4	Hot season demand profiles for 100 individual consumers.	189
5.5	Total Daily Appliance and Heating Demands	190
5.6	Performance of various consumer-to-generator ratios for the $L_i^{max} = 10kW$ case	192
5.7	Case 1 Results (unconstrained, gas heat, hot season, 10kW generators)	195
5.8	Case 2 Results (unconstrained, gas heat, hot season, 50kW generators)	195
5.9	Case 3 Results (constrained, gas heat, hot season, 10kW generators)	197
5.10	Case 4 Results (decentralized unconstrained optimization, gas heat, hot season, 10kW generators)	199

5.11	Case 5 Results (decentralized unconstrained optimization, gas heat, hot season, 50kW generators)	199
5.12	Case 6 Results (unconstrained, gas heat, cold season, 10kW generators)	201
5.13	Case 7 Results (unconstrained, gas heat, cold season, 50kW generators)	201
5.14	Case 8 Results (unconstrained, electric heat, hot season, 50kW generators)	202
5.15	Case 9 Results (unconstrained, electric heat, cold season, 50kW generators)	203
6.1	Decision-making hierarchy in process control and operations.	206
6.2	Slot-based continuous-time schedule and transitions between products [96, 22].	212
6.3	Scale-bridging models for integrated scheduling and control [77, 25].	214
6.4	A time slot in the DSBM - based integrated scheduling and control formulation.	218
6.5	Optimal setpoint schedule and transition dynamics as determined by $P1$ and $P2$.	222
7.1	Flowsheet for the cryogenic air separation unit for the production of N_2 [211, 44].	229
7.2	Hierarchy of decisions for process operation.	231
7.3	Left: a hypothetical process is allowed to settle to steady state prior to changing the setpoint resulting in a dynamically feasible transition. Right: the process does not reach steady state prior to the setpoint change, potentially leading to temporary constraint violations during the transition.	233
7.4	Comparison of integrated scheduling and control modeling choices	244
7.5	Optimal steady state values of the manipulated variables as a function of production setpoint level	263
7.6	Manipulated variable trajectories and production rate during setpoint changes	265
7.7	Data from the ERCOT day ahead market[88] for electricity market prices on September 9-11, 2013.	275

7.8	Simulated “historical” process transition data and corresponding predictions from the scheduling-oriented low-order dynamic model. Several constraints are violated throughout the production horizon by following the schedule determined using the static process model. These include the impurity level, P_{zone1}/P_{dew} and the condenser/reboiler minimum temperature driving force, and the reboiler holdup endpoint constraint.	280
7.9	Optimal production schedule and inventory holdup prediction from P1, P2, and P3.	285
7.10	Evolution of impurity levels predicted by solving P2 and P3. Notice that a ‘back-off’ upper bound constraint of 1800ppm is imposed. . .	286
7.11	A comparison of the low-order model predictions (labeled P3 low-order) and the trajectories of the corresponding variables in the full-order process model during implementation of the optimal schedule determined using P3 (labeled P3 full-order)	288
7.12	Comparison of rQCs when optimal schedules produced by P2 and P3 are simulated on the detailed model	290
7.13	rPCs comparison when optimal schedules produced by P2 and P3 are simulated on the detailed model	291
7.14	Results comparison between optimal schedules determined using the low-order process model (P3) and the detailed process model	292
8.1	Process management decisions. Scheduling with a dynamic process model is an alternate method for obtaining a production schedule. .	295
8.2	Closing the scheduling loop in response to dynamic disturbances using scheduling approaches with dynamic models and process monitoring.	299
8.3	Opportunities for providing additional information to model based fault detection schemes	307
8.4	State equipment network for the batch flowshop process.	309
8.5	Variations in batch operations for making the same product	316
8.6	Gantt charts for optimal production schedules for the base case and two faults subject to Requirement 1. Arrows indicate when the rescheduling calculation is triggered and performed.	322
8.7	Variable trajectories under Requirement 1.	323
8.8	Profit and production levels under Requirement 1.	323
8.9	Variable trajectories under Requirement 2.	324
8.10	Profit and production levels under Requirement 2.	324

8.11	Day-ahead market prices for July 10 - July 15 2013 from the Electricity Reliability Council of Texas (http://www.ercot.com/mktinfo/)	330
8.12	Production rate setpoint. The nominal flowrate (demand) is 20mol/s and the production rate may vary by up to 20%	330
8.13	Evolution of inventory level. The storage capacity is limited to 200 kmol.	331
8.14	Evolution of impurity levels in the product. The maximum allowed level is 2000ppm	332
8.15	Evolution of reboiler/condenser temperature driving force. The temperature difference must be greater than 1.85K.	332

Chapter 1

Overview

1.1 Motivation

The complex infrastructure for delivering energy (e.g., electricity and natural gas) to residential, commercial, and industrial consumers must be both reliable and efficient. The latter is particularly important to modern sustainability initiatives centered on energy conservation and reducing associated environmental impacts. Consider that in the United States in 2014, approximately 4000 billion kWhr of electricity was produced and consumed. Nearly three quarters of this amount is used by residential and commercial customers (i.e., buildings) while the remainder is attributed to industrial consumers. In order to produce this large amount of electricity, a variety of fuel sources are used. A quarter of US electricity production used natural gas as the primary fuel, which accounts for 30% of the total US natural gas consumption. The remainder of NG consumption is attributed to residential/commercial customers (for heating, cooking, etc.) and industrial customers (for plant fuel, CHP, etc.).

It is desirable to reduce the total yearly consumption of primary and secondary energy sources. Design modifications at all points in the energy supply chain are an obvious source of energy savings, especially considering that old and

inefficient technology contribute to the high consumption levels mentioned above. For example, the widespread use of energy efficient LED lights and appliances, the construction of integrated power plants (e.g., cogeneration and trigeneration facilities), and incorporation of renewable energy sources (e.g., wind and solar) into the power generation portfolio, are design modifications that ultimately reduce the losses associated with electricity generation and consumption.

There is another opportunity for improving the efficiency of the electric (and natural gas) grid, but to see it we must consider the *dynamics* of the system. The supply in this case follows the demand of energy, and neither are constant; rather, there are seasonal, daily, hourly, and even minute-wise fluctuations in energy consumption. Furthermore, the overall efficiency of the energy supply chain varies greatly with these changes in demand. The operators of generation facilities are well aware of this behavior, which is reflected in the spot market prices for electricity (prices may increase up to a hundred times when the grid is stressed). This is due to the use of less efficient spare generation capacity (often referred to as “peaking plants”) to meet this temporary demand rise, often incurring additional economic and environmental costs. Also, renewable generation technology, which is weather dependent, has dynamic characteristics which are not naturally aligned with consumer demand patterns (e.g., solar electricity production may peak earlier in the day than demand, and wind electricity is produced in large quantities at night when demand is low).

Demand response and demand management strategies aim to alter the electricity demand pattern of consumers such that the resulting load on the grid can

be satisfied efficiently. Typically, users are given a financial incentive to shift their demand to off-peak times via tiered energy price structures, which entail charging higher rates during the peak period. In order for this to be a ‘win-win’ scenario, where the stress on the electric grid is reduced and consumers lower their energy costs, advanced control strategies are needed to manage consumer energy use. This is not a trivial matter, since in order to react to fast-changing energy markets and hourly price changes, process and energy systems must be flexible and agile while maintaining stability. This is a significant departure from steady state operating regimes, and mandates that all decision-making should be aware of the system dynamics. In this dissertation, we develop dynamic optimization-based control and scheduling strategies which are capable of orchestrating the operation of processes and energy systems in the residential, commercial, and industrial sectors (specifically, residential/commercial buildings and grid-dependent chemical processes).

1.2 Problem Formulations and Solution Approaches

Our optimization-based control and scheduling strategies are inspired by Model Predictive Control (MPC), a technique originally developed by the chemical industry. MPC consists of solving an optimization problem (typically with an objective of setpoint tracking) repeatedly over a finite moving horizon, using a model of the process dynamics (e.g., first order transfer functions or state space models which relate the process inputs and outputs) to anticipate future behavior. With this formulation it is possible to account for system operating constraints and forecasts of disturbances. The result of the optimization is a sequence of control

moves which minimize the objective and satisfy the constraints. The first control move in the sequence is implemented, and then the problem is solved again for the shifted prediction horizon.

Since its introduction, MPC has been extended to include nonlinear process models (which are presumably more accurate representations of the system behavior) and economic objective functions, which aim to improve the controller performance and overall process economics. Such formulations are of interest for managing the process and energy systems discussed above because of their ability to account for complex dynamics and energy prices. From an implementation point of view, the limiting step in applying these formulations is the solution time. The optimization must converge within the control interval. This is difficult to guarantee when the optimization formulation involves a nonlinear objective, model, constraints, or some combination thereof. While desirable for increased accuracy, detailed nonlinear models render the optimization difficult to solve because of:

- Large number of model states (increases problem size)
- Stiff equations (cause numerical difficulties)
- Dynamic variables must be discretized in order to use a simultaneous solution approach (further increases the problem size)
- Multiple time scales (requires a long prediction horizon with short time steps for the control interval, and high execution frequency)
- Possible presence of integer decisions (results in a mixed integer optimization)

Note that when integer decisions are present the resulting formulation may have a natural connection to process scheduling, since the integer decisions and economic objective may relate to scheduling in addition to control concerns. Historically, production scheduling and process control have been carried out by separate entities of a company, and scheduling decisions are made using limited information on the processes dynamic behavior. Coordination of interactions between scheduling and control is desirable to improve overall process economics and ensure dynamic feasibility, which is an important consideration when process operating conditions will be changes frequently in response to fluctuating energy prices or other economic considerations.

In this dissertation, we address these challenges through:

- **Meaningful model reduction.** We identify computationally tractable dynamic model forms for the application at hand using a variety of methods, ranging from perturbation analysis for first-principles models, to data-driven empirical modeling.
- **Hierarchical controller designs.** We demonstrate how a thorough understanding of the system dynamics and operating space can be leveraged to design hierarchical control systems for energy management.
- **Novel scheduling formulations.** We incorporate low-order dynamic models into scheduling calculations, which is a significant departure from tradition scheduling approaches which use tabulated process data to approximate processing times.

Ultimately, the control and scheduling formulations take the form of a (mixed-integer nonlinear) dynamic optimization (DO), in which there are (partial and/or ordinary) differential equations in the set of equality constraints. DO solutions methods are based on either discretizing all differential equations so that a simultaneous solution algorithm can be applied, or using a sequential solution algorithm which integrates the model equations and associated sensitivities in time. In this dissertation, we demonstrate both approaches; the former using collocation of finite elements and the GAMS programming language along with open source and commercial optimization solvers, and the latter using either the gPROMS software package or control vector parametrization techniques built in Matlab.

1.3 Guide to Chapters

This dissertation is divided in two parts:

Part I contains four chapters related to hierarchical control of process and buildings. The first two chapters are related to the control of energy-integrated systems. Chapter 2 presents a theoretical foundation for analyzing a process with significant energy recovery/recycle, and performing model reduction using singular-perturbation arguments. A supervisory nonlinear MPC layer is designed and demonstrated in a case study involving a reactor feed-effluent heat exchanger process. In Chapter 3 we apply these principles to a building HVAC system which has an energy recovery device and substantial thermal mass. We show that the dynamics evolve over three distinct time scales, and derive reduced order models for each which are physically meaningful and tied to the building design data.

Economic and tracking nonlinear MPC objectives are demonstrated in a closed-loop simulation, subject to disturbances, and the calculations are easily performed in real-time using the reduced model. The ability to shift demand is seen under the economic objective, which suggests that this model and associated control structure show potential for rapid and cost-effective deployment as a model-based energy management approach for single-zone buildings

In chapter 4, we present a hierarchical energy management strategy for buildings with thermal energy storage. The distinct dynamics of the storage system and load requirements are amenable to a composite fast-slow controller design, where the slow controller takes the form of economic MPC. To alleviate challenges associated with integer decisions, we take advantage of the cyclical nature of charging/discharging modes to create a NLP formulation of the problem. There are three case studies in Chapter 4 which demonstrate different types of thermal storage (chilled water vs. latent energy storage using phase change material) and modeling approaches (detailed vs. lumped models). We show that our approach can achieve significant cost savings for consumers subject to time-of-use pricing, outperform (reasonable) operating heuristics even under uncertainty in forecasting building loads, and have low solution times, showing real incentive for online implementation in a recursive, moving horizon fashion.

In chapter 5, we turn our attention to distributed generation. In order to evaluate the consequence of tighter interdependency between the electricity and natural gas grids at intersection points near consumers due to widespread adoption of micro-scale generation, we develop a model of a small neighborhood with

several micro-turbines and establish the optimal operation of all units to achieve load-leveling on the electric grid. Several cases are presented which establish the effect of limitations on the current infrastructure capacity. We consider the difference between centralized and decentralized operation of the ensemble of distributed generation units. Our results suggest that centralized decision-making and coordination of the operation of the generators results in optimal demand profiles, and it is difficult to replicate this behavior with a decentralized decision-making scheme whereby the operation of each generator is managed separately. This observation is also expected to hold for the widespread use of strategies considered in chapters 3 and 4, which only consider the behavior of a single customer.

Part II contains 3 chapters on scheduling with dynamic models. In chapter 6 we summarize the foundations of integrated scheduling, control, and dynamics using first-principles models, and present our initial work on scheduling using data-driven and discrete models. We propose a scheduling formulation using a hybrid representation of time, whereby the dynamic process model is presented in discrete time and the scheduling problem uses a continuous time formulation. We introduced a novel ‘reverse integrated error’ concept to determine transition times between products in a continuous process.

In chapter 7 an industrial scale air separation case study is used to demonstrate the importance of accounting for process dynamics in a flexible manufacturing scenario. We develop a scheduling formulation that includes dynamic information on product quality, production rate, and a subset of variables relevant to process operating constraints. The novelty of our contribution consists of us-

ing scheduling-oriented low-order dynamic models that predict the closed-loop dynamics of relevant constrained process variables in response to scheduled operating point changes. Central to this effort, we introduce a new methodology for selecting the variables relevant to the scheduling calculation, and define the scope (i.e., model inputs and outputs) of the scheduling-oriented dynamic models. We highlight importance of accounting for process constraints, historically only of interest to the control domain, at the scheduling level. When applied to the air separation case study this framework is able to achieve significant cost savings in the time-of-use pricing environment while ensuring that the plant operation remains feasible.

In chapter 8 we turn our attention to rescheduling using dynamic process models. We emphasize differences between disturbances to the scheduling and process layer, and highlight how the latter can lead to changes in processing times or recipes, transition times, and the overall product wheel. We propose a novel event-driven rescheduling framework which relies on existing methods for process monitoring and fault detection, identification, and reconstruction, which diagnose specific changes in processing conditions. The integrated scheduling and dynamics framework is well suited for accounting for these differences; the new operating paradigm can be accounted for by altering constraint boundaries for the dynamic model in a rescheduling calculation. We also discuss periodic rescheduling strategies and the use of moving horizon scheduling, and compare their formulation to economic MPC strategies. A case study is presented showing how online state estimation can be used to update the dynamic model in the moving horizon scheduling

formulation.

Finally, a summary of contributions in the areas of model reduction, hierarchical controller design, and scheduling with dynamic models is given in Chapter 9, in addition to recommendations for future research directions.

Part I

Hierarchical Control for Processes and the Built Environment

Chapter 2

Nonlinear Model Predictive Control of Energy-Integrated Processes

The material in this chapter has been published in [27] and [26].

2.1 Introduction

Improving energy efficiency has become particularly important in industrial processes in view of recent increases in energy prices, growing environmental concerns, and regulatory pressure. The principle of energy integration is quite simple, and it entails identifying, within the process, sets of energy sources and corresponding energy sinks between which energy can be transferred without violating thermodynamic principles. Energy recovery has in general focused on thermal energy (heat or refrigeration) with the aim of minimizing the use of external utility streams. These ideas apply directly to other energy forms, and significant reductions in capital and operating costs have been reported as a consequence of using integrated designs in the chemical industry (see *e.g.*, [18] for an overview).

Chemical processes are typically comprised of multiple units interconnected through material streams and energy exchange in a main process flux. Intuitively, integration creates physical connections between process units that are not part

of the main flux and can thus give rise to feedback interactions which increase the complexity of the dynamic behavior of the process. These effects (including inverse response and open-loop instabilities) are by now well-documented in the literature [54, 228, 187, 184, 148] and make the control of integrated process systems especially challenging. Furthermore, in the recent past, market conditions have become increasingly dynamic, dictating that the control of integrated chemical processes go beyond simple regulation around a steady state, and towards efficient transient operation within a broader operating envelope.

In this paper, we focus on the optimal control of energy-integrated process systems. We concentrate on a particular class of systems in which energy recycling and recovery are significant in comparison to any available external energy sources and/or sinks. In our previous work [18, 127], we have shown that the simultaneous presence of energy flows of different magnitudes causes model stiffness and is at the origin of a time scale separation in the system dynamics, and derived reduced-order descriptions of the dynamics in each time scale. In the present paper, we demonstrate that the fast component of the dynamics of such systems is stable in general and asymptotically stable in practical systems. Then, we use these results to develop a hierarchical controller design, consisting of i) a linear control system to exponentially stabilize the fast dynamics and, ii) a nonlinear model predictive controller for the slow dynamics. We demonstrate that this composite approach guarantees exponential stability for the overall system.

Subsequently, we utilize the proposed control approach to formulate a strategy for optimal *energy management and reduction of energy use* at the system

level, and establish a parallel between the approach introduced in this paper and economic model predictive control [8, 114].

A case study considering a simulated Reactor-Feed Effluent Heat Exchanger process is presented to illustrate the theoretical developments.

2.2 Process Systems with Energy Recovery

In general, a process system is a system of finite dimensions, whose states can be used to define a set of additive *inventory* variables [94, 21]. In the present paper we will focus on enthalpy, θ , as a state variable representative of process energy use, and consider a process system consisting of N (lumped) units in series (Figure 2.1). The feedback connection in the figure represents the energy transfer between the last and the first units, making the process under consideration an energy-integrated one. Note that, for now, we are not concerned with the transfer mechanism or its physical implementation, we simply assume that transfer is thermodynamically feasible and occurs at a rate $Q_{in} \equiv Q_{out}$.

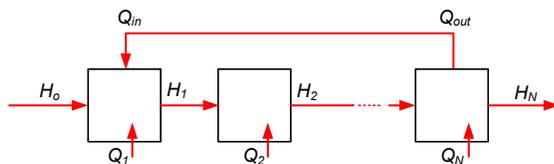


Figure 2.1: Prototype energy-integrated system [127, 18]

Then, we assume that each of the units $i = 2, \dots, N-1$ receives energy from the upstream unit $(i-1)$ at a rate H_{i-1} and, transfers energy to the downstream $(i+1)$ unit at a rate H_i via a material stream. Additionally, units 1 and N

exchange energy in the same manner with the environment (e.g., other processes) at rates H_o and H_N . Energy flows Q_i , $i = 1, \dots, N$, account for unit-level energy sources or sinks, including local heating or cooling via heat exchange or chemical reactions.

Our interest at this point is to elucidate the effect of this recycle connection on the dynamics of the unit enthalpies $\boldsymbol{\theta} = [\theta_1 \dots \theta_i \dots \theta_N]^T$, $\boldsymbol{\theta} \in \mathcal{Q} \subset \mathbb{R}^N$. To this end, under the assumptions listed above, we can write a mathematical model of the process system under consideration as [18, 127]:

$$\dot{\boldsymbol{\theta}} = \sum_{i=0,N} \gamma_i(\boldsymbol{\theta})H_i + \sum_{i=1}^N \phi_i(\boldsymbol{\theta})Q_i + \sum_{i=1}^{N-1} \gamma_i(\boldsymbol{\theta})H_i + \gamma_q^{in}(\boldsymbol{\theta})Q_{in} + \gamma_q^{out}(\boldsymbol{\theta})Q_{out} \quad (2.1)$$

and $\gamma_i(\boldsymbol{\theta})$, $\gamma_q^j(\boldsymbol{\theta})$ and $\phi_i(\boldsymbol{\theta})$ being appropriately defined vector functions.

Using the steady state values (denoted by the subscript s) of the energy flows, we can define the dimensionless variables $\omega_i = H_i/H_{i,s}$, $i = 1, \dots, N$, $\omega_{in} = Q_{in}/Q_{in,s}$ and $\omega_{out} = Q_{out}/Q_{out,s}$. Let us also define the following dimensionless quantities: $l_i = H_{i,s}/H_{1,s}$, $i = 1, \dots, N-1$, $m_{in} = Q_{in,s}/H_{1,s}$ and $m_{out} = Q_{out,s}/H_{1,s}$. We can thus rewrite (2.1) as:

$$\begin{aligned} \dot{\boldsymbol{\theta}} = & H_{1,s} \left[\gamma_q^{in}(\boldsymbol{\theta})m_{in}\omega_{in} + \gamma_q^{out}(\boldsymbol{\theta})m_{out}\omega_{out} + \sum_{i=1}^{N-1} \gamma_i(\boldsymbol{\theta})l_i\omega_i \right] \\ & + \sum_{i=0,N} \gamma_i(\boldsymbol{\theta})H_i + \sum_{i=1}^N \phi_i(\boldsymbol{\theta})Q_i \end{aligned} \quad (2.2)$$

Let us now concentrate on the dynamics of the system (2.2). First, we make the following general observation regarding the steady-state of this system: in order to achieve a high degree of energy integration, the amount of energy recovered from unit N and transferred to unit 1 should be significant in comparison to the energy input to the system either via the material input stream or through local heat

generation or local heat exchange. Furthermore, the energy recovered must be used in the process, and thus internal energy flow rates between units 1 and N should be of the same magnitude as the rate at which energy is recovered.

These observations can be formalized in the following assumptions [18, 127]:

Assumption 2.1. *The internal energy flow rates $H_{i,s}$, $i = 1 \dots N-1$ are of similar magnitude (i.e., $l_i = \frac{H_{i,s}}{H_{1,s}} = \mathcal{O}(1)$)*

Assumption 2.2. *The magnitude of the energy flow rate $Q_{in} \equiv Q_{out}$ is similar to the magnitude of the internal energy flows H_i , $i = 1 \dots N-1$, and we have:*

$$\frac{Q_{in,s}}{H_{1,s}} \equiv \frac{Q_{out,s}}{H_{1,s}} = m_{in} = m_{out} = \mathcal{O}(1) \quad (2.3)$$

Assumption 2.3. *The internal energy flows are much larger than the inlet and outlet energy flows $H_{o,s}$ and $H_{N,s}$, that is:*

$$\frac{H_{o,s}}{H_{1,s}} \ll 1 \quad \text{and} \quad \frac{H_{N,s}}{H_{1,s}} \ll 1 \quad (2.4)$$

Assumption 2.4. *The energy inputs Q_i , $i = 1 \dots N$ to the individual units are of similar magnitude to the inlet energy flow. Equivalently, $\frac{Q_{i,s}}{H_{o,s}} = \mathcal{O}(1)$ and $\frac{Q_{i,s}}{H_{1,s}} \ll 1$.*

We also make the following (non-restrictive, but convenient) assumption:

Assumption 2.5. *The energy holdups θ_i are defined (e.g., in terms of enthalpy or internal energy) such that $\theta_i \geq 0$.*

Assumptions 2.3 and 2.4 are in effect the key to tight energy integration, as they stipulate (as suggested above) that the amount of energy recovered from the process *exceed* the amount of energy input to the process from external sources. Furthermore, Assumption 2.3 suggests that a figure of merit for energy integration can be defined in the form of an energy recovery number [18] as:

$$Rc = \frac{1}{\varepsilon} = \frac{H_{1,s}}{H_{o,s}} \quad (2.5)$$

Based on the above, we can rewrite the model (2.2) as:

$$\begin{aligned} \dot{\boldsymbol{\theta}} &= \frac{1}{\varepsilon} H_{o,s} \left[\gamma_q^{out}(\boldsymbol{\theta}) m_{out} \omega_{out} + \gamma_q^{in}(\boldsymbol{\theta}) m_{in} \omega_{in} + \sum_{i=1}^{N-1} \gamma_i(\boldsymbol{\theta}) l_i \omega_i \right] \\ &+ \sum_{i=0,N} \gamma_i(\boldsymbol{\theta}) H_i + \sum_{i=1}^N \phi_i(\boldsymbol{\theta}) Q_i \end{aligned} \quad (2.6)$$

Or, in a more compact form, we can write (2.6) as:

$$\dot{\boldsymbol{\theta}} = \boldsymbol{\Gamma}^s(\boldsymbol{\theta}) \boldsymbol{\nu}^s + \frac{1}{\varepsilon} \boldsymbol{\Gamma}^l(\boldsymbol{\theta}) \boldsymbol{\nu}^l \quad (2.7)$$

where $\boldsymbol{\nu}^s \in \mathcal{U}^s \subset \mathbb{R}^{m^s}$, $\boldsymbol{\nu}^s = [H_o/H_{o,s} \ H_N/H_{N,s} \ Q_1/Q_{1,s} \ Q_N/Q_{N,s}]^T$ is a vector of scaled variables that correspond to the inlet and outlet energy flows and $\boldsymbol{\nu}^l \in \mathcal{U}^l \subset \mathbb{R}^{m^l}$ is a vector of scaled variables corresponding to the (much larger) internal and recycle energy flows, with $\boldsymbol{\nu}^l = [\omega_1 \dots \omega_{N-1} \ \omega_{in} \ \omega_{out}]^T$. $\boldsymbol{\Gamma}^s$, $\boldsymbol{\Gamma}^l$ are matrices of appropriate dimensions.

Remark 2.1. *In order to preserve the brevity of the presentation, we have not included material balance considerations in the system models utilized above. However, the proposed framework can easily accommodate the inclusion of material*

balance equations when necessary (see., e.g., [18], as well as the case study later in the paper).

Remark 2.2. *Some (but not all) of the components of the vectors $\boldsymbol{\nu}^s$ and $\boldsymbol{\nu}^l$ constitute potential manipulated inputs for the system (2.7). For instance, the flow rates of the material streams connecting units, or the rates of heat input or heat removal from each unit can, in principle, be adjusted directly. On the other hand, the rates of heat generation by chemical reaction depend on the process conditions (temperature, composition, pressure) in each unit and are thus not available as manipulated inputs for controller design. These points are further illustrated in the case study in Section 2.5.*

The representation in (2.7) provides a different perspective on process systems with energy integration. In effect, based on the energy integration figure of merit ε , we can isolate processes with tight energy integration as the class of systems for which $\varepsilon \ll 1$ (see Assumption 2.3). Clearly, such systems are interesting from the point of view of process economics, since increased reliance on energy recuperated from the process outlet reduces the use of external energy sources and diminishes operating costs.

From a mathematical point of view, if $\varepsilon \ll 1$, the generic model (2.7) is in a nonstandard singularly perturbed form [95, 176, 146]. The dynamics of such systems evolve over multiple time scales, a feature that must be accounted for in controller design (which, in principle, should be carried out separately in each time scale, an approach referred to as *composite control* [142, 57, 147]). In the following

section, we derive separate reduced-order models of the fast and slow dynamics of the process system under consideration; controller design will be addressed in the sequel.

2.3 Dynamic Analysis and Model Reduction

In order to elucidate the potential two time scale behavior of such systems, we extend the methods proposed in [127]. First, we obtain a description of the fast dynamics. To this end, we define a fast (“stretched”) time scale $\tau = t/\varepsilon$, in which the process model (2.7) becomes:

$$\frac{d\boldsymbol{\theta}}{d\tau} = \varepsilon\boldsymbol{\Gamma}^s(\boldsymbol{\theta})\boldsymbol{\nu}^s + \boldsymbol{\Gamma}^l(\boldsymbol{\theta})\boldsymbol{\nu}^l \quad (2.8)$$

We proceed using singular perturbation arguments, and consider the limit $\varepsilon \rightarrow 0$, which, from a physical perspective, corresponds to the ideal case of total heat integration via infinite energy recycling. This results in an expression for the fast dynamics, of the form:

$$\frac{d\boldsymbol{\theta}}{d\tau} = \boldsymbol{\Gamma}^l(\boldsymbol{\theta})\boldsymbol{\nu}^l \quad (2.9)$$

Recalling that inventories are additive, we can also define the total energy stored in the system as:

$$\theta_{tot} = \sum_{i=1}^N \theta_i \quad (2.10)$$

Intuitively, θ_{tot} is governed exclusively by the small external energy flows and the (small) energy generation terms, i.e., the vector $\boldsymbol{\nu}^s$ which captures energy exchange

between the system and the environment. Equivalently, the (large) internal ν^l flows do not affect the total energy stored in the system. From a mathematical point of view, it is straightforward to verify that the dynamics of θ_{tot} are independent of ν^l . Note, too, that the fast dynamics in (2.9) are independent of ν^s .

Based on these observations, we are now in a position to state our first result.

Proposition 2.1. *The fast component of the dynamics of systems with significant energy recycle is stable.*

Proof. Let $V^l = \theta_{tot}$. Based on Assumption 2.5, we have $V(\boldsymbol{\theta}) > 0$ and $V^l(\mathbf{0}) = 0$, and therefore the total energy holdup of the system is a Lyapunov function. Our reasoning above indicates that $\frac{dV^l}{d\tau} = 0$ and therefore the fast dynamics in (2.9) are **stable**. ■

In what follows, we will make a physical argument and a control argument to derive stronger stability results.

Corollary 2.1. *Considering the structure of the system and the developments above, we can write a differential equation describing the evolution of the total enthalpy of the process in the fast time scale as:*

$$\frac{d\theta_{tot}}{d\tau} = \frac{1}{H_{1,s}}(Q_{in} - Q_{out}) \quad (2.11)$$

Our assertion that $Q_{in} \equiv Q_{out}$ is based on the ideal case where all energy recovered from the outlet of the system can be recycled to the inlet. This, however, may not be

true for real systems, where thermodynamic limitations (e.g., heat transfer) would preclude the entire quantity of energy recovered from the last unit to be transferred to the first. We would thus have $Q_{out} > Q_{in}$ and, assuming that $H_{1,s} > 0$, $\frac{dV^l}{d\tau} \leq 0$. Consequently, the fast dynamics of process systems with significant energy recycling is **asymptotically stable**.

Corollary 2.2. *The inputs ν^l can be set, e.g., via a linear state feedback law*

$$\nu^l = \mathbf{K}\theta \tag{2.12}$$

*Assuming that the gains \mathbf{K} can be chosen so that the matrix $\Gamma^l(\theta)\mathbf{K}\theta$ is Hurwitz for any $\theta \in \mathcal{Q}$, the fast dynamics of systems with significant energy recycling becomes **exponentially stable**.*

We now continue our singular perturbation analysis and proceed with a derivation of an expression for the slow dynamics of the process following closely the arguments in [18, 127]. First, we note that the differential equations in the fast subsystem (2.9) are not linearly independent, given that there exists a variable (the total energy holdup) that does not depend on the internal energy flows. This confirms the existence of a slow component of the system dynamics, and also indicates that the solution of the system of algebraic equations that describes the quasi-steady state of the fast dynamics,

$$0 = \Gamma^l(\theta)\mathbf{K}\theta \tag{2.13}$$

is not an equilibrium point in the state space, but rather an equilibrium subspace where the slower dynamics evolves. Based on the physical arguments above, the

dimension of this subspace is equal to 1. To proceed with our analysis, we must characterize this manifold and ensure that a set of linearly independent constraints can be extracted from (2.13). To this end, we make the following assumption [18, 127]

Assumption 2.6. *There exist a full column rank matrix $\mathbf{B}(\boldsymbol{\theta}) \in \mathbb{R}^{\mathcal{N} \times \mathcal{N}-1}$ and a matrix $\tilde{\boldsymbol{\Gamma}}^l(\boldsymbol{\theta}) \in \mathbb{R}^{(\mathcal{N}-1) \times m^l}$ with linearly independent rows, such that $\boldsymbol{\Gamma}^l(\boldsymbol{\theta})$ can be rewritten as*

$$\boldsymbol{\Gamma}^l(\boldsymbol{\theta}) = \mathbf{B}(\boldsymbol{\theta})\tilde{\boldsymbol{\Gamma}}^l(\boldsymbol{\theta}) \quad (2.14)$$

We note that numerous examples (see, e.g., [18]) have shown that this assumption holds true for process systems of practical interest.

We can now substitute (2.14) in the original model (2.7), and consider the model in the same limit case, $\varepsilon \rightarrow 0$, in the original time scale t and under the linearly independent constraints isolated from (2.13), obtaining:

$$\begin{aligned} \dot{\boldsymbol{\theta}} &= \boldsymbol{\Gamma}^s(\boldsymbol{\theta})\boldsymbol{\nu}^s + \mathbf{B}(\boldsymbol{\theta}) \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \tilde{\boldsymbol{\Gamma}}^l(\boldsymbol{\theta})\mathbf{K}\boldsymbol{\theta} \\ 0 &= \tilde{\boldsymbol{\Gamma}}^l(\boldsymbol{\theta})\mathbf{K}\boldsymbol{\theta} \end{aligned} \quad (2.15)$$

The terms $\lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \tilde{\boldsymbol{\Gamma}}^l(\boldsymbol{\theta})\mathbf{K}\boldsymbol{\theta}$ (which, based on Equation (2.6), represent differences between large internal energy flows) become indeterminate (but still finite) in the slow time scale. These terms constitute an additional set of algebraic variables in the model of the slow dynamics. Defining these variables as $\mathbf{z} = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \tilde{\boldsymbol{\Gamma}}^l(\boldsymbol{\theta})\mathbf{K}\boldsymbol{\theta}$, the reduced-order model of the slow dynamics becomes:

$$\begin{aligned} \dot{\boldsymbol{\theta}} &= \boldsymbol{\Gamma}^s(\boldsymbol{\theta})\boldsymbol{\nu}^s + \mathbf{B}(\boldsymbol{\theta})\mathbf{z} \\ 0 &= \tilde{\boldsymbol{\Gamma}}^l(\boldsymbol{\theta})\mathbf{K}\boldsymbol{\theta} \end{aligned} \quad (2.16)$$

The differential-algebraic model (2.16) of the slow dynamics has a high index, since the algebraic variables \mathbf{z} cannot be computed directly from the algebraic equations of the model. Thus, in order to compute \mathbf{z} and obtain a state-space realization of the slow dynamics, the algebraic constraints of Equation (2.16) must be differentiated at least once with respect to the state variables $\boldsymbol{\theta}$.

After one differentiation, we obtain:

$$\mathbf{z} = -L_{\mathbf{B}(\boldsymbol{\theta})} \left(\tilde{\Gamma}^l(\boldsymbol{\theta}) \mathbf{K} \boldsymbol{\theta} \right)^{-1} L_{\Gamma^s(\boldsymbol{\theta})} \left(\tilde{\Gamma}^l(\boldsymbol{\theta}) \mathbf{K} \boldsymbol{\theta} \right) \boldsymbol{\nu}^s \quad (2.17)$$

noting that the assumption made in Corollary 2.2 ensures that the matrix $L_{\mathbf{B}(\boldsymbol{\theta})} \left(\tilde{\Gamma}^l(\boldsymbol{\theta}) \mathbf{K} \boldsymbol{\theta} \right)$ is invertible¹

An ODE representation of the slow dynamics can be obtained by substituting (2.17) in (2.16). This representation is not of minimal order, and a coordinate transformation must be used to obtain a minimal-order state-space realization. A potential transformation involves the energy balance equations and the corresponding quasi-steady state constraints [147], i.e.,:

$$\begin{bmatrix} \zeta \\ \boldsymbol{\eta} \end{bmatrix} = \mathbf{T}(\boldsymbol{\theta}) \begin{bmatrix} \boldsymbol{\delta}(\boldsymbol{\theta}) \\ \tilde{\Gamma}^l(\boldsymbol{\theta}) \mathbf{K} \boldsymbol{\theta} \end{bmatrix} \quad (2.18)$$

which yields a reduced-order model of the slow dynamics of the form:

$$\begin{aligned} \dot{\zeta} &= \frac{\partial \boldsymbol{\delta}}{\partial \boldsymbol{\theta}} \boldsymbol{\Gamma}^s(\boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\mathbf{T}^{-1}(\zeta)} \boldsymbol{\nu}^s + \frac{\partial \boldsymbol{\delta}}{\partial \boldsymbol{\theta}} \mathbf{B}(\boldsymbol{\theta}) \mathbf{z} \Big|_{\boldsymbol{\theta}=\mathbf{T}^{-1}(\zeta)} \\ \boldsymbol{\eta} &= 0 \end{aligned} \quad (2.19)$$

or, in a more compact form,

$$\dot{\zeta} = \hat{\mathbf{f}}(\zeta) + \hat{\mathbf{G}}(\zeta) \boldsymbol{\nu}^s \quad (2.20)$$

¹The *Lie derivative* (or *directional derivative*) of scalar function $a(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ along vector function $\mathbf{b}(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ (with $\mathbf{x} \in \mathbb{R}^n$) is defined as: $L_{\mathbf{b}}a(\mathbf{x}) = \frac{\partial a}{\partial \mathbf{x}}(\mathbf{x}) \mathbf{b}(\mathbf{x})$.

The aforementioned discussion on the dimension of the equilibrium manifold of the fast dynamics leads us to conclude that $\zeta \in \mathcal{Q}^1 \subset \mathbb{R}^1$ (i.e., the slow component of the dynamics is, in effect, one-dimensional), and (2.20) constitutes a (multiple-input, single output) representation of the slow dynamics of the system.

Remark 2.3. *Figure 2.1 presents the structural prototype for several processes of elevated industrial interest, including systems with feed-effluent heat exchangers, which are used to transfer heat from the process products to the feed stream, either at high temperatures (a case that will be discussed in detail in Section 2.5) or under cryogenic conditions (e.g., air separation units [281]). Systems with work exchange are also widespread, using either reciprocating or centrifugal exchange devices. The former are extensively employed in reverse osmosis water desalination processes [97], while the latter (which consist of a feed compressor driven by expanding a high-pressure product in a turbine installed on a common shaft) are a common occurrence in air separation [281] and natural gas liquefaction processes [138]. The analysis of the dynamics of systems with work exchange can be carried out in a manner similar to the one outlined above.*

2.4 Supervisory Nonlinear Model Predictive Control

The implementation of the linear feedback controller (2.12) leads to the exponential stabilization of the *fast dynamics*. A standard result in nonlinear systems theory allows us to exploit this stability property to design a controller for the entire system in the *slow time scale*, based solely on the slow model (2.20), following the *composite control* paradigm [142] (see also [53, 52]).

We formulate the system-wide control problem as the following Nonlinear Model Predictive Controller (NMPC), aimed at computing the optimal inputs $\boldsymbol{\nu}^{s*}(t)$ that provide exponential stability to the slow dynamics (2.20). To this end, we define the stage cost

$$\mathcal{L}(\boldsymbol{\zeta}(t), \boldsymbol{\nu}^s(t)) = \frac{1}{2} [S\boldsymbol{\zeta}(t)^2 + \boldsymbol{\nu}^s(t)^T \mathbf{R}\boldsymbol{\nu}^s(t)] \quad (2.21)$$

the terminal cost

$$F(\boldsymbol{\zeta}(t)) = \frac{1}{2} P\boldsymbol{\zeta}(t)^2 \quad (2.22)$$

and the cost function

$$J(t, \boldsymbol{\zeta}, \boldsymbol{\nu}^s) = F(\boldsymbol{\zeta}(t+T)) + \int_t^{t+T} \mathcal{L}(\boldsymbol{\zeta}(s), \boldsymbol{\nu}^s(s)) ds \quad (2.23)$$

where T is the prediction time horizon, $P > 0$ and $S > 0$ are penalty coefficients and $\mathbf{R} \succ 0$ is a positive-definite matrix that captures the *cost* associated with each energy input from external sources. Thus, we seek to identify the control signal $\boldsymbol{\nu}^{s*}$ as the solution to the optimization problem

$$\begin{aligned} \boldsymbol{\nu}^{s*} &= \underset{\boldsymbol{\nu}^s}{\operatorname{arg\,min}} J \\ \text{s.t. } \dot{\boldsymbol{\zeta}} &= \hat{\mathbf{f}}(\boldsymbol{\zeta}) + \hat{\mathbf{G}}(\boldsymbol{\zeta})\boldsymbol{\nu}^s, \\ \boldsymbol{\nu}^s &\in \mathcal{U}^s \subset \mathbb{R}^{m^s}, \\ \boldsymbol{\zeta} &\in Z \subset \mathbb{R}, \\ \boldsymbol{\zeta}(t+T) &\in Z_f \subset \mathbb{R} \end{aligned} \quad (2.24)$$

where the last three constraints require that the inputs and the state belong to their respective admissible sets, and the state at the end of the prediction time horizon belongs to the terminal set Z_f [178].

Proposition 2.2. *Assuming that:*

1. the set Z_f is positively invariant with respect to $\boldsymbol{\nu}^s$

2. there exists a local feedback controller $\boldsymbol{\nu}_k^s = k(\boldsymbol{\zeta})$ defined in Z_f such that F is a local Lyapunov function, i.e.,

$$\dot{F}(\boldsymbol{\zeta}) + \mathcal{L}(\boldsymbol{\zeta}) \leq 0 \quad (2.25)$$

3. the prediction horizon T is sufficiently long for $\boldsymbol{\zeta}^*(t+T) \subset Z_f$, where the $*$ symbol denotes the closed-loop state trajectory under the controller (2.24)

4. the function $\psi(\boldsymbol{\zeta}, \boldsymbol{\nu}^s) = \hat{\mathbf{f}}(\boldsymbol{\zeta}) + \hat{\mathbf{G}}(\boldsymbol{\zeta})\boldsymbol{\nu}^s$ is Lipschitz in $\boldsymbol{\zeta}$, with Lipschitz constant L , i.e., $\|\psi(\boldsymbol{\zeta}, \boldsymbol{\nu}^s)\| \leq L\|\boldsymbol{\zeta}\|$

the slow component of the system dynamics (2.20) is exponentially stable under the control law (2.24).

Proof. The proof follows the arguments in [122] as well as the developments made in [108] for SISO systems, and expands the reasoning to the MISO case of interest. Let us consider the Lyapunov function candidate:

$$V(t, \boldsymbol{\zeta}) = \min_{\boldsymbol{\nu}^s} J(t, \boldsymbol{\zeta}, \boldsymbol{\nu}^s) \quad (2.26)$$

We first establish the *asymptotic* stability of the slow dynamics under the proposed controller. To this end, we compute the time derivative of the Lyapunov function, \dot{V} , as:

$$\dot{V} = \lim_{\Delta t \rightarrow 0} \frac{V(t + \Delta t, \boldsymbol{\zeta}(t + \Delta t)) - V(t, \boldsymbol{\zeta}(t))}{\Delta t} \quad (2.27)$$

Let $\hat{\boldsymbol{\nu}}(s)$, $s \in [t, \infty)$ be the input sequence consisting of the optimal input $\boldsymbol{\nu}^{s*}$, for the period of time $s \in [t, t+T)$, and the input $\boldsymbol{\nu}_k^s(s)$ computed by the local controller at times $s \geq t+T$. Then, since $V(t + \Delta t, \boldsymbol{\zeta}) \leq J(t + \Delta t, \boldsymbol{\zeta}, \hat{\boldsymbol{\nu}}^s)$, we have

$$\dot{V} \leq \lim_{\Delta t \rightarrow 0} \frac{J(t + \Delta t, \boldsymbol{\zeta}(t + \Delta t), \hat{\boldsymbol{\nu}}^s) - V(t, \boldsymbol{\zeta}(t))}{\Delta t} \quad (2.28)$$

Now,

$$\begin{aligned}
J(t + \Delta t, \zeta(t + \Delta t), \hat{\nu}^s) &= F(\zeta(t + \Delta t + T)) & (2.29) \\
&+ \int_{t+\Delta t}^{t+\Delta t+T} \mathcal{L}(\zeta, \hat{\nu}^s) ds \\
&= F(\zeta(t + \Delta t + T)) + \int_{t+\Delta t}^{t+T} \mathcal{L}(\zeta, \nu^{s*}) ds \\
&+ \int_{t+T}^{t+\Delta t+T} \mathcal{L}(\zeta, \nu_k^s) ds
\end{aligned}$$

and therefore

$$\begin{aligned}
J(t + \Delta t, \zeta, \hat{\nu}^s) - V(t, \zeta) &= F(\zeta(t + \Delta t + T)) + \int_{t+\Delta t}^{t+T} \mathcal{L}(\zeta, \nu^{s*}) ds & (2.30) \\
&+ \int_{t+T}^{t+\Delta t+T} \mathcal{L}(\zeta, \nu_k^s) ds \\
&- F(\zeta(t + T)) - \int_t^{t+T} \mathcal{L}(\zeta, \nu^{s*}) ds \\
&= F(\zeta(t + \Delta t + T)) + \int_{t+T}^{t+\Delta t+T} \mathcal{L}(\zeta, \nu_k^s) ds \\
&- F(\zeta(t + T)) - \int_t^{t+\Delta t} \mathcal{L}(\zeta, \nu^{s*}) ds
\end{aligned}$$

Based on assumption 2 in the proposition, we have $\dot{F}(\zeta) + \mathcal{L}(\zeta, \nu_k^s) \leq 0$, $\forall s \in [T, \infty)$. Using this expression, we can write:

$$\int_{t+T}^{t+\Delta t+T} \dot{F}(\zeta(s)) ds + \int_{t+T}^{t+\Delta t+T} \mathcal{L}(\zeta, \nu_k^s) ds \leq 0 \quad (2.31)$$

and thus

$$F(\zeta(t + \Delta t + T)) - F(\zeta(t + T)) + \int_{t+T}^{t+\Delta t+T} \mathcal{L}(\zeta, \nu_k^s) ds \leq 0 \quad (2.32)$$

Substituting all of the above in (2.28), we get

$$\dot{V} \leq \lim_{\Delta t \rightarrow 0} \frac{\int_t^{t+\Delta t} -\mathcal{L}(\zeta, \nu^{s*}(s)) ds}{\Delta t}$$

$$\begin{aligned}
&\leq -\frac{1}{2}S\zeta(t)^2 - \frac{1}{2}\boldsymbol{\nu}^{s*T}\mathbf{R}\boldsymbol{\nu}^{s*} \\
&\leq -\frac{1}{2}S\zeta(t)^2
\end{aligned}$$

The Lyapunov function V thus meets the conditions in Theorem 4.9 in [137], and, consequently, the slow component of the system dynamics is *asymptotically* stable.

In what follows, we use the Lipschitz properties of the function ψ to establish bounds on V as follows:

- we have $\boldsymbol{\nu}^s(t)^T\mathbf{R}\boldsymbol{\nu}^s(t) > 0$ (\mathbf{R} is positive definite) and $F(\zeta(t+T)) > 0$. Therefore, we have $V(t, \zeta) \geq \frac{1}{2} \int_t^{t+T} S\zeta(s)^2 ds$.

Then, since ψ is Lipschitz, we have that $\|\zeta(s)\| \geq \|\zeta(t)\| \exp[-L(s-t)]$, $s \in [t, \infty)$ (see, *e.g.*, Theorem 3.4 in [137]), and:

$$\begin{aligned}
V(t, \zeta) &\geq \frac{1}{2} \int_t^{t+T} S\zeta(s)^2 ds & (2.33) \\
&\geq \frac{1}{2} S\zeta(t)^2 \int_t^{t+T} e^{-2L(s-t)} ds \\
&\geq \frac{1 - e^{-2LT}}{4L} S\zeta(t)^2
\end{aligned}$$

which provides a *lower bound* for V .

- Let $\bar{\zeta}(s)$ be the state at time s starting from an initial condition ζ_0 and with zero energy input, $\boldsymbol{\nu}^s(t) = \mathbf{0}$. As $V(t, \zeta)$ is the minimum of J , $V(t, \zeta) \leq J(t, \bar{\zeta}, 0)$. From the Lipschitz property of ψ , it follows that $\|\bar{\zeta}(s)\| \leq \|\zeta(t)\| \exp[L(s-t)]$, $s \in [t, \infty)$. Consequently, we have:

$$\begin{aligned}
V(t, \zeta) &\leq \frac{1}{2} P\bar{\zeta}(t+T)^2 + \frac{1}{2} \int_t^{t+T} S\bar{\zeta}(s)^2 ds \\
&\leq \frac{1}{2} P e^{2LT} \zeta(t)^2 + \frac{1}{2} S\zeta(t)^2 \int_t^{t+T} e^{2L(s-t)} ds
\end{aligned}$$

$$\leq \left[\frac{1}{2}Pe^{2LT} + S\frac{e^{2LT} - 1}{4L} \right] \zeta(t)^2$$

providing an *upper bound* for V .

With these bounds (and recalling the asymptotic stability property established above), the function V satisfies the assumptions of Theorem 4.10 in [137], and therefore the controller (2.24) guarantees the exponential stability of the slow dynamics (2.20). ■

Remark 2.4. *The fulfillment of Assumptions 1-3 in Proposition 2.2 guarantees asymptotic stability [178], while Assumption 4 [108] is the key component for establishing the exponential stability property of the controller (2.24).*

Following these developments, we can state:

Theorem 2.1. *Consider an integrated process system of the type (2.1) with high energy recycling, under the composite control consisting of the linear feedback (2.12) and the MISO NMPC (2.24).*

The closed-loop the dynamics of the system are described by:

$$\dot{\zeta} = \tilde{\mathbf{f}}(\zeta, \boldsymbol{\eta}) \tag{2.34}$$

$$\varepsilon \dot{\boldsymbol{\eta}} = \tilde{\mathbf{g}}(\zeta, \boldsymbol{\eta}, \varepsilon) \tag{2.35}$$

Let us assume [137] that $\tilde{\mathbf{f}}(\mathbf{0}, \mathbf{0}) = \mathbf{0}$ and $\tilde{\mathbf{g}}(\mathbf{0}, \mathbf{0}, \varepsilon) = \mathbf{0}$, and that $\mathbf{0} = \tilde{\mathbf{g}}(\zeta, \boldsymbol{\eta}, 0)$ has an isolated root $\boldsymbol{\eta} = \mathbf{p}(\zeta)$, such that $\mathbf{p}(\mathbf{0}) = \mathbf{0}$. Then, there exists $\varepsilon^ > 0$ such that the origin ($\boldsymbol{\theta} = 0$) of the original system (2.1) is exponentially stable in \mathcal{Q} for all $0 < \varepsilon < \varepsilon^*$.*

Proof. The proof follows the arguments of Theorem 11.4 in [137]. By converse Lyapunov theorems, there exists a Lyapunov function V^s for the closed-loop slow dynamics², such that:

$$c_1 \|\zeta\|^2 \leq V^s(\zeta) \leq c_2 \|\zeta\|^2 \quad (2.36)$$

$$\frac{\partial V^s}{\partial \zeta} \cdot \tilde{\mathbf{f}}(\zeta, \mathbf{p}(\zeta)) \leq -c_3 \|\zeta\|^2 \quad (2.37)$$

$$\left\| \frac{\partial V^s}{\partial \zeta} \right\| \leq c_4 \|\zeta\| \quad (2.38)$$

with $c_i > 0$, $i = 1, \dots, 4$.

Observing that (2.35) is a slowly varying system with respect to ζ (see, e.g., Lemma 9.8 in [137]), we can write a Lyapunov function for the fast subsystem, $V^f(\zeta, \bar{\eta})$, with

$$\bar{\eta} = \eta - \mathbf{p}(\zeta) \quad (2.39)$$

such that:

$$b_1 \|\bar{\eta}\|^2 \leq V^f(\zeta, \bar{\eta}) \leq b_2 \|\bar{\eta}\|^2 \quad (2.40)$$

$$\frac{\partial V^f}{\partial \bar{\eta}} \cdot \tilde{\mathbf{g}}(\zeta, \bar{\eta} + \mathbf{p}(\zeta), 0) \leq -b_3 \|\bar{\eta}\|^2 \quad (2.41)$$

$$\left\| \frac{\partial V^f}{\partial \bar{\eta}} \right\| \leq b_4 \|\bar{\eta}\| \quad (2.42)$$

$$\left\| \frac{\partial V^f}{\partial \zeta} \right\| \leq b_5 \|\bar{\eta}\|^2 \quad (2.43)$$

where $b_i > 0$, $i = 1, \dots, 5$.

²We note that V^s is not necessarily the same as the function V defined in (2.26)

Under the new coordinates (2.39), (2.34)-(2.35) becomes:

$$\dot{\zeta} = \tilde{\mathbf{f}}(\zeta, \bar{\eta} + \mathbf{p}(\zeta)) \quad (2.44)$$

$$\varepsilon \dot{\bar{\eta}} = \tilde{\mathbf{g}}(\zeta, \bar{\eta} + \mathbf{p}(\zeta), \varepsilon) - \varepsilon \frac{\partial \tilde{\mathbf{p}}}{\partial \zeta} \tilde{\mathbf{f}}(\zeta, \bar{\eta} + \mathbf{p}(\zeta)) \quad (2.45)$$

We can define a “composite” Lyapunov function candidate for (2.44)-(2.45) as:

$$V^{fs} = V^f + V^s \quad (2.46)$$

We observe that:

- $\tilde{\mathbf{g}}$ is Lipschitz in ε linearly in the states $\zeta, \bar{\eta}$:

$$\|\tilde{\mathbf{g}}(\zeta, \bar{\eta} + \mathbf{p}(\zeta), \varepsilon) - \tilde{\mathbf{g}}(\zeta, \bar{\eta} + \mathbf{p}(\zeta), 0)\| \leq \varepsilon L_1 (\|\zeta\| + \|\bar{\eta}\|) \quad (2.47)$$

- since $\tilde{\mathbf{f}}(\zeta, \mathbf{p}(\zeta))$ and $\mathbf{p}(\zeta)$ vanish at $\zeta = \mathbf{0}$, we have

$$\left\| \tilde{\mathbf{f}}(\zeta, \bar{\eta} + \mathbf{p}(\zeta)) - \tilde{\mathbf{f}}(\zeta, \bar{\eta}) \right\| \leq L_2 \|\bar{\eta}\| \quad (2.48)$$

$$\left\| \tilde{\mathbf{f}}(\zeta, \mathbf{p}(\zeta)) \right\| \leq L_3 \|\zeta\| \quad (2.49)$$

$$\left\| \frac{\partial \tilde{\mathbf{p}}}{\partial \zeta} \right\| \leq k_1 \quad (2.50)$$

Based on the above, it can be verified that the derivative of V^{fs} along the trajectories (2.44)-(2.45) fulfils:

$$\dot{V}^{fs} \leq -a_1 \|\zeta\|^2 - \frac{a_2}{\varepsilon} \|\bar{\eta}\|^2 + a_3 \|\bar{\eta}\|^2 + a_4 \|\zeta\| \|\bar{\eta}\| + a_5 \|\zeta\| \|\bar{\eta}\|^2 + a_6 \|\bar{\eta}\|^3 \quad (2.51)$$

with positive a_1 and a_2 , and nonnegative a_3, \dots, a_6 . For all bounded $\bar{\boldsymbol{\eta}}$ with $\|\bar{\boldsymbol{\eta}}\| \leq \rho$, this simplifies to:

$$\dot{V}^{fs} \leq -a_1 \|\boldsymbol{\zeta}\|^2 - \frac{a_2}{\varepsilon} \|\bar{\boldsymbol{\eta}}\|^2 + a_7 \|\bar{\boldsymbol{\eta}}\|^2 + 2a_8 \|\boldsymbol{\zeta}\| \|\bar{\boldsymbol{\eta}}\| \quad (2.52)$$

$$= - \begin{bmatrix} \|\boldsymbol{\zeta}\| \\ \|\bar{\boldsymbol{\eta}}\| \end{bmatrix}^T \begin{bmatrix} a_1 & -a_8 \\ -a_8 & (a_2/\varepsilon) - a_7 \end{bmatrix} \begin{bmatrix} \|\boldsymbol{\zeta}\| \\ \|\bar{\boldsymbol{\eta}}\| \end{bmatrix} \quad (2.53)$$

Thus, there exists an $\varepsilon^* > 0$ such that for all $0 < \varepsilon < \varepsilon^*$ for which the above quadratic form is positive definite, and we have (see, e.g., Theorem 4.10 in [137])

$$\dot{V}^{fs} \leq -2\gamma V^{fs} \quad (2.54)$$

for $\gamma > 0$. Consequently, we have:

$$V^{fs}(t, \boldsymbol{\zeta}(t), \bar{\boldsymbol{\eta}}(t)) \leq e^{-2\gamma(t-t_0)} V^{fs}(t_0, \boldsymbol{\zeta}(t_0), \bar{\boldsymbol{\eta}}(t_0)) \quad (2.55)$$

and, from the properties of the Lyapunov functions V^s and V^f , we have:

$$\left\| \begin{bmatrix} \boldsymbol{\zeta}(t) \\ \bar{\boldsymbol{\eta}}(t) \end{bmatrix} \right\| \leq K_1 e^{-2\gamma(t-t_0)} \left\| \begin{bmatrix} \boldsymbol{\zeta}(t_0) \\ \bar{\boldsymbol{\eta}}(t_0) \end{bmatrix} \right\| \quad (2.56)$$

Recalling that $\bar{\boldsymbol{\eta}} = \boldsymbol{\eta} - \mathbf{p}(\boldsymbol{\zeta})$ and $\|\mathbf{p}(\boldsymbol{\zeta})\| \leq k_1 \|\boldsymbol{\zeta}\|$, we obtain:

$$\left\| \begin{bmatrix} \boldsymbol{\zeta}(t) \\ \boldsymbol{\eta}(t) \end{bmatrix} \right\| \leq K_2 e^{-2\gamma(t-t_0)} \left\| \begin{bmatrix} \boldsymbol{\zeta}(t_0) \\ \boldsymbol{\eta}(t_0) \end{bmatrix} \right\| \quad (2.57)$$

which captures the desired exponential stability property. ■

Remark 2.5. *The role of the control law in (2.24) can be interpreted based on the relative values of the penalty coefficients P , R and S . From this perspective, (2.24)*

can be construed as either a NMPC controller designed for disturbance rejection or setpoint tracking (this is the case, for example, when $P/S = \mathcal{O}(1)$ and $R = \mathbf{diag}[\lambda_1 \dots \lambda_{m^s}]$, with $\lambda_i/\lambda_j = \mathcal{O}(1)$ and $P/\lambda_i = \mathcal{O}(1)$), or a real-time optimization of energy use (in which case $P/S = \mathcal{O}(1)$ and $P/\lambda_i \ll 1$). Note that in the latter case, control moves which consist of increasing the energy input to the system are strongly penalized.

The framework developed above can also be extended to account for predicted values of time-varying energy-relevant disturbances (e.g., weather, real-time or time-of-day changes in energy utility costs), thereby providing a predictive, rather than reactive, approach to managing system energy use.

Remark 2.6. A purely economic control objective that only accounts for the cost of the external energy inputs (i.e., having $P = S = 0$) can be employed to develop a version of the supervisory controller. However, this approach will only guarantee the asymptotic stability [8] in the slow time scale and the overall stability result in Theorem 2.1 is no longer valid. Note, however, that stability is still likely to be achieved in practical situations.

2.5 Case Study: Control of a Reactor Feed-Effluent Heat Exchanger Process

We consider a process with energy recovery as shown in Figure 2.2. The process consists of an adiabatic continuously stirred tank reactor (CSTR) and a feed-effluent heat exchanger (FEHE). Such process configurations are frequently used in industry for carrying out reactions that occur at high temperature but

are only mildly exothermic, when the role of the FEHE is to pre-heat the raw material stream based on recovering heat (including the heat of reaction) from the reactor effluent stream³. A fired heater is also present in the process for improved operability: its role is to ensure that the temperature of the reactor feed is at a desired value that ensures, e.g., maximum conversion. A bypass stream is available as a further means to adjust the operation of the FEHE. We assume that the reaction $A \rightarrow B$, which is first-order with rate constant $k = k_0 e^{(-E/RT)}$, occurs in the reactor. The component A is fed to the process at a volumetric flow rate F_{in} , with the inlet stream having molar concentration c_{A0} .

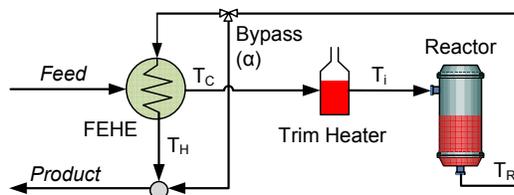


Figure 2.2: Reactor-FEHE process with fired furnace and exchanger bypass stream.

In order to derive a model for this process, we assume that the physical properties of the process streams (notably, heat capacity, density and thermal conductivity) do not vary significantly with temperature. Also, we consider that heat exchange in the FEHE occurs without phase change (no evaporation or condensation take place) and, further, that the quantity of material contained in both channels of the heat exchanger does not vary in time (i.e., constant holdup). Under these assumptions, we can write the process model as [18]:

³The use of FEHEs is in general avoided when strongly exothermic reactions are involved due to the risk of temperature runaway in the presence of disturbances that cause an imbalance in the rate at which the heat recovered from the effluent is absorbed by the feed stream.

$$\begin{aligned}
\frac{dV}{dt} &= F_{in} - F \\
\frac{dc_A}{dt} &= \frac{F_{in}}{V}(c_{Ao} - c_A) - kc_A \\
\frac{dT_R}{dt} &= \frac{F_{in}}{V}(T_i - T_R) - \frac{1}{\rho C_p}kc_A\Delta H \\
\frac{\partial T_H}{\partial t} &= -v_H \frac{\partial T_H}{\partial z} - \frac{UA}{\rho C_p} \frac{T_H - T_C}{V_H} \\
\frac{\partial T_C}{\partial t} &= -v_C \frac{\partial T_C}{\partial z} + \frac{UA}{\rho C_p} \frac{T_H - T_C}{V_C} \\
\frac{dT_i}{dt} &= \frac{F_{in}}{V_f}(T_{C_{z=L}} - T_i) + \frac{Q_H}{\rho C_p V_f}
\end{aligned} \tag{2.58}$$

The temperature in the FEHE is distributed in the axial (z) direction, and, in order to complete the model, we define the boundary conditions for T_H , the temperature in the hot leg of the exchanger, and, respectively, for T_C , the temperature in the cold leg, as $T_{H_{z=L}} = T_R$, $T_{C_{z=0}} = T_{in}$, with T_{in} being the temperature of the feed stream, and T_R the reactor temperature. We denote by T_{exit} the temperature of the product stream exiting the process, and by T_i the temperature of the stream leaving the trim heater. Let $v_H = F_{in}(1 - \alpha)/A_H$ and $v_C = F_{in}/A_C$, where α is the bypass ratio. Finally, we denote the holdups of the reactor, the FEHE cold and hot sides, and the trim furnace by V , V_H , V_C and V_f , respectively, and the duty of the furnace by Q_H .

The nominal values of the process variables and parameters are presented in Table 2.1.

By performing an energy balance calculation, we can show that the temperature values in the two sides of the heat exchanger at the “cold end” (i.e., at the inlet of the feed stream) are very similar (a situation often referred to as a “tem-

Table 2.1: Process parameters (adapted from [18])

k_o	1.2667×10^7	s^{-1}	α	0.1	
E	142870.0	J/mol	T_R	922.0	K
ΔH	-54.828	kJ/mol	T_i	909.62	K
ρC_p	4.184×10^6	$J/m^3/K$	T_{exit}	301.83	K
UA	83680	W/K	$T_{C,z=L}$	864.59	K
c_A	55.19	mol/l	T_{in}	300.0	K
Q_H	108.7	kW	V	0.1	m^3
F_{in}	5.77×10^{-4}	m^3/s	V_H	0.1	m^3
F	5.77×10^{-4}	m^3/s	V_C	0.09	m^3
c_{Ao}	1000.0	mol/m^3	V_f	0.01	m^3
L	5.0	m			

perature pinch” [157]). Evidently, this is very desirable and is indicative of the fact that the process is of the type discussed above, having high energy recovery.

For this process, we can define the energy recovery number as

$$Rc = H_{rec,s}/[F_{in}\rho C_p(T_{in} - T_{ref})]_s \quad (2.59)$$

with T_{ref} being a reference temperature. Following the developments in Section 2.3, we also define the following $\mathcal{O}(1)$ quantities, where $j \in \{R, i, C\}$:

$$\begin{aligned} k_j &= \frac{[F_{in}\rho C_p(T_j - T_{ref})]_s}{H_{rec,s}} \\ u_j &= \frac{F_{in}\rho C_p(T_j - T_{ref})}{[F_{in}\rho C_p(T_j - T_{ref})]_s} \end{aligned} \quad (2.60)$$

and $u_{rec} = H_{rec}/H_{rec,s}$.

Following the procedure outlined earlier in the paper, we define the stretched, fast time variable $\tau = t/\varepsilon$. Subsequently, we can rewrite the model in the fast time scale and, by considering the limit $\varepsilon \rightarrow 0$, we obtain a description of the fast dynamics of the process in the form of Equation (2.9):

$$\begin{aligned}
\frac{dV}{d\tau} &= 0 \\
\frac{dc_A}{d\tau} &= 0 \\
\frac{dT_R}{d\tau} &= \frac{[F_{in}(T_{in} - T_{ref})]_s}{V} (k_i u_i - k_R u_R) \\
\frac{dT_H}{d\tau} &= \frac{[F_{in}(T_{in} - T_{ref})]_s}{V_H} (k_R u_R - u_{rec}) \\
\frac{dT_C}{d\tau} &= \frac{[F_{in}(T_{in} - T_{ref})]_s}{V_C} (-k_C u_C + u_{rec}) \\
\frac{dT_i}{d\tau} &= \frac{[F_{in}(T_{in} - T_{ref})]_s}{V_f} (k_C u_C - k_i u_i)
\end{aligned} \tag{2.61}$$

Subsequently, using the analysis tools presented earlier in the paper, a description of the slow dynamics of the process can be obtained [18] in the form :

$$\begin{aligned}
\dot{V} &= F_{in} - F \\
\dot{c}_A &= \frac{F_{in}}{V} (c_{Ao} - c_A) - k_o e^{(-E/RT_R)} c_A \\
\dot{T}_R &= [-k c_A T_R \Delta H V U A + \rho C_p T_{in} U A (F_{in} - F) \\
&\quad + 8 (\rho C_p)^2 F F_{in} T_R (\alpha - 1) + \rho C_p F T_{in} \alpha U A \\
&\quad - \alpha \rho C_p T_R F U A + Q_H U A] / \text{DEN} \\
\text{DEN} &= \rho C_p [U A (V_C + V_f + V) + 8 V_h F_{in} \rho C_p (1 - \alpha)]
\end{aligned} \tag{2.62}$$

which constitutes a state-space realization of the form (2.20), with $\zeta = T_R$ and $\omega^s = Q_H$.

In order to carry out numerical simulations for the subsequent examples, the spatial derivatives in the plant model in Eq. (2.58) were discretized using a backwards finite difference scheme on a grid of 301 nodes, while the supervisory controller was based on the reduced-order model of the slow dynamics in Eq. (2.62). Lower and upper bounds of, respectively, 910K and 940K were imposed on the reactor temperature T_R . In order to achieve offset-free tracking, we utilized a model extension

following the ideas proposed in [204, 203], which effectively endows the NMPC controller with integral action. We note that this approach does not change the dynamic analysis and control results presented above.

2.5.1 Example 1

We performed simulations considering a production increase situation, whereby the feed flow rate is increased by 10% at time $t = 0.25\text{h}$, along with an increase in the feed temperature to 310K . We considered both a tracking controller (in which case the tuning parameters in Eq. (2.24) were set to $P = S = 10$ and $R = 10^{-6}$ as discussed in Remark 2.5), and an economics-oriented optimization, in which case R was increased to 10 and $P = S = 10^{-6}$, indicating that the control objective is the minimization of the total amount of energy used. The prediction horizon was $T = 10$ minutes. Disturbance measurements were assumed to be unavailable to the controller.

Figure 2.3 shows the response of the reactor-FEHE process to these disturbances under the supervisory NMPC using the tracking objective defined above, with $T_{R,sp} = 922\text{K}$. Note that the controller exhibits excellent performance. The simulation results obtained using the economic objective function are shown in Figure 2.4. Again, the proposed control strategy exhibits very good performance; as expected, the reactor temperature is maintained at the lower acceptable bound, which, intuitively, translates into minimizing the heating rate Q_H in the furnace. We note, however, that the lower bound on the temperature is violated briefly when disturbances occur; this can be easily interpreted in light of the fact that

the direction of the disturbance is such that the temperature of the reactor drops (see the open-loop response shown in Figure 2.5), and a purely feedback control strategy cannot counter this effect.

To further elucidate this matter, we performed an additional closed-loop simulation assuming that measurements are available for both the flow and the feed temperature disturbance. In this case, the reactor temperature returns within the admissible bounds faster (Figure 2.6), but - intuitively - at the cost of a more aggressive control action.

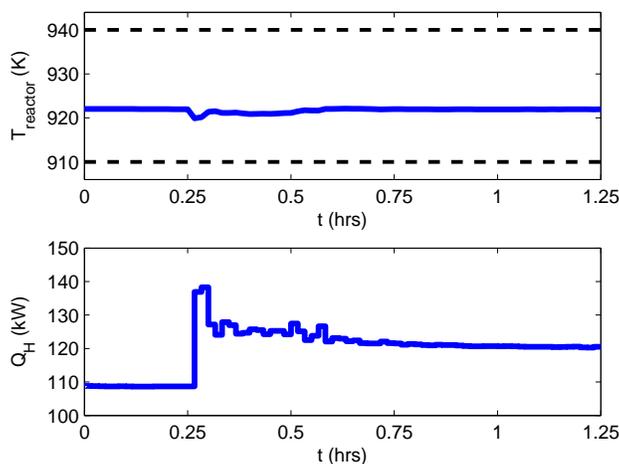


Figure 2.3: Evolution of the reactor temperature and furnace heat duty using a tracking objective

2.5.2 Example 2

We simulated a production increase scenario, with a 7% feed flow rate increase at $t = 1h$, along with an increase in the feed temperature to $310K$. We considered both a tracking controller (the tuning parameters in (2.24) were set to $P = S = 10$

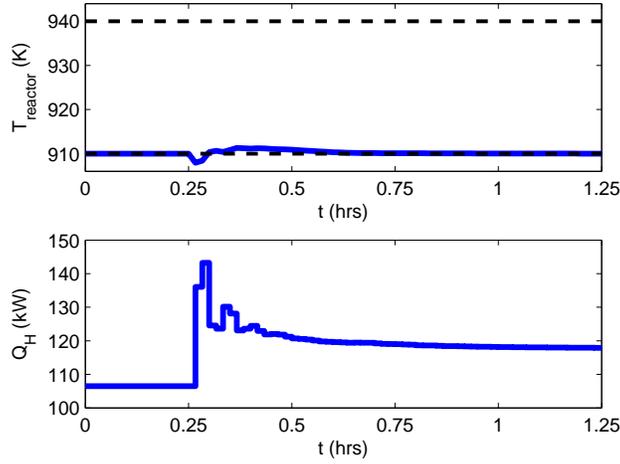


Figure 2.4: Evolution of the reactor temperature and furnace heat duty using an economic objective

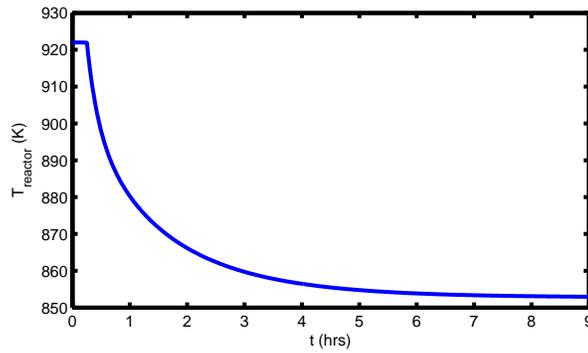


Figure 2.5: Open-loop response of the reactor temperature for an increase in both feed flow rate and inlet temperature

and $R = 10^{-6}$ as discussed in Remark 2.5), and an economics-oriented optimization, with $R = 10$ and $P = S = 10^{-6}$. The prediction horizon was $T = 10$ minutes. We solved the problem using a standard “delta-input” formulation, with $|dQ/dt| < 0.3\text{kW}/\text{min}$ as an additional constraint. In simulating the tempera-

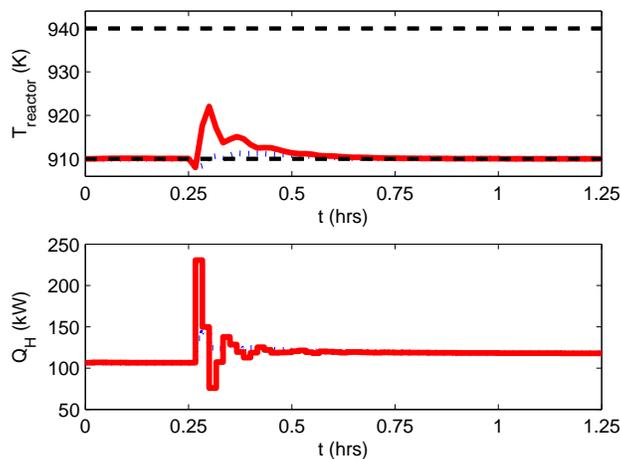


Figure 2.6: Evolution of the reactor temperature and furnace heat duty using an economic objective

ture tracking scenario, we also considered a $2K$ rise in the reactor temperature setpoint, $T_{R,sp}$, to compensate the conversion decrease associated with the rise in throughput.

Figure 2.7 shows the response of the reactor-FEHE process to these disturbances under the supervisory NMPC using the tracking objective defined in the previous section. Note that the controller exhibits excellent performance. The simulation results obtained using the economic objective function are shown in Figure 2.8. As expected, the reactor temperature is at the lower acceptable bound, which, intuitively, translates into minimizing the heating rate QH in the furnace.

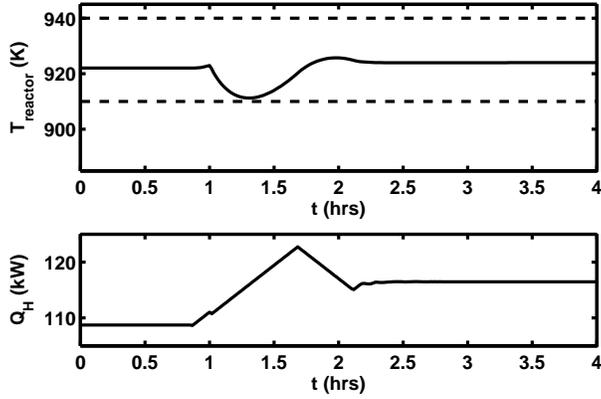


Figure 2.7: Evolution of the reactor temperature and furnace heat duty using a tracking objective

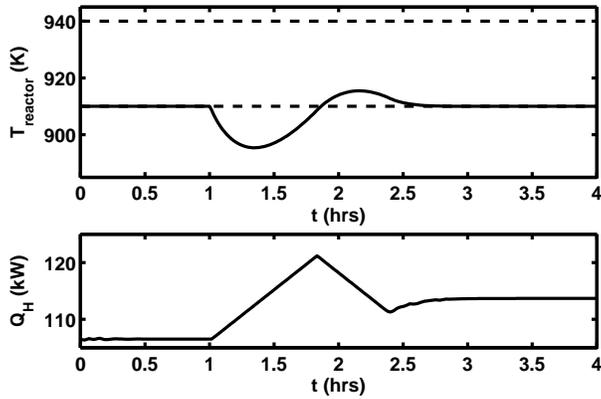


Figure 2.8: Evolution of the reactor temperature and furnace heat duty using an economic objective

2.6 Conclusions

In this paper, we considered a class of systems with significant energy recovery. Extending our previous results concerning the two time scale dynamics of such systems, we demonstrated that the fast component of the dynamics of such sys-

tems is stable in general and asymptotically stable in systems of practical interest. We used this result to develop a composite controller design, consisting of i) a linear control system to exponentially stabilize the fast dynamics and, ii) a nonlinear model predictive controller for the slow dynamics, and demonstrated that it guarantees exponential stability for the overall system. Through a transparent definition of the penalty terms in the NMPC objective function, we showed that the proposed control approach can be translated into a strategy for optimal energy management and reduction of energy use at the system level, and established a parallel between the approach introduced in this paper and economic model predictive control. Finally, we illustrated our theoretical developments with a simulation case study.

Chapter 3

Nonlinear Model Reduction and Model Predictive Control of Residential Buildings with Energy Recovery

The material in this chapter has been published in [253] and [254]

3.1 Introduction

Over 60% of the total electricity generated in the United States is used in residential and commercial buildings [266]. Daily variations in building energy utilization are at the origin of significant fluctuations in grid power demand. Meeting such changes in demand requires the use of additional (intermediate and peaking) power generation facilities, which operate at lower efficiencies, higher costs and higher environmental footprints than their base load counterparts. To reduce the variability of the energy demand posed by buildings, an understanding of the building *dynamics*, and of their impact on the design of *predictive, proactive model-based control and energy management strategies*, is essential.

Buildings are dynamical systems evolving in a time-varying environment defined by fluctuations in weather, occupancy and energy prices. In the presence of such fluctuations, heating, ventilation and air conditioning (HVAC) systems must

maintain occupant comfort within acceptable limits at a minimum cost. Operational decisions include the start-up and shutdown of heating and air-conditioning equipment units, as well as dynamic utilization of potentially available renewable energy and energy storage facilities. Decisions regarding the latter elements are especially crucial in the smart grid environment [240].

Empirical or ad-hoc decision-making by operators in building management is typically *reactive* in nature and based on the natural occupant response to variations in temperature throughout a given day. This leads to unnecessarily large energy consumption at peak times, with a considerable impact on building electricity costs [65, 245, 190]. Given the significant energy costs of buildings, it is quite surprising that on-off or simple linear controllers for temperature regulation are still the norm in practical implementations [233]. The reluctance of practitioners towards more advanced control and energy management techniques has traditionally been motivated by their relatively high cost, by robustness and real-time execution concerns, and by a potential lack of transparency for occupants and operators [233]. However, rapidly rising energy prices and the need for advanced control in the smart grid environment have spurred a surge in research regarding the application of optimization and model-based controllers to the building energy management problem [233, 289, 181].

The peak energy demand of buildings can be decreased significantly by adopting a *proactive* energy management approach, based on forecasting and optimizing energy generation, storage and utilization over a time horizon in the future [40, 65]. This requires i) *predicting weather and other factors (e.g., human activity) that*

impact energy consumption, ii) using these predictions and a model of the building to *anticipate energy needs* and, iii) optimizing energy use and reallocating loads via an appropriate *operating schedule* for the HVAC and energy storage systems.

Of particular interest is the use of model predictive control (MPC), originally developed by the chemical industry [218], owing to its ability to address an economic or temperature tracking objective while satisfying constraints on the system (e.g., occupant comfort levels) and accounting for forecasts of disturbances (e.g., changes in weather, occupancy). The implementation of these intuitive concepts poses, however, a number of significant difficulties, both practical and fundamental. In particular, MPC implementations are vitally dependent on the availability of an accurate building model that can be run rapidly, repeatedly and reliably in real-time. Detailed design models such as those underlying advanced building simulation software (e.g., EnergyPlus [263], eQUEST [123], TRNSYS [141]) are of considerable size and complexity and do not lend themselves easily to use in *on-line* calculations.

Several approaches have been employed for the identification of models consisting of approximate representations of key components of the building dynamics, which can then be used to provide the prediction of future states in the MPC framework.

A number of works [104, 163, 164, 109, 224, 102, 103, 289, 195] discuss the use of models with simplified structures such as resistance-capacitance (RC) networks. Parameters are identified from building operating data or detailed simulations as a means to simplify the modeling process. However, the derived models are typ-

ically linear and therefore exhibit extrapolation difficulties, because buildings are inherently nonlinear systems. Reduction approaches such as balanced realizations [110] have been applied to obtain linear and nonlinear reduced-order models of RC networks that retain a small number of states which evolve in a similar (typically slow) time scale [104, 163].

Data-driven methods, including neural network models and genetic optimization algorithms have lower online computational requirements but require long training periods, have poor extrapolation performance and lose physical insight [189, 144, 64]. Data obtained from detailed simulation software like EnergyPlus [263] are often used to generate linear state-space models, which simplify the online optimization calculation [197, 186, 162, 163, 215, 286, 285]. Even in cases where nonlinear system identification is performed, the result may eventually be linearized for use in MPC [196].

While all of these models are convenient for MPC implementations, their derivation is a laborious process that must be repeated for every new application; furthermore, they are inherently inaccurate when the operating conditions deviate from those considered during system identification [285]. Finally, the model variables themselves are devoid of physical meaning and carry no explicit correlation with building design parameters (e.g., geometry, material properties, etc.).

On the other hand, *first-principles, nonlinear models* preserve their accuracy over a broader range of operating conditions. To fully exploit this property, the MPC problem should also be formulated in a nonlinear context (i.e., nonlinear MPC - NMPC). Solving large-scale nonlinear optimization problems in real-time

is problematic. In this context, *nonlinear model reduction*, i.e., the derivation of low-dimensional nonlinear models that are suitable for online computations, is extremely important. However, results using NMPC are rather scarce and varied in their approach [216].

In this paper, we introduce an aggregate modeling approach and a singular perturbations-based framework for the dynamic analysis and model reduction of building models. While such techniques have been –more or less formally– used in the past [136, 274, 85], applications have relied largely on ad-hoc arguments. Working with a prototype building model, we present a theoretical justification of the empirically observed multiple time scale dynamic response of buildings, and develop a mathematically rigorous methodology for deriving reduced-order models for the dynamics in each time scale. Our analysis accounts for the potential use of Heat Recovery Ventilators (HRVs) [225, 159], and we show that their presence leads to the emergence of a dynamic behavior with *three* time scales, including an overall, system-wide component which involves *both* the building *and* the HVAC system. A simulation case study demonstrates the use of the derived reduced-order models in the synthesis of a nonlinear predictive model-based optimal energy management strategy for a single-zone test building situated on the University of Texas campus. The proposed controller exhibits excellent performance, can easily be executed in real-time and its implementation results in a shifted peak demand with potential cost savings compared to conventional control strategies.

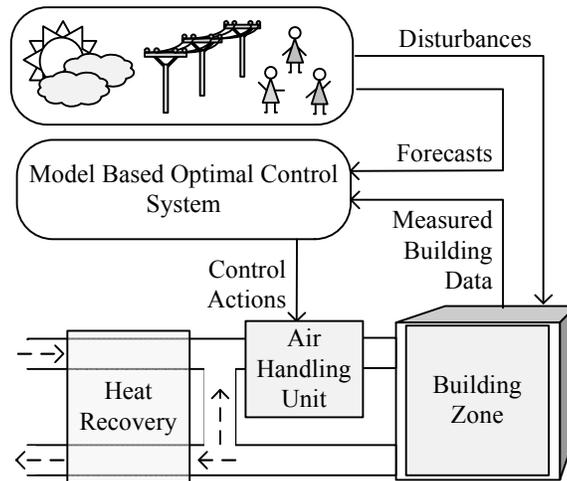


Figure 3.1: Prototype single-zone building HVAC model and control system

3.2 System Description

We consider a prototype single-zone building and its HVAC system, as shown in Figure 3.1, which is based on one unit of the UT Thermal Façade Laboratory, shown in Figure 3.2. Five of the laboratory’s surfaces are well insulated walls, and the entire south façade is a window. Regarding the function of the HVAC system units, outside air is pre-heated or pre-cooled with the zone exhaust air in a Heat Recovery Ventilator (HRV). This reduces the amount of heating or cooling that has to be provided by the Air Handling Unit (AHU). The fresh air exiting the HRV is heated or cooled in the AHU, then fed to the interior space or zone. Air then exits the interior zone; part of it may be recycled and mixed with the fresh air feed, while the remainder is directed through the HRV to the atmosphere. The building details and interaction with the environment are shown in Figure 3.3, and values for the building parameters (geometric dimensions, physical properties of



Figure 3.2: University of Texas Thermal Façade Laboratory (photo courtesy of the UT School of Architecture).

materials, etc.) are presented in the Appendix. The outer wall material is cement and the inner material is an insulating layer.

The operation of the system is subject to disturbances, which include fluctuations in outdoor air temperature, changes in cloud cover and solar irradiation through windows, and to variations in the heat generated by the building occupants (including human presence, the use of lighting, etc.).

Note that a single-zone *does not* necessarily correspond to a *single room*; a zone can be thought of as the airspace handled using the same HVAC system. In effect, this representation can be used to capture the dynamics of a building with multiple rooms, and possibly extending to a much larger size, e.g., a typical residential home served by a single HVAC system.

We are interested in identifying the fundamental dynamic features of this system and their influence on the design of an advanced control system. For this

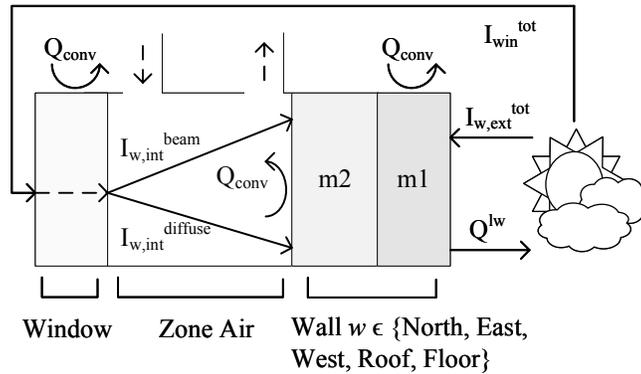


Figure 3.3: Interactions between the building zone and environment. The temperature of the air in the zone is influenced by conductive heat transfer through the window and walls (only one wall is shown) and convective heat transfer at these surfaces. The structural element temperatures are influenced by solar radiation (I), long wave radiation, and the outdoor air temperature.

analysis, we propose using a *lumped-parameter representation* of the prototype building. Lumped parameter models are derived from grouping particular structural elements with similar properties in order to yield a model with fewer states but reasonable accuracy. In this paper we will consider two possible lumped representations (note that coarse-graining, or the derivation of such aggregate models, is a research topic in its own right): i) a *minimal* aggregate model, consisting of a single thermal mass representative of the entire building structure, and ii) a more *granular* aggregate model, which represents each wall as two thermal masses in series. The minimal aggregate model will be analyzed in detail because of its simplicity. Then the same procedure will be applied to the more detailed model, which will be used in the case study later in the paper.

3.2.1 Minimal Aggregate Model

Based on the assumption that the air within the different system units is sufficiently well-mixed, and on the assumption that the building structural elements are sufficiently conductive for their temperature to be uniform, we propose the *minimal aggregate* model in Figure 3.4.

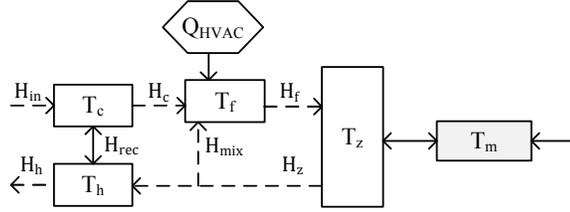


Figure 3.4: Minimal aggregate model. Dashed lines represent mass flow while solid lines represent heat transfer. Each block labeled T_i corresponds to a state in the model, and the heat streams of enthalpy H_i show the connections between the lumped building elements. Q_{HVAC} represents the heat removed from the zone feed air stream by the AHU.

Here, we assume that the operating conditions of the building are such that the heat capacity, density and thermal conductivity of the air and building materials do not vary considerably. To simplify the dynamic analysis, we do not include any complex interactions between the structural elements and the environment (*e.g.*, nonlinear terms like convection coefficient correlations, long wave radiation). Under these assumptions, the dynamics of the system in Figure 3.4 can be described by the following set of energy balance equations:

$$\frac{dT_z}{dt} = (H_f - H_z - U_m A_m (T_z - T_m) - Q_{dist}) / \rho_{air} C_{p,air} V_z \quad (3.1)$$

$$\frac{dT_f}{dt} = (H_c + H_{mix} - H_f + Q_{HVAC}) / \rho_{air} C_{p,air} V_f \quad (3.2)$$

$$\frac{dT_c}{dt} = (H_{in} - H_c - H_{rec}) / \rho_{air} C_{p,air} V_C \quad (3.3)$$

$$\frac{dT_h}{dt} = (H_z - H_h - H_{mix} + H_{rec}) / \rho_{air} C_{p,air} V_H \quad (3.4)$$

$$\frac{dT_m}{dt} = (U_m A_m (T_z - T_m) + U_{ext} A_{ext} (T_{ext} - T_m)) / \rho_m C_{p,m} V_m \quad (3.5)$$

where $H_i = F_i \rho C_{p,air} T_i$ is the enthalpy of air flow at point i , $U_m A_m$ are the heat transfer coefficient and heat transfer area between the zone air and the building structural elements, and $U_{ext} A_{ext}$ represents heat transfer from the lumped structural elements to the outside air. Note that the feed steam H_{in} is actually defined using the outdoor temperature T_{ext} . Q_{dist} represents any heat flux that may act as a disturbance to the indoor temperature. The notation is summarized in Table 3.7.

The calculation of the amount of heat recovered in the HRV is, evidently, crucial for accurate modeling, but can be done in several ways and will be discussed later.

3.2.2 Time-scale Analysis

We begin by observing that there is an inherent and significant discrepancy in the parameters of the model (3.1-3.5). Namely, the heat capacity and density of the building structural elements are significantly higher than the heat capacity and density of the air circulated through the building. Thus, we can define:

$$\varepsilon_1 = \frac{\rho_{air} C_{p,air}}{\rho_m C_{p,m}}$$

and note that, based on the above observation, $\varepsilon_1 \ll 1$. We also capture the effect of energy recovery in the HRV on the system dynamics, and define

$$\varepsilon_2 = \frac{H_{in}^{ss}}{H_{rec}^{ss}}$$

as the ratio of steady state input to recycled air enthalpy,

$$\kappa_i = \frac{H_i^{ss}}{H_{rec}^{ss}}$$

as the ratio of steady state enthalpy of stream i to the steady state rate of heat recycling, and

$$\mu_i = \frac{H_i}{H_i^{ss}}$$

as the ratio of enthalpy of stream i to its steady state value.

Using this notation, the model (3.1-3.5) becomes:

$$\varepsilon_1 \frac{dT_z}{dt} = \frac{-U_m A_m (T_z - T_m) - Q_{dist} + \frac{1}{\varepsilon_2} H_{in}^{ss} (\kappa_f \mu_f - \kappa_z \mu_z)}{\rho_m C_{p,m} V_z} \quad (3.6)$$

$$\varepsilon_1 \frac{dT_f}{dt} = \frac{Q_{HVAC} + H_{mix} + \frac{1}{\varepsilon_2} H_{in}^{ss} (\kappa_c \mu_c - \kappa_f \mu_f)}{\rho_m C_{p,m} V_f} \quad (3.7)$$

$$\varepsilon_1 \frac{dT_c}{dt} = \frac{H_{in} + \frac{1}{\varepsilon_2} H_{in}^{ss} (-\mu_{rec} - \kappa_c \mu_c)}{\rho_m C_{p,m} V_c} \quad (3.8)$$

$$\varepsilon_1 \frac{dT_h}{dt} = \frac{-H_h - H_{mix} + \frac{1}{\varepsilon_2} H_{in}^{ss} (\kappa_z \mu_z + \mu_{rec})}{\rho_m C_{p,m} V_h} \quad (3.9)$$

$$\frac{dT_m}{dt} = \frac{U_m A_m (T_z - T_m) + U_{ext} A_{ext} (T_{ext} - T_m)}{\rho_m C_{p,m} V_m} \quad (3.10)$$

Owing to the presence of the small parameter ε_1 , the system (3.6-3.10) is in a singularly perturbed form [147]; more specifically, since ε_1 multiplies the derivatives

of a subset of state variables, the system is in a *standard* singularly perturbed form. Standard theory (see, e.g., [142]) indicates that the system will have a two time scale behavior, with the temperatures of the air in the different units, i.e., the vector $T_{air} = [T_z \ T_f \ T_c \ T_h]^T$ being the *fast* variables, and the temperature of the building structure, T_m , being the slow variable. Singular perturbation theory also allows us to derive separate models for the dynamics in each time scale. In the limit $\varepsilon_1 \rightarrow 0$, which physically corresponds to an infinitely large structure heat capacity (or an infinitely small air heat capacity), an expression for the slow dynamics (i.e., for the evolution of T_m) can be obtained in the form of a differential-algebraic equation (DAE) system

$$0 = \mathbf{f}_T(T_{air}, T_m) \quad (3.11)$$

$$\frac{dT_m}{dt} = \frac{U_m A_m (T_z - T_m) + U_{ext} A_{ext} (T_{ext} - T_m)}{\rho_m C_{p,m} V_m} \quad (3.12)$$

where, for convenience of notation, we used \mathbf{f}_T to denote the right-hand-side of (3.6), (3.7), (3.8) and (3.9). Equation (3.11) can be solved to obtain expressions for T_{air} , which can be subsequently substituted in the differential equation (3.12) to obtain a state-space realization (ODE representation) of the slow dynamics which are, evidently, associated with the thermal mass.

Subsequently, by defining the stretched, fast time variable $\tau_1 = t/\varepsilon_1$, and considering the same limit case $\varepsilon_1 \rightarrow 0$, we obtain

$$\frac{dT_z}{d\tau_1} = \frac{-U_m A_m (T_z - T_m) - Q_{dist} + \frac{1}{\varepsilon_2} H_{in}^{ss} (\kappa_f \mu_f - \kappa_z \mu_z)}{\rho_m C_{p,m} V_z} \quad (3.13)$$

$$\frac{dT_f}{d\tau_1} = \frac{Q_{HVAC} + H_{mix} + \frac{1}{\varepsilon_2} H_{in}^{ss} (\kappa_c \mu_c - \kappa_f \mu_f)}{\rho_m C_{p,m} V_f} \quad (3.14)$$

$$\frac{dT_c}{d\tau_1} = \frac{H_{in} + \frac{1}{\varepsilon_2} H_{in}^{ss} (-\mu_{rec} - \kappa_c \mu_c)}{\rho_m C_{p,m} V_c} \quad (3.15)$$

$$\frac{dT_h}{d\tau_1} = \frac{-H_h - H_{mix} + \frac{1}{\varepsilon_2} H_{in}^{ss} (\kappa_z \mu_z + \mu_{rec})}{\rho_m C_{p,m} V_h} \quad (3.16)$$

$$\frac{dT_m}{d\tau_1} = 0 \quad (3.17)$$

which represents a description of the *fast* (boundary layer) dynamics associated with the air temperature, when the slow variable T_m is at quasi-steady state.

Let us now focus on the effect of the HRV (and, in general, of energy *recovery and recycling*) on the dynamics of the building. To this end, recall that the term ε_2 , present in the above model of the fast dynamics, represents the ratio of the energy flow rate associated with the air stream to the energy flow within the HRV, i.e., the amount of energy recovered from the exhaust air. Based on the value of ε_2 , we can distinguish two specific cases:

- $\varepsilon_2 \gg 1$, i.e., very low energy recovery
- $\varepsilon_2 \ll 1$, i.e., significant energy recovery

The first case corresponds, in practice, to the absence of an HRV (in this case, its capital cost would not be justified by the small energy savings realized). The second case is in fact the case of practical interest because of very large potential energy savings. Note that a third, intermediate case (i.e., $\varepsilon_2 = \mathcal{O}(1)$) can also be discussed, which corresponds to achieving some modicum of energy recovery. This situation is typically not of practical interest because energy savings would be offset by capital expenditure.

Let us now reconsider the fast dynamics (Equations 3.13-3.16) in the high energy

recovery case. Owing to the presence of $\varepsilon_2 \ll 1$, the model (3.13-3.16) is in a *nonstandard* singularly perturbed form [147], having thus itself the potential to exhibit a two time scale response. Note that in this case κ_i and μ_i are $\mathcal{O}(1)$, so the rate at which energy is circulated throughout the system is the same magnitude as the rate at which energy is recycled ($\kappa_i\mu_i = \mathcal{O}(1)$). We proceed by considering the fast dynamics of this system (and, by extension, the *fastest* dynamics of the entire building). To this end, let us define the *fastest* time scale $\tau_2 = \tau_1/\varepsilon_2 = t/(\varepsilon_1\varepsilon_2)$, where the model becomes:

$$\frac{dT_z}{d\tau_2} = \frac{\varepsilon_2(-U_m A_m(T_z - T_m) - Q_{dist}) + H_{in}^{ss}(\kappa_f\mu_f - \kappa_z\mu_z)}{\rho_m C_{p,m} V_z} \quad (3.18)$$

$$\frac{dT_f}{d\tau_2} = \frac{\varepsilon_2(Q_{HVAC} + H_{mix}) + H_{in}^{ss}(\kappa_c\mu_c - \kappa_f\mu_f)}{\rho_m C_{p,m} V_f} \quad (3.19)$$

$$\frac{dT_c}{d\tau_2} = \frac{\varepsilon_2 H_{in} + H_{in}^{ss}(-\mu_{rec} - \kappa_c\mu_c)}{\rho_m C_{p,m} V_c} \quad (3.20)$$

$$\frac{dT_h}{d\tau_2} = \frac{\varepsilon_2(-H_h - H_{mix}) + H_{in}^{ss}(\kappa_z\mu_z + \mu_{rec})}{\rho_m C_{p,m} V_h} \quad (3.21)$$

and consider the limit case $\varepsilon_2 \rightarrow 0$, which corresponds to the (ideal and, evidently, highly desirable) case of an infinitely large rate of energy recycling. In this case, we obtain a description of the fastest dynamics of the air in the building:

$$\frac{dT_z}{d\tau_2} = \frac{(\kappa_f\mu_f - \kappa_z\mu_z)}{\rho_m C_{p,m} V_z} \quad (3.22)$$

$$\frac{dT_f}{d\tau_2} = \frac{(\kappa_c\mu_c - \kappa_f\mu_f)}{\rho_m C_{p,m} V_f} \quad (3.23)$$

$$\frac{dT_c}{d\tau_2} = \frac{(-\mu_{rec} - \kappa_c\mu_c)}{\rho_m C_{p,m} V_c} \quad (3.24)$$

$$\frac{dT_h}{d\tau_2} = \frac{(\kappa_z\mu_z + \mu_{rec})}{\rho_m C_{p,m} V_h} \quad (3.25)$$

Note that not all the quasi-steady state constraints,

$$0 = \kappa_f \mu_f - \kappa_z \mu_z \quad (3.26)$$

$$0 = \kappa_c \mu_c - \kappa_f \mu_f \quad (3.27)$$

$$0 = -\mu_{rec} - \kappa_c \mu_c \quad (3.28)$$

$$0 = \kappa_z \mu_z + \mu_{rec} \quad (3.29)$$

corresponding to the above equations are linearly independent; in effect, the first equation above can easily be obtained from the sum of the last three. This suggests that the fast dynamics (3.22-3.25) do not have an isolated equilibrium point; rather, they will feature an equilibrium subspace (manifold), in which a slower dynamic component evolves. We will refer to this dynamic component as the *intermediate* dynamics of the building. To obtain an explicit description of the intermediate dynamics, we consider the same limit case $\varepsilon_2 \rightarrow 0$ in the time scale τ_1 , and under the quasi-steady constraints corresponding to the fastest dynamic component, i.e., Equations (3.27-3.29):

$$\frac{dT_z}{d\tau_1} = \frac{1}{\rho_m C_{p,m} V_z} [-U_m A_m (T_z - T_m) - Q_{dist} + z_1] \quad (3.30)$$

$$\frac{dT_f}{d\tau_1} = \frac{1}{\rho_m C_{p,m} V_f} [Q_{HVAC} + H_{mix} + z_2] \quad (3.31)$$

$$\frac{dT_c}{d\tau_1} = \frac{1}{\rho_m C_{p,m} V_c} [H_{in} + z_3] \quad (3.32)$$

$$\frac{dT_h}{d\tau_1} = \frac{1}{\rho_m C_{p,m} V_h} [-H_h - H_{mix} - (z_1 + z_2 + z_3)] \quad (3.33)$$

where $z_1 = \lim_{\varepsilon_2 \rightarrow 0} \frac{1}{\varepsilon_2} (\kappa_f \mu_f - \kappa_z \mu_z)$, $z_2 = \lim_{\varepsilon_2 \rightarrow 0} \frac{1}{\varepsilon_2} (\kappa_c \mu_c - \kappa_f \mu_f)$, and $z_3 = \lim_{\varepsilon_2 \rightarrow 0} \frac{1}{\varepsilon_2} (-\mu_{rec} - \kappa_c \mu_c)$ are indeterminate (yet finite) algebraic variables. Equations (3.30-3.33),

along with the (linearly independent) constraints (3.27-3.29) constitute a DAE model of the intermediate dynamics of the building. This DAE system is of high index, as the algebraic variables z_i cannot be computed directly from the algebraic constraints¹. A single differentiation of the constraints is, however, sufficient to obtain an expression for \mathbf{z} , which can then be used to derive a state-space realization (ODE representation) of the intermediate dynamics. The dimensions of the differential equation system and algebraic constraints indicate that this ODE will be *one-dimensional*².

3.2.3 Energy Management and Control System Design

Remark 3.1. *The developments above demonstrate that the use of energy recovery in buildings (a highly desirable feature from an energy efficiency standpoint) is at the origin of a dynamic behavior with three distinct time scales; (1) the temperatures of the structural elements evolve over a long time horizon, (2) the air temperature in the building zone evolves in a fast time scale and, finally, (3) an intermediate dynamics emerges as a result of energy integration. It can be shown [129] that the intermediate dynamics are associated with the total energy content of the air in the building.*

¹The index of a DAE system is defined as the minimum number of times the algebraic equations must be differentiated in order to obtain an expression for time derivative of the algebraic variables \mathbf{z} [147, 142, 18].

²The dimensionality of the slow variables ζ is related to the difference between the number of states $\mathbf{x} \in \mathbb{R}^n$ and algebraic constraints $\mathbf{z} \in \mathbb{R}^p$. In this case, there are 4 states and 3 linearly independent algebraic constraints, and the dimension of the ODE representation of the slow dynamics will be at most one.

The design of controllers for systems with multiple time scale dynamics should be addressed separately for each time scale, an approach known as composite control [142]. Along with the developments above, this suggests that the control and energy management of buildings should consider three separate tiers of control action, addressing control objectives i) in each element of the HVAC system, ii) over an intermediate time horizon that concerns the overall energy content of the air and, iii) related to the temperature of the structural elements. Our main interest in this paper is to identify, in a generic fashion, the structure of an overall, building-level *energy management strategy*. To this end, we make the following observations:

- intuitively, the dynamics in each time scale are stable (with no energy input, the building will eventually reach thermal equilibrium with its environment).
- controlling the temperature of the structural elements is not of practical interest, which eliminates the need for designing a controller for the dynamics in the slowest time scale. However, the significant thermal inertia (reflected in the slow dynamics) of the structural elements should be accounted for.
- it may not be necessary to control the temperature in each element of the HVAC system (or in each room of a zone). The temperature/energy use in one of these elements may not have a significant impact on the *total* building energy use.

The above observations suggest that managing building energy use should consider the evolution of the total energy stored in the air mass of the building, as well as the energy stored in the building structural elements. The models of the intermediate and slow dynamics should be used as a basis for synthesizing model-

based control structures that accomplish these goals. To this end, the availability of explicit low-dimensional state-space realizations of the dynamics in each time scale provides significant flexibility for controller design, and allows for the use of both inversion- and optimization-based controllers.

3.2.4 Model reduction

Because the intermediate dynamic is associated with the total energy content of the air in the building, this dynamic component will impact energy consumption. It should be explicitly characterized and accounted for in the controller design. The procedure for deriving the state space representation of the intermediate dynamics of the nonstandard singularly perturbed system (3.6 - 3.10) is described in [18, 27]. It involves a nonlinear coordinate transformation $\mathbf{T}(\mathbf{x}) = [\tilde{T}_z \ \boldsymbol{\eta}]^T$, where \mathbf{x} includes the states in the DAE model, \tilde{T}_z represents the evolution of variable T_z in the intermediate time scale, and $\boldsymbol{\eta}$ the fast variables. Note that, in order to proceed with this step, an explicit expression (i.e., in terms of the temperatures T_{air}) of the energy flow rate H_{rec} must be provided. We use the Underwood approximation [261] to define the driving force for heat transfer in the HRV:

$$H_{rec} = U_{rec}A_{rec} \left(\frac{(T_c - T_z)^{1/3} + (T_{ext} - T_h)^{1/3}}{2} \right)^{1/2} \quad (3.34)$$

The Underwood approximation has better numerical properties than the log mean temperature difference (LMTD) $\frac{\Delta T_A - \Delta T_B}{\ln(\Delta T_A / \Delta T_B)}$ in this situation because of the potential to have $\Delta T_A = T_c - T_z = T_{ext} - T_h = \Delta T_B$, in which case the LMTD is indeterminate.

Subsequently, we compute an ODE representation of the zone temperature in the DAE model accounting for overall temperature dynamics in the building (see, e.g., [19] for details) as:

$$\frac{d\tilde{T}_z}{dt} = \frac{8\tilde{T}_z F_{in}^2 (\rho_{air} C_{p,air})^2 + U_{rec} A_{rec} (Q_{dist} + Q_{HVAC} + U_m A_m (T_m - \tilde{T}_z))}{\rho_{air} C_{p,air} (U_{rec} A_{rec} (V_c + V_{HVAC} + V_z) - 8F_{in} V_h \rho_{air} C_{p,air})} \quad (3.35)$$

This model is non-stiff and can be used as a basis for controller design.

3.3 Case Study

We consider the problem of managing HVAC energy use in the UT Thermal Facade Laboratory, which we represent using a detailed model comprised of the partial-differential equations (3.36 - 3.46) presented in section 3.3.1. A finite difference discretization of this model yields an ODE representation consisting of over 400 ordinary differential equations, which is evidently too large for use in practical online optimization calculations. In response to this challenge, we will design our control system based on an aggregate model of similar structure as the model discussed in Section 3.2.1, but featuring a more detailed (and hence, more realistic) representation of the thermal properties and dynamics the wall elements of the building. This model, which we refer to as a *granular* aggregate model, is described in section 3.3.2. Table 3.1 summarizes the models developed in this paper.

3.3.1 Detailed System Model

Figure 3.3 shows the details of the building structure and interactions with the outdoor environment. The south facing wall is a single window, and each remaining

Table 3.1: Model naming conventions

Name	Equations
Minimal aggregate model	3.1 - 3.5
Granular aggregate model	3.54 - 3.60
Reduced order minimal model	3.35 and 3.5
Reduced order granular model	3.74 and 3.58 - 3.60
Detailed full-order system model	3.36 - 3.46

wall (including the roof and floor) has an inner insulation and outer cement layer. The insulation is such that the majority of the heat transferred into the air space comes from sunlight through the window. The parameters defining the components of this model are in Tables 3.2, 3.3, 3.4, and 3.5. The sets

$$w \in Wall = \{North, East, West, Roof, Floor\}$$

$$win \in Win = \{South\}$$

$$s \in Surfaces = Wall \cap Win$$

are used in the detailed system description, presented in Equations (3.36 - 3.46). In order to simulate the detailed building system, the PDEs (3.38), (3.39), (3.43), (3.44) were each discretized using a finite-difference approach and, respectively, 100, 100, 10, and 30 nodes. The final model therefore includes over 400 states.

$$\frac{dT_z}{dt} = \frac{H_{fed} - H_z + \sum_{w \in W} U_w^{in} A_w (T_w^{m2} - T_z) + \sum_{win \in Win} U_{win}^{in} A_{win} (T_{win} - T_z)}{\rho_{air} C_{p,air} V_z} \quad (3.36)$$

$$\frac{dT_f}{dt} = (H_c + H_{mix} - H_{fed} + Q_{HVAC}) / (\rho_{air} C_{p,air} V_{HVAC}) \quad (3.37)$$

$$\frac{\partial T_c}{\partial t} = -\nu_x \frac{\partial T_c}{\partial x} + \frac{U_{rec} A_{rec}}{\rho C_p V_c} (T_c - T_h) \quad (3.38)$$

$$\frac{\partial T_h}{\partial t} = -\nu_x \frac{\partial T_h}{\partial x} + \frac{U_{rec} A_{rec}}{\rho C_p V_c} (T_h - T_c) \quad (3.39)$$

$$T_c|_{x=0} = T_{ext} \quad (3.40)$$

$$T_h|_{x=l} = T_z \quad (3.41)$$

$$\frac{dT_{win}}{dt} = \frac{\alpha_{win} I_{win}^{out,tot} + I_{win}^{out,lw} - U_{win}^{in} A_{win} (T_{win} - T_z) - U_{win}^{out} A_{win} (T_{win} - T_{ext})}{\rho_{win} C_{p,win} V_{win}} \quad (3.42)$$

$\forall win \in Win$

$$\frac{\partial T_w^{m1}}{\partial t} = \left(U_w^{m1} \frac{\partial^2 T_w^{m1}}{\partial x^2} \right) / (\rho_w C_{p,w} V_w)^{m1} \quad (3.43)$$

$$\frac{\partial T_w^{m2}}{\partial t} = \left(U_w^{m2} \frac{\partial^2 T_w^{m2}}{\partial x^2} \right) / (\rho_w C_{p,w} V_w)^{m2} \quad (3.44)$$

$$\frac{dT_w^{m1}}{dt} |^{out} = \alpha_w^{m1} I_w^{out,tot} + I_w^{out,lw} - U_w^{out} A_w (T_w^{m1} - T_{ext}) \quad \forall w \in W \quad (3.45)$$

$$\frac{dT_w^{m2}}{dt} |^{in} = \alpha_w^{m2} I_w^{in,beam} + I_w^{in,fraction} - U_w^{in} A_w (T_w^{m2} - T_z) \quad \forall w \in W \quad (3.46)$$

The main natural disturbances to the building environment include changes in outdoor temperature and solar radiation, which is summarized in the following set of equations that define the remaining parameters in the system (3.36 - 3.46). Based on the calculation methods in EnergyPlus, in this model all beam radiation transmitted by the window is assumed to fall on the floor. The diffuse radiation transmitted is distributed among the interior surfaces by area. Thermal radiation between the interior surfaces is negligible, but the model does account for interactions between the exterior surfaces and the sky and ground (Equations 3.52 and 3.53) assuming a view factor of 0.5. The values of $I_s^{out,diff}$, $I_s^{out,beam}$, $I_{win}^{diff\ transmitted}$, $I_{win}^{beam\ transmitted}$, T_{sky} , and T_{ground} may be obtained over any time period using an EnergyPlus model and appropriate TMY data.

$$I_s^{out,tot} = (I_s^{out,diff} + I_s^{out,beam}) A_s \quad (3.47)$$

$$I_w^{in,diff} = I_{win}^{diff\ transmitted} \frac{\alpha_w^{m2} A_w}{\sum_w (\alpha_w^{m2} A_w) + (\tau_{win} A_{win})} \quad (3.48)$$

$$I_w^{in,beam} = 0 \quad \forall w \neq Floor \in Wall \quad (3.49)$$

$$I_{Floor}^{in,beam} = I_{win}^{beam\,transmitted} A_{Floor} \quad (3.50)$$

$$I_w^{out,lw} = I_{w \rightarrow sky}^{lw} + I_{w \rightarrow ground}^{lw} \quad (3.51)$$

$$I_{w \rightarrow sky}^{lw} = \frac{1}{2} \varepsilon_w \varepsilon_{sky} A_w \sigma (T_w^4 - T_{sky}^4) \quad (3.52)$$

$$I_{w \rightarrow ground}^{lw} = \frac{1}{2} \varepsilon_w \varepsilon_{ground} A_w \sigma (T_w^4 - T_{ground}^4) \quad (3.53)$$

Table 3.2: HVAC Properties

par.	val.	units
F_{in}	0.11	m^3/s
V_C	0.08	m^3
V_H	0.08	m^3
V_{HVAC}	0.7	m^3
$U_{rec} A_{rec}$	20	W/K
$\rho_{air} C_{pair}$	1214	$J/m^3 K$

Table 3.3: Wall Dimensions

par.	val.	units
A_N	12.96	m^2
$A_{E,W}$	10.51	m^2
$A_{R,F}$	14.41	m^2

Table 3.4: Wall Material Properties

Property	Outside (m1)	Inside (m2)	Units
ρC_p	2.16E6	5.5E5	J/m^3K
U	1	0.173	W/m^2K
U_N	2	7	W/mK
U_E	2.4	9.3	W/mK
U_W	2.1	7	W/mK
U_R	1.3	10.6	W/mK
U_F	3.2	2.9	W/mK
<i>thickness</i>	0.013	0.244	<i>m</i>
α	0.7	0.7	
ε	0.95		

Table 3.5: Window Properties

par.	val.	units
U^{in}	1.7	W/mK
U^{out}	6.3	W/mK
A	12.96	m^2
<i>thickness</i>	0.0057	<i>m</i>
ρC_p	2.1E6	J/m^2K
α	0.16	
τ	0.77	

Table 3.6: Estimated parameters used in the granular model

Parameter	Value (W/K)
$U_{rec}A_{rec}$ (coarse model)	20
$U_{rec}A_{rec}$ (reduced coarse model)	20E3
$U_w A_w \forall w \in W$	0.173

Table 3.7: Nomenclature

Variable	Description
T	Temperature
H	Enthalpy
ρ	Density
C_p	Heat capacity
V	Volume
Q_{HVAC}	Cooling load
I	Solar radiation
α	Absorption coefficient
τ	Transmission coefficient
U	Heat transfer coefficient
Subscript	Description
w	Wall
win	Window
z	Zone
f	Zone feed
c	HRV feed leg
h	HRV exhaust leg
in	Inlet stream
mix	Air recycle stream
rec	Energy recycle
Superscript	Description
$m1$	Material 1 in wall w
$m2$	Material 2 in wall w
ss	Steady state
in	Interior wall or window surface
out	Exterior wall or window surface
tot	Combined diffuse and beam solar radiation
$beam$	Direct (beam) solar radiation
$diff$	Diffuse solar radiation
lw	Long wave radiation

3.3.2 Granular Aggregate Model

The main difference between the simple prototype model used for our theoretical derivations in Section 3.2.1 and the granular model proposed in this case study is that the granular model uses a separate thermal mass to represent each material in every wall shown in Figure 3.3. This extension of the minimal model provides a more accurate representation of the different wall layer materials. In the model equations (3.54-3.60), two thermal mass elements ($m1$ and $m2$) are modeled for each wall $w \in W = \{North, East, West, Roof, Floor\}$. The south window is assumed to have a uniform temperature because it is a single glass pane, and is labeled $win = \{South\}$.

The physical properties of each thermal mass are based on the construction of the physical system discussed in the Appendix.

$$\frac{dT_z}{dt} = \frac{H_f - H_z + \sum_{w \in W} U_w^{in} A_w (T_w^{m2} - T_z) + U_{win}^{in} A_{win} (T_{win} - T_z)}{\rho_{air} C_{p,air} V_z} \quad (3.54)$$

$$\frac{dT_f}{dt} = (H_c + H_{mix} - H_f + Q_{HVAC}) / (\rho_{air} C_{p,air} V_{HVAC}) \quad (3.55)$$

$$\frac{dT_c}{dt} = (H_{in} - H_c - H_{rec}) / (\rho_{air} C_{p,air} V_c) \quad (3.56)$$

$$\frac{dT_h}{dt} = (H_z - H_h - H_{mix} + H_{rec}) / (\rho_{air} C_{p,air} V_h) \quad (3.57)$$

$$\frac{dT_{win}}{dt} = \frac{\alpha_{win} I_{win}^{out,tot} + I_{win}^{out,lw} - U_{win}^{in} A_{win} (T_{win} - T_z) - U_{win}^{out} A_{win} (T_{win} - T_{ext})}{\rho_{win} C_{p,win} V_{win}} \quad (3.58)$$

$$\frac{dT_w^{m1}}{dt} = \frac{\alpha_w^{m1} I_w^{out,tot} + I_w^{out,lw} - U_w^{out} A_w (T_w^{m1} - T_{ext}) - U_w A_w (T_w^{m1} - T_w^{m2})}{\rho_w^{m1} C_{p,w}^{m1} V_w^{m1}} \quad \forall w \in W \quad (3.59)$$

$$\frac{dT_w^{m2}}{dt} = \frac{\alpha_w^{m2} I_w^{in,beam} + I_w^{in,diff} - U_w^{in} A_w (T_w^{m2} - T_z) + U_w A_w (T_w^{m1} - T_w^{m2})}{\rho_w^{m2} C_{p,w}^{m2} V_w^{m2}} \quad \forall w \in W \quad (3.60)$$

I denotes various types of radiation (either solar or thermal) at each surface and the calculation of these values is described in detail in the Appendix. U_w^{in} and U_w^{out} represent convective heat transfer coefficients for the inside and outside surfaces of each wall, and U_w is the coefficient for heat transfer between the two thermal mass elements in each wall.

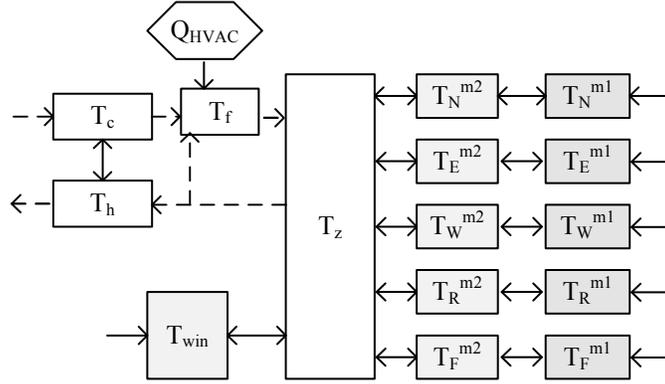


Figure 3.5: Granular aggregate model. Dashed lines represent mass flow while solid lines represent heat transfer. The enthalpy labels are not shown but are identical to those in the minimal aggregate model.

All model parameters are based on the building structure and are defined in the Appendix. Two parameters in the minimal and associated reduced model that cannot be specified based on the physical properties of the detailed system are $U_{rec}A_{rec}$ and U_wA_w , which represent the conduction coefficients and area for, respectively, the legs of the heat recovery device and the two wall layers. One method for the identification of these values would involve a rigorous parameter

estimation procedure using data measured in a real building or obtained from an EnergyPlus simulation. For this work the parameters were established by manually comparing the open loop behavior of the lumped and detailed system models. It is important to note that the value of $U_{rec}A_{rec}$ in the coarse and reduced coarse models is not the same. This is a result of the model reduction procedure, and can be attributed to the fact that the state corresponding to air temperature in the reduced model accounts for all of the air in the system, not just the air in the zone. Consequently, the value of $U_{rec}A_{rec}$ in the reduced model must be increased. The values of these additional parameters are summarized in Table 3.6.

3.3.3 Analysis and Model Reduction

Following the results in section 3.2.1 for the minimal model, we extend the analysis and model reduction for the granular model with several thermal mass elements. In this case the choice of the perturbation parameter ε_1 is not as straightforward because there is not a *single* mass to compare to the thermal inertia of air.

Let

$$\varepsilon_1^k = \frac{\rho_{air}C_{p,air}}{\rho_k C_{p,k}} \ll 1 \quad (3.61)$$

where $k \in \kappa$ is the set of indices for all thermal mass elements directly connected to the zone (corresponding to states labeled $T_{w \in W}^{m2}$ and T_{win} in Figure 3.5) and

$$\varepsilon_1^g = \frac{\rho_{air}C_{p,air}}{\rho_g C_{p,g}} \ll 1 \quad (3.62)$$

where $g \in \gamma$ is the set of indices for all thermal mass elements in contact with the outside air and the interior masses (i.e., states labeled $T_{w \in W}^{m1}$ in Figure 3.5). We

will show that as long there are only small differences in the material properties of each mass, the system dynamics will exhibit the same three time scales identified in section 3.2.2.

We begin with the assertion that the materials for every structural element have sufficiently similar properties so that we can define the $\mathcal{O}(1)$ quantities δ , such that

$$\text{Comparison between inner materials: } \delta_{k_1 k_2} = \frac{\varepsilon_1^{k_1}}{\varepsilon_1^{k_2}} = \frac{\rho_{k_1} C_{p,k_1}}{\rho_{k_2} C_{p,k_2}} = \mathcal{O}(1) \quad (3.63)$$

$$\text{Comparison between inner and outer materials: } \delta_{gk} = \frac{\varepsilon_1^g}{\varepsilon_1^k} = \frac{\rho_g C_{p,g}}{\rho_g C_{p,g}} = \mathcal{O}(1) \quad (3.64)$$

We designate ε_1^K with $K \in \kappa$ as the desired reference structural element for the entire system (*i.e.*, we wish to write all the model equations in terms of this single perturbation parameter). Using

$$\rho_k C_{p,k} = \frac{\rho_{air} C_{p,air}}{\varepsilon_1^k} = \frac{\rho_{air} C_{p,air}}{\varepsilon_1^K} \frac{1}{\delta_{kK}} \quad (3.65)$$

$$\rho_g C_{p,g} = \frac{\rho_{air} C_{p,air}}{\varepsilon_1^g} = \frac{\rho_{air} C_{p,air}}{\varepsilon_1^K} \frac{1}{\delta_{gK}} \quad (3.66)$$

we can rewrite Equations (3.54) - (3.60) as

$$\begin{aligned} \frac{dT_z}{dt} = & \frac{\sum_w U_w^{in} A_w (T_w^{m2} - T_z) + U_{win}^{in} A_{win} (T_{win} - T_z)}{\rho_{air} C_{p,air} V_z} + \\ & \frac{\frac{H_{in,ss}}{\varepsilon_2} (\kappa_f \mu_f - \kappa_z \mu_z)}{\rho_{air} C_{p,air} V_z} \end{aligned} \quad (3.67)$$

$$\frac{dT_f}{dt} = \left(H_{mix} + Q_{HVAC} + \frac{H_{in,ss}}{\varepsilon_2} (\kappa_c \mu_c - \kappa_f \mu_f) \right) / (\rho_{air} C_{p,air} V_{HVAC}) \quad (3.68)$$

$$\frac{dT_c}{dt} = (H_{in} + \frac{H_{in,ss}}{\varepsilon_2} (-\mu_{rec} - \kappa_c \mu_c)) / (\rho_{air} C_{p,air} V_c) \quad (3.69)$$

$$\frac{dT_h}{dt} = (-H_h - H_{mix} + \frac{H_{in,ss}}{\varepsilon_2}(\kappa_z \mu_z + \mu_{rec})) / (\rho_{air} C_{p,air} V_h) \quad (3.70)$$

$$\frac{1}{\varepsilon_1^K} \frac{1}{\delta_{winK}} \frac{dT_{win}}{dt} = \frac{\alpha_{win} I_{win}^{out,tot} + I_{win}^{out,lw}}{\rho_{air} C_{p,air} V_{win}} + \frac{-U_{win}^{in} A_{win} (T_{win} - T_z) - U_{win}^{out} A_{win} (T_{win} - T_{ext})}{\rho_{air} C_{p,air} V_{win}} \quad (3.71)$$

$$\frac{1}{\varepsilon_1^K} \frac{1}{\delta_{w^{m1}K}} \frac{dT_w^{m1}}{dt} = \frac{\alpha_w^{m1} I_w^{out,tot} + I_w^{out,lw}}{\rho_{air} C_{p,air} V_w^{m1}} + \frac{-U_w^{out} A_w (T_w^{m1} - T_{ext}) - U_w A_w (T_w^{m1} - T_w^{m2})}{\rho_{air} C_{p,air} V_w^{m1}} \quad \forall w \in W \quad (3.72)$$

$$\frac{1}{\varepsilon_1^K} \frac{1}{\delta_{w^{m2}K}} \frac{dT_w^{m2}}{dt} = \frac{\alpha_w^{m2} I_w^{in,beam} + I_w^{in,diff}}{\rho_{air} C_{p,air} V_w^{m2}} + \frac{-U_w^{in} A_w (T_w^{m2} - T_z) + U_w A_w (T_w^{m1} - T_w^{m2})}{\rho_{air} C_{p,air} V_w^{m2}} \quad \forall w \in W \quad (3.73)$$

Multiplying by ε_1^K results in a system of equations in the same form as (3.6-3.10), where the perturbation parameter multiplies the derivative of the fast variables T_{air} and the remaining derivatives of the thermal mass element temperatures are multiplied by $\mathcal{O}(1)$ quantities and therefore constitute slow variables in the system. Comparing (3.6) to (3.67) we see that the main difference is the addition of convective heat transfer terms to the system, which will appear collectively in the reduced order model, shown below. Following the same methods used to derive Equation (3.35), the Underwood approximation (3.34) was again used to model heat recovery in the HRV and an ODE representation of the DAE model associated with the intermediate dynamics of Equations (3.67 - 3.73) was computed as:

$$\begin{aligned} \frac{d\tilde{T}_z}{dt} = & [8\tilde{T}_z F_{in}^2 (\rho_{air} C_{p,air})^2 + U_{rec} A_{rec} (Q_{dist} + Q_{HVAC} + \sum_w U_w^{in} A_w (T_w^{m2} - \tilde{T}_z) \\ & + U_{win}^{in} A_{win} (T_{win} - \tilde{T}_z))] / \text{DEN} \end{aligned} \quad (3.74)$$

$$\text{DEN} = \rho_{air} C_{p,air} (U_{rec} A_{rec} (V_c + V_{HVAC} + V_z) - 8F_{in} V_h \rho_{air} C_{p,air})$$

Equations (3.74) and (3.35) differ only in the number of terms in the numerator that relate to the structural elements. This result is significant because for any single air zone, the reduced order model for the intermediate air temperature dynamics will have the same form regardless of the coarse-graining approach used to construct the lumped parameter model of the building structure. Also, while the reduction in the number of states in the model is less significant in this case (there are still a significant number of ODEs representing the thermal mass elements), the equation lacks the multiple time scale behavior that contributes to the stiffness of the original model and is therefore well suited for online control and optimization calculations. This point is demonstrated in the following section.

3.3.4 Model Stiffness Analysis

Designing inversion or optimization-based controllers using stiff models can be problematic [18]. To demonstrate the benefits of the proposed model reduction and time scale decomposition methodology we have linearized the detailed, granular, and reduced granular models about a common operating point in order to calculate the eigenvalues and condition number of each system. This calculation was performed assuming constant outdoor air temperature and solar radiation on the surfaces and no occupants in the zone. The eigenvalues of the system matrices are shown in Figure 3.6 and the condition numbers of the actual, granular, and reduced order granular systems are, respectively, $1.58E7$, $6.08E7$ and $1.72E3$. We make the following observations:

- The condition number of the full-order model system is slightly less than that of the coarse model. While the smallest eigenvalue is the same for both, the eigenvalues associated with slower states are larger in the full-order system than the granular model. This is likely a result of the discretization of the PDEs describing temperature in the wall materials.
- The condition number of the reduced order granular model is significantly less than that of the granular model. This is a natural result of the time scale decomposition and can be expected to yield significant benefits in controller design. Evidently, the use of this model in controller synthesis is preferred.
- The eigenvalues were calculated at several operating points with very similar results. This supports the fact that the stiffness of the system is a structural property and not a function of the operating point.

3.3.5 Energy Management

The reduced order granular model (3.74) was used to formulate the energy management problem as a moving-horizon optimal control problem, following the economic model-predictive control paradigm (see, e.g., [226]):

$$\min_{Q_{HVAC}} J = \int_{t_0}^{t_0+NM} C(t)Q_{HVAC}(t)dt \quad (3.75)$$

s.t. reduced order granular model equations

$$\begin{aligned} T_z^{min} &\leq \tilde{T}_z \leq T_z^{max} \\ Q_{HVAC}^{min} &\leq Q_{HVAC} \leq Q_{HVAC}^{max} \\ \left(\frac{dQ_{HVAC}}{dt}\right)^{min} &\leq Q_{HVAC}(t+M) - Q_{HVAC}(t) \leq \left(\frac{dQ_{HVAC}}{dt}\right)^{max} \end{aligned} \quad (3.76)$$

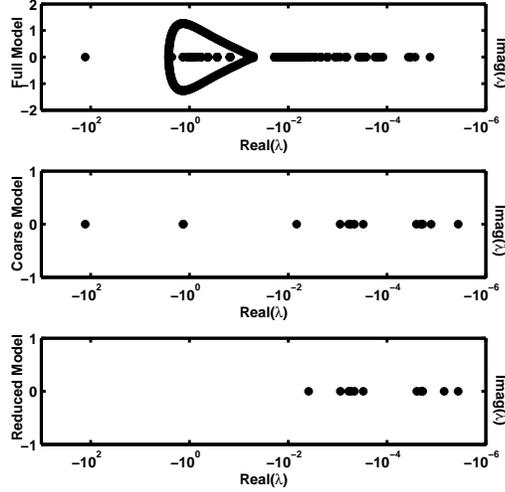


Figure 3.6: Model eigenvalues for 404-dimensional full-order model, the 15-dimensional granular aggregate model, and the 12-dimensional reduced-order aggregate model

The optimization objective, shown in Equation (3.75), is the total cost of energy input to the AHU over the prediction horizon, where M is the sampling rate and N is the number of samples taken over the prediction horizon. $C(t)$ represents the unit cost of energy. The constraints include bounds on the zone temperature, cooling load, and the rate of change of the cooling load.

We will compare the performance of this economic objective to a temperature tracking objective (3.77) subject to the same constraints (3.76)³.

$$\min_{Q_{HVAC}} J = \int_{t_0}^{t_0+NM} (T_z - T_z^{SP})^2 dt \quad (3.77)$$

³Note that both economic and tracking terms could easily be included in a single MPC weighted objective [27]

In order to account for the inherent plant-model mismatches, we extended the model with a disturbance term \mathbf{d} (see, e.g., [204, 285]), defined as the time integral of the discrepancy between the state predicted by the reduced model and the corresponding state measured from the plant:

$$\dot{\mathbf{x}} = \begin{bmatrix} \dot{\boldsymbol{\theta}} \\ \dot{\mathbf{d}} \end{bmatrix} = \begin{bmatrix} \mathbf{f}(\boldsymbol{\theta}) + \mathbf{D}\mathbf{d} \\ \boldsymbol{\theta}_{measured} - \boldsymbol{\theta} \end{bmatrix} \quad (3.78)$$

where $\boldsymbol{\theta}$ is the reduced granular model and \mathbf{D} is a matrix of weight parameters. This approach can be construed as providing the NMPC controller with *integral action*; however, given the nature of the moving horizon formulation (3.75), the role of the extended model is to eliminate plant-model mismatch rather than to provide offset-free tracking. In the following simulations only the zone temperature is measured and \mathbf{D} is defined below. It has the same number of rows and columns as the number of states in the reduced granular model (3.74 and 3.58 - 3.60), and the nonzero elements correspond to states that are also measured in the actual system.

$$\mathbf{D} = \begin{bmatrix} 10^{-5} & [0]_{1 \times 11} \\ [0]_{11 \times 1} & [0]_{11 \times 11} \end{bmatrix} \quad (3.79)$$

3.3.6 Results and Discussion

The nonlinear MPC (NMPC) controller described in the previous section is used to control the full-order system described in the Appendix, which consists of a set of PDEs discretized using a finite-difference approach to yield over 400 equations. Disturbances to the system include fluctuations in outside temperature and solar radiation on each surface. These profiles are shown in Figure 3.7 for a typical

August day in Austin, Texas, and are part of TMY2 weather data tabulated using EnergyPlus. There is an increase in energy price from 3pm - 6pm, shown in Figure 3.8, which is representative of peak market prices in Austin in the summer months. The controller has complete knowledge of the weather and pricing information, and its ability to deal with incorrect forecasts is evaluated through the use of unmeasured changes in the occupant profiles. Three different occupancy patterns are considered and shown in Figure 3.9. In the first case there are no occupants so we will directly observe the effect of the cost increase. The next two cases both have 2 occupants (corresponding to a 200W heat load) entering the zone, and in the third case the occupancy forecast is incorrect.

The NMPC settings when using the economic objective are: $M = 20min.$, $N = 24$ (i.e., the prediction horizon is 8hrs long), $Q_{HVAC}^{max} = 0W$, $Q_{HVAC}^{min} = -3000W$, $T_z^{min} = 21^\circ C$, $T_z^{max} = 26^\circ C$. Under the tracking objective, the bounds are tightened to $T_z^{min} = 23^\circ C$, $T_z^{max} = 25^\circ C$ and $T_z^{SP} = 24^\circ C$.

We evaluated the proposed energy management strategy through simulations. Figure 3.10a shows the zone temperature subject to the three occupant patterns when using the proposed economic strategy and the tracking NMPC.

Owing to the availability of accurate weather forecasts, the controller resorts to precooling, anticipating impending disturbances to the system regardless of the occupant scenario or objective function. In Figure 3.10a(a), the proposed energy management strategy results in a significant decrease in the zone temperature between hours 7 – 15, a measure taken to counteract the large increase in energy costs in the early-afternoon hours. For the tracking case, temperature is always

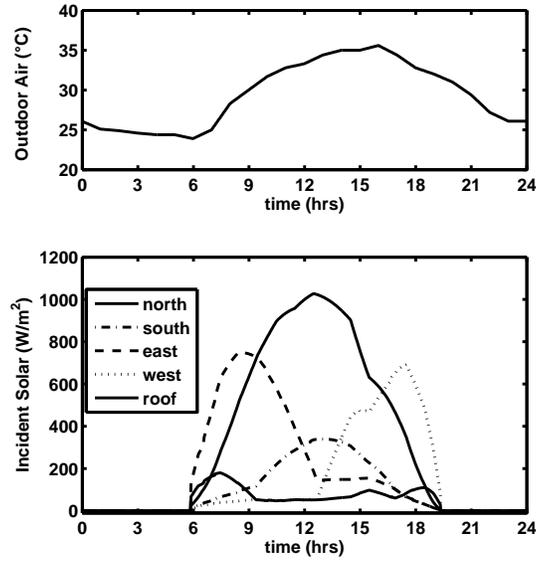


Figure 3.7: Weather Profile for August 1

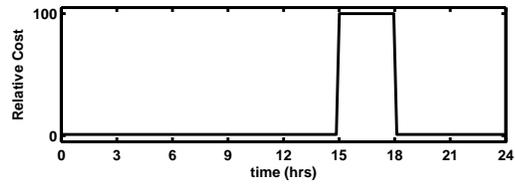


Figure 3.8: Energy Cost

well within the specified bounds.

For the economic case, the increased energy consumption and operating cost for this time interval is compensated for by the cost savings incurred during hours 15 – 18. We note that the overall cost savings compared to the tracking case are negligible for the studies shown here (the maximum savings were in the case with no occupants and were approximately 2.5%, and this figure depends on the price structure and setpoint value). This is because the thermal mass in the Thermal

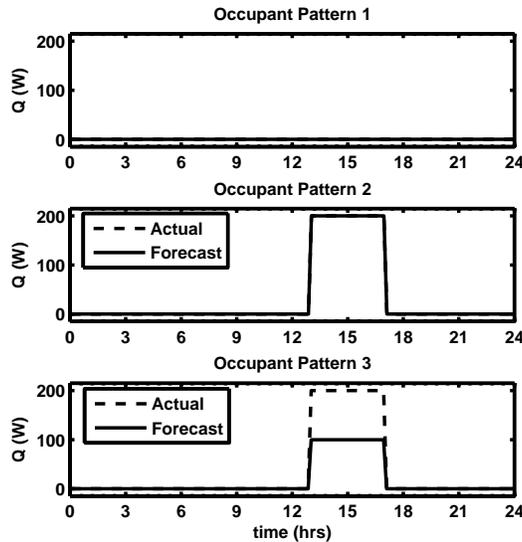


Figure 3.9: Occupancy Disturbances

Façade Laboratory is relatively small, so in spite of the precooling action the controller cannot significantly reduce the cooling load during the peak cost period by storing refrigeration in the building walls (see Figure 3.11). It is therefore reasonable to expect a larger cost savings and shift in energy demand in residential homes or larger commercial buildings. Regardless of the savings incurred, this example demonstrates the ability of economic NMPC to shift peak loads. The reduced-order model used in the optimization calculation enabled the NMPC actions to be computed within the controller sample time. We recognize that the continuous nature of the HVAC system modeled here is not representative of HVAC systems that operate only on an on/off basis; the energy management problem in the presence of such devices involves the formulation and solution of problem (3.75 - 3.76) in a mixed-integer context, and will be the subject of our future work.

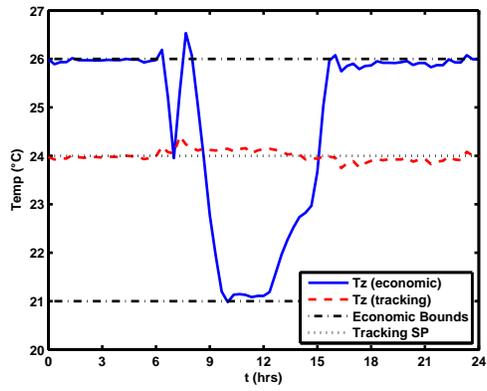
We also note from the simulation results that there are limited time intervals where the constraints on the zone temperature are violated. We attribute this to the inherent plant-model mismatch (e.g., Figure 3.12, where we note that the model predictions are still within the comfort bounds when the full-order system slightly deviates from them). For the cases with occupants entering the zone, when the model receives an incorrect occupant forecast, the closed-loop behavior is similar to the case when there are no occupants (Figure 3.10(c) is more similar to 3.10(a) than 3.10(b)). When the correct forecast is provided there is an even longer precooling period in the economic case because of the anticipated larger heat inputs to the system.

3.4 Conclusions

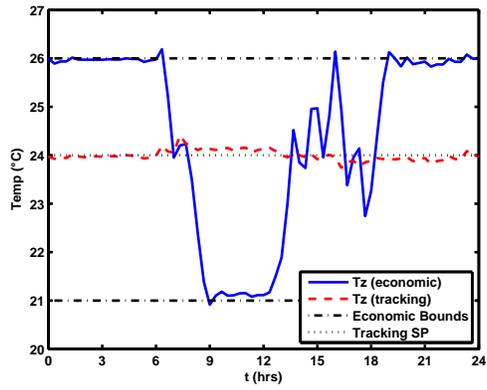
In this paper, we have addressed the building energy management problem through the use of model-based predictive control, and focused on buildings which are equipped with energy recovery devices. Our initial focus has been on modeling, where we used singular perturbation theory to demonstrate that the dynamics of buildings with energy recovery evolve over three distinct time scales: a slow time scale, associated with the temperature of the structural elements, a fast time scale, associated with the air temperatures in the building and HVAC system units, and an intermediate dynamics that emerges as a result of the use of energy recovery and involves the total energy stored in the air present in the entire building and HVAC system. We have shown that this is true for a single zone and air-loop, regardless of the number of structural elements involved in the model. Singular

perturbation arguments were then used to derive reduced-order models for the dynamics in each time scale of a residential (single-zone) building. These models are physically meaningful, and we are able to draw from the building design data to obtain values for the model parameters.

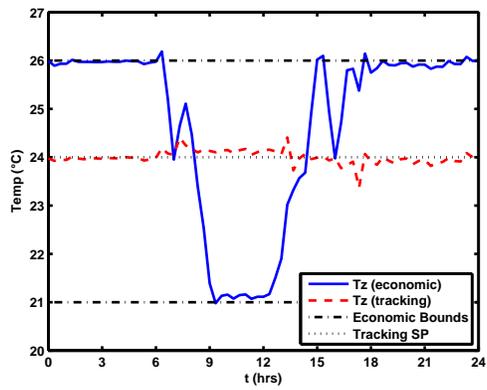
Economic and tracking objectives in an NMPC environment using the reduced-order models were compared in simulations subject to different occupancy profiles and forecasts. The single-zone building studied is representative of the University of Texas Thermal Façade Laboratory. As expected, cost savings, though small, and the ability to shift demand were seen under the economic objective. Most importantly, the calculations are easily performed in real-time using the reduced model. The derivations in the paper suggest that this model and associated control structure are applicable to a broader palette of single zone buildings, which we consider to be a very encouraging result towards the rapid and cost-effective deployment of model-based energy management approaches for small, residential buildings.



(a) Occupant Pattern 1



(b) Occupant Pattern 2



(c) Occupant Pattern 3

Figure 3.10: Zone Temperature Profiles

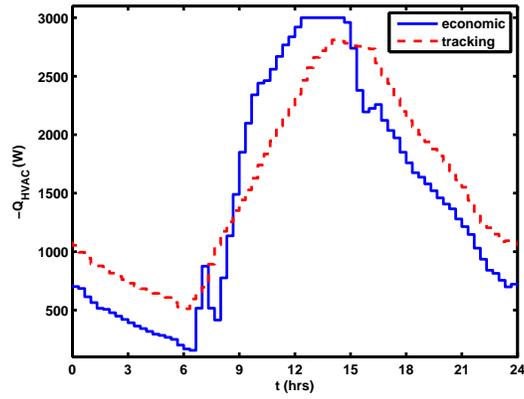


Figure 3.11: AHU Power for Occupant Profile 1

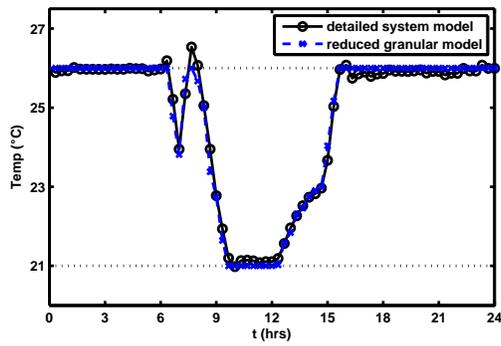


Figure 3.12: Plant-Model Mismatch Using the Economic Objective for Occupant Profile 1

Chapter 4

Hierarchical Scheduling and Control Strategies for Thermal Energy Storage Systems

The material in this chapter has been published in [252, 260, 256]

4.1 Introduction

Over 70% of the electricity use in the US can be attributed to commercial and residential buildings [267]. In addition to being large electricity consumers, buildings use power at a variable rate over the course of a day. In general, demand from buildings exhibits a peak in the afternoon hours. Electricity producers and grid operators typically use spare generation capacity (often referred to as “peaking plants”) to meet this temporary demand rise, often incurring additional economic and environmental costs [240]. Demand response (DR) strategies aim to alter the electricity demand pattern of buildings [4], typically focusing on “load leveling,” i.e., lowering peak demand and distributing power requirements more evenly during the day. Demand Response initiatives incentivize users to shift their demand to off-peak times via tiered energy price structures, which typically entail charging higher rates during the peak period (see the example in Figure 4.1 for real-time prices in the ERCOT market). In this context, energy storage technologies are

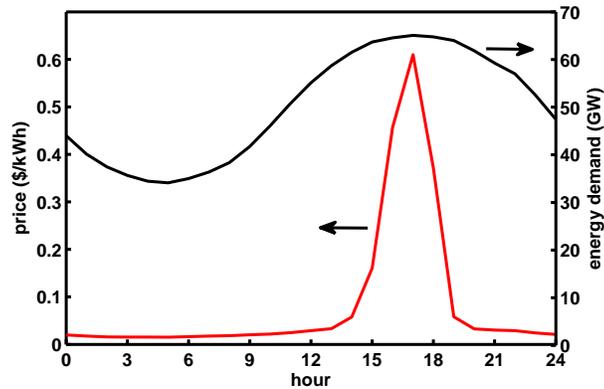


Figure 4.1: Energy Reliability Council of Texas (ERCOT) demand and day ahead settlement point prices for June 25, 2012 (Source: www.ercot.com)

key to enabling HVAC users to change their demand patterns without sacrificing comfort [248].

Energy storage is an important component of peak shifting and load leveling strategies: energy can be generated in excess of demand and stored during off-peak hours, and used to supplement the generation capacity during peak times. In this manner, generators can operate at or close to their peak efficiency. Evidently, the use of energy storage for load leveling purposes is dependent on the availability of an energy storage facility of adequate capacity. At present, grid-level installations of storage batteries are cost-prohibitive, and efforts have concentrated on using storage to modulate the energy use of large consumers. Since HVAC systems account for over 20% of the energy use of buildings [267], and methods for HVAC control and building temperature regulation remain fairly basic and are centered on simple linear feedback control [233], HVAC energy management has become a primary target for load leveling initiatives.

Thermal energy storage (TES) technologies provide a viable and cost-effective means of shifting electricity demands for HVAC loads. In particular, residential and commercial buildings offer multiple modes of thermal energy storage. On the one hand, structural elements (walls, foundations and foundation slabs, floors, etc.) in some building designs constitute a significant thermal mass that can be used to retain and release energy [253, 166, 234, 38]. Additionally, heating, ventilation and air conditioning (HVAC) systems can be designed or retrofitted with tanks that store cooling or heating agents (e.g., chilled or hot water) to help improve operating efficiency and flexibility [117, 66]. Notably in the case of a predominantly cooling climate (i.e., in hot weather), building HVAC systems can be operated in a pre-cooling mode, where the structure is cooled beyond the usual limit during the (morning) off-peak hours; in turn, this allows the HVAC system to operate at lower intensity during the peak (afternoon) hours. Storage tanks are also typically charged at night, and the chilled water is used to supplement HVAC operation at peak times.

Maximizing the load-leveling benefits afforded by building thermal energy storage does, however, require a close coordination of the charge/discharge events with external factors such as grid load (which, in turn, dictates energy prices in real-time price structures), weather, and building occupancy. This, in turn, calls for the development of advanced optimization-based control strategies, which are capable of orchestrating the operation of the HVAC system and TES elements under the uncertainty provided by the aforementioned factors.

In this chapter, we introduce a framework for developing such energy manage-

ment strategies. Both “active” (e.g., storage tanks) and “passive” (e.g., building structural elements) TES systems are considered. Regarding active storage, we present case studies involving cold water and latent energy storage using phase change material (PCM). PCM-based systems are advantageous due to their compact size [237]. However, their operation is complicated by the nonlinear behavior induced by the melting/solidification cycles [6], and the development of strategies for optimally integrating active PCM TES systems in building operations remains a pressing and open problem since the majority of the literature is focused on the design and control of passive PCM elements incorporated into the building structure [237, 180].

4.2 Background: Optimal Control Strategies for Buildings

Engaging HVAC systems in DR strategies involves operating the system at a higher level than actually needed during the off-peak hours, and storing the excess energy (heating or refrigeration) for use during peak times. Since storing energy incurs some losses, load leveling carries a penalty in terms of increased overall energy consumption. Incentives for implementing DR strategies are often provided in the form of time-of-use (TOU) pricing, where customers are subject to a time-sensitive rate structure similar to the market price for electricity, with energy prices typically much higher at peak times than in off-peak periods.

Shifting HVAC electricity demands in a TOU pricing framework calls for the use of an advanced control system, which, i) can account for the dynamic prices and minimize operating costs through the use of TES, while, ii) meeting indoor comfort

(i.e., temperature) requirements for the building occupants and, iii) preserving the integrity of the equipment and abiding by all operating constraints. Initially developed for applications in the chemical industry, Model Predictive Control (MPC) provides an ideal framework for addressing these goals and has recently become of major interest in the buildings sector [115]. It consists of a dynamic optimization that is repeatedly solved over a finite, rolling time horizon [227] that yields trajectories of system inputs.

The development of such control strategies for buildings is a challenging task, owing to several (potentially conflicting) requirements. The controller, must i) account for a long time horizon (at least equal to the period of time for which electricity price and weather predictions are available), but may, ii) potentially be executed frequently, such that control decisions are made over an appropriate time scale to account for short-term fluctuations in, e.g., weather and occupancy. Moreover, iii) the controller must include sufficient information about the building dynamics (i.e., incorporate an appropriate model), and iv) set the operating mode (e.g., charge, discharge) of TES storage systems, which, in turn, involves making integer, rather than continuous, decisions.

Most industrial applications of MPC rely on a tracking objective function, formulated in terms of minimizing the discrepancy between a (subset) of the system states and their corresponding target values (setpoints). In the case of HVAC systems, rather than tracking a prescribed temperature setpoint, it is more advantageous to formulate the controller objective function to account directly for operating cost. In this case, the indoor temperature is allowed to vary within

specified upper and lower comfort limits, and serves as an important handle for reshaping the energy consumption profile of the building. This approach is referred to as Economic MPC (E-MPC), and it has been increasingly adopted in buildings energy management research and applications [197, 166, 162, 181]. By accounting for the costs associated with dynamic changes in ambient conditions, occupancy and energy prices, E-MPC can outperform steady-state optimization and heuristics-based approaches from an economic perspective [10, 9].

Using economic MPC for advanced energy management in buildings thus consists of solving an optimization problem repeatedly over a prediction time horizon T that is periodically shifted forward in time. The objective of the problem is formulated in terms of operating cost (based on energy consumption), and the optimal value is established subject to (*s.t.*) a set of constraints, which includes the dynamic model of the building and equipment and comfort limitations:

$$\begin{aligned}
 \min_{\mathbf{u}, \mathbf{m}} \text{Cost} &= \int_{t_o}^T f(\text{electricity prices, electricity consumption}) dt \\
 \text{s.t.} & \text{ dynamic building model equations} \\
 & \text{HVAC system model} \\
 & \text{TES system model} \\
 & \text{occupant comfort limits} \\
 & \text{equipment limitations} \\
 & \text{disturbance forecasts}
 \end{aligned} \tag{4.1}$$

While the details of the problem formulation and associated solution strategies are elaborated on later in the paper, we discuss here a few of the challenges of E-

MPC for buildings. Specifically, the solution of (4.1) is complicated by the multiple time-scale nature of the system, comprising:

- a slow component, owing to longer-term trends imposed by daily weather changes and fluctuations in energy prices. Fully accounting for this dynamic component typically requires extending the prediction horizon T to a full day (24 h) period
- a fast component, which is associated with short-term fluctuations in occupancy, insolation and ambient conditions. Properly dealing with the short-term dynamics requires controller outputs to be recalculated frequently; typical applications reduce this sampling time to a few minutes, which, in turn, requires that controller calculations be completed within this time frame.

Moreover, since both continuous (\mathbf{u}) and discrete (\mathbf{m}) decision variables are typically present, and the system model must capture the system dynamics (represented by ordinary or partial differential equations), the formulation (4.1) is classified as (difficult to solve) a mixed-integer dynamic optimization (MIDO).

Given that an optimal value for each manipulated variable (or “control handle” \mathbf{u}, \mathbf{m}) of the system must be determined at every sample time over the entire time horizon, meeting these requirements considerably increases the size of problem (4.1). Additional complications arise from potential nonlinearities and discontinuities in the set of differential equations describing the building, HVAC and TES models (the latter, particularly due to capturing the melting/solidification events

if latent thermal storage is used). Moreover, building models can be themselves of large scale, compounding the dimensionality challenge.

Past research efforts seeking an efficient solution of the E-MPC controllers for buildings follow two main directions. On the one hand, *model reduction* is centered on simplifying the model elements to decrease the problem size. Methods for the derivation of low-order models that are suitable for online calculations have been proposed, e.g., in [164, 195, 104, 163]. On the other hand, modifications of the optimization problem formulation can improve execution efficiency. These include, e.g., utilizing the mixed logical dynamical (MLD) system paradigm [42]. Move blocking strategies [172], consisting of using the same value for a decision variable for all the time samples that are close to the end of the prediction horizon T (thereby decreasing the granularity of the problem), were shown to reduce problem size without significant economic penalty in terms of the value of the objective function in (4.1) [166, 167]. In a different vein, the use of simultaneous (rather than sequential) numerical solution approaches for the E-MPC optimization problem [9] was shown to lead to improvements in computational performance and robustness when dealing with problems involving unstable and ill-conditioned systems.

In this paper, we discuss a framework for developing such energy management strategies. Our work considers using “active” storage systems, as an extension of the available body of work that considers “passive” storage in buildings’ structural elements whose operation can only be modulated indirectly by altering the indoor air temperature. In particular, we focus on cold water storage and latent energy storage in phase change materials (PCMs) embedded in, e.g., chilled water storage

tanks [249, 206]. This affords us the opportunity to explore the equation-oriented modeling, simulation and control challenges posed by the nonlinearity associated with the solidification/melting phenomena that accompany the TES system charging and discharging events. Active TES systems feature multiple *operating modes* (including charging, discharging and an inactive/off state), and coordinating their operation with that of HVAC systems entails determining the optimal timing for activating each mode of operation. Heuristics are often used [116, 305], but a more sophisticated management approach is evidently desirable [66]. Model Predictive Control (MPC) is ideally suited to this end.

The novelty of our approach consists of performing a time-scale decomposition of the problem, followed by the design of a hierarchical control structure, comprising i) of a scheduling calculation in the slow time scale, which establishes the operating pattern of the thermal energy storage system, and, ii) a control system (e.g., MPC) operating in the fast time scale, which addresses short-time disturbances related to occupancy and ambient conditions. There are several examples [252, 152, 168, 284] in the literature that discuss use of hierarchical control schemes tailored to the use of storage in different time scales. In this work, we generalize this idea and present a framework for controlling passive and active TES in buildings. Focusing on the active storage management, we propose a reformulation of the optimization formulation that makes this problem tractable online and in real time.

4.3 Proposed Method: Optimization Framework for Energy Management

A schematic of a building and HVAC system with thermal energy storage is shown in Figure 4.2. This will form the basis for our discussion on energy management. The system can be described by a model of the form:

$$\mathbf{f}_1(t, \dot{\mathbf{x}}_B, \mathbf{x}_B, \mathbf{x}_P, \mathbf{u}_P, \mathbf{m}_P, \mathbf{d}) = \mathbf{0} \quad (\text{dynamic building model equations})$$

$$\mathbf{f}_2(t, \dot{\mathbf{x}}_P, \mathbf{x}_P, \mathbf{x}_B, \mathbf{x}_A, \mathbf{u}_P, \mathbf{m}_P, \mathbf{d}) = \mathbf{0} \quad (\text{air-side HVAC model equations})$$

$$\mathbf{f}_3(t, \dot{\mathbf{x}}_A, \mathbf{x}_P, \mathbf{x}_A, \mathbf{u}_A, \mathbf{m}_A, \mathbf{d}) = \mathbf{0} \quad (\text{coolant-side HVAC and TES model equations})$$

where \mathbf{x}_B is a vector of variables representing the temperature of the air at points inside the building and the temperatures of the structural elements of the building, \mathbf{x}_P represents the temperature of the air within the HVAC system, and \mathbf{x}_A represents the temperatures of the coolant in the HVAC system and TES. Subscripts P and A refer to passive and, respectively, active TES components. \mathbf{d} are the external disturbances (e.g., weather, occupants, etc.) to the system. The functions $\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3$ are systems of differential equations. Note that the equations are coupled to reflect the heat exchange between the components of the physical system, i.e., the terms $\mathbf{x}_B, \mathbf{x}_P,$ and \mathbf{x}_A appear in more than one function.

The time-varying decision variables of the optimization problem are divided into two subsets:

- continuous variables $\mathbf{u}_P \in \mathbb{R}^i$ and $\mathbf{u}_A \in \mathbb{R}^j$, which comprise, e.g., temperature setpoints and flow rates on the air-side and coolant-side of the HVAC system.

- binary variables $\mathbf{m}_P \in \mathbb{B}^k$ and $\mathbf{m}_A \in \mathbb{B}^l$, which represent the valve configurations and on/off equipment states that correspond to each distinct operating mode (as described in Section 4.7.1).

4.3.1 Centralized Optimization Formulation

We begin by describing a centralized control configuration where all decisions regarding the control of thermostat setpoints, as well as concerning the operation of the TES are made by a single optimization entity that follows the paradigm in (4.1), as illustrated in Figure 4.2. The grey area designated ‘Load’ shows the heat load that must be removed by the coolant part of the HVAC system.

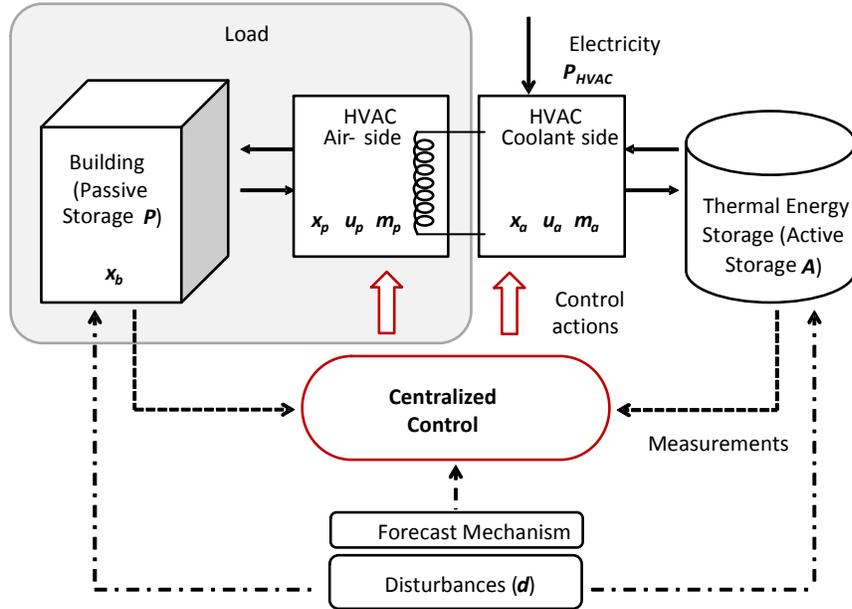


Figure 4.2: Centralized EMPC

In this case, the E-MPC problem (4.1) for building systems with TES becomes:

$$\min_{\mathbf{u}_P, \mathbf{m}_P, \mathbf{u}_A, \mathbf{m}_A} J = \int_{t_0}^{t_0+T} C(t) P_{HVAC}(t, \mathbf{u}_P, \mathbf{m}_P, \mathbf{u}_A, \mathbf{m}_A) dt \quad (4.2a)$$

$$s.t. \quad \mathbf{f}_1(t, \dot{\mathbf{x}}_B, \mathbf{x}_B, \mathbf{x}_P, \mathbf{u}_P, \mathbf{m}_P, \mathbf{d}) = \mathbf{0} \quad (4.2b)$$

(dynamic building model equations)

$$\mathbf{f}_2(t, \dot{\mathbf{x}}_P, \mathbf{x}_P, \mathbf{x}_B, \mathbf{x}_A, \mathbf{u}_P, \mathbf{m}_P, \mathbf{d}) = \mathbf{0} \quad (4.2c)$$

(air-side HVAC model equations)

$$\mathbf{f}_3(t, \dot{\mathbf{x}}_A, \mathbf{x}_P, \mathbf{x}_A, \mathbf{u}_A, \mathbf{m}_A, \mathbf{d}) = \mathbf{0} \quad (4.2d)$$

(coolant-side HVAC and TES model equations)

$$\mathbf{d} = \mathbf{f}_4(\mathbf{t}) \quad (\text{disturbance profiles}) \quad (4.2e)$$

$$\mathbf{x}_B^{low} < \mathbf{x}_B < \mathbf{x}_B^{high} \quad (\text{occupant comfort limits}) \quad (4.2f)$$

$$\mathbf{u}_P^{low} < \mathbf{u}_P < \mathbf{u}_P^{high} \quad (\text{equipment limitations}) \quad (4.2g)$$

$$\mathbf{u}_A^{low} < \mathbf{u}_A < \mathbf{u}_A^{high} \quad (4.2h)$$

$$\sum \mathbf{m}_P = 1 \quad \sum \mathbf{m}_A = 1 \quad (\text{logical constraints}) \quad (4.2i)$$

The optimization objective function J represents the cost over the prediction horizon T of power consumption in the HVAC system, P_{HVAC} , subject to the time-variable price structure $C(t)$.

The dynamic models of the building, TES, and HVAC systems are embedded in the problem formulation in constraints (4.2b) - (4.2d), and the disturbance forecasts are provided by (4.2e). Additional constraints (4.2f) - (4.2i) capture occupant

comfort bounds on the indoor air (zone) temperature, as well as equipment limitations. The latter include both physical limits (e.g., maximum and minimum air and coolant flow rates), and reliability and durability constraints (e.g., preventing short cycles for the HVAC equipment).

We refer to (4.2) as a “centralized” control strategy due to the fact that it is tasked with finding the optimal values of *all* manipulated variables/control handles (i.e., $\mathbf{u}_P, \mathbf{m}_P, \mathbf{u}_A, \mathbf{m}_A$) that pertain to managing the operation of *both* the building HVAC and TES systems. The solution of this problem for a warm day typically involves pre-cooling the building (the elements of \mathbf{x}_B representing indoor air temperature are at their lower bound \mathbf{x}_B^{low}) in order to further cool the structural elements, and charging the thermal storage tank (which corresponds to a specific value of \mathbf{m}_A and \mathbf{u}_A) for a period of time before the price increase [297, 117]. Then, during the peak price period, the structural elements can act as a heat sink to help maintain indoor comfort (the elements of \mathbf{x}_B representing indoor air temperature are usually allowed to drift to their upper bound \mathbf{x}_B^{high}) and the TES is used instead of the chiller to maintain the desired coolant temperature.

In the generic formulation (4.2), the control problem is infinite-dimensional, and a discretization step is required to obtain a solution. To this end, the prediction horizon T is divided into N intervals of length M , such that $T = NM$. The solution then consists of finding the optimal values of the decision variables for each of the N intervals assuming that each decision variable remains constant over the duration of an interval. The problem is solved on a moving time horizon, such that the solution for the first time interval is implemented in the system, the

prediction horizon is shifted in time by M time units, and the solution process is repeated.

As alluded earlier, in order to take full advantage of the predictive component of the controller, the time horizon T in (4.2) should be at least as long as the time span over which energy prices are provided (a 24-hour period in most energy markets). On the other hand, a short value of M must be used (typically, $M = 10$ minutes) in order to recompute control decisions and respond to disturbances (such as changes in occupancy and ambient conditions) as they occur. These two requirements increase the number of control intervals N , and, in turn, the total number of optimization decision variables (which is equal to $N \times (i + j + k + l)$, where i , j , k , and l are the dimensions of the decision variables - see the definitions at the start of section 4.3) over the entire time horizon. A large number of integer and continuous decision variables, coupled with a complex building model, renders this problem difficult to solve in the amount of time available to compute a control move (i.e., within the time span M) [32, 82].

4.3.2 Hierarchical Energy Management Strategy

The above observations motivate us to seek a new strategy for dealing with the dimensionality challenges related to solving the optimal energy management and control problem (4.2). To this end, we build on the developments in [252], and propose a hierarchical control strategy based on time scale decomposition. To begin, we make the following (reasonable) assumptions:

Assumption 4.1. *Electricity prices are known at least one day in advance for a*

duration T_p .

Assumption 4.2. *The interaction between the air-side and water-side of the HVAC system is defined in terms of the exchange of heat through a cooling coil (or a set of cooling coils).*

Assumption 4.3. *Owing to the fact that charging and discharging the TES involves changing the temperature and phase of a significant thermal mass, the dynamics of the TES are slower than those of the air-side HVAC.*

These assumptions indicate that the active thermal storage on the coolant-side of the HVAC system is distinct from the passive storage in the building in terms of both the dynamics, the dominant time scale, and level of uncertainty in operation. In turn, this suggests that a hierarchical control strategy, consisting of separate tiers of control action, is suitable for separately managing the use of active and passive storage in a building.

We develop the hierarchy of control actions based on the (centralized) problem formulation (4.2). We begin by noting that the solution to the problem (4.2) calculated at $t = 0$ will remain optimal over the course of a day if the disturbance variables (i.e., energy prices, occupancy and changes in ambient conditions) do not deviate from their forecast values. Based on Assumption 4.1 (i.e., that day-ahead energy prices are known with certainty), we note that the occupancy and ambient condition disturbances can typically be regarded as short-term fluctuations around a (forecast) mean. It is thus reasonable to assume that the solution calculated

at $t = 0$ will remain close to the optimal value even in the presence of such disturbances.

Using this observation and Assumptions 4.2 and 4.3, we conclude that controlling the operation (i.e., charging and discharging) of the active TES system is closely related to the day-ahead energy price profile and, as a consequence, should be addressed over the longer time horizon (slower time scale) for which these prices are available. Conversely, short-term fluctuations in occupancy and ambient conditions call for the use of a controller acting over a faster time scale. Consequently, we separate the centralized control structure (4.2) into two layers. The slow layer consists of an optimal *scheduling* calculation for the active TES in the slow time scale. The fast control layer operates over a shorter time horizon to reject short-term disturbances (Figure 4.3). Notice that the thermal load on the coolant side is essentially determined by the performance of the fast control layer.

A hierarchical control approach is widely-accepted for controlling systems with slow and fast dynamics [235]. The interactions between the control layers (discussed later in Section 4.3.2.3) distinguish this method from a distributed control system, where the individual controllers act over a similar time scale and share (limited) information with each other. Distributed controllers have been applied over the fast time scale for the control of multiple zones in large buildings (e.g., [185]). Below, we discuss in detail the two layers of control action in the hierarchical system, starting with the optimal scheduling system.

Remark 4.1. *We note that hierarchical control schemes consisting of multiple MPC controllers acting over different time horizons have been recently proposed in*

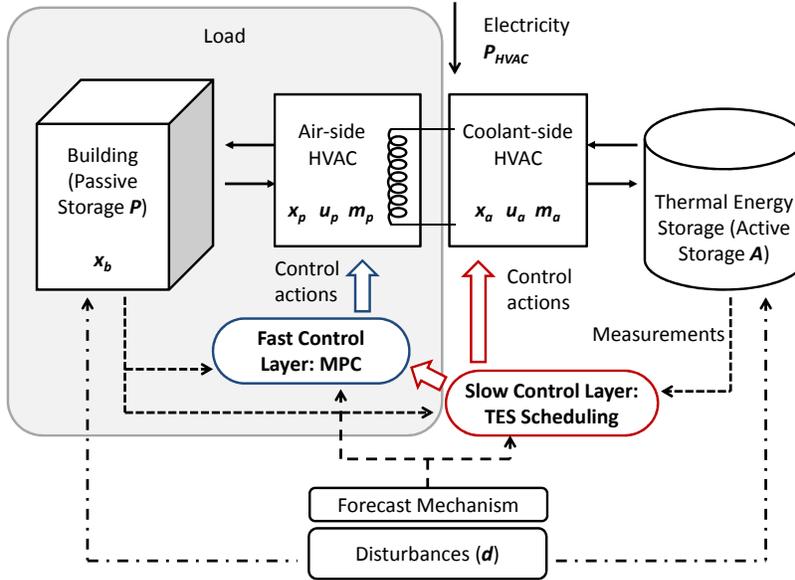


Figure 4.3: Hierarchical energy management scheme

the building systems literature, (e.g., [152]), as well as in the areas of microgrid (e.g., [301]) and process control (e.g., [90]). In the latter, the slow (economic) controller has a longer prediction horizon and less accurate forecasts than the fast controller, with the mission of the latter being to implement the setpoints computed in the slow layer.

4.3.2.1 Slow Layer: TES Scheduling System

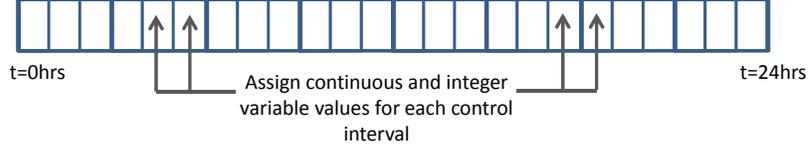
As we observed above, the centralized EMPC problem (4.2) could, in principle, be solved only once during the course of the day, namely, when energy prices become available for the next 24 hour period. This effectively transforms the op-

timal control problem into a scheduling one, consisting of optimally synchronizing the operation of the active TES system with weather, energy price and occupancy fluctuations. Thus, we define the optimal scheduling problem based on the original EMPC formulation (4.2).

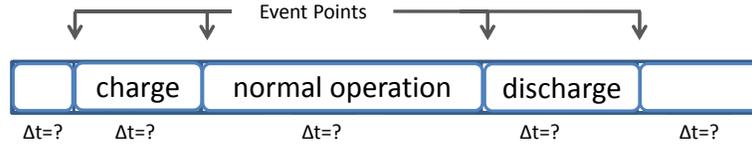
While this eliminates the challenge of solving a large, mixed-integer nonlinear optimization problem *in closed loop* and in *real-time*, it does not alleviate the difficulty of obtaining a solution. Our experience suggests that this difficulty is largely due to the need to use integer variables to capture the operating modes of the HVAC equipment as well as to the challenges related to modeling the phase transformation phenomena in the PCM-based TES system. To address the former, we propose reformulating problem (4.2) as nonlinear program that has only continuous decision variables. Modeling phase change is discussed in Section 4.7.2.4.

To begin, we make the key observation that the operating sequence of the active TES system is cyclical and the succession of the operating modes is known a-priori: charging must occur before discharging, and the two events are naturally separated by periods of time when the system is idle. This is represented in Figure 4.4a. Furthermore, it is reasonable to assume that the charging and discharging steps are uninterrupted, in the sense that a situation in which one would begin charging the storage and then stop for a period of time before resuming charging is not expected. This observation allows us to reformulate the mixed-integer dynamic optimization (MIDO) for operating the TES system by focusing our attention on the timing and duration of the operating events, rather than on their succession. This eliminates the discrete decisions associated with deciding the operating model

of the system (Figure 4.4b).



- (a) Mixed integer formulation for establishing the operating mode and continuous decision variables. In this example there are hourly control intervals.



- (b) Continuous representation of the operating mode assignment problem. Rather than using integer variables to select the operating state of the system at each (discretized) time point, the durations of the events in the established cyclical operating pattern are optimized.

Figure 4.4: Cyclical mode transition behavior.

The continuous reformulation of the optimal scheduling problem is shown in equation (4.3).

$$\min_{\mathbf{u}_A, t_1, \dots, t_v} J^S = \int_0^{T_{slow}} C(t) P_{HVAC}(t, \overline{\mathbf{u}}_P, \overline{\mathbf{m}}_P, \mathbf{u}_A, S) dt \quad (4.3a)$$

$$s.t. \quad \mathbf{f}_1(t, \dot{\mathbf{x}}_B, \mathbf{x}_B, \mathbf{x}_P, \overline{\mathbf{u}}_P, \overline{\mathbf{m}}_P, \mathbf{d}) = \mathbf{0} \quad (4.3b)$$

(dynamic building model equations)

$$\mathbf{f}_2(t, \dot{\mathbf{x}}_P, \mathbf{x}_P, \mathbf{x}_B, \mathbf{x}_A, \overline{\mathbf{u}}_P, \overline{\mathbf{m}}_P, \mathbf{d}) = \mathbf{0} \quad (4.3c)$$

(air-side HVAC model equations)

$$\mathbf{f}_3(t, \mathbf{x}_A, \mathbf{x}_P, \mathbf{x}_A, \mathbf{u}_A, \mathbf{m}_A, \mathbf{d}) = \mathbf{0}$$

(coolant-side HVAC and TES model equations) (4.3d)

$$\mathbf{d} = \mathbf{f}_4(\mathbf{t}) \quad (\text{disturbance profiles}) \quad (4.3e)$$

$$\mathbf{u}_A^{low} < \mathbf{u}_A < \mathbf{u}_A^{high} \quad (\text{equipment limitations}) \quad (4.3f)$$

$$S = [\{s_1 \rightarrow \dots \rightarrow s_v\}, \{t_1, \dots, t_v\}], s_i \in \mathbf{m}_A \quad (4.3g)$$

$$\sum_{i=1}^v t_i = T_{slow} \quad (4.3h)$$

where \mathbf{u}_P and \mathbf{m}_P are fixed to values $\overline{\mathbf{u}_P}$ and $\overline{\mathbf{m}_P}$ (i.e., the assumed values of the continuous and discrete decision variables in the air side of the HVAC system) and are not part of the decision variable set. In (4.3g), we use s_i to denote an event of duration t_i in the schedule S with v total events. The schedule is given by the pair $S = (\{s_1 \rightarrow \dots \rightarrow s_v\}, \{t_1, \dots, t_v\})$. The nature and order of each event (i.e., the elements of S) are pre-specified by assigning an operating state from the set \mathbf{m}_A to each event s_i . On the other hand, the event durations $t_i, i \in 1, \dots, v$ become part of the decision variables. The prediction horizon is $T^{slow} \approx T_p$, and an additional constraint (4.3h) ensures that the total duration of all events is the same as the prediction horizon.

4.3.2.2 Fast Layer: Air-Side Controller

The fast control layer is tasked with addressing the control objectives related to the fast dynamics of the system, that is, the air-side of the HVAC. The corresponding control problem can be formulated as:

$$\min_{\mathbf{u}_P, \mathbf{m}_P} J^F \quad (4.4a)$$

$$s.t. \quad \mathbf{f}_1(t, \dot{\mathbf{x}}_B, \mathbf{x}_B, \mathbf{x}_P, \mathbf{u}_P, \mathbf{m}_P, \mathbf{d}) = \mathbf{0} \quad (\text{dynamic building model equations}) \quad (4.4b)$$

$$\mathbf{f}_2(t, \dot{\mathbf{x}}_P, \mathbf{x}_P, \mathbf{x}_B, \overline{\mathbf{x}}_A, \mathbf{u}_P, \mathbf{m}_P, \mathbf{d}) = \mathbf{0} \quad (\text{air-side HVAC model equations}) \quad (4.4c)$$

$$\mathbf{d} = \mathbf{f}_4(\mathbf{t}) \quad (\text{disturbance profiles}) \quad (4.4d)$$

$$\mathbf{x}_B^{low} < \mathbf{x}_B < \mathbf{x}_B^{high} \quad (\text{occupant comfort limits}) \quad (4.4e)$$

$$\mathbf{u}_P^{low} < \mathbf{u}_P < \mathbf{u}_P^{high} \quad (\text{equipment limitations}) \quad (4.4f)$$

$$\sum \mathbf{m}_P = 1 \quad (\text{logic constraints}) \quad (4.4g)$$

Equation (4.4) does not include the model of the TES system and the related control decisions ($\overline{\mathbf{x}}_A$ approximates the values of the variables on the coolant side). This reduces the size and complexity of (4.4), which can be solved and implemented online. The objective function J^F can be formulated in two different ways: as a tracking or an economic objective. Both of these MPC objectives have been discussed extensively in the literature [9, 179, 27], and are summarized below:

- *Tracking* objective

$$J^F = \int_{t_0}^{t_0+T_{fast}} (\mathbf{x}_P - \mathbf{x}_P^{sp})^2 dt \quad (4.5)$$

where \mathbf{x}_P comprises, e.g., the indoor air temperature(s). This objective seeks to minimize the difference between the indoor temperature(s) and given temperature setpoint(s). The setpoints can either be defined by the occupants or, in the case of the hierarchical control strategy discussed in this paper, computed by the optimal scheduling system.

- *Economic* objective, whereby the fast controller attempts to minimize operating cost.

$$J^F = \int_{t_0}^{t_0+T_{fast}} C(t)P_{HVAC}(t, \mathbf{u}_P, \mathbf{m}_P, \mathbf{u}_A, S)dt \quad (4.6)$$

In this case, the temperature of the building zone is no longer steered towards the desired setpoint; rather, it is allowed to fluctuate within upper and lower comfort limits. This approach was shown to result in a precooling behavior (see, e.g., [297, 135]). The lower comfort limit is reached during the off-peak morning hours, when excess refrigeration is used to precool the building structure. Subsequently, temperatures are allowed to drift to the upper comfort limit during the peak hours in order to save energy.

Remark 4.2. *The choice of controller for the fast time scale is largely driven by the operating conditions of the building. When disturbance forecasts are accurate, tracking the temperature setpoint with a simple feedback controller would maintain the system close to the optimal solution calculated by solving (4.3). On the other hand, when significant fluctuations in weather and occupancy may be present, choosing an advanced control option is indicated.*

Remark 4.3. *The stability of the proposed hierarchical, composite control scheme can be inferred from the stability of the individual controllers in the different time scales (see, e.g., the arguments in [27]). Since the scheduling layer provides the setpoints for the low level controllers but does not operate in closed-loop, the issue of stability is related to the implementation of the lower level controllers. Stability*

can therefore be guaranteed as long as the systems remains closed-loop stable in the presence of setpoint changes and known disturbances.

4.3.2.3 Interaction and Information Exchange Between Control Tiers

The optimal scheduling problem (4.3) relies on the assumption that the behavior of the building and the air-side loop is known, so \mathbf{u}_P and \mathbf{m}_P are fixed to values $\overline{\mathbf{u}_P}$ and $\overline{\mathbf{m}_P}$. The models \mathbf{f}_1 and \mathbf{f}_2 are still used to approximate the load to be removed by the coolant loop, which is modeled in equation \mathbf{f}_3 . Ultimately, this assumption results in a model that approximates the ‘Load’ to the coolant-side of the HVAC system. It is therefore possible to avoid modeling the building and air-side HVAC system if the ‘Load’ to the coolant-side can be modeled directly. This approach further simplifies the optimization problem for scheduling TES usage to:

$$\min_{\mathbf{u}_A, t_1, \dots, t_v} J^S = \int_0^{T_{slow}} C(t) P_{HVAC}(t, Load, \mathbf{u}_A, S) dt \quad (4.7a)$$

$$s.t. \quad \overline{\mathbf{f}_3}(t, \mathbf{x}_A, \mathbf{x}_A, \mathbf{u}_A, \mathbf{m}_A, Load) = 0 \quad (4.7b)$$

$$Load = f_5(t) \quad (4.7c)$$

$$\mathbf{u}_A^{low} < \mathbf{u}_A < \mathbf{u}_A^{high} \quad (4.7d)$$

$$S = [\{s_1 \rightarrow \dots \rightarrow s_v\}, \{t_1, \dots, t_v\}], s_i \in \mathbf{m}_A \quad (4.7e)$$

$$\sum_{i=1}^v t_i = T_{slow} \quad (4.7f)$$

where $\overline{\mathbf{f}_3}$ represents the model of the cooling loop based directly on the load, rather than on the interactions between \mathbf{x}_P and \mathbf{x}_A . In practice, the *Load* value (or the estimates $\overline{\mathbf{u}_P}$ and $\overline{\mathbf{m}_P}$) can be determined based on the recorded behavior of a

previous day. Equation (4.7c) provides the load forecast and replaces the need to model disturbances directly. This approach is represented graphically in Figure 4.5.

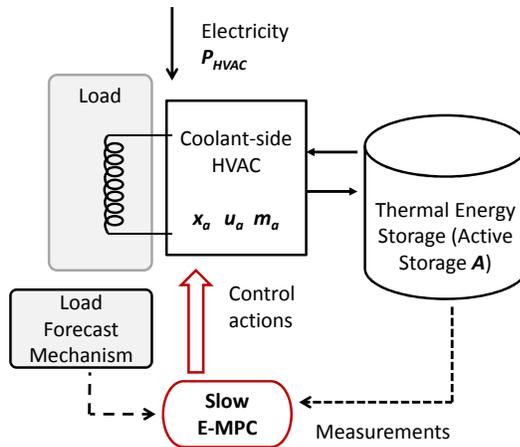


Figure 4.5: The slow layer of the energy management scheme using approximates of the load instead of a detailed building model.

In case there are significant day-to-day changes in the behavior of the system, a schedule update/rescheduling scheme can be easily implemented based on i) transferring \mathbf{u}_P , \mathbf{m}_P , and/or *Load* data in real-time to the scheduling layer, ii) monitoring the discrepancy between \mathbf{u}_P , \mathbf{m}_P , or *Load* and the corresponding estimates $\overline{\mathbf{u}_P}, \overline{\mathbf{m}_P}$, and the load forecast (4.7c) and, iii) re-solving (4.3) or (4.7) when the discrepancy (or its integral in time) exceeds a specified threshold. This rescheduling approach is facilitated by the fact that the continuous formulation of the problems (4.3) and (4.7) are likely easy to solve online.

The information exchange between the two control tiers can, however, be designed to be bidirectional. Thus, in addition to the \mathbf{u}_P and \mathbf{m}_P data transfer from

the fast controller to the scheduling layer, it is possible to include the decision variables related to the building and air-side HVAC and their associated models in the optimal scheduling problem. The optimal values of, e.g., the thermostat setpoints in the building zone, calculated at the scheduling level, could serve as an initial guess for the fast control layer calculations at the start of the day.

Remark 4.4. *Scheduling and control for process and energy systems have traditionally been considered separately (and often independently of each other) owing to the fact that they involve different time horizons and levels of modeling detail. Their integration has recently become of interest owing to potential economic benefits; however, the development of a generic and transparent methodology for the integration of scheduling and control remains an open question (see, e.g., the review in [24]). The framework discussed above can be construed as a “top-down” approach for integrating scheduling and control. Furthermore, our developments suggest that energy systems with storage represent excellent potential test beds for new control/scheduling integration techniques.*

4.4 Solution Methods

Obtaining a solution to a MIDO (which is the form of all of our scheduling/control tiers) is very challenging, particularly when the mathematical model is complex and the time horizon under consideration is large. In addition to the presence of binary decision variables, the solution is complicated by the model discontinuity associated with the temperature-enthalpy relationship for the PCM behavior. Methods for solving MIDOS fall into two broad categories.

Simultaneous methods for solving MDOs have found extensive use in the solution of nonlinear model predictive control problems [9]. They rely on discretizing the equations of the system’s mathematical model using, e.g., finite difference approximations or collocation on finite elements. This converts the MDO into a large-scale mixed-integer nonlinear program (MINLP), which can be solved with existing solvers. Challenges associated with simultaneous methods include managing the size of the resulting MINLP, as well as finding appropriate starting points for the solution algorithm when the system dynamics are complex [32].

However, an additional complication of using a simultaneous approach arises from dealing with the IF-THEN statements in the temperature-enthalpy relationship for PCM (this will be demonstrated in the case study). These cannot be directly accounted for in the MINLP formulation; rather, they must be converted to logical conditions based on binary variables using a “big-M” formulation and disjunctive programming [230, 151, 120]. In turn, this introduces a significant number of additional binary variables (in this case, three for each time discretization point), considerably increasing the computational resources and time required to solve the problem and severely limiting the potential for obtaining a solution on-line and in real-time.

By contrast, **sequential** solution approaches [277] alternate between the time integration of the model equations and appropriate sensitivities over the time horizon of the problem, and the solution of a simpler NLP. While less widespread in the solution of control problems, sequential methods provide a more natural approach for dealing with the discontinuities in the PCM model and thus represent

the method of choice for the following case studies.

4.5 Overview of Case Studies

The following three sections present case studies demonstrating various aspects of the hierarchical energy management system applied to thermal storage applications. The case studies are summarized in Table 4.1.

Table 4.1: TES Case Study Summary

Section	System Description	Demonstrated Control Layer
4.6	Chilled water TES tank and building which is based on the UT Thermal Facade Lab	Centralized optimization (formulation (4.2) in Section 4.3.1) but with variables only of the class $\mathbf{u}_P, \mathbf{u}_A, \mathbf{m}_A$. The cyclical operation observation is applied to transform the centralized MINLP to a NLP (this is successful because there are no decision variables of the class \mathbf{m}_P , which would still be present after the transformation). This case study demonstrates the behavior of both the TES and the fast air temperature dynamics.
4.7	PCM TES tank and building which is based on the UT Thermal Facade Lab	Slow layer using a model of the building load (formulation (4.7)).
4.8	PCM TES tank and building load profiles based on a commercial building simulation	Slow layer using a model of the building load but without the cyclical operation assumption to transform the problem to an NLP. The formulation is similar to that in (4.7), but binary variables of the class \mathbf{m}_A are still present.

Table 4.2: Operating Modes

Mode	V1	V2	V3	V4	Chiller	CC
m_1 - Cooling Only	1-2	open	closed	any	on	on
m_2 - Charging Only	1-3	open	open	3-2	on	off
m_3 - Cooling & Charging	1-2&3	open	open	3-2	on	on
m_4 - Cooling & Discharging	1-2	closed	open	2-1	off	on

flow to or from the tank. In Mode 2 (TES charging), coolant is fed into the tank from the bottom. In Mode 3, the coolant flow is split between the tank and the cooling coil. In Mode 4 (TES discharging), coolant from the tank flows out at the bottom and is cycled through the cooling coil and the chiller (with the latter turned off). These four modes can be selected via the binary vector $\mathbf{m} = [m_1 m_2 m_3 m_4]^T$, which is a special ordered set of type one (SOS1).

4.6.1 TES Model

The chilled water storage tank is assumed to be well-mixed in the radial direction (i.e., no temperature gradients along the radius). Furthermore, a narrow axial zone with high axial temperature gradients (thermocline) is assumed to exist at all times, separating the chilled water at the bottom from warmer return water at the top. The tank has height h and area A . The tank temperature can be modeled as:

$$(\rho C_p)_c V_{tank} \frac{\partial T_{tank}}{\partial t} = -v \frac{\partial T_{tank}}{\partial z} + k \frac{\partial^2 T_{tank}}{\partial z^2} \quad (4.8)$$

where the velocity is defined as a function of the operating mode as $v = (m_2 + m_3)F_4/A + m_4F_3/A$. As the temperature gradients outside the thermocline are minimal, we neglect the effect of thermal conductivity in the tank in this work.

We define the state of charge based on the enthalpy of the tank as:

$$H_{max} = \int_0^h (\rho C_p)_c (284) dz$$

$$\begin{aligned}
H_{min} &= \int_0^h (\rho C_p)_c (277) dz \\
\text{TES charge} &= \max\left(0, \frac{H_{max} - \int_0^h (\rho C_p)_c T_{tank}(t,z) dz}{H_{max} - H_{min}}\right) \in (0, 1)
\end{aligned} \tag{4.9}$$

with H_{max} and H_{min} being the enthalpies corresponding to uniform temperature distributions at the maximum (i.e., $284K$) and, respectively, minimum (i.e., $277K$) chilled water temperatures possible given the limits on the system. Note that a negative TES charge state may be reported when the temperature at the top of the tank rises slightly above the upper limit set in equation (4.9). This is not a concern because in the following day the tank will still be able to charge to full capacity.

The tank model was discretized using a finite-difference approach on a grid consisting of 100 nodes. We note that modeling the tank is a non-trivial matter because the direction of flow and velocity change when switching between charging and discharging modes. This necessitates switching between backwards and forwards discretization schemes to solve the PDE (4.8), depending on the operating mode.

4.6.2 Building Model

The single zone building is modeled as a well mixed air volume in contact with a thermal mass block that represents the structural elements (Figure 4.7).

The indoor air (T_z) and thermal mass (T_m) temperatures are given by Equations (4.10) and (4.11), where $Q_m = (UA)_m(T_m - T_z)$ represents the total heat transfer through the building envelope, $Q_{ext} = (UA)_{ext}(T_{ext} - T_m)$ represents the

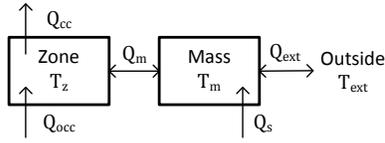


Figure 4.7: Model of the building zone

interaction of the structural elements with the environment (with T_{ext} the outside air temperature), Q_{occ} represents heat generated through the presence of occupants and their associated behavior, and Q_s accounts for incident solar radiation on the structural elements. Q_{cc} is the heat removed by the cooling coil as explained in the next section.

$$(\rho C_p)_{air} V_{zone} \frac{dT_z}{dt} = Q_m + Q_{occ} - (1 - m_2) Q_{cc} \quad (4.10)$$

$$\rho_m C_{p,m} V_m \frac{dT_m}{dt} = Q_{ext} + Q_s - Q_m \quad (4.11)$$

4.6.3 Low Level Controllers

Within each mode described in Table 4.2, PI controllers manipulate several continuous variables to maintain the operating setpoints. Because of the thermal stratification in the storage tank and the need to maintain a temperature gradient across the cooling coil, it is necessary to regulate the flow rate coming out of the cooling coil such that the temperature is $5^\circ C$ greater than the tank temperature. The flowrate F_2 is therefore manipulated based on deviations in the cooling coil temperature, T_{cc} , from the setpoint T_{cc}^{sp} . This constraint is also included in the formulation of the control problem in the next section in order to limit the total

temperature rise seen in the cooling loop during the discharge mode.

$$V_{cc} \frac{dT_{cc}}{dt} = F_3(T_{chill} - T_{cc}) + (1 - m_2) \frac{Q_{cc}}{(\rho C_p)_c} \quad (4.12)$$

The flows around valve 1 are given below. F_{1a} is a constant representing the steady state chiller flow rate. The parameter $\alpha \in [0, 1]$ is determined based on the operating mode and reflects the position of valve 1 as listed in Table 4.2 (in charging mode, $\alpha = 0$, and in discharge and cooling modes $\alpha = 1$. In the charge and cool mode α is anywhere in the range $(0, 1)$).

$$F_{1b} = F_{1a} - F_2$$

$$F_3 = \alpha F_2$$

$$F_4 = (1 - \alpha) F_2$$

A PI controller is used to regulate T_{chill} , the temperature of the stream exiting the chiller, by manipulating Q_{chill} , the heat removed inside the unit. The temperature exiting the chiller, T_{chill} is shown in equation (4.13).

$$\begin{aligned} V_{chill} \frac{dT_{chill}}{dt} &= (F_{1b} T_{chill} + (m_1 + m_3) F_3 T_{cc} + \\ &(m_2 + m_3) F_4 T_{tank}(h) + m_4 F_3 T_{tank}(0) \\ &- F_{1a} T_{chill}) - (1 - m_4) \frac{Q_{chill}}{(\rho C_p)_c} \end{aligned} \quad (4.13)$$

We assume that the cooling coil removes all the heat necessary to maintain the zone temperature at the desired setpoint, and do not model the physical aspects of the cooling coil dynamics, which can be quite complex. The heat removal Q_{cc} is computed as the output of a PI controller that maintains the zone temperature at its setpoint, T_z^{sp} .

4.6.4 Hierarchical Control Implementation

The slow controller is based on the formulation in (4.3). We will compare two different schedule structures, that differ in the choice of charging mode. In S^I , the tank is charged using m_2 (charge only), while in S^{II} , m_3 (charge and cool) is used. This results in the operating sequences shown in Equation (4.14) and (4.15) with $n = 5$ stages over a 24 hour horizon starting at midnight:

$$S^I = (m_1 \rightarrow m_2(\text{charge}) \rightarrow m_1 \rightarrow m_4(\text{discharge}) \rightarrow m_1, t_1, t_2, t_3, t_4, t_5) \quad (4.14)$$

$$S^{II} = (m_1 \rightarrow m_3(\text{cool \& charge}) \rightarrow m_1 \rightarrow m_4(\text{discharge}) \rightarrow m_1, t_1, t_2, t_3, t_4, t_5) \quad (4.15)$$

The formulation for the slow MPC (Equation 4.3) is shown below. The chiller load contribution in the objective function is squared to improve the convergence of the optimization solver. Furthermore, while the operating mode in each stage is fixed, we allow for a finite number of changes of T_z^{sp} in each stage.

$$\min_{T_z^{sp}, t_1, \dots, t_5} J = \int_0^{24} C(t) Q_{chill}^2(t, T_z^{sp}, S) dt \quad (4.16)$$

s.t. building model equations in previous section

$$297 < T_z^{sp} < 302$$

$$295 < T_z < 305$$

$$T_{cc} < 285$$

$$\sum_{i=1}^5 t_i = 24$$

In the formulation (4.16) there are two bounds related to the indoor temperature. $T_z^{sp, max} = 302K$ is the upper limit for the zone setpoint temperature, and $T_z^{max} = 305K$ is the upper limit for the actual zone temperature. It is necessary to specify the latter because when the charge only mode m_2 is used there is no means for heat removal in the zone to maintain the specified value of T_z^{sp} . T_z^{max} therefore

represents a comfort limit that must be satisfied even in the charge only mode when it is not possible to track T_z^{sp} .

In this case study, we assume that there is no plant-model mismatch (i.e., the models used in both the optimization and the system being controlled are the same). We do, however, consider there are different forecasts provided to the optimization and the system. The values of T_z^{sp} calculated by the slow MPC can thus be assumed to be near-optimal (barring the effect of inaccurate disturbance forecasts) and can be imposed by the distributed controllers described above, which act as the fast control layer (alternatively, a MPC controller can be implemented to control the dynamics in the fast time scale).

4.6.5 Model Inputs

Figure 4.8 depicts the disturbance profiles for the single zone building. The ‘actual’ values are provided to the system being controlled (i.e., the plant), while the ‘forecast’ profiles are used by the model in the optimization calculations. A heat input associated with about 4 occupants acts as a disturbance to the zone, and is not provided to the supervisory controller. The weather data are based on TMY2 (Typical Meteorological Year) data for a July day. Q_s is the net radiation on all building surfaces at a given time, and this profile was obtained through an EnergyPlus [263] simulation of the TFL.

Model parameters are presented in Table 4.3. There are two different storage tank volumes that will be compared to see the effect on total cost of having surplus or insufficient storage capacity.

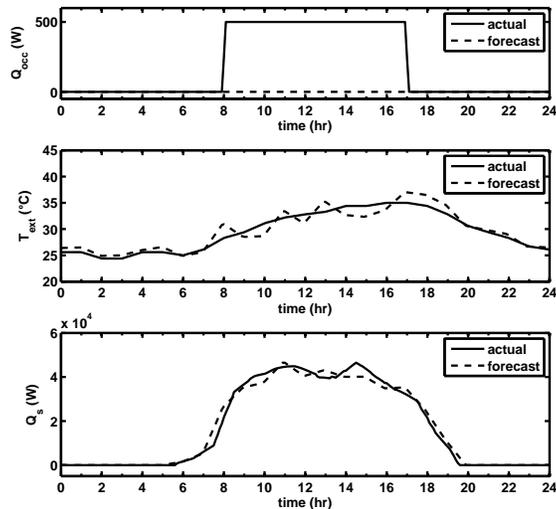


Figure 4.8: Disturbance profiles. The ‘actual’ values are provided to the system being controlled (i.e., the plant), while the ‘forecast’ profiles are used by the model in the optimization calculations.

The cost profile is shown in Figure 4.9 and is representative of a real-time pricing structure. We note that fifty-fold cost changes between peak- and off-peak times are typical in the Southern US (see Figure 4.1, as well as [65]).

4.6.6 Results and Discussion

The system was modeled and solved using the gPROMS dynamic optimization solver [217] using a control vector parametrization approach [278]. The solutions to (4.16) when there is surplus storage ($V_{tank} = 1m^3$) are shown in Figure 4.10, and when there is insufficient storage ($V_{tank} = 0.75m^3$) in Figure 4.11. We note that the optimal discharge time (i.e., the time when TES switches to mode 4–discharging) coincides with the start time of peak energy pricing in both cases. This indicates

Table 4.3: Parameter values

Parameter	Value	
F_{1a}	8	gal/min
V_{cc}	0.03	m ³
V_{zone}	500	m ³
V_m	100	m ³
V_{tank} (large)	1	m ³
V_{tank} (small)	0.75	m ³
h	1	m
k	0	W/m ² K
$(\rho C_p)_c$	4e6	J/m ³ K
$\rho_m C_{p,m}$	2.2e6	J/m ³ K
$(\rho C_p)_{air}$	1214	J/m ³ K
$(UA)_m$	75	W/K
$(UA)_{ext}$	100	W/K
T_{chill}^{sp}	279	K
T_{cc}^{sp}	284	K

that an operating heuristic replicating this behavior would be correct for systems of this type, regardless of the amount of active storage available.

Table 4.4: Objective function value at $t = 24$ hours (as $J/10^7$)

	Surplus Storage (S^I)	Insufficient Storage (S^I)	Insufficient Storage (S^{II})
Model	2.94	3.77	4.31
Plant	3.04	4.70	6.73

When there is sufficient storage capacity in the TES system, the optimization results suggest using only the active storage and not take advantage of the passive storage in the building walls through precooling behavior. In fact, the TES system is only charged to the extent required to cover the peak-price time interval, as seen by comparing the charge state in Figures 4.10 and 4.11. On the other hand, when

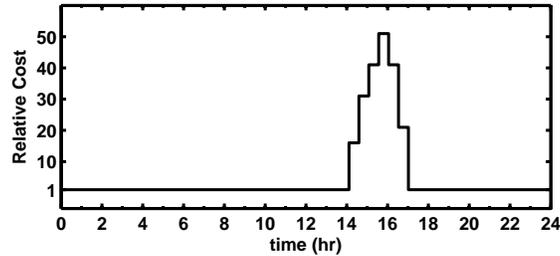


Figure 4.9: Cost profile

there is insufficient active TES to span the entire cost increase period, precooling of the building structure is needed directly before the price increase takes place (Figure 4.11). There is also an effort to precool the thermal mass in the brief period before charging the TES, as shown in both Figures 4.10 and 4.11. This is because the charge time is ultimately limited by the zone temperature constraint.

When the charging and cooling mode (m_3) is incorporated in the schedule, this is no longer a limiting factor. This is demonstrated in Figure 4.12, where the charge and cool mode is applied to the case with insufficient storage. It should be noted that when both operations are performed simultaneously the tank cannot be charged using the maximum flow rate (i.e., $\alpha > 0$. In this example, $\alpha = 0.1$ was used.). Therefore the time needed to charge the tank will increase compared to the results using the charge only mode, which has the fastest rate of charging. Charging the tank using the charge and cool mode took nearly double the charge time in the charge only mode, and the tank does not fully charge because of an additional practical constraint that limits the time allotted for this task is four hours. However, the discharge scenario is nearly identical so the savings during the peak period remain relatively constant. Differences arise because without the

tank fully charged, there is an even lower utilization of the active TES and more precooling is required to minimize the total operating costs.

The 24h-operating costs for the three cases are shown in table 4.4. The comparison between the plant and model rows highlights the effect of incorrect forecasts (especially the lack of occupancy predictions) used in the calculation. As expected, the total cost when there is surplus storage is lower than in the case when not all the needed storage capacity is available. The actual cost incurred by the “plant” is higher than that predicted by the “model” because of the incorrect forecasts. The difference is more pronounced when there is insufficient storage because the chiller must actually be turned on during the peak cost period (which is not anticipated by the model) in order to maintain the desired temperature setpoint.

We note that the solution to (4.16) was found in approximately ten minutes for each case on a regular desktop computer. This is sufficiently fast for a ‘slow’ MPC where the optimization is only solved once per day. In fact, in light of this relatively short solution time it would be possible to repeat the calculation more than once throughout the day, although this is not necessary. We conjecture that similarly low computing times would be required for the fast MPC calculation, and conclude that the proposed methodology shows real incentive for practical implementation.

Remark 4.5. *In all the scenarios explored in this case study, the presence of the TES system leads to a significant reduction (or complete elimination) of energy consumption in the building during the peak price hours (note the low values of Q_{chill} in Figures 4.10-4.12). This suggests that implementing energy storage on a*

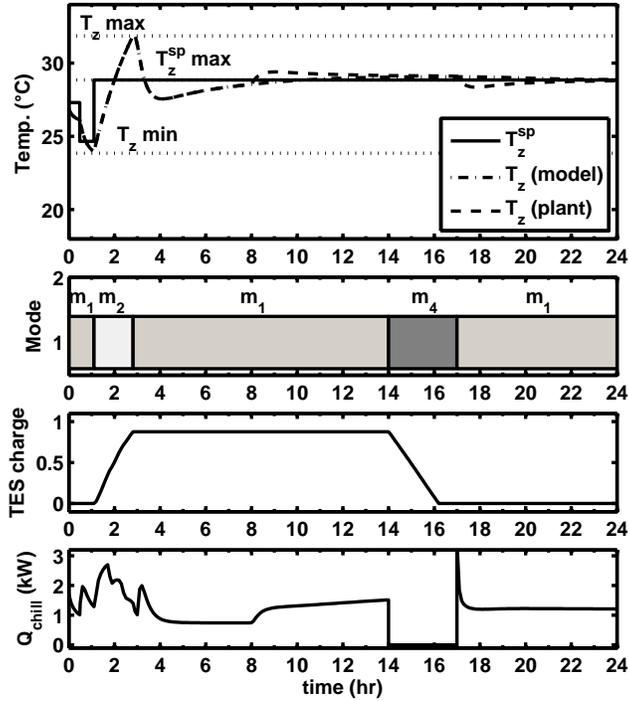


Figure 4.10: From top to bottom: Zone temperature and setpoint in the optimization model and plant, operating mode, TES charge state, and chiller load for the surplus storage case.

large scale would be a significant enabling factor (together with appropriate coordination mechanisms) for mitigating the variability of the load that buildings place on the grid and achieving a “load leveling” effect.

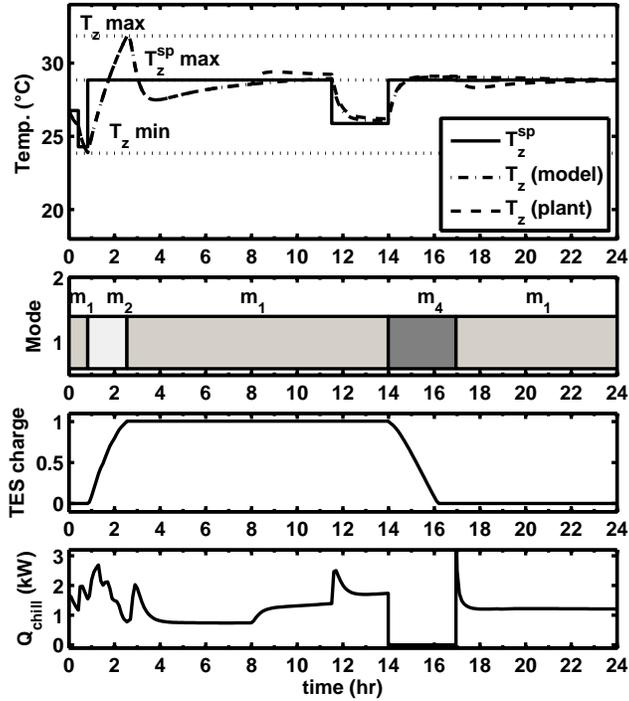


Figure 4.11: From top to bottom: Zone temperature and setpoint in the optimization model and plant, operating mode, TES charge state, and chiller load for the insufficient storage case.

4.7 Case Study: PCM TES Scheduling with Detailed Dynamic Model

In this section, we present a detailed simulation case study of the slow control layer for TES scheduling. The fast control layer and use of MPC for indoor comfort has been well represented in the literature (see the review papers [3, 295]). For this reason, we maintain our focus on the operation of active TES and the design of the Slow MPC control layer. The Thermal Facade Laboratory at the School of

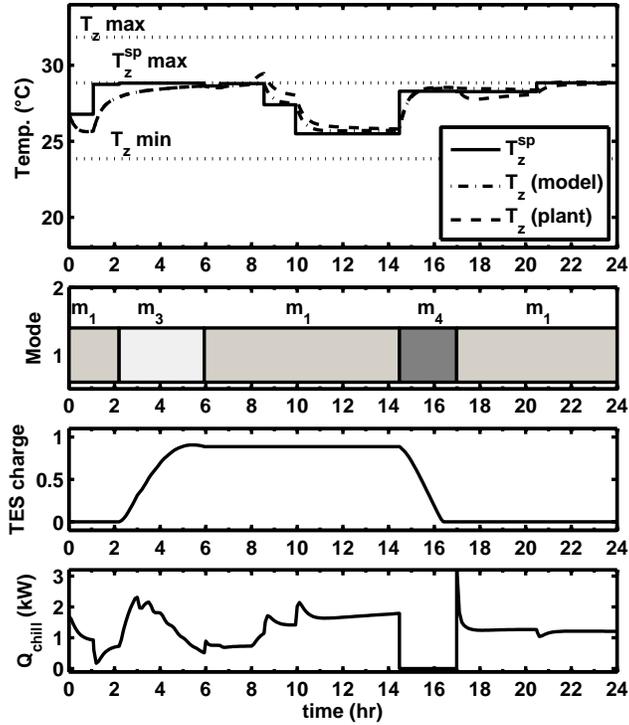


Figure 4.12: From top to bottom: Zone temperature and setpoint in the optimization model and plant, operating mode, TES charge state, and chiller load for the insufficient storage case when using the 'charge and cool' mode instead of 'charge only'.

Architecture of the University of Texas at Austin is the inspiration for our model of a building with a complex and flexible HVAC system with energy storage. The case study is focused on the operation of the coolant loop and the solution of a TES scheduling problem of the type in (4.7), so the construction of the building is not discussed in detail. For information on the design of the Thermal Facade Lab, we direct the reader to [262], and for information on the construction of the

storage Tank, [36, 37] present the parameters and results of the preliminary design analysis.

4.7.1 System Description

The system (see Figure 4.13) consists of a single-zone building, equipped with a chiller-based HVAC system. Chilled coolant is produced by an electric chiller and circulated through a cooling coil placed inside the building zone. The chilled coolant can also be (partially) directed into a storage tank filled with PCM capsules. The coolant circuit can operate in four distinct, discrete modes:

1. a normal mode (m_1), where the chiller and cooling coil are used to remove heat from the zone
2. a charging mode (m_2) where the refrigeration generated by the chiller is stored in the form of latent heat, by freezing the PCM capsules in the storage tank
3. a combined charging and cooling mode (m_3)
4. a discharge mode (m_4) where the TES is used in place of the chiller to remove the heat from the coolant loop. The chiller may turn on in this mode as well, depending on the building cooling demand

The system configurations (defined in terms of valve positions and equipment on/off states) associated with each operating mode are summarized in Table 4.5. The implementation of each operating mode (and switching between them) should

be monitored through a building automation system that satisfies the ASHRAE guidelines (e.g., Guideline 13: Specifying Building Automation Systems). The heat load removed from the building zone using the cooling coil is labeled Q_{bldg} (this is the *Load* in the formulation (4.7)). The calculation of this quantity can be performed using detailed models of the building physics, occupant behavior, and weather effects or approximations/predictions of the heat load itself (see Section 4.7.2.5).

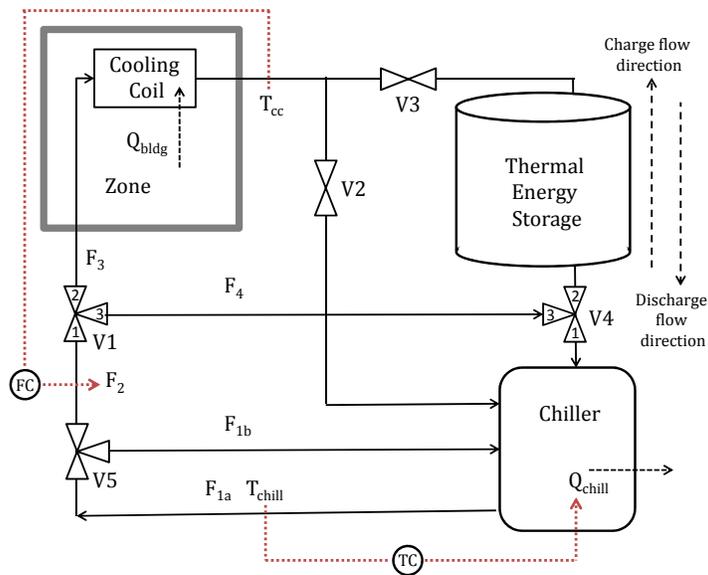


Figure 4.13: Flexible configuration of the coolant circuit in the University of Texas Thermal Facade Lab HVAC system and additional PI controllers (control loops are indicated by red dotted lines). Small dashed lines show heat transfer in/out of the system

Table 4.5: Operating Modes

Mode	V1	V2	V3	V4	Chiller	Cooling Coil
m_1 - Cooling Only	1-2	Open	Closed	Any	On	On
m_2 - Charging Only	1-3	Open	Open	3-2	On	Off
m_3 - Cooling & Charging	1-2&3	Open	Open	3-2	On	On
m_4 - Cooling & Discharging	1-2	Closed	Open	2-1	On or Off	On

A cylindrical storage tank packed with PCM-filled cylindrical elements is currently under construction at the TFL. A schematic is provided in Figure 4.14; chilled coolant supplied by the chiller flows through the interstitial space surrounding the encapsulated PCM. Energy is stored by removing latent heat from the material (i.e., freezing), and released by circulating warmer coolant from the coil outlet through the charged tank (which leads to melting the PCM).

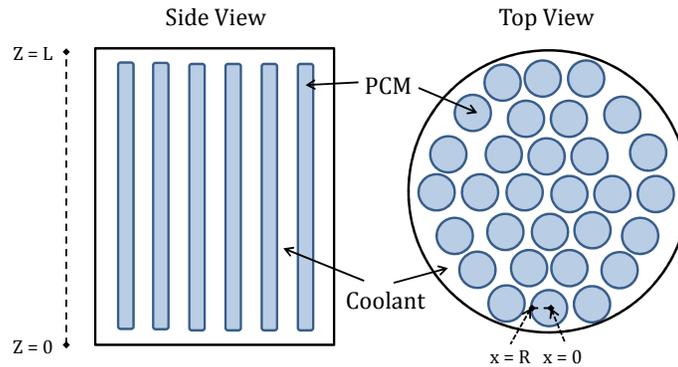


Figure 4.14: Configuration of the PCM TES tank.

While the primary focus of this paper is on *active* TES with PCM, it is worth noting that PCM can also be incorporated in a building's walls and air vents to increase the amount of *passive* TES. For both active and passive PCM storage, the selection of material properties (e.g., PCM melt temperature) and system geometry (e.g., size of the PCM elements and configuration of coolant flow around the PCM) are important design decisions [243, 303, 290, 180]. Several active storage medium geometries have been investigated in the literature. For example, [249] studies a PCM filled tank with a winding coil that contains the heat transfer fluid. [206] considers a large active PCM tank with encapsulated spheres, and references several other experimental efforts. The design choices affect the dynamics of the PCM behavior, which must be accounted for in a sophisticated controller for the system.

4.7.2 Equation-Oriented System Modeling

The parameters used in all of the following equations are listed in Table 4.8. The set of equations described here is general enough such that any coolant or PCM could be modeled by applying the appropriate parameter values, which may be established based on experiments or known physical properties. For this study, we have used the known physical properties of water and ice to specify the model parameters, and the tank has been sized such that the chiller can remain off for several hours even on a hot day. The properties of the TES tank being designed for the Thermal Facade Lab have recently been characterized in [37], which uses a paraffin for the PCM with a melting point around $5^{\circ}C$ and CPVC tubing for the

encapsulation.

4.7.2.1 Coolant Loop Model

Figure 4.13 describes the configuration the coolant loop of the Thermal Facade Laboratory. Two essential low-level control loops are implemented:

- The chiller temperature (T_{chill}) is regulated by a PI controller (TC in Figure 4.13) that manipulates the load Q_{chill} .

$$Q_{chill}(t) = Q_{chill}^0 + k_{D1}(T_{chill} - T_{chill}^{SP}) + k_{I1} \int_0^t (T_{chill} - T_{chill}^{SP}) \quad (4.17)$$

The setpoint is set to a value lower than the PCM melting temperature during the charging modes m_1 and m_3 (T_{chill}^{lowSP}) in order to freeze the PCM, and a slightly higher value in the other modes (T_{chill}^{highSP}):

$$T_{chill}^{SP} = (m_1 + m_3)T_{chill}^{lowSP} + (m_2 + m_4)T_{chill}^{highSP} \quad (4.18)$$

The flow rate of the coolant through the chiller (F_{1a}) is kept constant. Thus, the dynamics of T_{chill} depend on the mode of operation:

$$\begin{aligned} (\rho C_p)_c V_{chill} \frac{dT_{chill}}{dt} = & (\rho C_p)_c \left[F_{1b} T_{chill} + (m_1 + m_3) F_3 T_{cc} \right. \\ & \left. + (m_2 + m_3) F_4 T_{tank}(L) + m_4 F_3 T_{tank}(0) - F_{1a} T_{chill} \right] \end{aligned} \quad (4.19)$$

- In order to ensure that heat transfer between the building zone and the cooling coil is possible, a PI controller is used during modes m_1 and m_3 to regulate the temperature of the coolant exiting the cooling coil. The

temperature is maintained at the desired setpoint T_{cc}^{SP} by modulating the flow rate F_2 .

The PI controller is not used during the discharge mode m_4 . Instead, the flow rate F_2 is fixed during m_4 to $F_2^{discharge}$ (this parameter was selected based on the corresponding rate of freezing/melting the PCM). If the PI controller were used during the discharge mode, it is possible that the changes in the flow rate of coolant through the tank could hinder the effectiveness of the PCM storage (e.g., maintaining a set cooling coil temperature typically requires increasing flow rate F_2 , which reduces the amount of heat transfer from the PCM to the coolant during the discharge mode).

$$F_2 = (m_1 + m_3) \left(F_2^0 + k_{D2}(T_{cc} - T_{cc}^{SP}) + k_{I2} \int_0^t (T_{cc} - T_{cc}^{SP}) \right) + m_4 F_2^{discharge} \quad (4.20)$$

The resulting cooling coil dynamics are shown in Equation (4.21).

$$(\rho C_p)_c V_{cc} \frac{dT_{cc}}{dt} = (\rho C_p)_c F_3 (T_{chill} - T_{cc}) + (1 - m_2) Q_{bldg} \quad (4.21)$$

The values of the flow rates at other points in the cooling loop during each mode are summarized in Table 4.6, and reflect the valve positions listed in Table 4.5. We assume that in the combined charge and cool mode the coolant is equally divided between the coolant coil and the storage tank, so that $F_3 = F_4 = 0.5F_2$.

In the charge only mode m_2 , it is not possible to guarantee comfort levels inside the building because there is no heat removed by the cooling coil (see Equation (4.21)). For this reason, we will use the charge and cool mode (m_3) when freezing the PCM in the following examples instead of m_2 .

Table 4.6: Flow conditions. F_2 is calculated using the option indicated. The charge only mode (m_2) is not used in the current operation of the coolant loop.

Mode	F_2	F_3	F_4
m_1 - Cooling Only	PI	$= F_2$	$= 0$
m_3 - Cooling & Charging	PI	$= 0.5F_2$	$= 0.5F_2$
m_4 - Cooling & Discharging	$F_2^{discharge}$	$= F_2$	$= 0$

4.7.2.2 Modeling of PCM-based Active Thermal Energy Storage

In general, PCM-based TES systems consist of a tank filled with encapsulated PCM elements. The shape of the encapsulation shells (spheres, cylinders) and the layout of the elements within the tank are chosen to provide a high packing density (which, in turn, maximizes the energy storage density), while ensuring a uniform distribution of the coolant flow through the tank and appropriate surface area for heat exchange between the flowing liquid and the PCM elements.

In this work, without loss of generality, we consider a cylindrical storage tank that is packed with PCM-filled cylindrical rods. Water flows through the interstitial space surrounding the pipes (Figure 4.14). This geometry is based on the design in [36], which is currently being constructed at the Thermal Facade Laboratory. In developing the mathematical model of the tank, we assume that:

- the packing is arranged symmetrically with respect to the central axis of the tank, and the coolant is evenly dispersed into the free-flow area upon entering the tank
- the longitudinal temperature profiles in every coolant channel and PCM element are the same

- the conductivity of the shell of the PCM elements is high and does not limit heat transfer between the coolant and the PCM itself.
- the tank is well insulated and there are no heat losses to the environment

The tank has height L and total interstitial area A , and each of the N_{pcm} PCM elements has a radius R (Figure 4.14).

The equation-oriented model of this system is divided into two domains:

4.7.2.3 The Coolant Domain

The coolant temperature can be modeled as:

$$(\rho C_p)_c V_{tank} \frac{\partial T_{tank}}{\partial t} = -v \frac{\partial T_{tank}}{\partial z} + k_c \frac{\partial^2 T_{tank}}{\partial z^2} - Q_{pcm} \quad (4.22)$$

where the velocity is defined as a function of the operating mode as

$$v = (m_2 + m_3)F_4/A + m_4F_3/A \quad (4.23)$$

and V_{tank} represents the total volume of coolant in the tank. We assume that evolution of the temperature is dominated by convection because of the nonzero velocity in the interstitial areas and, as a consequence, we neglect the effect of thermal conduction in the coolant flow.

4.7.2.4 The PCM domain

Assuming that the thermal conductivity of the shells of the PCM elements is high enables us to focus our modeling efforts on the PCM domain. Here, we

rely on to a two-dimensional spatially distributed model to capture significant gradients present in the axial and radial directions. It is advantageous to model the PCM in terms of enthalpy rather than temperature, to avoid discontinuities in the temperature derivatives when phase change occurs. We construct the model below using a “mushy region” approximation of the interface between the solid and liquid phases, based on the developments in [237, 290, 6, 84].

We begin by describing the evolution of the enthalpy of the PCM (also see Figure 4.14):

$$\rho \frac{\partial H_{pcm}}{\partial t} = \frac{\partial}{\partial z} \left(k \frac{\partial T_{pcm}}{\partial z} \right) + \frac{1}{x} \frac{\partial}{\partial x} \left(kx \frac{\partial T_{pcm}}{\partial x} \right) \quad (4.24)$$

with boundary conditions:

$$T_{pcm}(z, R) = T_{tank}(z) \quad (4.25)$$

$$\frac{\partial T_{pcm}(z, 0)}{\partial x} = 0 \quad (4.26)$$

$$\frac{\partial H_{pcm}(0, x)}{\partial z} = 0 \quad (4.27)$$

$$\frac{\partial H_{pcm}(L, x)}{\partial z} = 0 \quad (4.28)$$

which specify that the temperatures of the coolant and PCM are the same at the PCM/coolant interface (4.25), capture symmetry in the radial direction (4.26) and, reasonably, assume that there is no heat transfer through the top and bottom of the PCM (Equations 4.27 and 4.28).

The rate of heat transfer between the coolant and PCM is computed as

$$Q_{pcm} = N_{pcm} \times k(z, R) \frac{\partial T_{pcm}(z, R)}{\partial x} \quad (4.29)$$

The phase of the PCM at each point in space is inferred from the corresponding enthalpy:

$$phase = \begin{cases} \text{solid} & \text{if } C_{p,s}T_m < H_{pcm}(z, x) \\ \text{phase change} & \text{if } C_{p,s}T_m - \Lambda < H_{pcm}(z, x) < C_{p,s}T_m \\ \text{liquid} & \text{if } H_{pcm}(z, x) < C_{p,s}T_m - \Lambda \end{cases} \quad (4.30)$$

Based on the phase, the temperature (T_{PCM}) and conductivity (k) at each point can be calculated as:

$$T_{pcm}(z, x) = \begin{cases} \frac{H_{pcm}(z, x)}{C_{p,s}} & \text{if solid} \\ T_m & \text{if phase change} \\ \frac{H_{pcm}(z, x) + (C_{p,l} - C_{p,s})T_m - \Lambda}{C_{p,l}} & \text{if liquid} \end{cases} \quad (4.31)$$

$$k(z, x) = \begin{cases} k_s & \text{if solid} \\ (1 - s(z, x))k_s + s(z, x)k_l & \text{if phase change} \\ k_l & \text{if liquid} \end{cases} \quad (4.32)$$

$$s(z, x) = \begin{cases} 1 & \text{if solid} \\ 1 - \frac{H_{pcm}(z, x) - C_{p,s}T_m}{\Lambda} & \text{if phase change} \\ 0 & \text{if liquid} \end{cases} \quad (4.33)$$

where it is assumed that the density of the PCM does not change significantly as a consequence of phase change, and the solid fraction $0 \leq s \leq 1$ is used to create a smooth (“mushy region” [6]) approximation of the evolution of thermal conductivity between the liquid and solid domains.

Furthermore, the state of charge for the storage system is defined using the total enthalpy of the PCM (4.36), the minimum total enthalpy (4.34), and the maximum total enthalpy (4.35):

$$H_{pcm}^{min} = \rho C_{p,s} T_{pcm}^{min} V_{pcm} \quad (4.34)$$

$$H_{pcm}^{max} = H^{min} + \rho C_{p,l} (T_{pcm}^{max} - T_m) V_{pcm} + \rho V_{pcm} \Lambda \quad (4.35)$$

$$H_{pcm}^{total} = \int_0^L \int_0^R H_{pcm}(z, r) 2\pi r dr dz \quad (4.36)$$

$$Charge = \frac{H_{pcm}^{max} - H_{pcm}^{total}}{H_{pcm}^{max} - H_{pcm}^{min}} \quad (4.37)$$

The temperatures T_{PCM}^{min} and T_{pcm}^{max} are selected such that the state of *Charge* in Equation (4.37) varies from 0 to 1. Note that the values of T_{pcm}^{min} and T_{pcm}^{max} are not equal to the melting point T_m because the storage can be cooled (or heated) below (or above) the melting temperature. Therefore the *Charge* metric includes both sensible and latent heat, although the majority of the storage capacity is associated with latent heat.

4.7.2.5 Building Load and Cost Model

We capture the building dynamics in the form of the thermal load profile Q_{bldg} , which aggregates the heat duty of the cooling system due to changes in weather, insolation, and the zone occupancy. This is effectively the load required to achieve thermal comfort in the building. Figure 4.15 shows two potential load profiles, representing high and low demand that might occur on, respectively, a hot or mild day. They were approximated based on the known insolation pattern for the UT Facade Lab, which has been simplified into well-defined steps to facilitate the interpretation of the simulation results. These profiles are representative of the load patterns typically seen in a commercial building, with occupants present (and thus and a larger cooling demand) during the daytime hours, and a flat base load over night.

The price schedules in Figure 4.16 provide the incentive for load shifting. We

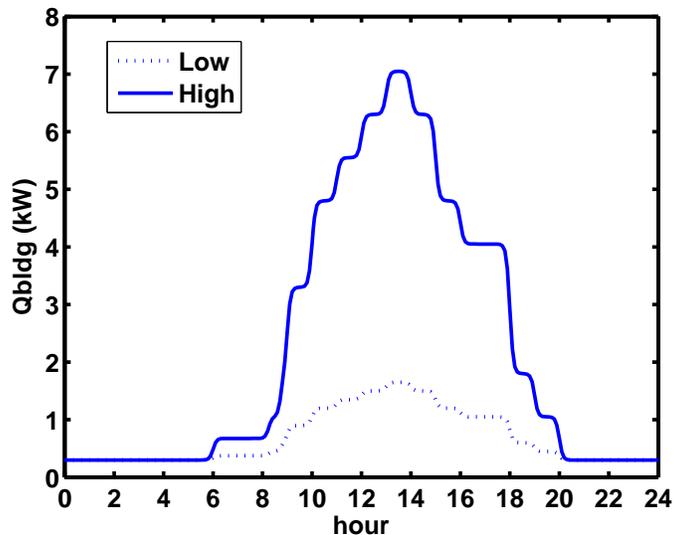


Figure 4.15: Building thermal load profile.

consider two cost profiles (presented on a relative basis); in the first case (a “two-hour peak”) the energy cost rises from the baseline by a factor of ten from hours 14 to 16. The second case has a “4 hour peak” price schedule where the cost increases during the hours 12 to 16. We will compare the effect of different load and cost scenarios on the optimal use of TES for this system.

4.7.3 Controller Design

Following the developments in Section 4.3, we define the operating sequence (4.38) to be used in formulating the optimal scheduling problem (4.39).

$$S = [\{m_1 \rightarrow m_3(\text{cool \& charge}) \rightarrow m_1 \rightarrow m_4(\text{discharge}) \rightarrow m_1\}, \quad (4.38)$$

$$\{t_1, t_2, t_3, t_4, t_5\}]$$

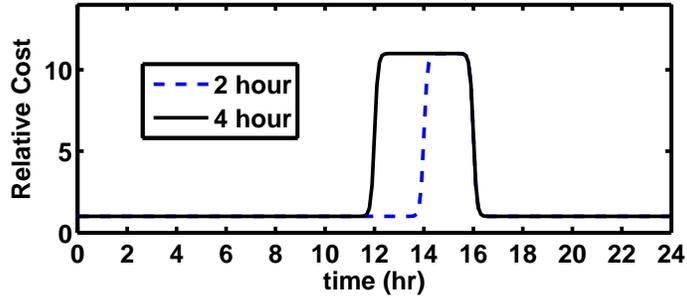


Figure 4.16: Real-time energy price profiles for a 24-hour time period. In the first case the energy cost rises from the baseline by a factor of ten from hours 14 to 16. The second case has a broader, four-hour peak price schedule, where the cost increases during the hours 12 to 16

$$\min_{t_1, \dots, t_5} J = \int_0^{24} C(t) Q_{chill}^2(t, S) dt \quad (4.39)$$

s.t. Storage tank model: Equations (4.22) – (4.37)

Coolant loop model: Equations (4.19) – (4.21) and Table 4.6

Q_{build} forecast: Figure 4.15

$C(t)$ forecast: Figure 4.16

Mode schedule in Equation (4.38)

$$\sum_{i=1}^5 t_i = 24$$

The formulation (4.39) follows the development of (4.7) in section 4.3.2.3 and is much simpler than the general problem (4.3). This can be attributed to the use of the Q_{bldg} profile for the modeling heat load on the coolant loop instead of a detailed model of the building and air-side of the HVAC system.

4.7.4 Results and discussion

In order to deal with the infinite-dimensional PCM model (4.24), we discretize the spatial domain using orthogonal collocation on finite elements, with four elements in the radial direction, and using central differences with 10 nodes in the axial direction. We note that modeling the coolant flow is a non-trivial matter because the direction of flow and velocity change when alternating between charging and discharging modes. This necessitates switching between backwards and forwards discretization schemes to solve the partial differential equation (4.22), depending on the operating mode. Following spatial discretization, we employ the sequential solution approach proposed by [277, 278] to solve the scheduling problem (4.39), using the software package gPROMS [217].

In order to evaluate the performance of the proposed energy management approach, we first define a baseline as the system performance (in terms of operating cost) under the time of use pricing scheme without the use of TES. In this case, the chiller is responsible for heat removal from the system at all times and will not turn off during the price peak. We will refer to this baseline operation as *base operation* in the sequel.

Second, we also define a *heuristic* to establish the timing of the charge/discharge events in (4.38). The operating guidelines are: charge the storage for 10 hours, and wait to begin using it until the price increases. Then, continue using the storage until it is depleted at the rate of discharge given by Table 4.6.

Finally, we present an analysis of the effect of incorrect forecasts on the effec-

tiveness of the optimal schedule. This entails simulating the implementation of the optimal schedule with a load profile that is different from the forecasted load used in the scheduling calculation. We also include a discussion on the effect of the system design, and limitations that this creates for the charge and discharge cycles.

4.7.4.1 Discussion: Optimal Scheduling vs. Heuristic Operation

Table 4.7 shows the values of the objective function (4.39) under the baseline, heuristic, and optimal control, along with the improvements made to the baseline under the latter two methods. Figure 4.17 shows the optimal schedule calculated for the four combinations of the load levels and pricing schemes (solid lines), and compares this result to the behavior under heuristic control (dashed lines).

In Figure 4.17, we observe that the total storage capacity is not used when the load is low (Figures 4.17a and 4.17b) and the storage is used outside of the peak pricing hours when possible. This lowers costs by not using the chiller during times when the *load* is high, in addition to times when the *price* is high. This behavior is especially evident in the extreme case of 4 hour peak pricing with a high load (Figure 4.17d) when it is not possible for the storage use to completely span the peak price period. Instead, the storage is used during the time when the load is highest. These features are not accounted for the by the heuristic, where delaying the use of the storage from the time of the price increase (Figure 4.17d) leads to a higher cost. The same is true for low load days, when the peak demand may occur before the price increase.

In Figures 4.17a and 4.17b, the discharge operation under the heuristic lasts much longer than the optimal schedule due to not using the storage during the peak load, and also overcharging the storage compared to the optimal case. There is also a small inverse response seen in the *Charge* metric towards the end of the discharge period. This is caused by the temperature of the return coolant from the cooling coil being lower than the PCM temperature, so the storage tank charge is slightly increased. However this results in the chiller requiring more energy to decrease the temperature of the coolant to its setpoint. Switching back to the normal operating mode before this occurs would be more cost effective.

In the optimal cases, there are instances where the chiller may turn on shortly before the end of the discharge operation (e.g., around hour 16 in Figures 4.17a, 4.17b, 4.17c). As the storage is depleted, the rate of heat exchange with the PCM decreases because the outer layers of the PCM elements have melted and the thermal conductivity of the liquid phase is lower than that of the solid. The chiller turns on during this period in order to maintain the temperature at its desired setpoint (the spike seen in Q_{chill} at the end of this period can be attributed to the controller tuning). Note, however, that the chiller load is still lower than the load in the baseline case during this period.

The optimal controller outperforms the heuristic, with lower total costs and a higher reduction to the baseline operating costs in Table 4.7. This does not mean that the heuristic is a poor controller - there are still substantial improvements over the baseline costs. However, it is possible that the heuristic may result in wasted energy by overcharging the storage tank when this is not necessary (Figure

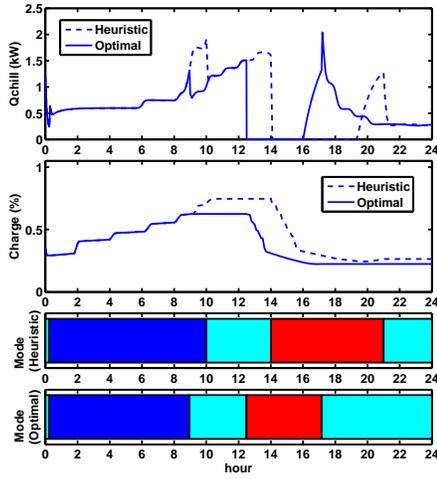
Table 4.7: Total cost (scaled by $1e7$) and the reduction of the baseline cost through the use of optimal control and heuristics

	Low Load 2hr Peak (Fig. 4.17a)	Low Load 4hr Peak (Fig. 4.17b)
Base	5.41	10.34
Optimal <i>Reduction</i>	1.24 <i>77%</i>	1.22 <i>88%</i>
Heuristic <i>Reduction</i>	1.83 <i>66%</i>	1.69 <i>84%</i>
	High Load 2hr Peak (Fig. 4.17c)	High Load 4hr Peak (Fig. 4.17d)
Base	89.37	177.59
Optimal <i>Reduction</i>	19.05 <i>79%</i>	64.05 <i>64%</i>
Heuristic <i>Reduction</i>	22.22 <i>75%</i>	73.56 <i>59%</i>

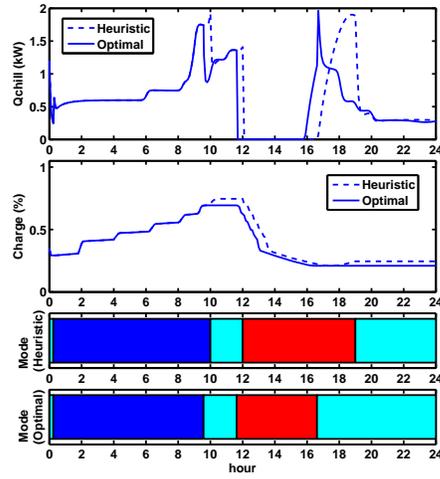
4.17a). We conclude that the optimized schedule is more effective than the heuristic because the system model provides an accurate insight into the behavior of the storage system, including PCM dynamics.

4.7.4.2 Assumptions and Potential Design Limitations

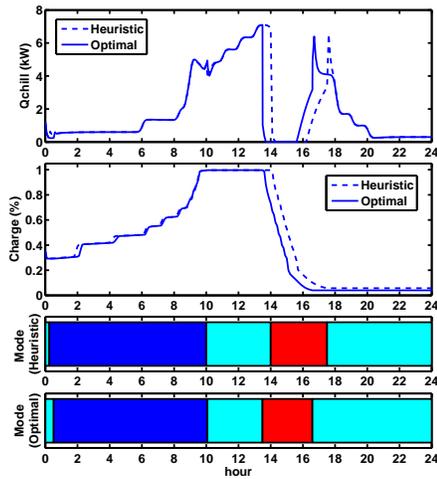
While a similar amount of time is spent charging the storage tank in both the low and high load cases of Figure 4.17, the *Charge* variable does not reach the same level. This can be attributed to the difference in the flow rate for charging between the two cases. Recall that the flow rate in the charge and cool mode is set by a PI controller (described in Section 4.7.2.1). This configuration results in the flow rate being proportional to the load needed to maintain the cooling coil exit



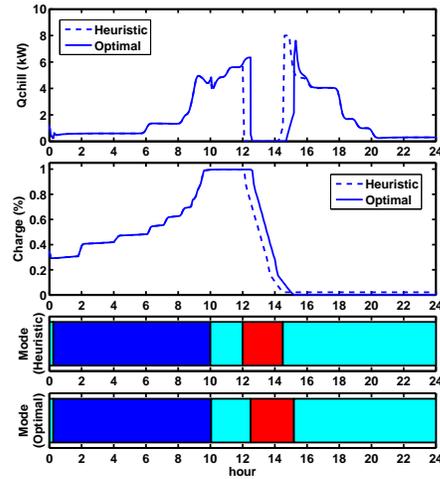
(a) Low load profile, 2 hour peak price schedule



(b) Low load profile, 4 hour peak price schedule



(c) High load profile, 2 hour peak price schedule



(d) High load profile, 4 hour peak price schedule

Figure 4.17: Comparison of optimal results and heuristic behavior for the two load profiles subject to the 2 and 4 hour peak price schedule (peak from hours 14-16 or 12-16). The chiller power, TES charge level, and active mode are shown for each optimal and heuristic controller. In the mode subplot, blue represents charging (m_3) and red discharging (m_4). All other periods are the normal operating mode (m_1).

temperature, which is not the same in the two load profiles considered. Therefore, the charge rate on the low load day is half of what it is on the high load day. In a real-life situation, more consideration would be given to the design of the flow controllers in the system to ensure that they minimize the charging time and maximize the discharging time regardless of the ambient conditions.

Also, notice that the initial condition for the storage *Charge* level for the optimal and heuristic cases in Figure 4.17 is not zero. We made the assumption of a non-zero initial storage level because, as can be observed on the low load days, it is possible that at the end of the day the storage amount will be not return to the original level. One situation where this behavior would be expected is when the prediction horizon is extended to two days and the prices on the first day are in general lower than those on the second day. In this case, it would be beneficial to overcharge the storage on the first day and save some stored energy for use on the second day. On the third day in this scenario, the storage would need to be charged starting from a completely empty state.

4.7.4.3 Effect of Forecasting Error

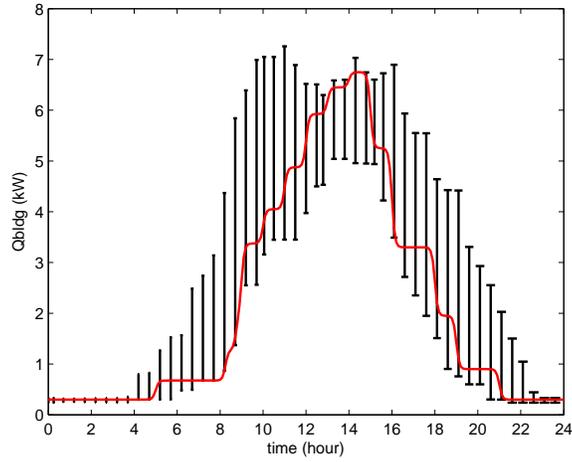
In practice, it is unlikely that the forecasts of the load used by the scheduling calculation will completely accurate. It is thus necessary to consider the effect of mismatch between the forecasted load and the load actually imposed on the system. We present this analysis for the high load (4 hour peak price) combination only as it is the most energy-demanding scenario.

Specifically, we consider a situation where the high load profile (Figure 4.15)

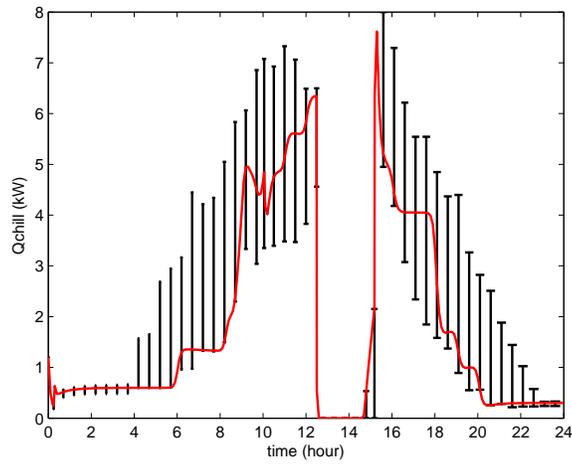
is used to determine the optimal schedule in Figure 4.17d. The actual load in the building is different when this schedule is implemented. Figure 4.18a shows a range of Q_{build} values (based on a sample set of ten different Q_{build} profiles) that are variations on the high load profile from Figure 4.15. The variations have a peak that occurs either before or after the forecast peak, and the magnitude may be about the same or lower than the forecast peak.

For each of the ten Q_{build} variations a new base cost (the cost when storage is not used) was calculated and the resulting sub-optimal cost reduction was calculated. These ranged from 29% to 63% with a median reduction of 55%. While lower than the optimal savings of 64% from Table 4.7, these values still indicate a substantial reduction in the energy costs under the peak pricing schedule.

Figure 4.18b shows the range of chiller demand under the different loads under the optimal schedule. Notice that there are some cases where the chiller is still off at the end of the discharge mode (i.e., the lower bound of the error bars is at 0 around hour 15). This indicates that the storage was not fully discharged, which means there was the potential to continue using it and have a higher cost reduction. Ultimately, this demonstrates the importance of rescheduling in the face of uncertainty. Because of the low computational times associated with the NLP reformulation of the scheduling problem, having a shorter rolling horizon (say, every 2 hours instead of once per day) is feasible.



(a) The high Q_{bldg} profile from Figure 4.15 (shown in red) was used to determine the optimal schedule. A set of ten alternate random Q_{bldg} profiles fall within the range described by the black vertical bars. The vertical bars show the maximum and minimum values of the ten alternate profiles at each time point.



(b) The Q_{chill} profiles when subject to the nominal schedule from Figure 4.17d. The red line corresponds to the nominal profile, while the results for the ten alternate load cases fall within the range shown by the black bars.

Figure 4.18: The effect of error in the Q_{bldg} forecast

Table 4.8: Case Study Parameters

Name	Description	Value
$(\rho C_p)_c$	Density and heat capacity of the coolant	2.00e6 J/m ³ K
ρ	Density of the PCM	1.00e3 kg/m ³
$C_{p,s}$	Solid PCM heat capacity	2.11e3 J/kgK
$C_{p,l}$	Liquid PCM heat capacity	4.18e3 J/kgK
Λ	Latent heat of PCM	333.00e3 J/kg
F_{1a}	Flow rate exiting the chiller	30.3 l/min
$F_2^{discharge}$	Flow rate in the coolant loop during the discharge mode	3.9 l/min
F_2^0	Nominal flow rate in the PI controller for F_2	11.4 l/min
k_s	Solid PCM conductivity	2.22 W/mK
k_l	Liquid PCM conductivity	0.56 W/mK
k_{D1}	Chiller PI controller proportional gain	300 kW/K
k_{I1}	Chiller PI controller integrator gain	500 kW/K
k_{D2}	Cooling coil PI controller proportional gain	0.50 kW/K
k_{I2}	Cooling coil PI controller integrator gain	0.70 kW/K
Q_{chill}^o	Nominal chiller heat removal rate	500 kW
$T_{chill}^{low SP}$	Chiller exit temperature low setpoint	260 K
$T_{chill}^{high SP}$	Chiller exit temperature high setpoint	279 K
T_{cc}^{SP}	Cooling coil exit temperature setpoint	284 K
T_m	Melt temperature of PCM	273 K
V_{chill}	Chiller volume	.00284 m ³
V_{cc}	Cooling coil volume	0.03 m ³
V_{tank}	Volume of the coolant flow channels in the tank	0.50 m ³
R	Radius of PCM cylinders	1.59 cm
L	Tank height	1.00 m
N_{pcm}	Number of PCM cylinders	100

4.8 Case Study: PCM TES Scheduling with a Low Order Model

In this example, we rely on a prototype system comprised of a heat pump - used for charging the storage and cooling the building - and a PCM TES as our basis

Table 4.9: Variable Nomenclature

Name	Description
$Charge(t)$	State of charge of the PCM tank
$C(t)$	Time-varying price schedule
$F_{1b}(t)$	Flow recycled to the chiller
$F_2(t)$	Flow before the split at valve 1
$F_3(t)$	Flow to the cooling coil
$F_4(t)$	Flow to the storage tank
$H_{pcm}(t, z, x)$	Enthalpy of the PCM
$k(t, z, x)$	Conductivity of the PCM
$m_i(t)$	Operating modes of the coolant loop (i=1,2,3,4)
$P_{HVAC}(t)$	Electricity consumption in the chiller
$Q_{chill}(t)$	Heat removed by the chiller
$Q_{pcm}(t, z)$	Heat transfer between the PCM and coolant
$Q_{bldg}(t)$	Load removed by the cooling from the building
$s(t, z, x)$	Solid fraction of the PCM
S	Cyclical schedule of operating modes
$T_{chill}(t)$	Chiller temperature
$T_{cc}(t)$	Cooling coil temperature
$T_{tank}(t, z)$	Temperature of coolant in the tank
$T_{pcm}(t, z, x)$	Temperature of the PCM
$T_{chill}^{SP}(t)$	Chiller setpoint temperature
$v(t)$	Velocity of coolant in the tank
t_0	Start time of the horizon
T	Prediction horizon length
T_p	Horizon of price forecast
T^{slow}	Prediction horizon in the TES scheduling calculation
T^{fast}	Prediction horizon in the Fast MPC

for formulating and solving the optimal operation problem for PCM TES systems. The system components are modeled based on industrial equipment included in a test facility currently under construction, and are assumed to be in use to meet the thermal requirements of a large commercial office building.

Based on the daily demands and total storage capacity, we delineate several classes of load characteristics and identify those where advanced decision-making and control can lead to economic and energy gains. A second contribution consists of formulating the optimal operation problem of the heat pump/PCM TES ensemble as a mixed integer dynamic optimization (MIDO), with the goal of minimizing total cost subject to comfort constraints. Our formulation captures the different potential operating modes (e.g., charging, discharging, and off) of the system via binary variables, which can considerably complicate the computation of an optimal solution. We propose a novel numerical solution strategy based on a sequential approach, showing that its computational implementation results in fast solution times with potential for real-time implementation. We present simulation results comparing optimal operation with different price schedule profiles typically encountered in the United States and European countries, and discuss the influence of energy price variability on the choice of operating strategy for the energy storage system.

4.8.1 System Description

4.8.1.1 Building

The system under consideration (which is representative of a 3000 square meter building currently under development in Lyon, France) includes a heat pump with a variable speed compressor and an electronically controlled expansion valve, a compact PCM storage device allowing the system to answer peak demands on short periods when demand may exceed the heat pump capacity, and a water chiller

linking the heat exchanger, the cold production group and the storage device. The coolant loop and water loop inside the building are shown in Figure 4.19.

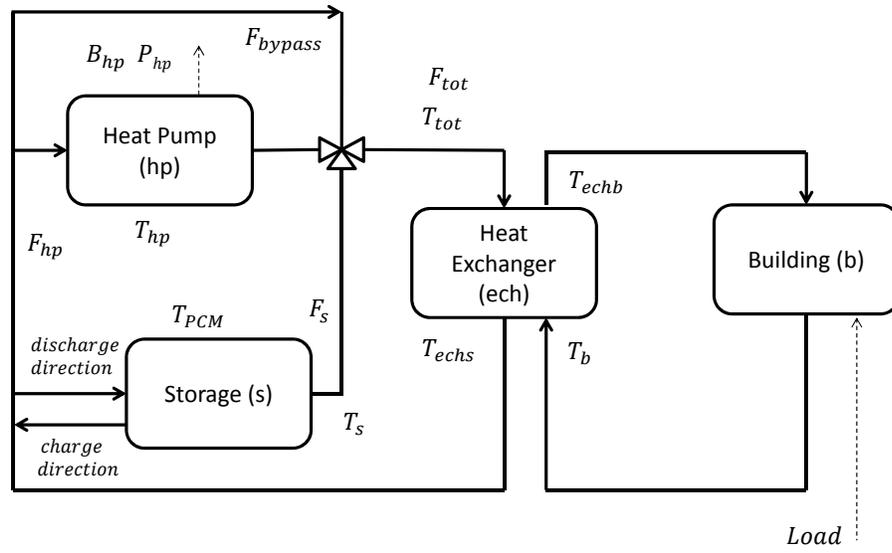


Figure 4.19: Coolant loops in the building system including the active PCM TES. The charging and discharging configurations are shown.

In this system, the flow configuration of the coolant loop and flow direction inside the PCM storage device changes depending on the operating mode (charging or discharging). Equations (4.40) to (4.46) model the coolant and water loops in the system. The building thermal load profiles ($Load$) are based on a detailed thermal simulation of the commercial building located in Lyon over one year with a one hour time step.

$$\rho_f C_{pf} V_{hp} \frac{dT_{hp}}{dt} = M_1 [F_{hp} C_{pf} (T_s - T_{hp} - P_{hp})]$$

$$+ M_2[F_{hp}C_{pf}(T_{echs} - T_{hp}) - B_{hp}P_{hp}] \quad (4.40)$$

$$\begin{aligned} \rho_f C_{pf} V_s \frac{dT_s}{dt} = & M_1[F_{hp}C_{pf}(T_{hp} - T_s) + K_p(T_{PCM} - T_s)] \\ & + M_2[F_s C_{pf}(T_{echs} - T_s) + K_p(T_{PCM} - T_s)] \end{aligned} \quad (4.41)$$

$$\begin{aligned} \rho_f C_{pf} V_{ech} \frac{dT_{echs}}{dt} = & M_2[F_{tot}C_{pf}(T_{tot} - T_{echs}) \\ & + K(T_{echs} - T_{echb})] \end{aligned} \quad (4.42)$$

$$\rho C_p V_{ech} \frac{dT_{echb}}{dt} = F_b C_p (T_b - T_{echb}) - K(T_{echs} - T_{echb}) \quad (4.43)$$

$$\rho C_p V_{bat} \frac{dT_b}{dt} = F_b C_p (T_{echb} - T_b) + Load(t) \quad (4.44)$$

$$F_{tot} = F_{hp} + F_s + F_{bypass} \quad (4.45)$$

$$\frac{dT_{tot}}{dt} = M_2 \left(-T_{tot} + \frac{F_{hp}T_{hp} + F_s T_s}{(F_{hp} + F_s)} \right) \quad (4.46)$$

The model states and control variables are described in Table 4.10. We note that, in the above model, flow configurations (modes) are represented mathematically using binary variables M_1 and M_2 . Possible modes are charging ($M_1 = 1$) and discharging ($M_2 = 1$), of which only one can be nonzero at any point in time (see (4.49)). The operation of the heat pump is defined in terms of a binary variable, B_{hp} , with the non-zero value denoting the “on” state of the pump and $B_{hp} = 0$ (in either charging or discharging mode) representing the system “off” state.

Table 4.10: Variables

States	Unit	Description
T_{hp}	$^{\circ}C$	Temperature of the heat pump
T_s	$^{\circ}C$	Temperature of the TES fluid
T_{PCM}	$^{\circ}C$	Temperature of the PCM
T_{tot}	$^{\circ}C$	Heat exchanger input temp. (source side)
T_{echs}	$^{\circ}C$	Heat exchanger output temp. (source side)
T_{echb}	$^{\circ}C$	Heat exchanger output temp. (building side)
T_b	$^{\circ}C$	Heat exchanger input temp. (building side)
H_{pcm}	J/kg	Enthalpy of the PCM
Decision Var.		
P_{hp}	kW	Cooling load removed by the heat pump
B_{bp}		Binary variable for heat pump on/off state
M_1, M_2		Binary variables for the charge/discharge configuration
F_s	kg/s	Storage flow rate in discharge mode
Disturbance		
Load	kW	Thermal load removed from the building

4.8.1.2 PCM Storage

We use a lumped approximation for the PCM in the storage system, modeling the evolution of the system total enthalpy as a function of time. The phase regime of the system (liquid, solid and melting/freezing) is defined as a function of enthalpy, and is used to select the appropriate temperature-enthalpy correlation.

$$M_n \frac{dH_{pcm}}{dt} = K_p (T_s - T_{pcm}) \quad (4.47)$$

$$T_{pcm} = \begin{cases} \frac{H_{PCM}}{C_{p,s}} & \text{if solid} \\ T_{melt} & \text{if phase change} \\ \frac{H_{PCM} + (C_{p,l} - C_{p,s})T_{melt} - L_f}{C_{p,l}} & \text{if liquid} \end{cases} \quad (4.48)$$

We chose to model the PCM enthalpy rather than temperature as a dynamic variable because enthalpy is a continuous function of temperature during the phase change process, whereas temperature is only a C^0 function (i.e., its time derivatives are discontinuous) [6].

We note here that, while the lumped approximation does not accurately capture the melt front observed during a phase change, it can provide a reasonable approximation to the overall dynamics of the system, which are of interest in deriving an optimal operation strategy. To this end, experimental data were used to tune the heat transfer coefficient K_p in (4.41) so that the times for melting and freezing approximated by the lumped model accurately represent the storage system.

4.8.1.3 Constraints

The constraint (4.49) guarantees that the optimization selects only one active mode at a time for the flow configuration.

$$M_1 + M_2 = 1 \quad \forall t \tag{4.49}$$

Bounds are imposed on the temperature of the PCM in the TES system and the heat pump:

$$T_{PCM}^{min} < T_{PCM} < T_{PCM}^{max} \tag{4.50}$$

$$T_{hp}^{min} < T_{PCM} < T_{hp}^{max} \tag{4.51}$$

The power level of the pump is defined by P_{hp} , which has a non-zero lower bound denoting that the pump should only operate in a range where its efficiency is

reasonably high.

$$P_{hp}^{min} < P_{hp} < P_{hp}^{max} \text{ if } B_{hp} = 1 \quad (4.52)$$

The water flow rate to the storage during the discharge mode is limited by (4.53). During the charge mode its value is fixed to the (constant) heat pump flowrate F_{hp} . This is automatically enforced through equation (4.41).

$$F_s^{min} < F_s < F_s^{max} \text{ if } M_2 = 1 \quad (4.53)$$

We assume that the required thermal load to be removed from the building in order to satisfy comfort constraints is known (this is the variable $Load$). We also assume that this load will always be satisfied by the devices in the water loop (hence, $Load$ is added directly to T_b in (4.44)). However, we must still include a constraint that guarantees all of the thermal load will be removed from the water loop so there is always a driving force for heat removal from the building. Therefore, we limit the building output temperature using (4.54).

$$T_b^{min} < T_b < T_b^{max} \quad (4.54)$$

4.8.2 Analysis of Operation Scenarios

We define a set of operating scenarios based on comparing the total daily demand to the total storage capacity available and to the maximum heat pump capacity. The following (reasonable) assumptions are used when defining each scenario:

- Approximate thermal load profiles are available for each day of the year with hourly measurements $Load(i)$.

- The daily total energy demand D can be found using the hourly load profiles for each day. $D = \sum_{i=1}^{24} 3600 Load(i)$.
- The maximum load that can be removed by the heat pump is $P_{hp}^{max} = 55kW$. There are days where the maximum $Load(i)$ value will exceed this limit.
- The maximum storage capacity is $E_{storage}^{max}$, and this has been sized based on the need to meet the load requirements on the most extreme day of the year, when the heat pump alone is not sufficient.
- To satisfy the temperature constraints on the system, nearly all of the building thermal load must be removed from the coolant loop using the heat pump and storage.

Based on the above, we define the operating scenarios:

Case 1: On *low days* the storage is sufficient to supply the total demand (4.55) and the heat pump is not needed.

$$D < E_{storage}^{max} \quad (4.55)$$

Case 2: On *average days* the storage is not able to provide enough power for the whole day (4.56) but the heat pump does not reach the power limit (4.57).

$$D > E_{storage}^{max} \quad (4.56)$$

$$Load(i) < P_{hp}^{max} \quad \forall i \quad (4.57)$$

Case 3: On *limited days* days, the storage cannot provide enough power for the whole day (4.56) and the power limit is reached by the heat pump.

$$\exists i \in [1 \dots 24] \text{ Load}(i) > P_{hp}^{max} \quad (4.58)$$

Case 4: On *extreme days*, all the same criteria for a limited day are met, and the total load exceeding P_{hp}^{max} is near the total storage capacity. We differentiate this case from the limited day because, as we will see below, it leads to a significant loss in degrees of freedom for the operation of the PCM TES system.

An example of each scenario is shown in figure 4.20. These profiles were obtained from a detailed simulation of the building involved in the project.

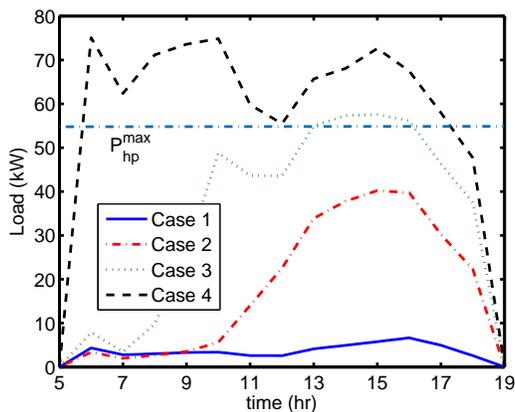


Figure 4.20: Types of load profiles, shown for the hours where *Load* is nonzero.

Figure 4.21 shows two potential cost profiles. The “peak schedule” is representative of a time of use (TOU) profile frequently encountered in the United States, where there is a sharp increase in the price during the late afternoon. The “flat schedule” is a day-night profile where there is a higher cost during the day and

a low cost overnight; this profile is more frequently encountered in countries in the European Union. The flat schedule is based on the rates at the French utility company (Electricité de France), while the peak schedule has been designed so that the total daily cost will be similar to that when subject to the flat schedule. We note that in both cases, the energy price fluctuations generally follow a natural trend in grid demand, i.e., the prices increase when grid demand increases (i.e., during daytime) and are low during low-demand periods (i.e., at night).

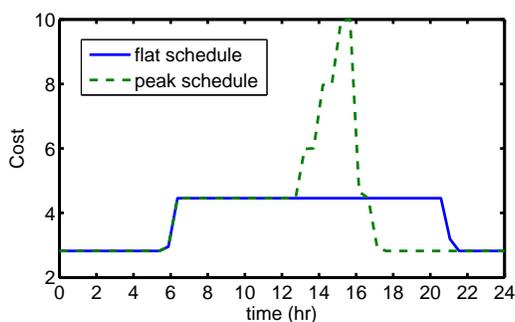


Figure 4.21: Comparison of electricity price profiles (cost in EU cents/kWh)

Based on the discussion above concerning load profiles and the corresponding operating scenarios, we make the following observations regarding the use of an advanced control system for the building under consideration.

- The control actions for a Case 1 day can be easily determined heuristically: since it is possible to meet all of the load requirements using the PCM TES system, the optimal charging and discharging pattern is to charge the storage at night and use as much as necessary during the day. This would be true regardless of the cost profile, and it is unlikely that the use of an advanced control system would significantly alter this operating strategy.

- Similarly, the operating schedule on the Case 4 day can be specified empirically. Since the storage system is sized to meet the amount of load in excess of the heat pump capacity, and the constraints on the system dictate the load must be satisfied, control actions are limited to operating the heat pump at or near full capacity and using the storage (which is charged during the previous night) to meet the excess load.

In light of the above, we will focus our efforts in the remainder of the paper on defining an optimal control strategy for Case 2 and Case 3 type days. In the climate scenario considered (Lyon, France), approximately 50 days of the year fall into these categories (the majority of the load profiles belong to Case 1). The number of such days is, however, much higher in a warmer climate, such as the (subtropical) climate of Austin, TX and many other locations in the Southwestern United States.

4.8.3 Optimal Operation Problem

The goal of the control system is to minimize operating cost by marshaling the operation of the PCM TES and heat pump in the face of fluctuations in load and energy prices described above. This can be formulated as an optimization problem, aimed at minimizing an objective function (4.59) that reflects the total cost of operating the heat pump over the time horizon T (which is typically 24 hours), subject to the system model and operating constraints introduced above:

$$\min_{M_1, M_2, B_{hp}, P_{hp}, F_s} \int_0^T (C(t)[B_{hp}(t)P_{hp}(t)])^2 \quad (4.59)$$

s.t. Model (4.40) - (4.48)

Constraints (4.49) - (4.54)

where $C(t)$ denotes electricity cost as a function of time (Figure 4.21).

The decision variables in (4.59) include the operating modes M_1, M_2 , heat pump variables B_{hp}, P_{hp} , and the storage flow rate F_s . Given a price profile, the optimization should determine the optimal charging/ discharging pattern for the thermal storage in addition to the other operating variables. In this work, we assume complete and accurate knowledge of the *Load* profiles so there is no uncertainty when making the control decisions. Note that modifications to traditional MPC formulations for building control can account for optimization in the face of uncertainty [169].

We assume that the cost of charging does not change from day to day, that charging happens at night and that it is a continuous process (in the sense that the TES system will not be charged during the day and there are no frequent switches from charging to discharging and back) [252]. Note that this does not mean that the storage system will be discharged continuously. It is possible for the flow rate of chilled water from the storage system to be set to zero even when in the discharge mode, so the timing of using the storage tank is not certain. Therefore, the time horizon T in (4.59) does not need to span the entire day. Rather, it only needs to span the period when the system is in the discharging configuration. At the minimum, this should include the period θ where the cost is higher than the base value and when *Load* is nonzero. Based on the above, we

propose the following simplification of problem (4.59), whereby we designate the times T' and t where $T' - t \geq \theta$ as the interval of interest, and set the operating mode of the system to discharge, i.e., $M_2 = 1$. This results in the following mixed integer dynamic optimization (MIDO) problem:

$$\begin{aligned} \min_{B_{hp}, P_{hp}, F_s} \int_t^{T'} (C(t)[B_{hp}(t)P_{hp}(t)])^2 & \quad (4.60) \\ \text{s.t. Model (4.40) - (4.48)} & \\ \text{Constraints (4.49) - (4.54)} & \\ M_1 = 0 & \\ M_2 = 1 & \end{aligned}$$

Equation (4.60) provides a complete description of the PCM TES and heat pump operating problem for the scenarios of interest. However, in view of the complications posed by integer decision variables (which we discuss further below), we also study a relaxation of this problem, which assumes that the heat pump is not subject to the minimum power constraint (4.52), allowing for a continuous transition between the off state ($P_{hp} = 0$) and the variable power in the on state. This formulation is shown below.

$$\begin{aligned} \min_{P_{hp}, F_s} \int_t^{T'} (C(t)[B_{hp}(t)P_{hp}(t)])^2 & \quad (4.61) \\ \text{s.t. Model (4.40) - (4.48)} & \\ \text{Constraints (4.49) - (4.54) excluding (4.52)} & \\ M_1 = 0 & \\ M_2 = 1 & \end{aligned}$$

$$B_{hp} = 1$$

We obtained the solution of (4.60) following the control vector parametrization approach [277, 278]. We divided the time horizon into 1 hour control intervals, using piecewise constant approximations for the continuous decision variables in each interval. We dealt with integer variables using the outer approximation proposed by Duran and Grossmann [82]. The time horizon spans the hours where the load is nonzero, so $t = 5$ and $T' = 19$. We implemented intermediate point (path) constraints of the form $y(t) < y_{lim}$ in section 4.8.1.3 using an integral transformation [278]:

$$\int_t^{T'} \max(0, y - y_{lim}) < \gamma \quad (4.62)$$

where $0 < \gamma$. To avoid discontinuities in (4.62), a differentiable approximation of the max function [132] was used:

$$\max(0, f(y)) \approx \frac{1}{2}[\sqrt{f(y)^2 + \beta^2} + f(y)] \quad (4.63)$$

4.8.4 Results

The optimization problem (4.60) (as well as the relaxation (4.61)) was implemented in gPROMS [217] and solved following the method described above. The system parameters are shown in Table 4.12. The objective values for all of the simulations are listed in Table 4.11. Results for the optimizations using Case 3 load profiles are discussed. We did verify the optimization for the Case 1, 2, and 4 profiles and the results were consistent with the assumptions in Section 4.8.2.

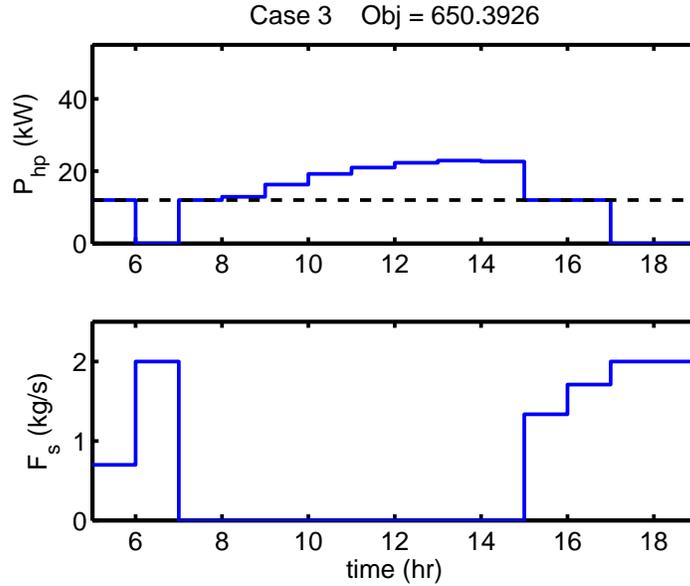


Figure 4.22: Results of the MINLP for the Case 3 load using the flat cost schedule. The dashed line shows the P_{hp}^{min} limit.

We begin by comparing the results to the MINLP formulation on a Case 3 load for the flat (Fig. 4.22) and peak (Fig. 4.23) cost profiles. In both cases, the storage is completely discharged by the end of the simulation time. As expected, more storage is used during the hours 13-16 in the peak profile case when the price is higher than the flat profile. Also, notice the TES is used at the beginning and end of the time horizon regardless of the price structure. This can be explained by comparing P_{hp} in Figure 4.23 to the result of the NLP optimization with a peak cost profile, shown in Figure 4.24. There are times when P_{hp} is lower than the minimum operating level in the constraint (4.52) (recall this constraint is relaxed in the NLP problem) and because of this extra freedom the objective for the NLP is lower than the MINLP. This is because a non-zero lower bound on the heat pump

power makes it more cost effective to use the storage, rather than the heat pump, to satisfy a low load requirement. Without a significant price peak, the storage is mainly used to satisfy high demands (hours 15-17) and very low demands that the heat pump cannot satisfy at as low of a cost as the storage. This demonstrates the benefits of using an optimal controller as opposed to a heuristic for managing the TES on days when there is significant freedom in the operating decisions - the controller behavior may change day-to-day depending on the load patterns.

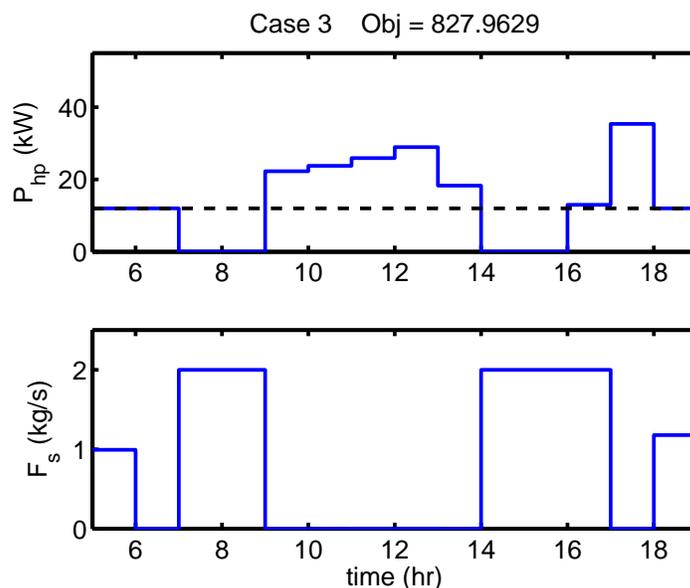


Figure 4.23: Results of the MINLP for the Case 3 load using the peak cost schedule. The dashed line shows the P_{hp}^{min} limit.

It is worth noting that both MINLP solutions were obtained in about 8 minutes on a desktop computer (exact times are in Table 4.11). This provides a significant incentive for real-time, recursive implementation of the proposed methodology.

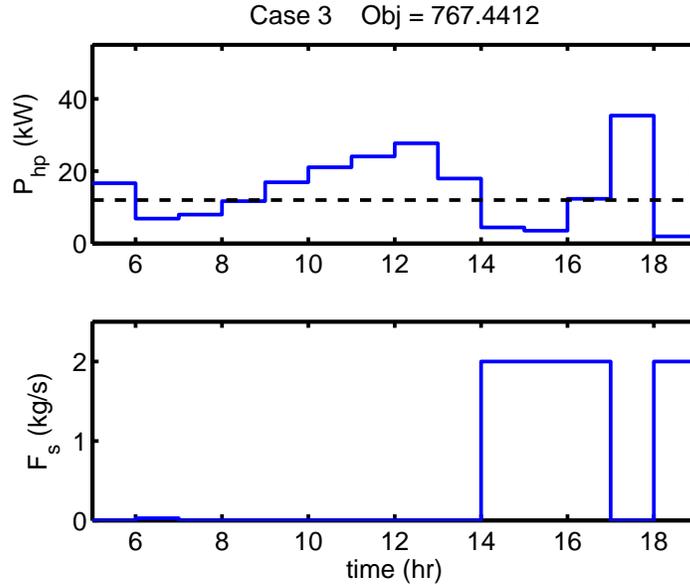


Figure 4.24: Results of the NLP for the Case 3 load using the peak cost schedule. The dashed line shows the P_{hp}^{min} limit, which has been relaxed.

Table 4.11: Objective Values and Solution Times

Load Case	Optim. Type	Cost Profile	Scaled Objective	Ob-jective	Computation Time	
3	MINLP	flat	650		455 sec.	(Fig. 4.22)
3	MINLP	peak	828		486 sec.	(Fig. 4.23)
3	NLP	peak	767		31 sec.	(Fig. 4.24)

4.9 Conclusion

Motivated by the ability to improve the operation of the electric grid through managing the demand patterns of buildings, we focused on devising optimal strategies for coordinating charging and discharging events of passive and active storage with the operation of the grid. We proposed a hierarchical energy management

Table 4.12: Parameters

Name	Value	Unit	Description
F_{hp}	2	kg/s	Heat pump flow rate
F_{tot}	4	kg/s	Total flow rate
F_b	2	kg/s	Building water loop flow rate
V_{ech}	0.05	m^3	Heat exchanger volume
V_s	2	m^3	Storage total volume
V_{hp}	.05	m^3	Heat pump volume
V_{bldg}	4	m^3	Building water volume
ρ	1000	kg/m^3	Water density
ρ_f	1061	kg/m^3	R410A density
C_p	4185	$J/kg K$	Water heat capacity
$C_{(p,f)}$	1690	J/kgK	R410A heat capacity
$C_{(p,nS)}$	2060	J/kgK	Solid PCM heat capacity
K_p	8400	W/K	UA value for the PCM
K	15000	W/K	Heat exchanger UA value
L_f	333000	J	Latent heat
M_s	$\rho_f V_s(0.4)$	kg	Mass of fluid in the storage
M_n	$\rho V_s(0.9)(0.6)$	kg	Mass of PCM in storage (60% efficient packing) accounting for PCM spheres only being about 90% full
P_{hp}^{min}	12	kJ	
P_{hp}^{max}	55	kJ	
F_s^{min}	0	kg/s	
F_s^{max}	2	kg/s	
T_b^{max}	17	$^{\circ}C$	
T_b^{min}	5	$^{\circ}C$	
T_n^{min}	-3	$^{\circ}C$	
T_{hp}^{min}	-30	$^{\circ}C$	
γ	10		
β	2		

strategy, based on controlling the use of active and passive TES in the building and HVAC system in different time scales. We showed that the operation of active TES systems must account for a longer time horizon, and formulated the optimal operation of active TES as an optimal scheduling problem. We proposed a continuous reformulation of this scheduling problem, that can be solved efficiently in a short amount of time, while accounting for the complexity and discontinuities associated with a detailed model of the phase change material-based storage. Extensive case studies for thermal storage using chilled water and phase change material demonstrated that our approach can achieve significant cost savings, outperform (reasonable) operating heuristics even under uncertainty in forecasting building loads, and have low solution times, showing real incentive for online implementation in a recursive, moving horizon fashion.

Chapter 5

The Effect of Distributed Electricity Generation Using Natural Gas on the Electric and Natural Gas Grid

The material in this chapter has been submitted for publication [258].

5.1 Introduction

Electricity and natural gas (NG) are supplied to residential, commercial, and industrial customers using complex networks that are in many ways interconnected. For example, natural gas fired power plants are used for load following on the electric grid [155]. Alternate methods for efficient electricity generation (e.g., CHP, microturbines, or fuel cells) have a range of operating capacities and can rely on NG as a primary energy source. It is possible to envision a future scenario where a large number of these units, ranging from the kW to MW production scale, are deployed in a region, thereby creating an increased dependency of the electric grid on the NG grid at multiple points along the transmission and distribution supply chains. This scenario appears as an attractive option for planning future residential or mixed-use communities under the assumption that natural gas prices remain low [161]. However, the potential changes in the interactions between the

electricity and NG grids that arise in this situation should be carefully evaluated.

In this work, we focus specifically on this issue, and address the question “*What is the consequence of tighter interdependency between the electric and NG grids through intersection points spanning utility-scale electricity generation through residential use of distributed generation?*” We develop several case studies illustrating the local requirements of both the electric and NG grids for a small neighborhood model considering multiple configurations of NG-based generators and various weather conditions and operating restrictions. Data for customer energy demands are based on national energy surveys and information provided by the Electricity Reliability Council of Texas (ERCOT). We formulate and solve the model for the operation of the ensemble of consumers and generators as a mixed-integer linear program, and analyze the resulting optimal energy usage profiles in terms of the ability to flatten the load placed by the host neighborhood on the electric grid.

5.2 Background

5.2.1 Overview of the Electric and Natural Gas Grids

The electricity and natural gas grids are both comprised of a transmission and distribution system, with components (e.g., substations for electricity and compression stations for NG) which ensure the safe and efficient delivery of energy to consumers. However, the similarities stop here. The dynamic characteristics of the electric grid are very different than those of the natural gas grid. The electric grid must operate with the supply and consumption of electricity in perfect balance;

large imbalances of electricity supply and demand lead to changes in voltage and frequency which cause (potentially catastrophic) failures [5]. On the other hand, natural gas is compressible and can therefore be stored in underground facilities or “linepacked.” Linepacking is the temporary increase in pressure in the distribution pipelines to act as a buffer for daily demands larger than the base load [268, 229]. Because natural gas can be easily stored, the natural gas grid has slower dynamics than the electric grid and it does not have to operate with a perfect instantaneous balance of supply and demand.

Planning for both short and long-term future production and demand is an important aspect of the operation of both grids for safety, reliability, and economic reasons [201]. It is particularly important for the electric grid because of the lack of cost-effective grid-scale storage (see [81] for a review of grid-scale battery technology). Consequently, the planning mechanisms for the two grids are structured differently.

The need to balance supply and time-varying demand on the electric grid has influenced the technology portfolio used by generation facilities and utility companies. Base load power generation (i.e., plants that run 24/7 and produce a constant amount of power) relies on hydro-electric, nuclear, coal, and natural gas combined cycle (NGCC) plants [269], while load-following power generation relies almost exclusively on natural gas-fired plants because of their ability to be ramped up and down quickly [155, 269, 150]. Natural gas-based electricity generation has lower greenhouse gas emissions than coal-fired power plants [124], so natural gas is a preferable fuel source for centralized electricity production for environmental

reasons. In the US, the fraction of electricity produced using NG has been steadily increasing over the past decade due to decreasing natural gas prices (attributed to the development of hydraulic fracturing of shale). In fact, approximately 35% of natural gas consumed in the US in 2014 was used for electricity generation, and 31% for industrial processes, while the remainder was used by residential and commercial consumers [270].

The variability of demand poses special challenges to grid operations. When the load on the grid is too high to be satisfied by typical generation resources, special “peaking plants” may be brought online for brief periods to meet the extreme demand level [269, 240]. In many cases, these are inefficient and tend to produce more CO₂ emissions than base load power plants. The desire to lower the peak load on the electric grid has led to the development of demand response (DR) and demand management (DM) programs, which encourage consumers to curtail or shift their electricity use from peak times [4, 202]. These programs are typically based on a time-dependent price schedule (e.g., time-of-use pricing) which penalizes consumers for electricity consumption during peak demand times. As a consequence, there has been a significant amount of research in how consumers should (or do) behave in such a scenario. For example, optimal scheduling of appliance usage [154] and the control of thermal energy storage in residential and commercial buildings [165, 253, 252] have been investigated as load-shifting strategies. Flattening electric grid load using batteries has been examined at both the grid scale [80] and locally at individual homes [35].

Another difficulty in operating the electric grid is the increased variability and

uncertainty in supply introduced by the increase in use of renewable generation technology, since their production levels are weather dependent [101, 207]. Natural gas based electricity generation can be used to mitigate the uncertainty of supply associated with the increased presence of renewable sources of electricity. This is –again– due to the flexibility natural gas provides for production to be ramped up or down quickly to satisfy short term load balancing needs. Distributed generation (discussed in the next section) can also aid in this effort.

5.2.2 Distributed Generation Technology

Distributed generation (DG) describes small-scale electricity generation technology located close to consumers, which leads to minimal losses associated with transmission (which can be up to 6% [272]). DG is a broad term and can refer to units producing from a few kW of electricity up to the MW scale [1]. In a similar vein, microgrids [125, 300] comprise a set of DG units and consumers, and have the ability to operate independently from the traditional electric grid. These are elements of the so-called 'smart grid' [240, 93], and aim to address the peak demand and intermittency issues associated with the incorporation of renewable energy sources in the generation mix at the grid level. DG can also be used as an ancillary service, which helps increase the reliability of the electric grid [131]. There is a large body of literature dedicated to the design and operation of DG technology in this context but, true to the scope of our paper, we will briefly review the NG-fired DG area.

Combined Heat and Power (CHP) and Combined Cooling Heating and Power

(CCHP) refer to electricity generation systems which also recover and use waste heat, increasing the overall efficiency. A detailed review of CCHP strategies is provided in [293], along with representative system designs – classified as micro-scale (under 20kW), small-scale (20kW - 1MW), medium-scale (1MW -10MW), and large-scale (over 10MW). Micro-CHP can be used by residential customers [72], while commercial customers may employ small-scale CHP [171]. Large-scale designs for planned communities and microgrids should also account for the presence of other energy sources, like photovoltaics (PV) [198, 300]. The “prime mover” for electricity generation in CCHP tends to be natural gas-based combustion turbines [293]. Solid Oxide Fuel Cells (SOFC) using natural gas as their fuel source are a promising technology for the small to large scale because of their high efficiencies and low environmental impacts [291, 56]. Hybrid micro-turbine and SOFC systems have also been proposed [68]. Micro-turbines (defined by [86] as under 500kW in size) are well-suited for DG applications because of their small size, high (greater than 80%) efficiencies, and flexible startup/load following capabilities [86]. In addition to being used for CHP, they can operate as a baseload or peak-shaving unit, or as a standby power source in case of emergencies [86]. Both micro-turbine based CHP systems and fuel cells used for distributed generation can have lower greenhouse gas emissions than centralized generation technology [246].

The design and operation of *individual* systems is a major focus of the literature on DG. The design and operation of single CHP systems were discussed, e.g., in [198, 73, 232], the design and operation of a grid-connected microgrid was considered in [107], and investment opportunities in DG for industrial facilities were

analyzed in [213]. Nearly all of the works mentioned consider capital investment and operating expenses. Another important research avenue is the optimal placement of DG units with respect to the distribution grid [98, 100, 283]. Available studies typically assume the generator acts as a baseload facility by a utility or grid operator. Several studies consider the implications of the use of DG resources on the electric grid stability (i.e., power quality and reliability) [2, 30] but do not focus on their potential for load-shifting applications or their cumulative effect on the dynamics of the electric and natural gas grids.

5.2.3 Points of Connectivity

We identify three different points of connectivity between the natural gas and electric grid, summarized in Figure 5.1. The first interconnection point between the natural gas and electric grids is at the supply side, where grid-scale power plants use natural gas (whether for load following or to help reduce renewable variability). Also, the natural gas grid relies on electricity for compressors that create the pressure gradients required to transport the gas through the pipeline network [92].

Interdependencies between the electric and natural gas grid should be understood in order to prepare for emergencies, where a small failure in one system may impact the other (see, e.g., the discussion in [200]). There are significant security and reliability concerns that arise from the interdependency of the supply of natural gas and the electric grid, and a large body of literature addresses this topic (e.g., [92, 118, 155]). Moreover, there has been a significant amount of research re-

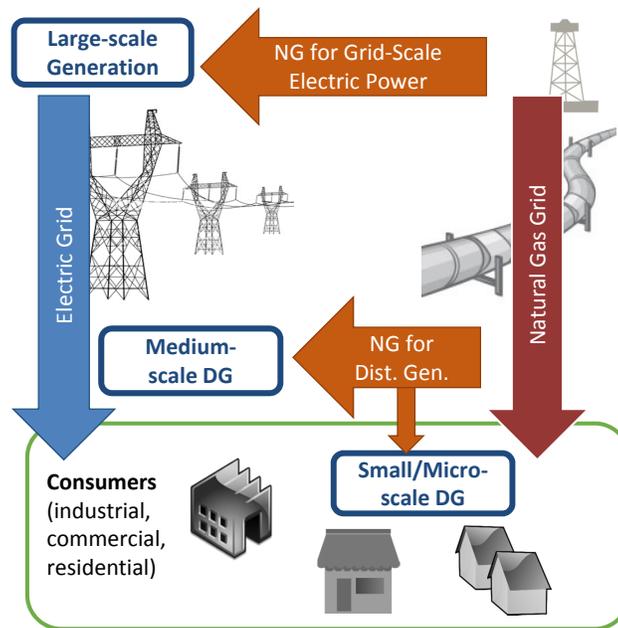


Figure 5.1: Points of connectivity between the electric and natural gas grids

garding the markets at this level, because it is important to capture the interaction between the energy sector and fuel availability [31]. For example, by coordinating the planning and scheduling of the electric and natural gas grids, market inefficiencies can be reduced and operating costs lowered [150]. This is itself a challenging undertaking owing to the multi-scale nature of the integrated system, with the dynamics of the electric grid being much faster than the evolution of the natural gas grid [118]. The simultaneous scheduling of power generation and NG supply (using detailed models of both electricity transmission and NG transport, e.g., as in [158]) has been demonstrated to improve the economic performance of the two systems [51, 55]. The literature review presented above shows that the current interest on the interdependency of the gas and electric grid is primarily focused

on the market considerations, and security and reliability concerns associated with the first point of (supply-side) connectivity.

The next point of connection is labeled “medium-scale DG” in Figure 5.1. Here, we are referring to distributed generation units with capacities $> 1\text{MW}$, such as CHP systems serving entire neighborhoods. These systems are likely operated by a utility and may still be connected to the distribution grid. Also, microgrids which still purchase some power from the central electric grid can be viewed similarly. Most of the literature reviewed in the previous section is dedicated to the design and operation of such systems for demand management, or considering their effect on the stability of the electric grid. Typically, these studies only consider a single generator serving a large group of consumers. They do not extend their analysis to the effects of many distributed CHP plants on the electric and natural gas grids.

Finally, small-scale distributed generation serving only a single consumer or small number of consumers is the last point of connectivity in Figure 5.1, and these DG units may even be located behind the meter for either/both electricity and natural gas (i.e., owned and operated by the consumer rather than a utility). Similar to the medium-scale DG, there is significant ongoing research in the design and operation of individual units. Recently, the interactions of a single small-scale generator with both grids was been investigated from a dynamic perspective [294].

The operation of an ensemble of small-scale distributed natural gas-based electricity generation units for load shifting, and their effect on the interdependency of the two grids has not been explicitly characterized in the literature, and will be considered in this paper. Specifically, we aim to examine (i) the effect of the

presence of several distributed generation units in a single area on the local grid load, (ii) the effect of distributed generation unit size on load patterns, and (iii) using the integration of the units within a residential community to flatten overall electric grid load.

To support this study, we develop a model of a future residential community that obtains electricity from the traditional electric grid and a set of small-scale ($< 50\text{kW}$) natural gas distributed generation units. We assume that the units are already in place, and do not need to consider capital investment. These units are used for local load following and peak load shifting for the electric grid. We will investigate the effect of the size of the natural gas distributed generation units and their ability to achieve a flat grid load within the limitations of the current NG grid infrastructure, while accounting for seasonal effects.

5.3 System Model

We develop a model of electricity and NG consumption for a residential neighborhood. In this model, a neighborhood is defined as a group of generators and the associated home consumers. Figure 5.2 shows a sample connectivity map of a single generator to the allocated consumers within the neighborhood. Each generator can only deliver electricity to its subset of consumers, while all consumers can receive electricity from the grid.

We assume that the neighborhood consists of 100 consumers (homes), and each consumer is connected to a single generator, the natural gas grid, and the electric grid. In Section 5.3.1 we present the mathematical description of the neighborhood

energy consumption and assumptions involved in developing it. This model will be used in an optimization formulation presented in Section 5.3.2 to establish the optimal operation of the distributed generation units under different scenarios. The parameters of the model (e.g., demand profiles, number of generators, assignment of houses to each generator) are discussed in Section 5.3.3.

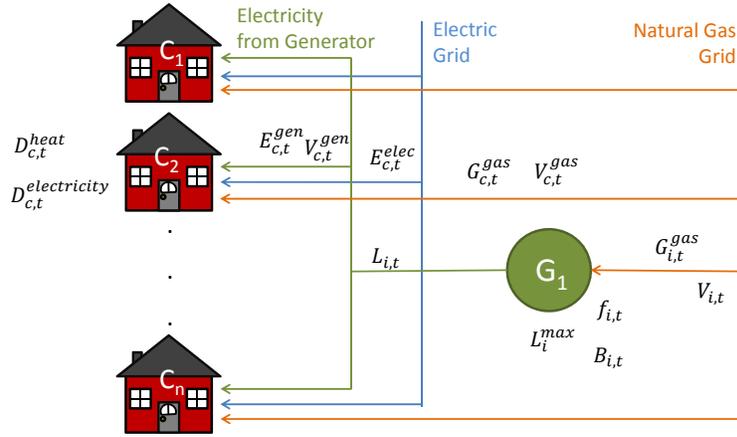


Figure 5.2: Structure of residential neighborhood grids with distributed generation units

5.3.1 Model formulation

The variable nomenclature and definitions are presented in Table 5.1. Each consumer $c \in C$ (where $c = 1 \dots N_c$ and C is the set of consumers) has a predicted total electricity demand for appliances and cooling $D_{c,t}^{electricity}$ and heating demand $D_{c,t}^{heat}$ at time t . Equation (5.1) states each of the consumers' total electricity demand must be satisfied using electricity from the grid $E_{c,t}^{elec}$ and the generator $E_{c,t}^{gen}$.

$$D_{c,t}^{electricity} = E_{c,t}^{elec} + E_{c,t}^{gen} \quad \forall c, t \quad (5.1)$$

Table 5.1: Model Parameters and Variables

Indices and Sets	
$c \in C$	Consumers
$S_i \subseteq C$	Consumers mapped to generator i
$i \in I$	Generators
$t \in T$	Time steps
Parameters	
Δt	Time interval of each time step (0.25 hr)
f_i^{min}	Minimum fractional operating level of natural gas generator (0.8)
f_i^{max}	Maximum fractional operating level of natural gas generator (1.0)
L_i^{max}	Load capacity of NG generator (kW)
ϵ_i	Efficiency of NG generator ($\frac{BTU}{kW hr}$)
$D_{c,t}^{heat}$	Consumer demand for heating (kW)
$D_{c,t}^{electricity}$	Consumer demand for electricity (kW)
HHV	High heating value ($\frac{BTU}{ft^3}$)
$NG^{Flow Cap}$	Volumetric flow rate capacity ($120 \frac{ft^3}{hr}$)
$NG^{Tot Cap}$	Volumetric consumption capacity ($35,880 \frac{ft^3}{hr}$)
M	Constant used in 'big M' reformulations
Continuous Variables	
$E_{c,t}^{elect}$	Electricity used by each consumer from the electric grid (kW)
$E_{c,t}^{gen}$	Electricity used by each consumer from the distributed generation (kW)
$G_{c,t}^{gas}$	Gas used by each consumer for heating (kW)
$L_{i,t}$	Electricity produced by each generator (kW)
$f_{i,t}$	Generator operating level (-)
$V_{i,t}$	Flow rate of NG consumed by each generator ($\frac{ft^3}{hr}$)
$V_{c,t}^{gen}$	Flow rate of natural gas attributed to each consumer for distributed generation ($\frac{ft^3}{hr}$)
$V_{c,t}^{gas}$	Flow rate of natural gas attributed to each consumer for heating ($\frac{ft^3}{hr}$)
$V_{c,t}^{tot}$	Total flow rate of natural gas attributed to each consumer for heating and distributed generation ($\frac{ft^3}{hr}$)
$E_t^{gridTot}$	Total electricity used by the entire neighborhood
E^{excess}	Electric load on the grid relative to a target load
α	Target electric grid load (kW)
N_{events}	Number of on/off events
Binary Variables	
$B_{i,t}$	NG generator ON (1) or OFF (0)

We initially assume that the entire demand for heating is met by gas supplied from the natural gas grid $G_{c,t}^{gas}$ (i.e., no electric heating). This is represented in Equation (5.2).

$$D_{c,t}^{heat} = G_{c,t}^{gas} \quad \forall c, t \quad (5.2)$$

The amount of electricity produced by generator $i \in I$ is divided among its subset of consumers, $S_i \subseteq C$. The sum of all the electricity used by the consumers within a generator's consumer subset is equal to the total amount of electricity produced, $L_{i,t}$, by generator i at time t . Equation (5.3) represents this division of generator electricity among the consumers. Note that we do not impose any constraints on this division; it is therefore possible to have an unequal split of electricity among the consumers assigned to a generator.

$$L_{i,t} = \sum_{c \in S_i} E_{c,t}^{gen} \quad \forall i, t \quad (5.3)$$

The operating state of each generator is defined by a binary variable $B_{i,t}$, which represents whether generator i is on (i.e., $B_{i,t} = 1$) or off (i.e., $B_{i,t} = 0$) at time t , and the fractional operating level $f_{i,t}$, which represents the operating level with respect to the maximum electricity production L_{max} . Equation (5.4) defines the amount of electricity produced by generator i .

$$L_{i,t} = f_{i,t} L_i^{max} B_{i,t} \quad \forall i, t \quad (5.4)$$

The generator operating level $f_{i,t}$ is constrained, as shown in Equation (5.5), so that each generator operates efficiently when turned on. We specify f^{min} and

f^{max} so that generators can operate at 80-100% of their maximum capacity.

$$f^{min} \leq f_{i,t} \leq f^{max} \quad (5.5)$$

For practical reasons, the generators should not repeatedly turn on and off since this causes excessive wear. Equations (5.6) and (5.7) impose that generators remain on for at least an hour and a half once they are turned on, and stay off for at least an hour once they are turned off, respectively.

$$N_{on}(B_{i,t} - B_{i,t-1}) \leq \sum_{\theta=1}^{N_{on}} B_{i,t+\theta} \quad N_{on} = \frac{1.5}{\Delta t} - 1 \quad \forall i, t \quad (5.6)$$

$$N_{off}(B_{i,t-1} - B_{i,t}) \leq N_2 - \sum_{\theta=1}^{N_{off}} B_{i,t+\theta} \quad N_{off} = \frac{1}{\Delta t} - 1 \quad \forall i, t \quad (5.7)$$

Every generator load has an associated volumetric flowrate $V_{i,t}$ of natural gas that is calculated based on the generator efficiency ϵ_i and the generator fractional operating level. Equation (5.8) provides the total natural gas flow rate required for generator i to produce the required amount of electricity under its current operating level, based on the energy content of natural gas (high heating value $HHV = 1027 \frac{Btu}{ft^3}$).

$$V_{i,t} = \frac{L_{i,t}\epsilon_i}{HHV} \quad \forall i, t \quad (5.8)$$

The portion of natural gas used by a generator attributed to each of the connected consumers, $V_{c,t}^{gen}$, is based on the consumer's electricity consumption from the generator $E_{c,t}^{gen}$.

$$V_{c,t}^{gen} = \frac{E_{c,t}^{gen} B_{i,t}\epsilon_i}{HHV} \quad \forall i, c \in S_i, t \quad (5.9)$$

Additionally, the volumetric flow that each consumer uses for gas heating is defined as:

$$V_{c,t}^{gas} = \frac{G_{c,t}^{gas}}{HHV} \quad \forall c, t \quad (5.10)$$

so that the total flow of natural gas used by each consumer $V_{c,t}^{tot}$ is:

$$V_{c,t}^{tot} = V_{c,t}^{gen} + V_{c,t}^{gas} \quad \forall c, t \quad (5.11)$$

The current pipeline infrastructure may limit the amount of natural gas that can be delivered to the generators. The pipe diameter at various points along the supply chain imposes a restriction on the flow rate of natural gas. To account for this possibility, equations (5.12) and (5.13) constrain the flow rate and total consumption, respectively, of a generator i .

The parameter $NG^{Flow\ Cap}$ is determined by pipe material, length, and pressure drop. In practice, these properties may vary greatly among consumers. For our model, we have selected a fairly conservative maximum flow rate of $120\ ft^3/hr$ for the parameter $NG^{Flow\ Cap}$ for all generators (this is reasonable for a maximum allowable pressure of 0.5psi). $NG^{Tot\ Cap}$ is based on the maximum observed gas use in the winter, when demand for gas is at its peak. We assumed a value of 115% of the natural gas consumption in an average Texas household during the winter ($NG^{Tot\ Cap}$), which is $35,880\ ft^3/day$ for a group of 100 consumers (this metric was based on the average monthly natural gas residential consumption for Texas[273] and average number of residential consumers [271]).

$$\sum_{c \in S_i} V_{c,t}^{gen} \leq NG^{Flow\ Cap} \quad \forall c, t \quad (5.12)$$

$$\sum_{all\ t} \sum_c V_{c,t}^{gen} \Delta t \leq NG^{Tot\ Cap} \quad \forall c \quad (5.13)$$

The main decision variables when establishing the optimal operation of generators are $f_{i,t}$ and $B_{i,t}$. Note that equations (5.4) and (5.9) contain bilinear terms (comprising the product of two variables, one binary and the other continuous). A linear model is preferred to simplify the optimization calculation. Consequently, we transform the bilinear terms into a set of equivalent linear inequalities using a “big M” technique. Equation (5.4) is rewritten as:

$$L_{i,t} \geq f_{i,t} L_i^{max} - M_1(1 - B_{i,t}) \quad \forall i, t \quad (5.14a)$$

$$L_{i,t} \leq f_{i,t} L_i^{max} + M_1(1 - B_{i,t}) \quad \forall i, t \quad (5.14b)$$

$$L_{i,t} \leq M_1 B_{i,t} \quad \forall i, t \quad (5.14c)$$

and Equation (5.9) is rewritten as:

$$V_{c,t}^{gen} \geq \frac{E_{c,t}^{gen} \epsilon_i}{HHV} - M_2(1 - B_{i,t}) \quad \forall c, t \quad (5.15a)$$

$$V_{c,t}^{gen} \leq \frac{E_{c,t}^{gen} \epsilon_i}{HHV} + M_2(1 - B_{i,t}) \quad \forall c, t \quad (5.15b)$$

$$V_{c,t}^{gen} \leq M_2 B_{i,t} \quad \forall c, t \quad (5.15c)$$

We set M_1 and M_2 to a value 10% greater than the maximum generator load to ensure a tight formulation of the optimization problem.

5.3.2 Optimization Problem Formulation

The purpose of the set of distributed generators is to improve the overall operation of the electric grid. To this end, we aim to flatten the load on the electric grid

over the course of a day, which will help reduce the peak load. We will consider a 24 hour time horizon with $\Delta t = 15$ minute time steps. The optimization calculation will be used to establish the operating pattern of the generators such that a flatter load profile is achieved. Rather than minimizing the peak load directly (e.g., as in [67]), we seek to minimize the excess energy, E_{excess} , defined as the difference between the grid load and a target load, α , over the time horizon considered.

$$E^{excess} = \sum_{all\ t} |E_t^{grid\ Tot} - \alpha| * \Delta t \quad kWhr \quad (5.16)$$

Note that α is also a decision variable and we do not penalize any particular α value. In other words, we are not concerned with the magnitude of α but rather focus on imposing an effective use of the generation capacity. This includes reducing the number of times the generators turn on/off to avoid unnecessary wear on the units (this was also discussed in relation to equations (5.6) and (5.7)). Equation (5.17) is used to determine N^{events} , the number of times the set of generators either turn on or off.

$$N^{events} = \sum_{all\ i} \sum_{t=2}^{T_{bill}} |B_{i,t} - B_{i,t-1}| \quad (5.17)$$

We will include this term in the optimization objective, with the added benefit of tightening the formulation (owing to the large number of decision variables, which is equal to the number of time steps \times number of generators \times number of decision variables, degenerate solutions may arise, indicating that are multiple ways to reach the same excess energy level).

The model is examined in three different cases: an unconstrained case, a constrained case, and a decentralized case. We evaluate these cases for different sea-

sonal demand profiles and configurations of gas vs. electric heating. The objective is to minimize the excess energy observed on the electric grid and the number of on/off events. The two objectives are weighted (the weighting factors are W_{excess} and W_{events}) to create a single objective function. The choice of weighting factors will be discussed after the optimization problem formulations are presented below.

Unconstrained case

The unconstrained case is the base case for establishing the operation of the generator units, with no limit on the amount or on the rate of consumption of natural gas. This is a centralized approach: the operation of all generators over the course of the day is assumed to be coordinated centrally and hence established simultaneously. Equations included in the optimization problem for the unconstrained case are described in the formulation (5.18):

$$\begin{aligned}
 \min_{B_{i,t}, f_{i,t}, \alpha} \quad & W_{excess} E^{excess} + W_{events} N^{events} \\
 \text{s.t.} \quad & (5.1), (5.2) \quad \text{Demand definition and constraint} \\
 & (5.3) \quad \text{Relate generator load to electricity used by consumer} \\
 & (5.5) \quad \text{Limitations on generator operating level} \\
 & (5.6), (5.7) \quad \text{Minimum amount of time to remain on/off} \\
 & (5.8), (5.11) \quad \text{Natural gas flow rate to each generator and consumer} \\
 & (5.10), (5.15) \quad \text{Total natural gas used for heating and electricity} \\
 & \quad \quad \quad \text{per consumer} \\
 & (5.14) \quad \text{Determine generator load} \\
 & (5.16), (5.17) \quad \text{Define variables in objective}
 \end{aligned}$$

Constrained Case

The unconstrained case is idealistic since it does not account for limitations due to the current NG distribution infrastructure. In the constrained case, we include

Equations (5.13) and (5.12), leading to the problem (5.18):

$$\begin{array}{ll}
 \min_{B_{i,t}, f_{i,t}, \alpha} & W_{excess} E^{excess} + W_{events} N^{events} \\
 \text{s.t.} & (5.1), (5.2) \quad \text{Demand definition and constraint} \\
 & (5.3) \quad \text{Relate generator load to electricity used by consumer} \\
 & (5.5) \quad \text{Limitations on generator operating level} \\
 & (5.6), (5.7) \quad \text{Minimum amount of time to stay on/off} \\
 & (5.8), (5.11) \quad \text{Natural gas flow rate to each generator and consumer} \\
 & (5.10), (5.15) \quad \text{Total natural gas used for heating and electricity} \\
 & \quad \quad \quad \text{per consumer} \\
 & (5.12), (5.13) \quad \text{Infrastructure volumetric flow rate and consumption} \\
 & \quad \quad \quad \text{constraints} \\
 & (5.14) \quad \text{Determine generator load} \\
 & (5.16), (5.17) \quad \text{Define variables in objective}
 \end{array}$$

Decentralized Case

In the constrained and unconstrained cases, centralized decision making is implied (the operation of all the generators is determined in a single calculation which would be performed on a centralized computer, likely by the utility). This may not be feasible in practical situations. For example, consider a scenario in which the generators are owned and operated by the consumers instead of a utility company. In this case, the consumers would be less likely to coordinate their actions or willing to relinquish control of their assets, and instead focus on minimizing their own utility bills.

In order to understand how centralized operation affects the performance and grid flattening objective, we compare the two scenarios already described to a decentralized optimization which considers a single generator, so that the set of consumers involved is $S_i \in C$. Because the formulation involves only a single

generator, excess energy is redefined as:

$$E_i^{excess} = \sum_{all\ t} \left| \sum_{c \in S_i} E_{c,t}^{elec} - \alpha_i \right| * \Delta t \quad kWhr \quad (5.18)$$

and equation (5.17) is redefined as:

$$N_i^{events} = \sum_{t=2}^{T_{bill}} |B_{i,t} - B_{i,t-1}| \quad (5.19)$$

The decentralized optimization formulation is unconstrained (i.e., similar to the formulation (5.18) because it does not include the constraints based on current infrastructure as in (5.18)), and described in Equation (5.20) below. In this case, the optimization will have to be solved once for each generator i .

$$\begin{aligned} \min_{B_{i,t}, f_{i,t}, \alpha} \quad & W_{excess} E_i^{excess} + W_{events} N_i^{events} \\ \text{s.t.} \quad & (5.1), (5.2) \quad \text{Demand definition and constraint} \\ & (5.3) \quad \text{Relate generator load to electricity used by consumer} \\ & (5.5) \quad \text{Limitations on generator operating level} \\ & (5.6), (5.7) \quad \text{Minimum amount of time to stay on/off} \\ & (5.8), (5.11) \quad \text{Natural gas flow rate to each generator and consumer} \\ & (5.10), (5.15) \quad \text{Total natural gas used for heating and electricity} \\ & \quad \quad \quad \text{per consumer} \\ & (5.14) \quad \text{Determine generator load} \\ & (5.18), (5.19) \quad \text{Define variables in objective} \end{aligned} \quad (5.20)$$

Note that when we present the results of the decentralized case, the sum of all the individual generators' operation is used to analyze the effect of decentralized operation relative to centralized operation.

Electric Heating Case

The previous cases assume that consumers only rely on gas heating. When the consumers are assumed to have electric heating, equations (5.1) and (5.2) are extended to:

$$D_{c,t}^{electricity} + D_{c,t}^{heat} = E_{c,t}^{elec} + E_{c,t}^{gen} \quad \forall c, t \quad (5.21)$$

$$0 = G_{c,t}^{gas} \quad \forall c, t \quad (5.22)$$

Equation (5.10) is no longer needed so Equation (5.11) becomes:

$$V_{c,t}^{tot} = V_{c,t}^{gen} \quad \forall c, t \quad (5.23)$$

The equations included in the optimization formulation of the unconstrained electric heating case are described in Equation (5.24).

$$\begin{array}{ll} \min_{B_{i,t}, f_{i,t}, \alpha} & W_{excess} E^{excess} + W_{events} N^{events} \\ \text{s.t.} & (5.21), (5.22) \quad \text{Demand definition and constraint} \\ & (5.3) \quad \text{Relate generator load to electricity used by consumer} \\ & (5.5) \quad \text{Limitations on generator operating level} \\ & (5.6), (5.7) \quad \text{Minimum amount of time to stay on/off} \\ & (5.8), (5.23) \quad \text{Natural gas flow rate to each generator and consumer} \\ & (5.15) \quad \text{Find total natural gas used for electricity per consumer} \\ & (5.14) \quad \text{Determine generator load} \\ & (5.16), (5.17) \quad \text{Define variables in objective} \end{array} \quad (5.24)$$

Objective Function Weighting Factor

Different ratios of weighting factors were examined to select values for the objective function used in the case studies. Figure 5.3 shows the resulting excess energy and number of events for the solution of the unconstrained formulations for

both gas and electric heating when the weighting factors are varied. It is possible to achieve very low excess energy levels when E^{excess} is heavily weighted, however there was a large number of on/off events and we observed several degenerate solutions. For this reason, we have selected weighting factors of $W_{excess} = 1$ and $W_{events} = 25$ ($W_2/W_1 = 25$ in Figure 5.3).

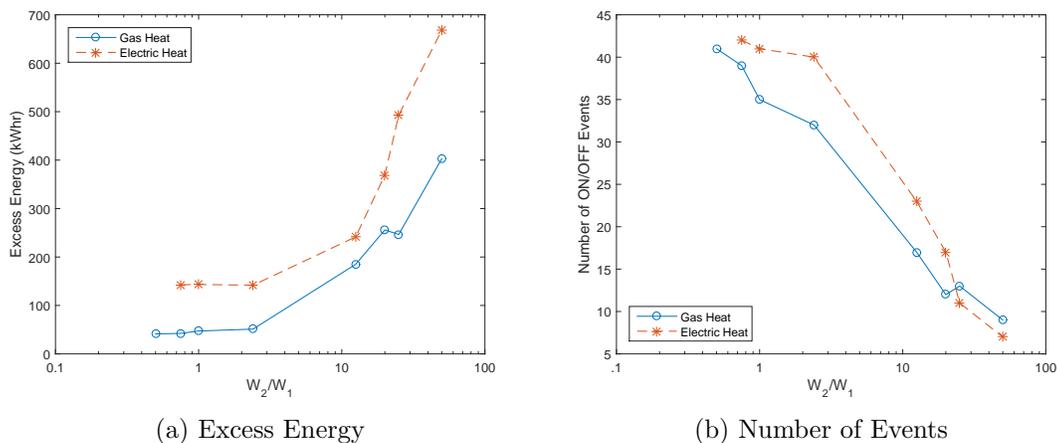


Figure 5.3: Varying Weighting Factors

5.3.3 Model Parameters

The main parameters that define the neighborhood are the number of consumers ($c \in C$), consumer demand profiles ($D_{c,t}^{electricity}$ and $D_{c,t}^{heat}$), number of generators ($i \in I$), generator capacity (L_i^{max}) and efficiency (ϵ_i), and connectivity between houses and generators ($S_i \subseteq C$). We present below the parameters for all of the case studies considered in the results section.

Consumer Demand Profile

The variables $D_{c,t}^{electricity}$ and $D_{c,t}^{heat}$ define the customer demand profiles. For consistency, we use kW as the unit for both, even though in the US it is customary to express the heat load in Btu. This is important for $D_{c,t}^{heat}$, as the heating demand may be satisfied using either electricity or gas, in which case the efficiency of the heating unit is accounted for to convert the power to the required amount of natural gas (this is discussed further shortly). We will develop profiles for a 24 hour period, for both the hot and cold seasons.

The representative demand profiles for electricity and natural gas (for heating) in a neighborhood consisting of 100 consumers are developed from EIA historical data and ERCOT typical load profiles for residential consumers in Texas. The EIA Residential Energy Consumption Survey [265] provided annual energy consumption statistics by end-use (space cooling, space heating, water heating, cooking, and other). We divided these data into seasonal and daily consumption levels, and assumed that the space cooling and other categories would be accounted for in $D_{c,t}^{electricity}$, while the space and water heating and cooking categories are accounted for in $D_{c,t}^{heat}$. ERCOT provides representative hourly load profiles for different types of consumers [87], which we used to represent the dynamic trends of the consumer appliance demand patterns. For heating, we assumed that the demand is higher in the morning and evening when generating the representative dynamic pattern. Based on these sources, we created a representative daily load profile for the appliance and heating requirements of a single residence in a hot and cold season, and then generated random perturbations on this profile in order

to obtain the $D_{c,t}^{electricity}$ and $D_{c,t}^{heat}$ profiles for one hundred consumers. An example of this is shown in Figure 5.4 for the hot season. The general equation used to create the load perturbations is shown below. For each consumer, the uniformly distributed random number R_c^A is used to shift the mean load while $R_{c,t}^B$ is used to scale the reference load at each time point.

$$D_{c,t} = R_c^A + R_{c,t}^B D_t^{ref} \quad (5.25)$$

The total neighborhood demand for appliances and heating (i.e., $\sum_c D_{c,t}^{electricity}$ and $\sum_c D_{c,t}^{heat}$ in the hot and cold season are shown in Figure 5.5). In order to calculate the amount of gas required (this is applicable in the gas heating scenario), we assume that all the equipment has efficiency ratings of 70% (based on reported efficiency values for commercially available water and space heaters). In the case of electric heating, we assume there are negligible conversion losses.

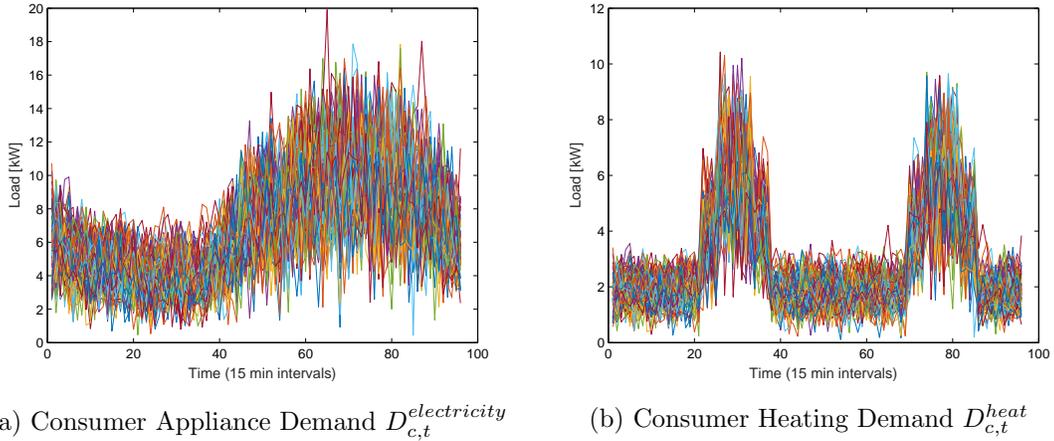


Figure 5.4: Hot season demand profiles for 100 individual consumers.

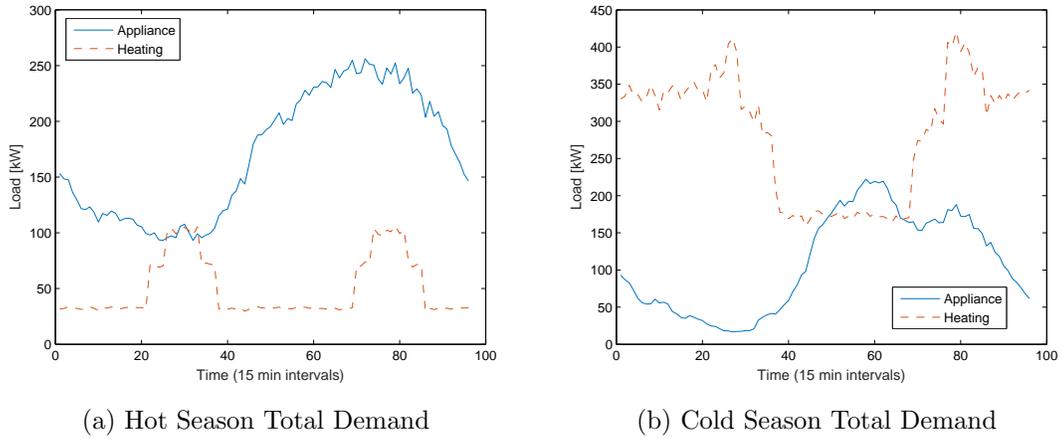


Figure 5.5: Total Daily Appliance and Heating Demands

Generator Size

The capacity of the generators used in this model is an important design consideration. Oversized generators may have a minimum load capacity (i.e., 80% of the maximum capacity) higher than the demand level during most of the day, and as a result will not turn on very often. Undersized generators serving a demand higher than their capacity will likely always be turned on to lower the load on the grid. Finding the appropriate generator size is therefore necessary to ensure that dynamic load levelling can be achieved.

In the following case studies, we will consider two generator sizes: $L_i^{max} = 10kw$ with $\epsilon_i = 15,031 \frac{BTU}{kWhr}$ and $L_i^{max} = 50kw$ with $\epsilon_i = 13,308 \frac{BTU}{kWhr}$. The generator size was determined based on observed peak demand for different groupings of houses. The selection of the generator sizes was carried out in conjunction with the mapping between consumers and generators, and is described below.

Consumer and Generator Configuration

The number of consumers serviced by an individual generator affects the operating state of the distributed generation system. When too few consumers are allocated to a generator, the combined demand is not high enough to allow the generator to turn on at its lowest operating level. When too many consumers are using the same generator, the generator will operate at maximum capacity for most of the day. Such a configuration would not allow for load following with the generators.

The ideal number of generators (and number of consumers assigned to each generator) is that which results in the lowest excess energy levels (as defined by equation (5.16)). To establish these conditions, the problem (5.18) was solved considering the hot season demand profiles for a variable number of generators. We assume that the same number of consumers is allocated to each of the generators (when the result of the division is not an integer, some generators may serve an additional consumer). The results for the $L_i^{max} = 10kW$ case are shown in Figure 5.6, which plots the excess energy as a function of the ratio of consumers to generators (e.g., the consumer:generator=10 point indicates that there is 1 generator serving 10 consumers or, equivalently, that there are 10 generators for the 100 consumers considered). Based on this analysis, we decided to consider 16 generators, each of which serves between 6-7 consumers. A similar analysis was performed for the $L_i^{max} = 50kW$ case, which showed that three generators are sufficient, with each generator serving 33 or 34 consumers.

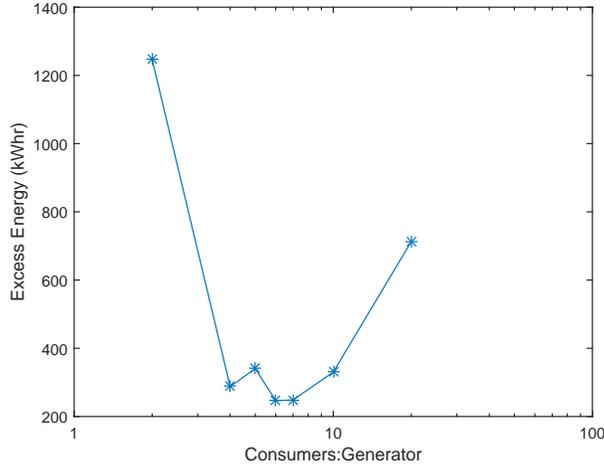


Figure 5.6: Performance of various consumer-to-generator ratios for the $L_i^{max} = 10kW$ case

5.4 Results

There is a large combination of potential scenarios to evaluate based on the choice of optimization model, hot or cold season demand, generator configuration, and electric or gas heating option. Table 5.2 lists the scenarios that are discussed here. This set of results was selected as the most practically meaningful. First, we discuss the unconstrained, constrained, and decentralized formulation solutions for the hot season. Then, we compare these data to the system behavior in the cold season. Finally, we evaluate the impact of electric heating.

5.4.1 Reporting Approach

Important output variables are presented in a results table for each case: the number of on/off events (N^{events}), the total volume of natural gas used for electric-

Table 5.2: Scenario Definitions

Results Section	Optimization Formulation	Heat Source	Season	Generator Size (kW)	Label
Sec. 5.4.2	Unconstrained (5.18)	Gas	Hot	10	Case1
	Unconstrained (5.18)	Gas	Hot	50	Case2
Sec. 5.4.3	Constrained (5.18)	Gas	Hot	10	Case3
Sec. 5.4.4	Decentralized (5.20)	Gas	Hot	10	Case4
	Decentralized (5.20)	Gas	Hot	50	Case5
Sec. 5.4.5	Unconstrained (5.18)	Gas	Cold	10	Case6
	Unconstrained (5.18)	Gas	Cold	50	Case7
Sec. 5.4.6	Unconstrained (5.24)	Electric	Hot	50	Case8
	Unconstrained (5.24)	Electric	Cold	50	Case9

ity generation over the time horizon ($V^{gen,tot}$), the total natural gas used for gas heating equipment ($V^{gas,tot}$), the maximum natural gas flow rate to the generators ($V^{gen,Max}$), the maximum natural gas flow rate for heating equipment ($V^{gas,Max}$), the excess energy (E^{excess}), and the target grid load (α). These values of these variables for each case are summarized in Table 5.3.

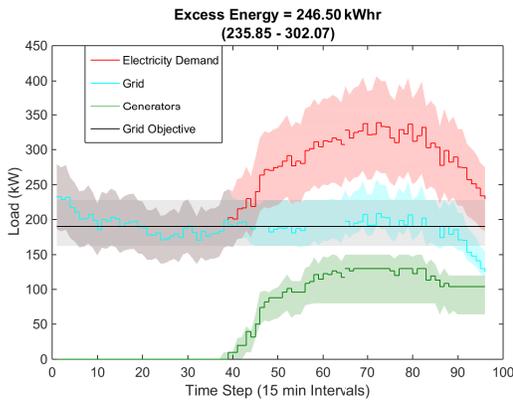
The solution for each scenario is also described with two plots: (a) the neighborhood demand and power use by energy source for each hour of the day, and (b) a Gantt chart of the generator operating levels throughout the day. The demand curve is equivalent to the sum of the natural gas generation load curve and electric grid load curve. The shaded, transparent area around each curve represents the range of solutions for different demand levels, which can be interpreted as the

size of the operating space or the effect of uncertainty for the given season. This was determined by solving the optimization problem for four other demand curves corresponding to 80, 90, 110, and 120% of the baseline demand profile determined in Section 5.3.3.

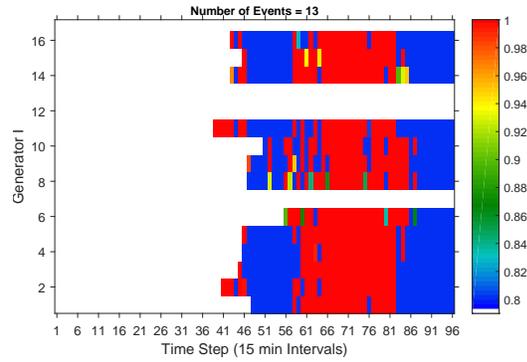
Figure (b) shows the generator operating schedules. Any white space within the chart indicates that the generator is off and the colored squares show the changes in operating level for each time step, where the color map represents ramping from 80% to 100% of the generator capacity.

Table 5.3: Summary of Case Results

Label	Number of on/off Events	Excess Energy (E^{excess})	Grid Target (α)	Total NG Consumed by Generators	Total NG for Heating Equipment	Max. Generator Flow
–	–	<i>kW-hr</i>	<i>kW</i>	<i>ft³/day</i>	<i>ft³/day</i>	<i>ft³/hr</i>
Case1	13	246.50	190.15	21823	31.79	146.4
Case2	4	180.57	187.39	23121	31.79	767.0
Case3	13	359.79	200.50	18889	31.79	120.0
Case 4	16	593.45	178.2	25728	31.79	146.4
Case 5	3	387.48	183.11	25323	31.79	767.0
Case6	22	469.93	65.90	15694	255.51	146.4
Case7	6	349.61	43.98	21936	255.51	767.0
Case8	5	297.09	230.36	25014	0.00	767.0
Case9	6	486.39	268.13	23673	0.00	767.0

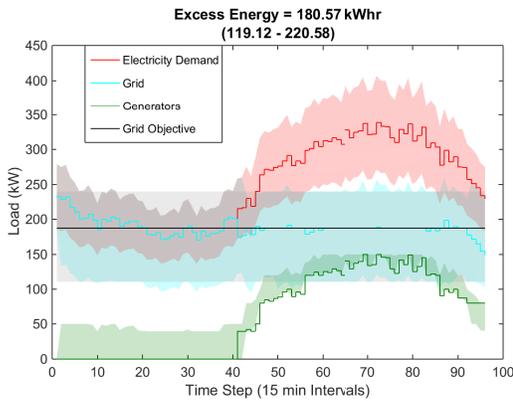


(a) Neighborhood Power Load by Source

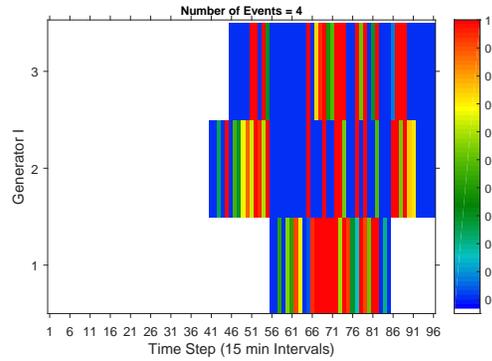


(b) Generator Operating Level

Figure 5.7: Case 1 Results (unconstrained, gas heat, hot season, 10kW generators)



(a) Neighborhood Power Load by Source



(b) Generator Operating Level

Figure 5.8: Case 2 Results (unconstrained, gas heat, hot season, 50kW generators)

5.4.2 Unconstrained Case Results

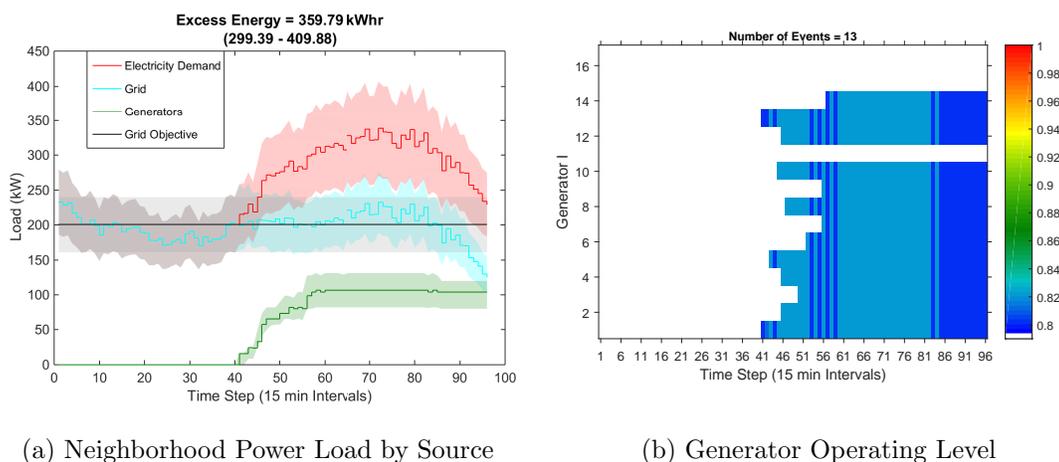
Figures 5.7 and 5.8 show the total load and generator behavior for the unconstrained cases 1 and 2. Interestingly, the excess energy and α value are lower in case 2 than case 1. This is unexpected since the total capacity for the 16 small

generators is 160kW, while the total capacity of the three larger generators is only 150kW. Also, the large number of small generators conceivably has the potential to achieve more precise load following by staggering the on/off time of each generator to reflect small changes in demand. This was not the case because, as can be seen in Figure 5.7b, not all of the generators are utilized. This is due to the value of α and the fact that the objective is formulated to achieve a flat load profile without penalizing the magnitude of the load. Also, because the 50kW generators have a larger operating range (40-50kW) than the 10kW generators (8-10kW), the larger generators have greater flexibility and can more broadly vary their operating range without incurring an on/off event that is penalized in the objective function. This is also shown in the Gantt charts of Figure 5.7 and 5.8. The 50kW generators experienced broader changes in their operating levels than the 10kW generators.

5.4.3 Constrained Case Results

Figure 5.9 shows the results of Case 3, whereby a constraint is imposed on the amount of natural gas delivered to the neighborhood to reflect potential natural gas infrastructure limitations. We did not consider a constrained version of the 50kW generator configurations based on the assumption that a dedicated NG delivery infrastructure would be implemented for larger generators. In both Case 1 and Case 2, there is a substantial increase in the amount of natural gas being delivered to the generators compared to a neighborhood lacking this type of distributed generation. As a result, the excess energy in Case 3 increases by 46% relative to the unconstrained case 1. The generators are forced to operate at only 82% capacity

due to NG availability limitations (constraint (5.12) is active). The results in Table 5.3 show that the total natural gas consumption is always less than the maximum allowable consumption, $35,880 \frac{ft^3}{hr}$, while the flow rate to each generator reaches the pipe capacity of $120 \frac{ft^3}{hr}$. As a result, the generator on/off times are further apart in Case 3 than in Case 1 (compare Figures 5.9b and 5.7b) to help with load following. It is clear that establishing the appropriate generator size depends on potential supply constraints in addition to the demand pattern.



(a) Neighborhood Power Load by Source (b) Generator Operating Level

Figure 5.9: Case 3 Results (constrained, gas heat, hot season, 10kW generators)

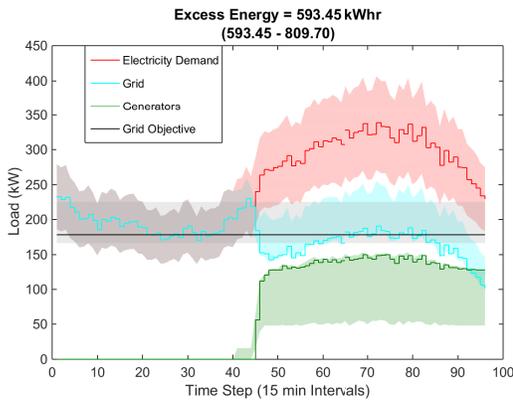
5.4.4 Decentralized Case Results

Figures 5.10 and 5.11 show the results for Cases 4 and 5. The decentralized operation of the set of generators shows a substantial increase in excess energy compared to the centralized optimization results (141% and 115% increase, respectively, compared to Cases 1 and 2). This is due to the lack of coordination of on/off times for all the generators, which causes a sharp drop in the grid load

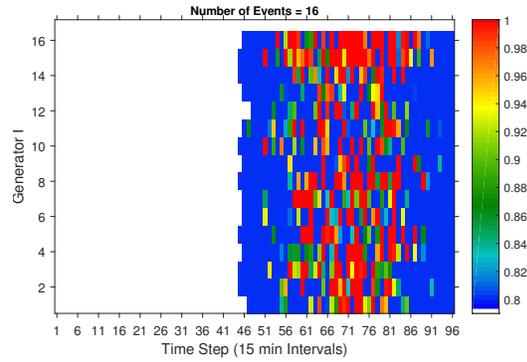
when all generators turn on at about the same time. A similar behavior is to be expected if the objective of each consumer were rewritten in terms of unilateral cost minimization, in which case it is likely that every consumer would decrease its demand during the peak pricing period (potentially causing a secondary rebound peak [202, 140]). This behavior is not desirable from the utility operators’ perspective (see, e.g., the challenges associated with operating the grid in the California “duck curve” scenario [41]), and these observations demonstrate the important role of a centralized decision making entity within the context of building energy management, particularly in the context of residential buildings. As such, the development of communication protocols (e.g., OpenADR [99] and cloud-based information sharing architectures [139]) between the utility and consumers participating in demand response programs is an ongoing effort. For the interested reader, we note that the utility-level economic implications of customer-owned DG were been discussed in [83], and the broad question of ownership of commonly pooled DG resources has been reviewed in [292].

5.4.5 Cold Season Cases

The results for Cases 6 and 7 are shown, respectively, in Figures 5.12 and 5.13. The electricity demand profile for the cold season has a different load pattern than the hot season, and overall the demand is lower at all times during the day. Because of these differences, the generators are underutilized, especially in the scenario involving the 10kW generators. One reason for this is the large slope of the demand curve affects the optimal value of α , and in this situation the optimal

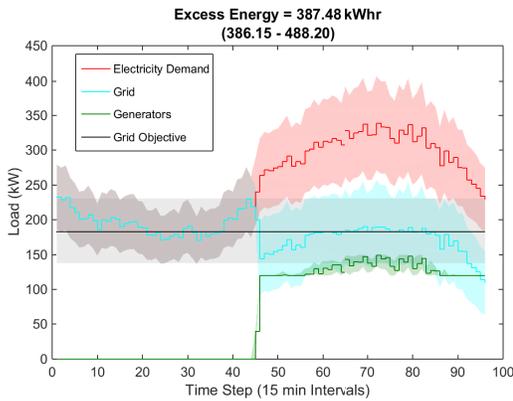


(a) Neighborhood Power Load by Source

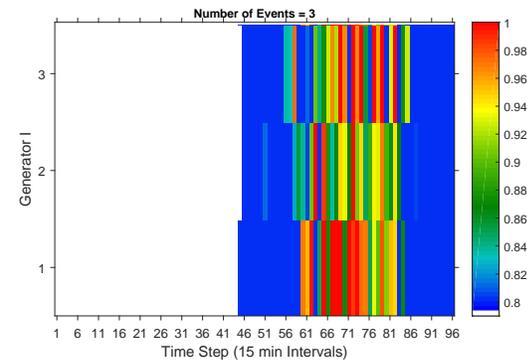


(b) Generator Operating Level

Figure 5.10: Case 4 Results (decentralized unconstrained optimization, gas heat, hot season, 10kW generators)



(a) Neighborhood Power Load by Source



(b) Generator Operating Level

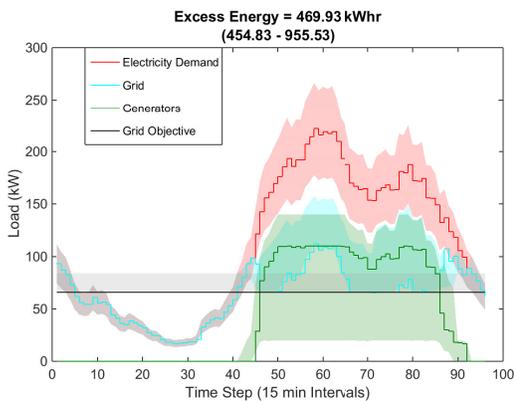
Figure 5.11: Case 5 Results (decentralized unconstrained optimization, gas heat, hot season, 50kW generators)

α value is low enough that not all the generators are needed.

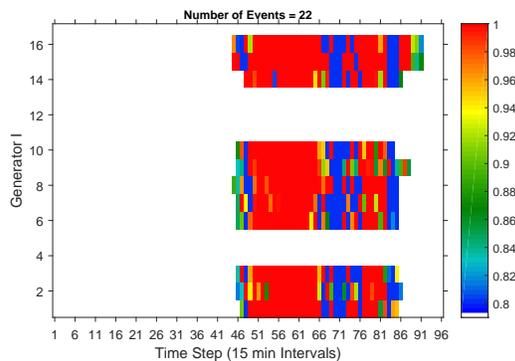
Another reason is that the generator size choices and the allotment of consumers to generators may not be optimal. Houses with low demand should be grouped

with houses that have high demand so there is enough total demand within a consumer set to allow the corresponding generator to turn on. Grouping houses with dissimilar demand profiles might also allow for more variation in the generator fractional operating level, which can improve load following. In this sense, the results presented here represent a “worst-case scenario,” in which the allocation of consumers to generators is random. In some cases, the combined demand of all the houses assigned to certain generators is lower than the minimum operating level of the corresponding generator for large periods of time (a similar behavior was observed for the constrained version of Case 6, not shown), which indicates that the generators are oversized for this season. The lack of staggering of times when the generators switch on is not desired, and this behavior is different from that observed in Case 1. This is explained by noting that all of the houses are increasing their demand around the same time and reach the minimum operating threshold of the generator (notice the steep slope of the demand curve around time step 45).

However, in the 50kW case (Case 7) the three generators are all used and the resulting excess energy and grid target α are much lower than in Case 6. There is still a similar issue regarding the inability to stagger the times when the generators switch on, which results in a steep drop in the load on the grid. This may be undesirable from the grid-operating point of view, and could be addressed by adding an explicit penalty in the optimization formulation on the rate of change of the grid load.

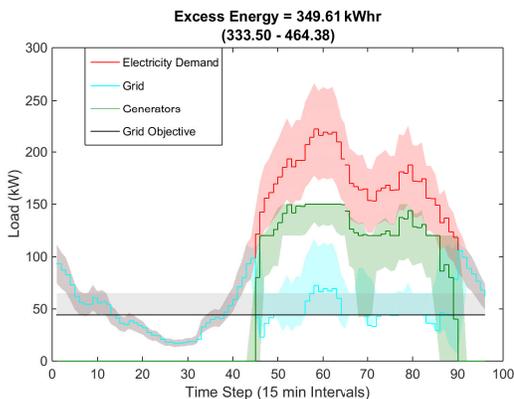


(a) Neighborhood Power Load by Source

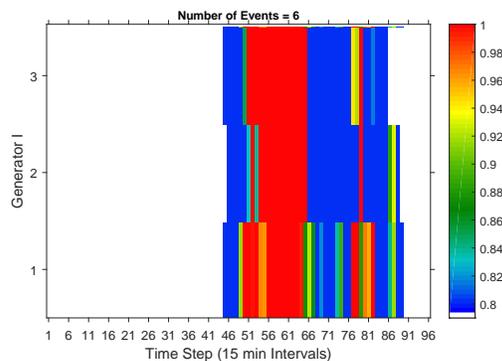


(b) Generator Operating Level

Figure 5.12: Case 6 Results (unconstrained, gas heat, cold season, 10kW generators)



(a) Cold Season Neighborhood Power Load by Source (BC4)



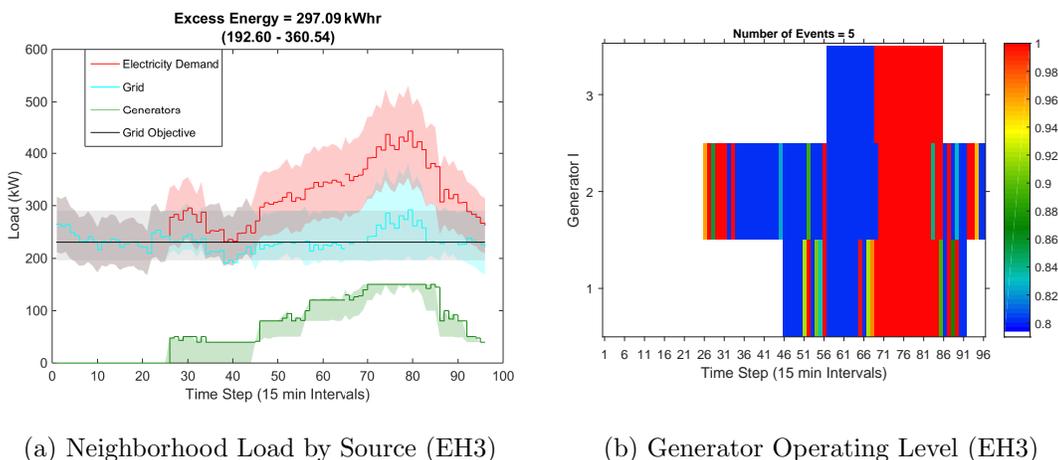
(b) Cold Season Operating Level (BC4)

Figure 5.13: Case 7 Results (unconstrained, gas heat, cold season, 50kW generators)

5.4.6 Electric Heating Cases

If the neighborhood is designed such that all the consumers rely on electricity for heating (rather than natural gas), the total electricity demand is higher throughout

the day and has multiple peaks due to the bimodal heating demand profile (see the red curves in Figures 5.14a and 5.15a for the hot and cold season electricity demand). The results for Cases 8 and 9 are shown in Figures 5.14 and 5.15. First, it is clear that the generators are undersized for these demand levels because they can not perform load following in the afternoon and they spend a significant portion of time operating at their maximum capacity. As such, their predominant function is to lower the baseload. However, in the cold season the demand in the morning is so high because of the added heat load that a generator turns on (Figure 5.15b), and this behavior was not observed in any of the previous cases.

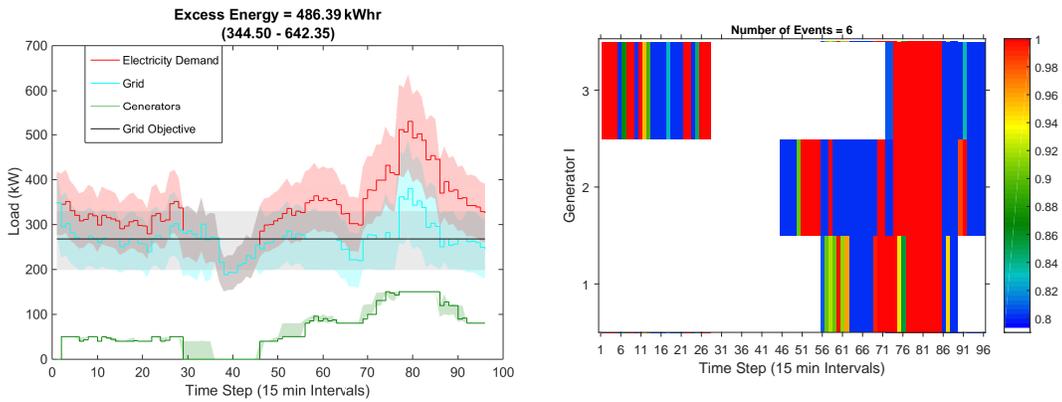


(a) Neighborhood Load by Source (EH3) (b) Generator Operating Level (EH3)

Figure 5.14: Case 8 Results (unconstrained, electric heat, hot season, 50kW generators)

5.5 Conclusion

In this paper, we considered the interaction of the natural gas and electricity grids at the residential level, the lowest level of reticulation of the two networks. In



(a) Neighborhood Load by Source (EH4)

(b) Generator Operating Level (EH4)

Figure 5.15: Case 9 Results (unconstrained, electric heat, cold season, 50kW generators)

particular, we considered the case of gas-fired distributed electricity generation in conjunction with home natural gas use (i.e., cooking and heating), under realistic demand scenarios and limitations posed by the natural gas distribution infrastructure. We demonstrated that it is indeed possible to use a large number of small scale generators to flatten the electricity demand of a residential neighborhood. Our optimization-based framework suggests that the system design should consider demand dynamics and limitations of current infrastructure, and the effects of mapping consumers to generators. Moreover, our results suggest that centralized decision-making and coordination of the operation of the generators results in optimal demand profiles, and it is difficult to replicate this behavior with a decentralized decision-making scheme whereby the operation of each generator is managed separately. While a large number of small scale generators is useful for localized load balancing, the need for a centralized control system and the ease of

designing and operating a smaller number of larger generators indicates that using higher-capacity DG units (in our case, 50kW) is an attractive investment for the future electric and natural-gas grids.

Part II

Scheduling with Dynamic Process Models

Chapter 6

Scheduling with Data-Driven Process Models

The material in this chapter has been published in [257] and [259].

6.1 Introduction

Chemical processes are complex systems whose operation spans multiple time and space horizons. Decision-making in the chemical supply chain has historically abided by the time scales of process and market phenomena (Figure 6.1), with separate decision tiers dedicated to production management functions (i.e., planning and scheduling), and to process control (including the supervisory and regulatory control layers).

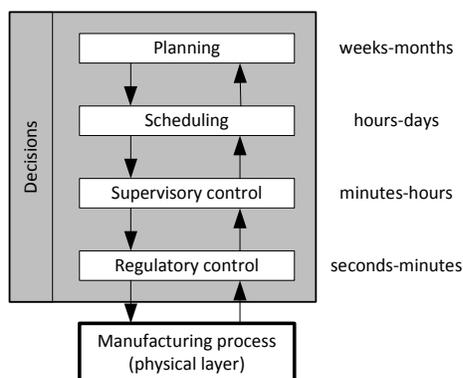


Figure 6.1: Decision-making hierarchy in process control and operations.

More recently, however, researchers have emphasized the potential benefits of exchanging information and integrating decisions between the tiers of the operation hierarchy in Figure 6.1. The current economic environment features rapid changes in market conditions, e.g., fluctuations in product demand, variations of the cost and availability of raw materials and energy. As a consequence, production management decisions are increasingly made at time intervals that overlap with the time scale of the process dynamics (see, e.g., [210, 46] for a case study focusing on the air separation industry).

Under these circumstances, a tighter coordination between the production management and control layers of the decision-making hierarchy in Figure 6.1, and in particular between production scheduling and supervisory process control, becomes critical [22]. Integrating scheduling and control is, however, a difficult task. On the one hand, there are organizational challenges that stem from the fact that production scheduling and, respectively, process control, are generally the prerogatives of different divisions of a commercial enterprise. These entities have different goals, operating tools and, more importantly, personnel with different backgrounds and expertise [238].

On the other hand, the integration of scheduling and control is faced with significant technical challenges, related to the need to, i) consider a sufficiently long (scheduling) time horizon to capture the process economics, ii) account for discrete variables that are inherent to production management decisions, and, iii) carry out the relevant calculation online and in real-time, in a manner that can guarantee the closed-loop properties (stability, disturbance rejection) that are expected to be

satisfied by the control system.

The majority of research efforts aimed at integrating scheduling and control of continuous processes that have been reported in the literature follow a top-down approach [22]. This consists of embedding dynamic information (i.e., the dynamic process model) in the scheduling framework, with the purpose of improving the accuracy of estimating the transition times (and associated costs) between different product grades and/or production rates. The resulting integrated scheduling and control formulation is a mixed-integer dynamic optimization (MIDO) problem, whose solution can become computationally costly as the model size and the length of the scheduling horizon increase [111]. Both simultaneous [96, 250, 309] and sequential methods [50, 214] have been proposed to solve this problem, along with several decomposition strategies [251, 60] for reducing computation time. The reader is directed to the recent survey [22] for more details on the current status of pathways for integrating scheduling and control.

In our previous work [77, 25, 210, 259], we introduced a new paradigm for integrating scheduling and supervisory control. Our approach consisted of deriving a low-order model of the closed-loop behavior of the process. This representation, which we refer to as a scale-bridging model (SBM), is then embedded in the scheduling formulation and used to evaluate the impact of process dynamics on scheduling decisions. All of the works cited rely on continuous time models.

In this work, we introduce a novel method for incorporating discrete-time models of the process dynamics in the integrated scheduling and control framework. We use autoregressive with exogenous inputs (ARX) models of the closed-loop process

response as a discrete-time scale-bridging model (DSBM). Based on these DSBMs, we define a hybrid discrete-continuous formulation of the integrated scheduling and control problem, and propose a computationally efficient reverse integral error criterion for establishing the completion of transitions between products.

6.2 Problem Formulation

In this work, we focus on continuous process systems that can be described using an input-affine nonlinear model (6.1)

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{G}(\mathbf{x})\mathbf{u} \quad (6.1a)$$

$$\mathbf{y} = \mathbf{h}(\mathbf{x}) \quad (6.1b)$$

where $\mathbf{x} \in D_x \subset \mathbb{R}^{n_x}$ are the differential (state) variables, $\mathbf{u} \in D_u \subset \mathbb{R}^{n_u}$ are manipulated inputs available for control, $\mathbf{f} : D_x \rightarrow D_x$ and \mathbf{G} is a $n_x \times n_u$ dimensional function with every column $g_j : D_x \rightarrow D_x$. The process quality variables are defined based on outputs $\mathbf{y} \in D_y \subset \mathbb{R}^{n_y}$, with $\mathbf{h} : D_x \rightarrow D_y$.

The process is assumed to make $N_p \geq 1$ products, which are defined in terms of the corresponding values of the quality variables. Thus, each product has distinct values of the quality variables, \mathbf{y}_i , with $i = 1, \dots, N_p$.

For each product i : the price π_i , demand rate $\delta_{r,i}$ and inventory cost $c_{storage,i}$, are assumed to be known. The (possibly time-varying and product-dependent) production rate $q(t)$ is defined as a function of the states \mathbf{x} and inputs \mathbf{u} . Without significant loss of generality, we assume there is no input-output multiplicity and

that the state space D_x is such that each product \mathbf{y}_i can be reached from at least one other product $\mathbf{y}_{i'}$ via a continuous path in the phase plane.

6.2.1 Scheduling

For the continuous processes under consideration, production scheduling consists of optimizing the order (production cycle) in which the N_p products are made, and identifying the production time for each product, maximizing an objective function defined in terms of the operational profit [96, 25]:

$$J = \frac{1}{T_m} \sum_{i=1}^{N_p} \sum_{s=1}^{N_s} \omega_{i,s} [\pi_i - c_{storage,i}(T_m - t_s^f)] \quad (6.2)$$

Following the approach in [96, 25, 77], we assume that production of each product i occurs in successive time intervals (time slots), with only one product produced in each time slot. For simplicity, a cyclical production scenario in which a product is made only once during the production cycle is considered. This implies that the number N_s of production slots is equal to the number of products, i.e., $N_s = N_p$.

The binary variables $z_{i,s} \in \{0, 1\}$ are used for the assignment of products to slots. The above assumptions can be expressed in terms of $z_{i,s}$ using the following assignment constraints imposed on (6.2):

$$\sum_{i=1}^{N_p} z_{i,s} = 1 \quad \forall s \quad (6.3a)$$

$$\sum_{s=1}^{N_s} z_{i,s} = 1 \quad \forall i \quad (6.3b)$$

which require that only one product be made in each time slot (6.3a), and, that a product i can only be made once in each production cycle (6.3b).

We will utilize a continuous-time formulation, in which the start time $t_s^s \geq 0$ and end time $t_s^f > 0$ of each slot define the production time $t_{i,s}^p$ in slot s . Along with the transition time τ_s , which is the (sequence- and product-dependent) time needed to switch from producing product i' in slot $s - 1$ to making product i in slot s , these variables are used to define the production timing constraints:

$$t_s^f = t_s^s + \tau_s + \sum_{i=1}^{N_p} t_{i,s}^p \quad \forall s \quad (6.3c)$$

Additional timing constraints are included to match end and start times of subsequent slots, fix the start and end times of the initial and final slots, and limit slot duration:

$$t_s^s = t_{s-1}^f \quad \forall s \neq 1 \quad (6.3d)$$

$$t_{N_s}^f = T_m \quad (6.3e)$$

$$t_1^s = 0 \quad (6.3f)$$

$$t_{i,s}^p \leq z_{i,s} t_{max}^p \quad \forall i, s \quad (6.3g)$$

The production time and slot assignment of each product can then be used to impose constraints on demand satisfaction:

$$\begin{aligned} \omega_i &\geq \delta_{r,i} T_c \quad \forall i \\ \omega_i &\leq \lambda_i \delta_{r,i} T_c \quad \forall i \\ \omega_{i,s} &= q_i t_{i,s}^p \quad \forall i, \forall s \\ \omega_i &= \sum_{s=1}^{N_s} \omega_{i,s} \end{aligned} \quad (6.3h)$$

6.2.2 Interaction of Scheduling and Process Dynamics

In a continuous process, the transition between products manufactured in different time slots does not happen instantaneously, as illustrated in Figure 6.2. Rather, such changeovers depend on both the dynamic characteristics of the process, and on the performance of its control system. The transition times τ_s in the scheduling problem formulation (6.2)-(6.3) capture the duration of such transitions (e.g., in the form of a transition table with nonzero entries for all possible transitions).

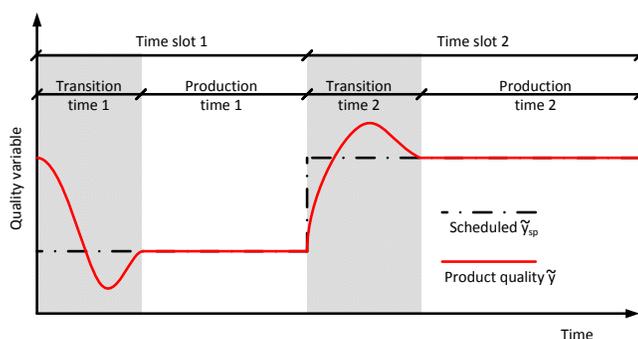


Figure 6.2: Slot-based continuous-time schedule and transitions between products [96, 22].

In the conventional approach to production scheduling, the parameters τ_s are constant and thus constitute a static representation of the process response; beyond this information, conventional scheduling calculations are agnostic to the process dynamics. This shortcoming can be addressed by augmenting the optimization problem with the dynamic model of the process, (6.1). The process inputs \mathbf{u} become an additional set of decision variables, and the transition times

are determined from the actual process dynamics:

$$\begin{aligned}
 & \underset{\mathbf{u}, \mathbf{z}, t_s^p, t_s^s, t_s^f}{max} \quad J & (6.4) \\
 & s.t. \quad \text{scheduling constraints (6.3)} \\
 & \quad \quad \text{dynamic process model (6.1)} \\
 & \quad \quad \mathbf{x}, \mathbf{y}, \mathbf{u} \in D_{\mathbf{x}, \mathbf{y}, \mathbf{u}}
 \end{aligned}$$

This mixed-integer dynamic optimization (MIDO) problem is computationally expensive [111], especially when the process model is large and nonlinear. Additionally, solving (6.4) at the beginning of the production cycle will result in an *open-loop* control solution $\mathbf{u}(t)$, which is optimal in the nominal case (in the sense that it results in the optimal production sequence, including optimal changeovers, when imposed on the process), but may lead to significant performance degradation or even loss of stability in the case of plant-model mismatch or in the presence of disturbances.

6.3 Integrated Scheduling and Control of Continuous Processes

Based on the system definition and problem formulation presented above, we can infer that:

i) in advanced manufacturing scenarios, where information exchange between the levels of the process decision-making hierarchy is prevalent, process dynamics must be considered at the scheduling stage. This becomes especially true in highly dynamic market conditions, when product changeovers are frequent and transition

times are not negligible compared to production times in each slot.

ii) the optimal production schedule should be implemented in closed-loop, i.e., a control system should be in place to account for the presence of disturbances and/or plant-model mismatch.

Moreover, observation ii) suggests that it is of interest for scheduling models to incorporate an expression of *closed-loop* process dynamics, whereby the *setpoint* of the control system becomes a decision variable in the scheduling calculation.

6.3.1 Scale-Bridging Models

In our previous work [77, 25], we have introduced the concept of a scale-bridging model (SBM) as precisely this representation of the closed-loop process dynamics (Figure 6.3).

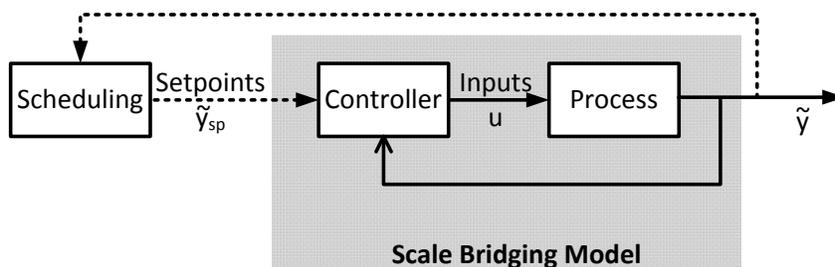


Figure 6.3: Scale-bridging models for integrated scheduling and control [77, 25].

More specifically, the SBM framework relies on defining $\tilde{\mathbf{y}}$ as the subset of the process quality variables \mathbf{y} that are relevant to scheduling, and capturing the dynamic behavior of $\tilde{\mathbf{y}}$ in response to changes in the relevant setpoints $\tilde{\mathbf{y}}_{sp}$. A SBM

thus takes the form:

$$\mathbf{0} = \Psi(\tilde{\mathbf{y}}_{sp}, \tilde{\mathbf{y}}, \dot{\tilde{\mathbf{y}}}, \ddot{\tilde{\mathbf{y}}}, \dots) \quad (6.5)$$

The use of SBMs to represent the process dynamics in scheduling calculations is well motivated by extensive previous research[18], where it was shown that the slow dynamics (i.e., over time scales relevant to scheduling) of quality variables such as product purity, total material inventory, production rate, can be captured with low dimensional models.

Assuming that a SBM is available, the integrated scheduling and control problem takes the form:

$$\begin{aligned} \max_{\tilde{\mathbf{y}}_{sp}, \mathbf{z}, t_s^p, t_s^s, t_s^f} \quad & J & (6.6) \\ \text{s.t.} \quad & \text{scheduling constraints (6.3)} \\ & \text{scale-bridging model (6.5)} \\ & \tilde{\mathbf{y}} \in D_{\tilde{\mathbf{y}}} \end{aligned}$$

In our previous work [77], we have shown that, under certain conditions on the dynamics of the process (6.1), geometric control ideas [71] can be used to derive a (nonlinear) control law that imposes, for each quality variable, a linear input-output behavior of the form

$$\tilde{y} + \sum_{k=1}^{r_k} \beta_k \frac{d^k \tilde{y}}{dt^k} = \tilde{y}_{sp} \quad (6.7)$$

which represents the desired SBM, and can be used for accounting for process dynamics and controller performance at the scheduling stage. Here r_k represents the relative degree of the system and β_k are tuning parameters.

6.4 Handling Discrete-Time Dynamic Process Models

In this work, we introduce a novel method for incorporating discrete-time models of the process dynamics in the integrated scheduling and control framework. We use autoregressive with exogenous inputs (ARX) models of the closed-loop process response as a discrete-time scale-bridging model (DSBM). Based on these DSBMs, we define a hybrid discrete-continuous formulation of the integrated scheduling and control problem, and propose a computationally efficient reverse integral error criterion for establishing the completion of transitions between products.

However, first-principles dynamic process models are often not available in practical applications, and data-driven process models are used instead.

6.4.1 Discrete-time Scale Bridging Models

In this section, we rely on ARX models that capture the closed-loop evolution of the process output \mathbf{y} and are driven by the exogenous input \mathbf{y}^{sp} , i.e., the controller setpoint. Our discussion will be limited to single input, single output (SISO) systems for simplicity, but can be easily generalized to the multiple input, multiple output (MIMO) case. Thus, we will assume that an ARX DSBM that describes the closed-loop input-output behavior of the process can be identified from available historical process data:

$$y_{t,s} = \sum_{i=1}^{n_a} a_i y_{t-i,s} + \sum_{j=1}^{n_b} b_j y_{t-j,s}^{sp} \quad ; \quad t_i = t_{i-1} + \delta \quad (6.8)$$

where the output at time t in slot s ($y_{t,s}$) is calculated based on n_a previous output points, n_b previous values of the setpoint, with a sampling interval δ .

Using this model for the purpose of integrating scheduling and control is, however, not a simple matter of substituting equation (6.5) in formulation (6.6) with the ARX SBM (6.8). Specifically, when using the continuous-time models, the MIDO (6.6) is transformed into a mixed-integer nonlinear program by discretizing the differential equations in time using a finite-element or finite-difference scheme with elements of equal but variable length [33]. The discretization spans exactly the length of the transition period. This in turn facilitates computing the transition durations τ_s , which are the key information provided by the dynamic model. On the other hand, this strategy is not applicable when using the ARX DSBM, since the sampling interval of the model is *fixed* and it cannot be used as a decision variable in the optimization.

In order to deal with this challenge, we develop below a hybrid formulation for integrating scheduling and control, whereby the dynamic process model is expressed in discrete time, and the remainder of the scheduling framework preserves its continuous-time formulation. This framework is shown in Figure 6.4. Our formulation is predicated on using the ARX DSBM to represent a *fixed* span T_H (where T_H is chosen such that it is longer than any of the transition times observed in historical data) of the process dynamics in the scheduling problem. The time horizon contains $N_H = T_H/\delta$ time points. This dynamic representation captures the evolution of the process output y following a change in the setpoint y^{sp} , and we make the following assumptions: (i) The tracking and stabilizing properties of the control law ensure the output of the process will reach a vicinity ε of the new steady state corresponding to the desired product in slot s at some time $\tau_s < T_H$,

i.e., $e_{t,s} < \varepsilon \quad \forall \tau_s \leq t < T_H$. (ii) The error $e_{t,s}$ may be computed using the absolute or squared deviation of the output from its setpoint, $(y_s^{sp} - y_{t,s})$.

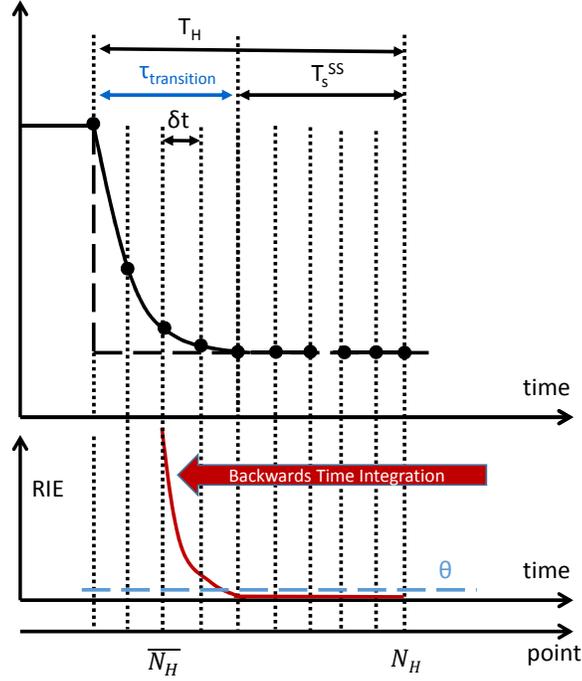


Figure 6.4: A time slot in the DSBM - based integrated scheduling and control formulation.

Determining the transition time τ_s is a key issue in this formulation. This task could, in principle, be accomplished by i) defining, at each time point (t,s) , a binary variable which becomes true when the error $e_{t,s}$ is lower than an established threshold and, ii) using the earliest time point where the binary variable becomes true as the transition time value. This approach does, however, have some disadvantages: first, it calls for the use of a large number $(N_s \cdot N_H)$ of binary variables, which complicates the solution of the problem, and, second, it can lead to erroneous

results in the case when the process response exhibits overshoot and/or oscillatory behavior, in which case the end of the process transition would be misidentified.

Instead, we propose identifying the end of the transition period τ_s by “looking backwards” from the end of the time span T_H for $N_H - \bar{N}_H$ points, where $\bar{N}_H > 1$. We thus compute a *reverse integrated error (RIE)*:

$$RIE_{t,s} = \sum_{i=t}^{N_H} e_{i,s} \quad \forall t \geq \bar{N}_H \quad (6.9)$$

We introduce the threshold parameter θ to identify the time at which the process is still in the vicinity of the desired steady state in the respective slot. Specifically, while the RIE is still below the threshold, the transition is complete and the process is assumed to be at steady state. The corresponding time points are catalogued using the binary variable $b_{t,s}$, which is true when the process is in the neighborhood of the steady state (i.e., $b_{t,s} = 1$ if $RIE_{t,s} < \theta$). Then, T_s^{SS} is defined based on the total number of points that are below the threshold (6.10), and the transition time is the remaining portion of the forecast horizon.

$$T_s^{SS} = \delta \sum_{t=1}^{N_H} b_{t,s} \quad ; \quad \tau_s = T_H - T_s^{SS} \quad (6.10)$$

Thus, integrating scheduling and control using the proposed ARX DSBM involves solving the following optimization problem:

$$\begin{aligned} \min_{z_{k,s}, \omega_k, t_s^p, t_s^b, t_s^e} \quad & J & (6.11) \\ \text{s.t.} \quad & \text{Scheduling constraints (6.3)} \end{aligned}$$

ARX model (6.8) and RIE calculation (6.9)

Transition time and production time calculation (6.10)

6.4.2 Case Study

We consider the non-isothermal CSTR discussed by [96]. The reactor is capable of producing four products ($P_k = [p_1, p_2, p_3, p_4]$). Product properties are described in [78]. The system model comprises two equations, describing the evolution of dimensionless concentration $C = C/C_{in}$ and dimensionless temperature $T = T/T_{in}$ [96]. The manipulated variable for this system is the coolant flow rate F_c , which can be altered to modify the concentration C of the reactor product. We use an input-output linearizing controller with integral action [70], to impose critically damped second-order linear behavior with time constant 0.86 h. The aim of the scheduling problem is to determine the optimal production sequence, processing time for each product, production quantities, and total production time, such that the operational profit is maximized. We will compare the FDS approach (referred to as problem $P1$) to the scheduling with an ARX model and RIE conditions (referred to as problem $P2$). The MIDO corresponding to the scheduling problem using the dynamic process model was reformulated as MINLP via orthogonal collocation on finite elements following the approach in [78].

Setting up $P2$ (Equation (6.11)) required identifying an ARX DSBM. Data for system identification were obtained by simulating the process in closed loop subject to a sequence of random step changes in concentration setpoint. The target concentration values ranged from 0.09–0.24 and the switching times between steps were between 20 and 40 hours to ensure a sufficient settling time. The sampling time was $\delta = 5$ minutes. The ARX model coefficients were identified using the *arx* function in MATLAB with $n_a = n_b = 10$. The ARX model uses C as the output

and the concentration target/setpoint C_k^{SS} as the exogenous input. To complete the formulation of $P2$, we considered a horizon $T_H = 10h$ with $N_H = 120$ points, $\bar{N}_H = 40$ points, and the sampling interval $\delta = 5min$. We chose to formulate the endpoint constraint and RIE in terms of the absolute error, $ARIE(t) = \sum_{i=t}^{N_H} |y_s^{sp} - y_{t,s}|$; the threshold θ was set to $\theta = 0.04$.

Problems $P1$ and $P2$ are MINLPs and were solved using GAMS with SBB as the MINLP solver, CONOPT as the NLP solver, and CPLEX as the MIP solver. Figure 6.5 shows the concentration setpoint schedule and anticipated transition dynamics. Both solutions identify the same production schedule ($p_1 \rightarrow p_2 \rightarrow p_3 \rightarrow p_4$), production quantities, and production times. They also produce very similar profit (\$5542 and, respectively, \$5515) and makespan values (82.63h for $P1$ and 82.78h for $P2$). This indicates that the ARX model is effective in representing the system dynamics. Figure 6.5 shows the closed loop implementation of the schedule determined in $P2$. The profit derived from this closed-loop implementation (\$5,407) compares favorably with the results above given the presence of a small plant-model mismatch.

The computational effort required to solve $P2$ is higher than for $P1$ (96.60s vs. 10.61s); this is due to the increased number of binary variables present in the problem formulation (16 in $P1$ vs. 336 in $P2$). However, our recent work [210] suggests that the computational effort of SBM-based integration of scheduling and control does not increase significantly with problem size, owing to the fact that the dimension of the SBM $P2$ is small compared to the number of process *states* that must be included in $P1$. We therefore expect that the proposed approach will be

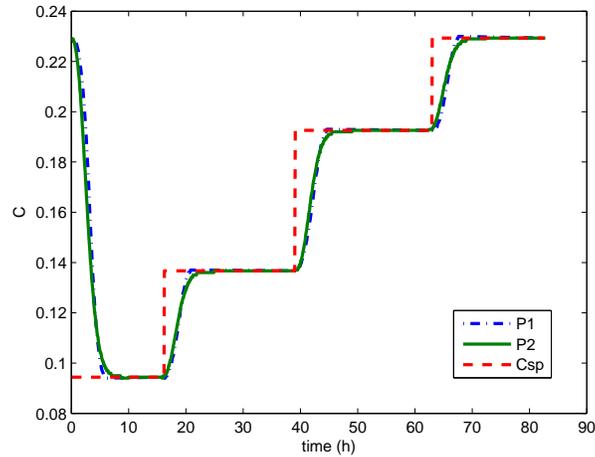


Figure 6.5: Optimal setpoint schedule and transition dynamics as determined by $P1$ and $P2$.

very effective in the case of large-scale processes having complex, high-dimensional models.

6.5 Conclusions

Motivated by the prevalence of data-driven models in practice, we discussed the integration of scheduling and control using data-driven models of the closed-loop process behavior. We formulated the integrated scheduling and control problem using a hybrid representation of time, whereby the (ARX) dynamic process model is presented in discrete time and the scheduling problem uses a continuous time formulation. We introduced the reverse integrated error concept to determine transition times between products. A CSTR case study was used to demonstrate our framework.

Chapter 7

Optimal Process Operations in Fast-Changing Electricity Markets: Framework for Scheduling with Low-Order Dynamic Models and an Air Separation Application

The material in this chapter has been published in [210].

7.1 Introduction

Fast-changing markets and the emergence of responsive, on-demand manufacturing require that production schedule changes occur over time scales comparable to the time constants of process systems. In turn, this requires that the dynamic characteristics and performance of the process and its control system be accounted for at the production scheduling stage [175, 106, 23].

However, embedding dynamics and control information in scheduling calculations has proven to be a difficult task [23]. Production scheduling and process control are typically carried out by separate entities of a company, and the coordination of interactions between the two functions is often challenging [239]. Significant challenges also arise from the need to account for the wide range of time scales involved in making scheduling and control decisions, and the corresponding

requirement to balance long-term prediction with real-time execution [23].

Intuitively, optimal scheduling decisions should consider process dynamics in markets where prices (and therefore operating costs and profit margins) change at a high frequency, *i.e.*, over time intervals that are comparable to the dominant time constant of the process. A typical example is the electricity market, where, owing to deregulation, prices change over short time spans (typically hours or even minutes). Participation in a real-time electricity market is often voluntary for industrial sites, who also have the option of entering fixed contracts with their utility suppliers. Variable pricing options become attractive when a site can quickly modulate its production rate by, i) increasing production and storing excess product when energy prices are low and, ii) using stored product to meet customer demand when prices are high and production rates are reduced.

Conventional methods for calculating optimal production schedules rely on tabulated transition information between (a set of discrete) operating points, coupled with steady state process models or production recipes[173]. It is implicitly assumed that the process is at a steady state prior to a production target change, and that it reaches steady state again before a new such change is made. However, in the case where market fluctuations are rapid and occur at a frequency comparable to the (slowest) dynamic modes of the process, this assumption is not valid, and may result in the calculation of a dynamically infeasible sequence of transitions (see Figure 7.3 below for an illustration of this phenomenon).

In this paper, we develop a scheduling formulation that includes dynamic information on product quality, production rate, and a subset of variables relevant

to process operating constraints. The novelty of our contribution consists of using scheduling-oriented low-order dynamic models that predict the closed-loop dynamics of relevant constrained process variables in response to scheduled operating point changes. Central to this effort, we introduce a new methodology for selecting the variables relevant to the scheduling calculation, and define the scope (i.e., model inputs and outputs) of the scheduling-oriented dynamic models. We also discuss the motivation for identifying such models from (closed-loop) historical operating data. These models are then integrated in the scheduling problem, which is formulated as a (mixed-integer) dynamic optimization aimed at maximizing profit over the scheduling time horizon.

The theoretical concepts are illustrated with an air separation unit (ASU) application. Here, taking advantage of a variable, time-of-day electricity price structure requires imposing production turn-up and turndown by exploiting the dynamic agility of the plant, and optimally utilizing the available cryogenic liquid product storage capacity to satisfy product demand at all times.

The paper is organized as follows: in the next section, we describe air separation units, and summarize the literature concerning their variable-rate operation. This motivating example is followed by a review of scheduling using dynamic models and associated computational challenges. We then introduce the proposed framework for scheduling using low-order models of the closed-loop process dynamics and demonstrate its implementation in an air separation unit case study.

7.2 Motivating Example: Air Separation Units

The purified components of air are an important feedstock for many manufacturing processes. For example, oxygen is used for steel production and in the chemical industry for production of ethylene oxide [91], and nitrogen gas serves as an inert replacement for air in the food and metals industries. Cryogenic distillation is the preferred method of separating air into its constituent gases when high production rates and moderate to high purities are required [282].

Air separation units (ASUs) have a very high energy consumption, and typically use electricity to drive the compressors that are used to handle and compress the air feed stream. The industrial gas sector utilized 19.4 TWh of electricity in 2010, or about 2.5% of the amount consumed by the entire manufacturing sector in the U.S. [264] Numerous publications and patents have contributed novel design concepts to minimize the nominal electricity use through tight process integration, more efficient unit operations, etc. [48, 74]

In a different vein, investigations have suggested improving operating economics by taking advantage of the deregulation of electricity markets, which has resulted in fast and significant fluctuations in electricity prices [182, 211]. This, in turn, requires exploiting the agility and switchability of the process, *i.e.*, frequently changing process outputs in response to electricity price changes [45, 304]. In principle, this calls for ramping up production rates during low electricity price periods and storing the excess products as cryogenic liquid. Then, stored liquid can be vaporized to satisfy gas demand while reducing production rates when electricity prices increase [121, 182, 134, 183, 308, 211].

Several studies have investigated variable production rate ASU operation. Zhu et al. [308] considered the optimal operation of an ASU subject to time-varying electricity prices and uncertain product demand over the course of a day using a multiperiod formulation to capture the uncertainty. A detailed steady-state nonlinear model was used and the process dynamics were approximated via fixed transition times. Miller et al. [182] considered the variable operation of ASUs producing liquid and gaseous products when subject to hourly electricity price variations. They computed the maximum-to-minimum energy price ratio that defines the profitability boundaries of a plant changing production rates to take advantage of time-varying electricity prices. The estimates were based on a simplified static plant model which used an ideal work calculation to compute the minimum power requirement of the ASU. Depending on the economic assumptions made, a ratio between two and seven (maximum-to-minimum electricity price) was required to render variable-rate production profitable [182]. In our recent work [211], we developed a similar design blueprint for a variable-capacity ASUs, showing that the design of the multistream heat exchanger may limit the agility of the process. Ierapetritou et al.[121] and Karwan and Kebli [134] also investigated a variable production rate ASU with a liquid storage tank, relying on simplified steady-state linear models to represent process performance. Mitra et al. [183] extended these results by considering the transition behavior and various limitations on production during the transitions, relying however on a linear problem formulation. All of these works suggested that modulating ASU operation (in particular, production rates) when subject to time-varying electricity prices can result in significant cost

savings, with the benefits increasing as the gap between peak and off-peak energy prices becomes wider.

The settling time (*i.e.*, the time to reach steady state after a change in process inputs or controller setpoints) for ASUs is typically in the order of hours. When utility prices (and, consequently production rate targets) change at a high (*e.g.*, hourly) frequency, accurate dynamic models of relevant process variables should be utilized to ensure that a sequence of scheduled production rate transitions is feasible and optimal. Cao et al. [44, 45, 47] presented initial results on dynamic modeling and optimization of ASU operations using a large-scale, first-principles, detailed dynamic process model. However, their results only focus on the optimal trajectory of individual production rate transitions, and do not consider multiple optimally scheduled production rate changes over an economically-relevant time horizon.

Motivated by the above, in this work, we study the integration of dynamics and control information in scheduling calculations for ASUs operating under fast-changing and highly variable market conditions. In particular, we will present a case study focusing on the cryogenic air separation process flowsheet shown in Figure 7.1 [45, 211]. The process utilizes a single cryogenic distillation column for producing high purity nitrogen. Inlet air at ambient conditions is compressed to 6.8 bar and is subsequently cooled to 300K in an auxiliary heat exchanger. The inlet air passes through the primary multistream heat exchanger (PHX) where the product and waste streams provide cooling. A portion of the air stream is removed from the PHX as a superheated vapor and sent through a turbine to generate elec-

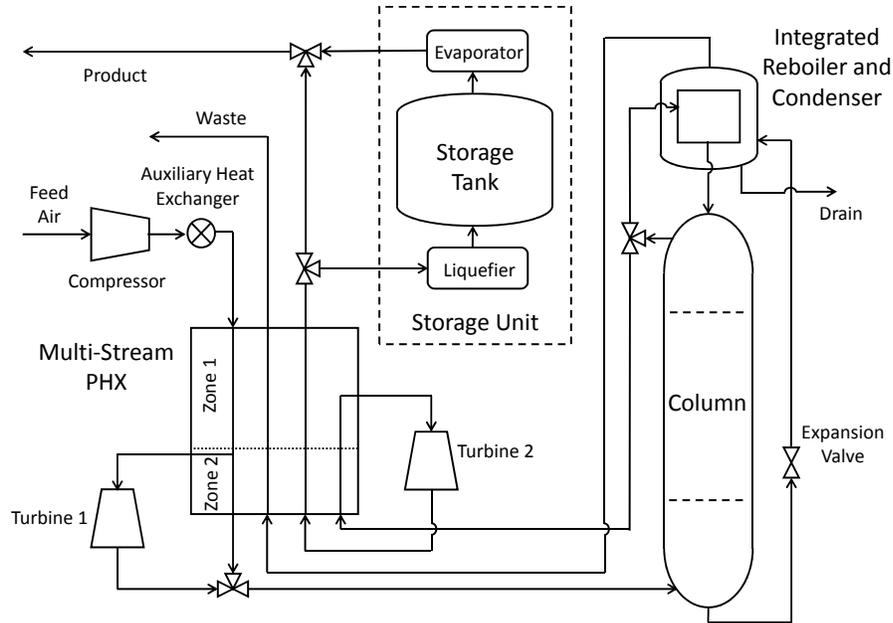


Figure 7.1: Flowsheet for the cryogenic air separation unit for the production of N_2 [211, 44].

tricity, and the balance is liquefied in the PHX. The vapor and liquid air streams are then fed to the bottom of the cryogenic distillation column. Adiabatic expansion of the bottoms provides cooling via the Joule-Thomson effect and is used to condense the vapor at the top of the column in an integrated reboiler/condenser. The low pressure waste stream from the reboiler and the high pressure gas product stream are returned to the PHX to cool the inlet air stream. To ensure full utilization of the available refrigeration, the product stream is expanded in Turbine 2 and repassed through the heat exchanger.

To further modulate plant production capacity, a separate nitrogen liquefier

is included in the process flowsheet along with a liquid nitrogen storage tank. These allow the process to meet gas nitrogen demand with (regasified) stored liquid nitrogen when electricity prices are high and production rate is decreased. During periods of low electricity price, production can be increased to build up liquid nitrogen inventory.

7.3 Scheduling under Dynamic Constraints

In this work, we consider a process system with a single product stream and storage capability. The role of production scheduling within the hierarchy of process operation decisions is illustrated in Figure 7.2. All nomenclature is described in Table 7.1. The planning step is performed by business units aiming to meet contractual agreements related to product specifications and quantities. In the planning step, \bar{y} , which represents (long-term) targets of product grade and quantity, is established. Scheduling then determines the optimal sequence of production setpoints (y_p^{sp}) and inventory utilization targets (α^{sp} represents the fractional split of the process output which bypasses the storage unit, and y_{inv}^{sp} represents the inventory output setpoint) which minimizes operating costs and satisfies \bar{y} . Most available scheduling approaches rely on capturing process dynamics in the simplest possible form, and typically assume that the plant output (y_p) is able to meet new target values (y_p^{sp}) in a well-defined, transition period whose duration is invariable. With this assumption, the process control and operation layers (at the bottom of Figure 7.2) can be ignored when making scheduling decisions.

As alluded to earlier, this assumption is not valid when setpoint changes are

made in response to time-varying market conditions which change at frequencies comparable to (or even higher than) the (slowest) dynamic modes of the process. In this case, the process dynamics and operating constraints must be considered to ensure that the sequence of transitions determined by the scheduling layer is feasible from a process *dynamics* point of view. This is illustrated in Figure 7.3, where in step change sequence 1, a hypothetical process is allowed to settle to steady state prior to making another setpoint change, resulting in two feasible transitions. In step change sequence 2, the process does not settle to steady state prior to making another setpoint change. The result is a violation of the upper constraint bound for the output y .

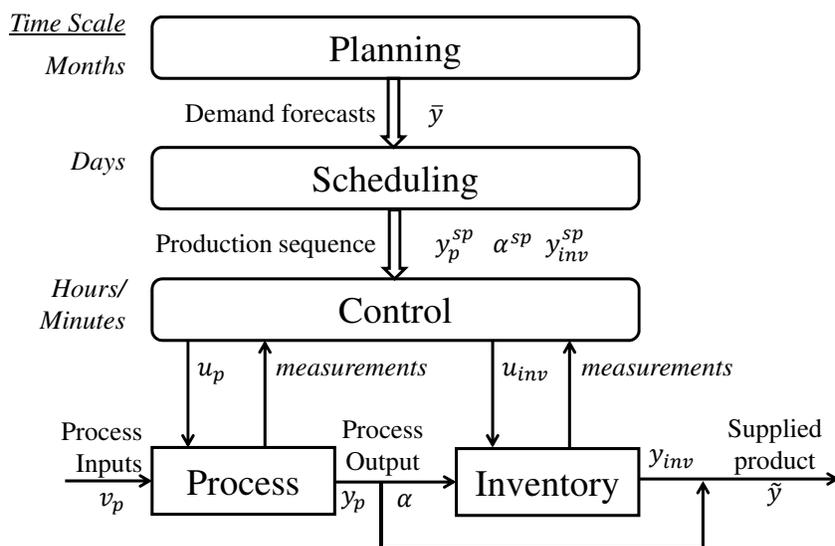


Figure 7.2: Hierarchy of decisions for process operation.

The general problem for production scheduling using dynamic models for a continuous process with product storage capacity and known product demand

rates can be stated as:

Problem Statement

Given:

- Schedule and production information
 - Time horizon
 - Product specifications and demand rates (product quality and quantity constraints)
 - Feedstock and utility prices or price forecasts
 - Length or number of production time slots
- Inventory information
 - Dynamic model of storage system
 - Inventory constraints
- Process information
 - Dynamic process model
 - Process operating constraints

Determine:

- Optimal schedule: production rate and rate of inventory accumulation/ depletion in each time slot

- Optimal operating cost and/or profit for the given time horizon

We describe each of these elements in detail below:

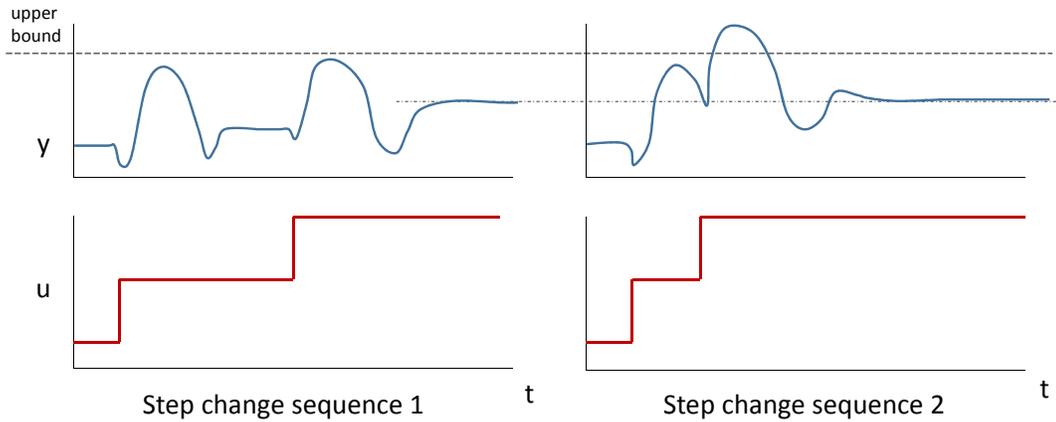


Figure 7.3: Left: a hypothetical process is allowed to settle to steady state prior to changing the setpoint resulting in a dynamically feasible transition. Right: the process does not reach steady state prior to the setpoint change, potentially leading to temporary constraint violations during the transition.

Schedule and Production Information

Time horizon and feedstock/utility prices: The time horizon for scheduling (or makespan) T_m should span the duration for which accurate price forecasts $p(t)$ and product specifications and demand rates are expected to be available. For example, in the context of electric utilities, energy prices may be known with reasonable certainty over a 3 day time horizon [121]. Predicted prices for a longer time horizon are less certain and likely necessitate optimal scheduling decisions under uncertainty and/or re-scheduling over a moving horizon [302].

Table 7.1: Nomenclature

Variable	Description
α	Fraction of process output which bypasses the storage unit
α^{sp}	Setpoint for α
n	Index for scheduling time slots
N_e	Number of time slots in the scheduling horizon
p	Price profile
t	Time
τ	Slot duration
T_m	Makespan/schedule horizon
J	Objective function
\bar{y}	Production target characteristics
\tilde{y}	Product stream characteristics
\hat{y}	Reduced set of scheduling-relevant product variables
y_p	Process output
y_p^{sp}	Setpoint for process output
\hat{y}_p	Reduced set of scheduling-relevant process variables
y_{inv}	Inventory output
y_{inv}^{sp}	Setpoint for inventory output
u_{inv}	Inventory manipulated variables
v_{inv}	Inventory inputs
x_p	Process states
u_p	Process manipulated variables
v_p	Process inputs
z_p	Process algebraic variables
\hat{v}_p	Reduced set of scheduling-relevant process input variables
\hat{w}_p	Reduced set of scheduling-relevant process states

Product specifications and demand rates: \bar{y} provides the desired product specifications and demand rates, which may be either constant in time or time-varying. Without loss of generality, we assume in this work that there is a single product stream. Product quality and production rate constraints are enforced on charac-

teristics, \tilde{y} , of the product stream:

$$h_{product}(\tilde{y}, \bar{y}, t) \leq 0 \quad \forall t \quad (7.1)$$

This set of constraints requires the process output to match the targets set by the production planning layer. We refer to (7.1) as *quality and quantity constraints*, or **QCs**.

Note that we have formulated the demand using a rate rather than a fixed quantity with a known due date. We have chosen this approach because it is more appropriate for the ASU case study, where product is supplied continuously via pipelines.

Time slots: In this case, the time slots are defined by the periods over which the process setpoints are fixed. We assume that there are N_e slots (designated by the superscript $n = 1, 2, \dots, N_e$), each of duration τ^n . Equation (7.2) shows the calculation of the start and end points of the event slots within the scheduling horizon:

$$t_{end}^n = t_{start}^n + \tau^n \quad (7.2a)$$

$$t_{start}^n = t_{end}^{n-1} \quad (7.2b)$$

$$t_{start}^1 = 0 \quad (7.2c)$$

$$t_{end}^{N_e} = T_m \quad (7.2d)$$

Depending on the application, τ_n may be fixed or included in the decision variable set. During each slot, the setpoints $y_p^{sp,n}$, $\alpha^{sp,n}$, and $y_{inv}^{sp,n}$ are the scheduling decision variables. The discrete-time sequence of setpoints is converted to a

continuous-time setpoint signal using:

$$y_p^{sp}(t) = y_p^{sp,n} \quad \forall n, t \in [t_{start}^n \ t_{end}^n) \quad (7.3a)$$

$$\alpha^{sp}(t) = \alpha^{sp,n} \quad \forall n, t \in [t_{start}^n \ t_{end}^n) \quad (7.3b)$$

$$y_{inv}^{sp}(t) = y_{inv}^{sp,n} \quad \forall n, t \in [t_{start}^n \ t_{end}^n) \quad (7.3c)$$

$$y_p^{sp}(T_m) = y_p^{sp,N_e} \quad (7.3d)$$

$$\alpha^{sp}(T_m) = \alpha^{sp,N_e} \quad (7.3e)$$

$$y_{inv}^{sp}(T_m) = y_{inv}^{sp,N_e} \quad (7.3f)$$

Inventory Information

Dynamic model of storage system: In its most general form, the storage system can be modeled as a differential-algebraic equation (DAE) system of the form:

$$f_{inv}(\dot{x}_{inv}, x_{inv}, z_{inv}, u_{inv}, v_{inv}, t) = 0 \quad (7.4a)$$

$$g_{inv}(x_{inv}, z_{inv}, u_{inv}, v_{inv}, t) = 0 \quad (7.4b)$$

$$u_{inv} = K_{inv}(x_{inv}, z_{inv}, y_{inv}^{sp}, t) \quad (7.4c)$$

where $f_{inv} : \mathbb{R}^{2n_{x_{inv}} + n_{z_{inv}} + n_{u_{inv}} + n_{v_{inv}}} \rightarrow \mathbb{R}^{n_{x_{inv}}}$ and $g_{inv} : \mathbb{R}^{n_{x_{inv}} + n_{z_{inv}} + n_{u_{inv}} + n_{v_{inv}}} \rightarrow \mathbb{R}^{n_{z_{inv}}}$ are the differential and algebraic equations in the storage system model, respectively. The variables of the stream exiting the storage system (*e.g.*, flowrate, composition, temperature) are represented by y_{inv} where $\{y_{inv}\} \in [\{x_{inv}\} \cup \{z_{inv}\}]$ and $\{\cdot\}$ denotes set membership. The inlet stream to the storage system, characterized by v_{inv} , is a function of the split fraction $\alpha \in [0, 1]$ of the process outlet stream and the states of the process outlet stream, y_p :

$$v_{inv} = g_{split}(\alpha, y_p) \quad (7.5)$$

where $g_{split_i} = y_{p_i}$ for all of the intensive states (i) of the process outlet stream (*e.g.*, pressure, temperature, composition), and $g_{split_e} = (1 - \alpha)y_{p_e}$ for all of the extensive states (e) of the process outlet stream (*e.g.*, flowrate). The controller K_{inv} modulates the variables u_{inv} , which are typically limited to the flowrate of the stream exiting the storage system (with the setpoint y_{inv}^{sp} defined by the scheduling layer).

While the storage system model presented here is general, we note that in practice (and as will be illustrated in the case study later) the model of such systems are typically fairly simple, with a low number of differential and algebraic variables and corresponding conservation equations.

Holdup restrictions: Knowledge of the inventory is required so that the following *inventory constraints (ICs)* can be enforced at all times.

$$h_{inv}(x_{inv}, z_{inv}, u_{inv}, t) \leq 0 \quad (7.6)$$

Constraints (7.6) include maximum/minimum inventory levels, restrictions on the rate of accumulation/depletion, etc.

Process Information

Dynamic process model: In a dynamic scheduling problem formulation a dynamic process model should be utilized which captures the transient behavior of relevant process variables. In the most general case, the process dynamics can be described by a DAE system of the form [18]:

$$f_p(\dot{x}_p, x_p, z_p, v_p, u_p, t) = 0 \quad (7.7a)$$

$$g_p(x_p, z_p, v_p, u_p, t) = 0 \quad (7.7b)$$

$$u_p = K_p(x_p, z_p, v_p, y_p^{sp}, t) \quad (7.7c)$$

where $f_p : \mathbb{R}^{2n_{x_p} + n_{z_p} + n_{u_p} + n_{v_p}} \rightarrow \mathbb{R}^{n_{x_p}}$ and $g_p : \mathbb{R}^{n_{x_p} + n_{z_p} + n_{u_p} + n_{v_p}} \rightarrow \mathbb{R}^{n_{z_p}}$ are the differential and algebraic equations in the process model, respectively. The process outlet stream variables are defined by y_p where $\{y_p\} \in [\{x_p\} \cup \{z_p\}]$. The manipulated process variables, u_p , are set through either an explicit control law, as in (7.7c), or an advanced control system for which a closed-form control law may not be available. Additionally, we define the set of variables v_p which correspond to the process inputs (utilities and raw materials).

The variables corresponding to the product stream supplied to the customers are computed accounting for the mixing of the outputs of the process and the storage system:

$$\tilde{y} = g_{mix}(y_p, y_{inv}, \alpha) \quad (7.8)$$

For example, the material flowrate supplied to customers for the process in Figure 7.2 is given by: $\tilde{F} = \alpha F_p + F_{inv}^{out}$.

Process operating constraints: The process operating constraints are given by:

$$h_{process}(x_p, z_p, u_p, v_p, y_p^{sp}, t) \leq 0 \quad (7.9)$$

We refer to constraints (7.9) as the *process operating constraints* (**PCs**). These constraints ensure that the operation is feasible and meets safety and equipment restrictions throughout the scheduling horizon. Note that these constraints are different from the **QCs** in (7.1), and that the set of **PCs** is typically larger than the set of **QCs**.

Optimal Scheduling under Dynamic Constraints

The optimal scheduling problem utilizing the detailed dynamic process model can be expressed as:

$$\begin{aligned} & \underset{y_p^{sp,n}, \alpha^{sp,n}, y_{inv}^{sp,n}}{\text{minimize}} & J &= \int_0^{T_m} \phi(p, v_p, y_p, y_{inv}, \tilde{y}) dt & (7.10) \\ & \text{subject to} & & & \end{aligned}$$

Time slots (7.2) – (7.3)

Process model (7.7)

Inventory model (7.4)

Product split (7.5)

Product mixing (7.8)

ICs (7.6)

QCs (7.1)

PCs (7.9)

The objective function, J , which represents the integral of process operating costs ϕ over the course of the scheduling horizon, depends on the process input and output variable trajectories, v_p and y_p , respectively, in addition to the inventory outputs y_p and y_{inv} . We have framed the objective in terms of minimizing operating cost. However, profit maximization could also be considered. The optimization decision variables are the process and inventory quality and quantity setpoint trajectories. In this formulation, we assume that a control law determines the manipulated variable trajectories, but $u(t)$ may also be included in the

optimization objective and decision variable set.

Remark 7.1. *Problem (7.10) assumes that the same product is made at a time-varying production rate. In the case when the process is capable of producing multiple products (or a similar product at different quality grades), an additional set of binary decision variables must be introduced, along with appropriate constraints to define the allocation of products to the production time slots. The modeling approaches discussed in this paper can be naturally extended to this type of problem, and this will constitute the object of future work.*

7.3.1 Challenges of Scheduling under Dynamic Constraints

Equation (7.10) reveals one of the principal challenges posed by optimal scheduling problems under dynamic constraints: detailed, first-principles process models are almost invariably highly-dimensional and highly nonlinear. This makes it very challenging to solve this problem in an amount of time that is sufficiently short to make the solution useful in a practical situation [28, 33]. As a consequence, it is natural to explore the development of *scheduling-oriented* low-order dynamic models that can be embedded in this scheduling formulation. We discuss specific issues related to this task below.

7.3.1.1 Problem Size

The formulation (7.10) falls under the category of *integrated scheduling and dynamic optimization* problems, which is a relatively recently proposed approach for improving process economics [23, 112]. Many case studies reported in the

literature which implement integrated scheduling and dynamic optimization focus on relatively low-dimensional systems, where the dynamic models have a small number of state variables. In this situation, problem (7.10) can typically be solved directly in a reasonable amount of time. This is the case in several works that demonstrate the benefits of scheduling with dynamic knowledge for continuous processes (*e.g.*, polymerization reactors [76, 75, 310, 58, 214]) and batch processes where a dynamic model is used for some of the units (*e.g.*, reactors [61, 192, 193] and separation units [192]) in the sequence of operations.

However, when applied to large-scale, complex systems, the increased computational load caused by using a detailed model can render the problems untractable in a practical time frame, especially if the need to perform rescheduling arises [143]. For this reason, obtaining scheduling-oriented low-order dynamic models representative of process dynamics has recently received attention [76, 75, 126]. Low-order dynamic model-based scheduling is compared to scheduling using a detailed process model in Figure 7.4. The model highlighted in Figure 7.4b is an input-output model that relates the output of the scheduling layer (*i.e.*, the process operating targets and controller setpoints) to the output of the process, and provides predictions of the dynamic behavior of the process when executing the schedule. The scheduling-oriented low-order dynamic model should be designed to have (significantly) fewer states and nonlinearities than the detailed process model (7.7).

There are two broad approaches to deriving low-order dynamic process models. While a comprehensive critical exploration of the extensive literature available on

this topic is beyond the scope of our paper, we provide a brief review below:

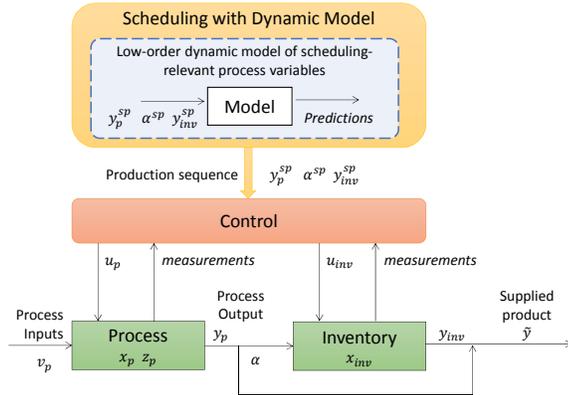
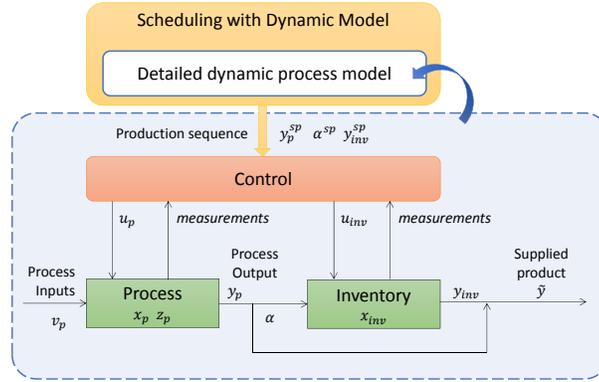
- **Model reduction**, which assumes that a (high-dimensional, first-principles) detailed dynamic process model is available. The derivation of low-order models can then proceed via several avenues: asymptotic analyses based on physical insight and singular perturbation arguments (e.g., [147, 18]), or null-space projection methods [191, 299] are often employed for system models exhibiting multiple-time scale dynamics to eliminate stiffness and reduce the number of states, resulting in a lower-dimensional differential-algebraic equation (DAE). The system can then be solved as-is, or a state-space realization (equivalent ODE representation) can be derived. The advantage presented by such approaches is that they result in models with physically meaningful states. However, these methods can be laborious and their application does require physical insight. Absent such information, empirical nonlinear model reduction methods are also available; these include the use of balanced empirical gramians [110], and the use of empirical eigenfunctions via proper orthogonal decomposition [149]. Empirical methods have the disadvantage of producing models whose states are not physically meaningful.
- Conversely, **system identification** techniques are required when a high-fidelity system model is not available as a starting point. System identification involves deriving a process model from operating data, which are collected in a set of tests during either open- or closed-loop operation. The tests consist of exciting the system inputs, typically by applying step changes; the trend is towards

increasing the efficiency of this process (which can be costly and time consuming) by exciting all inputs *simultaneously* via pseudo-random input sequences, either binary [306] (when the purpose is the identification of a linear model) or multi-level [29](when a nonlinear model is desired). The collected data are then used to perform the system identification/model fitting process. We direct the reader to the text by Ljung[160] for a thorough overview of system identification techniques, and to the book by Zhu [307] for a process systems-centric perspective.

Developing and maintaining high fidelity process models requires considerable technical expertise and financial resources[236] and, consequently, such models are not always available in practical scenarios. Data-driven dynamic system modeling remains widespread in industrial use and motivates our choice of using system identification approaches to develop the scheduling-oriented low-order models used in this work. We note, however, that the framework we propose below is generic, and lends itself naturally to the use of models derived via model reduction when such models are available.

7.3.1.2 Choice of PCs and QCs

The selection of variables and information to be included in the scheduling-oriented low-order dynamic model of a large and complex process is an important consideration. In a situation where the process model is relatively small (*i.e.*, n_{x_p} is low), the ratio of the number of product quality and production rate-related variables to the total number of state variables is close to unity, $\frac{n_{y_p}}{n_{x_p}} \approx 1$, indicating



(b) Scheduling with low-order dynamic models

Figure 7.4: Comparison of integrated scheduling and control modeling choices

that the majority of the state variables are relevant to the scheduling calculation and they likely appear in the $h_{product}$ constraints ((7.1)).

For a more complex process, it is intuitive that $\frac{n_{yp}}{n_{xp}} = \epsilon \ll 1$, *i.e.*, the number of variables relevant to the scheduling calculation is much lower than the number of states. While the selection of variables relevant to the *product quality and production rate constraints* (7.1) may be straightforward, the choice of process variables

relevant to the *process operating constraints* (7.9) requires further analysis. It is likely that only a subset of the PCs are relevant from the scheduling perspective, because not all operating constraints are near their bounds in transient operation. Tracking the PCs is required to ensure feasibility of the process operation throughout the execution of the schedule, so it is crucial to determine the (minimal) set of scheduling-relevant PCs.

These observations provide the motivation for the developments below. Previous efforts [23] on low-order dynamic modeling for scheduling applications focused mainly on multi-product processes with constraints in the form of QCs (*i.e.*, constraints related directly to the production output, manipulated variables, or a measurable operating state such as temperature). In this work, we address the challenge of ensuring that the process operating constraints (PCs) (7.9) are also satisfied throughout the execution of the schedule, while significantly reducing the problem size. Specifically, we explore the development of scheduling-oriented low-order dynamic models which capture the dynamic behavior of the process inputs, outputs and operating constraints in response to production rate and product grade changes. Our models are data-driven and in the single-input, multiple-output format, which, in addition to the reduction in the number of variables, present the advantage of promoting sparsity.

7.4 Scheduling with rPCs and rQCs

Our approach is based on constructing a set of low-order models which accurately describe the dynamics of scheduling-relevant product quality and process

variables, and can be used in the scheduling calculation in lieu of the detailed dynamic model. Our framework for scheduling with scheduling-oriented low-order dynamic models consists of the following steps:

Variable Selection

1. Establish the set of relevant variables to be represented in the scheduling-oriented low-order dynamic models
 - a. Determine the subset of PCs and QCs that are active during production transitions and/or steady state operation
 - b. Identify the subset of variables whose dynamics are relevant to the PCs and QCs
 - c. Identify the subset of process input and output variables whose dynamics are relevant to the scheduling objective

Model Identification

2. Identify low-order dynamic models of the variables selected in Step 1.
 - a. Obtain historical process transition data for model identification
 - b. Determine model form and model parameters for each variable of interest

Scheduling

3. Solve the integrated scheduling and dynamic optimization problem using the low-order dynamic models

In the next sections we describe in detail the steps outlined above.

7.4.1 Step 1: Selection of Variables for Scheduling-Oriented Low-Order Dynamic Models

Step 1a: The selection of operating constraints and process variables relevant to the scheduling calculation is based on the following conjecture:

Conjecture 7.1. *In a complex process with multiple operating constraints related to the process performance, efficiency, and safety, the subset of constraints relevant to the scheduling calculation are the constraints that closely approach or reach their corresponding bounds during steady state operation and/or during transitions between operating points. The behavior of the process variables involved in this subset of active constraints should be captured in scheduling-oriented low-order dynamic models, which are then embedded in the scheduling calculation.*

The operating constraints that are not near their bounds during static or transient operation are not relevant to the optimal scheduling calculation, and thus variable trajectories related to these constraints do not need to be predicted. The subset of PCs and QCs to be included in the low-order problem formulation following the analysis outlined in Conjecture 7.1 are designated, respectively, $\hat{h}_{process}$ and $\hat{h}_{product}$. Note that $dim(\hat{h}_{process}) \leq dim(h_{process})$ and $dim(\hat{h}_{product}) \leq dim(h_{product})$.

The identification of active constraints can be carried out using historical operating data which are often available. These data can be supplemented with a limited set of system identification experiments. Together, these data provide a more complete set of information concerning the operating space of the plant. We note here that the use of a mix of historical data and identification campaign experiments is typical for all data-driven approaches currently employed in model-based process operations, including, *e.g.*, MPC and process scheduling using steady-state models (where it is common practice to define a convex envelope of the operating region of the process). If a high quality model is available, the identification of rPCs and rQCs becomes a matter of exploring the operating domain of the plant through simulations. Alternatively, the tangent space of the Jacobian of the full-order model (if available) can be studied to identify constraints which become active throughout this set of dynamic simulations [194]. However, this approach will not identify variables which approach their constraint bounds but never reach them during the simulation (and could reach or violate these bounds in practical scenarios). Further, neither approach can identify constraints that may become active in new operating scenarios that are very different from those for which data are available (or from those considered in the simulation set mentioned above).

Step 1b: The variables relevant to the subset of QCs are denoted by \hat{y} , and it is necessary to define a subset of process output variables $\{\hat{y}_p\} \subseteq \{y_p\}$ where $\{\cdot\}$ denotes set membership, which are used to calculate \hat{y} :

$$\hat{y} = g_{mix}(\hat{y}_p, y_{inv}, \alpha) \quad (7.11)$$

For example, for the air separation process discussed earlier, the production rate,

impurity concentration and product pressure are QCs which must be satisfied throughout the scheduling horizon. However, we assume that the outlet pressure does not change, and thus is not relevant to the scheduling calculation. Thus, \hat{y} and \hat{y}_p would not include product pressure.

The set of reduced product quantity and quality constraints (rQCs) is thus given by:

$$\hat{h}_{product}(\hat{y}, \bar{y}, t) \leq 0 \quad (7.12)$$

where \bar{y} is the product demand forecast set by the planning layer.

Additionally, we define $\{\hat{w}_p\} \in \{x_p\} \cup \{z_p\}$ as the subset of state and algebraic variables included in the reduced PCs. Without loss of generality and for notation convenience, \hat{w}_p and \hat{y}_p are defined such that they do not have any variables in common (*i.e.*, $\{\hat{w}_p\} \cap \{\hat{y}_p\} = \emptyset$). The reduced process operating constraints (rPCs) are given by:

$$\hat{h}_{process}(\hat{w}_p, y_p^{sp}, t) \leq 0 \quad (7.13)$$

Step 1c: The input and output variables present in the objective (7.10) are v_p , y_p , y_{inv} and \tilde{y} . We assume that the variables \hat{y}_p and \hat{y} , identified as the reduced set of constrained output variables, contain the states of the output stream required to calculate the objective function (otherwise, these sets should be augmented appropriately). We also assume that the inventory model defined in (7.4) is itself low-dimensional and does not require a low-order dynamic model, so it will be used in the original form. The only remaining variable subset to define is \hat{v}_p , the subset of scheduling-relevant input variables. These may include raw material feed

rates or the heating or cooling rate applied to the process. In regards to the ASU case study, the feed air is obtained free of cost and therefore does not need to be included in \hat{v}_p ; however, the operating cost depends on electric power usage which is in turn a function of the feed flow rate; thus, a variable corresponding to the electricity use should be identified.

Evaluating each constraint in (7.13) may require predictions of several variables. Rather than modeling each of these variables individually, \hat{w}_p can represent meaningful groupings of variables (*e.g.*, dimensionless groups/numbers) in order to minimize the required number of low-order dynamic models.

Remark 7.2. *Conjecture 7.1 is conceptually similar to the ideas behind the “self optimizing control” framework [242, 14]. Specifically, determining a self optimizing control structure involves identifying variables that are at or close to their bounds during normal operation and thus may have a considerable impact on economic performance. The control structure is designed such that these variables are controlled at their setpoints. The number of such self-optimizing variables is lower than the number of process outputs (see the previous discussion on the value of the $\frac{n_{yp}}{n_{xp}}$ ratio). Our work extends self-optimizing ideas to the scheduling level, by selecting sets of schedule-relevant process variables which, when maintained within their prescribed limits, ensure that the schedule remains feasible from a dynamic point of view.*

7.4.2 Step 2: Identification of Scheduling-Oriented Low-Order Dynamic Models

In this subsection, we address Step 2 of the proposed approach, which consists of identifying the scheduling-oriented low-order dynamic models used to predict the dynamic behavior of the variables \hat{y}_p , \hat{v}_p , and \hat{w}_p in response to desired changes in the process output.

Step 2a: Low-order dynamic models can be identified either from appropriate transformations of first principles models (see, *e.g.*, [20, 15, 14, 16, 17, 128, 253, 18]) or constructed from experimental data using existing system identification techniques. When available, historical data recorded from step-wise production transitions that cover the process operating space can provide a rich and cost-effective source of information for model identification [170].

Step 2b: The inputs to the set of scheduling-oriented low-order dynamic models are the decision variables of the scheduling calculation, *i.e.*, the process output setpoints y_p^{sp} (since we are using the original model of the storage system, there is no need for low-order models with α^{sp} and y_{inv}^{sp} as inputs). The outputs of the low-order dynamic models are the trajectories of the process input and output variables \hat{v}_p and, respectively, \hat{y}_p , and the trajectories of the constrained process variables \hat{w}_p :

$$F_R(\hat{v}_p, \dot{\hat{v}}_p, y_p^{sp}, t) = 0 \quad (7.14a)$$

$$G_R(\hat{y}_p, \dot{\hat{y}}_p, y_p^{sp}, t) = 0 \quad (7.14b)$$

$$H_R(\hat{w}_p, \dot{\hat{w}}_p, y_p^{sp}, t) = 0 \quad (7.14c)$$

Remark 7.3. *The scheduling-oriented low-order dynamic models are identified from closed-loop data, and inherently reflect the actions and performance of the control system of the plant. Thus, it is not necessary to explicitly model the manipulated variable trajectories, u_p , in the low-order dynamic model.*

Remark 7.4. *In the general case, the low-order models are multiple-input multiple-output, where the inputs are the trajectories of the production targets (e.g., the product purity and product flowrate setpoints), and the outputs are the variables relevant to the scheduling calculation. In the case study presented later in the paper, the models are single-input multiple-output owing to the fact that the product flowrate target is the only time-varying setpoint.*

Remark 7.5. *The proposed framework is agnostic to the format of the scheduling-oriented low-order dynamic models (7.14), and can accommodate any class of dynamic model (linear or nonlinear, continuous- or discrete-time, based on first-principles or empirical). These models should represent the system with a level of accuracy that is appropriate for the application domain (see, e.g., the discussions in Ljung [160], p. 7, and in Seborg et al.[236], p. 16). We will illustrate this point further in the case study presented later in the paper, where we rely on several types of Hammerstein-Wiener models to capture the process dynamics.*

7.4.3 Step 3: Optimal Scheduling Under Low-Order Dynamic Constraints

The scheduling problem utilizing low-order models of the process dynamics can be formulated as:

$$\begin{aligned} & \underset{y_p^{sp,n}, \alpha^{sp,n}, y_{inv}^{sp,n}}{\text{minimize}} & \hat{J} &= \int_0^{T_m} \phi(p(t), \hat{v}(t), \hat{y}_p(t), y_{inv}(t), \hat{\tilde{y}}(t)) & (7.15) \\ & \text{subject to} & & & \end{aligned}$$

Time slots (7.2) – (7.3)

Low-order dynamic process models (7.14a), (7.14b), (7.14c)

Inventory model (7.4)

Production split ratio (7.5)

Mixing (7.11)

ICs (7.6)

rQCs (7.12)

rPCs (7.13)

Problem (7.15) differs from the formulation using the full-order dynamic model (7.10) in several important ways. First, only a subset of the process operating constraints (those that are active or close to their bounds – $\hat{h}_{process}$) are included in the low-order problem formulation. Moreover, the relevant variable trajectory predictions, $\hat{v}_p(t)$, $\hat{y}_p(t)$ and $\hat{w}_p(t)$ are determined using the low-order dynamic models rather than the detailed dynamic process model (7.7).

The inclusion of PCs in addition to QCs distinguishes this formulation from those of the type (7.16):

$$\underset{y_p^{sp,n}, \alpha^{sp,n}, y_{inv}^{sp,n}}{\text{minimize}} \quad \hat{J} = \int_0^{T_m} \phi(p(t), \hat{v}_p(t), \hat{y}_p(t), y_{inv}(t), \hat{\tilde{y}}(t)) \quad (7.16)$$

subject to

Time slots (7.2) – (7.3)

Low-order dynamic process models (7.14a) and (7.14b)

Inventory model (7.4)

Production split ratio (7.5)

Mixing (7.11)

ICs (7.6)

rQCs (7.12)

which are often encountered in the literature, but only consider QCs, and do not consider the process operating constraints and the trajectories of the relevant process variables \hat{w}_p . Problem formulation (7.16) implicitly assumes that a control system will prevent violation of the PCs throughout the production horizon. If this assumption is not valid, the optimal schedule will be infeasible *from a dynamic point of view*. We will compare the results of the formulations (7.15) and (7.16) in the case study.

Remark 7.6. *The key to the proposed dynamic scheduling method is the identification of accurate low-order dynamic models. In the ideal scenario where the low-order dynamic models predict the same trajectories (of the variables relevant to the scheduling calculation) as the detailed process model, the optimal schedule obtained using formulation (7.10) is equivalent to the schedule determined using the formulation given in (7.15), but likely requires much lower computational resources.*

Remark 7.7. *In practice, the low-order dynamic models do not (and most likely, cannot) provide perfect predictions of the scheduling relevant dynamics. This can be due to limitations in the model functional forms, gradual process drift, or unanticipated disturbances. Nevertheless, we note that practitioners in industry face these challenges every day in one of the most successful applications of optimization in process operations, namely, model predictive control. Notably, most MPC implementations rely on a linear and thus (given the nonlinear nature of chemical processes) inherently inaccurate model, to make closed-loop, real-time process operating decisions. Mechanisms such as disturbance modeling [205] have been devised to compensate for unmeasured disturbances and/or unmodeled dynamics, while controller performance assessment techniques are implemented to detect (and potentially correct) for other causes of performance degradation [288]. In spite of these apparent impediments, MPC has become the de-facto standard for advanced process management, with thousands of implementations reported in the last comprehensive survey carried out by Qin and Badgwell (2003) [221], which is already a decade old.*

We expect that similar disturbance modeling and performance assessment techniques can be applied to this proposed scheduling framework with dynamic process models. For example, the proposed dynamic scheduling framework can be implemented in a rolling horizon fashion by periodically recalculating the schedule when new price forecasts are available, or when disturbances or process drift are detected. In effect, the recursive application of the proposed framework can be construed as a new optimal operation paradigm that unites production scheduling and (economic)

model predictive control, and is the subject of ongoing research [11].

7.5 Case Study

We present a case study based on the the motivating example introduced earlier. In the next section, we describe the dynamic model of the air separation unit (ASU). Then, the scheduling-oriented low-order dynamic models proposed above are derived, and the main scheduling problem is presented in addition to some comparative formulations. Finally, the solution of the problem and simulation results are presented and discussed.

7.5.1 ASU Model

In this section, we briefly discuss the mathematical models describing the dynamic behavior of the unit operations of the ASU process shown in Figure 7.1. The detailed dynamic process model is discussed in detail in the thesis by [130], which is in turn based on the models developed by Cao et al.[43, 45].

Distillation Column Model

The cryogenic distillation column model is based on the work by Huang et al. [119]. We assume that (i) the inlet air stream contains only three gases: 78% N_2 , 21% O_2 and 1% Ar , (ii) the vapor phase behaves as an ideal gas, (iii) the material is well-mixed on every stage, (iv) vapor-liquid equilibrium (VLE) is established on each stage, and (v) the column is well insulated and there are no heat losses. The column consists of 30 equilibrium stages and the condenser operating pressure is 6.4

bar with a 0.2 bar linear pressure drop along the column. The phase equilibrium is modeled using an activity model for the non-ideal liquid phase:

$$y_{ij}P_i = \gamma_{ij}P_{ij}^{sat}x_{ij} \quad (7.17)$$

where index i and j represent the stage number and component, respectively. The vapor pressure, P_{ij}^{sat} is determined using Antoine's equation [296] and the activity coefficients, γ_{ij} are determined using the Margules equations [113].

The Material, Equilibrium, Summation and Heat (MESH) equations which describe each equilibrium stage constitute an index-2 system of differential algebraic equations (DAEs). The high index is due to the fact that the vapor flow from each stage (an algebraic variable) is not present in the algebraic equations, and thus cannot be solved for directly. Using the procedure outlined in [119] the index was reduced to one.

Integrated Reboiler/Condenser Model

The liquid at the bottom of the column is expanded adiabatically to 2.5 bar to condense the vapor at the top of the column in a heat-integrated reboiler/condenser. The model for the integrated reboiler/condenser is adapted from Cao [43].

The following assumptions are made for the condenser model: (i) fast dynamics (*i.e.*, material or energy accumulation are not considered) (ii) the condensed liquid is saturated, (iii) the outlet liquid composition is the same as the composition of the vapor inlet from the top of the column, and (iii) the heat duty required to

condense the vapor inlet stream can be supplied completely by the reboiler.

The reboiler is modeled as an equilibrium stage, with an additional heat input equivalent to the condenser heat duty. A proportional controller is implemented to maintain the reboiler liquid level by manipulating the liquid drain rate. The liquid waste drain rate is typically very small in order to minimize energy loss from the process [43].

Primary Heat Exchanger (PHX) Model

The PHX is a brazed aluminum plate-fin multistream heat exchanger and the corresponding model is adapted from the structure described by Cao [43]. The model consists of two zones (see Figure 7.1) which are delimited by the location where the inlet air gas stream is withdrawn from the PHX. The fraction of vapor removed prior to zone 2 is a manipulated variable at the control level. The two zones correspond, respectively, to sensible and latent heat removal from the inlet air stream, and the corresponding temperature changes of the product and waste streams. The first zone is further discretized into 50 segments, while the second zone, in which a portion of the inlet air stream is completely liquefied, is modeled by a single lumped energy balance equation to simplify the phase transformation calculations. Within each zone, the geometry of the channels created by the plates/fins is accounted for when calculating energy accumulation of each stream in each finite volume [43].

Compressor and Turbine Models

The compressor is the main energy consumer in the air separation process. Generators are coupled to the turbine expanders used in the process and serve to partially meet the compressor power demand. We assume that the dynamics of turbines, compressors and generators are fast and these units can be modeled using steady state equations. In order to calculate the power demand of the compressor (W_c), and power generated by the two turbines (W_{t1}, W_{t2}), we assume that the compression and expansion are polytropic processes with corresponding head and efficiencies calculated using the approach presented in Chapter 10 of [105].

Liquefier and Liquid Storage Tank Model

A liquefier is included in the process to liquefy a portion of the gaseous nitrogen product. A liquid nitrogen storage tank accumulates the liquefied nitrogen and an evaporator vaporizes the liquid before delivering the gas to customers. We assume that the physical dimensions of the liquefier are much smaller than those of the plant, and, unlike the ASU, the liquefier does not contain any significant material holdups (e.g., sumps). As a consequence, it is to be expected that the dynamics of the liquefier are much faster than those of the plant itself. As a consequence, we model the liquefier using the steady state versions of the corresponding material and energy balance equations. Further, we assume that the liquefier operates in an ideal refrigeration cycle with a constant 40% efficiency. The liquefier power demand (W_l) is computed based on the net work of the compressor and turbine. We assume that the evaporator is at ambient conditions and does not require any

additional energy input to operate.

The storage tank is sized such that, when full, it holds enough nitrogen in the tank to satisfy the demand rate with the plant operating at its lowest production level for 10 hours. The holdup, M_{inv} , is given by:

$$\frac{dM_{inv}}{dt} = F_{inv}^{in} - F_{inv}^{out} \quad (7.18a)$$

$$M_{inv}(0) = M_{inv}^0 \quad (7.18b)$$

The scheduling problem formulation requires the storage system model and constraints (ICs). The storage tank model consists of a mass balance equation (7.18), and the holdup M_{inv} is constrained such that the inventory is always greater or equal to zero, and never exceeds the maximum storage capacity M_{inv}^{max} :

$$0 \leq M_{inv} \leq M_{inv}^{max} \quad \forall t \quad (7.19)$$

Additionally, it is required that the holdup at the end of the scheduling horizon T_m be greater or equal to a minimum terminal value (M_{min}):

$$M_{inv}(T_m) \geq M_{min} \quad (7.20)$$

In this case, we set M_{min} equal to $M_{inv}(0)$ to ensure that inventory is not depleted throughout the horizon. Note that when the scheduling framework is implemented in a rolling horizon fashion, it is beneficial to fix the terminal constraint to ensure recursive feasibility and stability [7].

We use simple heuristics to determine α and F_{inv}^{out} based on the demand rate \bar{F} , such that demand is satisfied exactly:

$$\alpha = \begin{cases} \frac{\bar{F}}{F_p} & \text{if } F_p \geq \bar{F} \\ 1 & \text{if } F_p < \bar{F} \end{cases} \quad (7.21)$$

$$F_{inv}^{out} = \begin{cases} 0 & \text{if } F_p \geq \bar{F} \\ \bar{F} - F_p & \text{if } F_p < \bar{F} \end{cases} \quad (7.22)$$

The inlet flowrate to the storage system is calculated by the split equation:

$$F_{inv}^{in} = (1 - \alpha)F_p \quad (7.23)$$

and the product flowrate is given by the mixing equation:

$$\tilde{F} = \alpha F_p + F_{inv}^{out} \quad (7.24)$$

By including these heuristics, y_{inv}^{sp} and α^{sp} are not decision variables in our optimization formulation, leaving the process production rate F_p as the main scheduling decision variable.

Process Operation

We assume a constant nitrogen demand of 20mol/s at a purity greater than 99.8%, which corresponds to a total impurity (oxygen and argon) concentration of less than 2000 ppm. We assume that the production rate can deviate by up to $\pm 20\%$ from the nominal value. The total power required to operate the plant $P^{total}(t)$ is given by:

$$P^{total}(t) = W_c(t) + W_l(t) - W_{t1}(t) - W_{t2}(t) \quad (7.25)$$

In this case, the net work of the compressor, liquefier, and turbines is proportional to the flow rate through each unit. This is due to the fact that the process operates at constant pressure between production rate changes, and the inlet temperatures do not change significantly, which results in a nearly constant polytropic head [130].

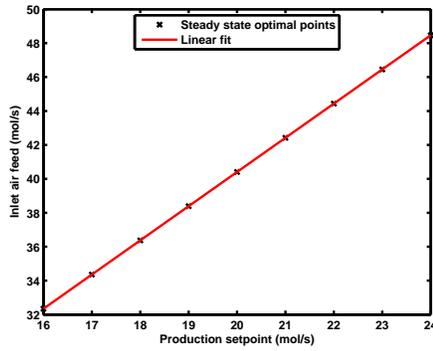
We assume that electricity is purchased from a utility company at market rates which fluctuate hourly, but are forecasted accurately for a three day horizon. In order to minimize operating costs, the production level will be lowered during high price periods, and increased during low price periods with the assumption that production rate setpoints may only change hourly.

Additionally, we assume that transitions between production levels are handled using a heuristic that mimics an operator's approach to adjusting the manipulated variables. The transition control heuristics are described in detail in the next section.

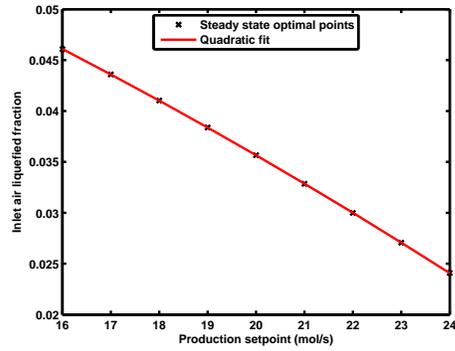
Transition Control

The manipulated variables of the process are the feed air flow rate, F_{air}^{in} , the split of the inlet air liquefied in the PHX, K_{PHX} , and the column reflux ratio, R_{col} . To mimic the actions of an operator, a heuristic control law was set to adjust the manipulated variables through any possible production rate change sequence. First, the steady state values of the manipulated variables were determined at 9 different steady-state production rates such that i) the production rate matched the target, ii) the impurity level was 500 ppm, and iii) the energy consumption was minimized. Polynomial curves were fitted to approximate the optimal values across the entire range of possible production rate setpoints. These can be seen in Figure 7.5.

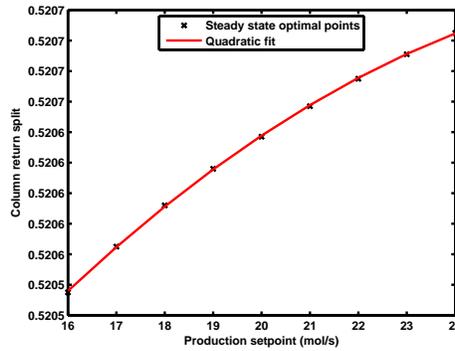
The piecewise linear control heuristic for determining the trajectory of the manipulated variables during a production setpoint transition between any two points



(a) Inlet air feed



(b) Inlet air liquefied fraction



(c) Column return split

Figure 7.5: Optimal steady state values of the manipulated variables as a function of production setpoint level

within $\pm 20\%$ of the nominal production flowrate was determined such that the deviations between the production flowrate and the setpoint are minimized. The trajectories of the reflux ratio, R_{col} , and the fraction of inlet air liquefied in the PHX, K_{PHX} , consist of a piecewise linear function with 2 segments, where the intermediate point, or peak of the trajectory, is determined as a function of the magnitude and direction of the setpoint change. The piecewise linear control law

(7.7c) for R_{col} at time slot n is given by:

$$\frac{dR_{col}}{dt} = \begin{cases} \frac{R_{col}^{P,n} - R_{col}^{SS,n-1}}{\frac{1}{2}T^{P,n}} & \text{if } t_{start}^n \leq t < t_{start}^n + \frac{1}{2}T^{P,n} \\ \frac{R_{col}^{SS,n} - R_{col}^{P,n}}{\frac{1}{2}T^{P,n}} & \text{if } t_{start}^n + \frac{1}{2}T^{P,n} \leq t < t_{start}^n + T^{P,n} \\ 0 & \text{if } t_{start}^n + T^{P,n} \leq t < t_{end}^n \end{cases} \quad (7.26)$$

where $R_{col}^{P,n}$ is the peak of the trajectory in time slot n which occurs at $t = t_{start}^n + \frac{1}{2}T^{P,n}$ (notice the ‘‘peak’’ of the column return split trajectory ($R_{col}^{P,n}$) in Figure 7.6), and $R_{col}^{SS,n}$ is the steady state optimal value at time slot n , which is obtained using the polynomial fit in Figure 7.5c. $T^{P,n}$ is the length of the transition time for the manipulated variables in time slot n . Likewise, the piecewise linear control law (7.7c) for K_{PHX} at time slot n is given by:

$$\frac{dK_{PHX}}{dt} = \begin{cases} \frac{K_{PHX}^{P,n} - K_{PHX}^{SS,n-1}}{\frac{1}{2}T^{P,n}} & \text{if } t_{start}^n \leq t < t_{start}^n + \frac{1}{2}T^{P,n} \\ \frac{K_{PHX}^{SS,n} - K_{PHX}^{P,n}}{\frac{1}{2}T^{P,n}} & \text{if } t_{start}^n + \frac{1}{2}T^{P,n} \leq t < t_{start}^n + T^{P,n} \\ 0 & \text{if } t_{start}^n + T^{P,n} \leq t < t_{end}^n \end{cases} \quad (7.27)$$

again, note the ‘‘peak’’ of the inlet air liquefied fraction trajectory ($K_{PHX}^{P,n}$) in Figure 7.6. The piecewise linear control law ((7.7c)) for F_{air}^{in} is:

$$\frac{dF_{air}^{in}}{dt} = \begin{cases} \frac{F_{air}^{in,SS,n} - F_{air}^{in,SS,n-1}}{\frac{1}{2}T^{P,n}} & \text{if } t_{start}^n \leq t < t_{start}^n + \frac{1}{2}T^{P,n} \\ 0 & \text{if } t_{start}^n + \frac{1}{2}T^{P,n} \leq t < t_{end}^n \end{cases} \quad (7.28)$$

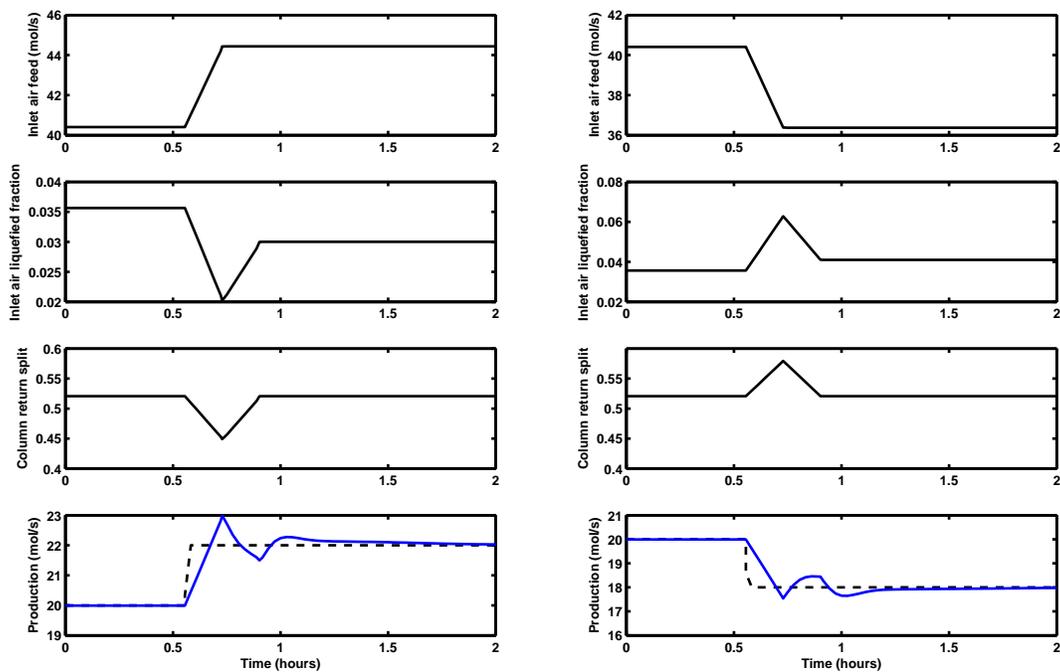
$R_{col}^{P,n}$, $K_{PHX}^{P,n}$, and $T^{P,n}$ are determined as a function of the production rate target change ($F_p^{sp,n} - F_p^{sp,n-1}$):

$$R_{col}^{P,n} = a_R(F_p^{sp,n} - F_p^{sp,n-1}) + b_R \quad (7.29a)$$

$$K_{PHX}^{P,n} = a_K(F_p^{sp,n} - F_p^{sp,n-1}) + b_K \quad (7.29b)$$

$$T^{P,n} = a_T(F_p^{sp,n} - F_p^{sp,n-1}) + b_T \quad (7.29c)$$

where a_R , a_K , a_T , b_R , b_K , and b_T are constants. As described in Equation (7.28), the inlet air flowrate changes to its new steady state value in half the manipulated variable transition time ($\frac{1}{2}T^{P,n}$). The trajectories of the manipulated variables and the corresponding production rates are illustrated in Figure 7.6 for a 10% increase (Figure 7.6a) and decrease (Figure 7.6b) in the setpoint. The production (the control variable) overshoots the setpoint, but quickly settles to the desired value.



(a) 10% increase in production setpoint

(b) 10% decrease in production setpoint

Figure 7.6: Manipulated variable trajectories and production rate during setpoint changes

Remark 7.8. *The control law above is open-loop, and thus offset-free tracking cannot be guaranteed. However, it is representative of operator actions during pro-*

duction setpoint changes (which are often carried out manually). Additional PI controllers would likely be implemented to ensure offset-free control of the production flow rate once the transition is complete and the process is near the target production rate.

7.5.2 Scheduling-Oriented Low-Order Dynamic Modeling of ASU Dynamics

We follow the procedure outlined in the theoretical part of the paper to derive the relevant rPCs, rQCs, and associated variables for the ASU process.

Product Quality and Production Rate Constraints (QCs)

Step 1a requires that the relevant product quality and production rate constraints (QCs) be identified:

QC1 The total production flowrate must be greater than or equal to the demand throughout the production horizon

QC2 Impurity levels in the product stream must be lower than the maximum allowed value

QC3 The product stream pressure must be maintained close to 1.3bar throughout the schedule

QC1 It is assumed that a perfect forecast of product demand \bar{F} is given throughout the horizon (here we assume that it is constant) and it must be satisfied at all

times using the product stream \tilde{F} , which was defined in (7.24):

$$\tilde{F} \geq \bar{F} \quad \forall t \quad (7.30)$$

QC2 The impurity level (\tilde{I}) in the product is reported in parts per million (ppm) and calculated based on the mole fraction of nitrogen, x_{p,N_2} , in the stream as $\tilde{I} = 10^6(1 - x_{p,N_2})$. The impurity level is required to be below a threshold level $I^{max} = 2000ppm$.

$$\tilde{I} = \frac{F_{inv}^{out} I_{inv} + F_p I_p}{\tilde{F}} \leq I^{max} \quad \forall t \quad (7.31)$$

Given the critical nature of this constraint, and to compensate for possible inaccuracies in the low-order impurity model, a “back-off” is used, setting the threshold to $\hat{I}^{max} = 1800ppm$ during scheduling calculations based on the low-order models. We note that such “back-off” from active constraints is implemented in many practical situations to avoid infeasible operation in the presence of disturbances or model error. We refer the reader to the works by Aske et al.[12] and the earlier work by Narraway and Perkins[188] for more details.

Rather than constraining the impurity levels in the stream supplied to the customer (\tilde{I}), we constrain the process output impurity (I_p):

$$I_p \leq \hat{I}^{max} \quad \forall t \quad (7.32)$$

While (7.32) is a more restrictive constraint than imposing a bound on impurity levels in the stream supplied to the customer, it guarantees that constraint (7.31) is satisfied and avoids the need to model the impurity concentration in the storage system, I_{inv} .

QC3 The pressure of the product stream must be within specified limits:

$$\bar{P}_l \leq \tilde{P} \leq \bar{P}_u \quad (7.33)$$

In the process model, the pressure in each piece of equipment is fixed to ensure that the outlet pressure is 1.3 bar at all times. As a result, constraint QC3 is inherently satisfied. Thus, only constraints QC1 and QC2 are relevant to the scheduling problem, and (7.30) and (7.32) constitute the rQCs.

Process Operating Constraints (PCs)

Step 1a also requires that we identify the scheduling-relevant process operating constraints (PCs). These are:

PC1 No weeping in the distillation column

PC2 The liquid level in the sump of the column should never go to zero or beyond the maximum capacity

PC3 No surging in the compressor

PC4 The liquid level in the reboiler should never go to zero or beyond the maximum capacity

PC5 The reboiler holdup at the end of the scheduling horizon should be at least as much as the initial holdup

PC6 All streams in the first zone of the PHX must be in the gas phase

PC7 The air stream exiting the second zone of the PHX must be in the liquid phase

PC8 No flooding in the distillation column

PC9 The temperature driving force across the integrated reboiler/condenser must be greater than a lower limit

Through observation of the process transition data (which are discussed later in the paper), it was determined that constraints PC1 - PC4 are not active during steady state operation or during transitions. These constraints therefore present no limitation to the feasibility of the production schedule, and modeling the corresponding dynamic behaviors during transitions is not required. Constraints PC5 - PC9 may, however, become active, and therefore make up the set of rPCs, which will be described below.

PC5 To ensure that refrigeration stored within the process is not depleted, a constraint must be enforced that requires the reboiler holdup to be greater or equal to a terminal value:

$$M_{reb}(T_m) \geq M_{reb}^{min} \quad (7.34)$$

As with the inventory constraint, we select M_{reb}^{min} equal to the inventory level at the start time, $M_{reb}(0)$.

PC6 and PC7 As discussed earlier in the paper, the primary heat exchanger is divided into two zones, whereby only sensible heat is removed from the incoming air stream in zone 1, and the inlet air is condensed in zone 2. To ensure that no

phase transformations occur in zone 1, the air feed stream at the outlet of zone 1 is constrained to be in the vapor phase at all times by requiring the outlet pressure (P_{zone1}) to be less than 96% of the dew pressure of air at the outlet temperature (P_{dew}):

$$\frac{P_{zone1}}{P_{dew}} \leq 0.96 \quad \forall t \quad (7.35)$$

Additionally, a constraint is imposed to ensure that the air feed stream at the outlet of zone 2 is in the liquid phase:

$$\frac{P_{zone2}}{P_{bubble}} \geq 1.05 \quad \forall t \quad (7.36)$$

where P_{zone2} is the pressure at the outlet of zone 2 and P_{bubble} is the corresponding bubble point pressure.

PC8 Flooding of distillation columns is a faulty operating state that occurs when liquid from a stage (or a subset of stages) is carried to the stages immediately above due to an excessively large vapor flow rate. Flooding drastically reduces stage efficiencies and increases the column pressure drop. We impose a flooding constraint based on the work of Coulson and Richardson (as described in [241]) and Cao [43]. The flooding fraction for each tray i , $\delta_i^{flooding}$, is constrained to be below 97%:

$$\delta_i^{flooding} = \frac{u_i}{u_{flooding,i}} < 0.97 \quad \forall i, t \quad (7.37)$$

where u_i is the vapor velocity at tray i [m/s], and $u_{flooding,i}$ is the flooding velocity at tray i [m/s] (the velocity calculations are detailed in the thesis by [130]). However, evaluating the flooding fraction for each tray is not necessary because there are typically only a few trays in the column that are susceptible to flooding,

and flooding usually happens simultaneously on these trays. We therefore define the maximum flooding level in the column, $\delta_{max}^{flooding}$, and use this single variable to evaluate the flooding constraint.

$$\delta_{max}^{flooding} = \max_i(\delta_i^{flooding}) < 0.97 \quad \forall t \quad (7.38)$$

PC9 To ensure that the heat flow rate $Q = UA(T_{condenser} - T_{reboiler})$ required to condense the liquid in the condenser is positive, a minimum temperature driving force constraint must be met at all times:

$$T_{condenser} - T_{reboiler} \geq \Delta T_{min} \quad \forall t \quad (7.39)$$

The minimum approach temperature is set at $\Delta T_{min} = 1.9^\circ C$.

Equations (7.34), (7.35), (7.36), (7.38) and (7.39) represent the set of rPCs.

Step 1b consists of identifying the variables whose trajectories are relevant to the rQCs and rPCs. To ensure that the rQCs are satisfied throughout the scheduling horizon, a dynamic model of F_p is required to evaluate constraint QC1, and we use a dynamic model of the evolution of the product output impurity (I_p) to evaluate constraint QC2. Similarly, several variable trajectories must be predicted to evaluate the rPCs. Constraint PC5 can be evaluated by using the level in the reboiler. Constraints PC6 and PC7 can be evaluated from the ratio of the pressure to the dew and bubble point pressures, respectively (*e.g.*, instead of modeling P_{zone1} and P_{dew} separately and then evaluating the constraint). The flooding constraints PC8 can be evaluated from a single variable ($\delta_{max}^{flooding}$), and constraint (7.39) can be evaluated from the temperature difference across the reboiler/condenser.

Step 1c requires the selection of variables relevant to the calculation of the scheduling objective function. The objective function in this case study captures the total electricity cost throughout the production horizon, which is described in equation (7.25). As stated earlier in the paper, the rate of electricity consumption/production in the compressors/turbines is proportional to the flowrates of the respective units, and the level of electricity consumption in the liquefier is proportional to the flow rate of the material directed to the storage system. We define the following correlations for electricity consumption:

$$W_c = \gamma_c F_{feed} \quad (7.40a)$$

$$W_{t1} = \gamma_{t1} F_{feed} \quad (7.40b)$$

$$W_{t2} = \gamma_{t2} F_p \quad (7.40c)$$

$$W_l = \gamma_l F_{inv}^{in} \quad (7.40d)$$

where γ_c , γ_{t1} , γ_{t2} , and γ_l are parameters computed based on the polytropic heads, which are nearly constant irrespective of the production rate. F_{inv}^{in} is calculated from F_p in equation (7.5). Thus, the variables required to evaluate the objective function are F_{feed} and F_p .

Based on the discussion above, the variables whose dynamic behavior is relevant to the scheduling problem are listed and categorized in Table 7.2.

7.5.3 Building ASU Low-Order Dynamic Models

In this section, we describe the system identification approach we followed to establish the low-order dynamic models for each variable in \hat{y}_p , \hat{w}_p , and \hat{v}_p in Table

Table 7.2: Variables described by the scheduling-oriented low-order dynamic models

Variable type	Variables
\hat{y}_p (process outputs)	F_p I_p
\hat{v}_p (process inputs)	F_{feed}
\hat{w}_p (constrained process variables)	M_{reb} $\delta_{max}^{flooding}$ P_{zone2}/P_{bub} P_{zone1}/P_{dew} $T_{condenser} - T_{reboiler}$

7.2.

7.5.3.1 Data Collection

A training dataset is required to identify scheduling-relevant low-order dynamic models. These data should describe the dynamic behavior of the process variables during production transitions. This information can be obtained from historical process data when available, or obtained from a system identification campaign at the plant site. The dataset should include a series of setpoint transitions which cover the range of potential operation setpoints, and the sampling frequency must be high enough to capture all of the process time scales. In particular, we refer to historic data collected from operations that did involve production target transitions of the kind considered in the scheduling procedure. Such production rate transitions are in many cases effectively equivalent to the step tests imposed during dedicated system identification experiments and, as such, the resulting historical

data can be used for the purpose of building the proposed low-order dynamic models.

For this work, process transition data were generated using the detailed process model described in the previous section. Specifically, to generate a training dataset with sufficiently rich dynamic information, a production target trajectory was determined by optimizing the production schedule using a simplified (static) process model, where it was assumed that the production flowrate and impurity levels match their setpoints following each transition. This formulation is representative of a conventional scheduling approach, where process dynamics are not explicitly modeled and the transition information is captured in tabulated transition time data. The schedule optimization problem used to create this “static” schedule is described below:

$$\begin{aligned}
& \underset{F_p^{sp,n}}{\text{minimize}} && \phi = \int_0^T p(t)(\gamma_c F_p^{sp}(t) - \gamma_{t1} \hat{F}_{feed} - \gamma_{t2} F_p^{sp} + \gamma_l F_{inv}^{in}) dt && (7.41) \\
& \text{subject to} && \text{Time slots (7.2) – (7.3)} \\
& && \text{Inventory model: (7.18)} \\
& && \text{Split: (7.21) and (7.23)} \\
& && \text{Mixing: (7.22) and (7.24)} \\
& && \text{ICs: (7.19) – (7.20)} \\
& && \text{Demand Constraint: (7.30)} \\
& && abs(F_P^{sp,n+1} - F_p^{sp,n}) \leq 2mol/s && (7.42) \\
& && 0.8 \leq F_p^{sp,n} / F_p^{nom} \leq 1.2 && (7.43) \\
& && \hat{F}_{feed} = \beta F_p^{sp} && (7.44)
\end{aligned}$$

where F_{feed} is predicted from F_p^{sp} using the static model in equation (7.44). The electricity price data used in this optimization and throughout the case study are given in Figure 7.7.

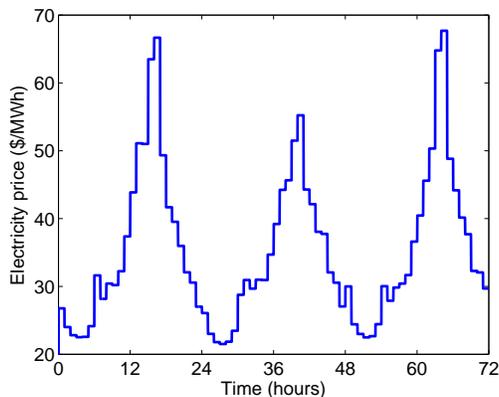


Figure 7.7: Data from the ERCOT day ahead market[88] for electricity market prices on September 9-11, 2013.

To take process dynamics into account, we implemented constraint (7.42), which limits the product flow rate target changes to $2\text{mol}/s$ (or 10% of the nominal production rate) over the course of an hour, and constraint (7.43), which constrains the production rate to within $\pm 20\%$ of the nominal value. Separate production target step change experiments indicated that any change of $\pm 2\text{mol}/s$ from any production level within $\pm 20\%$ of the nominal rate is feasible if the system is allowed to settle to steady state. However, as will be shown below, it cannot be guaranteed that any *sequence* of step changes is feasible if the system does not reach steady state prior to a new step change.

The optimal production rate target sequence determined by solving (7.41) is shown in Figure 7.8a. Notice that it covers the entire range of allowable production

rate levels (*i.e.*, $\pm 20\%$ of the nominal product flowrate). The optimal schedule was implemented on the detailed dynamic model using the transition control heuristic discussed in the Transition Control section in the ASU Model description. The corresponding trajectories of the variables of interest (given in Table 7.2) during the prediction horizon make up the training dataset which was used to identify the desired scheduling-oriented low-order dynamic models.

It is important to note that, while the manipulated variable trajectories reach their new steady state values within one hour, the time to steady state for the slowest process variables (specifically, impurity levels) is much longer. Despite individual $\pm 20\%$ changes in the production rate being feasible when the process is given time to settle to steady state, the implementation of production setpoint sequence obtained from solving (7.41) results in constraint violations during production transitions because the process dynamics are not adequately taken into account. Specifically, the product impurity levels and the reboiler/condenser temperature driving force constraints are violated on several occasions (see Figure 7.8). Nevertheless, the results presented in Figure 7.8 constitute a valid dataset for model identification purposes.

7.5.3.2 Low-Order Dynamic Model Identification

To determine good initial estimates of the model order and time constants for variables with longer dynamics, an additional dataset was generated, where the production rate transitions occur infrequently (*i.e.*, enough time is allowed to elapse for all process variables to reach steady state). Note that this dataset was

not used to train the models; rather, it assisted with the pre- and post-processing phases of our system identification effort, in that the data were used to establish the appropriate model form and order, and, respectively, to validate the models once the identification was completed. In practice, this strategy would correspond to the additional (limited) set of step test experiments required to build the scheduling-relevant low-order dynamic models as described earlier in the paper.

Low-order dynamic models were identified for all variables in Table 7.2. The models were trained using the historical transition data given in Figure 7.8 using the System Identification Toolbox in MATLAB [177]. Continuous time, nonlinear Hammerstein-Wiener models of the general form

$$u' = \Psi(y_p^{sp}) \quad (7.45a)$$

$$\dot{\bar{x}} = A\bar{x} + Bu' \quad (7.45b)$$

$$y' = C\bar{x} \quad (7.45c)$$

$$\hat{z}_p = \Phi(y') \quad (7.45d)$$

were identified for each variable. Here, Ψ and Φ are the input and output non-linearity functions, respectively, and A , B and C are the matrices of the linear state-space model. \hat{z}_p represents the output variable ($\hat{z}_p = [\hat{y}_p, \hat{v}_p, \hat{w}_p]$). Input nonlinearities were represented as piecewise linear functions:

$$\Psi(y_p^{sp}) = \frac{pw_{i+1} - pw_i}{bp_{i+1} - bp_i} (y_p^{sp} - bp_i) + pw_i \quad \text{if } bp_i < y_p^{sp} \leq bp_{i+1} \quad (7.46)$$

where $i \in I$ is the set of piecewise linear segments, bp_i is the breakpoint at segment i , and pw_i is the value of the input function for segment i . The output nonlinearities

were represented either as polynomials of the form:

$$\Phi(y') = \sum_{j=0}^N n_j y'^j \quad (7.47)$$

where n_j are the polynomial coefficients, or by piecewise linear functions:

$$\Phi(y') = \frac{pw_{i+1} - pw_i}{bp_{i+1} - bp_i} (y' - bp_i) + pw_i \quad \text{if } bp_i < y' \leq bp_{i+1} \quad (7.48)$$

To fit the models, the order of the linear state space models, the number of piecewise segments, and the order of the polynomials were adjusted in a trial-and-error fashion, with the model that resulted in the closest fit (i.e., lowest normalized mean square error - NMSE) being retained for each variable. An overview of the resulting scheduling-oriented low-order dynamic models is given in Table 7.3, while full details are provided as supplementary material. The predicted outputs of each low-order dynamic model, along with the corresponding variable trajectories computed using the full-order model, are shown in Figure 7.8.

The statistical analysis of our models attests to their quality, as reflected by the high normalized mean square error (NMSE) values provided in Table 7.3 for the training and validation data. Note that the NMSE values for the validation data in some cases are higher than the training data; this is due, in part, to the fact that the validation data has a longer time horizon with fewer switches, and reflects a high quality prediction of the steady state gain. The low NMSE in the prediction of impurity for the validation dataset reflects the need to implement a constraint “back-off” as discussed above (Equation (7.32)).

Table 7.3: Identified model details

Variable	Number of piecewise linear input segments	Linear system order	Output model type	Output polynomial order	Number of piecewise linear output segments	NMSE (training)	NMSE (validation)
F_p	5	3	Polynomial	2	–	0.96	0.99
I_p	4	4	PW Linear	–	6	0.82	0.52
F_{feed}	3	2	Polynomial	2	–	0.99	0.99
M_{reb}	3	4	Polynomial	1	–	0.78	0.75
$\delta_{max}^{flooding}$	5	5	Polynomial	2	–	0.91	0.92
P_{zone2}/P_{bub}	8	4	Polynomial	2	–	0.78	0.72
P_{zone1}/P_{dew}	2	8	PW Linear	–	6	0.83	0.97
$(T_{condenser} - T_{reboiler})$	9	4	Polynomial	2	–	0.69	0.84

7.5.4 ASU Scheduling Under Low-Order Dynamic Constraints

The complete ASU scheduling formulation (based on the general form (7.15)) using scheduling-oriented low-order dynamic models for product and process dynamics is given by:

$$\begin{aligned} & \underset{F_p^{sp,n}}{\text{minimize}} && \phi = \int_0^{T_m} p(t)(\gamma_c \hat{F}_p(t) - \gamma_{t1} \hat{F}_{feed} - \gamma_{t2} \hat{F}_p + \gamma_l \hat{F}_{inv}^{in}) dt && (7.49) \\ & \text{subject to} && \text{Time slots (7.2) – (7.3)} \\ & && \hat{y}_p \text{ models of the form (7.45)} \\ & && \hat{v}_p \text{ models of the form (7.45)} \\ & && \hat{w}_p \text{ models of the form (7.45)} \\ & && \text{Inventory model: (7.18)} \\ & && \text{Split: (7.21) and (7.23)} \\ & && \text{Mixing: (7.22) and (7.24)} \end{aligned}$$

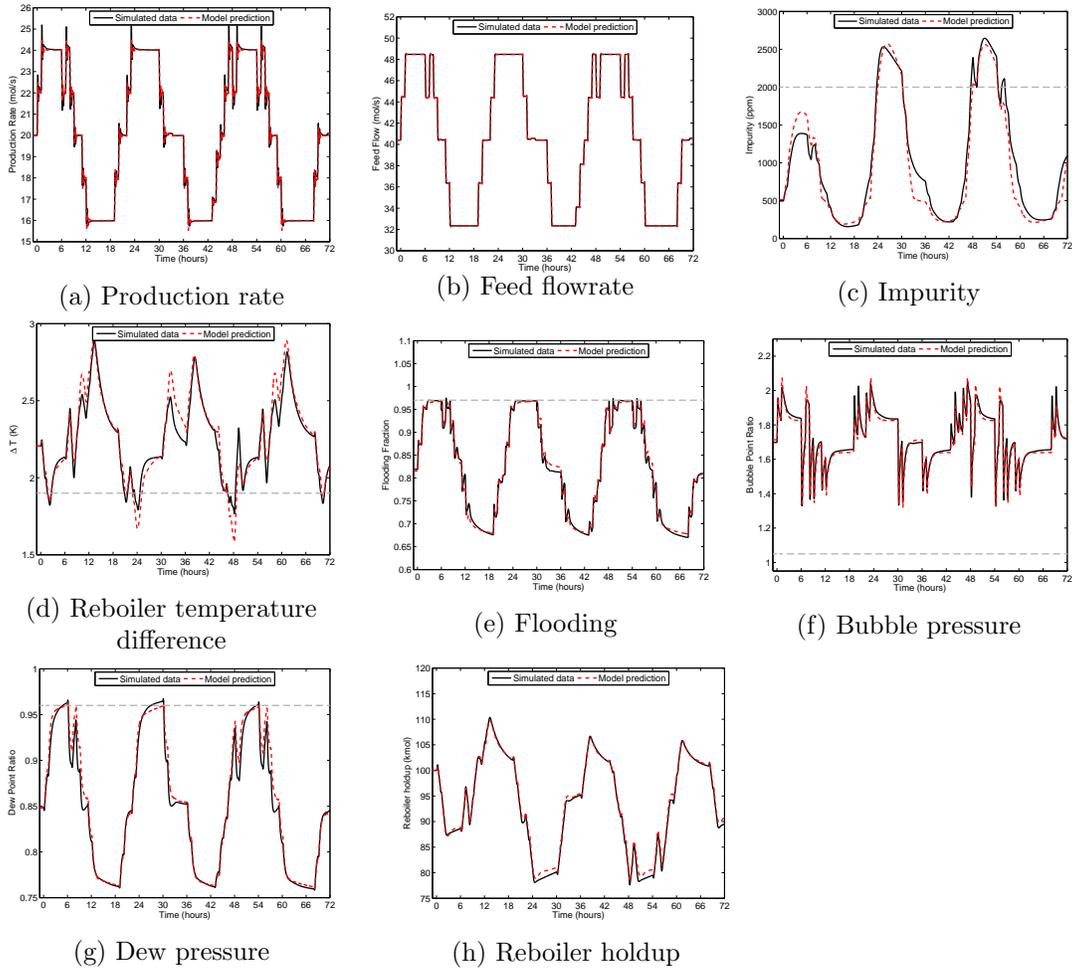


Figure 7.8: Simulated “historical” process transition data and corresponding predictions from the scheduling-oriented low-order dynamic model. Several constraints are violated throughout the production horizon by following the schedule determined using the static process model. These include the impurity level, P_{zone1}/P_{dew} and the condenser/reboiler minimum temperature driving force, and the reboiler holdup endpoint constraint.

ICs: (7.19) and (7.20)

rQCs: (7.30) and (7.32)

rPCs: (7.35), (7.36), (7.38), (7.39), (7.34)

$$0.8 \leq F_p^{sp,n} / F_p^{nom} \leq 1.2$$

where $p(t)$ is the price forecast profile throughout the horizon $T_m = 72h$. The decision variables, $F_p^{sp,n}$, are continuous and are bounded between $\pm 20\%$ of the nominal production flow rate (F_p^{nom}). The setpoint can change hourly, *i.e.*, $N_e = 72$ and $\tau_n = 1h \quad \forall \quad n$.

7.6 Results and Discussion

In this section we present and compare the results of the optimal schedules determined using:

P1 Formulation (7.41), using a static process. This is referred to as problem P1.

P2 An additional problem formulation that only considers the rQCs.

$$\begin{aligned} & \underset{F_p^{sp,n}}{\text{minimize}} && \phi = \int_0^T p(t)(\gamma_c \hat{F}_p(t) - \gamma_{t1} \hat{F}_{feed} - \gamma_{t2} \hat{F}_p + \gamma_l \hat{F}_{inv}) dt && (7.50) \\ & \text{subject to} && \text{Time slots (7.2) – (7.3)} \\ & && \hat{y}_p \text{ models of the form (7.45)} \\ & && \hat{v}_p \text{ models of the form (7.45)} \\ & && \text{Inventory model: (7.18)} \\ & && \text{Split: (7.21) and (7.23)} \\ & && \text{Mixing: (7.22) and (7.24)} \\ & && \text{ICs: (7.19) – (7.20)} \end{aligned}$$

rQCs: (7.30) – (7.32)

$$0.8 \leq F_p^{sp,n} / F_p^{nom} \leq 1.2$$

We refer to this as problem P2. The solution of P2 (which is based on the general expression ((7.16))) will illustrate the effects of excluding process operating constraints from the dynamic scheduling problem formulation.

P3 Formulation (7.49), which includes a process model for the rQCs and rPCs. We refer to this as problem P3.

P4 A scheduling and dynamic optimization calculation including the full-order process model based on (7.10). The solution time of P4 will be compared to that of problems P2 and P3 using low-order dynamic models to evaluate the improved computational performance of our approach.

The characteristics of each of these formulations are summarized in Table 7.4. Note that the model of the air separation process used in P4 is highly coupled and nonlinear and involves 6094 equations and 430 state variables. In comparison, the P2, P3, and P1 have nearly two orders of magnitude fewer equations and states.

The problems were implemented and solved in gPROMS [217] using a sequential dynamic optimization solver. All calculations were performed on a 64 bit Windows system with Intel Core i7-2600 CPU at 3.40 GHz and 16 GB RAM. The convergence tolerances used for changes in the objective function and constraint violations were, respectively, 10^{-3} and 10^{-9} . Sequential solution approaches alternate between the time integration of the DAE model equations and appropriate

Table 7.4: Summary of problem formulations.

Formulation	Description	Problem Size
P1	Scheduling using steady-state correlations (7.41)	1 differential variable
P2	Scheduling with a dynamic process model and constraints on rQCs (7.50)	10 differential variables
P3	Scheduling with dynamic process model and constraints on rQCs and rPCs (7.49)	51 differential variables
P4	Scheduling with a detailed dynamic process model	430 differential variables, 6094 algebraic variables

sensitivities over the time horizon of the problem, and the solution of an (MI)NLP. Discontinuities such as those in equations (7.26)-(7.28), (7.46), and (7.48). Discontinuities are dealt with at the integrator level by i) identifying the time point where the discontinuity occurs and, ii) performing a reinitialization calculation at that point, which involves imposing state and sensitivity continuity, and finding new consistent values for the algebraic variables. The reader is referred to the papers by Vassiliadis [277, 278], and the books by Brennan, Campbell and Petzold [39], and Cellier and Koffman [49] for further details.

The solutions of P1-P4 are discussed in the following sections. First, the optimal production target sequence determined from formulations P1-P3 will be compared and analyzed, then comparisons of the predictions of the rQCs and rPCs given by the low-order model will be compared with the results given by simulation of the production schedule on the full-order model, and finally a comparison of the variable trajectories given by all three schedules implemented on the full-order model will be presented.

7.6.1 Optimal Sequence of Production Rate and Inventory Level Setpoints for Problems P1, P2, and P3

The schedule was optimized for a three-day time horizon using the price profile shown in Figure 7.7, which are historical hourly prices from the Energy Reliability Council of Texas [88] (ERCOT) in the month of September 2013. The initial guess for each production schedule was a constant production rate equal to the product demand rate (because the models are nonlinear and nonconvex, it is possible that the solutions are only locally optimal). The optimal operating costs at the end of the horizon are shown in Table 7.5 (both the cost predictions using the low-order process models and the actual costs computed using the full-order model are presented).

Intuitively, as more constraints are introduced, the expected operating cost increases. The optimized schedules result in significant cost savings (about 2.7% over the 3 day horizon) in comparison to a production scenario where the production rate is constant and no storage is utilized (the energy prices reflect moderate-weather days in September in Texas). Intuitively, days with higher temperatures will have higher price variability owing, *e.g.*, to increased energy consumption by air conditioning systems [255, 198] and are expected to result in even higher electricity cost savings compared to an operating scenario based on a constant production rate.

The optimal production setpoint schedules determined by P1, P2, and P3 are shown in Figure 7.9. The solution of P1 results in the most aggressive schedule, with the production setpoint rate of change constraint (7.42) active during sig-

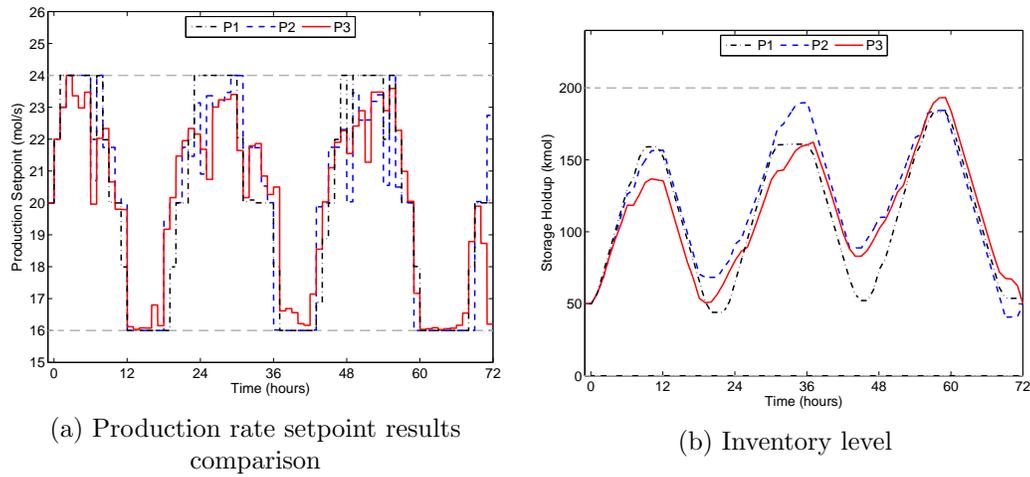


Figure 7.9: Optimal production schedule and inventory holdup prediction from P1, P2, and P3.

nificant production rate changes. The optimal schedules determined by solving problems P2 and P3 are more conservative in the approach and departure from the maximum production level. This is mainly owing to the presence rQCs, and specifically to the predicted value of I_p (see Figure 7.10) being at or near its bound (recall $I^{max} = 1800\text{ppm}$). In the case of P3, several rPCs (ΔT_{reb} , P_{zone1}/P_{dew} , $\delta_{max}^{flooding}$, and the final reboiler holdup) are also near their bounds, which results in a more conservative (but feasible) production setpoint schedule. The predicted trajectories for these variables are shown in Figure 7.11.

The importance of the storage unit can be seen in Figure 7.9b. Stored product is used to satisfy demand when prices are at their highest. Also, the three-day horizon increases the flexibility of the schedule and provides a buffer for days when energy prices are at their highest (e.g., in day 3).

Remark 7.9. Note that it is possible to solve the dynamic optimization problems

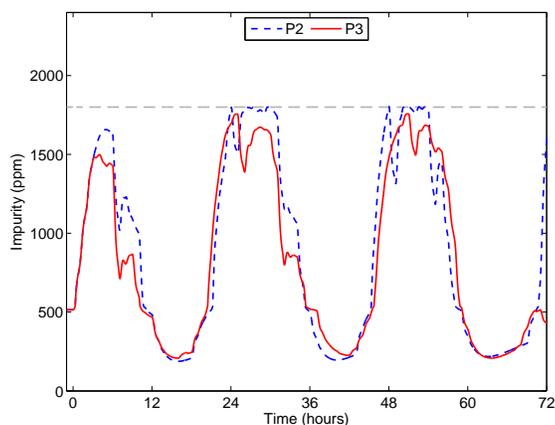


Figure 7.10: Evolution of impurity levels predicted by solving P2 and P3. Notice that a 'back-off' upper bound constraint of 1800ppm is imposed.

associated with our integrated scheduling/dynamic formulation using a simultaneous approach, which entails discretization of the time domain to convert a DAE system to a large system of algebraic equations. The resulting algebraic model can be solved as a (mixed integer) nonlinear program. Employing a simultaneous approach in the present case study (especially for P4) would require overcoming several significant hurdles. First, the dimension of the state-space of the models, and the need to capture multiple time scale dynamics over long time horizons would result in a nonlinear program of extremely large scale. Second, the challenge of finding a feasible initial guess for such models (an issue that has been time and again been emphasized in the literature [34]) is amplified by specific and severe nonlinearities associated with, e.g., phase transitions and the computation of phase equilibria. Motivated by the above, we opted for a sequential solution strategy [276] and, for consistency in evaluating the results and computational effort, we chose to use the same strategy for all the dynamic optimization problems considered.

7.6.2 Implementation on Full-Order Model

The optimal schedules illustrated in Figure 7.9a were implemented and simulated in the detailed dynamic model to evaluate the efficacy of the proposed low-order model scheduling method. The control heuristic discussed in section 7.5.1 was used to set the trajectories of the manipulated variables. The implementation of P1 (the static schedule) on the detailed model is shown in Figure 7.8. A comparison of the low-order model predictions and the trajectories of the corresponding variables in the detailed process model during the optimal schedule determined in P3 is shown in Figure 7.11.

As seen in Figure 7.11c, the model for impurity levels under-predicts on several occasions, but the actual constraint bound of 2000ppm is never violated. The low-order model of the condenser/reboiler temperature driving force (Figure 7.11d) provides a very accurate prediction. The corresponding constraint is near its bound often throughout the time horizon, which limits the production rate from reaching the maximum level. Additionally, the liquid level in the reboiler is completely replenished before the end of the horizon.

The optimal schedules determined with formulations P2 and P3 were implemented on the detailed ASU model to evaluate the effect of considering the rPCs in the scheduling formulation. The rQCs are plotted in Figure 7.12 where Figure 7.12a shows the product impurity throughout the schedules in both P2 and P3, and Figure 7.12b shows the fraction of the production that is sent to storage throughout each schedule. The impurity level in P2 violates the constraint bound by 200 ppm for several hours on days 2 and 3 (due to the underprediction by the

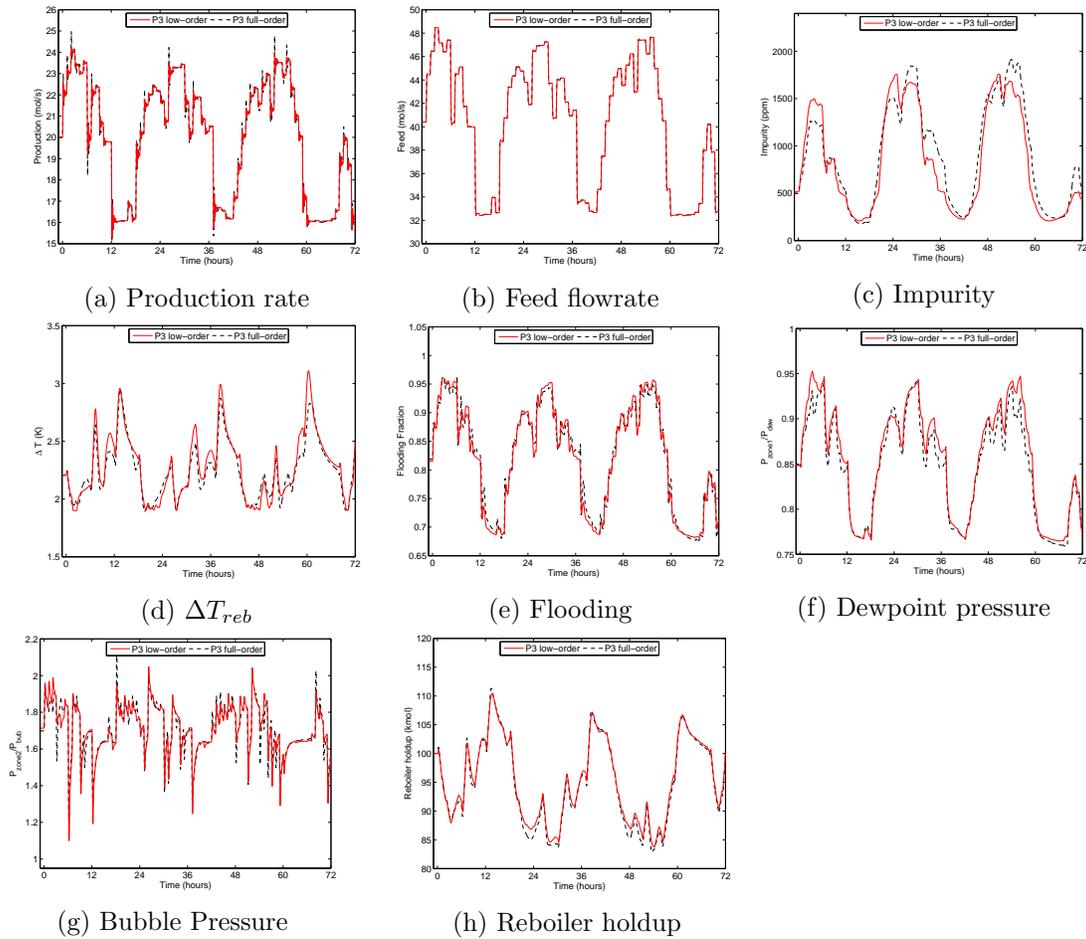


Figure 7.11: A comparison of the low-order model predictions (labeled P3 low-order) and the trajectories of the corresponding variables in the full-order process model during implementation of the optimal schedule determined using P3 (labeled P3 full-order)

low-order model).

The rPCs are plotted in Figure 7.13. Recall that here the schedule in P2 is optimized without considering the rPCs. The bubble point constraint in zone 2 of the PHX does not approach the constraint bound in either P2 or P3. Constraints

Table 7.5: Optimal solutions of problems P1-P4. The cost values are for the 3 day horizon.

Formulation	Predicted optimal cost (using low-order model) (\$)	Actual electricity cost (using full-order model) (\$)	Savings compared to constant production rate	Constraint violations when implemented	Solution times
P1	21466	21470	3.2%	Yes	0.33h
P2	21500	21504	3.1%	Yes	0.83h
P3	21584	21584	2.7%	No	1.2h
P4	21520	21520	3.0%	No	97h
Constant production rate	22187	22187	–	–	–

ΔT_{min}^{reb} and P_{zone1}/P_{dew} along with the flooding constraint are violated on several occasions indicating that the schedule is not physically realizable by the process. Additionally, the reboiler holdup is depleted by about 10% at the end of the P2 horizon indicating that the process is effectively losing refrigeration over time. This justifies the need to include rPCs in the scheduling problem formulation.

7.6.3 Comparison of P3 and P4

Finally, we compare the result of P3 to P4, *i.e.*, the optimization of the schedule using the low order model and full-order dynamic process model.

$$\begin{aligned}
 &\underset{F_p^{sp,n}}{\text{minimize}} && \phi = \int_0^T p(t)(\gamma_c F_p(t) - \gamma_{t1} F_{feed} - \gamma_{t2} F_p + \gamma_t F_{inv}^{in}) dt && (7.51) \\
 &\text{subject to} && \text{Time slots (7.2) – (7.3)} \\
 &&& \text{Detailed process model (7.7)}
 \end{aligned}$$

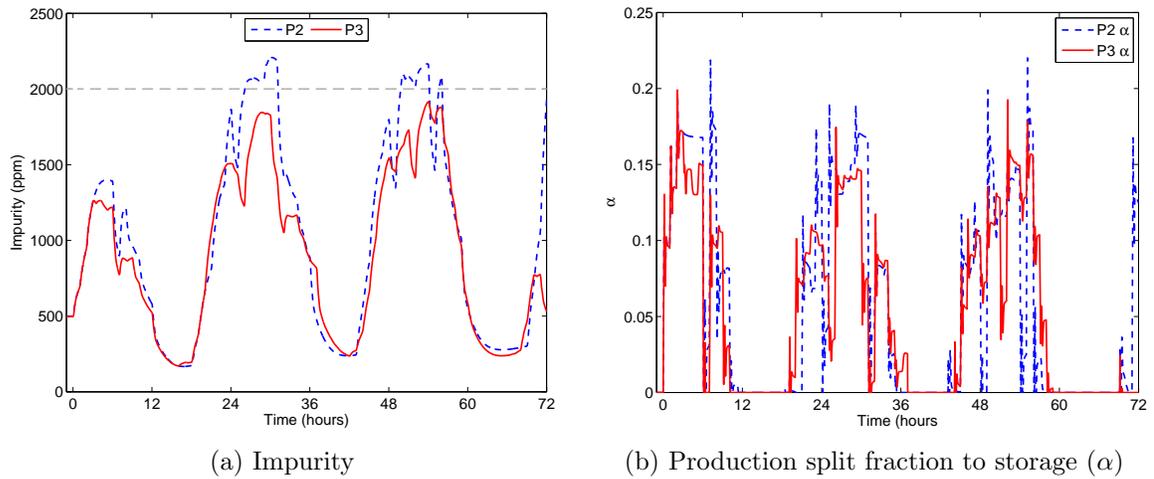


Figure 7.12: Comparison of rQCs when optimal schedules produced by P2 and P3 are simulated on the detailed model

Inventory model: (7.18)

Split: (7.21) and (7.23)

Mixing: (7.22) and (7.24)

ICs: (7.19) – (7.20)

QCs: (7.30) – (7.32)

PCs: (7.35), (7.36), (7.38), (7.39), (7.34)

$$0.8 \leq F_p^{sp,n} / F_p^{nom} \leq 1.2$$

The optimal production setpoint schedules and inventory holdup are compared in Figures 7.14a and 7.14b. The schedules are very similar, with the discrepancies stemming from the under-prediction of the impurity concentration in P3 (Figure 7.14c), and the lack of backoff constraint used in P4 since the model assumes perfect knowledge of the system in this situation. As a result, the cost savings

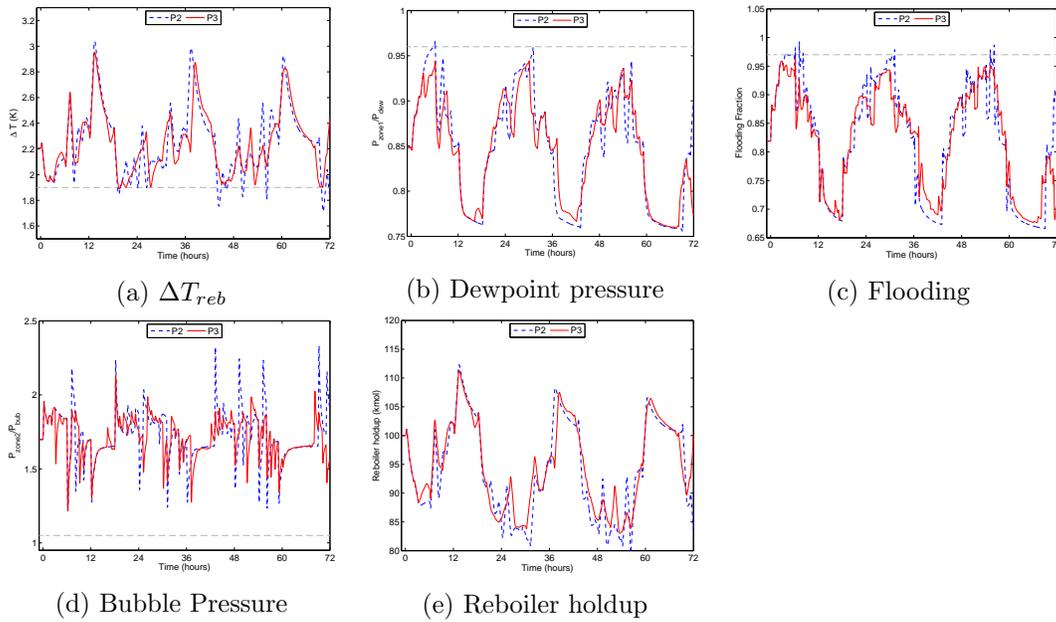


Figure 7.13: rPCs comparison when optimal schedules produced by P2 and P3 are simulated on the detailed model

associated with P3 (2.7%) is about 90% of the savings achieved using the detailed process model and P4 (3.0%). This similarity in overall savings combined with a feasible production schedule validates the quality of our low-order models. The important difference between the solutions of P3 and P4 is in the computation time; P3 required 1.2 hours to solve, while the solution of P4 required 97 hours (see Table 7.5). This reduction in computation time demonstrates the potential for executing P3 in a periodic or rolling-horizon fashion, and quickly updating the schedule in response to identified disturbances.

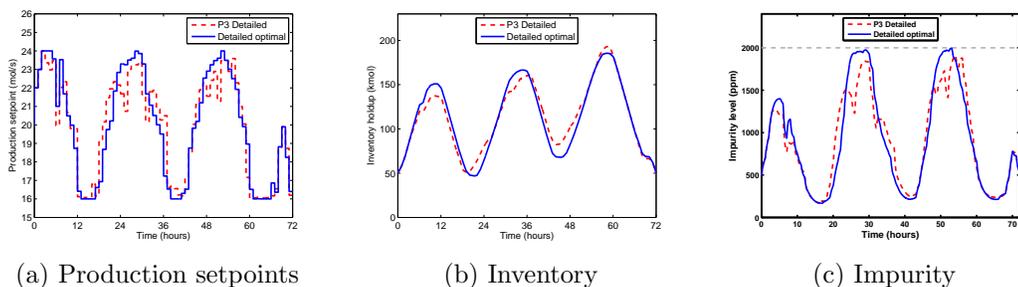


Figure 7.14: Results comparison between optimal schedules determined using the low-order process model (P3) and the detailed process model

7.7 Conclusion

In this work, we proposed a novel framework for optimal scheduling of continuous processes when subjected to high market variability, using scheduling-oriented low-order dynamic process models identified from historical process transition data. The scope of these models encompasses product quality, production rates, and variables relevant to product and process constraints that are near their limits during transitions between production targets. Low-order dynamic models of these variables are identified from historical process transition data. However, our framework is generic and can accommodate low/reduced-order representations of detailed process models if available. The low-order models predict the dynamic response of the process inputs, outputs and operating constraints to changes in the production setpoints.

The scheduling-oriented low-order models are then used in an optimal scheduling problem aimed at minimizing operating costs (or maximizing profit) by adjusting the product setpoints and utilizing the available inventory. This renders

the scheduling problem aware of the process dynamics and operating constraints without the need for a detailed process model.

Our framework represents a significant departure from the traditional scheduling approaches, which use static process models with tabulated transition information between discrete operating levels to determine the optimal production schedules. It allows for continuous (rather than discrete) changes to the production targets and ensures that the transitions are dynamically feasible. Additionally, the smaller dimension of the process models make this approach computationally efficient.

We applied the proposed optimal scheduling method to an air separation unit producing nitrogen. The ASU is outfitted with a nitrogen liquefier and a liquid nitrogen storage tank, which enable the process to adjust the production rate in response to variable electricity prices while satisfying product demand by re-gasifying the liquid nitrogen inventory. The optimal schedule results in a 2.7% savings in electricity cost over the course of 3 days in comparison to a scenario where the production rate is constant. The optimal production schedule was compared to the result produced using a static process model and a model which only accounts for the product (not process) constraints. The resulting schedules are dynamically feasible only in the case in which both product quality constraints and process operating constraints are included in the low-order dynamic model, emphasizing the importance of defining the scope of scheduling-oriented process models. Additionally, we show that the results determined using the low-order process model are very similar to the optimal schedule determined using a full-order dynamic process model, but are obtained in two orders of magnitude less time.

Chapter 8

Closing the Scheduling Loop: Rescheduling Formulations Incorporating Dynamic Process Models, Fault Detection-Based Triggers, and Moving Horizon Concepts

The second case study in this chapter is published in [209]. The majority of the material in this chapter is in preparation for publication.

8.1 Introduction

The objective of scheduling within the range of management decisions (shown in Figure 8.1) for multi-product processes is to establish the most profitable production sequence and inventory levels over the course of the scheduling horizon. Recently, cross-functional coordination between the layers of decision making in Figure 8.1 has been shown to improve the overall economics of chemical process systems [106, 275]. In particular, the *integration of scheduling and control* or *integration of scheduling and process dynamics* has been the subject of many recent reviews [24, 63, 112] and is a major ongoing research thrust. Research efforts aim to incorporate knowledge of process dynamics and control actions within scheduling calculations, typically by creating a dynamic model of the process (e.g., a set of differential equations or a data-driven discrete model) and linking the model to

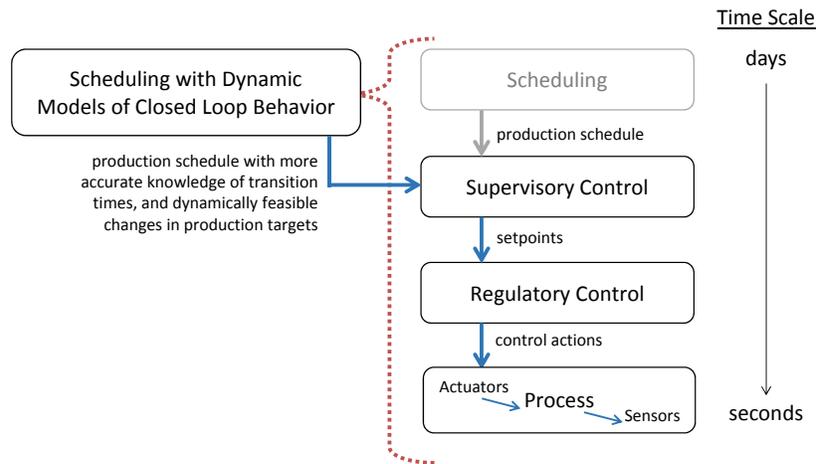


Figure 8.1: Process management decisions. Scheduling with a dynamic process model is an alternate method for obtaining a production schedule.

key scheduling decisions. Accounting for process dynamics explicitly in scheduling is a significant departure from traditional scheduling calculations (e.g., those reviewed in [174] for chemical processes), where process information is represented using tabulated performance data (e.g., transition tables) or heuristics. Scheduling with a dynamic process model (highlighted in Figure 8.1 as an alternate method for obtaining a production schedule) is desirable because it can provide more accurate knowledge of transition times, operating costs, and the feasibility of certain transitions.

Rescheduling is necessary when the current schedule cannot be executed as planned, usually due to an unforeseen event. Methods for scheduling under uncertainty [156, 69] are employed across a variety of sectors [279] to minimize the need

to reschedule by creating schedules that are ‘robust’ to differences in processing times and unforeseen events. However, rescheduling is unavoidable in the event of a major disturbance that impacts the ability to make a subset of products. Such ‘major disturbances’ typically have a direct impact on the scheduling calculation formulation (e.g., new/rush order arrival or machine break-down), and can therefore be viewed as disturbances to the scheduling layer in Figure 8.1. These have been extensively studied in the literature (the reviews [280, 231, 13, 199] provide a summary of rescheduling formulations and circumstances). There are several examples of rescheduling due to scheduling layer disturbances in the chemical process literature using reactive (i.e., event-driven, where the need to reschedule is ‘triggered’ by an adverse event) and periodic (i.e., the schedule is regenerated at a regular interval in a moving horizon fashion, using updated information on the performance of the schedule in the previous interval) techniques using static process information.

On the other hand, *disturbances to the process layer* can cause changes in the product quality and processing times, which may also necessitate rescheduling [280]. Addressing events that impact the process dynamics (e.g., disturbances affecting the operation of one unit within the system, without leading to an obvious failure) has received less attention, under the likely implicit assumption that their impact will be addressed by the local control system. Unlike disturbances to the scheduling layer, these are a result of the uncertainty present *within* a process [212, 143] and may be a result of drift in process parameters over time or sudden unmeasured changes in operating conditions. The development of integrated

scheduling and control provides novel opportunities for rescheduling in response to disturbances at the process level, as opposed to disturbances to the scheduling level. Recent works have shown that predictions of the state variables through surrogate models can be used to define and confirm the need to reschedule process operations [310, 62]. The general state of the literature regarding the role of dynamic models in event driven and periodic rescheduling is summarized these in Table 8.1. In general, the development of a unified approach for exploiting the wealth of information that can be derived from accounting for process dynamics at the scheduling stage, and from monitoring the closed-loop process performance during scheduling execution, remains an open question.

Table 8.1: Rescheduling Strategies

	with dynamic model		without dynamic model	
Disturbance type	Event-driven	Periodic	Event-driven	Periodic
Schedule layer	There is some literature here		accounted for in the traditional rescheduling literature	
Dynamic	This is the focus of our work		not possible to account for these explicitly using traditional rescheduling methods	

In this work, we will demonstrate novel rescheduling concepts that are possible when scheduling with a dynamic process model. In particular, we emphasize differences between disturbances to the scheduling and process layer, and highlight how the latter can lead to changes in processing times or recipes, transition times, and the overall product wheel. For this reason, the scheduling layer should be

aware of changes in the dynamic process behavior. We propose an event-driven rescheduling framework using existing methods for process monitoring and fault detection, identification, and reconstruction, which diagnose specific changes in processing conditions, as triggers. The integrated scheduling and dynamics framework is well suited for accounting for these differences; we propose that the new operating paradigm can be accounted for by altering constraint boundaries for the dynamic model in a rescheduling calculation. In this context, rescheduling effectively closes the loop at the scheduling level in response to the presence of faults at the process level. We also discuss a moving horizon scheduling formulation, which uses a state observer to track the evolution of model variables and reduce plant-model mismatch. A discussion on the differences between moving horizon scheduling and economic MPC is provided to highlight this contribution.

8.2 Rescheduling in Response to Dynamic Disturbances

8.2.1 Motivation

Figure 8.2 highlights the presence of three different types of disturbances, which can appear at the scheduling layer or process layer in the decision making hierarchy. Rescheduling in response to failures (e.g., machine breakdown or any event resulting in the inability to operate the process and possibly requiring downtime to perform maintenance) and schedule disturbances (e.g, rush orders, order cancellations, or due date changes) have been extensively studied within the operations research community. These disturbances are similar in that they provide an obvious rescheduling trigger, and are not related to the process dynamic.

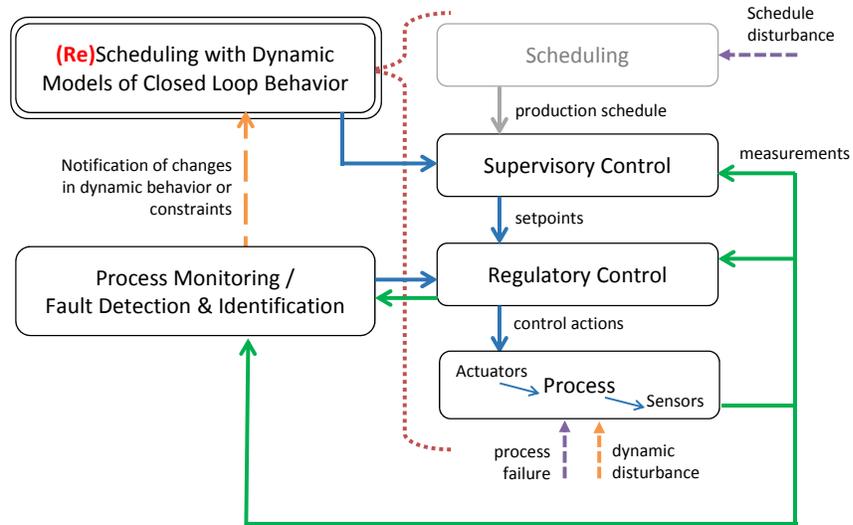


Figure 8.2: Closing the scheduling loop in response to dynamic disturbances using scheduling approaches with dynamic models and process monitoring.

In this paper we focus on dynamic disturbances at the process level in Figure 8.2. Dynamic disturbances are those that alter the plant operating conditions but do not lead to a failure. Examples of these events include valve stiction, controller saturation, changes in raw material properties, or drifts in process parameters and operating conditions. While these incidents do not completely stop the process, they will likely limit the operating range (i.e., tighten process constraints on inputs and/or states). The consequences of this are:

- (A) some products may no longer be feasible
- (B) some predicted processing times or transition times may change (they will likely be longer than expected)

(C) some transitions may no longer be feasible.

Notice that item (A) affects the product wheel, item (B) affects the makespan, and item (C) affects the sequence of the original schedule. Therefore, dynamic incidents at the control/process level can have serious consequences at the scheduling level. Also, evaluating each of these faults requires dynamic information (for a continuous process, steady state information may also be useful). Even though changes in processing times are frequently counted amongst reasons for rescheduling in the literature, methods tailored specifically for rescheduling in response to dynamic disturbances have received less direct attention. This is due to the fact that (i) in a traditional static or recipe based scheduling calculation, it is not possible predict how dynamic disturbances impact product recipes or feasibility. Such information would need to be established offline. Also, (ii) scheduling under uncertainty can be used to increase the robustness of schedules to dynamic disturbances and avoid rescheduling. Finally, (iii) it is often assumed that these disturbances at the process level will be mitigated by the control system so that the potential dynamic differences when executing the schedule are negligible. Note that the last assumption may not always hold in the presence of actuator faults, thereby providing further support for incorporating dynamic models of closed-loop process behavior into scheduling calculations.

Given the potential for dynamic disturbances to impact the outcomes of implementing a predefined schedule, rescheduling can be useful to update the expected processing times as the operating environment changes. We will leverage

the availability of scheduling formulations incorporating dynamic models to provide a unified approach for rescheduling, which can account for such changes in the operating conditions. A crucial component of this framework is a feedback link between changes at the process level and the scheduling level, which (i) identifies dynamic disturbances that may impact the predictions from the dynamic model in the scheduling calculation, and (ii) accounts for these disturbances in future scheduling efforts. The method behind each of these tasks is summarized in the following sections.

8.2.2 Framework

To begin, we assume a scheduling system is in place which follows the integrated scheduling and control paradigm (i.e., the scheduling calculation includes a model of the closed loop process dynamics [210]). The general scheduling formulation is given in Table 8.2.

The problem is a dynamic optimization (DO) because of the presence of a dynamic process model, and will be a mixed-integer DO in the most general case. The type of process model used in this environment has a significant impact on the quality and feasibility of the resulting schedule, the difficulty of solving the associated optimization problem and choice of solution algorithm, and the opportunities for rescheduling.

Regarding the form of the process model, this may be a detailed first principles model or a low-order model of scheduling relevant process variables (also referred to as a scale bridging model - SBM [76]). For complex processes using

Table 8.2: General Scheduling with Dynamic Model Problem Statement

Given:

Timing Information

- Time horizon
- Event timing and schedule characteristics

Material Information

- Product requirements
- Demand quantities or rates
- Prices of raw materials, utilities, and products
- Inventory restrictions

Dynamic Process Knowledge

- Process model
- Operating constraints

Establish:

Optimal production sequence/schedule that maximizes the profit or minimizes the operating costs of the process

optimization-based controllers, it is not desirable to use a detailed process model in the scheduling calculation because of the problem size, nonlinearity, and presence of nested optimization formulations. For this reason, SBMs are identified from data for use in scheduling calculations. In this work, we assume that the SBM of the general form

$$\mathbf{0} = \mathbf{f}(\mathbf{y}_{sp}, \mathbf{w}, \dot{\mathbf{w}}, \ddot{\mathbf{w}}, \dots, \mathbf{p}, t) \quad (8.1)$$

is used in the formulation in Table 8.2, along with the operating constraints

$$\mathbf{w}^{min} \leq \mathbf{w} \leq \mathbf{w}^{max} \quad (8.2)$$

$$\mathbf{g}(\mathbf{w}, \dot{\mathbf{w}}, t) \geq \mathbf{0} \quad (8.3)$$

where \mathbf{y}_{sp} represents the scheduling decision variables for product sequence, and \mathbf{w} is the vector of scheduling relevant variables that are modeled using the SBM

(8.1). The variables \mathbf{w} are subject to the bounds (8.2) and nonlinear constraints (8.3), which may include rate of change limits on the derivatives $\dot{\mathbf{w}}$. For a detailed discussion on the derivation of such SBMs and the selection of scheduling relevant variables, we refer the reader to [210].

8.2.2.1 Process Monitoring Provides Rescheduling Triggers

It is intuitive that dynamic disturbances which necessitate rescheduling are those which impact the model or constraints associated with the scheduling relevant variables \mathbf{w} , and any sensor or actuator faults that impact the variable \mathbf{w} . Such disturbances are expected to cause changes to the behavior of the process as predicted by the formulation in Table 8.2, and should be accounted for at the scheduling layer. The remaining question is ‘how will the dynamic disturbances be identified?’ We believe that existing methods for process monitoring [298], fault detection/identification and diagnosis [170, 220], fault reconstruction [79, 219], and state/disturbance observers [223, 145] can provide information relevant to rescheduling in response to dynamic disturbances.

In fact, these methods are likely already in place at many plant sites. Quantifying changes in process characteristics and operating conditions (which we have called dynamic disturbances) have been identified as a top priority for maintaining the accuracy of virtual sensors [133] and is important for controller performance assessment [287]. There is a clear industrial motivation for identifying dynamic disturbances and relaying this information to several levels of the decision making hierarchy. Ultimately, closing the loop at the scheduling level may just be a mat-

ter of relaying the information from process monitoring efforts to the rescheduling layer, which must be flexible enough to account for such disruptions.

The link between process monitoring/fault diagnosis and rescheduling effectively provides feedback about changes within the process level to the scheduling level. This is similar to the feedback provided by measurements to the regulatory and supervisory control layers in Figure 8.2, which plays a crucial role in ensuring optimal performance of the control system. We believe similar improvements can be made in scheduling through the act of rescheduling in response to dynamic disturbances.

Consider that in the literature on scheduling incorporating models of process dynamics there are case studies demonstrating its application to air separation unit [210], multi-product CSTRs [257, 310] and polymerization reactors [214, 76, 250], catalytic cracking [126], and multi-unit batch processes [59, 192]. For each of these applications, there are parallel case studies in the process monitoring literature which seek to identify faults or changes in conditions that are relevant to the models the corresponding scheduling formulation.

For example, the use of fault reconstruction to identify the magnitude of a leak in a reboiler [79] is of interest to the model of reboiler holdup in [210] In [244], an observer is used to estimate changes in the reaction kinetic parameters and also to detect a disturbance in the feed temperature for a styrene polymerization reactor. Statistical PM methods and fault reconstruction are used to identify changes in feed concentration and temperature for a CSTR in [153]. These are all dynamic disturbances that are relevant to the multi-product reactors in [257, 310,

214, 76, 250]. [244] and [208] demonstrate, respectively, the use of an observer or statistical fault detection method to address actuator faults or changes in physical properties in a fluid catalytic cracking unit. Fault detection for batch systems is also a well-established research area, and there are numerous case studies in the literature which demonstrate methods for detecting sensor and actuator faults for batch systems.

These examples serve to justify our claims that existing methods for process monitoring can detect information relevant to the models and constraints in the scheduling formulation, and that the link between these components is plausible and warrants further study.

8.2.2.2 Formulation

The last piece of the framework shown in Figure 8.2 is the formulation of the rescheduling calculation. Once the process monitoring/fault detection step has identified a dynamic disturbance and communicated it to the scheduling layer, how will this information be incorporated in the optimization? The presence of a process model provides the ability to *explicitly* account for changes in constraints or parameters which may alter processing/transition times and the feasibility of operating states. Thus, it is possible to update the model parameters in (8.1) and constraints (8.2) and (8.3) on affected variables with new values as identified through process monitoring and fault detection. This is shown below, where $\hat{\mathbf{p}}$, $\hat{\mathbf{w}}$, $\hat{\mathbf{g}}$ denote updated model components.

$$\mathbf{0} = \mathbf{f}(\mathbf{y}_{sp}, \mathbf{w}, \dot{\mathbf{w}}, \ddot{\mathbf{w}}, \dots, \hat{\mathbf{p}}, t) \quad (8.4)$$

$$\hat{\mathbf{w}}^{min} \leq \mathbf{w} \leq \hat{\mathbf{w}}^{max} \quad (8.5)$$

$$\hat{\mathbf{g}}(\mathbf{w}, \dot{\mathbf{w}}, t) \geq \mathbf{0} \quad (8.6)$$

With these model updates, it is possible to easily evaluate the feasibility of the remainder of the schedule and perform rescheduling (e.g., a right shift, partial, or complete reschedule). This is a major benefit of scheduling with a dynamic model compared to static scheduling approaches, where it is not straightforward to use heuristics to predict how a variety of dynamic disturbances impact product recipes or feasibility (such information would need to be established offline).

Depending on the complexity of the process, it may be beneficial to evaluate the feasibility of the product wheel in the new operating conditions before resolving the scheduling problem. For example, for a continuous process this involves identifying the new stable state-state operating space and ensuring that all products are still within this region. For a large multi-unit batch process, a simplified dynamic optimization for the affected unit may be evaluated.

8.2.2.3 Opportunities for Model Based Process Monitoring

The connection between process modeling and rescheduling does not need to be a one-way link. As shown in Figure 8.3, it is possible that the predictions of variable trajectories from a scheduling calculation with a dynamic model can be of use for model-based process monitoring. Although, the value of this information may be at odds with the desire to reduce scheduling complexity with SBM forms. Consider that if a detailed process model is incorporated into scheduling, this

would provide a large quantity of information for PM, whereas a SBM containing only a few scheduling-relevant variables provides limited information.

Regardless, it is important to recognize that the models used in scheduling with dynamics should be synchronized with those used for (model-based) process control and process monitoring. Communication of changes made to any of the models due to dynamic disturbances is needed to ensure all decision making is done with up-to-date information and minimal error.

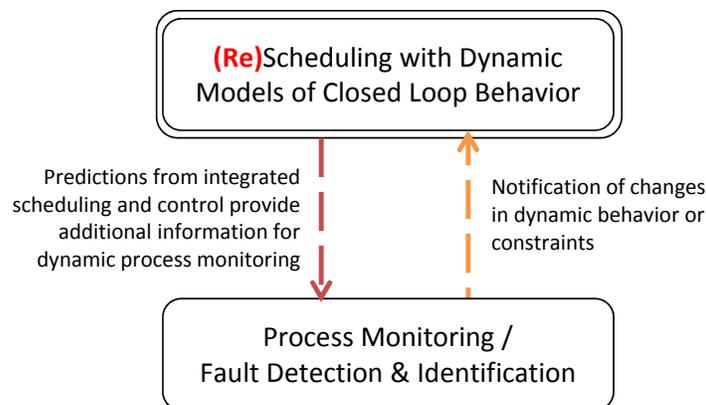


Figure 8.3: Opportunities for providing additional information to model based fault detection schemes

8.2.3 Summary

To summarize, the framework that we are utilizing to close the scheduling loop consists of:

1. Process monitoring to quantify the change in operating conditions and con-

straints due to dynamic disturbances. Some of the methods discussed in section 8.2.2.1 are assumed to be already in place and capable of quantifying dynamic disturbances which impact the scheduling relevant variables.

2. A feasibility analysis on the dynamic process model, to evaluate the effect of the change in operating conditions. As discussed in section 8.2.2.2, this step may be useful for complex processes.
3. Updating model parameters and constraints to reflect the changes at the process layer, and performing a rescheduling calculation to update the schedule.

This is, to our knowledge, the first proposed use of process monitoring for updating dynamic models incorporated into scheduling calculations. Because we are relying on existing process monitoring techniques, demonstrating the fault detection and reconstruction steps are not the focus of this paper. In the case studies in the following section we aim to highlight the impact of dynamic disturbances and demonstrate how, once identified through process monitoring, they are easily accounted for through rescheduling with a dynamic process model. Our focus will employ schedule repair methods (e.g., right shifts or partial rescheduling), which make minimal changes to the original schedule to avoid nervousness. These developments effectively close the scheduling loop through initiating rescheduling calculations under the impetus of changes in external factors that affect the operation of the process.

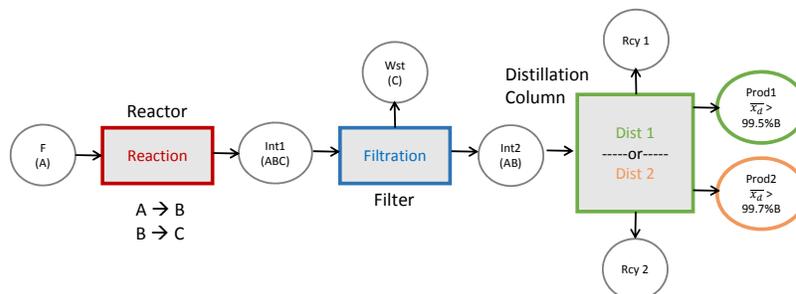


Figure 8.4: State equipment network for the batch flowshop process.

8.2.4 Case Study

We present a case study for rescheduling in a multi-unit sequential (flowshop) batch process based on the example originally proposed in [192]. The following equations and nomenclature are developed and presented in detail in [192], where the benefits of scheduling with a dynamic model compared to a fixed recipe are shown by evaluating the operating costs and profit of the two schedules. In this work, we provide an overview of the problem formulation and use this scenario to demonstrate our rescheduling framework.

The state equipment network (SEN) is shown in Figure 8.4. The reactions $A \rightarrow B$ and $B \rightarrow C$ take place in the reactor, which receives a feed of pure A . The first intermediate stream (with components A , B , and C) is sent through a filtration step, which is assumed to completely remove component C . The second intermediate (containing only B and C) is fed to a batch distillation column. The column can be run in two distinct states to produce product 1 (the total distillate has $\geq 99.5\%B$) or product 2 (the total distillate has $\geq 99.7\%B$). A

slot based scheduling formulation with N time slots is developed incorporating dynamic models of the reactor and distillation column, while the filtration step is modeled using algebraic equations.

Table 8.3: Nomenclature - Indices and Sets

Indices/Sets	Description
$j \in J$	units
$r \in R$	materials
$s \in S$	operating states
$n \in N$	event slots
$c \in C$	chemical components
$R_{prod} \subseteq R$	final products
$R_{raw} \subseteq R$	raw materials
$J_s \subseteq J$	units with operating state s
$J_r^p \subseteq J$	units that produce material r
$J_r^c \subseteq J$	units that consume material r
$R_s^p \subseteq R$	materials produced in state s
$R_s^c \subseteq R$	materials consumed in state s
$R_j^p \subseteq R$	materials produced in unit j
$R_j^c \subseteq R$	materials consumed in unit j
$S_r^p \subseteq S$	states that produce material r
$S_r^c \subseteq S$	states that consume material r
$S_j \subseteq S$	states that are possible in unit j
$C_r \subseteq C$	components of material r

8.2.4.1 Scheduling Problem Formulation

The MIDO for scheduling is presented in equations (8.7) - (8.36). The nomenclature is explained in Tables 8.3 and 8.4. The optimization problem aims to maximize profit, and the scheduling constraints describe a manufacturing scenario with no changeover delays and no fulfillment deadlines, and uses the unlimited intermediate storage policy. For details regarding their construction, we refer the

Table 8.4: Nomenclature - Parameters and Variables

Parameters	Description
\mathcal{N}	cardinality of set N
P_r	price of material r
H	scheduling horizon
B_j^{min}, B_j^{max}	bounds on batch size
E_r^{min}, E_r^{max}	bounds on excess material
\bar{E}_r	initial amount
$\eta_{r,c,1}$	initial composition
$\bar{\phi}_{j,r,c}, \bar{\eta}_{r,c}$	quality requirements
Variables	Description
$w_{j,s,n}$	binary variable indicating state of unit j in slot n
$F_{j,n}$	operating cost
$E_{r,n}$	excess material
$E_{r,0}$	initial excess material
$R_{j,r,n}^p$	production of r by j in slot n
$R_{j,r,n}^c$	consumption of r by j in slot n
$b_{j,n}$	batch size
$T_{j,n}$	slot start time
$Tp_{j,n}$	processing duration
$z_{j,n}$	differential variables
$y_{j,n}$	algebraic variables
τ	normalized time
$\eta_{r,c,n}$	composition of material r
$\phi_{j,r,c,n}$	composition of material r produced by j

reader to the original paper describing the case study [192].

$$\max_{b_{j,n}, w_{s,j,n}, u_{j,n}(\tau)} \text{Profit} = \text{Sales} - \text{Cost}_{\text{raw materials}} - \text{Cost}_{\text{unit operations}} \quad (8.7)$$

$$\text{Sales} = \sum_{r \in R_{\text{prod}}} P_r (E_{r,\mathcal{N}} + \sum_{j \in J_r^p} R_{j,r,\mathcal{N}}^p) \quad (8.8)$$

$$\text{Cost}_{\text{raw materials}} = \sum_{r \in R_{\text{raw}}} Pr E_{r,0} \quad (8.9)$$

$$\text{Cost}_{\text{unit operations}} = \sum_{j \in J, n \in N} F_{j,n} \quad (8.10)$$

Subject to:

Assignment constraints

$$\sum_{w \in S_j} w_{j,s,n} \leq 1 \quad \forall j \in J, n \in N \quad (8.11)$$

Material balances

$$E_{r,n} = E_{r,n+1} + \sum_{j \in J_r^p} R_{j,r,n-1}^p - \sum_{j \in J_r^c} R_{j,r,n}^c \quad \forall r \in R, n \in N, n > 1 \quad (8.12)$$

$$E_{r,1} = E_{r,0} - \sum_{j \in J_r^c} R_{j,r,1}^c \quad \forall r \in R \quad (8.13)$$

$$E_{r,0} = \bar{E}_r \quad \forall r \in R' \subset R \quad (8.14)$$

$$\sum_{r \in R_j^p} R_{j,r,n}^p = \sum_{r \in R_j^c} R_{j,r,n}^c \quad \forall j \in J, n \in N \quad (8.15)$$

Capacity constraints

$$\sum_{s \in S_j} w_{j,s,n} B_j^{\min} \leq b_{j,n} \leq \sum_{s \in S_j} w_{j,s,n} B_j^{\max} \quad \forall j \in J, n \in N \quad (8.16)$$

$$E_r^{min} \leq E_r^n \leq E_r^{max} \quad \forall r \in R, \forall n \in N \quad (8.17)$$

Timing constraints

$$T_{j,n+1} \geq T_{j,n} + Tp_{j,n} \quad \forall j \in J, n \in N, n < N \quad (8.18)$$

$$T_{j',n'} \geq T_{j,n} + Tp_{j,n} - H \left(2 - \sum_{s \in S_j \cap S_r^p} w_{j,s,n} - \sum_{s' \in S_{j'} \cap S_r^c} w_{j',s',n'} \right) \\ \forall r \in R, j \in J_r^p, j' \in J_r^c, j \neq j', n \in N, n' \in N, n < n' \leq N \quad (8.19)$$

$$T_{j,n} \leq H \quad \forall j \in J, n \in N \quad (8.20)$$

$$T_{j,N} + Tp_{j,N} \leq H \quad \forall j \in J \quad (8.21)$$

$$Tp_{j,s}^{min} \leq Tp_{j,n} \leq Tp_{j,s}^{max} \quad \forall j \in J, s \in S_j, n \in N \quad (8.22)$$

$$\sum_{n \in N} Tp_{j,n} \leq H \quad \forall j \in J \quad (8.23)$$

Unit models

$$\frac{dz_{j,n}(\tau)}{d\tau} = f_{j,s}(z_{j,n}(\tau), y_{j,n}(\tau), u_{j,n}(\tau)) Tp_{j,n} \quad (8.24)$$

$$0 = g_{j,s}(z_{j,n}(\tau), y_{j,n}(\tau), u_{j,n}(\tau)) \quad (8.25)$$

$$z_{j,n}(0) = Z_{j,s}(b_{j,n}, \eta_{r,c,n}) \quad (8.26)$$

Unit operating constraints

$$z_{j,s}^{min} \leq z_{j,n}(\tau) \leq z_{j,s}^{max} \quad (8.27)$$

$$y_{j,s}^{min} \leq y_{j,n}(\tau) \leq y_{j,s}^{max} \quad (8.28)$$

$$u_{j,s}^{min} \leq u_{j,n}(\tau) \leq u_{j,s}^{max} \quad (8.29)$$

Unit operating cost

$$F_{j,n} = h_{j,s}(u_{j,n}(\tau), Tp_{j,n}, b_{j,n}) \quad (8.30)$$

Material quality and constraints

$$\eta_{r,c,n} = \frac{\eta_{r,c,n-1}E_{r,n-1} + \sum_{j \in J_r^p} \phi_{j,r,c,n-1}R_{j,r,n-1}^p}{E_{r,n-1} + \sum_{j \in J_r^p} R_{j,r,n-1}^p} \quad (8.31)$$

$$\forall r \in R, c \in C_r, n \in N, n > 1$$

$$f(\eta_{r,c,n}, \eta_{r,c}^-) \geq 0 \quad \forall r \in R, c \in C_r, n \in N \quad (8.32)$$

Link equations between unit models and schedule variables

$$\phi_{j,r,c,n} = \Phi(z_{j,n}(\tau), y_{j,n}(\tau)) \quad (8.33)$$

$$R_{j,r,n}^p = R_{j,s}^p(z_{j,n}(1), z_{j,n}(1)) \quad (8.34)$$

$$R_{j,r,n}^c = R_{j,s}^c(z_{j,n}(0), z_{j,n}(0)) \quad (8.35)$$

Disjunctions

$$\left[\begin{array}{c} w_{j,s,n} = 1 \\ \text{Unit models} \\ \text{Unit operating constraints} \\ \text{Linking equations} \\ \text{Unit operating cost} \end{array} \right]_{s \in S_j} \bigvee \left[\begin{array}{c} \sum_{s \in S_j} w_{j,s,n} = 0 \\ R_{j,r,n}^p = 0 \\ R_{j,r,n}^c = 0 \\ F_{j,n} = 0 \\ Tp_{j,n} = 0 \end{array} \right] \quad \forall j \in J, n \in N, \tau \in [0, 1] \quad (8.36)$$

In this formulation, the manipulated variables are part of the decision variable set. Therefore, this is not a true closed loop representation of the process dynamics. However, under the assumption that the controller dynamics are very fast the open loop model is a reasonable approximation of the closed loop performance.

Note that one of the main benefits to incorporating dynamic unit models into a batch scheduling calculation is the ability to alter the properties of intermediate streams and unit recipes simultaneously with the schedule optimization. Considering the entire system operation can result in improved economic performance compared to optimizing units individually and using fixed recipes when scheduling. This is demonstrated in Figure 8.5, which shows two trajectories (A and B) for the evolution of a quality variable for a sequential batch process made up of two units and a schedule with two event slots. The duration of the event slots for each trajectory and the intermediate value of the quality variable are different, but at the end of the schedule the quality variable is the same. One trajectory may be favored over the other because it has a lower operating cost. Furthermore, it is possible to adjust these recipes online as part of rescheduling when a dynamic model is present. If the initial conditions at the start of the schedule changes, or if there is a disturbance during the first event slot, it is possible to update the schedule *and* recipe such that the target quality is still achieved. These capabilities will be demonstrated in the case study.

8.2.4.2 System Model

The reactor concentration model is shown in (8.37). Here, v is a scaled temperature and $\beta = 3.875 \times 10^{-3}$ is a scaled kinetic coefficient. v is the manipulated variable for controlling the reactor, and the initial batch size $b_{react,n}$ is a decision variable.

$$\frac{dC_A}{dt} = -vC_A^2 \quad (8.37a)$$

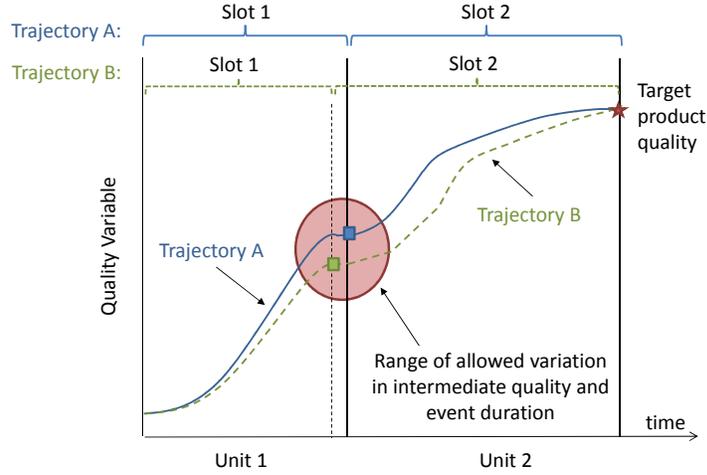


Figure 8.5: Variations in batch operations for making the same product

$$\frac{dC_B}{dt} = vC_A^2 - \beta v^2 C_B \quad (8.37b)$$

$$\frac{dC_C}{dt} = \beta v^2 C_B \quad (8.37c)$$

The binary distillation model includes the differential equations for the bottoms amount S (8.38a) and B composition x_b (8.38b), constant vapor flow rate V (8.38c), mass balance at the top of the column (8.38d) - (8.38e), tray composition balances (8.38f) - (8.38h), and phase equilibrium (8.38i) - (8.38j). The product concentration \bar{x}_d is described by (8.38k). Notice that the initial conditions depend on the concentration and amount of the Int2 stream. The reflux ratio R is the only manipulated variable for the column, which has $N = 4$ trays.

$$\frac{dS}{dt} = L - V, S(0) = b_{dist,n} = R_{Int2}^{prod} \quad (8.38a)$$

$$\frac{dx_b}{dt} = \frac{V(x_b - x_d)}{S(R + 1)}, x_b(0) = C_B^{Int2} \quad (8.38b)$$

$$V_n = kb_{dist,n} \quad (8.38c)$$

$$R = L/D \quad (8.38d)$$

$$V = L + D \quad (8.38e)$$

$$Lx_1 + Vy_1 = Lx_2 + Vy_b \quad (8.38f)$$

$$Lx_i + Vy_i = Lx_{i+1} + Vy_{i-1} \quad \forall i = 2 \dots N \quad (8.38g)$$

$$Lx_N + Vy_d = Lx_D + Vy_{N-1} \quad (8.38h)$$

$$y_i = \frac{\alpha x_i}{1 + (\alpha - 1)x_i} \quad \forall i = 1 \dots N \quad (8.38i)$$

$$y_b = \frac{\alpha x_b}{1 + (\alpha - 1)x_b} \quad (8.38j)$$

$$\bar{x}_d = \frac{S(0)x_b(0) - S(t)x_b(t)}{\int_0^t D dt} \quad (8.38k)$$

where $\alpha = 2.46$ and $k = 1.646$. The filter model is given in (8.39). The initial batch size is completely separated into a waste stream containing only C and the Int2 stream containing A and B . The variables R^{cons} and R^{prod} represent the amount produced and consumed of each stream.

$$R_{Int1}^{cons} = b_{filt,n} = b_{react,n} \quad (8.39a)$$

$$R_{Wst}^{prod} = C_C^{Int1} b_{filt,n} \quad (8.39b)$$

$$R_{Int2}^{prod} = (1 - C_C^{Int1}) b_{filt,n} \quad (8.39c)$$

$$T_{filt,n}^p = .8 + .02b_{filt,n} \quad (8.39d)$$

The operating costs (8.40) associated with the reactor and distillation column are based on the processing time Tp and utility prices p associated with the manipu-

lated variables. The filter operating cost is based only on the batch size.

$$F_{react,n} = p^{react} b_{react,n} \int_0^{Tp_{react,n}} u(t) dt \quad (8.40a)$$

$$F_{filt,n} = R_{Int2}^{prod} + 4R_{Wst}^{prod} \quad (8.40b)$$

$$F_{dist,n} = p^{dist} VT p_{dist,n} \quad (8.40c)$$

where $p^{react} = 0.3$ and $p^{dist} = 1.5$. The operating constraints are of the form (8.2) initially

$$v^{min} \leq v \leq v^{max} \quad (8.41a)$$

$$R^{min} \leq R \leq R^{max} \quad (8.41b)$$

where $v^{min} = 1.8$, $v^{max} = 7.9$, $R^{min} = 2$, and $R^{max} = 7$.

The model for the reactor (8.37) is an ODE system and the model for the distillation (8.38) is a DAE system. The filter model (8.39) is a simple linear function of the batch size only. These equations constitute the unit model (8.24) - (8.26) in the general optimization formulation. The general operating constraints (8.29) and costs (8.30) for this system are given in (8.41) and (8.40). In addition, the material cost/sale prices are

$$P_{feed} = 5 \quad (8.42a)$$

$$P_{prod1} = 30 \quad (8.42b)$$

$$P_{prod2} = 45 \quad (8.42c)$$

and the parameters for the bounds (8.22) on slot duration for each unit are

$$1.5 \leq Tp_{react,n} \leq 3 \quad \forall s \in S_{react} \quad (8.43a)$$

$$1.5 \leq Tp_{dist,n} \leq 3 \quad s = dist1 \quad (8.43b)$$

$$1.125 \leq Tp_{dist,n} \leq 2.5 \quad s = dist2 \quad (8.43c)$$

and the bounds on batch size and excess material quantity are the same for all units/materials, where $B^{min} = 30kg$, $B^{max} = 60kg$, $E^{min} = 0kg$, and $E^{max} = 400kg$. We use a scheduling horizon of $H = 8hours$. In this formulation, there is no penalty associated with the makespan so the optimal schedule will typically extend over the entire horizon in order to maximize profit.

The differential equations in the MIDO are discretized using orthogonal collocation of finite elements to convert the problem into a MINLP [32]. We use 8 finite elements per event slot with 2 collocation points each. The resulting MINLP is solved in less than a minute using GAMS and the SBB solver, with CPLEX as the MILP solver and CONOPT as the nonlinear subsolver. The decision variables for the unit recipes (v and R) are allowed to change for every finite element.

8.2.4.3 Results and Discussion

We consider two fault events in addition to the ‘Base’ schedule:

- **Base:** This is the nominal solution to the scheduling problem, obtained before the schedule is executed or any fault occurs.
- **Fault 1:** At $t=1hr$ the existing PM efforts indicate that there is a fault in the reactor heating jacket (e.g., loss of steam quality) and it is not possible to operate the reactor at temperatures greater than $v^{max} = 6$. We assume

there is no delay in the PM step and that the scheduling layer is notified of the fault at $t=1\text{hr}$.

- **Fault 2:** The second fault is identified by batch PM techniques using recent batch data, in which the final product concentration is slightly below the target concentration. The PM efforts indicate that the reflux ratio setpoints are not being tracked perfectly by the control system (e.g., due to valve stiction) and that in order to maintain controller performance the reflux setpoint changes should not be greater than $0.25/\text{hr}$. This fault information is relayed to the scheduling layer before the start of the third event slot, and is accounted for in an additional constraint $R(k) - R(k - 1) < 0.25$, where k denotes the finite element.

Each fault will trigger a reschedule. When the rescheduling calculation is performed, constraints are added to ensure the quality of the updated schedule. We consider separately two different requirements:

- **Requirement 1:** We require the final amount of product produced to be equal to that predicted in the Base case.
- **Requirement 2:** We require the duration of the schedule to be the same as that in the Base case. There is no penalty for producing less product than the amount predicted for the Base case.

Figure 8.6 presents a Gantt chart comparing the optimal production schedules for the Base case to those obtained after Fault 1 and Fault 2 when Requirement 1 is

imposed in the rescheduling calculation. In order for the production schedule to be feasible after each fault occurrence, the scheduling horizon and makespan H must be extended (see the T_{end} values reported on each plot in 8.6). Figure 8.7 shows the corresponding optimal values of the manipulated variables v and R during each unit operation. After Fault 1, only the second reactor batch is impacted because the scaled temperature is well below the upper bound at time $t=1\text{hr}$. Slot $n=2$ must be extended to reach the desired level of conversion in place of having a high initial temperature, and minor adjustments are made to the distillation recipe in the fourth event slot. After Fault 2, the distillation recipes in the third and fourth event slots are adjusted to compensate for the restriction on rate of change of the reflux ratio. The first distillation operation is particularly affected since the original recipe called for a large increase in the reflux ratio towards the end of the processing time. The expected product quantity and profit for all three cases are reported in Figure 8.8. The amount of product 1 and product 2 is the same after each rescheduling calculation, however the expected profit is slightly lower due to the longer processing times.

The optimal variable trajectories when Requirement 2 is imposed are shown in Figure 8.9. The slot durations and total makespan are the same as the Base case in Figure 8.6. After each fault, the recipes for each unit operation adjust in a similar fashion to the changes observed when imposing Requirement 1. The average reactor temperature for the second reactor batch is higher than that for the original recipe in order to reach an appropriate conversion level without extending the event duration. This is the main cause of the large drop in profit reported in

Figure 8.10 after Fault 1.

In each of these cases, we demonstrate that obtaining a new, feasible schedule can be accomplished relatively easily once the relevant fault information has been communicated from the process control layer to the scheduling layer in Figure 8.2. Furthermore, depending on how long a fault was present before it was detected, it is likely that past schedules were not accurate representations of what was actually happening when the schedule was executed. Using this framework, the scheduling layer can be made aware of changes in the process and prepare for future production accordingly. This serves to ‘close the loop’ at the scheduling level.

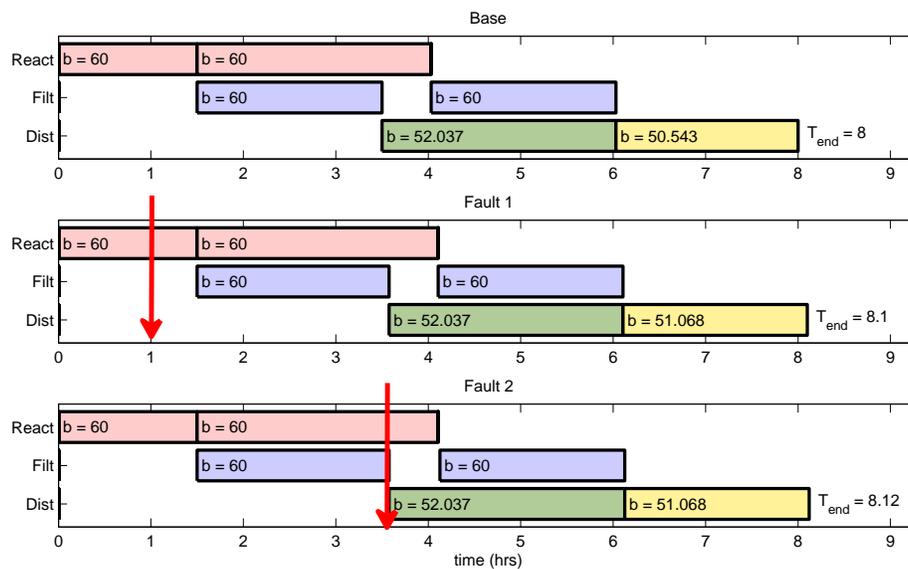


Figure 8.6: Gantt charts for optimal production schedules for the base case and two faults subject to Requirement 1. Arrows indicate when the rescheduling calculation is triggered and performed.

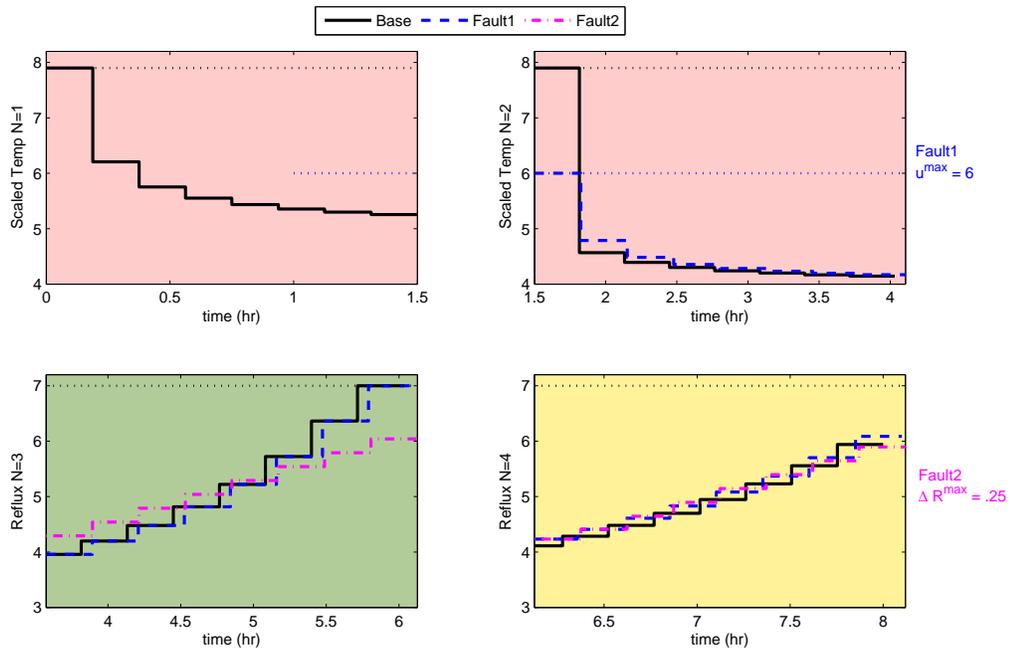


Figure 8.7: Variable trajectories under Requirement 1.

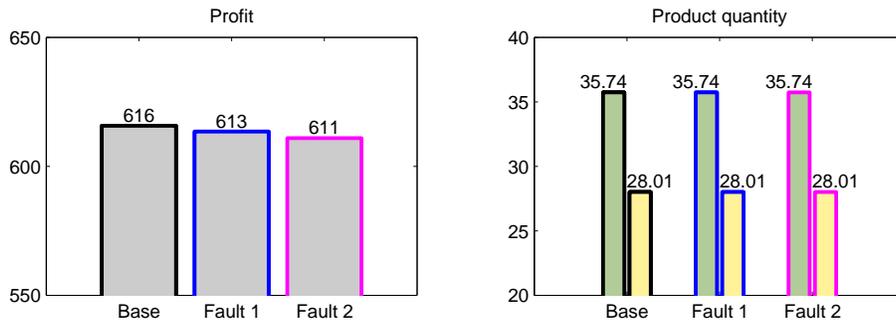


Figure 8.8: Profit and production levels under Requirement 1.

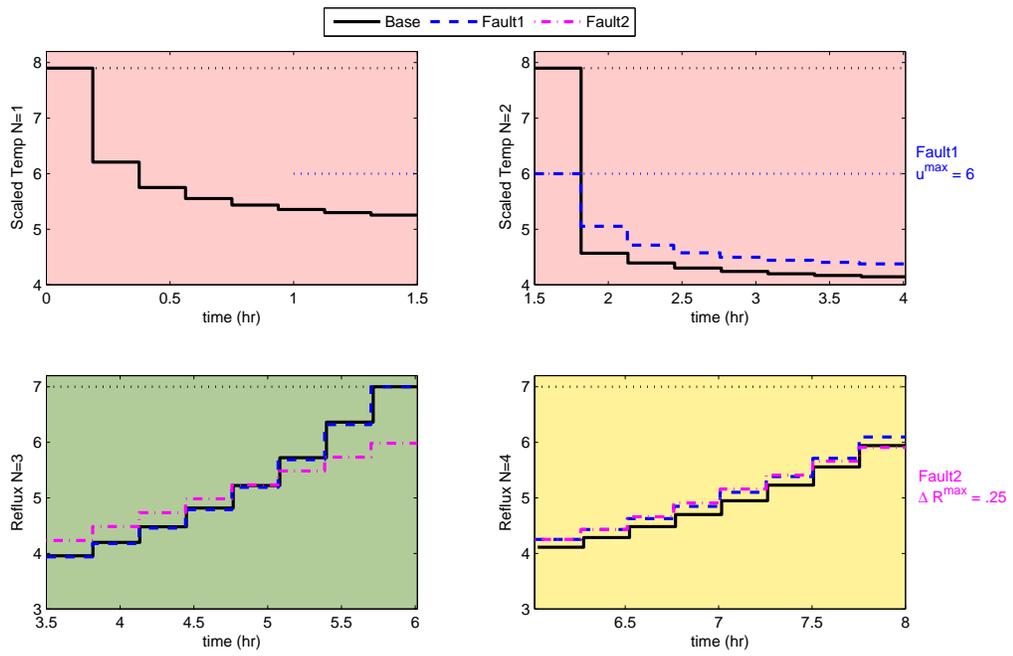


Figure 8.9: Variable trajectories under Requirement 2.

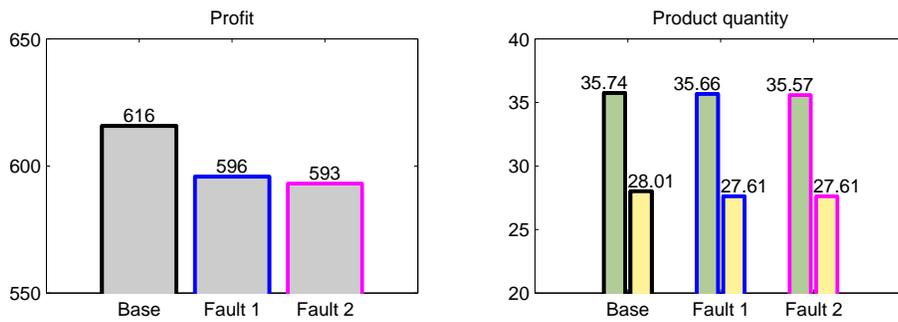


Figure 8.10: Profit and production levels under Requirement 2.

8.3 Moving Horizon Scheduling with Low-Order Dynamic Models

The solution of the scheduling problems formulated above, whether using a full-order process model or using the low-order scale-bridging models, provides a sequence of production targets that covers the entire horizon. While sufficient in the ideal case when demand and price forecasts are available for an extended amount of time and the process model is a perfect representation of the plant dynamics, this solution is likely suboptimal in practical situations when prices may fluctuate and the predicted process dynamics may not be accurate owing to disturbances and plant-model mismatch.

The above observations are best captured by drawing an analogy with control systems. Namely, determining the schedule for the entire horizon H amounts to an “open loop” solution which lacks process feedback. Thus “closing the loop” for productions scheduling requires incorporating process feedback. Our approach is based on the receding horizon control framework (and, more specifically, on economic model predictive control [9, 89]), in that we propose implementing the optimal schedule on a periodic basis, with the time horizon correspondingly shifted in time.

This effectively amounts to a moving horizon scheduling, with the price and demand forecasts being updated at each time point. Moreover, measurements of the process states are used to update the states of the model at each time point via an observer. Our moving horizon scheduling framework is described in the algorithm below:

1. Optimize the schedule using the dynamic model over the horizon for which price and demand forecasts are available
 2. Implement the optimal schedule on the plant (here we use a detailed model to represent the plant) and track measurements of scheduling-relevant process variables
 3. Use a state observer to determine the trajectories of the states of the scale-bridging models
 4. When new price and demand forecasts are available, or a disturbance is detected (this would be an event-driven rescheduling trigger), return to the scheduling problem and:
 - Shift the time horizon by the amount of time elapsed since the previous update
 - Update the initial conditions of the low-order model with the states of the observer
 - Update price and demand forecasts
 - Update endpoint constraints for inventory levels (to avoid depletion of product inventory)
 5. Return to step 1
-

8.3.1 Comparison to Economic MPC

While the objective of the two methods is the same, i.e., to minimize operating costs over a forecast horizon while meeting operational constraints, there are several

important differences between our proposed moving horizon scheduling framework using SBMs (or any dynamic model for that matter) and economic MPC (EMPC).

(1) *Model Structure*: The SBMs used in this work are single-input multi-output, and capture the *closed-loop* dynamics of the scheduling-relevant variables in response to a product target sequence, i.e., the input is the production target (product flowrate or quality) and the outputs are the evolution of the scheduling-relevant variables assuming that a (supervisory) control system is effectively guiding the plant throughout the production target changes. In contrast, EMPC models are multi-input multi-output, where the inputs are the manipulated variables (or set-points for distributed control loops) and the outputs are the controlled variables throughout the process.

SBMs have a single input and only several outputs resulting in a relatively small, sparse model [210]. MPC systems often accounts for m manipulated variables and n control variables (where m and n can be on the order of tens to hundreds) resulting in a much larger, non-sparse process model [222].

(2) *Execution Frequency*: An EMPC system must ensure that the process is stabilized throughout the horizon, and thus, it must be updated frequently (re-optimize the manipulated variable trajectories) to compensate for high frequency disturbances or plant-model mismatch. In contrast, the moving horizon scheduling framework assumes that a process control system is guiding (and stabilizing) the plant throughout transitions, and thus, the schedule can be re-optimized infrequently (e.g., when new price or demand forecasts are available). Conceivably, constraints could be imposed to ensure that the process remains within stability

limits of the control system.

8.3.2 Case Study: Moving Horizon ASU Scheduling

We consider the air separation unit (ASU) example used in Chapter 7 and use this to demonstrate the moving horizon scheduling concept.

Recall that the optimization problem is formulated to include the low-order dynamic models of scheduling-relevant variables and constraints:

$$\underset{y_p^{sp,n}, \alpha^{sp,n}, y_{inv}^{sp,n}}{\text{minimize}} \quad J = \int_0^{T_m} \phi(w, y_{inv}, \tilde{y}) dt \quad (8.44)$$

Subject to: *Time slots* (7.2)

Inventory model (7.4)

Production split/mixing ratio (7.5)

Process model:

Low-order dynamic process models (SBMs)

for each identified variable w

Constraints:

Inventory: (7.6)

$$\text{Quality: } \hat{h}_{product}(w, \bar{y}, t) \leq 0 \quad (8.45)$$

$$\text{Process: } \hat{h}_{process}(w, y_p^{sp}, t) \leq 0 \quad (8.46)$$

This formulation alters the original problem considering the full-order dynamic process model by, i) considering only the scheduling relevant subset of product quality and process operating constraints ($\hat{h}_{product}$ and $\hat{h}_{process}$). Additionally, ii)

the relevant variable trajectory predictions, $w(t)$, are determined using low-order dynamic models rather than the detailed dynamic process model (7.7).

We relied on Hammerstein-Weiner models of the form (8.47) to capture the nonlinear dynamic behaviour of the eight scheduling relevant variables (refer to Chapter 7). Here, Ψ and Φ are nonlinear transformations of the input (i.e., the production setpoints) and output. The matrices A , B and C describe a linear state-space model. The single output variable w contains dynamic information of a variable relevant to the scheduling calculation.

$$u' = \Psi(y_p^{sp}) \quad (8.47a)$$

$$\dot{\bar{x}} = A\bar{x} + Bu' \quad (8.47b)$$

$$y' = C\bar{x} \quad (8.47c)$$

$$w = \Phi(y') \quad (8.47d)$$

In conjunction with the Hammerstein-Wiener models (8.47), we use a high-gain (Luenberger) observer to update the model states \bar{x} :

$$u' = \Psi(y_p^{sp}) \quad (8.48a)$$

$$\dot{\bar{x}} = A\bar{x} + Bu' + L(\hat{w} - w) \quad (8.48b)$$

$$y' = C\bar{x} \quad (8.48c)$$

$$w = \Phi(y') \quad (8.48d)$$

where \hat{w} is the measurement of the scheduling relevant variable. The states \bar{x} are assumed to be observable.

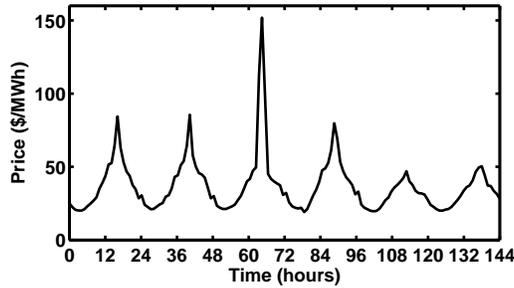


Figure 8.11: Day-ahead market prices for July 10 - July 15 2013 from the Electricity Reliability Council of Texas (<http://www.ercot.com/mktinfo/>)

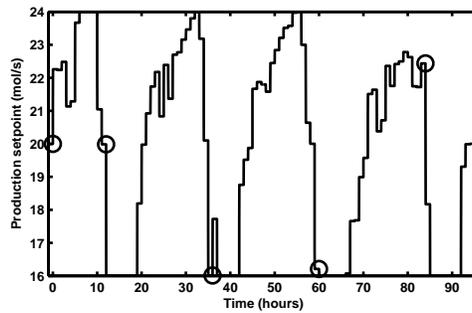


Figure 8.12: Production rate setpoint. The nominal flowrate (demand) is 20mol/s and the production rate may vary by up to 20%

8.3.2.1 Results and Discussion

We assume that accurate 2-day forecasts of time-variable electricity prices are available daily at noon (Figure 8.11). Thus, the algorithm introduced at the start of the section was implemented with a rescheduling period of 1 day (when new forecasts are available) with a two-day prediction horizon. A terminal point constraint [247] ($M_{inv} \geq 0.5M_{inv}^{max}$) was implemented to stabilize the inventory level throughout the horizon. The optimal production setpoint and inventory levels to meet a constant demand of 20mol/s with impurity levels below 2000ppm are shown

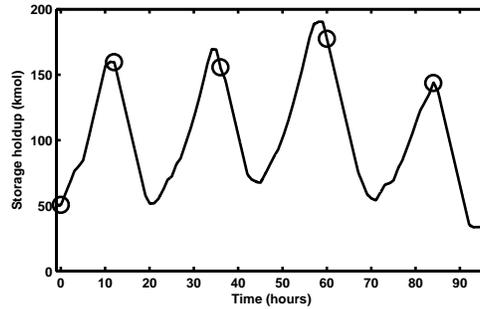


Figure 8.13: Evolution of inventory level. The storage capacity is limited to 200 kmol.

in Figures 8.12 and 8.13. Markers indicate points where the schedule was recalculated. While there are eight scheduling-relevant constraints and models, we will focus our discussion on the impurity and temperature driving force across the reboiler/condenser because these are examples of variables that evolve, respectively, over very slow and fast time scales and are near their bounds.

As expected, the production flowrate setpoint increases when prices are low, thus the inventory level in the liquid nitrogen storage tank rises. Then, when prices are high, the production target setpoint is lowered to the minimum allowed value and stored product is used to satisfy the remaining demand. Note that the maximum production rate (24mol/s) is only reached on several brief occasions; this is due to the fact that several operating constraints are near their bound when production rate is increased significantly in a short time interval. The impurity level and reboiler/condenser temperature difference, shown in Figures 8.14 and 8.15, are near their bounds at these times. Overall, the variable capacity operation results in an electricity cost savings of \$1620 (or 4.85%) over the 4-day horizon in

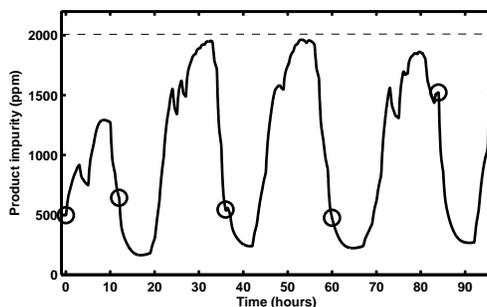


Figure 8.14: Evolution of impurity levels in the product. The maximum allowed level is 2000ppm

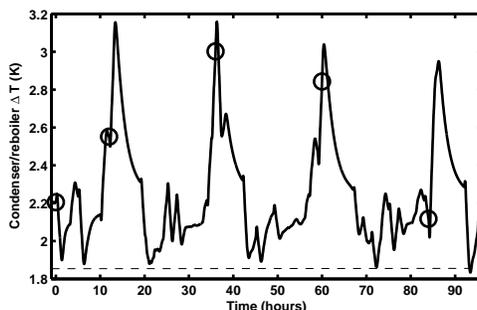


Figure 8.15: Evolution of reboiler/condenser temperature driving force. The temperature difference must be greater than 1.85K.

comparison to a constant production profile set at the nominal rate.

In this case study, we assume that there are no disturbances, and plant-model mismatch is minimal. However, in reality, processes are always subject to disturbances, and this framework allows for a rapid rescheduling calculation online when a disturbance is detected. We note here that the average CPU time for each optimization calculation was around 30 min on a desktop computer, a significant reduction compared with the computational effort required to solve the equivalent problem based on the full-order model of the process.

8.4 Conclusion

In this chapter, we emphasized differences between disturbances to the scheduling and process layer, and highlight how the latter can lead to changes in processing times or recipes, transition times, and the overall product wheel. We proposed a novel event-driven rescheduling framework which relies on existing methods for process monitoring and fault detection, identification, and reconstruction, which diagnose specific changes in processing conditions. The integrated scheduling and dynamics framework is well suited for accounting for these differences; the new operating paradigm can be accounted for by altering constraint boundaries for the dynamic model in a rescheduling calculation. A case study for a sequential batch process was used to demonstrate the ease of rescheduling once a dynamic disturbance has been identified. We also discussed periodic rescheduling strategies and the use of moving horizon scheduling, and compared the dynamic models used in this formulation to those used in economic MPC strategies.

Chapter 9

Conclusion and Future Directions

9.1 Summary of Contributions

In this thesis, we have demonstrated how embedding dynamics and control considerations in operational optimization decisions can improve the economic performance and energy management of processes and energy systems. Specifically, we considered the built environment (Chapters 3, 4, 5) and chemical processes (Chapters 2, 6, 7, and 8). The desire to participate in demand response initiatives and take advantage of time of use pricing schemes motivates the development of advanced decision making frameworks that can leverage differences in price (by increasing energy consumption at times when prices are low, and reduce consumption when prices are high) to achieve lower daily operating costs (Chapters 3, 4, 7). Such load shifting behavior, and the use of distributed generation resources (Chapter 5), can result in overall improvements in the efficiency of the energy supply chain by aligning demand with the availability of renewable resources and use of efficient generation technologies. In addition, manufacturing processes can lower their operating costs and improve predictions of event durations accounting for transition dynamics in continuous (Chapter 6 and 7) and batch (Chapter 8) production scheduling.

Incorporating a dynamic model into an optimization-based controller or scheduling formulation necessitates solving a dynamic optimization. Solution strategies rely on either (i) discretization of any differential equations in the constraint set to convert the dynamic optimization into a (large) program, which can be solved using a simultaneous solver, or (ii) using a sequential solution approach. In both cases, the presence of a nonlinear model with a large number of states, integer decision variables, and multiple time scales can render the optimization problem not solvable in a practical amount of time. We address these challenges through *meaningful model reduction, hierarchical controller design, and novel scheduling formulations*.

Model reduction. We identify computationally tractable dynamic model forms for the application at hand. In Chapters 2 and 3 we use singular perturbation to derive low-order, non-stiff models of the process dynamics which are relevant to energy management. Chapters 3 and 4 also use aggregate models of a building envelope and heat load in place of a high fidelity model of the building geometry and occupant behavior. In chapters 6 and 7 system identification methods are employed to develop low-order models of a subset of variables whose dynamics are relevant to the optimization objective and constraints. A framework for selecting variables and dynamics of interest to operational decisions was presented in Chapter 7.

Hierarchical controller designs. In Part I we demonstrate how a thorough understanding of the system dynamics and operating space can aid the design of hierarchical control systems for energy management. In Chapter 2 we propose a

supervisory nonlinear MPC layer for optimizing the trajectories of the slow process dynamics, while a linear controller is used to guarantee stability of the fast dynamics. This composite controller design is applicable to a general class of processes with tight energy integration. Based on this result, we design a nonlinear MPC for energy management in buildings with energy recovery in Chapter 3. In this case, we focus on the slow dynamics because the fast dynamics are inherently stable. In Chapter 4 we justify a composite control system for buildings with thermal energy storage based on the different dominant time scales for active and passive storage in the building and HVAC system, and the differences in minimum prediction horizon length for the economic and tracking objectives. In Chapter 5 we consider the difference between centralized and decentralized operation of the ensemble of distributed generation units. This is an important consideration for anticipating the effect on the electric grid of widespread adoption of the strategies proposed in Chapters 3 and 4.

Novel scheduling formulations. In part II we incorporate low-order dynamic models into scheduling calculations, which is a significant departure from tradition scheduling approaches which use tabulated process data. In Chapter 6 we developed a novel continuous-time scheduling formulation using a discrete-time process model, and introduced a novel ‘reverse integrated error’ concept to determine transition times between products. Chapter 7 presents a general framework for the selection of scheduling-relevant variables and an industrial-scale case study to demonstrate the benefits of scheduling with a dynamic scale bridging model. In chapter 8 we propose using information from existing process monitoring methods

for fault detection and state estimation to update parameter values and constraint formulations in the dynamics models incorporated into scheduling. These efforts aim to ‘close the scheduling loop’ using feedback on the process conditions.

Collectively, these efforts span operational decisions involving regulatory control, supervisory control, and scheduling, and demonstrate the ability to reduce the operating costs of residential, commercial, and industrial consumers. These improvements are attributed to the additional information provided by a dynamic model compared to heuristics or tabulated process data. In each case considered, we observed reductions in solution times for the optimization problems at hand due to the desirable computational characteristics of the models, thereby providing real incentive for online implementation in a moving horizon fashion. Ultimately, these strategies for managing the dynamics of processes and energy systems will have positive effects throughout the energy supply chain.

9.2 Future Research Directions

The work in Part I has furthered the study of building energy management and coordination of distributed generation, however we have considered a relatively small control volume for all of these studies (i.e., individual buildings or small neighborhoods). Future efforts should focus on a larger geographic area in order to quantify the exact benefits (reductions in primary energy consumption, associated changes in emissions, etc.) of widespread adoption of advanced control strategies and participation in demand response programs, rather than extrapolate such conclusions from small data sets. This is an analytical effort, instead of

controller or schedule design, so the tradeoff between accuracy and solution speed when modeling consumer behavior and decision-making should favor accuracy and quantify uncertainties. This model should be a holistic picture of residential, consumer, and industrial consumer behavior and the dynamics of the electric grid and electricity generation, and should take advantage of existing software (e.g., the recently developed FESTIV model by NREL, for multi time-scale modeling of power systems).

Similarly, the combined operation of electric and natural gas grid dynamics provides an interesting path forward, and is an emerging research thrust in the literature. The huge discrepancies in dominant time constant, robustness to uncertainties in demand, and market structure for the two systems make the model development for an area the size of the ERCOT grid a challenging task, which is probably worthy of high performance computing. A high fidelity model of both grids would provide a test bed for evaluating new market structures and the ability to investigate the optimal operation of the combined system.

There are several possible extensions of the scheduling concepts presented in Part II. It would be valuable to apply the framework for selection of scheduling-relevant variables to other grid-dependent processes with storage (e.g., production of pulp and paper, ammonia, etc.) and analyze their cumulative impact on the electric grid. The frameworks for moving horizon scheduling and rescheduling using process monitoring-based triggers should be extended beyond this proof-of-concept stage. This entails developing case studies which demonstrate all aspects of the framework in detail. This will be a highly interdisciplinary undertaking, requiring

expertise in the fields of process monitoring, control, and scheduling. As is the case with integrating scheduling and control, a major challenge will be effectively communicating these concepts to each individual research community. A rigorous mathematical definition of the framework with a unified notation for discussing and sharing information between process monitoring, control, and scheduling would serve to reduce the communication barrier.

Dissemination of Research

Journal Publications

1. C. R. Touretzky, M. Baldea, *Closing the Scheduling Loop: Rescheduling Formulations Incorporating Dynamic Process Models, Fault-Detection Based Triggers, and Moving Horizon Concepts*, In Preparation.
2. R. Pattison, C. R. Touretzky, I. Harjunkoski, M. Baldea, *Moving Horizon Scheduling of Process Operations Under Dynamic Constraints*, AIChE Journal, Under Review, 2016.
3. C. R. Touretzky, D. McGuffin, J. Ziesmer, M. Baldea, *The Effect of Distributed Electricity Generation Using Natural Gas on the Electric and Natural Gas Grid*, Applied Energy, Under Review, 2016.
4. R. Pattison, C. R. Touretzky, T. Johansson, I. Harjunkoski, M. Baldea, *Production Scheduling with Dynamic Low-Order Surrogate Models: Framework and Application to an Air Separation Unit*, Industrial & Engineering Chemistry Research, Accepted, 2016.
5. C. R. Touretzky, M. Baldea, *A Hierarchical Scheduling and Control Strategy for Thermal Energy Storage Systems*, *Energy and Buildings*, Vol. 110, pp. 94-107, Jan. 2016.
6. C. R. Touretzky, R. Patil, *Building-Level Power Demand Forecasting Using Building Specific Inputs: Development and Applications*, Applied Energy, Vol. 147, pp. 466-477, March 2014.
7. C. R. Touretzky, M. Baldea, *Integrating Scheduling and Control for Economic MPC of Buildings with Thermal Energy Storage*, Journal of Process Control - Special Issue on Economic MPC, Vol. 24, pp. 1292-1300, Aug. 2014.
8. C. R. Touretzky, M. Baldea, *Nonlinear Model Reduction and MPC of Residential Buildings with Energy Recovery*, Journal of Process Control Special Issue on Energy Efficient Buildings, Vol. 24, pp. 723-739, June 2014.

9. M. Baldea, C. Touretzky, *Nonlinear Model Predictive Control of Energy-Integrated Process Systems*, Systems and Control Letters, Vol. 62, pp. 723-731, Sept. 2013.

Conference Publications

1. C. R. Touretzky, M. Baldea, I. Harjunoski, *A Framework for Integrated Scheduling and Control Using Empirical Discrete-Time Models*, Proceedings of 2016 European Symposium on Computer Aided Process Engineering (ESCAPE), Accepted, July 2016.
2. R. Pattison, C. R. Touretzky, I. Harjunoski, M. Baldea, *Moving Horizon Scheduling of an Air Separation Unit Under Fast-Changing Energy Prices*, Proceedings of 2016 Conference on Dynamics Control and Process Systems (DYCOPS), Accepted, June 2016.
3. C. R. Touretzky, R. Pattison, J. Park, M. Baldea, I. Harjunoski, *Handling Input Dynamics in Integrated Scheduling and Control*, Proceedings of 2016 IEEE Conference on Automation, Quality, Testing, and Robotics (AQTR), Submitted, May 2016.
4. C. R. Touretzky, A. Salliot, L. Lefevre, M. Baldea, *Optimal Operation of Phase Change Material Thermal Energy Storage for a Commercial Building*, Proceedings of 2015 American Control Conference (ACC), pp. 980-985, Aug. 2015.
5. C. R. Touretzky, R. Patil, *A General Power Modeling Framework for Individual Building Demand Management*, Proceedings of 2014 ASME Dynamic System and Control Conference (DSCC), pp. 1-9, Oct. 2014.
6. C. R. Touretzky, M. Baldea, *Model Reduction and Nonlinear MPC for Energy Management in Buildings*, Proceedings of 2013 American Control Conference (ACC), pp. 443-448, June 2013.
7. M. Baldea, C. Touretzky, *Concurrent Nonlinear Predictive Control and Economic Management of Energy-Integrated Systems*, Proceedings of 2013 European Control Conference (ECC), 2062-2067, July 2013.

Patents

- Patil, R., Sharma, R. and Touretzky, C., NEC Laboratories America, Inc., POWER MODELING BASED BUILDING DEMAND MANAGEMENT SYSTEM. U.S. Patent Application 20,150,268,650., issued September 24, 2015.

Presentations

* indicates the presenting author.

- American Institute of Chemical Engineers (AIChE) Meetings and Workshops
 1. C. R. Touretzky*, I. Harjunkoski, M. Baldea, *Fault Detection-Based Triggers for Rescheduling of Batch Process Operations*, Annual Meeting, Salt Lake City, UT, Nov. 2015.
 2. C. R. Touretzky*, I. Harjunkoski, M. Baldea, *A Hybrid Scheduling Formulation for Integrated Scheduling and Process Control using Discrete-Time Dynamic Models*, Annual Meeting, Salt Lake City, UT, Nov. 2015.
 3. C. R. Touretzky*, D. McGuffin, J. Ziesmer, M. Baldea, *Securing our Energy Supply: The Effect of Distributed Electricity Generation Using Natural Gas on the Interdependency of the Electric and Natural Gas Grids*, Annual Meeting, Salt Lake City, UT, Nov. 2015.
 4. C. R. Touretzky*, R. Pattison, T. Johansson, I. Harjunkoski, M. Baldea, *Integrated Scheduling and Dynamic Optimization of a Cryogenic Air Separation Unit Subject to Time-Varying Electricity Prices*, Annual Meeting, Salt Lake City, UT, Nov. 2015.
 5. C. R. Touretzky*, M. Baldea, *Predictive Strategies for Coordinating Passive and Active Thermal Energy Storage in Buildings*, Spring Meeting, Austin, TX, May 2015.
 6. C. R. Touretzky, M. Baldea*, *Dynamics and Control of Energy Utilization in Buildings with Thermal Storage*, Annual Meeting, San Francisco, CA, Nov. 2013.
 7. W. Cole*, C. R. Touretzky, M. Baldea, T. Edgar, *Reduced Order Modeling Strategies for Optimal Energy Management in Buildings*, Workshop on Smart Grid for the Chemical Process Industry, Chicago, IL, Sept. 2013.

8. C. R. Touretzky*, M. Baldea, *Proactive Strategies for Building Energy Management in the Smart Grid*, Spring Meeting, San Antonio, TX, May 2013.
 9. C. R. Touretzky*, M. Baldea, *Modeling and Model Predictive Control Strategies for Building Energy Management*, Annual Meeting, Pittsburgh, PA, Nov 2012. ¹
- American Control Conference (ACC) Annual Meetings
 10. C. R. Touretzky*, A. Salliot, L. Lefevre, M. Baldea, *Optimal Operation of Phase Change Material Thermal Energy Storage for a Commercial Building*, Chicago, IL, July 2015.
 11. C. R. Touretzky*, M. Baldea, *Model Reduction and Nonlinear MPC for Energy Management in Buildings*, Washington D.C., June 2013.
 - American Society of Mechanical Engineers (ASME) Dynamic Systems and Control Conference (DSCC)
 12. C. R. Touretzky*, R. Patil, *A General Power Modeling Framework for Individual Building DM*, San Antonio, TX, Oct. 2014.
 13. C. R. Touretzky*, M. Baldea, *Frontiers in Building HVAC Controls: Optimal Energy Management for Buildings Using Thermal Energy Storage*, San Antonio, TX, Oct. 2014.
 - Texas Wisconsin California Control Consortium (TWCCC)
 14. C. R. Touretzky*, I. Harjunoski, M. Baldea, *Fault Detection-Based Triggers for Rescheduling of Batch Process Operations*, Austin, TX, March, 2016.
 15. C. R. Touretzky*, I. Harjunoski, M. Baldea, *Integration of Production Scheduling and Model Predictive Control*, Austin, TX, March, 2015.
 16. C. R. Touretzky*, M. Baldea, *Optimal Energy Management for Buildings Using Thermal Energy Storage*, Madison, WI, Oct. 2014.
 - Next Generation Photovoltaics (NGPV) NSF I/UCRC Center

¹Recipient of Computing and Systems Technology (CAST) division travel grant and best student presentation award.

17. C. R. Touretzky*, I. Harjunoski, M. Baldea, Fort Collins, *Integration of Scheduling and Control for Process Operations using Renewable Energy* CO, Sept. 2015.
 18. C. R. Touretzky*, J. Du, J. Park, M. Baldea, I. Harjunoski, *Integrated Production Scheduling and Model Predictive Control*, Austin, TX, March 2015.
 19. C. R. Touretzky*, J. Du, J. Park, M. Baldea, I. Harjunoski, *Integration of Production Scheduling and Control in the Polymer Supply Chain*, Fort Collins, CO, Sept. 2014.
- Poster Competitions
20. C. Touretzky*, Energy Forum Poster Presentation, University of Texas at Austin, Energy Week, Feb. 2015. ²
 21. C. Touretzky*, Energy Forum Poster Presentation, University of Texas at Austin, Energy Week, Feb. 2013

²Recipient of first place award in the Energy Storage and Efficiency category.

Bibliography

- [1] T. Ackermann, G. Andersson, and L. Söder. Distributed generation: a definition. *Electrical Power Systems Research*, 57:195 – 204, 2000.
- [2] T. Ackermann and V. Knyazkin. Interaction between distributed generation and the distribution network: operation aspects. In *Transmission and Distribution Conference and Exhibition 2002: Asia Pacific. IEEE/PES*, volume 2, pages 1357–1362. IEEE, 2002.
- [3] A. Afram and F. Janabi-Sharifi. Theory and applications of HVAC control systems - a review of model predictive control (MPC). *Building and Environment*, 72:343–355, 2014.
- [4] M. H. Albadi and E.-S. E. F. Demand response in electricity markets: An overview. *Proc. IEEE Power Eng. Soc. Gen. Meet.*, pages 1–5, 2007.
- [5] R. Albert, I. Albert, and G. Nakarado. Structural vulnerability of the north american power grid. *Physical Review E*, 69, 2004.
- [6] V. Alexiades and A. D. Soloman. Mathematical modeling of melting and freezing processes. *Hemisphere Publishing corporation: Washington D.C.*, 1993.
- [7] D. A. Allan, J. B. Rawlings, and T. A. Badgwell. Closed-loop properties of a mixed scheduling and control problem. *AIChE Annual Meeting, 2015*, November 2015.
- [8] R. Amrit, J. Rawlings, and D. Angeli. Economic optimization using model predictive control with a terminal cost. *Annual Reviews in Control*, 35:178–186, 2011.
- [9] R. Amrit, J. B. Rawlings, and L. T. Biegler. Optimizing process economics online using model predictive control. *Comput. Chem. Eng.*, 58:334–343, 2013.
- [10] D. Angeli, R. Amrit, and J. B. Rawlings. On average performance and stability of economic model predictive control. *IEEE Trans. Automat. Contr.*, 57(7):1615–1626, 2012.

- [11] D. Angeli, R. Amrit, and J. B. Rawlings. On average performance and stability of economic model predictive control. *Automatic Control, IEEE Transactions on*, 57(7):1615–1626, 2012.
- [12] E. M. B. Aske, S. Strand, and S. Skogestad. Coordinator {MPC} for maximizing plant throughput. *Comput. Chem. Eng.*, 32(12):195 – 204, 2008. Process Systems Engineering: Contributions on the State-of-the-Art Selected extended Papers from {ESCAPE} '16/PSE 2006.
- [13] H. Aytug, M. A. Lawley, K. McKay, S. Mohan, and R. Uzsoy. Executing production schedules in the face of uncertainties: A review and some future directions. *European Journal of Operational Research*, 161(1):86–110, 2005.
- [14] M. Baldea, A. Araujo, S. Skogestad, and P. Daoutidis. Dynamic considerations in the synthesis of self-optimizing control structures. *AIChE J.*, 54:1830–1841, 2008.
- [15] M. Baldea and P. Daoutidis. Model reduction and control of reactor–heat exchanger networks. *J. Proc. Contr.*, 16:265–274, 2006.
- [16] M. Baldea and P. Daoutidis. Control of integrated process networks—a multi-time scale perspective. *Comput. Chem. Eng.*, 31:426–444, 2007.
- [17] M. Baldea and P. Daoutidis. Dynamics and control of process networks with high energy throughput. *Comput. Chem. Eng.*, 32:1964–1983, 2008.
- [18] M. Baldea and P. Daoutidis. *Dynamics and Control of Integrated Process Systems*. Cambridge University Press, Cambridge, UK, 2012.
- [19] M. Baldea and P. Daoutidis. *Dynamics and Nonlinear Control of Integrated Process Systems*. Cambridge University Press, 2012.
- [20] M. Baldea, P. Daoutidis, and A. Kumar. Dynamics and control of integrated networks with purge streams. *AIChE J.*, 52:1460–1472, 2006.
- [21] M. Baldea, N. El-Farra, and B. Ydstie. Dynamics and control of chemical process networks: Integrating physics, communication and computation. *Comput. Chem. Eng.*, <http://dx.doi.org/10.1016/j.compchemeng.2012.05.016>.
- [22] M. Baldea and I. Harjunoski. Integrated production scheduling and process control: A systematic review. *Comput. Chem. Eng.*, 71:377–390, 2014.

- [23] M. Baldea and I. Harjunoski. Integrated production scheduling and process control: A systematic review. *Comput. Chem. Eng.*, 71:377–390, 2014.
- [24] M. Baldea and I. Harjunoski. A systematic review of the integration of production scheduling and process control. *Comput. Chem. Eng.*, 71:377–390, 2014.
- [25] M. Baldea, J. Park, J. Du, and I. Harjunoski. Integrated production scheduling and model predictive control of continuous processes. *AIChE J.*, 61(12):4179–4190, 2015.
- [26] M. Baldea and C. R. Touretzky. Concurrent nonlinear predictive control and economic management of energy-integrated systems. In *2013 European Control Conference (ECC)*, pages 2062–2067. IEEE, 2013.
- [27] M. Baldea and C. R. Touretzky. Nonlinear model predictive control of energy integrated process systems. *Sys. & Contr. Lett.*, 62(9):723 – 731, 2013.
- [28] V. Bansal, V. Sakizlis, R. Ross, J. D. Perkins, and E. N. Pistikopoulos. New algorithms for mixed-integer dynamic optimization. *Comput. Chem. Eng.*, 27(5):647–668, 2003.
- [29] H. Barker, A. Tan, and K. Godfrey. Design of multilevel perturbation signals with harmonic properties suitable for nonlinear system identification. *IEEE Proc. Contr. Theory and Applications*, 151(2):145–151, 2004.
- [30] S. Barsali, M. Ceraolo, P. Pelacchi, and D. Poli. Control techniques of dispersed generators to improve the continuity of electricity supply. In *Power Engineering Society Winter Meeting, 2002. IEEE*, volume 2, pages 789–794. IEEE, 2002.
- [31] M. Bazilian, M. Miller, R. Detchon, M. Liebreich, W. Blyth, M. Futch, V. Modi, L. Jones, B. Barkett, M. Howells, I. MacGill, D. M. Kammen, T. Mai, M. Wittenstein, S. Aggarwal, M. OMalley, J. Carvallo, M. Welsch, G. Pugh, R. Weston, and D. J. Arent. Accelerating the global transformation to 21st century power systems. *The Elec. J.*, 26:39–51, 2013.
- [32] L. Biegler. *Nonlinear programming: concepts, algorithms, and applications to chemical processes*. SIAM, Philadelphia, 2010.

- [33] L. T. Biegler. An overview of simultaneous strategies for dynamic optimization. *Chemical Engineering and Processing: Process Intensification*, 46(11):1043–1053, 2007.
- [34] L. T. Biegler. *Nonlinear programming: concepts, algorithms, and applications to chemical processes*, volume 10. SIAM, 2010.
- [35] A. Bilodeau and K. Agbossou. Control analysis of renewable energy system with hydrogen storage for residential applications. *Journal of Power Sources*, 162:757 – 764, 2005.
- [36] S. Bourne and A. Novoselac. Compact phase change based thermal stores: Experimental apparatus, methodology, and results. *ASHRAE 2014 Winter Meeting*, 2014.
- [37] S. Bourne and A. Novoselac. Compact pcm-based thermal stores for shifting peak cooling loads. *Building Simulation*, 8(6):673–688, 2015.
- [38] J. E. Braun, K. W. Montgomery, and N. Chaturvedi. Evaluating the performance of building thermal mass control strategies. *HVAC&R Research*, 7(4):403–428, 2001.
- [39] K. E. Brenan, S. L. Campbell, and L. R. Petzold. *Numerical solution of initial-value problems in differential-algebraic equations*, volume 14. Siam, 1996.
- [40] California Energy Commission. Source energy & environmental impacts of thermal energy storage. Report P500-95-005, California Energy Commission, 1996.
- [41] California ISO. What the duck curve tells us about managing a green grid, accessed Feb. 19, 2016.
- [42] E. Camacho, D. Ramirez, D. Limon, D. Muñoz de la Peña, and T. Alamo. Model predictive control techniques for hybrid systems. *Annual Reviews in Control*, 34(1):21–31, Apr. 2010.
- [43] Y. Cao. Design for dynamic performance: Application to an air separation unit. *Master’s Thesis, McMaster University*, 2011.
- [44] Y. Cao, C. Swartz, and M. Baldea. Design for dynamic performance: Application to an air separation unit. *American Control Conference*, pages 2683–2688, 2011.

- [45] Y. Cao, C. Swartz, M. Baldea, and S. Blouin. Optimization-based assessment of design limitations to air separation plant agility in demand response scenarios. *Journal of Process Control*, 33:37–48, 2015.
- [46] Y. Cao, C. Swartz, M. Baldea, and S. Blouin. Optimization-based assessment of design limitations to air separation plant agility in demand response scenarios. *J. Proc. Contr.*, <http://dx.doi.org/10.1016/j.jprocont.2015.05.002> .
- [47] Y. Cao, C. L. E. Swartz, J. Flores-Cerrillo, and J. Ma. Dynamic modeling and collocation-based model reduction of cryogenic air separation units. *AIChE J.*, 2016.
- [48] W. F. Castle. Air separation and liquefaction: recent developments and prospects for the beginning of the new millennium. *International Journal of Refrigeration*, 25(1):158–172, 2002.
- [49] F. E. Cellier and E. Kofman. *Continuous system simulation*. Springer Science & Business Media, 2006.
- [50] C. Chatzidoukas, C. Kiparissides, J. Perkins, and E. Pistikopoulos. Optimal grade transition campaign scheduling in a gas-phase polyolefin FBR using mixed integer dynamic optimization. *Computer Aided Chemical Engineering*, 14:71–76, 2003.
- [51] M. Chaudry, N. Jenkins, and G. Strbac. Multi-time period combined gas and electricity network optimisation. *Electric Power Systems Research*, 78(7):1265 – 1279, 2008.
- [52] X. Chen, M. Heidarinejad, J. Liu, and P. D. Christofides. Composite fast-slow mpc design for nonlinear singularly perturbed systems. *AIChE Journal*, 58(6):1802–1811, 2012.
- [53] X. Chen, M. Heidarinejad, J. Liu, D. M. de la Peña, and P. D. Christofides. Model predictive control of nonlinear singularly perturbed systems: Application to a large-scale process network. *Journal of Process Control*, 21(9):1296–1305, 2011.
- [54] Y. Chen and C. Yu. Design and control of heat integrated reactors. *Ind. Eng. Chem. Res.*, 42:2791–2808, 2003.

- [55] N.-Y. Chiang and V. M. Zavala. Large-scale optimal control of interconnected natural gas and electrical transmission systems. *Applied Energy*, 168:226–235, 2016.
- [56] L. Chick, M. Weimar, G. Whyatt, and M. Powell. The case for natural gas fueled solid oxide fuel cell power systems for distributed generation. *Fuel Cells*, 15(1):49–60, 2015.
- [57] P. Christofides and P. Daoutidis. Compensation of measurable disturbances for two-time-scale nonlinear systems. *Automatica*, 32(11):1553–1573, 1996.
- [58] Y. Chu and F. You. Integration of scheduling and control with online closed-loop implementation: Fast computational strategy and large-scale global optimization algorithm. *Comput. Chem. Eng.*, 47:248–268, 2012.
- [59] Y. Chu and F. You. Integrated scheduling and dynamic optimization of sequential batch processes with online implementation. *AIChE Journal*, 59(7):2379–2406, 2013.
- [60] Y. Chu and F. You. Integration of production scheduling and dynamic optimization for multi-product CSTRs: Generalized Benders decomposition coupled with global mixed-integer fractional programming. *Comput. Chem. Eng.*, 58:315–333, 2013.
- [61] Y. Chu and F. You. Integrated planning, scheduling, and dynamic optimization for batch processes: MINLP model formulation and efficient solution methods via surrogate modeling. *Ind. Eng. Chem. Res.*, 53:13391–13411, 2014.
- [62] Y. Chu and F. You. Moving horizon approach of integrating scheduling and control for sequential batch processes. *AIChE J.*, 60(5), 2014.
- [63] Y. Chu and F. You. Model-based integration of control and operations: Overview, challenges, advances, and opportunities. *Comput. Chem. Eng.*, 2015.
- [64] B. Coffey, F. Haghghat, E. Morofsky, and E. Kutrowski. A software framework for model predictive control with GenOpt. *Energy and Buildings*, 42:1084–1092, 2010.

- [65] W. Cole, T. Edgar, and A. Novoselac. Use of model predictive control to enhance flexibility of thermal storage cooling systems. *Proc. of Amer. Contr. Conf.*, pages 2788–2793, 2012.
- [66] W. J. Cole, K. M. Powell, and T. F. Edgar. Optimization and advanced control of thermal energy storage systems. *Reviews in Chemical Engineering*, 28:81–99, 2012.
- [67] W. J. Cole, J. D. Rhodes, W. Gorman, K. X. Perez, M. E. Webber, and T. F. Edgar. Community-scale residential air conditioning control for effective grid management. *Applied Energy*, 130:428–436, 2014.
- [68] P. Costamagna, L. Magistri, and A. Massardo. Design and part-load performance of a hybrid system based on a solid oxide fuel cell reactor and a micro gas turbine. *Journal of Power Sources*, 96(2):352–368, 2001.
- [69] R. L. Daniels and P. Kouvelis. Robust scheduling to hedge against processing time uncertainty in single-stage production. *Management Science*, 41(2):363–376, 1995.
- [70] P. Daoutidis and C. Kravaris. Dynamic output feedback control of minimum-phase nonlinear processes. *Chem. Eng. Sci.*, 47:837, 1992.
- [71] P. Daoutidis and C. Kravaris. Dynamic output feedback control of minimum-phase multivariable nonlinear processes. *Chem. Eng. Sci.*, 49:433–447, 1994.
- [72] M. De Paepe, P. D’Herdt, and D. Mertens. Micro-chp systems for residential applications. *Energy Conversion Management*, 47:3435–3446, 2006.
- [73] N. A. Diangelakis, S. Avraamidou, and E. N. Pistikopoulos. Decentralized multiparametric model predictive control for domestic combined heat and power systems. *Industrial & Engineering Chemistry Research*, 2015.
- [74] A. Dowling and L. Biegler. A framework for efficient large scale equation-oriented flowsheet optimization. *Comput. Chem. Eng.*, 72:3–20, 2015.
- [75] J. Du, J. Park, I. Harjunkoski, and M. Baldea. Integrated production scheduling and model predictive control for continuous process systems. *Submitted to Comput. Chem. Eng.*, 2014.

- [76] J. Du, J. Park, I. Harjunoski, and M. Baldea. A time scale bridging approach for integrating production scheduling and process control. *Submitted to Comput. Chem. Eng.*, 2014.
- [77] J. Du, J. Park, I. Harjunoski, and M. Baldea. Time scale bridging approaches for integration of production scheduling and process control. *Comput. Chem. Eng.*, 79:59–69, 2015.
- [78] J. Du, J. Park, I. Harjunoski, and M. Baldea. Time scale bridging approaches for integration of production scheduling and process control. *Comput. Chem. Eng.*, 79:59–69, 2015.
- [79] R. Dunia and S. Joe Qin. Subspace approach to multidimensional fault identification and reconstruction. *AIChE Journal*, 44(8):1813–1831, 1998.
- [80] B. Dunn, H. Kamath, and J. Tarascon. Electrical energy storage for the grid: A battery of choices. *Science*, 334:928 – 935, 2011.
- [81] B. Dunn, H. Kamath, and J.-M. Tarascon. Electrical energy storage for the grid: a battery of choices. *Science*, 334(6058):928–935, 2011.
- [82] M. Duran and I. Grossmann. An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Mathematical Programming*, 36(3):307–339, 1986.
- [83] R. C. Duthu, D. Zimmerle, T. H. Bradley, and M. J. Callahan. Evaluation of existing customer-owned, on-site distributed generation business models. *The Electricity Journal*, 27(1):42–52, 2014.
- [84] Y. Dutil, D. R. Rousse, N. B. Salah, S. Lassue, and L. Zalewski. A review on phase-change materials: mathematical modeling and simulations. *Renewable and Sustainable Energy Reviews*, 15(1):112–130, 2011.
- [85] B. Eisenhower, Z. O’Neill, S. Narayanan, V. Fonoberov, and I. Mezic. A methodology for meta-model based optimization in building energy models. *Energy and Buildings*, 47:292–301, 2012.
- [86] W. El-Khattam and M. Salama. Distributed generation technologies, definitions and benefits. *Electric power systems research*, 71(2):119–128, 2004.

- [87] Electricity Reliability Council of Texas (ERCOT). Backcasted (actual) load profiles – historical (accessed feb. 19, 2016), 2005.
- [88] Electricity Reliability Council of Texas (<http://www.ercot.com/mktinfo/>).
- [89] M. Ellis, H. Durand, and P. Christofides. A tutorial review of economic model predictive control methods. *J. Proc. Contr.*, 28:1156–1178, 2014.
- [90] M. Ellis, M. Heidarinejad, and P. D. Christofides. Economic model predictive control of nonlinear singularly perturbed systems. *Journal of Process Control*, 23(5):743–754, 2013.
- [91] J. Emsley. Oxygen. nature’s building blocks: An az guide to the elements, 2001.
- [92] B. C. Erdener, K. A. Pambour, R. B. Lavin, and B. Dengiz. An integrated simulation model for analysing electricity and gas systems. *Int. J. Elec. Power & Energ. Sys.*, 61:410–420, 2014.
- [93] H. Farhangi. The path of the smart grid. *Power and energy magazine, IEEE*, 8(1):18–28, 2010.
- [94] C. A. Farschman, K. P. Viswanath, and B. E. Ydstie. Process systems and inventory control. *AIChE J.*, 44(8):1841–1857, 1998.
- [95] N. Fenichel. Geometric singular perturbation theory for ordinary differential equations. *J. Diff. Equat.*, 31:53, 1979.
- [96] A. Flores-Tlacuahuac and I. E. Grossman. Simultaneous cyclic scheduling and control of a multiproduct CSTR. *Ind. Eng. Chem. Res.*, 45:6698–6712, 2006.
- [97] M. Forstmeier, F. Mannerheim, F. J. d’Amato, M. Shah, Y. Liu, M. Baldea, and A. Stella. Feasibility study on wind-powered desalination. *Desalination*, 203:463–470, 2007.
- [98] P. S. Georgilakis and N. D. Hatziargyriou. Optimal distributed generation placement in power distribution networks: models, methods, and future research. *Power systems, IEEE transactions on*, 28(3):3420–3428, 2013.
- [99] G. Ghatikar and R. Bienert. Smart grid standards and systems interoperability: a precedent with openadr. *proceedings of the Grid Interop Forum*, pages 1–7, 2011.

- [100] S. Ghosh, S. P. Ghoshal, and S. Ghosh. Optimal sizing and placement of distributed generation in a network system. *International Journal of Electrical Power & Energy Systems*, 32(8):849–856, 2010.
- [101] G. Giannakoudis, A. Papadopoulos, P. Seferlis, and S. Voutetakis. Optimum design and operation under uncertainty of power systems using renewable energy sources and hydrogen storage. *International Journal of Hydrogen Energy*, 35:872–891, 2010.
- [102] S. Ginestet, T. Bouache, K. Limam, and G. Lindner. Thermal identification of building multilayer walls using reflective Newton algorithm applied to quadrupole modelling. *Energy and Buildings*, 60:139–145, May 2013.
- [103] M. Gouda, S. Danaher, and C. Underwood. Building thermal model reduction using nonlinear constrained optimization. *Building and Environment*, 37(12):1255–1265, 2002.
- [104] S. Goyal and P. Barooah. A method for model-reduction of non-linear thermal dynamics of multi-zone buildings. *Energy and Buildings*, 47:332–340, 2012.
- [105] D. Green and R. Perry. *Perry’s Chemical Engineers’ Handbook, Eighth Edition*. Chemical Engineers Handbook. Mcgraw-hill, 2007.
- [106] I. E. Grossmann. Enterprise-wide optimization: A new frontier in process systems engineering. *AIChE J.*, 51:1846–1857, 2005.
- [107] W. Gu, Z. Wu, R. Bo, W. Liu, G. Zhou, W. Chen, and Z. Wu. Modeling, planning and optimal energy management of combined cooling, heating and power microgrid: A review. *International Journal of Electrical Power & Energy Systems*, 54:26–37, 2014.
- [108] K. Guemghar, B. Srinivasan, P. Mullhaupt, and D. Bonvin. Analysis of cascade structure with predictive control and feedback linearisation. *IEE Proceedings-Control Theory and Applications*, 152:317, 2005.
- [109] M. Gwerder, B. Lehmann, J. Tödttli, V. Dorer, and F. Renggli. Control of thermally-activated building systems (TABS). *Applied Energy*, 85(7):565–581, July 2008.

- [110] J. Hahn and T. Edgar. An improved method for nonlinear model reduction using balancing of empirical gramians. *Comput. Chem. Eng.*, 26(10):1379–1397, 2002.
- [111] I. Harjunoski, C. Maravelias, P. Bongers, P. Castro, S. Engell, I. Grossmann, J. Hooker, C. Méndez, G. Sand, and J. Wassick. Scope for industrial applications of production scheduling models and solution methods. *Comput. Chem. Eng.*, 62:161–193, 2014.
- [112] I. Harjunoski, R. Nyström, and A. Horch. Integration of scheduling and control - theory or practice? *Comput. Chem. Eng.*, 33(12):1909–1918, 2009.
- [113] A. Harmens. Vapour-liquid equilibrium N_2 - A_2 - O_2 for lower argon concentrations. *Cryogenics*, 10(5):406–409, 1970.
- [114] M. Heidarinejad, J. Liu, and P. Christofides. Economic model predictive control of nonlinear process systems using lyapunov techniques. *AIChE J.*, 58:855–870, 2011.
- [115] G. P. Henze. Model predictive control for buildings: a quantum leap? *Journal of Building Performance Simulation*, 6(3):157–168, 2013.
- [116] G. P. Henze, B. Biggar, D. Kohn, and M. P. Becker. Optimal design and operation of a thermal storage system for a chilled water plant serving pharmaceutical buildings. *Energy and Buildings*, 40:1004–1019, 2008.
- [117] G. P. Henze, C. Felsmann, and G. Knabe. Evaluation of optimal control for active and passive building thermal storage. *Int. J. Therm. Sci.*, 43(2):173–183, 2004.
- [118] P. J. Hibbard and T. Schatzki. The interdependence of electricity and natural gas: Current factors and future prospects. *The Elec. J.*, 25:6–17, 2012.
- [119] R. Huang, V. M. Zavala, and L. T. Biegler. Advanced step nonlinear model predictive control for air separation units. *Journal of Process Control*, 19(4):678–685, 2009.
- [120] C. Hui. Optimizing chemical processes with discontinuous function - a novel formulation. *Comput. Chem. Eng.*, pages S479–S482, 1999.
- [121] M. Ierapetritou, D. Wu, J. Vin, P. Sweeney, and M. Chigirinskiy. Cost minimization in an energy-intensive plant using mathematical programming approaches. *Ind. Eng. Chem. Res.*, 41(21):5262–5277, 2002.

- [122] A. Jadbabaie, J. Yu, and J. Hauser. Unconstrained receding-horizon control of nonlinear systems. *IEEE Trans. Automat. Contr.*, 46(5):776–783, 2001.
- [123] James J. Hirsch & Associates and Lawrence Berkeley National Laboratory. eQUEST: the QUick Energy Simulation Tool.
- [124] P. Jaramillo, W. M. Griffin, and H. S. Matthews. Comparative life-cycle air emissions of coal, domestic natural gas, lng, and sng for electricity generation. *Environmental Science & Technology*, 41(17):6290–6296, 2007.
- [125] H. Jiayi, J. Chuanwen, and X. Rong. A review on distributed energy resources and microgrid. *Renewable and Sustainable Energy Reviews*, 12(9):2472–2483, 2008.
- [126] Y. Jin, J. Li, W. Du, and F. Qian. Integrated operation and cyclic scheduling optimization for an ethylene cracking furnaces system. *Ind. Eng. Chem. Res.*, 54(15):3844–3854, 2015.
- [127] S. Jogwar, M. Baldea, and P. Daoutidis. Dynamics and control of process networks with large energy recycle. *Ind. Eng. Chem. Res.*, 48(13):6087–6097, 2009.
- [128] S. Jogwar, M. Baldea, and P. Daoutidis. Dynamics and control of process networks with large energy recycle. *Ind. Eng. Chem. Res.*, 48:6087–6097, 2009.
- [129] S. Jogwar, M. Baldea, and P. Daoutidis. Tight energy integration: Dynamic impact and control advantages. *Comput. Chem. Eng.*, 34:1457–1466, 2010.
- [130] T. Johansson. Integrated scheduling and control of an air separation unit subject to time-varying electricity prices. *Master’s Thesis, KTH Royal Institute of Technology*, 2015.
- [131] G. Joos, B. Ooi, D. McGillis, F. Galiana, and R. Marceau. The potential of distributed generation to provide ancillary services. In *Power Engineering Society Summer Meeting, 2000. IEEE*, volume 3, pages 1762–1767. IEEE, 2000.
- [132] R. S. Kamath, L. T. Begler, and I. Grossmann. Modeling multistream heat exchangers with and without phase changes for simultaneous optimization and heat integration. *AIChE J.*, 28(1):190–204, 2012.
- [133] M. Kano and K. Fujiwara. Virtual sensing technology in process industries: trends and challenges revealed by recent industrial applications. *Journal of chemical engineering of Japan*, 46(1):1–17, 2013.

- [134] M. Karwan and M. Kebli. Operations planning with real time pricing of a primary input. *Computers & operations research*, 34(3):848–867, 2007.
- [135] K. R. Keeney and J. E. Braun. Application of building precooling to reduce peak cooling requirements. *ASHRAE Transactions*, 103(1):463–469, 1997.
- [136] K. Keesman, D. Peters, and L. Lukasse. Optimal climate control of a storage facility using local weather forecasts. *Contr. Eng. Prac.*, 11(5):505–516, 2003.
- [137] H. K. Khalil. *Nonlinear Systems*. Prentice-Hall, Upper Saddle River, NJ, third edition, 2002.
- [138] Y. Kikkawa, M. Nakamura, and S. Sugiyama. Development of liquefaction process for natural gas. *Journal of Chemical Engineering of Japan*, 30(4):625–630, 1997.
- [139] H. Kim, Y.-J. Kim, K. Yang, and M. Thottan. Cloud-based demand response for smart grid: Architecture and distributed algorithms. In *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*, pages 398–403. IEEE, 2011.
- [140] S. Kishore and L. V. Snyder. Control mechanisms for residential electricity demand in smartgrids. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 443–448. IEEE, 2010.
- [141] Klein, S. A., et. al, Solar Energy Laboratory, University of Wisconsin, Madison, USA. TRNSYS: A Transient System Simulation Program.
- [142] P. V. Kokotovic, H. K. Khalil, and J. O’Reilly. *Singular Perturbations in Control: Analysis and Design*. Academic Press, London, 1986.
- [143] G. M. Kopanos and E. N. Pistikopoulos. Reactive scheduling by a multiparametric programming rolling horizon framework: a case of a network of combined heat and power units. *Ind. Eng. Chem. Res.*, 53:4366–4386, 2014.
- [144] M. Krarti. An Overview of Artificial Intelligence-Based Methods for Building Energy Systems. *J. Solar Energ. Eng.*, 125(3):331–342, 2003.
- [145] C. Kravaris, V. Sotiropoulos, C. Georgiou, N. Kazantzis, M. Xiao, and A. J. Krener. Nonlinear observer design for state and disturbance estimation. *Systems & Control Letters*, 56(11):730–735, 2007.

- [146] A. Kumar, P. D. Christofides, and P. Daoutidis. Singular perturbation modeling of nonlinear processes with non-explicit time-scale separation. *Chem. Eng. Sci.*, 53:1491–1504, 1998.
- [147] A. Kumar and P. Daoutidis. *Control of Nonlinear Differential Equation Systems*, volume 397 of *Research Notes in Mathematics Series*. Chapman & Hall/CRC, 1999.
- [148] S. Lakshminarayanan, K. Onodera, and G. Madhukar. Recycle effect index: A measure to aid in control system design for recycle processes. *Ind. Eng. Chem. Res.*, 43:1499–1511, 2004.
- [149] Y.-d. Lang, A. Malacina, L. T. Biegler, S. Munteanu, J. I. Madsen, and S. E. Zitney. Reduced order model based on principal component analysis for process simulation and optimization. *Energy & Fuels*, 23(3):1695–1706, 2009.
- [150] A. Lee, O. Zinaman, J. Logan, M. Bazilian, D. Arent, and R. L. Newmark. Interactions, complementarities and tensions at the nexus of natural gas and renewable energy. *The Elec. J.*, 25:38–48, 2012.
- [151] S. Lee and I. E. Grossmann. Logic-based modeling and solution of nonlinear discrete/continuous optimization problems. *Annals of Operations Research*, 139:267–288, 2005.
- [152] A. Lefort, R. Bourdais, G. Ansanay-Alex, and H. Guguen. Hierarchical control method applied to energy management of a residential house. *Energy and Buildings*, 64:53–61, 2013.
- [153] G. Li, S. J. Qin, Y. Ji, and D. Zhou. Reconstruction based fault prognosis for continuous processes. *Control Engineering Practice*, 18(10):1211–1219, 2010.
- [154] N. Li, L. Chen, and S. H. Low. Optimal demand response based on utility maximization in power networks. *Proc. of IEEE Power and Energy Society General Meeting*, 2011.
- [155] T. Li, M. Eremia, and M. Shahidehpour. Interdependency of natural gas network and power system security. *IEEE Trans. Power Sys.*, 23:1817–1824, 2008.
- [156] Z. Li and M. G. Ierapetritou. Process scheduling under uncertainty: reviews and challenges. *Comput. Chem. Eng.*, 32(4):715–727, 2008.

- [157] B. Linnhoff and E. Hindmarsh. The pinch design method for heat exchanger networks. *Chem. Eng. Sci.*, 38:745–763, 1983.
- [158] C. Liu, M. Shahidehpour, and J. Wang. Coordinated scheduling of electricity and natural gas infrastructures with a transient model for natural gas flow. *Chaos*, 21, 2011.
- [159] J. Liu, W. Li, J. Liu, and B. Wang. Efficiency of energy recovery ventilator with various weathers and its energy saving performance in a residential apartment. *Energy and Buildings*, 42(1):43–49, Jan. 2010.
- [160] L. Ljung. System identification: Theory for the user, second edition. *PTR Prentice Hall Information and System Sciences Series*, 1999.
- [161] J. Logan, A. Lopez, T. Mai, C. Davidson, M. Bazilian, and D. Arent. Natural gas scenarios in the u.s. power sector. *Energy Economics*, 40:183–195, 2013.
- [162] J. Ma, S. Qin, T. Salsbury, and P. Xu. Demand reduction in building energy systems based on economic model predictive control. *Chem. Eng. Sci.*, 67(1):92–100, 2012.
- [163] J. Ma, S. J. Qin, and T. Salsbury. Model predictive control of building energy systems with balanced model reduction. *Proc. of Amer. Contr. Conf.*, pages 3681–3686, 2012.
- [164] Y. Ma, G. Anderson, and F. Borrelli. A distributed predictive control approach to building temperature regulation. *Proc. of Amer. Contr. Conf.*, pages 2089–2094, 2011.
- [165] Y. Ma, F. Borrelli, B. Hancey, B. Coffey, S. Bengea, and P. Haves. Model predictive control for the operation of building cooling systems. *IEEE Transactions on Control Systems Technology*, 20:796 – 803, 2012.
- [166] Y. Ma, F. Borrelli, B. Hancey, B. Coffey, S. Bengea, A. Packard, M. Wetter, and P. Haves. Model predictive control for the operation of building cooling systems. *IEEE Trans. Cont. Syst. Technol.*, 20(3):796–803, 2012.
- [167] Y. Ma, F. Borrelli, B. Hancey, A. Packard, and S. Bortoff. Model Predictive Control of Thermal Energy Storage in Building Cooling Systems. In *Proceedings of the 48th IEEE Conference on Decision and Control*, pages 392–397, 2009.

- [168] Y. Ma, A. Kelman, and F. Daly, A. Borrelli. Predictive control for energy efficient buildings with thermal storage. *IEEE Control Systems Magazine*, pages 44–64, 2012.
- [169] M. Maasoumy, M. Razmara, M. Shahbakhti, and A. Sangiovanni Vincentelli. Handling model uncertainty in model predictive control for energy efficient buildings. *Energy and Buildings*, 77:377–392, 2014.
- [170] J. MacGregor and A. Cinar. Monitoring, fault diagnosis, fault-tolerant control, and optimization: Data driven methods. *Comput. Chem. Eng.*, 47:111–120, 2012.
- [171] P. Mago, L. Chamra, and A. Hueffed. A review on energy, economical, and environmental benefits of the use of chp systems for small commercial buildings for the north american climate. *International Journal of Energy Research*, 33(14):1252–1265, 2009.
- [172] F. Maneti. Considerations on nonlinear model predictive control techniques. *Comput. Chem. Eng.*, 35:2491–2509, 2011.
- [173] C. T. Maravelias. General framework and modeling approach classification for chemical production scheduling. *AIChE J.*, 58(6):1812–1828, 2012.
- [174] C. T. Maravelias. General framework and modeling approach classification for chemical production scheduling. *AIChE J.*, 58(6):1812–1828, 2012.
- [175] C. T. Maravelias and C. Sung. Integration of production planning and scheduling: Overview, challenges and opportunities. *Comput. Chem. Eng.*, 33(12):1919–1930, 2009.
- [176] R. Marino and P. Kokotovic. A geometric approach to nonlinear singular perturbed control systems. *Automatica*, 24:31–41, 1988.
- [177] MathWorks, Inc. MATLAB R2014. <http://www.mathworks.com/products/matlab/>, 1994 - 2015.
- [178] D. Mayne, J. Rawlings, C. Rao, and P. Scokaert. Constrained model predictive control: Stability and optimality. *Automatica*, 36:789–814, 2000.
- [179] D. Q. Mayne, J. Rawlings, C. V. Rao, and P. O. M. Scokaert. Constrained model predictive control: Stability and optimality. *Automatica*, 36:789–814, 2000.

- [180] S. A. Memon. Phase change materials integrated in building walls: A state of the art review. *Renewable and Sustainable Energy Reviews*, 31:870–906, 2014.
- [181] D. Mendoza-Serrano and D. J. Chmielewski. HVAC control using infinite-horizon economic MPC. *51st IEEE Conference on Decision and Control*, pages 6963–6968, Dec. 2012.
- [182] J. Miller, W. Luyben, and S. Blouin. Economic incentive for intermittent operation of air separation plants with variable power costs. *Ind. Eng. Chem. Res.*, 47(4):1132–1139, 2008.
- [183] S. Mitra, I. E. Grossmann, J. M. Pinto, and N. Arora. Optimal production planning under time-sensitive electricity prices for continuous power-intensive processes. *Comput. Chem. Eng.*, 38:171–184, 2012.
- [184] P. Mizsey, N. Hau, N. Benko, I. Kalmar, and Z. Fonyo. Process control for energy integrated distillation schemes. *Comput. Chem. Eng.*, 22:427–434, 1998.
- [185] P.-D. Moroan, R. Bourdais, D. Dumur, and J. Buisson. Building temperature regulation using a distributed model predictive control. *Energy and Buildings*, 42(9):1445–1452, 2010.
- [186] P. Morosan, R. Bourdais, D. Dumur, and J. Buisson. A distributed MPC strategy based on Benders decomposition applied to multi-source multi-zone temperature regulation. *J. Proc. Contr.*, 21(5):729–737, 2011.
- [187] J. Morud and S. Skogestad. Dynamic behavior of integrated plants. *J. Proc. Contr.*, 6(2–3):145–156, 1996.
- [188] L. T. Narraway and J. D. Perkins. Selection of process control structure based on linear economic dynamics. *Ind. Eng. Chem. Res.*, 32(1):2681, 1993.
- [189] A. Neto and F. Fiorelli. Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. *Energy and Buildings*, 40(12):2169–2176, 2008.
- [190] T. A. Nguyen and M. Aiello. Energy intelligent buildings based on user activity: A survey. *Energy and Buildings*, 56:244–257, Jan. 2013.

- [191] Y. Nie, L. T. Biegler, C. M. Villa, and J. M. Wassick. Reactor modeling and recipe optimization of polyether polyol processes: Polypropylene glycol. *AIChE Journal*, 59(7):2515–2529, 2013.
- [192] Y. Nie, L. T. Biegler, and J. M. Wassick. Integrated scheduling and dynamic optimization of batch processes using state equipment networks. *AIChE J.*, 58(11):3416–3432, 2012.
- [193] Y. Nie, L. T. Biegler, J. M. Wassick, and C. M. Villa. Extended discrete-time resource task network formulation for the reactive scheduling of a mixed batch/continuous process. *Ind. Eng. Chem. Res.*, 53:17112 – 17123, 2014.
- [194] J. Nocedal and S. Wright. *Numerical optimization*. Springer Science & Business Media, 2006.
- [195] F. Oldewurtel, A. Parisio, C. Jones, D. Gyalistras, M. Gwerder, V. Stauch, B. Lehmann, and M. Morari. Use of model predictive control and weather forecasts for energy efficient building climate control. *Energy and Buildings*, 45:15–27, 2012.
- [196] F. Oldewurtel, A. Parisio, C. Jones, M. Morari, D. Gyalistras, M. Gwerder, V. Stauch, B. Lehmann, and K. Wirth. Energy efficient building climate control using stochastic model predictive control and weather predictions. *Proc. of Amer. Contr. Conf.*, pages 5100–5105, 2010.
- [197] F. Oldewurtel, A. Ulbig, A. Parisio, G. Andersson, and M. Morari. Reducing peak electricity demand in building climate control using real-time pricing and model predictive control. *49th IEEE Conference on Decision and Control*, pages 1927–1932, 2010.
- [198] A. D. Ondeck, T. F. Edgar, and M. Baldea. Optimal operation of a residential district-level combined photovoltaic/natural gas power and cooling system. *Applied Energy*, 156:593–606, 2015.
- [199] D. Ouelhadj and S. Petrovic. A survey of dynamic scheduling in manufacturing systems. *Journal of Scheduling*, 12(4):417–431, 2009.
- [200] M. Ouyang. Review on modeling and simulation of interdependent critical infrastructure systems. *Reliability Engineering and System Safety*, 121:43–60, 2014.

- [201] T. Overbye and J. D. Weber. Visualizing the electric grid. *Spectrum, IEEE*, 38(2):52–58, 2001.
- [202] P. Palensky and D. Dietrich. Demand side management: Demand response, intelligent energy systems, and smart loads. *Industrial Informatics, IEEE Transactions on*, 7(3):381–388, 2011.
- [203] G. Pannocchia and A. Bemporad. Combined design of disturbance model and observer for offset-free model predictive control. *Automatic Control, IEEE Transactions on*, 52(6):1048–1053, 2007.
- [204] G. Pannocchia and J. B. Rawlings. Disturbance models for offset-free model-predictive control. *AIChE J.*, 49(2):426–437, 2003.
- [205] G. Pannocchia and J. B. Rawlings. Disturbance models for offset-free model-predictive control. *AIChE J.*, 49(2):426–437, 2003.
- [206] R. Parameshwaran, S. Harikrishnan, and S. Kalaiselvam. Energy efficient PCM-based variable air volume air conditioning system for modern buildings. *Energy and Buildings*, 42(8):1353–1360, 2010.
- [207] H. Park and R. Baldick. Transmission planning under uncertainties of wind and load: sequential approximation approach. *Power Systems, IEEE Transactions on*, 28(3):2395–2402, 2013.
- [208] K. Patan and J. Korbicz. Fault detection in catalytic cracking converter by means of probability density approximation. *Engineering Applications of Artificial Intelligence*, 20(7):912–923, 2007.
- [209] R. Pattison, C. Touretzky, T. Johansson, I. Harjunoski, and M. Baldea. Moving horizon scheduling of an air separation unit under fast-changing energy proces. *DYCOPS 2016*, page Accepted, 2016.
- [210] R. Pattison, C. Touretzky, T. Johansson, I. Harjunoski, and M. Baldea. Production scheduling with low-order dynamic models: Framework and an air separation application. *Ind. Eng. Chem. Res.*, page Accepted, 2016.
- [211] R. C. Pattison and M. Baldea. Optimal design of air separation plants with variable electricity pricing. *Foundations of Computer-Aided Process Design (FOCAPD)*, pages 393–398, 2014.

- [212] M. L. Pinedo. *Scheduling: Theory, Algorithm, Systems*. 2008.
- [213] M. Pipattanasomporn, M. Willingham, and S. Rahman. Implications of on-site distributed generation for commercial/industrial facilities. *Power Systems, IEEE Transactions on*, 20(1):206–212, 2005.
- [214] A. Prata, J. Oldenburg, A. Kroll, and W. Marquardt. Integrated scheduling and dynamic optimization of grade transitions for a continuous polymerization reactor. *Comput. Chem. Eng.*, 32(3):463–476, 2008.
- [215] S. Prívará, Z. Váa, D. Gyalistras, J. Cigler, C. Sagerschnig, M. Morari, and L. Ferkl. Modeling and Identification of a Large Multi-Zone Office Building. *IEEE Multi-Conference on Systems and Control*, pages 55–60, 2011.
- [216] S. Prívará, Z. Váa, E. Žáčková, and J. Cigler. Building modeling: Selection of the most appropriate model for predictive control. *Energy and Buildings*, 55:341–350, Dec. 2012.
- [217] Process Systems Enterprise, gPROMS ModelBuilder (<http://www.psenderprise.com/gproms.html>).
- [218] S. Qin and T. Badgwell. A survey of industrial model predictive control technology. *Contr. Eng. Prac.*, 11(7):733–764, 2003.
- [219] S. J. Qin. Statistical process monitoring: basics and beyond. *Journal of chemometrics*, 17(8-9):480–502, 2003.
- [220] S. J. Qin. Survey on data-driven industrial process monitoring and diagnosis. *Annual Reviews in Control*, 36(2):220–234, 2012.
- [221] S. J. Qin and T. A. Badgwell. A survey of industrial model predictive control technology. *Control engineering practice*, 11(7):733–764, 2003.
- [222] S. J. Qin and T. A. Badgwell. A survey of industrial model predictive control technology. *Control engineering practice*, 11:733–764, 2003.
- [223] A. Radke and Z. Gao. A survey of state and disturbance observers for practitioners. In *American Control Conference, 2006*, pages 6–pp. IEEE, 2006.
- [224] A. P. Ramallo-González, M. E. Eames, and D. A. Coley. Lumped parameter models for building thermal modelling: An analytic approach to simplifying complex multi-layered constructions. *Energy and Buildings*, 60:174–184, May 2013.

- [225] M. Rasouli, C. J. Simonson, and R. W. Besant. Applicability and optimum control strategy of energy recovery ventilators in different climatic conditions. *Energy and Buildings*, 42(9):1376–1385, Sept. 2010.
- [226] J. Rawlings, D. Bonn e, J. J orgensen, A. Venkat, and S. J orgensen. Unreachable Setpoints in Model Predictive Control. *IEEE Trans. Automat. Contr.*, 53(9):2209–2215, 2008.
- [227] J. B. Rawlings. Tutorial overview of model predictive control. *IEEE Control Systems*, 20(3):38–52, 2000.
- [228] F. Reyes and W. Luyben. Steady-state and dynamic effects of design alternatives in heat-exchanger/ furnace/ reactor processes. *Ind. Eng. Chem. Res.*, 39:3335–3346, 2000.
- [229] R. Z. R os-Mercado and C. Borraz-S anchez. Optimization problems in natural gas transportation systems: A state-of-the-art review. *Applied Energy*, 147:536–555, 2015.
- [230] J. P. Ruiz and I. E. Grossmann. Generalized disjunctive programming: A framework for formulation and alternative algorithms for minlp optimization. *The IMA Volumes in Mathematics and its Applications*, 24:93–115, 2012.
- [231] I. Sabuncuoglu and S. Goren. Hedging production schedules against uncertainty in manufacturing environment with a review of robustness and stability research. 41(17):4211–4231, 2003.
- [232] M. Salimi, H. Ghasemi, M. Adelpour, and S. Vaez-ZAdeh. Optimal planning of energy hubs in interconnected energy systems: a case study for natural gas and electricity. *Generation, Transmission & Distribution, IET*, 9(8):695–707, 2015.
- [233] T. Salsbury. A Survey of Control Technologies in the Building Automation Industry. *16th IFAC World Congress*, pages 1396–1407, 2005.
- [234] T. Salsbury, P. Mhaskar, and S. J. Qin. Predictive control methods to improve energy efficiency and reduce demand in buildings. *Computers & Chemical Engineering*, 51:77–85, Apr. 2013.
- [235] R. Scattolini. Architectures for distributed and hierarchical Model Predictive Control A review. *J. Proc. Contr.*, 19(5):723–731, 2009.

- [236] D. E. Seborg, T. F. Edgar, D. A. Mellichamp, and F. J. Doyle III. *Process dynamics and control*. Wiley, 2010.
- [237] A. Sharma, V. V. Tyagi, and D. Chen, C. R. Buddhi. Review on thermal energy storage with phase change materials and applications. *Renewable and Sustainable Energy Reviews*, 13:318–345, 2009.
- [238] D. Shobrys and D. White. Planning, scheduling and control systems: why cannot they work together. *Comput. Chem. Eng.*, 26(2):149–160, 2002.
- [239] D. E. Shobrys and D. C. White. Planning, scheduling and control systems: why cannot they work together. *Comput. Chem. Eng.*, 26(2):149 – 160, 2002.
- [240] J. Sirola and T. Edgar. Process energy systems: Control, economic, and sustainability objectives. *Comput. Chem. Eng.*, 47:134–144, Dec. 2012.
- [241] R. Sinnott and G. Towler. *Chemical Engineering Design*. Elsevier, 5th edition, 2009.
- [242] S. Skogestad. Plantwide control: the search for the self-optimizing control structure. *J. Proc. Contr.*, 10(5):487–507, 2000.
- [243] N. Soares, J. J. Costa, A. R. Gaspar, and P. Santos. Review of passive PCM latent heat thermal energy storage systems towards buildings’ energy efficiency. *Energy and Buildings*, 59:82–103, 2013.
- [244] O. A. Sotomayor and D. Odloak. Observer-based fault diagnosis in chemical plants. *Chemical Engineering Journal*, 112(1):93–108, 2005.
- [245] M. Sourbron, R. De Herdt, T. Van Reet, W. Van Passel, M. Baelmans, and L. Helsen. Efficiently produced heat and cold is squandered by inappropriate control strategies: A case study. *Energy and Buildings*, 41(10):1091–1098, Oct. 2009.
- [246] N. Strachan and A. Farrell. Emissions from distributed vs. centralized generation: The importance of system performance. *Energy Policy*, 34:2677–2689, 2005.
- [247] K. Subramanian, C. T. Maravelias, and J. B. Rawlings. A state-space model for chemical production scheduling. *Comput. Chem. Eng.*, 47:97–110, 2012.

- [248] Y. Sun, S. Wang, F. Xiao, and D. Gao. Peak load shifting control using different cold thermal energy storage facilities in commercial buildings: A review. *Energy Convers. Manage.*, 71:101–114, 2013.
- [249] N. H. S. Tay, M. Belusko, and F. Bruno. Experimental investigation of tubes in a phase change thermal energy storage system. *Applied Energy*, 90:288–297, 2012.
- [250] S. Terrazas-Moreno, A. Flores-Tlacuahuac, and I. Grossmann. Simultaneous cyclic scheduling and optimal control of polymerization reactors. *AIChE J.*, 53(9):2301–2315, 2007.
- [251] S. Terrazas-Moreno, A. Flores-Tlacuahuac, and I. Grossmann. Lagrangean heuristic for the scheduling and control of polymerization reactors. *AIChE J.*, 54(1):163–182, 2008.
- [252] C. Touretzky and M. Baldea. Integrating scheduling and control for economic MPC of buildings with energy storage. *J. Proc. Contr. Economic MPC Special Issue*, 24(8):1292–1300, 2014.
- [253] C. Touretzky and M. Baldea. Nonlinear Model Reduction and Model Predictive Control of Residential Buildings with Energy Recovery. *J. Proc. Contr. Energy Efficient Buildings Special Issue*, 24(6):723–739, 2014.
- [254] C. R. Touretzky and M. Baldea. Model reduction and nonlinear MPC for energy management in buildings. *Proc. of Amer. Contr. Conf.*, pages 461–466, 2013.
- [255] C. R. Touretzky and M. Baldea. Integrating scheduling and control for economic MPC of buildings with energy storage. *J. Proc. Contr.*, 24:1292–1300, 2014.
- [256] C. R. Touretzky and M. Baldea. A hierarchical scheduling and control strategy for thermal energy storage systems. *Energy and Buildings*, 110:94–107, 2016.
- [257] C. R. Touretzky, I. Harjunkoski, and M. Baldea. A framework for integrating scheduling and control using discrete-time dynamic process models. *ESCAPE*, 2016.
- [258] C. R. Touretzky, D. McGuffin, J. Ziesmer, and M. Baldea. The effect of distributed electricity generation using natural gas on the electric and natural gas grid. *Applied Energy*, under review, 2016.

- [259] C. R. Touretzky, R. P. Pattison, J. Park, I. Harjunoski, and M. Baldea. Handling input dynamics in integrated scheduling and control. *Submitted to IEEE AQTR*, 2016.
- [260] C. R. Touretzky, A. A. Salliot, L. Lefevre, and M. Baldea. Optimal operation of phase-change thermal energy storage for a commercial building. *American Control Conference (ACC)*, pages 980–985, 2015.
- [261] A. Underwood. Simple formula to calculate mean temperature difference. *Chemical Engineering*, 77(13):192, 1970.
- [262] University of Texas at Austin School of Architecture. Thermal facade lab.
- [263] U.S. Department of Energy. EnergyPlus Energy Simulation Software.
- [264] US EIA. Manufacturing energy consumption survey: Total consumption of electricity, 2010.
- [265] U.S. Energy Information Administration. Household energy use in texas (accessed feb. 19, 2016), 2009.
- [266] U.S. Energy Information Administration. Annual energy review 2011, 2012.
- [267] U.S. Energy Information Administration. Annual energy outlook 2013, 2013.
- [268] U.S. Energy Information Administration. About U.S. natural gas pipelines, accessed Feb. 19, 2016.
- [269] U.S. Energy Information Administration. Electric generator dispatch depends on system demand and the relative cost of operation, accessed Feb. 19, 2016.
- [270] U.S. Energy Information Administration. Natural gas consumption by end use, accessed Feb. 19, 2016.
- [271] U.S. Energy Information Administration. Number of natural gas residential consumers in texas, accessed Feb. 19, 2016.
- [272] U.S. Energy Information Administration. State electricity profiles, table 10. supply and disposition of electricity, 1990-2013, accessed Feb. 19, 2016.

- [273] U.S. Energy Information Administration. Texas natural gas residential consumption, accessed Feb. 19, 2016.
- [274] E. van Henten and J. Bontsema. Singular perturbation methods applied to a variational problem in greenhouse climate control. *Proceedings of 31st Conference on Decision and Control*, pages 3068–3069, 1992.
- [275] V. Varma, G. Reklaitis, G. Blau, and J. F. Pekny. Enterprise-wide modeling & optimization—An overview of emerging research challenges and opportunities. *Comput. Chem. Eng.*, 31:692–711, 2007.
- [276] V. S. Vassiliadis, R. W. H. Sargent, and C. C. Pantelides. Solution of a class of multistage dynamic optimization problems. 1. problems without path constraints. *Ind. Eng. Chem. Res.*, 33(9):2111–2122, 1994.
- [277] V. S. Vassiliadis, R. W. H. Sargent, and C. C. Pantelides. Solutions of a class of multistage dynamic optimization problems part 1. *Ind. Eng. Chem. Res.*, 33:2111–2122, 1994.
- [278] V. S. Vassiliadis, R. W. H. Sargent, and C. C. Pantelides. Solutions of a class of multistage dynamic optimization problems part 2. *Ind. Eng. Chem. Res.*, 33:2123–2133, 1994.
- [279] P. M. Verderame, J. A. Elia, J. Li, and C. A. Floudas. Planning and scheduling under uncertainty: a review across multiple sectors. *Industrial & engineering chemistry research*, 49(9):3993–4017, 2010.
- [280] G. E. Vieira, J. W. Herrmann, and E. Lin. Rescheduling manufacturing systems: a framework of strategies, policies, and methods. *Journal of Scheduling*, 6:39–62, 2003.
- [281] D. Vinson. Air separation control technology. *Comput. Chem. Eng.*, 30(10-12):1436–1446, 2006.
- [282] D. Vinson. Air separation control technology. *Computers & Chemical Engineering*, 30(10):1436–1446, 2006.
- [283] R. Viral and D. Khatod. Optimal planning of distributed generation systems in distribution system: A review. *Renewable and Sustainable Energy Reviews*, 16(7):5146–5165, 2012.

- [284] J. Široký, F. Oldewurtel, J. Cigler, and S. Prívará. Experimental analysis of model predictive control for an energy efficient building heating system. *Applied Energy*, 88(9):3079–3087, 2011.
- [285] M. Wallace, B. Das, P. Mhaskar, J. House, and T. Salsbury. Offset-free model predictive control of a vapor compression cycle. *J. Proc. Contr.*, 22(7):1374–1386, Aug. 2012.
- [286] M. Wallace, R. McBride, S. Aumi, P. Mhaskar, J. House, and T. Salsbury. Energy efficient model predictive building temperature control. *Chem. Eng. Sci.*, 69(1):45–58, 2012.
- [287] S. Wang and M. Baldea. Autocovariance-based mpc model mismatch estimation for siso systems. In *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on*, pages 3032–3037. IEEE, 2015.
- [288] S. Wang and B. M. Autocovariance-based mpc model mismatch estimation for siso systems. *Proceedings of the IEEE Conference on Decision and Control*, 2015.
- [289] S. Wang and Z. Ma. Supervisory and Optimal Control of Building HVAC Systems: A Review. *HVAC&R Research*, 14(1):3–32, 2008.
- [290] A. Waqas and Z. U. Din. Phase change material (pcm) storage for free cooling in buildings - a review. *Renewable and Sustainable Energy Reviews*, 18:607–625, 2013.
- [291] M. C. Williams, J. P. Strakey, and S. C. Singhal. Us distributed generation fuel cell program. *Journal of Power Sources*, 131(1):79–85, 2004.
- [292] M. Wolsink. The research agenda on social acceptance of distributed generation in smart grids: Renewable as common pool resources. *Renewable and Sustainable Energy Reviews*, 16(1):822–835, 2012.
- [293] J.-y. Wu, J.-l. Wang, and S. Li. Multi-objective optimal operation strategy study of micro-cchp system. *Energy*, 48(1):472–483, 2012.
- [294] X. Xu, H. Jia, H.-D. Chiang, D. Yu, and D. Wang. Dynamic modeling and interaction of hybrid natural gas and electricity supply system in microgrid. *Power Systems, IEEE Transactions on*, 30(3):1212–1221, 2015.

- [295] L. Yang, H. Yan, and J. C. Lam. Thermal comfort and building energy consumption implications - a review. *Applied Energy*, 115(15):164–173, 2014.
- [296] C. Yaws. *The Yaws Handbook of Vapor Pressure: Antoine coefficients*. Elsevier Science, 2015.
- [297] R. Yin, P. Xu, M. Piette, and S. Kiliccote. Study on Auto-DR and pre-cooling of commercial buildings with thermal mass in california. *Energy and Buildings*, 42(7):967–975, 2010.
- [298] S. Yin, S. X. Ding, X. Xie, and H. Luo. A review on basic data-driven approaches for industrial process monitoring. *Industrial Electronics, IEEE Transactions on*, 61(11):6418–6428, 2014.
- [299] M. Yu, D. C. Miller, and L. T. Biegler. Dynamic reduced order models for simulating bubbling fluidized bed adsorbers. *Industrial & Engineering Chemistry Research*, 54(27):6959–6974, 2015.
- [300] M. Zachar and P. Daoutidis. Understanding and predicting the impact of location and load on microgrid design. *Energy*, 90:1005–1023, 2015.
- [301] R. Zamora and A. K. Srivastava. Controls for microgrids with storage: Review, challenges, and research needs. *Renewable and Sustainable Energy Reviews*, 14:2009–2018, 2010.
- [302] Q. Zhang, I. E. Grossmann, C. F. Heuberger, A. Sundaramoorthy, and J. M. Pinto. Air separation with cryogenic energy storage: Optimal scheduling considering electric energy and reserve markets. *AIChE J.*, 61(5):1547–1558, 2015.
- [303] D. Zhou, C. Y. Zhao, and Y. Tian. Review on thermal energy storage with phase change materials (PCMs) in building applications. *Applied Energy*, 92:593–605, 2012.
- [304] L. Zhu, Z. Chen, X. Chen, Z. Shao, and J. Qian. Simulation and optimization of cryogenic air separation units using a homotopy-based backtracking method. *Separation and Purification Technology*, 67(3):262–270, 2009.
- [305] N. Zhu, S. Wang, Z. Ma, and Y. Sun. Energy performance and optimal control of air-conditioned buildings with envelopes enhanced by phase change materials. *Energy Convers. Manage.*, 52:3197–3205, 2011.

- [306] Y. Zhu. Multivariable process identification for mpc: the asymptotic method and its applications. *J. Proc. Contr.*, 8(2):101–115, 1998.
- [307] Y. Zhu. *Multivariable system identification for process control*. Elsevier, Oxford, UK, 2001.
- [308] Y. Zhu, S. Legg, and C. Laird. Optimal design of cryogenic air separation columns under uncertainty. *Comput. Chem. Eng.*, 34:1377–1384, 2010.
- [309] J. Zhuge and M. Ierapetritou. Simultaneous scheduling and control with closed loop implementation on parallel units. *Proceedings Foundations of Computer-Aided Process Operations (FOCAPO)*, 2012.
- [310] J. Zhuge and M. G. Ierapetritou. Integrtaion of scheduling and control with closed loop implementation. *Ind. Eng. Chem. Res.*, 51(25):8550–8565, 2012.

Vita

Cara Rose Touretzky is from Bridgewater, New Jersey. She received the Honors Bachelor in Chemical Engineering degree, magna cum laude, from the University of Delaware in 2011, with minors in mathematics, chemistry, and economics. She entered the PhD program in the McKetta Department of Chemical Engineering at the University of Texas at Austin in the fall of 2011 and was a founding member of Dr. Baldea's Process and Energy Systems Engineering research group. She was partially funded through an EPA STAR Graduate Fellowship and Cockrell School of Engineering Fellowship. After graduating, Cara will join the Advanced Control Group at Exxon Mobil Research and Engineering.

Permanent address: CaraTour@gmail.com

This dissertation was typeset with L^AT_EX by the author.